Project for Data Science with R Programming – Project 2

Project Name: **Comcast Telecom Consumer Complaints.**

**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.

The analysis results to be provided with insights wherever applicable.

Observations after my analysis:

There are more no of complaints during the month of June.

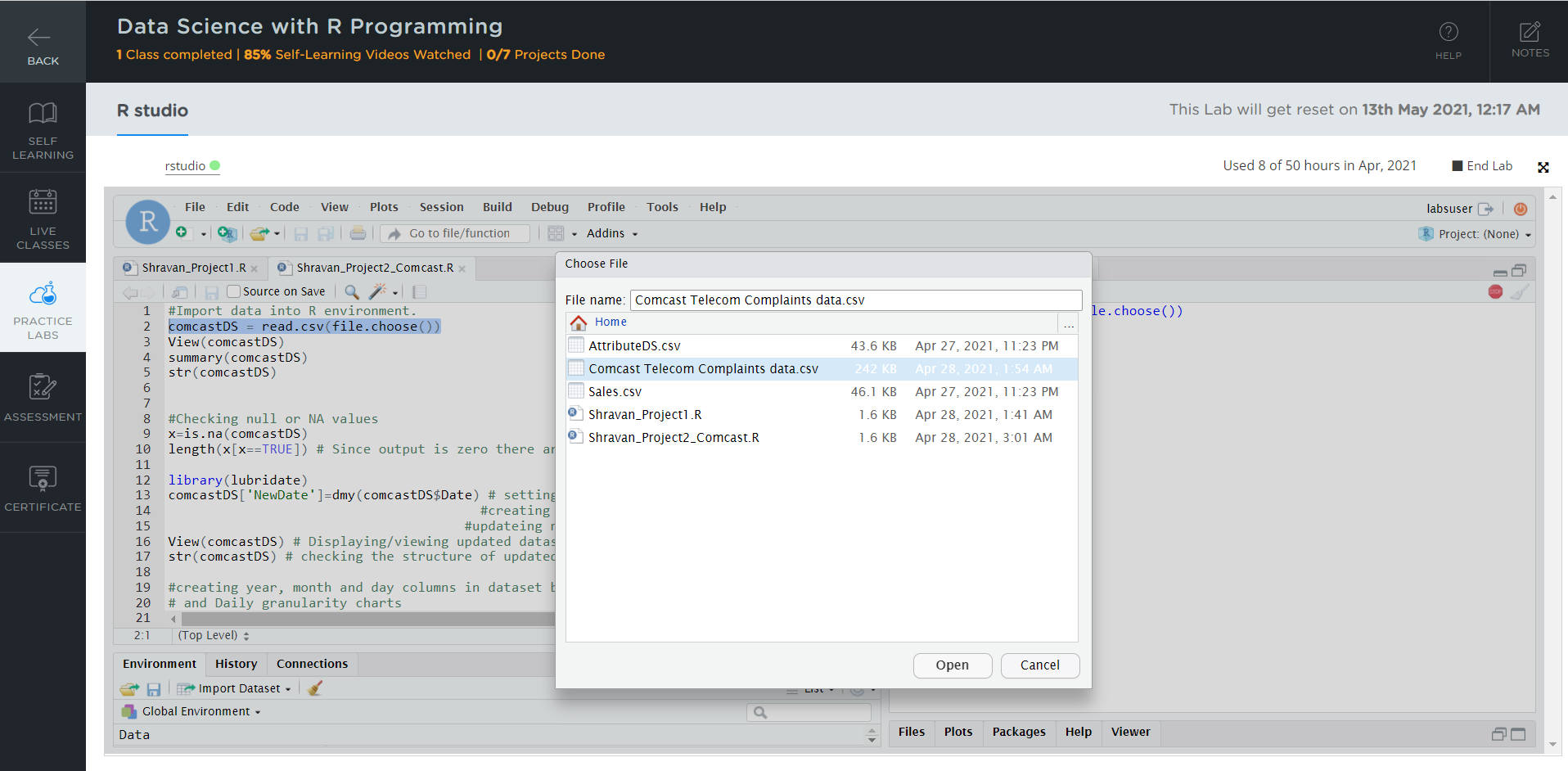
Majority of complaints are on data caps, Internet, billing and services.

Majority of complaints are in Georgia and California.

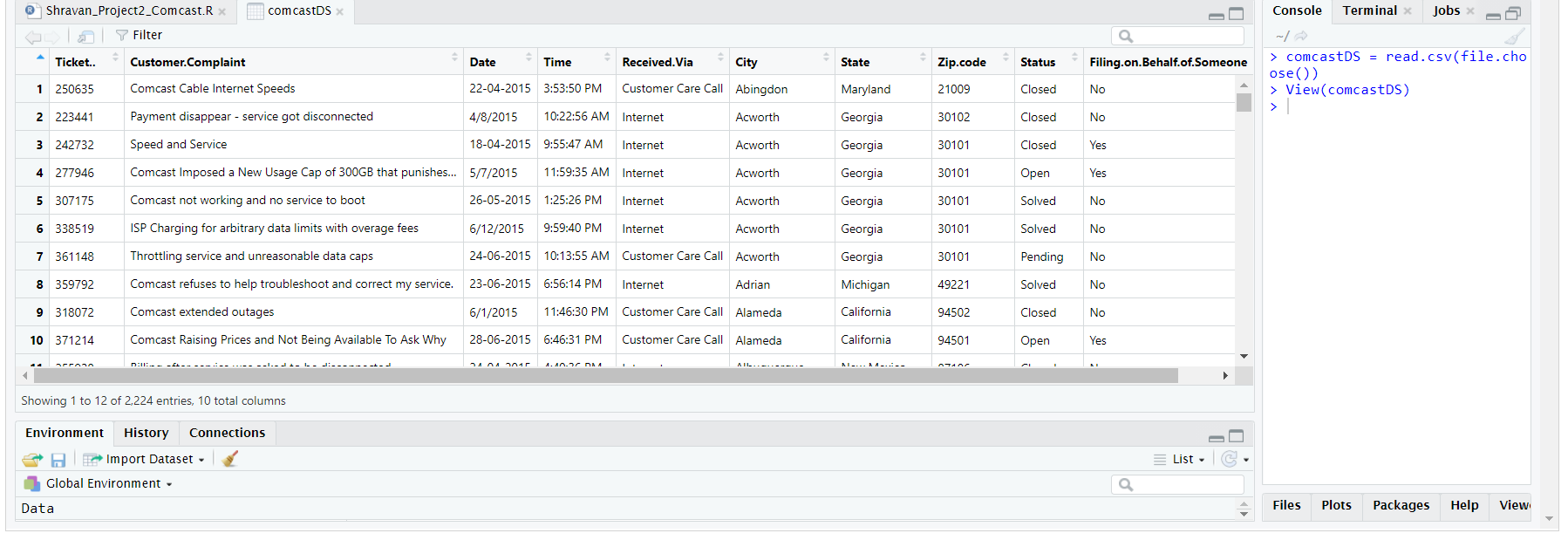
Most of the complaints are coming from customer care calls rather than internet.

There are still 255 complaints open in customer care call category and 262 open complaints in Internet category.

