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 (https://review.udacity.com/#!/rubrics/2508/view) 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!

Table of Contents

In [153]:
```
library(ggplot2)
library(dplyr)
library(lubridate)
```

In [154]:
```
ny = read.csv('new_york_city.csv')
wash = read.csv('washington.csv')
chi = read.csv('chicago.csv')
```

Visual assessment: Look at the 'head' of each dataframe to get a sense of the data it holds.

In [155]:
```
head(ny)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type	Ge
5688089	2017-06-11 14:55:05	2017-06-11 15:08:21	795	Suffolk St & Stanton St	W Broadway & Spring St	Subscriber	Ma
4096714	2017-05-11 15:30:11	2017-05-11 15:41:43	692	Lexington Ave & E 63 St	1 Ave & E 78 St	Subscriber	Ma
2173887	2017-03-29 13:26:26	2017-03-29 13:48:31	1325	1 Pl & Clinton St	Henry St & Degraw St	Subscriber	Ma
3945638	2017-05-08 19:47:18	2017-05-08 19:59:01	703	Barrow St & Hudson St	W 20 St & 8 Ave	Subscriber	Fe
6208972	2017-06-21 07:49:16	2017-06-21 07:54:46	329	1 Ave & E 44 St	E 53 St & 3 Ave	Subscriber	Ma
1285652	2017-02-22 18:55:24	2017-02-22 19:12:03	998	State St & Smith St	Bond St & Fulton St	Subscriber	Ma

In [156]:
```
head(wash)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type
1621326	2017-06-21 08:36:34	2017-06-21 08:44:43	489.066	14th & Belmont St NW	15th & K St NW	Subscriber
482740	2017-03-11 10:40:00	2017-03-11 10:46:00	402.549	Yuma St & Tenley Circle NW	Connecticut Ave & Yuma St NW	Subscriber
1330037	2017-05-30 01:02:59	2017-05-30 01:13:37	637.251	17th St & Massachusetts Ave NW	5th & K St NW	Subscriber
665458	2017-04-02 07:48:35	2017-04-02 08:19:03	1827.341	Constitution Ave & 2nd St NW/DOL	M St & Pennsylvania Ave NW	Customer
1481135	2017-06-10 08:36:28	2017-06-10 09:02:17	1549.427	Henry Bacon Dr & Lincoln Memorial Circle NW	Maine Ave & 7th St SW	Subscriber
1148202	2017-05-14 07:18:18	2017-05-14 07:24:56	398.000	1st & K St SE	Eastern Market Metro / Pennsylvania Ave & 7th St SE	Subscriber

In [157]:
```
head(chi)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type	Ge
1423854	2017-06-23 15:09:32	2017-06-23 15:14:53	321	Wood St & Hubbard St	Damen Ave & Chicago Ave	Subscriber	Ma
955915	2017-05-25 18:19:03	2017-05-25 18:45:53	1610	Theater on the Lake	Sheffield Ave & Waveland Ave	Subscriber	Fe
9031	2017-01-04 08:27:49	2017-01-04 08:34:45	416	May St & Taylor St	Wood St & Taylor St	Subscriber	Ma
304487	2017-03-06 13:49:38	2017-03-06 13:55:28	350	Christiana Ave & Lawrence Ave	St. Louis Ave & Balmoral Ave	Subscriber	Ma
45207	2017-01-17 14:53:07	2017-01-17 15:02:01	534	Clark St & Randolph St	Desplaines St & Jackson Blvd	Subscriber	Ma
1473887	2017-06-26 09:01:20	2017-06-26 09:11:06	586	Clinton St & Washington Blvd	Canal St & Taylor St	Subscriber	Ma

Programmatic assessment

- Check the dimensions of each dataframe
- Check the data types each dataframe holds
- Check for duplicates in each dataframe, remove if necessary.
- Check for null values in each dataframe, remove if necessary.

Check the dimensions of each dataframe

In [158]:
```
dim(ny)
```

