

Udacity R Assignment - Bikesharing

James Nguyen

28/04/2022

Contents

1	Bike Share Data	2
2	Question 1	5
3	Question 2	9
4	Question 3	10
5	Question 4	10

1 Bike Share Data

Over the past decade, bicycle-sharing systems have been growing in number and popularity in cities across the world. Bicycle-sharing systems allow users to rent bicycles on a very short-term basis for a price. This allows people to borrow a bike from point A and return it at point B, though they can also return it to the same location if they'd like to just go for a ride. Regardless, each bike can serve several users per day.

Thanks to the rise in information technologies, it is easy for a user of the system to access a dock within the system to unlock or return bicycles. These technologies also provide a wealth of data that can be used to explore how these bike-sharing systems are used.

In this project, you will use data provided by *Motivate*, a bike share system provider for many major cities in the United States, to uncover bike share usage patterns. You will compare the system usage between three large cities: Chicago, New York City, and Washington, DC.

The questions to explore include; 1. What are popular times of travel? i.e. common month? common day of week? common hour? 2. What are the most popular stations and trips? 3. What are the statistical breakdowns for trip duration? 4. What are the counts of each user type?

Set-up and initial synthesis of the dataset is required in order to prepare for the questions later to explore. Installing the required packages followed by setting up the directory and importing the data will occur first. Some preliminary data manipulation will also occur to improve ease and efficiency of which we examine the data.

```
knitr::opts_chunk$set(echo = TRUE)
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## Warning: package 'tibble' was built under R version 4.1.3
```

```
## Warning: package 'tidyr' was built under R version 4.1.3
```

```
## Warning: package 'readr' was built under R version 4.1.3
```

```
## Warning: package 'purrr' was built under R version 4.1.3
```

```
## Warning: package 'stringr' was built under R version 4.1.3
```

```
## Warning: package 'forcats' was built under R version 4.1.3
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.1.3
```

```
#Establishing the working directory and importing the data
```

```
setwd("C:/Users/James Nguyen/Desktop/R/Udacity/Bikeshare")
```

```
ny <- read.csv('new-york-city.csv')
```

```
wash <- read.csv('washington.csv')
```

```
chi <- read.csv('chicago.csv')
```

```
head(ny)
```

```
##           X           Start.Time           End.Time Trip.Duration
## 1 5688089 2017-06-11 14:55:05 2017-06-11 15:08:21          795
## 2 4096714 2017-05-11 15:30:11 2017-05-11 15:41:43          692
## 3 2173887 2017-03-29 13:26:26 2017-03-29 13:48:31         1325
## 4 3945638 2017-05-08 19:47:18 2017-05-08 19:59:01          703
## 5 6208972 2017-06-21 07:49:16 2017-06-21 07:54:46          329
## 6 1285652 2017-02-22 18:55:24 2017-02-22 19:12:03          998
##           Start.Station           End.Station User.Type Gender Birth.Year
## 1 Suffolk St & Stanton St W Broadway & Spring St Subscriber Male      1998
## 2 Lexington Ave & E 63 St           1 Ave & E 78 St Subscriber Male      1981
## 3           1 Pl & Clinton St      Henry St & Degraw St Subscriber Male      1987
## 4 Barrow St & Hudson St           W 20 St & 8 Ave Subscriber Female    1986
## 5           1 Ave & E 44 St           E 53 St & 3 Ave Subscriber Male      1992
## 6 State St & Smith St      Bond St & Fulton St Subscriber Male      1986
```

```
head(wash)
```

```
##           X           Start.Time           End.Time Trip.Duration
## 1 1621326 2017-06-21 08:36:34 2017-06-21 08:44:43        489.066
## 2  482740 2017-03-11 10:40:00 2017-03-11 10:46:00        402.549
## 3 1330037 2017-05-30 01:02:59 2017-05-30 01:13:37        637.251
## 4  665458 2017-04-02 07:48:35 2017-04-02 08:19:03       1827.341
## 5 1481135 2017-06-10 08:36:28 2017-06-10 09:02:17       1549.427
## 6 1148202 2017-05-14 07:18:18 2017-05-14 07:24:56        398.000
##           Start.Station
## 1           14th & Belmont St NW
## 2           Yuma St & Tenley Circle NW
## 3           17th St & Massachusetts Ave NW
## 4           Constitution Ave & 2nd St NW/DOL
## 5 Henry Bacon Dr & Lincoln Memorial Circle NW
## 6           1st & K St SE
##           End.Station User.Type
## 1           15th & K St NW Subscriber
## 2           Connecticut Ave & Yuma St NW Subscriber
## 3           5th & K St NW Subscriber
## 4           M St & Pennsylvania Ave NW Customer
## 5           Maine Ave & 7th St SW Subscriber
## 6 Eastern Market Metro / Pennsylvania Ave & 7th St SE Subscriber
```

```
head(chi)
```

```
##           X           Start.Time           End.Time Trip.Duration
## 1 1423854 2017-06-23 15:09:32 2017-06-23 15:14:53          321
## 2  955915 2017-05-25 18:19:03 2017-05-25 18:45:53         1610
## 3   9031 2017-01-04 08:27:49 2017-01-04 08:34:45          416
## 4  304487 2017-03-06 13:49:38 2017-03-06 13:55:28          350
## 5   45207 2017-01-17 14:53:07 2017-01-17 15:02:01          534
## 6 1473887 2017-06-26 09:01:20 2017-06-26 09:11:06          586
##           Start.Station           End.Station User.Type Gender
## 1 Wood St & Hubbard St      Damen Ave & Chicago Ave Subscriber Male
## 2 Theater on the Lake Sheffield Ave & Waveland Ave Subscriber Female
```

```
## 3      May St & Taylor St      Wood St & Taylor St Subscriber Male
## 4 Christiana Ave & Lawrence Ave St. Louis Ave & Balmoral Ave Subscriber Male
## 5      Clark St & Randolph St Desplaines St & Jackson Blvd Subscriber Male
## 6 Clinton St & Washington Blvd      Canal St & Taylor St Subscriber Male
## Birth.Year
## 1      1992
## 2      1992
## 3      1981
## 4      1986
## 5      1975
## 6      1990
```

```
##First wish to combine all working datasets
```

```
#Washington lacks a gender and birth-year column, so we will first create a NA column
```

```
wash$Gender <- NA #'NONE' or any character types do not work as the bind_rows function needs matching
```

```
wash$Birth.Year <- NA
```

```
#Create a location column to retain city split
```

```
ny$location <- 'NY'
```

```
wash$location <- 'WASH'
```

```
chi$location <- 'CHI'
```

```
#Combine all datasets together into ALL, rbind as the variables are the same
```

```
ALL <- bind_rows (ny, wash, chi)
```

