

# John Hopkins University – Data Science Specialization – Regression Models Course – Appendix to Solution of Project 1

*Dr. Guy Cohen*

*August 19, 2015*

## Effect of Transmission Type on Fuel Economy in 1973-74 Cars

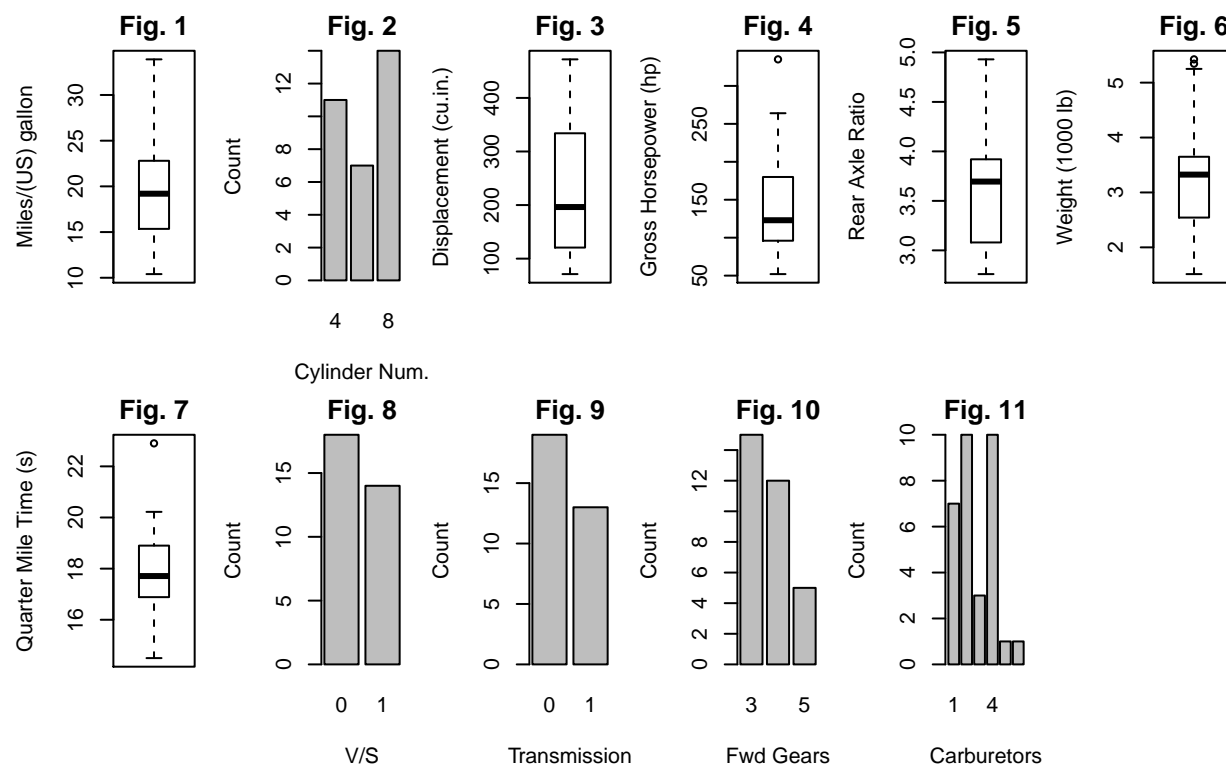
### Appendix

#### Exploratory Data Analysis

We first convert the factor variables to factor type.

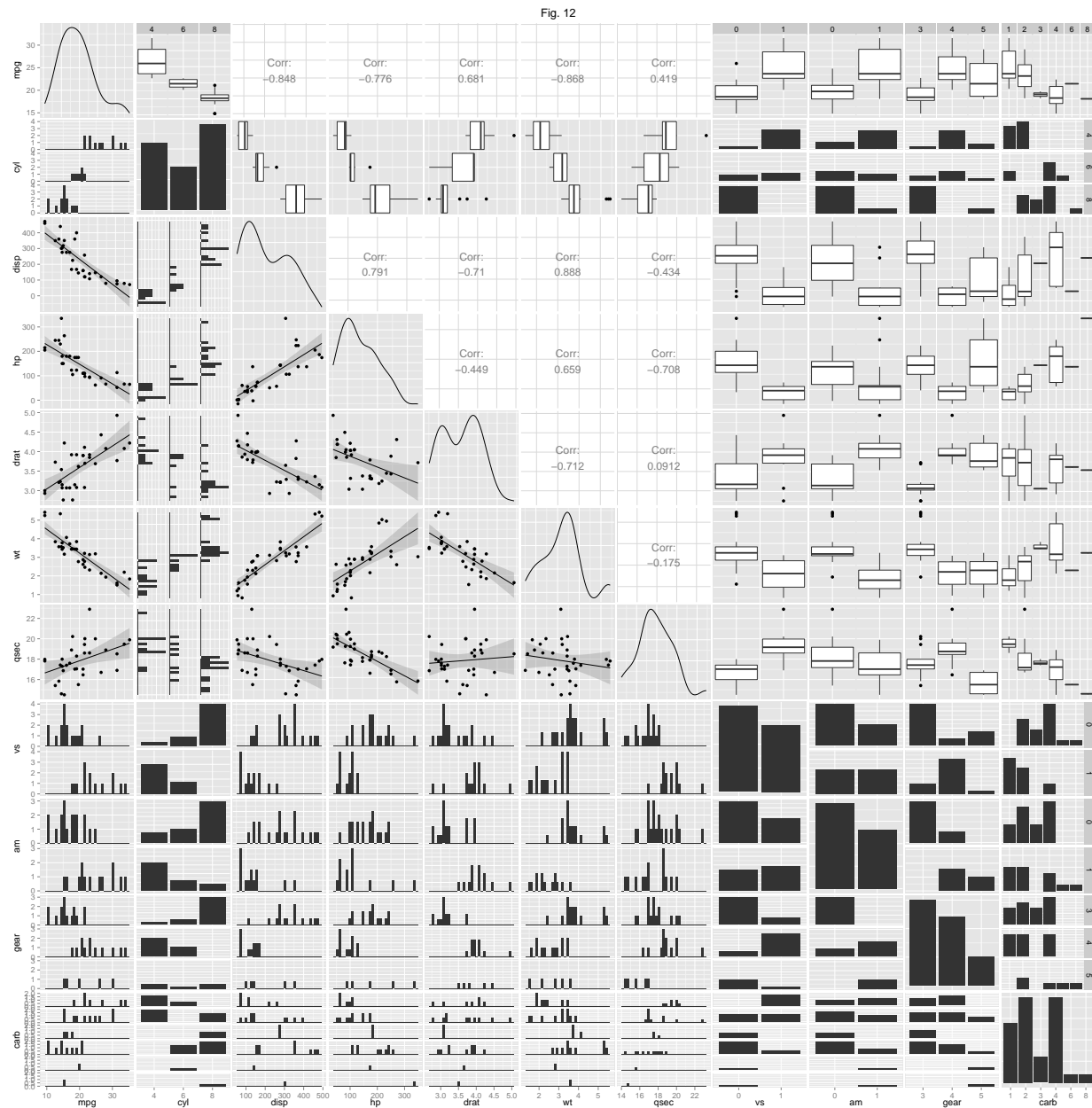
```
data(mtcars)
mtcars2 <- mtcars
mtcars2[,c("cyl", "vs", "am", "gear", "carb")] <-
  as.data.frame(lapply(mtcars2[,c("cyl", "vs", "am", "gear", "carb")], as.factor))
```

Second, we plot boxplots of continuous variables and barplots for categorical ones.



We plot a pairs plot to see how the variables relate to each other.

```
library(ggplot2)
library(GGally)
ggpairs(mtcars2, lower = list(continuous = "smooth"), params = c(method = "loess"), title = "Fig. 12")
```

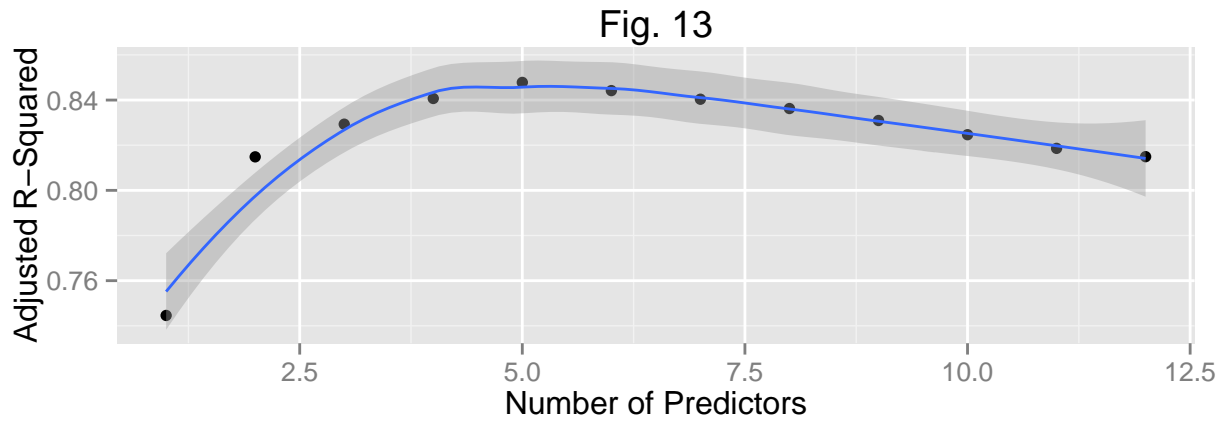


## Detailed Solution

We plot the adjusted R-squared for models of various number of predictors found using forward selection.

```
library(leaps)
library(ggplot2)
regfit.fwd=regsubsets(mpg ~ ., data = mtcars2, nvmax = 12, method="forward")
qplot(x,y,data = data.frame(x=1:12,y=summary(regfit.fwd)$adjr2),
```

```
geom = c("point", "smooth"), xlab = "Number of Predictors",
ylab = "Adjusted R-Squared", main = "Fig. 13")
```



We fit a linear model using the parameters found and plot the residual plot.

```
fit <- lm(mpg ~ I(cyl == "6") + hp + wt + I(am == "1"), data = mtcars2)
par(mfrow = c(2,2))
plot(fit)
```

