# Project on Data Science with R Programming

Name of the project:

**Comcast Telecom Consumer Complaints**

Project Description:

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 the Internet and customer care calls.

Source Code with graphs an insights:

#loading the file in R enviroment

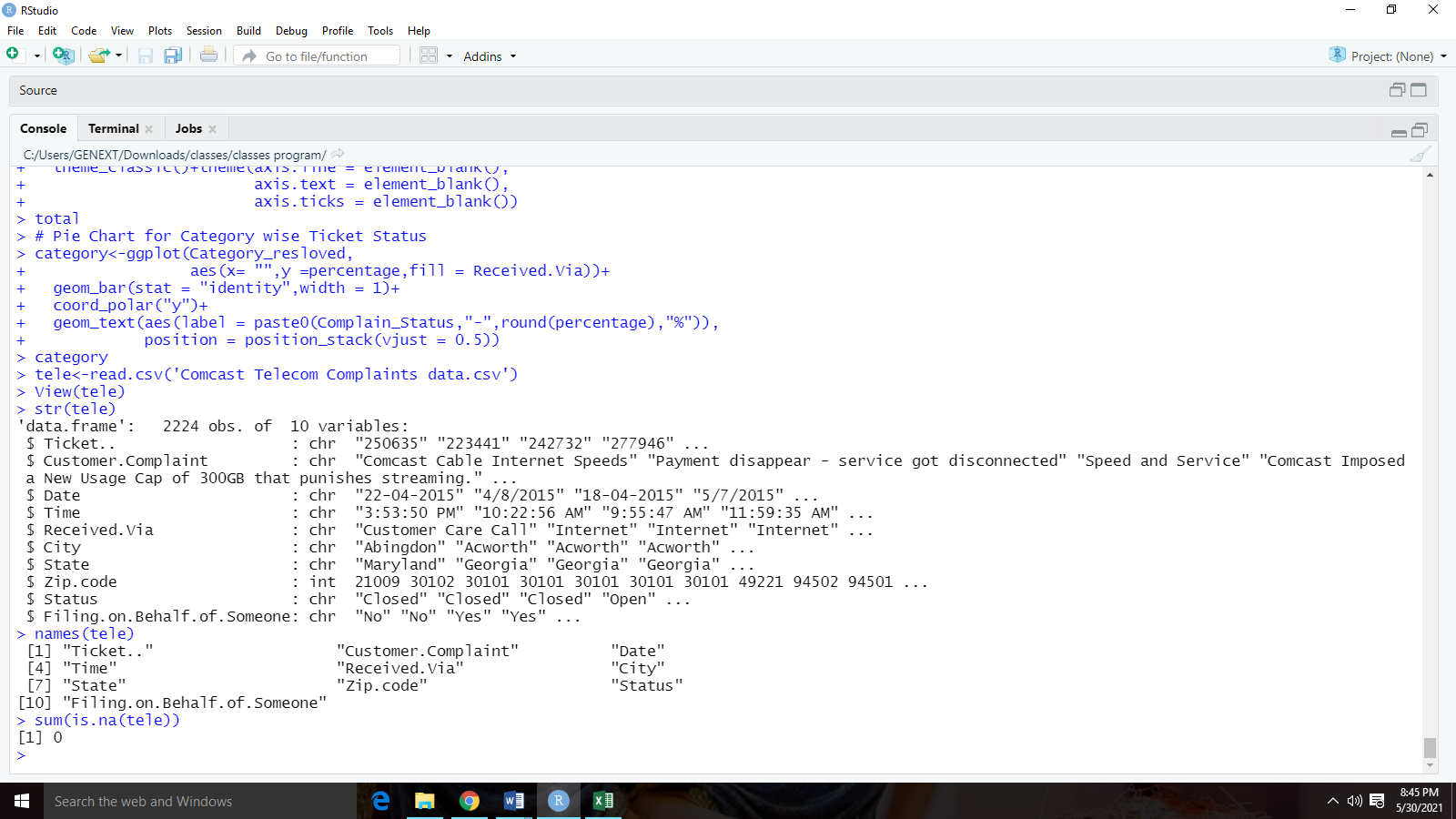
tele<-read.csv('Comcast Telecom Complaints data.csv')

View(tele)

str(tele)

names(tele)

sum(is.na(tele))

Output:

**Insights:** This shows that there is no missing values in dataset, so now data is tidy and available to process further or do EDA based on requirement. • Processing Date.

