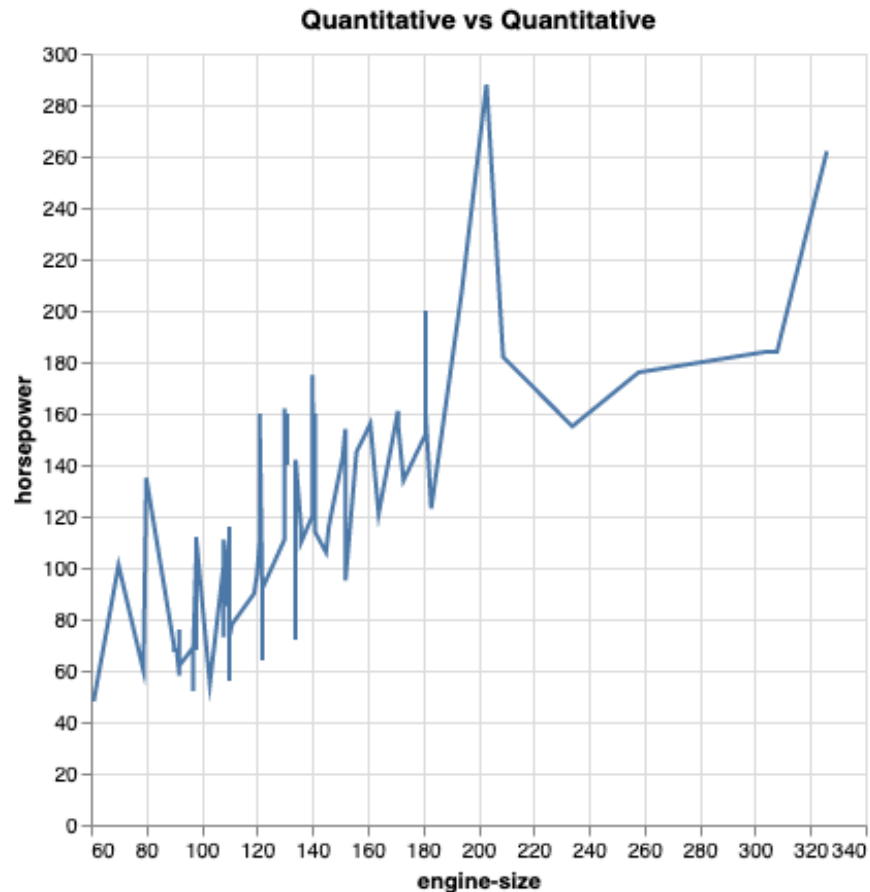Atiya Mirza
02/25/22
CSc 474

<div align="center">Assignment 1</div>

Quantitative vs. Quantitative
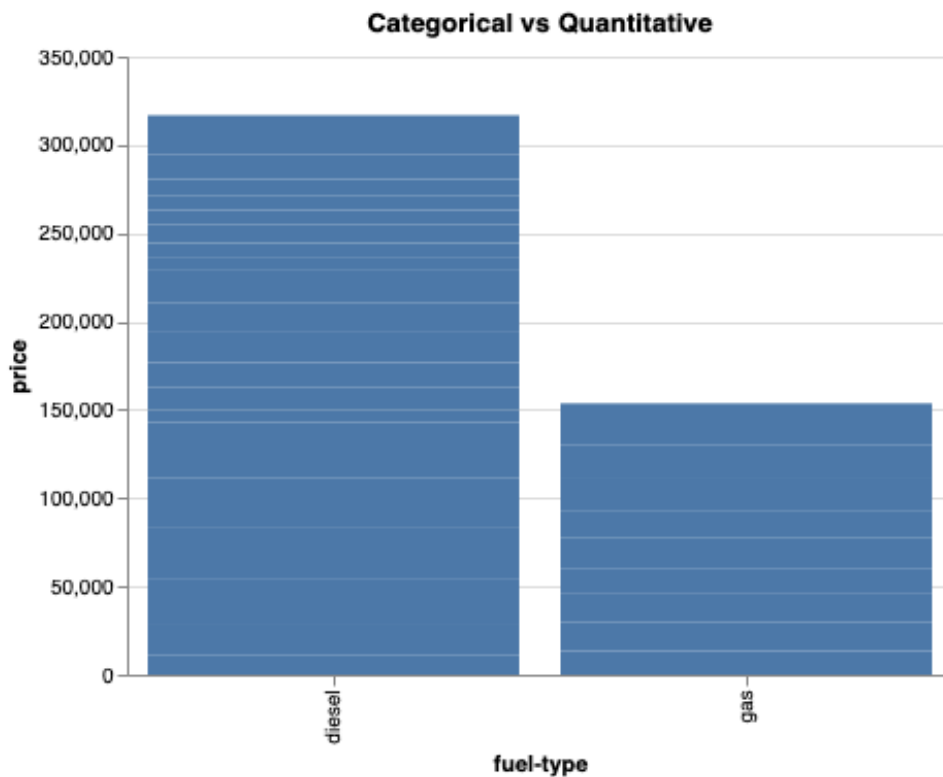


```
{
    "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
    "title": "Quantitative vs Quantitative",
    "data": {
      "url": "https://raw.githubusercontent.com/atiyamirza7/csc474/main/imports-85.json"
    },
    "width": 375,
    "height": 375,
    "mark": "line",
    "encoding": {
      "x": {"field": "engine-size", "type": "quantitative"},
      "y": {"field": "horsepower", "type": "quantitative"}
    }
}
```

