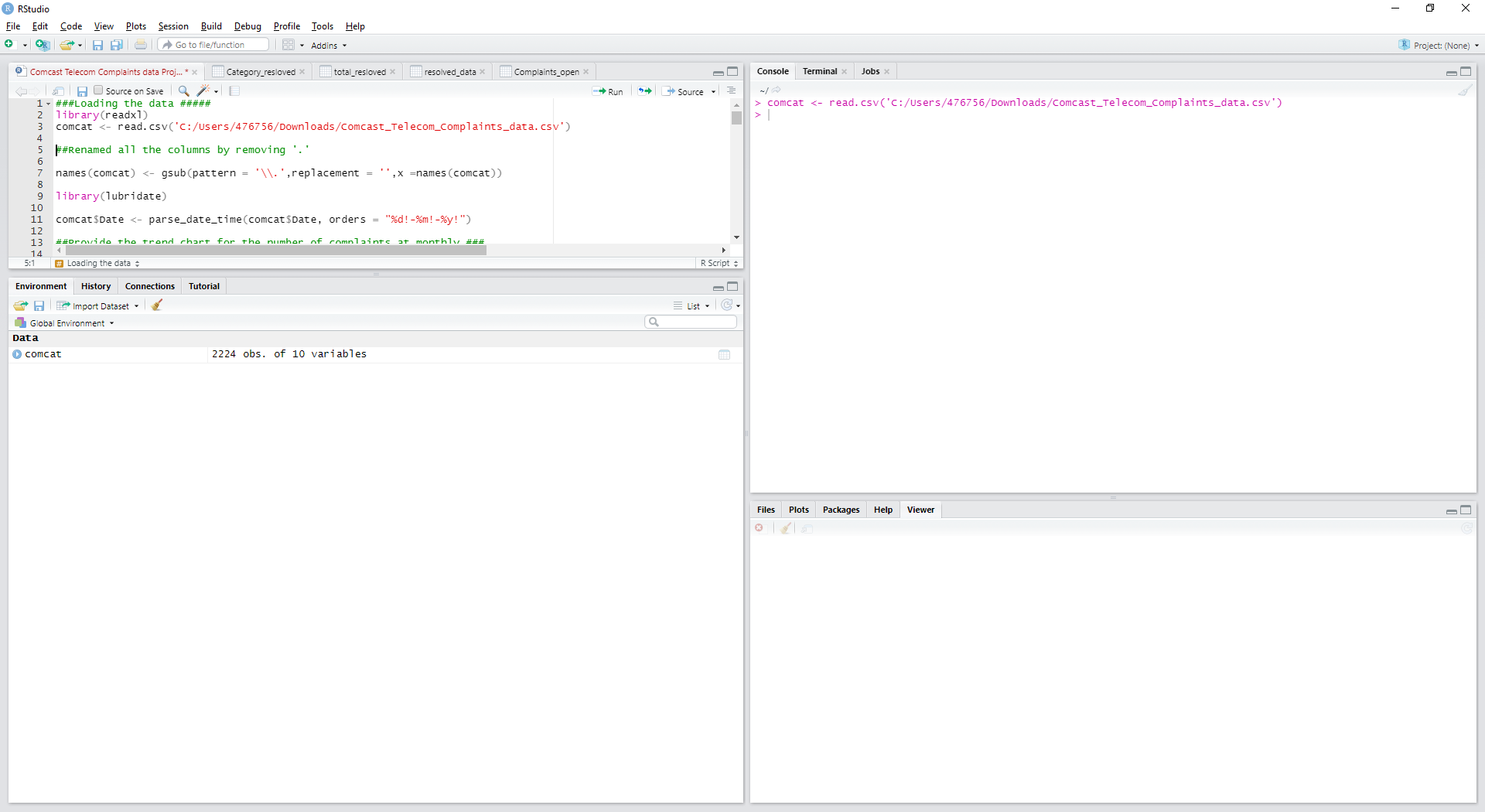
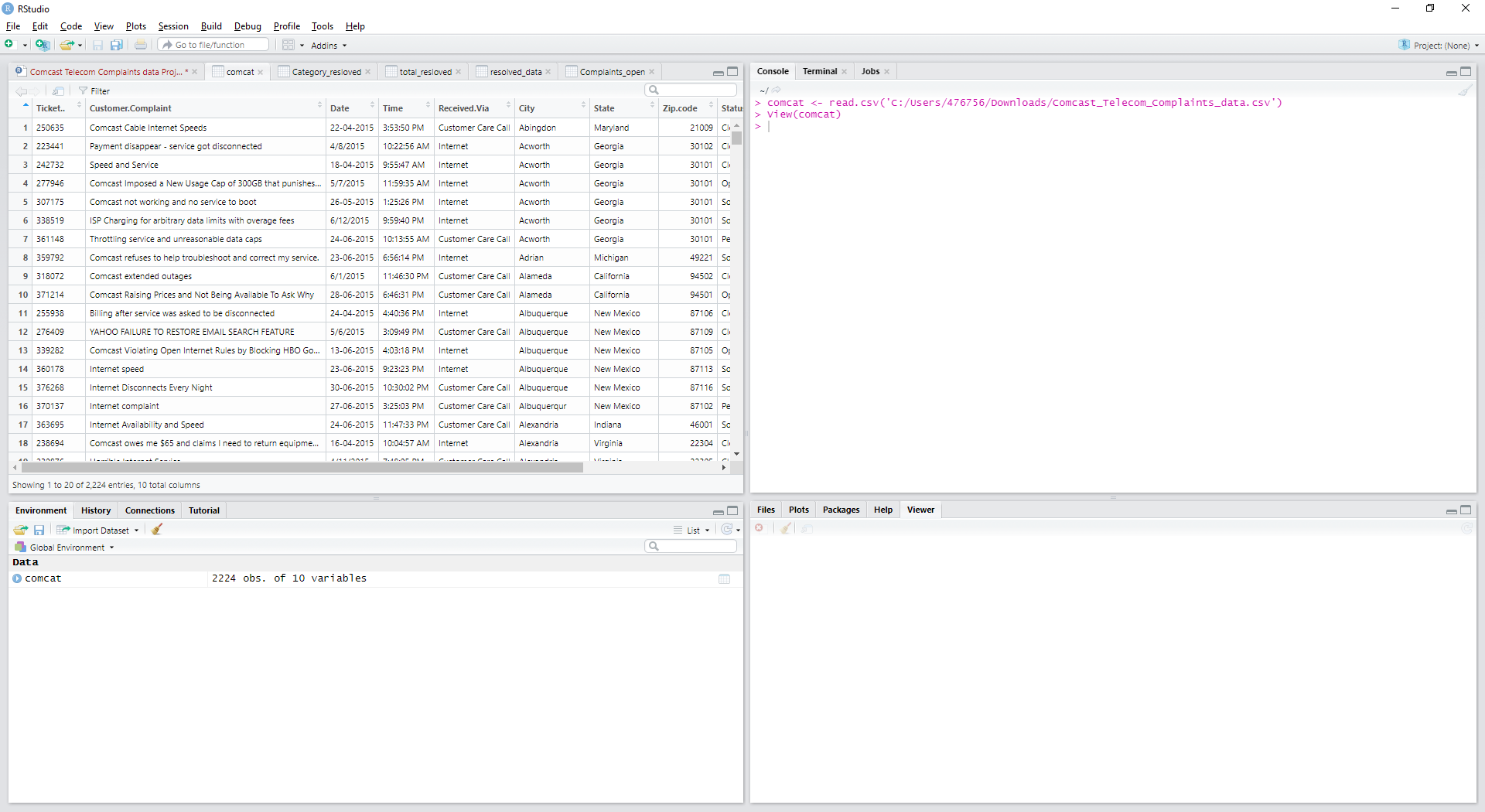
**Loading the dataset into R Studio**

