Added data to the rstudio

divvy\_trips\_2019\_q1<-read.csv("Divvy\_Trips\_2019\_Q1.csv")

divvy\_trips\_2020\_q1<-read.csv("Divvy\_Trips\_2020\_Q1.csv")

Filtered the data for tripduration for customer usertype by single use as noted in each set of data

customer\_filter\_divvy2019 <- filter(divvy\_trips\_2019\_q1, usertype == "Customer")

customer\_filter\_divvy2020 <- filter(divvy\_trips\_2020\_q1, member\_casual == "casual")

subscriber\_filter\_divvy2019 <- filter(divvy\_trips\_2019\_q1, usertype == "Subscriber")

subscriber\_filter\_divvy2020 <- filter(divvy\_trips\_2020\_q1, member\_casual == "member")

rename column data for customer\_filter\_divvy2019 usertype to member\_casual, from\_station\_name to Start\_station\_name and trip\_id to ride\_id, to match customer\_filter\_divvy2020

customer\_filter\_divvy2019 <- customer\_filter\_divvy2019 %>% rename\_at("usertype", ~"member\_casual")

customer\_filter\_divvy2019 <- customer\_filter\_divvy2019 %>% rename\_at("trip\_id", ~"ride\_id")

customer\_filter\_divvy2019 <- customer\_filter\_divvy2019 %>% rename\_at("from\_station\_name", ~"start\_station\_name")

customer\_filter\_divvy2019 <- customer\_filter\_divvy2019 %>% rename\_at("to\_station\_name", ~"end\_station\_name")

subscriber\_filter\_divvy2019 <- subscriber\_filter\_divvy2019 %>% rename\_at("usertype", ~"member\_casual")

subscriber\_filter\_divvy2019 <- subscriber\_filter\_divvy2019 %>% rename\_at("trip\_id", ~"ride\_id")

subscriber\_filter\_divvy2019 <- subscriber\_filter\_divvy2019 %>% rename\_at("from\_station\_name", ~"start\_station\_name")

subscriber\_filter\_divvy2019 <- subscriber\_filter\_divvy2019 %>% rename\_at("to\_station\_name", ~"end\_station\_name")

replace customer with casual in customer\_filter\_divvy2019 column member\_casual

customer\_filter\_divvy2019['member\_casual'][customer\_filter\_divvy2019['member\_casual']== 'Customer'] <- 'casual'

replace Subscriber with member in subscriber\_filter\_divvy2019 column member\_casual

subscriber\_filter\_divvy2019['member\_casual'][subscriber\_filter\_divvy2019['member\_casual']== 'Subscriber'] <- 'member'

Add column tripduration to customer\_filter\_divvy2020 and subscriber\_filter\_divvy2020

customer\_filter\_divvy2020$tripduration <- ""

subscriber\_filter\_divvy2020$tripduration <- ""

Add data to tripduration column of customer\_filter\_divvy2020 and subscriber\_filter\_divvy2020

customer\_filter\_divvy2020$tripduration <- difftime(customer\_filter\_divvy2020$ended\_at, customer\_filter\_divvy2020$started\_at, units = "mins")

subscriber\_filter\_divvy2020$tripduration <- difftime(subscriber\_filter\_divvy2020$ended\_at, subscriber\_filter\_divvy2020$started\_at, units = "mins")

add column year to customer\_filter\_divvy2019, customer\_filter\_divvy2020, subscriber\_filter\_divvy2019, and subscriber\_filter\_divvy2020 with the year of data collection, and as numeric

customer\_filter\_divvy2020$year <- "2020"

customer\_filter\_divvy2019$year <- "2019"

customer\_filter\_divvy2020$year <- as.numeric(customer\_filter\_divvy2020$year)

customer\_filter\_divvy2019$year <- as.numeric(customer\_filter\_divvy2019$year)

subscriber\_filter\_divvy2020$year <- "2020"

subscriber\_filter\_divvy2019$year <- "2019"

subscriber\_filter\_divvy2020$year <- as.numeric(subscriber\_filter\_divvy2020$year)