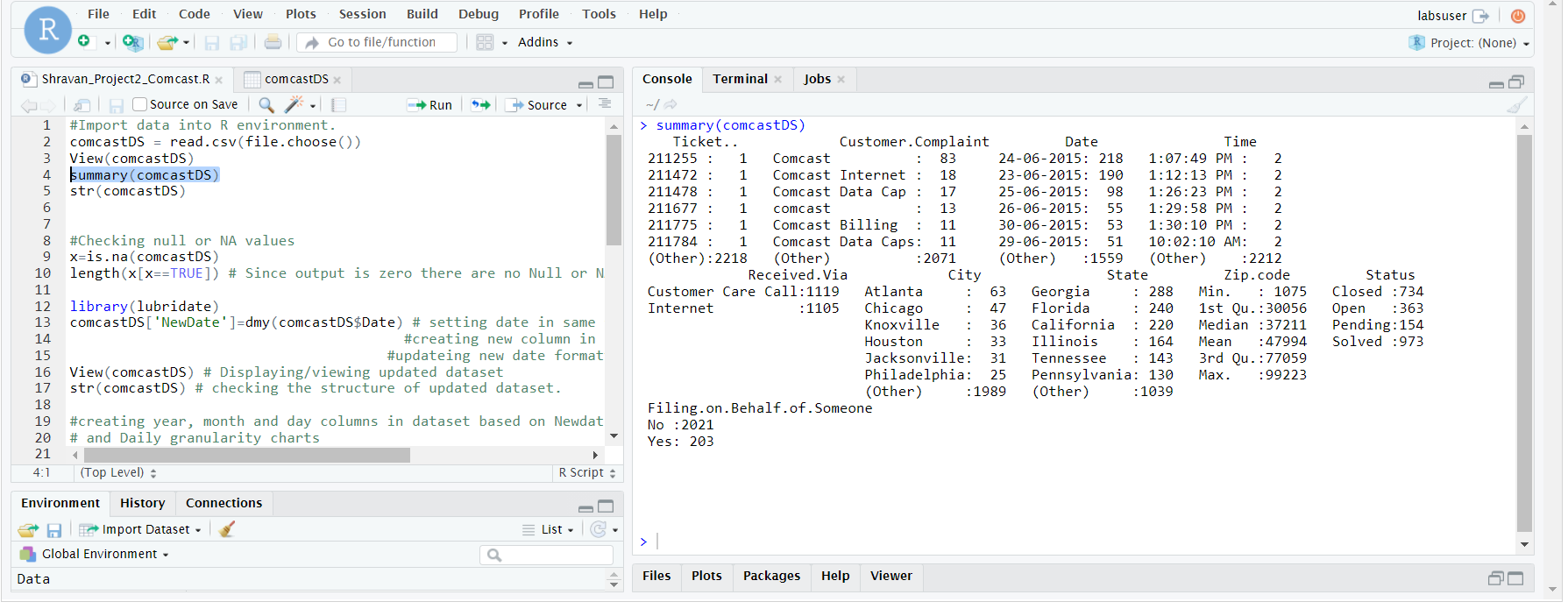
Import data into R environment.



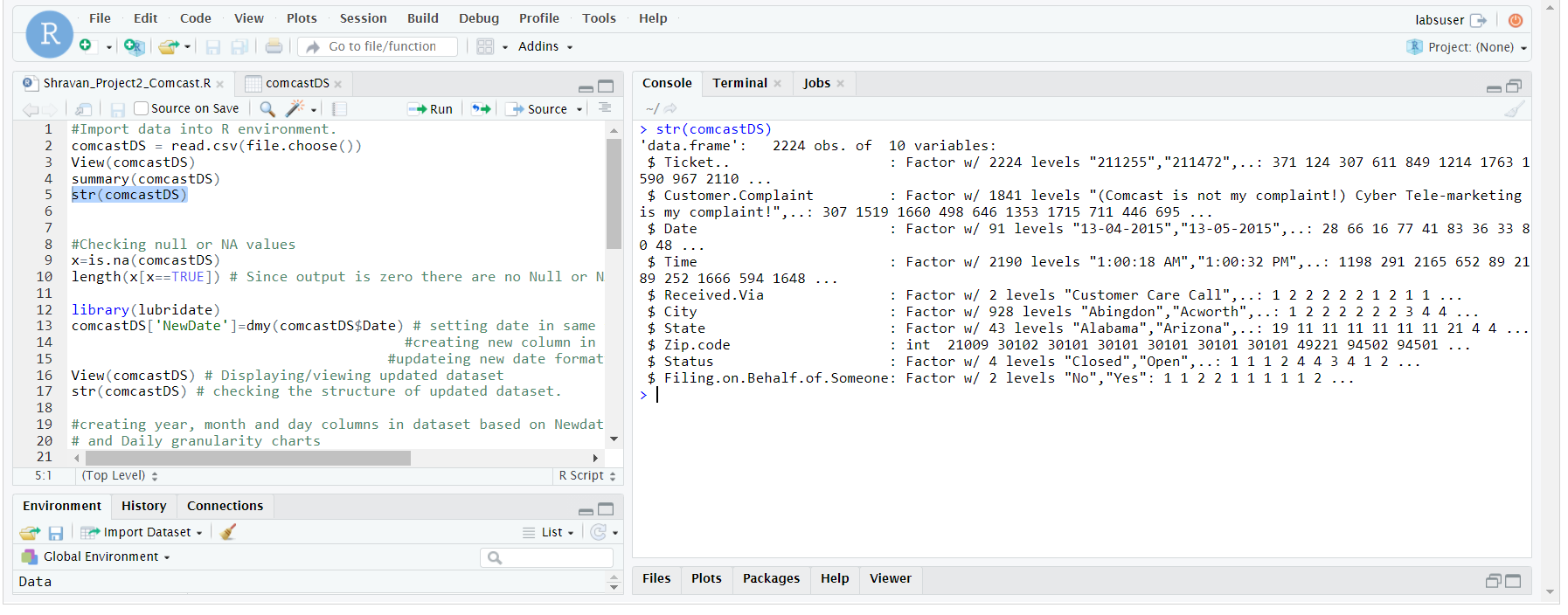
Initial view of dataset after importing csv file into R studio



