logistic_regression.R

sigsp

2022-03-03

```r
## Author: Stephen E. Porter
## Title: Logistic Regression
## Course: WGU D208: Predictive Modeling
## Instructor: Dr. William Sewell
options(warn=-1)

# Libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(broom)
library(caret)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

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

```r
library(ggplot2)
library(fastDummies)
library(car)
```

```
## 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
```

```r
library(Hmisc)
```

```
## Loading required package: survival

##
## Attaching package: 'survival'

## The following object is masked from 'package:caret':
##
##     cluster

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```r
library(cowplot)

# Import CSV as data frame
df <- read.csv(file = 'C:/WGU/D208 Predictive Modeling/PA Task 2/churn_clean.csv')

# Checking for nulls
sapply(df, function(x) sum(is.na(x)))
```

```
##            CaseOrder          Customer_id          Interaction
##                    0                    0                    0
##                  UID                 City                State
##                    0                    0                    0
##               County                  Zip                  Lat
##                    0                    0                    0
##                  Lng           Population                 Area
##                    0                    0                    0
##             TimeZone                  Job             Children
##                    0                    0                    0
##                  Age               Income              Marital
##                    0                    0                    0
##               Gender                Churn   Outage_sec_perweek
##                    0                    0                    0
##                Email             Contacts Yearly_equip_failure
##                    0                    0                    0
##               Techie             Contract           Port_modem
##                    0                    0                    0
##               Tablet      InternetService                Phone
##                    0                    0                    0
##             Multiple       OnlineSecurity         OnlineBackup
##                    0                    0                    0
```

```
##     DeviceProtection          TechSupport          StreamingTV
##                    0                    0                    0
##      StreamingMovies     PaperlessBilling        PaymentMethod
##                    0                    0                    0
##               Tenure        MonthlyCharge    Bandwidth_GB_Year
##                    0                    0                    0
##                Item1                Item2                Item3
##                    0                    0                    0
##                Item4                Item5                Item6
##                    0                    0                    0
##                Item7                Item8
##                    0                    0
```

```
dim(df)
```

```
## [1] 10000    50
```

```
str(df)
```

```
## 'data.frame':    10000 obs. of  50 variables:
##  $ CaseOrder           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Customer_id         : chr  "K409198" "S120509" "K191035" "D90850" ...
##  $ Interaction         : chr  "aa90260b-4141-4a24-8e36-b04ce1f4f77b" "fb76459f-c047-4a9d-8af9-e0f7d4a
##  $ UID                 : chr  "e885b299883d4f9fb18e39c75155d990" "f2de8bef964785f41a2959829830fb8a" "
##  $ City                : chr  "Point Baker" "West Branch" "Yamhill" "Del Mar" ...
##  $ State               : chr  "AK" "MI" "OR" "CA" ...
##  $ County              : chr  "Prince of Wales-Hyder" "Ogemaw" "Yamhill" "San Diego" ...
##  $ Zip                 : int  99927 48661 97148 92014 77461 31030 37847 73109 34771 45237 ...
##  $ Lat                 : num  56.3 44.3 45.4 33 29.4 ...
##  $ Lng                 : num  -133.4 -84.2 -123.2 -117.2 -95.8 ...
##  $ Population          : int  38 10446 3735 13863 11352 17701 2535 23144 17351 20193 ...
##  $ Area                : chr  "Urban" "Urban" "Urban" "Suburban" ...
##  $ TimeZone            : chr  "America/Sitka" "America/Detroit" "America/Los_Angeles" "America/Los_An
##  $ Job                 : chr  "Environmental health practitioner" "Programmer, multimedia" "Chief Fin
##  $ Children            : int  0 1 4 1 0 3 0 2 2 1 ...
##  $ Age                 : int  68 27 50 48 83 83 79 30 49 86 ...
##  $ Income              : num  28562 21705 9610 18925 40074 ...
##  $ Marital             : chr  "Widowed" "Married" "Widowed" "Married" ...
##  $ Gender              : chr  "Male" "Female" "Female" "Male" ...
##  $ Churn               : chr  "No" "Yes" "No" "No" ...
##  $ Outage_sec_perweek  : num  7.98 11.7 10.75 14.91 8.15 ...
##  $ Email               : int  10 12 9 15 16 15 10 16 20 18 ...
##  $ Contacts            : int  0 0 0 2 2 3 0 0 2 1 ...
##  $ Yearly_equip_failure: int  1 1 1 0 1 1 1 0 3 0 ...
##  $ Techie              : chr  "No" "Yes" "Yes" "Yes" ...
##  $ Contract            : chr  "One year" "Month-to-month" "Two Year" "Two Year" ...
##  $ Port_modem          : chr  "Yes" "No" "Yes" "No" ...
##  $ Tablet              : chr  "Yes" "Yes" "No" "No" ...
##  $ InternetService     : chr  "Fiber Optic" "Fiber Optic" "DSL" "DSL" ...
##  $ Phone               : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Multiple            : chr  "No" "Yes" "Yes" "No" ...
##  $ OnlineSecurity      : chr  "Yes" "Yes" "No" "Yes" ...
##  $ OnlineBackup        : chr  "Yes" "No" "No" "No" ...
##  $ DeviceProtection    : chr  "No" "No" "No" "No" ...
##  $ TechSupport         : chr  "No" "No" "No" "No" ...
##  $ StreamingTV         : chr  "No" "Yes" "No" "Yes" ...
```

```
##  $ StreamingMovies     : chr  "Yes" "Yes" "Yes" "No" ...
##  $ PaperlessBilling    : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ PaymentMethod       : chr  "Credit Card (automatic)" "Bank Transfer(automatic)" "Credit Card (aut
##  $ Tenure              : num  6.8 1.16 15.75 17.09 1.67 ...
##  $ MonthlyCharge       : num  172 243 160 120 150 ...
##  $ Bandwidth_GB_Year   : num  905 801 2055 2165 271 ...
##  $ Item1               : int  5 3 4 4 4 3 6 2 5 2 ...
##  $ Item2               : int  5 4 4 4 4 3 5 2 4 2 ...
##  $ Item3               : int  5 3 2 4 4 3 6 2 4 2 ...
##  $ Item4               : int  3 3 4 2 3 2 4 5 3 2 ...
##  $ Item5               : int  4 4 4 5 4 4 1 2 4 5 ...
##  $ Item6               : int  4 3 3 4 4 3 5 3 3 2 ...
##  $ Item7               : int  3 4 3 3 4 3 5 4 4 3 ...
##  $ Item8               : int  4 4 3 3 5 3 5 5 4 3 ...
```

