Predicting Brand Company Tweet Performance Through Textual Analytics

Student: Pendleton Tompkins

DSA-5900-001: 4 Credit Hours

Spring 2021

Faculty Sponsor: Dr. Doyle Yoon

Table of Contents

Introduction

Social media has rapidly emerged as perhaps the most prominent medium through which companies may increase their brand awareness. Twitter, a social media platform garnering over 330 million monthly active users, serves as an ideal environment for companies to engage with consumers (Statista). However, unlike traditional advertising mediums (such as television, billboards, etc.), on Twitter, the size of the audience which ultimately sees a certain "tweet" (message) crafted by a company is related to how favorably the contents of the tweet are received by other users on Twitter. This is because tweets, initially, only appear in the home timeline, or "feed," of users following the account which posted the tweet. However, Twitter users in the initial audience can "retweet" the post, causing it to appear in the feed of any Twitter users following them. This implies that any given tweet may potentially propagate through a massive network of people, much larger than its initial audience.

However, for this process to occur, people must enjoy the contents of the tweet enough to motivate them to retweet it. This differs enormously from traditional advertising mediums. For example, regardless of the time or effort that a company puts into creating a TV commercial, if they book a slot during a Super Bowl football game, they're guaranteed to reach a massive audience. The size of the audience has no relationship with the contents of the commercial. Regardless of whether the commercial is exceptionally good or bad in quality, everyone watching the super bowl would also watch the commercial. In summary, conventional advertisement is purchased in advance with little chance for expansion, whereas social media advertisement is earned by the quality of its content, with the potential for enormous expansion.

Throughout the course of the past semester Dr. Yoon and I have worked towards creating computer programs capable of predicting the number of "likes," as well as the number of "retweets," which any tweet posted by verified "brand company" Twitter accounts will ultimately generate. Specifically, we have worked towards creating programs capable of generating these predictions prior to tweet publication, given only the contents of the unpublished tweet and a sample of the company's previous tweets. Successful completion of this goal would allow companies to determine, beforehand, whether or not a certain message is worthy of posting on Twitter. Having the capability to predict likes and retweets is essentially synonymous with having the means to predict the number of people who will have a positive response to the company's tweet. This capability could serve as a free, quick focus group for advertisement and branding campaigns performed through Twitter.

In turn, this capability could increase the speed at which a company's Twitter account grows, as posts reaching larger audiences are likely to yield more followers than posts reaching smaller audiences. Finally, the programs could be used to improve 'customer service response' type tweets, and as an additional avenue for gaining insights and feedback regarding customer satisfaction. Responding to questions, comments, and concerns posted online about a company is an important component of every company's official social media accounts. The programs could be useful in determining which formats of responses satisfy a customer's questions or comments, and which formats generally leave customers unsatisfied.

The implementation of this project not only has significance towards business objectives, it has significance towards research objectives as well. To accurately predict likes and retweets we first had to explore the question of "what are the factors which separate a popular tweet from an underperforming tweet?" Exploring this question has led to many valuable insights into human behavior in the online realm, as a number of sub-questions had to be addressed sufficiently before being able to even begin providing answers to the overarching, parent question. These sub-questions include, but are not limited to:

Does including media attachments (links, pictures, videos) influence tweet popularity? What effect does the sentiment (positive, negative, or neutral) of a tweet have on its popularity? Does the inclusion of a hashtag have an effect on tweet popularity? Are these separating factors uniform across all companies? Or, are they specific to either each company, or the industry to which the company belongs? Finally, can the same variables and models be used to predict both likes and retweets? Or, does the psychology behind someone 'liking' a tweet vs. 'retweeting' it differ enough such that separate models and variables must be used?

Objectives

The primary project objective was to determine the methods, variables and models that can most accurately predict the final number of likes and retweets of a brand company's tweet. This implied a sub-objective of first identifying the specific factors which are most closely related with a tweet's popularity. Our aim was to identify, or create, meaningful independent variables drawn through textual analysis. Furthermore, a secondary sub-objective was to determine the model-variable combination which generalizes best in predicting likes and retweets across tweets sourced from a variety of brand company Twitter accounts.

Providing a definitive answer to the secondary sub-objective proved to be much easier than doing so for the initial sub-objective. For the secondary sub-objective, all model-variable combinations were trained and tested on the exact same data. Numeric accuracy measures were reported, allowing for fair and direct comparison. In contrast, the initial sub-objective was always inherently associated with some level of “noise” because there could never be certainty that the only difference between the sets of tweets under examination was solely the variable of interest.

For example, after splitting a company’s tweets into two groups, one containing hashtags and one not, and after finding statistically significant differences between the two groups, one still may not definitively conclude that these differences are due to the presence of a hashtag. It’s entirely possible that an unidentified factor better explains these differences. Or, more likely, that a combination of many factors is what truly explains these differences. To account for this, we fall short of definitively concluding that causal relationships exist between the independent and dependent variables. Rather, we simply note that statistically significant differences exist between the groups. Afterwards we posit that, potentially, one underlying factor explaining these differences is the independent variable under consideration.

Data

Ingestion:

Data used for this project is comprised of tweets retrieved from ten separate brand company Twitter accounts. These ten companies are as follows: Amazon, BMW, Coca Cola, Disney, Google, McDonald’s, Mercedes-Benz, Microsoft, Samsung, and Toyota. Data was gathered directly from the source by utilizing ‘Orange3’ data mining software to scrape the Twitter API. Data was collected such that a sample of, approximately, each company’s past 3,000 tweets (as of December 1st, 2020) was available for analysis. Moreover, each company is associated with its own unique dataset, rather than storing data for all companies in a single file.

In total, 32,247 rows of data were collected across 18 variables. However, very few of the initially collected variables were made use of, as many are either blank (such as ‘location,’ ‘latitude,’ and ‘longitude’) or static fields (‘Author’, ‘Author Description,’ ‘Author Verified,’ ‘Author Name,’ etc.). Of the 18 initial variables, only the target variables (‘Number of Likes,’ ‘Number of Retweets’), the tweet

'Content' variable (textual contents of each tweet), as well as the 'Language' and 'In Reply To' variables (for sub-setting purposes) are absolutely required.

Exploration:

Upon initial viewing of Amazon data one observation became readily apparent; there are two distinct categories of tweets. The first category may be referred to as "official tweets" (abbreviated as OT). These are standalone tweets, which appear directly in the "feed" of every user following the posting company's account. Official tweets average a much higher number of likes and retweets than their counterparts, "in reply to tweets" (abbreviated as IRT). Tweets belonging to the IRT category occur when another Twitter user posts a question, comment, or concern related to a company, and the company Twitter account replies to address the issue. In a manner of speaking, these may be thought of as customer service tweets.

The key difference between these two tweet categories is the size of the initial audience. Official tweets are meant for anybody and everybody to see, as they automatically appear in the feed of every user following the account. IRT tweets, on the other hand, are geared more towards the individual level. These tweets only automatically appear in the feed of the user being replied to, as well as any users happening to "follow" both parties engaged in the conversation. Although this is a very important component of a brand company's Twitter account, relative to official tweets, IRT tweets average very few likes and retweets.

Making this observation did not require any level of expertise in the field of data science and is in no way a groundbreaking revelation. Rather, it would be more aptly categorized as being "obvious." However, the distinct tweet categories are important to note, as this observation affected the manner in which analysis was performed. For example, of the ten companies listed above, only four companies (Amazon, BMW, Mercedes-Benz, and Microsoft) were found to have a sufficient volume of tweets belonging to both categories to allow for separate analysis. Of the remaining six companies, three were found to only have enough IRT tweets for analysis (Coca Cola, Google, and McDonald's), while the remaining three (Disney, Samsung, and Toyota) were found to only have sufficient official tweets for analysis.

Table 1: Tweet Category Summary Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
Amazon OT	247	353.12	914.57	59.27	146.21
Amazon IRT	2927	3.44	41.40	0.57	3.67
BMW OT	806	1063.79	826.36	111.72	105.53
BMW IRT	2347	2.21	12.14	0.14	0.74
Coca Cola IRT	3192	21.27	1084.98	3.15	172.01
Disney OT	2741	3224.17	7974.21	726.10	3180.35
Google IRT	3036	6.99	359.90	2.16	114.97
McDonald's IRT	3210	7.27	63.02	0.13	1.00
Mercedes OT	589	1109.54	1076.01	117.85	234.32
Mercedes IRT	2538	1.46	20.26	0.21	3.56
Microsoft OT	1266	762.94	2555.89	129.52	438.76
Microsoft IRT	1391	54.06	444.47	1.99	17.18
Samsung OT	2924	446.35	1701.40	45.85	232.52
Toyota OT	2558	111.52	676.12	24.76	148.70

After unintentionally discovering the first observation of the exploration process, our team next desired to examine the relationship between likes and retweets. To do so, scatterplots were created with the number of retweets for each tweet along the x-axis, and the number of likes along the y-axis, as in the following two examples:

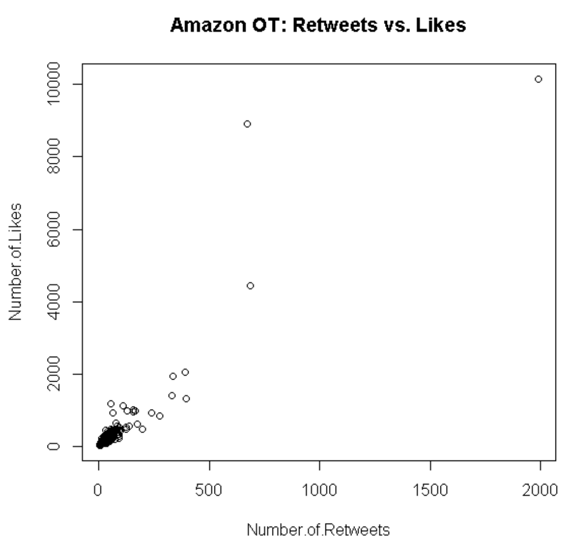


Figure 1: Scatterplot of Amazon OT Retweets vs. Likes

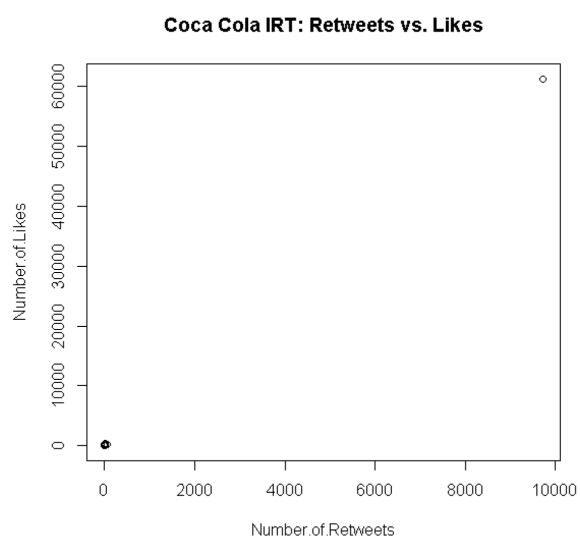


Figure 2: Scatterplot of Coca Cola IRT Retweets vs. Likes

The Coca Cola example was selected to showcase another discovery made through creating these figures, many of the datasets contain a few *extremely* large outliers. To save on space, similar figures for other companies will be placed in section 1 of the appendix. Visually, there appears to be a linear relationship between retweets and likes for both OT and IRT tweet categories. This appeared to remain true for all companies under consideration, although more so for some than others. However, this is nothing more than an observational remark.

In order to quantify the degrees of association, a statistical test was needed. Generally, calculating a correlation coefficient through use of Pearson's correlation test is an appropriate method. However, Pearson's is a parametric test. Meaning, Pearson's assumes that data are normally distributed. This will be discussed in detail at a later point, but collected data have severe issues of non-normality. Therefore, a non-parametric test of associativity was required, leading our team to Spearman's correlation test. This test makes no assumptions about the distributions of data. Calculating Spearman rank correlation (rho) values resulted in the following table:

Table 2: Correlation Measures of Retweets vs. Likes

Data	Likes: Shapiro's p	Retweets: Shapiro's p	Spearman's p	Spearman's Rho
Amazon OT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.84
Amazon IRT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.50
BMW OT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.96
BMW IRT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.30
Coca Cola IRT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.32
Disney OT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.98
Google IRT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.18
McDonald's IRT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.33
Mercedes OT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.92
Mercedes IRT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.30
Microsoft OT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.83
Microsoft IRT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.49
Samsung OT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.83
Toyota OT	< 2.2e-16	< 2.2e-16	< 2.2e-16	0.84

The null hypothesis of the Shapiro-Wilk test is that data are normally distributed, while the alternative hypothesis is that data are not normally distributed. Shapiro-Wilk tests resulting in p-values smaller than 0.05 imply that we may reject the null hypothesis and accept the alternative hypothesis of non-normality. Therefore, we may reject the null hypothesis, for both likes and retweets, across all companies and tweet categories – and conclude that data are distributed non-normally. This isn't a desirable quality but serves as justification for use of Spearman's correlation test.

The null hypothesis of Spearman's correlation test is a rho value of 0. In other words, the null hypothesis is that no monotonic relationship exists between the variables under consideration (likes and retweets), while the alternative hypothesis is that some monotonic relationship exists. In all cases, we may reject the null and conclude at least some degree of association exists between likes and retweets.

Interestingly, for all sets of official tweets there seems to be a strongly positive relationship between likes and retweets. However, for all sets of IRT tweets the relationship may only be characterized as moderately positive, or weakly positive in Google's case. It's possible that this is due to a larger volume of tied values existing within IRT tweet categories, but precautions were taken to ensure that a tie-corrected coefficient was returned from Spearman's test.

Still having the goal of identifying or creating meaningful independent variables through textual analysis, our team next decided to produce visualizations for wordcounts, the most commonly appearing words, as well as bigrams and trigrams for each company and tweet category. To do so, we first had to make some considerations about what specific data to examine. A majority of the datasets contain retweets made by their respective companies. Meaning, tweets from other users were collected as tweets from the company. Removing these retweets was simple, as all retweets - and only retweets - begin with "RT @". Furthermore, retweets were removed before producing all the tables and visualizations shown above as well, as they aren't technically company account tweets.

One major difference in data at this step, however, is that only tweets interpretable to an English-speaking audience were used in analysis. This implies that company tweets written in Spanish, Russian, German, etc. were removed from data. Only tweets whose 'language' field is marked either 'EN' (for English) or 'UND' (for undecided) were retained. Tweets whose language field is marked as 'UND' were retained because these tweets consist solely of some combination of '@username' mentions, emojis, and links, all of which are interpretable to an English reader.

After narrowing down all data to only consist of tweets interpretable to an English audience, further considerations were still required. A vitally important step in textual analysis is textual pre-processing and standardization. For the following analysis, major steps in textual pre-processing and standardization include: removal of links, removal of '@username' mentions, removal of non-alphanumeric characters (emojis, etc.), stop-word removal, converting all text to lowercase, and removing excess whitespace created from these processes.

Further steps in textual standardization occurred as well, such as: possessive apostrophe removal ("Disney's" transforms to "Disney"), apostrophe removal ("you're" transforms to "youre"), replacing ampersands with the word 'and', replacing remaining '@' symbols with the word 'at', and removing commas followed by exactly 3 numeric digits. Apostrophe and comma removal had to be performed because tokenization (the process of converting a body of text into the individual word tokens which it's comprised of) does not handle punctuation in inputs very well. For example, the phrase "you're our 1,000th customer!" would tokenize to something along the lines of [you, re, our, 1, 000, customer], while the phrase "youre our 1000th customer" would tokenize to [youre, our, 1000, customer].

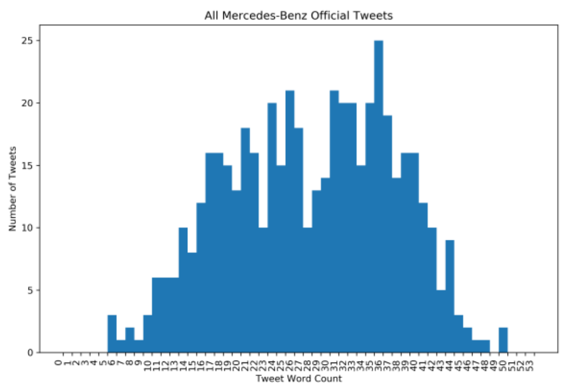


Figure 3: Mercedes OT Wordcounts

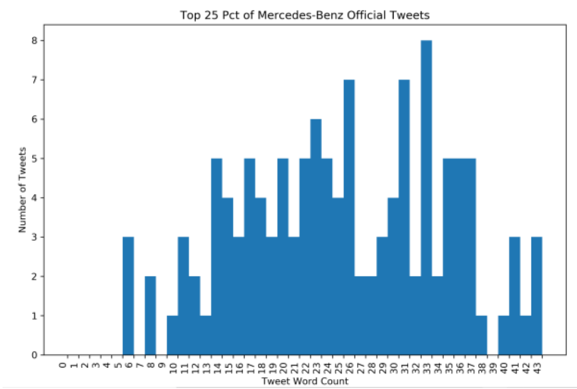


Figure 4: Mercedes Top Performing OT Wordcounts

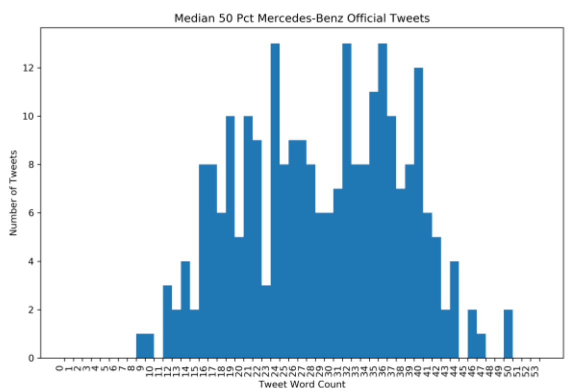


Figure 5: Mercedes Median Performing OT Wordcounts

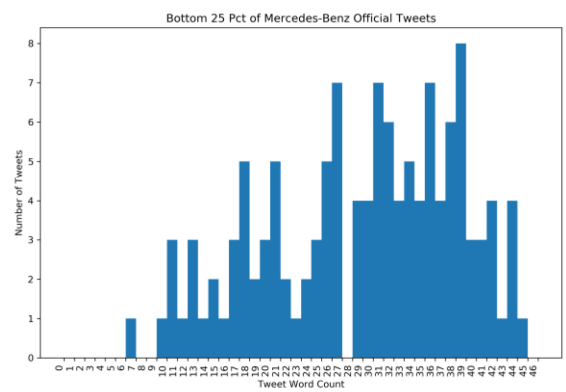


Figure 6: Mercedes Bottom Performing OT Wordcounts

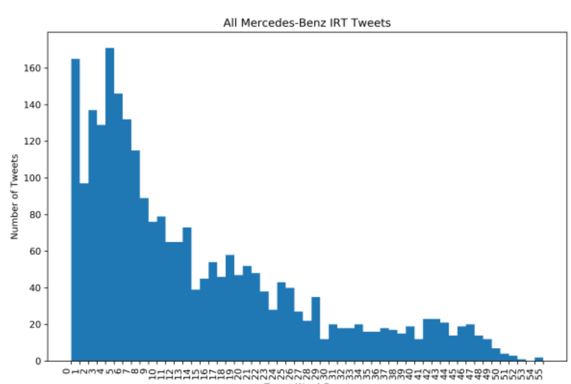


Figure 7: Mercedes IRT Wordcounts

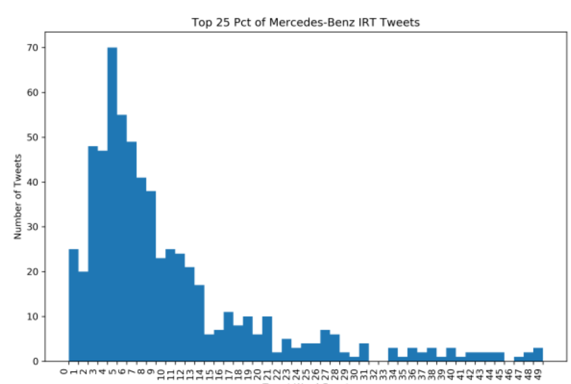


Figure 8: Mercedes Top Performing IRT Wordcounts

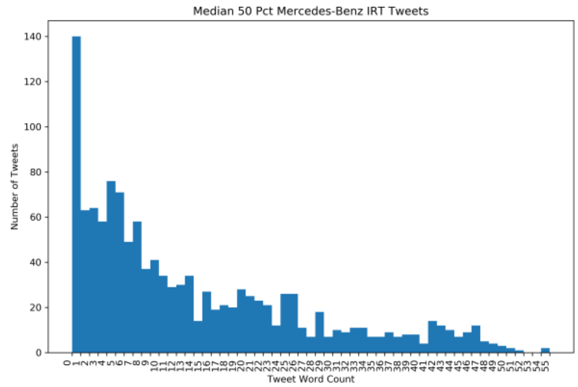


Figure 9: Mercedes Median Performing IRT Wordcounts

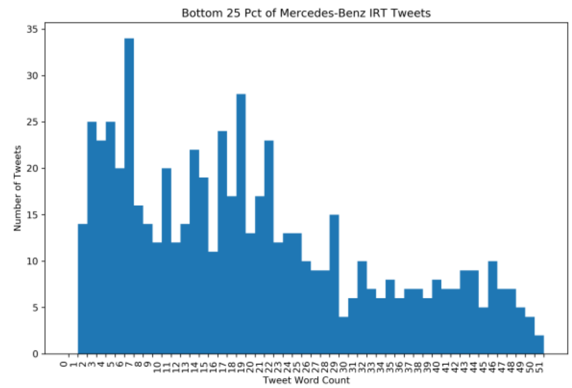


Figure 10: Mercedes Bottom Performing IRT Wordcounts

Similar figures for other companies will be placed in section 2 of the appendix. The 'top,' 'median,' and 'bottom' performance categories are split according to the number of likes generated by each tweet. Thus, the 'top performing OT category' in figure 4 consists of the top 25% most liked Mercedes official tweets, and so on. Mercedes data was selected as example data because it's fairly representative of the overall results. For official tweet categories, examining performance categories for differences in wordcount distributions seemed to be an exercise in futility.

For IRT categories, however, it looked like the top performing tweets, generally, are shorter, while the bottom performing tweets are generally longer. Furthermore, this only seemed to be true for companies who have tangible products (Amazon, BMW, Coca Cola, McDonald's, and Mercedes). Companies who primarily produce software or utility type services, such as Google and Microsoft, don't seem to follow this pattern.

This trend may be a result of tangible products inherently providing more opportunities for the average consumer to express their satisfaction online, which may then be adequately replied to with a short (in length) message from the company. Unsatisfied customers, on the other hand, require longer responses from the company to adequately address their concerns. It seems logical that satisfied customers are more apt to 'like' a response than unsatisfied customers are, no matter the length of the response. However, as noted above, satisfied customers themselves are more likely to receive a short response, while unsatisfied customers are more likely to receive a relatively lengthy response. Our team has deemed this phenomenon the "recipient effect," as wordcount seems to be more of an indication towards the emotions of the recipient than a true explanation for the number of likes an IRT tweet receives.

Figure 11: Example Short Amazon IRT Tweet

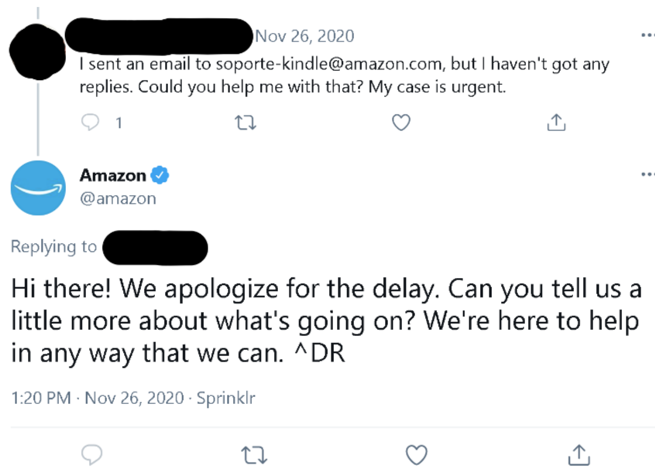


Figure 12: Example Lengthy Amazon IRT Tweet

Table 3: IRT Tweet Category Wordcounts

Data	Number Tweets	Average Wordcount	Total Words	Unique Words
Amazon Top IRT	723	13.51	9767	1538
Amazon Median IRT	1446	16.98	24556	2292
Amazon Bottom IRT	723	20.50	14822	1535
BMW Top IRT	570	13.99	7972	1177
BMW Median IRT	1140	20.82	23732	1396
BMW Bottom IRT	570	24.86	14171	908
Coca Cola Top IRT	795	21.27	16908	1745
Coca Cola Median IRT	1588	28.07	44575	2385
Coca Cola Bottom IRT	795	29.64	23562	1518
Google Top IRT	758	23.42	17750	846
Google Median IRT	1513	23.84	36067	1170
Google Bottom IRT	758	23.74	17998	660
McDonald's Top IRT	799	9.58	7656	1187
McDonald's Median IRT	1596	12.41	19804	1579
McDonald's Bottom IRT	799	13.93	11132	1130
Mercedes Top IRT	632	9.80	6192	1276
Mercedes Median IRT	1262	13.45	16979	1678
Mercedes Bottom IRT	632	19.24	12159	1292
Microsoft Top IRT	341	4.94	1683	661
Microsoft Median IRT	679	4.19	2843	984
Microsoft Bottom IRT	341	3.62	1235	528

After calculating wordcounts for all companies and categories, lists of the top 20 most commonly occurring words were produced as well. For each company, these lists were scanned against one another to search for words uniquely common to a particular category, as in the following for Amazon data:

Table 4: Amazon Top 20 Commonly Occurring (Words, Counts) by Category

All Tweets	Official Tweets	Top 25% Official	Median 50% Official	Bottom 25% Official	All IRT Tweets	Top 25% IRT	Median 50% IRT	Bottom 25% IRT
send, 939	amazon, 86	amazon, 28	amazon, 40	amazon, 18	send, 938	love, 132	send, 408	send, 418
love, 732	new, 30	delivery, 11	new, 19	holiday, 12	love, 723	send, 112	love, 356	details, 266
details, 628	see, 26	customers, 8	day, 17	shop, 8	details, 624	happy, 104	details, 282	deliveringsmiles, 239
like, 481	delivery, 25	thank, 8	see, 15	check, 8	deliveringsmiles, 468	thanks, 102	please, 230	love, 236
deliveringsmiles, 481	help, 23	ok, 8	help, 14	help, 7	like, 466	like, 87	thanks, 229	holiday, 199
please, 437	holiday, 22	family, 7	get, 13	favorite, 7	please, 434	hear, 84	like, 224	season, 175
holiday, 420	get, 22	new, 7	check, 11	new, 6	thanks, 405	details, 76	happy, 196	like, 155
thanks, 412	check, 20	like, 6	delivery, 10	gits, 6	holiday, 398	hope, 72	deliveringsmiles, 194	please, 137
happy, 386	today, 19	today, 6	items, 10	app, 6	happy, 377	please, 67	holiday, 168	surprise, 92
season, 357	day, 19	support, 6	holiday, 10	live, 6	season, 347	thank, 62	hear, 163	help, 77
hear, 323	people, 19	world, 6	customers, 9	season, 5	hear, 322	sharing, 57	thank, 151	happy, 77
help, 296	customers, 18	see, 6	people, 9	today, 5	help, 273	help, 52	season, 147	well, 77
thank, 285	support, 16	learn, 5	today, 9	shopping, 5	thank, 271	great, 50	help, 144	hear, 75
surprise, 238	family, 15	introducing, 5	free, 8	teamed, 5	surprise, 233	glad, 50	great, 119	thanks, 74
hope, 232	like, 15	year, 5	need, 8	support, 4	hope, 231	day, 44	hope, 118	something, 65
great, 231	thank, 14	employees, 5	around, 8	donate, 4	great, 227	enjoying, 44	surprise, 106	sounds, 65
day, 192	learn, 14	proud, 5	store, 8	beauty, 4	sharing, 186	shout, 38	sharing, 96	thank, 58
time, 188	items, 13	get, 5	home, 8	see, 4	time, 179	new, 37	day, 93	great, 58
sharing, 187	around, 13	people, 5	visit, 8	well, 4	well, 175	looks, 36	time, 91	time, 58
well, 187	deliveringsmiles, 13	community, 5	deals, 7	get, 4	day, 173	enjoy, 35	shout, 85	festive, 57

Table 5: Amazon Words Uniquely Common to Categories

All Tweets	Official Tweets	Top 25% Official	Median 50% Official	Bottom 25% Official	All IRT Tweets	Top 25% IRT	Median 50% IRT	Bottom 25% IRT
		ok	free	shop		glad		something
		world	need	favorite		enjoying		sounds
		introducing	store	gifts		looks		festive
		year	home	app		enjoy		
		employees	visit	live				
		proud	deals	shopping				
		community		teamed				
				donate				
				beauty				

Similar tables for other companies will be placed in section 3 of the appendix. Additionally, the top 20 commonly occurring bigrams and trigrams for all companies and categories were also produced at this stage, all of which will be placed in section 3 of the appendix as well. Over time, the high degree of difficulty in identifying some explanatory variable through the above means became increasingly evident. It may perhaps be possible to do so for a single company; but identifying commonalities, or common themes, across multiple companies and tweet categories is a different task altogether.

In an attempt to extract more information, the above processes were repeated with lemmatization as an included step. This only resulted in more hours being spent vainly scouring over word lists, bigrams, trigrams, and the same ultimate conclusion – that this is an interesting but highly impractical method for generating hypotheses regarding explanatory variables in Twitter data.

Preparation:

Making use of 'Orange3' data mining software to scrape the Twitter API proved to be an imperfect, although recoverable, method of collecting Twitter data. Issues stemming from data collection were very minor but had to be addressed during the analysis above. For example, all ampersands in the original tweets were transformed from '&' to '&' during data collection. The same may be said for "greater than" and "lesser than" symbols, which were transformed to '>' and '<' during collection, respectively. These issues were easily resolved by finding and replacing all occurrences with their appropriate symbols in the datafiles themselves. Other, more minor, issues were present as well (such as handling of 'smart' apostrophes and quotations), none of which require discussion.

After failing to generate a hypothesis or identify some common theme relevant to all companies in textual exploration above, it became apparent that further exploration would require the use of more general variables, applicable to all companies and categories. Our team next decided to create a binary "contains link" variable across all data. Creating this variable was simple, as a regular expression was used to scan each tweet for the string "https://", a common sequence shared by all links in all data. After doing so, the following descriptive statistics were produced, again only using tweets interpretable to an English audience:

Table 6: Link Descriptive Statistics by Company and Category

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
Amazon OT: Link	221	359.76	963.59	60.74	153.60
Amazon OT: No Link	26	296.65	236.49	46.77	50.40
Amazon IRT: Link	834	3.58	45.19	0.51	5.65
Amazon IRT: No Link	2058	3.42	40.13	0.60	2.48
BMW OT: Link	585	1279.85	775.40	135.57	108.19
BMW OT: No Link	211	450.64	650.43	44.72	63.07
BMW IRT: Link	500	1.61	16.07	0.12	0.89
BMW IRT: No Link	1780	2.43	11.02	0.15	0.70
Coca Cola IRT: Link	376	163.35	3161.08	25.89	501.17
Coca Cola IRT: No Link	2802	2.31	14.79	0.11	1.09
Disney OT: Link	2629	3199.32	8061.77	721.37	3226.00
Disney OT: No Link	91	4045.71	5926.06	920.54	1972.35
Google IRT: Link	1220	0.55	10.48	0.05	0.92
Google IRT: No Link	1809	11.24	466.13	3.52	148.92
McDonalds IRT: Link	1181	0.53	11.09	0.02	0.14
McDonalds IRT: No Link	2013	11.23	78.85	0.20	1.25
Mercedes OT: Link	505	1118.62	1147.70	121.48	251.80
Mercedes OT: No Link	0	NA	NA	NA	NA
Mercedes IRT: Link	120	0.86	1.40	0.11	0.38
Mercedes IRT: No Link	2406	1.50	20.81	0.22	3.66
Microsoft OT: Link	1014	613.82	2371.72	119.80	433.80
Microsoft OT: No Link	247	1353.54	3151.72	166.57	460.35
Microsoft IRT: Link	29	70.41	133.35	4.38	13.05
Microsoft IRT: No Link	1332	54.18	453.51	1.96	17.44
Samsung OT: Link	2421	533.10	1857.92	52.48	254.68
Samsung OT: No Link	479	27.14	57.41	13.80	30.25
Toyota OT: Link	2535	111.64	679.13	24.80	149.36
Toyota OT: No Link	4	15	7.12	2.25	0.96

Table 6 makes it clear that the effect, or lack thereof, potentially stemming from including a link in tweets is company - as well as category - specific. For example, BMW official tweets seem to experience a boost in performance with link inclusion, but BMW IRT tweets seem to have better performance when no links are embedded within their contents. Generally, this pattern applies to all OT and IRT tweet categories. However, it may not be considered a universal truth. Furthermore, some of the rows in the table above contain incredibly large standard deviations, implying huge spreads in data and that extreme outliers may be affecting results. This will be discussed in greater detail at a later point.

Following creation of the binary ‘contains link’ variable, a new variable was created, equal to the count of the number of hashtags contained in each tweet. This “number hashtags” variable was also created through use of a regular expression, with a considerable amount of trial and error. Obtaining an accurate count of the number of hashtags contained in each tweet was a much more difficult process than expected. This is because 'true hashtags' had to be separated from numeric pound signs, while still allowing for numeric characters to be contained within a hashtag's string. For example, the use of a pound sign in '#1' or 'reference number #1290374' differs from the use of a pound sign in '#Daytona500' or '#1A.' The former uses of the pound sign would be read aloud as "number," while the latter usages would be read aloud as "hashtag."

After creating the appropriate regular expression and hand-checking a large sample of results to ensure validity, a 'number hashtags' variable was created throughout all data. Following, a table similar to Table 6 above was created. However, issues with extreme outliers still mired the data; and separating data based on the number of hashtags, rather than a binary quality, made for extremely uneven sample sizes in some cases.

The results again seem to be highly dependent on the specific company and category of tweets under examination. One observation, however, was that the largest changes in tweet performance generally seem to occur between sets of tweets containing 0 hashtags, and those containing 1 hashtag. While still being company and category specific, in general, increasing from a positive number of hashtags to the next integer did not have a great impact on tweet performance. Regardless, due to the large number of rows, descriptive statistics for the 'number hashtags' variable will be placed in section 4 of the appendix.

Following hashtag analysis, our team decided to next examine the possibility that a tweet's sentiment (positive, negative, or neutral) has an impact on its performance. One limitation, however, is that sentiment, or anything remotely related to sentiment, is not available to scrape from the Twitter API. This implies that to make use of a 'supervised' sentiment analysis algorithm would require manually labeling a large number of tweets with their perceived sentiment as training data. This task would be time consuming, as well as subjective. Therefore, an 'unsupervised' sentiment analysis algorithm was needed, as this type of algorithm does not require training data.

Research into the optimal sentiment analysis model for short-texts, such as Twitter data, eventually led to the IBM-Watson Analyzer. In "Evaluating Unsupervised Sentiment Analysis Tools Using Labeled Data" a study was conducted in which three sentiment analyzers (Textblob, VaderSentiments, and IBM-Watson Analyzer) were tested on 2,748 rows of labelled, short-text data. Data consisted of reviews sourced from Yelp, Amazon, and IMDB. The results of the study found that IBM-Watson outperformed Textblob and VaderSentiments analyzers across all evaluation metrics that were used (accuracy, precision, and recall). Furthermore, IBM-Watson was found to have better performance across all metrics than an ensemble method, created through use of a “majority wins” ruleset of the above three analyzers (Anita).

Our team believed that these reviews and Twitter data held enough similarities for the results to be applicable, as both may be classified as short texts. However, before feeding data into the IBM-Watson Analyzer, textual pre-processing and standardization had to again be performed. For the most part, textual pre-processing in sentiment analysis was similar to textual pre-processing for wordcount and common word analysis mentioned earlier. Be that as it may, some key differences were made when pre-processing textual data for sentiment analysis, beginning with handling of emojis in data.

Earlier in wordcount analysis, emojis were removed during textual pre-processing and standardization, as they are not relevant to wordcounts, bigrams, trigrams, or lists of commonly occurring words. In sentiment analysis, however, emojis potentially contain vital information regarding the intended sentiment of the text to which they belong. For this reason, emojis were instead replaced with textual representations during sentiment analysis. For example, emojis taking the shape of a red heart were replaced with the text ‘red heart’.

Other key differences in textual pre-processing include negation handling, abbreviation replacement, lack of stop-word removal, and hashtag removal. Removing certain words in sentiment analysis, such as the word “don’t,” can greatly affect results. For example, the phrase “I don’t enjoy this music” is opposite from the phrase “I enjoy this music.” Therefore, abbreviations (such as “don’t”) are replaced with their full form (converts to “do not”) and then stop-words (such as “not”) are kept within text. Furthermore, stop-words were kept within textual contents because the IBM-Watson Analyzer takes the context of words into account (Porutiu). Finally, hashtags were elected to be removed because they commonly are a concatenation of multiple words or refer to the name of some event. Including these would provide the analyzer with no additional information.

Table 7: Sentiment Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
Amazon OT: Positive	175	328.50	719.47	52.74	76.02
Amazon OT: Neutral	39	198.54	121.08	34.44	28.45
Amazon OT: Negative	32	369.81	761.83	64.94	122.60
Amazon IRT: Positive	2528	3.23	40.01	0.53	2.48
Amazon IRT: Neutral	254	5.51	60.16	1.03	9.66
Amazon IRT: Negative	108	4.06	22.05	0.42	0.58
BMW OT: Positive	394	1298.02	807.92	138.48	118.86
BMW OT: Neutral	185	1284.88	791.91	131.22	89.66
BMW OT: Negative	36	1239.58	863.03	127.81	86.87
BMW IRT: Positive	1252	2.68	12.01	0.16	0.78
BMW IRT: Neutral	894	1.07	4.76	0.08	0.41
BMW IRT: Negative	102	3.17	12.92	0.16	0.50
Coca Cola IRT: Positive	1897	35.03	1407.35	5.27	223.12
Coca Cola IRT: Neutral	1012	1.02	13.83	0.04	0.37
Coca Cola IRT: Negative	268	1.51	6.43	0.04	0.26
Disney OT: Positive	1585	2930.65	4393.96	610.57	1369.30
Disney OT: Neutral	723	3889.09	13366.97	1025.29	5750.34
Disney OT: Negative	377	3134.73	5152.39	644.57	1289.29
Google IRT: Positive	194	103.12	1423.38	32.79	454.75
Google IRT: Neutral	2830	0.34	6.62	0.03	0.47
Google IRT: Negative	4	1.00	0.82	0.25	0.5
McDonald’s IRT: Positive	1364	7.37	75.98	0.13	1.18
McDonald’s IRT: Neutral	1184	6.96	46.33	0.12	0.78
McDonald’s IRT: Negative	640	7.71	60.40	0.15	0.96
Mercedes OT: Positive	403	1140.86	1245.18	127.55	279.74
Mercedes OT: Neutral	81	1049.84	628.91	102.26	66.77
Mercedes OT: Negative	21	957.05	624.81	79.05	53.67
Mercedes IRT: Positive	1606	1.76	23.74	0.24	3.68
Mercedes IRT: Neutral	579	1.11	15.29	0.23	4.24
Mercedes IRT: Negative	330	0.68	2.17	0.07	0.31
Microsoft OT: Positive	718	542.94	1915.91	103.88	400.34
Microsoft OT: Neutral	375	960.81	1870.99	153.31	288.88
Microsoft OT: Negative	148	1227.84	5289.55	185.30	800.38
Microsoft IRT: Positive	626	49.31	357.05	1.72	11.13
Microsoft IRT: Neutral	473	37.50	149.16	1.69	10.35
Microsoft IRT: Negative	247	95.72	859.93	3.18	33.67
Samsung OT: Positive	1504	612.89	1933.84	56.38	245.04
Samsung OT: Neutral	1258	281.79	1443.42	34.25	214.15
Samsung OT: Negative	132	204.73	1008.76	43.23	272.93
Toyota OT: Positive	1415	129.08	906.75	29.14	199.34
Toyota OT: Neutral	878	88.23	66.42	18.78	15.09
Toyota OT: Negative	242	94.34	73.80	21.14	18.61

The results of table 7 again imply that target variable behavior is dependent on the specific company and tweet category under consideration. No widespread statements may be made about potential sentiment effects which are applicable to all companies. However, it’s promising that in many cases it looks possible that the mean expected likes or retweets may potentially differ based on tweet sentiment.

The final item to discuss in feature creation is topic modeling. As with sentiment analysis, an immediate limitation is that the selected algorithm must again be unsupervised. ‘Topic’ labels are not a pre-existing feature in data and could not be created manually for the same reasons as in sentiment analysis. Research into topic modelling algorithms optimal for Twitter data consistently led to the same model, the “Biterm Topic Model” (BTM).

Many of the classic, best-performing topic modelling algorithms are not suitable for Twitter data, as their intended usage is for documents containing at least several hundred words. BTM, however, was created specifically for use with short-texts and to overcome the difficulties that traditional topic modelling algorithms experience when faced with sparse data. In “An Evaluation of Topic Modelling Techniques for Twitter” a study compared the performances of four topic modelling algorithms (LDA, LDA-U, BTM, W2V-GMM) on Twitter data.

Throughout all algorithms tested, BTM predicted topics with minimum distance within topic clusters and maximum distance between topic clusters. Furthermore, on average, BTM produced the best topic coherence scores across all models tested (Jonsson). Additionally, in “A Biterm Topic Model for Short Texts” another study was conducted whose results suggest that BTM produces more coherent topics than LDA, LDA-U, and a mixture model by statistically significant margins (Yan).

Although encouraging, one major difference still remained between the aforementioned studies and our desired usage of BTM. These studies had the benefit of using Twitter data with pre-defined numbers of topics. In sentiment analysis, the number of different categories is inherent (positive, neutral, negative). In topic modelling, one must also first determine the optimal number of topics to split data into. For example, BTM cannot simply be expected to assign Amazon’s official tweets into separate topic categories. Rather, an additional parameter must be supplied, allowing BTM to assign Amazon’s official tweets into ‘K’ separate topic categories, where ‘K’ is a positive integer. Thus, the question becomes: “How does one select the optimal K value for each company, and how may that selection be justified?”

Despite extensive research into the subject, no method, process, or heuristic for identifying or estimating the optimal number of topics in BTM modelling seems to exist. The studies above needn’t worry themselves with doing so, as they had pre-defined values prior to modelling. The absence of results from research into the subject suggests that, collectively, there hasn’t been a great deal of effort put forth by the academic community in identifying a process which can quickly identify optimal ‘K’ values to be used in BTM modelling. However, these methods do exist for other topic modelling algorithms, such as “k-means clustering.”

In k-means clustering, traditionally accepted, quick heuristics for identifying the optimal number of topics to split data into are ‘elbow’ plots and usage of ‘silhouette scores.’ Curious as to whether these may be useful in BTM modelling as well, a “quick” (but still computationally expensive) version of BTM modelling was created. For each company and tweet category, elbow plots and silhouette scores were generated using K values from 3 to 20, followed by “quick” BTM modelling for every K value. Afterwards, results were examined to see if any relationship might exist between optimal K values suggested by elbow plots and/or silhouette scores with the K value producing the best average topic coherence in “quick” BTM.

Topic coherence is essentially a measure of the semantic similarity between the top words assigned to each topic. Average topic coherence is simply the average coherence across all ‘K’ topics used in modelling. Topic coherence scores are useful because they assist in distinguishing topics which are semantically interpretable from topics which are no more than artifacts of statistical inference (Kapadia). Simply stated, the more similar the top words associated with a topic are to one another, the better the topic coherence score. Through most traditional methods, coherence scores are output as decimals and larger values are considered better. The evaluation algorithm used in BTM modelling, however, reports coherence scores as negative, continuous values. The reasonings for this are unknown to our team, but having less negative topic coherence scores in BTM modelling is desired (Yan). For example, a topic coherence of -100 would be preferred over a topic coherence of -110.

To save on space, elbow plots, silhouette scores, and their corresponding topic coherence scores will be placed in section 5 of the appendix. Unfortunately, no discernible relationship or pattern was found to exist between traditional heuristics for estimating the optimal number of topics and the ‘K’ value producing the least negative average topic coherence score in “quick” versions of BTM modelling. Ultimately, for each company and tweet category, the ‘K’ value to be used in the full BTM modelling process was simply the number of topics producing the lowest average topic coherence score during “quick” BTM estimation. K values ranged from three to twenty, rather than from two to twenty, because comparing average topic coherence scores is not a robust process, thereby potentially biasing selected K towards lower values. Furthermore, our team felt it was unlikely that any company or tweet category could be adequately captured using two topics alone, but three topics could be reasonable in some cases.

Table 8: BTM Estimation and Modelling, Average Topic Coherence Values

Data	Best Number of Topics	Estimated Avg. Coherence	Avg. Coherence
Amazon OT:	17	-67.41	-75.51
Amazon IRT:	3	-80.70	-79.22
BMW OT:	3	-57.27	-83.12
BMW IRT:	8	-44.11	-53.70
Coca Cola IRT:	3	-48.12	-50.01
Disney OT:	3	-105.79	-111.76
Google IRT:	4	-37.01	-49.95
McDonald's IRT:	6	-81.94	-68.03
Mercedes OT:	11	-80.43	-84.96
Mercedes IRT:	6	-76.98	-69.92
Microsoft OT:	3	-104.35	-108.14
Microsoft IRT:	20	-67.79	-72.74
Samsung OT:	4	-127.05	-128.18
Toyota OT:	14	-119.42	-121.34

For the most part, the results of table 8 show that the estimated average topic coherence in "quick" BTM modelling is reasonably close to the final average topic coherence produced when performing the full BTM modelling process, using the same value of K. Without a doubt, discrepancies exist between the estimates and final values. However, having found no alternatives for estimating the optimal number of topics, the 'K' values in table 8 were considered final.

Table 9: Disney OT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	disneyplus, stream, original, start, new, series, come, movie, november, 12	-88.43
1	disney, new, day, episode, disneymagicmoments, mickey, anniversary, story, celebrate, resort	-143.08
2	theater, see, new, get, film, disney, ticket, march, watch, pixaronward	-103.78

Table 9 shows the results of BTM modelling for Disney OT data. Similar tables for other companies and tweet categories will be placed in section 6 of the appendix. After examining the top words associated with each topic, as well as examining individual tweets assigned to each topic, our team was often able to identify an underlying theme associated with a particular topic. For example, topic 0 of Disney data has an underlying theme of streaming announcements. Regardless of whether it be a movie or TV series, tweets assigned to topic 0 generally seem to be announcements of what was currently (at the time) available for streaming on Disney platforms. Topic 1, for the most part, pertains to ‘classic’ Disney media and attractions. Generally, any tweet mentioning the anniversary of an old Disney TV show or movie, or any tweet referring to a physical Disney location (such as Disney World), was assigned to topic 1. Finally, tweets assigned to topic 2 have a definite underlying theme of being announcements for upcoming Disney movies.

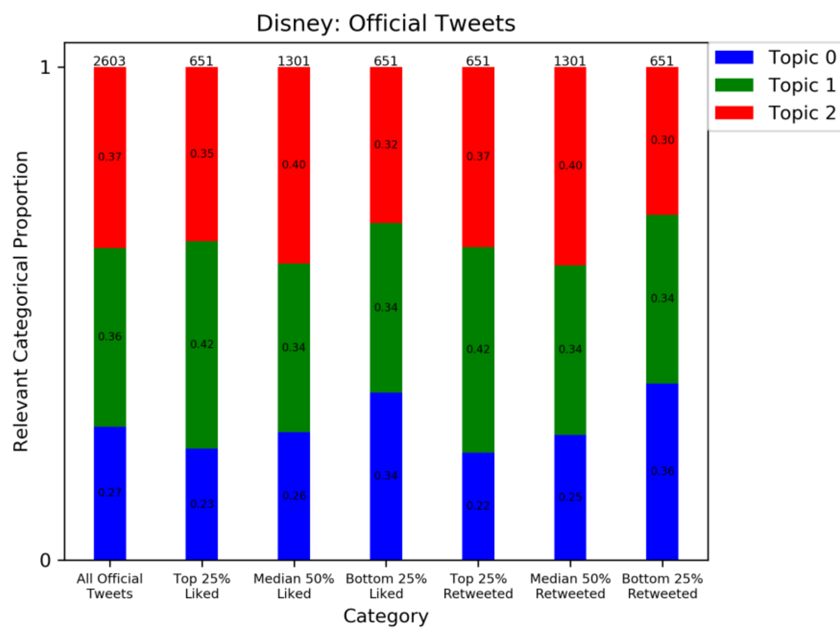


Figure 13: Disney OT Topic Proportions by Performance Category

Table 10: Disney OT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	705	0.27	2729.79	5275.55	554.14	1466.47
1	943	0.36	2936.63	3531.99	568.07	926.51
2	955	0.37	3685.44	11842.48	984.96	5100.34

Results similar to figure 13 and table 10 for other companies will be placed in section 6 of the appendix as well. Overall, BTM topic modelling looks to be impressive, but the results are inarguably imperfect. There are instances in which a tweet may seemingly fit perfectly with a certain topic yet be erroneously assigned to a different topic altogether. Furthermore, in many cases, no underlying theme or commonalities were able to be discerned by our team within topics. At other times, it seemed as if topic modelling results for a particular company would have been improved by combining similar topics or increasing the number of topics to allow for more separation. However, no alterations to topic modelling results were made by our team, as using estimated average topic coherence was the only objective measure available.

Before moving on, it's worth mentioning that the textual pre-processing and standardization applied to data before supplying tweets to BTM modelling was very similar to textual pre-processing and standardization initially referenced in wordcount/common word analysis. However, lemmatization was always performed in BTM modelling, whereas the lemmatized results for wordcounts and common words were ultimately discarded. Furthermore, in BTM modelling each tweet must contain at least three words after standardization and stop-word removal. This implied that all 'UND' tweets had to be removed from analysis, as they contain no legitimate words, and all English tweets containing less than three words after pre-processing and stop-word removal had to be discarded as well.

Methodology

Techniques: Variable Analysis

To determine whether a particular variable may or may not be useful in predicting likes and retweets, our team needed to find appropriate statistical tests for the occasion. Considering our desire was to compare average tweet performance after splitting based on some condition, a statistical test comparing observed means seemed most appropriate. Furthermore, since no tweet may be assigned to more than one group, and the tweets in one group provide no information regarding tweets in other groups, a statistical test comparing observed means of independent samples was required.

Usually, such comparison calls for a t-test (when comparing exactly two groups) or usage of ANOVA (when comparing more than two groups). However, similar to Pearson's correlation test, these are parametric techniques and require that data are normally distributed. As previously mentioned, severe issues of non-normality are present in data, making these tests ineligible. Furthermore, in a vast majority of cases, no transformation was found which made data normal, implying that these tests could not even be used on transformed data. This remained true despite testing multiple log-transformations, square root transformations, cube root transformations, fourth root transformations, as well as Inverse Hyperbolic Sine transformations. Over time, it became apparent that no transformation would be able to normalize data containing such large volumes of 0 values, or extremely long right-tails across all data. Very often, both issues are present in our data.

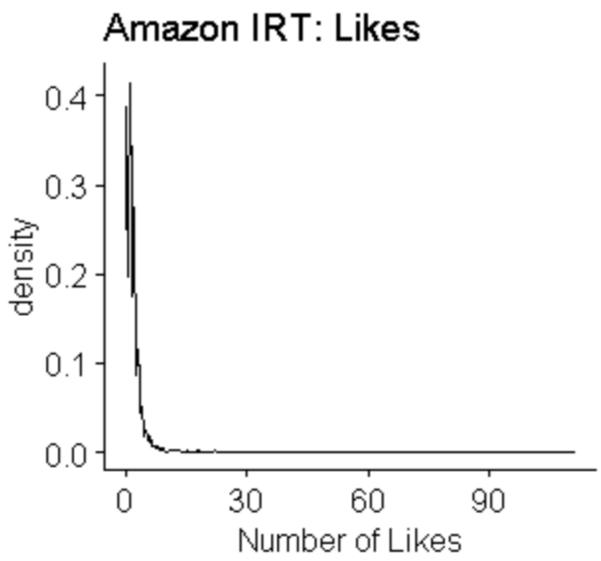


Figure 14: Amazon IRT Density Plot of Likes

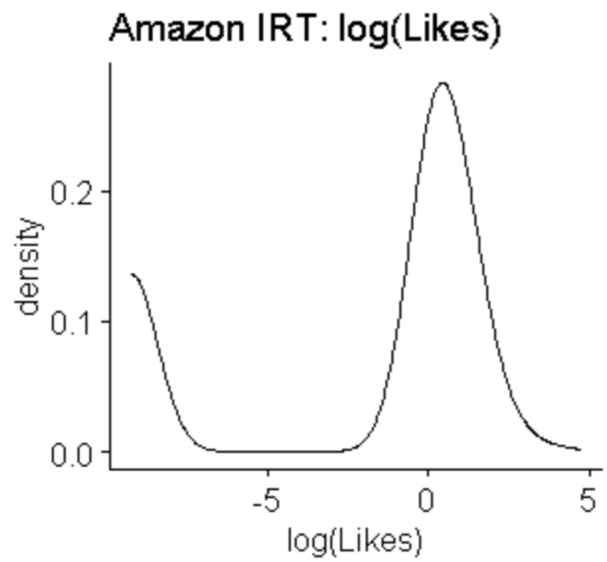


Figure 15: Amazon IRT Density Plot of Log(Likes)

Procedure: Variable Analysis

As a first step in variable analysis, and only variable analysis, extreme outliers were removed from our data. For the purposes of this project, extreme outliers are considered to be tweets receiving 5 standard deviations above their company and tweet category's mean expected likes or retweets. A value of 5 standard deviations was selected because, traditionally, receiving 3 standard deviations above mean expectations qualifies an observation as an outlier. Initially, 4 standard deviations were used, but that seemed to remove too much data. Furthermore, data in figures 14 and 15 had already undergone this process. So, they're not even fully representative of the true degree of non-normality in Amazon IRT data. Additionally, since the findings from variable analysis might have potentially influenced the variables selected for use in modelling, tweets not able to be assigned a topic in BTM modelling were discarded from analysis.

Considering parametric tests were not available for use, non-parametric alternatives had to be researched. In place of t-tests, a non-parametric alternative is to use Mann-Whitney U tests. In place of ANOVA, a non-parametric alternative is to use Kruskal-Wallis tests. If the results of Kruskal-Wallis are significant, this may be followed by Dunn's test to examine which specific groups differ from one another. Furthermore, it was decided to apply Dunn's test with a Bonferroni correction to account for the inflated Type I Error rates stemming from multiple comparisons.

However, to be quite honest, receiving statistically significant results from these tests doesn't reveal much more information than the fact that the distributions under examination differ in some manner. Considering the volatility of Twitter data, even after removing extreme outliers, results tended to be significant. In order to extract more information and insights, and to increase the threshold for concluding that statistically significant differences exist, further research was put into the matter. This eventually led to something known as "shift function" analysis.

Shift function analysis differs from traditional analysis in that it goes beyond comparing just the central tendencies of two or more groups. In "Plotting with Confidence: Graphical Comparisons of Two Populations" the author believed it to be unreasonable to assume, a priori, that two distributions would only differ in central tendency. Thus, the original version of the shift function was born; a systematic approach in which differencing is utilized to compare the average differences between observations belonging to the same quantile of two separate groups (Doksum). More simply, it estimates how far a certain quantile in one distribution would need to be shifted up or down to match its counterpart in another distribution.

Years later, improvements were made to add to the power of shift function analysis. The version of the shift function ultimately used in analysis for this project makes use of the Harrell-Davis quantile estimator, along with percentile bootstrapping techniques of estimating decile standard error, to compute confidence intervals of quantile differences between distributions. Furthermore, this technique takes measures to ensure that the Type I Error rate stays near 0.05, despite multiple comparisons being made (Rousselet).

Considering the shift function isn't a traditionally utilized method in data analysis, and that the traditional Mann-Whitney U and Kruskal-Wallis tests don't provide a great deal of information, it was decided that the two tests would be used in tandem with one another. Furthermore, in order to ensure that no erroneous conclusions might still be made, visualizations were produced to further examine results. These visualizations are produced using log-transformed data, as that creates better separation in visualized results, but all analysis is conducted on non-transformed data. In the end, results had to be statistically significant in Mann-Whitney U or Kruskal-Wallis tests, followed by statistically significant results from shift function analysis, as well as producing convincing visualizations in order to conclude that differences exist between the distributions under examination. Only then do we hypothesize that, potentially, one underlying factor explaining these differences is the variable under examination.

Techniques: Modelling Analysis

There is no shortage of techniques or models available which may be utilized in predicting likes and retweets. More or less, any model capable of generating predictions for a continuous target variable could be used for these purposes. In “Prediction of Likes and Retweets Using Text Information Retrieval” the authors compare performances between Logistic Regression, SVM, Random Forest, Neural Network, and Multinomial Naïve Bayes models (Daga). However, in this study, the target variables were transformed to categorical data representing ranges of likes or retweets, implying that all models were used for classification purposes rather than regression. As our team’s desire was to generate continuous predictions, their findings were not directly applicable, but they serve as a good starting point for model selection.

Of the five models listed above, only Multinomial Naïve Bayes cannot be converted to regression purposes. Due to this, Naïve Bayes was dropped from consideration for testing and the remaining four models were selected for testing. However, as logistic regression implies that data would need to first undergo some form of transformation, linear regression was substituted in its place.

The study mentioned above found that Random Forest models have the best accuracy in predicting retweets. For this reason, another forest-based model was selected for testing, Gradient Boosted Decision Trees. Finally, K-Nearest Neighbors was selected for use as the sixth and final model to be tested, in order to increase the variety of models used in this analysis. Ultimately, the six models selected for testing are: Linear Regression, Random Forest, Neural Networks, K-Nearest Neighbors, SVR (the regression version of SVM), and Gradient Boosted Decision Trees.

Procedure: Modelling Analysis

For all companies and tweet categories, data was separated randomly such that 80% of tweets reside in a training dataset, and the remaining 20% of tweets reside in a testing dataset. This was the only method which could guarantee that, across all companies, tweet categories, and models used, the exact same training and testing data would be supplied. Furthermore, for OT data, every combination of the following variables was used in modelling: link inclusion, hashtag inclusion, sentiment, and BTM topic assignment.

There are 15 unique ways to combine the above four variables. Additionally, six different models were used in testing, implying that each OT dataset was tested on 90 different model-variable combinations (for both likes and retweets, individually). For IRT data, the same four variables were used, as well as an additional variable representing the number of words in each tweet. Including a fifth variable makes for 31 unique variable combinations, using 6 separate models implies that each IRT dataset was tested on a total of 186 model-variable combinations (for both likes and retweets, individually).

All models, besides Linear Regression, underwent hyperparameter tuning prior to generating predictions. For OT hyperparameter tuning, 10-fold cross validation with 3 repeats was employed along with a grid search to locate optimal hyperparameters during training. For IRT data, the training had to be relaxed slightly, only employing 5-fold cross validation with 2 repeats during the training process' grid search for optimal hyperparameters.

Mean absolute error (MAE) was selected as the accuracy measure to compare model-variable performances within datasets. After modelling, for each company and tweet category, all 90 (for OT, 186 for IRT) model-variable combinations were ranked according to MAE. Precautions were taken such that, in the event of an exact tie, rank preference would be given to the model-variable combination containing fewer explanatory variables. After generating these rankings for each individual dataset, the average ranking of each model-variable combination was calculated. This calculation allowed us to aggregate results for each model-variable combination into a single performance measure, with inputs coming from all companies.

Results and Analysis

Variable Analysis Results:

Including all visualizations produced during variable analysis would add a considerable amount of length to an already lengthy report. For that reason, examples will be shown, along with explanations for interpretation, but most results will be presented tabularly. All visualizations and results, however, will be included in the appendix.

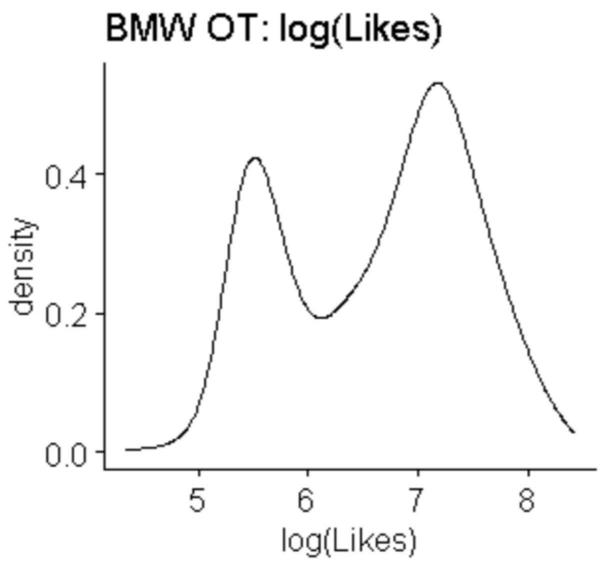


Figure 16: BMW OT Density Plot of Log(Likes)

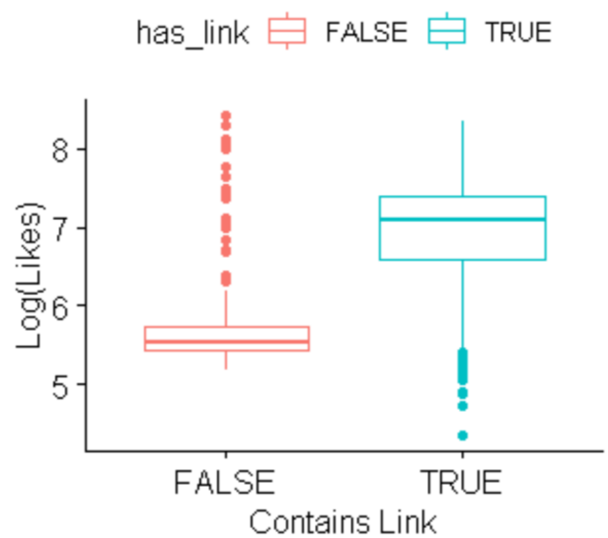


Figure 17: BMW OT Boxplot Separated by Link Inclusion

$`Group1 - Group2`

	q	Group1	Group2	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	386.0766	211.9932	174.0834	113.6116	259.8908	0.050000000	0
2	0.2	620.7758	221.5255	399.2503	331.5165	479.8915	0.025000000	0
3	0.3	830.3078	231.0865	599.2213	519.2971	683.3035	0.016666667	0
4	0.4	1019.0553	241.1312	777.9241	690.8800	863.1770	0.012500000	0
5	0.5	1187.4348	251.1989	936.2359	855.9810	1001.3014	0.010000000	0
6	0.6	1335.6774	270.1269	1065.5505	979.0295	1135.2630	0.008333333	0
7	0.7	1496.0910	296.8627	1199.2283	1098.6947	1334.5266	0.007142857	0
8	0.8	1788.8848	347.1739	1441.7109	1283.9145	1582.8427	0.006250000	0
9	0.9	2240.9671	694.5463	1546.4208	677.4410	1955.7307	0.005555556	0

Figure 18: BMW OT Confidence Intervals Produced by Shift Function for Link - Likes

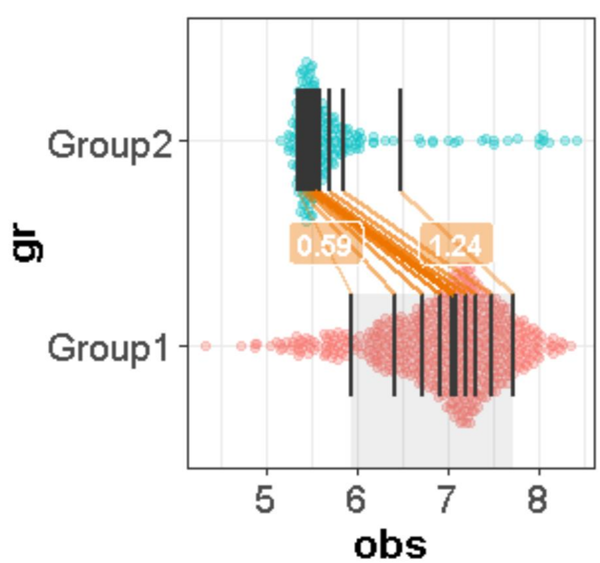


Figure 19: BMW OT Shift Function Visualization for Link - Likes

Above are the four most important visualizations produced when examining BMW OT data for potential effects of link inclusion on the 'number of likes received' target variable (all other figures for link analysis will be placed in section 7 of the appendix). For each company and tweet category, density plots and boxplots of log-transformed data are produced, as in figures 16 and 17. Afterwards, despite not being shown above, a Mann-Whitney U test is performed (on non-transformed data) to assess the null hypothesis that both distributions are equivalent. Receiving statistically significant results from Mann-Whitney U, as is the case for BMW, implies that we may reject the null hypothesis and conclude the two distributions differ in some manner. Afterwards, the shift function is utilized to further examine these differences.

Figure 18 represents the confidence intervals produced by shift function analysis for BMW OT data on the number of likes target variable. In all shift function analysis, 'group 1' represents tweets containing the variable of interest (links in this case), while 'group 2' represents tweets not containing the variable of interest. The confidence intervals represent the degree and direction which a particular quantile in group 2 would need to be shifted to match its counterpart in group 1 data. For example, we may say with 95% confidence that the first quantile of group 2 (BMW OT tweets not containing links) would need to be shifted up by 113 to 259 'likes' to match the first quantile of group 1 (BMW OT tweets containing links).

Therefore, since none of the confidence intervals in figure 18 contain the value 0, we may say with 95% confidence that every quantile of group 2 (BMW OT tweets not containing links) would need to be shifted up by significant (non-zero) amounts of 'likes' to match their counterparts in group 1 (BMW OT tweets containing links). And potentially, one underlying factor explaining these differences is the inclusion of a link in BMW OT tweets.

Lastly, figure 19 is a visualization produced during shift function analysis using log-transformed data instead of non-transformed data. The numbers contained within figure 19 may be interpreted as follows: Positive shifts are labelled in orange, while negative shifts are labelled in purple. The '1.24' value may be used to determine how much one would need to multiply group 2's mean expected likes in quantile 9 to match group 1's mean expected likes in quantile 9. Taking '$e^{1.24}$' yields a value of about 3.456. Multiplying 694 (group 2's mean expected likes in quantile 9) by this value yields 2398, group 1's mean expected likes within the 9th quantile (small differences due to rounding).

Table 11: Analysis for Potential Link Effect on OT Data Likes

Data	Mann-Whitney's p	Quantile Differences?	Quantile Locations	Conclusion
Amazon OT: Likes	0.2382	NA	NA	No evidence
BMW OT: Likes	< 2.2e-16	Yes	(1,2,3,4,5,6,7,8,9)	Differences exist
Disney OT: Likes	0.00115	Yes	(3,4,5,6,7)	Differences exist
Mercedes OT: Likes	NA	NA	NA	Insufficient data
Microsoft OT: Likes	1.719e-05	Yes	(1,2,4,5,6,7,8,9)	Differences exist
Samsung OT: Likes	< 2.2e-16	Yes	(1,2,3,4,5,6,7,8,9)	Differences exist
Toyota OT: Likes	NA	NA	NA	Insufficient data

Table 12: Analysis for Potential Link Effect on OT Data Retweets

Data	Mann-Whitney's p	Quantile Differences?	Quantile Locations	Conclusion
Amazon OT: Retweets	0.9568	NA	NA	No evidence
BMW OT: Retweets	< 2.2e-16	Yes	(1,2,3,4,5,6,7,8,9)	Differences exist
Disney OT: Retweets	0.002865	Yes	(3,4,5,6)	Differences exist
Mercedes OT: Retweets	NA	NA	NA	Insufficient data
Microsoft OT: Retweets	0.05122	NA	NA	No evidence
Samsung OT: Retweets	< 2.2e-16	Yes	(1,2,3,4,5,6,7,8,9)	Differences exist
Toyota OT: Retweets	NA	NA	NA	Insufficient data

Unfortunately, neither Mercedes nor Toyota OT data contain enough tweets without links for fair comparison. Otherwise, there seems to be potential that link inclusion has some effect on the expected number of likes for all OT data, besides Amazon. Interestingly, however, there is no evidence that differences exist between the distributions of retweets of Microsoft OT tweets containing links and Microsoft OT tweets not containing links. This is despite differences existing between Microsoft's two distributions of likes in 8 of the 9 quantiles. Furthermore, fewer quantile differences exist in Disney OT retweet data than in Disney OT likes data. This may perhaps imply that the effect of including a link in tweets differs between likes and retweets for OT data.

Table 13: Analysis for Potential Link Effect on IRT Data Likes

Data	Mann-Whitney's p	Quantile Differences?	Quantile Locations	Conclusion
Amazon IRT: Likes	< 2.2e-16	Yes	(2,3,4,5,6,7,8,9)	Differences exist
BMW IRT: Likes	< 2.2e-16	Yes	(4,5,6,7,8,9)	Differences exist
Coca Cola IRT: Likes	1.85e-09	Yes	(6,7,8,9)	Differences exist
Google IRT: Likes	3.639e-06	Yes	(8,9)	Insufficient evidence
McDonald's IRT: Likes	< 2.2e-16	Yes	(5,6,7,8,9)	Differences exist
Mercedes IRT: Likes	0.226	NA	NA	No evidence
Microsoft IRT: Likes	NA	NA	NA	Insufficient data

Table 14: Analysis for Potential Link Effect on IRT Data Retweets

Data	Mann-Whitney's p	Quantile Differences?	Quantile Locations	Conclusion
Amazon IRT: Retweets	< 2.2e-16	Yes	(6,7,8,9)	Differences exist
BMW IRT: Retweets	0.002906	No	NA	No evidence
Coca Cola IRT: Retweets	0.01973	No	NA	No evidence
Google IRT: Retweets	0.522	NA	NA	No evidence
McDonald's IRT: Retweets	< 2.2e-16	Yes	(9)	Insufficient evidence
Mercedes IRT: Retweets	0.9482	NA	NA	No evidence
Microsoft IRT: Retweets	NA	NA	NA	Insufficient data

It's very interesting that, for IRT likes, differences exist for 4 of the 6 datasets having sufficient data for analysis. For IRT retweets, however, we may conclude that differences exist only for 1 of these 6 datasets. McDonald's IRT retweet data are said to have insufficient evidence because the visualizations produced weren't all too convincing, and statistically significant differences were only found between the 9th quantiles of the two groups. The same may be said for Google IRT likes data, with weak differences in the 8th quantile as well. It seems that link inclusion may have more potential for affecting the number of likes an IRT tweet receives than for affecting the number of retweets an IRT tweet receives.

Table 15: Analysis for Potential Hashtag Effect on OT Data Likes

Data	Mann-Whitney's p	Quantile Differences?	Quantile Locations	Conclusion
Amazon OT: Likes	0.2246	NA	NA	No evidence
BMW OT: Likes	< 2.2e-16	Yes	(1,2,3,4,5,6,7,8,9)	Differences exist
Disney OT: Likes	< 2.2e-16	Yes	(2,3,4,5,6,7,8,9)	Differences exist
Mercedes OT: Likes	0.94	NA	NA	No evidence
Microsoft OT: Likes	2.58e-10	Yes	(1,4,5,6,7,8,9)	Differences exist
Samsung OT: Likes	2.185e-06	Yes	(5,6,7,8,9)	Differences exist
Toyota OT: Likes	< 2.2e-16	Yes	(1,2,3,4,5,6,7,8,9)	Differences exist

Table 16: Analysis for Potential Hashtag Effect on OT Data Retweets

Data	Mann-Whitney's p	Quantile Differences?	Quantile Locations	Conclusion
Amazon OT: Retweets	0.4289	NA	NA	No evidence
BMW OT: Retweets	< 2.2e-16	Yes	(1,2,3,4,5,6,7,8,9)	Differences exist
Disney OT: Retweets	< 2.2e-16	Yes	(2,3,4,5,6,7,8,9)	Differences exist
Mercedes OT: Retweets	0.393	NA	NA	No evidence
Microsoft OT: Retweets	0.5655	NA	NA	No evidence
Samsung OT: Retweets	1.487e-09	Yes	(3,4,5,6,7,8,9)	Differences exist
Toyota OT: Retweets	1.523e-14	Yes	(1,2,3,4,5,6,7,8,9)	Differences exist

For hashtags, no examples are shown because this analysis followed the exact same basic process as link analysis. All hashtag visualizations will be placed in section 8 of the appendix. Furthermore, it was decided that a binary hashtag variable would be used, rather than an integer variable representing the number of hashtags contained in each tweet, because a previous study found no relation between number hashtags and tweet popularity (Feng). Also, our exploratory analysis seemed to indicate that differences mostly occur between having 0 and any positive number of hashtags. However, it's very interesting that, again, conclusions may differ between OT likes data and OT retweets data for the same company.

Table 17: Analysis for Potential Hashtag Effect on IRT Data Likes

Data	Mann-Whitney's p	Quantile Differences?	Quantile Locations	Conclusion
Amazon IRT: Likes	< 2.2e-16	Yes	(2,3,4,5,6,7,8,9)	Differences exist
BMW IRT: Likes	NA	NA	NA	Insufficient data
Coca Cola IRT: Likes	NA	NA	NA	Insufficient data
Google IRT: Likes	NA	NA	NA	Insufficient data
McDonald's IRT: Likes	NA	NA	NA	Insufficient data
Mercedes IRT: Likes	1.782e-11	Yes	(2,3,4,5,6,7,8,9)	Differences exist
Microsoft IRT: Likes	NA	NA	NA	Insufficient data

Table 18: Analysis for Potential Hashtag Effect on IRT Data Retweets

Data	Mann-Whitney's p	Quantile Differences?	Quantile Locations	Conclusion
Amazon IRT: Retweets	< 2.2e-16	Yes	(6,7,8)	Differences exist
BMW IRT: Retweets	NA	NA	NA	Insufficient data
Coca Cola IRT: Retweets	NA	NA	NA	Insufficient data
Google IRT: Retweets	NA	NA	NA	Insufficient data
McDonald's IRT: Retweets	NA	NA	NA	Insufficient data
Mercedes IRT: Retweets	1.463e-07	Yes	(7,8,9)	Differences exist
Microsoft IRT: Retweets	NA	NA	NA	Insufficient data

Unfortunately, it seems that hashtag usage within IRT data is not very common. For example, only 5 Google IRT tweets contain hashtags, which isn't enough for fair comparison. However, it's highly interesting that, for Amazon and Mercedes, quantile differences exist everywhere besides the first quantile for likes data. But, within retweet data, quantile differences only exist in the right-tails of distributions. This repeating pattern implies that independent variables, if at all related to target variables, potentially have more of an effect on the 'likes' target variable than the 'retweets' target variable.

Table 19: Analysis for Potential Sentiment Effect on OT Data Likes

Data	Kruskal-Wallis' p	Dunn's Test Results	Differing Quantile Locations	Conclusion
Amazon OT: Likes	0.406	NA	NA	No evidence
BMW OT: Likes	0.7641	NA	NA	No evidence
Disney OT: Likes	0.2025	NA	NA	No evidence
Mercedes OT: Likes	0.717	NA	NA	No evidence
Microsoft OT: Likes	4.821e-05	(Neutral, Positive)	(5,6,7,8,9)	Differences exist
Samsung OT: Likes	< 2.2e-16	(Neutral, Positive) (Negative, Positive)	(1,2,3,4,5,6,7,8,9) (4,5,6,7,8,9)	Differences exist
Toyota OT: Likes	4.294e-07	(Neutral, Positive) (Negative, Positive)	(3,4,5,6,7,8) (4,5,6,7)	Differences exist

Table 20: Analysis for Potential Sentiment Effect on OT Data Retweets

Data	Kruskal-Wallis' p	Dunn's Test Results	Differing Quantile Locations	Conclusion
Amazon OT: Retweets	0.3388	NA	NA	No evidence
BMW OT: Retweets	0.9894	NA	NA	No evidence
Disney OT: Retweets	0.2193	NA	NA	No evidence
Mercedes OT: Retweets	0.2676	NA	NA	No evidence
Microsoft OT: Retweets	0.003562	(Neutral, Positive)	(7,8,9)	Differences exist
Samsung OT: Retweets	1.105e-13	(Neutral, Positive) (Negative, Positive)	(1,2,3,4,5,6,7,8,9) (8,9)	Differences exist
Toyota OT: Retweets	0.009108	(Negative, Positive)	None	No evidence

All visualizations produced during sentiment analysis will be placed in section 9 of the appendix. However, sentiment appears to be a less promising explanatory variable than link inclusion or hashtag inclusion for OT data. Furthermore, we again see less datasets containing differences within OT retweet data than with OT likes data, and we see the number of differing quantiles is fewer.

Table 21: Analysis for Potential Sentiment Effect on IRT Data Likes

Data	Kruskal-Wallis' p	Dunn's Test Results	Differing Quantile Locations	Conclusion
Amazon IRT: Likes	4.688e-08	(Neutral, Positive)	(3,4,5,6,7,8,9)	Differences exist
BMW IRT: Likes	< 2.2e-16	(Negative, Neutral) (Positive, Neutral)	(1,2,3,4,5,6,7,8,9) (3,4,5,6,7,8,9)	Differences exist
Coca Cola IRT: Likes	< 2.2e-16	(Negative, Neutral) (Positive, Neutral)	(5,6,7,8,9) (6,7,8,9)	Differences exist
Google IRT: Likes	0.003446	(Positive, Neutral)	(6,7,8,9)	Insufficient evidence
McDonald's IRT: Likes	0.0001567	(Negative, Positive) (Neutral, Positive)	(6,7) (6,7)	Insufficient evidence
Mercedes IRT: Likes	< 2.2e-16	(Negative, Positive) (Neutral, Positive)	(4,5,6,7) (4,5,6,7)	Differences exist
Microsoft IRT: Likes	0.4112	NA	NA	No evidence

Table 22: Analysis for Potential Sentiment Effect on IRT Data Retweets

Data	Kruskal-Wallis' p	Dunn's Test Results	Differing Quantile Locations	Conclusion
Amazon IRT: Retweets	0.8144	NA	NA	No evidence
BMW IRT: Retweets	9.489e-05	(Neutral, Positive)	(9)	Insufficient evidence
Coca Cola IRT: Retweets	2.35e-05	(Neutral, Positive)	(9)	Insufficient evidence
Google IRT: Retweets	0.4473	NA	NA	No evidence
McDonald's IRT: Retweets	0.4084	NA	NA	No evidence
Mercedes IRT: Retweets	3.024e-05	(Negative, Positive) (Neutral, Positive)	(9) (9)	Insufficient evidence
Microsoft IRT: Retweets	0.602	NA	NA	No evidence

Sentiment may potentially have an effect on the number of likes an IRT tweet receives, but it does not seem to have any effect on the number of retweets an IRT tweet receives. For Google and McDonald's IRT likes, they are said to contain insufficient evidence due to uneven sample sizes between sentiment categories and unconvincing visualizations produced.

Furthermore, the same analysis process regarding sentiment variables was applied towards BTM topic assignments. However, the large number of topics made for uneven sample sizes, as well as making it difficult for no differences to exist between any quantiles of any two tweet topic groups. For these reasons, all topic modeling results will be stored in section 10 of the appendix. Overall, it seems the effects which individual variables potentially have on tweet performance vary by company and tweet category. Furthermore, it seems to be more difficult to 'move the needle' for retweets than for likes, regardless of whether OT or IRT data is under examination.

Modelling Analysis Results:

Table 23: Top 10 Model-Variable Combination Rankings for OT Likes

Model-Variable Combination	Average Rank
(SVR: Hashtag)	10.29
(SVR: Link, Hashtag)	18.00
(SVR: Link)	18.43
(KNN: Hashtag)	18.71
(KNN: Link, Hashtag)	20.43
(KNN: Link)	28.71
(RF: Link, Hashtag, Topic)	29.14
(KNN: Sentiment)	30.14
(SVR: Link, Hashtag, Sentiment, Topic)	31.29
(SVR: Link, Sentiment, Topic)	32.57

Table 24: Top 10 Model-Variable Combination Rankings for OT Retweets

Model-Variable Combination	Average Rank
(SVR: Hashtag)	2.14
(KNN: Link, Sentiment)	11.86
(SVR: Link, Hashtag)	12.86
(SVR: Link)	14.00
(KNN: Link)	16.29
(KNN: Sentiment)	19.00
(KNN: Link, Hashtag)	19.14
(KNN: Hashtag)	21.14
(KNN: Hashtag, Sentiment)	28.43
(Neural Network: Hashtag, Sentiment)	32.57

For both predicting OT likes and retweets, SVR using hashtag alone has the best average ranking across all 7 OT datasets tested. For OT likes, SVR with hashtag receives an average ranking of about 10 out of all 90 models tested. For OT retweets, SVR with hashtag achieves an average ranking of about 2 out of all 90 models tested. Furthermore, SVR's superior performance in MAE scores, generally, was not by a negligible amount. SVR was clearly able to do more with relatively limited input information than all other models tested for OT data.

Table 25: Top 10 Model-Variable Combination Rankings for IRT Likes

Model-Variable Combination	Average Rank
(SVR: Hashtag)	22.29
(SVR: Link, Hashtag)	23.57
(KNN: Sentiment)	27.14
(KNN: Link)	30.00
(SVR: Link)	31.14
(KNN: Topic)	38.86
(SVR: Link, Hashtag, Topic, Number words)	40.29
(KNN: Link, Topic)	40.57
(KNN: Link, Hashtag)	40.86
(KNN: Hashtag)	41.14

Table 26: Top 10 Model-Variable Combination Rankings for IRT Retweets

Model-Variable Combination	Average Rank
(KNN: Topic)	25.00
(KNN: Link, Topic)	30.14
(KNN: Link, Hashtag, Topic, Number words)	33.57
(KNN: Link)	39.00
(KNN: Hashtag, Topic, Number words)	40.57
(KNN: Link, Sentiment, Number words)	42.26
(KNN: Link, Sentiment)	45.43
(KNN: Link, Hashtag, Sentiment, Topic, Number words)	50.29
(KNN: Sentiment, Topic)	51.43
(KNN: Hashtag, Sentiment, Topic)	52.00

In predicting IRT data, SVR with hashtag again has the best average ranking for likes, achieving an average rank of about 22 out of 186 model-variable combinations tested. However, K-Nearest Neighbors regression using BTM topic assignments has the best average ranking in predicting IRT retweets. This implies that, at least for IRT data, it may be necessary to use different models and variables in predicting retweets than for likes. However, for both OT and IRT tables, SVR and KNN seem to comprise a majority of the top 10 ranked models. All modelling results will be placed in section 11 of the appendix.

Deliverables

It would be misleading to say that we have been able to build prediction models accurate enough for commercial use in the near future. However, the contents of this paper should serve as a solid foundation for whomever next tries to solve the complexities behind predicting brand company tweet performance. The results clearly indicate that potential variable effects, as well as the best model-variable combination, vary by company, tweet category, and target variable. However, the results also indicate that for any given company attempting to build Twitter prediction models for OT data, the best jumping off point is to use Support Vector Regression with either hashtag inclusion, link inclusion, or both. For predicting IRT likes, the same general recommendation may be made. However, BTM topic modelling with K-Nearest Neighbors regression would serve as the optimal jumping off point in predicting IRT retweets of any given company.

This paper presents a novel usage of shift function analysis with applications towards Twitter data. This analysis may be used in tandem with more traditional methods to further examine how any two (or more) given distributions differ, and at times may make up for shortcomings in non-parametric tests. Overall, the findings indicate that none of the potential explanatory variables tested have uniform, universal effects. If such a variable exists, it seems unlikely that it may be created through traditional textual analytics techniques such as sentiment and topic modelling.

However, the fact that no variable's results or potential effects were found to be uniform and universal doesn't imply that results weren't significant for individual companies and tweet categories. For example, if I were the head of social media at BMW and my objective was to maximize the number of likes each official tweet receives, I would use this analysis to guide tweet format and content. I would recommend including links and hashtags in each BMW official tweet, as those are both associated with statistically significant boosts regarding the number of likes variable.

Furthermore, I would recommend for BMW to generally tweet about new cars, along with links/other resources for the average consumer (non-automobile enthusiast) to learn more about BMW's products and services. This is because BTM topic modelling uncovered that tweets belonging to this format are associated with a statistically significant boost when compared to BMW tweets belonging to other topic categories. Finally, sentiment seems to have no effect on the number of likes a BMW official tweet receives, so the BMW social media team may word the tweets however they see best fit. There's no need to take precautions to ensure that the sentiment of a BMW official tweet is labelled as negative, neutral, or positive, as they all seem to have similar performance.

Additionally, BTM topic modelling seems to have great potential as a method for structuring and examining Twitter data. As an example, here are the BTM topic modelling results for tweets belonging to the Amazon IRT category:

Table 27: Amazon IRT BTM Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	love thanks hear happy enjoy thank use share hope please	-102.58
1	send detail deliveringsmiles holiday love season like surprise please well	-48.68
2	please sorry order help detail know hear account provide like	-86.40

Despite having no prior experience with Amazon's Twitter, our team was able to identify a common theme for each of Amazon's 3 IRT tweet topic categories. Topic 0 tweets may be thought of as "positive feedback responses." These are exactly the kind of short responses uncovered during exploration of the "recipient effect," and the tweet shown in figure 11 was correctly classified as belonging to topic 0. Topic 1 tweets are very similar to topic 0 tweets, but they are generally followed by Amazon asking the recipient to 'send' their 'details' so that Amazon can 'send' them a 'surprise.' These tweets also generally seem to occur during the 'holiday season' and are often followed by the hashtag 'deliveringsmiles.' Our team has monikered topic 1 tweets as "long positive feedback responses." Finally, topic 2 tweets are your classic customer service tweets. Something has gone wrong for someone and Amazon is attempting to correct the issue. The tweet shown in figure 12 was correctly classified as belonging to topic 2.

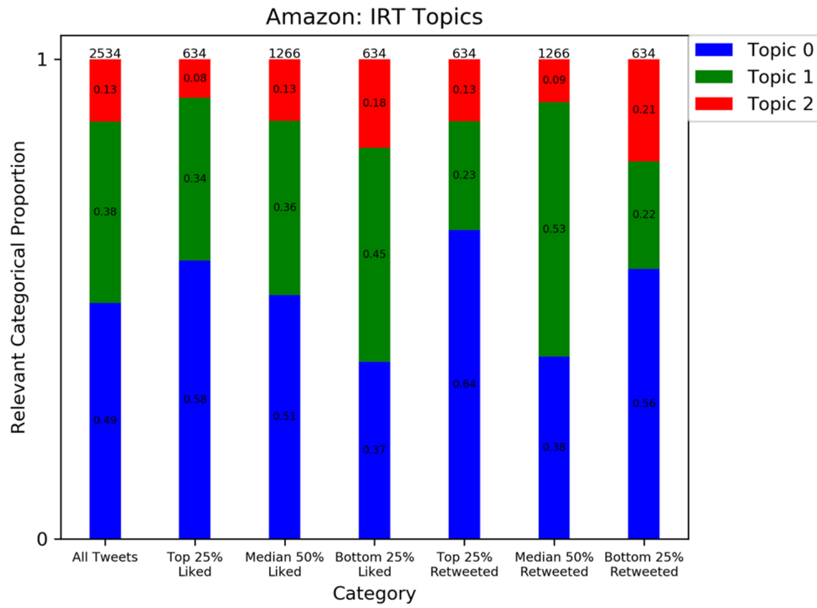
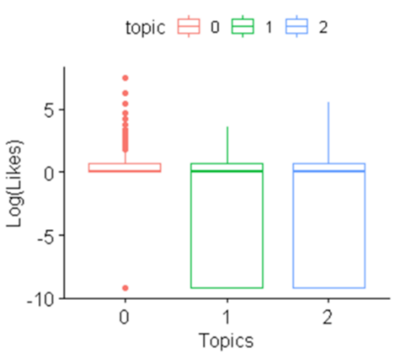
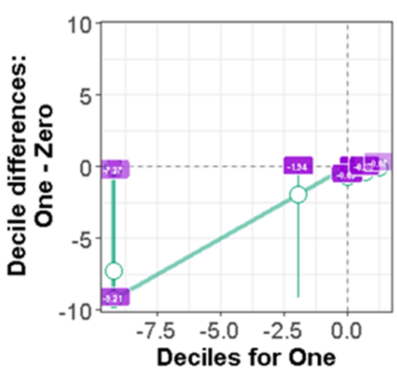


Figure 20: Amazon IRT Topic Proportions by Performance Category

In the stacked bar-charts shown in figure 20 we can see that topic 0 tweets (blue) make up approximately 49% of all Amazon IRT tweets but constitute about 58% of Amazon's most liked IRT tweets, as well as 64% of Amazon's most retweeted IRT tweets. Topic 1 tweets (green), surprisingly decrease in representation when looking at either the most liked or most retweeted Amazon IRT performance categories. Topic 2 tweets (red), unsurprisingly are over-represented within both bottom performance categories.





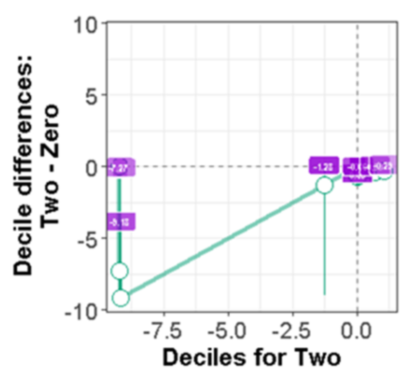


Figure 21: Amazon IRT Topic Log-Likes *Figure 22: Amazon IRT Topics 1 & 0 Shifts* *Figure 23: Amazon IRT Topics 2 & 0 Shifts*

```
$`One - Zero`
    q          One      Zero    difference      ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.0000000000 -1.825797e-10  0.000000e+00 0.0062500000   0.732
2 0.2 0.000000e+00 0.7888855 -0.7888855020 -9.999164e-01 -2.005888e-02 0.0015625000   0.000
3 0.3 7.961445e-09 1.0000000 -0.9999999920 -1.000000e+00 -9.962028e-01 0.0012500000   0.000
4 0.4 7.891276e-01 1.0000000 -0.2108724027 -9.747310e-01 -1.034521e-04 0.0010416667   0.000
5 0.5 1.000000e+00 1.0008158 -0.0008157556 -3.993929e-01 -7.188884e-10 0.0008928571   0.000
6 0.6 1.000000e+00 1.9999620 -0.9999615301 -1.000000e+00 -8.533926e-01 0.0007812500   0.000
7 0.7 1.950473e+00 2.0000000 -0.0495267632 -9.523744e-01 -2.395845e-06 0.0006944444   0.000
8 0.8 2.003960e+00 2.9503309 -0.9463708415 -9.998867e-01  5.422125e-02 0.0020833333   0.005
9 0.9 3.562680e+00 3.8117935 -0.2491131489 -9.625729e-01  8.191151e-01 0.0031250000   0.598

$`Two - Zero`
    q          Two      Zero    difference      ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.0000000000 -1.825797e-10  0.000000e+00 0.0031250000   0.741
2 0.2 7.213563e-12 0.7888855 -0.7888855020 -9.999564e-01 -1.702210e-02 0.0010416667   0.000
3 0.3 3.278985e-03 1.0000000 -0.9967210148 -1.000000e+00 -2.113664e-01 0.0007812500   0.000
4 0.4 8.607182e-01 1.0000000 -0.1392817533 -9.937108e-01 -2.442152e-05 0.0006250000   0.000
5 0.5 9.999991e-01 1.0008158 -0.0008166721 -6.006074e-01 -1.503621e-08 0.0005208333   0.000
6 0.6 1.000000e+00 1.9999620 -0.9999620160 -1.000000e+00 -7.229880e-01 0.0004464286   0.000
7 0.7 1.046199e+00 2.0000000 -0.9538011890 -1.000081e+00 -1.086811e-01 0.0003906250   0.000
8 0.8 1.994057e+00 2.9503309 -0.9562740504 -1.739037e+00 -8.124780e-02 0.0003472222   0.000
9 0.9 2.871102e+00 3.8117935 -0.9406919763 -1.950072e+00  9.086744e-02 0.0015625000   0.010
```

Figure 24: Amazon IRT Topic Shift Function Confidence Intervals

Figure 22 shows a visualization produced during shift function analysis when comparing the distributions of likes between topic 0 and topic 1 tweets above. The first quantile of topic 0 tweets would need to be shifted down by an insignificant margin to match the first quantile of topic 1 tweets. This may be confirmed by noting that the confidence interval for the first quantile between topic 1 and topic 0 tweets (shown in figure 24) contains the value 0, implying that it's entirely possible no shift is required for these two quantiles to match.

The second quantile of topic 0 tweets, however, would need to be shifted down by a significant margin to match the second quantile of topic 1 tweets, while the third quantile would need to be shifted down by an even larger degree. Past that point, every shift is still significant, but decreases in magnitude for each next quantile. The same may be said when comparing topic 2 and topic 0 tweets in figure 23. These results indicate that, surprisingly, it may not be worth the additional effort to create 'long positive feedback responses.' Recipients respond worse to these than regular "positive feedback responses" by a statistically significant margin, and topic 1 tweets have similar performance to topic 2 tweets – the classic customer service responses. Our team holds the belief that using the BTM topic estimation process outlined in this paper, followed by discussions with the company's social media team to further refine results, could offer great potential as a method for examining Twitter data to uncover actionable insights.

Self-Assessment

Personally, my primary individual learning objective was to advance my proficiency in working with textual data. I of course had to rely heavily on techniques learned when taking Textual Analytics, but I also had to learn a great deal about NLP on my own for this project. To be honest, I'm not sure that in between finishing Textual Analytics and beginning this project I had done one thing that could be considered textual analytics. Furthermore, I think that the one and only DSA core class that I didn't necessarily make use of in this project was Bayesian Statistics. Not to say that I wouldn't have liked to in some fashion, but I can't identify any elements of the project having inspiration from Bayesian Statistics. Overall, I certainly accomplished my personal objective of becoming proficient in analyzing textual data. Furthermore, I learned quite a bit along the way regarding various data science project components such as organization, time-management, and planning. Finally, this was a 4-hour credit course for me, and it was a research project. The work was sponsored by Dr. Doyle Yoon (Associate Professor at the Gaylord College of Journalism and Mass Communication, Associate Professor at the Data Science and Analytics in the Gallogly College of Engineering, dyoon@ou.edu).

References

Anita, O. (2019, December 23). Evaluating Unsupervised Sentiment Analysis Tools Using Labeled Data. Retrieved March 22, 2021, from https://heartbeat.fritz.ai/evaluating-unsupervised-sentiment-analysis-tools-using-labeled-data-8d4bb1336cee

Daga, I., Gupta, A., Vardhan, R., & Mukherjee, P. (2020, May 13). Prediction of likes and retweets using text information retrieval. Retrieved April 28, 2021, from https://www.sciencedirect.com/science/article/pii/S1877050920304129

Doksum, K. A., & Sievers, G. L. (1976, December). *Plotting with Confidence: Graphical Comparisons of Two Populations* [Scholarly project]. In *JSTOR*. Retrieved April 27, 2021, from http://pages.stat.wisc.edu/~doksum/papers/confidence_plots_1976.pdf

Feng, R., Hou, Y., Mahmood, Z., & Zeng, B. (n.d.). Predicting Popularity. Retrieved April 28, 2021, from http://belindazeng.github.io/goingviral/

Jonsson, E., & Stolee, J. (n.d.). *An Evaluation of Topic Modelling Techniques for Twitter* [Scholarly project]. Retrieved from https://www.cs.toronto.edu/~jstolee/projects/topic.pdf

Kapadia, S. (2020, December 29). Evaluate Topic Models: Latent Dirichlet Allocation (LDA). Retrieved April 27, 2021, from https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0

Porutiu, H. (n.d.). An Introduction to Watson Natural Language Processing. Retrieved April 26, 2021, from https://developer.ibm.com/technologies/artificial-intelligence/articles/introduction-to-watson-natural-language-processing/

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2017, January 01). Beyond differences In means: Robust graphical methods to compare two groups in neuroscience. Retrieved April 28, 2021, from https://www.biorxiv.org/content/10.1101/121079v2.full

Statista. (2021, January 27). *Twitter: Number of monthly active users 2010-2019*. https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Yan, Xiaohui & Guo, Jiafeng & Lan, Yanyan & Cheng, Xueqi. (2013). A biterm topic model for short texts. WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web. 1445-1456. 10.1145/2488388.2488514.

Appendix

Section 1: Correlation Scatterplots

Section 1 of the appendix contains scatterplots visualizing the relationship between retweets and likes for all companies and tweet categories under consideration.

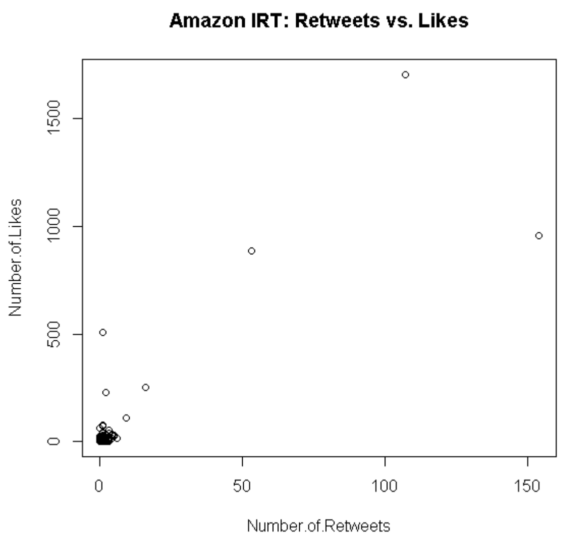


Figure 25.1: Scatterplot of Amazon IRT Retweets vs. Likes

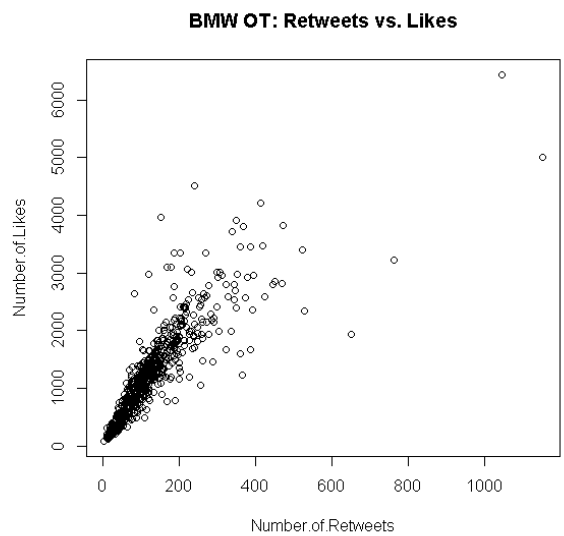


Figure 1.2: Scatterplot of BMW OT Retweets vs. Likes

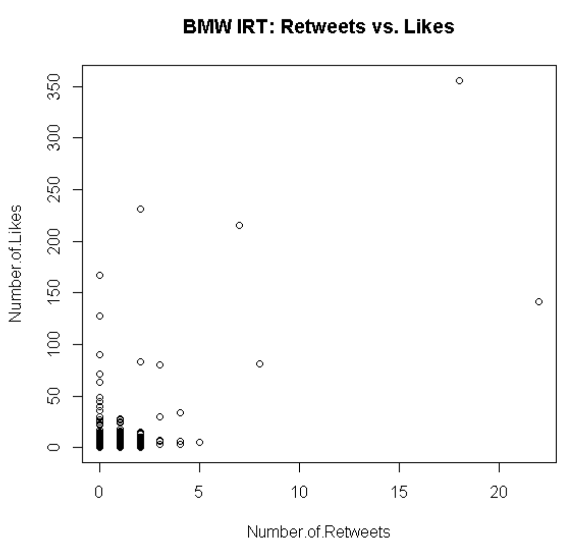


Figure 1.3: Scatterplot of BMW IRT Retweets vs. Likes

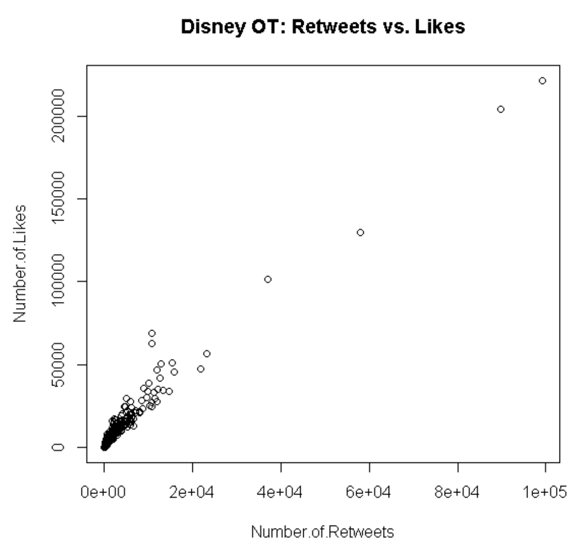


Figure 1.4: Scatterplot of Disney OT Retweets vs. Likes

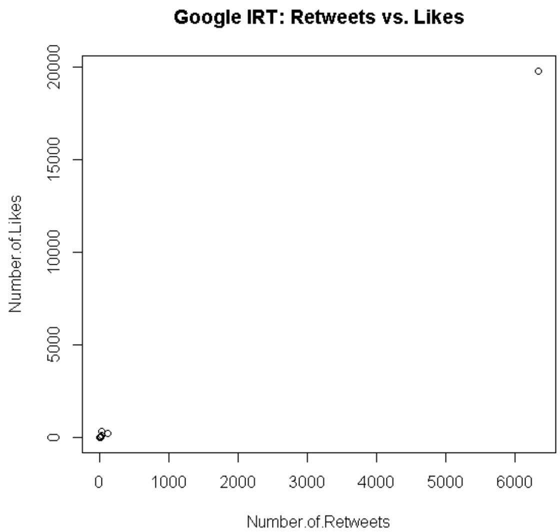


Figure 1.5: Scatterplot of Google IRT Retweets vs. Likes

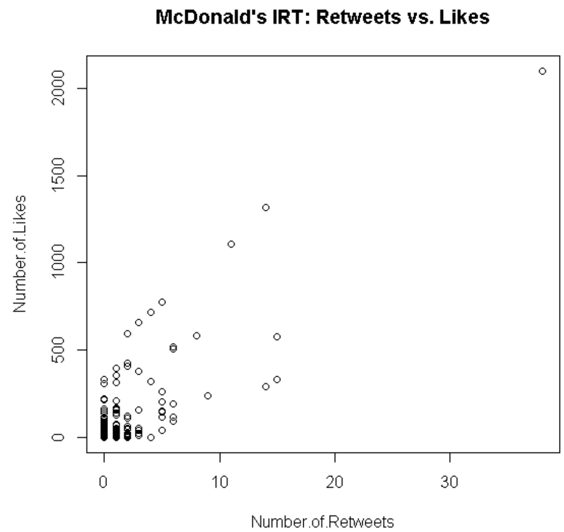


Figure 1.6: Scatterplot of McDonald's IRT Retweets vs. Likes

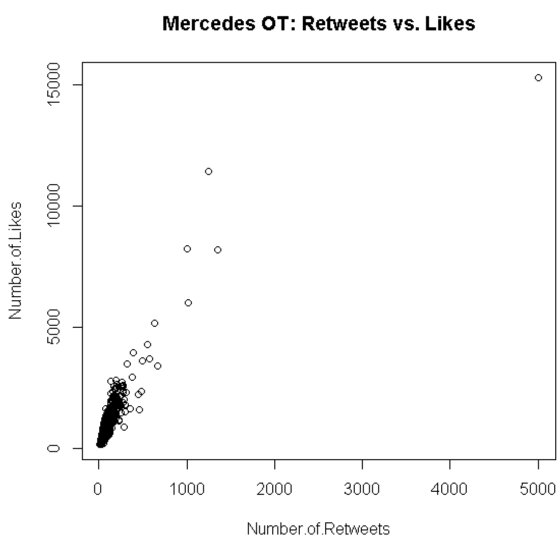


Figure 1.7: Scatterplot of Mercedes OT Retweets vs. Likes

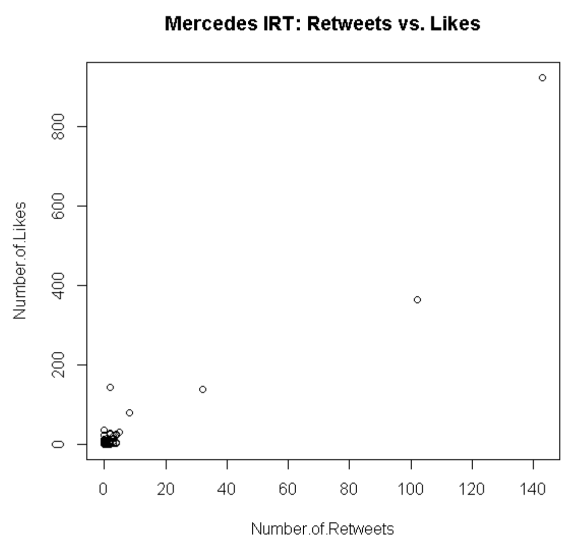


Figure 1.8: Scatterplot of Mercedes IRT Retweets vs. Likes

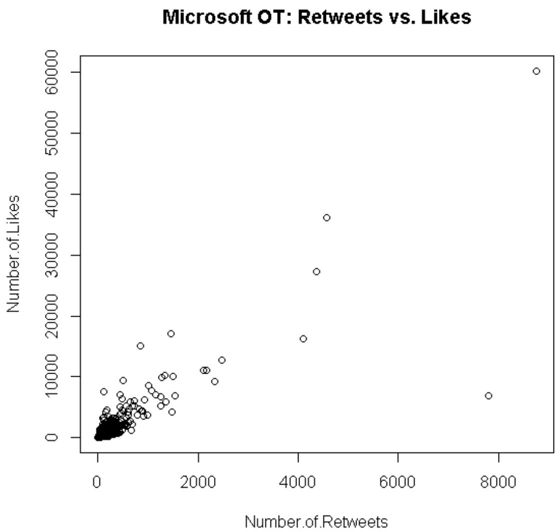


Figure 1.9: Scatterplot of Microsoft OT Retweets vs. Likes

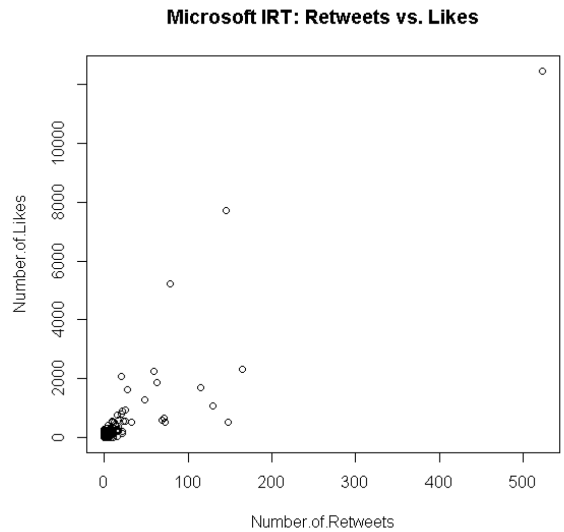


Figure 1.10: Scatterplot of Microsoft IRT Retweets vs. Likes

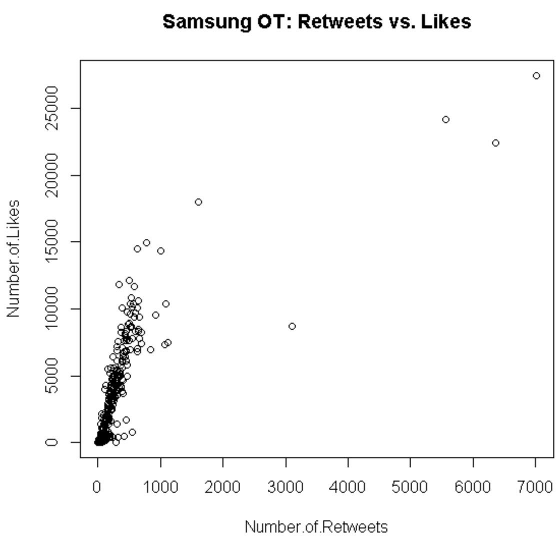


Figure 1.11: Scatterplot of Samsung OT Retweets vs. Likes

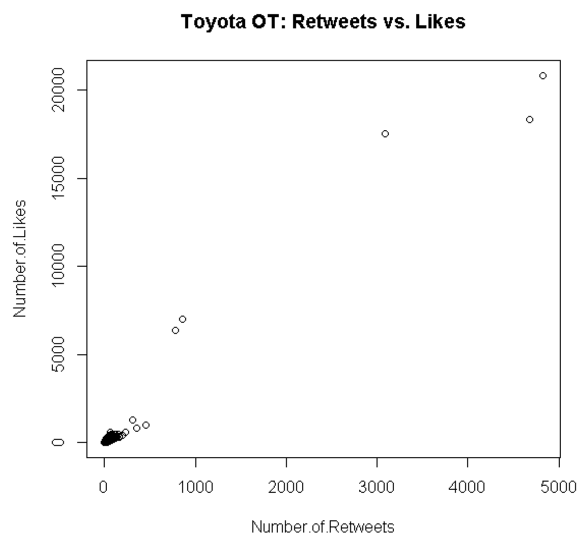


Figure 1.12: Scatterplot of Toyota OT Retweets vs. Likes

<u>Section 2:</u> Wordcount Histograms

Section 2 of the appendix contains histograms representing the number of words per tweet, for all companies and tweet categories under consideration. Furthermore, the ‘performance categories’ are based on the number of likes each tweet receives.

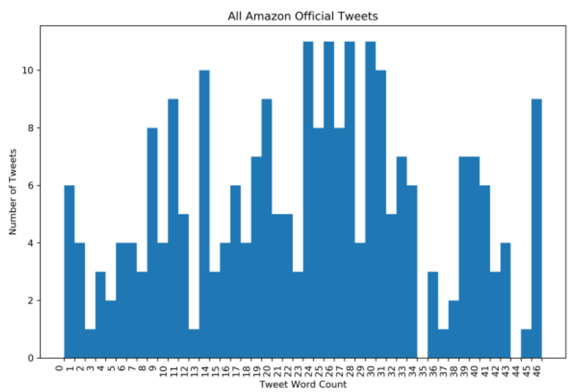


Figure 26.1: Amazon OT Wordcounts

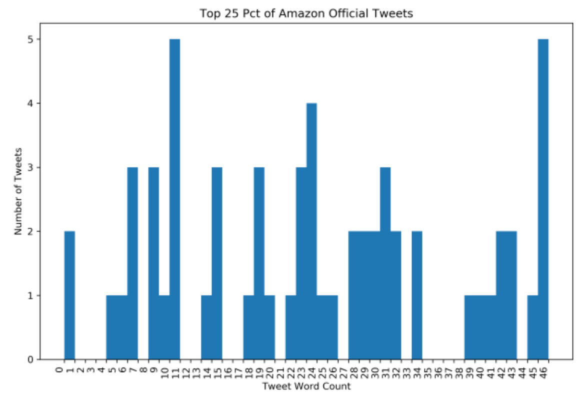


Figure 2.2: Amazon Top OT Wordcounts

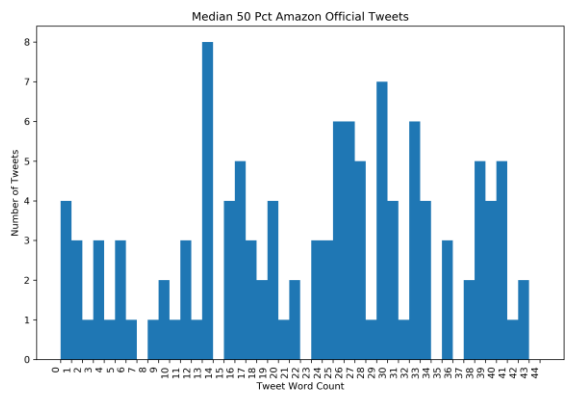


Figure 2.3: Amazon Median OT Wordcounts

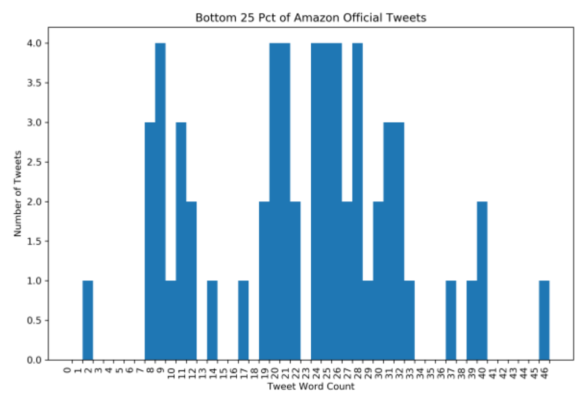


Figure 2.4: Amazon Bottom OT Wordcounts

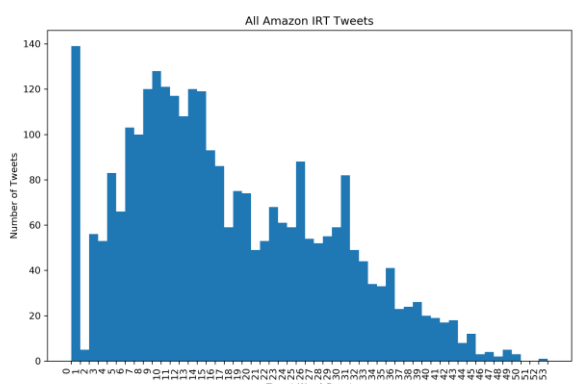


Figure 2.5: Amazon IRT Wordcounts

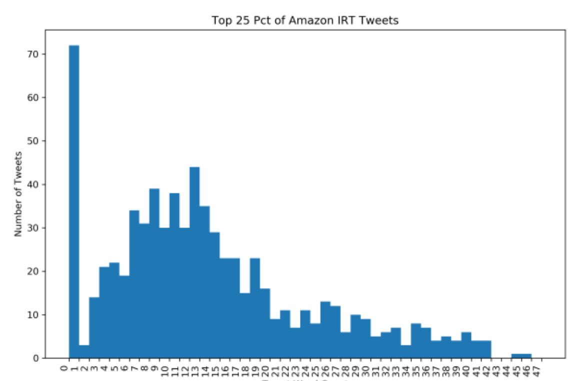


Figure 2.6: Amazon Top IRT Wordcounts

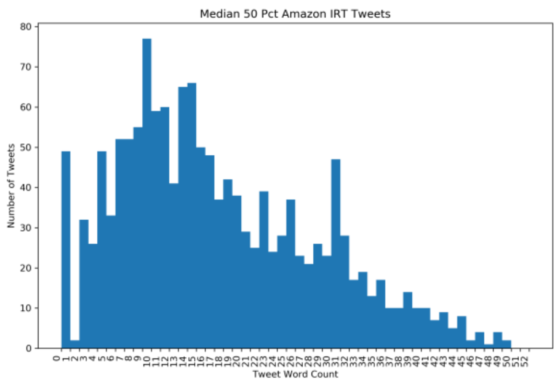


Figure 2.7: Amazon Median IRT Wordcounts

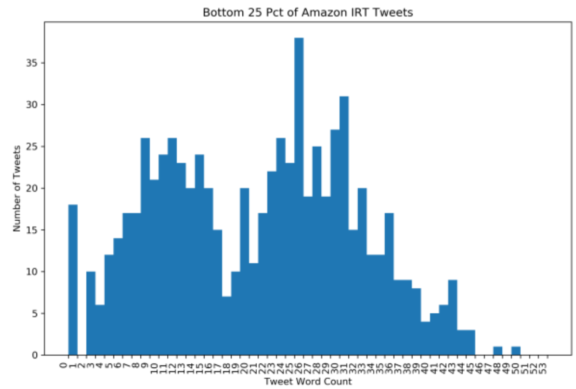


Figure 2.8: Amazon Bottom IRT Wordcounts

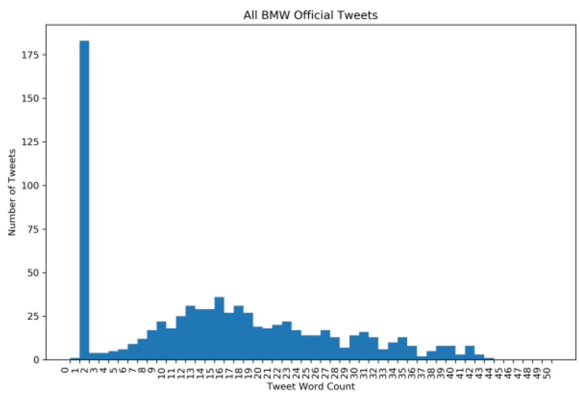


Figure 2.9: BMW OT Wordcounts

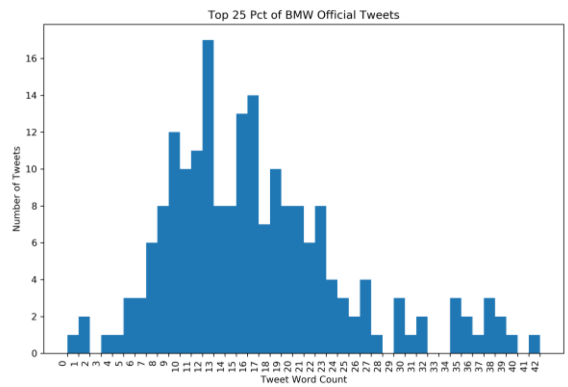


Figure 2.10: BMW Top OT Wordcounts

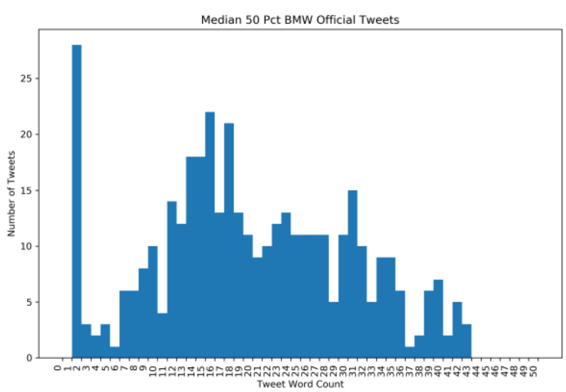


Figure 2.11: BMW Median OT Wordcounts

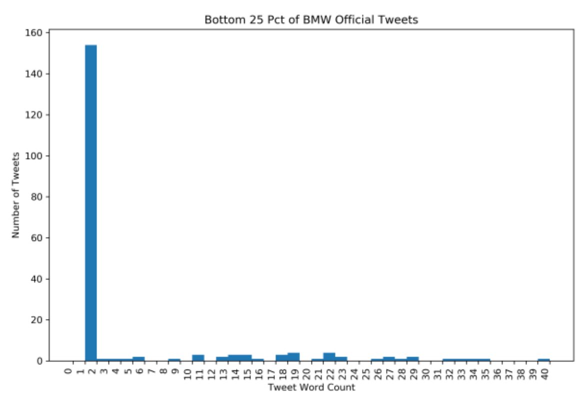


Figure 2.12: BMW Bottom OT Wordcounts

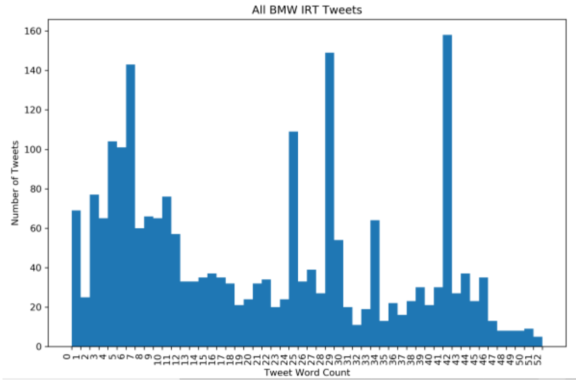


Figure 2.13: BMW IRT Wordcounts

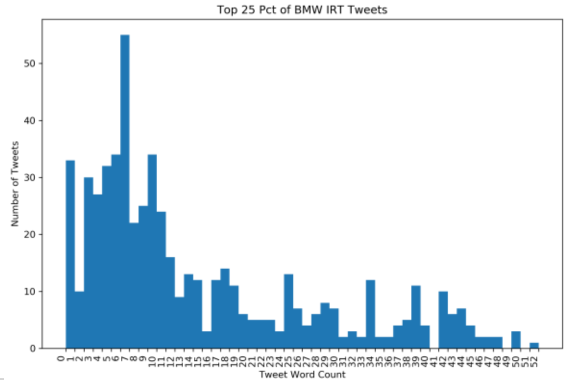


Figure 2.14: BMW Top IRT Wordcounts

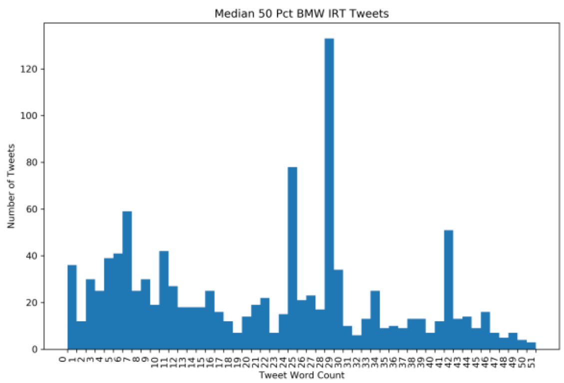


Figure 2.15: BMW Median IRT Wordcounts

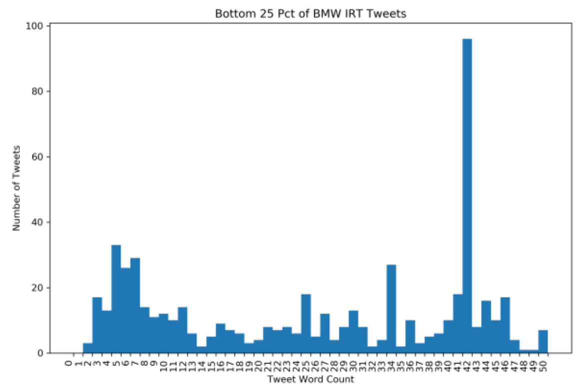


Figure 2.16: BMW Bottom IRT Wordcounts

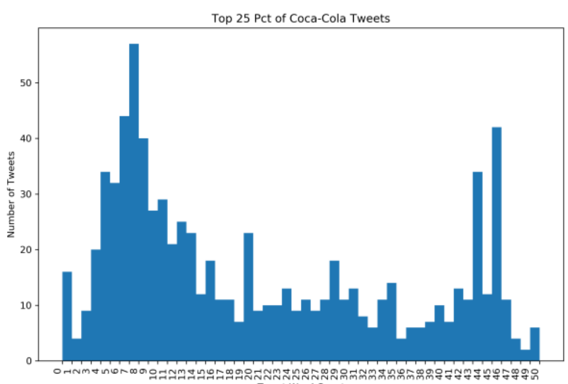


Figure 2.17: Coca Cola Top IRT Wordcounts

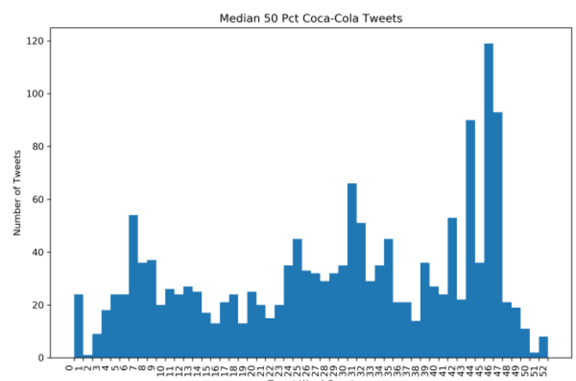


Figure 2.18: Coca Cola Median IRT Wordcounts

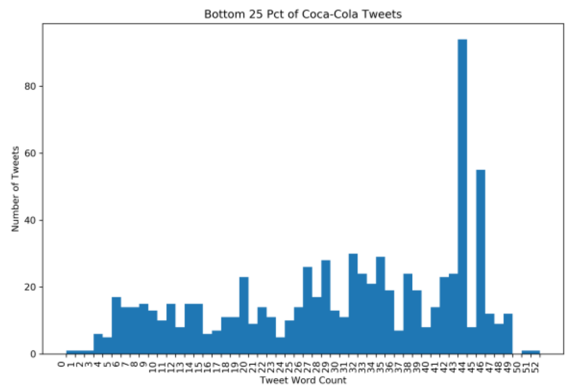


Figure 2.19: Coca Cola Bottom IRT Wordcounts

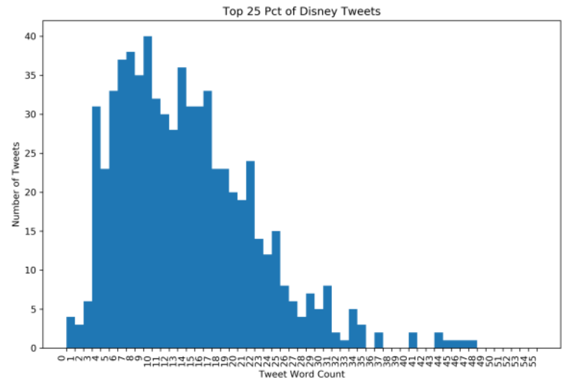


Figure 2.20: Disney Top OT Wordcounts

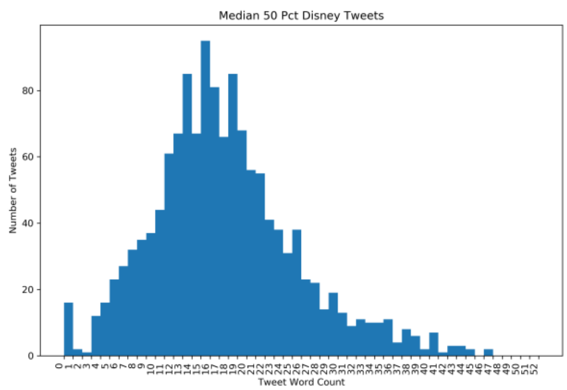


Figure 2.21: Disney Median OT Wordcounts

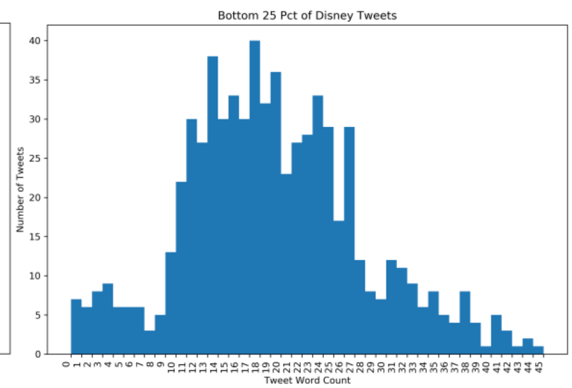


Figure 2.22: Disney Bottom OT Wordcounts

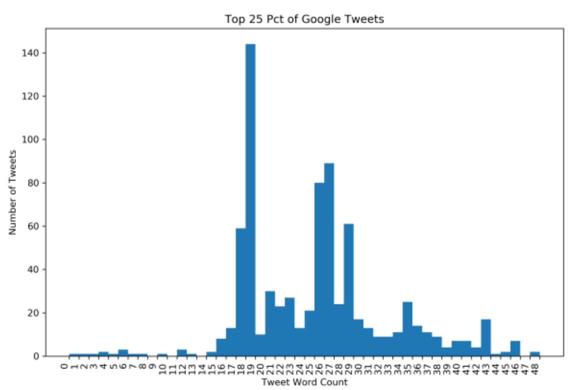


Figure 2.23: Google Top IRT Wordcounts

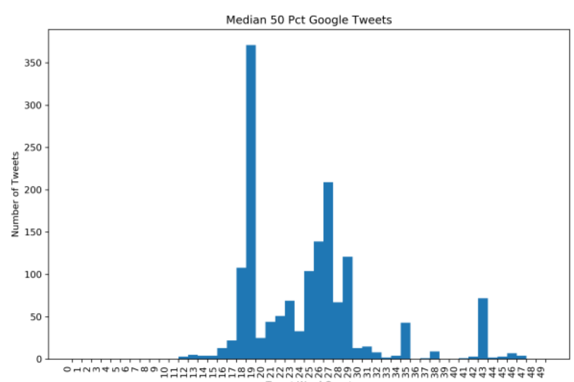


Figure 2.24: Google Median IRT Wordcounts

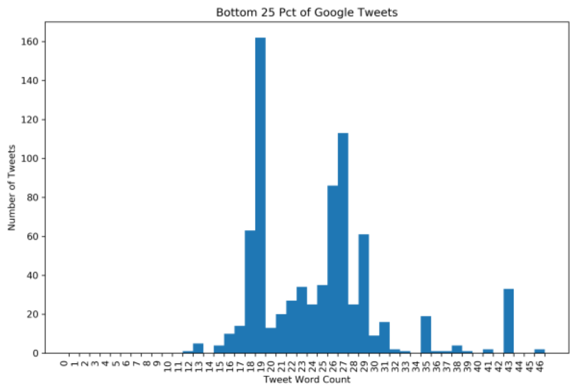


Figure 2.25: Google Bottom IRT Wordcounts

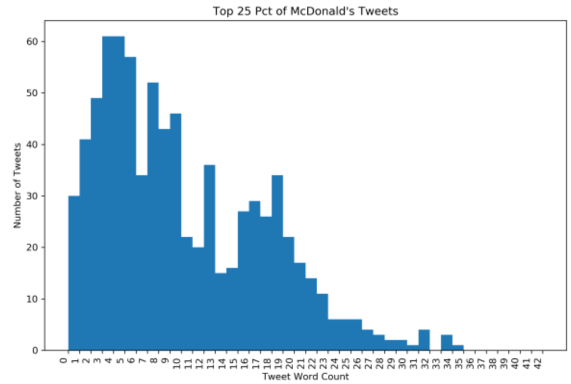


Figure 2.26: McDonald's Top IRT Wordcounts

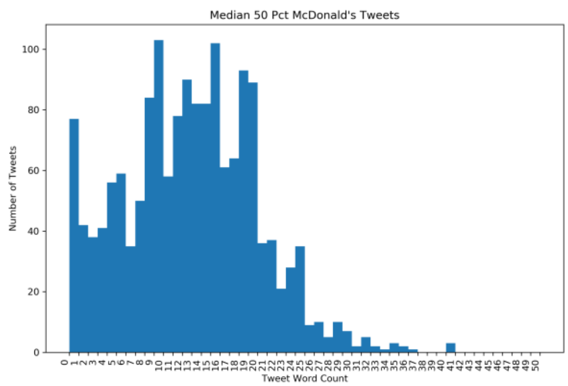


Figure 2.27: McDonald's Median IRT Wordcounts

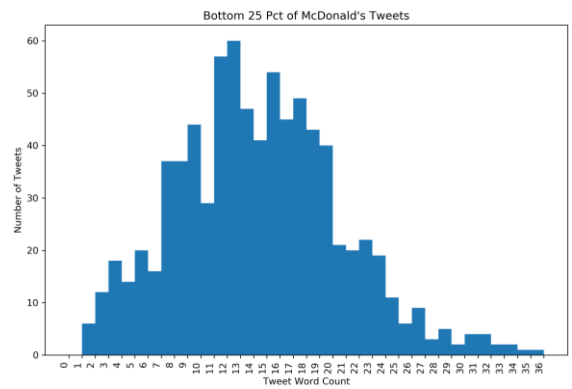


Figure 2.28: McDonald's Bottom IRT Wordcounts

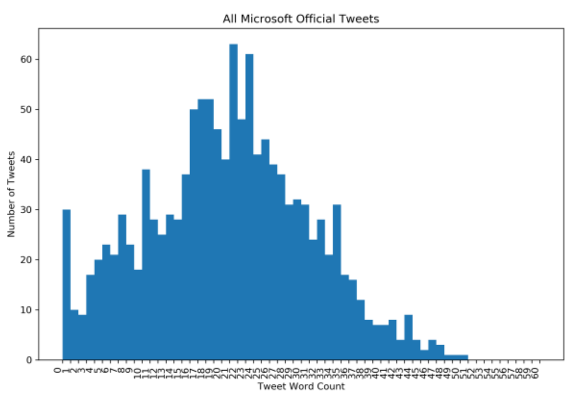


Figure 2.29: Microsoft OT Wordcounts

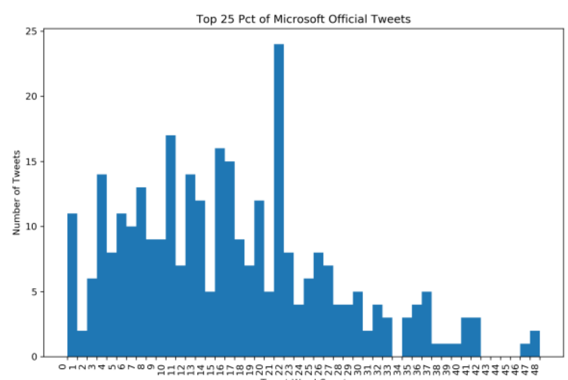


Figure 2.30: Microsoft Top OT Wordcounts

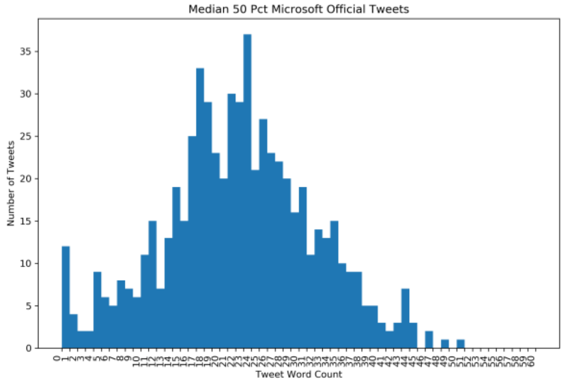


Figure 2.31: Microsoft Median OT Wordcounts

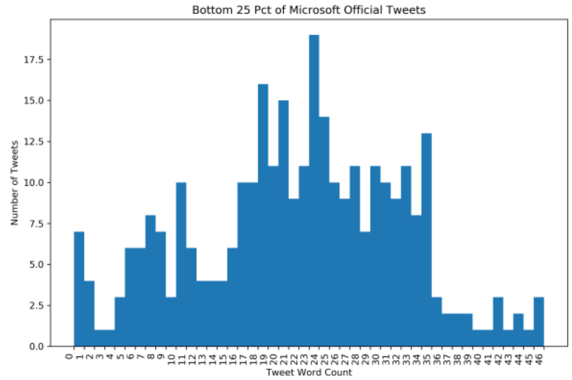


Figure 2.32: Microsoft Bottom OT Wordcounts

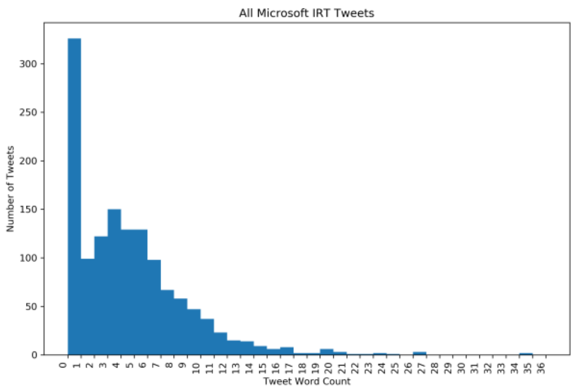


Figure 2.33: Microsoft IRT Wordcounts

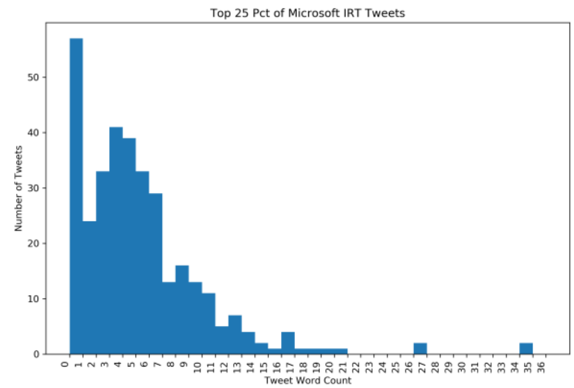


Figure 2.34: Microsoft Top IRT Wordcounts

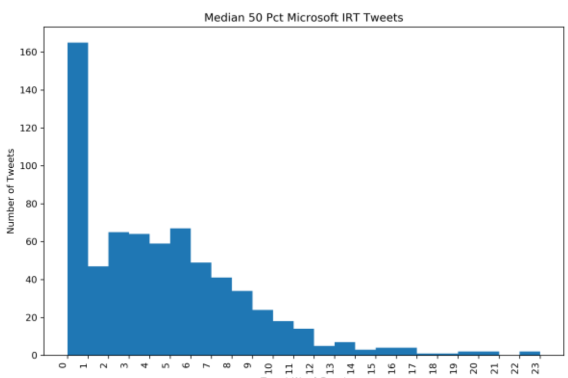


Figure 2.35: Microsoft Median IRT Wordcounts

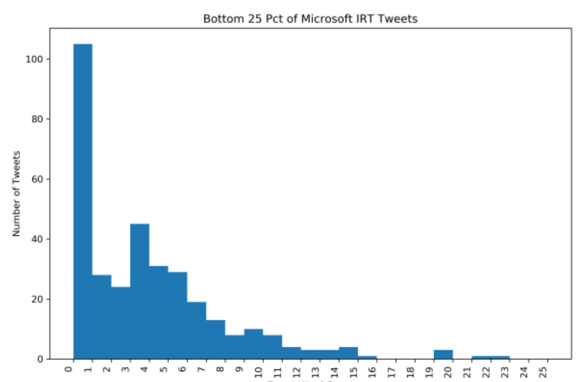


Figure 2.36: Microsoft Bottom IRT Wordcounts

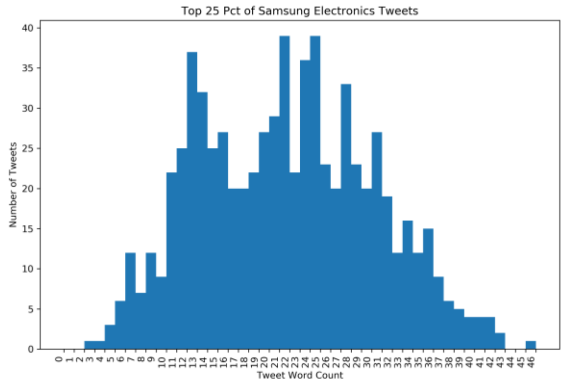


Figure 2.37: Samsung Top OT Wordcounts

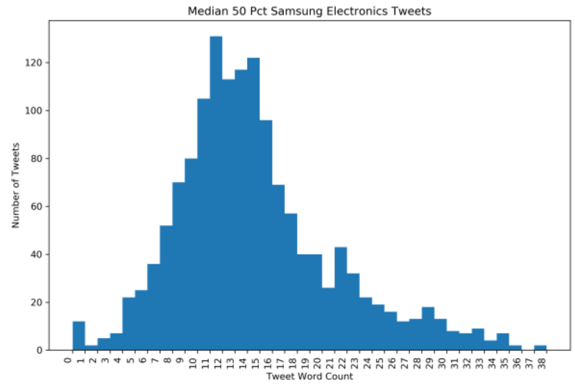


Figure 2.38: Samsung Median OT Wordcounts

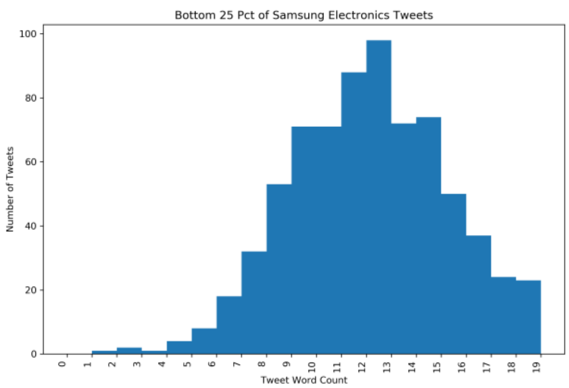


Figure 2.39: Samsung Bottom OT Wordcounts

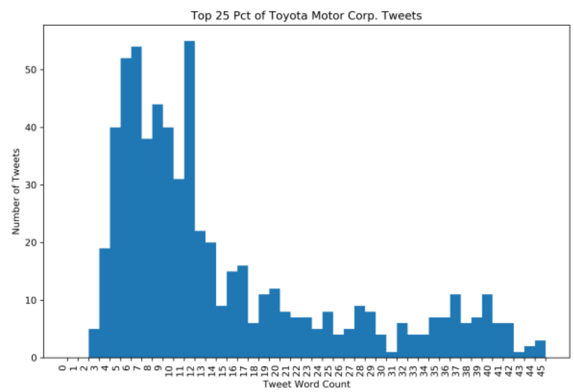


Figure 2.40: Toyota Top OT Wordcounts

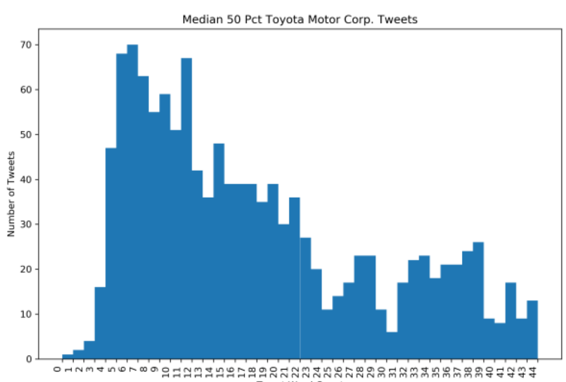


Figure 2.41: Toyota Median OT Wordcounts

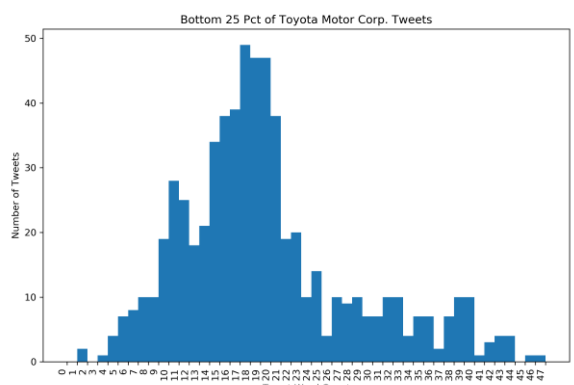


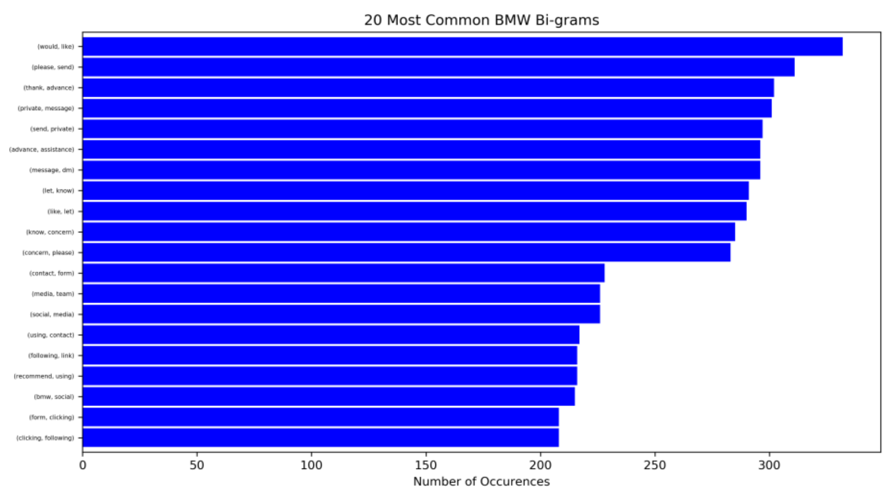
Figure 2.42: Toyota Bottom OT Wordcounts

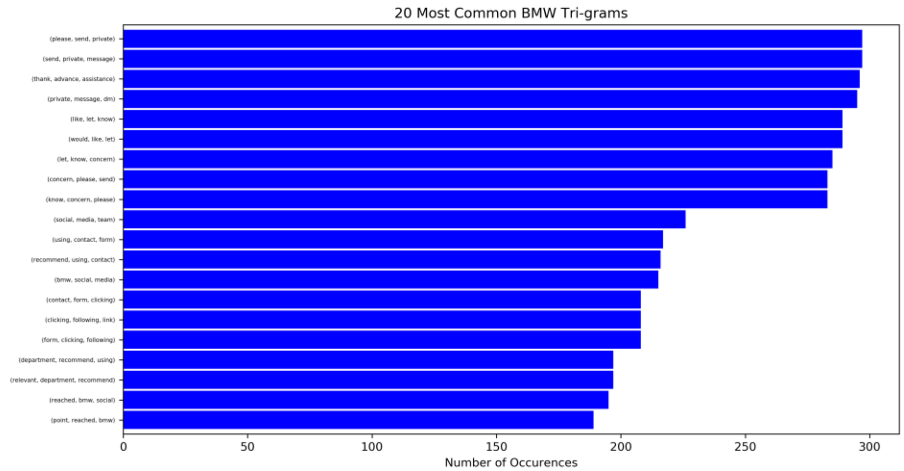
<u>Section 3:</u> Common Words, Bigrams and Trigrams

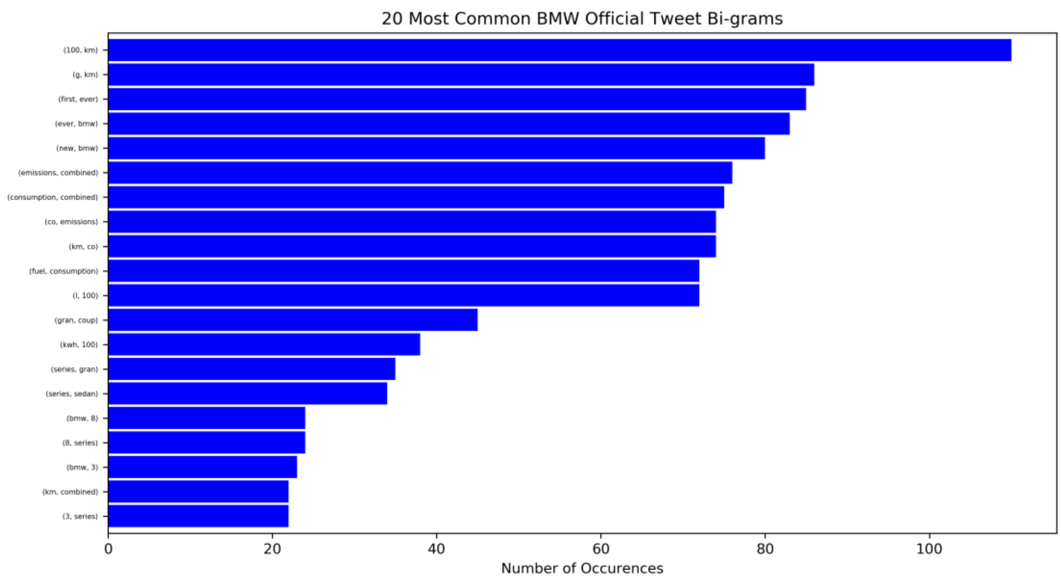
Section 3 of the appendix contains wordcount descriptive statistics, and the top 20 commonly occurring bigrams/trigrams for all companies and tweet categories under consideration. Furthermore, this section contains the top 20 commonly occurring words for each company, as well as the words uniquely common to each tweet/performance category.

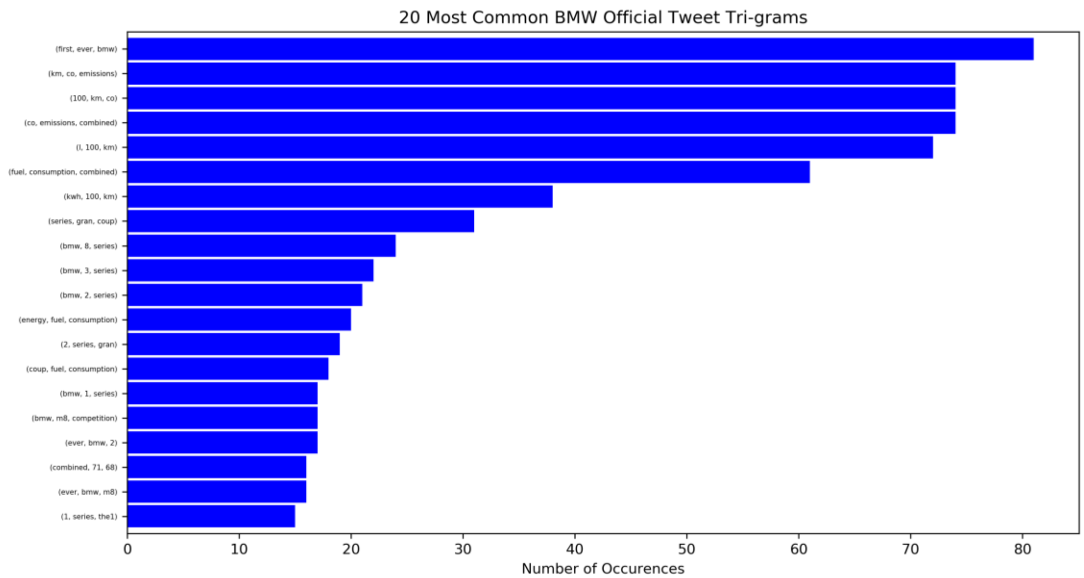
BMW

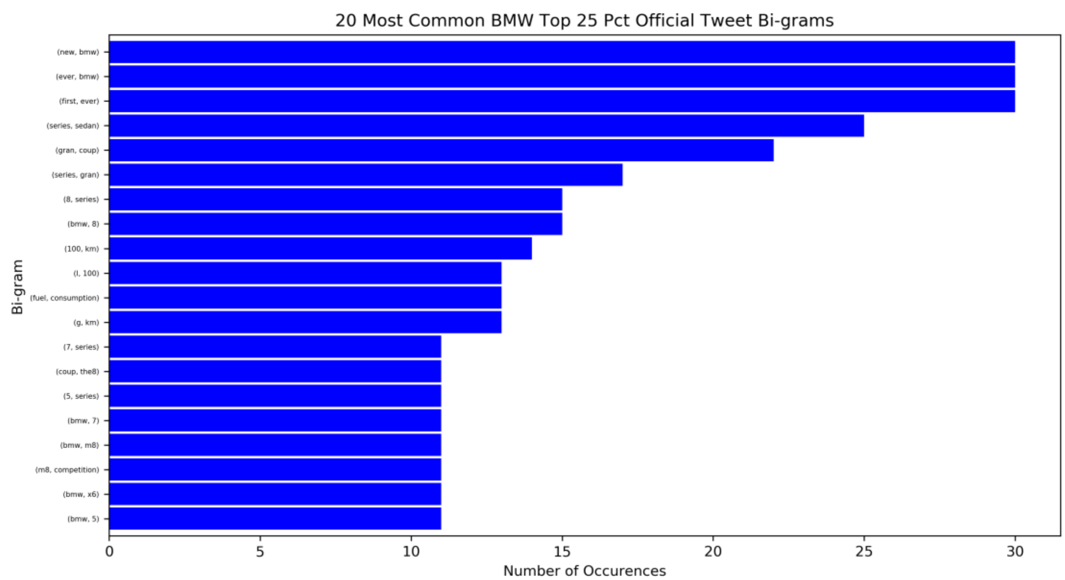
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	3076	18.85	57994	3505
Official Tweets:	796	15.18	12085	2410
Top 25% Official:	199	16.44	3271	965
Median 50% Official:	398	19.58	7791	1754
Bottom 25% Official:	199	5.10	1014	546
IRT Tweets:	2280	20.14	45909	2006
Top 25% IRT:	570	13.99	7972	1177
Median 50% IRT:	1140	20.82	23732	1396
Bottom 25% IRT:	570	24.86	14171	908

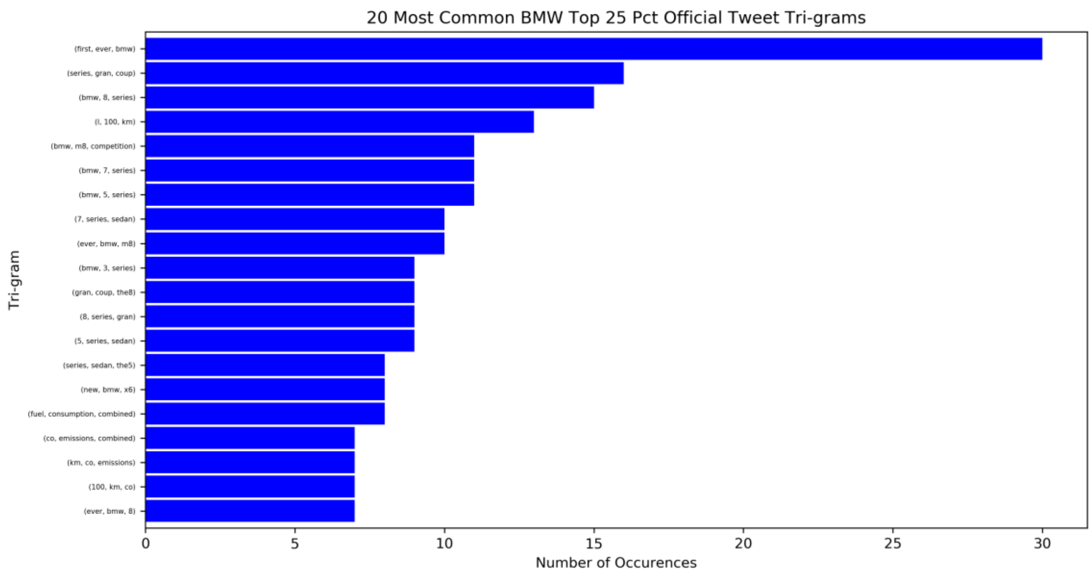


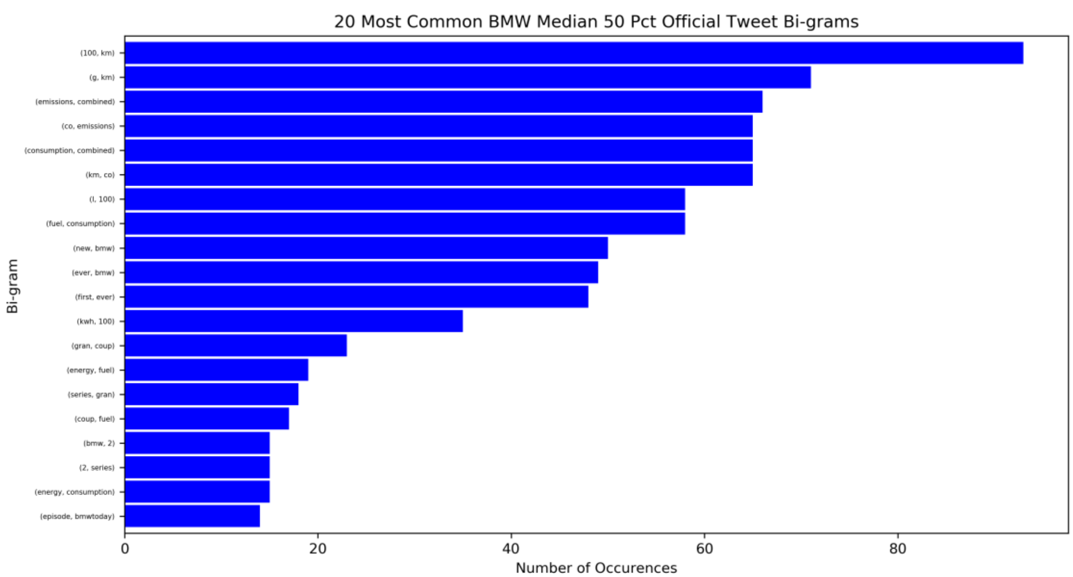

20 Most Common BMW Tri-grams
(please, send, private)
(send, private, message)
(thank, advance, assistance)
(private, message, dm)
(like, let, know)
(would, like, let)
(let, know, concern)
(concern, please, send)
(know, concern, please)
(social, media, team)
(using, contact, form)
(recommend, using, contact)
(bmw, social, media)
(contact, form, clicking)
(clicking, following, link)
(form, clicking, following)
(department, recommend, using)
(relevant, department, recommend)
(reached, bmw, social)
(point, reached, bmw)
0
50
100
150
200
250
300
Number of Occurences

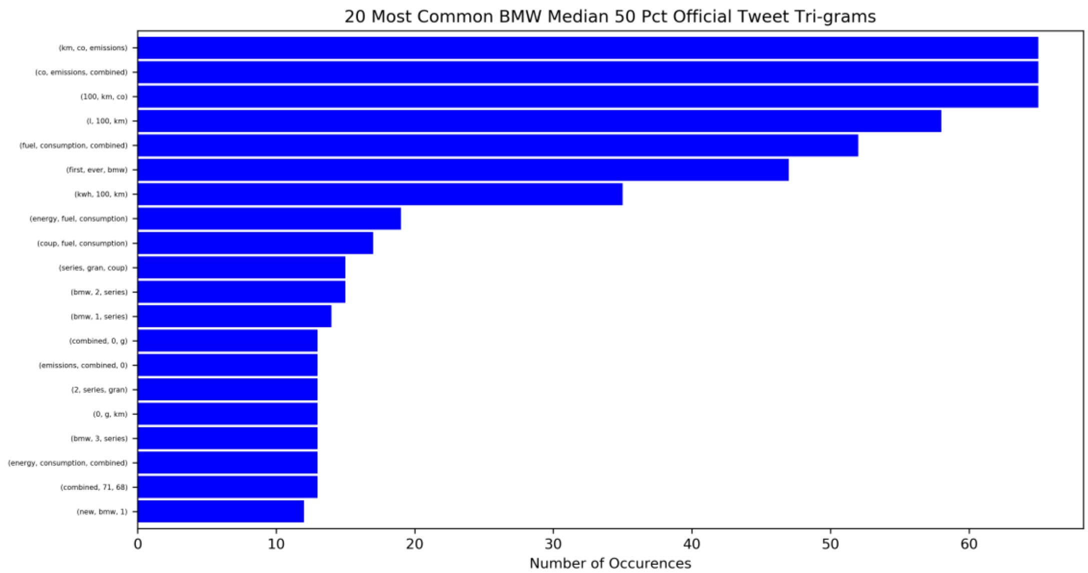

20 Most Common BMW Official Tweet Bi-grams
(100, km)
(g, km)
(first, ever)
(ever, bmw)
(new, bmw)
(emissions, combined)
(consumption, combined)
(co, emissions)
(km, co)
(fuel, consumption)
(l, 100)
(gran, coup)
(kwh, 100)
(series, gran)
(series, sedan)
(bmw, 8)
(8, series)
(bmw, 3)
(km, combined)
(3, series)
0
20
40
60
80
100
Number of Occurences

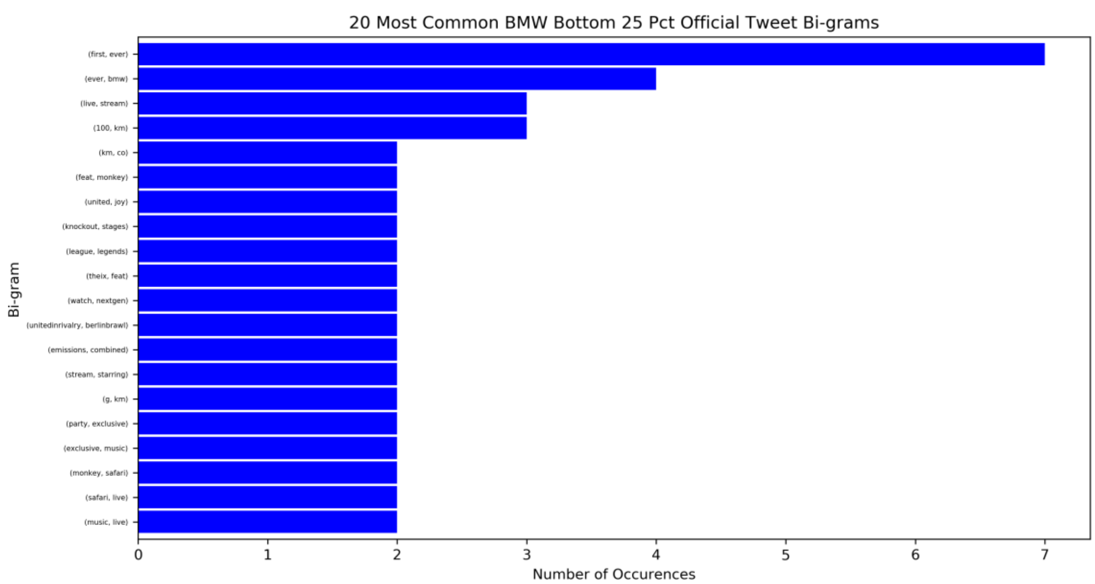

20 Most Common BMW Official Tweet Tri-grams
(first, ever, bmw)
(km, co, emissions)
(100, km, co)
(co, emissions, combined)
(l, 100, km)
(fuel, consumption, combined)
(kwh, 100, km)
(series, gran, coup)
(bmw, 8, series)
(bmw, 3, series)
(bmw, 2, series)
(energy, fuel, consumption)
(2, series, gran)
(coup, fuel, consumption)
(bmw, 1, series)
(bmw, m8, competition)
(ever, bmw, 2)
(combined, 71, 68)
(ever, bmw, m8)
(1, series, the1)
0
10
20
30
40
50
60
70
80
Number of Occurences

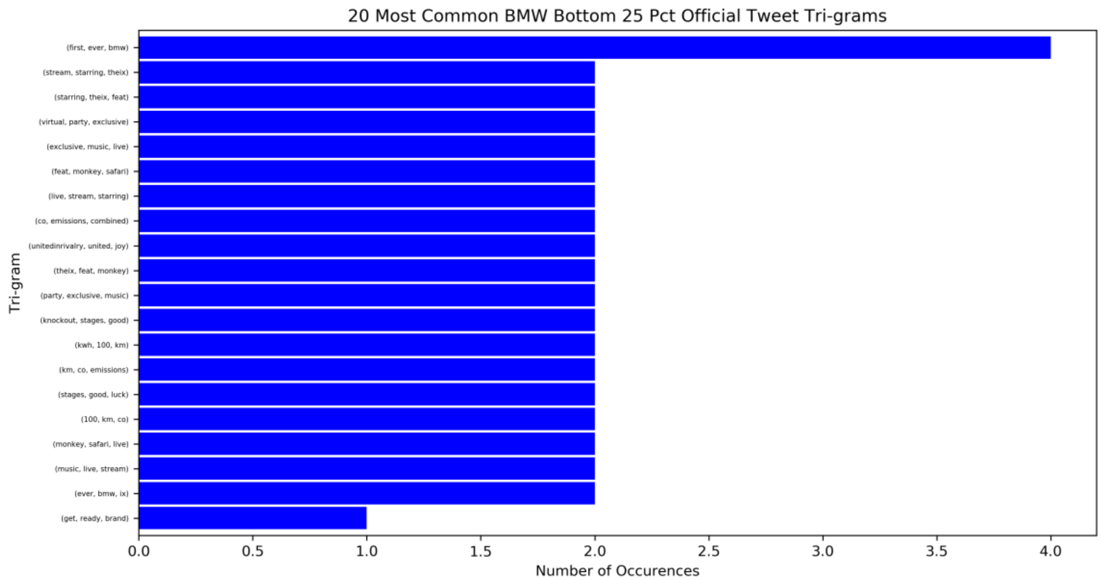

20 Most Common BMW Top 25 Pct Official Tweet Bi-grams
Bi-gram
(new, bmw)
(ever, bmw)
(first, ever)
(series, sedan)
(gran, coup)
(series, gran)
(8, series)
(bmw, 8)
(100, km)
(l, 100)
(fuel, consumption)
(g, km)
(7, series)
(coup, the8)
(5, series)
(bmw, 7)
(bmw, m8)
(m8, competition)
(bmw, x6)
(bmw, 5)
0
5
10
15
20
25
30
Number of Occurences

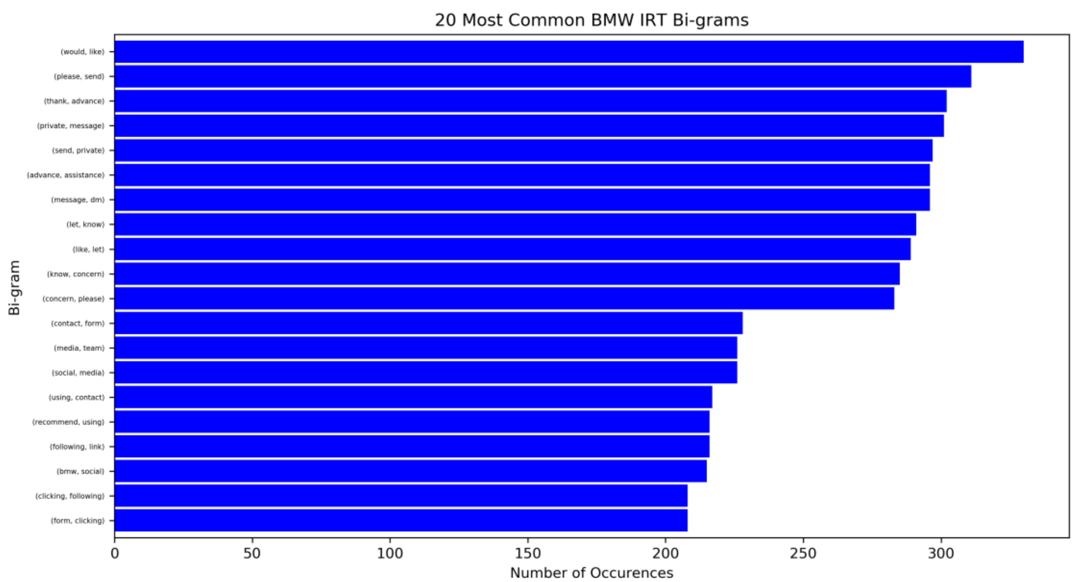

20 Most Common BMW Top 25 Pct Official Tweet Tri-grams
(first, ever, bmw)
(series, gran, coup)
(bmw, 8, series)
(l, 100, km)
(bmw, m8, competition)
(bmw, 7, series)
(bmw, 5, series)
(7, series, sedan)
(ever, bmw, m8)
(bmw, 3, series)
(gran, coup, the8)
(8, series, gran)
(5, series, sedan)
(series, sedan, the5)
(new, bmw, x6)
(fuel, consumption, combined)
(co, emissions, combined)
(km, co, emissions)
(100, km, co)
(ever, bmw, 8)
Tri-gram
0
5
10
15
20
25
30
Number of Occurences

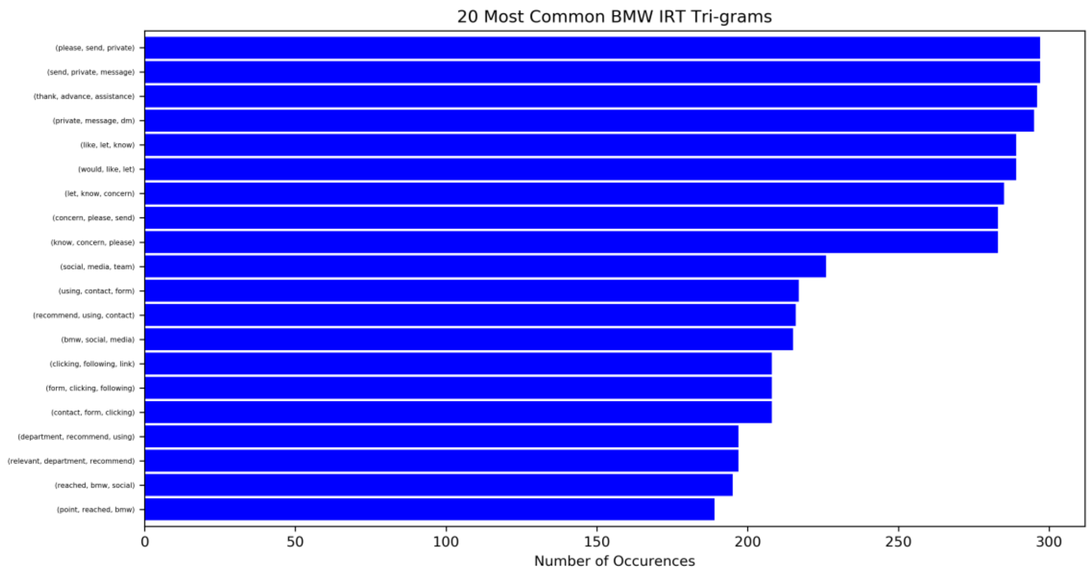

20 Most Common BMW Median 50 Pct Official Tweet Bi-grams
(100, km)
(g, km)
(emissions, combined)
(co, emissions)
(consumption, combined)
(km, co)
(l, 100)
(fuel, consumption)
(new, bmw)
(ever, bmw)
(first, ever)
(kwh, 100)
(gran, coup)
(energy, fuel)
(series, gran)
(coup, fuel)
(bmw, 2)
(2, series)
(energy, consumption)
(episode, bmwtoday)
Bi-gram
0
20
40
60
80
Number of Occurences

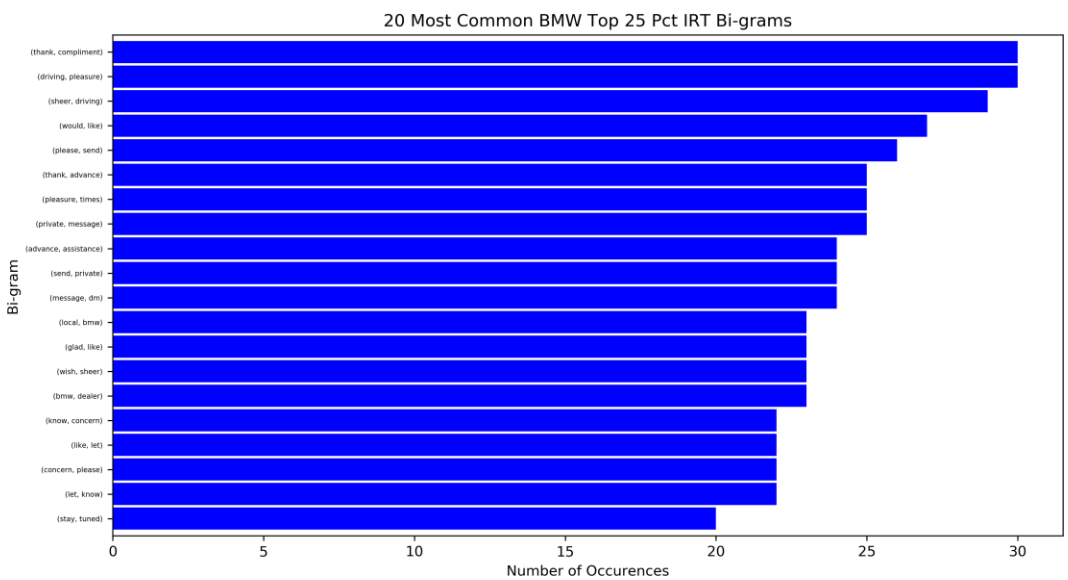

20 Most Common BMW Median 50 Pct Official Tweet Tri-grams
(km, co, emissions)
(co, emissions, combined)
(100, km, co)
(l, 100, km)
(fuel, consumption, combined)
(first, ever, bmw)
(kwh, 100, km)
(energy, fuel, consumption)
(coup, fuel, consumption)
(series, gran, coup)
(bmw, 2, series)
(bmw, 1, series)
(combined, 0, g)
(emissions, combined, 0)
(2, series, gran)
(0, g, km)
(bmw, 3, series)
(energy, consumption, combined)
(combined, 71, 68)
(new, bmw, 1)
0
10
20
30
40
50
60
Number of Occurences

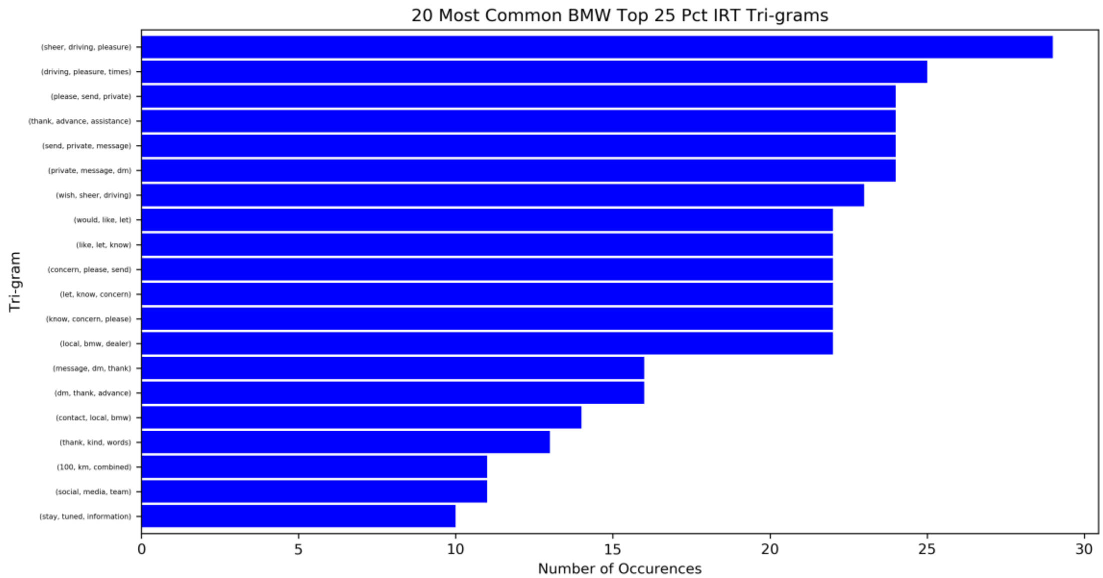

20 Most Common BMW Bottom 25 Pct Official Tweet Bi-grams
Bi-gram
(first, ever)
(ever, bmw)
(live, stream)
(100, km)
(km, co)
(feat, monkey)
(united, joy)
(knockout, stages)
(league, legends)
(theix, feat)
(watch, nextgen)
(unitedinrivalry, berlinbrawl)
(emissions, combined)
(stream, starring)
(g, km)
(party, exclusive)
(exclusive, music)
(monkey, safari)
(safari, live)
(music, live)
0
1
2
3
4
5
6
7
Number of Occurences

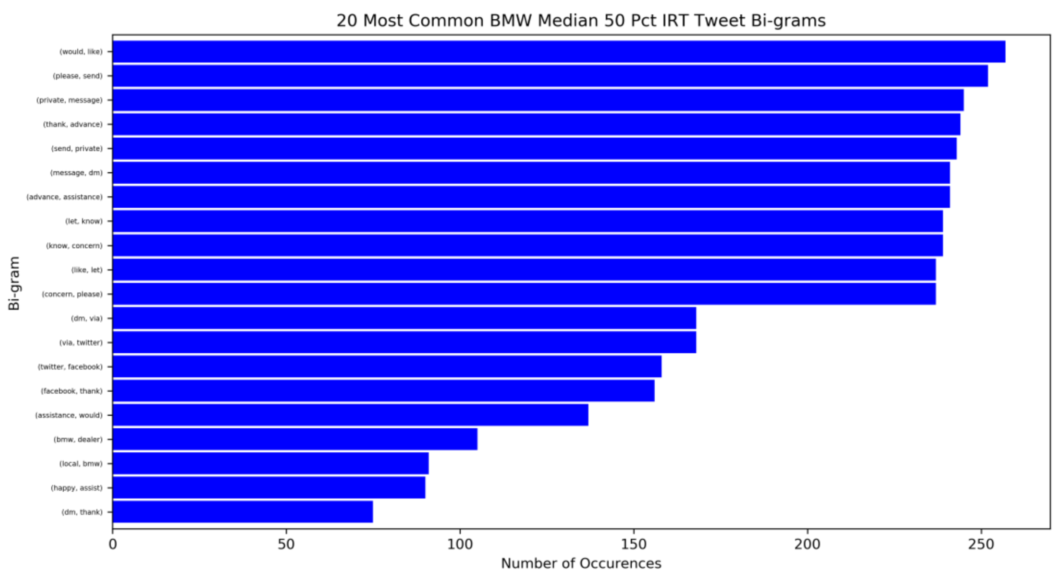

20 Most Common BMW Bottom 25 Pct Official Tweet Tri-grams
Tri-gram
(first, ever, bmw)
(stream, starring, theix)
(starring, theix, feat)
(virtual, party, exclusive)
(exclusive, music, live)
(feat, monkey, safari)
(live, stream, starring)
(co, emissions, combined)
(unitedinrivalry, united, joy)
(theix, feat, monkey)
(party, exclusive, music)
(knockout, stages, good)
(kwh, 100, km)
(km, co, emissions)
(stages, good, luck)
(100, km, co)
(monkey, safari, live)
(music, live, stream)
(ever, bmw, ix)
(get, ready, brand)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Number of Occurences

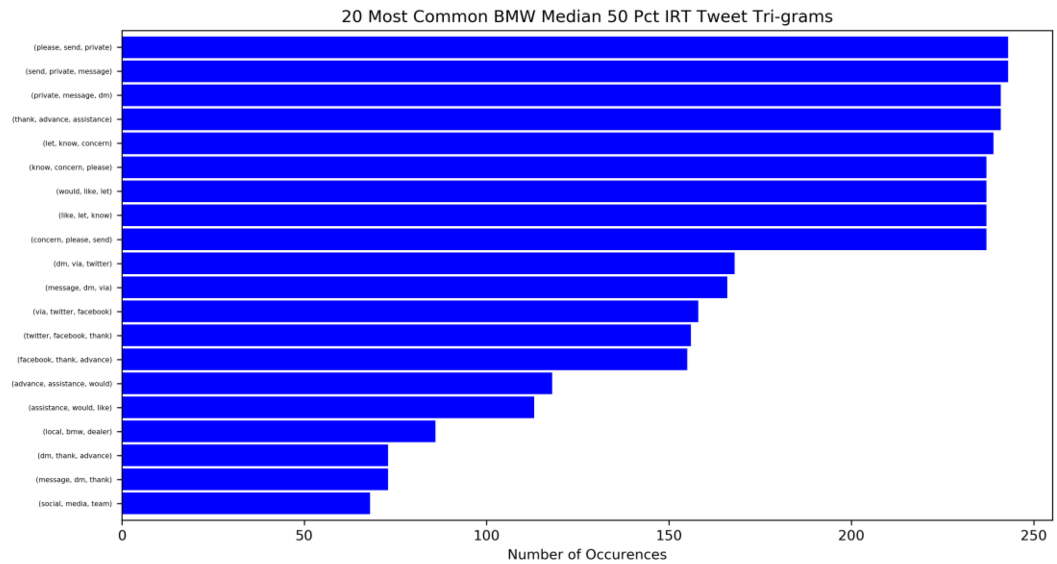

20 Most Common BMW IRT Bi-grams
Bi-gram
(would, like)
(please, send)
(thank, advance)
(private, message)
(send, private)
(advance, assistance)
(message, dm)
(let, know)
(like, let)
(know, concern)
(concern, please)
(contact, form)
(media, team)
(social, media)
(using, contact)
(recommend, using)
(following, link)
(bmw, social)
(clicking, following)
(form, clicking)
0
50
100
150
200
250
300
Number of Occurences

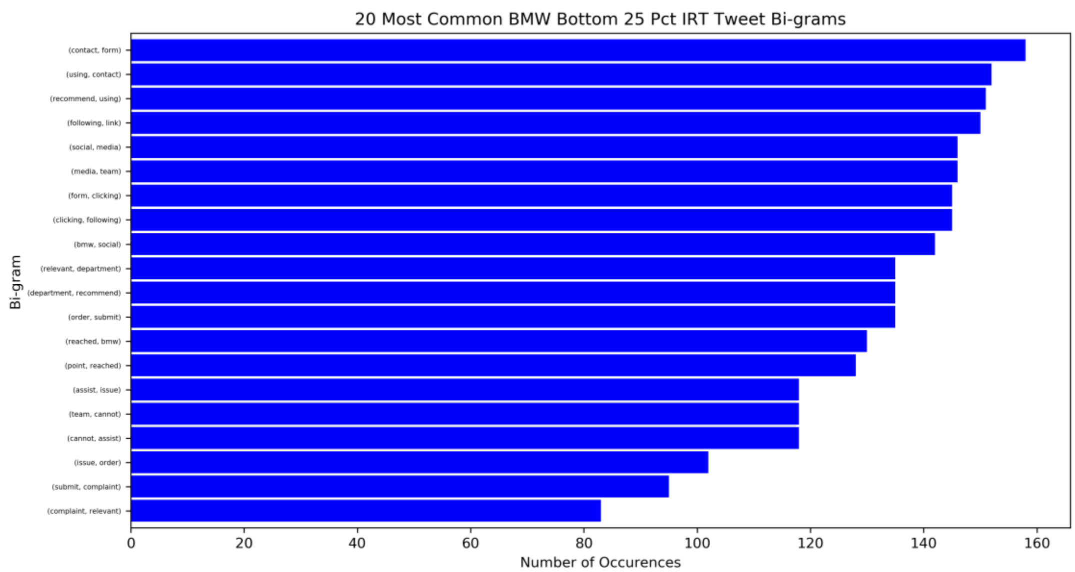

20 Most Common BMW IRT Tri-grams
(please, send, private)
(send, private, message)
(thank, advance, assistance)
(private, message, dm)
(like, let, know)
(would, like, let)
(let, know, concern)
(concern, please, send)
(know, concern, please)
(social, media, team)
(using, contact, form)
(recommend, using, contact)
(bmw, social, media)
(clicking, following, link)
(form, clicking, following)
(contact, form, clicking)
(department, recommend, using)
(relevant, department, recommend)
(reached, bmw, social)
(point, reached, bmw)
0
50
100
150
200
250
300
Number of Occurences

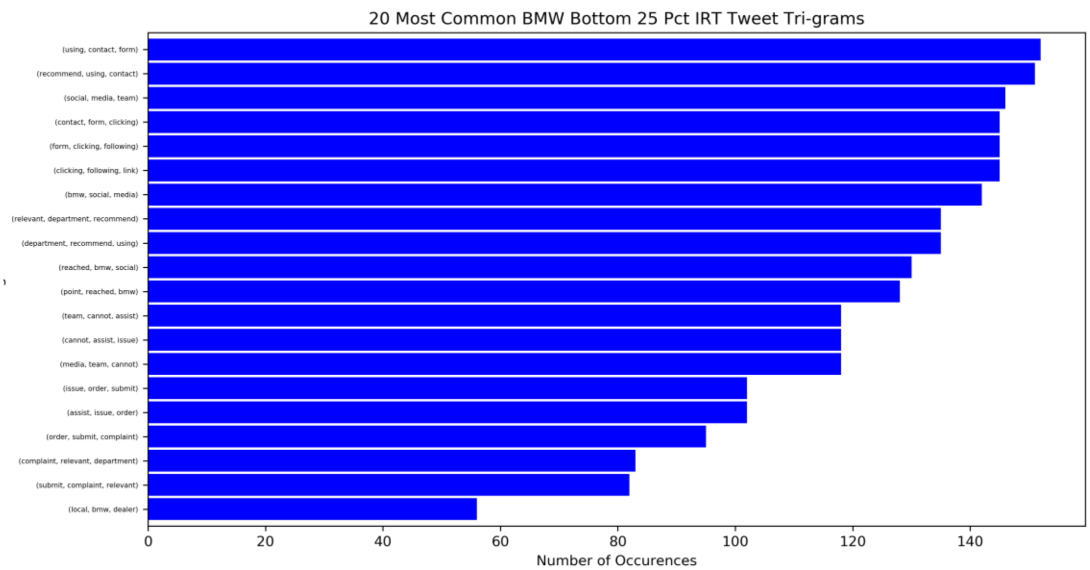

20 Most Common BMW Top 25 Pct IRT Bi-grams
Bi-gram
(thank, compliment)
(driving, pleasure)
(sheer, driving)
(would, like)
(please, send)
(thank, advance)
(pleasure, times)
(private, message)
(advance, assistance)
(send, private)
(message, dm)
(local, bmw)
(glad, like)
(wish, sheer)
(bmw, dealer)
(know, concern)
(like, let)
(concern, please)
(let, know)
(stay, tuned)
0
5
10
15
20
25
30
Number of Occurences


20 Most Common BMW Top 25 Pct IRT Tri-grams
(sheer, driving, pleasure)
(driving, pleasure, times)
(please, send, private)
(thank, advance, assistance)
(send, private, message)
(private, message, dm)
(wish, sheer, driving)
(would, like, let)
(like, let, know)
(concern, please, send)
(let, know, concern)
(know, concern, please)
(local, bmw, dealer)
(message, dm, thank)
(dm, thank, advance)
(contact, local, bmw)
(thank, kind, words)
(100, km, combined)
(social, media, team)
(stay, tuned, information)
Tri-gram
0
5
10
15
20
25
30
Number of Occurences


20 Most Common BMW Median 50 Pct IRT Tweet Bi-grams
(would, like)
(please, send)
(private, message)
(thank, advance)
(send, private)
(message, dm)
(advance, assistance)
(let, know)
(know, concern)
(like, let)
(concern, please)
(dm, via)
(via, twitter)
(twitter, facebook)
(facebook, thank)
(assistance, would)
(bmw, dealer)
(local, bmw)
(happy, assist)
(dm, thank)
Bi-gram
0
50
100
150
200
250
Number of Occurences


20 Most Common BMW Median 50 Pct IRT Tweet Tri-grams
(please, send, private)
(send, private, message)
(private, message, dm)
(thank, advance, assistance)
(let, know, concern)
(know, concern, please)
(would, like, let)
(like, let, know)
(concern, please, send)
(dm, via, twitter)
(message, dm, via)
(via, twitter, facebook)
(twitter, facebook, thank)
(facebook, thank, advance)
(advance, assistance, would)
(assistance, would, like)
(local, bmw, dealer)
(dm, thank, advance)
(message, dm, thank)
(social, media, team)
0
50
100
150
200
250
Number of Occurences


20 Most Common BMW Bottom 25 Pct IRT Tweet Bi-grams
Bi-gram
(contact, form)
(using, contact)
(recommend, using)
(following, link)
(social, media)
(media, team)
(form, clicking)
(clicking, following)
(bmw, social)
(relevant, department)
(department, recommend)
(order, submit)
(reached, bmw)
(point, reached)
(assist, issue)
(team, cannot)
(cannot, assist)
(issue, order)
(submit, complaint)
(complaint, relevant)
0
20
40
60
80
100
120
140
160
Number of Occurences



BMW Top Words

All Tweets	Official Tweets	Top 25% Official	Median 50% Official	Bottom 25% Official	All IRT Tweets	Top 25% IRT	Median 50% IRT	Bottom 25% IRT
bmw, 1558	bmw, 538	bmw, 178	bmw, 342	nextgen, 18	bmw, 1020	bmw, 176	bmw, 505	bmw, 338
thank, 830	km, 197	series, 61	km, 164	bmw, 17	thank, 828	thank, 157	thank, 505	assist, 234
like, 548	combined, 175	new, 42	combined, 144	live, 10	like, 530	like, 92	would, 360	team, 216
would, 509	series, 125	first, 41	‘100’, 98	watch, 9	would, 501	please, 58	like, 357	contact, 208
please, 500	‘100’, 115	sedan, 34	consumption, 74	first, 8	please, 498	glad, 53	please, 351	thank, 167
assist, 424	new, 114	coup, 33	emissions, 72	podcast, 7	assist, 424	would, 43	message, 294	recommend, 167
know, 407	first, 107	ever, 32	‘g’, 72	ever, 7	message, 400	compliment, 41	know, 283	form, 158
message, 400	ever, 93	competition, 28	new, 70	berlinbrawl, 7	contact, 396	know, 40	assistance, 267	using, 155
contact, 396	consumption, 90	combined, 27	co, 65	km, 6	know, 393	happy, 38	send, 257	following, 151
team, 362	g, 87	km, 27	series, 63	unitedinrivalry, 6	team, 358	message, 34	dm, 252	link, 151
assistance, 347	emissions, 86	gran, 23	fuel, 59	joy, 5	assistance, 345	contact, 33	private, 245	social, 146
send, 324	coup, 76	‘the7’, 17	first, 59	like, 4	send, 320	driving, 31	let, 244	media, 146
let, 318	co, 74	‘the8’, 16	l..(L), 58	learn, 4	dm, 308	let, 30	advance, 244	clicking, 145
dm, 308	fuel, 73	‘8’, 16	ever, 54	combined, 4	let, 304	assist, 30	concern, 241	submit, 143
km, 307	l..(L), 72	design, 16	coup, 43	world, 4	advance, 302	pleasure, 30	via, 186	order, 138
advance, 302	competition, 66	iaa19, 15	competition, 38	go, 4	private, 301	send, 29	twitter, 168	department, 138
private, 301	sedan, 53	luxury, 15	energy, 36	good, 4	concern, 291	sheer, 29	facebook, 160	relevant, 136
concern, 291	gran, 50	consumption, 14	kwh, 35	listen, 4	recommend, 277	times, 29	assist, 159	reached, 130

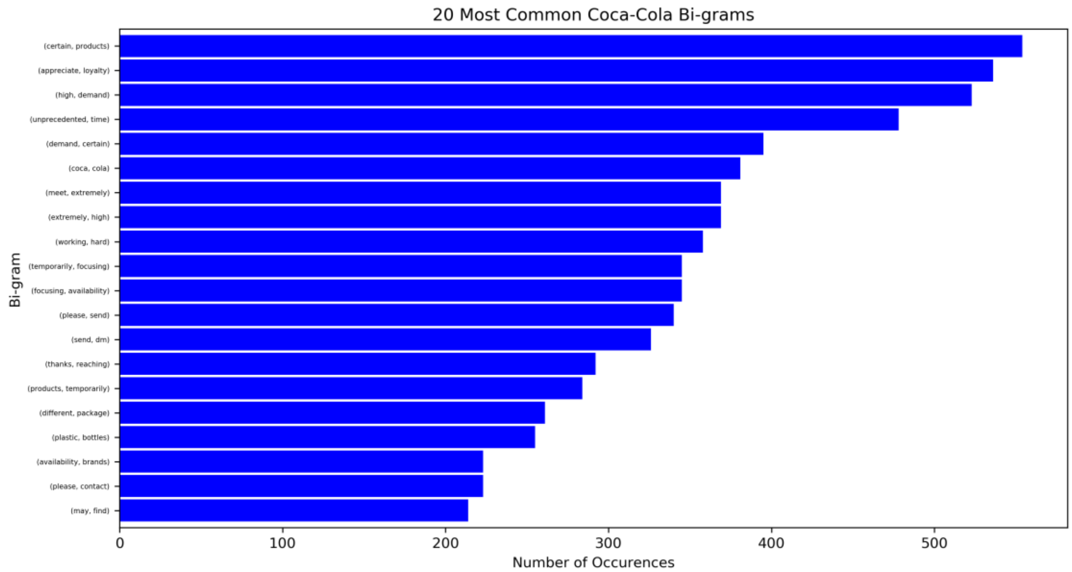
recommend, 277	driving, 40	‘100’, 14	driving, 28	theix, 4	form, 228	assistance, 27	contact, 154	point, 128
combined, 271	energy, 39	car, 13	gran, 27	way, 3	social, 227	local, 27	team, 119	cannot, 121

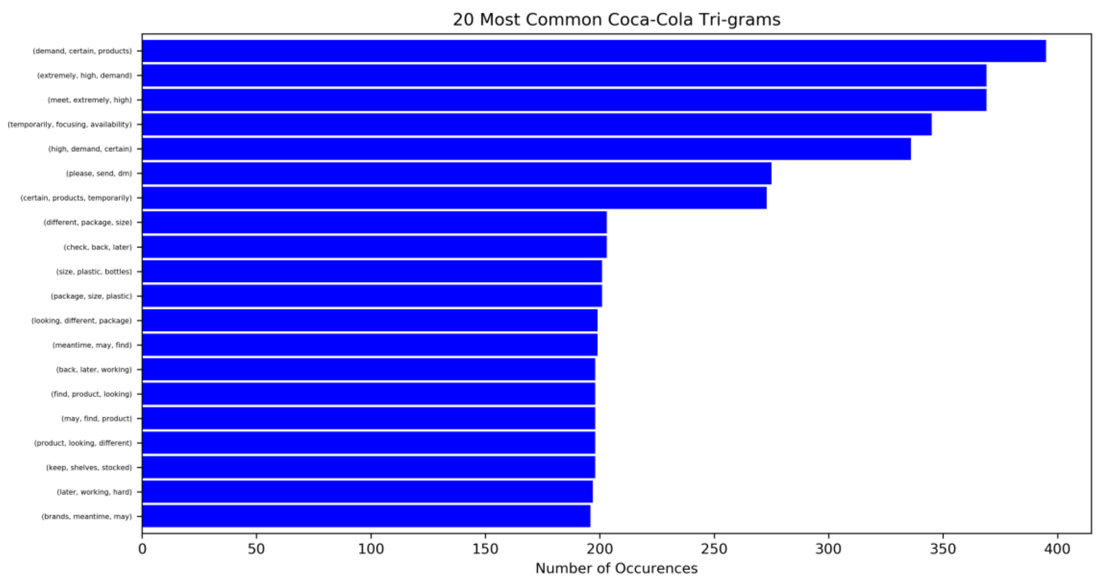
-Following are the words which are uniquely common to each of the categories above:

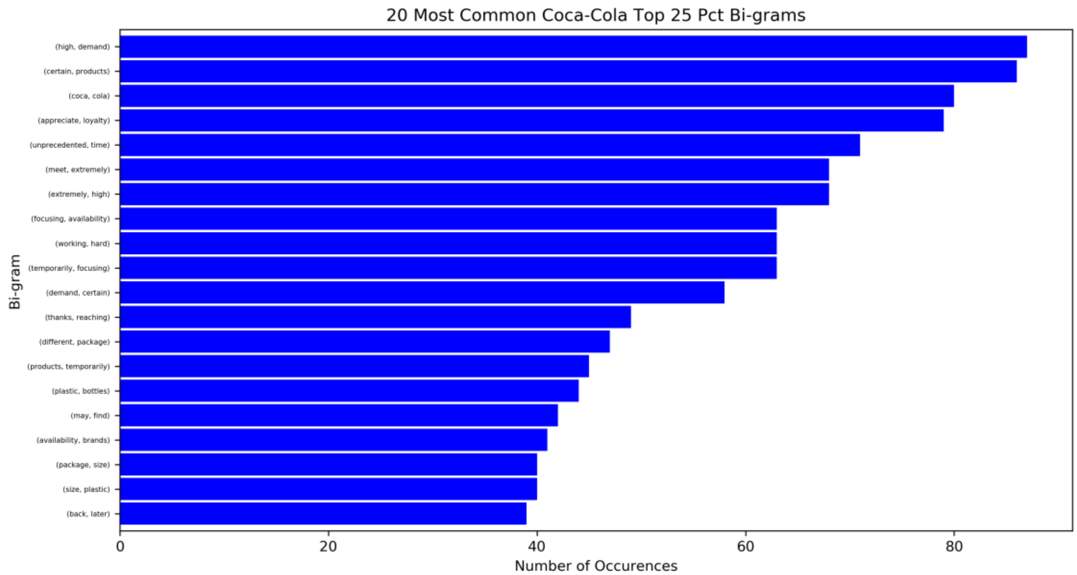
All Tweets	Official Tweets	Top 25% Official	Median 50% Official	Bottom 25% Official	All IRT Tweets	Top 25% IRT	Median 50% IRT	Bottom 25% IRT
		the7	kwh	nextgen		glad	via	using
		the8		live		compliment	twitter	following
		8		watch		happy	facebook	link
		design		podcast		pleasure		media
		iaa19		berlinbrawl		sheer		clicking
		luxury		unitedinrivalry		times		submit
		car		joy		local		order
				learn				department
				world				relevant
				go				reached
				good				point
				listen				cannot
				theix				
				way				

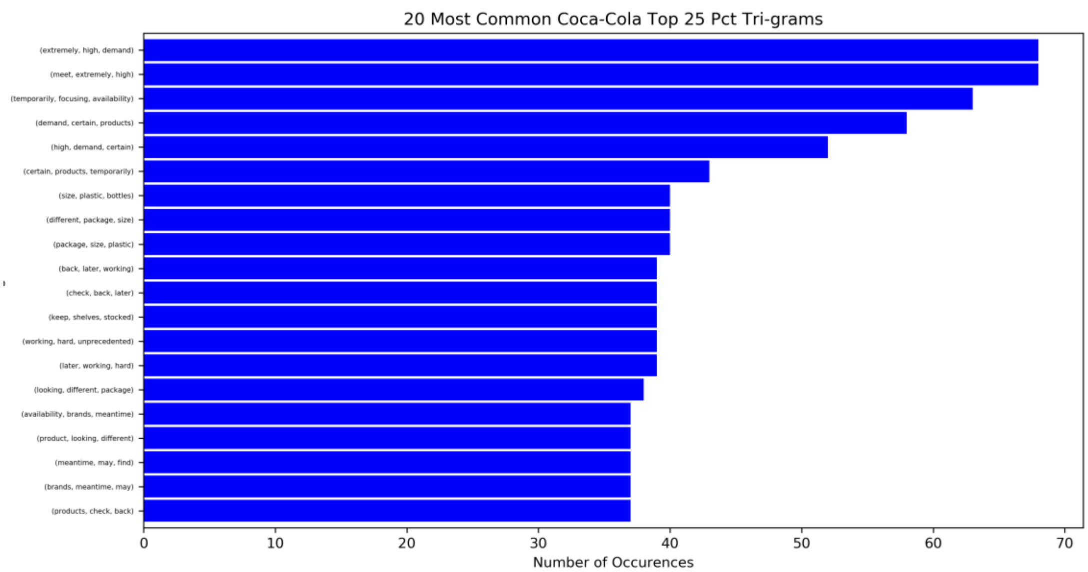
Coca Cola

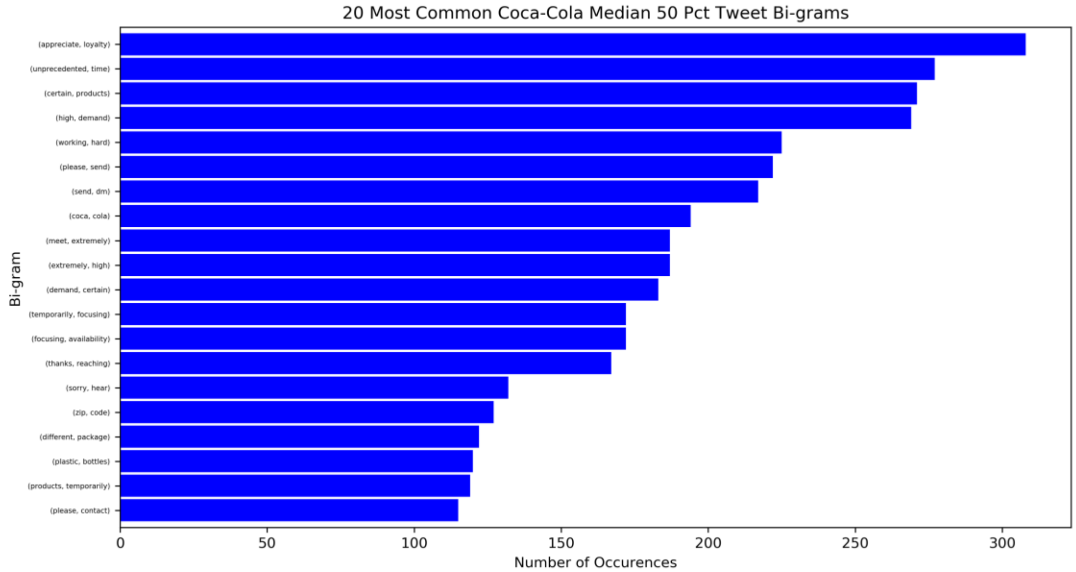
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	3206	26.62	85331	3549
Top 25% IRT:	795	21.26792	16908	1745
Median 50% IRT:	1588	28.0699	44575	2385
Bottom 25% IRT:	795	29.63774	23562	1518

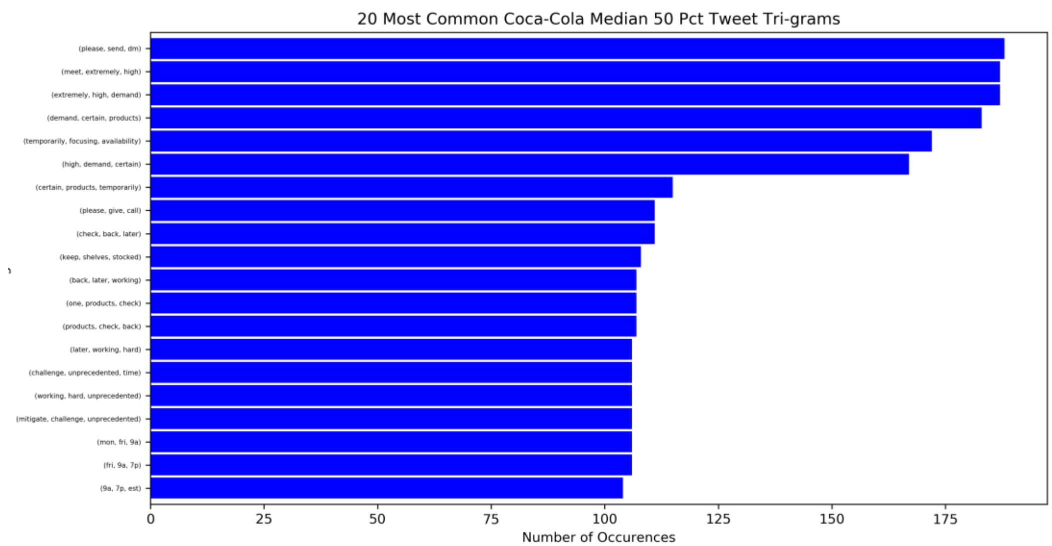

20 Most Common Coca-Cola Bi-grams
Bi-gram
(certain, products)
(appreciate, loyalty)
(high, demand)
(unprecedented, time)
(demand, certain)
(coca, cola)
(meet, extremely)
(extremely, high)
(working, hard)
(temporarily, focusing)
(focusing, availability)
(please, send)
(send, dm)
(thanks, reaching)
(products, temporarily)
(different, package)
(plastic, bottles)
(availability, brands)
(please, contact)
(may, find)
0
100
200
300
400
500
Number of Occurences

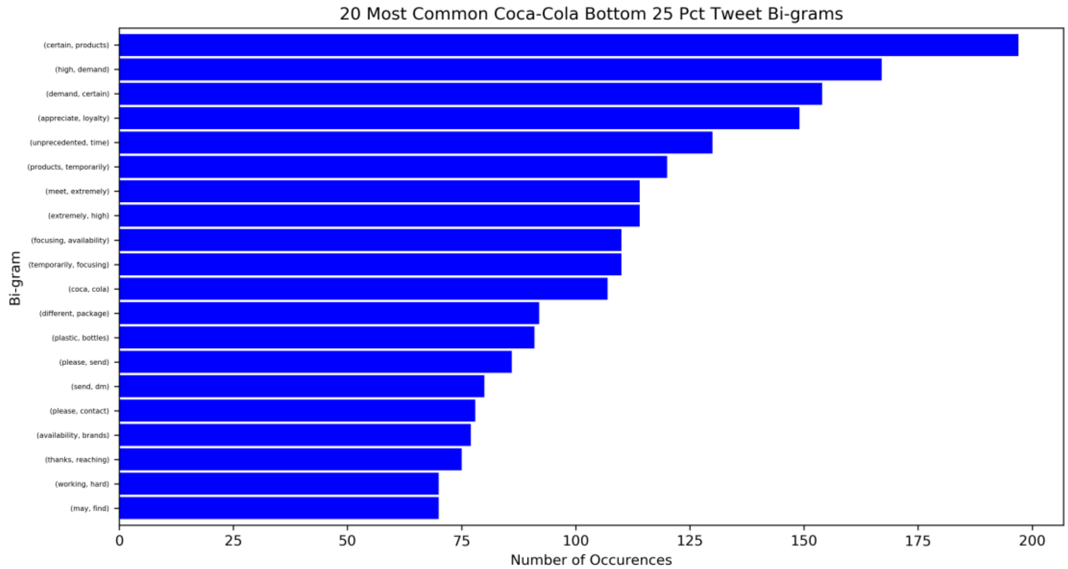

20 Most Common Coca-Cola Tri-grams
(demand, certain, products)
(extremely, high, demand)
(meet, extremely, high)
(temporarily, focusing, availability)
(high, demand, certain)
(please, send, dm)
(certain, products, temporarily)
(different, package, size)
(check, back, later)
(size, plastic, bottles)
(package, size, plastic)
(looking, different, package)
(meantime, may, find)
(back, later, working)
(find, product, looking)
(may, find, product)
(product, looking, different)
(keep, shelves, stocked)
(later, working, hard)
(brands, meantime, may)
0
50
100
150
200
250
300
350
400
Number of Occurences

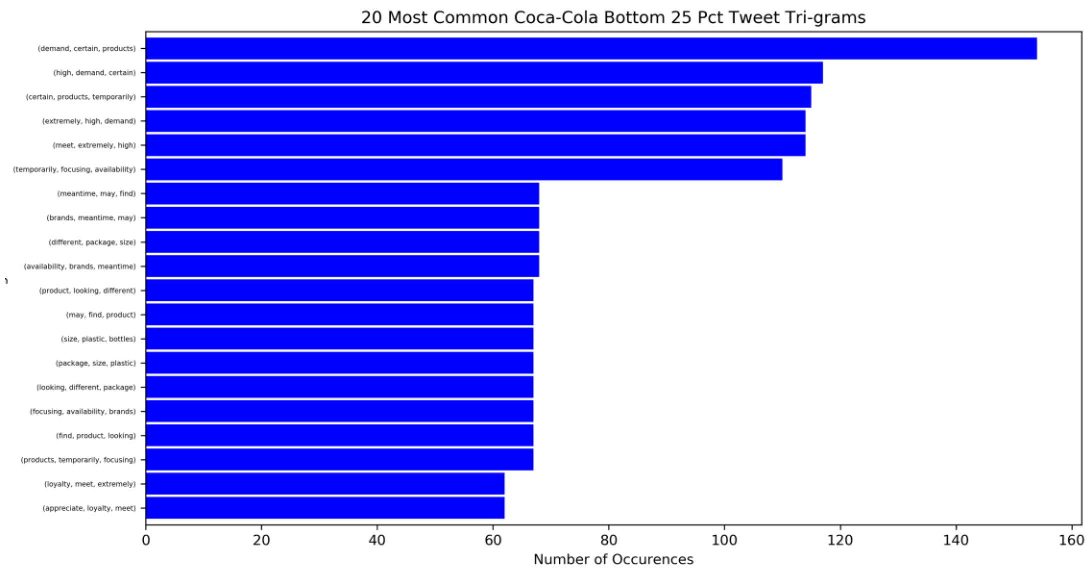

20 Most Common Coca-Cola Top 25 Pct Bi-grams
Bi-gram
(high, demand)
(certain, products)
(coca, cola)
(appreciate, loyalty)
(unprecedented, time)
(meet, extremely)
(extremely, high)
(focusing, availability)
(working, hard)
(temporarily, focusing)
(demand, certain)
(thanks, reaching)
(different, package)
(products, temporarily)
(plastic, bottles)
(may, find)
(availability, brands)
(package, size)
(size, plastic)
(back, later)
0
20
40
60
80
Number of Occurences


20 Most Common Coca-Cola Top 25 Pct Tri-grams
(extremely, high, demand)
(meet, extremely, high)
(temporarily, focusing, availability)
(demand, certain, products)
(high, demand, certain)
(certain, products, temporarily)
(size, plastic, bottles)
(different, package, size)
(package, size, plastic)
(back, later, working)
(check, back, later)
(keep, shelves, stocked)
(working, hard, unprecedented)
(later, working, hard)
(looking, different, package)
(availability, brands, meantime)
(product, looking, different)
(meantime, may, find)
(brands, meantime, may)
(products, check, back)
0
10
20
30
40
50
60
70
Number of Occurences


20 Most Common Coca-Cola Median 50 Pct Tweet Bi-grams
Bi-gram
(appreciate, loyalty)
(unprecedented, time)
(certain, products)
(high, demand)
(working, hard)
(please, send)
(send, dm)
(coca, cola)
(meet, extremely)
(extremely, high)
(demand, certain)
(temporarily, focusing)
(focusing, availability)
(thanks, reaching)
(sorry, hear)
(zip, code)
(different, package)
(plastic, bottles)
(products, temporarily)
(please, contact)
0
50
100
150
200
250
300
Number of Occurences


20 Most Common Coca-Cola Median 50 Pct Tweet Tri-grams
(please, send, dm)
(meet, extremely, high)
(extremely, high, demand)
(demand, certain, products)
(temporarily, focusing, availability)
(high, demand, certain)
(certain, products, temporarily)
(please, give, call)
(check, back, later)
(keep, shelves, stocked)
(back, later, working)
(one, products, check)
(products, check, back)
(later, working, hard)
(challenge, unprecedented, time)
(working, hard, unprecedented)
(mitigate, challenge, unprecedented)
(mon, fri, 9a)
(fri, 9a, 7p)
(9a, 7p, est)
0
25
50
75
100
125
150
175
Number of Occurences





Coca Cola Top Words

All Tweets	Top 25%	Median 50%	Bottom 25%
products, 1245	products, 193	products, 707	products, 345
please, 1030	hi, 135	please, 637	demand, 305
hi, 970	demand, 133	hi, 532	hi, 303
demand, 938	thanks, 132	demand, 500	please, 272
time, 657	please, 120	time, 369	temporarily, 216
thanks, 653	time, 108	appreciate, 351	certain, 201
appreciate, 631	appreciate, 99	working, 337	thanks, 197
temporarily, 594	temporarily, 99	thanks, 323	appreciate, 181
certain, 568	high, 91	loyalty, 318	time, 180
loyalty, 551	certain, 89	unprecedented, 297	high, 172
high, 536	working, 86	temporarily, 279	loyalty, 151
working, 511	loyalty, 82	certain, 278	beverages, 145

unprecedented, 510	coca, 80	high, 273	unprecedented, 135
sorry, 437	cola, 80	send, 270	availability, 130
send, 425	unprecedented, 78	hard, 263	contact, 127
availability, 417	availability, 77	dm, 257	assistance, 126
hard, 413	sorry, 77	sorry, 256	plastic, 125
dm, 395	like, 76	reaching, 219	meet, 119
reaching, 388	may, 75	availability, 210	meantime, 117
cola, 382	coke, 74	keep, 207	extremely, 114

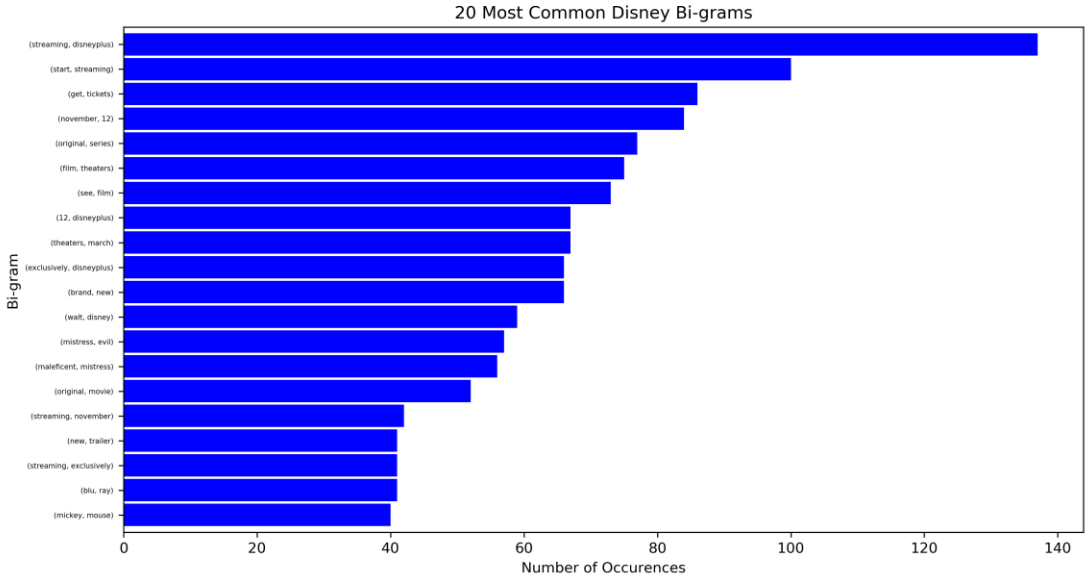
-Following are the words uniquely common to each list above:

All Tweets	Top 25%	Median 50%	Bottom 25%
	coca	keep	beverages
	like		contact
	may		assistance
	coke		plastic
			meet
			meantime
			extremely

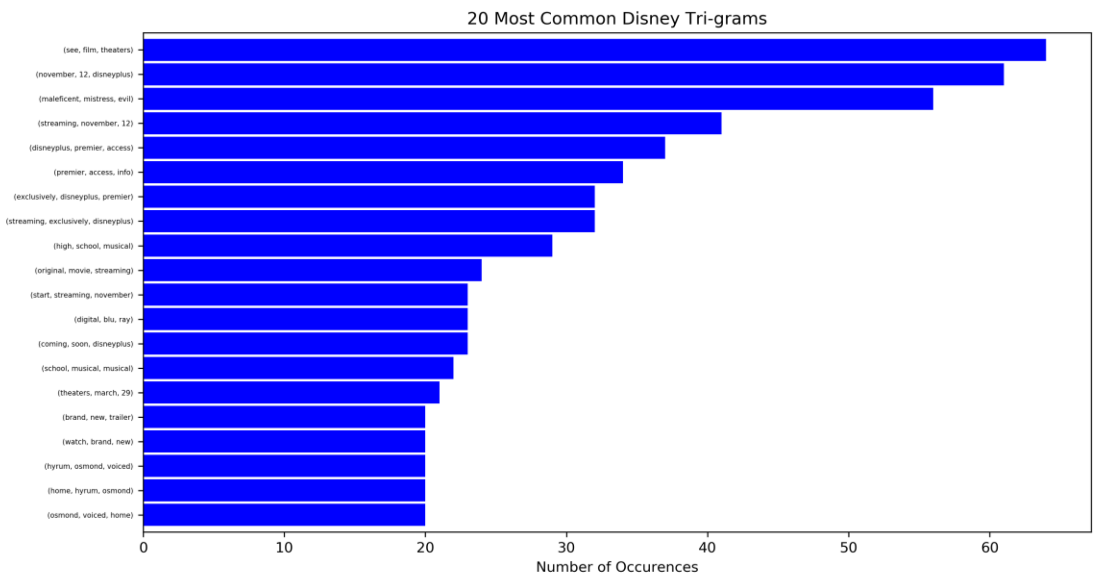
Disney:

It seems that Disney does not have enough IRT data for split analysis. Analysis will be performed using all Disney tweets, including the two 'IRT' tweets.

