TUN Data Challenge 2018:
Bike MS

Prepared by

Tong Wen Liang, Samuel

Table of Contents

Data Preparation

Quality Issue 1 – Multiple Delimiters

1.1 Data Issue

In the 2016 Donation file, 2017 Donations file and the 2013-2017 Bike Teams file, we noted that there were 16 rows, 10 rows and two rows respectively with multiple delimiters (coma and left parenthesis) specifying the boundaries between variables, which caused the values of these rows to be shifted to the left after import into JMP.

These rows can be identified by performing data filter on the columns on the right to identify any anomalies in values, and to view them in data view to correct these errors.

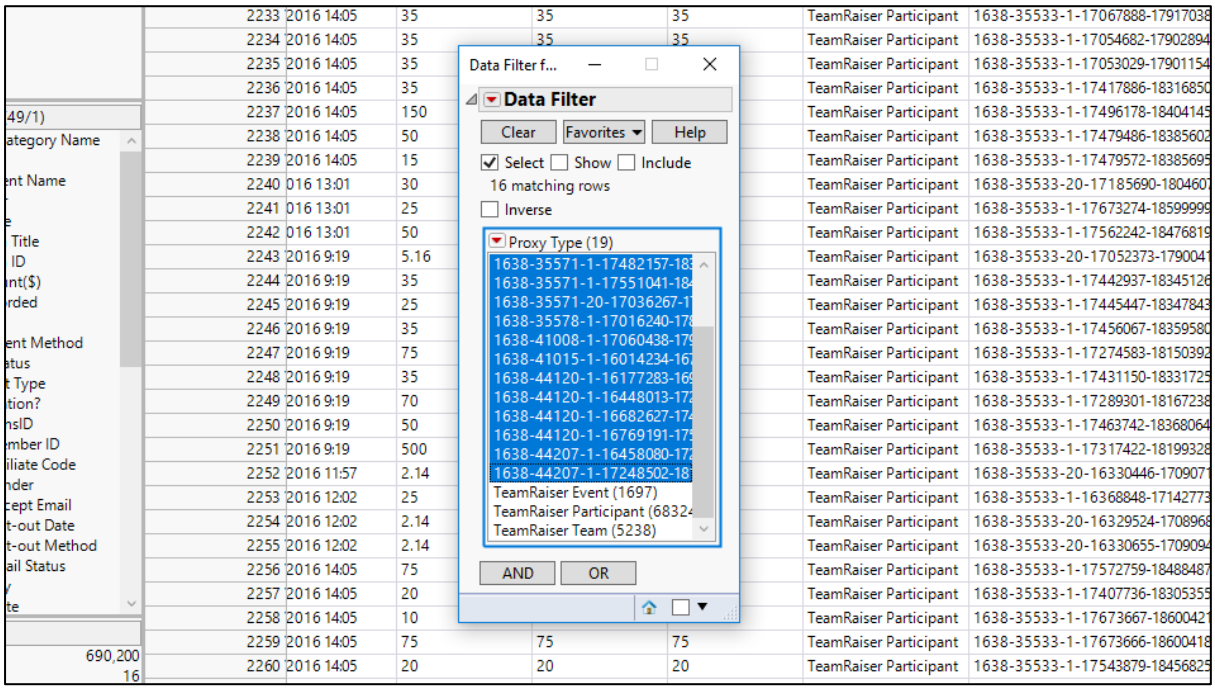


Figure 1: Data Filter of User Confirmation Code on 2016 Donations File

	Security Category Name	Event ID	Public Event Name	Fiscal Year	E
1	ALC Bike Events	27775	Bike MS: Tour de Beach 2016	2016	9/17/201
2	MNM Bike Events	27006	Bike MS: C.H. Robinson MS 150 Ride 2016	2016	6/10/201
3	NYN Bike Events	26036	Bike MS: New York City 2015	2016	10/4/201
4	OHA Bike Events	27178	Bike MS: Buckeye Breakaway 2016	2016	8/6/2016
5	PAE Bike Events	27060	Bike MS: Bike to the Bay 2016	2016	9/10/201
6	PAE Bike Events	27061	Bike MS: City to Shore Ride 2016	2016	9/24/201
7	PAE Bike Events	27061	Bike MS: City to Shore Ride 2016	2016	9/24/201
8	PAE Bike Events	27061	Bike MS: City to Shore Ride 2016	2016	9/24/201
9	PAE Bike Events(27061	Bike MS: City to Shore Ride 2016	2016	9/24/2016 9:27	PAE Bike D
10	TXH Bike Events	27003	2016 BP MS 150	2016	4/16/201
11	TXH Bike Events	27003	2016 BP MS 150	2016	4/16/201
12	TXH Bike Events	27003	2016 BP MS 150	2016	4/16/201
13	TXH Bike Events	27003	2016 BP MS 150	2016	4/16/201
14	TXH Bike Events(27149	Bike MS: Sam's Club Round-Up Ride 2016	2016	4/30/2016 12:49	TXH 2016
15	TXH Bike Events	27149	Bike MS: Sam's Club Round-Up Ride 2016	2016	4/30/201
16	VAR Bike Events	27059	Bike MS: Colonial Crossroads 2016	2016	6/4/2016

Figure 2: Data view of 16 affected rows on 2016 Donations File

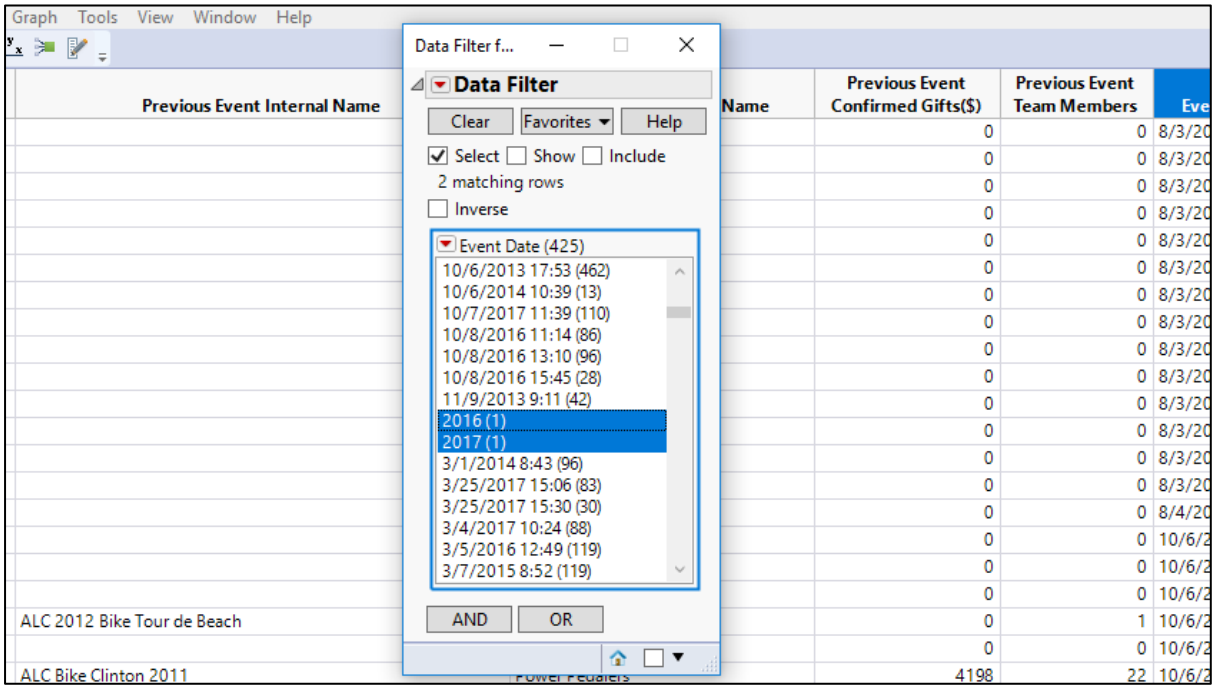


Figure 3: Data Filter of Event Date on 2013-2017 Bike Teams File

1.2 Implications

As these affected rows will be shifted to the left, some of the rows which contain significant information would greatly affect analysis. For example, the Gift Amount ($) in two of the 16 affected rows in the 2016 Donation file are significantly large and may cause the result of any analysis on the Gift Amount ($) to be skewed.

9	TXH Bike Events	27003	35	2016 BP MS 150
10	TXH Bike Events(27149	•	3549432960	2016
11	TXH Bike Events	27149	50	Bike MS: Sam's Club Round-Up Ride 20
12	VAR Bike Events	27059	3545734980	Bike MS: Colonial Crossroads 2016

Figure 4: Example of significant values that may cause results to be skewed during analysis due to presence of multiple delimiters during file import.

Quality Issue 2 – Duplicated Entries

2.1 Data Issue

2016 & 2017 Bike Donations File

According to the Field Definitions, we noted that the User Confirmation Code refers to a unique transaction ID. After summarising the data based on the User Confirmation Code, there are 1,966 duplicated entries.

Cleaned End of Field) By (User Confirmation Code) - JMP Pro

ls DOE Analyze Graph Tools View Window Help

	User Confirmation Code	N Rows	N(User Confirmation ...
1956	1638-42893-1-16101833-16832796	2	2
1957	1638-42893-1-16102884-16833950	2	2
1958	1638-42893-1-16103560-16834676	2	2
1959	1638-42893-1-16105273-16836479	2	2
1960	1638-42893-1-16106630-16837912	2	2
1961	1638-42893-1-16139221-16873903	2	2
1962	1638-42893-1-16139254-16873938	2	2
1963	1638-42893-1-16170581-16909076	2	2
1964	1638-44118-1-16158939-16895904	2	2

Figure 5: Example of duplicated entries in data view

We will first concatenate both 2016 and 2017 Donations files before joining the new concatenated file with itself by matching all columns and checking on “drop multiple” checkboxes to remove all duplicated rows.

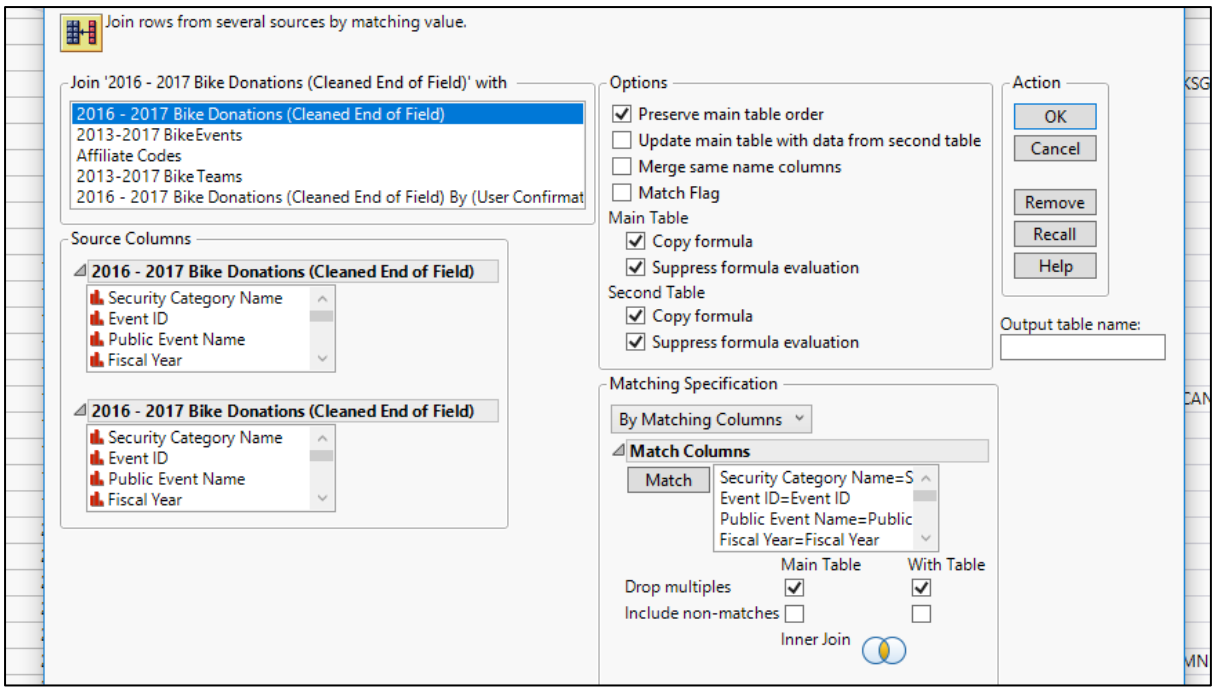
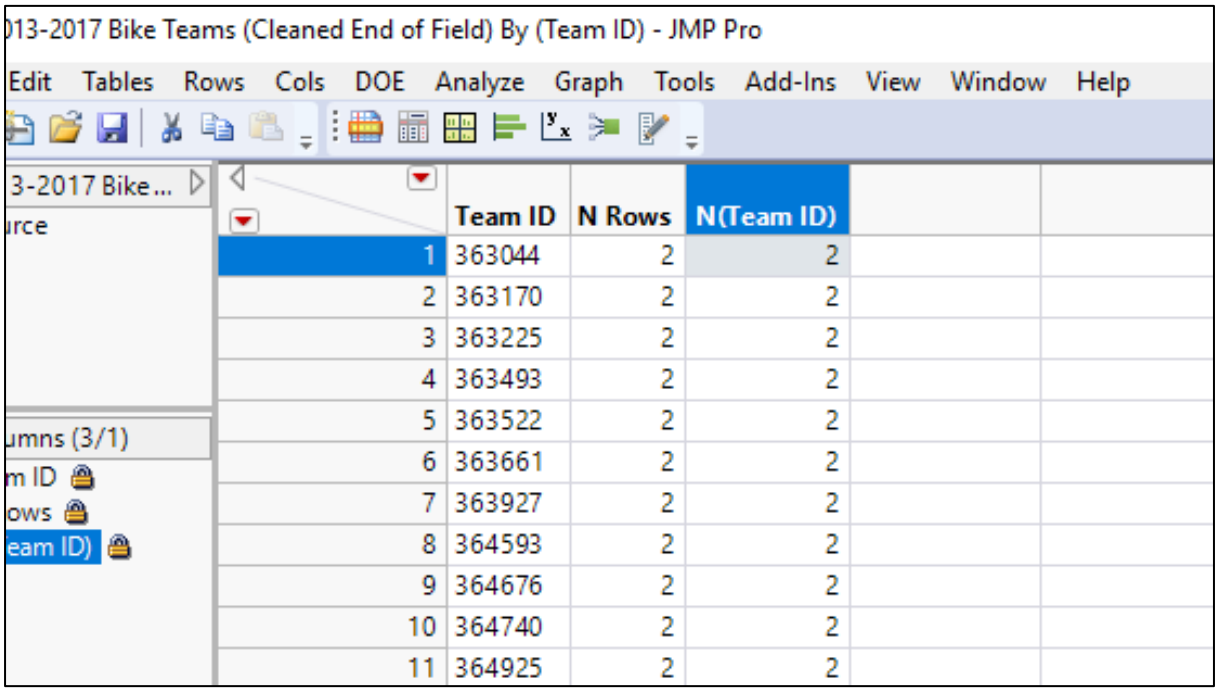


Figure 6: Example of a method to remove duplicated rows

2013-2017 Bike Teams File

In the Teams File, "Team ID" and "Team Name" should be unique to each row. We will first summarise both columns to identify that there are duplicated entries of both, before filtering by descending order to identify total number of duplicates.