subscriber\_filter\_divvy2019$year <- as.numeric(subscriber\_filter\_divvy2019$year)

remove columns that are not needed

customer\_filter\_divvy2019[ , c(2,3,4,6,8,11,12)] <- list(NULL)

customer\_filter\_divvy2020[ , c(2:4,6,8:12)] <- list(NULL)

subscriber\_filter\_divvy2019[ , c(2,3,4,6,8,11,12)] <- list(NULL)

subscriber\_filter\_divvy2020[ , c(2:4,6,8:12)] <- list(NULL)

Export data frames to excel to edit it

write.csv(customer\_filter\_divvy2019, "customer\_filter\_divvy2019.csv")

write.csv(customer\_filter\_divvy2020, "customer\_filter\_divvy2020.csv")

write.csv(subscriber\_filter\_divvy2019, "subscriber\_filter\_divvy2019.csv")

write.csv(subscriber\_filter\_divvy2020, "subscriber\_filter\_divvy2020.csv")

in excel mutate tripduration in customer\_filter\_divvy2019 and subscriber\_filter\_divvy2019 by dividing by 60

in excel copy customer\_filter\_divvy2019 information into customer\_filter\_divvy2020, align information and save as casual\_bike\_case\_study.csv and subscriber\_bike\_case\_study.csv

in excel find and replace & with and

upload edited data frames to RStudio

casual\_bike\_case\_study\_splitting <- read.csv("casual\_bike\_case\_study.csv")

subscriber\_bike\_case\_study\_Splitting <- read.csv("subscriber\_bike\_case\_study.csv")

casual\_bike\_case\_study <- read.csv("casual\_bike\_case\_study.csv")

subscriber\_bike\_case\_study <- read.csv("subscriber\_bike\_case\_study.csv")

remove rows using subset,

casual\_bike\_case\_study <- subset(casual\_bike\_case\_study, tripduration >= 60)

subscriber\_bike\_case\_study <- subset(subscriber\_bike\_case\_study, tripduration >= 60)

turn tripduration in bike\_case\_study into hours

casual\_bike\_case\_study <- casual\_bike\_case\_study %>% mutate(tripduration = tripduration/60)

subscriber\_bike\_case\_study <- subscriber\_bike\_case\_study %>% mutate(tripduration = tripduration/60)

create jitter plots with ggplot2 to look at the longest length of borrow for casual and subscriber

ggplot(data = casual\_bike\_case\_study) +

geom\_jitter(aes(x=start\_station\_name, y= tripduration))+

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))

ggplot(data = subscriber\_bike\_case\_study) +

geom\_jitter(aes(x=start\_station\_name, y= tripduration))+

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))

after reviewing the jitterplot, remove more rows by having a tripduration of more than 12 hours

casual\_bike\_case\_study <- subset(casual\_bike\_case\_study, tripduration >= 12)

subscriber\_bike\_case\_study <- subset(subscriber\_bike\_case\_study, tripduration >= 12)

View jitter plot again

ggplot(data = casual\_bike\_case\_study) +

geom\_jitter(aes(x=start\_station\_name, y= tripduration))+

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))

ggplot(data = subscriber\_bike\_case\_study) +

geom\_jitter(aes(x=start\_station\_name, y= tripduration))+

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))

remove outliers by looking at only stations with less than 400 hour tripduration

casual\_bike\_case\_study <- subset(casual\_bike\_case\_study, tripduration <= 400)

subscriber\_bike\_case\_study <- subset(subscriber\_bike\_case\_study, tripduration <= 400)

remove rows where a station is used less than 10 times for casual but more than 5 for subcriber

casual\_bike\_case\_study <- casual\_bike\_case\_study %>% group\_by(start\_station\_name) %>% filter(n()>10) %>% ungroup()

subscriber\_bike\_case\_study <- subscriber\_bike\_case\_study %>% group\_by(start\_station\_name) %>% filter(n()>5) %>% ungroup()

make the year column factor in both versions of subscriber\_bike\_case\_study and subscriber\_bike\_case\_study2

casual\_bike\_case\_study$year <- as.factor(casual\_bike\_case\_study$year)

subscriber\_bike\_case\_study$year <- as.factor(subscriber\_bike\_case\_study$year)

