**COMCAST TELECOM CONSUMER COMPLAINTS**

**Submitted by-**

**Name : Prerit Pushp**

**Course : Data Science With R**

**Program : Data Scientist Master Program**

**Submitted to-**

**SimpliLearn**

**BUSINESS SCENARIO**

**DESCRIPTION OF DATA**

Comcast is an American global telecommunication company. The firm has been providing terrible customer service. They continue to fall short despite repeated promises to improve. Only last month (October 2016) the authority fined them a $2.3 million, after receiving over 1000 consumer complaints.

The existing database will serve as a repository of public customer complaints filed against Comcast. It will help to pin down what is wrong with Comcast's customer service.

**Data Dictionary**

Ticket #: Ticket number assigned to each complaint

Customer Complaint: Description of complaint

Date: Date of complaint

Time: Time of complaint

Received Via: Mode of communication of the complaint

City: Customer city

State: Customer state

Zipcode: Customer zip

Status: Status of complaint

Filing on behalf of someone

**Analysis Task**

- Import data into R environment.  
- Provide the trend chart for the number of complaints at monthly and daily granularity levels.  
- Provide a table with the frequency of complaint types.

* Which complaint types are maximum i.e., around internet, network issues, or across any other domains.

- Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.  
- Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:

* Which state has the maximum complaints
* Which state has the highest percentage of unresolved complaints

- Provide the percentage of complaints resolved till date, which were received through theInternet and customer care calls.

The analysis results to be provided with insights wherever applicable.

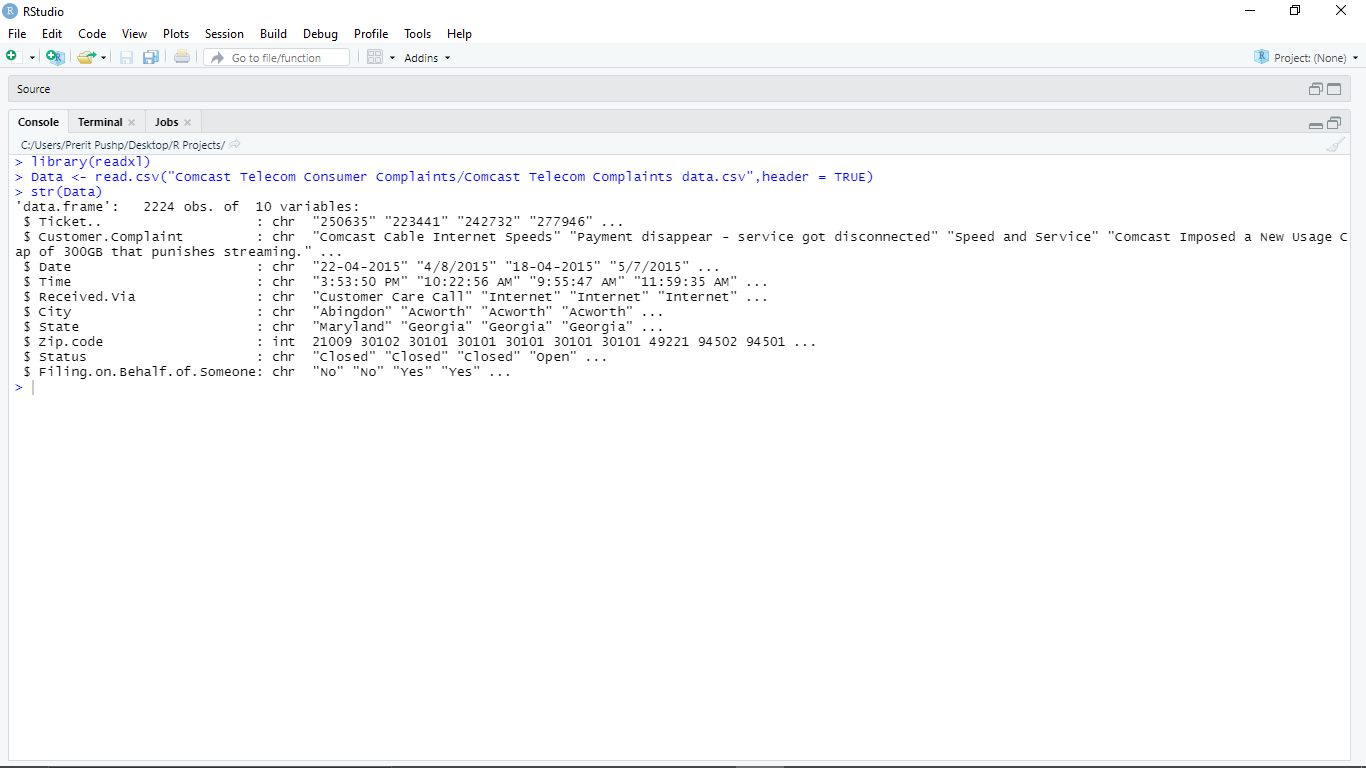
1. Importing the Data into R environment

We will be importing the Data into the R environment by using readXl library.

library(readxl)

Data <- read.csv("Comcast Telecom Consumer Complaints/Comcast Telecom Complaints data.csv",header = TRUE)

str(Data)



2. Clearing The Data

Now, to perform this analysis we will extract the month from the given Date column and create a new column called Month where we store all the months when complaints are filed. So, before extracting month we need to put the dates in the Date column in the correct format. By using the Lubridate library we will transform all the dates in a particular format which will make it easy for us to extract month and make analysis later.

library(lubridate)

li<-parse\_date\_time(x = Data$Date,

orders = c("d m y", "d B Y", "m/d/y"),

locale = Sys.getlocale("LC\_TIME"))