	Team ID	N Rows	N(Team ID)
1	363044	2	2
2	363170	2	2
3	363225	2	2
4	363493	2	2
5	363522	2	2
6	363661	2	2
7	363927	2	2
8	364593	2	2
9	364676	2	2
10	364740	2	2
11	364925	2	2

Figure 7: Example of duplicated entries in data view

Thereafter, we joined the Team file with itself to remove duplicated entries.

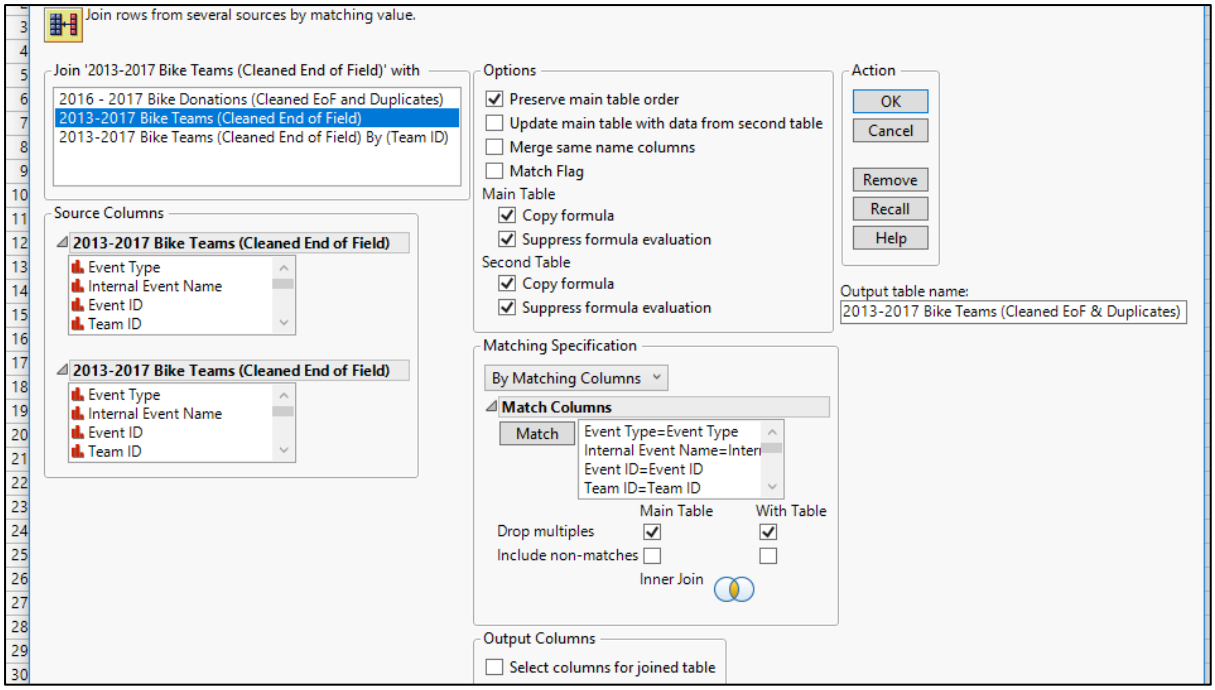


Figure 8: *Example of a method to remove duplicated rows*

2.2 Implications

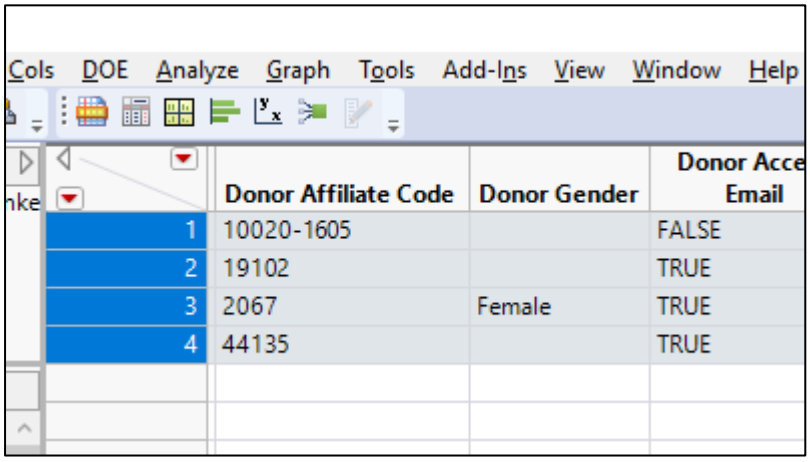
Duplicated entries could cause the distortion of the dataset, especially when it is required to sum all values in the entire column. For example, when it is required to obtain the total amount of gift amount, the duplicated entries will cause the gift amount to be overstated.

Quality Issue 3 – Data Inconsistency

3.1 Data Issue

Misspellings / Inconsistency in Naming Categorical Data

In the 2016 and 2017 Donations file, we noted that there were multiple errors and inconsistencies in the values of several data fields. For example, under Donor Affiliate Code, the values should be a string of three letters. However, the following errors were noted:



	Donor Affiliate Code	Donor Gender	Donor Acce Email
1	10020-1605		FALSE
2	19102		TRUE
3	2067	Female	TRUE
4	44135		TRUE

Figure 9: Examples of errors / inconsistent values

Similarly, under the teams file, we noted that there were several naming inconsistencies under the “Team Division” header:

7	Beer/Brewery	Beer/Brewery
4	Bike Club	Bike Club
15	Bike Shop	Bike Shop
1	Bike Shops	Bike Shops
7	Civic Team	Civic Team
9	Club/Organization	Club/Organization
6873	Corporate	Corporate
813	Corporation	Corporation
2207	Family and Friends	Family and Friends
154	Family/Friends	Family/Friends
1	Frien`s and Family	Frien`s and Family
240	Friend and Family	Friend and Family
19039	Friends and Family	Friends and Family
11	Ohana	Ohana
12	Ohana and Friends	Ohana and Friends
6	Open	Open
3	Open Team	Open Team
10	Organization	Organization
11	Organization (Clubs, Civic Groups, etc.)	Organization (Clubs, Civic Groups, etc.)
222	Organization (Clubs; Civic Groups; etc.)	Organization (Clubs; Civic Groups; etc.)
3	Organization (Clubs; Civic Groups; Place of Worship;	Organization (Clubs; Civic Groups; Place of Worship;
628	Other	Other
11	Place of worship	Place of worship
58	Place of Worship	Place of Worship
1	Religious	Religious

Figure 10: Examples of inconsistency in naming

Inconsistent Header Across Different File

In the National Team Activity File, the values under the header "Event Type" are different from the values under "Event Type" of the Teams file.

Event Categor	Event Type	Local
Bike	MS 150 (2 Day Bike)	missio
Bike	One Day Bike	Breaki
Bike	Bike 1 Day	YaYa
Bike	One Day Bike	Serina
Bike	MS 150 (2 Day Bike)	Team2
Bike	MS 150 (2 Day Bike)	24 Hou

	Event Type	
2249	Bike	Texas
2250	Bike	Texas
2251	Bike	Texas
2252	Bike	Texas
2253	Bike	Texas

Figure 11 & 12: Examples of values under "Event Type" in National Team Activity File (left) and Teams File (right)

Similarly, the header "Event Name" in National Activity File is inconsistent with the "Public Event Name" or "Internal Event Name" in the Donations file, Events file and Teams file.

3.2 Implications

Misspellings / Inconsistency in Naming Categorical Data

Misspellings / inconsistency in naming categorical values could cause difficulty in determining how to attribute values to a specific categorical value. For example, it would be difficult to attribute total amount of donations to a Company name if the name of the same company is inconsistent.

Inconsistent Category/Header Name Across Different File

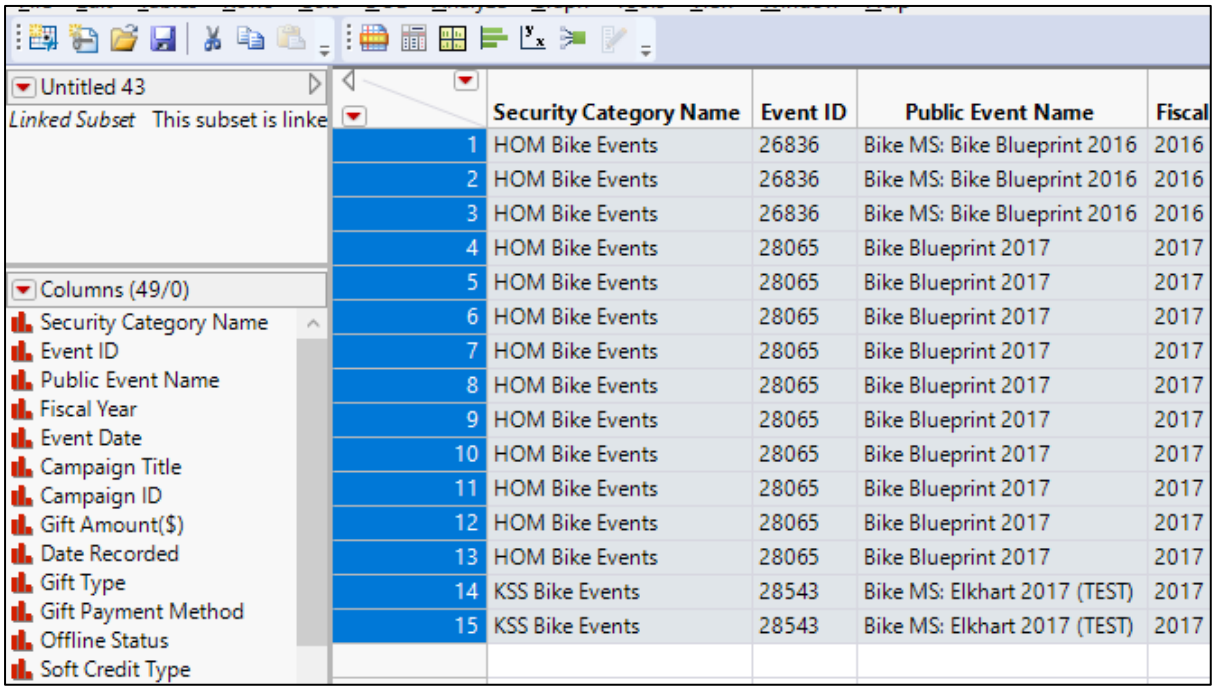
If separate files have the same header names but have different values reflected under these headers, it would prevent ease of data transfer between files to allow for better analysis of data.

Quality Issue 4 – Missing Data

4.1 Data Issue

Security Category Code Not Found in Affiliate Code

Based on 2016 and 2017 Donation files, there are several Security Category Code data that were captured in these files, but not reflected under the "Affiliate Codes" file.

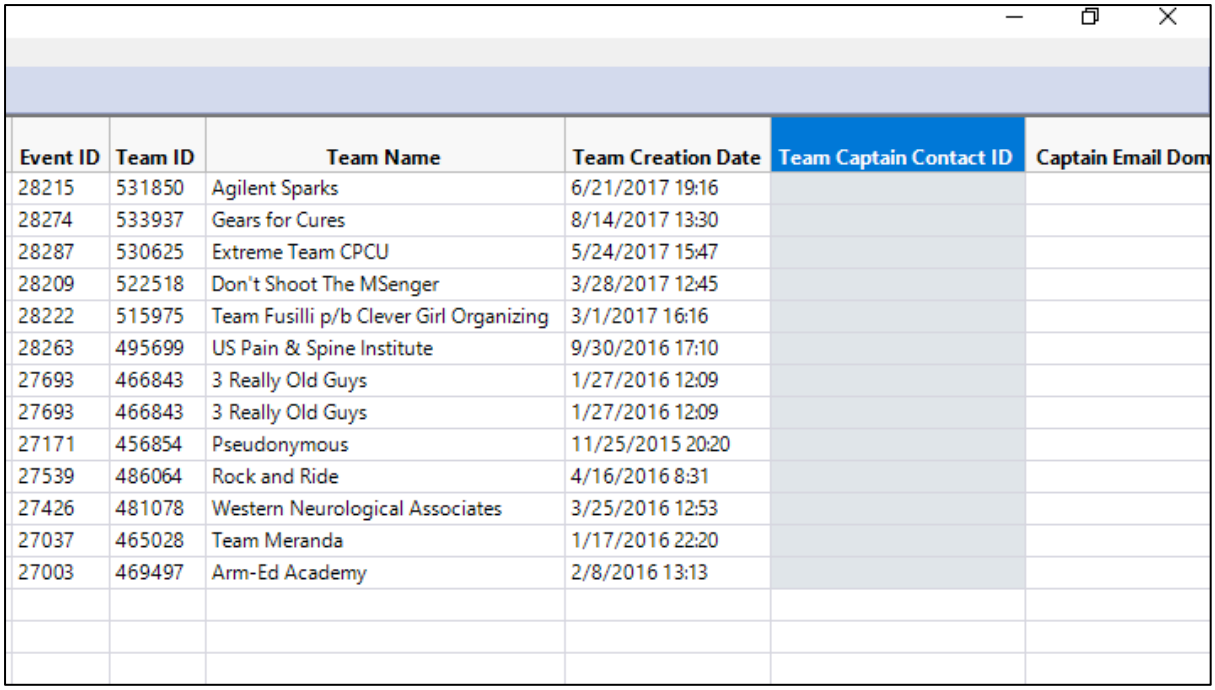


	Security Category Name	Event ID	Public Event Name	Fiscal
1	HOM Bike Events	26836	Bike MS: Bike Blueprint 2016	2016
2	HOM Bike Events	26836	Bike MS: Bike Blueprint 2016	2016
3	HOM Bike Events	26836	Bike MS: Bike Blueprint 2016	2016
4	HOM Bike Events	28065	Bike Blueprint 2017	2017
5	HOM Bike Events	28065	Bike Blueprint 2017	2017
6	HOM Bike Events	28065	Bike Blueprint 2017	2017
7	HOM Bike Events	28065	Bike Blueprint 2017	2017
8	HOM Bike Events	28065	Bike Blueprint 2017	2017
9	HOM Bike Events	28065	Bike Blueprint 2017	2017
10	HOM Bike Events	28065	Bike Blueprint 2017	2017
11	HOM Bike Events	28065	Bike Blueprint 2017	2017
12	HOM Bike Events	28065	Bike Blueprint 2017	2017
13	HOM Bike Events	28065	Bike Blueprint 2017	2017
14	KSS Bike Events	28543	Bike MS: Elkhart 2017 (TEST)	2017
15	KSS Bike Events	28543	Bike MS: Elkhart 2017 (TEST)	2017

Figure 13: Security Category Name in Donations file, but not in "Affiliate Codes" file

