9.15.23 & 9.19.23

Topic: Pandas Fundamentals

Dataframes

- A data frame is a pandas object that is used to store a dataset
- Information is organized in rows and columns
- Dataframes can be created from a dictionary of lists; keys become the column headers and list is the values of that column.

```
In [2]: df = pd.DataFrame(
   ...:     {
   ...:         "Name": [
   ...:             "Braund, Mr. Owen Harris",
   ...:             "Allen, Mr. William Henry",
   ...:             "Bonnell, Miss. Elizabeth",
   ...:         ],
   ...:         "Age": [22, 35, 58],
   ...:         "Sex": ["male", "male", "female"],
   ...:     }
   ...: )
   ...:
```

Series

- A series is a pandas object used to create dataframes
- Seen as a one-dimensional list of data; think of it as a single column in a data frame.

Indexing into Dataframes

- df.loc[row_index, col_label]
- df.iloc[row_index, col_index]

Selection

- Process of accessing a subset of a dataframe. You can select subsets using loc and iloc

df.loc[initial row index : final row index, [column labels separated by commas]]

- loc for row indices are both inclusive. For iloc, the final row and column index is not included but initial indices are included.

df.iloc[initial row index : final row index, initial column index : final column index]

- Just putting ":" for the index range means all of them are included; ignoring one side of the colon means it goes all the way to the end.

Filtering

- Selecting values of a dataset where certain conditions are true → **df[*condition*]**
    - It creates a new dataframe only displaying the row where the condition is true.
    - and in pandas when inside the square brackets for conditions → &
    - Or in pandas when inside the square brackets for conditions → |

Combining Dataframes

- Concatenate: naively combines along an axis
- Merge: combine through shared column
- Join: combine using shared indices
    - Inner join
    - Left join
    - Right join
    - Outer join