View jitter plot again with color

ggplot(data = casual\_bike\_case\_study) +

geom\_jitter(aes(x=start\_station\_name, y= tripduration, col=year))+

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Trip Duration by Departure Station Casual Riders", subtitle = "12 - 400 hours, 10+ departures", x= "Departure Station Name", y= "Trip Duration in Hours")

ggplot(data = subscriber\_bike\_case\_study) +

geom\_jitter(aes(x=start\_station\_name, y= tripduration, col=year))+

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Trip Duration by Departure Station Subscribers", subtitle = "12 - 400 hours, 5+ departures", x= "Departure Station Name", y= "Trip Duration in Hours")

save Jitter plot to add to report on which stations to target with ads for membership status

Create a data frame for each street identified in the jitter plots

Dearborn\_and\_Erie <- subset(casual\_bike\_case\_study, start\_station\_name == "Dearborn St and Erie St")

Lake\_Shore\_and\_Monroe <- subset(casual\_bike\_case\_study, start\_station\_name == "Lake Shore Dr and Monroe St")

Shedd\_Aquarium <- subset(casual\_bike\_case\_study, start\_station\_name == "Shedd Aquarium")

Streeter\_and\_Grand <- subset(casual\_bike\_case\_study, start\_station\_name == "Streeter Dr and Grand Ave")

Wabash\_and\_Grand <- subset(casual\_bike\_case\_study, start\_station\_name == "Wabash Ave and Grand Ave")

create frequency data frames for each of the new data frames made by subsets

frequency\_table\_Dearborn <- as.data.frame(table(Dearborn\_and\_Erie$start\_station\_name,Dearborn\_and\_Erie$end\_station\_name))

frequency\_table\_Lake\_Shore <- as.data.frame(table(Lake\_Shore\_and\_Monroe$start\_station\_name,Lake\_Shore\_and\_Monroe$end\_station\_name))

frequency\_table\_Shedd\_Aquarium <- as.data.frame(table(Shedd\_Aquarium$start\_station\_name,Shedd\_Aquarium$end\_station\_name))

frequency\_table\_Streeter <- as.data.frame(table(Streeter\_and\_Grand$start\_station\_name,Streeter\_and\_Grand$end\_station\_name))

frequency\_table\_Wabash <- as.data.frame(table(Wabash\_and\_Grand$start\_station\_name,Wabash\_and\_Grand$end\_station\_name))

change column names to match previous data frames

colnames(frequency\_table\_Dearborn) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_table\_Lake\_Shore) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_table\_Shedd\_Aquarium) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_table\_Streeter) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_table\_Wabash) <- c("departing\_station","arriving\_station","frequency")

Make Bar Charts for each frequency table.

ggplot(frequency\_table\_Dearborn, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "forestgreen") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Dearborn and Erie", subtitle = "Casual Rider, 10+ Departures, 12-400 hour rental", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_table\_Lake\_Shore, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "forestgreen") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Lake Shore and Monroe",subtitle = "Casual Rider, 10+ Departures, 12-400 hour rental",x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_table\_Shedd\_Aquarium, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "forestgreen") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Shedd Aquarium",subtitle = "Casual Rider, 10+ Departures, 12-400 hour rental",x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_table\_Streeter, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "forestgreen") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Streeter Dr and Grand Ave",subtitle = "Casual Rider, 10+ Departures, 12-400 hour rental",x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_table\_Wabash, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "forestgreen") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Wabash Ave and Grand Ave",subtitle = "Casual Rider, 10+ Departures, 12-400 hour rental",x = "arriving station", y = "Number of Arrivals")

Save all the plots made to be able to show the Cyclist to show in a slide later for my arguements

