***Overview***

In this project, you will make use of R to explore data related to bike share systems for three major cities in the United States—Chicago, New York City, and Washington. You will write code to import the data and answer interesting questions about it by computing descriptive statistics and making visualizations!

***Bike Share Data***

Over the past decade, bicycle-sharing systems have been growing in number and popularity in cities across the world. Bicycle-sharing systems allow users to rent bicycles on a very short-term basis for a price. This allows people to borrow a bike from point A and return it at point B, though they can also return it to the same location if they'd like to just go for a ride. Regardless, each bike can serve several users per day.

Thanks to the rise in information technologies, it is easy for a user of the system to access a dock within the system to unlock or return bicycles. These technologies also provide a wealth of data that can be used to explore how these bike-sharing systems are used.

In this project, you will use data provided by Motivate(opens in a new tab), a bike share system provider for many major cities in the United States, to uncover bike share usage patterns. You will compare the system usage between three large cities: Chicago, New York City, and Washington, DC.

***The Datasets***

Randomly selected data for the first six months of 2017 are provided for all three cities. All three of the data files contain the same core six (6) columns:

Start Time (e.g., 2017-01-01 00:07:57)

End Time (e.g., 2017-01-01 00:20:53)

Trip Duration (in seconds - e.g., 776)

Start Station (e.g., Broadway & Barry Ave)

End Station (e.g., Sedgwick St & North Ave)

User Type (Subscriber or Customer)

The Chicago and New York City files also have the following two columns:

Gender

Birth Year

The original files are much larger and messier, and you don't need to download them, but they can be accessed here if you'd like to see them (Chicago(opens in a new tab), New York City(opens in a new tab), Washington(opens in a new tab)). These files had more columns and they differed in format in many cases. Some data wrangling(opens in a new tab) has been performed to condense these files to the above core six columns to make your analysis and the evaluation of your R skills more straightforward.

***Statistics Computed***

There are a number of different areas of interest available in this dataset. Here are some ideas we came up with for exploration, but feel free to explore any question you are interested in.

***#1 Popular times of travel (i.e., occurs most often in the start time)***

What is the most common month?

What is the most common day of week?

What is the most common hour of day?

***#2 Popular stations and trip***

What is the most common start station?

What is the most common end station?

What is the most common trip from start to end (i.e., most frequent combination of start station and end station)?

***#3 Trip duration***

What is the total travel time for users in different cities?

What is the average travel time for users in different cities?

***#4 User info***

What are the counts of each user type?

What are the counts of each gender (only available for NYC and Chicago)?

What are the earliest, most recent, most common year of birth (only available for NYC and Chicago)?

***The Files***

You will answer your questions of interest about the Motivate datasets using R. To help guide your work in this project, a template with helper code and comments is provided in a notebook file on the next page. You will need the three city dataset files too:

chicago.csv

new\_york\_city.csv

washington.csv