

Lab: Spatial & Temporal Data

04/18/2019

Data set for this lab

- Data sets used for this lab are available in the git repo
- NYC taxi data, originally come from: <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>
 - 2016 June
 - Check TaxiPrepare for sampling process
- US Unemployment data
- US state map data
- <https://github.com/python-visualization/folium/tree/master/examples/data>

Explore Taxi Data

- Spatiotemporal data, including location (longitude, latitude) and time (date and time)
- Data from 2 vendors
- 10000 records (after sampling)

Do you need a map? Why?

Try to answer the following questions

Questions?

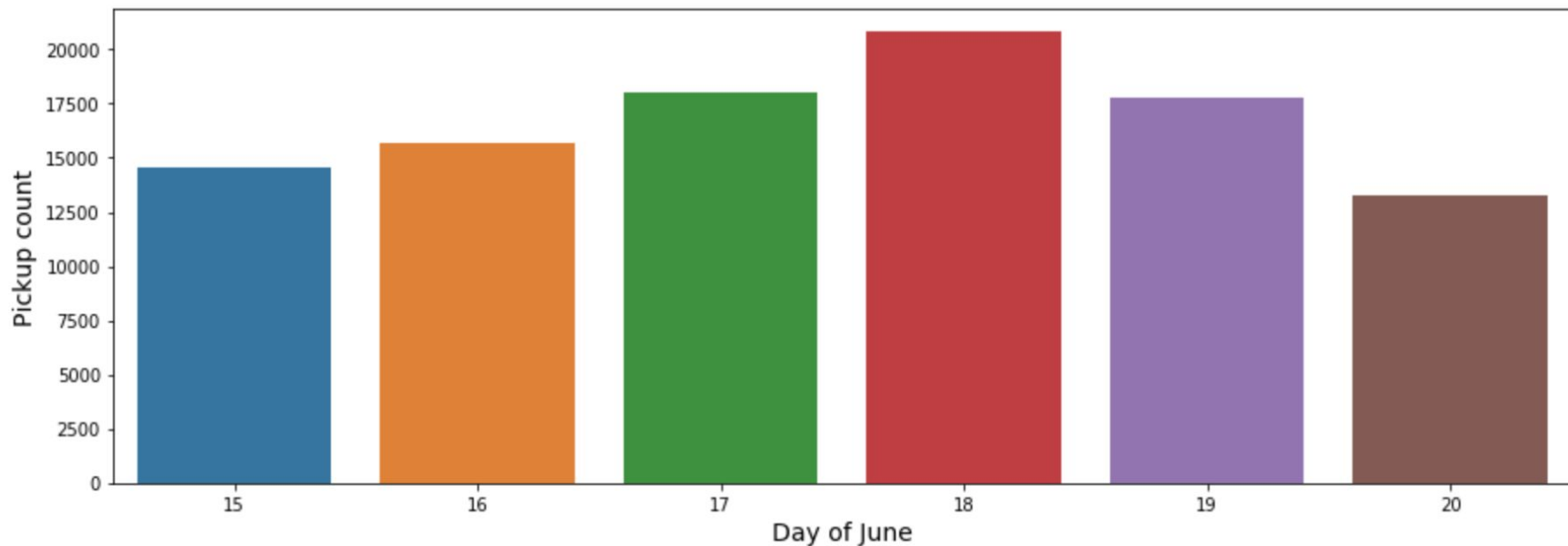
- How does the total pick-up change over time (date, weekday, hour)?
- How does the trip duration change over time?
- How does the trip duration from different vendors change over time ?
- What is the relationship between trip duration and number of passengers?
- How does the total pick-up distribute in NYC?
- How does the pick-up distribution change over time?

How does the total pick-up change over time
(date, weekday, hour)?

- Let's Sketch

How does the total pick-up change over date?

```
f = plt.figure(figsize=(15,5))  
  
sns.countplot(x='day', data=df)  
  
plt.xlabel('Day of June', fontsize=14)  
  
plt.ylabel('Pickup count', fontsize=14)  
  
plt.show()
```

