

Thursday, November 5, 2020 | **Class #10**

# pandas and xarray

**Working with CSV and netCDF data files**

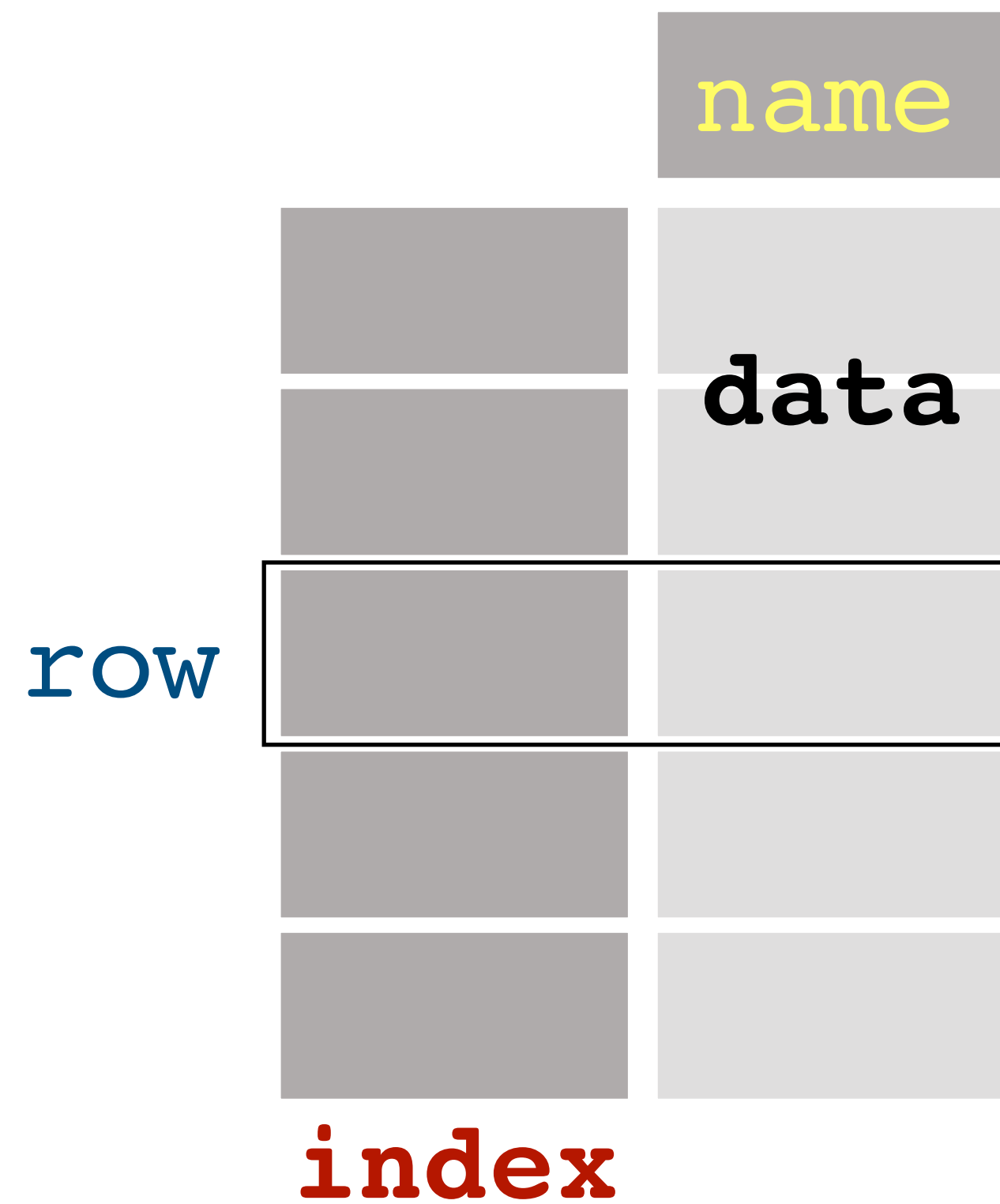