#converting the required variables as factor

tele$Received.Via=as.factor(tele$Received.Via)

tele$City=as.factor(tele$City)

tele$State=as.factor(tele$State)

tele$Status=as.factor(tele$Status)

tele$Filing.on.Behalf.of.Someone=as.factor(tele$Filing.on.Behalf.of.Someone)

#converting the date column to in date format

library(lubridate)

tele$Date<-dmy(tele$Date)

#Creating the chart for the number of complaints at monthly and daily granularity levels.

library(dplyr)

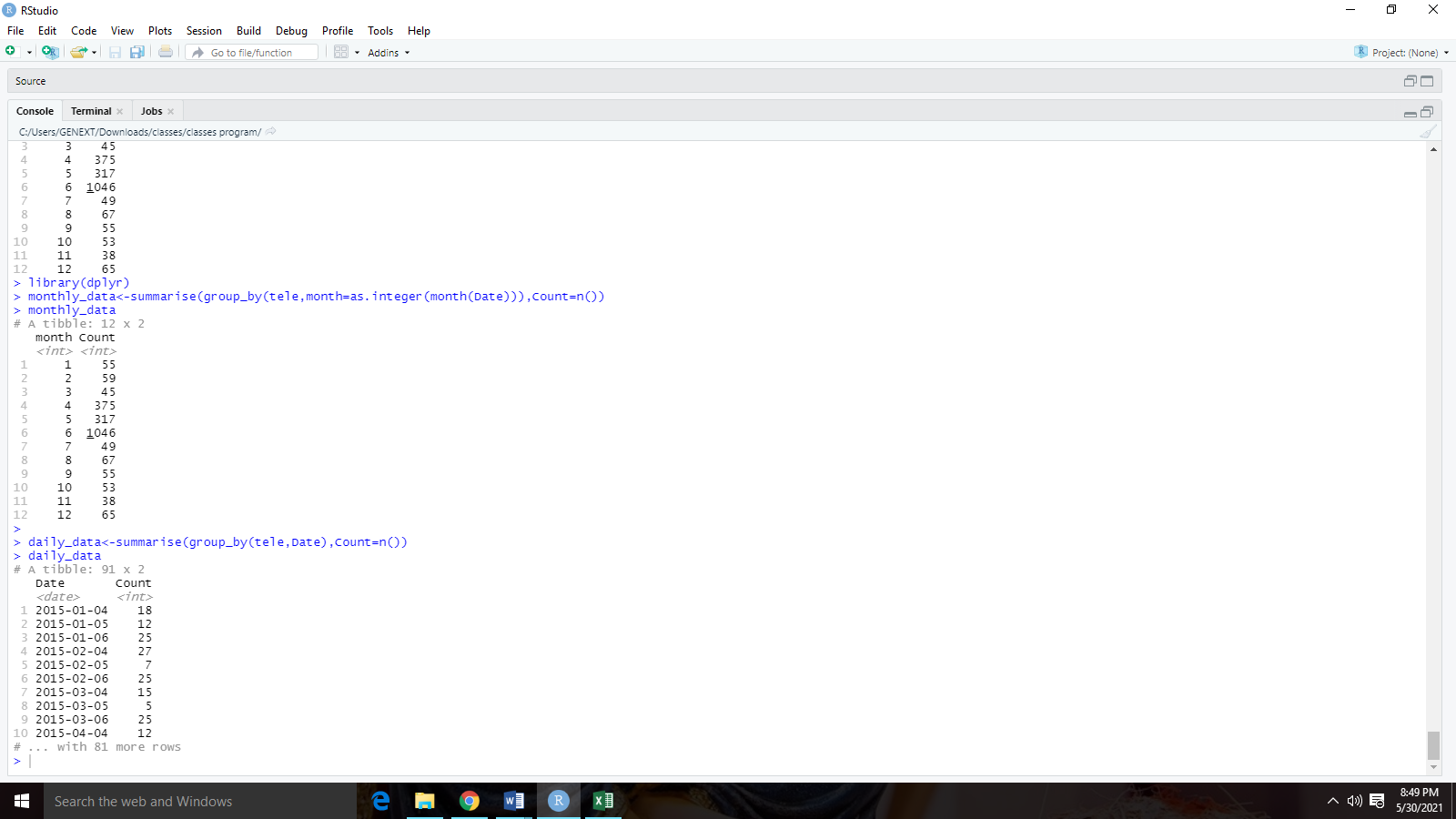
monthly\_data<-summarise(group\_by(tele,month=as.integer(month(Date))),Count=n())