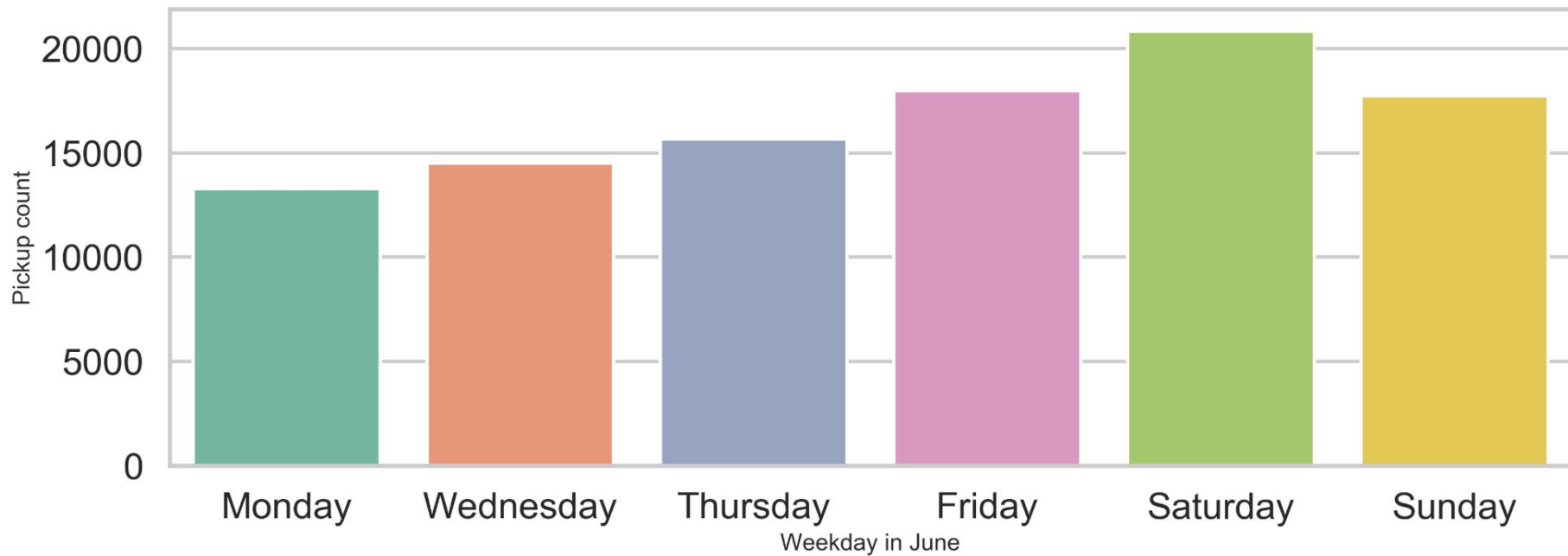


* Here we just use the function “countplot” here which generates the bar chart above. We can also generate line chart after pre-processing.

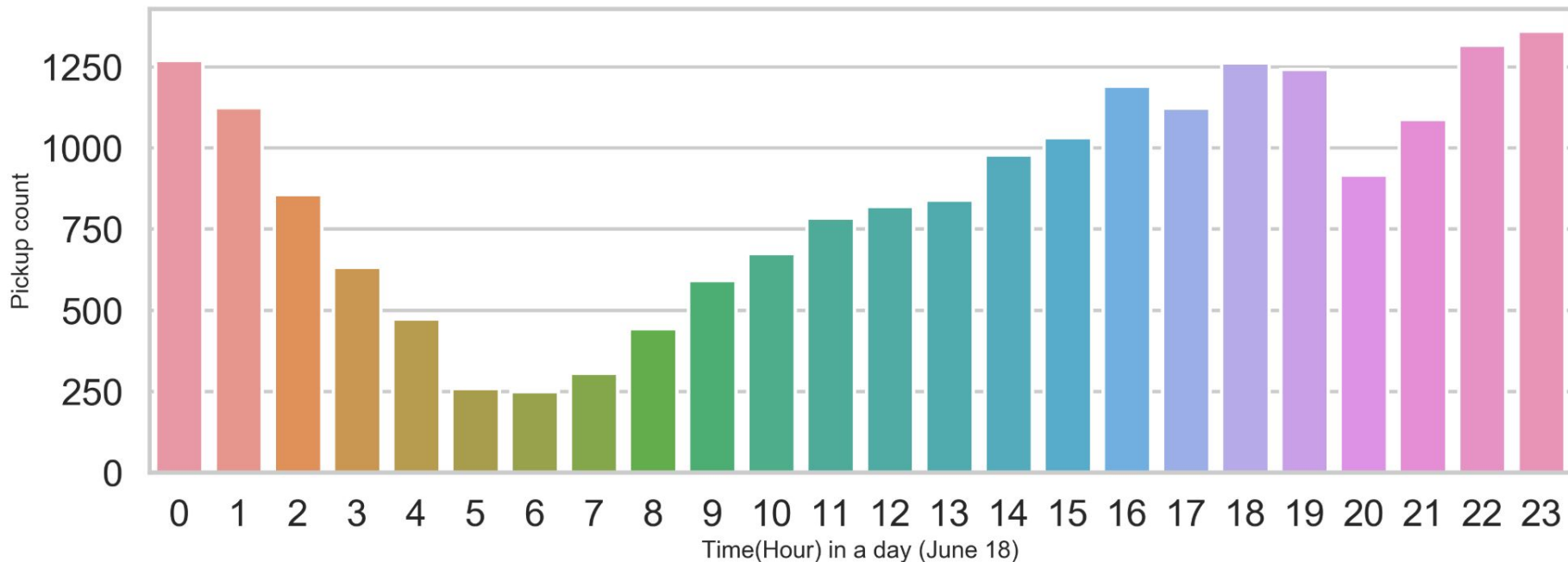
How does the total pick-up change over weekday?

```
values_week = ['Monday', 'Wednesday', 'Thursday', 'Friday', 'Saturday',  
               'Sunday']
```

```
f, axes = plt.subplots(nrows=1, ncols=1, figsize=(15,5))  
  
sns.countplot(x='weekday', data=df, palette='Set2')  
  
plt.xlabel('Weekday in June', fontsize=14)  
  
plt.ylabel('Pickup count', fontsize=14)  
  
axes.set_xticklabels(values_week)  
  
plt.show()
```



