

Midterm Exam

ST 597 / Spring 2016
University of Alabama

Cars Data

The file <http://bama.ua.edu/~mdporter2/st597/data/cars.txt> contains a sample of US automobiles that were for sale in the 1990's.

1. Read the data into R creating a data frame named `cars`.
2. Fix the `weight` column by removing the extra *lbs* characters and converting to a number (integer or numeric).
3. Add a new column to the `cars` data frame that calculates torque according to the equation

$$\text{horsepower} = \frac{\text{torque} \times \text{rpm}}{5252}$$

and name it `torque`.

Descriptive Statistics

1. What proportion of cars are `Small` (`type` column)?
2. How many cars have a highway MPG (`mpg.highway`) at least 40% (1.4 times) greater than city MPG (`mpg.city`)?
3. Calculate the average `price` and median `rpm` for each `type` of car. Repeat for the US cars (`origin` is USA).
4. Explain why `a` is different than `b`?

```
a = cars %>% filter(price >= mean(price))
b = cars %>% group_by(origin) %>% filter(price >= mean(price))
```

5. Calculate the mean `luggage.room`.

Graphics

1. Create a scatterplot of the `cars` data with:
 - `mpg.city` on the x-axis and `price` on the y-axis
 - color the points according to `type`
 - set the shape of the points according to `origin`
 - set the size of the points according to `rpm`
 - add a smooth curve fit with a line color of `red` and fill color of `yellow`

2. Generate this bar graph of the frequency of **type** faceted by **origin**. Note that the order of the bars corresponds to the average **enginesize** of the car type.

