KHANH TRAN
INFO 250
Date: August 30, 2020

Final Report

Tasks

1. Revise your visualization based on your reflections and my comments.
2. Publish your visualization somewhere that can be seen by the public (including me).
    - You should NOT just publish the standalone graph or dashboard.
    - Your work MUST be published along with some dedicated descriptions of this project for your readers, such as a blog post with the graph or link to the dashboard, or at least some simple descriptions along with the graph in your portfolio.
    - More examples will be shown and discussed later.

Deliverable: A written report with

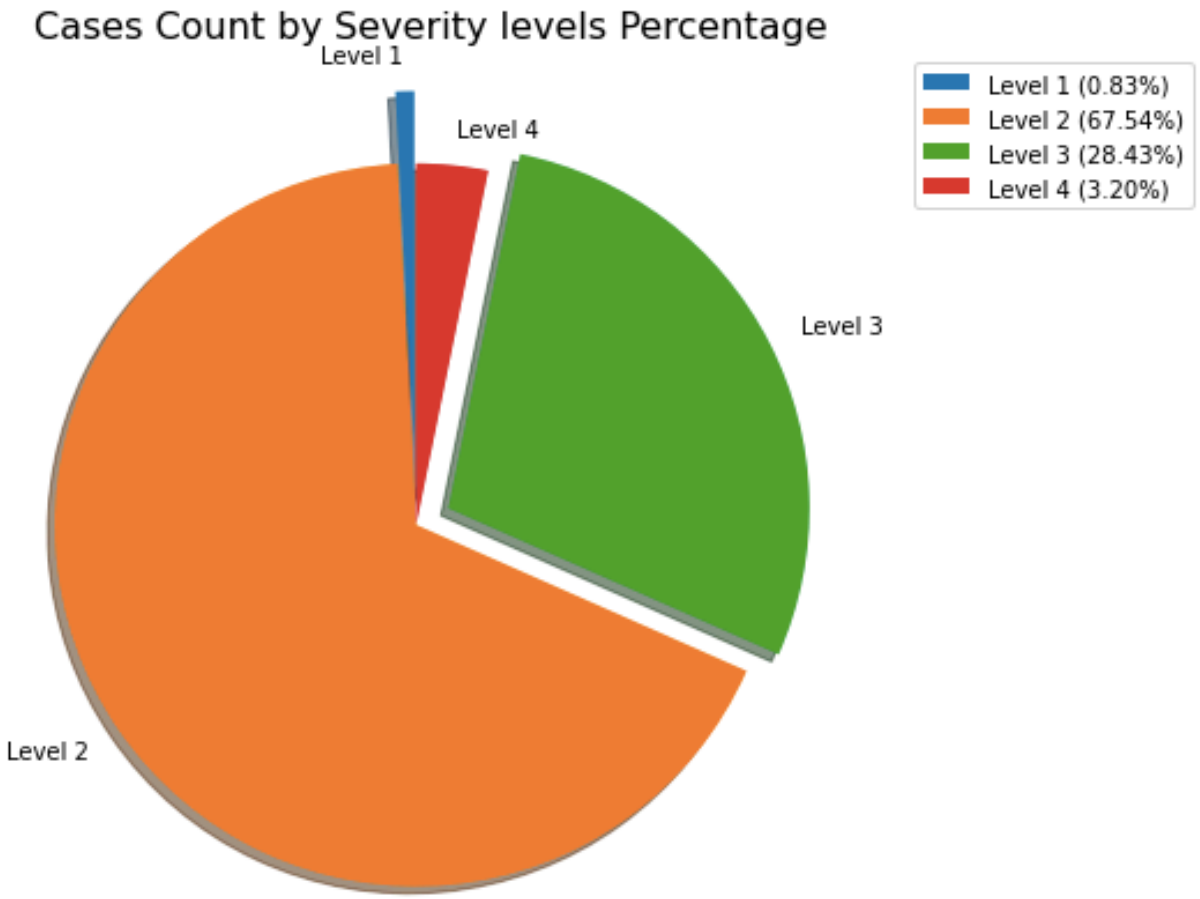
- (1) the link to your published visualization;
- (2) what changes did you make after the initial design and why;
- (3) for those of you decided to use a different dataset as proposed in project 2A, please illustrate the basic planning elements (as what you included in your 2A report) in the report; and
- (4) what are the most and least useful aspect of this course do you see?