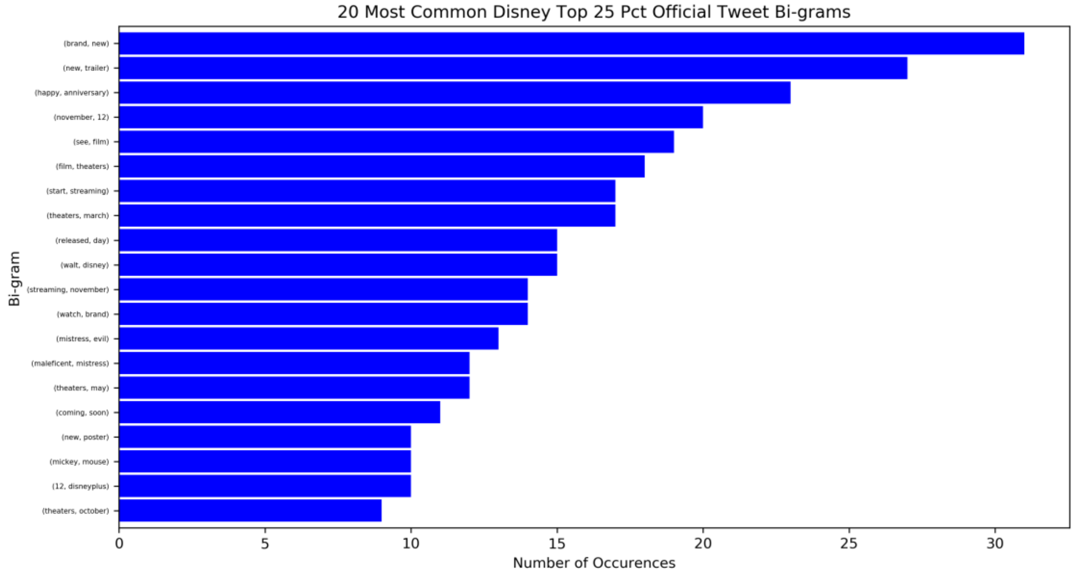
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	2722	16.71	45472	4911
Top 25 OT%:	681	13.61	9269	2099
Median 50 OT%:	1360	17.17	23352	3476
Bottom 25 OT%:	681	18.88	12854	2321

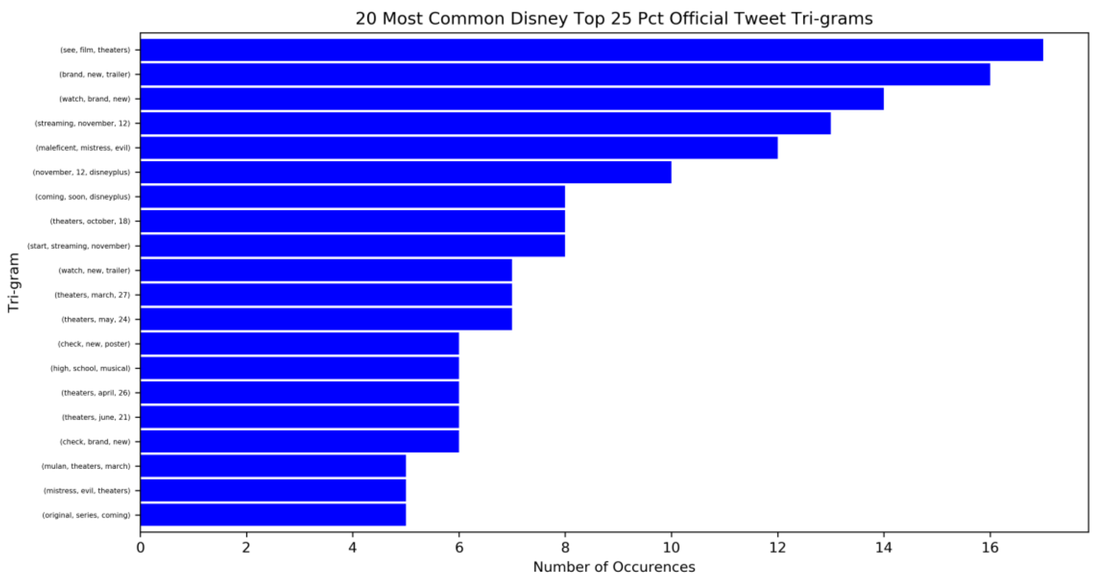


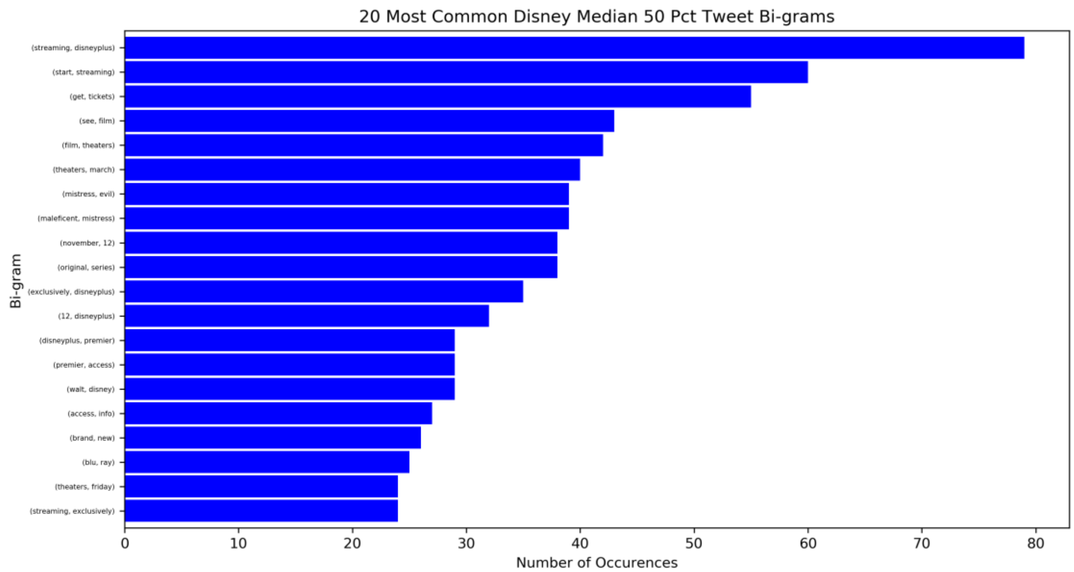
20 Most Common Disney Bi-grams
(streaming, disneyplus)
(start, streaming)
(get, tickets)
(november, 12)
(original, series)
(film, theaters)
(see, film)
(12, disneyplus)
(theaters, march)
(exclusively, disneyplus)
(brand, new)
(walt, disney)
(mistress, evil)
(maleficent, mistress)
(original, movie)
(streaming, november)
(new, trailer)
(streaming, exclusively)
(blu, ray)
(mickey, mouse)
Bi-gram
0
20
40
60
80
100
120
140
Number of Occurences

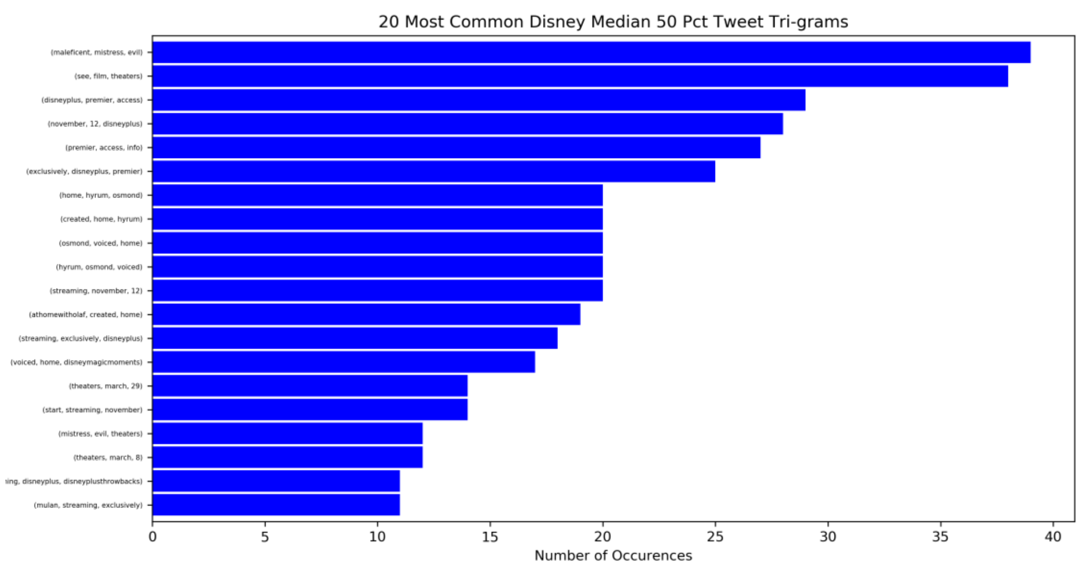


20 Most Common Disney Tri-grams
(see, film, theaters)
(november, 12, disneyplus)
(maleficent, mistress, evil)
(streaming, november, 12)
(disneyplus, premier, access)
(premier, access, info)
(exclusively, disneyplus, premier)
(streaming, exclusively, disneyplus)
(high, school, musical)
(original, movie, streaming)
(start, streaming, november)
(digital, blu, ray)
(coming, soon, disneyplus)
(school, musical, musical)
(theaters, march, 29)
(brand, new, trailer)
(watch, brand, new)
(hyrum, osmond, voiced)
(home, hyrum, osmond)
(osmond, voiced, home)
0
10
20
30
40
50
60
Number of Occurences

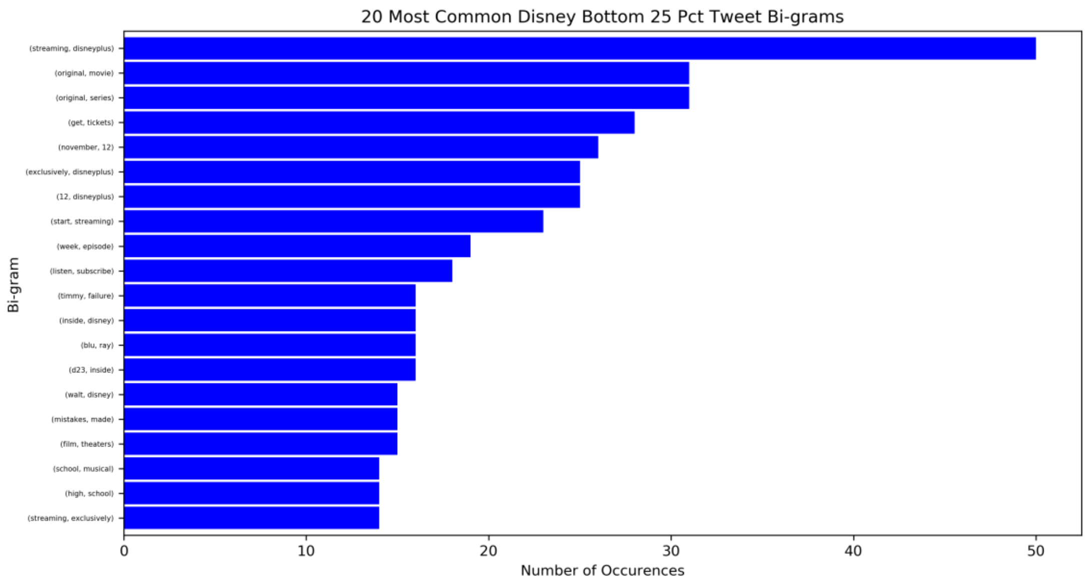

20 Most Common Disney Top 25 Pct Official Tweet Bi-grams
(brand, new)
(new, trailer)
(happy, anniversary)
(november, 12)
(see, film)
(film, theaters)
(start, streaming)
(theaters, march)
(released, day)
(walt, disney)
(streaming, november)
(watch, brand)
(mistress, evil)
(maleficent, mistress)
(theaters, may)
(coming, soon)
(new, poster)
(mickey, mouse)
(12, disneyplus)
(theaters, october)
Bi-gram
0
5
10
15
20
25
30
Number of Occurences

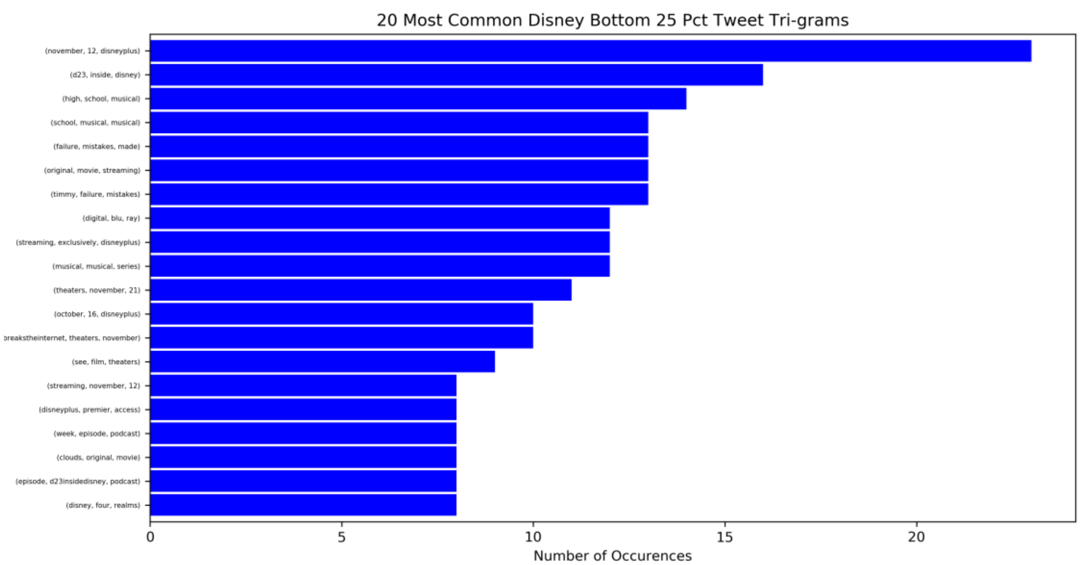

20 Most Common Disney Top 25 Pct Official Tweet Tri-grams
(see, film, theaters)
(brand, new, trailer)
(watch, brand, new)
(streaming, november, 12)
(maleficent, mistress, evil)
(november, 12, disneyplus)
(coming, soon, disneyplus)
(theaters, october, 18)
(start, streaming, november)
(watch, new, trailer)
(theaters, march, 27)
(theaters, may, 24)
(check, new, poster)
(high, school, musical)
(theaters, april, 26)
(theaters, june, 21)
(check, brand, new)
(mulan, theaters, march)
(mistress, evil, theaters)
(original, series, coming)
Tri-gram
0
2
4
6
8
10
12
14
16
Number of Occurences





-Note: The first word in the second to last trigram did not fit on the screen. I believe this word is "streaming".





Note: There's something before 'breakstheinternet', but it did not fit on screen. Thus, the trigram is 'Xbreakstheinternet', where I cannot see 'X.'

Disney Top Words

All Tweets	Top 25%	Median 50%	Bottom 25%
disneyplus, 594	new, 107	disneyplus, 296	disneyplus, 206
theaters, 450	disneyplus, 92	theaters, 267	streaming, 142
streaming, 411	theaters, 91	streaming, 221	disney, 141
new, 405	disney, 71	see, 218	new, 113
disney, 395	day, 64	new, 185	theaters, 93
see, 365	see, 63	disney, 183	see, 85
get, 204	happy, 60	get, 107	get, 80
original, 182	anniversary, 57	film, 97	original, 72
film, 174	streaming, 48	original, 87	series, 72
series, 171	watch, 38	tickets, 87	episode, 52
watch, 158	trailer, 37	series, 74	movie, 50

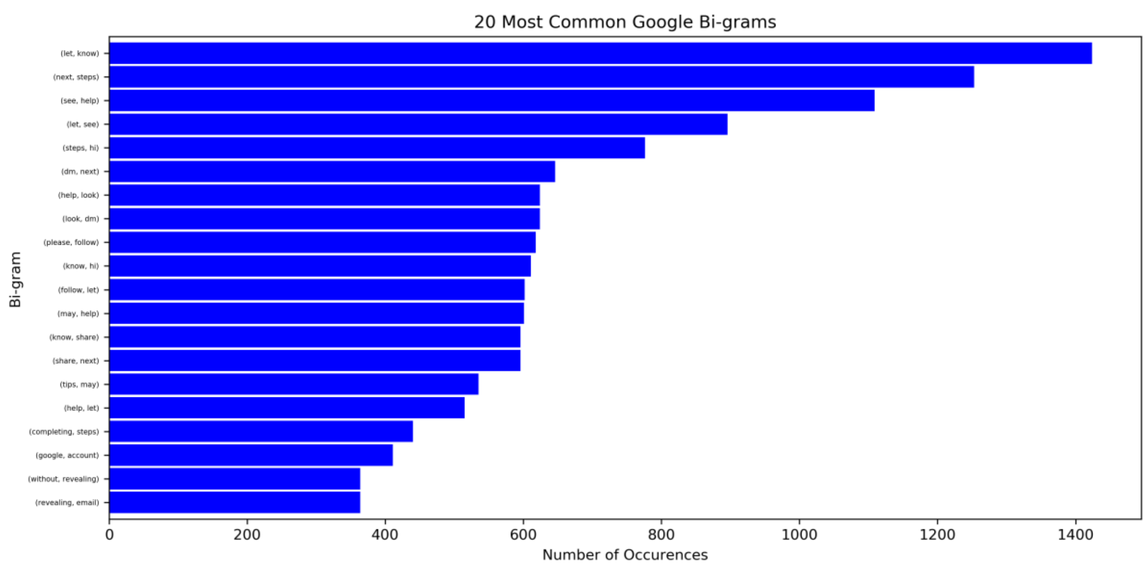
day, 150	coming, 37	watch, 71	watch, 49
world, 142	world, 35	home, 71	film, 45
november, 139	november, 33	world, 71	november, 44
one, 132	film, 32	start, 69	one, 43
movie, 131	thelionking, 31	one, 67	podcast, 39
tickets, 128	brand, 31	magic, 64	tickets, 37
coming, 122	frozen2, 28	disneymagicmoments, 63	world, 36
magic, 116	series, 25	november, 62	pixaronward, 35
start, 116	toystory4, 24	movie, 62	story, 35

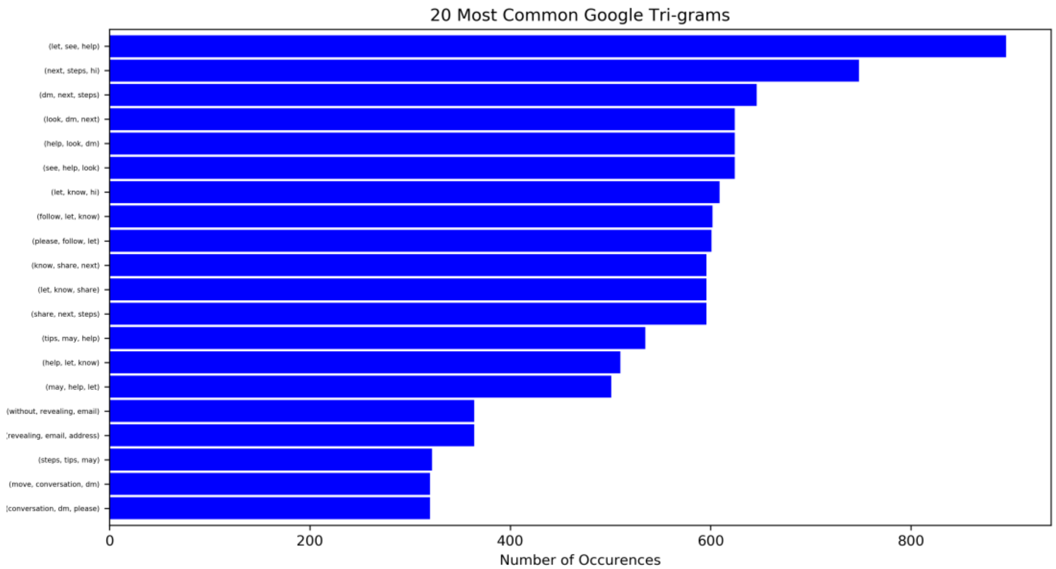
-Following are the words uniquely common to each list above:

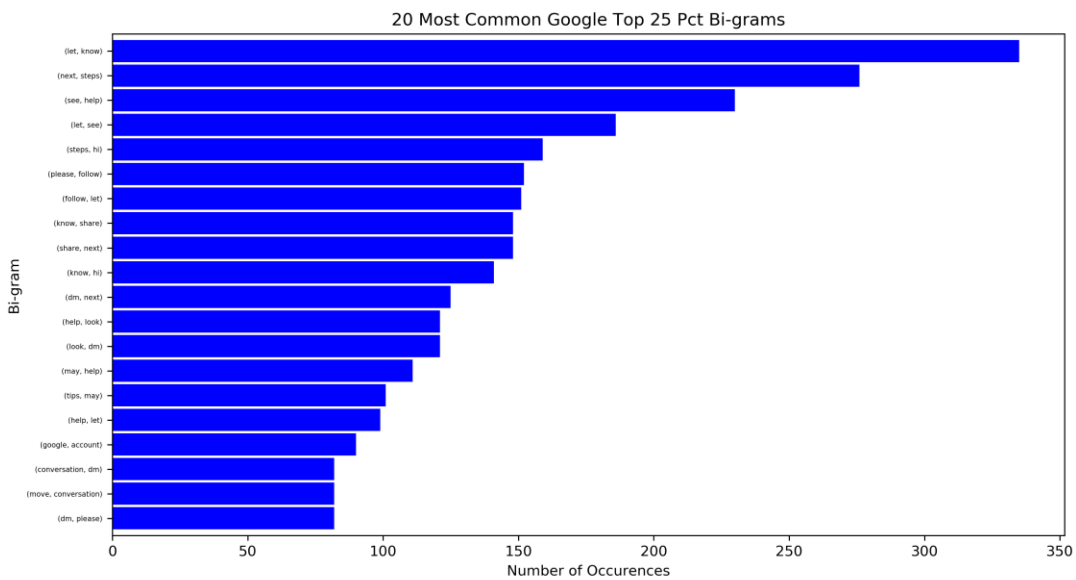
All Tweets	Top 25%	Median 50%	Bottom 25%
	happy	home	episode
	anniversary	disneymagicmoments	podcast
	trailer		pixaronward
	thelionking		story
	brand		
	frozen2		
	toystory4		

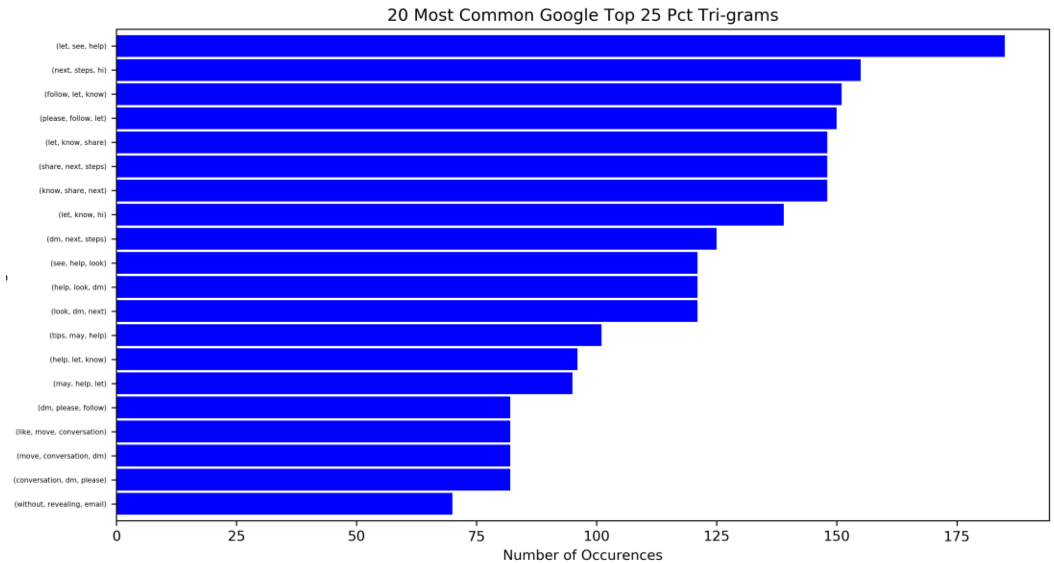
Google

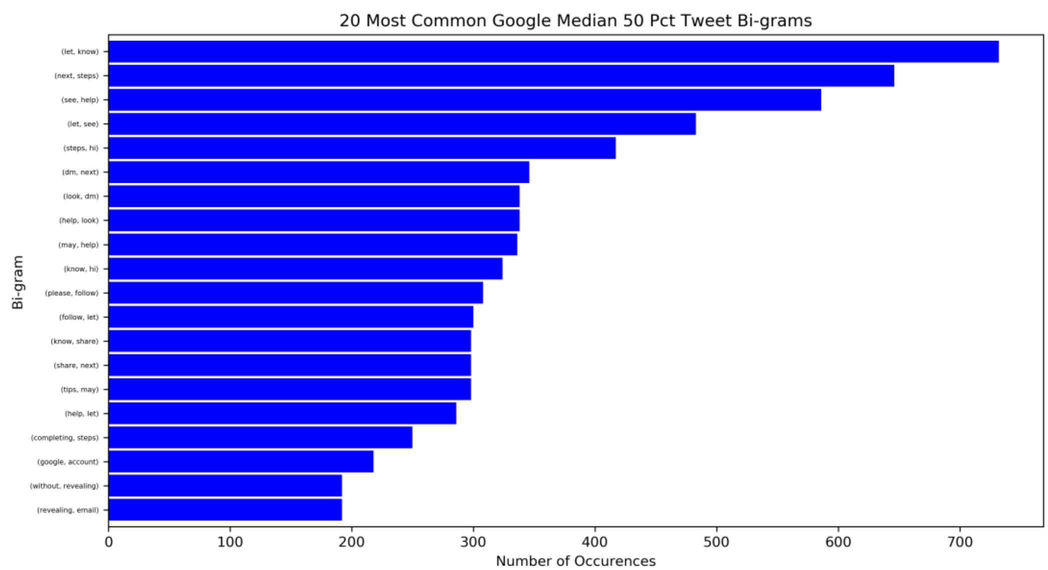
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	3160	23.97	75738	2723
Top 25 IRT%:	758	23.41689	17750	846
Median 50 IRT%:	1513	23.83807	36067	1170
Bottom 25 IRT%:	758	23.74406	17998	660

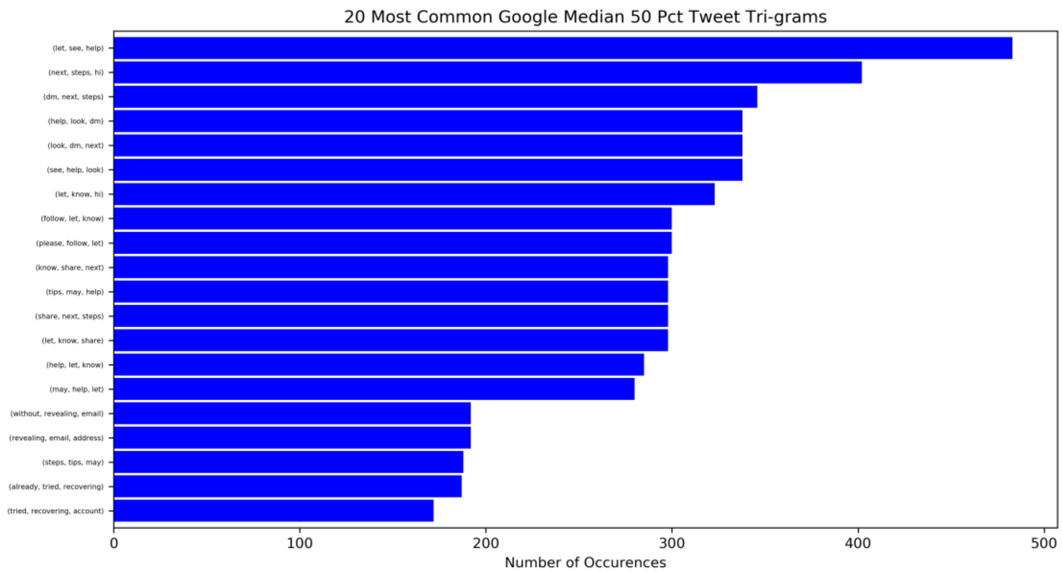



20 Most Common Google Tri-grams
(let, see, help)
(next, steps, hi)
(dm, next, steps)
(look, dm, next)
(help, look, dm)
(see, help, look)
(let, know, hi)
(follow, let, know)
(please, follow, let)
(know, share, next)
(let, know, share)
(share, next, steps)
(tips, may, help)
(help, let, know)
(may, help, let)
(without, revealing, email)
revealing, email, address)
(steps, tips, may)
(move, conversation, dm)
conversation, dm, please)
0
200
400
600
800
Number of Occurences

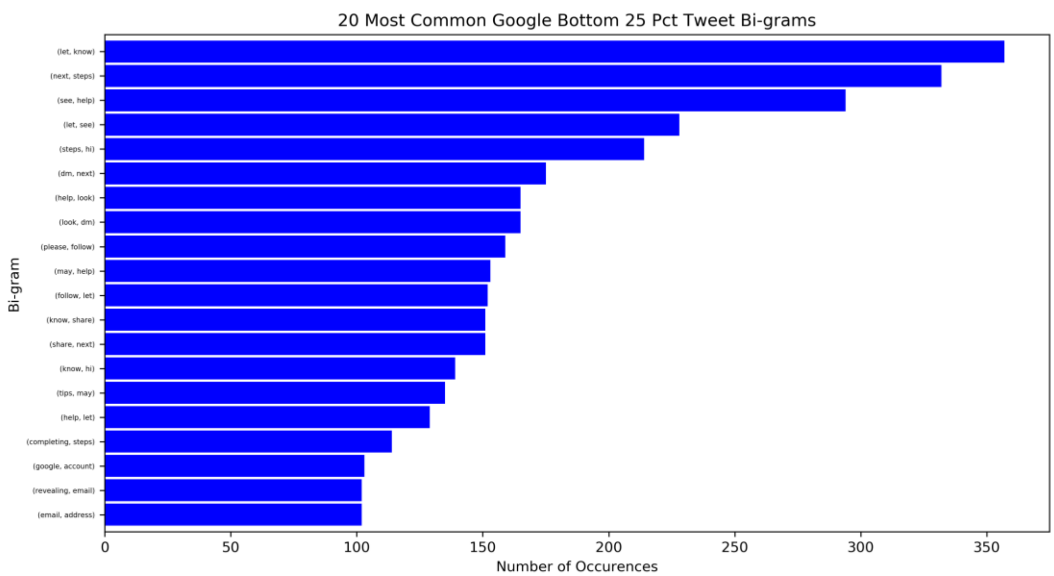

20 Most Common Google Top 25 Pct Bi-grams
Bi-gram
(let, know)
(next, steps)
(see, help)
(let, see)
(steps, hi)
(please, follow)
(follow, let)
(know, share)
(share, next)
(know, hi)
(dm, next)
(help, look)
(look, dm)
(may, help)
(tips, may)
(help, let)
(google, account)
(conversation, dm)
(move, conversation)
(dm, please)
0
50
100
150
200
250
300
350
Number of Occurences


20 Most Common Google Top 25 Pct Tri-grams
(let, see, help)
(next, steps, hi)
(follow, let, know)
(please, follow, let)
(let, know, share)
(share, next, steps)
(know, share, next)
(let, know, hi)
(dm, next, steps)
(see, help, look)
(help, look, dm)
(look, dm, next)
(tips, may, help)
(help, let, know)
(may, help, let)
(dm, please, follow)
(like, move, conversation)
(move, conversation, dm)
(conversation, dm, please)
(without, revealing, email)
0
25
50
75
100
125
150
175
Number of Occurences

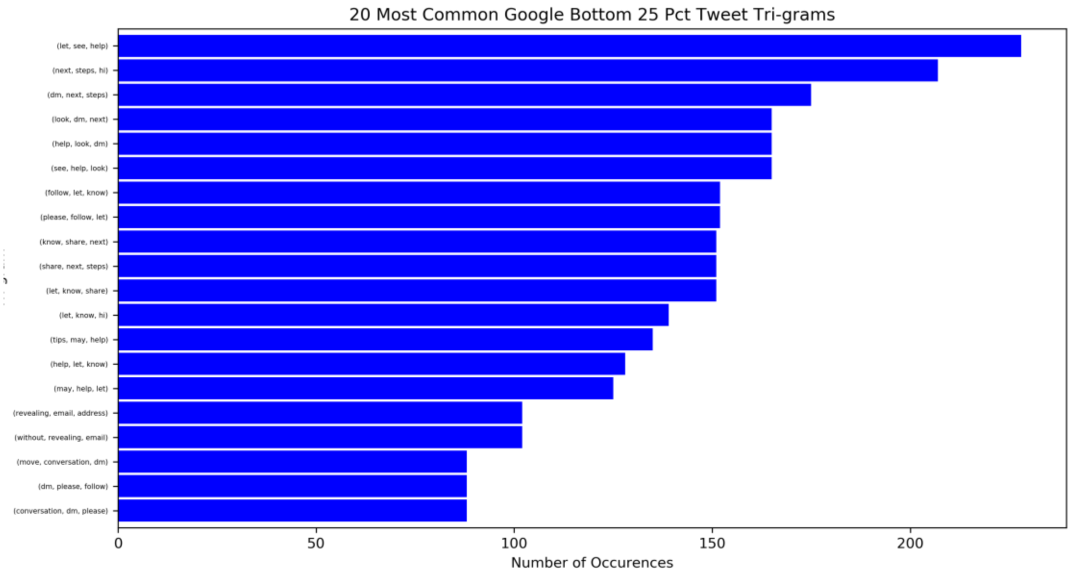

20 Most Common Google Median 50 Pct Tweet Bi-grams
Bi-gram
(let, know)
(next, steps)
(see, help)
(let, see)
(steps, hi)
(dm, next)
(look, dm)
(help, look)
(may, help)
(know, hi)
(please, follow)
(follow, let)
(know, share)
(share, next)
(tips, may)
(help, let)
(completing, steps)
(google, account)
(without, revealing)
(revealing, email)
0
100
200
300
400
500
600
700
Number of Occurences



20 Most Common Google Median 50 Pct Tweet Tri-grams
(let, see, help)
(next, steps, hi)
(dm, next, steps)
(help, look, dm)
(look, dm, next)
(see, help, look)
(let, know, hi)
(follow, let, know)
(please, follow, let)
(know, share, next)
(tips, may, help)
(share, next, steps)
(let, know, share)
(help, let, know)
(may, help, let)
(without, revealing, email)
(revealing, email, address)
(steps, tips, may)
(already, tried, recovering)
(tried, recovering, account)
0
100
200
300
400
500
Number of Occurences



20 Most Common Google Bottom 25 Pct Tweet Bi-grams
Bi-gram
(let, know)
(next, steps)
(see, help)
(let, see)
(steps, hi)
(dm, next)
(help, look)
(look, dm)
(please, follow)
(may, help)
(follow, let)
(know, share)
(share, next)
(know, hi)
(tips, may)
(help, let)
(completing, steps)
(google, account)
(revealing, email)
(email, address)
0
50
100
150
200
250
300
350
Number of Occurences



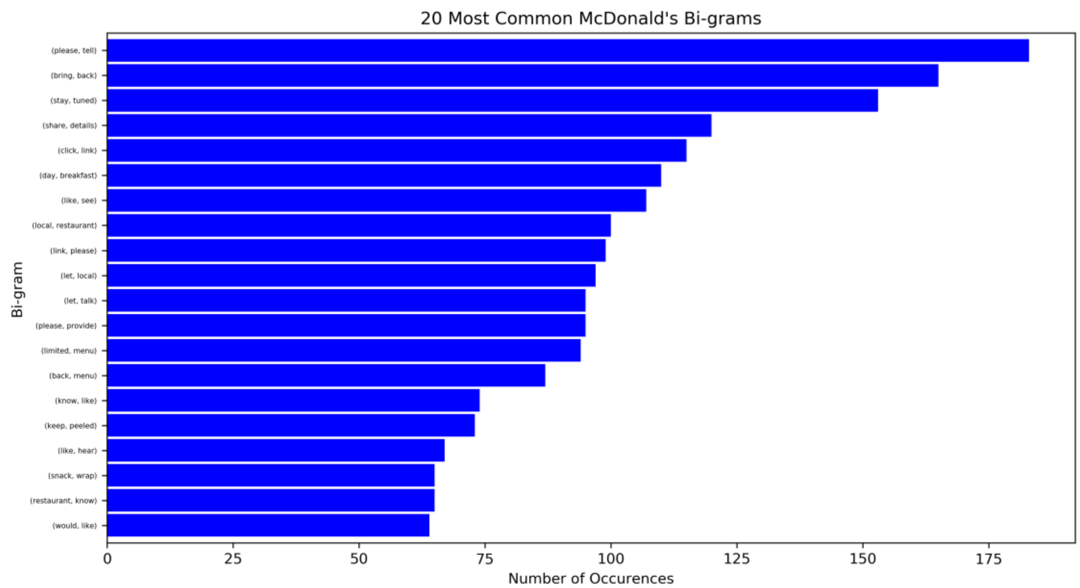
Google Top Words

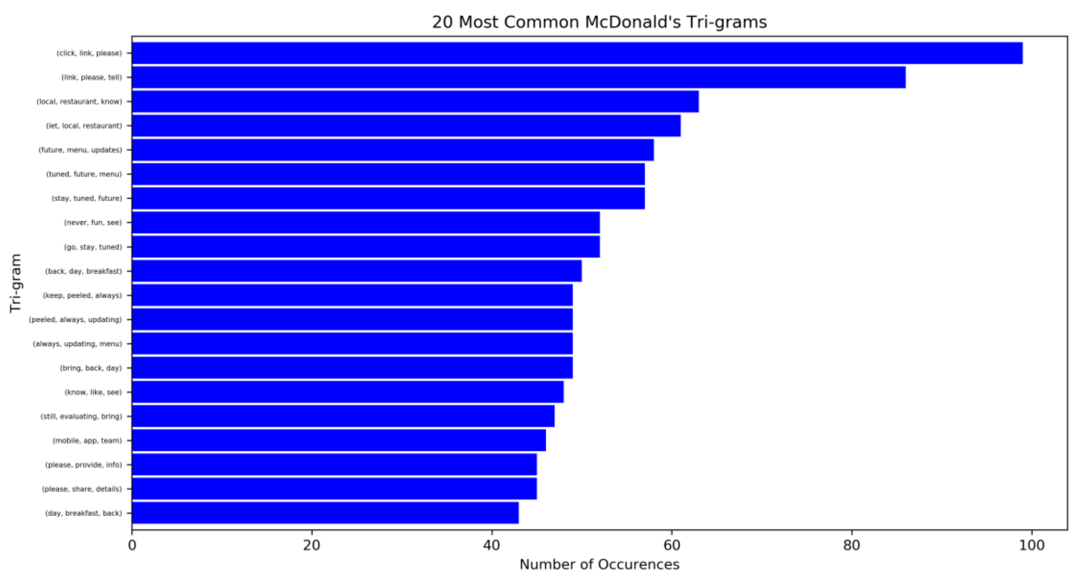
All Tweets	Top 25%	Median 50%	Bottom 25%
hi, 2417	let, 531	hi, 1283	hi, 622
let, 2338	hi, 512	let, 1221	let, 587
help, 2001	help, 410	help, 1071	help, 520
steps, 1862	steps, 390	steps, 80	steps, 492
know, 1535	know, 362	know, 789	know, 384
dm, 1271	dm, 279	dm, 657	dm, 336
next, 1255	next, 278	next, 646	next, 332
see, 1184	see, 264	see, 614	see, 307
account, 1078	account, 229	account, 557	account, 272
may, 852	google, 179	may, 479	may, 217
try, 722	please, 163	try, 391	try, 201
tips, 701	may, 155	tips, 381	tips, 178
please, 667	follow, 155	look, 344	please, 177
look, 639	share, 152	please, 328	look, 170
follow, 626	tips, 141	follow, 311	follow, 161
google, 609	try, 129	confirm, 304	share, 152
share, 604	look, 125	share, 301	google, 143
confirm, 555	confirm, 113	google, 287	confirm, 137
well, 456	like, 98	completing, 250	well, 126
completing, 440	well, 91	well, 239	completing, 114

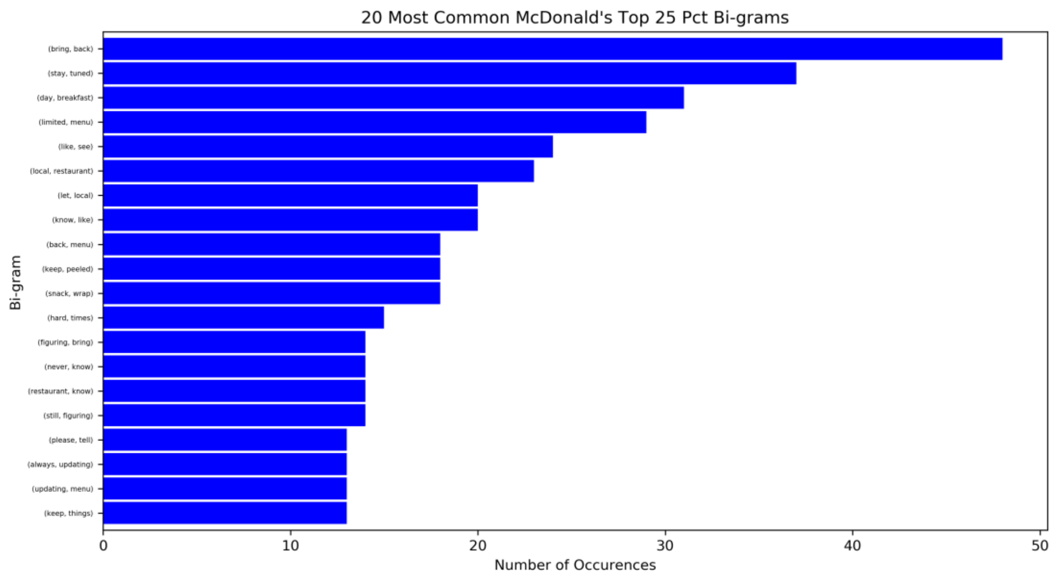
-It seems that the large volume of pre-formatted tweets is making it difficult to extract much more than the words in that tweet for Google. There are no words uniquely common to any list, except 'like' is uniquely common to the top 25% of Google tweets. It seems worthwhile to address large volumes of pre-formatted tweets in wordcount analysis. Otherwise, the presence of the same tweet over and over again imbalances wordcounts.

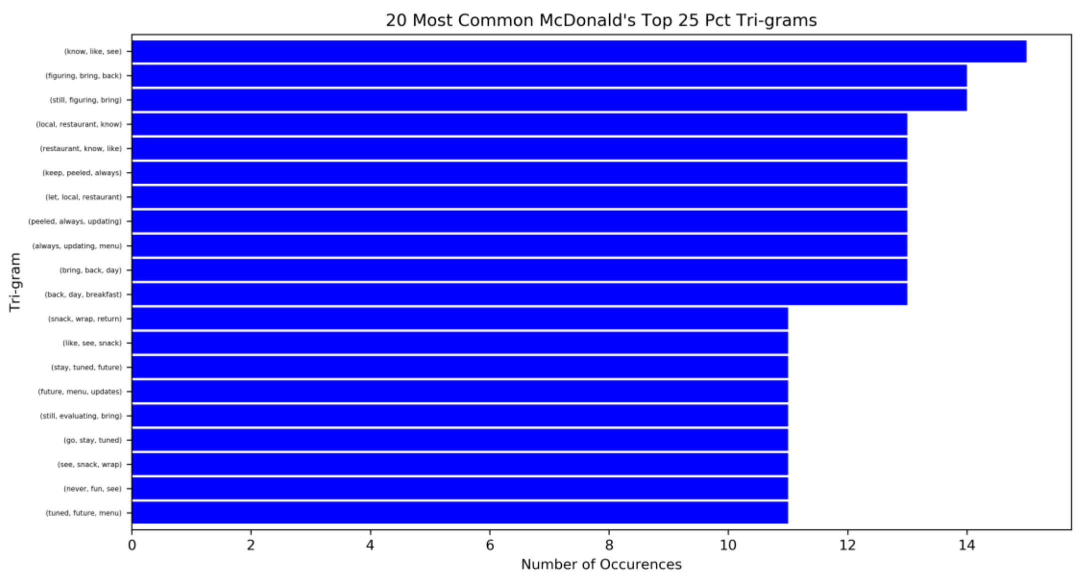
McDonalds

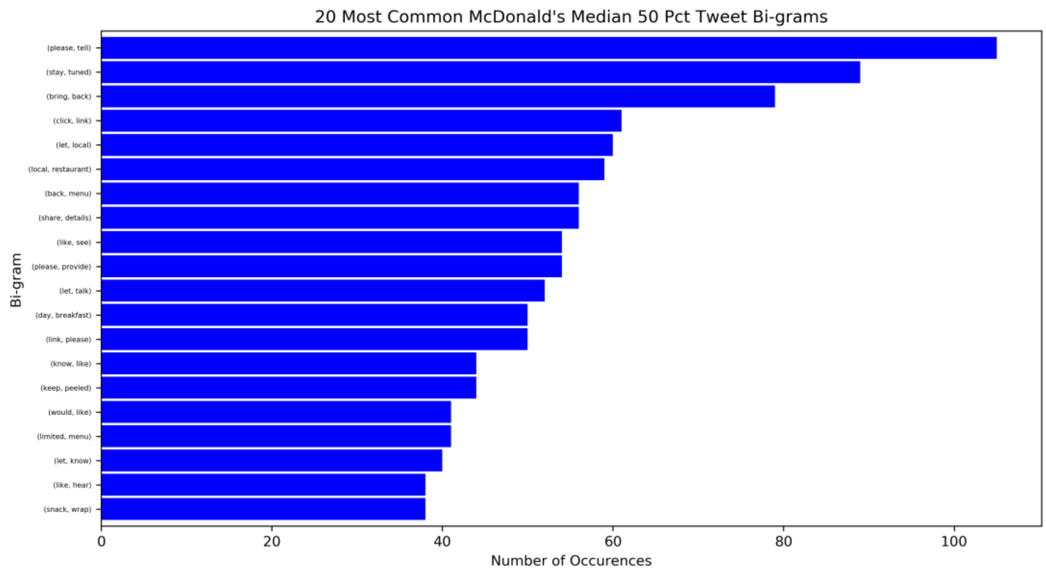
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	3206	12.10	38757	2431
Top 25% IRT:	799	9.581977	7656	1187
Median 50% IRT:	1596	12.40852	19804	1579
Bottom 25% IRT:	799	13.93242	11132	1130

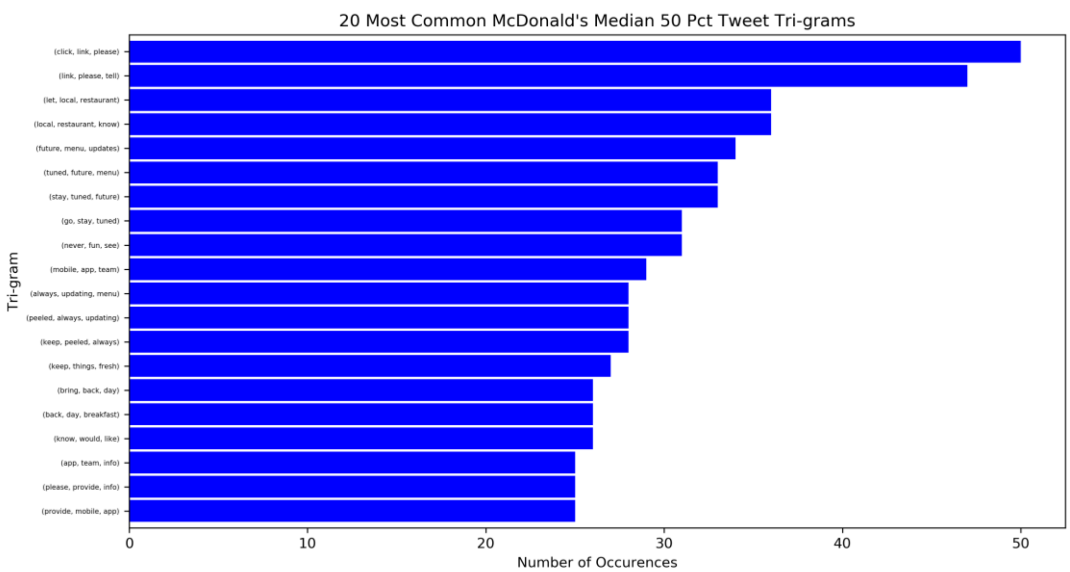


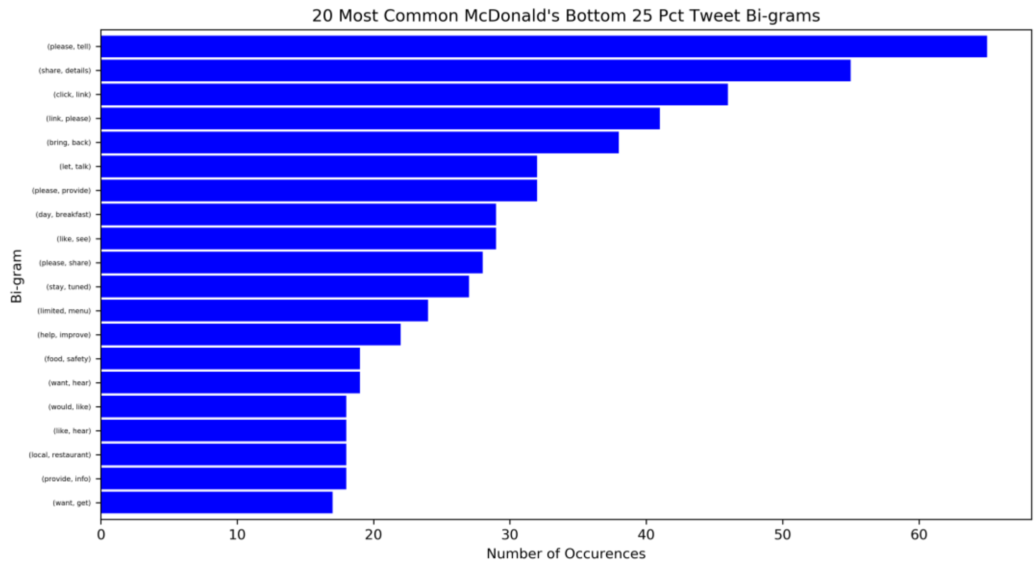


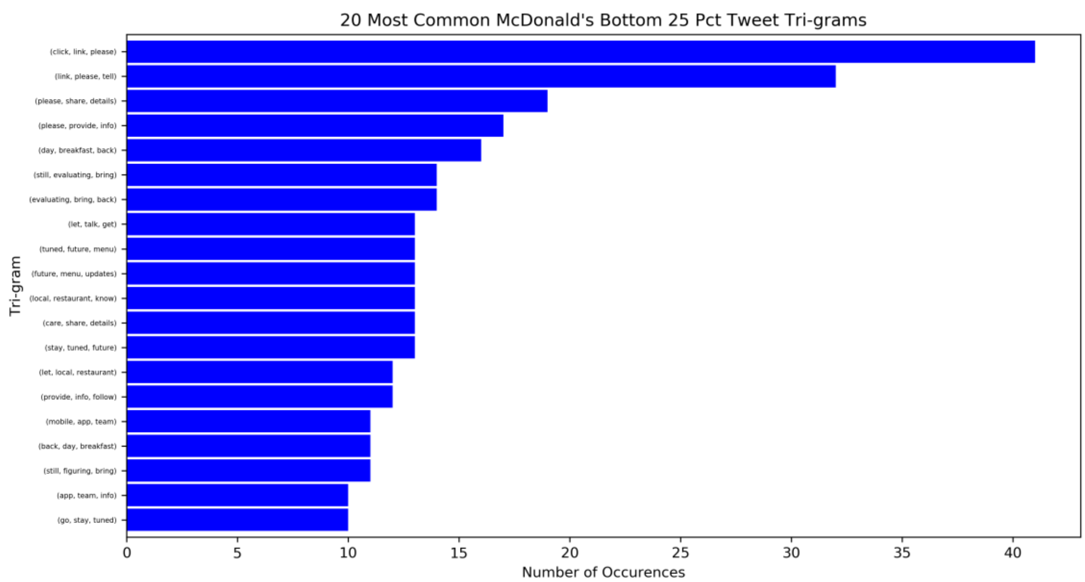

20 Most Common McDonald's Top 25 Pct Bi-grams
(bring, back)
(stay, tuned)
(day, breakfast)
(limited, menu)
(like, see)
(local, restaurant)
(let, local)
(know, like)
(back, menu)
(keep, peeled)
(snack, wrap)
(hard, times)
(figuring, bring)
(never, know)
(restaurant, know)
(still, figuring)
(please, tell)
(always, updating)
(updating, menu)
(keep, things)
Bi-gram
0
10
20
30
40
50
Number of Occurences


20 Most Common McDonald's Top 25 Pct Tri-grams
(know, like, see)
(figuring, bring, back)
(still, figuring, bring)
(local, restaurant, know)
(restaurant, know, like)
(keep, peeled, always)
(let, local, restaurant)
(peeled, always, updating)
(always, updating, menu)
(bring, back, day)
(back, day, breakfast)
(snack, wrap, return)
(like, see, snack)
(stay, tuned, future)
(future, menu, updates)
(still, evaluating, bring)
(go, stay, tuned)
(see, snack, wrap)
(never, fun, see)
(tuned, future, menu)
Tri-gram
0
2
4
6
8
10
12
14
Number of Occurences


20 Most Common McDonald's Median 50 Pct Tweet Bi-grams
Bi-gram
Number of Occurences
(please, tell)
(stay, tuned)
(bring, back)
(click, link)
(let, local)
(local, restaurant)
(back, menu)
(share, details)
(like, see)
(please, provide)
(let, talk)
(day, breakfast)
(link, please)
(know, like)
(keep, peeled)
(would, like)
(limited, menu)
(let, know)
(like, hear)
(snack, wrap)
0
20
40
60
80
100


20 Most Common McDonald's Median 50 Pct Tweet Tri-grams
Tri-gram
Number of Occurences
(click, link, please)
(link, please, tell)
(let, local, restaurant)
(local, restaurant, know)
(future, menu, updates)
(tuned, future, menu)
(stay, tuned, future)
(go, stay, tuned)
(never, fun, see)
(mobile, app, team)
(always, updating, menu)
(peeled, always, updating)
(keep, peeled, always)
(keep, things, fresh)
(bring, back, day)
(back, day, breakfast)
(know, would, like)
(app, team, info)
(please, provide, info)
(provide, mobile, app)
0
10
20
30
40
50





McDonalds Top Words

All Tweets	Top 25%	Median 50%	Bottom 25%
tell, 579	menu, 112	tell, 321	tell, 218
menu, 457	back, 96	menu, 249	please, 160
please, 444	like, 74	please, 242	details, 125
like, 432	know, 67	like, 240	like, 118
back, 419	see, 64	back, 226	back, 97
know, 322	still, 62	know, 176	menu, 96
let, 299	bring, 58	let, 173	let, 80
see, 260	let, 47	see, 138	know, 80
details, 245	never, 47	local, 133	share, 77
local, 215	day, 46	things, 109	link, 72
bring, 199	keep, 44	keep, 105	want, 71
get, 196	local, 43	get, 105	get, 69
things, 176	please, 42	details, 103	hear, 62
good, 174	stay, 42	bring, 96	see, 58

keep, 174	tell, 40	stay, 95	good, 56
never, 174	tuned, 37	good, 93	help, 55
want, 171	hi, 37	never, 93	click, 46
hear, 171	new, 35	tuned, 89	provide, 46
still, 169	breakfast, 35	always, 84	bring, 45
stay, 167	limited, 34	hi, 83	right, 44

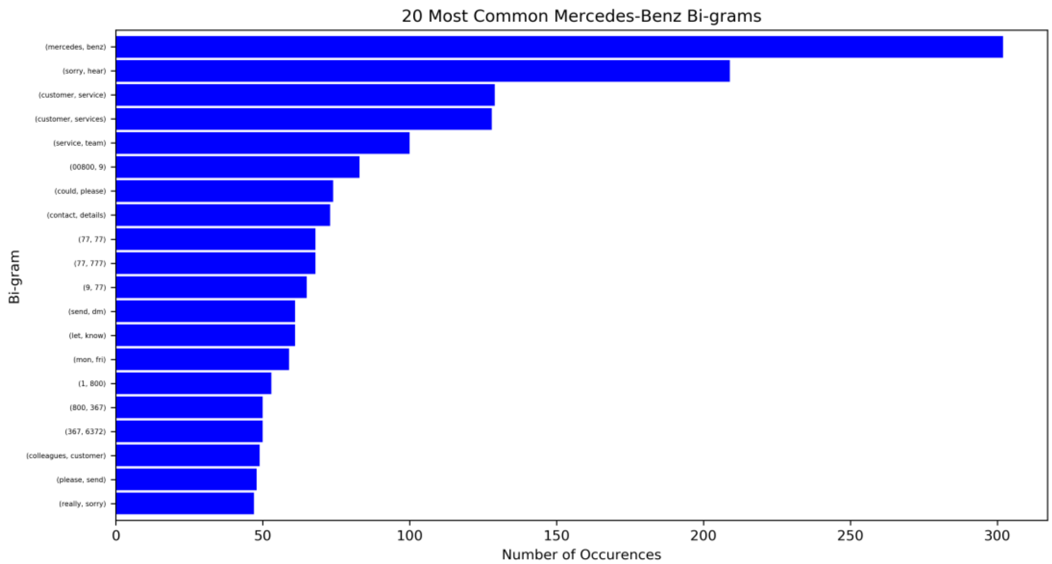
-Following are the words uniquely common to each of the above categories:

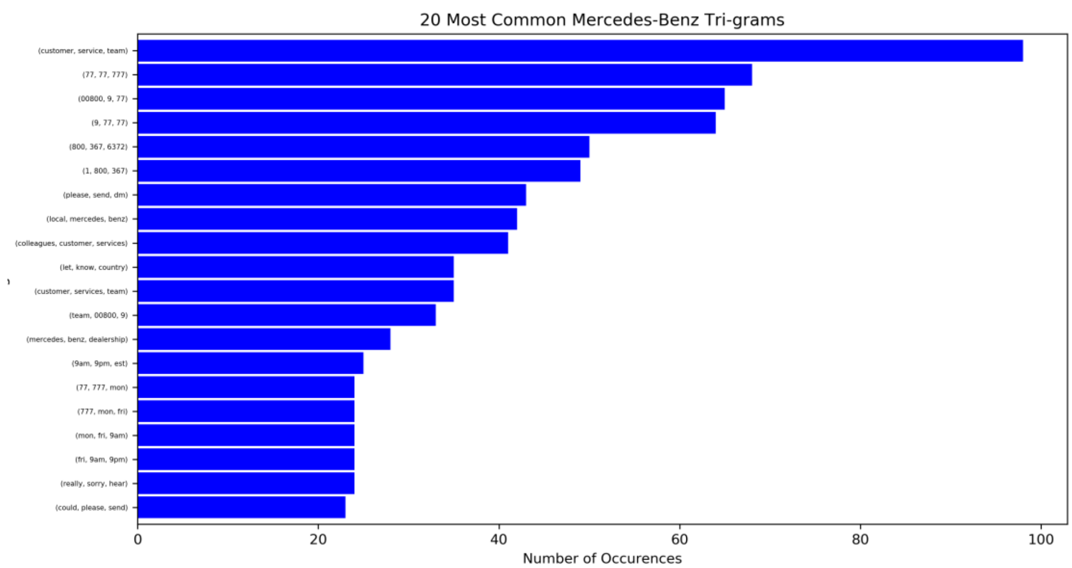
All Tweets	Top 25%	Median 50%	Bottom 25%
	day	always	share
	new		link
	breakfast		help
	limited		click
			provide
			right

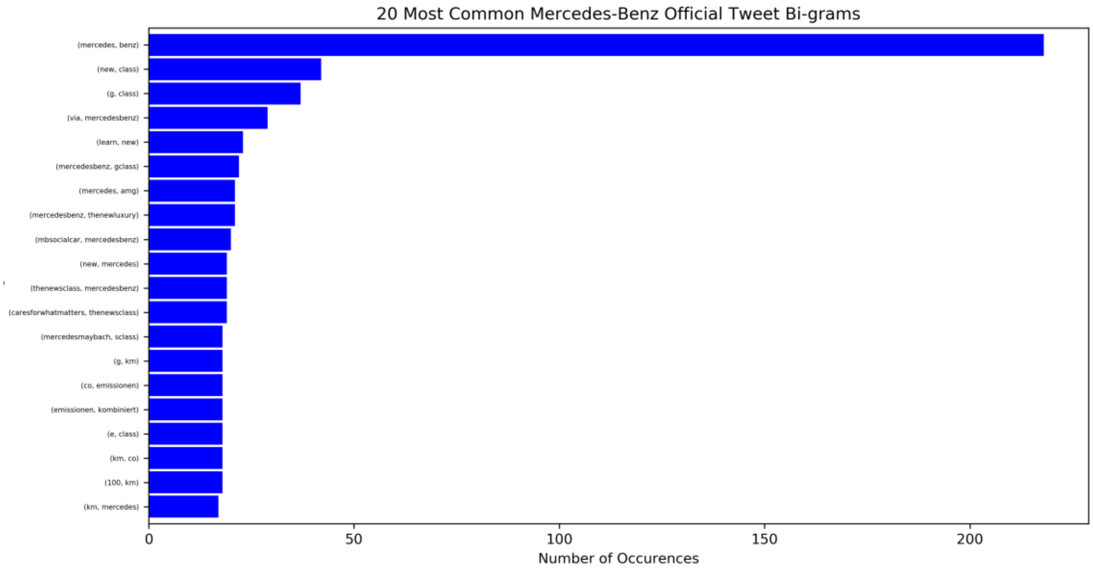
Mercedes

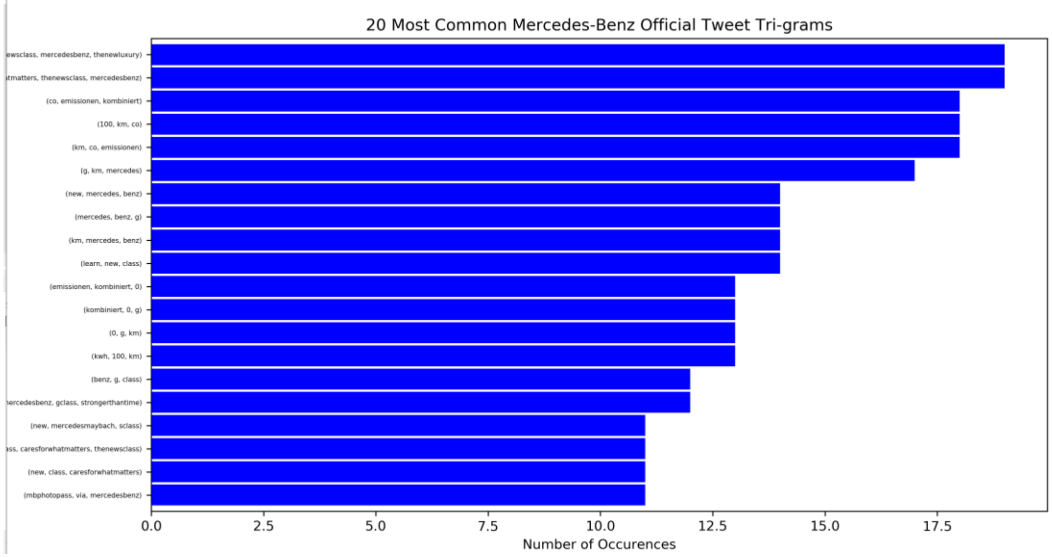
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	3031	16.21	49134	4213
Official Tweets:	505	27.39	13832	2529
Top 25% Official:	127	24.61	3125	983
Median 50% Official:	251	28.22	7084	1691
Bottom 25% Official:	127	28.64	3637	1145
IRT Tweets:	2526	13.98	35302	2616
Top 25% IRT:	632	9.80	6192	1276
Median 50% IRT:	1262	13.45	16979	1678
Bottom 25% IRT:	632	19.24	12159	1292

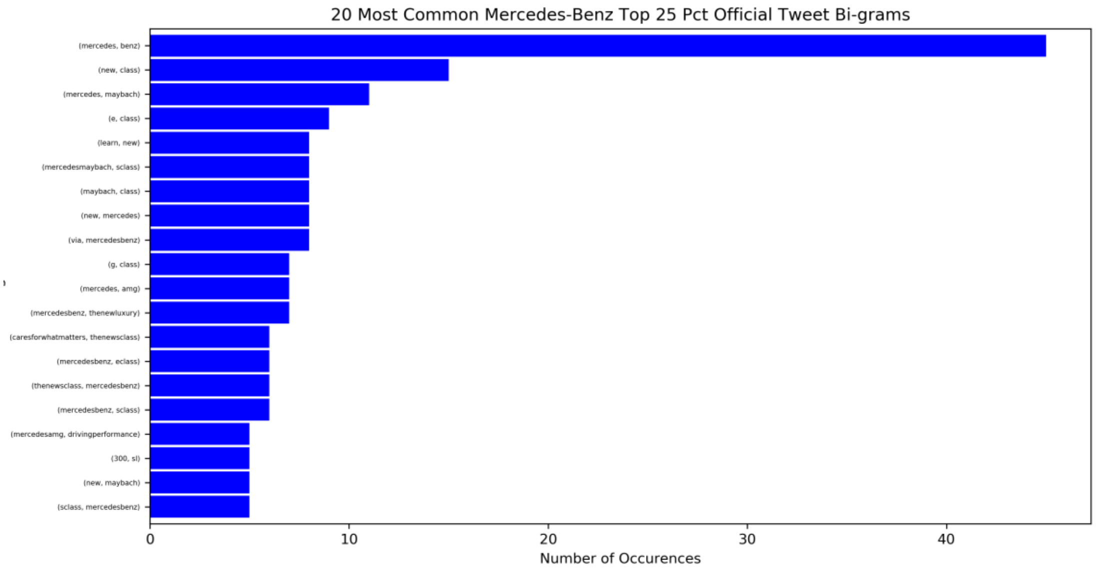
-Mercedes seems to be another company for which the best IRT tweets, on average, are shorter than median performing and poor performing IRT tweets.

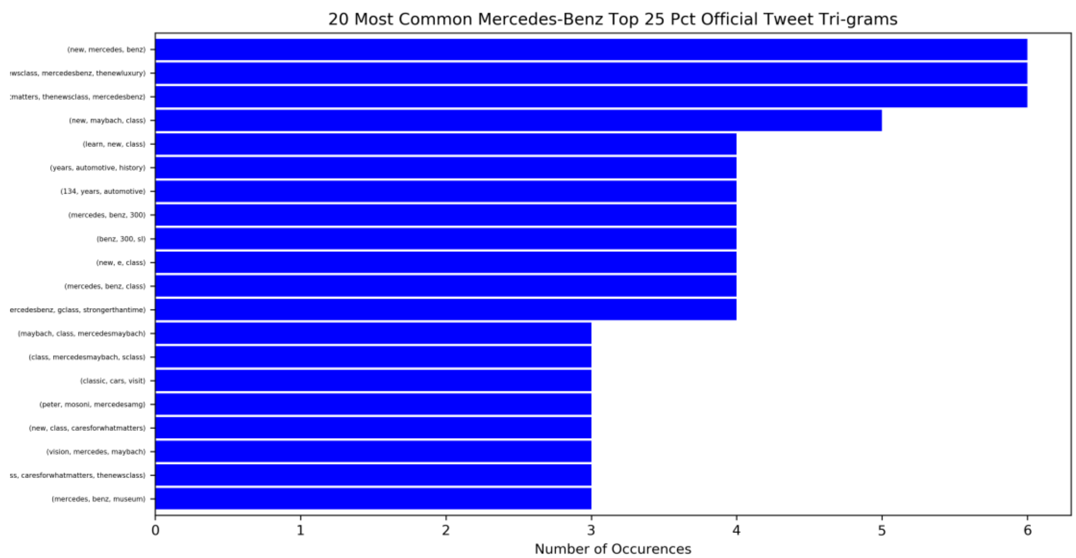

20 Most Common Mercedes-Benz Bi-grams
(mercedes, benz)
(sorry, hear)
(customer, service)
(customer, services)
(service, team)
(00800, 9)
(could, please)
(contact, details)
(77, 77)
(77, 777)
(9, 77)
(send, dm)
(let, know)
(mon, fri)
(1, 800)
(800, 367)
(367, 6372)
(colleagues, customer)
(please, send)
(really, sorry)
Bi-gram
0
50
100
150
200
250
300
Number of Occurences

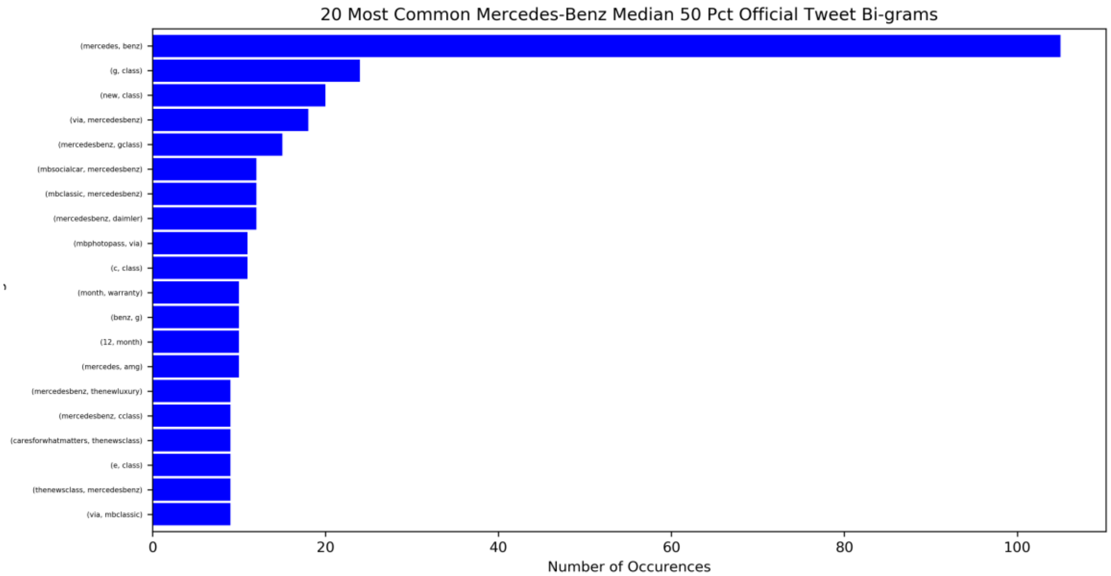

20 Most Common Mercedes-Benz Tri-grams
(customer, service, team)
(77, 77, 777)
(00800, 9, 77)
(9, 77, 77)
(800, 367, 6372)
(1, 800, 367)
(please, send, dm)
(local, mercedes, benz)
(colleagues, customer, services)
(let, know, country)
(customer, services, team)
(team, 00800, 9)
(mercedes, benz, dealership)
(9am, 9pm, est)
(77, 777, mon)
(777, mon, fri)
(mon, fri, 9am)
(fri, 9am, 9pm)
(really, sorry, hear)
(could, please, send)
0
20
40
60
80
100
Number of Occurences

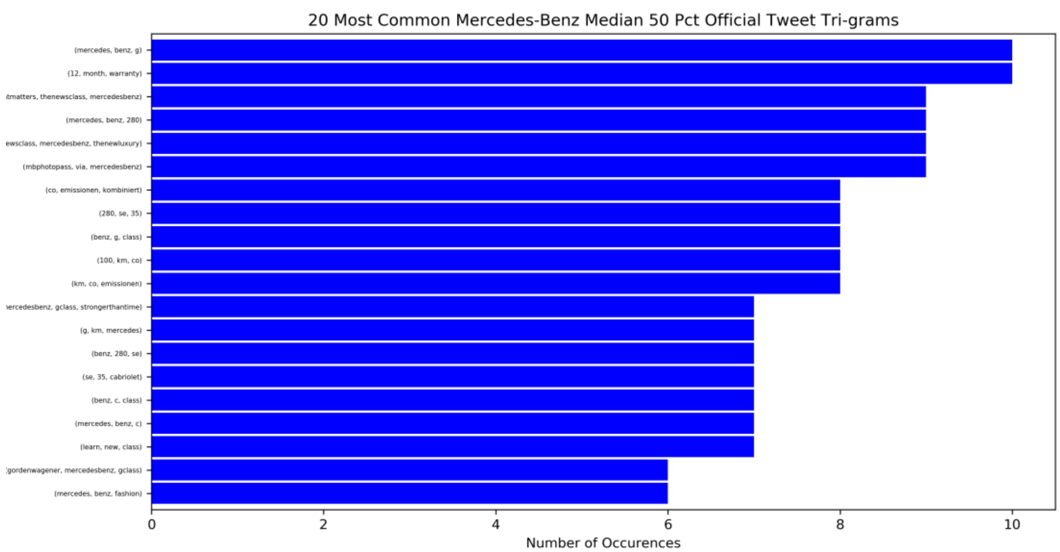

20 Most Common Mercedes-Benz Official Tweet Bi-grams
(mercedes, benz)
(new, class)
(g, class)
(via, mercedesbenz)
(learn, new)
(mercedesbenz, gclass)
(mercedes, amg)
(mercedesbenz, thenewluxury)
(mbsocialcar, mercedesbenz)
(new, mercedes)
(thenewsclass, mercedesbenz)
(caresforwhatmatters, thenewsclass)
(mercedesmaybach, sclass)
(g, km)
(co, emissionen)
(emissionen, kombiniert)
(e, class)
(km, co)
(100, km)
(km, mercedes)
0
50
100
150
200
Number of Occurences

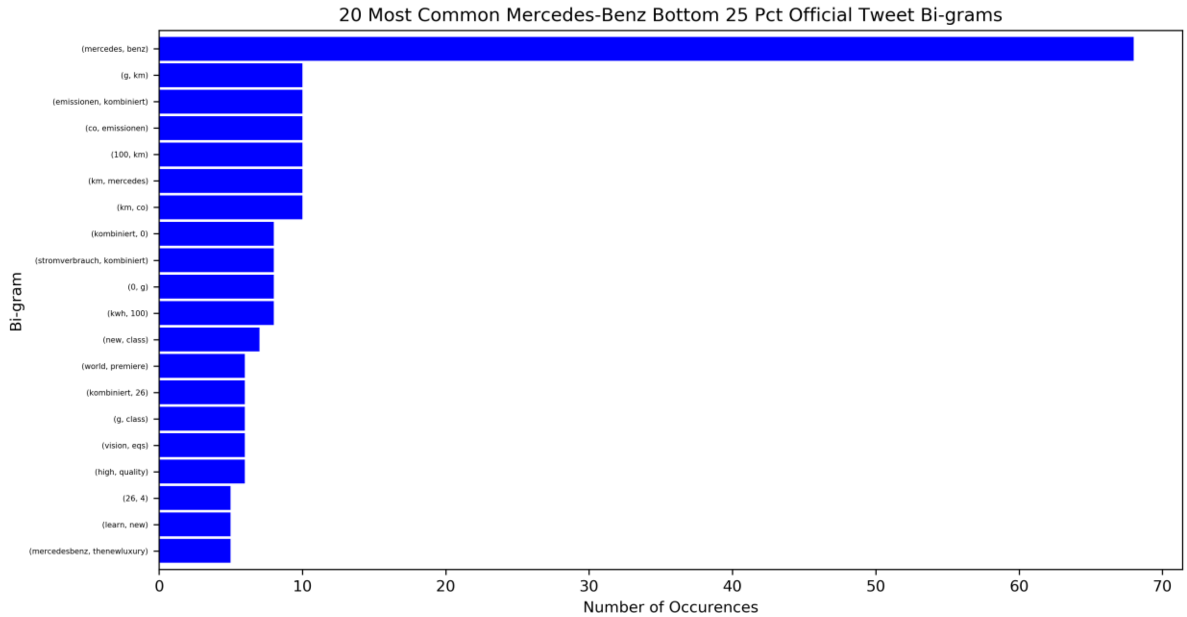

20 Most Common Mercedes-Benz Official Tweet Tri-grams
ewsclass, mercedesbenz, thenewluxury)
tmatters, thenewsclass, mercedesbenz)
(co, emissionen, kombiniert)
(100, km, co)
(km, co, emissionen)
(g, km, mercedes)
(new, mercedes, benz)
(mercedes, benz, g)
(km, mercedes, benz)
(learn, new, class)
(emissionen, kombiniert, 0)
(kombiniert, 0, g)
(0, g, km)
(kwh, 100, km)
(benz, g, class)
ercedesbenz, gclass, strongerthantime)
(new, mercedesmaybach, sclass)
ss, caresforwhatmatters, thenewsclass)
(new, class, caresforwhatmatters)
(mbphotopass, via, mercedesbenz)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Number of Occurences

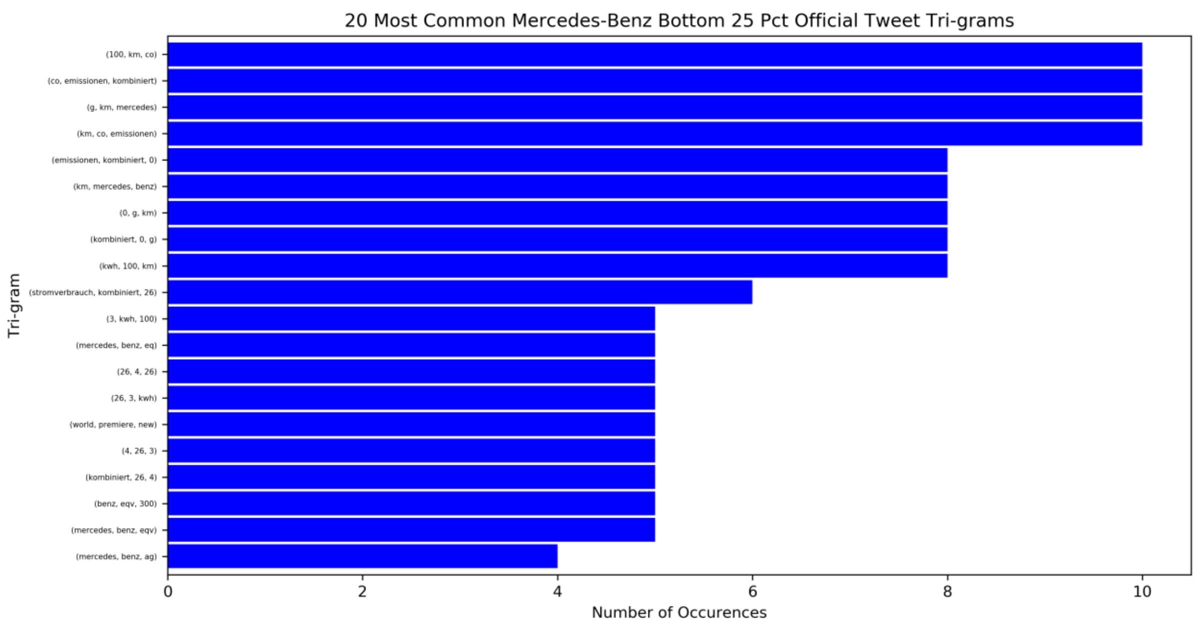

20 Most Common Mercedes-Benz Top 25 Pct Official Tweet Bi-grams
(mercedes, benz)
(new, class)
(mercedes, maybach)
(e, class)
(learn, new)
(mercedesmaybach, sclass)
(maybach, class)
(new, mercedes)
(via, mercedesbenz)
(g, class)
(mercedes, amg)
(mercedesbenz, thenewluxury)
(caresforwhatmatters, thenewsclass)
(mercedesbenz, eclass)
(thenewsclass, mercedesbenz)
(mercedesbenz, sclass)
(mercedesamg, drivingperformance)
(300, sl)
(new, maybach)
(sclass, mercedesbenz)
0
10
20
30
40
Number of Occurences

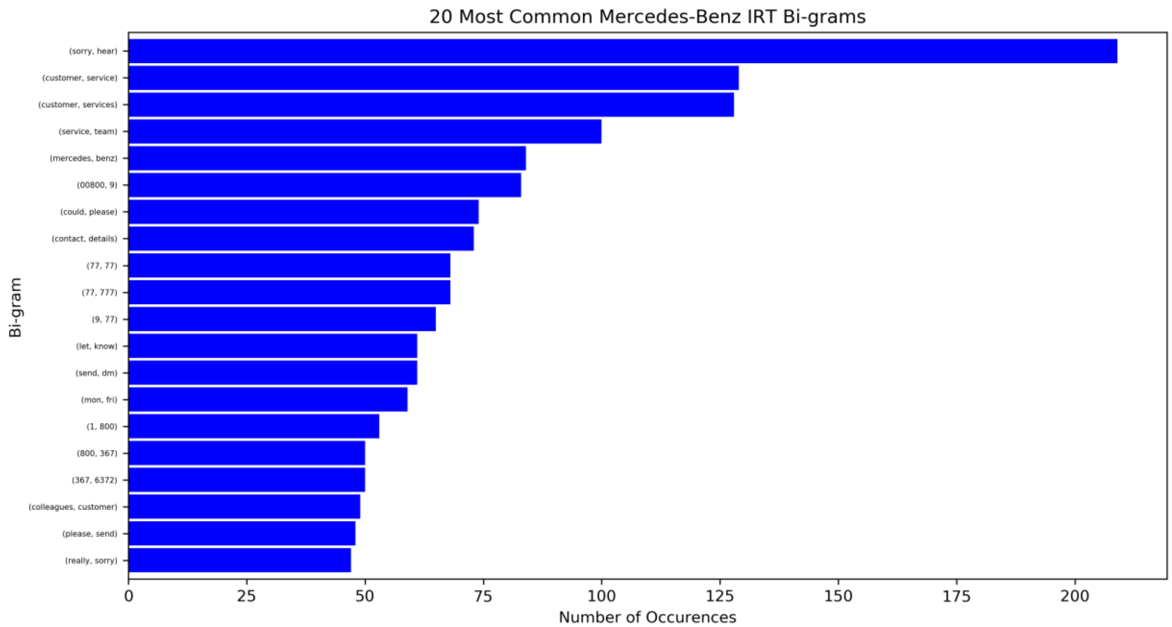

20 Most Common Mercedes-Benz Top 25 Pct Official Tweet Tri-grams
(new, mercedes, benz)
wsclass, mercedesbenz, thenewluxury)
matters, thenewsclass, mercedesbenz)
(new, maybach, class)
(learn, new, class)
(years, automotive, history)
(134, years, automotive)
(mercedes, benz, 300)
(benz, 300, sl)
(new, e, class)
(mercedes, benz, class)
ercedesbenz, gclass, strongerthantime)
(maybach, class, mercedesmaybach)
(class, mercedesmaybach, sclass)
(classic, cars, visit)
(peter, mosoni, mercedesamg)
(new, class, caresforwhatmatters)
(vision, mercedes, maybach)
ss, caresforwhatmatters, thenewsclass)
(mercedes, benz, museum)
0
1
2
3
4
5
6
Number of Occurences

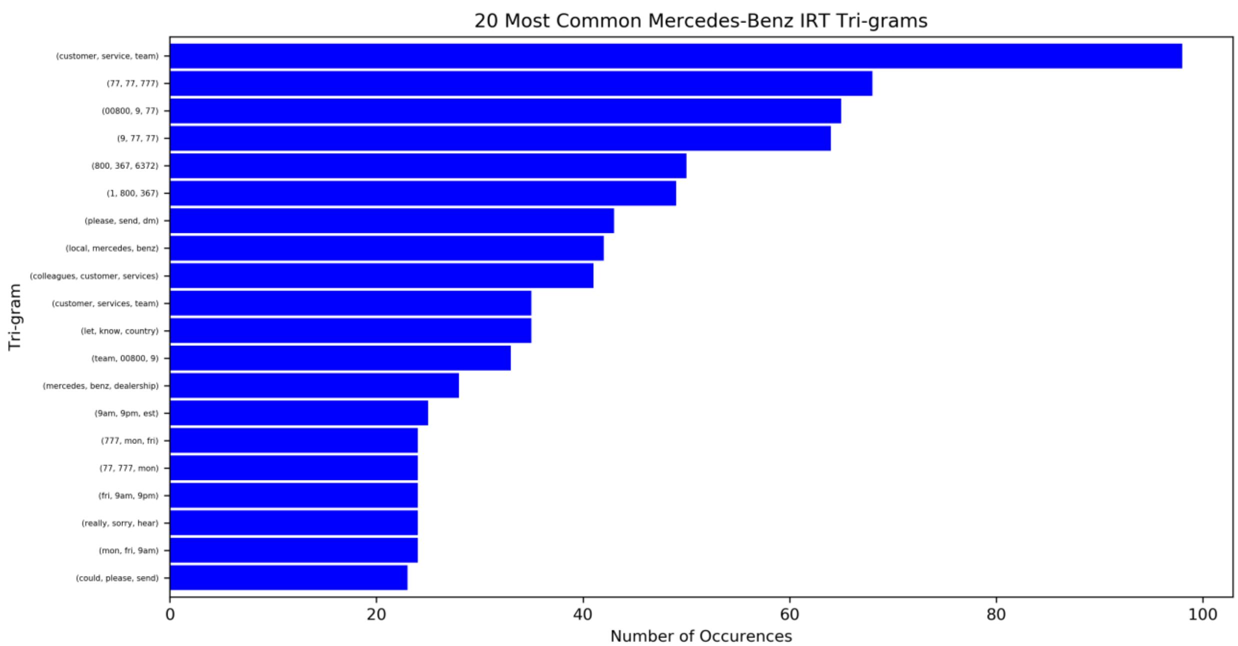

20 Most Common Mercedes-Benz Median 50 Pct Official Tweet Bi-grams
(mercedes, benz)
(g, class)
(new, class)
(via, mercedesbenz)
(mercedesbenz, gclass)
(mbsocialcar, mercedesbenz)
(mbclassic, mercedesbenz)
(mercedesbenz, daimler)
(mbphotopass, via)
(c, class)
(month, warranty)
(benz, g)
(12, month)
(mercedes, amg)
(mercedesbenz, thenewluxury)
(mercedesbenz, cclass)
(caresforwhatmatters, thenewsclass)
(e, class)
(thenewsclass, mercedesbenz)
(via, mbclassic)
0
20
40
60
80
100
Number of Occurences

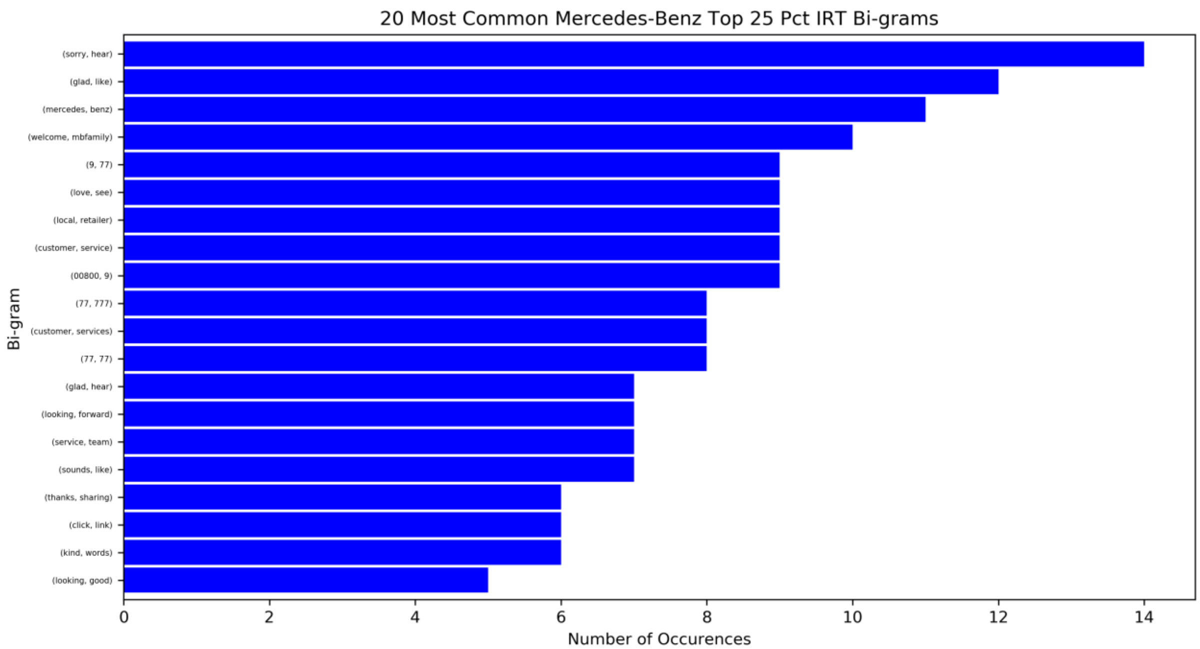

20 Most Common Mercedes-Benz Median 50 Pct Official Tweet Tri-grams
(mercedes, benz, g)
(12, month, warranty)
tmatters, thenewsclass, mercedesbenz)
(mercedes, benz, 280)
ewsclass, mercedesbenz, thenewluxury)
(mbphotopass, via, mercedesbenz)
(co, emissionen, kombiniert)
(280, se, 35)
(benz, g, class)
(100, km, co)
(km, co, emissionen)
ercedesbenz, gclass, strongerthantime)
(g, km, mercedes)
(benz, 280, se)
(se, 35, cabriolet)
(benz, c, class)
(mercedes, benz, c)
(learn, new, class)
gordenwagener, mercedesbenz, gclass)
(mercedes, benz, fashion)
0
2
4
6
8
10
Number of Occurences


20 Most Common Mercedes-Benz Bottom 25 Pct Official Tweet Bi-grams
(mercedes, benz)
(g, km)
(emissionen, kombiniert)
(co, emissionen)
(100, km)
(km, mercedes)
(km, co)
(kombiniert, 0)
(stromverbrauch, kombiniert)
(0, g)
(kwh, 100)
(new, class)
(world, premiere)
(kombiniert, 26)
(g, class)
(vision, eqs)
(high, quality)
(26, 4)
(learn, new)
(mercedesbenz, thenewluxury)
Bi-gram
0
10
20
30
40
50
60
70
Number of Occurences

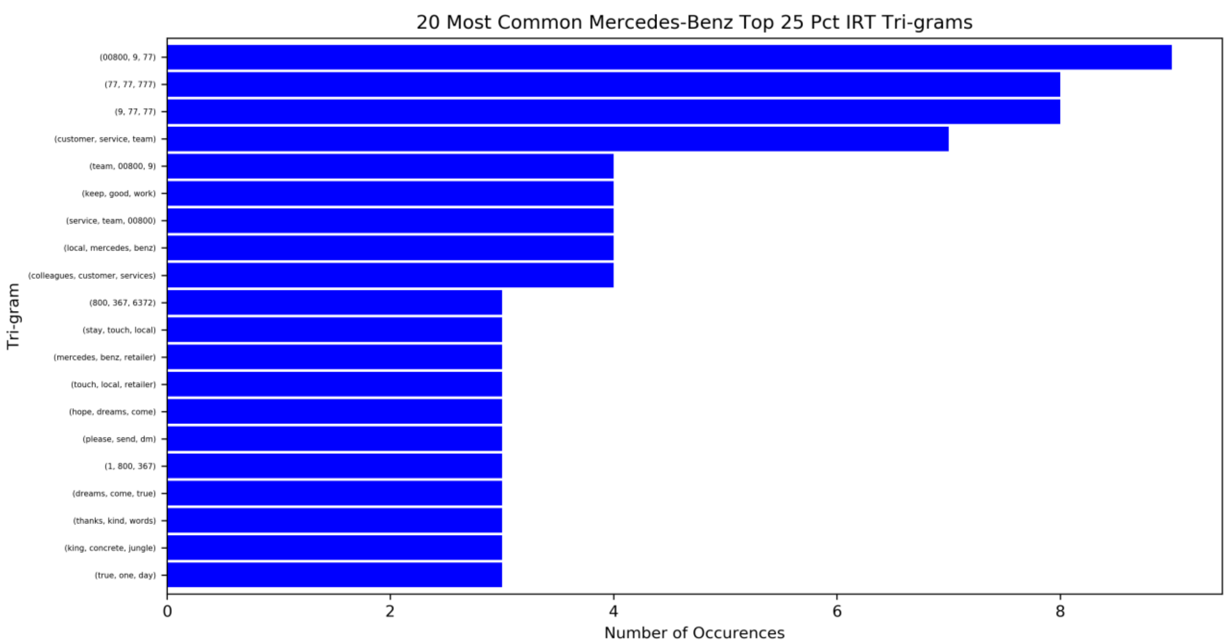

20 Most Common Mercedes-Benz Bottom 25 Pct Official Tweet Tri-grams
(100, km, co)
(co, emissionen, kombiniert)
(g, km, mercedes)
(km, co, emissionen)
(emissionen, kombiniert, 0)
(km, mercedes, benz)
(0, g, km)
(kombiniert, 0, g)
(kwh, 100, km)
(stromverbrauch, kombiniert, 26)
(3, kwh, 100)
(mercedes, benz, eq)
(26, 4, 26)
(26, 3, kwh)
(world, premiere, new)
(4, 26, 3)
(kombiniert, 26, 4)
(benz, eqv, 300)
(mercedes, benz, eqv)
(mercedes, benz, ag)
Tri-gram
0
2
4
6
8
10
Number of Occurences


20 Most Common Mercedes-Benz IRT Bi-grams
Bi-gram
Number of Occurences
(sorry, hear)
(customer, service)
(customer, services)
(service, team)
(mercedes, benz)
(00800, 9)
(could, please)
(contact, details)
(77, 77)
(77, 777)
(9, 77)
(let, know)
(send, dm)
(mon, fri)
(1, 800)
(800, 367)
(367, 6372)
(colleagues, customer)
(please, send)
(really, sorry)
0
25
50
75
100
125
150
175
200

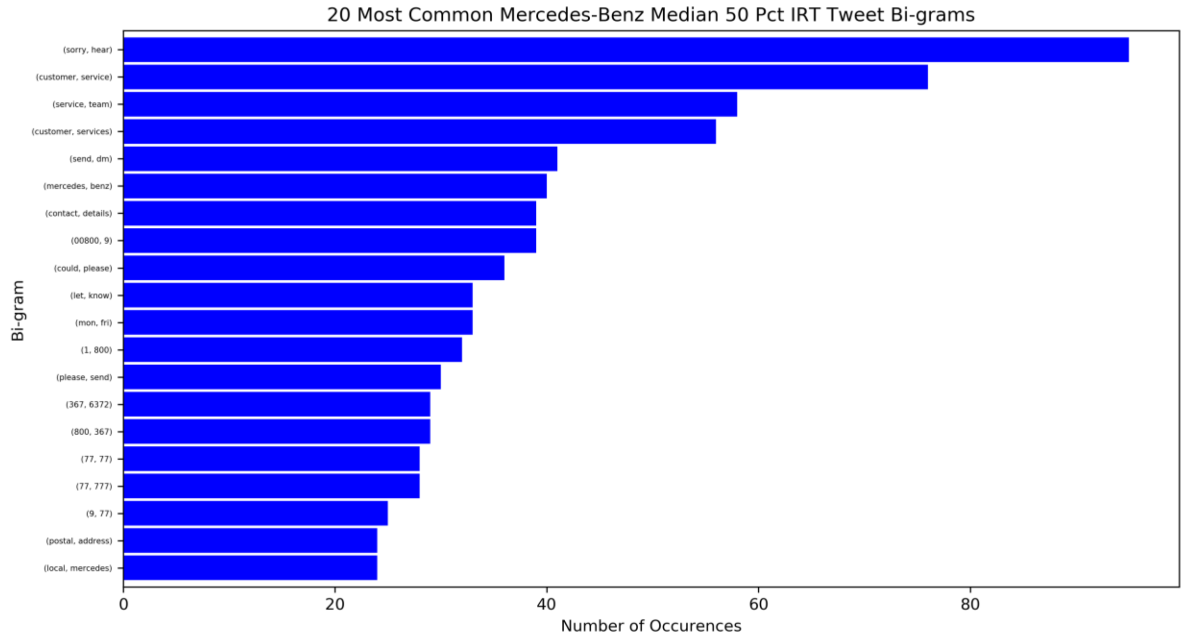

20 Most Common Mercedes-Benz IRT Tri-grams
Tri-gram
Number of Occurences
(customer, service, team)
(77, 77, 777)
(00800, 9, 77)
(9, 77, 77)
(800, 367, 6372)
(1, 800, 367)
(please, send, dm)
(local, mercedes, benz)
(colleagues, customer, services)
(customer, services, team)
(let, know, country)
(team, 00800, 9)
(mercedes, benz, dealership)
(9am, 9pm, est)
(777, mon, fri)
(77, 777, mon)
(fri, 9am, 9pm)
(really, sorry, hear)
(mon, fri, 9am)
(could, please, send)
0
20
40
60
80
100

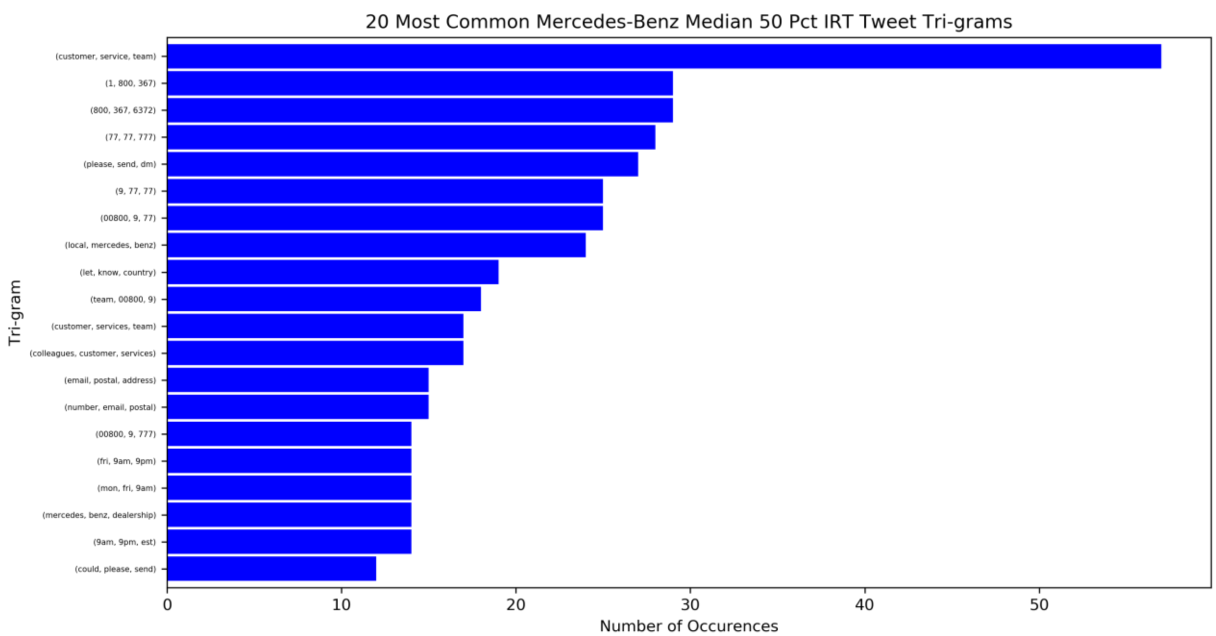


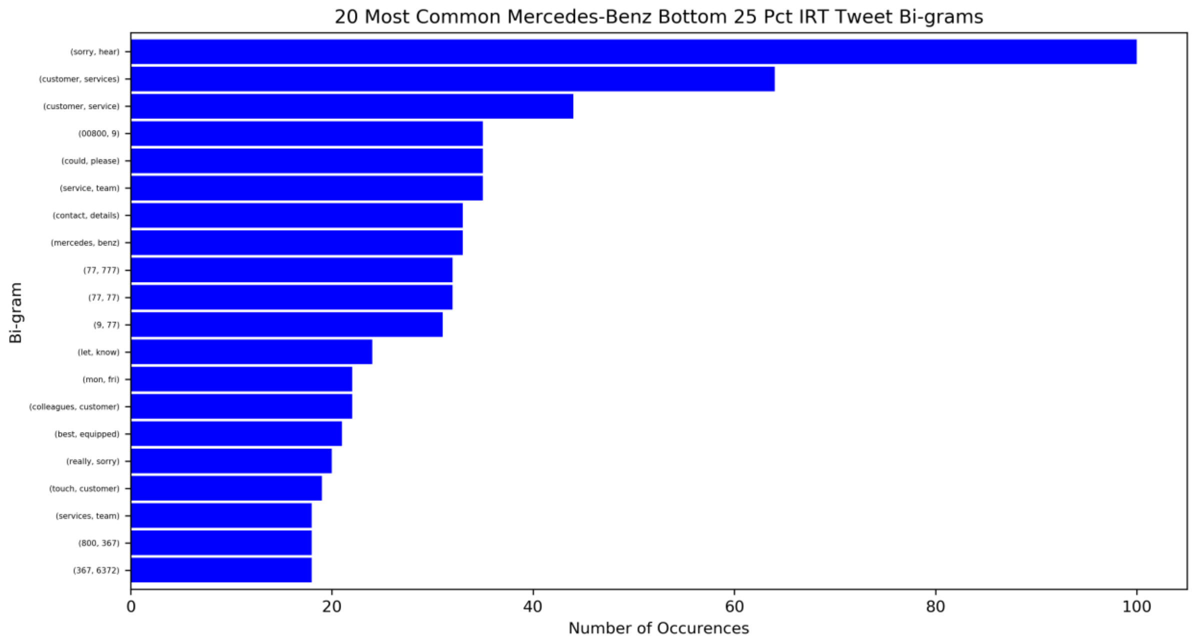
20 Most Common Mercedes-Benz Top 25 Pct IRT Bi-grams
Bi-gram
Number of Occurences
(sorry, hear)
(glad, like)
(mercedes, benz)
(welcome, mbfamily)
(9, 77)
(love, see)
(local, retailer)
(customer, service)
(00800, 9)
(77, 777)
(customer, services)
(77, 77)
(glad, hear)
(looking, forward)
(service, team)
(sounds, like)
(thanks, sharing)
(click, link)
(kind, words)
(looking, good)
0
2
4
6
8
10
12
14

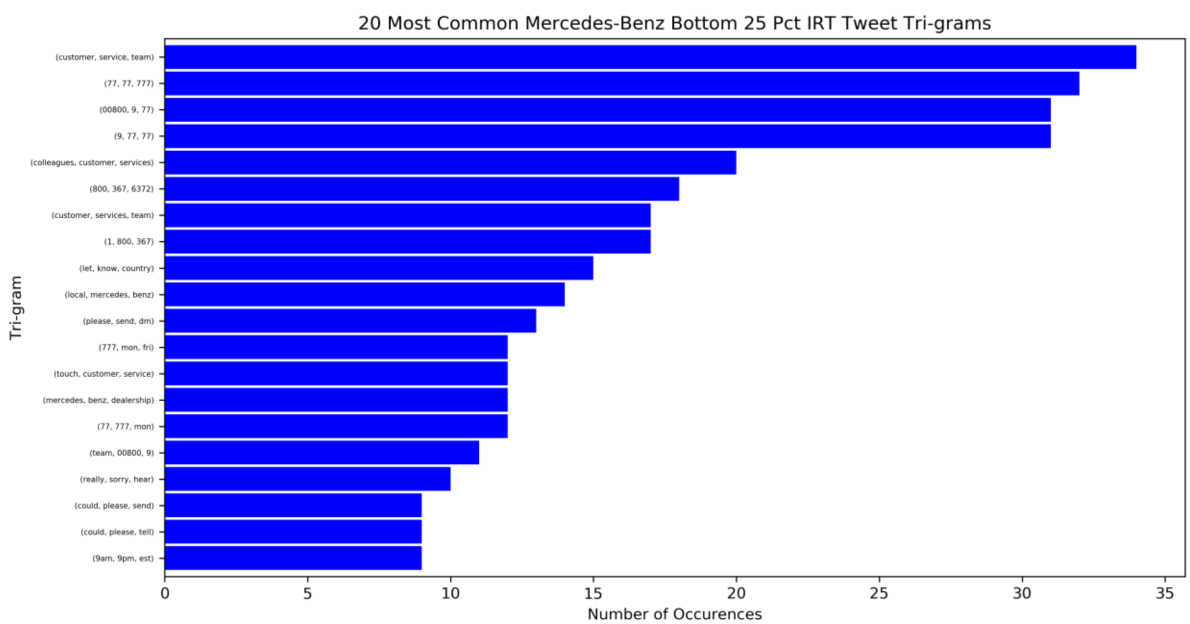


20 Most Common Mercedes-Benz Top 25 Pct IRT Tri-grams
Tri-gram
Number of Occurences
(00800, 9, 77)
(77, 77, 777)
(9, 77, 77)
(customer, service, team)
(team, 00800, 9)
(keep, good, work)
(service, team, 00800)
(local, mercedes, benz)
(colleagues, customer, services)
(800, 367, 6372)
(stay, touch, local)
(mercedes, benz, retailer)
(touch, local, retailer)
(hope, dreams, come)
(please, send, dm)
(1, 800, 367)
(dreams, come, true)
(thanks, kind, words)
(king, concrete, jungle)
(true, one, day)
0
2
4
6
8


20 Most Common Mercedes-Benz Median 50 Pct IRT Tweet Bi-grams
(sorry, hear)
(customer, service)
(service, team)
(customer, services)
(send, dm)
(mercedes, benz)
(contact, details)
(00800, 9)
(could, please)
(let, know)
(mon, fri)
(1, 800)
(please, send)
(367, 6372)
(800, 367)
(77, 77)
(77, 777)
(9, 77)
(postal, address)
(local, mercedes)
Bi-gram
0
20
40
60
80
Number of Occurences


20 Most Common Mercedes-Benz Median 50 Pct IRT Tweet Tri-grams
(customer, service, team)
(1, 800, 367)
(800, 367, 6372)
(77, 77, 777)
(please, send, dm)
(9, 77, 77)
(00800, 9, 77)
(local, mercedes, benz)
(let, know, country)
(team, 00800, 9)
(customer, services, team)
(colleagues, customer, services)
(email, postal, address)
(number, email, postal)
(00800, 9, 777)
(fri, 9am, 9pm)
(mon, fri, 9am)
(mercedes, benz, dealership)
(9am, 9pm, est)
(could, please, send)
Tri-gram
0
10
20
30
40
50
Number of Occurences





Mercedes Top Words

All Tweets	**Official Tweets**	**Top 25% Official**	**Median 50% Official**	**Bottom 25% Official**	**All IRT Tweets**	**Top 25% IRT**	**Median 50% IRT**	**Bottom 25% IRT**
mercedes, 381	mercedesbenz, 300	mercedes, 66	mercedesbenz, 169	mercedes, 80	sorry, 345	thanks, 46	sorry, 164	sorry, 156
sorry, 346	mercedes, 267	mercedesbenz, 65	mercedes, 121	benz, 72	hear, 323	glad, 42	team, 154	hear, 140
hear, 323	benz, 224	class, 52	benz, 107	mercedesbenz, 67	team, 291	love, 36	hear, 149	team, 119

mercedesbenz, 312	new, 166	new, 48	class, 79	new, 40	customer, 261	like, 34	customer, 133	please, 111
benz, 309	class, 153	benz, 45	new, 77	class, 22	please, 258	hear, 34	please, 130	customer, 111
team, 296	learn, 77	learn, 22	via, 43	kombiniert, 20	colleagues, 248	great, 29	colleagues, 118	colleagues, 108
customer, 262	via, 73	luxury, 20	car, 42	km, 20	thanks, 202	hope, 27	contact, 102	contact, 70
please, 259	design, 63	sclass, 20	get, 42	g, 18	contact, 189	see, 25	thanks, 94	touch, 66
colleagues, 249	mbclassic, 63	design, 20	learn, 38	learn, 17	like, 155	sorry, 25	service, 86	services, 65
thanks, 210	g, 63	via, 19	amg, 37	next, 14	service, 149	happy, 25	like, 83	‘77’, 64
new, 197	amg, 62	Maybach, 19	mbclassic, 37	sclass, 13	thank, 146	thank, 25	dm, 72	thanks, 63
contact, 189	get, 62	amg, 17	g, 36	get, 13	hi, 143	good, 23	thank, 70	hi, 63
class, 187	luxury, 56	mercedesamg, 15	design, 33	mbclassic, 12	best, 141	best, 23	details, 69	dm, 53
like, 181	car, 56	car, 14	luxury, 31	world, 12	‘77’, 137	colleagues, 22	send, 69	service, 53
service, 161	sclass, 55	mbclassic, 14	sclass, 22	eqs, 12	touch, 136	one, 22	would, 68	details, 51
best, 150	mercedesamg, 44	road, 13	gle, 21	eq, 12	services, 134	mbfamily, 20	best, 68	help, 51
thank, 147	km, 39	vision, 13	cabriolet, 21	via, 11	dm, 132	sure, 19	hi, 63	thank, 51
hi, 143	kombiniert, 36	mercedesmaybach, 12	classic, 21	shopping, 11	details, 128	contact, 18	services, 60	best, 50
get, 142	vision, 35	drivingperformance, 11	coup, 20	‘26’, 11	local, 119	please, 18	know, 59	know, 46
‘77’, 137	drivingperformance, 34	one, 10	mercedesamg, 20	‘100’, 11	would, 118	team, 18	local, 58	local, 44

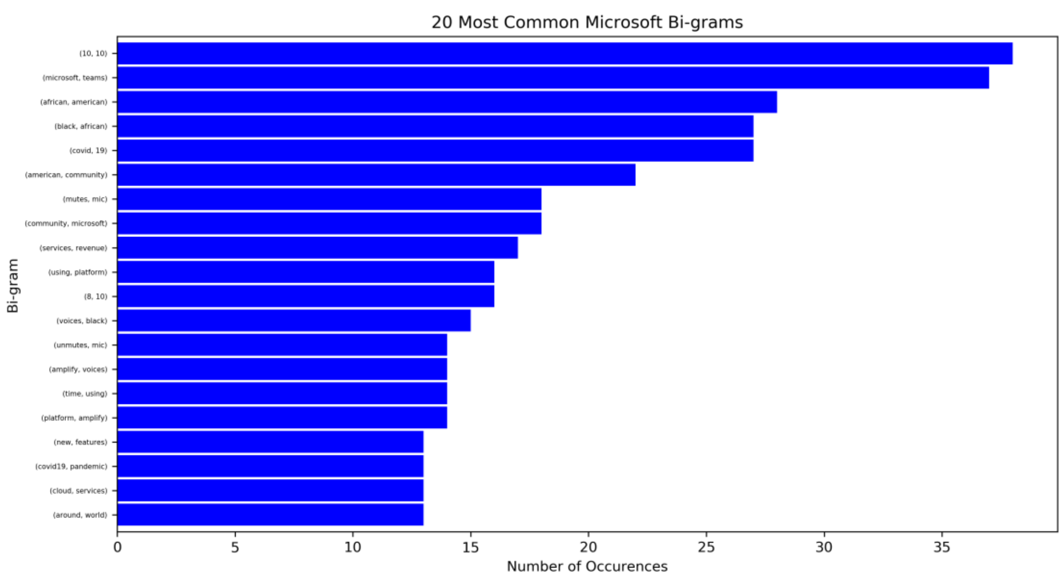
-Following are the words uniquely common to each of the above lists:

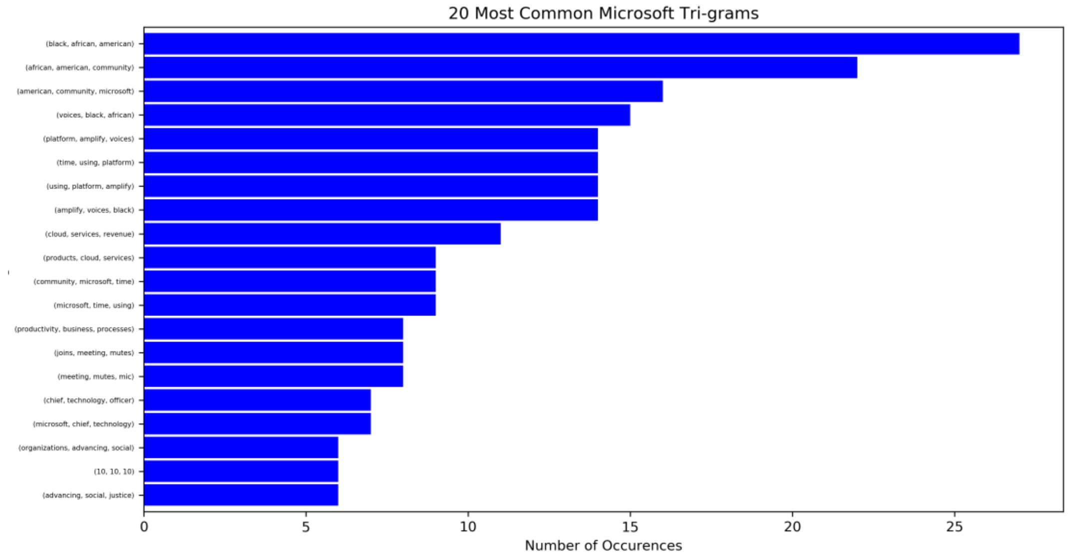
All Tweets	Official Tweets	Top 25% Official	Median 50% Official	Bottom 25% Official	All IRT Tweets	Top 25% IRT	Median 50% IRT	Bottom 25% IRT
		Maybach	gle	next		glad	send	help
		road	cabriolet	world		love		
		mercedesmaybach	classic	eqs		great		
			coup	eq		hope		
				shopping		see		
				26		happy		
				100		good		
						mbfamily		
						sure		

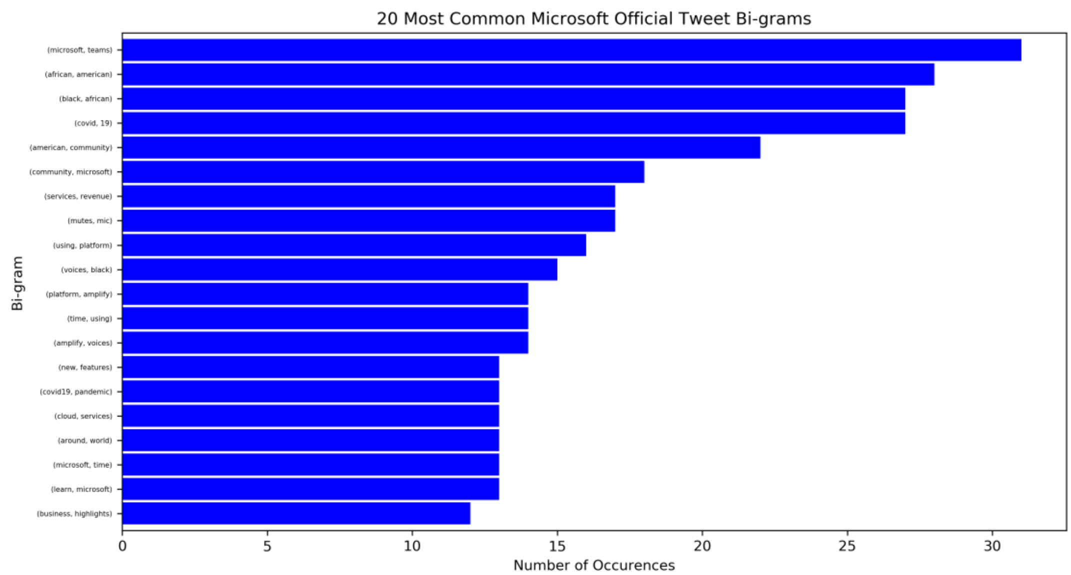
-Interestingly, the word ‘one’ is uniquely common to both top 25% lists, but no other lists.

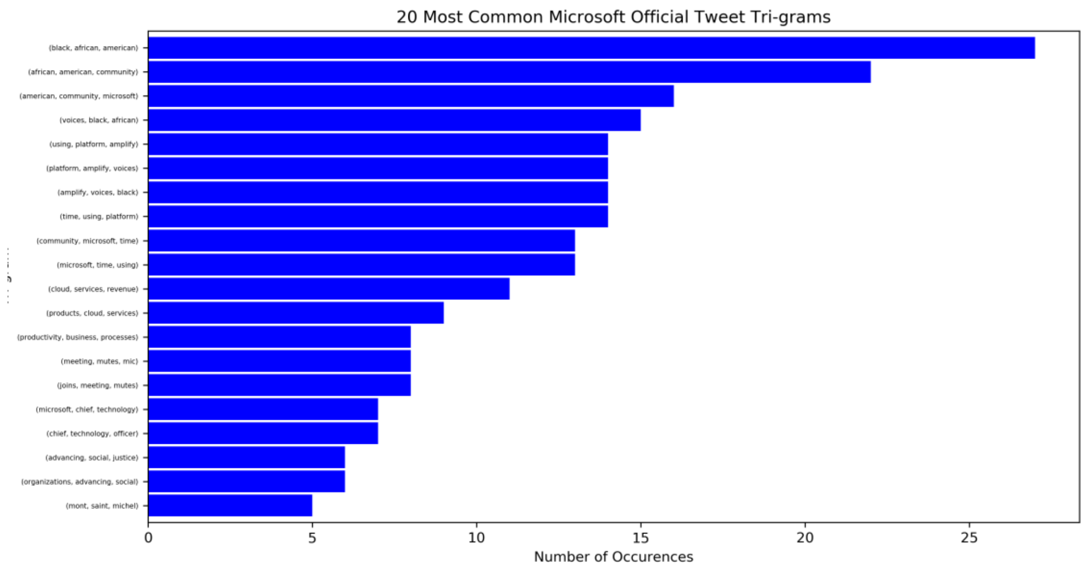
Microsoft

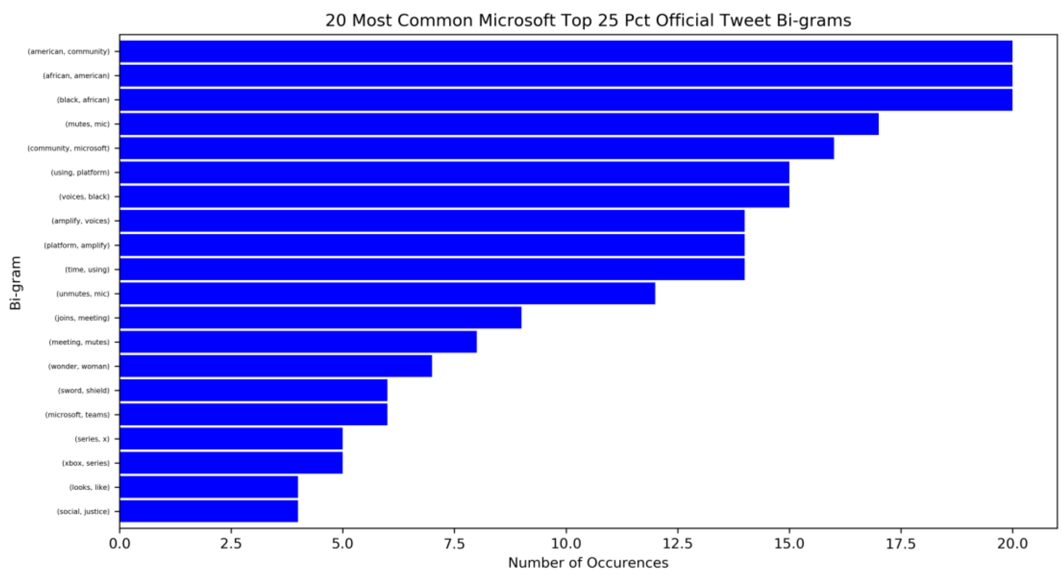
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	2622	12.00	31468	4910
Official Tweets:	1261	20.38	25701	4204
Top 25% Official:	316	16.09	5086	1453
Median 50% Official:	629	21.99	13831	2900
Bottom 25% Official:	316	21.47	6786	1972
IRT Tweets:	1361	4.24	5767	1557
Top 25% IRT:	341	4.94	1683	661
Median 50% IRT:	679	4.19	2843	984
Bottom 25% IRT:	341	3.62	1235	528

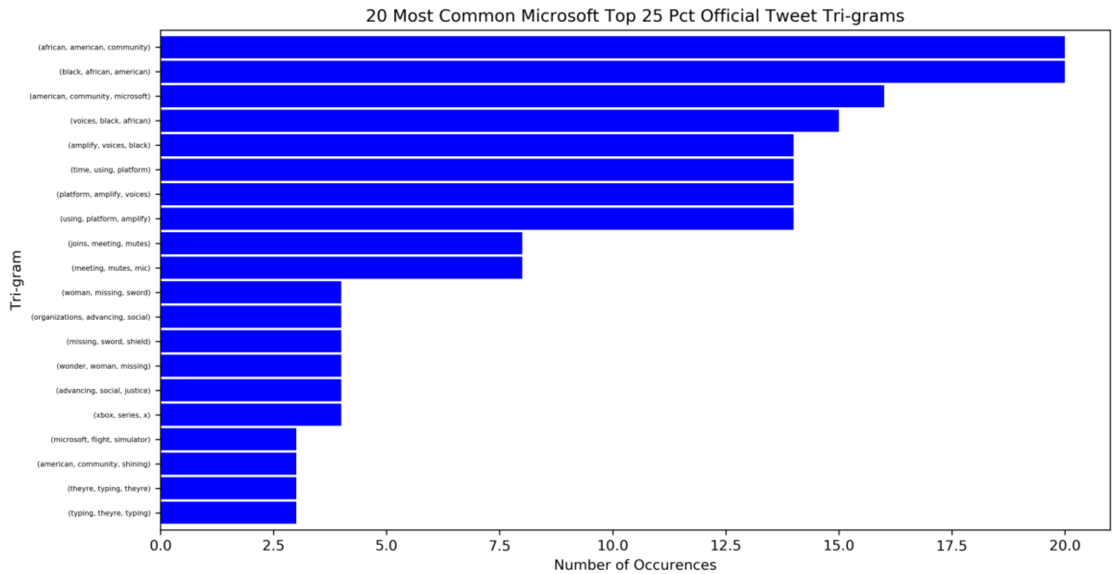


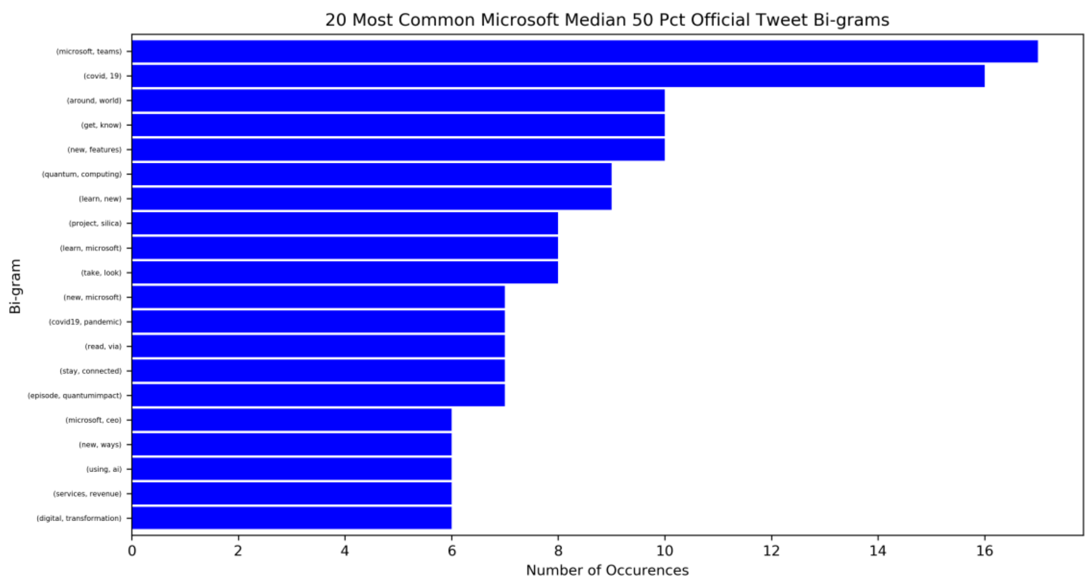

20 Most Common Microsoft Tri-grams
(black, african, american)
(african, american, community)
(american, community, microsoft)
(voices, black, african)
(platform, amplify, voices)
(time, using, platform)
(using, platform, amplify)
(amplify, voices, black)
(cloud, services, revenue)
(products, cloud, services)
(community, microsoft, time)
(microsoft, time, using)
(productivity, business, processes)
(joins, meeting, mutes)
(meeting, mutes, mic)
(chief, technology, officer)
(microsoft, chief, technology)
(organizations, advancing, social)
(10, 10, 10)
(advancing, social, justice)
0
5
10
15
20
25
Number of Occurences

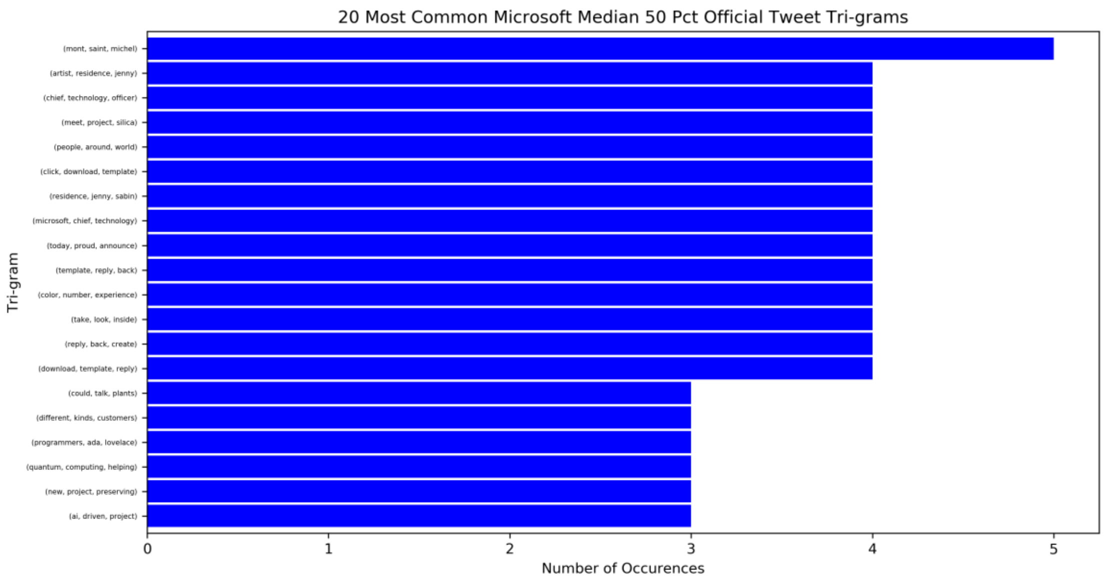

20 Most Common Microsoft Official Tweet Bi-grams
Bi-gram
(microsoft, teams)
(african, american)
(black, african)
(covid, 19)
(american, community)
(community, microsoft)
(services, revenue)
(mutes, mic)
(using, platform)
(voices, black)
(platform, amplify)
(time, using)
(amplify, voices)
(new, features)
(covid19, pandemic)
(cloud, services)
(around, world)
(microsoft, time)
(learn, microsoft)
(business, highlights)
0
5
10
15
20
25
30
Number of Occurences

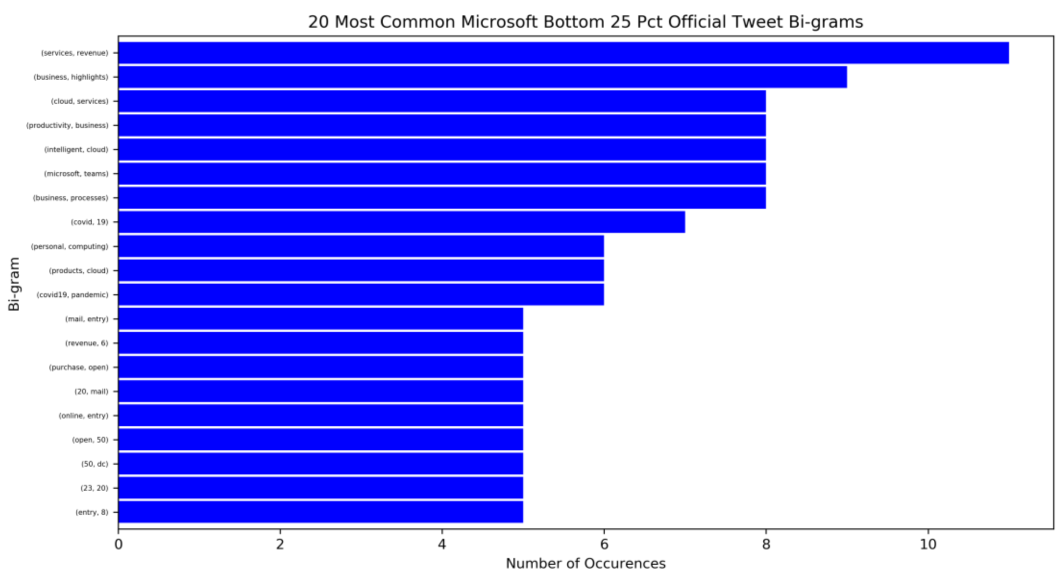

20 Most Common Microsoft Official Tweet Tri-grams
(black, african, american)
(african, american, community)
(american, community, microsoft)
(voices, black, african)
(using, platform, amplify)
(platform, amplify, voices)
(amplify, voices, black)
(time, using, platform)
(community, microsoft, time)
(microsoft, time, using)
(cloud, services, revenue)
(products, cloud, services)
(productivity, business, processes)
(meeting, mutes, mic)
(joins, meeting, mutes)
(microsoft, chief, technology)
(chief, technology, officer)
(advancing, social, justice)
(organizations, advancing, social)
(mont, saint, michel)
0
5
10
15
20
25
Number of Occurences

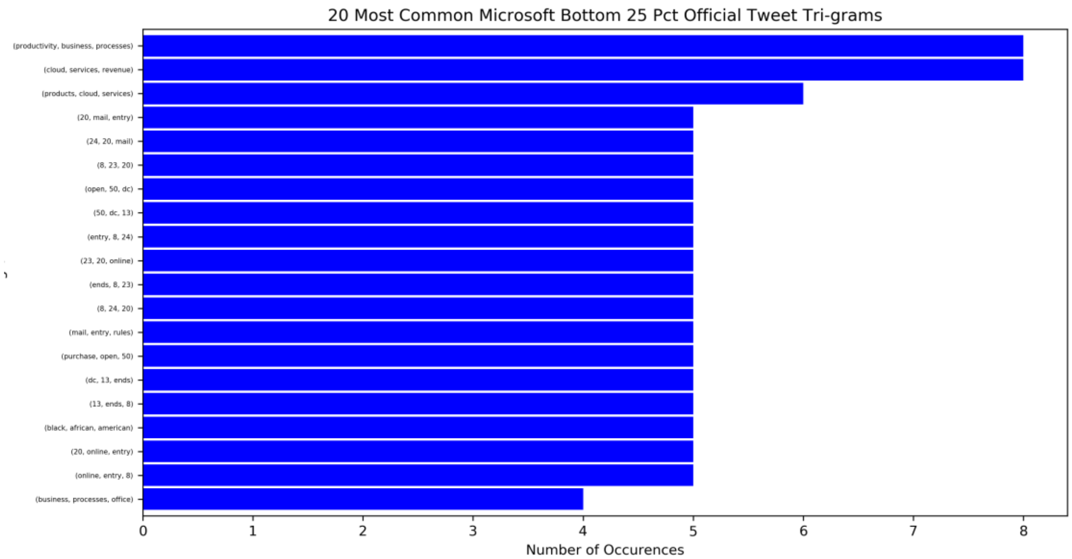

20 Most Common Microsoft Top 25 Pct Official Tweet Bi-grams
Bi-gram
(american, community)
(african, american)
(black, african)
(mutes, mic)
(community, microsoft)
(using, platform)
(voices, black)
(amplify, voices)
(platform, amplify)
(time, using)
(unmutes, mic)
(joins, meeting)
(meeting, mutes)
(wonder, woman)
(sword, shield)
(microsoft, teams)
(series, x)
(xbox, series)
(looks, like)
(social, justice)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Number of Occurences


20 Most Common Microsoft Top 25 Pct Official Tweet Tri-grams
Tri-gram
(african, american, community)
(black, african, american)
(american, community, microsoft)
(voices, black, african)
(amplify, voices, black)
(time, using, platform)
(platform, amplify, voices)
(using, platform, amplify)
(joins, meeting, mutes)
(meeting, mutes, mic)
(woman, missing, sword)
(organizations, advancing, social)
(missing, sword, shield)
(wonder, woman, missing)
(advancing, social, justice)
(xbox, series, x)
(microsoft, flight, simulator)
(american, community, shining)
(theyre, typing, theyre)
(typing, theyre, typing)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Number of Occurences

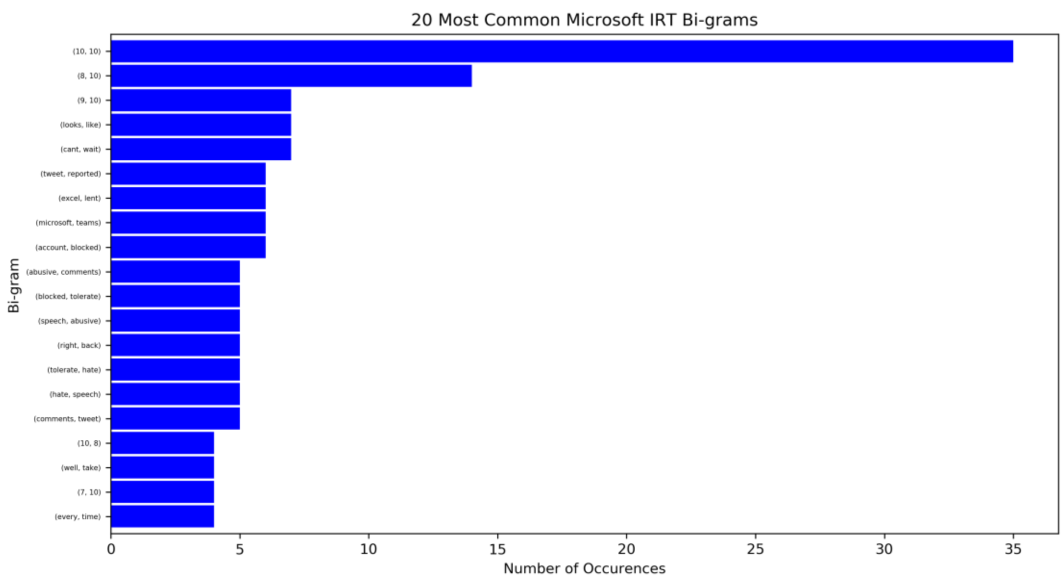

20 Most Common Microsoft Median 50 Pct Official Tweet Bi-grams
Bi-gram
(microsoft, teams)
(covid, 19)
(around, world)
(get, know)
(new, features)
(quantum, computing)
(learn, new)
(project, silica)
(learn, microsoft)
(take, look)
(new, microsoft)
(covid19, pandemic)
(read, via)
(stay, connected)
(episode, quantumimpact)
(microsoft, ceo)
(new, ways)
(using, ai)
(services, revenue)
(digital, transformation)
0
2
4
6
8
10
12
14
16
Number of Occurences


20 Most Common Microsoft Median 50 Pct Official Tweet Tri-grams
(mont, saint, michel)
(artist, residence, jenny)
(chief, technology, officer)
(meet, project, silica)
(people, around, world)
(click, download, template)
(residence, jenny, sabin)
(microsoft, chief, technology)
(today, proud, announce)
(template, reply, back)
(color, number, experience)
(take, look, inside)
(reply, back, create)
(download, template, reply)
(could, talk, plants)
(different, kinds, customers)
(programmers, ada, lovelace)
(quantum, computing, helping)
(new, project, preserving)
(ai, driven, project)
Tri-gram
0
1
2
3
4
5
Number of Occurences

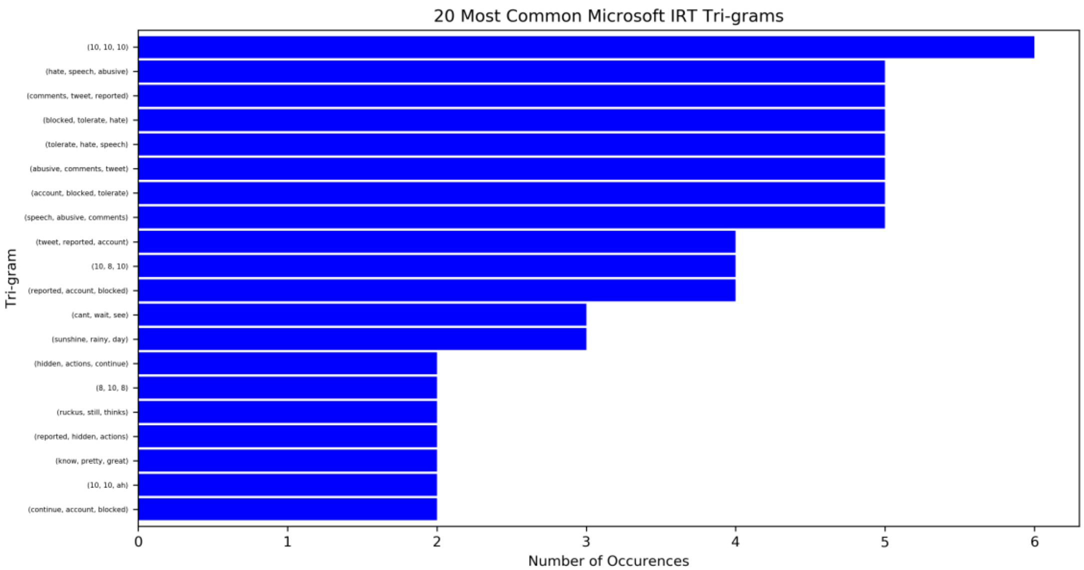

20 Most Common Microsoft Bottom 25 Pct Official Tweet Bi-grams
(services, revenue)
(business, highlights)
(cloud, services)
(productivity, business)
(intelligent, cloud)
(microsoft, teams)
(business, processes)
(covid, 19)
(personal, computing)
(products, cloud)
(covid19, pandemic)
(mail, entry)
(revenue, 6)
(purchase, open)
(20, mail)
(online, entry)
(open, 50)
(50, dc)
(23, 20)
(entry, 8)
Bi-gram
0
2
4
6
8
10
Number of Occurences

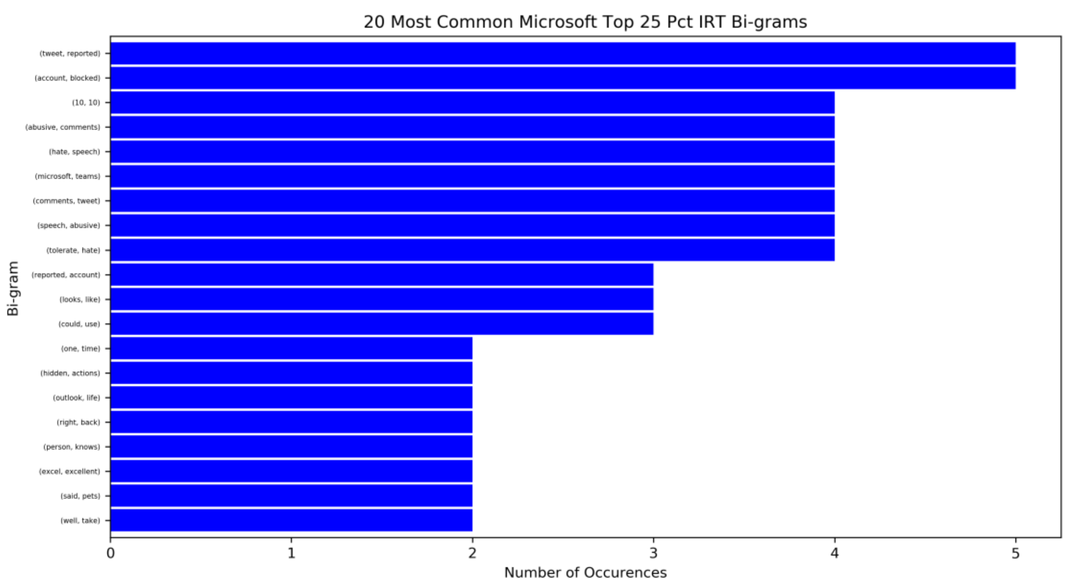


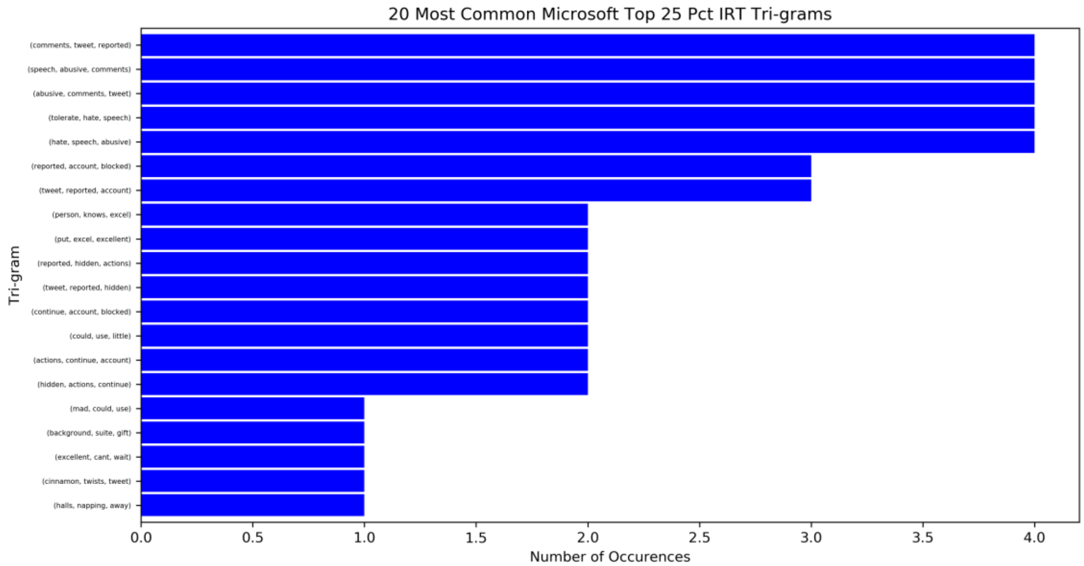
20 Most Common Microsoft Bottom 25 Pct Official Tweet Tri-grams
(productivity, business, processes)
(cloud, services, revenue)
(products, cloud, services)
(20, mail, entry)
(24, 20, mail)
(8, 23, 20)
(open, 50, dc)
(50, dc, 13)
(entry, 8, 24)
(23, 20, online)
(ends, 8, 23)
(8, 24, 20)
(mail, entry, rules)
(purchase, open, 50)
(dc, 13, ends)
(13, ends, 8)
(black, african, american)
(20, online, entry)
(online, entry, 8)
(business, processes, office)
0 1 2 3 4 5 6 7 8
Number of Occurences



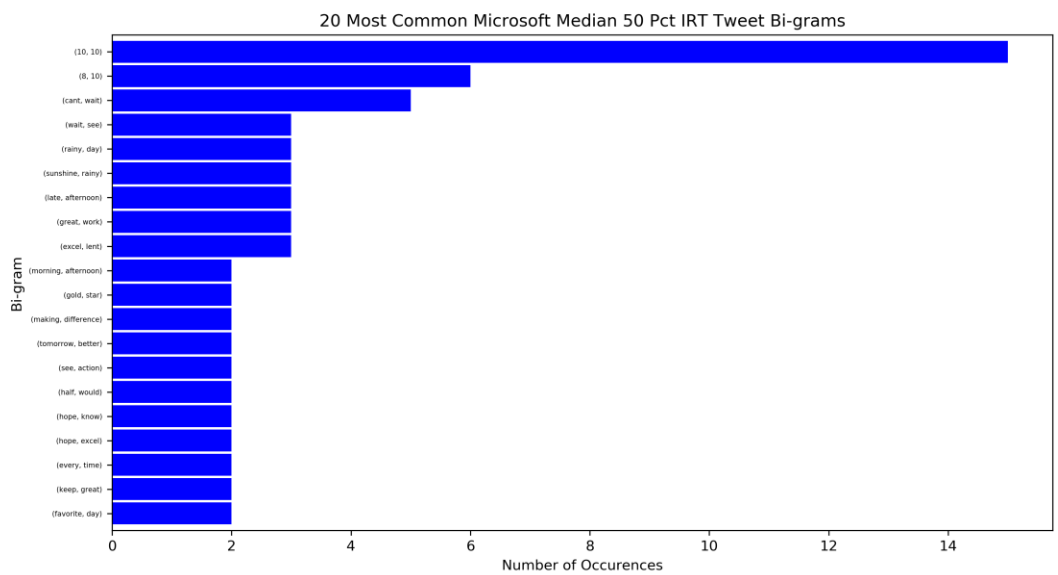
20 Most Common Microsoft IRT Bi-grams
Bi-gram
(10, 10)
(8, 10)
(9, 10)
(looks, like)
(cant, wait)
(tweet, reported)
(excel, lent)
(microsoft, teams)
(account, blocked)
(abusive, comments)
(blocked, tolerate)
(speech, abusive)
(right, back)
(tolerate, hate)
(hate, speech)
(comments, tweet)
(10, 8)
(well, take)
(7, 10)
(every, time)
0 5 10 15 20 25 30 35
Number of Occurences


20 Most Common Microsoft IRT Tri-grams
Tri-gram
Number of Occurences
(10, 10, 10)
(hate, speech, abusive)
(comments, tweet, reported)
(blocked, tolerate, hate)
(tolerate, hate, speech)
(abusive, comments, tweet)
(account, blocked, tolerate)
(speech, abusive, comments)
(tweet, reported, account)
(10, 8, 10)
(reported, account, blocked)
(cant, wait, see)
(sunshine, rainy, day)
(hidden, actions, continue)
(8, 10, 8)
(ruckus, still, thinks)
(reported, hidden, actions)
(know, pretty, great)
(10, 10, ah)
(continue, account, blocked)
0
1
2
3
4
5
6

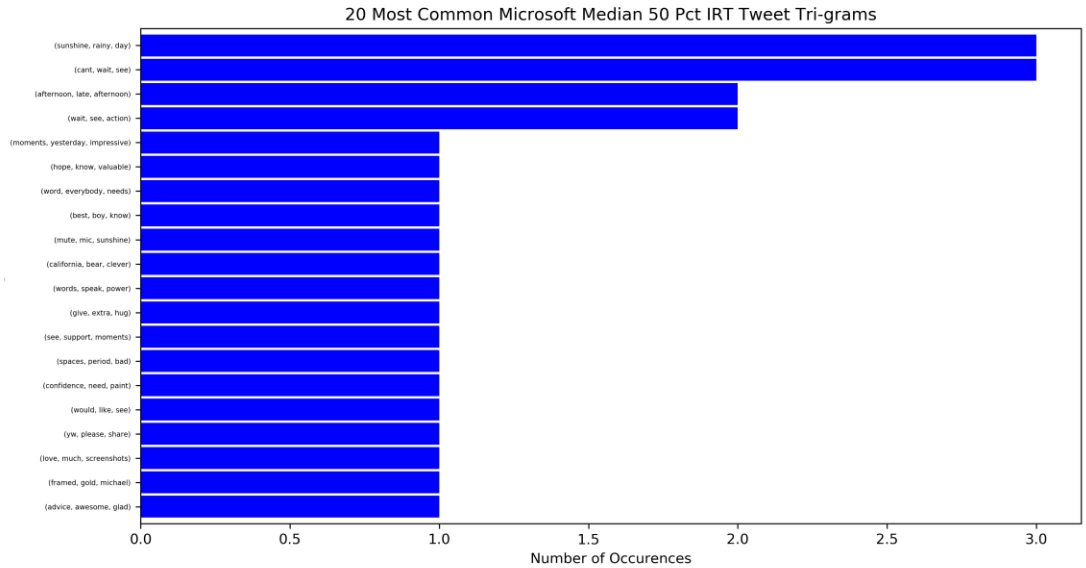

20 Most Common Microsoft Top 25 Pct IRT Bi-grams
Bi-gram
Number of Occurences
(tweet, reported)
(account, blocked)
(10, 10)
(abusive, comments)
(hate, speech)
(microsoft, teams)
(comments, tweet)
(speech, abusive)
(tolerate, hate)
(reported, account)
(looks, like)
(could, use)
(one, time)
(hidden, actions)
(outlook, life)
(right, back)
(person, knows)
(excel, excellent)
(said, pets)
(well, take)
0
1
2
3
4
5

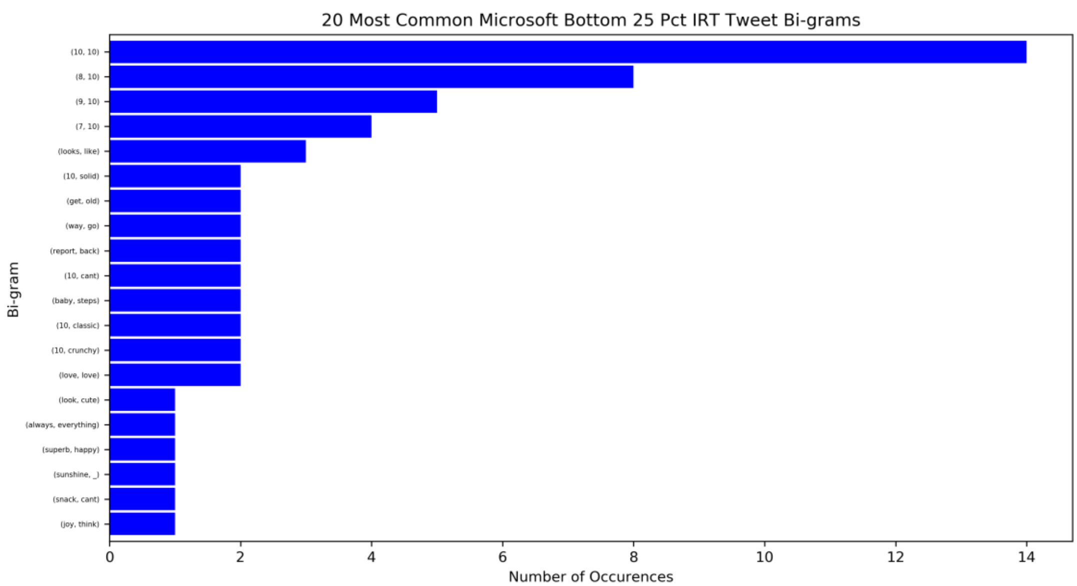


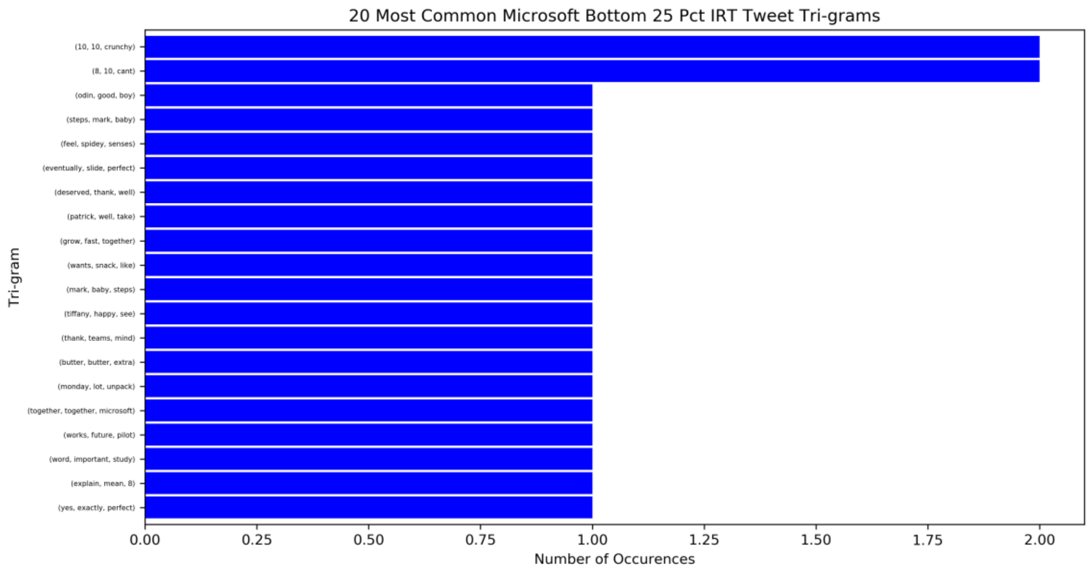
20 Most Common Microsoft Top 25 Pct IRT Tri-grams
Tri-gram
Number of Occurences
(comments, tweet, reported)
(speech, abusive, comments)
(abusive, comments, tweet)
(tolerate, hate, speech)
(hate, speech, abusive)
(reported, account, blocked)
(tweet, reported, account)
(person, knows, excel)
(put, excel, excellent)
(reported, hidden, actions)
(tweet, reported, hidden)
(continue, account, blocked)
(could, use, little)
(actions, continue, account)
(hidden, actions, continue)
(mad, could, use)
(background, suite, gift)
(excellent, cant, wait)
(cinnamon, twists, tweet)
(halls, napping, away)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0



20 Most Common Microsoft Median 50 Pct IRT Tweet Bi-grams
Bi-gram
Number of Occurences
(10, 10)
(8, 10)
(cant, wait)
(wait, see)
(rainy, day)
(sunshine, rainy)
(late, afternoon)
(great, work)
(excel, lent)
(morning, afternoon)
(gold, star)
(making, difference)
(tomorrow, better)
(see, action)
(half, would)
(hope, know)
(hope, excel)
(every, time)
(keep, great)
(favorite, day)
0
2
4
6
8
10
12
14


20 Most Common Microsoft Median 50 Pct IRT Tweet Tri-grams
(sunshine, rainy, day)
(cant, wait, see)
(afternoon, late, afternoon)
(wait, see, action)
(moments, yesterday, impressive)
(hope, know, valuable)
(word, everybody, needs)
(best, boy, know)
(mute, mic, sunshine)
(california, bear, clever)
(words, speak, power)
(give, extra, hug)
(see, support, moments)
(spaces, period, bad)
(confidence, need, paint)
(would, like, see)
(yw, please, share)
(love, much, screenshots)
(framed, gold, michael)
(advice, awesome, glad)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Number of Occurences


20 Most Common Microsoft Bottom 25 Pct IRT Tweet Bi-grams
Bi-gram
(10, 10)
(8, 10)
(9, 10)
(7, 10)
(looks, like)
(10, solid)
(get, old)
(way, go)
(report, back)
(10, cant)
(baby, steps)
(10, classic)
(10, crunchy)
(love, love)
(look, cute)
(always, everything)
(superb, happy)
(sunshine, _)
(snack, cant)
(joy, think)
0
2
4
6
8
10
12
14
Number of Occurences



Microsoft Top Words

All Tweets	Official Tweets	Top 25% Official	Median 50% Official	Bottom 25% Official	All IRT Tweets	Top 25% IRT	Median 50% IRT	Bottom 25% IRT
microsoft, 292	microsoft, 282	microsoft, 76	learn, 156	learn, 71	‘10’, 96	thank, 11	‘10’, 43	‘10’, 45
learn, 245	learn, 244	mic, 31	microsoft, 141	microsoft, 65	thank, 32	could, 9	thank, 14	great, 10
new, 149	new, 146	meeting, 24	new, 89	msignite, 48	love, 29	time, 9	day, 14	hi, 9
msignite, 132	msignite, 132	time, 23	ai, 80	revenue, 45	like, 29	like, 9	love, 14	like, 8
‘10’, 120	ai, 118	black, 23	world, 70	new, 37	great, 26	one, 8	hey, 14	‘8’, 8
ai, 118	world, 97	community, 23	msignite, 67	cloud, 34	right, 24	great, 8	like, 12	thank, 7
one, 107	help, 92	one, 22	people, 53	ai, 32	one, 23	love, 8	well, 11	back, 6
world, 102	one, 84	african, 20	help, 51	help, 28	hi, 22	right, 8	good, 11	love, 6
help, 95	people, 81	american, 20	technology, 49	technology, 27	well, 21	‘10’, 8	know, 11	right, 6
people, 89	technology, 76	new, 20	one, 47	billion, 24	day, 21	excel, 7	hi, 10	looks, 5
technology, 77	microsoftteams, 71	today, 19	microsoftteams, 46	business, 24	see, 21	see, 7	right, 10	go, 5
work, 77	using, 66	using, 18	work, 39	world, 21	hey, 20	teams, 7	one, 10	yes, 5
microsoftteams, 72	work, 64	mutes, 17	msbuild, 36	work, 20	good, 19	know, 7	see, 10	please, 5
time, 70	revenue, 60	learn, 17	covid19, 33	innovation, 17	know, 19	tweet, 6	thanks, 8	well, 5
like, 70	msbuild, 60	msignite, 16	using, 32	tech, 16	time, 19	work, 5	got, 8	one, 5

using, 68	cloud, 55	platform, 16	future, 31	people, 16	could, 15	well, 5	great, 8	‘9’, 5
today, 62	today, 54	voices, 15	data, 30	microsoftteams, 16	excel, 15	take, 5	work, 8	pizza, 5
msbuild, 62	time, 51	year, 15	discover, 29	using, 16	teams, 15	always, 5	true, 7	good, 4
revenue, 60	year, 50	amplify, 14	azure, 28	covid19, 15	yes, 15	two, 5	would, 7	hey, 4
teams, 59	covid19, 49	msbuild, 14	learning, 28	teams, 15	cant, 14	microsoft, 5	afternoon, 7	got, 4

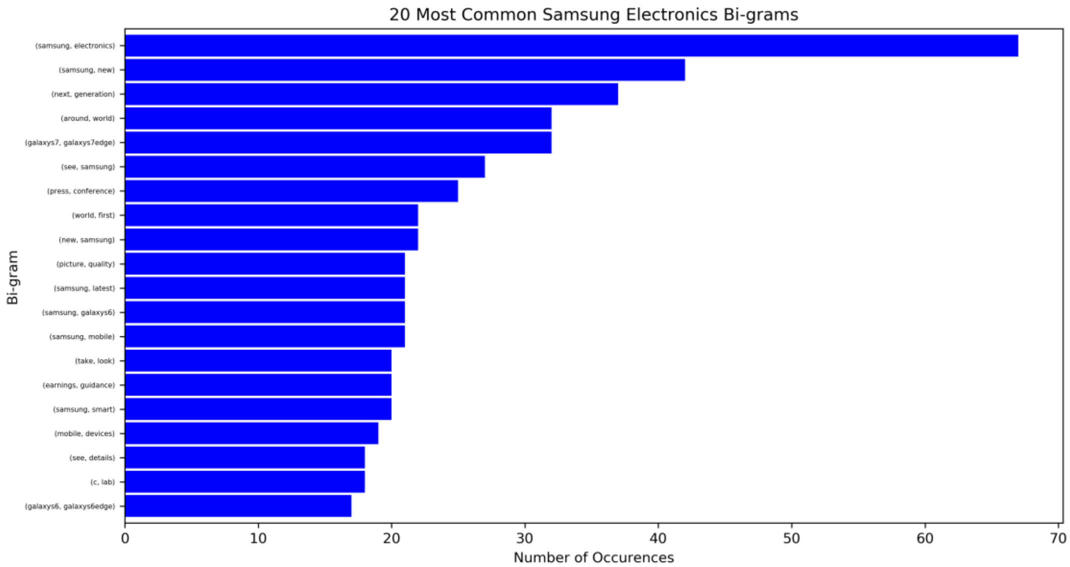
-Following are the words uniquely common to each of the lists above:

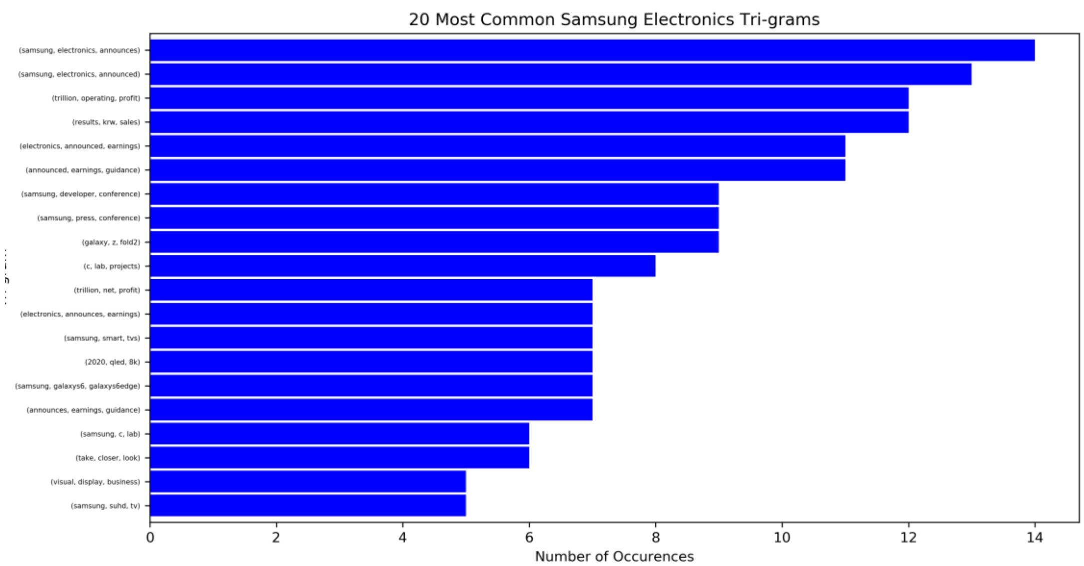
All Tweets	Official Tweets	Top 25% Official	Median 50% Official	Bottom 25% Official	All IRT Tweets	Top 25% IRT	Median 50% IRT	Bottom 25% IRT
		mic	future	billion	cant	tweet	thanks	8
		meeting	data	business		take	true	back
		black	discover	innovation		always	would	looks
		community	azure	tech		two	afternoon	go
		african	learning					please
		american						9
		mutes						pizza
		platform						
		voices						
		amplify						

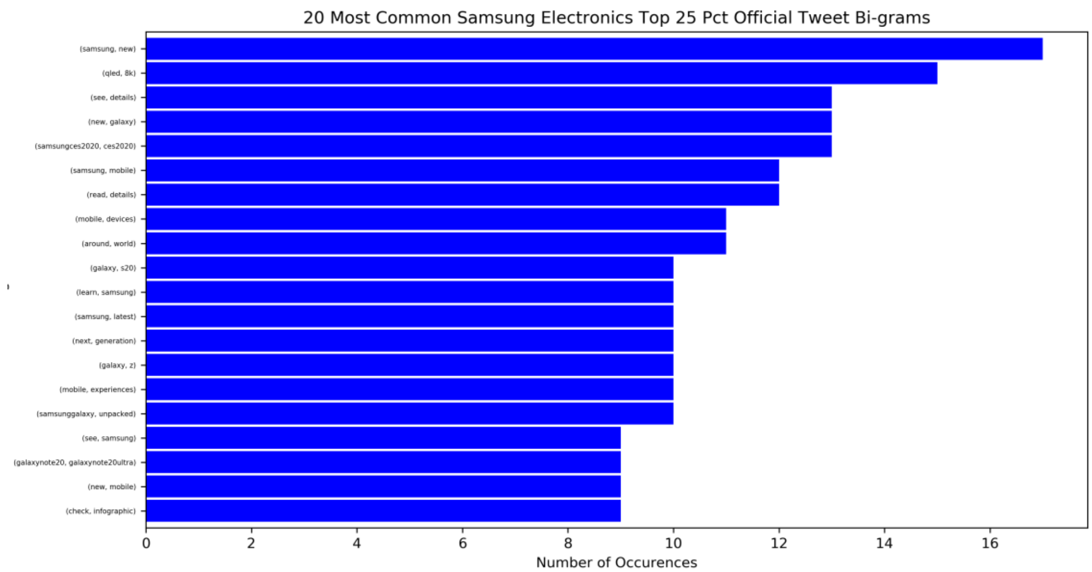
Samsung

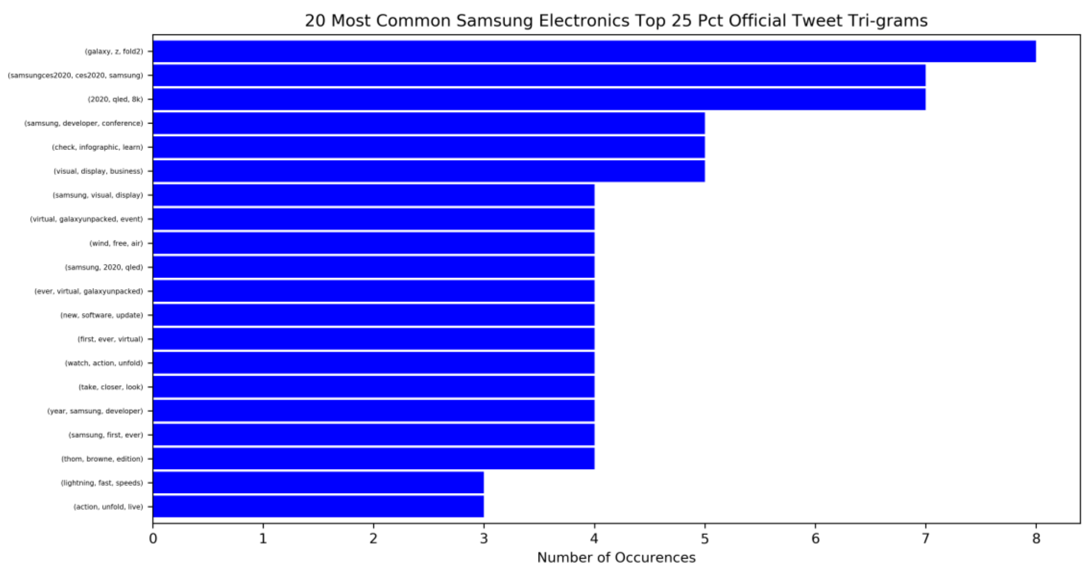
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	2914	15.30	44575	6379
Top 25% OT:	725	21.2469	15404	3111
Median 50% OT:	1450	14.16483	20539	4252
Bottom 25% OT:	725	11.67172	8462	2575

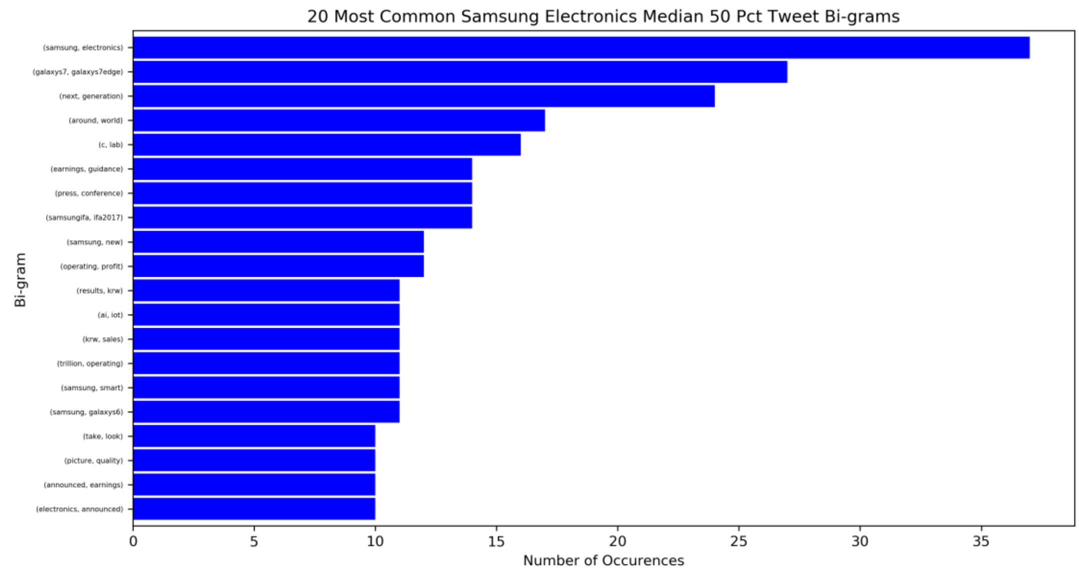
-Samsung data definitely seems to display some form of relationship between tweet popularity and average tweet length. The differences between categories are relatively large, and there are ample tweets contained within each category.

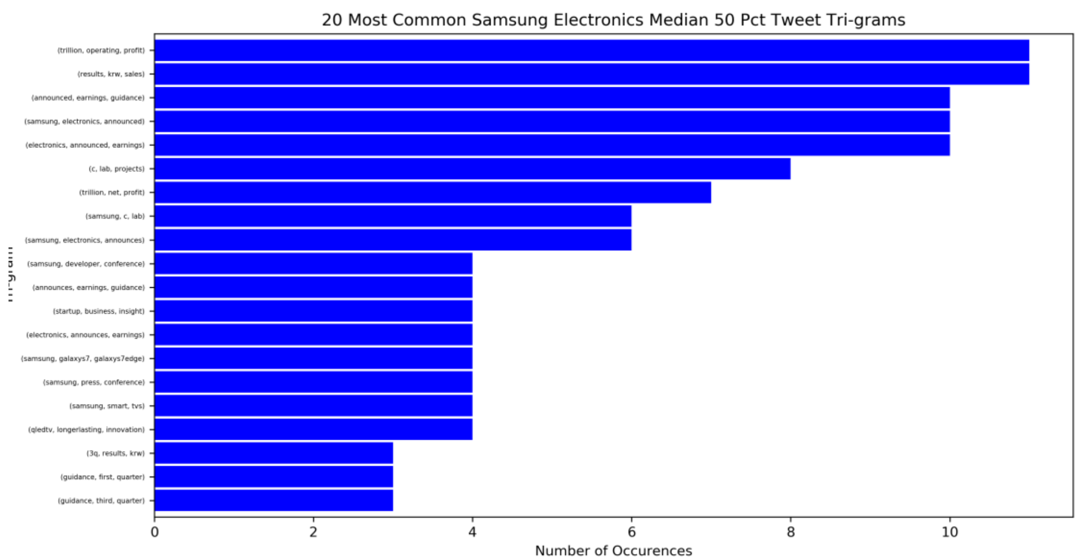

20 Most Common Samsung Electronics Bi-grams
Bi-gram
(samsung, electronics)
(samsung, new)
(next, generation)
(around, world)
(galaxys7, galaxys7edge)
(see, samsung)
(press, conference)
(world, first)
(new, samsung)
(picture, quality)
(samsung, latest)
(samsung, galaxys6)
(samsung, mobile)
(take, look)
(earnings, guidance)
(samsung, smart)
(mobile, devices)
(see, details)
(c, lab)
(galaxys6, galaxys6edge)
0
10
20
30
40
50
60
70
Number of Occurences

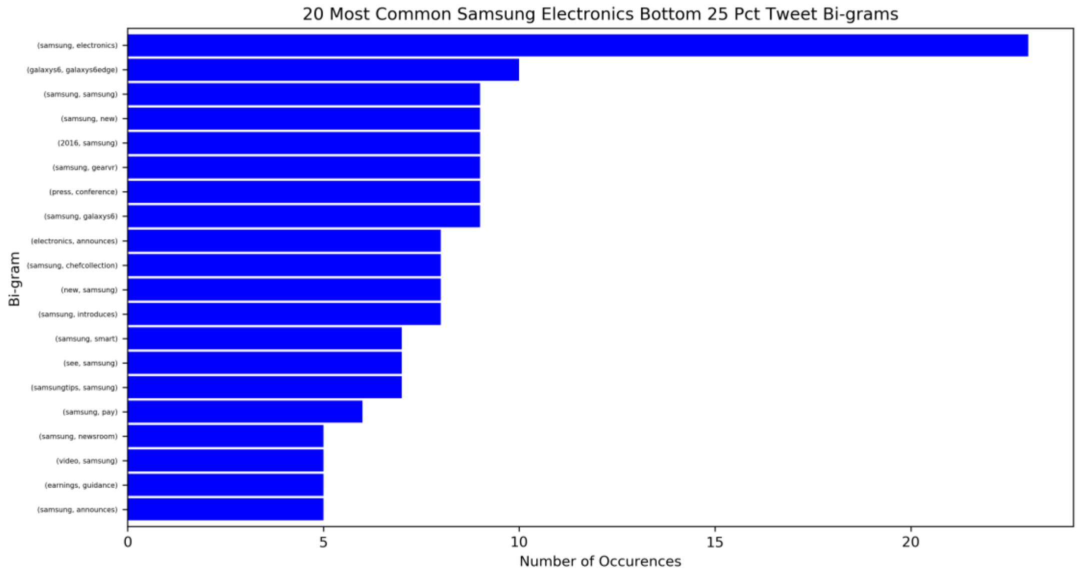

20 Most Common Samsung Electronics Tri-grams
(samsung, electronics, announces)
(samsung, electronics, announced)
(trillion, operating, profit)
(results, krw, sales)
(electronics, announced, earnings)
(announced, earnings, guidance)
(samsung, developer, conference)
(samsung, press, conference)
(galaxy, z, fold2)
(c, lab, projects)
(trillion, net, profit)
(electronics, announces, earnings)
(samsung, smart, tvs)
(2020, qled, 8k)
(samsung, galaxys6, galaxys6edge)
(announces, earnings, guidance)
(samsung, c, lab)
(take, closer, look)
(visual, display, business)
(samsung, suhd, tv)
0
2
4
6
8
10
12
14
Number of Occurences

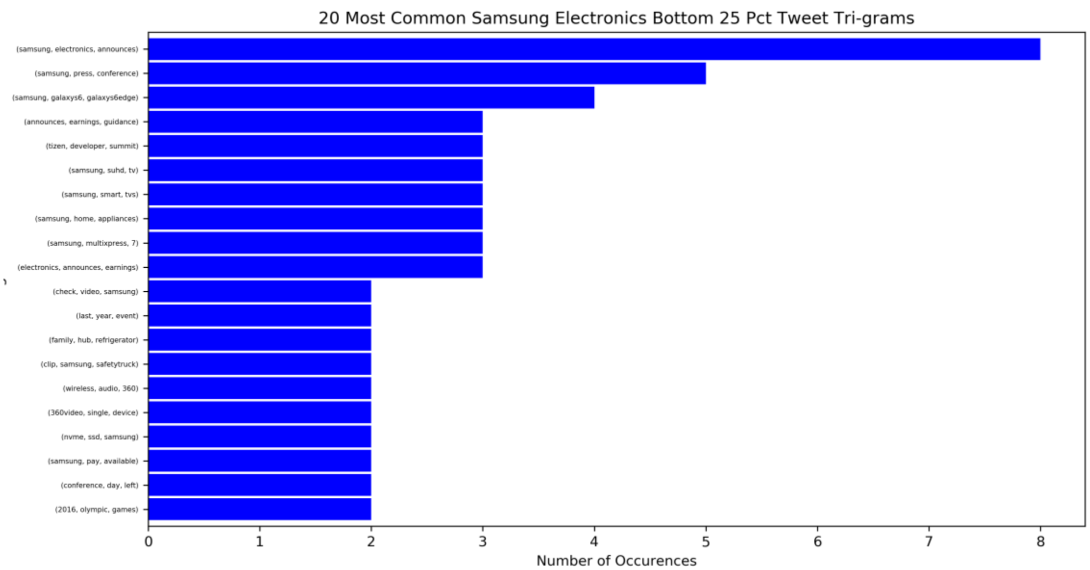

20 Most Common Samsung Electronics Top 25 Pct Official Tweet Bi-grams
(samsung, new)
(qled, 8k)
(see, details)
(new, galaxy)
(samsungces2020, ces2020)
(samsung, mobile)
(read, details)
(mobile, devices)
(around, world)
(galaxy, s20)
(learn, samsung)
(samsung, latest)
(next, generation)
(galaxy, z)
(mobile, experiences)
(samsunggalaxy, unpacked)
(see, samsung)
(galaxynote20, galaxynote20ultra)
(new, mobile)
(check, infographic)
0
2
4
6
8
10
12
14
16
Number of Occurences


20 Most Common Samsung Electronics Top 25 Pct Official Tweet Tri-grams
(galaxy, z, fold2)
(samsungces2020, ces2020, samsung)
(2020, qled, 8k)
(samsung, developer, conference)
(check, infographic, learn)
(visual, display, business)
(samsung, visual, display)
(virtual, galaxyunpacked, event)
(wind, free, air)
(samsung, 2020, qled)
(ever, virtual, galaxyunpacked)
(new, software, update)
(first, ever, virtual)
(watch, action, unfold)
(take, closer, look)
(year, samsung, developer)
(samsung, first, ever)
(thom, browne, edition)
(lightning, fast, speeds)
(action, unfold, live)
0
1
2
3
4
5
6
7
8
Number of Occurences


20 Most Common Samsung Electronics Median 50 Pct Tweet Bi-grams
(samsung, electronics)
(galaxys7, galaxys7edge)
(next, generation)
(around, world)
(c, lab)
(earnings, guidance)
(press, conference)
(samsungifa, ifa2017)
(samsung, new)
(operating, profit)
(results, krw)
(ai, iot)
(krw, sales)
(trillion, operating)
(samsung, smart)
(samsung, galaxys6)
(take, look)
(picture, quality)
(announced, earnings)
(electronics, announced)
Bi-gram
0
5
10
15
20
25
30
35
Number of Occurences


20 Most Common Samsung Electronics Median 50 Pct Tweet Tri-grams
(trillion, operating, profit)
(results, krw, sales)
(announced, earnings, guidance)
(samsung, electronics, announced)
(electronics, announced, earnings)
(c, lab, projects)
(trillion, net, profit)
(samsung, c, lab)
(samsung, electronics, announces)
(samsung, developer, conference)
(announces, earnings, guidance)
(startup, business, insight)
(electronics, announces, earnings)
(samsung, galaxys7, galaxys7edge)
(samsung, press, conference)
(samsung, smart, tvs)
(qledtv, longerlasting, innovation)
(3q, results, krw)
(guidance, first, quarter)
(guidance, third, quarter)
0
2
4
6
8
10
Number of Occurences





Samsung Top Words

All Tweets	Top 25%	Median 50%	Bottom 25%
samsung, 1334	samsung, 332	samsung, 547	samsung, 455
new, 506	new, 205	new, 248	new, 52
see, 171	galaxy, 87	technology, 78	tv, 35
5g, 161	see, 85	ai, 75	galaxys6, 32
mobile, 157	features, 80	5g, 75	experience, 28
galaxy, 151	mobile, 77	experience, 63	iot, 27
technology, 144	5g, 74	tv, 63	check, 27
experience, 139	samsunggalaxy, 61	world, 60	mobile, 27
ai, 133	ai, 58	see, 59	see, 27
world, 131	users, 56	design, 58	technology, 25
design, 130	learn, 52	first, 57	smart, 23
features, 128	design, 50	mobile, 53	home, 23
first, 123	world, 48	galaxy, 50	‘2016’, 23
tv, 121	experience, 48	solutions, 50	first, 23

check, 114	devices, 48	iot, 49	world, 23
next, 90	check, 45	next, 47	electronics, 23
latest, 89	camera, 44	features, 45	design, 22
home, 86	first, 43	smart, 45	app, 19
iot, 86	smartphone, 43	check, 42	future, 18
learn, 85	read, 40	galaxys7, 41	video, 18

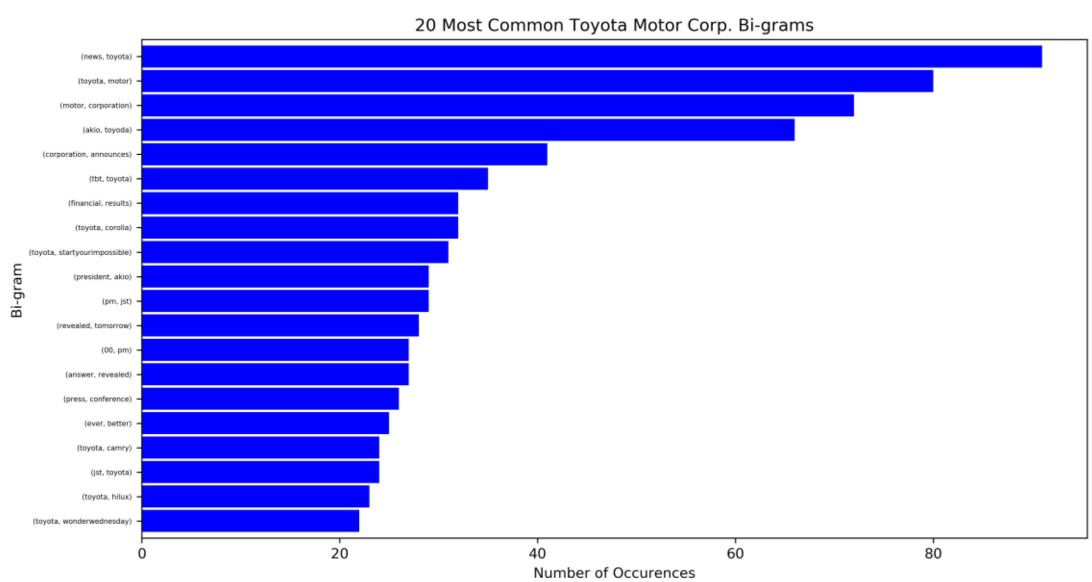
-Following are the word uniquely common to each of the lists above:

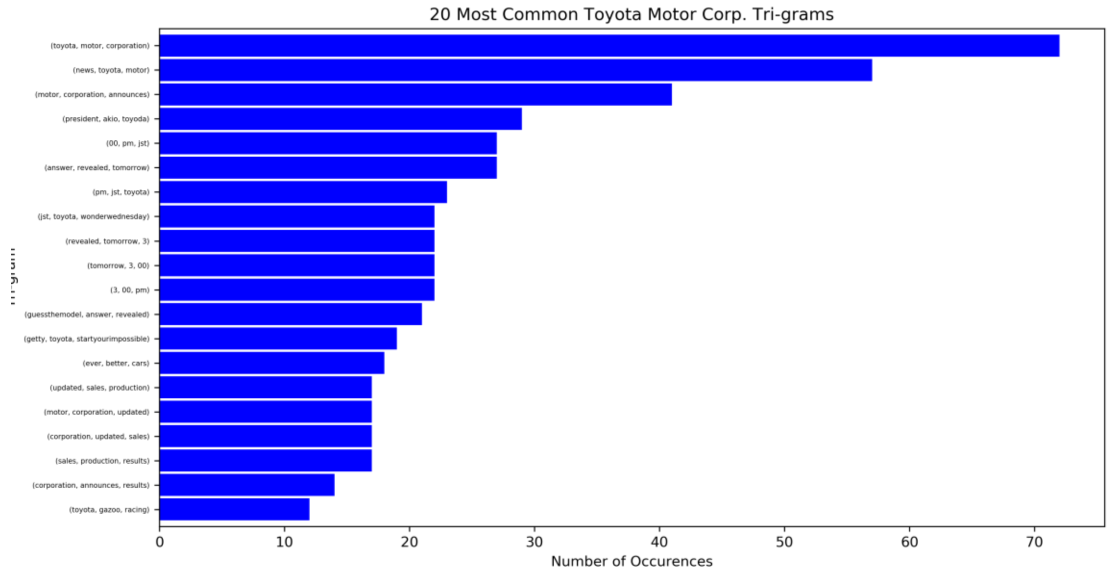
All Tweets	Top 25%	Median 50%	Bottom 25%
latest	samsunggalaxy	solutions	galaxys6
	users	galaxys7	2016
	devices		electronics
	camera		app
	smartphone		future
	read		video

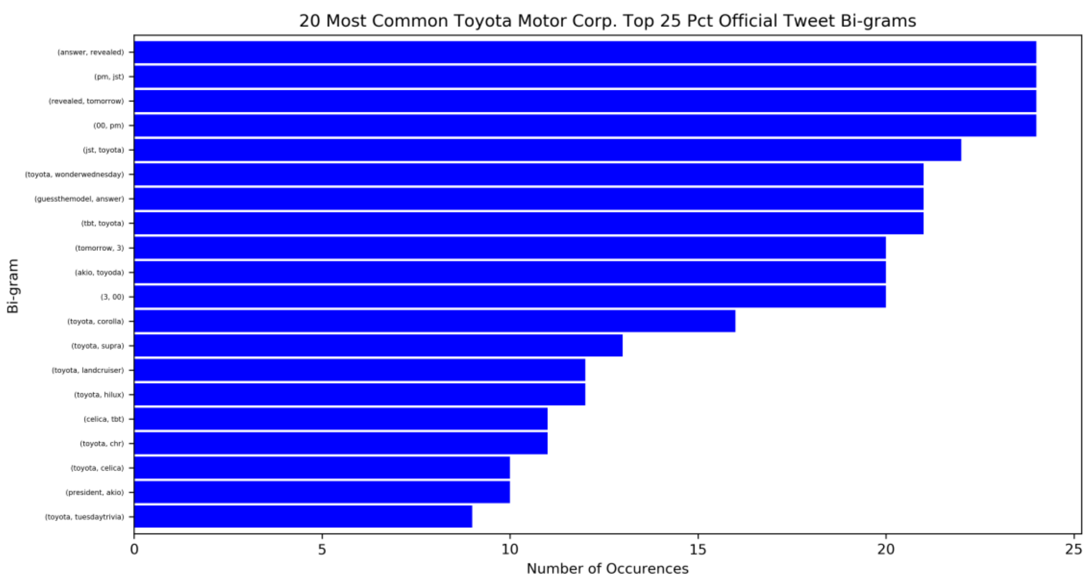
Toyota

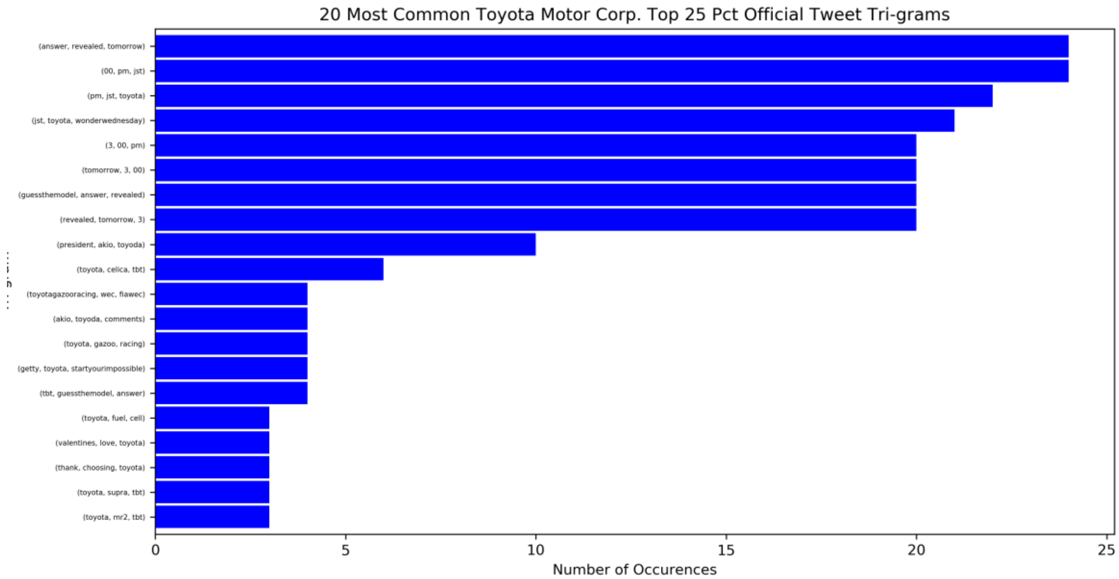
Tweet Category	No. Tweets	Avg. Word Count	Total Words	Unique Words
All Tweets:	2545	17.08	43466	5781
Top 25% OT:	635	14.14646	8983	2172
Median 50% OT:	1269	17.57053	22297	4070
Bottom 25% OT:	635	19.0189	12077	2850

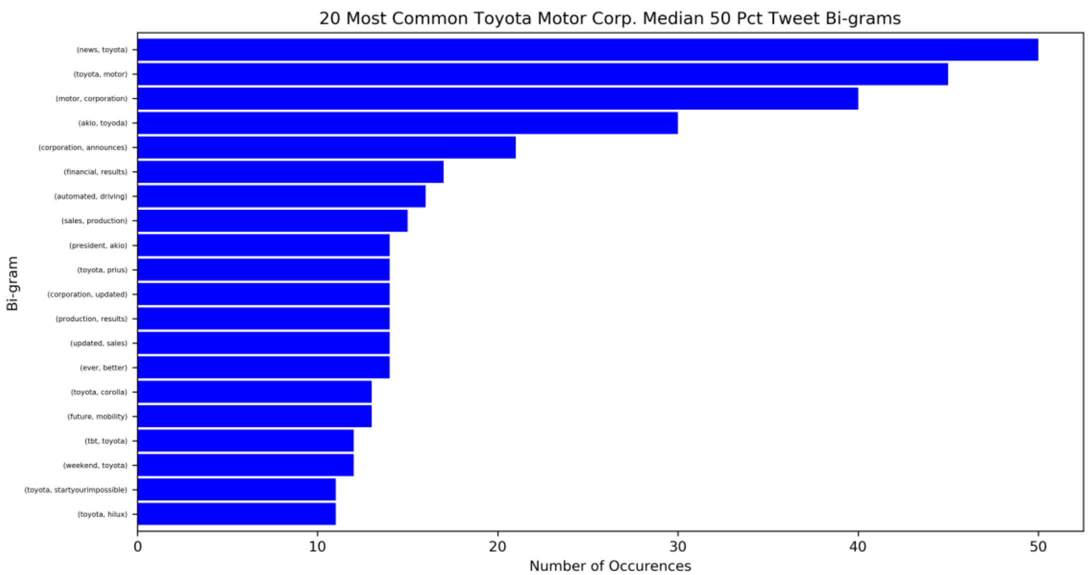
-It seems that Toyota data is displaying about the opposite behavior as seen in Samsung data.

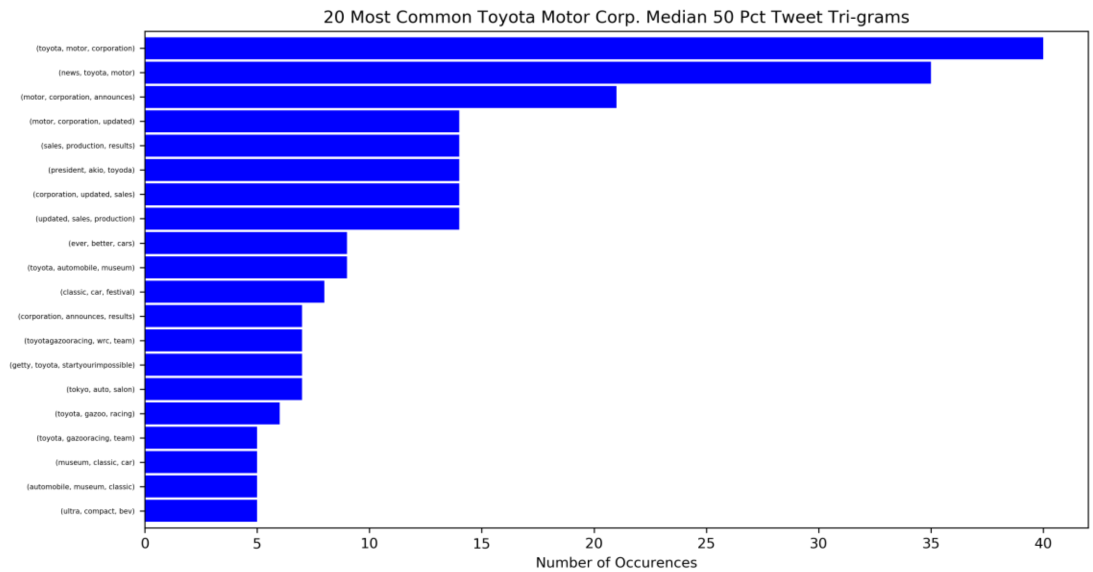


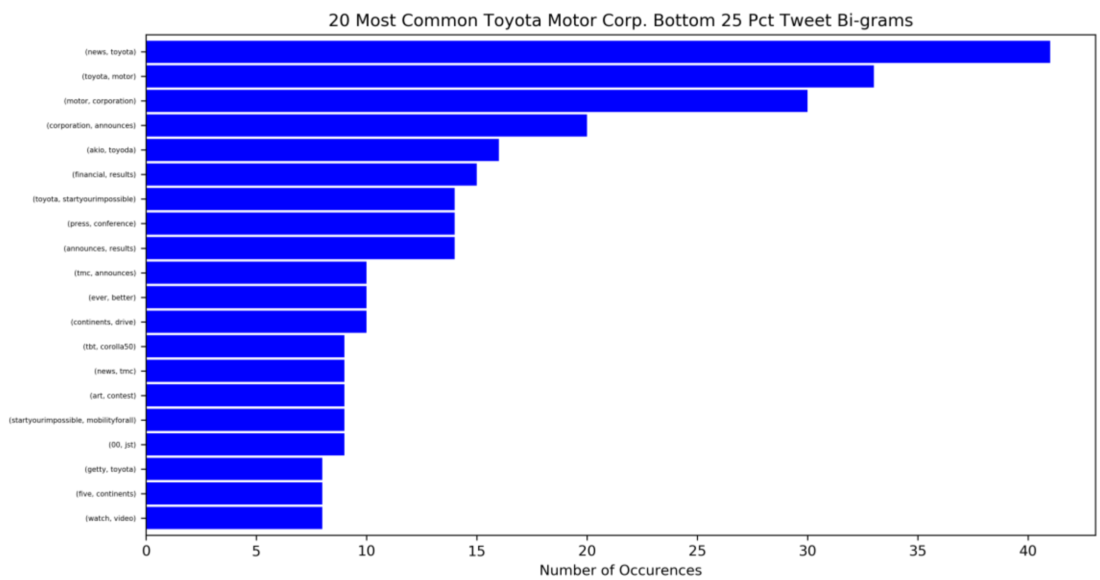

20 Most Common Toyota Motor Corp. Tri-grams
(toyota, motor, corporation)
(news, toyota, motor)
(motor, corporation, announces)
(president, akio, toyoda)
(00, pm, jst)
(answer, revealed, tomorrow)
(pm, jst, toyota)
(jst, toyota, wonderwednesday)
(revealed, tomorrow, 3)
(tomorrow, 3, 00)
(3, 00, pm)
(guessthemodel, answer, revealed)
(getty, toyota, startyourimpossible)
(ever, better, cars)
(updated, sales, production)
(motor, corporation, updated)
(corporation, updated, sales)
(sales, production, results)
(corporation, announces, results)
(toyota, gazoo, racing)
0
10
20
30
40
50
60
70
Number of Occurences

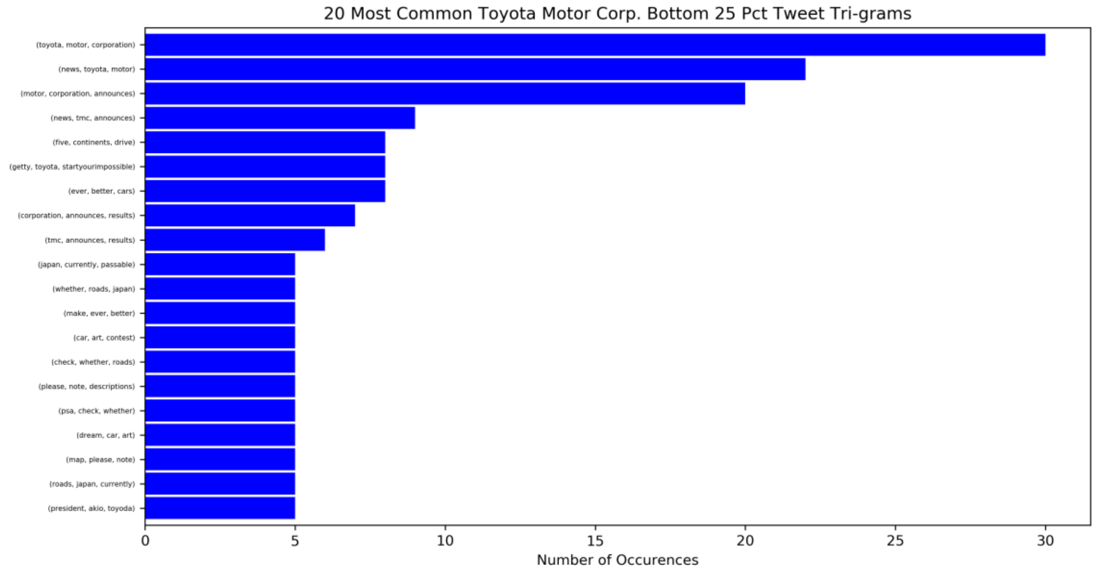

20 Most Common Toyota Motor Corp. Top 25 Pct Official Tweet Bi-grams
Bi-gram
(answer, revealed)
(pm, jst)
(revealed, tomorrow)
(00, pm)
(jst, toyota)
(toyota, wonderwednesday)
(guessthemodel, answer)
(tbt, toyota)
(tomorrow, 3)
(akio, toyoda)
(3, 00)
(toyota, corolla)
(toyota, supra)
(toyota, landcruiser)
(toyota, hilux)
(celica, tbt)
(toyota, chr)
(toyota, celica)
(president, akio)
(toyota, tuesdaytrivia)
0
5
10
15
20
25
Number of Occurences


20 Most Common Toyota Motor Corp. Top 25 Pct Official Tweet Tri-grams
(answer, revealed, tomorrow)
(00, pm, jst)
(pm, jst, toyota)
(jst, toyota, wonderwednesday)
(3, 00, pm)
(tomorrow, 3, 00)
(guessthemodel, answer, revealed)
(revealed, tomorrow, 3)
(president, akio, toyoda)
(toyota, celica, tbt)
(toyotagazooracing, wec, flawec)
(akio, toyoda, comments)
(toyota, gazoo, racing)
(getty, toyota, startyourimpossible)
(tbt, guessthemodel, answer)
(toyota, fuel, cell)
(valentines, love, toyota)
(thank, choosing, toyota)
(toyota, supra, tbt)
(toyota, mr2, tbt)
0
5
10
15
20
25
Number of Occurences


20 Most Common Toyota Motor Corp. Median 50 Pct Tweet Bi-grams
Bi-gram
(news, toyota)
(toyota, motor)
(motor, corporation)
(akio, toyoda)
(corporation, announces)
(financial, results)
(automated, driving)
(sales, production)
(president, akio)
(toyota, prius)
(corporation, updated)
(production, results)
(updated, sales)
(ever, better)
(toyota, corolla)
(future, mobility)
(tbt, toyota)
(weekend, toyota)
(toyota, startyourimpossible)
(toyota, hilux)
0
10
20
30
40
50
Number of Occurences


20 Most Common Toyota Motor Corp. Median 50 Pct Tweet Tri-grams
(toyota, motor, corporation)
(news, toyota, motor)
(motor, corporation, announces)
(motor, corporation, updated)
(sales, production, results)
(president, akio, toyoda)
(corporation, updated, sales)
(updated, sales, production)
(ever, better, cars)
(toyota, automobile, museum)
(classic, car, festival)
(corporation, announces, results)
(toyotagazooracing, wrc, team)
(getty, toyota, startyourimpossible)
(tokyo, auto, salon)
(toyota, gazoo, racing)
(toyota, gazooracing, team)
(museum, classic, car)
(automobile, museum, classic)
(ultra, compact, bev)
0
5
10
15
20
25
30
35
40
Number of Occurences


20 Most Common Toyota Motor Corp. Bottom 25 Pct Tweet Bi-grams
(news, toyota)
(toyota, motor)
(motor, corporation)
(corporation, announces)
(akio, toyoda)
(financial, results)
(toyota, startyourimpossible)
(press, conference)
(announces, results)
(tmc, announces)
(ever, better)
(continents, drive)
(tbt, corolla50)
(news, tmc)
(art, contest)
(startyourimpossible, mobilityforall)
(00, jst)
(getty, toyota)
(five, continents)
(watch, video)
0
5
10
15
20
25
30
35
40
Number of Occurences



Toyota Top Words

All Tweets	Top 25%	Median 50%	Bottom 25%
toyota, 1453	toyota, 421	toyota, 725	toyota, 307
new, 264	tbt, 113	new, 124	new, 73
tbt, 197	new, 66	car, 78	news, 60
car, 149	supra, 50	mobility, 74	japan, 44
japan, 138	toyota86, 42	tbt, 69	car, 42
news, 127	corolla, 39	learn, 66	motor, 41
mobility, 121	guessthemodel, 39	news, 64	mobility, 38
one, 113	japan, 38	weekend, 62	world, 38
motor, 107	landcruiser, 34	prius, 58	results, 37
team, 105	back, 32	one, 58	startyourimpossible, 36
learn, 105	car, 29	team, 57	watch, 36
corolla, 102	wonderwednesday, 28	motor, 57	cars, 34
world, 100	camry, 26	japan, 56	announces, 34
future, 94	celica, 25	hilux, 55	corporation, 31
see, 90	answer, 25	future, 53	one, 31
cars, 89	one, 24	driving, 53	team, 30
yariswrc, 87	revealed, 24	hydrogen, 53	‘2017’, 28
hydrogen, 86	tomorrow, 24	ready, 52	see, 26
weekend, 86	‘00’, 24	toyotatimes, 50	learn, 26
hilux, 85	pm, 24	yariswrc, 49	corolla, 25

-Following are the words uniquely common to each list above:

All Tweets	Top 25%	Median 50%	Bottom 25%
	supra	prius	results
	toyota86	driving	startyourimpossible
	guessthemodel	ready	watch
	landcruiser	toyotatimes	announces
	back		corporation
	wonderwednesday		2017

	camry		
	celica		
	answer		
	revealed		
	tomorrow		
	00		
	pm		

Section 4: Hashtag Descriptive Statistics

Section 4 of the appendix contains descriptive statistics for each company and tweet category, based on the number of hashtags contained within tweets.

Table 4.1: Amazon Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
OT with 0 #	138	443.9493	1199.14	71.35507	189.2509
OT with 1 #	98	241.8367	248.2469	44.62245	55.17068
OT with 2 #	10	210.8	145.4845	40.6	32.98215
OT with 3 #	1	148	NA	13	NA
IRT with 0 #	2292	3.239092	30.18182	0.61911	3.471055
IRT with 1 #	599	4.342237	69.95776	0.388982	4.417654
IRT with 2 #	1	2	NA	0	NA
IRT with 3 #	0	NA	NA	NA	NA

Table 4.2: BMW Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
OT with 0 #	210	429.9762	604.5451	44.55238	71.41582
OT with 1 #	145	1312.807	1013.025	140.9172	140.9349
OT with 2 #	341	1322.716	708.089	134.2258	94.77254
OT with 3 #	83	1107.831	577.1646	129.5542	78.09036
OT with 4 #	12	992.5833	504.0277	113.9167	70.70741
OT with 5#	3	2069.667	1083.41	275.3333	179.7229
OT with 6 #	2	1013	203.6468	120	22.62742
IRT with 0 #	2277	2.254282	12.31774	0.140975	0.746917
IRT with 1 #	3	1	1	0	0
IRT with 2 #	0	NA	NA	NA	NA
IRT with 3 #	0	NA	NA	NA	NA
IRT with 4 #	0	NA	NA	NA	NA
IRT with 5 #	0	NA	NA	NA	NA
IRT with 6 #	0	NA	NA	NA	NA

Table 4.3: Coca Cola Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
IRT with 0 #	3168	21.40436	1089.081	3.169508	172.6583
IRT with 1 #	10	8.1	16.07932	0.2	0.421637

Table 4.4: Disney Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
OT w/ 0 #	956	3763.977	5131.915	724.1768	1187.424
OT w/ 1 #	1237	3423.001	10830.68	904.1431	4591.565
OT w/ 2 #	456	1782.039	2255.127	312.9254	596.155
OT w/ 3 #	52	1866.904	1686.916	373.8269	437.957
OT w/ 4 #	11	2013	781.4035	411.6364	255.61
OT w/ 5 #	6	1279.667	1304.823	183.8333	187.5819
OT w/ 6 #	1	3410	NA	562	NA
OT w/ **14** #	1	3634	NA	1190	NA

Table 4.5: Google Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
IRT with 0 #	3023	0.383063	6.721126	0.033741	0.596859
IRT with 1 #	6	3306.5	8092.87	1056.333	2585.518

Table 4.6: McDonald's Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
IRT with 0 #	3193	7.268713	63.17498	0.130911	1.000827
IRT with 1 #	1	21	NA	1	NA

Table 4.7: Mercedes Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
OT w/ 0#	59	1317.373	1680.861	156.2034	220.4997
OT w/ 1#	50	958.78	928.5006	93	102.9222
OT w/ 2#	63	1252.46	1135.041	131.4603	151.1467
OT w/ 3#	134	1195.381	1518.123	145.8955	442.8014
OT w/ 4#	103	1043.524	619.6401	100.9903	69.66439
OT w/ 5#	51	889.8039	469.0228	87.05882	55.64114
OT w/ 6#	27	1103.926	459.9071	105.9259	55.35753
OT w/ 7#	12	1034.833	684.3577	122.0833	74.29115
OT w/ 8#	2	979.5	736.0982	104.5	101.1163
OT w/ 9#	2	855	619.4255	66.5	23.33452
OT w/ 10#	2	698	472.3473	69.5	41.7193
IRT w/ 0#	2312	1.477941	21.21397	0.211505	3.729354
IRT w/ 1#	203	1.305419	2.893372	0.20197	0.539156
IRT w/ 2#	10	0.9	0.875595	0.5	0.707107
IRT w/ 3#	0	NA	NA	NA	NA
IRT w/ 4#	0	NA	NA	NA	NA
IRT w/ 5#	1	11	NA	0	NA

Table 4.8: Microsoft Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
OT w/ 0#	601	1120.476	3379.96	164.5391	509.8319
OT w/ 1#	554	420.6498	1327.141	86.86462	213.1071
OT w/ 2#	103	418.699	1502.551	74.62136	106.1949
OT w/ 3#	3	2390	3861.608	2640	4473.045
IRT w/ 0#	1352	54.86021	450.5394	2.018491	17.41186
IRT w/ 1#	9	4.888889	7.406829	0.555556	0.881917

Table 4.9: Samsung Hashtag Descriptive Statistics

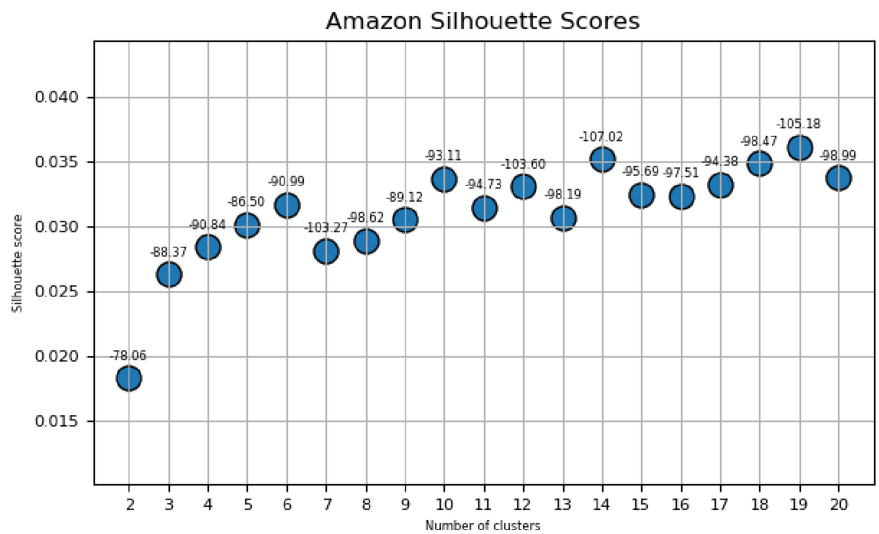
Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
OT w/ 0#	533	175.0263	1044.013	20.53471	73.00413
OT w/ 1#	973	441.4234	1531.53	42.57657	139.5485
OT w/ 2#	800	486.2188	1793.823	47.5125	221.1941
OT w/ 3#	381	513.8976	1954.719	58.77953	339.5973
OT w/ 4#	143	722.8322	1902.103	56.35664	126.0619
OT w/ 5#	47	1130.681	4091.66	198.234	1018.703
OT w/ 6#	16	927.25	2078.925	67.375	127.2472
OT w/ 7#	3	4790.333	4153.885	522.6667	556.7785
OT w/ 8#	3	3440.667	3803.068	286.6667	187.6761
OT w/ 9#	1	25	NA	9	NA

Table 4.10: Toyota Hashtag Descriptive Statistics

Data	No. Observations	Average Likes	SD: Likes	Average Retweets	SD: Retweets
OT w/ 0#	169	41.31953	38.18429	12.33136	11.76239
OT w/ 1#	661	75.69138	65.03611	18.14977	15.97404
OT w/ 2#	726	134.9545	1027.769	31.33058	248.8814
OT w/ 3#	533	159.3752	861.0892	31.29831	144.7173
OT w/ 4#	250	95.06	63.38257	20	16.92269
OT w/ 5#	105	93.15238	75.45832	20.45714	24.48666
OT w/ 6#	59	102.7797	72.94024	23.71186	23.1301
OT w/ 7#	22	88.77273	50.27916	20.18182	14.5623
OT w/ 8#	9	125	81.24808	29.22222	19.60725
OT w/ 9#	3	77.33333	53.00314	20	18.08314
OT w/ 10#	2	109.5	20.5061	21	1.414214

<u>Section 5:</u> Elbow Plots and Silhouette Scores

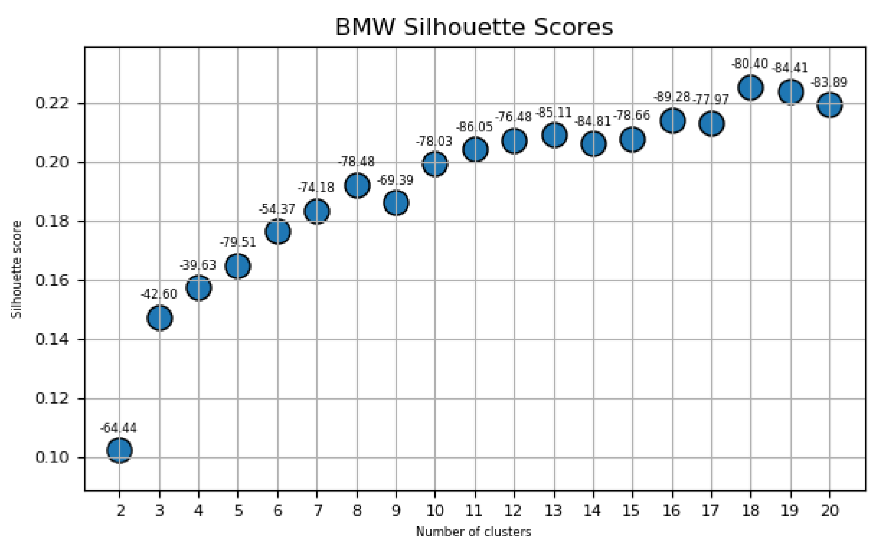
Note: These visualizations were produced before making the decision to totally separate OT and IRT tweet categories, rather than have tweet category be a binary variable. Thus, the images produced here don’t align with the final BTM topic assignments. However, they do show that, for 8 of the 10 original topic assignments, the optimal estimated number of topics was within 1 position of a 15% or greater change in silhouette scores.



-Of the images produced above, Amazon was the only company which matched the original plots. Thus, the same analysis still applies as earlier. To make things easier, I'll copy/paste that analysis below:

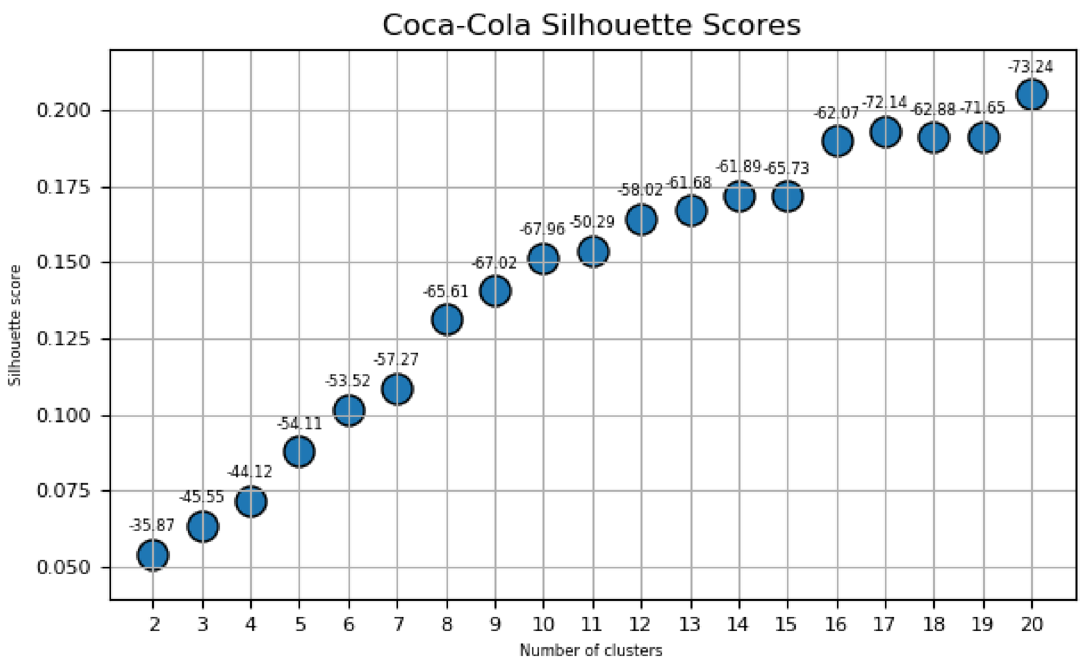
"For Amazon, it seems that k values two places behind a drop in silhouette scores produce better average topic coherence than k values immediately prior to a drop, as well as k values which constitute a drop in silhouette scores. For example, k = 5 performs better than k = 6 (1 value prior to a drop) and k = 7 (the k value which displays a drop in silhouette scores). The same may be said for k's equal to 9, 10, 11, k's equal to 11, 12, 13, and k's equal to 18, 19, 20. However, this isn't the case for k's equal to 13, 14, 15, but those values are located at the end of a sequence of rises and drops in silhouette scores."

-I'm writing this portion after having examined figures for all data. One thing I noticed was that, at least for the final few companies, their optimal k was located in the vicinity of a 15%+ increase or decrease in silhouette scores. This does not quite hold true for Amazon data.



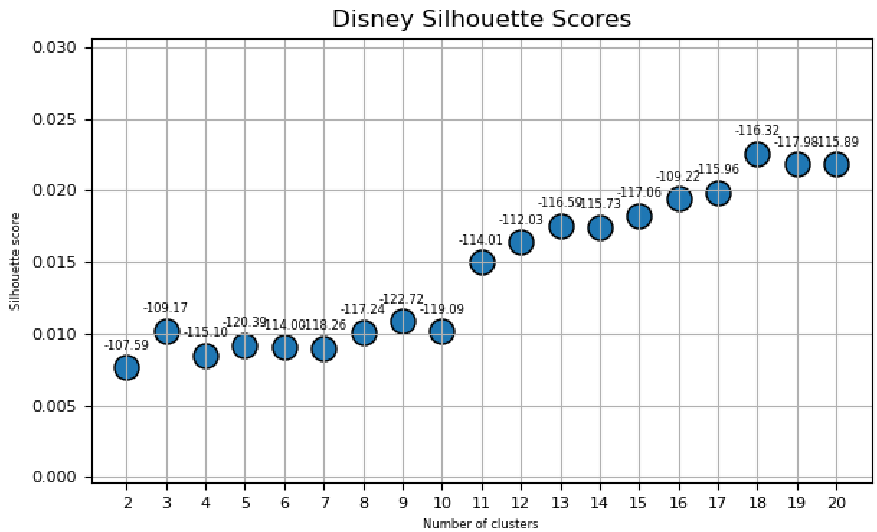
-For BMW, the Amazon pattern doesn't seem to continue. The best performing number of clusters is located at k = 4, well before a drop in silhouette scores. Also, the Amazon pattern only seems to be true for some of the drops (best performing out of the trio is located two positions prior to drop). I'm beginning to become doubtful that silhouette scores are a viable way of reducing the number of topics requiring estimation in order to locate an 'optimal' k value.

-Going back through this, it seems that BMW's optimal k is located within the vicinity of a significant change, as there's a 44.21% increase in silhouette scores from k = 2 to k = 3 topics.



-Coca Cola does not seem to adhere to the Amazon pattern either. The best estimated average topic coherence for Coca Cola lies at k = 4 topics, well before any drop in silhouette scores.

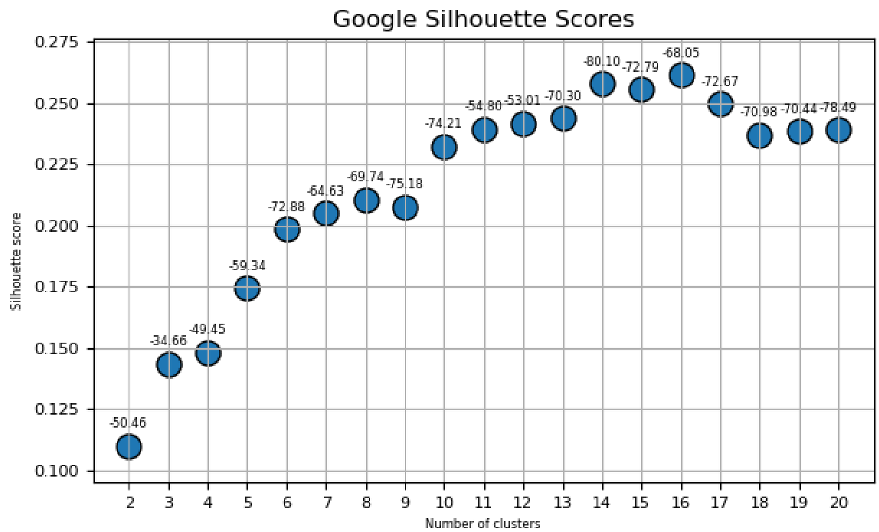
-However, going back through this, Coca Cola's best performing k value is located in the vicinity of a significant change in silhouette scores. There's a 17.18% increase from k = 2 to k = 3 topics, as well as a 22.68% increase from k = 4 to k = 5 topics.



