Jupyter Notebook & Pandas

Command and Edit Mode

- **Command mode** - Click inside cell, Hit Escape Key, Blue line on left
- Shortcuts
  a - add cell above
  b – add cell below
  m – change to markdown cell
- Edit mode – double click inside cell, Green line on left
  Shift + Enter – runs code

Pandas

- import package and give it an alias

import pandas as pd

We will be working heavily with panda’s data types: series and dataframes

Creating dataframes from scratch

- Provide data and column names

```
df = pd.DataFrame(data = [[1,2,3],
                          [4,5,6],
                          [7,8,9]],
                  columns = ['column1', 'column2', 'column3'])
```

Importing csv and excels files into a dataframe

- CSV’s

df_csv = pd.read_csv(“filepath/filename.csv”)

- Excel files

df_excel = pd.read_excel(“filepath/filename.xlsx”)

Subsetting dataframes

- Isolating one column

df.columnName

df[‘columnName’]

- Accessing multiple columns – two square brackets
- df[[‘column1’, ‘column2’, ‘column3’]]

Subsetting continued

- Indexing rows – start : stop (stops at one row BEFORE specified stopping point, i.e. if you want to stop at row 9, put 10)

df[4:10] #returns rows with index 4 – 9

df[4:] #putting no stopping point returns all rows after starting point

df[:10] #returns rows with index 0 through 9

df[:-2] # returns all rows except last 2

Useful Functions

- Summary statistics of a dataframe

```
df.describe() # for continuous variables
df.describe(include=['O']) # catergorical variables

df.column.mean()
df.column.std()
df.column.median()
df.column.value_counts() #count for each category group
```