OCEAN 215 | Autumn 2020

**Ethan Campbell** and Katy Christensen

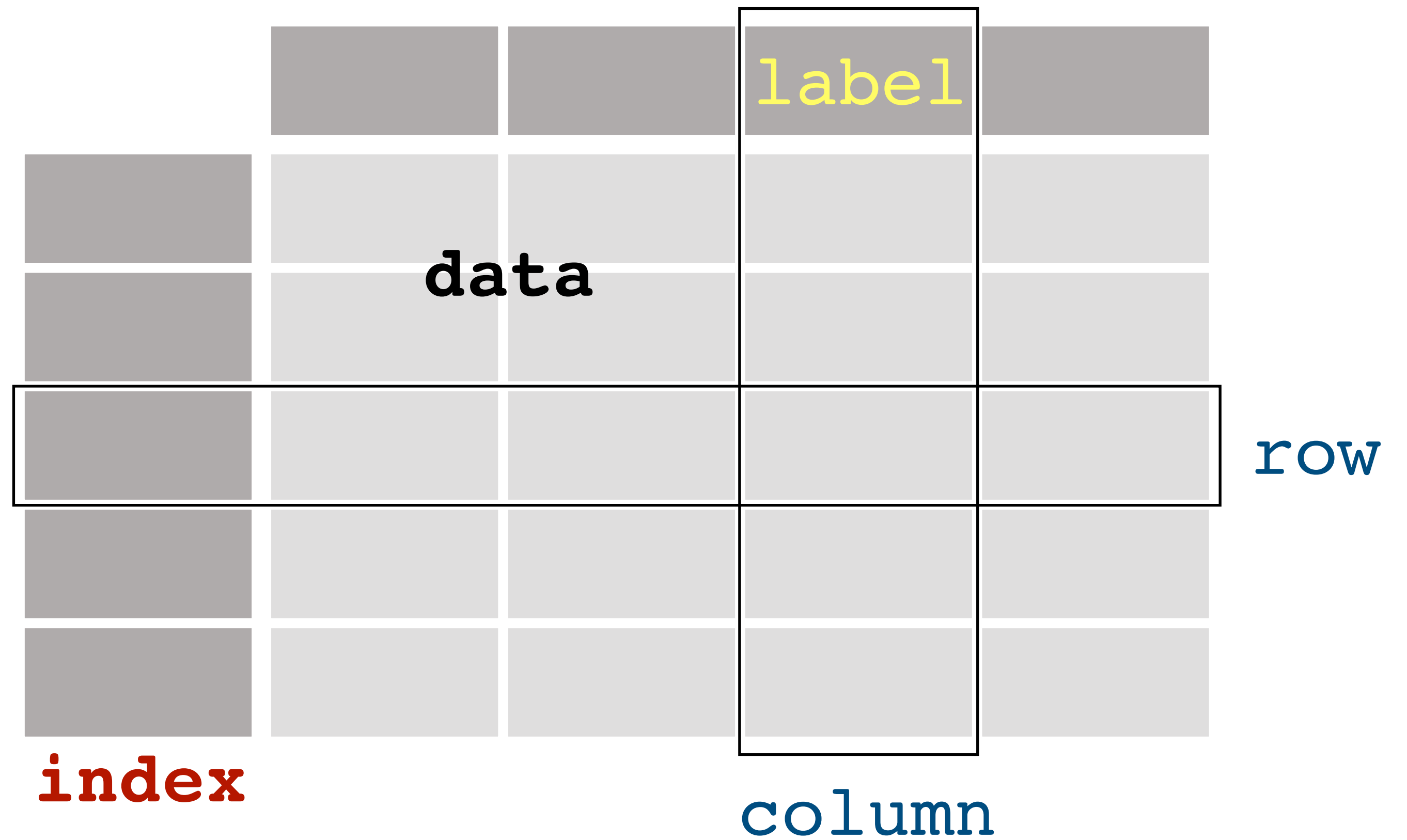
# pandas handles tabular data (tables or spreadsheets)

