

# Vega Lite Visualizations

Visual Analytics DSE I2700

Ivan Miller  
Fall 2021

# 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?

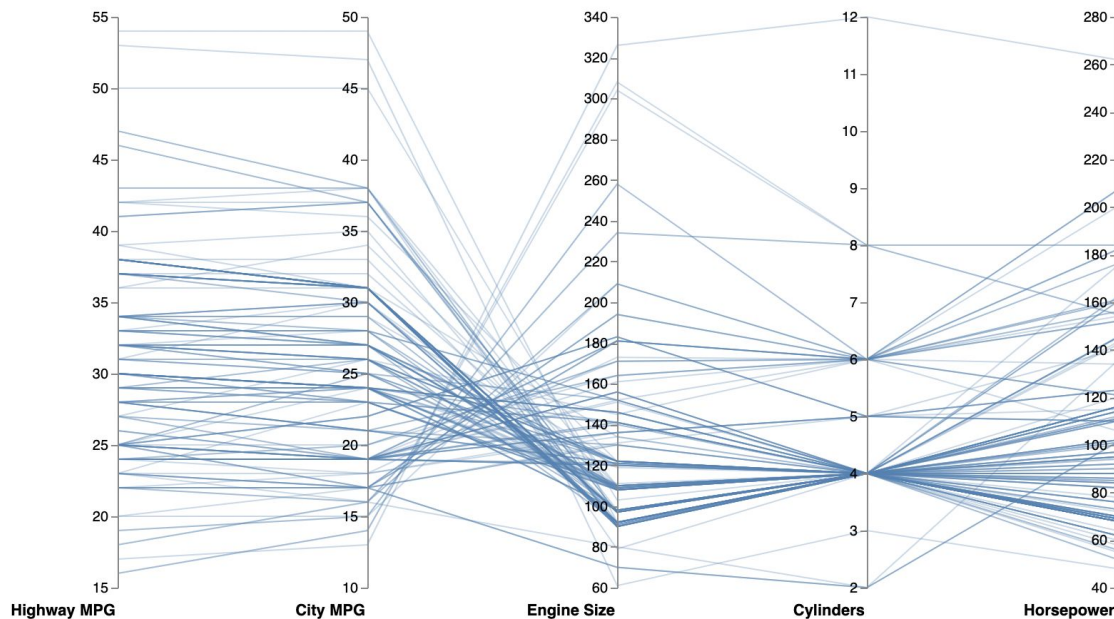
# Engines

Quantitative vs. Quantitative

# Parallel Coordinates Plot on Several Dimensions

The parallel coordinates plot allows us to spot relationships between different variables and make initial assumptions about the size of the engine, number of cylinders, and horsepower of a car, how they could be connected and consequently affect the MPG.

We would use this chart to draft initial questions that we would then try to answer while exploring relationships between different variables at greater detail.

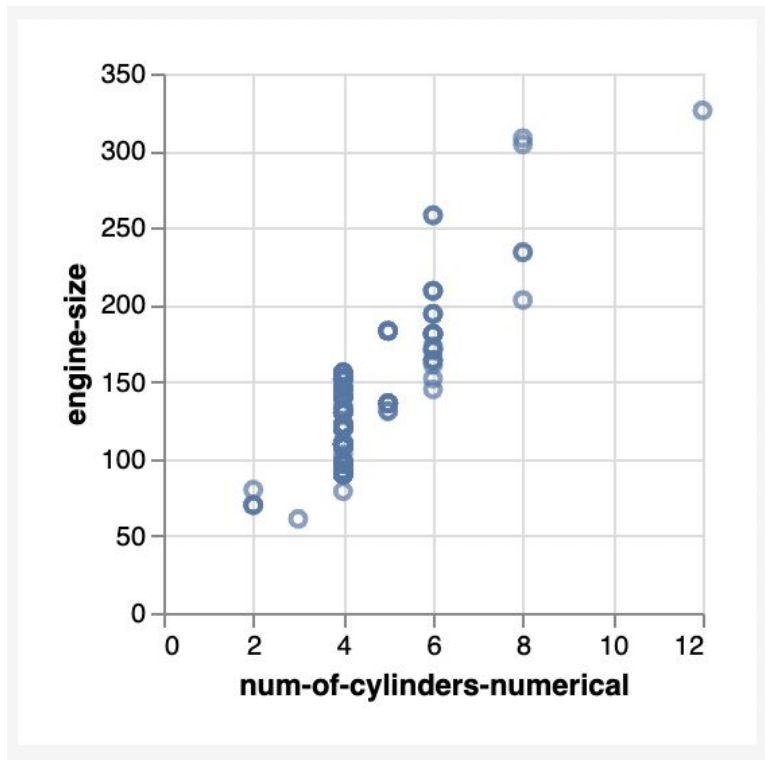


Example of questions that could arise:

1. What are the relationship between the variables describing engines of cars (engine-size, number of cylinders, horsepower)?
2. In what range of engine-sizes does the majority of the cars in the dataset fall?
3. Are there any clusters of cars by the number of cylinders that could be identified in a parallel coordinates plot the presented here?
4. How many cylinders do the cars with the most horsepower have?
5. How does the engine size affect MPG?
6. What is the relationship between city and highway MPG?