###Loading the data #####

library(readxl)

comcat <- read.csv('C:/Users/476756/Downloads/Comcast\_Telecom\_Complaints\_data.csv')



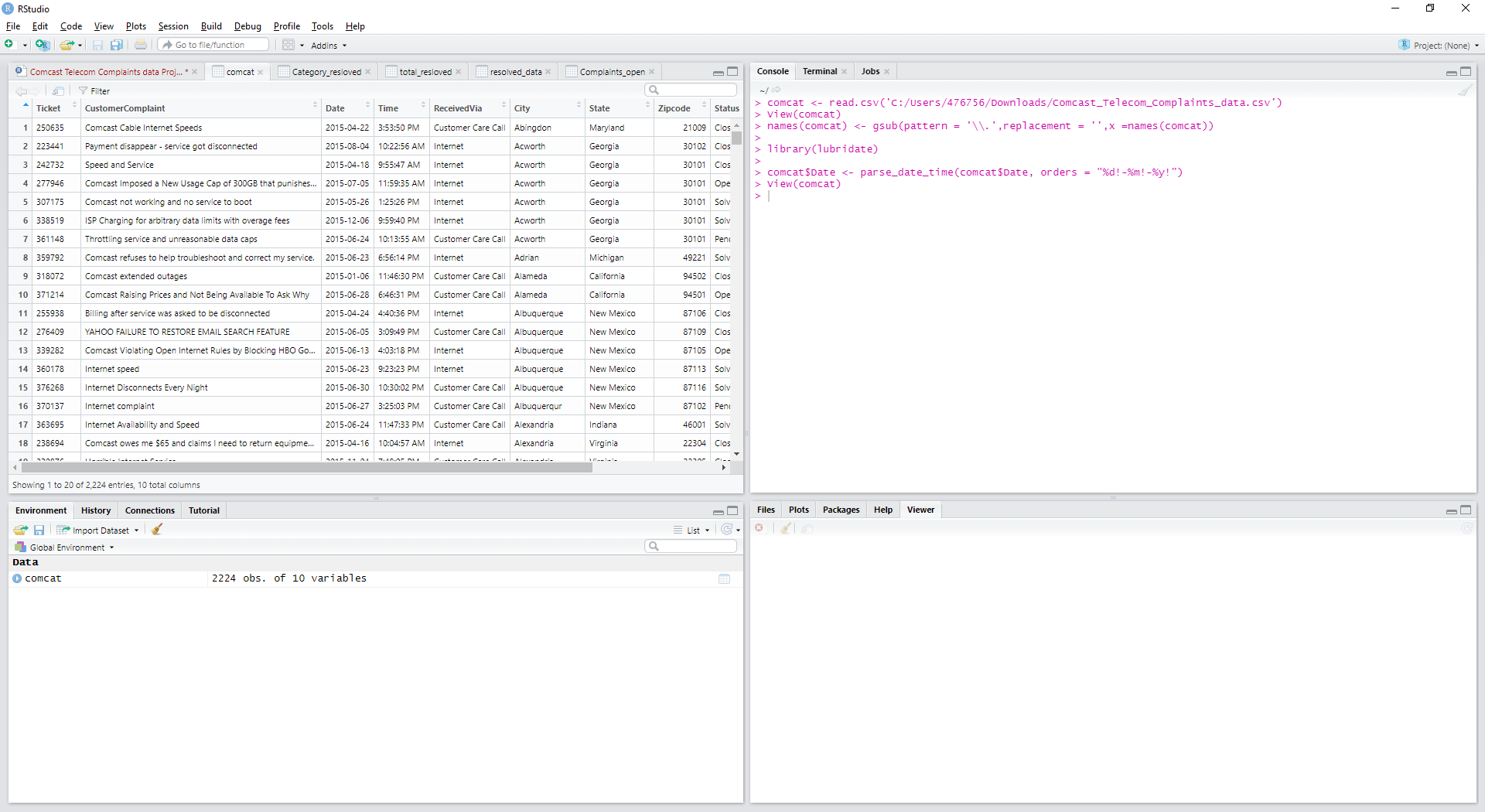


**##Renamed all the columns by removing '.'**

names(comcat) <- gsub(pattern = '\\.',replacement = '',x =names(comcat))

library(lubridate)

comcat$Date <- parse\_date\_time(comcat$Date, orders = "%d!-%m!-%y!")