```r
# Renaming unclear columns named Item1 through Item8 for improved readability &
# confirming they have been renamed correctly

df <- df %>%
  rename(
    Response = Item1,
    Fix = Item2,
    Replacement = Item3,
    Reliability = Item4,
    Options = Item5,
    Respectful = Item6,
    Courteous = Item7,
    Listening = Item8
  )

colnames(df)
```

```
##  [1] "CaseOrder"            "Customer_id"          "Interaction"
##  [4] "UID"                  "City"                 "State"
##  [7] "County"               "Zip"                  "Lat"
## [10] "Lng"                  "Population"           "Area"
## [13] "TimeZone"             "Job"                  "Children"
## [16] "Age"                  "Income"               "Marital"
## [19] "Gender"               "Churn"                "Outage_sec_perweek"
## [22] "Email"                "Contacts"             "Yearly_equip_failure"
## [25] "Techie"               "Contract"             "Port_modem"
## [28] "Tablet"               "InternetService"      "Phone"
## [31] "Multiple"             "OnlineSecurity"       "OnlineBackup"
## [34] "DeviceProtection"     "TechSupport"          "StreamingTV"
## [37] "StreamingMovies"      "PaperlessBilling"     "PaymentMethod"
## [40] "Tenure"               "MonthlyCharge"        "Bandwidth_GB_Year"
## [43] "Response"             "Fix"                  "Replacement"
## [46] "Reliability"          "Options"              "Respectful"
## [49] "Courteous"            "Listening"
```

```r
# Several columns will not be useful in analysis and therefore will be dropped.
to_drop <- c('CaseOrder', 'Customer_id', 'Interaction', 'UID', 'City',
             'County', 'Zip', 'Lat', 'Lng', 'TimeZone', 'Job')

dfDropped = df[,!(names(df) %in% to_drop)]
```

```r
str(dfDropped)
```

```
## 'data.frame':    10000 obs. of  39 variables:
##  $ State               : chr  "AK" "MI" "OR" "CA" ...
##  $ Population          : int  38 10446 3735 13863 11352 17701 2535 23144 17351 20193 ...
##  $ Area                : chr  "Urban" "Urban" "Urban" "Suburban" ...
##  $ Children            : int  0 1 4 1 0 3 0 2 2 1 ...
##  $ Age                 : int  68 27 50 48 83 83 79 30 49 86 ...
##  $ Income              : num  28562 21705 9610 18925 40074 ...
##  $ Marital             : chr  "Widowed" "Married" "Widowed" "Married" ...
##  $ Gender              : chr  "Male" "Female" "Female" "Male" ...
##  $ Churn               : chr  "No" "Yes" "No" "No" ...
##  $ Outage_sec_perweek  : num  7.98 11.7 10.75 14.91 8.15 ...
##  $ Email               : int  10 12 9 15 16 15 10 16 20 18 ...
##  $ Contacts            : int  0 0 0 2 2 3 0 0 2 1 ...
##  $ Yearly_equip_failure: int  1 1 1 0 1 1 1 0 3 0 ...
##  $ Techie              : chr  "No" "Yes" "Yes" "Yes" ...
##  $ Contract            : chr  "One year" "Month-to-month" "Two Year" "Two Year" ...
##  $ Port_modem          : chr  "Yes" "No" "Yes" "No" ...
##  $ Tablet              : chr  "Yes" "Yes" "No" "No" ...
##  $ InternetService     : chr  "Fiber Optic" "Fiber Optic" "DSL" "DSL" ...
##  $ Phone               : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Multiple            : chr  "No" "Yes" "Yes" "No" ...
##  $ OnlineSecurity      : chr  "Yes" "Yes" "No" "Yes" ...
##  $ OnlineBackup        : chr  "Yes" "No" "No" "No" ...
##  $ DeviceProtection    : chr  "No" "No" "No" "No" ...
##  $ TechSupport         : chr  "No" "No" "No" "No" ...
##  $ StreamingTV         : chr  "No" "Yes" "No" "Yes" ...
##  $ StreamingMovies     : chr  "Yes" "Yes" "Yes" "No" ...
##  $ PaperlessBilling    : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ PaymentMethod       : chr  "Credit Card (automatic)" "Bank Transfer(automatic)" "Credit Card (aut
##  $ Tenure              : num  6.8 1.16 15.75 17.09 1.67 ...
##  $ MonthlyCharge       : num  172 243 160 120 150 ...
##  $ Bandwidth_GB_Year   : num  905 801 2055 2165 271 ...
##  $ Response            : int  5 3 4 4 4 3 6 2 5 2 ...
##  $ Fix                 : int  5 4 4 4 4 3 5 2 4 2 ...
##  $ Replacement         : int  5 3 2 4 4 3 6 2 4 2 ...
##  $ Reliability         : int  3 3 4 2 3 2 4 5 3 2 ...
##  $ Options             : int  4 4 4 5 4 4 1 2 4 5 ...
##  $ Respectful          : int  4 3 3 4 4 3 5 3 3 2 ...
##  $ Courteous           : int  3 4 3 3 4 3 5 4 4 3 ...
##  $ Listening           : int  4 4 3 3 5 3 5 5 4 3 ...
```

```r
# Creating dummy variables for categorical columns
dfReg <- dummy_cols(dfDropped, remove_selected_columns = TRUE)
names(dfReg) <- gsub(" ", "_", names(dfReg))
names(dfReg) <- gsub("-", "_", names(dfReg))
names(dfReg) <- gsub("[()]", "_", names(dfReg))
str(dfReg)
```

```
## 'data.frame':    10000 obs. of  118 variables:
##  $ Population                           : int  38 10446 3735 13863 11352 17701 2535 23144 17351 201
##  $ Children                             : int  0 1 4 1 0 3 0 2 2 1 ...
##  $ Age                                  : int  68 27 50 48 83 83 79 30 49 86 ...
```