# Engine Size vs Number of Cylinders

The scatter plot below shows a positive correlation between the number of cylinders and the size of the engine: generally, the higher the number of cylinders the larger the engine itself should be:



Example of questions that could be answered:

1. What is the relationship between the engine-size and number of cylinders? - *they are positively correlated*
2. Engines with how many cylinders seem to be the most represented in this dataset? - *four*

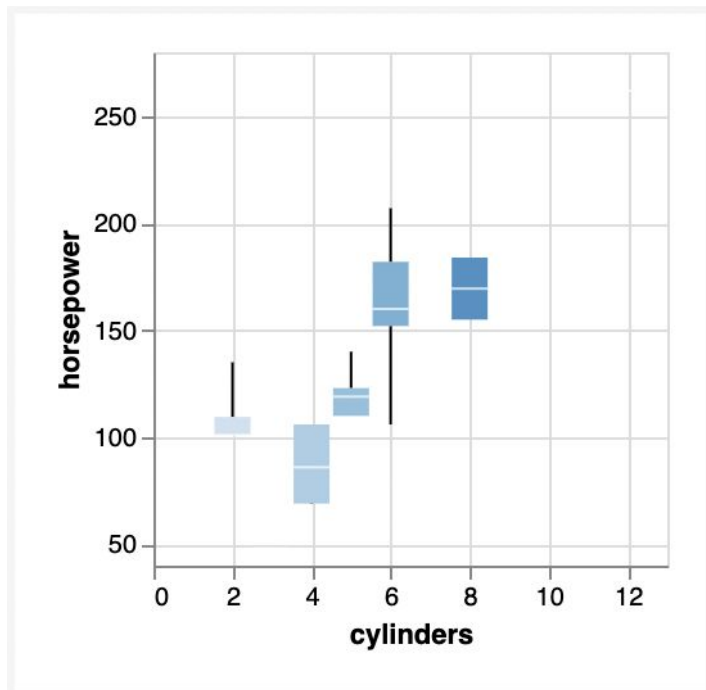
Code:

```
{
  "$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": "point",
  "encoding": {
    "x": {"field": "num-of-cylinders-numerical", "type": "quantitative"},
    "y": {"field": "engine-size", "type": "quantitative"}
  }
}
```

Preview on [Vega.GitHub.IO](https://vega.github.io)

# Horsepower vs Number of Cylinders

The boxplot visualization below shows a positive correlation between the number of cylinders in an engine and its horsepower. Having more cylinders would likely translate into an increase in horsepower, with several rather powerful (above 100hp) two-cylinder engines being an exception of that rule.



Examples of questions that this chart could answer:

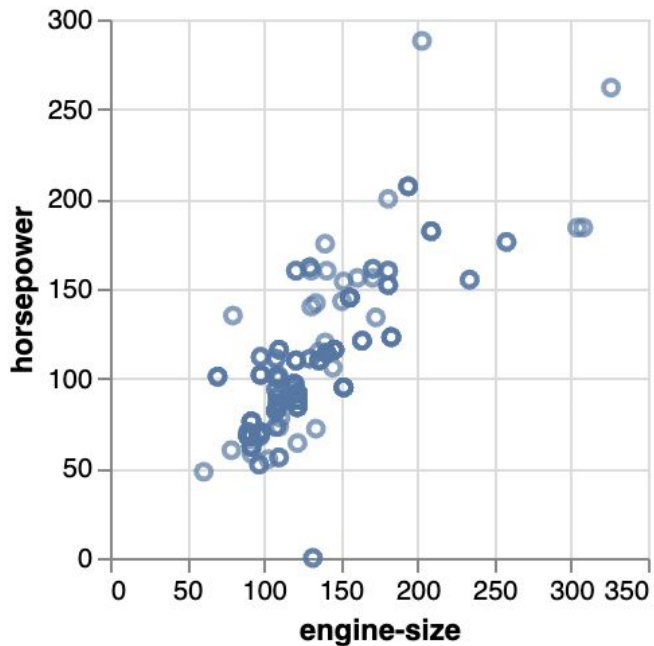
1. What is the relationship between the number of cylinders and horsepower? - *positive correlation: when number of cylinders increases, the horsepower increases as well*
2. How many cylinders do the cars with the most horsepower have? - *8, as the only car with 12 cylinders was not visualized*
3. What is the median horsepower of the cars with 4 cylinders? - *approximately 87 hp*

