Jupyter Notebook & Pandas

Why Jupyter Notebook?

- Easier to use
- In-line processing – see results pretty and immediately
- A standard of DS
- Comes with DS packages already installed

Command and Edit Mode

Command Mode

- Click inside cell and hit Escape
  - a - add cell above
  - b – add cell belo w
  - m – change to markdown

Edit Mode

- Double click inside cell
- Shift + Enter runs code

pandas

- Your primary data science package
  - Read in and view data
  - Stores data in a data frame
  - Data wrangling
  - Summary statistics

- You must import it to use it!

```
import pandas as pd
```

Creating dataframes from Scratch

- You 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 Excel Files

- CSVs

```
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

df[['column1', 'column2', 'column3']]

Multiple Summary Statistics

- Continuous

```
df.describe()
```

- Categorical

```
df.describe(include=['O'])
```

IndividualSummary Statistics

- Mean

df.column.mean()

- Standard Deviation

df.column.std()

- Median

df.column.median()

- Frequencies

df.column.value_counts()