In this report, I will discuss about my changes to the visualizations based on the comments from the professor on my previous work. I will go through the graphs in the exact order they show up in my previous work. The titles of the graphs are:

- "Cases count by Severity levels Percentage" (Pie chart)
- "Cases count by Start Hours" (Column chart)
- "Cases Across the Country Identified by Severity levels" (Interactive Geographical Map)
- "Cases count per State" (Column chart)
- "Top 10 States with Highest Cases count vs. Severity levels" (Scatterplot)
- "Temperature vs. Distance vs. Severity levels" (Scatterplot)
- "Cases count when Rain and Not Rain" (Column chart)
- "Cases Counts based on Months and Simplified Weather Conditions" (Stacked Column chart)
- "Cases Counts based on Months and Simplified Weather Conditions (No Others)" (Stacked Column chart)

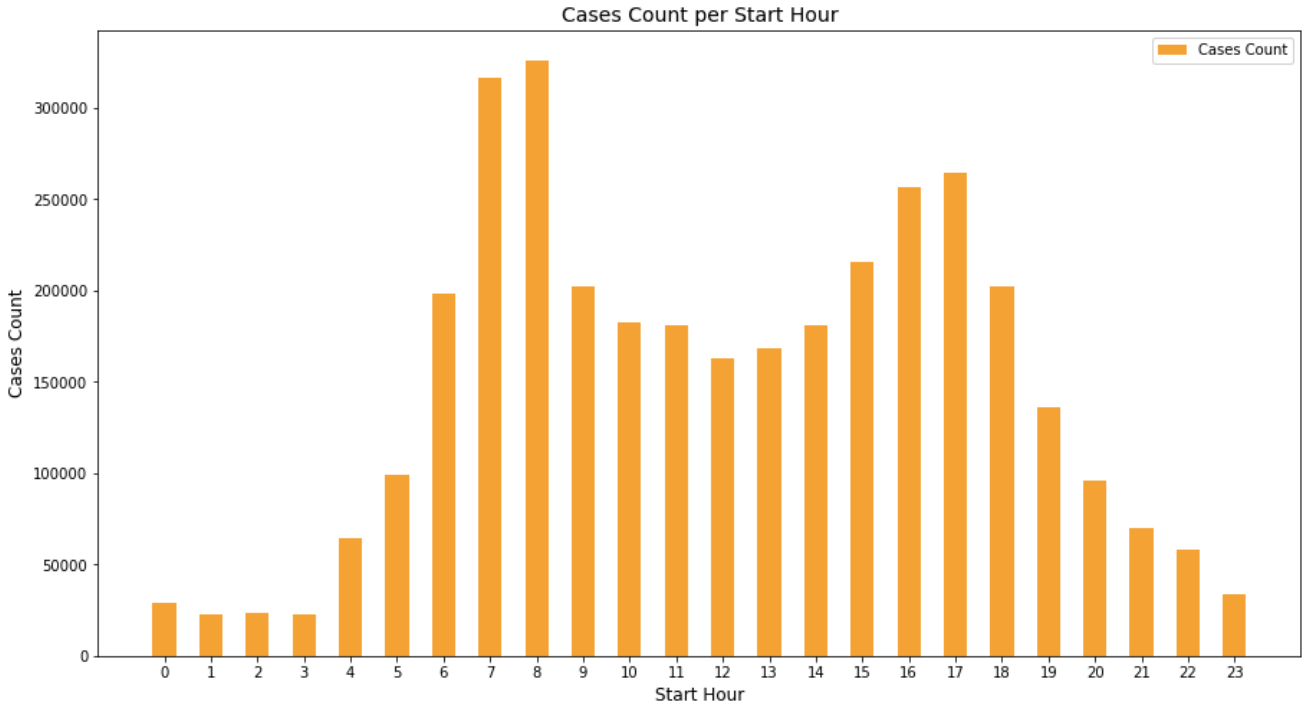
I. Changes to the Visualizations

1. Cases count by Severity levels Percentage.



This is a pie chart that I used in my previous assignment to illustrate the proportion of each severity level. Even though it has well shown the proportions of each class, it failed to be considered a good graph. It was lack of a title, a legend, and clear labels. For this graph, I have added a title, a legend, and made my labels clearer. I removed the percentages from the graph and put the numbers in the legend. By doing that, I think the readers will not find the graph visually confusing. I also changed the class names from “1,2,3,4” to “Level 1, Level 2, ...” in order to deliver a clearer set of labels.