```
54770  9
```

In [159]:
```
dim(wash)
```

```
89051  7
```

In [160]:
```
dim(chi)
```

```
8630  9
```

Check data types of each dataframe

In [161]:
```
str(ny)
```

```
'data.frame':   54770 obs. of  9 variables:
 $ X             : int  5688089 4096714 2173887 3945638 6208972 12856
52 1675753 1692245 2271331 1558339 ...
 $ Start.Time    : Factor w/ 54568 levels "2017-01-01 00:17:01",..: 4
5448 32799 17316 31589 49688 10220 13390 13509 18111 12449 ...
 $ End.Time      : Factor w/ 54562 levels "201","2017-01-01 00:30:56"
,..: 45432 32783 17295 31567 49668 10204 13364 13505 18092 12422 ...
 $ 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 5 521 325 309 151 245 ...
 $ End.Station  : Factor w/ 638 levels "","1 Ave & E 16 St",..: 613
8 362 558 269 107 389 110 151 243 ...
 $ 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 [162]:
```
str(chi)
```

```
'data.frame':   8630 obs. of  9 variables:
 $ X             : int  1423854 955915 9031 304487 45207 1473887 9619
16 65924 606841 135470 ...
 $ Start.Time    : Factor w/ 8624 levels "2017-01-01 00:40:14",..: 78
76 5303 73 1721 267 8173 5347 368 3376 795 ...
 $ End.Time      : Factor w/ 8625 levels "2017-01-01 00:46:32",..: 78
76 5303 73 1722 267 8173 5346 368 3376 796 ...
 $ 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 22 255 374 420 ...
 $ End.Station  : Factor w/ 471 levels "","2112 W Peterson Ave",..:
132 381 469 409 151 70 467 251 200 118 ...
 $ 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 ...
```

In [163]:
```
str(wash)
```

```
'data.frame':   89051 obs. of  7 variables:
 $ X             : int  1621326 482740 1330037 665458 1481135 1148202
1594275 1601832 574182 327058 ...
 $ Start.Time    : Factor w/ 81223 levels "","2017-01-01 00:11:00",..
: 74753 19510 59964 26708 67716 50891 73381 73775 23142 13333 ...
 $ End.Time      : Factor w/ 81217 levels "","2017-01-01 00:14:00",..
: 74744 19473 59981 26732 67753 50918 73397 73775 23114 13350 ...
 $ Trip.Duration: num  489 403 637 1827 1549 ...
 $ Start.Station: Factor w/ 478 levels "","10th & E St NW",..: 27 47
8 66 221 278 84 368 82 71 60 ...
 $ End.Station  : Factor w/ 479 levels "","10th & E St NW",..: 47 21
9 144 312 315 239 162 376 51 308 ...
 $ User.Type    : Factor w/ 3 levels "","Customer",..: 3 3 3 2 3 3 3
3 3 3 ...
```

Check for duplicates in each dataframe, remove if necessary

In [164]:
```
# Check for duplicates in 'ny'
duplicates_ny = sum(duplicated(ny))
print(duplicates_ny)
```

```
[1] 0
```

In [165]:
```
# Check for duplicates in 'wash'
duplicates_wash = sum(duplicated(wash))
print(duplicates_wash)
```

```
[1] 0
```

In [166]:
```
# Check for duplicates in 'chi'
duplicates_chi = sum(duplicated(chi))
print(duplicates_chi)
```

```
[1] 0
```

There are no duplicates so lets continue.

Check for null values in each dataframe, remove if necessary.

In [167]:
```
# Check for null values in 'ny'
nulls_ny = colSums(is.na(ny))
print(nulls_ny)
```

```
            X    Start.Time      End.Time Trip.Duration Start.Statio
n
            0             0             0             1
0
  End.Station     User.Type        Gender    Birth.Year
            0             0             0          5218
```

In [168]:
```
# Remove null values from 'ny'
ny = na.omit(ny)
dim(ny)
```

```
49552  9
```

In [169]:
```
# Check for null values in 'chi'
nulls_chi = colSums(is.na(chi))
print(nulls_chi)
```

```
            X    Start.Time      End.Time Trip.Duration Start.Statio
n
            0             0             0             0
0
  End.Station     User.Type        Gender    Birth.Year
            0             0             0          1747
```

In [170]:
```
# Remove null values in 'chi'
chi = na.omit(chi)
dim(chi)
```

```
6883  9
```

In [171]:
```
# Check for null values in 'wash'
nulls_wash = colSums(is.na(wash))
print(nulls_wash)
```

```
            X    Start.Time      End.Time Trip.Duration Start.Statio
n
            0             0             0             1
0
  End.Station     User.Type
            0             0
```

In [172]:
```
# Remove null values from 'wash'
wash = na.omit(wash)
dim(wash)
```

```
89050  7
```

Further Wrangling

- Add columns 'Gender' and Birh.Year' to dataframe 'wash'
- Add a 'City' column to each dataframe
- Combine the three dataframes
- Convert 'Start.Time' and 'End.Time' to 'datetime' (POSIXct) format
- Change 'Trip.Duration' from seconds to minutes
- Add an 'Age' column

Add 'Gender' and 'Birth.Year' to columns to 'wash' dataframe, fill with 'NA' values.