-Disney data was handled a little differently than for others. Ultimately, I didn't consider k = 2 topics as an eligible value, based off personal opinion that there's no way any of these companies only discuss two topics throughout all tweets collected. The second lowest estimated average topic coherence lies at k = 3 topics, but I chose k = 16 topics for Disney data. This is because the differences in average topic coherence are minimal between the two (-109.17 vs. -109.22), and I believe the added difficulty in achieving (practically) the same score, with over 5x as many topics, is an indication that k = 16 topics is more correct for Disney.

-With that being said, Disney data seems to adhere to the Amazon pattern. The optimal k value is found two locations prior to a drop in silhouette scores. And, when a drop occurs, the best performing k value is always located two positions prior to the drop. Perhaps there are a few exceptions to that for minimal drops (like the drop between 5 and 6).

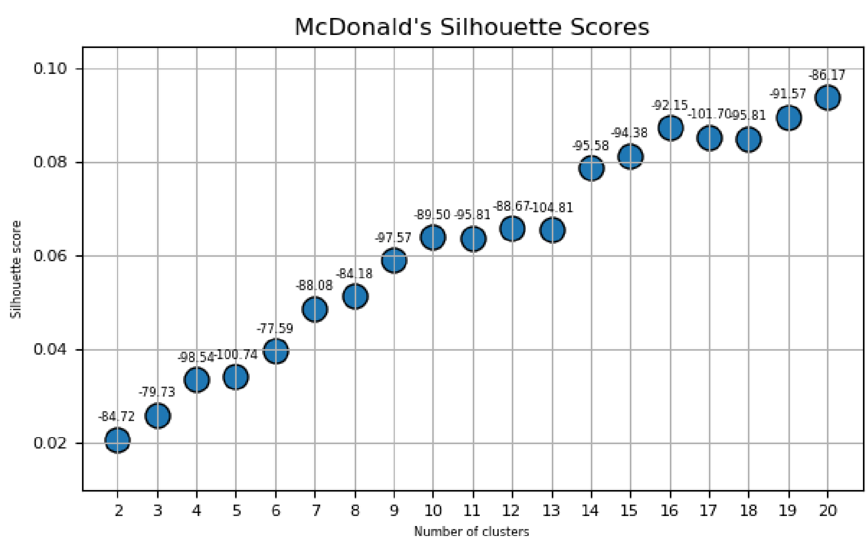
-Going back through here, Disney's optimal k value does not seem to be located within the vicinity of a significant change in silhouette scores.



-For Google, the best performing k value is located at k = 3 topics, well before any drop in silhouette scores. The Amazon pattern seems to hold true for cases in which there are drops in silhouette scores. However, I'm not sure of the usefulness in identifying that k values two positions prior to a drop are, generally, the best performing out of the trio, when even better performing k values may be found at random spots in the plot. For example, k = 13 outperforms k = 14 and 15, but k = 12 and 11 outperform all of them and aren't located near any drops.

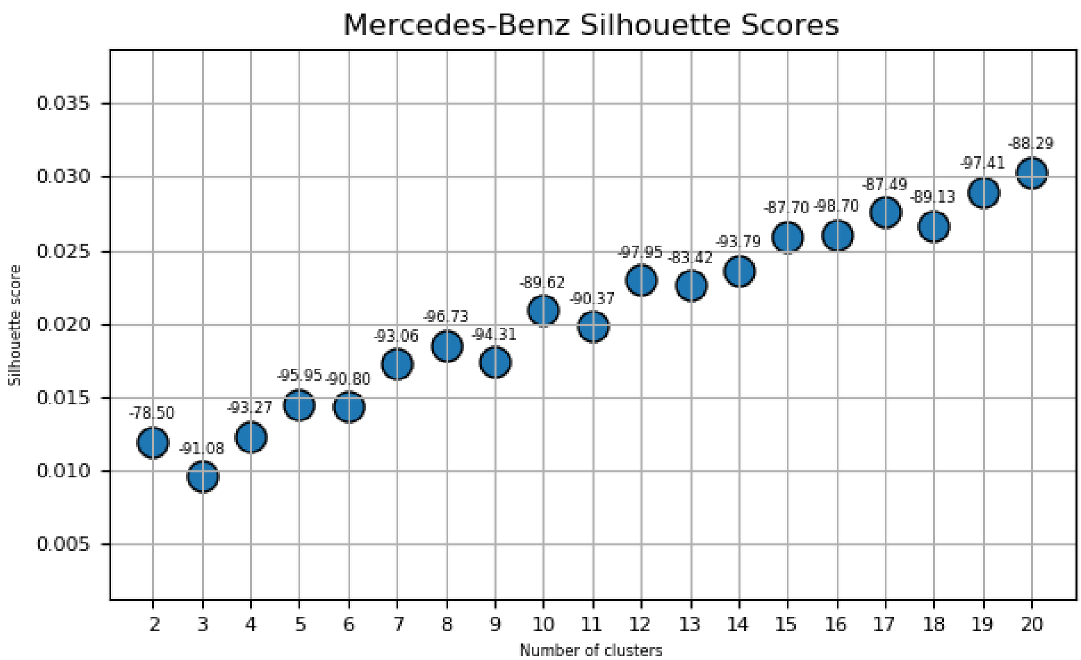
-I suppose it's possible that including low k values is muddying the waters. These probably perform better because it's easier to produce a lower average topic coherence with a lower number of topics. Perhaps a heuristic might be, when a significant change in silhouette scores occurs (either positive or negative), check the values located around the significant change.

-Going back through, Google's optimal k is located in the vicinity of a significant change in silhouette scores. There's a 30.47% increase in silhouette scores from k = 2 to k = 3 topics.



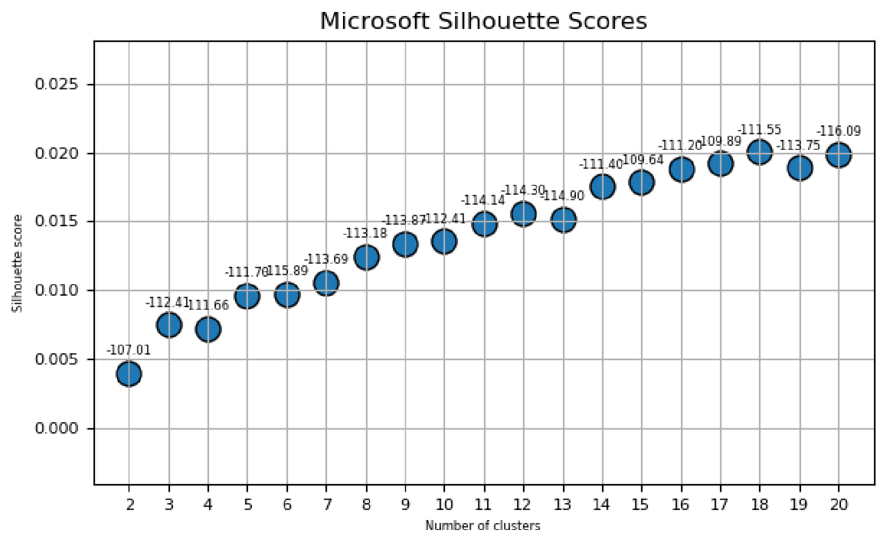
-McDonalds' best k value is located at k = 6 topics, well before any drop in silhouette scores. However, it seems that there's a significant increase in silhouette scores from k = 6 to k = 7 topics. Perhaps the 'heuristic' mentioned just above (testing values around any significant change in silhouette scores) would work for McDonalds data.

-Going back through, McDonalds' optimal k is located in the vicinity of two separate significant changes in silhouette scores.

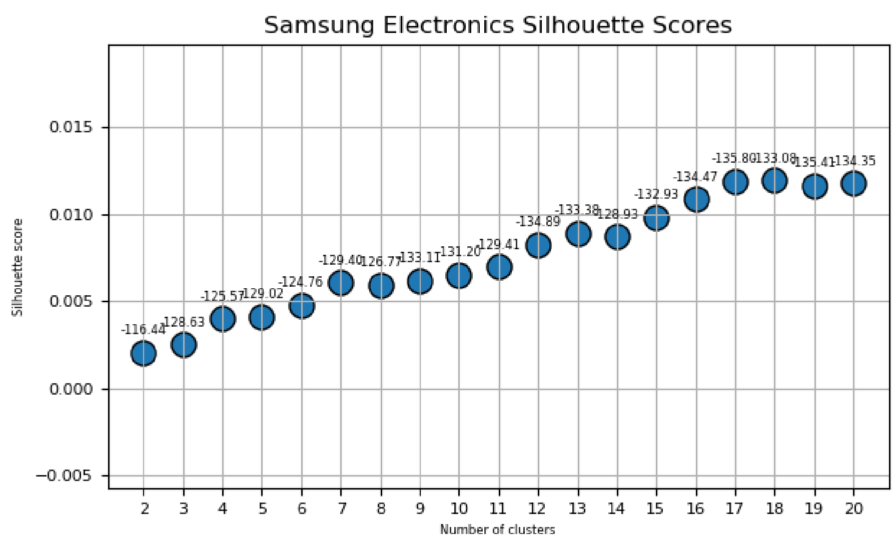


-Mercedes' best performing k value is located at k = 13 topics, which, if you define the change in silhouette scores from k = 11 to k = 12 as significant (hard to determine visually), is located in the vicinity of a significant change in silhouette scores.

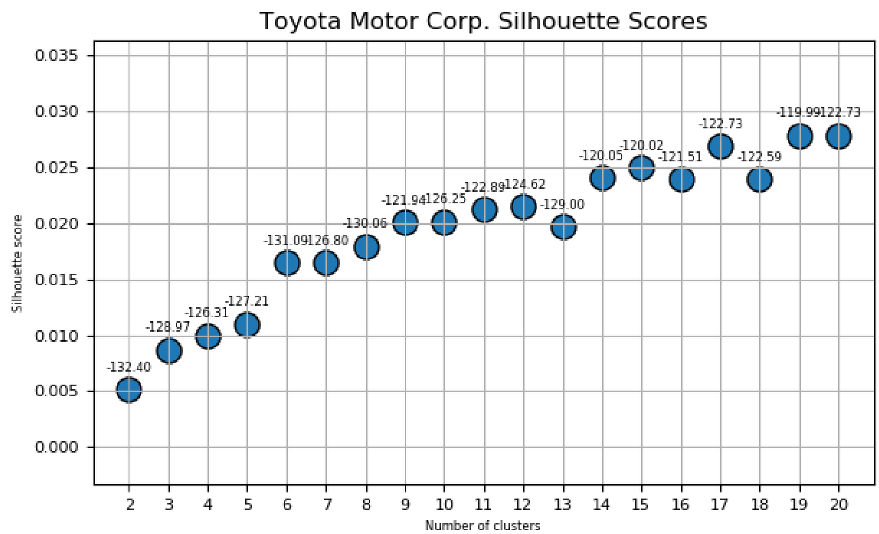
-I went ahead and checked the results which are printed to console and saw that there was about a 16.12% increase in silhouette scores from k = 11 to k = 12 topics. Perhaps defining a 15% change as 'significant' is suitable?



-For Microsoft, the best performing k value is located at k = 15 topics. Furthermore, there's a 15.82% increase in silhouette scores from k = 13 to k = 14 topics. Thus, Microsoft's optimal k value is located within the vicinity of a significant change in silhouette scores.



-For Samsung, the best performing k value is located at k = 6 topics (when k = 2 topics is not considered eligible). Furthermore, there's a 15.70% increase in silhouette scores from k = 5 to k = 6 topics, putting Samsung's optimal k in the vicinity of a significant change.



-Toyota's optimal k value is located at k = 19 topics. There's a 16.53% increase in silhouette scores from k = 18 to k = 19 topics, putting Toyota's optimal k in the vicinity of a significant change in silhouette scores.

Section 6: BTM Topic Modelling Results

Section 6 contains final BTM topic assignments and their top words. Furthermore, stacked bar charts of BTM topic performance category proportions, as well as BTM topic descriptive statistics are contained here as well. Finally, all notes taken while identifying underlying themes of topic categories are also in section 6. However, the remainder of BTM topic analysis is contained in section 10 of the appendix.

Table 6.1: Amazon OT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	support product live help day check new amazonlive chef 8pm	-84.22
1	people safe mask keep customer get covid community need safety	-60.32
2	check see get amazon delivery story go member love well	-98.41
3	amazon new make plan first help like team resolution never	-97.04
4	provide join location visit get important weve detail find registration	-37.50
5	story see work blog amazon support proud like deliver family	-60.49
6	amazon work day honor family create protect help path year	-84.45
7	amazon associate make community serve driver see team happy holiday	-85.58
8	amazon delivery help check see gift story app proud gear	-97.78
9	doorstep deliver team partner bring primeday delivery world pre go	-72.13

10	amazon free new seattle next climatepledge support offer oak view	-83.20
11	amazon holiday raise awareness amazongoesgold childhood cancer fun gift kid	-65.43
12	amazon today new prime september provide free get start day	-86.88
13	need community item wish list donate kid affect partner Bahamas	-62.60
14	amazon 19 store covid new help product support buy mission	-70.72
15	amazon small business holiday season make store item favorite charity	-80.00
16	home help family stream stay people music thank make amazon	-60.57

-Tweets belonging to topic 0 seem to mostly be ads for Amazon products and services.

-Tweets belonging to topic 1 are certainly COVID related tweets, with perhaps a few tweets erroneously assigned to topic 1.

-I was unable to ascertain an underlying theme for topic 2.

-Tweets belonging to topic 3 seemed to be light-hearted, non-product related, Amazon promotional tweets.

-About half the tweets I saw assigned topic 4 were election/voting related tweets. The other half seemed to be stories/profiles covering Amazon employees, or other people working on innovative changes.

-Tweets assigned topic 5 definitely have an underlying theme of covering employee stories/profiles, as some of the tweets assigned topic 4 were.

-I was unable to ascertain an underlying theme to topic 6.

-I was unable to ascertain an underlying theme to topic 7.

-Tweets assigned to topic 8 seem to have an underlying theme regarding ‘deliveries’ and efforts to keep Amazon’s carbon footprint as low as possible.

-Tweets assigned topic 9 all seem to contain the hashtag ‘#PrimeDay’.

-No underlying theme found for topic 10.

-No obvious underlying theme found for topic 11.

-No obvious underlying theme found for topic 12.

-No obvious underlying theme found for topic 13.

-Tweets assigned topic 14 seem to be regarding disastrous events including, but not limited to, COVID.

-Tweets assigned to topic 15 seem to have an underlying theme of charity and/or supporting small businesses.

-Tweets assigned topic 16 seem to have an underlying theme of streaming/downloading media/music.

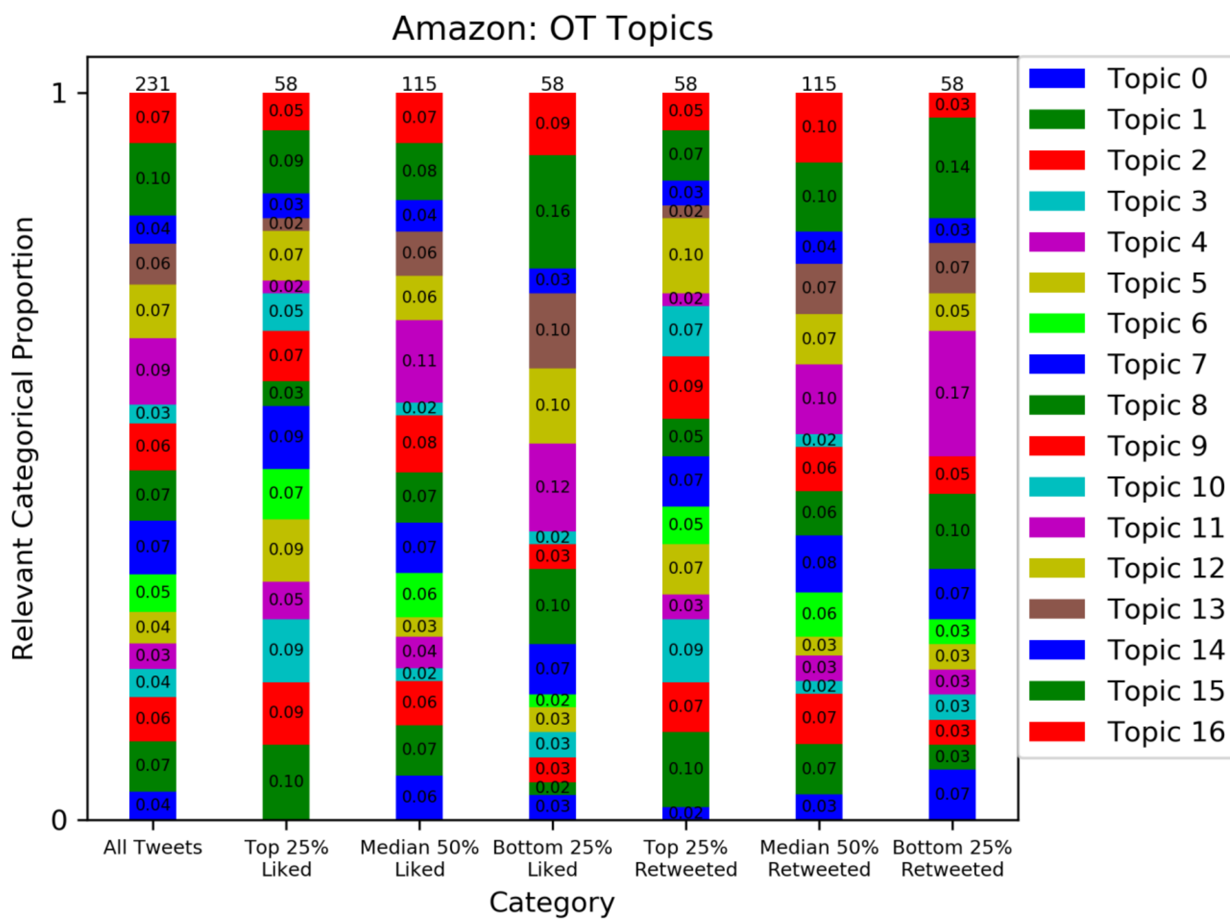


Table 6.2: Amazon OT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	9	0.038961	166.3333	50.75431	31.22222	14.85579
1	16	0.069264	311.25	216.0199	52.25	36.08231
2	14	0.060606	284.1429	225.1956	53.71429	39.56383
3	9	0.038961	858	1384.528	139.4444	210.3516
4	8	0.034632	255	129.4968	38.875	19.93875
5	10	0.04329	277.3	144.2976	56.3	53.86413
6	12	0.051948	321.0833	357.2848	60.5	87.33062
7	17	0.073593	234.2941	188.8325	46.05882	61.94299
8	16	0.069264	176.9375	104.5935	34.125	23.11385
9	15	0.064935	262.8667	132.3846	41.4	20.1239
10	6	0.025974	317.8333	178.0881	77	52.58517
11	21	0.090909	196.4286	231.0798	25.47619	11.52657
12	17	0.073593	283.7647	274.8737	55.17647	60.77544
13	13	0.056277	163.3077	65.70691	27.46154	16.03442
14	9	0.038961	207.2222	107.4742	39.44444	25.72018
15	23	0.099567	662.3913	1854.408	82.65217	162.0639
16	16	0.069264	356.125	520.2187	61.5	91.39803

Table 6.3: Amazon IRT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	love thanks hear happy enjoy thank use share hope please	-102.58
1	send detail deliveringsmiles holiday love season like surprise please well	-48.68
2	please sorry order help detail know hear account provide like	-86.40

-Tweets belonging to topic 0 seem to be ‘positive feedback responses.’ In other words, Amazon is replying to a satisfied customer.

-Tweets belonging to topic 1 are also ‘positive feedback responses’, but with two possible differences. These tweets, generally, seem to be followed either by a question, further engaging with the customer, or an offer to send the customer a small gift/surprise through a ‘secure link.’

-Tweets assigned to this topic seem to be classic cases of customer service. In other words, there’s some issue and Amazon is working with the customer to address it.

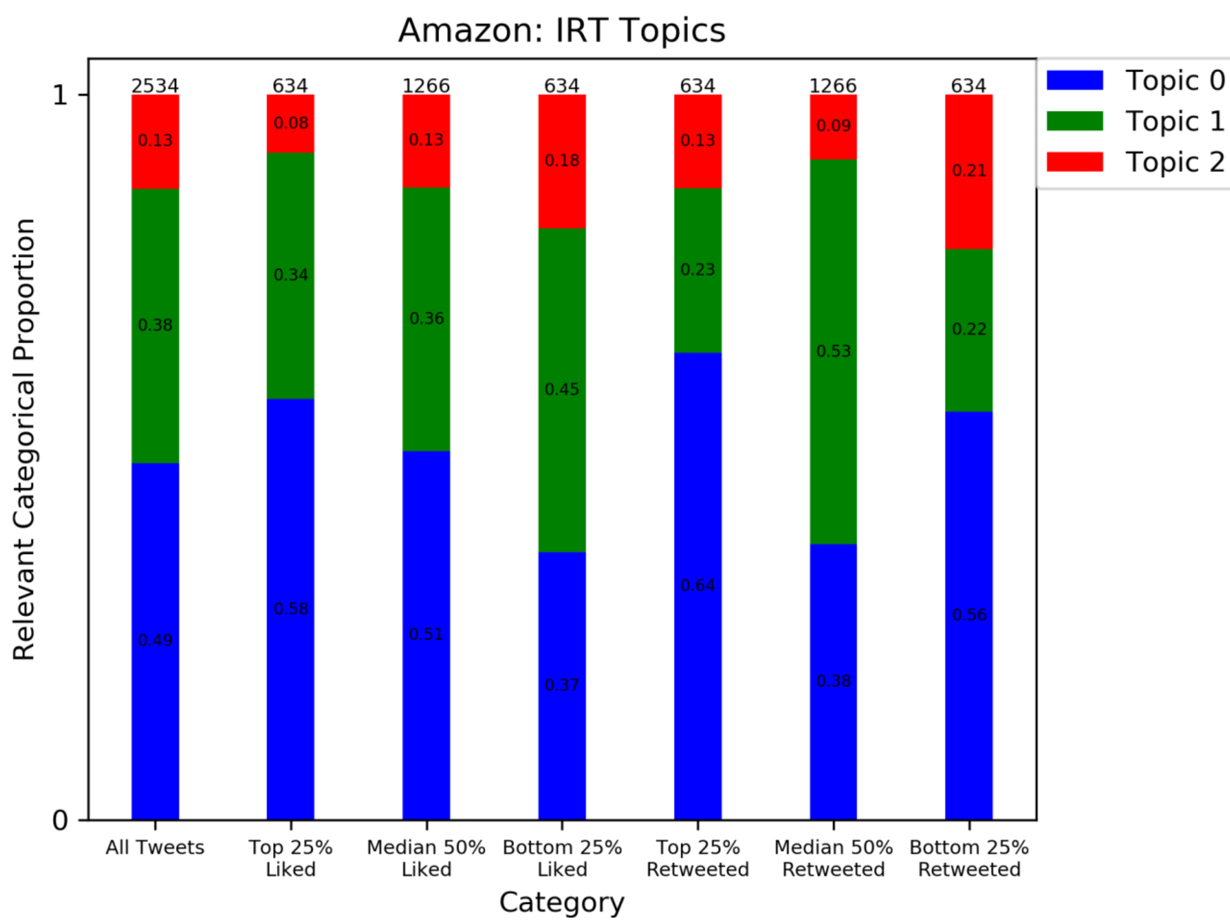


Table 6.4: Amazon IRT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	1247	0.492107	3.970329	50.94506	0.64154	3.106815
1	958	0.378058	1.649269	3.335342	0.265136	0.582446
2	329	0.129834	2.294833	14.90952	0.534954	1.05323

Table 6.5: BMW OT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	bmw combine 100 consumption km emission fuel co energy new	-34.43
1	bmw series unitedinrivalry new world next drive get time driving	-127.14
2	bmw first series ever coup gran new competition the8 them8	-87.79

-Tweets assigned topic 0 seem to, generally, discuss the specifications of vehicles.

-Tweets assigned topic 1 seem to, generally, be tweets in which BMW is attempting to interact/connect with their followers.

-Tweets assigned topic 2 seem to be, generally, regarding new BMW vehicles and offering people resources to learn more about BMW and its products/services.

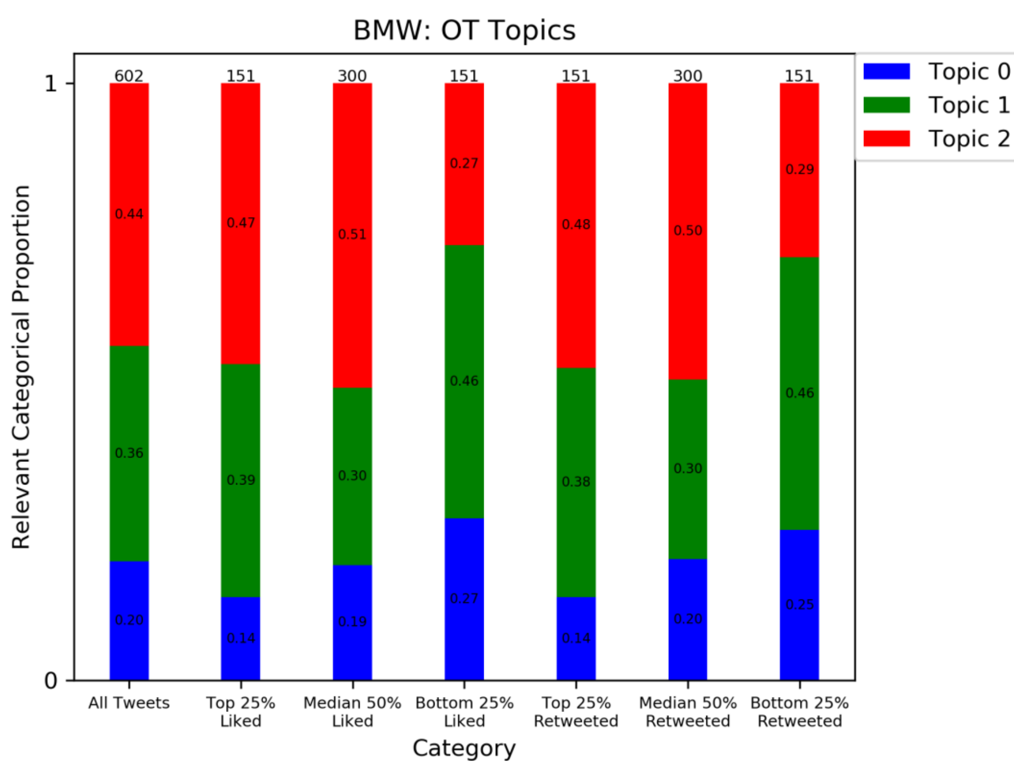


Table 6.6: BMW OT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	120	0.199336	1086.508	660.6297	115.6833	92.29146
1	217	0.360465	1277.106	930.305	138.0968	126.6364
2	265	0.440199	1413.026	727.8116	144.9208	100.3499

Table 6.7: BMW IRT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	bmw vehicle production please information comment thank stay early interest	-108.68
1	thank pleasure wish drive sheer time glad like share fan	-91.80
2	bmw km combine 100 consumption emission fuel co2 engine	-38.30
3	bmw please market contact local dealer availability visit may price	-63.60
4	bmw contact dealer happy assist local would directly suggest thank	-53.20
5	contact team form use recommend link following assist click submit	-14.97
6	thank like message send would know please dm advance private	-17.69
7	bmw colleague team contact phone drive country reach assist suggest	-41.38

-Topic 0: Perhaps an underlying theme of ‘positive feedback responses.’

-Topic 1: Perhaps could be merged with topic 0, as these seem to be positive feedback responses as well.

-Topic 2: Perhaps an underlying theme of ‘technical responses.’ In other words, BMW is really getting into the technical details of their vehicles in these replies.

-Topic 3: Unable to identify underlying theme. Maybe 1/3 of these tweets I saw assigned topic 3 were regarding contacting a local dealer.

-Topic 4: This topic certainly has an underlying theme of asking customers to ‘please contact their local dealer.’

-Topic 5: Seems that topic 5 mostly, or perhaps solely, consists of a repeated tweet in which the BMW social media team is informing the customer that they are no longer able to assist with the issue.

-Topic 6: Underlying theme of requesting to continue a conversation with a customer through 'direct messages' (DM).

-Topic 7: Underlying theme of putting consumers in touch with BMW's 'colleagues' to further assist with their specific questions/issues/requests.

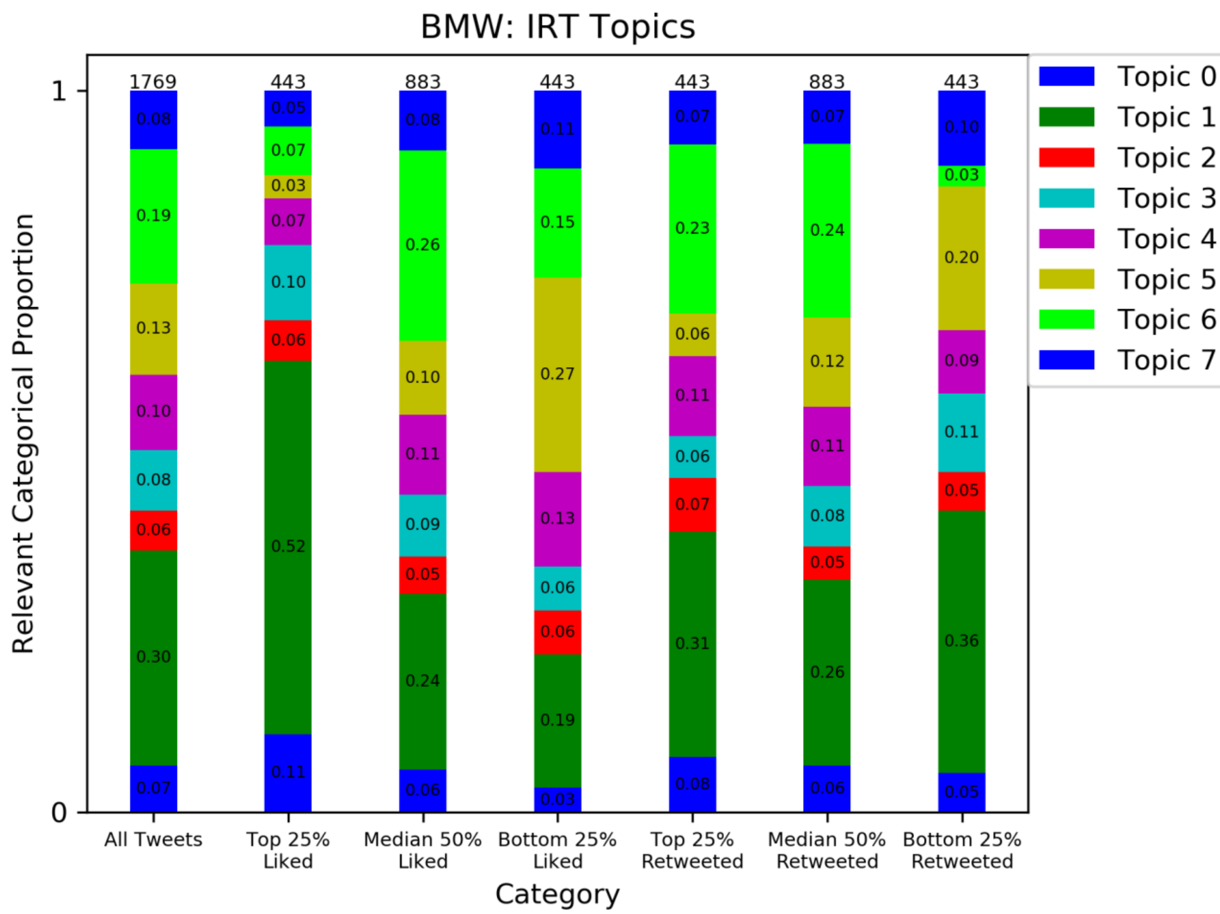


Table 6.8: BMW IRT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	115	0.065008	2.191304	4.932697	0.2	0.565065
1	526	0.297343	3.747148	16.2587	0.212928	1.08685
2	98	0.055399	1.091837	1.753269	0.102041	0.304258
3	149	0.084228	1.550336	3.572485	0.080537	0.427386
4	185	0.104579	0.772973	1.827425	0.081081	0.292885
5	223	0.12606	0.273543	0.711279	0.03139	0.290854
6	330	0.186546	0.518182	1.75765	0.030303	0.188555
7	143	0.080837	1.27972	3.91531	0.118881	0.523928

Table 6.9: Coca Cola IRT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	hi coca cola product please send store dm thanks sorry	-71.27
1	product demand focus appreciate temporarily high certain loyalty time unprecedented	-20.10
2	please reach thanks call like hear 800 provide give contact	-58.67

-Topic 0: There is perhaps an underlying theme of 'unavailability' associated with topic 0.

-Topic 1: Definite underlying theme of not being able to produce all products during this 'unprecedented time' (COVID, I assume) and that Coca Cola is focusing on their products with the highest demand.

-Topic 2: Underlying theme of either providing a phone number for the consumer to contact or requesting that the conversation be continued through DMs.

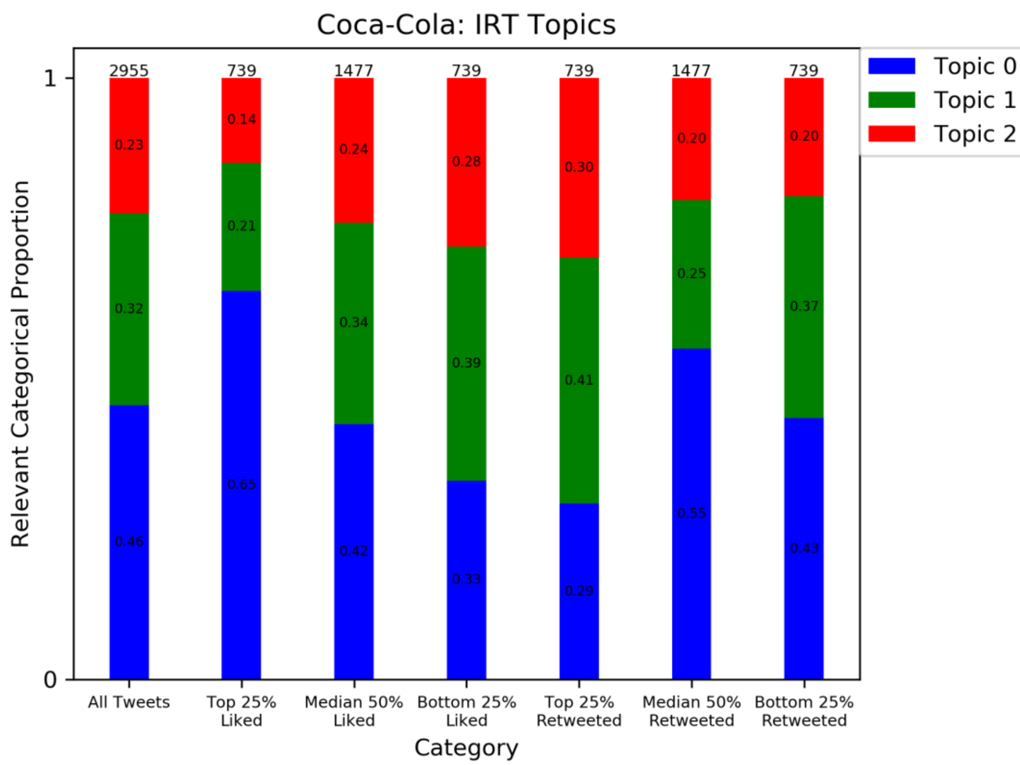


Table 6.10: Coca Cola IRT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	1348	0.456176	49.12834	1669.514	7.352374	264.684
1	941	0.318443	0.758767	8.614821	0.068013	1.575628
2	666	0.225381	0.540541	5.752463	0.033033	0.243018

Table 6.11: Google IRT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	hi step let know try help tip may account confirm	-40.89
1	help email address reveal well without hi guide may message	-30.66
2	hi help account google suggest hope know security see extra	-97.55
3	dm next let step hi see help know please follow	-30.70

-Topic 0: Underlying theme of Google prodding for more information from the consumer regarding their issue. In other words, Google is trying to help the consumer but has a few questions regarding their problem.

-Topic 1: These seem similar to topic 0, except all tweets belonging to topic 1 seem to include the phrase "without revealing your email address."

-Topic 2: Perhaps an underlying theme of security/privacy, and I noticed quite a few of these tweets mentioned 'Google One.'

-Topic 3: Underlying theme of requesting to continue conversation with customer through DMs.

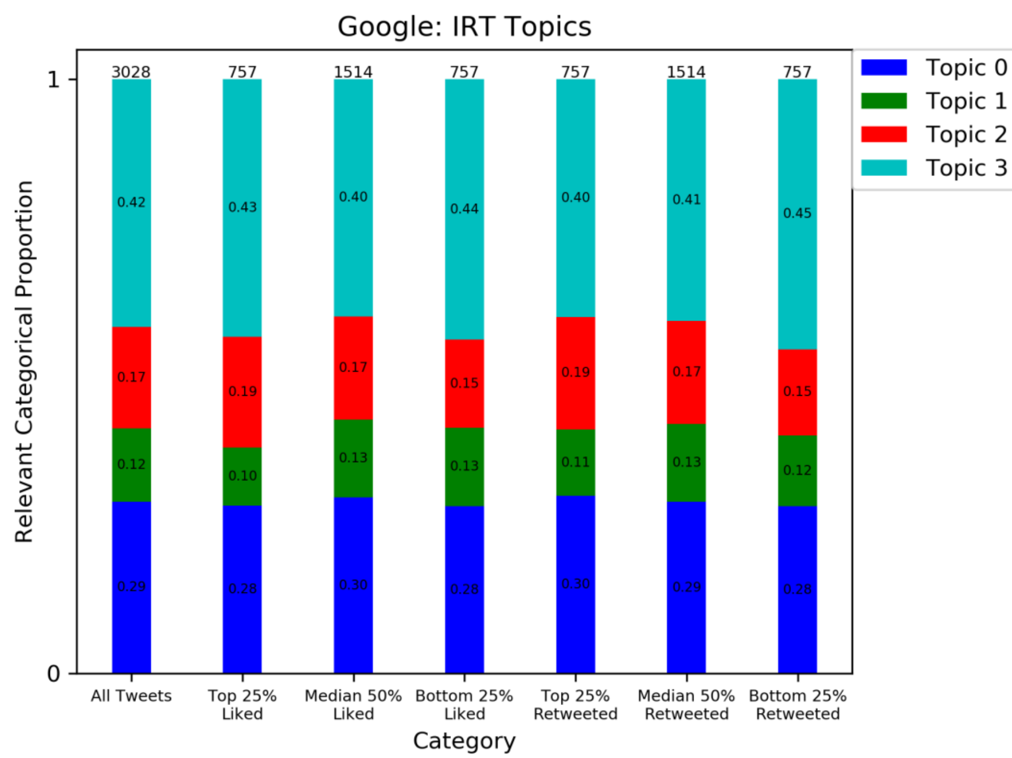


Table 6.12: Google IRT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	876	0.2893	0.215753	0.897121	0.018265	0.133984
1	373	0.123184	0.150134	0.37956	0.005362	0.073127
2	516	0.17041	39.68023	872.8856	12.39147	278.8373
3	1263	0.417107	0.209818	0.467144	0.022169	0.157685

Table 6.13: McDonald's IRT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	menu future never update go tune stay see limited fun	-49.64
1	tell please detail share link like get click want would	-99.78
2	back menu bring still day breakfast time thing stay keep	-74.79
3	please provide info app team mobile look like tell hear	-74.38
4	let know local like menu restaurant back see vary snack	-68.47
5	new may give find fave option chance upset definitely disappear	-41.14

-Topic 0: Unable to identify underlying theme.

-Topic 1: Underlying theme of 'classic customer service' type tweets. In other words, there was an actual issue or something that needed to be reported/addressed. Sometimes, however, the customer service may seem to be positive, although not generally.

-Topic 2: Underlying theme of 'changes to the menu.'

-Topic 3: This topic should probably be merged with topic 1, they seem about the exact same to me.

-Topic 4: Unable to identify underlying theme. Seemed to be a mixture of all topics above.

-Topic 5: I would dub these 'glass half full' tweets. As in topic 2, these seem to mostly be related to menu changes. However, these tweets oftentimes seem to be followed by a positive, such as "to give you something new to vibe to" or "perhaps you'll find a new favorite!"

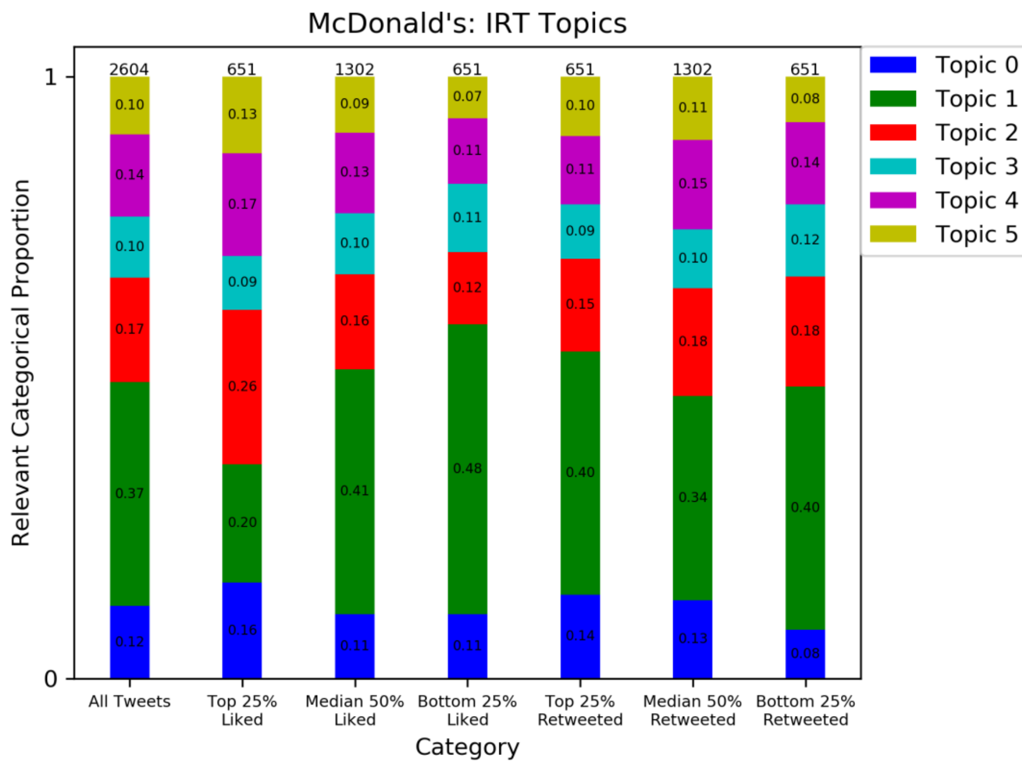


Table 6.14: McDonald's IRT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	314	0.120584	1.44586	8.740413	0.092357	0.400964
1	969	0.37212	2.080495	25.80801	0.0258	0.223503
2	452	0.173579	1.679204	8.543418	0.068584	0.253026
3	264	0.101382	12.20833	88.563	0.238636	1.453939
4	356	0.136713	15.43258	129.4311	0.292135	2.255271
5	249	0.095622	7.437751	43.45146	0.100402	0.350633

Table 6.15: Mercedes OT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	mercedesbenz mercedes benz new eq switchtoeq enjoyelectric one worlds2020 design	-107.75
1	mercedes mercedesbenz benz mercedesamg via amg drivingperformance get mbclassic chic	-94.18
2	mercedesbenz class car mercedes halloween make history via happy technology	-113.14
3	mercedesbenz benz eq mercedes gorden design switchtoeq wagener thenewluxury gclass	-98.47

4	mercedesbenz mercedes benz sclass new class get world learn luxury	-82.30
5	mercedesbenz new class learn caresforwhatmatters thenewluxury thenewsclass mercedesmaybach sclass Mercedes	-77.12
6	mercedesbenz mercedes class benz coup new design learn model via	-79.31
7	mercedesbenz benz mercedes class new car model learn sclass get	-81.66
8	mercedes benz mercedesbenz mbclassic car classic alltimestars get month 280	-75.11
9	mercedes co km emissionen kombiniert 100 benz kwh stromverbrauch 4matic	-38.64
10	mercedes benz mercedesbenz via mbclassic car sl 300 get visit	-86.94

-Topic 0: Unable to find underlying theme for topic 0.

- Topic 1: Unable to find underlying theme.

- Topic 2: There is perhaps an underlying theme related to car design and aesthetics associated with topic 2.

- Topic 3: There is perhaps an underlying theme of ‘luxury’ vehicles associated with topic 3.

- Topic 4: Perhaps underlying themes of ‘adventure’ and luxury vehicles again.

- Topic 5: Again, these seem to be regarding luxury vehicles, as well as Mercedes tweets posted to serve as evidence of how they’re rising above the competition (the extra features and such that Mercedes offers).

- Topic 6: Unable to find underlying theme for topic 6.

- Topic 7: Unable to find underlying theme for topic 7.

- Topic 8: There is perhaps an underlying ‘artistic’ theme associated with topic 8.

- Topic 9: No obvious underlying theme found for topic 9.

- Topic 10: Tweets assigned to topic 10 seem to mostly be related to older Mercedes models and classic cars.

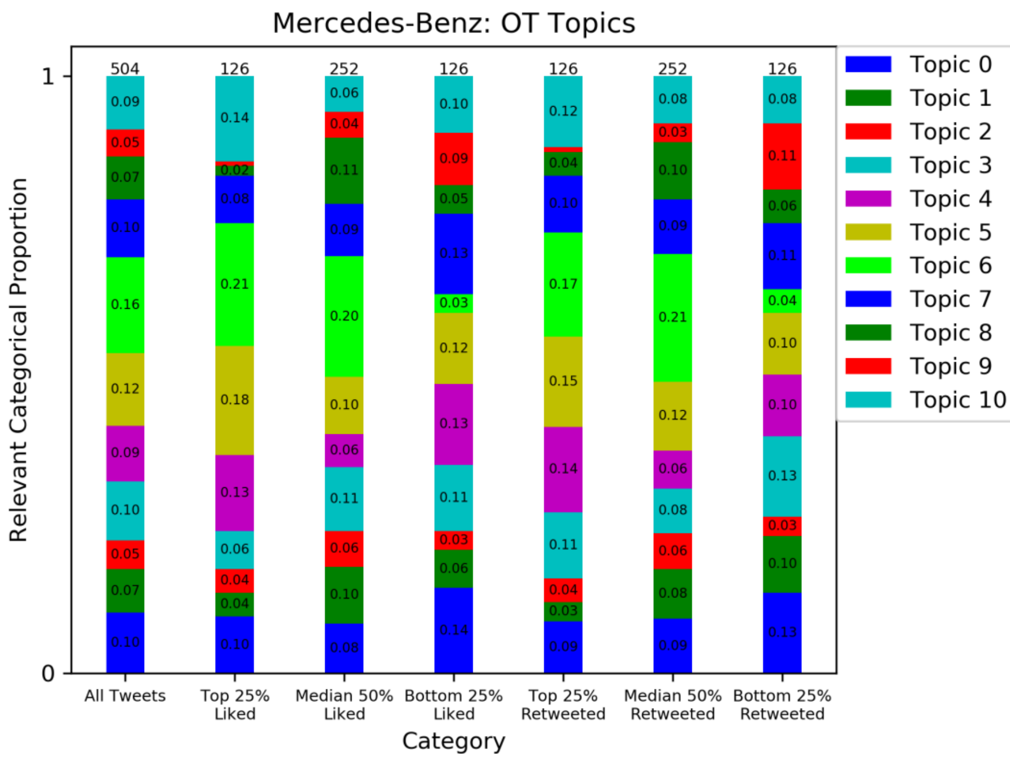


Table 6.16: Mercedes OT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	51	0.10119	1138.804	1231.442	127.2353	176.6543
1	37	0.073413	883.5135	398.8723	77.97297	42.2877
2	24	0.047619	1512.583	2190.898	156.5	250.8835
3	50	0.099206	900.96	562.3436	93.88	60.74894
4	47	0.093254	1390.489	2226.033	236.2979	723.656
5	61	0.121032	1134.59	677.5184	116.918	88.92924
6	81	0.160714	1257.63	545.098	112.5309	54.02733
7	49	0.097222	942.7551	547.4318	96.93878	73.88262
8	36	0.071429	805.9167	436.9173	86.69444	53.03951
9	23	0.045635	633.3913	406.0914	56.6087	38.26551
10	45	0.089286	1463.156	1674.163	154.1111	239.5434

Table 6.17: Mercedes IRT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	sorry colleague team service customer hear contact touch best please	-62.70
1	sorry service customer team hear 800 colleague 367 6372 best	-70.42
2	great hope one see star get like look glad welcome	-122.46
3	local know country please retailer let mercedes thanks contact benz	-71.05
4	00800 777 team 77 customer service mon fri sorry reach	-43.69
5	dm send detail please number address contact full email postal	-49.20

-Topic 0: Perhaps underlying theme of being ‘classic customer service’ tweets.

-Topic 1: Probably should be merged with topic 0, as these seem to be the same type of tweets.

-Topic 2: Perhaps underlying theme of being ‘positive feedback responses.’

-Topic 3: Unable to identify underlying theme.

-Topic 4: Perhaps underlying theme of redirecting questions/concerns to the appropriate team/division at Mercedes.

-Topic 5: Underlying theme of continuing conversation through DMs.

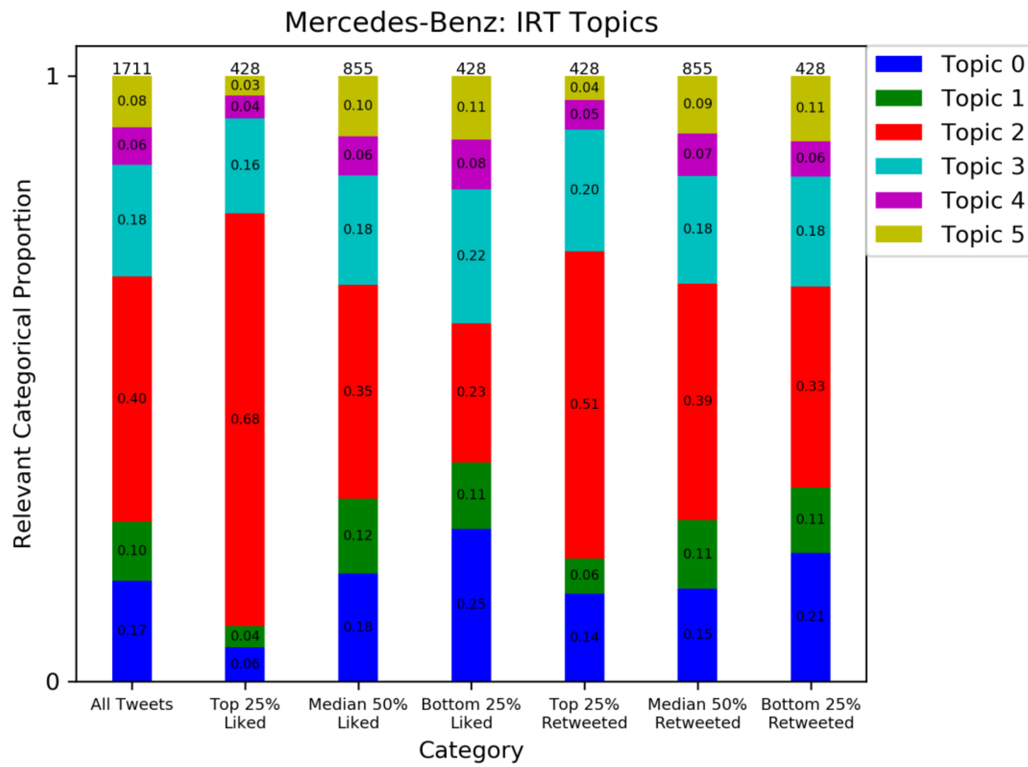


Table 6.18: Mercedes IRT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	284	0.165985	0.172535	0.439023	0.007042	0.08377
1	168	0.098188	1.14881	10.77013	0.220238	2.479679
2	692	0.404442	1.489884	6.600764	0.171965	0.585073
3	316	0.184687	3.484177	52.03665	0.512658	8.045298
4	106	0.061952	0.566038	2.927762	0.018868	0.194257
5	145	0.084746	0.2	0.596285	0.013793	0.117036

Table 6.19: Microsoft OT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	microsoft learn new help microsoftteams make people experience one msignite	-111.74
1	learn microsoft new ai use help world msignite work technology	-94.81
2	learn microsoft cloud business new revenue msignite service help billion	-117.88

-Topic 0: Tweets assigned to topic 0 seem to generally be tweets regarding Microsoft Teams, or other Microsoft tweets which aren't too heavy on the technical side of things.

-Topic 1: Tweets assigned to topic 1 seem to be, generally, regarding what new products/services Microsoft is working on to help people/the world.

-Topic 2: Tweets assigned to topic 2 seem to be, generally, about 'non-revolutionary' ways in which you can, or Microsoft is, helping people. In other words, they're not about cutting-edge technology being designed to help people, rather they're about what anyone can reasonably do to help.

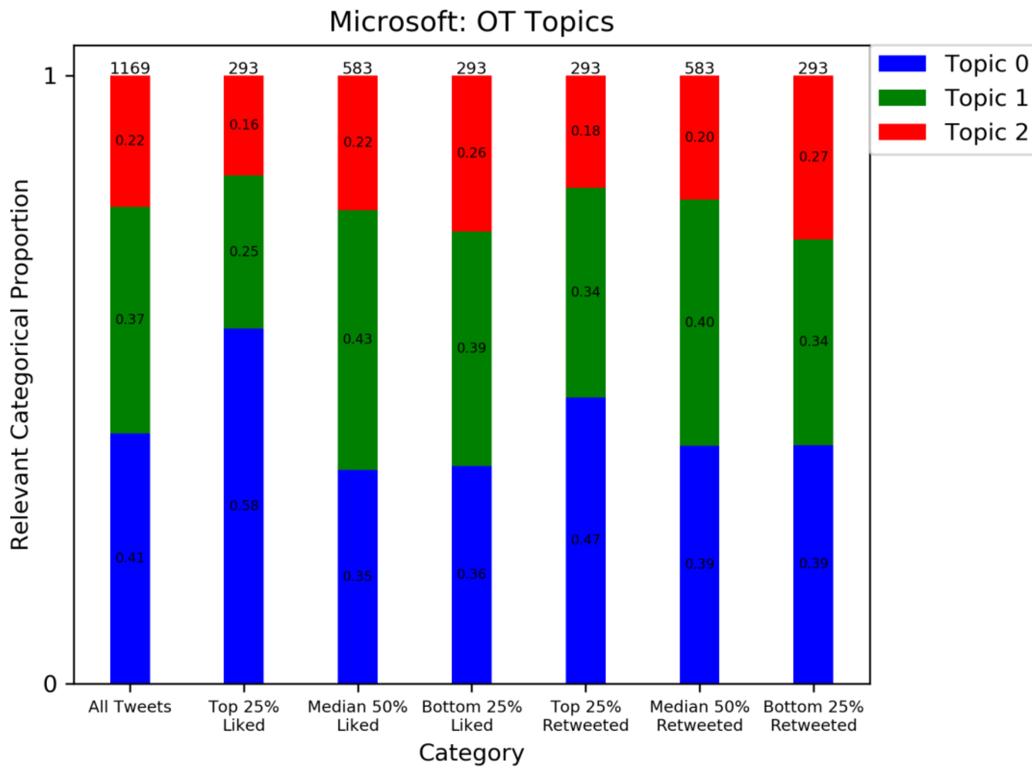


Table 6.20: Microsoft OT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	481	0.411463	795.7256	2173.86	119.0894	323.3222
1	436	0.372968	519.7156	1394.38	105.1789	226.5164
2	252	0.215569	614.6389	2066.757	135.7579	585.0361

Table 6.21: Microsoft IRT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	day get hope deserve back word heart friday thank turn	-88.63
1	10 afternoon snack late morning deep thought cant wrong evening	-67.21
2	little could go use something everything positive reinforcement love gif	-69.04
3	10 really extra appreciate outlook coffee cup go like work	-81.41
4	one thank world time see love conquer kindness act random	-76.28
5	people 10 bring know like would great feel pretty think	-94.97
6	good get love hi make sunshine dog point like keep	-94.38
7	say could weve yes pet enjoy human josh powerpoint presentation	-64.41
8	like look happy hey well way try seem rock take	-90.20
9	great see work hi look appreciate favorite lose like one	-96.19
10	let go know chat challenge taste great every gifs occasion	-68.96
11	still important agree use doggo ruckus think human feature background	-38.85
12	take get know day two excel time team must everyone	-93.79
13	team theme mute call high time hop sorry base minute	-38.02
14	10 great pizza excel right pretty old choice excellent lent	-95.19
15	wait expert cant well keep skill everybody see need team	-70.14
16	someone perfect 10 meeting make one ask look need like	-100.16
17	feature team photo hey philip adorable chat george dm link	-36.80
18	tweet block report account abusive comment tolerate speech hate action	-10.11
19	thanks team make today last minute glad connect two entire	-80.11

Unfortunately, due to the large number of topics, no further exploration was done for Microsoft IRT regarding topic themes.

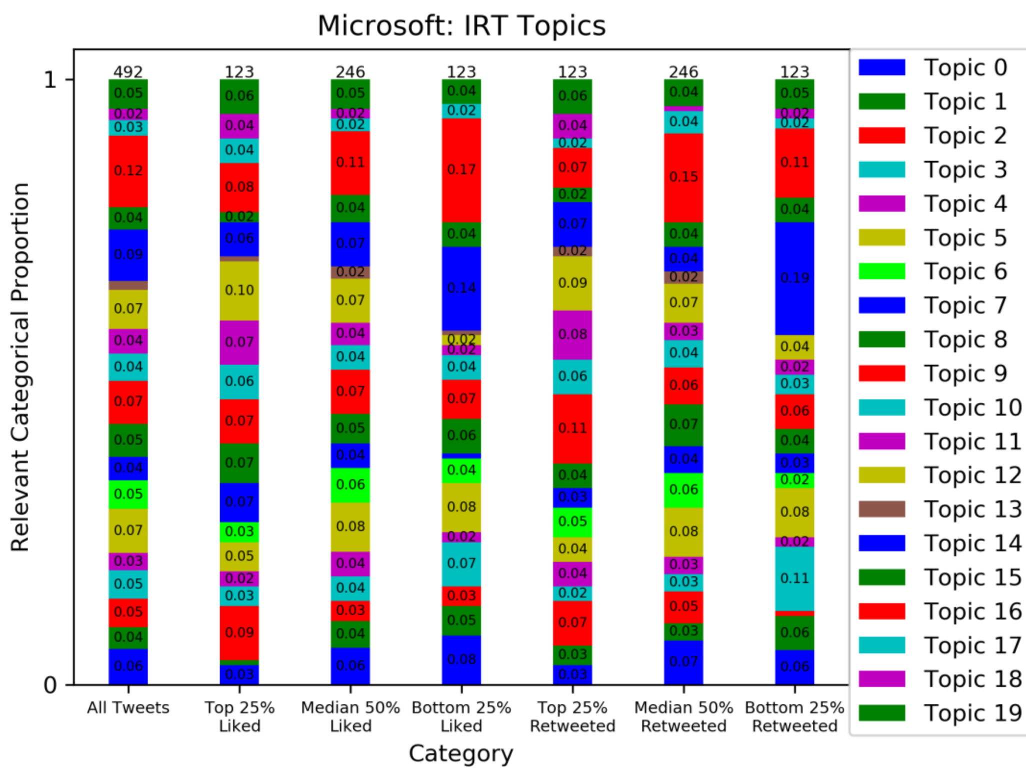


Table 6.22: Microsoft IRT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	29	0.058943	6.034483	8.666367	0.172414	0.468201
1	18	0.036585	8.833333	21.17226	0.5	1.043185
2	23	0.046748	281.3913	1080.726	5.652174	16.6067
3	23	0.046748	14.52174	33.78793	0.521739	1.410015
4	14	0.028455	53.14286	139.1645	5.785714	19.0795
5	36	0.073171	20.22222	73.86711	0.5	2.171241
6	23	0.046748	49.43478	186.9696	1.130435	4.351639
7	19	0.038618	29.73684	53.74822	0.315789	0.671038
8	27	0.054878	97.25926	316.0058	1.814815	6.082997
9	35	0.071138	13.77143	20.91592	0.742857	1.038745
10	22	0.044715	31.13636	47.24838	0.863636	1.780729
11	20	0.04065	151	510.2155	9.85	36.65382
12	32	0.065041	130	414.9513	5.6875	15.8754
13	7	0.014228	19.85714	27.16878	0.571429	1.133893
14	42	0.085366	13.69048	31.65861	0.52381	1.214508
15	18	0.036585	11.77778	25.13857	0.333333	0.685994
16	58	0.117886	87.98276	367.9099	1.931034	8.843398
17	13	0.026423	19.23077	22.97881	0.230769	0.599145
18	9	0.018293	22.11111	20.94901	3.111111	4.807402
19	24	0.04878	12.91667	15.40445	1	2.284161

Table 6.23: Samsung OT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	samsung new feature tv design experience user device offer see	-114.67
1	samsung new 5g solution technology mobile first generation next ai	-121.57
2	samsung new galaxy see experience take make check get watch	-137.07
3	samsung technology new ai innovation year world see electronics award	-139.42

-Topic 0: Tweets assigned topic 0, generally, seem to be regarding Samsung products which are not mobile phones. However, a few tweets regarding mobile phones are erroneously assigned to topic 0 anyways.

-Topic 1: Perhaps an underlying theme of ‘cutting-edge’ technology associated with topic 1.

-Topic 2: Tweets assigned to topic 2 seem to be regarding mobile phones and/or any Samsung ‘Galaxy’ product.

-Topic 3: Perhaps an underlying theme of being ‘Samsung announcements.’

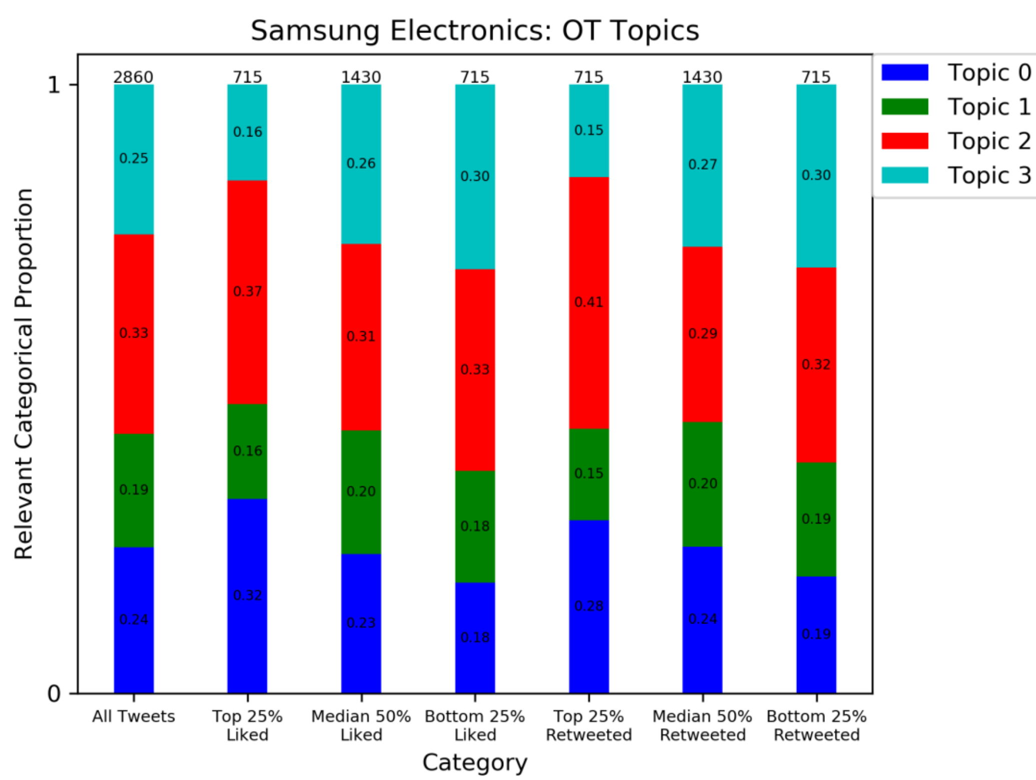


Table 6.24: Samsung OT Topic Descriptive Statistics

Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	686	0.23986	608.2464	1709.846	54.04956	162.1472
1	534	0.186713	383.073	1479.12	34.07491	97.65952
2	936	0.327273	481.7981	1948.232	56.36004	296.9165
3	704	0.246154	324.4886	1555.942	35.63494	272.3089

Table 6.25: Toyota OT Topic Modelling Results

Topic	Top 10 Common Words	Coherence
0	toyota 00 jst wonderwednesday guessthemodel tomorrow answer reveal mobilityforall pm	-73.06
1	toyota toyoda tmstoyota akio concept press tms2019 live new tune	-119.52
2	toyota car startyourimpossible dream world art mobilityforall contest come make	-123.22
3	toyota car race make gr new hybrid toyotagazooracing sport 2018	-130.50
4	toyota hilux check road long japan new priusphv drive teamacp	-144.32
5	toyota new plant hydrogen world customer vehicle japan factoryfriday mirai	-122.14
6	toyota tbt car new japan look back model classic check	-136.81
7	toyota tbt day corolla toyotatimes car go one know landcruiser	-139.29
8	toyota make vehicle drive year project 000 team use aim	-136.17
9	toyota hydrogen use toyotatimes system power emission future new society	-118.79
10	toyota mobility new car learn service vehicle drive technology japan	-110.63
11	toyota motor result news corporation announces financial time update production	-94.53
12	toyota learn car new drive vehicle technology ai future center	-127.66
13	toyota yariswrc team wrc toyotagazooracing startyourimpossible win getty congratulation action	-122.14

-Topic 0: Tweets assigned topic 0 seem to be associated with some form of '#GuessTheModel' hashtag campaign.

-Topic 1: Unable to identify underlying theme.

-Topic 2: Unable to identify underlying theme.

-Topic 3: Unable to identify underlying theme for topic 3.

-Topic 4: Unable to identify underlying theme.

-Topic 5: Perhaps an underlying theme of Toyota factories and plants.

-Topic 6: Perhaps an underlying theme of 'classic' cars and earlier times.

-Topic 7: Unable to identify underlying theme.

-Topic 8: Perhaps underlying theme of 'what Toyota is doing to improve their cars.'

-Topic 9: Perhaps an underlying theme of 'going green' associated with topic 9.

-Topic 10: Unable to identify underlying theme.

-Topic 11: Seems to be an underlying theme regarding the 'financials' of Toyota.

-Topic 12: Perhaps an underlying theme of futuristic technology.

-Topic 13: Seems to be associated, mostly, with some racing event that happened.

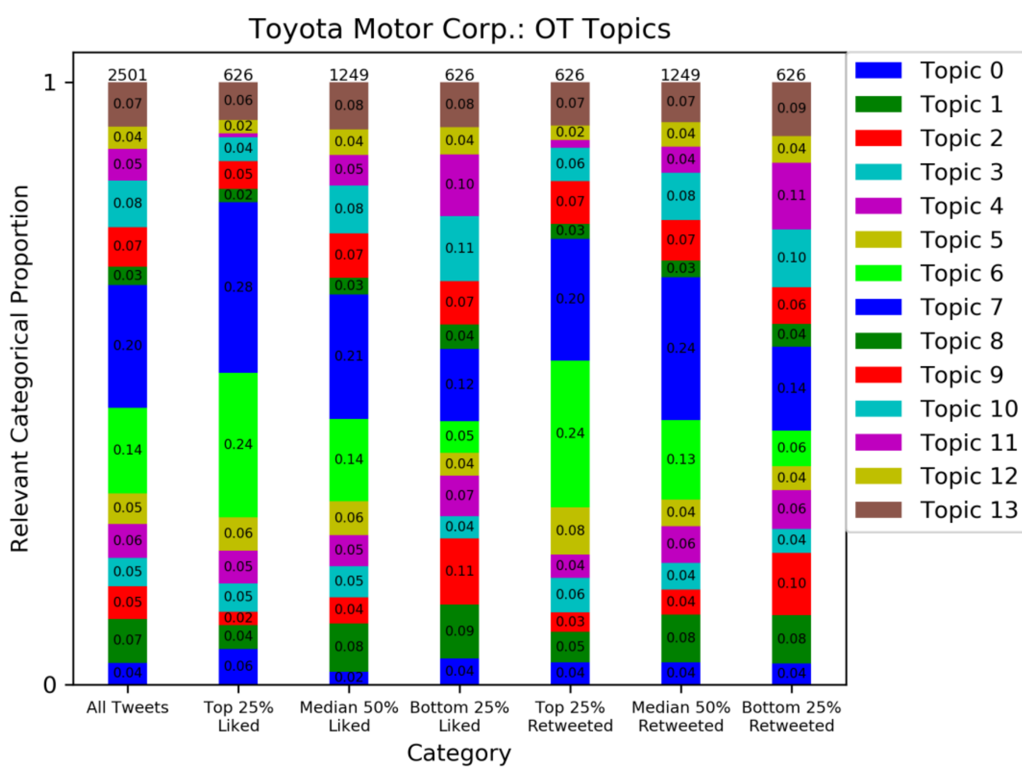


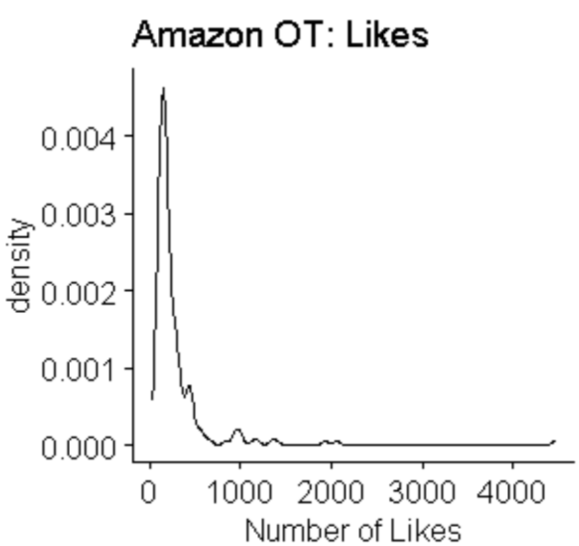
Table 6.26: Toyota OT Topic Descriptive Statistics

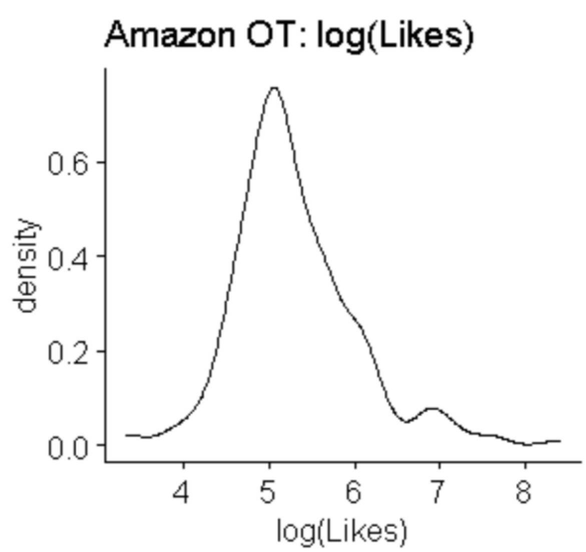
Topic	No. Obs.	Data Proportion	Avg. Likes	SD: Likes	Avg. Retweets	SD: Retweets
0	91	0.036385	101.6044	82.3947	16.65934	9.654268
1	181	0.072371	66.62431	62.38449	16.59669	16.73246
2	137	0.054778	53.66423	55.48904	14.60584	17.86401
3	117	0.046781	84.05128	63.61249	20.51282	16.4047
4	140	0.055978	81.59286	74.37325	17.19286	14.60457
5	129	0.051579	87.43411	65.54815	21.7907	15.03521
6	355	0.141943	114.2338	87.59982	25.95211	30.51573
7	510	0.203918	100.9824	70.22466	20.50784	18.65663
8	74	0.029588	65.82432	51.11828	16.17568	13.86269
9	166	0.066373	67.18675	44.88553	18.75904	14.33292
10	192	0.076769	59.21354	44.50909	15.58333	11.10508
11	131	0.052379	48.67176	44.20972	11.20611	12.91196
12	94	0.037585	75.26596	130.7827	19.08511	32.29257
13	184	0.073571	461.3696	2486.693	96.32065	546.138

<u>Section 7:</u> Link Inclusion Analysis

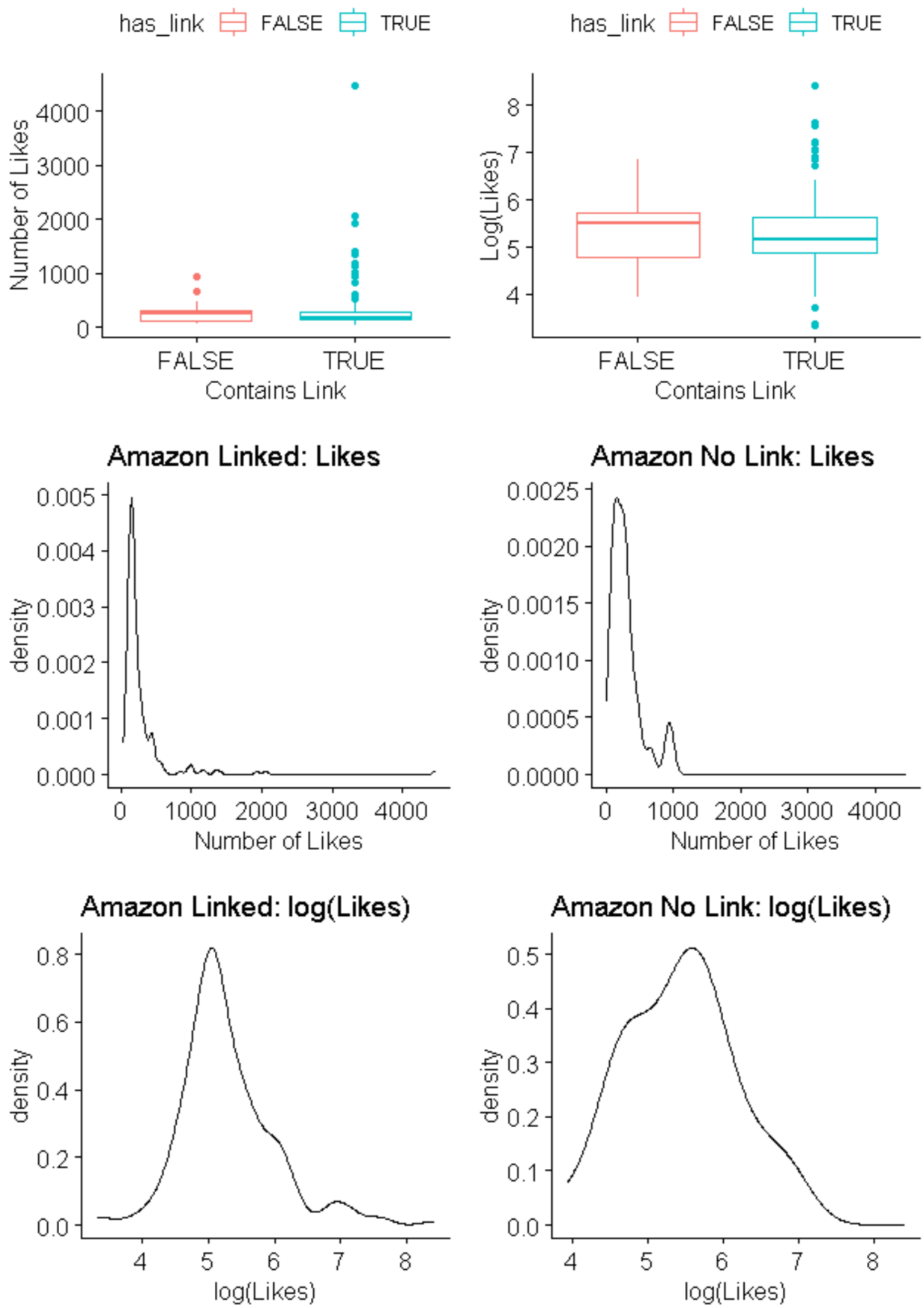
Section 7 contains the visualizations, notes, results, and conclusions made when examining companies and tweet categories for a potential link inclusion effect.

Amazon Official: Number of Likes





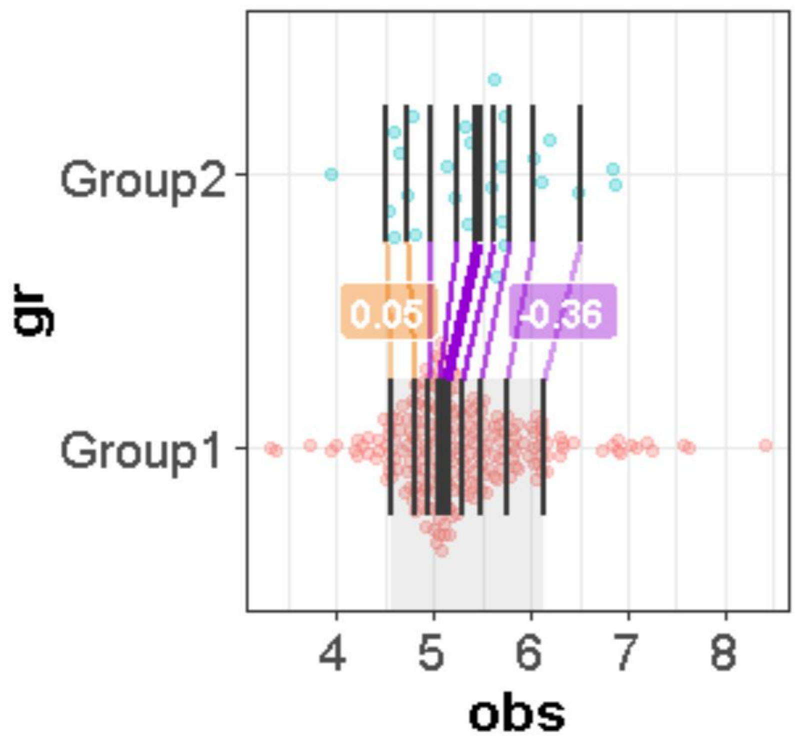
The log distribution may appear normal, but it did not pass a Shapiro-Wilk normality test.


has_link FALSE TRUE
Number of Likes
4000
3000
2000
1000
0
FALSE
TRUE
Contains Link
has_link FALSE TRUE
Log(Likes)
8
7
6
5
4
FALSE
TRUE
Contains Link
Amazon Linked: Likes
density
0.005
0.004
0.003
0.002
0.001
0.000
0 1000 2000 3000 4000
Number of Likes
Amazon No Link: Likes
density
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0 1000 2000 3000 4000
Number of Likes
Amazon Linked: log(Likes)
density
0.8
0.6
0.4
0.2
0.0
4 5 6 7 8
log(Likes)
Amazon No Link: log(Likes)
density
0.5
0.4
0.3
0.2
0.1
0.0
4 5 6 7 8
log(Likes)

The bottom right distribution may be considered normal, but none of the other distributions above passed a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 2443.5, p-value = 0.2382
alternative hypothesis: true location shift is not equal to 0
```

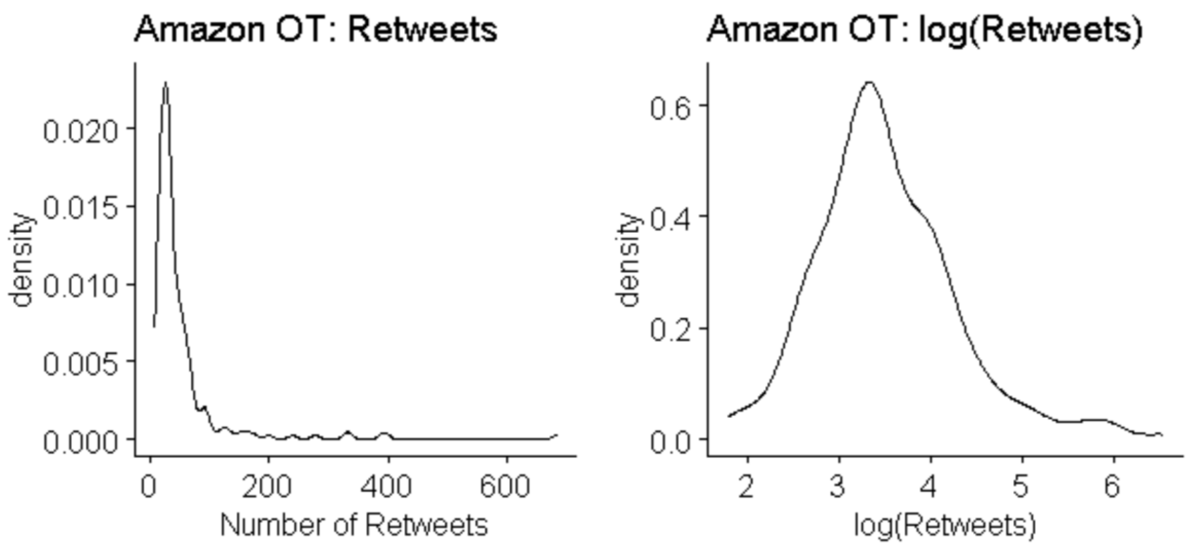
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'like' distributions of both populations are equal. Performing a shift function yields the following:



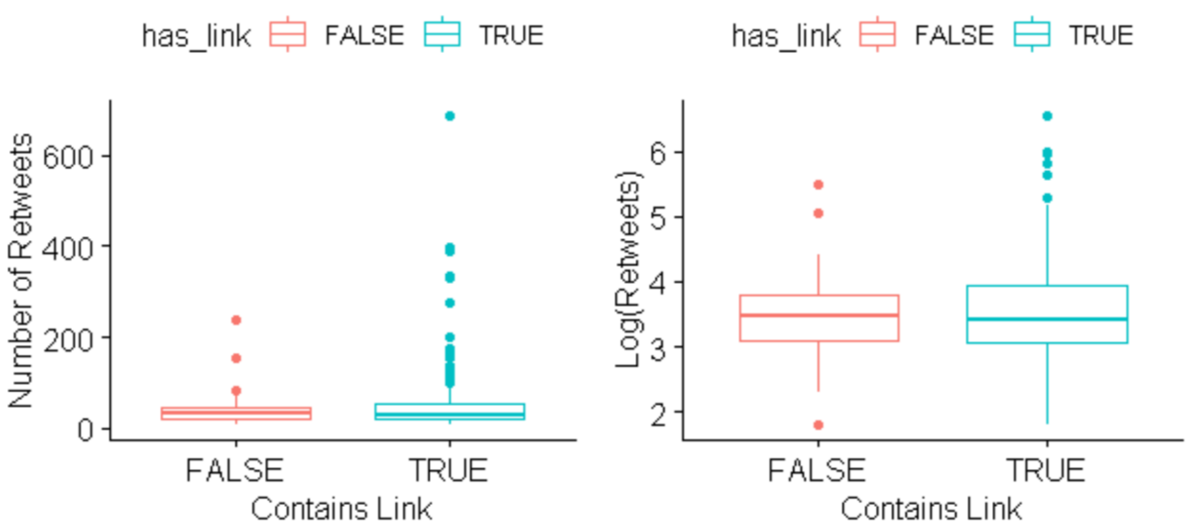

```
$`Group1 - Group2`
    q    Group1    Group2  difference   ci_lower  ci_upper      p_crit p_value
1 0.1  95.89842  93.21329    2.685137  -29.04673  33.94615 0.050000000   0.921
2 0.2 121.66556 114.12391    7.541643  -66.24627  33.87030 0.025000000   0.793
3 0.3 139.73977 149.03621   -9.296440 -108.18201  40.30363 0.016666667   0.745
4 0.4 156.92272 195.67753  -38.754816 -127.32554  38.70022 0.012500000   0.318
5 0.5 171.33101 241.70592  -70.374908 -187.11629  28.46126 0.007142857   0.069
6 0.6 199.27786 282.26865  -82.990787 -260.88723  16.75478 0.005555556   0.033
7 0.7 243.17405 329.25702  -86.082963 -345.15466  30.81659 0.006250000   0.039
8 0.8 321.31633 435.42382 -114.107488 -495.52555  85.71980 0.008333333   0.198
9 0.9 464.45413 700.27929 -235.825163 -492.41989 210.92103 0.010000000   0.239
```

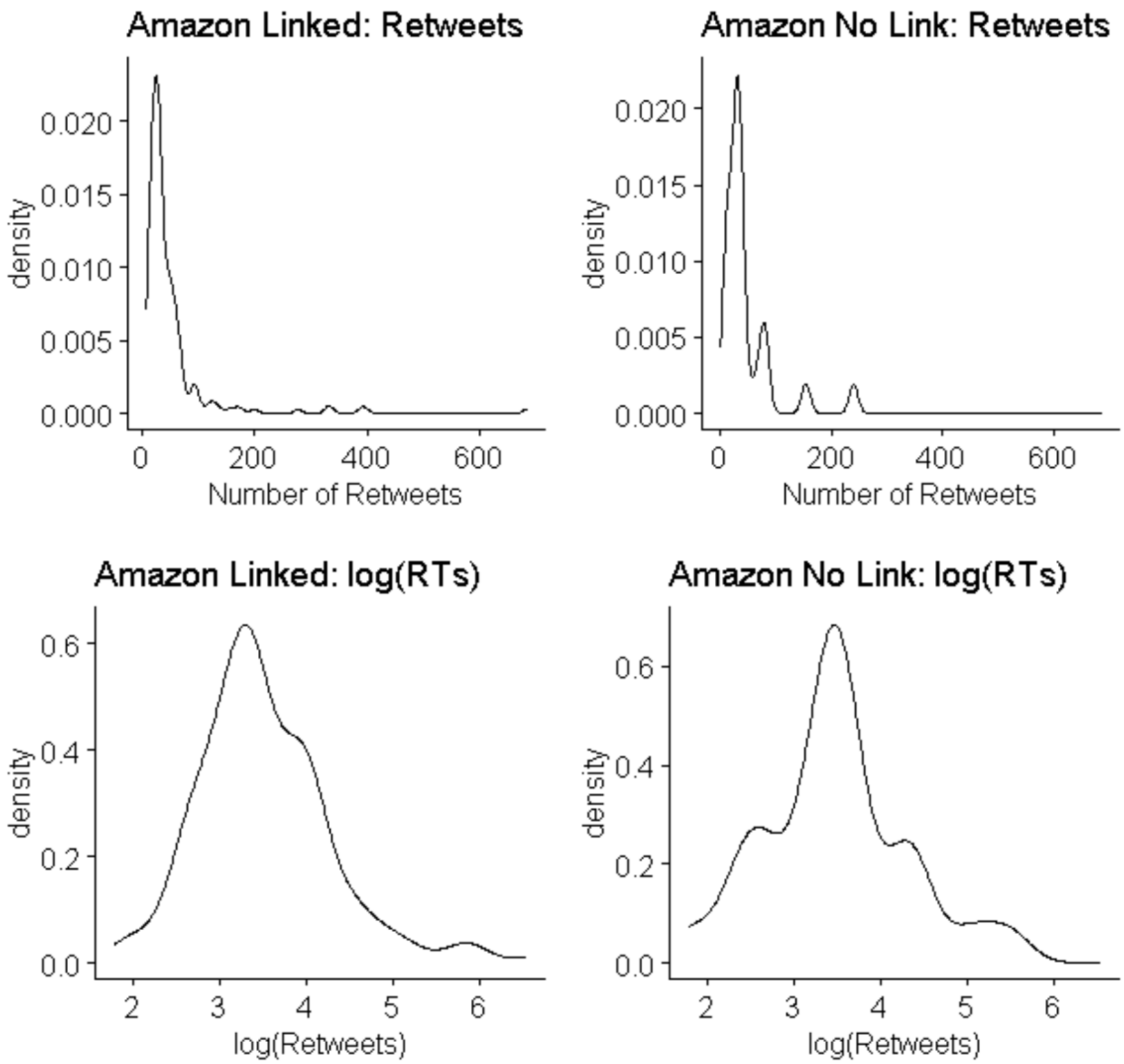
Statistics are computed using original data, visualizations are produced using log-transformed data such that resulting figures are spread out nicely. Considering each confidence interval contains the value 0, we may not conclude any quantile differences in tweet performance between the two groups. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of likes are equal to one another. **Inclusion of a link does not seem to have a statistically significant effect on the number of likes which an Amazon official tweet receives.**

Amazon Official: Number of Retweets



Again, neither of the above distributions passed a Shapiro-Wilk normality test.

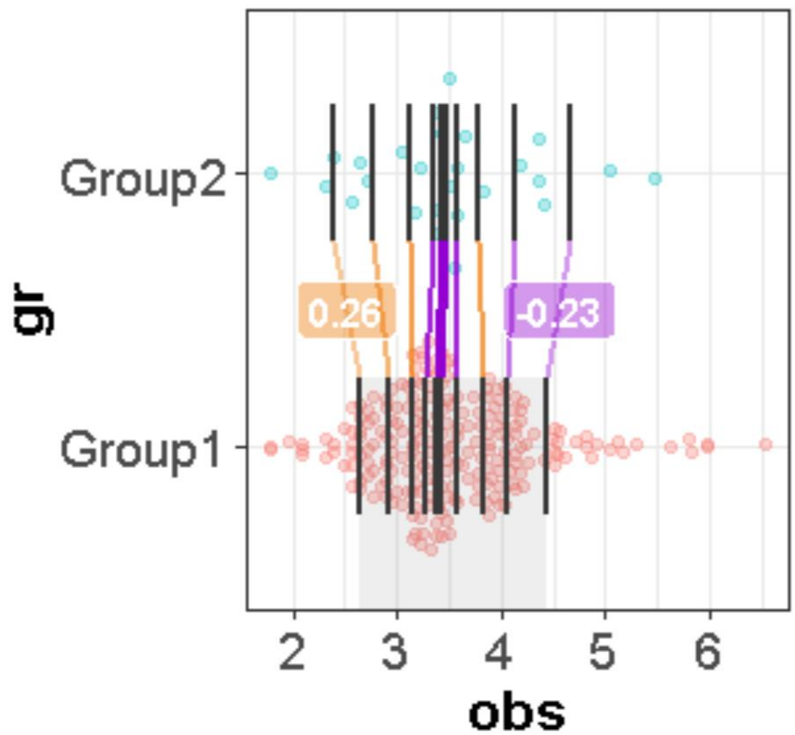




Again, only the bottom right distribution passes a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 2828, p-value = 0.9568
alternative hypothesis: true location shift is not equal to 0
```

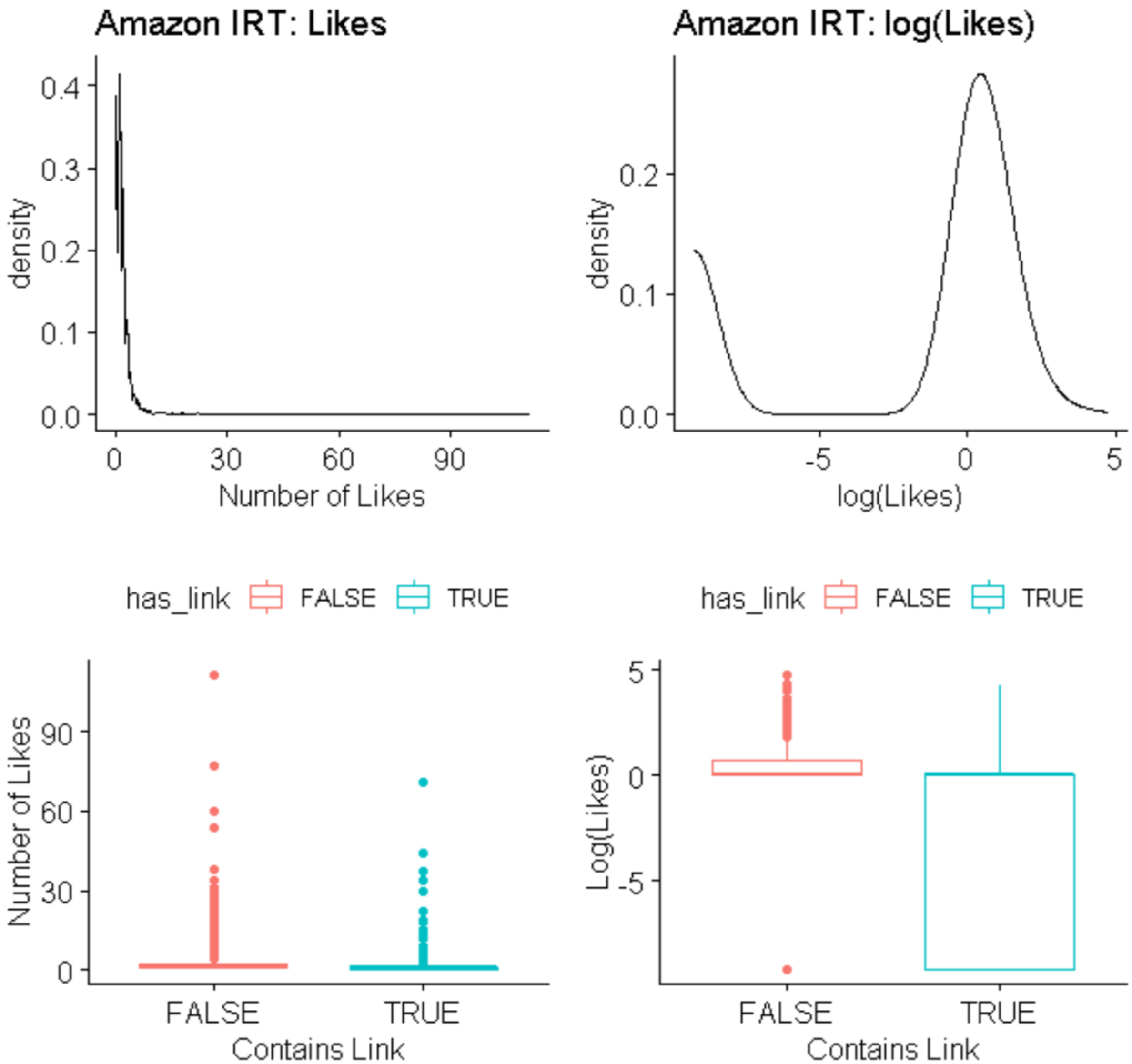
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of both populations are equal. Performing a shift function yields the following:

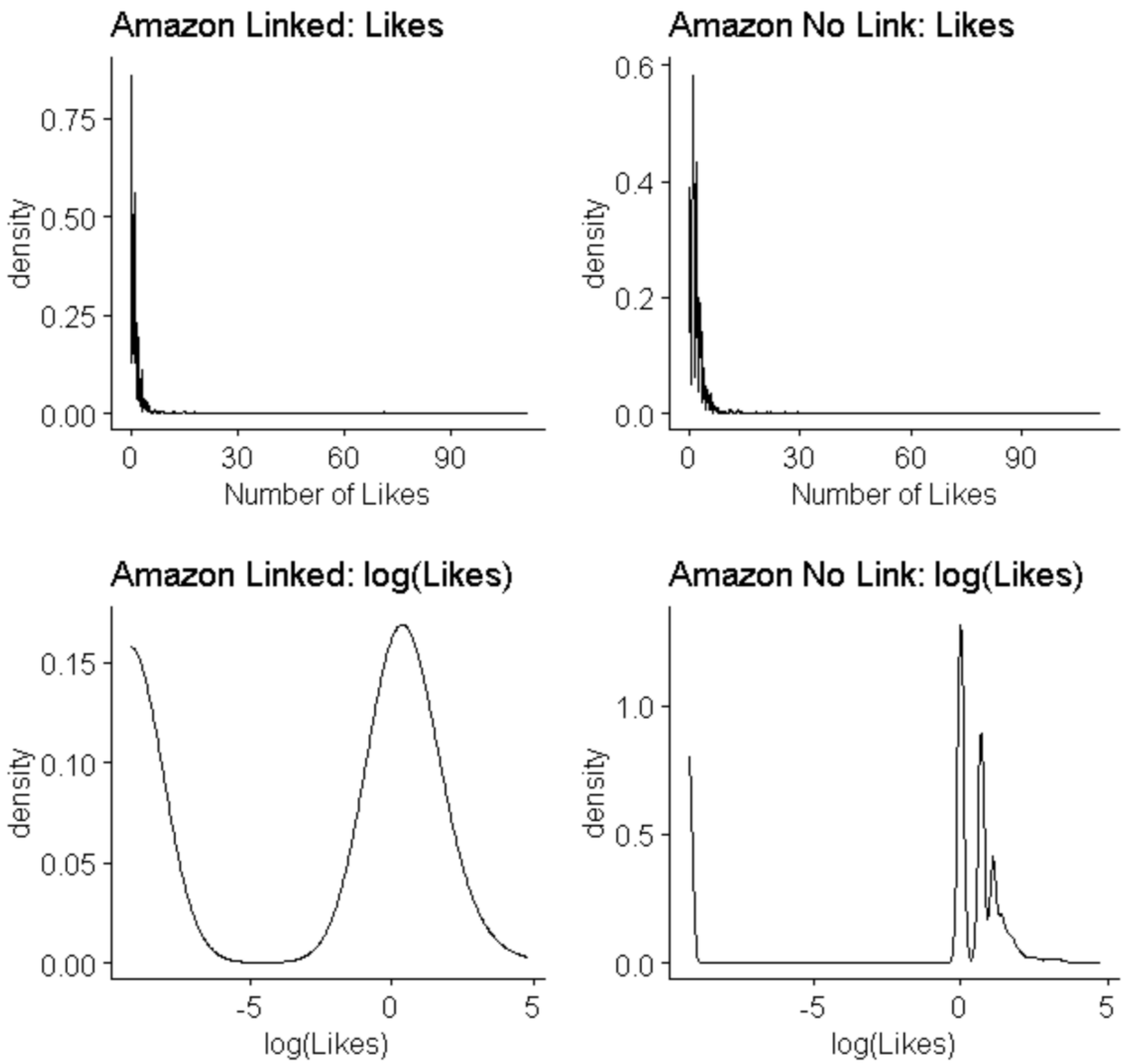



```
$`Group1 - Group2`
  q   Group1    Group2  difference   ci_lower  ci_upper      p_crit p_value
1 0.1 14.08351  11.25452   2.8289959   -9.475107  8.110818 0.005555556   0.386
2 0.2 18.51573  16.64774   1.8679894  -10.293939  9.387837 0.012500000   0.722
3 0.3 23.05320  23.32170  -0.2684947   -8.933698  9.230156 0.025000000   0.930
4 0.4 26.21925  28.59791  -2.3786563  -10.408910  9.928078 0.010000000   0.552
5 0.5 30.18199  32.01479  -1.8328033  -22.721380  7.366196 0.007142857   0.533
6 0.6 35.43684  35.92561  -0.4887755  -28.525380 10.815493 0.016666667   0.874
7 0.7 46.64981  45.59501   1.0548018  -29.820385 14.900905 0.050000000   0.960
8 0.8 57.49178  66.63696  -9.1451776  -92.597725 25.289176 0.008333333   0.597
9 0.9 85.77515 118.08556 -32.3104135 -146.368652 51.332518 0.006250000   0.493
```

Considering each confidence interval contains the value 0, we may not conclude any quantile differences in tweet performance between the two groups. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of retweets are equal to one another. **Inclusion of a link does not seem to have a statistically significant effect on the number of retweets which an Amazon official tweet receives.**

Amazon IRT: Number of Likes

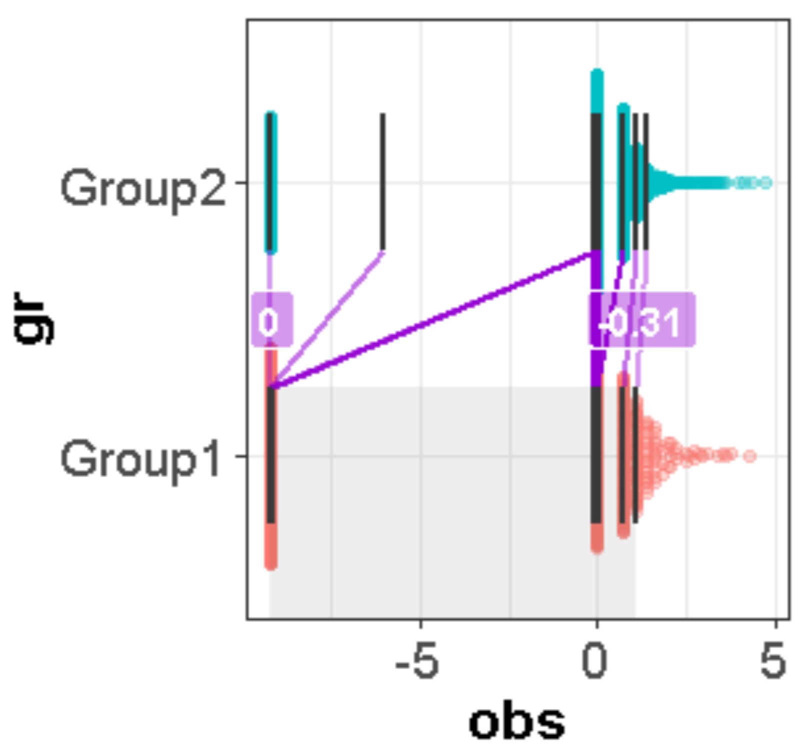




Visually, looks like Amazon IRT tweets containing links have a higher proportion of observations yielding 0 likes than Amazon IRT tweets not containing links. Furthermore, it looks as if the right tails may differ. Unsurprisingly, none of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 592082, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

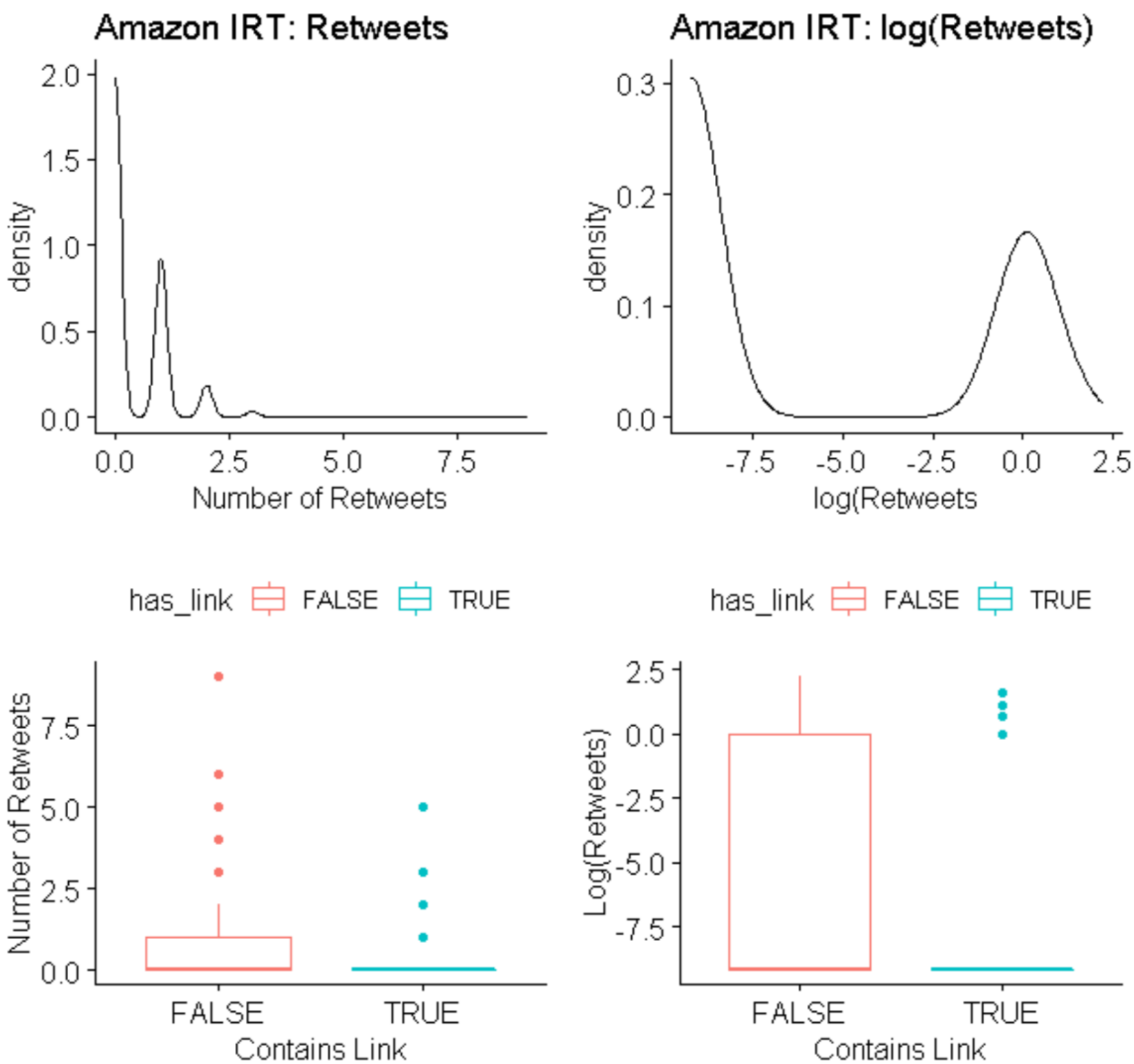
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



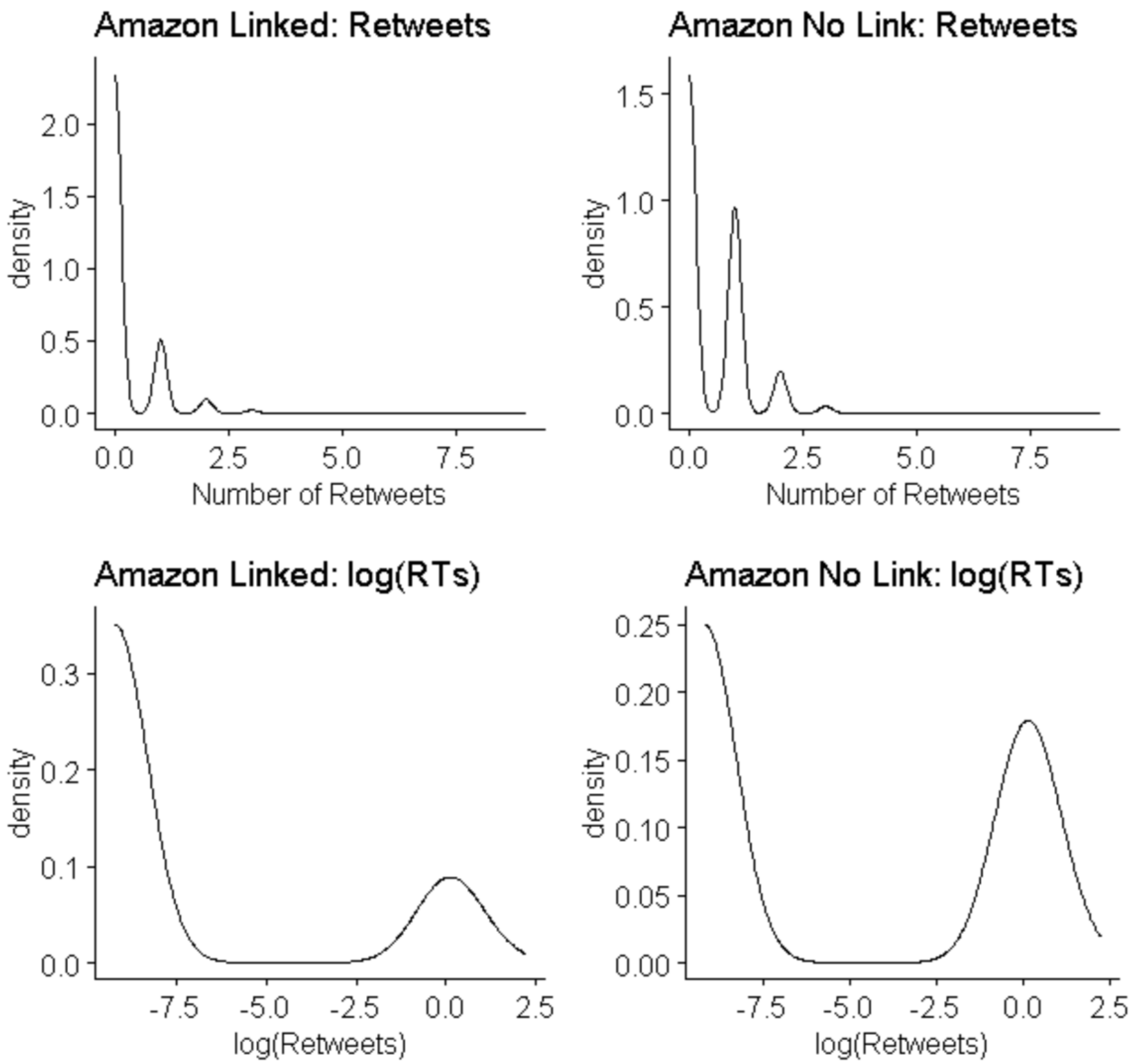

```
$`Group1 - Group2`
    q      Group1    Group2   difference   ci_lower      ci_upper      p_crit p_value
1 0.1 0.00000000 0.0000000  0.000000000  0.0000000  0.000000e+00 0.050000000       1
2 0.2 0.00000000 0.3421827 -0.342182656 -0.9623915 -4.689598e-03 0.025000000       0
3 0.3 0.00000000 1.0000000 -1.000000000 -1.0000000 -1.000000e+00 0.016666667       0
4 0.4 0.00154553 1.0000000 -0.998454470 -1.0000000 -7.202739e-01 0.012500000       0
5 0.5 0.99781840 1.0000286 -0.002210181 -0.3386421 -1.628988e-06 0.010000000       0
6 0.6 1.00000000 1.9999998 -0.999999825 -1.0000000 -9.960900e-01 0.008333333       0
7 0.7 1.00002240 2.0000000 -0.999977600 -1.0000000 -9.327979e-01 0.007142857       0
8 0.8 1.99587650 2.9970612 -1.001184675 -1.5352038 -4.868363e-01 0.006250000       0
9 0.9 2.93999894 3.9967130 -1.056714085 -1.8891901 -6.437519e-01 0.005555556       0
```

Judging by the confidence intervals, we can say with 95% confidence that the 2nd through 9th quantiles of group 2 (Amazon IRT tweets not containing links) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Amazon IRT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a link in Amazon IRT tweets.**

Amazon IRT: Number of Retweets

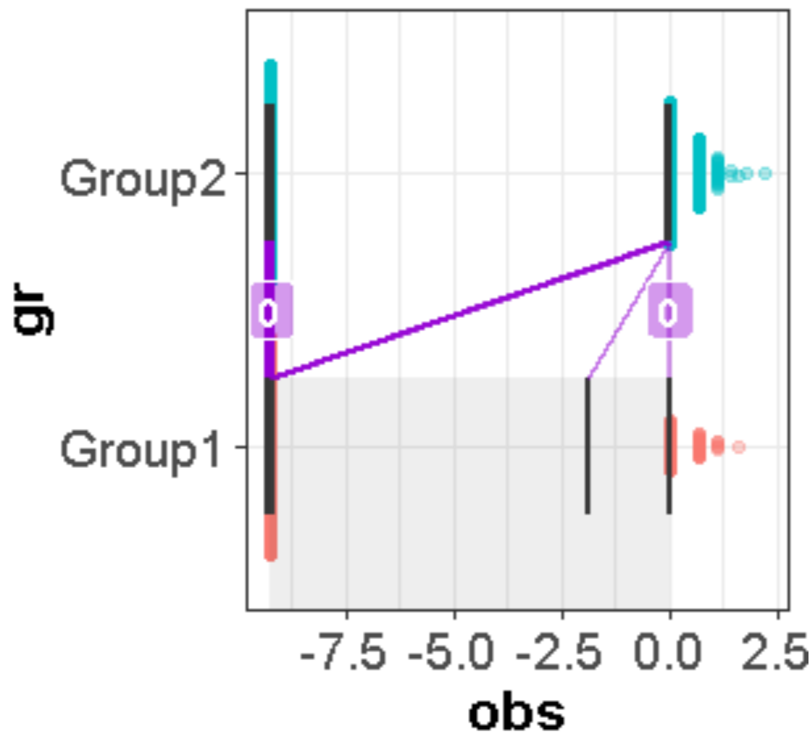


Visually, looks as if almost all Amazon IRT tweets containing links yield 0 retweets.




```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 666987, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

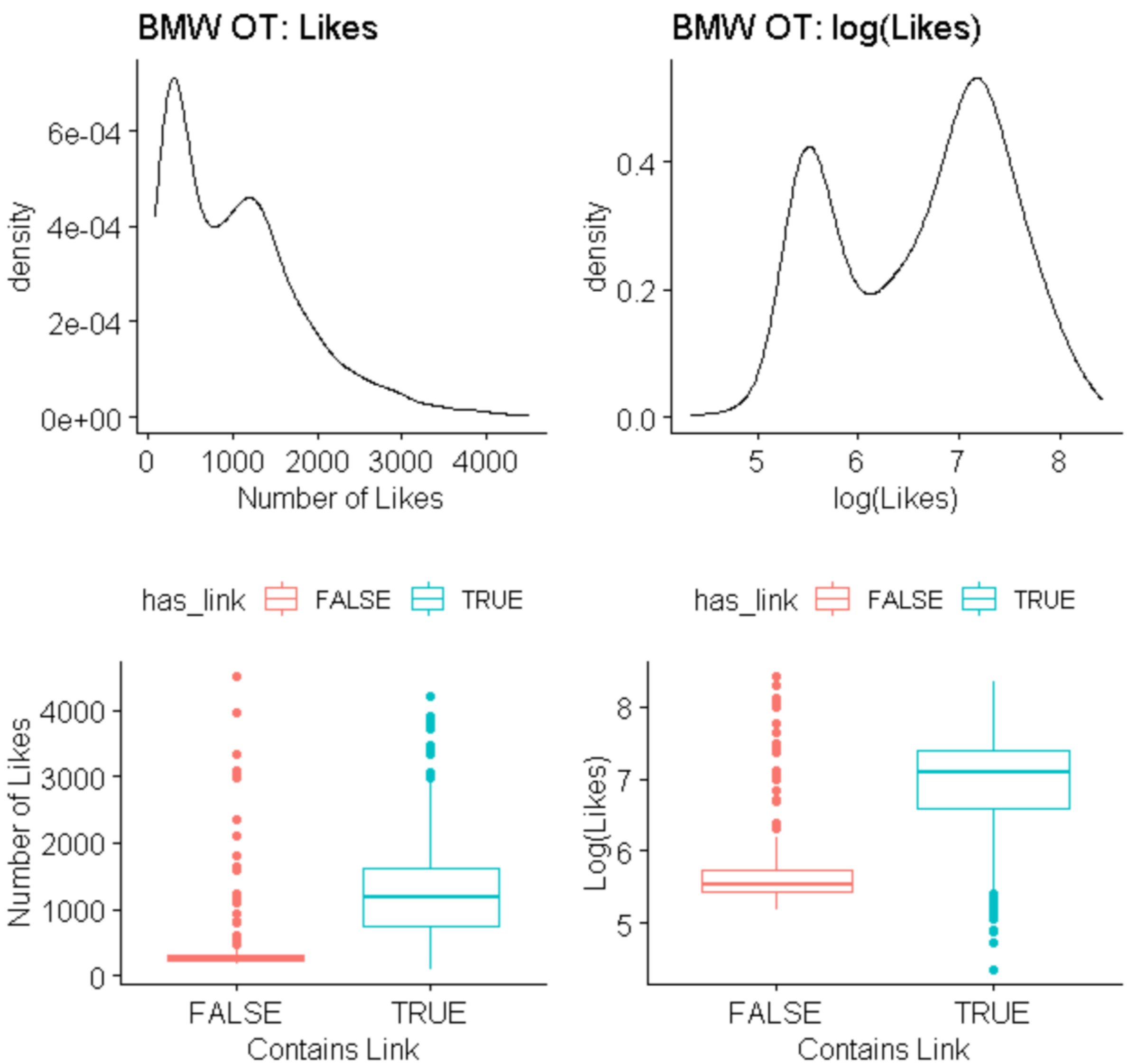
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

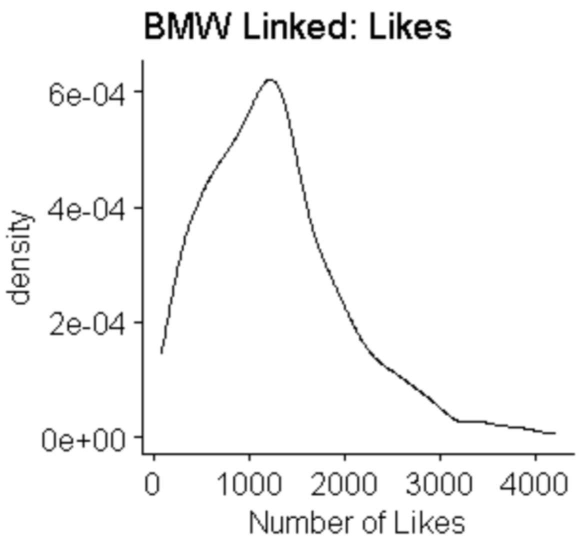


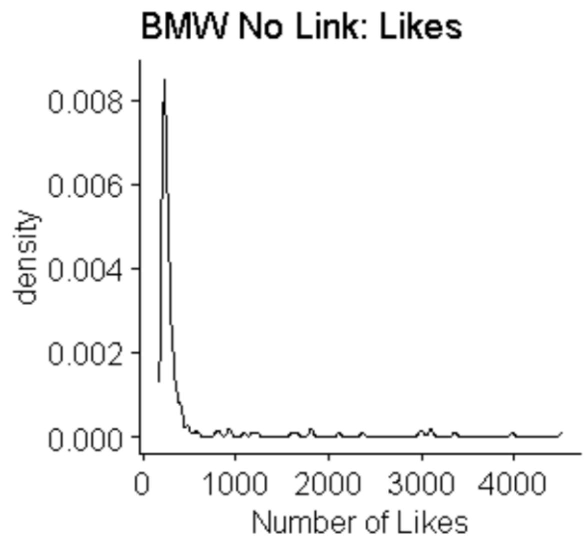

```
$`Group1 - Group2`
    q       Group1       Group2    difference      ci_lower      ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.050000000   1.000
2 0.2 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.025000000   1.000
3 0.3 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.016666667   1.000
4 0.4 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.012500000   1.000
5 0.5 0.000000e+00 1.513087e-10 -1.513087e-10 -8.566602e-05  0.000000e+00 0.010000000   0.027
6 0.6 0.000000e+00 9.977264e-01 -9.977264e-01 -1.000000e+00 -6.315669e-01 0.007142857   0.000
7 0.7 1.155095e-09 1.000000e+00 -1.000000e+00 -1.000000e+00 -9.994322e-01 0.006250000   0.000
8 0.8 7.989903e-01 1.000000e+00 -2.010097e-01 -9.770647e-01 -3.992748e-04 0.005555556   0.000
9 0.9 1.000000e+00 1.006443e+00 -6.442520e-03 -6.389794e-01 -1.149557e-07 0.008333333   0.002
```

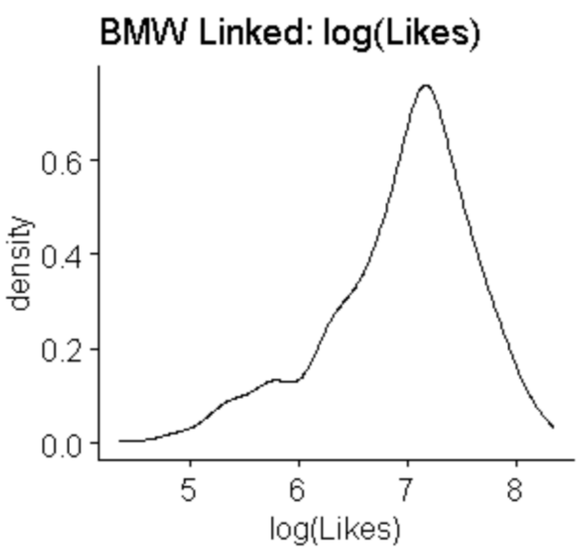
Judging by the confidence intervals, we can say with 95% confidence that the 6th through 9th quantiles of group 2 (Amazon IRT tweets not containing links) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Amazon IRT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets, and the right halves specifically), and potentially one underlying factor explaining these differences is the inclusion of a link in Amazon IRT tweets.**

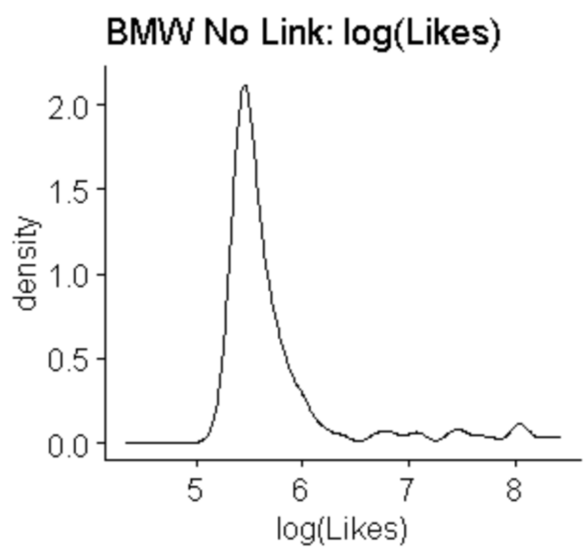
BMW Official: Number of Likes





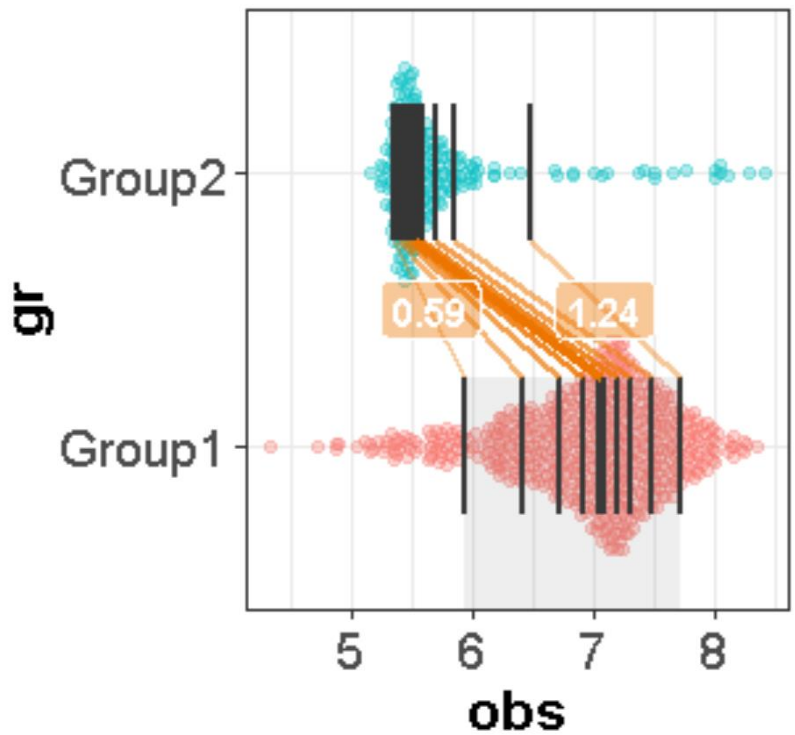







```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 107900, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

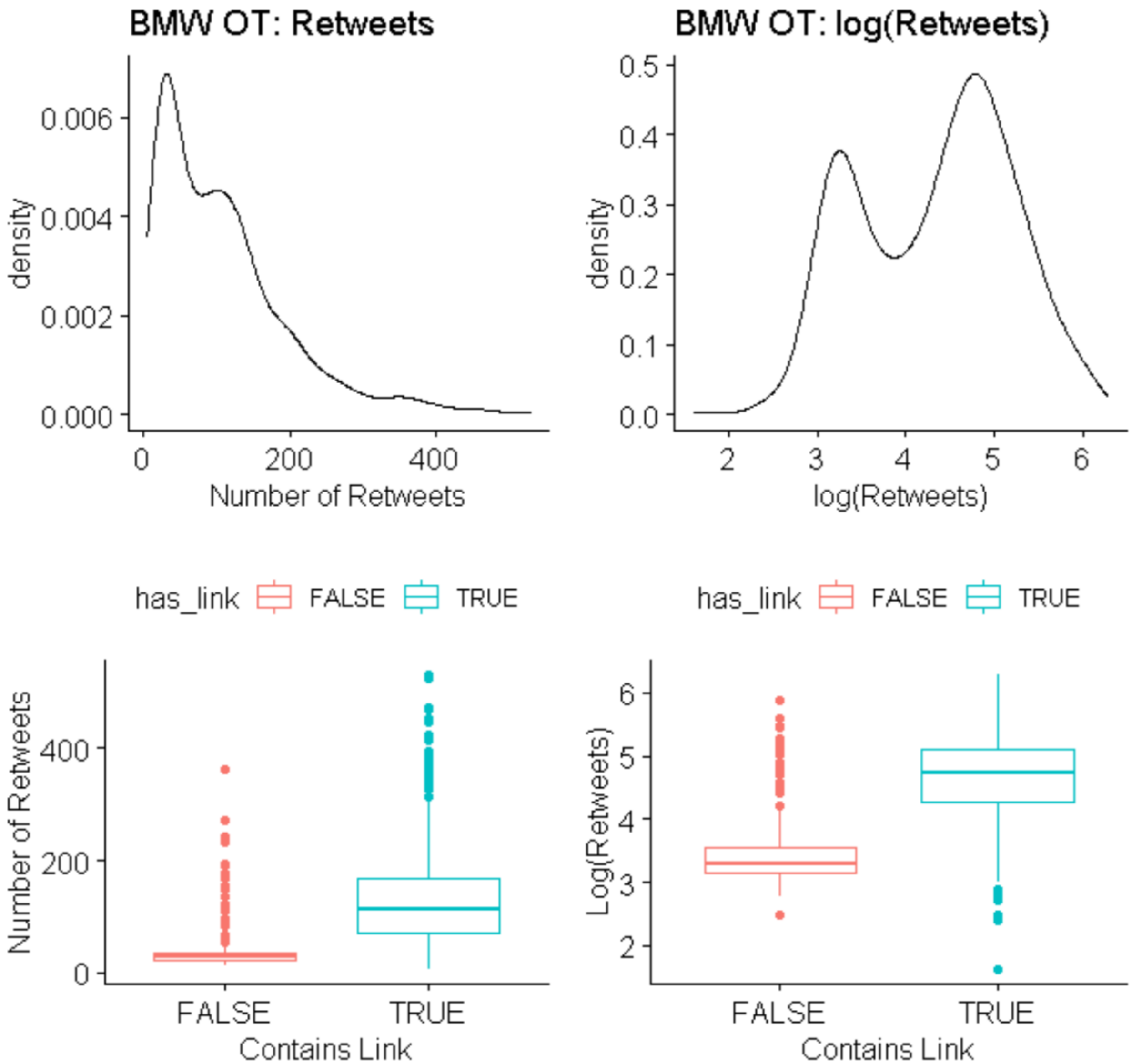


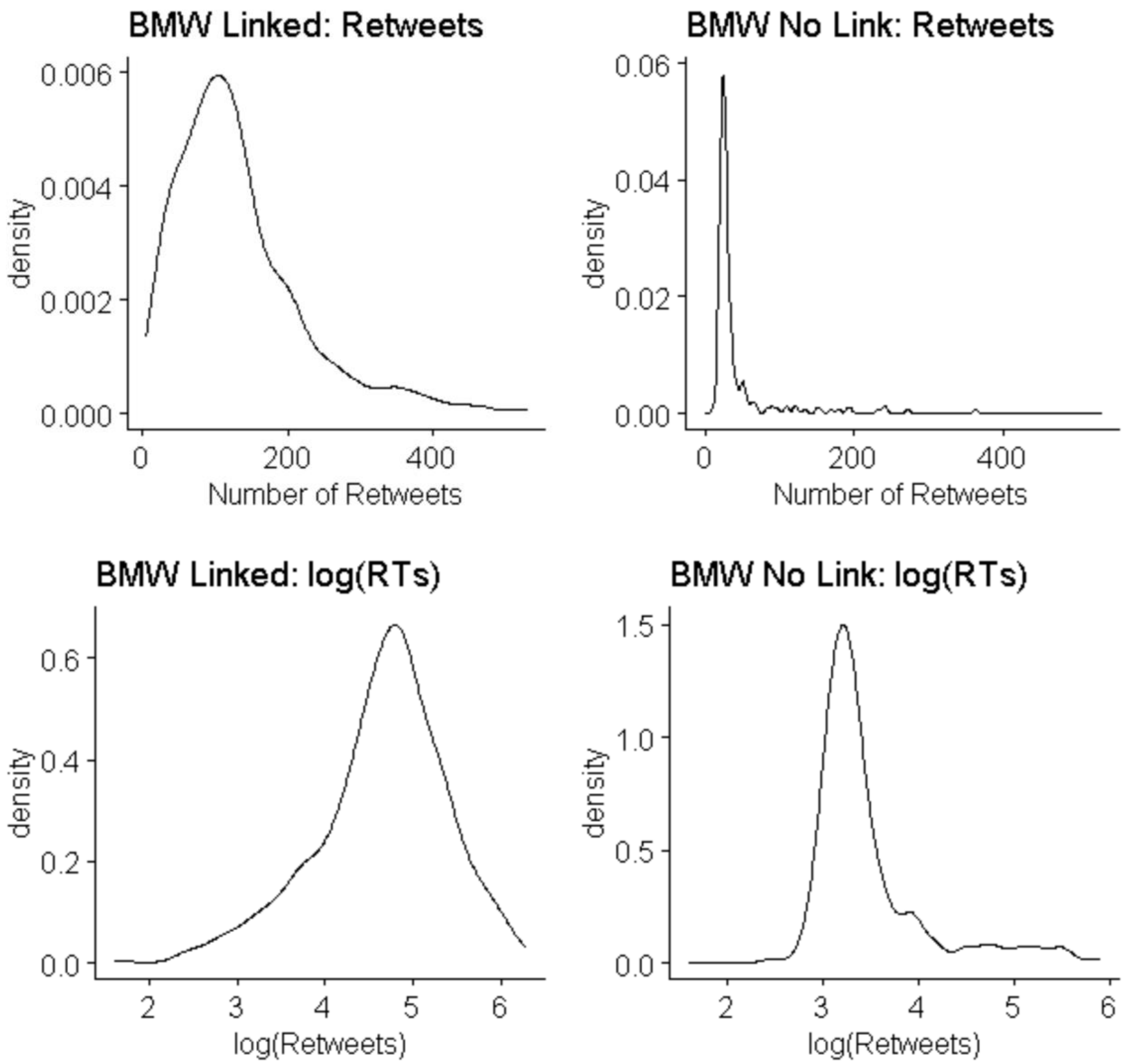
$`Group1 - Group2`

	q	Group1	Group2	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	386.0766	211.9932	174.0834	113.6116	259.8908	0.050000000	0
2	0.2	620.7758	221.5255	399.2503	331.5165	479.8915	0.025000000	0
3	0.3	830.3078	231.0865	599.2213	519.2971	683.3035	0.016666667	0
4	0.4	1019.0553	241.1312	777.9241	690.8800	863.1770	0.012500000	0
5	0.5	1187.4348	251.1989	936.2359	855.9810	1001.3014	0.010000000	0
6	0.6	1335.6774	270.1269	1065.5505	979.0295	1135.2630	0.008333333	0
7	0.7	1496.0910	296.8627	1199.2283	1098.6947	1334.5266	0.007142857	0
8	0.8	1788.8848	347.1739	1441.7109	1283.9145	1582.8427	0.006250000	0
9	0.9	2240.9671	694.5463	1546.4208	677.4410	1955.7307	0.005555556	0

With 95% confidence we can say that every quantile of group 2 (BMW OT tweets not containing links) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (BMW OT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a link in BMW OT tweets.**

BMW Official: Number of Retweets

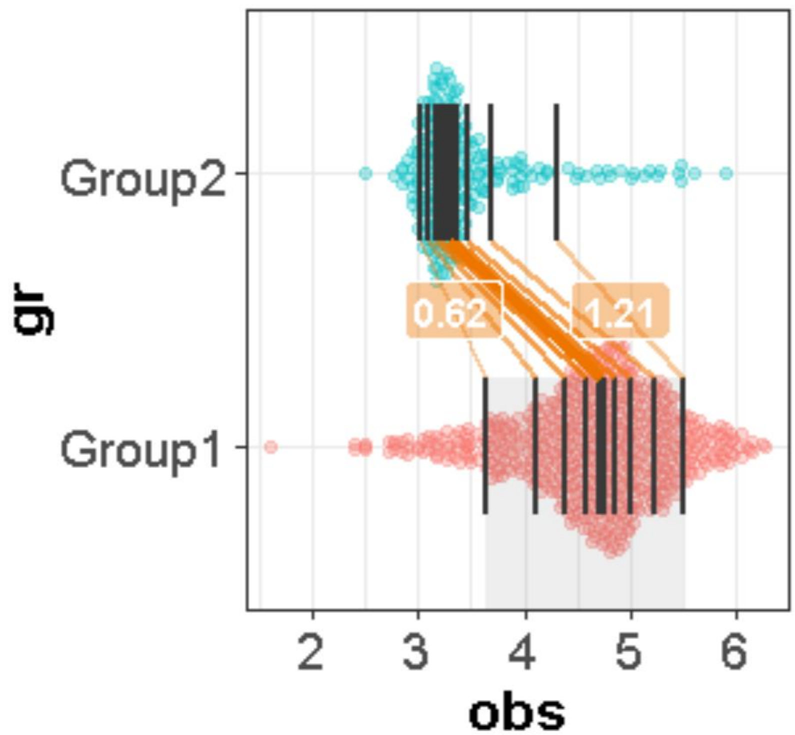




None of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 107430, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

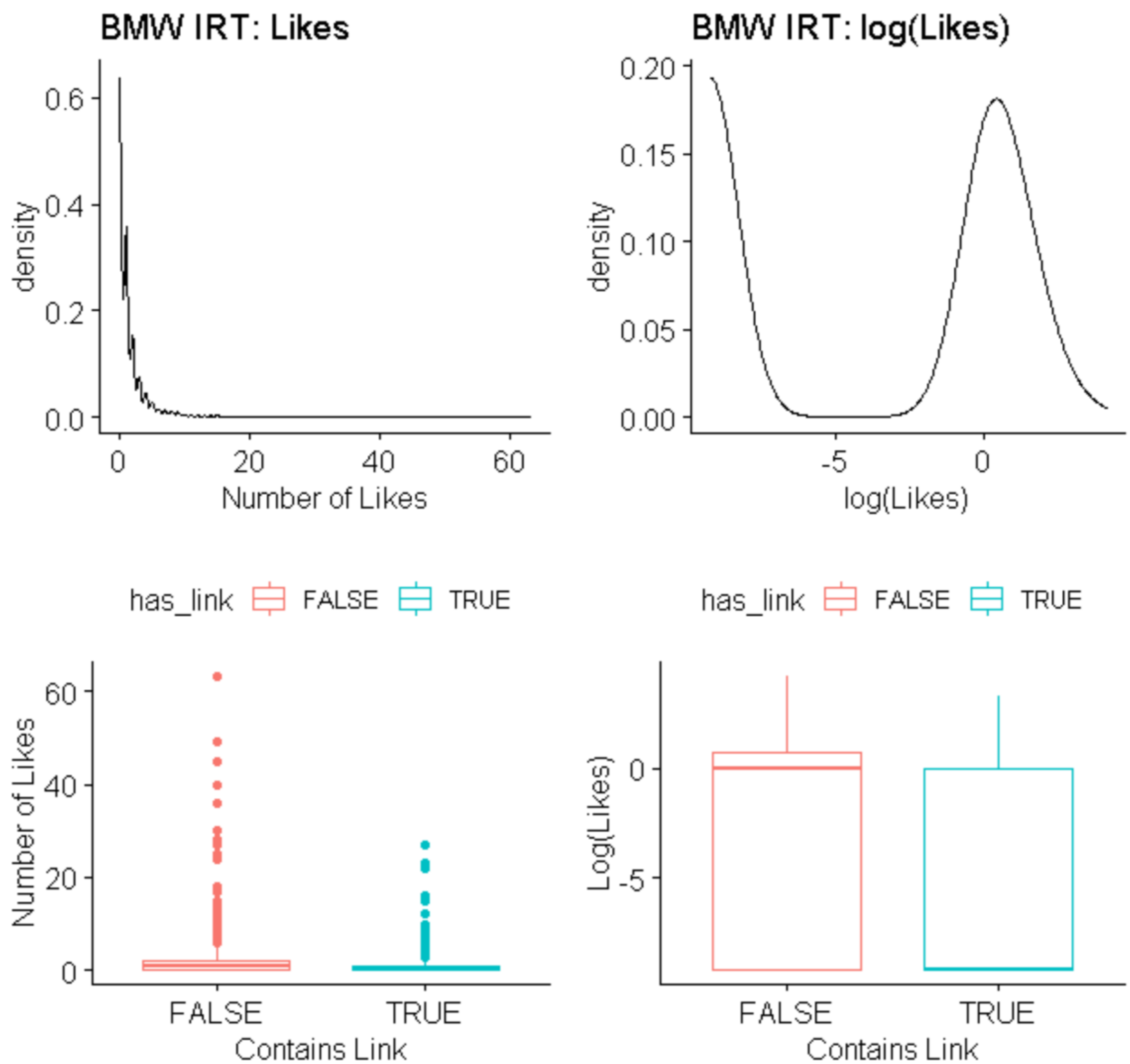
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

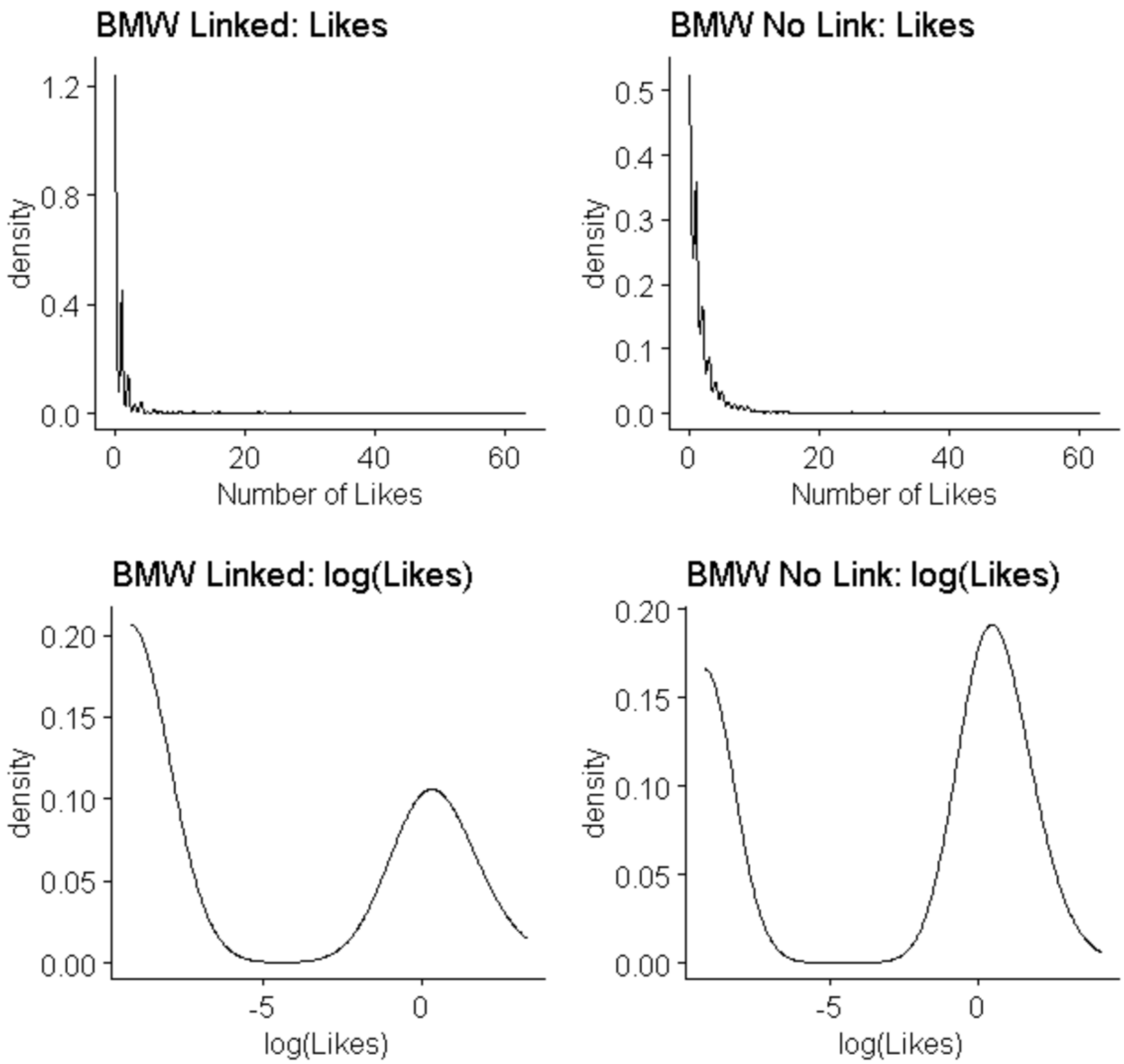



```
$`Group1 - Group2`
    q    Group1   Group2 difference  ci_lower  ci_upper      p_crit p_value
1 0.1  38.06661 20.38177   17.68484  12.16675  22.34625 0.050000000       0
2 0.2  60.53162 22.21710   38.31451  30.44920  46.21252 0.025000000       0
3 0.3  81.05291 23.63741   57.41550  49.43110  64.37101 0.016666667       0
4 0.4  97.35104 25.11866   72.23239  63.97297  79.32422 0.012500000       0
5 0.5 113.63186 26.80721   86.82466  77.64022  95.50846 0.010000000       0
6 0.6 130.00458 28.68006  101.32452  91.38023 110.52201 0.008333333       0
7 0.7 150.53029 31.86760  118.66270 105.90546 134.20796 0.007142857       0
8 0.8 188.61021 40.31699  148.29322 123.87732 166.12717 0.006250000       0
9 0.9 246.34675 76.13202  170.21474 104.89435 213.79653 0.005555556       0
```

With 95% confidence we can say that every quantile of group 2 (BMW OT tweets not containing links) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (BMW OT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets), and potentially one underlying factor explaining these differences is the inclusion of a link in BMW OT tweets.**

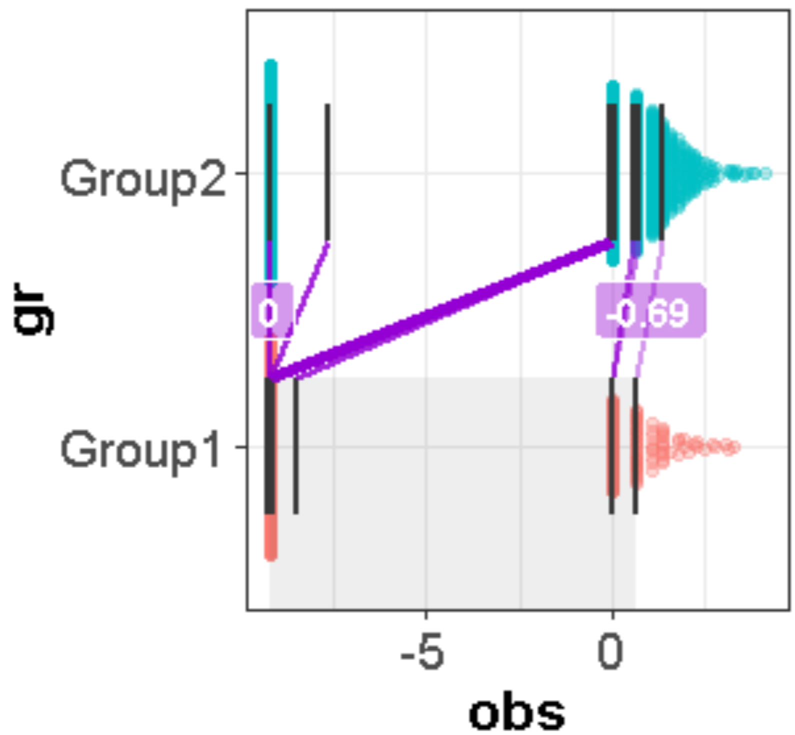
BMW IRT: Number of Likes






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 327793, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

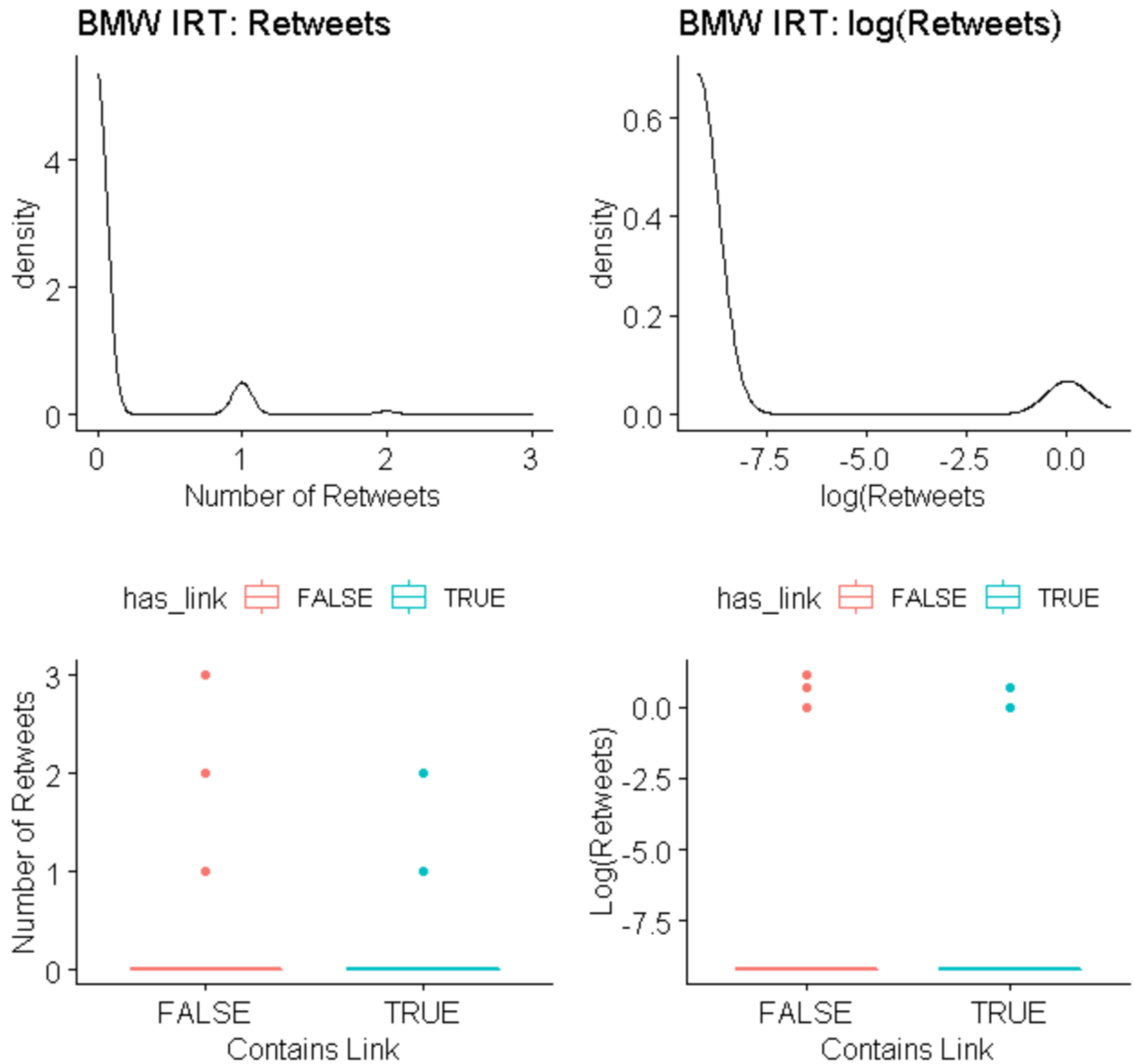


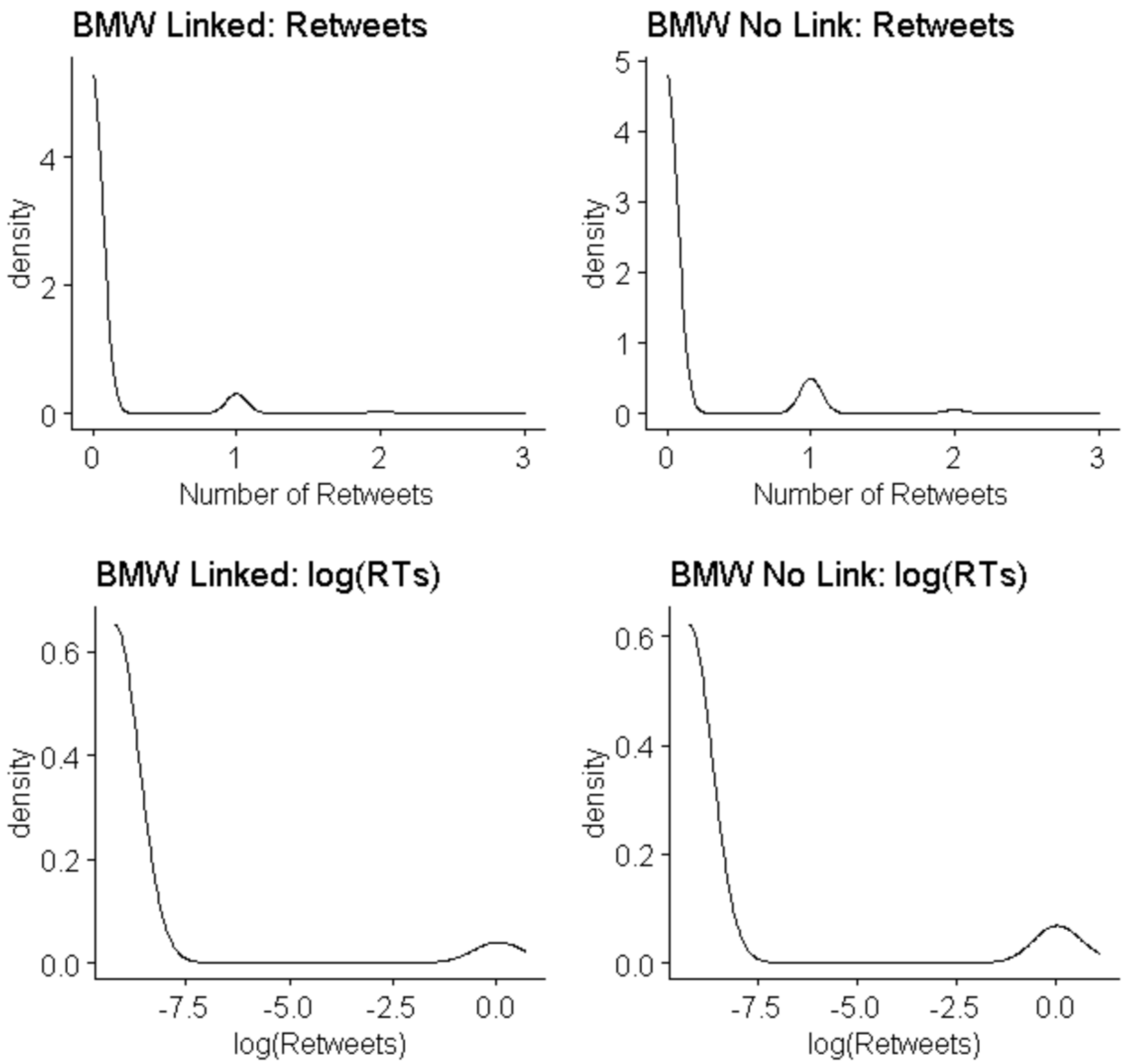
$`Group1 - Group2`

	q	Group1	Group2	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	0.000000e+00	0.0000000	0.0000000	0.000000e+00	0.0000000000	0.050000000	1.000
2	0.2	0.000000e+00	0.0000000	0.0000000	0.000000e+00	0.0000000000	0.025000000	1.000
3	0.3	0.000000e+00	0.0000000	0.0000000	-6.938894e-14	0.0000000000	0.016666667	0.921
4	0.4	0.000000e+00	0.1680928	-0.1680928	-9.292904e-01	-0.0001911919	0.012500000	0.000
5	0.5	1.080136e-09	1.0000000	-1.0000000	-1.000000e+00	-0.9996445082	0.010000000	0.000
6	0.6	7.271542e-02	1.0000000	-0.9272846	-9.999830e-01	-0.1320263839	0.008333333	0.000
7	0.7	9.993477e-01	1.8984814	-0.8991337	-1.160709e+00	-0.0927675833	0.007142857	0.000
8	0.8	1.000462e+00	2.1209258	-1.1204642	-1.924625e+00	-0.8229836069	0.006250000	0.000
9	0.9	2.012294e+00	4.0153456	-2.0030513	-2.702316e+00	-1.1668266384	0.005555556	0.000

We can say with 95% confidence that the 4th through 9th quantiles of group 2 (BMW IRT tweets not containing links) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (BMW IRT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a link in BMW IRT tweets.**

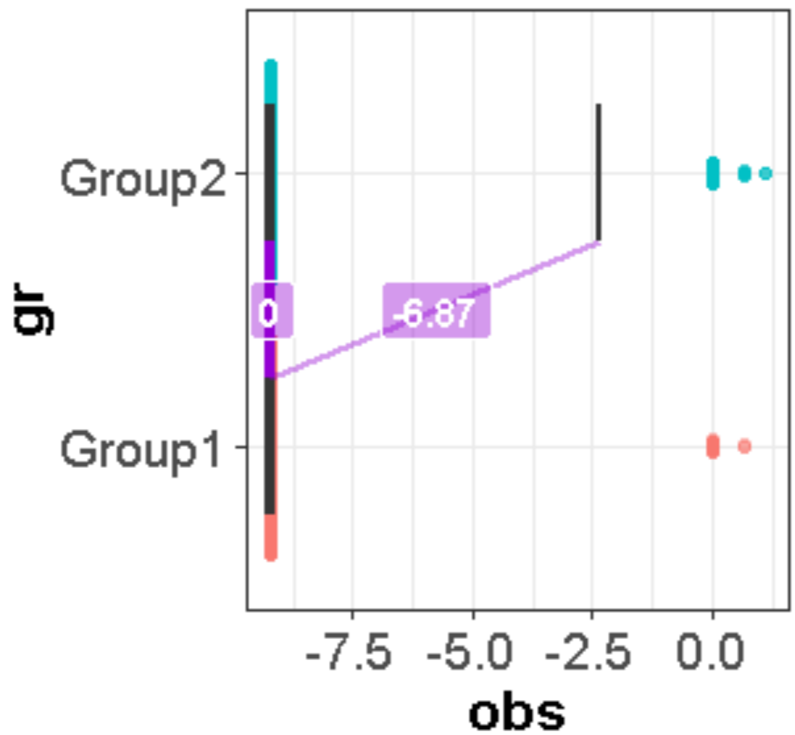
BMW IRT: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 419867, p-value = 0.002906
alternative hypothesis: true location shift is not equal to 0
```

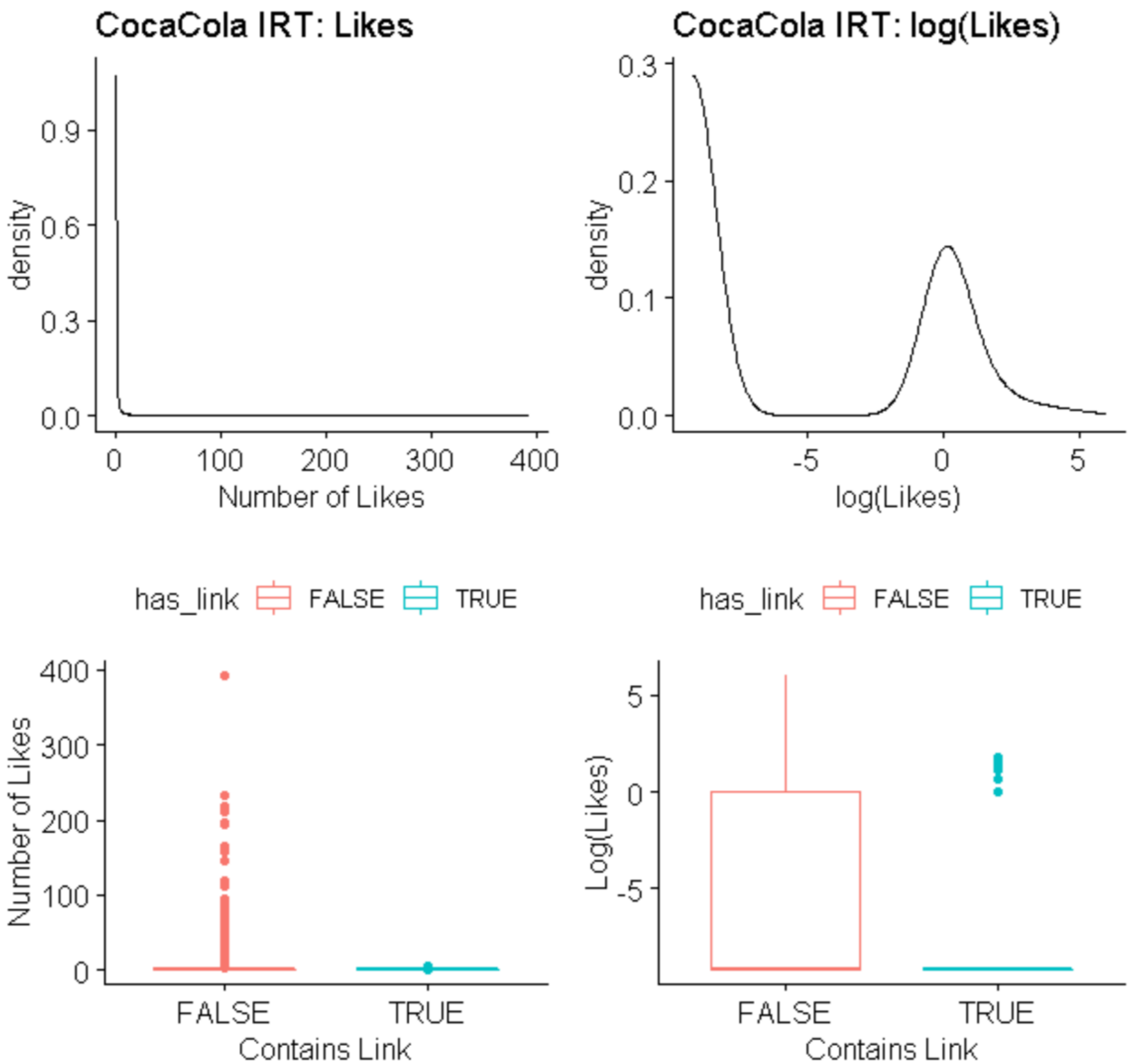
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

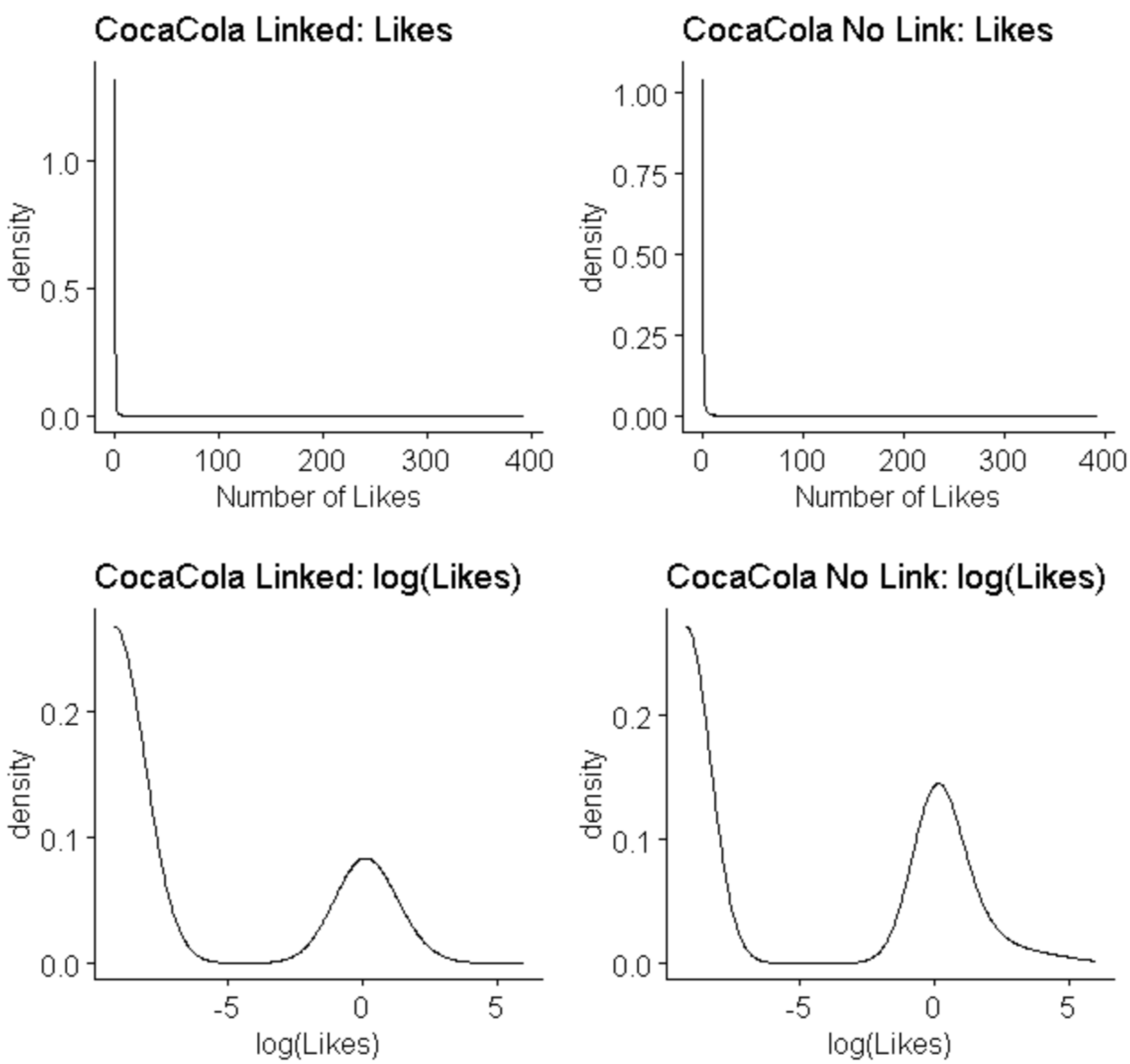



```
$`Group1 - Group2`
    q       Group1   Group2 difference    ci_lower     ci_upper      p_crit p_value
1 0.1 0.0000000000 0.000000  0.0000000   0.0000000 0.000000e+00 0.050000000  1.0000
2 0.2 0.0000000000 0.000000  0.0000000   0.0000000 0.000000e+00 0.025000000  1.0000
3 0.3 0.0000000000 0.000000  0.0000000   0.0000000 0.000000e+00 0.016666667  1.0000
4 0.4 0.0000000000 0.000000  0.0000000   0.0000000 0.000000e+00 0.012500000  1.0000
5 0.5 0.0000000000 0.000000  0.0000000   0.0000000 0.000000e+00 0.010000000  1.0000
6 0.6 0.0000000000 0.000000  0.0000000   0.0000000 0.000000e+00 0.008333333  1.0000
7 0.7 0.0000000000 0.000000  0.0000000   0.0000000 0.000000e+00 0.007142857  1.0000
8 0.8 0.0000000000 0.000000  0.0000000   0.0000000 5.218048e-14 0.006250000  0.9715
9 0.9 0.0003764862 0.746183 -0.7458065  -0.9989367 2.218448e-03 0.005555556  0.0100
```

Considering each confidence interval contains, or is, the value 0, we may not conclude any quantile differences in tweet performance between the two groups. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of retweets are equal to one another. **Inclusion of a link does not seem to have a statistically significant effect on the number of retweets which a BMW IRT tweet receives.**

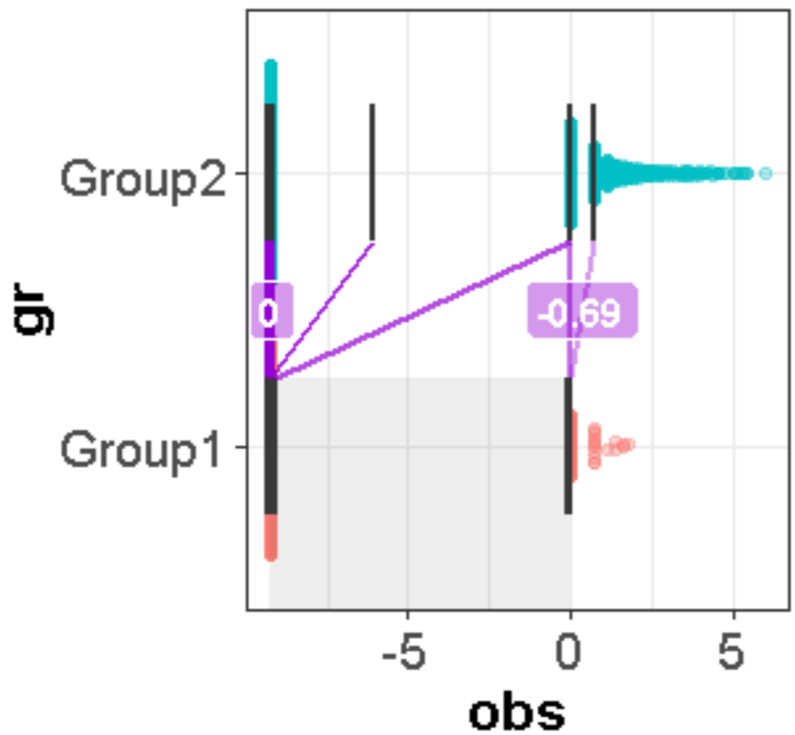
Coca Cola IRT: Number of Likes






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 438963, p-value = 1.85e-09
alternative hypothesis: true location shift is not equal to 0
```

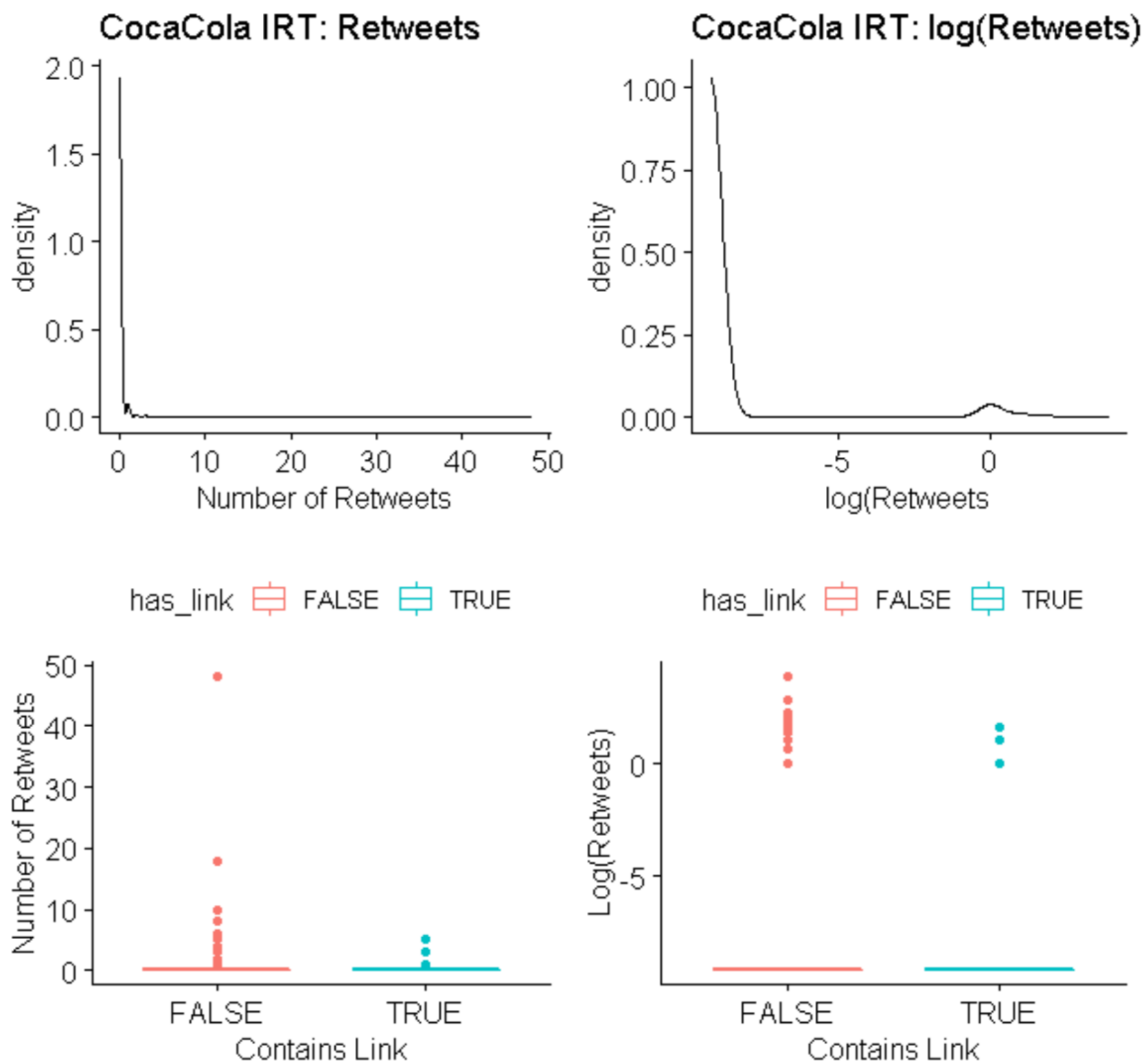
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

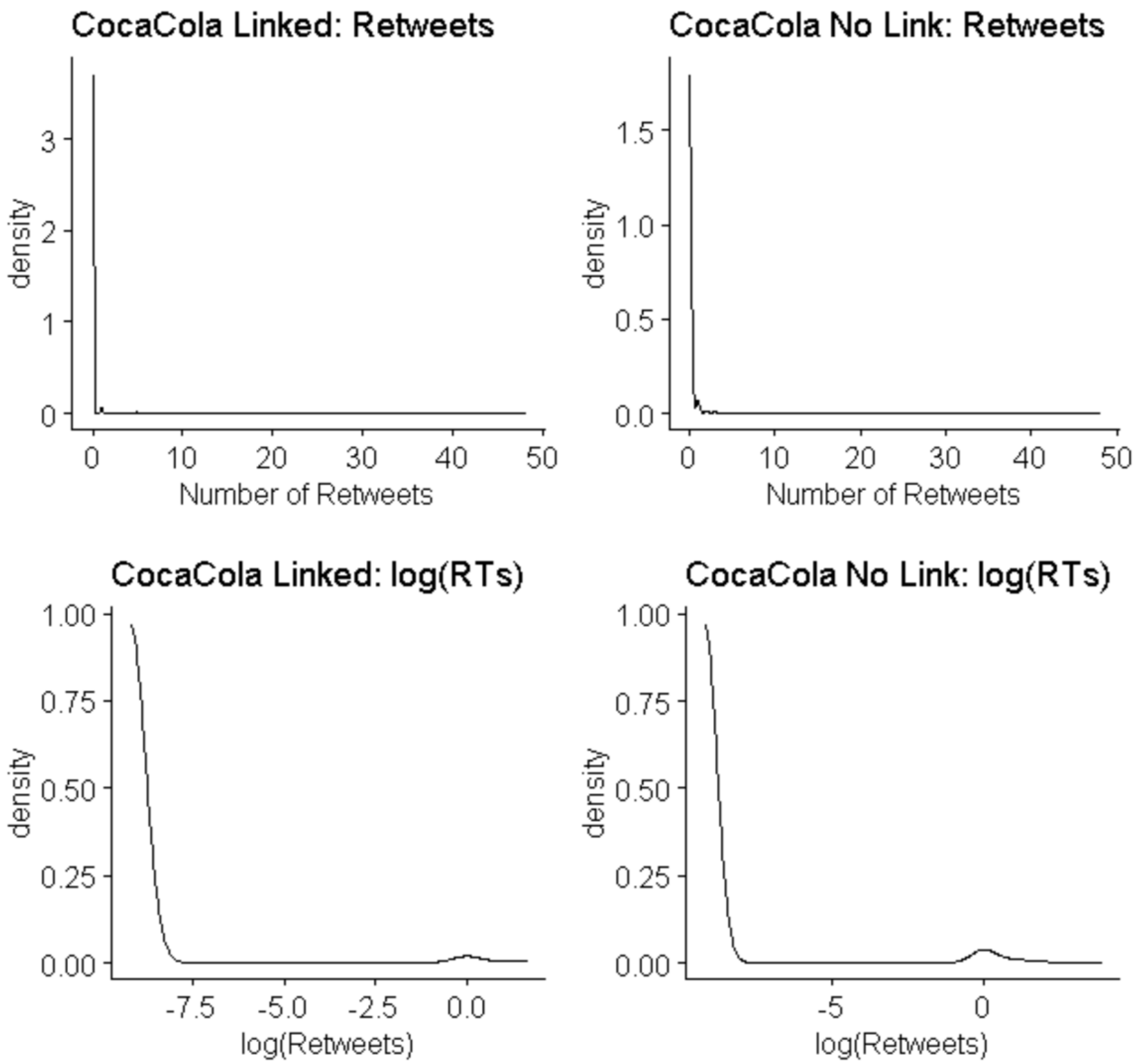



```
$`Group1 - Group2`
    q       Group1    Group2  difference   ci_lower     ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.00000000  0.0000000  0.000000e+00 0.050000000  1.0000
2 0.2 0.000000e+00 0.0000000  0.00000000  0.0000000  0.000000e+00 0.025000000  1.0000
3 0.3 0.000000e+00 0.0000000  0.00000000  0.0000000  0.000000e+00 0.016666667  1.0000
4 0.4 0.000000e+00 0.0000000  0.00000000  0.0000000  0.000000e+00 0.012500000  1.0000
5 0.5 0.000000e+00 0.0000000  0.00000000  0.0000000  7.882583e-15 0.010000000  0.9845
6 0.6 5.152367e-11 0.3392944 -0.33929441 -0.9813506 -1.965307e-03 0.008333333  0.0000
7 0.7 1.163476e-02 1.0000000 -0.98836524 -0.9999997 -3.418634e-01 0.007142857  0.0000
8 0.8 9.871627e-01 1.0000000 -0.01283735 -0.6426973 -7.988904e-07 0.006250000  0.0000
9 0.9 1.000006e+00 2.0000046 -0.99999856 -1.0463229 -9.086893e-01 0.005555556  0.0000
```

From the confidence intervals we can say, with 95% confidence, that the 6^{th} through 9^{th} quantiles of group 2 (Coca Cola IRT tweets not containing links) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Coca Cola IRT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes, and the right tails specifically), and potentially one underlying factor explaining these differences is the inclusion of a link in Coca Cola IRT tweets.**

Coca Cola IRT: Number of Retweets

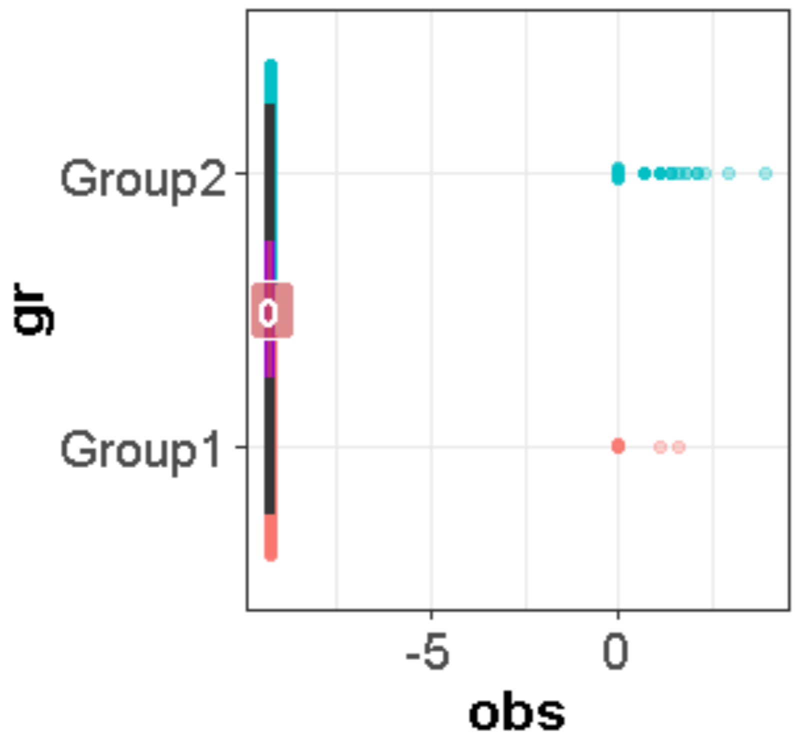




I do not know why I cannot get the x-axes to match between the two visualizations directly above, but note the differences.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 510948, p-value = 0.01973
alternative hypothesis: true location shift is not equal to 0
```

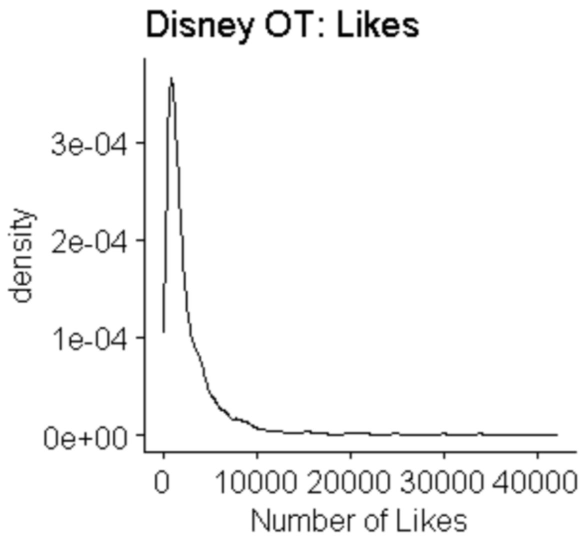
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



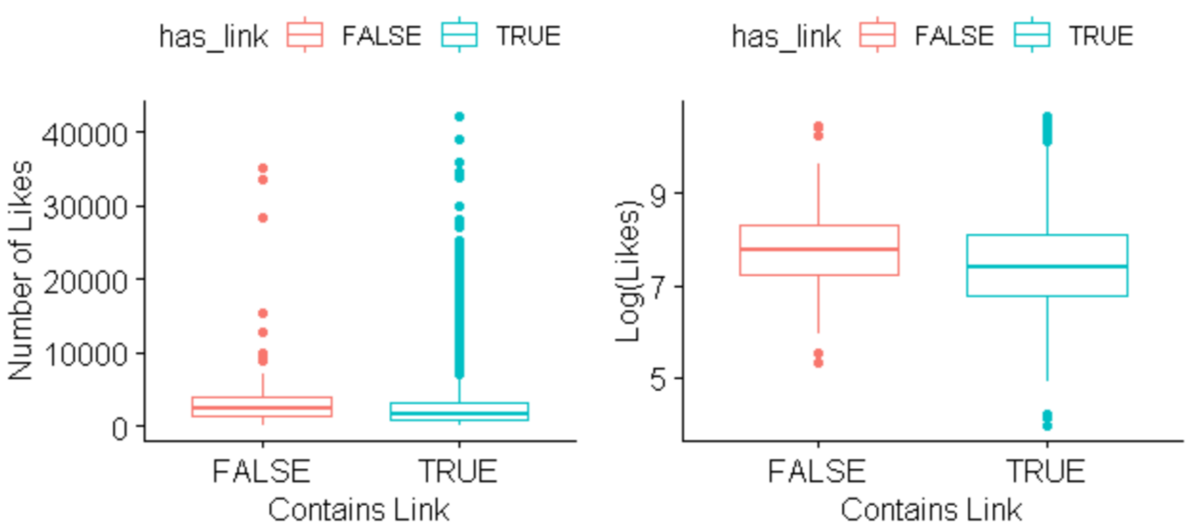

```
$`Group1 - Group2`
    q       Group1 Group2   difference     ci_lower     ci_upper      p_crit p_value
1 0.1 0.000000e+00      0 0.000000e+00  0.00000e+00 0.000000e+00 0.050000000   1.000
2 0.2 0.000000e+00      0 0.000000e+00  0.00000e+00 0.000000e+00 0.025000000   1.000
3 0.3 0.000000e+00      0 0.000000e+00  0.00000e+00 0.000000e+00 0.016666667   1.000
4 0.4 0.000000e+00      0 0.000000e+00  0.00000e+00 0.000000e+00 0.012500000   1.000
5 0.5 0.000000e+00      0 0.000000e+00  0.00000e+00 0.000000e+00 0.010000000   1.000
6 0.6 0.000000e+00      0 0.000000e+00  0.00000e+00 0.000000e+00 0.008333333   1.000
7 0.7 0.000000e+00      0 0.000000e+00  0.00000e+00 0.000000e+00 0.007142857   1.000
8 0.8 0.000000e+00      0 0.000000e+00  0.00000e+00 0.000000e+00 0.006250000   1.000
9 0.9 3.268497e-13      0 3.268497e-13 -8.21565e-15 2.119059e-05 0.005555556   0.224
```

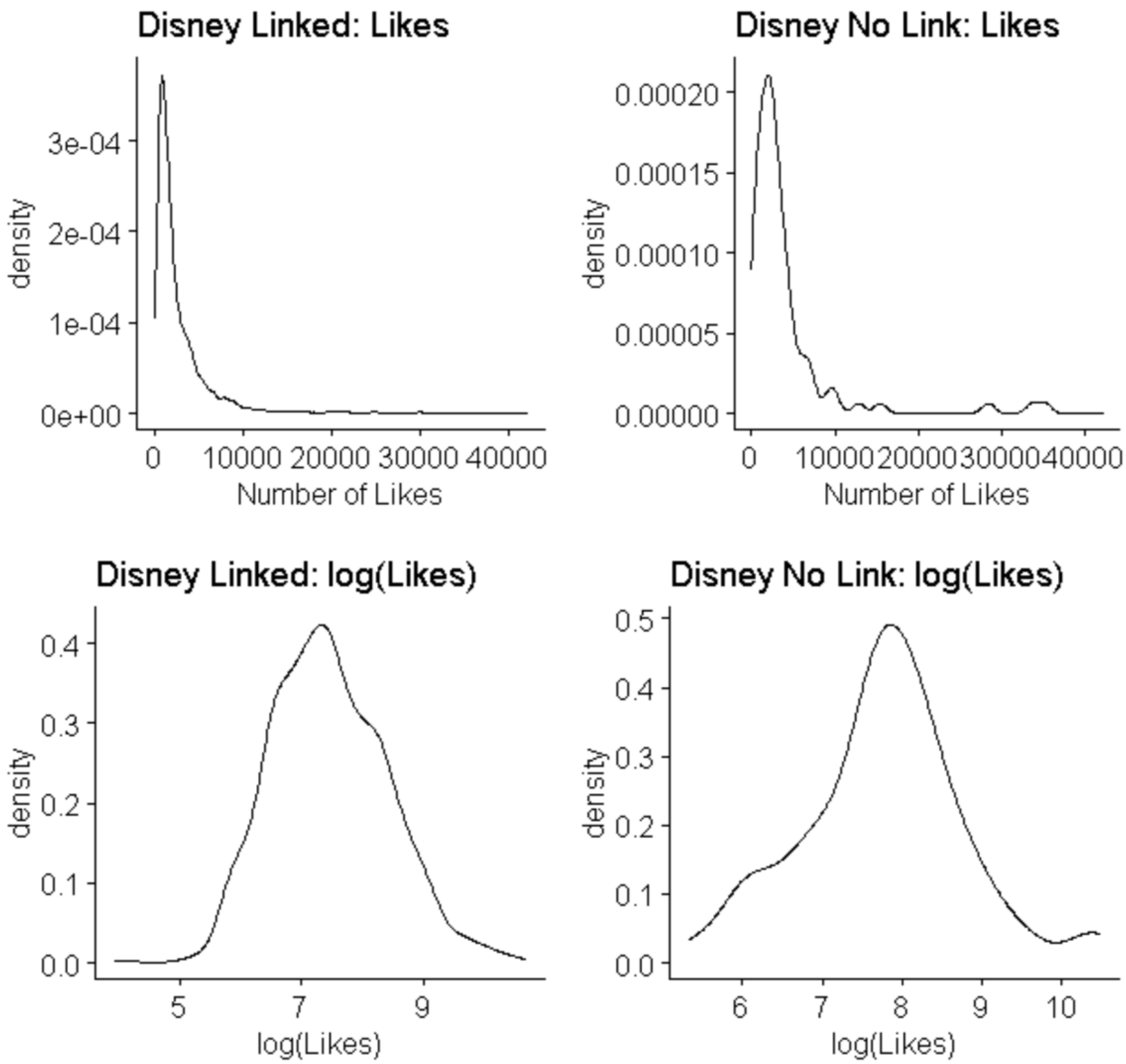
However, from the confidence intervals we can see that no significant differences between the quantiles of the two groups exist. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of retweets are equal to one another. **Inclusion of a link does not seem to have a statistically significant effect on the number of retweets which a Coca Cola IRT tweet receives.**

Disney Official: Number of Likes



While the log distribution may appear normal, it does not pass a Shapiro-Wilk normality test.

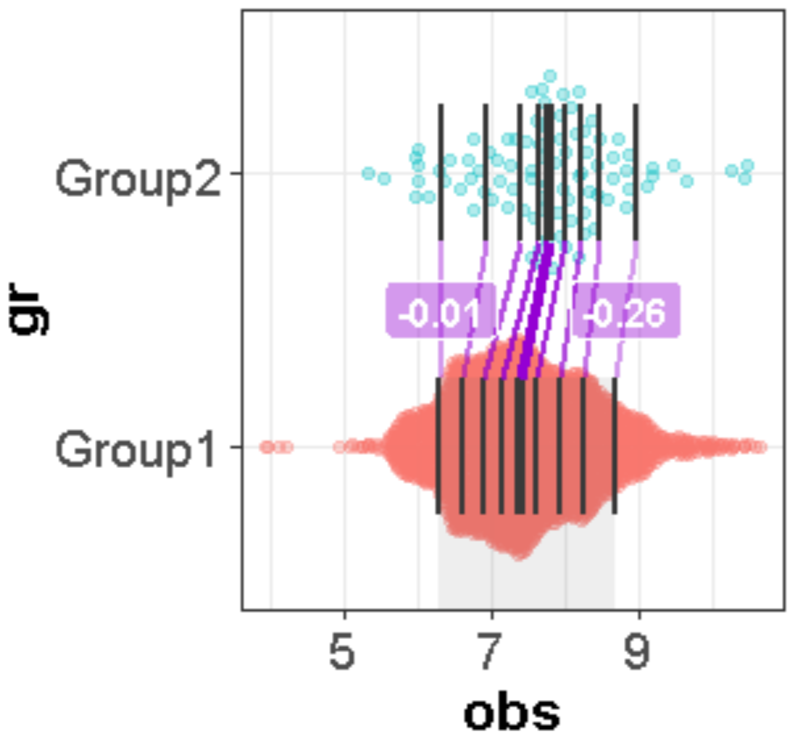




Again, I'm having trouble getting the x-axes to align between the two plots above, but note the differences. Furthermore, only the bottom right distribution passes a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 95236, p-value = 0.00115
alternative hypothesis: true location shift is not equal to 0
```

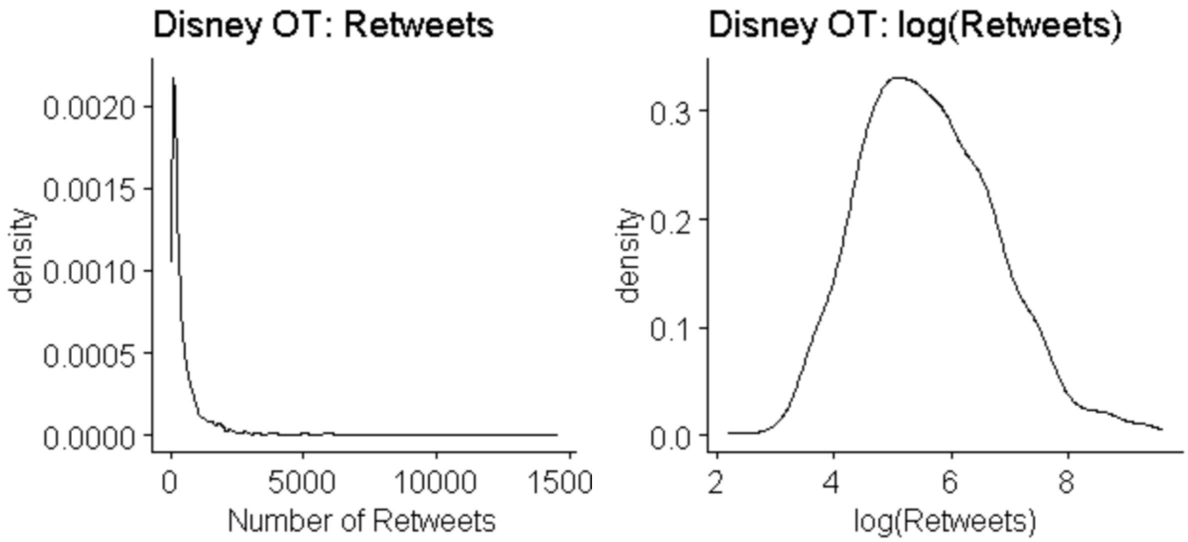
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



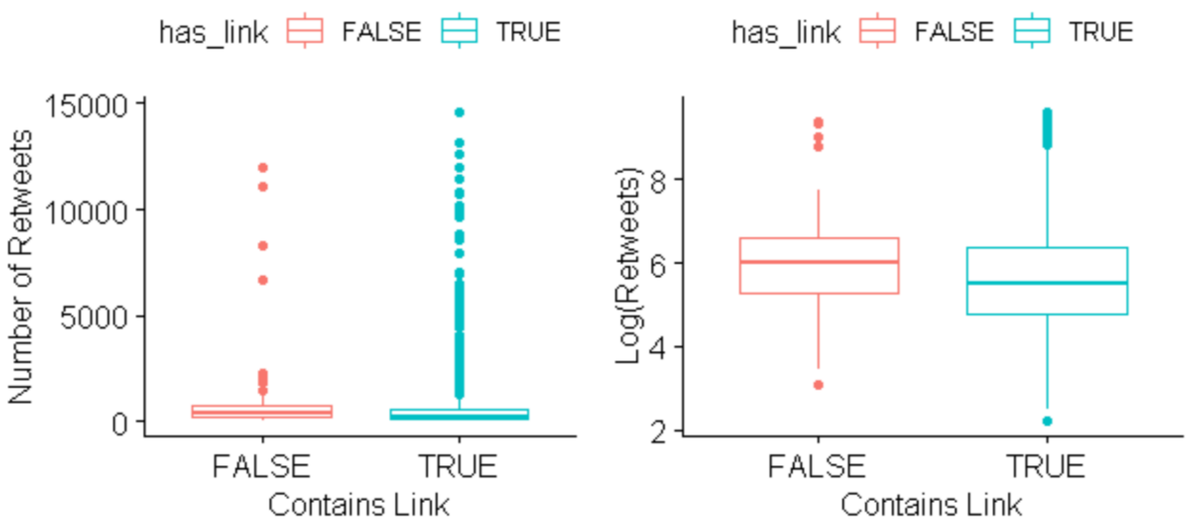

```
$`Group1 - Group2`
    q    Group1    Group2   difference    ci_lower    ci_upper      p_crit p_value
1 0.1  551.0306  575.6454    -24.61483   -343.5085   131.92948 0.050000000   0.776
2 0.2  750.1101 1055.8133   -305.70318   -961.1617    93.13302 0.012500000   0.059
3 0.3  994.0098 1636.5283   -642.51856  -1171.4128   -43.78663 0.010000000   0.005
4 0.4 1278.9038 2085.0215   -806.11768  -1370.9753 -226.84535 0.007142857   0.000
5 0.5 1626.8677 2448.2374   -821.36965  -1582.4745 -304.93665 0.006250000   0.000
6 0.6 2055.7025 2978.1925   -922.49002  -1823.1418 -249.82861 0.005555556   0.000
7 0.7 2789.1769 3692.6871   -903.51015  -2126.5248   -36.98568 0.008333333   0.006
8 0.8 3899.9892 4833.0179   -933.02867  -3136.3935   255.37007 0.016666667   0.084
9 0.9 6013.3768 8009.0073 -1995.63052 -10179.7451   803.20765 0.025000000   0.143
```

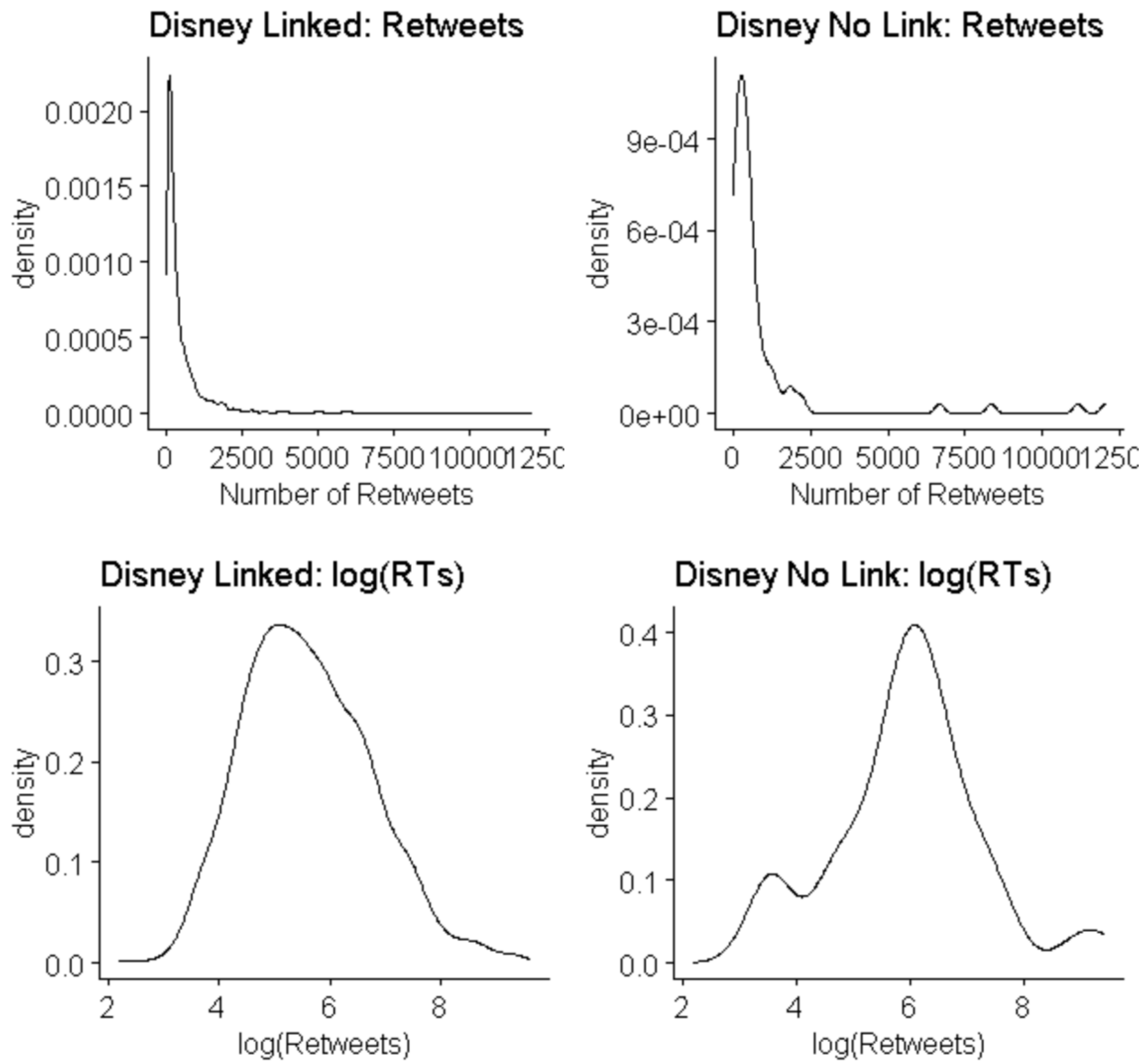
From the confidence intervals we can say, with 95% confidence, that the 3rd through 7th quantiles of group 2 (Disney OT tweets not containing links) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Disney OT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes, and the mid quantiles specifically), and potentially one underlying factor explaining these differences is the inclusion of a link in Disney official tweets.**

Disney Official: Number of Retweets



While the log distribution of retweets may appear normal, it does not pass a Shapiro-Wilk normality test.