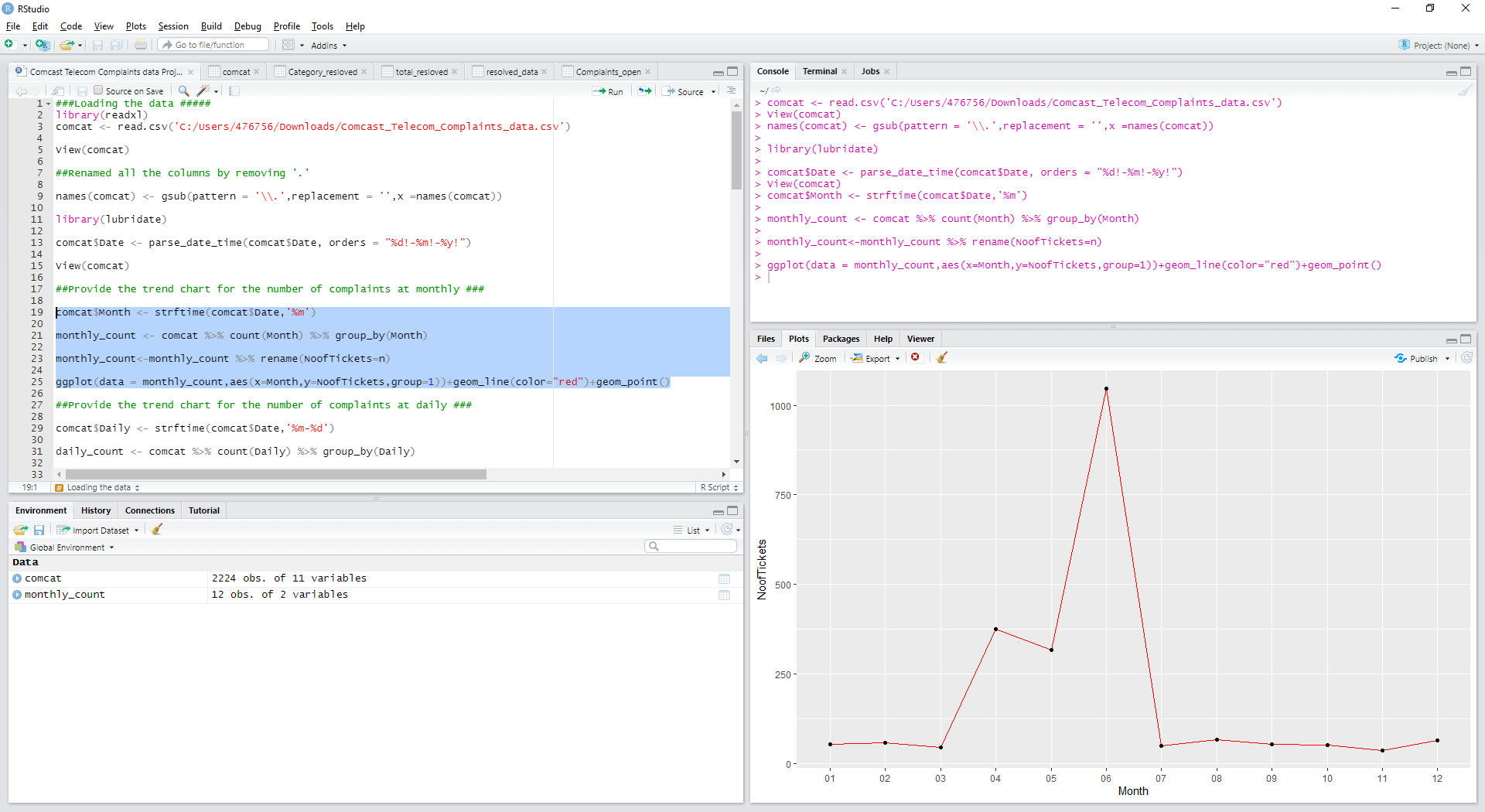
**##Provide the trend chart for the number of complaints at monthly ###**

comcat$Month <- strftime(comcat$Date,'%m')

monthly\_count <- comcat %>% count(Month) %>% group\_by(Month)

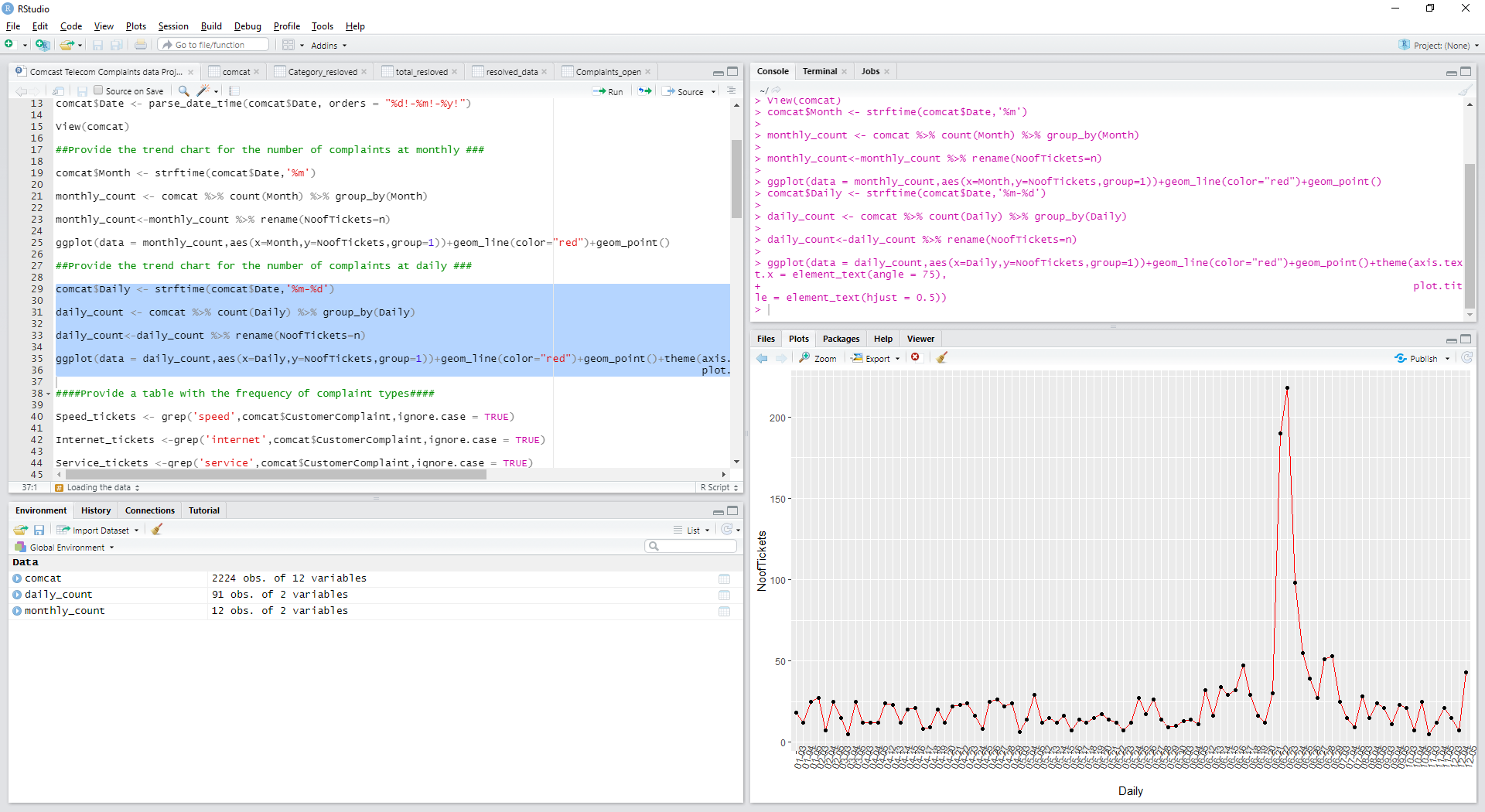
monthly\_count<-monthly\_count %>% rename(NoofTickets=n)