Code:

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json" ,
  "data": { "url":
    "https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/car_s_parallel.csv" },
  "mark": { "type": "boxplot", "extent": "min-max" },
  "encoding": {
    "x": { "field": "cylinders", "type": "quantitative" },
    "color": { "field": "cylinders", "type":
      "quantitative", "legend": null },
    "y": { "field": "horsepower", "type": "quantitative", "scale":
      { "zero": false } }
  }
}
```

# Horsepower vs 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? - *positive correlation*
2. In what range of engine-sizes does the majority of the cars in the dataset fall? - *under 150*

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"}  
  }  
}
```

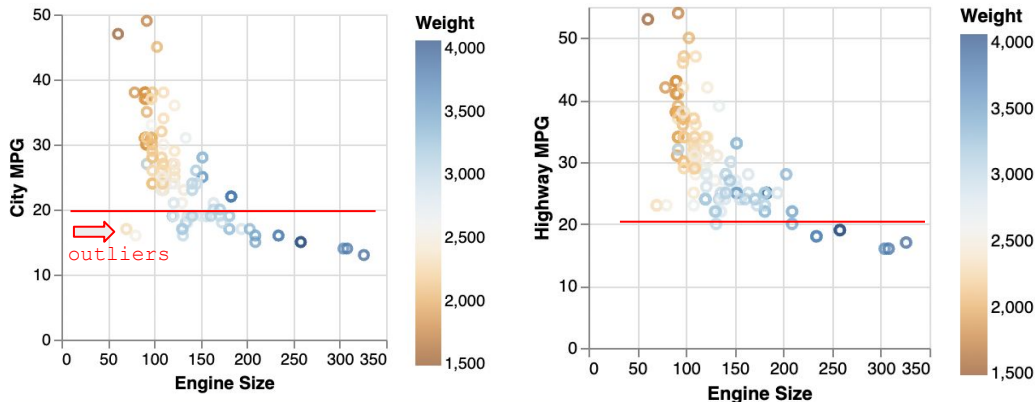
Preview on [Vega.GitHub.IO](https://vega.github.io)

# Engine Size/Weight vs MPG

The scatter plot below shows a negative correlation between the size of the engine and mileage per gallon one should expect from it. Smaller engines usually get a much higher MPG and when the size of the engine increases the MPG is getting lower (with an exception of several outliers indicated on the chart).

Consequently, an increase in engine size leads to an overall increase in a curb-weight of the car (see the color bar on the right for reference)

Using 20 miles-per-gallon as a threshold allows us to see that highway MPG is noticeably higher across the board, with a vast majority of the cars in the dataset usually showing above 20 MPG (indicated on the chart with a red line).



Example of questions that could be answered:

1. What is the relationship between the engine-size and MPG?  
- *negative correlation*
2. What is the approximate city MPG of the largest engine? - *12*
3. What is the approximate highway MPG of the smallest engine? - *53*

Code:

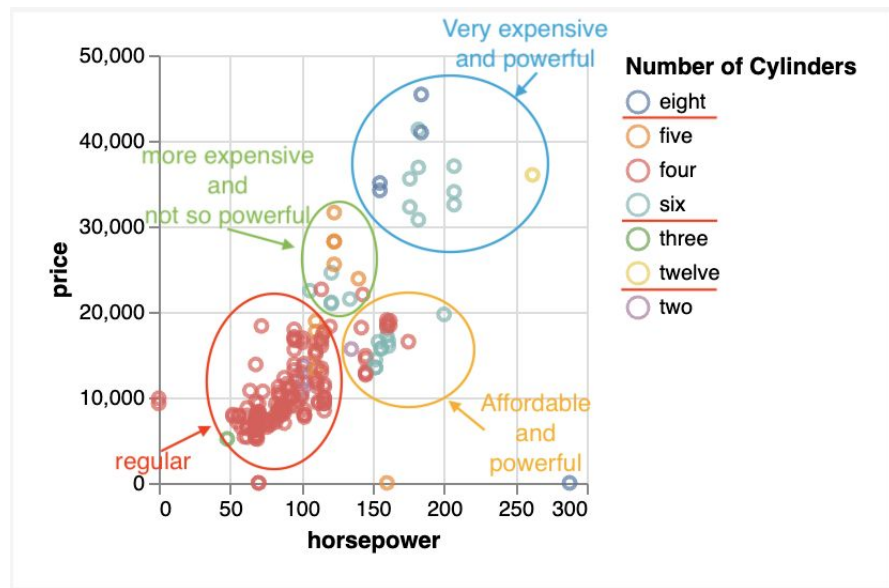
```
{
  "$schema":
    "https://vega.github.io/schema/vega-lite/v5.json",
  "data": {"url":
    "https://raw.githubusercontent.com/mivan000/dse_i2700/main/hwl/cars_clean.csv"},
  "mark": "point",
  "encoding": {
    "x": {"field": "engine-size", "type": "quantitative",
    "title": "Engine Size"},
    "y": {"field": "highway-mpg", "type":
    "quantitative", "title": "Highway MPG"},
    "color": {"field": "curb-weight", "type":
    "quantitative", "title": "Weight",
    "scale": {"domainMid": 2555, "range": "diverging"}}
  }}
```



