Student Name: Jesseca Johnson

Course: DSC 501: Introduction to Data Science

Dataset Descriptive Statistics with Excel

Contents

Research Question and Variables

My research question is: How early was COVID-19 spreading in three different locations based on Google Trends data?

My independent variable is time (specifically September 2018-February 2020).

My dependent variables are Google Trends index scores (discussed further below).

Null hypothesis:

H_0 = There was no significant increase in Google searches for COVID-19 symptoms in late 2019 or the first two months of 2020.

Alternative hypothesis:

H_a = There was a significant increase in Google searches for at least one COVID-19 symptom in late 2019 or the first two months of 2020.

Project Review

Since Week 2, I have had to change my project, originally intended to examine Baidu search results in two cities in China, to examine areas outside China, due to the difficulty in obtaining a Baidu account without a legitimate, non-VoIP Chinese phone number. After looking at Google Trends search results for over 20 cities and regions, I identified three areas that appeared on visual inspection to possibly have had a significant increase in searches for at least one covid symptom in late 2019 or early 2020:

1. Seattle-Tacoma, US
2. Île-de-France, France (includes Paris metro area)
3. Manila, Philippines

Search Criteria

The eight original search terms have been narrowed to four due to topics on Google Trends returning searches about concepts rather than solely individual terms (more on this below):

1. Diarrhea
2. Shortness of breath
3. Anosmia (loss of smell)
4. Ageusia (loss of taste)

Google Search Engine Market Share

In 2020, Google was the most popular search engine in the United States with an 89% market share (Search Engine Market Share United States of America January-December 2020, StatCounter, 2021), the most popular search engine in France with a 92% market share (Search Engine Market Share France January-December 2020, Stat Counter, 2021), and the most popular search engine in the Philippines with a 96% market share (Search Engine Market Share Philippines January-December 2020, StatCounter, 2021).

Google Trends Indices

My project uses Google Trends indices, a ratio scale of 0-100 of how many searches were performed on any given term or topic over time, by week. This is what the indices represent according to Google Trends:

Interest over time

Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term (Google Trends, 2021).

Language and Topic Considerations

Google distinguishes between "terms," which return results only in the original language that was searched for, and "topics," which return results in any language (Trends Help, n. d.). Since at least three languages will be involved in the project, I searched for topics rather than terms.

Time Scope

This project compares 26 weeks in 2018-19 with the same weeks in 2019-20 beginning in September and ending at the end of February. These have been labeled 1-26, with 1 representing the first week of September and 26 representing the last week of February in each year.

I. Descriptive Statistics – Seattle-Tacoma, US

Google Trends Data (Google Trends, 2021)

Seattle-Tacoma, US										
Timeframe			Diarrhea		Shortness of Breath		Anosmia		Ageusia	
Week	Week Start Date		2018-19	2019-20	2018-19	2019-20	2018-19	2019-20	2018-19	2019-20
1	9/2/2018	9/1/2019	48	62	0	0	2	0	0	0
2	9/9/2018	9/8/2019	50	45	0	0	0	0	0	1
3	9/16/2018	9/15/2019	53	49	0	0	2	0	0	1
4	9/23/2018	9/22/2019	34	69	0	0	0	0	0	1
5	9/30/2018	9/29/2019	52	48	0	0	2	0	0	0
6	10/7/2018	10/6/2019	60	61	0	0	2	2	0	0
7	10/14/2018	10/13/2019	43	55	0	0	0	2	0	2
8	10/21/2018	10/20/2019	48	39	0	0	0	0	0	0
9	10/28/2018	10/27/2019	58	52	0	0	2	2	0	0
10	11/4/2018	11/3/2019	45	62	0	0	0	0	0	0
11	11/11/2018	11/10/2019	56	53	0	0	0	0	3	0
12	11/18/2018	11/17/2019	52	56	0	0	0	0	0	0
13	11/25/2018	11/24/2019	48	69	0	0	0	0	0	0
14	12/2/2018	12/1/2019	51	60	0	0	2	1	0	0
15	12/9/2018	12/8/2019	46	51	0	0	0	0	0	0
16	12/16/2018	12/15/2019	37	46	0	0	0	2	0	0
17	12/23/2018	12/22/2019	57	66	0	0	0	2	0	0
18	12/30/2018	12/29/2019	75	54	0	0	2	0	0	0
19	1/6/2019	1/5/2020	46	50	0	0	2	1	2	0
20	1/13/2019	1/12/2020	41	71	0	0	0	1	0	1
21	1/20/2019	1/19/2020	70	49	0	0	0	0	0	0
22	1/27/2019	1/26/2020	50	54	0	0	0	0	0	1
23	2/3/2019	2/2/2020	55	53	0	0	3	0	3	0
24	2/10/2019	2/9/2020	43	64	0	0	0	0	0	0
25	2/17/2019	2/16/2020	83	67	0	0	0	1	0	0
26	2/24/2019	2/23/2020	51	67	0	0	2	1	0	1