```
##  $ Income                                : num  28562 21705 9610 18925 40074 ...
##  $ Outage_sec_perweek                    : num  7.98 11.7 10.75 14.91 8.15 ...
##  $ Email                                 : int  10 12 9 15 16 15 10 16 20 18 ...
##  $ Contacts                              : int  0 0 0 2 2 3 0 0 2 1 ...
##  $ Yearly_equip_failure                  : int  1 1 1 0 1 1 1 0 3 0 ...
##  $ Tenure                                : num  6.8 1.16 15.75 17.09 1.67 ...
##  $ MonthlyCharge                         : num  172 243 160 120 150 ...
##  $ Bandwidth_GB_Year                     : num  905 801 2055 2165 271 ...
##  $ Response                              : int  5 3 4 4 4 3 6 2 5 2 ...
##  $ Fix                                   : int  5 4 4 4 4 3 5 2 4 2 ...
##  $ Replacement                           : int  5 3 2 4 4 3 6 2 4 2 ...
##  $ Reliability                           : int  3 3 4 2 3 2 4 5 3 2 ...
##  $ Options                               : int  4 4 4 5 4 4 1 2 4 5 ...
##  $ Respectful                            : int  4 3 3 4 4 3 5 3 3 2 ...
##  $ Courteous                             : int  3 4 3 3 4 3 5 4 4 3 ...
##  $ Listening                             : int  4 4 3 3 5 3 5 5 4 3 ...
##  $ State_AK                              : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ State_AL                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_AR                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_AZ                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_CA                              : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ State_CO                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_CT                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_DC                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_DE                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_FL                              : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ State_GA                              : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ State_HI                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_IA                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_ID                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_IL                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_IN                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_KS                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_KY                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_LA                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MA                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MD                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_ME                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MI                              : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ State_MN                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MO                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MS                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MT                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NC                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_ND                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NE                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NH                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NJ                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NM                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NV                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NY                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_OH                              : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ State_OK                              : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ State_OR                              : int  0 0 1 0 0 0 0 0 0 0 ...
```

```
##  $ State_PA                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_PR                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_RI                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_SC                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_SD                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_TN                              : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ State_TX                              : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ State_UT                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_VA                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_VT                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_WA                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_WI                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_WV                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_WY                              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Area_Rural                            : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Area_Suburban                         : int  0 0 0 1 1 0 1 1 1 0 ...
##  $ Area_Urban                            : int  1 1 1 0 0 1 0 0 0 0 ...
##  $ Marital_Divorced                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Marital_Married                       : int  0 1 0 1 0 0 0 1 0 1 ...
##  $ Marital_Never_Married                 : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Marital_Separated                     : int  0 0 0 0 1 0 0 0 1 0 ...
##  $ Marital_Widowed                       : int  1 0 1 0 0 0 1 0 0 0 ...
##  $ Gender_Female                         : int  0 1 1 0 0 1 0 1 0 1 ...
##  $ Gender_Male                           : int  1 0 0 1 1 0 1 0 0 0 ...
##  $ Gender_Nonbinary                      : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Churn_No                              : int  1 0 1 1 0 1 0 0 1 1 ...
##  $ Churn_Yes                             : int  0 1 0 0 1 0 1 1 0 0 ...
##  $ Techie_No                             : int  1 0 0 0 1 1 0 0 1 1 ...
##  $ Techie_Yes                            : int  0 1 1 1 0 0 1 1 0 0 ...
##  $ Contract_Month_to_month               : int  0 1 0 0 1 0 1 1 1 0 ...
##  $ Contract_One_year                     : int  1 0 0 0 0 1 0 0 0 0 ...
##  $ Contract_Two_Year                     : int  0 0 1 1 0 0 0 0 0 1 ...
##  $ Port_modem_No                         : int  0 1 0 1 0 0 1 1 0 0 ...
##  $ Port_modem_Yes                        : int  1 0 1 0 1 1 0 0 1 1 ...
##  $ Tablet_No                             : int  0 0 1 1 1 1 1 1 1 1 ...
##  $ Tablet_Yes                            : int  1 1 0 0 0 0 0 0 0 0 ...
##  $ InternetService_DSL                   : int  0 0 1 1 0 0 1 1 1 0 ...
##  $ InternetService_Fiber_Optic           : int  1 1 0 0 1 0 0 0 0 1 ...
##  $ InternetService_None                  : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Phone_No                              : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ Phone_Yes                             : int  1 1 1 1 0 1 1 0 1 1 ...
##  $ Multiple_No                           : int  1 0 0 1 1 0 1 1 1 1 ...
##    [list output truncated]
```

```r
dim(dfReg)
```

```
## [1] 10000   118
```

```r
# Split dfReg into training and testing subsets
set.seed(22)
trainId = createDataPartition(dfReg$Churn_Yes, times = 1, p = 0.7, list = FALSE)

dfTrain = dfReg[trainId,]
dfTest = dfReg[-trainId,]
```

```r
# Summary Statistics
summary(dfTrain)
```

```
##    Population          Children          Age             Income
##  Min.   :     0.0   Min.   : 0.00   Min.   :18.00   Min.   :   643.2
##  1st Qu.:   733.8   1st Qu.: 0.00   1st Qu.:35.00   1st Qu.: 19349.8
##  Median :  2867.0   Median : 1.00   Median :53.00   Median : 33304.9
##  Mean   :  9683.6   Mean   : 2.06   Mean   :52.89   Mean   : 40080.0
##  3rd Qu.: 12933.0   3rd Qu.: 3.00   3rd Qu.:71.00   3rd Qu.: 53249.4
##  Max.   :111850.0   Max.   :10.00   Max.   :89.00   Max.   :256998.4
##  Outage_sec_perweek     Email          Contacts     Yearly_equip_failure
##  Min.   : 0.09975   Min.   : 1.00   Min.   :0.000   Min.   :0.0000
##  1st Qu.: 8.02346   1st Qu.:10.00   1st Qu.:0.000   1st Qu.:0.0000
##  Median : 9.98095   Median :12.00   Median :1.000   Median :0.0000
##  Mean   : 9.97626   Mean   :12.02   Mean   :1.004   Mean   :0.3939
##  3rd Qu.:11.93163   3rd Qu.:14.00   3rd Qu.:2.000   3rd Qu.:1.0000
##  Max.   :21.20723   Max.   :23.00   Max.   :7.000   Max.   :6.0000
##      Tenure       MonthlyCharge    Bandwidth_GB_Year    Response
##  Min.   : 1.000   Min.   : 79.98   Min.   : 155.5    Min.   :1.00
##  1st Qu.: 7.851   1st Qu.:139.98   1st Qu.:1233.9    1st Qu.:3.00
##  Median :28.095   Median :169.94   Median :3018.9    Median :3.00
##  Mean   :34.282   Mean   :172.84   Mean   :3377.1    Mean   :3.49
##  3rd Qu.:61.354   3rd Qu.:202.44   3rd Qu.:5570.3    3rd Qu.:4.00
##  Max.   :71.994   Max.   :290.16   Max.   :7159.0    Max.   :7.00
##       Fix        Replacement     Reliability       Options        Respectful
##  Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:3.00   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000
##  Median :4.00   Median :3.000   Median :4.000   Median :3.000   Median :4.000
##  Mean   :3.51   Mean   :3.491   Mean   :3.507   Mean   :3.485   Mean   :3.516
##  3rd Qu.:4.00   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :7.00   Max.   :8.000   Max.   :7.000   Max.   :7.000   Max.   :8.000
##    Courteous       Listening        State_AK           State_AL
##  Min.   :1.000   Min.   :1.000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:3.000   1st Qu.:3.000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :4.000   Median :4.000   Median :0.000000   Median :0.00000
##  Mean   :3.517   Mean   :3.508   Mean   :0.008143   Mean   :0.01857
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :7.000   Max.   :8.000   Max.   :1.000000   Max.   :1.00000
##     State_AR          State_AZ          State_CA          State_CO
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01843   Mean   :0.01214   Mean   :0.05486   Mean   :0.01586
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##     State_CT           State_DC           State_DE           State_FL
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000
##  Median :0.000000   Median :0.000000   Median :0.000000   Median :0.000
##  Mean   :0.007286   Mean   :0.001429   Mean   :0.002143   Mean   :0.033
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000
##  Max.   :1.000000   Max.   :1.000000   Max.   :1.000000   Max.   :1.000
##     State_GA          State_HI        State_IA          State_ID
##  Min.   :0.00000   Min.   :0.000   Min.   :0.00000   Min.   :0.000000
```

```
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.00000   Median :0.000   Median :0.00000   Median :0.000000
##  Mean   :0.02314   Mean   :0.003   Mean   :0.02714   Mean   :0.008714
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.000000
##  Max.   :1.00000   Max.   :1.000   Max.   :1.00000   Max.   :1.000000
##     State_IL          State_IN          State_KS          State_KY
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.04143   Mean   :0.02414   Mean   :0.01929   Mean   :0.02357
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##     State_LA          State_MA          State_MD        State_ME
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.000   Median :0.00000
##  Mean   :0.01386   Mean   :0.01586   Mean   :0.012   Mean   :0.01171
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000   Max.   :1.00000
##     State_MI        State_MN        State_MO          State_MS
##  Min.   :0.000   Min.   :0.000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.000   Median :0.000   Median :0.00000   Median :0.00000
##  Mean   :0.028   Mean   :0.027   Mean   :0.02929   Mean   :0.01229
##  3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.000   Max.   :1.000   Max.   :1.00000   Max.   :1.00000
##     State_MT           State_NC          State_ND        State_NE
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000
##  Median :0.000000   Median :0.00000   Median :0.000   Median :0.00000
##  Mean   :0.009429   Mean   :0.02843   Mean   :0.012   Mean   :0.01986
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000   Max.   :1.000   Max.   :1.00000
##     State_NH           State_NJ       State_NM          State_NV
##  Min.   :0.000000   Min.   :0.00   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.00   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.000000   Median :0.00   Median :0.00000   Median :0.000000
##  Mean   :0.008857   Mean   :0.02   Mean   :0.01157   Mean   :0.005429
##  3rd Qu.:0.000000   3rd Qu.:0.00   3rd Qu.:0.00000   3rd Qu.:0.000000
##  Max.   :1.000000   Max.   :1.00   Max.   :1.00000   Max.   :1.000000
##     State_NY          State_OH          State_OK        State_OR
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000   Min.   :0.00
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00
##  Median :0.00000   Median :0.00000   Median :0.000   Median :0.00
##  Mean   :0.05471   Mean   :0.03514   Mean   :0.021   Mean   :0.01
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000   Max.   :1.00
##     State_PA          State_PR        State_RI        State_SC
##  Min.   :0.00000   Min.   :0.000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000   Median :0.000   Median :0.00000
##  Mean   :0.05471   Mean   :0.003   Mean   :0.002   Mean   :0.01314
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000   Max.   :1.000   Max.   :1.00000
```

```
##     State_SD           State_TN          State_TX          State_UT
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.000000   Median :0.00000   Median :0.00000   Median :0.000000
##  Mean   :0.008571   Mean   :0.01943   Mean   :0.05971   Mean   :0.006714
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000
##  Max.   :1.000000   Max.   :1.00000   Max.   :1.00000   Max.   :1.000000
##     State_VA          State_VT        State_WA          State_WI
##  Min.   :0.00000   Min.   :0.000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000   Median :0.00000   Median :0.00000
##  Mean   :0.02857   Mean   :0.007   Mean   :0.01714   Mean   :0.02329
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000   Max.   :1.00000   Max.   :1.00000
##     State_WV        State_WY       Area_Rural     Area_Suburban
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.000   Median :0.000   Median :0.0000   Median :0.0000
##  Mean   :0.024   Mean   :0.004   Mean   :0.3391   Mean   :0.3324
##  3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##    Area_Urban     Marital_Divorced Marital_Married  Marital_Never_Married
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.3284   Mean   :0.2136   Mean   :0.1889   Mean   :0.1943
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  Marital_Separated Marital_Widowed  Gender_Female     Gender_Male
##  Min.   :0.0000    Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000    1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000    Median :0.0000   Median :0.0000   Median :0.000
##  Mean   :0.2024    Mean   :0.2009   Mean   :0.4981   Mean   :0.479
##  3rd Qu.:0.0000    3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000    Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##  Gender_Nonbinary     Churn_No        Churn_Yes        Techie_No
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000
##  Median :0.00000   Median :1.0000   Median :0.0000   Median :1.000
##  Mean   :0.02286   Mean   :0.7267   Mean   :0.2733   Mean   :0.833
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##    Techie_Yes    Contract_Month_to_month Contract_One_year Contract_Two_Year
##  Min.   :0.000   Min.   :0.000           Min.   :0.0000    Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000           1st Qu.:0.0000    1st Qu.:0.0000
##  Median :0.000   Median :1.000           Median :0.0000    Median :0.0000
##  Mean   :0.167   Mean   :0.548           Mean   :0.2051    Mean   :0.2469
##  3rd Qu.:0.000   3rd Qu.:1.000           3rd Qu.:0.0000    3rd Qu.:0.0000
##  Max.   :1.000   Max.   :1.000           Max.   :1.0000    Max.   :1.0000
##  Port_modem_No    Port_modem_Yes     Tablet_No        Tablet_Yes
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000   Median :1.0000   Median :0.0000
##  Mean   :0.5173   Mean   :0.4827   Mean   :0.7063   Mean   :0.2937
```