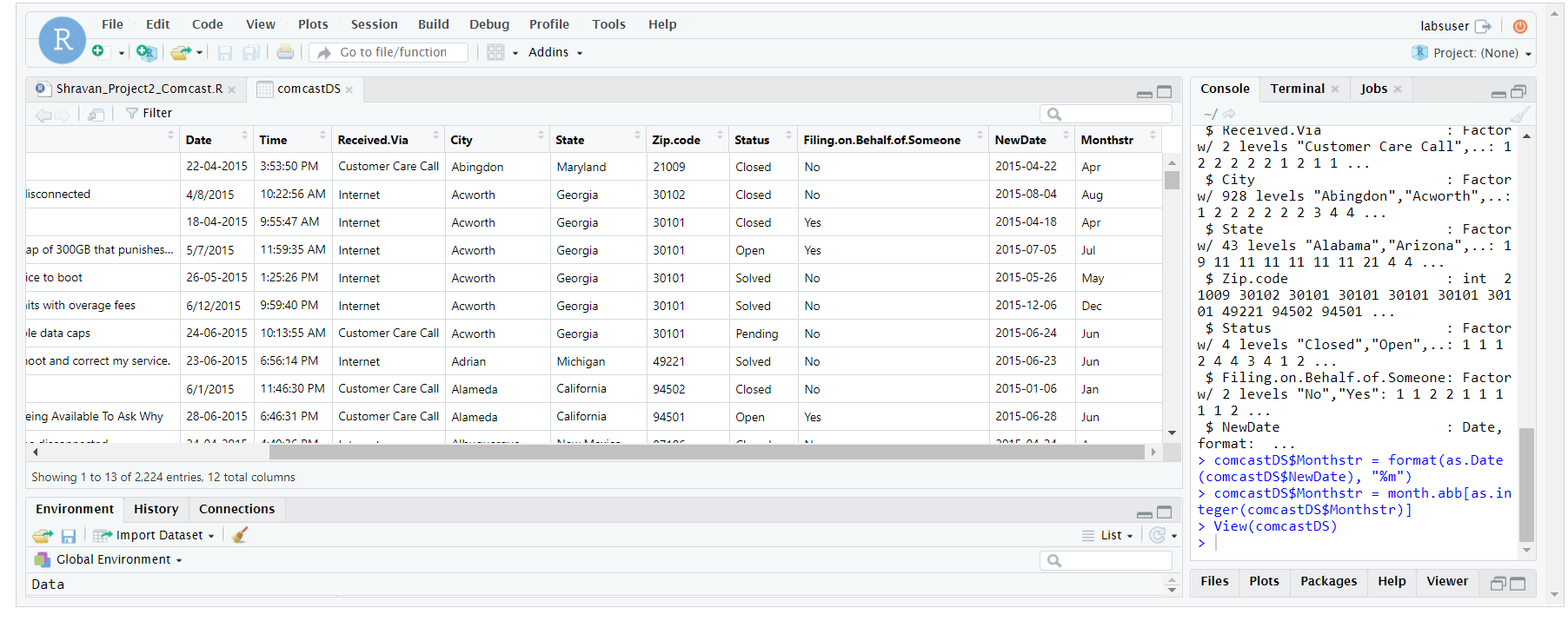
Summary



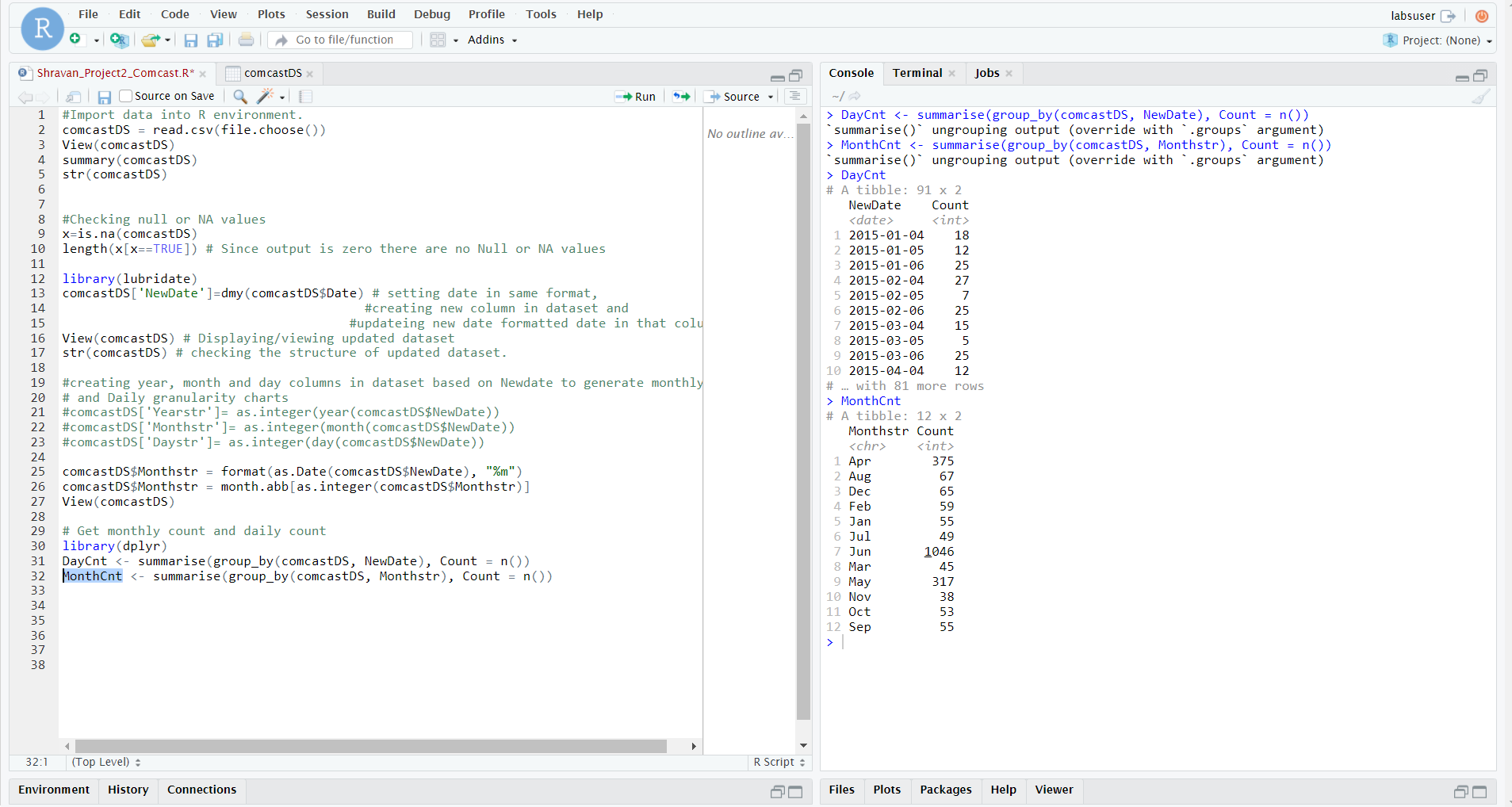
Structure



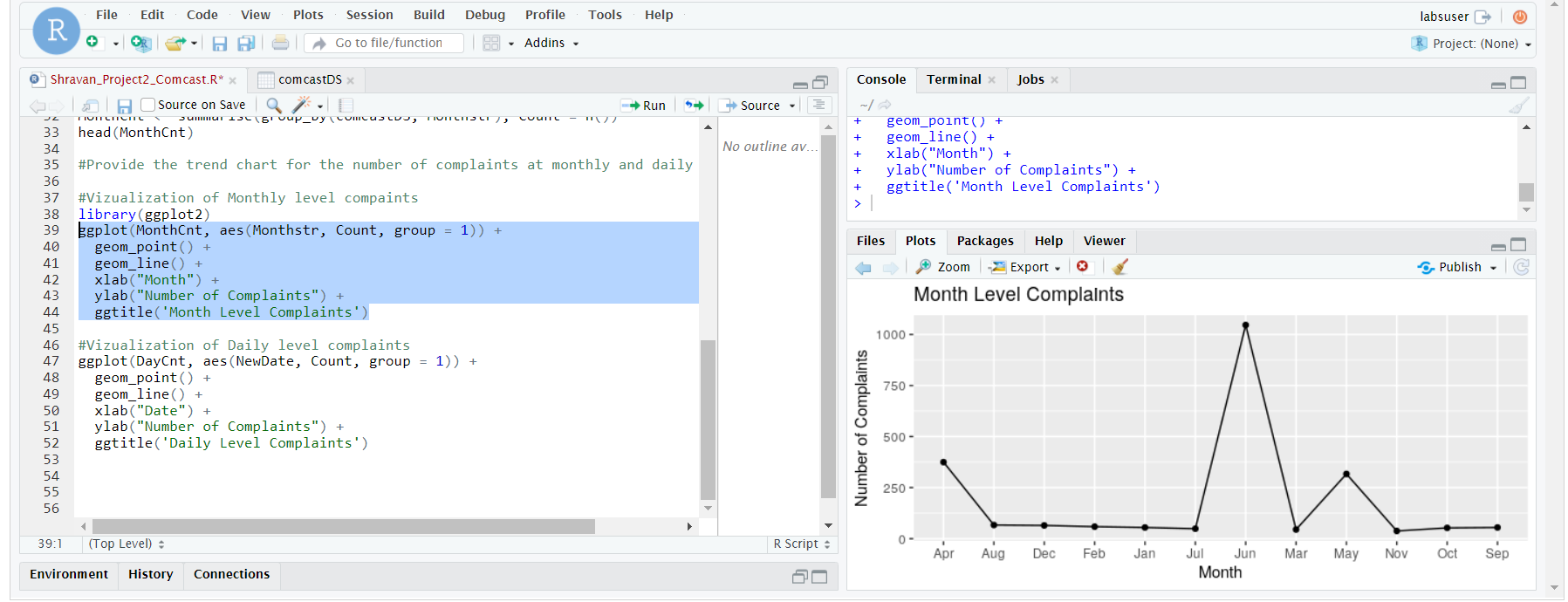
Screenshot of updated dataset after adding formatted date and month columns