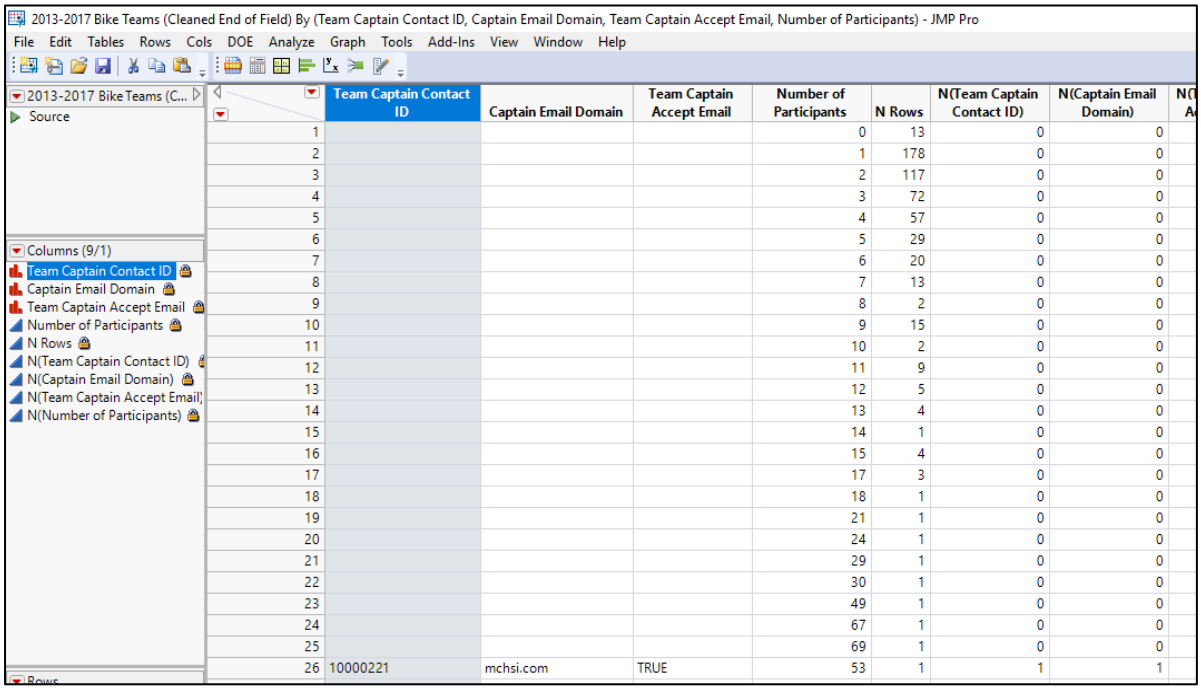
Blank Fields

Under the Teams file, by tabulating the variables "Team Captain Contact ID", "Captain Email Domain", "Team Captain Accept Email", and "Number of Participants", the following missing data were identified.



Event ID	Team ID	Team Name	Team Creation Date	Team Captain Contact ID	Captain Email Dom
28215	531850	Agilent Sparks	6/21/2017 19:16		
28274	533937	Gears for Cures	8/14/2017 13:30		
28287	530625	Extreme Team CPCU	5/24/2017 15:47		
28209	522518	Don't Shoot The MSenger	3/28/2017 12:45		
28222	515975	Team Fusilli p/b Clever Girl Organizing	3/1/2017 16:16		
28263	495699	US Pain & Spine Institute	9/30/2016 17:10		
27693	466843	3 Really Old Guys	1/27/2016 12:09		
27693	466843	3 Really Old Guys	1/27/2016 12:09		
27171	456854	Pseudonymous	11/25/2015 20:20		
27539	486064	Rock and Ride	4/16/2016 8:31		
27426	481078	Western Neurological Associates	3/25/2016 12:53		
27037	465028	Team Meranda	1/17/2016 22:20		
27003	469497	Arm-Ed Academy	2/8/2016 13:13		

Figure 14: 13 instances of team without captain information and no participants



	Team Captain Contact ID	Captain Email Domain	Team Captain Accept Email	Number of Participants	N Rows	N(Team Captain Contact ID)	N(Captain Email Domain)
1				0	13	0	0
2				1	178	0	0
3				2	117	0	0
4				3	72	0	0
5				4	57	0	0
6				5	29	0	0
7				6	20	0	0
8				7	13	0	0
9				8	2	0	0
10				9	15	0	0
11				10	2	0	0
12				11	9	0	0
13				12	5	0	0
14				13	4	0	0
15				14	1	0	0
16				15	4	0	0
17				17	3	0	0
18				18	1	0	0
19				21	1	0	0
20				24	1	0	0
21				29	1	0	0
22				30	1	0	0
23				49	1	0	0
24				67	1	0	0
25				69	1	0	0
26	10000221	mchsi.com	TRUE	53	1	1	1

Figure 15: 549 instances of team without captain information and range of 1 to 178 number of participants

4.2 Implications

Security Category Code Not Found in Affiliate Code

The "Affiliate Code" file should be promptly updated to always ensure data completeness.

Blank Fields

Incomplete data would cause quality of analysis to be poor, where results obtained would not be representative of the total population.

Quality Issue 5 – Multiple Variables Under Single Header

5.1 Data Issue

Under the "Participation Type Name" in the Participants file, there are multiple variables captured in the column under a single header.

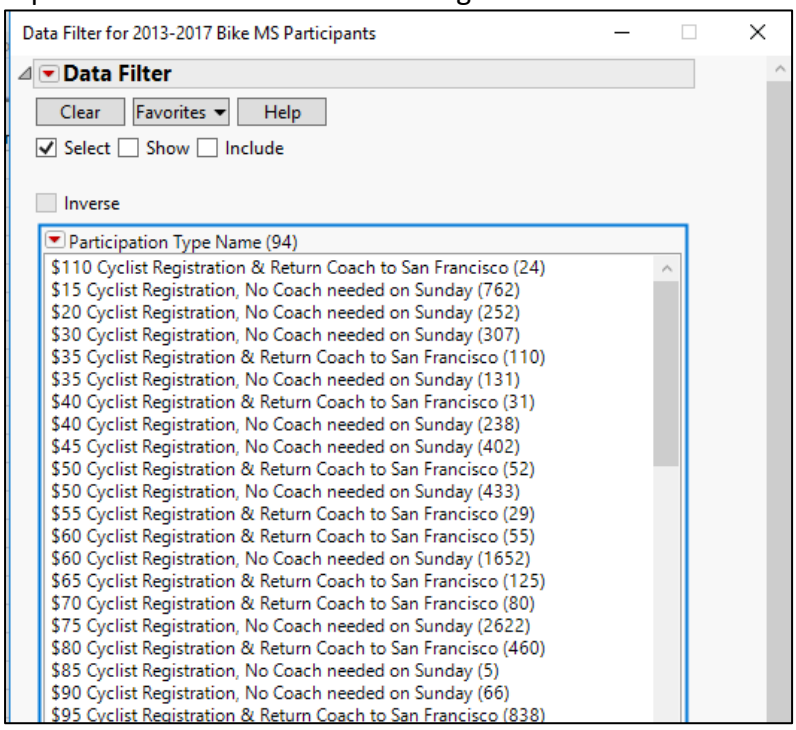


Figure 16: Data filter of Participation Type Name showing multiple variables under a single header

5.2 Implications

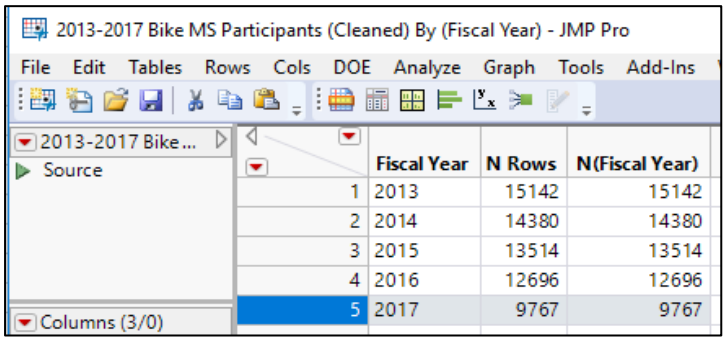
Multiple variables in a single header could cause difficulty in retrieving just a single variable. Therefore, more fields should be created to better segment customers according to their needs.

Data Analysis and Interpretation of Analysis Results

Insight 1 - Discrepancy of Total Number of Participants in Participants File and in Bike Teams File

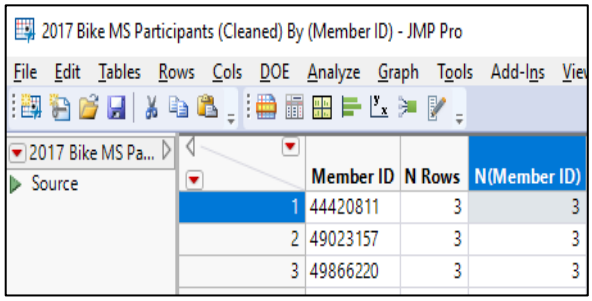
1.1 Data Analysis

From the Participant File, the total number of participants from 2013 to 2017 was 65,499 in aggregate. By summarising data by year, there were 9,767 individual participants taking part in events in 2017.



	Fiscal Year	N Rows	N(Fiscal Year)
1	2013	15142	15142
2	2014	14380	14380
3	2015	13514	13514
4	2016	12696	12696
5	2017	9767	9767

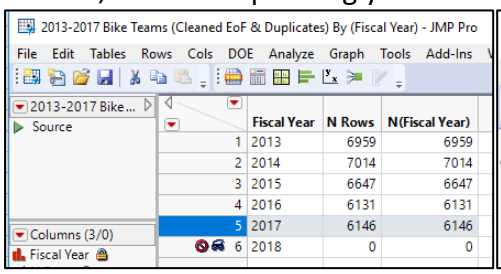
By summarising "Member ID" on the Participant File, we noted that it is possible for one member to be involved in multiple teams. By selecting data view of one sample, we can confirm that one member can be involved in multiple teams in a year.



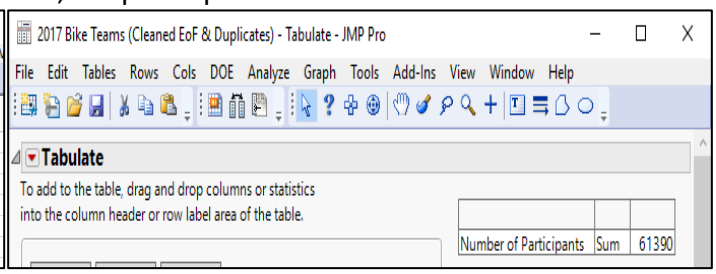
	Member ID	N Rows	N(Member ID)
1	44420811	3	3
2	49023157	3	3
3	49866220	3	3

Team ID	Contact ID	Member ID
519274	2637545	44420811
518015	2637545	44420811
467986	2637545	44420811

From the Bike Teams file, by summarising data by year, we noted that there are 6,146 teams in 2017, and correspondingly there are 61,390 participants in the team file.



	Fiscal Year	N Rows	N(Fiscal Year)
1	2013	6959	6959
2	2014	7014	7014
3	2015	6647	6647
4	2016	6131	6131
5	2017	6146	6146
6	2018	0	0



1.2 Interpretation of Analysis Results

Based on the above, there is a discrepancy of 51,623 participants between both files for FY2017. Additionally, the total number of participants as reported by NMSS could be inflated as one participant can be involved in multiple teams. Furthermore, the total number of participants indicated in the participant file from 2013 to 2017 was 65,499, which is already less than the total number of participants in 2017 at 74,000 as reported by NMSS.

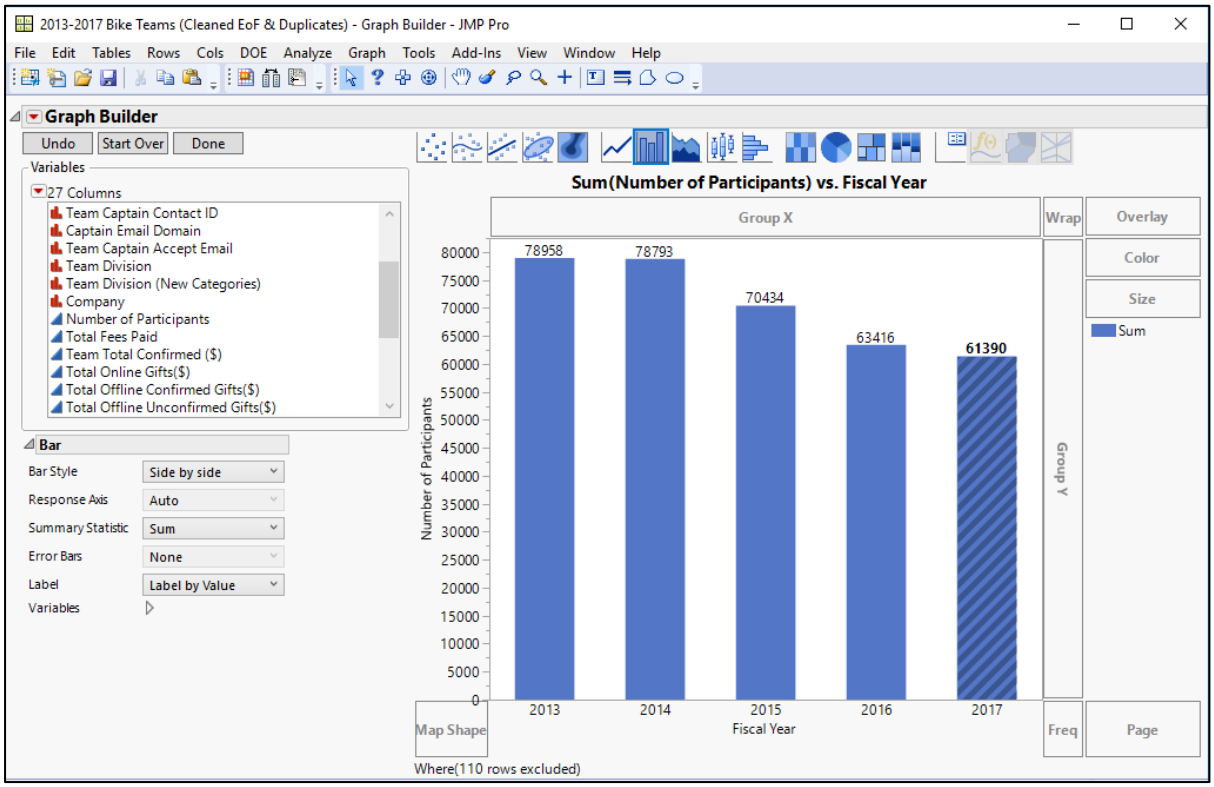
Therefore, the participant file might be incomplete based on the discrepancy in number of types of events indicated in the participant file against the event file. Proper quality control should be established to ensure that all data are complete in the participant file.

However, the number of teams indicated on the Bike Teams file at 6,146 is close to the 6,150 number of teams as reported by NMSS.

