Part II - Ford GoBike Data Set Analysis

by Shanshan Chu

Investigation Overview

> For the given data set which contains trip information for a bike-sharing system, my goal to know more time and location related information for the collected trips, as well as the user information. Such data exploration of this data set will give us some insights about the bike-sharing market in SF area.

Key Findings

> User background: The most users for Ford GoBike are around 20 to 40 years old, male. Most of the users are subscribers. Using time: The most bike using time is on weekdays during rush hours for work. Usage duration: despite those extremely long and short usage, most users use the bike less than 40 minutes. Gender and user type does not have a large impact on the duration time.

Dataset Overview

> The data set that I analyzed includes information about individual rides made in a bike-sharing system (Ford GoBike) covering the greater San Francisco Bay area. The data includes information about each trip for Ford GoBike (e.g. the start and end time, the start and ending station) and user information (e.g. user's subscription status and ages).

In [35]:
```
# import all packages and set plots to be embedded inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb

%matplotlib inline

# suppress warnings from final output
import warnings
warnings.simplefilter("ignore")
```

In [36]:
```
# load in the dataset into a pandas dataframe
```

```
df = pd.read_csv('GoBikeDataClean.csv')
df.head()
```

Out[36]:

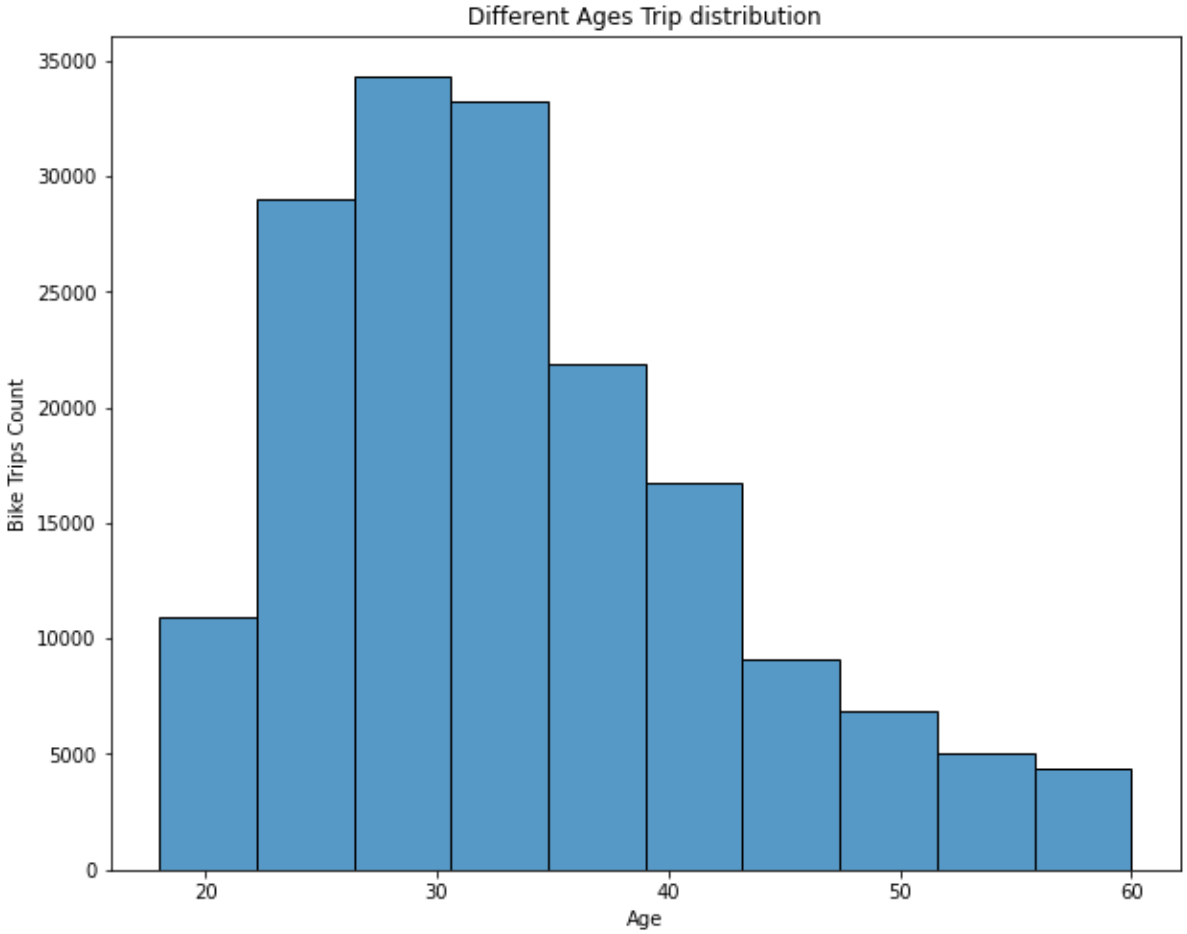
	Unnamed: 0	start_time	end_time	start_station_name	end_station_name	bike_id	u
0	0	2019-02-28 17:32:10.145	2019-03-01 08:01:55.975	Montgomery St BART Station (Market St at 2nd St)	Commercial St at Montgomery St	4902	
1	2	2019-02-28 12:13:13.218	2019-03-01 05:24:08.146	Market St at Dolores St	Powell St BART Station (Market St at 4th St)	5905	
2	3	2019-02-28 17:54:26.010	2019-03-01 04:02:36.842	Grove St at Masonic Ave	Central Ave at Fell St	6638	S
3	4	2019-02-28 23:54:18.549	2019-03-01 00:20:44.074	Frank H Ogawa Plaza	10th Ave at E 15th St	4898	S
4	5	2019-02-28 23:49:58.632	2019-03-01 00:19:51.760	4th St at Mission Bay Blvd S	Broadway at Kearny	5200	S