monthly\_data

daily\_data<-summarise(group\_by(tele,Date),Count=n())

daily\_data

**Output:**



library(ggplot2)

ggplot(monthly\_data,aes(x=month,y=Count,label=Count)) + geom\_line(colour='blue',linetype=1)+

geom\_text()+

scale\_x\_continuous(breaks = monthly\_data$month)+

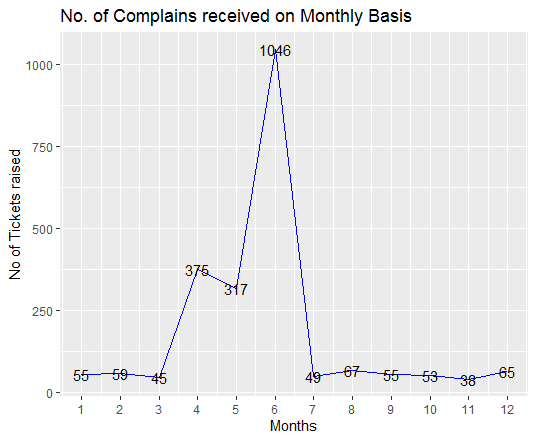
labs(title = "No. of Complains received on Monthly Basis",x='Months',y="No of Tickets raised")

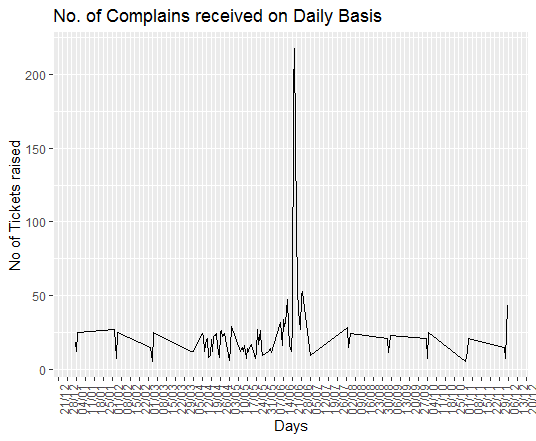
ggplot(daily\_data,aes(x=as.POSIXct(Date),y=Count))+geom\_line()+

theme(axis.text.x =element\_text(angle=90) )+

scale\_x\_datetime(breaks = "1 weeks",date\_labels='%d/%m')+

labs(title ='No. of Complains received on Daily Basis',x='Days',y="No of Tickets raised" )

Output as graphs:

Insights: As we can clearly see from the graph that the complaint was increased drastically in the month of June due to some reason.

Insight: With the help of the daily report chart it is clearly observed the complaint increased in the second half of the month of June.

#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

library(tidyverse)

network\_tickets<-contains(tele$Customer.Complaint,match='network',ignore.case = T)

internet\_tickets<-contains(tele$Customer.Complaint,match ='internet',ignore.case = T)

bill\_tickets<-contains(tele$Customer.Complaint,match='bill',ignore.case = T)

email\_tickets<-contains(tele$Customer.Complaint,match="email",ignore.case = T)

charge\_tickets<-contains(tele$Customer.Complaint,match='charge',ignore.case = T)

tele$ComplaintType[network\_tickets]<-'Network'

tele$ComplaintType[internet\_tickets]<-'Internet'

tele$ComplaintType[bill\_tickets]<-'Billing'