***Figure 1**: Google Trends indices for "diarrhea," "shortness of breath," "anosmia," and "ageusia" searches in Seattle-Tacoma in September-February 2018-19 and September-February 2019-20*

Because of the high number of zeroes for "shortness of breath," "anosmia" and "ageusia," I have excluded those variables for this location. The following statistics will describe searches over time for "diarrhea" only.

Descriptive Statistics – "Diarrhea"

Seattle-Tacoma Google Searches for "Diarrhea"			
2018-19		*2019-20*	
Mean	52	Mean	56.62
Standard Error	2.14	Standard Error	1.69
Median	50.5	Median	54.5
Mode	48	Mode	62
Standard Deviation	10.91	Standard Deviation	8.60
Sample Variance	119.04	Sample Variance	73.93
Kurtosis	2.00	Kurtosis	-0.89
Skewness	1.19	Skewness	0.02
Range	49	Range	32
Minimum	34	Minimum	39
Maximum	83	Maximum	71
Sum	1352	Sum	1472
Count	26	Count	26

***Figure 1a**: Descriptive statistics for Google Trends indices for "diarrhea" searches in Seattle-Tacoma in September-February 2018-19 and September-February 2019-20*

The correlation coefficient -.03, indicating that there is virtually no relationship between the 2018-19 and 2019-20 indices. Although the mean, median, and mode are higher and the range and standard deviation are smaller in 2019-20, it is difficult to tell whether any of these are significant in terms of the time of year without performing some different calculations specific to time series analyses. This will be better to do in Alteryx, but Excel can give us a preview via a line chart. It looks like there may have been a significant overall increase starting the week of October 20 (week 8) and then again during the week of January 20, 2020 (week 21), with some peaks and valleys in between, but we will need to investigate further to determine whether this was actually significant. The peaks during the weeks of November 24 (week 13) and January 12 (week 20) seem to be significant as well with such a large difference vs. the previous year (48 vs. 69 in November and 41 vs. 71 in January).