Create subset data frames for "Ada St and Washington Blvd" and "Wabash Ave and Grand Ave" based on the stations from trip duration subscriber jitter plot

subscriber\_Ada\_and\_Washington <- subset(subscriber\_bike\_case\_study, start\_station\_name == "Ada St and Washington Blvd")

subscriber\_Wabash\_and\_Grand <- subset(subscriber\_bike\_case\_study, start\_station\_name == "Wabash Ave and Grand Ave")

create frequency charts

frequency\_table\_ada\_subscriber <- as.data.frame(table(subscriber\_Ada\_and\_Washington$start\_station\_name,subscriber\_Ada\_and\_Washington$end\_station\_name))

frequency\_table\_Wabash\_subscriber <- as.data.frame(table(subscriber\_Wabash\_and\_Grand$start\_station\_name,subscriber\_Wabash\_and\_Grand$end\_station\_name))

change column names to match previous data frames

colnames(frequency\_table\_ada\_subscriber) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_table\_Wabash\_subscriber) <- c("departing\_station","arriving\_station","frequency")

create bar charts via ggplot2 for subcriber frequency tables

ggplot(frequency\_table\_ada\_subscriber, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "gold") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Ada St and Washington Blvd",subtitle = "Subscribers, 5+ Departures, 12-400 hour rental", x = "Arriving Station", y = "Number of Arrivals")

ggplot(frequency\_table\_Wabash\_subscriber, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "gold") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Wabash Ave and Grand Ave",subtitle = "Subscribers, 5+ Departures, 12-400 hour rental",x = "Arriving Station", y = "Number of Arrivals")

Look at the frequency numbers for arrival stations of the full subscriber\_bike\_case\_study.csv and casual\_bike\_case\_study.csv

frequency\_table\_subscriber\_arrival\_bike\_case\_study <- as.data.frame(table(subscriber\_bike\_case\_study\_splitting$end\_station\_name))

frequency\_table\_casual\_arrival\_bike\_case\_study <- as.data.frame(table(casual\_bike\_case\_study\_splitting$end\_station\_name))

Rename column names for understandability in frequency tables

colnames(frequency\_table\_subscriber\_arrival\_bike\_case\_study) <- c("arriving\_station","frequency")

colnames(frequency\_table\_casual\_arrival\_bike\_case\_study) <- c("arriving\_station","frequency")

remove frequency >1000 from the casual frequency table and >7500 from the subscriber frequency table.

frequency\_table\_casual\_bike\_case\_study <- subset(frequency\_table\_casual\_arrival\_bike\_case\_study, frequency > 1000)

frequency\_table\_subscriber\_bike\_case\_study <- subset(frequency\_table\_subscriber\_arrival\_bike\_case\_study, frequency > 7500)

create a chart of the frequency\_table\_subscriber\_bike\_case\_study to see most popular arriving station

ggplot(frequency\_table\_subscriber\_bike\_case\_study, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "skyblue") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Top Arrival Station",subtitle = "7500+ Subscribers Q1 2019+2020",x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_table\_casual\_bike\_case\_study, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "skyblue") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Top Arrival Station",subtitle = "1000+ Casual Users Q1 2019+2020",x = "arriving station", y = "Number of Arrivals")

Look at the frequency numbers for arrival stations of the full subscriber\_bike\_case\_study and bike\_case\_study

frequency\_table\_subscriber\_departure\_bike\_case\_study <- as.data.frame(table(subscriber\_bike\_case\_study$start\_station\_name))

frequency\_table\_casual\_departure\_bike\_case\_study <- as.data.frame(table(casual\_bike\_case\_study$start\_station\_name))

Rename column names for understandability in frequency tables

colnames(frequency\_table\_subscriber\_departure\_bike\_case\_study) <- c("departure\_station","frequency")

colnames(frequency\_table\_casual\_departure\_bike\_case\_study) <- c("departure\_station","frequency")

remove frequency >1000 from the casual frequency table and >7500 from the subscriber frequency table.

