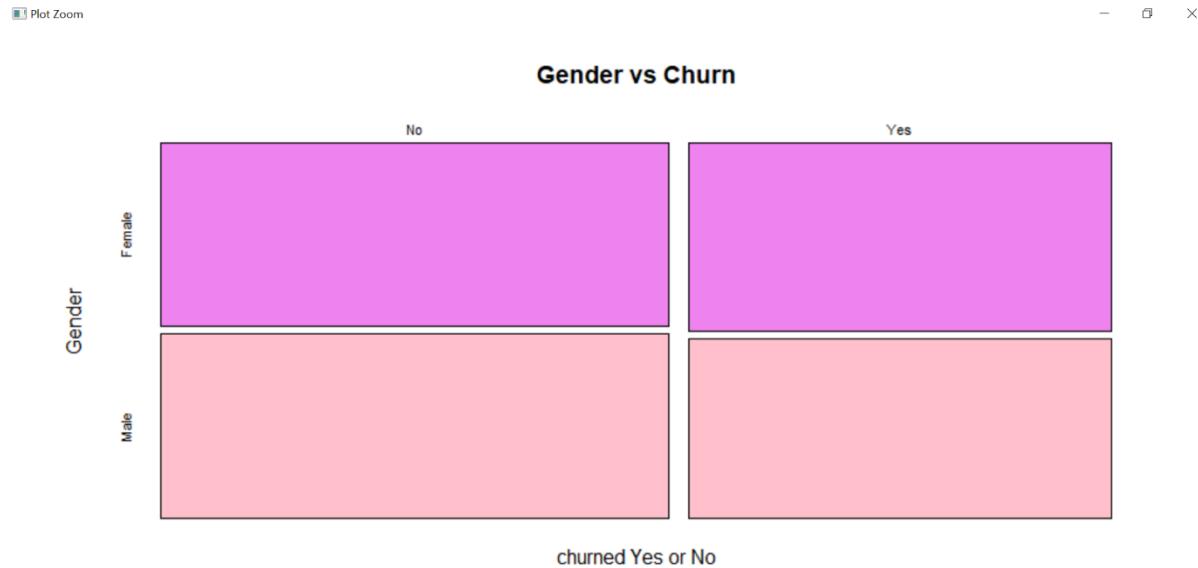


Churn analysis VIZUALIZATION



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

```

34 colSums(is.na(df))
35 #dataset contain no NA values.
36
37 length(df$customerID)
38
39 ## Vizualization ##
40 #Number of Female and Male
41 count(df$gender)
42
43 #Gender vs Churn
44
45 plot(table(df$churn, df$gender), col=c("red", "green"),
46 + xlab = "churned Yes or No", ylab = "Gender", main = "Gender vs Churn")
47

```

47:1 (Top Level) R Script

Console C:/Users/Shiva/Rprog/ ↵

```

x freq
1 Female 6216
2 Male 6119
> plot(table(df$churn, df$gender), col=c("red", "green"),
+ xlab = "churned Yes or No", ylab = "Gender", main = "Gender vs Churn")
> |

```

Gender vs Churn

Gender

No Yes

Female

Male

churned Yes or No

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

```

50
51
52
53
54 partner, col=c("red", "green"),
55 d Yes or No", ylab = "Partner", main = "Partner vs Churn")
56
57
58
59
60 e
61 = "purple", xlab = "Tenure of the connection", main = "Distribution of tenure"
62

