

Explore_bikeshare_data

July 2, 2020

0.0.1 Explore Bike Share Data

For this project, your goal is to ask and answer three questions about the available bikeshare data from Washington, Chicago, and New York. This notebook can be submitted directly through the workspace when you are confident in your results.

You will be graded against the project [Rubric](#) by a mentor after you have submitted. To get you started, you can use the template below, but feel free to be creative in your solutions!

```
In [12]: ny = read.csv('new_york_city.csv')
        wash = read.csv('washington.csv')
        chi = read.csv('chicago.csv')
```

```
In [6]: head(ny)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station
5688089	2017-06-11 14:55:05	2017-06-11 15:08:21	795	Suffolk St & Stanton St	W Broadw
4096714	2017-05-11 15:30:11	2017-05-11 15:41:43	692	Lexington Ave & E 63 St	1 Ave & E 7
2173887	2017-03-29 13:26:26	2017-03-29 13:48:31	1325	1 Pl & Clinton St	Henry St &
3945638	2017-05-08 19:47:18	2017-05-08 19:59:01	703	Barrow St & Hudson St	W 20 St & 8
6208972	2017-06-21 07:49:16	2017-06-21 07:54:46	329	1 Ave & E 44 St	E 53 St & 3
1285652	2017-02-22 18:55:24	2017-02-22 19:12:03	998	State St & Smith St	Bond St &

```
In [17]: head(wash)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station
1621326	2017-06-21 08:36:34	2017-06-21 08:44:43	489.066	14th & Belmont St NW	
482740	2017-03-11 10:40:00	2017-03-11 10:46:00	402.549	Yuma St & Tenley Circle NW	
1330037	2017-05-30 01:02:59	2017-05-30 01:13:37	637.251	17th St & Massachusetts Ave NW	
665458	2017-04-02 07:48:35	2017-04-02 08:19:03	1827.341	Constitution Ave & 2nd St NW/DOL	
1481135	2017-06-10 08:36:28	2017-06-10 09:02:17	1549.427	Henry Bacon Dr & Lincoln Memorial	
1148202	2017-05-14 07:18:18	2017-05-14 07:24:56	398.000	1st & K St SE	

```
In [18]: head(chi)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station
1423854	2017-06-23 15:09:32	2017-06-23 15:14:53	321	Wood St & Hubbard St	Dan
955915	2017-05-25 18:19:03	2017-05-25 18:45:53	1610	Theater on the Lake	She
9031	2017-01-04 08:27:49	2017-01-04 08:34:45	416	May St & Taylor St	Wo
304487	2017-03-06 13:49:38	2017-03-06 13:55:28	350	Christiana Ave & Lawrence Ave	St.
45207	2017-01-17 14:53:07	2017-01-17 15:02:01	534	Clark St & Randolph St	Des
1473887	2017-06-26 09:01:20	2017-06-26 09:11:06	586	Clinton St & Washington Blvd	Car

0.0.2 Question 1

Which gender from New York spend the most total time riding? What is the average time spent per ride by gender?

```
In [13]: ##Loading relevant r-packages
library(magrittr)
library(dplyr)

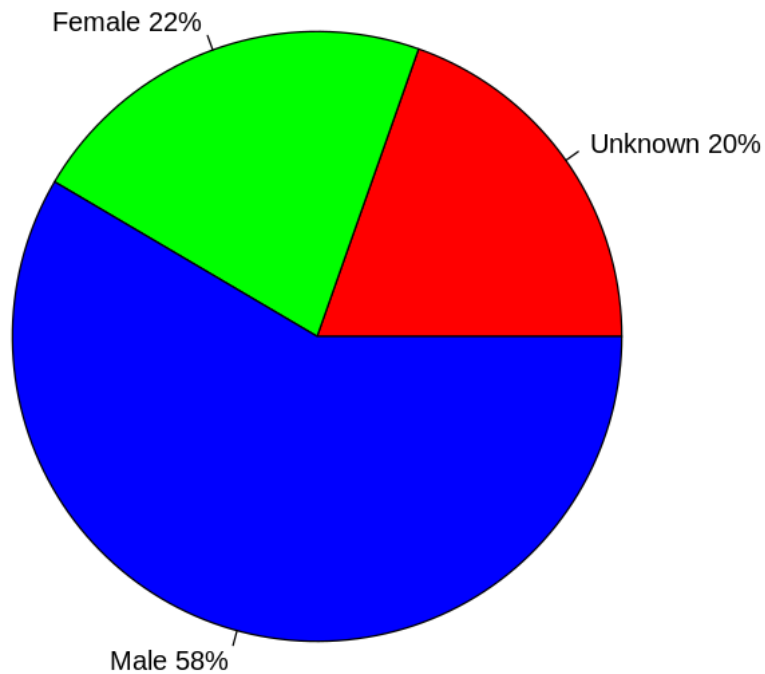
##Finding the gender with the most rides
highest.rider.ny = ny %>% group_by(Gender) %>% summarise(Total.Trip.Duration = sum(Trip
highest.rider.ny