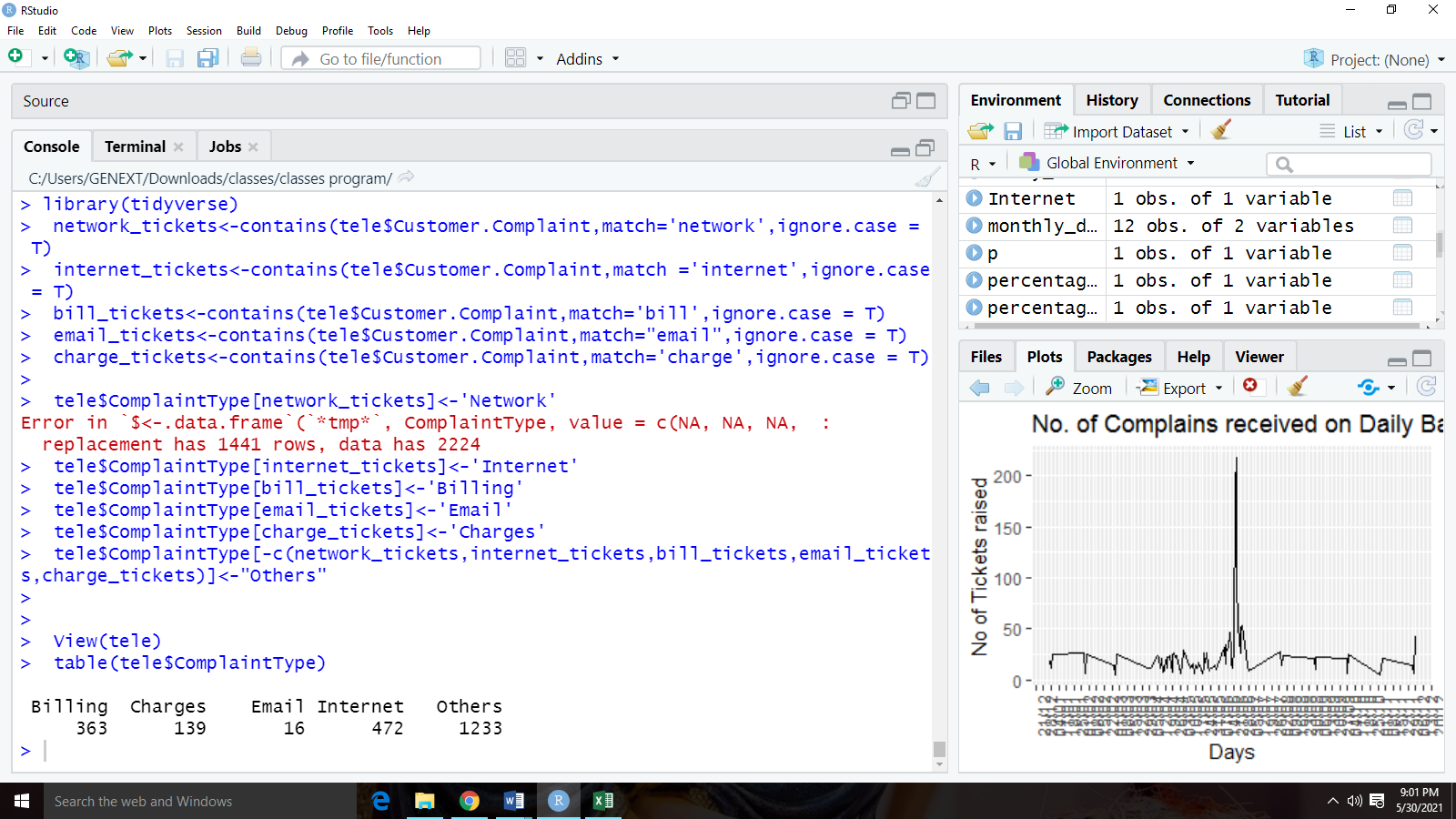
tele$ComplaintType[email\_tickets]<-'Email'

tele$ComplaintType[charge\_tickets]<-'Charges'

tele$ComplaintType[-c(network\_tickets,internet\_tickets,bill\_tickets,email\_tickets,charge\_tickets)]<-"Others"

View(tele)

table(tele$ComplaintType)

 Output:

Insight: As we can observe that there are some complaints from different-different categories and we combined them into one, i.e.- others. So most of the complaints are related to Internet issue.

