Telecom Churn Analysis

```r
library(readxl)
teldata <- read.csv("D:/Spring 2019/Multi Analysis/WA_Fn-UseC_-Telco-
Customer-Churn.csv")

summary(teldata)
```

```
##      customerID       gender      SeniorCitizen    Partner    Dependents
## 0002-ORFBO:   1   Female:3488   Min.   :0.0000   No :3641   No :4933
## 0003-MKNFE:   1   Male  :3555   1st Qu.:0.0000   Yes:3402   Yes:2110
## 0004-TLHLJ:   1                 Median :0.0000
## 0011-IGKFF:   1                 Mean   :0.1621
## 0013-EXCHZ:   1                 3rd Qu.:0.0000
## 0013-MHZWF:   1                 Max.   :1.0000
## (Other)   :7037
##     tenure        PhoneService          MultipleLines       InternetService
## Min.   : 0.00   No : 682    No              :3390   DSL        :2421
## 1st Qu.: 9.00   Yes:6361    No phone service: 682   Fiber optic:3096
## Median :29.00               Yes             :2971   No         :1526
## Mean   :32.37
## 3rd Qu.:55.00
## Max.   :72.00
##
##             OnlineSecurity              OnlineBackup
## No                 :3498   No                 :3088
## No internet service:1526   No internet service:1526
## Yes                :2019   Yes                :2429
##
##
##
##
##            DeviceProtection             TechSupport
## No                 :3095   No                 :3473
## No internet service:1526   No internet service:1526
## Yes                :2422   Yes                :2044
##
##
##
##
##              StreamingTV               StreamingMovies
## No                 :2810   No                 :2785
## No internet service:1526   No internet service:1526
## Yes                :2707   Yes                :2732
##
##
##
##
```

```
##           Contract    PaperlessBilling                  PaymentMethod
##   Month-to-month:3875  No :2872        Bank transfer (automatic):1544
##   One year      :1473  Yes:4171        Credit card (automatic)  :1522
##   Two year      :1695                  Electronic check         :2365
##                                        Mailed check             :1612
##
##
##
##   MonthlyCharges    TotalCharges     Churn
##   Min.   : 18.25   Min.   :  18.8   No :5174
##   1st Qu.: 35.50   1st Qu.: 401.4   Yes:1869
##   Median : 70.35   Median :1397.5
##   Mean   : 64.76   Mean   :2283.3
##   3rd Qu.: 89.85   3rd Qu.:3794.7
##   Max.   :118.75   Max.   :8684.8
##                    NA's   :11
```

```r
#install.packages("h2o")
library(h2o)
```

```
## Warning: package 'h2o' was built under R version 3.5.2

##
## ----------------------------------------------------------------------
##
## Your next step is to start H2O:
##     > h2o.init()
##
## For H2O package documentation, ask for help:
##     > ??h2o
##
## After starting H2O, you can use the Web UI at http://localhost:54321
## For more information visit http://docs.h2o.ai
##
## ----------------------------------------------------------------------

##
## Attaching package: 'h2o'

## The following objects are masked from 'package:stats':
##
##     cor, sd, var

## The following objects are masked from 'package:base':
##
##     %*%, %in%, &&, ||, apply, as.factor, as.numeric, colnames,
##     colnames<-, ifelse, is.character, is.factor, is.numeric, log,
##     log10, log1p, log2, round, signif, trunc
```

```r
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:h2o':
##
##      hour, month, week, year

library(purrr)

##
## Attaching package: 'purrr'

## The following object is masked from 'package:data.table':
##
##      transpose

library(ggplot2)
#install.packages("hrbrthemes")
library(hrbrthemes)

## Warning: package 'hrbrthemes' was built under R version 3.5.2

## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use
these themes.

##       Please use hrbrthemes::import_roboto_condensed() to install Roboto
Condensed and

##       if Arial Narrow is not on your system, please see
http://bit.ly/arialnarrow

library(tidyverse)

## -- Attaching packages ------------------------------------------------------
---------------------------- tidyverse 1.2.1 --

## v tibble  1.4.2      v dplyr   0.7.6
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## -- Conflicts ---------------------------------------------------------------
---------------------- tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()

library(MASS)

##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select

library(car)

## Warning: package 'car' was built under R version 3.5.2

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some

library(e1071)
library(caret)

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 3.5.2

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
##     lift

#install.packages("cowplot")
library(cowplot)

## Warning: package 'cowplot' was built under R version 3.5.2

##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggplot2':
##
##     ggsave

library(caTools)
library(pROC)

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'
```

```
## The following object is masked from 'package:h2o':
##
##     var

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```r
#install.packages("ggcorrplot")
library(ggcorrplot)
```

```
## Warning: package 'ggcorrplot' was built under R version 3.5.2
```

```r
str(teldata)
```

```
## 'data.frame':    7043 obs. of  21 variables:
##  $ customerID      : Factor w/ 7043 levels "0002-ORFBO","0003-MKNFE",..:
5376 3963 2565 5536 6512 6552 1003 4771 5605 4535 ...
##  $ gender          : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 1 2 1 1
2 ...
##  $ SeniorCitizen   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Partner         : Factor w/ 2 levels "No","Yes": 2 1 1 1 1 1 1 1 2 1
...
