**Comcast Telecom Customer Complaints**

**Analysis:**

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.

**Code:**

**# Selecting the working Directory**

setwd("D:/Simplilearn/Project DataSets/2/1567503160\_comcasttelecomcomplaintsdata")

getwd()

**# Importing the dataset**

Comcast <- read.csv('Complaints.csv',head = TRUE,sep = ',')

View(Comcast)

str(Comcast)

**# Let us check if there is any missing data**

Comcastna <- is.na('Comcast')

length(Comcastna[Comcastna == T])

**# As per the results there is no missing values in the dataset**

library(lubridate)

Comcast$Date <- dmy(Comcast$Date)

**# Let us extract the monthly and daily count tickets**

library(dplyr)

monthly\_tickets <-

summarise(group\_by(Comcast,month = as.integer(month(Date))), count = n())

**# Let us remove the NA values**

monthly\_tickets <- na.omit(monthly\_tickets)

daily\_tickets <- summarise(group\_by(Comcast,Date),count = n())

daily\_tickets <- na.omit(daily\_tickets)

monthly\_tickets <- arrange(monthly\_tickets,month)

**# Let us plot the monthly and daily complaints to perform a comparison**

library(ggplot2)

**# Monthly tickets**

ggplot(data = monthly\_tickets,aes(month,count,label = count))+

geom\_line()+ geom\_point(size = 0.5)+geom\_text()+

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

labs(title = "Monthly Ticket Count",x= "Months",y ="No. of Tickets")+

theme(plot.title = element\_text(hjust = 0.5))

**# Daily Tickets**

ggplot(data = daily\_tickets,aes(as.POSIXct(Date),count))+

geom\_line()+geom\_point(size = 1)+

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

labs(title = "Daily Ticket Count",x= "Days",y ="No. of Tickets")+

theme(axis.text.x = element\_text(angle = 75),

plot.title = element\_text(hjust = 0.5))

**# Complaint Type Processing**

**# To find which types of Complaints are maximum**

network\_issues<- contains(Comcast$Customer.Complaint,match="network",ignore.case = T)

internet\_issues<-

contains(Comcast$Customer.Complaint,match = "internet",ignore.case = T)

billing\_issues<-

contains(Comcast$Customer.Complaint,match = "billing",ignore.case = T)

charges\_issues <-

contains(Comcast$Customer.Complaint,match = "charge", ignore.case = T)

email\_issues <-

contains(Comcast$Customer.Complaint,match = "email", ignore.case = T)

Comcast$ComplaintType[internet\_issues]<- "Internet"

Comcast$ComplaintType[network\_issues] <- "Network"

Comcast$ComplaintType[billing\_issues] <- "billing"

Comcast$ComplaintType[charges\_issues] <- "Charges"

Comcast$ComplaintType[email\_issues] <- "Email"

Comcast$ComplaintType[c(internet\_issues,network\_issues,billing\_issues,charges\_issues,email\_issues)] <- "Others"

table(Comcast$ComplaintType)

**# Let us create a new categorical variable with value as Open and Closed**

open\_complaints <- (Comcast$Status == "Open"| Comcast$Status=="Pending")

closed\_complaints <- (Comcast$Status=="Closed"| Comcast$Status=="Solved")

Comcast$ComplaintStatus[open\_complaints] <- "Open"

Comcast$ComplaintStatus[closed\_complaints] <- "Closed"

**# Even after changing it finding for NA values**

na\_vector <- is.na(Comcast)

length(na\_vector[na\_vector==T])

Comcast <- subset(Comcast,!is.na(Comcast$ComplaintStatus))

**# State wise Complaints in a stacked bar chart**

library(stringi)

library(ggpubr)

Comcast <- group\_by(Comcast,State,ComplaintStatus)

chart\_data <- summarise(Comcast,Count = n())

ggplot(as.data.frame(chart\_data), mapping = aes(State,Count))+

geom\_col(aes(fill = ComplaintStatus),width = 0.95)+

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

axis.title.y = element\_text(size = 15),

axis.title.x = element\_text(size = 15),

title = element\_text(size = 16,colour = "Red"),

plot.title = element\_text(hjust = 0.5))+

labs (title = "Ticket Status", x = "States",y = "No of Tickets",

fill= "Status")

**# To Provide the percentage of complaints resolved till date,**

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

**# To know Which state has the maximum complaints**

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

resolved <- group\_by(Comcast, ComplaintStatus)

total\_resolved <- summarise(resolved,percentage=(n()/nrow(resolved)))

resolved <- group\_by(Comcast,Received.Via,ComplaintStatus)

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

**# To visualize the results**

**# Let's plot this in a pie chart**

par(mfrow = c(1,2))

