Regression Models Course Project

Executive Summary

This project looks at the mtcars dataset to determine whether gas mileage is better for manual transmission cars or automatic transmission cars. While the mean gas milage for manual transmission cars is substantially higher than the gas milage for automatic, the correlation is fairly low. Evidence for a distinct higher mpg performance for manual drive cars compared with other more correlated variables was inconclusive.

Exploratory Analysis

As a first measure, we take the mean of the gas mileage for automatic transmission cars and compare it to the mean for manual transmission cars.

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

```
##     0     1
## 17.15 24.39
```

The mean gas mileage for automatic transmission is about 17 mpg, and for manual it is about 24 mpg. this is a difference of 7 mpg, or 41 percent.

However, there seems to be correlations between engine displacement and mpg, number of cylinders and mpg. A quick computation of correlations shows that there are stronger correlations between other variables and mpg than transmission type and mpg:

```
ghostmpg <- mtcars
for (i in 2:11) {
     ghostmpg[i]<-mtcars$mpg
}
lotsa <- cor(ghostmpg, mtcars)
correlations <- lotsa[1,]
correlations
```

```
##     mpg     cyl    disp      hp    drat      wt    qsec      vs      am
##  1.0000 -0.8522 -0.8476 -0.7762  0.6812 -0.8677  0.4187  0.6640  0.5998
##    gear    carb
##  0.4803 -0.5509
```

The correlations show that the largest magnitude correlation is to wt (a negative correlation), followed by number of cylinders.

The remaining analysis will compare linear regressions of wt vs. mpg, for automatic and manual transmissions, and see if there is a difference bewteen the automatic transmission cars and manual transmission cars.

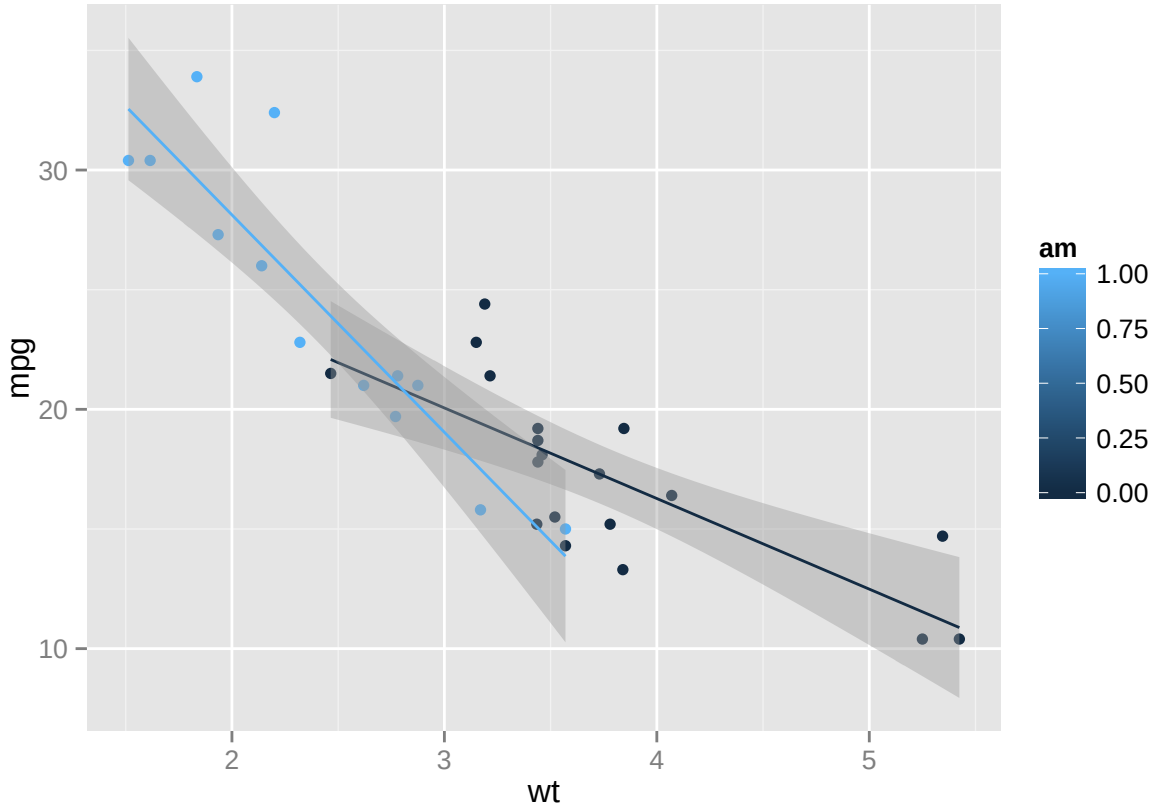
Comparing linear regressions for manual and automatic