##Plotting a Pie Chart for the representation
slices = c(52884891, 59218793, 157801564)
lbls = c("Unknown", "Female", "Male")

parts = round(slices/sum(slices)*100)
lbls = paste(lbls, parts)
lbls = paste(lbls,"%",sep="")
pie(slices,labels = lbls, col=rainbow(length(lbls)),
    main="Total Ride Time by Gender")
```

Gender	Total.Trip.Duration
	NA
Female	10651022
Male	28604515

Total Ride Time by Gender



```
In [14]: ##Loading relevant r-packages
library(magrittr)
library(dplyr)

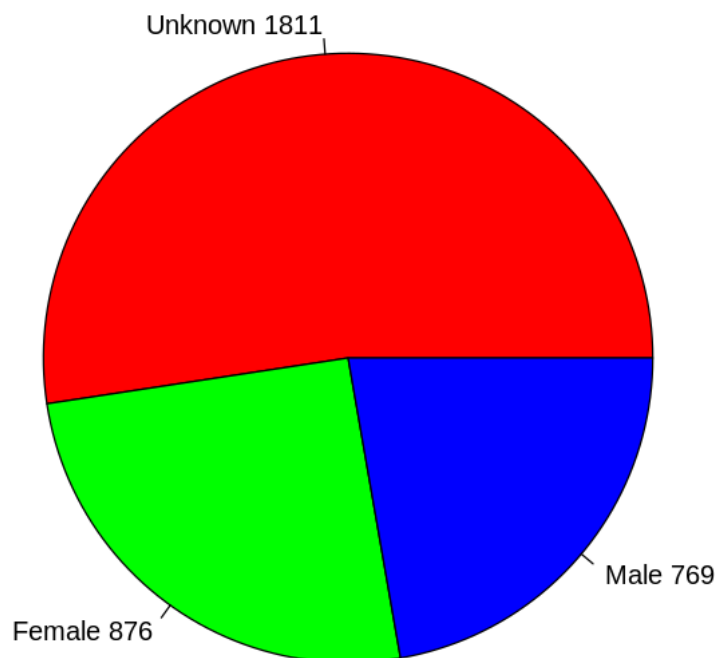
##Finding the gender which spends the most riding time
average.travel.ny = ny %>% group_by(Gender) %>% summarise(Total.Trip.Duration = mean(Trip.Duration))

##Plotting a Pie Chart for the representation
slices = c(1811, 875.9785, 768.9179)
lbls = c("Unknown", "Female", "Male")
parts = round(slices)
lbls = paste(lbls, parts)
```

```
pie(slices,labels = lbls, col=rainbow(length(lbls)),
    main="Average Ride Time per trip by Gender")
```

Gender	Total.Trip.Duration
	NA
Female	875.9785
Male	768.9179

Average Ride Time per trip by Gender



We can see that males from new york travel more than females as their total travel time constitute 58% of the overall total travel time—more than the remaining two gender categories combined.

Even though males in total spent a higher time travelling, they spent the least time travelling PER travel of 774 minutes where as females spend an average of 887 minutes

0.0.3 Question 2

What are the top three start locations that bikers in Chicago frequently start their riding journey? Based on the proportion of trips starting from there, are they the favourite starting points for bike trips in Chicago?

```
In [15]: ##Reading the Chicago bikeshare data
        chi = read.csv('chicago.csv')

