Day 1: Data Munging

Gabriella Gerlach

December 9, 2020

1 Data Transformations

Associated screen cast: link

For each question, add the requisite Python code to implement it. Assume that previous code in the same assignment are still available in the Python environment

1. Load the mtcars dataset into Python
2. Create a new column that stores fuel efficiency in km/l. You will take the mpg column and perform the transformation km/l = mpg * 1.6 / 3.8 and store it in this new column

2 Data Concatenation

Associated screen cast: link

1. Load the gapminder data set into Python
2. Extract the first 5 rows of the data set
3. Extract the country, Per Capita GDP, life expectancy columns into a new DataFrame
4. Extract the rows of the dataset that correspond to Canada

Hint: note the file type of the gapminder file (tsv = tab separated variables; read in as: `pd.read_csv("file", sep="t")`)

3 Data Merging

Associated screen cast: link

1. Load the 4 survey data files into Python
2. Perform an innter join of the survey data (`survey_survey.csv`) with the visited data
3. Perform an outer join of the survey data (`survey_survey.csv`) with the person data

4 Tidy Data

Associated screen cast: link

1. Load the weather data into Python
2. Describe how you would transform the data to make it tidy

5 Reshaping Data

Associated screen cast: link

Provide code to transform the weather data into a tidy format

6 Split-apply-combine

Associated screen cast: link

1. Load the gapminder data into Python
2. Provide the code to fine the median per Capita GDP by continent and year