A question for comparing two quantitative attributes based on the given dataset is "How does the size of a car's engine affect its horsepower?" I chose this question because horsepower is a

measurement of how powerful a vehicle's engine is, so I wanted to know if the power is correlated to the size of the engine itself. When creating a visualization for two quantitative attributes, there are several suitable options including scatterplot, line chart, and dot plot. I decided to implement the data in a line chart because they are useful for finding general trends between two attributes since the points are all connected by a line. Based on the question and the visualization shown above, we can conclude that generally speaking, the horsepower of a vehicle increases as the engine size gets bigger. We know this to be true because the line graph exhibits a positive trend between the two attributes. This suggests that the larger a car's engine is, the more powerful it will be. Furthermore, I thought it was interesting to see such a big jump in horsepower when the engine size is between 180 and 200 and then immediately a big drop in horsepower when the engine size is between 200 and 220. I'm not knowledgeable enough about cars to know why this would occur, but perhaps it could mean that ~200 is the optimal engine size to maximize horsepower.

Categorical vs. Quantitative

```json
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "title": "Categorical vs Quantitative",
  "data": {
    "url": "https://raw.githubusercontent.com/atiyamirza7/csc474/main/imports-85.json"
  },
  "width": 400,
  "height": 300,
  "mark": "bar",
  "encoding": {
    "x": {"field": "fuel-type", "type": "nominal"},
    "y": {"field": "price", "type": "quantitative"}
  }
}
```
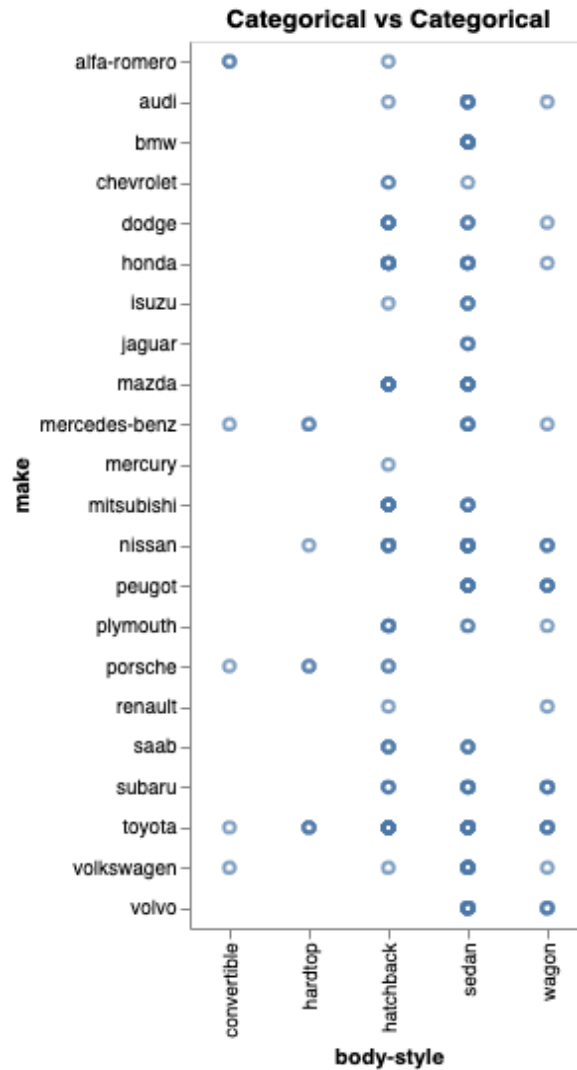
A question for comparing one categorical and one quantitative attribute could be "Are diesel-fuel or gas-fuel cars more expensive to buy?" This question can be answered with a proper visualization using the 'fuel-type' and 'price' attributes from the dataset. When creating a visualization for both categorical and quantitative attributes, options may include a bar chart, pie chart, or polar area chart. I thought a bar chart would be the best option since there are only two categorical attributes and the quantitative values can be easily expressed with the height of the bars. Based on the question and visualization shown above, it's evident that diesel fueled vehicles are generally much more expensive than gas fueled vehicles. The bar graph suggests that on average, cars with diesel-fuel cost more than double the price of cars with gas-fuel. This makes sense since diesel is known to be more expensive fuel than gas.

Categorical vs. Categorical

```json
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "title": "Categorical vs Categorical",
  "data": {
    "url": "https://raw.githubusercontent.com/atiyamirza7/csc474/main/imports-85.json"
  },
  "width": 200,
  "height": 450,
  "mark": "point",
  "encoding": {
    "x": {"field": "body-style", "type": "nominal"},
    "y": {"field": "make", "type": "nominal"}
  }
}
```

Categorical vs Categorical

A question for comparing two categorical attributes is "Which car brand/style combinations are available for car buyers?" To phrase this question in terms of the dataset, how many of the 5 different car styles exist for each of the 22 car brands? Answering this question requires the use of the 'make' and 'body-style' attributes. Initially, I struggled to find a suitable visualization for these two categorical attributes because most of the charts I looked at were solely for quantitative values or for a mix of the two. I ended up choosing a scatter plot, which may seem like an unconventional option as it is typically used to represent two quantitative attributes but it was the most effective visualization for trying to answer my question. To read the visualization shown above, you can look at the point corresponding to the car brand and body style that you are interested in and if there is a dot present, then you know that the brand has that style of car, otherwise it does not. For example, we know Toyota has the most style options because there is a dot above every column of the horizontal axis whereas BMW has the least style options since there is only one dot in the column representing sedans. Thus, this visualization answers the question because it allows us to view all the different car makes and how many of the 5 body styles are available for every make.