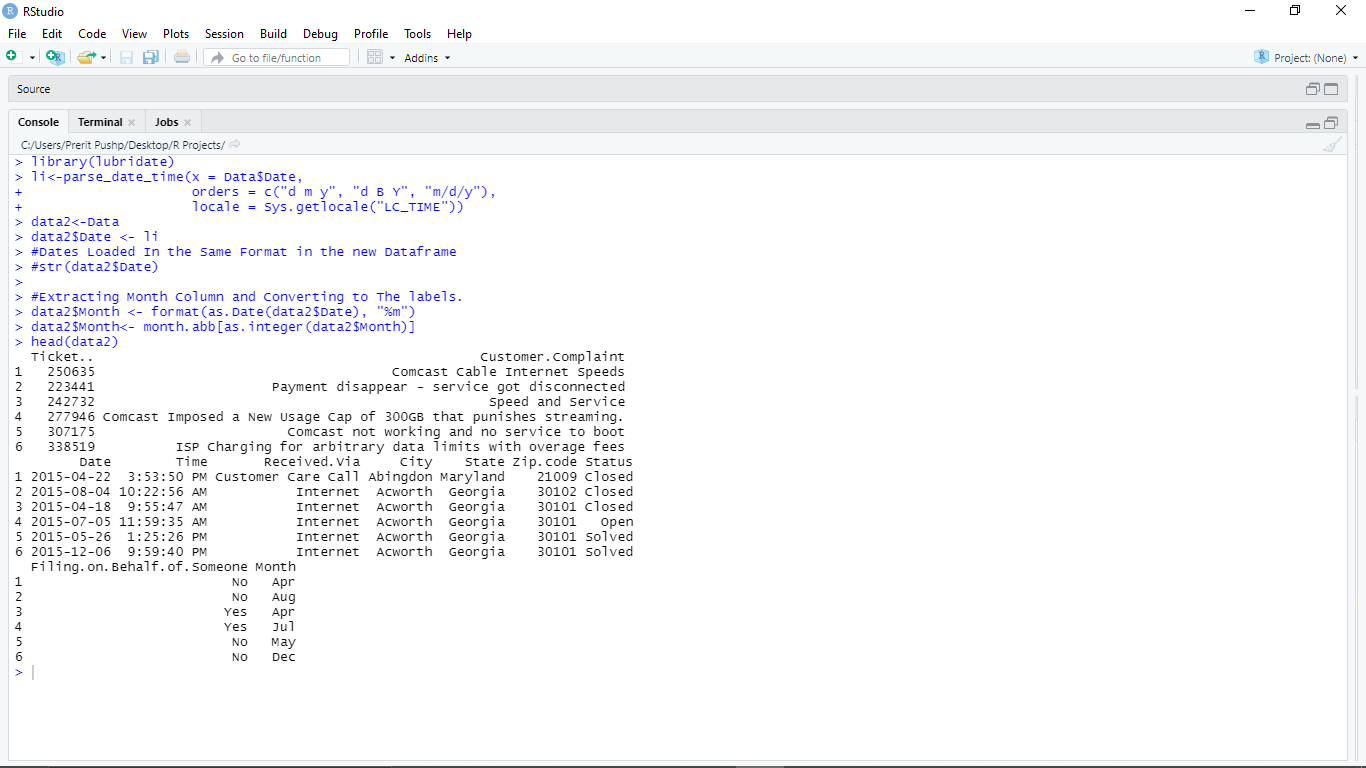
data2<-Data

data2$Date <- li

data2$Month <- format(as.Date(data2$Date), "%m")

data2$Month<- month.abb[as.integer(data2$Month)]

head(data2)



## 3. Analysis Of Data

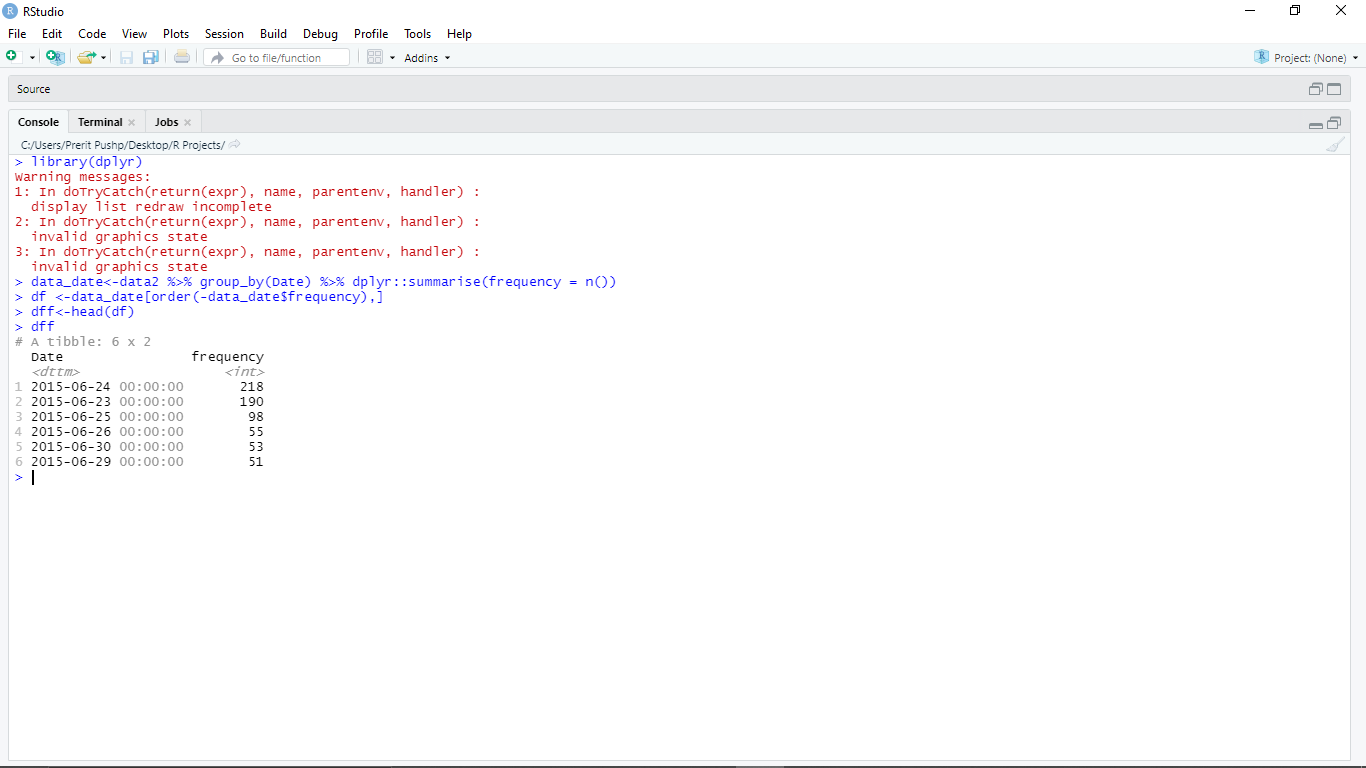
library(dplyr)

data\_date<-data2 %>% group\_by(Date) %>% dplyr::summarise(frequency = n())

df <-data\_date[order(-data\_date$frequency),]

dff<-head(df)

dff



#### **Analysis:- This Above Data Frame Tells me that, on June 24 Comcast was reported with 218 complaints, particularly indicating a bad day for company to fix up many issues on that day.**

library(ggplot2)

