Explore_bikeshare_data

September 21, 2023

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]: library(ggplot2)
```

```
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	Ca

0.0.2 Question 1

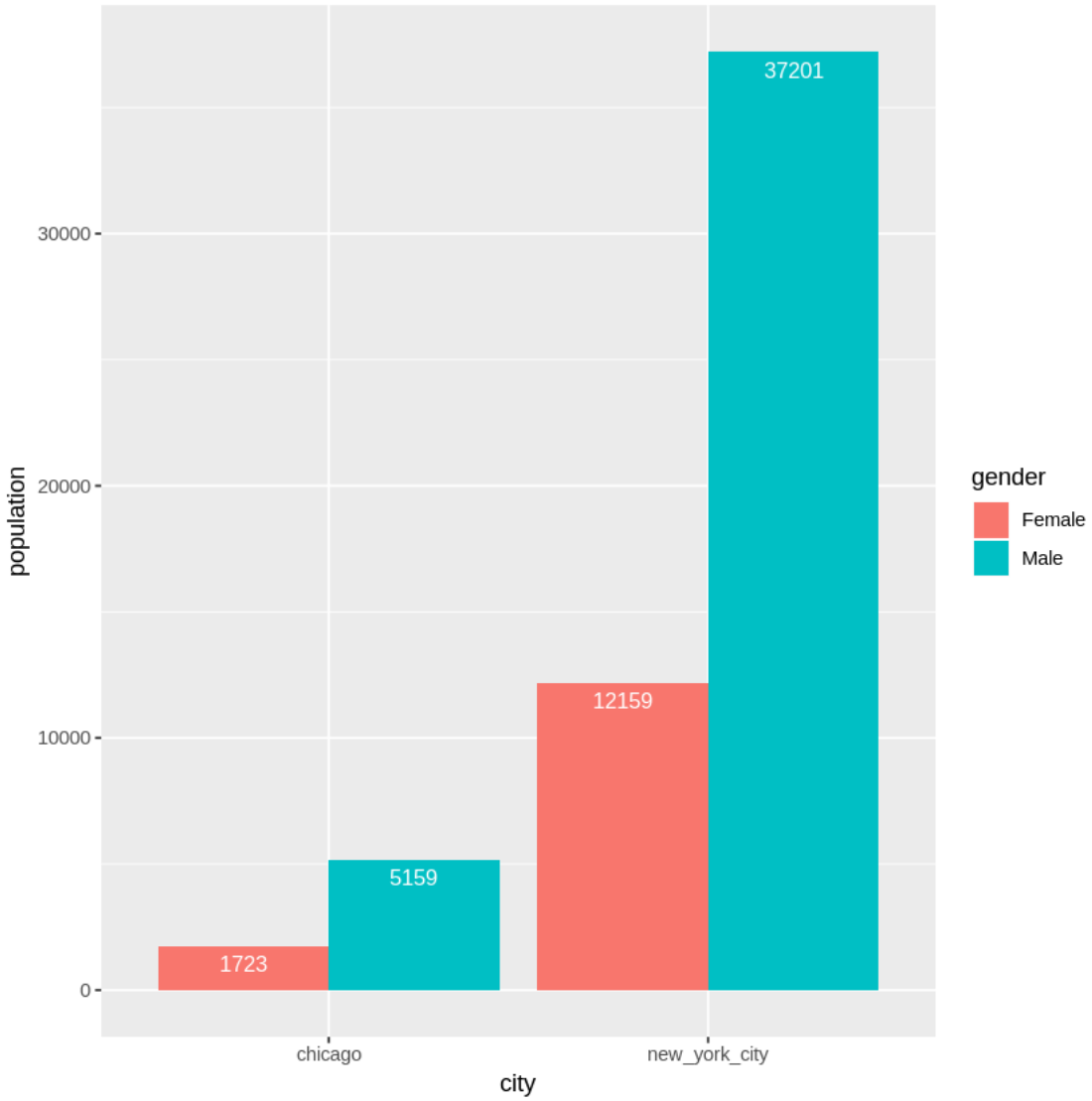
Your question 1 goes here.

0.0.3 What is the gender distribution of bikeshare users in New York City and Chicago.

In [6]:
```
# Your solution code goes here
```

In [7]:
```
gender_data <- data.frame(city=rep(c("new_york_city", "chicago"),each=2),
                          gender=rep(c("Male", "Female"),2),
                          population=c(sum(ny["Gender"] == "Male"),
                                       sum(ny["Gender"] == "Female"),
                                       sum(chi["Gender"] == "Male"),
                                       sum(chi["Gender"] == "Female")))
```

In [8]:
```
ggplot(gender_data, aes(x=city, y=population, fill=gender))+
geom_bar(stat = "identity", position=position_dodge())+
geom_text(aes(label=population), vjust=1.6, color="white",
            position = position_dodge(0.9), size=3.5)
```



Summary of your question 1 results goes here.