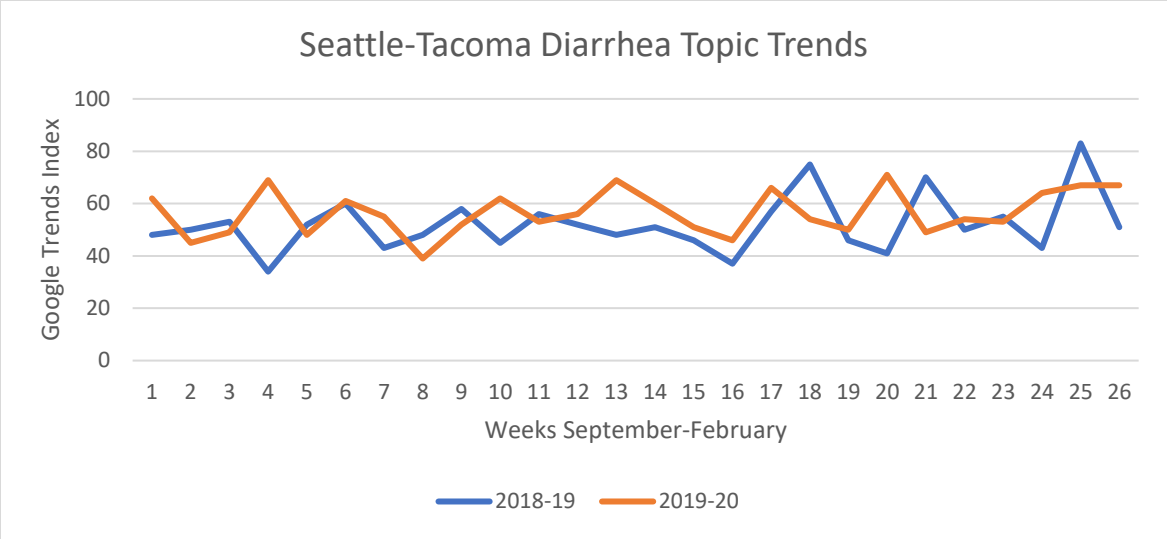

***Figure 1b**: Line chart for Google Trends indices for "diarrhea" searches in Seattle-Tacoma in September-February 2018-19 and September-February 2019-20*

II. Descriptive Statistics – Île-de-France, France

Note: Île-de-France is the most populated region of France with over 12 million people and includes the Paris metro area (Statista, 2021).

Google Trends Data (Google Trends, 2021)

Île-de-France, France										
Timeframe			Diarrhea		Shortness of Breath		Anosmia		Ageusia	
Weeknum	Week Start Date		2018-19	2019-20	2018-19	2019-20	2018-19	2019-20	2018-19	2019-20
1	9/2/2018	9/1/2019	47	49	0	0	2	0	0	0
2	9/9/2018	9/8/2019	50	42	0	1	0	2	2	0
3	9/16/2018	9/15/2019	54	46	0	0	2	0	2	0
4	9/23/2018	9/22/2019	58	56	0	0	2	0	1	1
5	9/30/2018	9/29/2019	53	47	0	0	4	0	2	1
6	10/7/2018	10/6/2019	51	40	0	0	2	0	1	1
7	10/14/2018	10/13/2019	60	39	0	0	2	2	0	1
8	10/21/2018	10/20/2019	55	45	0	0	0	1	2	2
9	10/28/2018	10/27/2019	61	44	0	0	3	0	2	2
10	11/4/2018	11/3/2019	43	50	1	0	0	1	0	1
11	11/11/2018	11/10/2019	45	51	0	0	1	0	1	0
12	11/18/2018	11/17/2019	46	69	0	0	1	2	2	0
13	11/25/2018	11/24/2019	61	44	0	0	0	2	0	0
14	12/2/2018	12/1/2019	62	44	0	1	1	0	2	1
15	12/9/2018	12/8/2019	60	54	1	0	2	1	2	3
16	12/16/2018	12/15/2019	54	47	0	1	1	0	4	2
17	12/23/2018	12/22/2019	55	91	0	0	4	1	4	3
18	12/30/2018	12/29/2019	65	77	0	2	3	4	2	1
19	1/6/2019	1/5/2020	57	53	0	1	3	0	0	2
20	1/13/2019	1/12/2020	54	57	0	1	1	1	0	2
21	1/20/2019	1/19/2020	55	42	0	0	0	0	3	1
22	1/27/2019	1/26/2020	49	47	1	1	1	1	1	2
23	2/3/2019	2/2/2020	53	54	0	0	1	0	2	2
24	2/10/2019	2/9/2020	55	51	1	0	1	2	1	2
25	2/17/2019	2/16/2020	57	46	1	0	0	1	0	2
26	2/24/2019	2/23/2020	48	67	1	1	2	2	0	3

***Figure 2**: Google Trends indices for "diarrhea," "shortness of breath," "anosmia," and "ageusia" searches in Île-de-France in September-February 2018-19 and September-February 2019-20*

Because of the high number of zeroes for "shortness of breath," I have excluded that variable in my calculations.

