Explore_bikeshare_data

March 30, 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 [2]: ny = read.csv('new_york_city.csv')
        wash = read.csv('washington.csv')
        chi = read.csv('chicago.csv')
```

```
In [3]: 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 Broadwa
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 [4]: head(wash)
```

X	Start.Time	End.Time	Trip.Duration	Start.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 [5]: head(chi)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End
1423854	2017-06-23 15:09:32	2017-06-23 15:14:53	321	Wood St & Hubbard St	Da
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

What is the most common month in Chicago?

In [6]:
```
library(ggplot2)
library(plyr)

#Exploring
head(chi)
tail(chi)
str(chi)
summary(chi)
names(chi)

#Checking the frequency of each month
count(chi$start_month)

#Creating a mode function for the starting months.
mode_month <- function(m){
  uniqm <- unique(m)
  uniqm[which.max(tabulate(match(m, uniqm)))]
}

#Calculating which month appeared the most.
mode_month(chi$start_month)

#Converting start.time to be date format
Date <- as.Date(chi$Start.Time)
head(Date)

#Adding a column for chi for the start date
chi$start_date <- Date

#Adding a column for chi for the month only of the start date
chi$start_month <- months(chi$start_date)


#Creating a bar chart for the start month
ggplot(data = chi, aes(x = start_month)) +
  geom_bar(fill = 'blue', color = 'black') +
  ggtitle("Number of Counts in Chicago Per Month") +
  labs(x = "Starting Month", y = "Number of Counts")
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End
1423854	2017-06-23 15:09:32	2017-06-23 15:14:53	321	Wood St & Hubbard St	Da
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

	X	Start.Time	End.Time	Trip.Duration	Start.Station
8625	397518	2017-03-24 16:52:16	2017-03-24 16:57:57	341	Southport Ave & Waveland Av
8626	879494	2017-05-18 05:06:50	2017-05-18 05:22:10	920	Artesian Ave & Hubbard St
8627	360389	2017-03-19 07:21:29	2017-03-19 07:27:18	349	Wabash Ave & Roosevelt Rd
8628	858496	2017-05-16 17:03:24	2017-05-16 17:31:12	1668	Ashland Ave & Harrison St
8629	777620	2017-05-10 08:53:03	2017-05-10 08:54:32	89	Western Ave & Leland Ave
8630	1230561	2017-06-11 14:52:13	2017-06-11 15:42:33	3020	Waba

```
'data.frame':        8630 obs. of  9 variables:
 $ X            : int  1423854 955915 9031 304487 45207 1473887 961916 65924 606841 135470 ...
 $ Start.Time   : Factor w/ 8624 levels "2017-01-01 00:40:14",..: 7876 5303 73 1721 267 8173 534
 $ End.Time     : Factor w/ 8625 levels "2017-01-01 00:46:32",..: 7876 5303 73 1722 267 8173 534
 $ Trip.Duration: int  321 1610 416 350 534 586 281 723 689 493 ...
 $ Start.Station: Factor w/ 472 levels "2112 W Peterson Ave",..: 468 424 291 80 103 119 22 255 3
 $ End.Station  : Factor w/ 471 levels "","2112 W Peterson Ave",..: 132 381 469 409 151 70 467 2
 $ User.Type    : Factor w/ 3 levels "","Customer",..: 3 3 3 3 3 3 3 2 3 3 ...
 $ Gender       : Factor w/ 3 levels "","Female","Male": 3 2 3 3 3 3 2 1 3 3 ...
 $ Birth.Year   : num  1992 1992 1981 1986 1975 ...
```

```
       X                       Start.Time                   End.Time
 Min.   :     36   2017-01-24 07:40:32:   2   2017-04-16 13:16:52:   2
 1st Qu.: 386722   2017-04-22 13:16:25:   2   2017-04-26 16:29:26:   2
 Median : 773554   2017-05-27 15:17:50:   2   2017-05-21 16:20:56:   2
 Mean   : 776721   2017-06-10 13:29:41:   2   2017-05-27 09:58:21:   2
 3rd Qu.:1171266   2017-06-20 17:05:11:   2   2017-06-25 14:51:35:   2
 Max.   :1551248   2017-06-21 13:18:52:   2   2017-01-01 00:46:32:   1
                   (Other)            :8618   (Other)            :8619
 Trip.Duration                          Start.Station
 Min.   :   60.0   Streeter Dr & Grand Ave     : 210
 1st Qu.:  394.2   Lake Shore Dr & Monroe St   : 140
 Median :  670.0   Clinton St & Washington Blvd: 120
 Mean   :  937.2   Clinton St & Madison St     : 102
 3rd Qu.: 1119.0   Canal St & Adams St         : 101
 Max.   :85408.0   Michigan Ave & Oak St       :  98
                   (Other)                    :7859
                      End.Station        User.Type       Gender
 Streeter Dr & Grand Ave     : 233             :   1         :1748
 Clinton St & Madison St     : 145   Customer  :1746   Female:1723
 Theater on the Lake         : 131   Subscriber:6883   Male  :5159
 Lake Shore Dr & Monroe St   : 115
```