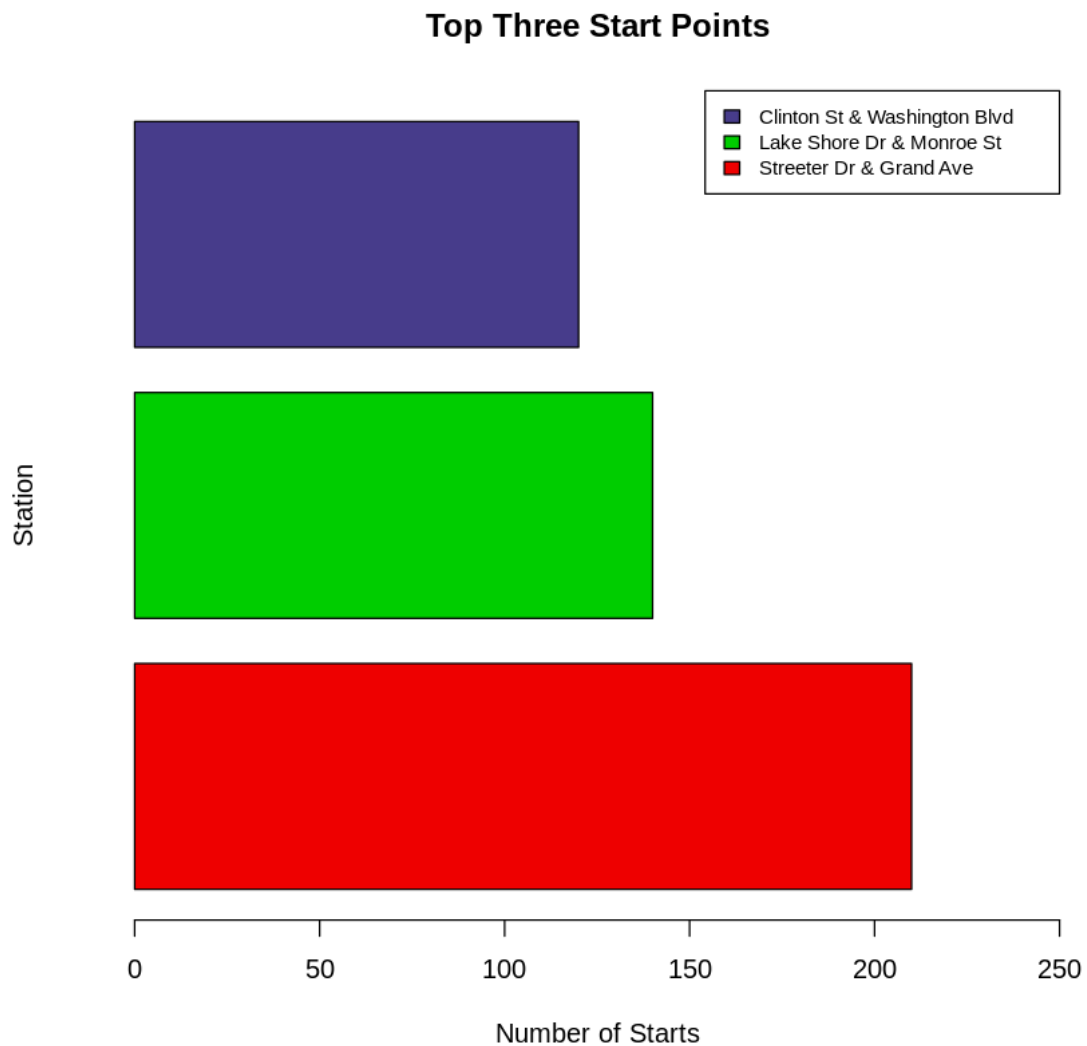
        ##Loading relevant r-packages
        library(magrittr)
        library(dplyr)

        ##Finding the top three(3) favourite Start Stations
        oldest.rider.chi = chi %>% group_by(Start.Station) %>% summarise(Number.of.bike.trips=n)
        arranged.oldest.rider.chi=oldest.rider.chi %>% arrange(desc(Number.of.bike.trips))
        top.arranged.oldest.rider.chi = arranged.oldest.rider.chi %>% top_n(3)
        top.arranged.oldest.rider.chi

        ##Plotting a Bar Graph
        numbik = c(210,140,120)
        barplot(nummik, main = "Top Three Start Points", xlab = "Number of Starts", ylab = "Sta
```

Selecting by Number.of.bike.trips

Start.Station	Number.of.bike.trips
Streeter Dr & Grand Ave	210
Lake Shore Dr & Monroe St	140
Clinton St & Washington Blvd	120



```
In [30]: ##Reading the Chicago bikeshare data
chi = read.csv('chicago.csv')

##Loading relevant r-packages
library(magrittr)
library(dplyr)

##Finding the percentage of the top three start points and the rest out of the total data
total = nrow(chi)
numbik = c(210,140,120,total-(210+140+120))

##Using a function to find the percentage
```

```

numbik_to_percent = function(temp_N) {
  temp_P = (temp_N/total)*100
  return(temp_P)
}

##Generating the values from the function
slice_for_pie=numbik_to_percent(numbik)
slice_for_pie

