

Jupyter Data Science Workflow

From exploratory analysis to reproducible science

Jake VanderPlas University of Washington eScience Institute

```
[22] URL = 'https://data.seattle.gov/api/views/65db-xm6k/rows.csv?
      accessType=DOWNLOAD'
```

```
[23] from urllib.request import urlretrieve
      urlretrieve(URL, 'Fremont.csv')
```

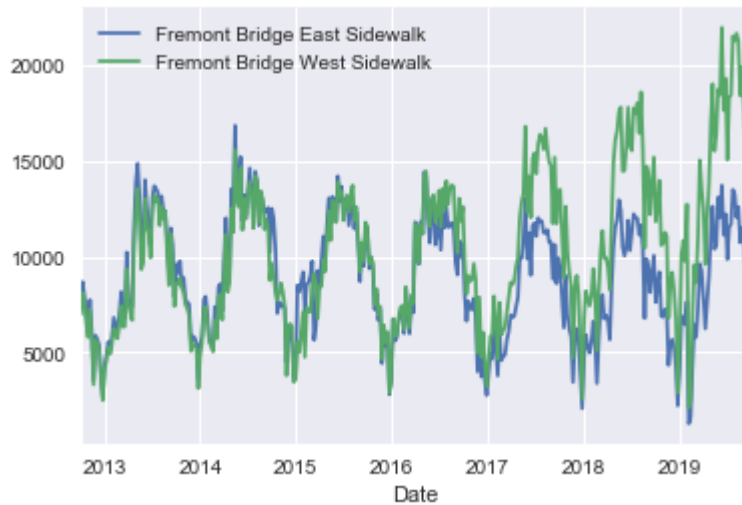
```
('Fremont.csv', <http.client.HTTPMessage at 0x11ac605c0>)
```

```
[24] import pandas as pd
      data = pd.read_csv('Fremont.csv', index_col='Date',
      parse_dates=True)
      data.head()
```

	Fremont Bridge East Sidewalk	Fremont Bridge West Sidewalk
Date		
2015-02-24 02:00:00	3.0	3.0
2019-01-01 00:00:00	0.0	9.0
2019-01-01 01:00:00	2.0	22.0
2016-02-15 00:00:00	3.0	3.0
2019-01-01 02:00:00	1.0	11.0

```
[25] %matplotlib inline
```