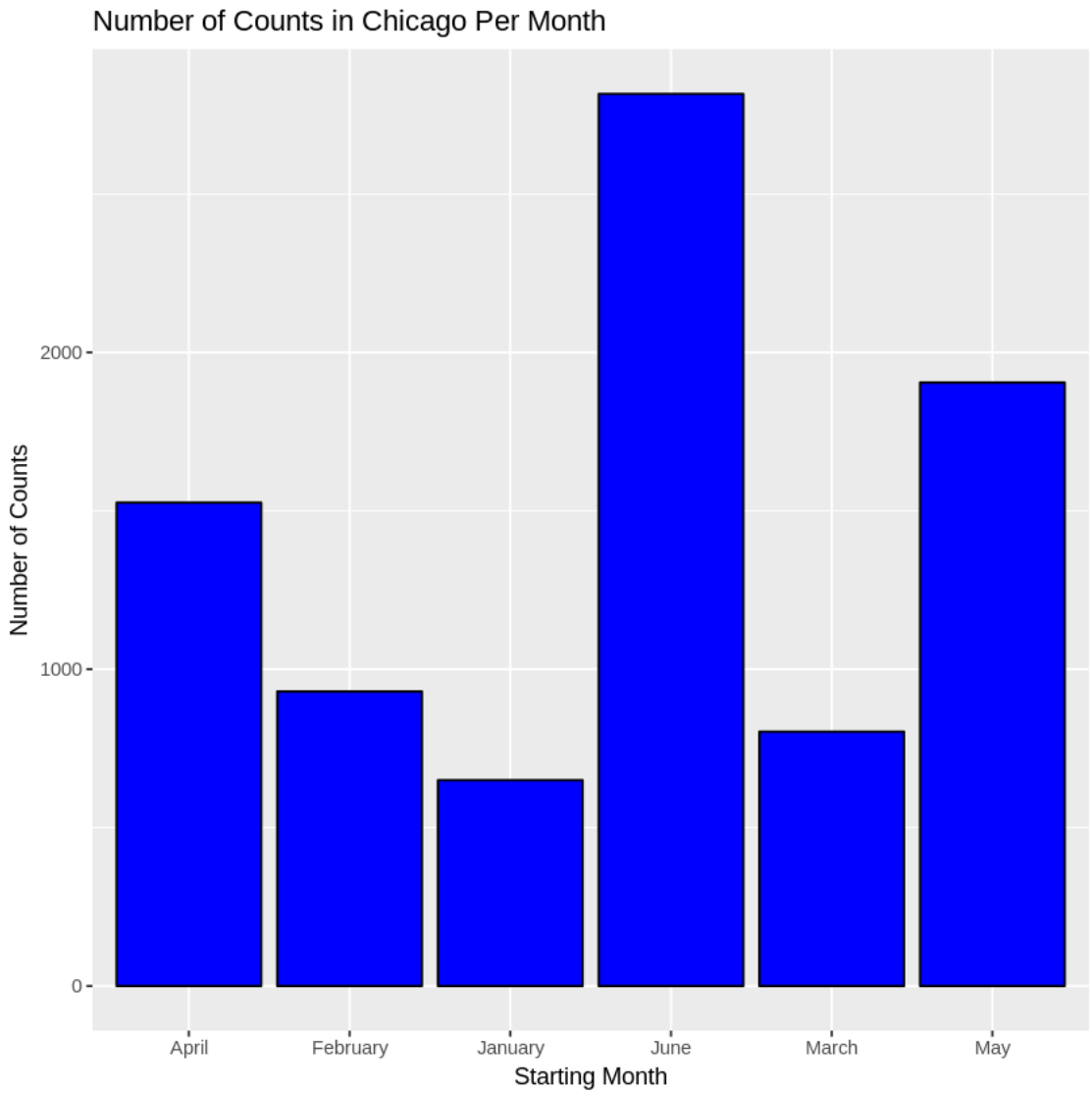
```
 Clinton St & Washington Blvd: 109
 Lake Shore Dr & North Blvd  : 102
 (Other)                     :7795
   Birth.Year
 Min.   :1899
 1st Qu.:1975
 Median :1984
 Mean   :1981
 3rd Qu.:1989
 Max.   :2002
 NA's   :1747
```