Descriptive Statistics – "Diarrhea"

Île-de-France Google Searches for "Diarrhea"			
2018-19		*2019-20*	
Mean	54.15	Mean	52
Standard Error	1.11	Standard Error	2.36
Median	54.5	Median	48
Mode	55	Mode	47
Standard Deviation	5.64	Standard Deviation	12.04
Sample Variance	31.82	Sample Variance	145.04
Kurtosis	-0.54	Kurtosis	3.68
Skewness	-0.15	Skewness	1.86
Range	22	Range	52
Minimum	43	Minimum	39
Maximum	65	Maximum	91
Sum	1408	Sum	1352
Count	26	Count	26

***Figure 2a**: Descriptive statistics for Google Trends indices for "diarrhea" searches in Île-de-France in September-February 2018-19 and September-February 2019-20*

The correlation coefficient is -.001, indicating that there is virtually no relationship between the 2018-19 and 2019-20 indices. Although the mean, median, and mode appear to be similar and expected, the much higher standard deviation, range, and maximum for the 2019-20 data warrants investigation. The line chart indicates an increase in searches from the week of October 13, 2019 (week 7), which peaked the week of November 17 (week 12). There was a subsequent drastic increase in searches starting the week of December 15, 2019 (week 16) reaching a low point for the week of January 19, 2020 (week 21), only to escalate again. Could COVID-19 already have been spreading in France in October or December 2019?

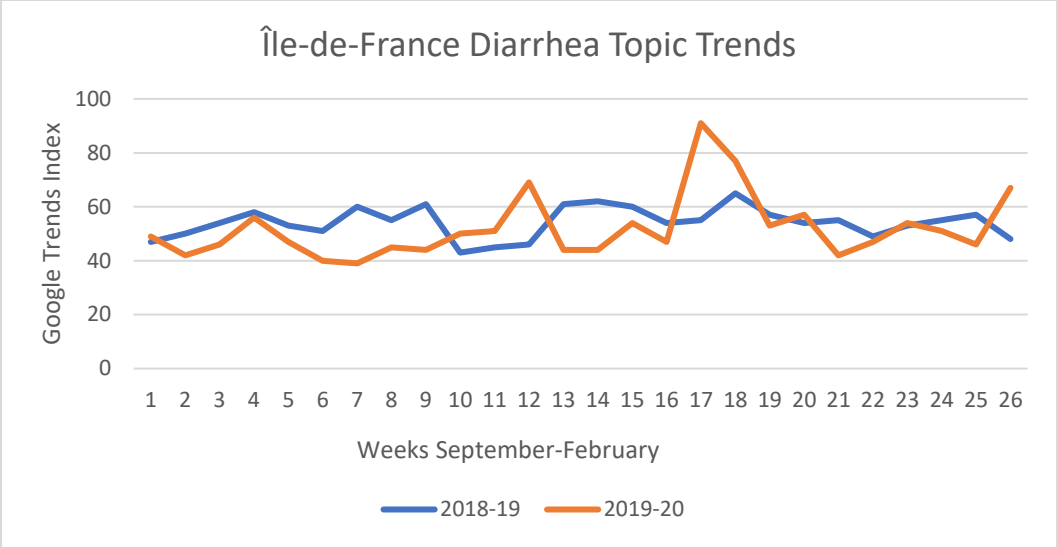

***Figure 2b**: Line chart for Google Trends indices for "diarrhea" searches in Île-de-France in September-February 2018-19 and September-February 2019-20*

Descriptive Statistics – "Anosmia" (loss of smell)