User background

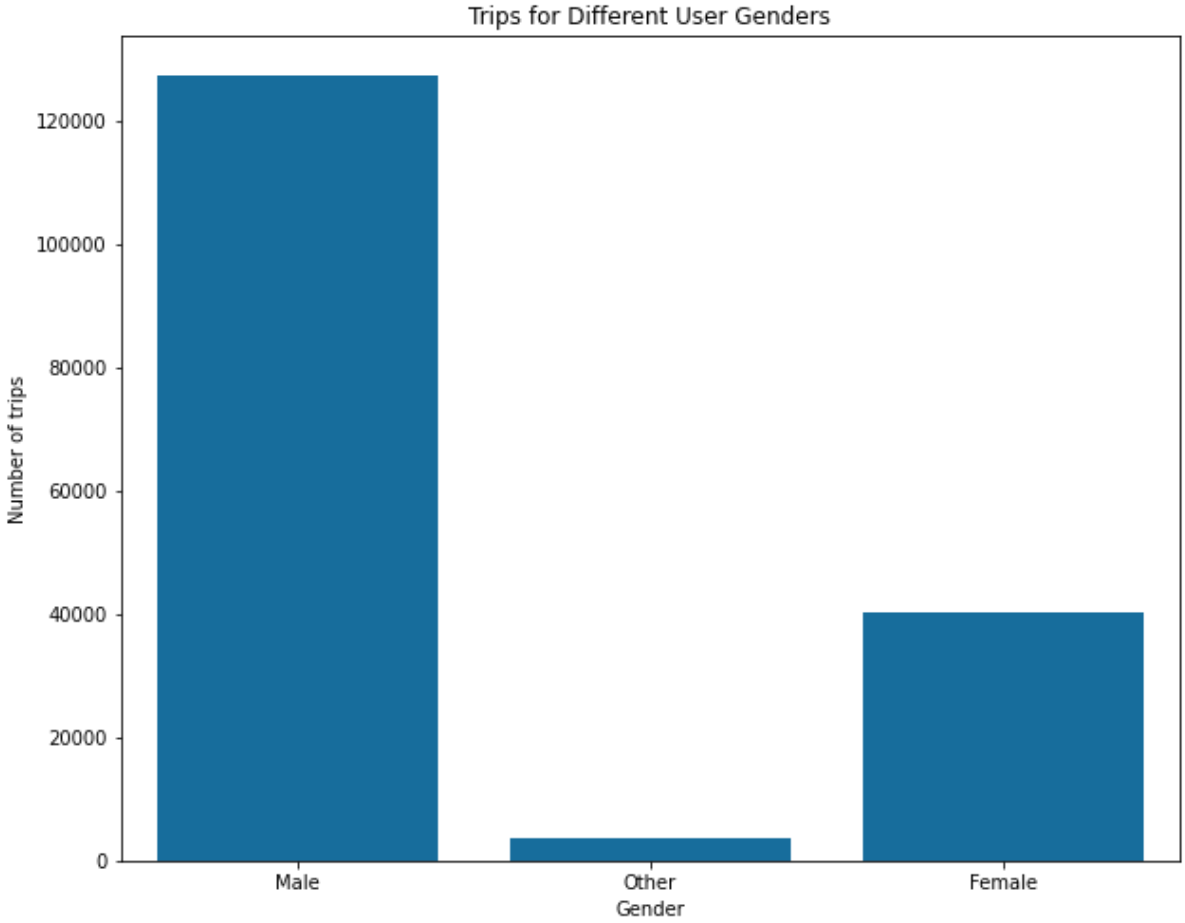
> For this part, I will present my findings about user background. I will show the users' age distribution, genders, and subscribe status.