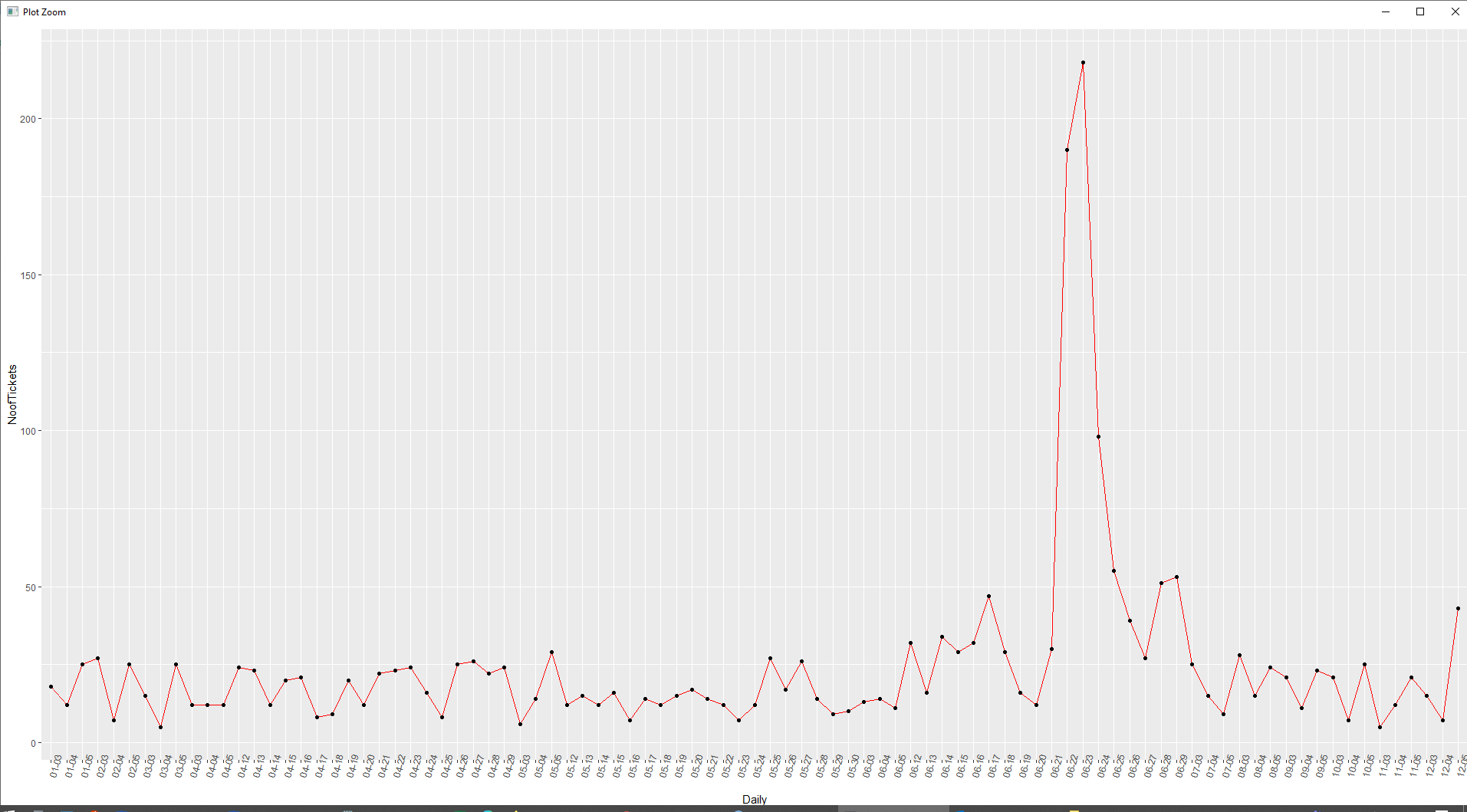
ggplot(data = monthly\_count,aes(x=Month,y=NoofTickets,group=1))+geom\_line(color="red")+geom\_point()



It is evident from the above graph that no of tickets is maximum in the month of June.

**##Provide the trend chart for the number of complaints at daily ###**





It is evident from the above graph that no of tickets is maximum on 24th June.

**####Provide a table with the frequency of complaint types####**

Categorized the different type of complaints based on the keyword from ‘CusomterComplaint’ column

Speed\_tickets <- grep('speed',comcat$CustomerComplaint,ignore.case = TRUE)

Internet\_tickets <-grep('internet',comcat$CustomerComplaint,ignore.case = TRUE)

Service\_tickets <-grep('service',comcat$CustomerComplaint,ignore.case = TRUE)

Billing\_tickets <-grep('billing',comcat$CustomerComplaint,ignore.case = TRUE)

Network\_tickets <-grep('network',comcat$CustomerComplaint,ignore.case = TRUE)

comcat$ComplaintType[Speed\_tickets] <- "Speed"

comcat$ComplaintType[Internet\_tickets] <- "internet"

comcat$ComplaintType[Service\_tickets] <- "service"