Île-de-France, France Google searches for "Anosmia"			
2018-19		*2019-20*	
Mean	1.5	Mean	0.88
Standard Error	0.24	Standard Error	0.20
Median	1	Median	1
Mode	1	Mode	0
Standard Deviation	1.21	Standard Deviation	1.03
Sample Variance	1.46	Sample Variance	1.07
Kurtosis	-0.42	Kurtosis	1.56
Skewness	0.52	Skewness	1.19
Range	4	Range	4
Minimum	0	Minimum	0
Maximum	4	Maximum	4
Sum	39	Sum	23
Count	26	Count	26

***Figure 2c**: Descriptive statistics for Google Trends indices for "anosmia" searches in Île-de-France in September-February 2018-19 and September-February 2019-20*

The correlation coefficient is -.01, which again illustrates that there is no relationship between the two variables. There does not appear to be anything in the descriptive statistics that would indicate a significant increase in searches in 2019-20 over the previous year. This is supported by the line chart; although there was an increase starting the week of December 15, it seems to follow the previous year's trend. At any rate, the scale appears to be quite small and probably insignificant.

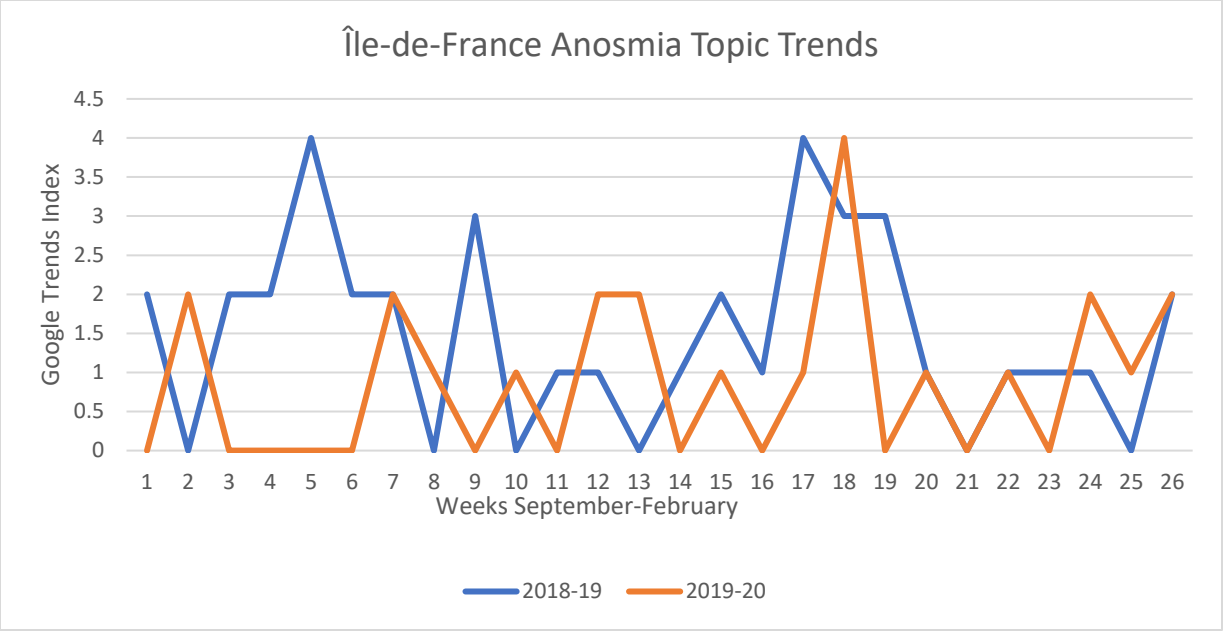

***Figure 2d**: Line chart for Google Trends indices for "anosmia" searches in Île-de-France in September-February 2018-19 and September-February 2019-20*

Descriptive Statistics – "Ageusia" (loss of taste)