In [37]:

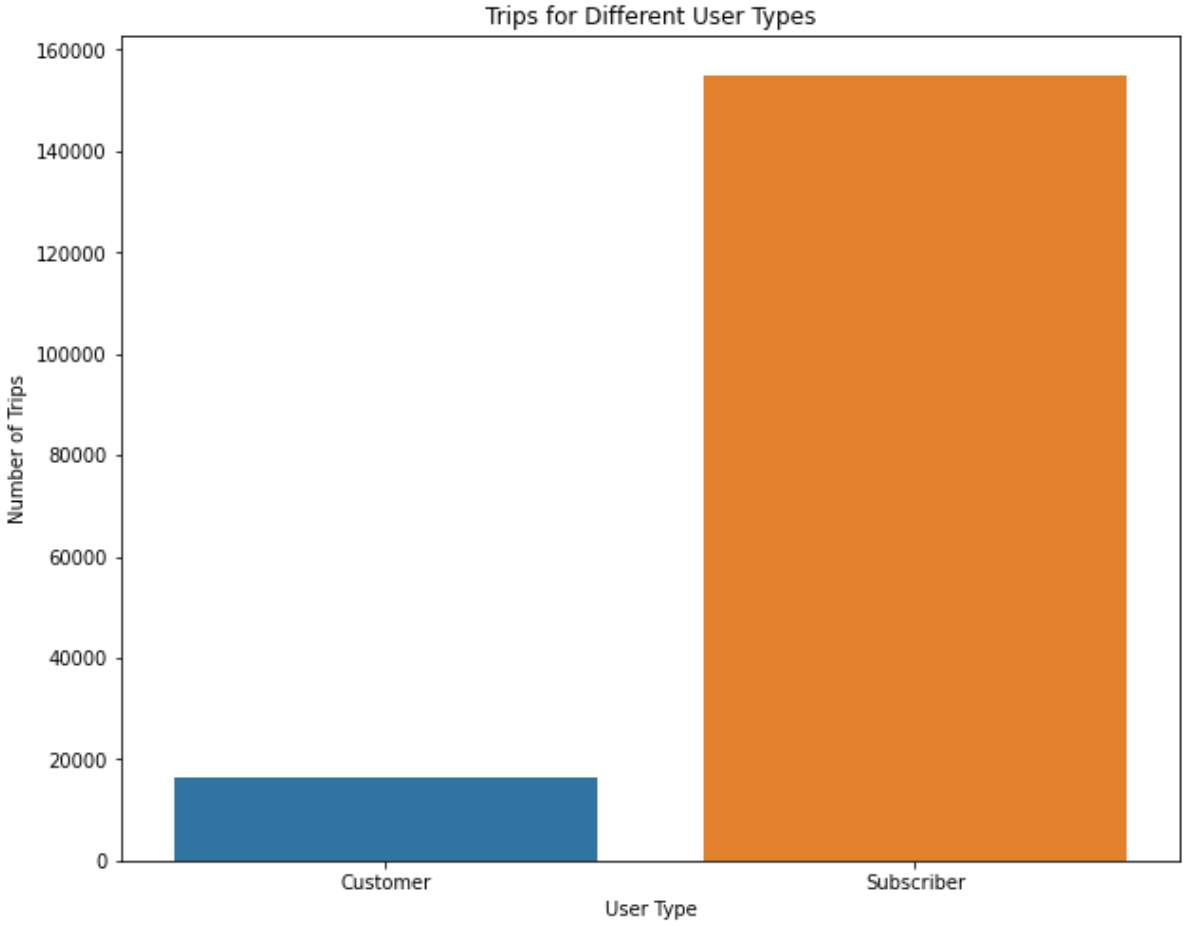
```
# Set color
color = sb.color_palette('colorblind')[0]

plt.figure(figsize=(10,8))
sb.histplot(x = 'member_age', data = df,bins=10)
plt.title('Different Ages Trip distribution')
plt.xlabel('Age')
plt.ylabel('Bike Trips Count');
```