# Horsepower (& number of Cylinders) vs Price of the Car

On top of higher MPG (see previous charts) 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? - *four*
2. Name engine types in terms of numbers of cylinders that are the most expensive - *six, eight, twelve*

Code:

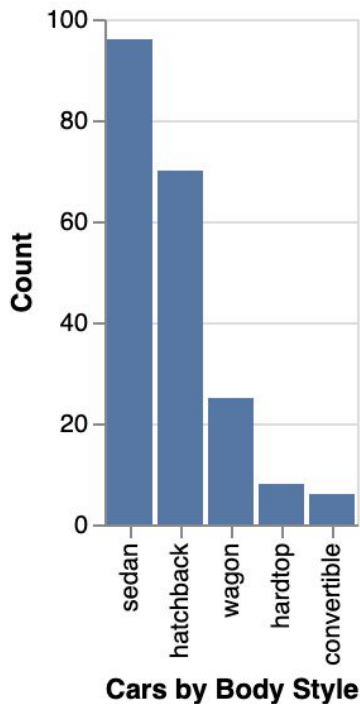
```
{  
  "$schema":  
    "https://vega.github.io/schema/vega-lite/v5.json",  
  "data": {  
    "url":  
      "https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/cars_clean.csv"  
  },  
  "mark": "point",  
  "encoding": {  
    "y": {"field": "price", "type": "quantitative"},  
    "x":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

# Body Styles

Categorical vs. Quantitative

# Cars by Body Style

A bar chart breaking down the cars in the dataset by their body type shows that the vast majority of vehicles are sedans and hatchbacks, Wagons are represented much less and a very few of the cars are hardtops and convertibles.



Example of questions that could be answered:

1. What is the least/most popular body type? - [convertible/sedan](#)
2. Approximately how many cars are sedans? - [97](#)

Code:

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

} Preview on [Vega.GitHub.IO](#)

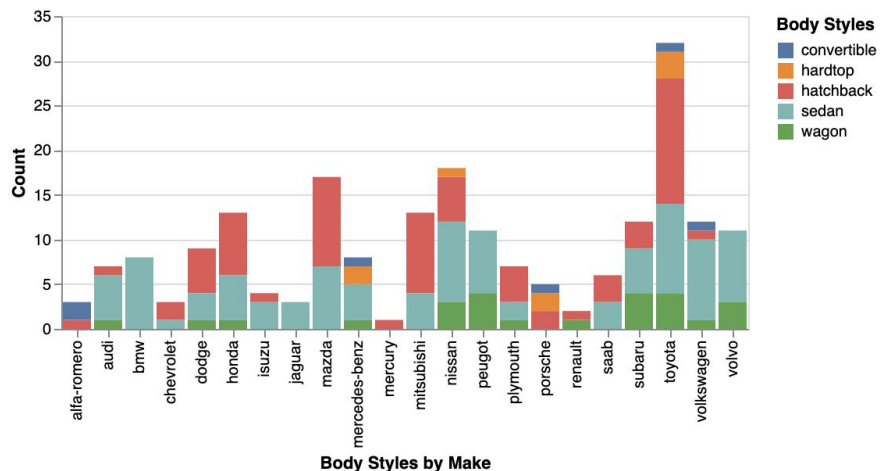
# Body Styles by Make

A more detailed view with each car make broken out by body style:

- Sedans are dominating the market (18 brands have them). Two automakers (BMW and Jaguar) produce sedans exclusively and only 4 do not have one in their lineup (Alfa Romeo, Mercury, Porsche, Renault).
- Hatchbacks follow very closely with 17 automakers producing them. Several brands even make more hatchbacks than cars of any other body style: Dodge, Honda, Mazda, Mitsubishi, Plymouth, and Toyota.
- Wagons are not as popular but 12/22 brands (55%) have them in their lineups.
- Hardtops and convertibles seem to be niche products with only 4 and 5 automakers respectively producing at least one of each.

Hardtops: MB, Nissan, Porsche, Toyota.

Convertibles: Alfa-Romeo, MB, Porsche, Toyota, Volkswagen.



Example of questions that could be answered:

1. Which automaker produces the highest number of cars?  
- *Toyota*
2. Which automaker produces most of the Convertibles?  
- *Alfa-Romeo*

Code:

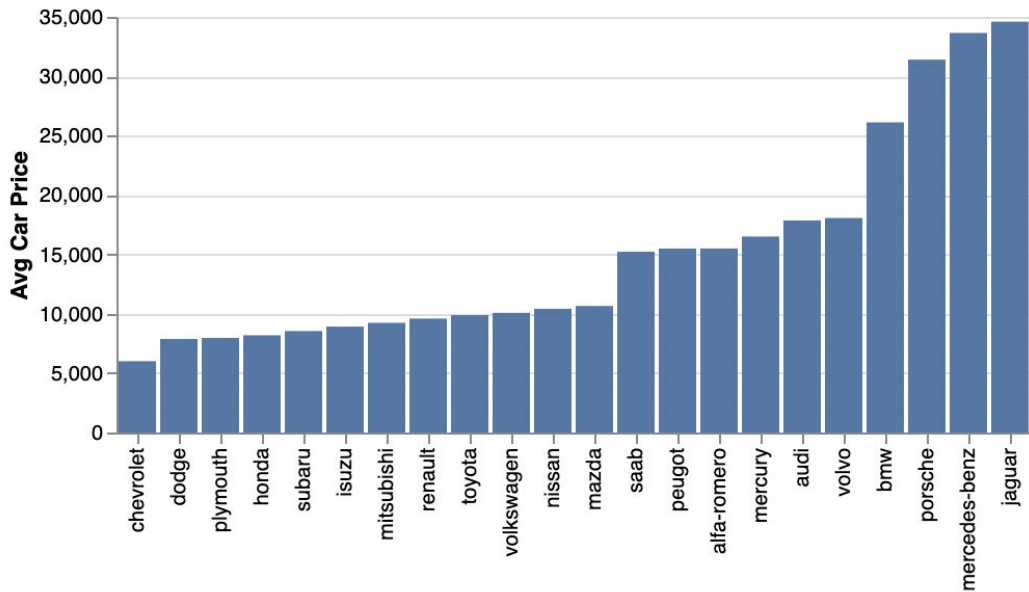
```
{ "$schema": "https://vega.github.io/schema/vega-lite/v5.json" ,  
  "data": { "url":  
    "https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/  
cars_clean.csv"  
  }, "mark": "bar",  
  "encoding": {  
    "x": { "field": "make", "type": "nominal",  
      "title": "Body Styles by Make"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

