# CSE-564: Visualization

## Mini Project #1 Report

Saurabh Sanjay Agrawal (113278626)

### **Project Overview**:

- This project focuses on the visualization of CSV data in the form of a bar chart for categorical variables, a histogram for numerical variables and a scatterplot for both categorical and numerical variables.

### **Data:**

- The data used for this project is 'Car Rental Data' in the USA and it is obtained from following source in Kaggle: https://www.kaggle.com/kushleshkumar/cornell-car-rental-dataset

### **Attributes description:**

Following are the attributes I chose from the dataset which I thought would be interesting to visualize and analyse:

- Categorical
    1. State
        - Represents the State in which the rental vehicle is used
        - The distinct States in the data: Georgia (GA), Texas (TX), Tennessee (TN), Maryland (MD), Colorado (CO), Florida (FL), New Jersey (NJ), California (CA), Louisiana (LA), Utah (UT)
    2. Vehicle Type
        - Represents the type of rental vehicle
        - Distinct Vehicle Types in the data: SUV, Car, MiniVan, Truck, Van
    3. Fuel Type
        - Represents the type of fuel used by the rental vehicle
        - Distinct Fuel Types in the data: Gasoline, Hybrid, Electric, Diesel
    4. Brand
        - Represents the brand of the rental vehicle
        - Distinct Brands in the data: Acura, Alfa Romeo, Aston Martin, Audi, Bentley, BMW, Buick, Cadillac, Chevrolet, Chrysler, Dodge, Ferrari, FIAT, Ford, Genesis, GMC, Honda, HUMMER, Hyundai, Infiniti, Jaguar, Jeep, Kia, Lamborghini, Land Rover, Lexus, Lincoln, Lotus, Maserati,

Mazda, McLaren, Mercedes-Benz, MINI, Mitsubishi, Nissan, Polaris, Pontiac, Porsche, Ram, Rolls Royce, Saturn, Scion, smart, Subaru, Suzuki, Tesla, Toyota, Volkswagen, Volvo

5. Manufacture Year
    - Represents the manufacture year of the rental vehicle
    - Distinct Manufacturing Years in the data: 1961, 1980, 2000, 2004 - 2020

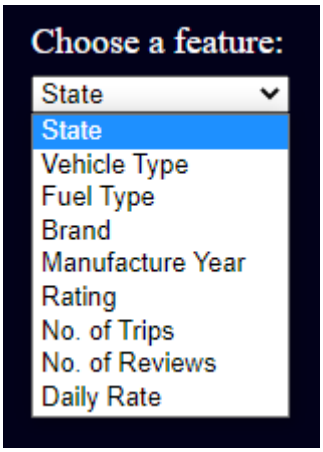
- Numerical
    1. Rating
        - Represents the rating given to the rental vehicle from 1 to 5
    2. No. of Trips
        - Represents the number of trips taken by the rental vehicle
    3. No. of Reviews
        - Represents the number of reviews for the rental vehicle
    4. Daily Rate
        - Represents the daily rate in USD on the rental vehicle

## Reason for the choice of dataset:

- I thought this data is interesting because of my start-up mind-set to explore trends in Vehicle Renting Businesses and to find market gaps if any.
- Plus, this dataset matches the requirement of having 15 dimensions (attributes) with a good mixture of both categorical and numerical variables.

## Features implemented in this project:

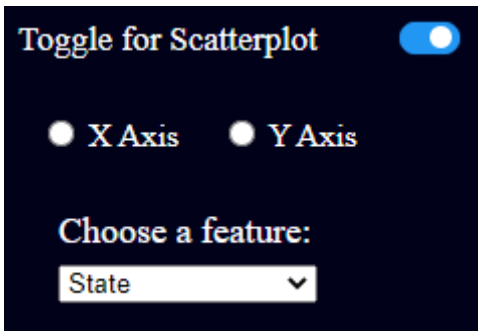
1. Dropdown Menu



- Using this dropdown menu, the user can select a variable for the X-Axis, and based on whether the variable is categorical or numerical, a bar chart or a histogram will be plotted having Y-Axis as the Frequency.
- **Note**: I have managed to use the same dropdown menu for plotting Bar Chart, Histogram, and Scatterplot.