```

62:1 (Top Level) R Script

Console C:/Users/Shiva/Rprog/ ↵

```

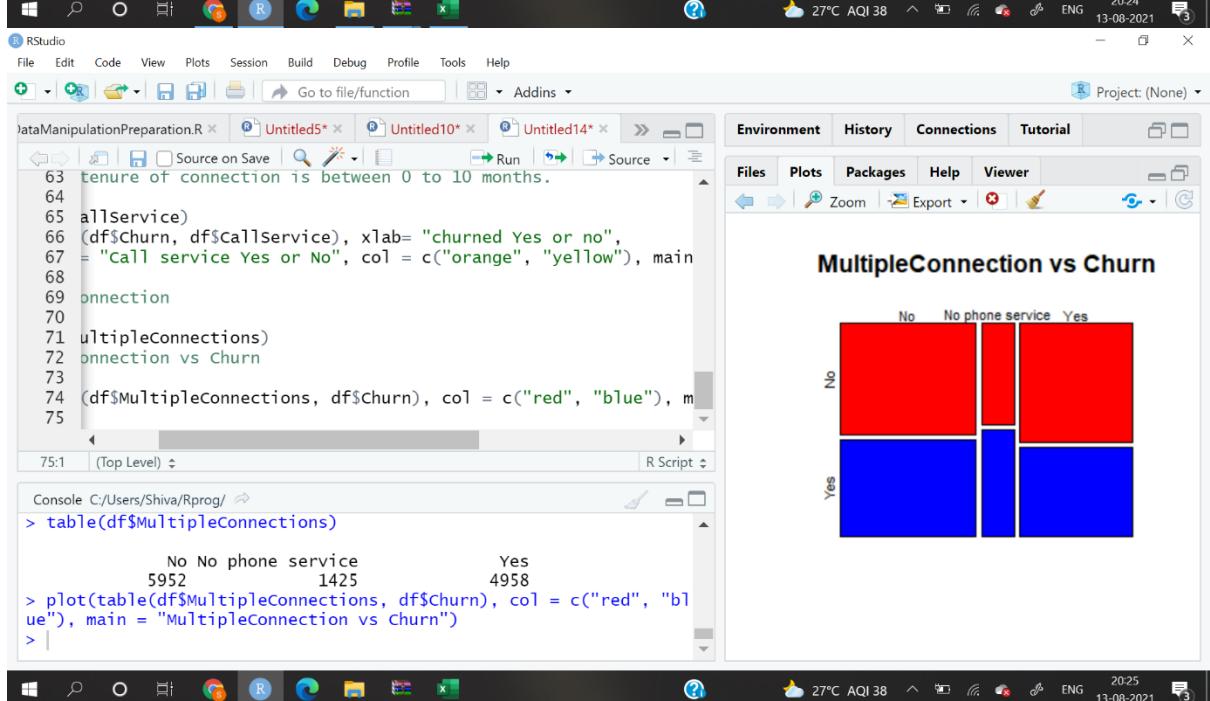
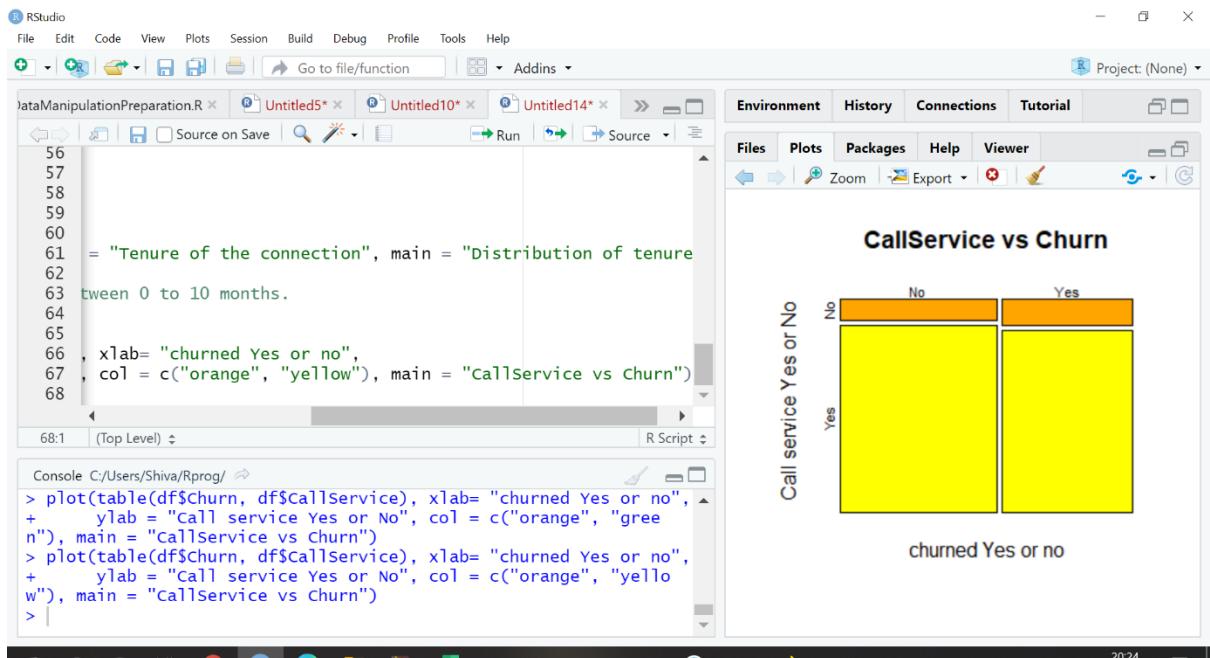
> #Histogram of tenure
> hist(df$tenure, col = "orange", xlab = "Tenure of the connection",
+ main = "Distribution of tenure")
> #Histogram of tenure
> hist(df$tenure, col = "purple", xlab = "Tenure of the connection",
+ main = "Distribution of tenure")
> |

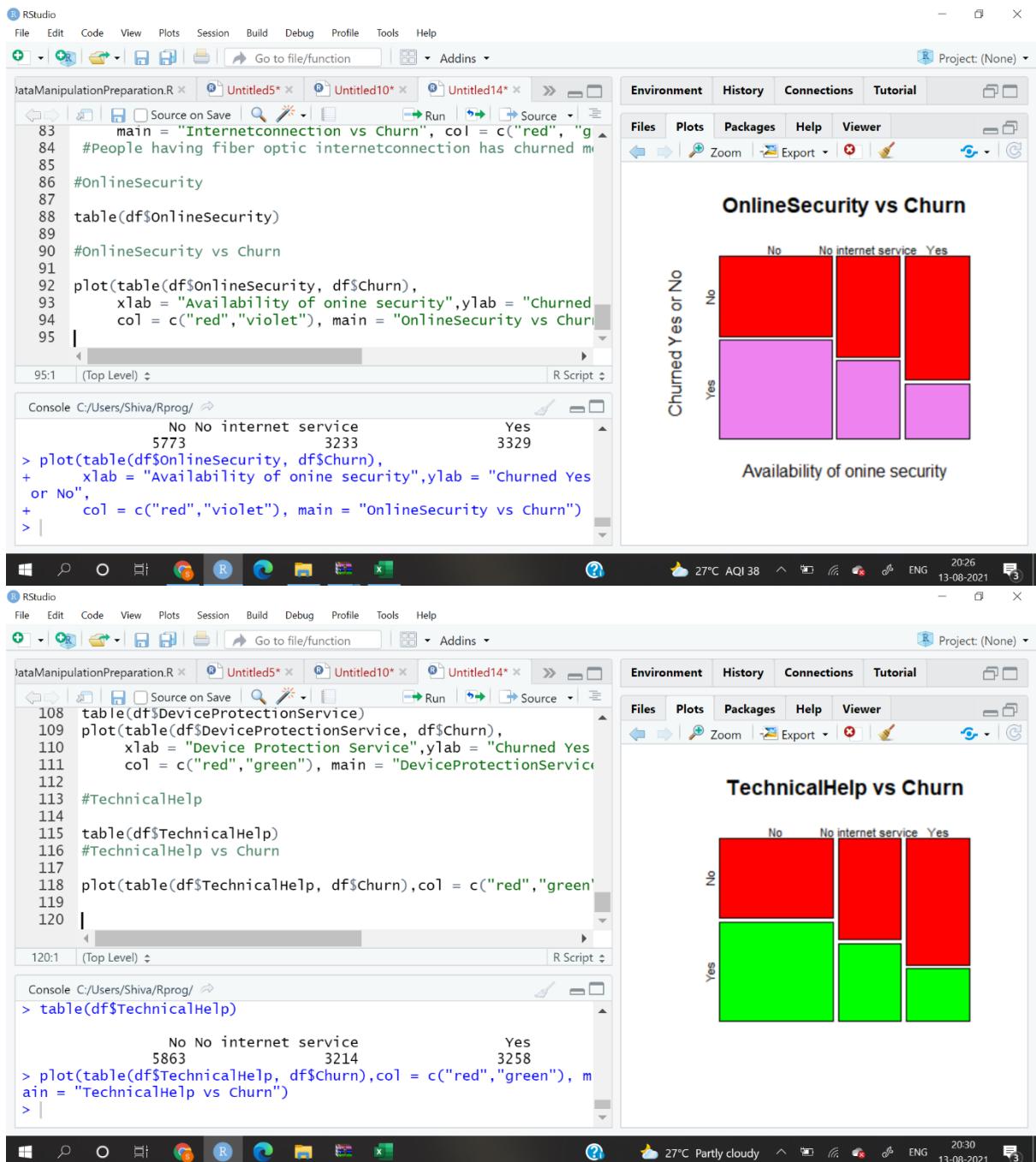
```

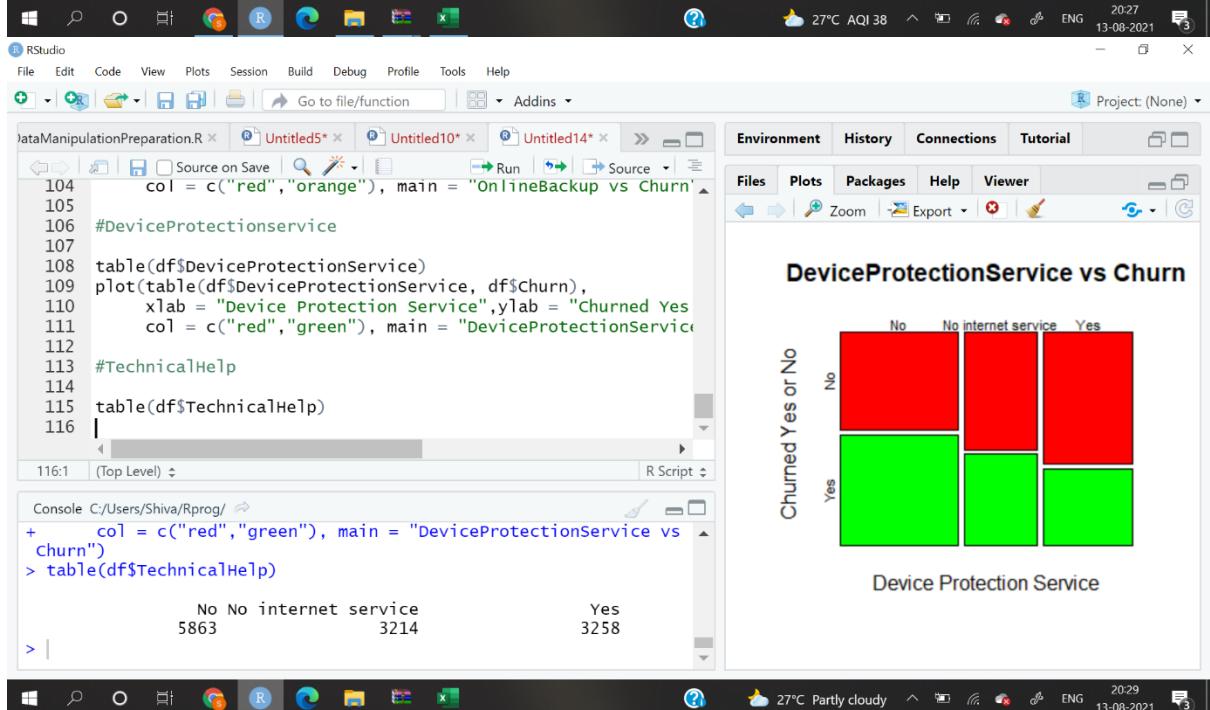
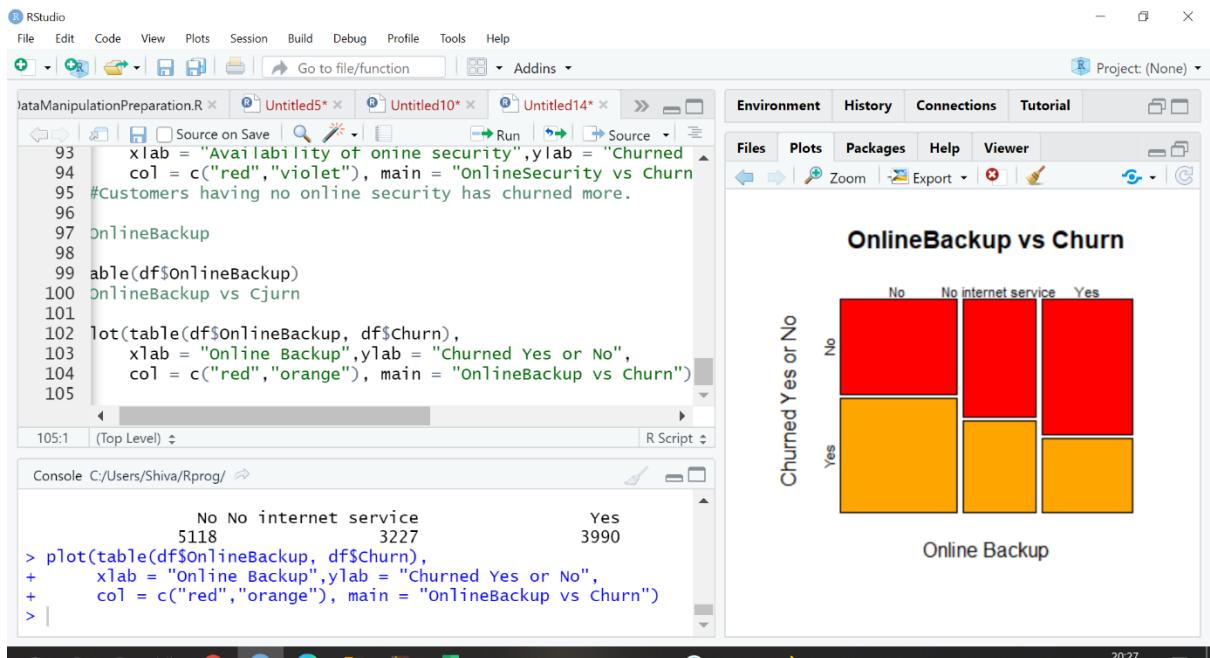
Distribution of tenure

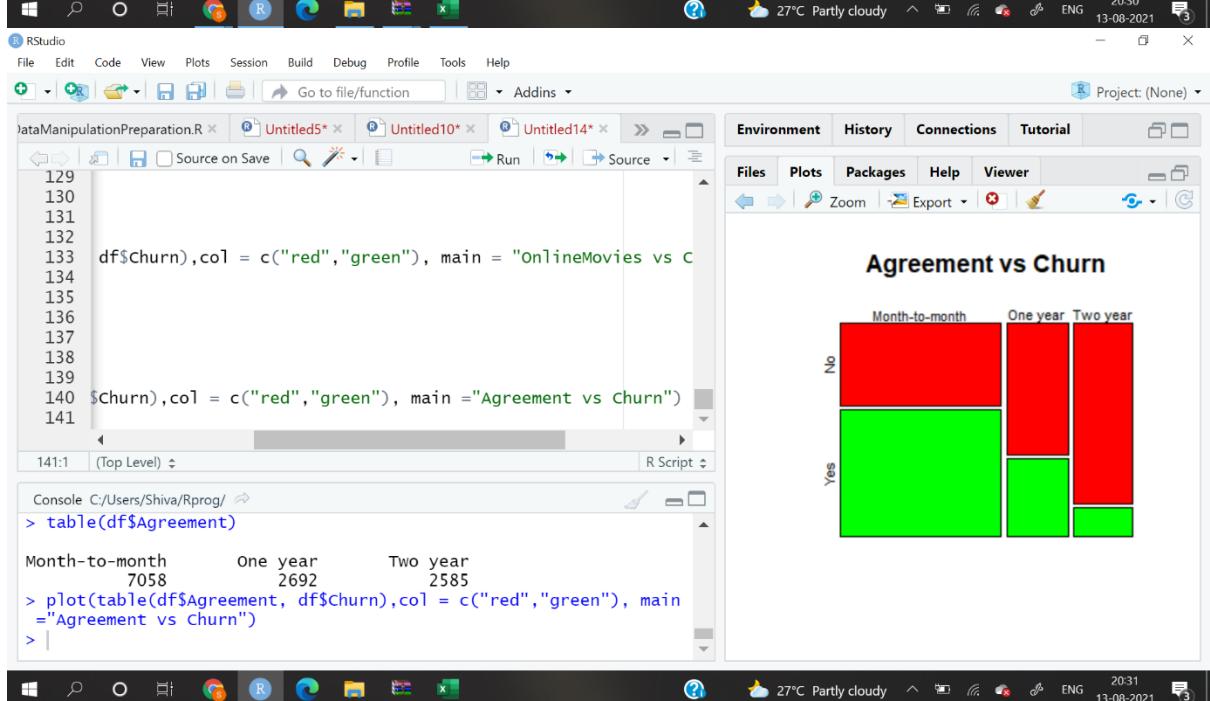
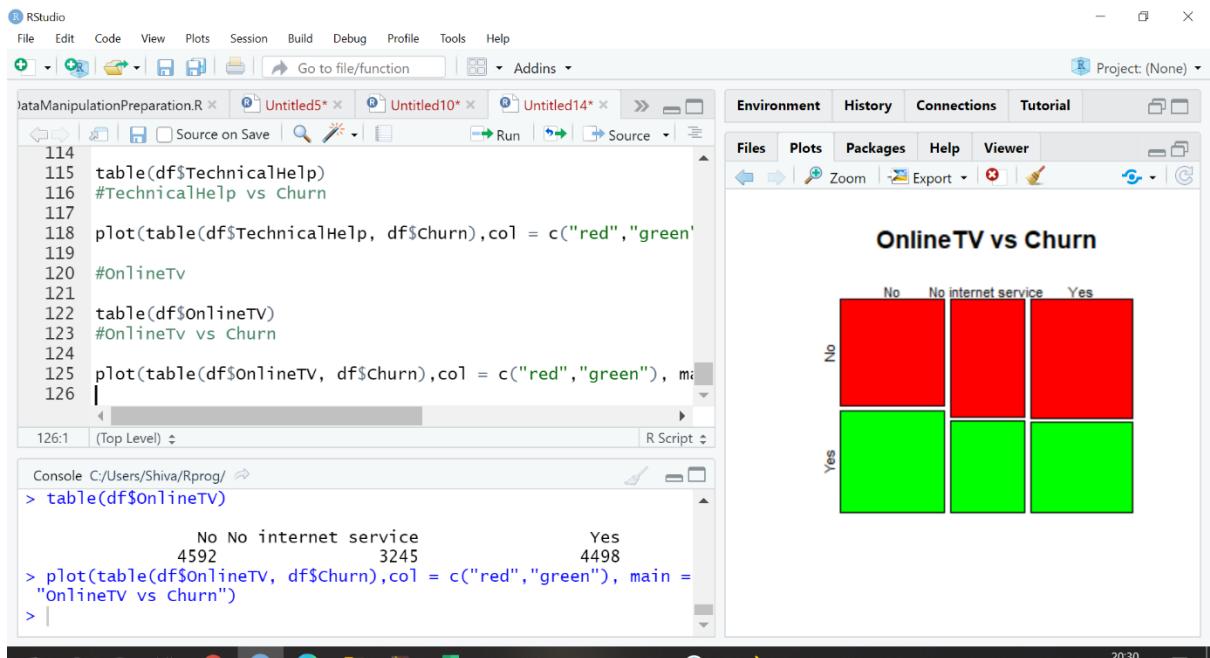
Frequency

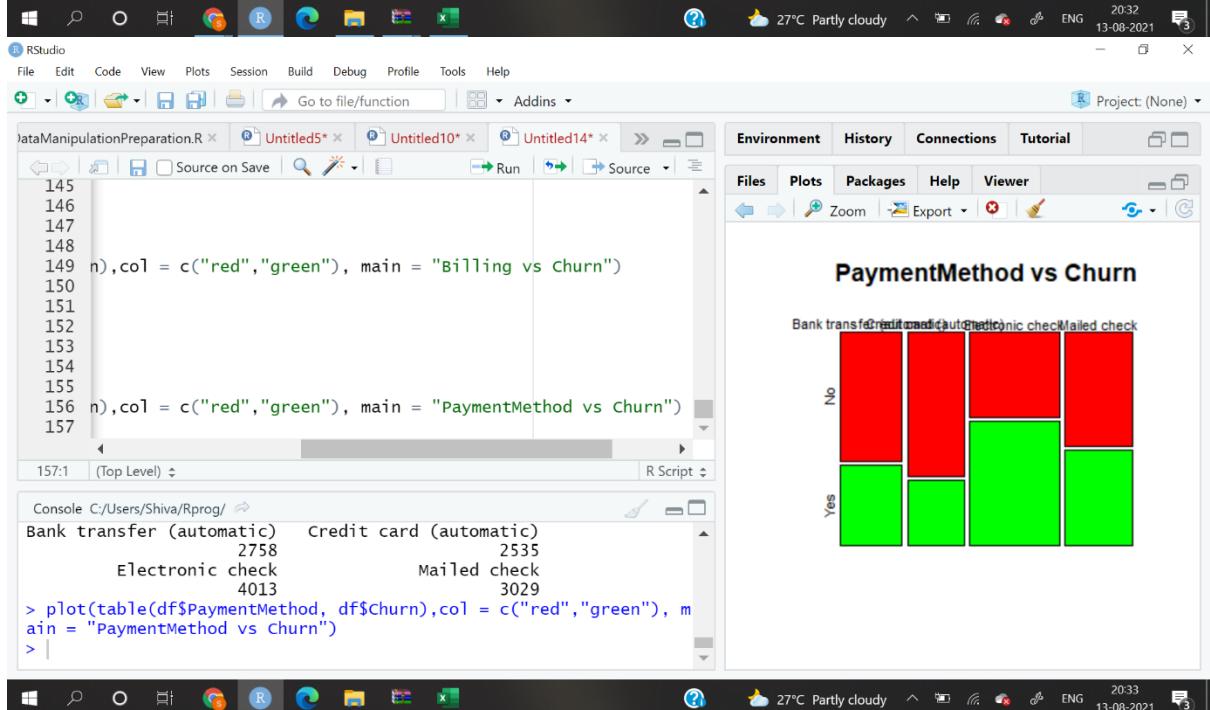
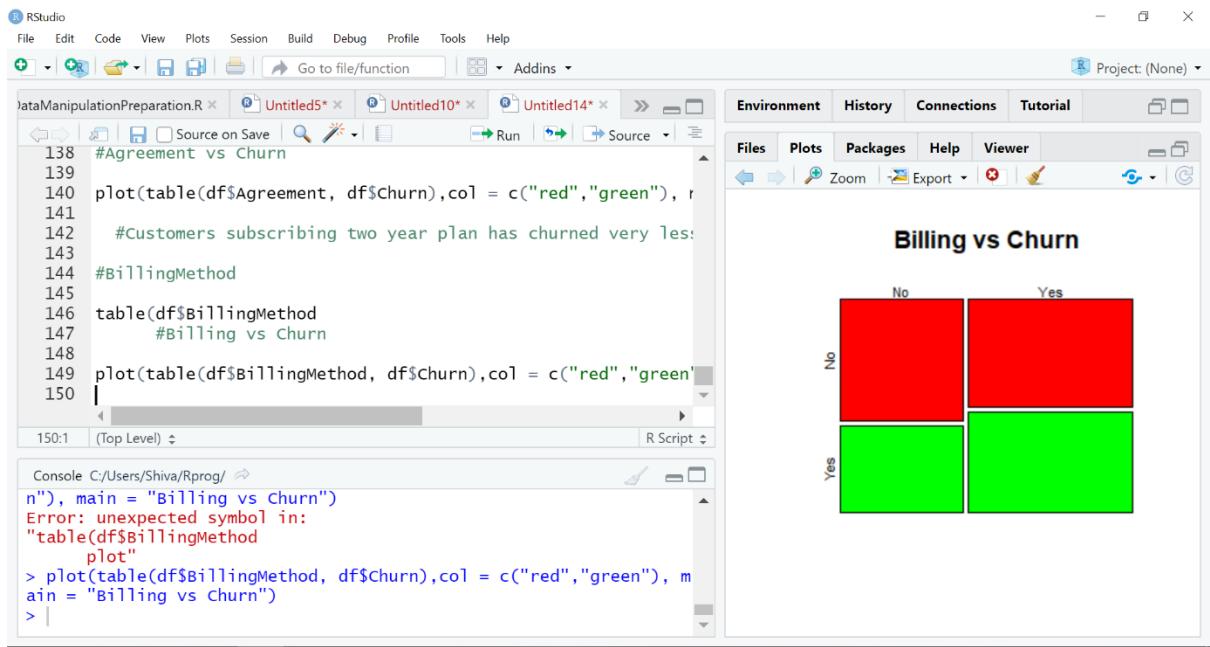
Tenure of the connection