##  $ Dependents      : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 2 1 1 2
...
##  $ tenure          : int  1 34 2 45 2 8 22 10 28 62 ...
##  $ PhoneService    : Factor w/ 2 levels "No","Yes": 1 2 2 1 2 2 2 1 2 2
...
##  $ MultipleLines   : Factor w/ 3 levels "No","No phone service",..: 2 1 1
2 1 3 3 2 3 1 ...
##  $ InternetService : Factor w/ 3 levels "DSL","Fiber optic",..: 1 1 1 1 2
2 2 1 2 1 ...
##  $ OnlineSecurity  : Factor w/ 3 levels "No","No internet service",..: 1 3
3 3 1 1 1 3 1 3 ...
##  $ OnlineBackup    : Factor w/ 3 levels "No","No internet service",..: 3 1
3 1 1 1 3 1 1 3 ...
##  $ DeviceProtection: Factor w/ 3 levels "No","No internet service",..: 1 3
1 3 1 3 1 1 3 1 ...
##  $ TechSupport     : Factor w/ 3 levels "No","No internet service",..: 1 1
1 3 1 1 1 1 3 1 ...
##  $ StreamingTV     : Factor w/ 3 levels "No","No internet service",..: 1 1
1 1 3 3 1 3 1 ...
##  $ StreamingMovies : Factor w/ 3 levels "No","No internet service",..: 1 1
1 1 3 1 1 3 1 ...
##  $ Contract        : Factor w/ 3 levels "Month-to-month",..: 1 2 1 2 1 1 1
1 1 2 ...
##  $ PaperlessBilling: Factor w/ 2 levels "No","Yes": 2 1 2 1 2 2 2 1 2 1
...
##  $ PaymentMethod   : Factor w/ 4 levels "Bank transfer (automatic)",..: 3
4 4 1 3 3 2 4 3 1 ...
##  $ MonthlyCharges  : num  29.9 57 53.9 42.3 70.7 ...
```

```
##  $ TotalCharges    : num  29.9 1889.5 108.2 1840.8 151.7 ...
##  $ Churn           : Factor w/ 2 levels "No","Yes": 1 1 2 1 2 2 1 1 2 1
...
```

```
head(teldata)
```

```
##   customerID gender SeniorCitizen Partner Dependents tenure PhoneService
## 1 7590-VHVEG Female             0     Yes         No      1           No
## 2 5575-GNVDE   Male             0      No         No     34          Yes
## 3 3668-QPYBK   Male             0      No         No      2          Yes
## 4 7795-CFOCW   Male             0      No         No     45           No
## 5 9237-HQITU Female             0      No         No      2          Yes
## 6 9305-CDSKC Female             