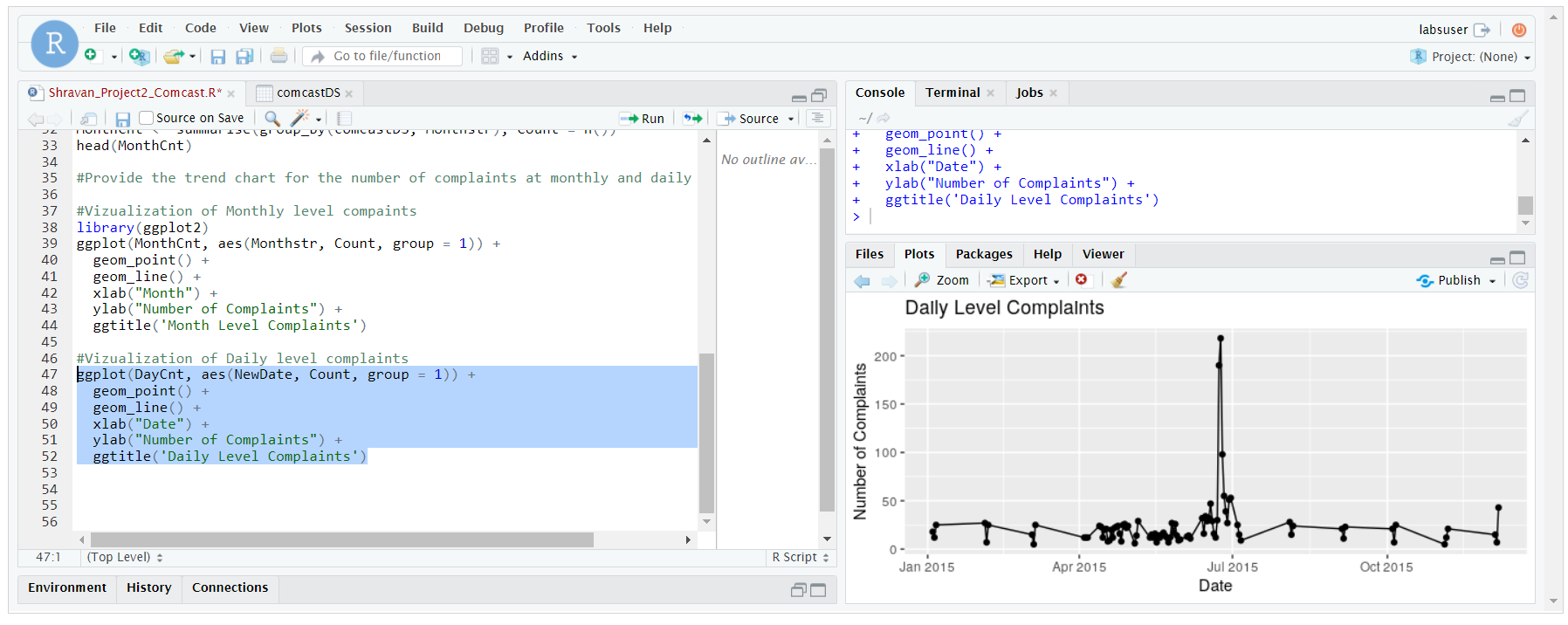
Below screenshot shows Daily count and Monthly count of complaints



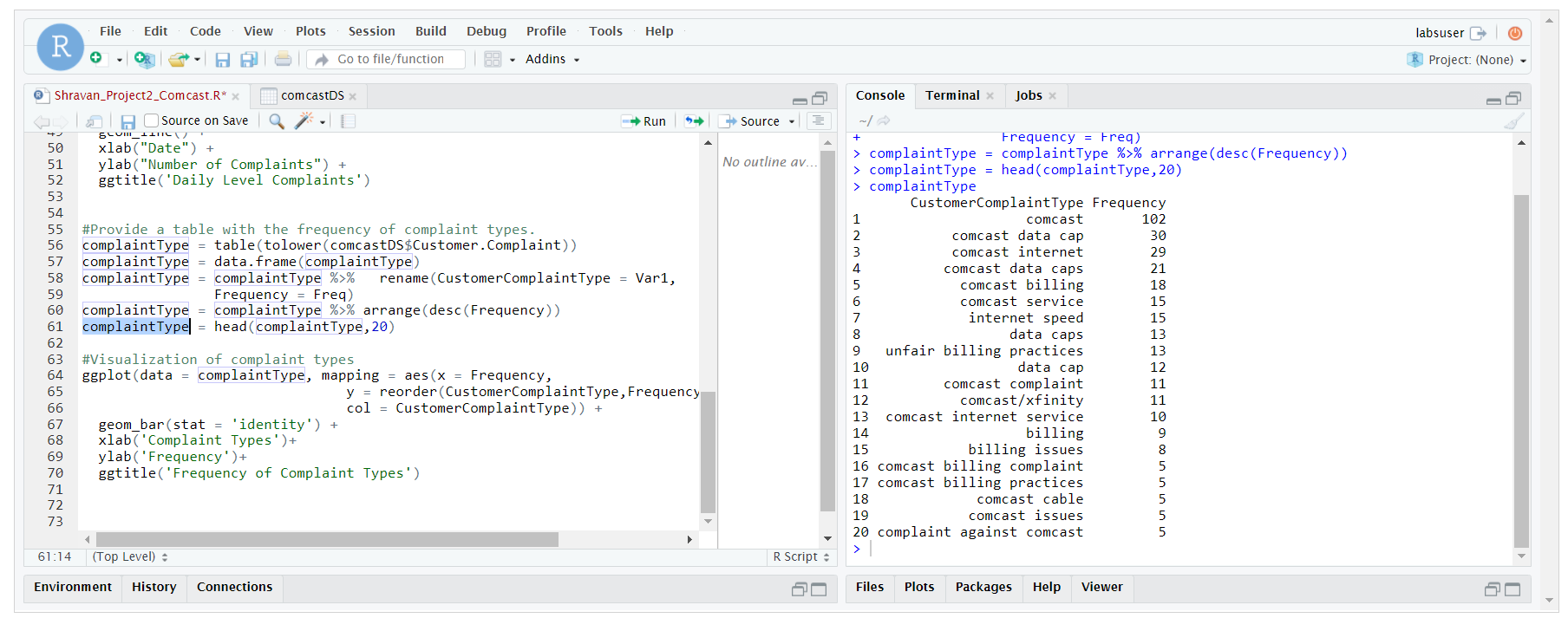
Graph to display monthly level complaints



