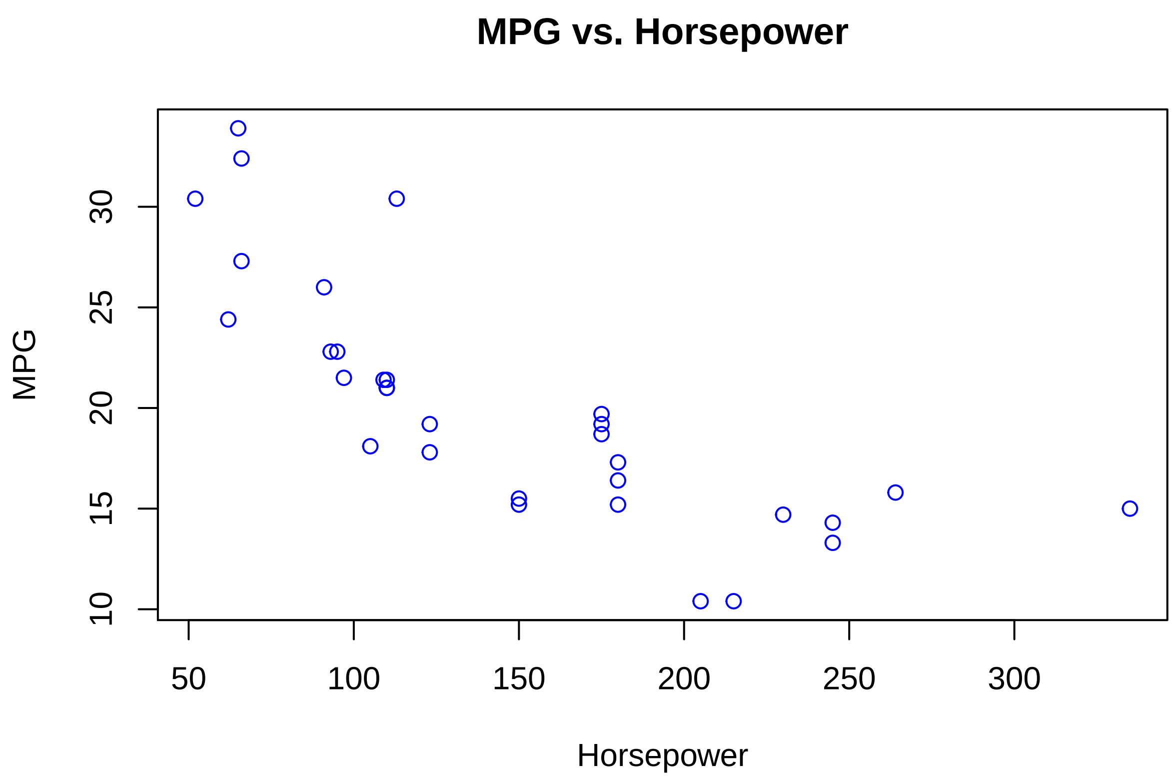
Homework Sample

Chris Conlan

September 17, 2015

Problem 1

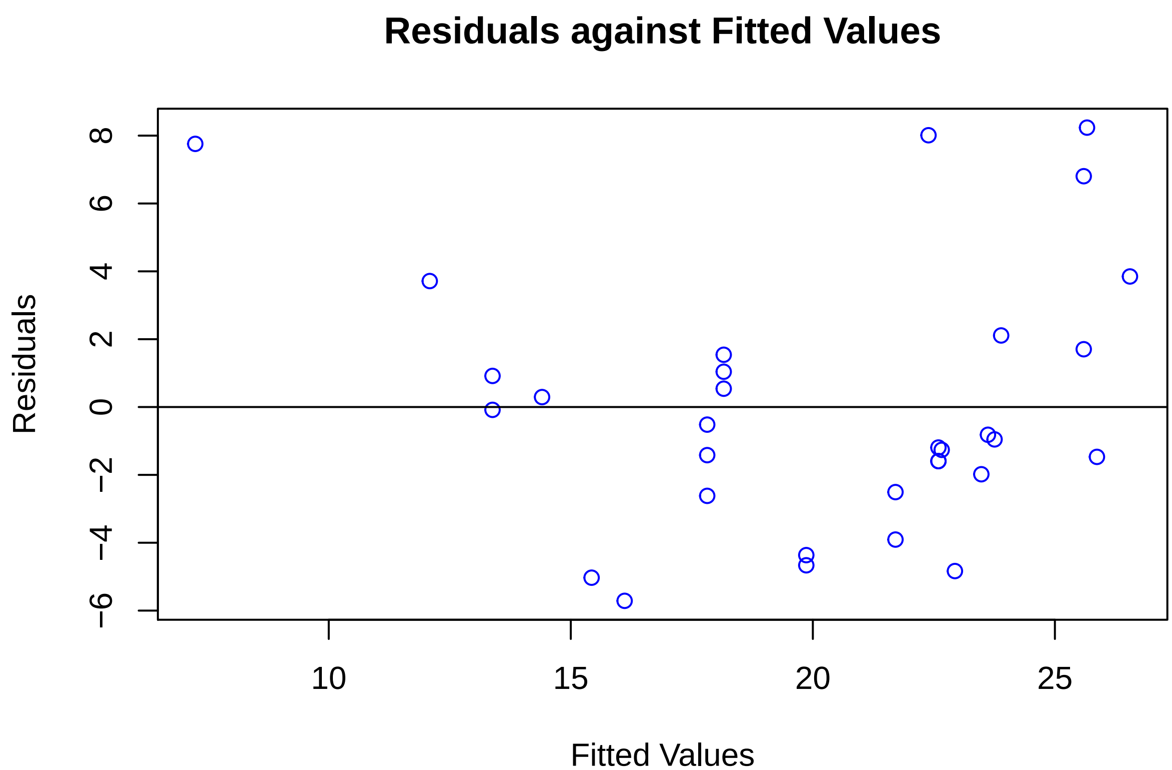
```r
CAR <- mtcars
plot(y = CAR$mpg, x = CAR$hp, col = 4,
     main = "MPG vs. Horsepower",
     ylab = "MPG",
     xlab = "Horsepower")
```



There are outliers in the x-axis in *Horsepower* which will likely have large pull on the regression line. It is probably a good idea to remove outliers to increase accuracy in the shortened domain.

Part (b)

```
model <- lm(mpg ~ hp, data = CAR)
plot(y = model$residuals, x = model$fitted.values, col = 4,
     main = "Residuals against Fitted Values",
     ylab = "Residuals",
     xlab = "Fitted Values")
abline(h = 0, col = 1)
```



Assumptions of mean-zero residuals, constant variance hold well in this plot. There is some shape to the error distribution. It curves up in the right tail end.

Part (c)

The estimated regression coefficient $\hat{\beta}_1 =$ -0.068.

Part (d)

$R^2 = 0.602$.

Part (e)

```
LOWCAR <- CAR[CAR$hp <= 250, ]
newmodel <- lm(mpg ~ hp , data = LOWCAR)
```

Part (f)

$\hat{\beta}_1$ = -0.09 for the new model, and $\hat{\beta}_1$ = -0.068 for the model with outliers. The outliers had a strong upward pull on the slope of the line.

Part (g)

The MSE is 10.86 as opposed to 13.99 previously. This is significant given deletion of only 2 of 32 rows.

Part (h)

Removing the outliers in this case helped the data conform to a linear model. I would remove the outliers to better fit the model within $(0, 250)$ horsepower, and avoid predicting outside of that interval.