In [173]:
```
# Add 'Gender' and 'Birth.Year' columns to 'wash' dataframe, fill with
'NA' values.
wash$Gender = NA
wash$Birth.Year = NA
```

In [174]:
```
# Check for success
head(wash)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type
1621326	2017-06-21 08:36:34	2017-06-21 08:44:43	489.066	14th & Belmont St NW	15th & K St NW	Subscriber
482740	2017-03-11 10:40:00	2017-03-11 10:46:00	402.549	Yuma St & Tenley Circle NW	Connecticut Ave & Yuma St NW	Subscriber
1330037	2017-05-30 01:02:59	2017-05-30 01:13:37	637.251	17th St & Massachusetts Ave NW	5th & K St NW	Subscriber
665458	2017-04-02 07:48:35	2017-04-02 08:19:03	1827.341	Constitution Ave & 2nd St NW/DOL	M St & Pennsylvania Ave NW	Customer
1481135	2017-06-10 08:36:28	2017-06-10 09:02:17	1549.427	Henry Bacon Dr & Lincoln Memorial Circle NW	Maine Ave & 7th St SW	Subscriber
1148202	2017-05-14 07:18:18	2017-05-14 07:24:56	398.000	1st & K St SE	Eastern Market Metro / Pennsylvania Ave & 7th St SE	Subscriber

Add a 'City' column to each dataframe.

In [175]:
```
# Add 'City' column to 'ny' dataframe
ny$City = 'New York'
```

In [176]:
```
# Check to see if successful
head(ny)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type	Ge
5688089	2017-06-11 14:55:05	2017-06-11 15:08:21	795	Suffolk St & Stanton St	W Broadway & Spring St	Subscriber	Ma
4096714	2017-05-11 15:30:11	2017-05-11 15:41:43	692	Lexington Ave & E 63 St	1 Ave & E 78 St	Subscriber	Ma
2173887	2017-03-29 13:26:26	2017-03-29 13:48:31	1325	1 Pl & Clinton St	Henry St & Degraw St	Subscriber	Ma
3945638	2017-05-08 19:47:18	2017-05-08 19:59:01	703	Barrow St & Hudson St	W 20 St & 8 Ave	Subscriber	Fe
6208972	2017-06-21 07:49:16	2017-06-21 07:54:46	329	1 Ave & E 44 St	E 53 St & 3 Ave	Subscriber	Ma
1285652	2017-02-22 18:55:24	2017-02-22 19:12:03	998	State St & Smith St	Bond St & Fulton St	Subscriber	Ma

In [177]:
```
# Add 'City' column to 'chi' dataframe
chi$City = 'Chicago'
```

In [178]:
```
# Check to see if successful
head(chi)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type	Ge
1423854	2017-06-23 15:09:32	2017-06-23 15:14:53	321	Wood St & Hubbard St	Damen Ave & Chicago Ave	Subscriber	Ma
955915	2017-05-25 18:19:03	2017-05-25 18:45:53	1610	Theater on the Lake	Sheffield Ave & Waveland Ave	Subscriber	Fe
9031	2017-01-04 08:27:49	2017-01-04 08:34:45	416	May St & Taylor St	Wood St & Taylor St	Subscriber	Ma
304487	2017-03-06 13:49:38	2017-03-06 13:55:28	350	Christiana Ave & Lawrence Ave	St. Louis Ave & Balmoral Ave	Subscriber	Ma
45207	2017-01-17 14:53:07	2017-01-17 15:02:01	534	Clark St & Randolph St	Desplaines St & Jackson Blvd	Subscriber	Ma
1473887	2017-06-26 09:01:20	2017-06-26 09:11:06	586	Clinton St & Washington Blvd	Canal St & Taylor St	Subscriber	Ma

In [179]:
```
# Add 'City' column to 'wash' dataframe
wash$City = 'Washington, D.C.'
```

In [180]:
```
# Check to see if successful
head(wash)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type
1621326	2017-06-21 08:36:34	2017-06-21 08:44:43	489.066	14th & Belmont St NW	15th & K St NW	Subscriber
482740	2017-03-11 10:40:00	2017-03-11 10:46:00	402.549	Yuma St & Tenley Circle NW	Connecticut Ave & Yuma St NW	Subscriber
1330037	2017-05-30 01:02:59	2017-05-30 01:13:37	637.251	17th St & Massachusetts Ave NW	5th & K St NW	Subscriber
665458	2017-04-02 07:48:35	2017-04-02 08:19:03	1827.341	Constitution Ave & 2nd St NW/DOL	M St & Pennsylvania Ave NW	Customer
1481135	2017-06-10 08:36:28	2017-06-10 09:02:17	1549.427	Henry Bacon Dr & Lincoln Memorial Circle NW	Maine Ave & 7th St SW	Subscriber
1148202	2017-05-14 07:18:18	2017-05-14 07:24:56	398.000	1st & K St SE	Eastern Market Metro / Pennsylvania Ave & 7th St SE	Subscriber