frequency\_table\_casual\_departure\_bike\_case\_study <- subset(frequency\_table\_casual\_departure\_bike\_case\_study, frequency > 1000)

frequency\_table\_subscriber\_departure\_bike\_case\_study <- subset(frequency\_table\_subscriber\_departure\_bike\_case\_study, frequency > 7500)

create a chart of the departure frequency tables to see most popular departure station

ggplot(frequency\_table\_subscriber\_departure\_bike\_case\_study, aes(x = departure\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "red") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Top Departure Stations" ,subtitle = "7500+ Subscribers Q1 2019+2020",x = "Departure station", y = "Number of Departures")

ggplot(frequency\_table\_casual\_departure\_bike\_case\_study, aes(x = departure\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "red") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Top Departure Stations",subtitle = "1000+ Casual Users Q1 2019 +2020",x = "Departure station", y = "Number of Departures")

Create a data frame for each station identified in the departure bar charts

Casual Deparutre Data Frames

HQ\_QR\_departure\_casual <- subset(casual\_bike\_case\_study\_splitting, start\_station\_name== "HQ QR")

Lake\_Shore\_and\_Monroe\_departure\_casual<-subset(casual\_bike\_case\_study\_splitting, start\_station\_name== "Lake Shore Dr and Monroe St")

Michigan\_and\_Oak\_departure\_casual<-subset(casual\_bike\_case\_study\_splitting, start\_station\_name== "Michigan Ave and Oak St")

Millennium\_Park\_departure\_casual<-subset(casual\_bike\_case\_study\_splitting, start\_station\_name== "Millennium Park")

Shedd\_Aquarium\_departure\_casual<-subset(casual\_bike\_case\_study\_splitting, start\_station\_name== "Shedd Aquarium")

Streeter\_and\_Grand\_departure\_casual<-subset(casual\_bike\_case\_study\_splitting, start\_station\_name== "Streeter Dr and Grand Ave")

subscriber departure data frames

Canal\_and\_Adams\_departure\_subscriber <-subset(subscriber\_bike\_case\_study\_splitting, start\_station\_name=="Canal St and Adams St")

Canal\_and\_Madison\_departure\_subscriber <-subset(subscriber\_bike\_case\_study\_splitting, start\_station\_name=="Canal St and Madison St")

Clinton\_and\_Madison\_departure\_subscriber <-subset(subscriber\_bike\_case\_study\_splitting, start\_station\_name=="Clinton St and Madison St")

Clinton\_and\_washington\_departure\_subscriber <-subset(subscriber\_bike\_case\_study\_splitting, start\_station\_name=="Clinton St and Washington Blvd")

Columbus\_and\_Randolph\_departure\_subscriber <-subset(subscriber\_bike\_case\_study\_splitting, start\_station\_name=="Columbus Dr and Randolph St")

Kingsbury\_and\_Kinzie\_departure\_subscriber <-subset(subscriber\_bike\_case\_study\_splitting, start\_station\_name=="Kingsbury St and Kinzie St")

create frequency data frames for each of the new data frames made by subsets

casual departure frequency frames

frequency\_HQ\_QR\_departure\_casual <- as.data.frame(table(HQ\_QR\_departure\_casual$start\_station\_name,HQ\_QR\_departure\_casual$end\_station\_name))

frequency\_Lake\_Shore\_and\_Monroe\_departure\_casual <- as.data.frame(table(Lake\_Shore\_and\_Monroe\_departure\_casual$start\_station\_name,Lake\_Shore\_and\_Monroe\_departure\_casual$end\_station\_name))

frequency\_Michigan\_and\_Oak\_departure\_casual <- as.data.frame(table(Michigan\_and\_Oak\_departure\_casual$start\_station\_name,Michigan\_and\_Oak\_departure\_casual$end\_station\_name))

frequency\_Millennium\_Park\_departure\_casual <- as.data.frame(table(Millennium\_Park\_departure\_casual$start\_station\_name,Millennium\_Park\_departure\_casual$end\_station\_name))