2. Cases count by Start Hours



I didn't make any changes to this graph since the professor didn't point out any flaws.

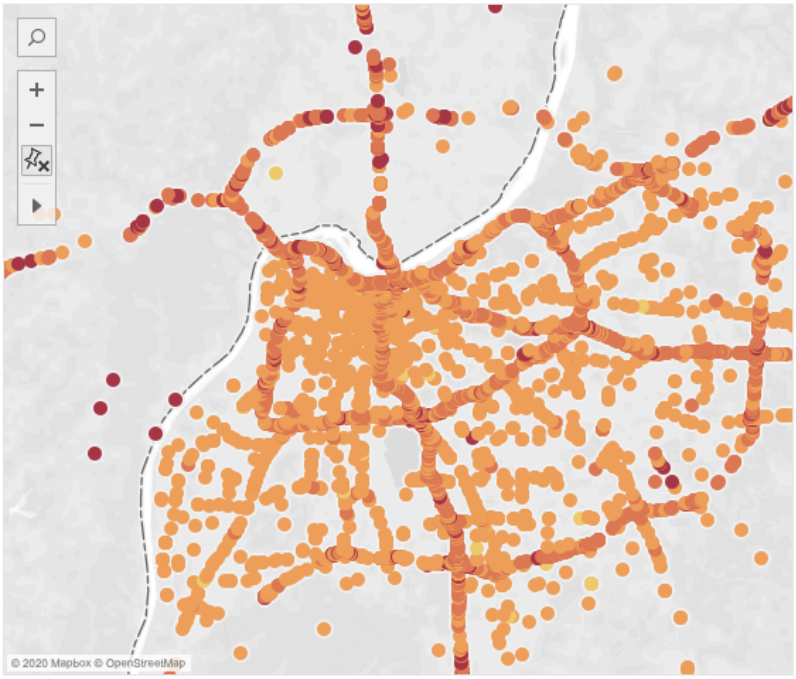
3. Cases Across the Country Identified by Severity levels

For this graph, check out this link to my Tableau Public page: https://public.tableau.com/profile/khanh.tran4682#!/vizhome/INFO250FinalSeverityvsLocationsofAccidents/Sheet1

The only comment that I got on my old static graph that I created using seaborn and matplotlib packages in Python is that it was unreadable and that I should make it an interactive graph in order to improve users' experience. The fact that my dataset contains over 3.5 million records made the map overcrowded with data points which was not a good idea if the user decided to zoom in to explore cases in smaller areas. With Tableau, I was able to create an interactive version where users can zoom in and out and explore cases in any area they want. I kept the same color choices for the severity levels.