Combine the three dataframes

In [181]:
```
# Combine the three dataframes, check to see if successful
merged = rbind(ny, chi, wash)
head(merged)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type	Ge
5688089	2017-06-11 14:55:05	2017-06-11 15:08:21	795	Suffolk St & Stanton St	W Broadway & Spring St	Subscriber	Ma
4096714	2017-05-11 15:30:11	2017-05-11 15:41:43	692	Lexington Ave & E 63 St	1 Ave & E 78 St	Subscriber	Ma
2173887	2017-03-29 13:26:26	2017-03-29 13:48:31	1325	1 Pl & Clinton St	Henry St & Degraw St	Subscriber	Ma
3945638	2017-05-08 19:47:18	2017-05-08 19:59:01	703	Barrow St & Hudson St	W 20 St & 8 Ave	Subscriber	Fe
6208972	2017-06-21 07:49:16	2017-06-21 07:54:46	329	1 Ave & E 44 St	E 53 St & 3 Ave	Subscriber	Ma
1285652	2017-02-22 18:55:24	2017-02-22 19:12:03	998	State St & Smith St	Bond St & Fulton St	Subscriber	Ma

In [182]:
```
tail(merged)
```

	X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Ty
89045	1371409	2017-06-02 06:02:14	2017-06-02 06:14:27	732.245	Columbus Circle / Union Station	L'Enfant Plaza / 7th & C St SW	Custom
89046	1484340	2017-06-10 10:58:09	2017-06-10 11:25:58	1669.700	M St & New Jersey Ave SE	4th St & Madison Dr NW	Custom
89047	555788	2017-03-22 18:46:00	2017-03-22 19:04:00	1082.789	8th & H St NW	21st & I St NW	Subscr
89048	739004	2017-04-09 04:00:22	2017-04-09 04:09:54	571.879	Eckington Pl & Q St NE	Columbus Circle / Union Station	Subscr
89049	1214907	2017-05-19 09:00:53	2017-05-19 09:07:38	404.152	1st & M St NE	1st & Rhode Island Ave NW	Subscr
89050	1419806	2017-06-06 04:27:33	2017-06-06 04:49:59	1345.911	10th & Florida Ave NW	Georgetown Harbor / 30th St NW	Custom

In [183]:
```
dim(merged)
```

145485 10

Convert 'Start.Time' and 'End.Time' to 'datetime' (POSIXct) format

In [184]:
```
# Convert 'Start.Time' and 'End.Time' to new data format
merged$Start.Time <- as.POSIXct(merged$Start.Time, format = "%Y-%m-%d %H:%M:%S")
merged$End.Time <- as.POSIXct(merged$End.Time, format = "%Y-%m-%d %H:%M:%S")
```

In [185]:
```
# Check for success
str(merged)
```