Insight 2 – Inaccurate Reported Participant Retention Rate

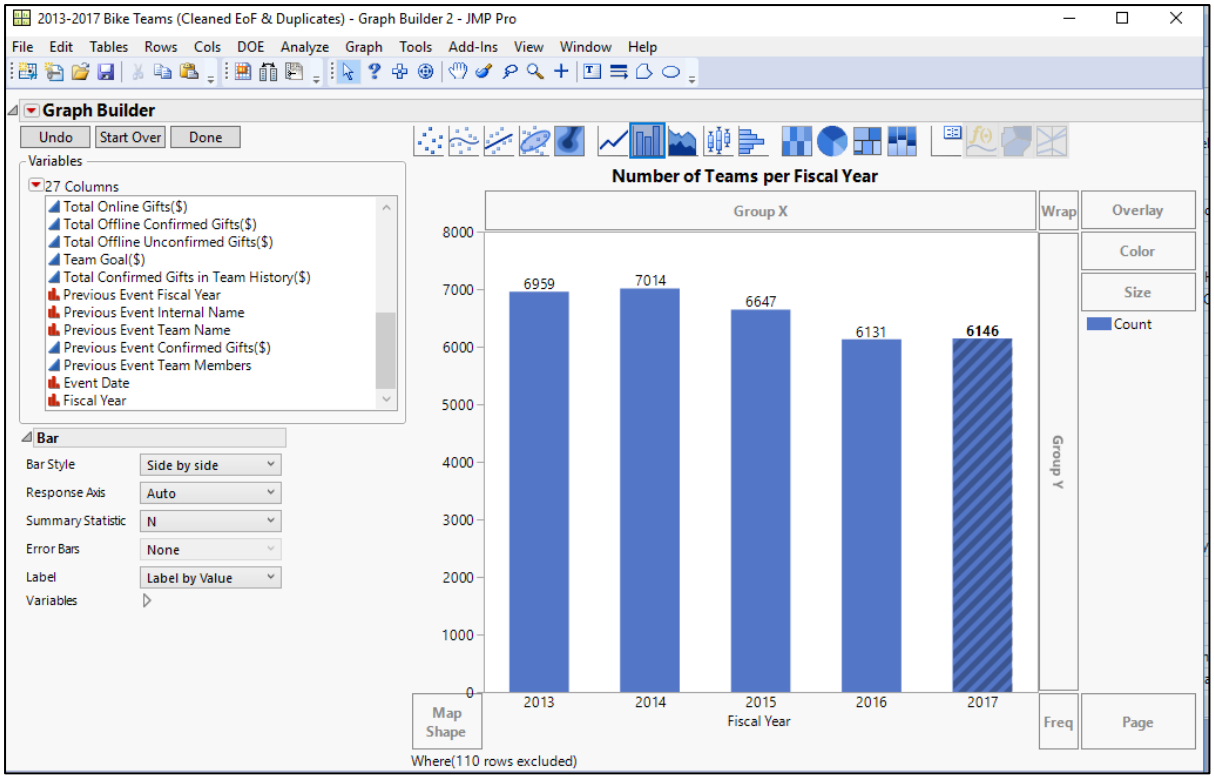
2.1 Data Analysis

According to NMSS, it aims to increase new participants by 40,000 in 2018, which accounts of an 8.9% increase from the reported 74,000 participants.



Based on our analysis of the total number of participants in aggregate, by considering participants who were active in multiple teams, we can see that the total number of participants in 2013 and 2014 were relatively consistent with a minor drop of 0.2% from 2013 to 2014, before significantly decreasing by 19.5% from 2014 to 2016. From 2016 to 2017, the rate of attrition has decreased, but the number of participants is still decreasing by 3.2%.

Additionally, NMSS reported that it aims to increase the total number of teams in 2018 to 6,489, which is an expected goal of 5.5% growth.



Based on the bar chart above, the number of teams increased slightly from year 2013 to 2014. However, like the trend of participations per year, the number of teams began to decrease by 12.6% from year 2014 to 2016. However, from 2016 to 2017, Bike MS has seen a slight increase in the number of teams by 0.2%.

2.2 Interpretation of Analysis Results

Bike MS has seen a substantial decrease in participation over the years. Even though the rate of decline has slowed from 2014 to 2017, the goal of an increase in participants by 8.9% in 2018 would require Bike MS to acquire a substantial number of participants.

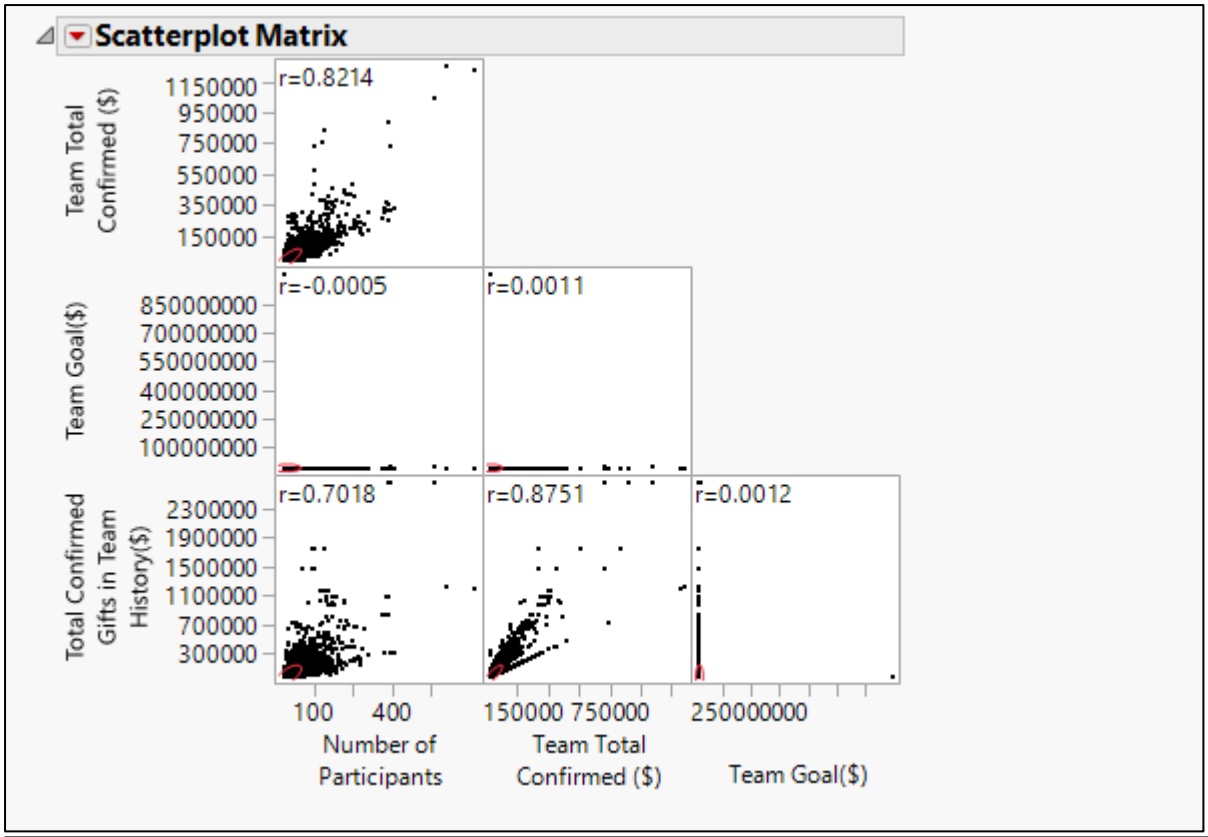
Additionally, even though Bike MS has seen a significant drop in the number of teams from 2014 to 2016, the trend started to reverse, albeit slightly, in 2017. This is in line with NMSS goal of an increase in number of teams. However, NMSS' goal of 5.5% growth in team number from 2017 to 2018 is significantly higher than the actual growth seen in 2017 at 0.2%.

Insight 3 – Correlation between Number of Participants, Donation Received, Team Goal and Amount of Donations Received Historically

3.1 Data Analysis

Next, having understood Bike MS's participant acquisition trend over the years, we are interested to identify if an increase in the number of participants would increase the total amount of donations received.

Additionally, we are interested in identifying other factors which may provide further insights on increasing donation amount.



Pairwise Correlations

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
Total Confirmed Gifts in Team History($)	Team Total Confirmed ($)	0.8751	32897	0.8725	0.8776	<.0001*
Team Total Confirmed ($)	Number of Participants	0.8214	32897	0.8178	0.8248	<.0001*
Total Confirmed Gifts in Team History($)	Number of Participants	0.7018	32897	0.6963	0.7073	<.0001*
Total Confirmed Gifts in Team History($)	Team Goal($)	0.0012	32897	-0.0096	0.0120	0.8267
Team Goal($)	Team Total Confirmed ($)	0.0011	32897	-0.0097	0.0119	0.8403
Team Goal($)	Number of Participants	-0.0005	32897	-0.0113	0.0103	0.9330

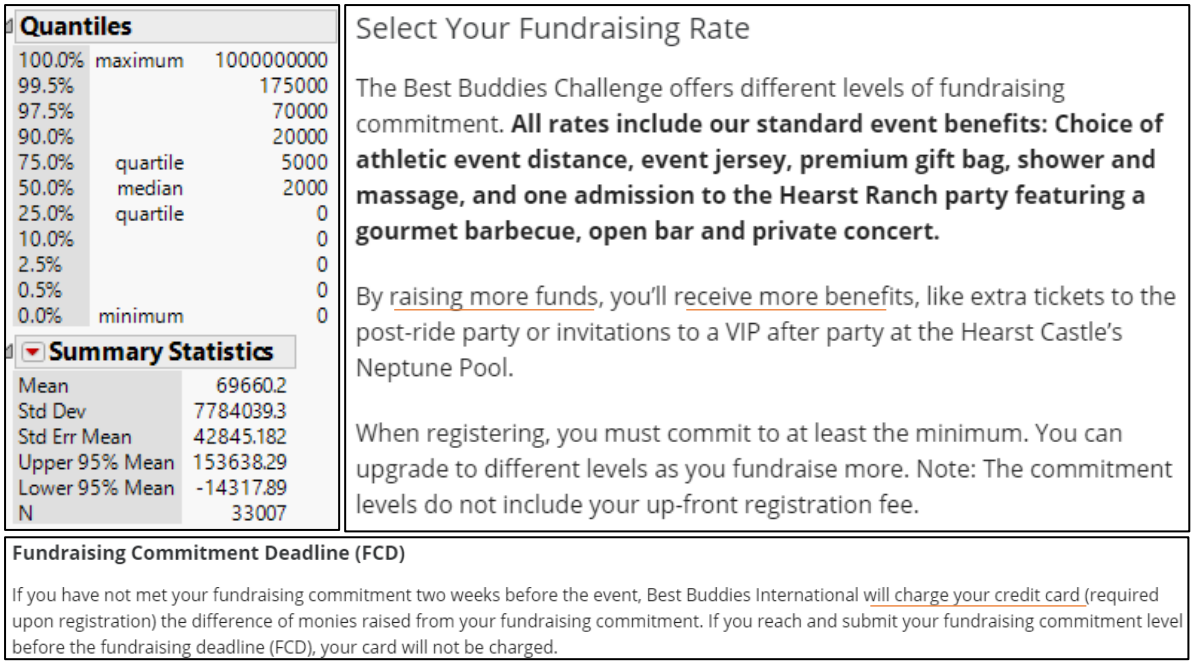
By performing an exploratory Multivariate Correlation Analysis, we can see that historical amount of donation received is highly positively correlated with the team's current amount of donations received. Additionally, we see that the total number of participants does have a relatively high positive correlation to the total amount of donations received.

Conversely, we can see that the amount of team goal has no correlation with the actual amount of donations received historically, nor currently. Additionally, it shows that even with an increase of participants, the amount of team goal does not increase.

3.2 Interpretation of Analysis Results

The above proves that by increasing participation rate, the total amount of donations will also increase.

It is interesting to note that despite allowing teams to set goals, it does not effectively lead to an increase in the total amount of donations received. Based on the distribution analysis below, we can see that there are several abnormally high goal set. Realistic goals should be set to encourage participants to achieve desired donation results.



Quantiles

100.0%	maximum	1000000000
99.5%		175000
97.5%		70000
90.0%		20000
75.0%	quartile	5000
50.0%	median	2000
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	69660.2
Std Dev	7784039.3
Std Err Mean	42845.182
Upper 95% Mean	153638.29
Lower 95% Mean	-14317.89
N	33007

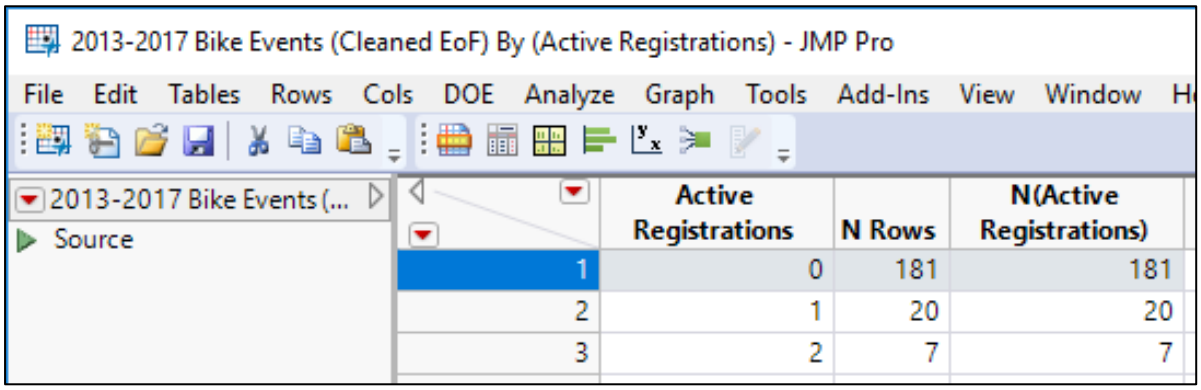
Additionally, based on an excerpt extracted from one of Bike MS' competitors in California, Best Buddies, we can see that they encourage participants to raise more funds by providing benefits. Also, they have established a requirement for their participants to commit to their fundraising goals. Should they fail to meet their goal, they will be charged for the net amount unfulfilled.

Insight 4 – Events and Donations by Region

4.1 Data Analysis

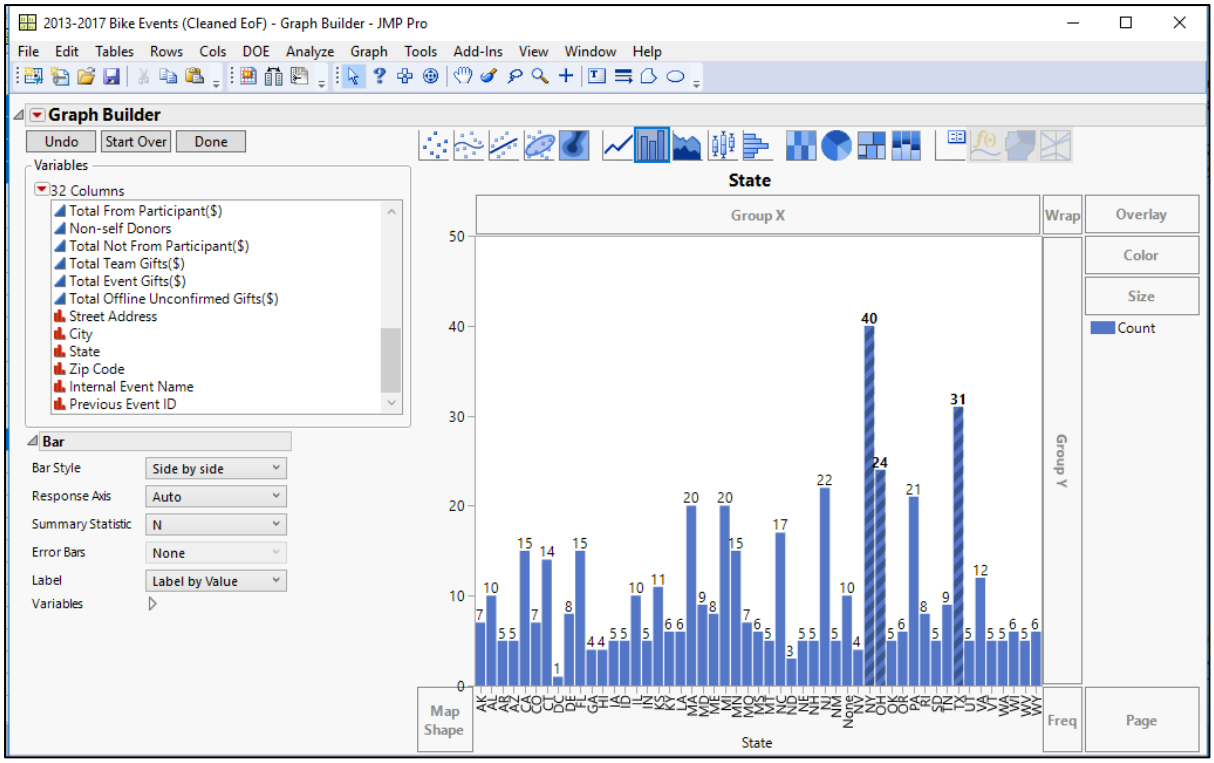
Events by Region

Next, we are interested in identifying the correlation between Events and Donations. By excluding list of events with no active registrations, we can see that there were 519 events held from 2013 to 2017.

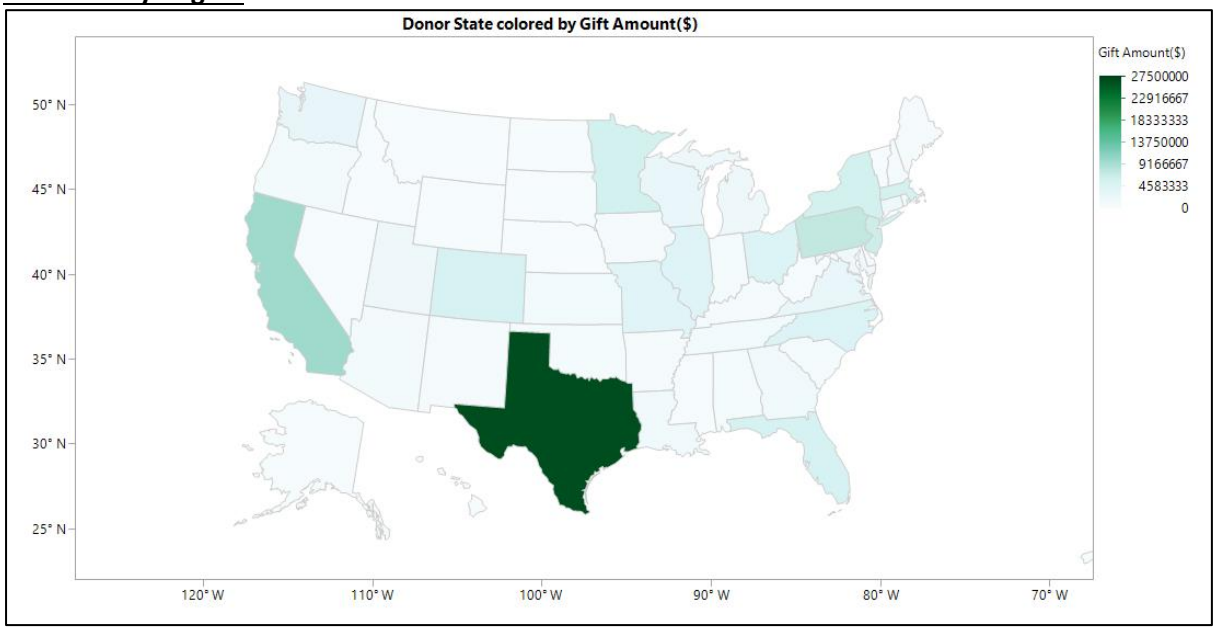


	Active Registrations	N Rows	N(Active Registrations)
1	0	181	181
2	1	20	20
3	2	7	7

From ther Bar Chart below, we can see that from 2013 to 2017, the top three states that holds more events than the other states are New York (NY) at 40 events, Texas (TX) at 31 events, and Ohio (OH) at 24 events. In total, the events held in these three states accounted for 18.3% of the total number of events held from 2013 to 2017.



Donation by Region



Based on the graph above, it shows that from 2013 to 2017, Texas accounts for the highest amount of donations received at $26,697,690.74. The second highest amount of donations received ($9,902,655.09) was from California. The third highest amount of donations received ($7,563,487.92) was from Pennsylvania.

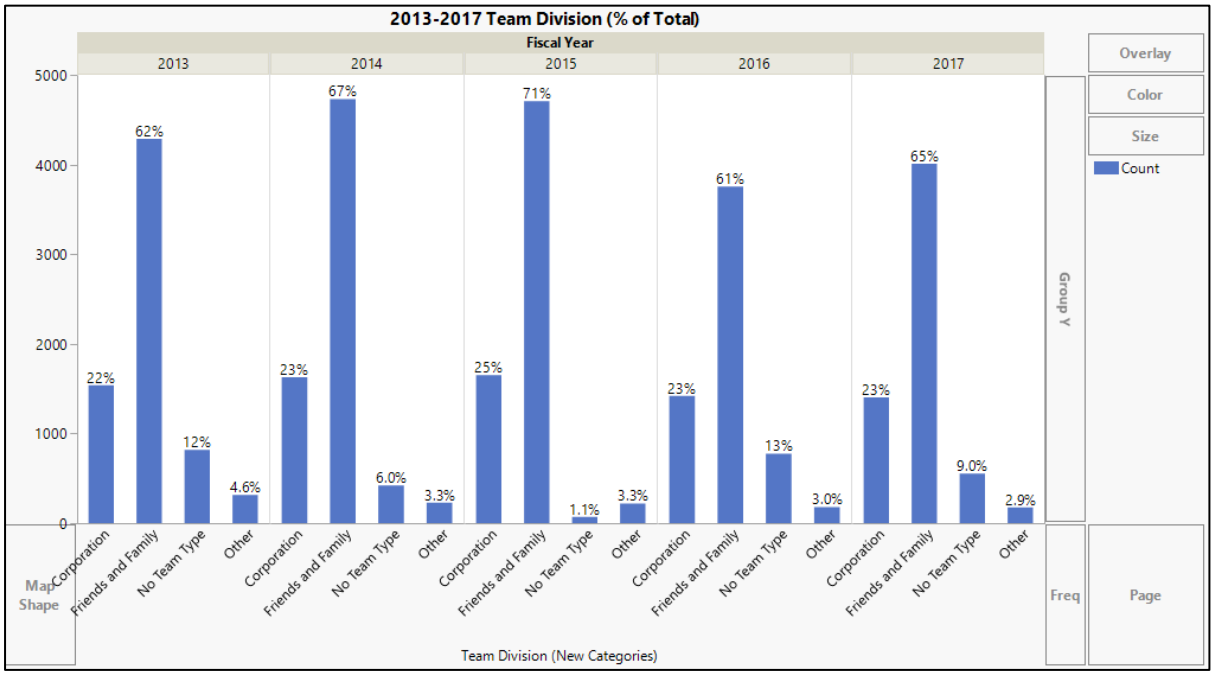
4.2 Interpretation of Analysis Results

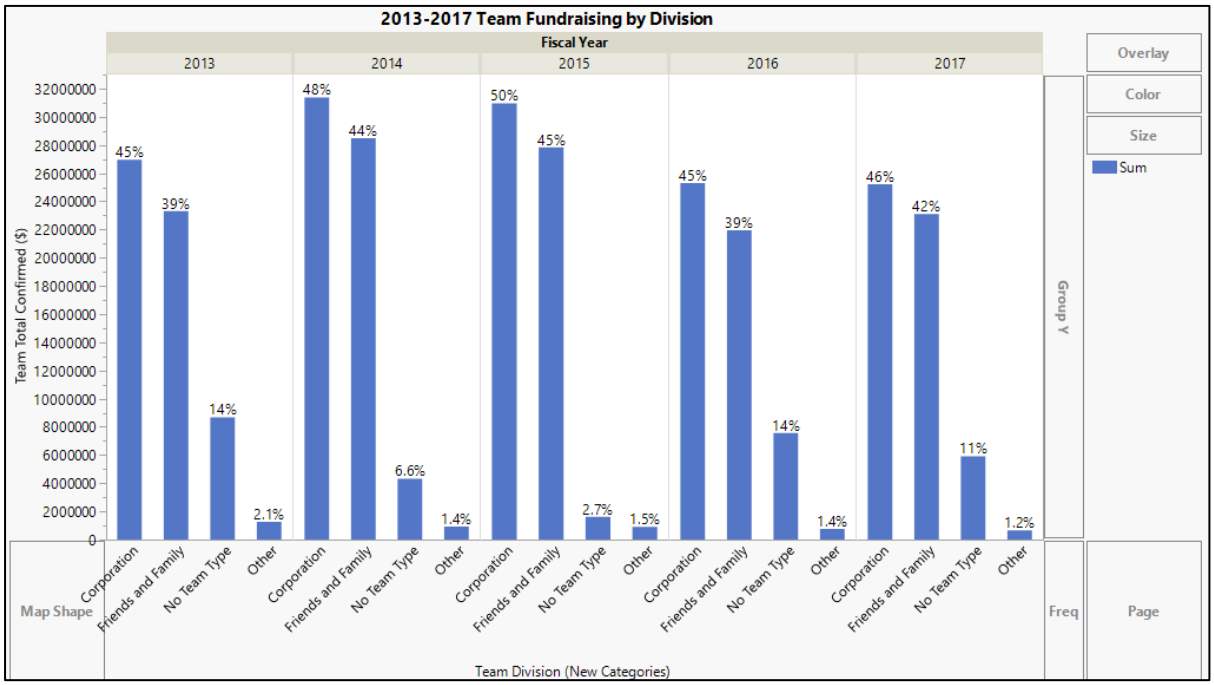
Even though Texas holds the second highest number of events, it received the highest amount of donations. Additionally, even though most of the events are held in New York, and the third highest number of events were held in Ohio, California and Pennsylvania accounts for the second and third highest amount of donations received respectively.

Insight 5 – Fundraising Effort by Corporation

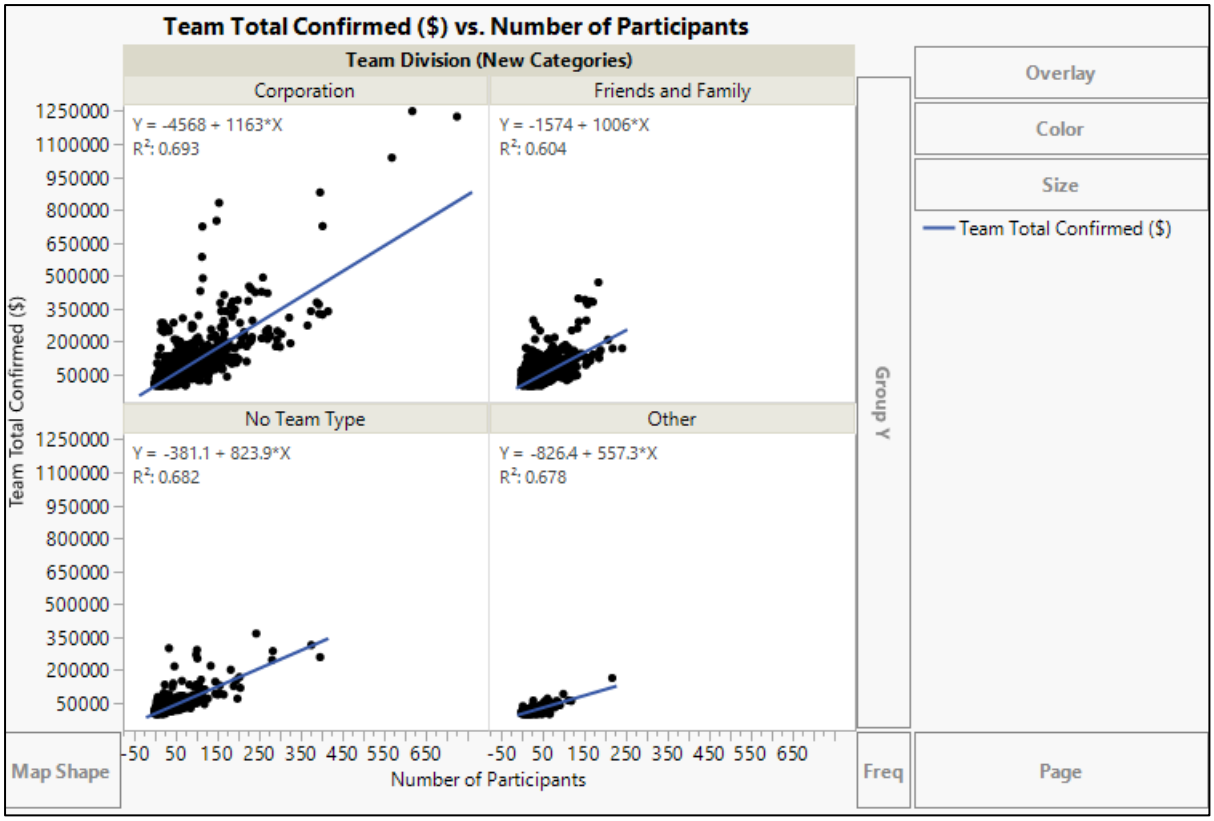
5.1 Data Analysis

The following bar charts depicts the percentage participation team division type and the amount of fund raised by each division category:

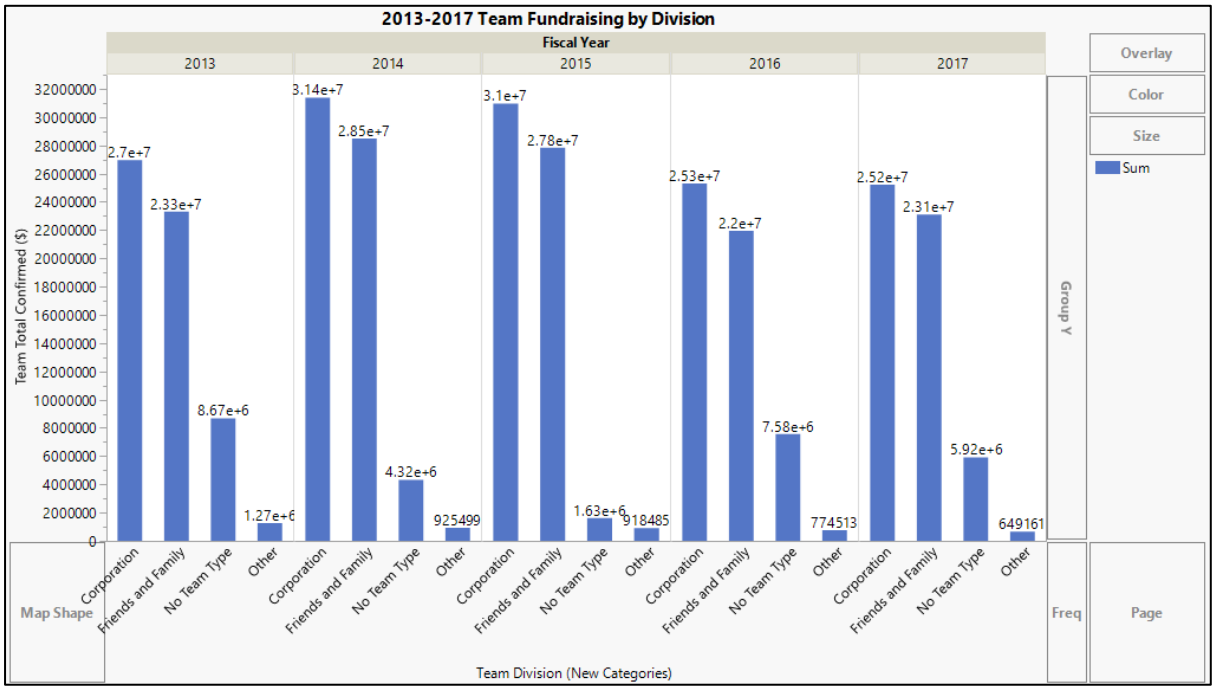




By analysing the total amount of fund raised, we can see that over the years, corporations have been contributing the most in each year, despite representing only about 22% to 25% of the team division.

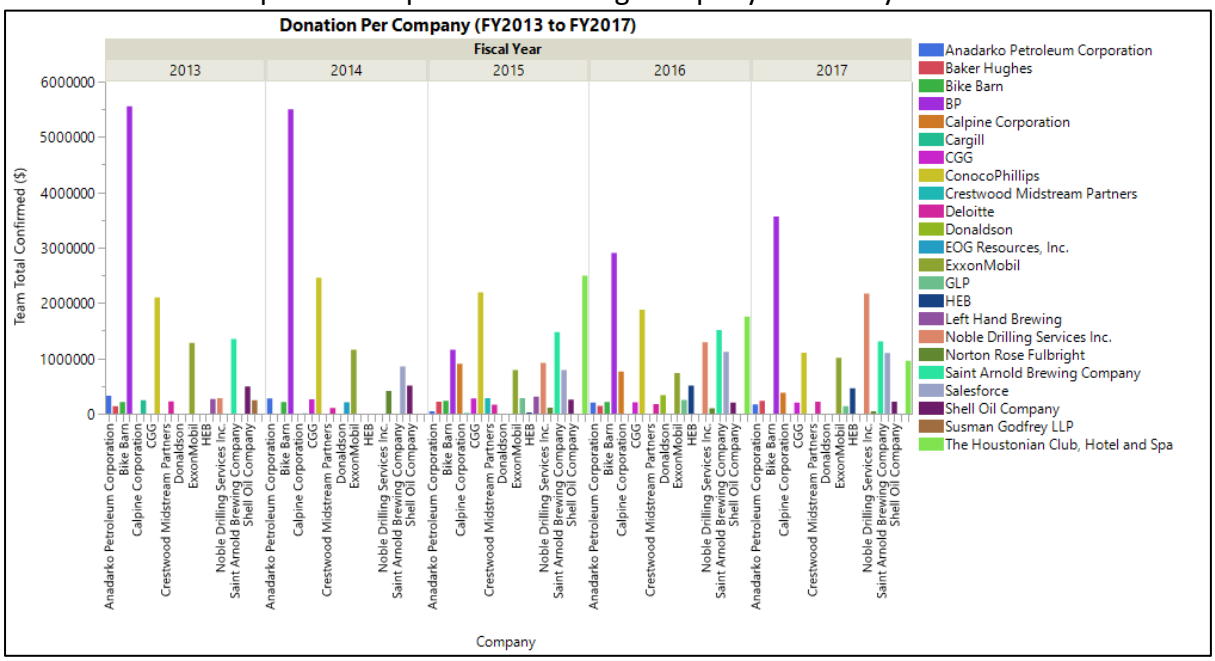


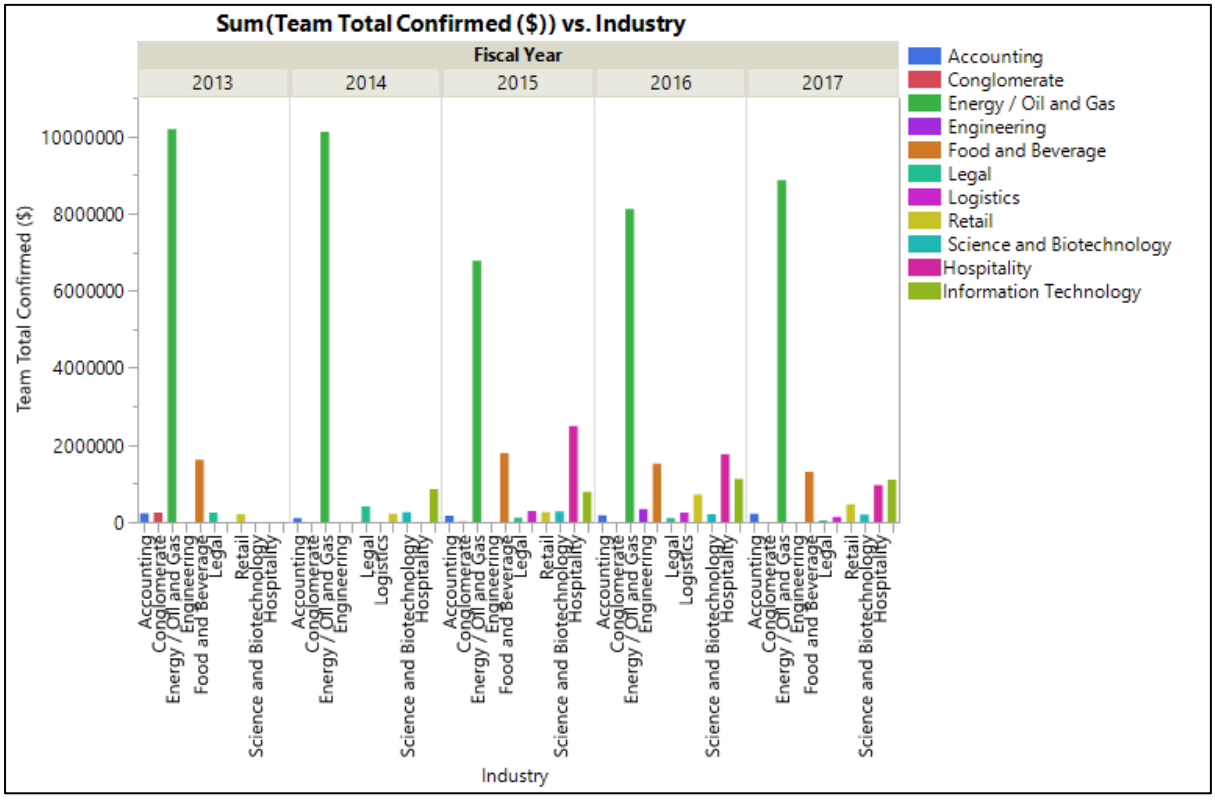
Based on the above, we can confirm our analysis that the number of corporate participants is moderately highly correlated with the total amount of donations received. Additionally, we can see that corporate participants show a higher number of significant outliers that contribute large amount of donations, resulting in a higher total amount of donations received when compared with the other divisions.



However, it is important to note that in 2017, the amount of fund raised has decreased significantly both in absolute amount by corporations, as well as by percentage total of fund raised by all divisions in 2017.

Next, we are interested in identifying the top participating corporations that have managed to raise the most amount of fund over the years. By selecting rows only from each year and summarising the data by “Company” and “Team Total Confirmed ($)”, we are able to identify the top 10 contributing Company for each year. As some of the fields in the Company column are empty, we have identified the team name to locate the team page on the Bike MS webpage to identify the name of the corporation of which the team represents. The following bar charts below depicts the top 10 contributing Company for each year:





Based on the bar charts above, we can see that British Petroleum (BP) has consistently been the top contributor to the team total confirmed ($) from 2013 to 2017, with the exception of 2015. Additionally, we can see that despite seeing a decrease of 79.1% in BP’s team total

confirmed ($) from 2013 to 2015, BP has shown a significant increase of 206.8% from 2015 to 2017.

Next, by analysing team total confirmed ($) by industry type, we can see that the Energy / Oil and Gas Industry has consistently been the highest contributor to the team total confirmed over the years, which is consistent with the finding above, where BP is the top contributor to the team total confirmed ($).

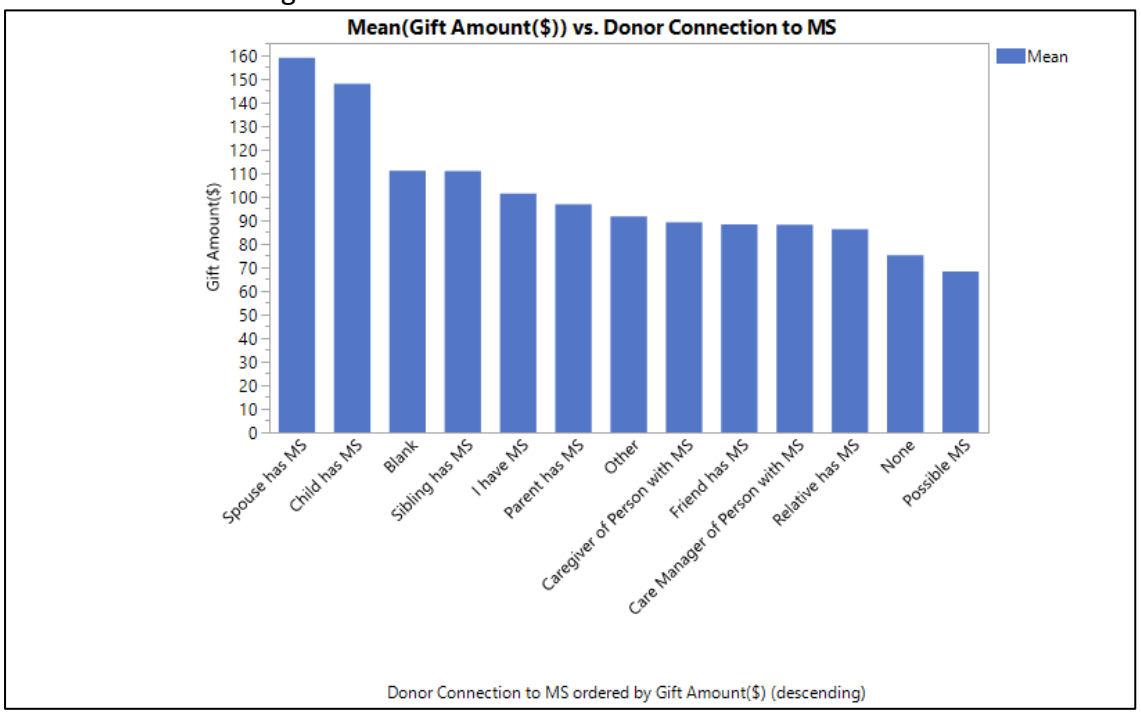
5.2 Interpretation of Analysis Results

Despite representing a small percentage of the total team division type, corporations have generally been contributing most to the fundraising effort. Additionally, we noted that most of the corporations are from the Energy / Oil and Gas Industry

Insight 6 – Analysis of Donation Amount and Donor's Connection to MS

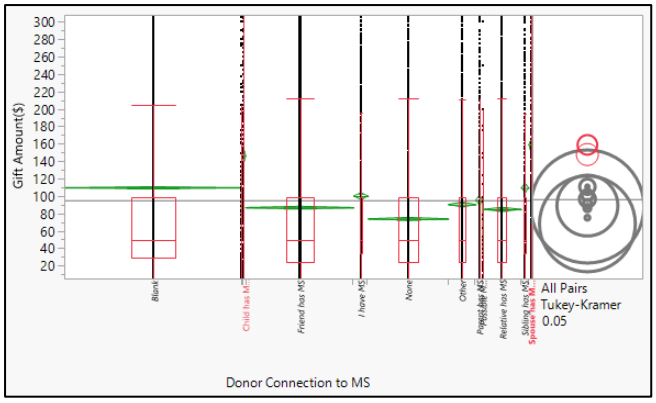
6.1 Data Analysis

Next, we are interested in identifying the profile of the donors, to understand the factors which would encourage them to donate to Bike MS.



Based on the bar chart above, we can see that donors are more likely to donate a higher amount when their spouse or child has MS.

By performing a One-way ANOVA analysis with All Pairs, Tukey HSD test based on the variables "Gift Amount" and "Donor's Connection to MS", we can confirm that the above finding is true.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Donor Connection to MS	12	325026024	27085502	176.7654	<.0001*
Error	1315719	2.0161e+11	153228.57		
C. Total	1315731	2.0193e+11			

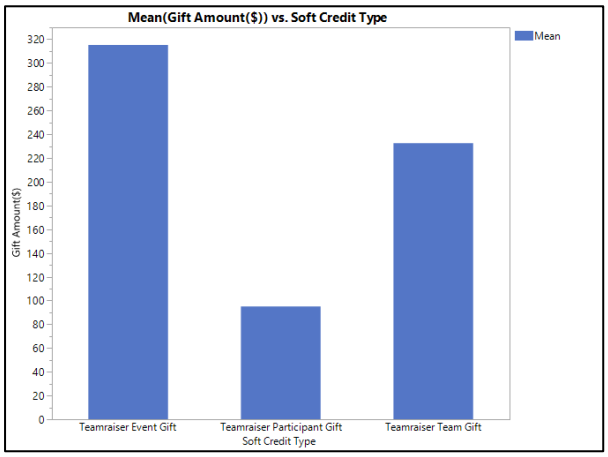
Means for Oneway Anova

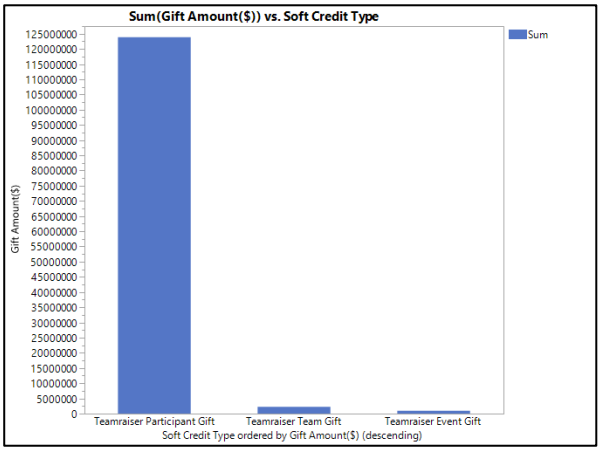
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Possible MS	556	68.122	16.601	35.59	100.66
None	225707	75.137	0.824	73.52	76.75
Relative has MS	107415	86.029	1.194	83.69	88.37
Care Manager of Person with MS	399	88.025	19.597	49.62	126.43
Friend has MS	300368	88.095	0.714	86.70	89.49
Caregiver of Person with MS	1378	89.079	10.545	68.41	109.75
Other	76611	91.562	1.414	88.79	94.33
Parent has MS	20011	96.624	2.767	91.20	102.05
I have MS	40729	101.255	1.940	97.45	105.06
Sibling has MS	21455	110.762	2.672	105.52	116.00
Blank	497479	110.875	0.555	109.79	111.96
Child has MS	10450	147.705	3.829	140.20	155.21
Spouse has MS	13174	158.801	3.410	152.12	165.49

Std Error uses a pooled estimate of error variance

Donor's Preferred Choice to Donate on Behalf Of

Based on the bar chart below, we can see that on average, donors tend to donate a higher amount to Events per person. However, most of the donations are made to participants.





