Python for Data Analysis and Visualization

Instructor: Claudia Carroll

Class 4: Data manipulation with pandas

In-Class Exercises

**Exercise One:**

Create a new dataframe called SAFI\_subset from the SAFI\_results.csv that contains the columns C01\_respondent\_roof\_type, C02\_respondent\_wall\_type, C02\_respondent\_wall\_type\_other, and C03\_respondent\_floor\_type. Calculate the percentage of cells in the dataframe that are null.

Hint: Arithmetic operators in python are:

* + (addition)
* - (subtraction)
* \* (multiplication)
* / (division)

**Exercise Two:**

1. Read in the SAFI\_results.csv dataset.
2. Get a list of the different C01\_respondent\_roof\_type values.
3. Groupby C01\_respondent\_roof\_type and describe the results.

**Exercise Three:**

Examine the contents of the SN7577i\_aa and SN7577i\_bb csv files using Excel or equivalent. Then, using the SN7577i\_aa and SN7577i\_bb csv files, create a dataframe which is the result of an outer join using the Id column to join on. What do you notice about the column names in the new dataframe?