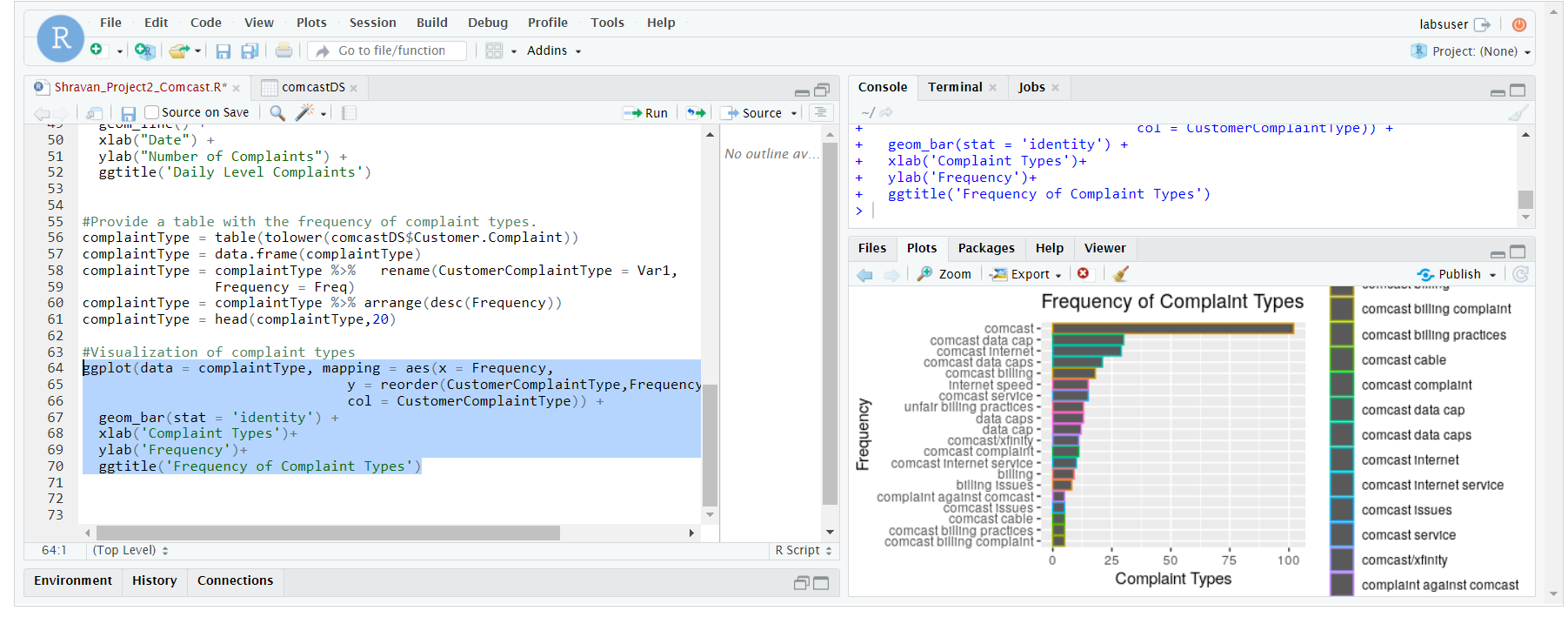
Graph to display Daily level complaints



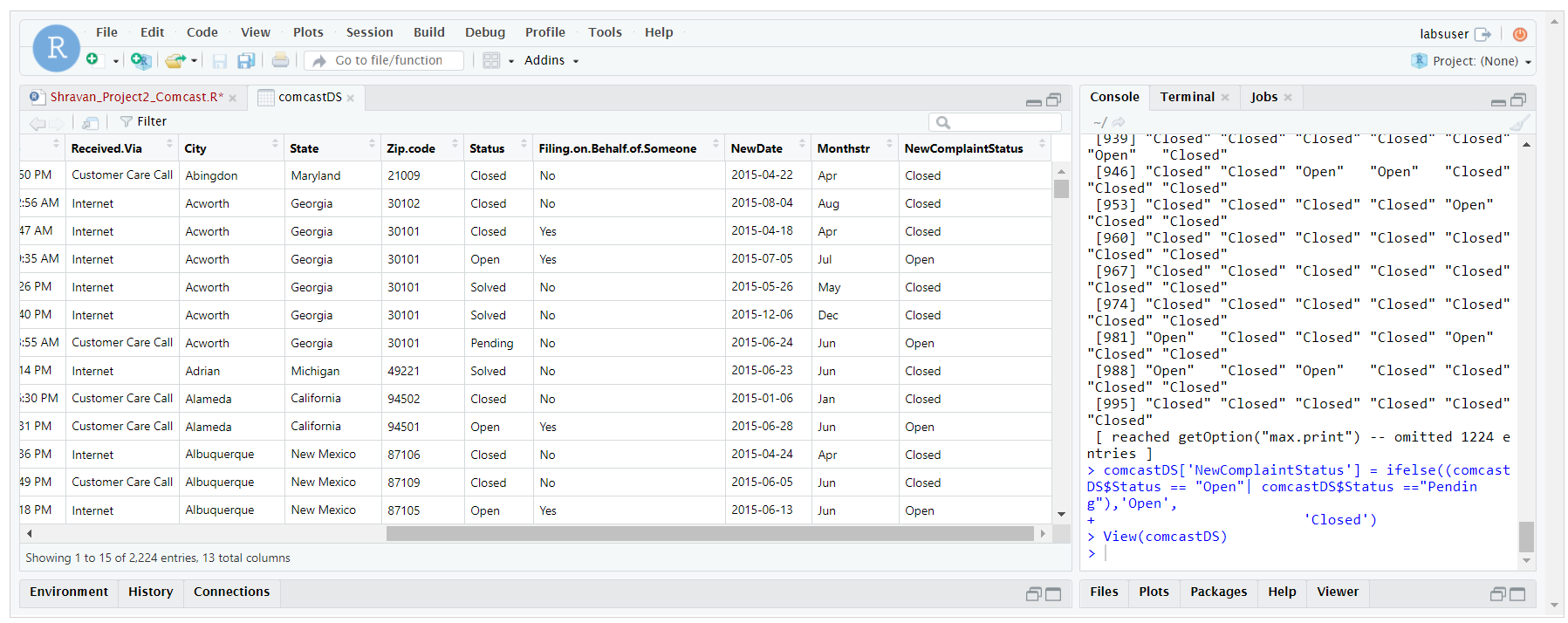
Provide a table with the frequency of complaint types.



