Vega-Lite Visualizations

Visual Analytics DSE I2700

Ivan Miller

Assignment

1. Download the Automobile Data Set provided here: http://archive.ics.uci.edu/ml/datasets/Automobile.
2. Parse the data to automatically extract the following attributes: make (3), fuel-type (4), body-style (7), curb-weight (14), num-of-cylinders (16), engine-size (17), horsepower (22), city-mpg (24), highway-mpg (25), and price (26).
3. Map each data attribute to a suitable visual encoding.
4. Implement the visual encodings of the individual attributes.
5. Develop strategies for visually encoding combinations of attributes (and implement where feasible).
6. What questions can you answer with your visual representations?
7. Discuss your findings.

Choose three main questions comparing:

1. Categorical vs. Quantitative
2. Quantitative vs. Quantitative
3. Categorical vs. Categorical

Choose proper visualization from vegalite from single view examples and implement it using this data

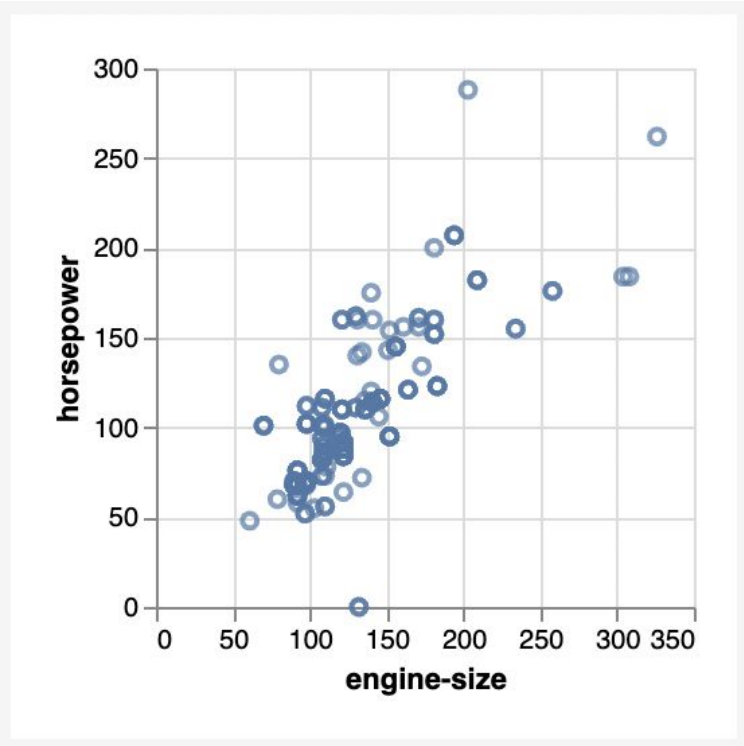
Describe the results of the generated visualization: What kind of relationships you see on data given the questions that you have formulated above?

Engines

Quantitative vs. Quantitative

Horsepower by Engine Size

The scatter plot below shows a positive correlation between the size of the engine and its horsepower: generally, the larger the engine, the more powerful the car is and vice versa.



Example of questions that could be answered:

1. What is the relationship between the engine-size and horsepower?
2. In what range of engine-sizes does the majority of the cars in the dataset falls?

Code:

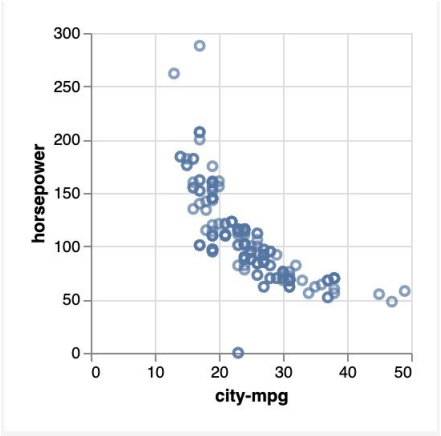
```
{
 "$schema":
"https://vega.github.io/schema/vega-lite/v5.json",
 "data": {
   "url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw
1/cars_clean.csv"
 },
 "mark": "point",
 "encoding": {
   "x": {"field": "engine-size", "type": "quantitative"},
   "y": {"field": "horsepower", "type": "quantitative"}
 }
}
```