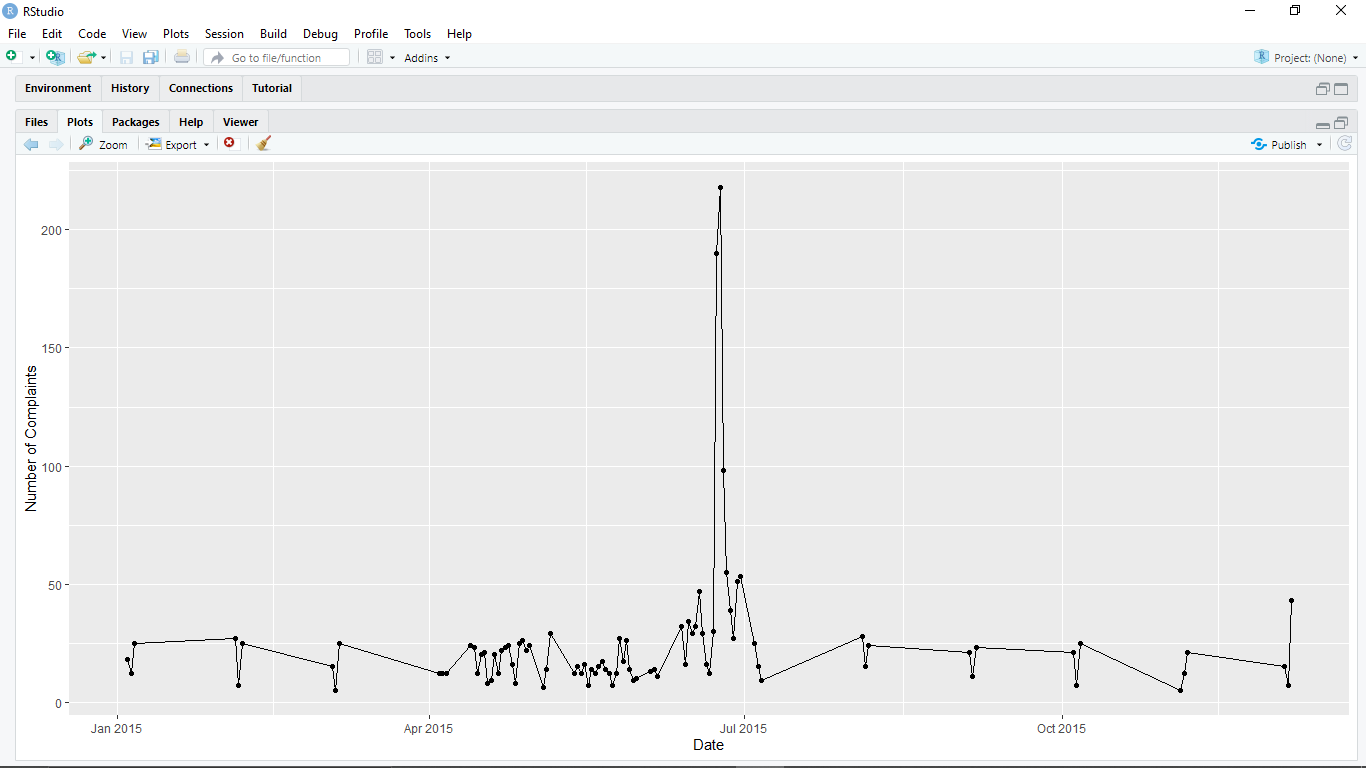
ggplot(data\_date, aes(Date, frequency, group = 1)) +

geom\_point() +

geom\_line() +

xlab("Date") +

ylab("Number of Complaints")



#### **Analysis:- Clearly, from the above Trend Graph, we can easily say that in the month of JUNE 2015, Comcast got reported with Maximum Number of complaints.**

library(ggplot2)

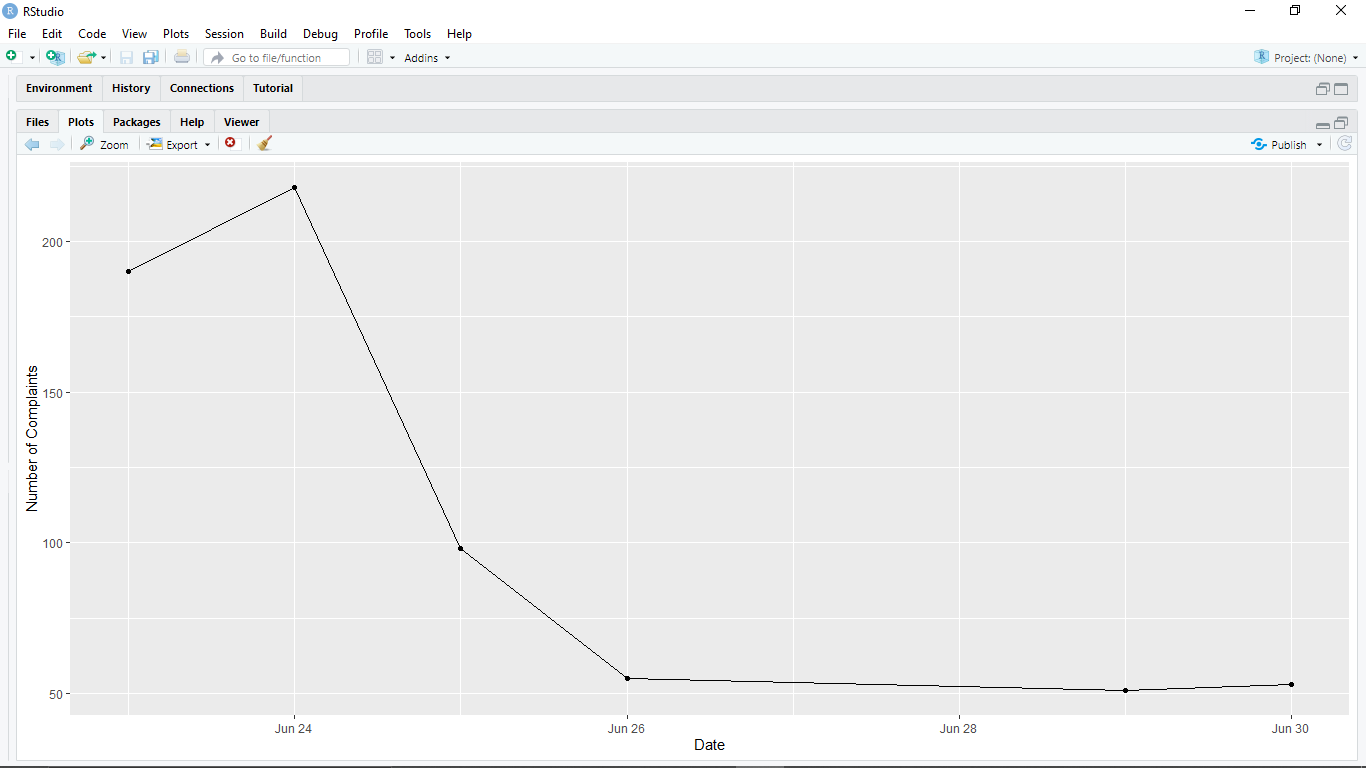
ggplot(dff, aes(Date, frequency, group = 1)) +

geom\_point() +

geom\_line() +

xlab("Date") +

ylab("Number of Complaints")