How does the total pick-up change over time (hour)?



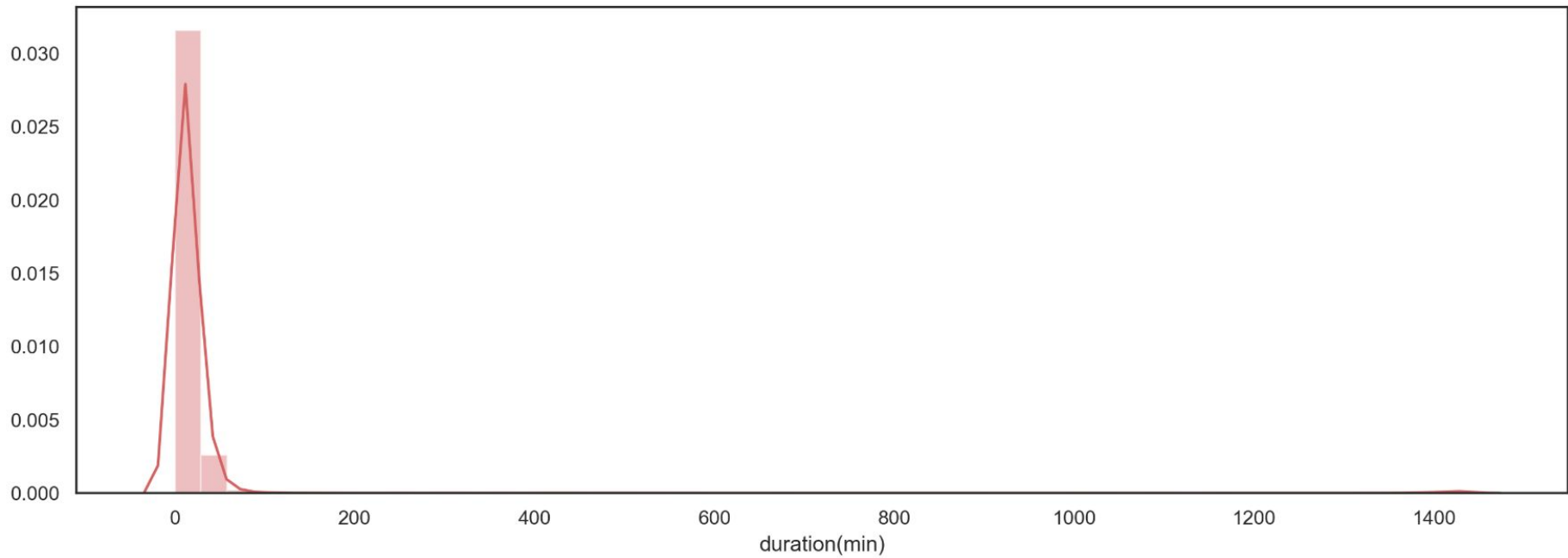
How does the trip duration change over time?