total <- ggplot(total\_resolved,

aes(x= "",y =percentage,fill = ComplaintStatus))+

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

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

geom\_text(aes(label = paste0(round(percentage\*100),"%")),

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

**# To get the visualized result of categorized wise Ticket status**

category<-ggplot(Category\_resloved,

aes(x= "",y =percentage,fill = ComplaintStatus))+

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

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

geom\_text(aes(label = paste0(Received.Via,"-",round(percentage\*100),"%")),

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

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

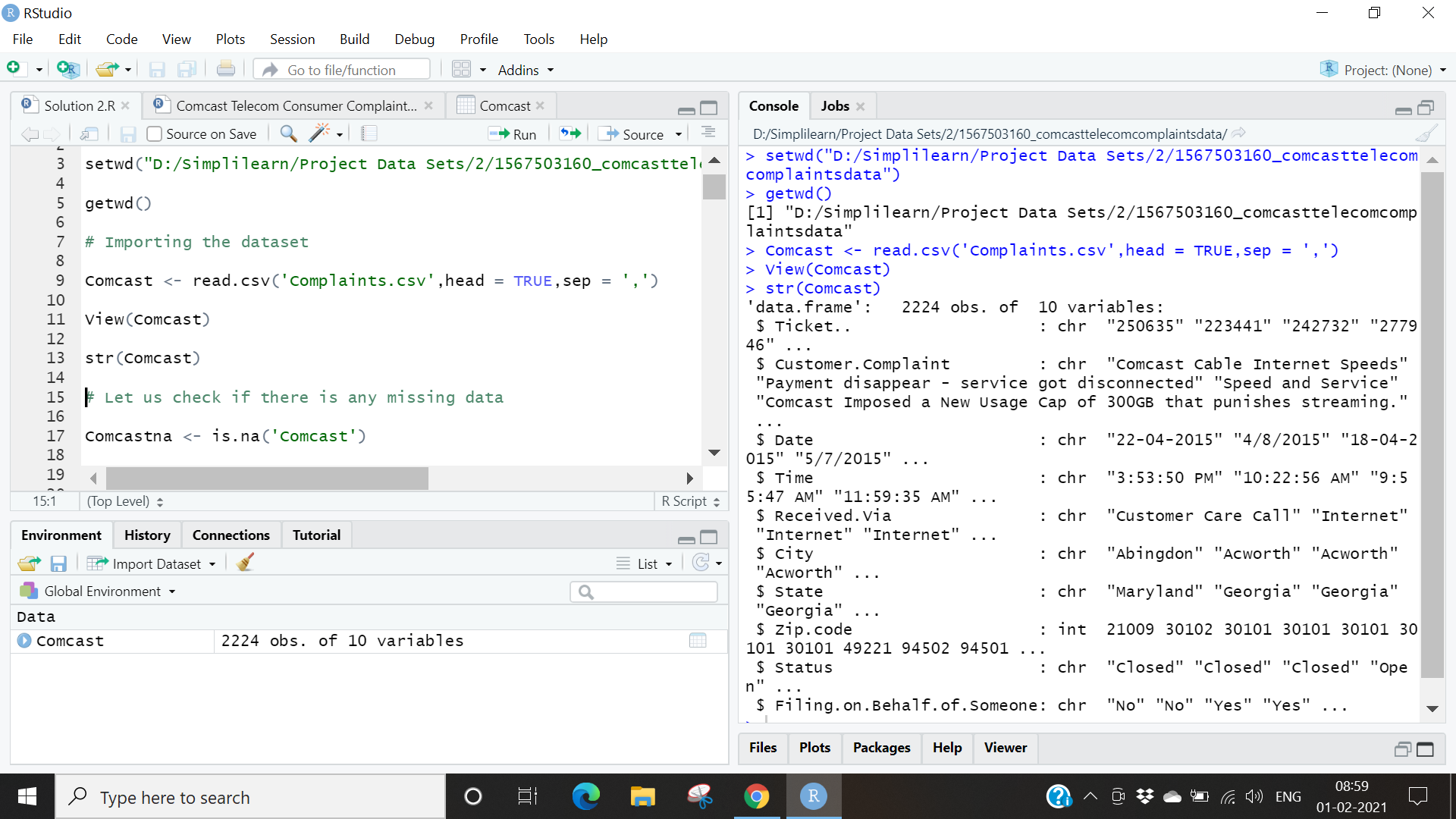
theme\_classic()+theme(axis.line = element\_blank(),