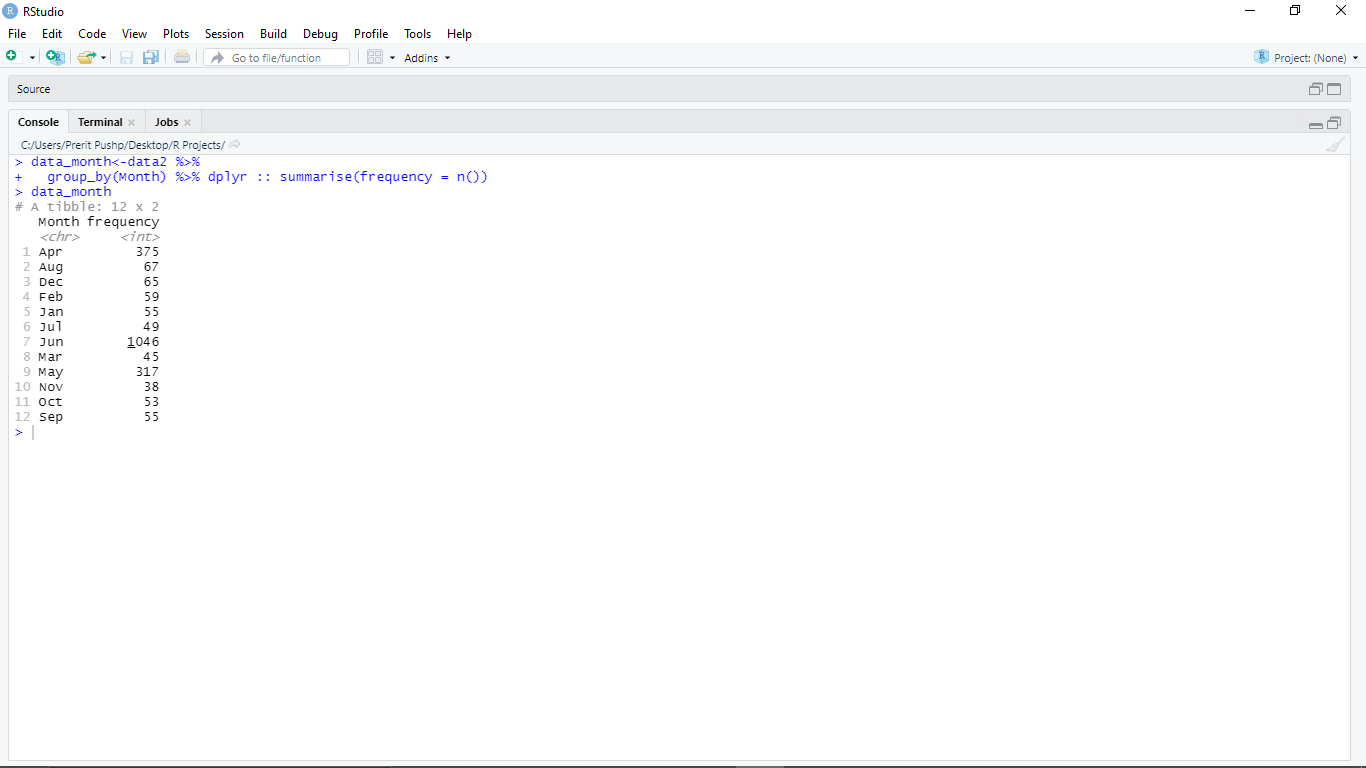


#### **Analysis:- Clearly you can see on June 24, Company got reported with many complaints. This is the following trend for a few observations from the month of June.**

data\_month<-data2 %>%

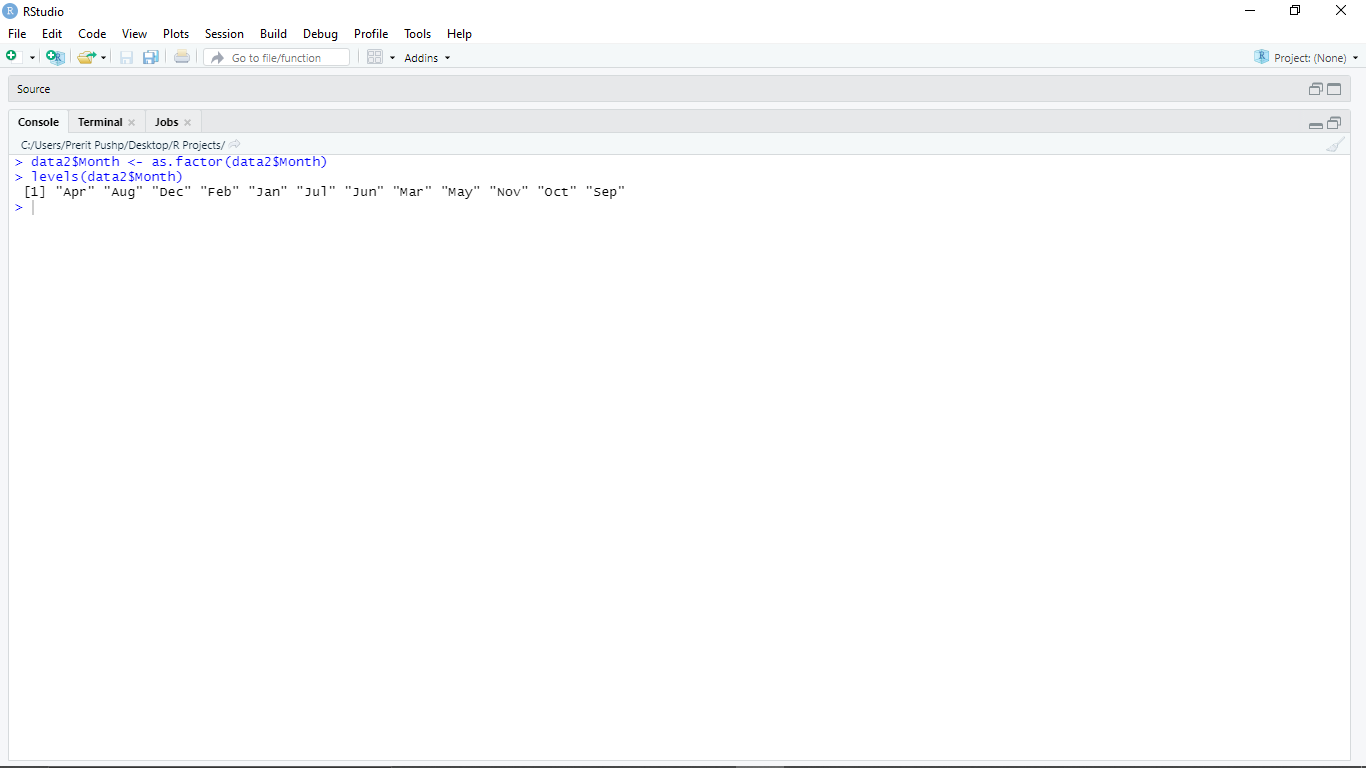
group\_by(Month) %>% dplyr :: summarise(frequency = n())

data\_month



data2$Month <- as.factor(data2$Month)

levels(data2$Month)



library(ggplot2)