```
'data.frame':   145485 obs. of  10 variables:
 $ X            : int  5688089 4096714 2173887 3945638 6208972 12856
52 1675753 1692245 1558339 2744874 ...
 $ Start.Time   : POSIXct, format: "2017-06-11 14:55:05" "2017-05-11
15:30:11" ...
 $ End.Time     : POSIXct, format: "2017-06-11 15:08:21" "2017-05-11
15:41:43" ...
 $ Trip.Duration: num  795 692 1325 703 329 ...
 $ Start.Station: Factor w/ 1585 levels "","1 Ave & E 16 St",..: 522
406 10 93 5 521 325 309 245 326 ...
 $ End.Station  : Factor w/ 1586 levels "","1 Ave & E 16 St",..: 613
8 362 558 269 107 389 110 243 407 ...
 $ User.Type    : Factor w/ 3 levels "","Customer",..: 3 3 3 3 3 3 3
3 3 3 ...
 $ Gender       : Factor w/ 3 levels "","Female","Male": 3 3 3 2 3 3
3 3 3 3 ...
 $ Birth.Year   : num  1998 1981 1987 1986 1992 ...
 $ City         : chr  "New York" "New York" "New York" "New York" .
..
 - attr(*, "na.action")= 'omit' Named int  9 11 20 24 33 37 39 53 61
66 ...
  ..- attr(*, "names")= chr  "9" "11" "20" "24" ...
```

Change 'Trip.Duration' from seconds to minutes.

In [186]:
```
# Perform the conversion by dividing the values in 'Trip.Duration' by
'60'
merged$Trip.Duration <- merged$Trip.Duration / 60
```

In [187]:
```
# Check to see if successful
head(merged)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type	Ge
5688089	2017-06-11 14:55:05	2017-06-11 15:08:21	13.250000	Suffolk St & Stanton St	W Broadway & Spring St	Subscriber	Ma
4096714	2017-05-11 15:30:11	2017-05-11 15:41:43	11.533333	Lexington Ave & E 63 St	1 Ave & E 78 St	Subscriber	Ma
2173887	2017-03-29 13:26:26	2017-03-29 13:48:31	22.083333	1 Pl & Clinton St	Henry St & Degraw St	Subscriber	Ma
3945638	2017-05-08 19:47:18	2017-05-08 19:59:01	11.716667	Barrow St & Hudson St	W 20 St & 8 Ave	Subscriber	Fe
6208972	2017-06-21 07:49:16	2017-06-21 07:54:46	5.483333	1 Ave & E 44 St	E 53 St & 3 Ave	Subscriber	Ma
1285652	2017-02-22 18:55:24	2017-02-22 19:12:03	16.633333	State St & Smith St	Bond St & Fulton St	Subscriber	Ma

Add an 'Age' column to the dataframe.

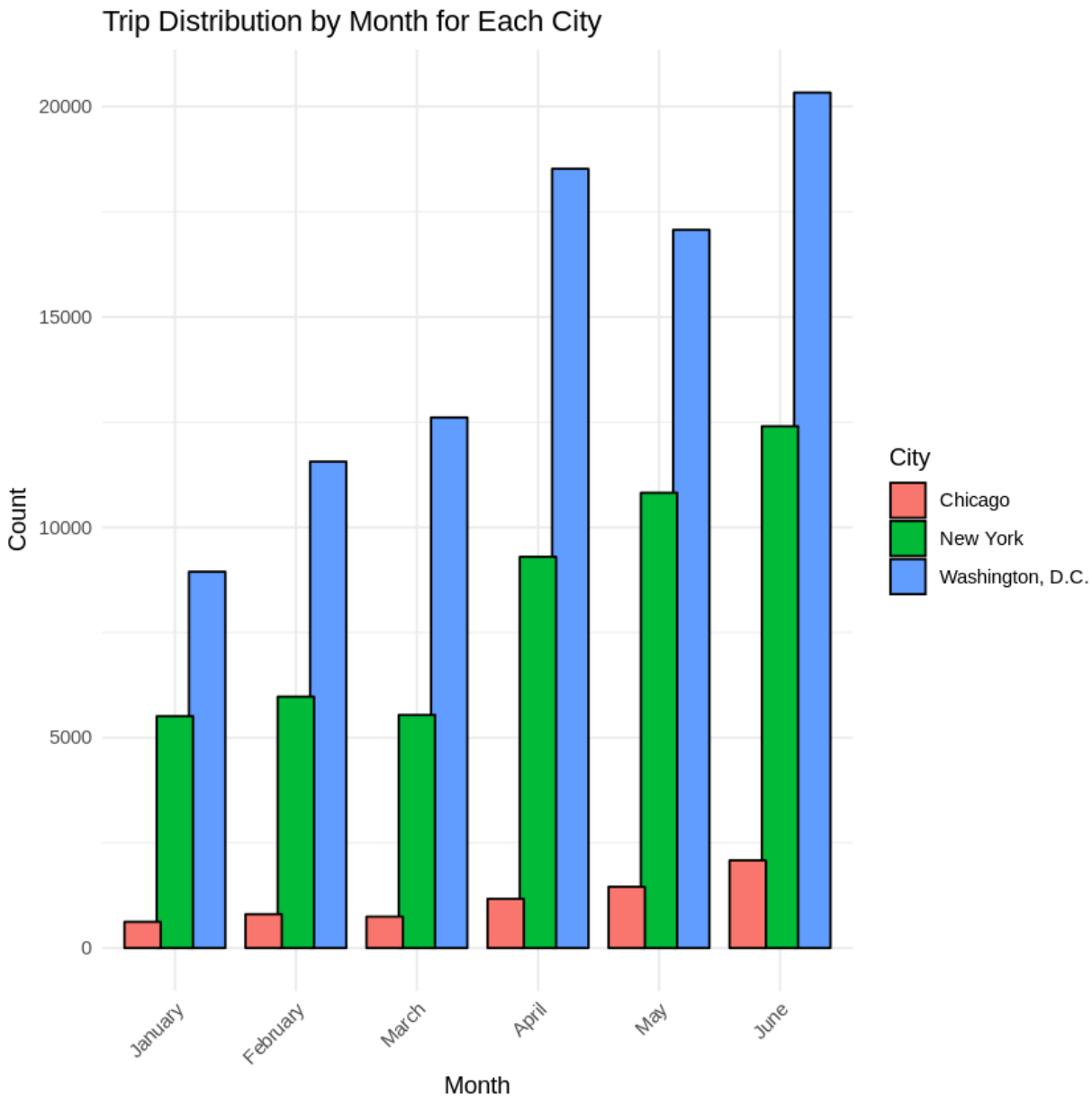
In [188]:
```
# Add column 'Age' to dataframe, which is 2017(when the dataframe is f
rom) - 'Birth.Year
merged$Age = 2017 - merged$Birth.Year
head(merged)
```

