جامعة الأخوين

ⵜⴰⵙⴷⴰⵡⵉⵜ ⵏ ⵓⵅⴰⵡⵉⵏ

AL AKHAWAYN UNIVERSITY

Spring 2022

Data Engineering and Visualization

CSC 5356 – 01

Project #4 Report

Omar El Amrani

Supervised by: Dr. Tajjedine

Submission Date: 03/05/2022

What: The data that we want to process

For the provided Jupyter notebooks, I worked with two CVS datasets. The first one describes the relationship between fictional characters. The first column specifies the source character (of the relationship), the second one specifies the target character, the third column showcases the strength of the relationship between the two characters using a scale from 1 to 50, and the last column shows how many relationships does the source character have with the other characters.

The second dataset is also a CSV file that describes US cars. It shows the brand, price, status, year of production, mileage, and the color of the cars. This dataset contains information about 3000 real cars, making it fairly large.

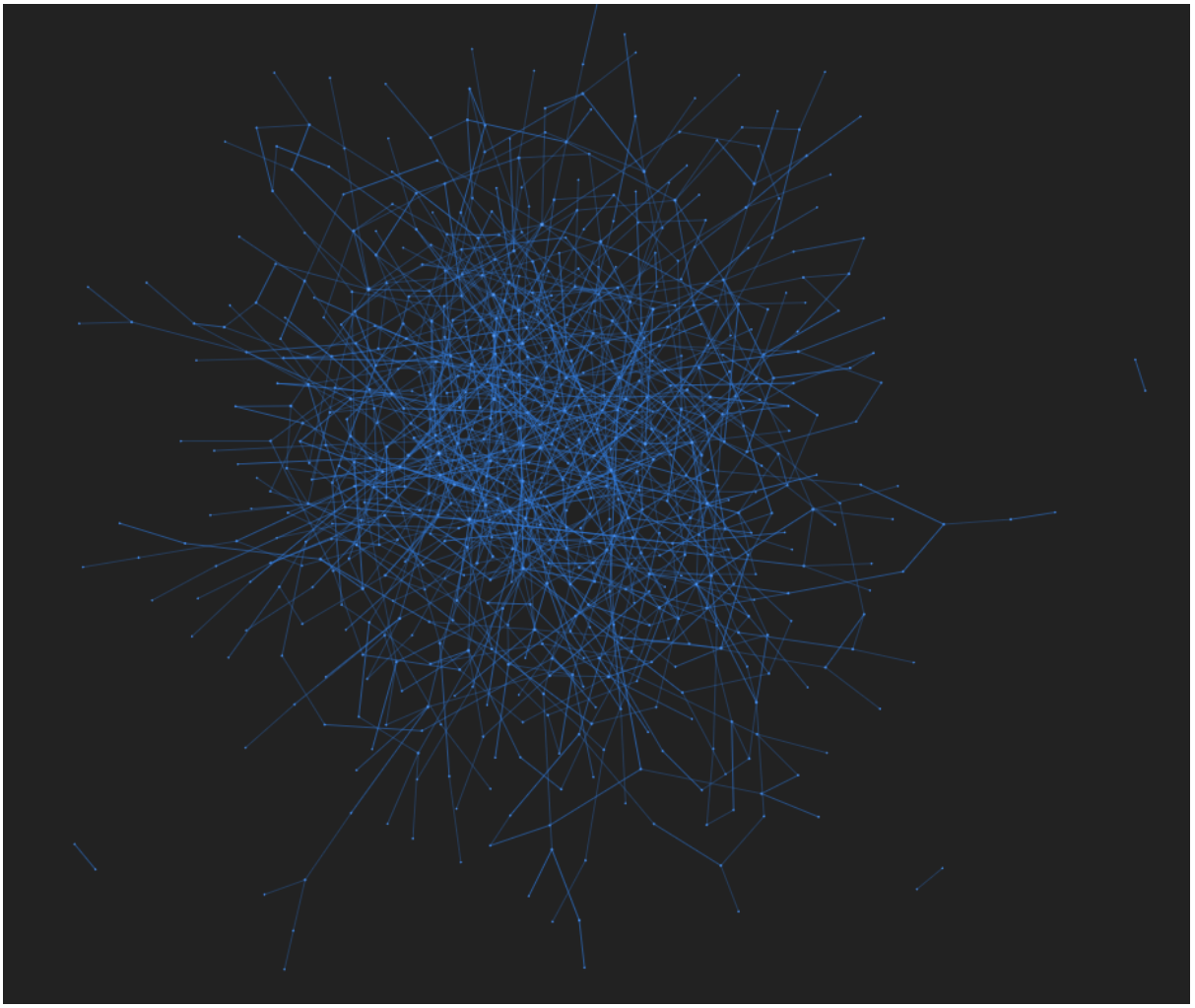
Why: The intended use of the visualization

For this project, I want to focus on the ***present*** aspect of visualization actions. For the first dataset, I want to first showcase the relationship between the characters in a graphical manner. The second task I want to tackle is to show the correlation between the number of relationships and the strength of relationships. In other words, how does, *on average*, the strength of relationships change with the number of relationships a certain character has.