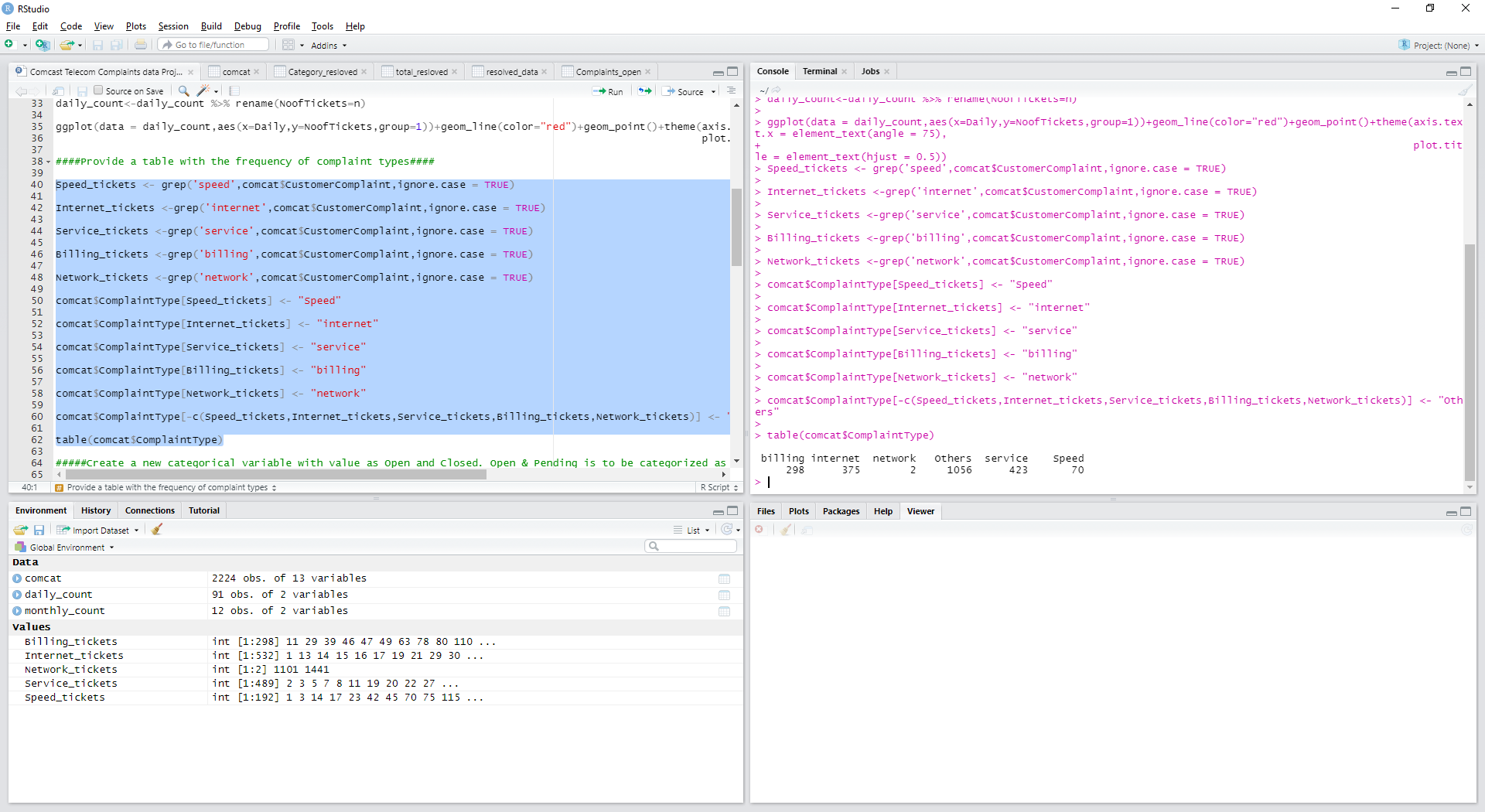
comcat$ComplaintType[Billing\_tickets] <- "billing"

comcat$ComplaintType[Network\_tickets] <- "network"

**####Frequency of complaint types.########**

comcat$ComplaintType[-c(Speed\_tickets,Internet\_tickets,Service\_tickets,Billing\_tickets,Network\_tickets)] <- "Others"

table(comcat$ComplaintType)



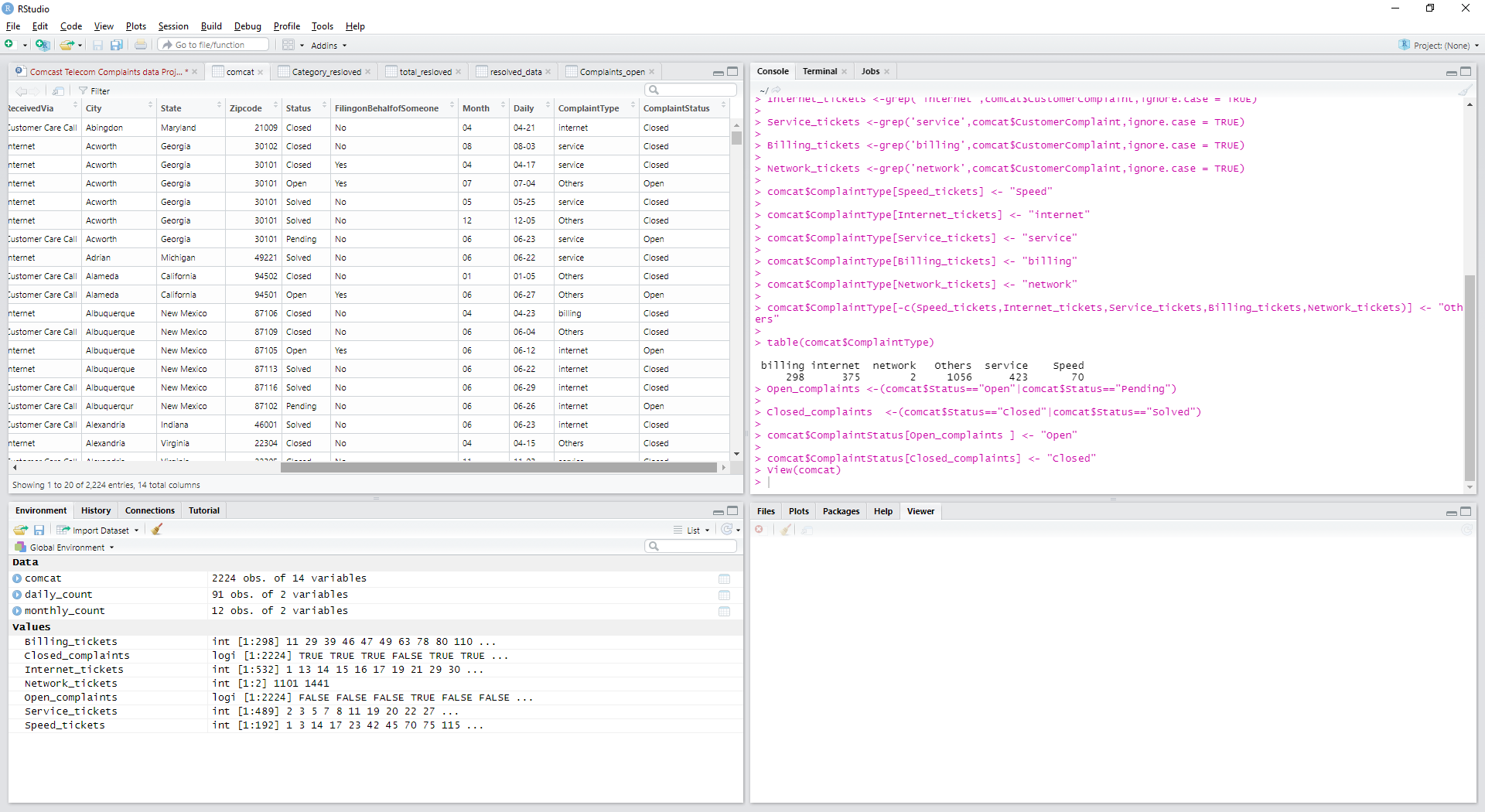
**#####Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.###**