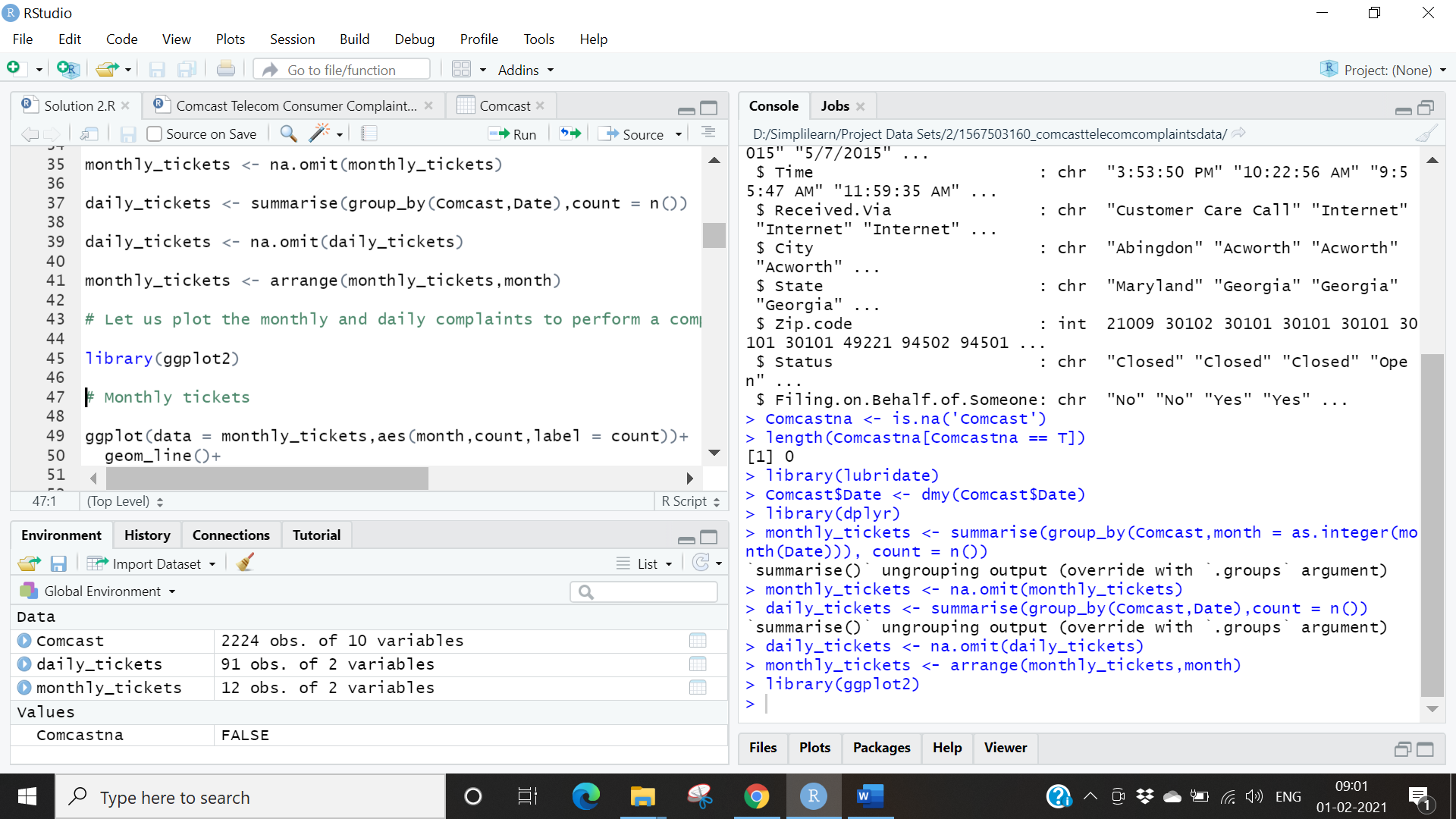
axis.text = element\_blank(),

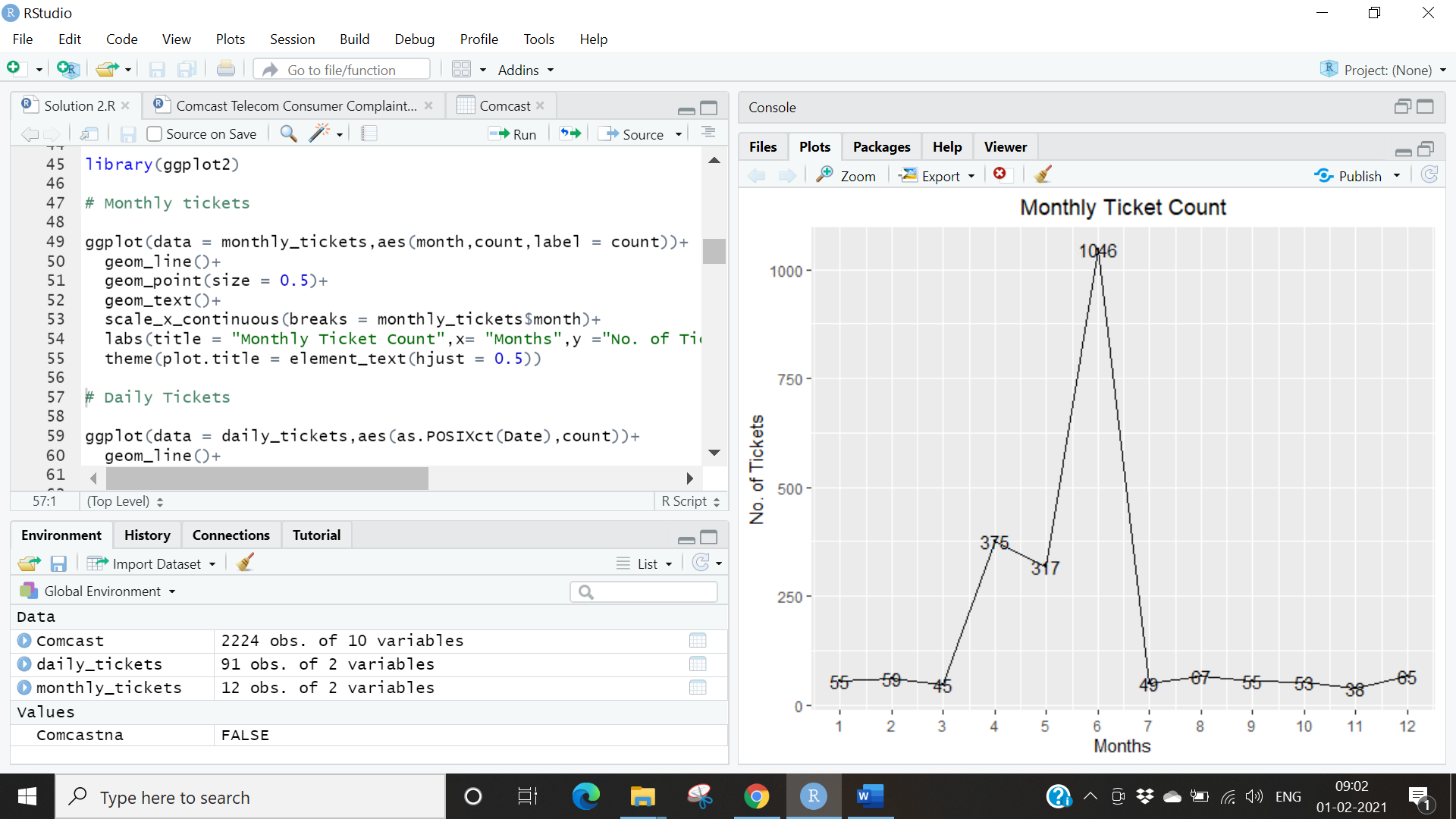
axis.ticks = element\_blank())