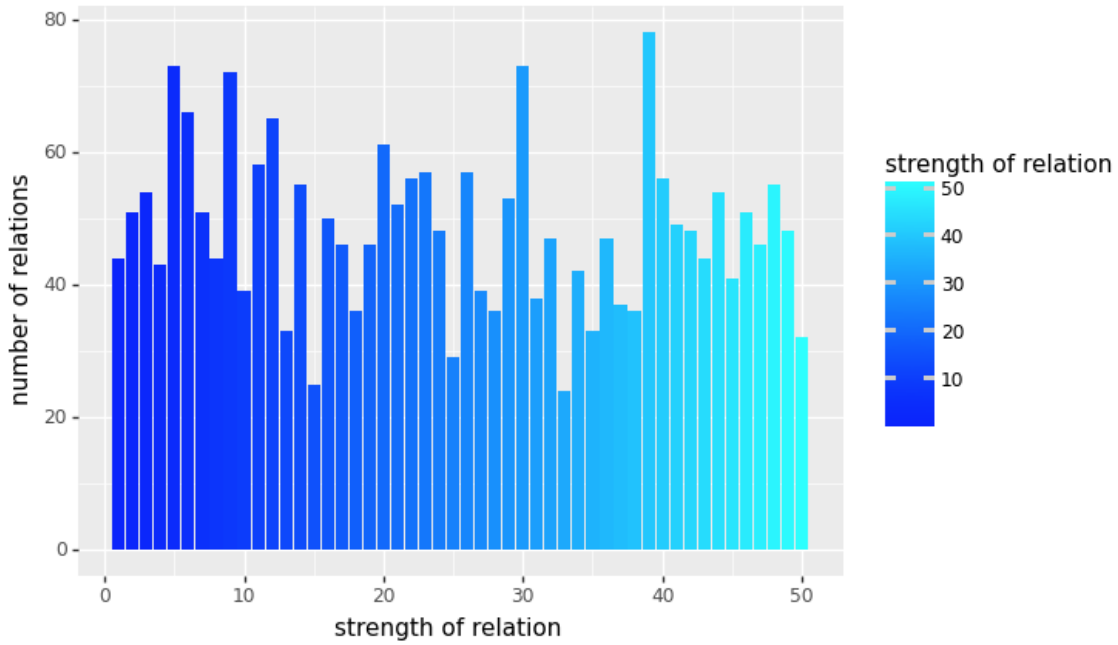
For the second dataset, I want to show the correlation between the mileage of a car and its price. My second task will be to showcase the information about the price of a car depending on its brand.

How: What the user sees

For the first dataset, we want to show a graph representation of the relationships. We will use the thickness of our connections to showcase the weight of each relationship, and the size of the node to showcase nodes that have more relationships.



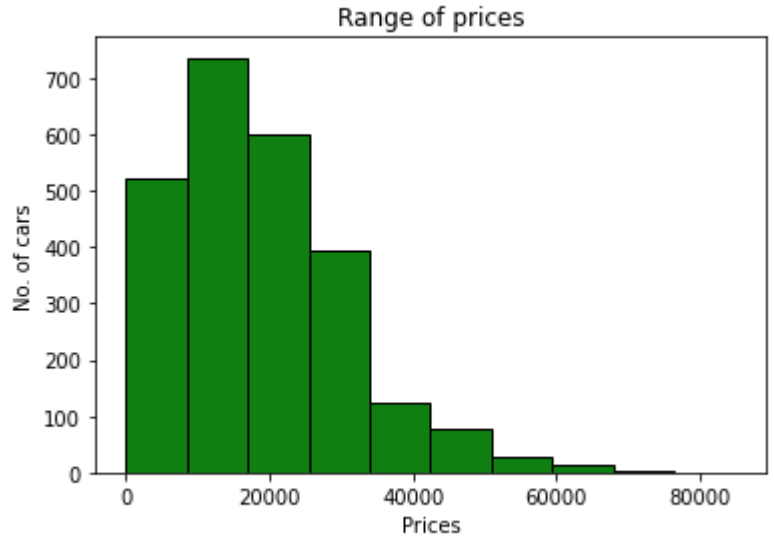
For our second task, we want to showcase the correlation between the number of relationships and the strength of the relationships. In other words, we want to see how does the strength of relationships change on average when the number of relationship changes.



In the above visualization, we showcase the strength of relationship both by using the x dimension of the bar-graph idiom, and also through the color gradient channel.

For the second dataset, we wanted to see how the number of car changes as we increase the price range. As such, we use a simple histogram as our idiom that will showcase the change in the number of car relative to the price.

```
Text(0, 0.5, 'No. of cars')
```



For our second task, we want to see how is the price range of cars affected by the different brands. We want to use boxplot that shows us the mean, the standard deviation, and the outliers. We will further encode the different brand through a color scheme for more visual clarity.