CS 5821
February 2, 2017
Mariia Kravtsova
Assignment 2

1. Null Hypothesis: There is no relation between TV, radio, newspaper and sales.
   $\beta1 = \beta2 = \beta3 = 0$
   Given the p-values in Table 3.4 we can see that null hypothesis fails for TV and radio. However, we do not reject null hypothesis for newspaper.

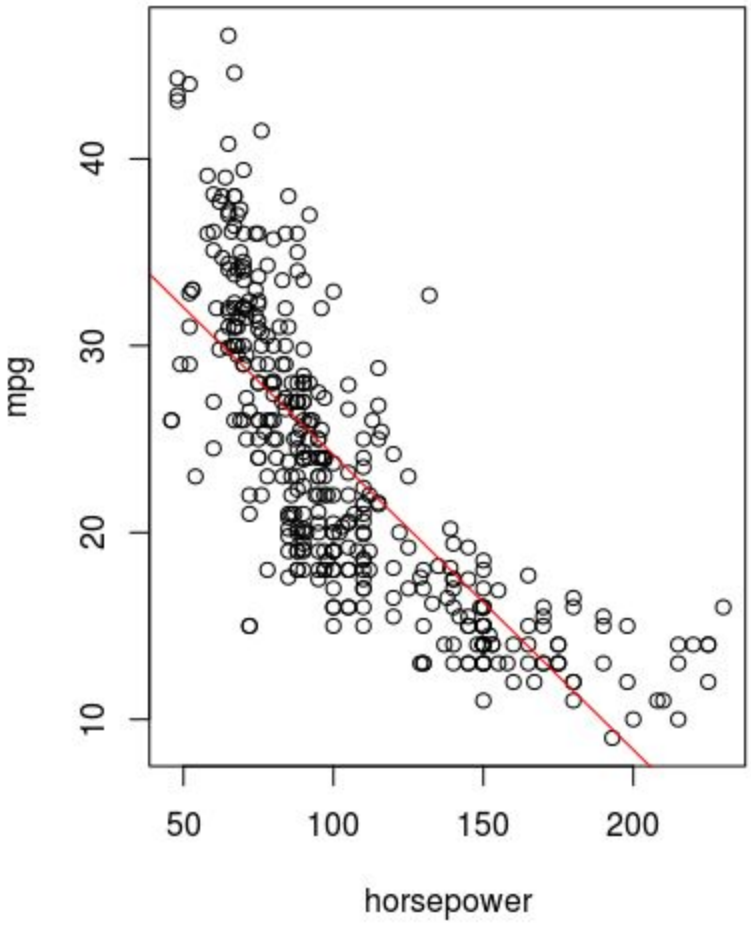
3. **a**. y(estimate) = 50+20*GPA+0.07*IQ+35*Gender+0.01*GPA*IQ-10*GPA*Gender
   y (estimate female) = 50+20*GPA+0.07*IQ+35*1+0.01*GPA*IQ-10*GPA*1 =
   = 85 + 10*GPA + 0.07IQ + 0.01*GPA*IQ
   y (estimate male) = 50 + 20*GPA + 0.07IQ + 0.01*GPA*IQ

   The baseline for male salary is higher than female because the GPA multiplier is higher. Hence, (iii) For a fixed value of IQ and GPA, males earn more on average than females provided that the GPA is high enough is correct.

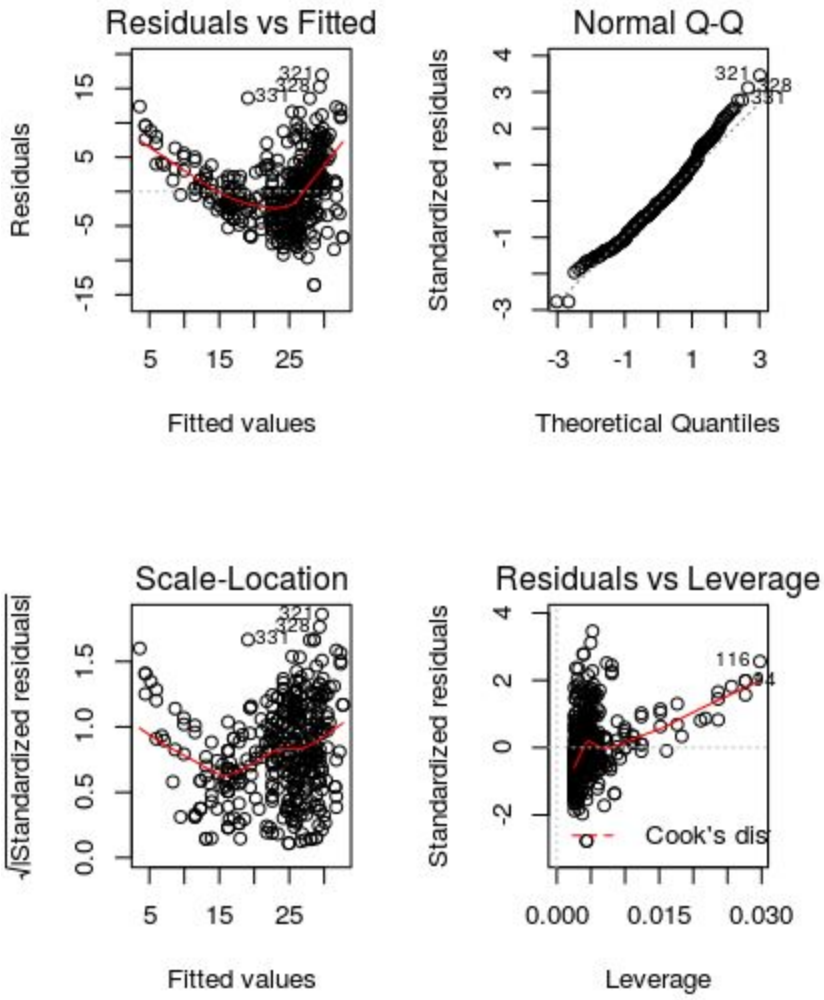
   b. y (estimate) = 85 + 10*4.0 + 0.07*110 + 0.01*4.0*110 = 137.1
   c. False. The coefficient for GPA/IQ does not provide us with any evidence to draw conclusions about the interaction effect.

4. **a**. In terms of flexibility the training RSS will decrease with a cubic fit.
   b. In this case the RSS will be higher because there will be less points to fit, so it will most likely overfit the data and will be less accurate.
   c. A flexible model will fit the training closer, and will reduce train RSS.
   d. We don't have enough information to tell if the test RSS will be lower than cubic or closer to linear.

8. **a.** Based on the p-value 2.2e-16 there is definitely a relationship between predictor and the response. The R^2 is 60.49% of the variability in mpg that can be explained using horsepower. The coefficient of horsepower is negative, that points that the relationship is negative - the more horsepower in the car, the less fuel efficient it will be. The predicted mpg associated with a horsepower of 98 is in the interval 14.8 - 34.12, fit of 24.26. The 95% confidence interval is between 24.44 - 25.44 and prediction interval between 15.28 - 34.59.



b.



c.

There are high leverage points at 116 as we see in the plot of Residuals vs Leverage, which provides non linearity in the data. We also see those high points being a problem in other plots.