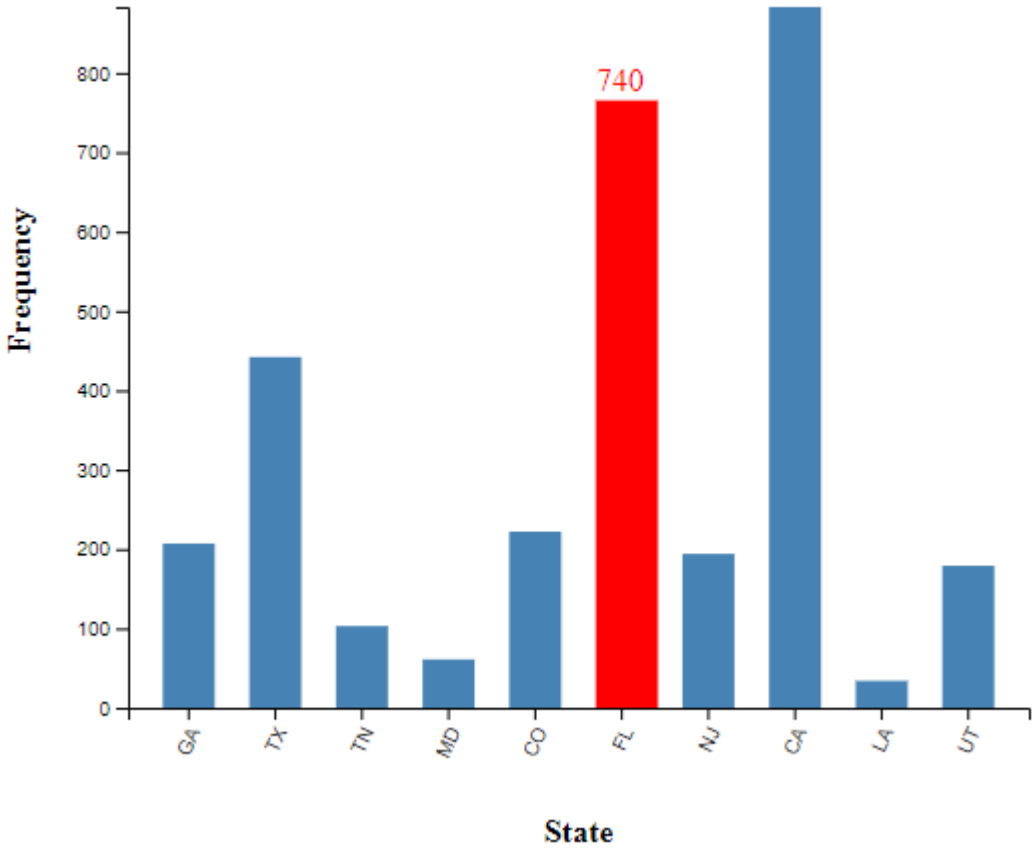
2. Toggle and Radio buttons for Scatterplot


- User can make use of this toggle button if he/she wants to plot a scatterplot.
- After toggling this button, the user will get two radio buttons, one for X-Axis and one for Y-Axis.




- Then the user can select the variable for X-Axis by checking the X-Axis Radio button first and then selecting a variable from the dropdown. Similarly, for selecting a variable from the same dropdown menu for Y-Axis, the user needs to check Y-Axis radio button and then select a variable from the dropdown menu.
- The default variable for X and Y Axis is “State”.
- The user can choose the same variable for both the axes from the same dropdown.