Let's Sketch!

Plot Duration

- Generate Duration first
 - $\text{Duration} = \text{Drop_off_time} - \text{Pick_up_time}$
- Code provided in the notebook

- Still use “distplot” function

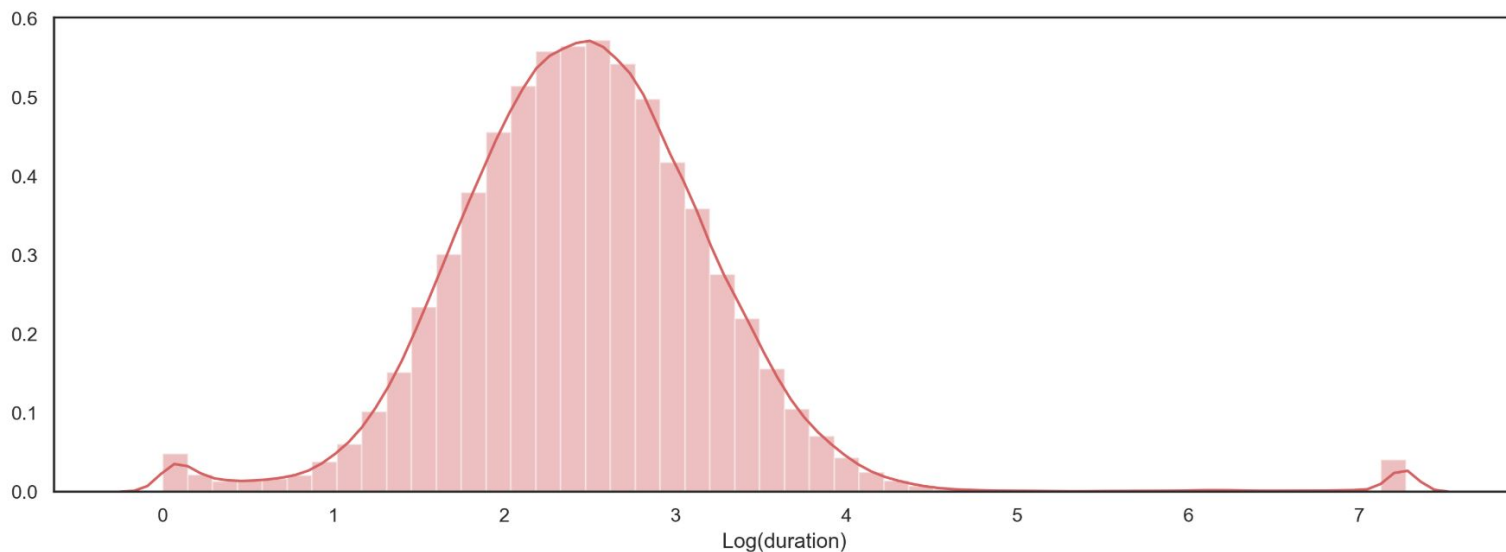


Cannot really gain much insights here...

Use log to differentiate the durations

Render distribution plot of the following array:

```
np.log(df['duration'].values+1)
```



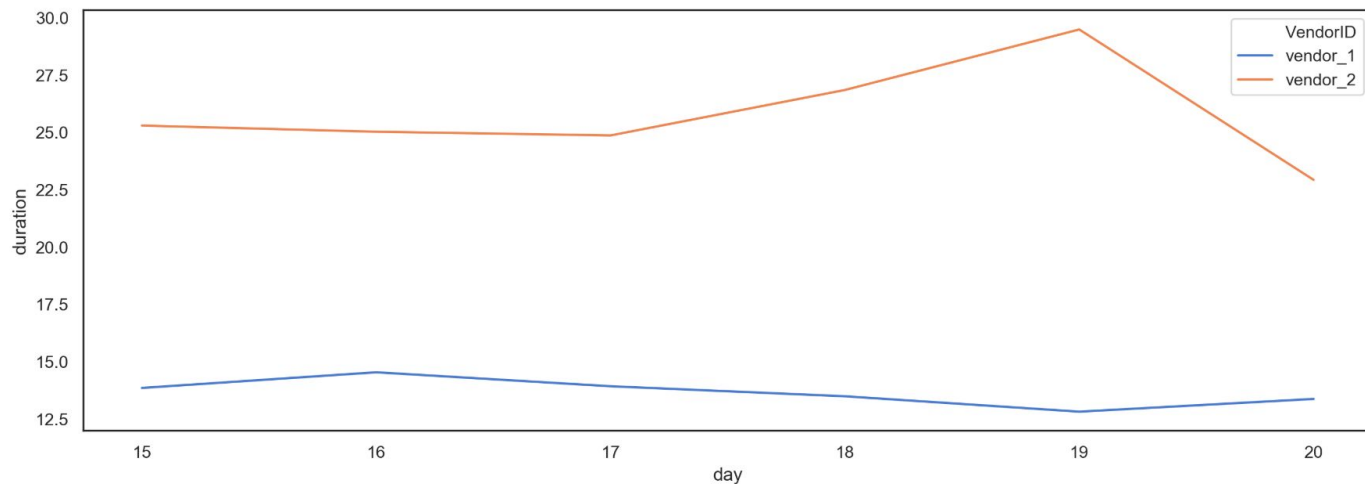
Visualize the mean trip duration changes
over time from different vendors