Preview on [Vega.GitHub.IO](https://vega.github.io)

# Average Car Price by Make

A bar chart visualising average price of a car by make allows us to easily identify three distinctive clusters in the dataset:

- **Average price above above \$25K:** BMW, Porsche, Mercedes-Benz, Jaguar
- **\$11K - 20K:** Saab, Peugeot, Alfa-Romeo, Mercury, Audi, Volvo
- **Under ~\$11K:** the rest of the car manufacturers



Preview on [Vega.GitHub.IO](https://vega.github.io)

Example of questions that could be answered:

1. Name 3 automakers producing the most expensive cars on average. - *Porsche, Mercedes-Benz, Jaguar*
2. Which brand on average produces the most affordable cars in the dataset? - *Chevrolet*

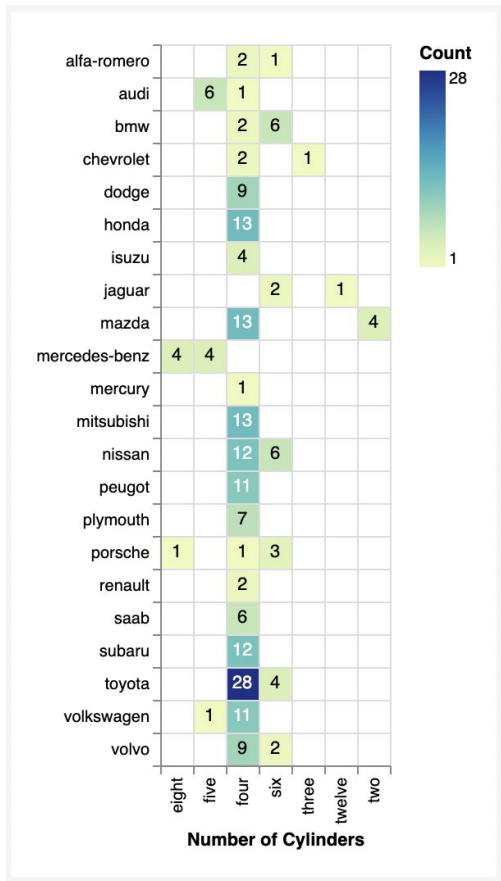
Code:

```
{ "$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",  
      "sort": "y", "title": "" },  
    "y": { "field": "price", "aggregate":  
      "mean", "type": "quantitative", "title": "Avg Car Price" }  
  }  
}
```