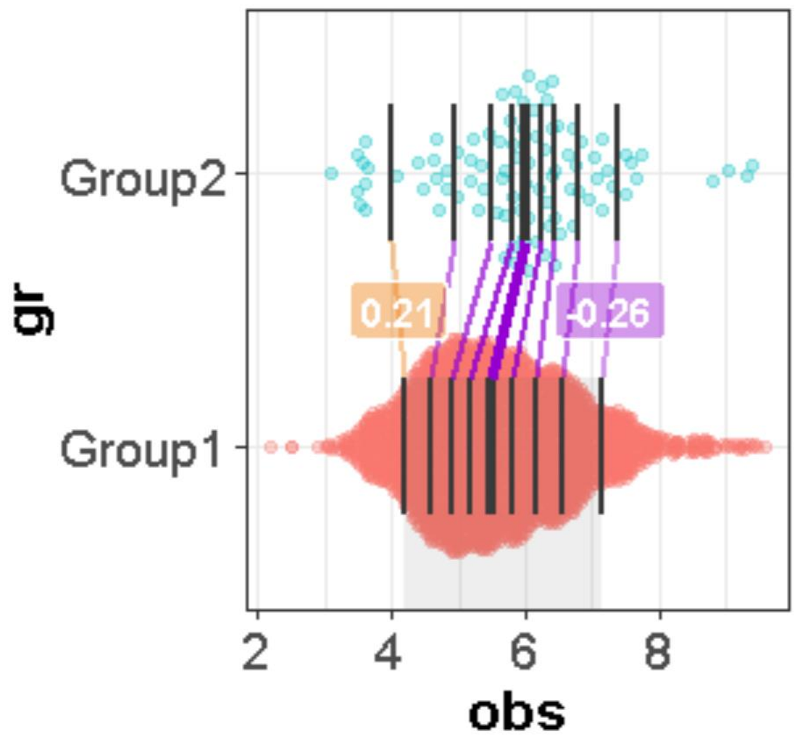




None of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 97210, p-value = 0.002865
alternative hypothesis: true location shift is not equal to 0
```

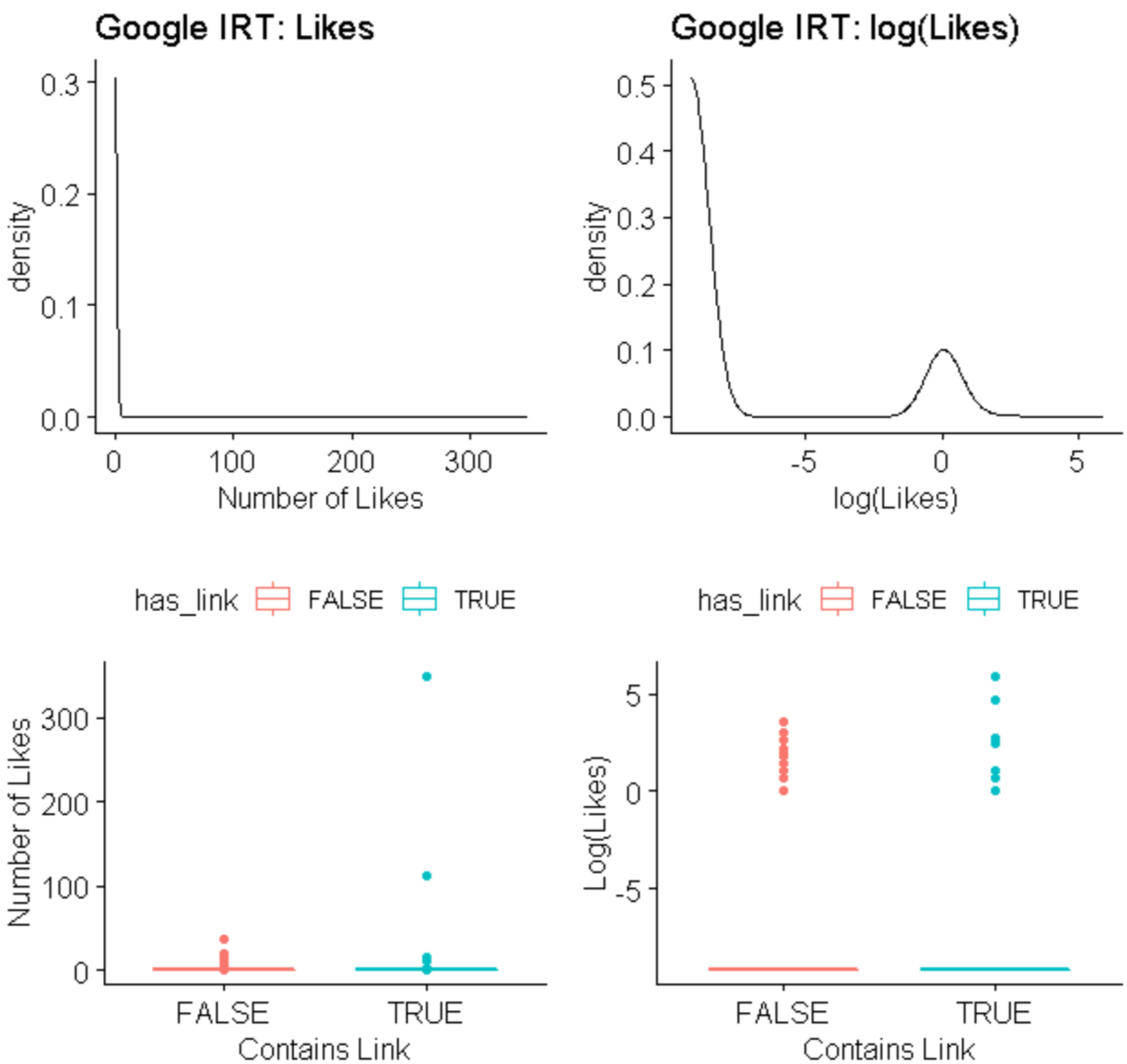
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

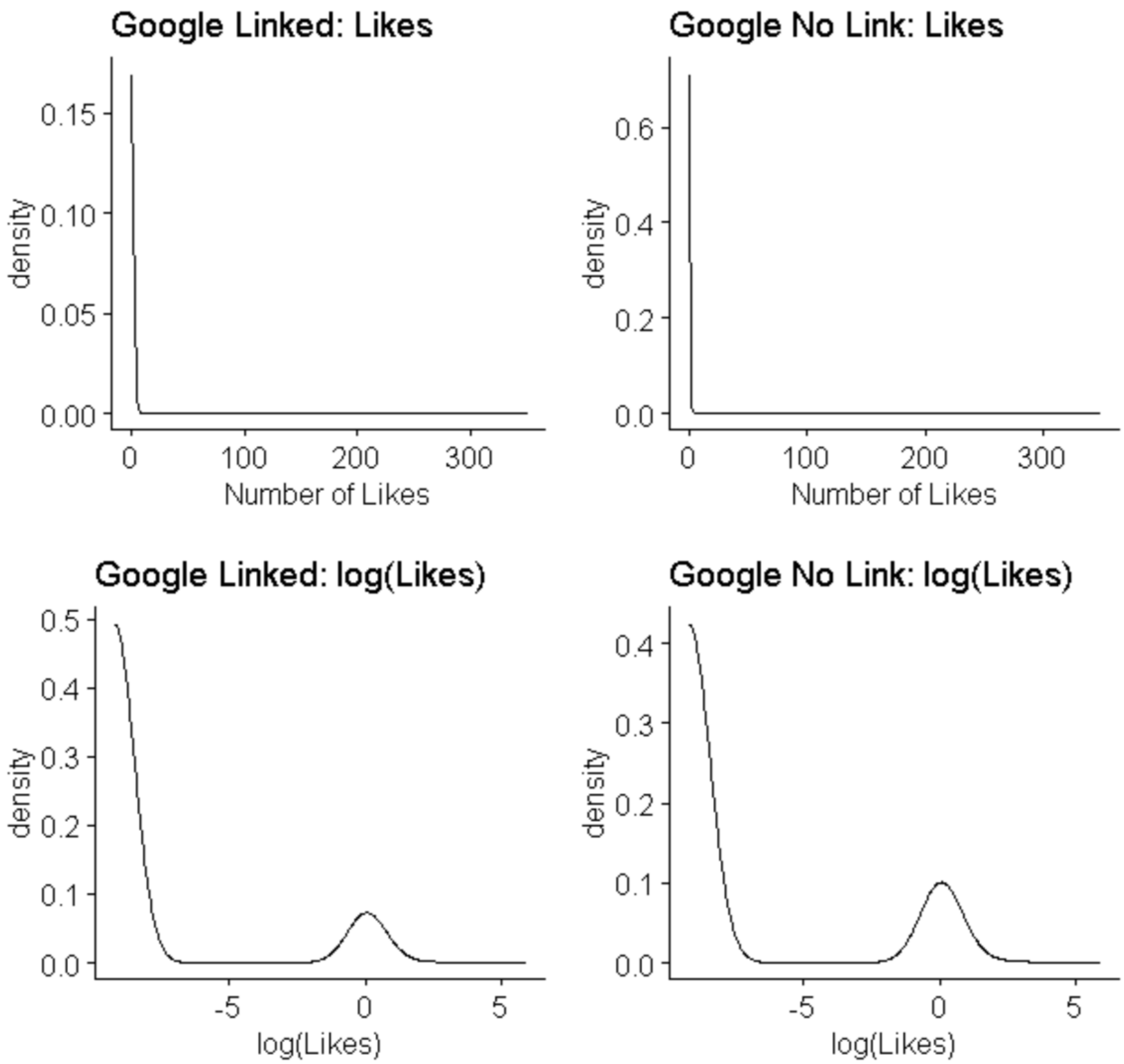



```
$`Group1 - Group2`
     q     Group1     Group2  difference    ci_lower   ci_upper      p_crit p_value
1 0.1   66.84446   59.43106    7.413395   -48.11041  30.833894 0.050000000   0.835
2 0.2   98.49059  143.66518  -45.174589  -152.36547  24.426220 0.012500000   0.110
3 0.3  135.15328  245.87726 -110.723981  -217.56448  -7.292960 0.008333333   0.005
4 0.4  180.15585  329.54578 -149.389931  -260.40907 -45.508418 0.007142857   0.000
5 0.5  242.07645  405.18274 -163.106287  -303.37297 -64.954953 0.006250000   0.000
6 0.6  336.07767  510.69045 -174.612780  -328.35124 -42.805694 0.005555556   0.000
7 0.7  474.96781  637.88364 -162.915829  -467.95993   3.588967 0.010000000   0.013
8 0.8  712.22624  919.86817 -207.641930  -759.69564  88.505234 0.016666667   0.140
9 0.9 1253.65952 1712.39504 -458.735516 -3624.04874 244.602119 0.025000000   0.197
```

We can say with 95% confidence that the 3rd, 4th, 5th, and 6th quantiles of group 2 (Disney OT tweets not containing links) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Disney OT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets, and the mid quantiles specifically), and potentially one underlying factor explaining these differences is the inclusion of a link in Disney official tweets.**

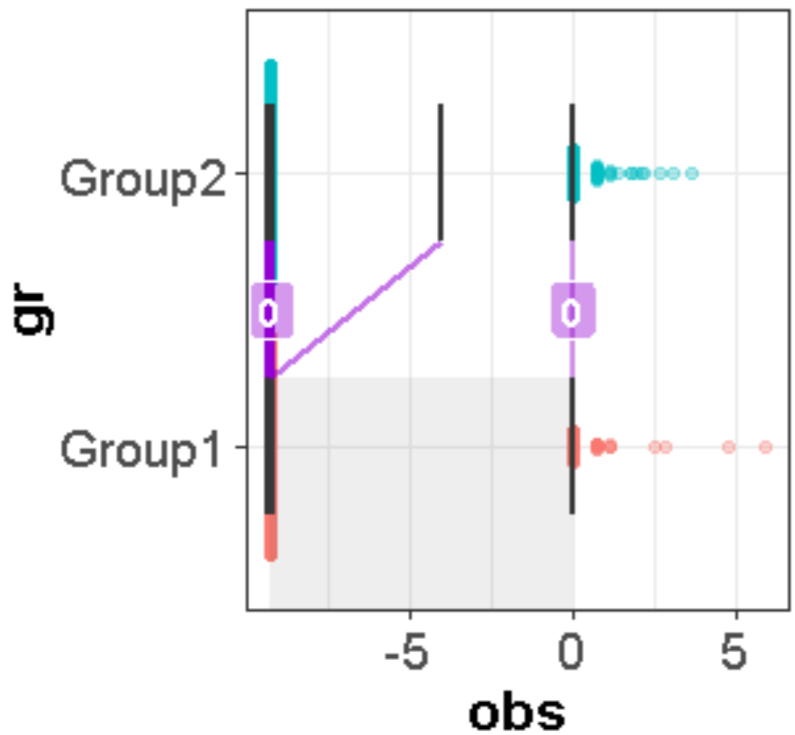
Google IRT: Number of Likes






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 1030827, p-value = 3.639e-06
alternative hypothesis: true location shift is not equal to 0
```

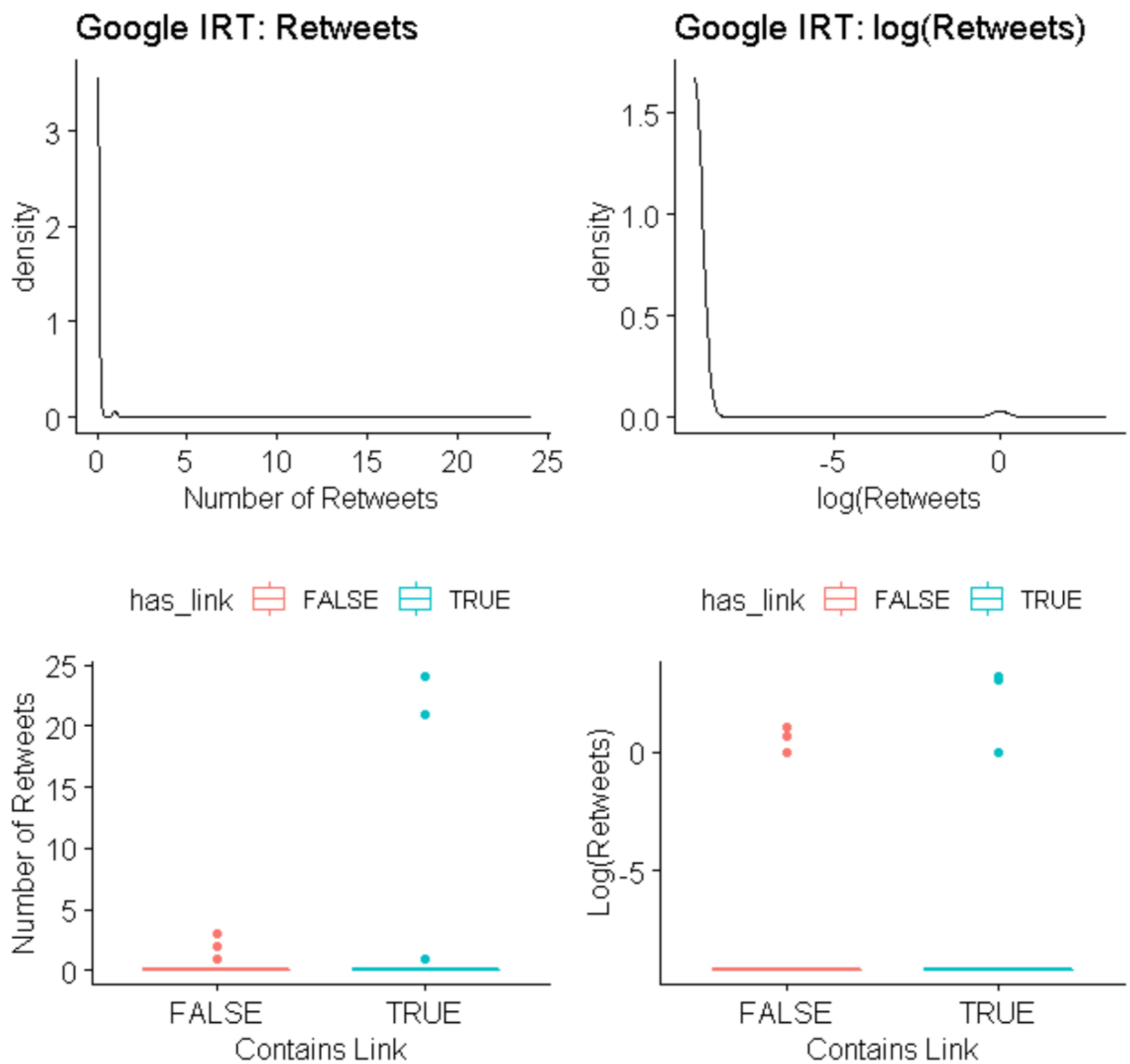
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

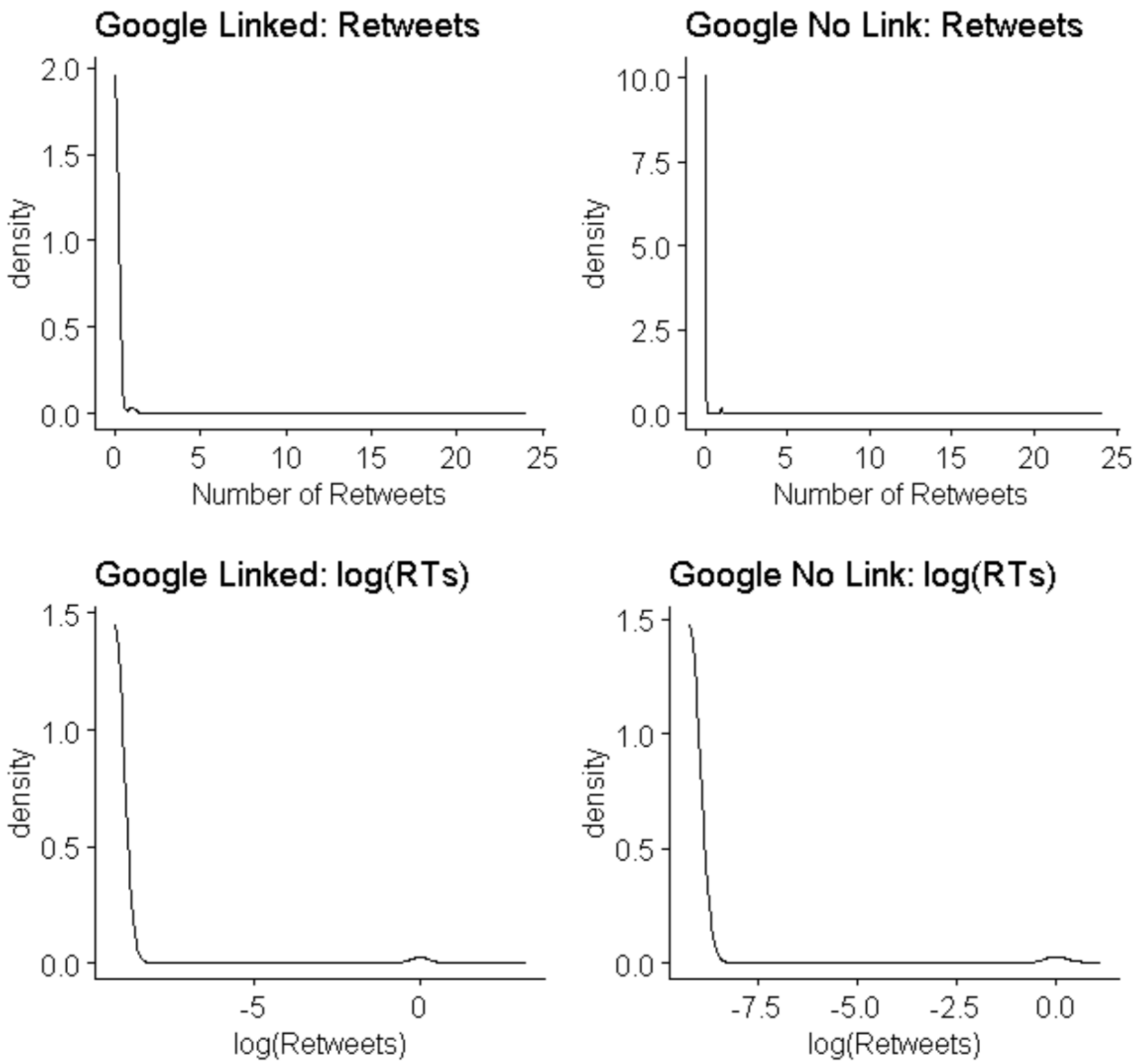



```
$`Group1 - Group2`
    q       Group1    Group2    difference      ci_lower      ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.000000e+00  0.000000e+00  0.000000e+00 0.050000000   1.000
2 0.2 0.000000e+00 0.0000000  0.000000e+00  0.000000e+00  0.000000e+00 0.025000000   1.000
3 0.3 0.000000e+00 0.0000000  0.000000e+00  0.000000e+00  0.000000e+00 0.016666667   1.000
4 0.4 0.000000e+00 0.0000000  0.000000e+00  0.000000e+00  0.000000e+00 0.012500000   1.000
5 0.5 0.000000e+00 0.0000000  0.000000e+00  0.000000e+00  0.000000e+00 0.010000000   1.000
6 0.6 0.000000e+00 0.0000000  0.000000e+00  0.000000e+00  0.000000e+00 0.008333333   1.000
7 0.7 0.000000e+00 0.0000000  0.000000e+00 -1.764366e-12  0.000000e+00 0.007142857   0.944
8 0.8 4.074179e-10 0.5606616 -5.606616e-01 -9.980187e-01 -8.523077e-03 0.006250000   0.000
9 0.9 9.999422e-01 1.0000000 -5.775516e-05 -7.949629e-02 -1.493209e-10 0.005555556   0.000
```

We can say with 95% confidence that the 8th and 9th quantiles of group 2 (Google IRT tweets not containing links) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Google IRT tweets containing links). However, the visualizations produced above are not convincing. **Inclusion of a link seems to have no effect on the number of likes a Google IRT tweet receives.**

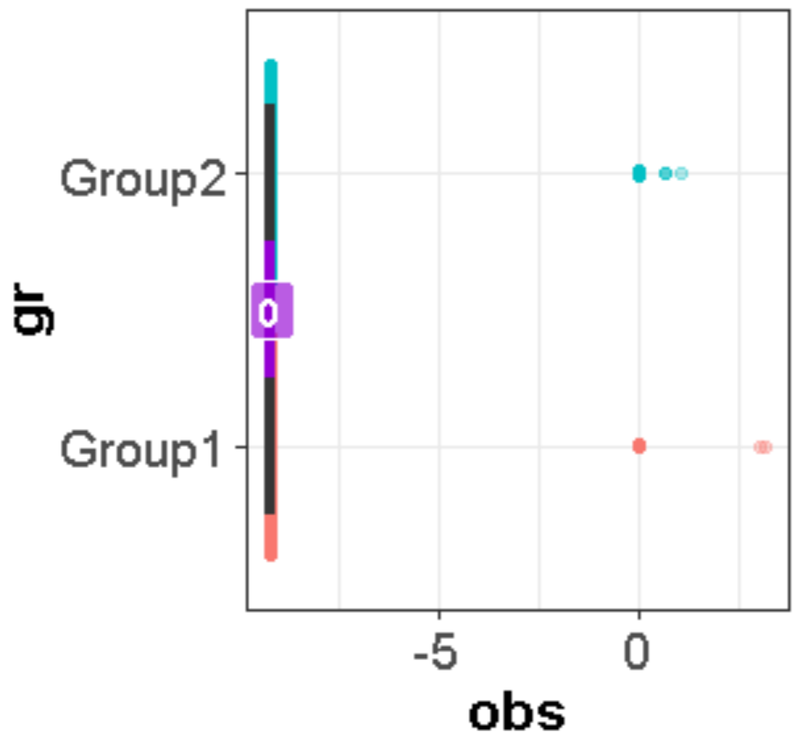
Google IRT: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 1099293, p-value = 0.522
alternative hypothesis: true location shift is not equal to 0
```

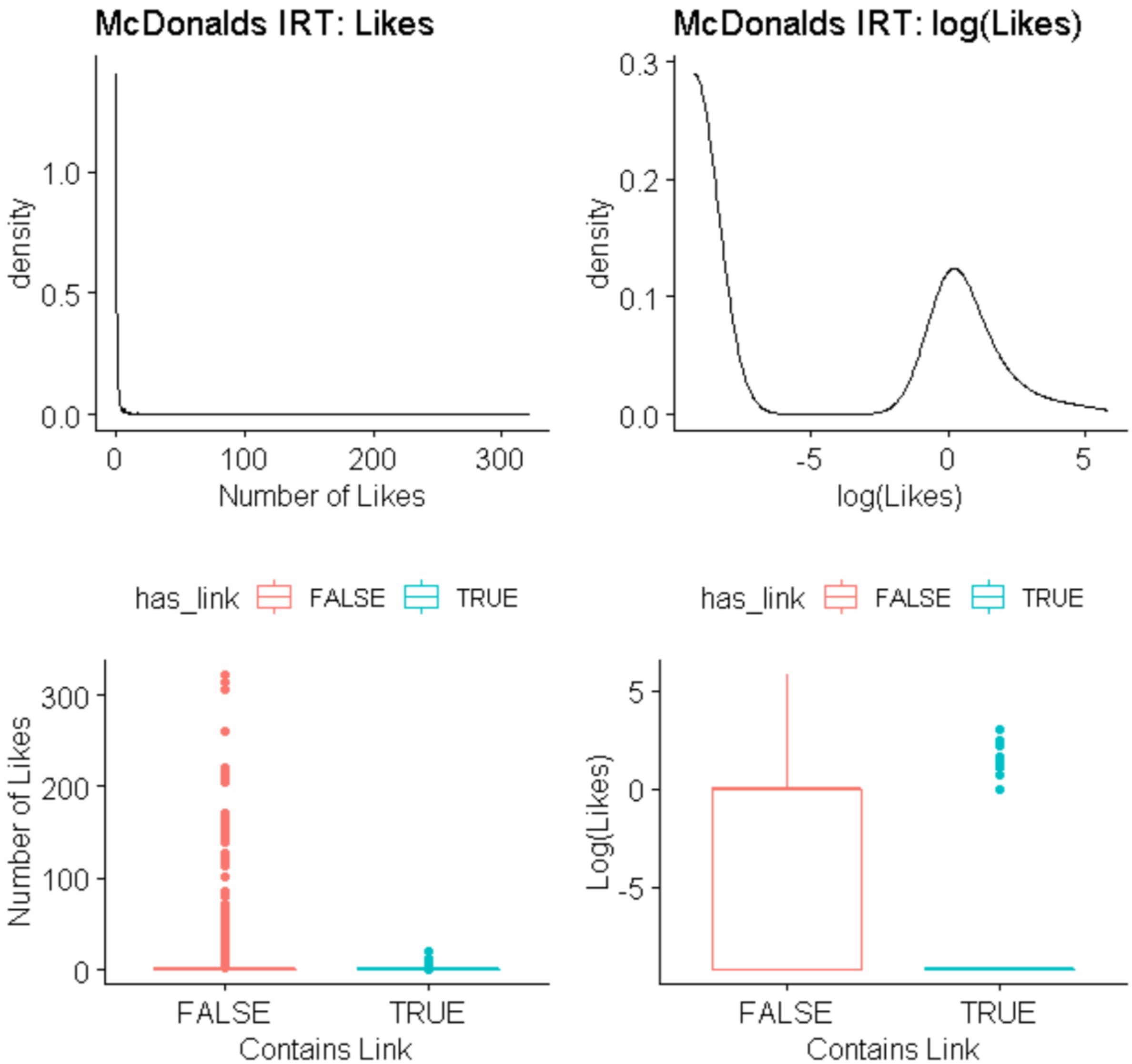
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of both populations are equal. Performing a shift function yields the following:

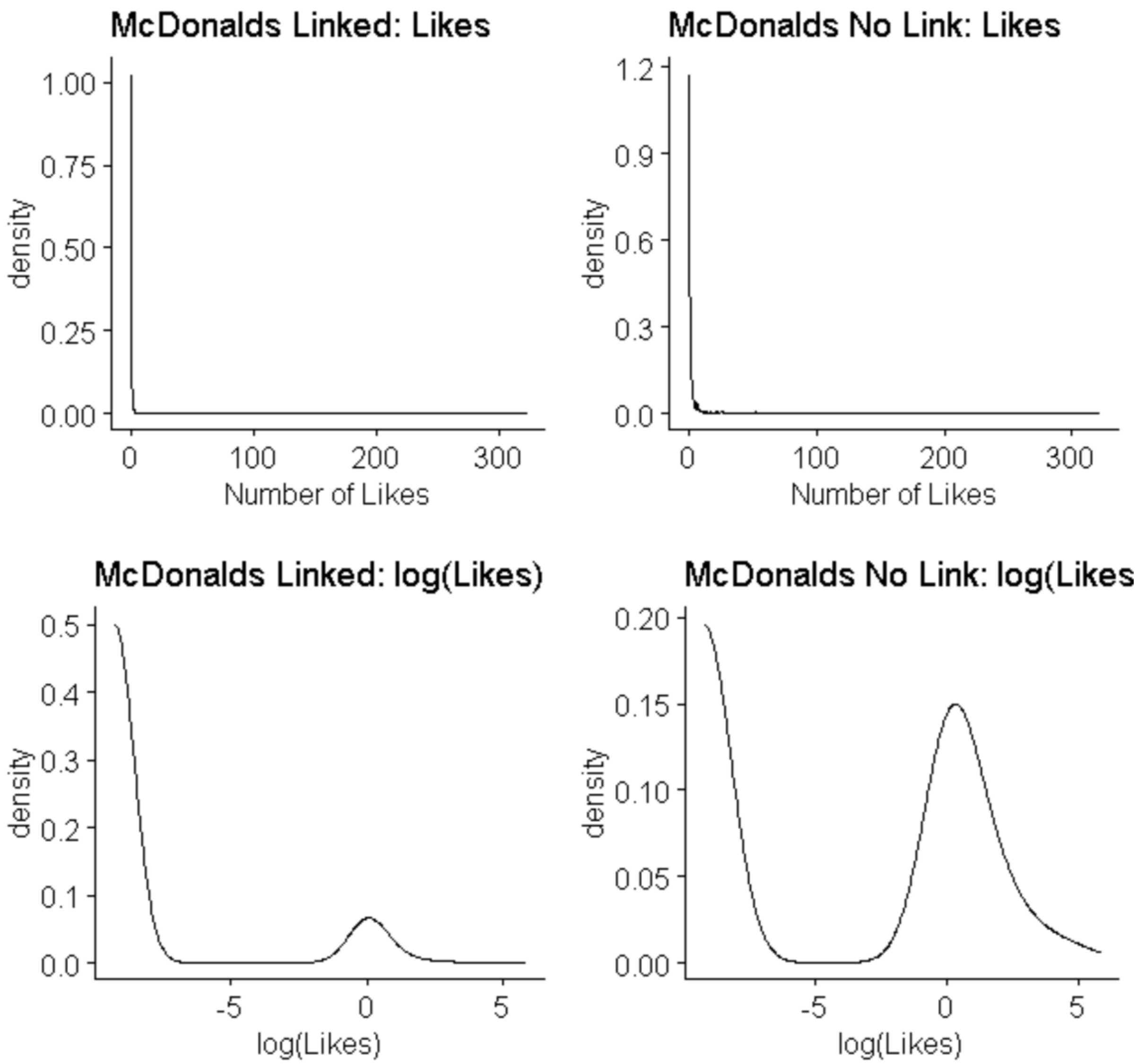



```
$`Group1 - Group2`
    q Group1 Group2 difference ci_lower ci_upper      p_crit p_value
1 0.1      0      0          0        0        0 0.050000000       1
2 0.2      0      0          0        0        0 0.025000000       1
3 0.3      0      0          0        0        0 0.016666667       1
4 0.4      0      0          0        0        0 0.012500000       1
5 0.5      0      0          0        0        0 0.010000000       1
6 0.6      0      0          0        0        0 0.008333333       1
7 0.7      0      0          0        0        0 0.007142857       1
8 0.8      0      0          0        0        0 0.006250000       1
9 0.9      0      0          0        0        0 0.005555556       1
```

Inclusion of a link does not seem to have a statistically significant effect on the number of retweets which a Google IRT tweet receives. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of retweets are equal to one another. Furthermore, it appears that Google has quite a bit of trouble getting anyone to retweet their IRT tweets, regardless of whether they contain a link or not.

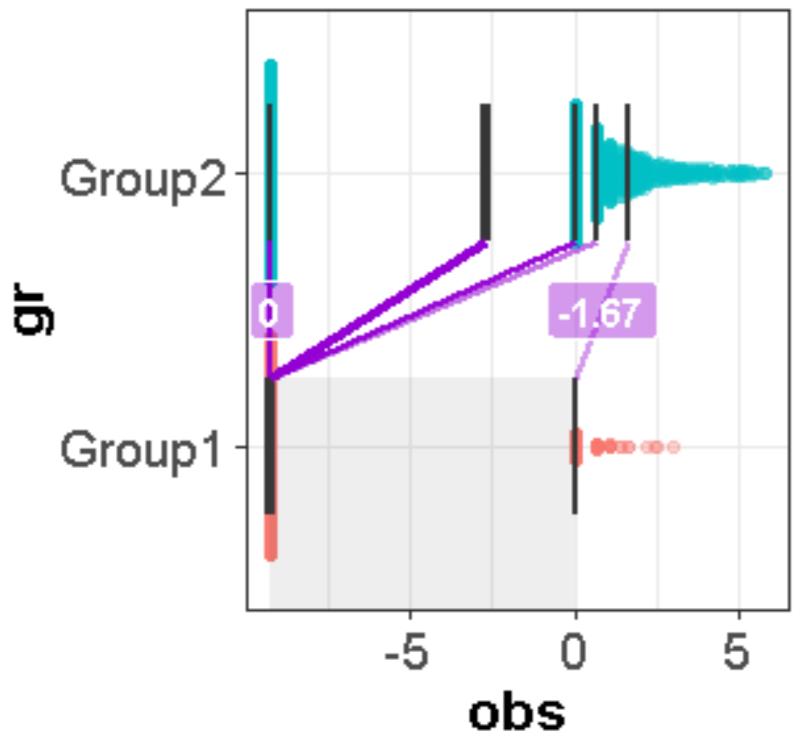
McDonalds IRT: Number of Likes






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 709590, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

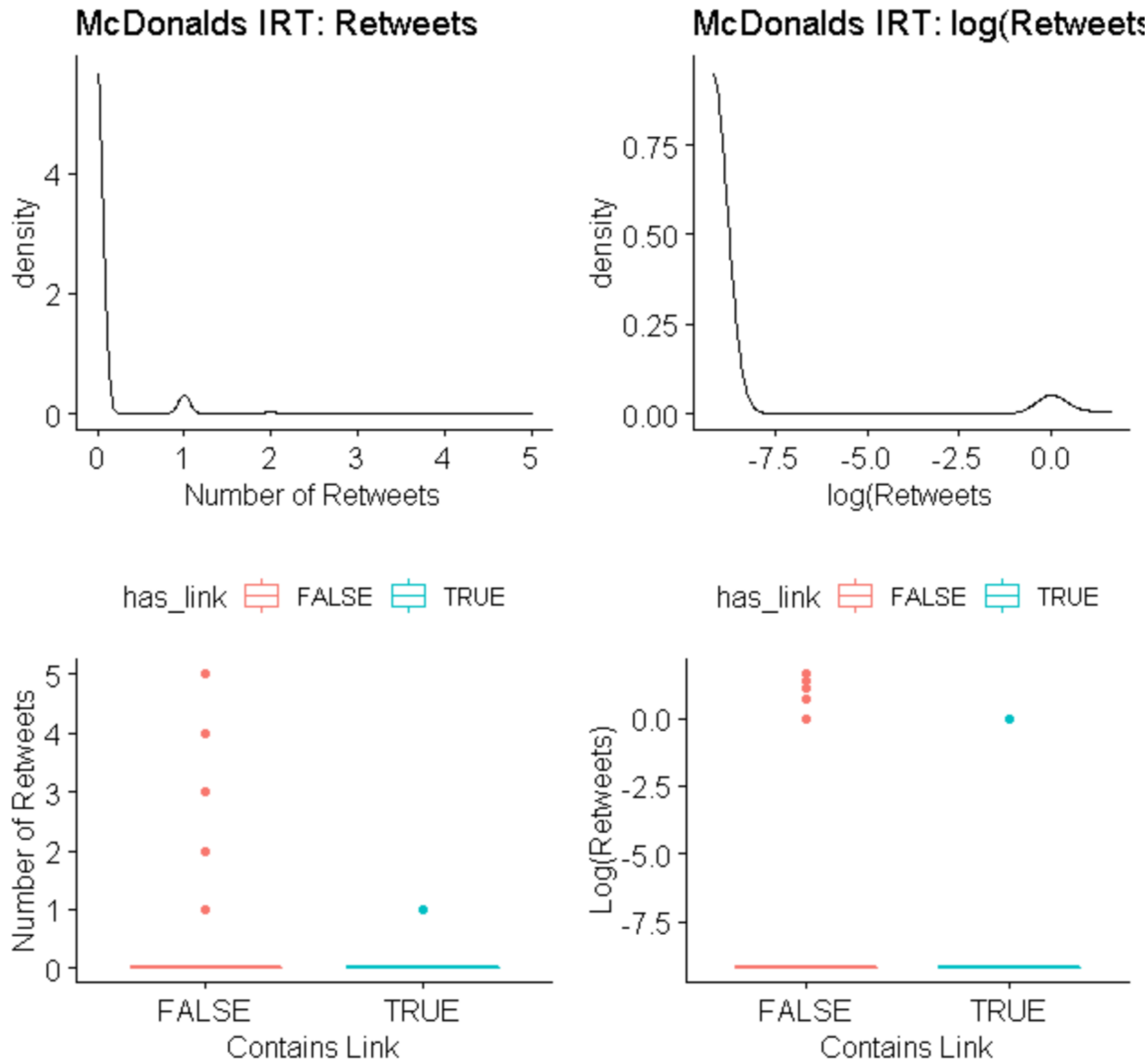
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

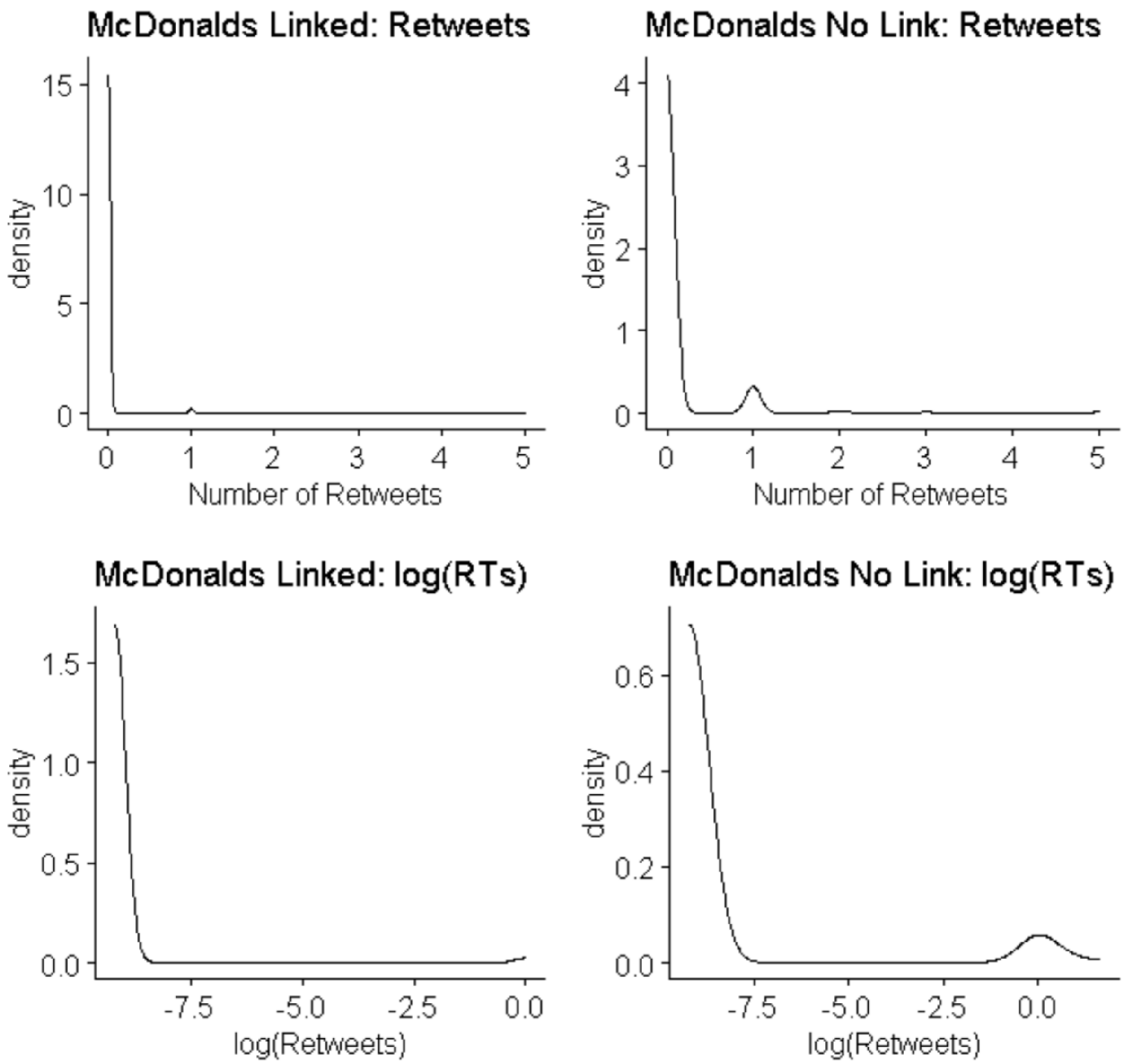



```
$`Group1 - Group2`
    q       Group1    Group2 difference      ci_lower    ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.050000000  1.0000
2 0.2 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.025000000  1.0000
3 0.3 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.016666667  1.0000
4 0.4 0.000000e+00 0.0000000  0.0000000 -7.171217e-10  0.00000000 0.012500000  0.5505
5 0.5 0.000000e+00 0.7047426 -0.7047426 -9.988658e-01 -0.02175412 0.010000000  0.0000
6 0.6 0.000000e+00 1.0000000 -1.0000000 -1.000000e+00 -1.00000000 0.008333333  0.0000
7 0.7 0.000000e+00 1.0000000 -1.0000000 -1.000002e+00 -1.00000000 0.007142857  0.0000
8 0.8 8.551604e-12 1.9986385 -1.9986385 -2.039033e+00 -1.59283212 0.006250000  0.0000
9 0.9 9.992349e-01 5.2861205 -4.2868857 -5.748545e+00 -3.20965634 0.005555556  0.0000
```

We can say, with 95% confidence, that the 5th through 9th quantiles of group 2 (McDonalds IRT tweets not containing links) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (McDonalds IRT tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes, and the right halves specifically), and potentially one underlying factor explaining these differences is the inclusion of a link in McDonalds IRT tweets.**

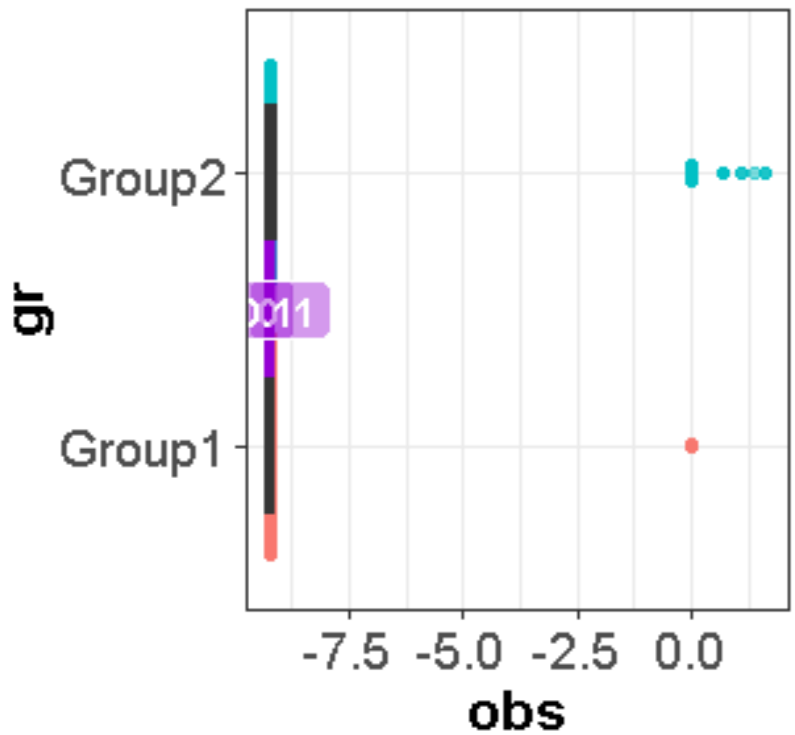
McDonalds IRT: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 1089504, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

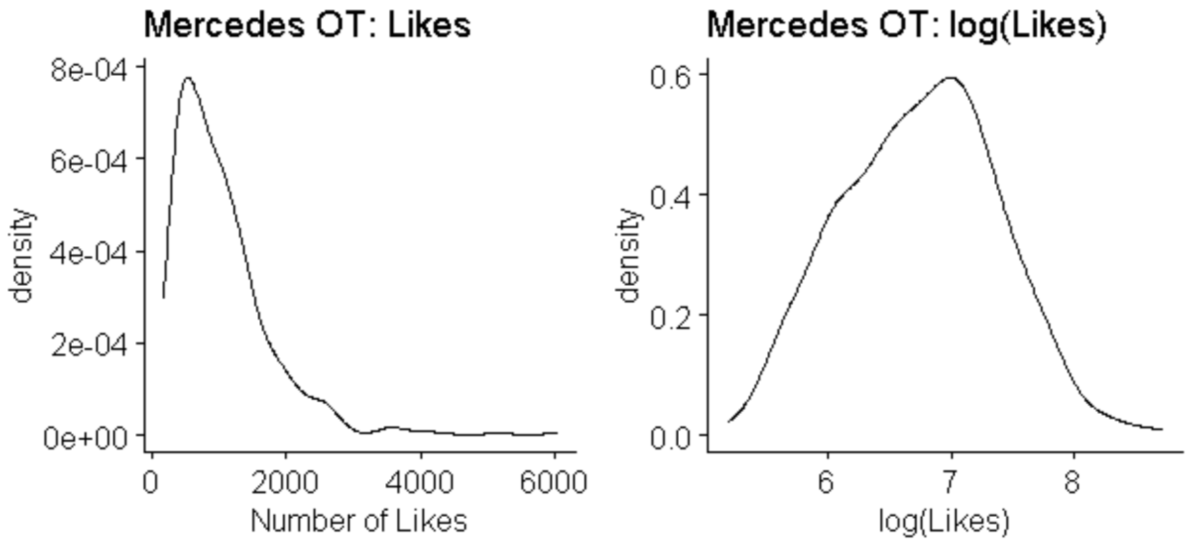
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



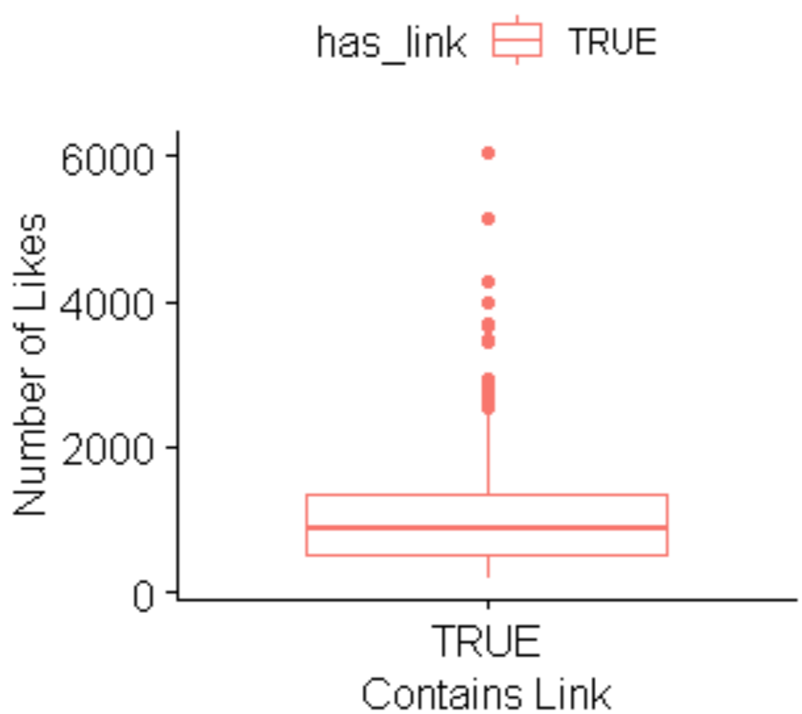

```
$`Group1 - Group2`
    q Group1     Group2  difference   ci_lower     ci_upper      p_crit p_value
1 0.1      0 0.00000000  0.00000000  0.0000000  0.000000e+00 0.050000000       1
2 0.2      0 0.00000000  0.00000000  0.0000000  0.000000e+00 0.025000000       1
3 0.3      0 0.00000000  0.00000000  0.0000000  0.000000e+00 0.016666667       1
4 0.4      0 0.00000000  0.00000000  0.0000000  0.000000e+00 0.012500000       1
5 0.5      0 0.00000000  0.00000000  0.0000000  0.000000e+00 0.010000000       1
6 0.6      0 0.00000000  0.00000000  0.0000000  0.000000e+00 0.008333333       1
7 0.7      0 0.00000000  0.00000000  0.0000000  0.000000e+00 0.007142857       1
8 0.8      0 0.00000000  0.00000000  0.0000000  0.000000e+00 0.006250000       1
9 0.9      0 0.01222561 -0.01222561 -0.7045939 -8.032865e-08 0.005555556       0
```

We can say, with 95% confidence, that the 9th quantile of group 2 (McDonalds IRT tweets not containing links) would need to be shifted down by a significant (non-zero) amount to match its counterparts in group 1 (McDonalds IRT tweets containing links). However, the visualizations produced make it seem as if this could just be happenstance, or a result of uneven sample sizes. **Inclusion of a link seems to have no effect on the number of retweets a McDonalds IRT tweet receives..**

Mercedes Official: Number of Likes

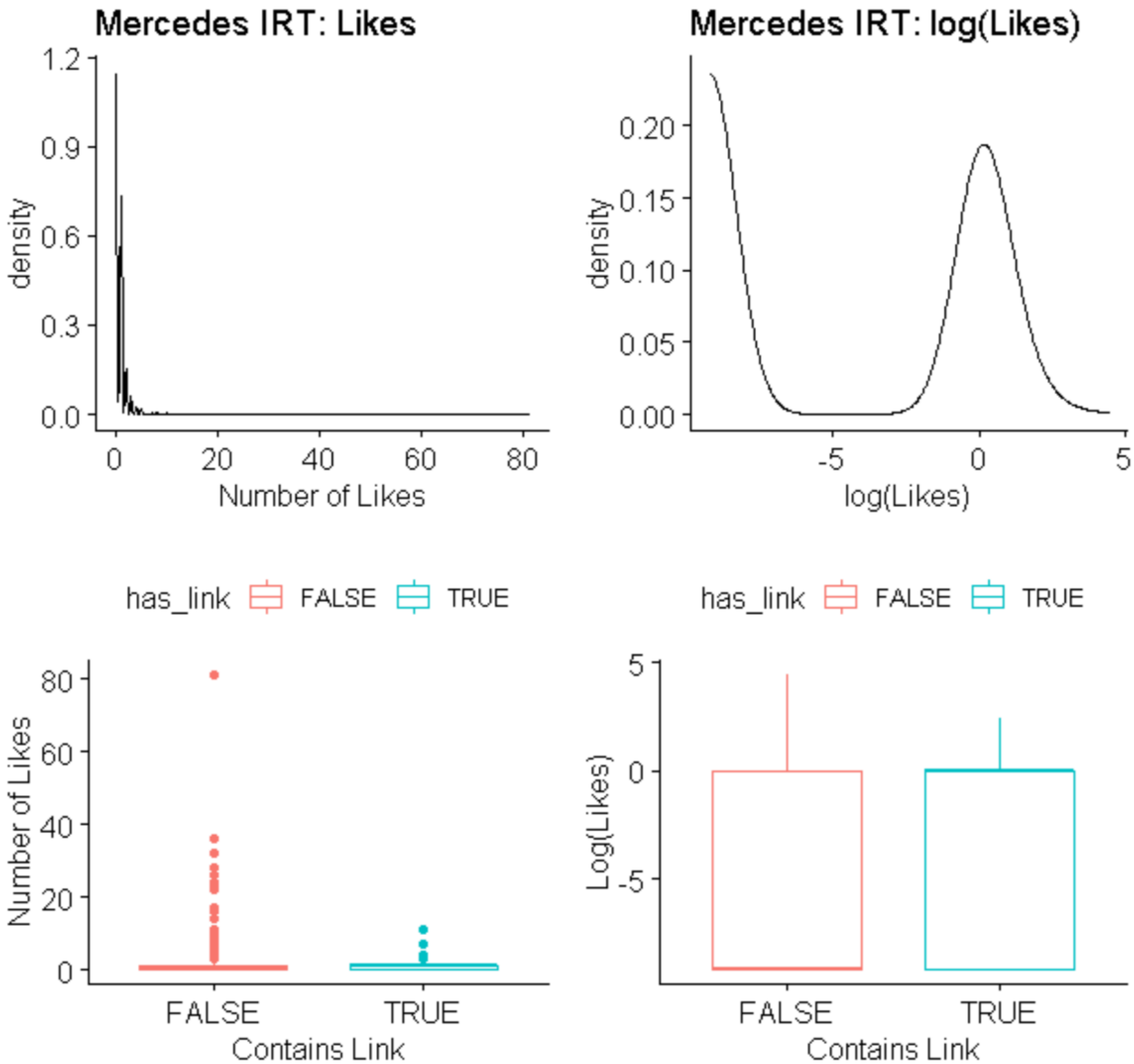


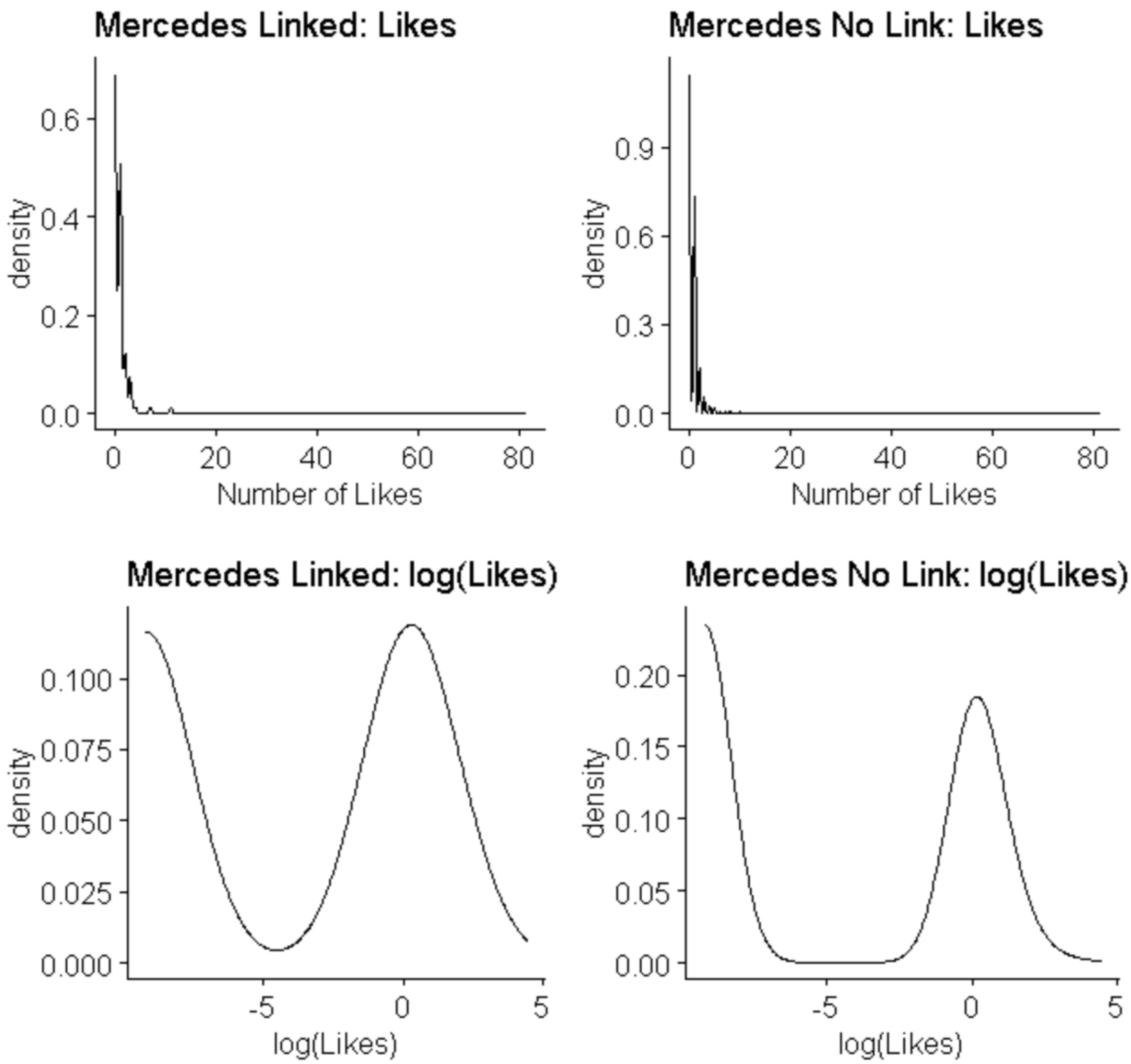
The log distribution of likes just barely does not pass a Shapiro-Wilk normality test.



Apparently, all 501 of Mercedes' official tweets which were not considered extreme outliers include a link, making this level of analysis impossible for Mercedes official tweets.

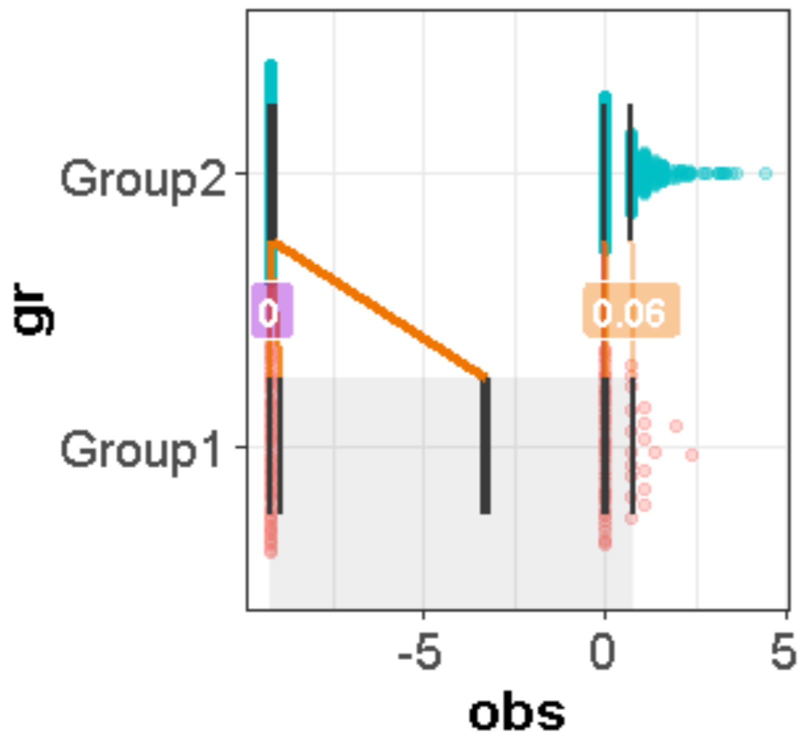
Mercedes IRT: Number of Likes






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 152601, p-value = 0.226
alternative hypothesis: true location shift is not equal to 0
```

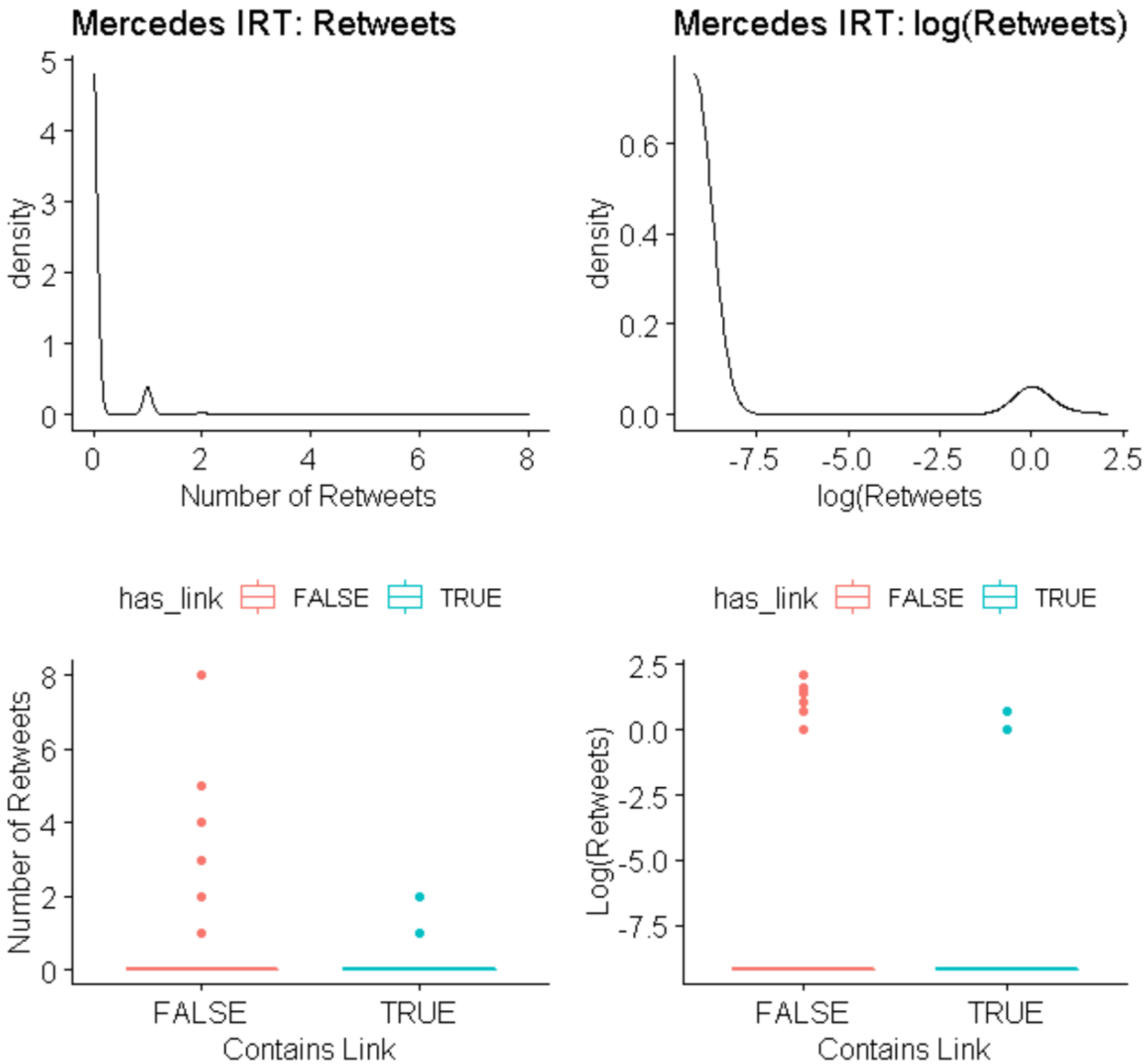
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘like’ distributions of both populations are equal. Performing a shift function yields the following:

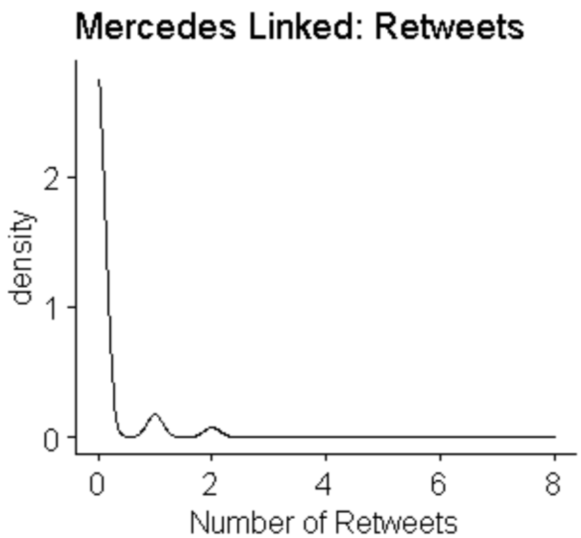


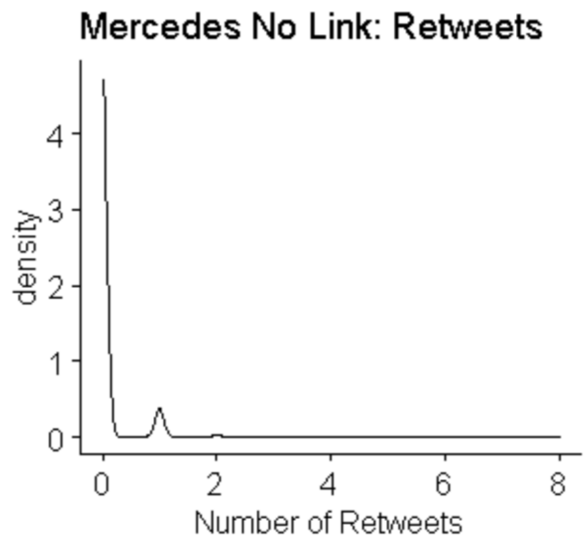

```
$`Group1 - Group2`
    q       Group1      Group2    difference      ci_lower     ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000000  0.000000e+00  0.000000e+00 1.619815e-13 0.050000000   0.842
2 0.2 3.896594e-11 0.000000000  3.896594e-11  0.000000e+00 2.892201e-05 0.010000000   0.050
3 0.3 1.907954e-05 0.000000000  1.907954e-05  4.217515e-11 3.971844e-02 0.006250000   0.000
4 0.4 3.211213e-02 0.000000000  3.211213e-02  1.891952e-06 8.701271e-01 0.005555556   0.000
5 0.5 6.428218e-01 0.001092373  6.417294e-01 -7.454777e-02 9.970553e-01 0.008333333   0.021
6 0.6 9.950621e-01 1.000000000 -4.937914e-03 -4.950585e-01 5.487895e-05 0.012500000   0.058
7 0.7 1.000056e+00 1.000000000  5.587094e-05 -3.753745e-03 4.917349e-02 0.025000000   0.612
8 0.8 1.121637e+00 1.000000000  1.216375e-01  6.957420e-06 9.485614e-01 0.007142857   0.001
9 0.9 2.154467e+00 1.999483235  1.549839e-01 -7.032035e-01 9.913089e-01 0.016666667   0.544
```

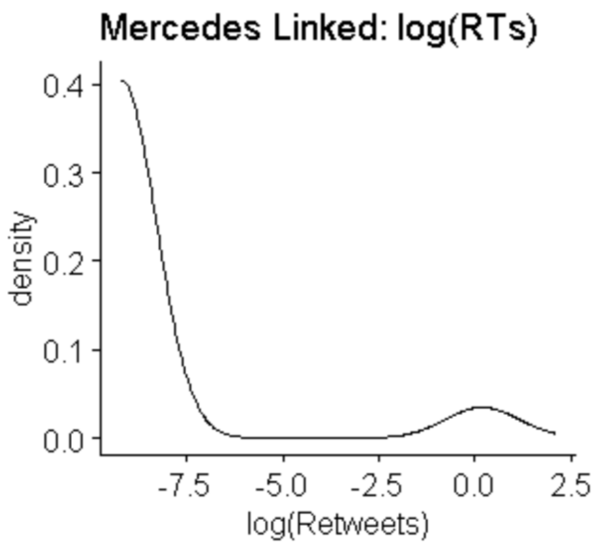
Despite some confidence intervals indicating statistical significance, I don't feel quite comfortable concluding such, given the insignificant results stemming from the Mann-Whitney U test. **Inclusion of a link does not seem to have a statistically significant effect on the number of likes which a Mercedes IRT tweet receives.**

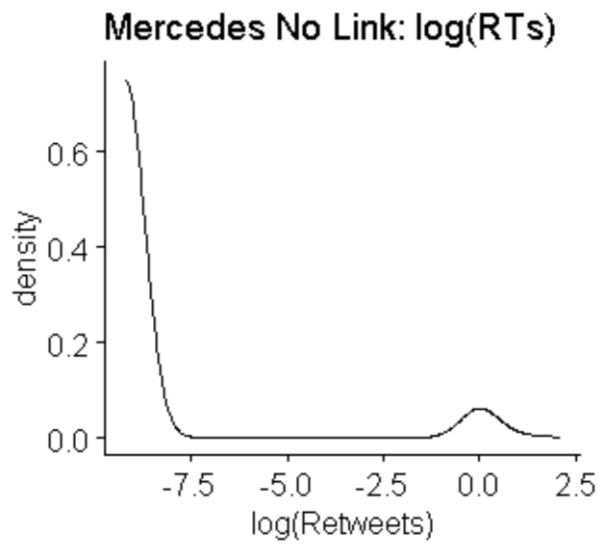
Mercedes IRT: Number of Retweets





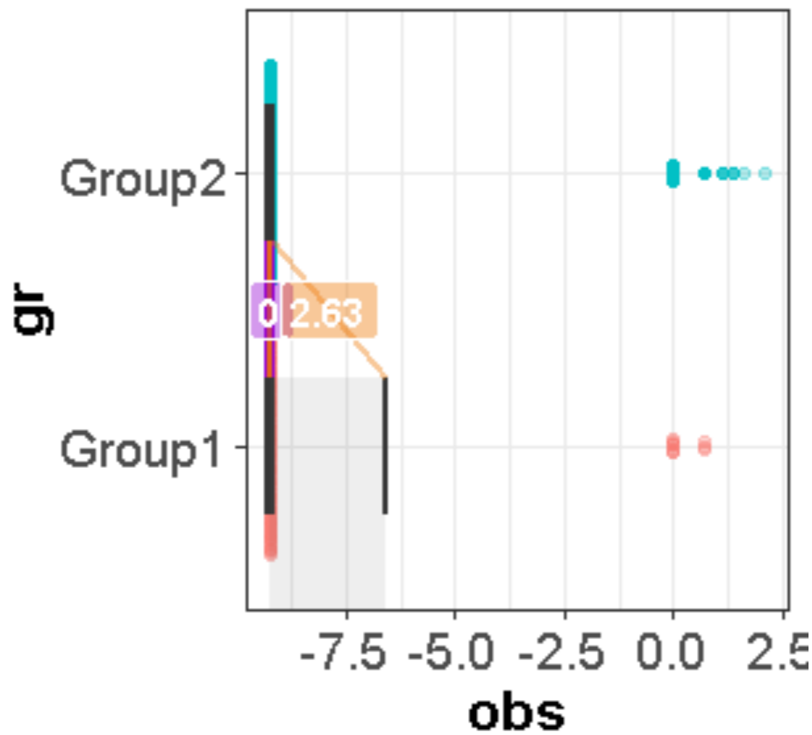







```
	Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 144362, p-value = 0.9482
alternative hypothesis: true location shift is not equal to 0
```

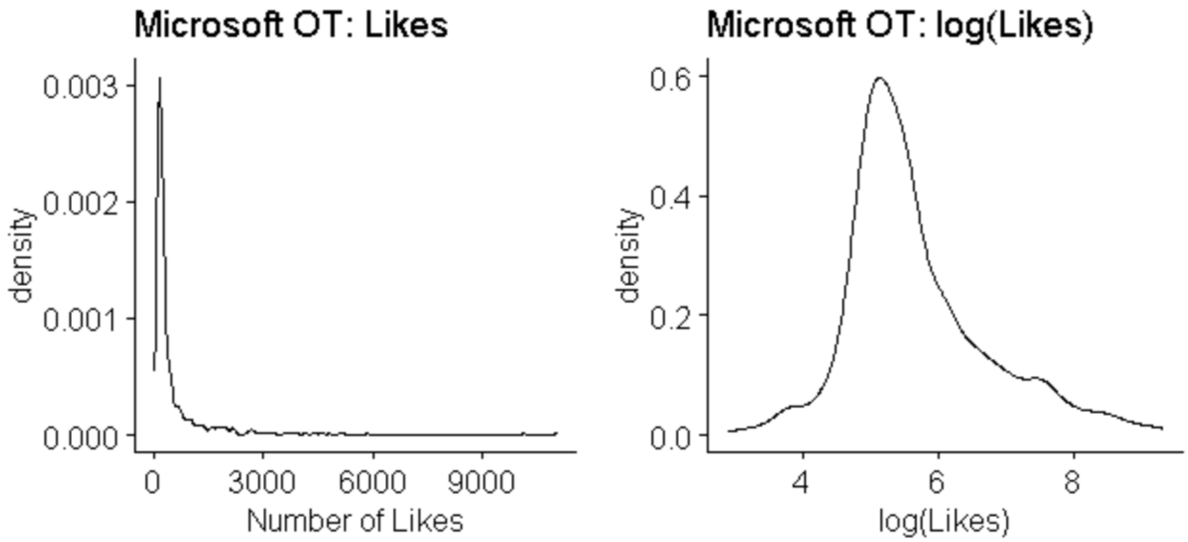
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of both populations are equal. Performing a shift function yields the following:



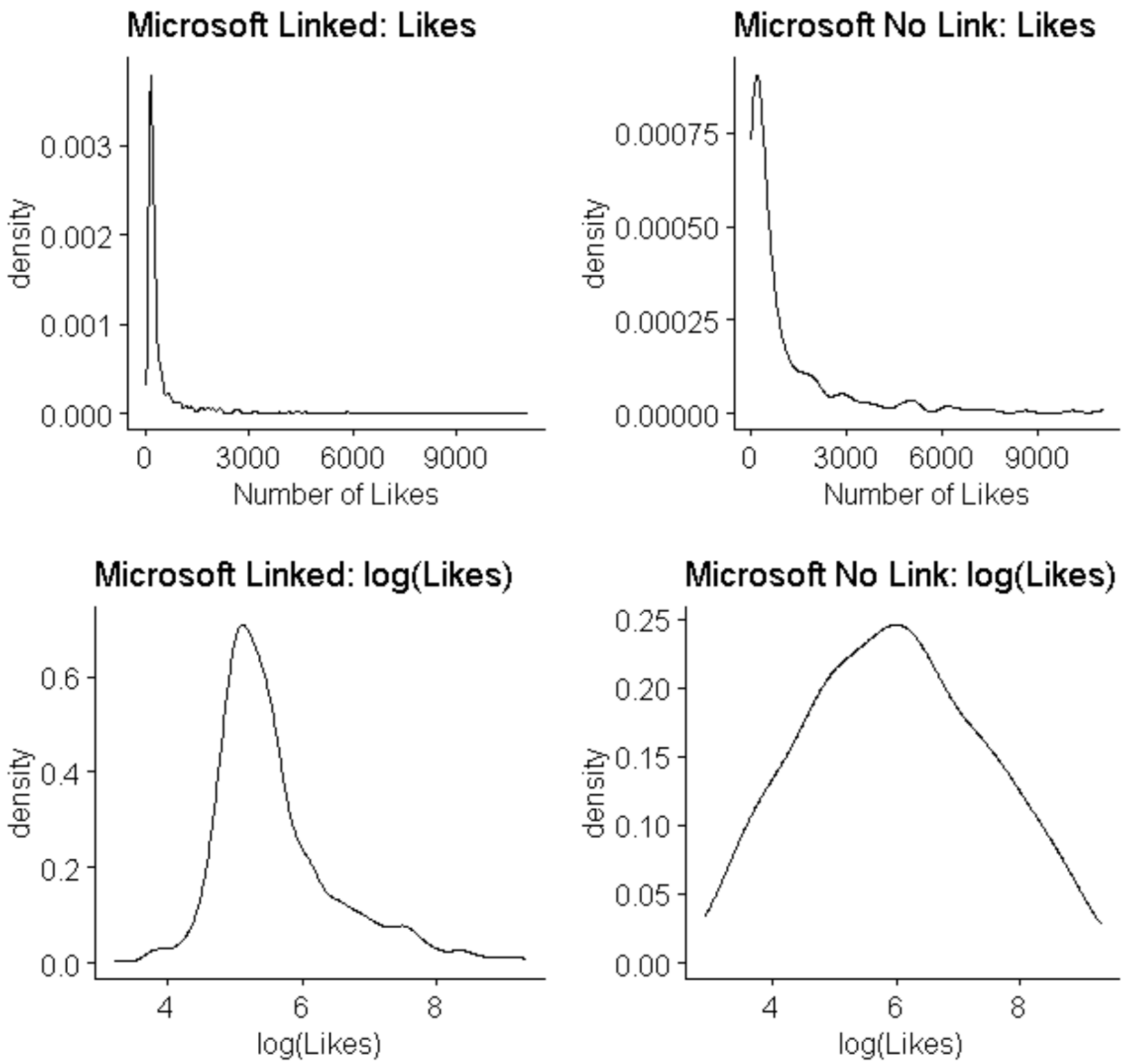

```
$`Group1 - Group2`
    q       Group1      Group2   difference      ci_lower     ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000000 0.000000e+00  0.000000e+00 0.000000e+00 0.050000000  1.0000
2 0.2 0.000000e+00 0.000000000 0.000000e+00  0.000000e+00 0.000000e+00 0.025000000  1.0000
3 0.3 0.000000e+00 0.000000000 0.000000e+00  0.000000e+00 0.000000e+00 0.016666667  1.0000
4 0.4 0.000000e+00 0.000000000 0.000000e+00  0.000000e+00 0.000000e+00 0.012500000  1.0000
5 0.5 0.000000e+00 0.000000000 0.000000e+00  0.000000e+00 0.000000e+00 0.010000000  0.9990
6 0.6 0.000000e+00 0.000000000 0.000000e+00  0.000000e+00 1.118865e-10 0.008333333  0.8690
7 0.7 4.017342e-12 0.000000000 4.017342e-12  0.000000e+00 5.652006e-05 0.006250000  0.1250
8 0.8 3.958361e-05 0.000000000 3.958361e-05  2.442491e-15 1.216305e-01 0.005555556  0.0025
9 0.9 2.873011e-01 0.001622656 2.856784e-01 -2.093808e-01 9.988301e-01 0.007142857  0.1130
```

Despite one confidence interval indicating statistical significance, I don't feel quite comfortable concluding such, given the insignificant results stemming from the Mann-Whitney U test. **Inclusion of a link does not seem to have a statistically significant effect on the number of retweets which a Mercedes IRT tweet receives.**

Microsoft Official: Number of Likes



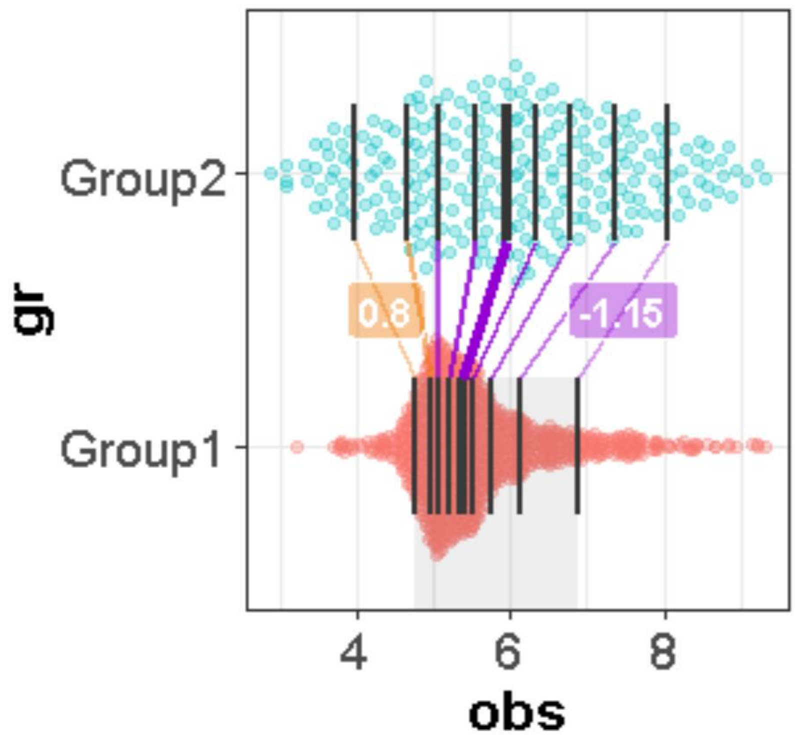
The log distribution does not pass a Shapiro-Wilk normality test.



Neither of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 101192, p-value = 1.719e-05
alternative hypothesis: true location shift is not equal to 0
```

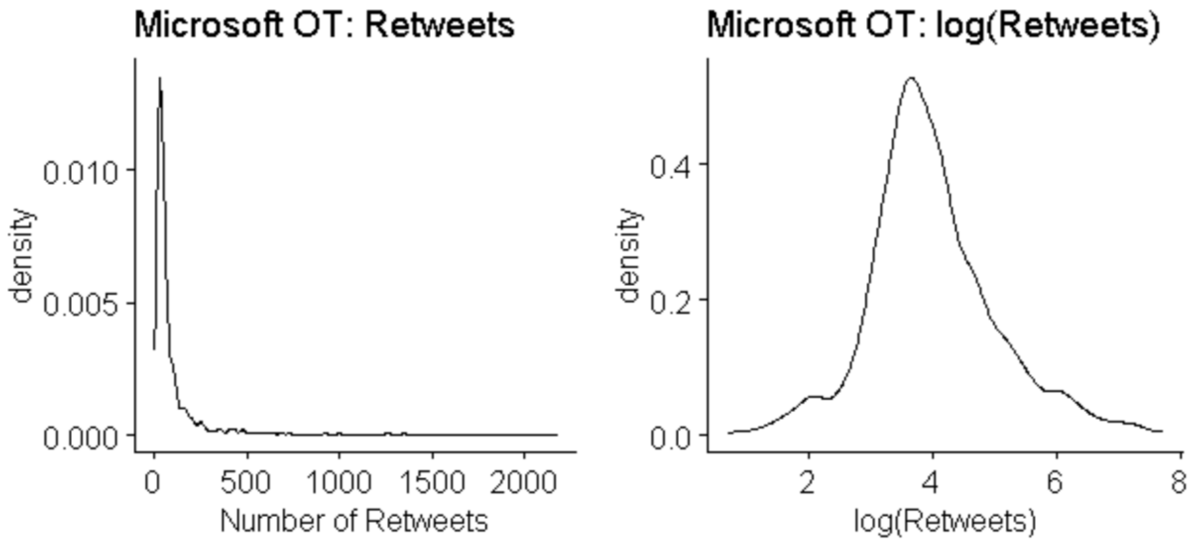
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



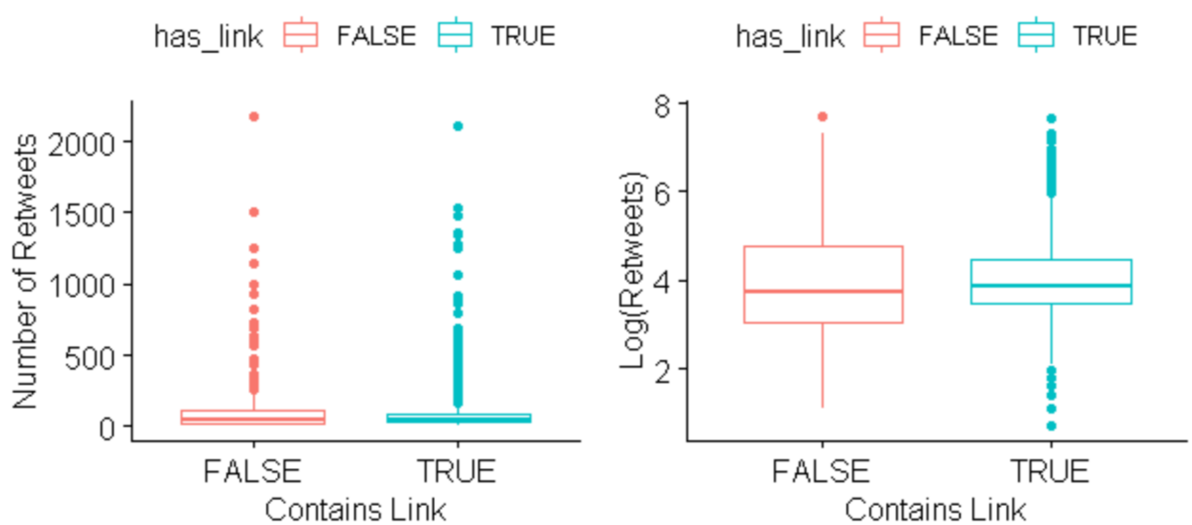

```
$`Group1 - Group2`
    q   Group1     Group2   difference     ci_lower     ci_upper      p_crit p_value
1 0.1 118.3146   53.96817    64.346425    37.160644    79.110830 0.012500000   0.000
2 0.2 141.8543  108.04155    33.812727     3.398456    65.060676 0.016666667   0.010
3 0.3 160.2222  161.69459    -1.472436   -58.475895    29.251678 0.050000000   0.925
4 0.4 184.8254  262.12779   -77.302348  -163.583942    -8.343386 0.025000000   0.015
5 0.5 214.6340  395.42579  -180.791772  -299.863715   -73.709465 0.010000000   0.000
6 0.6 250.8811  561.53113  -310.650054  -545.363847  -169.563473 0.008333333   0.000
7 0.7 313.1903  895.21207  -582.021728  -999.287045  -287.834340 0.007142857   0.000
8 0.8 466.9703 1612.29343 -1145.323089 -1853.232602  -564.672672 0.006250000   0.000
9 0.9 988.2740 3135.35597 -2147.081979 -3623.730230 -1094.010977 0.005555556   0.000
```

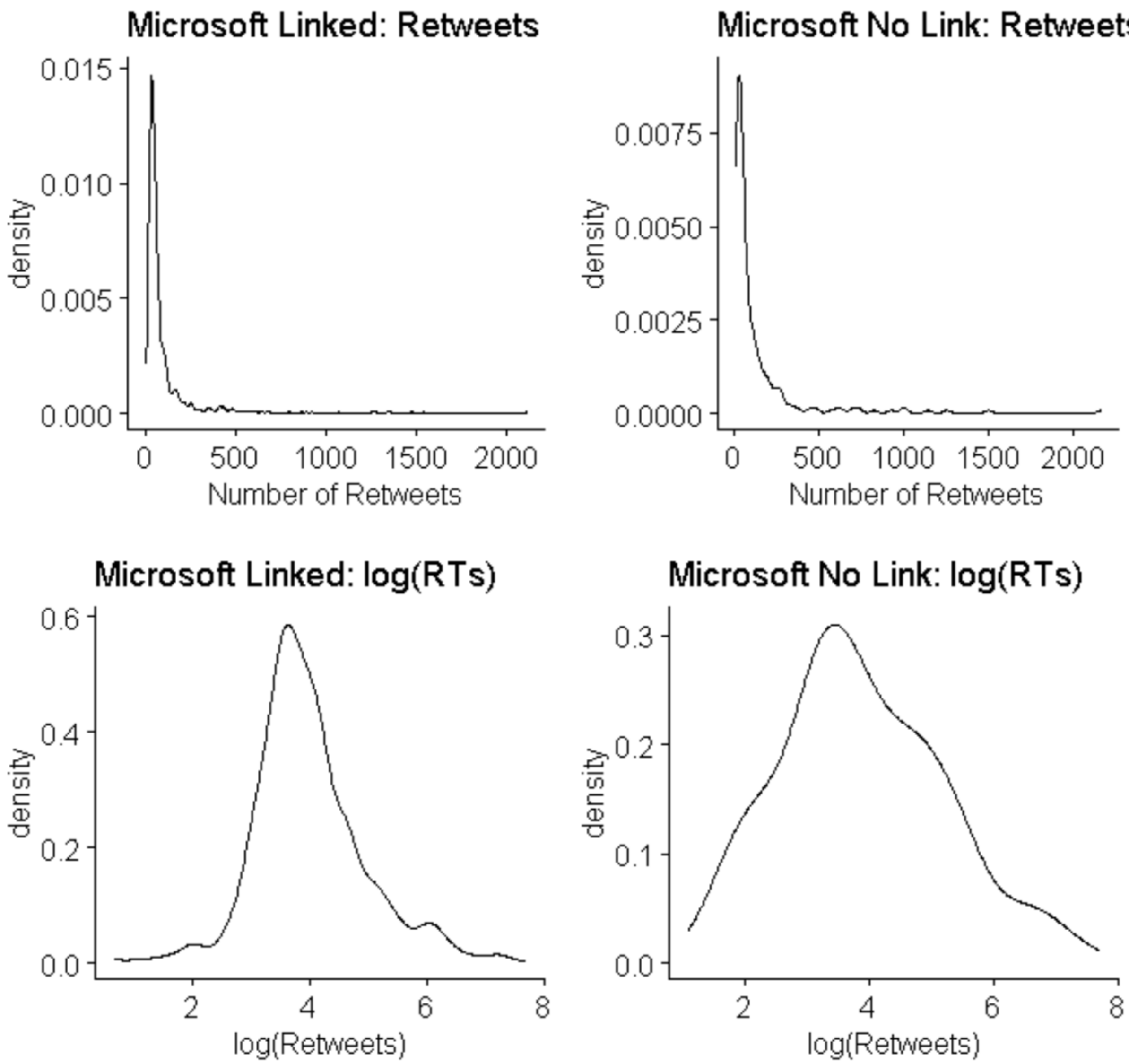
With 95% confidence we can say that the first and second quantiles of group 2 (Microsoft OT tweets not containing links) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Microsoft OT tweets containing links). Furthermore, the fourth through ninth quantiles of group 2 would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1. **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a link in Microsoft official tweets.**

Microsoft Official: Number of Retweets



The log distribution does not pass a Shapiro-Wilk normality test.

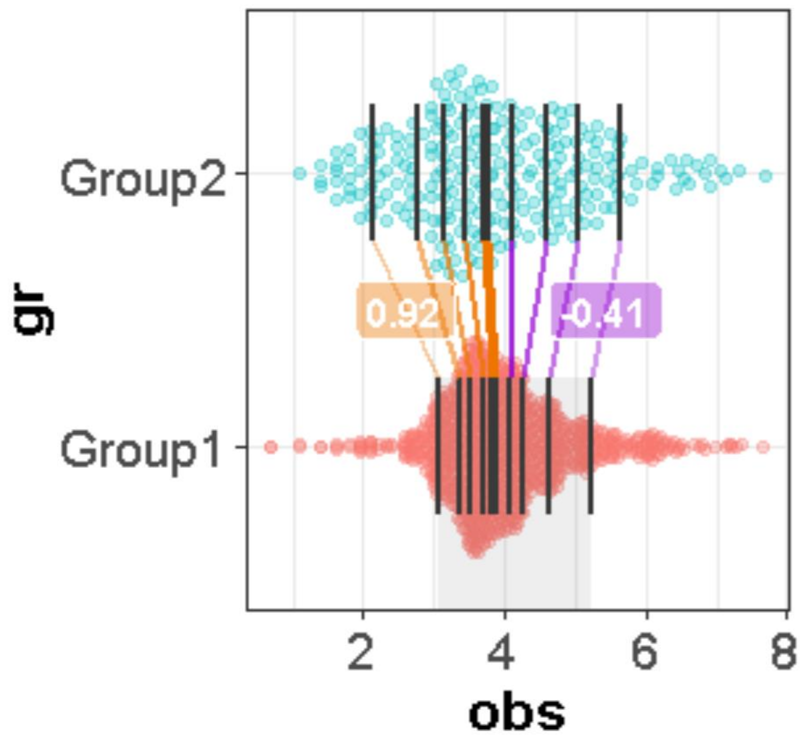




None of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 132856, p-value = 0.05122
alternative hypothesis: true location shift is not equal to 0
```

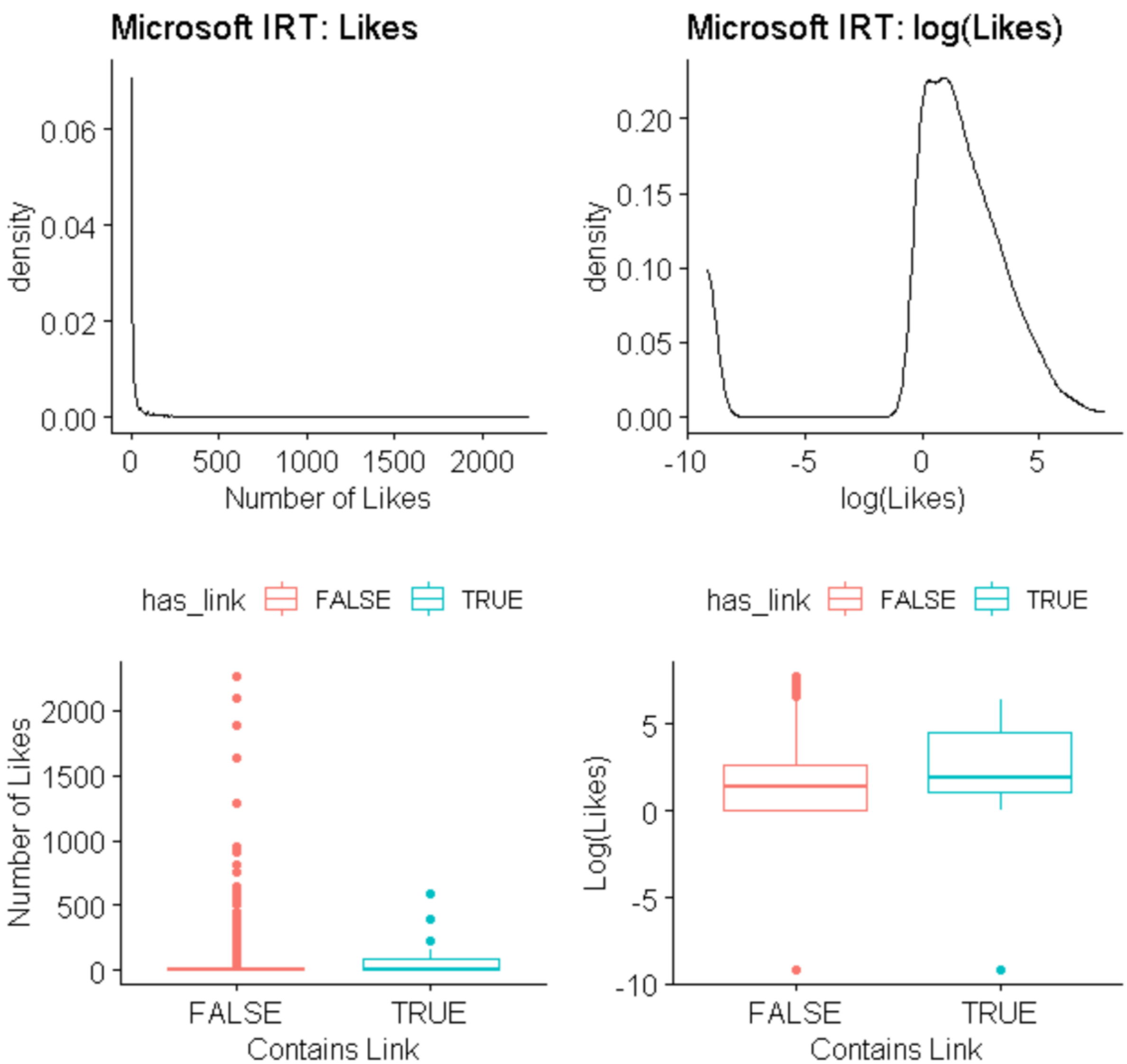
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of both populations are equal. Performing a shift function yields the following:

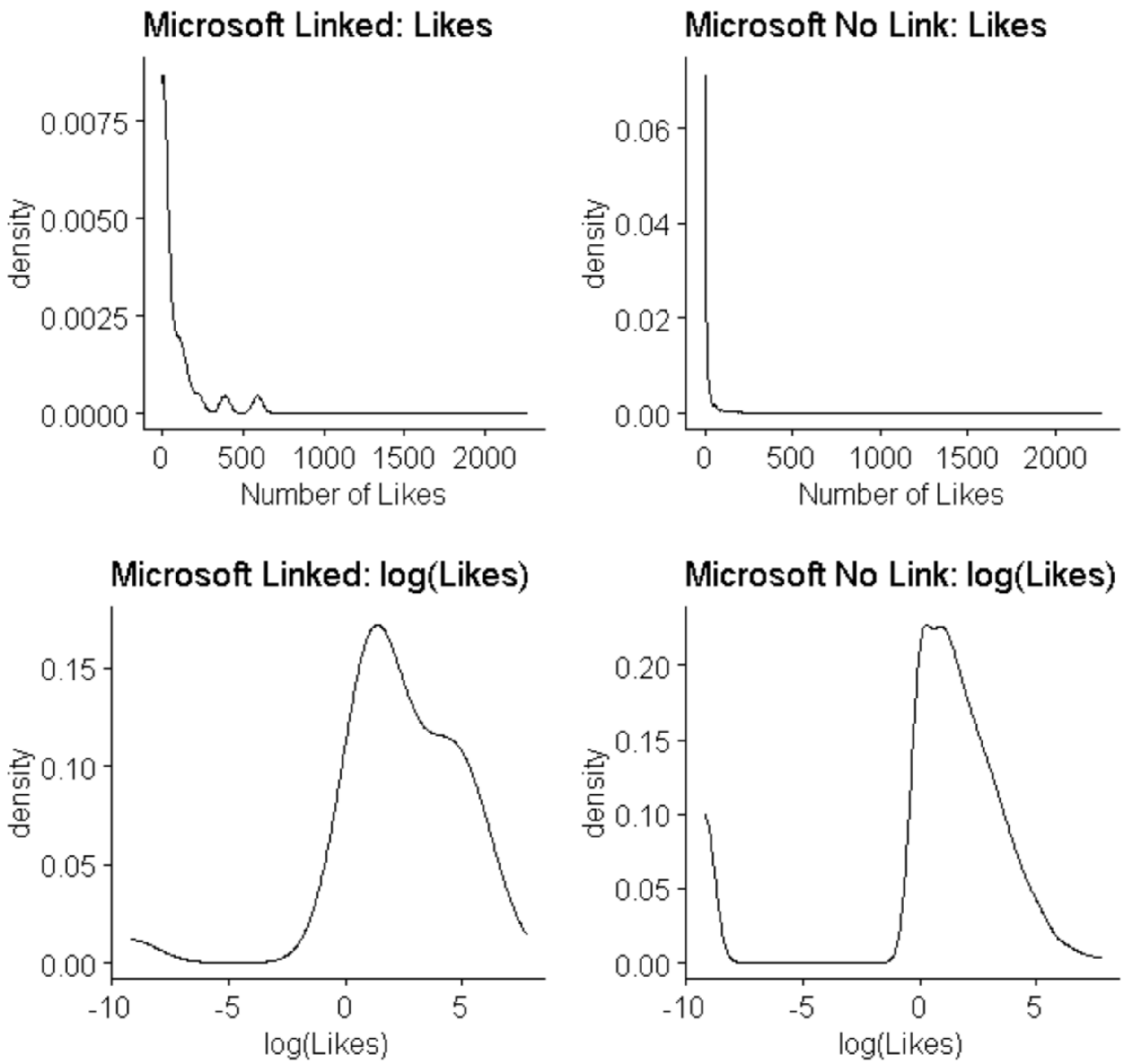



```
$`Group1 - Group2`
    q     Group1     Group2 difference      ci_lower  ci_upper      p_crit p_value
1 0.1   21.13591   8.445273  12.690638     8.5280755 14.984053 0.007142857   0.000
2 0.2   28.42145  16.087014  12.334440     6.0019065 17.796645 0.006250000   0.000
3 0.3   34.13847  23.464957  10.673512     4.1129264 15.714111 0.005555556   0.000
4 0.4   39.78808  30.925287   8.862792     0.7557015 15.144348 0.012500000   0.003
5 0.5   47.39474  42.157116   5.237621    -6.7551066 14.168148 0.025000000   0.262
6 0.6   57.89766  61.147525  -3.249870   -22.0658948  9.647708 0.050000000   0.639
7 0.7   71.13161  98.886580 -27.754966   -55.7715968  3.082336 0.016666667   0.028
8 0.8  103.21881 153.757182 -50.538373 -107.7761201 -7.289268 0.008333333   0.001
9 0.9  185.36325 283.702093 -98.338842 -324.5642479 -9.643746 0.010000000   0.006
```

We can say, with 95% confidence, that the 1st through 4th quantiles of group 2 (Microsoft OT tweets containing links) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Microsoft OT tweets not containing links). Furthermore, the 8th and 9th quantiles of group 2 would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1. Normally, I would take this as an indication that quantile differences exist between the two groups. However, considering we're lacking a statistically significant Mann-Whitney U test, I don't feel quite comfortable doing so. **Inclusion of a link does not seem to have a statistically significant effect on the number of retweets which a Microsoft official tweet receives.**

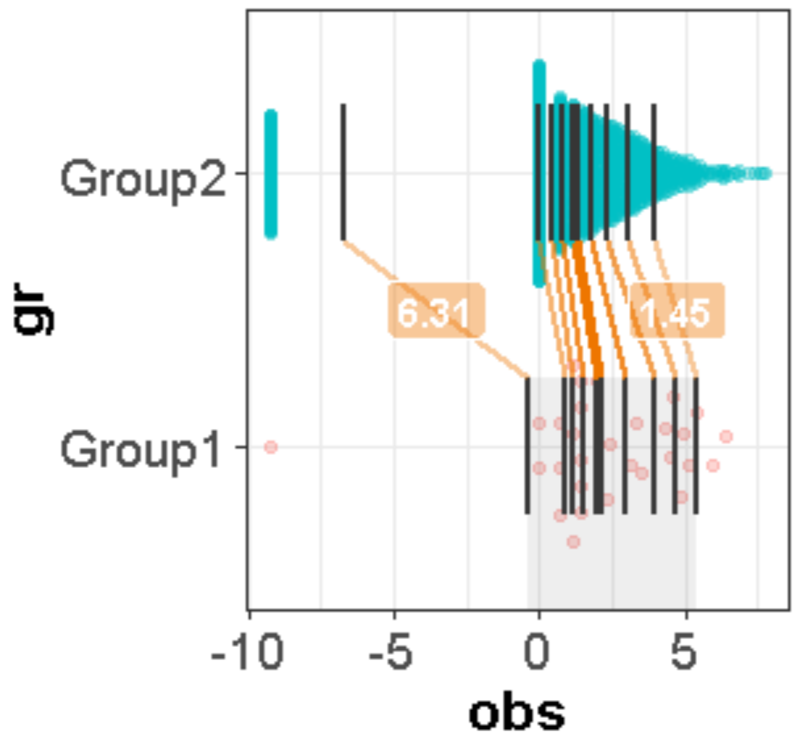
Microsoft IRT: Number of Likes






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 24700, p-value = 0.008108
alternative hypothesis: true location shift is not equal to 0
```

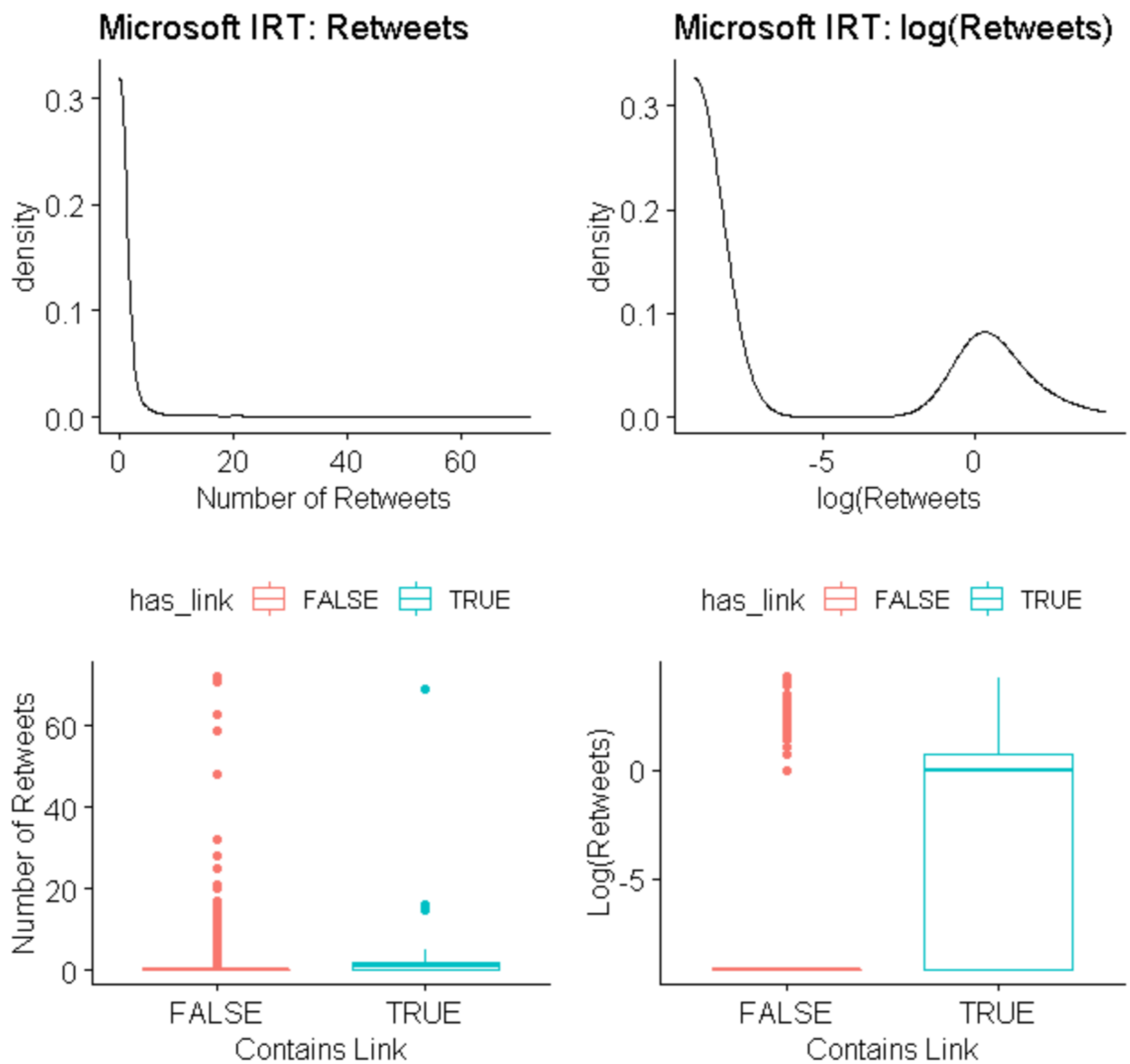
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

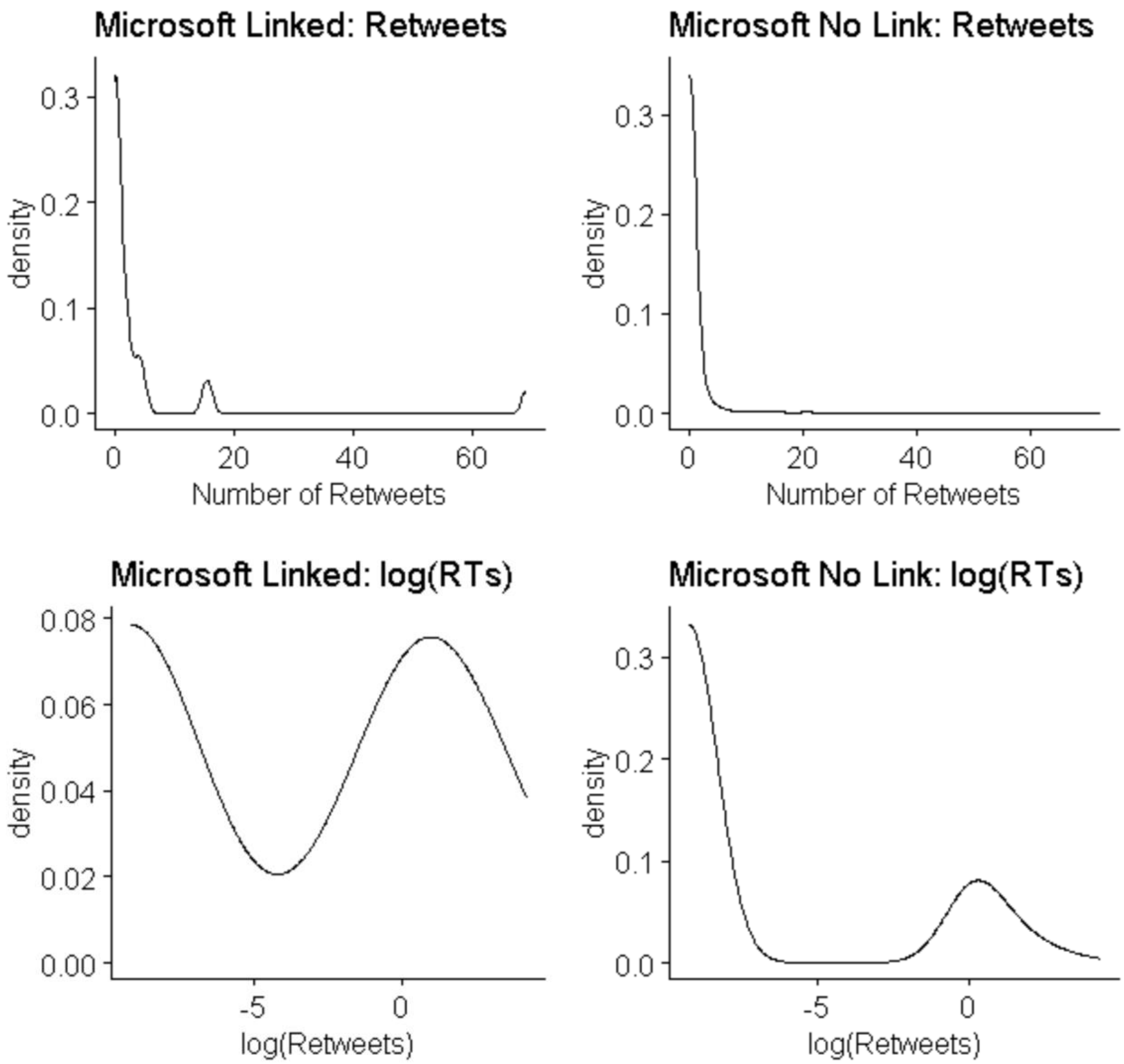



```
$`Group1 - Group2`
    q      Group1     Group2 difference    ci_lower   ci_upper      p_crit p_value
1 0.1    1.379483  0.2705111   1.108972 -0.007337394   2.426517 0.050000000   0.051
2 0.2    2.430848  1.0000000   1.430848  0.103086193   4.710982 0.010000000   0.003
3 0.3    3.392155  1.6491316   1.743023  0.045768289  11.633068 0.008333333   0.007
4 0.4    4.997148  2.1563145   2.840834  0.251950029  34.853791 0.006250000   0.002
5 0.5   10.714366  3.5984345   7.115931 -0.135881614  61.894597 0.016666667   0.028
6 0.6   28.451529  5.8705488  22.580981 -0.526388057  92.288551 0.025000000   0.035
7 0.7   65.952808 10.5882789  55.364529 -0.696093547 159.236291 0.012500000   0.016
8 0.8  121.283186 21.0487013 100.234484  2.562996565 318.800928 0.007142857   0.004
9 0.9  260.226397 53.5898280 206.636569 20.850825416 497.054987 0.005555556   0.002
```

Honestly, there are only 29 Microsoft IRT tweets containing links. I probably shouldn't have gone through with this analysis for Microsoft IRT data, there are 1325 tweets in the no link group. Disregard these results, as my conclusions are that the sample sizes differ too greatly for fair comparison.

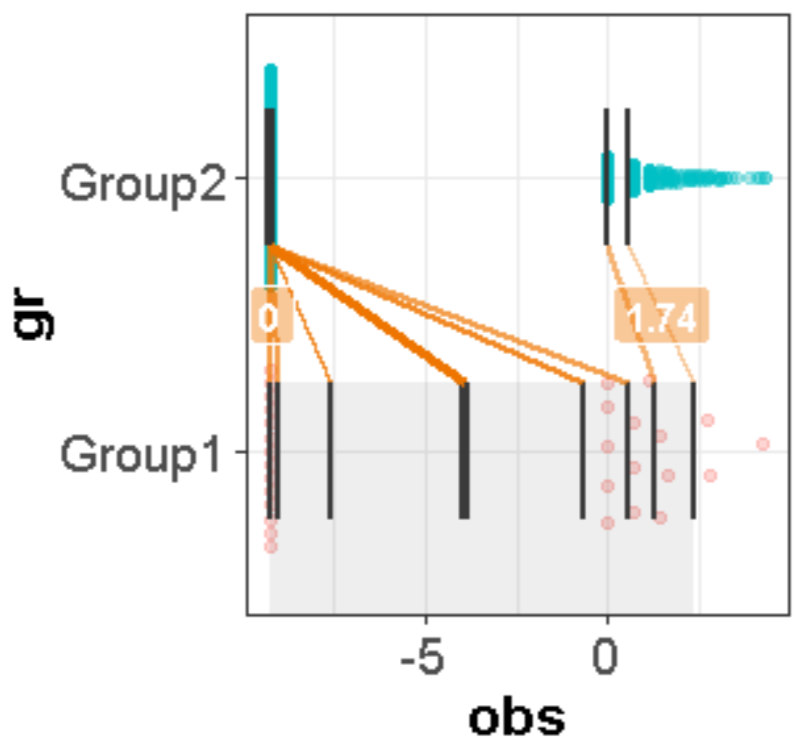
Microsoft IRT: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 25095, p-value = 0.0001898
alternative hypothesis: true location shift is not equal to 0
```

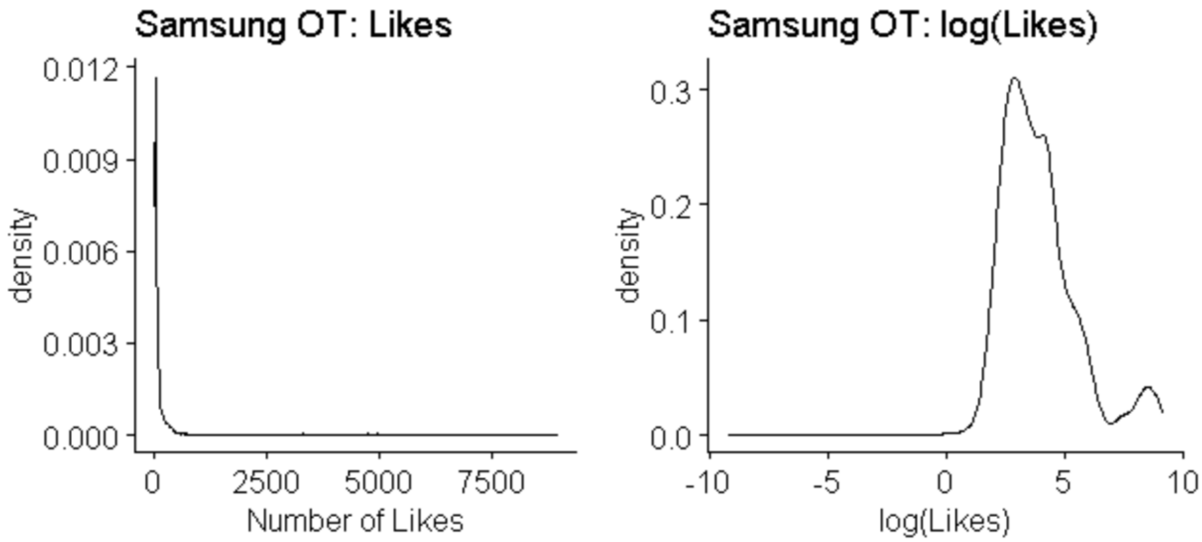
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



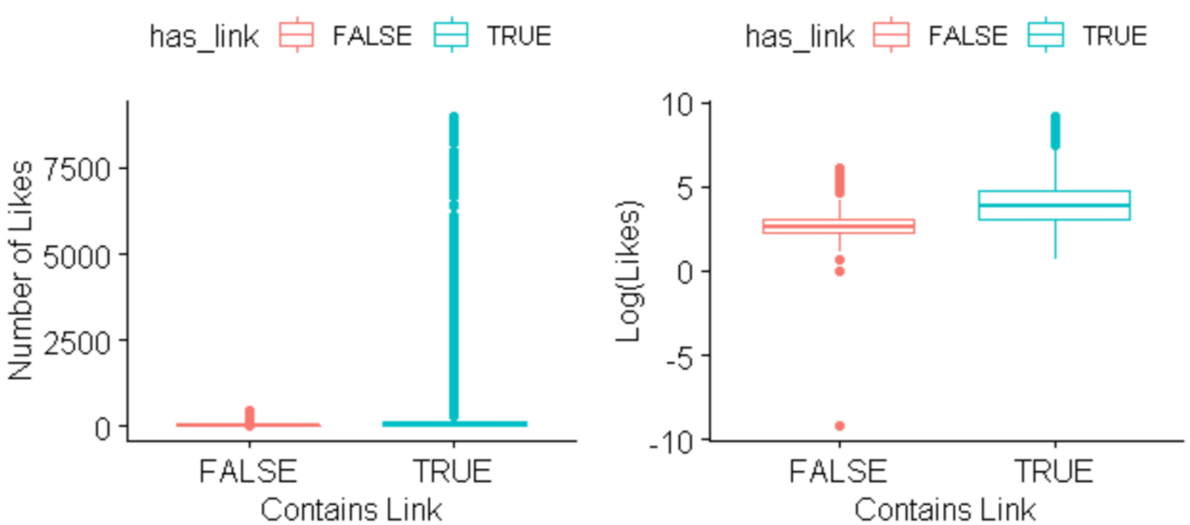

```
$`Group1 - Group2`
    q       Group1       Group2   difference     ci_lower    ci_upper      p_crit p_value
1 0.1 1.900567e-06 0.000000e+00 1.900567e-06 3.797851e-12  0.01702519 0.016666667   0.000
2 0.2 5.252839e-04 0.000000e+00 5.252839e-04 9.641442e-10  0.16282829 0.012500000   0.000
3 0.3 1.922214e-02 0.000000e+00 1.922214e-02 6.743519e-07  0.63887740 0.010000000   0.000
4 0.4 1.788713e-01 0.000000e+00 1.788713e-01 1.080200e-04  1.29812827 0.008333333   0.000
5 0.5 6.176710e-01 0.000000e+00 6.176710e-01 5.039510e-03  2.34319480 0.007142857   0.000
6 0.6 1.218085e+00 0.000000e+00 1.218085e+00 9.518052e-02  3.69367138 0.006250000   0.000
7 0.7 2.120295e+00 8.763325e-07 2.120294e+00 4.095127e-01  8.75745548 0.005555556   0.000
8 0.8 4.238132e+00 9.998848e-01 3.238248e+00 5.900592e-01 14.73753323 0.050000000   0.002
9 0.9 1.501861e+01 1.832641e+00 1.318596e+01 1.332877e+00 54.37409362 0.025000000   0.002
```

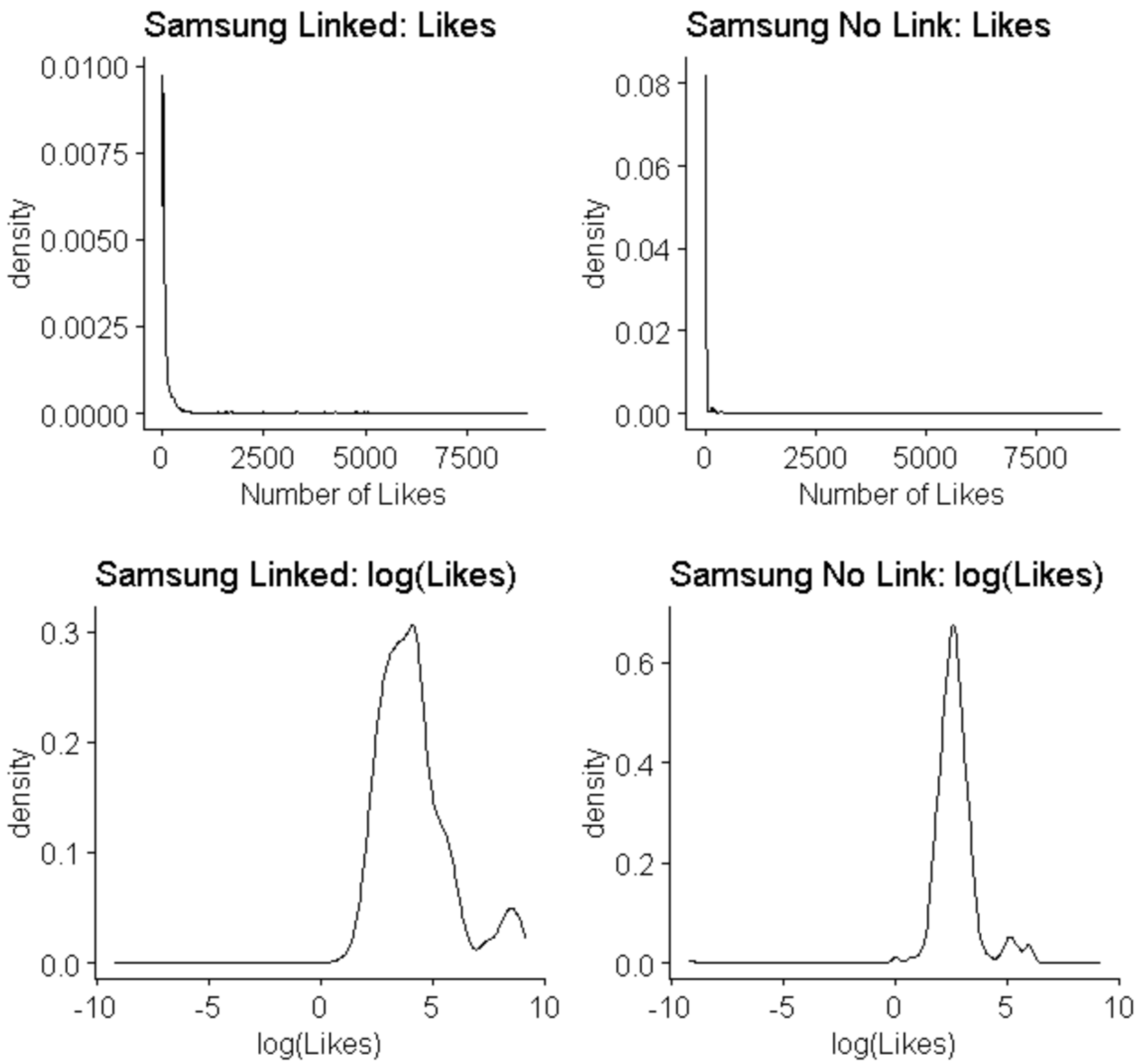
Again, I probably shouldn't have performed this analysis. The sample sizes between the two groups vary too widely for fair comparison.

Samsung Official: Number of Likes



The log distribution does not pass a Shapiro-Wilk normality test.

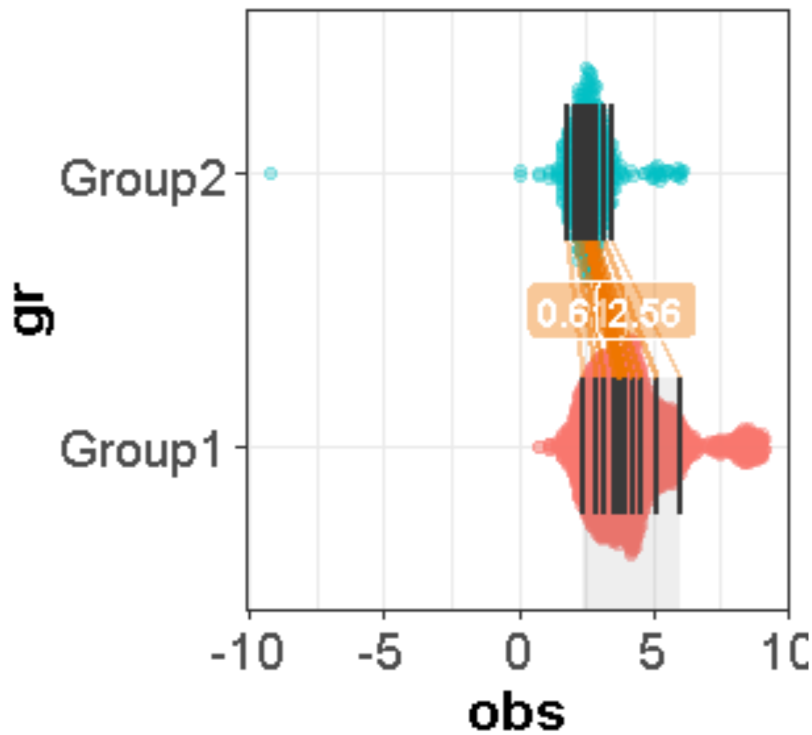




None of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Likes` and noLink_tweets$`Number of Likes`
W = 931063, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

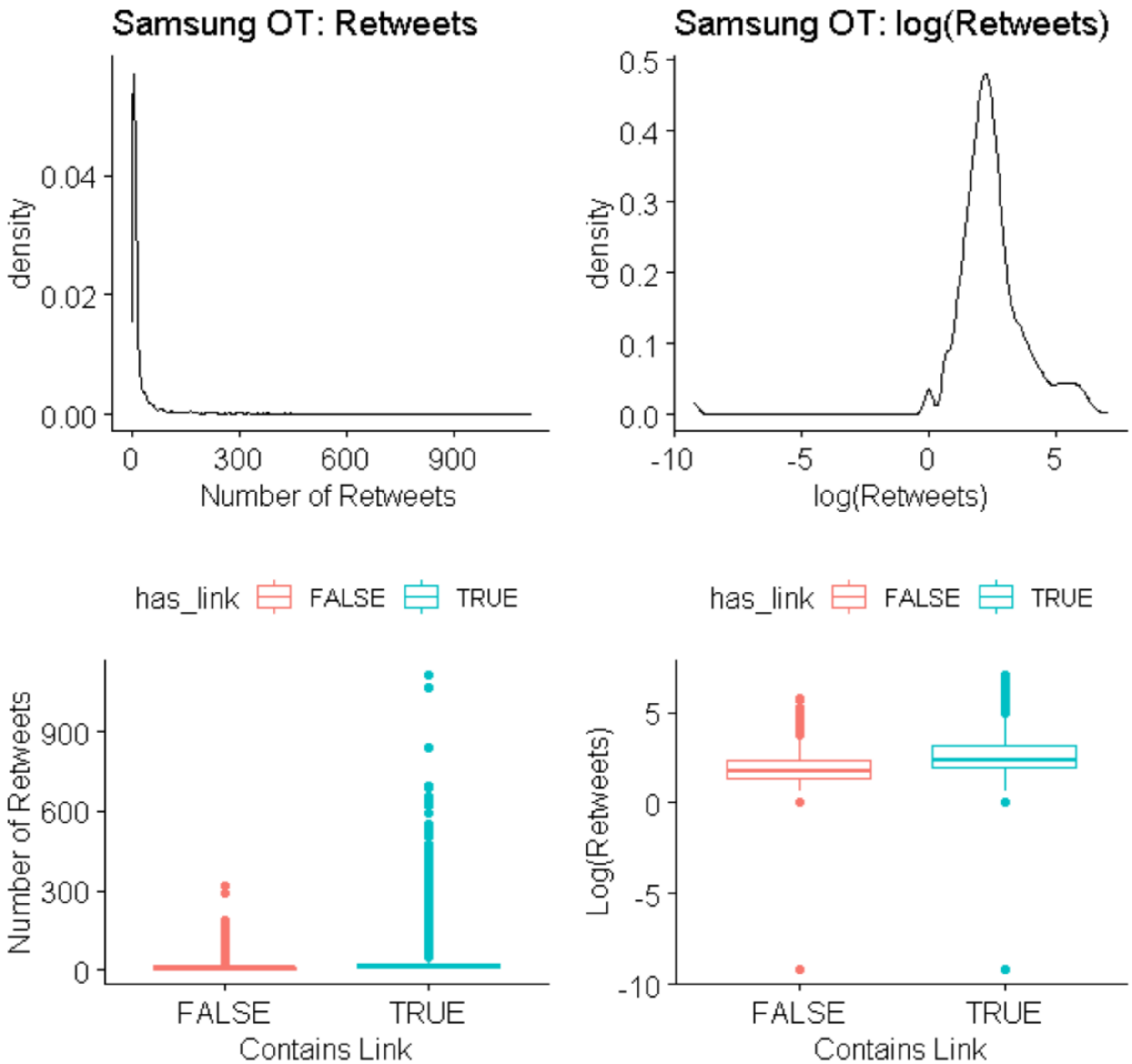
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

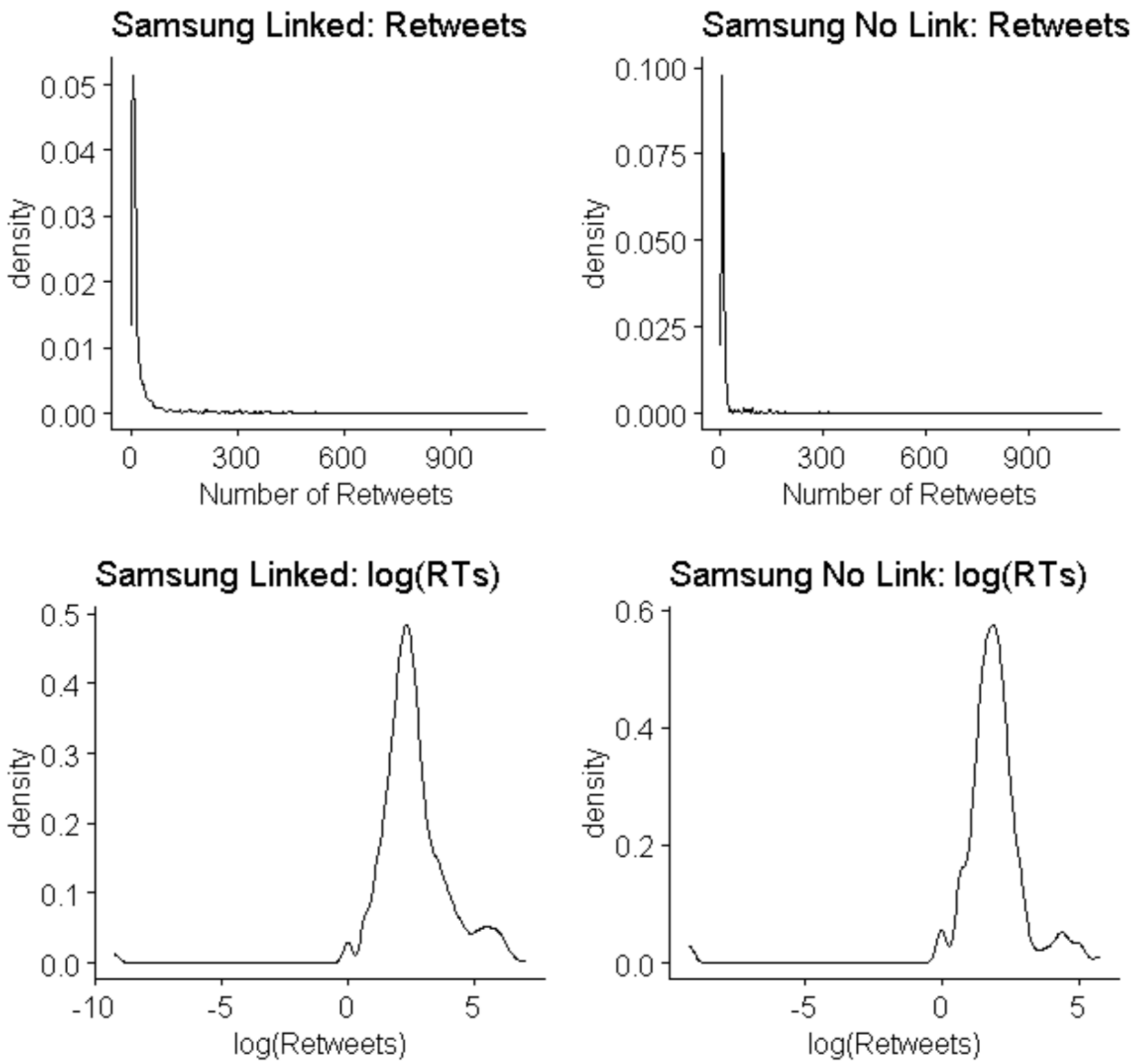



```
$`Group1 - Group2`
    q    Group1    Group2 difference   ci_lower   ci_upper     p_crit p_value
1 0.1  11.14318  6.085471   5.057712   4.117697   5.840121 0.050000000       0
2 0.2  16.97395  8.150222   8.823732   7.383972  10.372997 0.025000000       0
3 0.3  23.93147  9.637409  14.294064  12.268662  16.520260 0.016666667       0
4 0.4  34.28844 11.560194  22.728251  19.893832  25.591937 0.012500000       0
5 0.5  48.45508 13.120641  35.334439  30.629572  39.775449 0.010000000       0
6 0.6  65.79747 15.375862  50.421605  45.989665  55.495082 0.008333333       0
7 0.7  92.60555 17.957560  74.647988  65.608094  84.375097 0.007142857       0
8 0.8 172.50604 22.749330 149.756707 121.340581 177.883622 0.006250000       0
9 0.9 410.98344 31.843307 379.140133 301.364136 534.881804 0.005555556       0
```

We can say, with 95% confidence, that every quantile of group 2 (Samsung official tweets not containing links) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Samsung official tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a link in Samsung official tweets.**

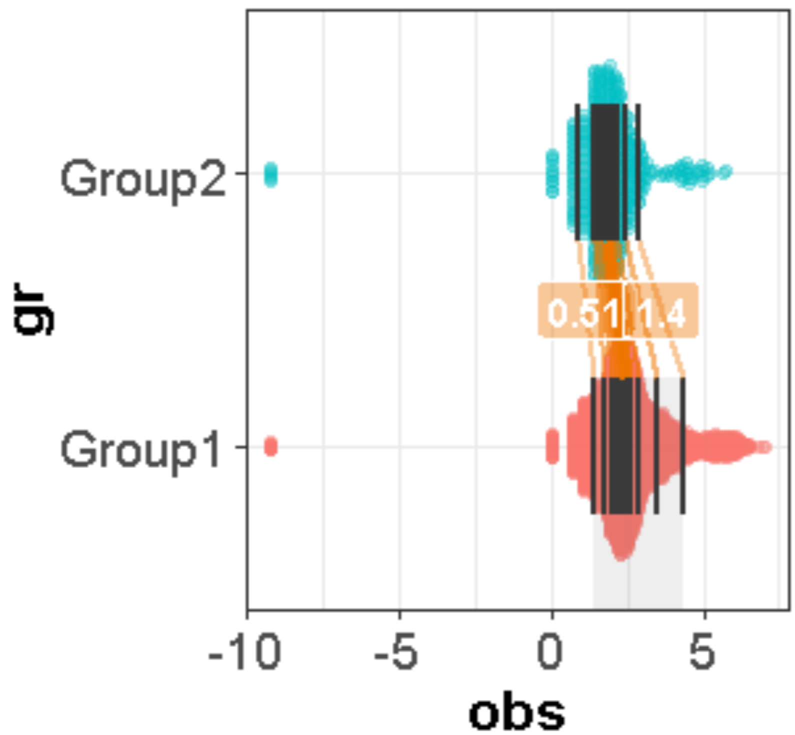
Samsung Official: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  link_tweets$`Number of Retweets` and noLink_tweets$`Number of Retweets`
W = 788304, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

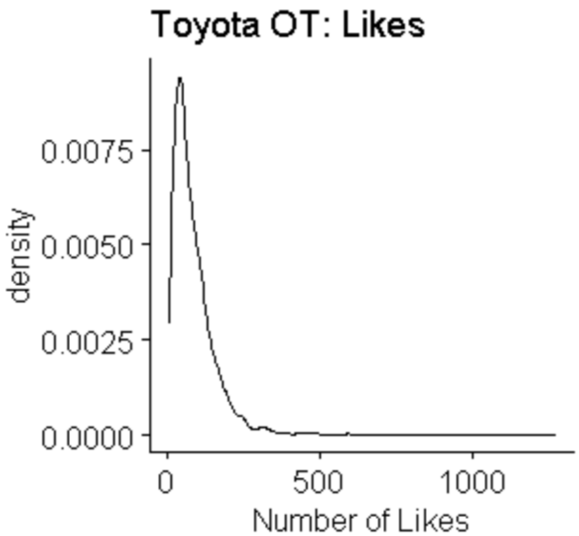
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

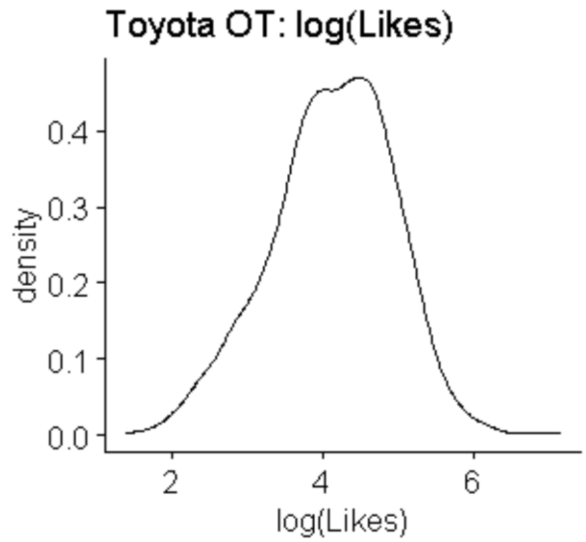



```
$`Group1 - Group2`
    q     Group1    Group2 difference   ci_lower   ci_upper      p_crit p_value
1 0.1   3.962812  2.417996   1.544816  0.7959389   1.946542 0.050000000       0
2 0.2   5.849140  3.914806   1.934334  1.2120792   2.712619 0.025000000       0
3 0.3   7.262157  4.562253   2.699904  2.0582176   3.704496 0.016666667       0
4 0.4   9.054108  5.373396   3.680712  3.0408014   4.584758 0.012500000       0
5 0.5  11.081243  6.409622   4.671621  3.9023891   5.708025 0.010000000       0
6 0.6  13.898680  7.696898   6.201782  5.0849442   7.484370 0.008333333       0
7 0.7  18.450156  9.157794   9.292362  7.3827601  11.293262 0.007142857       0
8 0.8  31.135107 11.674935  19.460173 14.4435797  24.455666 0.006250000       0
9 0.9  74.757975 18.488465  56.269510 37.0746690  76.895562 0.005555556       0
```

With 95% confidence we can say that every quantile of group 2 (Samsung official tweets not containing links) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Samsung official tweets containing links). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets), and potentially one underlying factor explaining these differences is the inclusion of a link in Samsung official tweets.**

Toyota Official: Number of Likes





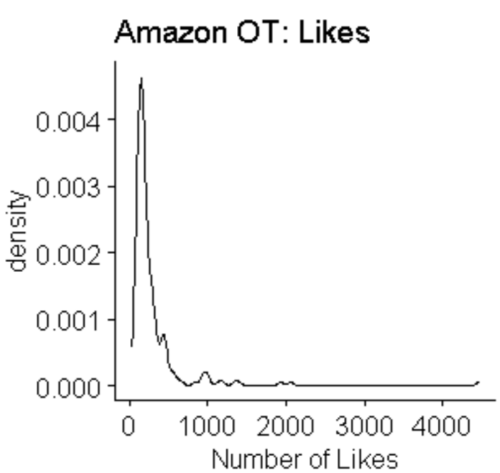
The log distribution does not pass a Shapiro-Wilk normality test.

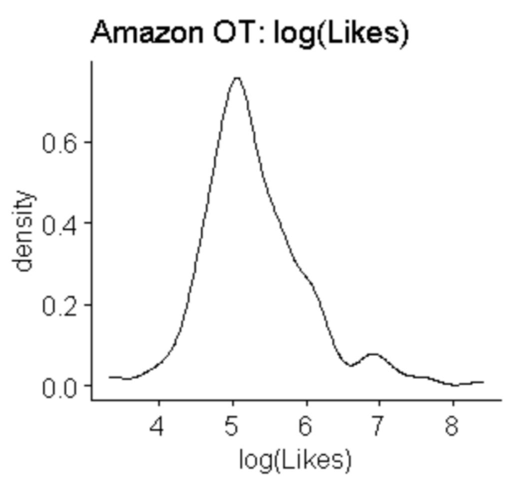
It appears that Toyota only has 4 official tweets (not considered extreme outliers at least) which do not contain links, which doesn't seem to be enough for analysis.

<u>Section 8:</u> Hashtag Inclusion Analysis

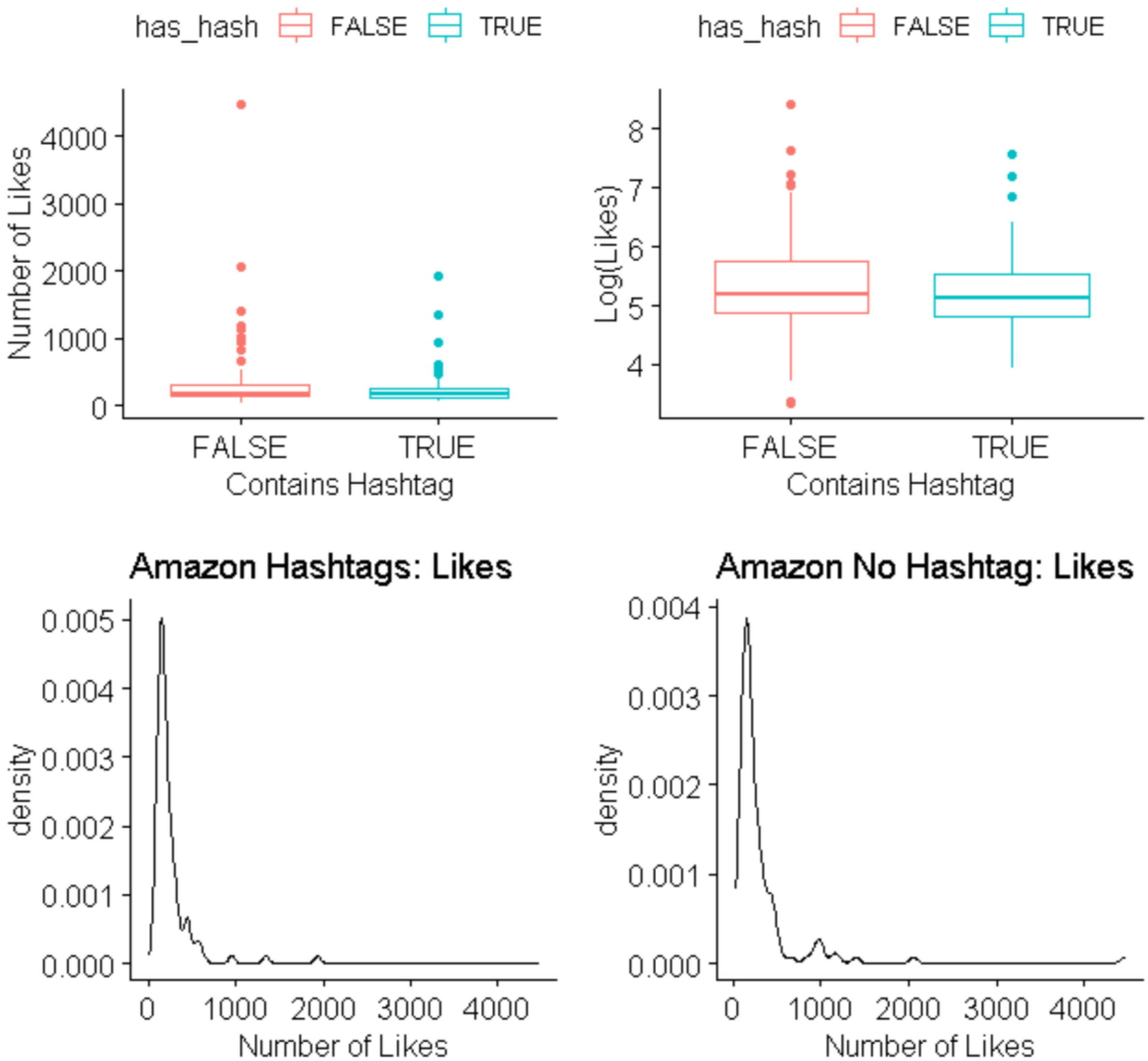
Section 8 contains the visualizations, notes, results, and conclusions made when examining companies and tweet categories for a potential hashtag inclusion effect.

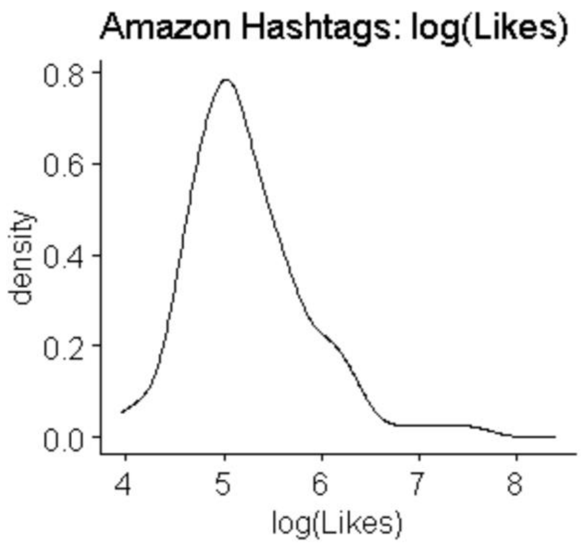
Amazon Official: Number of Likes

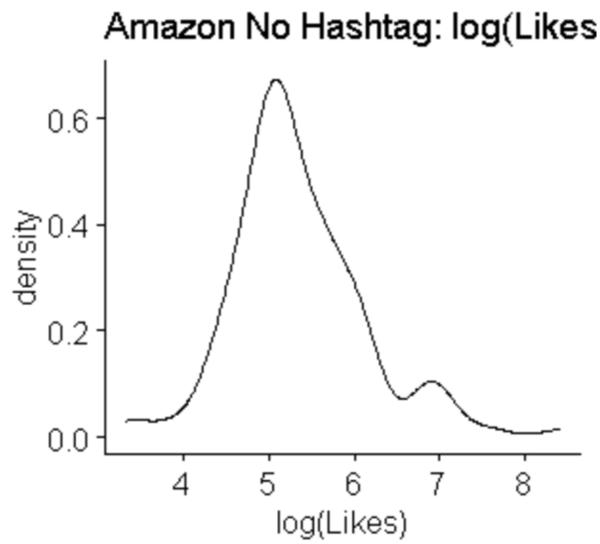




The log distribution does not pass a Shapiro-Wilk normality test.



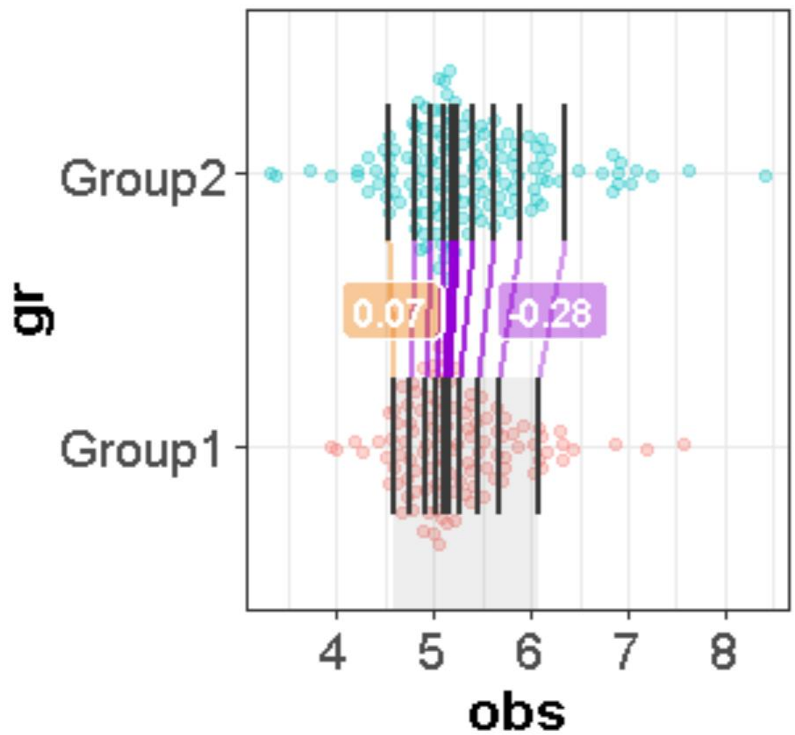




None of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 6742, p-value = 0.2246
alternative hypothesis: true location shift is not equal to 0
```

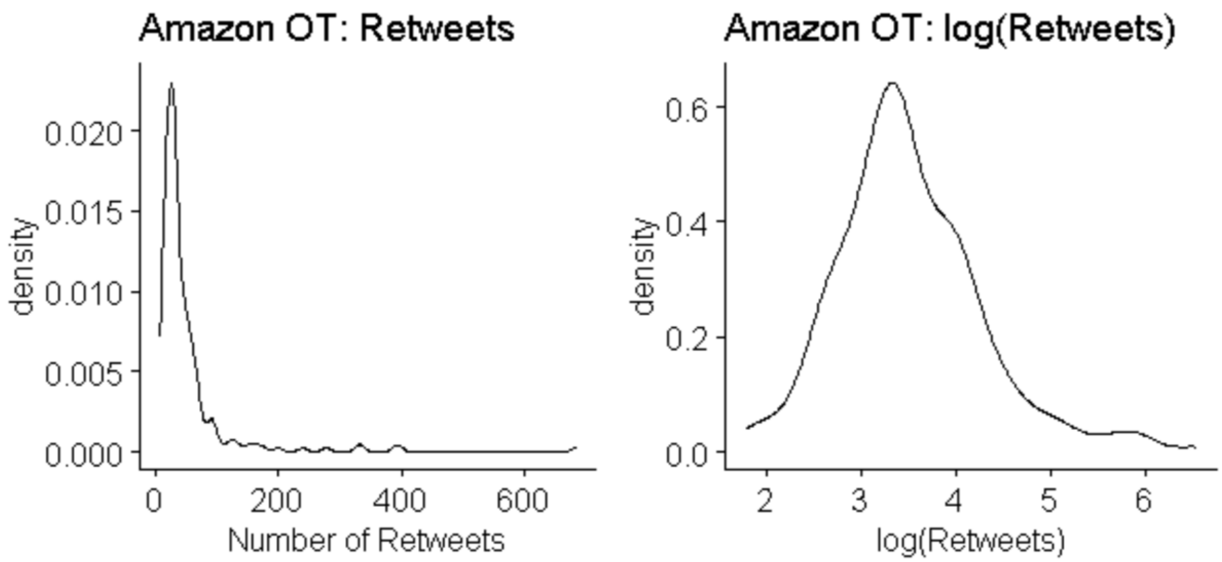
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'like' distributions of both populations are equal. Performing a shift function yields the following:



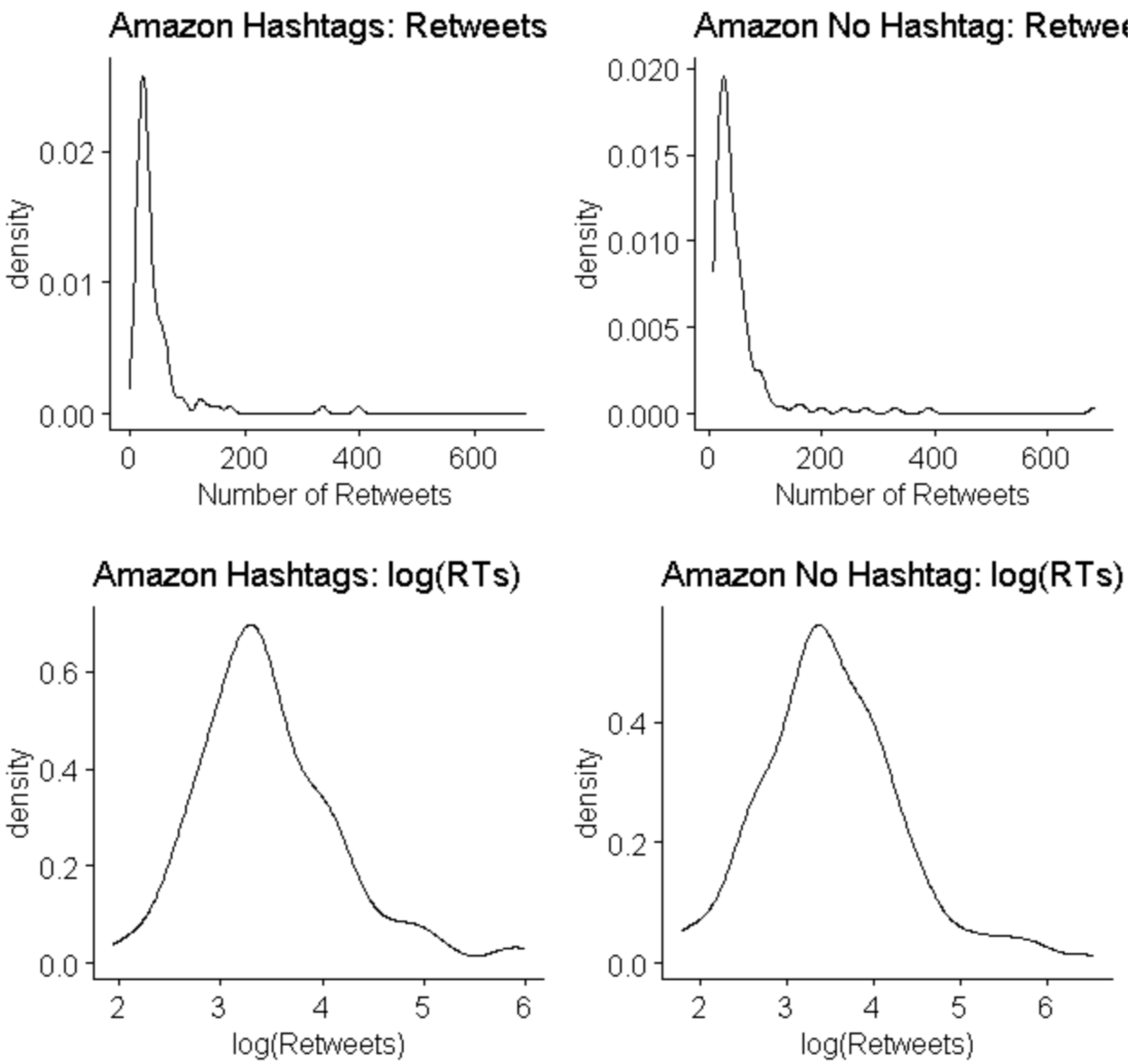

```
$`Group1 - Group2`
    q    Group1    Group2  difference   ci_lower  ci_upper      p_crit p_value
1 0.1  99.66126  92.79311    6.868148  -17.36612  28.52736 0.016666667   0.459
2 0.2 117.04043 122.84167   -5.801245  -23.24120  16.78218 0.050000000   0.584
3 0.3 135.79790 143.86414   -8.066245  -31.48238  16.37033 0.025000000   0.471
4 0.4 152.81510 163.57613  -10.761037  -38.62639  14.52081 0.008333333   0.242
5 0.5 167.72241 182.08411  -14.361694  -53.70326  23.23349 0.012500000   0.270
6 0.6 195.52539 221.35350  -25.828110  -88.19131  32.09096 0.010000000   0.266
7 0.7 232.85734 278.08510  -45.227759 -130.65913  32.44319 0.006250000   0.147
8 0.8 294.76100 366.40091  -71.639914 -219.82088  62.63146 0.005555556   0.158
9 0.9 436.48715 598.12946 -161.642309 -566.94330 127.34677 0.007142857   0.155
```

Here, group 1 represents Amazon official tweets containing hashtags, while group 2 represents Amazon official tweets not containing hashtags. Considering each confidence interval contains the value 0, we may not conclude any quantile differences in tweet performance between the two groups. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of likes are equal to one another. **Inclusion of a hashtag does not seem to have a statistically significant effect on the number of likes which an Amazon official tweet receives.**

Amazon Official: Number of Retweets



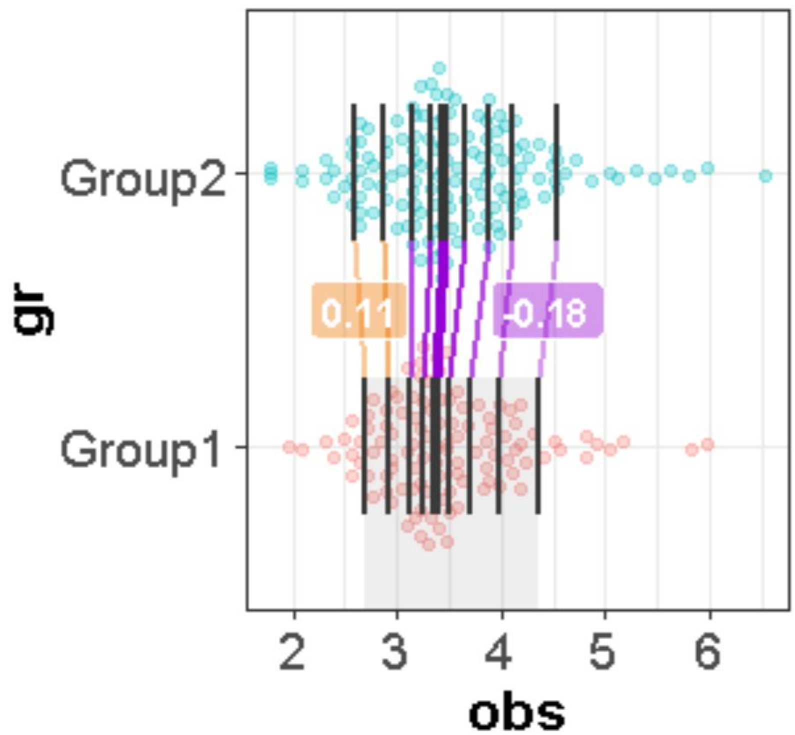
The log distribution does not pass a Shapiro-Wilk normality test.



None of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 6975.5, p-value = 0.4289
alternative hypothesis: true location shift is not equal to 0
```

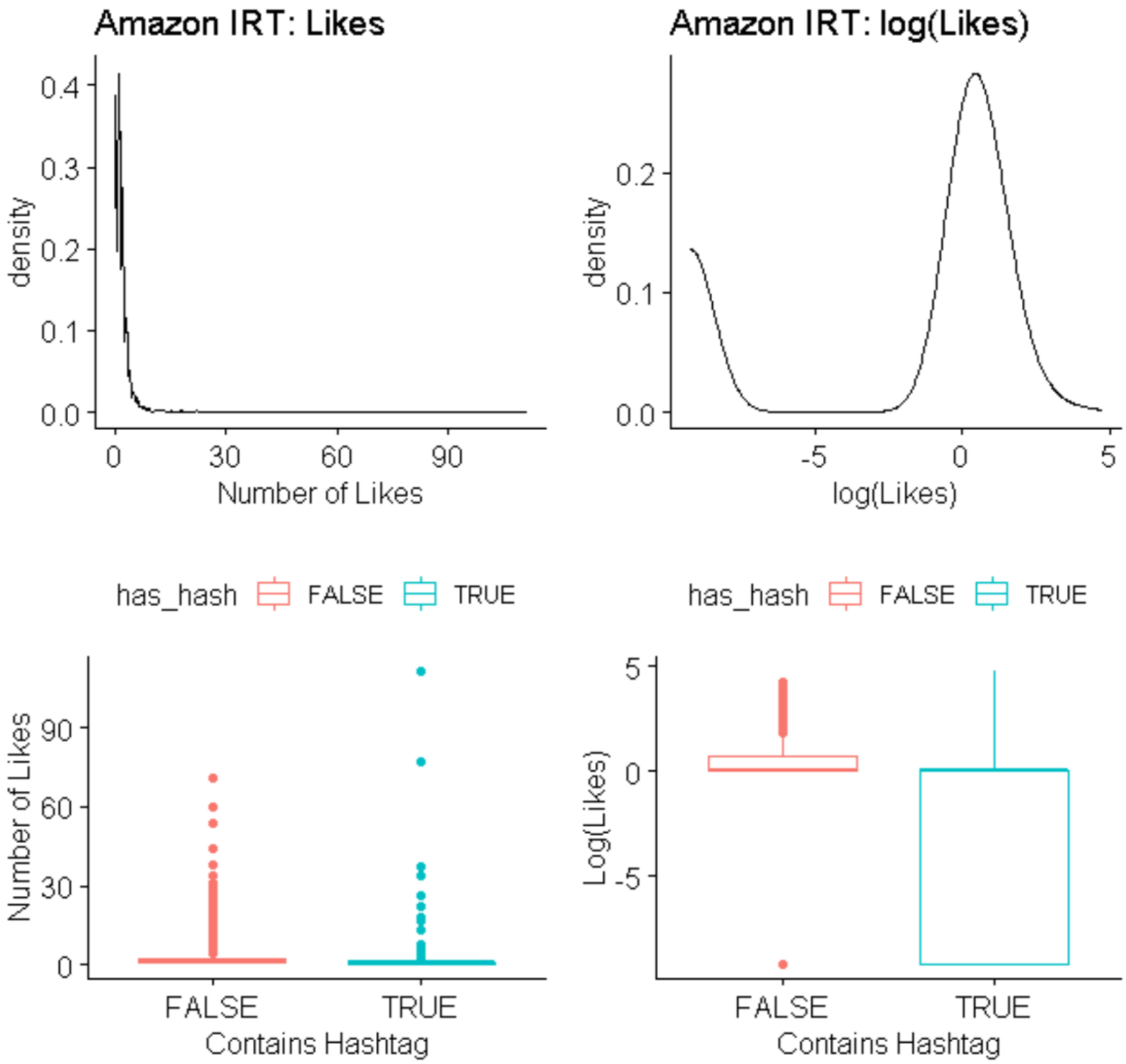
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of both populations are equal. Performing a shift function yields the following:

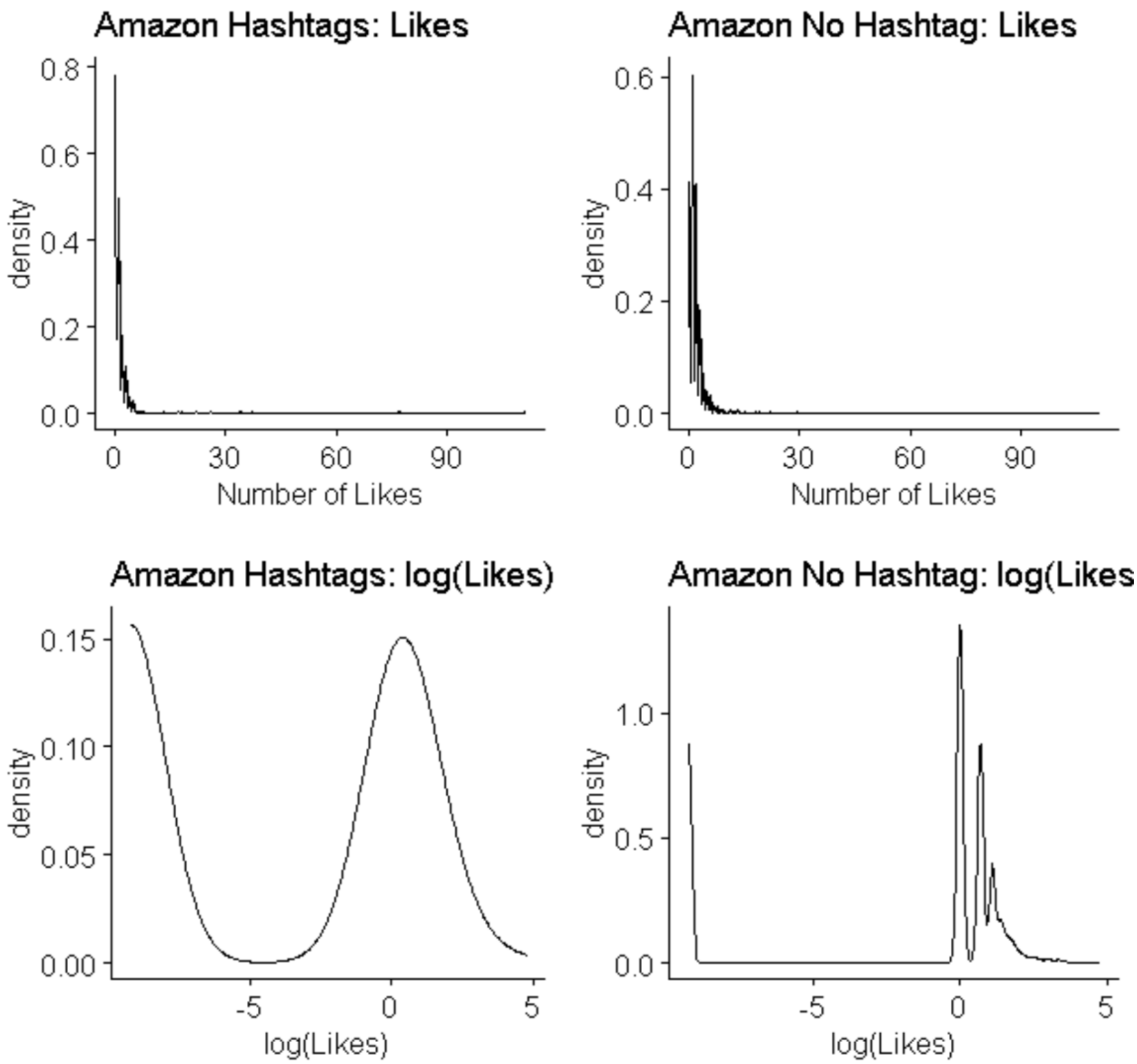



```
$`Group1 - Group2`
    q   Group1   Group2  difference   ci_lower  ci_upper      p_crit p_value
1 0.1 14.72986 13.24025   1.4896079  -2.169425  5.852477 0.008333333   0.237
2 0.2 18.68360 17.86466   0.8189457  -4.100244  5.185249 0.050000000   0.762
3 0.3 22.58922 23.41665  -0.8274287  -5.681657  4.223524 0.025000000   0.708
4 0.4 25.64961 27.67190  -2.0222986  -7.341593  3.305281 0.012500000   0.386
5 0.5 29.05746 31.55326  -2.4958017 -11.021125  3.649153 0.007142857   0.298
6 0.6 33.15327 38.84687  -5.6935949 -17.979056  5.880831 0.006250000   0.207
7 0.7 41.11491 48.86865  -7.7537393 -22.716966  8.199363 0.005555556   0.198
8 0.8 54.39280 61.17129  -6.7784957 -30.639223 10.796911 0.010000000   0.348
9 0.9 79.55389 94.39701 -14.8431154 -88.289663 45.065108 0.016666667   0.467
```

Considering each confidence interval contains the value 0, we may not conclude any quantile differences in tweet performance between the two groups. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of retweets are equal to one another. **Inclusion of a hashtag does not seem to have a statistically significant effect on the number of retweets which an Amazon official tweet receives.**

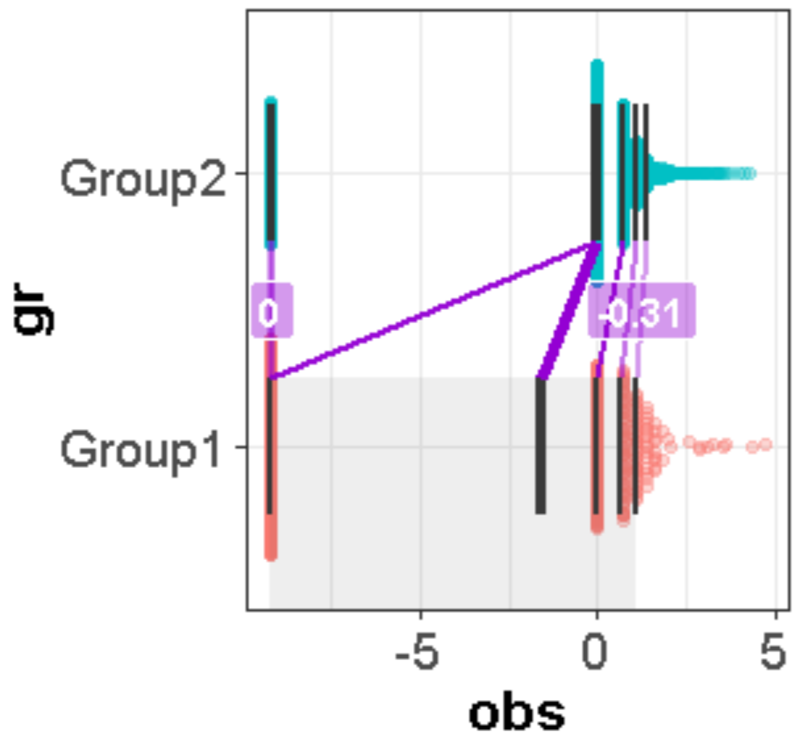
Amazon IRT: Number of Likes






```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 476326, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

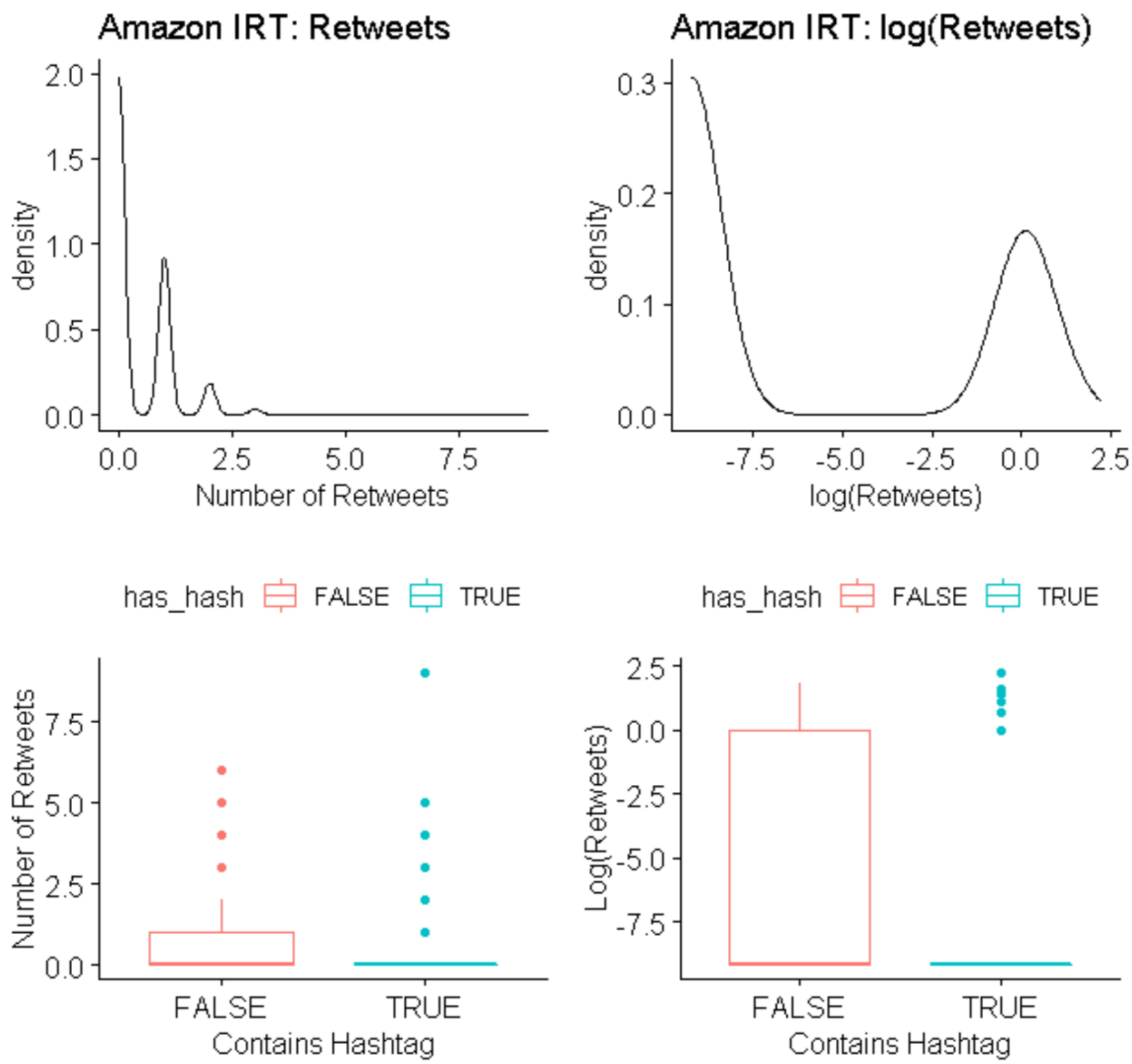
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

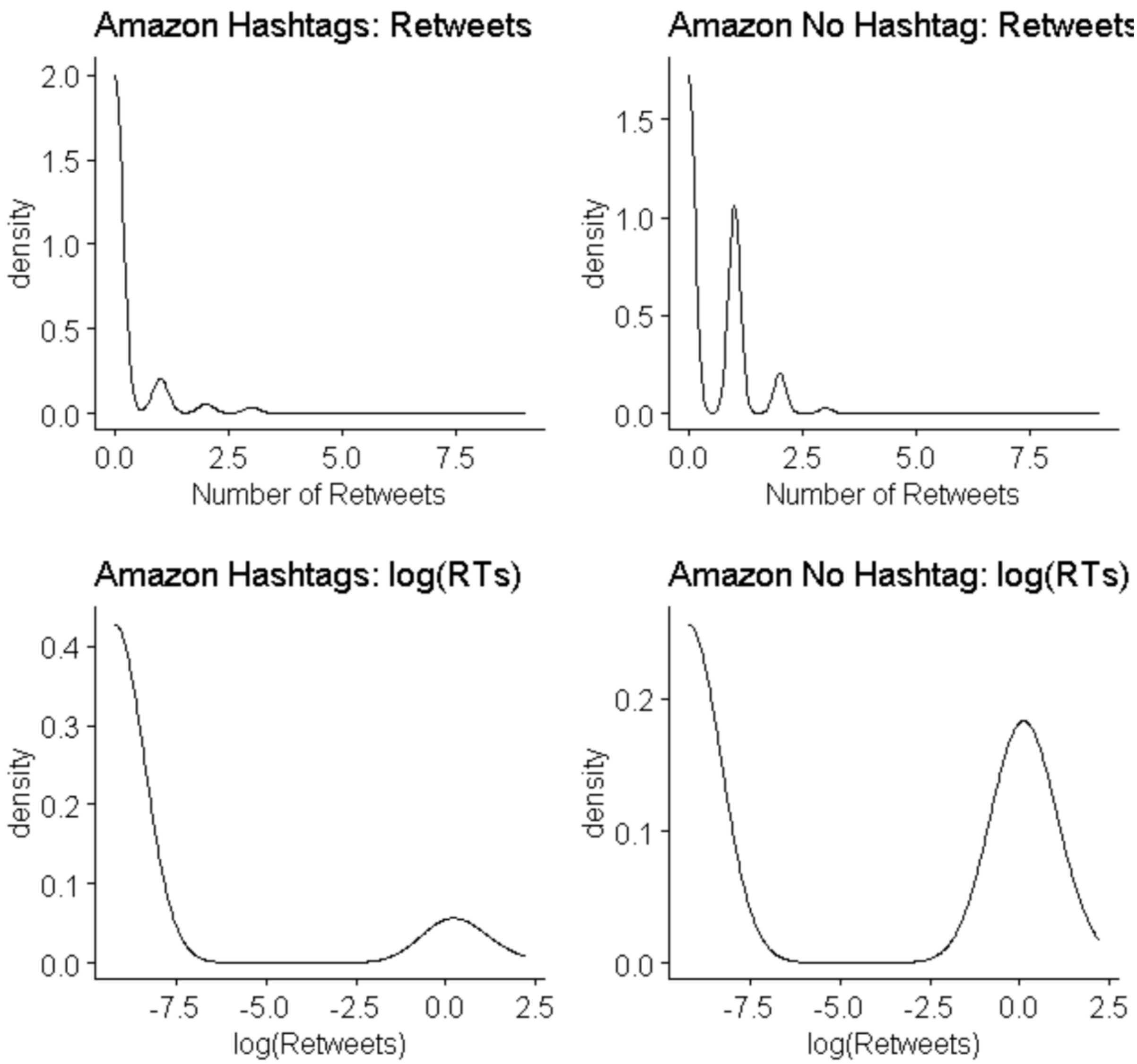



```
$`Group1 - Group2`
    q       Group1      Group2   difference   ci_lower      ci_upper      p_crit p_value
1 0.1 0.0000000000 0.000000000  0.000000000  0.0000000  0.000000e+00 0.050000000   1.000
2 0.2 0.0000000000 0.007372012 -0.007372012 -0.4214065 -2.802594e-06 0.025000000   0.000
3 0.3 0.0000000000 1.000000000 -1.000000000 -1.0000000 -1.000000e+00 0.016666667   0.000
4 0.4 0.0000341249 1.000000000 -0.999965875 -1.0000000 -9.369117e-01 0.012500000   0.000
5 0.5 0.8265178040 1.000000000 -0.173482196 -0.9349688 -3.258131e-04 0.010000000   0.000
6 0.6 0.9999999976 1.999709054 -0.999709056 -1.0003173 -7.817480e-01 0.008333333   0.000
7 0.7 1.0000457562 2.000000000 -0.999954244 -1.0000000 -9.067994e-01 0.007142857   0.000
8 0.8 1.9653258200 2.943479121 -0.978153301 -1.8118111 -1.904813e-01 0.006250000   0.000
9 0.9 2.9298916180 3.982939191 -1.053047573 -1.8864187 -2.913878e-01 0.005555556   0.002
```

From the confidence intervals we can say, with 95% confidence, that the 2nd through 9th quantiles of group 2 (Amazon IRT tweets not containing hashtags) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Amazon IRT tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Amazon IRT tweets.**

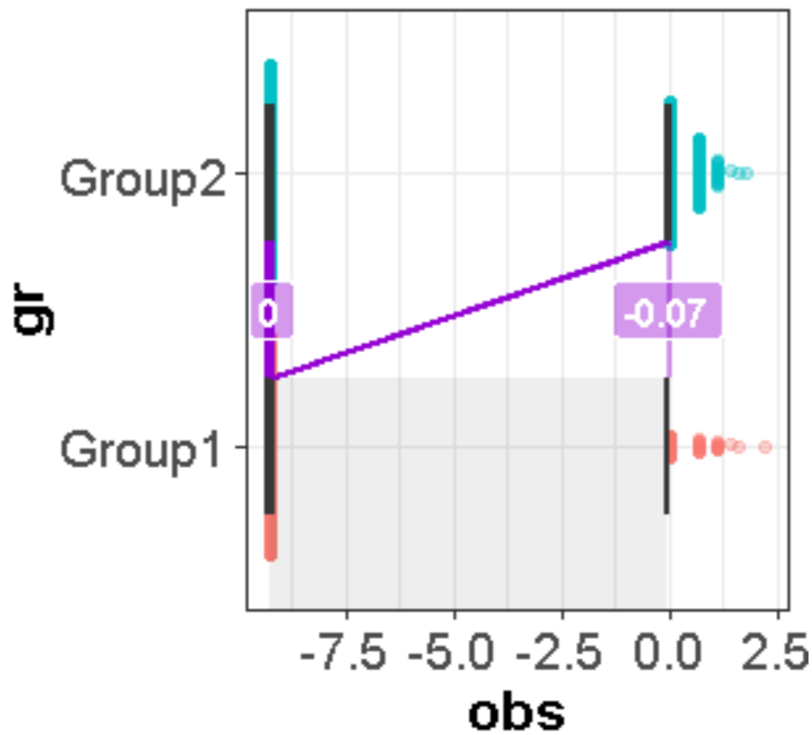
Amazon IRT: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 487593, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

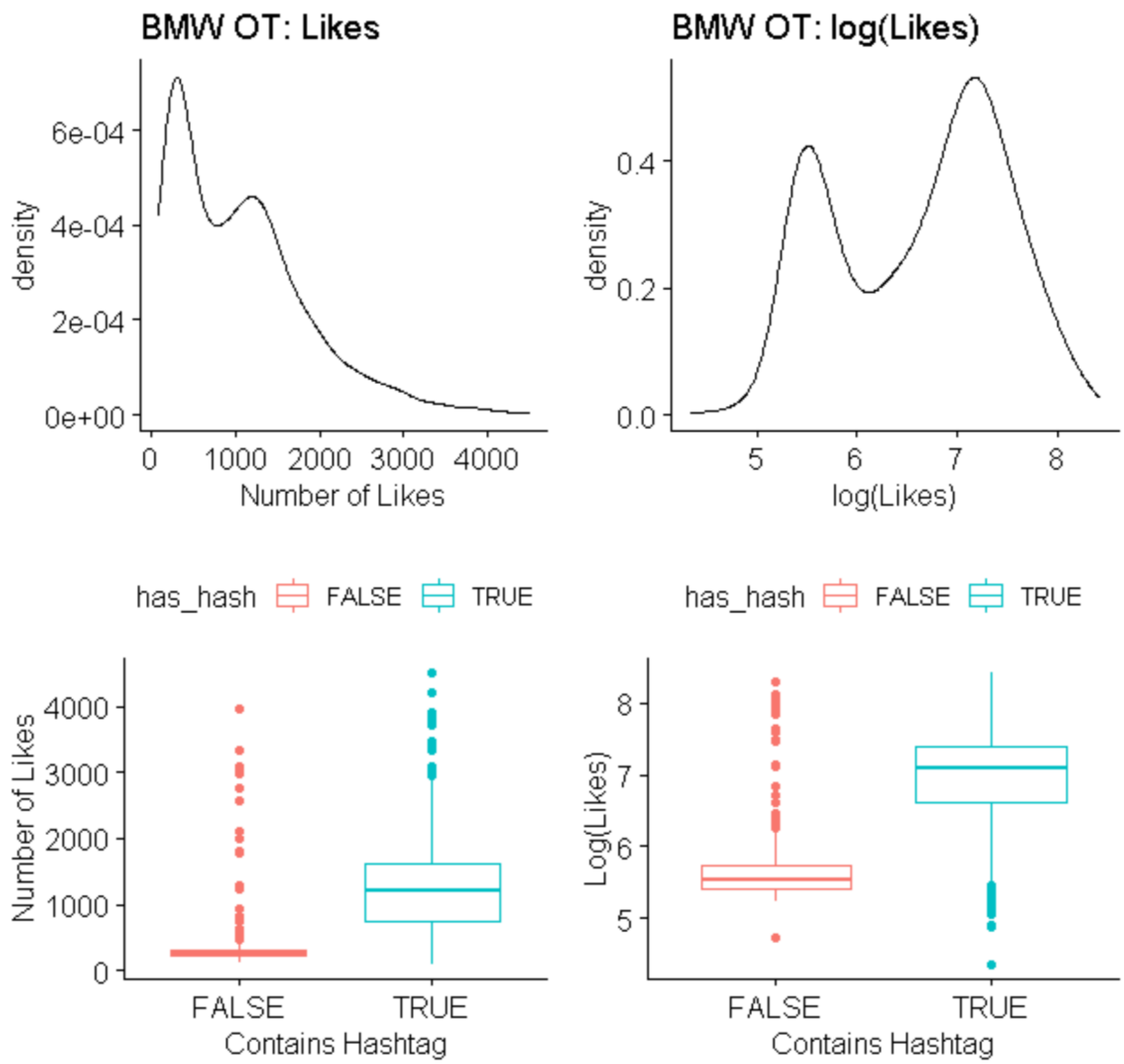


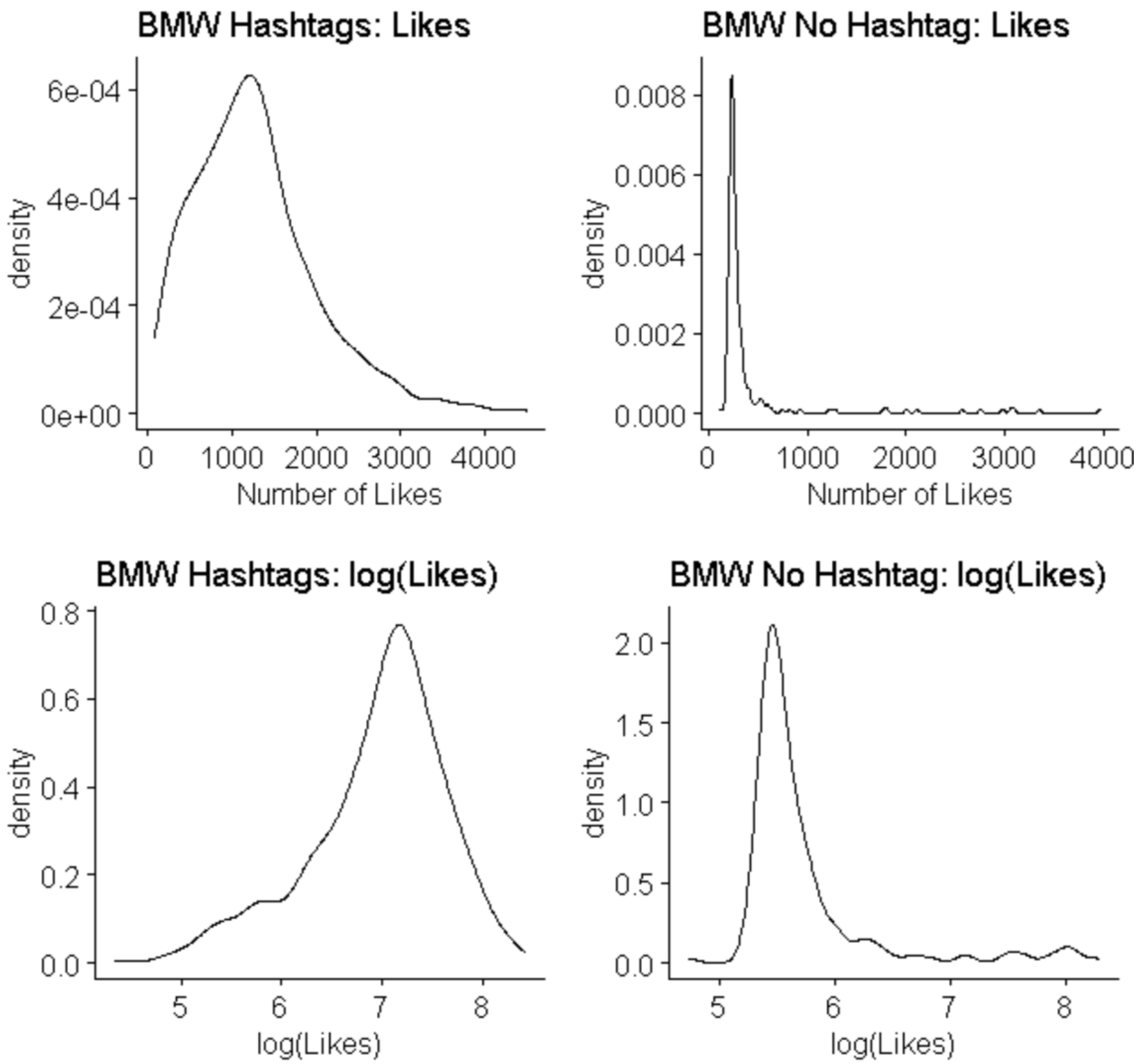
$`Group1 - Group2`

	q	Group1	Group2	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0000000000	0.050000000	1.000
2	0.2	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0000000000	0.025000000	1.000
3	0.3	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0000000000	0.016666667	1.000
4	0.4	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0000000000	0.012500000	1.000
5	0.5	0.000000e+00	5.648593e-12	-5.648593e-12	-1.051249e-05	0.0000000000	0.010000000	0.073
6	0.6	0.000000e+00	9.978203e-01	-9.978203e-01	-1.000000e+00	-0.6216948080	0.007142857	0.000
7	0.7	0.000000e+00	1.000000e+00	-1.000000e+00	-1.000000e+00	-1.0000000000	0.006250000	0.000
8	0.8	1.987155e-06	1.000000e+00	-9.999980e-01	-1.000000e+00	-0.9695088636	0.005555556	0.000
9	0.9	9.924814e-01	1.000198e+00	-7.716778e-03	-5.890047e-01	0.0002160393	0.008333333	0.022

From the confidence intervals we can say, with 95% confidence, that the 6th, 7th, and 8th quantiles of group 2 (Amazon IRT tweets not containing hashtags) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Amazon IRT tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets, and the right-tails specifically), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Amazon IRT tweets.**

BMW Official: Number of Likes

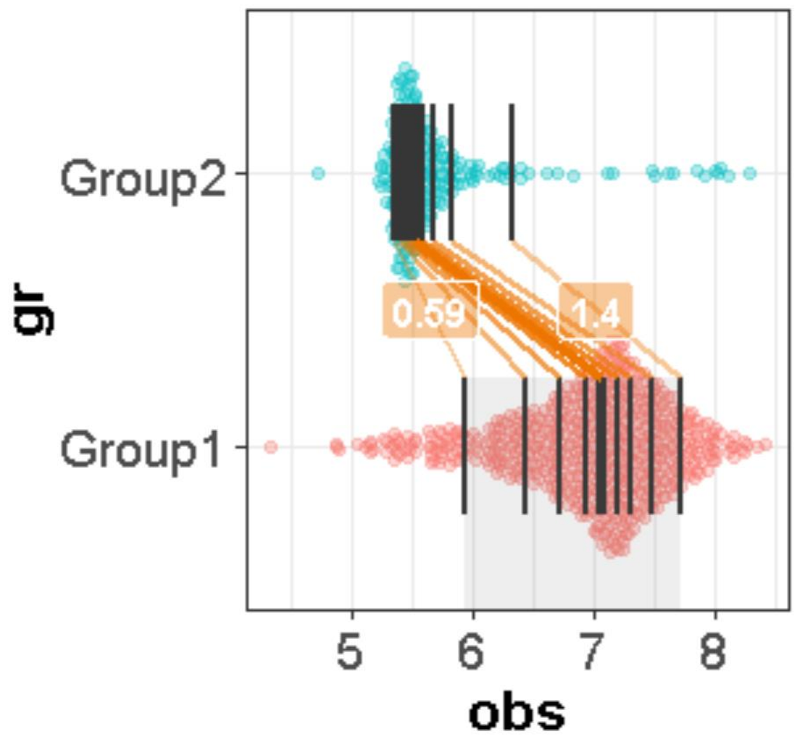




Neither of the above log distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 108742, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

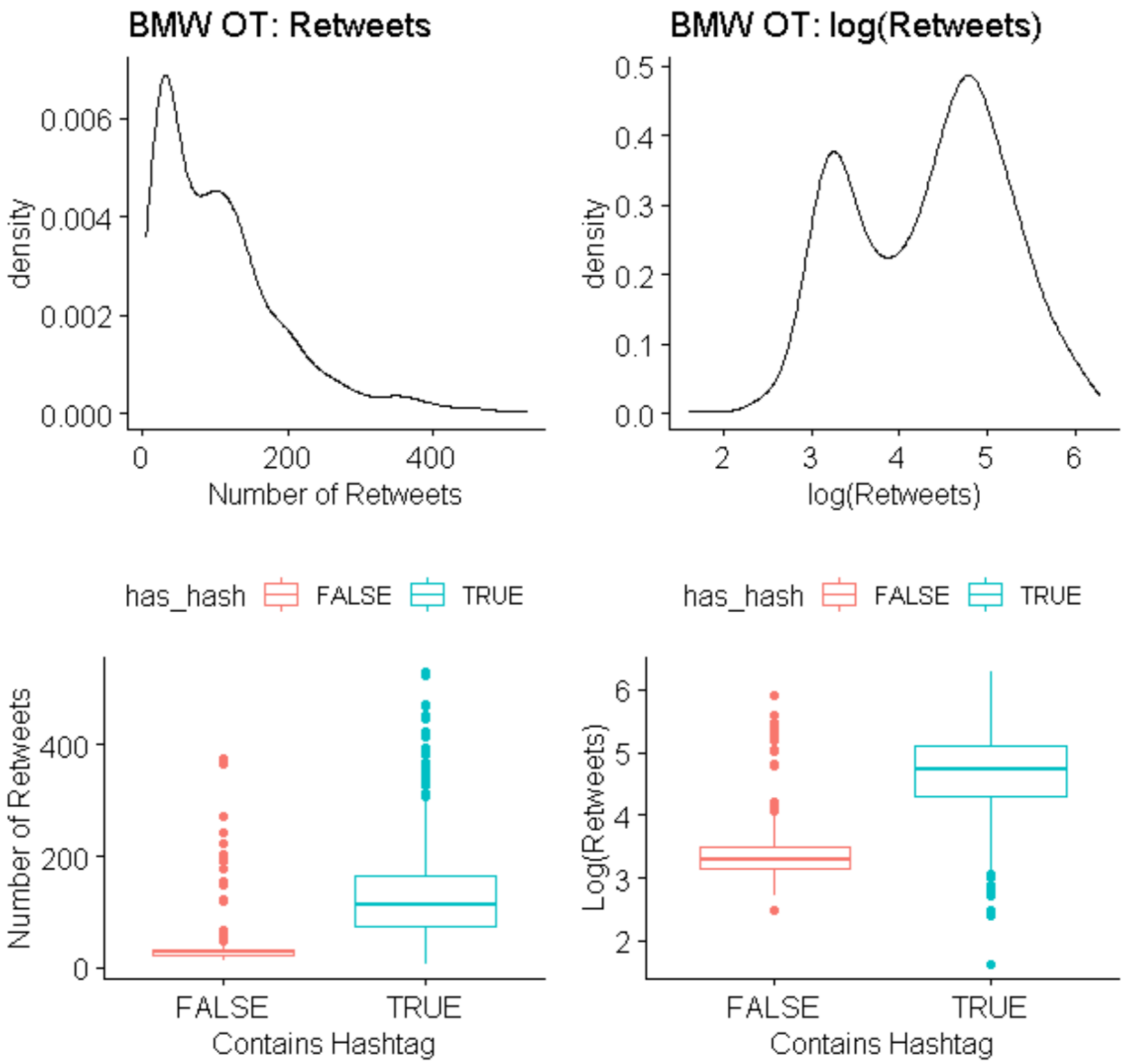
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

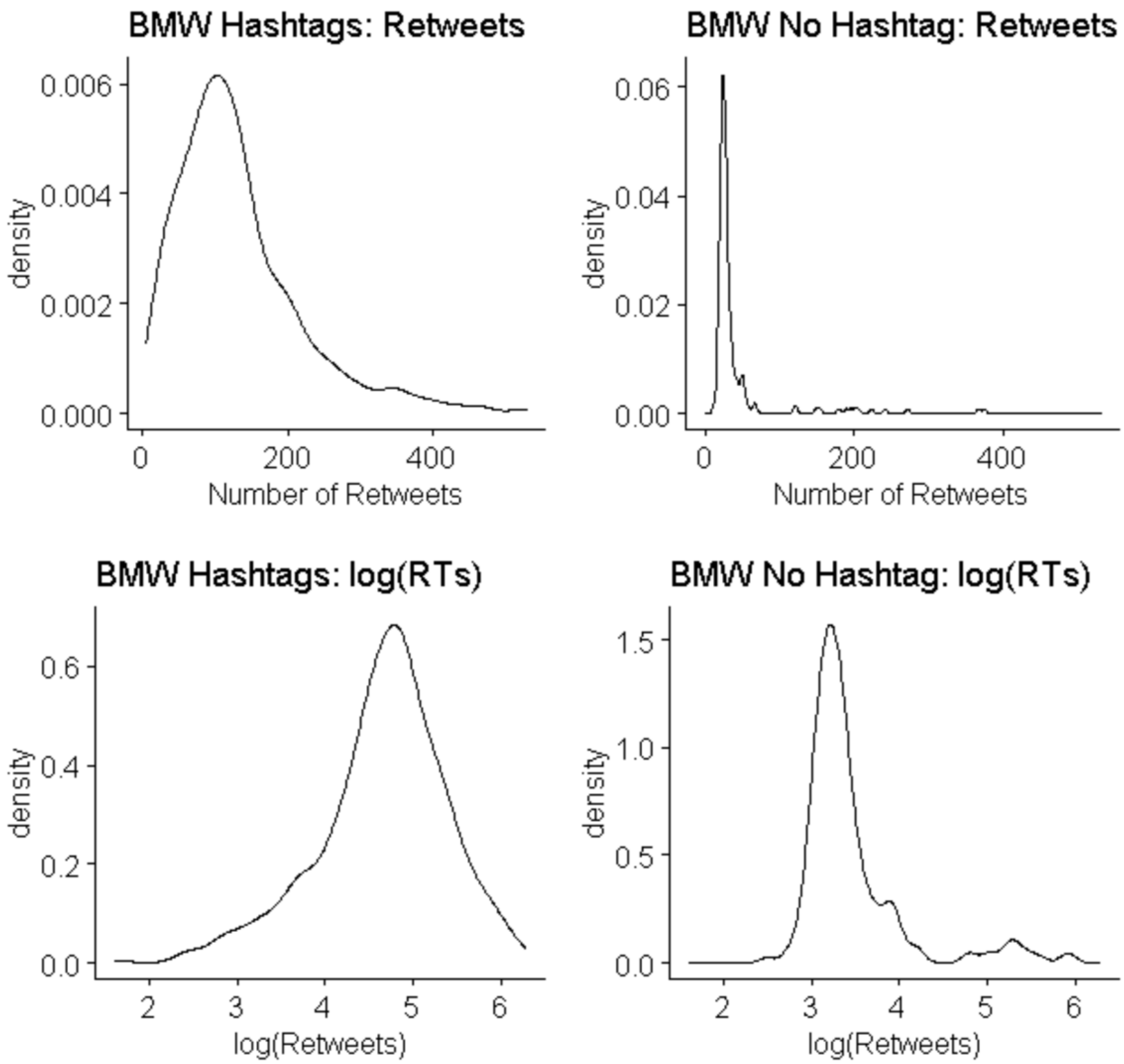



```
$`Group1 - Group2`
    q    Group1   Group2 difference  ci_lower  ci_upper      p_crit p_value
1 0.1  383.5908 211.3126   172.2781  119.0175  250.6592 0.050000000       0
2 0.2  629.4138 221.0670   408.3467  336.8106  497.4416 0.025000000       0
3 0.3  840.5046 230.4929   610.0117  535.8264  699.4748 0.016666667       0
4 0.4 1025.7019 240.4818   785.2201  706.5907  874.1722 0.012500000       0
5 0.5 1189.7387 250.6444   939.0943  851.9757 1003.8250 0.010000000       0
6 0.6 1339.8562 268.5974  1071.2588  978.9024 1138.0360 0.008333333       0
7 0.7 1502.8899 293.9934  1208.8964 1113.9968 1336.2160 0.007142857       0
8 0.8 1789.5132 341.3619  1448.1512 1296.7569 1586.9960 0.006250000       0
9 0.9 2250.6641 571.9856  1678.6785  744.1154 1989.8692 0.005555556       0
```

With 95% confidence we can say that every quantile of group 2 (BMW official tweets not containing hashtags) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (BMW official tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in BMW official tweets.**

BMW Official: Number of Retweets

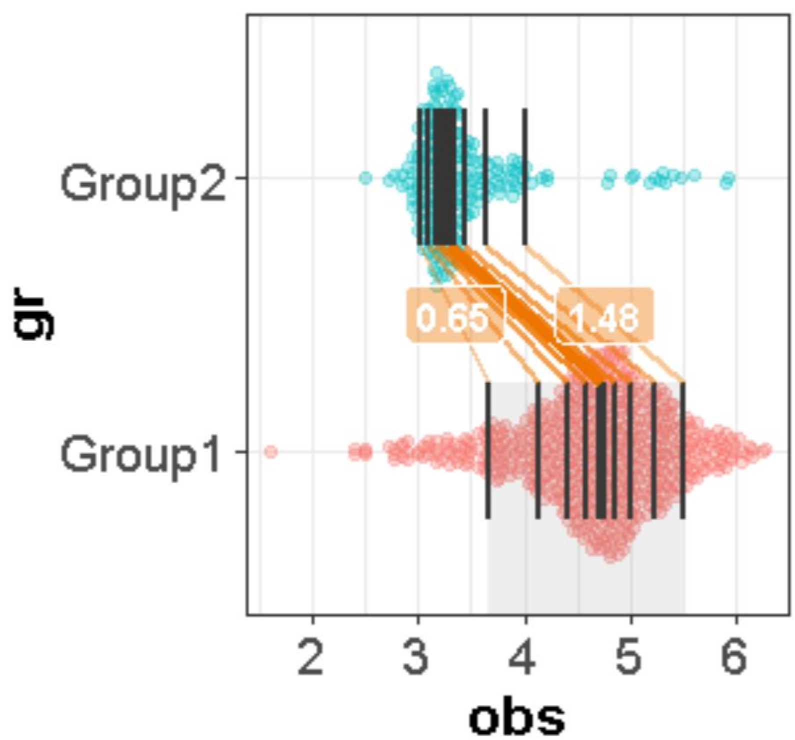




None of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 108272, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

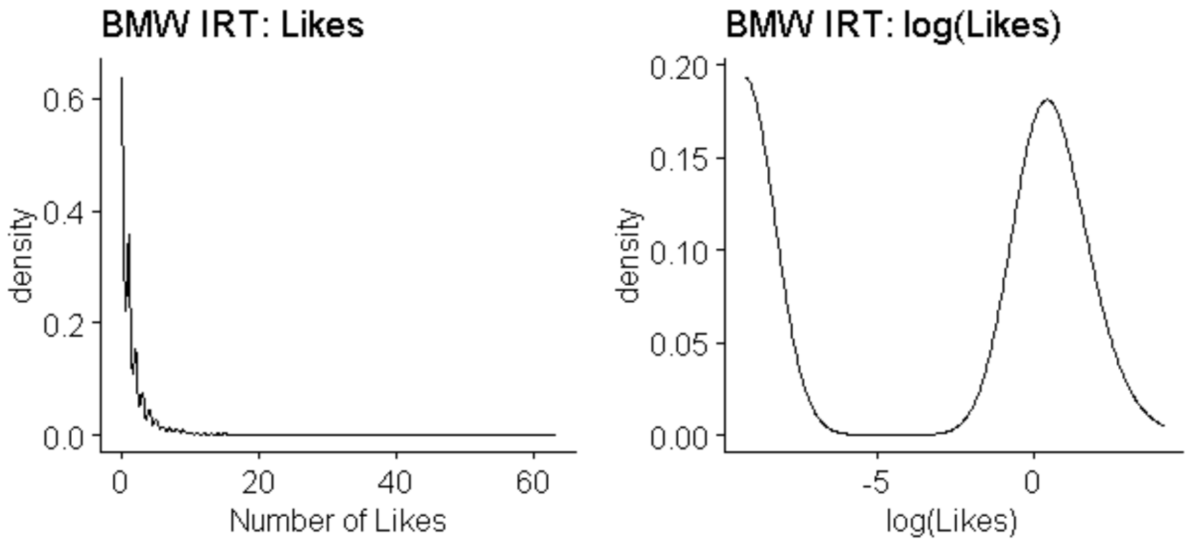


$`Group1 - Group2`

	q	Group1	Group2	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	39.15333	20.37210	18.78122	13.60320	23.73812	0.050000000	0
2	0.2	62.43536	22.19996	40.23540	33.59596	47.91132	0.025000000	0
3	0.3	82.46836	23.60962	58.85874	50.99612	65.05416	0.016666667	0
4	0.4	97.87396	25.01134	72.86262	64.79242	79.81122	0.012500000	0
5	0.5	113.56936	26.68810	86.88127	77.70403	95.22238	0.010000000	0
6	0.6	129.84273	28.50589	101.33685	91.77690	110.14558	0.008333333	0
7	0.7	149.97729	31.12967	118.84762	106.37833	135.27619	0.007142857	0
8	0.8	187.23206	37.79565	149.43641	126.81126	167.07926	0.006250000	0
9	0.9	245.14618	57.09201	188.05417	105.39539	222.61164	0.005555556	0

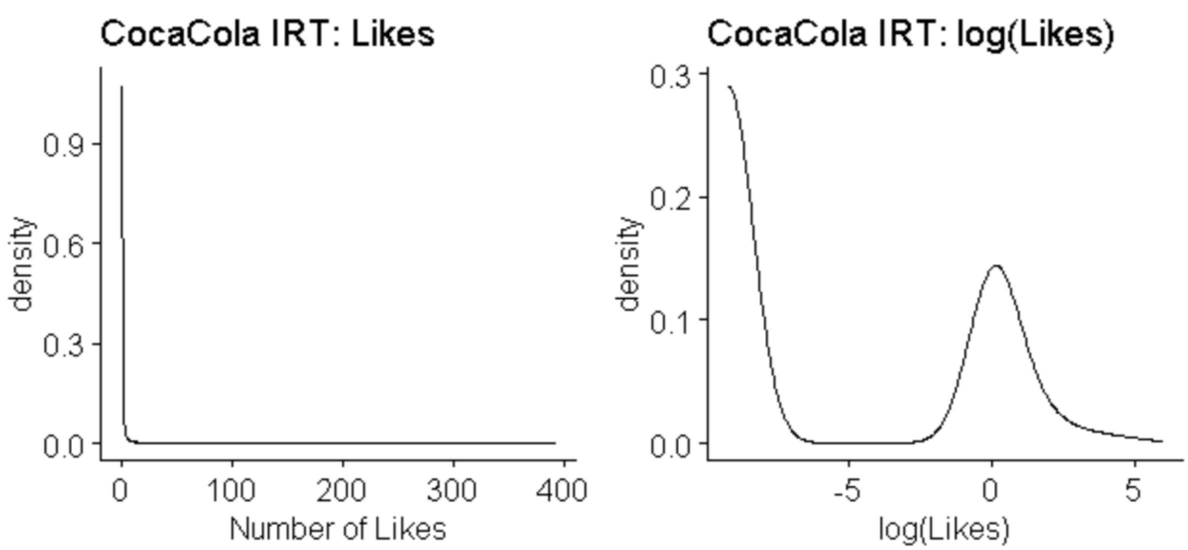
With 95% confidence we can say that every quantile of group 2 (BMW official tweets not containing hashtags) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (BMW official tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in BMW official tweets.**

BMW IRT: Number of Likes



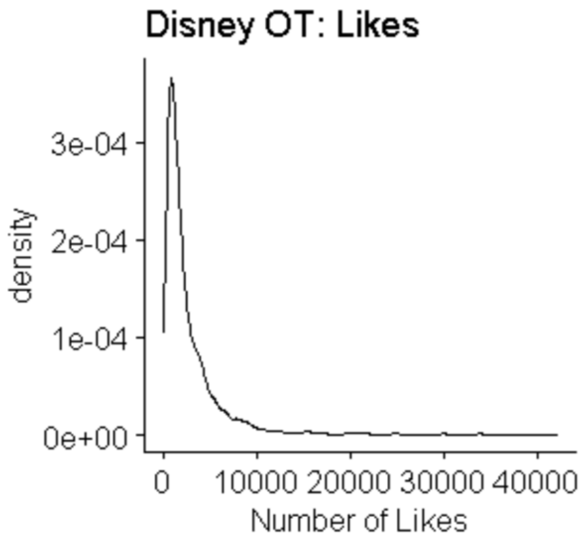
Apparently, there are only 3 BMW IRT tweets (not considered extreme outliers) containing hashtags, which is not enough for analysis.

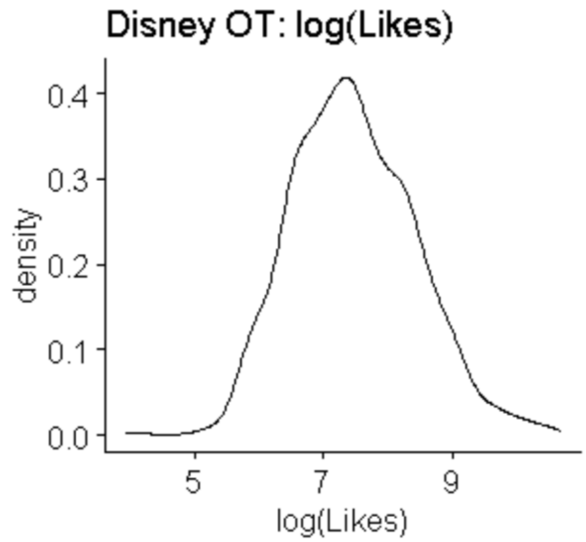
Coca Cola IRT: Number of Likes



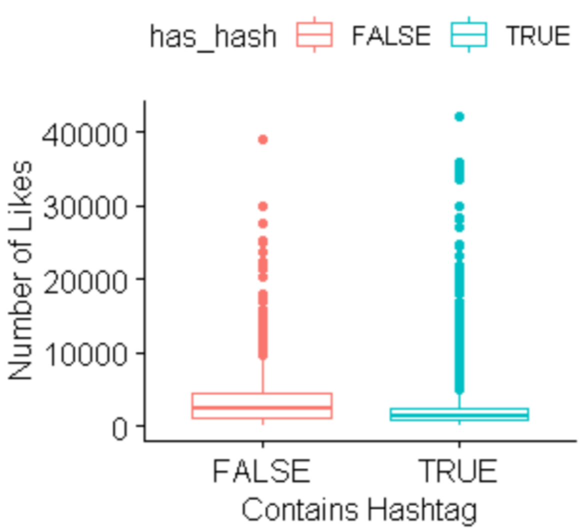
Apparently, Coca Cola only has 10 IRT tweets (not considered extreme outliers) containing hashtags, which isn't enough for analysis.

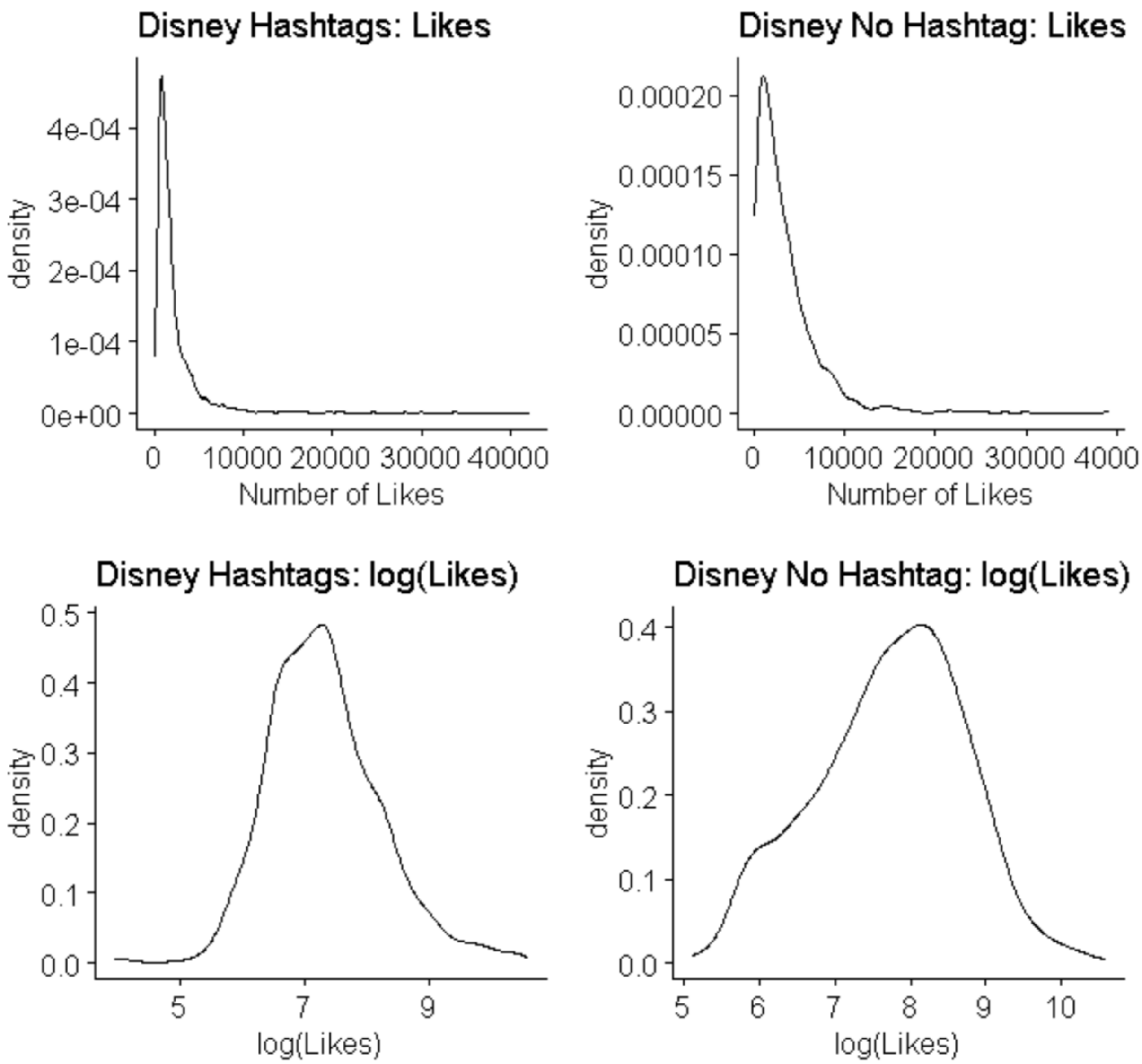
Disney Official: Number of Likes





The log distribution does not pass a Shapiro-Wilk normality test.

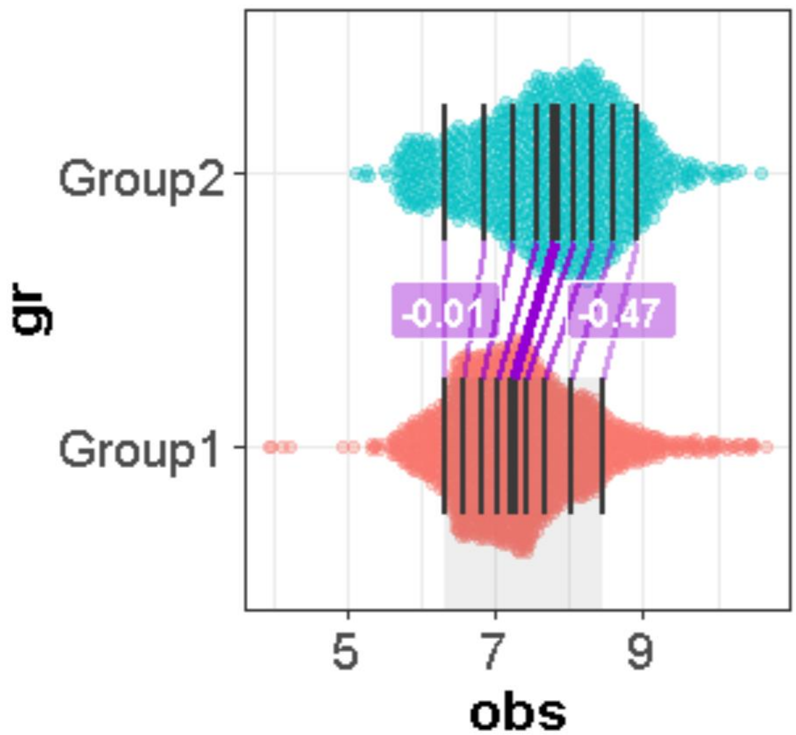




I'm unsure why I'm unable to get the x-axes on the same scale above, but note the differences. Furthermore, none of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 605999, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

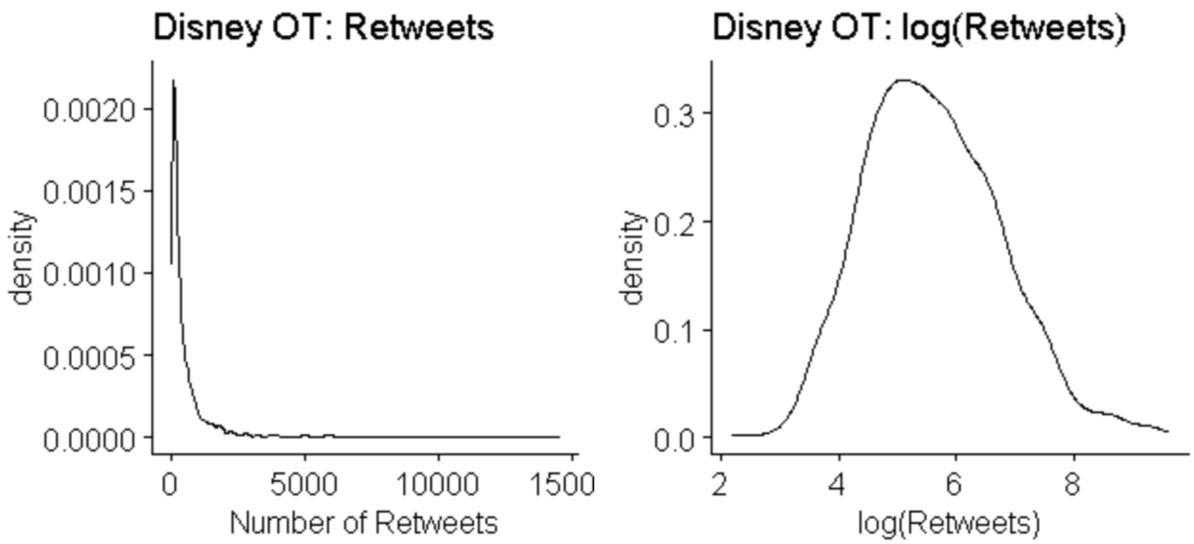
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



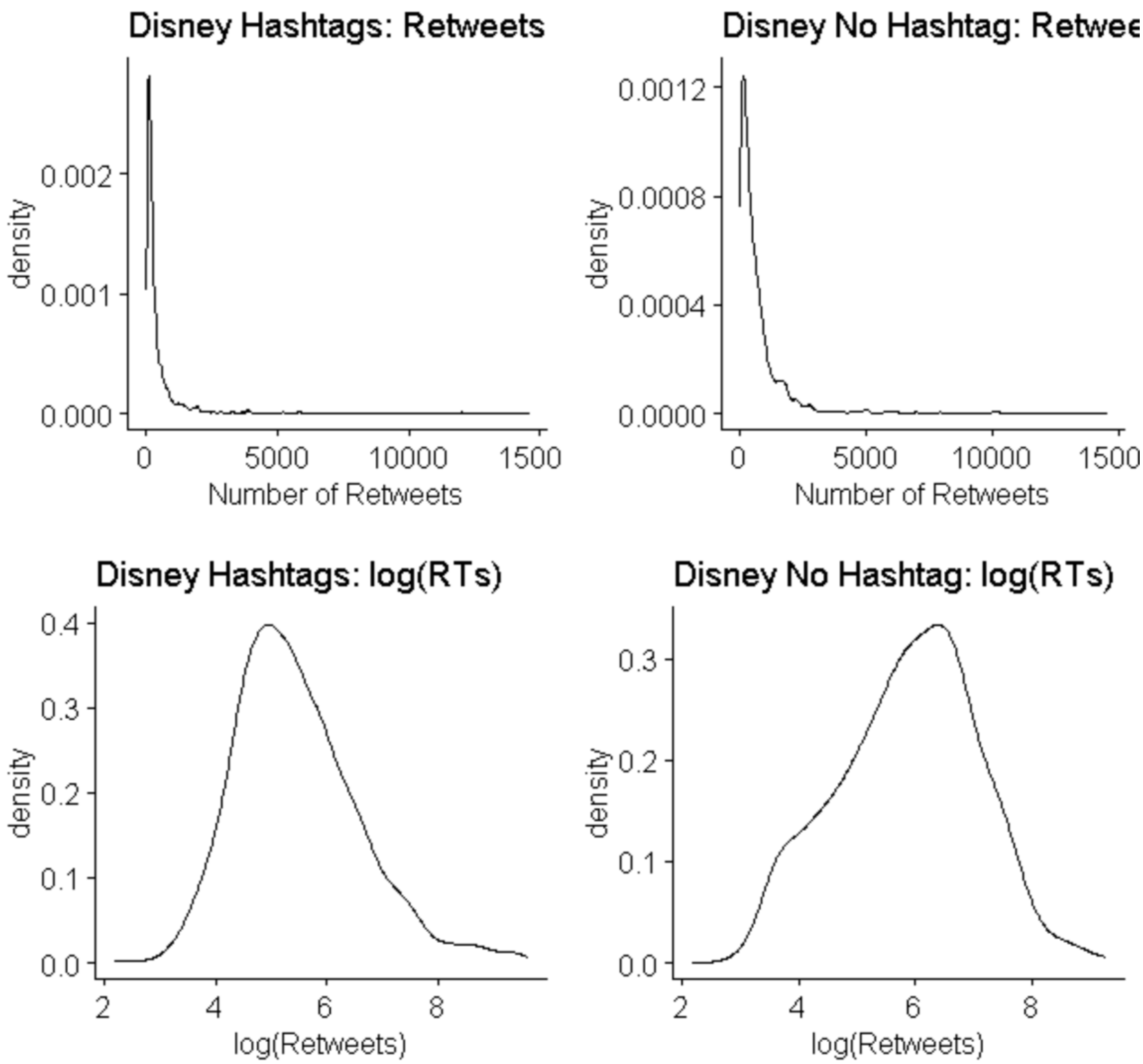

```
$`Group1 - Group2`
    q    Group1    Group2   difference    ci_lower    ci_upper       p_crit p_value
1 0.1  547.7839  556.0035    -8.219528   -80.67617    73.20292 0.050000000    0.81
2 0.2  716.6085  946.7390  -230.130506  -386.63812  -110.18965 0.025000000    0.00
3 0.3  902.7928 1402.4862  -499.693335  -685.04421  -330.01668 0.016666667    0.00
4 0.4 1131.8866 1926.7021  -794.815569  -972.07690  -588.84997 0.012500000    0.00
5 0.5 1405.7790 2511.6675 -1105.888415 -1399.20851  -812.77226 0.010000000    0.00
6 0.6 1692.8636 3220.7200 -1527.856359 -1873.10051 -1164.99414 0.008333333    0.00
7 0.7 2166.8757 4070.8170 -1903.941335 -2291.12096 -1527.27752 0.007142857    0.00
8 0.8 3055.6821 5302.1216 -2246.439562 -2874.42189 -1708.35349 0.006250000    0.00
9 0.9 4736.8488 7558.4735 -2821.624700 -3806.70178 -1721.48918 0.005555556    0.00
```

With 95% confidence we may say that every quantile of group 2 (Disney official tweets not containing hashtags), except for the first quantile, would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Disney official tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Disney official tweets.**

Disney Official: Number of Retweets



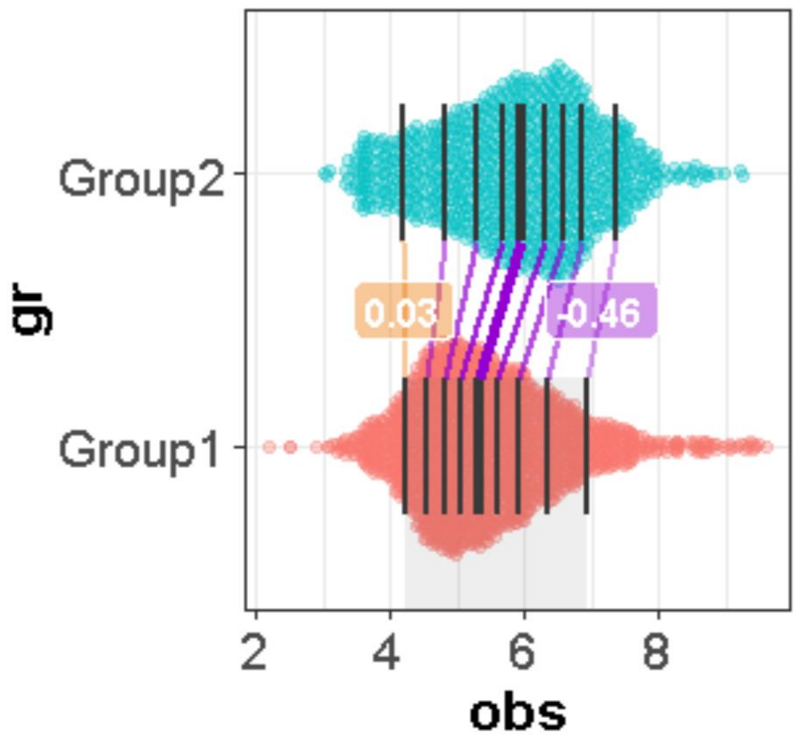
The log distribution does not pass a Shapiro-Wilk normality test.



Neither of the above log distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 637479, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

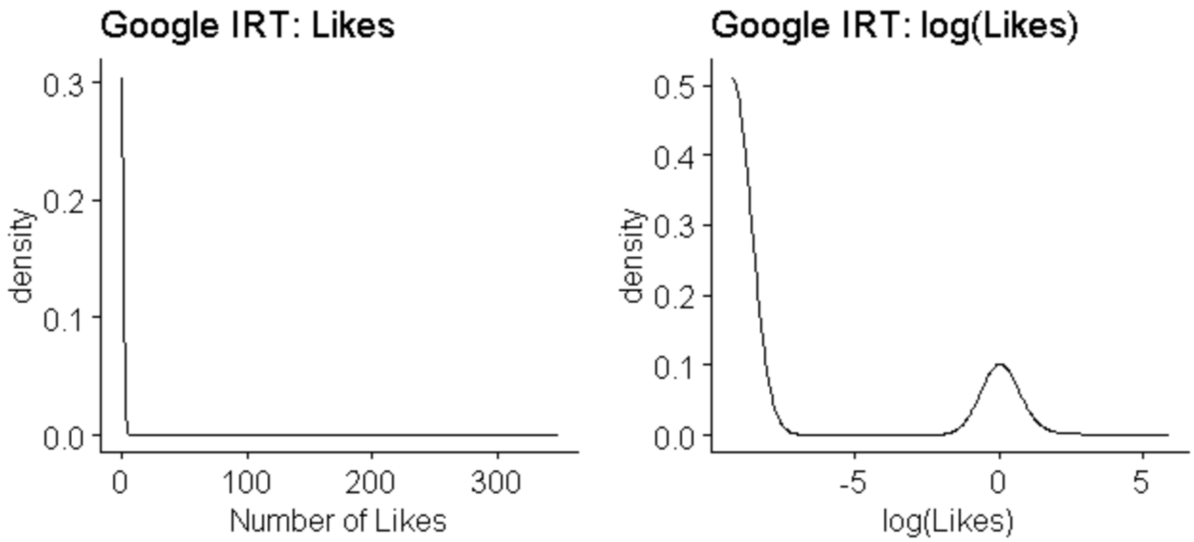
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:




```
$`Group1 - Group2`
    q     Group1     Group2  difference    ci_lower    ci_upper      p_crit p_value
1 0.1   67.23588   65.20766    2.028217   -7.834957   11.780664 0.050000000   0.689
2 0.2   94.55354  122.39216  -27.838624  -48.973442   -9.561685 0.025000000   0.000
3 0.3  122.70647  196.94430  -74.237836 -101.801733  -42.274366 0.016666667   0.000
4 0.4  155.23678  285.35902 -130.122241 -171.383156  -92.940026 0.012500000   0.000
5 0.5  202.71845  388.09477 -185.376319 -243.591881 -140.850328 0.010000000   0.000
6 0.6  265.39264  535.54827 -270.155630 -341.098601 -197.379719 0.008333333   0.000
7 0.7  369.03398  703.14868 -334.114703 -427.082502 -261.466637 0.007142857   0.000
8 0.8  557.69482  960.76023 -403.065415 -550.493129 -280.810347 0.006250000   0.000
9 0.9 1006.18135 1592.39404 -586.212685 -832.416341 -301.067658 0.005555556   0.000
```

With 95% confidence we may say that every quantile of group 2 (Disney official tweets not containing hashtags), except for the first quantile, would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (Disney official tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Disney official tweets.**

Google IRT: Number of Likes

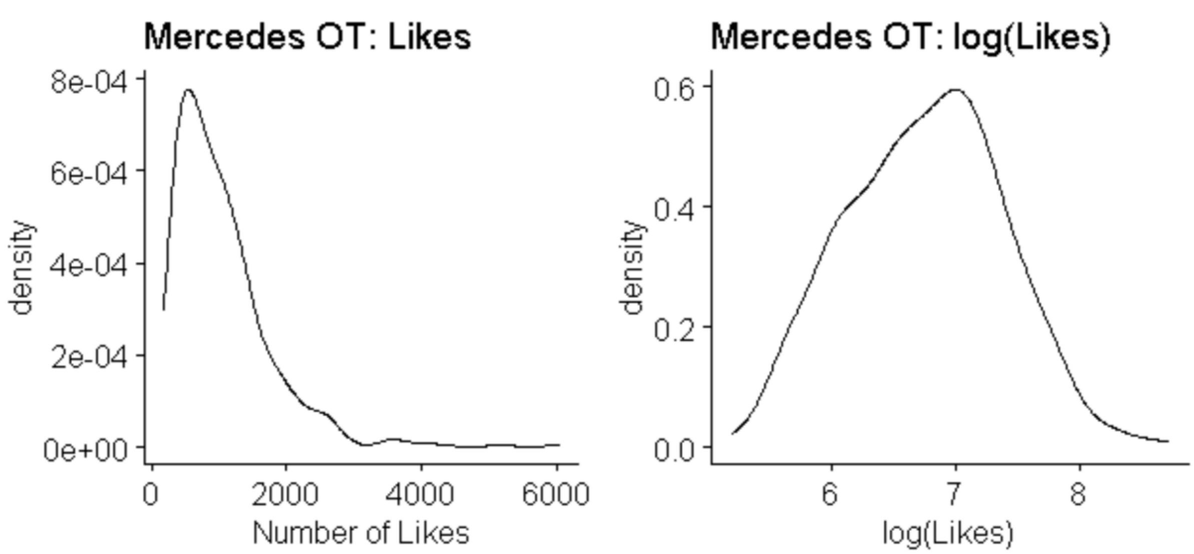


Apparently, Google only has 5 IRT tweets (not considered extreme outliers) containing hashtags, which isn't enough for analysis.

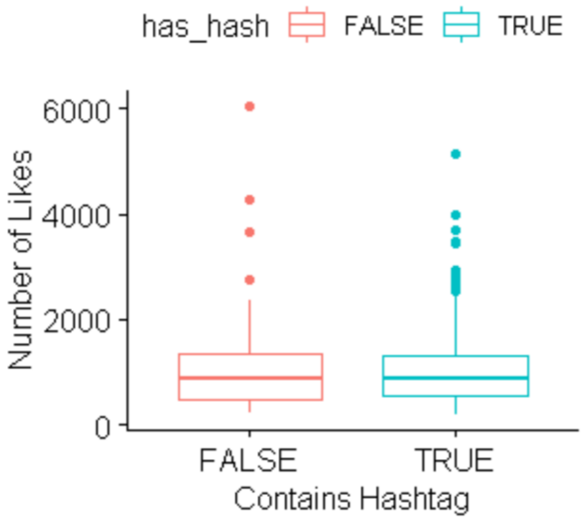
McDonalds IRT: Number of Likes

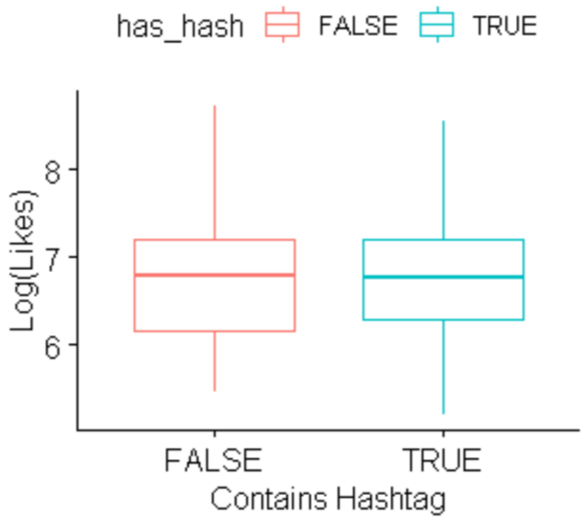
Apparently, McDonalds only has 1 IRT tweet (not considered extreme outliers) containing a hashtag, which isn't enough for analysis.

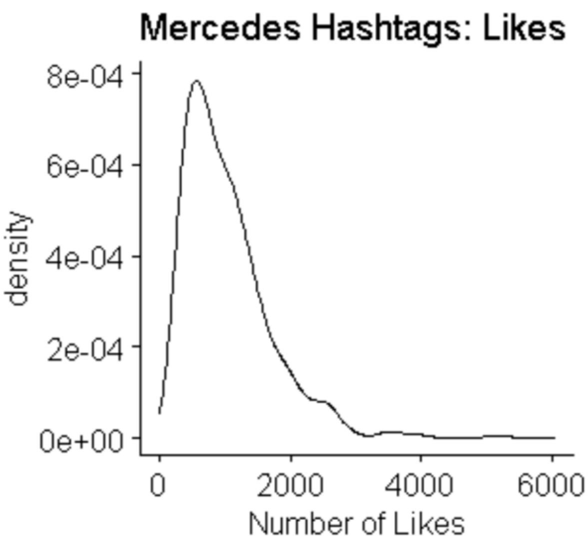
Mercedes OT: Number of Likes

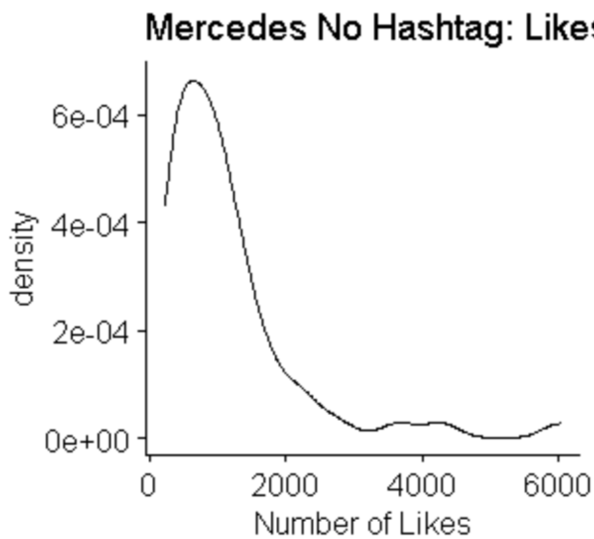


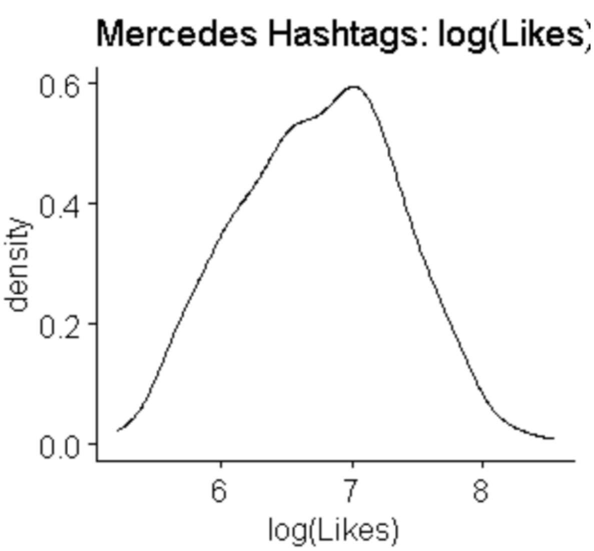
The log distribution barely does not pass a Shapiro-Wilk normality test.

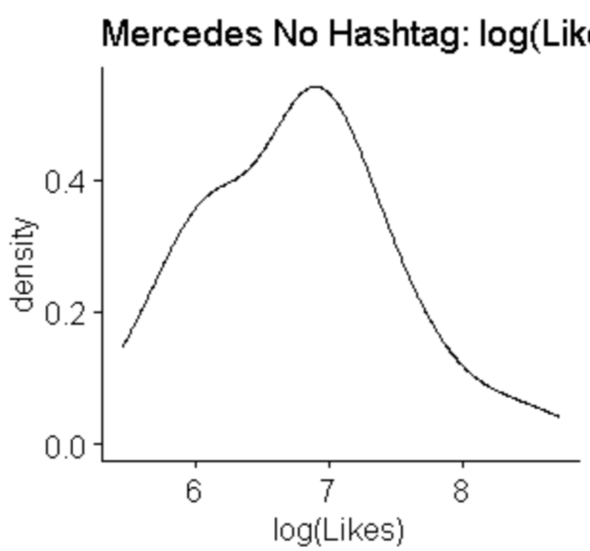

has_hash
FALSE
TRUE
Number of Likes
6000
4000
2000
0
FALSE
TRUE
Contains Hashtag


has_hash
FALSE
TRUE
Log(Likes)
8
7
6
FALSE
TRUE
Contains Hashtag


Mercedes Hashtags: Likes
density
8e-04
6e-04
4e-04
2e-04
0e+00
0
2000
4000
6000
Number of Likes


density
6e-04
4e-04
2e-04
0e+00
0
2000
4000
6000
Number of Likes

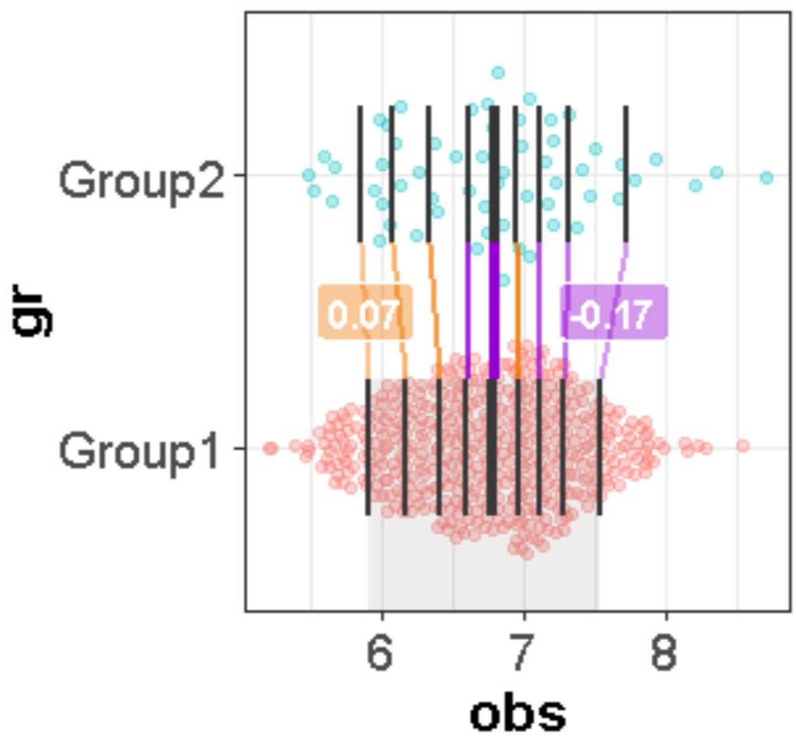

density
0.6
0.4
0.2
0.0
6
7
8
log(Likes)


density
0.4
0.2
0.0
6
7
8
log(Likes)

Surprisingly, both log distributions pass a Shapiro-Wilk normality test. However, I feel confident that my validation process is appropriate nonetheless.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 12926, p-value = 0.94
alternative hypothesis: true location shift is not equal to 0
```

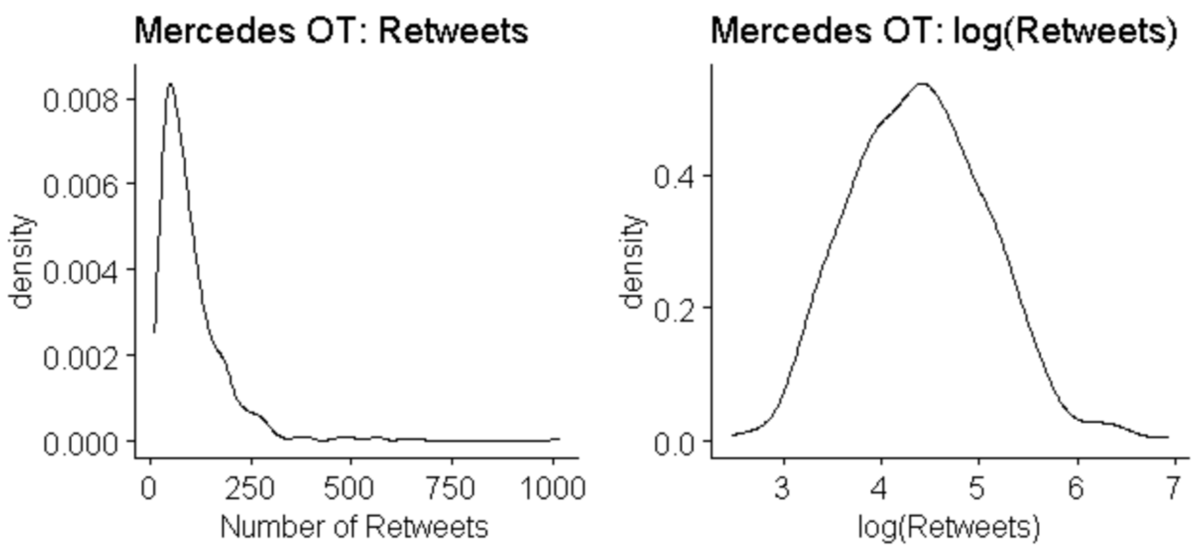
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'like' distributions of both populations are equal. Performing a shift function yields the following:



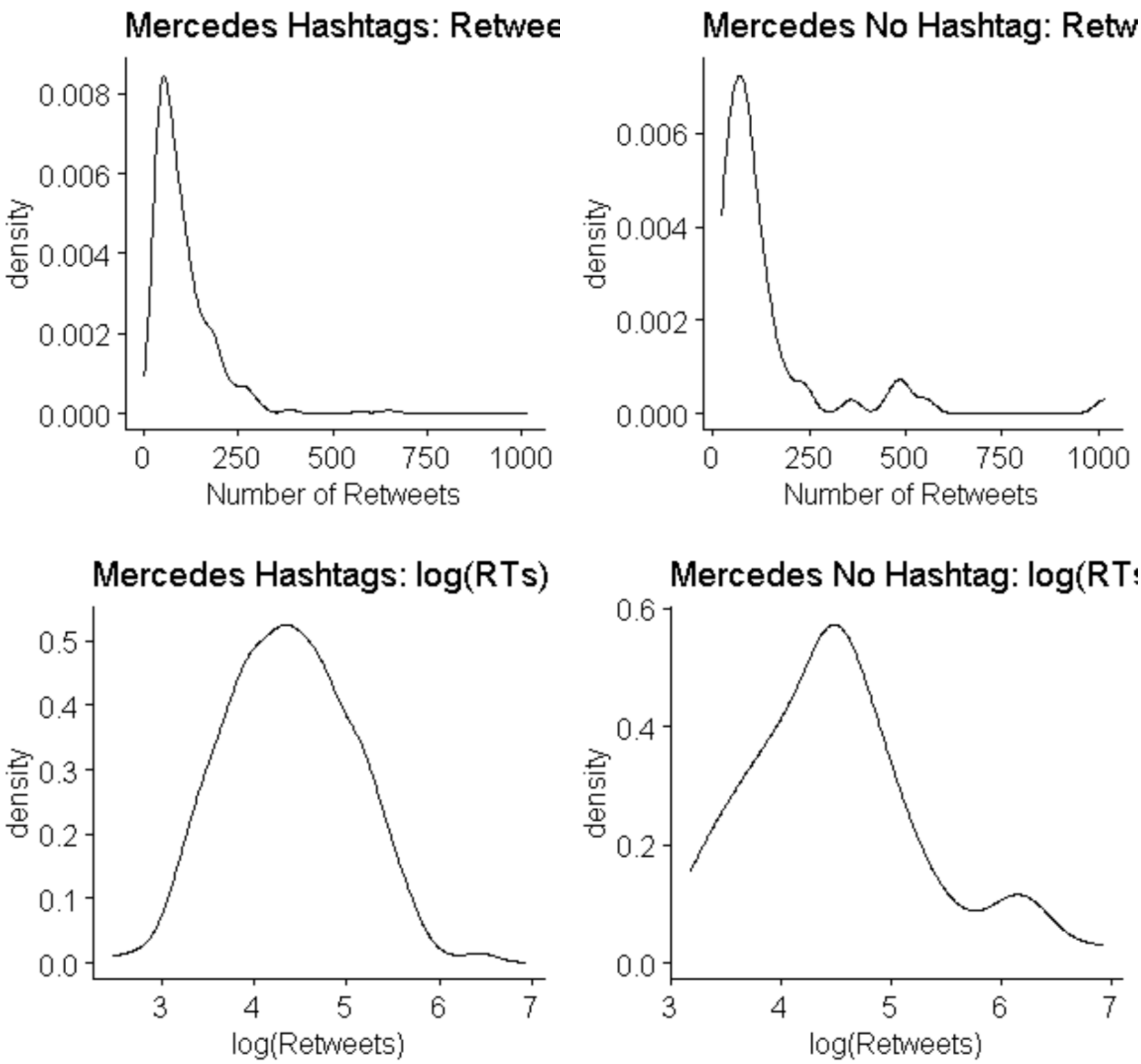

```
$`Group1 - Group2`
    q    Group1    Group2 difference    ci_lower ci_upper      p_crit p_value
1 0.1  369.1019  348.3772   20.72468   -89.49137 117.2707 0.010000000   0.741
2 0.2  477.6939  438.8903   38.80359  -170.52820 156.1147 0.006250000   0.480
3 0.3  607.3479  569.5437   37.80421  -232.87185 205.7399 0.008333333   0.729
4 0.4  724.8235  750.4081  -25.58460  -236.08399 209.7376 0.012500000   0.834
5 0.5  876.8272  888.7641  -11.93689  -232.23117 217.3284 0.025000000   0.893
6 0.6 1052.1023 1036.3239   15.77838  -277.90010 233.9365 0.016666667   0.898
7 0.7 1221.9460 1234.4167  -12.47073  -314.24307 237.8796 0.050000000   0.932
8 0.8 1454.0698 1531.3433  -77.27353  -876.07872 329.5880 0.007142857   0.668
9 0.9 1895.5245 2302.2205 -406.69600 -2442.87908 506.6404 0.005555556   0.327
```

Considering each confidence interval contains the value 0, we may not conclude any quantile differences in tweet performance between the two groups. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of likes are equal to one another. **Inclusion of a hashtag does not seem to have a statistically significant effect on the number of likes which a Mercedes official tweet receives.**

Mercedes Official: Number of Retweets



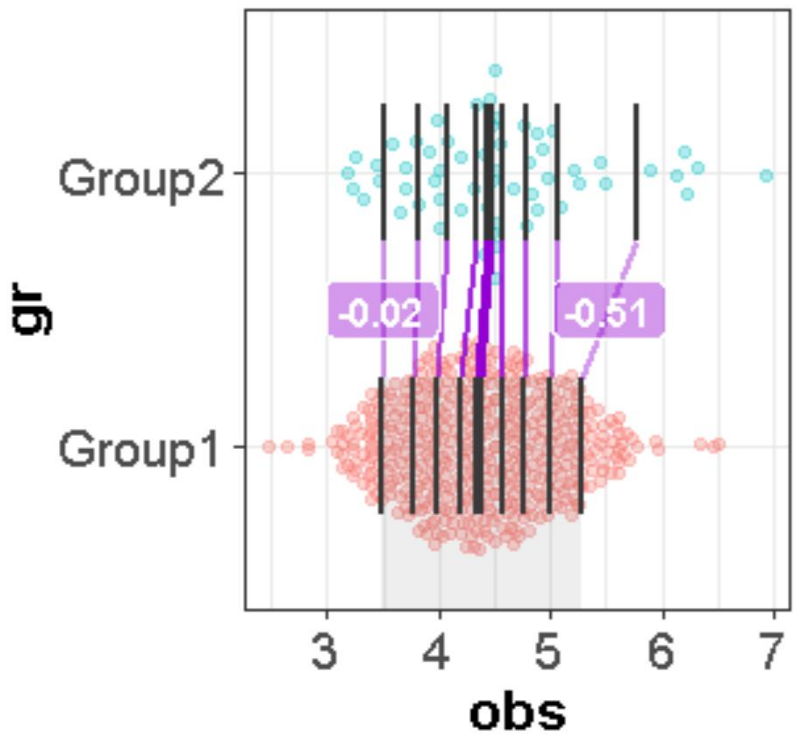
The log distribution does not pass a Shapiro-Wilk normality test.



