Descriptive statistics

Erasmus Q-Intelligence

Create a new R script (that is, a `.R` file) in RStudio which will contain the code to answer the questions below. For all questions, include the code that you use to answer the question in your R script file.

Remember to include comments (lines that start with a `#`) in this text file so that you can remember which code relates to which exercise. Save your work under an appropriate file name, and remember to resave regularly.

1 House Prices

Exercise 1.1 (Data loading and exploration)**.** Download the house price data in the file `forsale.Rds` from Canvas by saving it directly to disk. Store it in an appropriate folder.

(a) Load the `forsale.Rds` file into your R session and name it `forsale`. Include the code to do so in your R script.

(b) Open the resulting data frame object `forsale` using RStudio's viewer with the `View()` function (as always, include the code). Make sure you have reasonable answers to the following questions:

- What does a single observation (row) in this data set represent?
- Which variables do the data set contain?
- What units are these measured in?
- What types of variables are these?

(c) How many observations (rows) are there in this data set? How many variables (columns) are there? Include code to obtain these values.

(d) Use the `summary()` function on the data frame object to see what output this gives you. Note down some of the key points that you have learned about the data.

Exercise 1.2 (Univariate summary statistics)**.** Now that you are familiar with the basics of the data, let's answer some simple but important questions. Remember that you can extract a single column vector from a data frame using the `$` operator, as in `dataname$variablename`.

(a) What is the mean asking price of a house in this data set? Include code to calculate this using the `mean()` function.

(b) How does that compare to the median price? Can you say something about the expected skewness of the data based on the mean and median?.

(c) What is the asking price of the cheapest house? And of the most expensive house? Remember the functions `min()`, `max()` and `range()`. What results do these respective functions report?

(d) Calculate the range of the asking prices, as well as the standard deviation.

(e) What values do the variables `city`, `available_now`, `type_build` and `bedrooms` take? Make frequency tables of counts for each of these variables separately using the `table()` function. How many houses have six bedrooms? How many have five bedrooms?

(f) Use the `table()` function with two arguments (as in `table(x, y)`) to investigate the relation between `jacuzzi` and `sauna` in this data set. How many houses are there with both a jacuzzi and a sauna?