Intro to R Lab 2: Solutions

Simon Caton

Prep work

```
data(mtcars) #loads the built-in dataset
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$am <- factor(mtcars$am, labels=c("Automatic", "Manual"), levels=c(0,1))
```

1. Are there more automatic or manual cars?

```
table(mtcars$am)
```

```
##
## Automatic    Manual
##        19        13
```

2. Which car is heaviest, and which is fastest?

```
maxWeight <- max(mtcars$wt)
(mtcars[mtcars$wt>=maxWeight, ])
```

```
##                      mpg cyl disp  hp drat    wt  qsec vs        am gear
## Lincoln Continental 10.4   8  460 215    3 5.424 17.82  0 Automatic    3
##                     carb
## Lincoln Continental    4
```

```
minQSec <- min(mtcars$qsec)
(mtcars[mtcars$qsec==minQSec, ])
```

```
##                 mpg cyl disp  hp drat   wt qsec vs     am gear carb
## Ford Pantera L 15.8   8  351 264 4.22 3.17 14.5  0 Manual    5    4
```

Note putting brackets around an expression will print the results after it has been executed.

3. Do automatic or manual cars have on average a better mpg?

```
(meanMPG <- tapply(mtcars$mpg, mtcars$am, mean))
```

```
## Automatic    Manual
##  17.14737  24.39231
```

4. How many cars have above average hp?

```
meanHP <- mean(mtcars$hp)
aboveMeanHP <- mtcars[mtcars$hp > meanHP, ]
dim(aboveMeanHP)
```

```
## [1] 15 11
```

So we have 15 cars that have above average hp.

5. Of the cars that have above average hp, how many have 6 cylinders?

```
moreThanSix <- aboveMeanHP[aboveMeanHP$cyl == 6, ]
dim(moreThanSix)
```

```
## [1]  1 11
```

So we have 1 cars that have above average hp and 6 cylinders.

6. Of the cars that have above average hp, and 6 cylinders, how many are automatic?

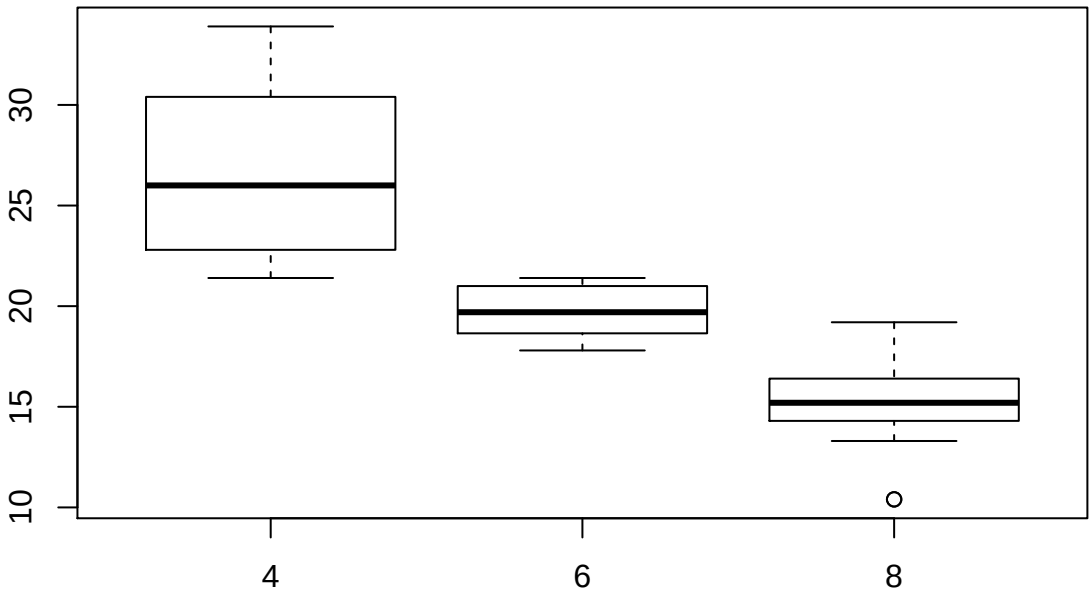
```
automatics <- moreThanSix[moreThanSix$am == "Automatic", ]
dim(automatics)
```

```
## [1]  0 11
```

So we have 0 automatic cars that have above average hp and 6 cylinders.

7. Make a boxplot of mpg split by no. of cylinders

```
boxplot(mpg ~ cyl, data=mtcars)
```



8. In the above boxplot are there outliers?

It looks like there aren't really any outliers, just the one with a very low MPG and 8 cylinders.