In [38]:
```
plt.figure(figsize = [10, 8])
sb.countplot(data=df, x='member_gender',color=color);
plt.title('Trips for Different User Genders')
plt.xlabel('Gender');
plt.ylabel('Number of trips');
```



In [39]:
```
plt.figure(figsize=(10,8))
sb.countplot(data=df, x='user_type')
plt.title('Trips for Different User Types')
plt.ylabel('Number of Trips')
plt.xlabel('User Type');
```

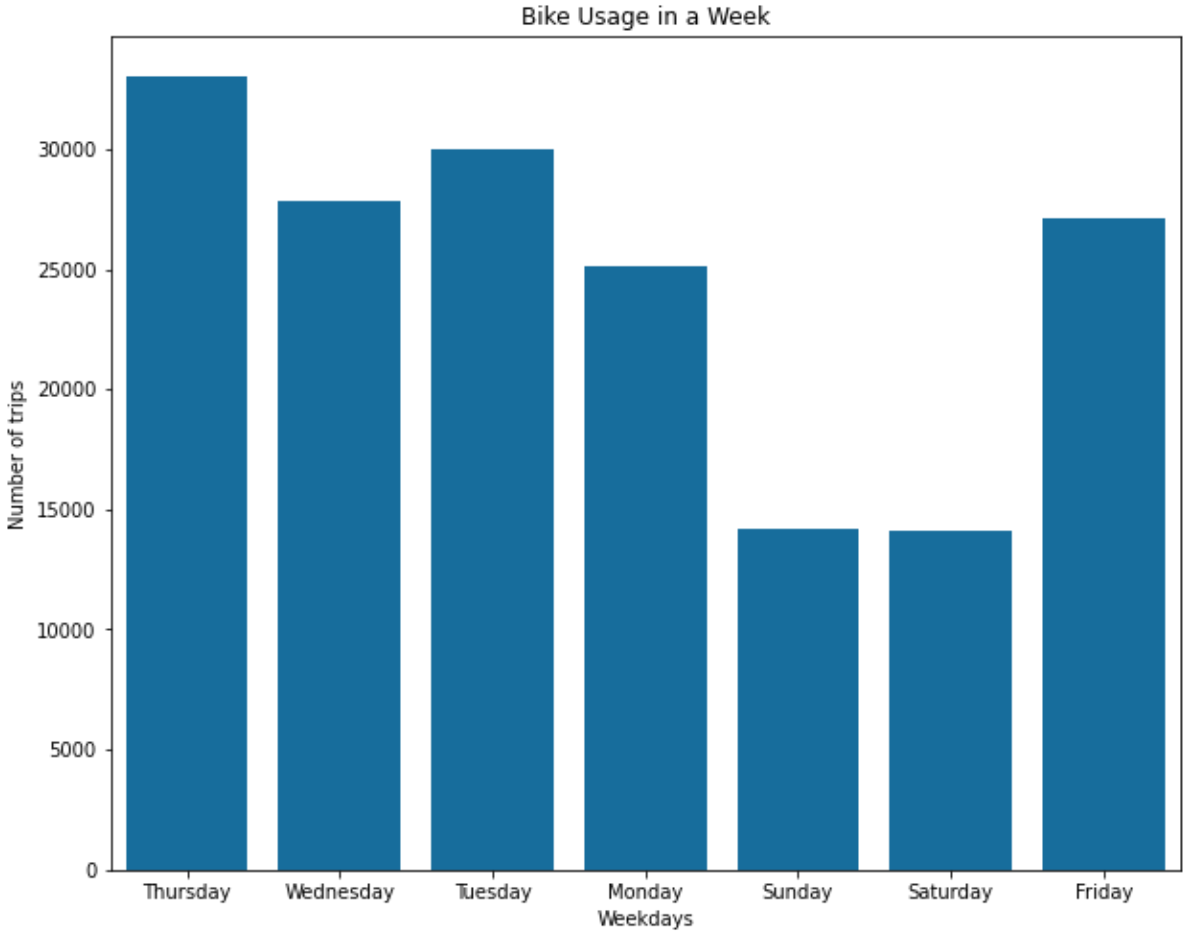


> See from the results, I found that the most users for Ford GoBike are around 20 to 40 years old, male. Most of the users are subscribers.