0.0.4 In both cities, we see more male bikeshare users than female bikeshare users. The female/male ratio is 33.40% in Chicage and 13.87% in New York City.

0.0.5 Question 2

Your question 2 goes here.

0.0.6 What is the trip duration among bikeshare subscriber user types in Washington?

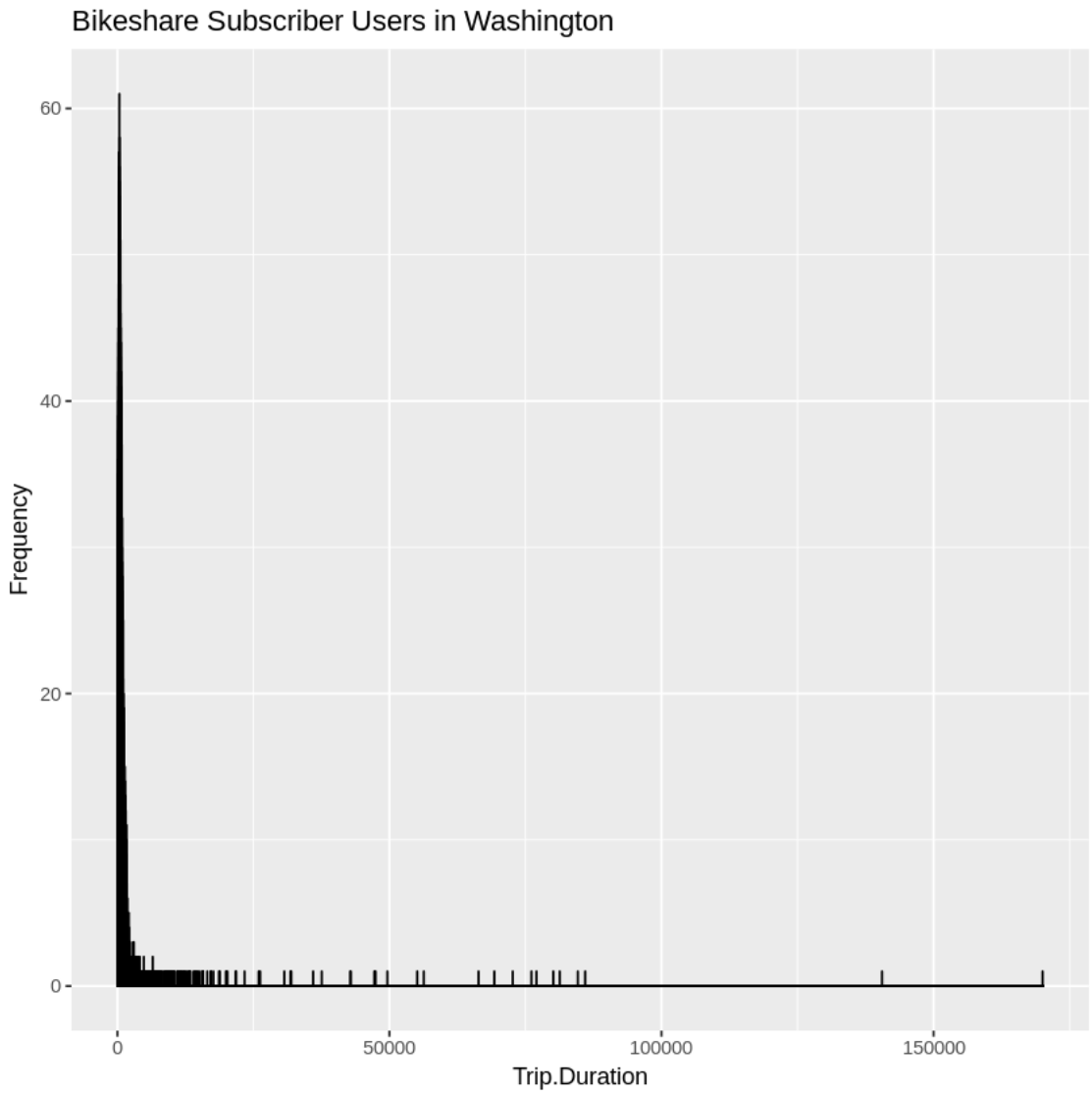
```
In [9]: # Your solution code goes here
```

```
In [10]: unique(wash["User.Type"])
```

	User.Type
1	Subscriber
4	Customer
89051	

In [11]:
```
data <- wash[wash["User.Type"]=="Subscriber",]["Trip.Duration"]

ggplot(data, aes(x=Trip.Duration)) +
  geom_histogram(binwidth=0.5, fill="skyblue", color="black", alpha=0.7) +
  labs(title="Bikeshare Subscriber Users in Washington", x="Trip.Duration", y="Frequenc
```



Summary of your question 2 results goes here.