Let's sketch

One possible solution:

```
f = plt.figure(figsize=(15,5))
```

```
sns.lineplot(data=df_v1and2, x='day', y='duration',  
hue='VendorID')
```



Explore the relationship between attributes

E.g.: duration and count_passenger

Let's sketch!

A possible solution here

```
sns.violinplot(x="Passenger_count", y="duration", hue="VendorID", data=df2)
```

*Try to set the parameter “split” as True

How does the total pick-up distribute in NYC?

Sketch them:

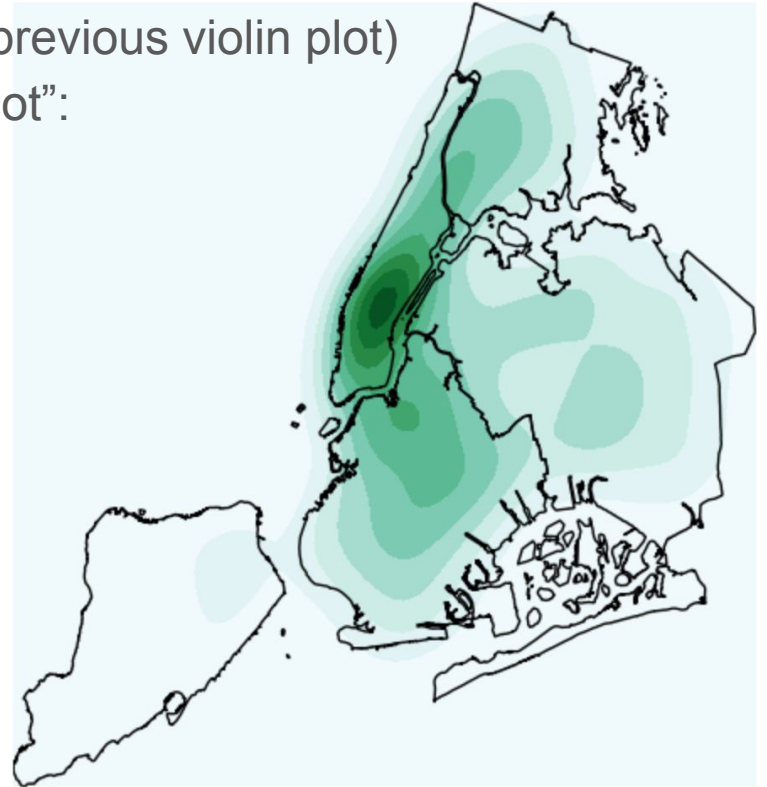
- I want to know how does it change over boroughs
- I want to know the density with higher granularity

Distribute over boroughs?

- Bar chart
- Choropleth map
- Dot map
- ...

Heatmap

- Kernel Density Estimation (as used in the previous violin plot)
- Check “kdeplot” provided by library “GeoPlot”:
 - <https://residentmario.github.io/geoplot/kdeplot.html>



Heatmap

A popular and effective library “folium”. We can add a layer of heatmap to the original real map. Check the code below:

```
# Create Monday Map
# latitude first, then longitude
monday = folium.Map([40.67, -73.97], zoom_start=10)
# Add heatMap
plugins.HeatMap(monday_trips[['Pickup_latitude',
'Pickup_longitude']].values, radius=15).add_to(monday)
# Print heatMap
monday
```

Try to plot the heatmap for drop off

Don't forget to add drop_off location to the dataframe

Choropleth Map

Now we change to a data set with unemployment rate in all states (2012 Oct.)