6.2 Interpretation of Analysis Results

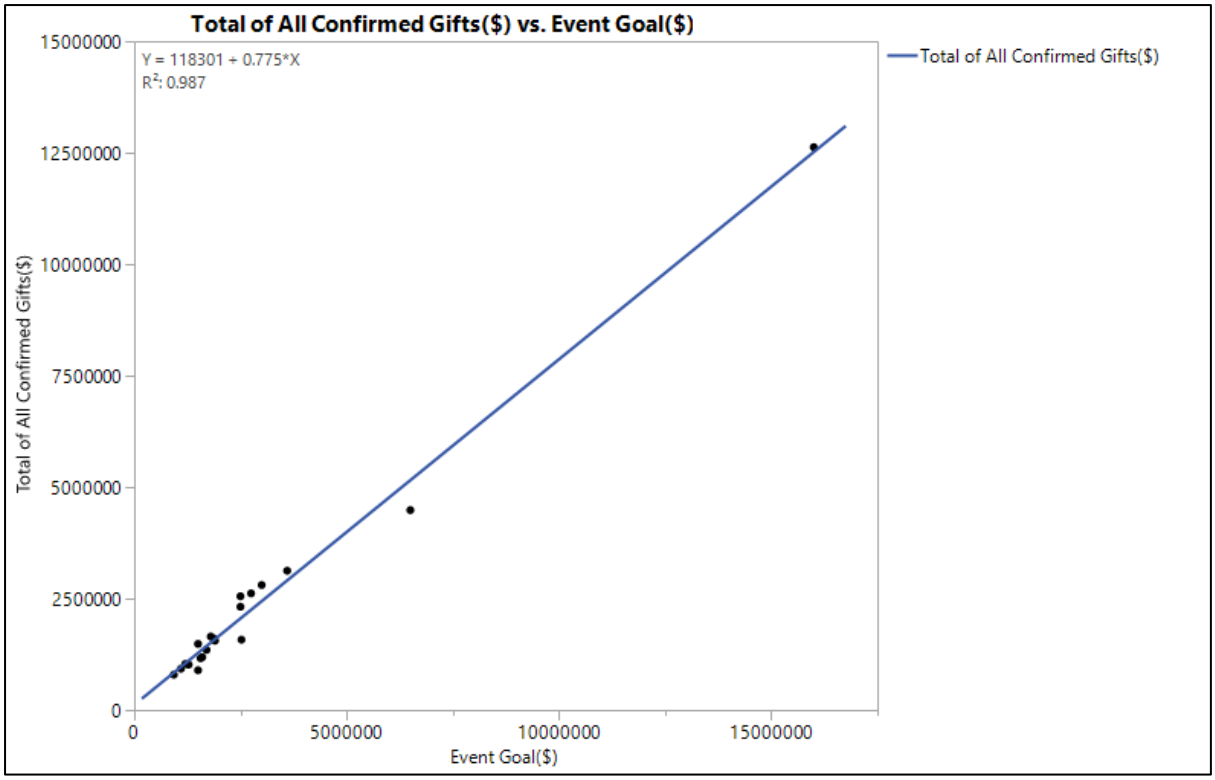
Based on the above, we can see that when one has a greater propensity to donate a larger amount on average to Bike MS when the donor has a close relationship with someone who has MS, more than when he/she is the one with MS or has possibility of MS.

Additionally, most of the donation efforts are made in support of participants of Bike MS.

Insight 7 – Bike MS Top 20 Events

7.1 Data Analysis

NMSS has established a 2018 budget where the following top 20 rides are aspiring to achieve. Accordingly, we are interested in identifying if the top 20 rides have been able to meet their targeted goal in 2017 by analysing the correlation between their event goal and their actual confirmed gifts.



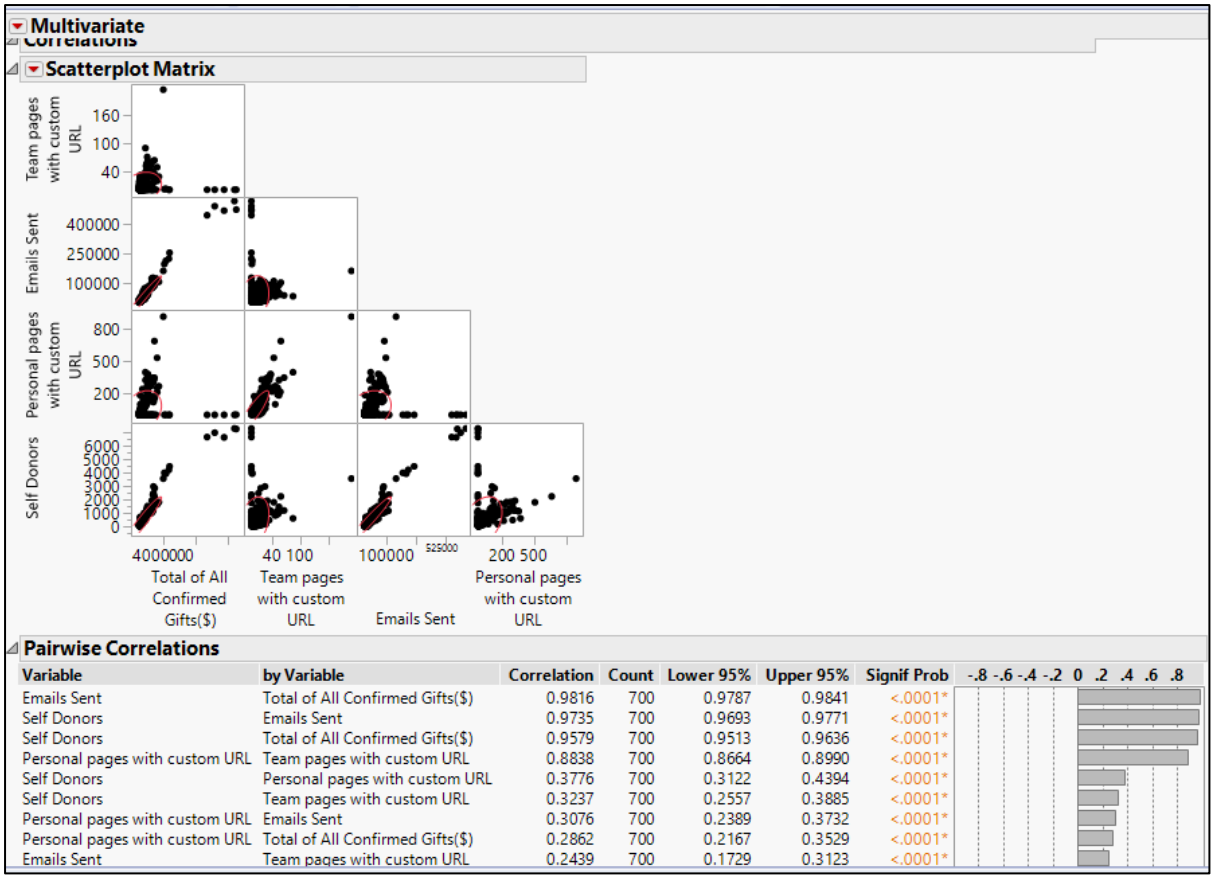
7.2 Interpretation of Analysis Results

Based on the above, we can see that in general, both variables indicate a strong correlation, where all rides have been able to, or close to meeting their goals.

Insight 8 – Correlation Between Positive Fundraising Behaviour and Amount of Fund Raised

8.1 Data Analysis

According to NMSS, participants are "more likely to satisfy fundraising minimums than others", when they use email tool, update personal page and self-donate. Therefore, by performing a multivariate analysis on the above-mentioned variables, we have derived with the following results:



Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
Emails Sent	Total of All Confirmed Gifts($)	0.9816	700	0.9787	0.9841	<.0001*
Self Donors	Emails Sent	0.9735	700	0.9693	0.9771	<.0001*
Self Donors	Total of All Confirmed Gifts($)	0.9579	700	0.9513	0.9636	<.0001*
Personal pages with custom URL	Team pages with custom URL	0.8838	700	0.8664	0.8990	<.0001*
Self Donors	Personal pages with custom URL	0.3776	700	0.3122	0.4394	<.0001*
Self Donors	Team pages with custom URL	0.3237	700	0.2557	0.3885	<.0001*
Personal pages with custom URL	Emails Sent	0.3076	700	0.2389	0.3732	<.0001*
Personal pages with custom URL	Total of All Confirmed Gifts($)	0.2862	700	0.2167	0.3529	<.0001*
Emails Sent	Team pages with custom URL	0.2439	700	0.1729	0.3123	<.0001*

8.2 Interpretation of Analysis Results

There is high correlation between the number of emails sent by participants, number of self-donors and the actual total amount of confirmed gifts.

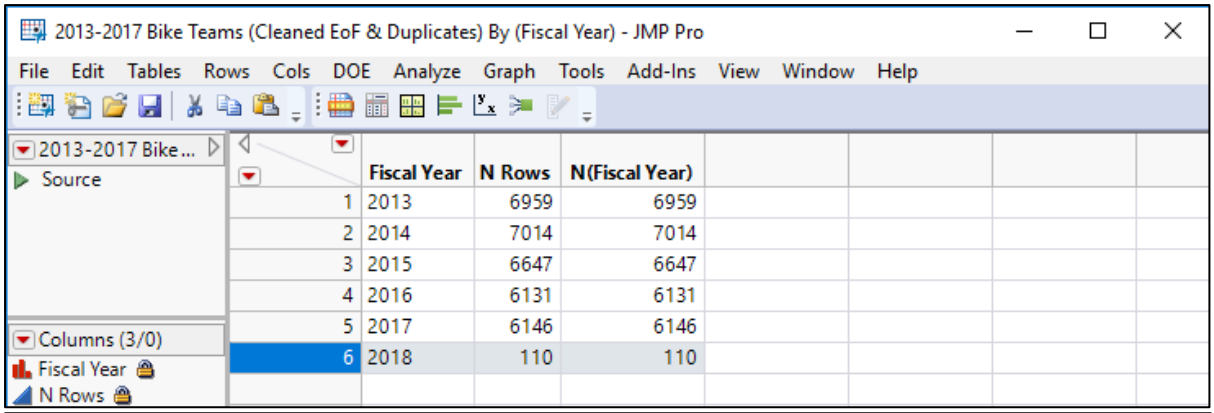
However, the results show that even if participants update their personal and team pages, there is no strong correlation with the total amount of confirmed gifts.

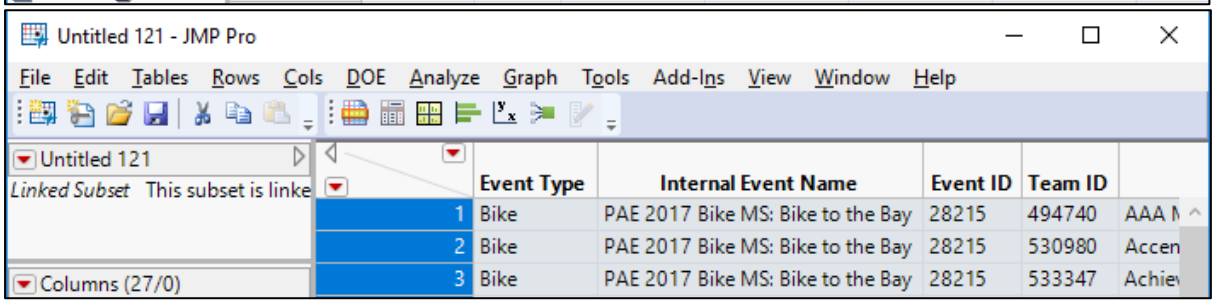
Insight 9 – Discrepancy of Total Offline Unconfirmed Gift Amount Between Bike Team and Bike Event Files

9.1 Data Analysis

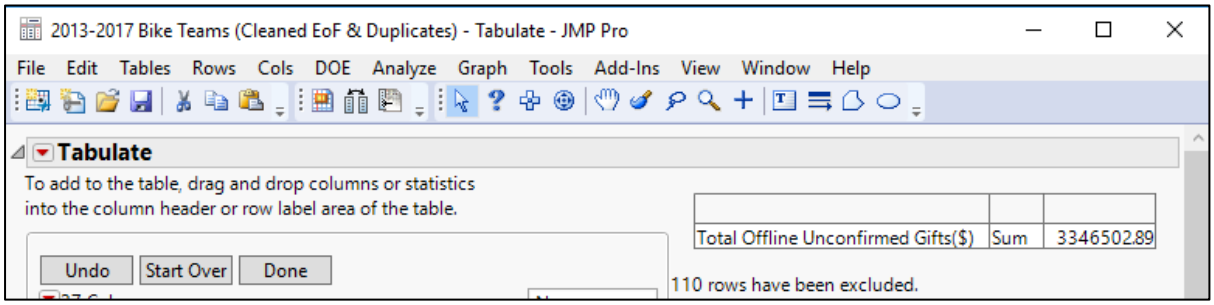
According to the field definitions in team, it states that the total offline unconfirmed gifts refers to the “amount of team’s donations received offline but never received by NMSS”. Additionally, the total offline unconfirmed gifts in events refers to “the sum of offline (cash/check) donations reported by participants, but never received by NMSS. Therefore, the total amount of offline unconfirmed gifts in the bike teams file should reconcile with the total amount of offline unconfirmed gifts in the bike events file.

As the Team file includes 110 FY 2018 rows that includes event ID 28215, which does not appear in the events file, we need to hide and exclude these 110 rows.