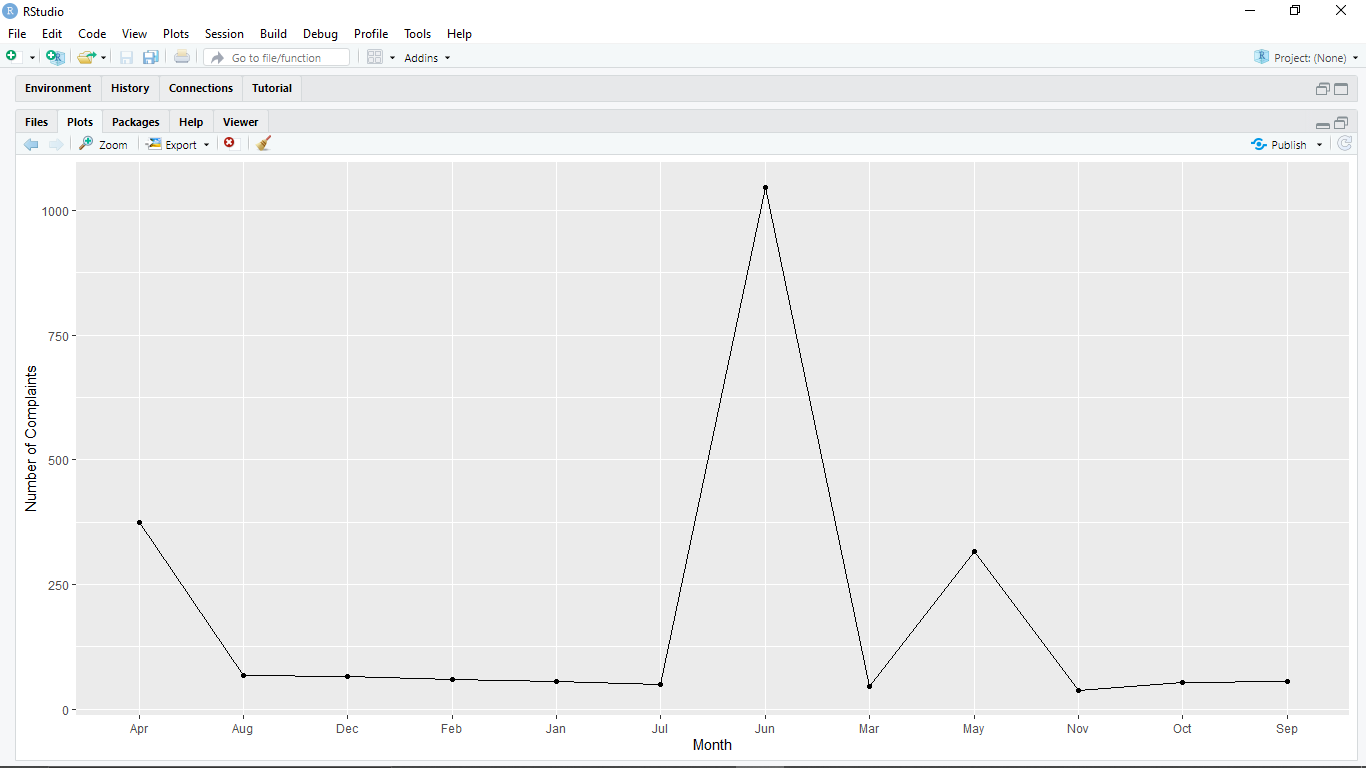
ggplot(data\_month, aes(Month, frequency, group = 1)) +

geom\_point() +

geom\_line() +

xlab("Month") +

ylab("Number of Complaints")



#### **By now we have quite a good amount of insights by which we can clearly say that in the Month of June, date 24 company was filed with a maximum number of complaints.**

### Frequency Table For Customer Complaints During Year 2015 - 2016 Period :-

library(dplyr)

#Converting All String Values to Lower, so as to Eliminate Duplication of Any Complaint

data3<-data2%>% mutate(Customer.Complaint = tolower(Customer.Complaint))

CustTable <- table(data3$Customer.Complaint)

CustTable <- data.frame(CustTable)

filtered<-CustTable %>%

rename(

CustomerComplaintType = Var1,

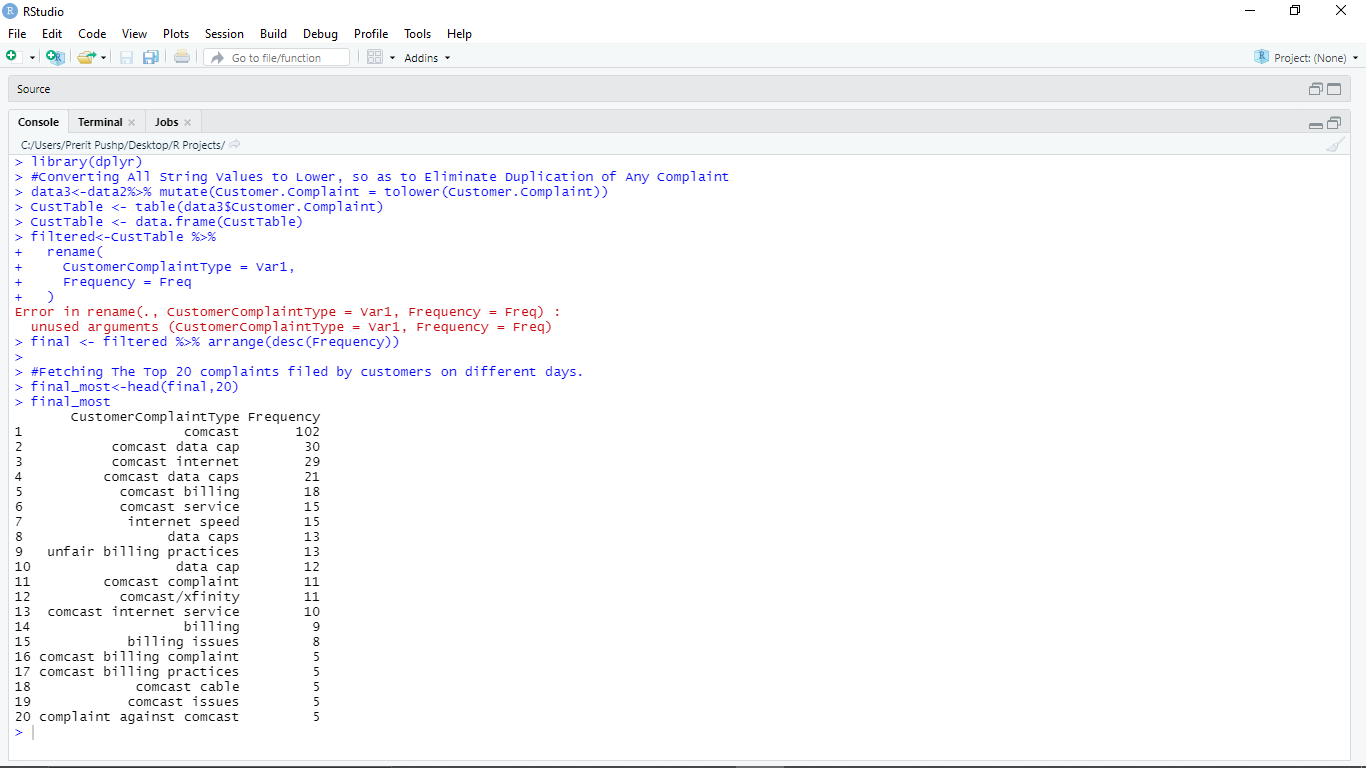
Frequency = Freq

)

final <- filtered %>% arrange(desc(Frequency))

final\_most<-head(final,20)