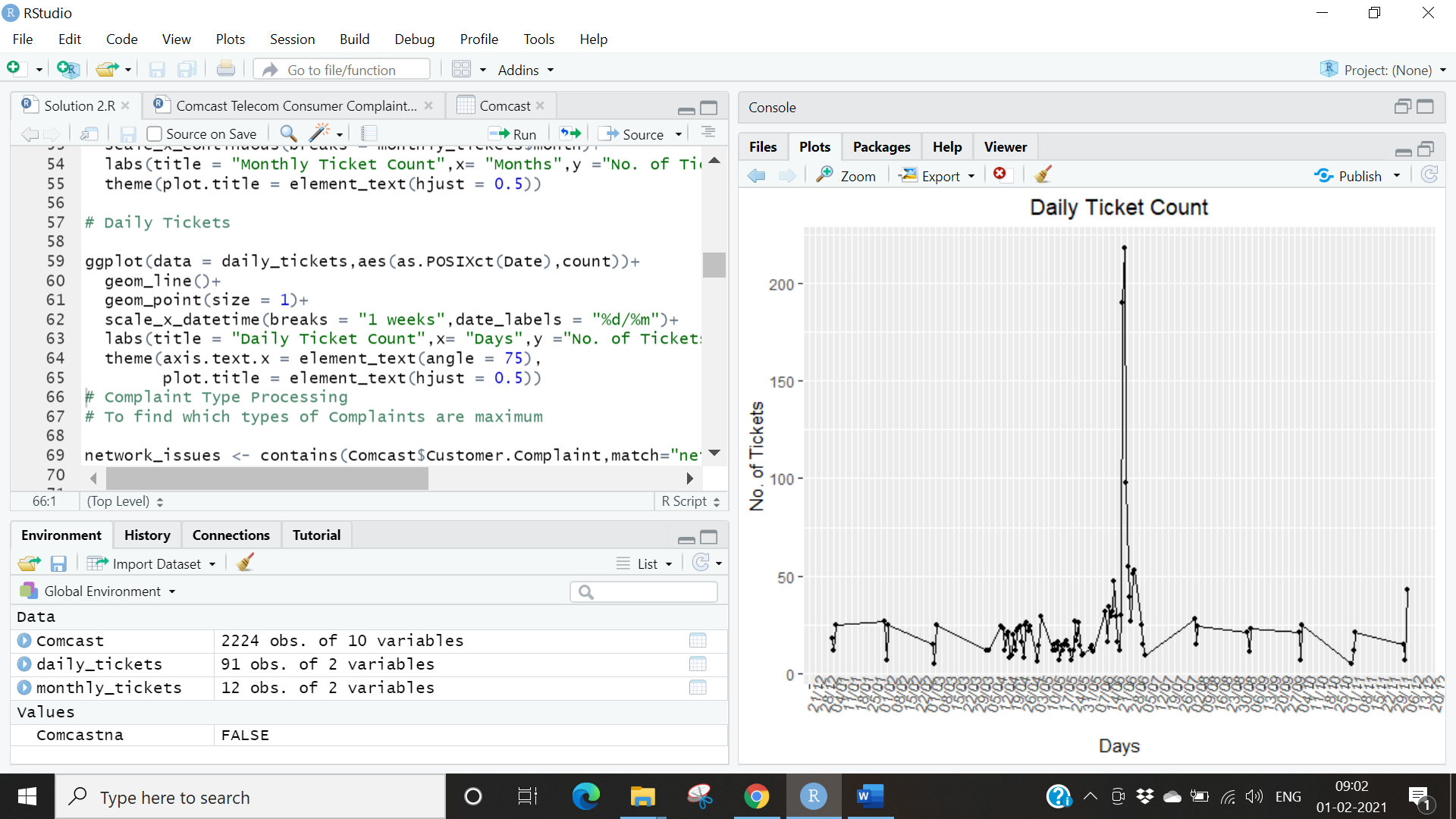
ggarrange(total,category,nrow = 1, ncol = 2)

