Programming for Analytics
Assignment 2

Unless otherwise specified, you may use any function, package or method to complete the requirements of this assignment. I expect you to adhere to the best practices we have discussed in class including using intuitive, consistent variable names, using comments in your code, piping commands together instead of making multiple variables, etc.

The deliverables are as follows:

- A fully self-contained R script, loading data in from http://asayanalytics.com/pierce_housing-pool

Your script should include everything needed to complete the tasks below. This includes setting up libraries (but NOT installing packages), downloading the data files and it should include all code required to manipulate the data and create the output.

Pierce County Washington Housing Data

You are an analyst working for the Pierce County, Washington County Auditor. This office is responsible for collecting and reporting numerous county level data, including the historical sales of every property in the county. The data provided to you for analysis is structured such that each row is a property, and each column describes something about the property:

Column	Definition
Unique_id	A unique identifier for the property
yr_blt	The year the current structure on the property was built
lotsize	Size of the lot (in square feet)
sqft	Size of the structure (in square feet)
bath	Number of bathrooms where each value corresponds to some combination of the following: • 0.25 represents a sink only. • 0.50 represents a sink and toilet. • 0.75 represents a sink, toilet and shower. • 1.00 represents a sink, toilet, shower and bathtub. (This is also defined as a full bathroom)
bedrms	The number of bedrooms
trans_value	The transaction value of the most recent sale of the property (in dollars)
pool	A dummy variable indicating whether the home has a pool (1 is TRUE, 0 is FALSE)

You have been asked to complete the following tasks.

Instructions:

You are required to use data from the following URL in a .CSV format. This link contains data from all the necessary property values: http://asayanalytics.com/pierce_housing-pool

Section	Expectation	Possible Points
Data Preparation and Simple Questions	**1.1** Import the data set from the source specified above **1.2** Create a new numeric vector in the data denoting the age of the current structure relative to the ***most recent*** build date in the data. For example, if the most recent build date was 2004, a property with a `yr_blt` variable of 2001 would be 3 years old. Do ***not*** hard code any value. Your code should still be fully functional, even when the data is updated in the future. **1.3** Create a new logical vector in the data denoting whether the home has more bedrooms than bathrooms. **1.4** Create a new object in the environment that answers the following question: How many homes have more bathrooms than bedrooms?	0.5
Directed Analysis	**2.** Create the following visualizations. For some visualizations, you may find it helpful to create additional vectors in the original data or additional data frames in your global environment. For each result, briefly interpret the visualization to help the reader draw the appropriate conclusions. **2.1** The distribution of transaction values. **2.2** The average lot size of homes with a pool compared to homes without a pool. **2.3** The relationship between transaction value and square footage of the home. **2.4** Illustrate how transaction value varies by whether the home has a pool. **2.5** The relationship between the number of full bathrooms (no partial units) and the number of bedrooms.	1.0
Self-Directed Analysis	**3.** Create a visualization to provide evidence for or against each of the following claims. Include a brief summary of the visual and interpret it within the context of the stated claim to draw a conclusion. **3.1** **"**Older homes are worth less than newer homes." **3.2** "If the number of bedrooms increases, the square footage of the home must also increase." **3.3** "Bigger homes with more bathrooms are more likely to have pools."	0.75
Formatting & Miscellaneous Requirements	**4.1** Proper coding style is followed, and code is well commented where necessary. **4.2** Coding is systematic: complicated problems are broken down into steps that are individually much simpler, code is efficient, uses proper naming conventions, and is not unnecessarily verbose. **4.3** Visuals are carefully tuned for the desired purpose. Each visual illustrates one primary point, is appropriately formatted (plot and axis titles, legend (if necessary), scales are appropriate, appropriate geoms used, etc.) Visuals should not be misleading or more difficult than necessary to understand or interpret. **4.4** Any statistics used are not misleading, for example using a mean as a representative statistic on a very right skewed distribution. Implied causality (X-Y) are sensical (where appropriate.) **4.5** Insights obtained from the analysis are thoroughly, yet succinctly, explained. The results should be easy to see and understandable to a lay person viewing your webpage.	0.25