1. 'X' 2. 'Start.Time' 3. 'End.Time' 4. 'Trip.Duration' 5. 'Start.Station' 6. 'End.Station' 7. 'User.Type' 8. 'Gender' 9. 'Birth.Year'

freq

```
NULL
```

1. 2017-06-23 2. 2017-05-25 3. 2017-01-04 4. 2017-03-06 5. 2017-01-17 6. 2017-06-26



We could see that June is the most common month in Chicago. There are about 100,000 number of counts in the month of June. May is the second most common month of about 60,000 number of counts. We could conclude as the month increase, the number of counts increase.

0.0.3 Question 2

How often do most people borrow bikes for in New York?

```
In [7]: library(ggplot2)

        #Exploring
        head(ny)
        tail(ny)
        str(ny)
```

```r
summary(ny)
names(ny)

#Creating an average function
avg_trip <- function(x){
  sum(x) / length(x)
}

#Finding the average of the trip duration
avg_trip(ny$Trip.Duration)

#Finding the median of the trip duration
median(ny$Trip.Duration)

#Creating a function for the mode of the trip duration
mode_trip <- function(y){
  uniqy <- unique(y)
  uniqy[which.max(tabulate(match(y, uniqy)))]
}

#Finding the mode of the trip duration, which could also answer our question.
mode_trip(ny$Trip.Duration)

#Creating a histogram of the trip duration
ggplot(data = ny, aes(x = Trip.Duration)) +
  geom_histogram(binwidth = 60, color = 'black', fill = 'green') +
  scale_x_continuous(limits = c(0,2000)) +
  ggtitle("Trip Duration Distribution in New York City") +
  labs(x = "Trip Duration per 60 Seconds", y = "Frequency")
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Statio
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 &

	X	Start.Time	End.Time	Trip.Duration	Start.Station
54765	1293888	2017-02-23 06:14:14	2017-02-23 06:23:32	558	E 27 St & 1 Ave
54766	642855	2017-01-28 16:44:18	2017-01-28 16:48:18	240	W 52 St & 9 Ave
54767	2157959	2017-03-29 06:30:35	2017-03-29 06:32:41	125	W 84 St & Columbus Ave
54768	5679624	2017-06-11 12:52:27	2017-06-11 12:58:35	367	8 Ave & W 33 St
54769	6762960	2017-06-30 07:48:34	2017-06-30 08:17:16	1722	Cathedral Pkwy & Broadway
54770	6078570	2017-06-18 16:20:21	201	NA	

```
'data.frame':        54770 obs. of  9 variables:
 $ X            : int  5688089 4096714 2173887 3945638 6208972 1285652 1675753 1692245 2271331 1
 $ Start.Time   : Factor w/ 54568 levels "2017-01-01 00:17:01",..: 45448 32799 17316 31589 49688
```