X	Start.Time	End.Time	Trip.Duration	Start.Station	End.Station	User.Type	Ge
5688089	2017-06-11 14:55:05	2017-06-11 15:08:21	13.250000	Suffolk St & Stanton St	W Broadway & Spring St	Subscriber	Ma
4096714	2017-05-11 15:30:11	2017-05-11 15:41:43	11.533333	Lexington Ave & E 63 St	1 Ave & E 78 St	Subscriber	Ma
2173887	2017-03-29 13:26:26	2017-03-29 13:48:31	22.083333	1 Pl & Clinton St	Henry St & Degraw St	Subscriber	Ma
3945638	2017-05-08 19:47:18	2017-05-08 19:59:01	11.716667	Barrow St & Hudson St	W 20 St & 8 Ave	Subscriber	Fe
6208972	2017-06-21 07:49:16	2017-06-21 07:54:46	5.483333	1 Ave & E 44 St	E 53 St & 3 Ave	Subscriber	Ma
1285652	2017-02-22 18:55:24	2017-02-22 19:12:03	16.633333	State St & Smith St	Bond St & Fulton St	Subscriber	Ma

Question 1

In each city, what month was the bikeshare program most used?

In [192]:
```
# Extract the month from the 'Start.Time' column and set the order
merged$Month <- factor(format(merged$Start.Time, "%B"), levels = month
.name[1:6])

# Create a bar plot with months in order and different colors for each
city
month_distribution_plot <- ggplot(merged, aes(x = Month, fill = City))
+
  geom_bar(position = position_dodge(width = 0.8), color = "black") +
  labs(title = "Trip Distribution by Month for Each City", x = "Month"
, y = "Count") +
  scale_x_discrete(limits = levels(merged$Month)) +  # Set the order o
f the months
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Print the combined plot
print(month_distribution_plot)
```



It appears that June was the most popular month to use the bikeshare service in each city. In fact, the distribution looks practically the same across all six months for each of the three cities we are analyzing.