2 Question 1

Question 1 examines the popular times of travel. Analysis is provided below to which was used to answer these questions.

1. What is the most common month? June was consistently seen as the most popular month across all locations.
2. What is the most common day of the week to travel? Wednesday was the most popular for Washington and New York, albeit Chicago saw Tuesday to be its most popular day.
3. What is the most common hour of the day to travel? Chicago and New York saw 5:00PM to be its common hour of travel which aligns with common peak-hour notions. Washington however saw 8:00AM to be its common hour.

```
class(ALL$Start.Time) #character
```

```
## [1] "character"
```

```
##Lubridate is able to format character types into date formats
```

```
#ALL$Start.Time <- as.POSIXlt(ALL$Start.Time, format="%d/%m/%Y")
```

```
ALL$Start.Time <- ymd_hms(ALL$Start.Time)
```

```
ALL$month <- month(ALL$Start.Time, label=TRUE)
```

```
ALL$day <- wday(ALL$Start.Time, label=TRUE)
```

```
ALL$hour <- hour(ALL$Start.Time)
```

```
by(ALL$month, ALL$location, summary)
```

```
## ALL$location: CHI
##   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
## 21809 32057 29639 51659 66755 98081     0     0     0     0     0     0
## -----
## ALL$location: NY
##   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
## 31882 34741 32164 58176 67015 76022     0     0     0     0     0     0
## -----
## ALL$location: WASH
##   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
## 30053 38932 41863 62620 58193 68339     0     0     0     0     0     0
```

```
by(ALL$day, ALL$location, summary)
```