0      No         No      8          Yes
##      MultipleLines InternetService OnlineSecurity OnlineBackup
## 1 No phone service             DSL             No          Yes
## 2               No             DSL            Yes           No
## 3               No             DSL            Yes          Yes
## 4 No phone service             DSL            Yes           No
## 5               No     Fiber optic             No           No
## 6              Yes     Fiber optic             No           No
##   DeviceProtection TechSupport StreamingTV StreamingMovies        Contract
## 1               No          No          No              No Month-to-month
## 2              Yes          No          No              No       One year
## 3               No          No          No              No Month-to-month
## 4              Yes         Yes          No              No       One year
## 5               No          No          No              No Month-to-month
## 6              Yes          No         Yes             Yes Month-to-month
##   PaperlessBilling            PaymentMethod MonthlyCharges TotalCharges
## 1              Yes         Electronic check          29.85        29.85
## 2               No             Mailed check          56.95      1889.50
## 3              Yes             Mailed check          53.85       108.15
## 4               No Bank transfer (automatic)          42.30      1840.75
## 5              Yes         Electronic check          70.70       151.65
## 6              Yes         Electronic check          99.65       820.50
##   Churn
## 1    No
## 2    No
## 3   Yes
## 4    No
## 5   Yes
## 6   Yes
```

Checking for NA values

```
colSums(is.na(teldata))
```

```
##       customerID           gender    SeniorCitizen          Partner
##                0                0                0                0
##       Dependents           tenure     PhoneService    MultipleLines
##                0                0                0                0
##  InternetService   OnlineSecurity     OnlineBackup DeviceProtection
```