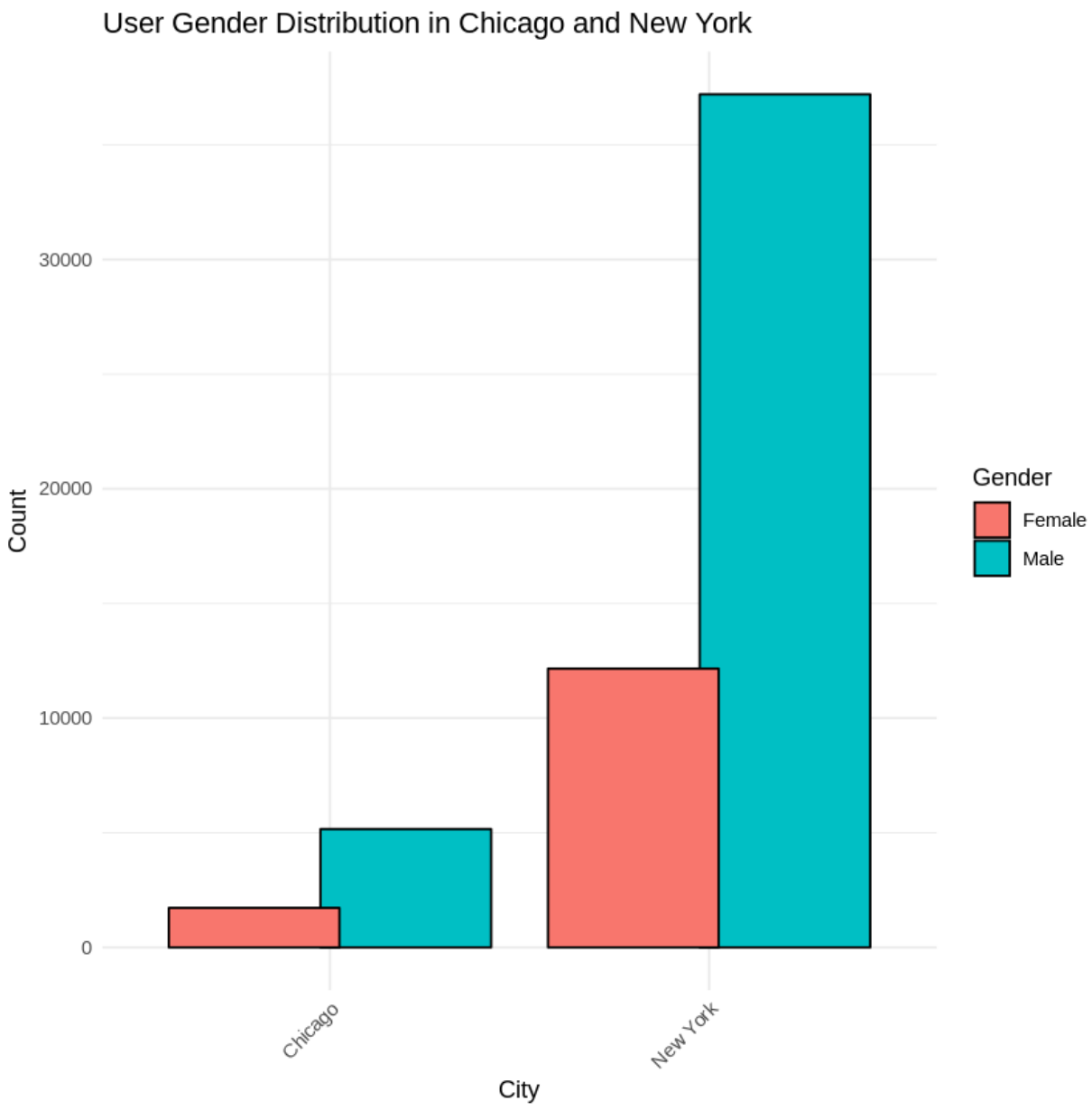
Question 2

What is the distribution of male and female users (only in New York and Chicago, as the data is missing from Washington, D.C.)?

In [193]:
```
# Filter the data for Chicago and New York and include only Male and F
emale
filtered_data <- merged %>%
  filter(City %in% c("Chicago", "New York"), Gender %in% c("Male", "Fe
male"))

# Create a grouped bar plot for the distribution of Male and Female us
ers in Chicago and New York
gender_distribution_plot <- ggplot(filtered_data, aes(x = City, fill =
Gender)) +
  geom_bar(position = position_dodge(width = 0.8), color = "black") +
  labs(title = "User Gender Distribution in Chicago and New York", x =
"City", y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Calculate value counts of Males and Females by City
gender_counts_by_city <- with(filtered_data, table(City, Gender))

# Print the value counts of Males and Females by City
print(gender_counts_by_city)
```

```
          Gender
City             Female  Male
  Chicago      0   1723  5159
  New York     0  12158 37201
```

In [194]:
```
# Print the grouped bar plot
print(gender_distribution_plot)
```



There is a large disparity between male and female users, with both cities having more than double the amount of male users than female. It would be interesting to compare these results with Washington, D.C., if only we had the data to compare.