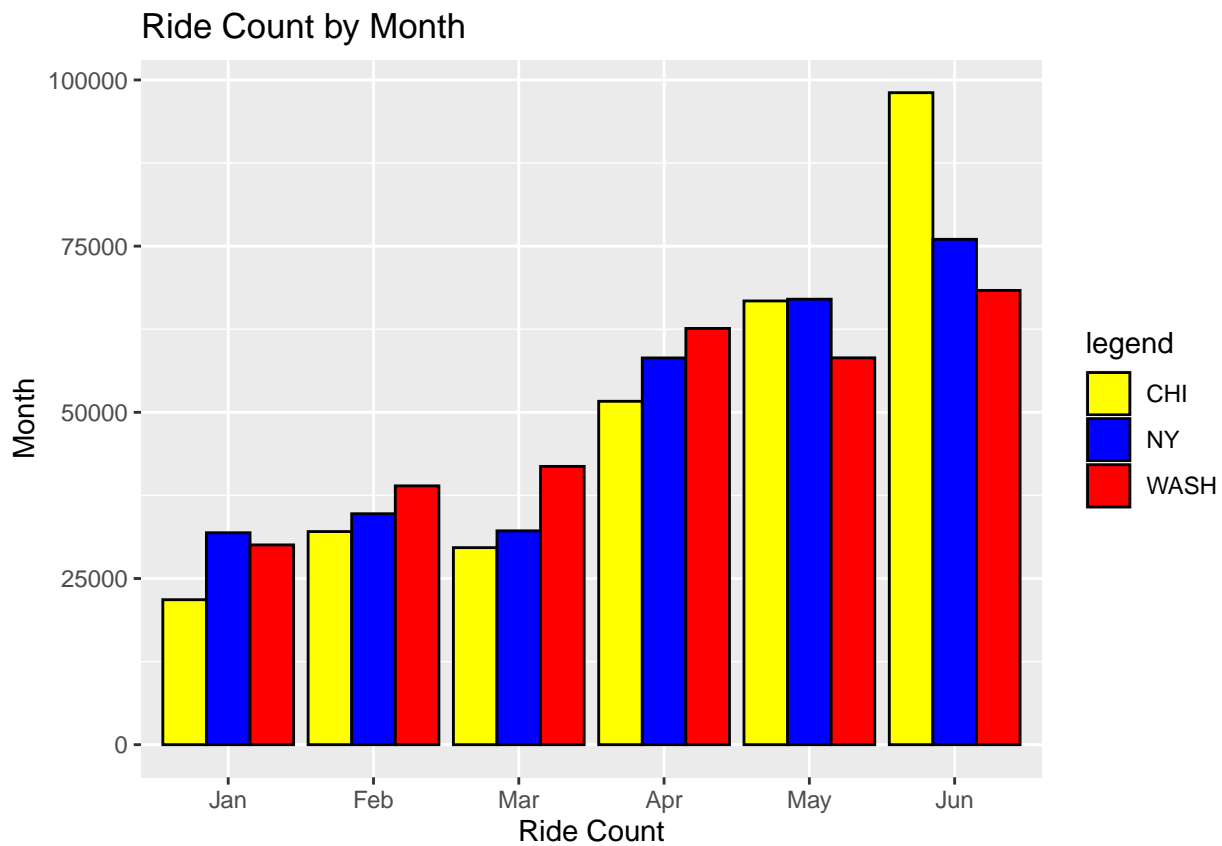
```
## ALL$location: CHI
##   Sun   Mon   Tue   Wed   Thu   Fri   Sat
## 38775 44881 45912 42530 43095 43922 40885
## -----
## ALL$location: NY
##   Sun   Mon   Tue   Wed   Thu   Fri   Sat
## 36151 41923 43752 52087 47497 44664 33926
## -----
## ALL$location: WASH
##   Sun   Mon   Tue   Wed   Thu   Fri   Sat
## 39576 39930 44519 48156 43946 43280 40593
```

```
by(ALL$hour, ALL$location, summary)
```