```
##                    0               0               0               0
##        TechSupport      StreamingTV  StreamingMovies        Contract
##                    0               0               0               0
## PaperlessBilling    PaymentMethod   MonthlyCharges     TotalCharges
##                    0               0               0              11
##             Churn
##                    0
```

Displaying NA values

```
telmissing<-teldata[is.na(teldata$TotalCharges),]
telmissing
```

```
##       customerID gender SeniorCitizen Partner Dependents tenure
## 489   4472-LVYGI Female             0     Yes        Yes      0
## 754   3115-CZMZD   Male             0      No        Yes      0
## 937   5709-LVOEQ Female             0     Yes        Yes      0
## 1083  4367-NUYAO   Male             0     Yes        Yes      0
## 1341  1371-DWPAZ Female             0     Yes        Yes      0
## 3332  7644-OMVMY   Male             0     Yes        Yes      0
## 3827  3213-VVOLG   Male             0     Yes        Yes      0
## 4381  2520-SGTTA Female             0     Yes        Yes      0
## 5219  2923-ARZLG   Male             0     Yes        Yes      0
## 6671  4075-WKNIU Female             0     Yes        Yes      0
## 6755  2775-SEFEE   Male             0      No        Yes      0
##       PhoneService     MultipleLines InternetService      OnlineSecurity
## 489             No No phone service             DSL                 Yes
## 754            Yes                No              No No internet service
## 937            Yes                No             DSL                 Yes
## 1083           Yes               Yes              No No internet service
## 1341            No No phone service             DSL                 Yes
## 3332           Yes                No              No No internet service
## 3827           Yes               Yes              No No internet