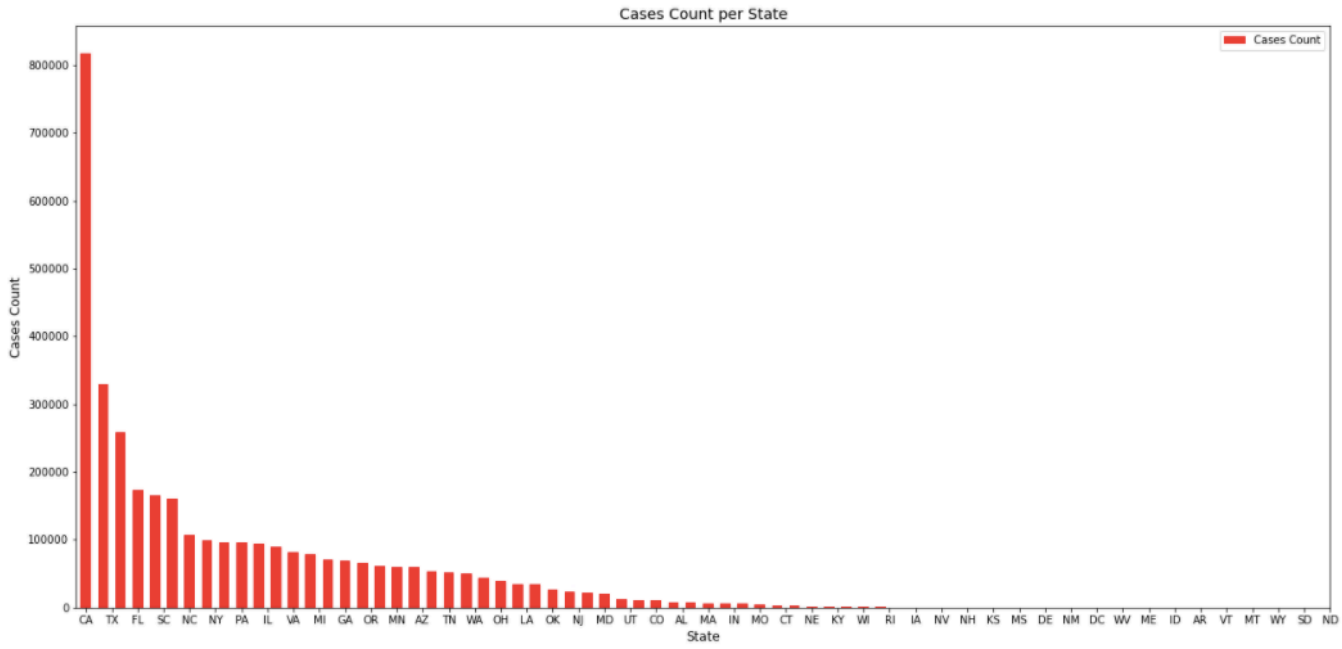
Apart from these changes, I have found myself a couple interesting things when playing around with the new version. I notice that if we zoom in the map, especially in the clusters, we may find the streets pop up from the map thanks to the pattern of the accidents. The graph below shows how we can figure out where the big cities and the roads, which I think are highways, that connect them together are on the map. I call them the "spiders". This was when I zoomed in Kentucky.

There is more to this finding. When I further zoomed in the one of the big cities, I found out that the streets in that city were also glowed up thanks to the color pattern.



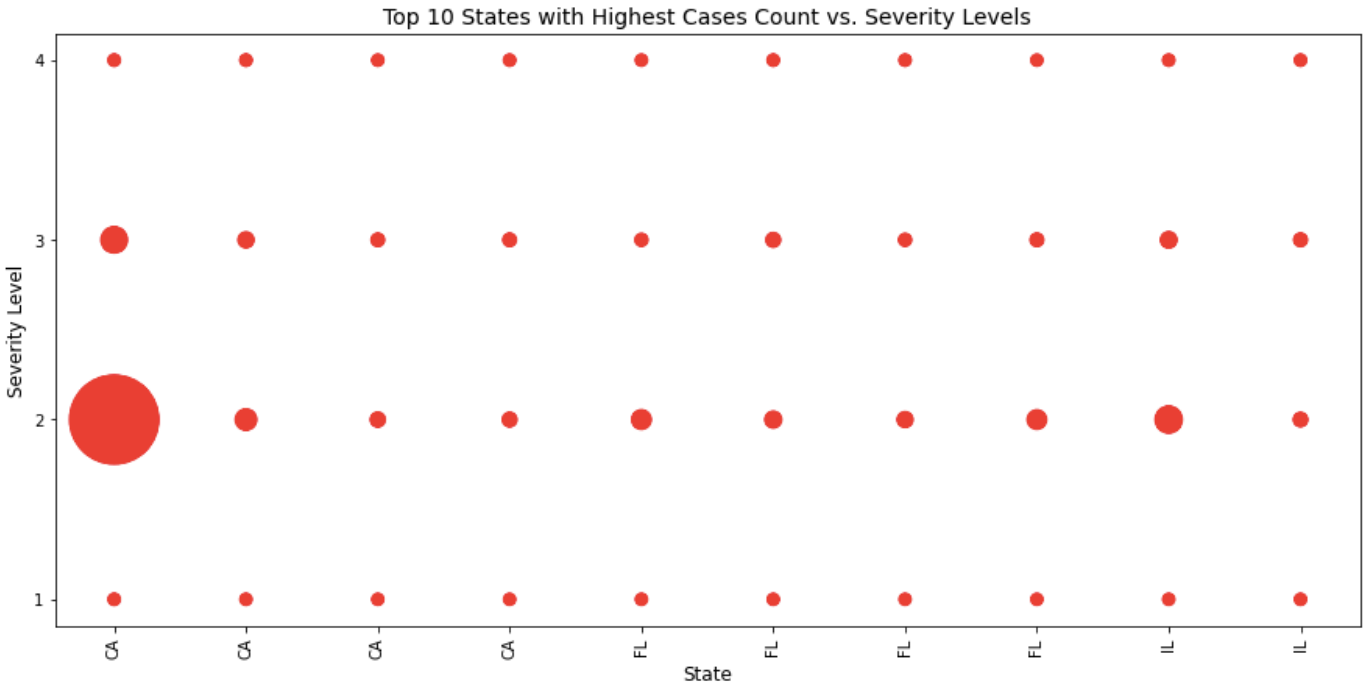
This surely looks like a spider web to me! The users can try this out on my interactive map to explore other parts of the country. However, the loading speed will be quite slow as a result of the high volume of data points.

