COMM 581 - Assignment #1
Data visualization and basic coding in R

Name: _____ **Total: 10 marks**
Due date: Sept. 12, 2015 (11pm)

Background:

You are choosing a house to buy in a new town. You gather some data about houses that are up for sale in the neighborhood that you are interested in. For this assignment, you will import this dataset into RStudio and examine the data to see what your house choices are.

The dataset for this assignment is called "Real_Estate_Sales_Data.csv" and is accompanied by a script file to open in RStudio called "2016-09-06_Assignment_01_script.R". The script file includes code to run, as well as comments to help you use R to find answers to the questions.

Answers to questions can be written in numeric form if the answer is simply a number (2 decimal places). If the question requires more interpretation, please answer using complete sentences.

Instructions and questions (questions to be answered are in bold)

1. Save the script file and the data file in the same folder somewhere on your computer – remember where this is!
2. Open the data file in Excel to check the variable names, if there are blanks that need to be replaced with "NA", commas in numbers to be removed, etc.
3. Open the script file in RStudio.
4. Import the dataset using the read.csv command
5. **a. How many rows of data are there? What does this represent? (1 mark)**

   b. How many columns of data are there? What does this represent? (1 mark)

6. Look at the structure of the data using the str command. **Do these variable types make sense for these data? What might you want to change? (1 mark)**

7. Change the variable bedrooms to a factor and view the structure of the data. **How many levels does the variable bedrooms have? (0.5 marks)**

8. Use commands in R to find the following values: **(1.5 marks)**

	Sale Price	**Square footage**
Mean		
Median		
Range (min, max)		
S.D.		
Variance		

9. Create histograms for price and square footage. Export them as .jpg files. Insert these into a Word document. **Do these distributions look normal? Why or why not? (1 mark)**

10. Create a scatterplot of the relationship between square footage and price. Copy and paste into a word document. **Does it look like there is a relationship between these two variables? Does the relationship look linear? Draw a line or curve that you think would fit the data on top of the graph (you can use shapes in Word). (1.5 marks)**

11. Create a boxplot of the relationship between sale price and number of bedrooms, including the filled circles for means. Insert into your Word document. **Which of these categories has the mean almost equal to the median? What does this indicate about the distribution of price for that number of bedrooms? (1 mark)**

12. Use the table command to find out how many options there are for the different number of bedrooms. You are looking for a 3-bedroom house, **how many options do you have? (0.5 marks)**

13. Use the subset command to create a dataset that has only the houses you are interested in (3 bedrooms). Create a barplot to show the number of choices you have in each bathroom category. Insert this barplot into your Word document. You are also looking for at least two bathrooms. **How many choices do you have? (0.5 marks)**

14. Compare sale price to your bathroom options. Consider the houses that meet your criteria: 3 bedrooms, and at least 2 bathrooms. **For which number of bathrooms is price the least variable? How many houses are in this category? (0.5 marks)**