```
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  InternetService_DSL InternetService_Fiber_Optic InternetService_None
##  Min.   :0.0000      Min.   :0.0000              Min.   :0.0000
##  1st Qu.:0.0000      1st Qu.:0.0000              1st Qu.:0.0000
##  Median :0.0000      Median :0.0000              Median :0.0000
##  Mean   :0.3544      Mean   :0.4344              Mean   :0.2111
##  3rd Qu.:1.0000      3rd Qu.:1.0000              3rd Qu.:0.0000
##  Max.   :1.0000      Max.   :1.0000              Max.   :1.0000
##     Phone_No         Phone_Yes       Multiple_No      Multiple_Yes
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :1.0000   Median :1.0000   Median :0.0000
##  Mean   :0.09586   Mean   :0.9041   Mean   :0.5407   Mean   :0.4593
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  OnlineSecurity_No OnlineSecurity_Yes OnlineBackup_No  OnlineBackup_Yes
##  Min.   :0.0000    Min.   :0.0000     Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000    1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000    Median :0.0000     Median :1.0000   Median :0.0000
##  Mean   :0.6446    Mean   :0.3554     Mean   :0.5426   Mean   :0.4574
##  3rd Qu.:1.0000    3rd Qu.:1.0000     3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000    Max.   :1.0000     Max.   :1.0000   Max.   :1.0000
##  DeviceProtection_No DeviceProtection_Yes TechSupport_No   TechSupport_Yes
##  Min.   :0.0000      Min.   :0.0000       Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000      1st Qu.:0.0000       1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000      Median :0.0000       Median :1.0000   Median :0.0000
##  Mean   :0.5626      Mean   :0.4374       Mean   :0.6237   Mean   :0.3763
##  3rd Qu.:1.0000      3rd Qu.:1.0000       3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000      Max.   :1.0000       Max.   :1.0000   Max.   :1.0000
##  StreamingTV_No  StreamingTV_Yes StreamingMovies_No StreamingMovies_Yes
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000     1st Qu.:0.0000
##  Median :1.000   Median :0.000   Median :1.0000     Median :0.0000
##  Mean   :0.503   Mean   :0.497   Mean   :0.5107     Mean   :0.4893
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.0000     3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000     Max.   :1.0000
##  PaperlessBilling_No PaperlessBilling_Yes
##  Min.   :0.000       Min.   :0.000
##  1st Qu.:0.000       1st Qu.:0.000
##  Median :0.000       Median :1.000
##  Mean   :0.414       Mean   :0.586
##  3rd Qu.:1.000       3rd Qu.:1.000
##  Max.   :1.000       Max.   :1.000
##  PaymentMethod_Bank_Transfer_automatic_ PaymentMethod_Credit_Card__automatic_
##  Min.   :0.000                          Min.   :0.000
##  1st Qu.:0.000                          1st Qu.:0.000
##  Median :0.000                          Median :0.000
##  Mean   :0.225                          Mean   :0.202
##  3rd Qu.:0.000                          3rd Qu.:0.000
##  Max.   :1.000                          Max.   :1.000
##  PaymentMethod_Electronic_Check PaymentMethod_Mailed_Check
##  Min.   :0.0000                 Min.   :0.0000
##  1st Qu.:0.0000                 1st Qu.:0.0000
```

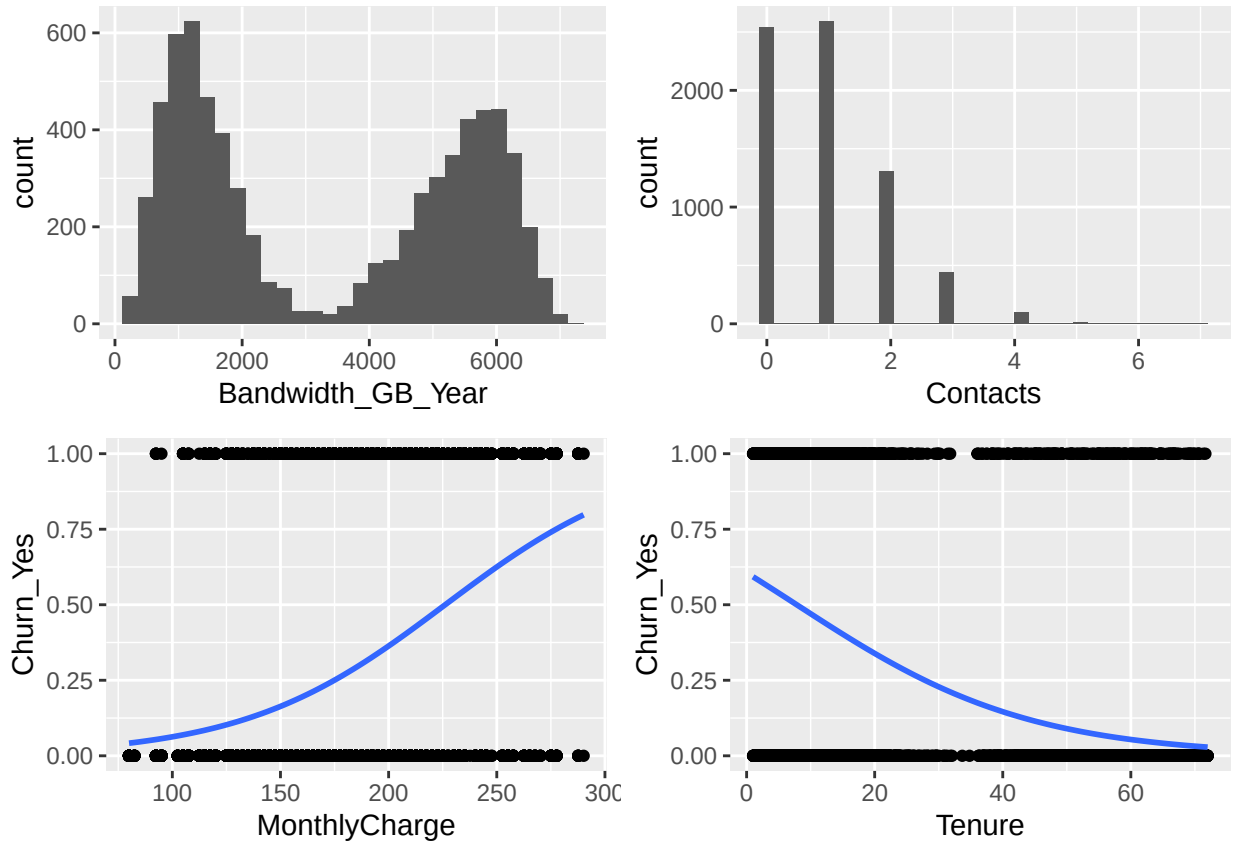
```
##  Median :0.0000                 Median :0.0000
##  Mean   :0.3399                 Mean   :0.2331
##  3rd Qu.:1.0000                 3rd Qu.:0.0000
##  Max.   :1.0000                 Max.   :1.0000
```