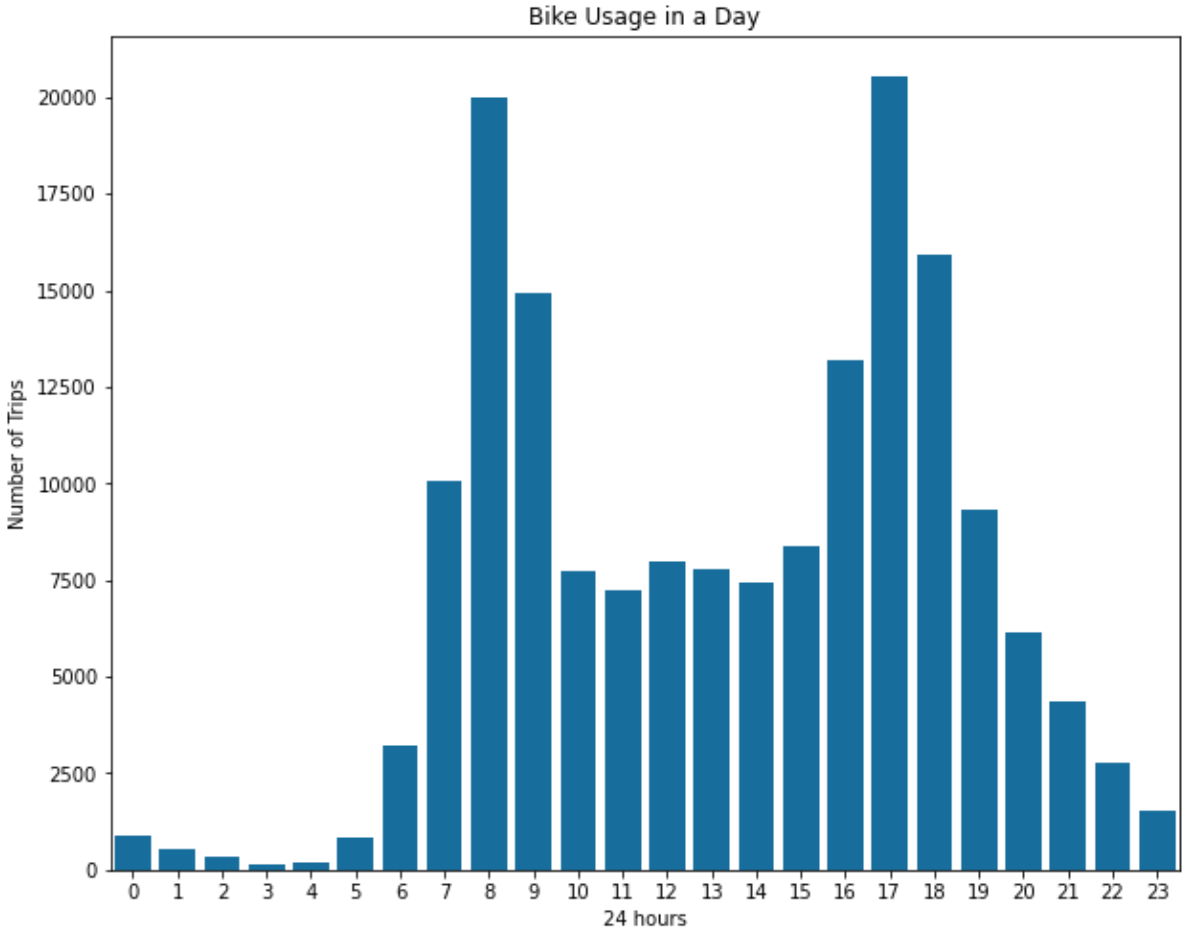
Using time

> For this part, I will show the users' start/end time in the scale of both 24 hours and 7 days.