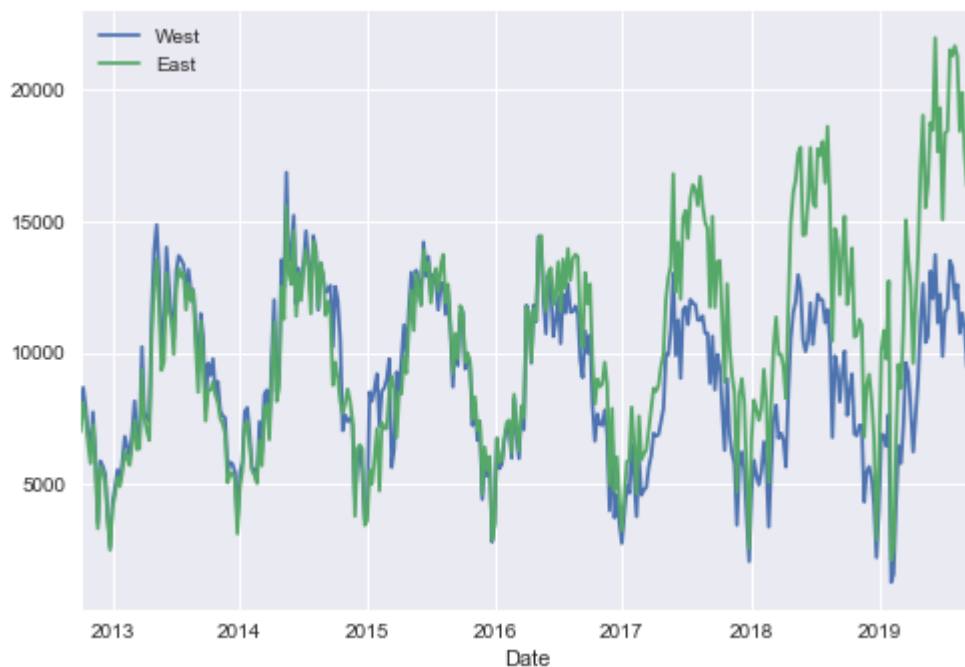
```
data.resample('3W').sum().plot();
```



```
[26] import matplotlib.pyplot as plt
plt.style.use('seaborn')

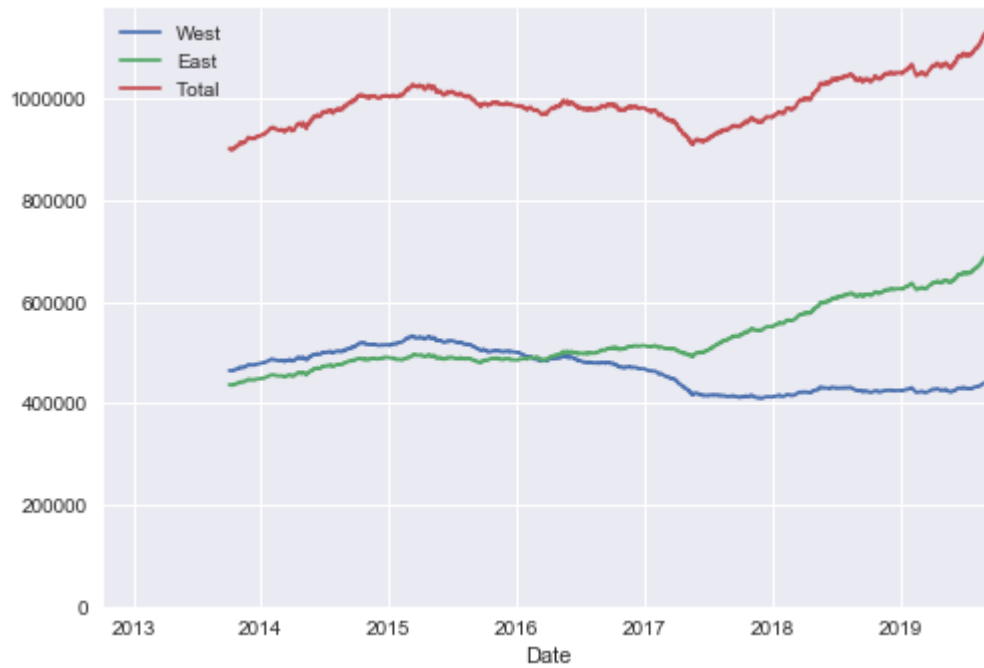
data.columns = ['West', 'East']

data.resample('W').sum().plot();
```

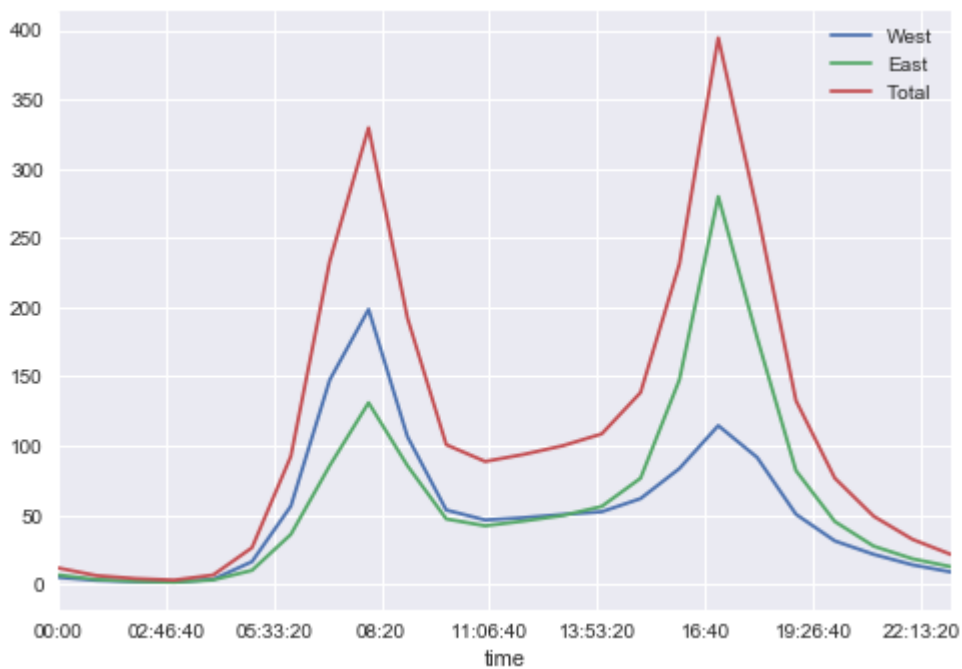


```
[6] data['Total'] = data['West'] + data['East']

ax = data.resample('D').sum().rolling(365).sum().plot();
ax.set_ylim(0, None);
```



```
[7] data.groupby(data.index.time).mean().plot();
```