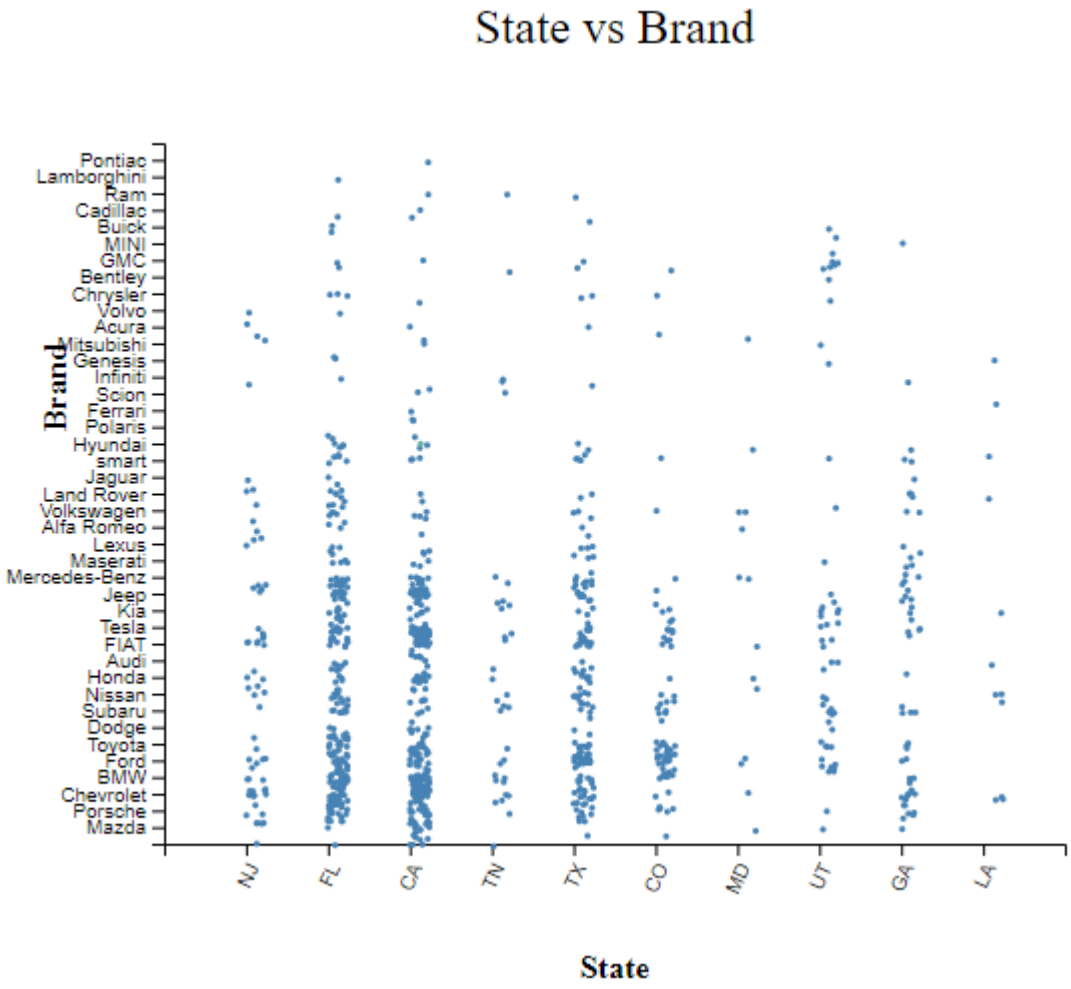
3. Highlight and display bar’s value on mouse hover on bar




- The user can easily get to know the bar's value by hovering on the bar with the help of this feature. It is implemented both for Bar Chart and Histogram.

4. Mouse click and drag left or right
    - In the case of numerical variables, i.e. for Histogram, the user can increase or decrease the bin size and number of bins with the help of this feature.
    - On mouse click and left drag, bin size will decrease and the number of bins will increase.
    - On mouse click and right drag, bin size will increase and the number of bins will decrease.
    - The default number of bins is 20. Minimum and maximum number of bins are 2 and 100 respectively.

5. Jitter in scatterplot
    - For categorical variables, I have added jitter to avoid data overlapping in case of scatterplots




**Technologies used:**

- Python is used for data cleansing and sampling.
- D3.js along with HTML and CSS is used for plotting all the charts.

**Video Link:** https://youtu.be/nah0iZQwHrE

**Code Execution:**

- Open the project in Visual Studio, go to index.html and host it as a live server