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

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

comcat$ComplaintStatus[Open\_complaints ] <- "Open"

comcat$ComplaintStatus[Closed\_complaints] <- "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###############**

comcat\_statewise <- group\_by(comcat,State,ComplaintStatus)

View(comcat\_statewise)

stackedchart <- summarise(comcat\_statewise,count=n())

View(stackedchart)

ggplot(stackedchart ,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),

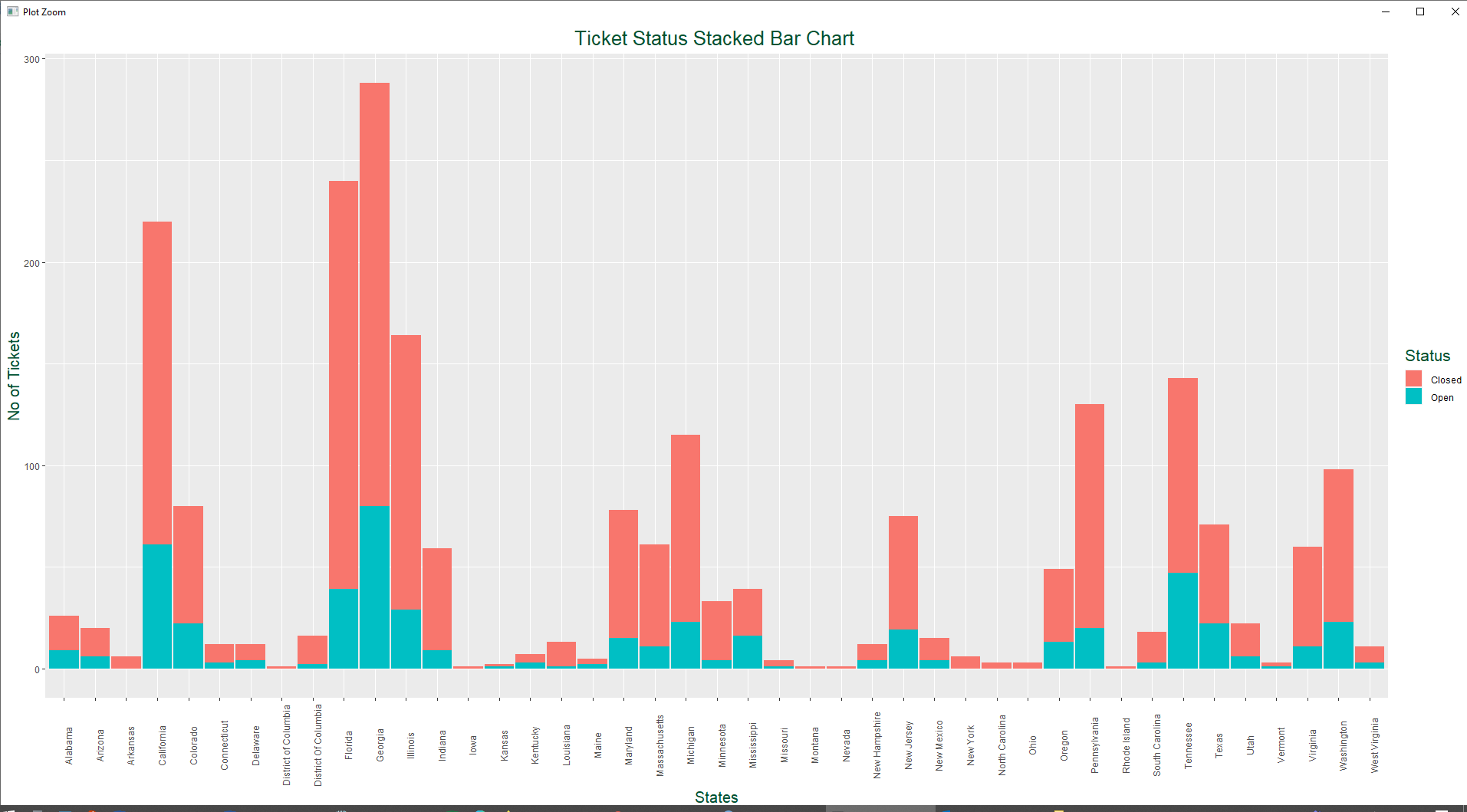
title = element\_text(size = 16,colour = "#005333FF"),