frequency\_Shedd\_Aquarium\_departure\_casual <- as.data.frame(table(Shedd\_Aquarium\_departure\_casual$start\_station\_name,Shedd\_Aquarium\_departure\_casual$end\_station\_name))

frequency\_Streeter\_and\_Grand\_departure\_casual <- as.data.frame(table(Streeter\_and\_Grand\_departure\_casual$start\_station\_name,Streeter\_and\_Grand\_departure\_casual$end\_station\_name))

subscriber departure frequency frames

frequency\_Canal\_and\_Adams\_departure\_subscriber <- as.data.frame(table(Canal\_and\_Adams\_departure\_subscriber$start\_station\_name,Canal\_and\_Adams\_departure\_subscriber$end\_station\_name))

frequency\_Canal\_and\_Madison\_departure\_subscriber <- as.data.frame(table(Canal\_and\_Madison\_departure\_subscriber$start\_station\_name,Canal\_and\_Madison\_departure\_subscriber$end\_station\_name))

frequency\_Clinton\_and\_Madison\_departure\_subscriber <- as.data.frame(table(Clinton\_and\_Madison\_departure\_subscriber$start\_station\_name,Clinton\_and\_Madison\_departure\_subscriber$end\_station\_name))

frequency\_Clinton\_and\_washington\_departure\_subscriber <- as.data.frame(table(Clinton\_and\_washington\_departure\_subscriber$start\_station\_name,Clinton\_and\_washington\_departure\_subscriber$end\_station\_name))

frequency\_Columbus\_and\_Randolph\_departure\_subscriber <- as.data.frame(table(Columbus\_and\_Randolph\_departure\_subscriber$start\_station\_name,Columbus\_and\_Randolph\_departure\_subscriber$end\_station\_name))

frequency\_Kingsbury\_and\_Kinzie\_departure\_subscriber <- as.data.frame(table(Kingsbury\_and\_Kinzie\_departure\_subscriber$start\_station\_name,Kingsbury\_and\_Kinzie\_departure\_subscriber$end\_station\_name))

change column names to match previous data frames

Casual Departure

colnames(frequency\_HQ\_QR\_departure\_casual) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Lake\_Shore\_and\_Monroe\_departure\_casual) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Michigan\_and\_Oak\_departure\_casual) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Millennium\_Park\_departure\_casual) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Shedd\_Aquarium\_departure\_casual) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Streeter\_and\_Grand\_departure\_casual) <- c("departing\_station","arriving\_station","frequency")

Subscriber Departure

colnames(frequency\_Canal\_and\_Adams\_departure\_subscriber) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Canal\_and\_Madison\_departure\_subscriber) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Clinton\_and\_Madison\_departure\_subscriber) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Clinton\_and\_washington\_departure\_subscriber) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Columbus\_and\_Randolph\_departure\_subscriber) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_Kingsbury\_and\_Kinzie\_departure\_subscriber) <- c("departing\_station","arriving\_station","frequency")

remove rows using subset

Casual Data Frames <50

frequency\_HQ\_QR\_departure\_casual <- subset(frequency\_HQ\_QR\_departure\_casual, frequency >= 50)

frequency\_Lake\_Shore\_and\_Monroe\_departure\_casual <- subset(frequency\_Lake\_Shore\_and\_Monroe\_departure\_casual, frequency >= 50)

frequency\_Michigan\_and\_Oak\_departure\_casual <- subset(frequency\_Michigan\_and\_Oak\_departure\_casual, frequency >= 50)

frequency\_Millennium\_Park\_departure\_casual <- subset(frequency\_Millennium\_Park\_departure\_casual, frequency >= 50)

frequency\_Shedd\_Aquarium\_departure\_casual <- subset(frequency\_Shedd\_Aquarium\_departure\_casual, frequency >= 50)

frequency\_Streeter\_and\_Grand\_departure\_casual <- subset(frequency\_Streeter\_and\_Grand\_departure\_casual, frequency >= 50)