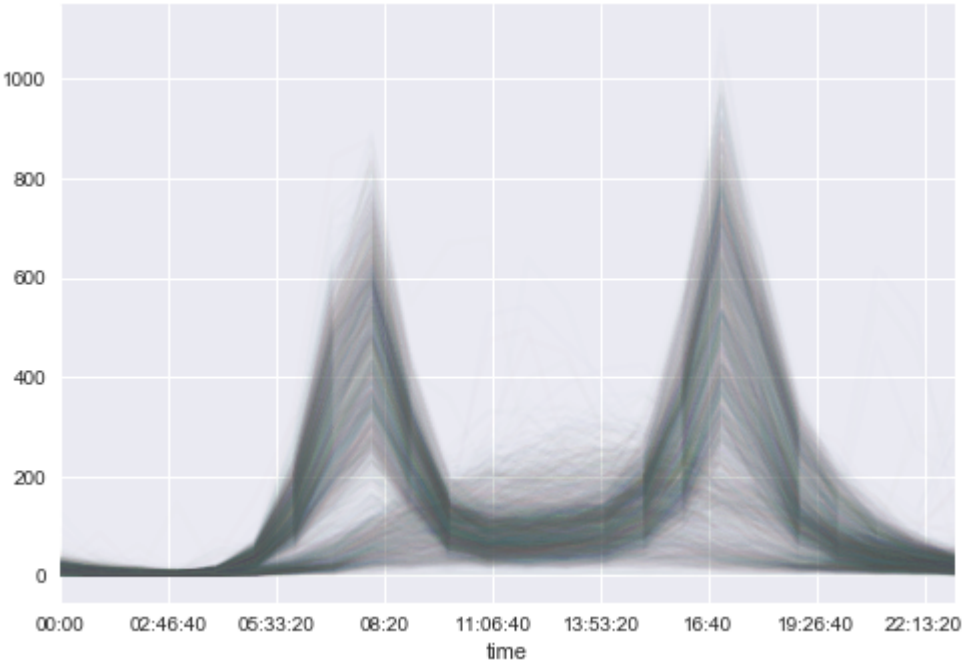


```
[8] pivoted = data.pivot_table('Total', index=data.index.time,
columns=data.index.date)
pivoted.iloc[:5, :5]
```

	2012-10-03	2012-10-04	2012-10-05	2012-10-06	2012-10-07
	2012-10-03	2012-10-04	2012-10-05	2012-10-06	2012-10-07
00:00:00	13.0	18.0	11.0	15.0	11.0

00:00:00	10.0	10.0	11.0	10.0	11.0
01:00:00	10.0	3.0	8.0	15.0	17.0
02:00:00	2.0	9.0	7.0	9.0	3.0
03:00:00	5.0	3.0	4.0	3.0	6.0
04:00:00	7.0	8.0	9.0	5.0	3.0

```
[9] pivoted.plot(legend=False, alpha=0.01);
```



```
[ ]
```