Explore_bikeshare_data

April 12, 2022

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

```
In [2]: 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 [3]: 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 [4]: 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	Ca

```
In [5]: # add Gender and Birth.Year to wash in order to bind
        add_columns <- c('Gender', 'Birth.Year')
        wash[ , add_columns] <- NA
```

```
In [6]: # Combine all datasets into one.
        combo <- rbind(ny, wash, chi)

        head(combo)
```

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 &

0.0.2 Question 1

What local time is the busiest hour before noon?

```
In [7]: # checking combo df column types for use
        str(combo)
```

```
'data.frame':        152451 obs. of  9 variables:
 $ X             : int  5688089 4096714 2173887 3945638 6208972 1285652 1675753 1692245 2271331 1
 $ Start.Time    : Factor w/ 143945 levels "2017-01-01 00:17:01",..: 45448 32799 17316 31589 4968
 $ End.Time      : Factor w/ 143948 levels "201","2017-01-01 00:30:56",..: 45432 32783 17295 3156
 $ Trip.Duration: num  795 692 1325 703 329 ...
 $ Start.Station: Factor w/ 1585 levels "","1 Ave & E 16 St",..: 522 406 10 93 5 521 325 309 151
 $ End.Station  : Factor w/ 1586 levels "","1 Ave & E 16 St",..: 613 8 362 558 269 107 389 110 1
 $ 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 ...
```

```
In [8]: # transform Start.Time to usable datetime
        # used from https://www.geeksforgeeks.org/convert-dataframe-column-to-datetime-in-r/
        combo[['Start.Time']] <- as.POSIXct(combo[['Start.Time']],
                                            format = "%Y-%m-%d %H:%M:%S")
```

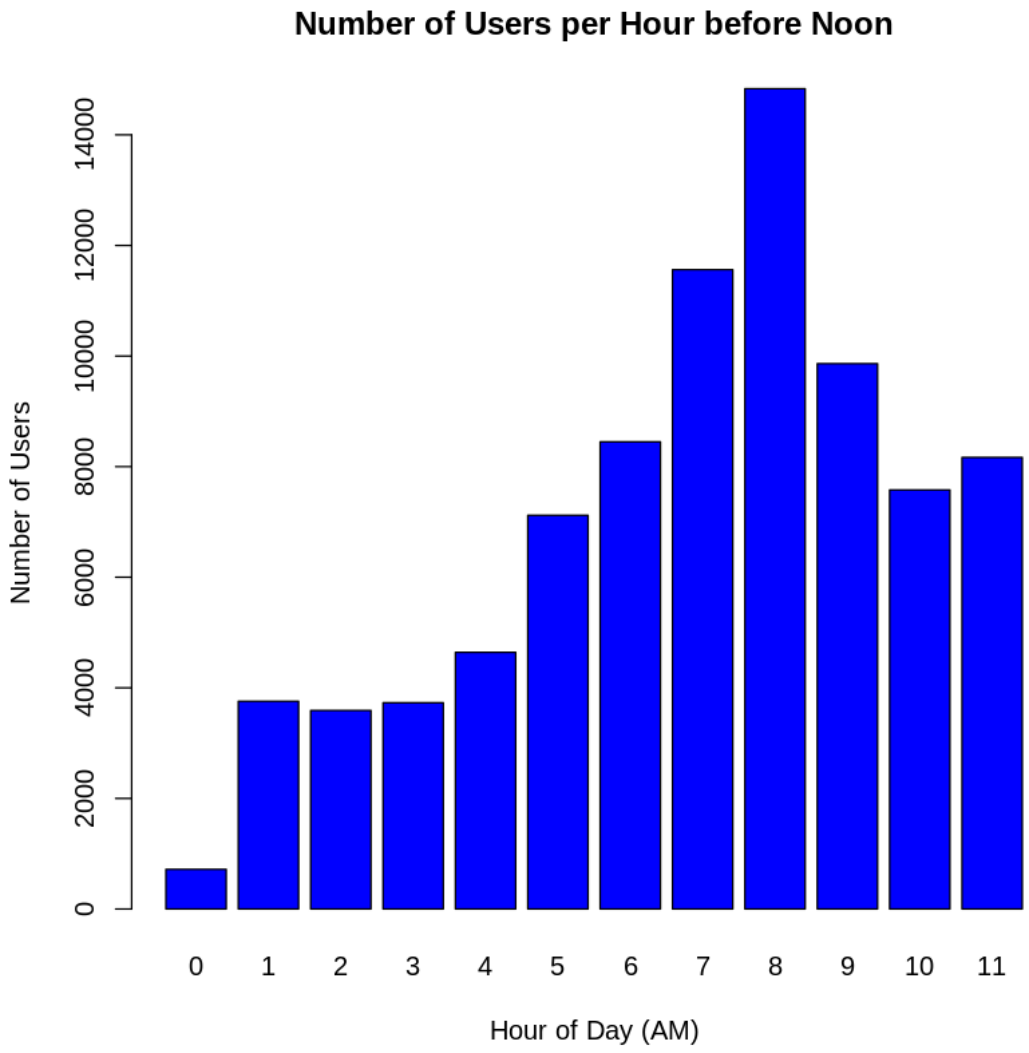
```
In [9]: # Create Hour.Start using extraction form Start.Time extraction
        add.hour.Start <- 'Hour.Start'
        combo[ , add.hour.Start] <- format(combo$Start.Time, format = "%H")
```

