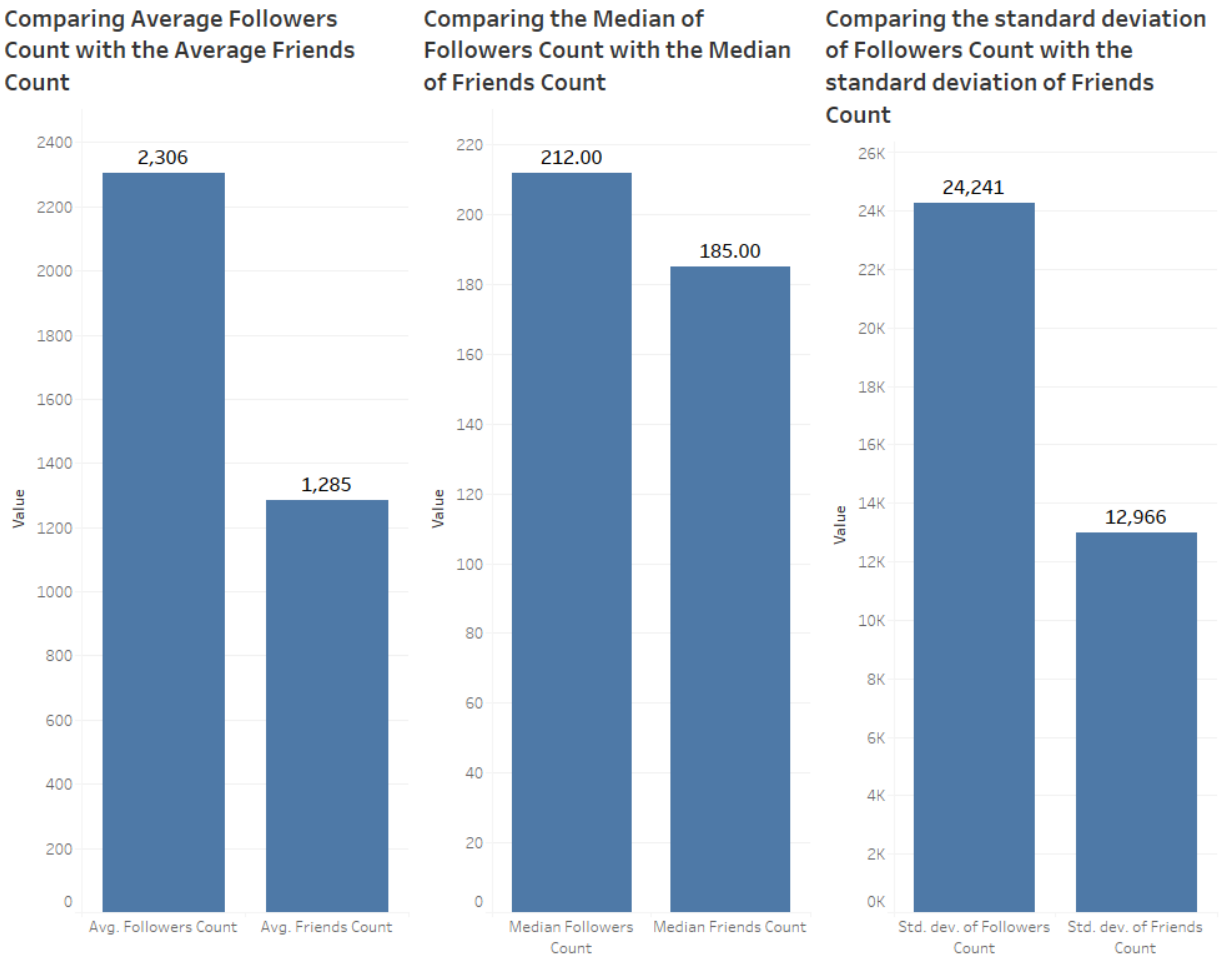
Assignment 3 **Mahdieh Asl Javadian**

The link to my Tableau Public:
https://public.tableau.com/app/profile/mahdieh2210/viz/BDAclass_Mahdie/Nov2015Tweets?publish=yes

Question 1



As we can see in the bar charts, the Standard Deviation of "followers count" is about 10 times greater than the Average of "followers counts"; this means that this data is dramatically skewed. To be more specific, it seems that a few numbers of the samples in this database, have a high number of followers (maybe they are bloggers, celebrities, or athletes …), so they have increased the Variance and subsequently the Standard Deviation. In addition to this, the Median for the "followers count" in 212, meaning that 50% of the samples have a follower count less than 212, and the other half are more than this amount. This means that in this database, most of the samples have a few numbers of followers, and even in the presence of the high follower samples, the Average of followers is 2306, attesting to this fact that low count followers are much more frequent than high count followers. Therefore, the data is drastically skewed having some limited samples that have increased the Standard Deviation, but they could not affect the Average so much, as they are limited.