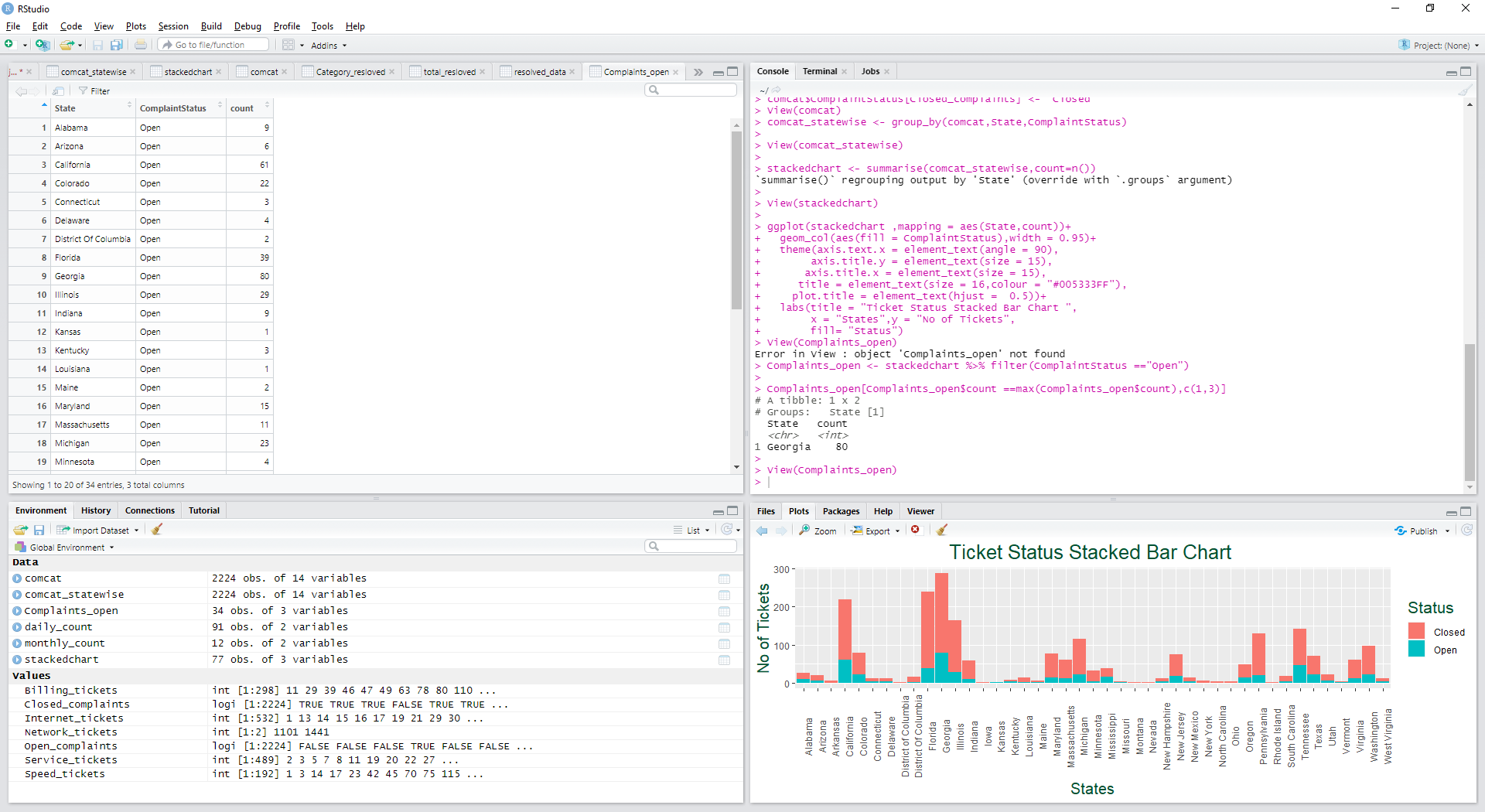
plot.title = element\_text(hjust = 0.5))+

labs(title = "Ticket Status Stacked Bar Chart ",

x = "States",y = "No of Tickets",

fill= "Status")





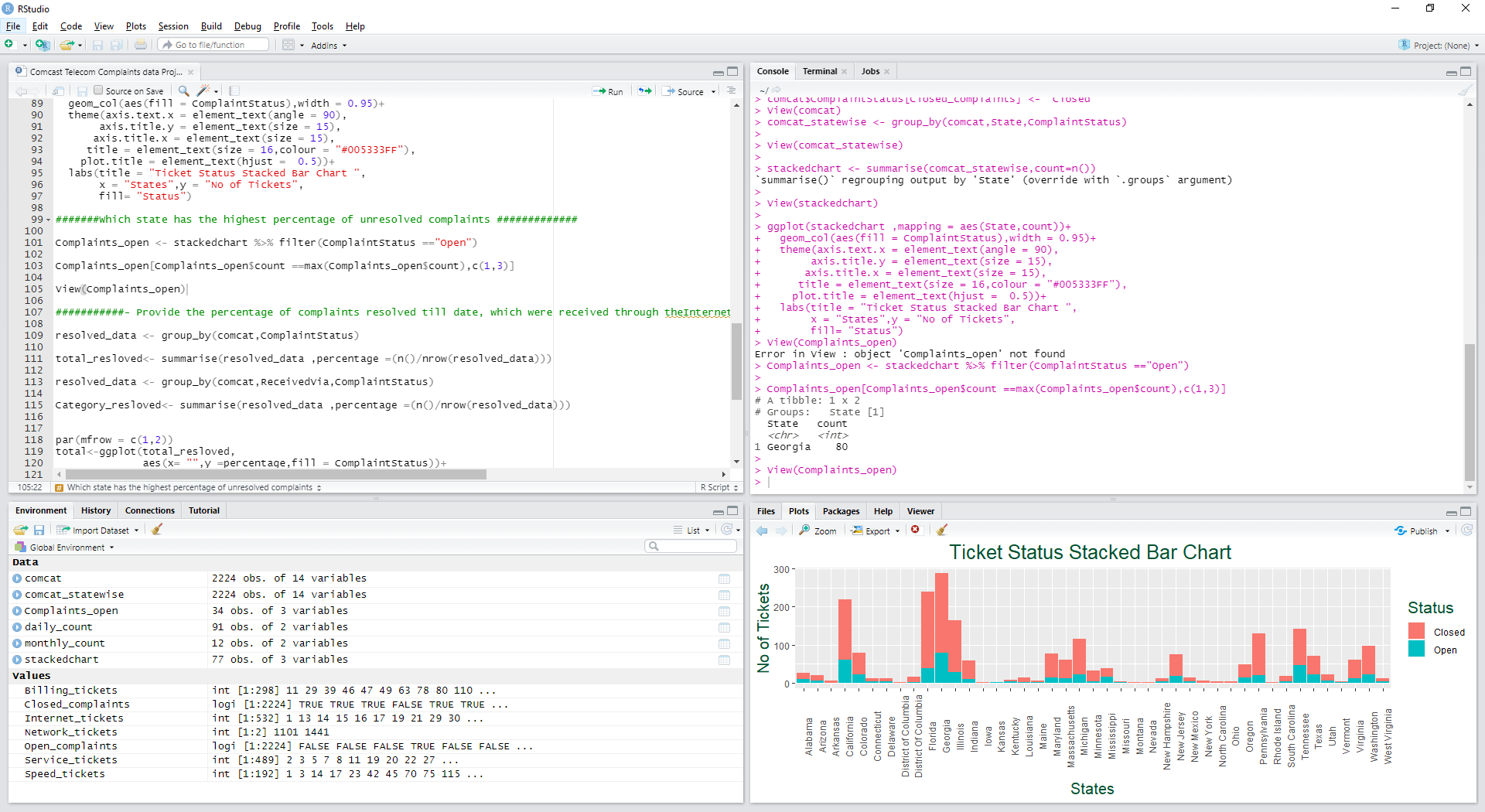
It is evident from the above graph that Georgia State have the most complaints. Florida is the second and California is the third in the list.