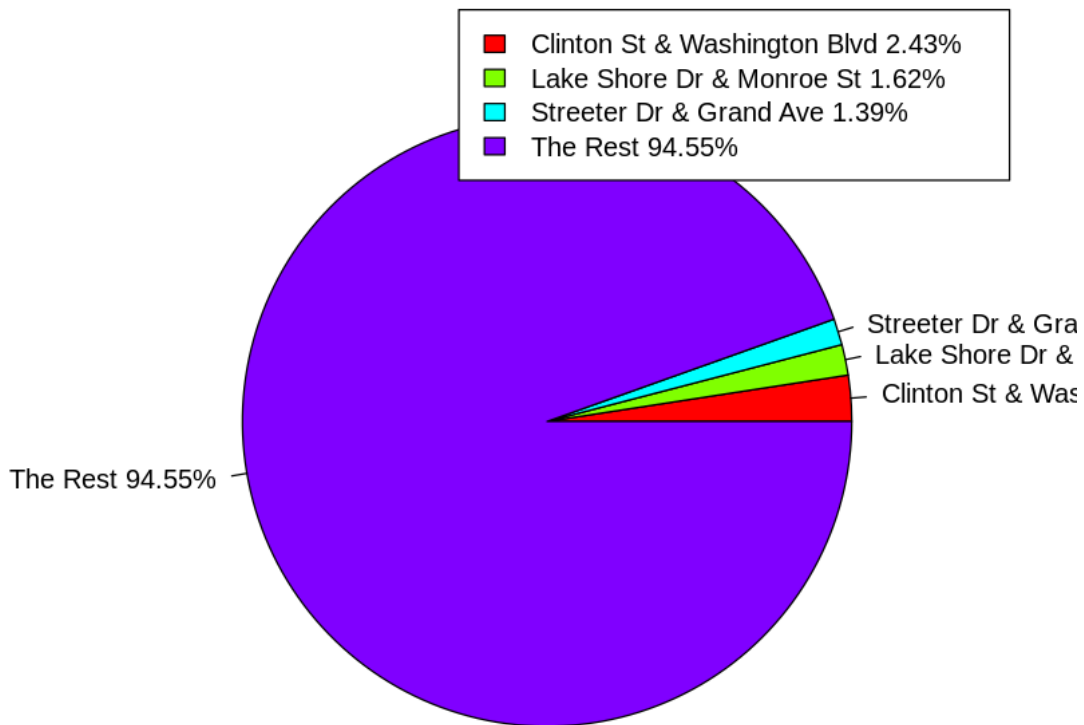
##Ploting a Pie Chart for the representation
slices = slice_for_pie
lbls = c("Clinton St & Washington Blvd", "Lake Shore Dr & Monroe St", "Streeter Dr & Gr

parts = round(slices/sum(slices)*100, digits=2)
lbls = paste(lbls, parts)
lbls = paste(lbls,"%",sep="")
pie(slices,labels = lbls, col=rainbow(length(lbls)),
    main="Percentage Representation of the Preferred Start Points")
legend("topright",legend=lbls, col=rainbow(length(lbls)),fill=rainbow(length(lbls)))

```

1. 2.43337195828505 2. 1.62224797219003 3. 1.39049826187717 4. 94.5538818076477

Percentage Representation of the Preferred Start Points



The top three start point turn out to be “Streeter Dr & Grand Avenue”, “Lake Shore DR & Monroe St” and “Clinton St & Washington Blvd”. 210 bike trips started from “Streeter Dr & Grand Avenue”, 140 trips started from “Lake Shore DR & Monroe St” and 120 trips started from “Clinton St & Washington Blvd”.

However, we could not say these three are favourite starting points since they do not together consist up to half of the start point. 2.43% of the bike trips started from “Streeter Dr & Grand Avenue”, 1.62% started from “Lake Shore DR & Monroe St” and 1.39% started from “Clinton St & Washington Blvd”. They just form 5.5% of the overall.

Thus though, these three are the top three, we cannot conclude they are the favourite start points.

0.0.4 Question 3

Which type of rider from Washington rides less?