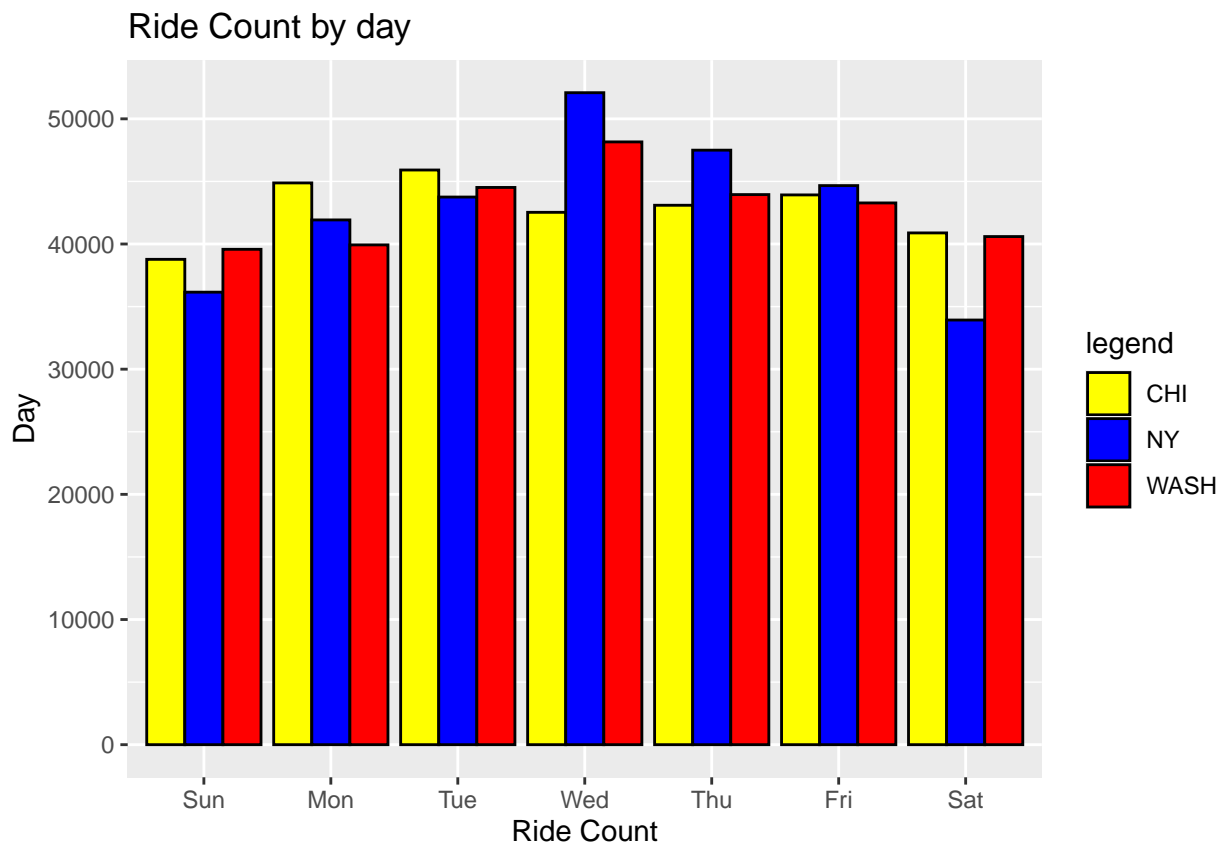
```
## ALL$location: CHI
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.00   10.00   14.00   13.69   17.00   23.00
## -----
## ALL$location: NY
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.00   10.00   15.00   13.93   18.00   23.00
## -----
## ALL$location: WASH
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   6.000   8.000   9.205  12.000  23.000
```

Creating visualisations

```
ggplot(aes(x = month, fill = location), data = ALL) +
  geom_bar(position = 'dodge', colour="black") +
  ggtitle('Ride Count by Month') +
  xlab('Ride Count') +
  ylab('Month') +
  scale_fill_manual("legend", values = c("CHI" = "Yellow", "NY" = "blue", "WASH" = "Red"))
```