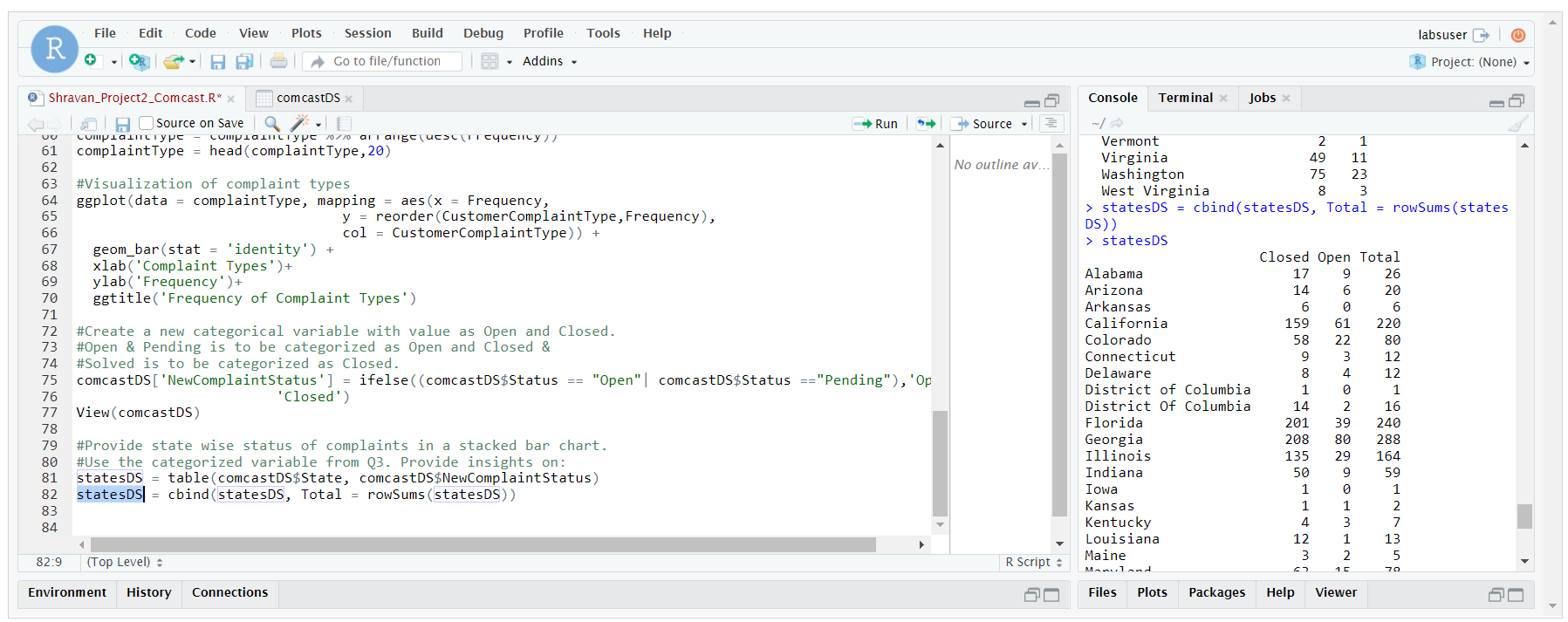
Visualization of Complaint Types

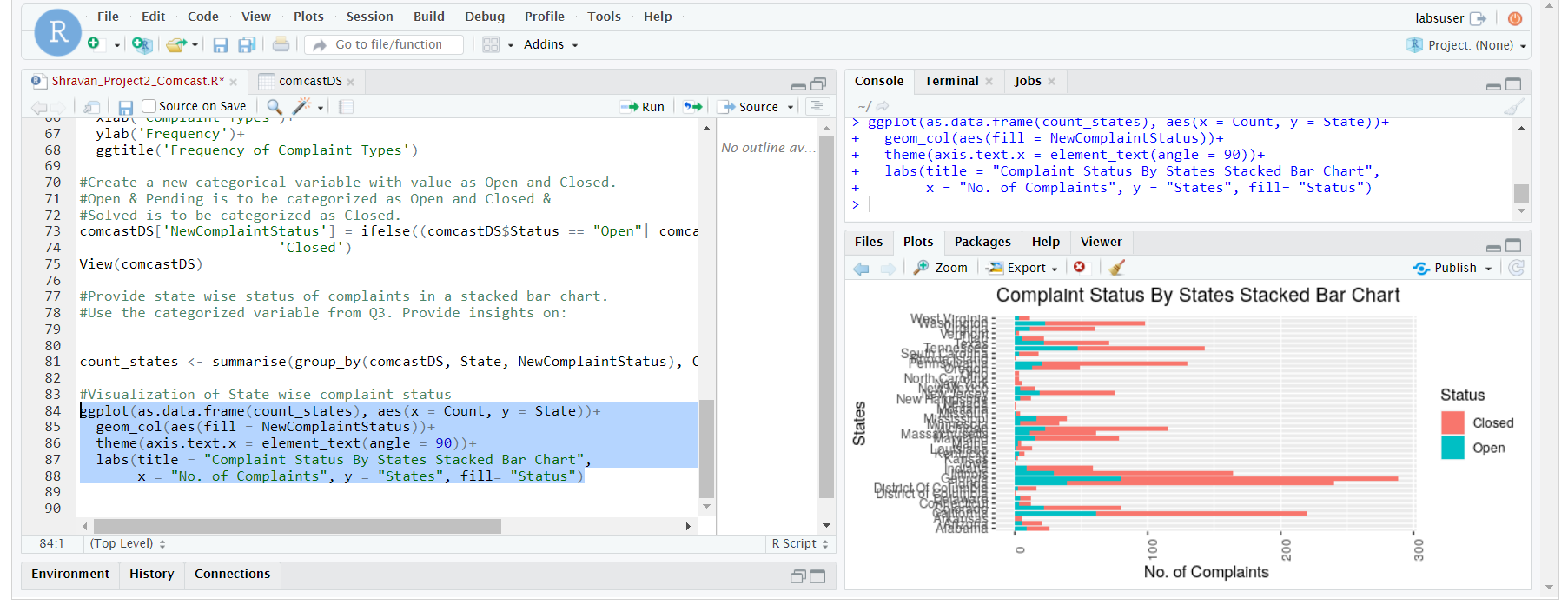


Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.