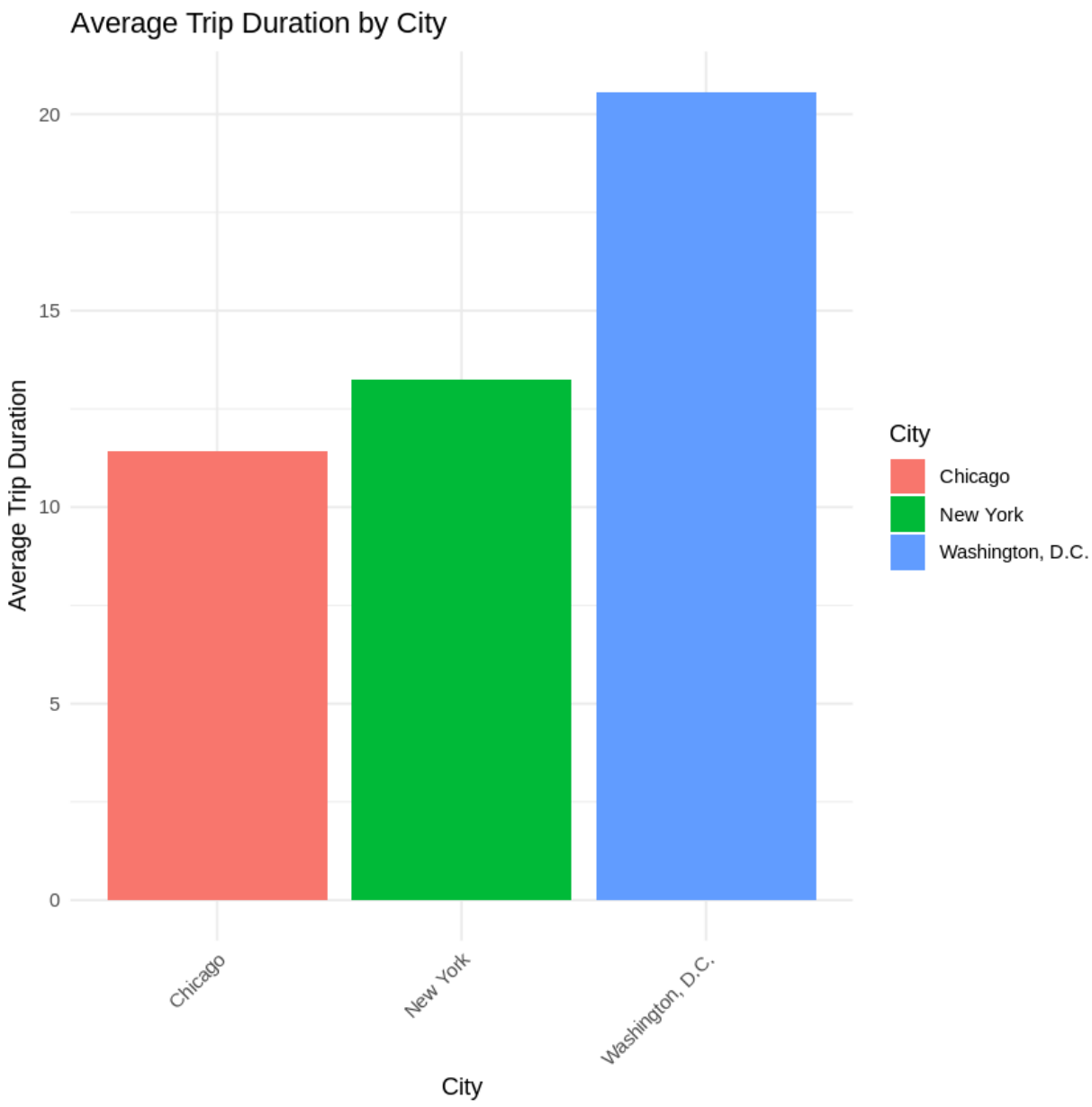
Question 3

What is the average trip duration by city?

In [195]:
```
# Calculate the mean Trip.Duration for each city
average_trip_duration <- aggregate(merged$Trip.Duration, by = list(mer
ged$City), FUN = mean)

# Rename the columns
colnames(average_trip_duration) <- c("City", "Average_Trip_Duration")


# Create a bar plot for the average Trip.Duration for each city
average_trip_duration_bar_plot <- ggplot(average_trip_duration, aes(x
= City, y = Average_Trip_Duration, fill = City)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Trip Duration by City", x = "City", y = "Avera
ge Trip Duration") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(average_trip_duration)
```

```
             City Average_Trip_Duration
1         Chicago              11.41712
2        New York              13.25573
3 Washington, D.C.              20.56589
```

In [196]:
```
# Print the bar plot
print(average_trip_duration_bar_plot)
```



While Chicago and New York have similar average trip durations at about 11 and 13 minutes, Washington, D.C. has a much longer average trip duration at just over 20 minutes.

Conclusions and limitations

This was an interesting and challenging project, which led to some interesting discoveries. The most interesting for me was the average trip duration being so much longer in Washington, D.C. Could it be the geography? Are there aggressive hills that cause slower travel? This would be interesting to explore further somehow. Although some insights were uncovered, there are limitations that we must be aware of. Some of the data was removed due to having 'null values', and the 'age' and 'birth year' data was missing from the Washington, D. C. data set. Nevertheless, this was an interesting project used to dip my toes into all that R has to offer.

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 (https://review.udacity.com/#!/rubrics/2508/view).

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