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

Complaints\_open <- stackedchart %>% filter(ComplaintStatus =="Open")

View(Complaints\_open)

Complaints\_open[Complaints\_open$count ==max(Complaints\_open$count),c(1,3)]



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

resolved\_data <- group\_by(comcat,ComplaintStatus)

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

resolved\_data <- group\_by(comcat,ReceivedVia,ComplaintStatus)

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

par(mfrow = c(1,2))

total<-ggplot(total\_resloved,

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())

# Pie Chart for Category 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(ReceivedVia,"-",round(percentage\*100),"%")),

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

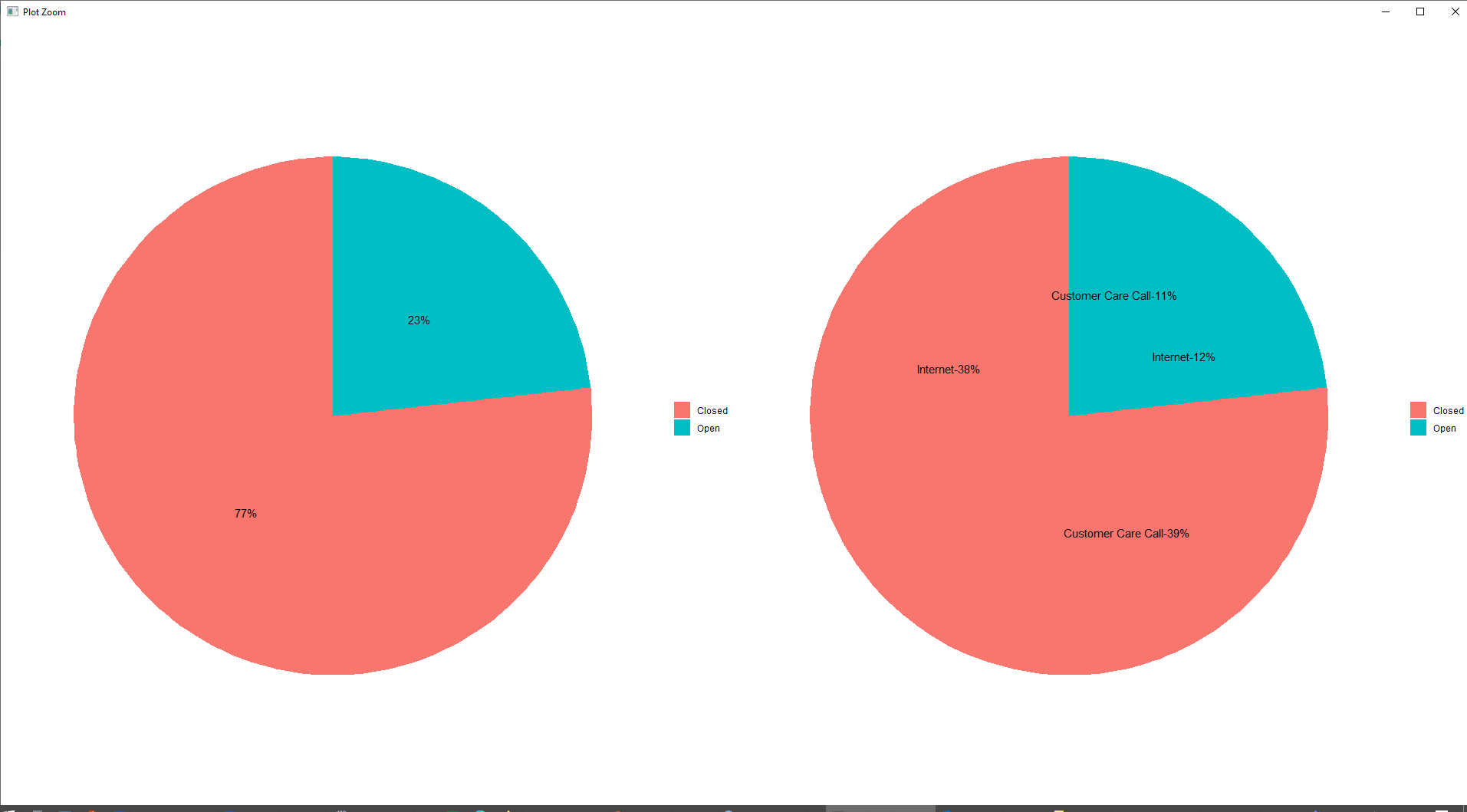
labs(x = NULL,y = NULL,fill = NULL)+

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

axis.text = element\_blank(),

axis.ticks = element\_blank())

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



From the above graphs, it is evident that 23% of the complaints are open and 77% of the complaints have been closed

In the closed complaints, 39% where related to Customer care call and 38% where related to Internet

In the Open complaints, 11% where related to Customer care call and 12% where related to Internet

**Insights:**

* During the June month, Comcast received high amount of complaints in which most of the complaints are related to internet service issue
* Highest amount of complaints are received from the state Georgia.
* The highest unresolved complaints are related from the state Georgia
* The total amount of resolved complaints are 77% in which 38% are received the internet and 39% are from the customer care calls.