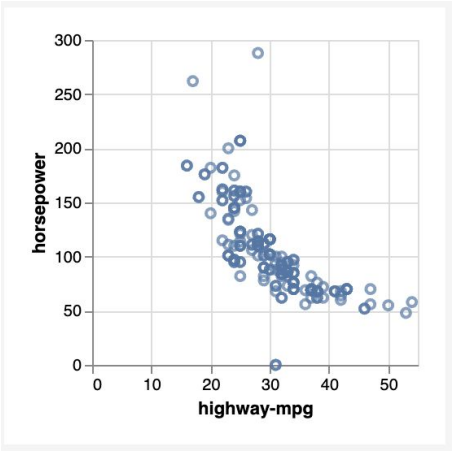
Preview on Vega.GitHub.IO

Horsepower vs MPG

MPG vs horsepower scatter plots show that there is a price to pay for a powerful engine: increase in horsepower usually leads to an increase in MPG (city MPG on the below picture).

Highway MPG is noticeably better with vast majority of the cars in the dataset showing above 20 MPG.





Example of questions that could be answered:

1. What is the relationship between horsepower and MPG
2. What is the expected increase in MPG for each 25hp?

Code:

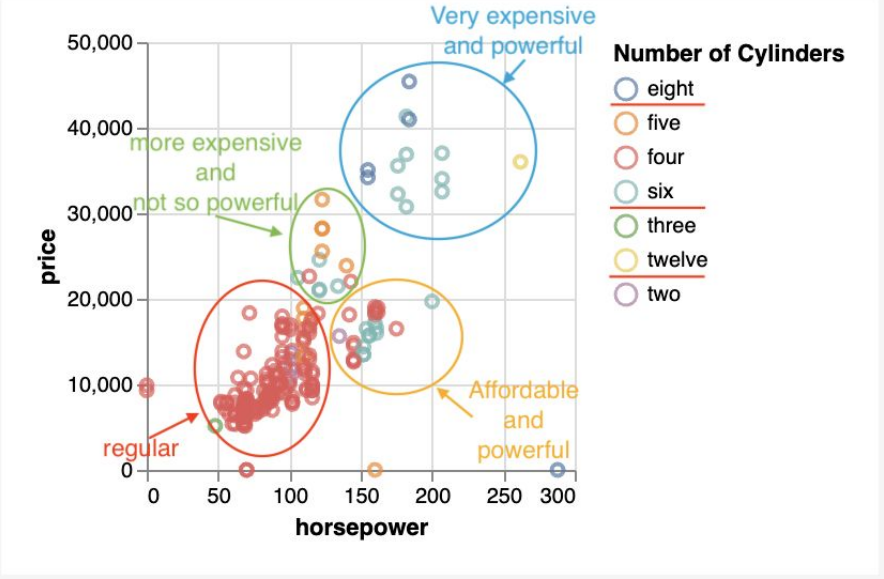
```
{
 "$schema":
"https://vega.github.io/schema/vega-lite/v5.json",
 "data": {
   "url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/ma
in/hw1/cars_clean.csv"
 },
 "mark": "point",
 "encoding": {
   "x": {"field": "city-mpg", "type": "quantitative"},
   "y": {"field": "horsepower", "type": "quantitative"}
 }
}
```

Preview on Vega.GitHub.IO - city, highway

Horsepower vs Price of the Car (+ num of Cylinders)

On top of higher MPG (see previous chart) powerful cars (those with 6-, 8-, and 12-cylinder engines) tend to be more expensive. Adding number of cylinders to the scatter plot allows us to see 4 distinct clusters:

1. Regular cars, primarily 4-cylinder engines, up to 20k and 50-120hp. Most popular cars in the dataset.
2. Still affordable (up to 20k) but powerful cars with 140 - 200hp. Mainly with 4 or 6-cylinder engines.
3. More expensive (~20-33k) but not so powerful cars (less than 150hp), mostly with 5- and 6-cylinder engines.
4. Very expensive (>30k) and powerful vehicles (~150-260hp) with large engines