---

## Series

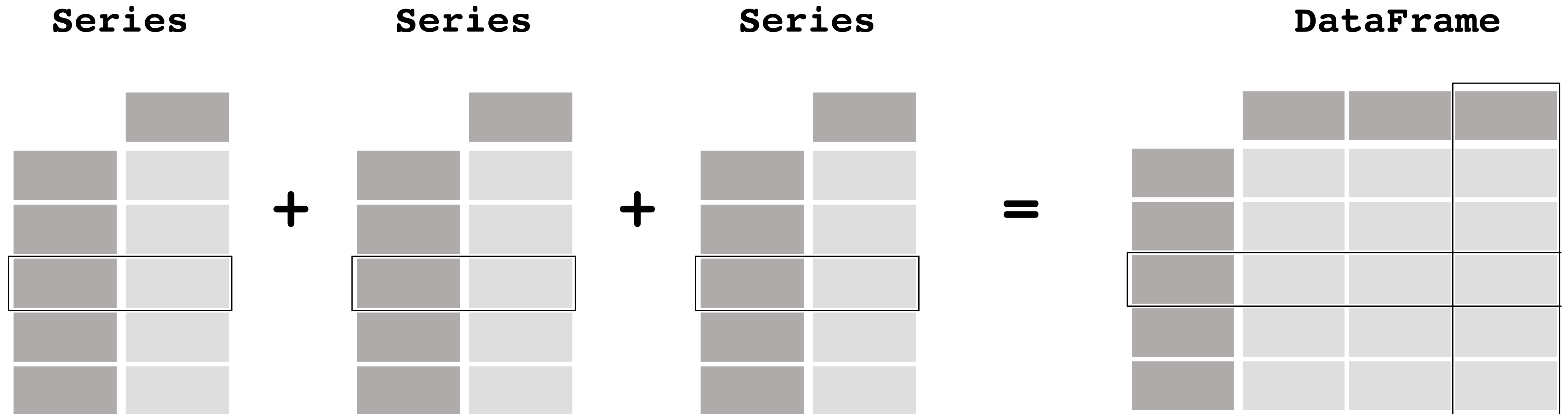


## DataFrame



# Two or more `Series` can be concatenated to become a `DataFrame`

---



```
pd.concat([s1, s2, s3, ...], axis=1)
```

# Getting information about a DataFrame

---

**.shape**

```
1 df.shape
```

```
(5, 2)
```

**.size**

```
1 df.size
```

```
10
```

**print()**

```
1 print(df)
```

```
      Temperature  Salinity
2016           4.1     35.5
2017           5.2     35.0
2018           6.3     34.5
2019           7.4     34.0
2020           8.5     33.5
```

**display()**

```
1 display(df)
```

```
      Temperature  Salinity
2016           4.1     35.5
2017           5.2     35.0
2018           6.3     34.5
2019           7.4     34.0
2020           8.5     33.5
```

**.describe()**

```
1 df.describe()
```

```
      Temperature  Salinity
count          5.000000  5.000000
mean           6.300000  34.500000
std            1.739253  0.790569
min            4.100000  33.500000
25%            5.200000  34.000000
50%            6.300000  34.500000
75%            7.400000  35.000000
max            8.500000  35.500000
```