The log distribution on the left passes a Shapiro-Wilk normality test, but the log distribution on the right does not.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 11961, p-value = 0.393
alternative hypothesis: true location shift is not equal to 0
```

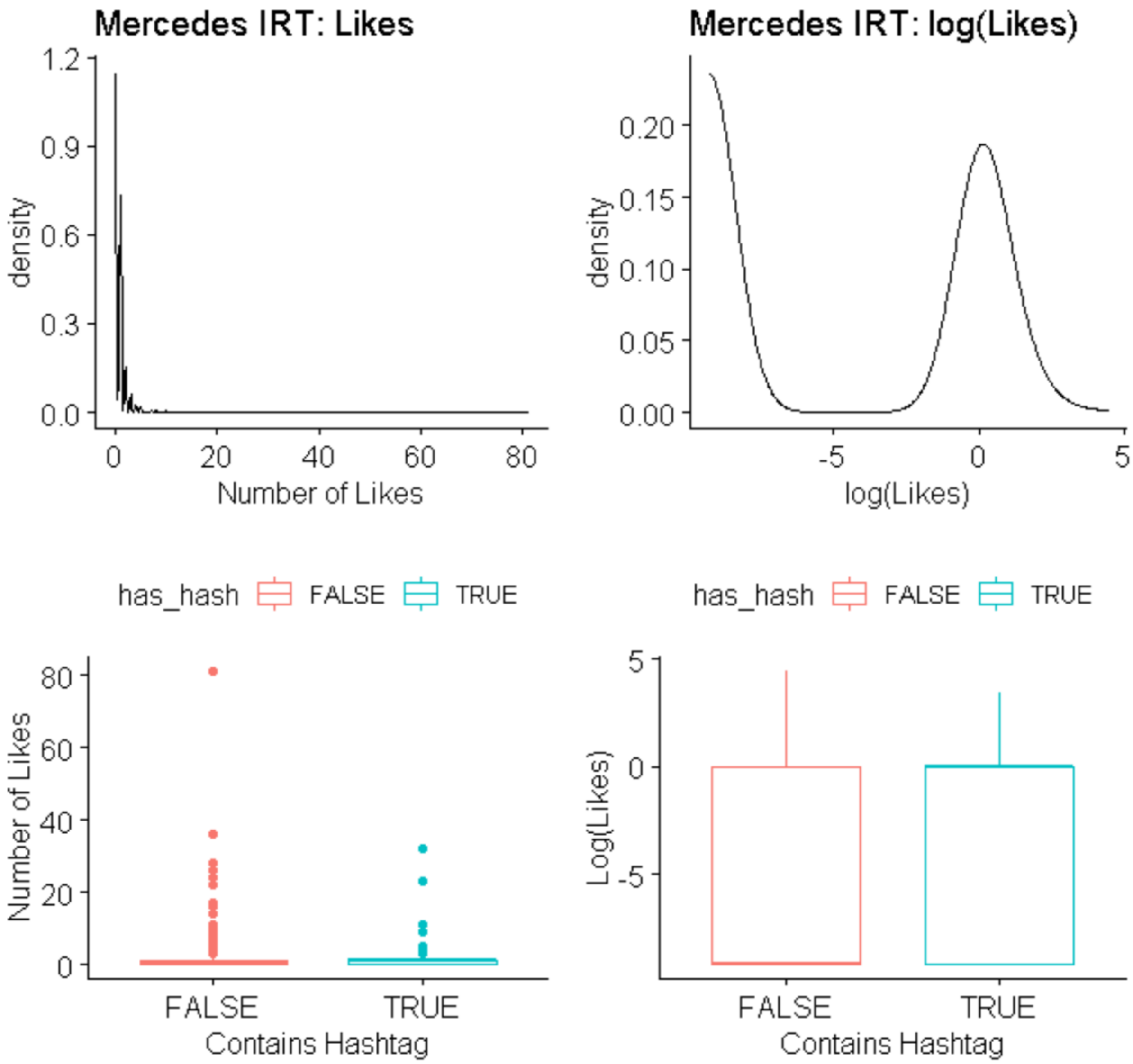
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘retweet’ distributions of both populations are equal. Performing a shift function yields the following:

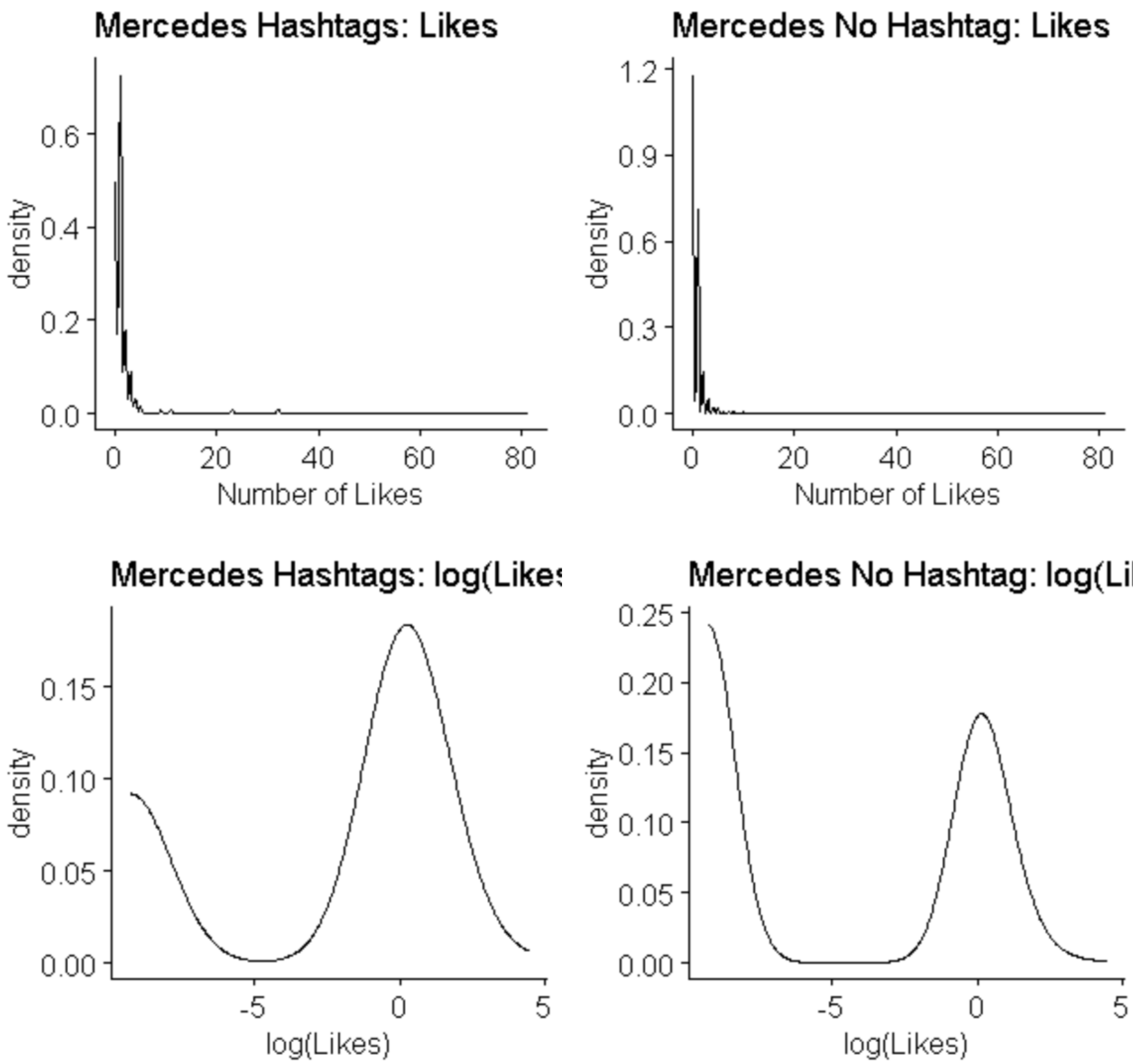



```
$`Group1 - Group2`
    q    Group1    Group2  difference   ci_lower  ci_upper      p_crit p_value
1 0.1  32.49817  33.66516   -1.166990  -14.43871  6.853046 0.025000000   0.757
2 0.2  43.52451  46.37243   -2.847927  -21.83410 10.411458 0.012500000   0.629
3 0.3  53.16950  59.53369   -6.364196  -31.41796 11.080655 0.008333333   0.416
4 0.4  65.92911  76.22012  -10.291012  -28.17032 13.253239 0.007142857   0.321
5 0.5  78.68715  86.96065   -8.273508  -35.32371 13.741499 0.006250000   0.207
6 0.6  96.04200  97.71614   -1.674145  -29.49536 13.658724 0.050000000   0.798
7 0.7 115.58136 121.70638   -6.125011  -63.23482 27.745311 0.016666667   0.726
8 0.8 148.34498 164.09263  -15.747653 -208.03647 39.776784 0.010000000   0.554
9 0.9 193.97759 347.53081 -153.553228 -478.26447 49.096063 0.005555556   0.104
```

Considering each confidence interval contains the value 0, we may not conclude any quantile differences in tweet performance between the two groups. We fail to reject the null hypotheses that the matching quantiles of the two groups distributions of retweets are equal to one another. **Inclusion of a hashtag does not seem to have a statistically significant effect on the number of retweets which a Mercedes official tweet receives.**

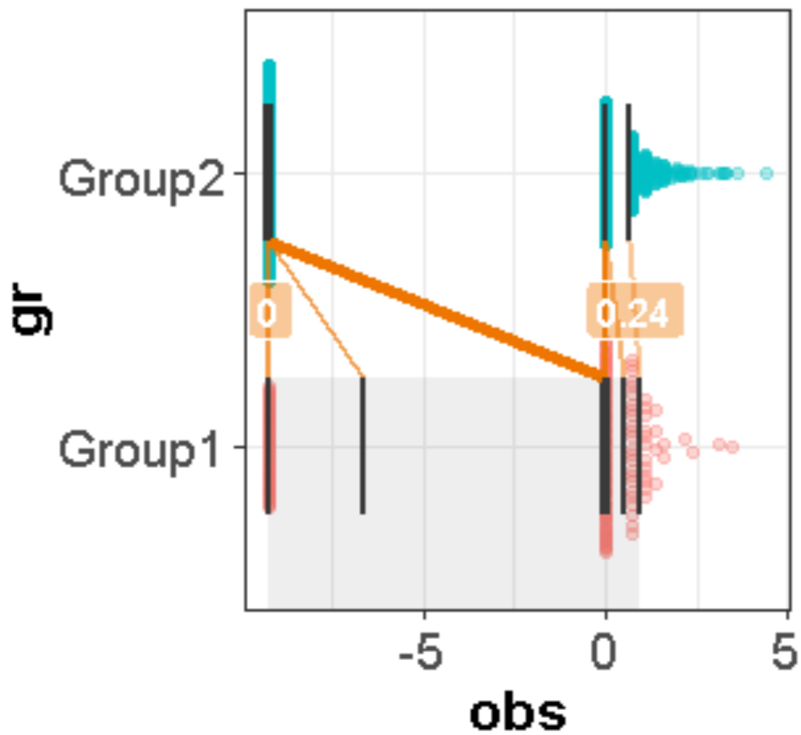
Mercedes IRT: Number of Likes






```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 308596, p-value = 1.782e-11
alternative hypothesis: true location shift is not equal to 0
```

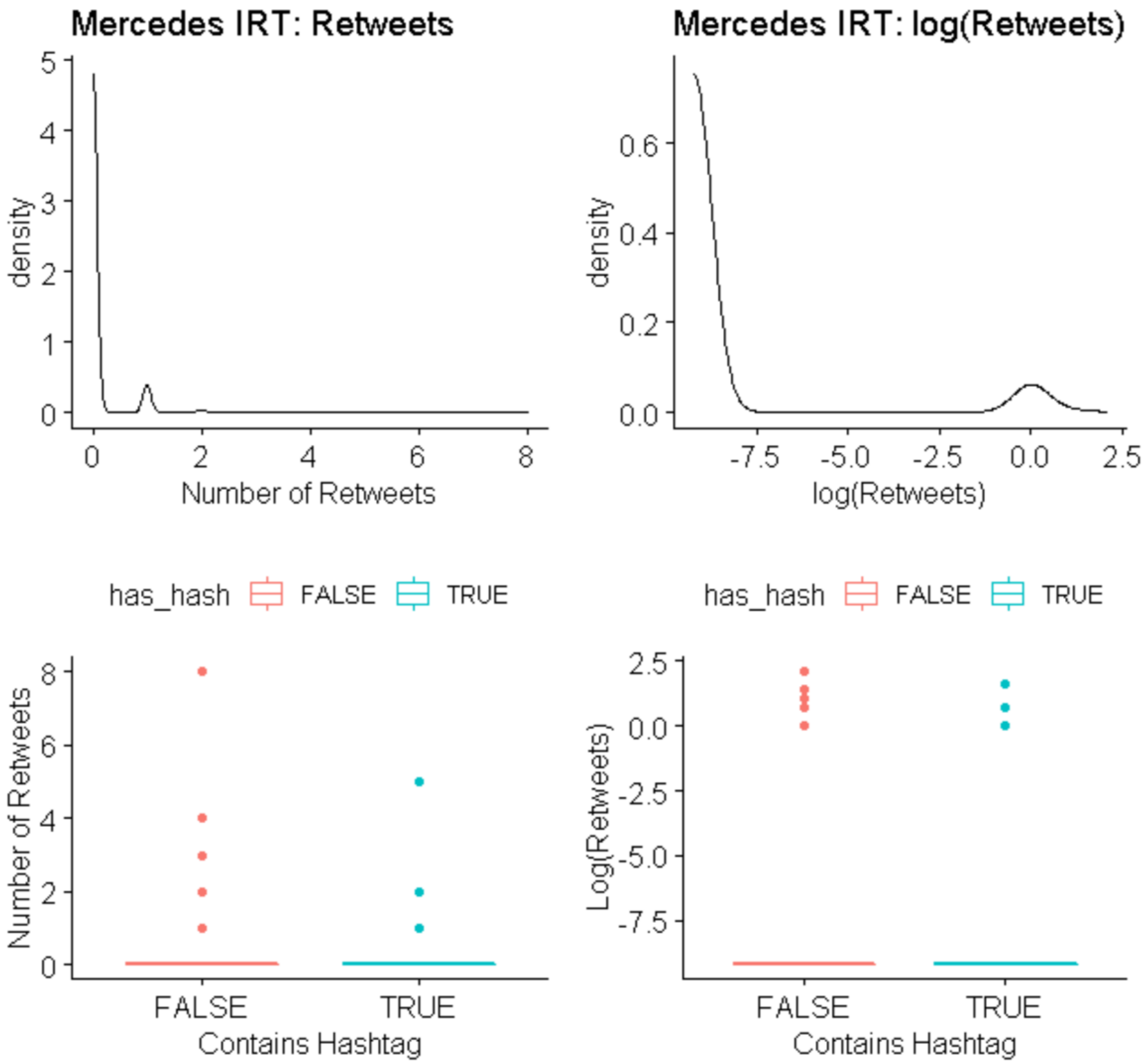
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

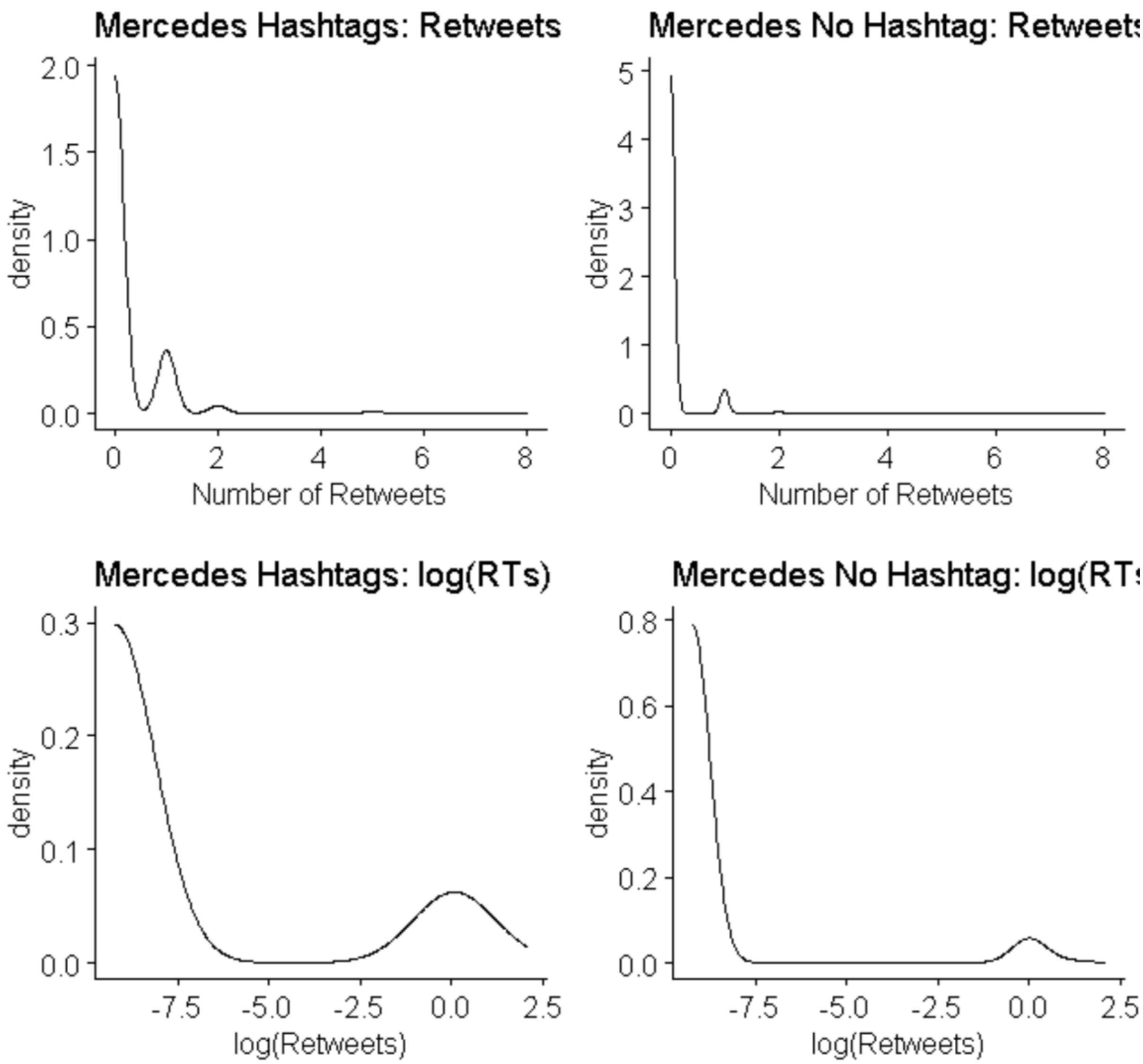



```
$`Group1 - Group2`
    q       Group1       Group2   difference     ci_lower     ci_upper      p_crit p_value
1 0.1 1.199041e-14 0.000000e+00 1.199041e-14 0.000000e+00 2.528313e-09 0.050000000   0.239
2 0.2 4.888983e-05 0.000000e+00 4.888983e-05 2.517593e-10 8.419911e-02 0.012500000   0.000
3 0.3 2.805385e-01 0.000000e+00 2.805385e-01 1.435098e-03 9.699925e-01 0.010000000   0.000
4 0.4 9.942510e-01 0.000000e+00 9.942510e-01 5.297322e-01 9.999999e-01 0.008333333   0.000
5 0.5 1.000000e+00 1.475768e-06 9.999985e-01 9.772146e-01 1.000000e+00 0.007142857   0.000
6 0.6 1.000000e+00 9.999997e-01 2.963445e-07 6.237455e-12 3.096151e-03 0.025000000   0.010
7 0.7 1.001924e+00 1.000000e+00 1.924310e-03 2.537024e-10 4.364457e-01 0.006250000   0.000
8 0.8 1.716269e+00 1.000000e+00 7.162686e-01 1.554086e-02 1.056211e+00 0.005555556   0.000
9 0.9 2.578597e+00 1.987186e+00 5.914115e-01 1.309707e-02 1.406738e+00 0.016666667   0.007
```

We can say, with 95% confidence, that the 2nd through 9th quantiles of group 2 (Mercedes IRT tweets not containing hashtags) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Mercedes IRT tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Mercedes IRT tweets.** It seems that, what's really occurring is that group 2 is dominated by tweets receiving 0 likes, while group 1 just has a fair number of those.

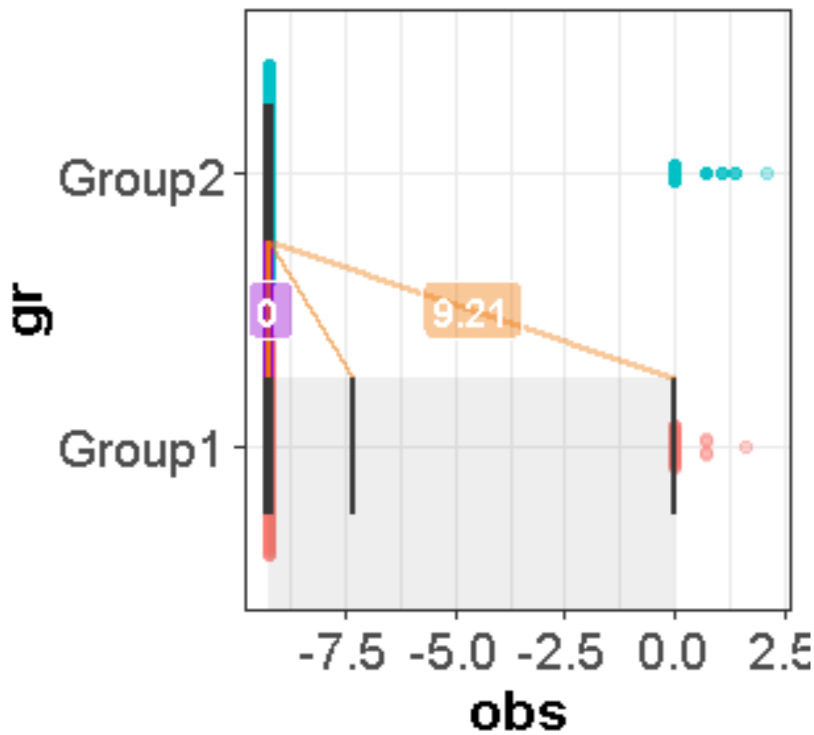
Mercedes IRT: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 272548, p-value = 1.463e-07
alternative hypothesis: true location shift is not equal to 0
```

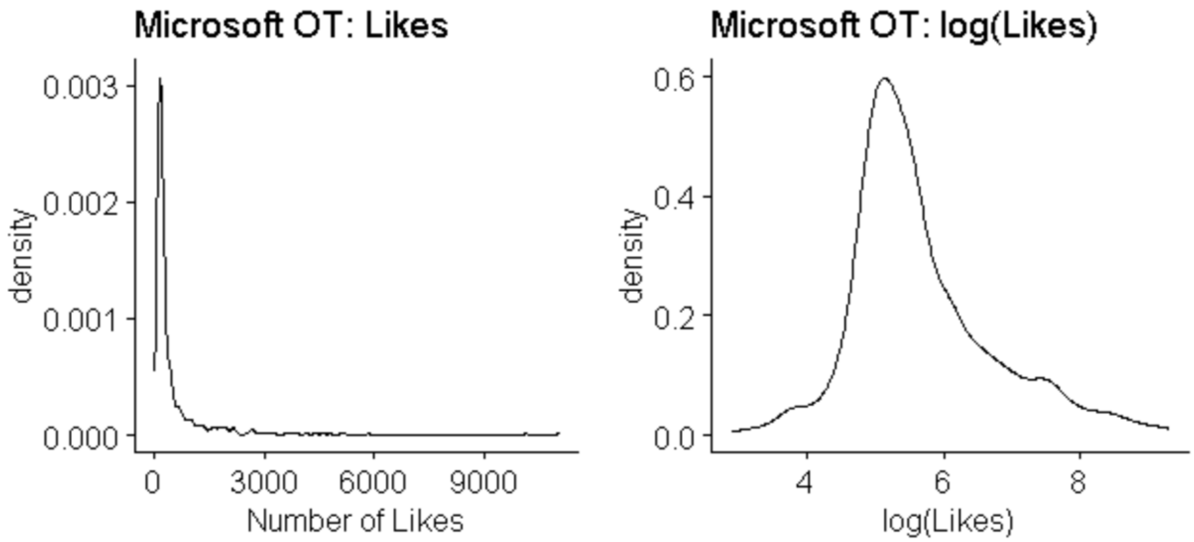
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



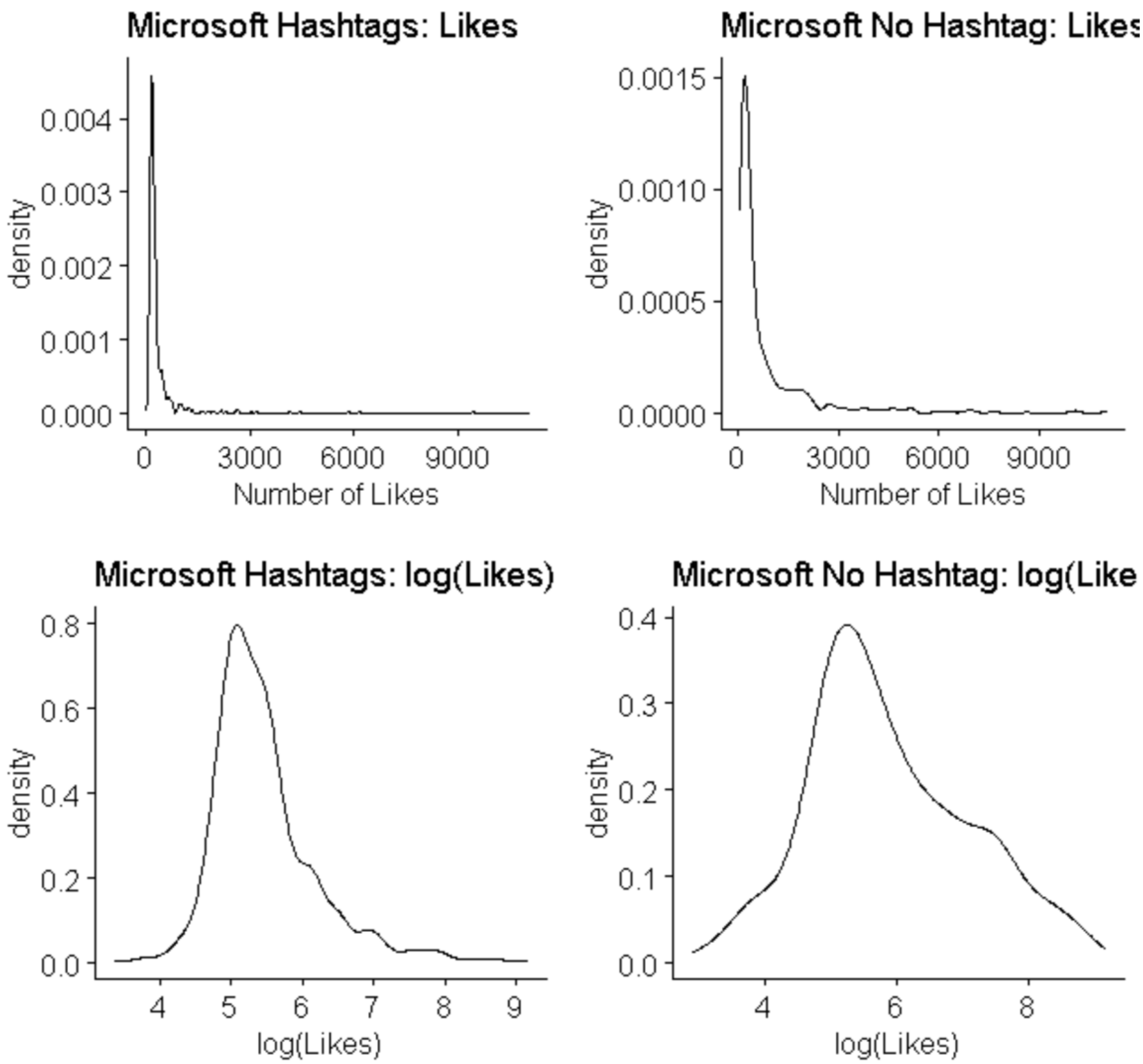

```
$`Group1 - Group2`
    q       Group1       Group2   difference     ci_lower     ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.050000000   1.000
2 0.2 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.025000000   1.000
3 0.3 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.016666667   1.000
4 0.4 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.012500000   1.000
5 0.5 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 9.992007e-16 0.010000000   0.986
6 0.6 1.276756e-14 0.000000e+00 1.276756e-14 0.000000e+00 1.145274e-06 0.008333333   0.273
7 0.7 6.838378e-06 0.000000e+00 6.838378e-06 6.183942e-14 3.000746e-02 0.007142857   0.000
8 0.8 2.084414e-01 0.000000e+00 2.084414e-01 1.121535e-04 9.544427e-01 0.006250000   0.000
9 0.9 9.994153e-01 3.222464e-06 9.994121e-01 6.617886e-01 1.006743e+00 0.005555556   0.000
```

With 95% confidence, we may say that the 7th through 9th quantiles of group 2 (Mercedes IRT tweets not containing hashtags) would need to be shifted up by significant (non-zero) amounts to match their counterpart in group 1 (Mercedes IRT tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets, and the right-tails specifically), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Mercedes IRT tweets.** However, again, it seems that the real story is that group 2 is dominated by tweets receiving 0 retweets.

Microsoft Official: Number of Likes



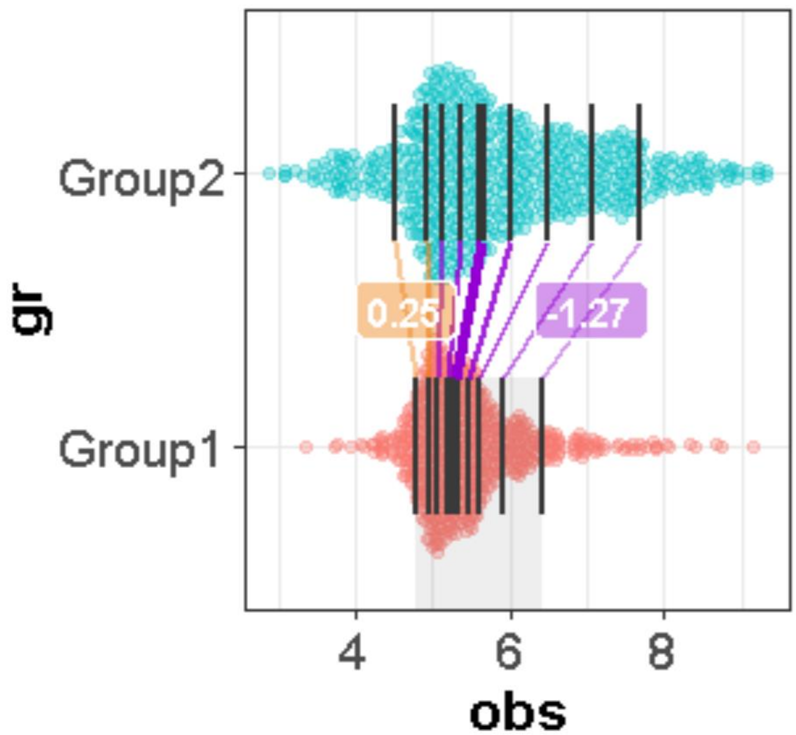
The log distribution does not pass a Shapiro-Wilk normality test.



Again, the x-axes don't perfectly align, so note the differences. Furthermore, none of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 155066, p-value = 2.58e-10
alternative hypothesis: true location shift is not equal to 0
```

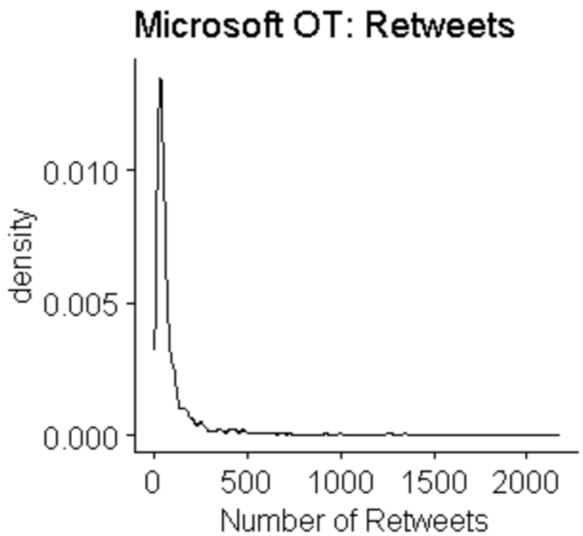
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

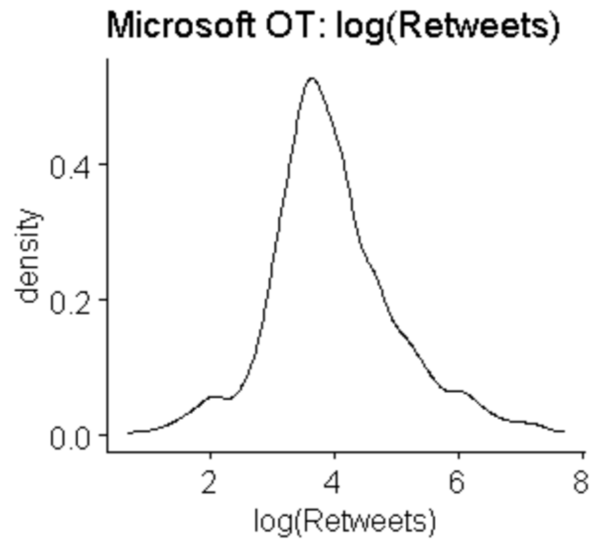



```
$`Group1 - Group2`
    q   Group1     Group2  difference     ci_lower     ci_upper      p_crit p_value
1 0.1 117.9351   92.38403    25.55109     5.104524    49.047789 0.016666667   0.001
2 0.2 139.5856  137.18744     2.39816   -10.178010    14.672599 0.050000000   0.721
3 0.3 155.9766  169.98232   -14.00567   -33.897604     1.479258 0.025000000   0.043
4 0.4 177.0421  217.21252   -40.17042   -74.526504   -12.576216 0.012500000   0.000
5 0.5 201.8240  281.17010   -79.34606  -137.159914   -41.720671 0.010000000   0.000
6 0.6 235.5729  407.42392  -171.85100  -273.799373   -99.146406 0.008333333   0.000
7 0.7 273.3133  661.26817  -387.95485  -575.468218  -226.294633 0.007142857   0.000
8 0.8 368.8219 1171.63471  -802.81282 -1190.954214  -495.261749 0.006250000   0.000
9 0.9 609.0454 2170.25537 -1561.21000 -2415.875227 -1148.939987 0.005555556   0.000
```

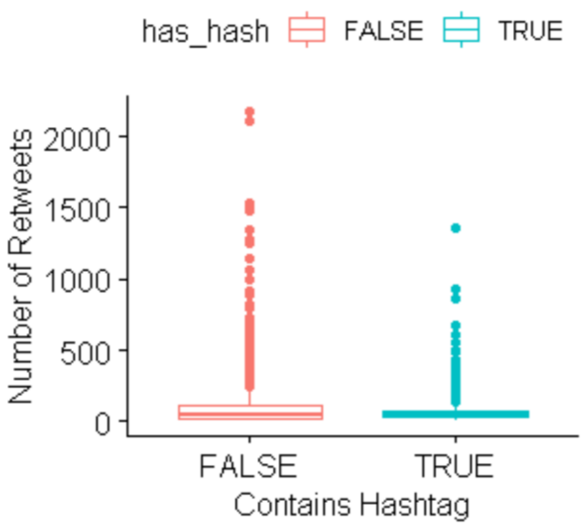
With 95% confidence we may say that the first quantile of group 2 (Microsoft official tweets not containing hashtags) would need to be shifted up by a significant (non-zero) amount to match its counterpart in group 1 (Microsoft official tweets containing hashtags). Furthermore, the 4th through 9th quantiles of group 2 would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1. **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Microsoft official tweets.**

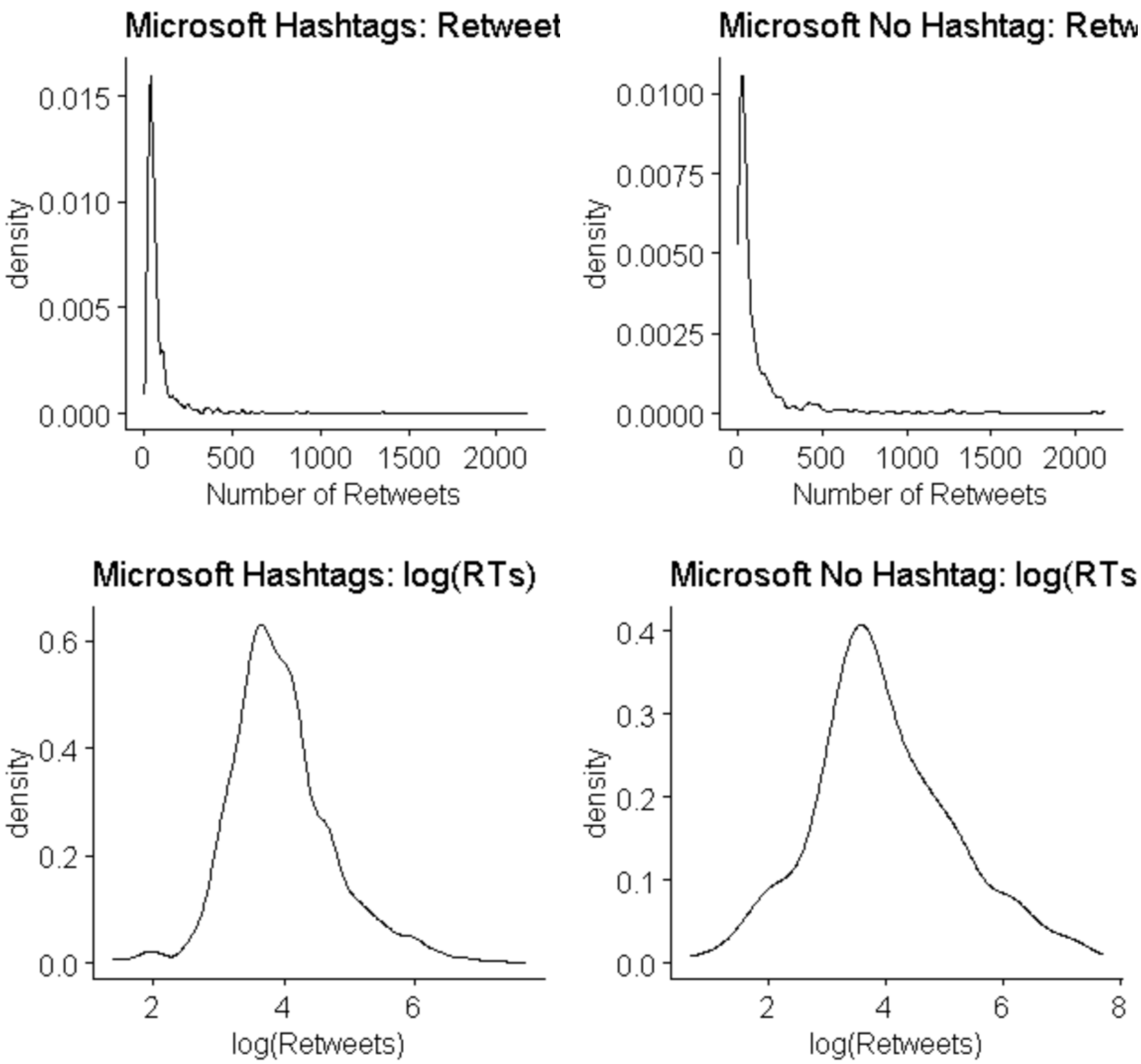
Microsoft Official: Number of Retweets





The log distribution does not pass a Shapiro-Wilk normality test.

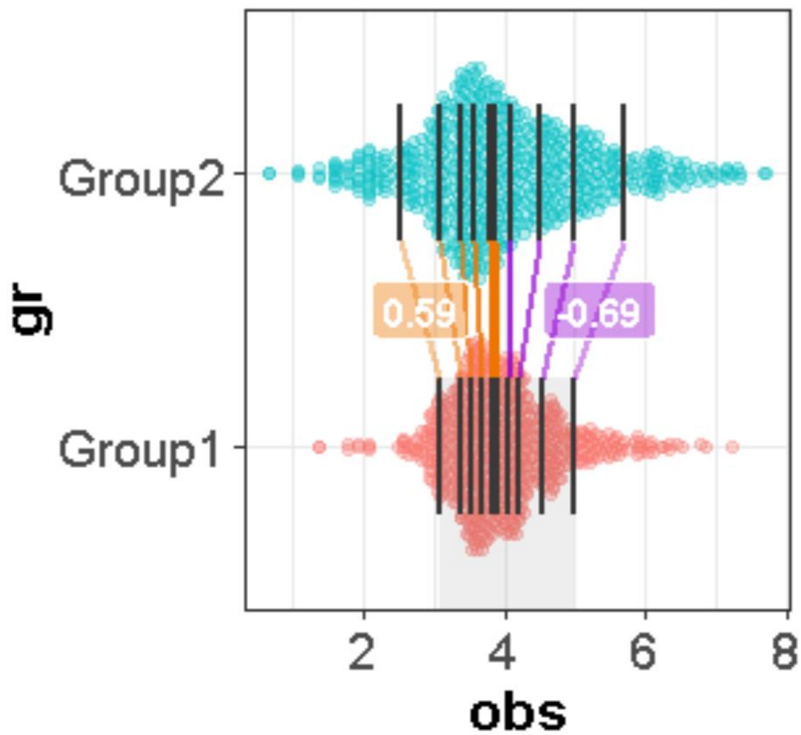




Again, unable to get x-axes to match, so note the differences. Furthermore, none of the above distributions pass a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 199129, p-value = 0.5655
alternative hypothesis: true location shift is not equal to 0
```

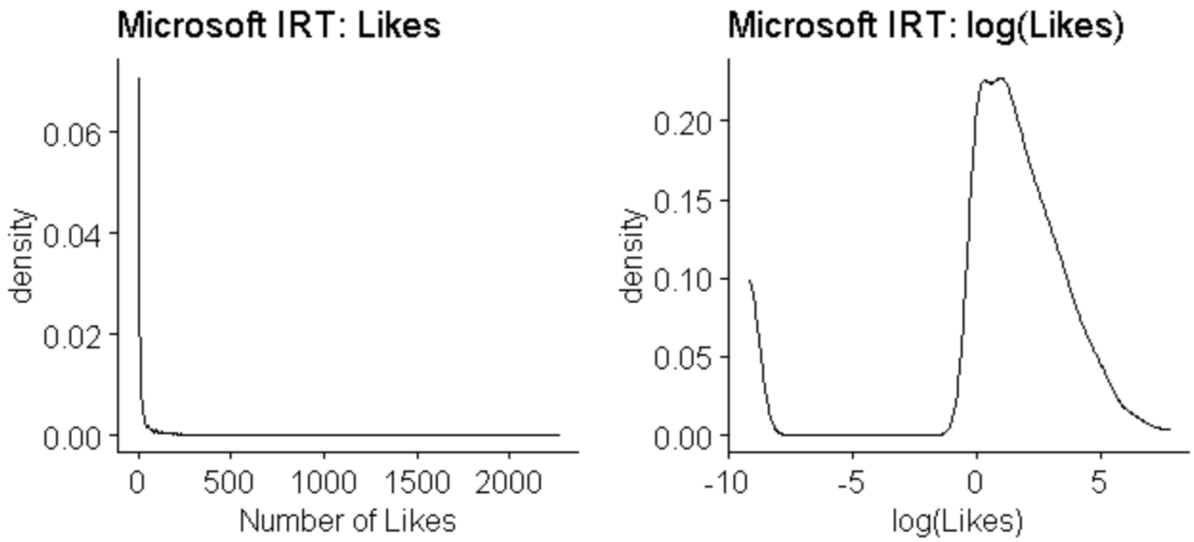
Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘retweet’ distributions of both populations are equal. Performing a shift function yields the following:




```
$`Group1 - Group2`
    q    Group1    Group2  difference    ci_lower   ci_upper      p_crit p_value
1 0.1  22.35501  12.53834    9.816672   5.9186443  13.848304 0.012500000   0.000
2 0.2  29.30109  22.01832    7.282773   3.5734703  11.094460 0.010000000   0.000
3 0.3  34.81051  29.33972    5.470784   1.8654607   9.588939 0.008333333   0.000
4 0.4  40.24720  36.20336    4.043839  -0.4608664   8.719668 0.016666667   0.038
5 0.5  47.45895  45.72755    1.731398  -3.8217578   7.385652 0.050000000   0.546
6 0.6  57.11580  60.74386   -3.628058 -15.1816002   5.567751 0.025000000   0.375
7 0.7  67.44227  91.48447  -24.042205 -45.2296156  -3.355942 0.007142857   0.000
8 0.8  93.67175 148.25008  -54.578331 -90.6544959 -16.387419 0.006250000   0.000
9 0.9 148.13881 297.38324 -149.244436 -283.4042504 -65.008428 0.005555556   0.000
```

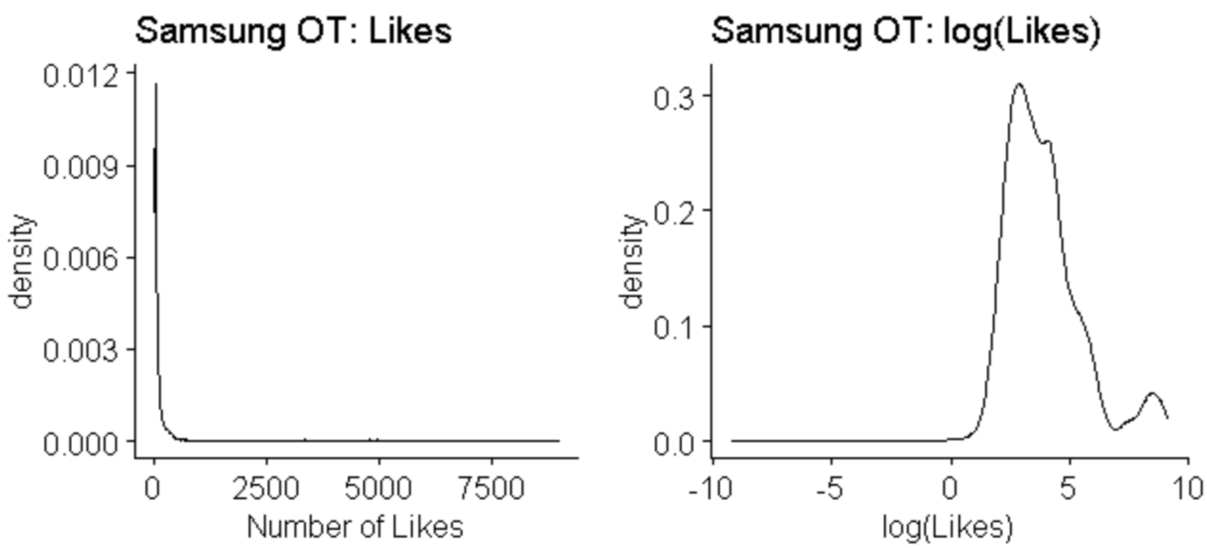
We can say, with 95% confidence, that the 1st, 2nd, and 3rd quantiles of group 2 (Microsoft official tweets not containing hashtags) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Microsoft official tweets containing hashtags). Furthermore, the 7th through 9th quantiles of group 2 would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1. However, considering we are lacking statistically significant Mann-Whitney U results as well, I don't feel comfortable making any such conclusions. **Inclusion of a hashtag does not seem to have a statistically significant effect on the number of retweets which a Microsoft official tweet receives.**

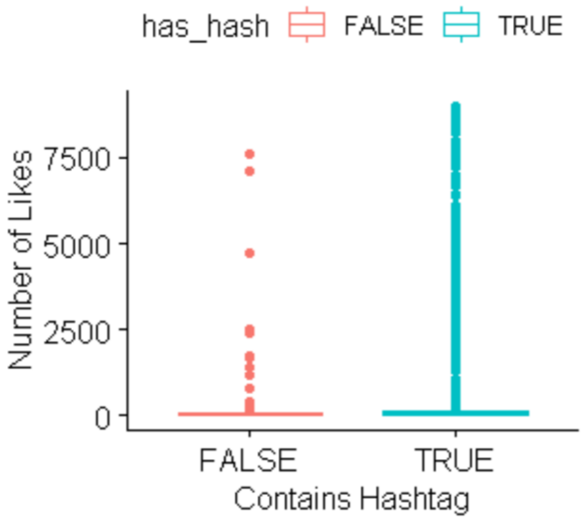
Microsoft IRT: Number of Likes



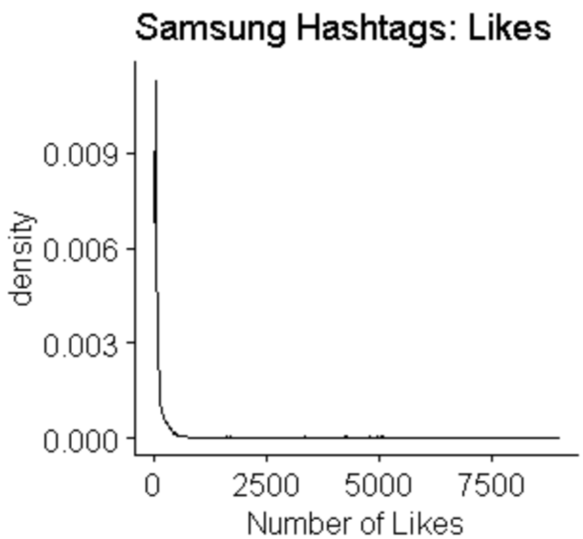
Apparently, there are only 9 Microsoft IRT tweets (not considered extreme outliers) containing hashtags, which isn't enough for analysis.

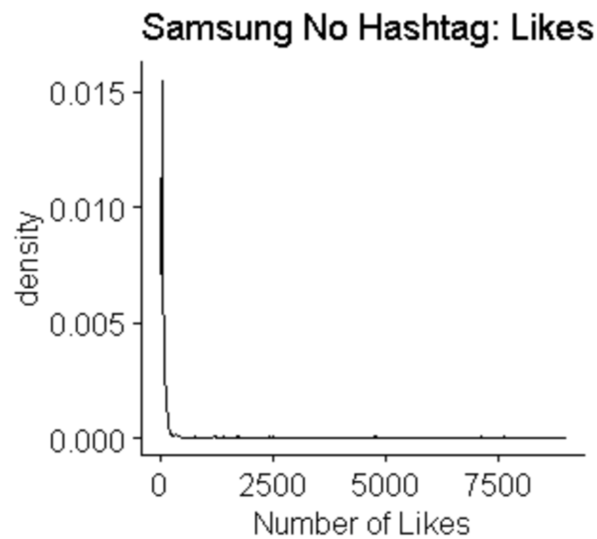
Samsung Official: Number of Likes

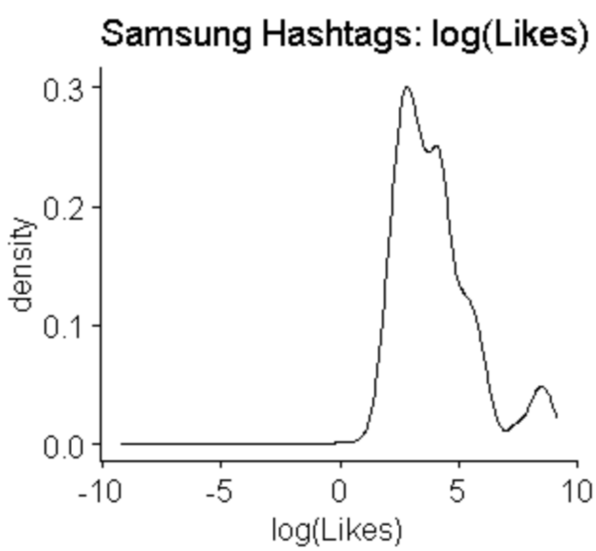


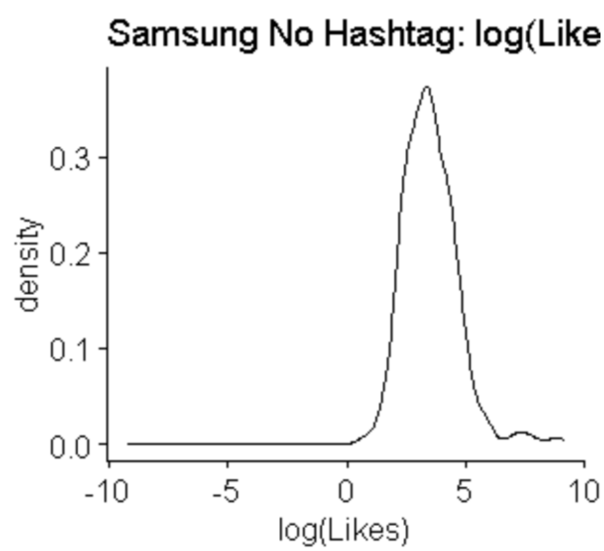

has_hash FALSE TRUE
Number of Likes
Contains Hashtag


has_hash FALSE TRUE
Log(Likes)
Contains Hashtag


Samsung Hashtags: Likes
density
Number of Likes

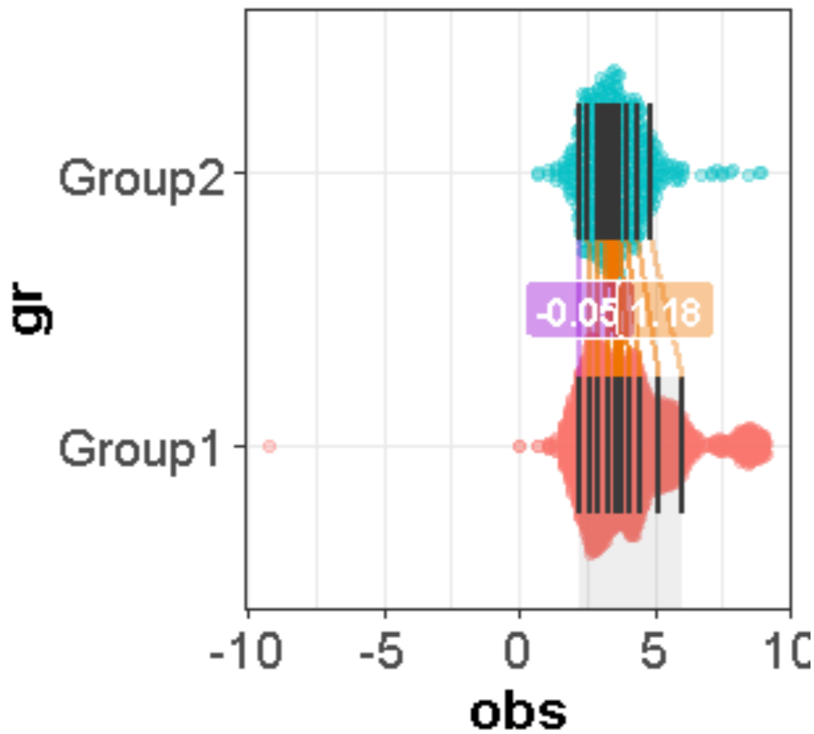

Samsung No Hashtag: Likes
density
Number of Likes


Samsung Hashtags: log(Likes)
density
log(Likes)


Samsung No Hashtag: log(Like
density
log(Likes)

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 703148, p-value = 2.185e-06
alternative hypothesis: true location shift is not equal to 0
```

Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

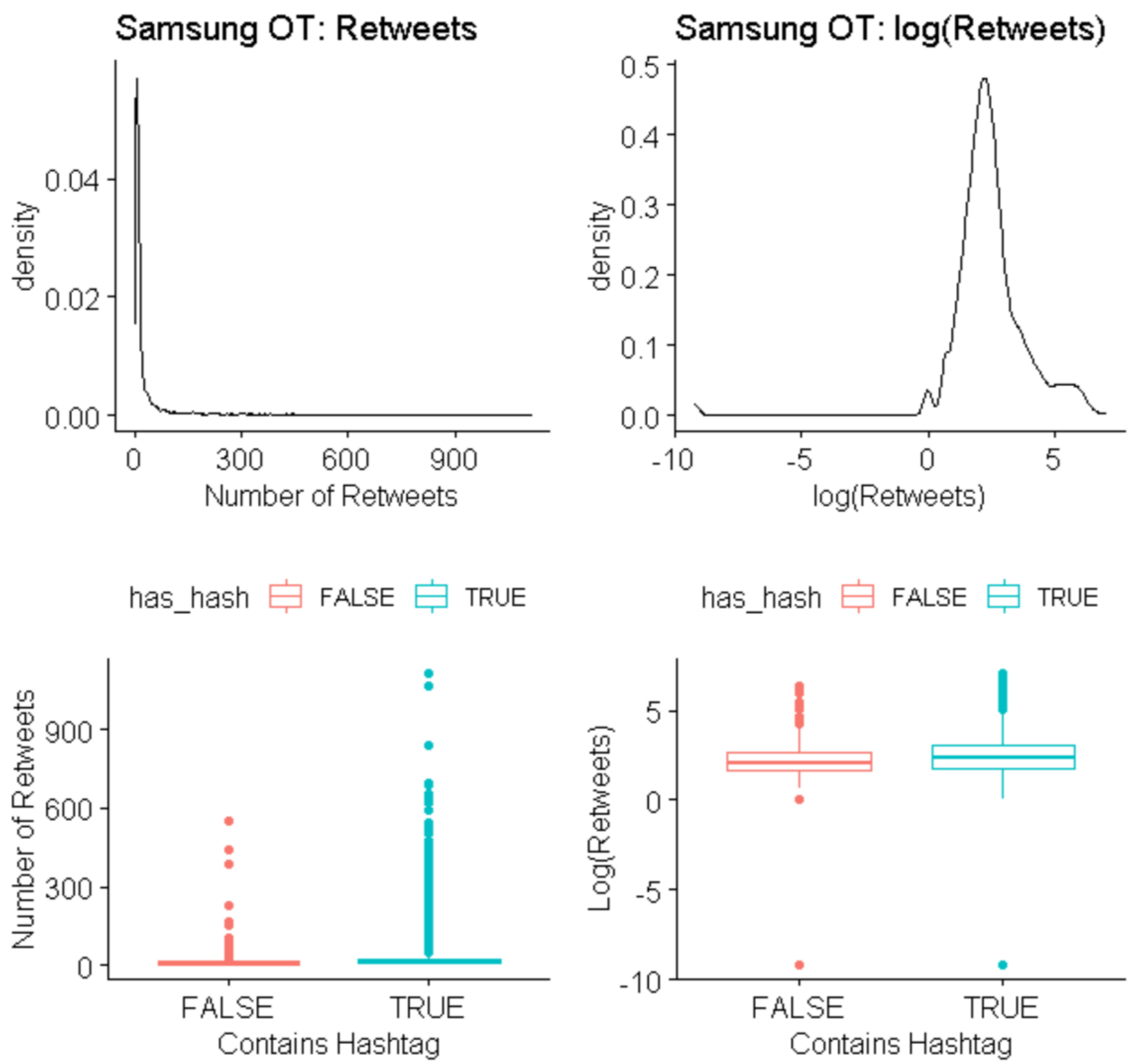


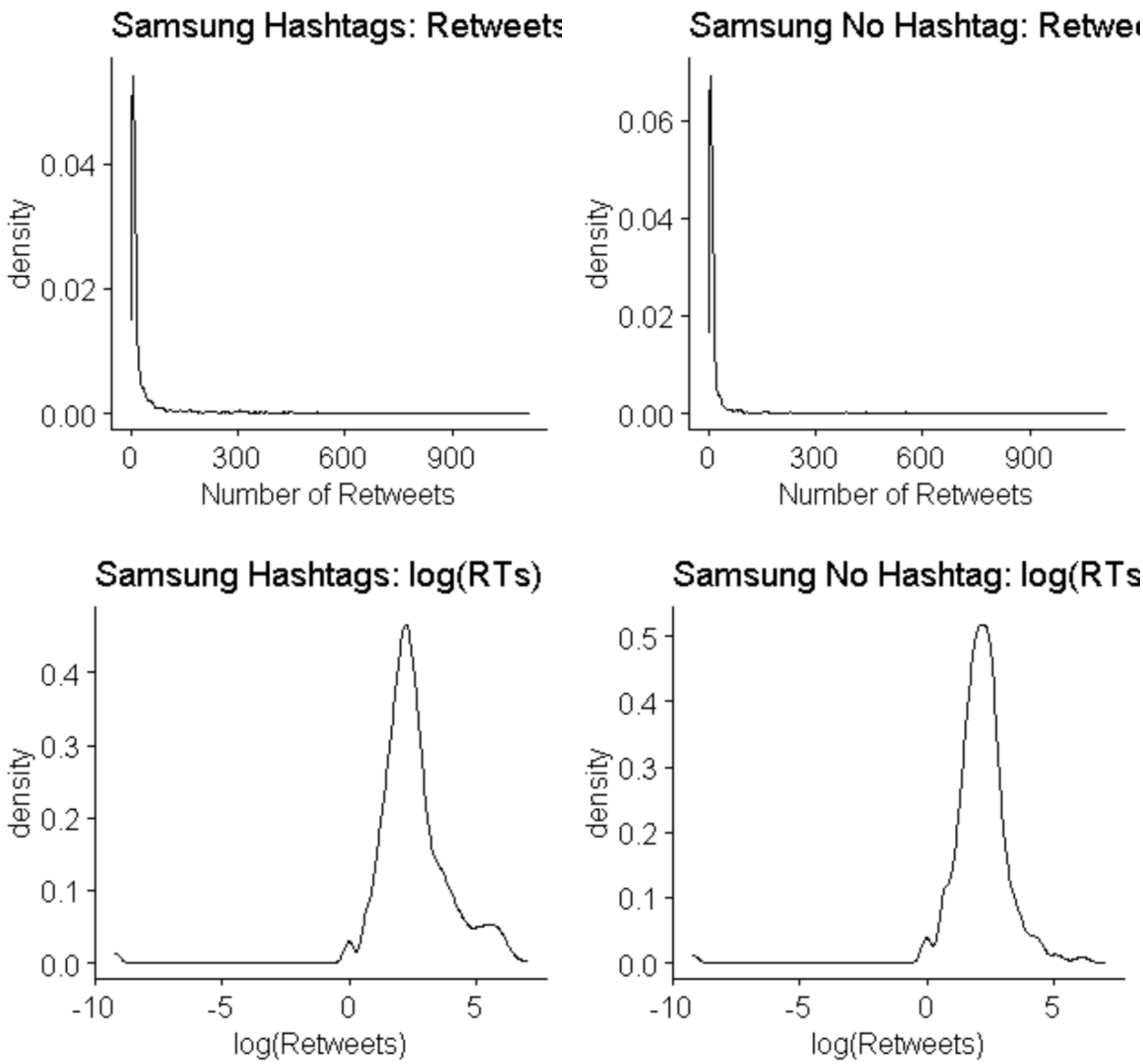

```
$`Group1 - Group2`
    q      Group1     Group2  difference   ci_lower    ci_upper      p_crit p_value
1 0.1    9.071984   9.529157  -0.4571725  -1.140081   0.8585556 0.025000000   0.448
2 0.2   13.357544  12.723451   0.6340923  -1.339493   2.5653068 0.016666667   0.360
3 0.3   18.451808  17.593889   0.8579197  -1.685259   2.9484780 0.050000000   0.450
4 0.4   25.813124  23.084941   2.7281835  -1.602088   6.4369660 0.012500000   0.113
5 0.5   38.640398  30.623750   8.0166477   2.201580  14.2393795 0.010000000   0.000
6 0.6   59.252171  38.063701  21.1884694  13.142038  27.5651713 0.008333333   0.000
7 0.7   86.029669  54.156534  31.8731354  16.951154  48.2838015 0.007142857   0.000
8 0.8  170.543874  75.978517  94.5653563  64.938300 124.9672990 0.006250000   0.000
9 0.9  399.656877 122.985962 276.6709160 203.008617 409.6213543 0.005555556   0.000
```

With 95% confidence, we may say that the 5th through 9th quantiles of group 2 (Samsung official tweets not containing hashtags) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Samsung official tweets containing hashtags). **Therefore, we may conclude**

that statistically significant differences exist between the two groups quantiles (in their distributions of likes, and the right halves specifically), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Samsung official tweets.

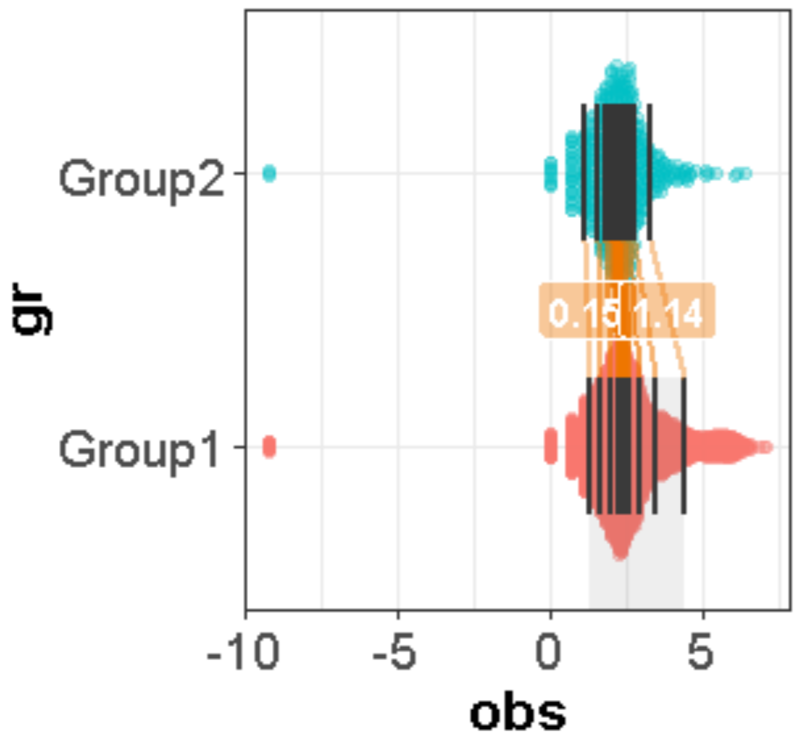
Samsung Official: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 725672, p-value = 1.487e-09
alternative hypothesis: true location shift is not equal to 0
```

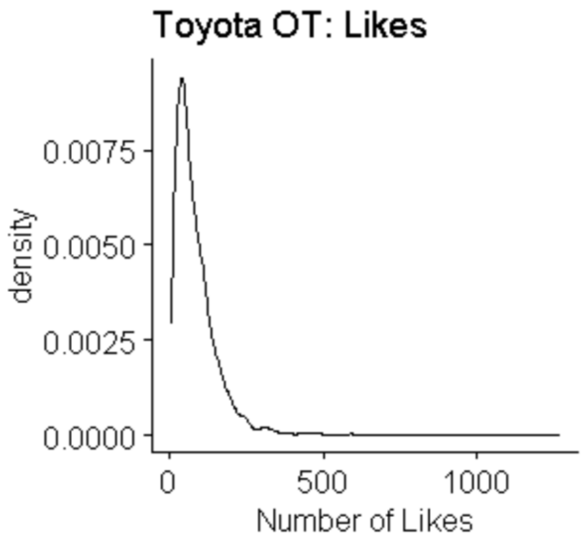
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

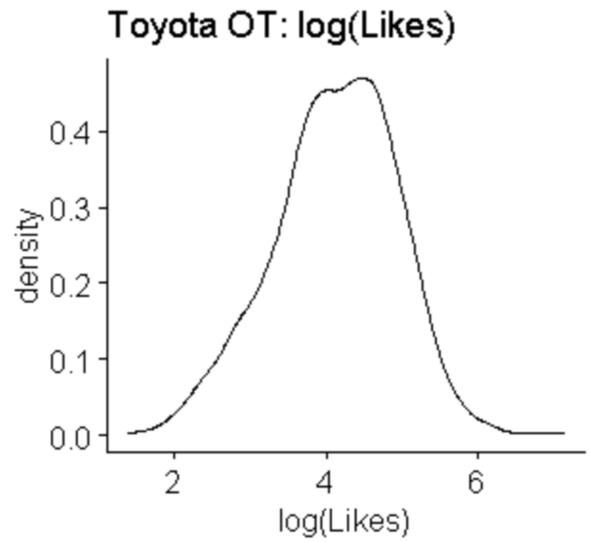



```
$`Group1 - Group2`
    q    Group1    Group2 difference     ci_lower  ci_upper      p_crit p_value
1 0.1  3.533908  3.032431  0.5014766 -0.277884211  1.110492 0.050000000   0.209
2 0.2  5.020979  4.661710  0.3592686 -0.008893717  1.235633 0.025000000   0.030
3 0.3  6.997632  5.915911  1.0817206  0.426841469  1.828171 0.016666667   0.001
4 0.4  8.823078  7.114981  1.7080969  0.519467305  2.406717 0.012500000   0.000
5 0.5 10.508266  8.479847  2.0284185  0.632893175  2.982375 0.010000000   0.000
6 0.6 13.285372 10.472387  2.8129849  1.410570020  4.268545 0.008333333   0.000
7 0.7 17.761566 12.720829  5.0407372  3.123583216  7.328672 0.007142857   0.000
8 0.8 31.212956 15.761365 15.4515903  9.677164448 21.524040 0.006250000   0.000
9 0.9 79.986154 25.696109 54.2900449 34.202284642 77.531734 0.005555556   0.000
```

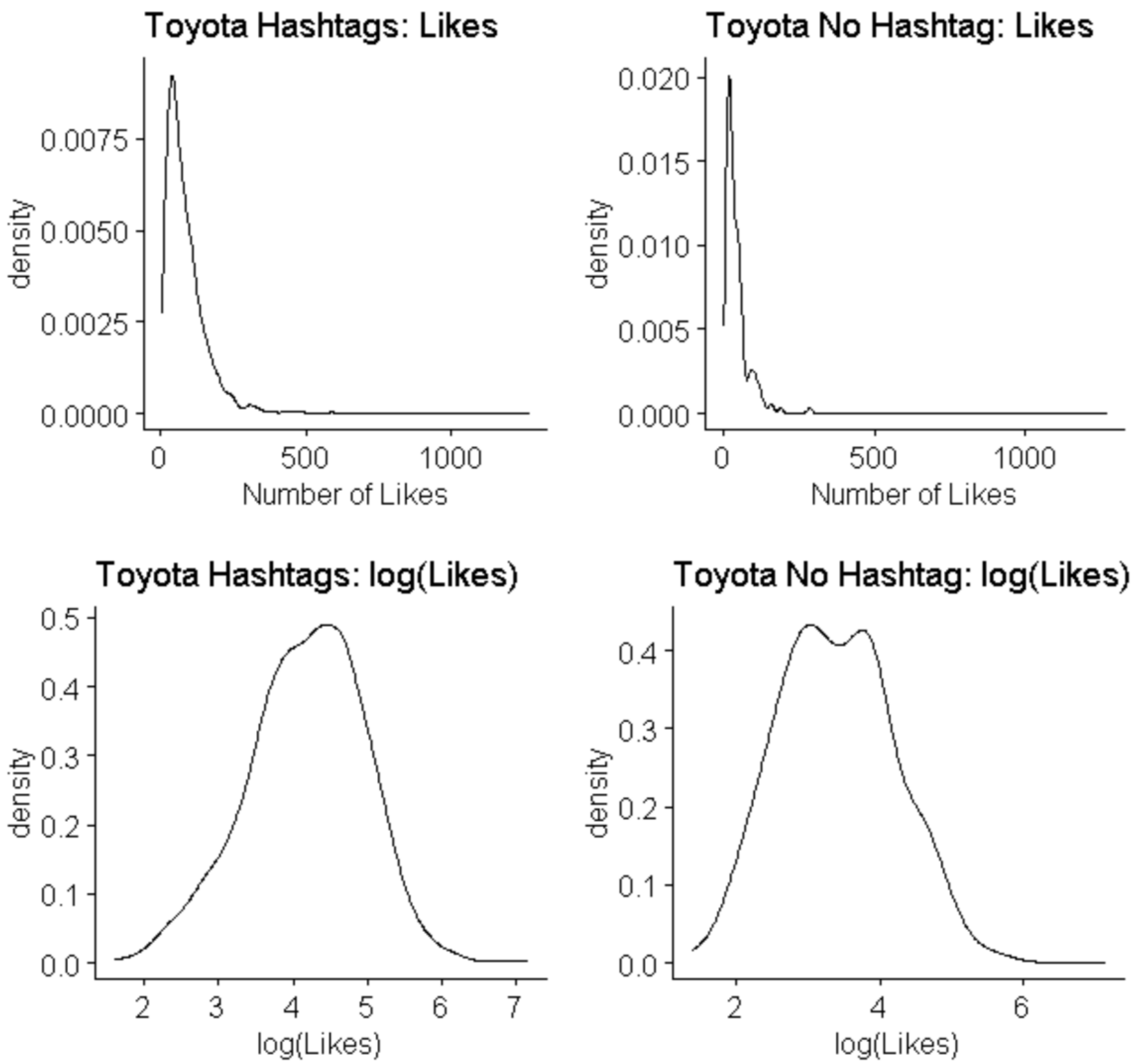
With 95% confidence we may say that the 3rd through 9th quantiles of group 2 (Samsung official tweets not containing hashtags) would need to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Samsung official tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Samsung official tweets.**

Toyota Official: Number of Likes





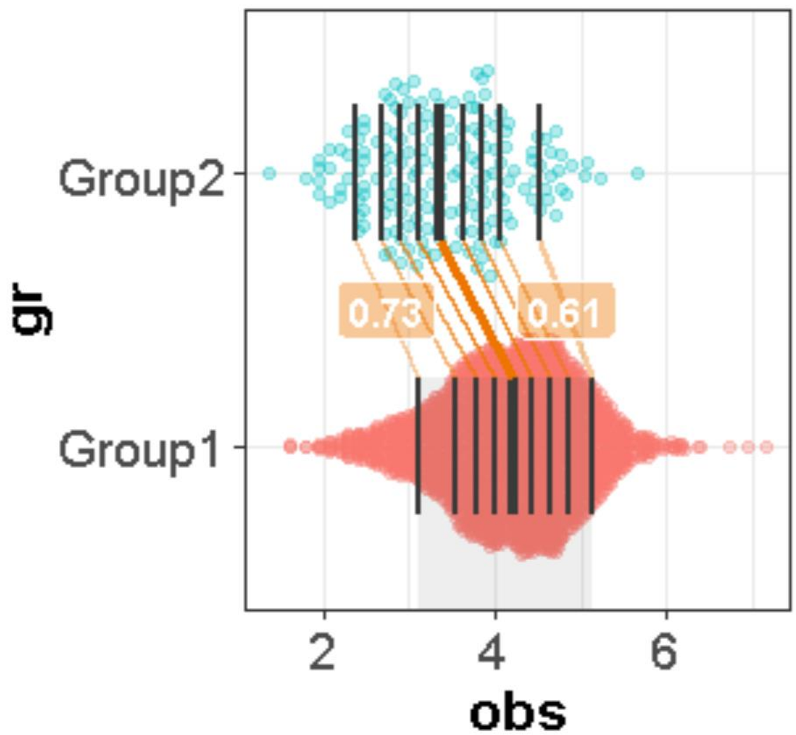
The log distribution does not pass a Shapiro-Wilk normality test.



Unable to get the x-axes to match, but note the differences. Furthermore, only the bottom right log distribution passes a Shapiro-Wilk normality test.

```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Likes` and noHashtag_tweets$`Number of Likes`
W = 301651, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

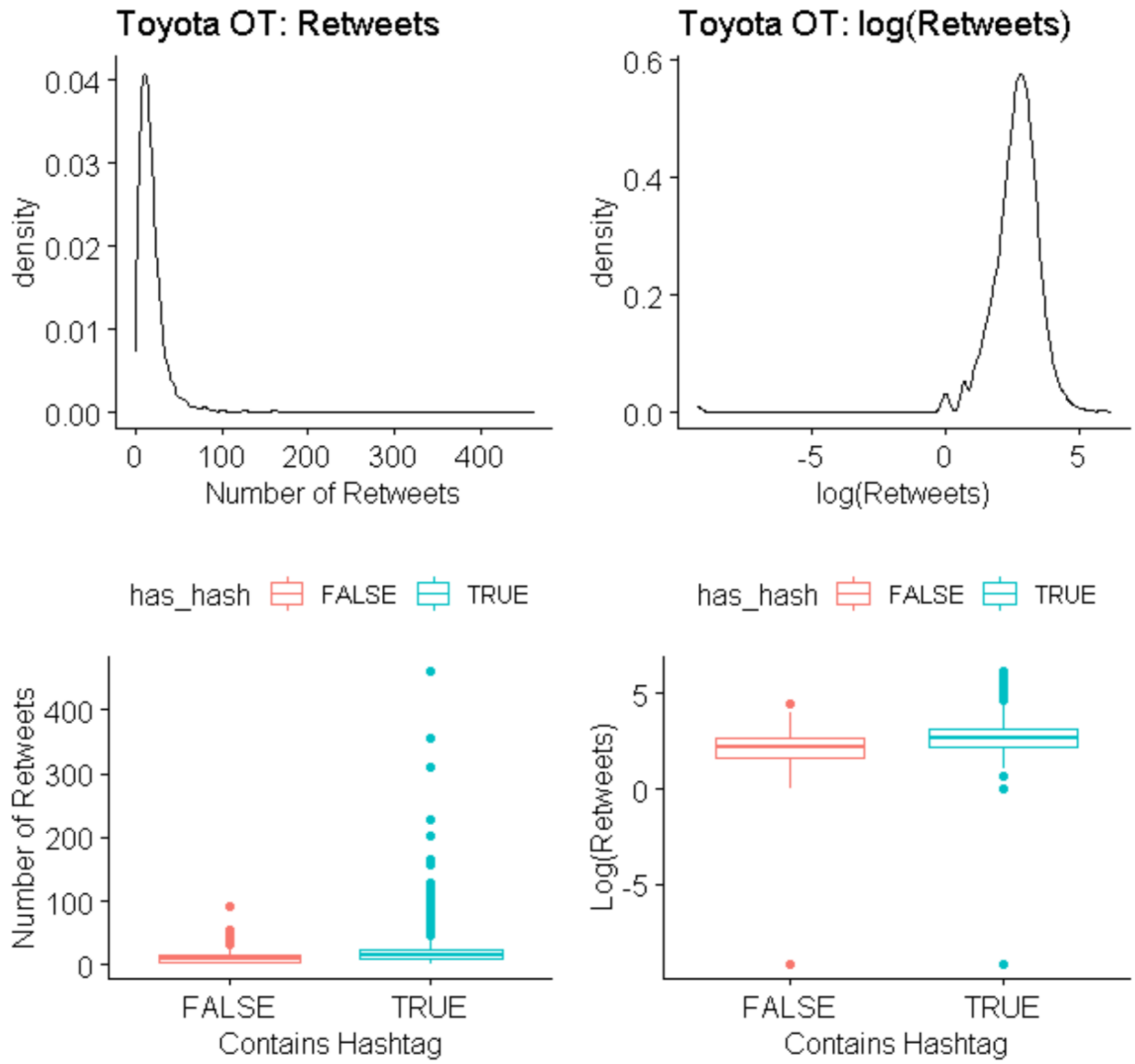
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

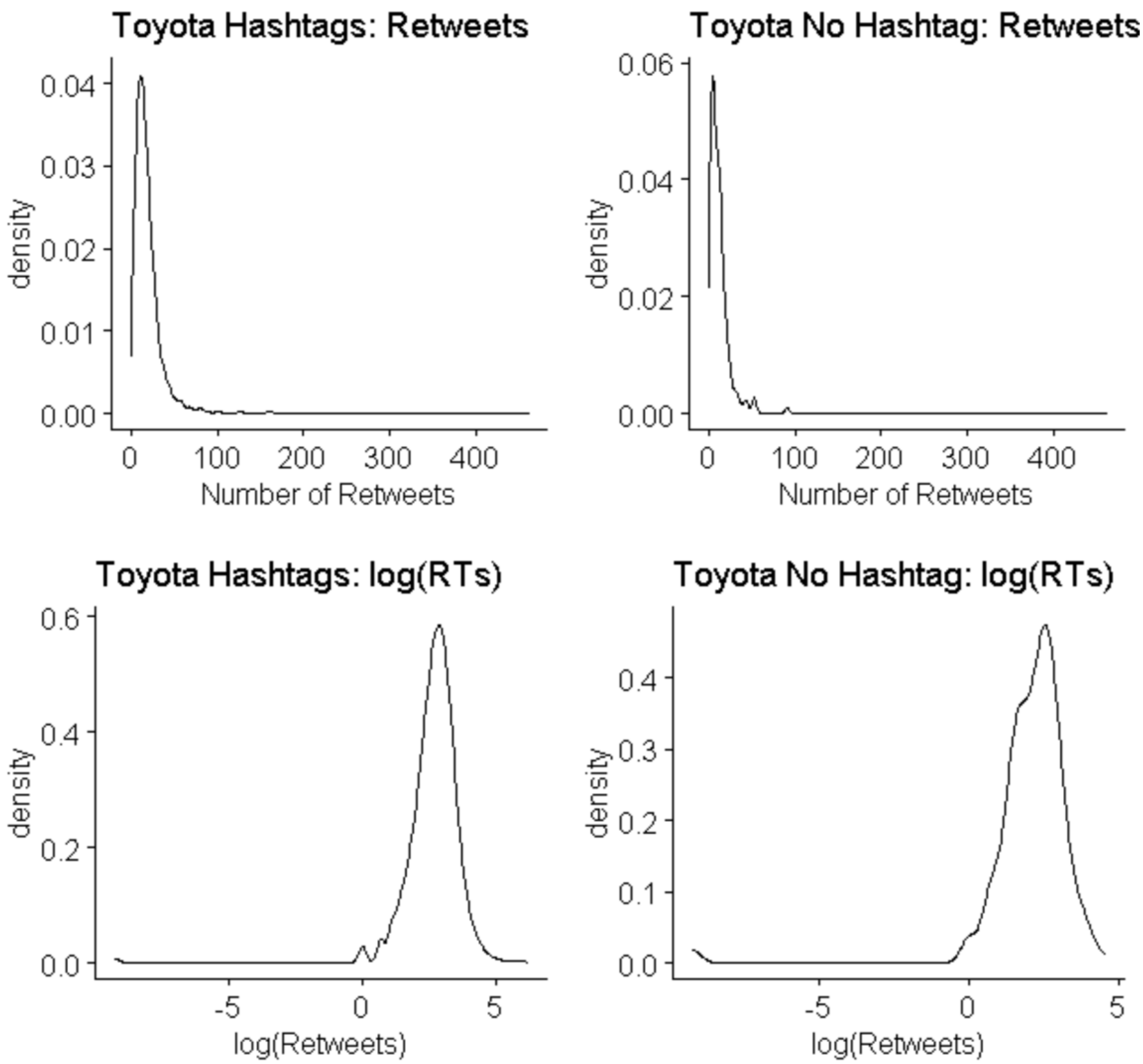



```
$`Group1 - Group2`
    q    Group1   Group2 difference  ci_lower  ci_upper      p_crit p_value
1 0.1  22.25734 10.78481   11.47253  9.209582  13.74978 0.050000000       0
2 0.2  34.27635 14.83636   19.43999 16.367646  22.60981 0.025000000       0
3 0.3  43.99858 18.31903   25.67955 20.916063  29.17087 0.016666667       0
4 0.4  54.82912 22.88390   31.94522 26.054446  36.76440 0.012500000       0
5 0.5  68.54021 28.95712   39.58309 30.028634  46.18338 0.010000000       0
6 0.6  83.38358 38.56074   44.82285 36.551152  54.96989 0.008333333       0
7 0.7 102.84531 46.88782   55.95749 47.279728  64.96496 0.007142857       0
8 0.8 127.30631 58.08566   69.22065 45.678759  81.45351 0.006250000       0
9 0.9 169.69596 92.62949   77.06647 56.159297 107.69077 0.005555556       0
```

With 95% confidence, we can say that every quantile of group 2 (Toyota official tweets not containing hashtags) would have to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Toyota official tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Toyota official tweets.**

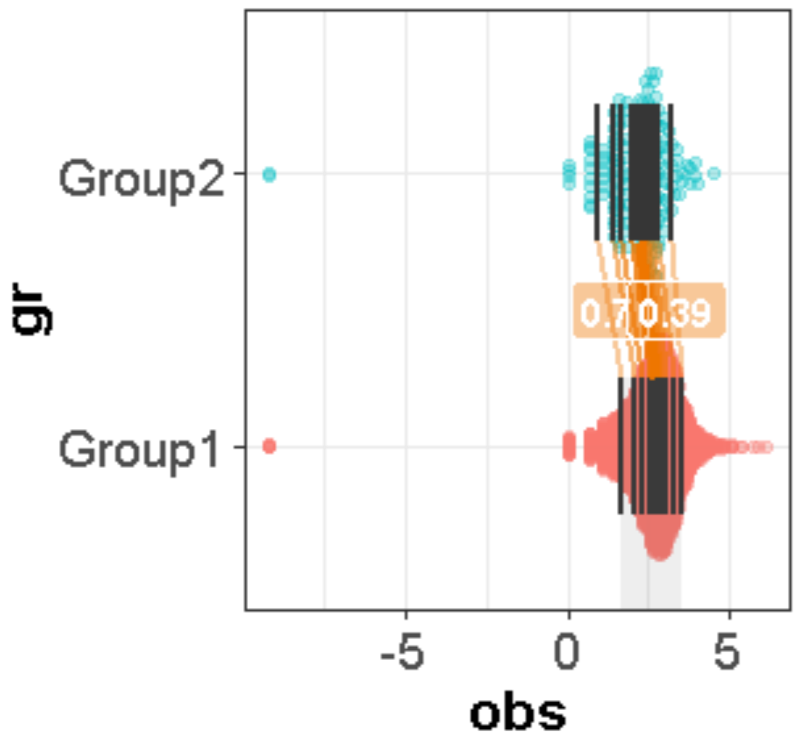
Toyota Official: Number of Retweets






```
        Wilcoxon rank sum test with continuity correction

data:  Hashtag_tweets$`Number of Retweets` and noHashtag_tweets$`Number of Retweets`
W = 270429, p-value = 1.523e-14
alternative hypothesis: true location shift is not equal to 0
```

Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:



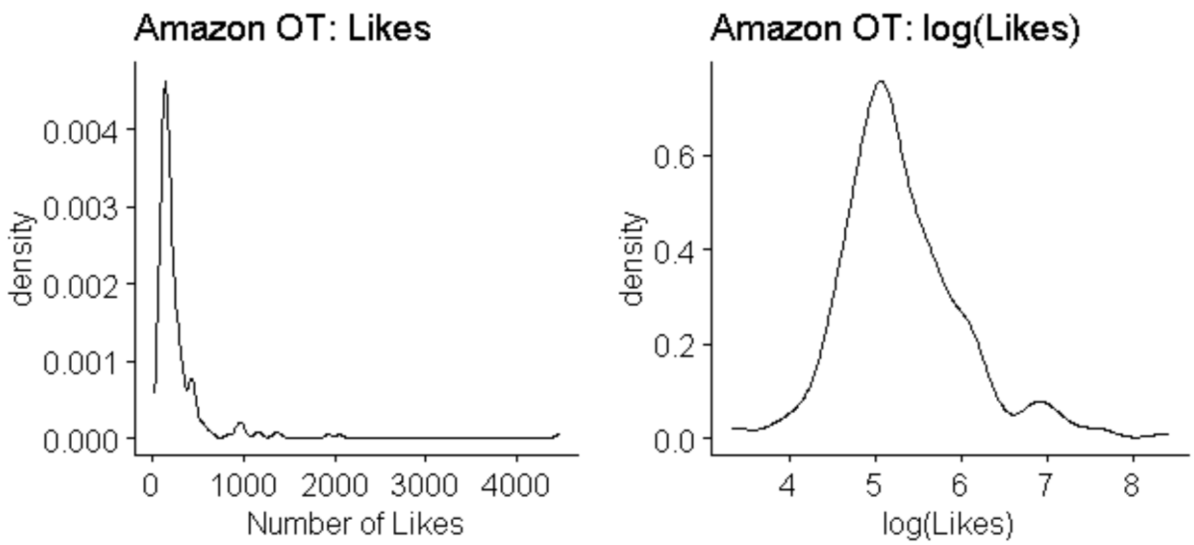

```
$`Group1 - Group2`
    q    Group1    Group2 difference ci_lower  ci_upper      p_crit p_value
1 0.1  5.176544  2.633368   2.543176 1.374437  3.590466 0.025000000   0.000
2 0.2  8.056278  4.249905   3.806374 2.944648  4.882509 0.016666667   0.000
3 0.3 10.470837  5.485158   4.985679 3.064815  6.217568 0.012500000   0.000
4 0.4 13.020027  7.359872   5.660156 3.682869  7.474020 0.010000000   0.000
5 0.5 15.458948  9.485033   5.973915 3.439765  8.096481 0.008333333   0.000
6 0.6 18.318132 11.854229   6.463903 4.208985  8.896528 0.007142857   0.000
7 0.7 21.723385 14.194665   7.528720 4.893077  9.820455 0.006250000   0.000
8 0.8 26.733113 17.284887   9.448226 4.991959 12.432035 0.005555556   0.000
9 0.9 35.721213 24.329234  11.391979 4.302365 15.750582 0.050000000   0.008
```

With 95% confidence, we can say that every quantile of group 2 (Toyota official tweets not containing hashtags) would have to be shifted up by significant (non-zero) amounts to match their counterparts in group 1 (Toyota official tweets containing hashtags). **Therefore, we may conclude that statistically significant differences exist between the two groups quantiles (in their distributions of retweets), and potentially one underlying factor explaining these differences is the inclusion of a hashtag in Toyota official tweets.**

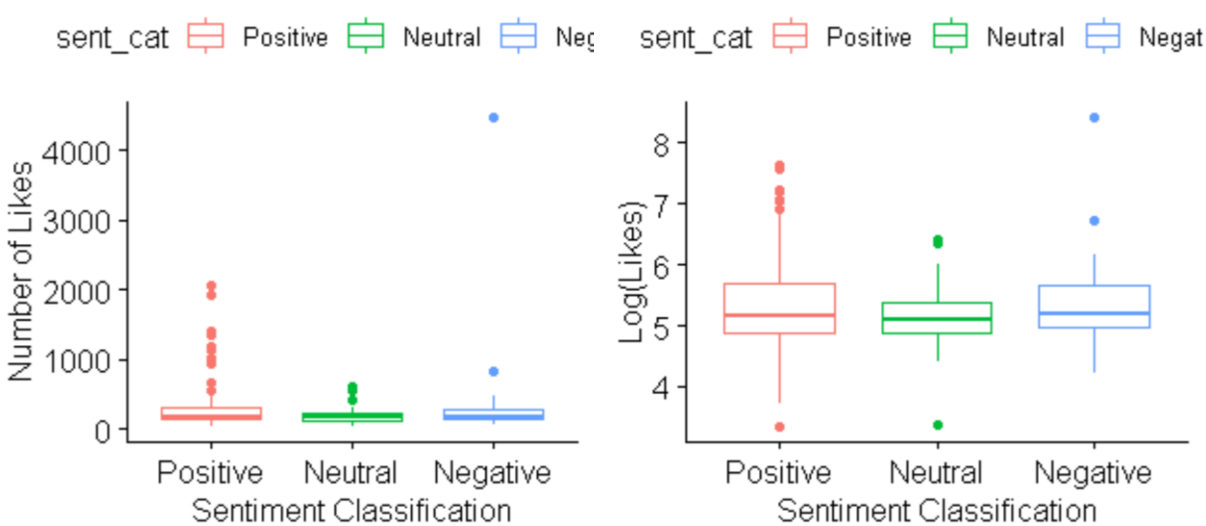
<u>Section 9:</u> Sentiment Analysis

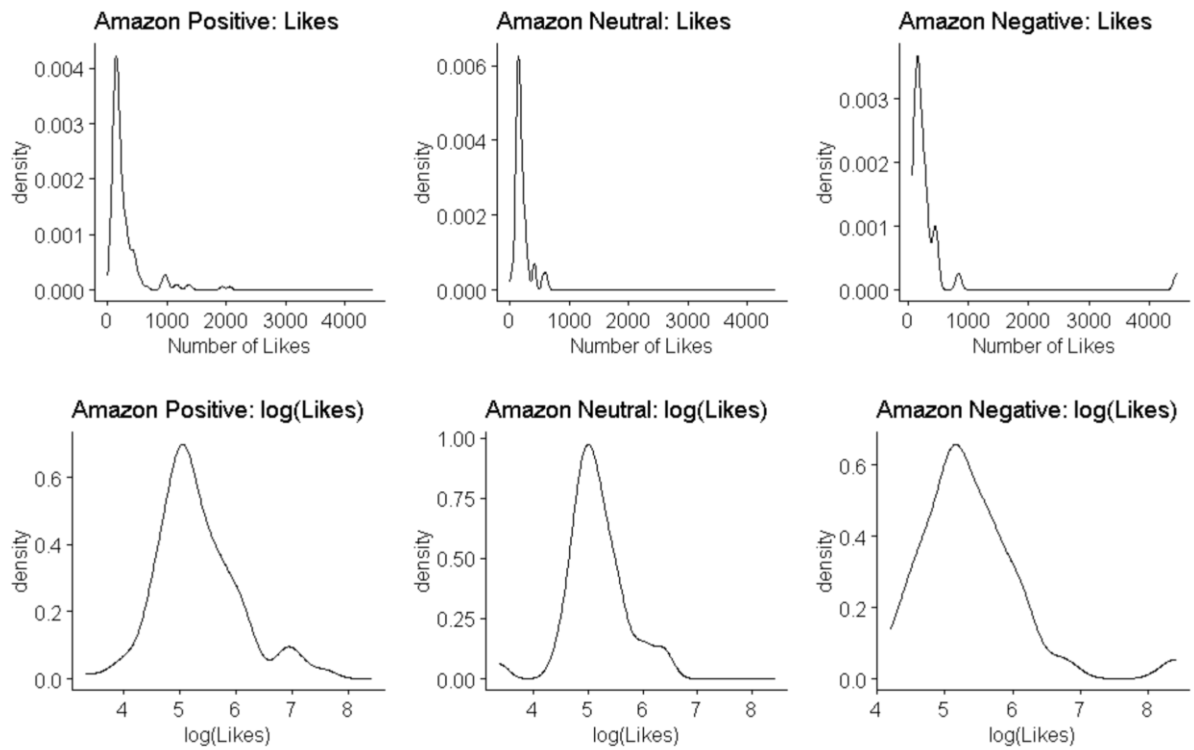
Section 9 contains the visualizations, notes, results, and conclusions made when examining companies and tweet categories for a potential sentiment effect.

Amazon Official: Number of Likes



The log distribution does not pass a Shapiro-Wilk normality test.





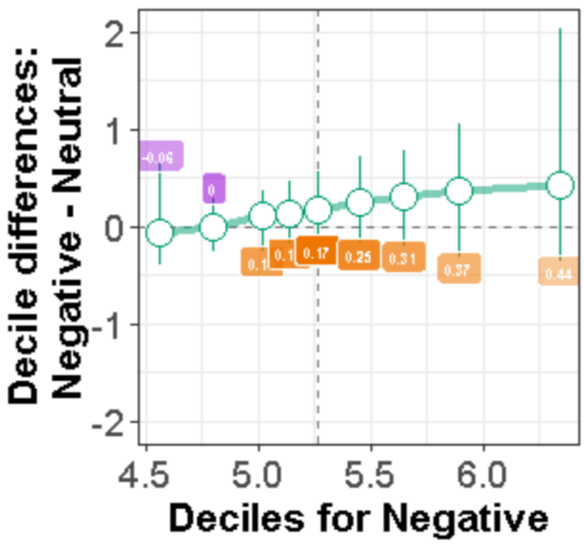
None of the above distributions pass a Shapiro-Wilk normality test.

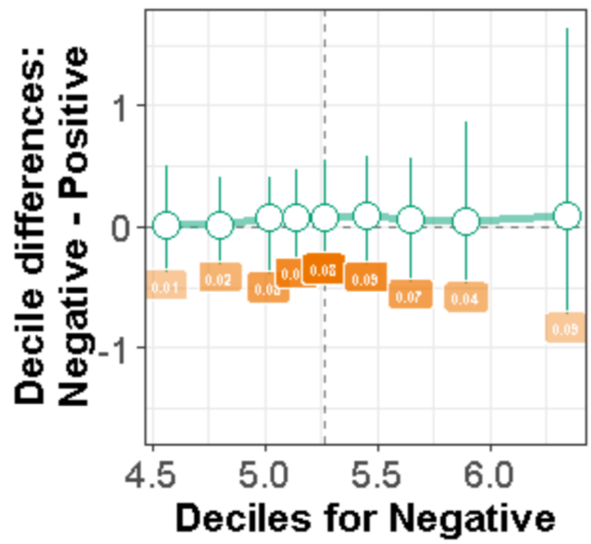
```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 1.8028, df = 2, p-value = 0.406
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'like' distributions of all populations are equal. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

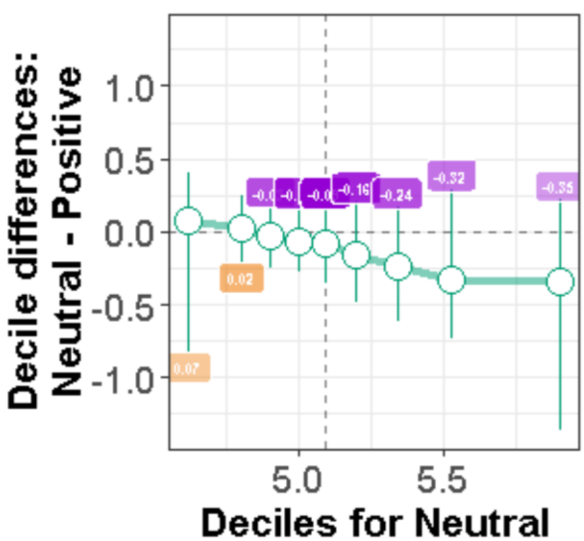
```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z    P.unadj     P.adj
1  Negative - Neutral  1.2715579 0.2035302 0.6105907
2 Negative - Positive  0.5504181 0.5820326 1.0000000
3  Neutral - Positive -1.1143103 0.2651461 0.7954383
```

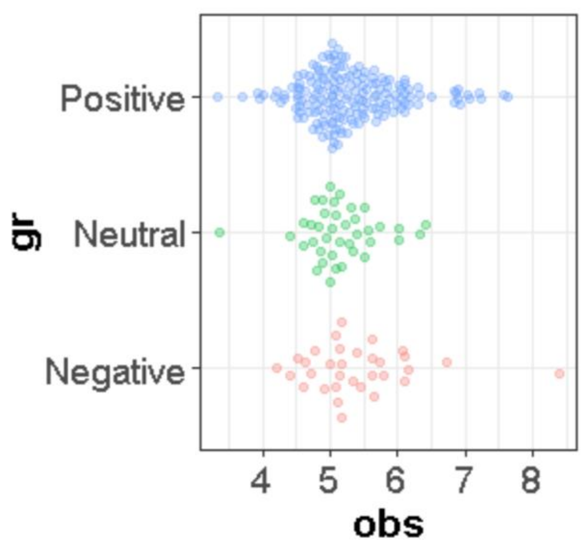
Further confirmation that all distributions are equal. Performing the shift function yields the following:





The figure on the above left indicates that the first quantile of neutral tweets would need to be down-shifted slightly to match the first quantile of negative tweets, the second quantiles match well, and the 3rd through 9th quantiles of neutral tweets would all need to be up-shifted, to a greater extent each time, to match the same quantiles in negative tweets. On the above right we can see that essentially every quantile of positive tweets would need to be up-shifted ever so slightly to match their counterparts within the set of negative tweets.





Above left we can see that the first two quantiles of positive tweets would need to be up-shifted slightly to match their counterparts in the set of neutral tweets, while essentially every other quantile would need to be down-shifted.

$`Negative - Neutral`

	q	Negative	Neutral	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	96.18539	102.9634	-6.777968	-37.87178	48.77371	0.025000000	0.815
2	0.2	123.35476	122.1635	1.191286	-31.84030	39.66753	0.050000000	0.911
3	0.3	151.96140	134.8821	17.079265	-29.84600	59.59080	0.016666667	0.447
4	0.4	171.41353	148.7208	22.692720	-28.21702	91.27872	0.012500000	0.212
5	0.5	194.66947	162.8574	31.812111	-24.22707	122.39654	0.008333333	0.153
6	0.6	235.59245	181.7105	53.881903	-34.82488	185.86902	0.007142857	0.113
7	0.7	287.28714	209.4347	77.852406	-64.12837	238.01164	0.005555556	0.105
8	0.8	370.54293	253.9943	116.548591	-75.83855	1348.38552	0.006250000	0.112
9	0.9	726.46638	377.7764	348.689954	-129.89459	3056.04339	0.010000000	0.155

$`Negative - Positive`

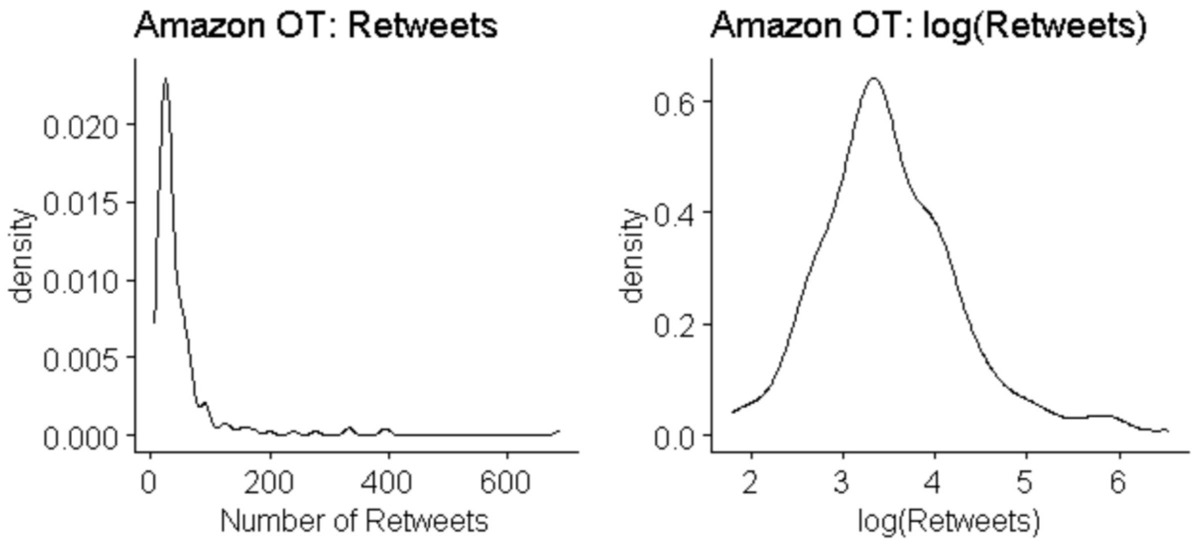
	q	Negative	Positive	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	96.18539	94.27942	1.905968	-26.76951	49.01391	0.010000000	0.771
2	0.2	123.35476	119.45555	3.899216	-36.65301	57.48177	0.005000000	0.787
3	0.3	151.96140	139.34591	12.615488	-46.40962	65.54580	0.001666667	0.532
4	0.4	171.41353	157.97372	13.439814	-44.28475	97.97382	0.001111111	0.320
5	0.5	194.66947	177.26227	17.407198	-36.21144	136.62371	0.001250000	0.453
6	0.6	235.59245	213.45435	22.138099	-76.37297	173.87675	0.001428571	0.543
7	0.7	287.28714	265.74624	21.540899	-102.07300	276.64799	0.002500000	0.642
8	0.8	370.54293	348.14165	22.401282	-162.95589	1281.05760	0.003333333	0.745
9	0.9	726.46638	528.45292	198.013458	-506.23827	3306.97532	0.002000000	0.636

$`Neutral - Positive`

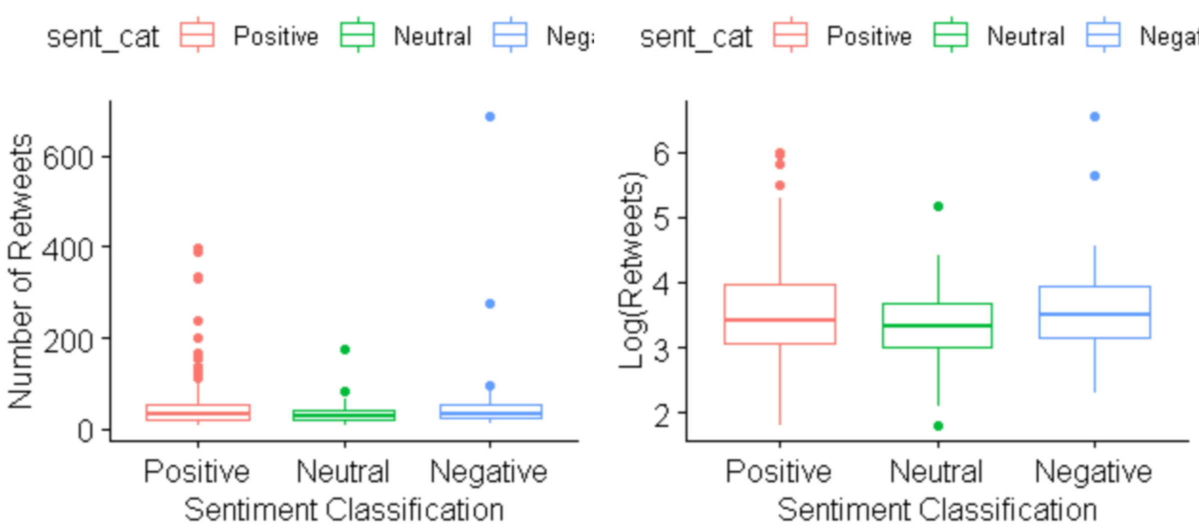
	q	Neutral	Positive	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	102.9634	94.27942	8.683936	-52.24344	44.77587	0.0006666667	0.561
2	0.2	122.1635	119.45555	2.707931	-30.66777	37.20868	0.0020000000	0.744
3	0.3	134.8821	139.34591	-4.463776	-37.85916	32.63833	0.0010000000	0.699
4	0.4	148.7208	157.97372	-9.252907	-47.75082	32.10347	0.0005000000	0.412
5	0.5	162.8574	177.26227	-14.404914	-70.03311	46.07115	0.0004000000	0.332
6	0.6	181.7105	213.45435	-31.743804	-127.52438	53.51822	0.0003333333	0.197
7	0.7	209.4347	265.74624	-56.311507	-162.12722	56.26189	0.0002222222	0.085
8	0.8	253.9943	348.14165	-94.147309	-244.88979	132.80051	0.0002500000	0.104
9	0.9	377.7764	528.45292	-150.676496	-728.17487	141.51408	0.0002857143	0.147

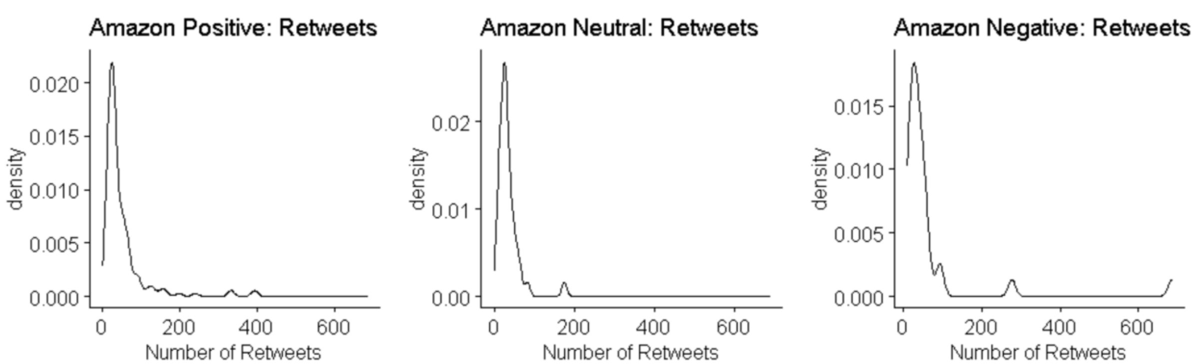
Considering each confidence interval contains the value 0, we may not conclude any quantile differences in tweet performance between any two groups. We fail to reject the null hypotheses that the matching quantiles of any two groups distributions of likes are equal to one another. **Sentiment does not seem to have a statistically significant effect on the number of likes which an Amazon official tweet receives.**

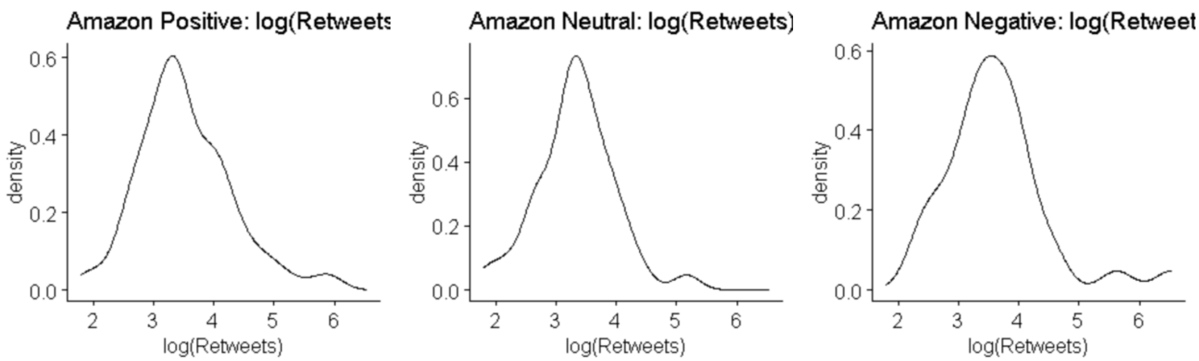
Amazon Official: Number of Retweets



The log distribution does not pass a Shapiro-Wilk normality test.







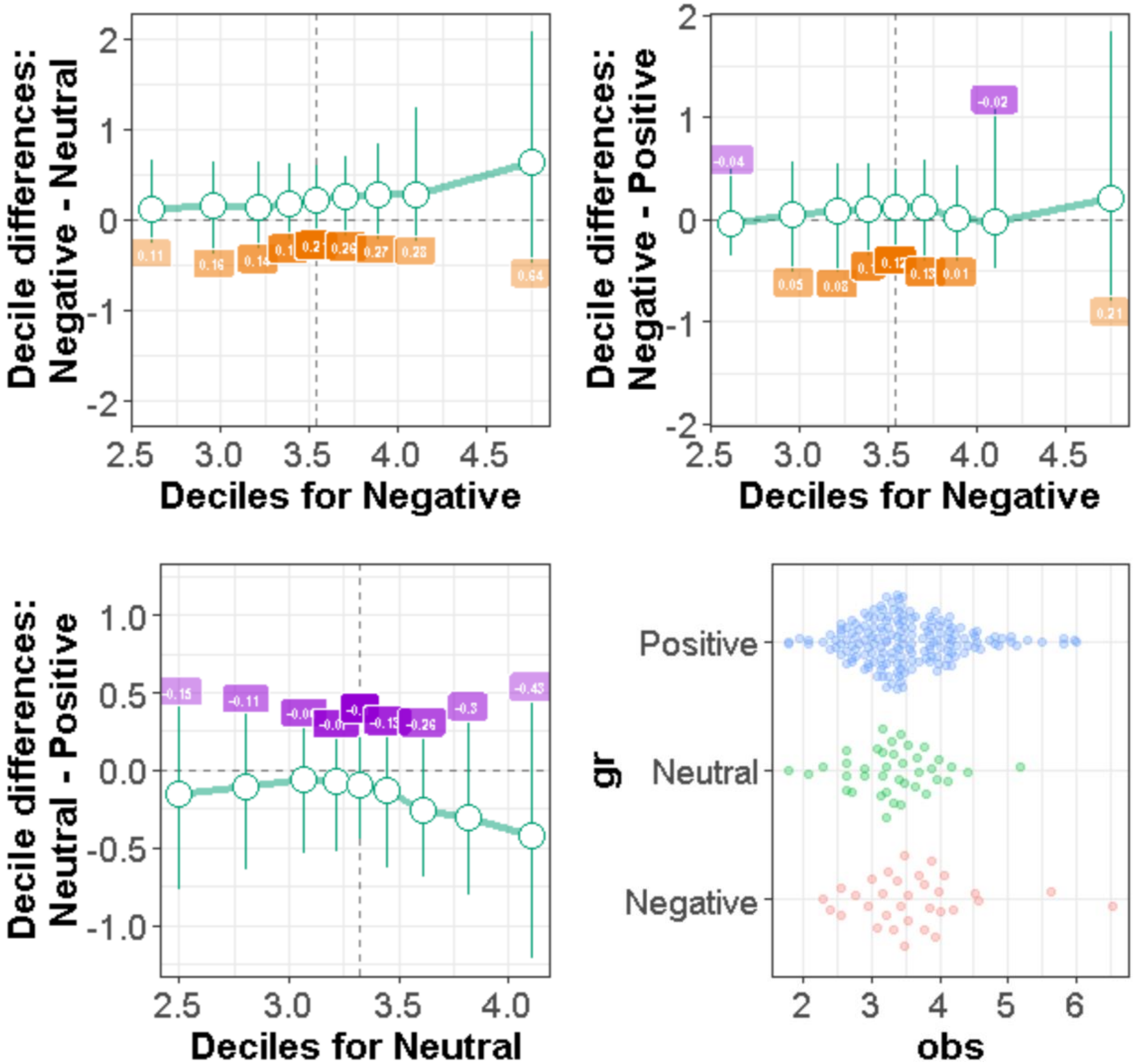
Only the middle distribution above passes a Shapiro-Wilk normality test.

```
> kruskal.test(`Number of Retweets` ~ sent_cat, data = tweets)

        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 2.1644, df = 2, p-value = 0.3388
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'like' distributions of all populations are equal. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z    P.unadj      P.adj
1  Negative - Neutral  1.3614599 0.1733684 0.5201052
2 Negative - Positive  0.5191783 0.6036364 1.0000000
3  Neutral - Positive -1.2692609 0.2043480 0.6130441
```

Further confirmation that all distributions are equal. Performing the shift function yields the following:

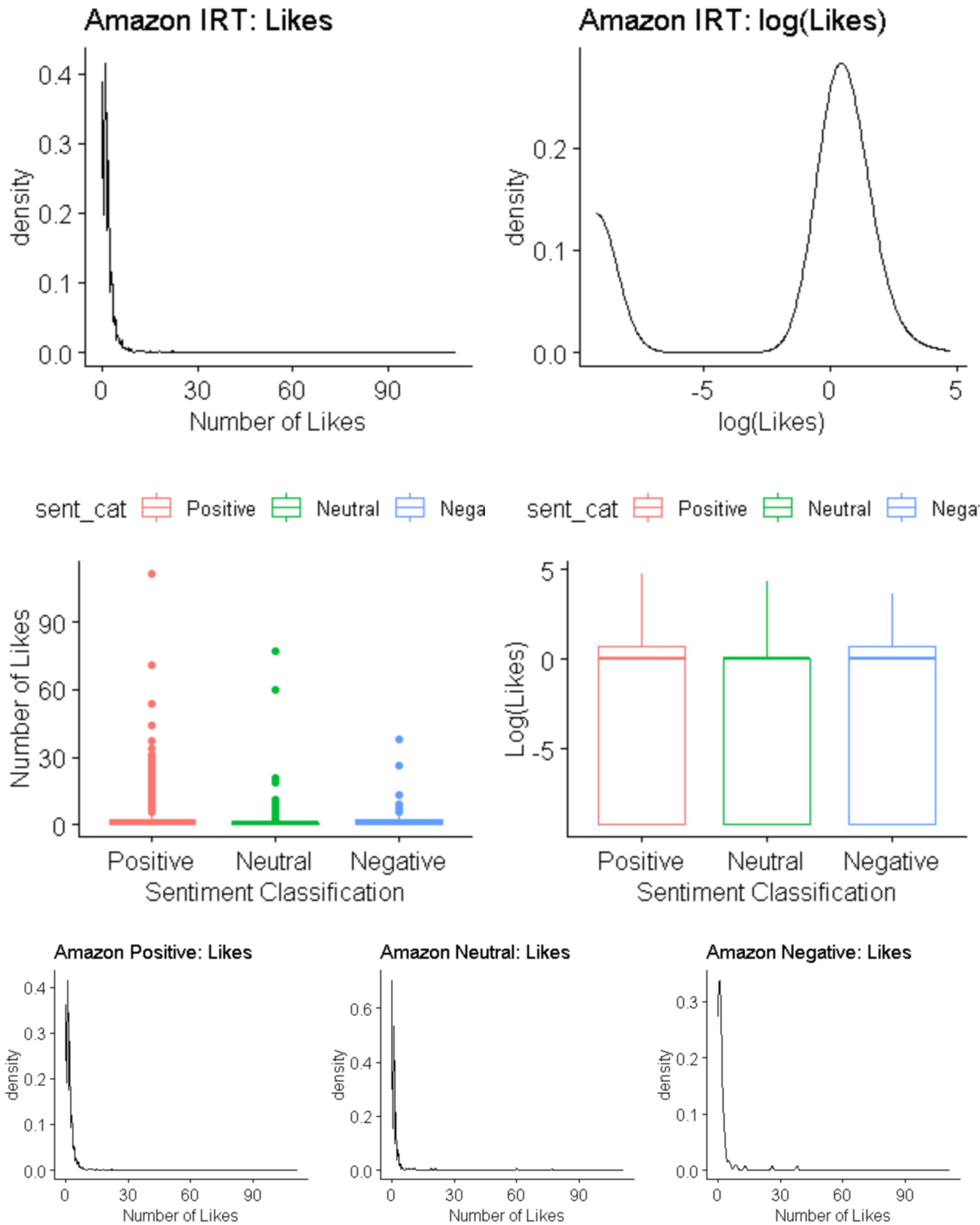


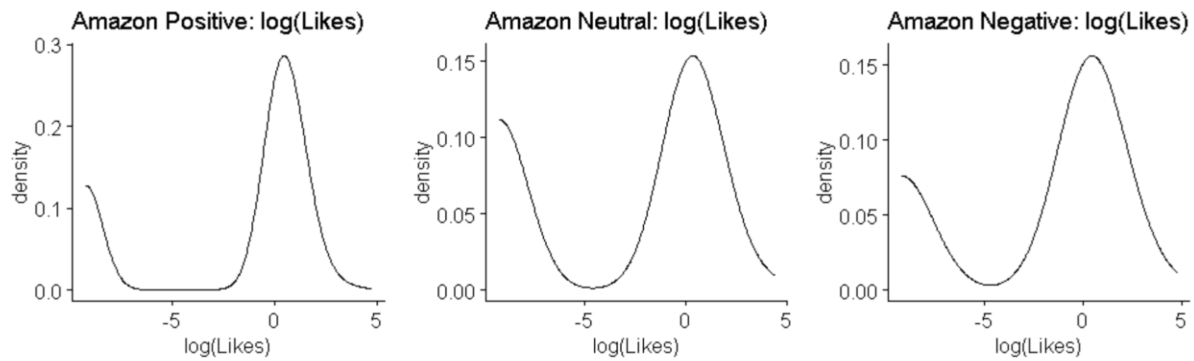
Decile differences:
Negative - Neutral
Deciles for Negative
Decile differences:
Negative - Positive
Deciles for Negative
Decile differences:
Neutral - Positive
Deciles for Neutral
Positive
Neutral
Negative
gr
obs

```
$`Negative - Neutral`
    q  Negative  Neutral difference   ci_lower   ci_upper      p_crit p_value
1 0.1  13.82269 12.44620   1.376491  -3.816095   9.257429 0.050000000   0.593
2 0.2  19.72717 16.75944   2.967731  -6.753782  11.549551 0.025000000   0.529
3 0.3  25.02204 21.75408   3.267963  -6.165738  13.973294 0.016666667   0.399
4 0.4  29.68710 24.94059   4.746505  -6.578807  17.137372 0.012500000   0.220
5 0.5  34.63257 27.84060   6.791963  -6.293503  21.807893 0.008333333   0.144
6 0.6  41.13932 31.61891   9.520412  -6.945603  27.769519 0.007142857   0.125
7 0.7  48.85568 37.17135  11.684331  -9.575180  50.442652 0.005555556   0.128
8 0.8  61.96665 45.75828  16.208375 -14.253690 243.242150 0.006250000   0.130
9 0.9 152.40616 62.59533  89.810828 -40.404766 490.891012 0.010000000   0.156

$`Negative - Positive`
    q  Negative Positive difference   ci_lower  ci_upper      p_crit p_value
1 0.1  13.82269 14.08616 -0.2634607  -4.491719  10.00190 0.010000000   0.973
2 0.2  19.72717 18.45258  1.2745862  -8.049255  12.72608 0.002500000   0.677
3 0.3  25.02204 22.87443  2.1476146  -8.358718  13.65181 0.002000000   0.544
4 0.4  29.68710 26.64423  3.0428639  -7.362852  16.91142 0.001250000   0.391
5 0.5  34.63257 30.51320  4.1193688  -7.663801  21.59011 0.001111111   0.347
6 0.6  41.13932 35.95523  5.1840828 -12.783816  30.62705 0.001428571   0.463
7 0.7  48.85568 47.95962  0.8960670 -19.582275  52.41395 0.003333333   0.845
8 0.8  61.96665 61.45869  0.5079658 -27.549416 210.10921 0.005000000   0.886
9 0.9 152.40616 94.51712 57.8890379 -74.507311 494.26576 0.001666667   0.518


$`Neutral - Positive`
    q  Neutral Positive difference   ci_lower  ci_upper       p_crit p_value
1 0.1 12.44620 14.08616  -1.639951  -8.585742  6.518637 0.0005555556   0.492
2 0.2 16.75944 18.45258  -1.693145  -8.881333  8.004133 0.0008333333   0.678
3 0.3 21.75408 22.87443  -1.120348  -9.469745  6.526190 0.0016666667   0.713
4 0.4 24.94059 26.64423  -1.703641 -13.062691  7.304281 0.0004166667   0.446
5 0.5 27.84060 30.51320  -2.672594 -10.443470  8.733305 0.0003333333   0.382
6 0.6 31.61891 35.95523  -4.336329 -19.582025 14.072361 0.0002777778   0.344
7 0.7 37.17135 47.95962 -10.788264 -28.589324 10.734744 0.0002083333   0.108
8 0.8 45.75828 61.45869 -15.700409 -49.016614 28.478383 0.0001851852   0.071
9 0.9 62.59533 94.51712 -31.921790 -98.756850 65.244582 0.0002380952   0.189
```

Considering each confidence interval contains the value 0, we may not conclude any quantile differences in tweet performance between any two groups. We fail to reject the null hypotheses that the matching quantiles of any two groups distributions of retweets are equal to one another. **Sentiment does not seem to have a statistically significant effect on the number of retweets which an Amazon official tweet receives.**

Amazon IRT: Number of Likes



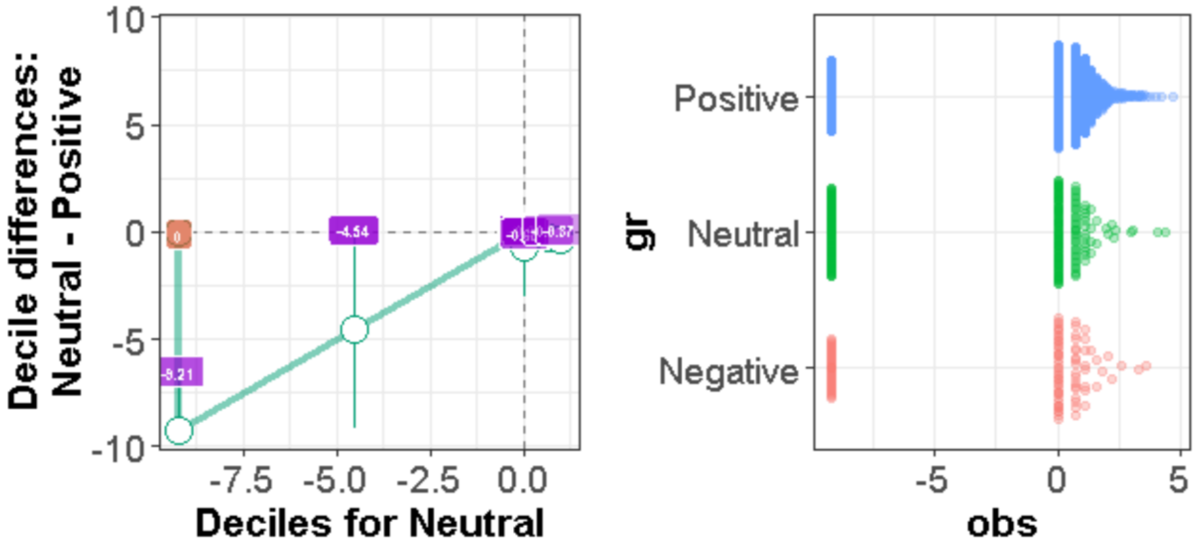



```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 33.751, df = 2, p-value = 4.688e-08
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison         Z      P.unadj        P.adj
1  Negative - Neutral  1.882929 5.970993e-02 1.791298e-01
2 Negative - Positive -1.591007 1.116079e-01 3.348238e-01
3  Neutral - Positive -5.673979 1.395184e-08 4.185551e-08
```

We fail to reject the null hypotheses for (Negative, Neutral) and (Negative, Positive) pairs. However, we may reject the null for the (Neutral, Positive) pairing and conclude that the 'like' distributions of these two populations differ. Performing a shift function to further examine these differences yields the following:



$`Negative - Neutral`

	q	Negative	Neutral	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	7.662024e-08	0.000000e+00	7.662024e-08	1.243450e-14	0.005057434	0.005555556	0.0005
2	0.2	4.926374e-03	4.655831e-12	4.926374e-03	1.746862e-07	0.519580679	0.006250000	0.0020
3	0.3	4.139798e-01	4.798837e-04	4.134999e-01	-1.307195e-01	0.994262168	0.008333333	0.0250
4	0.4	9.770327e-01	5.068428e-01	4.701899e-01	-3.451321e-01	0.976224550	0.016666667	0.1710
5	0.5	1.000090e+00	9.994083e-01	6.814183e-04	-1.197940e-02	0.121547905	0.050000000	0.3480
6	0.6	1.058044e+00	1.000000e+00	5.804362e-02	-3.129784e-05	0.848775915	0.007142857	0.0100
7	0.7	1.736251e+00	1.017019e+00	7.192319e-01	-2.904071e-01	1.075778120	0.010000000	0.0580
8	0.8	2.138561e+00	1.944508e+00	1.940533e-01	-2.531915e-01	1.360062785	0.012500000	0.1360
9	0.9	3.550398e+00	2.734536e+00	8.158622e-01	-5.937530e-01	4.499473075	0.025000000	0.2280

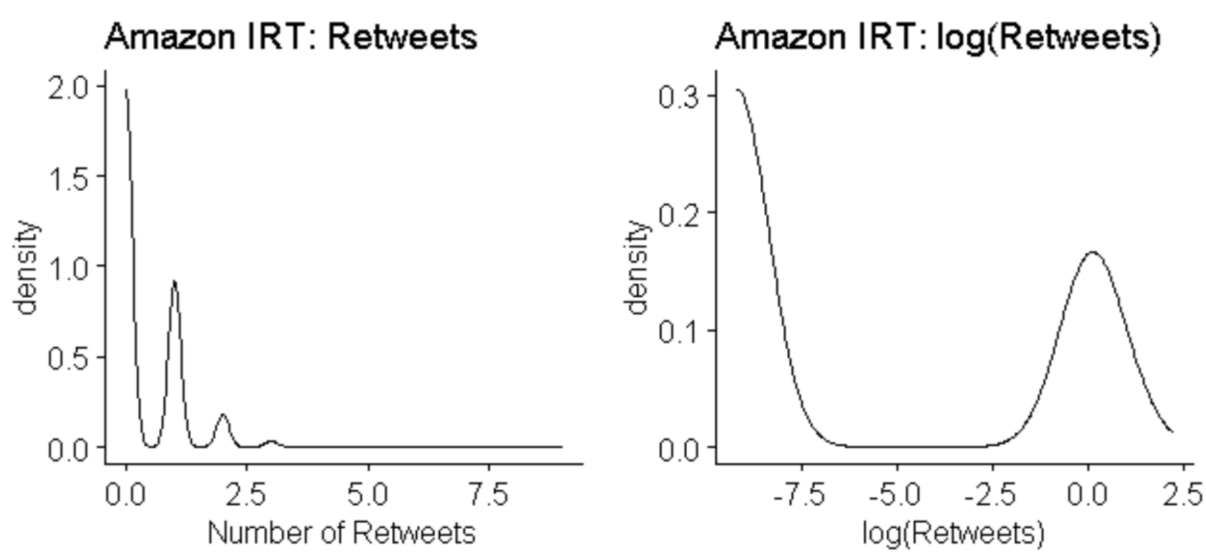
$`Negative - Positive`

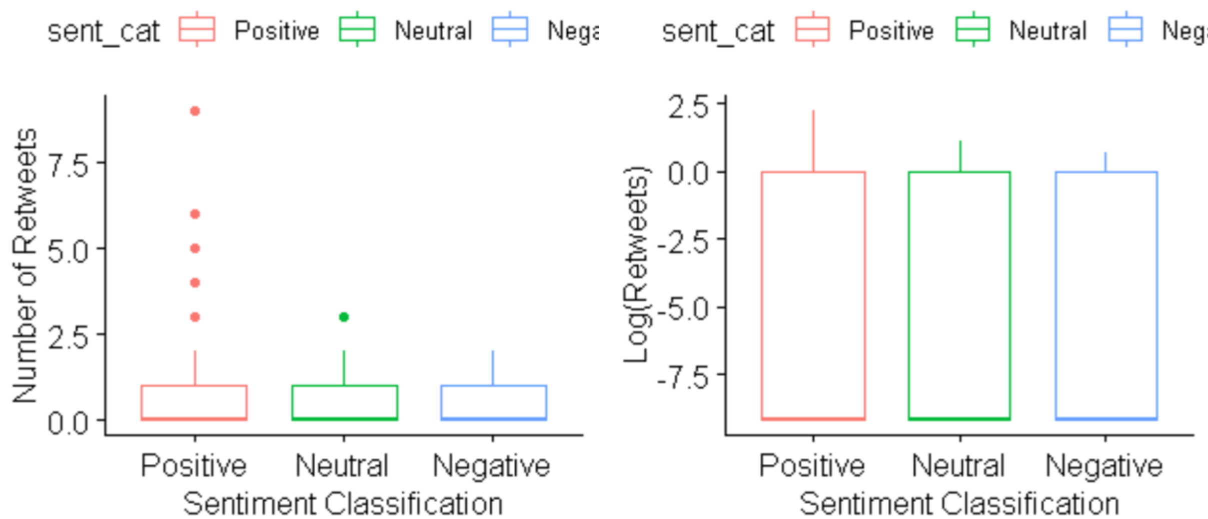
	q	Negative	Positive	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	7.662024e-08	0.000000e+00	7.662024e-08	1.243450e-14	0.005057434	0.003125000	0.0000
2	0.2	4.926374e-03	4.150014e-13	4.926374e-03	8.410128e-09	0.616599039	0.002777778	0.0010
3	0.3	4.139798e-01	9.999952e-01	-5.860154e-01	-9.986144e-01	-0.004503561	0.003571429	0.0020
4	0.4	9.770327e-01	1.000000e+00	-2.296732e-02	-7.984824e-01	0.001336000	0.004166667	0.0420
5	0.5	1.000090e+00	1.000000e+00	8.974887e-05	-3.173572e-02	0.048674424	0.025000000	0.8215
6	0.6	1.058044e+00	1.867120e+00	-8.090767e-01	-9.977220e-01	0.365817182	0.005000000	0.0550
7	0.7	1.736251e+00	2.000000e+00	-2.637493e-01	-9.789235e-01	0.116973421	0.006250000	0.0820
8	0.8	2.138561e+00	2.661437e+00	-5.228760e-01	-1.139491e+00	0.676421265	0.008333333	0.2990
9	0.9	3.550398e+00	3.901403e+00	-3.510047e-01	-1.637326e+00	3.659076325	0.012500000	0.8040

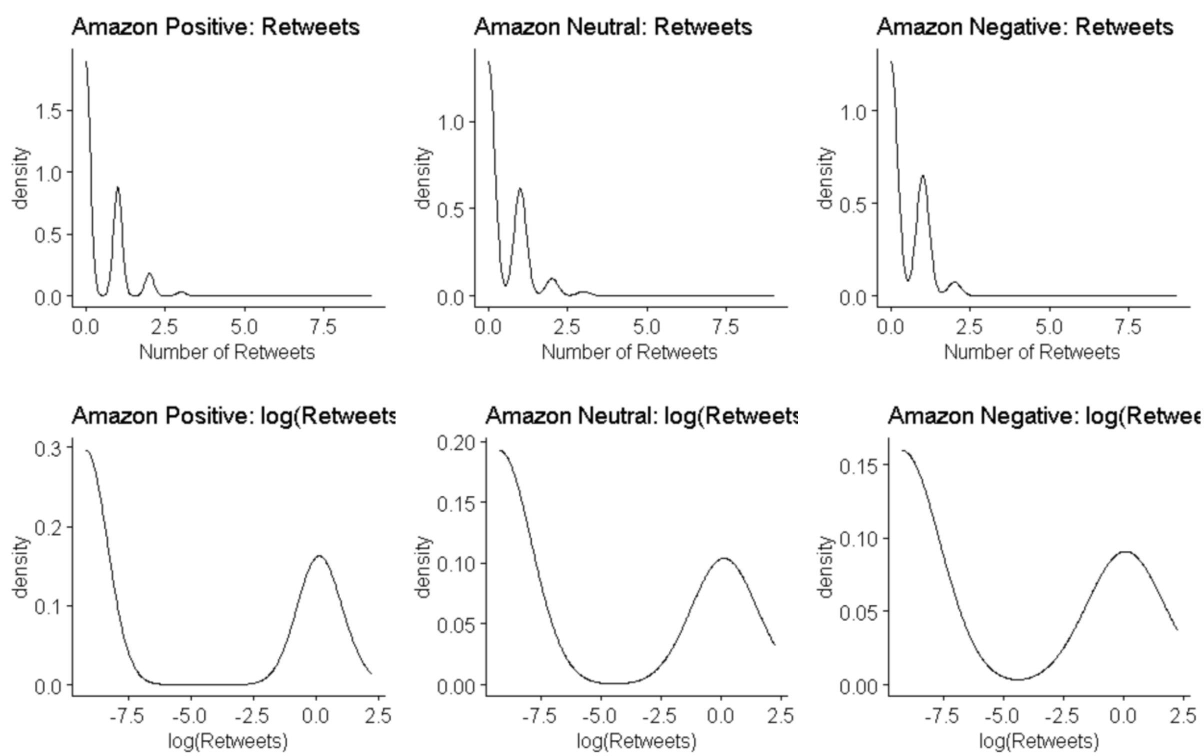
```
$`Neutral - Positive`
    q      Neutral     Positive    difference      ci_lower      ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.012500000  0.9965
2 0.2 4.655831e-12 4.150014e-13  4.240830e-12 -2.154029e-06  8.216712e-06 0.006250000  0.7915
3 0.3 4.798837e-04 9.999952e-01 -9.995153e-01 -1.000000e+00 -6.198741e-01 0.002083333  0.0000
4 0.4 5.068428e-01 1.000000e+00 -4.931572e-01 -9.994552e-01 -1.032842e-03 0.001785714  0.0000
5 0.5 9.994083e-01 1.000000e+00 -5.916694e-04 -4.749104e-01 -3.656420e-11 0.002500000  0.0015
6 0.6 1.000000e+00 1.867120e+00 -8.671203e-01 -9.999783e-01 -6.276561e-02 0.001562500  0.0000
7 0.7 1.017019e+00 2.000000e+00 -9.829812e-01 -1.000000e+00 -2.388084e-01 0.001388889  0.0000
8 0.8 1.944508e+00 2.661437e+00 -7.169294e-01 -1.759840e+00 -1.335972e-02 0.003125000  0.0020
9 0.9 2.734536e+00 3.901403e+00 -1.166867e+00 -1.948205e+00  3.078862e-01 0.004166667  0.0160
```

Only the (Neutral, Positive) pairing will be discussed, as that was the only pair which had statistically significant results for Dunn's test. We can say, with 95% confidence, that the 3rd through 9th quantiles of positive tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of neutral tweets. **Therefore, we may conclude that statistically significant differences exist between these two groups quantiles (in their distributions of likes), and potentially one underlying factor explaining these differences is the sentiment of Amazon IRT tweets (positive or neutral).**

Amazon IRT: Number of Retweets



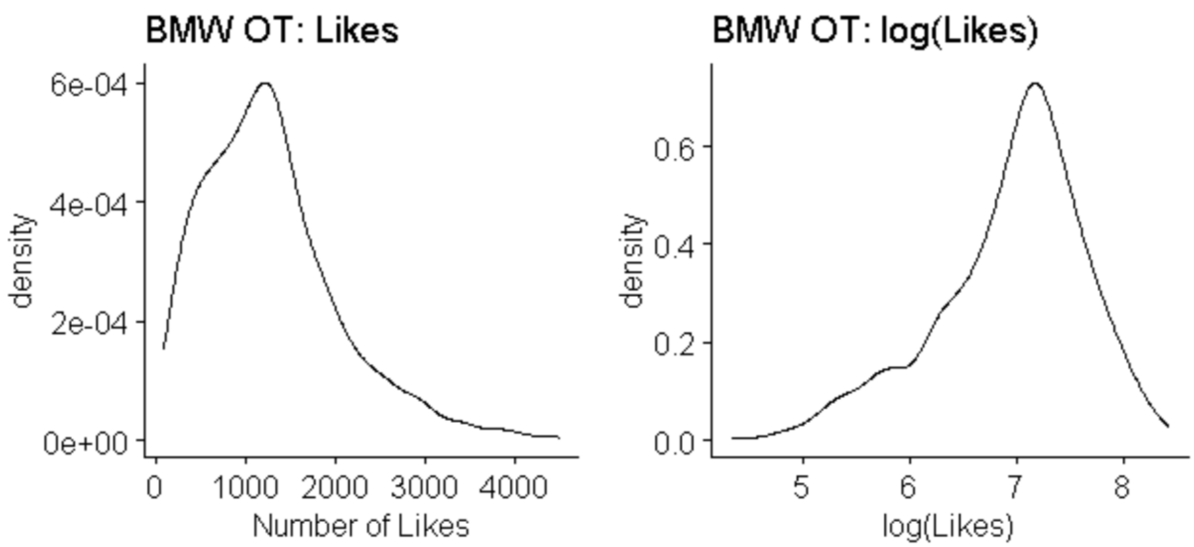




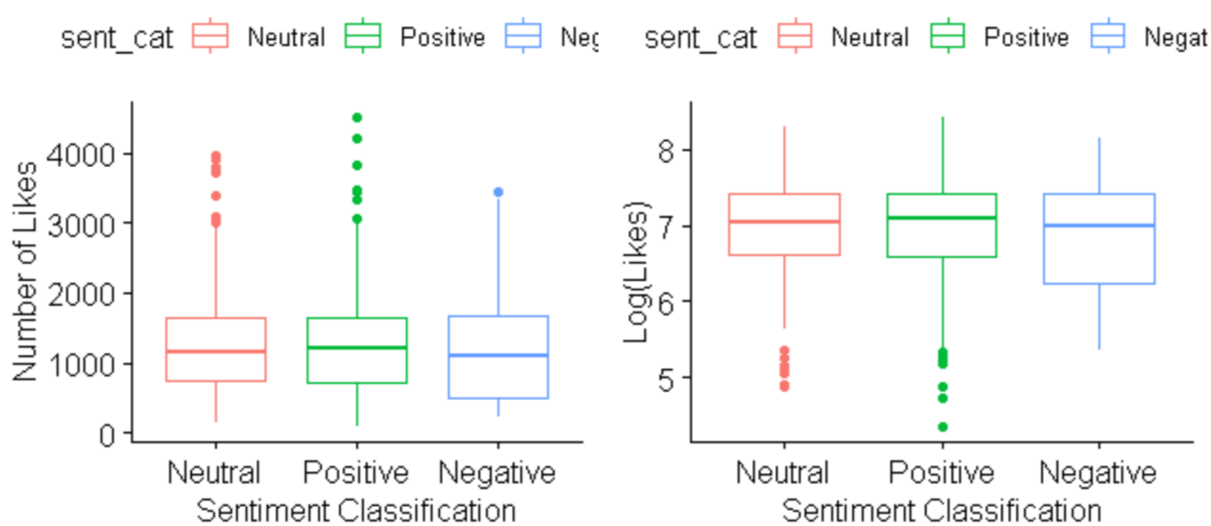

```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 0.41067, df = 2, p-value = 0.8144
```

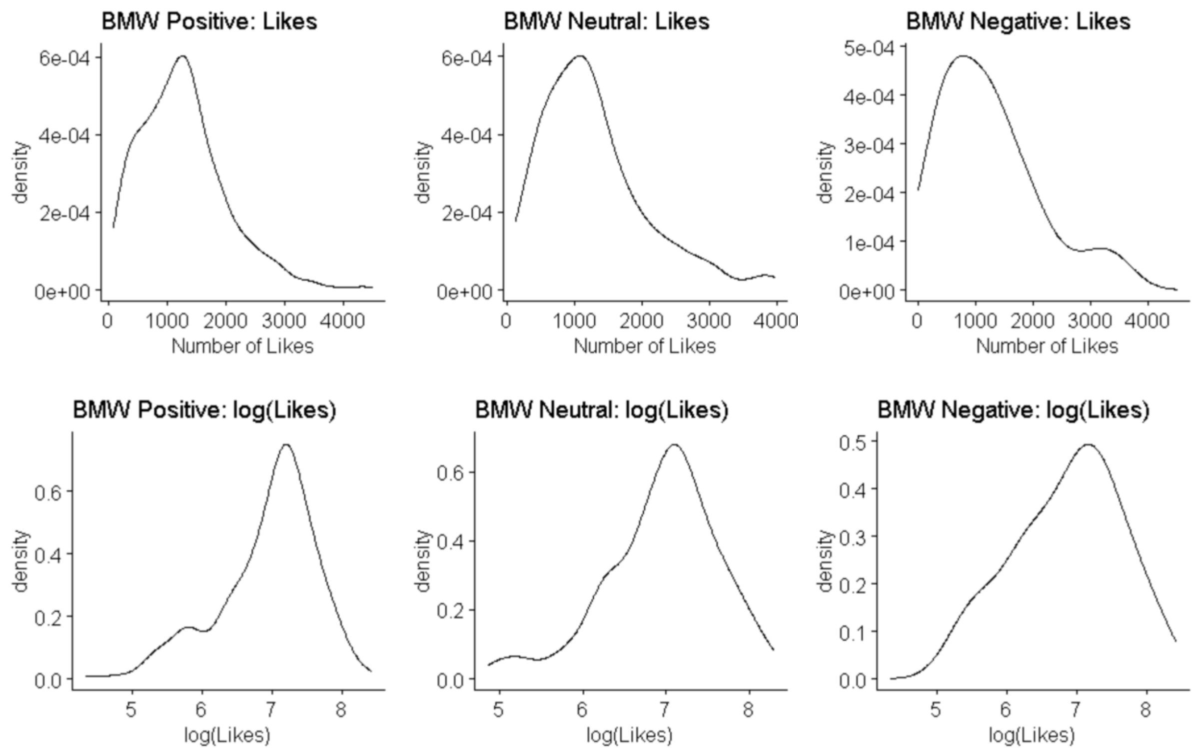
Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘retweet’ distributions of all populations are equal. Considering this implies that all further analysis is purely out of curiosity, and any statistically significant results would be ultimately disregarded, in an effort to save on time and space I will discontinue analysis at this point.

BMW Official: Number of Likes



The log distribution does not pass a Shapiro-Wilk normality test.



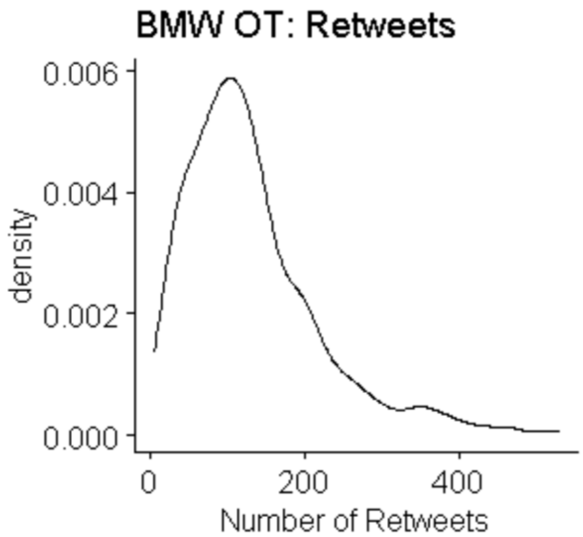


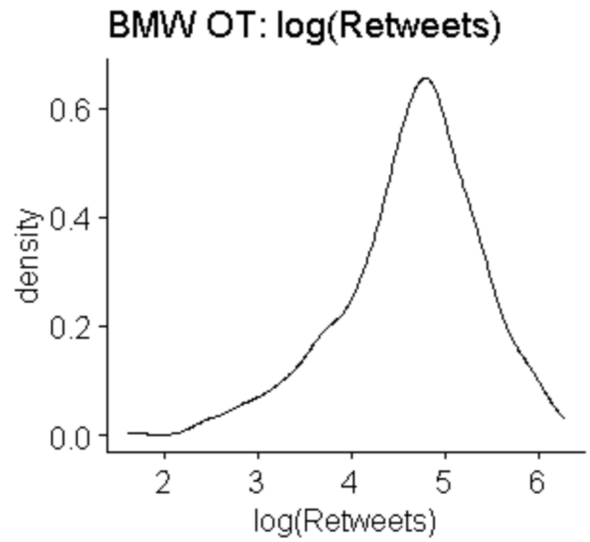
Only the log distribution on the above right passes a Shapiro-Wilk normality test.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 0.53815, df = 2, p-value = 0.7641
```

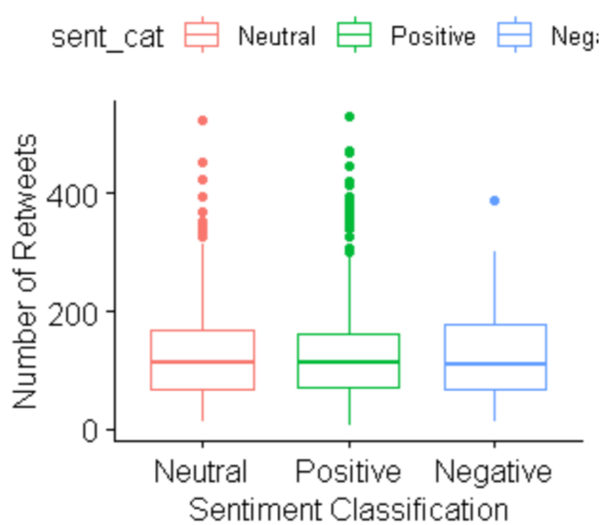
Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘like’ distributions of all populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of likes which a BMW official tweet receives.**

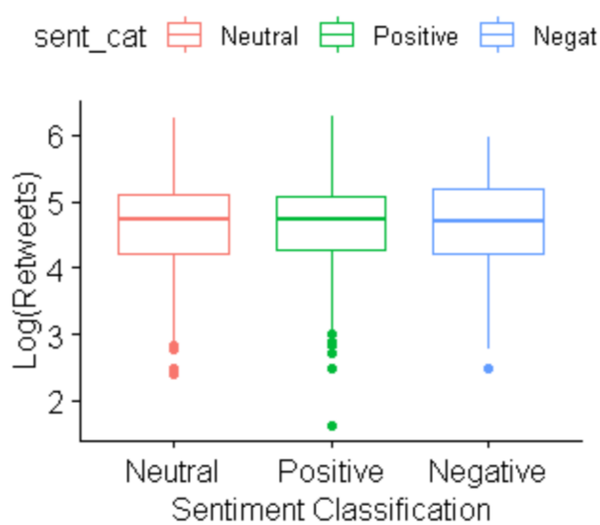
BMW Official: Number of Retweets

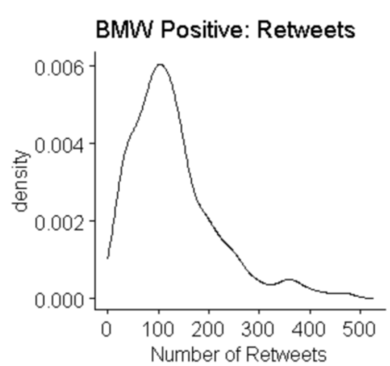


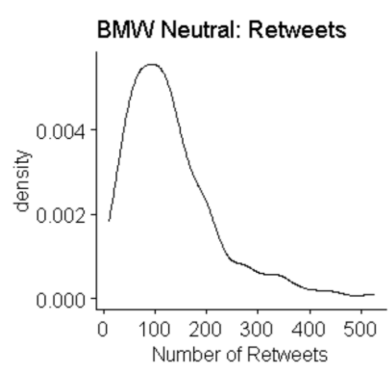


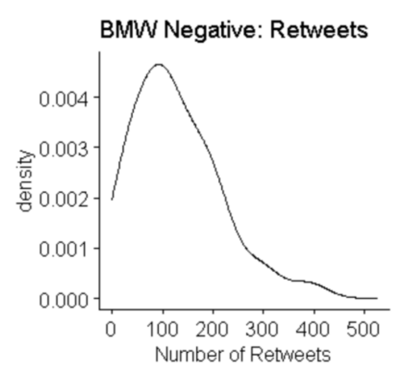
The log distribution does not pass a Shapiro-Wilk normality test.

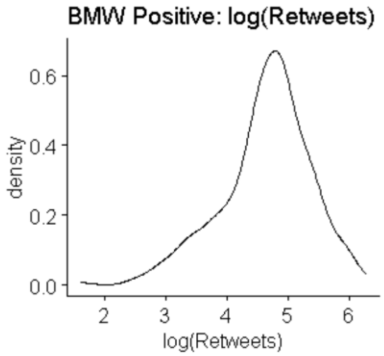


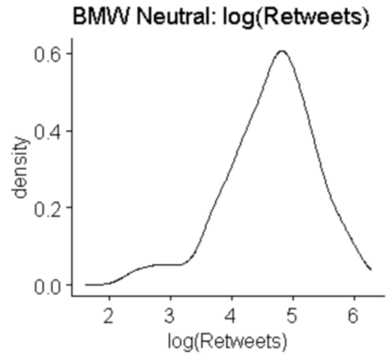


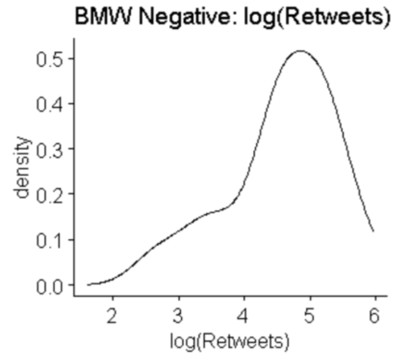










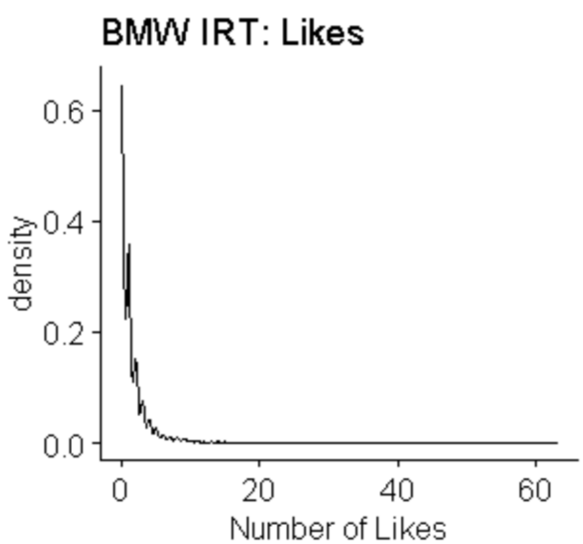


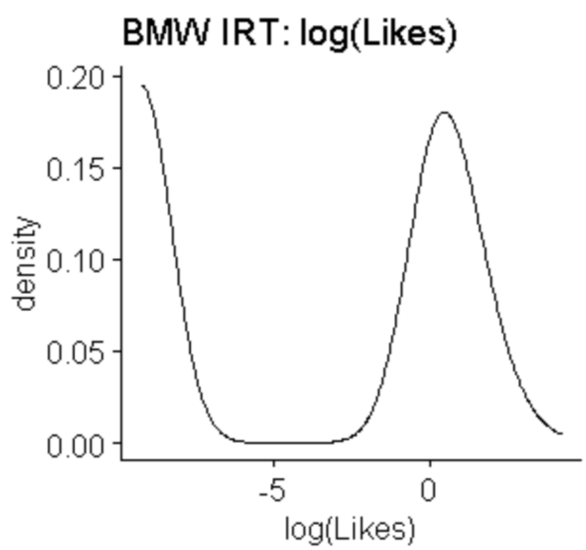
Again, only the above right log distribution is able to pass a Shapiro-Wilk normality test.

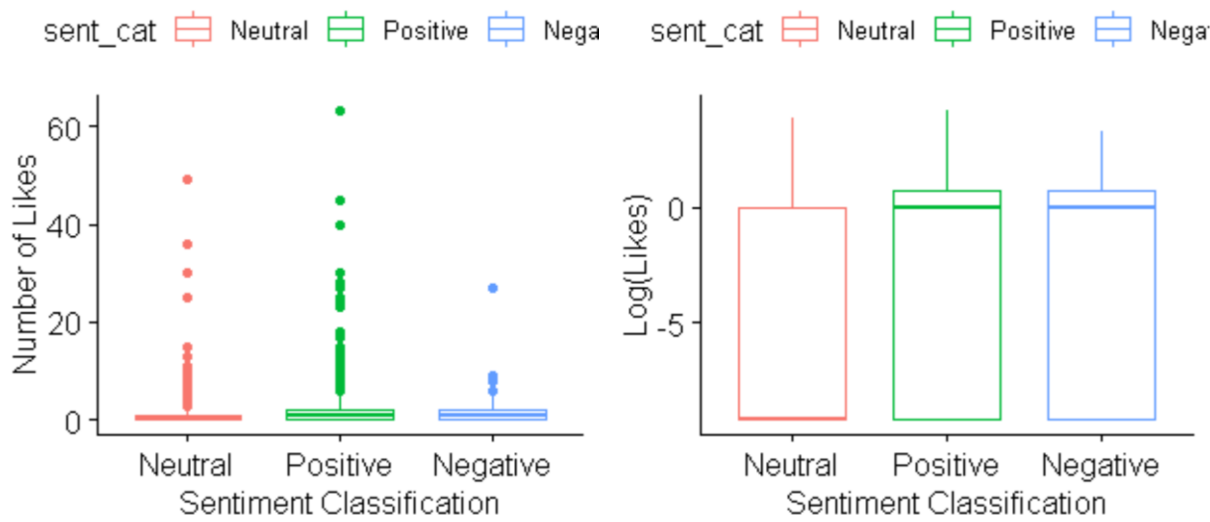
```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 0.021324, df = 2, p-value = 0.9894
```

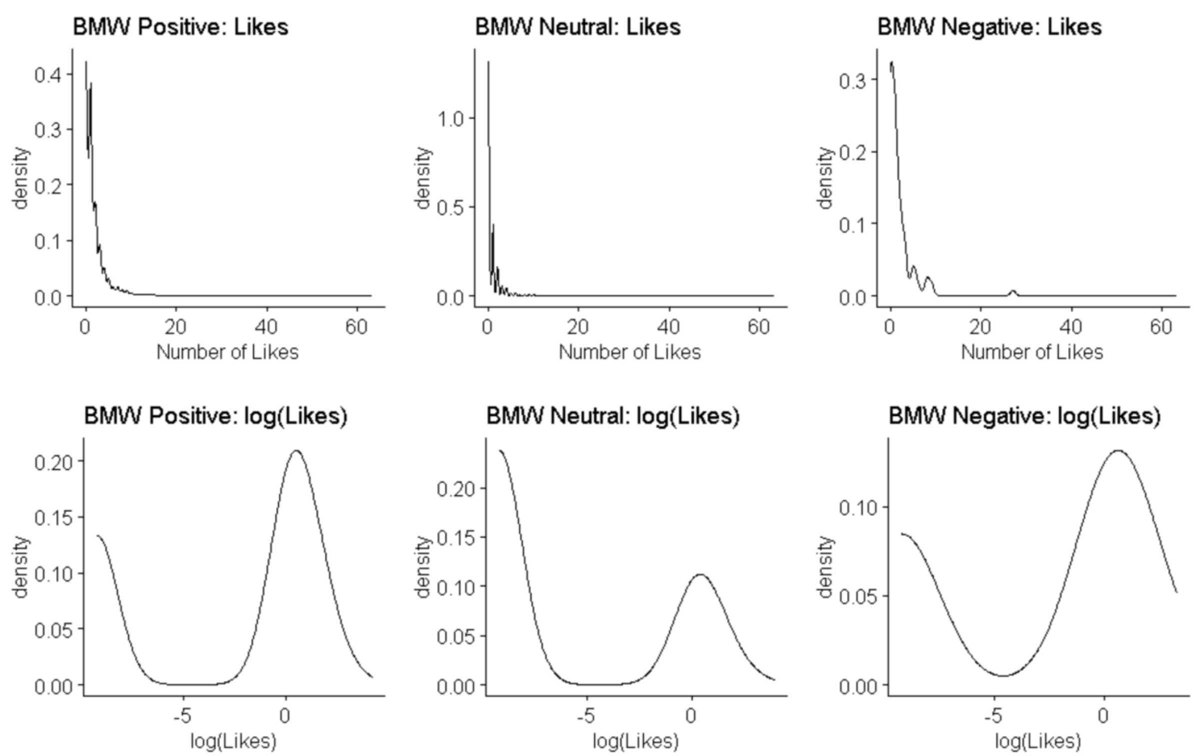
Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘retweet’ distributions of all populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a BMW official tweet receives.**

BMW IRT: Number of Likes







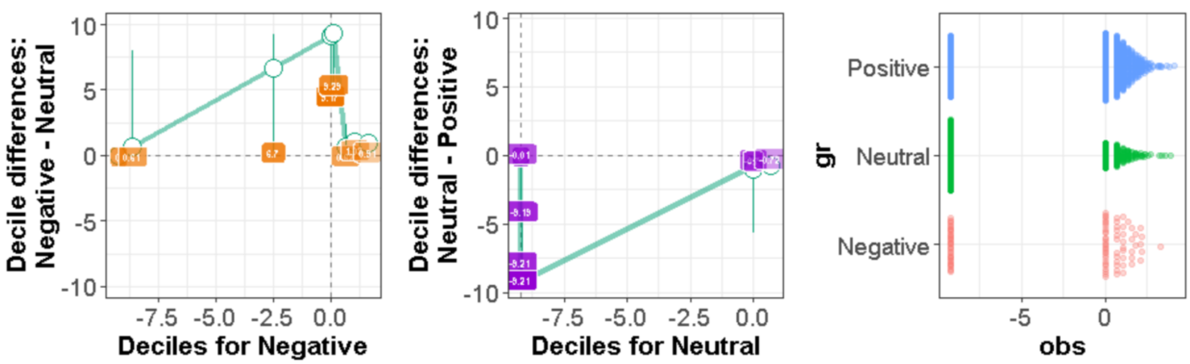



```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 192.53, df = 2, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison           Z       P.unadj         P.adj
1  Negative - Neutral   5.4925319 3.962120e-08 1.188636e-07
2 Negative - Positive  -0.2100748 8.336093e-01 1.000000e+00
3  Neutral - Positive -13.6936209 1.108450e-42 3.325350e-42
```

From the results of Dunn's test, we can see that we may reject the null for both the (Negative, Neutral) and the (Positive, Neutral) pairs. In other words, the distribution of likes for neutral tweets differs from the distributions of likes for other sentiment categories. Performing a shift function to further examine these differences yields the following:



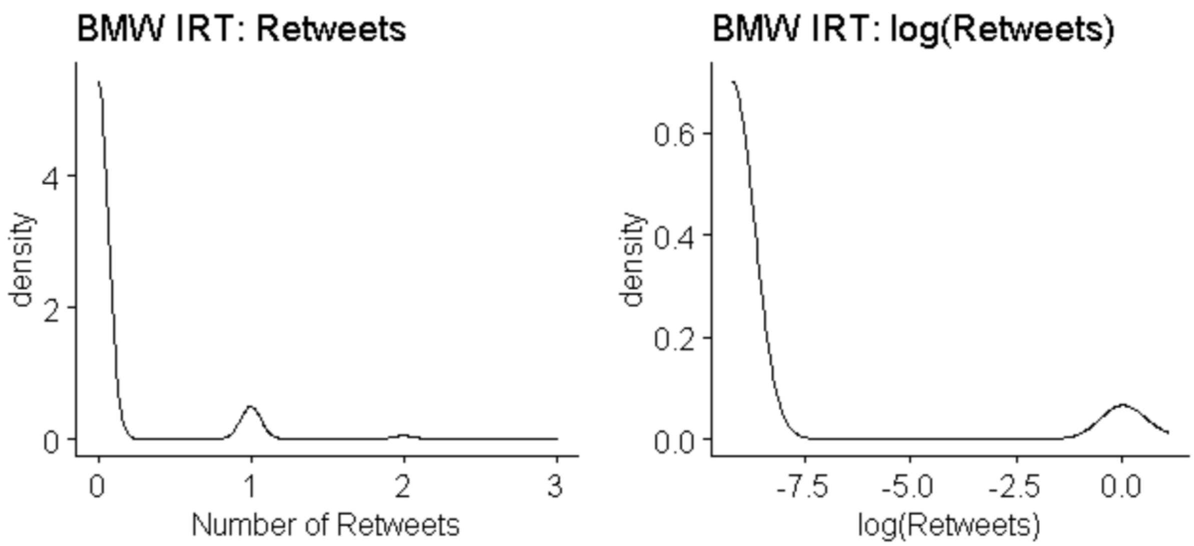

```
$`Negative - Neutral`
    q     Negative     Neutral   difference     ci_lower     ci_upper      p_crit p_value
1 0.1 1.933018e-10 0.000000000 1.933018e-10 2.220446e-16 3.163852e-06 0.050000000   0.018
2 0.2 8.964986e-05 0.000000000 8.964986e-05 2.007605e-09 4.650844e-02 0.025000000   0.000
3 0.3 6.622539e-02 0.000000000 6.622539e-02 4.862377e-05 8.081446e-01 0.016666667   0.000
4 0.4 7.277781e-01 0.000000000 7.277781e-01 3.396441e-02 1.000539e+00 0.012500000   0.000
5 0.5 9.970419e-01 0.000000000 9.970419e-01 5.794633e-01 1.273573e+00 0.010000000   0.000
6 0.6 1.152293e+00 0.002549085 1.149744e+00 7.254657e-01 2.012934e+00 0.008333333   0.000
7 0.7 1.919015e+00 0.999735441 9.192798e-01 3.396840e-02 1.922445e+00 0.007142857   0.000
8 0.8 2.864391e+00 1.000699641 1.863691e+00 6.995465e-01 3.688787e+00 0.006250000   0.000
9 0.9 5.158821e+00 2.022573769 3.136247e+00 7.398220e-01 5.922840e+00 0.005555556   0.000
```

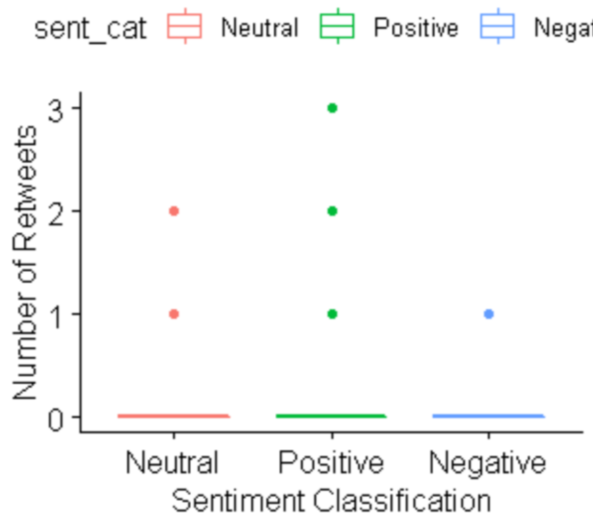
We can say, with 95% confidence, that every quantile of the set of neutral tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of negative tweets.

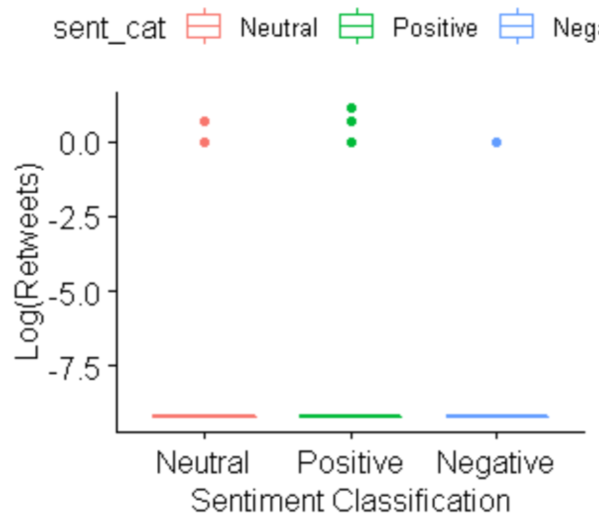
```
$`Neutral - Positive`
    q     Neutral     Positive    difference      ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000000 0.0000000000  0.0000000000  0.000000e+00  0.000000e+00 0.0013888889  1.0000
2 0.2 0.000000000 0.0000000000  0.0000000000 -1.465494e-14  0.000000e+00 0.0006944444  0.9975
3 0.3 0.000000000 0.0007951944 -0.0007951944 -6.163310e-01 -1.035481e-10 0.0004629630  0.0000
4 0.4 0.000000000 0.9999898862 -0.9999898862 -1.000000e+00 -8.542046e-01 0.0003472222  0.0000
5 0.5 0.000000000 1.0000000000 -1.0000000000 -1.000000e+00 -1.000000e+00 0.0002777778  0.0000
6 0.6 0.002549085 1.0000374298 -0.9974883448 -1.159081e+00 -4.546415e-01 0.0002314815  0.0000
7 0.7 0.999735441 1.9997227863 -0.9999873450 -1.534031e+00 -6.800172e-01 0.0001984127  0.0000
8 0.8 1.000699641 2.7500624032 -1.7493627621 -1.999836e+00 -9.005759e-01 0.0001736111  0.0000
9 0.9 2.022573769 4.1648829275 -2.1423091585 -3.206514e+00 -1.208613e+00 0.0001543210  0.0000
```

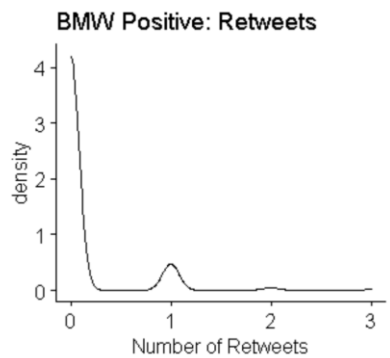
We can say, with 95% confidence, that the 3rd through 9th quantiles of the set of positive tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of neutral tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes), and potentially one underlying factor explaining these differences is the sentiment of BMW IRT tweets (neutral sentiment behaving differently).**

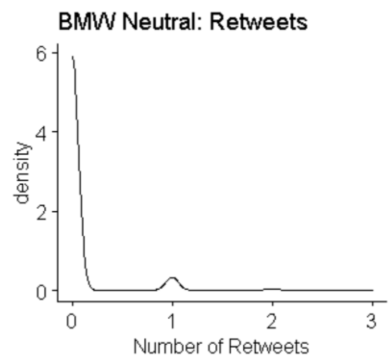
BMW IRT: Number of Retweets

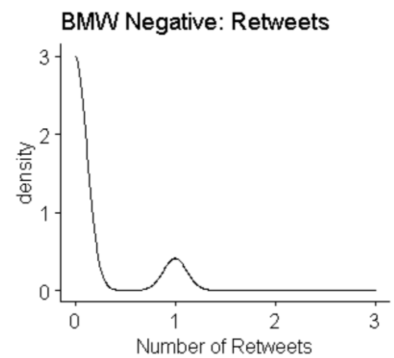


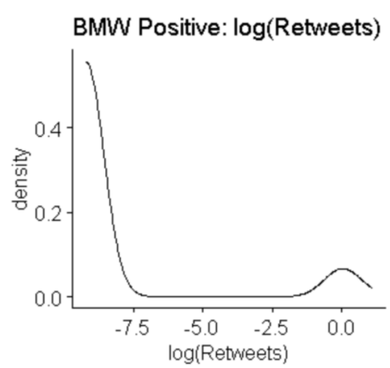


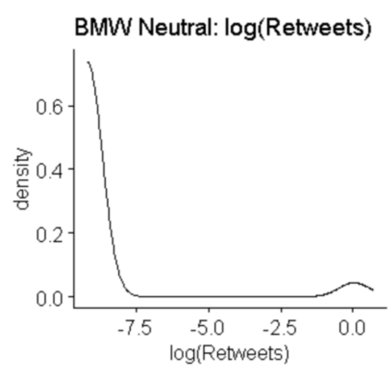


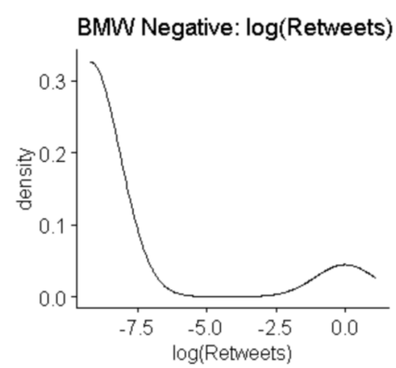










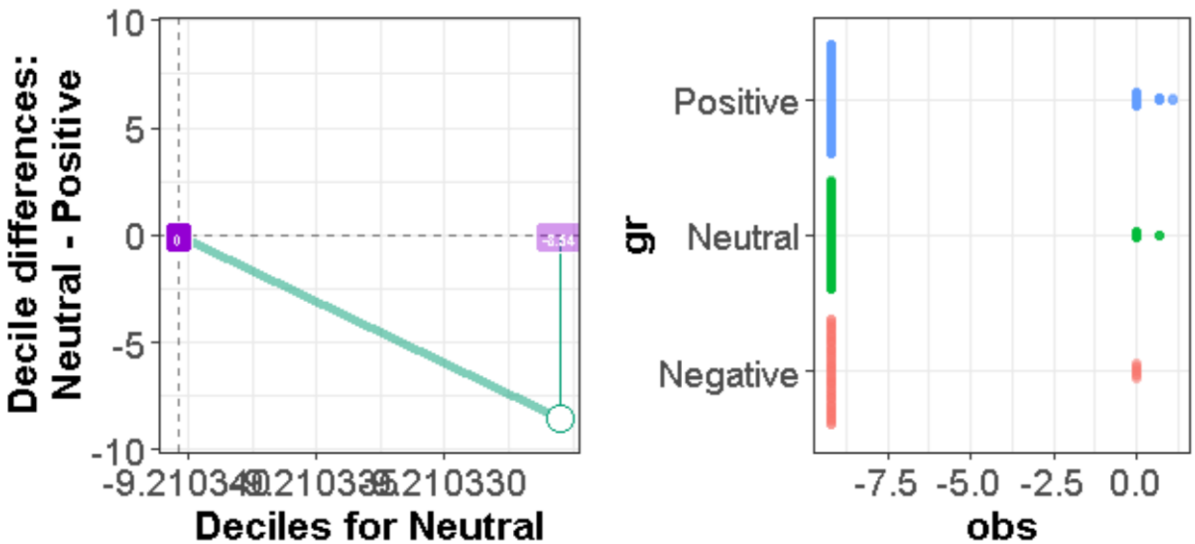



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 18.524, df = 2, p-value = 9.498e-05
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z      P.unadj        P.adj
1  Negative - Neutral  1.9468716 5.155013e-02 1.546504e-01
2 Negative - Positive  0.2033718 8.388445e-01 1.000000e+00
3  Neutral - Positive -4.1959039 2.717855e-05 8.153565e-05
```

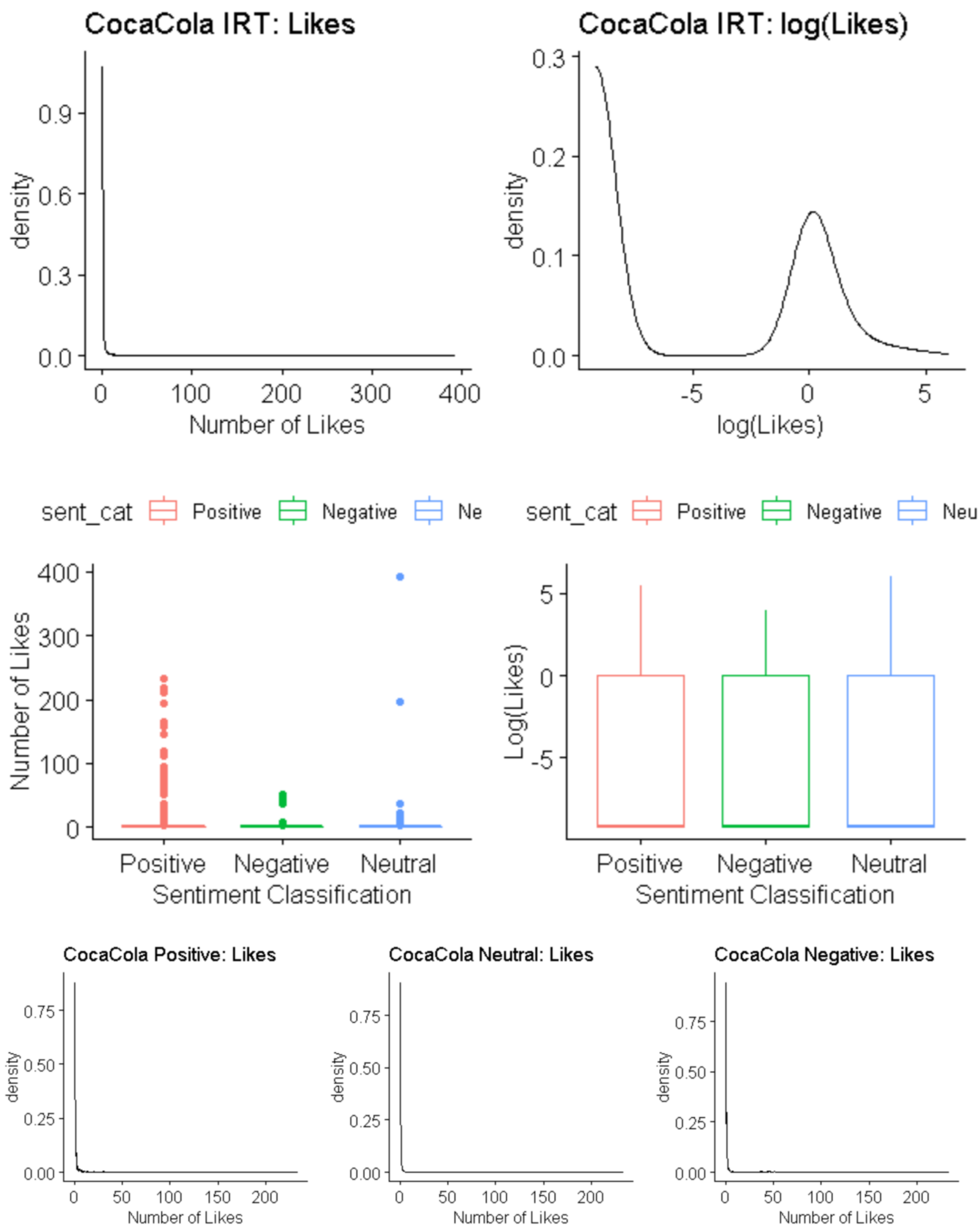
We may reject the null only for the (Neutral, Positive) pairing and conclude that the distribution of likes differs between these two groups. Performing a shift function to further analyze these differences yields the following:

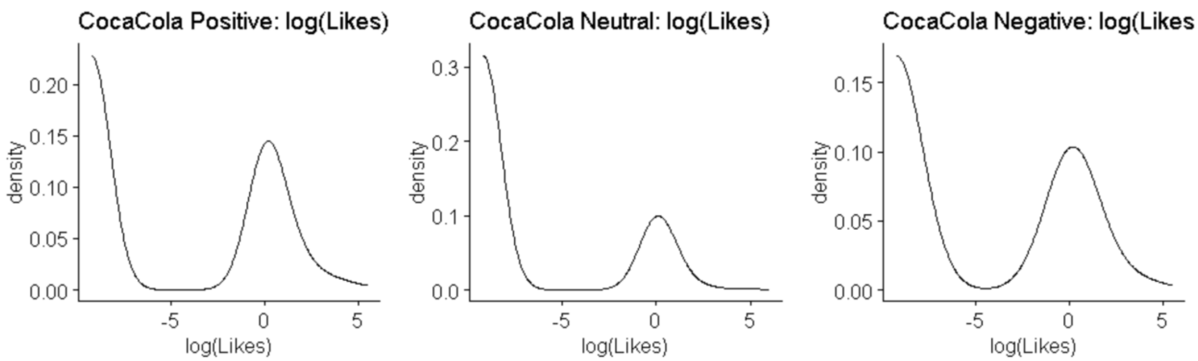



```
$`Neutral - Positive`
    q      Neutral  Positive difference      ci_lower    ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.0009259259   1.000
2 0.2 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.0004629630   1.000
3 0.3 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.0003086420   1.000
4 0.4 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.0002314815   1.000
5 0.5 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.0001851852   1.000
6 0.6 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.0001543210   1.000
7 0.7 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.00000000 0.0001322751   1.000
8 0.8 0.000000e+00 0.0000000  0.0000000 -4.397829e-08  0.00000000 0.0001157407   0.711
9 0.9 1.617224e-06 0.9275851 -0.9275835 -9.999982e-01 -0.05689788 0.0001028807   0.000
```

Given the images above, and the fact that differences are significant only at the 9th quantile, I’m going to consider these results erroneous. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a BMW IRT tweet receives.**

Coca Cola IRT: Number of Likes



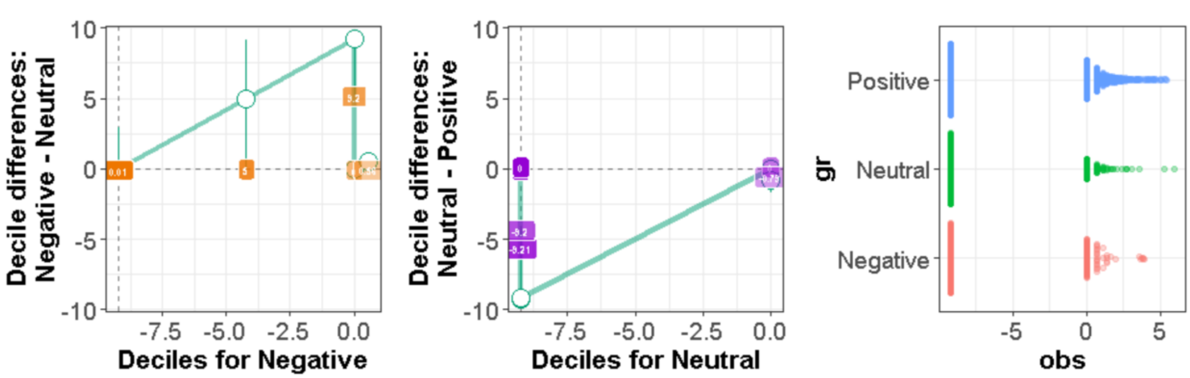



```
Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 110.05, df = 2, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z       P.unadj        P.adj
1  Negative - Neutral   4.445463 8.770269e-06 2.631081e-05
2 Negative - Positive  -1.567465 1.170061e-01 3.510184e-01
3  Neutral - Positive -10.472261 1.158430e-25 3.475291e-25
```

From the results of Dunn’s test, we can see that we may reject the null for both the (Negative, Neutral) and the (Positive, Neutral) pairs. In other words, the distribution of likes for neutral tweets differs from the distributions of likes for other sentiment categories. Performing a shift function to further examine these differences yields the following:



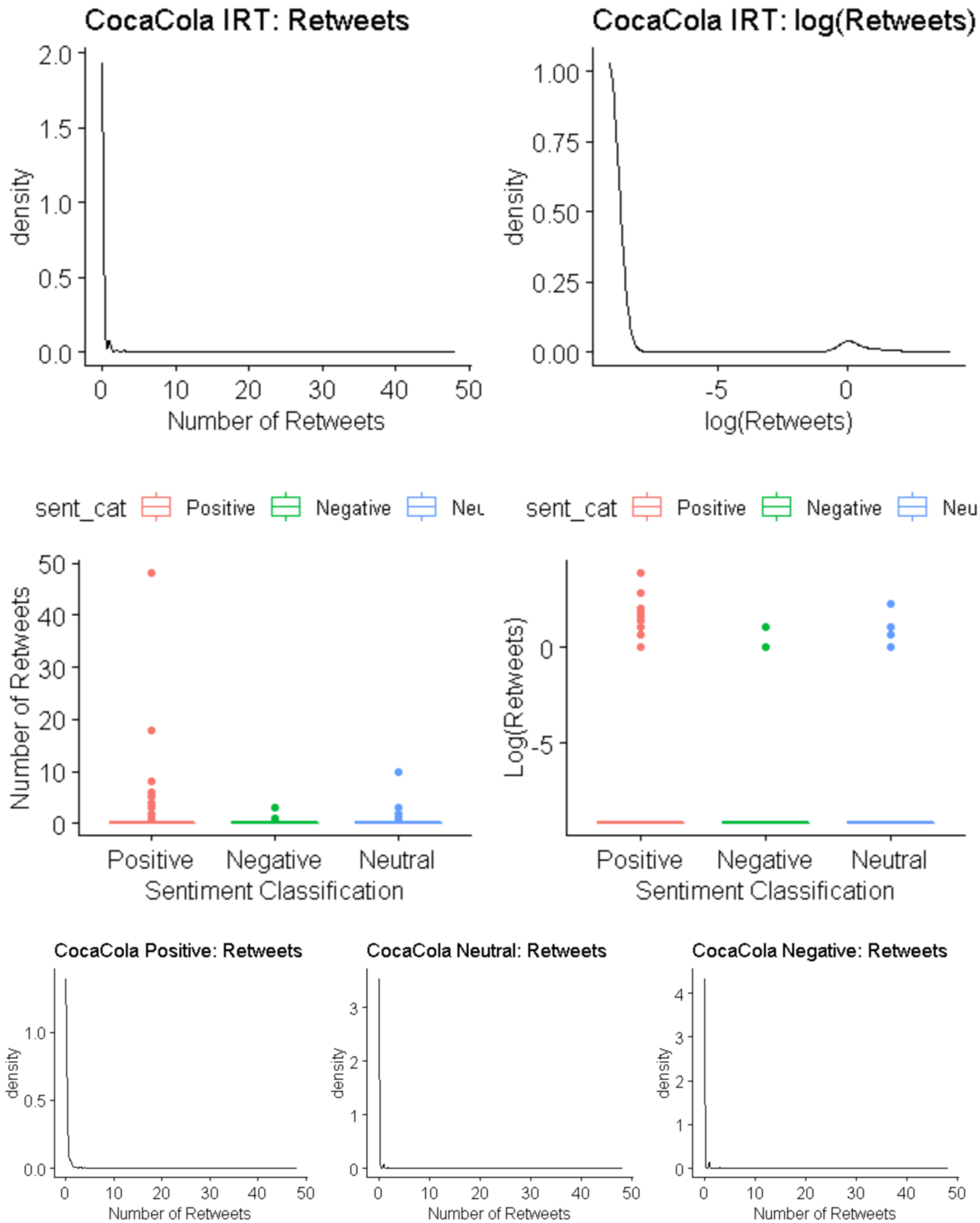
```
$`Negative - Neutral`
    q     Negative     Neutral   difference     ci_lower     ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000000 0.000000e+00 0.000000e+00 0.000000e+00 0.050000000  1.0000
2 0.2 0.000000e+00 0.000000000 0.000000e+00 0.000000e+00 0.000000e+00 0.025000000  1.0000
3 0.3 0.000000e+00 0.000000000 0.000000e+00 0.000000e+00 7.371881e-14 0.016666667  0.9305
4 0.4 3.856004e-11 0.000000000 3.856004e-11 0.000000e+00 4.682888e-05 0.012500000  0.0405
5 0.5 6.638375e-04 0.000000000 6.638375e-04 1.451423e-09 2.703912e-01 0.008333333  0.0000
6 0.6 5.430269e-01 0.000000000 5.430269e-01 4.883716e-03 9.975166e-01 0.007142857  0.0000
7 0.7 9.997853e-01 0.001338434 9.984468e-01 5.942224e-01 9.999997e-01 0.006250000  0.0000
8 0.8 1.000041e+00 0.999993006 4.807923e-05 1.184650e-09 8.964286e-02 0.010000000  0.0080
9 0.9 1.811517e+00 1.000000000 8.115173e-01 2.509126e-02 1.650075e+00 0.005555556  0.0000
```

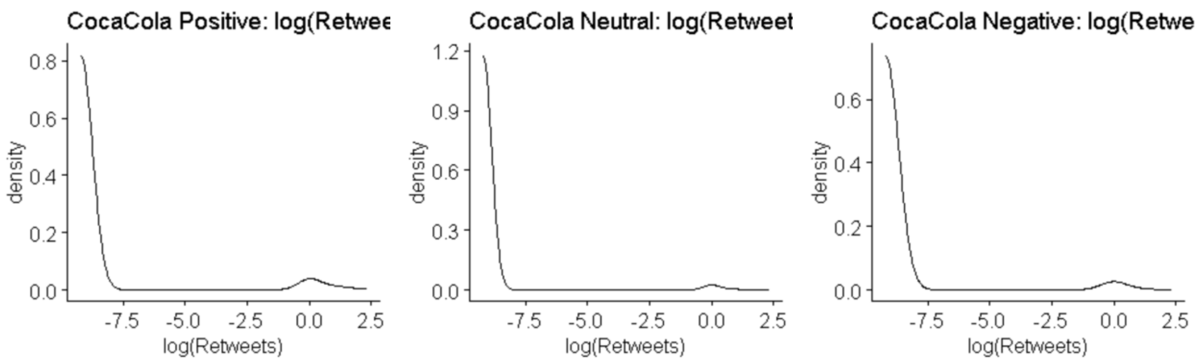
We can say, with 95% confidence, that the 5th through 9th quantiles of the set of neutral tweets would have to be shifted up by significant (non-zero) amounts to match their counterparts in the set of negative tweets.

```
$`Neutral - Positive`
    q     Neutral     Positive    difference    ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000000 0.000000e+00  0.000000e+00  0.00000000  0.000000e+00 0.0009259259   1.000
2 0.2 0.000000000 0.000000e+00  0.000000e+00  0.00000000  0.000000e+00 0.0004629630   1.000
3 0.3 0.000000000 0.000000e+00  0.000000e+00  0.00000000  0.000000e+00 0.0003086420   1.000
4 0.4 0.000000000 0.000000e+00  0.000000e+00  0.00000000  0.000000e+00 0.0002314815   1.000
5 0.5 0.000000000 7.362439e-08 -7.362439e-08 -0.01214702  0.000000e+00 0.0001851852   0.001
6 0.6 0.000000000 9.997799e-01 -9.997799e-01 -1.00000000 -7.074969e-01 0.0001543210   0.000
7 0.7 0.001338434 1.000000e+00 -9.986616e-01 -1.00000000 -4.044171e-01 0.0001322751   0.000
8 0.8 0.999993006 1.000001e+00 -8.402355e-06 -0.14286762 -1.350038e-10 0.0001157407   0.000
9 0.9 1.000000000 2.234578e+00 -1.234578e+00 -2.51528593 -1.000054e+00 0.0001028807   0.000
```

We can say, with 95% confidence, that the 6th through 9th quantiles of the set of positive tweets would have to be shifted down by significant (non-zero) amounts to match their counterparts in the set of neutral tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes), and potentially one underlying factor explaining these differences is the sentiment of Coca Cola IRT tweets (neutral sentiment behaving differently).** It seems that the set of neutral tweets has a higher proportion of tweets receiving 0 likes than other sentiment categories.

Coca Cola IRT: Number of Retweets



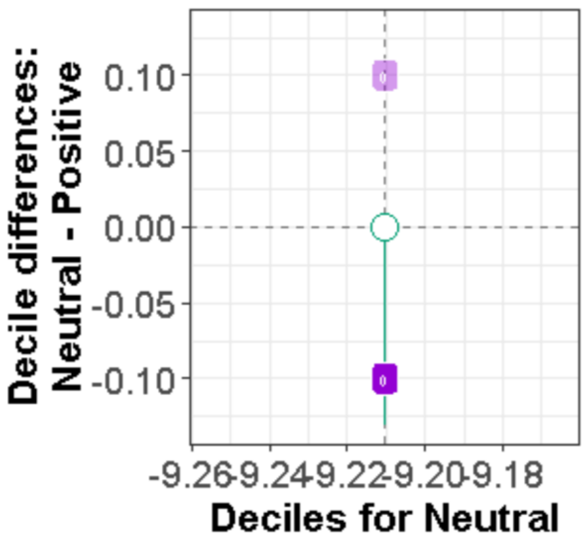


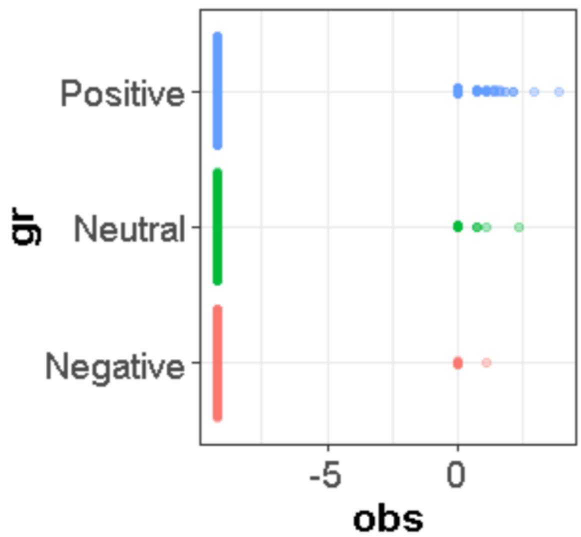

```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 21.317, df = 2, p-value = 2.35e-05
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison         Z      P.unadj        P.adj
1  Negative - Neutral  0.841724 3.999425e-01 1.000000e+00
2 Negative - Positive -1.815923 6.938225e-02 2.081468e-01
3  Neutral - Positive -4.529400 5.915131e-06 1.774539e-05
```

We may reject the null only for the (Neutral, Positive) pairing and conclude that the distribution of likes differs between these two groups. Performing a shift function to further analyze these differences yields the following:

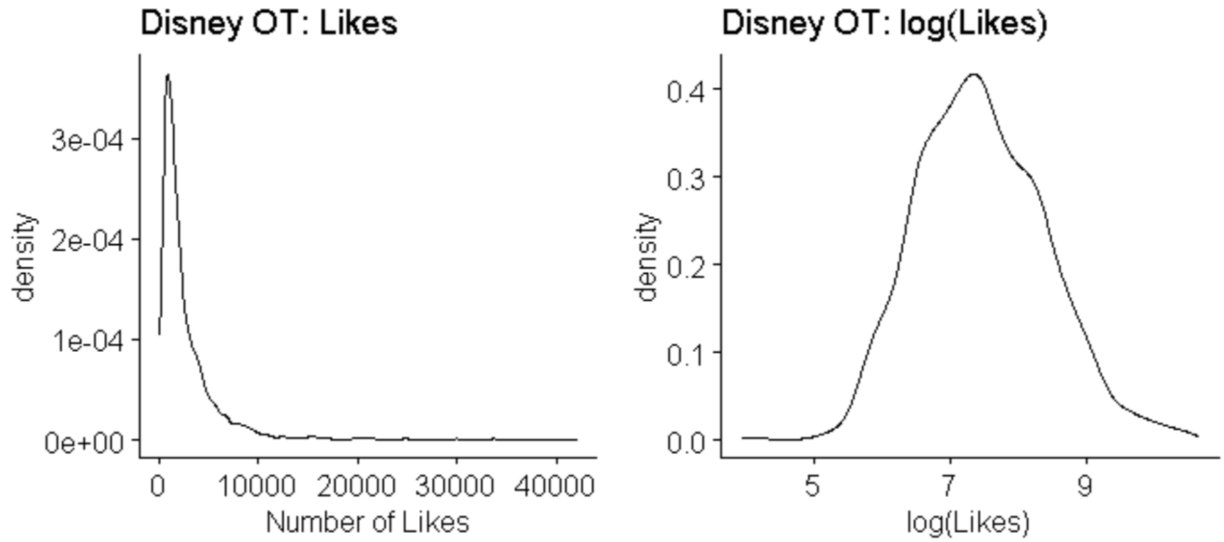




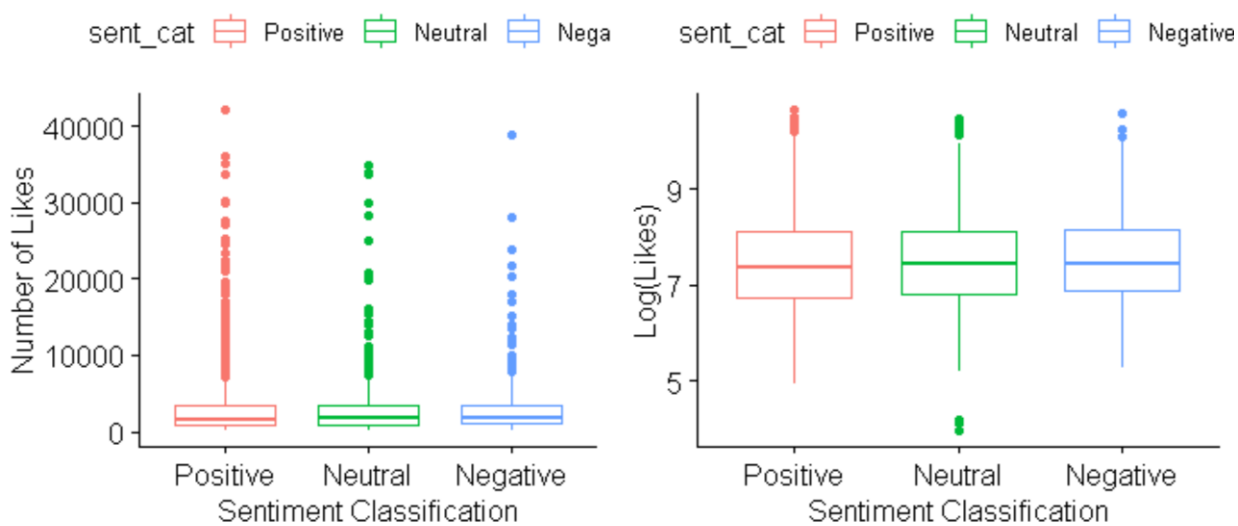

```
$`Neutral - Positive`
    q Neutral      Positive    difference     ci_lower ci_upper        p_crit p_value
1 0.1       0 0.000000e+00  0.000000e+00  0.000000000        0 6.172840e-04    1.00
2 0.2       0 0.000000e+00  0.000000e+00  0.000000000        0 3.086420e-04    1.00
3 0.3       0 0.000000e+00  0.000000e+00  0.000000000        0 2.057613e-04    1.00
4 0.4       0 0.000000e+00  0.000000e+00  0.000000000        0 1.543210e-04    1.00
5 0.5       0 0.000000e+00  0.000000e+00  0.000000000        0 1.234568e-04    1.00
6 0.6       0 0.000000e+00  0.000000e+00  0.000000000        0 1.028807e-04    1.00
7 0.7       0 0.000000e+00  0.000000e+00  0.000000000        0 8.818342e-05    1.00
8 0.8       0 0.000000e+00  0.000000e+00  0.000000000        0 7.716049e-05    1.00
9 0.9       0 1.819138e-10 -1.819138e-10 -0.001519907        0 6.858711e-05    0.03
```

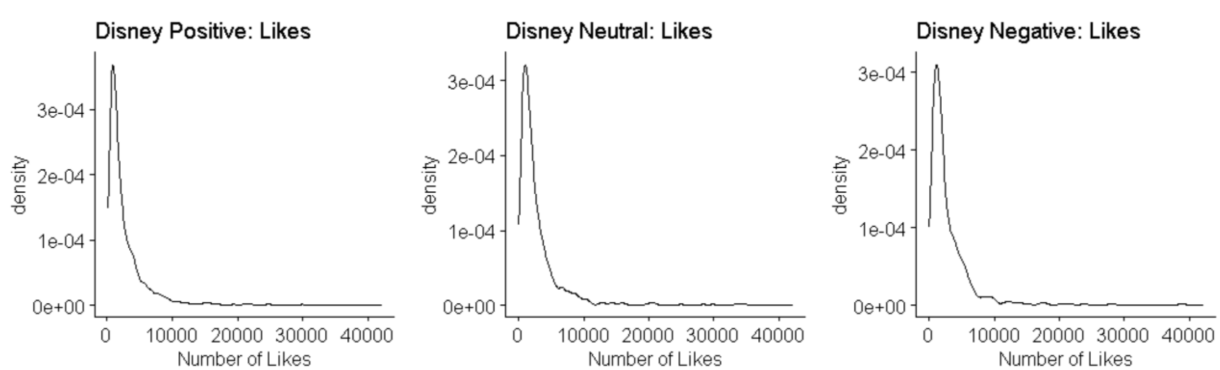
We can say, with 95% confidence, that the 9th quantile of the set of positive tweets would need to be shifted downy by a significant (non-zero) amount to match its counterpart in the set of neutral tweets. However, given the small size of the downshift, as well as the visualizations produced above, I consider these results to be dubious. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a Coca Cola IRT tweet receives.**

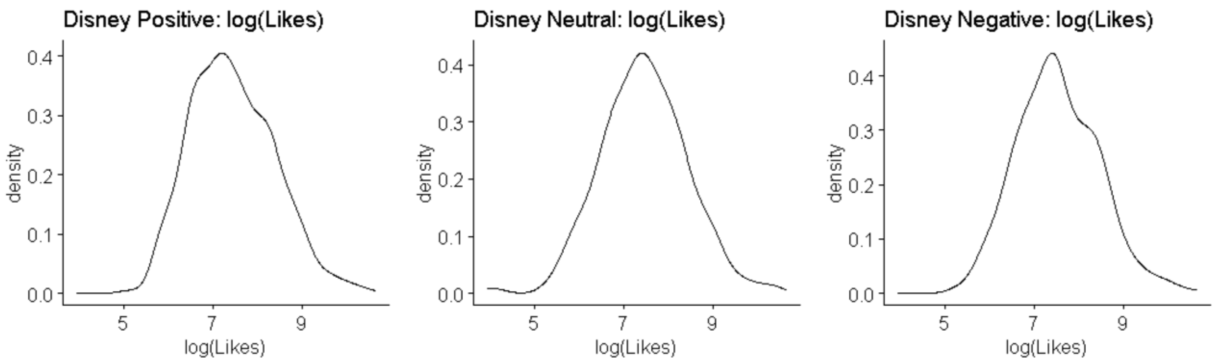
Disney Official: Number of Likes



The log distribution does not pass a Shapiro-Wilk normality test.





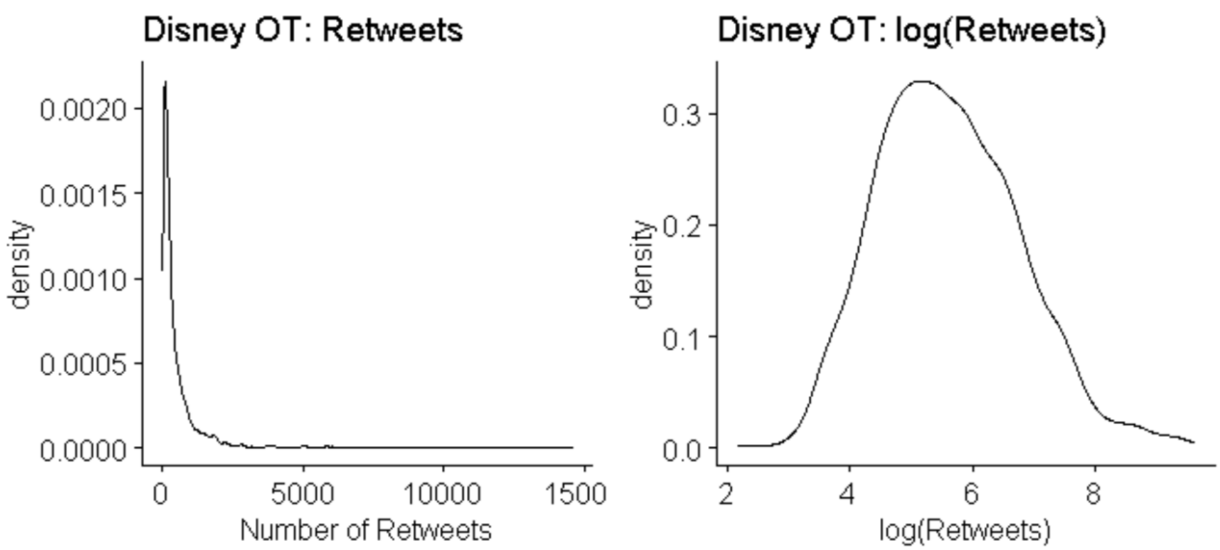


None of the above log distributions pass a Shapiro-Wilk normality test.