In [40]:
```
plt.figure(figsize=(10,8))
sb.countplot(x = 'weekday', data = df, color=color)
plt.title('Bike Usage in a Week')
plt.xlabel('Weekdays')
plt.ylabel('Number of trips');
```



In [41]:
```
plt.figure(figsize=(10,8))
sb.countplot(data = df, x='start_time_hour', color=color)
plt.title("Bike Usage in a Day")
plt.xlabel("24 hours")
plt.ylabel("Number of Trips")
plt.show();
```



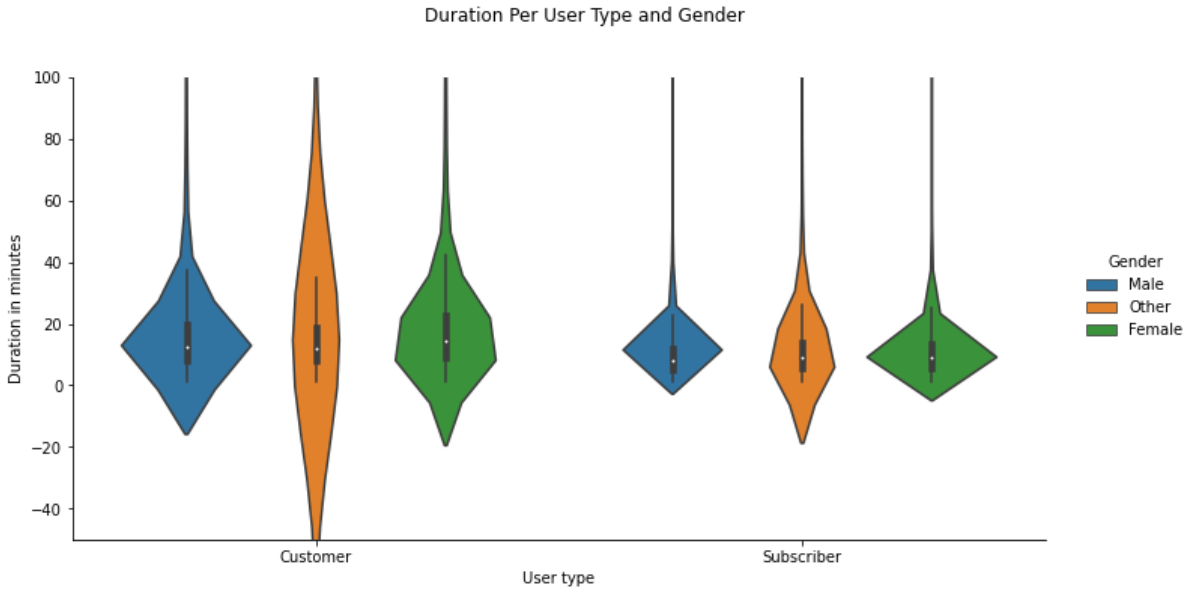
> See from the result, I found that the most bike using time is on weekdays during rush hours for work.

Usage duration

> For this part, I will use the plot of the duration for different gender and user types as example, to demonstrate the point that gender and user type does not have a large impact on the duration time.

In [42]:
```
plt.figure(figsize=(10,8))
graph = sb.catplot(data=df, x='user_type', y="duration_minute", hue="member_

plt.ylim([-50, 100])
graph.set_axis_labels("User type", "Duration in minutes")
graph._legend.set_title('Gender')
graph.fig.suptitle('Duration Per User Type and Gender', y=1.1);
```

```
<Figure size 720x576 with 0 Axes>
```



> See from the result, I found that despite those extremely long and short usage, most users use the bike less than 40 minutes. Gender and user type does not have a large impact on the duration time.

Generate Slideshow

Once you're ready to generate your slideshow, use the `jupyter nbconvert` command to generate the HTML slide show.

In [ ]:
```
# Covnert to slides
!jupyter nbconvert Part_II_slide_deck.ipynb --to slides --post serve --templ
```

```
[NbConvertApp] Converting notebook Part_II_slide_deck.ipynb to slides
[NbConvertApp] Writing 420133 bytes to Part_II_slide_deck.slides.html
[NbConvertApp] Redirecting reveal.js requests to https://cdnjs.cloudflare.com/ajax/libs/reveal.js/3.5.0
Serving your slides at http://127.0.0.1:8000/Part_II_slide_deck.slides.html
Use Control-C to stop this server
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: x-www-browser: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: firefox: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: iceweasel: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: seamonkey: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: mozilla: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: epiphany: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: konqueror: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: chromium-browser: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: google-chrome: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: www-browser: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: links2: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: elinks: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: links: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: lynx: not found
/usr/bin/xdg-open: 778: /usr/bin/xdg-open: w3m: not found
xdg-open: no method available for opening 'http://127.0.0.1:8000/Part_II_slide_deck.slides.html'
```

In [ ]: