CODE

Task 1 -Import data into R environment

Comcast = read.csv("Comcast Telecom Complaints data.csv")

str(Comcast)

Comcast$Date = as.Date(Comcast$Date,format="%d-%m-%Y")

View(Comcast)

library('dplyr')

#renaming columns

Comcast <- rename(Comcast, Mode = Received.Via, Type = Customer.Complaint)

Task 2 - Represent the trend chart for number of complaints at monthly and daily

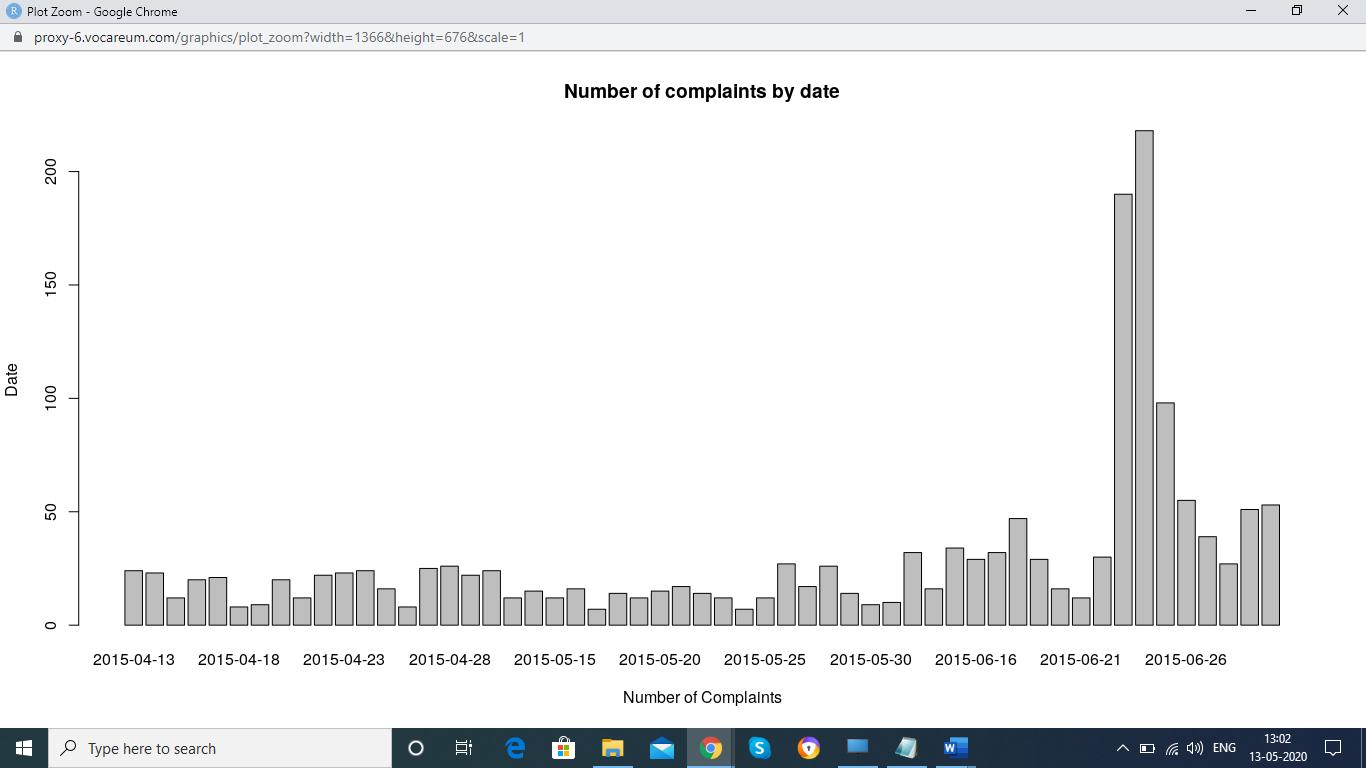
granularity levels

# Daily complaints chart

Freq = table(Comcast$Date)

Freq

barplot(Freq, ylab="Date",xlab="Number of Complaints",main="Number of complaints by date")



# Monthly complaints chart

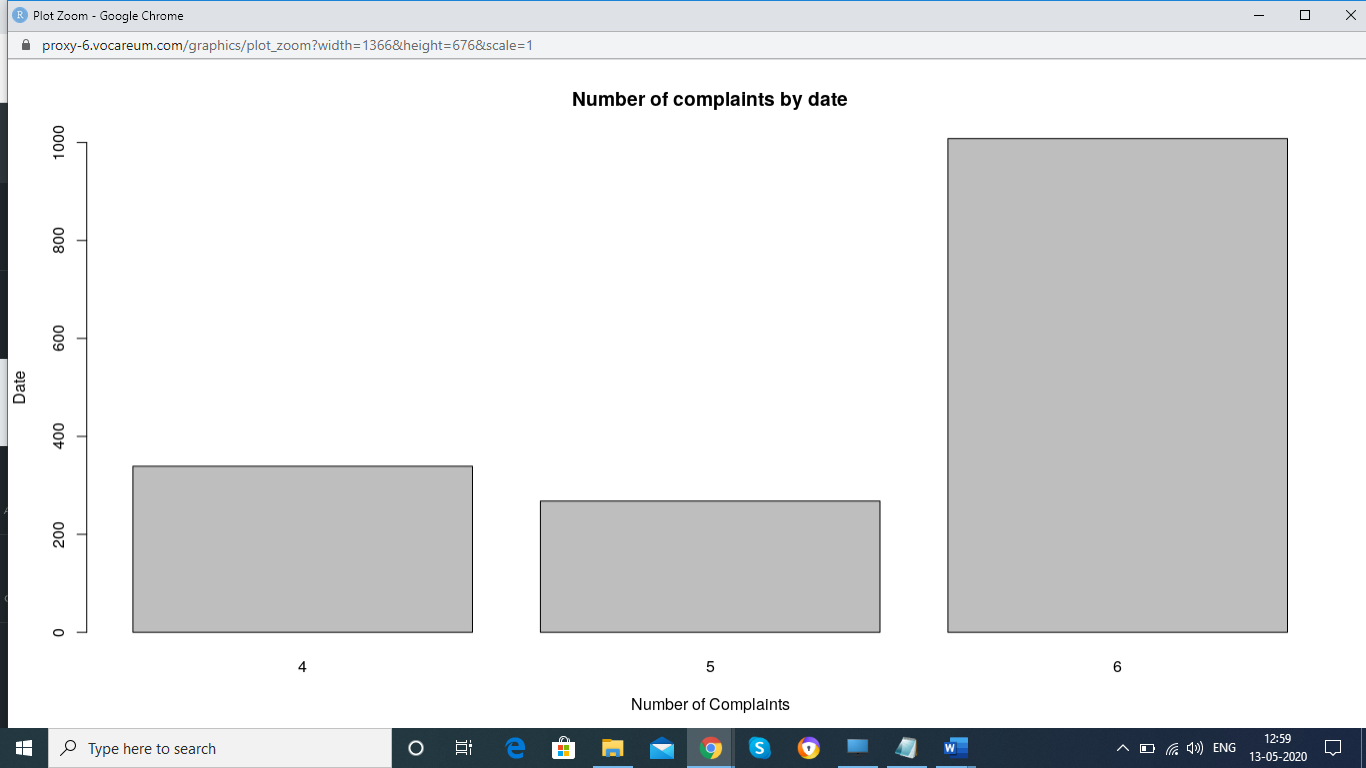
Comcast = transform(Comcast,Month = as.numeric(format(Date,format="%m")))

View(Comcast)

Freq = table(Comcast$Month)

Freq

barplot(Freq, ylab="Months",xlab="Number of Complaints",main="Number of complaints by Month")