#creating a new column for Open & Pending as Open and Closed & Solved as Closed.

tele$Complain\_Status<-ifelse(tele$Status == "Open"| tele$Status =="Pending",'Open','Closed')

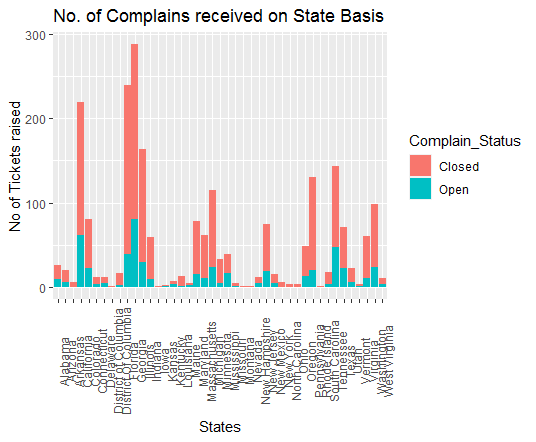
#creating the stacked bar for the no of complains received on the basis of cities

Complain<-summarise(group\_by(tele,State,Complain\_Status),Count=n())

ggplot(Complain,aes(x=State,y=Count,fill=Complain\_Status))+geom\_col()+

theme(axis.text.x = element\_text(angle = 90))+

labs(title ='No. of Complains received on State Basis',x='States',y="No of Tickets raised" )

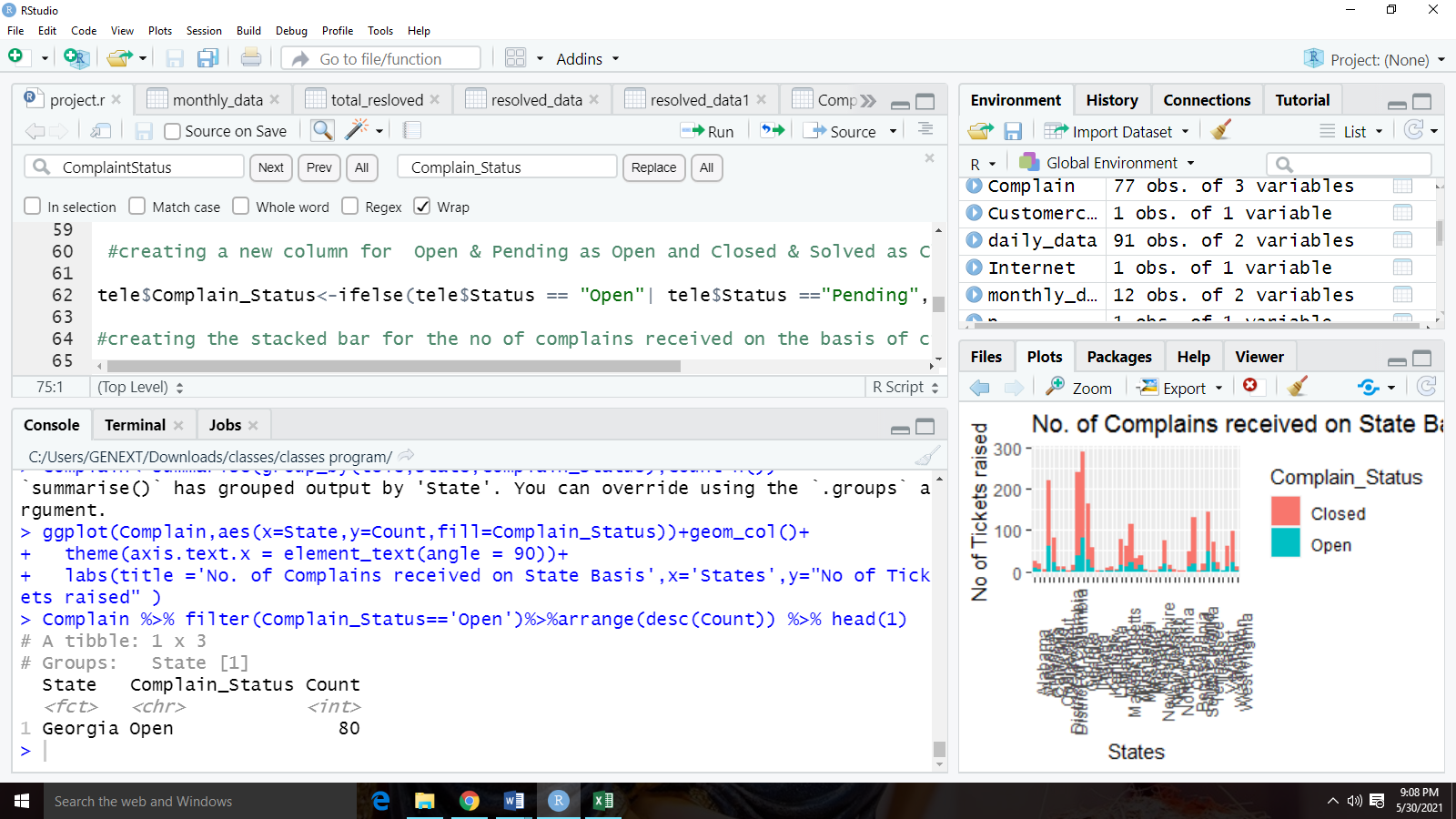
Output as graph:

Insight: Now it`s clearly shown that the highest number of complaints recorded from the state Georgia and the second highest number of complaints recorded from the state Florida.

#Calculating the State with highest unclosed\_data cases

Complain %>% filter(Complain\_Status=='Open')%>%arrange(desc(Count)) %>% head(1)

Output:



Insight: As seen from the table the maximum number of open cases is from the state of Georgia.

#Calculating the percentage of complaints closed\_data till date,

#which were received through the Internet and customer care calls.

resolved\_data <- group\_by(tele,Complain\_Status)

total\_resloved<- summarise(resolved\_data ,percentage =(n()/nrow(resolved\_data))\*100)

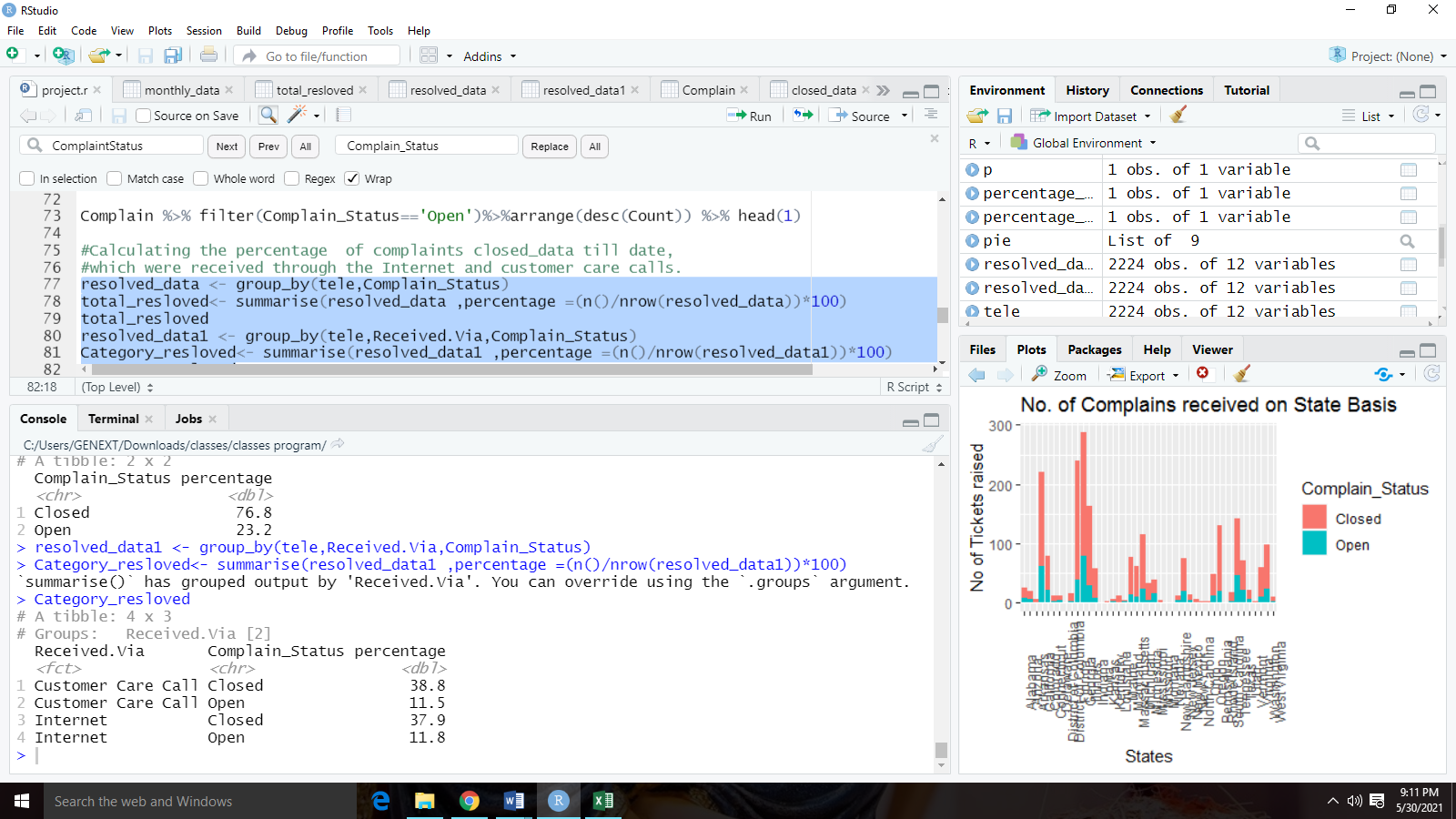
total\_resloved

resolved\_data1 <- group\_by(tele,Received.Via,Complain\_Status)

Category\_resloved<- summarise(resolved\_data1 ,percentage =(n()/nrow(resolved\_data1))\*100)

Category\_resloved

Output:



# Ploting Pie Chart for Total Resolved Vs Category Resolved

total<-ggplot(total\_resloved,

aes(x= "",y =percentage,fill = Complain\_Status))+

geom\_bar(stat = "identity",width = 1)+

coord\_polar("y",start = 0)+

geom\_text(aes(label = paste0(round(percentage),"%")),

position = position\_stack(vjust = 0.5))+

labs(x = NULL,y = NULL,fill = NULL)+

theme\_classic()+theme(axis.line = element\_blank(),