```r
# Data Visualizations
plot_grid(
  ggplot(dfTrain, aes(x=Bandwidth_GB_Year)) +
    geom_histogram(),

  ggplot(dfTrain, aes(x=Contacts)) +
    geom_histogram(),

  ggplot(dfTrain, aes(x=MonthlyCharge, y=Churn_Yes)) +
    geom_point() +
    stat_smooth(method = "glm", se=FALSE, method.args = list(family=binomial)),

  ggplot(dfTrain, aes(x=Tenure, y=Churn_Yes)) +
    geom_point() +
    stat_smooth(method = "glm", se=FALSE, method.args = list(family=binomial)),


  ncol = 2, nrow = 2
)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```




```
# Export prepared data sets
write.csv(dfTrain, "C:\\WGU\\D208 Predictive Modeling\\PA Task 2\\D208_dfTrain.csv", row.names = FALSE)
write.csv(dfTest, "C:\\WGU\\D208 Predictive Modeling\\PA Task 2\\D208_dfTest.csv", row.names = FALSE)

# Initial Model [Note the "Yes" and "No" columns are redundant, so only one will
# be used]
trainColsDf = dfTrain[, -which(names(dfTrain) %in% c("Churn_Yes", "Churn_No"))]
trainColNames <- colnames(trainColsDf)

predictorVars <- paste(trainColNames, collapse = " + ")

logRegForm <- paste("Churn_Yes ~ ", predictorVars, sep = "")

initialModel <- glm(formula = logRegForm, family = "binomial", data = dfTrain)
initialModel
```

```
##
## Call:  glm(formula = logRegForm, family = "binomial", data = dfTrain)
##
## Coefficients:
##                          (Intercept)                           Population
##                           -5.952e+00                            2.459e-07
##                             Children                                  Age
##                            9.252e-02                           -8.679e-03
##                               Income                   Outage_sec_perweek
##                            1.088e-06                           -1.179e-02
##                                Email                             Contacts
```

```
##                          -1.350e-02                          5.306e-02
##                 Yearly_equip_failure                             Tenure
##                          2.231e-02                          1.345e-01
##                       MonthlyCharge                   Bandwidth_GB_Year
##                          5.547e-02                         -3.091e-03
##                           Response                                Fix
##                          2.042e-02                         -4.005e-02
##                        Replacement                        Reliability
##                          1.282e-03                         -1.924e-02
##                            Options                        Respectful
##                         -1.592e-02                          5.526e-03
##                          Courteous                          Listening
##                         -2.130e-02                          8.620e-03
##                           State_AK                           State_AL
##                         -9.120e-01                         -1.033e+00
##                           State_AR                           State_AZ
##                         -1.150e+00                         -8.197e-01
##                           State_CA                           State_CO
##                         -6.196e-01                         -8.288e-01
##                           State_CT                           State_DC
##                         -7.755e-01                         -1.446e-01
##                           State_DE                           State_FL
##                         -1.066e+00                         -1.186e+00
##                           State_GA                           State_HI
##                         -3.891e-01                         -2.822e-01
##                           State_IA                           State_ID
##                         -9.981e-01                         -6.773e-01
##                           State_IL                           State_IN
##                         -6.190e-01                         -7.593e-01
##                           State_KS                           State_KY
##                         -5.372e-01                         -6.241e-01
##                           State_LA                           State_MA
##                         -9.076e-01                         -9.070e-01
##                           State_MD                           State_ME
##                         -3.658e-01                         -8.765e-01
##                           State_MI                           State_MN
##                         -6.360e-01                         -4.209e-01
##                           State_MO                           State_MS
##                         -5.389e-01                         -6.629e-01
##                           State_MT                           State_NC
##                         -2.343e-01                         -9.549e-01
##                           State_ND                           State_NE
##                         -9.166e-01                         -4.609e-01
##                           State_NH                           State_NJ
##                         -1.165e+00                         -7.358e-01
##                           State_NM                           State_NV
##                         -1.577e+00                         -1.045e+00
##                           State_NY                           State_OH
##                         -9.496e-01                         -7.178e-01
##                           State_OK                           State_OR
##                         -8.405e-01                         -4.001e-01
##                           State_PA                           State_PR
##                         -1.004e+00                         -1.070e+00
##                           State_RI                           State_SC
```

```