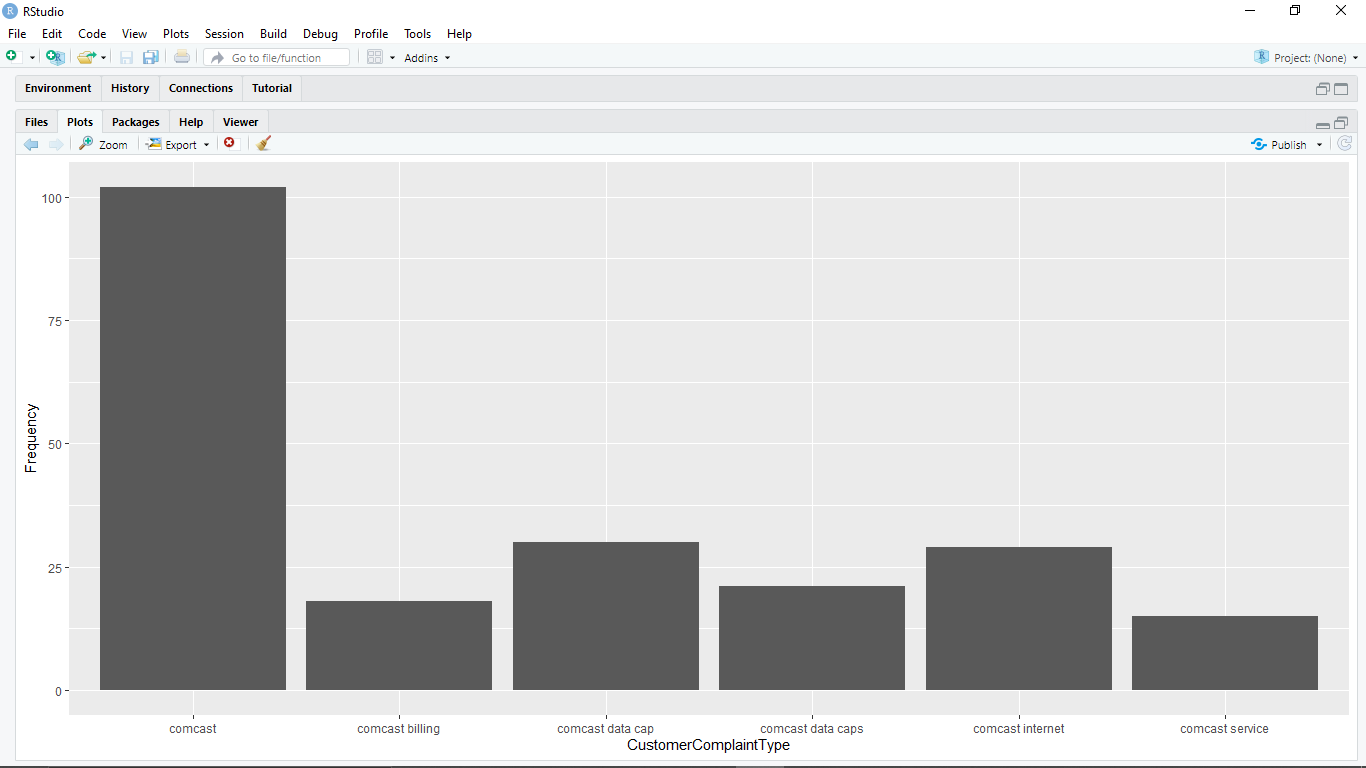
final\_most



library(ggplot2)

ggplot(head(final\_most,6), aes(CustomerComplaintType, Frequency)) +

geom\_bar(stat = "identity")



***Customer Are Mainly complaining about the Data Caps, Internet Speed, Billing Methods and Services that Comcast is Providing and Very few Cases were registered against Comcast Cable Services.***

library(stringr)

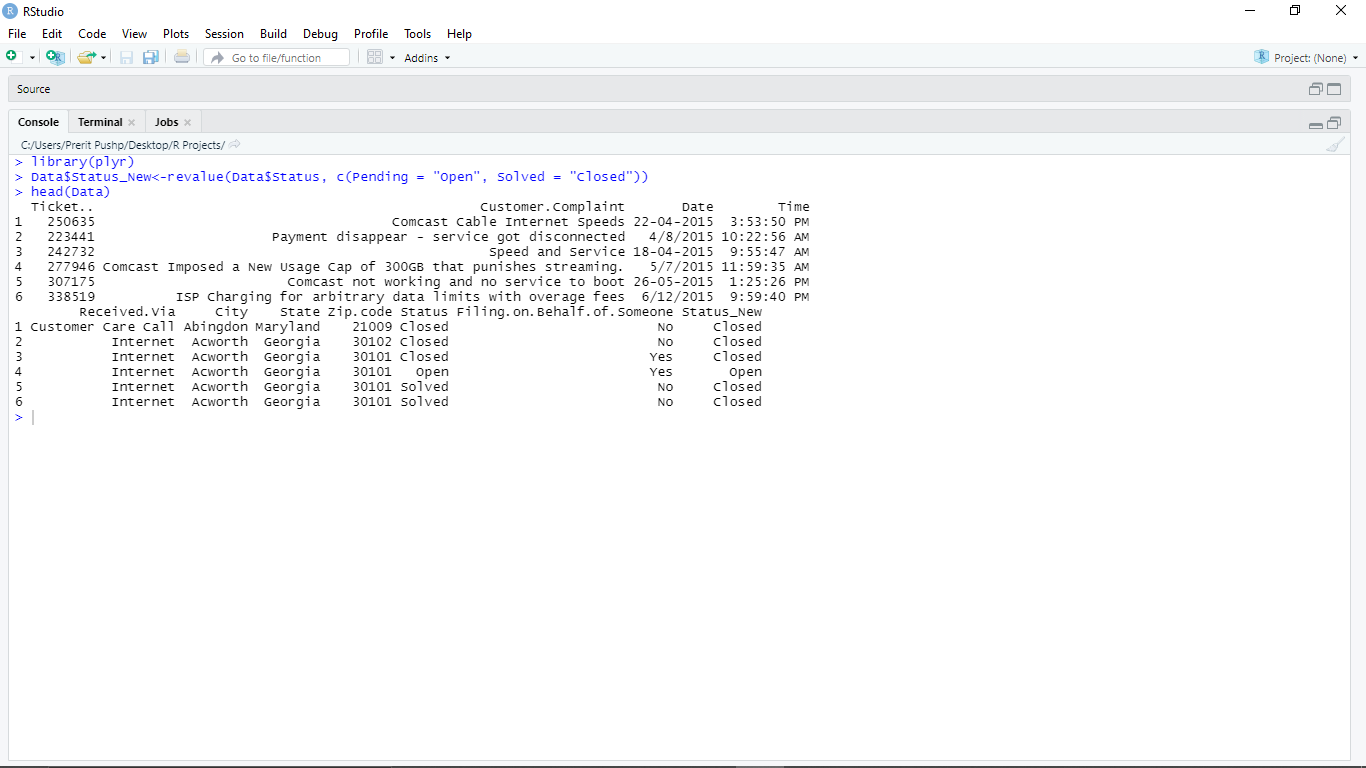
library(tidyverse)

levels(Data$Status)

library(plyr)

Data$Status\_New<-revalue(Data$Status, c(Pending = "Open", Solved = "Closed"))

head(Data)