# Insights- Most of complaints are received in the Month of June

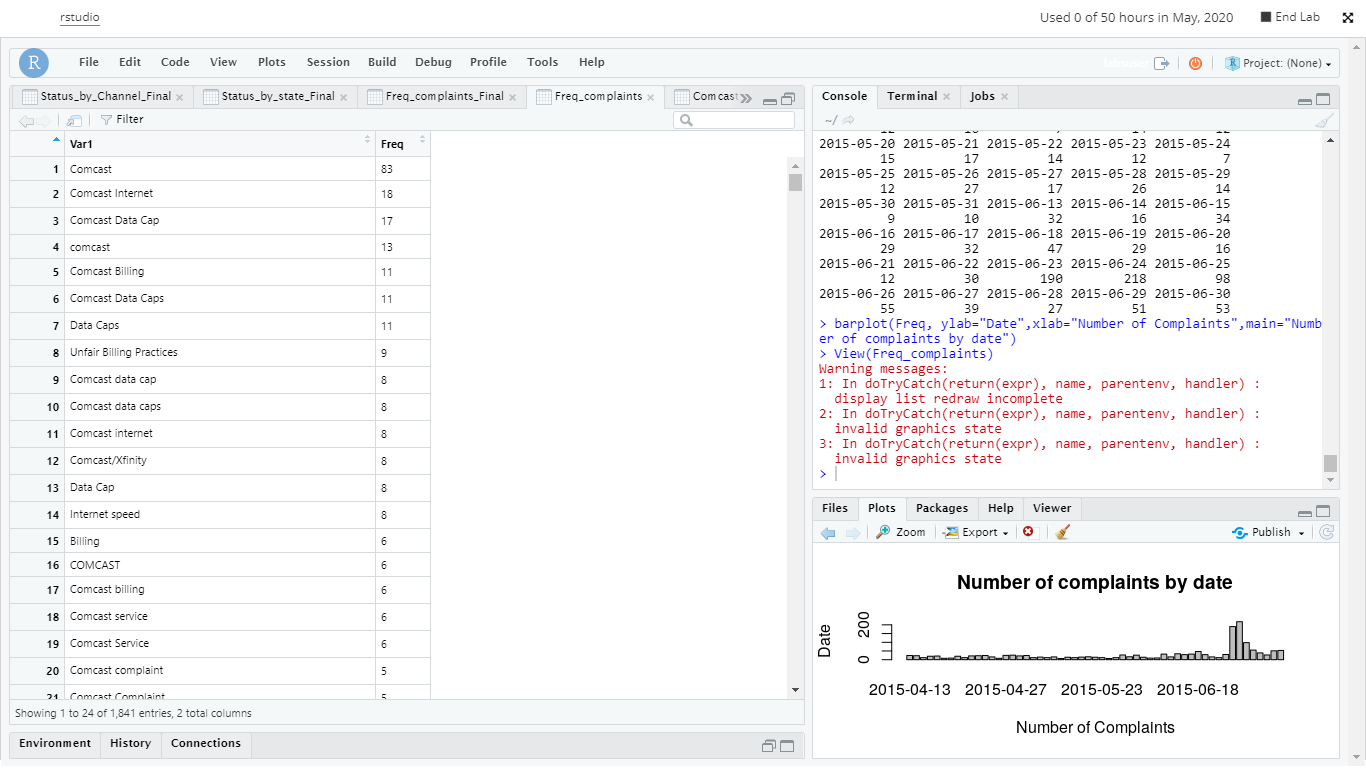
Task 3 - Provide a table with frequency of complaint types

Freq\_complaints = table(Comcast$Type)

View(Freq\_complaints)

Freq\_complaints = arrange(as.data.frame(Freq\_complaints),desc(Freq))

View(Freq\_complaints)



# Insight - Most of the complaints are around Internet and Data cap.

Task 4 - Create a new categorical variable with value as - Open and Closed. Open & Pending to be categorized as 'Open' and Closed & Solved to be categorized as 'Closed'

mutate(Comcast, Category=Status)

for(i in 1:nrow(Comcast)) {

if(Comcast$Status[i]=='Open') {

Comcast$Category[i]='Open'

} else if(Comcast$Status[i]=='Pending') {

Comcast$Category[i]='Open'

}else {

Comcast$Category[i]='Closed'

}

}

View(Comcast)

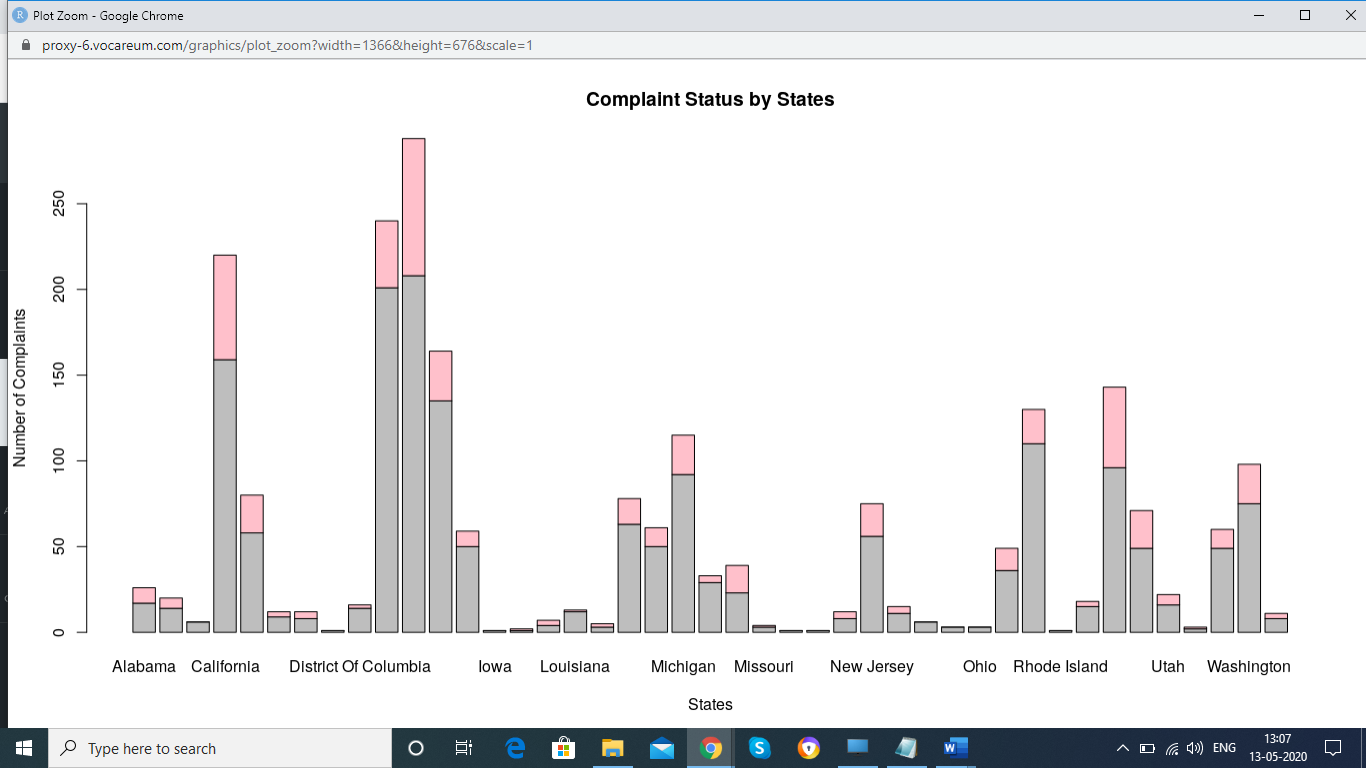
Task 5 - Provide State wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights

Freq1 = table(Comcast$Category,Comcast$State)

Freq1

barplot(Freq1,main="Complaint Status by States", xlab="States", ylab="Number of Complaints",

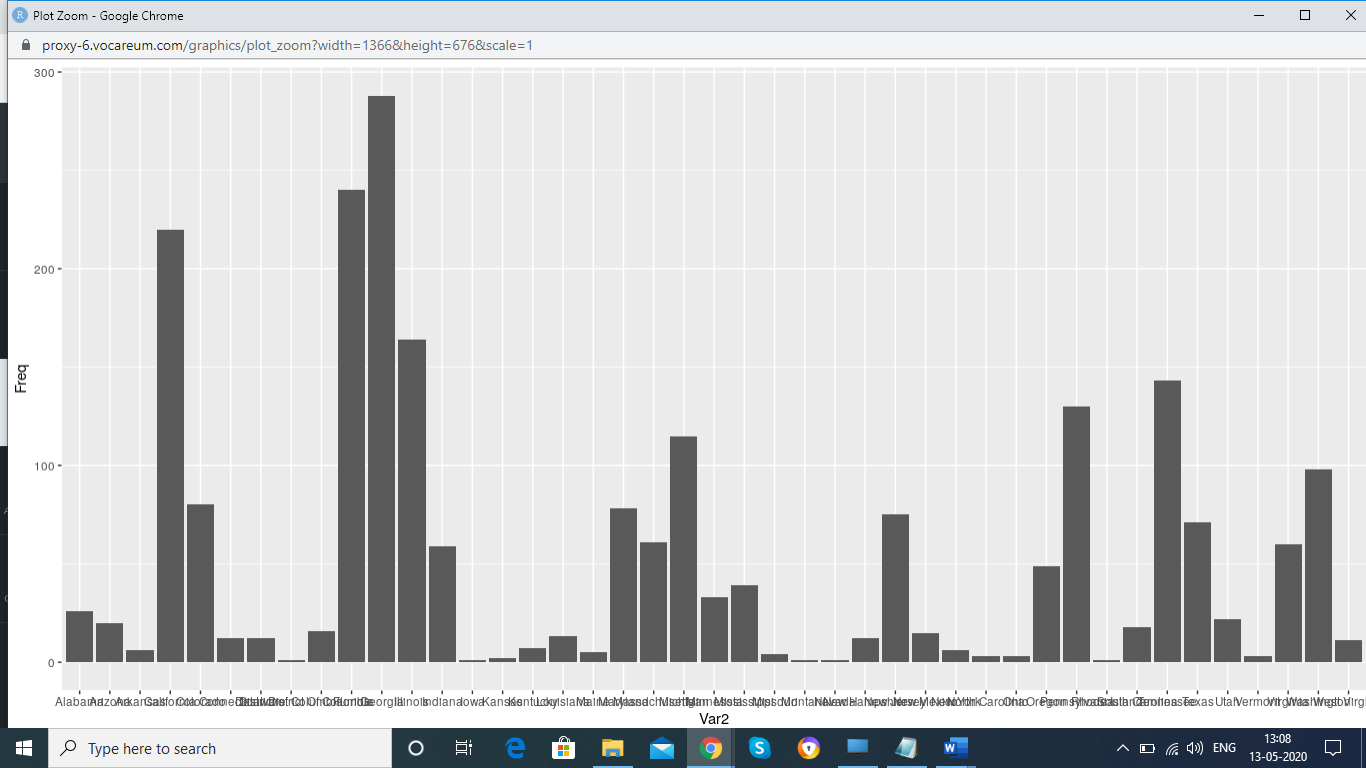
col=c("grey","pink"))



Freq2 = as.data.frame(Freq1)

library(ggplot2)

ggplot(Freq2,aes(x=Var2,y=Freq))+geom\_bar(stat="identity")



# Insights - Georgia has maximum complaints

# Which state has highest percentage of unresolved complaints

Freq2 = table(Comcast$State,Comcast$Category)

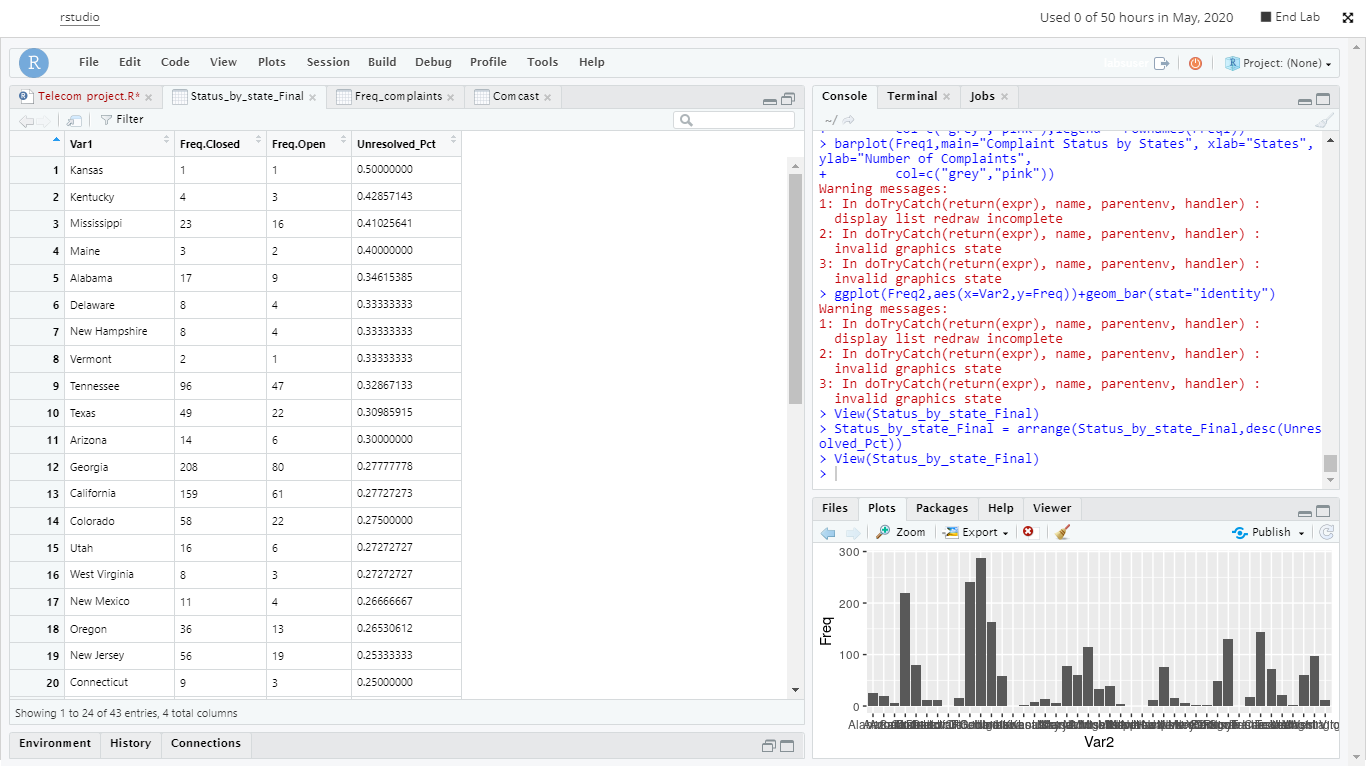
Freq3 = as.data.frame(Freq2)

Status\_by\_state = reshape(Freq3,idvar="Var1",timevar ='Var2',direction="wide")

Status\_by\_state\_Final = transform(Status\_by\_state,Unresolved\_Pct = (Freq.Open/(Freq.Open+Freq.Closed)))

Status\_by\_state\_Final = arrange(Status\_by\_state\_Final,desc(Unresolved\_Pct))

View(Status\_by\_state\_Final)



# Kansas has most unsolved complaints

Task 6- Provide percentage of complaint resolved till date for the complaints received through Internet and customer care calls