##                          -6.784e+00                          -2.242e-01
##                            State_SD                            State_TN
##                          -1.792e+00                          -1.199e-01
##                            State_TX                            State_UT
##                          -4.142e-01                          -2.125e-01
##                            State_VA                            State_VT
##                          -7.307e-01                          -1.091e+00
##                            State_WA                            State_WI
##                          -4.121e-01                          -5.068e-01
##                            State_WV                            State_WY
##                          -1.205e-01                                  NA
##                          Area_Rural                       Area_Suburban
##                           6.474e-02                           3.227e-02
##                          Area_Urban                    Marital_Divorced
##                                  NA                          -2.659e-01
##                     Marital_Married                 Marital_Never_Married
##                          -2.197e-01                          -2.952e-01
##                   Marital_Separated                      Marital_Widowed
##                          -1.668e-01                                  NA
##                       Gender_Female                          Gender_Male
##                           4.144e-01                           7.982e-01
##                    Gender_Nonbinary                           Techie_No
##                                  NA                          -1.089e+00
##                          Techie_Yes              Contract_Month_to_month
##                                  NA                           3.659e+00
##                   Contract_One_year                    Contract_Two_Year
##                           1.877e-01                                  NA
##                       Port_modem_No                       Port_modem_Yes
##                          -2.123e-01                                  NA
##                           Tablet_No                          Tablet_Yes
##                           4.148e-02                                  NA
##                   InternetService_DSL          InternetService_Fiber_Optic
##                           2.045e+00                          -1.733e+00
##                 InternetService_None                            Phone_No
##                                  NA                           3.818e-01
##                           Phone_Yes                         Multiple_No
##                                  NA                          -1.075e-01
##                         Multiple_Yes                    OnlineSecurity_No
##                                  NA                          -9.717e-03
##                  OnlineSecurity_Yes                      OnlineBackup_No
##                                  NA                           1.232e-01
##                     OnlineBackup_Yes                  DeviceProtection_No
##                                  NA                           2.520e-02
##                 DeviceProtection_Yes                      TechSupport_No
##                                  NA                           3.535e-01
##                      TechSupport_Yes                      StreamingTV_No
##                                  NA                          -1.327e+00
##                      StreamingTV_Yes                  StreamingMovies_No
##                                  NA                          -1.320e+00
##                  StreamingMovies_Yes                  PaperlessBilling_No
##                                  NA                          -2.609e-01
##                 PaperlessBilling_Yes  PaymentMethod_Bank_Transfer_automatic_
##                                  NA                          -1.106e-01
##  PaymentMethod_Credit_Card__automatic_        PaymentMethod_Electronic_Check
```

```
##                              2.480e-02                              4.497e-01
##             PaymentMethod_Mailed_Check
##                                     NA
##
## Degrees of Freedom: 6999 Total (i.e. Null);  6902 Residual
## Null Deviance:       8211
## Residual Deviance: 3025  AIC: 3221
```

```
summary(initialModel)
```

```
##
## Call:
## glm(formula = logRegForm, family = "binomial", data = dfTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6732  -0.2542  -0.0475   0.0834   3.5123
##
## Coefficients: (19 not defined because of singularities)
##                                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)                           -5.952e+00  1.906e+00  -3.123 0.001792
## Population                             2.459e-07  3.558e-06   0.069 0.944910
## Children                               9.252e-02  1.660e-01   0.557 0.577351
## Age                                   -8.679e-03  1.772e-02  -0.490 0.624270
## Income                                 1.088e-06  1.660e-06   0.656 0.512068
## Outage_sec_perweek                    -1.179e-02  1.572e-02  -0.750 0.453300
## Email                                 -1.350e-02  1.533e-02  -0.881 0.378403
## Contacts                               5.306e-02  4.733e-02   1.121 0.262236
## Yearly_equip_failure                   2.231e-02  7.513e-02   0.297 0.766474
## Tenure                                 1.345e-01  4.388e-01   0.306 0.759306
## MonthlyCharge                          5.547e-02  1.673e-02   3.315 0.000917
## Bandwidth_GB_Year                     -3.091e-03  5.356e-03  -0.577 0.563919
## Response                               2.042e-02  6.600e-02   0.309 0.757023
## Fix                                   -4.005e-02  6.257e-02  -0.640 0.522163
## Replacement                            1.282e-03  5.657e-02   0.023 0.981914
## Reliability                           -1.924e-02  5.091e-02  -0.378 0.705456
## Options                               -1.592e-02  5.380e-02  -0.296 0.767289
## Respectful                             5.526e-03  5.476e-02   0.101 0.919623
## Courteous                             -2.130e-02  5.269e-02  -0.404 0.686085
## Listening                              8.620e-03  4.891e-02   0.176 0.860093
## State_AK                              -9.120e-01  8.729e-01  -1.045 0.296101
## State_AL                              -1.033e+00  8.351e-01  -1.237 0.216273
## State_AR                              -1.150e+00  8.219e-01  -1.400 0.161551
## State_AZ                              -8.197e-01  8.845e-01  -0.927 0.354041
## State_CA                              -6.196e-01  7.736e-01  -0.801 0.423177
## State_CO                              -8.288e-01  8.383e-01  -0.989 0.322817
## State_CT                              -7.755e-01  9.014e-01  -0.860 0.389593
## State_DC                              -1.446e-01  1.274e+00  -0.114 0.909623
## State_DE                              -1.066e+00  1.313e+00  -0.812 0.417072
## State_FL                              -1.186e+00  7.941e-01  -1.494 0.135252
## State_GA                              -3.891e-01  8.038e-01  -0.484 0.628357
## State_HI                              -2.822e-01  1.052e+00  -0.268 0.788482
## State_IA                              -9.981e-01  7.992e-01  -1.249 0.211748
## State_ID                              -6.773e-01  8.944e-01  -0.757 0.448874
## State_IL                              -6.190e-01  7.796e-01  -0.794 0.427245
```