axis.text = element\_blank(),

axis.ticks = element\_blank())

total

# Pie Chart for Category wise Ticket Status

category<-ggplot(Category\_resloved,

aes(x= "",y =percentage,fill = Received.Via))+

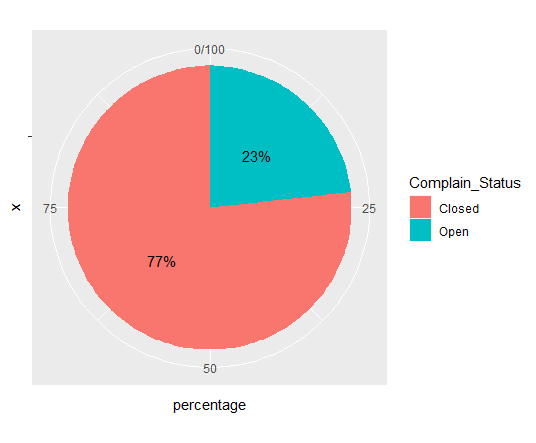
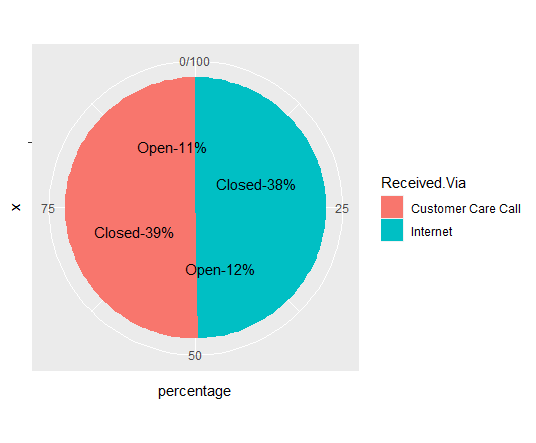
geom\_bar(stat = "identity",width = 1)+

coord\_polar("y")+

geom\_text(aes(label = paste0(Complain\_Status,"-",round(percentage),"%")),

position = position\_stack(vjust = 0.5))

category

Output as graphs:

With the help of above Chart of Total Resolved we can conclude that the total resolved complaints are 77% and open cases are 23%.From the Category Resolved the complaint received through Internet and Customer care call are same of 50%-50% in which closed case through Internet is 38% and open is 12% in case of Customer Care Call the closed cases 39% and open cases are 11%.

**Overall Insights From the Project:**

As per the above analysis we observe that in the 2nd half of the June month Comcast received high amount of complaints in which most of the complaints are related to internet service issue and the highest amount of complaints are received from the state Georgia. The highest unresolved complaints are related from the state Georgia and the total amount of resolved complaints are 77% in which 38% are received the internet and 39% are from the customer care calls.