Subscriber Data frames <250

frequency\_Canal\_and\_Adams\_departure\_subscriber <- subset(frequency\_Canal\_and\_Adams\_departure\_subscriber, frequency >= 250)

frequency\_Canal\_and\_Madison\_departure\_subscriber <- subset(frequency\_Canal\_and\_Madison\_departure\_subscriber, frequency >= 250)

frequency\_Clinton\_and\_Madison\_departure\_subscriber <- subset(frequency\_Clinton\_and\_Madison\_departure\_subscriber, frequency >= 250)

frequency\_Clinton\_and\_washington\_departure\_subscriber <- subset(frequency\_Clinton\_and\_washington\_departure\_subscriber, frequency >= 250)

frequency\_Columbus\_and\_Randolph\_departure\_subscriber <- subset(frequency\_Columbus\_and\_Randolph\_departure\_subscriber, frequency >= 250)

frequency\_Kingsbury\_and\_Kinzie\_departure\_subscriber <- subset(frequency\_Kingsbury\_and\_Kinzie\_departure\_subscriber, frequency >= 250)

Bar plot the arriving data frames made above to view the arriving stations

Casual Rider Bar Plots

ggplot(frequency\_HQ\_QR\_departure\_casual, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "purple") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from HQ QR", subtitle = "Casual Rider, 50+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Lake\_Shore\_and\_Monroe\_departure\_casual, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "purple") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Lake Shore and Monroe", subtitle = "Casual Rider, 50+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Michigan\_and\_Oak\_departure\_casual, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "purple") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Michigan and Oak", subtitle = "Casual Rider, 50+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Millennium\_Park\_departure\_casual, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "purple") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Millennium Park", subtitle = "Casual Rider, 50+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Shedd\_Aquarium\_departure\_casual, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "purple") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Shedd Aquarium", subtitle = "Casual Rider, 50+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Streeter\_and\_Grand\_departure\_casual, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "purple") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Streeter and Grand", subtitle = "Casual Rider, 50+ overall usage", x = "arriving station", y = "Number of Arrivals")

Subscriber Bar plots

ggplot(frequency\_Canal\_and\_Adams\_departure\_subscriber, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "gray") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 60, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Canal and Adams", subtitle = "Subscribers, 250+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Canal\_and\_Madison\_departure\_subscriber, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "gray") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Canal and Madison", subtitle = "Subscribers, 250+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Clinton\_and\_Madison\_departure\_subscriber, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "gray") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Clinton and Madison", subtitle = "Subscribers, 250+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Clinton\_and\_washington\_departure\_subscriber, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "gray") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 65, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Clinton and Washington", subtitle = "Subscribers, 250+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Columbus\_and\_Randolph\_departure\_subscriber, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "gray") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Columbus and Randolph", subtitle = "Subscribers, 250+ overall usage", x = "arriving station", y = "Number of Arrivals")

ggplot(frequency\_Kingsbury\_and\_Kinzie\_departure\_subscriber, aes(x = arriving\_station, y = frequency)) +

geom\_bar(stat = "identity", fill = "gray") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Arrivals to Various Stations from Kingsbury and Kinzie", subtitle = "Subscribers, 250+ overall usage", x = "arriving station", y = "Number of Arrivals")

Split data for users who use bikes for 45 or <

remove rows that have more than one hour

casual\_bike\_case\_study\_less\_than\_45 <- subset(casual\_bike\_case\_study\_splitting, tripduration <=1)

subscriber\_bike\_case\_study\_less\_than\_45 <- subset(subscriber\_bike\_case\_study\_splitting, tripduration <=1)

Turn hours back into minutes for tripduration

casual\_bike\_case\_study\_less\_than\_45$tripduration <- casual\_bike\_case\_study\_less\_than\_45$tripduration\*60

subscriber\_bike\_case\_study\_less\_than\_45$tripduration <- subscriber\_bike\_case\_study\_less\_than\_45$tripduration\*60