4. Cases count per State



For this column chart, I have spread out the bars and flipped the state labels right side up so that they are easy to read.

5. Top 10 States with Highest Cases count vs. Severity levels

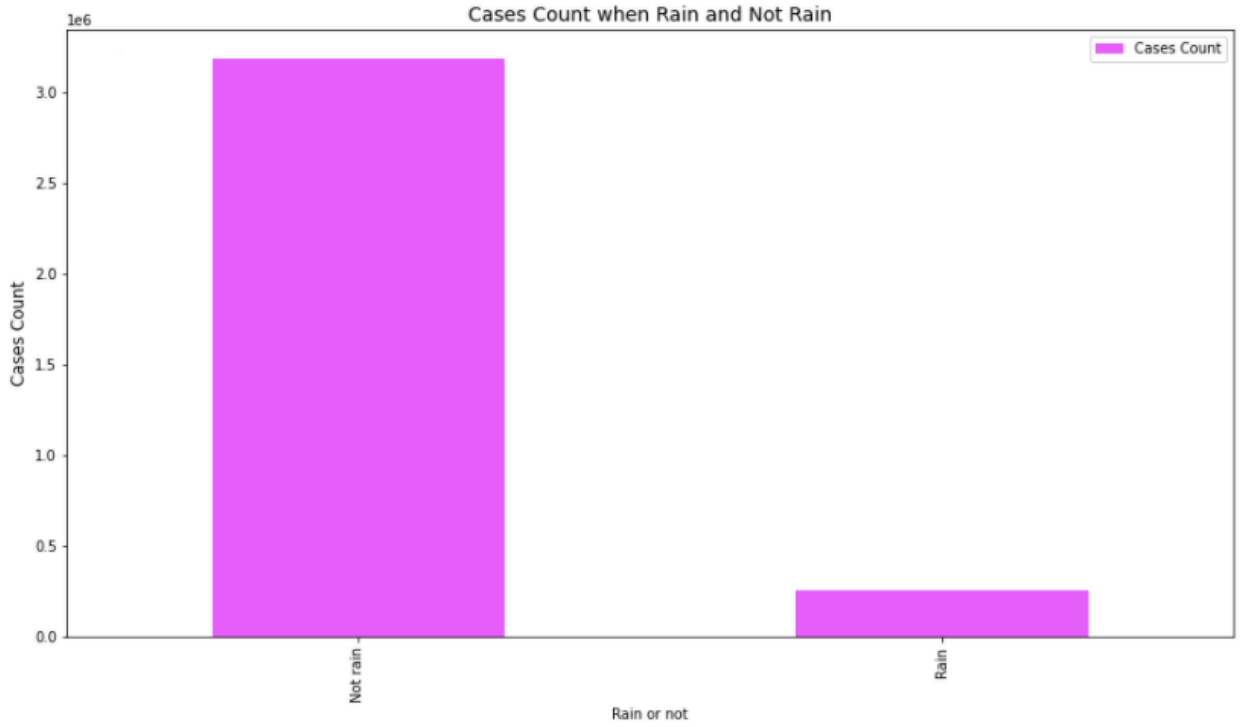


This graph wasn't commented on in my last assignment, but I decided to include it in my report since I think it is visually easy to understand and its patterns clearly do their job in delivering the message to the users. However, I realized that there was a flaw in the old version, which was the state labels being rotated 90 degrees. Here is the new version of the graph.