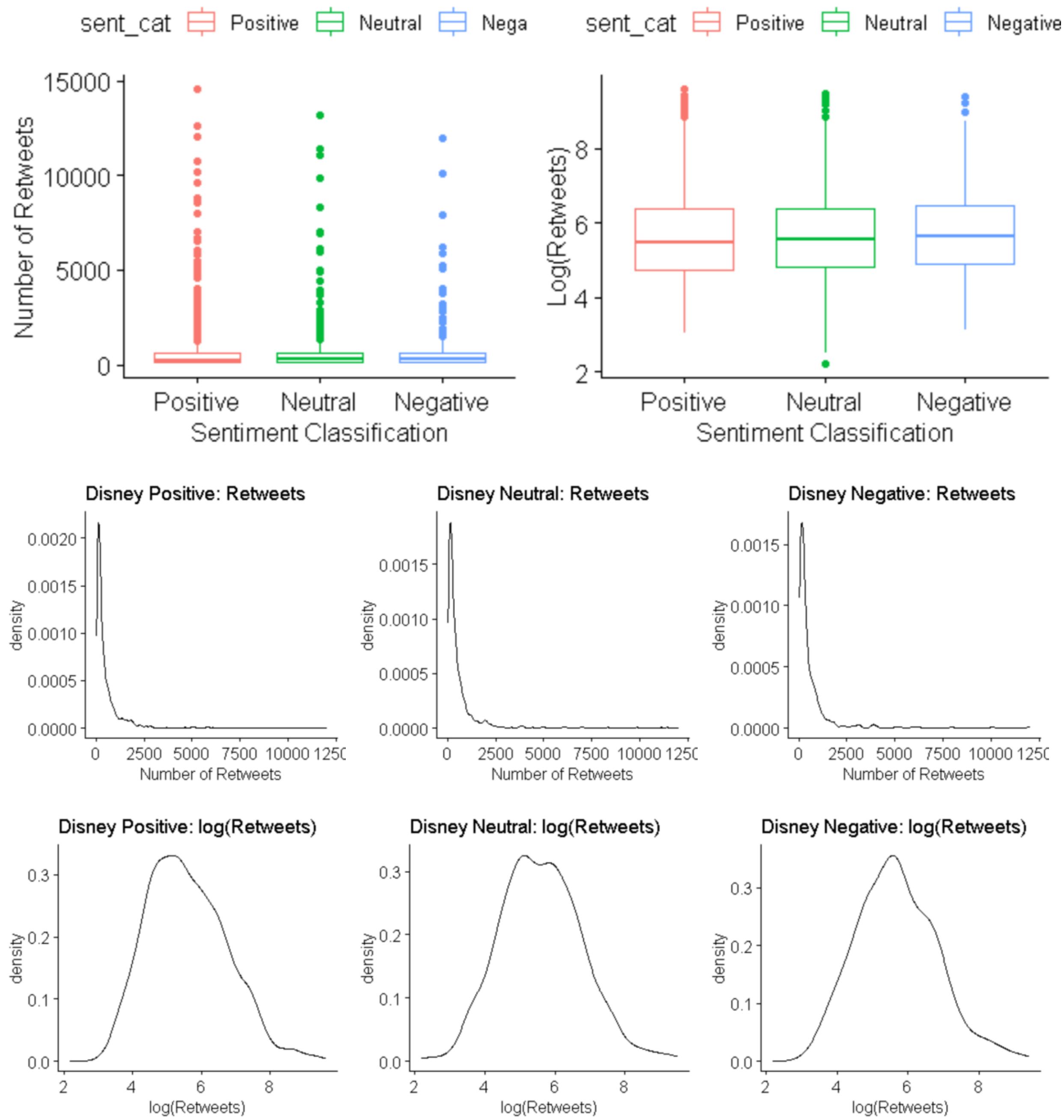
```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 3.1939, df = 2, p-value = 0.2025
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘like’ distributions of all populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of likes which a Disney official tweet receives.**

Disney Official: Number of Retweets



The log distribution does not pass a Shapiro-Wilk normality test.

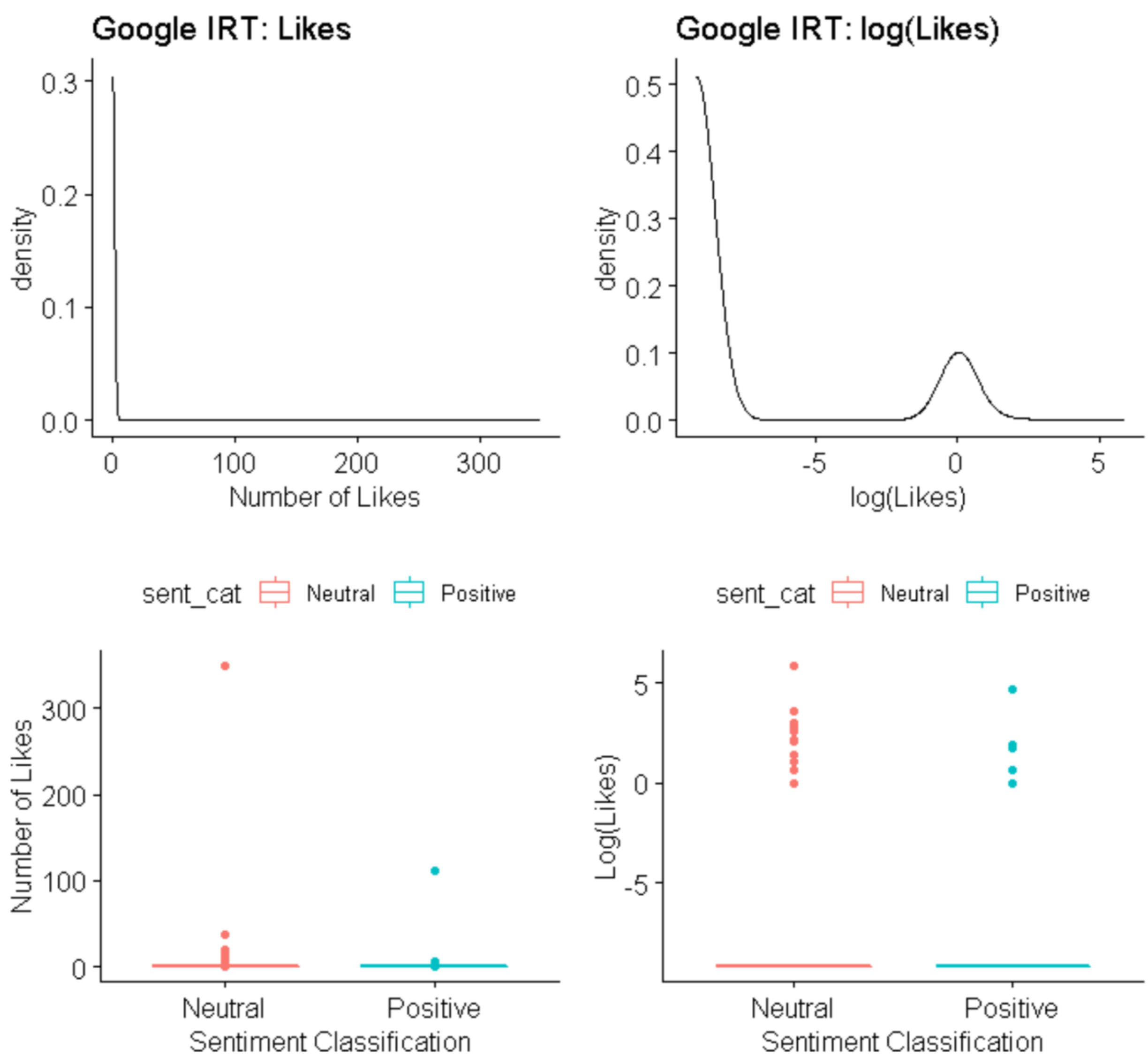


None of the above distributions pass a Shapiro-Wilk normality test.

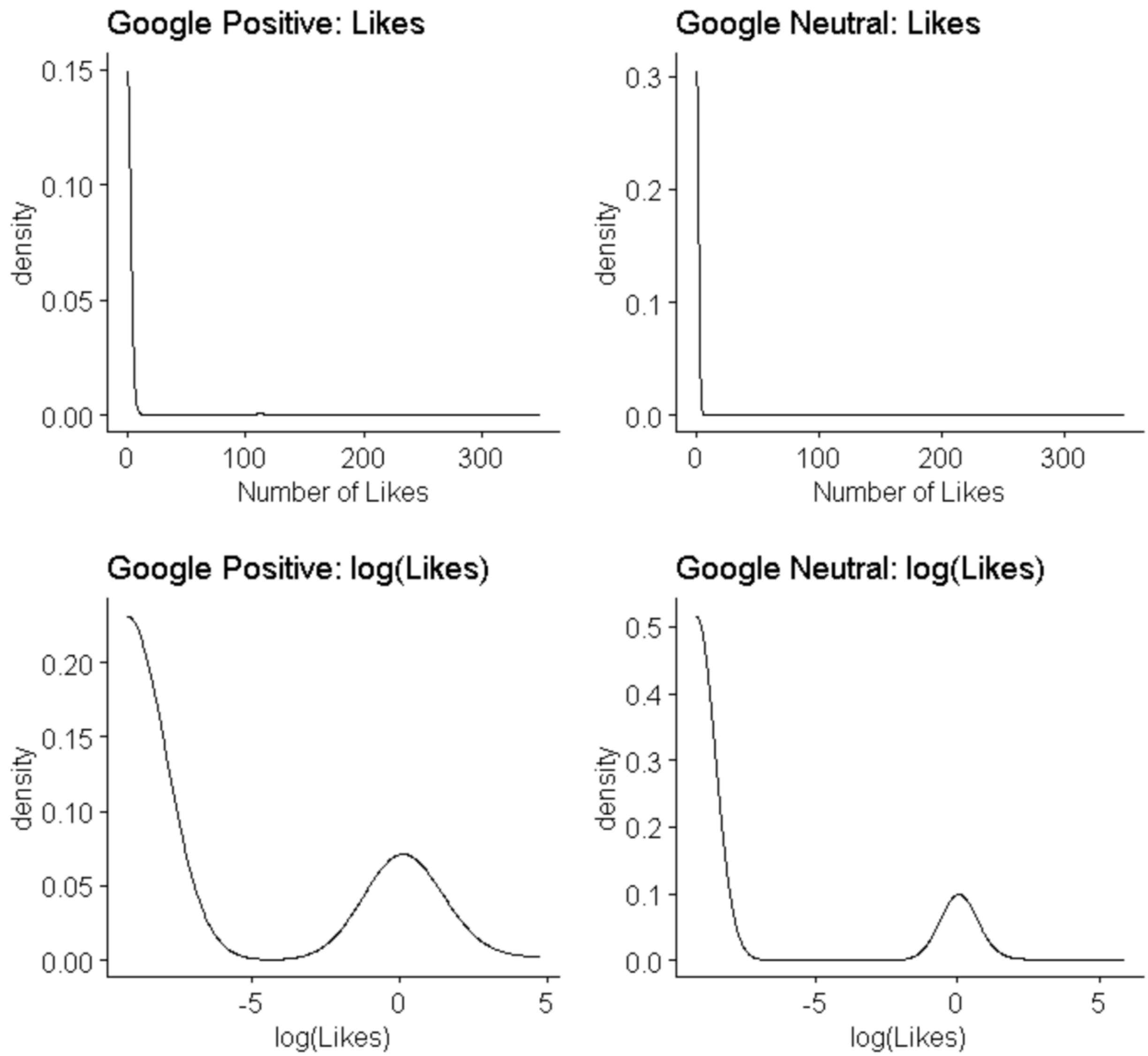
```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 3.0342, df = 2, p-value = 0.2193
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘retweet’ distributions of all populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a Disney official tweet receives.**

Google IRT: Number of Likes

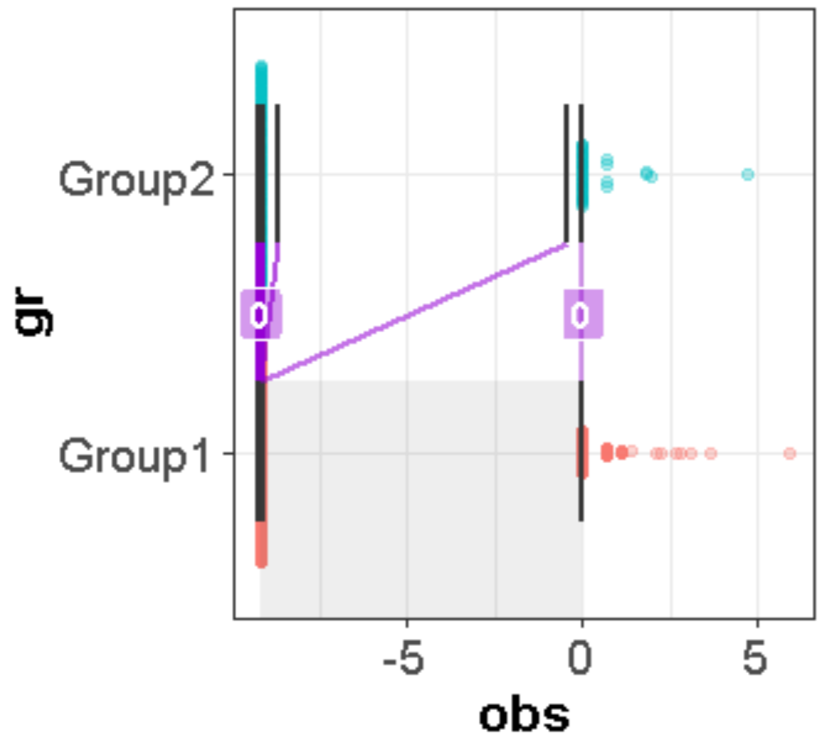


Google only has 4 tweets assigned a negative sentiment, leaving us with only enough tweets in neutral and positive categories for analysis.




```
        Wilcoxon rank sum test with continuity correction

data:  neut_tweets$`Number of Likes` and pos_tweets$`Number of Likes`
W = 250522, p-value = 0.003446
alternative hypothesis: true location shift is not equal to 0
```

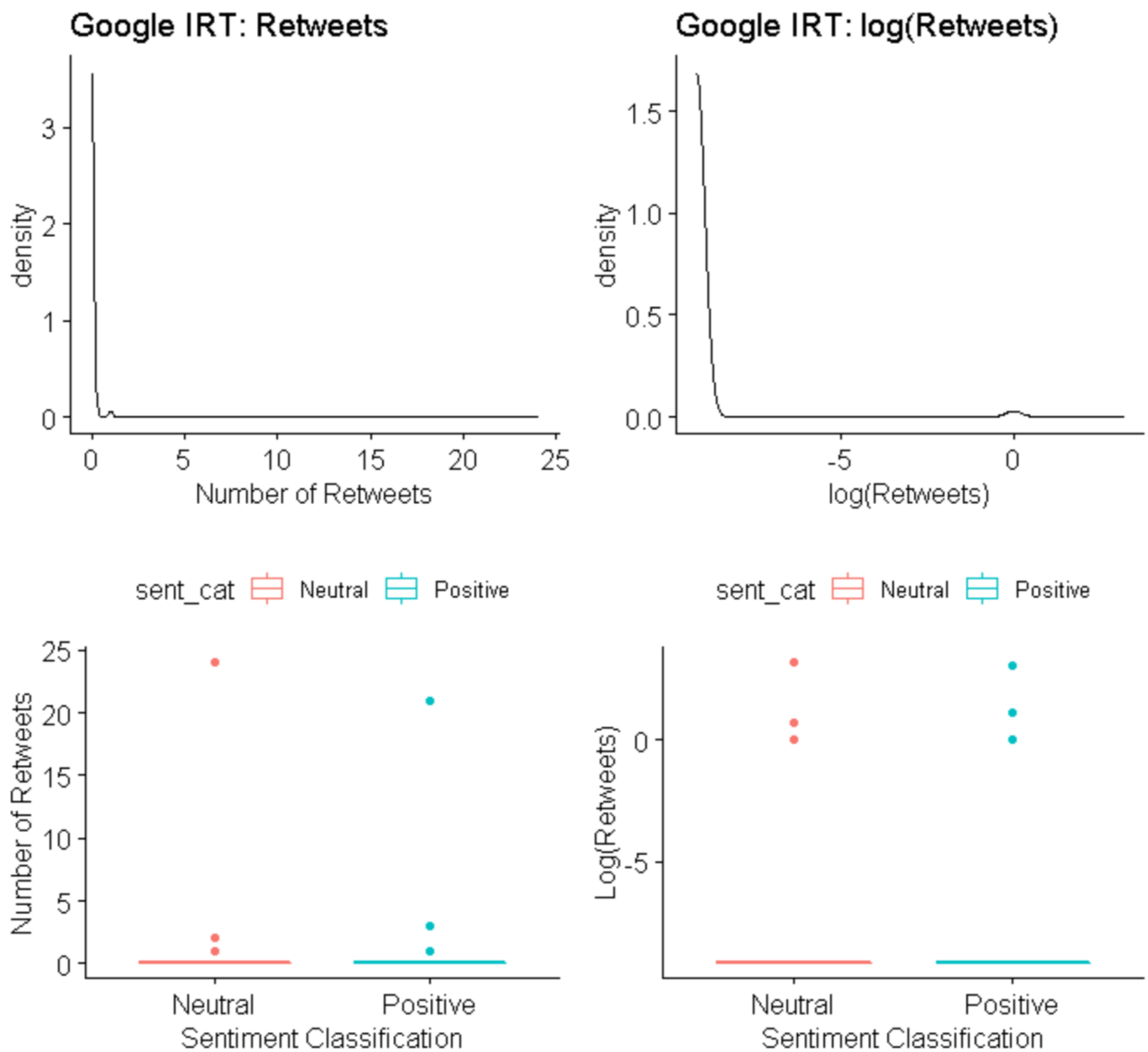
Performing a Mann-Whitney U test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of both populations are equal to one another. Performing a shift function to further analyze the differences produces the following results:

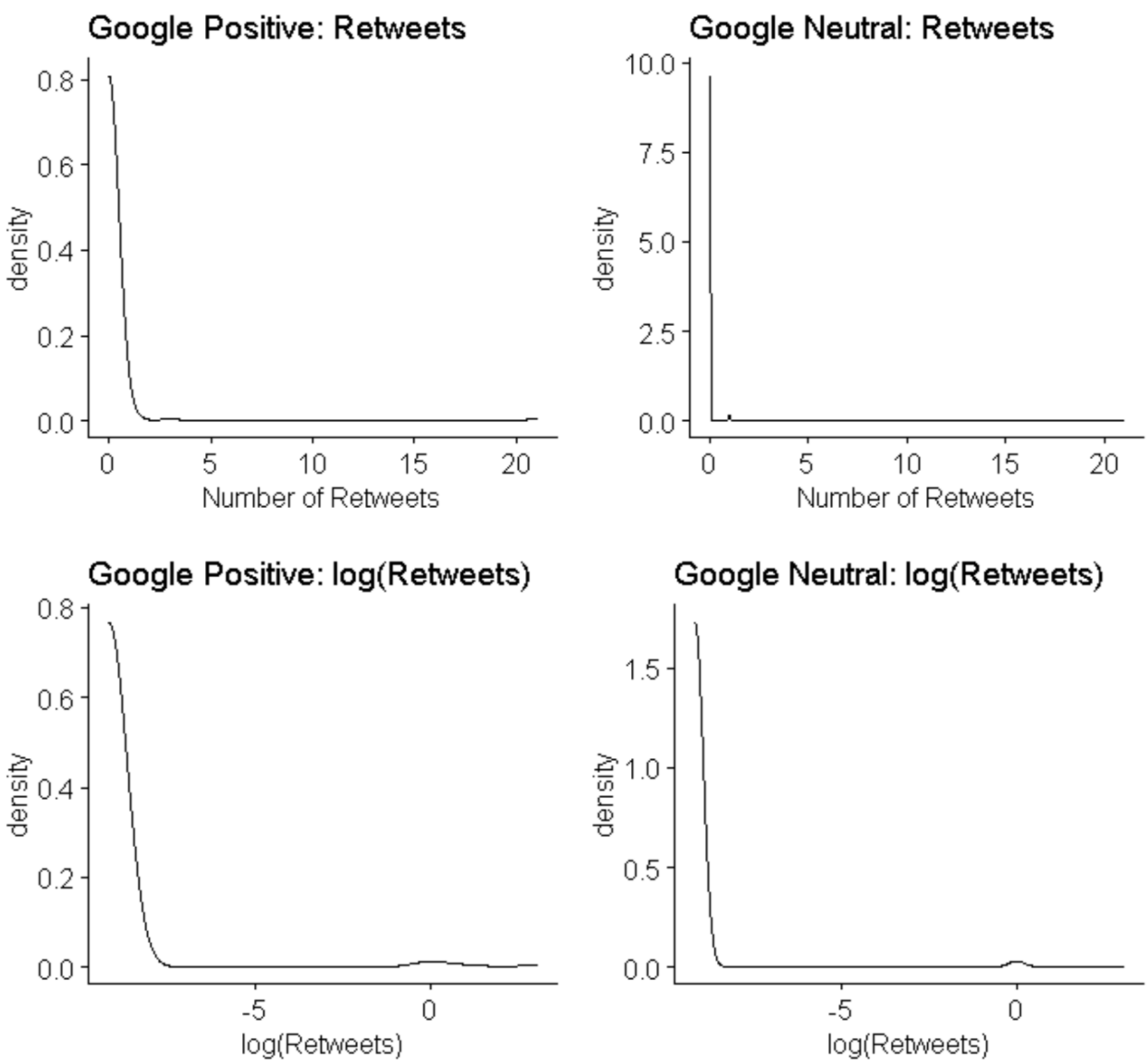



```
$`Group1 - Group2`
   q       Group1       Group2    difference      ci_lower      ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.050000000  1.0000
2 0.2 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.025000000  1.0000
3 0.3 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.016666667  1.0000
4 0.4 0.000000e+00 0.000000e+00  0.000000e+00 -8.881784e-16  0.000000e+00 0.012500000  0.9835
5 0.5 0.000000e+00 3.042011e-14 -3.042011e-14 -3.099371e-07  0.000000e+00 0.008333333  0.2360
6 0.6 0.000000e+00 2.009296e-06 -2.009296e-06 -2.381347e-02 -3.042011e-14 0.007142857  0.0010
7 0.7 0.000000e+00 5.530938e-02 -5.530938e-02 -8.321430e-01 -5.700737e-06 0.006250000  0.0000
8 0.8 8.297065e-06 9.495914e-01 -9.495831e-01 -9.999611e-01 -1.082847e-01 0.005555556  0.0000
9 0.9 1.000000e+00 1.000318e+00 -3.180093e-04 -2.199831e-01  1.174689e-03 0.010000000  0.3020
```

We can say, with 95% confidence, that the 6th through 9th quantiles of group 2 (positive tweets) would need to be shifted down by significant (non-zero) amounts to match their counterparts in group 1 (neutral tweets). However, the sample sizes between the two groups vary greatly (2830 neutral vs. 193 positive), and the visualizations produced above give me pause on drawing any conclusions. **Sentiment does not seem to have a statistically significant effect on the number of likes which a Google IRT tweet receives.**

Google IRT: Number of Retweets

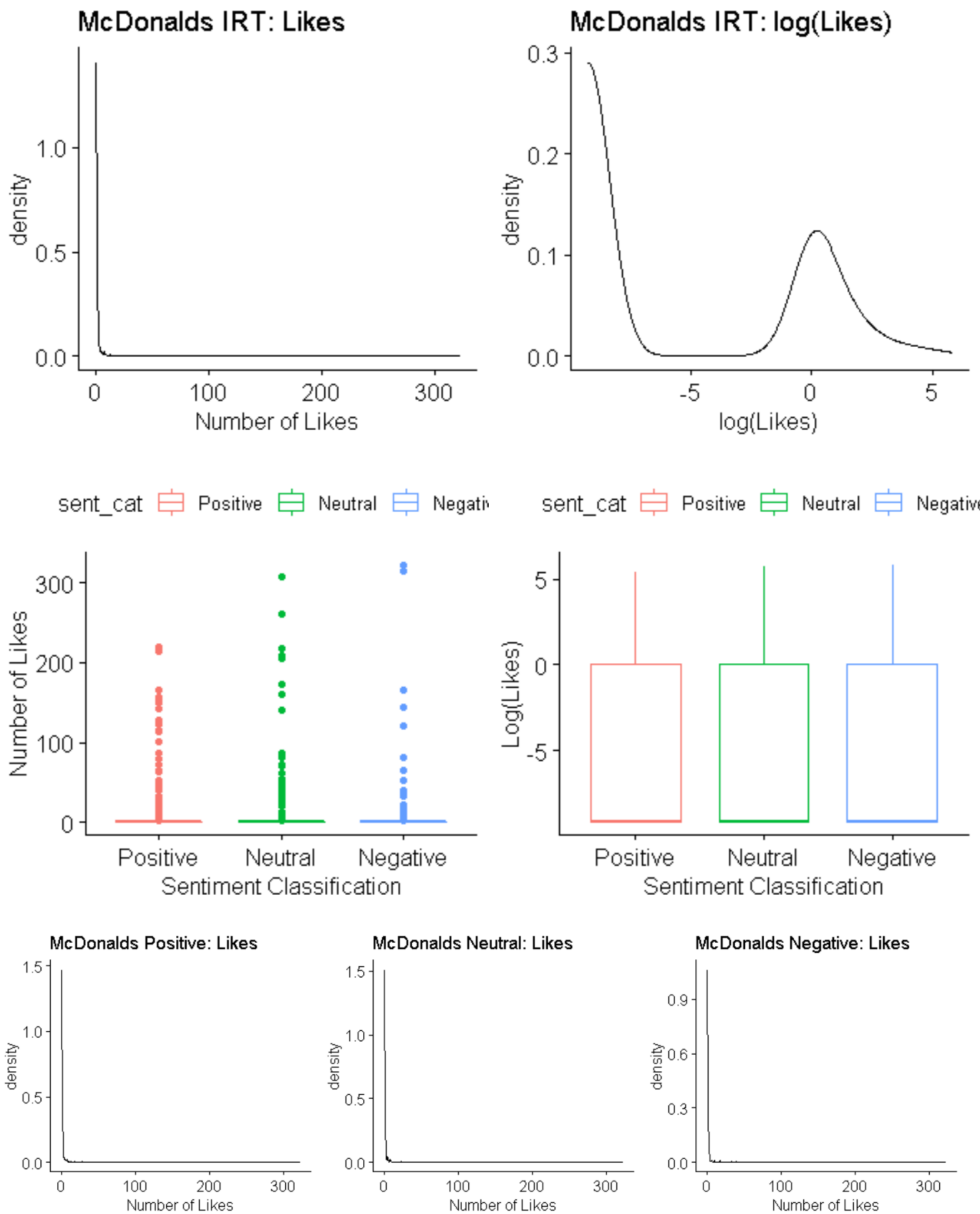


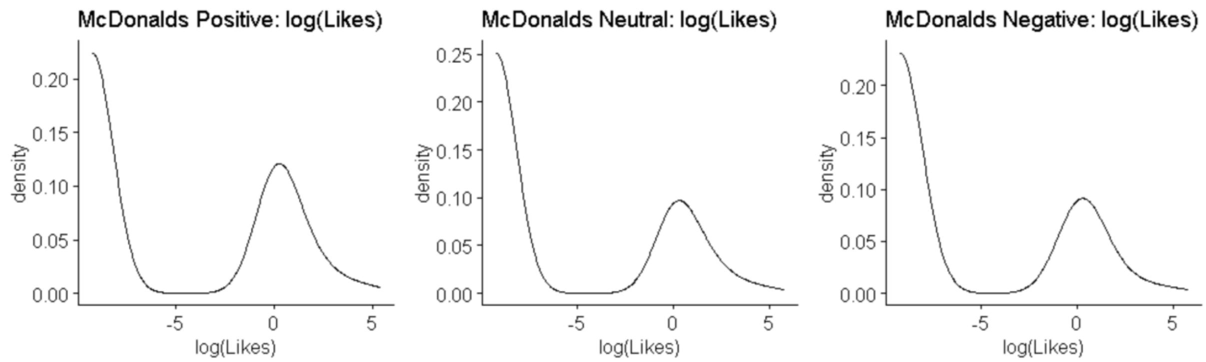



```
        Wilcoxon rank sum test with continuity correction

data:  neut_tweets$`Number of Retweets` and pos_tweets$`Number of Retweets`
W = 270994, p-value = 0.4473
alternative hypothesis: true location shift is not equal to 0
```

Performing a Mann-Whitney U test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘retweet’ distributions of both populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a Google IRT tweet receives.**

McDonalds IRT: Number of Likes



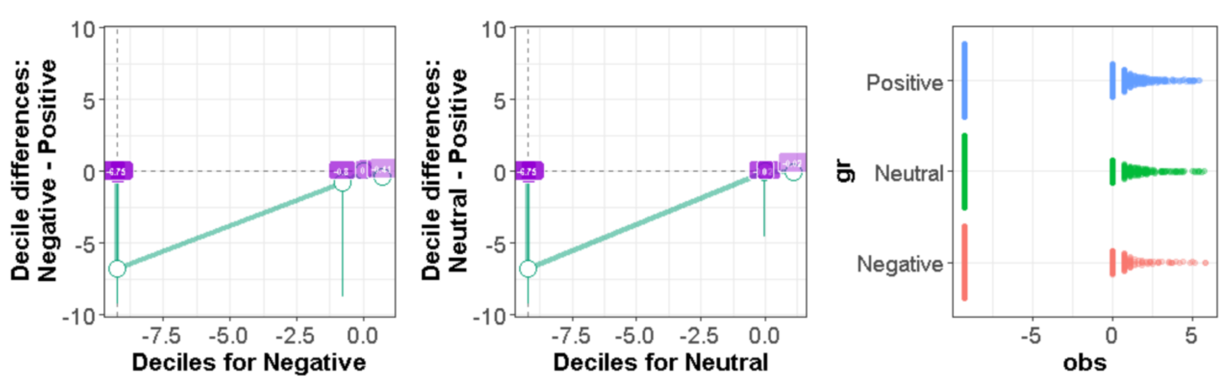



```
         Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 17.522, df = 2, p-value = 0.0001567
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z      P.unadj        P.adj
1  Negative - Neutral -0.7188422 0.4722381677 1.000000000
2 Negative - Positive -3.5573643 0.0003745945 0.001123783
3  Neutral - Positive -3.4036912 0.0006648187 0.001994456
```

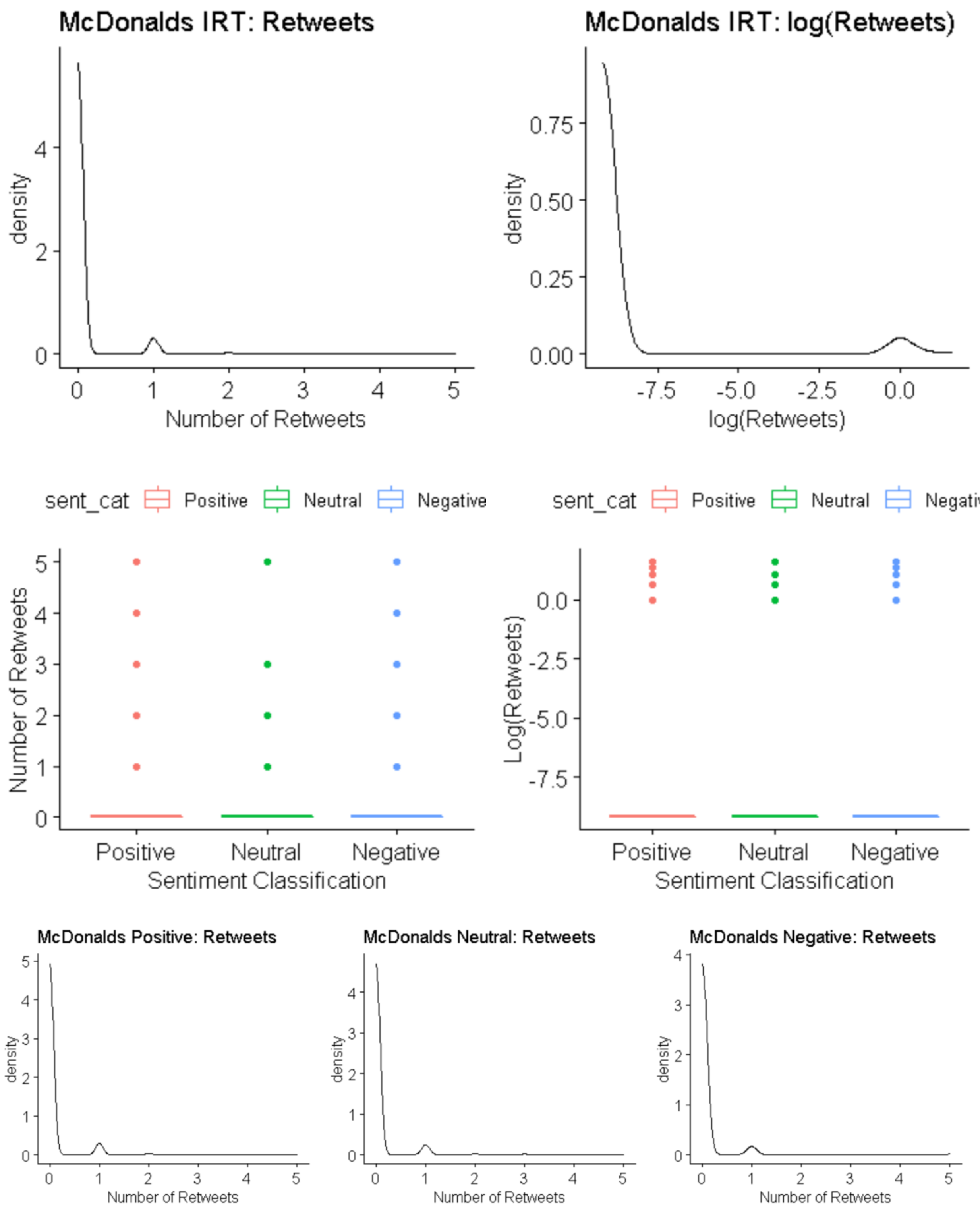
From the results of Dunn’s test, we can see that we may reject the null for both the (Negative, Positive) and the (Neutral, Positive) pairs. In other words, the distribution of likes for positive tweets differs from the distributions of likes for other sentiment categories. Performing a shift function to further examine these differences yields the following:

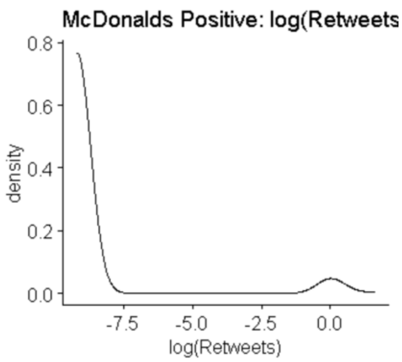


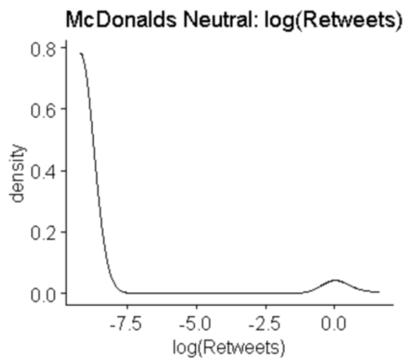
```
$`Negative - Positive`
    q     Negative     Positive    difference      ci_lower      ci_upper        p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.0055555556  1.0000
2 0.2 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.0027777778  1.0000
3 0.3 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.0018518519  1.0000
4 0.4 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.0013888889  1.0000
5 0.5 0.000000e+00 4.595435e-12 -4.595435e-12 -7.716925e-05  8.808210e-11 0.0009259259  0.1445
6 0.6 3.693625e-05 7.333732e-01 -7.333362e-01 -9.997216e-01 -7.992173e-03 0.0006944444  0.0020
7 0.7 9.132777e-01 1.000000e+00 -8.672232e-02 -9.218403e-01 -7.847707e-06 0.0006172840  0.0000
8 0.8 1.000004e+00 1.001772e+00 -1.767471e-03 -4.405036e-01  3.830093e-02 0.0011111111  0.2840
9 0.9 2.040375e+00 3.109482e+00 -1.069106e+00 -2.768249e+00  3.483570e-01 0.0007936508  0.0210

$`Neutral - Positive`
    q      Neutral     Positive    difference      ci_lower      ci_upper        p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 7.936508e-04  1.0000
2 0.2 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 3.968254e-04  1.0000
3 0.3 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 2.645503e-04  1.0000
4 0.4 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 1.984127e-04  1.0000
5 0.5 0.000000e+00 4.595435e-12 -4.595435e-12 -7.716925e-05  0.000000e+00 1.322751e-04  0.0675
6 0.6 5.558154e-06 7.333732e-01 -7.333676e-01 -9.996791e-01 -9.758774e-03 9.920635e-05  0.0000
7 0.7 9.975983e-01 1.000000e+00 -2.401732e-03 -6.674554e-01 -5.847151e-09 8.818342e-05  0.0000
8 0.8 1.000000e+00 1.001772e+00 -1.771722e-03 -7.233325e-01  5.236805e-03 1.133787e-04  0.0760
9 0.9 3.029336e+00 3.109482e+00 -8.014567e-02 -2.216415e+00  1.531606e+00 1.587302e-04  0.8690
```

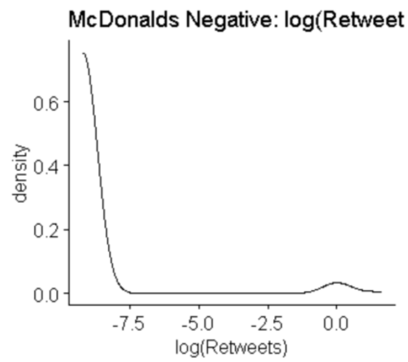
We can say, with 95% confidence, that the 6th and 7th quantiles of positive tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in both neutral and negative tweets. However, given how small the shifts are estimated to be, as well as the visualizations produced above, I don't feel comfortable drawing any conclusions. **Sentiment does not seem to have a statistically significant effect on the number of likes which a McDonalds IRT tweet receives.**

McDonalds IRT: Number of Retweets





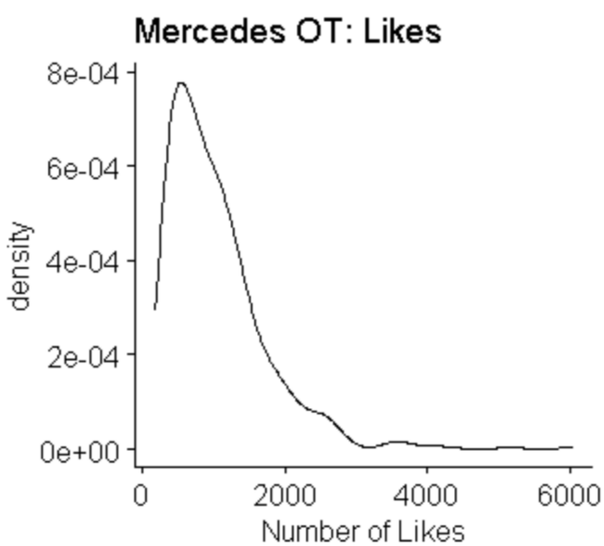


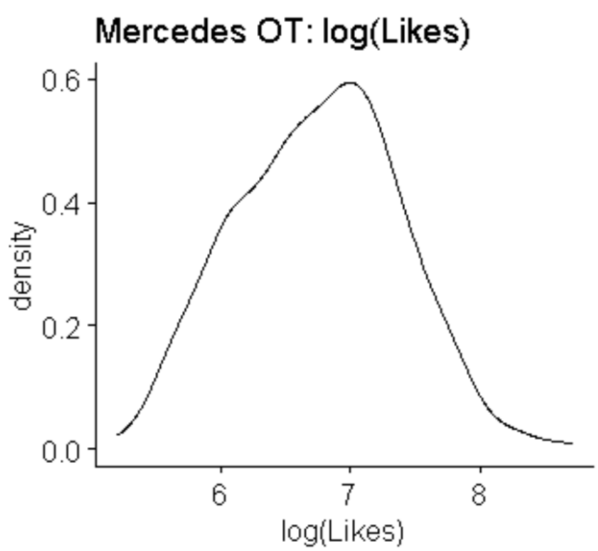



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 1.7911, df = 2, p-value = 0.4084
```

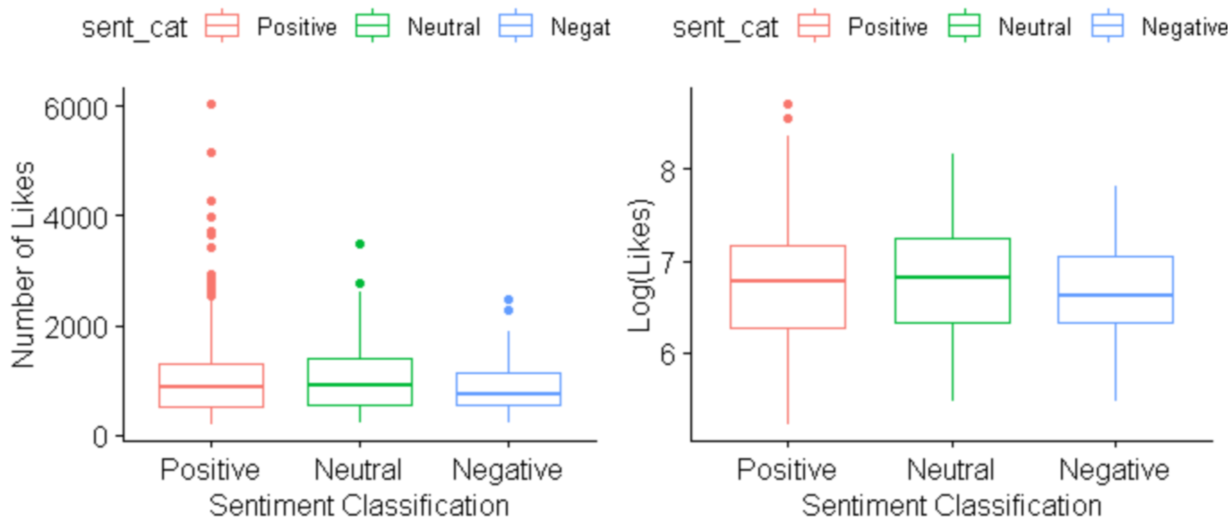
Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of all populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a McDonalds IRT tweet receives.**

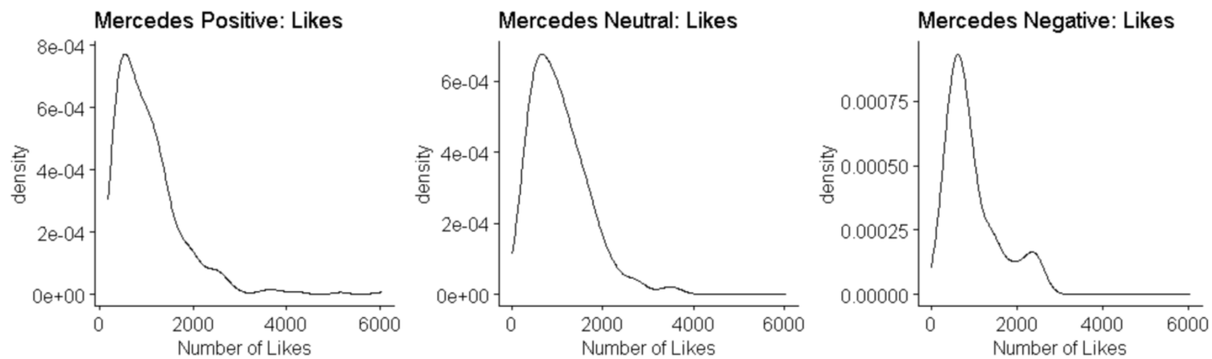
Mercedes Official: Number of Likes

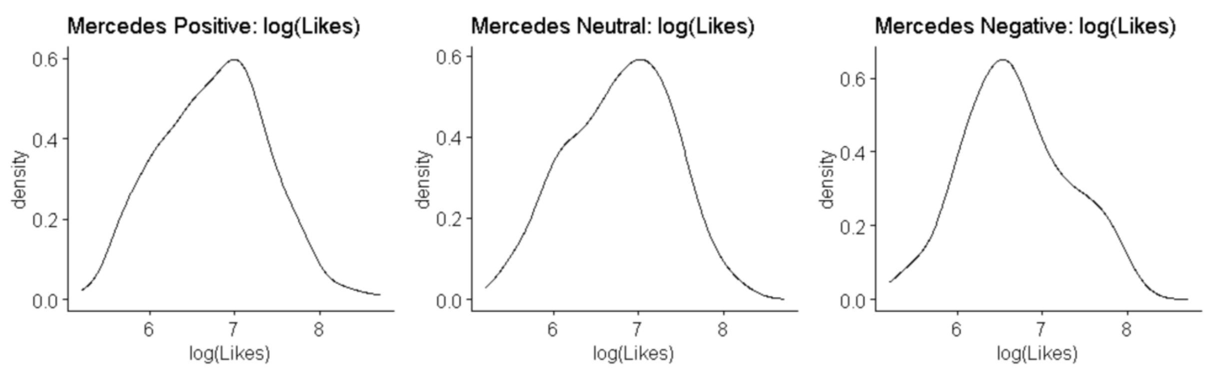




The log distribution barely does not pass a Shapiro-Wilk normality test.







All three log distributions above pass a Shapiro-Wilk normality test.

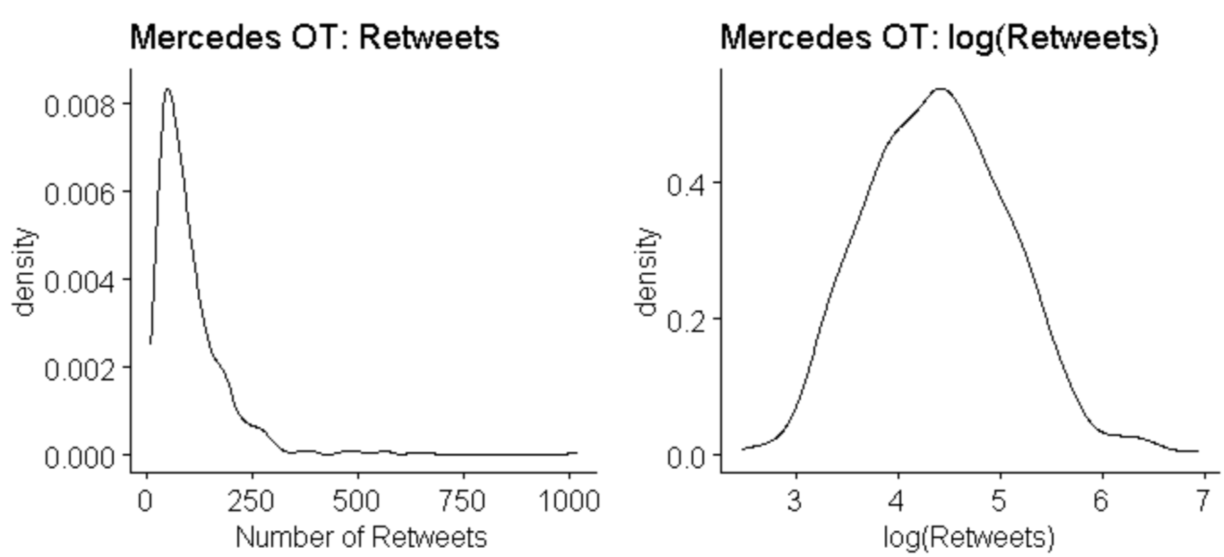
```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 0.66526, df = 2, p-value = 0.717
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'like' distributions of all populations are equal. However, since all 3 log distributions may be considered normal, it seems worthwhile to perform ANOVA as well:

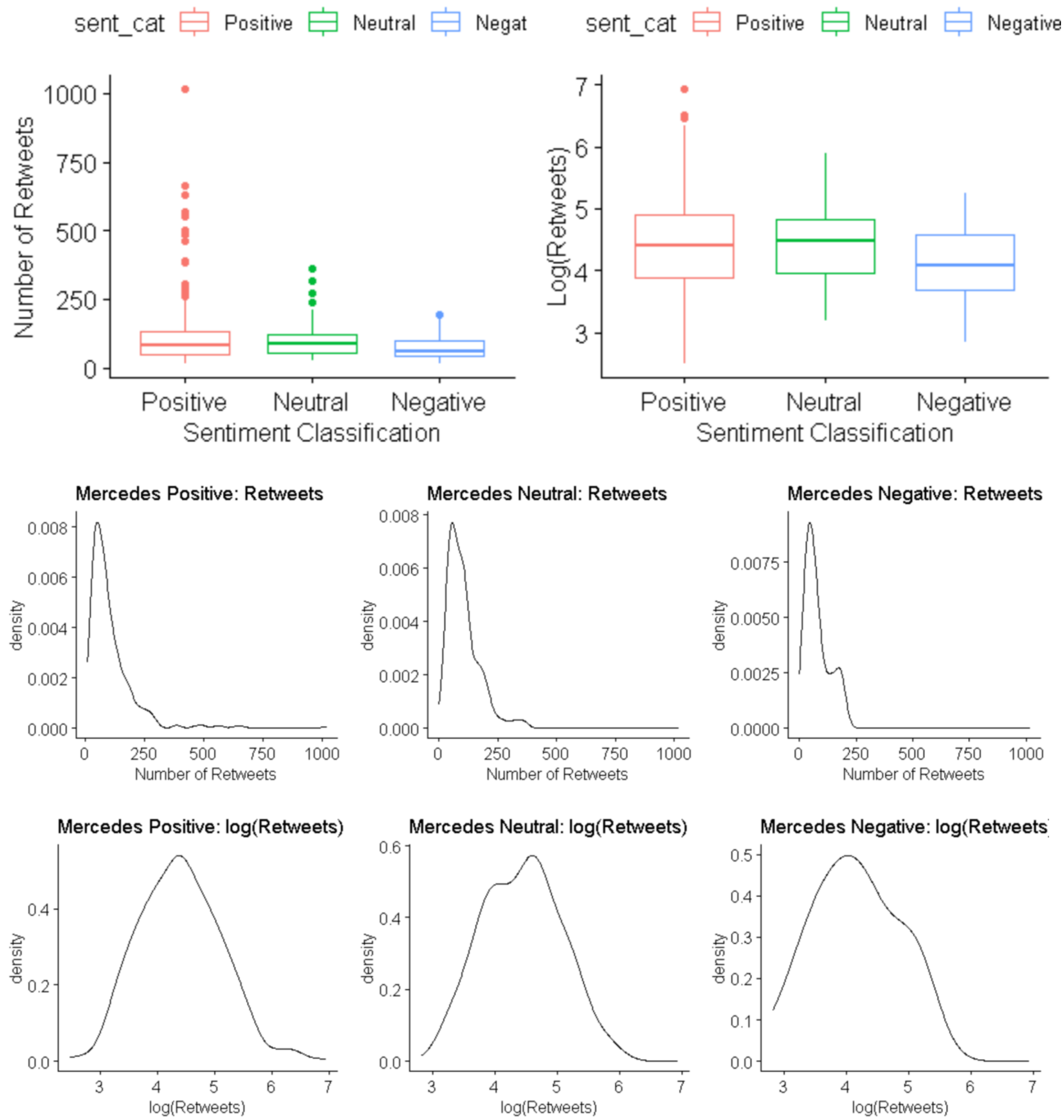
```
> #Since the log dists. are all normal, try ANOVA
> nova = aov(logLikes ~ sent_cat, data = tweets)
> summary(nova)
             Df Sum Sq Mean Sq F value Pr(>F)
sent_cat      2   0.18  0.0899   0.226  0.797
Residuals   498 197.76  0.3971
```

Performing ANOVA results in an insignificant p-value as well. Therefore, we fail to reject the null hypothesis that the mean expected number of likes are equivalent between all sentiment categories. **Sentiment does not seem to have a statistically significant effect on the number of likes which a Mercedes official tweet receives.**

Mercedes Official: Number of Retweets



The log distribution does not pass a Shapiro-Wilk normality test.

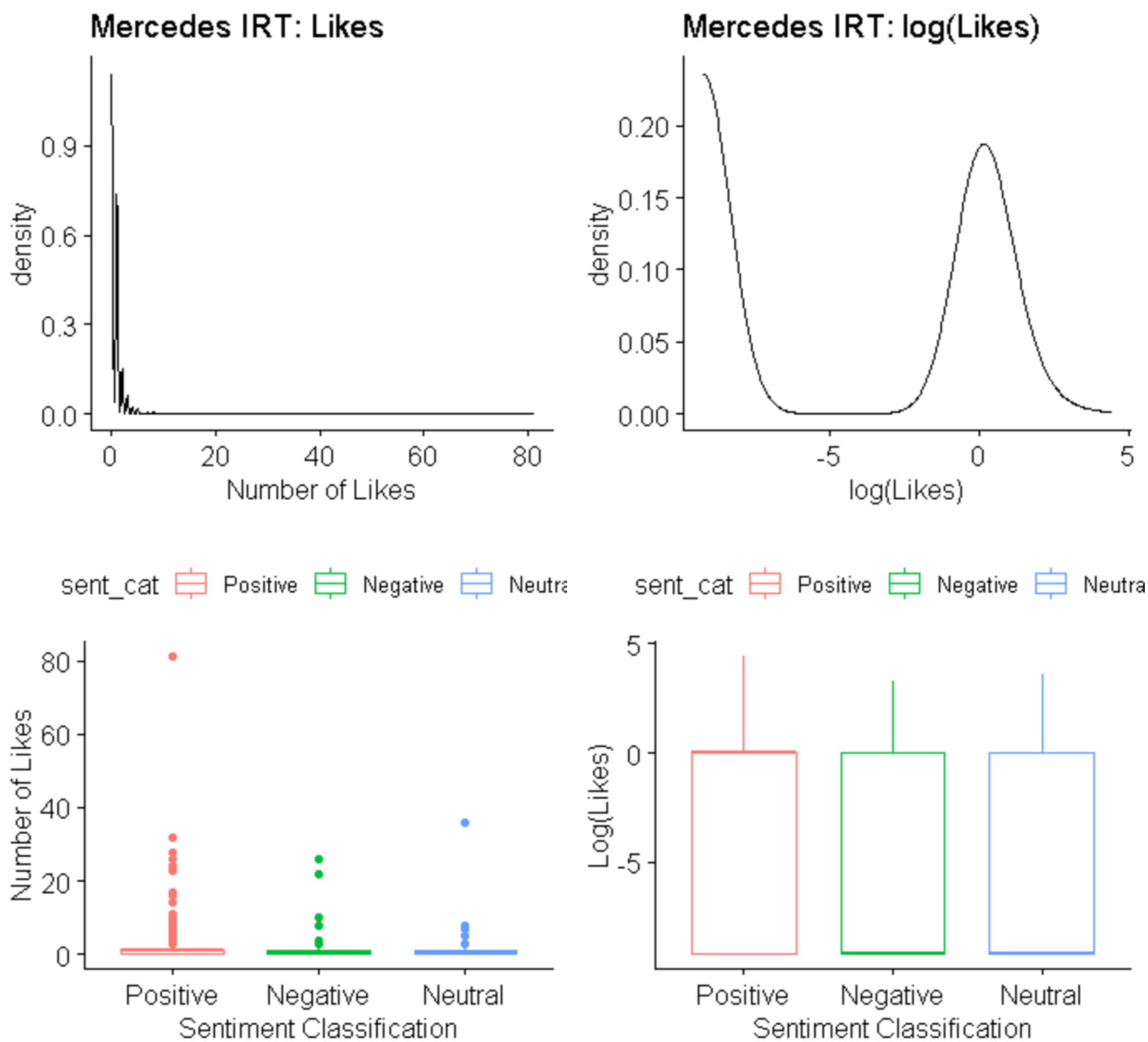


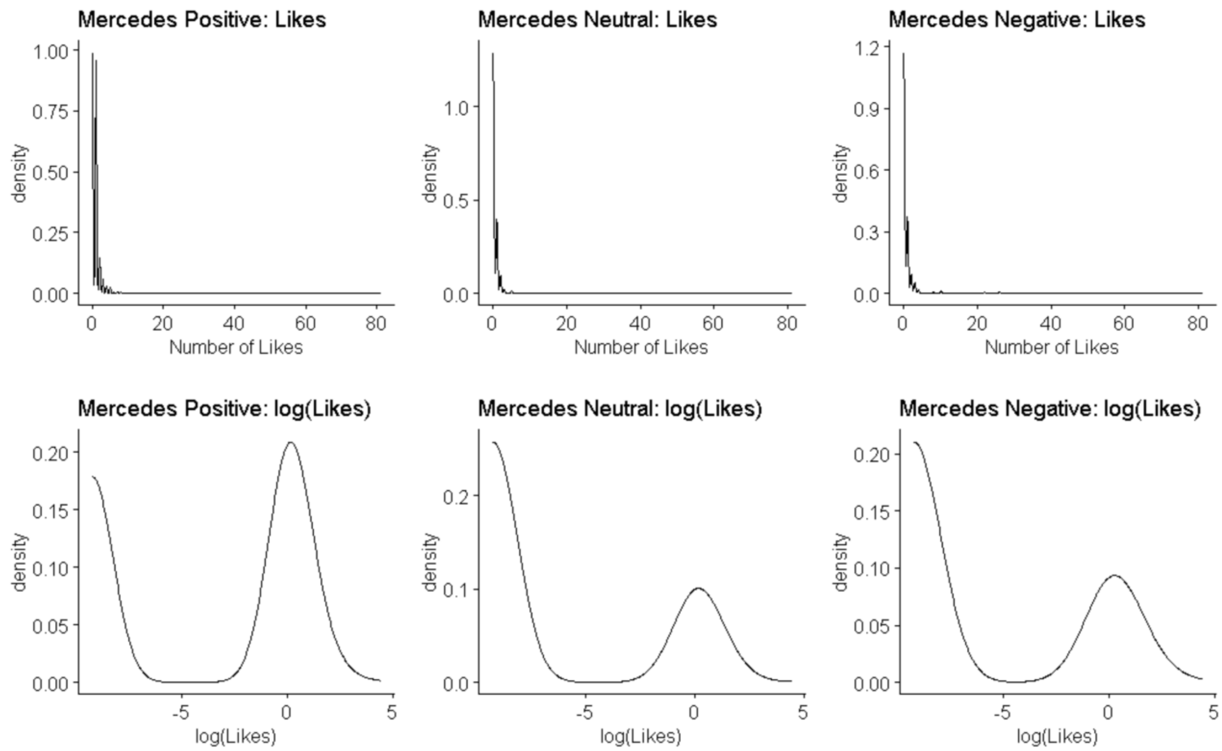
The above left log distribution does not pass a Shapiro-Wilk normality test, but the middle and right log distributions do.

```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 2.6368, df = 2, p-value = 0.2676
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘retweet’ distributions of all populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a Mercedes official tweet receives.**

Mercedes IRT: Number of Likes



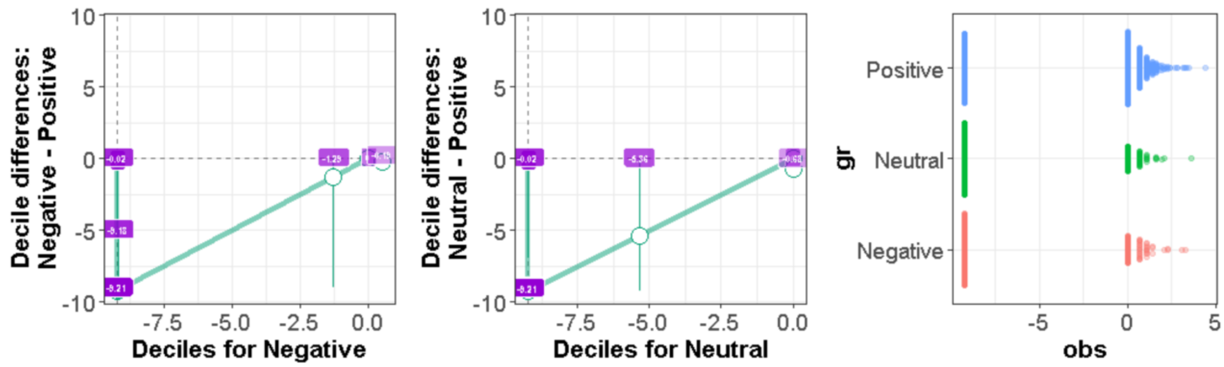



```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 138.39, df = 2, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z      P.unadj        P.adj
1  Negative - Neutral   1.247886 2.120729e-01 6.362188e-01
2 Negative - Positive  -7.130613 9.992256e-13 2.997677e-12
3  Neutral - Positive -10.658871 1.585105e-26 4.755315e-26
```

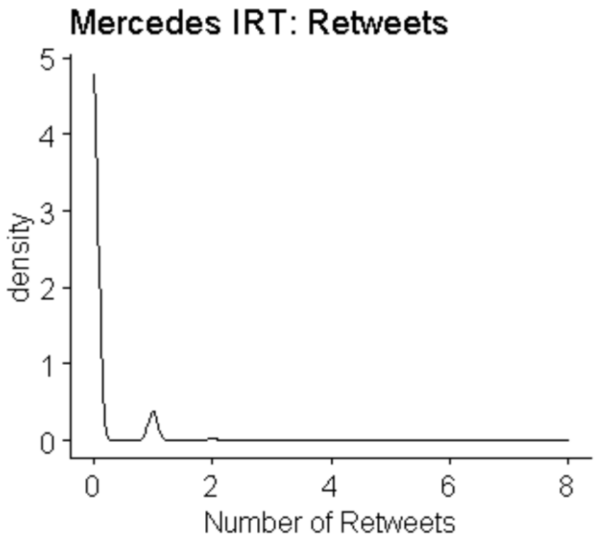
From the results of Dunn's test, we can see that we may reject the null for both the (Negative, Positive) and the (Neutral, Positive) pairs. In other words, the distribution of likes for positive tweets differs from the distributions of likes for other sentiment categories. Performing a shift function to further examine these differences yields the following:

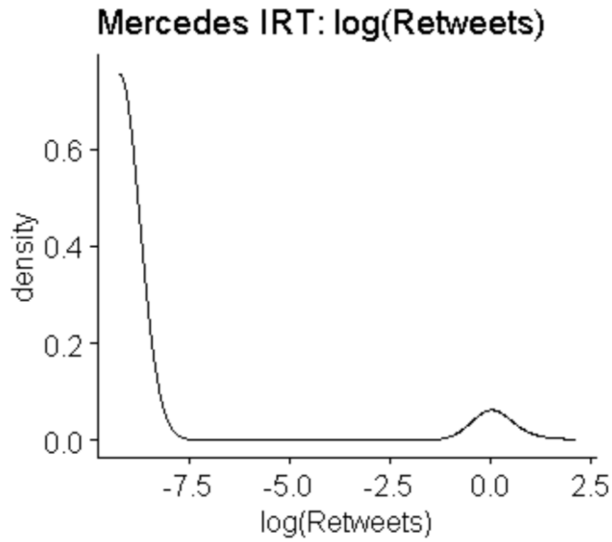


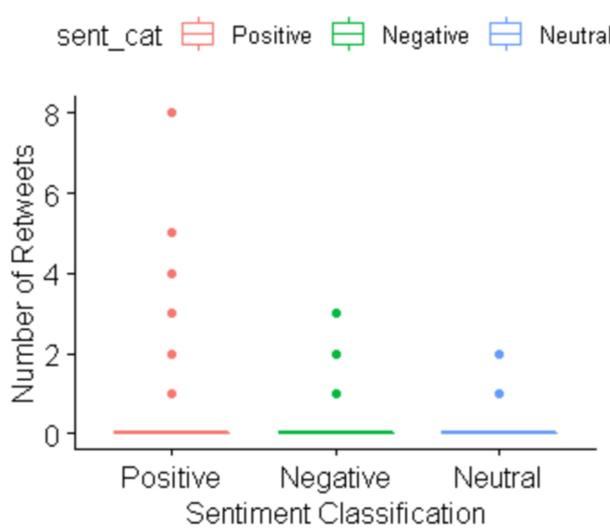

```
$`Negative - Positive`
    q     Negative    Positive    difference      ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.00000000  0.000000e+00  0.000000e+00  0.000000e+00 0.0062500000  1.0000
2 0.2 0.000000e+00 0.00000000  0.000000e+00  0.000000e+00  0.000000e+00 0.0031250000  1.0000
3 0.3 0.000000e+00 0.00000000  0.000000e+00 -5.551115e-16  0.000000e+00 0.0020833333  0.9985
4 0.4 0.000000e+00 0.00201806 -2.018060e-03 -6.976505e-01 -1.576486e-09 0.0010416667  0.0000
5 0.5 4.560330e-11 0.99999990 -9.999999e-01 -1.000000e+00 -9.851814e-01 0.0008928571  0.0000
6 0.6 2.853402e-03 1.00000000 -9.971466e-01 -1.000000e+00 -3.242502e-01 0.0007812500  0.0000
7 0.7 8.600597e-01 1.00000000 -1.399403e-01 -9.823941e-01 -1.148570e-04 0.0006944444  0.0000
8 0.8 1.000001e+00 1.00000000  7.203420e-07 -9.041667e-03  5.955954e-02 0.0015625000  0.9040
9 0.9 1.721489e+00 2.00000147 -2.785123e-01 -1.010542e+00  3.716089e-01 0.0012500000  0.1140

$`Neutral - Positive`
    q      Neutral    Positive    difference      ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.00000000  0.000000e+00  0.000000e+00  0.000000e+00 0.0012500000  1.0000
2 0.2 0.000000e+00 0.00000000  0.000000e+00  0.000000e+00  0.000000e+00 0.0006250000  1.0000
3 0.3 0.000000e+00 0.00000000  0.000000e+00 -5.551115e-16  0.000000e+00 0.0004166667  0.9995
4 0.4 0.000000e+00 0.00201806 -2.018060e-03 -6.228945e-01 -1.576486e-09 0.0002500000  0.0000
5 0.5 0.000000e+00 0.99999990 -9.999999e-01 -1.000000e+00 -9.659634e-01 0.0002083333  0.0000
6 0.6 4.902939e-08 1.00000000 -1.000000e+00 -1.000000e+00 -9.715160e-01 0.0001785714  0.0000
7 0.7 4.182342e-01 1.00000000 -5.817658e-01 -9.998718e-01 -1.973586e-03 0.0001562500  0.0000
8 0.8 9.999999e-01 1.00000000 -6.971327e-08 -1.973117e-02  2.595208e-09 0.0003125000  0.0080
9 0.9 1.020887e+00 2.00000147 -9.791146e-01 -1.361845e+00 -8.395417e-02 0.0001388889  0.0000
```

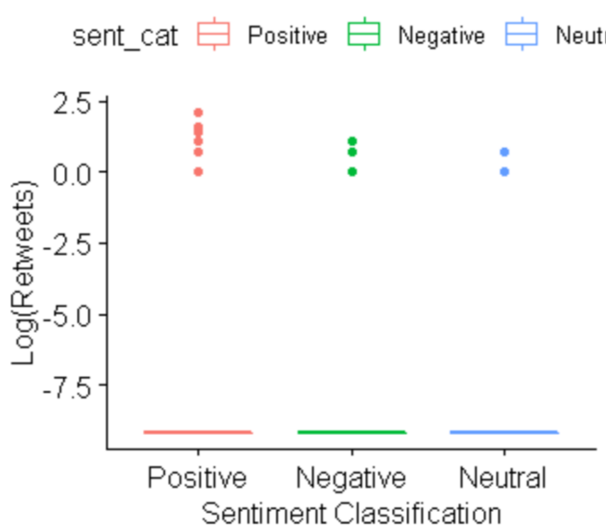
We can say, with 95% confidence, that the 4th through 7th quantiles of positive tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of negative tweets, as well as the set of neutral tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes, and the mid quantiles specifically), and potentially one underlying factor explaining these differences is the sentiment of Mercedes IRT tweets (positive sentiment behaving differently).** The real story here seems to be that positive Mercedes IRT tweets are more likely to receive a non-zero amount of likes than negative or neutral Mercedes IRT tweets are. We can see in the scatterplot above that positive tweets have a long bar at x = 0 (which back transforms to ‘number of likes’ = 1), while neutral and negative categories have very short bars at x = 0. All categories have long bars at x = -7.5 (which back transforms to 0 likes), but only positive tweets have a matching long bar at x = 0.

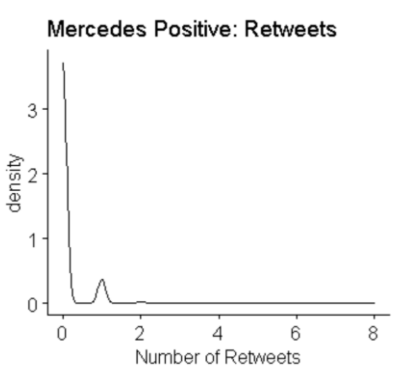
Mercedes IRT: Number of Retweets

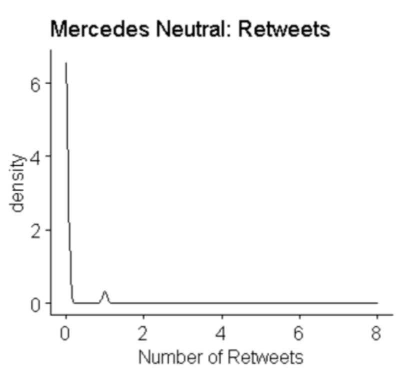


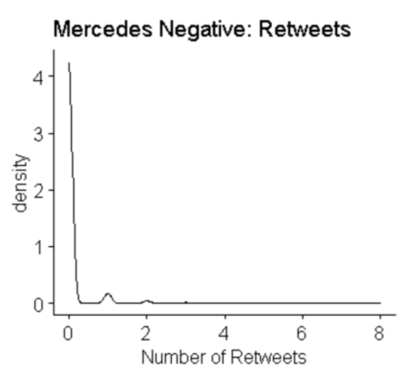


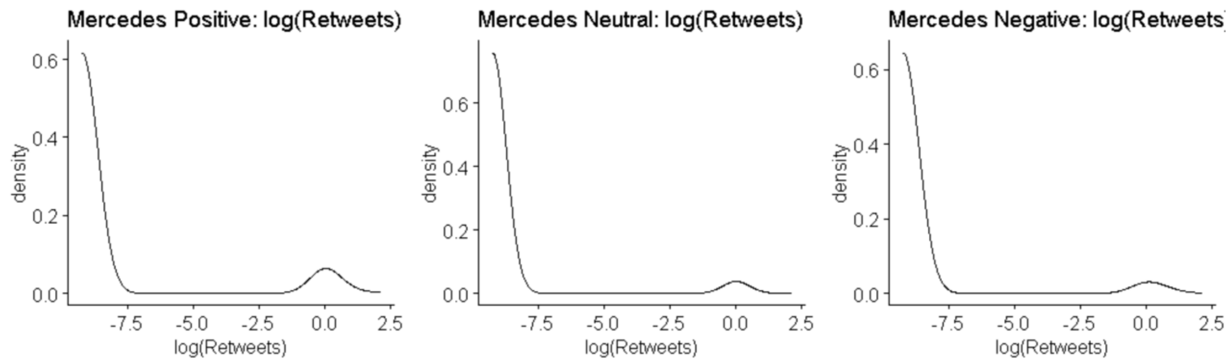










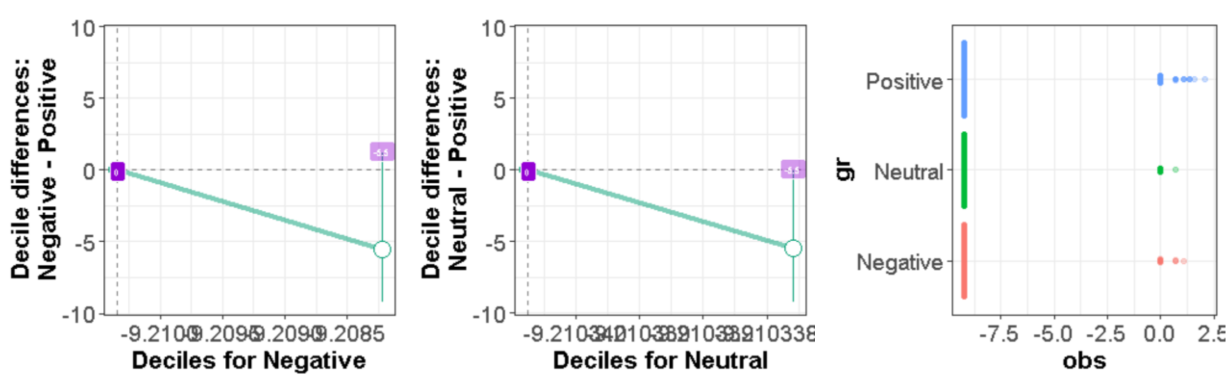



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 20.813, df = 2, p-value = 3.024e-05
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z       P.unadj         P.adj
1  Negative - Neutral  0.2038912 8.384385e-01 1.0000000000
2 Negative - Positive -2.9833719 2.850914e-03 0.0085527406
3  Neutral - Positive -4.0070347 6.148583e-05 0.0001844575
```

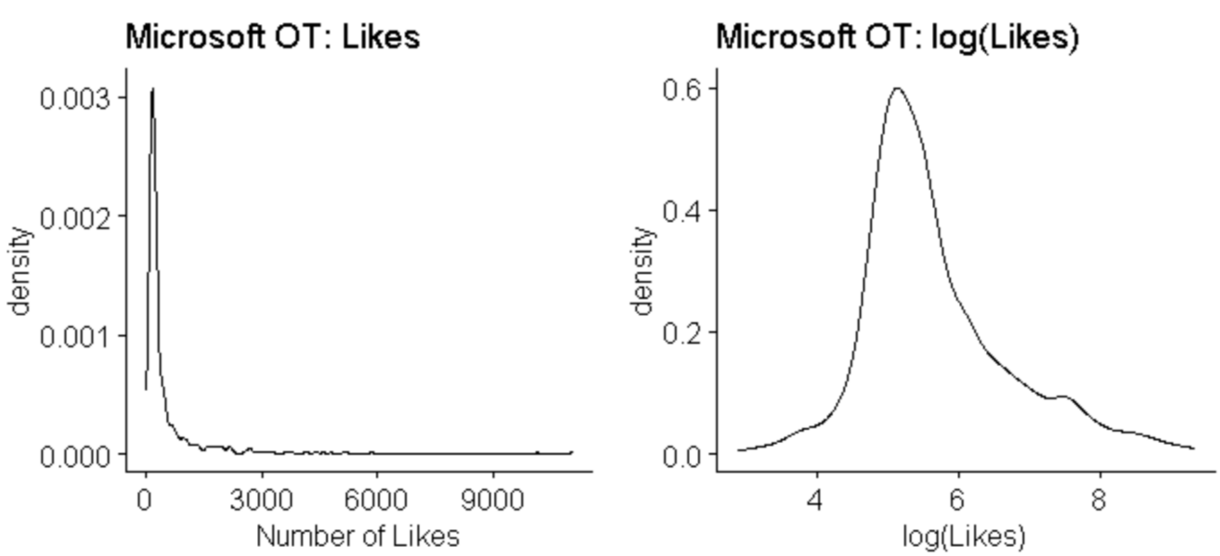
From the results of Dunn's test, we can see that we may reject the null for both the (Negative, Positive) and the (Neutral, Positive) pairs. In other words, the distribution of retweets for positive tweets differs from the distributions of retweets for other sentiment categories. Performing a shift function to further examine these differences yields the following:



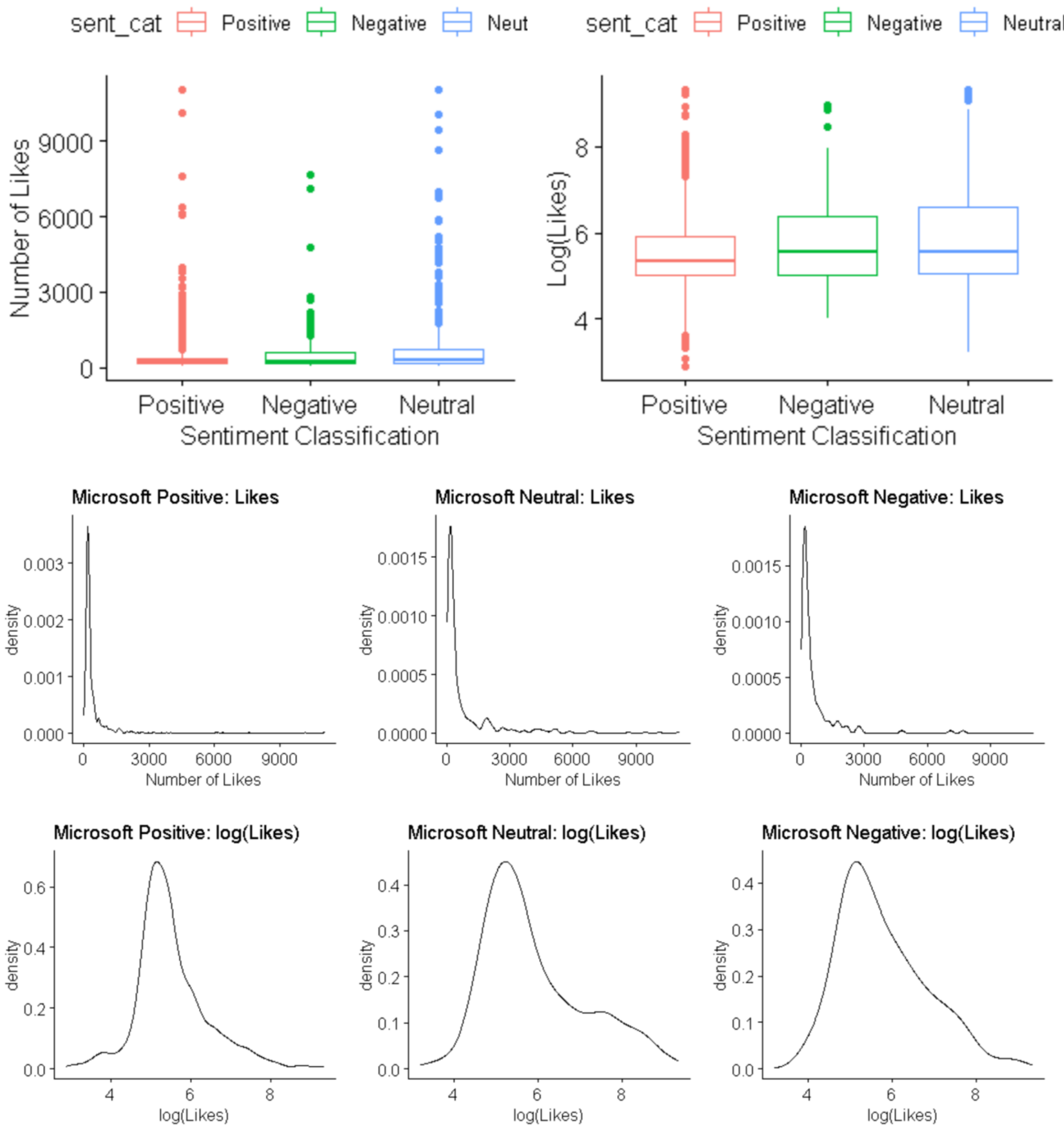
```
$`Negative - Positive`
    q     Negative  Positive difference   ci_lower     ci_upper       p_crit p_value
1 0.1 0.0000000000 0.0000000  0.0000000  0.0000000 0.000000e+00 0.0055555556   1.000
2 0.2 0.0000000000 0.0000000  0.0000000  0.0000000 0.000000e+00 0.0027777778   1.000
3 0.3 0.0000000000 0.0000000  0.0000000  0.0000000 0.000000e+00 0.0018518519   1.000
4 0.4 0.0000000000 0.0000000  0.0000000  0.0000000 0.000000e+00 0.0013888889   1.000
5 0.5 0.0000000000 0.0000000  0.0000000  0.0000000 0.000000e+00 0.0011111111   1.000
6 0.6 0.0000000000 0.0000000  0.0000000  0.0000000 0.000000e+00 0.0009259259   1.000
7 0.7 0.0000000000 0.0000000  0.0000000  0.0000000 0.000000e+00 0.0007936508   1.000
8 0.8 0.0000000000 0.0000000  0.0000000  0.0000000 3.155350e-09 0.0006944444   0.811
9 0.9 0.0002301449 0.5971144 -0.5968843 -0.9990088 2.750505e-02 0.0006172840   0.005

$`Neutral - Positive`
    q      Neutral  Positive difference      ci_lower     ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.000000000 6.172840e-04  1.0000
2 0.2 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.000000000 3.086420e-04  1.0000
3 0.3 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.000000000 2.057613e-04  1.0000
4 0.4 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.000000000 1.543210e-04  1.0000
5 0.5 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.000000000 1.234568e-04  1.0000
6 0.6 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.000000000 1.028807e-04  1.0000
7 0.7 0.000000e+00 0.0000000  0.0000000  0.000000e+00  0.000000000 8.818342e-05  1.0000
8 0.8 0.000000e+00 0.0000000  0.0000000 -2.220446e-16  0.000000000 7.716049e-05  0.9995
9 0.9 2.256223e-07 0.5971144 -0.5971142 -9.994481e-01 -0.001172859 6.858711e-05  0.0000
```

We can say, with 95% confidence, that the 9th quantile of positive tweets would need to be shifted down by significant (non-zero) amounts to match its counterparts in negative and neutral tweets. **Given the visualizations produced above, as well as the shifts being contained only to the 9th quantiles, sentiment does not seem to have an effect on the number of retweets a Mercedes IRT tweet receives.**

Microsoft Official: Number of Likes



The log distribution does not pass a Shapiro-Wilk normality test.



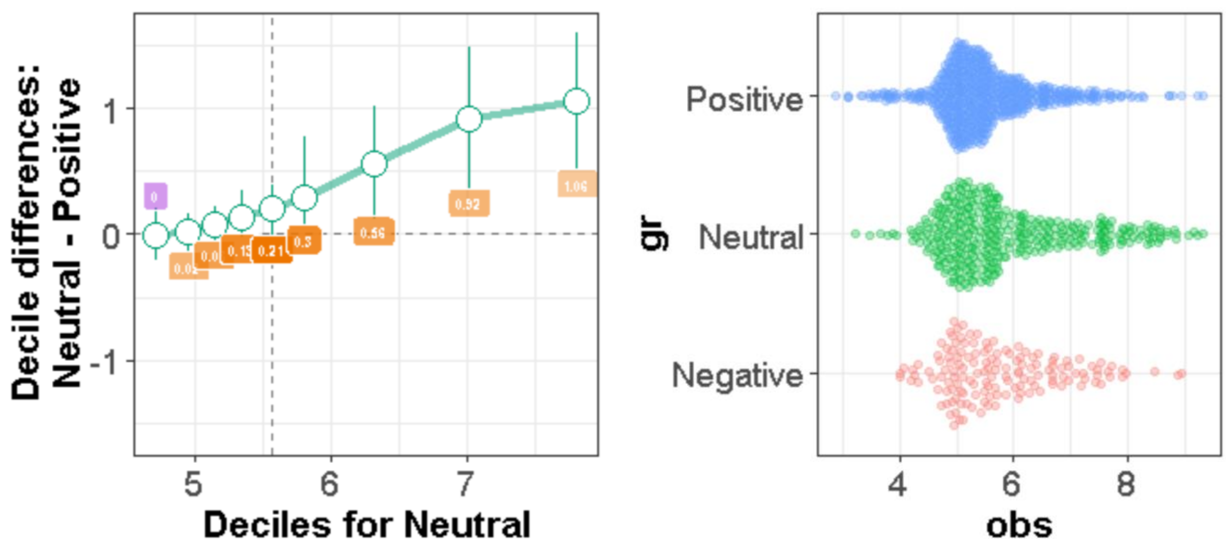
None of the above distributions pass a Shapiro-Wilk normality test.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 19.88, df = 2, p-value = 4.821e-05
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison         Z      P.unadj        P.adj
1  Negative - Neutral -0.587237 5.570445e-01 1.000000e+00
2 Negative - Positive  2.346190 1.896641e-02 5.689924e-02
3  Neutral - Positive  4.244706 2.188801e-05 6.566404e-05
```

From the results of Dunn's test, we can see that we may reject the null for only the (Neutral, Positive) pairing. In other words, the distribution of likes for positive tweets differs from the distributions of likes for neutral tweets. Performing a shift function to further examine these differences yields the following:

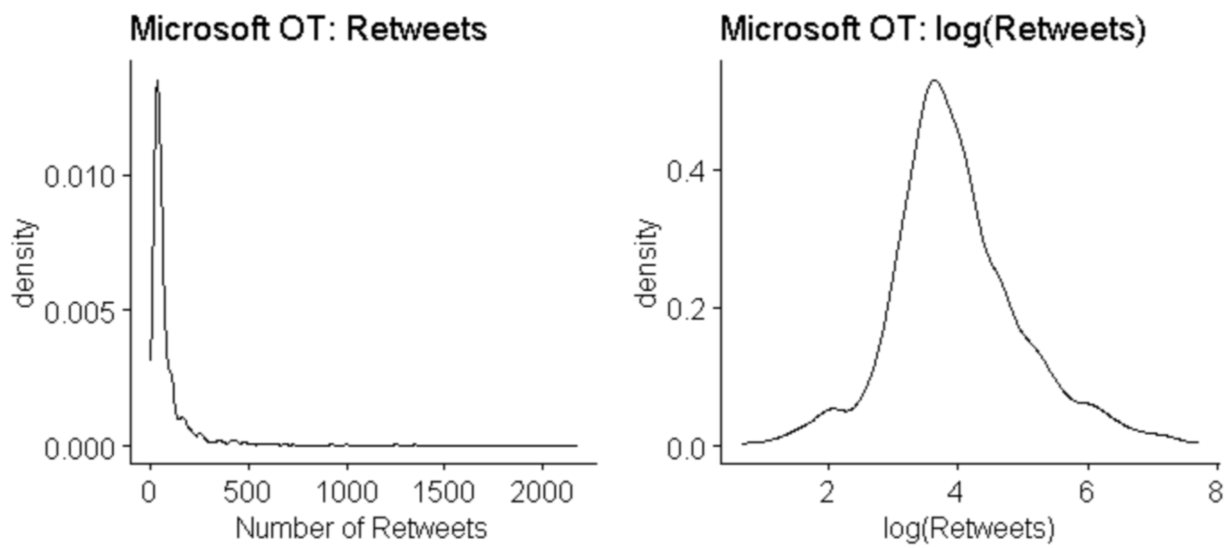



```
$`Neutral - Positive`
    q   Neutral Positive   difference   ci_lower    ci_upper       p_crit p_value
1 0.1  110.1826 110.4469   -0.2643342 -20.297158   24.45618 0.0009259259   0.959
2 0.2  140.9254 138.2131    2.7122800 -16.604558   23.70874 0.0004629630   0.686
3 0.3  169.9088 157.3528   12.5560584 -13.288851   43.23526 0.0003086420   0.114
4 0.4  207.3860 181.9140   25.4719437  -6.611908   69.55041 0.0002314815   0.013
5 0.5  261.6036 211.6855   49.9180928   4.057676   98.09271 0.0001851852   0.000
6 0.6  331.6240 245.0693   86.5546426   5.494829  238.09339 0.0001543210   0.000
7 0.7  556.2765 314.8933  241.3832766  17.789283  561.02243 0.0001322751   0.000
8 0.8 1135.9174 448.0099  687.9074850 235.979514 1433.13884 0.0001157407   0.000
9 0.9 2503.3761 857.7311 1645.6449212 833.543000 3156.58677 0.0001028807   0.000
```

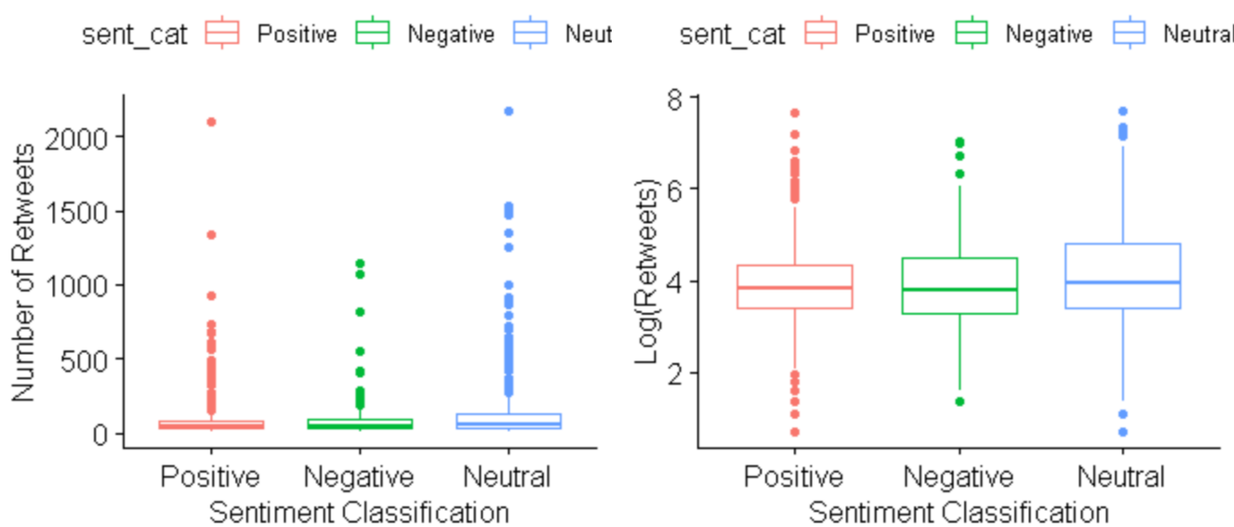
We can say, with 95% confidence, that the 5^{th} through 9^{th} quantiles of positive tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of neutral tweets.

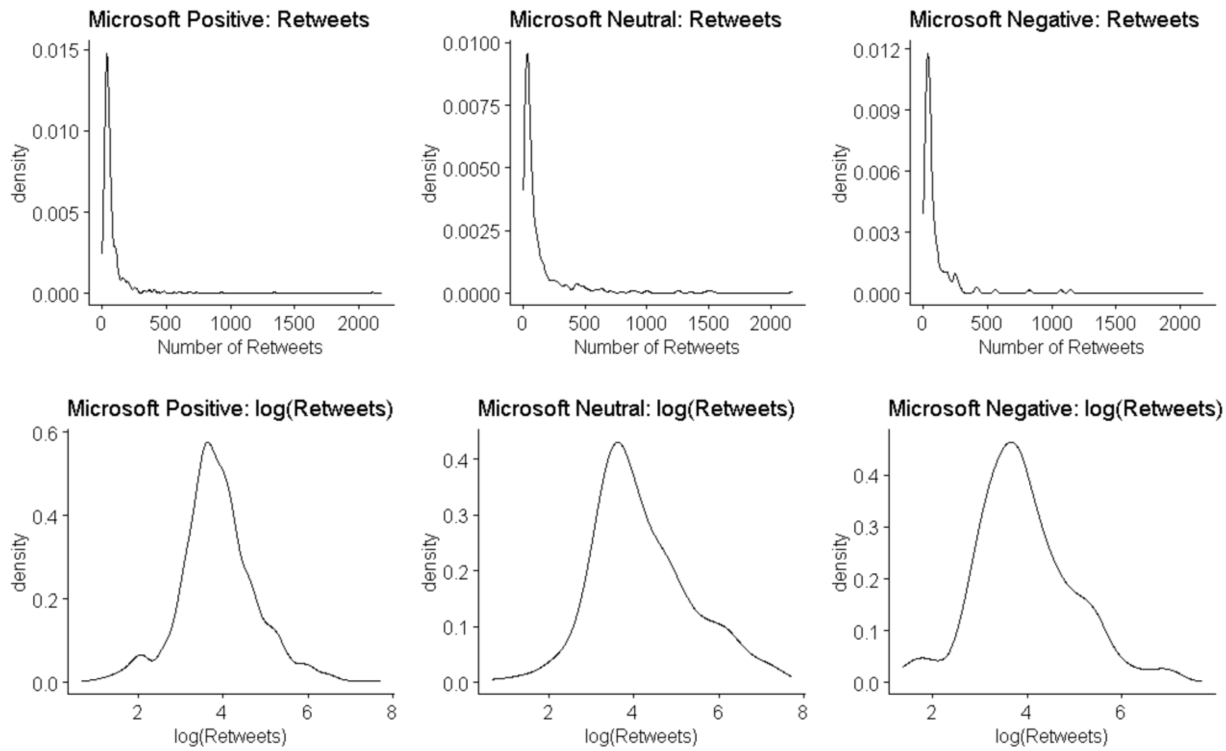
Therefore, we may conclude that statistically significant differences exist between the quantiles of the above group pairing (in their distributions of likes, and the right halves specifically), and potentially one underlying factor explaining these differences is the sentiment of Microsoft official tweets (positive sentiment behaving differently than neutral).

Microsoft Official: Number of Retweets



The log distribution does not pass a Shapiro-Wilk normality test.





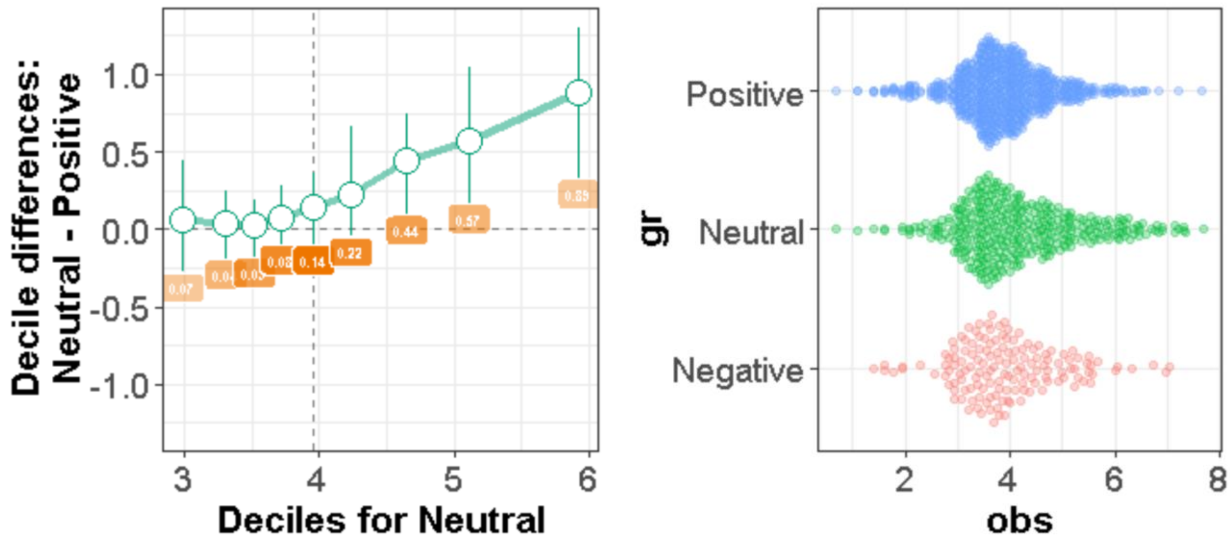
I'm unable to get the x-axes above to align perfectly in the log distributions, but note the differences. Furthermore, none of the above log distributions pass a Shapiro-Wilk normality test.

```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 11.275, df = 2, p-value = 0.003562
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z       P.unadj        P.adj
1  Negative - Neutral -2.0050594 0.0449566933 0.134870080
2 Negative - Positive  0.1554082 0.8764994844 1.000000000
3  Neutral - Positive  3.2930208 0.0009911713 0.002973514
```

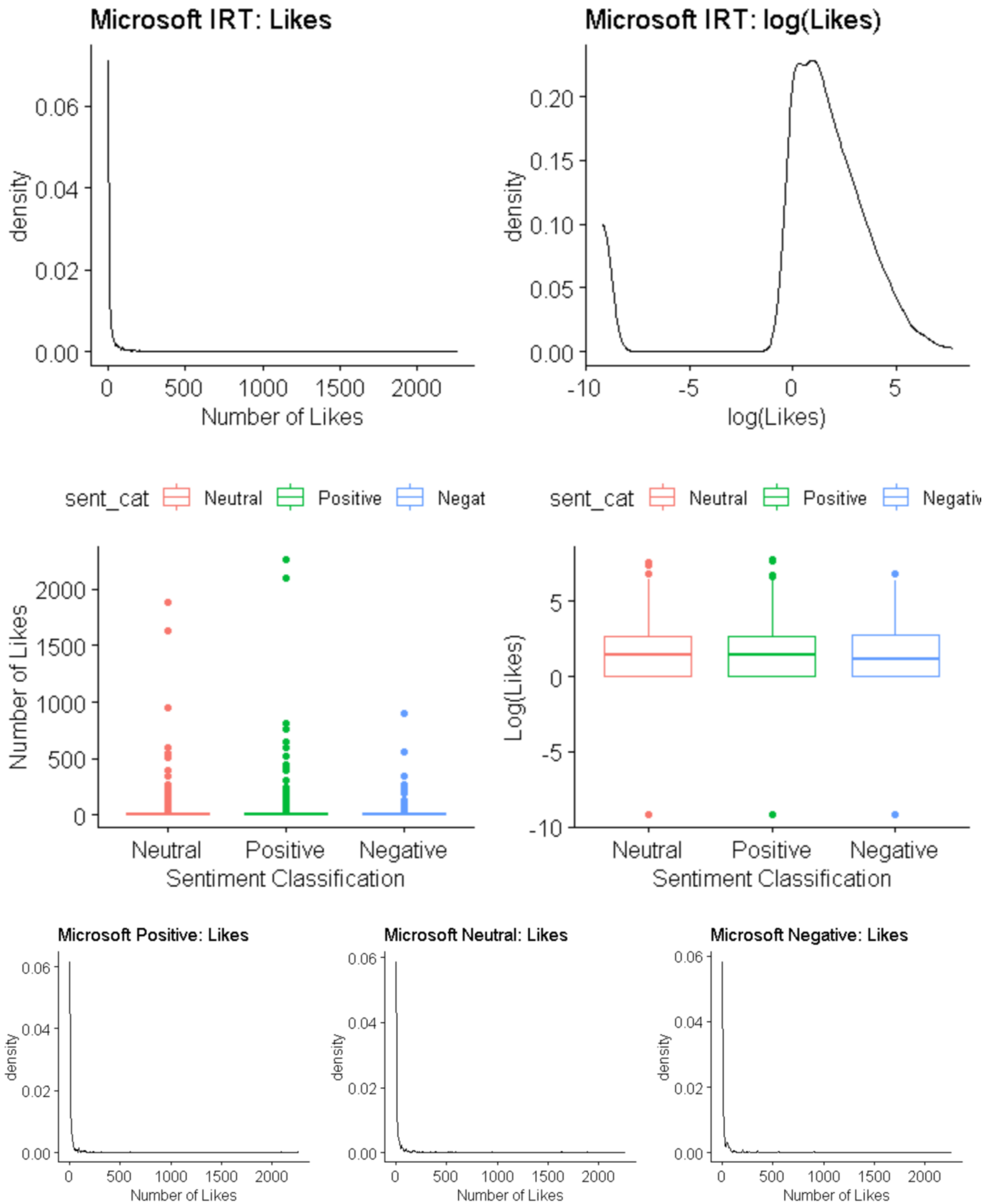
From the results of Dunn's test, we can see that we may reject the null for only the (Neutral, Positive) pairing. In other words, the distribution of retweets for positive tweets differs from the distributions of retweets for neutral tweets. Performing a shift function to further examine these differences yields the following:

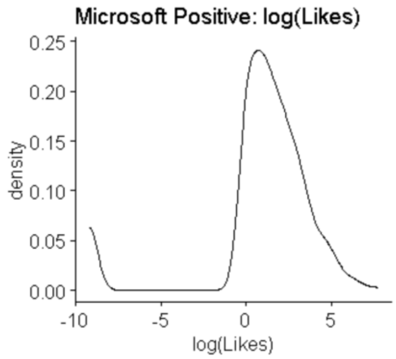


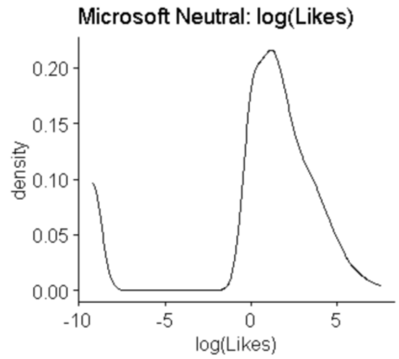

```
$`Neutral - Positive`
    q   Neutral  Positive difference  ci_lower   ci_upper       p_crit p_value
1 0.1  19.90823  18.67784   1.230392 -4.208483   6.954760 2.057613e-04   0.475
2 0.2  27.34279  26.19132   1.151470 -3.816161   6.326403 6.172840e-04   0.556
3 0.3  33.72140  32.68053   1.040868 -4.431214   7.289412 3.086420e-04   0.518
4 0.4  41.21770  38.02197   3.195723 -3.753055  11.197031 1.543210e-04   0.160
5 0.5  52.51672  45.44273   7.073987 -4.437209  21.712289 1.234568e-04   0.044
6 0.6  69.49983  55.38987  14.109953 -2.019535  39.089533 1.028807e-04   0.005
7 0.7 104.77346  67.06865  37.704805  7.931069  74.265612 8.818342e-05   0.000
8 0.8 166.56218  93.58954  72.972638 24.181248 168.938851 7.716049e-05   0.000
9 0.9 377.95514 154.17074 223.784406 39.383208 420.432599 6.858711e-05   0.000
```

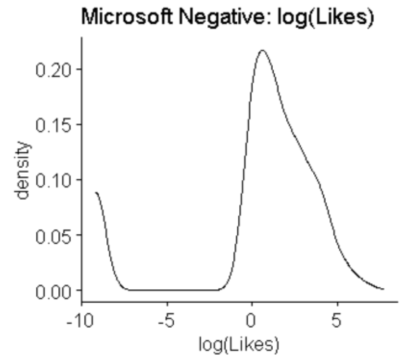
We can say, with 95% confidence, that the 7th through 9th quantiles of positive tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of neutral tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above group pairing (in their distributions of retweets, and the right-tails specifically), and potentially one underlying factor explaining these differences is the sentiment of Microsoft official tweets (positive sentiment behaving differently than neutral).**

Microsoft IRT: Number of Likes





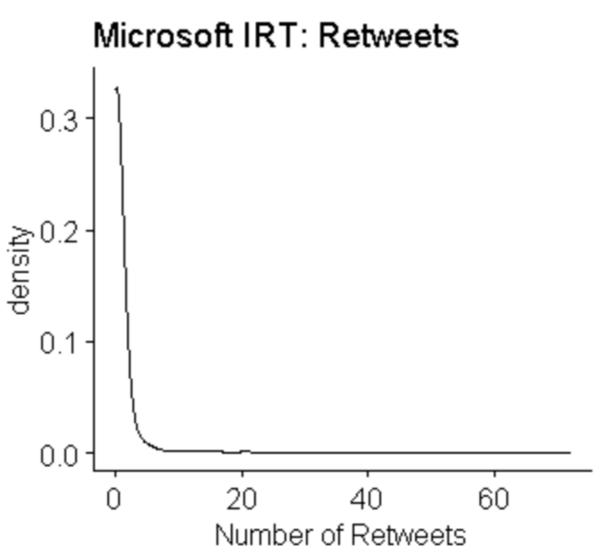


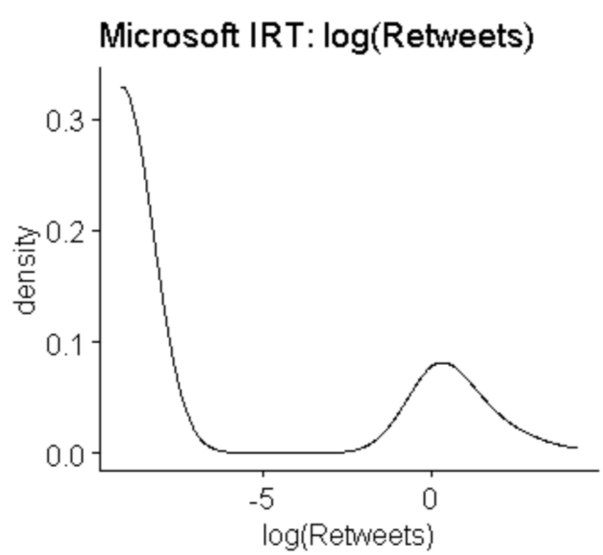


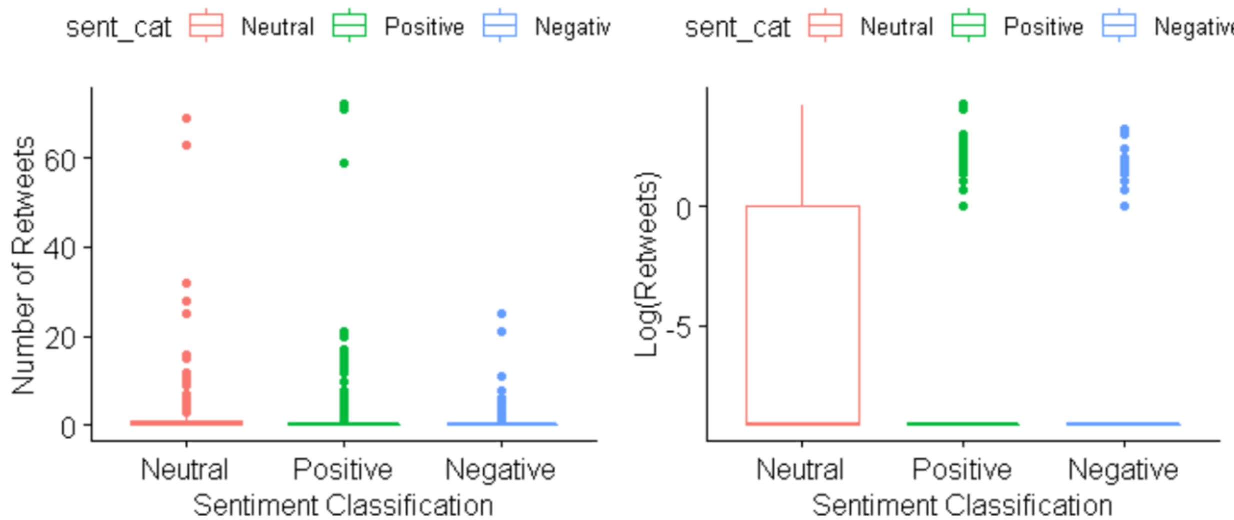

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 1.7774, df = 2, p-value = 0.4112
```

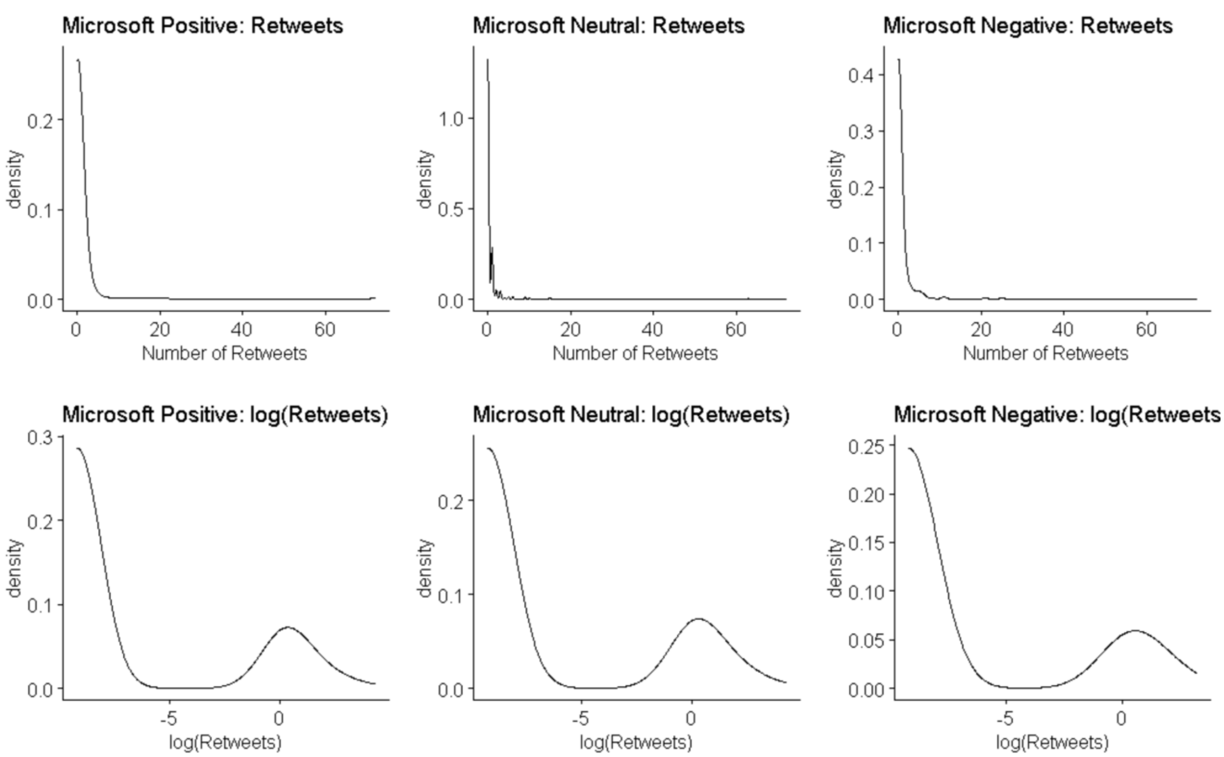
Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the ‘like’ distributions of all populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of likes which a Microsoft IRT tweet receives.**

Microsoft IRT: Number of Retweets





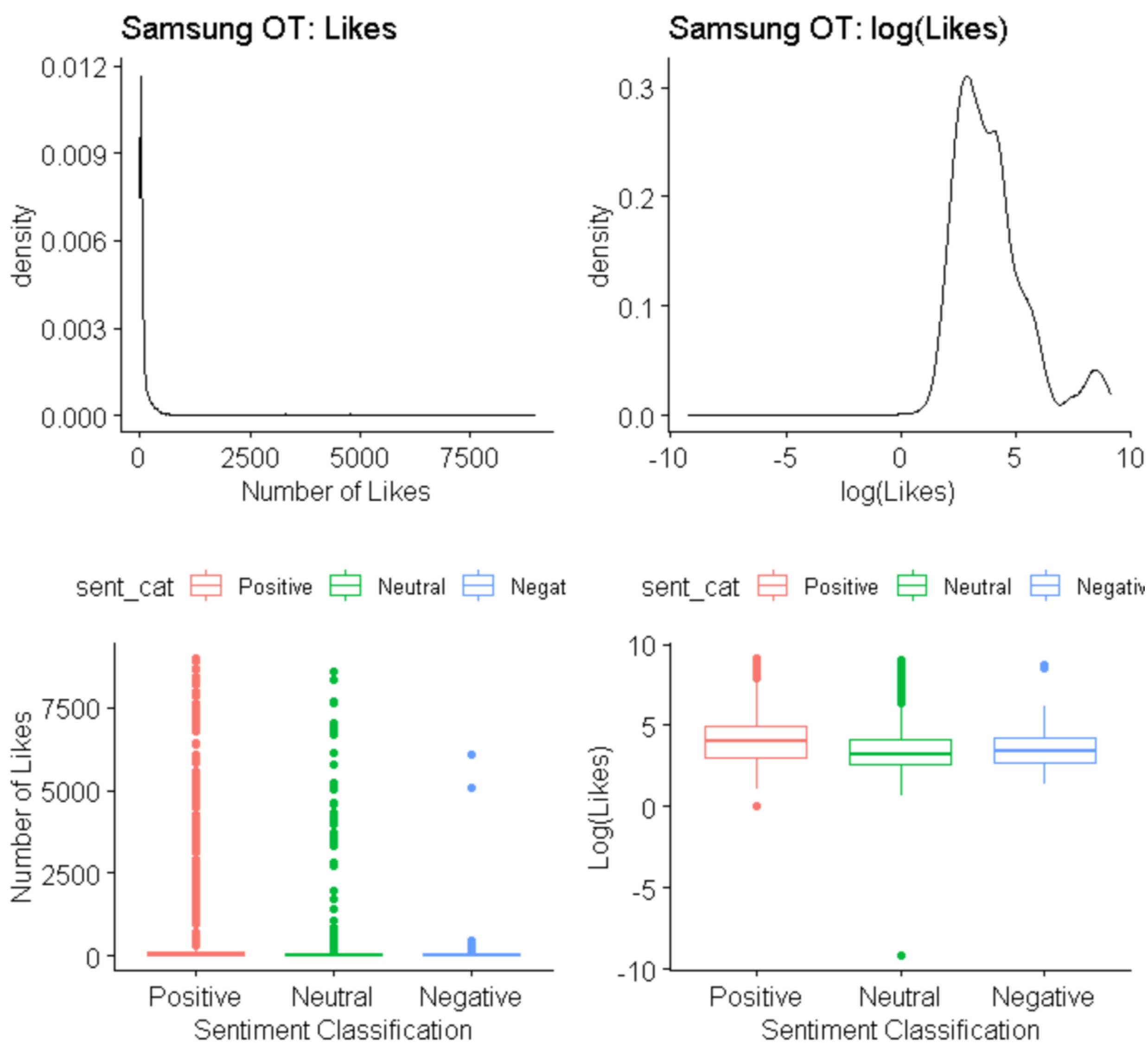


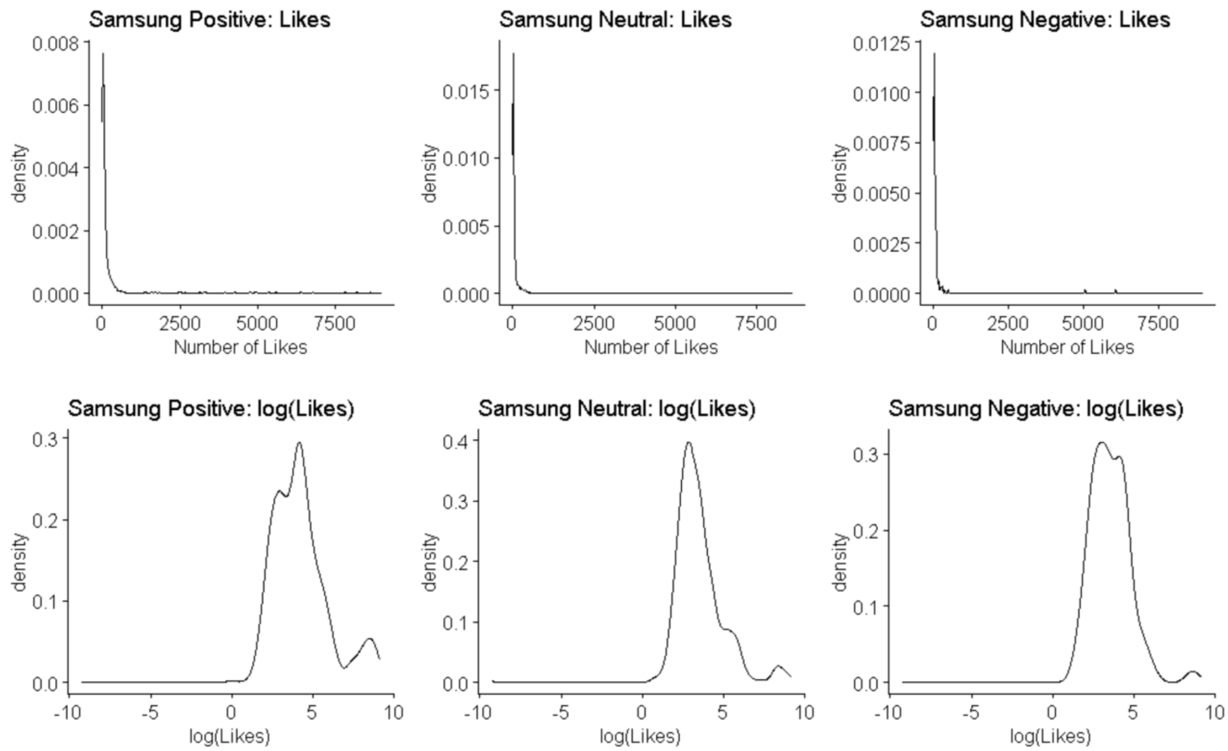



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 1.0152, df = 2, p-value = 0.602
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of all populations are equal. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a Microsoft IRT tweet receives.**

Samsung Official: Number of Likes





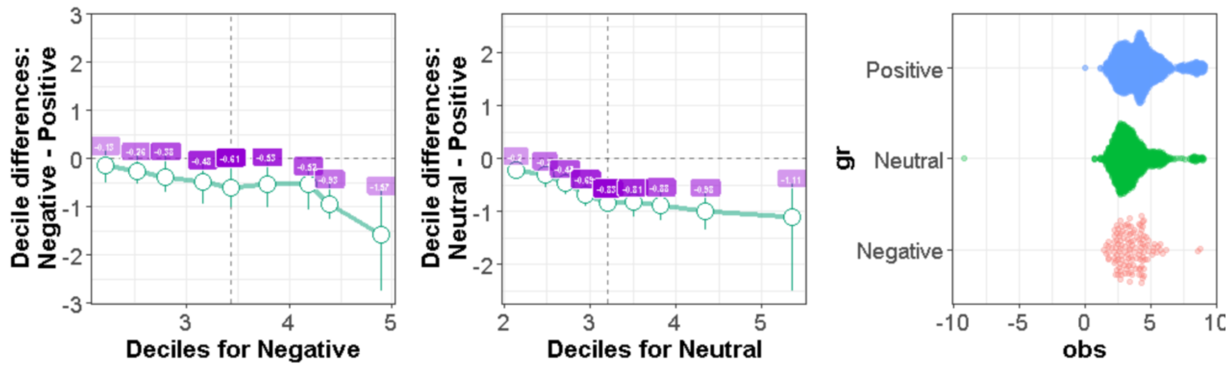
None of the above distributions pass a Shapiro-Wilk normality test.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 158.48, df = 2, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison          Z       P.unadj         P.adj
1  Negative - Neutral   1.088434 2.764037e-01 8.292111e-01
2 Negative - Positive  -4.148966 3.339803e-05 1.001941e-04
3  Neutral - Positive -12.461416 1.212000e-35 3.636000e-35
```

From the results of Dunn's test, we can see that we may reject the null for both the (Negative, Positive) and the (Neutral, Positive) pairs. In other words, the distribution of likes for positive tweets differs from the distributions of likes for other sentiment categories. Performing a shift function to further examine these differences yields the following:



$`Negative - Positive`

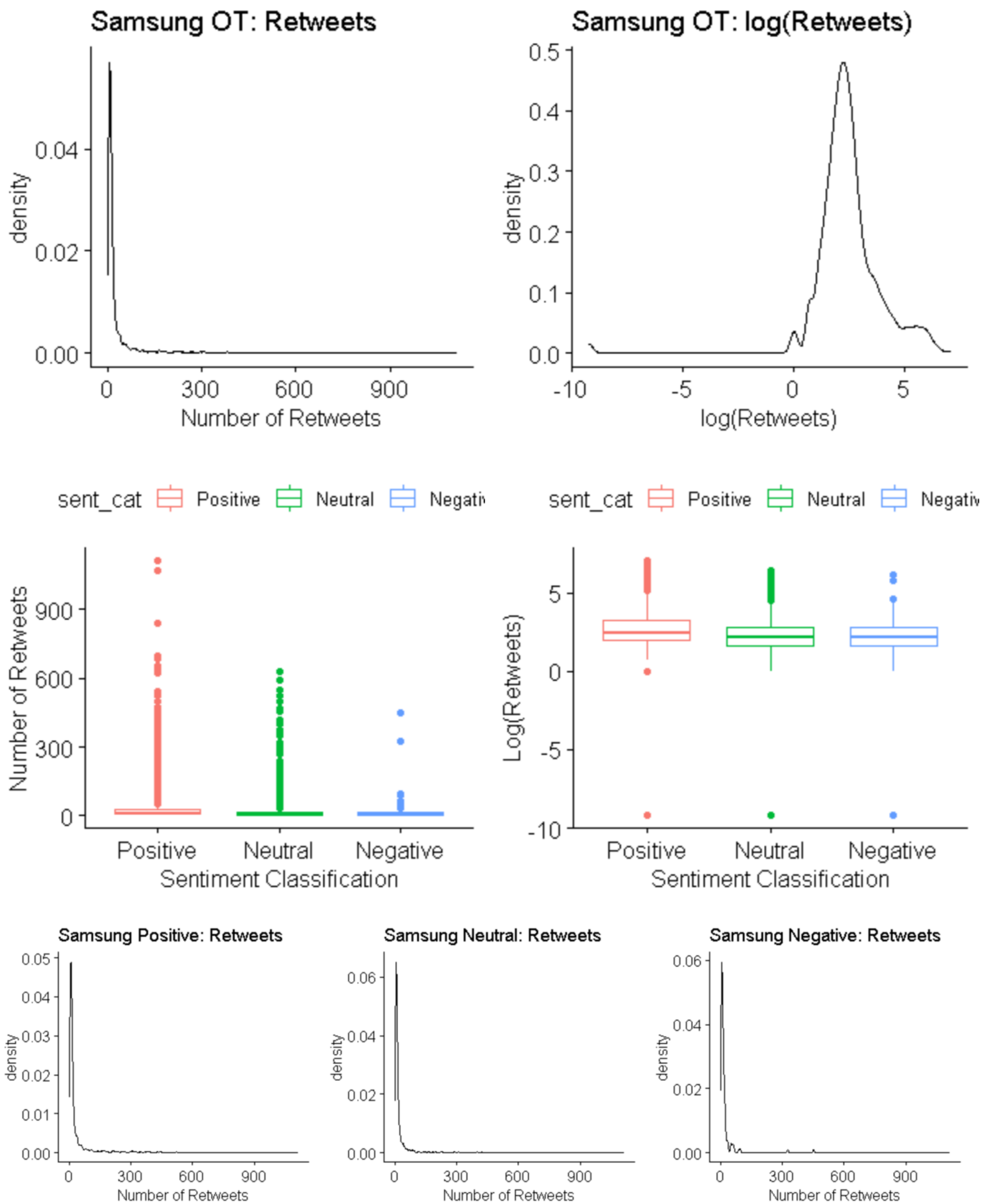
	q	Negative	Positive	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	9.155644	10.40005	-1.244406	-4.290241	1.356536	0.0083333333	0.200
2	0.2	12.423160	16.02277	-3.599610	-6.942458	1.558549	0.0041666667	0.022
3	0.3	16.526732	23.94518	-7.418450	-12.995148	1.513941	0.0027777778	0.011
4	0.4	23.710730	38.05845	-14.347718	-25.077174	-3.981005	0.0016666667	0.000
5	0.5	31.032762	56.90288	-25.870120	-37.362548	-8.231088	0.0013888889	0.000
6	0.6	45.163309	74.85324	-29.689927	-49.800974	-3.439788	0.0011904762	0.000
7	0.7	65.971310	110.92479	-44.953481	-84.638399	-20.756752	0.0010416667	0.000
8	0.8	81.127449	206.54927	-125.421826	-185.660036	-76.929976	0.0009259259	0.000
9	0.9	139.097903	678.44912	-539.351221	-1556.356716	-115.932518	0.0020833333	0.002

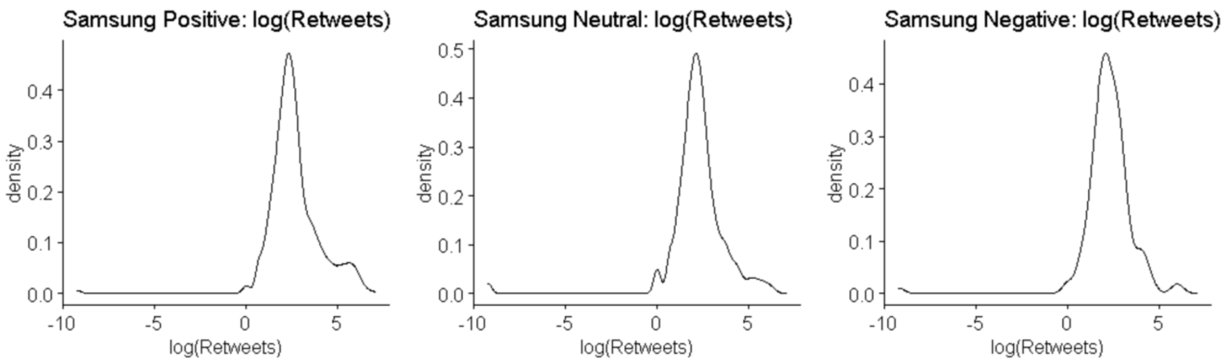
$`Neutral - Positive`

	q	Neutral	Positive	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	8.493774	10.40005	-1.906276	-3.361015	-0.3272804	0.0020833333	0
2	0.2	11.886974	16.02277	-4.135795	-6.742430	-2.1507612	0.0010416667	0
3	0.3	15.009429	23.94518	-8.935753	-14.546812	-5.6613538	0.0006944444	0
4	0.4	19.140884	38.05845	-18.917564	-27.713558	-12.9334073	0.0005208333	0
5	0.5	24.768451	56.90288	-32.134431	-39.552574	-24.5212452	0.0004166667	0
6	0.6	33.251627	74.85324	-41.601608	-52.729541	-31.9072214	0.0003472222	0
7	0.7	45.927738	110.92479	-64.997053	-88.505972	-44.9852208	0.0002976190	0
8	0.8	77.156285	206.54927	-129.392990	-200.041130	-72.6260407	0.0002604167	0
9	0.9	213.241133	678.44912	-465.207991	-1680.644235	-156.2109591	0.0002314815	0

From the top table we can say, with 95% confidence, that the 4th through 9th quantiles of positive tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of negative tweets. From the bottom table we can say, with 95% confidence, that every quantile of positive tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of neutral tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes), and potentially one underlying factor explaining these differences is the sentiment of Samsung official tweets (positive sentiment behaving differently).**

Samsung Official: Number of Retweets



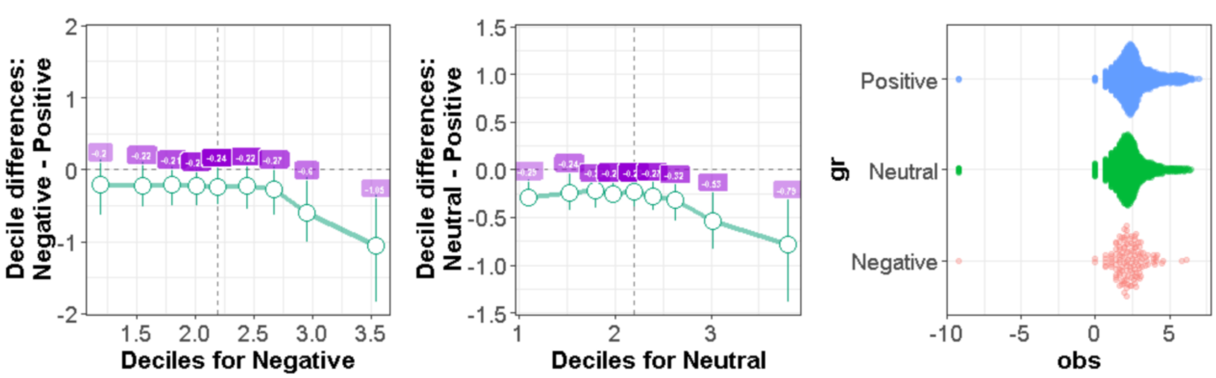



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 59.668, df = 2, p-value = 1.105e-13
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison           Z       P.unadj        P.adj
1  Negative - Neutral  0.05235443 9.582463e-01 1.000000e+00
2 Negative - Positive -3.11904101 1.814407e-03 5.443221e-03
3  Neutral - Positive -7.53477838 4.891649e-14 1.467495e-13
```

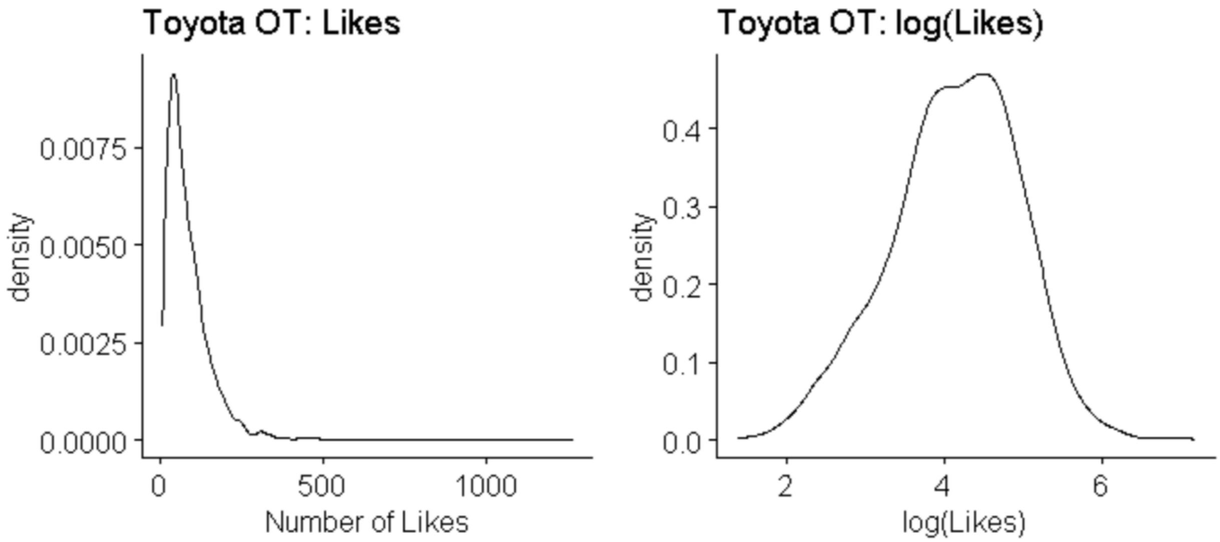
From the results of Dunn’s test, we can see that we may reject the null for both the (Negative, Positive) and the (Neutral, Positive) pairs. In other words, the distribution of retweets for positive tweets differs from the distributions of retweets for other sentiment categories. Performing a shift function to further examine these differences yields the following:



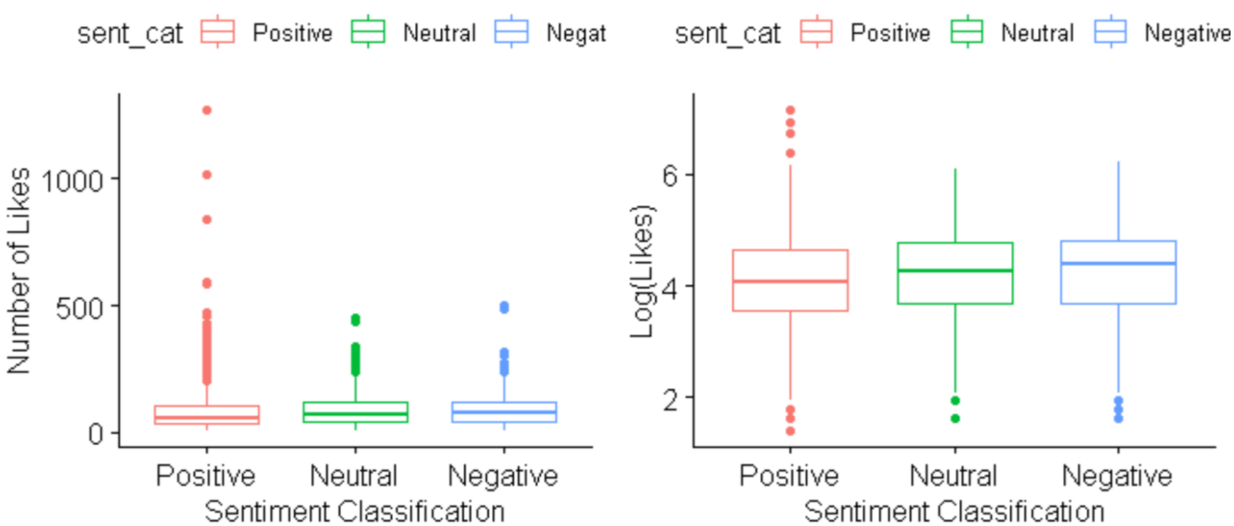
```
$`Negative - Positive`
    q  Negative  Positive  difference    ci_lower    ci_upper       p_crit p_value
1 0.1  3.321290  3.994179  -0.6728897   -1.850636   0.5713542 0.0055555556   0.157
2 0.2  4.705977  5.856812  -1.1508346   -2.273321   0.4882015 0.0027777778   0.031
3 0.3  6.094179  7.499267  -1.4050880   -3.108110   0.4139417 0.0013888889   0.019
4 0.4  7.445526  9.331018  -1.8854920   -3.943261   0.6071190 0.0007936508   0.006
5 0.5  8.934236 11.293095  -2.3588590   -4.512210   0.9398433 0.0018518519   0.023
6 0.6 11.520772 14.333157  -2.8123848   -6.136908   1.3152186 0.0011111111   0.032
7 0.7 14.516450 19.031542  -4.5150917   -9.719719   1.6090201 0.0009259259   0.007
8 0.8 19.196349 34.962634 -15.7662845  -27.014663  -1.0319981 0.0006944444   0.001
9 0.9 36.167493 99.362544 -63.1950509 -111.318631 -16.0063576 0.0006172840   0.001

$`Neutral - Positive`
    q   Neutral  Positive  difference    ci_lower     ci_upper       p_crit p_value
1 0.1  3.000493  3.994179  -0.9936859   -1.262102  -0.27738967 6.172840e-04       0
2 0.2  4.611809  5.856812  -1.2450023   -2.211995  -0.07731731 3.086420e-04       0
3 0.3  5.992818  7.499267  -1.5064484   -2.831997  -0.73147204 2.057613e-04       0
4 0.4  7.215744  9.331018  -2.1152737   -3.347384  -1.04578693 1.543210e-04       0
5 0.5  8.968753 11.293095  -2.3243425   -3.684423  -1.35293623 1.234568e-04       0
6 0.6 10.915802 14.333157  -3.4173550   -5.240000  -1.28812666 1.028807e-04       0
7 0.7 13.832291 19.031542  -5.1992509   -9.530429  -2.42705497 8.818342e-05       0
8 0.8 20.535681 34.962634 -14.4269533  -25.927158  -5.56524246 7.716049e-05       0
9 0.9 45.007563 99.362544 -54.3549807 -105.409089 -20.95693788 6.858711e-05       0
```

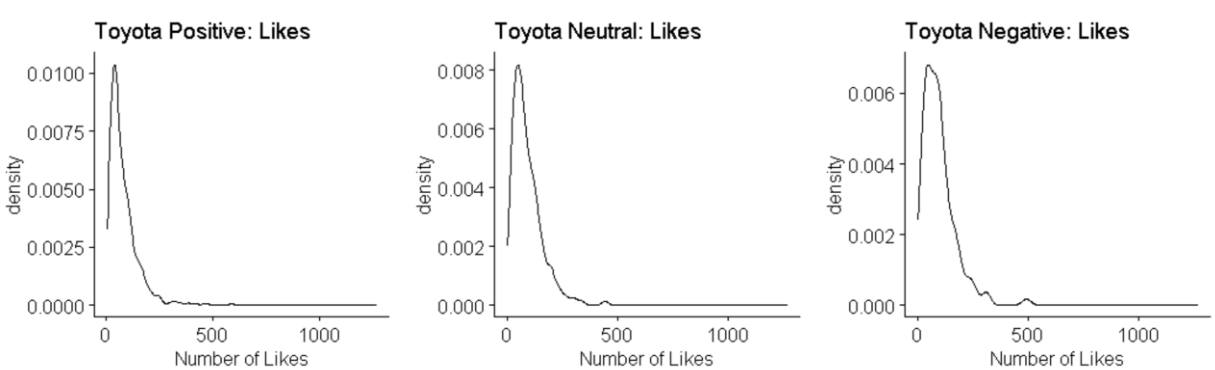
From the top table we can say, with 95% confidence, that the 8th and 9th quantiles of positive tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of negative tweets. From the bottom table we can say, with 95% confidence, that every quantile of positive tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of neutral tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of retweets), and potentially one underlying factor explaining these differences is the sentiment of Samsung official tweets (positive sentiment behaving differently).**

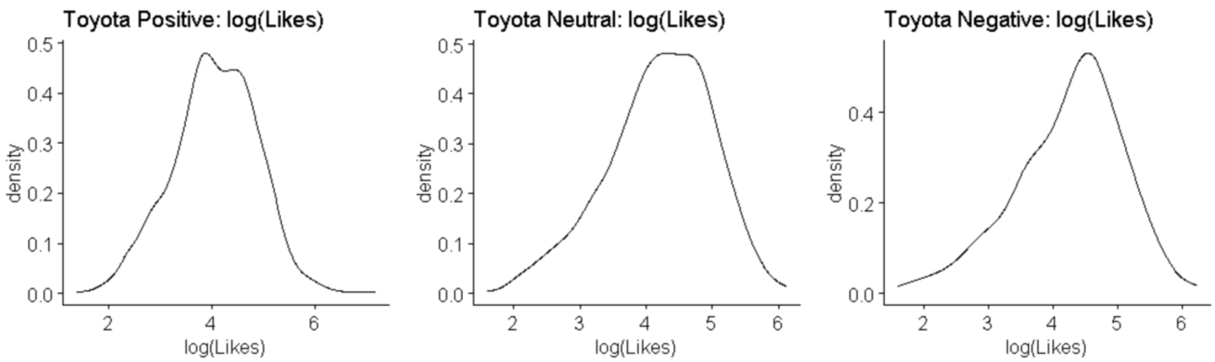
Toyota Official: Number of Likes



The log distribution does not pass a Shapiro-Wilk normality test.







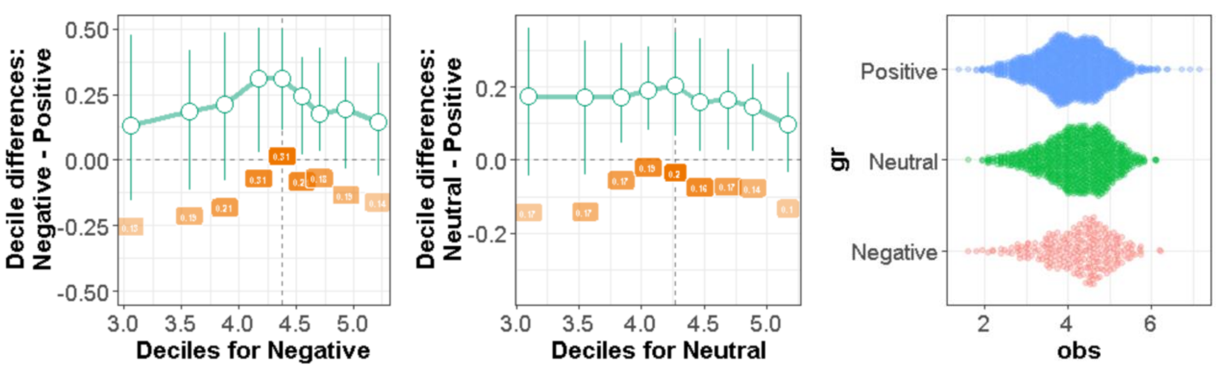
None of the above distributions pass a Shapiro-Wilk normality test.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by sent_cat
Kruskal-Wallis chi-squared = 29.322, df = 2, p-value = 4.294e-07
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison         Z       P.unadj        P.adj
1  Negative - Neutral 0.9791732 3.274944e-01 9.824833e-01
2 Negative - Positive 3.8645393 1.112991e-04 3.338972e-04
3  Neutral - Positive 4.6011674 4.201296e-06 1.260389e-05
```

From the results of Dunn's test, we can see that we may reject the null for both the (Negative, Positive) and the (Neutral, Positive) pairs. In other words, the distribution of likes for positive tweets differs from the distributions of likes for other sentiment categories. Performing a shift function to further examine these differences yields the following:



$`Negative - Positive`

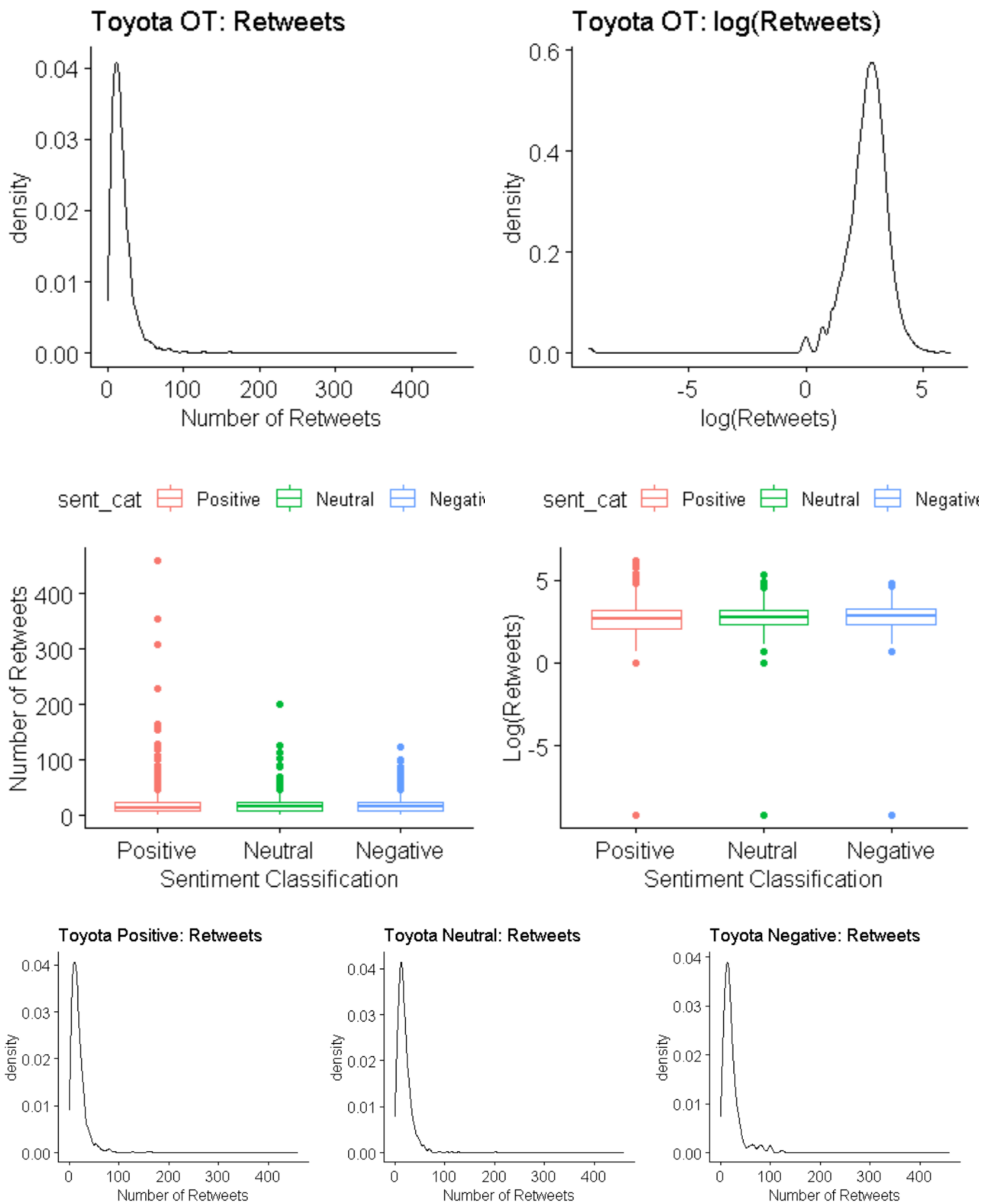
	q	Negative	Positive	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	21.38114	18.63194	2.749195	-3.114778	11.37460	0.010000000	0.247
2	0.2	35.46703	29.36828	6.098752	-2.561370	15.74105	0.005000000	0.035
3	0.3	48.55353	39.07162	9.481902	-2.402955	23.94710	0.002500000	0.016
4	0.4	65.05068	47.59847	17.452209	2.334596	30.96260	0.001666667	0.000
5	0.5	79.95044	58.35372	21.596718	5.749718	35.96468	0.001428571	0.000
6	0.6	94.64758	74.12008	20.527502	7.116225	36.54535	0.001250000	0.000
7	0.7	110.21571	92.23663	17.979077	1.524336	40.96477	0.001111111	0.000
8	0.8	139.05581	114.51084	24.544976	-3.020513	51.88808	0.002000000	0.006
9	0.9	183.67022	158.88805	24.782165	-6.098862	71.29811	0.003333333	0.023

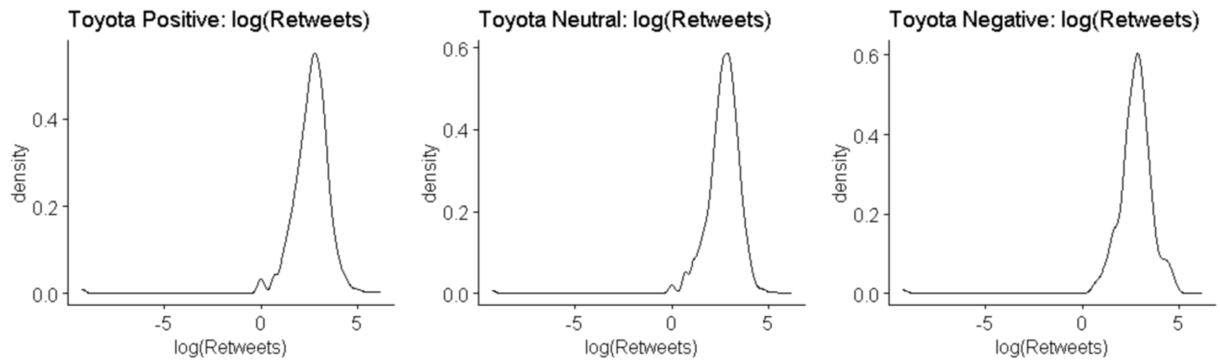
$`Neutral - Positive`

	q	Neutral	Positive	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	22.20992	18.63194	3.577983	-0.9887428	7.826644	0.0016666667	0.023
2	0.2	34.83772	29.36828	5.469439	-0.9093086	10.787694	0.0011111111	0.008
3	0.3	46.41774	39.07162	7.346120	1.6806660	14.654637	0.0008333333	0.000
4	0.4	57.60534	47.59847	10.006878	3.3271377	17.117762	0.0006666667	0.000
5	0.5	71.50298	58.35372	13.149257	4.0591686	22.214115	0.0005555556	0.000
6	0.6	86.85579	74.12008	12.735717	2.2559280	27.414723	0.0004761905	0.000
7	0.7	108.80661	92.23663	16.569979	2.1371791	27.729168	0.0004166667	0.000
8	0.8	132.27728	114.51084	17.766448	1.6245307	31.267253	0.0003703704	0.000
9	0.9	175.35073	158.88805	16.462683	-6.7817549	38.990632	0.0033333333	0.041

From the top table we can say, with 95% confidence, that the 4^{th} through 7^{th} quantiles of positive tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of negative tweets. From the bottom table we can say, with 95% confidence, that the 3^{rd} through 8^{th} quantiles of positive tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of neutral tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes, and the mid quantiles specifically), and potentially one underlying factor explaining these differences is the sentiment of Toyota official tweets (positive sentiment behaving differently).**

Toyota Official: Number of Retweets



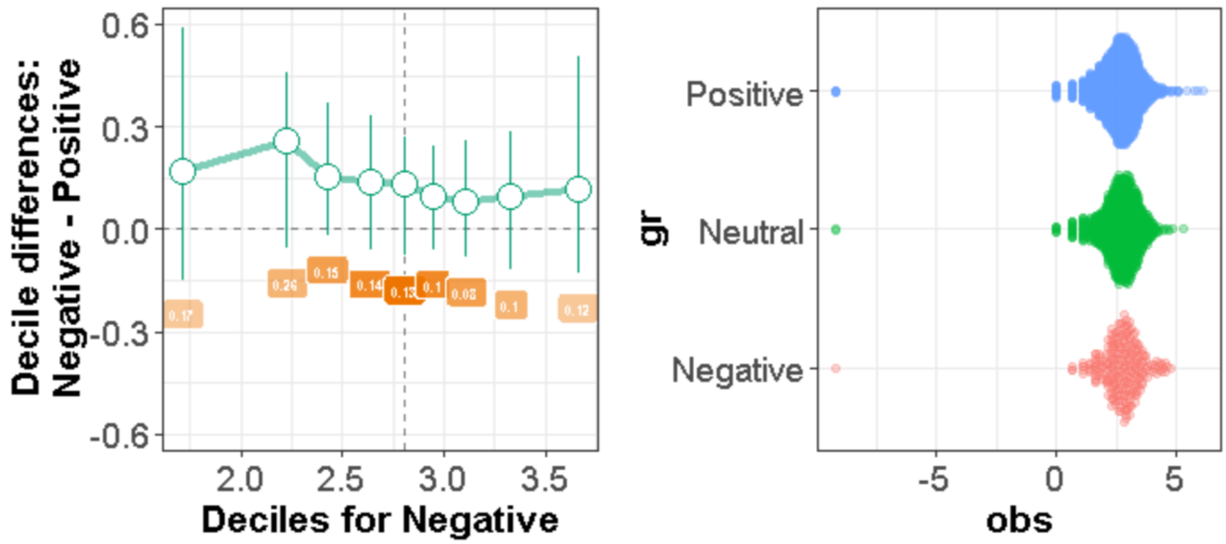



```
         Kruskal-Wallis rank sum test

data:  Number of Retweets by sent_cat
Kruskal-Wallis chi-squared = 9.3972, df = 2, p-value = 0.009108
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ factor(sent_cat), data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

           Comparison        Z      P.unadj      P.adj
1  Negative - Neutral 1.159241 0.246358036 0.73907411
2 Negative - Positive 2.580692 0.009860251 0.02958075
3  Neutral - Positive 2.219140 0.026477171 0.07943151
```

From the results of Dunn's test, we can see that we may reject the null for only the (Negative, Positive) pairing. In other words, the distribution of retweets for positive tweets differs from the distributions of retweets for negative, but not neutral, tweet categories. Performing a shift function to further examine these differences yields the following:



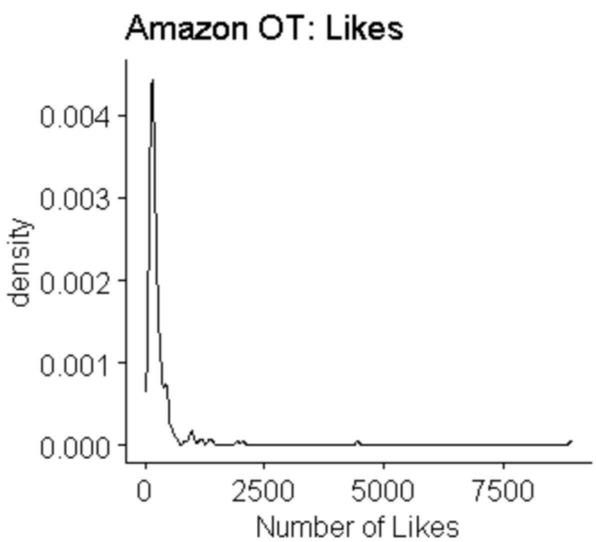

```
$`Negative - Positive`
    q  Negative  Positive difference    ci_lower  ci_upper       p_crit p_value
1 0.1  5.585462  4.697369  0.8880933 -0.64532576  3.622052 0.0013888889   0.054
2 0.2  9.298454  7.168265  2.1301893 -0.51097777  3.973181 0.0006944444   0.008
3 0.3 11.319046  9.697833  1.6212131 -0.06286769  4.142291 0.0006172840   0.002
4 0.4 13.941239 12.110592  1.8306472 -1.62196941  4.517480 0.0009259259   0.036
5 0.5 16.570509 14.500227  2.0702822 -0.78531539  4.241204 0.0007936508   0.029
6 0.6 19.002307 17.260754  1.7415532 -0.73444513  4.905570 0.0011111111   0.043
7 0.7 22.422332 20.714395  1.7079373 -1.93107510  6.514673 0.0018518519   0.143
8 0.8 27.940320 25.315481  2.6248398 -2.03185506  8.005857 0.0027777778   0.145
9 0.9 39.142090 34.634791  4.5072986 -3.75830617 23.726015 0.0055555556   0.180
```

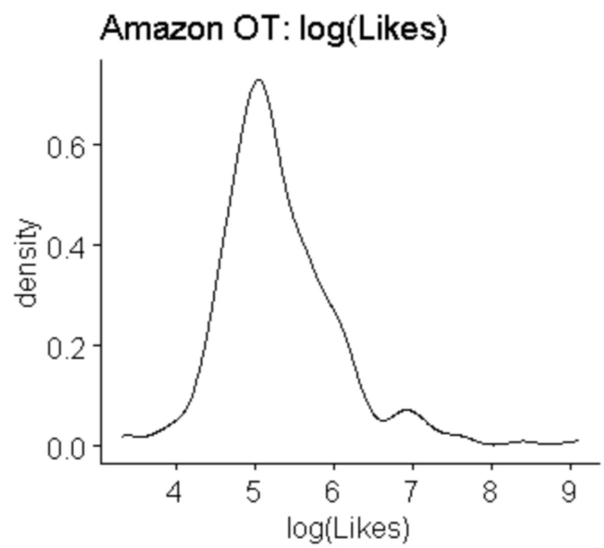
Considering each confidence interval contains the value 0, we fail to reject the null that all quantiles are equivalent. **Sentiment does not seem to have a statistically significant effect on the number of retweets which a Toyota official tweet receives.**

<u>Section 10:</u> BTM Topic Analysis

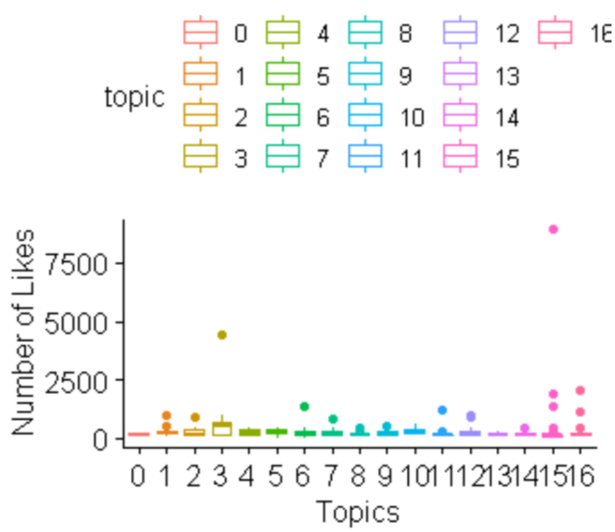
Section 10 contains the visualizations, notes, results, and conclusions made when examining companies and tweet categories for the potential that BTM topic modelling uncovers subject matter or underlying themes which help further explain the number of likes or retweets a tweet receives.

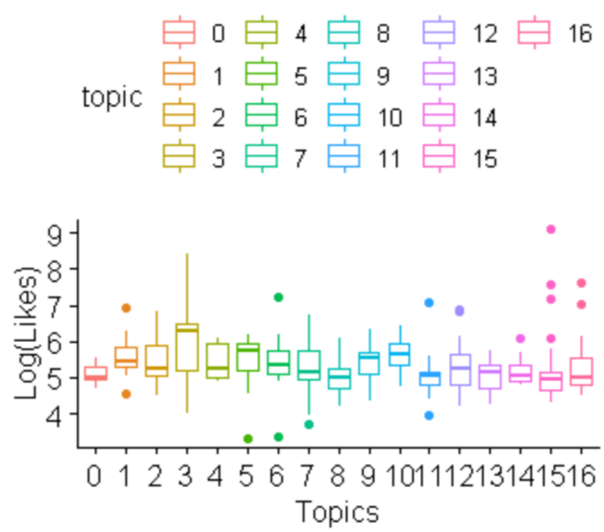
Amazon Official Topics: Number of Likes





The log distribution does not pass a Shapiro-Wilk normality test.






```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 28.52, df = 16, p-value = 0.02738
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

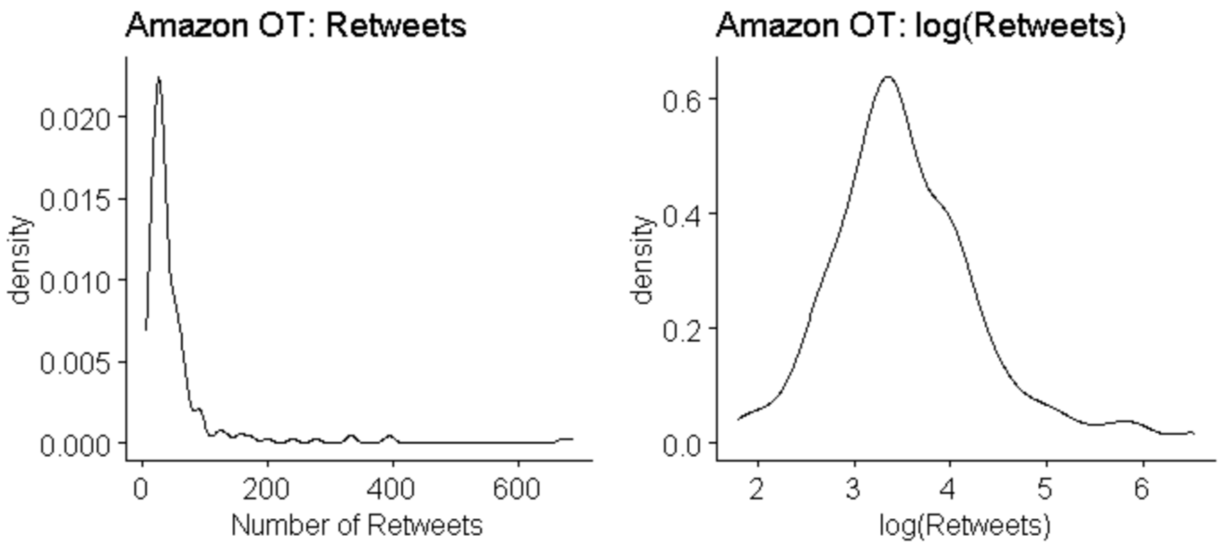
```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

    Comparison           Z      P.unadj     P.adj
1        0 - 1 -2.13113453 0.033078061 1.0000000
2       0 - 10 -1.73272186 0.083145117 1.0000000
3       1 - 10 -0.05274947 0.957931522 1.0000000
4       0 - 11  0.29451453 0.768364764 1.0000000
5       1 - 11  3.02948516 0.002449709 0.3331605
6      10 - 11  2.22626837 0.025996211 1.0000000
7       0 - 12 -0.78199659 0.434216575 1.0000000
8       1 - 12  1.62384281 0.104409362 1.0000000
9      10 - 12  1.24428995 0.213392869 1.0000000
10     11 - 12 -1.34771987 0.177748500 1.0000000
11      0 - 13  0.11119548 0.911461337 1.0000000
12      1 - 13  2.50724522 0.012167627 1.0000000
13     10 - 13  1.94802121 0.051412425 1.0000000
14     11 - 13 -0.19585937 0.844720239 1.0000000
15     12 - 13  1.00581645 0.314503875 1.0000000
16      0 - 14 -0.57843910 0.562967701 1.0000000
17      1 - 14  1.47670539 0.139754628 1.0000000
18     10 - 14  1.21535020 0.224232553 1.0000000
19     11 - 14 -0.97893290 0.327613133 1.0000000
20     12 - 14  0.12052617 0.904066347 1.0000000
21     13 - 14 -0.74002508 0.459284777 1.0000000
22      0 - 15  0.07133811 0.943128663 1.0000000
23      1 - 15  2.81382572 0.004895576 0.6657984
24     10 - 15  2.05331816 0.040041731 1.0000000
25     11 - 15 -0.29583130 0.767358919 1.0000000
26     12 - 15  1.09556326 0.273269949 1.0000000
27     13 - 15 -0.05812571 0.953648498 1.0000000
28     14 - 15  0.76486223 0.444353585 1.0000000
29      0 - 16 -0.54756066 0.583993607 1.0000000
30      1 - 16  1.86625970 0.062005041 1.0000000
31     10 - 16  1.43107360 0.152409122 1.0000000
32     11 - 16 -1.04112130 0.297819255 1.0000000
33     12 - 16  0.27048252 0.786789062 1.0000000
34     13 - 16 -0.74015081 0.459208491 1.0000000
35     14 - 16  0.10686848 0.914893321 1.0000000
36     15 - 16 -0.78698934 0.431288105 1.0000000
37       0 - 2 -1.28311818 0.199450676 1.0000000
38       1 - 2  0.92841419 0.353192763 1.0000000
39      10 - 2  0.74806161 0.454422999 1.0000000
40      11 - 2 -1.92893358 0.053739109 1.0000000
41      12 - 2 -0.62577333 0.531463627 1.0000000
42      13 - 2 -1.54849577 0.121502979 1.0000000
43      14 - 2 -0.64489475 0.518995393 1.0000000
44      15 - 2 -1.69997657 0.089135333 1.0000000
45      16 - 2 -0.87456397 0.381811198 1.0000000
46       0 - 3 -1.85347408 0.063814439 1.0000000
47       1 - 3  0.03416798 0.972743196 1.0000000
```

48	10 - 3	0.07492425	0.940274982	1.0000000
49	11 - 3	-2.48757463	0.012861747	1.0000000
50	12 - 3	-1.33753210	0.181049011	1.0000000
51	13 - 3	-2.12613423	0.033492084	1.0000000
52	14 - 3	-1.27503497	0.202296864	1.0000000
53	15 - 3	-2.29357546	0.021814896	1.0000000
54	16 - 3	-1.54940588	0.121284181	1.0000000
55	2 - 3	-0.76192091	0.446107209	1.0000000
56	0 - 4	-1.28850585	0.197569934	1.0000000
57	1 - 4	0.60476613	0.545334401	1.0000000
58	10 - 4	0.53164905	0.594969086	1.0000000
59	11 - 4	-1.78937270	0.073554814	1.0000000
60	12 - 4	-0.70843228	0.478676853	1.0000000
61	13 - 4	-1.50062561	0.133452425	1.0000000
62	14 - 4	-0.72733749	0.467019235	1.0000000
63	15 - 4	-1.59369554	0.111004245	1.0000000
64	16 - 4	-0.91902853	0.358080649	1.0000000
65	2 - 4	-0.17575118	0.860489425	1.0000000
66	3 - 4	0.50962812	0.610312017	1.0000000
67	0 - 5	-1.59837763	0.109958960	1.0000000
68	1 - 5	0.38095751	0.703234780	1.0000000
69	10 - 5	0.34628493	0.729128596	1.0000000
70	11 - 5	-2.21685084	0.026633286	1.0000000
71	12 - 5	-1.03390907	0.301178673	1.0000000
72	13 - 5	-1.86062776	0.062796762	1.0000000
73	14 - 5	-1.00491163	0.314939403	1.0000000
74	15 - 5	-2.01288848	0.044126363	1.0000000
75	16 - 5	-1.25586053	0.209166543	1.0000000
76	2 - 5	-0.44970587	0.652922535	1.0000000
77	3 - 5	0.30324665	0.761701900	1.0000000
78	4 - 5	-0.22832179	0.819396084	1.0000000
79	0 - 6	-1.40643155	0.159596016	1.0000000
80	1 - 6	0.70125113	0.483146307	1.0000000
81	10 - 6	0.58609317	0.557812906	1.0000000
82	11 - 6	-2.03804798	0.041545135	1.0000000
83	12 - 6	-0.78988160	0.429596916	1.0000000
84	13 - 6	-1.66965171	0.094988293	1.0000000
85	14 - 6	-0.78805414	0.430665038	1.0000000
86	15 - 6	-1.82031854	0.068710510	1.0000000
87	16 - 6	-1.02656961	0.304623170	1.0000000
88	2 - 6	-0.18294481	0.854841323	1.0000000
89	3 - 6	0.57501558	0.565280759	1.0000000
90	4 - 6	0.01297726	0.989645935	1.0000000
91	5 - 6	0.26677404	0.789643147	1.0000000
92	0 - 7	-0.70085496	0.483393536	1.0000000
93	1 - 7	1.71987397	0.085455353	1.0000000
94	10 - 7	1.31473020	0.188600593	1.0000000
95	11 - 7	-1.24519551	0.213059891	1.0000000
96	12 - 7	0.09752011	0.922313364	1.0000000
97	13 - 7	-0.91503016	0.360175796	1.0000000
98	14 - 7	-0.03938454	0.968583808	1.0000000
99	15 - 7	-0.99098459	0.321693106	1.0000000
100	16 - 7	-0.17445135	0.861510758	1.0000000
101	2 - 7	0.71845468	0.472476981	1.0000000
102	3 - 7	1.41867374	0.155994157	1.0000000
103	4 - 7	0.78644837	0.431604857	1.0000000
104	5 - 7	1.11784102	0.263634898	1.0000000

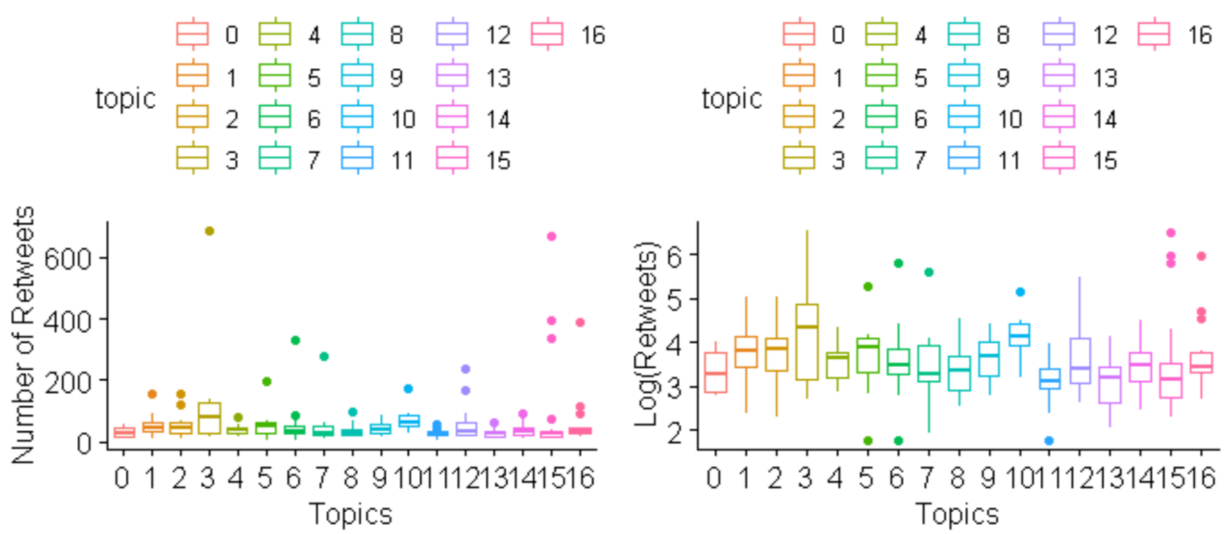
105	6 - 7	0.87859737	0.379619622	1.0000000
106	0 - 8	0.17857137	0.858274280	1.0000000
107	1 - 8	2.72201450	0.006488530	0.8824400
108	10 - 8	2.06309041	0.039104039	1.0000000
109	11 - 8	-0.12937685	0.897059463	1.0000000
110	12 - 8	1.13910654	0.254658721	1.0000000
111	13 - 8	0.07013273	0.944088013	1.0000000
112	14 - 8	0.83300050	0.404844445	1.0000000
113	15 - 8	0.14239629	0.886766989	1.0000000
114	16 - 8	0.85575480	0.392133421	1.0000000
115	2 - 8	1.70130165	0.088886358	1.0000000
116	3 - 8	2.27553791	0.022873681	1.0000000
117	4 - 8	1.61774940	0.105716615	1.0000000
118	5 - 8	2.00640714	0.044812828	1.0000000
119	6 - 8	1.81884460	0.068935141	1.0000000
120	7 - 8	1.04307537	0.296913387	1.0000000
121	0 - 9	-1.64715048	0.099527119	1.0000000
122	1 - 9	0.53832600	0.590352004	1.0000000
123	10 - 9	0.45280373	0.650690075	1.0000000
124	11 - 9	-2.40144629	0.016330406	1.0000000
125	12 - 9	-1.05050220	0.293487281	1.0000000
126	13 - 9	-1.96002577	0.049992778	1.0000000
127	14 - 9	-1.00043590	0.317099602	1.0000000
128	15 - 9	-2.17713117	0.029470782	1.0000000
129	16 - 9	-1.29758599	0.194429640	1.0000000
130	2 - 9	-0.39366822	0.693826011	1.0000000
131	3 - 9	0.42509653	0.670766305	1.0000000
132	4 - 9	-0.15623277	0.875849545	1.0000000
133	5 - 9	0.09774497	0.922134808	1.0000000
134	6 - 9	-0.19189755	0.847822456	1.0000000
135	7 - 9	-1.14492564	0.252239962	1.0000000
136	8 - 9	-2.13942515	0.032401252	1.0000000

No p-values resulting from Dunn’s test are statistically significant. Therefore, for each (i, j) topic group pairing, we fail to reject the null hypothesis that their distributions of likes are equal to one another. **BTM topic modeling does not seem to uncover subject matter or underlying themes within Amazon official tweets which help further explain their expected number of likes.** However, it seems possible that this may be due to the fact that Amazon only has 231 official tweets able to pass through BTM topic modeling, which are then divided into 17 topic groups.

Amazon Official Topics: Number of Retweets



The log distribution does not pass a Shapiro-Wilk normality test.




```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 28.005, df = 16, p-value = 0.03158
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

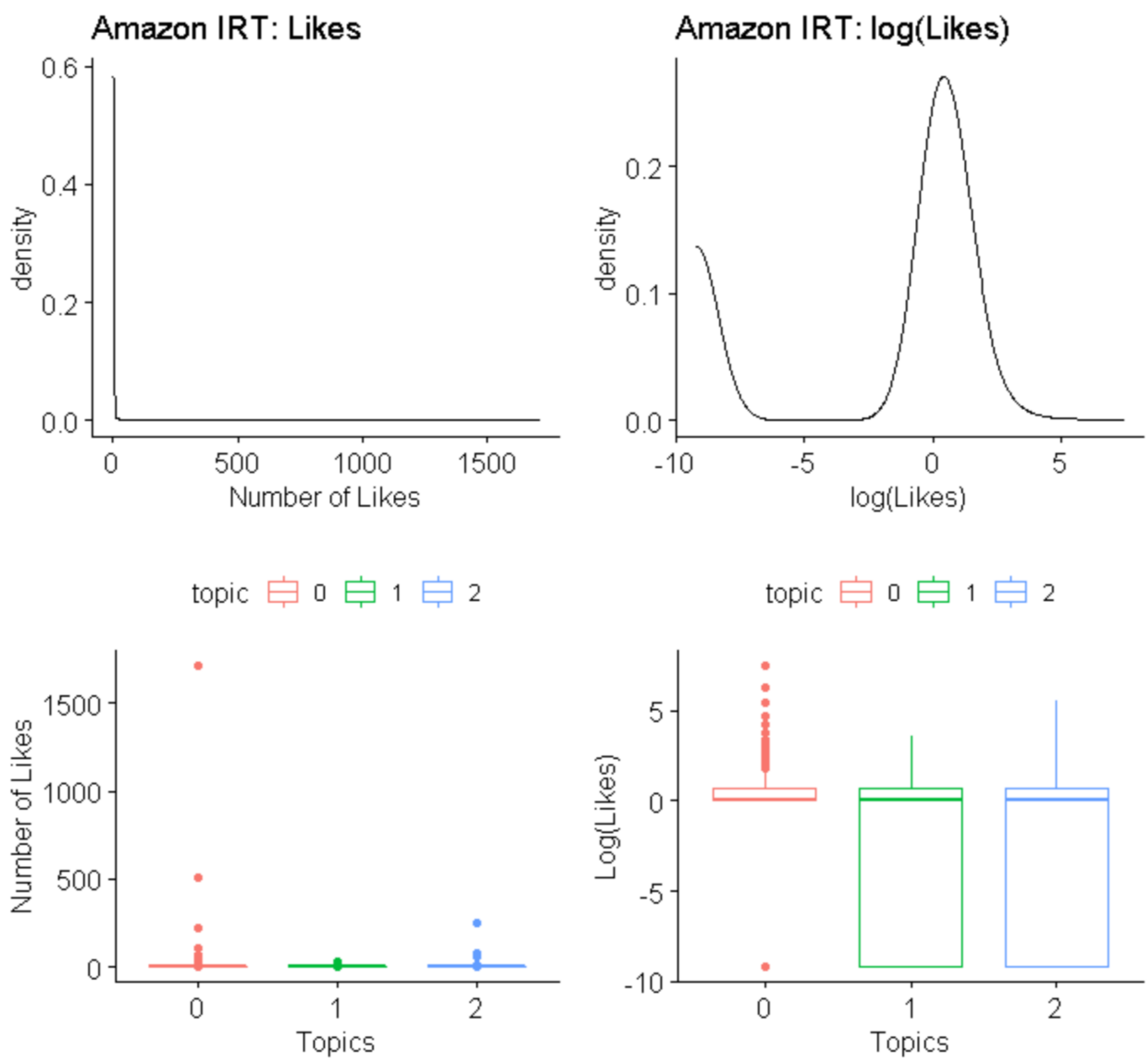
```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni"
)
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

    Comparison            Z      P.unadj     P.adj
1        0 - 1 -1.544497043 0.122467967 1.0000000
2       0 - 10 -2.161432944 0.030661912 1.0000000
3       1 - 10 -1.035347228 0.300506781 1.0000000
4       0 - 11  0.770432674 0.441043277 1.0000000
5       1 - 11  2.864280113 0.004179582 0.5684232
6      10 - 11  3.123982533 0.001784210 0.2426526
7       0 - 12 -0.737312027 0.460932620 1.0000000
8       1 - 12  0.974971152 0.329574570 1.0000000
9      10 - 12  1.758908881 0.078592986 1.0000000
10     11 - 12 -1.872431359 0.061146950 1.0000000
11      0 - 13  0.481605428 0.630086270 1.0000000
12      1 - 13  2.282785918 0.022442982 1.0000000
13     10 - 13  2.731270293 0.006309070 0.8580336
14     11 - 13 -0.278005492 0.781008143 1.0000000
15     12 - 13  1.391770467 0.163991916 1.0000000
16      0 - 14 -0.545047919 0.585720561 1.0000000
17      1 - 14  0.927845716 0.353487606 1.0000000
18     10 - 14  1.673927266 0.094144895 1.0000000
19     11 - 14 -1.415342069 0.156968226 1.0000000
20     12 - 14  0.114025886 0.909217273 1.0000000
21     13 - 14 -1.074134985 0.282762187 1.0000000
22      0 - 15  0.197141056 0.843717156 1.0000000
23      1 - 15  2.214920961 0.026765490 1.0000000
24     10 - 15  2.654112625 0.007951728 1.0000000
25     11 - 15 -0.760166910 0.447154822 1.0000000
26     12 - 15  1.192618064 0.233018988 1.0000000
27     13 - 15 -0.378474547 0.705078098 1.0000000
28     14 - 15  0.850630553 0.394974612 1.0000000
29      0 - 16 -0.910259179 0.362685841 1.0000000
30      1 - 16  0.747456490 0.454788057 1.0000000
31     10 - 16  1.587380431 0.112426509 1.0000000
32     11 - 16 -2.067919632 0.038647581 1.0000000
33     12 - 16 -0.216274084 0.828774109 1.0000000
34     13 - 16 -1.575046162 0.115245790 1.0000000
35     14 - 16 -0.293607853 0.769057575 1.0000000
36     15 - 16 -1.403151835 0.160571567 1.0000000
37       0 - 2 -1.493582676 0.135284730 1.0000000
38       1 - 2  0.014789137 0.988200406 1.0000000
39      10 - 2  1.026840832 0.304495419 1.0000000
40      11 - 2 -2.739090153 0.006160948 0.8378889
41      12 - 2 -0.925965229 0.354464050 1.0000000
42      13 - 2 -2.198973199 0.027879828 1.0000000
43      14 - 2 -0.892201565 0.372284903 1.0000000
44      15 - 2 -2.111161993 0.034758390 1.0000000
45      16 - 2 -0.707322436 0.479366126 1.0000000
46       0 - 3 -1.760381303 0.078343177 1.0000000
```

47	1	- 3	-0.447147048	0.654768891	1.0000000
48	10	- 3	0.586900040	0.557270843	1.0000000
49	11	- 3	-2.853343922	0.004326177	0.5883601
50	12	- 3	-1.275761010	0.202040014	1.0000000
51	13	- 3	-2.395341667	0.016604884	1.0000000
52	14	- 3	-1.215333385	0.224238964	1.0000000
53	15	- 3	-2.307764092	0.021012261	1.0000000
54	16	- 3	-1.081384912	0.279525930	1.0000000
55	2	- 3	-0.448742077	0.653617727	1.0000000
56	0	- 4	-0.757439024	0.448786884	1.0000000
57	1	- 4	0.636219732	0.524633214	1.0000000
58	10	- 4	1.427847857	0.153335655	1.0000000
59	11	- 4	-1.624641123	0.104239049	1.0000000
60	12	- 4	-0.149519875	0.881143428	1.0000000
61	13	- 4	-1.283801933	0.199211267	1.0000000
62	14	- 4	-0.228664873	0.819129399	1.0000000
63	15	- 4	-1.085513133	0.277694450	1.0000000
64	16	- 4	0.025924064	0.979317906	1.0000000
65	2	- 4	0.609380308	0.542272387	1.0000000
66	3	- 4	0.950381695	0.341918342	1.0000000
67	0	- 5	-1.251792021	0.210645658	1.0000000
68	1	- 5	0.169634103	0.865297901	1.0000000
69	10	- 5	1.092213148	0.274739425	1.0000000
70	11	- 5	-2.295883285	0.021682556	1.0000000
71	12	- 5	-0.680545748	0.496158965	1.0000000
72	13	- 5	-1.863897751	0.062336065	1.0000000
73	14	- 5	-0.692584656	0.488570242	1.0000000
74	15	- 5	-1.723061779	0.084877354	1.0000000
75	16	- 5	-0.485928606	0.627017786	1.0000000
76	2	- 5	0.152085605	0.879119418	1.0000000
77	3	- 5	0.554321087	0.579359121	1.0000000
78	4	- 5	-0.436625232	0.662383161	1.0000000
79	0	- 6	-0.839607110	0.401128711	1.0000000
80	1	- 6	0.715689649	0.474183004	1.0000000
81	10	- 6	1.537886501	0.124076369	1.0000000
82	11	- 6	-1.871316322	0.061301250	1.0000000
83	12	- 6	-0.175815244	0.860439093	1.0000000
84	13	- 6	-1.446516544	0.148032375	1.0000000
85	14	- 6	-0.256926376	0.797235609	1.0000000
86	15	- 6	-1.257331732	0.208633538	1.0000000
87	16	- 6	0.023679407	0.981108333	1.0000000
88	2	- 6	0.680980330	0.495883937	1.0000000
89	3	- 6	1.042319663	0.297263499	1.0000000
90	4	- 6	-0.004782113	0.996184440	1.0000000
91	5	- 6	0.478605907	0.632219015	1.0000000
92	0	- 7	-0.257833768	0.796535202	1.0000000
93	1	- 7	1.542433904	0.122968188	1.0000000
94	10	- 7	2.175151022	0.029618801	1.0000000
95	11	- 7	-1.266599389	0.205298571	1.0000000
96	12	- 7	0.576261149	0.564438679	1.0000000
97	13	- 7	-0.855300533	0.392384793	1.0000000
98	14	- 7	0.365452373	0.714773753	1.0000000
99	15	- 7	-0.574646827	0.565530176	1.0000000
100	16	- 7	0.783736836	0.433194543	1.0000000
101	2	- 7	1.473633411	0.140580321	1.0000000
102	3	- 7	1.755239268	0.079218390	1.0000000
103	4	- 7	0.610528794	0.541511575	1.0000000

```
104       5 - 7  1.176512377 0.239390189 1.0000000
105       6 - 7  0.700050184 0.483895965 1.0000000
106       0 - 8 -0.053757428 0.957128428 1.0000000
107       1 - 8  1.756853485 0.078942786 1.0000000
108      10 - 8  2.332869518 0.019654995 1.0000000
109      11 - 8 -0.992480610 0.320963134 1.0000000
110      12 - 8  0.808302665 0.418916369 1.0000000
111      13 - 8 -0.619284331 0.535729065 1.0000000
112      14 - 8  0.562893899 0.573507135 1.0000000
113      15 - 8 -0.306904326 0.758916196 1.0000000
114      16 - 8  1.009396995 0.312784279 1.0000000
115       2 - 8  1.682492579 0.092473363 1.0000000
116       3 - 8  1.937886664 0.052637051 1.0000000
117       4 - 8  0.798245131 0.424728252 1.0000000
118       5 - 8  1.371228335 0.170303788 1.0000000
119       6 - 8  0.910840620 0.362379357 1.0000000
120       7 - 8  0.240839913 0.809679194 1.0000000
121       0 - 9 -1.038118889 0.299214709 1.0000000
122       1 - 9  0.572711262 0.566840211 1.0000000
123      10 - 9  1.452172230 0.146453722 1.0000000
124      11 - 9 -2.202725350 0.027614112 1.0000000
125      12 - 9 -0.377608446 0.705721489 1.0000000
126      13 - 9 -1.706234476 0.087964432 1.0000000
127      14 - 9 -0.428736790 0.668114784 1.0000000
128      15 - 9 -1.552427393 0.120559993 1.0000000
129      16 - 9 -0.162590654 0.870840746 1.0000000
130       2 - 9  0.539323227 0.589663845 1.0000000
131       3 - 9  0.930047241 0.352346625 1.0000000
132       4 - 9 -0.159114988 0.873578280 1.0000000
133       5 - 9  0.336680860 0.736357493 1.0000000
134       6 - 9 -0.174225951 0.861687891 1.0000000
135       7 - 9 -0.935570905 0.349494161 1.0000000
136       8 - 9 -1.155573595 0.247855614 1.0000000
```

No p-values resulting from Dunn’s test are statistically significant. Therefore, for each (i, j) topic group pairing, we fail to reject the null hypothesis that their distributions of retweets are equal to one another. **BTM topic modeling does not seem to uncover subject matter or underlying themes within Amazon official tweets which help further explain their expected number of retweets.** However, again, we only have 231 tweets split across 17 topics.

Amazon IRT Topics: Number of Likes

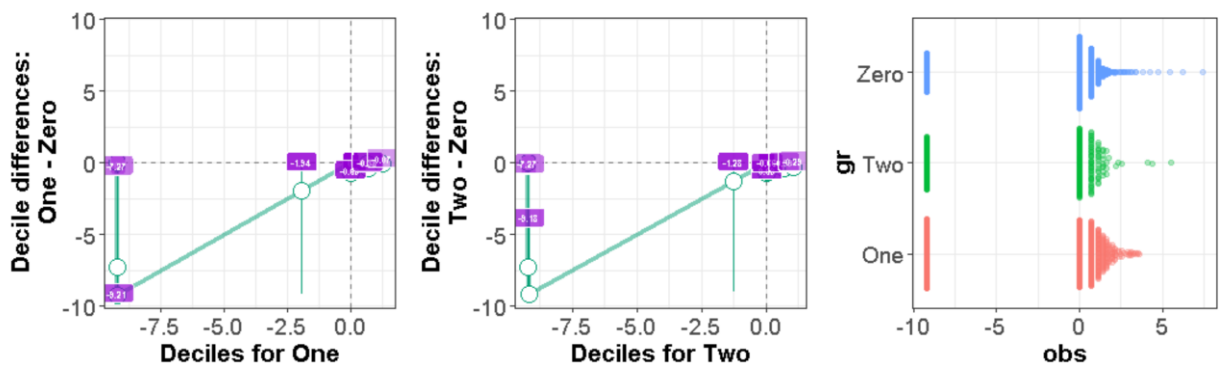



```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 97.796, df = 2, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison         Z       P.unadj        P.adj
1       0 - 1 8.568172 1.051409e-17 3.154226e-17
2       0 - 2 7.289008 3.122444e-13 9.367333e-13
3       1 - 2 1.309192 1.904692e-01 5.714076e-01
```

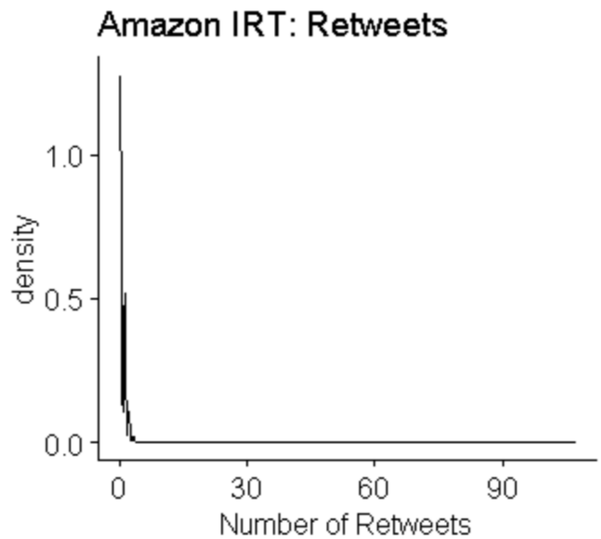
We fail to reject the null for the (topic 1, topic 2) pair. However, we may reject the null for the (topic 0, topic 1) and (topic 0, topic 2) pairs and conclude that the 'like' distributions of the topic groups within these pairs differ from one another. Performing a shift function to further examine these differences yields the following:

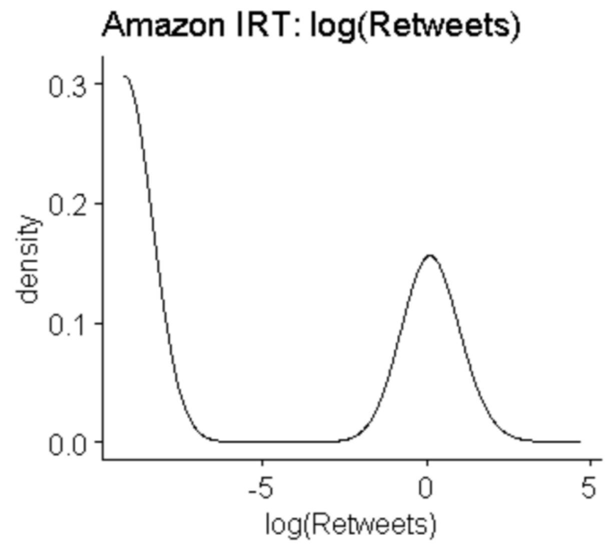


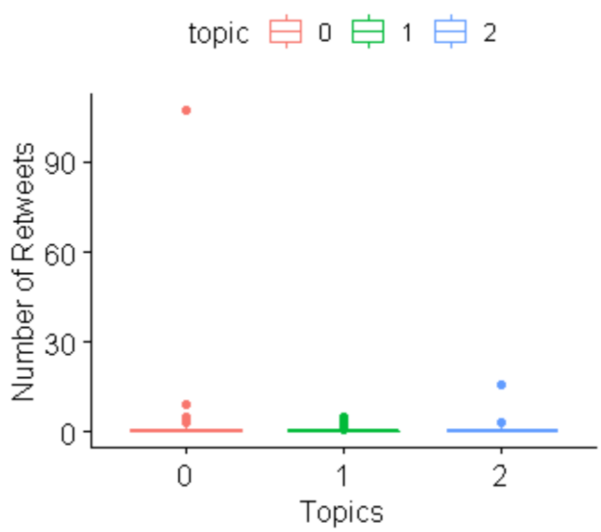
```
$`One - Zero`
    q          One      Zero    difference      ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.0000000000 -1.825797e-10  0.000000e+00 0.0062500000   0.732
2 0.2 0.000000e+00 0.7888855 -0.7888855020 -9.999164e-01 -2.005888e-02 0.0015625000   0.000
3 0.3 7.961445e-09 1.0000000 -0.9999999920 -1.000000e+00 -9.962028e-01 0.0012500000   0.000
4 0.4 7.891276e-01 1.0000000 -0.2108724027 -9.747310e-01 -1.034521e-04 0.0010416667   0.000
5 0.5 1.000000e+00 1.0008158 -0.0008157556 -3.993929e-01 -7.188884e-10 0.0008928571   0.000
6 0.6 1.000000e+00 1.9999620 -0.9999615301 -1.000000e+00 -8.533926e-01 0.0007812500   0.000
7 0.7 1.950473e+00 2.0000000 -0.0495267632 -9.523744e-01 -2.395845e-06 0.0006944444   0.000
8 0.8 2.003960e+00 2.9503309 -0.9463708415 -9.998867e-01  5.422125e-02 0.0020833333   0.005
9 0.9 3.562680e+00 3.8117935 -0.2491131489 -9.625729e-01  8.191151e-01 0.0031250000   0.598

$`Two - Zero`
    q          Two      Zero    difference      ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.0000000  0.0000000000 -1.825797e-10  0.000000e+00 0.0031250000   0.741
2 0.2 7.213563e-12 0.7888855 -0.7888855020 -9.999564e-01 -1.702210e-02 0.0010416667   0.000
3 0.3 3.278985e-03 1.0000000 -0.9967210148 -1.000000e+00 -2.113664e-01 0.0007812500   0.000
4 0.4 8.607182e-01 1.0000000 -0.1392817533 -9.937108e-01 -2.442152e-05 0.0006250000   0.000
5 0.5 9.999991e-01 1.0008158 -0.0008166721 -6.006074e-01 -1.503621e-08 0.0005208333   0.000
6 0.6 1.000000e+00 1.9999620 -0.9999620160 -1.000000e+00 -7.229880e-01 0.0004464286   0.000
7 0.7 1.046199e+00 2.0000000 -0.9538011890 -1.000081e+00 -1.086811e-01 0.0003906250   0.000
8 0.8 1.994057e+00 2.9503309 -0.9562740504 -1.739037e+00 -8.124780e-02 0.0003472222   0.000
9 0.9 2.871102e+00 3.8117935 -0.9406919763 -1.950072e+00  9.086744e-02 0.0015625000   0.010
```

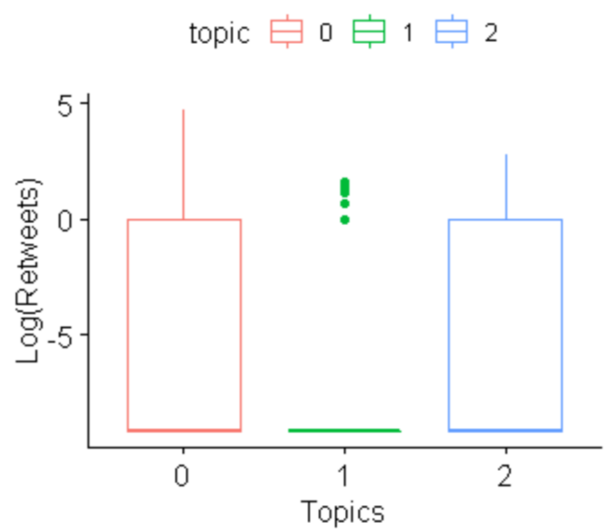
From the top table we can say, with 95% confidence, that the 2nd through 7th quantiles of topic 0 tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of topic 1 tweets. From the bottom table we can say, with 95% confidence, that the 2nd through 8th quantiles of topic 0 tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of topic 2 tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Amazon IRT tweets.** Looking back at my thoughts on Amazon IRT topics, it's very interesting that differences were found between topics 0 and 1 (I considered both to be 'positive feedback responses', but topic 0 was essentially short form while topic 1 was 'long' form). It's not surprising, however, that topic 2 has worse performance than topic 0, as topic 2 is considered to be 'classic customer service.' Furthermore, it seems that these significant results are due to the fact that topic 0 has a smaller proportion of tweets receiving 0 likes than topics 1 and 2 have.

Amazon IRT Topics: Number of Retweets







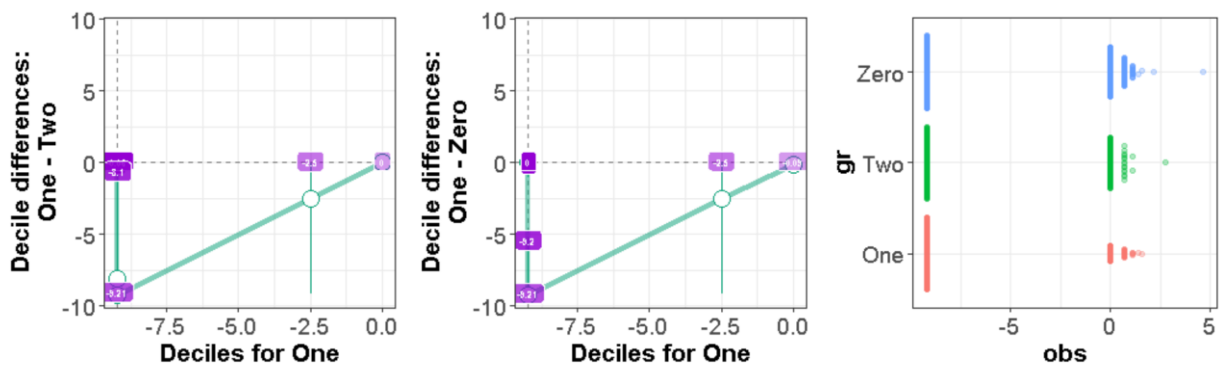



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 133.16, df = 2, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison          Z       P.unadj         P.adj
1      0 - 1 11.2045135 3.874767e-29 1.162430e-28
2      0 - 2  0.7429454 4.575147e-01 1.000000e+00
3      1 - 2 -6.8124813 9.592954e-12 2.877886e-11
```

We fail to reject the null for the (topic 0, topic 2) pair and conclude that their distributions of retweets are equal. However, we may reject the null for both the (topic 0, topic 1) and (topic 1, topic 2) pairings and conclude that the 'retweet' distributions of the topic groups within these pairs differ from one another. Performing a shift function to further examine these differences yields the following:

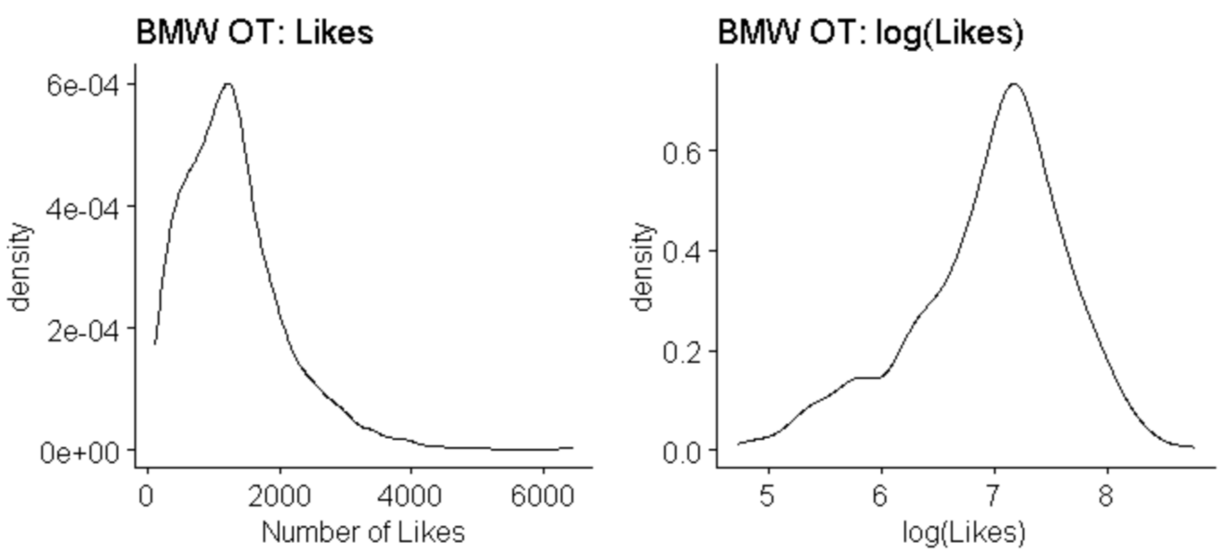



```
$`One - Two`
    q          One          Two    difference      ci_lower      ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.050000000  1.0000
2 0.2 0.000000e+00 0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00 0.025000000  1.0000
3 0.3 0.000000e+00 0.000000e+00  0.000000e+00 -2.309264e-14  0.000000e+00 0.016666667  0.9430
4 0.4 0.000000e+00 4.074472e-10 -4.074472e-10 -1.056389e-04  0.000000e+00 0.010000000  0.0205
5 0.5 0.000000e+00 6.305836e-03 -6.305836e-03 -4.780000e-01 -1.519129e-07 0.008333333  0.0000
6 0.6 0.000000e+00 8.789607e-01 -8.789607e-01 -9.999567e-01 -6.117691e-02 0.007142857  0.0000
7 0.7 9.381940e-12 9.999996e-01 -9.999996e-01 -1.000000e+00 -9.843422e-01 0.006250000  0.0000
8 0.8 7.285226e-01 1.000000e+00 -2.714774e-01 -9.915862e-01 -2.350908e-04 0.005555556  0.0000
9 0.9 1.000000e+00 1.000666e+00 -6.661259e-04 -3.089692e-01  6.617261e-07 0.012500000  0.0440

$`One - Zero`
    q          One         Zero    difference   ci_lower      ci_upper      p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.012500000   1.000
2 0.2 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.006250000   1.000
3 0.3 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.004166667   1.000
4 0.4 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.003125000   0.999
5 0.5 0.000000e+00 1.886494e-05 -1.886494e-05 -0.2394199 -1.210698e-12 0.002500000   0.000
6 0.6 0.000000e+00 9.986176e-01 -9.986176e-01 -1.0000000 -5.061495e-01 0.002083333   0.000
7 0.7 9.381940e-12 1.000000e+00 -1.000000e+00 -1.0000000 -9.999548e-01 0.001785714   0.000
8 0.8 7.285226e-01 1.000000e+00 -2.714774e-01 -0.9948682 -3.085564e-04 0.001562500   0.000
9 0.9 1.000000e+00 1.133094e+00 -1.330938e-01 -0.9683659 -9.123139e-06 0.001388889   0.000
```

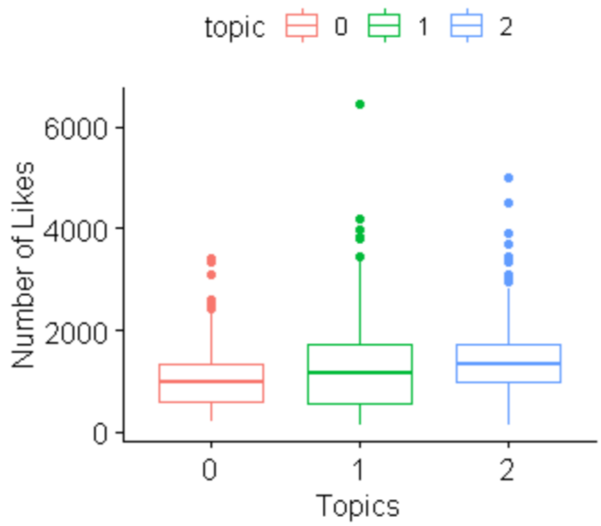
From the top table we can say, with 95% confidence, that the 5^{th} through 8^{th} quantiles of topic 2 tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in topic 1 tweets. From the bottom table we can say, with 95% confidence, that the 5^{th} through 9^{th} quantiles of topic 0 tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in topic 1 tweets. **Therefore, we may conclude that statistically significant differences**

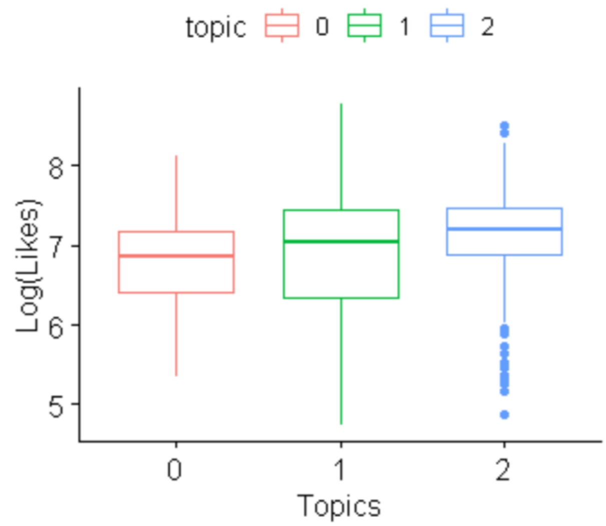
exist between the quantiles of the above two group pairings (in their distributions of retweets), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Amazon IRT tweets. It's very interesting that topic 1 is associated with the worst retweet performance, I would've assumed the opposite to be true before going into this. Furthermore, the real difference seems to be that topic 1 tweets very rarely receive a non-zero number of retweets.

BMW Official Topics: Number of Likes



The log distribution does not pass a Shapiro-Wilk normality test.



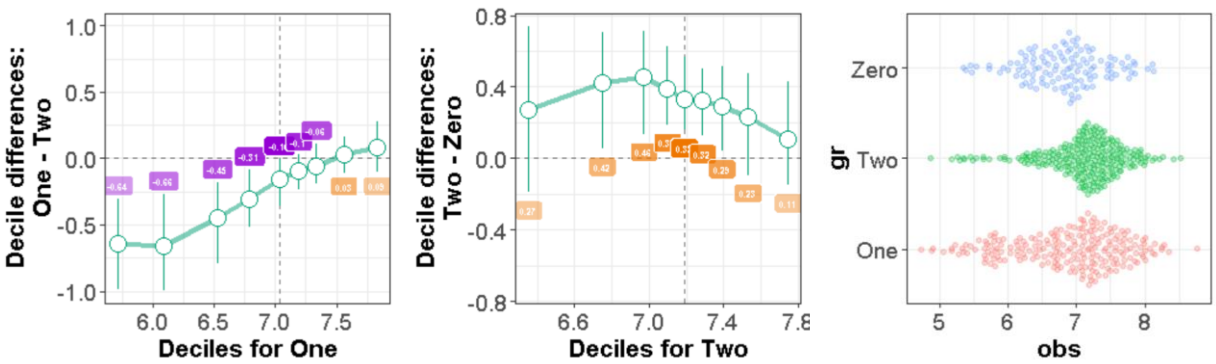



```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 25.499, df = 2, p-value = 2.903e-06
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison         Z       P.unadj        P.adj
1      0 - 1 -1.902429 5.711509e-02 1.713453e-01
2      0 - 2 -4.756872 1.966157e-06 5.898471e-06
3      1 - 2 -3.353075 7.991912e-04 2.397573e-03
```

We fail to reject the null for the (topic 0, topic 1) pair and conclude that their distributions of likes are equal. However, we may reject the null for both the (topic 0, topic 2) and (topic 1, topic 2) pairings and conclude that the 'like' distributions of the topic groups within these pairs differ from one another. Performing a shift function to further examine these differences yields the following:



$`One - Two`

	q	One	Two	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	302.7643	576.0709	-273.30659	-418.3220	-121.202449	0.008333333	0.000
2	0.2	443.4421	854.5826	-411.14056	-605.6901	-174.746033	0.007142857	0.000
3	0.3	686.1867	1066.8827	-380.69608	-601.9929	-154.560934	0.006250000	0.000
4	0.4	891.2745	1209.5690	-318.29448	-496.2328	-84.773019	0.005555556	0.000
5	0.5	1134.7384	1325.4283	-190.68991	-437.8550	-9.482992	0.010000000	0.007
6	0.6	1327.6275	1460.9036	-133.27612	-325.0825	36.623187	0.012500000	0.048
7	0.7	1536.3237	1621.5730	-85.24925	-285.2924	159.344959	0.025000000	0.433
8	0.8	1919.8836	1864.3405	55.54314	-201.4940	304.711552	0.050000000	0.655
9	0.9	2527.8092	2315.3081	212.50110	-302.7142	698.279453	0.016666667	0.362

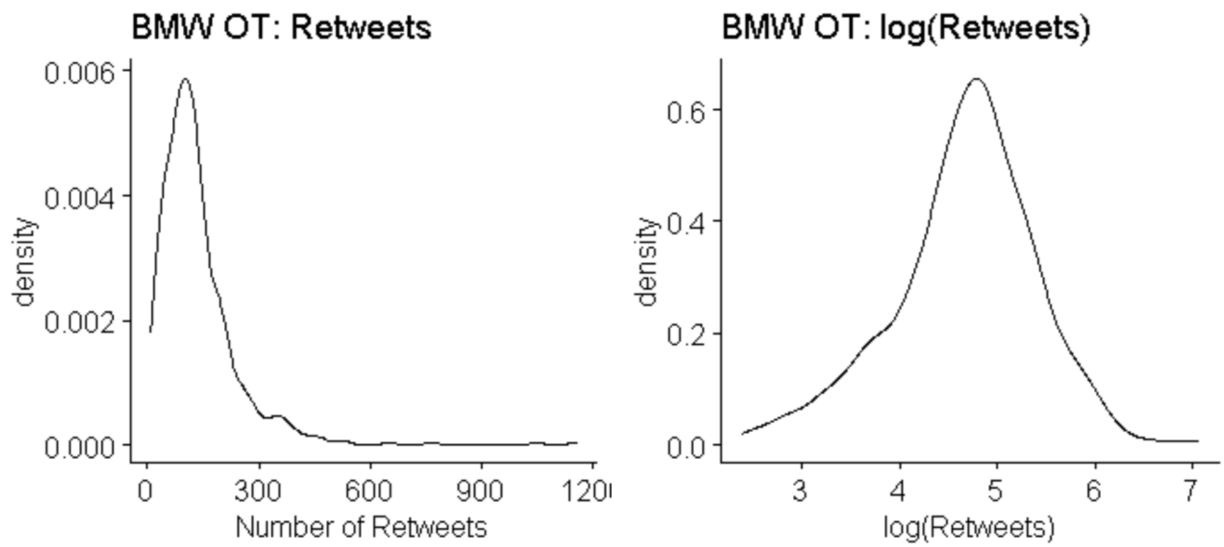
We can say, with 95% confidence, that the 1st through 5th quantiles of topic 2 tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in topic 1 tweets.

$`Two - Zero`

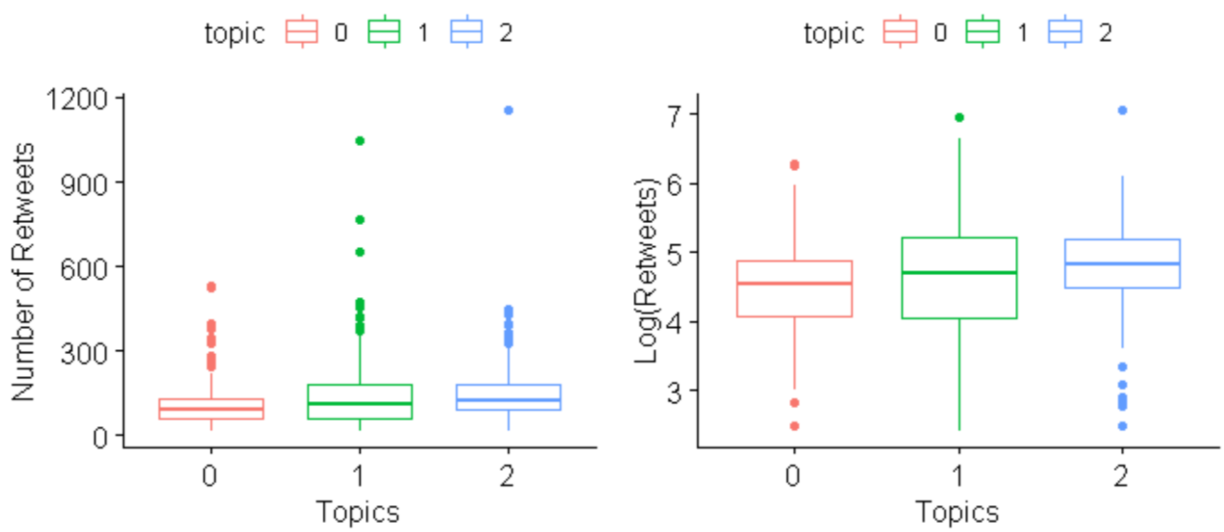
	q	Two	Zero	difference	ci_lower	ci_upper	p_crit	p_value
1	0.1	576.0709	441.7884	134.2825	-67.79265	362.4506	0.0013888889	0.050
2	0.2	854.5826	559.8645	294.7181	53.14397	487.3983	0.0010416667	0.000
3	0.3	1066.8827	676.3392	390.5436	140.32718	589.6760	0.0008333333	0.000
4	0.4	1209.5690	823.3299	386.2391	155.09790	584.3873	0.0006944444	0.000
5	0.5	1325.4283	955.3160	370.1123	202.20778	595.5194	0.0005952381	0.000
6	0.6	1460.9036	1056.3461	404.5575	166.96751	586.5023	0.0005208333	0.000
7	0.7	1621.5730	1213.5628	408.0101	94.20804	690.2493	0.0004629630	0.000
8	0.8	1864.3405	1483.7924	380.5480	-196.51638	793.5701	0.0020833333	0.028
9	0.9	2315.3081	2084.5921	230.7159	-321.82510	873.5306	0.0041666667	0.313

We can say, with 95% confidence, that the 2nd through 7th quantiles of topic 0 tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in topic 2 tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of BMW official tweets.** It seems that topic 2 (BMW tweets regarding new cars) tweets, generally, expect to receive the highest number of likes out of all topic categories.

BMW Official Topics: Number of Retweets



The log distribution does not pass a Shapiro-Wilk normality test.

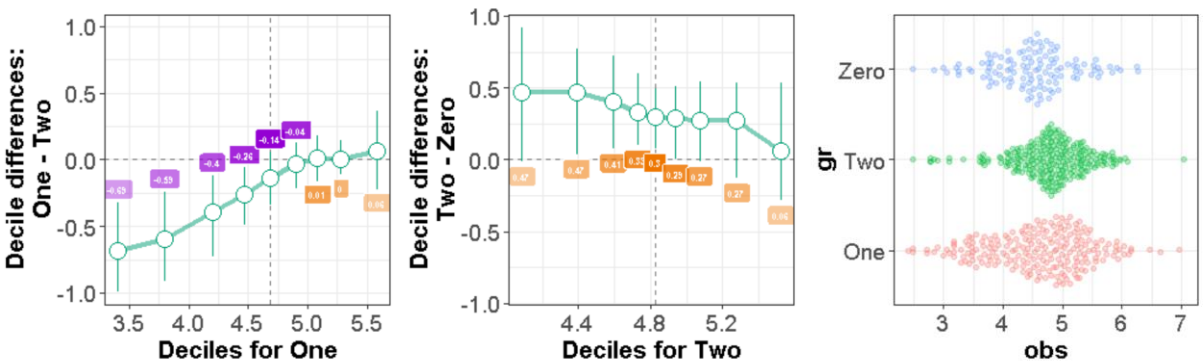



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 19.757, df = 2, p-value = 5.126e-05
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison         Z       P.unadj        P.adj
1      0 - 1 -1.888193 5.900005e-02 1.770002e-01
2      0 - 2 -4.259605 2.047885e-05 6.143654e-05
3      1 - 2 -2.773130 5.552001e-03 1.665600e-02
```

We fail to reject the null for the (topic 0, topic 1) pair and conclude that their distributions of retweets are equal. However, we may reject the null for both the (topic 0, topic 2) and (topic 1, topic 2) pairings and conclude that the 'retweet' distributions of the topic groups within these pairs differ from one another. Performing a shift function to further examine these differences yields the following:



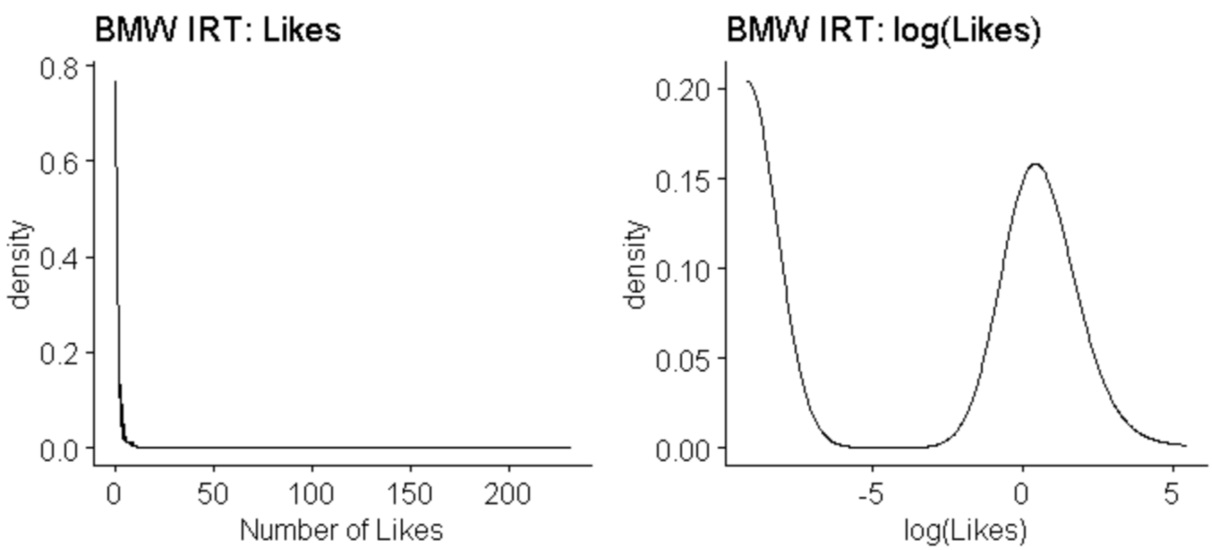

```
$`One - Two`
    q       One       Two  difference  ci_lower   ci_upper      p_crit p_value
1 0.1  30.09213  59.78536 -29.6932245 -41.78234 -13.123569 0.007142857   0.000
2 0.2  44.70659  80.72778 -36.0211900 -52.03531 -17.073063 0.006250000   0.000
3 0.3  66.85159  98.88586 -32.0342751 -53.83848 -10.305596 0.005555556   0.000
4 0.4  87.37221 113.06056 -25.6883511 -44.84379  -7.822910 0.008333333   0.001
5 0.5 108.40149 124.44610 -16.0446077 -39.34014   6.961957 0.010000000   0.082
6 0.6 135.01252 139.74989  -4.7373720 -29.32175  16.366596 0.016666667   0.613
7 0.7 161.78291 159.47226   2.3106432 -24.62971  27.993333 0.025000000   0.869
8 0.8 196.75467 195.99251   0.7621668 -24.02762  28.507227 0.050000000   0.897
9 0.9 267.80842 251.03106  16.7773666 -52.72588 105.647474 0.012500000   0.575
```

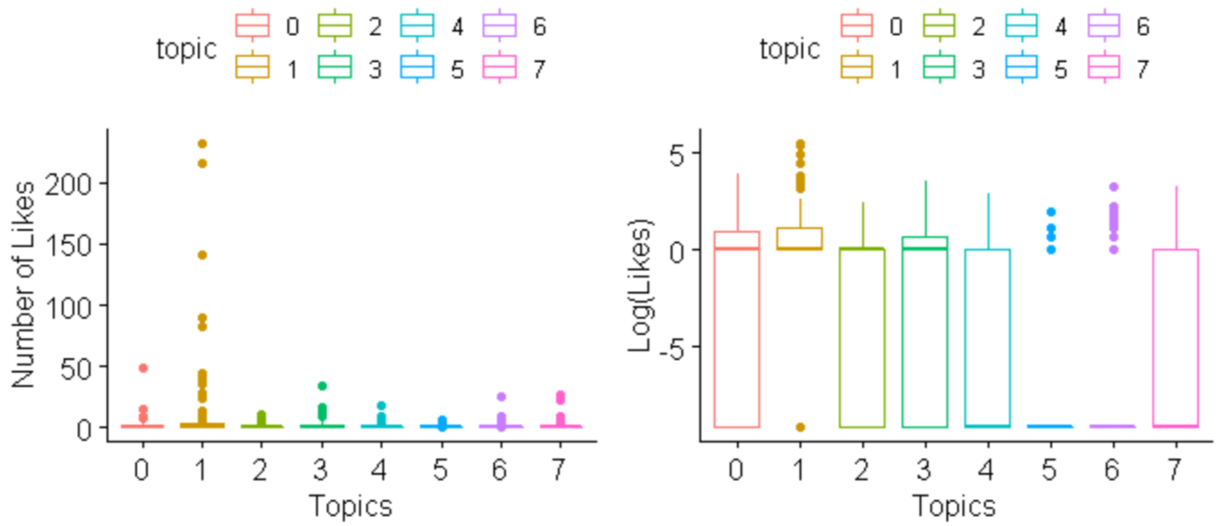
We can say, with 95% confidence, that the 1st through 4th quantiles of topic 2 would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of topic 1 tweets.