```
$ End.Time     : Factor w/ 54562 levels "201","2017-01-01 00:30:56",..: 45432 32783 17295 31567
$ Trip.Duration: int  795 692 1325 703 329 998 478 4038 5132 309 ...
$ Start.Station: Factor w/ 636 levels "","1 Ave & E 16 St",..: 522 406 10 93 5 521 325 309 151
$ End.Station  : Factor w/ 638 levels "","1 Ave & E 16 St",..: 613 8 362 558 269 107 389 110 15
$ User.Type    : Factor w/ 3 levels "","Customer",..: 3 3 3 3 3 3 3 3 2 3 ...
$ Gender       : Factor w/ 3 levels "","Female","Male": 3 3 3 2 3 3 3 3 1 3 ...
$ Birth.Year   : num  1998 1981 1987 1986 1992 ...
```

```
       X                         Start.Time                   End.Time
 Min.   :     47   2017-05-11 18:26:10:    3   2017-01-03 08:54:10:    2
 1st Qu.:1712425   2017-01-04 13:58:24:    2   2017-01-04 17:21:55:    2
 Median :3418634   2017-01-09 09:36:01:    2   2017-01-05 17:25:17:    2
 Mean   :3415873   2017-01-21 15:36:56:    2   2017-01-12 08:34:01:    2
 3rd Qu.:5123382   2017-01-21 17:49:59:    2   2017-01-12 09:41:54:    2
 Max.   :6816152   2017-01-21 20:08:29:    2   2017-01-12 20:34:42:    2
                   (Other)            :54757   (Other)            :54758
 Trip.Duration                    Start.Station
 Min.   :     61.0   Pershing Square North:  592
 1st Qu.:    368.0   W 21 St & 6 Ave      :  385
 Median :    610.0   Broadway & E 22 St   :  383
 Mean   :    903.6   E 17 St & Broadway   :  380
 3rd Qu.:   1051.0   West St & Chambers St:  364
 Max.   :1088634.0   W 20 St & 11 Ave     :  329
 NA's   :1           (Other)              :52337
               End.Station         User.Type        Gender        Birth.Year
 Pershing Square North:  556             :  119         : 5410   Min.   :1885
 E 17 St & Broadway   :  445   Customer  : 5558   Female:12159   1st Qu.:1970
 Broadway & E 22 St   :  427   Subscriber:49093   Male  :37201   Median :1981
 W 21 St & 6 Ave      :  365                                     Mean   :1978
 W 20 St & 11 Ave     :  344                                     3rd Qu.:1988
 W 38 St & 8 Ave      :  338                                     Max.   :2001
 (Other)              :52295                                     NA's   :5218
```

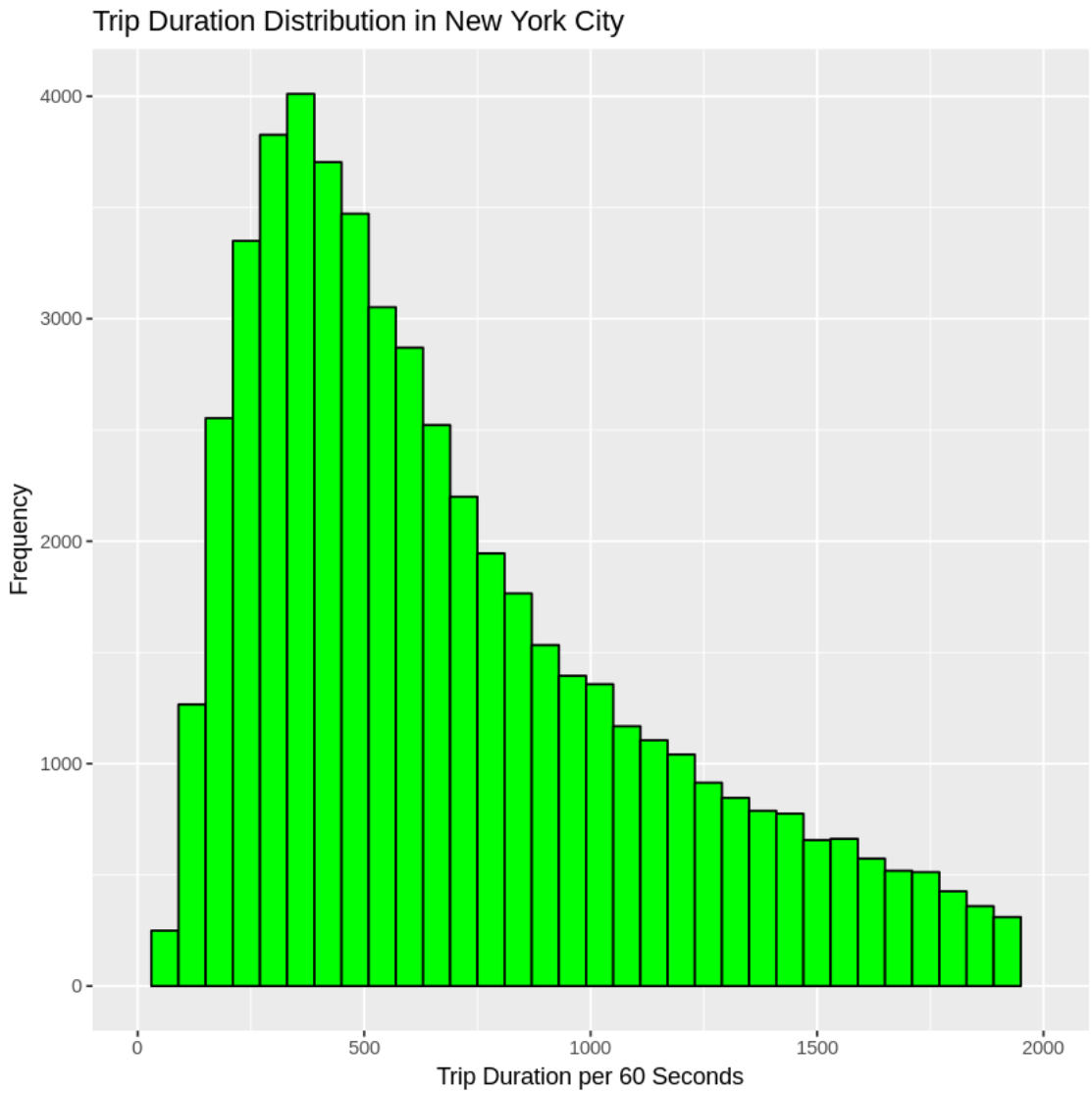
1. 'X' 2. 'Start.Time' 3. 'End.Time' 4. 'Trip.Duration' 5. 'Start.Station' 6. 'End.Station' 7. 'User.Type' 8. 'Gender' 9. 'Birth.Year'

<NA>

<NA>

388