6. Temperature vs. Distance vs. Severity levels

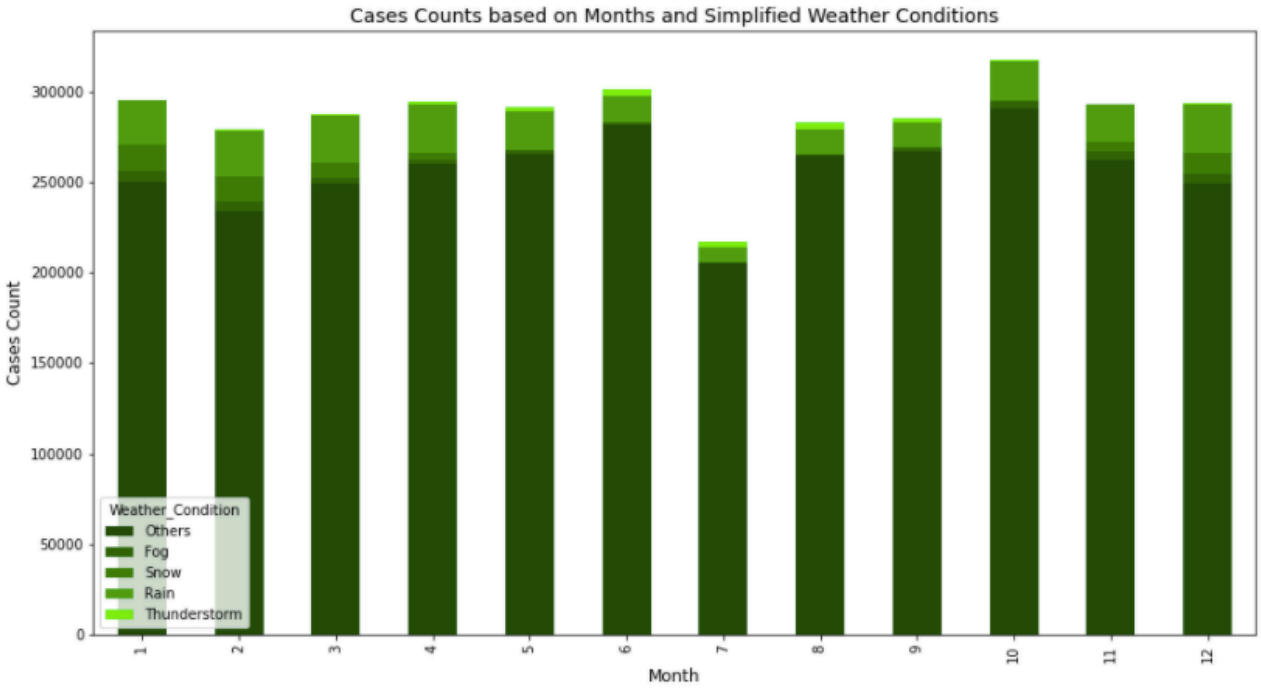
I added a title to this graph. It was clumsy of me not including a title in my old graph. Even though the old graph ran into the same problem as the US map which is an interactive version is recommended, I couldn't produce the same format as the old version via Tableau. That is why I'm keeping the graph this way. Even though it is better to use an interactive visualization, which is the case for any kind of graphs, I believe this case is not that unreadable like my old US map. This still illustrates the main features of the relationship between the specified attributes.