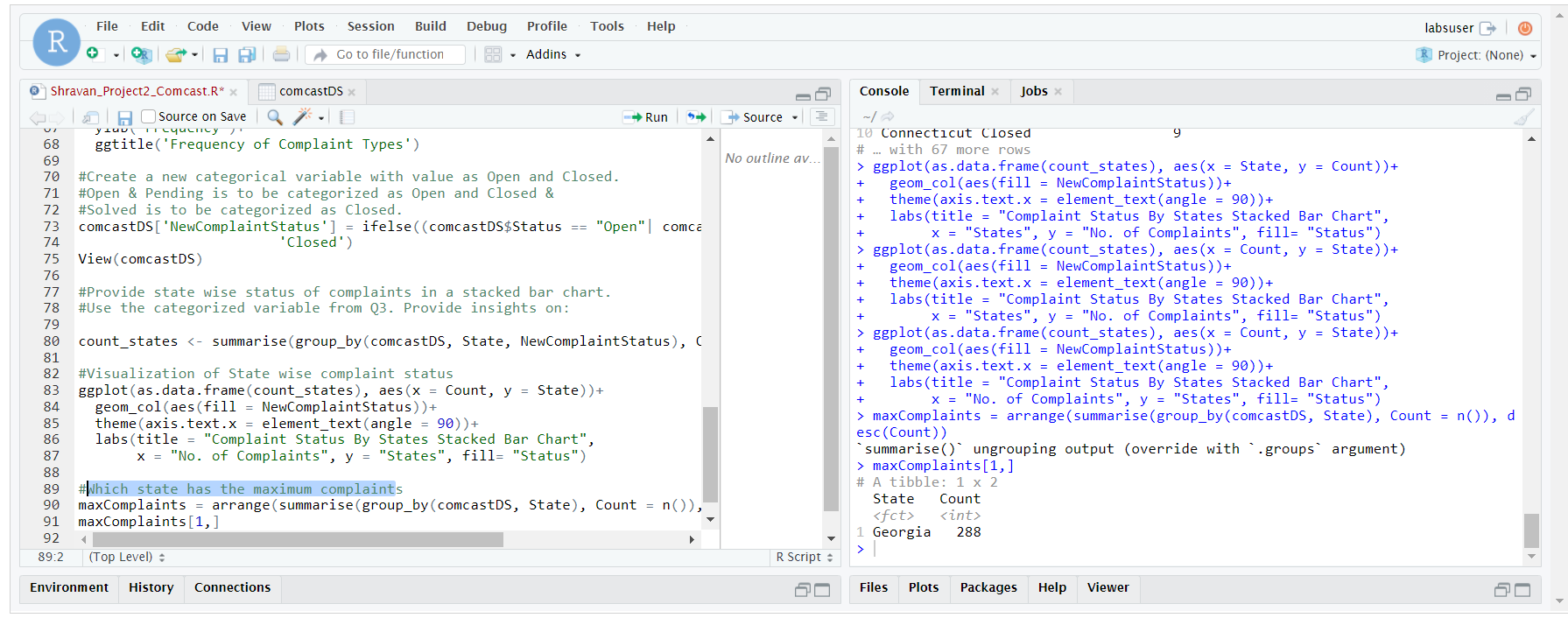


Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on

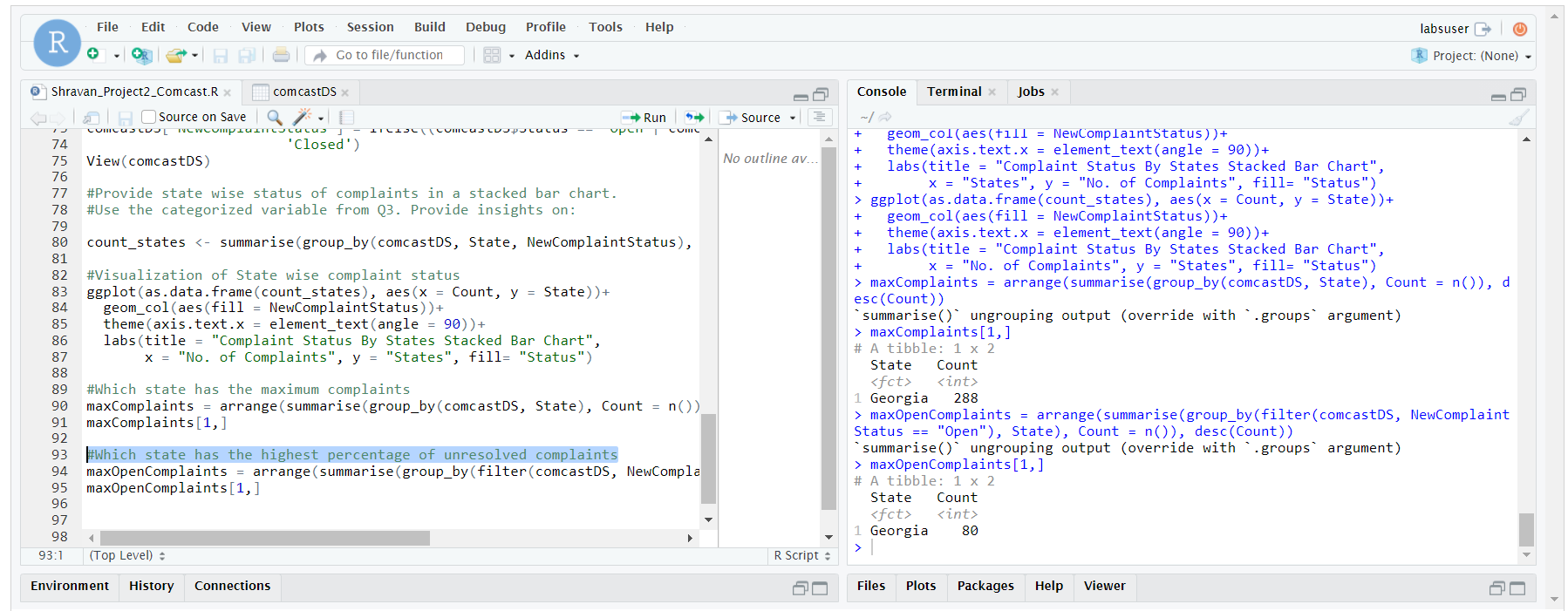


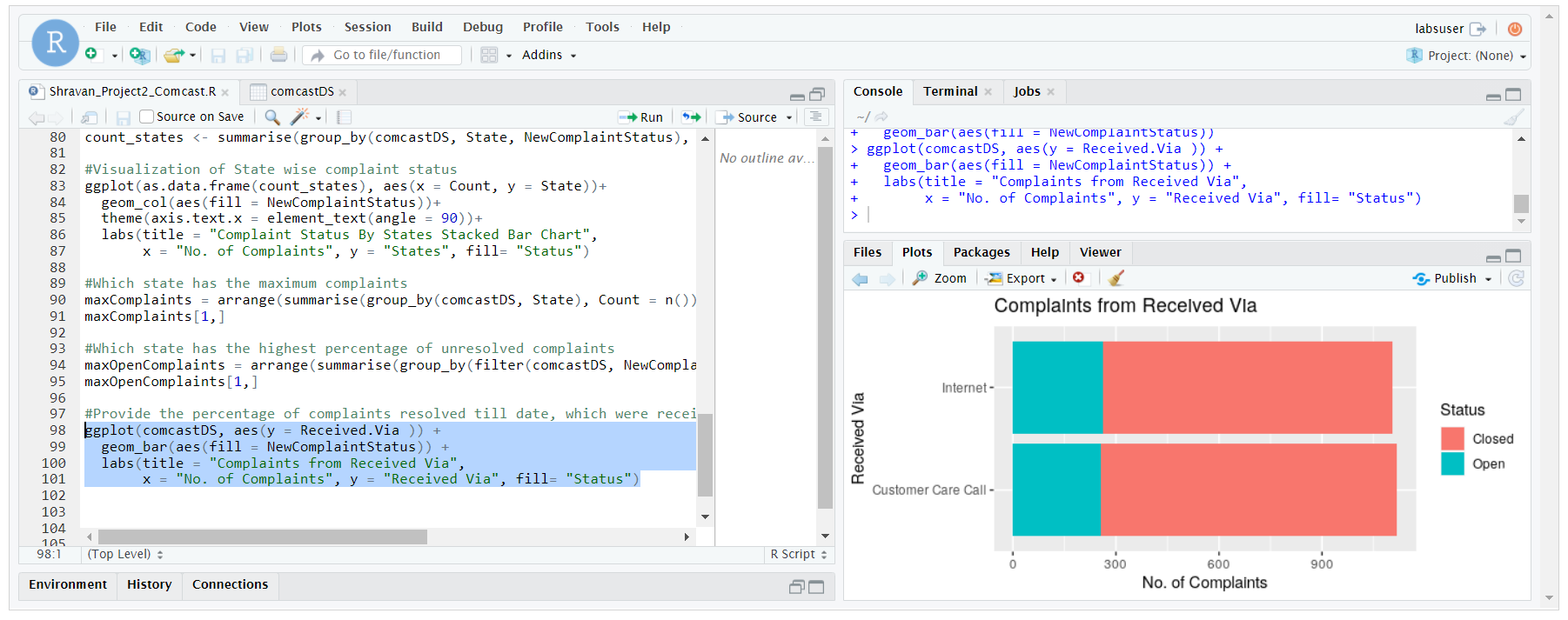


Which state has the maximum complaint



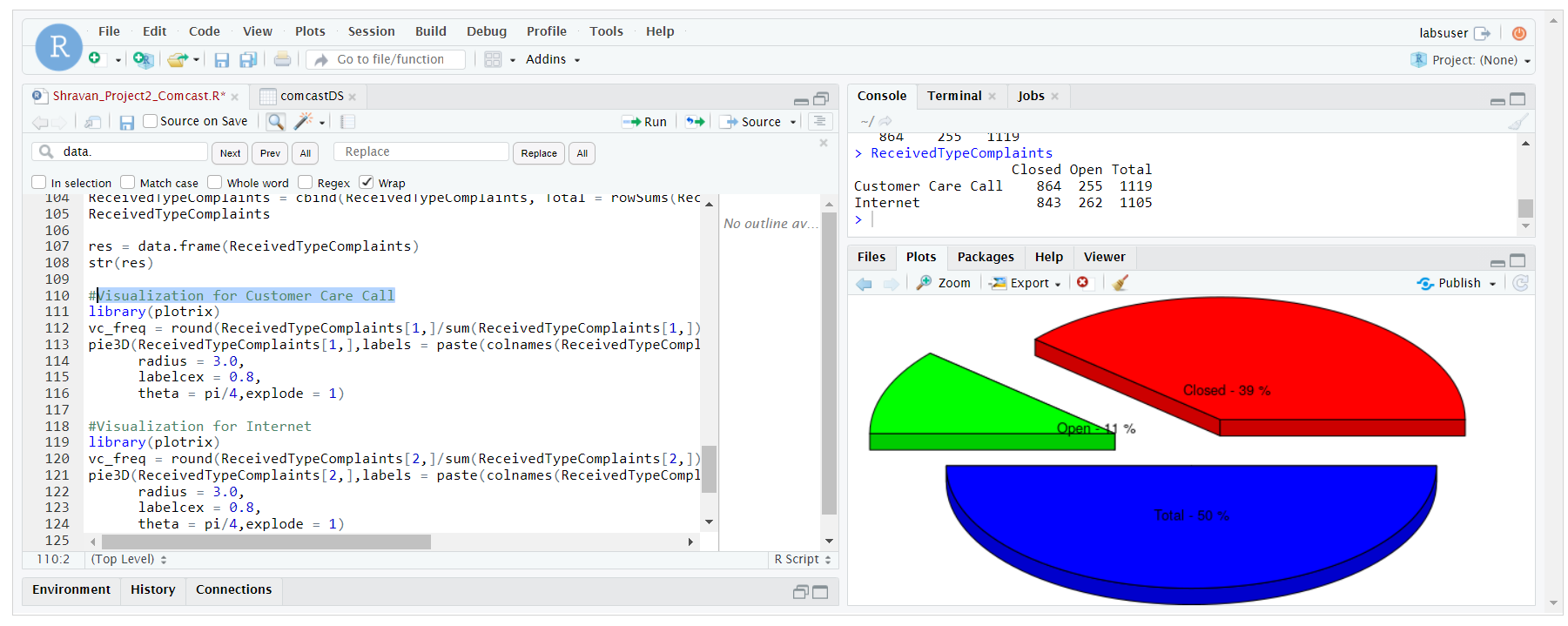
Which state has the highest percentage of unresolved complaints