```
Warning message:
Removed 2822 rows containing non-finite values (stat_bin).Warning message:
Removed 2 rows containing missing values (geom_bar).
```



Since the bin width was set to 60 seconds, this histogram shows that most of the frequency appears within 300 to 360 seconds. Our calculated mode is within that range.

0.0.4 Question 3

What is the average travel time for users in Washington based on their user type?

In [8]:
```
library(ggplot2)

#Exploring the summaries, which includes the average trip duration
head(wash)
tail(wash)
str(wash)
summary(wash)
```

```r
#Creating a filter just for the Customer user type
cust_filter <- wash$User.Type == "Customer"

#Creating a data frame for the Washington data with only the Customer user type
wash_cust <- wash[cust_filter,]

#Creating a filter just for the Subscriber user type
sub_filter <- wash$User.Type == "Subscriber"

#Creating a data frame for the Washington data with only the Subscriber user type
wash_sub <- wash[sub_filter,]

#Creating a function for the mean of the trip duration
avg_trip <- function(x){
  sum(x) / length(x)
}

#Finding average of Washington Customer user type trip duration
avg_trip(wash_cust$Trip.Duration)

#Finding average of Washington Subscriber user type trip duration
avg_trip(wash_sub$Trip.Duration)

#Creating a mean for trip duration for Customer user type
avg_cust <- mean(wash_cust$Trip.Duration)

#Creating a mean for trip duration for Subscriber user type
avg_sub <- mean(wash_sub$Trip.Duration)

#Creating a data with two average values
avg_user <- rbind(avg_cust, avg_sub)

#Creating a user type value
user_type <- rbind("Customer", "Subscriber")