service
## 4381           Yes                No              No No internet service
## 5219           Yes                No              No No internet service
## 6671           Yes               Yes             DSL                  No
## 6755           Yes               Yes             DSL                 Yes
##            OnlineBackup    DeviceProtection         TechSupport
## 489                  No                 Yes                 Yes
## 754  No internet service No internet service No internet service
## 937                 Yes                 Yes                  No
## 1083 No internet service No internet service No internet service
## 1341                Yes                 Yes                 Yes
## 3332 No internet service No internet service No internet service
## 3827 No internet service No internet service No internet service
## 4381 No internet service No internet service No internet service
## 5219 No internet service No internet service No internet service
## 6671                Yes                 Yes                 Yes
## 6755                Yes                  No                 Yes
##            StreamingTV    StreamingMovies Contract PaperlessBilling
```

```
## 489                     Yes                   No Two year              Yes
## 754  No internet service No internet service Two year              No
## 937                     Yes                  Yes Two year              No
## 1083 No internet service No internet service Two year              No
## 1341                    Yes                   No Two year              No
## 3332 No internet service No internet service Two year              No
## 3827 No internet service No internet service Two year              No
## 4381 No internet service No internet service Two year              No
## 5219 No internet service No internet service One year              Yes
## 6671                     Yes                   No Two year              No
## 6755                      No                   