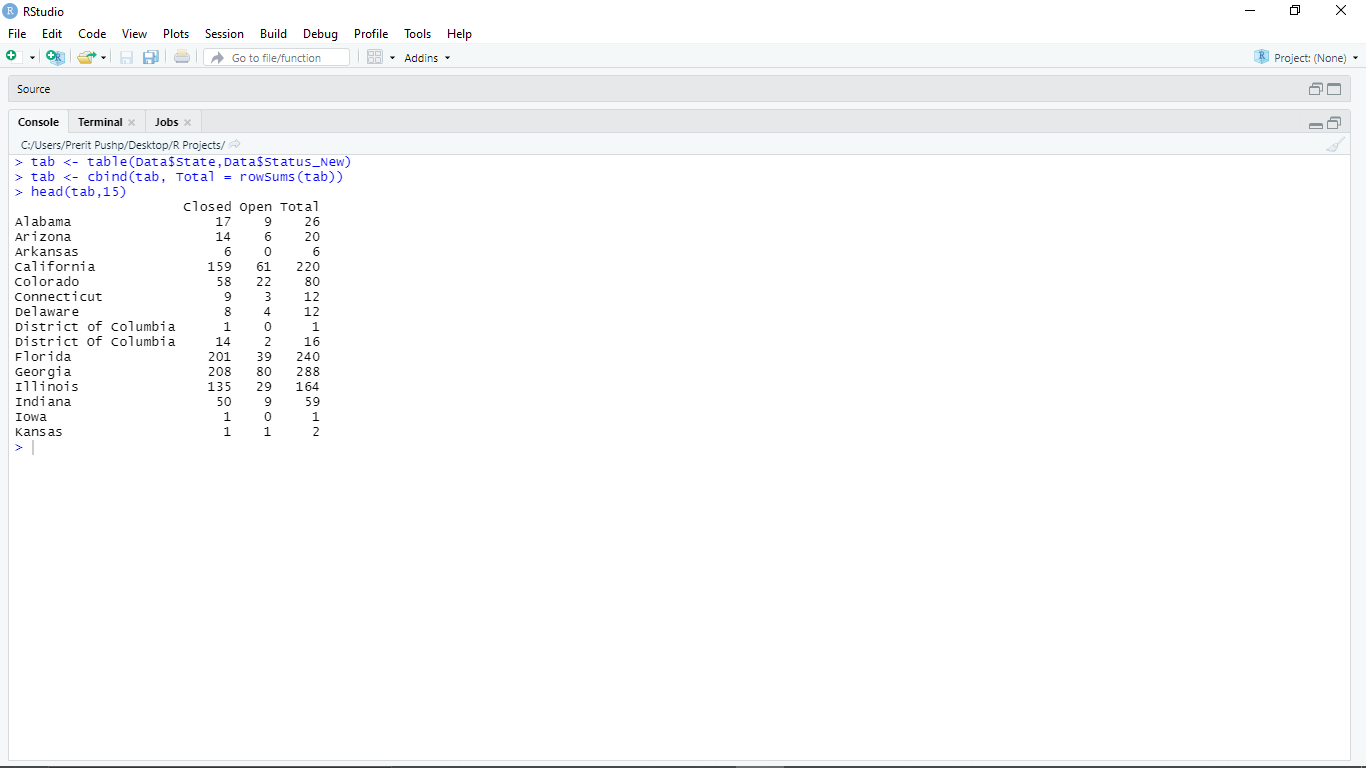


### ***Clearly looking at the above new created column Status1 has only two levels as required. We have merged Pending requests as Open and Solved Requests as Closed.***

tab <- table(Data$State,Data$Status\_New)

tab <- cbind(tab, Total = rowSums(tab))

head(tab,15)



library(gridExtra)

ggplot(Data, aes(y = State)) + geom\_bar(aes(fill = Status\_New))



#### **Analysis:- Clearly Looking at the chart we can say, Georgia and Florida are the Two where Comcast has a good number of Happy customers by solving the issues in bulk.**

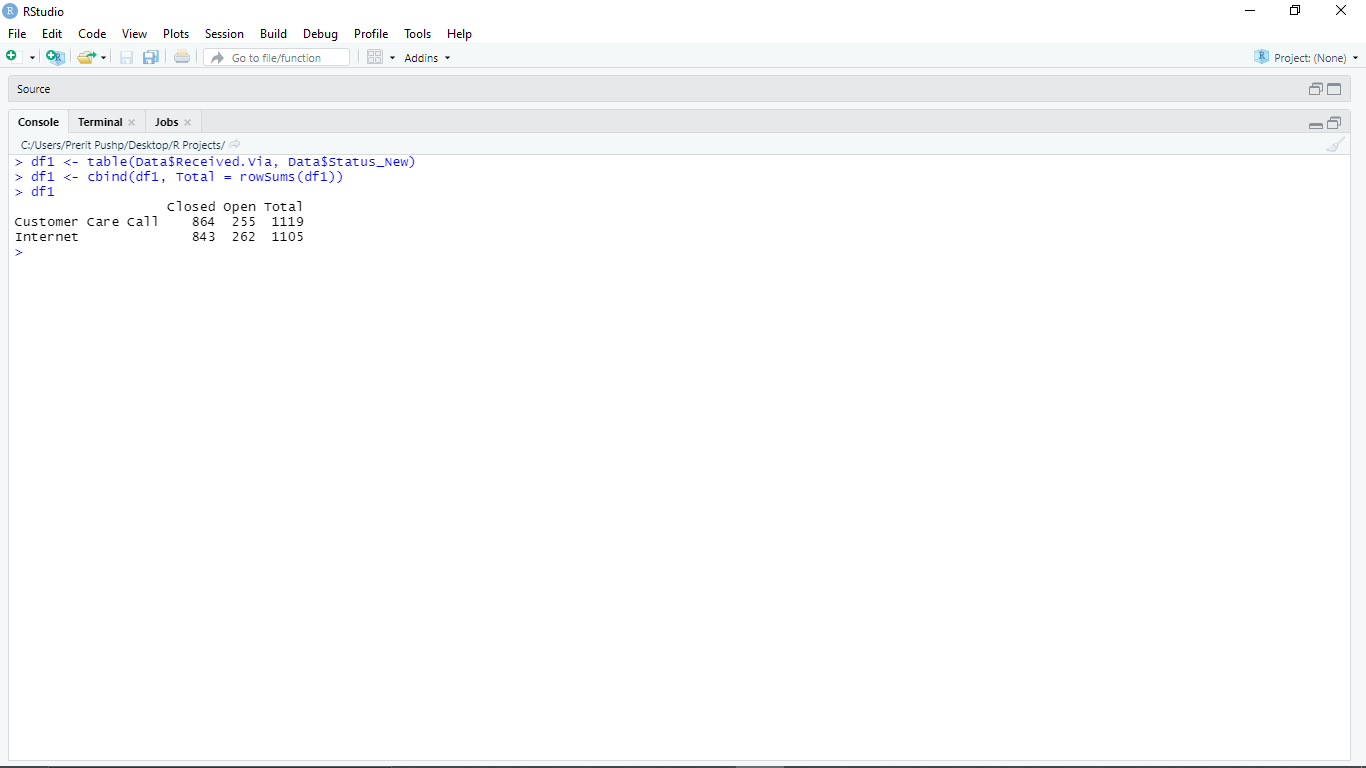
ggplot(Data, aes(y = Received.Via )) + geom\_bar(aes(fill = Status\_New))



df1 <- table(Data$Received.Via, Data$Status\_New)

df1 <- cbind(df1, Total = rowSums(df1))

df1



slices <- c(864, 255)

lbls <- c("Closed", "Open")