0.0.7 In the figure we see the trip duration distribution of Bikeshare subscriber users. And we see that most subscribers have low trip duration, but we do see individuals having very high trip duration.

0.0.8 Question 3

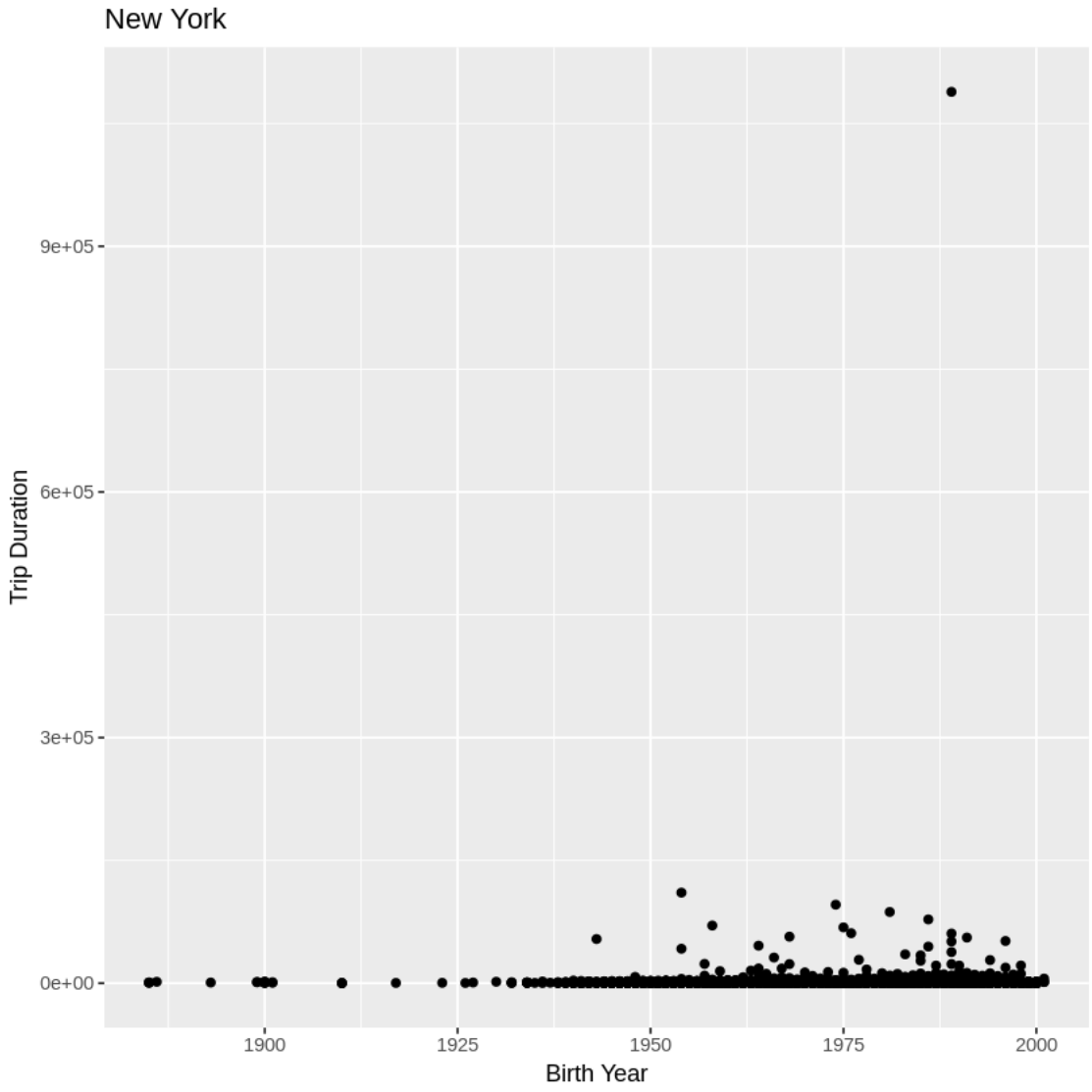
Your question 3 goes here.