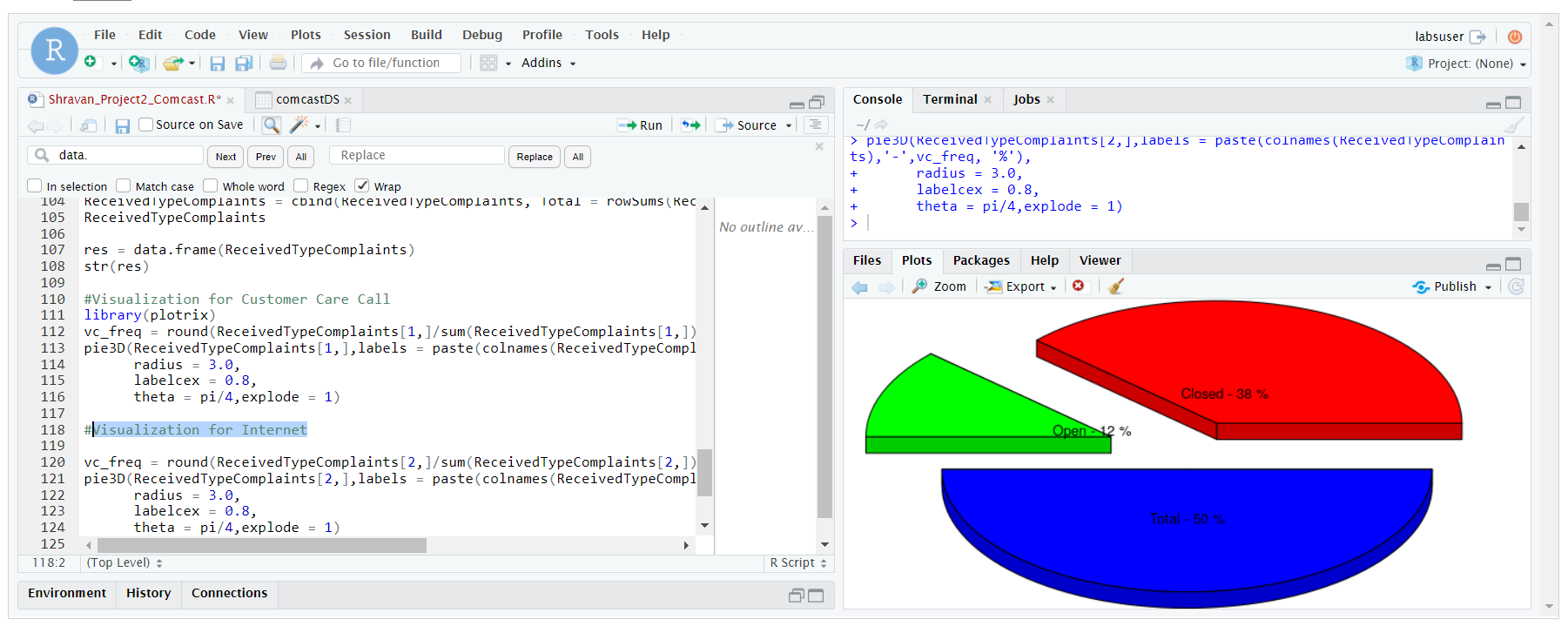
Provide the percentage of complaints resolved till date, which were received through the 

Internet and customer care calls.

Visualization for Customer Care Call



Visualization for Internet



Code in R Studio

#Import data into R environment.

comcastDS = read.csv(file.choose())

View(comcastDS)

summary(comcastDS)

str(comcastDS)

#Checking null or NA values

x=is.na(comcastDS)

length(x[x==TRUE]) # Since output is zero there are no Null or NA values

library(lubridate)

comcastDS['NewDate']=dmy(comcastDS$Date) # setting date in same format,

#creating new column in dataset and

#updateing new date formatted date in that column

View(comcastDS) # Displaying/viewing updated dataset

str(comcastDS) # checking the structure of updated dataset.

#creating year, month and day columns in dataset based on Newdate to generate monthly

# and Daily granularity charts

comcastDS$Monthstr = format(as.Date(comcastDS$NewDate), "%m")

comcastDS$Monthstr = month.abb[as.integer(comcastDS$Monthstr)]

View(comcastDS)

# Get monthly count and daily count

library(dplyr)

DayCnt <- summarise(group\_by(comcastDS, NewDate), Count = n())

MonthCnt <- summarise(group\_by(comcastDS, Monthstr), Count = n())

head(MonthCnt)

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

#Visualization of Monthly level compaints

library(ggplot2)

ggplot(MonthCnt, aes(Monthstr, Count, group = 1)) +

geom\_point() +

geom\_line() +

xlab("Month") +

ylab("Number of Complaints") +

ggtitle('Month Level Complaints')

#Visualization of Daily level complaints

ggplot(DayCnt, aes(NewDate, Count, group = 1)) +

geom\_point() +

geom\_line() +

xlab("Date") +

ylab("Number of Complaints") +

ggtitle('Daily Level Complaints')

#Provide a table with the frequency of complaint types.

complaintType = table(tolower(comcastDS$Customer.Complaint))

complaintType = data.frame(complaintType)

complaintType = complaintType %>% rename(CustomerComplaintType = Var1,

Frequency = Freq)

complaintType = complaintType %>% arrange(desc(Frequency))

complaintType = head(complaintType,20)

#Visualization of complaint types

ggplot(data = complaintType, mapping = aes(x = Frequency,

y = reorder(CustomerComplaintType,Frequency),

col = CustomerComplaintType)) +

geom\_bar(stat = 'identity') +

xlab('Complaint Types')+

ylab('Frequency')+

ggtitle('Frequency of Complaint Types')

#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.

comcastDS['NewComplaintStatus'] = ifelse((comcastDS$Status == "Open"| comcastDS$Status =="Pending"),'Open',

'Closed')

View(comcastDS)

#Provide state wise status of complaints in a stacked bar chart.

#Use the categorized variable from Q3. Provide insights on:

count\_states <- summarise(group\_by(comcastDS, State, NewComplaintStatus), Count = n())

#Visualization of State wise complaint status

ggplot(as.data.frame(count\_states), aes(x = Count, y = State))+

geom\_col(aes(fill = NewComplaintStatus))+

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

labs(title = "Complaint Status By States Stacked Bar Chart",

x = "No. of Complaints", y = "States", fill= "Status")

#Which state has the maximum complaints

maxComplaints = arrange(summarise(group\_by(comcastDS, State), Count = n()), desc(Count))

maxComplaints[1,]

#Which state has the highest percentage of unresolved complaints

maxOpenComplaints = arrange(summarise(group\_by(filter(comcastDS, NewComplaintStatus == "Open"), State), Count = n()), desc(Count))

maxOpenComplaints[1,]

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

ggplot(comcastDS, aes(y = Received.Via )) +

geom\_bar(aes(fill = NewComplaintStatus)) +

labs(title = "Complaints from Received Via",

x = "No. of Complaints", y = "Received Via", fill= "Status")

ReceivedTypeComplaints = table(comcastDS$Received.Via, comcastDS$NewComplaintStatus)

ReceivedTypeComplaints = cbind(ReceivedTypeComplaints, Total = rowSums(ReceivedTypeComplaints),make.row.name = T)

ReceivedTypeComplaints

res = data.frame(ReceivedTypeComplaints)

str(res)

#Visualization for Customer Care Call

library(plotrix)

vc\_freq = round(ReceivedTypeComplaints[1,]/sum(ReceivedTypeComplaints[1,])\*100,0)

pie3D(ReceivedTypeComplaints[1,],labels = paste(colnames(ReceivedTypeComplaints),'-',vc\_freq, '%'),

radius = 3.0,

labelcex = 0.8,

theta = pi/4,explode = 1)

#Visualization for Internet

library(plotrix)

vc\_freq = round(ReceivedTypeComplaints[2,]/sum(ReceivedTypeComplaints[2,])\*100,0)

pie3D(ReceivedTypeComplaints[2,],labels = paste(colnames(ReceivedTypeComplaints),'-',vc\_freq, '%'),

radius = 3.0,

labelcex = 0.8,

theta = pi/4,explode = 1)

­