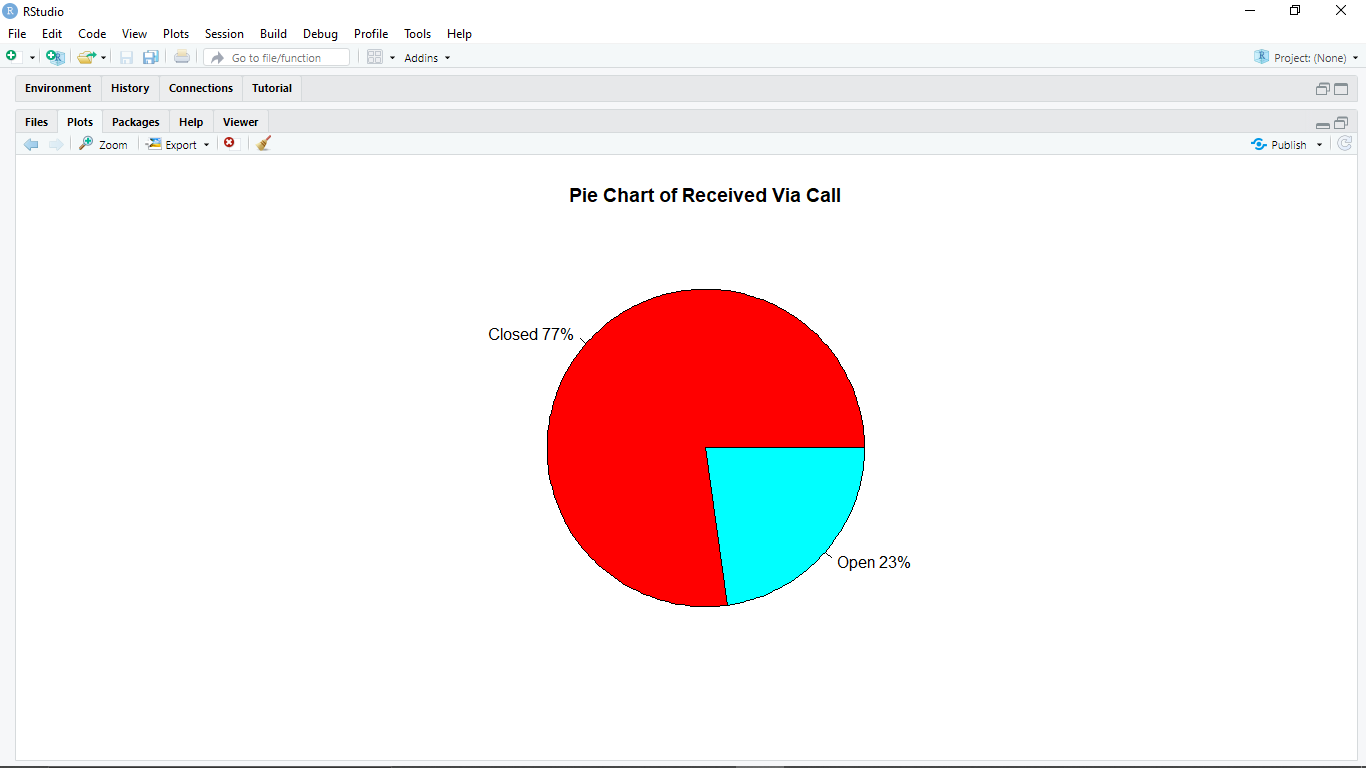
pct <- round(slices/sum(slices)\*100)

lbls <- paste(lbls, pct) # add percents to labels

lbls <- paste(lbls,"%",sep="") # ad % to labels

pie(slices,labels = lbls, col=rainbow(length(lbls)),

main="Pie Chart of Received Via Call")



***Above picture shows pie chart of Received via call complaint in percentage.***

slices <- c(843, 262)

lbls <- c("Closed", "Open")

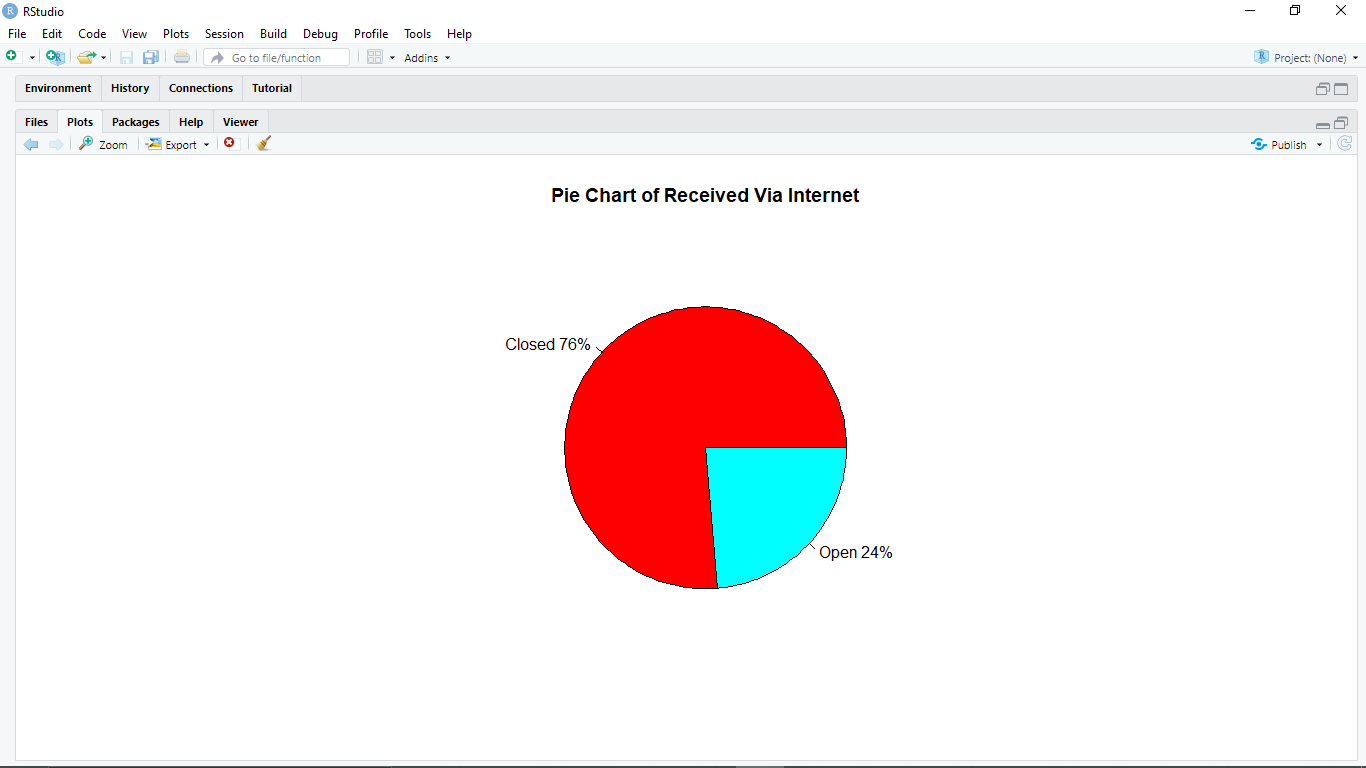
pct <- round(slices/sum(slices)\*100)

lbls <- paste(lbls, pct) # add percents to labels

lbls <- paste(lbls,"%",sep="") # ad % to labels

pie(slices,labels = lbls, col=rainbow(length(lbls)),

main="Pie Chart of Received Via Internet")



#### **Solutions**

1. The Company should Focus more on resolving complaints - Customer Are Mainly complaining about the Data Caps, Internet Speed, Billing Methods and Services that Comcast is Providing and Very few Cases were registered against Comcast Cable Services.
2. In Georgia and Florida company services are already Improving but, in States - California, Colorado and Illinois company should extend their resources in terms of the above-mentioned issues in order to improve their customer servicing.
3. During the month of June and the start of July, the Company reported lots of complaints, so as to for future reference they can keep this in check already so as to provide better services during these months. While working with their BPO clients to extend the staff during such days. Which ensures proper feedback for the particular arisen issue.

**THANK YOU**