Section 3: pandas and Asn2

Sam Lau
OH: Th 3:30-4:40pm CSB 115

1. All you need to know about numpy
2. Most of what you need to know about pandas
3. Collaborate on Asn2

Numpy

Numpy array = A Python list with better math.

```python
foo = np.array([20, 3, 0, -6])
foo
```

```
array([20,  3,  0, -6])
```

```python
foo + 5
```

```
array([25,  8,  5, -1])
```

(demo)

Pandas

Every column has a label

	plate	time	location
0	1275226	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
1	27529C	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
2	1158423	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
...	...	...	...
2742098	5X10319	12/19/2013 08:28:00 PM	(37.804148, -122.284861)
2742099	7D56240	12/19/2013 08:28:00 PM	(37.804096, -122.284635)
2742100	6JNM127	12/19/2013 08:28:00 PM	(37.80397, -122.284046)

2742101 rows × 3 columns

Every row also has a label

You can work with lots of data at once

Pandas

Some methods only work on columns (axis=1)

Some only work on rows (axis=0)

	plate	time	location
0	1275226	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
1	27529C	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
2	1158423	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
...	...	...	...
2742098	5X10319	12/19/2013 08:28:00 PM	(37.804148, -122.284861)
2742099	7D56240	12/19/2013 08:28:00 PM	(37.804096, -122.284635)
2742100	6JNM127	12/19/2013 08:28:00 PM	(37.80397, -122.284046)

2742101 rows × 3 columns

Some work on both (axis=0 or axis=1)

Pandas

pandas.DataFrame.drop

DataFrame.**drop**(*labels=None, axis=0, index=None, columns=None, level=None, inplace=False, errors='raise'*)

[source]

Drop specified labels from rows or columns.

Remove rows or columns by specifying label names and corresponding axis, or by specifying directly index or column names. When using a multi-index, labels on different levels can be removed by specifying the level.

Parameters:

labels : *single label or list-like*
　　Index or column labels to drop.

axis : *{0 or 'index', 1 or 'columns'}, default 0*
　　Whether to drop labels from the index (0 or 'index') or columns (1 or 'columns').

index, columns : *single label or list-like*
　　Alternative to specifying axis (`labels, axis=1` is equivalent to `columns=labels`).
　　New in version 0.21.0.

level : *int or level name, optional*
　　For MultiIndex, level from which the labels will be removed.

inplace : *bool, default False*
　　If True, do operation inplace and return None.

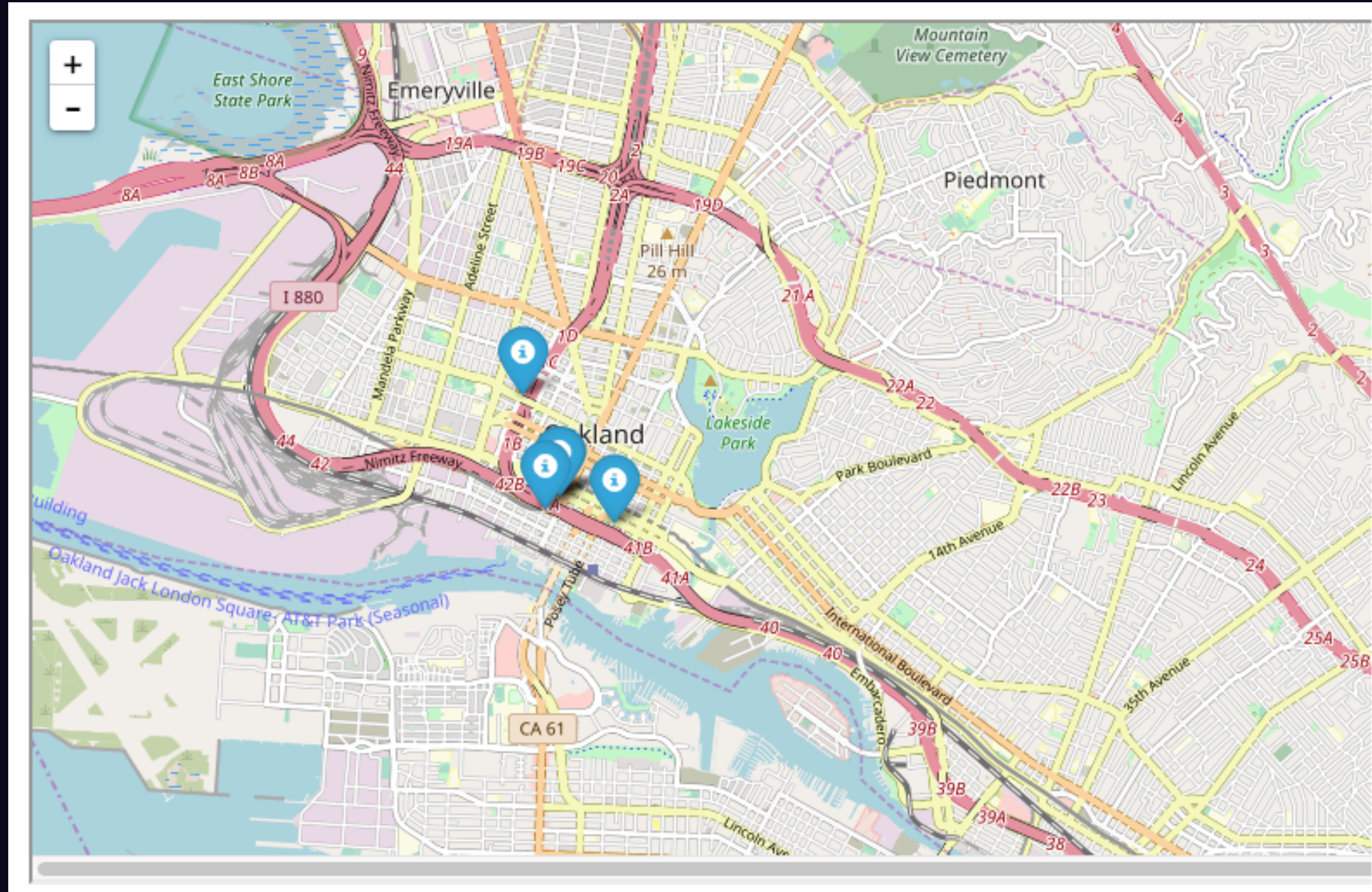# Pandas

You can almost do the entire assignment with `df.loc`

`df.loc[2, 'plate'] => '1158423'`

	plate	time	location
0	1275226	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
1	27529C	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
2	1158423	01/19/2011 02:06:00 AM	(37.798304999999999, -122.27574799999999)
...	...	...	...
2742098	5X10319	12/19/2013 08:28:00 PM	(37.804148, -122.284861)
2742099	7D56240	12/19/2013 08:28:00 PM	(37.804096, -122.284635)
2742100	6JNM127	12/19/2013 08:28:00 PM	(37.80397, -122.284046)

2742101 rows × 3 columns

(demo)

Next Time...

Go work on Asn2!