# Car Makes

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?  
- *Toyota*
2. What is the most popular engine size in terms of number of cylinders? - *four*

```
{ "$schema": "https://vega.github.io/schema/vega-lite/v5.json" ,  
  "data": { "url":  
    "https://raw.githubusercontent.com/mivan000/dse_i2700/main/hwl/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](https://vega.github.io/vega-lite/)

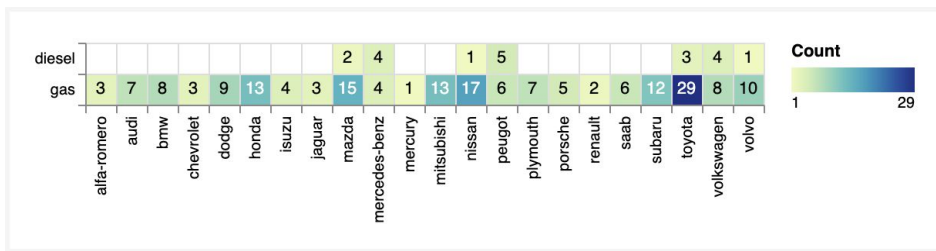
# Make by Fuel Type

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 that are represented in the dataset being diesel (5 out of 6).

Example of questions that could be answered:

1. Which automakers produce diesel cars?  
- *Mazda, Mercedes-Benz, Nissan, Peugeot, Toyota, Volkswagen, Volvo*
2. Which automaker has the highest relative share of diesel cars compared to all cars in their lineup?  
- *Mercedes-Benz, 50%*



```
{ "$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", "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
```



# Findings

There are different kinds of relationship between variables in dataset, please see some below:

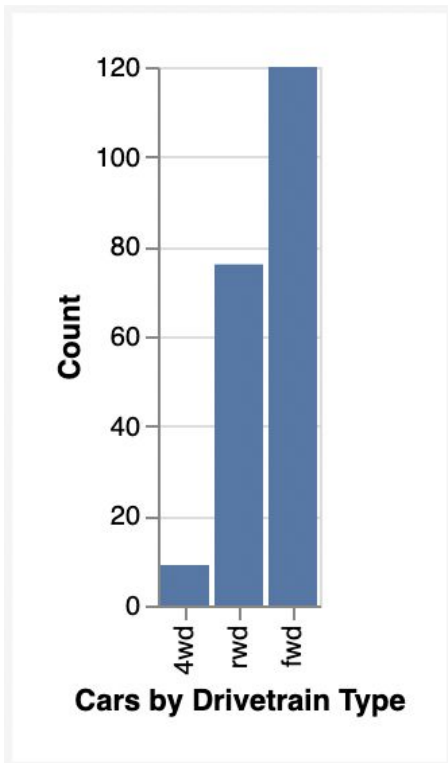
1. Engines with larger number of cylinders are more powerful and usually have much lower MPG.
2. Engines with a lower number of cylinders tend to be smaller and lead to the curb-weight of the car itself being lower.
3. There's also a strong relationship between horsepower and a price of a car when cars with more horsepower usually being more expensive.
4. Gasoline powered cars with four-cylinder engines are much more popular than diesel cars.
5. The most popular body styles of cars represented in the dataset are sedans and hatchbacks.

# Drivetrains

Bonus Categorical vs. Quantitative section

# 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?

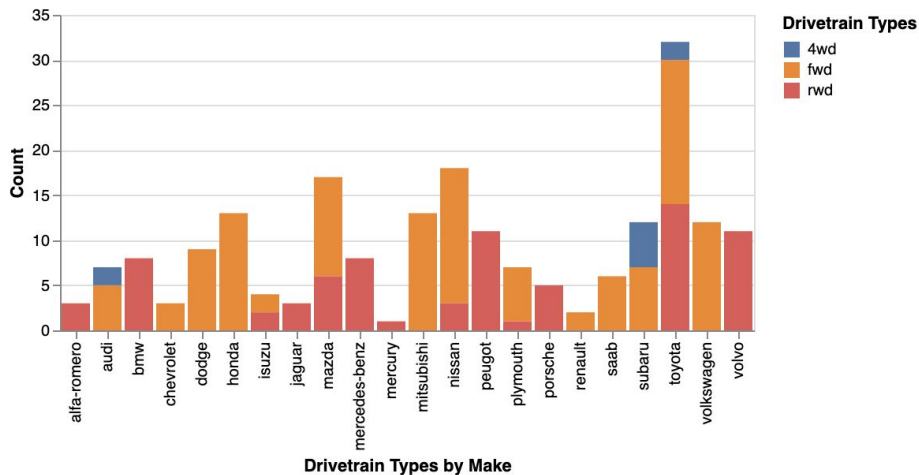
Code:

```
{ "$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?