Île-de-France, France Google searches for "Ageusia"			
2018-19		*2019-20*	
Mean	1.38	Mean	1.35
Standard Error	0.24	Standard Error	0.19
Median	1.5	Median	1
Mode	2	Mode	2
Standard Deviation	1.20	Standard Deviation	0.98
Sample Variance	1.45	Sample Variance	0.96
Kurtosis	-0.20	Kurtosis	-0.95
Skewness	0.52	Skewness	0.05
Range	4	Range	3
Minimum	0	Minimum	0
Maximum	4	Maximum	3
Sum	36	Sum	35
Count	26	Count	26

***Figure 2e**: Descriptive statistics for Google Trends indices for "ageusia" searches in Île-de-France in September-February 2018-19 and September-February 2019-20*

Again, we see nothing particularly meaningful here. Although the correlation coefficient is again close to zero at .15, indicating little relationship, the line chart seems to reflect the normal ups and

downs that one might expect throughout the year except for an increase during the week of February 16 (week 25).

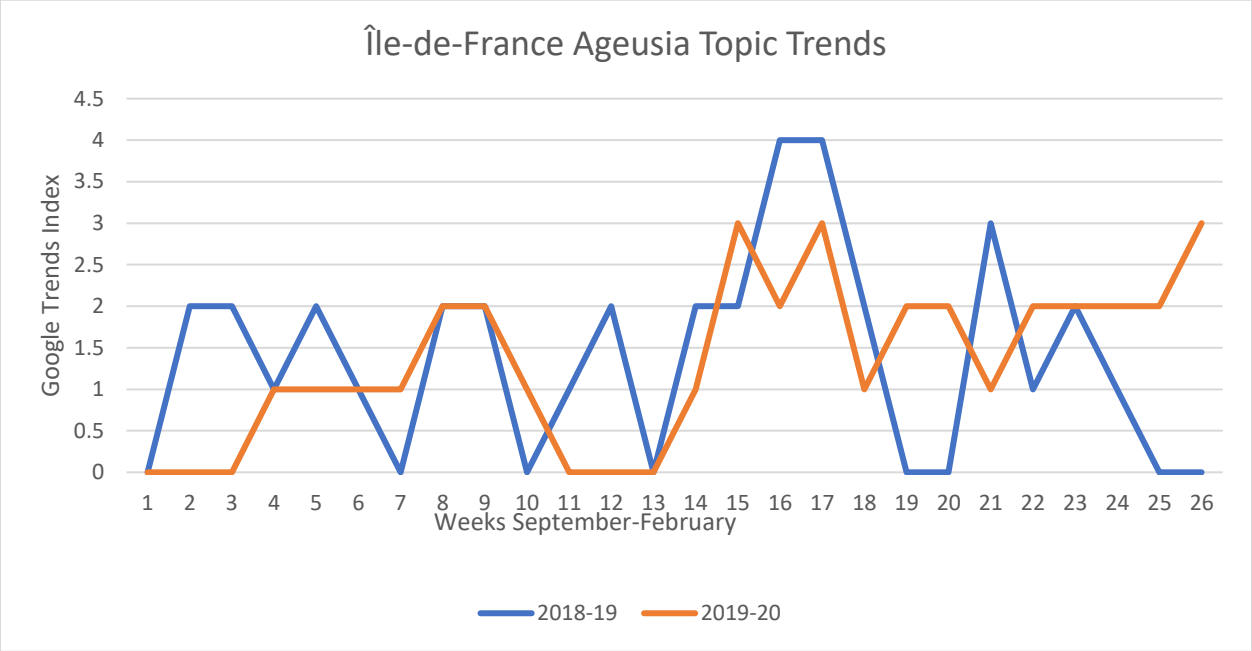

***Figure 2f**: Line chart for Google Trends indices for "ageusia" searches in Île-de-France in September-February 2018-19 and September-February 2019-20*

III. Descriptive Statistics – Manila, Philippines

Google Trends Data (Google Trends, 2021)