```
$`Two - Zero`
    q       Two      Zero difference   ci_lower  ci_upper       p_crit p_value
1 0.1  59.78536  37.86758   21.91778   0.212812  39.99277 0.0020833333   0.002
2 0.2  80.72778  50.75826   29.96952   6.396679  50.10504 0.0012500000   0.000
3 0.3  98.88586  66.22285   32.66301   7.725712  54.31633 0.0010416667   0.000
4 0.4 113.06056  81.53014   31.53042  10.714493  52.07050 0.0008928571   0.000
5 0.5 124.44610  92.63113   31.81497  12.333268  52.62279 0.0007812500   0.000
6 0.6 139.74989 104.77274   34.97716   6.722935  58.14950 0.0006944444   0.000
7 0.7 159.47226 121.72894   37.74332  -2.560294  69.27859 0.0015625000   0.003
8 0.8 195.99251 150.54668   45.44583 -33.478599  89.83464 0.0031250000   0.056
9 0.9 251.03106 237.26848   13.76258 -83.507817 103.83359 0.0062500000   0.661
```

We can say, with 95% confidence, that the 1st through 6th quantiles of topic 0 tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of topic 2 tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of retweets, and the left halves specifically), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of BMW official tweets.** Essentially, it seems that topic 2 tweets have a higher 'floor' than topic 0 and topic 1 tweets. In other words, the worst performing topic 2 tweets seem to receive more retweets than the worst performing topic 0 and topic 1 tweets.

BMW IRT Topics: Number of Likes





Here, the least represented topic is topic 2, which contains 98 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 374.02, df = 7, p-value < 2.2e-16
```

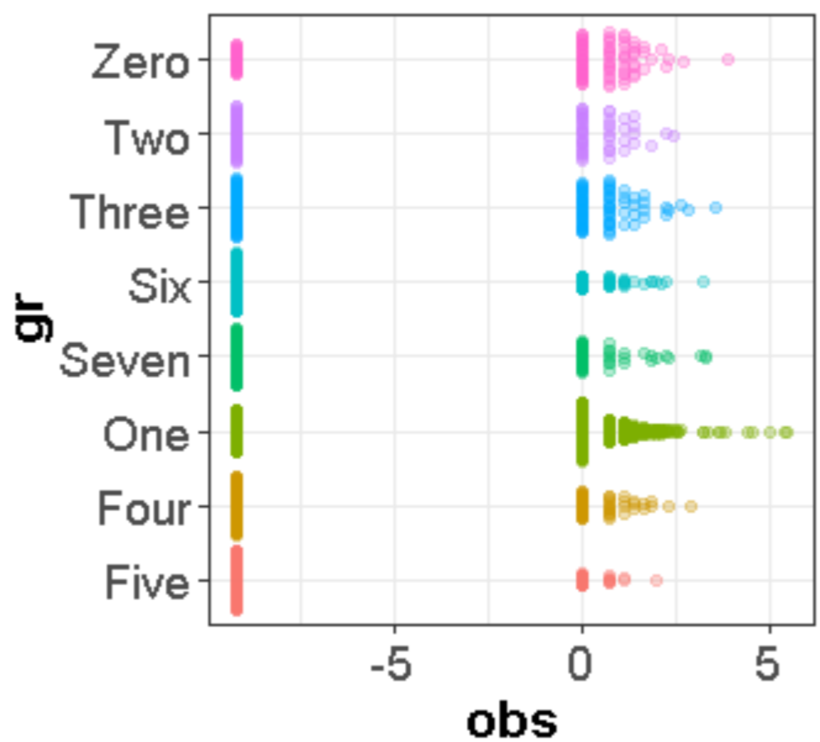
Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

   Comparison          Z      P.unadj        P.adj
1       0 - 1 -0.7356200 4.619620e-01 1.000000e+00
2       0 - 2  3.2277861 1.247522e-03 3.493063e-02
3       1 - 2  4.7214298 2.341924e-06 6.557388e-05
4       0 - 3  2.9577221 3.099214e-03 8.677798e-02
5       1 - 3  4.7719268 1.824718e-06 5.109211e-05
6       2 - 3 -0.5890861 5.558035e-01 1.000000e+00
7       0 - 4  6.2871008 3.234493e-10 9.056580e-09
8       1 - 4  9.6200411 6.580488e-22 1.842537e-20
9       2 - 4  2.4238897 1.535527e-02 4.299475e-01
10      3 - 4  3.4471697 5.664925e-04 1.586179e-02
11      0 - 5  9.4656314 2.917929e-21 8.170201e-20
12      1 - 5 14.5467693 6.122710e-48 1.714359e-46
13      2 - 5  5.3050433 1.126462e-07 3.154095e-06
14      3 - 5  6.8005465 1.042230e-11 2.918245e-10
15      4 - 5  3.4200363 6.261277e-04 1.753158e-02
16      0 - 6  9.0289197 1.733748e-19 4.854494e-18
17      1 - 6 15.0010158 7.230418e-51 2.024517e-49
18      2 - 6  4.6415323 3.458349e-06 9.683377e-05
```

```
19        3 - 6  6.1862326 6.161908e-10 1.725334e-08
20        4 - 6  2.5164898 1.185303e-02 3.318849e-01
21        5 - 6 -1.2571992 2.086815e-01 1.000000e+00
22        0 - 7  5.4864952 4.099865e-08 1.147962e-06
23        1 - 7  8.0897317 5.979630e-16 1.674296e-14
24        2 - 7  1.8565486 6.337541e-02 1.000000e+00
25        3 - 7  2.7341851 6.253488e-03 1.750977e-01
26        4 - 7 -0.5332132 5.938860e-01 1.000000e+00
27        5 - 7 -3.7288946 1.923216e-04 5.385004e-03
28        6 - 7 -2.9016413 3.712133e-03 1.039397e-01
```

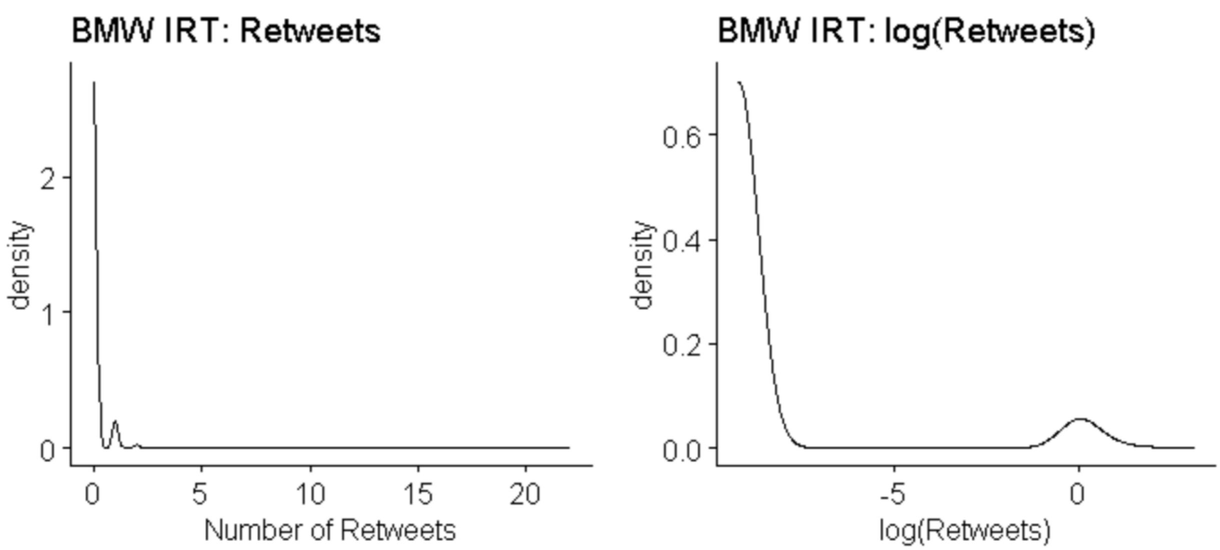
We may reject the null for a number of (i, j) topic group pairings. Performing a shift function to further examine differences yields the following: (In order to keep this document from becoming extremely cluttered, confidence intervals are stored in 'Topic_Appendix.docx').

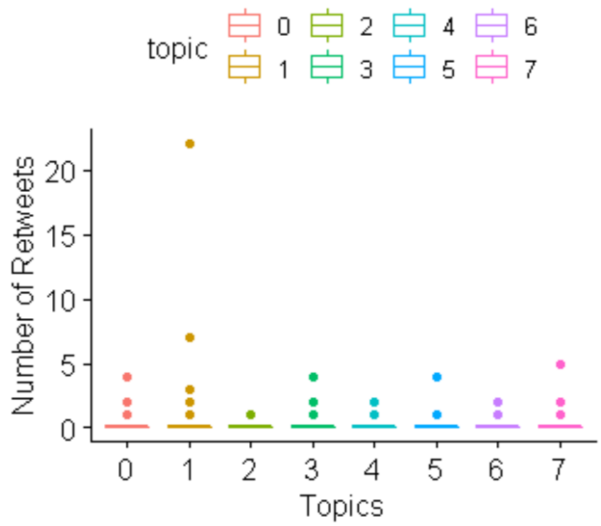
Note: Adding the 'Topic_Appendix' document would perhaps double the length of this appendix. Thus, confidence intervals stored in that document will not be included in this report.

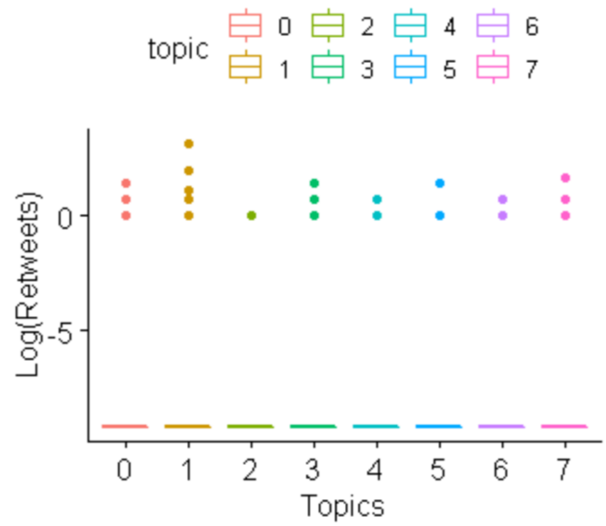


Topics 0 and 1 seem to be the only two topics which have less tweets receiving 0 likes than tweets receiving 1 like. This isn't too surprising, considering I dubbed both of these different variations of 'positive feedback responses.' There are too many topic group combinations to comment on each individual shift, but plenty are significant (non-zero). **Therefore, we may conclude that statistically significant differences exist between the quantiles of certain topic group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of BMW IRT tweets.**

BMW IRT: Number of Retweets







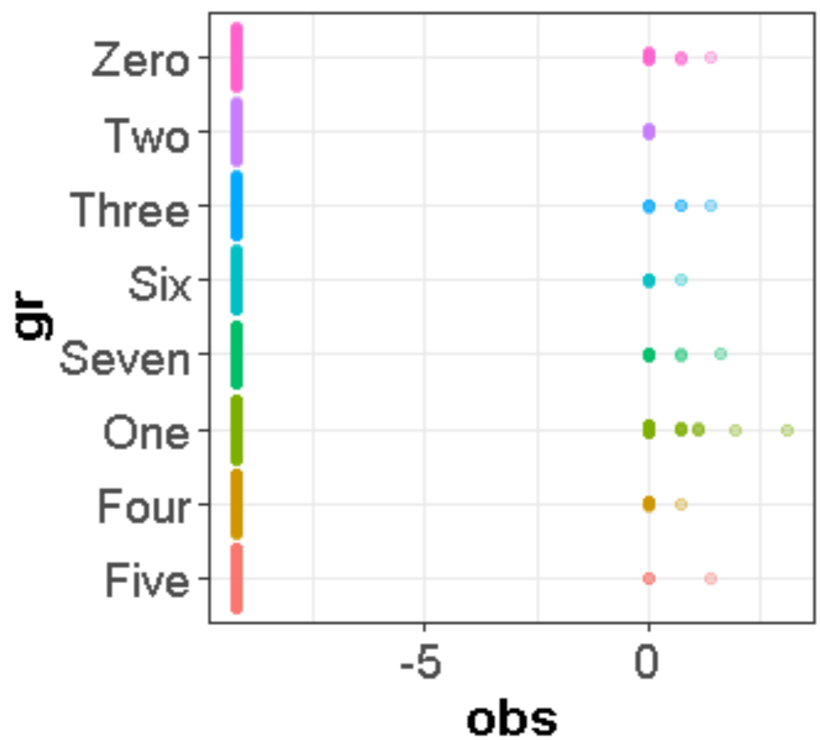

```
	Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 58.4, df = 7, p-value = 3.147e-10
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni"
)
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

   Comparison          Z       P.unadj         P.adj
1       0 - 1  0.2915119 7.706598e-01 1.000000e+00
2       0 - 2  1.2677631 2.048825e-01 1.000000e+00
3       1 - 2  1.3113438 1.897416e-01 1.000000e+00
4       0 - 3  2.9457991 3.221215e-03 9.019402e-02
5       1 - 3  3.6166549 2.984347e-04 8.356172e-03
6       2 - 3  1.4713295 1.412020e-01 1.000000e+00
7       0 - 4  2.2506546 2.440743e-02 6.834079e-01
8       1 - 4  2.7755804 5.510328e-03 1.542892e-01
9       2 - 4  0.7441539 4.567834e-01 1.000000e+00
10      3 - 4 -0.8938103 3.714234e-01 1.000000e+00
11      0 - 5  4.1308970 3.613504e-05 1.011781e-03
12      1 - 5  5.5592627 2.709167e-08 7.585668e-07
13      2 - 5  2.4749740 1.332459e-02 3.730885e-01
14      3 - 5  1.0263285 3.047368e-01 1.000000e+00
15      4 - 5  2.0813397 3.740282e-02 1.000000e+00
16      0 - 6  4.0762760 4.576271e-05 1.281356e-03
17      1 - 6  5.8583334 4.675354e-09 1.309099e-07
18      2 - 6  2.3219421 2.023606e-02 5.666096e-01
19      3 - 6  0.7675532 4.427527e-01 1.000000e+00
```

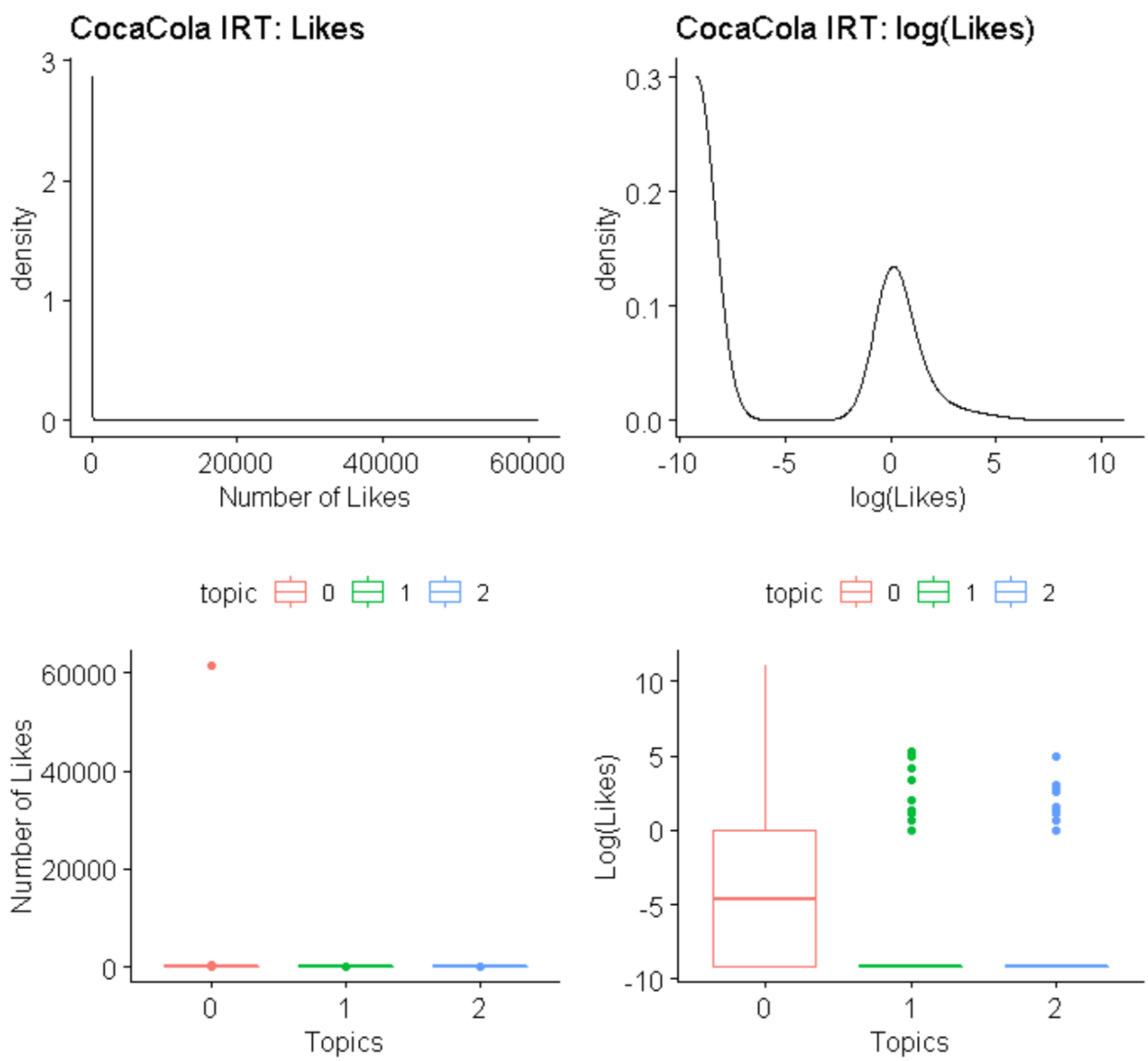
```
20       4 - 6  1.8960520 5.795316e-02 1.000000e+00
21       5 - 6 -0.3788121 7.048274e-01 1.000000e+00
22       0 - 7  2.0599882 3.939967e-02 1.000000e+00
23       1 - 7  2.4177396 1.561725e-02 4.372830e-01
24       2 - 7  0.6385286 5.231296e-01 1.000000e+00
25       3 - 7 -0.9193570 3.579089e-01 1.000000e+00
26       4 - 7 -0.0829641 9.338801e-01 1.000000e+00
27       5 - 7 -2.0182625 4.356393e-02 1.000000e+00
28       6 - 7 -1.8316955 6.699680e-02 1.000000e+00
```

We may reject the null for a number of (i, j) topic group pairings. Performing a shift function to further examine these differences yields the following (confidence intervals stored in 'Topic_Appendix'):



After examining the confidence intervals, as well as the images produced above, I don't feel comfortable making any conclusions for BMW IRT retweets. When differences are significant, they seem to be extremely small and are restrained only to the far right-tails of topic group pairings. Had the images produced above been more convincing, I may have felt comfortable drawing conclusions. However, it seems that all BMW IRT tweets expect 0 retweets, and any positive number of retweets could be considered an outlier or anomaly. **BTM topic modeling does not seem to uncover subject matter or underlying themes within BMW IRT tweets which help further explain their expected number of retweets.**

Coca Cola IRT: Number of Likes



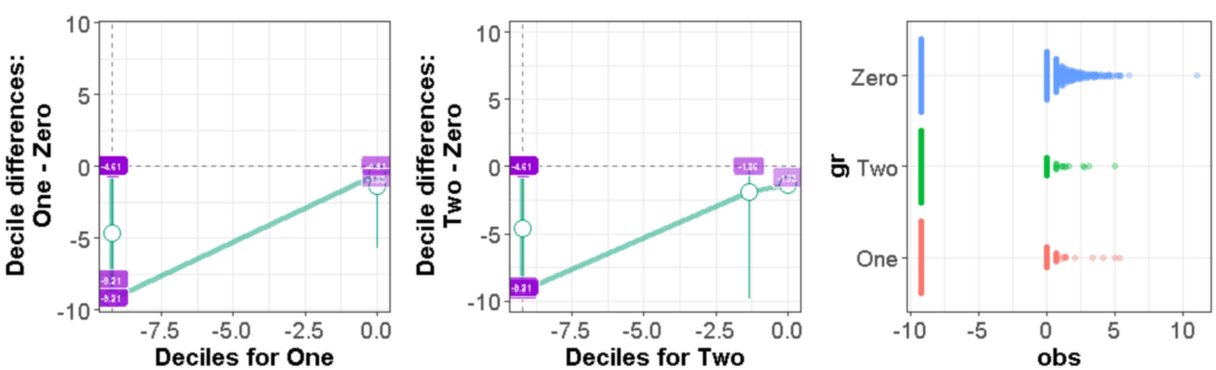
Topic 2 is the least represented above, containing 666 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 281.92, df = 2, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison        Z      P.unadj        P.adj
1      0 - 1 14.06393 6.327288e-45 1.898186e-44
2      0 - 2 13.71133 8.685376e-43 2.605613e-42
3      1 - 2  1.02667 3.045761e-01 9.137282e-01
```

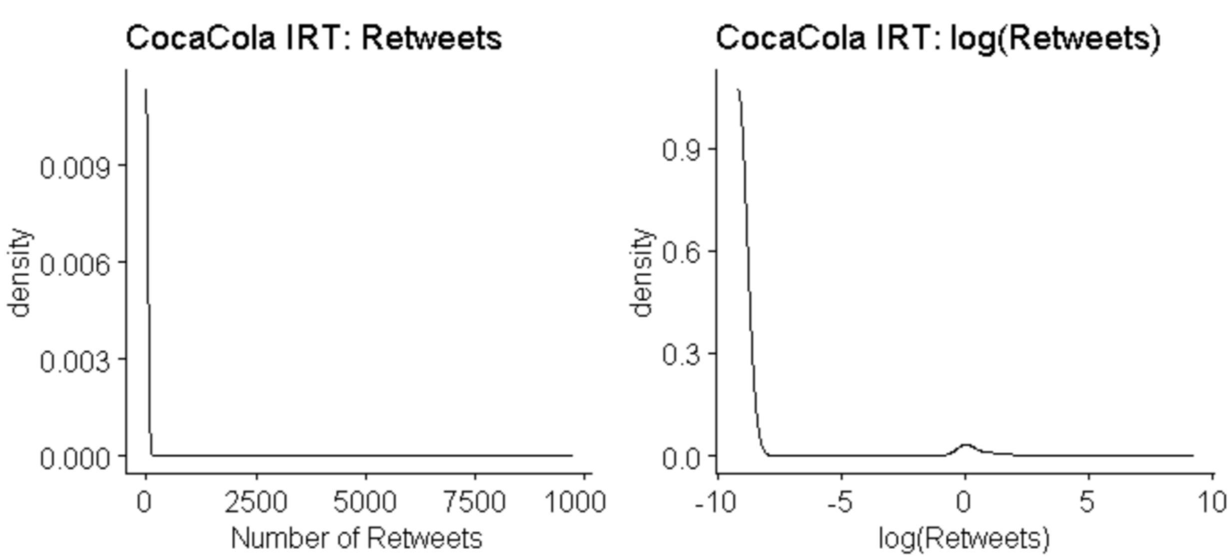
We fail to reject the null for the (topic 1, topic 2) pair and conclude that their distributions of likes are equal. However, we may reject the null for both the (topic 0, topic 1) and (topic 0, topic 2) pairings and conclude that the 'like' distributions of the topic groups within these pairs differ from one another. Performing a shift function to further examine these differences yields the following:

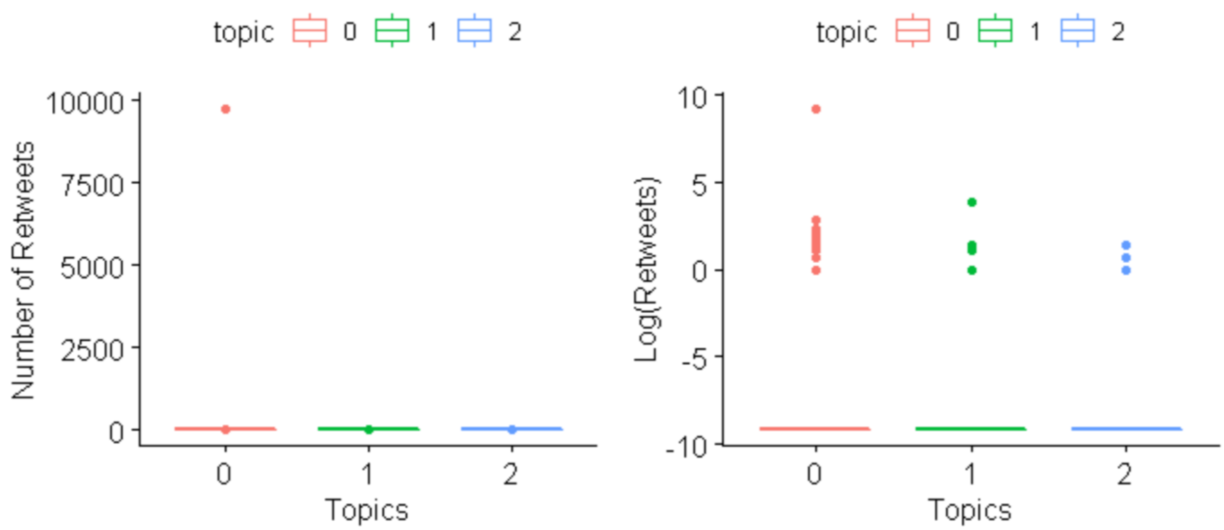


```
$`One - Zero`
    q          One         Zero    difference      ci_lower     ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000000  0.000000000 0.0062500000  1.0000
2 0.2 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000000  0.000000000 0.0031250000  1.0000
3 0.3 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000000  0.000000000 0.0020833333  1.0000
4 0.4 0.000000e+00 8.348877e-14 -8.348877e-14 -0.0000612831  0.000000000 0.0015625000  0.1725
5 0.5 0.000000e+00 5.000000e-01 -5.000000e-01 -0.9998981422 -0.001913328 0.0012500000  0.0000
6 0.6 0.000000e+00 1.000000e+00 -1.000000e+00 -1.0000000000 -0.999995501 0.0010416667  0.0000
7 0.7 2.866886e-05 1.000000e+00 -9.999713e-01 -1.0000000023 -0.845732954 0.0008928571  0.0000
8 0.8 9.990674e-01 1.719482e+00 -7.204148e-01 -1.2323555517 -0.007673756 0.0007812500  0.0000
9 0.9 1.000000e+00 4.020258e+00 -3.020258e+00 -5.0041254247 -1.549037079 0.0006944444  0.0000

$`Two - Zero`
    q          Two         Zero    difference      ci_lower      ci_upper       p_crit p_value
1 0.1 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000000  0.0000000000 6.944444e-04  1.0000
2 0.2 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000000  0.0000000000 3.472222e-04  1.0000
3 0.3 0.000000e+00 0.000000e+00  0.000000e+00  0.0000000000  0.0000000000 2.314815e-04  1.0000
4 0.4 0.000000e+00 8.348877e-14 -8.348877e-14 -0.0008694052  0.0000000000 1.736111e-04  0.1835
5 0.5 0.000000e+00 5.000000e-01 -5.000000e-01 -0.9995662916 -0.0009318375 1.388889e-04  0.0000
6 0.6 0.000000e+00 1.000000e+00 -1.000000e+00 -1.0000000000 -0.9999689132 1.157407e-04  0.0000
7 0.7 2.373326e-07 1.000000e+00 -9.999998e-01 -1.0000736988 -0.9696081631 9.920635e-05  0.0000
8 0.8 8.521724e-01 1.719482e+00 -8.673097e-01 -1.9500459350 -0.0236526138 8.680556e-05  0.0000
9 0.9 1.000000e+00 4.020258e+00 -3.020258e+00 -5.1509471042 -1.7368215938 7.716049e-05  0.0000
```

We can say, with 95% confidence, that the 5th through 9th quantiles of topic 0 tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of topic 1 tweets, as well as the set of topic 2 tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes, and the right halves specifically), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Coca Cola IRT tweets.**

Coca Cola IRT Topics: Number of Retweets



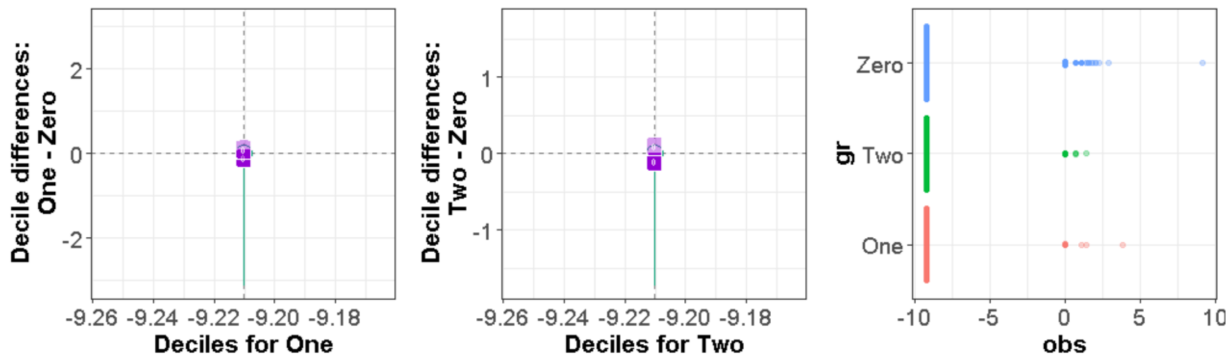



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 54.159, df = 2, p-value = 1.736e-12
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison         Z       P.unadj        P.adj
1      0 - 1  6.870719 6.387923e-12 1.916377e-11
2      0 - 2  4.984236 6.220720e-07 1.866216e-06
3      1 - 2 -1.101827 2.705369e-01 8.116106e-01
```

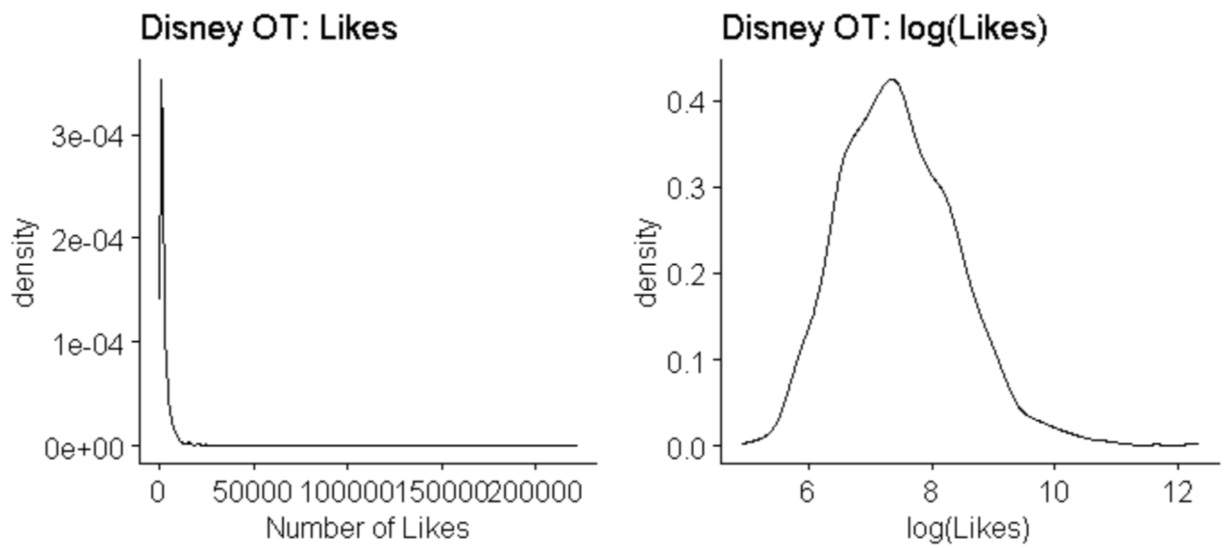
We fail to reject the null for the (topic 1, topic 2) pair and conclude that their distributions of retweets are equal. However, we may reject the null for both the (topic 0, topic 1) and (topic 0, topic 2) pairings and conclude that the 'retweet' distributions of the topic groups within these pairs differ from one another. Performing a shift function to further examine these differences yields the following:



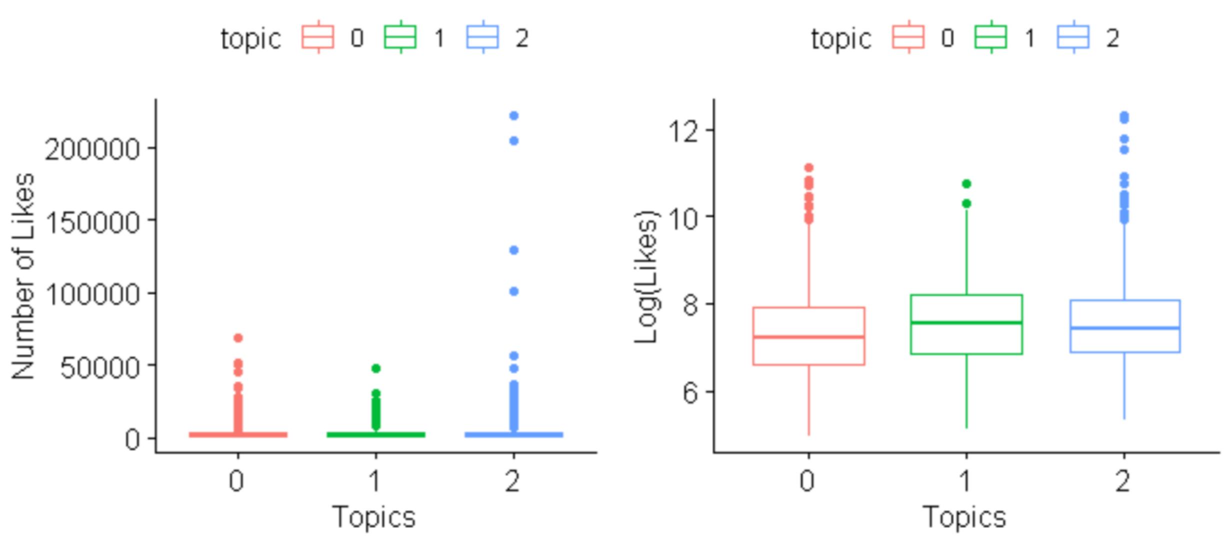

```
$`One - Zero`
    q One          Zero    difference   ci_lower     ci_upper       p_crit p_value
1 0.1   0 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.0055555556       1
2 0.2   0 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.0027777778       1
3 0.3   0 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.0018518519       1
4 0.4   0 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.0013888889       1
5 0.5   0 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.0011111111       1
6 0.6   0 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.0009259259       1
7 0.7   0 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.0007936508       1
8 0.8   0 0.000000e+00  0.000000e+00  0.0000000  0.000000e+00 0.0006944444       1
9 0.9   0 5.541096e-05 -5.541096e-05 -0.2145097 -2.708944e-13 0.0006172840       0

$`Two - Zero`
    q Two          Zero    difference   ci_lower    ci_upper       p_crit p_value
1 0.1   0 0.000000e+00  0.000000e+00  0.0000000 0.000000e+00 6.172840e-04   1.000
2 0.2   0 0.000000e+00  0.000000e+00  0.0000000 0.000000e+00 3.086420e-04   1.000
3 0.3   0 0.000000e+00  0.000000e+00  0.0000000 0.000000e+00 2.057613e-04   1.000
4 0.4   0 0.000000e+00  0.000000e+00  0.0000000 0.000000e+00 1.543210e-04   1.000
5 0.5   0 0.000000e+00  0.000000e+00  0.0000000 0.000000e+00 1.234568e-04   1.000
6 0.6   0 0.000000e+00  0.000000e+00  0.0000000 0.000000e+00 1.028807e-04   1.000
7 0.7   0 0.000000e+00  0.000000e+00  0.0000000 0.000000e+00 8.818342e-05   1.000
8 0.8   0 0.000000e+00  0.000000e+00  0.0000000 0.000000e+00 7.716049e-05   1.000
9 0.9   0 5.541096e-05 -5.541096e-05 -0.1633353 1.033502e-10 6.858711e-05   0.002
```

Considering the visualizations produced above, the small magnitudes of the shifts, and the fact that all shifts are contained to the 9th quantile of topic 0 tweets, I don't feel comfortable drawing any conclusions. Topic 0 contains the most tweets here, and these differences may be due to that fact, rather than underlying subject matter. **BTM topic modeling does not seem to uncover subject matter or underlying themes within Coca Cola IRT tweets which help further explain their expected number of retweets.**

Disney Official: Number of Likes



The log distribution does not pass a Shapiro-Wilk normality test.



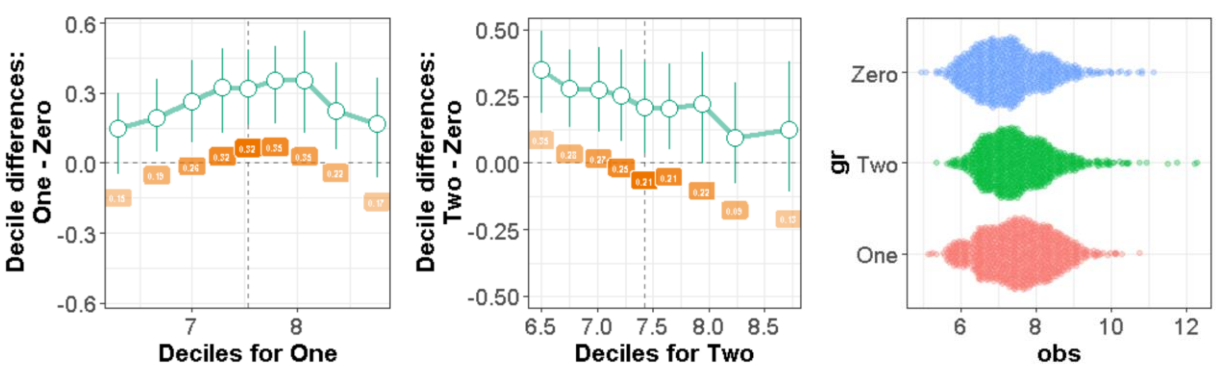
Topic 0 is the least represented above, containing 705 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 35.661, df = 2, p-value = 1.804e-08
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison          Z       P.unadj         P.adj
1      0 - 1 -5.6021261 2.117384e-08 6.352153e-08
2      0 - 2 -4.9081706 9.192986e-07 2.757896e-06
3      1 - 2  0.7669346 4.431204e-01 1.000000e+00
```

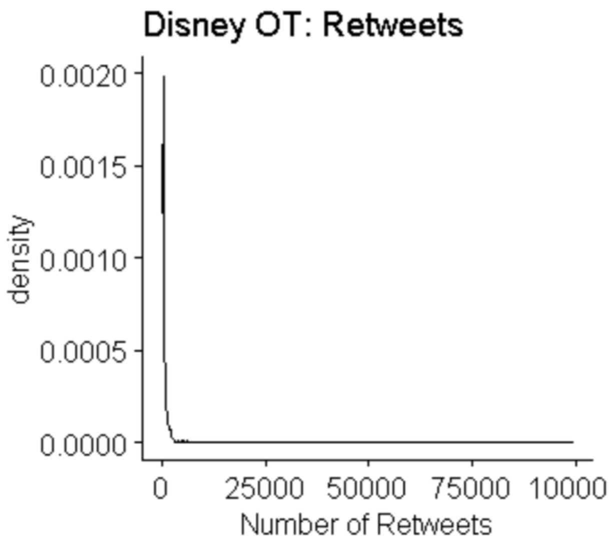
We fail to reject the null for the (topic 1, topic 2) pair and conclude that their distributions of likes are equal. However, we may reject the null for both the (topic 0, topic 1) and (topic 0, topic 2) pairings and conclude that the ‘like’ distributions of the topic groups within these pairs differ from one another. Performing a shift function to further examine these differences yields the following:

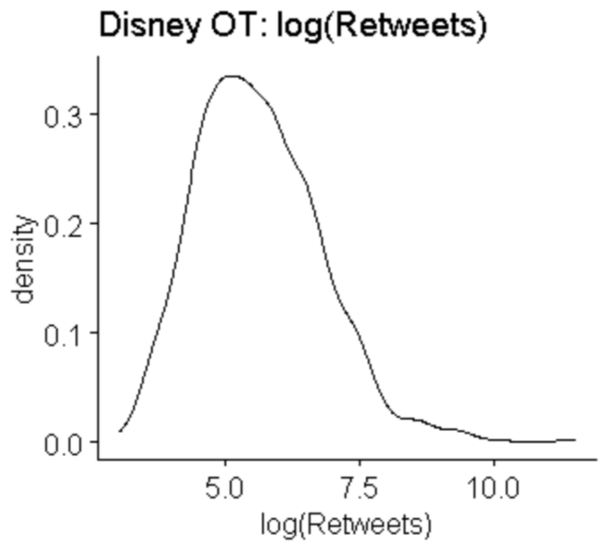


```
$`One - Zero`
    q       One      Zero difference   ci_lower  ci_upper      p_crit p_value
1 0.1  542.9804  466.9683   76.01213  -12.44219  159.2478 0.025000000   0.066
2 0.2  783.4906  646.9833  136.50727   33.22021  251.1048 0.008333333   0.001
3 0.3 1095.2864  843.2358  252.05056   88.02826  407.2532 0.006250000   0.000
4 0.4 1460.0330 1057.0782  402.95480  166.36523  610.2575 0.005000000   0.000
5 0.5 1869.0139 1355.8980  513.11597  254.09979  799.7599 0.004166667   0.000
6 0.6 2408.8432 1694.0518  714.79143  389.02694 1034.2694 0.003571429   0.000
7 0.7 3185.6130 2239.9687  945.64426  294.60235 1456.3726 0.003125000   0.000
8 0.8 4278.5602 3425.2971  853.26314  176.40277 1729.9985 0.002777778   0.000
9 0.9 6357.4382 5380.2619  977.17628 -372.42507 2151.2795 0.012500000   0.062

$`Two - Zero`
    q       Two      Zero difference   ci_lower  ci_upper      p_crit p_value
1 0.1  661.0453  466.9683   194.0770  118.45253  266.3824 0.003125000   0.000
2 0.2  853.6051  646.9833   206.6218   98.63797  320.1818 0.002500000   0.000
3 0.3 1107.8849  843.2358   264.6490  106.71407  412.8619 0.002083333   0.000
4 0.4 1360.1650 1057.0782   303.0868   97.41346  481.8137 0.001785714   0.000
5 0.5 1669.6890 1355.8980   313.7911   98.15538  568.6183 0.001562500   0.000
6 0.6 2081.0487 1694.0518   386.9969   75.57584  745.7733 0.001388889   0.000
7 0.7 2795.7430 2239.9687   555.7742   46.68479 1047.6024 0.004166667   0.003
8 0.8 3756.5499 3425.2971   331.2528 -298.91142 1039.8308 0.006250000   0.172
9 0.9 6100.0331 5380.2619   719.7712 -683.72154 2255.3857 0.012500000   0.198
```

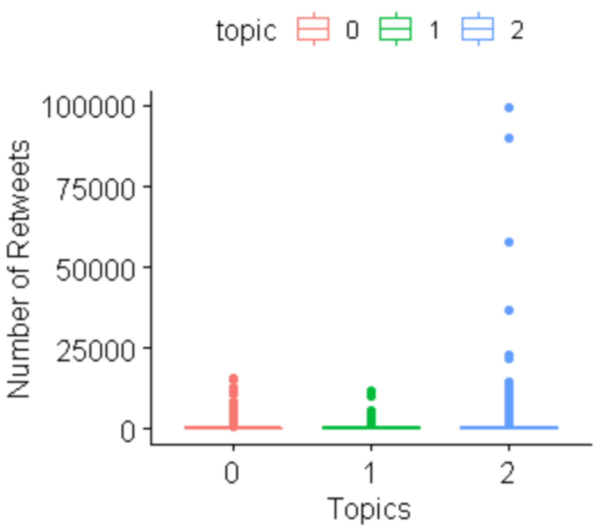
From the top table we can say, with 95% confidence, that the 2nd through 8th quantiles of topic 0 tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of topic 1 tweets. From the bottom table we can say, with 95% confidence, that the 1st through 7th quantiles of topic 0 tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of topic 2 tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Disney official tweets.**

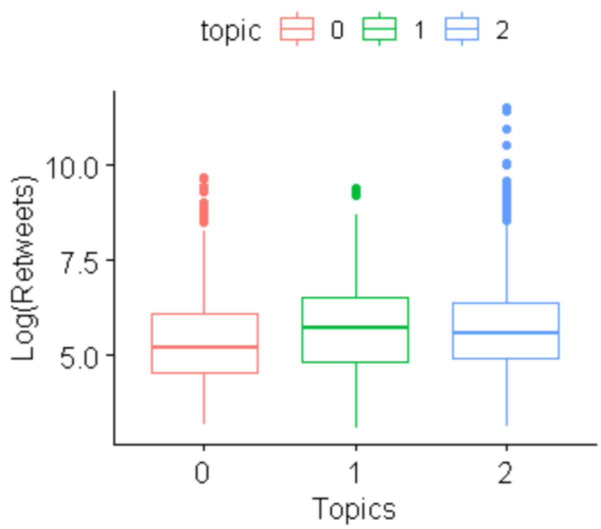
Disney Official Topics: Number of Retweets





The log distribution does not pass a Shapiro-Wilk normality test.



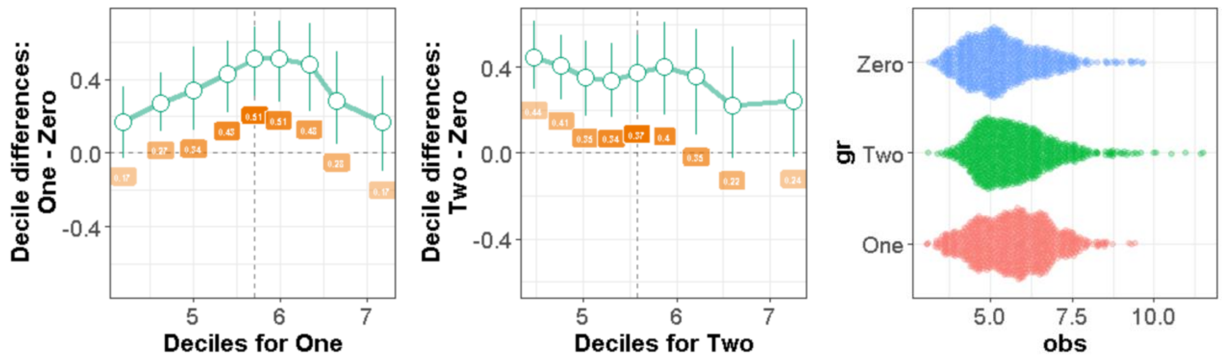



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 50.755, df = 2, p-value = 9.519e-12
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison           Z       P.unadj         P.adj
1      0 - 1 -6.32930488 2.462681e-10 7.388043e-10
2      0 - 2 -6.30998613 2.790605e-10 8.371814e-10
3      1 - 2  0.03936908 9.685961e-01 1.000000e+00
```

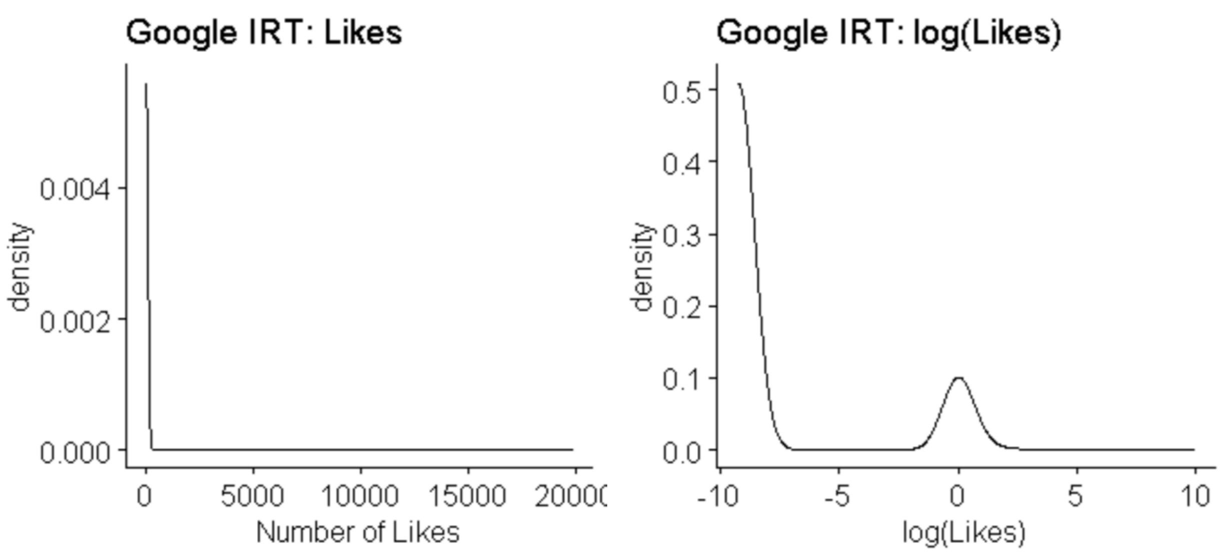
We fail to reject the null for the (topic 1, topic 2) pair and conclude that their distributions of retweets are equal. However, we may reject the null for both the (topic 0, topic 1) and (topic 0, topic 2) pairings and conclude that the 'retweet' distributions of the topic groups within these pairs differ from one another. Performing a shift function to further examine these differences yields the following:

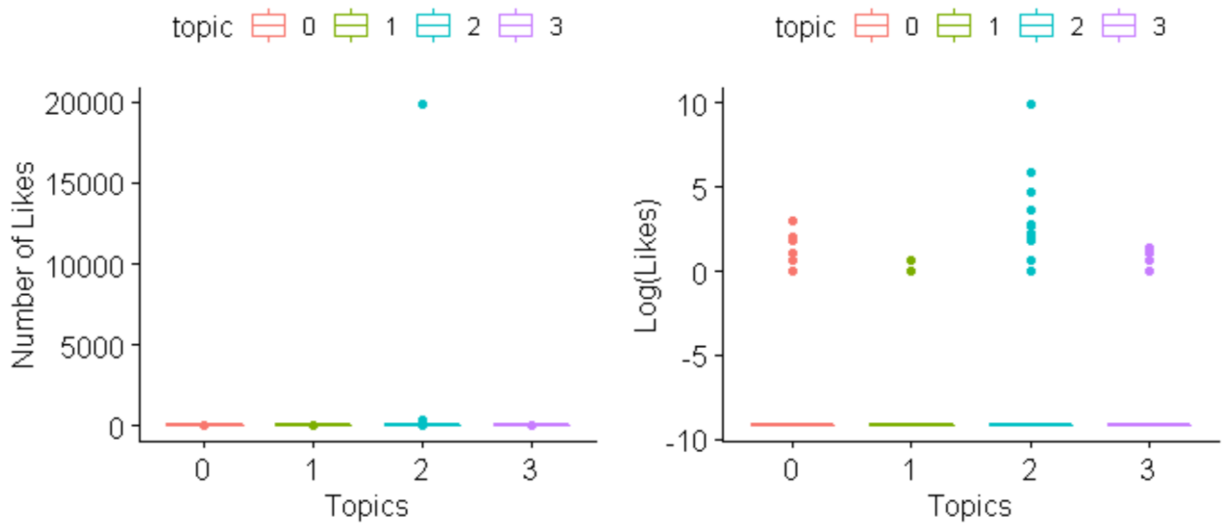



```
$`One - Zero`
    q       One       Zero difference    ci_lower  ci_upper      p_crit p_value
1 0.1   66.62892   56.24491   10.38401   -2.267680  22.87103 0.008333333   0.023
2 0.2  102.34968   78.00109   24.34859    9.041513  39.32664 0.005555556   0.000
3 0.3  149.31830  106.43332   42.88498   15.229040  73.56637 0.004166667   0.000
4 0.4  219.24320  143.20861   76.03459   38.473608 114.36693 0.003333333   0.000
5 0.5  301.63050  181.13172  120.49878   69.681067 166.94116 0.002777778   0.000
6 0.6  396.03434  238.26899  157.76535   91.492745 221.85041 0.002380952   0.000
7 0.7  563.07816  348.68687  214.39129   95.132301 316.23209 0.002083333   0.000
8 0.8  773.14110  584.95513  188.18597   18.804294 365.72631 0.001851852   0.000
9 0.9 1309.38482 1110.55505  198.82977 -151.658940 530.85475 0.016666667   0.145

$`Two - Zero`
    q        Two       Zero difference  ci_lower  ci_upper      p_crit p_value
1 0.1   87.57708   56.24491   31.33217  20.14116  42.36542 0.005555556   0.000
2 0.2  117.28518   78.00109   39.28409  23.69588  54.37744 0.004166667   0.000
3 0.3  151.22606  106.43332   44.79274  20.37724  68.12538 0.003333333   0.000
4 0.4  200.31640  143.20861   57.10780  27.10133  84.66770 0.002777778   0.000
5 0.5  263.05525  181.13172   81.92353  37.59242 131.86514 0.002380952   0.000
6 0.6  354.60622  238.26899  116.33723  47.61963 181.44025 0.002083333   0.000
7 0.7  497.07356  348.68687  148.38668  35.96535 243.69952 0.001851852   0.000
8 0.8  730.04408  584.95513  145.08895 -30.41245 339.54541 0.008333333   0.027
9 0.9 1414.37441 1110.55505  303.81936 -48.62893 720.69891 0.016666667   0.040
```

From the top table we can say, with 95% confidence, that the 2nd through 8th quantiles of topic 0 tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of topic 1 tweets. From the bottom table we can say, with 95% confidence, that the 1st through 7th quantiles of topic 0 tweets would need to be shifted up by significant (non-zero) amounts to match their counterparts in the set of topic 2 tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of retweets), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Disney official tweets.**

Google IRT Topics: Number of Likes





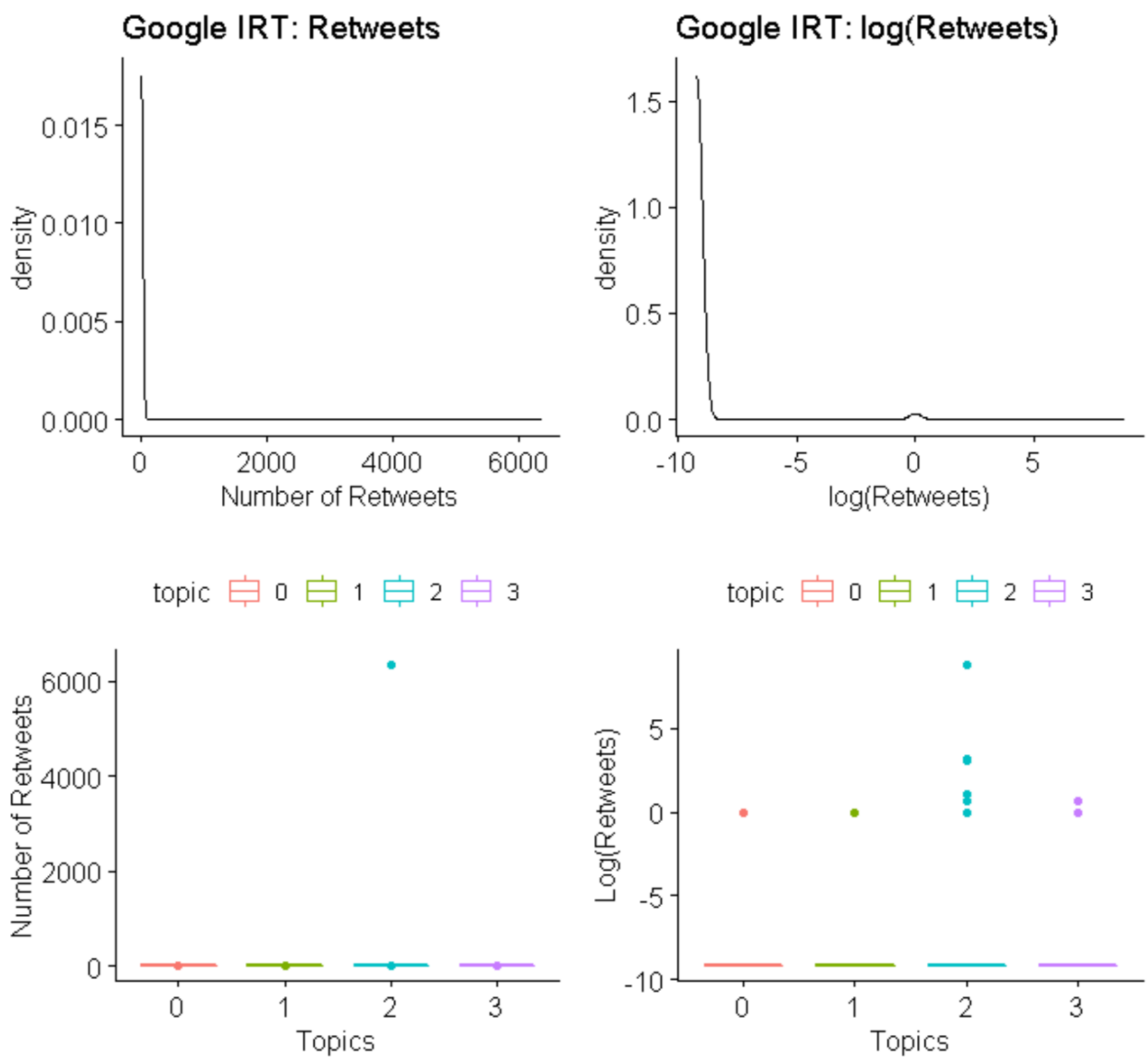

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 8.2643, df = 3, p-value = 0.04085
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison          Z     P.unadj     P.adj
1      0 - 1  0.7237141 0.46924126 1.0000000
2      0 - 2 -1.9746232 0.04831093 0.2898656
3      1 - 2 -2.2707054 0.02316482 0.1389889
4      0 - 3 -1.7628642 0.07792339 0.4675404
5      1 - 3 -2.0746182 0.03802193 0.2281316
6      2 - 3  0.6137503 0.53938034 1.0000000
```

No p-values resulting from Dunn's test are statistically significant. Therefore, for each (i, j) topic group pairing, we fail to reject the null hypothesis that their distributions of likes are equal to one another. **BTM topic modeling does not seem to uncover subject matter or underlying themes within Google IRT tweets which help further explain their expected number of likes.**

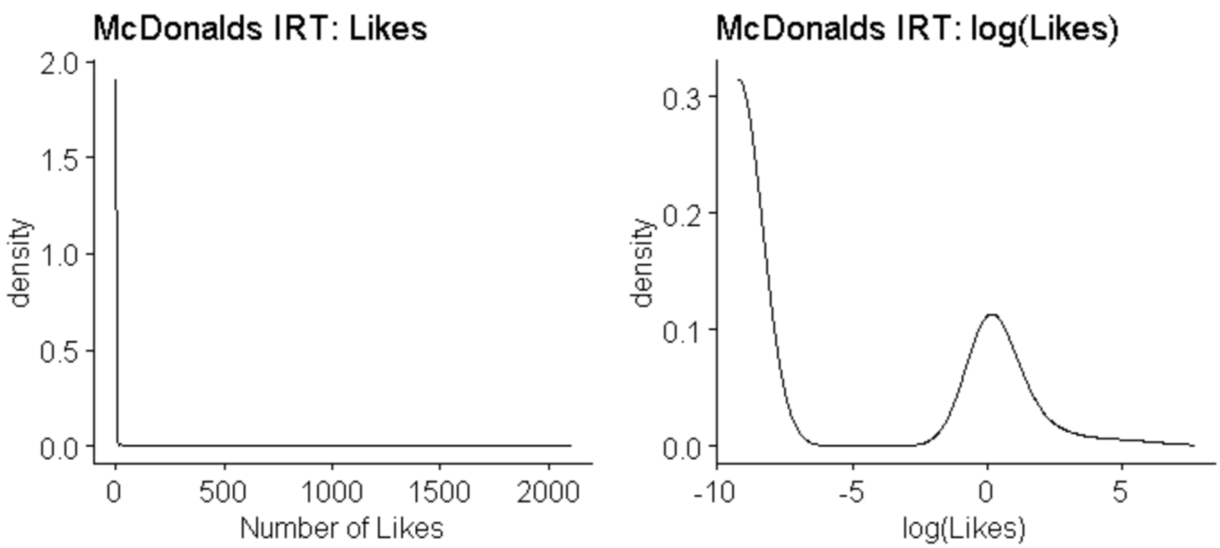
Google IRT Topics: Number of Retweets

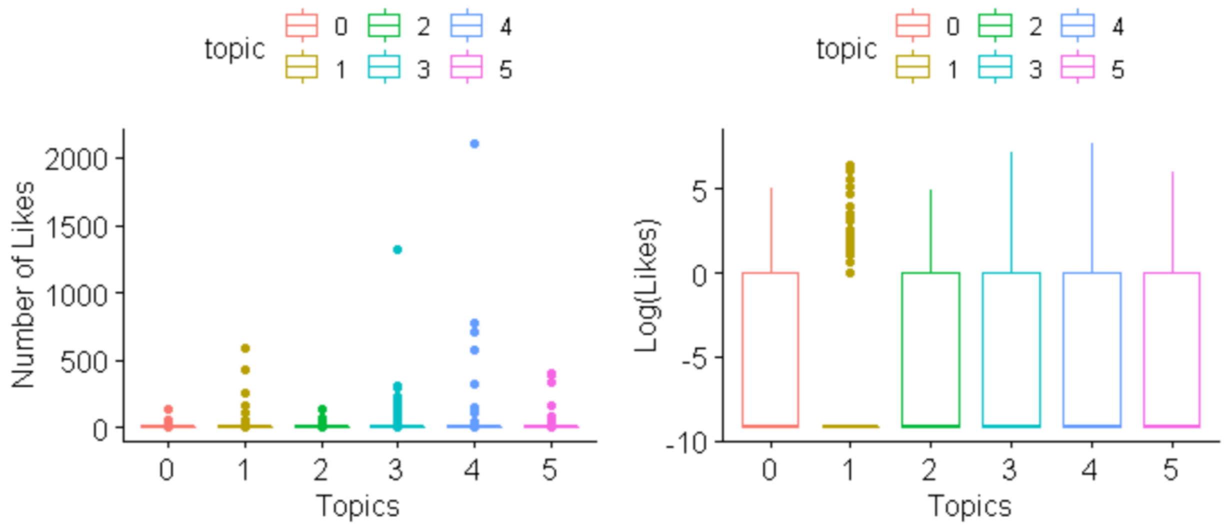



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 6.5963, df = 3, p-value = 0.08594
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of all populations are equal. **BTM topic modeling does not seem to uncover subject matter or underlying themes within Google IRT tweets which help further explain their expected number of retweets.**

McDonalds IRT Topics: Number of Likes





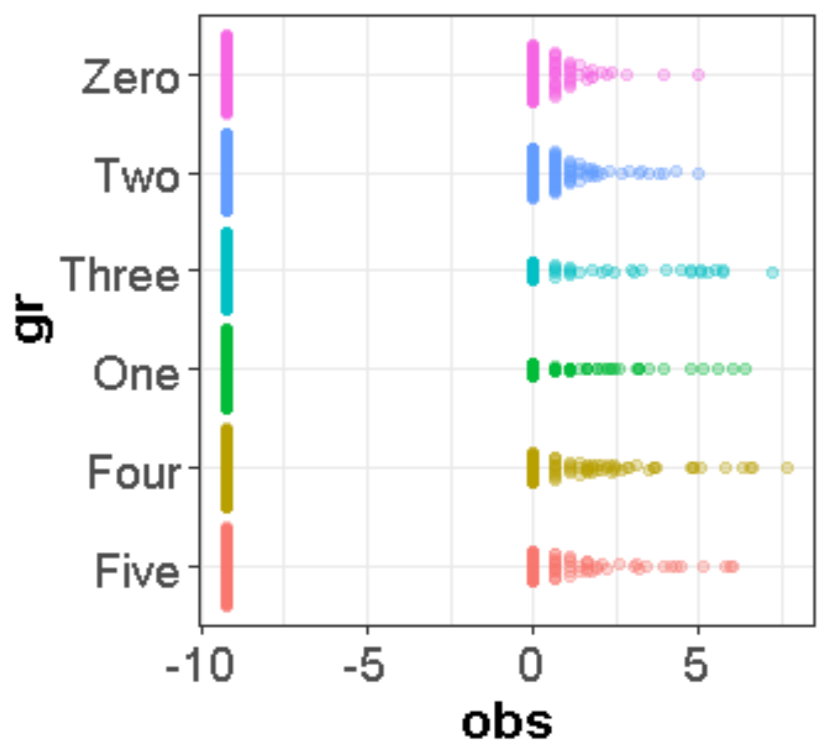
Topic 5 is the least represented above, containing 249 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 175.82, df = 5, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

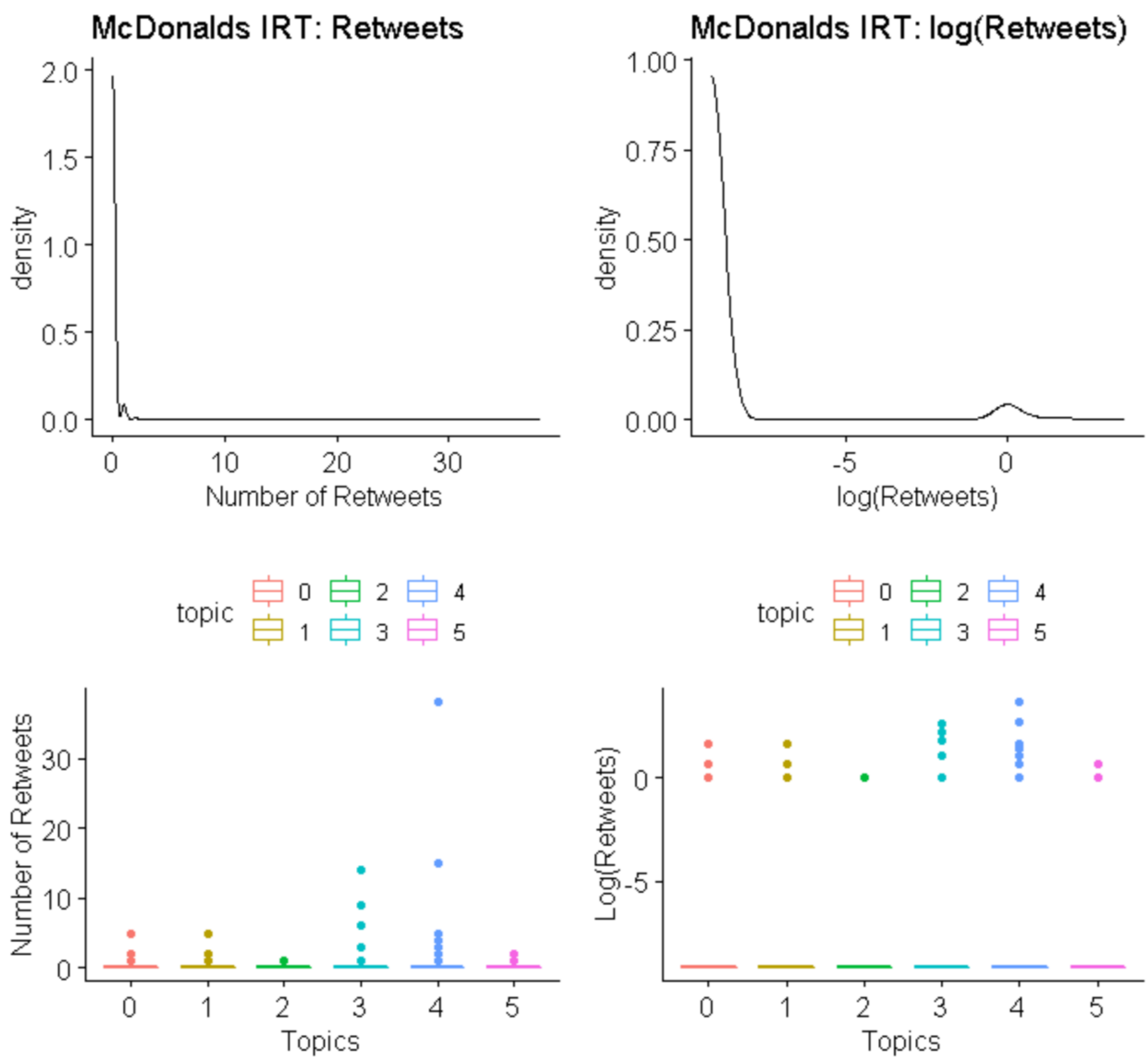
```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

   Comparison          Z      P.unadj        P.adj
1       0 - 1   9.4565998 3.181196e-21 4.771794e-20
2       0 - 2   0.2003528 8.412047e-01 1.000000e+00
3       1 - 2 -10.5225064 6.803910e-26 1.020587e-24
4       0 - 3   4.7244098 2.307846e-06 3.461769e-05
5       1 - 3  -3.1627870 1.562666e-03 2.343999e-02
6       2 - 3   4.9028160 9.447245e-07 1.417087e-05
7       0 - 4   2.2113453 2.701194e-02 4.051791e-01
8       1 - 4  -7.1459547 8.937253e-13 1.340588e-11
9       2 - 4   2.2082630 2.722594e-02 4.083892e-01
10      3 - 4  -2.7492578 5.973039e-03 8.959558e-02
11      0 - 5   1.5514971 1.207826e-01 1.000000e+00
12      1 - 5  -6.7898876 1.122209e-11 1.683314e-10
13      2 - 5   1.4817064 1.384184e-01 1.000000e+00
14      3 - 5  -2.9753465 2.926578e-03 4.389867e-02
15      4 - 5  -0.4786623 6.321789e-01 1.000000e+00
```

We may reject the null for a number of (i, j) topic group pairings. Performing a shift function to further examine these differences yields the following (confidence intervals stored in 'Topic_Appendix'):



There are too many topic group combinations to comment on every single shift, but plenty are significant (non-zero). It seems that some of these topic groups (one and three, for example) are much less likely to receive a non-zero amount of likes than tweets belonging to other topic groups. **Therefore, we may conclude that statistically significant differences exist between the quantiles of certain topic group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of McDonalds IRT tweets.**

McDonalds IRT Topics: Number of Retweets

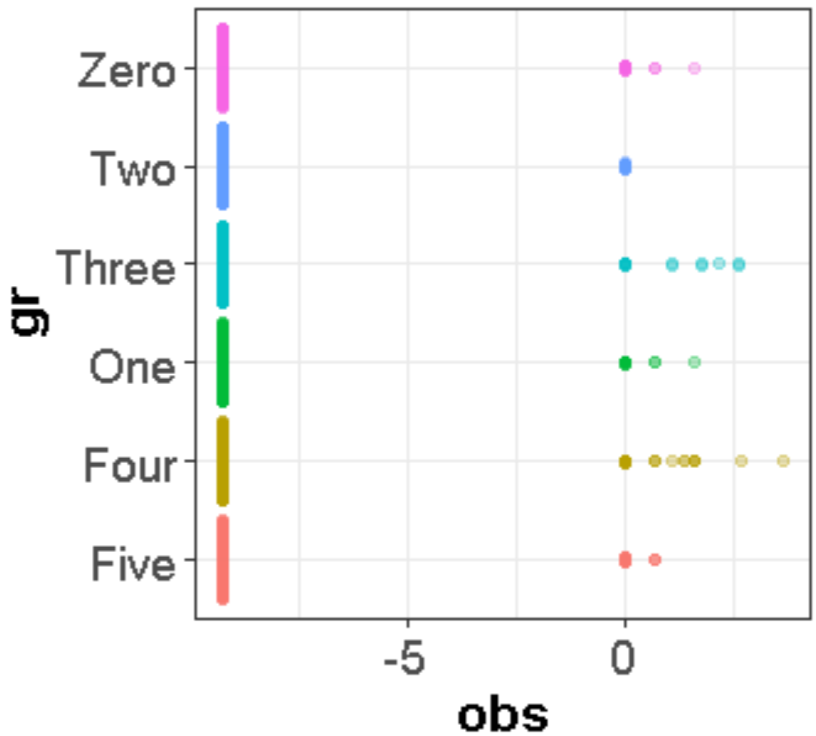


```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 34.737, df = 5, p-value = 1.698e-06
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘retweet’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

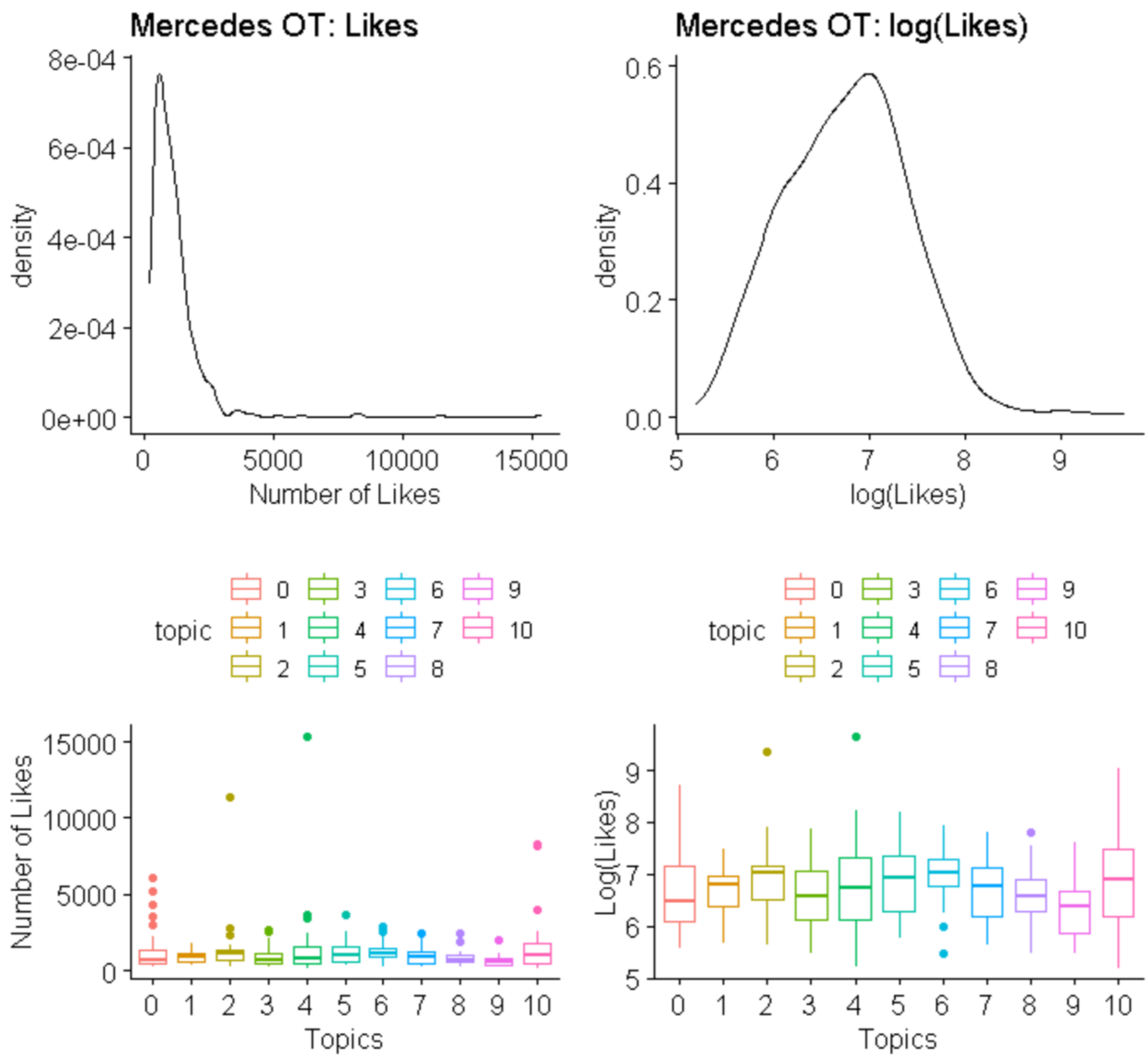
```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

   Comparison           Z       P.unadj       P.adj
1       0 - 1  3.70440502 2.118875e-04 0.003178312
2       0 - 2  0.31142502 7.554775e-01 1.000000000
3       1 - 2 -3.82150704 1.326386e-04 0.001989579
4       0 - 3  0.82073296 4.117984e-01 1.000000000
5       1 - 3 -2.47772774 1.322220e-02 0.198332996
6       2 - 3  0.58937811 5.556077e-01 1.000000000
7       0 - 4 -0.06461638 9.484794e-01 1.000000000
8       1 - 4 -3.96207868 7.430004e-05 0.001114501
9       2 - 4 -0.39346183 6.939784e-01 1.000000000
10      3 - 4 -0.90537332 3.652676e-01 1.000000000
11      0 - 5 -0.59637365 5.509256e-01 1.000000000
12      1 - 5 -4.09792962 4.168620e-05 0.000625293
13      2 - 5 -0.93113319 3.517847e-01 1.000000000
14      3 - 5 -1.34864947 1.774496e-01 1.000000000
15      4 - 5 -0.55201595 5.809374e-01 1.000000000
```

We may reject the null for a number of (i, j) topic group pairings. Performing a shift function to further examine these differences yields the following (confidence intervals stored in ‘Topic_Appendix’):



When shifts were significant, they were confined to the far-right tails only, and they were of very small magnitudes. Given that information, as well as the visualizations produced above, I don't feel comfortable drawing any conclusions. **BTM topic modeling does not seem to uncover subject matter or underlying themes within McDonalds IRT tweets which help further explain their expected number of retweets.**

Mercedes Official Topics: Number of Likes



Topic 9 is the least represented above, containing only 23 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 45.09, df = 10, p-value = 2.095e-06
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

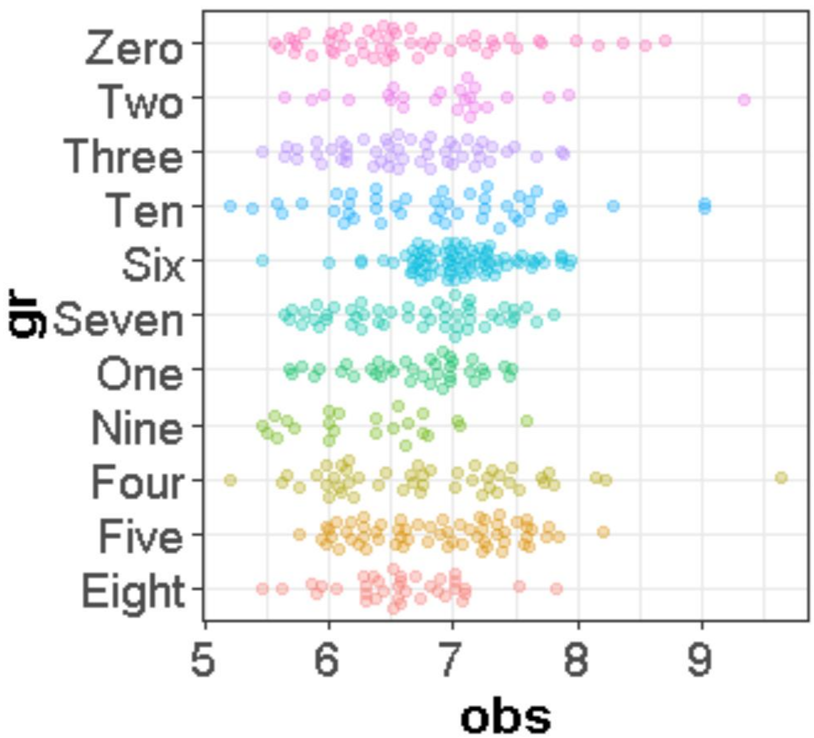
```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
```

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	0 - 1	-0.40345954	6.866102e-01	1.000000e+00
2	0 - 10	-1.81990378	6.877366e-02	1.000000e+00
3	1 - 10	-1.28462628	1.989229e-01	1.000000e+00
4	0 - 2	-1.73321098	8.305818e-02	1.000000e+00
5	1 - 2	-1.30451310	1.920587e-01	1.000000e+00
6	10 - 2	-0.22479462	8.221391e-01	1.000000e+00
7	0 - 3	-0.05650830	9.549369e-01	1.000000e+00
8	1 - 3	0.34991421	7.264031e-01	1.000000e+00
9	10 - 3	1.75670163	7.896868e-02	1.000000e+00
10	2 - 3	1.68240376	9.249057e-02	1.000000e+00
11	0 - 4	-1.07263996	2.834327e-01	1.000000e+00
12	1 - 4	-0.59040271	5.549207e-01	1.000000e+00
13	10 - 4	0.74474733	4.564245e-01	1.000000e+00
14	2 - 4	0.84559439	3.977791e-01	1.000000e+00
15	3 - 4	-1.01217813	3.114529e-01	1.000000e+00
16	0 - 5	-1.99575784	4.596029e-02	1.000000e+00
17	1 - 5	-1.39914365	1.617699e-01	1.000000e+00
18	10 - 5	-0.03287993	9.737703e-01	1.000000e+00
19	2 - 5	0.20899386	8.344530e-01	1.000000e+00
20	3 - 5	-1.92602507	5.410124e-02	1.000000e+00
21	4 - 5	-0.83358423	4.045153e-01	1.000000e+00
22	0 - 6	-3.97010207	7.184185e-05	3.951302e-03
23	1 - 6	-3.13744841	1.704253e-03	9.373390e-02
24	10 - 6	-1.81505719	6.951512e-02	1.000000e+00
25	2 - 6	-1.20756085	2.272162e-01	1.000000e+00
26	3 - 6	-3.88343114	1.029927e-04	5.664599e-03
27	4 - 6	-2.68750338	7.198837e-03	3.959361e-01
28	5 - 6	-1.95251461	5.087714e-02	1.000000e+00
29	0 - 7	-0.52068203	6.025883e-01	1.000000e+00
30	1 - 7	-0.07819105	9.376761e-01	1.000000e+00
31	10 - 7	1.29827630	1.941924e-01	1.000000e+00
32	2 - 7	1.30394840	1.922512e-01	1.000000e+00
33	3 - 7	-0.46220338	6.439355e-01	1.000000e+00
34	4 - 7	0.55213963	5.808527e-01	1.000000e+00
35	5 - 7	1.43099251	1.524324e-01	1.000000e+00
36	6 - 7	3.34578136	8.205107e-04	4.512809e-02
37	0 - 8	0.49910075	6.177084e-01	1.000000e+00
38	1 - 8	0.83626400	4.030064e-01	1.000000e+00
39	10 - 8	2.15046878	3.151815e-02	1.000000e+00
40	2 - 8	2.04034869	4.131561e-02	1.000000e+00
41	3 - 8	0.54849877	5.833495e-01	1.000000e+00
42	4 - 8	1.46978921	1.416189e-01	1.000000e+00
43	5 - 8	2.31870285	2.041115e-02	1.000000e+00
44	6 - 8	4.08531510	4.401704e-05	2.420937e-03
45	7 - 8	0.96943104	3.323302e-01	1.000000e+00
46	0 - 9	2.02302123	4.307096e-02	1.000000e+00
47	1 - 9	2.24175223	2.497739e-02	1.000000e+00
48	10 - 9	3.43450199	5.936437e-04	3.265040e-02
49	2 - 9	3.21167769	1.319623e-03	7.257928e-02
50	3 - 9	2.06139915	3.926498e-02	1.000000e+00
51	4 - 9	2.84909265	4.384411e-03	2.411426e-01
52	5 - 9	3.62420784	2.898485e-04	1.594166e-02
53	6 - 9	5.15424737	2.546516e-07	1.400584e-05
54	7 - 9	2.42239331	1.541865e-02	8.480259e-01

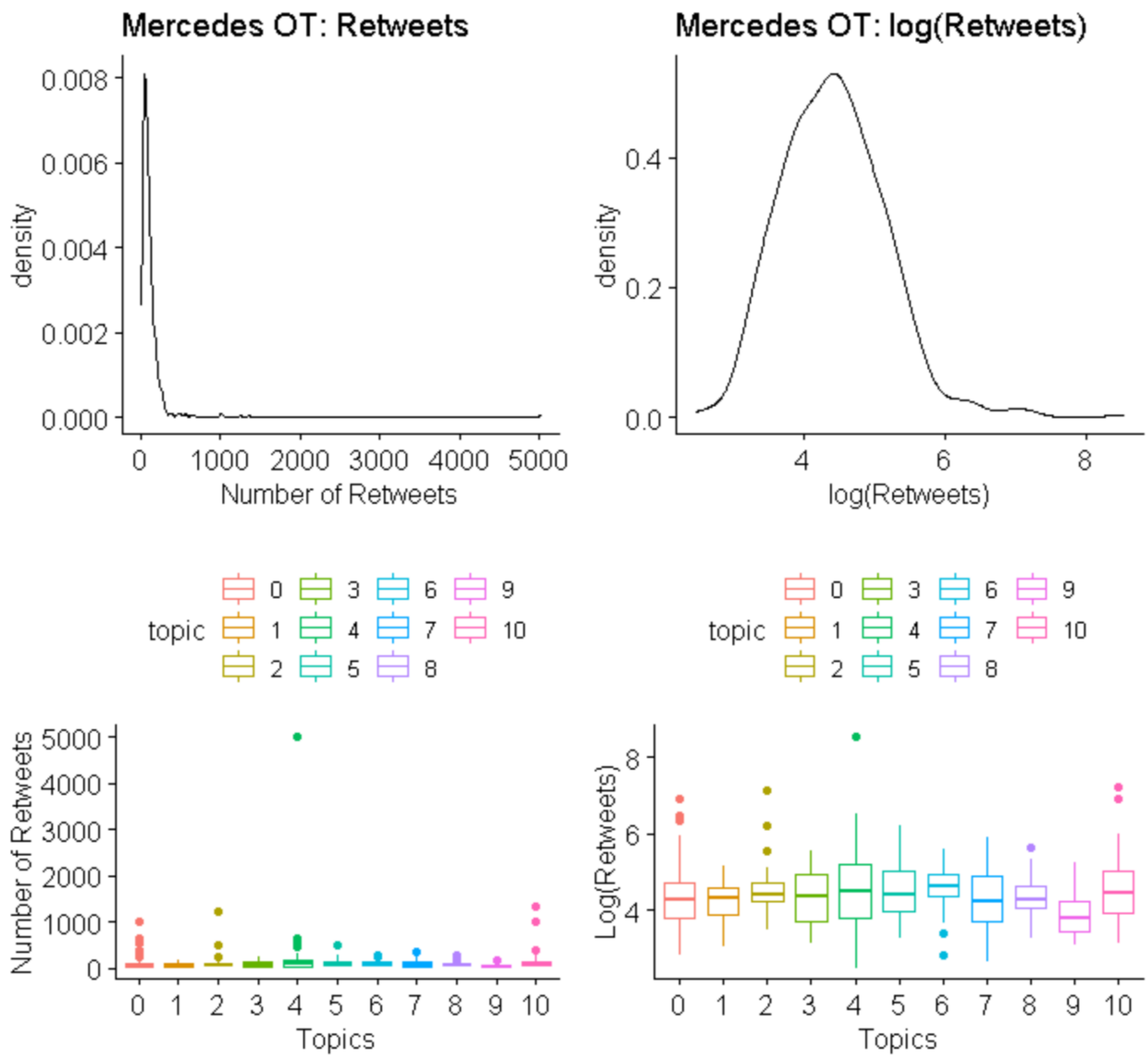
```
55        8 - 9  1.49650981 1.345209e-01 1.000000e+00
```

We may reject the null for a number of (i, j) topic group pairings. Performing a shift function to further examine these differences yields the following (confidence intervals stored in 'Topic_Appendix'):



There are too many topic group combinations to comment on each individual shift, but plenty are significant (non-zero). **Therefore, we may conclude that statistically significant differences exist between the quantiles of certain topic group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Mercedes official tweets.**

Mercedes Official Topics: Number of Retweets




```
              Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 32.759, df = 10, p-value = 0.000299
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

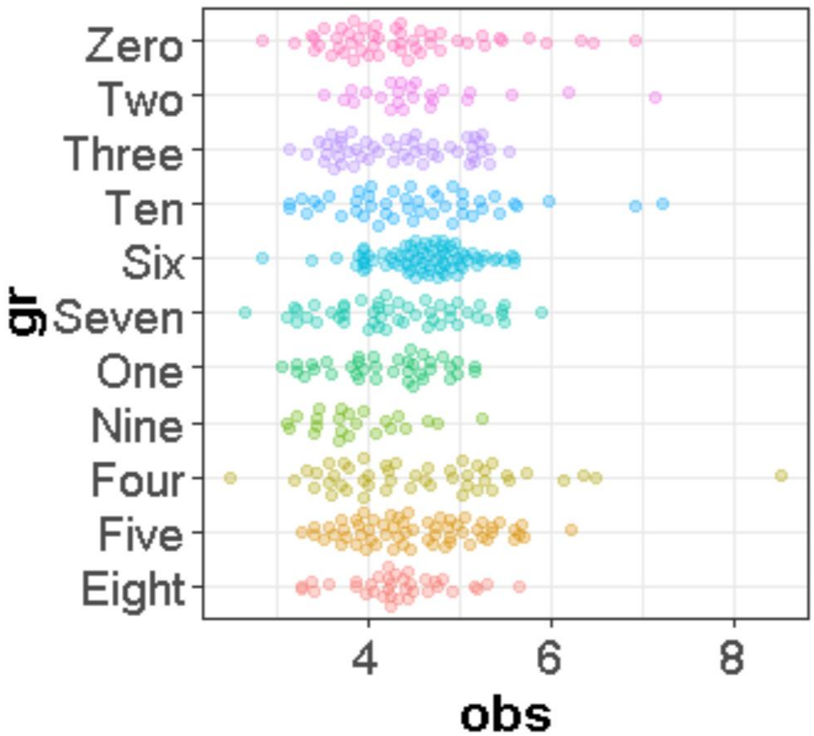
```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni"
)
Dunn (1964) Kruskal-Wallis multiple comparison
```

p-values adjusted with the Bonferroni method.

	Comparison	Z	P.unadj	P.adj
1	0 - 1	0.550647386	5.818754e-01	1.000000e+00
2	0 - 10	-1.193978221	2.324865e-01	1.000000e+00
3	1 - 10	-1.636208027	1.017961e-01	1.000000e+00
4	0 - 2	-1.179185306	2.383244e-01	1.000000e+00
5	1 - 2	-1.567389849	1.170236e-01	1.000000e+00
6	10 - 2	-0.188693565	8.503330e-01	1.000000e+00
7	0 - 3	-0.250700318	8.020458e-01	1.000000e+00
8	1 - 3	-0.778422877	4.363198e-01	1.000000e+00
9	10 - 3	0.945605817	3.443497e-01	1.000000e+00
10	2 - 3	0.974511282	3.298027e-01	1.000000e+00
11	0 - 4	-1.205732228	2.279208e-01	1.000000e+00
12	1 - 4	-1.650329213	9.887562e-02	1.000000e+00
13	10 - 4	0.001914703	9.984723e-01	1.000000e+00
14	2 - 4	0.191697413	8.479792e-01	1.000000e+00
15	3 - 4	-0.954410593	3.398758e-01	1.000000e+00
16	0 - 5	-1.547833031	1.216625e-01	1.000000e+00
17	1 - 5	-1.980070783	4.769558e-02	1.000000e+00
18	10 - 5	-0.251839831	8.011649e-01	1.000000e+00
19	2 - 5	-0.007445551	9.940594e-01	1.000000e+00
20	3 - 5	-1.277933655	2.012728e-01	1.000000e+00
21	4 - 5	-0.257038703	7.971489e-01	1.000000e+00
22	0 - 6	-2.813628467	4.898581e-03	2.694219e-01
23	1 - 6	-3.133993350	1.724448e-03	9.484462e-02
24	10 - 6	-1.391717510	1.640080e-01	1.000000e+00
25	2 - 6	-0.908152328	3.637977e-01	1.000000e+00
26	3 - 6	-2.519100602	1.176550e-02	6.471027e-01
27	4 - 6	-1.413330674	1.575585e-01	1.000000e+00
28	5 - 6	-1.234414724	2.170484e-01	1.000000e+00
29	0 - 7	-0.144386625	8.851952e-01	1.000000e+00
30	1 - 7	-0.678597666	4.973928e-01	1.000000e+00
31	10 - 7	1.042828358	2.970278e-01	1.000000e+00
32	2 - 7	1.055631735	2.911365e-01	1.000000e+00
33	3 - 7	0.104520940	9.167560e-01	1.000000e+00
34	4 - 7	1.052634482	2.925085e-01	1.000000e+00
35	5 - 7	1.380349957	1.674789e-01	1.000000e+00
36	6 - 7	2.619448649	8.807203e-03	4.843962e-01
37	0 - 8	-0.063140320	9.496548e-01	1.000000e+00
38	1 - 8	-0.566659635	5.709455e-01	1.000000e+00
39	10 - 8	1.030615639	3.027211e-01	1.000000e+00
40	2 - 8	1.055494878	2.911991e-01	1.000000e+00
41	3 - 8	0.165380756	8.686443e-01	1.000000e+00
42	4 - 8	1.038698843	2.989448e-01	1.000000e+00
43	5 - 8	1.331979058	1.828671e-01	1.000000e+00
44	6 - 8	2.442268849	1.459527e-02	8.027400e-01
45	7 - 8	0.068964341	9.450180e-01	1.000000e+00
46	0 - 9	2.554377877	1.063777e-02	5.850776e-01
47	1 - 9	1.968410933	4.902077e-02	1.000000e+00
48	10 - 9	3.455740082	5.487842e-04	3.018313e-02
49	2 - 9	3.199061956	1.378755e-03	7.583154e-02
50	3 - 9	2.744506322	6.060199e-03	3.333109e-01
51	4 - 9	3.479311725	5.027035e-04	2.764869e-02
52	5 - 9	3.822302978	1.322111e-04	7.271611e-03
53	6 - 9	4.844156594	1.271506e-06	6.993284e-05
54	7 - 9	2.652599735	7.987453e-03	4.393099e-01

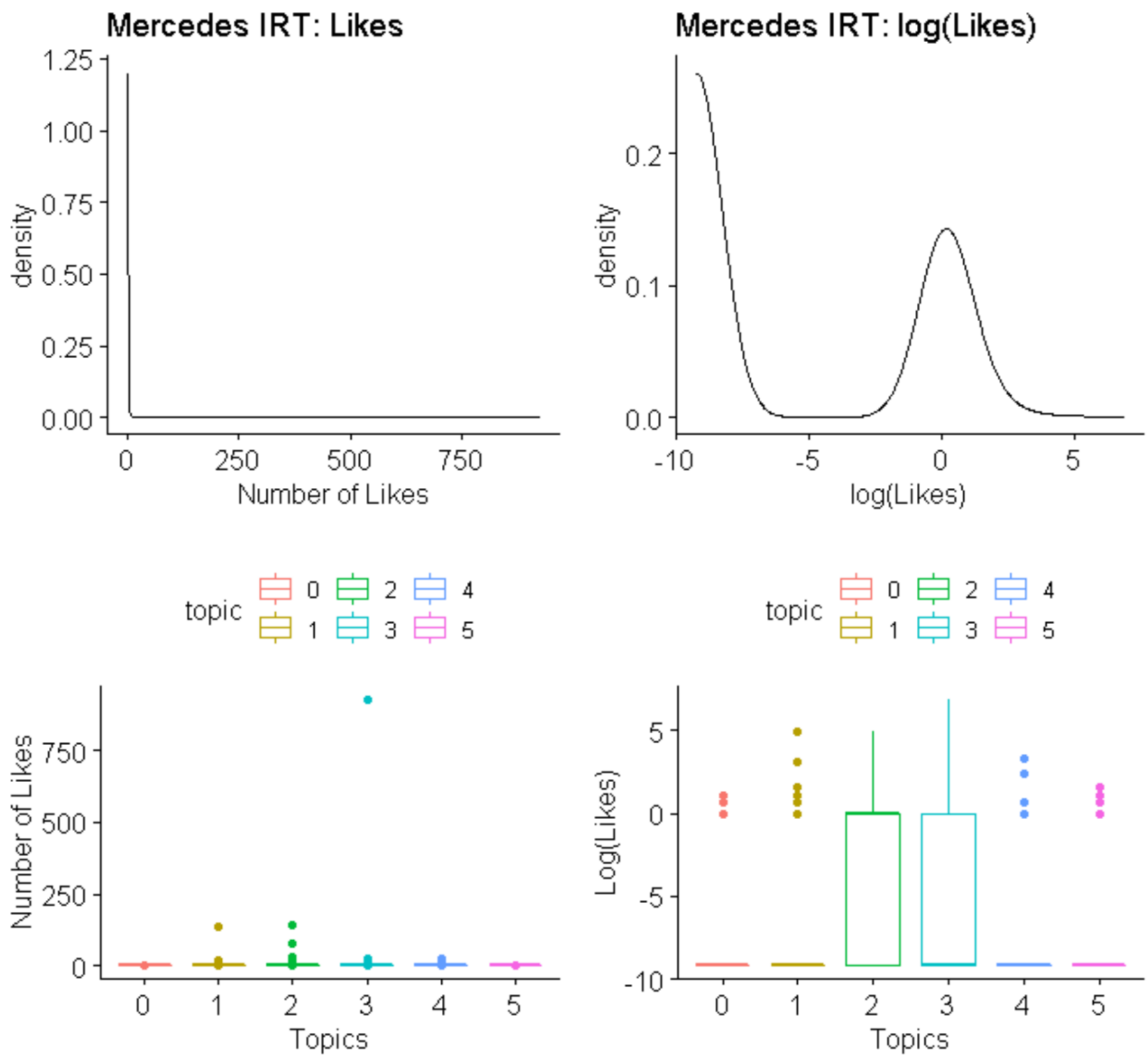
```
55       8 - 9  2.454973152 1.408951e-02 7.749231e-01
```

We may reject the null only for the (10, 9), (4, 9), (5, 9), and (6, 9) topic group pairings and conclude that their retweet distributions differ from one another. Performing a shift function to further examine these differences yields the following (confidence intervals stored in 'Topic_Appendix'):



Each topic group pair mentioned above is associated with statistically significant (non-zero) shifts at certain quantiles. **Therefore, we may conclude that statistically significant differences exist between the quantiles of certain topic group pairings (in their distributions of retweets), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Mercedes official tweets.** The main takeaway here seems to be that topic 9 tweets performed worse (in terms of retweets) than a few other tweet topic categories (on average).

Mercedes IRT Topics: Number of Likes



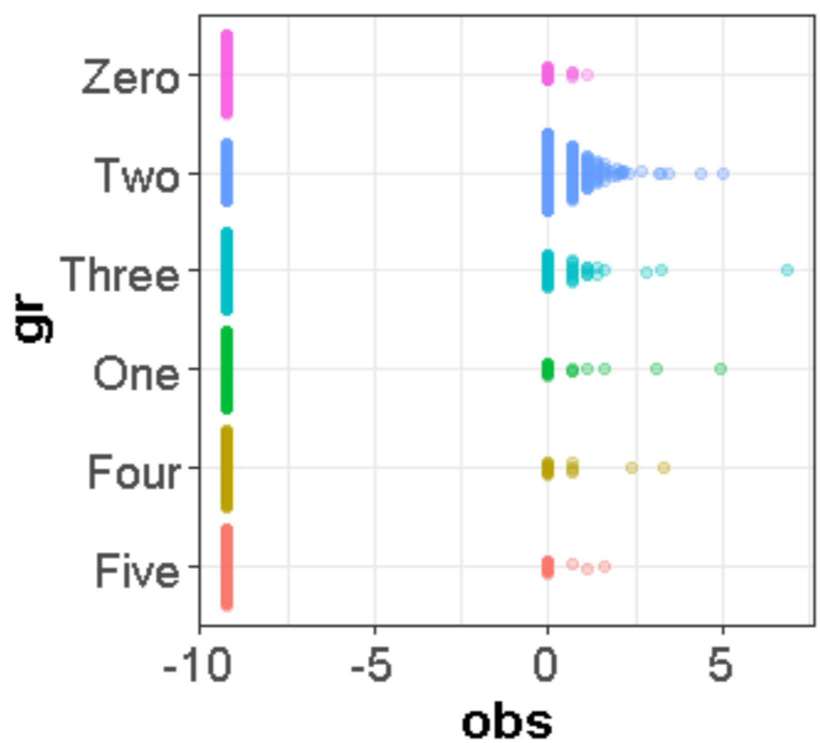
Topic 4 is the least represented above, containing 106 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 351.83, df = 5, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

   Comparison            Z       P.unadj        P.adj
1       0 - 1  -0.28212792 7.778454e-01 1.000000e+00
2       0 - 2 -14.39206309 5.803843e-47 8.705764e-46
3       1 - 2 -11.47291160 1.804826e-30 2.707240e-29
4       0 - 3  -4.34835951 1.371596e-05 2.057395e-04
5       1 - 3  -3.43610080 5.901514e-04 8.852270e-03
6       2 - 3   9.70154643 2.969630e-22 4.454446e-21
7       0 - 4  -0.70990485 4.777631e-01 1.000000e+00
8       1 - 4  -0.43002967 6.671741e-01 1.000000e+00
9       2 - 4   8.94921092 3.580325e-19 5.370487e-18
10      3 - 4   2.44777953 1.437396e-02 2.156094e-01
11      0 - 5  -0.04785722 9.618300e-01 1.000000e+00
12      1 - 5   0.19916005 8.421375e-01 1.000000e+00
13      2 - 5  11.05130076 2.160613e-28 3.240920e-27
14      3 - 5   3.49596908 4.723434e-04 7.085151e-03
15      4 - 5   0.59407225 5.524638e-01 1.000000e+00
```

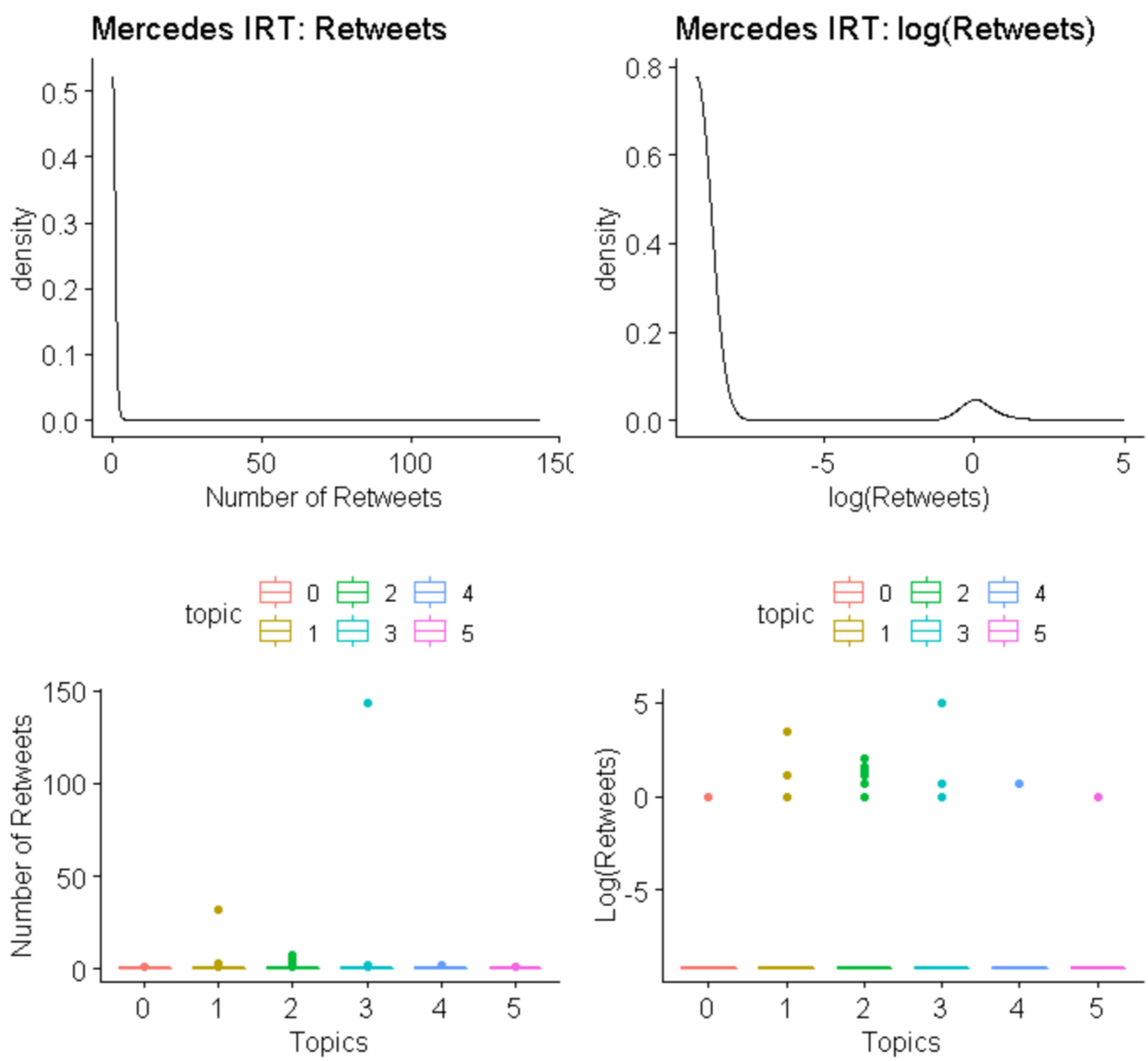
We may reject the null for a number of (i, j) topic group pairings. Performing a shift function to further examine these differences yields the following (confidence intervals stored in 'Topic_Appendix'):



There are too many topic group combinations to comment on each individual shift, but plenty are significant (non-zero). **Therefore, we may conclude that statistically significant differences exist**

between the quantiles of certain topic group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Mercedes IRT tweets.

Mercedes IRT Topics: Number of Retweets

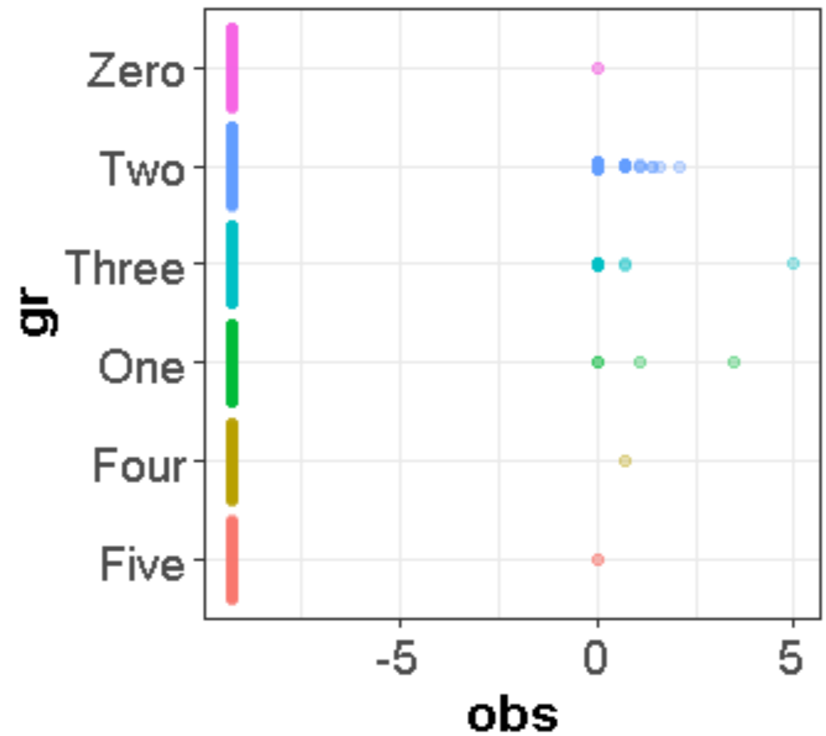



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 70.996, df = 5, p-value = 6.358e-14
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

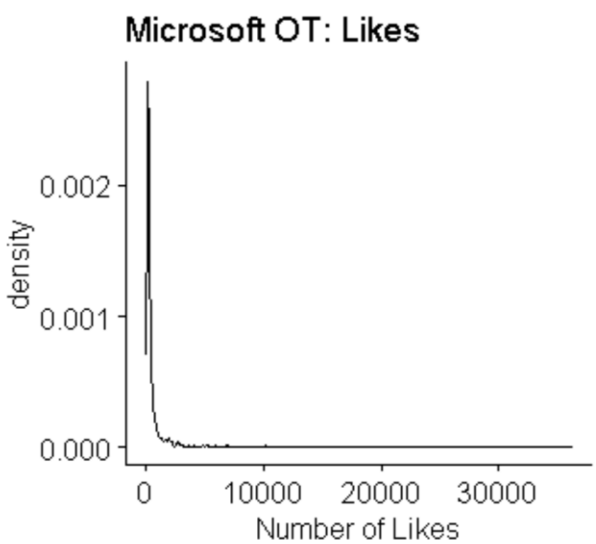
```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

   Comparison          Z       P.unadj        P.adj
1       0 - 1 -0.7202318 4.713823e-01 1.000000e+00
2       0 - 2 -6.6955619 2.148444e-11 3.222666e-10
3       1 - 2 -4.6709765 2.997713e-06 4.496569e-05
4       0 - 3 -2.4545805 1.410491e-02 2.115736e-01
5       1 - 3 -1.3677849 1.713794e-01 1.000000e+00
6       2 - 3  3.9936175 6.507281e-05 9.760921e-04
7       0 - 4 -0.1034907 9.175736e-01 1.000000e+00
8       1 - 4  0.4701823 6.382248e-01 1.000000e+00
9       2 - 4  4.4108747 1.029539e-05 1.544309e-04
10      3 - 4  1.6831495 9.234616e-02 1.000000e+00
11      0 - 5 -0.2625574 7.928918e-01 1.000000e+00
12      1 - 5  0.3820196 7.024468e-01 1.000000e+00
13      2 - 5  4.8728175 1.100179e-06 1.650268e-05
14      3 - 5  1.7337393 8.296435e-02 1.000000e+00
15      4 - 5 -0.1175287 9.064411e-01 1.000000e+00
```

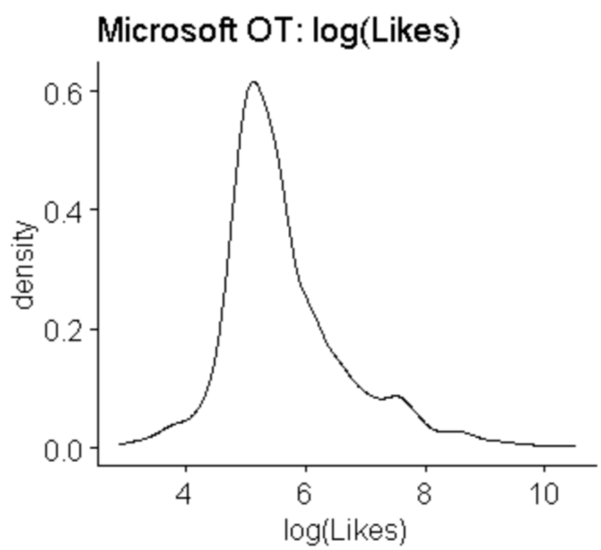
We may reject the null for each (i, 2) topic group pairing only and conclude that their distributions of retweets differ. Performing a shift function to further examine these differences yields the following (confidence intervals stored in 'Topic_Appendix'):

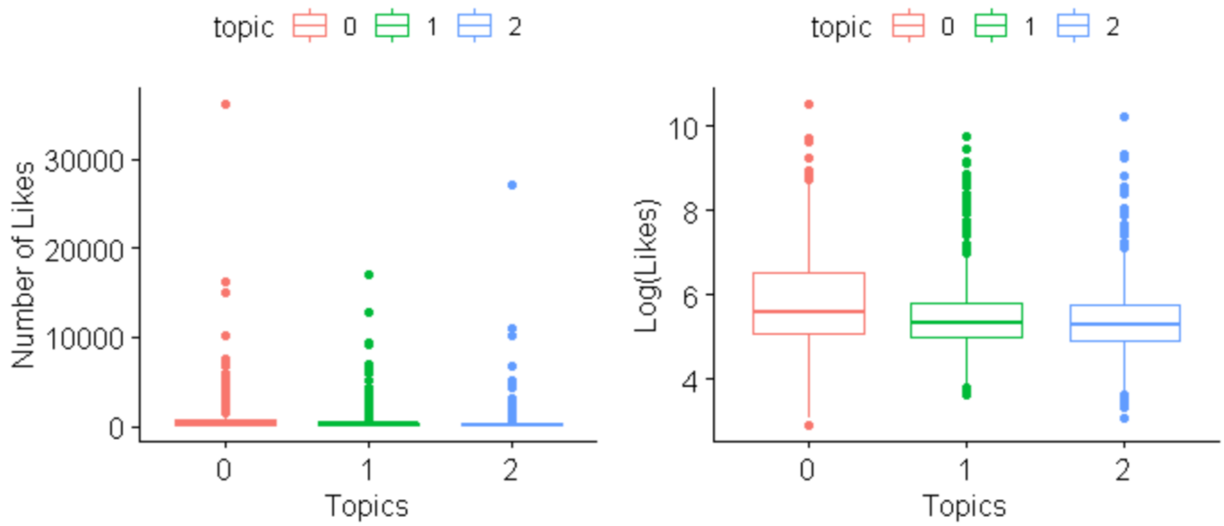


When shifts were significant, they were confined to the far-right tails only, and they were of very small magnitudes. Given that information, as well as the visualizations produced above, I don't feel comfortable drawing any conclusions. **BTM topic modeling does not seem to uncover subject matter or underlying themes within Mercedes IRT tweets which help further explain their expected number of retweets.**

Microsoft Official Topics: Number of Likes







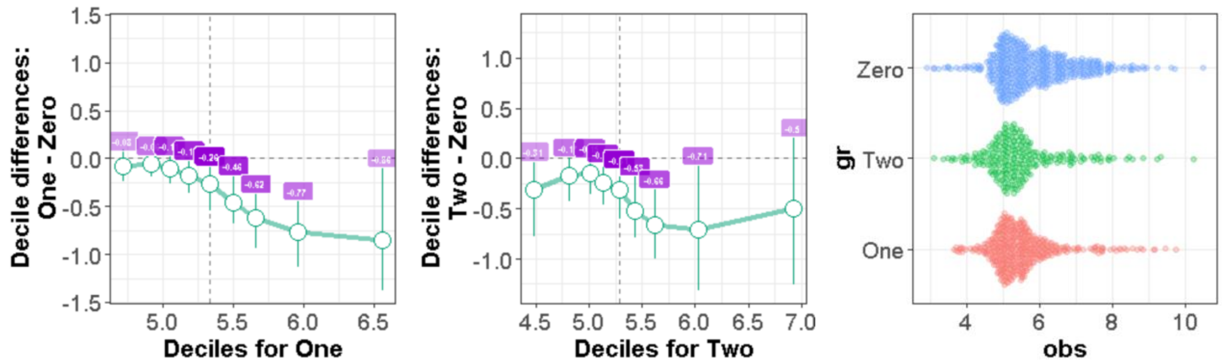
Topic 2 is the least represented above, containing 252 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 35.418, df = 2, p-value = 2.037e-08
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison        Z       P.unadj        P.adj
1      0 - 1 4.904496 9.366730e-07 2.810019e-06
2      0 - 2 5.066131 4.059836e-07 1.217951e-06
3      1 - 2 0.880189 3.787569e-01 1.000000e+00
```

We may reject the null for the (0, 1) and (0, 2) topic group pairings and conclude that their distributions of likes differ from one another. Performing the shift function for further examination yields the following:

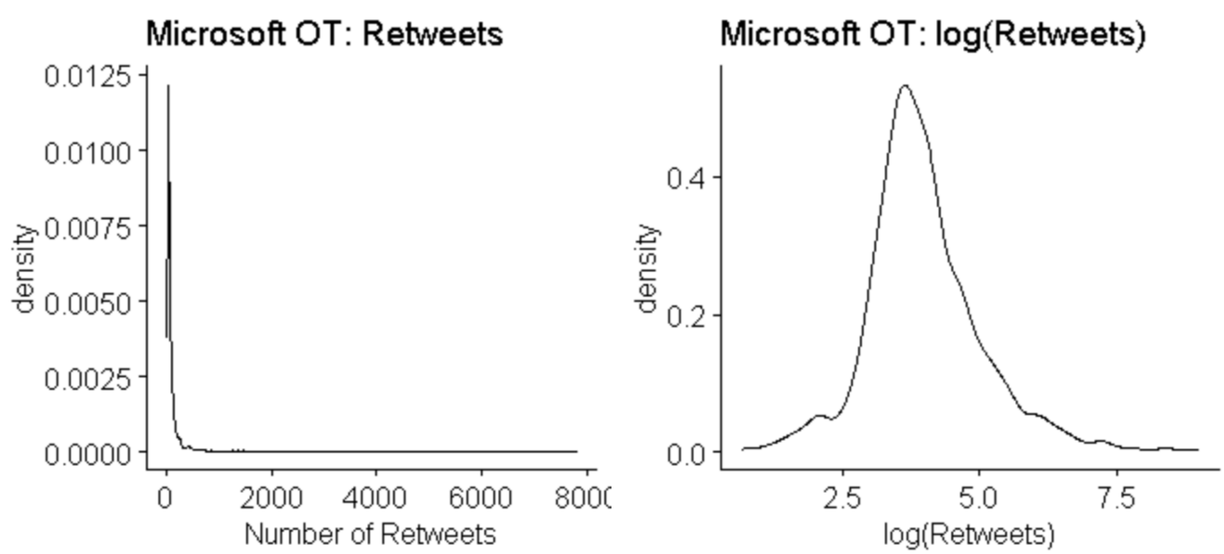


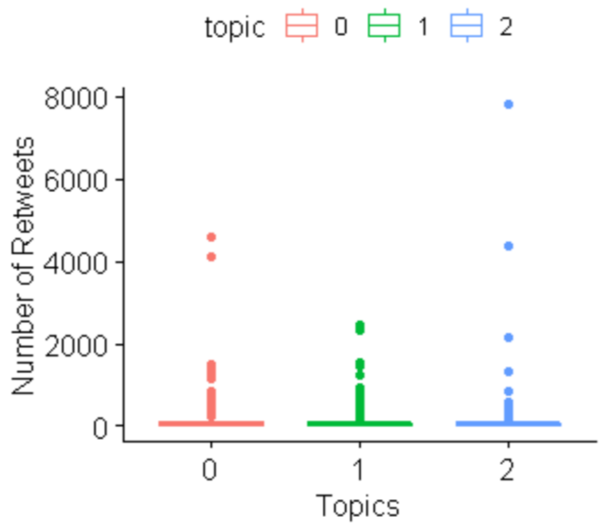

```
$`One - Two`
    q      One        Two  difference     ci_lower  ci_upper      p_crit p_va
lue
1 0.1 111.9285   89.08288   22.845666    -3.660640  44.87790 0.005555556   0.
017
2 0.2 137.0563  122.70693   14.349342    -8.519566  32.70541 0.006250000   0.
067
3 0.3 155.7576  149.94405    5.813591    -9.751344  25.50032 0.010000000   0.
313
4 0.4 178.8606  168.98430    9.876273   -16.751499  30.03409 0.012500000   0.
392
5 0.5 206.7202  197.42897    9.291204   -18.933217  39.43166 0.016666667   0.
420
6 0.6 245.1450  228.85177   16.293197   -18.945218  49.41378 0.007142857   0.
226
7 0.7 287.7613  275.44292   12.318397   -48.407232  63.05399 0.025000000   0.
592
8 0.8 389.4195  419.35617  -29.936713  -209.868416  88.67870 0.050000000   0.
669
9 0.9 720.3697 1041.27821 -320.908533 -1173.415706 364.85593 0.008333333   0.
210

$`One - Zero`
    q      One      Zero  difference    ci_lower    ci_upper       p_crit p_v
alue
1 0.1 111.9285  121.0735   -9.144977   -27.55904    7.927144 0.0083333333   0
.197
2 0.2 137.0563  145.6695   -8.613234   -25.90467    6.231580 0.0041666667   0
.100
3 0.3 155.7576  173.5658  -17.808162   -42.00094    3.621872 0.0027777778   0
.017
4 0.4 178.8606  214.5093  -35.648770   -68.97350   -4.392669 0.0016666667   0
.001
5 0.5 206.7202  268.5949  -61.874683  -138.55258  -14.219493 0.0013888889   0
.000
6 0.6 245.1450  388.2417 -143.096684  -236.36865  -39.691623 0.0011904762   0
.000
7 0.7 287.7613  535.3059 -247.544599  -407.94654 -131.414243 0.0010416667   0
.000
8 0.8 389.4195  843.5251 -454.105656  -810.37236 -228.627074 0.0009259259   0
.000
9 0.9 720.3697 1676.2860 -955.916330 -1455.85188 -135.343115 0.0020833333   0
.001
```

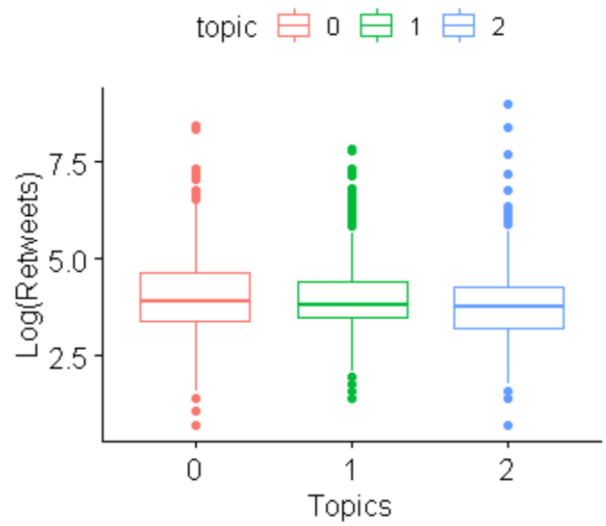
```
$`Two - Zero`
    q        Two      Zero difference    ci_lower     ci_upper       p_crit p
_value
1 0.1   89.08288  121.0735  -31.99064   -60.78775   -1.4511808 0.0005208333
0.001
2 0.2  122.70693  145.6695  -22.96258   -48.08028    1.3137364 0.0010416667
0.006
3 0.3  149.94405  173.5658  -23.62175   -57.84521    0.5945243 0.0006944444
0.003
4 0.4  168.98430  214.5093  -45.52504   -85.10797   -1.0360199 0.0004166667
0.001
5 0.5  197.42897  268.5949  -71.16589  -155.55941  -23.2475313 0.0003472222
0.000
6 0.6  228.85177  388.2417 -159.38988  -240.15158  -62.8126628 0.0002976190
0.000
7 0.7  275.44292  535.3059 -259.86300  -427.79085 -130.7611004 0.0002604167
0.000
8 0.8  419.35617  843.5251 -424.16894  -815.21942  -35.4000147 0.0002314815
0.000
9 0.9 1041.27821 1676.2860 -635.00780 -1280.87511  397.7357605 0.0020833333
0.057
```

From the middle table we can say, with 95% confidence, that the 4th through 9th quantiles of topic 0 tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of topic 1 tweets. From the bottom table we can say, with 95% confidence, that the 1st, as well as the 4th through 8th, quantiles of topic 0 tweets would need to be shifted down by significant (non-zero) amounts to match their counterparts in the set of topic 2 tweets. **Therefore, we may conclude that statistically significant differences exist between the quantiles of the above two group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Microsoft official tweets.**

Microsoft Official Topics: Number of Retweets



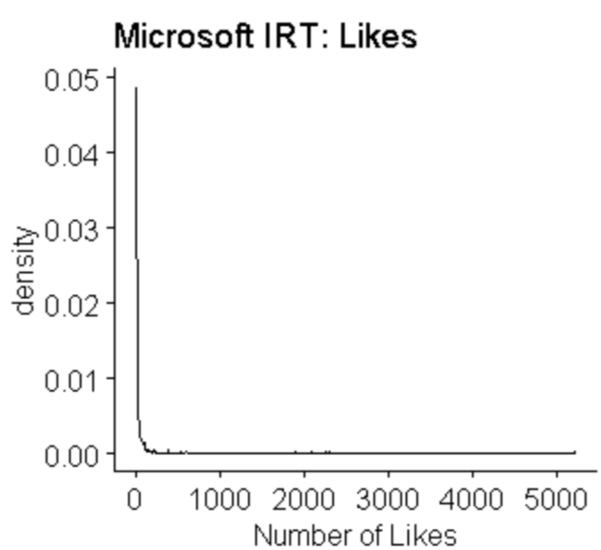


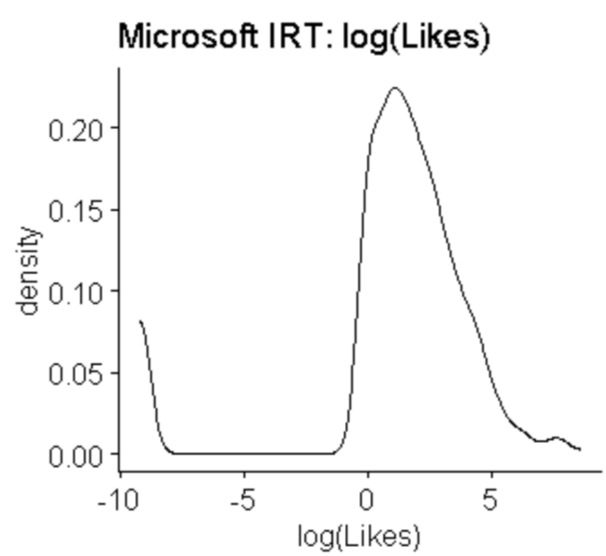


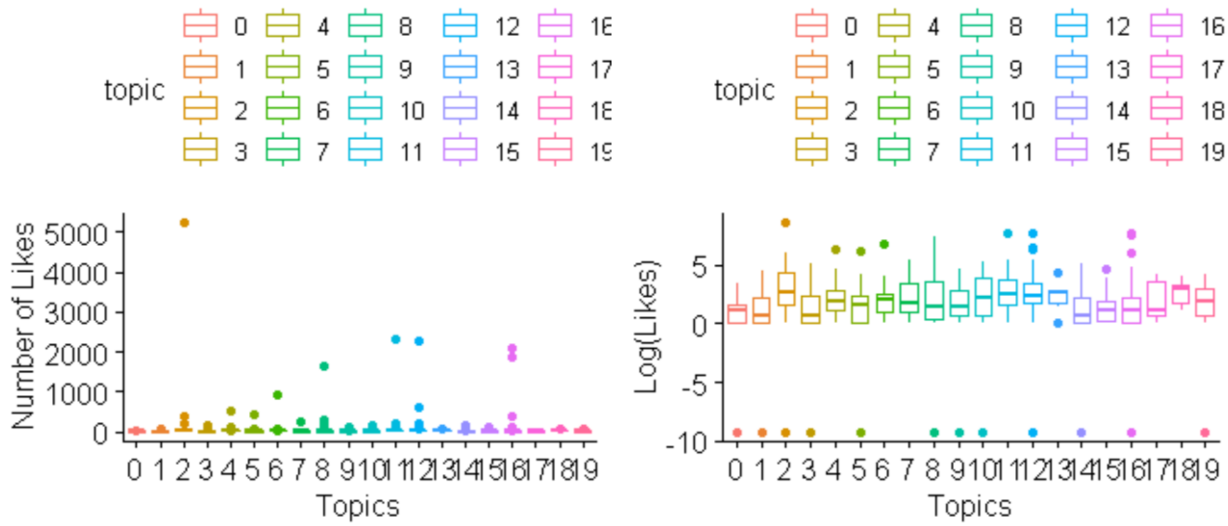

```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 5.5598, df = 2, p-value = 0.06205
```

Performing a Kruskal-Wallis test results in an insignificant p-value. Therefore, we fail to reject the null hypothesis that the 'retweet' distributions of all populations are equal. **BTM topic modeling does not seem to uncover subject matter or underlying themes within Microsoft official tweets which help further explain their expected number of retweets.**

Microsoft IRT Topics: Number of Likes





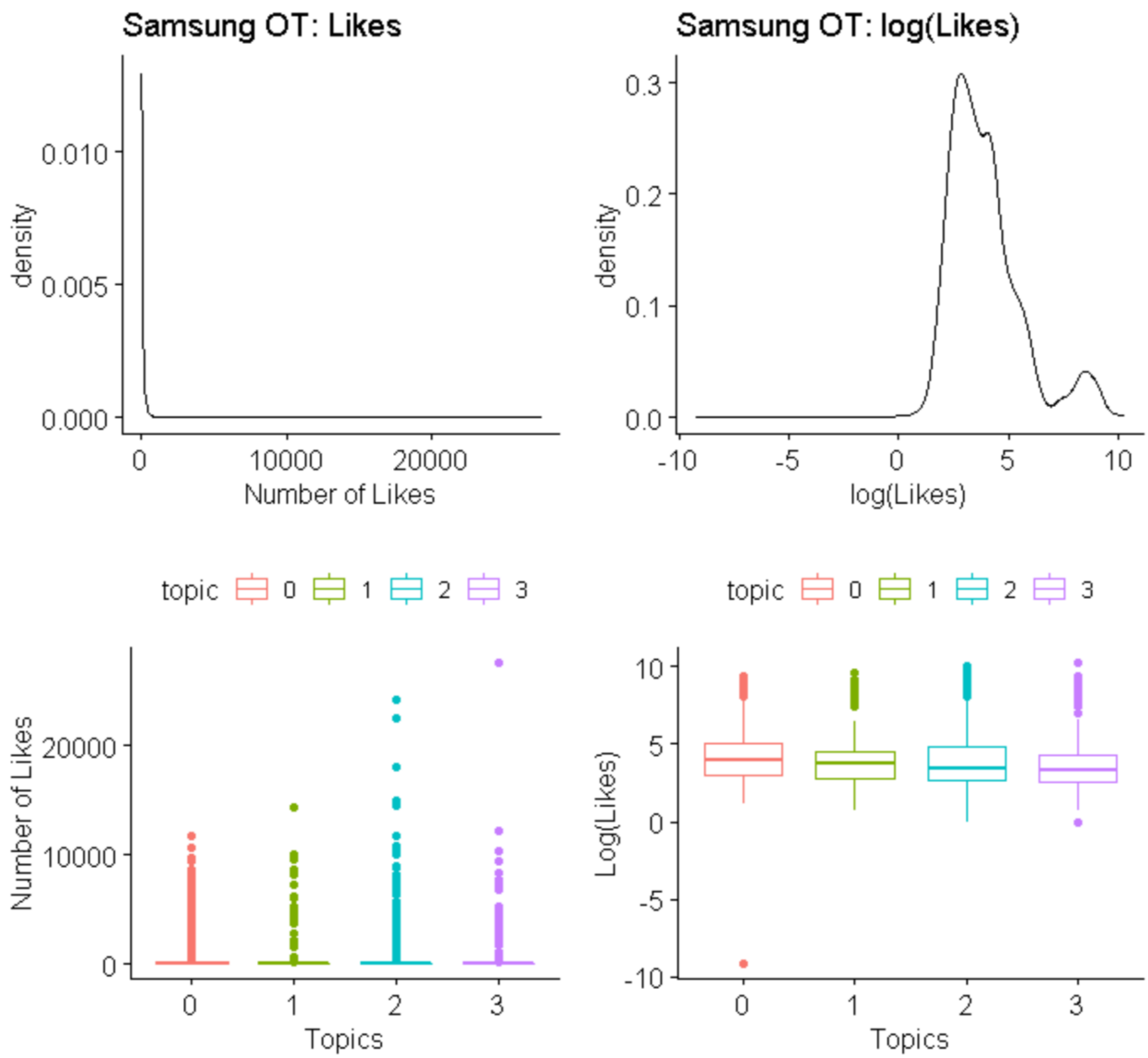



```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 54.413, df = 19, p-value = 2.858e-05
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

After further examination, there are too many topics and too few tweets for honest analysis here. Insufficient sample sizes.

Samsung Official Topics: Number of Likes



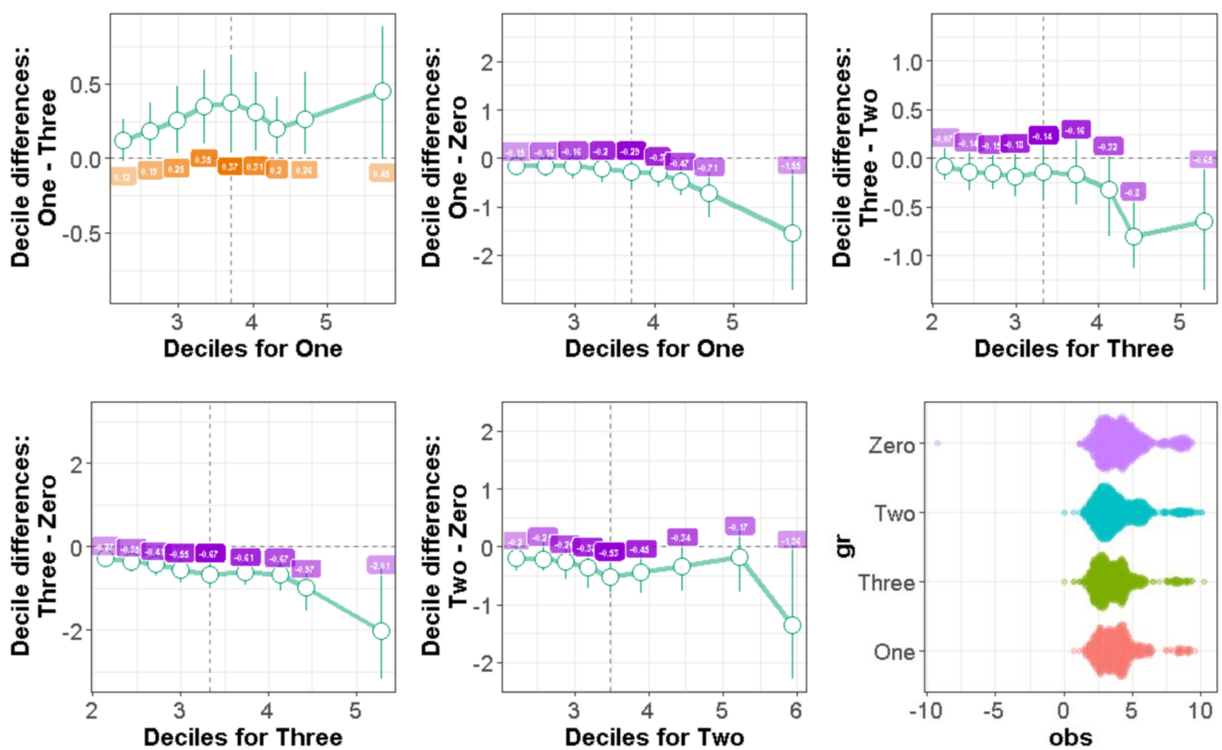
Topic 1 is the least represented above, containing 534 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 58.735, df = 3, p-value = 1.095e-12
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'like' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison         Z       P.unadj        P.adj
1      0 - 1 3.4535798 5.531987e-04 3.319192e-03
2      0 - 2 4.6162170 3.907983e-06 2.344790e-05
3      1 - 2 0.6031262 5.464248e-01 1.000000e+00
4      0 - 3 7.6324976 2.302488e-14 1.381493e-13
5      1 - 3 3.6623972 2.498661e-04 1.499196e-03
6      2 - 3 3.5571701 3.748714e-04 2.249229e-03
```

We may reject the null for every topic group pairing besides the (1, 2) topic group pair and conclude that each pairs distribution of likes differ from one another. Performing a shift function to further examine these differences yields the following:




```
$`One - Three`
    q       One     Three difference  ci_lower  ci_upper      p_crit p_va
lue
1 0.1  9.590497  8.506501   1.083996 -0.0475631  2.546247 0.025000000   0.
036
2 0.2 13.810438 11.459283   2.351155  0.3133488  4.727012 0.010000000   0.
003
3 0.3 19.686795 15.232105   4.454690  0.8746019  8.821553 0.012500000   0.
002
4 0.4 28.444667 19.956593   8.488074  2.8489210 15.742421 0.005555556   0.
000
```

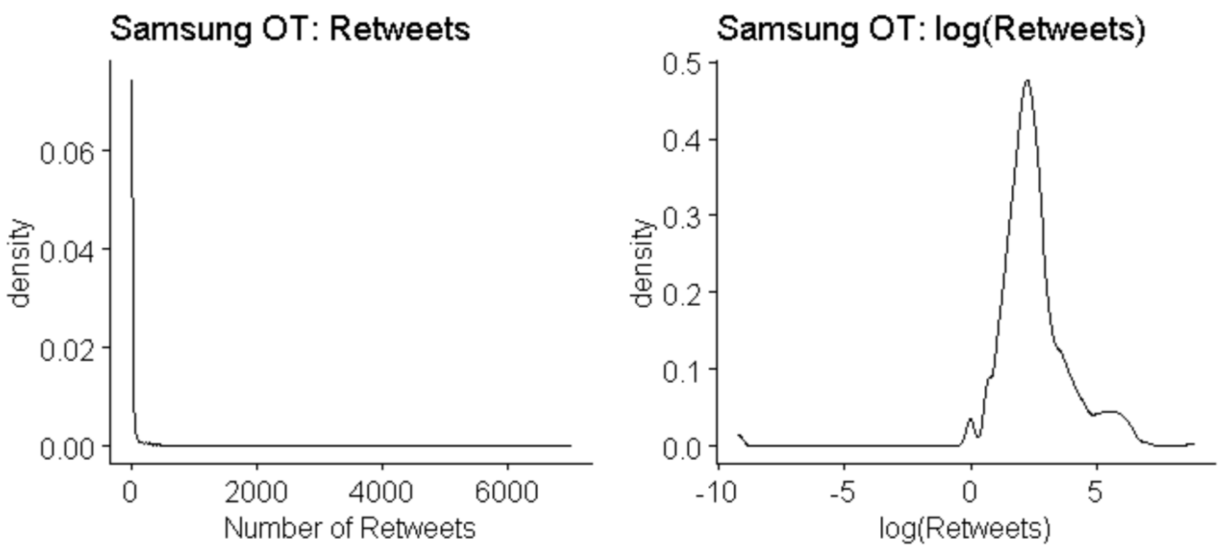
```
5 0.5  41.035780  28.254515  12.781265  2.6374997  24.846249 0.006250000   0.
000
6 0.6  57.227300  41.985874  15.241426  4.1173360  27.696326 0.008333333   0.
000
7 0.7  75.624813  62.107350  13.517463  1.5323237  28.550145 0.007142857   0.
003
8 0.8 109.809517  83.826519  25.982999  0.7656361  66.032957 0.016666667   0.
016
9 0.9 317.032809 200.796426 116.236383  0.3696298 262.363956 0.050000000   0.
050

$`One - Two`
    q        One        Two  difference    ci_lower   ci_upper      p_crit p_v
alue
1 0.1   9.590497   9.15687   0.4336274   -0.675169   1.684393 0.050000000   0
.479
2 0.2  13.810438  13.10983   0.7006058   -1.009816   2.526825 0.016666667   0
.328
3 0.3  19.686795  17.77178   1.9150183   -2.033650   6.804828 0.010000000   0
.215
4 0.4  28.444667  24.00763   4.4370336   -1.339191  11.054285 0.007142857   0
.050
5 0.5  41.035780  32.40328   8.6324952   -1.817993  18.552141 0.006250000   0
.024
6 0.6  57.227300  49.51923   7.7080687   -5.165129  20.635582 0.008333333   0
.106
7 0.7  75.624813  86.38330 -10.7584866  -39.996645  10.703662 0.012500000   0
.316
8 0.8 109.809517 186.95095 -77.1414315 -137.636832 -18.754310 0.005555556   0
.001
9 0.9 317.032809 381.01059 -63.9777828 -245.798298 106.803711 0.025000000   0
.321

$`One - Zero`
    q        One        Zero   difference     ci_lower     ci_upper      p_cri
t p_value
1 0.1   9.590497   11.18877    -1.598268    -3.187601    0.6051374 0.00833333
3   0.041
2 0.2  13.810438   16.22740    -2.416959    -5.669400    0.3895549 0.00625000
0   0.019
3 0.3  19.686795   23.01037    -3.323579    -8.401812    1.0214835 0.02500000
0   0.082
4 0.4  28.444667   34.72104    -6.276368   -15.013379    1.9319614 0.01250000
0   0.044
5 0.5  41.035780   55.00969   -13.973906   -27.663697    1.5470161 0.00500000
0   0.012
6 0.6  57.227300   77.31982   -20.092522   -40.928232   -3.5650376 0.00416666
7   0.001
7 0.7  75.624813  121.50067   -45.875859   -83.042119  -16.0183977 0.00357142
9   0.000
8 0.8 109.809517  223.33503  -113.525514  -227.819780  -36.8303491 0.00312500
0   0.000
9 0.9 317.032809 1586.59019 -1269.557380 -2992.844826 -183.6170993 0.00277777
8   0.001

$`Three - Two`
```

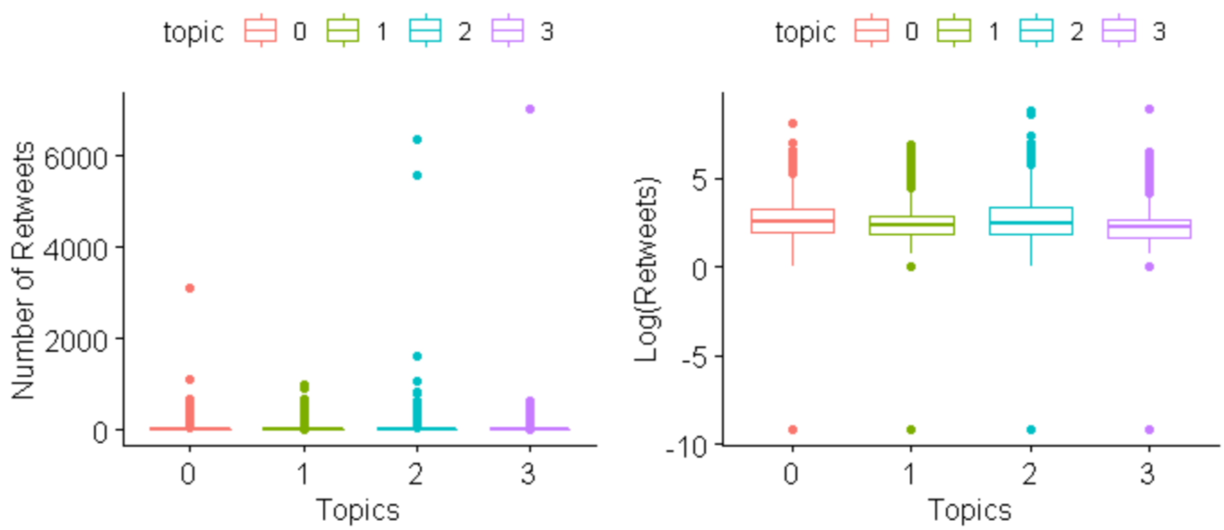
```
    q      Three        Two  difference    ci_lower    ci_upper       p_crit p
_value
1 0.1   8.506501   9.15687   -0.650369   -2.120303   0.7287803 0.0027777778
0.184
2 0.2  11.459283  13.10983   -1.650549   -3.978289   0.7178504 0.0006944444
0.020
3 0.3  15.232105  17.77178   -2.539672   -5.376128   0.3950164 0.0004629630
0.008
4 0.4  19.956593  24.00763   -4.051040   -9.058457   0.5193988 0.0005555556
0.005
5 0.5  28.254515  32.40328   -4.148770  -12.798302   4.2624310 0.0013888889
0.159
6 0.6  41.985874  49.51923   -7.533358  -23.084903   7.0926083 0.0009259259
0.120
7 0.7  62.107350  86.38330  -24.275949  -66.044615   0.6454082 0.0003968254
0.002
8 0.8  83.826519 186.95095 -103.124430 -175.947835 -49.8594691 0.0003086420
0.000
9 0.9 200.796426 381.01059 -180.214166 -503.943825 -17.3338089 0.0003472222
0.001

$`Three - Zero`
    q      Three        Zero   difference     ci_lower     ci_upper       p_cr
it p_value
1 0.1   8.506501   11.18877    -2.682264    -4.357257   -0.8012172 3.472222e-
04       0
2 0.2  11.459283   16.22740    -4.768114    -9.077848   -1.5874019 1.736111e-
04       0
3 0.3  15.232105   23.01037    -7.778269   -14.697370   -3.7905519 1.157407e-
04       0
4 0.4  19.956593   34.72104   -14.764442   -28.708054   -7.0085760 8.680556e-
05       0
5 0.5  28.254515   55.00969   -26.755171   -38.970148  -11.5957982 6.944444e-
05       0
6 0.6  41.985874   77.31982   -35.333948   -59.442534  -17.4458292 5.787037e-
05       0
7 0.7  62.107350  121.50067   -59.393322  -121.587304  -30.3305300 4.960317e-
05       0
8 0.8  83.826519  223.33503  -139.508513  -282.194198  -72.4951762 4.340278e-
05       0
9 0.9 200.796426 1586.59019 -1385.793763 -3751.561759 -228.4620425 3.858025e-
05       0

$`Two - Zero`
    q       Two        Zero   difference     ci_lower     ci_upper       p_cri
t p_value
1 0.1   9.15687   11.18877    -2.031895    -4.240992   0.43638280 1.286008e-0
5   0.003
2 0.2  13.10983   16.22740    -3.117565    -7.159694  -0.08968297 9.645062e-0
6   0.001
3 0.3  17.77178   23.01037    -5.238597   -12.511051  -1.11541042 6.430041e-0
6   0.000
4 0.4  24.00763   34.72104   -10.713402   -21.484217  -3.34686465 5.511464e-0
6   0.000
5 0.5  32.40328   55.00969   -22.606402   -35.845379  -7.91603192 4.822531e-0
6   0.000
```

```
6 0.6  49.51923   77.31982   -27.800590   -54.158565  -9.84799670 4.286694e-0
6   0.000
7 0.7  86.38330  121.50067   -35.117373   -80.582727  14.65899500 1.929012e-0
5   0.010
8 0.8 186.95095  223.33503   -36.384083  -192.306877  58.11450190 3.858025e-0
5   0.221
9 0.9 381.01059 1586.59019 -1205.579597 -3482.469371 -43.68589607 7.716049e-0
6   0.000
```

Each significant pair from Dunn's test is also associated with significant (non-zero) shifts. **Therefore, we may conclude that statistically significant differences exist between the quantiles of certain topic group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Samsung official tweets.**

Samsung Official Topics: Number of Retweets



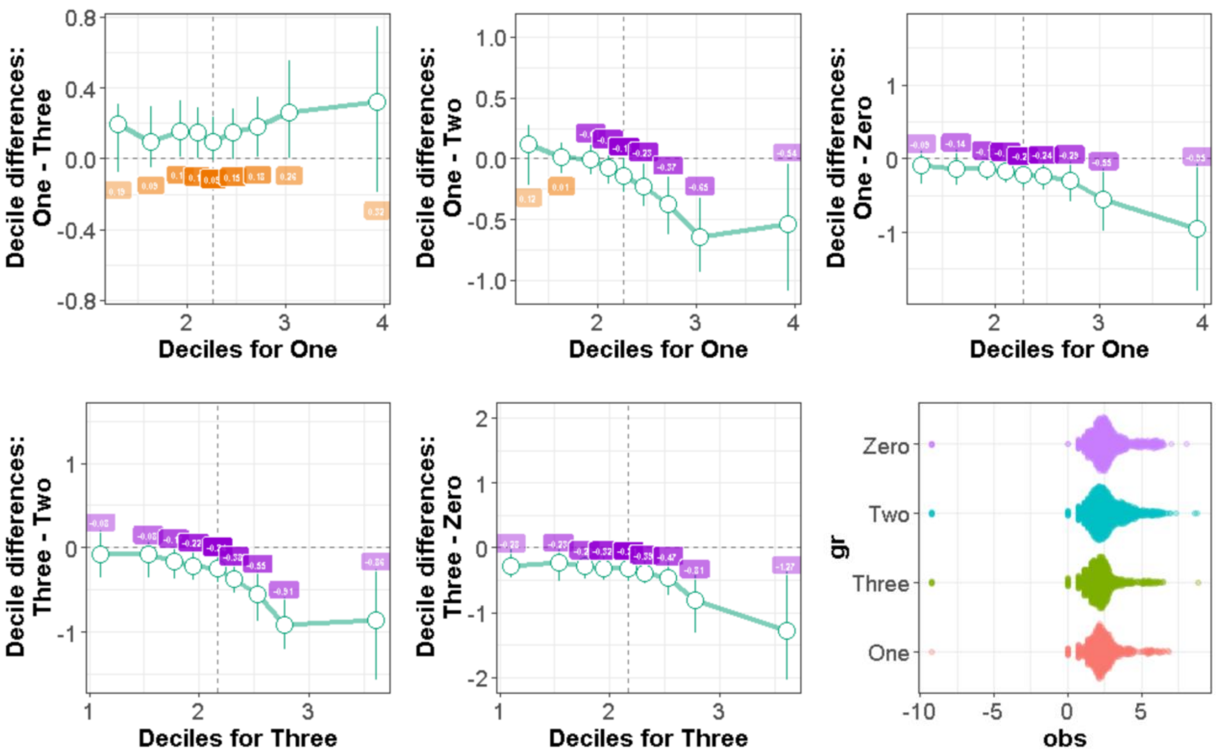



```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 63.731, df = 3, p-value = 9.37e-14
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni")
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

  Comparison         Z      P.unadj        P.adj
1      0 - 1  3.850639 1.178103e-04 7.068617e-04
2      0 - 2  1.479869 1.389081e-01 8.334488e-01
3      1 - 2 -2.726094 6.408879e-03 3.845327e-02
4      0 - 3  7.270014 3.594506e-13 2.156704e-12
5      1 - 3  2.924219 3.453215e-03 2.071929e-02
6      2 - 3  6.327107 2.497993e-10 1.498796e-09
```

We may reject the null for every topic group pairing besides the (0, 2) topic group pair and conclude that each pairs' distribution of likes differ from one another. Performing a shift function to further examine these differences yields the following:



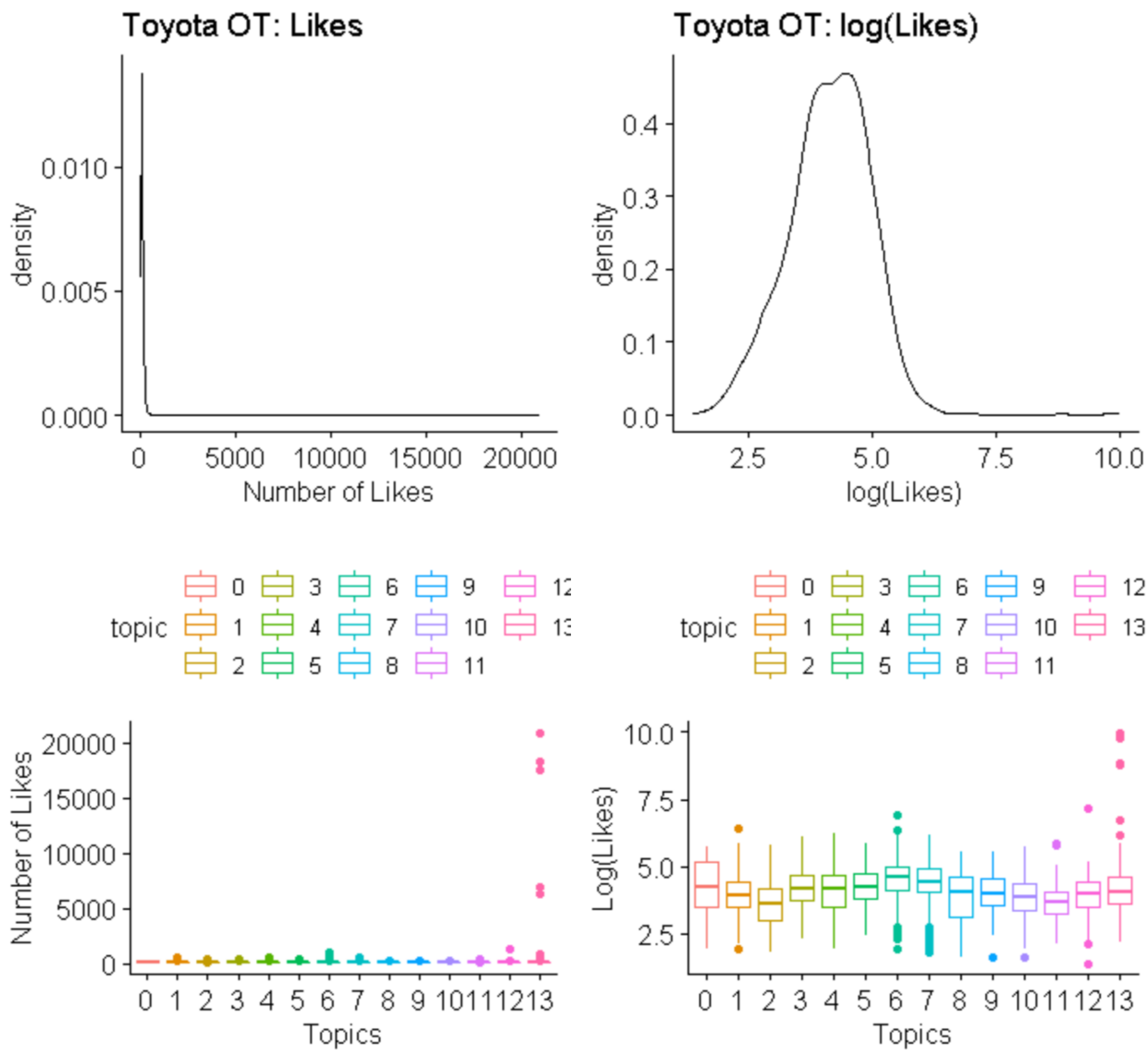

```
$`One - Three`
    q       One      Three difference    ci_lower  ci_upper      p_crit p_valu
e
1 0.1  3.677460  3.003909  0.6735513 -0.18715743  1.047555 0.016666667   0.05
4
2 0.2  5.107684  4.669126  0.4385582 -0.22729940  1.492462 0.025000000   0.12
1
3 0.3  6.883202  5.879876  1.0033260  0.08069199  2.020693 0.008333333   0.00
3
4 0.4  8.170142  7.031153  1.1389895  0.11959440  2.202693 0.005555556   0.00
3
5 0.5  9.654045  8.779621  0.8744242 -0.20569777  2.301272 0.012500000   0.04
5
6 0.6 11.782112 10.163736  1.6183765  0.07882874  3.339730 0.010000000   0.00
5
7 0.7 15.103821 12.621211  2.4826108  0.38035444  4.940928 0.006250000   0.00
2
8 0.8 20.919329 16.081776  4.8375535 -0.25950589 11.334438 0.007142857   0.01
0
9 0.9 52.046735 37.677304 14.3694316 -7.95113109 36.548488 0.050000000   0.21
6

$`One - Two`
    q       One       Two  difference   ci_lower   ci_upper      p_crit p_
value
1 0.1  3.677460  3.271689  0.40577176 -0.7847579 0.95806871 0.016666667
0.453
2 0.2  5.107684  5.054442  0.05324160 -0.6099627 0.70493428 0.050000000
0.878
3 0.3  6.883202  6.947461 -0.06425967 -0.9158173 0.69598596 0.025000000
0.798
```

```
4 0.4  8.170142  8.800770  -0.63062746  -1.7541486  0.64238652 0.012500000
0.246
5 0.5  9.654045 11.164534  -1.51048861  -3.1366311 -0.08125996 0.008333333
0.006
6 0.6 11.782112 14.801884  -3.01977233  -5.3689336 -0.72089155 0.007142857
0.001
7 0.7 15.103821 21.897544  -6.79372266 -13.0296578 -2.58064570 0.006250000
0.000
8 0.8 20.919329 39.921159 -19.00182955 -29.7599070 -9.78498445 0.005555556
0.000
9 0.9 52.046735 88.364490 -36.31775511 -75.5493189 -0.20631707 0.010000000
0.011

$`One - Zero`
    q       One       Zero  difference    ci_lower     ci_upper      p_crit p
_value
1 0.1  3.677460   3.983926  -0.3064653   -1.173274   0.47681253 0.010000000
0.253
2 0.2  5.107684   5.882595  -0.7749110   -1.946334   0.48509955 0.005000000
0.117
3 0.3  6.883202   7.865977  -0.9827757   -2.273026   0.18516310 0.003333333
0.014
4 0.4  8.170142   9.694327  -1.5241845   -2.949216   0.03524005 0.002500000
0.004
5 0.5  9.654045  12.000363  -2.3463185   -4.131879  -0.74157848 0.002000000
0.000
6 0.6 11.782112  14.969159  -3.1870466   -5.954636  -0.42510184 0.001666667
0.000
7 0.7 15.103821  20.284293  -5.1804719  -11.460660  -1.00669330 0.001428571
0.000
8 0.8 20.919329  36.430181 -15.5108513  -32.498675  -3.44456236 0.001250000
0.000
9 0.9 52.046735 135.369082 -83.3223467 -178.813318 -13.30148083 0.001111111
0.001

$`Three - Two`
    q     Three        Two  difference   ci_lower     ci_upper       p_crit p_
value
1 0.1  3.003909  3.271689  -0.2677795  -1.077543   0.35280003 0.0011111111
0.120
2 0.2  4.669126  5.054442  -0.3853166  -1.729177   0.26492888 0.0005555556
0.048
3 0.3  5.879876  6.947461  -1.0675856  -2.163422  -0.08866699 0.0002222222
0.000
4 0.4  7.031153  8.800770  -1.7696169  -3.197221  -0.10172468 0.0003703704
0.001
5 0.5  8.779621 11.164534  -2.3849128  -4.138833  -0.68578318 0.0001851852
0.000
6 0.6 10.163736 14.801884  -4.6381488  -7.434746  -2.31255837 0.0001587302
0.000
7 0.7 12.621211 21.897544  -9.2763334 -15.469907  -4.54349928 0.0001388889
0.000
8 0.8 16.081776 39.921159 -23.8393830 -35.994099 -14.81238935 0.0001234568
0.000
9 0.9 37.677304 88.364490 -50.6871867 -96.274411 -14.49481032 0.0002777778
0.000
```

```
$`Three - Zero`
    q     Three        Zero  difference    ci_lower      ci_upper       p_crit
p_value
1 0.1  3.003909   3.983926  -0.9800166   -1.903436  -0.006656401 1.388889e-04
0.001
2 0.2  4.669126   5.882595  -1.2134691   -2.879105  -0.086492425 2.777778e-04
0.000
3 0.3  5.879876   7.865977  -1.9861017   -3.734201  -1.024043198 9.259259e-05
0.000
4 0.4  7.031153   9.694327  -2.6631740   -4.599276  -0.939243353 6.944444e-05
0.000
5 0.5  8.779621  12.000363  -3.2207427   -5.054878  -1.702399974 5.555556e-05
0.000
6 0.6 10.163736  14.969159  -4.8054231   -7.460095  -2.653901022 4.629630e-05
0.000
7 0.7 12.621211  20.284293  -7.6630826  -17.720925  -3.311083266 3.968254e-05
0.000
8 0.8 16.081776  36.430181 -20.3484048  -39.968852  -9.891861679 3.472222e-05
0.000
9 0.9 37.677304 135.369082 -97.6917782 -211.382063 -27.240341197 3.086420e-05
0.000

$`Two - Zero`
    q       Two        Zero  difference    ci_lower   ci_upper       p_crit p_
value
1 0.1  3.271689   3.983926  -0.7122371   -1.353641  0.8022680 5.144033e-06
0.056
2 0.2  5.054442   5.882595  -0.8281526   -1.974228  0.5815799 4.409171e-06
0.073
3 0.3  6.947461   7.865977  -0.9185161   -2.242591  0.2310890 3.429355e-06
0.014
4 0.4  8.800770   9.694327  -0.8935570   -2.848404  0.6505409 6.172840e-06
0.074
5 0.5 11.164534  12.000363  -0.8358299   -3.215312  1.3465880 7.716049e-06
0.197
6 0.6 14.801884  14.969159  -0.1672743   -3.445051  3.0269873 3.086420e-05
0.875
7 0.7 21.897544  20.284293   1.6132508   -6.135220  9.0906425 1.028807e-05
0.471
8 0.8 39.921159  36.430181   3.4909783  -14.377171 18.1646904 1.543210e-05
0.472
9 0.9 88.364490 135.369082 -47.0045916 -166.342406 25.9006609 3.858025e-06
0.059
```

Each significant pair from Dunn’s test is also associated with significant (non-zero) shifts. **Therefore, we may conclude that statistically significant differences exist between the quantiles of certain topic group pairings (in their distributions of retweets), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Samsung official tweets.** However, it should be noted that these differences seem to be confined, generally, to the far right-tails of distributions.

Toyota Official Topics: Number of Likes



Topic 8 is the least represented above, containing 74 tweets.

```
        Kruskal-Wallis rank sum test

data:  Number of Likes by topic
Kruskal-Wallis chi-squared = 320.41, df = 13, p-value < 2.2e-16
```

Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the ‘like’ distributions of all populations are equal to one another. Performing Dunn’s test with a Bonferroni correction to account for multiple comparisons yields the following:

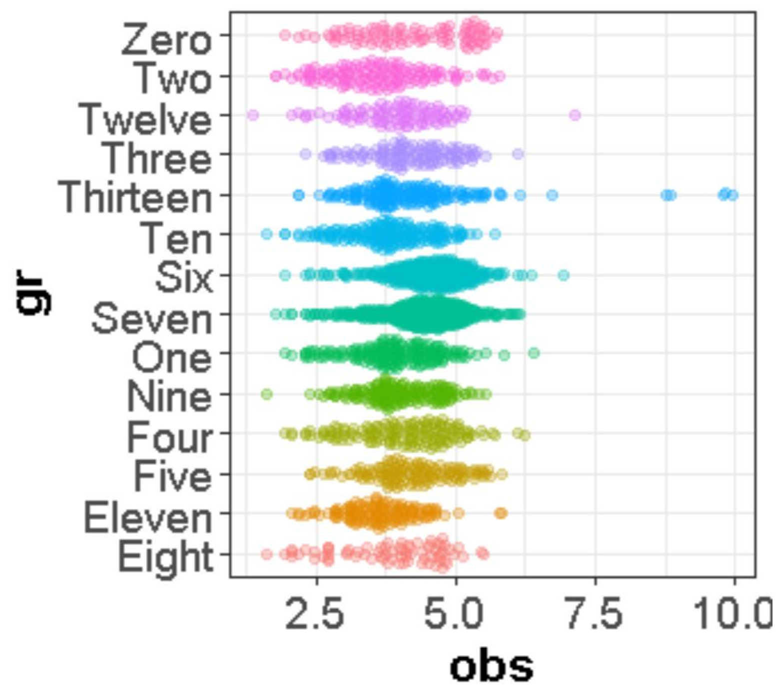
```
> dunnTest(`Number of Likes` ~ topic, data = tweets, method = "bonferroni")
```

```
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.
```

	Comparison	Z	P.unadj	P.adj
1	0 - 1	3.1586786	1.584862e-03	1.442224e-01
2	0 - 10	4.0046296	6.211472e-05	5.652440e-03
3	1 - 10	1.0014776	3.165960e-01	1.000000e+00
4	0 - 11	5.6395009	1.705438e-08	1.551948e-06
5	1 - 11	3.1704422	1.522071e-03	1.385085e-01
6	10 - 11	2.2937087	2.180724e-02	1.000000e+00
7	0 - 12	2.5502229	1.076541e-02	9.796519e-01
8	1 - 12	-0.2428028	8.081582e-01	1.000000e+00
9	10 - 12	-1.0694246	2.848784e-01	1.000000e+00
10	11 - 12	-2.9188473	3.513283e-03	3.197087e-01
11	0 - 13	1.4351979	1.512307e-01	1.000000e+00
12	1 - 13	-2.1204070	3.397173e-02	1.000000e+00
13	10 - 13	-3.1574134	1.591755e-03	1.448497e-01
14	11 - 13	-5.1231603	3.004567e-07	2.734156e-05
15	12 - 13	-1.5074442	1.316968e-01	1.000000e+00
16	0 - 2	5.4826723	4.189487e-08	3.812433e-06
17	1 - 2	2.9629422	3.047138e-03	2.772895e-01
18	10 - 2	2.0724756	3.822111e-02	1.000000e+00
19	11 - 2	-0.2303422	8.178259e-01	1.000000e+00
20	12 - 2	2.7357559	6.223717e-03	5.663583e-01
21	13 - 2	4.9405384	7.790713e-07	7.089549e-05
22	0 - 3	0.4989426	6.178198e-01	1.000000e+00
23	1 - 3	-2.8339096	4.598233e-03	4.184392e-01
24	10 - 3	-3.7509772	1.761467e-04	1.602935e-02
25	11 - 3	-5.5018731	3.757774e-08	3.419574e-06
26	12 - 3	-2.2041832	2.751146e-02	1.000000e+00
27	13 - 3	-0.9657135	3.341875e-01	1.000000e+00
28	2 - 3	-5.3360033	9.501768e-08	8.646609e-06
29	0 - 4	1.4523732	1.463979e-01	1.000000e+00
30	1 - 4	-1.8688546	6.164305e-02	1.000000e+00
31	10 - 4	-2.8262296	4.709949e-03	4.286054e-01
32	11 - 4	-4.7222052	2.333011e-06	2.123040e-04
33	12 - 4	-1.3459181	1.783289e-01	1.000000e+00
34	13 - 4	0.1037937	9.173331e-01	1.000000e+00
35	2 - 4	-4.5423262	5.563686e-06	5.062954e-04
36	3 - 4	1.0045642	3.151067e-01	1.000000e+00
37	0 - 5	0.3083269	7.578336e-01	1.000000e+00
38	1 - 5	-3.1564413	1.597070e-03	1.453334e-01
39	10 - 5	-4.1061319	4.023397e-05	3.661292e-03
40	11 - 5	-5.8641778	4.513635e-09	4.107408e-07
41	12 - 5	-2.4543239	1.411498e-02	1.000000e+00
42	13 - 5	-1.2341272	2.171555e-01	1.000000e+00
43	2 - 5	-5.6995129	1.201502e-08	1.093367e-06
44	3 - 5	-0.2156280	8.292778e-01	1.000000e+00
45	4 - 5	-1.2565899	2.089022e-01	1.000000e+00
46	0 - 6	-3.2407297	1.192242e-03	1.084940e-01
47	1 - 6	-8.6134151	7.091572e-18	6.453331e-16
48	10 - 6	-9.9398183	2.793363e-23	2.541960e-21
49	11 - 6	-11.2530586	2.236977e-29	2.035649e-27
50	12 - 6	-6.5159063	7.225208e-11	6.574939e-09
51	13 - 6	-6.2166112	5.080064e-10	4.622858e-08
52	2 - 6	-11.1576344	6.571695e-29	5.980242e-27
53	3 - 6	-4.2261914	2.376798e-05	2.162886e-03

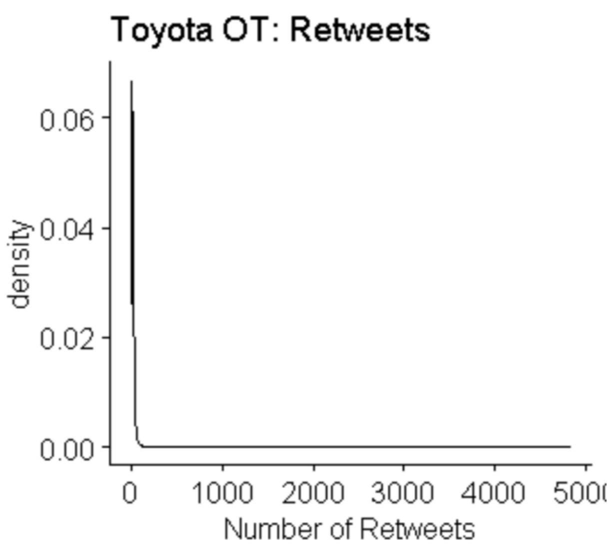
54	4 - 6	-5.7751302	7.689361e-09	6.997318e-07
55	5 - 6	-4.1144971	3.880245e-05	3.531023e-03
56	0 - 7	-1.8953058	5.805190e-02	1.000000e+00
57	1 - 7	-7.1843879	6.750894e-13	6.143314e-11
58	10 - 7	-8.5666622	1.065278e-17	9.694032e-16
59	11 - 7	-10.0588483	8.397491e-24	7.641717e-22
60	12 - 7	-5.2627614	1.419077e-07	1.291360e-05
61	13 - 7	-4.6467491	3.372069e-06	3.068582e-04
62	2 - 7	-9.9463084	2.617118e-23	2.381578e-21
63	3 - 7	-2.7843642	5.363277e-03	4.880582e-01
64	4 - 7	-4.3102006	1.631065e-05	1.484269e-03
65	5 - 7	-2.6167571	8.876948e-03	8.078023e-01
66	6 - 7	2.3885812	1.691357e-02	1.000000e+00
67	0 - 8	2.4978174	1.249605e-02	1.000000e+00
68	1 - 8	-0.1081300	9.138925e-01	1.000000e+00
69	10 - 8	-0.8673234	3.857648e-01	1.000000e+00
70	11 - 8	-2.6034999	9.227730e-03	8.397234e-01
71	12 - 8	0.1026253	9.182603e-01	1.000000e+00
72	13 - 8	1.5042345	1.325210e-01	1.000000e+00
73	2 - 8	-2.4292165	1.513149e-02	1.000000e+00
74	3 - 8	2.1629112	3.054801e-02	1.000000e+00
75	4 - 8	1.3597080	1.739224e-01	1.000000e+00
76	5 - 8	2.3917498	1.676827e-02	1.000000e+00
77	6 - 8	6.0393489	1.547373e-09	1.408110e-07
78	7 - 8	4.8769426	1.077428e-06	9.804594e-05
79	0 - 9	2.4322178	1.500668e-02	1.000000e+00
80	1 - 9	-0.8250559	4.093399e-01	1.000000e+00
81	10 - 9	-1.8155689	6.943652e-02	1.000000e+00
82	11 - 9	-3.8706508	1.085452e-04	9.877611e-03
83	12 - 9	-0.4477504	6.543333e-01	1.000000e+00
84	13 - 9	1.2454099	2.129811e-01	1.000000e+00
85	2 - 9	-3.6750537	2.377994e-04	2.163975e-02
86	3 - 9	2.0504106	4.032438e-02	1.000000e+00
87	4 - 9	1.0603806	2.889715e-01	1.000000e+00
88	5 - 9	2.3432926	1.911439e-02	1.000000e+00
89	6 - 9	7.4237063	1.138876e-13	1.036377e-11
90	7 - 9	5.9639216	2.462549e-09	2.240920e-07
91	8 - 9	-0.5275959	5.977798e-01	1.000000e+00

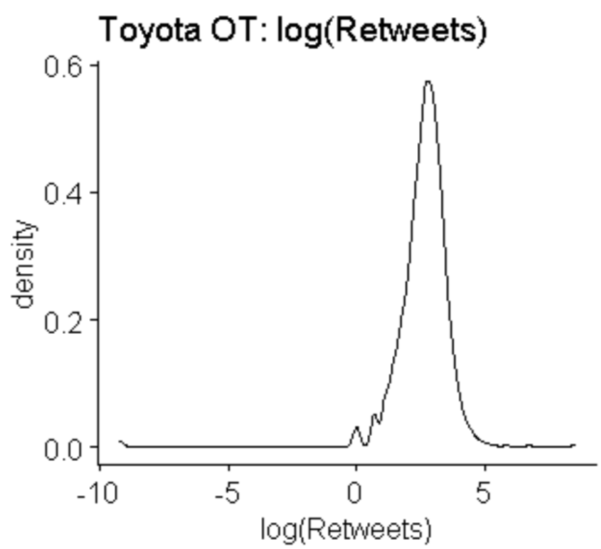
We may reject the null for a number of (i, j) topic group pairings. Performing a shift function to further examine these differences yields the following (confidence intervals stored in 'Topic_Appendix'):

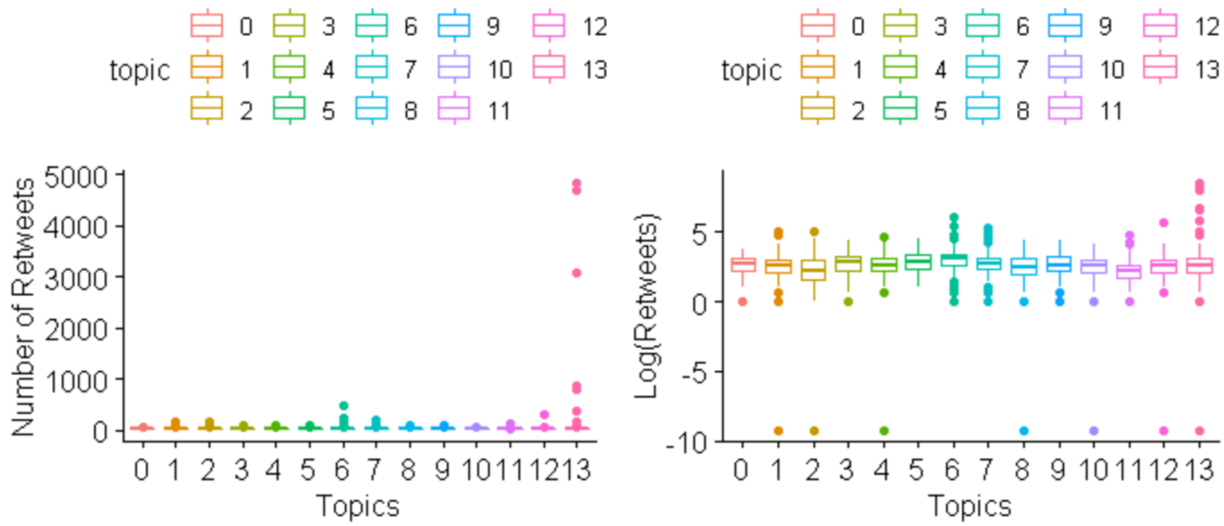


There are too many topic group combinations to comment on each individual shift, but plenty are significant (non-zero). **Therefore, we may conclude that statistically significant differences exist between the quantiles of certain topic group pairings (in their distributions of likes), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Toyota official tweets.**

Toyota Official Topics: Number of Retweets








```
        Kruskal-Wallis rank sum test

data:  Number of Retweets by topic
Kruskal-Wallis chi-squared = 221.16, df = 13, p-value < 2.2e-16
```

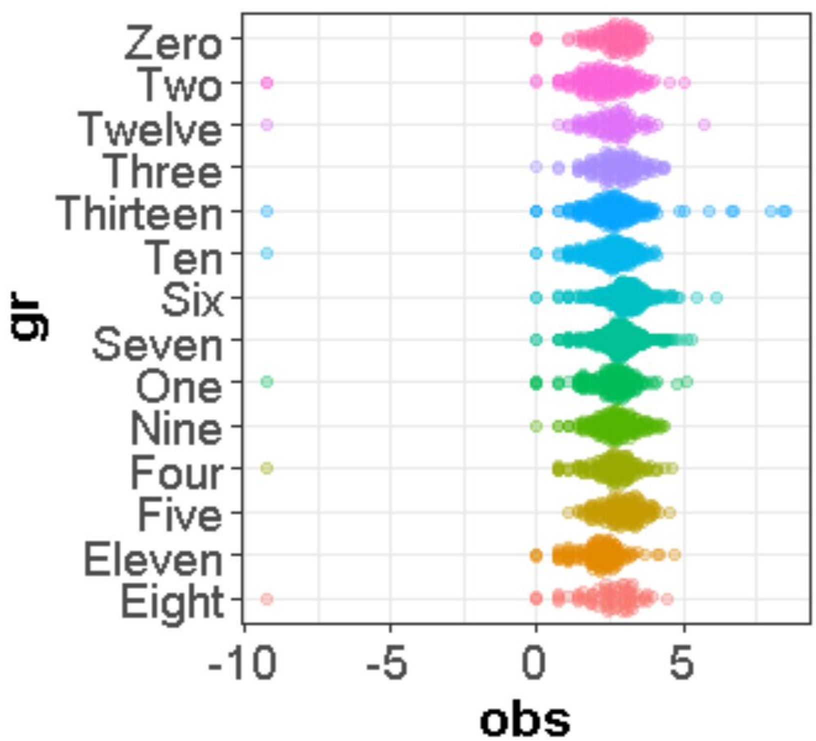
Performing a Kruskal-Wallis test results in a statistically significant p-value. Therefore, we may reject the null hypothesis that the 'retweet' distributions of all populations are equal to one another. Performing Dunn's test with a Bonferroni correction to account for multiple comparisons yields the following:

```
> dunnTest(`Number of Retweets` ~ topic, data = tweets, method = "bonferroni"
)
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Bonferroni method.

   Comparison           Z      P.unadj        P.adj
1       0 - 1  1.19562193 2.318441e-01 1.000000e+00
2      0 - 10  1.39039725 1.644083e-01 1.000000e+00
3      1 - 10  0.22498986 8.219872e-01 1.000000e+00
4      0 - 11  5.21423403 1.845783e-07 1.679663e-05
5      1 - 11  4.86367559 1.152257e-06 1.048554e-04
6     10 - 11  4.71756116 2.386887e-06 2.172067e-04
7      0 - 12  0.88831160 3.743732e-01 1.000000e+00
8      1 - 12 -0.18097267 8.563890e-01 1.000000e+00
9     10 - 12 -0.36793618 7.129208e-01 1.000000e+00
10    11 - 12 -4.29759756 1.726593e-05 1.571199e-03
11     0 - 13  0.58398163 5.592327e-01 1.000000e+00
12     1 - 13 -0.75275443 4.515975e-01 1.000000e+00
13    10 - 13 -0.98980593 3.222690e-01 1.000000e+00
14    11 - 13 -5.56976095 2.550891e-08 2.321311e-06
15    12 - 13 -0.44010803 6.598589e-01 1.000000e+00
16      0 - 2  3.31806552 9.064322e-04 8.248533e-02
17      1 - 2  2.60562764 9.170610e-03 8.345255e-01
18     10 - 2  2.42997132 1.510002e-02 1.000000e+00
```

19	11 - 2	-2.15092133	3.148241e-02	1.000000e+00
20	12 - 2	2.37494111	1.755175e-02	1.000000e+00
21	13 - 2	3.31316947	9.224507e-04	8.394301e-02
22	0 - 3	-0.74152935	4.583725e-01	1.000000e+00
23	1 - 3	-2.16893305	3.008777e-02	1.000000e+00
24	10 - 3	-2.39248908	1.673453e-02	1.000000e+00
25	11 - 3	-6.40867973	1.467852e-10	1.335745e-08
26	12 - 3	-1.69142947	9.075481e-02	1.000000e+00
27	13 - 3	-1.50945467	1.311826e-01	1.000000e+00
28	2 - 3	-4.38792170	1.144390e-05	1.041395e-03
29	0 - 4	0.78839238	4.304672e-01	1.000000e+00
30	1 - 4	-0.42189144	6.731043e-01	1.000000e+00
31	10 - 4	-0.63699926	5.241253e-01	1.000000e+00
32	11 - 4	-4.98030492	6.348417e-07	5.777059e-05
33	12 - 4	-0.18355582	8.543619e-01	1.000000e+00
34	13 - 4	0.27927141	7.800365e-01	1.000000e+00
35	2 - 4	-2.85045987	4.365606e-03	3.972701e-01
36	3 - 4	1.67496855	9.394040e-02	1.000000e+00
37	0 - 5	-2.01428460	4.397966e-02	1.000000e+00
38	1 - 5	-3.72658433	1.940922e-04	1.766239e-02
39	10 - 5	-3.97656532	6.991780e-05	6.362520e-03
40	11 - 5	-7.95970642	1.724480e-15	1.569276e-13
41	12 - 5	-2.99672593	2.728959e-03	2.483353e-01
42	13 - 5	-3.05303967	2.265360e-03	2.061478e-01
43	2 - 5	-5.90517336	3.522761e-09	3.205713e-07
44	3 - 5	-1.34808311	1.776317e-01	1.000000e+00
45	4 - 5	-3.12928653	1.752313e-03	1.594605e-01
46	0 - 6	-4.21287154	2.521445e-05	2.294515e-03
47	1 - 6	-7.10202794	1.229394e-12	1.118748e-10
48	10 - 6	-7.50092614	6.336849e-14	5.766533e-12
49	11 - 6	-11.80272778	3.778607e-32	3.438532e-30
50	12 - 6	-5.39363864	6.904500e-08	6.283095e-06
51	13 - 6	-6.27316573	3.537800e-10	3.219398e-08
52	2 - 6	-9.38287817	6.419593e-21	5.841830e-19
53	3 - 6	-3.67125513	2.413623e-04	2.196397e-02
54	4 - 6	-6.02380288	1.703658e-09	1.550329e-07
55	5 - 6	-2.13273626	3.294637e-02	1.000000e+00
56	0 - 7	-1.37949880	1.677410e-01	1.000000e+00
57	1 - 7	-3.59026071	3.303474e-04	3.006161e-02
58	10 - 7	-3.94395531	8.014860e-05	7.293523e-03
59	11 - 7	-8.86711551	7.506503e-19	6.830918e-17
60	12 - 7	-2.56241745	1.039463e-02	9.459113e-01
61	13 - 7	-2.69570016	7.024086e-03	6.391918e-01
62	2 - 7	-6.29433733	3.087158e-10	2.809314e-08
63	3 - 7	-0.52033766	6.028283e-01	1.000000e+00
64	4 - 7	-2.75794413	5.816614e-03	5.293119e-01
65	5 - 7	1.20511267	2.281598e-01	1.000000e+00
66	6 - 7	4.89032638	1.006689e-06	9.160872e-05
67	0 - 8	1.18836267	2.346906e-01	1.000000e+00
68	1 - 8	0.23462013	8.145036e-01	1.000000e+00
69	10 - 8	0.06624086	9.471861e-01	1.000000e+00
70	11 - 8	-3.61394079	3.015780e-04	2.744359e-02
71	12 - 8	0.35635528	7.215745e-01	1.000000e+00
72	13 - 8	0.80766526	4.192833e-01	1.000000e+00
73	2 - 8	-1.82092396	6.861842e-02	1.000000e+00
74	3 - 8	1.95022210	5.114965e-02	1.000000e+00
75	4 - 8	0.55563096	5.784632e-01	1.000000e+00

```
76       5 - 8   3.16655448 1.542564e-03 1.403734e-01
77       6 - 8   5.32922286 9.863391e-08 8.975685e-06
78       7 - 8   2.75733594 5.827445e-03 5.302975e-01
79       0 - 9  -0.17995090 8.571911e-01 1.000000e+00
80       1 - 9  -1.64811761 9.932854e-02 1.000000e+00
81      10 - 9  -1.89111126 5.860949e-02 1.000000e+00
82      11 - 9  -6.28950898 3.184717e-10 2.898092e-08
83      12 - 9  -1.19387531 2.325267e-01 1.000000e+00
84      13 - 9  -0.91840890 3.584048e-01 1.000000e+00
85       2 - 9  -4.09079391 4.298990e-05 3.912081e-03
86       3 - 9   0.66417205 5.065802e-01 1.000000e+00
87       4 - 9  -1.12972068 2.585939e-01 1.000000e+00
88       5 - 9   2.14940804 3.160207e-02 1.000000e+00
89       6 - 9   5.01490730 5.305904e-07 4.828373e-05
90       7 - 9   1.49411410 1.351458e-01 1.000000e+00
91       8 - 9  -1.49874109 1.339408e-01 1.000000e+00
```

We may reject the null for a number of (i, j) topic group pairings. Performing a shift function to further examine these differences yields the following (confidence intervals stored in 'Topic_Appendix'):



There are too many topic group combinations to comment on each individual shift, but plenty are significant (non-zero). **Therefore, we may conclude that statistically significant differences exist between the quantiles of certain topic group pairings (in their distributions of retweets), and potentially, one underlying factor explaining these differences is the common subject matter unveiled during BTM topic modeling of Toyota official tweets.**

<u>Section 11:</u> Modelling Results

Note: ‘NN’ refers to Neural Networks.

Section 11 displays each model-variable combination’s MAE, along with rank, for each company, tweet category, and target variable under consideration. All the individual company results which were conglomerated into average ranking results in tables 23-26 within the contents of the report are contained here.

Table 11.1: Amazon Model-Variable Combination Rankings for OT Data - Likes

Model-Variable Combination	MAE	Rank
(KNN: Hashtag)	400.7915	1
(KNN: Link, Hashtag)	401.1522	2
(SVR: Hashtag)	402.0544	3
(SVR: Link, Hashtag)	402.7596	4
(SVR: Link)	403.4097	5
(KNN: Topic)	404.5274	6
(KNN: Sentiment)	404.9084	7
(KNN: Link)	406.8308	8
(GBDT: Link, Topic)	407.8108	9
(KNN: Link, Topic)	407.9351	10
(GBDT: Sentiment, Topic)	409.5381	11
(NN: Topic)	409.652	12
(GBDT: Link, Hashtag)	410.1161	13
(GBDT: Hashtag, Topic)	410.5748	14
(GBDT: Topic)	411.3073	15
(SVR: Link, Topic)	411.5753	16
(SVR: Hashtag, Sentiment)	411.6046	17
(SVR: Link, Hashtag, Topic)	411.6219	18
(RF: Link, Hashtag, Topic)	411.8173	19
(SVR: Link, Hashtag, Sentiment)	411.9172	20
(LR: Hashtag, Sentiment, Topic)	412.0342	21
(SVR: Topic)	412.0742	22
(SVR: Hashtag, Topic)	412.1354	23
(SVR: Link, Sentiment, Topic)	412.1386	24
(RF: Topic)	412.1529	25
(SVR: Link, Hashtag, Sentiment, Topic)	412.2415	26
(SVR: Sentiment, Topic)	412.2663	27
(KNN: Hashtag, Sentiment)	412.2687	28
(KNN: Link, Sentiment, Topic)	412.3159	29
(SVR: Hashtag, Sentiment, Topic)	412.3234	30
(LR: Sentiment, Topic)	412.5853	31
(LR: Hashtag, Topic)	412.5864	32
(RF: Hashtag, Topic)	412.598	33
(NN: Hashtag, Topic)	412.6476	34
(GBDT: Link, Sentiment, Topic)	412.8972	35
(GBDT: Link, Hashtag, Topic)	412.9494	36
(LR: Link, Hashtag)	413.0986	37
(NN: Link, Hashtag)	413.1277	38
(RF: Link, Topic)	413.1496	39
(RF: Link, Sentiment, Topic)	413.1728	40
(LR: Topic)	413.1797	41
(GBDT: Hashtag)	413.2636	42
(RF: Hashtag, Sentiment, Topic)	413.4068	43
(KNN: Sentiment, Topic)	413.4437	44
(RF: Link, Hashtag, Sentiment, Topic)	413.5946	45
(RF: Sentiment, Topic)	413.6491	46
(GBDT: Hashtag, Sentiment, Topic)	413.7744	47
(GBDT: Link, Hashtag, Sentiment, Topic)	413.8484	48
(RF: Link, Hashtag)	414.7104	49

(RF: Hashtag)	414.9133	50
(LR: Hashtag)	415.0041	51
(NN: Hashtag)	415.0077	52
(LR: Link, Sentiment, Topic)	415.2441	53
(LR: Link, Hashtag, Sentiment, Topic)	415.4946	54
(LR: Link, Topic)	415.778	55
(LR: Link, Hashtag, Topic)	415.8092	56
(RF: Link)	415.9436	57
(LR: Link)	416.1059	58
(NN: Link)	416.2053	59
(GBDT: Link)	417.6802	60
(GBDT: Link, Sentiment)	418.038	61
(GBDT: Hashtag, Sentiment)	418.282	62
(NN: Link, Topic)	418.409	63
(NN: Link, Hashtag, Sentiment)	418.4324	64
(KNN: Link, Hashtag, Topic)	418.6173	65
(SVR: Link, Sentiment)	418.7264	66
(RF: Link, Hashtag, Sentiment)	418.8554	67
(SVR: Sentiment)	418.8711	68
(RF: Hashtag, Sentiment)	420.4272	69
(RF: Link, Sentiment)	420.9021	70
(KNN: Link, Hashtag, Sentiment)	421.5629	71
(NN: Link, Hashtag, Topic)	421.8525	72
(LR: Link, Hashtag, Sentiment)	421.8898	73
(NN: Hashtag, Sentiment)	422.3122	74
(NN: Sentiment, Topic)	422.6888	75
(LR: Hashtag, Sentiment)	423.2862	76
(NN: Hashtag, Sentiment, Topic)	423.4531	77
(RF: Sentiment)	423.7395	78
(GBDT: Link, Hashtag, Sentiment)	424.0369	79
(KNN: Link, Hashtag, Sentiment, Topic)	424.1144	80
(NN: Link, Sentiment)	424.1154	81
(KNN: Hashtag, Sentiment, Topic)	424.9259	82
(LR: Link, Sentiment)	425.2014	83
(KNN: Hashtag, Topic)	425.4426	84
(KNN: Link, Sentiment)	425.446	85
(LR: Sentiment)	426.4954	86
(NN: Sentiment)	426.5012	87
(GBDT: Sentiment)	426.5951	88
(NN: Link, Sentiment, Topic)	427.9828	89
(NN: Link, Hashtag, Sentiment, Topic)	438.612	90

Table 11.2: Amazon Model-Variable Combination Rankings for OT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(SVR: Hashtag)	49.44681	1
(SVR: Link, Hashtag)	49.51983	2
(SVR: Link)	49.52842	3
(KNN: Sentiment)	50.36059	4
(KNN: Link, Hashtag)	50.60817	5
(KNN: Link, Sentiment)	50.6506	6
(KNN: Link)	50.67698	7
(KNN: Hashtag)	50.72586	8
(KNN: Sentiment, Topic)	51.14225	9
(KNN: Link, Sentiment, Topic)	51.25545	10
(KNN: Topic)	51.35669	11
(KNN: Hashtag, Sentiment)	51.47237	12
(SVR: Link, Hashtag, Sentiment, Topic)	51.52523	13
(SVR: Hashtag, Sentiment, Topic)	51.54824	14

(SVR: Link, Sentiment, Topic)	51.57842	15
(SVR: Sentiment, Topic)	51.59918	16
(SVR: Link, Hashtag, Topic)	51.60138	17
(SVR: Link, Topic)	51.66466	18
(SVR: Topic)	51.6675	19
(SVR: Hashtag, Topic)	51.69583	20
(KNN: Link, Topic)	51.72041	21
(GBDT: Hashtag, Sentiment, Topic)	51.83543	22
(SVR: Hashtag, Sentiment)	51.99453	23
(LR: Hashtag, Sentiment, Topic)	52.01428	24
(GBDT: Link, Hashtag, Sentiment, Topic)	52.01442	25
(SVR: Link, Hashtag, Sentiment)	52.03697	26
(RF: Sentiment, Topic)	52.08735	27
(KNN: Link, Hashtag, Sentiment)	52.10111	28
(SVR: Sentiment)	52.11117	29
(RF: Link, Sentiment, Topic)	52.1118	30
(RF: Link, Hashtag, Topic)	52.1355	31
(GBDT: Link, Sentiment, Topic)	52.15799	32
(RF: Topic)	52.18255	33
(SVR: Link, Sentiment)	52.18497	34
(RF: Hashtag, Topic)	52.18707	35
(RF: Hashtag, Sentiment, Topic)	52.19311	36
(LR: Link, Hashtag, Sentiment, Topic)	52.20484	37
(KNN: Hashtag, Topic)	52.21599	38
(RF: Link, Hashtag, Sentiment, Topic)	52.22431	39
(RF: Link, Topic)	52.38215	40
(LR: Sentiment, Topic)	52.48003	41
(NN: Hashtag, Topic)	52.48377	42
(GBDT: Link, Hashtag)	52.49788	43
(GBDT: Hashtag, Topic)	52.51639	44
(NN: Link, Hashtag)	52.55981	45
(LR: Link, Sentiment, Topic)	52.63456	46
(GBDT: Hashtag)	52.65652	47
(LR: Link, Hashtag)	52.70887	48
(GBDT: Sentiment, Topic)	52.72071	49
(RF: Hashtag)	52.72095	50
(NN: Hashtag)	52.72295	51
(LR: Hashtag)	52.72349	52
(RF: Link, Hashtag)	52.7245	53
(GBDT: Link, Hashtag, Topic)	52.80764	54
(LR: Hashtag, Topic)	52.82526	55
(NN: Hashtag, Sentiment)	52.87465	56
(GBDT: Link)	52.90275	57
(RF: Link)	52.92281	58
(LR: Link)	52.92872	59
(NN: Link)	52.93247	60
(NN: Sentiment)	52.93751	61
(NN: Link, Sentiment)	52.97065	62
(KNN: Link, Hashtag, Topic)	52.97446	63
(LR: Link, Hashtag, Topic)	53.03241	64
(RF: Link, Hashtag, Sentiment)	53.06708	65
(NN: Topic)	53.08054	66
(GBDT: Topic)	53.12045	67
(RF: Link, Sentiment)	53.12734	68
(LR: Link, Hashtag, Sentiment)	53.13816	69
(LR: Hashtag, Sentiment)	53.15803	70
(RF: Hashtag, Sentiment)	53.17664	71
(KNN: Link, Hashtag, Sentiment, Topic)	53.22456	72
(LR: Topic)	53.24204	73
(GBDT: Link, Sentiment)	53.27176	74
(RF: Sentiment)	53.3639	75
(GBDT: Hashtag, Sentiment)	53.37894	76
(LR: Link, Topic)	53.39222	77

(LR: Link, Sentiment)	53.45711	78
(LR: Sentiment)	53.46417	79
(GBDT: Link, Hashtag, Sentiment)	53.52034	80
(NN: Link, Topic)	53.60191	81
(NN: Link, Hashtag, Sentiment)	53.60254	82
(KNN: Hashtag, Sentiment, Topic)	53.64126	83
(GBDT: Link, Topic)	53.82146	84
(GBDT: Sentiment)	53.91258	85
(NN: Link, Hashtag, Topic)	54.84128	86
(NN: Hashtag, Sentiment, Topic)	55.29993	87
(NN: Sentiment, Topic)	58.23734	88
(NN: Link, Sentiment, Topic)	58.81394	89
(NN: Link, Hashtag, Sentiment, Topic)	60.51407	90

Table 11.3: Amazon Model-Variable Combination Rankings for IRT Data - Likes

Model-Variable Combination	MAE	Rank
(SVR: Link, Hashtag, Sentiment)	1.25868	1
(SVR: Link, Hashtag, Topic)	1.266127	2
SVR: Hashtag, Topic, NW)	1.295894	3
(SVR: Link, Hashtag, NW)	1.29696	4
(SVR: Link, Topic)	1.301775	5
(SVR: Link, Sentiment)	1.301775	6
(SVR: Topic)	1.301775	7
(SVR: Hashtag, Sentiment)	1.301775	8
(SVR: Sentiment)	1.301775	9
(SVR: Hashtag, NW)	1.304015	10
(SVR: Sentiment, Topic)	1.304813	11
(SVR: Hashtag, Sentiment, NW)	1.30488	12
(SVR: NW)	1.305727	13
(SVR: Link, Hashtag, Topic, NW)	1.306992	14
(SVR: Link, Hashtag, Sentiment, NW)	1.310228	15
(SVR: Link, NW)	1.31342	16
(SVR: Sentiment, NW)	1.318787	17
(SVR: Topic, NW)	1.319235	18
(KNN: Sentiment)	1.327416	19
(SVR: Hashtag, Topic)	1.332692	20
(SVR: Hashtag)	1.340828	21
(SVR: Link, Sentiment, NW)	1.342258	22
(SVR: Link, Topic, NW)	1.346833	23
(SVR: Sentiment, Topic, NW)	1.347411	24
(SVR: Link, Sentiment, Topic, NW)	1.389521	25
(SVR: Link, Hashtag)	1.400487	26
(SVR: Link, Sentiment, Topic)	1.417604	27
(SVR: Link, Hashtag, Sentiment, Topic)	1.423301	28
(SVR: Hashtag, Sentiment, Topic, NW)	1.461886	29
(SVR: Link, Hashtag, Sentiment, Topic, NW)	1.46821	30
(SVR: Link)	1.475345	31
(SVR: Hashtag, Sentiment, Topic)	1.479164	32
(KNN: Topic)	1.517443	33
(KNN: Hashtag)	1.541251	34
(KNN: Link)	1.568166	35
(KNN: Link, Topic)	1.630585	36
(KNN: Link, Sentiment)	1.670877	37
(KNN: Link, NW)	1.746588	38
(KNN: NW)	1.754575	39
(KNN: Hashtag, Topic)	1.765303	40
(KNN: Link, Hashtag, Sentiment)	1.845158	41
(KNN: Link, Hashtag)	1.849064	42

(KNN: Hashtag, NW)	1.881225	43
(NN: Link, Hashtag, NW)	1.902035	44
(KNN: Sentiment, Topic)	1.96509	45
(KNN: Link, Sentiment, Topic)	1.986404	46
(KNN: Hashtag, Sentiment)	1.990095	47
(KNN: Hashtag, Sentiment, Topic)	2.036482	48
(KNN: Topic, NW)	2.098619	49
(KNN: Sentiment, NW)	2.183578	50
(KNN: Hashtag, Sentiment, NW)	2.193217	51
(KNN: Link, Topic, NW)	2.209956	52
(KNN: Link, Hashtag, Topic)	2.264553	53
(RF: Topic, NW)	2.331626	54
(GBDT: Sentiment, Topic)	2.334019	55
(KNN: Link, Sentiment, NW)	2.353301	56
(RF: Link, Sentiment, Topic)	2.400036	57
(RF: Sentiment, Topic, NW)	2.405951	58
(RF: Sentiment, Topic)	2.407948	59
(RF: Link, NW)	2.412873	60
(RF: Link, Topic, NW)	2.418577	61
(KNN: Link, Hashtag, Topic, NW)	2.421339	62
(RF: Link, Sentiment)	2.428255	63
(RF: Link, Sentiment, NW)	2.4303	64
(RF: Link, Topic)	2.435252	65
(RF: Link, Hashtag, Topic)	2.444208	66
(KNN: Link, Hashtag, Sentiment, Topic, NW)	2.453149	67
(RF: Hashtag, NW)	2.455658	68
(NN: Link, Sentiment, NW)	2.456235	69
(NN: Link, Hashtag, Sentiment, Topic, NW)	2.457096	70
(NN: Link, Hashtag, Topic, NW)	2.45801	71
(NN: Hashtag, Sentiment, NW)	2.458512	72
(NN: NW)	2.458842	73
(NN: Link, Hashtag, Topic)	2.458912	74
(NN: Topic, NW)	2.459081	75
(NN: Link, Hashtag, Sentiment)	2.459418	76
(NN: Link, Hashtag, Sentiment, Topic)	2.459576	77
(NN: Sentiment, Topic)	2.462896	78
(NN: Topic)	2.463169	79
(LR: NW)	2.464206	80
(NN: Sentiment, Topic, NW)	2.464471	81
(NN: Link, Sentiment, Topic, NW)	2.464612	82
(NN: Link, Topic)	2.466291	83
(RF: Link, Hashtag, Sentiment)	2.469197	84
(NN: Link, Topic, NW)	2.469723	85
(NN: Hashtag, Topic, NW)	2.470036	86
(RF: Sentiment, NW)	2.471123	87
(RF: Sentiment)	2.471291	88
(GBDT: Sentiment, NW)	2.474519	89
(NN: Link, NW)	2.475424	90
(NN: Link, Sentiment)	2.476103	91
(KNN: Link, Sentiment, Topic, NW)	2.476716	92
(NN: Hashtag)	2.478266	93
(NN: Sentiment)	2.478377	94
(NN: Hashtag, Sentiment)	2.478984	95
(RF: Link)	2.479363	96
(RF: Topic)	2.479931	97
(NN: Hashtag, NW)	2.479964	98
(RF: NW)	2.480441	99
(NN: Link, Hashtag, Sentiment, NW)	2.482329	100
(NN: Link, Hashtag)	2.493247	101
(RF: Link, Hashtag, Sentiment, Topic)	2.493972	102
(NN: Link)	2.496161	103
(NN: Sentiment, NW)	2.498123	104
(RF: Link, Sentiment, Topic, NW)	2.498966	105

(LR: Sentiment)	2.49938	106
(LR: Sentiment, NW)	2.499798	107
(RF: Link, Hashtag)	2.501312	108
(LR: Link)	2.506487	109
(RF: Link, Hashtag, NW)	2.50767	110
(RF: Link, Hashtag, Sentiment, Topic, NW)	2.517143	111
(RF: Hashtag, Sentiment, NW)	2.532916	112
(NN: Link, Sentiment, Topic)	2.533578	113
(RF: Hashtag, Sentiment)	2.534345	114
(LR: Link, Topic)	2.535018	115
(KNN: Sentiment, Topic, NW)	2.536539	116
(RF: Link, Hashtag, Sentiment, NW)	2.545682	117
(LR: Topic)	2.548942	118
(LR: Link, Sentiment)	2.549125	119
(RF: Hashtag, Sentiment, Topic)	2.562503	120
(RF: Hashtag, Sentiment, Topic, NW)	2.562887	121
(RF: Hashtag)	2.569616	122
(GBDT: Link, Sentiment, Topic)	2.574887	123
(NN: Hashtag, Sentiment, Topic)	2.586072	124
(LR: Link, Sentiment, Topic)	2.588911	125
(LR: Sentiment, Topic)	2.594695	126
(LR: Hashtag)	2.602406	127
(GBDT: Sentiment)	2.607665	128
(LR: Hashtag, NW)	2.614059	129
(GBDT: Topic, NW)	2.616616	130
(RF: Link, Hashtag, Topic, NW)	2.618278	131
(LR: Link, NW)	2.619936	132
(KNN: Hashtag, Sentiment, Topic, NW)	2.630503	133
(KNN: Hashtag, Topic, NW)	2.633176	134
(LR: Topic, NW)	2.636072	135
(KNN: Link, Hashtag, Sentiment, Topic)	2.642372	136
(KNN: Link, Hashtag, NW)	2.648511	137
(LR: Hashtag, Sentiment)	2.651675	138
(LR: Link, Sentiment, NW)	2.664233	139
(RF: Hashtag, Topic)	2.668079	140
(GBDT: Link, Sentiment)	2.670181	141
(LR: Sentiment, Topic, NW)	2.675459	142
(RF: Hashtag, Topic, NW)	2.676831	143
(GBDT: Hashtag, Sentiment)	2.68027	144
(LR: Hashtag, Sentiment, NW)	2.685564	145
(GBDT: Topic)	2.695382	146
(GBDT: NW)	2.724151	147
(GBDT: Link, Sentiment, Topic, NW)	2.724347	148
(LR: Link, Topic, NW)	2.744929	149
(GBDT: Link, Topic)	2.746671	150
(LR: Link, Sentiment, Topic, NW)	2.777512	151
(GBDT: Link)	2.909264	152
(GBDT: Sentiment, Topic, NW)	2.916828	153
(GBDT: Link, Hashtag, NW)	2.937699	154
(GBDT: Link, Hashtag, Topic)	2.940285	155
(GBDT: Hashtag, Sentiment, Topic, NW)	2.957301	156
(GBDT: Hashtag, Sentiment, Topic)	2.986992	157
(NN: Hashtag, Topic)	3.021076	158
(GBDT: Hashtag)	3.07213	159
(GBDT: Hashtag, NW)	3.14821	160
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	3.153079	161
(GBDT: Link, Hashtag, Topic, NW)	3.191551	162
(NN: Hashtag, Sentiment, Topic, NW)	3.227149	163
(GBDT: Hashtag, Sentiment, NW)	3.23081	164
(GBDT: Link, NW)	3.262171	165
(GBDT: Link, Hashtag, Sentiment)	3.312818	166
(LR: Hashtag, Topic, NW)	3.38936	167

(GBDT: Link, Hashtag, Sentiment, Topic)	3.390595	168
(LR: Hashtag, Topic)	3.403258	169
(LR: Hashtag, Sentiment, Topic, NW)	3.446274	170
(LR: Hashtag, Sentiment, Topic)	3.45861	171
(LR: Link, Hashtag)	3.476944	172
(LR: Link, Hashtag, NW)	3.503872	173
(GBDT: Link, Topic, NW)	3.522576	174
(LR: Link, Hashtag, Sentiment)	3.54191	175
(LR: Link, Hashtag, Sentiment, NW)	3.550879	176
(GBDT: Link, Sentiment, NW)	3.551891	177
(GBDT: Link, Hashtag, Sentiment, NW)	3.584539	178
(KNN: Link, Hashtag, Sentiment, NW)	3.616863	179
(LR: Link, Hashtag, Topic)	3.625049	180
(GBDT: Link, Hashtag)	3.699494	181
(LR: Link, Hashtag, Sentiment, Topic)	3.713741	182
(LR: Link, Hashtag, Topic, NW)	3.763709	183
(LR: Link, Hashtag, Sentiment, Topic, NW)	3.825214	184
(GBDT: Hashtag, Topic)	3.837282	185
(GBDT: Hashtag, Topic, NW)	3.838893	186

Table 11.4: Amazon Model-Variable Combination Rankings for IRT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(KNN: Link, Hashtag, Topic, NW)	0.509204	1
(KNN: Link, Hashtag, Sentiment, Topic, NW)	0.511098	2
(GBDT: Hashtag, Sentiment, Topic)	0.51589	3
(GBDT: Link, Sentiment, Topic)	0.519675	4
(LR: Link, Topic, NW)	0.521024	5
(LR: Link, Sentiment, Topic, NW)	0.521359	6
(NN: Link, Sentiment, Topic, NW)	0.521539	7
(NN: Hashtag, Sentiment, Topic)	0.521775	8
(NN: Link, Topic)	0.523385	9
(NN: Sentiment, Topic, NW)	0.523451	10
(NN: Link, Topic, NW)	0.524107	11
(NN: Link, Sentiment, Topic)	0.524511	12
(GBDT: Link, Sentiment, Topic, NW)	0.524602	13
(GBDT: Hashtag, Sentiment, Topic, NW)	0.526772	14
(LR: Link, Topic)	0.527224	15
(NN: Topic, NW)	0.527811	16
(LR: Link, Sentiment, Topic)	0.52808	17
(KNN: Hashtag, Topic, NW)	0.529125	18
(KNN: Hashtag, Sentiment, Topic, NW)	0.529496	19
(GBDT: Topic, NW)	0.531066	20
(NN: Sentiment, Topic)	0.532071	21
(LR: Topic, NW)	0.53317	22
(NN: Topic)	0.533264	23
(LR: Topic)	0.533469	24
(RF: Link, Hashtag, NW)	0.534087	25
(LR: Sentiment, Topic, NW)	0.53432	26
(LR: Sentiment, Topic)	0.534617	27
(KNN: Link, Sentiment, NW)	0.53487	28
(RF: Link, Hashtag, Topic)	0.535475	29
(GBDT: Sentiment, Topic, NW)	0.535766	30
(RF: Link, Topic)	0.536253	31
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	0.536345	32
(NN: Hashtag, Topic, NW)	0.536554	33
(KNN: Hashtag, Topic)	0.536572	34

(KNN: Hashtag, Sentiment, Topic)	0.537749	35
(KNN: Topic)	0.538167	36
(RF: Link, Hashtag, Sentiment, Topic, NW)	0.538268	37
(RF: Link, Topic, NW)	0.53836	38
(RF: Link, Hashtag)	0.538422	39
(GBDT: Topic)	0.538888	40
(KNN: Sentiment, Topic)	0.539002	41
(GBDT: Link, Topic, NW)	0.53909	42
(NN: Link, Sentiment)	0.539961	43
(KNN: Link, NW)	0.540355	44
(SVR: Link, Hashtag)	0.540493	45
(NN: Link, Sentiment, NW)	0.540619	46
(GBDT: Sentiment, Topic)	0.540892	47
(GBDT: Hashtag, Topic)	0.541118	48
(RF: Link, Sentiment, Topic)	0.54135	49
(NN: Link, Hashtag, Topic)	0.541381	50
(GBDT: Link, Topic)	0.541471	51
(GBDT: Link, Hashtag, Topic, NW)	0.542332	52
(RF: Link, Sentiment, Topic, NW)	0.542812	53
(RF: Link, Hashtag, Sentiment, Topic)	0.542987	54
(LR: Hashtag, Topic)	0.543178	55
(LR: Hashtag, Topic, NW)	0.543279	56
(GBDT: Link)	0.543474	57
(KNN: Link, Topic)	0.543524	58
(LR: Link, Sentiment, NW)	0.543528	59
(NN: Hashtag, Sentiment, NW)	0.543587	60
(RF: Topic, NW)	0.543679	61
(LR: Link, NW)	0.543695	62
(GBDT: Link, Sentiment)	0.543793	63
(LR: Hashtag, Sentiment, Topic, NW)	0.54387	64
(LR: Hashtag, Sentiment, Topic)	0.543936	65
(RF: Topic)	0.544308	66
(RF: Link, Hashtag, Sentiment, NW)	0.544464	67
(LR: Link, Hashtag, Sentiment, Topic, NW)	0.544499	68
(RF: Link, Hashtag, Topic, NW)	0.545138	69
(LR: Link, Hashtag, Topic, NW)	0.545277	70
(GBDT: Link, NW)	0.54574	71
(KNN: Sentiment, NW)	0.546286	72
(KNN: Hashtag, Sentiment, NW)	0.546632	73
(RF: Link, Hashtag, Sentiment)	0.547147	74
(RF: Hashtag, Topic, NW)	0.54789	75
(RF: Link)	0.548071	76
(KNN: Link, Sentiment, Topic)	0.548095	77
(NN: Link, NW)	0.548341	78
(NN: Link)	0.548385	79
(LR: Link)	0.548424	80
(LR: Link, Sentiment)	0.548549	81
(GBDT: Link, Hashtag, Topic)	0.549115	82
(KNN: Hashtag, NW)	0.549744	83
(GBDT: Link, Hashtag, NW)	0.550032	84
(RF: Link, NW)	0.550422	85
(RF: Hashtag, NW)	0.550498	86
(RF: Hashtag, Sentiment, Topic, NW)	0.550864	87
(RF: Sentiment, Topic, NW)	0.55124	88
(RF: Link, Sentiment, NW)	0.55124	89
(LR: Link, Hashtag, Sentiment, Topic)	0.551245	90
(LR: Link, Hashtag, Topic)	0.551472	91
(RF: Sentiment, Topic)	0.551519	92
(KNN: Link)	0.552588	93
(NN: Link, Hashtag, Sentiment)	0.552862	94
(KNN: NW)	0.553426	95
(RF: Hashtag, Topic)	0.553801	96
(RF: Hashtag, Sentiment, Topic)	0.554027	97

(KNN: Link, Sentiment)	0.554331	98
(KNN: Hashtag, Sentiment)	0.555225	99
(RF: Link, Sentiment)	0.556199	100
(KNN: Sentiment)	0.55853	101
(LR: Link, Hashtag, Sentiment, NW)	0.559311	102
(NN: Hashtag, Topic)	0.559851	103
(NN: Link, Hashtag, Topic, NW)	0.560426	104
(LR: Link, Hashtag, NW)	0.560687	105
(NN: Link, Hashtag, Sentiment, Topic)	0.561188	106
(GBDT: Link, Hashtag, Sentiment, NW)	0.562508	107
(GBDT: Link, Hashtag, Sentiment)	0.562567	108
(RF: NW)	0.562712	109
(KNN: Hashtag)	0.562773	110
(GBDT: Link, Sentiment, NW)	0.562909	111
(RF: Hashtag, Sentiment, NW)	0.563055	112
(LR: Link, Hashtag, Sentiment)	0.563566	113
(LR: Link, Hashtag)	0.564319	114
(GBDT: Sentiment, NW)	0.565222	115
(NN: Sentiment, NW)	0.565652	116
(NN: Hashtag, NW)	0.565908	117
(GBDT: Link, Hashtag)	0.566611	118
(KNN: Link, Sentiment, Topic, NW)	0.566889	119
(RF: Hashtag)	0.566914	120
(NN: Hashtag)	0.567172	121
(KNN: Topic, NW)	0.567773	122
(NN: Link, Hashtag, NW)	0.567919	123
(RF: Sentiment, NW)	0.568075	124
(LR: Hashtag, Sentiment)	0.568247	125
(LR: Hashtag)	0.568503	126
(NN: Hashtag, Sentiment)	0.568827	127
(LR: Hashtag, Sentiment, NW)	0.569157	128
(LR: Hashtag, NW)	0.569692	129
(NN: Link, Hashtag, Sentiment, Topic, NW)	0.569943	130
(GBDT: Hashtag, Sentiment)	0.570437	131
(RF: Hashtag, Sentiment)	0.570464	132
(NN: NW)	0.570602	133
(GBDT: Hashtag, Topic, NW)	0.571204	134
(KNN: Link, Topic, NW)	0.571239	135
(KNN: Link, Hashtag, Sentiment, Topic)	0.571432	136
(GBDT: Link, Hashtag, Sentiment, Topic)	0.572751	137
(GBDT: NW)	0.57281	138
(LR: NW)	0.574121	139
SVR: Hashtag, Topic, NW)	0.574145	140
(SVR: Hashtag, Sentiment, NW)	0.574183	141
(SVR: Link, Topic, NW)	0.574192	142
(SVR: Link, Sentiment, Topic)	0.574224	143
(SVR: Hashtag, NW)	0.574245	144
(SVR: Hashtag, Sentiment, Topic, NW)	0.574259	145
(SVR: Link, Sentiment, Topic, NW)	0.574271	146
(SVR: Sentiment, NW)	0.574272	147
(SVR: Topic, NW)	0.574293	148
(SVR: Link, Sentiment)	0.574303	149
(SVR: Hashtag, Topic)	0.574303	150
(SVR: Link, Topic)	0.574303	151
(SVR: Sentiment, Topic)	0.574303	152
(SVR: Hashtag, Sentiment)	0.574303	153
(SVR: Sentiment)	0.574303	154
(SVR: Topic)	0.574303	155
(SVR: NW)	0.574309	156
(SVR: Link, NW)	0.574311	157
(SVR: Sentiment, Topic, NW)	0.574329	158
(SVR: Link, Hashtag, NW)	0.574353	159
(SVR: Link, Hashtag, Topic)	0.574405	160

(SVR: Hashtag, Sentiment, Topic)	0.574435	161
(SVR: Link, Hashtag, Sentiment, NW)	0.574488	162
(SVR: Link, Hashtag, Topic, NW)	0.574491	163
(SVR: Link, Hashtag, Sentiment)	0.5745	164
(SVR: Link, Sentiment, NW)	0.574765	165
(LR: Sentiment, NW)	0.575056	166
(NN: Link, Hashtag)	0.575257	167
(NN: Sentiment)	0.577651	168
(SVR: Link, Hashtag, Sentiment, Topic, NW)	0.577796	169
(NN: Hashtag, Sentiment, Topic, NW)	0.577875	170
(RF: Sentiment)	0.577892	171
(KNN: Sentiment, Topic, NW)	0.577993	172
(SVR: Link, Hashtag, Sentiment, Topic)	0.578119	173
(LR: Sentiment)	0.578186	174
(GBDT: Sentiment)	0.582172	175
(GBDT: Hashtag, Sentiment, NW)	0.58572	176
(KNN: Link, Hashtag, Topic)	0.587761	177
(GBDT: Hashtag, NW)	0.590838	178
(KNN: Link, Hashtag, Sentiment)	0.597821	179
(KNN: Link, Hashtag)	0.606335	180
(GBDT: Hashtag)	0.608097	181
(SVR: Link)	0.609862	182
(SVR: Hashtag)	0.609862	183
(NN: Link, Hashtag, Sentiment, NW)	0.647531	184
(KNN: Link, Hashtag, Sentiment, NW)	0.655402	185
(KNN: Link, Hashtag, NW)	0.691346	186

Table 11.5: BMW Model-Variable Combination Rankings for OT Data - Likes

Model-Variable Combination	MAE	Rank
(SVR: Link, Sentiment, Topic)	534.3088	1
(SVR: Link, Hashtag, Sentiment, Topic)	546.4105	2
(GBDT: Link, Topic)	547.3904	3
(LR: Link, Topic)	549.6319	4
(LR: Link, Sentiment, Topic)	549.9754	5
(NN: Sentiment, Topic)	550.0994	6
(GBDT: Link, Sentiment, Topic)	550.1439	7
(LR: Link, Hashtag, Topic)	552.8593	8
(LR: Link, Hashtag, Sentiment, Topic)	553.1881	9
(KNN: Hashtag, Sentiment, Topic)	553.8135	10
(LR: Sentiment, Topic)	554.0423	11
(RF: Link, Sentiment, Topic)	554.1752	12
(GBDT: Sentiment, Topic)	554.2065	13
(RF: Link, Topic)	554.6072	14
(GBDT: Topic)	554.6161	15
(NN: Topic)	554.6376	16
(LR: Topic)	554.8478	17
(RF: Topic)	554.8858	18
(RF: Sentiment, Topic)	555.9046	19
(LR: Hashtag, Sentiment, Topic)	556.4149	20
(RF: Link, Sentiment)	556.6193	21
(NN: Link, Sentiment)	556.7957	22
(LR: Hashtag, Topic)	557.2092	23
(NN: Hashtag, Sentiment, Topic)	557.5787	24
(GBDT: Hashtag, Sentiment, Topic)	557.7002	25
(KNN: Sentiment, Topic)	558.5285	26
(RF: Hashtag, Sentiment, Topic)	558.6159	27