No Two year              Yes
##                 PaymentMethod MonthlyCharges TotalCharges Churn
## 489  Bank transfer (automatic)          52.55           NA   No
## 754             Mailed check          20.25           NA   No
## 937             Mailed check          80.85           NA   No
## 1083            Mailed check          25.75           NA   No
## 1341  Credit card (automatic)          56.05           NA   No
## 3332            Mailed check          19.85           NA   No
## 3827            Mailed check          25.35           NA   No
## 4381            Mailed check          20.00           NA   No
## 5219            Mailed check          19.70           NA   No
## 6671            Mailed check          73.35           NA   No
## 6755 Bank transfer (automatic)          61.90           NA   No
```

The tenture for the missing values in the total charges are all 0. This means they are new customers. Hence these can be removed.

```r
teldata <- na.exclude(teldata)
```

Checking if the NA vals were removed

```r
colSums(is.na(teldata))
```

```
##       customerID           gender     SeniorCitizen          Partner
##                0                0                 0                0
##       Dependents           tenure       PhoneService     MultipleLines
##                0                0                 0                0
##  InternetService   OnlineSecurity      OnlineBackup DeviceProtection
##                0                0                 0                0
##      TechSupport       StreamingTV    StreamingMovies          Contract
##                0                0                 0                0
## PaperlessBilling     PaymentMethod     MonthlyCharges      TotalCharges
##                0                0                 0                0
##            Churn
##                0
```

DATA VISUALIZATION

```r
library(ggplot2)
library(dplyr)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine

b1 <- ggplot(teldata, aes(gender,fill=Churn)) + geom_bar(position='fill')
#+theme(legend.position="none")

b2 <- ggplot(teldata, aes(SeniorCitizen, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b3 <- ggplot(teldata, aes(Partner, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b4 <- ggplot(teldata, aes(Dependents, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b5 <- ggplot(teldata, aes(PhoneService, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b6 <- ggplot(teldata, aes(MultipleLines, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b7 <- ggplot(teldata, aes(InternetService, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b8 <- ggplot(teldata, aes(OnlineSecurity, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b9 <- ggplot(teldata, aes(OnlineBackup, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b10 <- ggplot(teldata, aes(DeviceProtection, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")
b11 <- ggplot(teldata, aes(TechSupport, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b12 <- ggplot(teldata, aes(StreamingTV, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b13 <- ggplot(teldata, aes(StreamingMovies, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b14 <- ggplot(teldata, aes(Contract, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")

b15 <- ggplot(teldata, aes(PaperlessBilling, fill = Churn)) +
```

```
geom_bar(position='fill')+theme(legend.position="none")

b16 <- ggplot(teldata, aes(PaymentMethod, fill = Churn)) +
geom_bar(position='fill')+theme(legend.position="none")


grid.arrange(b1,b2,b3, ncol = 1)
```



```
grid.arrange(b4,b5,b6, ncol = 1)
```
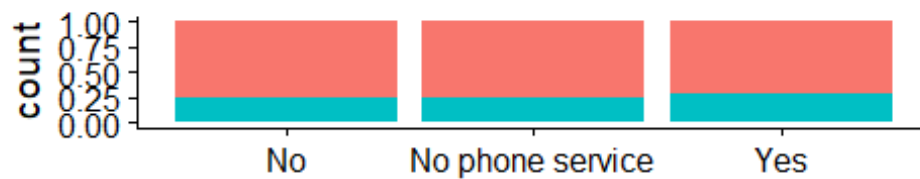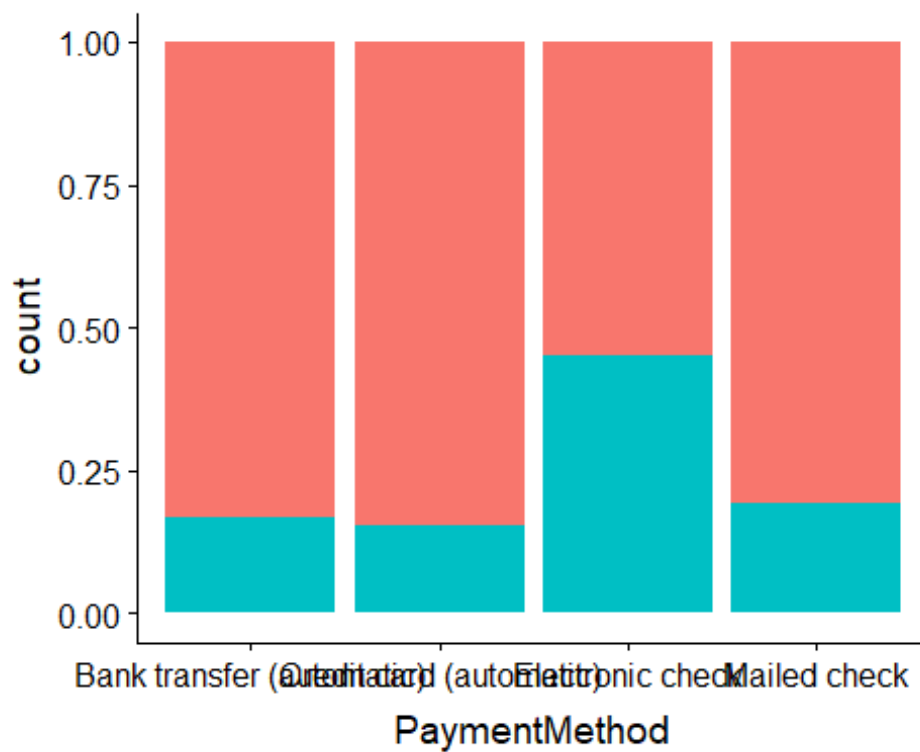
```
grid.arrange(b7,b8,b9,b10, ncol = 2)
```
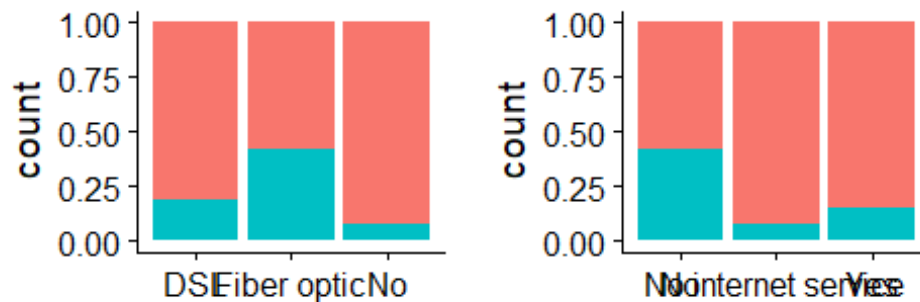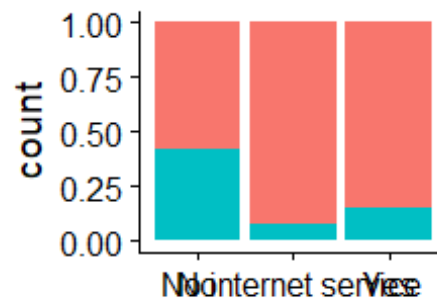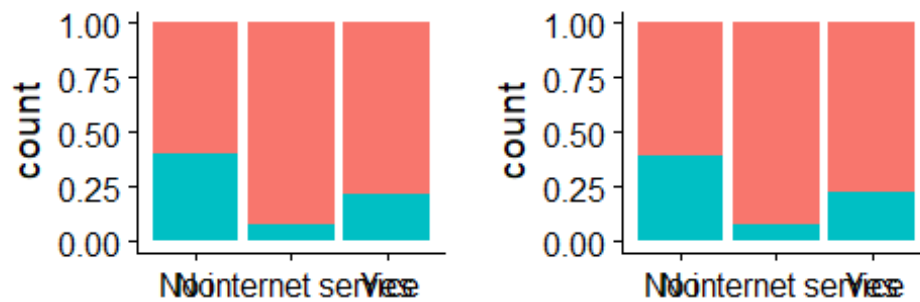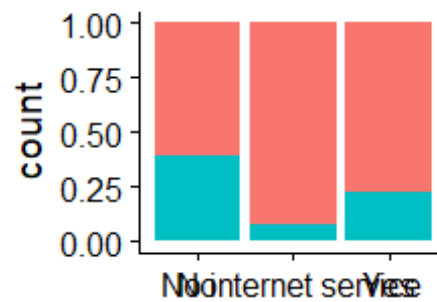


```
grid.arrange(b11,b12,b13,b14,b15, ncol = 2)
```

b16
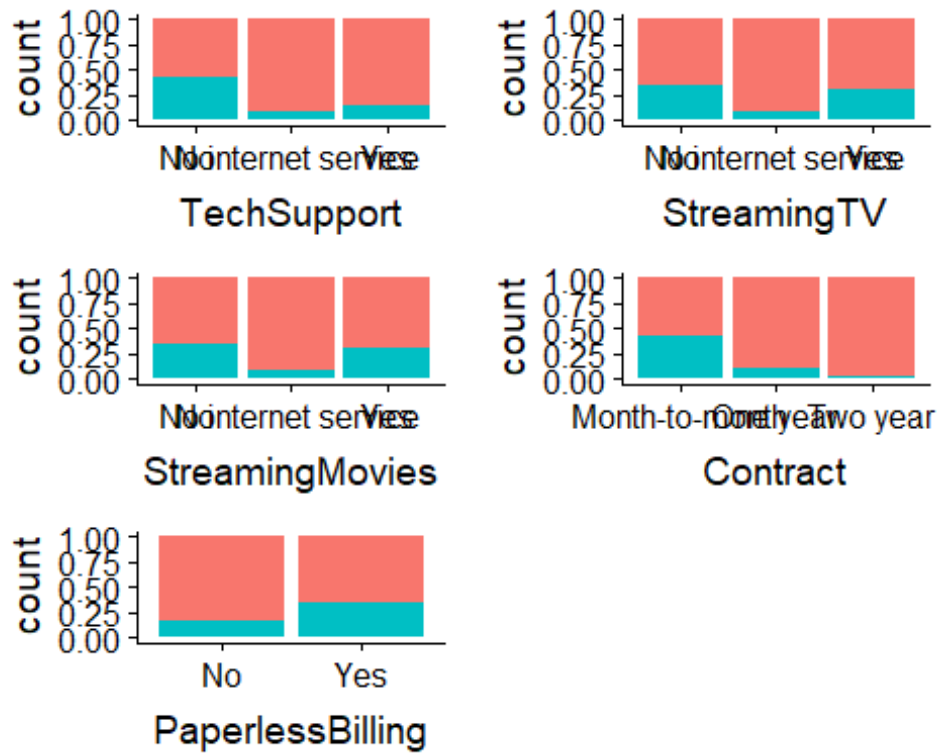