Manila, Philippines										
Timeframe			Diarrhea		Shortness of Breath		Anosmia		Ageusia	
Week	Week Start Date		2018-19	2019-20	2018-19	2019-20	2018-19	2019-20	2018-19	2019-20
1	9/2/2018	9/1/2019	52	56	0	0	2	2	0	0
2	9/9/2018	9/8/2019	49	59	0	0	4	0	0	0
3	9/16/2018	9/15/2019	53	53	0	0	2	2	0	0
4	9/23/2018	9/22/2019	59	49	0	0	0	0	2	0
5	9/30/2018	9/29/2019	44	46	0	0	0	0	0	0
6	10/7/2018	10/6/2019	57	54	0	0	2	0	0	2
7	10/14/2018	10/13/2019	46	48	0	0	0	0	0	0
8	10/21/2018	10/20/2019	55	69	0	0	0	0	0	2
9	10/28/2018	10/27/2019	42	42	0	0	0	0	0	0
10	11/4/2018	11/3/2019	52	57	0	0	2	2	0	0
11	11/11/2018	11/10/2019	52	68	0	0	0	2	0	0
12	11/18/2018	11/17/2019	76	52	0	0	0	0	0	0
13	11/25/2018	11/24/2019	45	65	0	0	0	0	0	0
14	12/2/2018	12/1/2019	56	48	0	0	0	2	0	0
15	12/9/2018	12/8/2019	44	54	0	0	2	4	0	0
16	12/16/2018	12/15/2019	66	79	0	0	0	0	0	0
17	12/23/2018	12/22/2019	73	80	0	0	2	0	2	0
18	12/30/2018	12/29/2019	57	84	0	0	0	2	0	0
19	1/6/2019	1/5/2020	80	100	0	0	0	2	2	2
20	1/13/2019	1/12/2020	60	66	0	0	0	0	2	0
21	1/20/2019	1/19/2020	87	62	0	0	0	2	0	0
22	1/27/2019	1/26/2020	61	95	0	0	0	0	0	2
23	2/3/2019	2/2/2020	73	77	0	0	2	0	2	0
24	2/10/2019	2/9/2020	73	78	0	0	0	0	0	0
25	2/17/2019	2/16/2020	64	70	0	0	0	0	0	2
26	2/24/2019	2/23/2020	76	81	0	0	0	0	0	0

***Figure 3**: Google Trends indices for "diarrhea," "shortness of breath," "anosmia," and "ageusia" searches in Metro Manila, Philippines in September-February 2018-19 and September-February 2019-20*

"Shortness of breath," "anosmia" and "ageusia" were omitted due to a high number of zeroes.

Descriptive Statistics – "Diarrhea"

Manila, Philippines Google searches for "Diarrhea"			
2018-19		*2019-20*	
Mean	59.69	Mean	65.08
Standard Error	2.44	Standard Error	3.04
Median	57	Median	63.5
Mode	52	Mode	54
Standard Deviation	12.45	Standard Deviation	15.49
Sample Variance	155.10	Sample Variance	239.83
Kurtosis	-0.67	Kurtosis	-0.42
Skewness	0.51	Skewness	0.57
Range	45	Range	58
Minimum	42	Minimum	42
Maximum	87	Maximum	100
Sum	1552	Sum	1692
Count	26	Count	26

***Figure 3a**: Descriptive statistics for Google Trends indices for "diarrhea" searches in Metro Manila, Philippines in September-February 2018-19 and September-February 2019-20*

The correlation coefficient is .57, indicating that there is a medium amount of correlation between the two variables. This would indicate some level of normalcy for the number of Google searches in 2019-20 versus the previous year. The high standard deviation for 2019-20 is notable, but again, we will not be able to draw any conclusions without using ETS or ARIMA measurements in Alteryx. In the line chart, it is interesting to note the increase in searches beginning the week of December 1 (week 14), increasing all the way to a peak of 100 the week of January 5 (week 19).