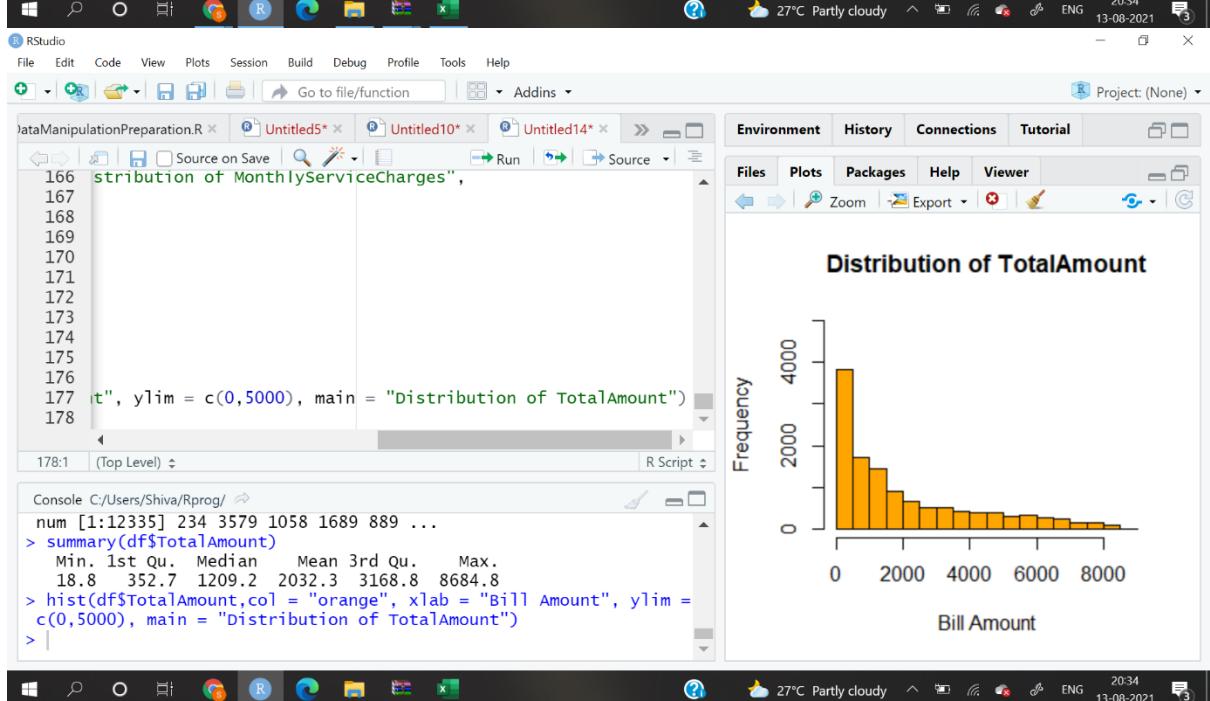
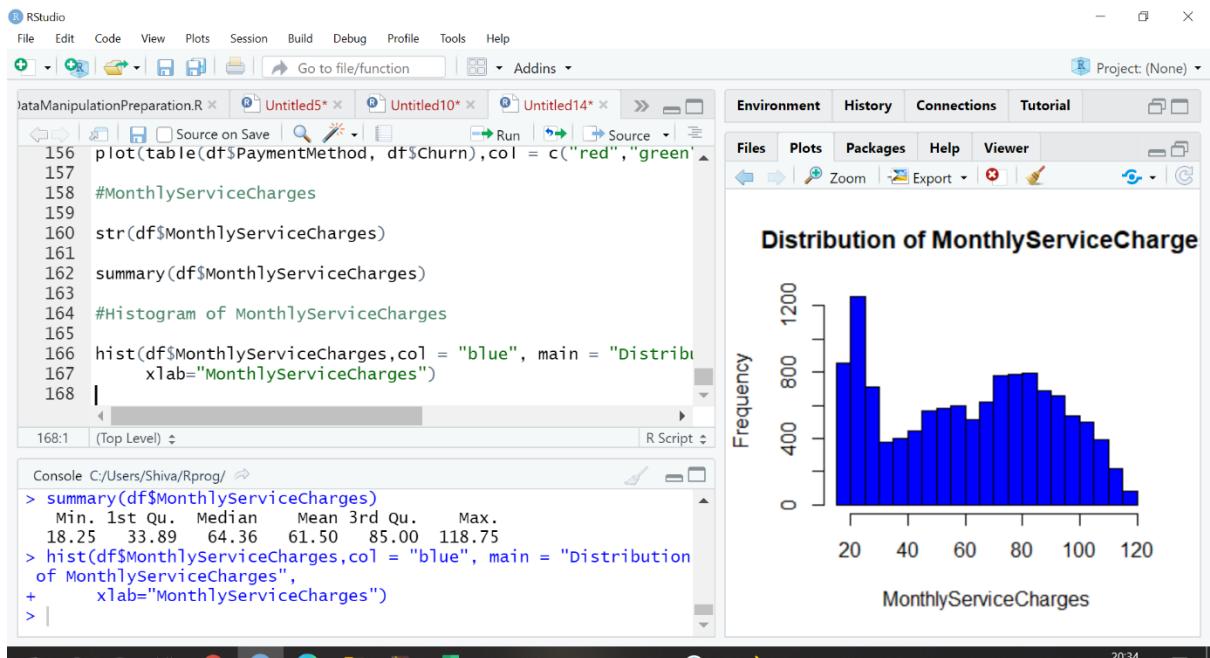