# Selecting data from a DataFrame using `.iloc [ ]` and `.loc [ ]`

---

## Selection by index:

```
<DataFrame> [ <column label(s)> ] . iloc [ <index or indices> ]
```

## Selection by label:

```
<DataFrame> [ <column label(s)> ] . loc [ <label or labels> ]
```

**Example:** `df [ 'Salinity' ] . loc [ 2019 ]`

# Reminder: convert the resulting `Series` to a NumPy array

---

```
s1.loc[2018:2020]
```

gives a `Series` object

```
s1.loc[2018:2020].values
```

gives a NumPy array

# Putting it all together

---

*Combine column extraction, selection by label, and applying a NumPy function*

Start with a DataFrame



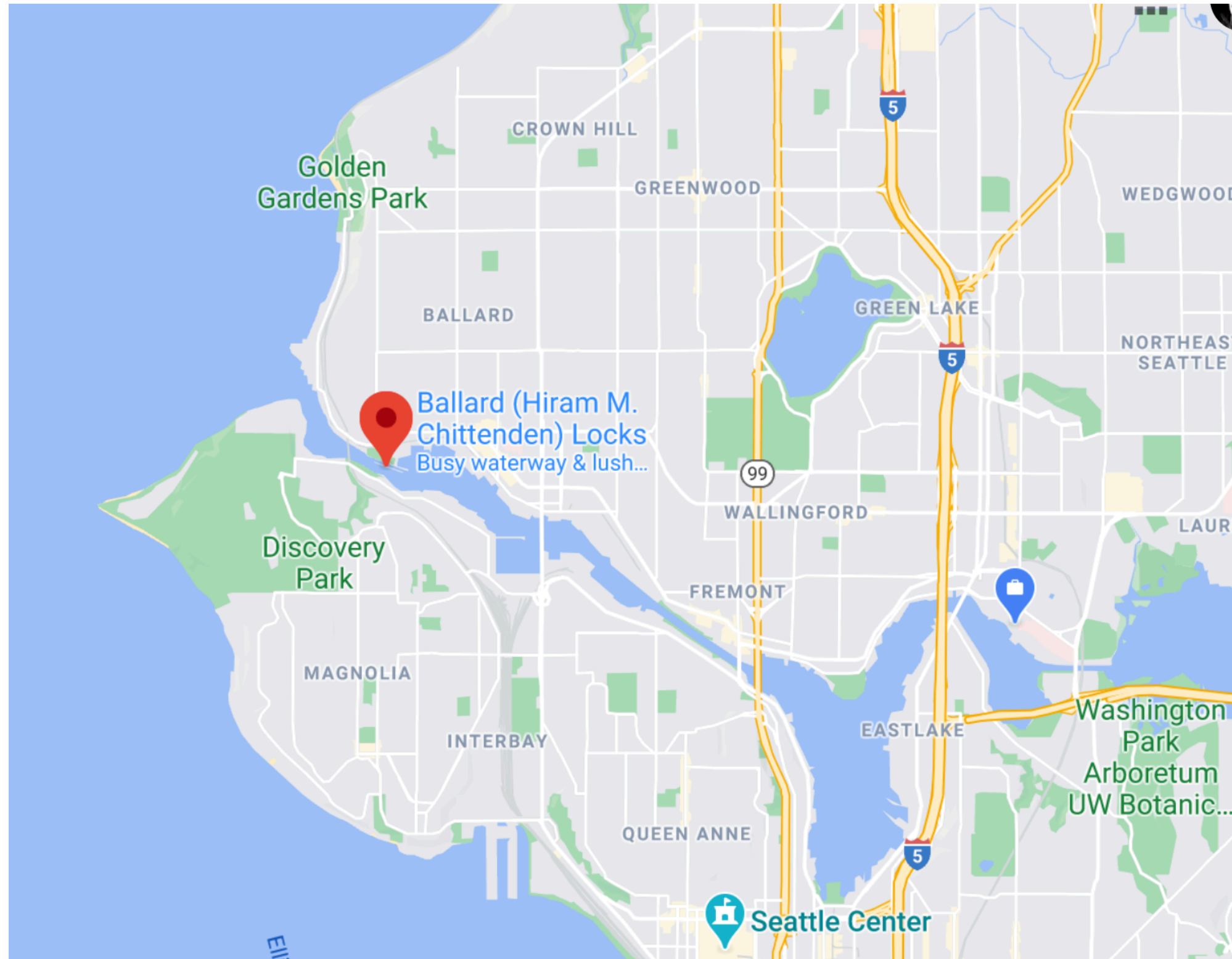
```
df['Salinity'].loc[2017:].mean()
```

This gives a Series

This gives a slice from that Series

This gives a single value: the average salinity from 2017 onwards

# pandas activity: Ballard Locks salmon counts



Google Doc with activities (also accessible from Canvas Modules or Google Drive folder):

**<https://tinyurl.com/OCEAN215-Class10>**