Theses interpretations are also true for the friends count. In order for comparing the “friends count” and the “followers count”, “followers count” have higher numbers than “friends count”, and it is due to the fact that people may be followed by others, but they may reject some friendship requests, so the number of friends are always less than the number of followers in all three sets.

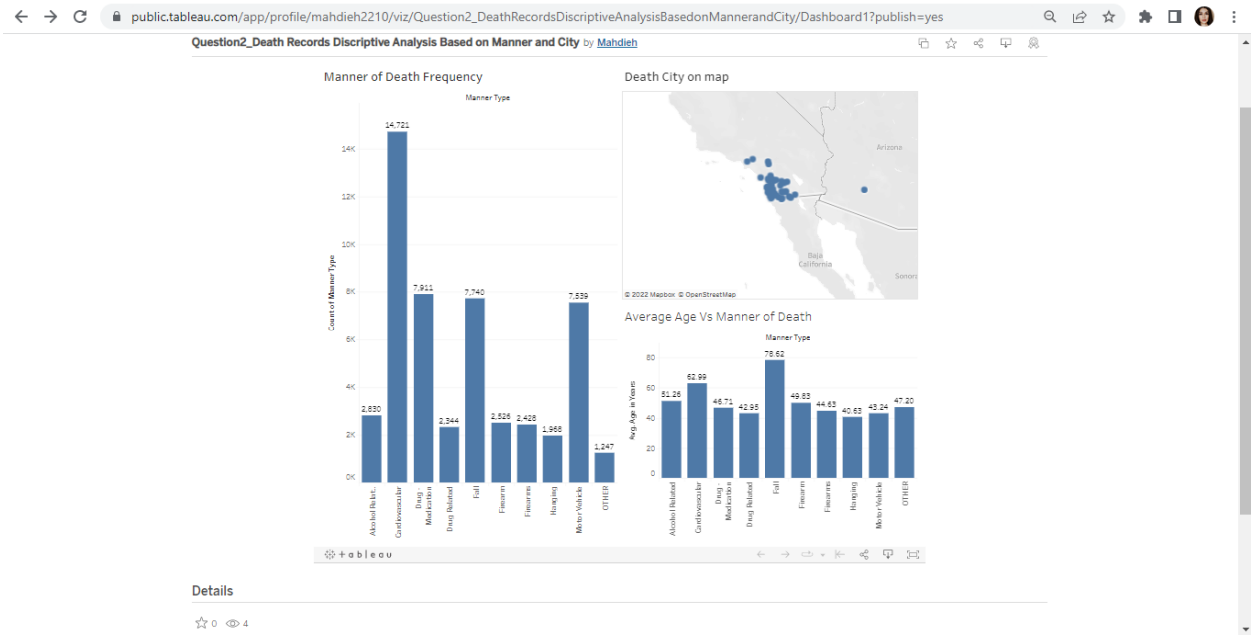
Question 2

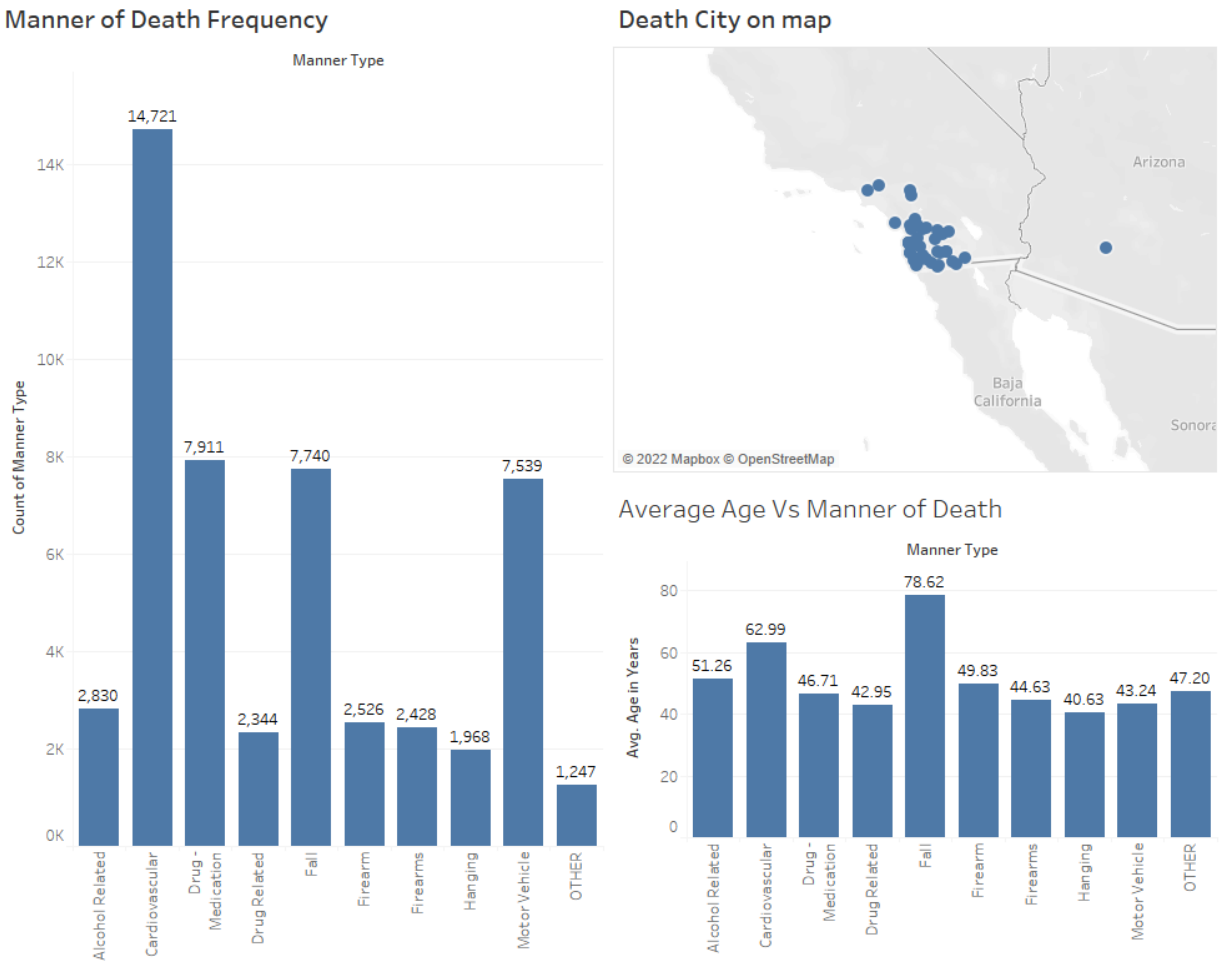
After opening the Medical_Examiner_Cases data to the Tableau, I decided to plot the mean of “Age in Years” for each manner of death (using “Manner Type” variable). In order to do so, I dragged “Manner Type” into the column section, and “Age in Years” into the row section; then Marked Measure>Mean for “Age in Years” in its setting. There were some null values in the “Age in Years” variable, so I used filtered feature and selected only the rows without a null value in “Age in Years”. I created this chart in the first worksheet.

Then I plotted each manner type vs its frequency, so I dragged the variable once in a row, and once in column. For the row section I marked Measure, and Count to show the frequency. I created this chart in the second worksheet.

Additionally, I dragged altitude and latitude variables in the column and row, and used the symbol map feature. Then I dragged Death city as the third variable, and the points were created on the map. I created this plot in the third worksheet. Finally, I dragged the worksheets into a dashboard, and set a name for it. After it was ready to publish, I used Server → Tableau Public → save to Tableau Public. Then entered my account info, and the dashboard was published in my public account after some minutes. This is the URL link to my Tableau Public (this dashboard):

https://public.tableau.com/app/profile/mahdieh2210/viz/Question2_DeathRecordsDiscriptiveAnalysisBasedonMannerandCity/Dashboard1?publish=yes





My dashboard in more details

Question 3

R has provided us with a platform to create dynamic data visualization through R-shiny, so that different people can access the tables, graphs or other info, comment on them, and interact with each other. While, in Tableau, we only can share the dashboards, and graphs in the Tableau Public and other people cannot make changes or comment on them. Also, R is open source and we can integrate it with other apps (for example through DeployR a server-based framework), while Tableau cannot be integrated with other applications.