```r
boxplot(teldata$TotalCharges,data=teldata, main="Total Charges")
```
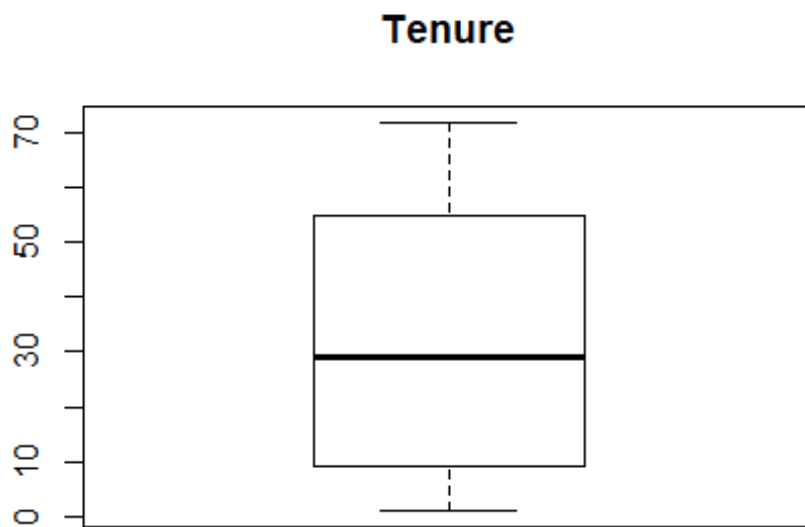
Total Charges
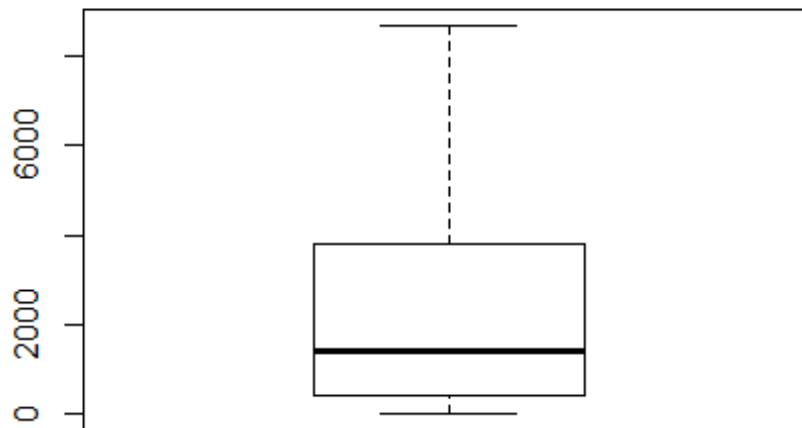

```
boxplot(teldata$MonthlyCharges,data=teldata, main="Monthly Charges")
```

Monthly Charges
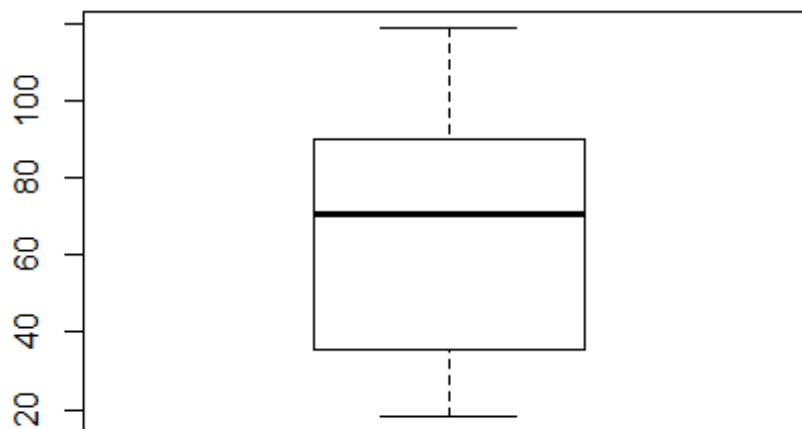

```
boxplot(teldata$tenure,data=teldata, main="Tenure")
```
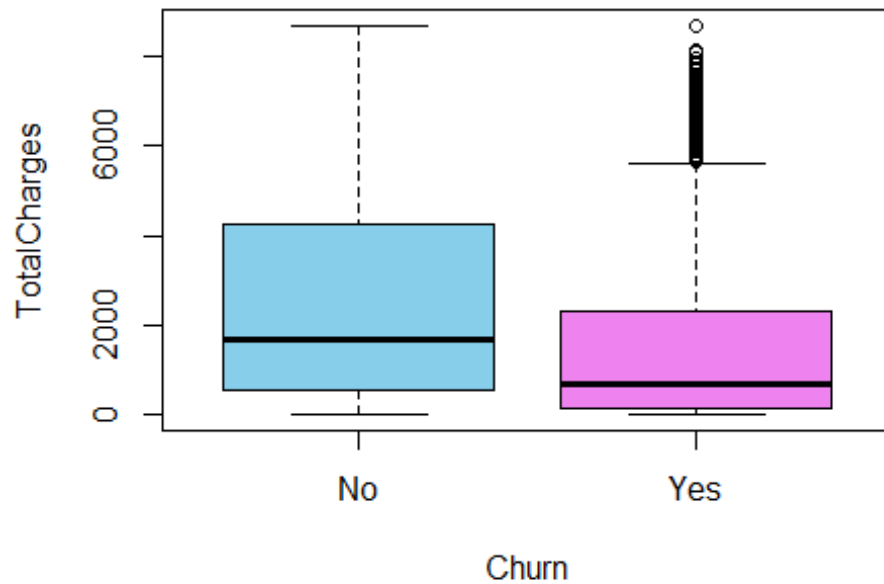
Tenure

The we look at the 3 variables based on churn.

```
b1 <- boxplot(tenure~Churn,data = teldata,col = c("skyblue","violet"), xlab
="Churn" , ylab = "tenure")
```
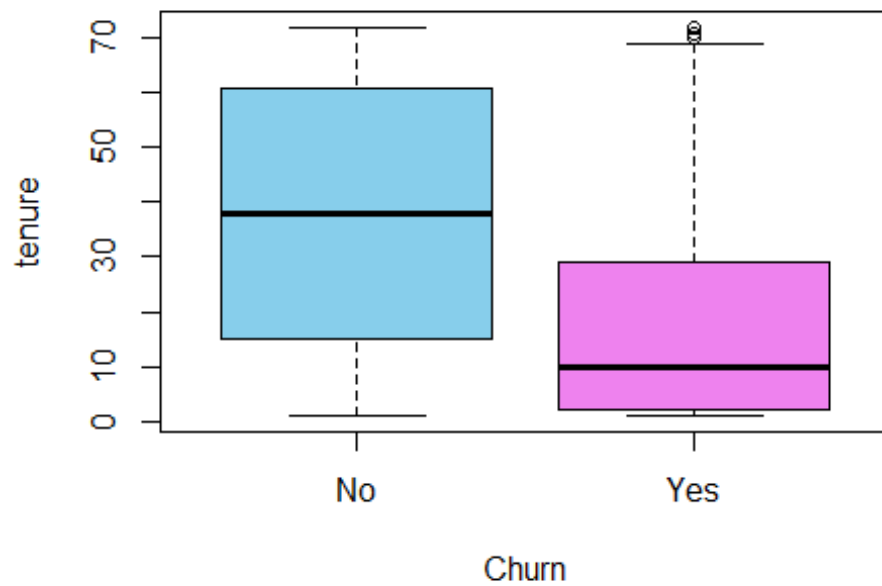
```
b2 <- boxplot(MonthlyCharges~Churn,data = teldata,col =
c("skyblue","violet"), xlab ="Churn" , ylab = "MonthlyCharges")
```
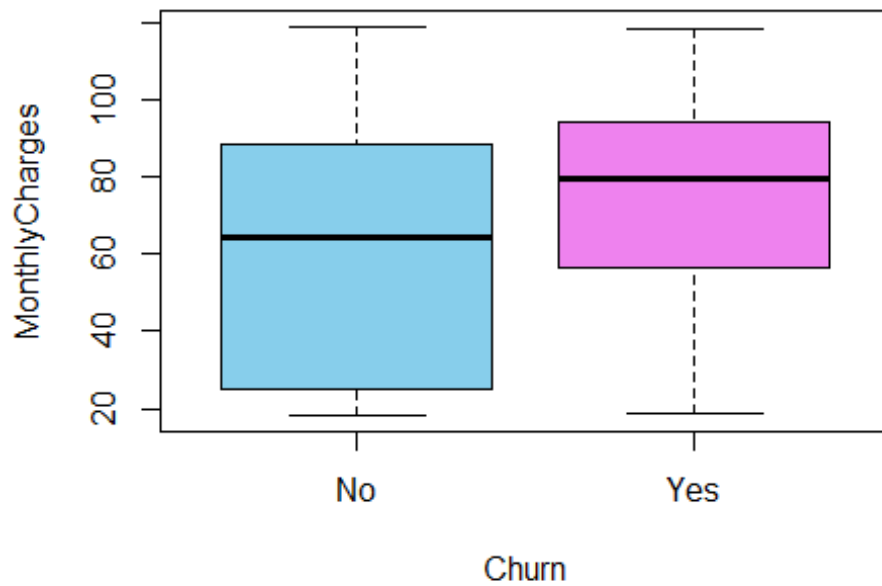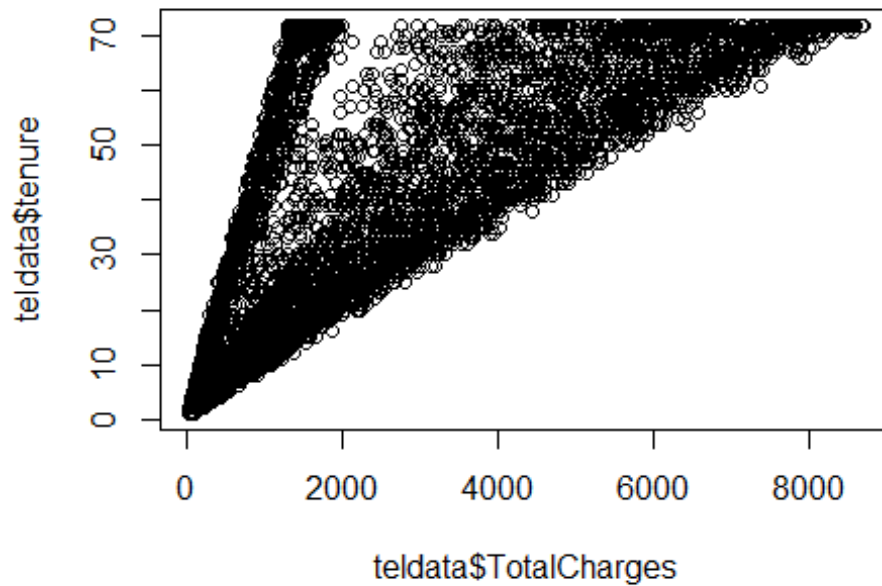
```
b3 <- boxplot(TotalCharges~Churn,data = teldata,col = c("skyblue","violet"),
xlab ="Churn" , ylab = "TotalCharges")
```



```
#ggplot(teldata, aes(xlab = Churn, ylab = TotalCharges)) + geom_boxplot(fill
= Churn)
 #grid.arrange(b1,b2,b3,ncol = 2)

plot(teldata$TotalCharges, teldata$tenure)
```
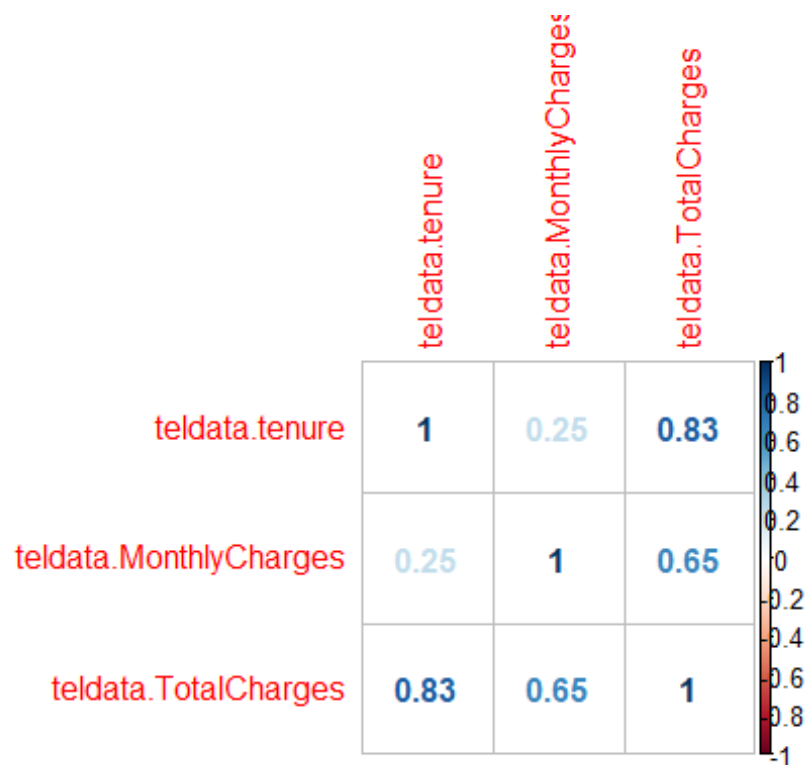
```r
library(corrplot)

## corrplot 0.84 loaded

cor_data <-
data.frame(teldata$tenure,teldata$MonthlyCharges,teldata$TotalCharges)
corr <- cor(cor_data)
corrplot(corr, method = "number")
```

Correlation matrix:

	teldata.tenure	teldata.MonthlyCharges	teldata.TotalCharges
teldata.tenure	1	0.25	0.83
teldata.MonthlyCharges	0.25	1	0.65
teldata.TotalCharges	0.83	0.65	1

```
hist(teldata$tenure, main="Tenure Distribution",
    xlab="Tenure (Months)")
```

Tenure Distribution
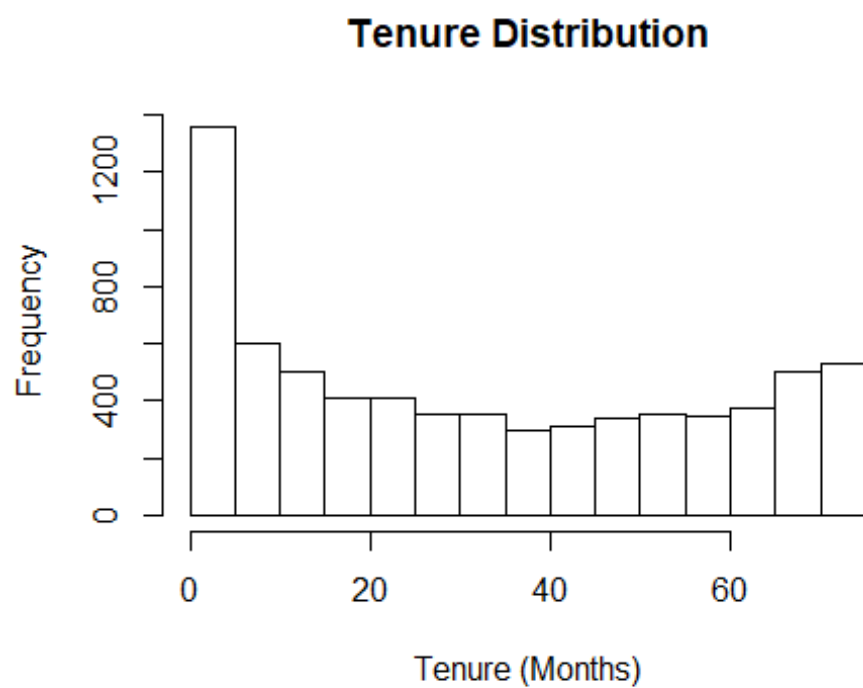
```
hist(teldata$MonthlyCharges, main="Distribution of Monthly Charges",
xlab="Monthly Charges")
```


Distribution of Monthly Charges
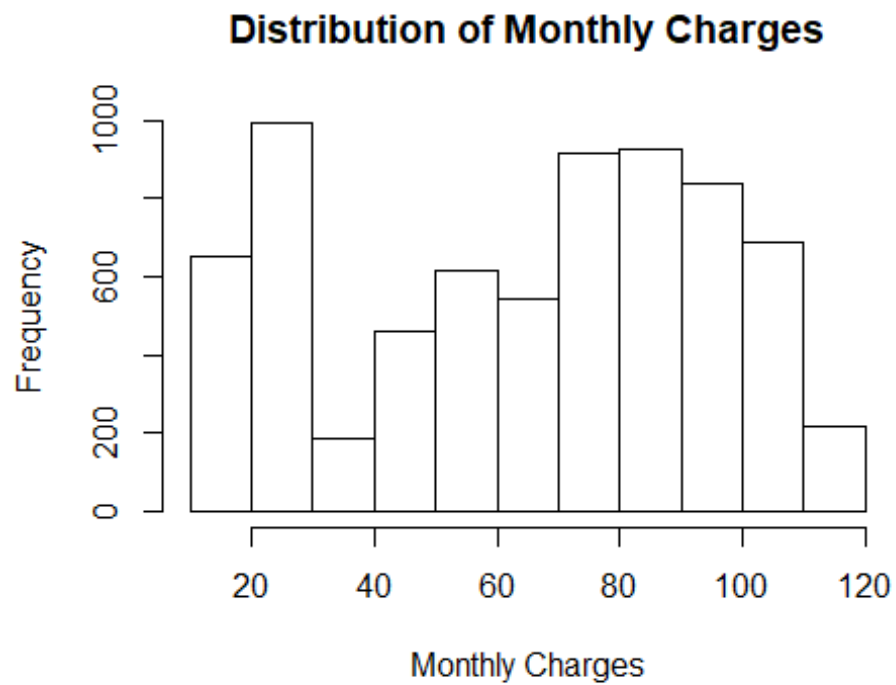
```
hist(teldata$TotalCharges, main="Distribution of Monthly Charges",
xlab="Monthly Charges")
```

Distribution of Monthly Charges