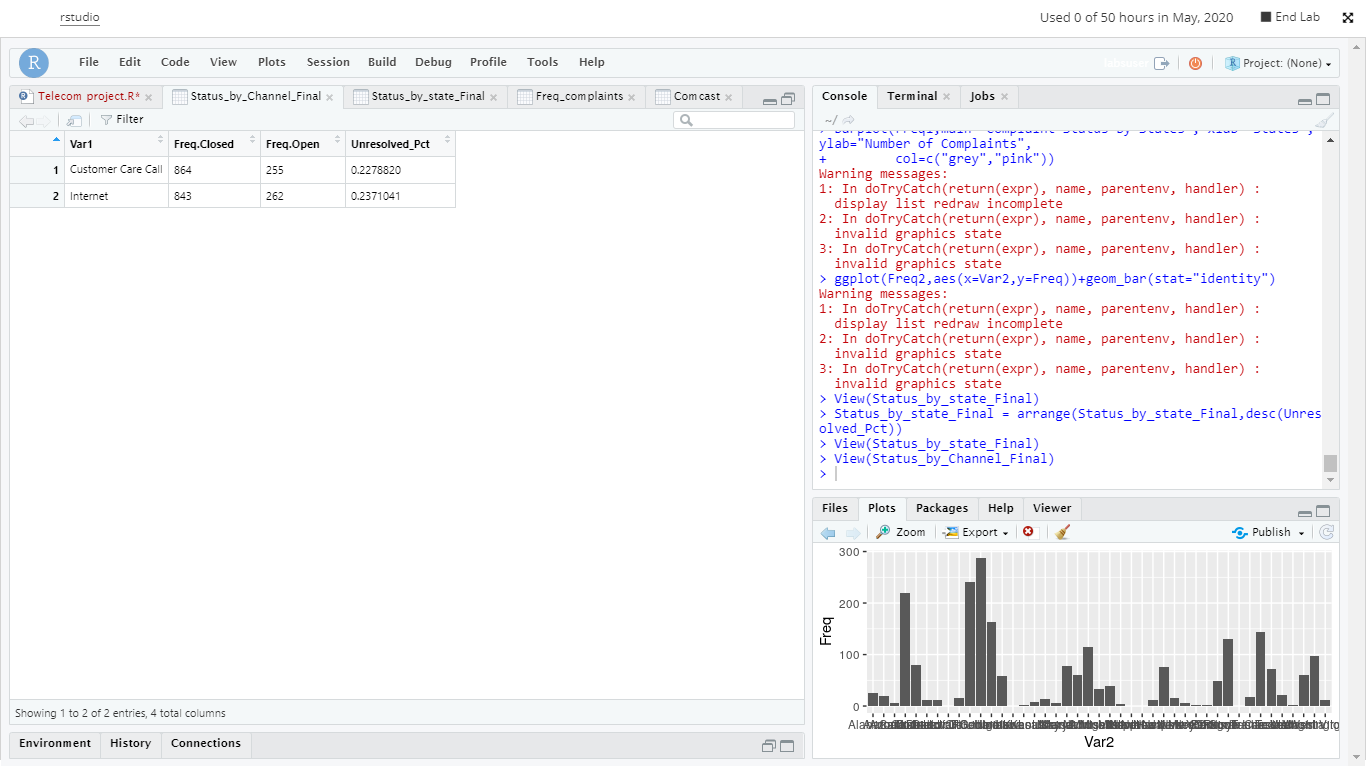
Freq4 = table(Comcast$Mode,Comcast$Category)

Freq4 = as.data.frame(Freq4)

Status\_by\_Channel = reshape(Freq4,idvar="Var1",timevar ='Var2',direction="wide")

Status\_by\_Channel\_Final = transform(Status\_by\_Channel,Unresolved\_Pct = (Freq.Open/(Freq.Open+Freq.Closed)))

View(Status\_by\_Channel\_Final)



# Insights -There is not much difference in complaint resolved status through internet or customer call