category

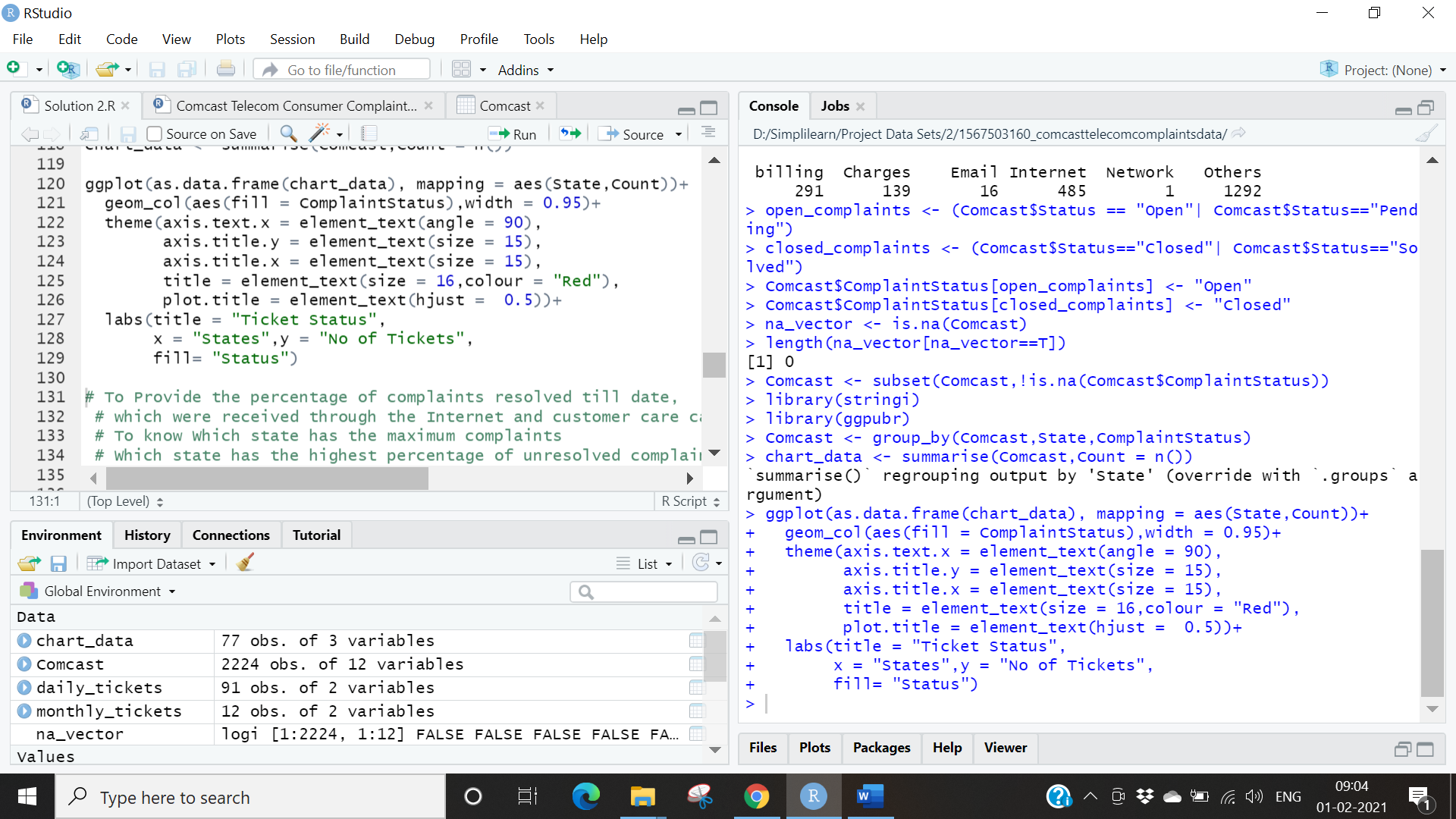


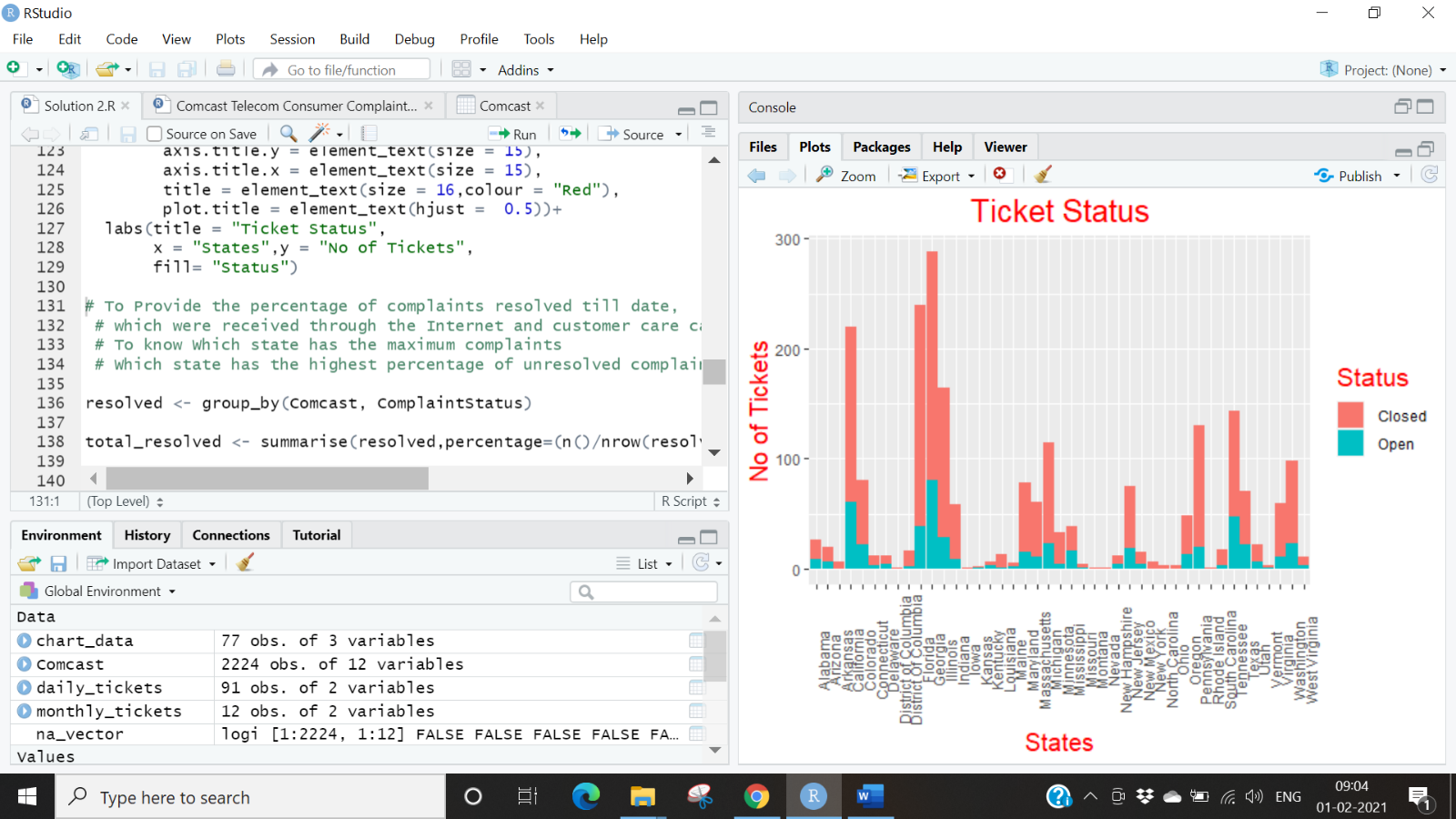


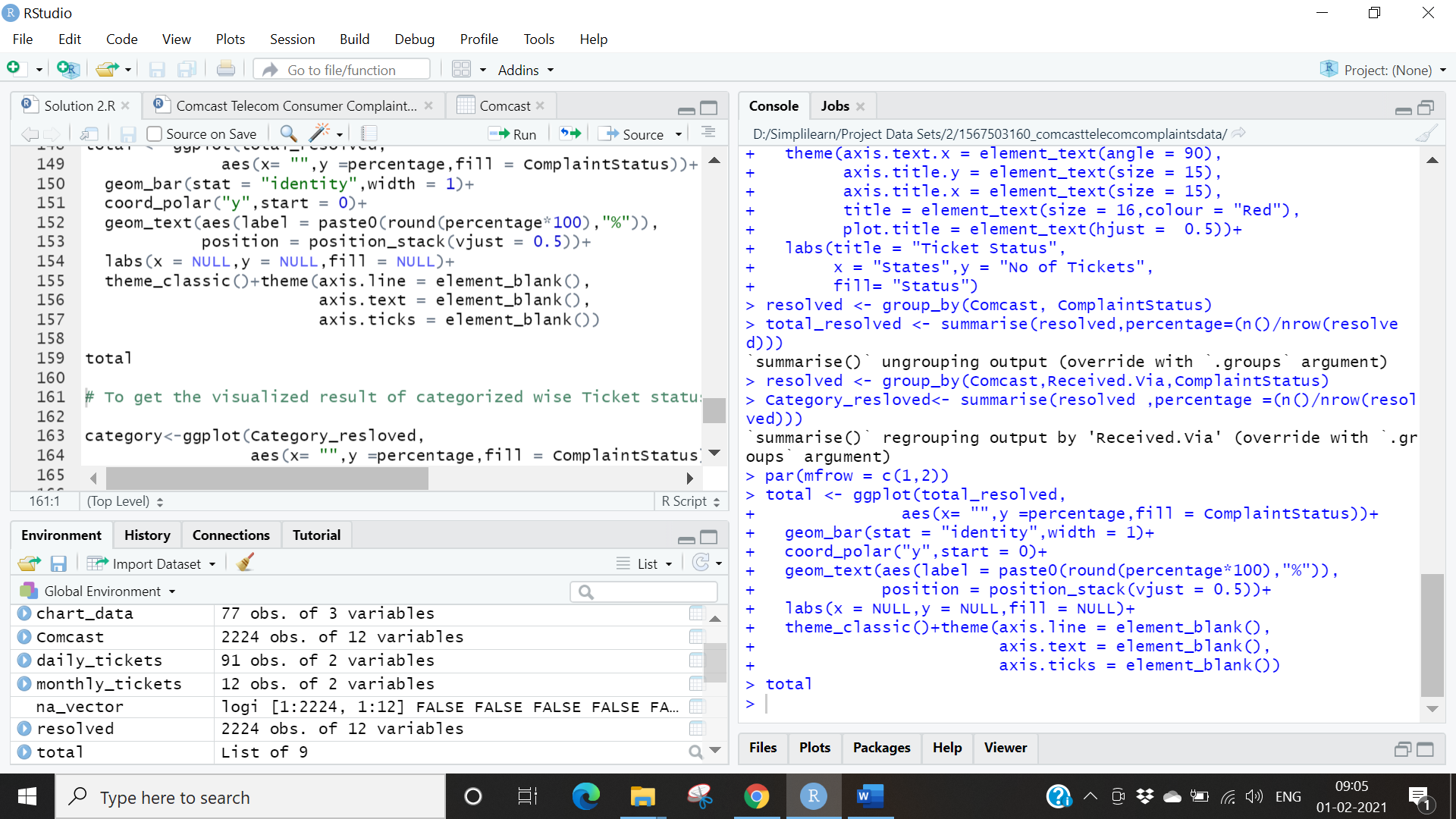


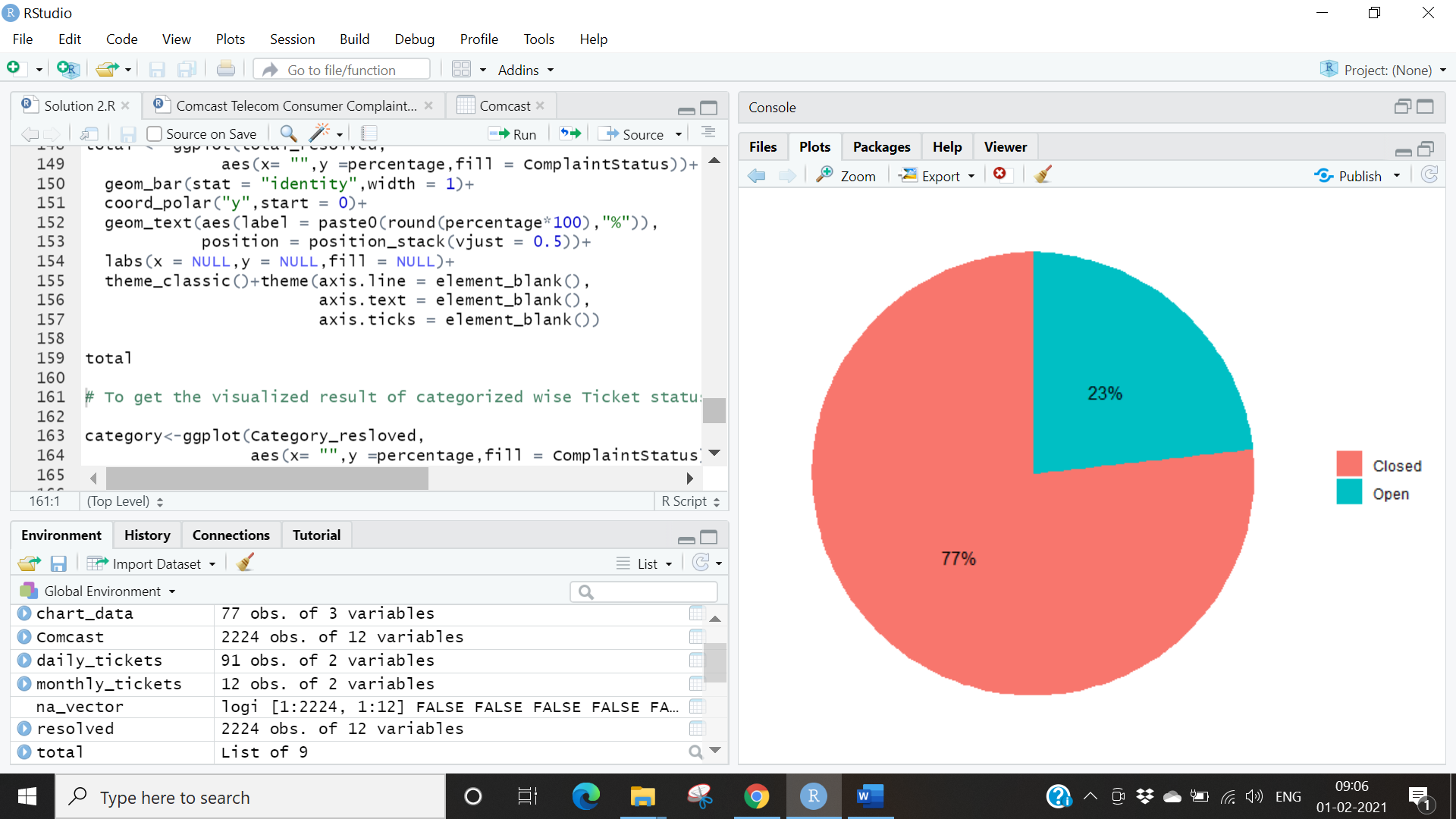


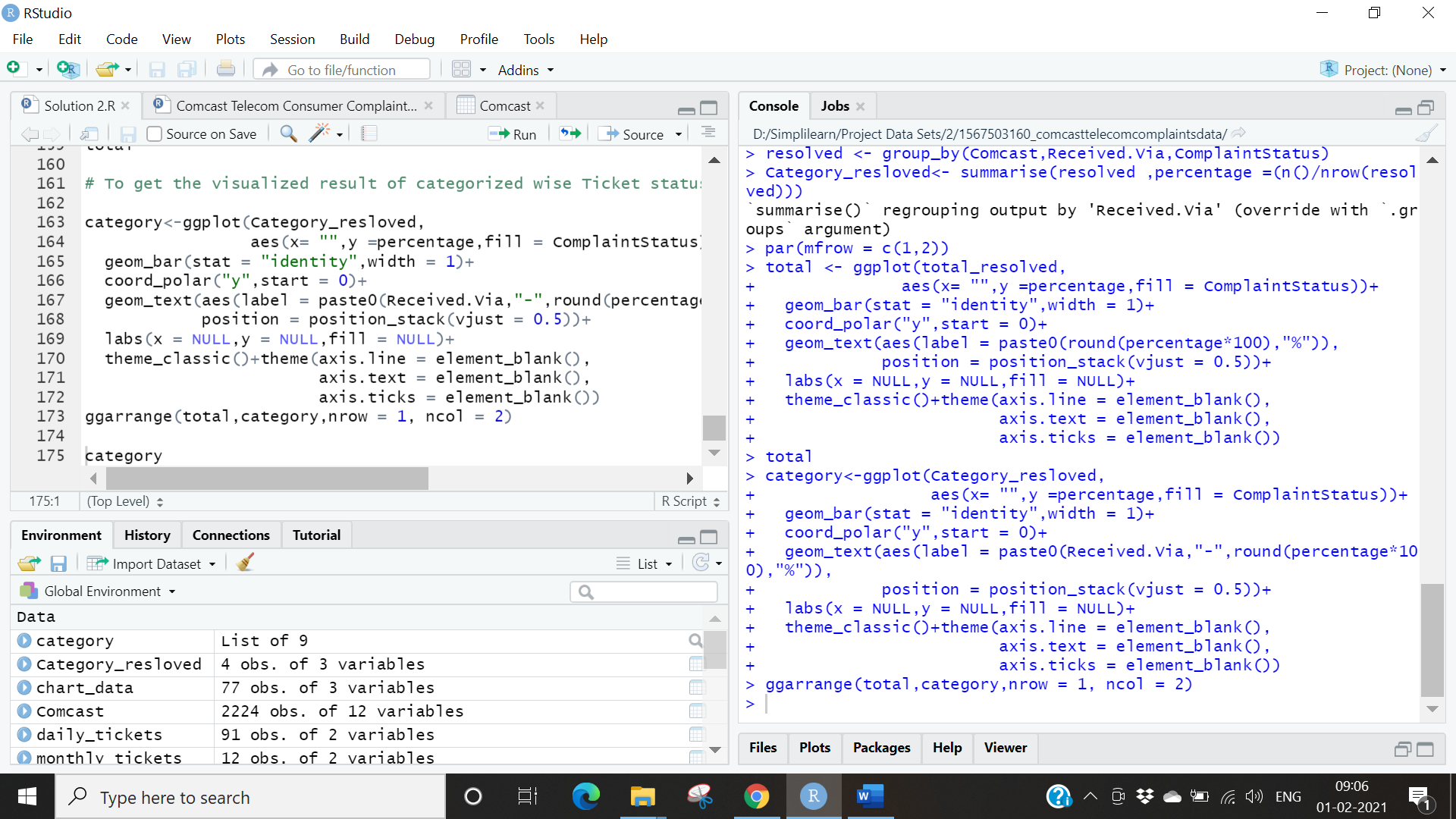


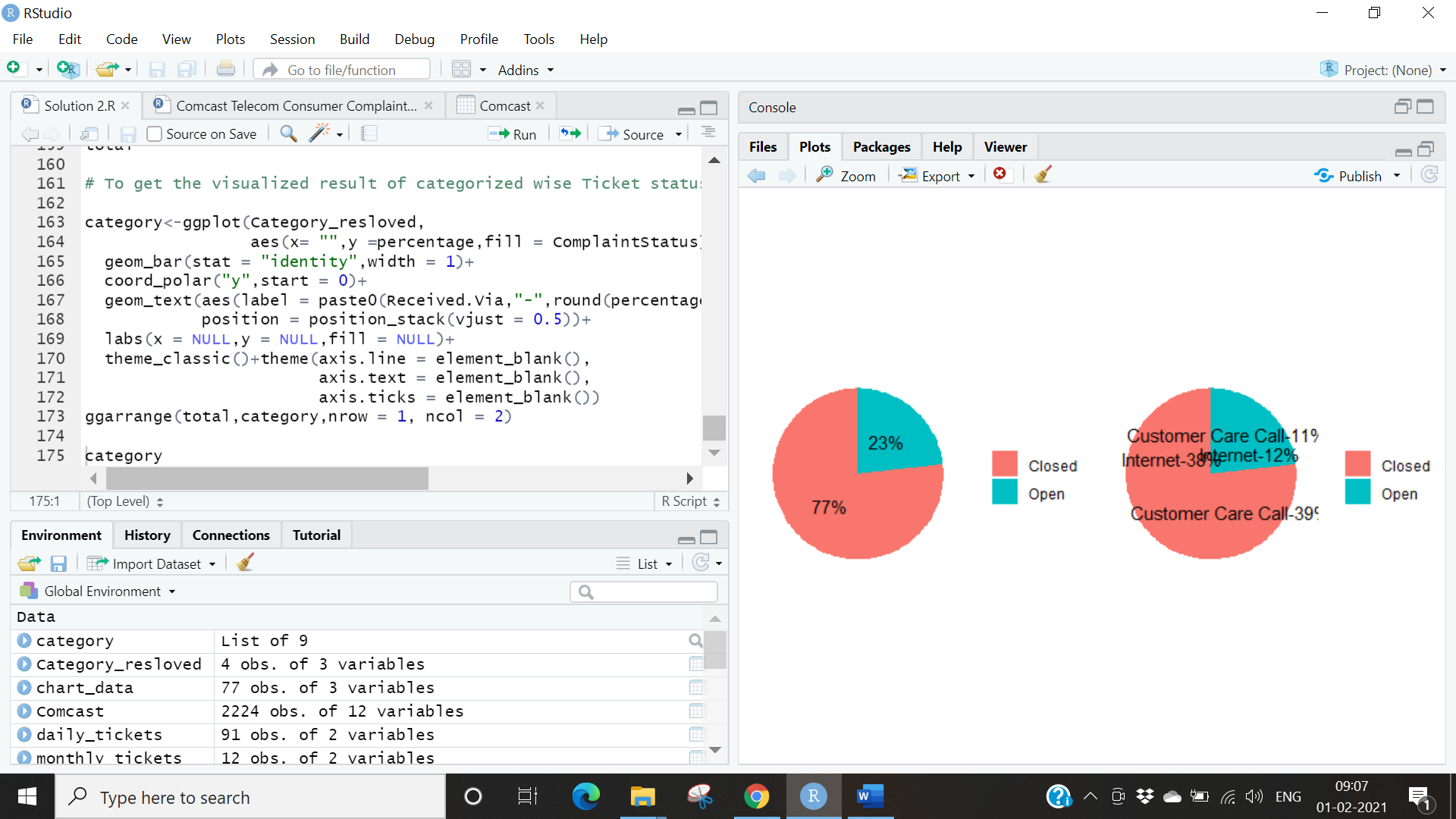


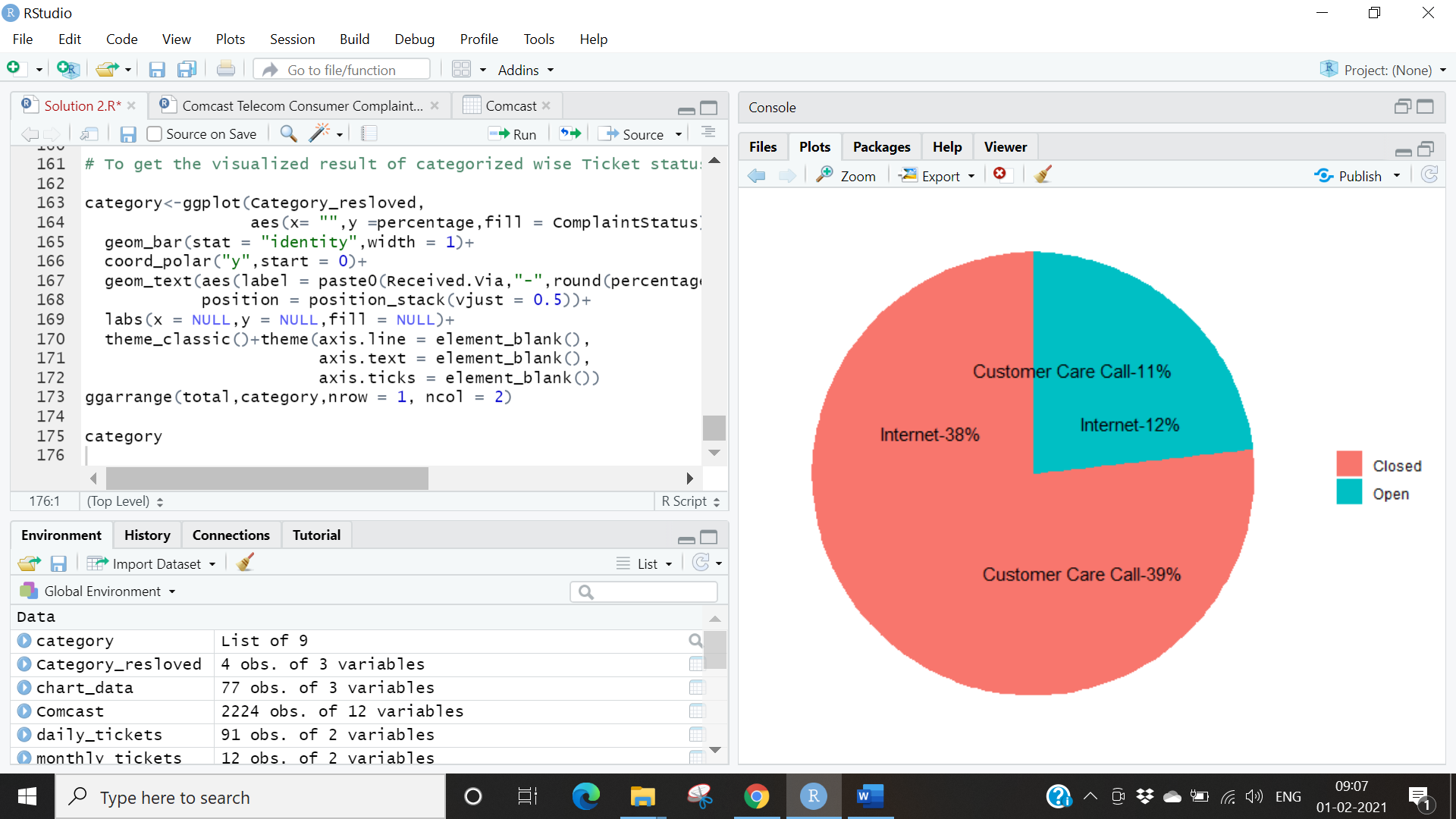












**Conclusion:**

* The Number of tickets started increasing drastically in the month of April, May and June.
* In the table most of the complaints are related to Internet issues. A lot of other categories of complaints were grouped under the “Others” category.
* We can observe in the Stacked chart the states where the number of tickets is the highest are in Georgia and Florida.
* From the Pie charts we observe that
  + - We can conclude that the resolved complaints are 77% in which 38% are received from the Internet and 39% from the customer care calls. Also, we can notice that there is 23% of complaints that are still unresolved and in which 12% are received from the Internet and 11% from the customer care calls.
    - In the 2nd half of the June month Comcast received high number of complaints in which most of the complaints are related to internet service issue and the highest number 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.