```

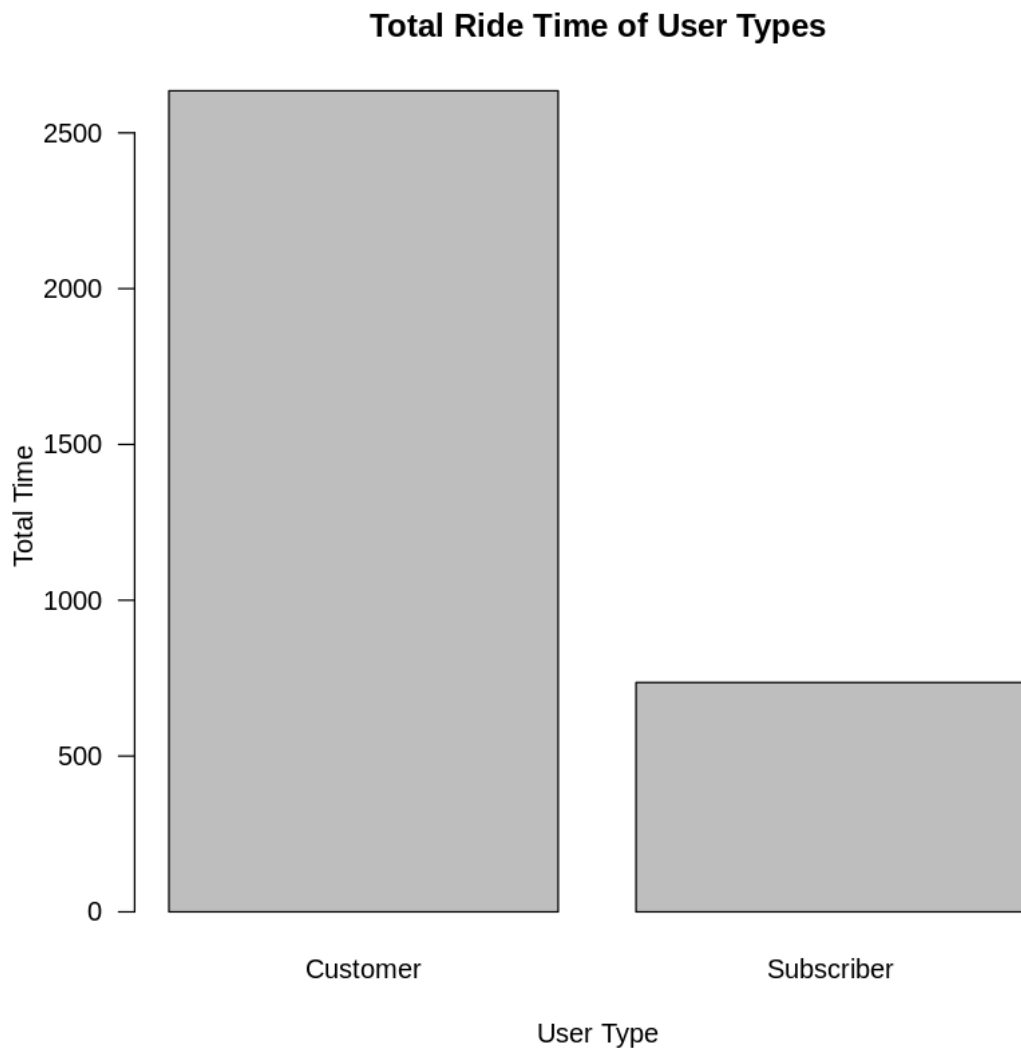
In [3]: ##Loading relevant r-packages
library(magrittr)
library(dplyr)

##Finding the user with the average ride per User Type
low.rides.wash = wash %>% group_by(User.Type) %>% summarise(Total.Trip.Duration = mean(T
low.rides.wash

##Plotting a Bar Graph
numbik=c(2635,736)
barplot(numbik, main = "Total Ride Time of User Types", xlab = "User Type", ylab = "Total

```

User.Type	Total.Trip.Duration
	NA
Customer	2634.429
Subscriber	733.326



It can be seen that Customers ride more on the average than Subscribers. The average ride time of customers in Washington is 2635 minutes where as subscribers average 736 per ride

0.1 Finishing Up

Congratulations! You have reached the end of the Explore Bikeshare Data Project. You should be very proud of all you have accomplished!

Tip: Once you are satisfied with your work here, check over your report to make sure that it satisfies all the areas of the [rubric](#).

0.2 Directions to Submit

Before you submit your project, you need to create a .html or .pdf version of this notebook in the workspace here. To do that, run the code cell below. If it worked correctly,

you should get a return code of 0, and you should see the generated .html file in the workspace directory (click on the orange Jupyter icon in the upper left).

Alternatively, you can download this report as .html via the **File > Download as** sub-menu, and then manually upload it into the workspace directory by clicking on the orange Jupyter icon in the upper left, then using the Upload button.

Once you've done this, you can submit your project by clicking on the "Submit Project" button in the lower right here. This will create and submit a zip file with this .ipynb doc and the .html or .pdf version you created. Congratulations!

```
In [31]: system('python -m nbconvert Explore_bikeshare_data.ipynb')
```