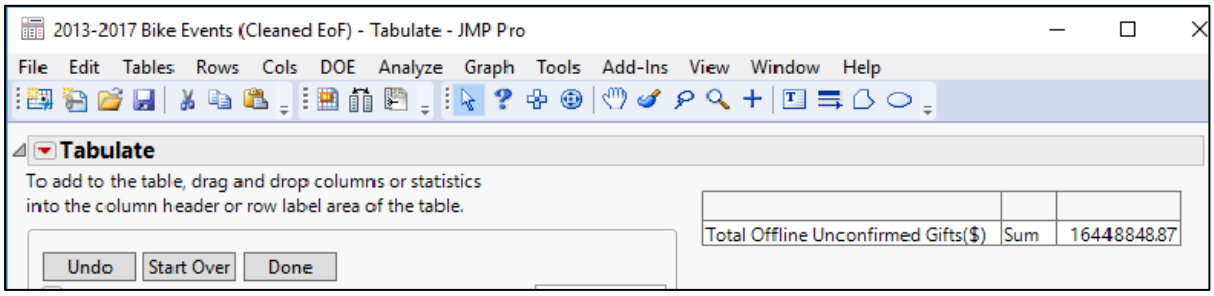




By tabulating the sum of total offline unconfirmed gift in team file from 2013 to 2017, the total amounted to $3,346,502.89

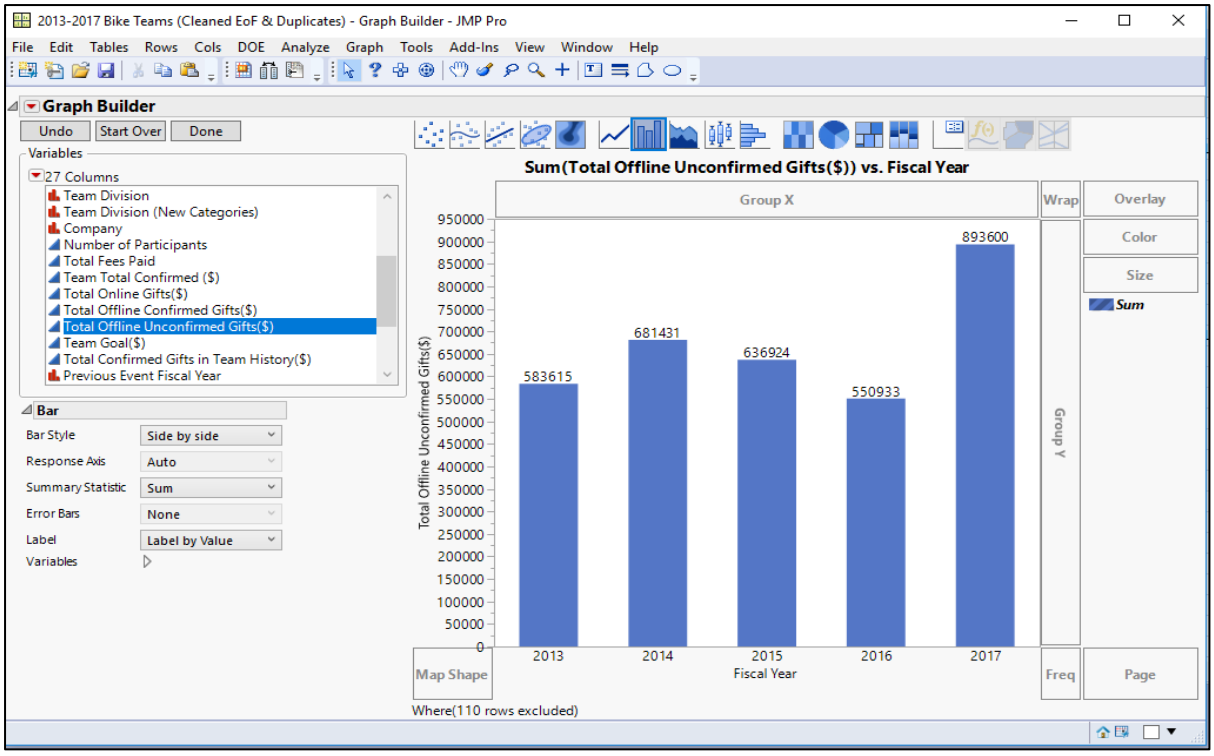


By tabulating the sum of total offline unconfirmed gift in events file from 2013 to 2017, the total amounted to $16,448,848.87

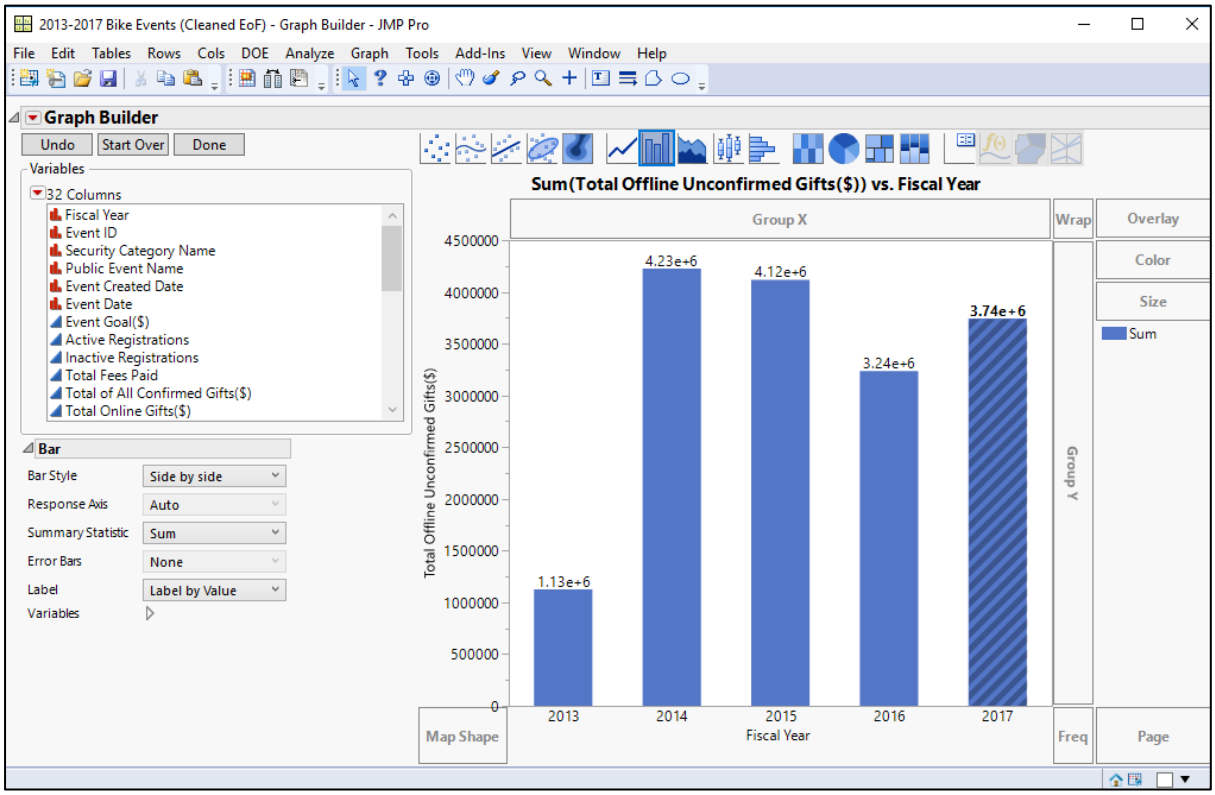


This shows a stark difference of $13,102,345.98, which is 391.5% of the amount in bike teams, and 79.7% of the amount in bike events.

By further analysing Total Offline Unconfirmed Gift in the Team file, we noted that trend increased briefly from 2013 to 2014 by 16.8%. From 2014 to 2016, trend decreased by 19.2% before increasing significantly by 62.2% between 2016 and 2017.



By analysing the same in the event file, we noted that trend increased significantly from 2013 to 2014 by 274.3%. From 2014 to 2016, trend decreased from by 23.4% before increasing again from 2016 to 2017 by 15.4%.



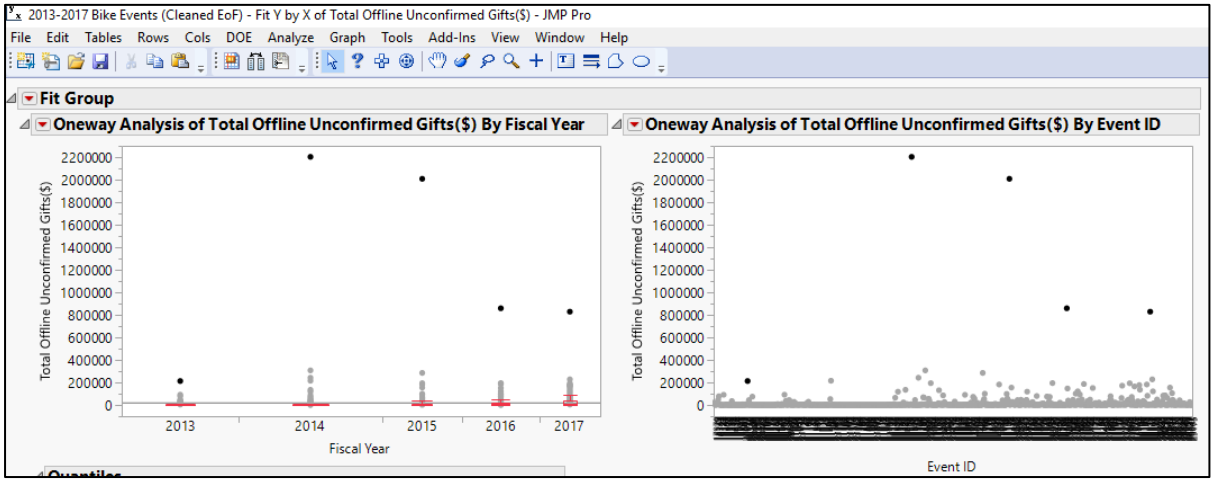
9.2 Interpretation of Analysis Results

The difference could mean that the consolidation of offline unconfirmed gift for all teams in each event are underreported, or that the total offline unconfirmed gift for events have been overreported.

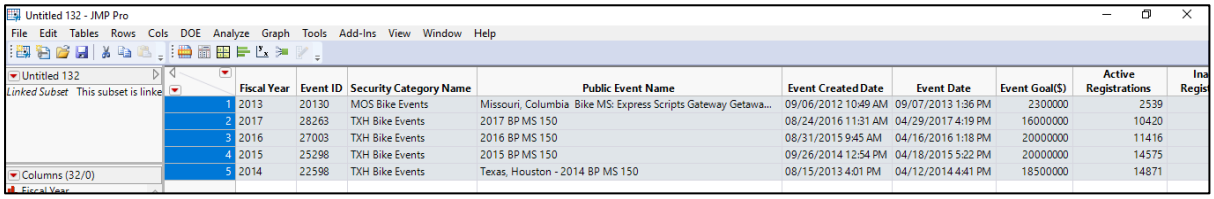
Insight 10 – Analysis of Highest Contributor to Total Offline Unconfirmed Gift by Events

10.1 Data Analysis

Since Events file shows a significantly higher amount of total offline unconfirmed gifts, further analysis should be performed to identify the event ID contributing the highest amount per year. Upon identification of the Event ID, further analysis can be performed to identify the total unreconcilable amount against the team file based on a Oneway Analysis of Total Offline Unconfirmed Gifts by Fiscal Year and by Event ID.

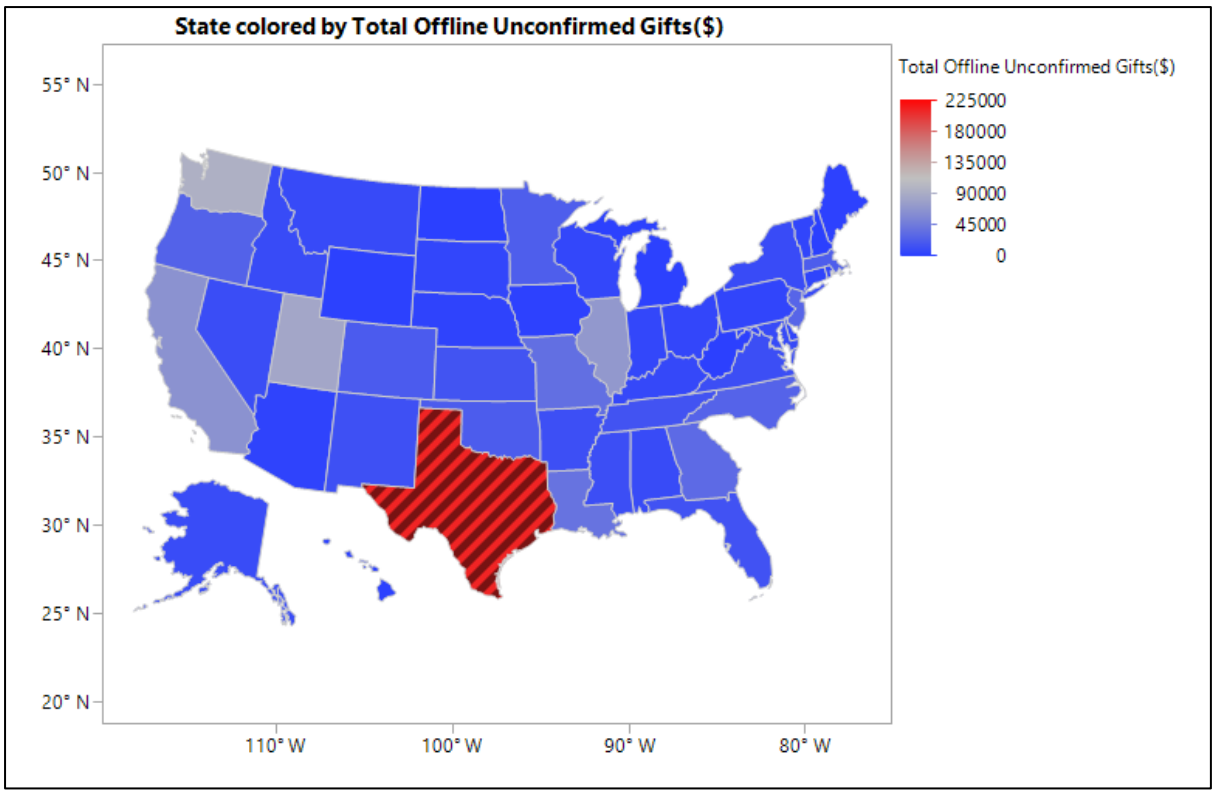


The following details the total offline unconfirmed gifts ($) in descending order per event ID and Fiscal year:



	Fiscal Year	Event ID	Security Category Name	Public Event Name	Event Created Date	Event Date	Event Goal($)	Active Registrations	Ina Regis
1	2013	20130	MOS Bike Events	Missouri, Columbia Bike MS: Express Scripts Gateway Getawa...	09/06/2012 10:49 AM	09/07/2013 1:36 PM	2300000	2539	
2	2017	28263	TXH Bike Events	2017 BP MS 150	08/24/2016 11:31 AM	04/29/2017 4:19 PM	16000000	10420	
3	2016	27003	TXH Bike Events	2016 BP MS 150	08/31/2015 9:45 AM	04/16/2016 1:18 PM	20000000	11416	
4	2015	25298	TXH Bike Events	2015 BP MS 150	09/26/2014 12:54 PM	04/18/2015 5:22 PM	20000000	14575	
5	2014	22598	TXH Bike Events	Texas, Houston - 2014 BP MS 150	08/15/2013 4:01 PM	04/12/2014 4:41 PM	18500000	14871	

Except for 2013, the top contributing events are held in Texas, as confirmed by the map below:



10.2 Interpretation of Analysis Results

We noted that except for 2013, the top contributing events for total offline unconfirmed gifts from 2014 to 2017 are from the same event held annually (TXH Bike Events – BP MS 150 (2014 to 2017)), held in Texas. Even though the annual BP MS 150 Bike event has decreasing total offline unconfirmed gifts from 2014 to 2017, it remains the highest contributing event per year towards the total offline unconfirmed gift amount. Given the fact that offline gifts are donated in form of cash or checks, the consistently high amount of unreconcilable offline unconfirmed gift on a yearly basis could point towards a possible internal control deficiency in cash/cheque handling.