To keep things simple we split the dataset:

```
xmissSplit <- split(mtcars, as.factor(mtcars$am))
autoCars <- as.data.frame(xmissSplit[1])
manualCars <- as.data.frame(xmissSplit[2])
names(autoCars) <- names(mtcars)
names(manualCars) <- names(mtcars)
sortCars <- rbind(autoCars, manualCars)
```

Next we'll plot wt vs mpg, grouping by transmission type:

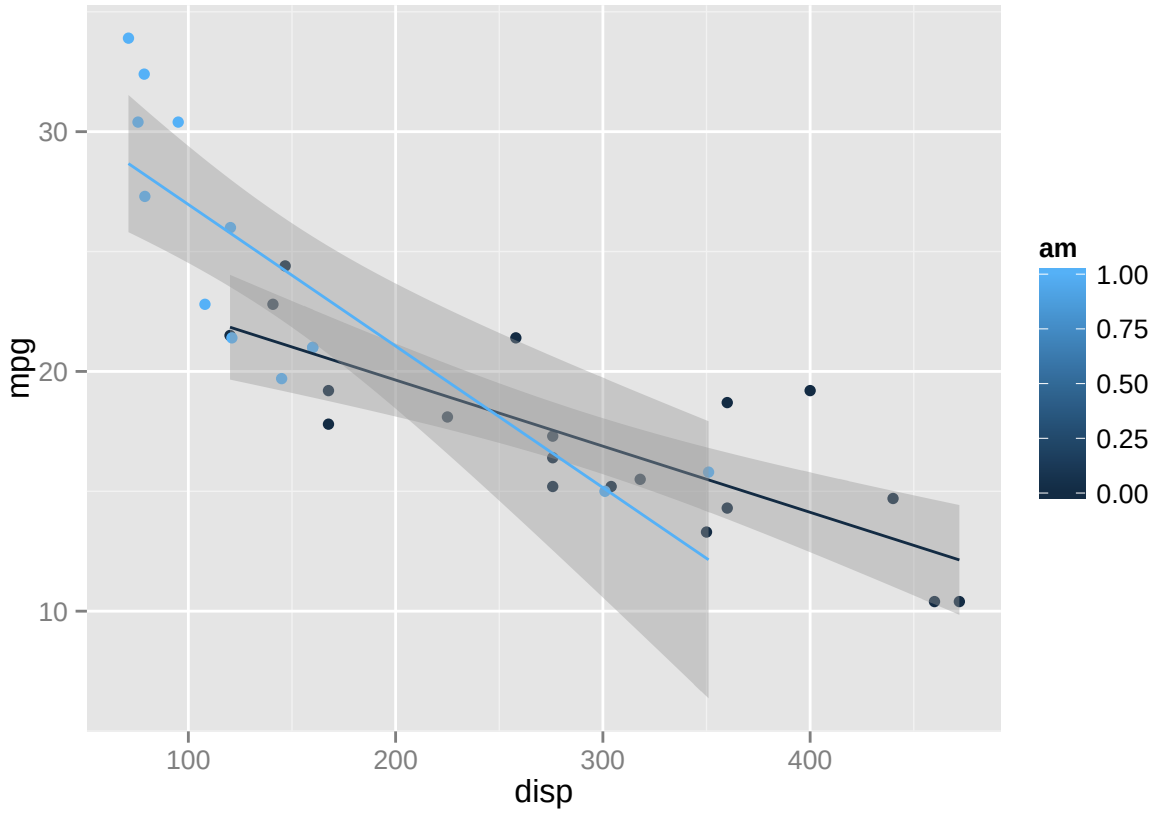
```
library(ggplot2)
qplot(data=sortCars,x=wt,y=mpg,color=am)+stat_smooth(formula = y~x, method="lm", data=autoCars)+stat_sm
```



The plots include linear regression lines with 95% confidence intervals for predicted values.

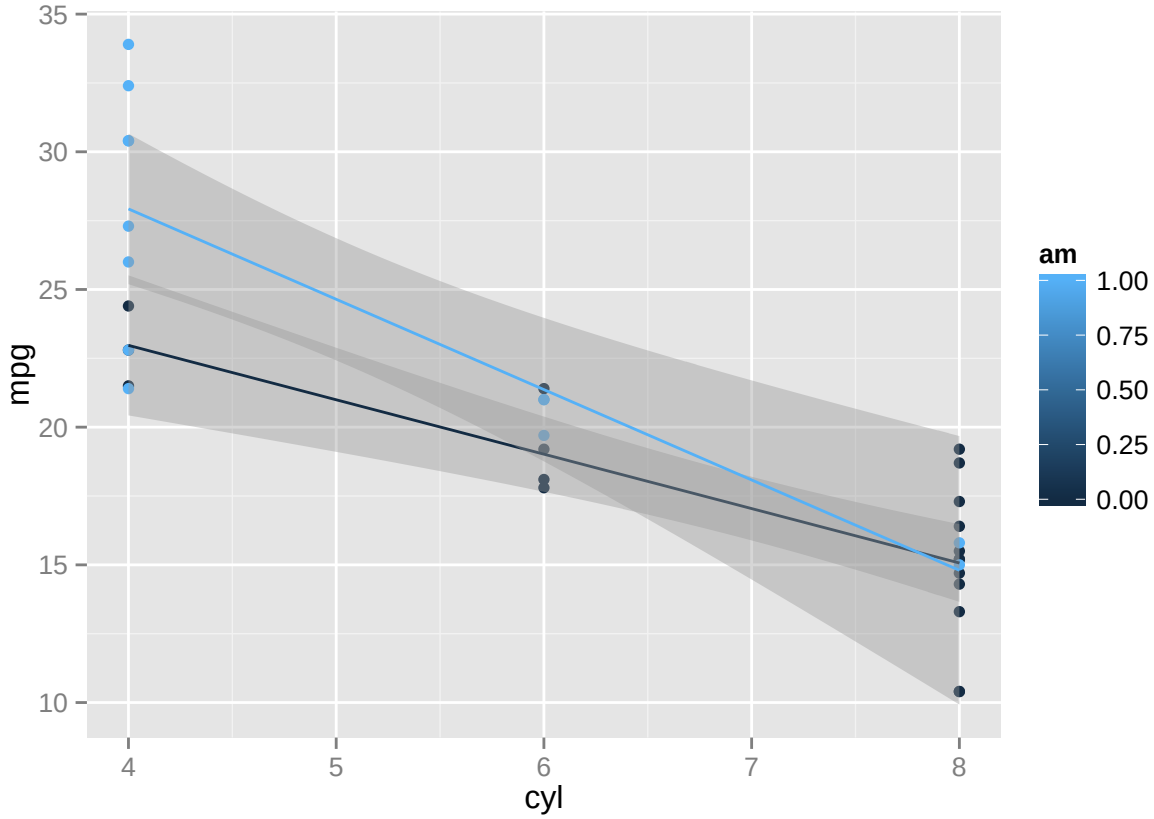
Because there is not much overlap between wt data and transmission data (the manuals are skewed to low weight, and the automatics are skewed to high wt), we will perform the same exercise with two other highly correlated variables, disp (engine size) and cyl (number of cylinders).

```
qplot(data=sortCars,x=disp,y=mpg,color=am)+stat_smooth(formula = y~x, method="lm", data=autoCars)+stat_
```



For cylinders:

```
qplot(data=sortCars,x=cyl,y=mpg,color=am)+stat_smooth(formula = y~x, method="lm", data=autoCars)+stat_sm
```



Given the overlap of confidence intervals it seems inconclusive that higher gas milage occurs with manual transmission given the effects of weight, displacement, or number of cylinders.