9. Take a random 50% sample of the dataset, and rerun questions above to see what changes

```
mysample <- mtcars[sample(1:nrow(mtcars), nrow(mtcars)/2, replace=FALSE),]
```

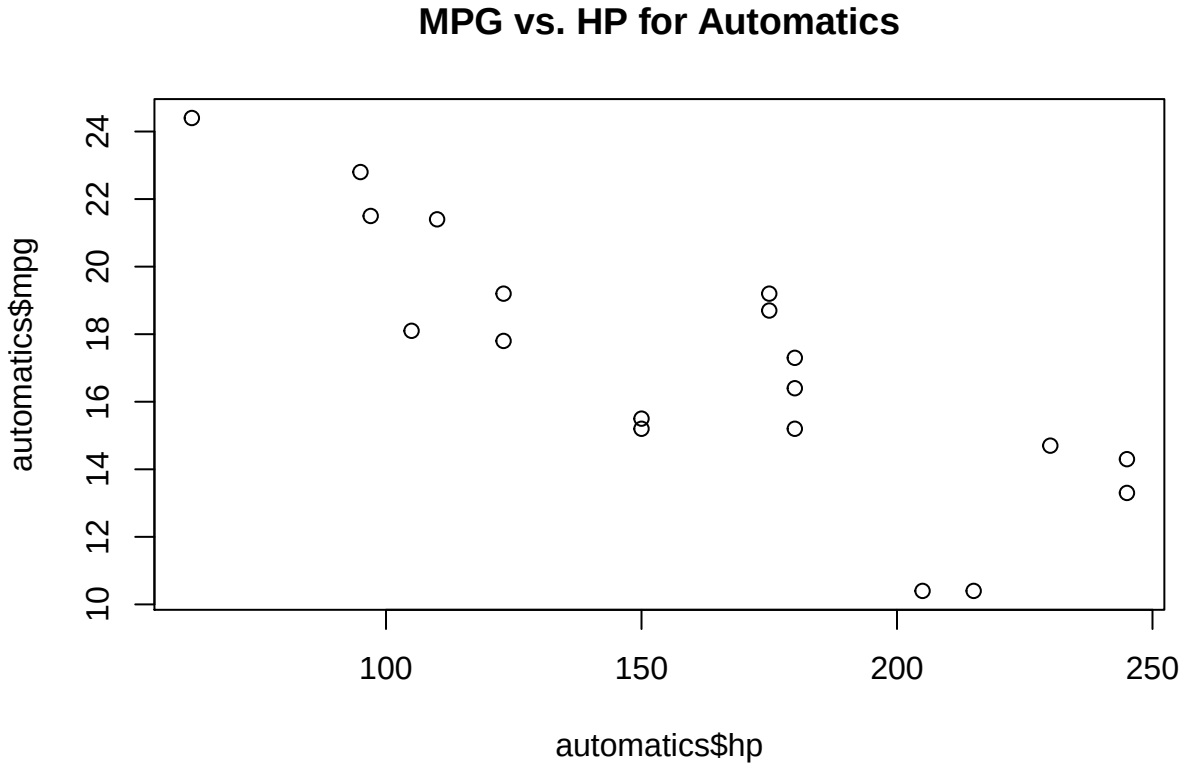
Rerun all solutions above, replacing all instances of *mtcars* with *mysample*.

10. Split the dataset in 2, one half containing only automatics, and one half containing only manual transmissions

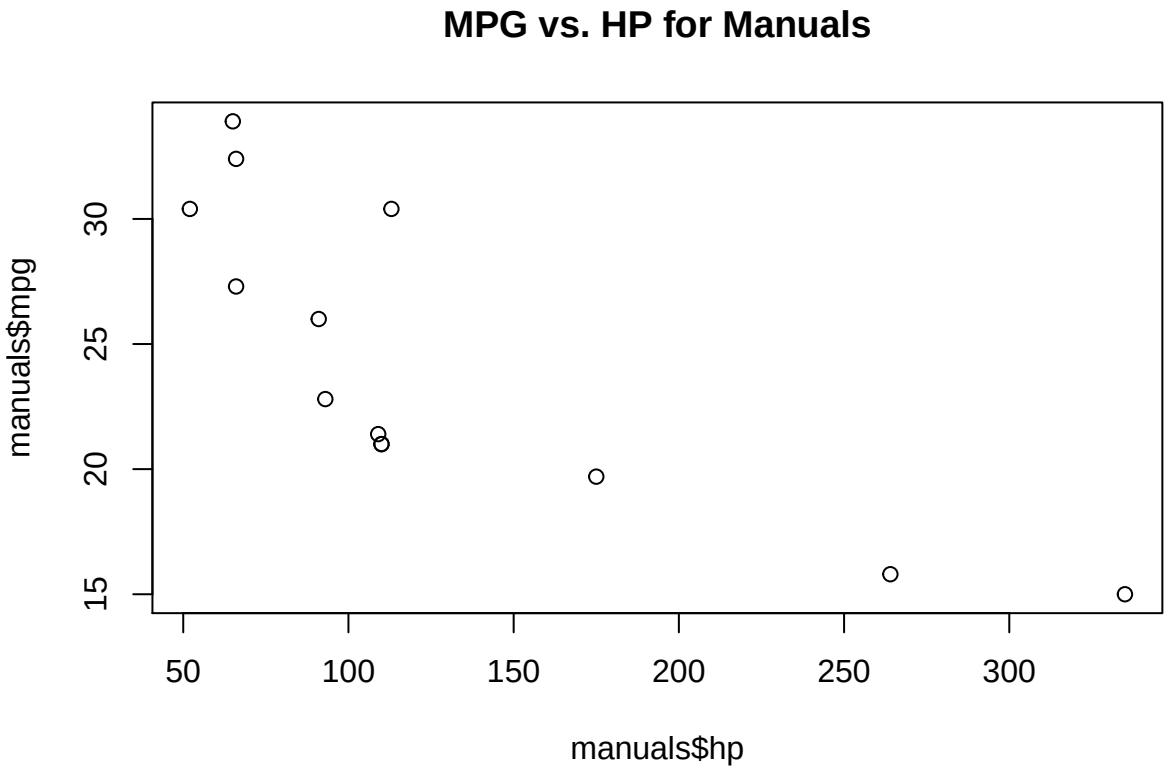
```
automatics <- mtcars[mtcars$am == "Automatic", ]
manuals <- mtcars[mtcars$am != "Automatic", ]
```