In R, we can use some functions to make some prediction models, and reflect the results on our graphs, while Tableau only gives us a descriptive analysis, with no tools to predict the future trends of the data. R can only make plots in its own environment while R have some tools to integrate it with other software

including R. R has lots of packages for plotting the data as we want, but the features that Tableau has provided is somehow fixed and organized.

On the other hand, Tabluea is more user-friendly and we can make perfect plots in short amount of time and so easy. The plots and charts made by Tableau are much better than R in terms of easy interpretation; for instance, we can work with geographical variables of the data using the symbol maps option, and it makes a great map to plot data on it.

Generally, I suppose R is better for more complex analysis, while Tableau is more efficient for ordinary people to visually interpret their data in any field.

Question 4

Unisex Names in the US by Nir Smilga:

This is a great Tableau story sharing some interesting information on the most frequent unisex names in the US. For each name, there is a pie chart that depicts men and women shares of the name, showing the gender frequency for the name. In addition to this, the frequency distribution based on the gender is presented; for example if the name frequency is quite equal in both genders, the distribution plot looks like a normal distribution, while for the name that are dramatically frequent in one gender, the distribution is skewed towards that gender. Jordan, Taylor, and Leslie, are the three most frequent unisex names in the US respectively. “Jordan” is more prevalent in men, while “Taylor” and “Leslie” are more frequent in women. Moreover, the trend of this using each name between 1922 till 2021 for both men and women is depicted. Using charts and colors gives a great visual assessment of the data. Another interesting point about this Tableau story is that it has been updated just a week ago!