(RF: Link, Hashtag, Sentiment, Topic)	559.4767	28
(SVR: Topic)	559.479	29
(NN: Link, Topic)	560.8243	30
(LR: Link, Sentiment)	560.846	31
(GBDT: Link, Sentiment)	560.8991	32
(RF: Link)	560.9292	33
(SVR: Sentiment, Topic)	560.9704	34
(SVR: Hashtag, Sentiment, Topic)	560.9762	35
(NN: Link)	561.1544	36
(LR: Link)	561.1976	37
(GBDT: Link)	562.5274	38
(KNN: Hashtag, Topic)	562.8344	39
(SVR: Link, Sentiment)	562.8507	40
(NN: Hashtag, Sentiment)	563.4603	41
(LR: Link, Hashtag, Sentiment)	563.7053	42
(LR: Link, Hashtag)	564.0191	43
(RF: Link, Hashtag, Topic)	564.1742	44
(RF: Hashtag, Topic)	564.4587	45
(KNN: Link)	565.5654	46
(NN: Link, Sentiment, Topic)	565.7878	47
(KNN: Link, Sentiment, Topic)	566.2347	48
(GBDT: Link, Hashtag)	567.5543	49
(KNN: Link, Hashtag, Topic)	567.6087	50
(NN: Sentiment)	568.1008	51
(SVR: Link, Hashtag, Sentiment)	568.1205	52
(RF: Hashtag, Sentiment)	568.1486	53
(LR: Sentiment)	568.1879	54
(SVR: Hashtag)	568.2148	55
(SVR: Link, Hashtag)	568.215	56
(SVR: Link)	568.2151	57
(RF: Sentiment)	568.3012	58
(GBDT: Sentiment)	568.3571	59
(NN: Hashtag)	568.466	60
(SVR: Sentiment)	569.067	61
(SVR: Hashtag, Sentiment)	569.1762	62
(RF: Link, Hashtag, Sentiment)	569.3687	63
(KNN: Sentiment)	570.1166	64
(LR: Hashtag, Sentiment)	570.9473	65
(KNN: Hashtag, Sentiment)	571.0404	66
(GBDT: Hashtag)	571.1423	67
(LR: Hashtag)	571.1734	68
(NN: Link, Hashtag, Sentiment)	571.4303	69
(GBDT: Link, Hashtag, Sentiment)	571.5927	70
(GBDT: Hashtag, Sentiment)	571.6723	71
(RF: Hashtag)	571.7171	72
(KNN: Link, Hashtag, Sentiment, Topic)	571.9372	73
(KNN: Topic)	572.4049	74
(GBDT: Link, Hashtag, Sentiment, Topic)	572.6732	75
(SVR: Hashtag, Topic)	572.6841	76
(SVR: Link, Hashtag, Topic)	573.3331	77
(KNN: Link, Topic)	573.8937	78
(SVR: Link, Topic)	574.2545	79
(NN: Hashtag, Topic)	574.3832	80
(GBDT: Link, Hashtag, Topic)	576.7551	81
(NN: Link, Hashtag, Topic)	577.7307	82
(KNN: Link, Sentiment)	578.4164	83
(GBDT: Hashtag, Topic)	578.4524	84
(KNN: Hashtag)	579.2142	85
(NN: Link, Hashtag, Sentiment, Topic)	579.2778	86
(KNN: Link, Hashtag, Sentiment)	586.1404	87
(RF: Link, Hashtag)	594.5466	88
(NN: Link, Hashtag)	595.2992	89
(KNN: Link, Hashtag)	599.6403	90

Table 11.6: BMW Model-Variable Combination Rankings for OT Data - Retweets

Model-Variable Combination	MAE	Rank
(SVR: Link)	67.92701	1
(SVR: Link, Hashtag)	67.96106	2
(SVR: Hashtag)	67.96109	3
(KNN: Topic)	68.59091	4
(KNN: Link, Sentiment, Topic)	68.59504	5
(GBDT: Sentiment, Topic)	68.96256	6
(NN: Sentiment, Topic)	68.9844	7
(GBDT: Hashtag, Topic)	69.13757	8
(NN: Link, Sentiment, Topic)	69.19418	9
(GBDT: Link, Sentiment, Topic)	69.23204	10
(SVR: Hashtag, Sentiment)	69.26677	11
(SVR: Link, Hashtag, Sentiment)	69.27615	12
(SVR: Sentiment)	69.27982	13
(SVR: Link, Sentiment)	69.27984	14
(GBDT: Link, Topic)	69.30075	15
(KNN: Hashtag, Sentiment, Topic)	69.31318	16
(NN: Topic)	69.33082	17
(KNN: Sentiment, Topic)	69.33167	18
(LR: Topic)	69.33785	19
(LR: Link, Topic)	69.36703	20
(NN: Link, Hashtag, Sentiment, Topic)	69.3686	21
(LR: Sentiment, Topic)	69.3991	22
(LR: Link, Sentiment, Topic)	69.43634	23
(NN: Link, Topic)	69.45204	24
(LR: Hashtag, Topic)	69.49493	25
(LR: Link, Hashtag, Topic)	69.50877	26
(KNN: Link, Sentiment)	69.52501	27
(LR: Hashtag, Sentiment, Topic)	69.55107	28
(LR: Link, Hashtag, Sentiment, Topic)	69.57087	29
(GBDT: Hashtag, Sentiment)	69.60885	30
(GBDT: Link, Hashtag, Sentiment, Topic)	69.65098	31
(GBDT: Link, Hashtag, Topic)	69.66437	32
(NN: Hashtag, Sentiment, Topic)	69.69801	33
(NN: Link, Hashtag, Sentiment)	69.70517	34
(NN: Hashtag, Sentiment)	69.72478	35
(RF: Link, Hashtag, Sentiment)	69.73321	36
(RF: Topic)	69.74716	37
(RF: Link, Sentiment)	69.75781	38
(GBDT: Topic)	69.77216	39
(NN: Hashtag, Topic)	69.77569	40
(RF: Hashtag, Sentiment)	69.78454	41
(RF: Sentiment, Topic)	69.78949	42
(RF: Link)	69.81525	43
(KNN: Hashtag, Topic)	69.82152	44
(RF: Link, Sentiment, Topic)	69.82589	45
(NN: Hashtag)	69.82654	46
(NN: Link)	69.82702	47
(LR: Link)	69.82922	48
(NN: Link, Hashtag)	69.82984	49
(NN: Sentiment)	69.83321	50
(GBDT: Hashtag, Sentiment, Topic)	69.8382	51
(NN: Link, Sentiment)	69.84049	52
(KNN: Hashtag, Sentiment)	69.85807	53

(KNN: Link, Hashtag, Topic)	69.88729	54
(GBDT: Link)	69.88806	55
(RF: Link, Hashtag)	69.92169	56
(LR: Link, Hashtag)	69.93701	57
(RF: Sentiment)	69.93714	58
(LR: Hashtag)	69.94096	59
(LR: Link, Sentiment)	69.94427	60
(LR: Sentiment)	69.94514	61
(RF: Hashtag)	69.95569	62
(GBDT: Link, Hashtag)	69.95829	63
(RF: Link, Topic)	69.9826	64
(RF: Link, Hashtag, Sentiment, Topic)	69.99814	65
(LR: Link, Hashtag, Sentiment)	70.04549	66
(LR: Hashtag, Sentiment)	70.04874	67
(RF: Hashtag, Sentiment, Topic)	70.0655	68
(KNN: Link)	70.06942	69
(GBDT: Link, Hashtag, Sentiment)	70.15129	70
(NN: Link, Hashtag, Topic)	70.1755	71
(GBDT: Link, Sentiment)	70.2938	72
(SVR: Link, Hashtag, Topic)	70.34119	73
(SVR: Hashtag, Sentiment, Topic)	70.3749	74
(RF: Hashtag, Topic)	70.3781	75
(SVR: Sentiment, Topic)	70.39735	76
(SVR: Link, Hashtag, Sentiment, Topic)	70.40167	77
(SVR: Link, Sentiment, Topic)	70.40883	78
(RF: Link, Hashtag, Topic)	70.41262	79
(KNN: Link, Topic)	70.43868	80
(GBDT: Sentiment)	70.47635	81
(GBDT: Hashtag)	70.49146	82
(SVR: Hashtag, Topic)	70.7793	83
(SVR: Link, Topic)	70.8126	84
(SVR: Topic)	70.9014	85
(KNN: Link, Hashtag, Sentiment)	70.98744	86
(KNN: Link, Hashtag)	71.05556	87
(KNN: Link, Hashtag, Sentiment, Topic)	71.10888	88
(KNN: Hashtag)	71.47429	89
(KNN: Sentiment)	71.85648	90

Table 11.7: BMW Model-Variable Combination Rankings for IRT Data - Likes

Model-Variable Combination	MAE	Rank
(KNN: Hashtag, Sentiment)	2.085061	1
(KNN: Sentiment)	2.087269	2
(KNN: Link)	2.095339	3
(KNN: Link, Hashtag)	2.096045	4
(KNN: Hashtag, Topic)	2.098983	5
(KNN: Topic)	2.102583	6
(KNN: Link, Hashtag, Topic)	2.102882	7
(KNN: Link, Topic)	2.103335	8
(KNN: Link, Hashtag, Sentiment)	2.153275	9
(KNN: Link, Sentiment)	2.15647	10
(KNN: Link, Sentiment, Topic)	2.184495	11
(KNN: Link, Hashtag, Sentiment, Topic)	2.184877	12
(KNN: Hashtag)	2.190174	13
(SVR: Hashtag)	2.211864	14
(SVR: Link)	2.211864	15
(SVR: Link, Hashtag)	2.211864	16
(KNN: Hashtag, Sentiment, Topic)	2.252512	17
(KNN: Sentiment, Topic)	2.255474	18

(NN: Hashtag, Sentiment, Topic, NW)	2.271098	19
(RF: Link, Hashtag, Sentiment, Topic, NW)	2.278946	20
(RF: Link, Sentiment, Topic, NW)	2.292971	21
(RF: Hashtag, Sentiment, Topic, NW)	2.296631	22
(KNN: Hashtag, Sentiment, Topic, NW)	2.300803	23
(KNN: Sentiment, Topic, NW)	2.301796	24
(RF: Sentiment, Topic, NW)	2.30338	25
(NN: Link, Sentiment, Topic, NW)	2.311824	26
(RF: Link, Hashtag, Topic, NW)	2.314543	27
(KNN: Link, Hashtag, Sentiment, Topic, NW)	2.341599	28
(RF: Link, Topic, NW)	2.343137	29
(RF: Hashtag, Topic, NW)	2.349466	30
(KNN: Topic, NW)	2.363808	31
(KNN: Hashtag, Topic, NW)	2.365436	32
(RF: Topic, NW)	2.37361	33
(NN: Hashtag, Sentiment, Topic)	2.374285	34
(GBDT: Topic)	2.37533	35
(NN: Hashtag, Topic, NW)	2.37975	36
(NN: Link, Hashtag, Sentiment, Topic, NW)	2.381782	37
(KNN: Link, Topic, NW)	2.382813	38
(RF: Link, Hashtag, Sentiment, Topic)	2.38414	39
(NN: Sentiment, Topic)	2.385328	40
(NN: Link, Hashtag, Sentiment, Topic)	2.386616	41
(GBDT: Link, Hashtag, Sentiment, Topic)	2.388519	42
(RF: Sentiment, Topic)	2.389491	43
(RF: Link, Sentiment, Topic)	2.390903	44
(RF: Hashtag, Sentiment, Topic)	2.392416	45
(NN: Link, Topic)	2.398639	46
(NN: Link, Hashtag, Topic)	2.398649	47
(NN: Link, Sentiment, Topic)	2.399196	48
(LR: Link, Topic)	2.401165	49
(LR: Link, Hashtag, Topic)	2.401756	50
(GBDT: Link, Hashtag, Topic)	2.404276	51
(GBDT: Link, Topic)	2.40459	52
(LR: Topic)	2.406826	53
(LR: Hashtag, Topic)	2.407418	54
(NN: Hashtag, Topic)	2.407597	55
(GBDT: Hashtag, Topic)	2.408983	56
(NN: Topic)	2.409851	57
(LR: Link, Sentiment, Topic)	2.411517	58
(RF: Link, Topic)	2.411863	59
(LR: Link, Hashtag, Sentiment, Topic)	2.412085	60
(LR: Sentiment, Topic)	2.415118	61
(LR: Hashtag, Sentiment, Topic)	2.415684	62
(RF: Topic)	2.417262	63
(KNN: Link, Sentiment, Topic, NW)	2.42004	64
(GBDT: Link, Hashtag, Sentiment, NW)	2.421996	65
(RF: Link, Hashtag, Topic)	2.425601	66
(NN: Link, Sentiment)	2.426736	67
(RF: Hashtag, Topic)	2.427915	68
(KNN: Link, Hashtag, Topic, NW)	2.42827	69
(NN: Link, Hashtag, Sentiment)	2.430224	70
(NN: Link, Topic, NW)	2.439314	71
(GBDT: Sentiment, Topic)	2.443049	72
(NN: Sentiment)	2.443912	73
(LR: Sentiment)	2.444789	74
(NN: Hashtag, Sentiment)	2.445217	75
(LR: Hashtag, Sentiment)	2.445433	76
(GBDT: Hashtag, Sentiment, Topic)	2.446075	77
(LR: Sentiment, NW)	2.446444	78
(LR: Hashtag, Sentiment, NW)	2.447259	79
(NN: Hashtag, Sentiment, NW)	2.448262	80
(LR: Link, Sentiment)	2.451019	81

(LR: Link, Hashtag, Sentiment)	2.451786	82
(KNN: Sentiment, NW)	2.456652	83
(GBDT: Link, Sentiment, Topic)	2.460014	84
(LR: Link, Sentiment, NW)	2.461045	85
(RF: Sentiment)	2.462525	86
(LR: Link, Hashtag, Sentiment, NW)	2.462542	87
(RF: Link, Sentiment)	2.466222	88
(NN: Link, Sentiment, NW)	2.466294	89
(GBDT: Hashtag, Sentiment)	2.46748	90
(RF: Link, Sentiment, NW)	2.4718	91
(NN: Link, Hashtag, Sentiment, NW)	2.473042	92
(KNN: Link, Hashtag, NW)	2.474086	93
(RF: Hashtag, Sentiment, NW)	2.476956	94
(KNN: Link, NW)	2.478373	95
(RF: Sentiment, NW)	2.479825	96
(RF: Link, Hashtag, Sentiment)	2.480957	97
(GBDT: Link, Sentiment)	2.481356	98
(GBDT: Link, Sentiment, Topic, NW)	2.482883	99
(GBDT: Hashtag, Sentiment, Topic, NW)	2.482883	99
(GBDT: Sentiment, NW)	2.482981	101
(GBDT: Sentiment)	2.483169	102
(RF: Hashtag, Sentiment)	2.485387	103
(NN: Sentiment, NW)	2.489275	104
(GBDT: Link, Hashtag, Sentiment)	2.493724	105
(RF: Link, Hashtag, Sentiment, NW)	2.494346	106
(NN: Link, Hashtag, NW)	2.501408	107
(KNN: Hashtag, Sentiment, NW)	2.501959	108
(KNN: Link, Hashtag, Sentiment, NW)	2.503284	109
(KNN: NW)	2.510887	110
(NN: Link, NW)	2.51468	111
(GBDT: Link, Topic, NW)	2.516141	112
(GBDT: Hashtag, Topic, NW)	2.516141	112
(KNN: Link, Sentiment, NW)	2.520049	114
(NN: Hashtag, NW)	2.533653	115
(GBDT: Hashtag, Sentiment, NW)	2.53826	116
(LR: Link, NW)	2.538977	117
(LR: Link, Hashtag, NW)	2.53945	118
(NN: Link)	2.540467	119
(LR: Link)	2.540513	120
(LR: Link, Hashtag)	2.541	121
(NN: Link, Hashtag)	2.54123	122
(RF: Link, Hashtag)	2.541269	123
(RF: Link)	2.543463	124
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	2.544906	125
(LR: Link, Topic, NW)	2.550008	126
(LR: Link, Hashtag, Topic, NW)	2.556117	127
(LR: Topic, NW)	2.55628	128
(LR: Link, Sentiment, Topic, NW)	2.557824	129
(GBDT: Link, Hashtag, NW)	2.56009	130
(GBDT: Sentiment, Topic, NW)	2.561843	131
(LR: Hashtag, Topic, NW)	2.562117	132
(GBDT: Link, Hashtag, Topic, NW)	2.563612	133
(LR: Link, Hashtag, Sentiment, Topic, NW)	2.56394	134
(LR: Sentiment, Topic, NW)	2.564168	135
(SVR: Link, Sentiment, Topic)	2.564567	136
(SVR: Link, Hashtag, Sentiment, Topic)	2.564567	137
(NN: Link, Hashtag, Topic, NW)	2.567132	138
(SVR: Sentiment, NW)	2.567705	139
(SVR: Hashtag, Sentiment, NW)	2.567705	140
(SVR: Sentiment, Topic)	2.568368	141
(SVR: Hashtag, Sentiment, Topic)	2.568368	142
(SVR: Sentiment, Topic, NW)	2.56841	143

(SVR: Hashtag, Sentiment, Topic, NW)	2.56841	144
(SVR: Link, Topic, NW)	2.568444	145
(SVR: Link, Hashtag, Topic, NW)	2.568444	146
(SVR: Sentiment)	2.568652	147
(SVR: Hashtag, Sentiment)	2.568652	148
(SVR: Link, Sentiment)	2.568652	149
(SVR: Link, Hashtag, Sentiment)	2.568652	150
(SVR: Link, Topic)	2.568652	151
(SVR: Link, Hashtag, Topic)	2.568652	152
(SVR: NW)	2.568652	153
(SVR: Hashtag, NW)	2.568652	154
(SVR: Topic)	2.568652	155
(SVR: Hashtag, Topic)	2.568652	156
(SVR: Link, Sentiment, Topic, NW)	2.568656	157
(SVR: Link, Hashtag, Sentiment, Topic, NW)	2.568656	158
(SVR: Link, Hashtag, Sentiment, NW)	2.568738	159
(SVR: Topic, NW)	2.568804	160
SVR: Hashtag, Topic, NW)	2.568804	161
(SVR: Link, Sentiment, NW)	2.568943	162
(SVR: Link, Hashtag, NW)	2.569445	163
(SVR: Link, NW)	2.569592	164
(LR: Hashtag, Sentiment, Topic, NW)	2.570037	165
(GBDT: Link, Hashtag)	2.570727	166
(LR: Hashtag, NW)	2.571692	167
(LR: NW)	2.571849	168
(NN: NW)	2.57229	169
(NN: Topic, NW)	2.578984	170
(RF: Link, NW)	2.587144	171
(RF: Link, Hashtag, NW)	2.590188	172
(GBDT: Link)	2.590506	173
(RF: Hashtag, NW)	2.596816	174
(GBDT: Link, Sentiment, NW)	2.597108	175
(KNN: Hashtag, NW)	2.602097	176
(NN: Sentiment, Topic, NW)	2.604956	177
(LR: Hashtag)	2.607225	178
(NN: Hashtag)	2.607387	179
(GBDT: NW)	2.607578	180
(RF: Hashtag)	2.60801	181
(GBDT: Topic, NW)	2.642832	182
(GBDT: Hashtag)	2.649185	183
(RF: NW)	2.649406	184
(GBDT: Hashtag, NW)	2.661527	185
(GBDT: Link, NW)	2.692948	186

Table 11.8: BMW Model-Variable Combination Rankings for IRT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(KNN: Sentiment)	0.135034	1
(KNN: Link, Hashtag, Sentiment)	0.135089	2
(KNN: Hashtag, Sentiment)	0.135722	3
(KNN: Hashtag, Topic)	0.136003	4
(KNN: Link, Hashtag, Topic)	0.136151	5
(KNN: Topic)	0.136513	6
(KNN: Link, Sentiment)	0.136515	7
(KNN: Link, Topic)	0.136647	8
(KNN: Hashtag, Sentiment, Topic)	0.142607	9
(KNN: Sentiment, Topic)	0.142724	10

(KNN: Link, Sentiment, Topic)	0.149533	11
(KNN: Link, Hashtag, Sentiment, Topic)	0.149567	12
(GBDT: Link, Sentiment)	0.152077	13
(GBDT: Hashtag, Sentiment)	0.152167	14
(GBDT: Sentiment)	0.155357	15
(KNN: Link)	0.155604	16
(KNN: Link, Hashtag)	0.156578	17
(KNN: Hashtag, NW)	0.159261	18
(KNN: Hashtag, Topic, NW)	0.159433	19
(KNN: NW)	0.159808	20
(KNN: Link, NW)	0.16086	21
(RF: NW)	0.16095	22
(KNN: Sentiment, NW)	0.162236	23
(RF: Link, Sentiment, Topic, NW)	0.16281	24
(RF: Link, Hashtag, Sentiment, Topic, NW)	0.163557	25
(KNN: Link, Hashtag, NW)	0.163563	26
(RF: Link, Hashtag, Topic, NW)	0.163841	27
(KNN: Topic, NW)	0.164437	28
(RF: Link, Topic, NW)	0.164507	29
(KNN: Sentiment, Topic, NW)	0.16452	30
(GBDT: Hashtag, Topic)	0.16477	31
(RF: Hashtag, Sentiment, Topic, NW)	0.165091	32
(KNN: Hashtag, Sentiment, NW)	0.1652	33
(RF: Hashtag, Topic, NW)	0.165303	34
(KNN: Link, Hashtag, Topic, NW)	0.165464	35
(RF: Sentiment, Topic, NW)	0.165483	36
(GBDT: Sentiment, Topic)	0.165759	37
(KNN: Link, Sentiment, Topic, NW)	0.165832	38
(KNN: Link, Hashtag, Sentiment, Topic, NW)	0.165862	39
(NN: Hashtag, Sentiment, Topic)	0.166068	40
(GBDT: Link, Hashtag, Sentiment, NW)	0.166532	41
(KNN: Link, Hashtag, Sentiment, NW)	0.166587	42
(NN: Link, Sentiment, Topic, NW)	0.166791	43
(GBDT: Link, Topic)	0.166927	44
(GBDT: Link, Sentiment, NW)	0.166959	45
(GBDT: Hashtag, Sentiment, NW)	0.166973	46
(GBDT: Sentiment, NW)	0.167156	47
(RF: Sentiment, NW)	0.167217	48
(NN: Sentiment, Topic)	0.167297	49
(RF: Link, Sentiment, NW)	0.167548	50
(GBDT: Sentiment, Topic, NW)	0.167653	51
(KNN: Link, Sentiment, NW)	0.167667	52
(GBDT: Link, Hashtag, Topic)	0.168293	53
(RF: Topic, NW)	0.168322	54
(NN: Sentiment, NW)	0.16841	55
(NN: Link, Hashtag, Sentiment, Topic)	0.168521	56
(GBDT: Link, Sentiment, Topic, NW)	0.168531	57
(GBDT: Hashtag, Sentiment, Topic, NW)	0.168531	57
(KNN: Link, Topic, NW)	0.168615	59
(RF: Link, NW)	0.168885	60
(KNN: Hashtag, Sentiment, Topic, NW)	0.169027	61
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	0.169028	62
(GBDT: Link, Hashtag, Topic, NW)	0.169417	63
(GBDT: Hashtag, Topic, NW)	0.169774	64
(GBDT: Link, Topic, NW)	0.169938	65
(RF: Link, Hashtag, Sentiment, NW)	0.169978	66
(GBDT: Topic, NW)	0.170032	67
(NN: Link, Hashtag, Sentiment, Topic, NW)	0.17005	68
(NN: Hashtag, Sentiment, NW)	0.170348	69
(LR: Link, Sentiment)	0.170666	70
(NN: Sentiment)	0.170679	71
(LR: Sentiment)	0.170709	72

(RF: Hashtag, Sentiment, NW)	0.170722	73
(LR: Link, Sentiment, NW)	0.170844	74
(LR: Link, Hashtag, Sentiment)	0.170859	75
(LR: Hashtag, Sentiment)	0.170897	76
(NN: Link, Sentiment)	0.170905	77
(LR: Sentiment, NW)	0.170906	78
(NN: Link, Hashtag, Sentiment)	0.170926	79
(NN: Hashtag, Sentiment)	0.170933	80
(NN: Topic, NW)	0.170977	81
(LR: Link, Hashtag, Sentiment, NW)	0.171004	82
(NN: Link, Topic)	0.171038	83
(NN: Link, Hashtag, Topic)	0.171045	84
(LR: Hashtag, Sentiment, NW)	0.17106	85
(NN: Link, Hashtag, Sentiment, NW)	0.1711	86
(GBDT: NW)	0.171361	87
(LR: Link, Sentiment, Topic)	0.171384	88
(LR: Sentiment, Topic)	0.171385	89
(NN: Hashtag, Sentiment, Topic, NW)	0.17144	90
(NN: Link, Sentiment, NW)	0.17144	91
(NN: Link, Hashtag, Topic, NW)	0.171636	92
(LR: Link, Hashtag, Sentiment, Topic)	0.171658	93
(LR: Hashtag, Sentiment, Topic)	0.171658	94
(LR: Sentiment, Topic, NW)	0.171664	95
(LR: Link, Sentiment, Topic, NW)	0.171665	96
(GBDT: Link, Hashtag, Sentiment, Topic)	0.171752	97
(LR: Hashtag, Sentiment, Topic, NW)	0.172001	98
(LR: Link, Hashtag, Sentiment, Topic, NW)	0.172002	99
(GBDT: Link, NW)	0.172012	100
(NN: Link, Topic, NW)	0.172093	101
(LR: Topic)	0.172191	102
(LR: Link, Topic)	0.172195	103
(NN: Topic)	0.172224	104
(NN: Hashtag, Topic)	0.17223	105
(GBDT: Link, Hashtag, NW)	0.172413	106
(GBDT: Hashtag, NW)	0.172442	107
(KNN: Hashtag)	0.172444	108
(LR: Topic, NW)	0.17245	109
(LR: Link, Topic, NW)	0.172454	110
(LR: Hashtag, Topic)	0.172472	111
(LR: Link, Hashtag, Topic)	0.172476	112
(NN: Sentiment, Topic, NW)	0.172703	113
(RF: Sentiment, Topic)	0.172732	114
(RF: Link, Hashtag, NW)	0.172743	115
(RF: Hashtag, NW)	0.172746	116
(LR: Hashtag, Topic, NW)	0.172792	117
(LR: Link, Hashtag, Topic, NW)	0.172796	118
(GBDT: Hashtag, Sentiment, Topic)	0.172845	119
(NN: Link, Hashtag, NW)	0.173797	120
(RF: Sentiment)	0.173809	121
(RF: Hashtag, Sentiment, Topic)	0.174177	122
(RF: Hashtag, Sentiment)	0.174194	123
(RF: Link, Sentiment, Topic)	0.174218	124
(NN: Link, Sentiment, Topic)	0.174219	125
(RF: Link, Sentiment)	0.174259	126
(RF: Link, Hashtag, Sentiment, Topic)	0.174714	127
(NN: Hashtag, NW)	0.174958	128
(RF: Link, Topic)	0.174978	129
(NN: Link, NW)	0.175003	130
(RF: Link, Hashtag, Sentiment)	0.175296	131
(GBDT: Link, Sentiment, Topic)	0.175561	132
(RF: Hashtag, Topic)	0.175875	133
(GBDT: Topic)	0.175902	134
(RF: Topic)	0.175961	135

(RF: Link, Hashtag, Topic)	0.176697	136
(LR: Link, NW)	0.178179	137
(LR: Link, Hashtag, NW)	0.178305	138
(NN: NW)	0.178513	139
(LR: NW)	0.178585	140
(LR: Hashtag, NW)	0.178691	141
(RF: Link)	0.17875	142
(LR: Link)	0.17914	143
(NN: Link)	0.17923	144
(NN: Link, Hashtag)	0.179312	145
(LR: Link, Hashtag)	0.179315	146
(RF: Hashtag)	0.179695	147
(GBDT: Link, Hashtag, Sentiment)	0.179928	148
(GBDT: Link, Hashtag)	0.180005	149
(NN: Hashtag, Topic, NW)	0.180367	150
(RF: Link, Hashtag)	0.180444	151
(NN: Hashtag)	0.180467	152
(LR: Hashtag)	0.180591	153
(GBDT: Link)	0.184075	154
(GBDT: Hashtag)	0.18499	155
(SVR: Link, NW)	0.196114	156
(SVR: Link, Hashtag, NW)	0.196114	157
(SVR: Sentiment, Topic, NW)	0.196158	158
(SVR: Link, Hashtag, Sentiment, Topic, NW)	0.196161	159
(SVR: Hashtag, Sentiment, Topic, NW)	0.196164	160
(SVR: Link, Sentiment, Topic, NW)	0.196202	161
(SVR: Link, Topic, NW)	0.196245	162
(SVR: Link, Hashtag, Topic, NW)	0.196245	163
(SVR: Sentiment)	0.196248	164
(SVR: Hashtag, Sentiment)	0.196248	165
(SVR: Link, Hashtag, Topic)	0.196248	166
(SVR: Link, Sentiment)	0.196248	167
(SVR: Link, Hashtag, Sentiment)	0.196248	168
(SVR: Link, Topic)	0.196248	169
(SVR: NW)	0.196248	170
(SVR: Hashtag, NW)	0.196248	171
(SVR: Topic)	0.196248	172
(SVR: Hashtag, Topic)	0.196248	173
(SVR: Sentiment, Topic)	0.196248	174
(SVR: Hashtag, Sentiment, Topic)	0.196248	175
(SVR: Topic, NW)	0.196258	176
SVR: Hashtag, Topic, NW)	0.196258	177
(SVR: Link, Sentiment, NW)	0.196309	178
(SVR: Link, Hashtag, Sentiment, NW)	0.196309	179
(SVR: Sentiment, NW)	0.19631	180
(SVR: Hashtag, Sentiment, NW)	0.19631	181
(SVR: Link, Hashtag, Sentiment, Topic)	0.200925	182
(SVR: Link, Sentiment, Topic)	0.200925	183
(SVR: Link)	0.242938	184
(SVR: Link, Hashtag)	0.242938	185
(SVR: Hashtag)	0.242938	186

Table 11.9: Coca Cola Model-Variable Combination Rankings for IRT Data - Likes

Model-Variable Combination	MAE	Rank
(SVR: Link, Hashtag, Sentiment, Topic, NW)	1.488938	1
(SVR: Link, Hashtag, Sentiment, NW)	1.529132	2
(SVR: Link, Hashtag, Topic, NW)	1.542005	3
(SVR: Link, Sentiment, NW)	1.549922	4

(SVR: Link, Sentiment, Topic, NW)	1.556261	5
(SVR: Link, Topic, NW)	1.574672	6
SVR: Hashtag, Topic, NW)	1.613043	7
(SVR: Hashtag, Sentiment, NW)	1.628287	8
(SVR: Sentiment, NW)	1.628602	9
(SVR: Link, Hashtag, Topic)	1.630315	10
(SVR: Link, Hashtag, Sentiment, Topic)	1.630415	11
(SVR: Hashtag, Sentiment, Topic, NW)	1.644608	12
(SVR: Hashtag, NW)	1.647498	13
(SVR: Link, Topic)	1.664129	14
(SVR: Link, NW)	1.664129	15
(SVR: NW)	1.664129	16
(SVR: Sentiment, Topic)	1.664129	17
(SVR: Topic, NW)	1.664129	18
(SVR: Link, Sentiment, Topic)	1.664129	19
(SVR: Topic)	1.664129	20
(SVR: Link, Sentiment)	1.664129	21
(SVR: Link, Hashtag, Sentiment)	1.664129	22
(SVR: Sentiment)	1.664129	23
(SVR: Hashtag, Sentiment)	1.664129	24
(SVR: Link, Hashtag, NW)	1.678458	25
(SVR: Hashtag, Sentiment, Topic)	1.692555	26
(SVR: Hashtag, Topic)	1.696	27
(SVR: Sentiment, Topic, NW)	1.698146	28
(SVR: Hashtag)	1.827411	29
(SVR: Link)	1.827411	30
(SVR: Link, Hashtag)	1.827411	31
(KNN: Link, Hashtag)	2.085448	32
(KNN: Link, Sentiment)	3.075577	33
(KNN: Link, Hashtag, Sentiment)	3.145983	34
(KNN: Link, Hashtag, Sentiment, Topic)	3.781759	35
(KNN: Hashtag, Sentiment, Topic)	3.804564	36
(KNN: Sentiment, Topic)	3.805334	37
(KNN: Link, Sentiment, Topic)	4.258217	38
(KNN: Link, Hashtag, NW)	4.604844	39
(KNN: Link, Hashtag, Topic, NW)	4.7425	40
(KNN: Link, Topic, NW)	4.777689	41
(KNN: Hashtag)	4.945167	42
(KNN: Link, NW)	6.265099	43
(KNN: Link, Hashtag, Sentiment, Topic, NW)	6.379737	44
(KNN: Link, Sentiment, Topic, NW)	6.488996	45
(KNN: NW)	6.600636	46
(KNN: Sentiment, Topic, NW)	6.682172	47
(KNN: Hashtag, NW)	6.738814	48
(KNN: Hashtag, Sentiment)	7.568382	49
(KNN: Hashtag, Topic, NW)	7.905099	50
(KNN: Sentiment)	7.939656	51
(KNN: Link, Sentiment, NW)	8.205663	52
(KNN: Link, Hashtag, Sentiment, NW)	8.223551	53
(KNN: Hashtag, Sentiment, Topic, NW)	8.557123	54
(KNN: Link, Topic)	9.445467	55
(KNN: Link, Hashtag, Topic)	9.88781	56
(KNN: Topic, NW)	10.24543	57
(KNN: Hashtag, Topic)	12.86089	58
(KNN: Topic)	13.70809	59
(KNN: Hashtag, Sentiment, NW)	13.80203	60
(KNN: Sentiment, NW)	13.8291	61
(GBDT: Hashtag)	14.68471	62
(NN: Link, Sentiment, NW)	20.31638	63
(RF: Hashtag, NW)	22.25835	64
(RF: Sentiment, NW)	22.88477	65
(RF: NW)	23.02478	66
(KNN: Link)	23.10163	67

(GBDT: NW)	23.17293	68
(RF: Topic, NW)	23.4957	69
(RF: Hashtag, Topic, NW)	24.46987	70
(RF: Hashtag, Sentiment, NW)	25.06572	71
(GBDT: Hashtag, Sentiment, Topic, NW)	25.49961	72
(RF: Link, Sentiment, NW)	25.52459	73
(RF: Topic)	25.59749	74
(RF: Sentiment, Topic)	25.64805	75
(GBDT: Link)	25.76927	76
(RF: Sentiment)	25.9962	77
(RF: Link, Hashtag, Sentiment, NW)	26.07837	78
(RF: Sentiment, Topic, NW)	26.10506	79
(LR: NW)	26.33979	80
(RF: Hashtag, Sentiment, Topic)	26.47878	81
(LR: Hashtag, NW)	26.48432	82
(RF: Link, Sentiment)	26.70716	83
(RF: Hashtag, Sentiment)	26.76376	84
(NN: Link, Sentiment)	27.13713	85
(RF: Link, Sentiment, Topic)	27.15271	86
(NN: Hashtag, NW)	27.16173	87
(NN: Link, Hashtag)	27.37617	88
(NN: Link, Hashtag, Sentiment, NW)	27.47514	89
(RF: Link, Topic)	27.48466	90
(NN: Link, Hashtag, Sentiment)	27.50797	91
(NN: Hashtag, Topic, NW)	27.55694	92
(NN: Topic, NW)	27.56383	93
(NN: Hashtag, Sentiment, NW)	27.58834	94
(NN: Link, Hashtag, Sentiment, Topic)	27.59466	95
(RF: Link, Hashtag, NW)	27.62052	96
(NN: Topic)	27.64916	97
(NN: Hashtag)	27.65693	98
(NN: Link, Hashtag, Topic, NW)	27.66558	99
(NN: Hashtag, Sentiment, Topic, NW)	27.68751	100
(LR: Topic)	27.72167	101
(NN: Sentiment, Topic, NW)	27.72481	102
(LR: Hashtag)	27.72871	103
(NN: Sentiment, Topic)	27.73153	104
(NN: Link, Topic, NW)	27.7316	105
(RF: Hashtag, Topic)	27.77307	106
(NN: Hashtag, Topic)	27.77553	107
(RF: Link, Hashtag, Topic, NW)	27.7972	108
(NN: Link, Hashtag, Topic)	27.82879	109
(NN: Sentiment)	27.83881	110
(RF: Link, Hashtag, Sentiment, Topic, NW)	27.89966	111
(NN: Link, Sentiment, Topic, NW)	27.90273	112
(LR: Hashtag, Topic)	27.90605	113
(LR: Sentiment)	27.96264	114
(NN: Link, Topic)	27.96812	115
(NN: Hashtag, Sentiment)	27.99201	116
(RF: Hashtag, Sentiment, Topic, NW)	28.00657	117
(RF: Link, Hashtag, Sentiment)	28.09366	118
(LR: Topic, NW)	28.10993	119
(LR: Hashtag, Sentiment)	28.1277	120
(NN: Link, Hashtag, NW)	28.13035	121
(NN: Link)	28.17045	122
(LR: Hashtag, Topic, NW)	28.29033	123
(RF: Hashtag)	28.33745	124
(RF: Link, Hashtag, Sentiment, Topic)	28.34593	125
(NN: Sentiment, NW)	28.35818	126
(RF: Link, Sentiment, Topic, NW)	28.48948	127
(NN: Link, NW)	28.65906	128
(GBDT: Sentiment, Topic, NW)	28.79669	129
(GBDT: Hashtag, NW)	28.83888	130

(NN: Hashtag, Sentiment, Topic)	29.29577	131
(GBDT: Hashtag, Topic, NW)	29.44344	132
(RF: Link, NW)	29.72066	133
(RF: Link, Hashtag, Topic)	29.797	134
(RF: Link, Hashtag)	29.9274	135
(NN: NW)	30.25289	136
(RF: Link, Topic, NW)	30.56093	137
(GBDT: Sentiment, NW)	30.94719	138
(LR: Link)	31.61294	139
(LR: Link, Hashtag)	31.68771	140
(GBDT: Topic, NW)	32.24823	141
(LR: Sentiment, NW)	32.35827	142
(RF: Link)	32.42534	143
(GBDT: Topic)	32.52254	144
(LR: Hashtag, Sentiment, NW)	32.6252	145
(GBDT: Sentiment)	32.71553	146
(GBDT: Hashtag, Sentiment)	34.46929	147
(GBDT: Link, Sentiment, Topic, NW)	34.81865	148
(GBDT: Link, Hashtag, NW)	35.01159	149
(GBDT: Link, Hashtag)	36.34078	150
(GBDT: Hashtag, Sentiment, NW)	37.38874	151
(GBDT: Hashtag, Topic)	37.61527	152
(NN: Link, Hashtag, Sentiment, Topic, NW)	37.88062	153
(LR: Sentiment, Topic, NW)	38.08875	154
(LR: Sentiment, Topic)	38.15629	155
(LR: Hashtag, Sentiment, Topic, NW)	38.44677	156
(LR: Hashtag, Sentiment, Topic)	38.51954	157
(GBDT: Sentiment, Topic)	40.81346	158
(GBDT: Link, Hashtag, Topic)	41.00347	159
(GBDT: Link, Hashtag, Topic, NW)	42.44949	160
(LR: Link, NW)	42.47886	161
(LR: Link, Hashtag, NW)	42.55416	162
(GBDT: Link, Hashtag, Sentiment, NW)	43.52373	163
(GBDT: Link, Hashtag, Sentiment, Topic)	44.54946	164
(GBDT: Link, Topic)	44.79042	165
(GBDT: Link, NW)	44.91467	166
(LR: Link, Topic)	45.9633	167
(LR: Link, Hashtag, Topic)	46.06993	168
(LR: Link, Topic, NW)	46.49724	169
(LR: Link, Hashtag, Topic, NW)	46.59595	170
(GBDT: Hashtag, Sentiment, Topic)	49.133	171
(NN: Link, Sentiment, Topic)	49.13625	172
(GBDT: Link, Topic, NW)	49.23935	173
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	55.3565	174
(GBDT: Link, Hashtag, Sentiment)	55.85937	175
(LR: Link, Sentiment, NW)	57.84184	176
(GBDT: Link, Sentiment)	57.97971	177
(LR: Link, Hashtag, Sentiment, NW)	58.10225	178
(LR: Link, Sentiment, Topic, NW)	58.52535	179
(GBDT: Link, Sentiment, NW)	58.75497	180
(LR: Link, Hashtag, Sentiment, Topic, NW)	58.81078	181
(LR: Link, Sentiment, Topic)	59.21061	182
(LR: Link, Sentiment)	59.2563	183
(LR: Link, Hashtag, Sentiment)	59.46599	184
(LR: Link, Hashtag, Sentiment, Topic)	59.4886	185
(GBDT: Link, Sentiment, Topic)	61.60104	186

Table 11.10: Coca Cola Model-Variable Combination Rankings for IRT Data - Retweets

Model-Variable Combination	MAE	Rank
(SVR: Sentiment, NW)	0.093063	1
(SVR: Hashtag, Sentiment, NW)	0.093063	2
(SVR: Link, Hashtag, Topic, NW)	0.093063	3
(SVR: Link, Sentiment)	0.093063	4
(SVR: Sentiment)	0.093063	5
(SVR: Link, Hashtag, Sentiment)	0.093063	6
(SVR: Hashtag, Sentiment)	0.093063	7
(SVR: Link, Topic, NW)	0.093063	8
(SVR: Topic, NW)	0.093063	9
SVR: Hashtag, Topic, NW)	0.093063	10
(SVR: NW)	0.093063	11
(SVR: Sentiment, Topic)	0.093063	12
(SVR: Link, NW)	0.093063	13
(SVR: Link, Hashtag, Sentiment, Topic)	0.093063	14
(SVR: Hashtag, Topic)	0.093063	15
(SVR: Hashtag, Sentiment, Topic)	0.093063	16
(SVR: Link, Hashtag, Topic)	0.093063	17
(SVR: Link, Hashtag, NW)	0.093063	18
(SVR: Link, Topic)	0.093063	19
(SVR: Link, Sentiment, Topic)	0.093063	20
(SVR: Topic)	0.093063	21
(SVR: Hashtag, NW)	0.093063	22
(SVR: Link, Sentiment, Topic, NW)	0.101625	23
(SVR: Link, Hashtag, Sentiment, Topic, NW)	0.121123	24
(SVR: Hashtag, Sentiment, Topic, NW)	0.122039	25
(SVR: Link, Hashtag, Sentiment, NW)	0.128133	26
(SVR: Link, Sentiment, NW)	0.13553	27
(SVR: Sentiment, Topic, NW)	0.139223	28
(KNN: Link, Sentiment)	0.149619	29
(KNN: Link, Hashtag, Sentiment, Topic)	0.149954	30
(KNN: NW)	0.150253	31
(KNN: Link, Topic)	0.152051	32
(KNN: Link, Sentiment, Topic)	0.153489	33
(KNN: Topic)	0.156273	34
(KNN: Link, Hashtag, Sentiment)	0.15641	35
(KNN: Link, Hashtag, Topic)	0.157131	36
(KNN: Hashtag, Sentiment, NW)	0.158472	37
(KNN: Hashtag, Topic)	0.158473	38
(KNN: Link, Sentiment, Topic, NW)	0.159146	39
(KNN: Link, Hashtag, Sentiment, Topic, NW)	0.15974	40
(KNN: Link, NW)	0.160354	41
(KNN: Link)	0.162146	42
(KNN: Link, Hashtag, NW)	0.162505	43
(KNN: Sentiment, NW)	0.163733	44
(KNN: Sentiment, Topic, NW)	0.165847	45
(KNN: Hashtag, Sentiment, Topic, NW)	0.167702	46
(KNN: Link, Hashtag)	0.169405	47
(KNN: Link, Hashtag, Topic, NW)	0.178886	48
(KNN: Sentiment, Topic)	0.181377	49
(KNN: Hashtag, Sentiment, Topic)	0.181411	50
(KNN: Hashtag, Topic, NW)	0.182734	51
(KNN: Topic, NW)	0.182987	52
(KNN: Link, Topic, NW)	0.185607	53
(KNN: Sentiment)	0.208432	54
(KNN: Hashtag, NW)	0.209712	55
(KNN: Hashtag, Sentiment)	0.213414	56
(KNN: Hashtag)	0.228176	57
(KNN: Link, Hashtag, Sentiment, NW)	0.620216	58
(KNN: Link, Sentiment, NW)	0.647216	59
(SVR: Link)	1.005076	60

(SVR: Link, Hashtag)	1.005076	61
(SVR: Hashtag)	1.005076	62
(GBDT: Hashtag)	1.021874	63
(GBDT: Link)	1.239001	64
(RF: Hashtag, NW)	3.340087	65
(RF: NW)	3.380182	66
(RF: Sentiment, NW)	3.481072	67
(GBDT: NW)	3.51423	68
(RF: Hashtag, Sentiment, Topic)	3.741796	69
(RF: Hashtag, Topic, NW)	3.775567	70
(RF: Topic, NW)	3.798889	71
(RF: Link, Hashtag, NW)	3.821398	72
(RF: Hashtag, Sentiment, NW)	3.852609	73
(RF: Sentiment, Topic)	3.861133	74
(RF: Topic)	3.867336	75
(RF: Sentiment)	3.886629	76
(RF: Link, Hashtag, Sentiment, NW)	3.915648	77
(NN: Hashtag, NW)	3.939258	78
(RF: Sentiment, Topic, NW)	3.962276	79
(RF: Link, Sentiment)	3.970664	80
(RF: Link, Hashtag, Sentiment, Topic, NW)	3.985237	81
(RF: Hashtag, Sentiment)	4.010063	82
(LR: NW)	4.013541	83
(RF: Link, Hashtag, Sentiment)	4.016283	84
(RF: Hashtag, Topic)	4.033487	85
(LR: Hashtag, NW)	4.036116	86
(NN: Link, NW)	4.043711	87
(NN: Link, Hashtag, Sentiment, Topic, NW)	4.049929	88
(NN: Sentiment, Topic, NW)	4.074546	89
(NN: Link, Hashtag, Sentiment, Topic)	4.108376	90
(RF: Link, Topic)	4.117833	91
(NN: Hashtag, Topic, NW)	4.126112	92
(NN: Hashtag, Sentiment, NW)	4.128891	93
(NN: Sentiment, Topic)	4.130836	94
(NN: Link, Hashtag, Sentiment)	4.15003	95
(NN: Hashtag, Topic)	4.152275	96
(RF: Link, Topic, NW)	4.15505	97
(NN: Link, Sentiment)	4.159492	98
(NN: Hashtag)	4.161019	99
(NN: Link, Hashtag)	4.161472	100
(NN: Hashtag, Sentiment, Topic, NW)	4.162376	101
(NN: Link, Hashtag, Topic, NW)	4.163495	102
(NN: Link, Hashtag, Topic)	4.165075	103
(NN: Topic)	4.172181	104
(NN: Link, Topic, NW)	4.175649	105
(LR: Hashtag)	4.177011	106
(RF: Hashtag, Sentiment, Topic, NW)	4.185	107
(NN: Link, Sentiment, NW)	4.186045	108
(LR: Sentiment)	4.198383	109
(NN: Link, Topic)	4.201664	110
(LR: Topic)	4.20615	111
(NN: Sentiment)	4.213108	112
(RF: Link, Sentiment, Topic)	4.220669	113
(NN: Hashtag, Sentiment)	4.221087	114
(LR: Hashtag, Sentiment)	4.224803	115
(RF: Link, Sentiment, NW)	4.229986	116
(LR: Hashtag, Topic)	4.236127	117
(RF: Link, Hashtag, Sentiment, Topic)	4.249399	118
(LR: Topic, NW)	4.264667	119
(NN: Link)	4.2734	120
(RF: Hashtag)	4.277225	121
(NN: Sentiment, NW)	4.281908	122
(RF: Link, Hashtag, Topic)	4.282361	123

(LR: Hashtag, Topic, NW)	4.292997	124
(RF: Link, Hashtag)	4.313366	125
(GBDT: Sentiment, Topic, NW)	4.325878	126
(NN: NW)	4.328608	127
(RF: Link, NW)	4.368716	128
(GBDT: Hashtag, Sentiment, NW)	4.400791	129
(GBDT: Hashtag, NW)	4.42533	130
(RF: Link, Sentiment, Topic, NW)	4.463594	131
(NN: Link, Sentiment, Topic, NW)	4.588522	132
(RF: Link, Hashtag, Topic, NW)	4.61557	133
(GBDT: Topic)	4.66406	134
(GBDT: Hashtag, Sentiment, Topic, NW)	4.723217	135
(LR: Link)	4.741128	136
(LR: Link, Hashtag)	4.753587	137
(NN: Hashtag, Sentiment, Topic)	4.763146	138
(GBDT: Sentiment, Topic)	4.765661	139
(NN: Link, Hashtag, NW)	4.817675	140
(RF: Link)	4.871645	141
(GBDT: Hashtag, Topic, NW)	4.938684	142
(LR: Sentiment, NW)	4.941963	143
(LR: Hashtag, Sentiment, NW)	4.984106	144
(NN: Link, Hashtag, Sentiment, NW)	5.061564	145
(NN: Link, Sentiment, Topic)	5.22226	146
(NN: Topic, NW)	5.447467	147
(GBDT: Sentiment, NW)	5.558583	148
(GBDT: Topic, NW)	5.728756	149
(GBDT: Hashtag, Sentiment)	5.750147	150
(LR: Sentiment, Topic)	5.8418	151
(LR: Sentiment, Topic, NW)	5.850409	152
(LR: Hashtag, Sentiment, Topic)	5.898927	153
(GBDT: Hashtag, Topic)	5.903721	154
(LR: Hashtag, Sentiment, Topic, NW)	5.906736	155
(GBDT: Link, Hashtag, Topic, NW)	5.927729	156
(GBDT: Link, Hashtag)	6.14503	157
(GBDT: Link, Hashtag, Sentiment, NW)	6.422961	158
(GBDT: Sentiment)	6.475513	159
(LR: Link, NW)	6.582224	160
(LR: Link, Hashtag, NW)	6.594539	161
(GBDT: Link, Hashtag, NW)	6.845649	162
(GBDT: Link, Hashtag, Sentiment, Topic)	6.901315	163
(GBDT: Link, NW)	7.077855	164
(LR: Link, Topic)	7.111435	165
(LR: Link, Hashtag, Topic)	7.128557	166
(LR: Link, Topic, NW)	7.239969	167
(GBDT: Link, Sentiment, Topic, NW)	7.252131	168
(LR: Link, Hashtag, Topic, NW)	7.255138	169
(GBDT: Hashtag, Sentiment, Topic)	7.269981	170
(GBDT: Link, Topic)	8.205237	171
(GBDT: Link, Sentiment, NW)	8.320226	172
(GBDT: Link, Topic, NW)	8.354159	173
(GBDT: Link, Hashtag, Topic)	8.462193	174
(LR: Link, Sentiment, NW)	9.04646	175
(LR: Link, Hashtag, Sentiment, NW)	9.0867	176
(LR: Link, Sentiment, Topic, NW)	9.154991	177
(LR: Link, Hashtag, Sentiment, Topic, NW)	9.198881	178
(LR: Link, Sentiment)	9.232994	179
(LR: Link, Sentiment, Topic)	9.249649	180
(LR: Link, Hashtag, Sentiment)	9.266759	181
(LR: Link, Hashtag, Sentiment, Topic)	9.293263	182
(GBDT: Link, Sentiment, Topic)	10.2479	183
(GBDT: Link, Hashtag, Sentiment)	10.50067	184
(GBDT: Link, Sentiment)	11.17441	185

(GBDT: Link, Hashtag, Sentiment, Topic, NW)	11.61405	186

Table 11.11: Disney Model-Variable Combination Rankings for OT Data - Likes

Model-Variable Combination	**MAE**	**Rank**
(SVR: Link, Hashtag)	1942.779	1
(SVR: Hashtag)	1943.78	2
(SVR: Link)	2000.928	3
(KNN: Link, Hashtag, Sentiment)	2087.964	4
(KNN: Link, Sentiment)	2099.995	5
(KNN: Sentiment)	2108.055	6
(KNN: Hashtag, Sentiment)	2131.085	7
(KNN: Link, Hashtag)	2138.161	8
(KNN: Link)	2147.15	9
(KNN: Hashtag)	2186	10
(GBDT: Link, Hashtag, Topic)	2381.976	11
(RF: Link, Hashtag)	2401.971	12
(NN: Link, Hashtag)	2407.025	13
(GBDT: Link, Hashtag, Sentiment)	2423.883	14
(NN: Link, Hashtag, Topic)	2429.111	15
(SVR: Link, Hashtag, Sentiment)	2433.284	16
(LR: Link, Hashtag)	2438.672	17
(RF: Link, Hashtag, Topic)	2442.468	18
(NN: Hashtag)	2444.241	19
(LR: Hashtag)	2444.298	20
(RF: Hashtag)	2445.527	21
(GBDT: Link, Hashtag)	2446.881	22
(SVR: Link, Hashtag, Topic)	2449.084	23
(SVR: Hashtag, Sentiment)	2457.431	24
(GBDT: Hashtag, Topic)	2461.321	25
(SVR: Link, Hashtag, Sentiment, Topic)	2462.977	26
(SVR: Link, Topic)	2464.225	27
(NN: Link, Hashtag, Sentiment, Topic)	2464.865	28
(SVR: Link, Sentiment)	2467.147	29
(SVR: Link, Sentiment, Topic)	2470.848	30
(GBDT: Link, Hashtag, Sentiment, Topic)	2470.965	31
(RF: Link, Hashtag, Sentiment, Topic)	2471.538	32
(SVR: Hashtag, Topic)	2471.948	33
(LR: Link, Hashtag, Sentiment)	2472.252	34
(SVR: Sentiment)	2473.091	35
(LR: Hashtag, Sentiment)	2477.994	36
(GBDT: Link, Topic)	2479.575	37
(RF: Link, Hashtag, Sentiment)	2480.824	38
(SVR: Hashtag, Sentiment, Topic)	2481.94	39
(NN: Hashtag, Topic)	2482.326	40
(RF: Hashtag, Topic)	2482.477	41
(RF: Hashtag, Sentiment)	2485.984	42
(SVR: Topic)	2486.063	43
(LR: Link)	2486.812	44
(SVR: Sentiment, Topic)	2486.994	45
(RF: Link)	2489.321	46
(NN: Hashtag, Sentiment)	2489.367	47
(NN: Link)	2490.773	48
(LR: Link, Hashtag, Topic)	2491.959	49
(NN: Link, Sentiment, Topic)	2493.337	50
(LR: Hashtag, Topic)	2500.012	51

(GBDT: Link)	2505.68	52
(NN: Hashtag, Sentiment, Topic)	2507.174	53
(RF: Link, Sentiment)	2507.4	54
(RF: Hashtag, Sentiment, Topic)	2512.444	55
(GBDT: Hashtag)	2515.348	56
(RF: Sentiment)	2516.065	57
(LR: Link, Hashtag, Sentiment, Topic)	2518.477	58
(NN: Link, Hashtag, Sentiment)	2519.281	59
(NN: Link, Topic)	2520.377	60
(NN: Link, Sentiment)	2521.445	61
(LR: Link, Sentiment)	2523.517	62
(LR: Hashtag, Sentiment, Topic)	2527.081	63
(NN: Sentiment)	2529.91	64
(LR: Sentiment)	2530.191	65
(RF: Link, Sentiment, Topic)	2530.634	66
(GBDT: Link, Sentiment)	2531.33	67
(RF: Link, Topic)	2532.361	68
(RF: Sentiment, Topic)	2534.793	69
(GBDT: Sentiment)	2539.037	70
(RF: Topic)	2541.542	71
(GBDT: Topic)	2546.979	72
(LR: Link, Topic)	2549.084	73
(LR: Topic)	2559.149	74
(NN: Topic)	2559.277	75
(NN: Sentiment, Topic)	2560.862	76
(GBDT: Hashtag, Sentiment)	2568.684	77
(LR: Link, Sentiment, Topic)	2580.56	78
(GBDT: Link, Sentiment, Topic)	2582.424	79
(LR: Sentiment, Topic)	2590.551	80
(GBDT: Sentiment, Topic)	2593.44	81
(GBDT: Hashtag, Sentiment, Topic)	2610.24	82
(KNN: Link, Topic)	2700.869	83
(KNN: Link, Hashtag, Topic)	2705.087	84
(KNN: Hashtag, Topic)	2798.407	85
(KNN: Topic)	2814.752	86
(KNN: Link, Sentiment, Topic)	2819.767	87
(KNN: Link, Hashtag, Sentiment, Topic)	2876.997	88
(KNN: Hashtag, Sentiment, Topic)	2915.984	89
(KNN: Sentiment, Topic)	2981.814	90

Table 11.12: Disney Model-Variable Combination Rankings for OT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(SVR: Link, Hashtag)	481.7841	1
(SVR: Hashtag)	483.1862	2
(SVR: Link)	491.8238	3
(KNN: Link, Hashtag, Topic)	497.6723	4
(KNN: Hashtag, Topic)	521.7441	5
(KNN: Link, Sentiment)	529.1684	6
(KNN: Sentiment)	538.6384	7
(KNN: Link)	538.6391	8
(KNN: Hashtag, Sentiment)	554.6243	9
(KNN: Link, Hashtag, Sentiment)	571.7397	10
(KNN: Link, Hashtag)	641.8516	11
(GBDT: Link)	683.283	12
(GBDT: Hashtag)	685.7706	13
(GBDT: Link, Hashtag)	689.3944	14
(NN: Link, Hashtag, Sentiment, Topic)	690.3798	15
(NN: Link)	690.7575	16

(NN: Hashtag, Sentiment)	691.3566	17
(NN: Link, Hashtag)	691.3814	18
(LR: Link)	691.525	19
(NN: Hashtag, Sentiment, Topic)	691.8538	20
(KNN: Hashtag)	692.2257	21
(GBDT: Sentiment, Topic)	692.3208	22
(RF: Link)	692.4615	23
(RF: Link, Hashtag)	692.4817	24
(NN: Sentiment, Topic)	692.808	25
(NN: Link, Hashtag, Sentiment)	693.0594	26
(LR: Link, Hashtag)	693.1584	27
(NN: Hashtag)	693.4844	28
(LR: Hashtag)	693.5632	29
(RF: Hashtag)	694.1796	30
(NN: Sentiment)	696.9211	31
(GBDT: Hashtag, Sentiment)	698.5706	32
(RF: Link, Hashtag, Sentiment, Topic)	699.2278	33
(RF: Link, Hashtag, Topic)	700.2359	34
(GBDT: Link, Hashtag, Sentiment, Topic)	701.3035	35
(RF: Hashtag, Topic)	701.7261	36
(NN: Link, Hashtag, Topic)	703.251	37
(RF: Hashtag, Sentiment, Topic)	703.8783	38
(RF: Link, Sentiment)	704.5575	39
(RF: Hashtag, Sentiment)	705.9041	40
(RF: Link, Topic)	706.2663	41
(RF: Link, Sentiment, Topic)	706.4815	42
(RF: Sentiment)	707.3441	43
(RF: Sentiment, Topic)	707.7896	44
(RF: Link, Hashtag, Sentiment)	708.8686	45
(GBDT: Link, Sentiment)	709.0221	46
(GBDT: Link, Hashtag, Sentiment)	710.4303	47
(NN: Hashtag, Topic)	710.9328	48
(GBDT: Link, Sentiment, Topic)	711.2019	49
(SVR: Link, Sentiment)	711.2424	50
(SVR: Link, Hashtag, Sentiment)	711.5079	51
(NN: Link, Topic)	711.7001	52
(SVR: Sentiment)	711.8474	53
(NN: Link, Sentiment)	713.3211	54
(SVR: Hashtag, Sentiment)	713.5131	55
(LR: Link, Hashtag, Sentiment)	713.6454	56
(SVR: Link, Hashtag, Topic)	713.8122	57
(SVR: Link, Sentiment, Topic)	714.2674	58
(LR: Hashtag, Sentiment)	714.2701	59
(LR: Link, Sentiment)	714.2979	60
(SVR: Link, Topic)	714.7089	61
(SVR: Hashtag, Topic)	714.9366	62
(RF: Topic)	714.9779	63
(LR: Sentiment)	715.0212	64
(GBDT: Topic)	715.0412	65
(SVR: Sentiment, Topic)	715.3461	66
(SVR: Link, Hashtag, Sentiment, Topic)	715.8433	67
(SVR: Topic)	716.1329	68
(LR: Link, Hashtag, Topic)	716.5616	69
(SVR: Hashtag, Sentiment, Topic)	716.9298	70
(GBDT: Hashtag, Sentiment, Topic)	717.1173	71
(LR: Hashtag, Topic)	718.3448	72
(GBDT: Hashtag, Topic)	719.6003	73
(LR: Link, Topic)	720.0614	74
(LR: Topic)	722.231	75
(NN: Topic)	722.2344	76
(GBDT: Sentiment)	725.6322	77
(NN: Link, Sentiment, Topic)	729.9584	78
(GBDT: Link, Topic)	734.394	79

(LR: Link, Hashtag, Sentiment, Topic)	739.5457	80
(LR: Hashtag, Sentiment, Topic)	741.5986	81
(LR: Link, Sentiment, Topic)	744.8137	82
(GBDT: Link, Hashtag, Topic)	746.9844	83
(LR: Sentiment, Topic)	747.2524	84
(KNN: Link, Hashtag, Sentiment, Topic)	859.5135	85
(KNN: Hashtag, Sentiment, Topic)	904.377	86
(KNN: Sentiment, Topic)	944.4506	87
(KNN: Link, Sentiment, Topic)	953.805	88
(KNN: Topic)	1110.703	89
(KNN: Link, Topic)	1127.803	90

Table 11.13: Google Model-Variable Combination Rankings for IRT Data - Likes

Model-Variable Combination	**MAE**	**Rank**
(KNN: Link)	0.20132	1
(SVR: Sentiment)	0.20132	2
(SVR: Sentiment, Topic)	0.20132	3
(SVR: Link, Topic)	0.20132	4
(SVR: Topic)	0.20132	5
(SVR: Link, Sentiment)	0.20132	6
(SVR: Hashtag)	0.20132	7
(SVR: Link, Hashtag)	0.201469	8
(SVR: Hashtag, Sentiment)	0.206612	9
(SVR: Link, Hashtag, Sentiment)	0.220111	10
(SVR: Topic, NW)	0.254948	11
(SVR: Link, Hashtag, NW)	0.257281	12
(SVR: Link, Hashtag, Topic)	0.264405	13
(SVR: Hashtag, NW)	0.264426	14
(SVR: Link, Hashtag, Topic, NW)	0.264937	15
(SVR: Hashtag, Sentiment, NW)	0.267832	16
(SVR: NW)	0.271564	17
(SVR: Hashtag, Sentiment, Topic)	0.272349	18
(SVR: Link, Hashtag, Sentiment, Topic)	0.273484	19
(SVR: Hashtag, Topic)	0.278016	20
(SVR: Link, Sentiment, Topic)	0.281113	21
(SVR: Link, Hashtag, Sentiment, NW)	0.282481	22
(SVR: Link, Hashtag, Sentiment, Topic, NW)	0.28506	23
(SVR: Link, Topic, NW)	0.286392	24
(SVR: Hashtag, Sentiment, Topic, NW)	0.29295	25
SVR: Hashtag, Topic, NW)	0.294591	26
(SVR: Sentiment, NW)	0.306215	27
(KNN: Sentiment)	0.310442	28
(KNN: Topic)	0.312964	29
(SVR: Link, NW)	0.319976	30
(SVR: Link, Sentiment, Topic, NW)	0.324644	31
(KNN: Link, Sentiment)	0.328664	32
(KNN: Link, Topic)	0.330572	33
(SVR: Link, Sentiment, NW)	0.331098	34
(KNN: Sentiment, Topic)	0.335258	35
(KNN: Link, Sentiment, Topic)	0.344629	36
(SVR: Sentiment, Topic, NW)	0.349035	37
(KNN: Link, NW)	0.353135	38
(KNN: Link, Sentiment, Topic, NW)	0.362391	39
(KNN: NW)	0.367974	40
(KNN: Link, Topic, NW)	0.376006	41
(KNN: Link, Hashtag, NW)	0.389235	42
(KNN: Link, Hashtag, Sentiment, Topic, NW)	0.412541	43
(KNN: Hashtag, Topic, NW)	0.419142	44

(KNN: Link, Sentiment, NW)	0.42577	45
(KNN: Link, Hashtag, Topic, NW)	0.427393	46
(KNN: Hashtag, Sentiment, Topic, NW)	0.429043	47
(KNN: Hashtag, NW)	0.448845	48
(NN: Link, Hashtag, Sentiment, NW)	0.480139	49
(NN: Link, Hashtag, Sentiment, Topic, NW)	0.485188	50
(NN: Hashtag, Sentiment, Topic, NW)	0.489098	51
(NN: Hashtag, Sentiment, NW)	0.49353	52
(NN: Hashtag, NW)	0.51564	53
(KNN: Link, Hashtag, Sentiment, NW)	0.554178	54
(NN: Hashtag, Topic, NW)	0.605192	55
(KNN: Sentiment, Topic, NW)	0.644375	56
(KNN: Hashtag, Sentiment, NW)	0.691515	57
(NN: Link, Hashtag, NW)	0.726698	58
(SVR: Link)	0.871287	59
(KNN: Sentiment, NW)	3.469413	60
(NN: Link, Hashtag, Topic, NW)	4.097737	61
(RF: Link, NW)	4.447255	62
(KNN: Topic, NW)	4.7019	63
(RF: Link, Sentiment, Topic, NW)	5.154723	64
(KNN: Link, Hashtag, Topic)	5.734018	65
(GBDT: Hashtag)	5.827244	66
(NN: NW)	7.706653	67
(RF: Link, Topic, NW)	7.792741	68
(RF: Topic)	7.921526	69
(RF: Hashtag, NW)	7.959197	70
(NN: Link, Topic, NW)	7.999575	71
(RF: Link, Hashtag, NW)	8.070833	72
(RF: Link, Sentiment, NW)	8.085348	73
(NN: Link, Topic)	8.352855	74
(NN: Link, Hashtag, Sentiment, Topic)	8.378542	75
(NN: Sentiment, Topic)	8.388495	76
(NN: Link, Sentiment, Topic, NW)	8.396675	77
(RF: Link, Topic)	8.427471	78
(NN: Hashtag)	8.430075	79
(NN: Topic, NW)	8.495228	80
(RF: Link)	8.535252	81
(NN: Sentiment)	8.573168	82
(NN: Link)	8.717524	83
(NN: Topic)	8.724229	84
(LR: Link)	8.725293	85
(LR: Topic)	8.7343	86
(RF: Link, Sentiment, Topic)	8.963516	87
(RF: Topic, NW)	9.178276	88
(GBDT: Link)	9.191435	89
(NN: Sentiment, Topic, NW)	9.246333	90
(RF: Link, Sentiment)	9.398978	91
(RF: Sentiment, Topic)	9.471614	92
(RF: Link, Hashtag, Topic, NW)	9.501252	93
(RF: Sentiment)	9.614374	94
(NN: Link, Sentiment, Topic)	9.628129	95
(RF: Link, Hashtag, Sentiment, NW)	9.931569	96
(RF: Sentiment, NW)	9.962187	97
(RF: Hashtag, Sentiment, NW)	10.24669	98
(RF: Sentiment, Topic, NW)	10.5589	99
(GBDT: Sentiment)	10.71994	100
(KNN: Hashtag, Sentiment, Topic)	10.88321	101
(NN: Link, NW)	11.10194	102
(KNN: Hashtag, Topic)	11.16099	103
(LR: Sentiment)	11.20578	104
(RF: Link, Hashtag, Sentiment, Topic, NW)	11.56325	105
(RF: Hashtag, Topic, NW)	11.62927	106
(RF: Link, Hashtag, Sentiment, Topic)	12.05262	107

(RF: Hashtag, Sentiment, Topic, NW)	12.14888	108
(RF: Link, Hashtag, Topic)	12.34265	109
(KNN: Link, Hashtag)	12.38486	110
(RF: Hashtag, Sentiment, Topic)	12.39347	111
(GBDT: NW)	12.40867	112
(RF: Hashtag, Topic)	12.4682	113
(LR: Sentiment, Topic)	12.53577	114
(RF: Link, Hashtag, Sentiment)	13.23767	115
(LR: NW)	13.29106	116
(RF: Link, Hashtag)	13.72492	117
(RF: Hashtag, Sentiment)	13.90883	118
(NN: Link, Sentiment)	13.95604	119
(GBDT: Link, Hashtag)	13.9563	120
(RF: NW)	14.50219	121
(GBDT: Sentiment, Topic)	15.12994	122
(KNN: Hashtag)	15.64102	123
(GBDT: Topic)	15.92961	124
(KNN: Hashtag, Sentiment)	15.97276	125
(RF: Hashtag)	16.32998	126
(NN: Sentiment, NW)	16.69592	127
(KNN: Link, Hashtag, Sentiment)	16.79867	128
(NN: Hashtag, Sentiment)	16.80113	129
(LR: Hashtag)	16.83077	130
(LR: Hashtag, Sentiment)	16.83771	131
(LR: Link, Hashtag)	16.84178	132
(LR: Link, Hashtag, Sentiment)	16.84798	133
(NN: Link, Hashtag, Sentiment)	17.20902	134
(NN: Link, Hashtag)	17.73618	135
(NN: Hashtag, Topic)	17.89919	136
(GBDT: Link, Sentiment, Topic)	18.01074	137
(LR: Link, NW)	18.1699	138
(GBDT: Sentiment, Topic, NW)	18.6762	139
(LR: Link, Topic)	18.78929	140
(LR: Link, Sentiment)	19.18205	141
(LR: Sentiment, NW)	19.22091	142
(LR: Link, Sentiment, Topic)	19.39221	143
(GBDT: Hashtag, Sentiment, Topic)	19.85708	144
(GBDT: Hashtag, Sentiment)	20.09072	145
(GBDT: Link, NW)	20.57969	146
(GBDT: Topic, NW)	20.66736	147
(LR: Hashtag, Topic)	20.73594	148
(LR: Link, Hashtag, Topic)	20.74074	149
(LR: Link, Hashtag, Sentiment, Topic)	20.87772	150
(LR: Hashtag, Sentiment, Topic)	20.8789	151
(GBDT: Link, Sentiment, NW)	21.3965	152
(GBDT: Sentiment, NW)	21.66979	153
(NN: Hashtag, Sentiment, Topic)	21.965	154
(LR: Topic, NW)	22.03053	155
(GBDT: Link, Topic, NW)	22.04459	156
(NN: Link, Hashtag, Topic)	22.26049	157
(GBDT: Hashtag, Topic)	22.64669	158
(GBDT: Link, Sentiment)	22.82305	159
(GBDT: Link, Hashtag, Sentiment, Topic)	23.39228	160
(GBDT: Link, Hashtag, NW)	23.86091	161
(GBDT: Hashtag, Sentiment, NW)	24.36465	162
(LR: Hashtag, NW)	24.53645	163
(LR: Hashtag, Sentiment, NW)	24.65377	164
(GBDT: Link, Hashtag, Topic, NW)	24.6609	165
(LR: Sentiment, Topic, NW)	25.07052	166
(LR: Link, Sentiment, NW)	25.3307	167
(LR: Link, Hashtag, NW)	25.91256	168
(LR: Link, Hashtag, Sentiment, NW)	25.92551	169

(GBDT: Link, Hashtag, Sentiment, Topic, NW)	26.08525	170
(NN: Link, Sentiment, NW)	26.74949	171
(LR: Link, Topic, NW)	28.59013	172
(LR: Hashtag, Topic, NW)	29.05633	173
(LR: Link, Hashtag, Topic, NW)	29.54418	174
(LR: Hashtag, Sentiment, Topic, NW)	29.83088	175
(GBDT: Link, Topic)	29.97829	176
(GBDT: Link, Hashtag, Topic)	30.23055	177
(LR: Link, Hashtag, Sentiment, Topic, NW)	30.31924	178
(LR: Link, Sentiment, Topic, NW)	30.35035	179
(GBDT: Link, Hashtag, Sentiment, NW)	30.74731	180
(GBDT: Link, Hashtag, Sentiment)	32.00093	181
(GBDT: Link, Sentiment, Topic, NW)	33.85662	182
(GBDT: Hashtag, Sentiment, Topic, NW)	35.83488	183
(GBDT: Hashtag, NW)	42.73013	184
(GBDT: Hashtag, Topic, NW)	50.51885	185
(KNN: Link, Hashtag, Sentiment, Topic)	65.75248	186

Table 11.14: Google Model-Variable Combination Rankings for IRT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(KNN: Topic)	0.029703	1
(SVR: Hashtag, Sentiment)	0.029703	2
(SVR: Sentiment)	0.029703	3
(SVR: Link, Hashtag, Sentiment)	0.029703	4
(SVR: Link, Sentiment)	0.029703	5
(SVR: Topic)	0.029703	6
(SVR: Link, Topic)	0.029703	7
(SVR: Sentiment, Topic)	0.029703	8
(SVR: Hashtag, Topic)	0.029847	9
(SVR: Link, Hashtag, Topic)	0.037183	10
(KNN: Sentiment)	0.037264	11
(KNN: NW)	0.041163	12
(KNN: Link, Topic)	0.043074	13
(SVR: Hashtag, Sentiment, Topic)	0.044691	14
(SVR: Link, Hashtag, Topic, NW)	0.048957	15
(KNN: Link)	0.052262	16
(SVR: Topic, NW)	0.054779	17
(SVR: NW)	0.055308	18
(SVR: Link, Topic, NW)	0.055753	19
(SVR: Hashtag, NW)	0.056068	20
SVR: Hashtag, Topic, NW)	0.057394	21
(KNN: Link, Sentiment, Topic)	0.057791	22
(SVR: Link, Hashtag, NW)	0.057952	23
(SVR: Hashtag, Sentiment, NW)	0.058486	24
(KNN: Link, NW)	0.058548	25
(SVR: Link, NW)	0.059923	26
(SVR: Link, Hashtag, Sentiment, Topic)	0.060126	27
(SVR: Link, Hashtag, Sentiment, Topic, NW)	0.061639	28
(SVR: Link, Sentiment, Topic, NW)	0.061784	29
(KNN: Link, Sentiment, Topic, NW)	0.062693	30
(SVR: Sentiment, Topic, NW)	0.063357	31
(SVR: Link, Sentiment, Topic)	0.065165	32
(SVR: Link, Hashtag, Sentiment, NW)	0.065643	33
(SVR: Sentiment, NW)	0.06759	34
(SVR: Hashtag, Sentiment, Topic, NW)	0.067877	35

(KNN: Link, Topic, NW)	0.072211	36
(SVR: Link, Sentiment, NW)	0.077454	37
(KNN: Link, Sentiment, NW)	0.080014	38
(KNN: Link, Sentiment)	0.083623	39
(KNN: Link, Hashtag, NW)	0.08909	40
(KNN: Link, Hashtag, Topic, NW)	0.09571	41
(KNN: Link, Hashtag, Sentiment, Topic, NW)	0.09736	42
(KNN: Hashtag, NW)	0.098597	43
(NN: Link, Hashtag, NW)	0.0997	44
(KNN: Hashtag, Topic, NW)	0.10396	45
(KNN: Hashtag, Sentiment, Topic, NW)	0.10396	46
(KNN: Sentiment, Topic)	0.106958	47
(KNN: Link, Hashtag, Sentiment, NW)	0.125473	48
(KNN: Sentiment, Topic, NW)	0.142951	49
(KNN: Hashtag, Sentiment, NW)	0.164656	50
(NN: Link, Hashtag, Sentiment, Topic, NW)	0.226243	51
(NN: Link, Hashtag, Sentiment, NW)	0.42188	52
(NN: Hashtag, NW)	0.697105	53
(NN: Link, Topic, NW)	0.950029	54
(SVR: Link)	0.976898	55
(SVR: Hashtag)	0.976898	56
(SVR: Link, Hashtag)	0.976898	57
(KNN: Sentiment, NW)	1.037683	58
(RF: Link, Sentiment, Topic, NW)	1.393537	59
(RF: Link, NW)	1.406481	60
(KNN: Topic, NW)	1.431998	61
(KNN: Link, Hashtag, Topic)	1.766545	62
(GBDT: Hashtag)	1.780246	63
(KNN: Hashtag, Sentiment, Topic)	1.876399	64
(NN: Link, NW)	2.334445	65
(RF: Link, Sentiment, NW)	2.382217	66
(NN: NW)	2.396609	67
(NN: Hashtag, Sentiment, Topic, NW)	2.462028	68
(RF: Link, Topic, NW)	2.470588	69
(RF: Topic)	2.474418	70
(NN: Link, Hashtag, Topic)	2.532993	71
(RF: Link, Hashtag, NW)	2.580465	72
(NN: Sentiment, NW)	2.612674	73
(NN: Link, Hashtag, Sentiment, Topic)	2.620162	74
(NN: Sentiment, Topic)	2.628613	75
(NN: Link, Topic)	2.642612	76
(NN: Hashtag, Sentiment)	2.642715	77
(NN: Topic, NW)	2.644476	78
(NN: Link, Sentiment, Topic)	2.649771	79
(NN: Hashtag)	2.651618	80
(NN: Sentiment)	2.667343	81
(RF: Link)	2.676194	82
(LR: Topic)	2.683581	83
(LR: Link)	2.736596	84
(NN: Link)	2.747284	85
(RF: Hashtag, NW)	2.766559	86
(RF: Link, Topic)	2.78824	87
(RF: Topic, NW)	2.828577	88
(NN: Topic)	2.828872	89
(RF: Sentiment, Topic)	2.865475	90
(GBDT: Link)	2.880312	91
(NN: Sentiment, Topic, NW)	2.886458	92
(RF: Link, Sentiment, Topic)	2.900501	93
(RF: Sentiment)	3.012493	94
(RF: Link, Sentiment)	3.015862	95
(RF: Sentiment, Topic, NW)	3.04381	96
(RF: Sentiment, NW)	3.178834	97
(RF: Link, Hashtag, Sentiment, Topic, NW)	3.270875	98

(RF: Hashtag, Sentiment, NW)	3.355057	99
(GBDT: Sentiment)	3.355624	100
(RF: Link, Hashtag, Sentiment, NW)	3.36425	101
(NN: Link, Sentiment, Topic, NW)	3.416369	102
(RF: Link, Hashtag, Topic, NW)	3.438175	103
(LR: Sentiment)	3.491473	104
(KNN: Hashtag, Topic)	3.540483	105
(RF: Hashtag, Topic, NW)	3.657298	106
(RF: Link, Hashtag, Sentiment, Topic)	3.8356	107
(RF: Link, Hashtag, Sentiment)	3.900899	108
(RF: Hashtag, Topic)	3.921931	109
(KNN: Link, Hashtag)	3.925911	110
(GBDT: NW)	3.931142	111
(LR: Sentiment, Topic)	3.939997	112
(RF: Hashtag, Sentiment, Topic, NW)	3.945383	113
(RF: Hashtag, Sentiment, Topic)	3.976773	114
(RF: Link, Hashtag, Topic)	4.046157	115
(LR: NW)	4.223323	116
(RF: Link, Hashtag)	4.327704	117
(GBDT: Link, Hashtag)	4.427889	118
(NN: Link, Sentiment)	4.434937	119
(RF: Hashtag, Sentiment)	4.491105	120
(RF: NW)	4.545411	121
(NN: Link, Sentiment, NW)	4.844413	122
(GBDT: Sentiment, Topic)	4.858918	123
(KNN: Hashtag)	4.907828	124
(KNN: Hashtag, Sentiment)	5.022595	125
(GBDT: Topic)	5.058176	126
(RF: Hashtag)	5.121459	127
(NN: Hashtag, Topic, NW)	5.166558	128
(NN: Hashtag, Sentiment, Topic)	5.2779	129
(LR: Link, Hashtag)	5.27986	130
(LR: Hashtag)	5.28078	131
(LR: Link, Hashtag, Sentiment)	5.283226	132
(LR: Hashtag, Sentiment)	5.284121	133
(KNN: Link, Hashtag, Sentiment)	5.327755	134
(NN: Link, Hashtag)	5.402973	135
(NN: Link, Hashtag, Sentiment)	5.427899	136
(GBDT: Link, Sentiment, Topic)	5.775544	137
(LR: Link, NW)	5.842053	138
(LR: Link, Topic)	5.940885	139
(NN: Hashtag, Topic)	5.949135	140
(NN: Hashtag, Sentiment, NW)	6.083146	141
(LR: Sentiment, NW)	6.128916	142
(LR: Link, Sentiment)	6.159675	143
(LR: Link, Sentiment, Topic)	6.180045	144
(GBDT: Hashtag, Sentiment)	6.338763	145
(GBDT: Hashtag, Sentiment, Topic)	6.344998	146
(GBDT: Topic, NW)	6.513916	147
(LR: Hashtag, Topic)	6.541695	148
(LR: Link, Hashtag, Topic)	6.552477	149
(LR: Hashtag, Sentiment, Topic)	6.58722	150
(LR: Link, Hashtag, Sentiment, Topic)	6.588713	151
(GBDT: Link, NW)	6.621023	152
(GBDT: Link, Sentiment, NW)	6.83835	153
(GBDT: Sentiment, NW)	6.967818	154
(LR: Topic, NW)	6.989266	155
(GBDT: Link, Topic, NW)	7.023495	156
(GBDT: Hashtag, Topic)	7.298561	157
(GBDT: Link, Sentiment)	7.322402	158
(GBDT: Link, Hashtag, Sentiment, Topic)	7.433245	159
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	7.486384	160

(GBDT: Link, Hashtag, NW)	7.663001	161
(GBDT: Hashtag, Sentiment, NW)	7.746294	162
(LR: Hashtag, NW)	7.840238	163
(LR: Hashtag, Sentiment, NW)	7.875789	164
(GBDT: Link, Hashtag, Topic, NW)	7.889674	165
(GBDT: Sentiment, Topic, NW)	7.900918	166
(LR: Sentiment, Topic, NW)	7.996623	167
(LR: Link, Sentiment, NW)	8.133375	168
(LR: Link, Hashtag, NW)	8.321256	169
(LR: Link, Hashtag, Sentiment, NW)	8.323875	170
(LR: Link, Topic, NW)	9.141958	171
(LR: Hashtag, Topic, NW)	9.275344	172
(GBDT: Link, Hashtag, Topic)	9.394047	173
(LR: Link, Hashtag, Topic, NW)	9.449857	174
(LR: Hashtag, Sentiment, Topic, NW)	9.505353	175
(GBDT: Link, Topic)	9.535416	176
(LR: Link, Hashtag, Sentiment, Topic, NW)	9.685587	177
(LR: Link, Sentiment, Topic, NW)	9.711223	178
(GBDT: Link, Hashtag, Sentiment, NW)	9.787522	179
(GBDT: Link, Hashtag, Sentiment)	10.25064	180
(GBDT: Link, Sentiment, Topic, NW)	10.78635	181
(NN: Link, Hashtag, Topic, NW)	11.05463	182
(GBDT: Hashtag, Sentiment, Topic, NW)	11.47156	183
(GBDT: Hashtag, NW)	13.6903	184
(GBDT: Hashtag, Topic, NW)	15.57718	185
(KNN: Link, Hashtag, Sentiment, Topic)	21.09241	186

Table 11.15: McDonald's Model-Variable Combination Rankings for IRT Data - Likes

Model-Variable Combination	**MAE**	**Rank**
(SVR: Link, Topic, NW)	5.399611	1
(SVR: Link, Hashtag, Topic, NW)	5.399611	2
(SVR: Topic, NW)	5.400429	3
SVR: Hashtag, Topic, NW)	5.400429	4
(SVR: Link, NW)	5.404909	5
(SVR: Link, Hashtag, NW)	5.404909	6
(SVR: Link, Sentiment)	5.408829	7
(SVR: Link, Hashtag, Sentiment)	5.408829	8
(SVR: NW)	5.408829	9
(SVR: Hashtag, NW)	5.408829	10
(SVR: Topic)	5.408829	11
(SVR: Hashtag, Topic)	5.408829	12
(SVR: Sentiment)	5.408829	13
(SVR: Hashtag, Sentiment)	5.408829	14
(SVR: Link, Sentiment, NW)	5.410643	15
(SVR: Link, Hashtag, Sentiment, NW)	5.410643	16
(SVR: Link, Topic)	5.42962	17
(SVR: Link, Hashtag, Topic)	5.42962	18
(SVR: Sentiment, NW)	5.432756	19
(SVR: Hashtag, Sentiment, NW)	5.432756	20
(SVR: Link, Sentiment, Topic, NW)	5.443725	21
(SVR: Link, Hashtag, Sentiment, Topic, NW)	5.443725	22
(SVR: Sentiment, Topic)	5.444133	23
(SVR: Hashtag, Sentiment, Topic)	5.444133	24
(SVR: Link, Sentiment, Topic)	5.451795	25
(SVR: Link, Hashtag, Sentiment, Topic)	5.451795	26
(SVR: Sentiment, Topic, NW)	5.471807	27

(SVR: Hashtag, Sentiment, Topic, NW)	5.471807	28
(SVR: Link)	5.737044	29
(SVR: Link, Hashtag)	5.737044	30
(SVR: Hashtag)	5.737044	31
(KNN: Link, Hashtag, Topic)	6.278857	32
(KNN: NW)	6.332011	33
(KNN: Hashtag)	6.362852	34
(KNN: Hashtag, Sentiment, NW)	6.470172	35
(KNN: Sentiment, NW)	6.49249	36
(KNN: Link, Topic)	6.549362	37
(KNN: Hashtag, NW)	6.567949	38
(KNN: Hashtag, Sentiment, Topic)	6.57461	39
(KNN: Sentiment, Topic)	6.576588	40
(KNN: Link, Sentiment)	6.61318	41
(KNN: Link, Hashtag, Sentiment)	6.613428	42
(KNN: Link, Hashtag, NW)	6.642917	43
(KNN: Link, NW)	6.696959	44
(KNN: Topic)	6.728116	45
(KNN: Sentiment)	6.752554	46
(KNN: Hashtag, Sentiment)	6.809731	47
(KNN: Hashtag, Topic)	6.816007	48
(KNN: Link, Sentiment, NW)	6.858192	49
(KNN: Link, Hashtag)	7.054983	50
(KNN: Link)	7.22232	51
(KNN: Link, Hashtag, Sentiment, Topic)	7.272728	52
(KNN: Link, Sentiment, Topic)	7.284521	53
(NN: Link, Sentiment, Topic, NW)	7.310815	54
(KNN: Link, Sentiment, Topic, NW)	7.474318	55
(NN: Hashtag, Sentiment, Topic, NW)	7.490149	56
(NN: Sentiment, Topic, NW)	7.582843	57
(KNN: Link, Hashtag, Sentiment, Topic, NW)	7.641985	58
(NN: Hashtag, Sentiment, NW)	7.711593	59
(KNN: Link, Hashtag, Topic, NW)	7.768573	60
(NN: Link, Sentiment, NW)	7.794167	61
(KNN: Hashtag, Sentiment, Topic, NW)	7.885457	62
(KNN: Link, Topic, NW)	7.906001	63
(KNN: Topic, NW)	7.954043	64
(NN: Link, Hashtag, Sentiment, NW)	7.964234	65
(NN: Sentiment, NW)	7.976257	66
(KNN: Sentiment, Topic, NW)	8.096887	67
(KNN: Link, Hashtag, Sentiment, NW)	8.097709	68
(RF: Link, Hashtag, Sentiment, Topic, NW)	8.230371	69
(GBDT: Hashtag, NW)	8.252135	70
(NN: Link, NW)	8.262874	71
(RF: Link, Hashtag, Topic, NW)	8.278612	72
(KNN: Hashtag, Topic, NW)	8.28215	73
(NN: Link, Sentiment, Topic)	8.28493	74
(RF: Link, Sentiment, Topic, NW)	8.310308	75
(NN: Link, Hashtag, NW)	8.349832	76
(GBDT: Sentiment)	8.351163	77
(RF: Link, Topic, NW)	8.369632	78
(GBDT: Hashtag, Topic)	8.55684	79
(NN: Link, Hashtag, Topic, NW)	8.563387	80
(RF: Hashtag, Topic, NW)	8.603868	81
(RF: Hashtag, Sentiment, Topic, NW)	8.622529	82
(RF: Topic, NW)	8.62789	83
(RF: Sentiment, Topic, NW)	8.65881	84
(GBDT: Link, Hashtag, Sentiment)	8.733555	85
(NN: Hashtag, Sentiment, Topic)	8.7371	86
(NN: Link, Topic)	8.762304	87
(GBDT: Hashtag, Sentiment)	8.76274	88
(GBDT: Link, Topic)	8.764711	89
(RF: Link, NW)	8.833298	90

(NN: Link, Hashtag, Sentiment, Topic)	8.84692	91
(NN: Link, Topic, NW)	8.849366	92
(GBDT: Link, Hashtag, Topic)	8.86699	93
(RF: Link, Sentiment, Topic)	8.870503	94
(NN: Hashtag, Topic)	8.910764	95
(RF: Link, Sentiment, NW)	8.91401	96
(RF: Link, Topic)	8.929248	97
(NN: Link, Hashtag, Topic)	8.932865	98
(RF: Sentiment, NW)	8.954464	99
(RF: Link, Hashtag, Sentiment, Topic)	8.956657	100
(NN: Sentiment, Topic)	8.982872	101
(RF: Link, Hashtag, Sentiment, NW)	8.982914	102
(RF: Link, Hashtag, NW)	9.001853	103
(RF: Topic)	9.004607	104
(LR: Topic)	9.024367	105
(LR: Hashtag, Topic)	9.024367	106
(NN: Topic)	9.026613	107
(RF: Hashtag, Sentiment, NW)	9.046706	108
(GBDT: Topic)	9.051789	109
(RF: Hashtag, NW)	9.057253	110
(RF: Sentiment, Topic)	9.064559	111
(RF: Link, Hashtag, Topic)	9.065303	112
(NN: Link, Hashtag, Sentiment, Topic, NW)	9.082716	113
(RF: Hashtag, Topic)	9.103845	114
(GBDT: Link, Sentiment)	9.130768	115
(RF: Hashtag, Sentiment, Topic)	9.140682	116
(RF: Link, Sentiment)	9.141907	117
(RF: NW)	9.14304	118
(GBDT: Link, Hashtag)	9.207895	119
(NN: Link, Hashtag, Sentiment)	9.210874	120
(NN: Link, Sentiment)	9.223262	121
(RF: Link, Hashtag)	9.225573	122
(RF: Link)	9.254395	123
(NN: Link, Hashtag)	9.265517	124
(NN: Link)	9.266949	125
(LR: Link)	9.271419	126
(LR: Link, Hashtag)	9.271419	127
(RF: Link, Hashtag, Sentiment)	9.280012	128
(RF: Hashtag)	9.297117	129
(RF: Sentiment)	9.310741	130
(RF: Hashtag, Sentiment)	9.33706	131
(NN: Hashtag)	9.352271	132
(NN: Topic, NW)	9.354412	133
(LR: Hashtag)	9.355495	134
(NN: Hashtag, Sentiment)	9.388556	135
(NN: Sentiment)	9.404356	136
(LR: Sentiment)	9.404745	137
(LR: Hashtag, Sentiment)	9.404745	138
(GBDT: NW)	9.474007	139
(GBDT: Link)	9.593109	140
(LR: Link, Sentiment)	9.621732	141
(LR: Link, Hashtag, Sentiment)	9.621732	142
(GBDT: Hashtag)	9.693086	143
(GBDT: Hashtag, Sentiment, NW)	9.806564	144
(NN: Hashtag, Topic, NW)	9.82454	145
(GBDT: Sentiment, Topic)	9.953185	146
(LR: Link, Topic)	9.960112	147
(LR: Link, Hashtag, Topic)	9.960112	148
(LR: NW)	9.981915	149
(LR: Hashtag, NW)	9.981915	150
(NN: NW)	10.0245	151
(GBDT: Sentiment, NW)	10.06443	152
(NN: Hashtag, NW)	10.06696	153

(GBDT: Hashtag, Sentiment, Topic)	10.12581	154
(GBDT: Link, NW)	10.26407	155
(LR: Sentiment, NW)	10.32044	156
(LR: Hashtag, Sentiment, NW)	10.32044	157
(GBDT: Link, Hashtag, Sentiment, Topic)	10.45947	158
(GBDT: Link, Sentiment, Topic)	10.46303	159
(LR: Topic, NW)	10.52167	160
(LR: Hashtag, Topic, NW)	10.52167	161
(LR: Sentiment, Topic)	10.57848	162
(LR: Hashtag, Sentiment, Topic)	10.57848	163
(GBDT: Hashtag, Sentiment, Topic, NW)	10.61087	164
(LR: Link, NW)	10.6215	165
(LR: Link, Hashtag, NW)	10.6215	166
(GBDT: Sentiment, Topic, NW)	10.6336	167
(LR: Link, Sentiment, NW)	10.63394	168
(LR: Link, Hashtag, Sentiment, NW)	10.63394	169
(GBDT: Link, Hashtag, Sentiment, NW)	10.67768	170
(GBDT: Link, Sentiment, Topic, NW)	10.86193	171
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	10.87922	172
(GBDT: Link, Sentiment, NW)	10.93876	173
(LR: Link, Sentiment, Topic)	10.96063	174
(LR: Link, Hashtag, Sentiment, Topic)	10.96063	175
(LR: Sentiment, Topic, NW)	11.01393	176
(LR: Hashtag, Sentiment, Topic, NW)	11.01393	177
(GBDT: Link, Hashtag, NW)	11.06621	178
(LR: Link, Topic, NW)	11.37794	179
(LR: Link, Hashtag, Topic, NW)	11.37794	180
(GBDT: Topic, NW)	11.38591	181
(GBDT: Link, Hashtag, Topic, NW)	11.40652	182
(LR: Link, Sentiment, Topic, NW)	11.50487	183
(LR: Link, Hashtag, Sentiment, Topic, NW)	11.50487	184
(GBDT: Hashtag, Topic, NW)	12.28553	185
(GBDT: Link, Topic, NW)	12.77532	186

Table 11.16: McDonald's Model-Variable Combination Rankings for IRT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(SVR: Link, Sentiment, NW)	0.109405	1
(SVR: Link, Hashtag, Sentiment, NW)	0.109405	2
(SVR: Sentiment, NW)	0.109405	3
(SVR: Hashtag, Sentiment, NW)	0.109405	4
(SVR: Link, NW)	0.109405	5
(SVR: Link, Hashtag, NW)	0.109405	6
(SVR: Link, Topic)	0.109405	7
(SVR: Link, Hashtag, Topic)	0.109405	8
(SVR: Sentiment, Topic)	0.109405	9
(SVR: Hashtag, Sentiment, Topic)	0.109405	10
(SVR: NW)	0.109405	11
(SVR: Hashtag, NW)	0.109405	12
(SVR: Topic)	0.109405	13
(SVR: Hashtag, Topic)	0.109405	14
(SVR: Link, Sentiment)	0.109405	15
(SVR: Link, Hashtag, Sentiment)	0.109405	16
(SVR: Topic, NW)	0.109405	17
SVR: Hashtag, Topic, NW)	0.109405	18
(SVR: Sentiment)	0.109405	19

(SVR: Hashtag, Sentiment)	0.109405	20
(SVR: Sentiment, Topic, NW)	0.109515	21
(SVR: Hashtag, Sentiment, Topic, NW)	0.109515	22
(SVR: Link, Topic, NW)	0.109632	23
(SVR: Link, Hashtag, Topic, NW)	0.109632	24
(SVR: Link, Sentiment, Topic)	0.113412	25
(SVR: Link, Hashtag, Sentiment, Topic)	0.113412	26
(SVR: Link, Sentiment, Topic, NW)	0.116629	27
(SVR: Link, Hashtag, Sentiment, Topic, NW)	0.116629	28
(KNN: Link, Hashtag, Topic)	0.132683	29
(KNN: Link, Topic)	0.133226	30
(NN: Link, Sentiment, NW)	0.172589	31
(KNN: Topic)	0.173423	32
(NN: Link, NW)	0.17344	33
(NN: Link, Hashtag, Topic, NW)	0.175168	34
(KNN: Link, Hashtag)	0.175213	35
(KNN: Link, Hashtag, Sentiment, Topic, NW)	0.177176	36
(KNN: Link, Sentiment, Topic, NW)	0.177196	37
(NN: Sentiment, Topic, NW)	0.181609	38
(KNN: Link, Hashtag, NW)	0.182006	39
(RF: Link, Hashtag, Topic, NW)	0.182911	40
(NN: Hashtag, Sentiment, NW)	0.182969	41
(RF: Link, Topic, NW)	0.183501	42
(NN: Hashtag, Topic, NW)	0.183508	43
(KNN: Link, Hashtag, Topic, NW)	0.183654	44
(KNN: Link, Topic, NW)	0.183722	45
(KNN: NW)	0.183899	46
(NN: Link, Sentiment, Topic)	0.184303	47
(NN: Link, Hashtag, NW)	0.184971	48
(NN: Link, Hashtag, Sentiment, NW)	0.187677	49
(KNN: Hashtag)	0.188232	50
(NN: Sentiment, NW)	0.188323	51
(RF: Link, Sentiment, Topic, NW)	0.188451	52
(KNN: Link, Hashtag, Sentiment, Topic)	0.188541	53
(KNN: Link, Sentiment, Topic)	0.188586	54
(NN: Link, Hashtag, Sentiment, Topic, NW)	0.189128	55
(RF: Link, Hashtag, Sentiment, Topic, NW)	0.189131	56
(KNN: Link)	0.189449	57
(KNN: Link, Hashtag, Sentiment, NW)	0.189568	58
(NN: Topic, NW)	0.189969	59
(KNN: Link, Sentiment, NW)	0.189994	60
(NN: Hashtag, Sentiment, Topic)	0.190144	61
(KNN: Topic, NW)	0.190152	62
(KNN: Hashtag, NW)	0.190173	63
(NN: Hashtag, NW)	0.190252	64
(RF: Link, Sentiment, Topic)	0.190269	65
(RF: Hashtag, Topic, NW)	0.190506	66
(KNN: Sentiment, NW)	0.190908	67
(KNN: Hashtag, Sentiment, NW)	0.190949	68
(NN: NW)	0.191172	69
(RF: Topic, NW)	0.191437	70
(KNN: Link, NW)	0.191956	71
(GBDT: Link, Hashtag)	0.191999	72
(RF: Link, Hashtag, Sentiment, Topic)	0.192115	73
(NN: Link, Topic)	0.192169	74
(GBDT: Link, Sentiment, Topic)	0.192448	75
(KNN: Sentiment, Topic)	0.192523	76
(KNN: Hashtag, Sentiment, Topic)	0.192656	77
(GBDT: Link)	0.192705	78
(NN: Link, Topic, NW)	0.192766	79
(GBDT: Link, Hashtag, Sentiment, Topic)	0.19282	80
(RF: Link, NW)	0.192865	81
(RF: Link, Topic)	0.193031	82

(KNN: Hashtag, Topic, NW)	0.193317	83
(NN: Sentiment, Topic)	0.193729	84
(RF: Hashtag, Sentiment, Topic, NW)	0.194411	85
(KNN: Link, Sentiment)	0.194435	86
(NN: Link, Hashtag, Topic)	0.194591	87
(GBDT: Topic)	0.194909	88
(RF: Link, Hashtag, Topic)	0.195171	89
(GBDT: Hashtag, Sentiment)	0.19538	90
(GBDT: Link, Sentiment)	0.195458	91
(GBDT: Sentiment)	0.195647	92
(RF: Link, Sentiment, NW)	0.196183	93
(LR: Topic)	0.196435	94
(LR: Hashtag, Topic)	0.196435	95
(NN: Topic)	0.196458	96
(RF: Sentiment, Topic, NW)	0.196567	97
(NN: Hashtag, Topic)	0.196583	98
(GBDT: Hashtag, NW)	0.19662	99
(NN: Link, Hashtag, Sentiment, Topic)	0.196689	100
(NN: Link, Sentiment)	0.196972	101
(RF: Topic)	0.197109	102
(GBDT: Hashtag, Topic)	0.197343	103
(NN: Link)	0.197433	104
(RF: Link, Hashtag, NW)	0.197446	105
(NN: Link, Hashtag)	0.197454	106
(NN: Link, Hashtag, Sentiment)	0.197479	107
(LR: Link)	0.197498	108
(LR: Link, Hashtag)	0.197498	109
(LR: Link, Sentiment)	0.197524	110
(LR: Link, Hashtag, Sentiment)	0.197524	111
(RF: Link, Sentiment)	0.19768	112
(RF: Link)	0.197807	113
(RF: Link, Hashtag, Sentiment, NW)	0.198107	114
(GBDT: Hashtag, Sentiment, NW)	0.198193	115
(RF: Hashtag, Topic)	0.198256	116
(RF: Sentiment, Topic)	0.19827	117
(KNN: Sentiment, Topic, NW)	0.198299	118
(KNN: Hashtag, Sentiment, Topic, NW)	0.198465	119
(RF: Sentiment, NW)	0.198641	120
(GBDT: Link, Hashtag, Topic)	0.198674	121
(KNN: Link, Hashtag, Sentiment)	0.198681	122
(RF: Hashtag, Sentiment, Topic)	0.199253	123
(KNN: Hashtag, Topic)	0.199318	124
(NN: Hashtag, Sentiment, Topic, NW)	0.199398	125
(RF: Link, Hashtag)	0.199406	126
(GBDT: NW)	0.199711	127
(RF: Hashtag, Sentiment, NW)	0.199748	128
(RF: Hashtag, NW)	0.199759	129
(RF: NW)	0.199961	130
(RF: Link, Hashtag, Sentiment)	0.200094	131
(RF: Sentiment)	0.20024	132
(RF: Hashtag, Sentiment)	0.201107	133
(LR: Sentiment)	0.201258	134
(LR: Hashtag, Sentiment)	0.201258	135
(NN: Hashtag, Sentiment)	0.201271	136
(NN: Sentiment)	0.201359	137
(NN: Link, Sentiment, Topic, NW)	0.201797	138
(RF: Hashtag)	0.201825	139
(LR: Hashtag)	0.202048	140
(GBDT: Sentiment, Topic)	0.202063	141
(NN: Hashtag)	0.20208	142
(GBDT: Link, Topic)	0.202836	143
(LR: NW)	0.20353	144
(LR: Hashtag, NW)	0.20353	145

(LR: Sentiment, Topic)	0.203542	146
(LR: Hashtag, Sentiment, Topic)	0.203542	147
(LR: Sentiment, NW)	0.20381	148
(LR: Hashtag, Sentiment, NW)	0.20381	149
(GBDT: Hashtag, Sentiment, Topic)	0.204035	150
(GBDT: Hashtag, Topic, NW)	0.205024	151
(KNN: Hashtag, Sentiment)	0.205123	152
(LR: Topic, NW)	0.206864	153
(LR: Hashtag, Topic, NW)	0.206864	154
(GBDT: Hashtag)	0.207164	155
(GBDT: Hashtag, Sentiment, Topic, NW)	0.208704	156
(LR: Link, Sentiment, NW)	0.208845	157
(LR: Link, Hashtag, Sentiment, NW)	0.208845	158
(LR: Sentiment, Topic, NW)	0.208981	159
(LR: Hashtag, Sentiment, Topic, NW)	0.208981	160
(LR: Link, NW)	0.209034	161
(LR: Link, Hashtag, NW)	0.209034	162
(KNN: Sentiment)	0.209753	163
(GBDT: Sentiment, NW)	0.210598	164
(GBDT: Sentiment, Topic, NW)	0.211419	165
(LR: Link, Topic)	0.213011	166
(LR: Link, Hashtag, Topic)	0.213011	167
(GBDT: Link, Hashtag, Sentiment)	0.213822	168
(GBDT: Link, Topic, NW)	0.213944	169
(GBDT: Link, NW)	0.214183	170
(GBDT: Link, Hashtag, NW)	0.216974	171
(LR: Link, Sentiment, Topic)	0.217155	172
(LR: Link, Hashtag, Sentiment, Topic)	0.217155	173
(GBDT: Link, Sentiment, NW)	0.21875	174
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	0.219223	175
(GBDT: Link, Hashtag, Sentiment, NW)	0.219827	176
(GBDT: Link, Sentiment, Topic, NW)	0.220292	177
(GBDT: Topic, NW)	0.221229	178
(LR: Link, Sentiment, Topic, NW)	0.22187	179
(LR: Link, Hashtag, Sentiment, Topic, NW)	0.22187	180
(GBDT: Link, Hashtag, Topic, NW)	0.222606	181
(LR: Link, Topic, NW)	0.22288	182
(LR: Link, Hashtag, Topic, NW)	0.22288	183
(SVR: Hashtag)	0.284837	184
(SVR: Link)	0.284837	185
(SVR: Link, Hashtag)	0.284837	186

Table 11.17: Mercedes Model-Variable Combination Rankings for OT Data - Likes

Model-Variable Combination	**MAE**	**Rank**
(SVR: Link)	436.505	1
(SVR: Hashtag)	436.5158	2
(SVR: Link, Hashtag)	436.5158	3
(KNN: Hashtag)	475.6317	4
(KNN: Link, Hashtag)	475.6317	5
(NN: Sentiment, Topic)	479.9155	6
(GBDT: Hashtag, Sentiment)	481.9998	7
(RF: Link, Hashtag, Topic)	487.1451	8
(RF: Hashtag, Topic)	487.3167	9
(GBDT: Sentiment, Topic)	488.8486	10
(RF: Topic)	489.6885	11

(KNN: Hashtag, Sentiment, Topic)	490.2166	12
(GBDT: Link, Topic)	490.6853	13
(RF: Link, Topic)	490.8227	14
(NN: Link, Sentiment, Topic)	492.0657	15
(GBDT: Hashtag)	492.7786	16
(GBDT: Link)	492.8032	17
(LR: Sentiment, Topic)	492.8045	18
(LR: Link, Sentiment, Topic)	492.8045	19
(GBDT: Link, Hashtag, Sentiment, Topic)	492.8312	20
(RF: Sentiment, Topic)	492.8669	21
(RF: Hashtag, Sentiment, Topic)	493.0687	22
(GBDT: Hashtag, Topic)	493.2882	23
(RF: Link, Hashtag, Sentiment, Topic)	493.7536	24
(GBDT: Link, Sentiment, Topic)	494.1518	25
(RF: Link, Sentiment, Topic)	495.6899	26
(GBDT: Link, Hashtag, Sentiment)	495.8732	27
(LR: Topic)	496.4434	28
(LR: Link, Topic)	496.4434	29
(NN: Link, Topic)	496.6062	30
(NN: Topic)	496.7732	31
(LR: Hashtag, Sentiment, Topic)	496.9771	32
(LR: Link, Hashtag, Sentiment, Topic)	496.9771	33
(GBDT: Link, Hashtag, Topic)	496.9964	34
(GBDT: Sentiment)	497.1673	35
(SVR: Sentiment)	497.6934	36
(SVR: Link, Sentiment)	497.6934	37
(SVR: Link, Hashtag, Sentiment)	497.7411	38
(SVR: Hashtag, Sentiment)	497.7887	39
(KNN: Link, Hashtag, Sentiment, Topic)	498.6628	40
(NN: Link, Hashtag)	498.6782	41
(RF: Link, Hashtag)	498.8008	42
(LR: Hashtag, Sentiment)	498.8329	43
(LR: Link, Hashtag, Sentiment)	498.8329	44
(NN: Hashtag)	498.9306	45
(LR: Hashtag)	499.0422	46
(LR: Link, Hashtag)	499.0422	47
(RF: Hashtag)	499.7126	48
(LR: Hashtag, Topic)	500.3642	49
(LR: Link, Hashtag, Topic)	500.3642	50
(LR: Sentiment)	500.3674	51
(LR: Link, Sentiment)	500.3674	52
(RF: Link, Hashtag, Sentiment)	500.5576	53
(RF: Hashtag, Sentiment)	500.6339	54
(GBDT: Hashtag, Sentiment, Topic)	500.7565	55
(RF: Sentiment)	501.8432	56
(NN: Sentiment)	501.8808	57
(RF: Link, Sentiment)	502.0575	58
(NN: Link, Sentiment)	502.063	59
(GBDT: Link, Sentiment)	502.104	60
(NN: Hashtag, Sentiment)	502.1739	61
(NN: Link)	502.2304	62
(LR: Link)	502.2501	63
(NN: Link, Hashtag, Sentiment)	502.3508	64
(SVR: Topic)	503.7653	65
(SVR: Link, Topic)	503.7653	66
(SVR: Link, Hashtag, Topic)	503.7981	67
(SVR: Hashtag, Topic)	503.8089	68
(RF: Link)	504.2698	69
(SVR: Hashtag, Sentiment, Topic)	505.5203	70
(SVR: Sentiment, Topic)	505.5224	71
(SVR: Link, Sentiment, Topic)	505.5224	72
(SVR: Link, Hashtag, Sentiment, Topic)	505.5569	73
(GBDT: Link, Hashtag)	506.8553	74

(KNN: Sentiment, Topic)	509.1132	75
(KNN: Link, Sentiment, Topic)	509.1132	76
(KNN: Hashtag, Topic)	509.8023	77
(KNN: Link, Hashtag, Topic)	509.8023	78
(KNN: Hashtag, Sentiment)	515.7372	79
(KNN: Link, Hashtag, Sentiment)	515.7372	80
(GBDT: Topic)	521.3567	81
(KNN: Topic)	523.3492	82
(KNN: Link, Topic)	523.3492	83
(NN: Link, Hashtag, Sentiment, Topic)	543.6922	84
(NN: Link, Hashtag, Topic)	547.5931	85
(NN: Hashtag, Sentiment, Topic)	548.1533	86
(NN: Hashtag, Topic)	557.5782	87
(KNN: Sentiment)	561.995	88
(KNN: Link, Sentiment)	561.995	89
(KNN: Link)	572.224	90

Table 11.18: Mercedes Model-Variable Combination Rankings for OT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(SVR: Hashtag)	51.98735	1
(SVR: Link, Hashtag)	52.01011	2
(SVR: Link)	52.13564	3
(KNN: Sentiment)	59.65347	4
(KNN: Link, Sentiment)	59.65347	5
(KNN: Link)	60.38614	6
(GBDT: Hashtag, Sentiment)	60.91096	7
(KNN: Hashtag)	61.72785	8
(KNN: Link, Hashtag)	61.72785	9
(SVR: Sentiment)	63.33562	10
(SVR: Link, Sentiment)	63.33562	11
(SVR: Hashtag, Sentiment)	63.43484	12
(SVR: Link, Hashtag, Sentiment)	63.46465	13
(NN: Hashtag, Sentiment, Topic)	64.6524	14
(SVR: Link, Hashtag, Sentiment, Topic)	64.78013	15
(SVR: Hashtag, Sentiment, Topic)	64.78022	16
(SVR: Sentiment, Topic)	64.92349	17
(SVR: Link, Sentiment, Topic)	64.92349	18
(KNN: Hashtag, Topic)	65.21632	19
(KNN: Link, Hashtag, Topic)	65.22929	20
(SVR: Topic)	65.24909	21
(SVR: Link, Topic)	65.24909	22
(SVR: Link, Hashtag, Topic)	65.45751	23
(KNN: Hashtag, Sentiment)	65.46733	24
(KNN: Link, Hashtag, Sentiment)	65.46733	25
(SVR: Hashtag, Topic)	65.47058	26
(NN: Hashtag, Topic)	66.15239	27
(RF: Hashtag, Sentiment, Topic)	66.36374	28
(GBDT: Link, Hashtag, Sentiment, Topic)	66.39557	29
(RF: Link, Hashtag, Sentiment, Topic)	66.47158	30
(RF: Sentiment, Topic)	66.74343	31
(NN: Link, Sentiment, Topic)	66.76793	32
(LR: Hashtag, Sentiment)	66.76927	33
(LR: Link, Hashtag, Sentiment)	66.76927	34
(RF: Link, Sentiment, Topic)	67.04333	35
(NN: Hashtag, Sentiment)	67.14394	36
(NN: Link, Sentiment)	67.22492	37

(LR: Sentiment)	67.24469	38
(LR: Link, Sentiment)	67.24469	39
(RF: Link, Hashtag, Topic)	67.27408	40
(RF: Hashtag, Sentiment)	67.37327	41
(RF: Hashtag, Topic)	67.5507	42
(RF: Link, Topic)	67.57309	43
(RF: Link, Hashtag, Sentiment)	67.60131	44
(KNN: Link, Hashtag, Sentiment, Topic)	67.69198	45
(KNN: Hashtag, Sentiment, Topic)	67.72961	46
(RF: Sentiment)	67.75866	47
(RF: Link, Hashtag)	67.83236	48
(GBDT: Hashtag)	67.88016	49
(RF: Link, Sentiment)	68.02259	50
(RF: Topic)	68.16254	51
(LR: Hashtag)	68.19026	52
(LR: Link, Hashtag)	68.19026	53
(NN: Hashtag)	68.23245	54
(RF: Hashtag)	68.30339	55
(NN: Link, Hashtag)	68.37587	56
(NN: Link, Hashtag, Sentiment)	68.44827	57
(GBDT: Sentiment, Topic)	68.6285	58
(NN: Sentiment, Topic)	68.72662	59
(NN: Sentiment)	68.74445	60
(NN: Link, Hashtag, Sentiment, Topic)	68.81308	61
(NN: Link)	68.81717	62
(LR: Link)	68.84208	63
(LR: Hashtag, Sentiment, Topic)	68.85109	64
(LR: Link, Hashtag, Sentiment, Topic)	68.85109	65
(NN: Link, Hashtag, Topic)	68.86185	66
(GBDT: Sentiment)	69.19695	67
(KNN: Sentiment, Topic)	69.22789	68
(KNN: Link, Sentiment, Topic)	69.22813	69
(RF: Link)	69.36745	70
(LR: Sentiment, Topic)	69.38128	71
(LR: Link, Sentiment, Topic)	69.38128	72
(GBDT: Link, Hashtag, Sentiment)	69.45997	73
(NN: Link, Topic)	69.47081	74
(GBDT: Hashtag, Topic)	69.58215	75
(LR: Hashtag, Topic)	69.74335	76
(LR: Link, Hashtag, Topic)	69.74335	77
(GBDT: Link, Hashtag)	69.75002	78
(KNN: Topic)	69.7508	79
(KNN: Link, Topic)	69.76598	80
(NN: Topic)	69.90964	81
(LR: Topic)	70.35399	82
(LR: Link, Topic)	70.35399	83
(GBDT: Link, Topic)	71.0296	84
(GBDT: Topic)	72.51376	85
(GBDT: Link, Hashtag, Topic)	72.53374	86
(GBDT: Link)	72.80279	87
(GBDT: Link, Sentiment)	72.81712	88
(GBDT: Link, Sentiment, Topic)	75.19053	89
(GBDT: Hashtag, Sentiment, Topic)	80.21971	90

Table 11.19: Mercedes Model-Variable Combination Rankings for IRT Data - Likes

Model-Variable Combination	**MAE**	**Rank**

(SVR: Topic)	0.998286	1
(SVR: Link, Topic)	0.998739	2
(SVR: Link, Hashtag, Topic)	0.999534	3
(SVR: Sentiment, Topic)	0.999784	4
(SVR: Topic, NW)	1.000088	5
(SVR: Link, Sentiment, Topic)	1.000252	6
(SVR: Link, Hashtag, Sentiment, Topic)	1.000362	7
(SVR: Hashtag, Topic)	1.000984	8
SVR: Hashtag, Topic, NW)	1.001609	9
(SVR: Hashtag, Sentiment, Topic)	1.00234	10
(SVR: Link, Topic, NW)	1.004653	11
(SVR: Link, Hashtag, Topic, NW)	1.008817	12
(SVR: Sentiment, Topic, NW)	1.013817	13
(SVR: Hashtag, Sentiment, Topic, NW)	1.018347	14
(SVR: Link, Sentiment, Topic, NW)	1.040462	15
(SVR: Link, Hashtag, Sentiment, Topic, NW)	1.042445	16
(SVR: Hashtag, Sentiment, NW)	1.044416	17
(SVR: Sentiment, NW)	1.044729	18
(KNN: Topic)	1.055317	19
(SVR: Link, Sentiment, NW)	1.057752	20
(SVR: Hashtag, NW)	1.065977	21
(SVR: Link, Hashtag, Sentiment, NW)	1.068008	22
(SVR: NW)	1.073059	23
(SVR: Link, Hashtag, NW)	1.079366	24
(KNN: Sentiment, Topic)	1.085989	25
(SVR: Link, NW)	1.095602	26
(SVR: Hashtag, Sentiment)	1.097933	27
(SVR: Link, Hashtag, Sentiment)	1.10766	28
(KNN: Hashtag, Topic)	1.109428	29
(KNN: Link, Sentiment, Topic)	1.116628	30
(SVR: Link, Sentiment)	1.123062	31
(SVR: Sentiment)	1.124565	32
(KNN: Link, Hashtag, Topic)	1.125702	33
(KNN: Hashtag, Sentiment, Topic)	1.137026	34
(KNN: Hashtag)	1.139258	35
(KNN: Hashtag, NW)	1.150774	36
(KNN: Link, Hashtag, NW)	1.152113	37
(KNN: Link, Hashtag, Sentiment, Topic)	1.152288	38
(KNN: Link, Hashtag)	1.152362	39
(KNN: Sentiment)	1.163625	40
(KNN: Link, Topic)	1.164046	41
(KNN: Link, Sentiment)	1.173178	42
(KNN: Link)	1.173504	43
(KNN: Hashtag, Sentiment)	1.18213	44
(KNN: Link, Hashtag, Sentiment)	1.191524	45
(KNN: NW)	1.22312	46
(KNN: Hashtag, Sentiment, NW)	1.252598	47
(KNN: Link, NW)	1.270438	48
(KNN: Sentiment, NW)	1.276838	49
(KNN: Hashtag, Topic, NW)	1.30465	50
(SVR: Link)	1.309038	51
(SVR: Hashtag)	1.309038	52
(SVR: Link, Hashtag)	1.309038	53
(KNN: Link, Sentiment, NW)	1.356656	54
(KNN: Link, Hashtag, Sentiment, NW)	1.398752	55
(GBDT: Link)	1.487559	56
(GBDT: Hashtag)	1.491885	57
(GBDT: Hashtag, Topic)	1.567829	58
(NN: Sentiment, NW)	1.64171	59
(GBDT: Hashtag, Sentiment, Topic)	1.652173	60
(GBDT: NW)	1.659947	61
(NN: Sentiment, Topic, NW)	1.66786	62
(NN: Hashtag, Sentiment, NW)	1.673643	63

(RF: Link, Hashtag)	1.680284	64
(NN: Link)	1.687864	65
(NN: Topic, NW)	1.689061	66
(LR: Link)	1.690465	67
(NN: Link, Sentiment, NW)	1.691682	68
(NN: Link, Hashtag, Sentiment, Topic)	1.693315	69
(NN: Link, Hashtag)	1.693873	70
(LR: Link, Hashtag)	1.694623	71
(RF: Link)	1.69477	72
(NN: Hashtag, Sentiment, Topic, NW)	1.699533	73
(NN: Link, Hashtag, Topic)	1.70056	74
(NN: Hashtag, Topic)	1.700843	75
(NN: Link, Topic, NW)	1.700942	76
(NN: Link, Sentiment, Topic, NW)	1.701049	77
(NN: Link, Hashtag, Sentiment, NW)	1.701486	78
(NN: Link, Topic)	1.701564	79
(NN: Link, Hashtag, NW)	1.701579	80
(NN: Sentiment, Topic)	1.70166	81
(NN: Hashtag, Topic, NW)	1.702005	82
(NN: Link, Hashtag, Sentiment)	1.702882	83
(LR: Hashtag)	1.706122	84
(NN: Hashtag)	1.708705	85
(RF: Sentiment)	1.709461	86
(RF: Hashtag)	1.70981	87
(NN: Hashtag, Sentiment)	1.709917	88
(NN: Hashtag, Sentiment, Topic)	1.713879	89
(NN: Link, Sentiment, Topic)	1.714622	90
(RF: Link, Sentiment)	1.714629	91
(NN: NW)	1.716138	92
(RF: Link, Hashtag, Sentiment)	1.716806	93
(NN: Link, NW)	1.718033	94
(NN: Topic)	1.71997	95
(NN: Hashtag, NW)	1.723663	96
(GBDT: Link, Sentiment, Topic)	1.725523	97
(RF: Hashtag, Sentiment)	1.737349	98
(GBDT: Link, Hashtag, Sentiment, Topic)	1.738795	99
(KNN: Link, Topic, NW)	1.76005	100
(RF: Sentiment, Topic)	1.760527	101
(RF: Hashtag, Sentiment, Topic)	1.763124	102
(RF: Link, Sentiment, NW)	1.764678	103
(RF: Topic)	1.765072	104
(NN: Link, Sentiment)	1.775771	105
(GBDT: Link, Topic)	1.78189	106
(RF: Link, Hashtag, Sentiment, NW)	1.782358	107
(NN: Sentiment)	1.782903	108
(RF: Link, Hashtag, Sentiment, Topic)	1.782963	109
(RF: Link, Hashtag, Topic)	1.788776	110
(GBDT: Link, Hashtag)	1.791246	111
(RF: Hashtag, Sentiment, NW)	1.791887	112
(LR: Sentiment)	1.795315	113
(RF: Link, Topic)	1.79538	114
(RF: Hashtag, Topic)	1.795558	115
(LR: Link, Sentiment)	1.798524	116
(RF: Link, Sentiment, Topic)	1.798823	117
(RF: Link, Hashtag, NW)	1.809755	118
(LR: Link, NW)	1.824611	119
(RF: Link, NW)	1.827485	120
(LR: NW)	1.831865	121
(GBDT: Link, Hashtag, Sentiment, NW)	1.83416	122
(LR: Hashtag, Sentiment)	1.837626	123
(RF: Hashtag, NW)	1.83842	124
(LR: Link, Hashtag, Sentiment)	1.840801	125
(GBDT: Topic)	1.848056	126

(LR: Link, Hashtag, NW)	1.85433	127
(NN: Link, Hashtag, Sentiment, Topic, NW)	1.861213	128
(LR: Hashtag, NW)	1.86305	129
(GBDT: Link, Hashtag, Sentiment)	1.873702	130
(LR: Topic)	1.88489	131
(RF: Sentiment, NW)	1.884993	132
(GBDT: Link, Sentiment)	1.88597	133
(NN: Link, Hashtag, Topic, NW)	1.890369	134
(LR: Hashtag, Topic)	1.891632	135
(LR: Link, Sentiment, NW)	1.897298	136
(LR: Sentiment, NW)	1.902179	137
(GBDT: Sentiment)	1.943542	138
(RF: NW)	1.955343	139
(LR: Link, Topic)	1.964586	140
(LR: Link, Hashtag, Sentiment, NW)	1.964884	141
(LR: Hashtag, Sentiment, NW)	1.96893	142
(GBDT: Link, Sentiment, NW)	1.970481	143
(LR: Link, Hashtag, Topic)	1.970589	144
(RF: Topic, NW)	1.988948	145
(GBDT: Hashtag, Sentiment, NW)	1.989949	146
(GBDT: Hashtag, Sentiment)	1.997615	147
(GBDT: Link, Hashtag, NW)	2.007181	148
(RF: Hashtag, Topic, NW)	2.071975	149
(GBDT: Sentiment, NW)	2.091763	150
(RF: Link, Topic, NW)	2.099564	151
(LR: Topic, NW)	2.121904	152
(LR: Sentiment, Topic)	2.122695	153
(LR: Hashtag, Topic, NW)	2.129321	154
(LR: Hashtag, Sentiment, Topic)	2.135581	155
(LR: Link, Topic, NW)	2.173292	156
(LR: Link, Sentiment, Topic)	2.178594	157
(LR: Link, Hashtag, Topic, NW)	2.181742	158
(LR: Link, Hashtag, Sentiment, Topic)	2.195336	159
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	2.235557	160
(RF: Link, Hashtag, Topic, NW)	2.249826	161
(GBDT: Hashtag, Sentiment, Topic, NW)	2.249846	162
(LR: Sentiment, Topic, NW)	2.279423	163
(GBDT: Link, Sentiment, Topic, NW)	2.282144	164
(RF: Sentiment, Topic, NW)	2.298291	165
(LR: Hashtag, Sentiment, Topic, NW)	2.298377	166
(RF: Link, Hashtag, Sentiment, Topic, NW)	2.322572	167
(LR: Link, Sentiment, Topic, NW)	2.327203	168
(RF: Link, Sentiment, Topic, NW)	2.331819	169
(LR: Link, Hashtag, Sentiment, Topic, NW)	2.344448	170
(RF: Hashtag, Sentiment, Topic, NW)	2.361416	171
(GBDT: Link, Hashtag, Topic, NW)	2.488342	172
(GBDT: Hashtag, NW)	2.501124	173
(GBDT: Topic, NW)	2.509766	174
(GBDT: Link, NW)	2.511486	175
(GBDT: Link, Hashtag, Topic)	2.543099	176
(GBDT: Sentiment, Topic, NW)	2.574331	177
(GBDT: Sentiment, Topic)	2.648403	178
(GBDT: Hashtag, Topic, NW)	2.656219	179
(GBDT: Link, Topic, NW)	2.67316	180
(KNN: Topic, NW)	2.951727	181
(KNN: Link, Hashtag, Topic, NW)	3.127478	182
(KNN: Link, Hashtag, Sentiment, Topic, NW)	3.630167	183
(KNN: Link, Sentiment, Topic, NW)	3.853183	184
(KNN: Hashtag, Sentiment, Topic, NW)	4.841445	185
(KNN: Sentiment, Topic, NW)	4.982894	186

Table 11.20: Mercedes Model-Variable Combination Rankings for IRT Data - Retweets

Model-Variable Combination	MAE	Rank
(SVR: Topic, NW)	0.177843	1
(SVR: Link, Sentiment, Topic)	0.177843	2
(SVR: Sentiment, Topic)	0.177843	3
(SVR: Hashtag, Topic)	0.177843	4
(SVR: Topic)	0.177843	5
(SVR: Sentiment)	0.177843	6
(SVR: Hashtag, Sentiment)	0.177843	7
(SVR: Link, Topic)	0.177843	8
(SVR: Sentiment, NW)	0.177843	9
(SVR: Link, Hashtag, Topic)	0.177843	10
(SVR: Link, Sentiment)	0.177843	11
(SVR: Link, Hashtag, Sentiment)	0.177843	12
(SVR: NW)	0.177843	13
(SVR: Hashtag, NW)	0.177852	14
(SVR: Link, Hashtag, Topic, NW)	0.177902	15
SVR: Hashtag, Topic, NW)	0.178131	16
(SVR: Link, Topic, NW)	0.178154	17
(SVR: Sentiment, Topic, NW)	0.178297	18
(SVR: Link, Hashtag, Sentiment, Topic)	0.17832	19
(SVR: Link, Sentiment, Topic, NW)	0.17837	20
(SVR: Hashtag, Sentiment, Topic, NW)	0.178593	21
(SVR: Hashtag, Sentiment, Topic)	0.178648	22
(SVR: Link, NW)	0.178659	23
(SVR: Link, Hashtag, Sentiment, Topic, NW)	0.178695	24
(SVR: Link, Hashtag, NW)	0.178724	25
(SVR: Link, Sentiment, NW)	0.178782	26
(SVR: Link, Hashtag, Sentiment, NW)	0.178801	27
(SVR: Hashtag, Sentiment, NW)	0.178942	28
(KNN: Hashtag, Sentiment)	0.206626	29
(KNN: Topic, NW)	0.20861	30
(KNN: Hashtag, NW)	0.209049	31
(KNN: Topic)	0.213027	32
(KNN: Sentiment, Topic)	0.214029	33
(KNN: Link, Sentiment, Topic)	0.218526	34
(KNN: Link, Topic, NW)	0.219653	35
(KNN: Hashtag)	0.219943	36
(KNN: Sentiment)	0.220664	37
(KNN: Link, Topic)	0.221632	38
(KNN: Link, Hashtag, Sentiment)	0.221669	39
(KNN: Hashtag, Topic)	0.223083	40
(KNN: Hashtag, Sentiment, Topic)	0.223468	41
(KNN: Link, Hashtag, Topic, NW)	0.223798	42
(KNN: Link, Hashtag, Topic)	0.22705	43
(KNN: Link, Hashtag)	0.227564	44
(KNN: Hashtag, Topic, NW)	0.23152	45
(KNN: Link, Sentiment)	0.232211	46
(KNN: Link, Hashtag, NW)	0.23624	47
(KNN: Link)	0.239074	48
(KNN: Hashtag, Sentiment, NW)	0.240341	49
(KNN: Link, Hashtag, Sentiment, NW)	0.242927	50
(KNN: Link, Sentiment, NW)	0.248764	51

(KNN: Sentiment, NW)	0.256725	52
(KNN: Link, Hashtag, Sentiment, Topic)	0.279125	53
(GBDT: Link, Hashtag, Topic)	0.305802	54
(GBDT: NW)	0.319368	55
(NN: Sentiment, NW)	0.331762	56
(NN: Hashtag, Sentiment, NW)	0.332775	57
(NN: Hashtag, Topic, NW)	0.339367	58
(RF: Link, Hashtag)	0.341905	59
(LR: Link, Hashtag)	0.342591	60
(NN: Link, NW)	0.342789	61
(NN: Link, Hashtag)	0.342794	62
(NN: Link, Hashtag, Sentiment, NW)	0.343178	63
(NN: Link, Sentiment, Topic, NW)	0.343307	64
(LR: Link)	0.343357	65
(NN: Link)	0.343714	66
(RF: Link)	0.34395	67
(NN: Link, Sentiment)	0.344013	68
(NN: Hashtag, Sentiment, Topic)	0.344133	69
(NN: Hashtag)	0.344317	70
(NN: Topic, NW)	0.344368	71
(LR: Hashtag)	0.344598	72
(NN: Link, Sentiment, NW)	0.344677	73
(NN: Link, Hashtag, Sentiment, Topic)	0.345023	74
(NN: Link, Hashtag, Topic)	0.345059	75
(RF: Hashtag)	0.345097	76
(NN: Hashtag, Sentiment, Topic, NW)	0.345134	77
(NN: Link, Hashtag, Sentiment)	0.345269	78
(NN: Hashtag, Topic)	0.345488	79
(RF: Sentiment)	0.345524	80
(NN: Sentiment, Topic)	0.345592	81
(NN: Link, Topic)	0.34583	82
(NN: Hashtag, Sentiment)	0.346372	83
(NN: NW)	0.346581	84
(NN: Link, Hashtag, NW)	0.346656	85
(NN: Link, Sentiment, Topic)	0.346663	86
(RF: Hashtag, Sentiment, NW)	0.347789	87
(NN: Sentiment)	0.348319	88
(NN: Link, Topic, NW)	0.348467	89
(RF: Link, Sentiment, NW)	0.348972	90
(LR: Sentiment)	0.349004	91
(NN: Hashtag, NW)	0.349312	92
(LR: Link, Sentiment)	0.349485	93
(GBDT: Link, Hashtag, Sentiment, Topic)	0.350507	94
(RF: Link, Hashtag, Sentiment)	0.351073	95
(LR: Hashtag, Sentiment)	0.351616	96
(NN: Link, Hashtag, Sentiment, Topic, NW)	0.351847	97
(RF: Link, Sentiment)	0.352088	98
(SVR: Hashtag)	0.352187	99
(SVR: Link, Hashtag)	0.352187	100
(SVR: Link)	0.352187	101
(LR: Link, Hashtag, Sentiment)	0.352203	102
(RF: Hashtag, Sentiment)	0.352884	103
(RF: Hashtag, NW)	0.35369	104
(NN: Link, Hashtag, Topic, NW)	0.353749	105
(RF: Link, Topic)	0.355537	106
(GBDT: Hashtag, Sentiment, Topic)	0.35585	107
(GBDT: Link, Sentiment, Topic)	0.356066	108
(RF: Link, Sentiment, Topic)	0.356358	109
(RF: Link, NW)	0.356644	110
(RF: Link, Hashtag, NW)	0.356689	111
(RF: Hashtag, Topic)	0.357355	112
(RF: Link, Hashtag, Sentiment, NW)	0.358507	113
(GBDT: Link, Sentiment, NW)	0.359039	114

(GBDT: Link, Hashtag, Sentiment, NW)	0.359144	115
(RF: Topic)	0.359228	116
(GBDT: Link, Hashtag)	0.360335	117
(RF: Sentiment, NW)	0.3607	118
(RF: Link, Hashtag, Sentiment, Topic)	0.360771	119
(GBDT: Hashtag, Sentiment, NW)	0.361028	120
(RF: Link, Hashtag, Topic)	0.361786	121
(RF: Hashtag, Sentiment, Topic)	0.362806	122
(LR: Hashtag, Topic)	0.363702	123
(NN: Topic)	0.363978	124
(LR: Topic)	0.364628	125
(RF: NW)	0.364804	126
(RF: Sentiment, Topic)	0.366892	127
(LR: Link, NW)	0.371364	128
(LR: NW)	0.372105	129
(LR: Link, Hashtag, NW)	0.373085	130
(LR: Hashtag, NW)	0.373826	131
(LR: Link, Sentiment, NW)	0.377715	132
(LR: Link, Hashtag, Topic)	0.379131	133
(LR: Sentiment, NW)	0.379547	134
(LR: Link, Topic)	0.379728	135
(GBDT: Link)	0.380112	136
(GBDT: Hashtag, Topic)	0.38017	137
(GBDT: Link, Hashtag, Sentiment)	0.381875	138
(LR: Link, Hashtag, Sentiment, NW)	0.382305	139
(GBDT: Hashtag)	0.383145	140
(NN: Sentiment, Topic, NW)	0.383456	141
(RF: Topic, NW)	0.383757	142
(LR: Hashtag, Sentiment, NW)	0.384138	143
(RF: Link, Topic, NW)	0.385165	144
(GBDT: Topic)	0.387067	145
(KNN: Link, NW)	0.389869	146
(GBDT: Link, Topic)	0.392497	147
(GBDT: Link, Hashtag, NW)	0.392571	148
(RF: Hashtag, Topic, NW)	0.393209	149
(GBDT: Hashtag, Sentiment, Topic, NW)	0.3948	150
(GBDT: Sentiment, Topic)	0.395772	151
(GBDT: Sentiment)	0.401445	152
(GBDT: Link, Sentiment)	0.401633	153
(GBDT: Link, Sentiment, Topic, NW)	0.401927	154
(GBDT: Link, NW)	0.403938	155
(GBDT: Hashtag, NW)	0.405584	156
(RF: Link, Sentiment, Topic, NW)	0.406459	157
(LR: Hashtag, Sentiment, Topic)	0.4068	158
(LR: Sentiment, Topic)	0.407029	159
(RF: Link, Hashtag, Topic, NW)	0.40712	160
(RF: Sentiment, Topic, NW)	0.408805	161
(GBDT: Sentiment, NW)	0.409629	162
(GBDT: Hashtag, Sentiment)	0.410919	163
(LR: Hashtag, Topic, NW)	0.411426	164
(LR: Topic, NW)	0.412013	165
(LR: Link, Hashtag, Sentiment, Topic)	0.416738	166
(LR: Link, Sentiment, Topic)	0.416763	167
(LR: Link, Hashtag, Topic, NW)	0.420176	168
(LR: Link, Topic, NW)	0.420606	169
(GBDT: Hashtag, Topic, NW)	0.421467	170
(RF: Hashtag, Sentiment, Topic, NW)	0.422739	171
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	0.423701	172
(GBDT: Link, Topic, NW)	0.423762	173
(LR: Sentiment, Topic, NW)	0.428948	174
(LR: Hashtag, Sentiment, Topic, NW)	0.428992	175
(RF: Link, Hashtag, Sentiment, Topic, NW)	0.43491	176

(LR: Link, Sentiment, Topic, NW)	0.436323	177
(LR: Link, Hashtag, Sentiment, Topic, NW)	0.436506	178
(KNN: NW)	0.442807	179
(GBDT: Sentiment, Topic, NW)	0.474344	180
(GBDT: Link, Hashtag, Topic, NW)	0.481219	181
(KNN: Link, Sentiment, Topic, NW)	0.511142	182
(KNN: Link, Hashtag, Sentiment, Topic, NW)	0.529983	183
(GBDT: Topic, NW)	0.540993	184
(KNN: Sentiment, Topic, NW)	0.595092	185
(KNN: Hashtag, Sentiment, Topic, NW)	0.865858	186

Table 11.21: Microsoft Model-Variable Combination Rankings for OT Data - Likes

Model-Variable Combination	MAE	Rank
(SVR: Link, Hashtag)	443.6042	1
(SVR: Link)	444.9915	2
(SVR: Hashtag)	446.6983	3
(KNN: Hashtag, Sentiment)	511.8536	4
(KNN: Link, Hashtag, Sentiment)	517.9941	5
(KNN: Link, Sentiment)	535.881	6
(SVR: Link, Hashtag, Sentiment, Topic)	553.9856	7
(SVR: Link, Sentiment, Topic)	562.5686	8
(SVR: Link, Hashtag, Topic)	562.8054	9
(NN: Link, Hashtag, Sentiment, Topic)	565.4859	10
(SVR: Link, Topic)	568.3071	11
(KNN: Link, Sentiment, Topic)	570.1646	12
(KNN: Sentiment)	571.476	13
(GBDT: Link, Sentiment)	574.6701	14
(GBDT: Link, Hashtag, Sentiment, Topic)	580.8838	15
(NN: Link, Sentiment, Topic)	581.5831	16
(KNN: Sentiment, Topic)	582.229	17
(KNN: Topic)	586.3357	18
(KNN: Link, Hashtag, Sentiment, Topic)	588.1274	19
(KNN: Hashtag)	588.7794	20
(NN: Hashtag, Sentiment, Topic)	589.0636	21
(KNN: Link, Hashtag)	592.9683	22
(NN: Link, Hashtag, Sentiment)	594.5176	23
(GBDT: Link, Hashtag, Topic)	598.9934	24
(GBDT: Link)	599.3212	25
(NN: Link, Hashtag, Topic)	600.1964	26
(KNN: Hashtag, Sentiment, Topic)	602.8909	27
(GBDT: Link, Topic)	605.6251	28
(NN: Link, Sentiment)	605.8085	29
(LR: Link, Sentiment, Topic)	605.8716	30
(GBDT: Hashtag, Sentiment, Topic)	607.4113	31
(NN: Link, Topic)	607.7089	32
(LR: Link, Hashtag, Sentiment, Topic)	608.0108	33
(KNN: Link)	612.9199	34
(LR: Link, Hashtag, Sentiment)	615.6528	35
(LR: Link, Sentiment)	616.6491	36
(RF: Link, Hashtag, Sentiment, Topic)	617.2148	37
(GBDT: Link, Sentiment, Topic)	618.5422	38
(LR: Link, Hashtag, Topic)	620.2332	39
(LR: Hashtag, Sentiment, Topic)	620.2625	40
(NN: Hashtag, Sentiment)	621.372	41
(RF: Link, Hashtag, Topic)	622.0221	42
(RF: Link, Hashtag, Sentiment)	624.0687	43

(GBDT: Link, Hashtag, Sentiment)	624.9884	44
(RF: Link, Sentiment, Topic)	625.186	45
(LR: Link, Topic)	625.5721	46
(GBDT: Sentiment, Topic)	626.6329	47
(GBDT: Hashtag, Sentiment)	627.9422	48
(NN: Link, Hashtag)	628.6699	49
(LR: Link, Hashtag)	628.7444	50
(GBDT: Hashtag, Topic)	630.0185	51
(NN: Hashtag, Topic)	630.2379	52
(RF: Link, Hashtag)	631.0904	53
(LR: Sentiment, Topic)	631.4227	54
(LR: Hashtag, Sentiment)	631.8812	55
(RF: Hashtag, Sentiment, Topic)	632.5939	56
(RF: Link, Topic)	634.5083	57
(RF: Link, Sentiment)	634.7298	58
(LR: Hashtag, Topic)	635.2163	59
(NN: Sentiment, Topic)	635.6023	60
(NN: Link)	636.79	61
(LR: Link)	637.0109	62
(RF: Link)	637.7113	63
(GBDT: Link, Hashtag)	638.9781	64
(RF: Hashtag, Topic)	641.6994	65
(RF: Hashtag, Sentiment)	642.865	66
(NN: Hashtag)	643.7096	67
(LR: Hashtag)	643.7137	68
(RF: Hashtag)	643.7161	69
(KNN: Link, Hashtag, Topic)	651.0059	70
(NN: Sentiment)	651.4812	71
(LR: Sentiment)	651.6256	72
(NN: Topic)	653.8034	73
(LR: Topic)	653.8764	74
(RF: Sentiment, Topic)	655.5192	75
(RF: Sentiment)	657.5569	76
(RF: Topic)	659.7973	77
(GBDT: Topic)	669.7415	78
(KNN: Link, Topic)	674.8237	79
(GBDT: Hashtag)	675.8134	80
(GBDT: Sentiment)	677.8737	81
(KNN: Hashtag, Topic)	682.8323	82
(SVR: Hashtag, Sentiment, Topic)	684.465	83
(SVR: Sentiment, Topic)	691.3038	84
(SVR: Hashtag, Sentiment)	693.4149	85
(SVR: Sentiment)	700.6023	86
(SVR: Hashtag, Topic)	700.858	87
(SVR: Link, Hashtag, Sentiment)	703.4252	88
(SVR: Topic)	704.8497	89
(SVR: Link, Sentiment)	705.7299	90

Table 11.22: Microsoft Model-Variable Combination Rankings for OT Data - Retweets

Model-Variable Combination	MAE	Rank
(SVR: Hashtag)	105.0513	1
(SVR: Link, Hashtag)	105.0825	2
(SVR: Link)	105.0831	3
(KNN: Sentiment)	113.0129	4
(KNN: Topic)	113.3154	5
(KNN: Hashtag, Sentiment)	120.6856	6
(KNN: Link, Hashtag)	121.6418	7
(KNN: Hashtag)	122.8159	8

(KNN: Link)	123.042	9
(KNN: Link, Sentiment)	123.8013	10
(KNN: Link, Hashtag, Sentiment)	127.188	11
(NN: Link, Hashtag, Sentiment, Topic)	127.6563	12
(GBDT: Hashtag, Sentiment)	128.5403	13
(GBDT: Link)	128.9212	14
(NN: Hashtag, Sentiment, Topic)	128.9986	15
(GBDT: Hashtag, Sentiment, Topic)	129.5987	16
(KNN: Link, Sentiment, Topic)	129.7684	17
(GBDT: Link, Hashtag, Topic)	130.9133	18
(NN: Link, Hashtag, Sentiment)	130.979	19
(NN: Hashtag, Sentiment)	130.9916	20
(GBDT: Link, Hashtag, Sentiment, Topic)	130.9926	21
(GBDT: Link, Sentiment)	131.2642	22
(LR: Link, Sentiment, Topic)	131.6125	23
(KNN: Link, Hashtag, Topic)	131.6508	24
(LR: Link, Sentiment)	131.8642	25
(LR: Link, Hashtag, Sentiment)	131.9064	26
(RF: Hashtag, Sentiment, Topic)	131.9449	27
(KNN: Link, Hashtag, Sentiment, Topic)	132.0101	28
(NN: Link, Sentiment, Topic)	132.0718	29
(SVR: Hashtag, Sentiment, Topic)	132.1493	30
(GBDT: Link, Hashtag, Sentiment)	132.1748	31
(SVR: Sentiment)	132.1851	32
(GBDT: Sentiment)	132.197	33
(LR: Link, Hashtag, Sentiment, Topic)	132.2011	34
(RF: Link, Hashtag, Sentiment, Topic)	132.2318	35
(NN: Link, Topic)	132.2535	36
(LR: Sentiment, Topic)	132.3314	37
(RF: Link, Hashtag, Topic)	132.3889	38
(SVR: Sentiment, Topic)	132.3892	39
(SVR: Link, Hashtag, Sentiment, Topic)	132.4156	40
(NN: Sentiment, Topic)	132.4958	41
(NN: Link, Sentiment)	132.5217	42
(RF: Link, Hashtag, Sentiment)	132.5647	43
(LR: Hashtag, Sentiment)	132.6263	44
(SVR: Hashtag, Sentiment)	132.7235	45
(LR: Hashtag, Sentiment, Topic)	132.7763	46
(RF: Hashtag, Sentiment)	132.8331	47
(NN: Sentiment)	132.8847	48
(RF: Link, Sentiment, Topic)	132.9096	49
(LR: Sentiment)	132.9409	50
(GBDT: Link, Topic)	132.9424	51
(GBDT: Hashtag)	133.031	52
(NN: Link, Hashtag, Topic)	133.1364	53
(RF: Sentiment)	133.1815	54
(SVR: Topic)	133.1816	55
(SVR: Link, Sentiment, Topic)	133.2031	56
(RF: Link)	133.2118	57
(NN: Link)	133.2849	58
(LR: Link)	133.2976	59
(LR: Link, Topic)	133.4808	60
(NN: Hashtag, Topic)	133.5129	61
(RF: Link, Topic)	133.5312	62
(GBDT: Hashtag, Topic)	133.5861	63
(RF: Link, Sentiment)	133.6111	64
(RF: Link, Hashtag)	133.6744	65
(SVR: Link, Hashtag, Sentiment)	133.733	66
(SVR: Hashtag, Topic)	133.7395	67
(NN: Link, Hashtag)	133.7503	68
(RF: Hashtag, Topic)	133.7689	69
(SVR: Link, Sentiment)	133.7996	70
(SVR: Link, Topic)	133.8075	71

(KNN: Hashtag, Sentiment, Topic)	133.8199	72
(LR: Link, Hashtag)	133.888	73
(LR: Link, Hashtag, Topic)	133.9169	74
(KNN: Sentiment, Topic)	133.9877	75
(SVR: Link, Hashtag, Topic)	134.0441	76
(RF: Sentiment, Topic)	134.1629	77
(GBDT: Link, Hashtag)	134.2565	78
(RF: Hashtag)	134.2788	79
(GBDT: Link, Sentiment, Topic)	134.3334	80
(LR: Hashtag, Topic)	134.4062	81
(LR: Hashtag)	134.4554	82
(NN: Hashtag)	134.4586	83
(NN: Topic)	134.6415	84
(LR: Topic)	134.7533	85
(RF: Topic)	135.2178	86
(GBDT: Topic)	135.2228	87
(KNN: Hashtag, Topic)	138.1357	88
(GBDT: Sentiment, Topic)	138.3481	89
(KNN: Link, Topic)	148.6071	90

Table 11.23: Microsoft Model-Variable Combination Rankings for IRT Data - Likes

Model-Variable Combination	MAE	Rank
(SVR: Link, Hashtag)	20.45458	1
(SVR: Hashtag)	20.46465	2
(SVR: Link)	20.48485	3
(KNN: Sentiment)	21.27273	4
(KNN: Hashtag, Sentiment)	21.40404	5
(KNN: NW)	21.66667	6
(KNN: Hashtag)	21.83838	7
(KNN: Hashtag, NW)	21.85859	8
(KNN: Link, Hashtag)	26.05499	9
(KNN: Link)	27.73999	10
(NN: NW)	46.19089	11
(NN: Hashtag, Sentiment, NW)	46.84685	12
(NN: Link, Hashtag, NW)	48.1205	13
(GBDT: Link, Hashtag, Topic)	48.19197	14
(NN: Topic, NW)	48.39862	15
(GBDT: Sentiment, Topic)	48.80884	16
(NN: Link, Sentiment, NW)	49.32891	17
(GBDT: Link, Sentiment, Topic, NW)	49.71195	18
(GBDT: Hashtag, Sentiment, Topic, NW)	49.71195	18
(KNN: Link, Sentiment, Topic, NW)	49.89776	20
(KNN: Link, Hashtag, Sentiment, Topic, NW)	50.22176	21
(GBDT: Hashtag, Topic, NW)	50.27559	22
(GBDT: Link, Topic, NW)	50.51663	23
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	50.68962	24
(GBDT: Link, Hashtag, Sentiment, Topic)	51.20762	25
(RF: Link, Hashtag, Sentiment, Topic)	51.55444	26
(NN: Hashtag, Sentiment, Topic)	51.60732	27
(GBDT: Link, Sentiment, Topic)	51.69097	28
(GBDT: Hashtag, Sentiment, Topic)	51.69097	28
(RF: Link, Sentiment, Topic)	51.74753	30
(GBDT: Topic, NW)	51.75261	31
(RF: Link, Hashtag, Topic)	51.99668	32
(RF: Hashtag, Sentiment, Topic)	52.18146	33

(RF: Topic)	52.20146	34
(NN: Sentiment, NW)	52.21309	35
(RF: Link)	52.23956	36
(RF: Link, Topic)	52.25821	37
(LR: Link, Hashtag)	52.38039	38
(NN: Link)	52.38634	39
(LR: Link)	52.40966	40
(RF: Link, Hashtag, Sentiment, NW)	52.40998	41
(LR: Link, Hashtag, NW)	52.41626	42
(LR: Link, NW)	52.44569	43
(NN: Link, Hashtag, Sentiment)	52.70268	44
(RF: Hashtag, Topic)	52.71192	45
(LR: Hashtag, NW)	52.72476	46
(NN: Hashtag, NW)	52.7333	47
(NN: Link, NW)	52.74595	48
(NN: Hashtag)	52.74671	49
(LR: NW)	52.76329	50
(RF: Link, Hashtag)	52.77975	51
(LR: Hashtag)	52.78586	52
(NN: Topic)	52.78654	53
(NN: Link, Hashtag)	52.79152	54
(NN: Link, Hashtag, Sentiment, NW)	52.80242	55
(NN: Link, Topic, NW)	52.80809	56
(NN: Hashtag, Topic, NW)	52.80817	57
(NN: Hashtag, Sentiment, Topic, NW)	52.87396	58
(RF: Link, Sentiment)	52.91525	59
(RF: Link, Hashtag, NW)	52.94062	60
(RF: Hashtag)	52.99788	61
(GBDT: Sentiment, Topic, NW)	52.9997	62
(RF: Sentiment, Topic)	53.00133	63
(GBDT: Link, Hashtag, Topic, NW)	53.00162	64
(NN: Sentiment)	53.07916	65
(RF: Hashtag, Sentiment, NW)	53.0827	66
(RF: Link, NW)	53.12108	67
(RF: Sentiment, NW)	53.31757	68
(GBDT: Link, Hashtag, NW)	53.42688	69
(NN: Hashtag, Sentiment)	53.8095	70
(RF: Hashtag, NW)	53.89875	71
(RF: Link, Sentiment, NW)	53.99689	72
(RF: NW)	54.01107	73
(KNN: Link, Topic)	54.08907	74
(RF: Sentiment)	54.24868	75
(KNN: Link, Hashtag, Topic)	54.26976	76
(RF: Link, Hashtag, Sentiment)	54.3759	77
(RF: Hashtag, Sentiment)	54.48035	78
(KNN: Hashtag, Topic)	54.48996	79
(NN: Link, Sentiment)	54.65784	80
(KNN: Topic)	54.68343	81
(LR: Link, Hashtag, Sentiment)	54.70697	82
(LR: Topic)	54.71193	83
(LR: Link, Sentiment)	54.72214	84
(LR: Link, Topic)	54.85749	85
(NN: Hashtag, Topic)	55.0512	86
(GBDT: Link, Hashtag, Sentiment)	55.05492	87
(LR: Topic, NW)	55.12311	88
(LR: Link, Hashtag, Sentiment, NW)	55.13747	89
(SVR: Link, Hashtag, Topic, NW)	55.14194	90
(NN: Link, Hashtag, Topic, NW)	55.14703	91
(LR: Link, Sentiment, NW)	55.15777	92
(LR: Link, Topic, NW)	55.19688	93
(LR: Hashtag, Sentiment)	55.20123	94
(LR: Sentiment)	55.22748	95
(LR: Hashtag, Topic)	55.32239	96

(LR: Link, Hashtag, Topic)	55.42341	97
(LR: Link, Hashtag, Topic, NW)	55.61143	98
(LR: Hashtag, Topic, NW)	55.63038	99
(SVR: Link, Topic, NW)	55.63244	100
(LR: Hashtag, Sentiment, NW)	55.6642	101
(SVR: Link, Hashtag, Sentiment, Topic, NW)	55.67169	102
(LR: Sentiment, NW)	55.69578	103
SVR: Hashtag, Topic, NW)	55.74867	104
(SVR: Link, Sentiment, Topic, NW)	55.91897	105
(GBDT: Link, Sentiment, NW)	55.92649	106
(NN: Link, Hashtag, Topic)	55.92834	107
(NN: Link, Topic)	55.94065	108
(RF: Topic, NW)	56.04884	109
(NN: Sentiment, Topic, NW)	56.07098	110
(SVR: Topic, NW)	56.36985	111
(SVR: Sentiment, Topic, NW)	56.49457	112
(LR: Sentiment, Topic, NW)	56.5361	113
(LR: Link, Sentiment, Topic, NW)	56.66155	114
(RF: Hashtag, Topic, NW)	56.70263	115
(SVR: Link, Sentiment)	56.71953	116
(SVR: Link, Hashtag, Sentiment)	56.71955	117
(SVR: Sentiment)	56.79761	118
(SVR: Hashtag, Sentiment)	56.79761	119
(SVR: Hashtag, Sentiment, NW)	56.83393	120
(SVR: Link, Hashtag, Sentiment, NW)	56.85372	121
(SVR: Sentiment, NW)	56.86785	122
(SVR: NW)	56.86972	123
(SVR: Link, Sentiment, NW)	56.87506	124
(SVR: Hashtag, NW)	56.87508	125
(SVR: Link, Hashtag, NW)	56.88453	126
(SVR: Link, NW)	56.91837	127
(LR: Sentiment, Topic)	56.99791	128
(SVR: Link, Topic)	57.02002	129
(LR: Hashtag, Sentiment, Topic, NW)	57.05231	130
(SVR: Link, Hashtag, Topic)	57.07297	131
(LR: Link, Hashtag, Sentiment, Topic, NW)	57.11028	132
(SVR: Hashtag, Sentiment, Topic, NW)	57.19773	133
(LR: Link, Sentiment, Topic)	57.21221	134
(SVR: Topic)	57.3888	135
(SVR: Hashtag, Topic)	57.45339	136
(SVR: Link, Sentiment, Topic)	57.49065	137
(SVR: Link, Hashtag, Sentiment, Topic)	57.57175	138
(LR: Hashtag, Sentiment, Topic)	57.61445	139
(SVR: Sentiment, Topic)	57.68749	140
(NN: Link, Hashtag, Sentiment, Topic)	57.70072	141
(LR: Link, Hashtag, Sentiment, Topic)	57.77783	142
(KNN: Link, Hashtag, Topic, NW)	57.85112	143
(SVR: Hashtag, Sentiment, Topic)	57.86298	144
(KNN: Hashtag, Sentiment, Topic)	57.87871	145
(GBDT: Link, Hashtag)	57.99292	146
(GBDT: Sentiment, NW)	58.74359	147
(NN: Sentiment, Topic)	59.29661	148
(GBDT: Topic)	59.46903	149
(GBDT: Hashtag, NW)	59.73964	150
(GBDT: Link, NW)	59.88897	151
(KNN: Link, Topic, NW)	60.12416	152
(RF: Link, Hashtag, Topic, NW)	60.12932	153
(GBDT: Hashtag, Topic)	60.339	154
(GBDT: Link, Hashtag, Sentiment, NW)	60.46805	155
(RF: Link, Topic, NW)	60.98181	156
(GBDT: Sentiment)	61.39952	157
(KNN: Sentiment, Topic, NW)	62.21659	158
(GBDT: Link, Sentiment)	63.03622	159

(GBDT: NW)	63.15524	160
(KNN: Link, Sentiment, Topic)	63.69917	161
(NN: Link, Sentiment, Topic)	63.7735	162
(KNN: Topic, NW)	64.25324	163
(KNN: Hashtag, Sentiment, Topic, NW)	64.31704	164
(KNN: Link, Hashtag, Sentiment, Topic)	64.36478	165
(GBDT: Hashtag, Sentiment)	64.47535	166
(GBDT: Link, Topic)	64.85097	167
(KNN: Hashtag, Topic, NW)	65.14702	168
(GBDT: Hashtag)	66.05969	169
(GBDT: Link)	66.14431	170
(RF: Sentiment, Topic, NW)	66.24653	171
(KNN: Hashtag, Sentiment, NW)	66.32326	172
(RF: Link, Sentiment, Topic, NW)	66.36609	173
(RF: Link, Hashtag, Sentiment, Topic, NW)	67.02317	174
(RF: Hashtag, Sentiment, Topic, NW)	67.05125	175
(KNN: Link, NW)	69.2756	176
(KNN: Link, Hashtag, NW)	70.12676	177
(GBDT: Hashtag, Sentiment, NW)	70.65373	178
(KNN: Sentiment, Topic)	73.31593	179
(KNN: Sentiment, NW)	81.89998	180
(KNN: Link, Hashtag, Sentiment)	82.55075	181
(KNN: Link, Sentiment, NW)	82.85357	182
(NN: Link, Hashtag, Sentiment, Topic, NW)	82.87528	183
(NN: Link, Sentiment, Topic, NW)	83.96499	184
(KNN: Link, Hashtag, Sentiment, NW)	84.54244	185
(KNN: Link, Sentiment)	85.99761	186

Table 11.24: Microsoft Model-Variable Combination Rankings for IRT Data - Retweets

Model-Variable Combination	MAE	Rank
(KNN: Link)	0.676768	1
(KNN: Hashtag)	0.686869	2
(KNN: Link, Hashtag)	0.686869	3
(KNN: Sentiment)	0.924804	4
(SVR: Link)	1.070707	5
(SVR: Link, Hashtag)	1.070707	6
(SVR: Hashtag)	1.070707	7
(KNN: Link, Sentiment, NW)	1.318883	8
(KNN: Sentiment, Topic, NW)	1.512207	9
(KNN: Link, Hashtag, Sentiment, Topic, NW)	1.525986	10
(KNN: Hashtag, Sentiment, Topic, NW)	1.544533	11
(KNN: Link, Sentiment, Topic, NW)	1.545802	12
(KNN: Link, Sentiment)	1.636364	13
(KNN: Link, Hashtag, Sentiment)	1.646465	14
(NN: Link, Sentiment, NW)	1.846839	15
(GBDT: Link)	1.865018	16
(GBDT: Link, Hashtag)	1.87685	17
(GBDT: Sentiment, Topic)	1.88198	18
(GBDT: Hashtag)	1.941234	19
(GBDT: Sentiment)	1.941604	20
(KNN: Topic, NW)	1.948901	21
(KNN: Hashtag, Sentiment)	1.961893	22
(KNN: Hashtag, Topic, NW)	1.962385	23
(KNN: Link, Hashtag, Topic, NW)	1.972501	24
(RF: Link, Hashtag, Sentiment)	1.989394	25
(RF: Link, Sentiment)	1.989917	26

(GBDT: Link, Hashtag, Topic, NW)	1.993534	27
(RF: Link, Sentiment, NW)	1.995788	28
(NN: Link, Hashtag, Sentiment)	2.003929	29
(GBDT: Hashtag, Sentiment)	2.006855	30
(GBDT: Sentiment, Topic, NW)	2.007403	31
(KNN: Link, Topic)	2.011223	32
(NN: Link, Sentiment)	2.011716	33
(KNN: Topic)	2.01321	34
(RF: Link, Hashtag, NW)	2.013552	35
(RF: Link, Hashtag, Sentiment, NW)	2.016158	36
(KNN: Hashtag, Topic)	2.017014	37
(NN: Link, Topic, NW)	2.018387	38
(RF: Link, Hashtag)	2.0204	39
(RF: Sentiment)	2.021079	40
(NN: Link, Hashtag, Topic)	2.022077	41
(RF: Hashtag, Sentiment)	2.023325	42
(NN: Hashtag)	2.024509	43
(NN: Link, Hashtag, Sentiment, NW)	2.02528	44
(NN: Link, Hashtag)	2.025548	45
(LR: Hashtag)	2.026731	46
(NN: Link)	2.028005	47
SVR: Hashtag, Topic, NW)	2.028726	48
(NN: Sentiment)	2.030582	49
(RF: Link, Hashtag, Topic)	2.031516	50
(LR: NW)	2.031724	51
(NN: Hashtag, Sentiment)	2.031726	52
(RF: Link, Hashtag, Sentiment, Topic)	2.032715	53
(LR: Hashtag, NW)	2.033849	54
(SVR: Link, Hashtag, Topic, NW)	2.033949	55
(NN: NW)	2.034825	56
(SVR: Link, Topic, NW)	2.035502	57
(SVR: Topic, NW)	2.039262	58
(LR: Sentiment)	2.042949	59
(NN: Hashtag, NW)	2.043105	60
(NN: Link, NW)	2.043321	61
(LR: Hashtag, Sentiment)	2.043772	62
(RF: Hashtag, Sentiment, Topic)	2.047322	63
(GBDT: Hashtag, NW)	2.047899	64
(RF: Sentiment, NW)	2.048177	65
(RF: Hashtag)	2.048397	66
(LR: Hashtag, Sentiment, NW)	2.049295	67
(SVR: Hashtag, Sentiment, Topic, NW)	2.050129	68
(LR: Sentiment, NW)	2.050228	69
(RF: Link, Sentiment, Topic)	2.050988	70
(RF: Link, Topic)	2.054256	71
(LR: Link)	2.054918	72
(RF: Topic)	2.055127	73
(RF: Link)	2.055263	74
(RF: Hashtag, Topic)	2.056055	75
(SVR: Sentiment, Topic, NW)	2.056368	76
(LR: Link, Hashtag)	2.056585	77
(SVR: Link, Hashtag, Sentiment, Topic, NW)	2.056727	78
(SVR: Link, Sentiment, Topic, NW)	2.057733	79
(RF: Sentiment, Topic)	2.058427	80
(RF: Hashtag, Sentiment, NW)	2.060116	81
(RF: Hashtag, NW)	2.062192	82
(GBDT: Sentiment, NW)	2.063976	83
(RF: Sentiment, Topic, NW)	2.064124	84
(LR: Link, NW)	2.064681	85
(LR: Link, Hashtag, NW)	2.06496	86
(RF: Hashtag, Sentiment, Topic, NW)	2.065419	87
(KNN: Hashtag, Sentiment, Topic)	2.066819	88
(GBDT: NW)	2.0671	89

(GBDT: Link, Hashtag, Sentiment)	2.069107	90
(KNN: Link, Topic, NW)	2.069701	91
(GBDT: Hashtag, Sentiment, NW)	2.072149	92
(SVR: Hashtag, Sentiment, NW)	2.07265	93
(SVR: Link, Sentiment, NW)	2.072796	94
(SVR: Sentiment, Topic)	2.073179	95
(SVR: Link, Hashtag, Sentiment, NW)	2.073646	96
(LR: Link, Hashtag, Sentiment)	2.073987	97
(SVR: Sentiment)	2.073992	98
(SVR: Hashtag, Sentiment)	2.073992	99
(SVR: Link, Sentiment)	2.073992	100
(SVR: Link, Hashtag, Sentiment)	2.073992	101
(SVR: Topic)	2.073992	102
(SVR: Hashtag, Topic)	2.073992	103
(KNN: Sentiment, Topic)	2.074036	104
(SVR: Hashtag, Sentiment, Topic)	2.074292	105
(SVR: Sentiment, NW)	2.07456	106
(SVR: Link, Topic)	2.075074	107
(SVR: Link, Hashtag, Topic)	2.075075	108
(LR: Link, Sentiment)	2.075163	109
(SVR: Link, Hashtag, NW)	2.076025	110
(SVR: Link, NW)	2.076215	111
(SVR: Link, Sentiment, Topic)	2.077119	112
(SVR: Link, Hashtag, Sentiment, Topic)	2.07714	113
(SVR: NW)	2.077371	114
(SVR: Hashtag, NW)	2.077646	115
(GBDT: Link, Hashtag, NW)	2.078046	116
(KNN: Link, Hashtag, Topic)	2.081999	117
(LR: Link, Hashtag, Sentiment, NW)	2.085204	118
(LR: Link, Sentiment, NW)	2.085495	119
(GBDT: Link, Sentiment, NW)	2.094027	120
(GBDT: Link, Hashtag, Sentiment, NW)	2.09442	121
(GBDT: Link, NW)	2.095181	122
(RF: Link, NW)	2.106704	123
(GBDT: Link, Sentiment)	2.130123	124
(RF: Topic, NW)	2.143639	125
(NN: Sentiment, NW)	2.147037	126
(RF: NW)	2.147857	127
(GBDT: Hashtag, Topic)	2.150638	128
(GBDT: Topic)	2.166529	129
(NN: Link, Hashtag, NW)	2.174867	130
(RF: Hashtag, Topic, NW)	2.17799	131
(NN: Hashtag, Sentiment, NW)	2.186662	132
(KNN: Link, Hashtag, Sentiment, Topic)	2.189771	133
(GBDT: Hashtag, Sentiment, Topic)	2.231754	134
(GBDT: Hashtag, Sentiment, Topic, NW)	2.232567	135
(GBDT: Topic, NW)	2.234611	136
(KNN: Link, Hashtag, Sentiment, NW)	2.237923	137
(GBDT: Link, Hashtag, Sentiment, Topic)	2.254446	138
(KNN: Link, Sentiment, Topic)	2.259481	139
(KNN: Link, NW)	2.259731	140
(GBDT: Link, Sentiment, Topic, NW)	2.263033	141
(NN: Topic)	2.270776	142
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	2.271433	143
(LR: Sentiment, Topic)	2.280844	144
(NN: Hashtag, Topic)	2.283847	145
(LR: Sentiment, Topic, NW)	2.285722	146
(LR: Topic)	2.292767	147
(LR: Topic, NW)	2.29567	148
(GBDT: Link, Sentiment, Topic)	2.308906	149
(KNN: Hashtag, Sentiment, NW)	2.311373	150
(LR: Hashtag, Sentiment, Topic)	2.313811	151

(LR: Hashtag, Sentiment, Topic, NW)	2.319406	152
(LR: Hashtag, Topic)	2.323679	153
(LR: Hashtag, Topic, NW)	2.328216	154
(NN: Hashtag, Sentiment, Topic)	2.343998	155
(GBDT: Link, Topic, NW)	2.3487	156
(NN: Sentiment, Topic)	2.349831	157
(RF: Link, Hashtag, Topic, NW)	2.351723	158
(RF: Link, Hashtag, Sentiment, Topic, NW)	2.353942	159
(KNN: NW)	2.366083	160
(LR: Link, Sentiment, Topic)	2.38608	161
(LR: Link, Sentiment, Topic, NW)	2.386501	162
(LR: Link, Hashtag, Sentiment, Topic)	2.391154	163
(LR: Link, Hashtag, Sentiment, Topic, NW)	2.392749	164
(LR: Link, Topic)	2.396328	165
(LR: Link, Topic, NW)	2.397377	166
(LR: Link, Hashtag, Topic)	2.400423	167
(LR: Link, Hashtag, Topic, NW)	2.40205	168
(RF: Link, Topic, NW)	2.406282	169
(RF: Link, Sentiment, Topic, NW)	2.419569	170
(KNN: Sentiment, NW)	2.474639	171
(KNN: Hashtag, NW)	2.502948	172
(KNN: Link, Hashtag, NW)	2.553319	173
(GBDT: Link, Topic)	2.572141	174
(NN: Link, Sentiment, Topic)	2.573953	175
(NN: Hashtag, Topic, NW)	2.598032	176
(GBDT: Hashtag, Topic, NW)	2.686082	177
(GBDT: Link, Hashtag, Topic)	2.721732	178
(NN: Link, Topic)	2.780054	179
(NN: Link, Hashtag, Sentiment, Topic)	2.90605	180
(NN: Topic, NW)	3.042881	181
(NN: Sentiment, Topic, NW)	3.439023	182
(NN: Link, Hashtag, Topic, NW)	4.372171	183
(NN: Hashtag, Sentiment, Topic, NW)	4.734276	184
(NN: Link, Sentiment, Topic, NW)	4.930985	185
(NN: Link, Hashtag, Sentiment, Topic, NW)	7.291852	186

Table 11.25: Samsung Model-Variable Combination Rankings for OT Data - Likes

Model-Variable Combination	MAE	Rank
(SVR: Link, Hashtag)	332.5615	1
(SVR: Link)	333.6084	2
(SVR: Hashtag)	336.6154	3
(KNN: Hashtag, Sentiment)	427.5374	4
(KNN: Hashtag)	438.3985	5
(KNN: Link)	445.9554	6
(KNN: Link, Hashtag)	456.8313	7
(KNN: Topic)	483.0702	8
(KNN: Hashtag, Topic)	490.3099	9
(KNN: Sentiment, Topic)	504.524	10
(KNN: Link, Hashtag, Topic)	522.0006	11
(KNN: Link, Hashtag, Sentiment)	523.299	12
(KNN: Link, Hashtag, Sentiment, Topic)	535.8904	13
(KNN: Link, Sentiment, Topic)	539.2602	14
(KNN: Sentiment)	539.3505	15
(KNN: Link, Topic)	544.2381	16
(KNN: Hashtag, Sentiment, Topic)	590.1408	17
(GBDT: Link, Hashtag)	607.5001	18

(GBDT: Link, Hashtag, Sentiment)	622.4108	19
(RF: Link, Hashtag, Sentiment)	625.1503	20
(NN: Link, Hashtag, Sentiment)	626.7154	21
(GBDT: Link, Sentiment)	629.3657	22
(RF: Link, Hashtag)	630.529	23
(RF: Link, Sentiment)	631.4776	24
(GBDT: Hashtag, Topic)	631.4803	25
(RF: Link, Hashtag, Sentiment, Topic)	632.8345	26
(NN: Link, Sentiment)	633.2264	27
(NN: Link, Hashtag)	633.2569	28
(NN: Hashtag, Sentiment, Topic)	634.8262	29
(RF: Hashtag)	635.3176	30
(LR: Link, Hashtag)	635.3971	31
(NN: Link, Hashtag, Sentiment, Topic)	635.6657	32
(NN: Hashtag)	635.6884	33
(LR: Hashtag)	635.7102	34
(RF: Link, Hashtag, Topic)	635.8532	35
(RF: Hashtag, Sentiment)	636.2627	36
(RF: Link, Sentiment, Topic)	636.5505	37
(RF: Link)	637.7066	38
(NN: Link)	637.7556	39
(LR: Link)	637.9215	40
(GBDT: Hashtag)	637.9474	41
(KNN: Link, Sentiment)	638.4682	42
(NN: Hashtag, Sentiment)	638.4829	43
(RF: Hashtag, Topic)	639.2844	44
(GBDT: Link)	639.506	45
(NN: Link, Sentiment, Topic)	639.6662	46
(NN: Link, Hashtag, Topic)	639.9233	47
(RF: Hashtag, Sentiment, Topic)	639.9247	48
(LR: Hashtag, Topic)	639.9779	49
(NN: Hashtag, Topic)	640.2396	50
(LR: Hashtag, Sentiment)	640.7484	51
(RF: Sentiment, Topic)	641.5799	52
(RF: Link, Topic)	641.6294	53
(GBDT: Hashtag, Sentiment)	641.8587	54
(NN: Link, Topic)	642.8807	55
(RF: Sentiment)	643.0468	56
(GBDT: Topic)	643.7316	57
(LR: Sentiment)	644.8575	58
(NN: Sentiment)	644.8821	59
(LR: Topic)	645.4612	60
(NN: Topic)	645.481	61
(RF: Topic)	645.8987	62
(NN: Sentiment, Topic)	646.2614	63
(GBDT: Sentiment, Topic)	647.2201	64
(LR: Sentiment, Topic)	647.5836	65
(GBDT: Link, Topic)	648.1405	66
(LR: Link, Hashtag, Sentiment)	650.7475	67
(LR: Hashtag, Sentiment, Topic)	650.7562	68
(LR: Link, Hashtag, Topic)	650.8914	69
(GBDT: Link, Hashtag, Sentiment, Topic)	657.7048	70
(LR: Link, Sentiment)	657.7461	71
(GBDT: Link, Hashtag, Topic)	658.1228	72
(LR: Link, Topic)	659.0636	73
(LR: Link, Hashtag, Sentiment, Topic)	659.1629	74
(LR: Link, Sentiment, Topic)	661.2242	75
(GBDT: Hashtag, Sentiment, Topic)	666.3677	76
(GBDT: Link, Sentiment, Topic)	671.6614	77
(GBDT: Sentiment)	680.0088	78
(SVR: Link, Topic)	701.8947	79
(SVR: Link, Hashtag, Topic)	702.0653	80
(SVR: Hashtag, Sentiment)	702.0894	81

(SVR: Sentiment, Topic)	702.1018	82
(SVR: Link, Sentiment, Topic)	702.2403	83
(SVR: Link, Hashtag, Sentiment, Topic)	702.3272	84
(SVR: Hashtag, Topic)	702.3611	85
(SVR: Topic)	702.402	86
(SVR: Sentiment)	702.4106	87
(SVR: Hashtag, Sentiment, Topic)	702.4432	88
(SVR: Link, Hashtag, Sentiment)	703.0678	89
(SVR: Link, Sentiment)	703.2044	90

Table 11.26: Samsung Model-Variable Combination Rankings for OT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(SVR: Link, Hashtag)	28.66479	1
(SVR: Link)	28.67308	2
(SVR: Hashtag)	28.83217	3
(KNN: Hashtag)	32.16764	4
(KNN: Hashtag, Sentiment)	32.68422	5
(KNN: Sentiment)	33.07397	6
(KNN: Link)	33.33615	7
(KNN: Link, Hashtag, Sentiment)	33.89189	8
(KNN: Link, Hashtag)	35.54406	9
(KNN: Link, Sentiment)	36.08769	10
(KNN: Hashtag, Topic)	37.09361	11
(KNN: Topic)	37.72451	12
(KNN: Hashtag, Sentiment, Topic)	39.97325	13
(KNN: Link, Hashtag, Sentiment, Topic)	40.36645	14
(KNN: Link, Topic)	42.88415	15
(KNN: Link, Hashtag, Topic)	43.64473	16
(KNN: Link, Sentiment, Topic)	43.695	17
(KNN: Sentiment, Topic)	44.45649	18
(GBDT: Hashtag)	48.54544	19
(GBDT: Link)	49.81295	20
(NN: Link, Hashtag, Sentiment)	51.09793	21
(RF: Link, Hashtag)	51.29142	22
(RF: Link, Hashtag, Sentiment, Topic)	51.40907	23
(GBDT: Link, Sentiment)	51.61112	24
(RF: Link, Hashtag, Sentiment)	51.70267	25
(NN: Link, Sentiment)	51.80556	26
(LR: Link, Hashtag)	51.85717	27
(NN: Link, Hashtag)	51.87597	28
(RF: Hashtag)	51.89184	29
(NN: Hashtag, Sentiment)	52.01897	30
(GBDT: Hashtag, Sentiment)	52.06841	31
(NN: Hashtag)	52.06973	32
(LR: Hashtag)	52.07787	33
(LR: Link, Hashtag, Sentiment)	52.07845	34
(NN: Link, Hashtag, Sentiment, Topic)	52.09914	35
(NN: Link, Sentiment, Topic)	52.1243	36
(RF: Link)	52.15404	37
(LR: Hashtag, Sentiment)	52.17661	38
(RF: Link, Hashtag, Topic)	52.24563	39
(GBDT: Link, Hashtag, Sentiment)	52.26874	40
(RF: Link, Sentiment, Topic)	52.31399	41
(NN: Link)	52.31415	42
(LR: Link)	52.33473	43
(RF: Hashtag, Sentiment)	52.40159	44
(RF: Hashtag, Sentiment, Topic)	52.41385	45
(GBDT: Topic)	52.44707	46
(GBDT: Hashtag, Sentiment, Topic)	52.45461	47
(NN: Hashtag, Topic)	52.46522	48
(RF: Link, Sentiment)	52.48879	49
(RF: Hashtag, Topic)	52.62146	50

(NN: Link, Hashtag, Topic)	52.67055	51
(NN: Link, Topic)	52.67535	52
(NN: Hashtag, Sentiment, Topic)	52.70029	53
(LR: Hashtag, Topic)	52.71987	54
(RF: Link, Topic)	52.7384	55
(NN: Sentiment)	52.75751	56
(RF: Sentiment)	52.76226	57
(LR: Sentiment)	52.76778	58
(NN: Sentiment, Topic)	52.84278	59
(RF: Sentiment, Topic)	52.87618	60
(LR: Hashtag, Sentiment, Topic)	52.92398	61
(LR: Link, Sentiment)	52.934	62
(LR: Topic)	53.0976	63
(NN: Topic)	53.10049	64
(LR: Sentiment, Topic)	53.11986	65
(RF: Topic)	53.13522	66
(LR: Link, Hashtag, Topic)	53.15414	67
(LR: Link, Hashtag, Sentiment, Topic)	53.22585	68
(GBDT: Link, Sentiment, Topic)	53.47428	69
(GBDT: Link, Hashtag)	53.48614	70
(GBDT: Sentiment)	53.64717	71
(LR: Link, Sentiment, Topic)	53.7805	72
(LR: Link, Topic)	53.81986	73
(GBDT: Sentiment, Topic)	54.12089	74
(GBDT: Link, Hashtag, Sentiment, Topic)	54.16169	75
(GBDT: Link, Hashtag, Topic)	54.22062	76
(GBDT: Hashtag, Topic)	56.12954	77
(GBDT: Link, Topic)	56.14077	78
(SVR: Hashtag, Sentiment)	56.60155	79
(SVR: Topic)	56.76048	80
(SVR: Link, Hashtag, Topic)	56.9493	81
(SVR: Sentiment)	56.95334	82
(SVR: Hashtag, Topic)	56.96798	83
(SVR: Sentiment, Topic)	57.00482	84
(SVR: Link, Sentiment)	57.01075	85
(SVR: Link, Sentiment, Topic)	57.01086	86
(SVR: Hashtag, Sentiment, Topic)	57.07777	87
(SVR: Link, Hashtag, Sentiment, Topic)	57.08605	88
(SVR: Link, Hashtag, Sentiment)	57.11918	89
(SVR: Link, Topic)	57.2337	90

Table 11.27: Toyota Model-Variable Combination Rankings for OT Data - Likes

Model-Variable Combination	MAE	Rank
(SVR: Link, Hashtag, Sentiment, Topic)	94.43972	1
(SVR: Link, Hashtag, Topic)	94.78529	2
(SVR: Link, Topic)	95.6029	3
(SVR: Hashtag)	97.2509	4
(SVR: Link, Hashtag, Sentiment)	97.33678	5
(KNN: Hashtag)	97.45213	6
(SVR: Link, Sentiment)	98.27049	7
(KNN: Link)	98.69154	8
(KNN: Link, Hashtag)	99.04115	9
(SVR: Link, Sentiment, Topic)	99.05146	10
(SVR: Hashtag, Topic)	99.17632	11
(SVR: Hashtag, Sentiment, Topic)	99.4221	12
(SVR: Topic)	99.68692	13
(SVR: Sentiment, Topic)	100.58	14
(SVR: Sentiment)	100.6988	15

(SVR: Hashtag, Sentiment)	100.7735	16
(GBDT: Link, Hashtag)	103.9173	17
(KNN: Sentiment)	104.0472	18
(KNN: Link, Sentiment)	104.0472	19
(GBDT: Link)	106.5968	20
(GBDT: Sentiment)	106.9184	21
(RF: Hashtag)	107.0682	22
(NN: Hashtag)	107.163	23
(LR: Hashtag)	107.228	24
(NN: Link, Hashtag)	107.233	25
(LR: Link, Hashtag)	107.2899	26
(RF: Link, Hashtag)	107.596	27
(RF: Link)	107.7541	28
(NN: Link, Sentiment)	107.8598	29
(NN: Link)	107.8993	30
(NN: Link, Topic)	107.9152	31
(RF: Link, Hashtag, Sentiment)	107.9198	32
(NN: Link, Hashtag, Sentiment)	107.9359	33
(NN: Hashtag, Sentiment)	107.9459	34
(LR: Link)	108.0266	35
(NN: Sentiment)	108.0609	36
(RF: Hashtag, Topic)	108.1284	37
(RF: Link, Hashtag, Topic)	108.3097	38
(RF: Hashtag, Sentiment)	108.5973	39
(RF: Link, Sentiment, Topic)	108.7091	40
(RF: Link, Topic)	108.8169	41
(RF: Sentiment, Topic)	108.8934	42
(GBDT: Link, Hashtag, Sentiment, Topic)	108.905	43
(GBDT: Hashtag)	108.9354	44
(RF: Sentiment)	109.0383	45
(RF: Link, Sentiment)	109.094	46
(RF: Link, Hashtag, Sentiment, Topic)	109.2323	47
(RF: Topic)	109.4203	48
(LR: Hashtag, Sentiment)	109.4958	49
(RF: Hashtag, Sentiment, Topic)	109.5219	50
(LR: Link, Hashtag, Sentiment)	109.5732	51
(LR: Sentiment)	109.6321	52
(LR: Link, Sentiment)	109.736	53
(GBDT: Hashtag, Sentiment)	111.7313	54
(GBDT: Link, Sentiment)	112.0805	55
(GBDT: Sentiment, Topic)	112.2173	56
(NN: Link, Hashtag, Sentiment, Topic)	113.7803	57
(GBDT: Link, Hashtag, Sentiment)	114.3043	58
(SVR: Link)	114.7385	59
(SVR: Link, Hashtag)	114.7441	60
(LR: Hashtag, Topic)	115.0348	61
(LR: Link, Hashtag, Topic)	115.0486	62
(NN: Sentiment, Topic)	115.2123	63
(NN: Topic)	115.4148	64
(LR: Topic)	115.623	65
(LR: Link, Topic)	115.6461	66
(NN: Link, Hashtag, Topic)	115.7624	67
(LR: Hashtag, Sentiment, Topic)	115.8348	68
(LR: Link, Hashtag, Sentiment, Topic)	115.8542	69
(NN: Hashtag, Sentiment, Topic)	116.0936	70
(LR: Sentiment, Topic)	116.2768	71
(LR: Link, Sentiment, Topic)	116.3063	72
(KNN: Hashtag, Sentiment)	116.3075	73
(KNN: Link, Hashtag, Sentiment)	116.3151	74
(GBDT: Topic)	116.75	75
(NN: Link, Sentiment, Topic)	116.8282	76
(KNN: Hashtag, Sentiment, Topic)	116.9989	77
(KNN: Link, Hashtag, Sentiment, Topic)	117.0177	78

(KNN: Sentiment, Topic)	117.1092	79
(KNN: Link, Sentiment, Topic)	117.2555	80
(NN: Hashtag, Topic)	117.33	81
(GBDT: Hashtag, Topic)	118.842	82
(GBDT: Link, Topic)	119.5605	83
(GBDT: Hashtag, Sentiment, Topic)	120.6919	84
(GBDT: Link, Sentiment, Topic)	121.0099	85
(KNN: Hashtag, Topic)	122.318	86
(KNN: Link, Hashtag, Topic)	122.3888	87
(KNN: Topic)	122.605	88
(KNN: Link, Topic)	122.6456	89
(GBDT: Link, Hashtag, Topic)	123.4751	90

Table 11.28: Toyota Model-Variable Combination Rankings for OT Data - Retweets

Model-Variable Combination	**MAE**	**Rank**
(SVR: Link, Hashtag, Topic)	18.60316	1
(SVR: Link, Hashtag, Sentiment, Topic)	18.64355	2
(SVR: Link, Topic)	18.67341	3
(SVR: Hashtag)	18.94711	4
(SVR: Link, Hashtag, Sentiment)	18.99087	5
(KNN: Link, Hashtag)	19.02201	6
(SVR: Link, Sentiment)	19.05513	7
(KNN: Link)	19.68231	8
(SVR: Link, Sentiment, Topic)	19.68706	9
(KNN: Hashtag)	19.68999	10
(SVR: Hashtag, Topic)	19.69508	11
(SVR: Topic)	19.74506	12
(SVR: Hashtag, Sentiment, Topic)	19.81625	13
(SVR: Hashtag, Sentiment)	19.85488	14
(SVR: Sentiment, Topic)	19.93345	15
(SVR: Sentiment)	19.97309	16
(GBDT: Link, Hashtag)	21.2504	17
(KNN: Sentiment)	21.34331	18
(KNN: Link, Sentiment)	21.34331	19
(GBDT: Hashtag)	21.50103	20
(RF: Hashtag)	21.51629	21
(NN: Hashtag)	21.5316	22
(LR: Hashtag)	21.55365	23
(NN: Link, Hashtag)	21.55493	24
(RF: Link, Hashtag, Sentiment)	21.56872	25
(LR: Link, Hashtag)	21.56999	26
(RF: Link)	21.57373	27
(GBDT: Sentiment)	21.59809	28
(GBDT: Link)	21.59888	29
(NN: Link, Hashtag, Sentiment, Topic)	21.59968	30
(NN: Sentiment, Topic)	21.60524	31
(NN: Sentiment)	21.60634	32
(NN: Link, Hashtag, Topic)	21.60824	33
(NN: Hashtag, Sentiment)	21.61153	34
(NN: Link, Sentiment)	21.61281	35
(NN: Link, Hashtag, Sentiment)	21.61654	36
(NN: Link)	21.61695	37
(RF: Link, Hashtag)	21.61782	38
(NN: Hashtag, Topic)	21.61923	39
(LR: Link)	21.63986	40
(GBDT: Link, Hashtag, Sentiment)	21.68708	41
(RF: Hashtag, Sentiment)	21.70849	42
(RF: Hashtag, Topic)	21.73586	43

(RF: Link, Hashtag, Topic)	21.76535	44
(RF: Link, Sentiment)	21.77657	45
(RF: Sentiment)	21.78139	46
(RF: Link, Sentiment, Topic)	21.83309	47
(RF: Link, Topic)	21.85837	48
(LR: Sentiment)	21.87971	49
(LR: Hashtag, Sentiment)	21.88121	50
(LR: Link, Hashtag, Sentiment)	21.90037	51
(LR: Link, Sentiment)	21.90457	52
(RF: Sentiment, Topic)	21.93928	53
(RF: Link, Hashtag, Sentiment, Topic)	22.01481	54
(RF: Topic)	22.02354	55
(RF: Hashtag, Sentiment, Topic)	22.06986	56
(GBDT: Link, Hashtag, Sentiment, Topic)	22.09703	57
(GBDT: Hashtag, Topic)	22.42992	58
(GBDT: Link, Topic)	22.45842	59
(LR: Hashtag, Topic)	23.27133	60
(LR: Link, Hashtag, Topic)	23.27898	61
(LR: Topic)	23.33046	62
(NN: Topic)	23.33789	63
(LR: Link, Topic)	23.33945	64
(LR: Hashtag, Sentiment, Topic)	23.38096	65
(LR: Link, Hashtag, Sentiment, Topic)	23.38988	66
(LR: Sentiment, Topic)	23.43822	67
(LR: Link, Sentiment, Topic)	23.44784	68
(NN: Link, Sentiment, Topic)	23.46354	69
(NN: Link, Topic)	23.60721	70
(KNN: Sentiment, Topic)	23.75369	71
(KNN: Link, Sentiment, Topic)	23.75853	72
(NN: Hashtag, Sentiment, Topic)	23.82535	73
(GBDT: Link, Sentiment)	23.87402	74
(GBDT: Hashtag, Sentiment)	23.92603	75
(GBDT: Link, Sentiment, Topic)	24.00221	76
(GBDT: Hashtag, Sentiment, Topic)	24.01191	77
(KNN: Hashtag, Sentiment, Topic)	24.03238	78
(KNN: Link, Hashtag, Sentiment, Topic)	24.03922	79
(SVR: Link, Hashtag)	24.12363	80
(GBDT: Topic)	24.75184	81
(GBDT: Link, Hashtag, Topic)	25.31264	82
(SVR: Link)	25.40619	83
(KNN: Topic)	25.44092	84
(KNN: Link, Topic)	25.44691	85
(KNN: Hashtag, Topic)	25.60981	86
(KNN: Link, Hashtag, Topic)	25.61469	87
(GBDT: Sentiment, Topic)	25.67418	88
(KNN: Link, Hashtag, Sentiment)	27.22638	89
(KNN: Hashtag, Sentiment)	27.22821	90

Table 281.29: Model-Variable Combination Rankings Across All OT Data - Likes

Model-Variable Combination	Average Rank
(SVR: Hashtag)	10.28571
(SVR: Link, Hashtag)	18
(SVR: Link)	18.42857
(KNN: Hashtag)	18.71429
(KNN: Link, Hashtag)	20.42857
(KNN: Link)	28.71429
(RF: Link, Hashtag, Topic)	29.14286
(KNN: Sentiment)	30.14286
(SVR: Link, Hashtag, Sentiment, Topic)	31.28571
(SVR: Link, Sentiment, Topic)	32.57143
(GBDT: Link, Topic)	34.14286
(RF: Link, Hashtag, Sentiment, Topic)	34.14286
(LR: Link, Hashtag)	35.85714
(GBDT: Link)	36.71429
(GBDT: Link, Hashtag)	36.71429
(KNN: Hashtag, Sentiment)	37.28571
(RF: Link, Sentiment, Topic)	38
(RF: Hashtag, Topic)	39.14286
(SVR: Link, Hashtag, Topic)	39.42857
(SVR: Link, Topic)	40.14286
(GBDT: Sentiment, Topic)	40.28571
(NN: Link, Hashtag)	40.42857
(RF: Link, Topic)	40.85714
(RF: Link, Hashtag)	42
(NN: Hashtag)	42.71429
(NN: Link, Topic)	43
(RF: Hashtag, Sentiment, Topic)	43
(GBDT: Link, Hashtag, Sentiment, Topic)	43.14286
(GBDT: Hashtag, Topic)	43.42857
(NN: Link, Sentiment)	44
(SVR: Link, Hashtag, Sentiment)	44
(LR: Hashtag)	44.42857
(GBDT: Link, Sentiment)	44.42857
(GBDT: Link, Hashtag, Sentiment)	44.42857
(RF: Hashtag)	44.57143
(RF: Topic)	44.57143
(LR: Hashtag, Sentiment, Topic)	44.57143
(KNN: Hashtag, Sentiment, Topic)	44.85714
(RF: Link, Hashtag, Sentiment)	45.14286
(SVR: Hashtag, Sentiment)	46.28571
(LR: Hashtag, Topic)	46.28571
(RF: Sentiment, Topic)	46.28571
(KNN: Link, Sentiment)	47
(LR: Sentiment, Topic)	47.14286
(LR: Link, Hashtag, Sentiment, Topic)	47.14286
(RF: Link, Sentiment)	47.28571
(NN: Topic)	47.42857
(LR: Link, Sentiment, Topic)	47.42857
(NN: Link, Hashtag, Sentiment)	47.57143

(KNN: Link, Hashtag, Sentiment)	47.57143
(LR: Link, Hashtag, Topic)	47.57143
(RF: Link)	47.71429
(NN: Link)	47.85714
(LR: Link)	48.42857
(NN: Link, Sentiment, Topic)	48.42857
(NN: Hashtag, Sentiment)	48.71429
(KNN: Sentiment, Topic)	48.71429
(GBDT: Hashtag)	49.42857
(LR: Link, Topic)	49.42857
(LR: Link, Hashtag, Sentiment)	49.42857
(KNN: Link, Sentiment, Topic)	49.42857
(GBDT: Link, Sentiment, Topic)	49.42857
(SVR: Topic)	49.57143
(GBDT: Link, Hashtag, Topic)	49.71429
(NN: Sentiment, Topic)	49.85714
(SVR: Sentiment, Topic)	51
(SVR: Hashtag, Sentiment, Topic)	51
(LR: Topic)	51.28571
(SVR: Link, Sentiment)	51.28571
(RF: Hashtag, Sentiment)	51.28571
(NN: Hashtag, Sentiment, Topic)	51.42857
(KNN: Topic)	51.71429
(GBDT: Hashtag, Sentiment)	53.28571
(LR: Hashtag, Sentiment)	53.57143
(SVR: Hashtag, Topic)	54.71429
(NN: Link, Hashtag, Sentiment, Topic)	55.28571
(SVR: Sentiment)	55.42857
(LR: Link, Sentiment)	55.42857
(KNN: Link, Hashtag, Sentiment, Topic)	55.85714
(GBDT: Topic)	56.14286
(NN: Link, Hashtag, Topic)	56.28571
(GBDT: Hashtag, Sentiment, Topic)	57.14286
(NN: Hashtag, Topic)	60.57143
(NN: Sentiment)	60.71429
(RF: Sentiment)	60.85714
(GBDT: Sentiment)	61.71429
(LR: Sentiment)	62.57143
(KNN: Link, Topic)	62.57143
(KNN: Link, Hashtag, Topic)	63.57143
(KNN: Hashtag, Topic)	66

Table 11.30: Model-Variable Combination Rankings Across All OT Data - Retweets

Model-Variable Combination	**Average Rank**
(SVR: Hashtag)	2.142857
(KNN: Link, Sentiment)	11.85714
(SVR: Link, Hashtag)	12.85714
(SVR: Link)	14
(KNN: Link)	16.28571
(KNN: Sentiment)	19
(KNN: Link, Hashtag)	19.14286
(KNN: Hashtag)	21.14286
(KNN: Hashtag, Sentiment)	28.42857
(NN: Hashtag, Sentiment)	32.57143
(SVR: Sentiment)	33.57143
(SVR: Hashtag, Sentiment)	34.14286
(KNN: Link, Hashtag, Sentiment)	36.71429
(SVR: Link, Hashtag, Sentiment)	37.42857

(GBDT: Hashtag, Sentiment)	37.71429
(NN: Link, Hashtag, Sentiment, Topic)	37.71429
(KNN: Link, Hashtag, Topic)	38.28571
(SVR: Link, Sentiment)	38.71429
(GBDT: Link, Hashtag, Sentiment, Topic)	39
(GBDT: Link)	39.14286
(NN: Link, Hashtag, Sentiment)	39.28571
(KNN: Link, Sentiment, Topic)	39.71429
(RF: Link, Hashtag, Sentiment, Topic)	39.85714
(GBDT: Hashtag)	40.28571
(RF: Link, Hashtag, Sentiment)	40.42857
(KNN: Topic)	40.57143
(NN: Link, Hashtag)	41.14286
(RF: Link, Sentiment, Topic)	41.28571
(KNN: Hashtag, Topic)	41.57143
(NN: Hashtag, Sentiment, Topic)	42.14286
(RF: Hashtag, Sentiment, Topic)	42.57143
(SVR: Link, Hashtag, Sentiment, Topic)	43.14286
(SVR: Hashtag, Sentiment, Topic)	43.42857
(NN: Hashtag, Topic)	43.57143
(RF: Link, Hashtag, Topic)	43.57143
(RF: Link, Hashtag)	43.71429
(NN: Link, Sentiment)	44
(NN: Sentiment, Topic)	44.28571
(LR: Link, Hashtag)	44.42857
(SVR: Sentiment, Topic)	44.71429
(RF: Link)	45
(NN: Hashtag)	45.14286
(SVR: Link, Sentiment, Topic)	45.71429
(NN: Link)	46
(RF: Hashtag)	46.57143
(RF: Hashtag, Sentiment)	46.57143
(SVR: Link, Hashtag, Topic)	46.85714
(LR: Hashtag)	47.14286
(LR: Link)	47.28571
(RF: Sentiment, Topic)	47.71429
(LR: Link, Hashtag, Sentiment)	48
(NN: Sentiment)	48.28571
(SVR: Topic)	48.57143
(NN: Link, Sentiment, Topic)	48.85714
(KNN: Sentiment, Topic)	49.42857
(SVR: Link, Topic)	49.85714
(RF: Hashtag, Topic)	50
(SVR: Hashtag, Topic)	50.28571
(RF: Link, Sentiment)	50.42857
(RF: Link, Topic)	50.42857
(LR: Hashtag, Sentiment)	51.57143
(GBDT: Link, Hashtag)	51.85714
(LR: Hashtag, Sentiment, Topic)	52.71429
(GBDT: Hashtag, Sentiment, Topic)	53.42857
(LR: Link, Sentiment)	53.71429
(LR: Link, Hashtag, Sentiment, Topic)	54.14286
(RF: Sentiment)	54.28571
(GBDT: Link, Hashtag, Sentiment)	54.57143
(GBDT: Sentiment, Topic)	55.14286
(LR: Link, Sentiment, Topic)	55.14286
(LR: Sentiment, Topic)	55.28571
(NN: Link, Topic)	55.57143
(RF: Topic)	55.85714
(KNN: Hashtag, Sentiment, Topic)	56.28571
(NN: Link, Hashtag, Topic)	56.71429
(GBDT: Hashtag, Topic)	56.85714
(LR: Sentiment)	57

(GBDT: Link, Sentiment)	57.14286
(GBDT: Link, Sentiment, Topic)	57.85714
(KNN: Link, Hashtag, Sentiment, Topic)	58.71429
(LR: Hashtag, Topic)	60.42857
(GBDT: Link, Hashtag, Topic)	61.57143
(LR: Link, Hashtag, Topic)	62.57143
(GBDT: Sentiment)	63.14286
(GBDT: Link, Topic)	64.28571
(NN: Topic)	64.42857
(LR: Link, Topic)	64.42857
(LR: Topic)	65.57143
(KNN: Link, Topic)	65.85714
(GBDT: Topic)	67.14286

Table 11.31: Model-Variable Combination Rankings Across All IRT Data - Likes

Model-Variable Combination	**Average Rank**
(SVR: Hashtag)	22.28571
(SVR: Link, Hashtag)	23.57143
(KNN: Sentiment)	27.14286
(KNN: Link)	30
(SVR: Link)	31.14286
(KNN: Topic)	38.85714
(SVR: Link, Hashtag, Topic, NW)	40.28571
(KNN: Link, Topic)	40.57143
(KNN: Link, Hashtag)	40.85714
(KNN: Hashtag)	41.14286
(SVR: Link, Topic, NW)	44.28571
SVR: Hashtag, Topic, NW)	44.85714
(KNN: Hashtag, Sentiment)	45.42857
(KNN: NW)	45.71429
(SVR: Link, Topic)	46
(KNN: Link, Hashtag, Topic)	46
(SVR: Topic, NW)	46.57143
(SVR: Link, Hashtag, Topic)	47
(SVR: Hashtag, Sentiment, NW)	47.57143
(SVR: Topic)	47.71429
(SVR: Link, Sentiment)	48
(SVR: Link, Hashtag, Sentiment)	48
(SVR: Sentiment, Topic)	48.42857
(SVR: Sentiment)	49.14286
(SVR: Hashtag, NW)	49.57143
(SVR: Hashtag, Sentiment)	49.85714
(SVR: Sentiment, NW)	50.14286
(SVR: Link, Hashtag, Sentiment, Topic, NW)	50.28571
(SVR: NW)	50.57143
(SVR: Link, Hashtag, Sentiment, NW)	51
(SVR: Link, Sentiment, Topic, NW)	51.28571
(SVR: Link, Hashtag, NW)	51.42857
(KNN: Hashtag, Topic)	51.71429
(SVR: Link, Hashtag, Sentiment, Topic)	52.28571
(SVR: Link, Sentiment, Topic)	53
(KNN: Link, Sentiment, Topic)	53.57143
(SVR: Hashtag, Topic)	54.14286
(KNN: Sentiment, Topic)	54.14286
(KNN: Link, Sentiment)	54.42857
(SVR: Link, Sentiment, NW)	54.42857
(SVR: Link, NW)	54.71429
(SVR: Sentiment, Topic, NW)	54.85714

(SVR: Hashtag, Sentiment, Topic, NW)	55
(SVR: Hashtag, Sentiment, Topic)	56.57143
(KNN: Hashtag, NW)	56.71429
(KNN: Hashtag, Sentiment, Topic)	60
(NN: Hashtag, Sentiment, NW)	61.71429
(KNN: Link, Hashtag, Sentiment, Topic, NW)	63.42857
(KNN: Link, Hashtag, Sentiment)	68.57143
(KNN: Link, NW)	68.85714
(KNN: Link, Topic, NW)	69.57143
(NN: Link, Hashtag, NW)	71.28571
(KNN: Link, Sentiment, Topic, NW)	71.28571
(RF: Link, Sentiment, Topic)	73.57143
(KNN: Sentiment, NW)	74.14286
(NN: Hashtag, Sentiment, Topic, NW)	74.28571
(NN: Link, Hashtag, Sentiment, NW)	75.42857
(KNN: Hashtag, Sentiment, NW)	75.71429
(NN: Link, Sentiment, NW)	76.85714
(RF: Link, Topic)	77.14286
(RF: Sentiment, Topic)	77.71429
(RF: Topic)	77.85714
(KNN: Hashtag, Topic, NW)	78.71429
(KNN: Link, Sentiment, NW)	78.85714
(NN: Hashtag, Topic, NW)	79
(NN: Link, Topic, NW)	79.42857
(KNN: Link, Hashtag, NW)	81.14286
(NN: Topic)	81.71429
(RF: Link, Sentiment, NW)	81.71429
(RF: Topic, NW)	83
(NN: Link, Hashtag, Sentiment, Topic)	84.14286
(RF: Link, Sentiment)	84.57143
(NN: Link, Topic)	84.57143
(KNN: Link, Hashtag, Topic, NW)	86
(KNN: Topic, NW)	86.85714
(RF: Hashtag, Sentiment, Topic)	86.85714
(RF: Link, Hashtag, Sentiment, Topic)	86.85714
(NN: Link, Sentiment, Topic, NW)	87.42857
(NN: Link, Hashtag, Sentiment)	88.28571
(NN: Sentiment, NW)	88.71429
(KNN: Link, Hashtag, Sentiment, Topic)	89.14286
(NN: Sentiment, Topic)	89.71429
(RF: Link, Hashtag, Topic)	89.85714
(NN: Topic, NW)	90.28571
(RF: Sentiment)	90.85714
(NN: Link, NW)	92
(RF: Sentiment, NW)	92
(NN: Hashtag, Sentiment, Topic)	92.14286
(RF: Link, Hashtag, Sentiment, NW)	92.42857
(NN: Hashtag, NW)	92.71429
(KNN: Sentiment, Topic, NW)	93.42857
(NN: Link)	93.71429
(RF: Hashtag, Sentiment, NW)	94.42857
(NN: Link, Hashtag, Topic)	95.14286
(NN: Sentiment)	95.42857
(NN: Link, Sentiment)	95.42857
(KNN: Hashtag, Sentiment, Topic, NW)	95.42857
(NN: Link, Hashtag, Topic, NW)	96.28571
(RF: Link)	96.42857
(LR: Topic)	96.71429
(NN: Sentiment, Topic, NW)	97
(RF: Link, Topic, NW)	97.14286
(RF: Hashtag, NW)	97.28571
(RF: Sentiment, Topic, NW)	97.28571
(LR: Link)	98

(NN: Link, Hashtag)	99.14286
(RF: Hashtag, Topic, NW)	99.14286
(NN: NW)	99.85714
(RF: Hashtag, Topic)	100.1429
(RF: Link, NW)	100.4286
(KNN: Link, Hashtag, Sentiment, NW)	100.4286
(NN: Hashtag, Sentiment)	101.1429
(NN: Hashtag, Topic)	101.7143
(RF: Link, Hashtag, Sentiment)	101.7143
(NN: Hashtag)	102.1429
(RF: Link, Hashtag)	102.8571
(RF: Hashtag, Sentiment)	103.7143
(RF: Link, Hashtag, NW)	104.4286
(RF: Link, Sentiment, Topic, NW)	104.8571
(NN: Link, Hashtag, Sentiment, Topic, NW)	104.8571
(LR: Sentiment)	106.1429
(RF: Link, Hashtag, Topic, NW)	106.4286
(GBDT: Sentiment, Topic)	106.7143
(NN: Link, Sentiment, Topic)	107.7143
(RF: Link, Hashtag, Sentiment, Topic, NW)	108.1429
(LR: NW)	109.1429
(GBDT: Hashtag, Sentiment, Topic)	113
(RF: Hashtag, Sentiment, Topic, NW)	113.7143
(RF: NW)	114.2857
(LR: Link, Hashtag)	114.4286
(LR: Hashtag)	115.4286
(GBDT: Link, Sentiment, Topic)	116.2857
(GBDT: Link, Hashtag, Sentiment, Topic)	116.5714
(LR: Hashtag, Sentiment)	117.1429
(LR: Hashtag, Topic)	117.2857
(GBDT: Link, Hashtag, Topic)	117.8571
(RF: Hashtag)	118.5714
(GBDT: Topic)	119
(GBDT: Hashtag)	119.8571
(GBDT: Hashtag, Topic)	120.2857
(LR: Link, Topic)	120.4286
(GBDT: Sentiment)	121.1429
(GBDT: Hashtag, Sentiment, Topic, NW)	122
(GBDT: Link)	122.2857
(LR: Link, Sentiment)	123.5714
(LR: Sentiment, NW)	123.5714
(LR: Hashtag, NW)	123.7143
(GBDT: NW)	123.8571
(LR: Link, NW)	125
(LR: Sentiment, Topic)	128.4286
(GBDT: Link, Topic)	129.2857
(LR: Link, Hashtag, Sentiment)	131.8571
(GBDT: Hashtag, Sentiment)	132.4286
(GBDT: Link, Hashtag, Sentiment)	132.7143
(GBDT: Sentiment, NW)	132.8571
(GBDT: Link, Sentiment, Topic, NW)	132.8571
(LR: Hashtag, Sentiment, NW)	133.2857
(LR: Link, Hashtag, Topic)	133.7143
(LR: Topic, NW)	133.8571
(LR: Link, Hashtag, NW)	136.5714
(GBDT: Sentiment, Topic, NW)	136.8571
(LR: Link, Sentiment, NW)	137.5714
(LR: Link, Sentiment, Topic)	139
(GBDT: Link, Sentiment)	140.2857
(GBDT: Topic, NW)	140.8571
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	140.8571
(GBDT: Link, Hashtag, NW)	141.2857
(GBDT: Link, Hashtag)	141.8571

(LR: Hashtag, Sentiment, Topic)	142.5714
(GBDT: Hashtag, Topic, NW)	143
(GBDT: Link, Topic, NW)	143.4286
(LR: Hashtag, Topic, NW)	144.1429
(LR: Link, Hashtag, Sentiment, NW)	144.1429
(GBDT: Link, Hashtag, Sentiment, NW)	147.5714
(GBDT: Link, Hashtag, Topic, NW)	148.2857
(LR: Link, Topic, NW)	149.1429
(LR: Sentiment, Topic, NW)	149.8571
(GBDT: Hashtag, NW)	150.2857
(LR: Link, Hashtag, Sentiment, Topic)	150.4286
(GBDT: Hashtag, Sentiment, NW)	151.5714
(LR: Link, Hashtag, Topic, NW)	155.7143
(LR: Link, Sentiment, Topic, NW)	157.5714
(GBDT: Link, Sentiment, NW)	158
(LR: Hashtag, Sentiment, Topic, NW)	162.7143
(GBDT: Link, NW)	163.4286
(LR: Link, Hashtag, Sentiment, Topic, NW)	166.1429

Table 11.32: Model-Variable Combination Rankings Across All IRT Data - Retweets

Model-Variable Combination	**Average Rank**
(KNN: Topic)	25
(KNN: Link, Topic)	30.14286
(KNN: Link, Hashtag, Topic, NW)	33.57143
(KNN: Link)	39
(KNN: Hashtag, Topic, NW)	40.57143
(KNN: Link, Sentiment, NW)	42.28571
(KNN: Link, Sentiment)	45.42857
(KNN: Link, Hashtag, Sentiment, Topic, NW)	50.28571
(KNN: Sentiment, Topic)	51.42857
(KNN: Hashtag, Sentiment, Topic)	52
(KNN: Link, Sentiment, Topic)	52.85714
(KNN: Sentiment)	53
(KNN: Topic, NW)	53.71429
(KNN: Hashtag, Topic)	54.57143
(SVR: Topic, NW)	60.85714
(SVR: Link, Topic, NW)	61.14286
SVR: Hashtag, Topic, NW)	61.42857
(KNN: Link, Hashtag)	62.28571
(SVR: Link, Hashtag, Topic, NW)	62.57143
(SVR: Sentiment)	64.14286
(SVR: Link, Sentiment)	64.42857
(SVR: Hashtag, Sentiment)	64.71429
(SVR: Sentiment, Topic)	64.71429
(KNN: Link, Topic, NW)	64.85714
(KNN: Link, Sentiment, Topic, NW)	65.28571
(KNN: Hashtag, Sentiment, NW)	65.71429
(KNN: Hashtag, NW)	66.42857
(SVR: Link, Topic)	66.85714
(SVR: Hashtag, Topic)	66.85714
(KNN: Link, Hashtag, Topic)	67
(SVR: Link, Hashtag, Sentiment)	67.28571
(SVR: Hashtag, Sentiment, NW)	67.57143
(SVR: Topic)	67.71429
(SVR: Hashtag, Sentiment, Topic, NW)	68
(NN: Link, Topic, NW)	68.14286

(SVR: Link, Hashtag, Topic)	68.42857
(SVR: Sentiment, NW)	68.57143
(SVR: Link, Sentiment, Topic, NW)	69.28571
(KNN: Hashtag, Sentiment)	69.42857
(NN: Link, Sentiment, NW)	69.42857
(KNN: Hashtag)	69.57143
(KNN: Sentiment, NW)	69.57143
(KNN: Link, NW)	69.71429
(KNN: Hashtag, Sentiment, Topic, NW)	69.71429
(SVR: Sentiment, Topic, NW)	70
(SVR: Link, NW)	70.14286
(SVR: NW)	70.42857
(SVR: Hashtag, NW)	71.14286
(SVR: Link, Hashtag, NW)	71.14286
(SVR: Hashtag, Sentiment, Topic)	71.85714
(SVR: Link, Hashtag, Sentiment, Topic, NW)	72.85714
(NN: Link, Hashtag, Topic)	73
(NN: Link, NW)	73.57143
(SVR: Link, Sentiment, Topic)	73.85714
(KNN: Link, Hashtag, Sentiment)	75
(SVR: Link, Hashtag, Sentiment, NW)	75
(SVR: Link, Sentiment, NW)	75.42857
(RF: Link, Sentiment, NW)	76
(RF: Link, Hashtag, NW)	76.42857
(NN: Link, Sentiment)	77
(KNN: NW)	77.57143
(KNN: Link, Hashtag, NW)	79.14286
(SVR: Link, Hashtag, Sentiment, Topic)	79.14286
(NN: Sentiment, Topic)	80.14286
(RF: Link, Hashtag, Sentiment, NW)	82
(KNN: Link, Hashtag, Sentiment, NW)	82.57143
(RF: Link, Topic, NW)	84
(NN: Hashtag, NW)	84.57143
(NN: Hashtag, Sentiment, NW)	84.71429
(GBDT: Link)	85.14286
(RF: Link, Topic)	85.28571
(NN: Sentiment, NW)	85.57143
(NN: Hashtag, Sentiment, Topic)	85.71429
(KNN: Link, Hashtag, Sentiment, Topic)	86.14286
(KNN: Sentiment, Topic, NW)	86.85714
(RF: Topic, NW)	87.28571
(NN: Link, Topic)	87.57143
(NN: Link, Hashtag, Sentiment)	88.28571
(RF: Link, Sentiment, Topic)	89
(NN: Link, Hashtag, Sentiment, NW)	89
(RF: Hashtag, Topic, NW)	90.14286
(RF: Link, Hashtag, Sentiment, Topic, NW)	90.28571
(NN: Topic, NW)	90.42857
(RF: Topic)	91
(RF: Link, Sentiment)	91
(RF: Sentiment, NW)	91.28571
(SVR: Link, Hashtag)	91.42857
(RF: Sentiment, Topic, NW)	91.57143
(NN: Link)	92.14286
(RF: Link, Sentiment, Topic, NW)	92.28571
(RF: Link, NW)	92.42857
(RF: Link, Hashtag, Sentiment)	92.57143
(RF: Link, Hashtag, Sentiment, Topic)	93
(RF: Hashtag, Sentiment, NW)	93.28571
(RF: Link, Hashtag)	93.71429
(GBDT: Sentiment, Topic)	93.71429
(RF: Link, Hashtag, Topic)	94.71429
(NN: Sentiment, Topic, NW)	95

(RF: Hashtag, NW)	95.42857
(NN: Hashtag, Sentiment)	95.57143
(NN: Link, Sentiment, Topic)	95.71429
(NN: Link, Sentiment, Topic, NW)	95.85714
(NN: NW)	96.42857
(GBDT: NW)	96.42857
(NN: Link, Hashtag, Sentiment, Topic, NW)	96.42857
(NN: Hashtag, Topic, NW)	97.14286
(NN: Link, Hashtag, Sentiment, Topic)	97.14286
(NN: Topic)	97.42857
(RF: Hashtag, Sentiment, Topic, NW)	97.42857
(LR: Topic)	98
(LR: Link)	98.28571
(NN: Link, Hashtag, NW)	98.57143
(RF: Link, Hashtag, Topic, NW)	98.57143
(RF: Sentiment, Topic)	99.14286
(RF: Link)	99.28571
(RF: NW)	100.1429
(NN: Sentiment)	100.8571
(NN: Hashtag)	101
(RF: Hashtag, Sentiment, Topic)	101.4286
(GBDT: Sentiment)	101.8571
(RF: Sentiment)	102
(GBDT: Hashtag, Sentiment)	103.2857
(RF: Hashtag, Topic)	103.7143
(RF: Hashtag, Sentiment)	105
(LR: Hashtag, Sentiment)	106
(LR: Sentiment)	106.1429
(GBDT: Link, Hashtag)	106.8571
(GBDT: Sentiment, Topic, NW)	107
(GBDT: Hashtag, Topic)	108.2857
(NN: Link, Hashtag)	108.5714
(NN: Hashtag, Topic)	109.4286
(SVR: Link)	110.2857
(LR: Link, Hashtag)	110.4286
(LR: Hashtag)	110.5714
(GBDT: Hashtag)	110.8571
(SVR: Hashtag)	111
(LR: Link, Sentiment)	112.1429
(GBDT: Link, Sentiment)	112.4286
(GBDT: Link, Sentiment, Topic)	112.5714
(RF: Hashtag)	113.7143
(GBDT: Topic)	113.7143
(LR: NW)	114.5714
(LR: Hashtag, Topic)	114.5714
(NN: Link, Hashtag, Topic, NW)	114.5714
(LR: Link, Hashtag, Sentiment)	115.8571
(NN: Hashtag, Sentiment, Topic, NW)	116.4286
(GBDT: Link, Hashtag, Topic, NW)	117.8571
(LR: Sentiment, Topic)	118.2857
(GBDT: Hashtag, Sentiment, Topic)	118.4286
(GBDT: Hashtag, Sentiment, Topic, NW)	118.5714
(GBDT: Link, Hashtag, Topic)	119.2857
(GBDT: Hashtag, Sentiment, NW)	120
(LR: Hashtag, NW)	121.2857
(GBDT: Link, Hashtag, Sentiment, Topic)	124
(LR: Link, NW)	124.4286
(LR: Topic, NW)	124.4286
(GBDT: Sentiment, NW)	124.7143
(LR: Sentiment, NW)	125.7143
(LR: Hashtag, Sentiment, NW)	125.7143
(GBDT: Topic, NW)	125.8571
(LR: Link, Sentiment, NW)	126.2857

(LR: Link, Topic)	126.8571
(GBDT: Link, Sentiment, NW)	127
(GBDT: Link, Sentiment, Topic, NW)	127.2857
(GBDT: Link, Hashtag, Sentiment, NW)	128.1429
(GBDT: Link, Topic)	129.4286
(GBDT: Hashtag, NW)	131.1429
(LR: Hashtag, Sentiment, Topic)	131.1429
(LR: Sentiment, Topic, NW)	131.2857
(LR: Link, Sentiment, Topic)	132.7143
(GBDT: Link, Hashtag, Sentiment, Topic, NW)	132.8571
(GBDT: Link, NW)	133.4286
(GBDT: Link, Topic, NW)	133.4286
(LR: Hashtag, Topic, NW)	134.4286
(LR: Link, Hashtag, Sentiment, NW)	135
(GBDT: Link, Hashtag, NW)	135.4286
(LR: Link, Hashtag, NW)	135.8571
(LR: Link, Topic, NW)	138.5714
(LR: Link, Sentiment, Topic, NW)	139.2857
(LR: Hashtag, Sentiment, Topic, NW)	139.8571
(LR: Link, Hashtag, Topic)	140.7143
(GBDT: Link, Hashtag, Sentiment)	145.1429
(LR: Link, Hashtag, Sentiment, Topic)	145.4286
(GBDT: Hashtag, Topic, NW)	146.1429
(LR: Link, Hashtag, Sentiment, Topic, NW)	149.1429
(LR: Link, Hashtag, Topic, NW)	150