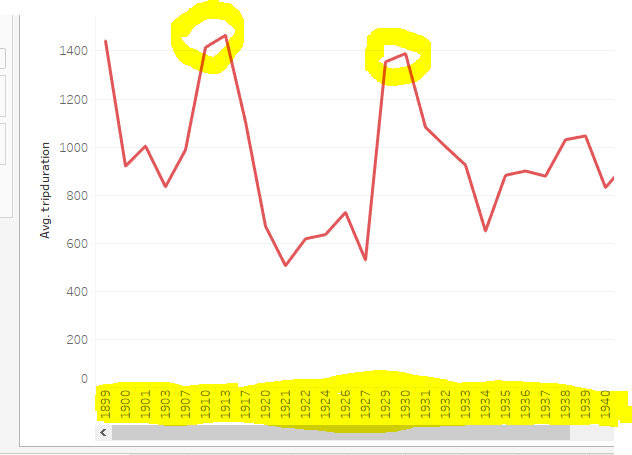
**The Misuse of Data**

**(See Tableau Story Page 1)**

Gentlemen, it is important to stay well informed; to not be mislead by a half-truth statement given to sway one’s mind in any given direction (buy a product, use a service, vote for a candidate, etc). This statement certainly holds throughout history, but as Americans living amidst a particularly volatile social and political environment, this statement of staying well informed seems to be more important now than ever.

The graph we see here (reference tableau story page 1) shows the average duration of bike trips taken per age group from August 2013 through December.

Directly on this word document below, I would like to draw attention to the volatility of average trip duration for those bikers born before 1940. Much more volatility than those born in subsequent decades. This graph can potentially be misleading to an uniformed viewer…



(**See Tableau Story Page 2)**

A salesman promoting bike trips hands you a brochure about biking with the colorful graph on story page 2, and tells you "the data clearly shows how bikers here born 1930 and before bike for longer than our young guns on the trails here," (a blatant half-truth). Back to my original point, it would be important to doubt the validity of something that sounds as strange as “30 year olds bike on average less time than those born in 1930 and earlier.”

**(See Tableau Story Page 3)**

We can see by this bar graph that the majority of data points for bikers’ birth years were collected for bikers born after 1940 by several orders of magnitude. With millions of data points collected, the average duration of bike trips for younger bikers normalized, whereas with a much smaller amount of data collected for elderly bikers (born before 1940), a much greater amount of volatility is apparent in the data for their average duration of biking time per trip.

Interesting to note also, how the amount of biking trips declined as evidence by the bar chart on bottom from August to December. From this part of the data we can infer biking is more popular in warmer months than colder months in and around Manhattan.

**Gender Discrepancies**

**(See Tableau Story Page 4)**

We can see from the Pie chart there were far more males than females in the data, 76% male 24% female, (with null values, those who did not disclose gender, not being counted in the graphs). Interesting to note, however, that despite lower participation, females (on average) took longer trips than males consistently from August through December. One would have thought with higher numbers of participation, males would bike for longer amounts of time and further distances, but the data clearly indicates otherwise.

**(See Tableau Story Page 5)**

Do males and females have varying biking practices? Map density shows that, when scaled, males and females began their biking trips at similar stations in numbers mostly around the center of the city, rather than outlying areas like Brooklyn. The data from map shows that neither males nor females particularly favored or disfavored any particular area of the island or part of Brooklyn on the mainland.

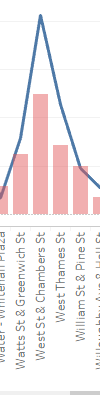
**(See Tableau Story Page 6)**

Likewise, the datapoints collected for males and females follow a similar distribution for age groups. From this we can infer that biking is more popular based on age, rather than gender.

**(See Tableau Story Page 7** men represented with red bar chart, women blue line graph)

Proportionally speaking, with two notable exceptions shown just below,





men and woman proportional to their total counts, start their biking journeys off at relatively similar places. Two notable exceptions to this are at West and Chambers Street, where women proportionally make up a large amount of the riders given their numbers, and conversely at 8th Ave and West 31st St men even given their numbers comprise a disproportionate amount of the riders.

It is also important to understand not just the proportion of riders, but how that proportion then translates into real numbers of trips for men and women. In the first graph above, where men proportionally comprise double the values on the graph, given men comprise triple the amount of data points, in real life terms this means roughly six times the amount of men begin their journeys from 8th Ave and West 31st.

**See Tableau Story Page 8**

Per the data, from August 2013 through December of that year, over 99% of those who reported their gender are also subscribers. It was only customers, those using temp passes, who did not disclose gender for the most part.

Out of millions of data points, only 17 ‘customers’ reported to be males (i.e. non-subscribers who reported their gender). This data shows how a lack of participation amongst users can make it more difficult to see gender discrepancies between non-subscribers. Also note how customers decline by a much larger percentage than subscribers do from August to December. This could potentially indicate subscribers live locally in New York, although more research would be needed to conclude that fully.

**SUMMARY;**

Part 1; Data can be used to mislead people, it is important to stay skeptical and think critically about what data is really saying and its validity.

Part 2; Gender discrepancies for biking were less than anticipated. Both genders showed equal amounts of subscriptions proportionally, and used the same stations proportionally as one another. However, where male and females differed is that three times the number of men used the bikes from August through December meaning men comprised roughly 76% of the data when those who did not disclose gender are taken out of the equation.