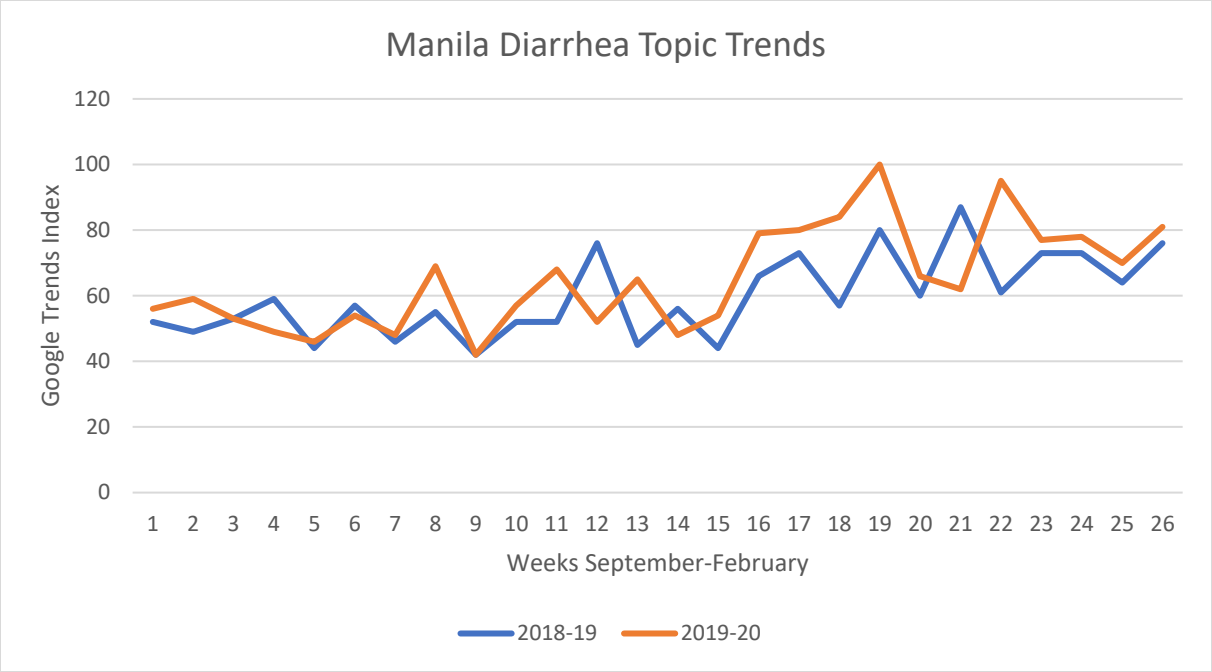

***Figure 3b**: Line chart for Google Trends indices for "diarrhea" searches in Metro Manila, Philippines in September-February 2018-19 and September-February 2019-20*

Works Cited

Google Trends. (2021, June 2). Google. Retrieved June 2, 2021 from https://trends.google.com/

Search Engine Market Share France, Jan-Dec 2020. (2020). StatCounter. Retrieved June 6, 2021 from https://gs.statcounter.com/search-engine-market-share/all/france/2020

Search Engine Market Share Philippines, Jan-Dec 2020. (2020). StatCounter. Retrieved June 6, 2021 from https://gs.statcounter.com/search-engine-market-share/all/philippines/2020

Search Engine Market Share United States of America, Jan-Dec 2020. (2020). StatCounter. Retrieved June 6, 2021 from https://gs.statcounter.com/search-engine-market-share/all/united-states-of-america/2020

Total population of France as of January 1st 2021, by region. (2021). Statista. Retrieved June 4, 2021 from https://www.statista.com/statistics/608761/population-of-france-by-region/

Trends Help. (n. d.). Google. Retrieved June 2, 2021 from https://support.google.com/trends/answer/4359550?hl=en