Example of questions that could be answered:

1. How many cylinders does the majority of engines has?
2. Name engine types in terms of numbers of cylinders that are the most expensive

Code:

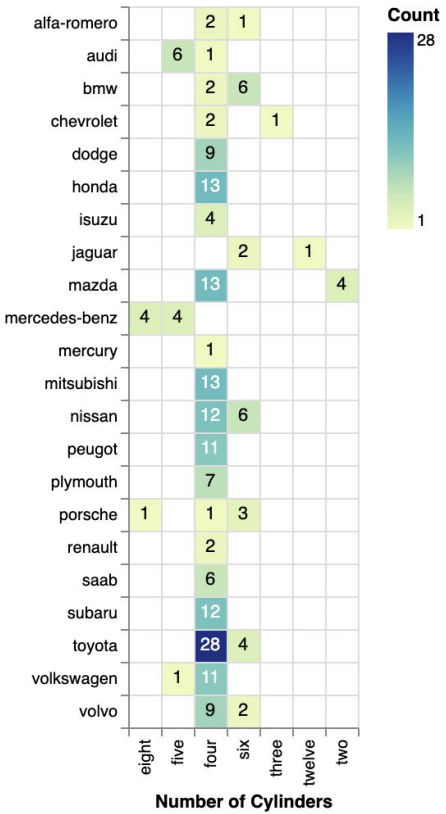
```
{
 "$schema":
"https://vega.github.io/schema/vega-lite/v5.json",
 "data": {
   "url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/mai
n/hw1/cars_clean.csv"
 },
 "mark": "point",
 "encoding": {
   "y": {"field": "price", "type": "quantitative"},
   "x": {"field": "horsepower", "type": "quantitative"},
   "color": {"field": "num-of-cylinders", "type":
"nominal", "title": "Number of Cylinders"}
 }
}
```

Preview on Vega.GitHub.IO

Make by ...

Categorical vs. Categorical

Make by Number of Cylinders in the Engine



This heatmap chart shows the breakdown of makes by number of cylinders in the engine.

Darker colors indicate higher number of records for a given make/engine type.

Count of records representing different engines for each automaker is layered on top of the heatmap.

If the total number of records is higher than 10 the font changes from black to white.

Example of questions that could be answered:

1. Which automaker produces the highest number of cars?
2. What is the most popular engine size in terms of number of cylinders?

```
{"$schema": "https://vega.github.io/schema/vega-lite/v5.json" ,
 "data": {"url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/cars_clean.csv" },
 "transform" : [
   {"aggregate" : [{"op": "count", "as": "num_cars"}],
     "groupby" : ["make", "num-of-cylinders" ]}
 ], "encoding" : {
   "y": {"field": "make", "type": "ordinal", "title":""},
   "x": {"field": "num-of-cylinders" , "type": "ordinal","title": "Number of
Cylinders"}
 }, "layer": [
   {"mark": "rect",
     "encoding" : {
       "color": {"field": "num_cars","type": "quantitative" ,"title": "Count",
         "legend": {"direction": "vertical", "gradientLength" : 120}
       }
     }
   },
   {"mark": "text",
     "encoding" : {
       "text": {"field": "num_cars", "type": "quantitative" },
       "color": { "condition" : {"test": "datum['num_cars'] > 10" , "value":
"white"}, "value": "black"}
     }
   }
 ],"config" : { "axis": {"grid": true, "tickBand" : "extent"}
 }
}Preview on Vega.GitHub.IO
```

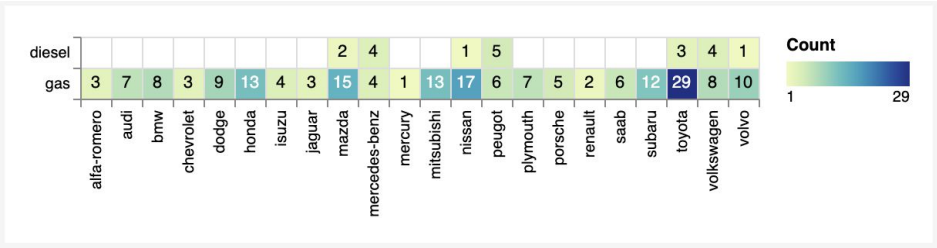
Make by Fuel Type

The vast majority of cars in the dataset are powered by gasoline engines with only 7 out of 22 automakers producing diesel cars: Mazda, Mercedes-Benz, Nissan, Peugeot, Toyota, Volkswagen, and Volvo.