7. Cases count when Rain and Not Rain



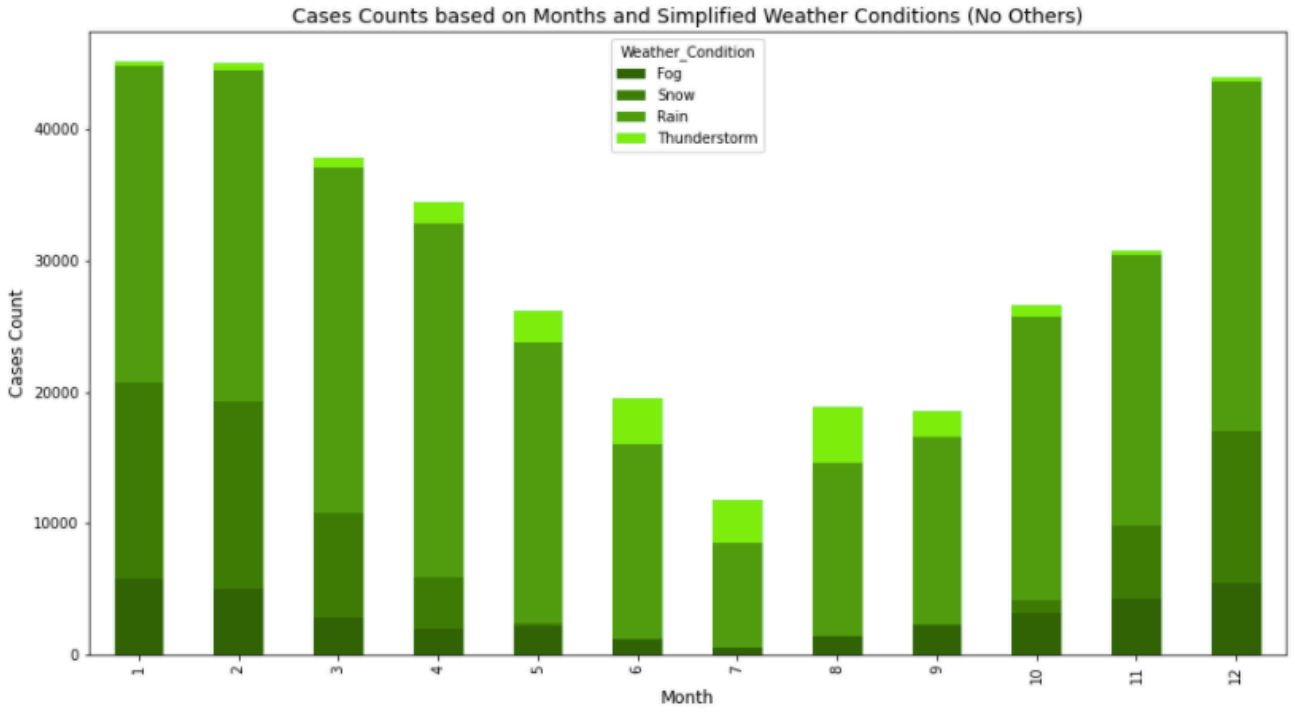
I included a label for the y-axis this time.

8. Cases Counts based on Months and Simplified Weather Conditions



I changed the label for the x-axis from “Start Month” to “Month” because it is just the month the accidents occurred; it isn’t a Start Month.

9. Cases Counts based on Months and Simplified Weather Conditions (No Others)



I did the same thing to this graph: changing the x-axis label from “Start_Month” to “Month”.

10. Link to my public dashboard.

It is not an actual dashboard. It is just where I post this work online. Make sure to check out my Tableau graph which is embedded in a different link above.

II. Course Evaluation

- **Useful aspects:**
    - The first thing that I love about this course is that the professor is a nice guy to work with, and I really love working with someone who is understandable, reasonable, and open to all of the suggestions from the students. He has tried his best to interact with his students, and that really paid off, so thank you Prof. Kai Li!

- I love learning about visualization. The fact that this course was created recently by Prof. Kai Li himself shows how helpful and necessary visualization skills are to the Drexel students. This course introduces the knowledge, the tools, and the practice that suit so well with my career path, so I feel thankful that I took this course.
- The structure of the class content was well-created. I never felt overwhelmed at any point of the course. That being said, the workload is no joke to anyone who thinks this course is an easy A.
- Great assignments. I don't like having too many assignments, but in this class, each assignment has a very different format and requirements. I didn't feel like doing repetitive stuff.

- **Not very useful aspects:**
  - I really want to give some good suggestions but I really don't know what to say. Prof. Kai Li really knows how to talk to his students, so that prevents him from making his students uncomfortable and uninspired.

---- *Thank you, Prof. Kai Li, for the wonderful term!*