11. For each half, plot mpg against hp using the plot function

```
plot(automatics$hp, automatics$mpg, main="MPG vs. HP for Automatics")
```

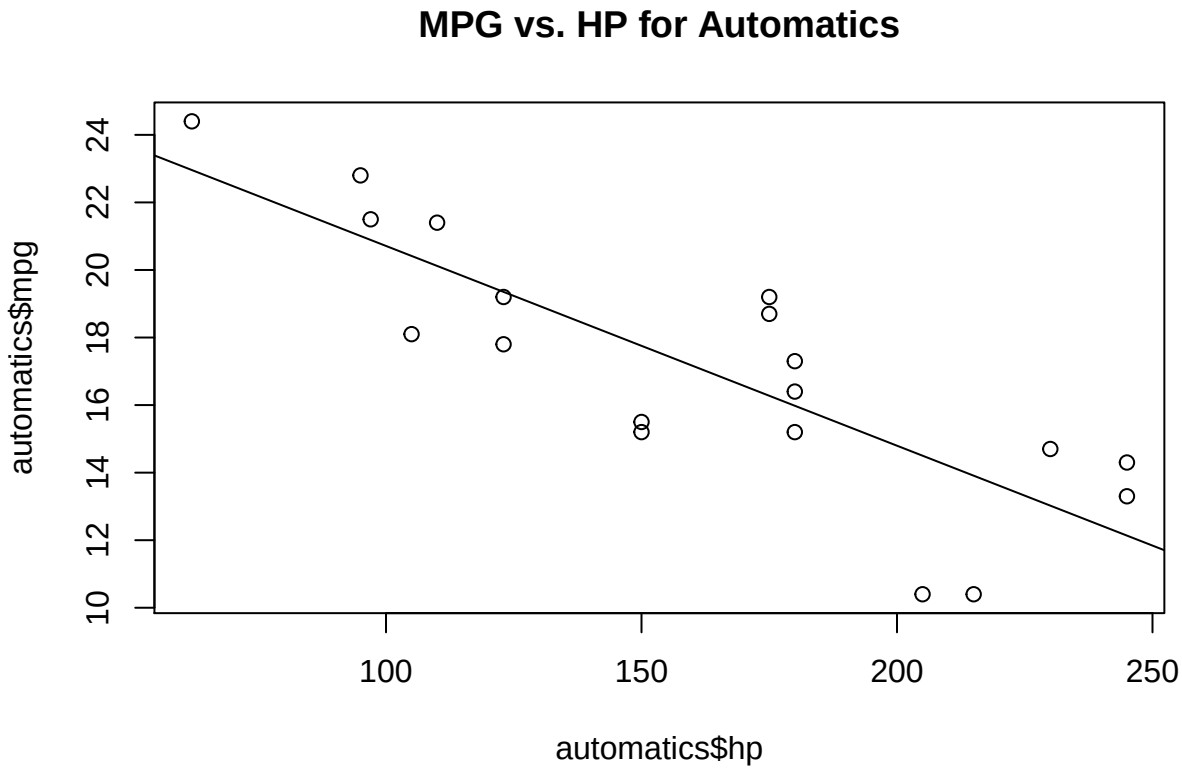



```
plot(manuals$hp, manuals$mpg, main="MPG vs. HP for Manuals")
```



12. Add an "abline" that "fits" a linear model between the two variables in the plots you just drew

```
lmAutomatic <- lm(mpg ~ hp, automatics)
plot(automatics$hp, automatics$mpg, main="MPG vs. HP for Automatics")
abline(lmAutomatic)
```




```
lmManual <- lm(mpg ~ hp, manuals)
plot(manuals$hp, manuals$mpg, main="MPG vs. HP for Manuals")
abline(lmManual)
```