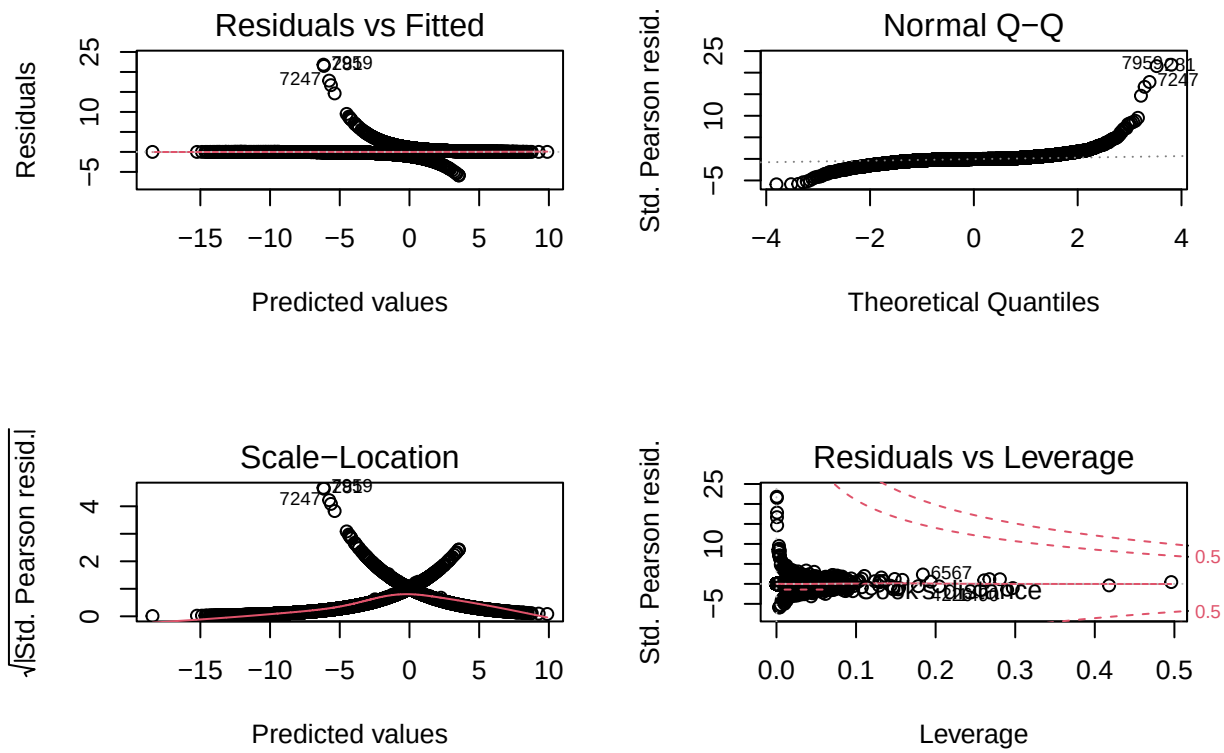
```
## State_IN                          -7.593e-01  8.049e-01  -0.943 0.345493
## State_KS                          -5.372e-01  8.344e-01  -0.644 0.519735
## State_KY                          -6.241e-01  8.039e-01  -0.776 0.437525
## State_LA                          -9.076e-01  8.287e-01  -1.095 0.273455
## State_MA                          -9.070e-01  8.308e-01  -1.092 0.274924
## State_MD                          -3.658e-01  8.463e-01  -0.432 0.665543
## State_ME                          -8.765e-01  8.533e-01  -1.027 0.304338
## State_MI                          -6.360e-01  7.941e-01  -0.801 0.423160
## State_MN                          -4.209e-01  7.918e-01  -0.532 0.595018
## State_MO                          -5.389e-01  7.864e-01  -0.685 0.493148
## State_MS                          -6.629e-01  8.294e-01  -0.799 0.424145
## State_MT                          -2.343e-01  8.754e-01  -0.268 0.789019
## State_NC                          -9.549e-01  7.922e-01  -1.205 0.228064
## State_ND                          -9.166e-01  8.274e-01  -1.108 0.267926
## State_NE                          -4.609e-01  8.135e-01  -0.567 0.570978
## State_NH                          -1.165e+00  8.992e-01  -1.296 0.195003
## State_NJ                          -7.358e-01  8.146e-01  -0.903 0.366343
## State_NM                          -1.577e+00  8.418e-01  -1.874 0.060958
## State_NV                          -1.045e+00  1.058e+00  -0.988 0.323312
## State_NY                          -9.496e-01  7.684e-01  -1.236 0.216551
## State_OH                          -7.178e-01  7.862e-01  -0.913 0.361288
## State_OK                          -8.405e-01  8.058e-01  -1.043 0.296909
## State_OR                          -4.001e-01  8.638e-01  -0.463 0.643242
## State_PA                          -1.004e+00  7.704e-01  -1.303 0.192665
## State_PR                          -1.070e+00  1.258e+00  -0.851 0.394958
## State_RI                          -6.784e+00  2.527e+00  -2.685 0.007257
## State_SC                          -2.242e-01  8.627e-01  -0.260 0.794968
## State_SD                          -1.792e+00  9.440e-01  -1.898 0.057708
## State_TN                          -1.199e-01  8.272e-01  -0.145 0.884738
## State_TX                          -4.142e-01  7.672e-01  -0.540 0.589284
## State_UT                          -2.125e-01  9.693e-01  -0.219 0.826446
## State_VA                          -7.307e-01  7.982e-01  -0.915 0.360006
## State_VT                          -1.091e+00  9.796e-01  -1.114 0.265188
## State_WA                          -4.121e-01  8.107e-01  -0.508 0.611179
## State_WI                          -5.068e-01  8.012e-01  -0.632 0.527082
## State_WV                          -1.205e-01  7.998e-01  -0.151 0.880216
## State_WY                                  NA         NA      NA       NA
## Area_Rural                         6.474e-02  1.149e-01   0.564 0.573066
## Area_Suburban                      3.227e-02  1.140e-01   0.283 0.777040
## Area_Urban                                NA         NA      NA       NA
## Marital_Divorced                  -2.659e-01  1.453e-01  -1.830 0.067299
## Marital_Married                   -2.197e-01  1.491e-01  -1.474 0.140610
## Marital_Never_Married             -2.952e-01  1.513e-01  -1.951 0.051067
## Marital_Separated                 -1.668e-01  1.474e-01  -1.132 0.257689
## Marital_Widowed                           NA         NA      NA       NA
## Gender_Female                      4.144e-01  3.577e-01   1.159 0.246643
## Gender_Male                        7.982e-01  5.730e-01   1.393 0.163611
## Gender_Nonbinary                          NA         NA      NA       NA
## Techie_No                         -1.089e+00  1.261e-01  -8.642  < 2e-16
## Techie_Yes                                NA         NA      NA       NA
## Contract_Month_to_month            3.659e+00  1.553e-01  23.567  < 2e-16
## Contract_One_year                  1.877e-01  1.676e-01   1.120 0.262654
## Contract_Two_Year                         NA         NA      NA       NA
## Port_modem_No                     -2.123e-01  9.371e-02  -2.265 0.023486
```

```
## Port_modem_Yes                              NA         NA      NA       NA
## Tablet_No                             4.148e-02  1.037e-01   0.400 0.689201
## Tablet_Yes                                  NA         NA      NA       NA
## InternetService_DSL                   2.045e+00  2.021e+00   1.012 0.311602
## InternetService_Fiber_Optic          -1.733e+00  5.629e-01  -3.079 0.002075
## InternetService_None                        NA         NA      NA       NA
## