1. Download the data file from the link provided for question 1.

Check that the names of the variables in the data file are:

"Age" "Photos" "Location"

Each row in the data corresponds to a unique user account on a photo sharing social networking site. The data has variables denoting the age of the user, the number of photos and the primary location (we have 4 states in the data: AZ, CA, NV, OR) of the user.

1. What is the sample size of this dataset?
2. Report the correlation between Age and Photos. Round your answer to 3 decimal places. What kind of linear association does it suggest (write only few words)?
3. Make a barplot of Location. Based on the plot, what is the most represented Location?
4. Make a histogram of Age. Use 10 breaks. What is the shape of the histogram?
5. Report the median Age
6. Make side-by-side boxplot of Photos for different Location. What do you see?
7. Copy and paste your R code for this data set here
8. Download the data file from the link provided for question 2.

Check that the names of the variables in the data file are:

"Sales" "Price" "Ratings"

1. What is sample size of this dataset?
2. Report the average Ratings? Round your answer to 3 decimal places.
3. Fit a linear regression model to predict Sales by Price and Ratings. Write down the equation of the predicting line as (fill in the blanks with the coefficients and their correct signs)

1. Is the model overall significant at 5% level or should you collect more explanatory variables? What is the value of the corresponding test statistic?
2. What percentage of variation in Sales is explained by the model?
3. What is the estimate for noise standard deviation in the model? Round your answer to 3 decimal places.
4. Based on the model, what is the predicted sales of a product which is priced at 80 dollars and has ratings 4.5?
5. What business implications can be made from the signs of the slope coefficients of Price and Ratings? Write at most two sentences.
6. Copy and paste your R code for this data set here