#Creating a data frame
avg_trip_user <- data.frame(avg_user, user_type)
avg_trip_user

#Creating a bar graph for Washington trip duration by user type.
ggplot(data = avg_trip_user, aes(x = user_type, y = avg_user)) +
  geom_col(fill = 'blue', color = 'black') +
  ggtitle("Average Trip Duration by User Type") +
  labs(x = "User Type", y = "Average Trip Duration in Seconds")
```

X	Start.Time	End.Time	Trip.Duration	Start.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

	X	Start.Time	End.Time	Trip.Duration	Start.Station	E
89046	1484340	2017-06-10 10:58:09	2017-06-10 11:25:58	1669.700	M St & New Jersey Ave SE	4
89047	555788	2017-03-22 18:46:00	2017-03-22 19:04:00	1082.789	8th & H St NW	2
89048	739004	2017-04-09 04:00:22	2017-04-09 04:09:54	571.879	Eckington Pl & Q St NE	C
89049	1214907	2017-05-19 09:00:53	2017-05-19 09:07:38	404.152	1st & M St NE	1
89050	1419806	2017-06-06 04:27:33	2017-06-06 04:49:59	1345.911	10th & Florida Ave NW	C
89051	132			NA		

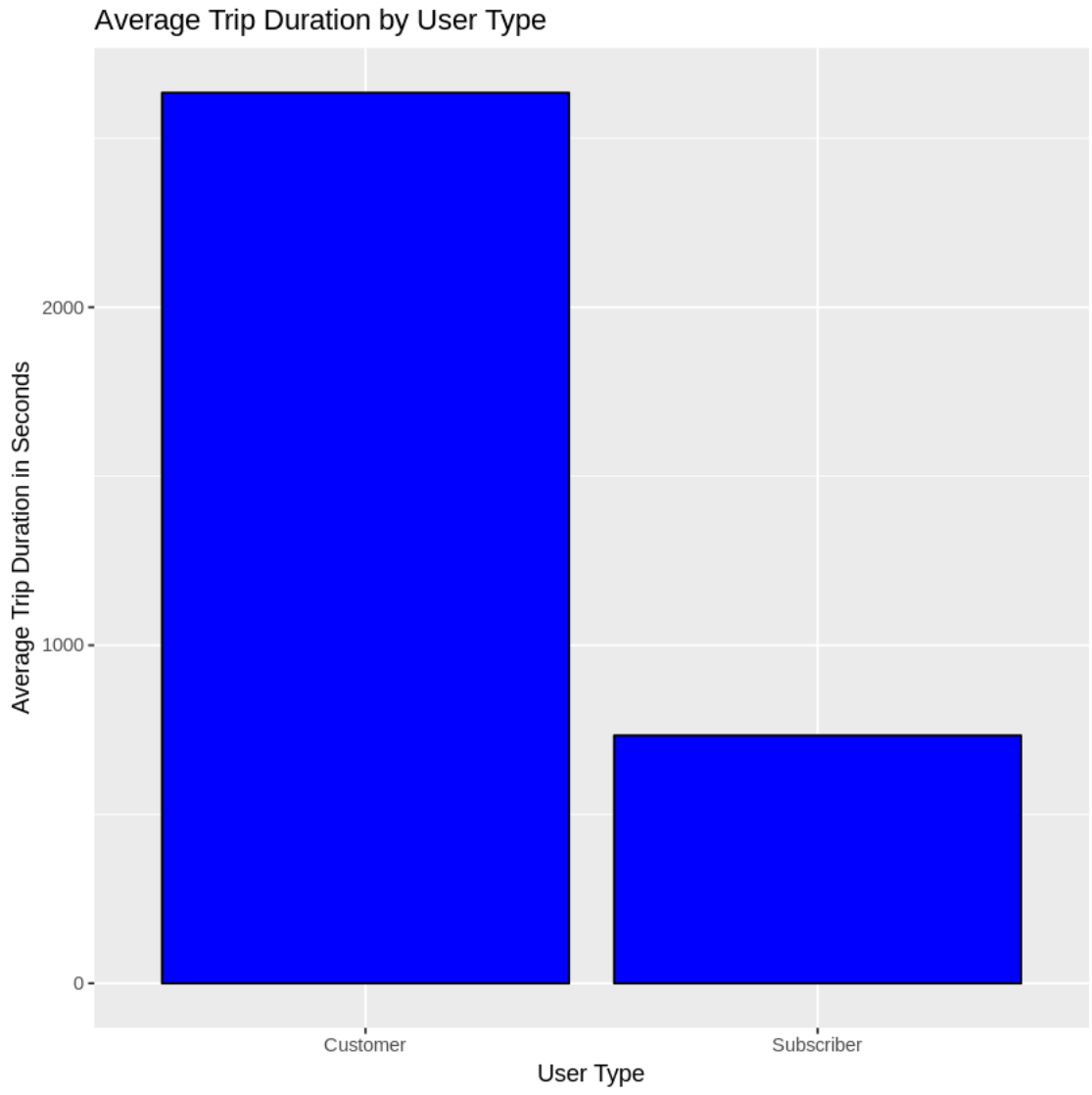
```
'data.frame':        89051 obs. of  7 variables:
 $ X            : int  1621326 482740 1330037 665458 1481135 1148202 1594275 1601832 574182 3270
 $ Start.Time   : Factor w/ 81223 levels "","2017-01-01 00:11:00",..: 74753 19510 59964 26708 67
 $ End.Time     : Factor w/ 81217 levels "","2017-01-01 00:14:00",..: 74744 19473 59981 26732 67
 $ Trip.Duration: num  489 403 637 1827 1549 ...
 $ Start.Station: Factor w/ 478 levels "","10th & E St NW",..: 27 478 66 221 278 84 368 82 71 60
 $ End.Station  : Factor w/ 479 levels "","10th & E St NW",..: 47 219 144 312 315 239 162 376 51
 $ User.Type    : Factor w/ 3 levels "","Customer",..: 3 3 3 2 3 3 3 3 3 3 ...
```

