Introduction to Pandas: Takeaways ↱

Syntax

- Loading a CSV file into pandas:

```
import pandas as pd
df = pd.read_csv('filename.csv', sep=',', encoding='UTF8')
```

- Setting a column as index:

```
df = pd.read_csv('filename.csv', index_col=0)

df = df.set_index('ColName')

df.set_index('ColName', inplace=True)
```

- The first few rows:

```
df.head(num_rows)
```

- The last few rows:

```
df.tail(num_rows)
```

- Accessing columns:

```
df[ColLabel]
df.loc[:, ColLabel]

df.iloc[:, col_index]
```

- Accessing rows:

```
df.loc[RowLabel]
df.loc[RowLabel, :]

df.iloc[row_index]
df.iloc[row_index, :]
```

- Accessing a specific cell:

```
df.at[RowLabel, ColLabel]
df.iat[row_index, col_index]
```

Concepts

- Pandas dataframes are similar to 2-dimensional ndarrays but allow us to store multiply types of values.
- Pandas dataframes have two properties to access the values. The `DataFrame.loc` and `DataFrame.iloc` properties. The first works with row and column labels. The second works with indexes, in the same way as ndarrays did.
- When reading a CSV file with headers, be default pandas will use these as column labels. The default row labels are the same as the row indexes.

Resources

- Pandas documentation