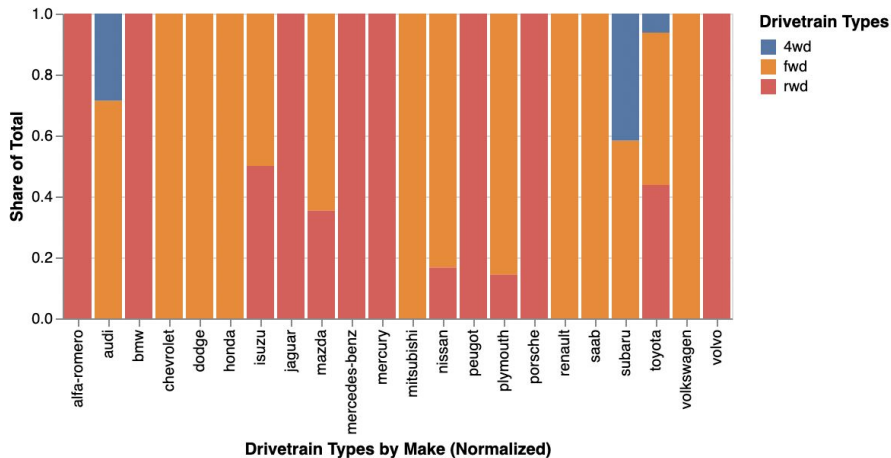
Code:

```
{ "$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?

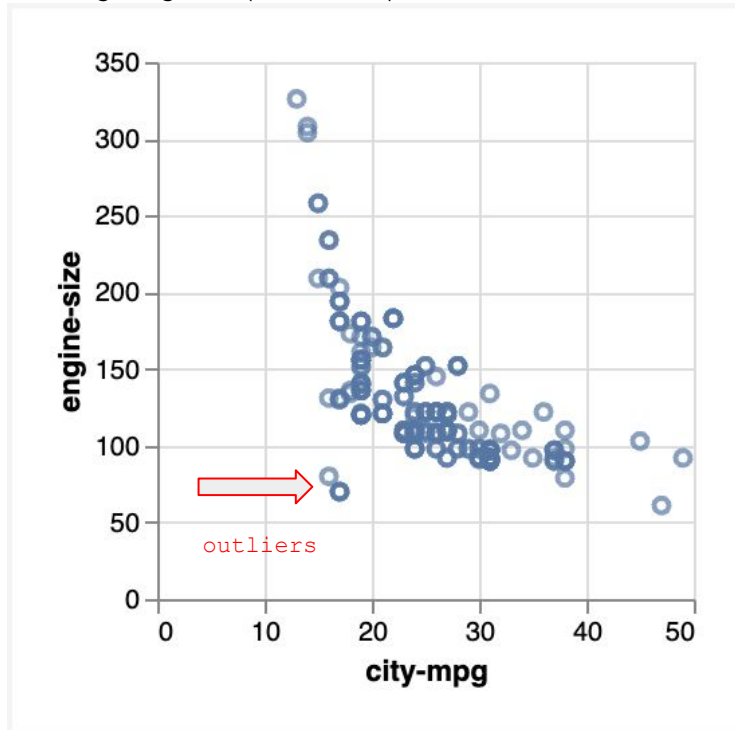
Code:

```
{ "$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](https://vega.github.io)

# Engine Size vs MPG

The scatter plot below shows a negative correlation between the size of the engine and mileage per gallon one should expect from it. Smaller engines usually get a much higher MPG and when the size of the engine increases the MPG is getting lower (with an exception of two outliers indicated on the crart).



Example of questions that could be answered:

1. What is the relationship between the engine-size and city MPG?  
- *negative correlation*
2. What is the approximate MPG of the largest engine? - *12*
3. What is the approximate MPG of the smallest engine? - *47*

Code:

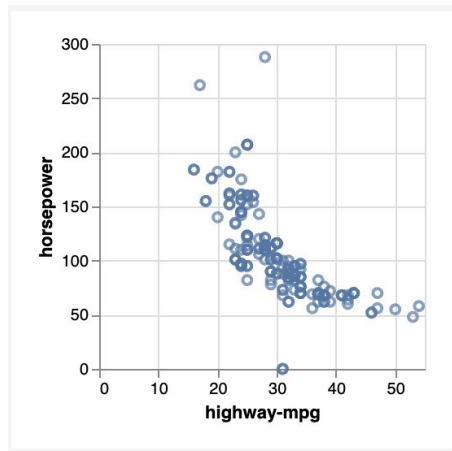
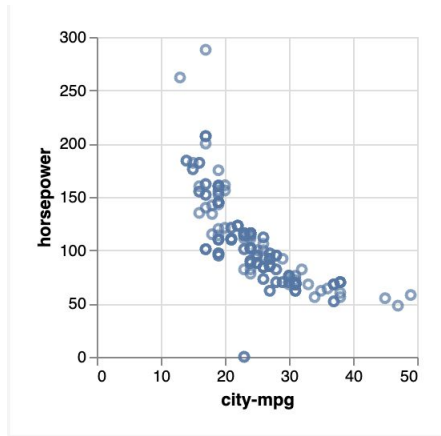
```
{
  "$schema":
    "https://vega.github.io/schema/vega-lite/v5.json",
  "data": {
    "url":
      "https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw
1/cars_clean.csv"
  },
  "mark": "point",
  "encoding": {
    "x": {"field": "city-mpg", "type": "quantitative"},
    "y": {"field": "engine-size",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

Preview on [Vega.GitHub.IO](https://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 (indicated on the chart with a red line).



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:

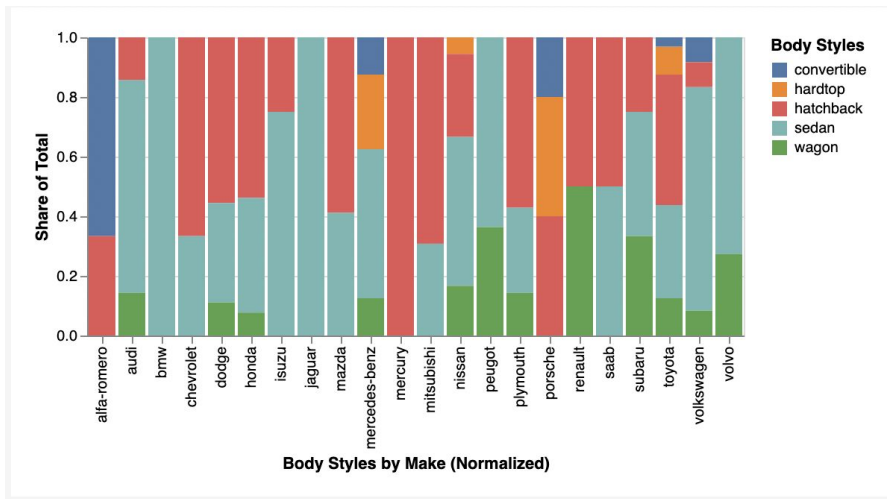
```
{
  "$schema":
    "https://vega.github.io/schema/vega-lite/v5.json",
  "data": {
    "url":
      "https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/cars_clean.csv"
  },
  "mark": "point",
  "encoding": {
    "x": {"field": "city-mpg", "type": "quantitative"},
    "y": {"field": "horsepower",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

Preview on [Vega.GitHub.IO - city](https://vega.github.io/vega-lite/?data=https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/cars_clean.csv), [highway](https://vega.github.io/vega-lite/?data=https://raw.githubusercontent.com/mivan000/dse_i2700/main/hw1/cars_clean.csv)

# Body Styles by Make (Normalized)

Normalized version of the same chart allows us to have a clearer picture of a share of a given body style in each automaker's lineup:

- 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:

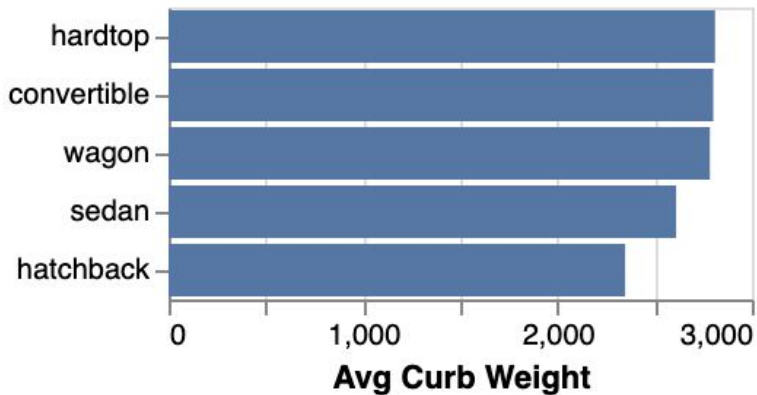
```
{ "$schema":  
  "https://vega.github.io/schema/vega-lite/v5.json",  
  "data": { "url":  
    "https://raw.githubusercontent.com/mivan000/dse_i2700/main  
/hw1/cars_clean.csv"  
  }, "mark": "bar",  
  "encoding": {  
    "x": {  
      "field": "make", "type": "nominal",  
      "title": "Body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