Peugeot and Mercedes-Benz stand out from the above group even more with approximately half of their cars represented in the dataset being diesel (5 out of 6).

Example of questions that could be answered:

1. Which automakers produce diesel cars?
2. Which automaker has the highest relative share of diesel cars compared to all cars in their lineup?



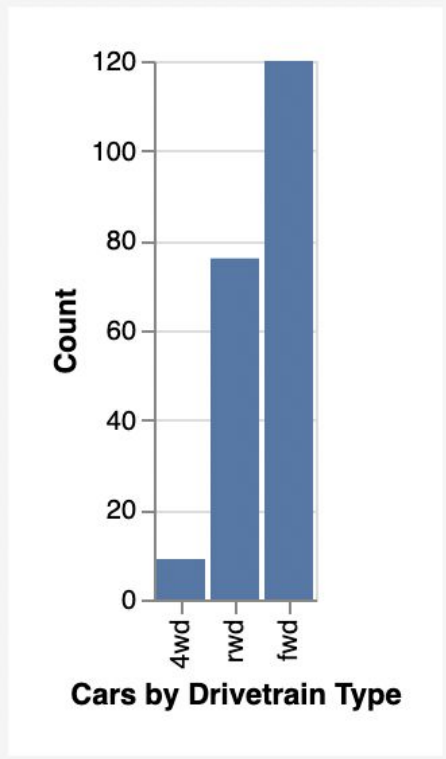

```
{"$schema": "https://vega.github.io/schema/vega-lite/v5.json"  ,
 "data": {"url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/cars_clean.cs
v"},
 "transform" : [
   { "aggregate" : [{"op": "count", "as": "num_cars"}],
     "groupby" : ["make", "fuel-type"]}
 ], "encoding" : {
   "x": {"field": "make", "type": "ordinal", "title": ""},
   "y": { "field": "fuel-type", "type": "ordinal","title": ""}
 }, "layer": [
   {"mark": "rect",
     "encoding": {
       "color": {
         "field": "num_cars","type": "quantitative","title": "Count",
         "legend": {"direction": "horizontal", "gradientLength" : 80}
       }
     }
   },
   {"mark": "text",
     "encoding": {
       "text": {"field": "num_cars", "type": "quantitative" },
       "color": { "condition" : {"test": "datum['num_cars'] > 10" , "value":
"white"},"value": "black"}
     }
   }
 ], "config": {"axis": {"grid": true, "tickBand": "extent"}}
}Preview on Vega.GitHub.IO
```

Drivetrain

Categorical vs. Quantitative

Cars by Drivetrain Type

The distribution of cars in the dataset by the type of the drivetrain shows that the vast majority of vehicles are front-wheel drive, about 30% less are rear-wheel drive and a very few of the cars represented in the dataset are four-wheel drive.



Example of questions that could be answered:

1. What is the least/most popular drivetrain?
2. Approximately how many cars have front-wheel drive?

Code:

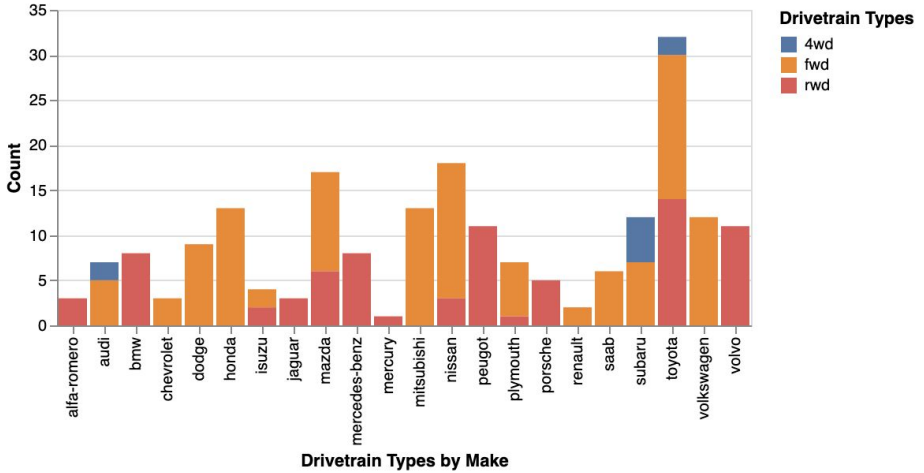
```
{"$schema": "https://vega.github.io/schema/vega-lite/v5.json",
 "data": {
   "url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/cars_clean.csv"
 },"mark": "bar",
 "encoding": {
   "x": {
     "field": "drive-wheels", "type": "nominal",
     "sort": "y", "title": "Cars by Drivetrain Type"
   },
   "y": {
     "aggregate": "count", "type": "quantitative",
     "title": "Count"
   }
 }
}
```

Preview on Vega.GitHub.IO

Drivetrain Types by Make

A more detailed view with each car make broken out by drivetrain type:

- Cars with front-wheel drive are dominating the market with 7 automakers producing them exclusively: Chevrolet, Dodge, Honda, Mitsubishi, Renault, Saab, and Volkswagen.
- 8 automakers produce only rear- wheel drive cars: Alfa-Romeo, BMW, Jaguar, Mercedes-Benz, Mercury, Peugeot, Porsche, and Volvo. However, the absolute share of RWD cars seems to be lower.
- Only 3 automakers: Audi, Subaru, and Toyota make cars with four-wheel drive. Subaru produces more 4WD cars than Audi and Toyota combined.



Example of questions that could be answered:

1. Which automaker produces the highest number of cars?
2. Which automaker produces most of the 4WD cars?

Code:

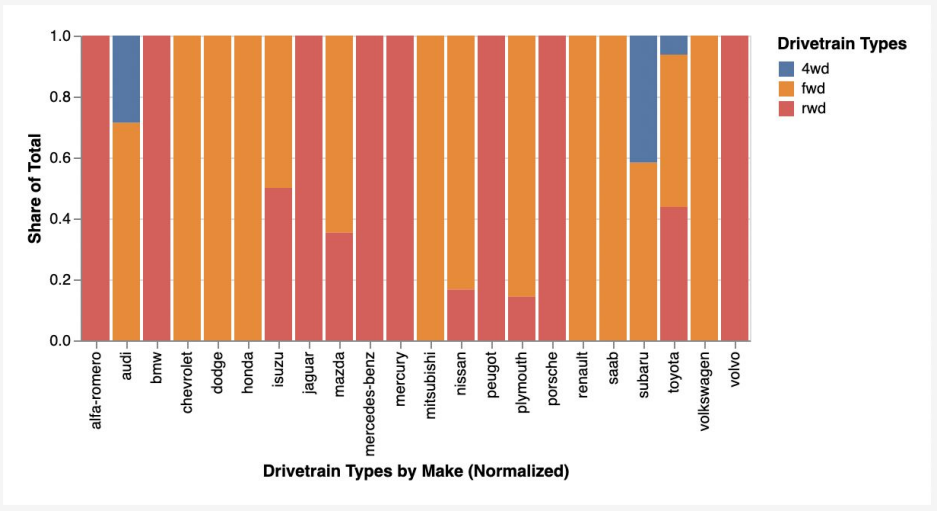
```
{"$schema":
"https://vega.github.io/schema/vega-lite/v5.json",
 "data": {"url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/main
/hw1/cars_clean.csv"
 },"mark": "bar",
 "encoding": {
   "x": {"field": "make", "type": "nominal",
     "title": "Drivetrain Types by Make" },
   "y": {
     "aggregate": "count", "type": "quantitative",
"title": "Count"
   },
   "color": {"field": "drive-wheels", "type": "nominal",
"title": "Drivetrain Types"}
 }
}
```