```
       X                        Start.Time                   End.Time
 Min.   :      7   2017-02-19 12:19:00:    6   2017-03-09 17:54:00:    7
 1st Qu.: 434587   2017-02-20 11:35:00:    6   2017-03-28 18:11:00:    7
 Median : 872858   2017-02-24 17:46:00:    6   2017-01-13 17:48:00:    6
 Mean   : 873881   2017-03-01 08:20:00:    6   2017-01-31 08:49:00:    6
 3rd Qu.:1313305   2017-03-02 08:39:00:    6   2017-02-13 18:09:00:    6
 Max.   :1751392   2017-03-09 17:31:00:    6   2017-02-20 11:38:00:    6
                   (Other)            :89015   (Other)            :89013
 Trip.Duration                                  Start.Station
 Min.   :    60.3   Columbus Circle / Union Station     : 1700
 1st Qu.:   410.9   Lincoln Memorial                    : 1546
 Median :   707.0   Jefferson Dr & 14th St SW           : 1488
 Mean   :  1234.0   Massachusetts Ave & Dupont Circle NW: 1219
 3rd Qu.:  1233.2   Jefferson Memorial                  : 1068
 Max.   :904591.4   15th & P St NW                      : 1040
 NA's   :1          (Other)                            :80990
                                                 End.Station
 Columbus Circle / Union Station                       : 1767
 Jefferson Dr & 14th St SW                             : 1603
 Lincoln Memorial                                      : 1514
 Massachusetts Ave & Dupont Circle NW                  : 1344
 Smithsonian-National Mall / Jefferson Dr & 12th St SW: 1103
 15th & P St NW                                        : 1077
```

```
(Other)                                                 :80643
     User.Type
          :     1
Customer  :23450
Subscriber:65600
```

2634.42892149254
733.326013094512

	avg_user	user_type
avg_cust	2634.429	Customer
avg_sub	733.326	Subscriber



The average of the Washington trip duration based on Customer user type is 2635.14 seconds. The average of the Washington trip duration based on Subscriber user type is 735.75 seconds. We could see that there is a significant difference of 1899.39 seconds between the mean trip duration of Customers and Subscribers.

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 is 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** submenu, 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 [1]: system('python -m nbconvert Explore_bikeshare_data.ipynb')
```