Remove rows with with more than 45 minutes of use

casual\_bike\_case\_study\_less\_than\_45 <- subset(casual\_bike\_case\_study\_less\_than\_45, tripduration <=45)

subscriber\_bike\_case\_study\_less\_than\_45 <- subset(subscriber\_bike\_case\_study\_less\_than\_45, tripduration<=45)

Remove rows for less than 30 minutes on subscriber data

subscriber\_bike\_case\_study\_less\_than\_45 <- subset(subscriber\_bike\_case\_study\_less\_than\_45, tripduration>=30)

remove rows with data less than 0

casual\_bike\_case\_study\_less\_than\_45 <- subset(casual\_bike\_case\_study\_less\_than\_45, tripduration >=0)

subscriber\_bike\_case\_study\_less\_than\_45 <- subset(subscriber\_bike\_case\_study\_less\_than\_45, tripduration >=0)

Remove results with HQ QR & HUBBARD TEST

casual\_bike\_case\_study\_less\_than\_45<-casual\_bike\_case\_study\_less\_than\_45[!grepl("HQ QR", casual\_bike\_case\_study\_less\_than\_45$end\_station\_name),]

casual\_bike\_case\_study\_less\_than\_45<-casual\_bike\_case\_study\_less\_than\_45[!grepl("HUBBARD ST BIKE CHECKING (LBS-WH-TEST)", casual\_bike\_case\_study\_less\_than\_45$end\_station\_name),]

casual\_bike\_case\_study\_less\_than\_45<-casual\_bike\_case\_study\_less\_than\_45[!grepl("HQ QR", casual\_bike\_case\_study\_less\_than\_45$start\_station\_name),]

casual\_bike\_case\_study\_less\_than\_45<-casual\_bike\_case\_study\_less\_than\_45[!grepl("HUBBARD ST BIKE CHECKING (LBS-WH-TEST)", casual\_bike\_case\_study\_less\_than\_45$start\_station\_name),]

crate a frequency data frame for a 45 min window

frequency\_casual\_bike\_case\_study\_less\_than\_45 <- as.data.frame(table(casual\_bike\_case\_study\_less\_than\_45$start\_station\_name,casual\_bike\_case\_study\_less\_than\_45$end\_station\_name))

frequency\_subscriber\_bike\_case\_study\_less\_than\_45 <-as.data.frame(table(subscriber\_bike\_case\_study\_less\_than\_45$start\_station\_name,subscriber\_bike\_case\_study\_less\_than\_45$end\_station\_name))

Change coloumn names

colnames(frequency\_casual\_bike\_case\_study\_less\_than\_45) <- c("departing\_station","arriving\_station","frequency")

colnames(frequency\_subscriber\_bike\_case\_study\_less\_than\_45) <- c("departing\_station","arriving\_station","frequency")

change the frequency greater than for each frame

frequency\_casual\_bike\_case\_study\_less\_than\_45 <- subset(frequency\_casual\_bike\_case\_study\_less\_than\_45, frequency >=4)

frequency\_subscriber\_bike\_case\_study\_less\_than\_45 <- subset(frequency\_subscriber\_bike\_case\_study\_less\_than\_45, frequency >=650)

make a plot of new data to see who uses a bike for less than 45 minutes

ggplot(data=frequency\_casual\_bike\_case\_study\_less\_than\_45) +

geom\_point(aes(x = departing\_station, y = arriving\_station, col=frequency)) +

scale\_color\_gradient(low = "blue", high = "red")+

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Most Used Stations, Departure to Arrival", subtitle = "Casual Riders, 45 minutes or less, 5+ frequency", x = "Departing Station", y = "Arriving Station", color="Most Used")

ggplot(data=frequency\_subscriber\_bike\_case\_study\_less\_than\_45) +

geom\_point(aes(x = departing\_station, y = arriving\_station, col=frequency)) +

scale\_color\_gradient(low = "blue", high = "red")+

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust=1, hjust = 1))+

labs(title = "Most Used Stations, Departure to Arrival", subtitle = "Subscribers, 45 minutes or less, 650+ frequency", x = "Departing Station", y = "Arriving Station", color="Most Used")