Many libraries provide support for choropleth map in Python:

- Folium (we just used for the heatmap)
- Plotly
- ...

Choropleth Map

```
# Initialize the map:
m = folium.Map(location=[37, -102],
zoom_start=3)

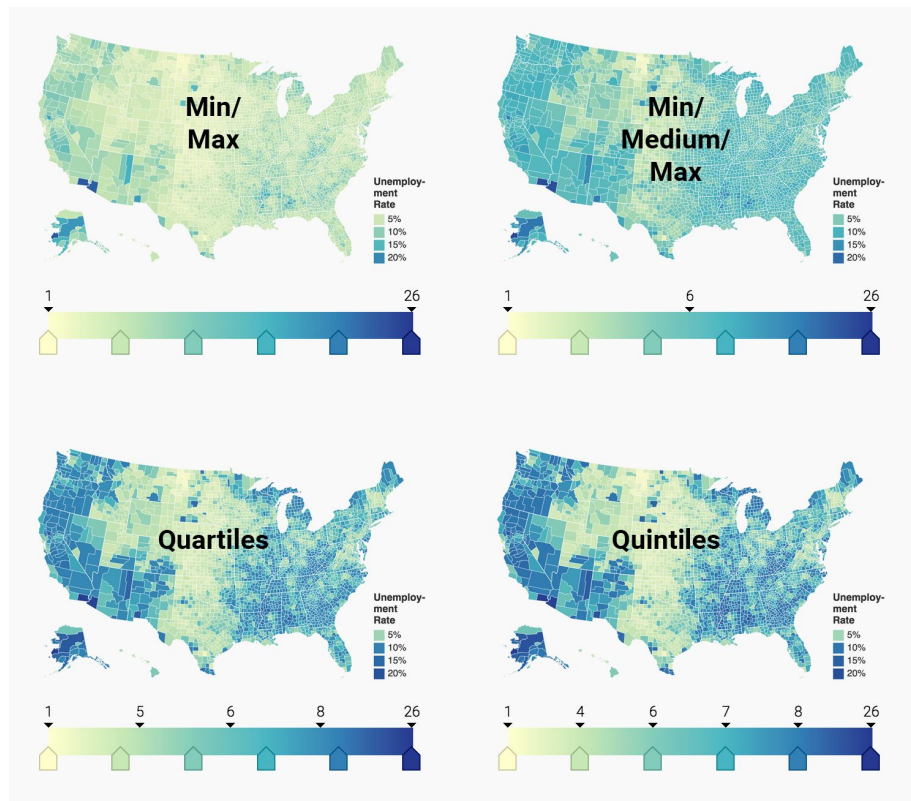
# Add the color for the choropleth:
m.choropleth(
    geo_data='us-states.json',
    name='choropleth',
    data=us_df,
    columns=['State', 'Unemployment'],
    key_on='feature.id',
    fill_color='YlGn',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Unemployment Rate (%)'
)
folium.LayerControl().add_to(m)
```

Geo_data: the map rendering

Name: decide the map type

Fill_color: palette are predefined in RColorBrewer palettes.

Try different color scales



```
m2.choropleth(  
    geo_data='us-states.json',  
    name='choropleth',  
    data=us_df,  
    columns=['State', 'Unemployment'],  
    key_on='feature.id',  
    fill_color='YlGn',  
    fill_opacity=0.7,  
    line_opacity=0.2,  
    legend_name='Unemployment Rate (%)',  
    threshold_scale=stops  
)
```

Set the parameter “threshold_scale”

Finish the code to calculate quartiles.

Dot map

- Add dot to an existing map, supported by most plot libraries
- Here we can something else:
 - Scatter plot

```
p = monday_trips.plot(  
    kind='scatter',  
    x='Pickup_longitude',  
    y='Pickup_latitude',  
    color='white',  
    xlim=(-74.06,-73.77),  
    ylim=(40.61, 40.91),  
    s=.02,  
    alpha=.6)
```

How does the pick-up
distribution change over time?

Sketch it

Check one possible solution in class

Check other things you can do with map

- Draw lines on the map (e.g. flight path): <https://plot.ly/python/lines-on-maps/>
- Use different projection for a map: <https://matplotlib.org/basemap/users/mapsetup.html>
- Bubble map (dot map) using folium: <https://python-graph-gallery.com/313-bubble-map-with-folium/>
- Cartogram using geoplot: Example of population estimates for African countries http://geopandas.org/gallery/plotting_with_geoplot.html
- ...