Preview on Vega.GitHub.IO

Drivetrain Types by Make (Normalized)

Normalized version of the same chart allows us to have a clearer picture of a share of a given drivetrain type for each automaker:

- Even though on previous charts we saw that RWD cars weren't leading in terms of absolute numbers, the abundance of red color on this chart highlights the fact that the number of automakers producing exclusively RWD is the highest (8 solid red bars).
- FWD. Runner up in terms of exclusivity (7 automakers only produce FWD cars). In fact, most of the automakers favor front-wheel driven cars and FWD cars have the highest share of the market.
- 4WD cars only represent a fraction of the dataset (about 3% based on the information on prev charts). Interesting that, the share of 4WD cars among Subaru's vehicles is quite significant (~45%), ~28% for Audi.



Example of questions that could be answered:

1. Which automakers produce only FWD/RWD cars?
2. Which automakers produce 4WD cars?

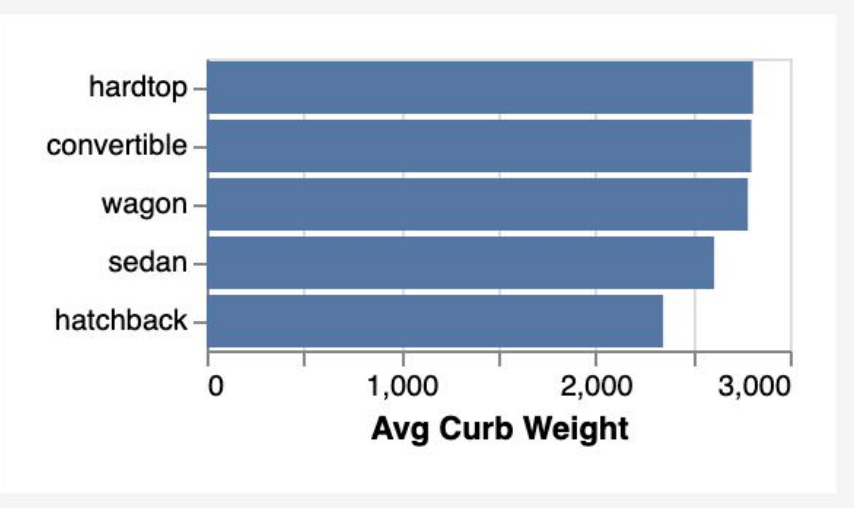
Code:

```
{"$schema":
"https://vega.github.io/schema/vega-lite/v5.json",
 "data": {"url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/main
/hw1/cars_clean.csv"
 },"mark": "bar",
 "encoding": {
   "x": {
     "field": "make", "type": "nominal",
     "title": "Drivetrain Types by Make (Normalized)" },
   "y": {
     "aggregate": "count", "type": "quantitative",
"title": "Share of Total", "stack": "normalize" },
   "color": {"field": "drive-wheels", "type": "nominal",
"title": "Drivetrain Types"}
 }
}
```

Preview on Vega.GitHub.IO

Curb Weight by Body Style

Hatchbacks tend to be the lightest among other body types.
There's no significant difference in average weights of hardtops, convertibles or wagon body styles.



Example of questions that could be answered:

1. Which body style tends to be the lightest on average?
2. On average is there a difference in curb weight of a hardtop, convertible or wagon?

Code:

```
{"$schema":
"https://vega.github.io/schema/vega-lite/v5.json",
 "data": {
   "url":
"https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw
1/cars_clean.csv"
 },
 "mark": "bar",
 "encoding": {
   "y": {
     "field": "body-style", "type": "nominal",
     "sort": "-x","title": ""},
   "x": { "field": "curb-weight", "aggregate": "mean",
"type": "quantitative", "title": "Avg Curb Weight"}
 }
```

} Preview on Vega.GitHub.IO