0.0.9 What is the relation between trip duration and birth year in NYC and Chicago?

```
In [12]: # Your solution code goes here
```

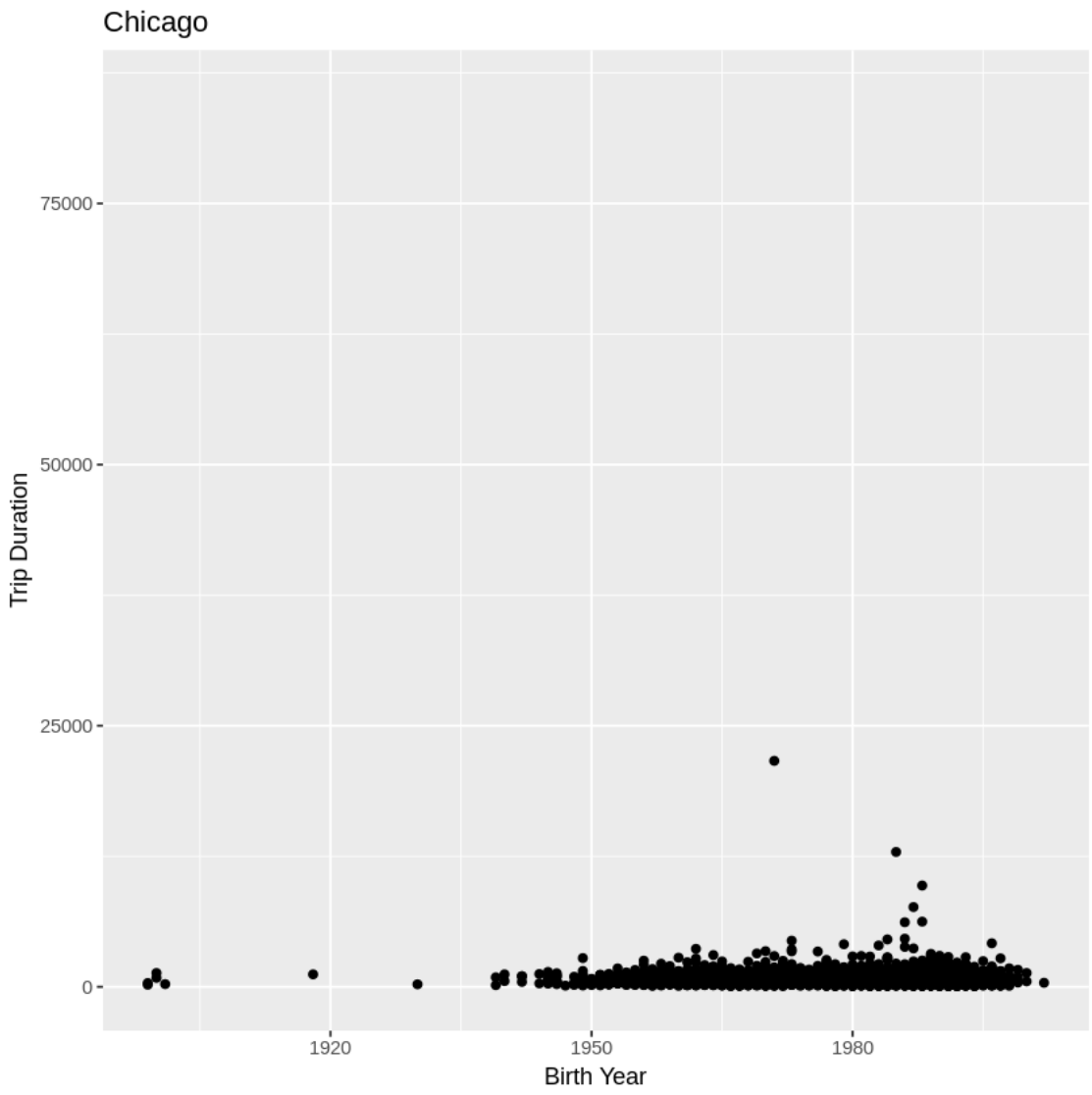
```
In [13]: ggplot(ny, aes(x=Birth.Year, y=Trip.Duration)) +
           geom_point() + labs(title="New York", x="Birth Year", y="Trip Duration")
```

```
Warning message:
Removed 5218 rows containing missing values (geom_point).
```



```
In [14]: ggplot(chi, aes(x=Birth.Year, y=Trip.Duration)) +
           geom_point() + labs(title="Chicago", x="Birth Year", y="Trip Duration")
```

```
Warning message:
Removed 1747 rows containing missing values (geom_point).
```



Summary of your question 3 results goes here.

0.0.10 In the above two figures, we see the distribution of bikeshare users of their birth year and trip duration from both New York City and Chicago. As we can see, from both cities we have users with extremely old birth year, most likely due to error when registration. Users born in the 80s and 90s in both cities has the highest average trip duration, which makes sense as their age is around 20-30.

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 [16]:
```
system('python -m nbconvert Explore_bikeshare_data.ipynb')
```