RStudio

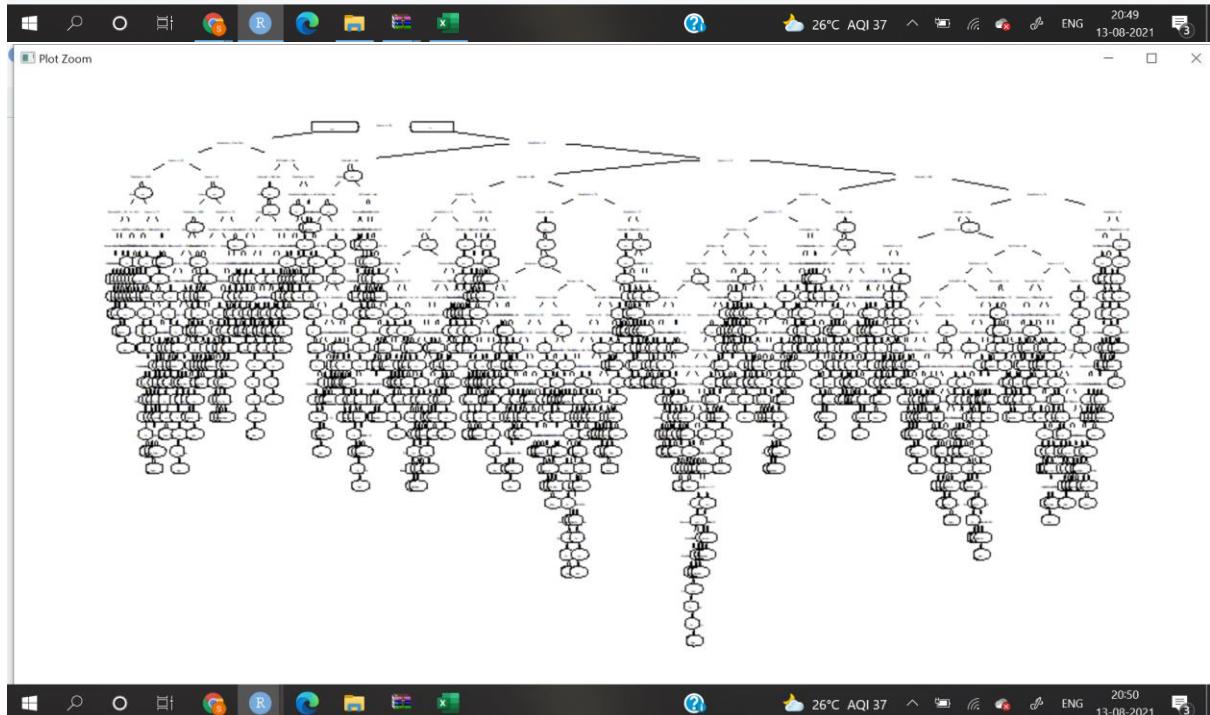
File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

```
219 #Dimension of training and testing datasets.
220
221 noquote("Dimension of training dataset:"); dim(train)
222
223 noquote("Dimension of testing dataset:"); dim(test)
224
225 #full grown tree model
226 dev.off()
227 full_model_tree <- rpart(Churn~, train, method = "class",minsplit=0,cp=0)
228 #plotting model
229
230 plot(full_model_tree)
231
```

231:1 (Top Level) R Script

```
Console C:/Users/Shiva/Rprog/
> dev.off()
null device
1
> full_model_tree <- rpart(Churn~, train, method = "class",minsplit=0,cp=0)
> plot(full_model_tree)
>
```



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

```
> train(churn~, data = train, method = "rpart", trControl = numt
1ds, tuneGrid = cpGrid)
CART

8635 samples
  17 predictor
   2 classes: 'No', 'Yes'

No pre-processing
Resampling: Cross-Validated (100 fold)
Summary of sample sizes: 8549, 8549, 8549, 8549, 8549, 8549, ...
Resampling results across tuning parameters:

  cp    Accuracy   Kappa
  0.01  0.7605194  0.5179565
  0.02  0.7504324  0.4985202
  0.03  0.6964814  0.4056634
  0.04  0.7057342  0.4280418
  0.05  0.7057342  0.4280418

Accuracy was used to select the optimal model using the
largest value.
The final value used for the model was cp = 0.01.
> |
```

Environment History Connections Tutorial

Files Plots Packages Help Viewer

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

DataManipulationPreparation.R

```
263 cpGrid <- expand.grid(cp=seq(0.01,.05,0.01))
264
265 cpGrid
266
267 ### Checking the cross validation accuracy for cp parameters
268
269 train(churn~, data = train, method = "rpart", trControl = n
270
271 ## Pruning the full_model_tree(fully grown model)
272
273 pruned_tree <- prune(full_model_tree, cp = 0.01)
274 prp(pruned_tree)
275
```

Console C:/Users/Shiva/Rprog/

```
4444444 0.75555556) *
  61) MonthlyServiceCharges< 43.34037 440 118 Yes (0.268
18182 0.73181818) *
  31) InternetConnection=Fiber optic, No 2307 472 Yes (0.20
459471 0.79540529) *
> prp(pruned_tree)
> |
```

Environment History Connections Tutorial

Files Plots Packages Help Viewer