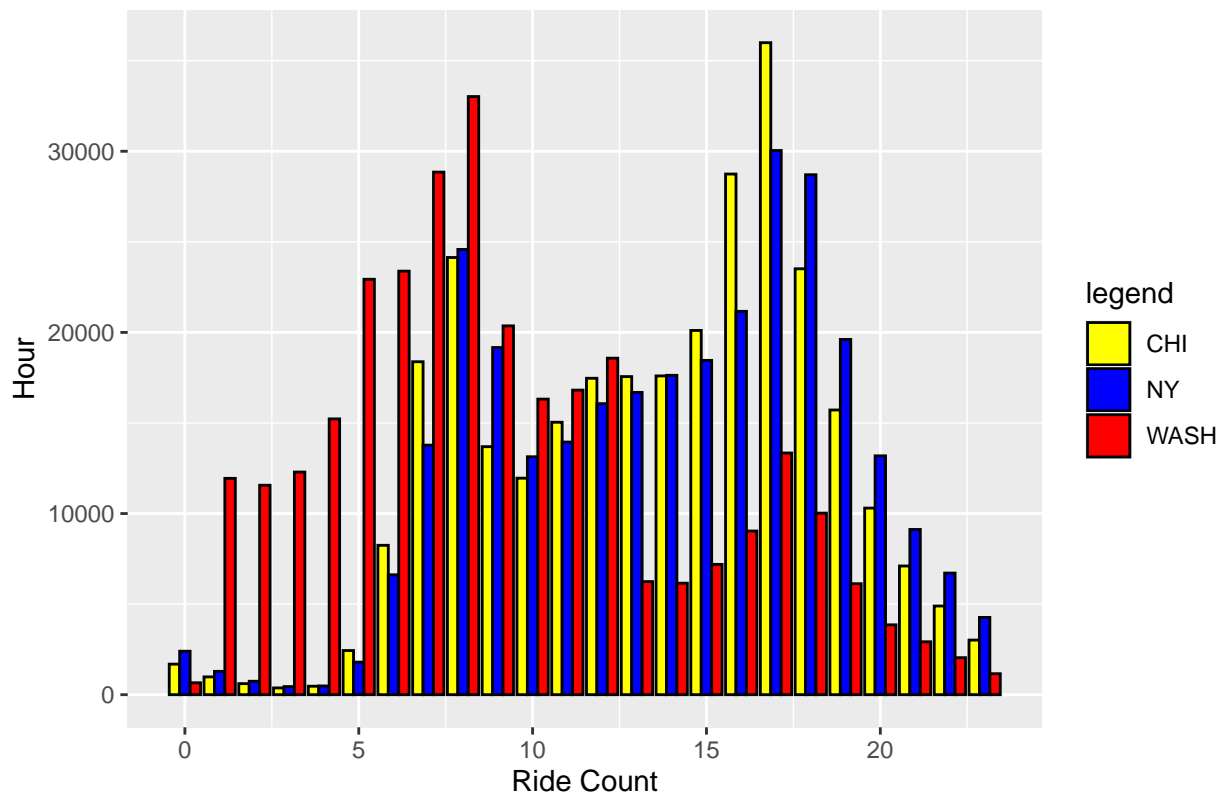


```
ggplot(aes(x = day, fill = location), data = ALL) +
  geom_bar(position = 'dodge', colour="black") +
  ggtitle('Ride Count by day') +
  xlab('Ride Count') +
  ylab('Day') +
  scale_fill_manual("legend", values = c("CHI" = "Yellow", "NY" = "blue", "WASH" = "Red"))
```



```
ggplot(aes(x = hour, fill = location), data = ALL) +
  geom_bar(position = 'dodge', colour="black") +
  ggtitle('Ride Count by Hour') +
  xlab('Ride Count') +
  ylab('Hour') +
  scale_fill_manual("legend", values = c("CHI" = "Yellow", "NY" = "blue", "WASH" = "Red"))
```


Ride Count by Hour



3 Question 2

Question 2 examines the popular stations and trips. Analysis is provided below to which was used to answer these questions.

1. What is the common start station?
2. What is the most common end station?
3. What is the most common trip from start to end?

```
by(ALL$Start.Station, ALL$location, summary)
```

```
## ALL$location: CHI
##   Length   Class      Mode
##   300000 character character
## -----
## ALL$location: NY
##   Length   Class      Mode
##   300000 character character
## -----
## ALL$location: WASH
##   Length   Class      Mode
##   300000 character character
```

4 Question 3

Question 3 aims to analyse the trip duration. Analysis is provided below to which was used to answer these questions.

1. What is the total travel time for users in different cities? Chicago users have travelled a total of 78020 hours (rounded to nearest integer), New York users have travelled for a total of 74974 hours and Washington users have travelled for a total of 103107 hours.
2. What is average travel time for users in different cities? Chicago users travel an average of 15.6 minutes. New York users travel for an average of 15 minutes and Washington users travel for an average of 20.6 minutes.

```
ALL$Trip.Duration <- ALL$Trip.Duration/3600
by(ALL$Trip.Duration, ALL$location, sum)
```

```
## ALL$location: CHI
## [1] 78019.94
## -----
## ALL$location: NY
## [1] 74973.68
## -----
## ALL$location: WASH
## [1] 103106.7
```

```
ALL$Trip.Duration <- ALL$Trip.Duration*60
by(ALL$Trip.Duration, ALL$location, summary)
```

```
## ALL$location: CHI
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   6.55   11.17   15.60   18.75 1437.07
## -----
## ALL$location: NY
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.02   6.13   10.15   14.99   17.57 35929.58
## -----
## ALL$location: WASH
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   6.844   11.775   20.621   20.490 20594.370
```

5 Question 4

Question 4 examines the user types. Analysis is provided below to which was used to answer these questions.

1. What are the counts of each user type?
2. What are the counts of each gender? (only available for NYC and Chicago)
3. What are the earliest, most recent, most common year of birth (Only available for NYC and Chicago)

```
#y(ALL$User.Type, ALL$location, count)
```