```
In [10]: # Change Hour.Start to int/numeric
         combo$Hour.Start <- as.numeric(combo$Hour.Start)
```

```
In [11]: # subset times before noon
         b.noon <- combo[combo$Hour.Start < 12, ]
         table(b.noon$Hour.Start)
```

```
    0     1     2     3     4     5     6     7     8     9    10    11
  718  3757  3591  3731  4642  7120  8449 11562 14835  9863  7579  8167
```

In [12]:
```
# Visual of Users per hour
barplot(table(b.noon$Hour.Start),
main = "Number of Users per Hour before Noon",
xlab = "Hour of Day (AM)",
ylab = "Number of Users",
col = "blue")
```



While there is use of the service at all hours, the most service uses before noon occur starting at 8(am).

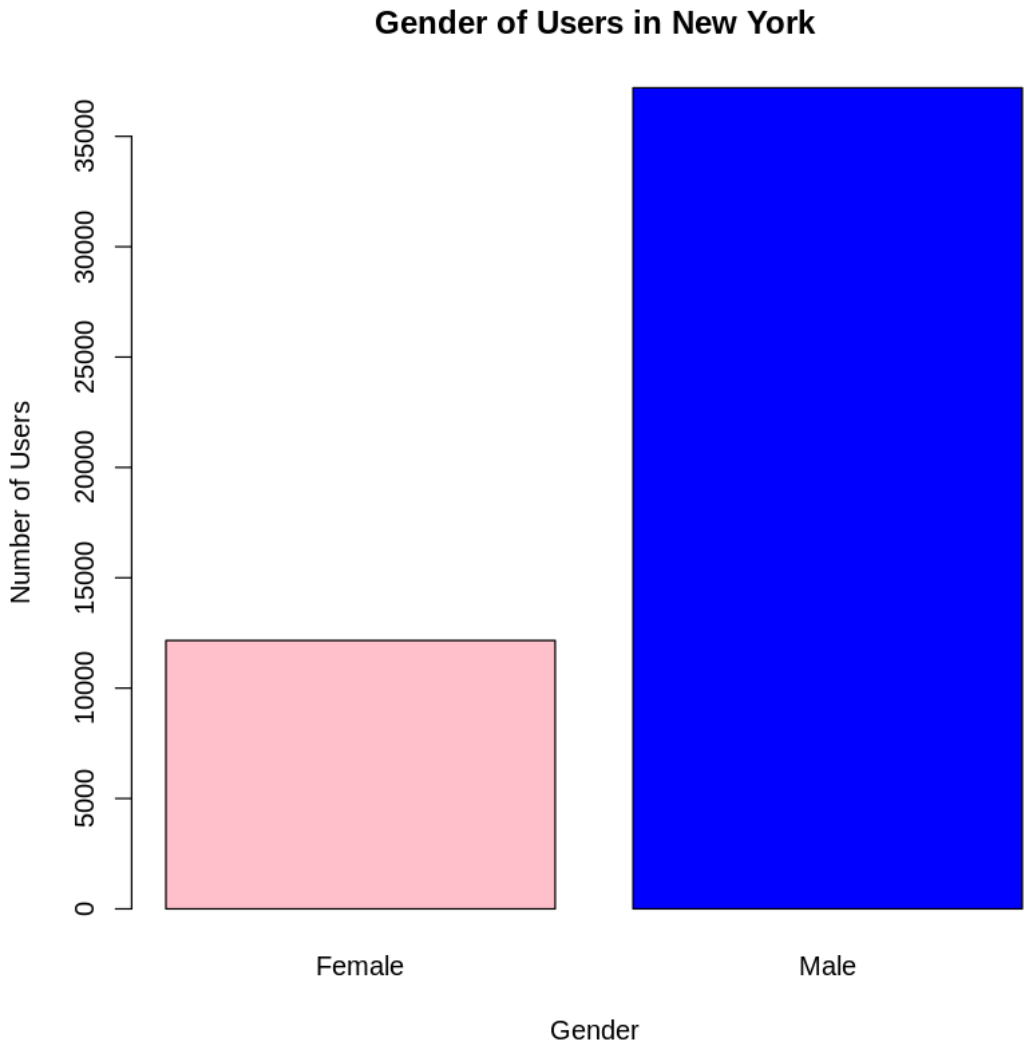
0.0.3 Question 2

In each city, do women or men use the service more?

```
In [13]: # New York
         table(ny$Gender)
```

```
       Female    Male
  5410  12159   37201
```

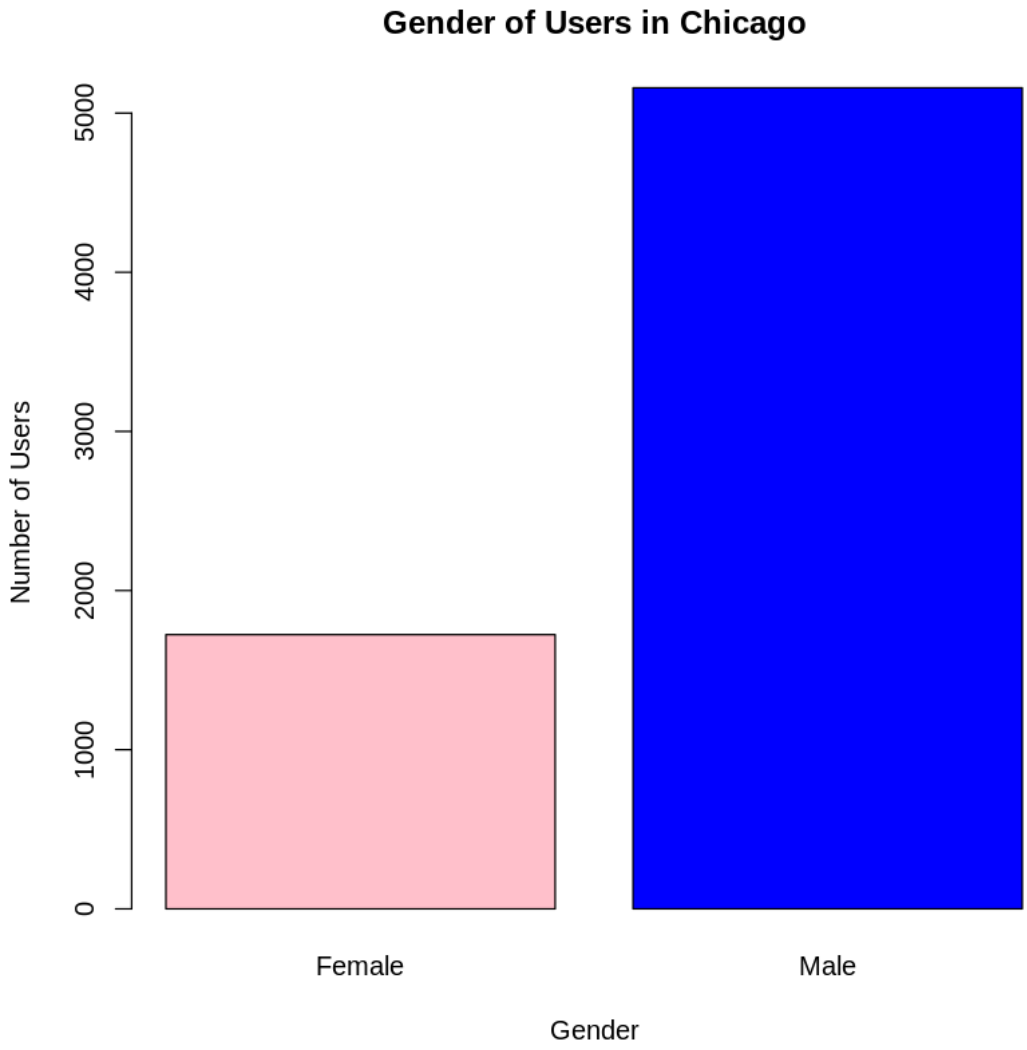
```
In [14]: # Visual of New York Gender Users
         barplot(table(ny$Gender)[2:3],
         main = "Gender of Users in New York",
         xlab = "Gender",
         ylab = "Number of Users",
         col = c("pink", "blue"))
```



```
In [15]: # Chicago
         table(chi$Gender)
```

```
       Female    Male
  1748   1723    5159
```

```
In [16]: # Visual of CHicago Gender Users
         barplot(table(chi$Gender)[2:3],
         main = "Gender of Users in Chicago",
         xlab = "Gender",
         ylab = "Number of Users",
         col = c("pink", "blue"))
```



```
In [17]: # Washington
         table(wash$Gender)
```

```
< table of extent 0 >
```

In New York, Female users count 12,159, while Male users count 37,201. There are more Male users in New York. In Chicago, Female users count 1,723, while Male users count 5,159. There are more Male users than Female users in Chicago. Washington does not have Gender information included.

0.0.4 Question 3

What are the quartile ages of the age-documented users of the service?

```
In [18]: # create new column of dataframe using current year (2022) minus birth.Year
         add.age <- 'Age'
         combo[ , add.age] <- 2022 - combo['Birth.Year']
         head(combo)
```

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 [19]: # summary used for single column
         summary(combo$Age)[1:6]
```

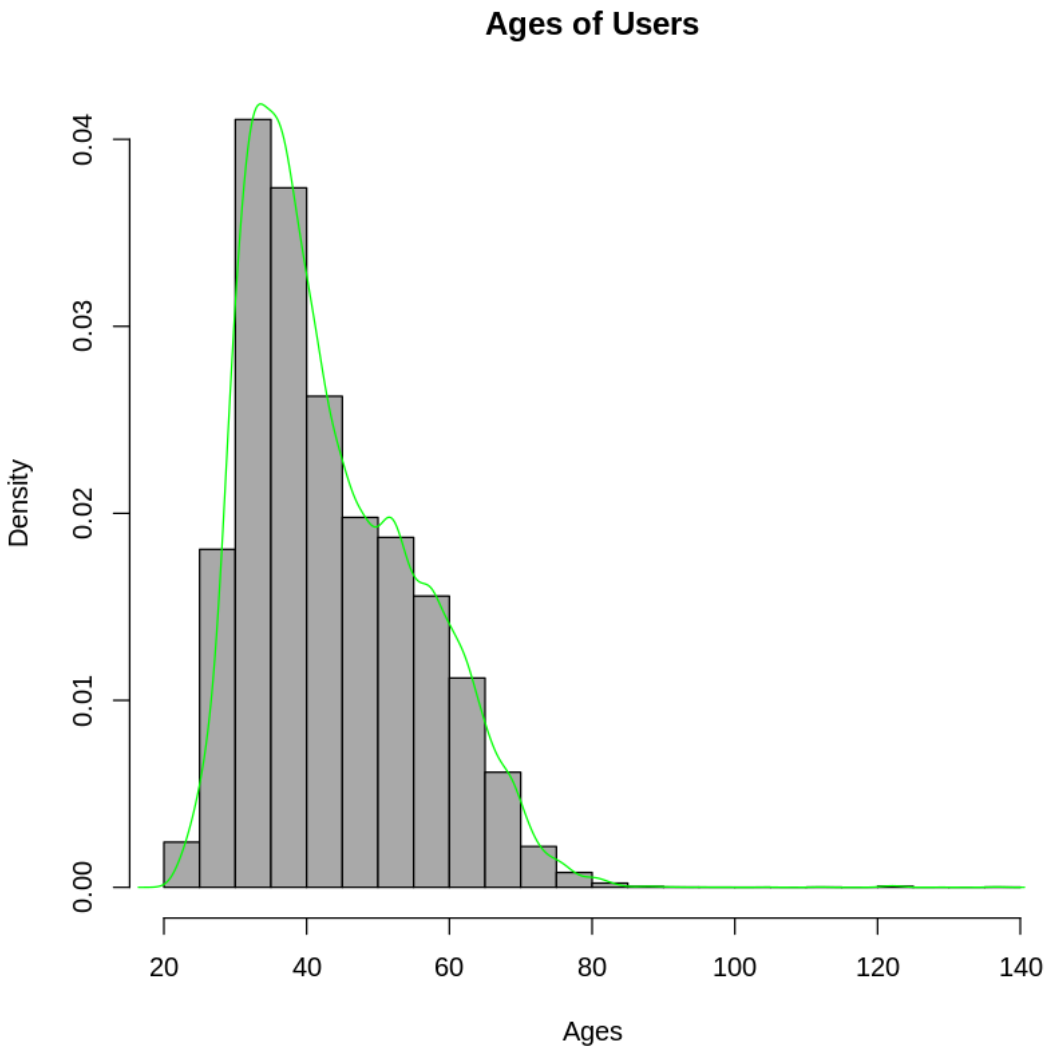
```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
 20.00000  34.00000  41.00000  43.45934  52.00000 137.00000
```

```
In [20]: summary(combo$Age)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  20.00   34.00   41.00   43.46   52.00  137.00   96016
```

```
In [21]: # Visual of Age
         hist(combo$Age,prob=T,
             main = "Ages of Users",
         xlab = "Ages",
         col = "darkgrey")

         points(density(combo$Age, na.rm=TRUE),type="l",col="green")
```



The minimum age of users is 20. The 25th percentile of users is 34. The median age of users is 41. The 75th percentile of users is 52. The oldest user is 137.

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 [23]:
```
system('python -m nbconvert Explore_bikeshare_data.ipynb')
```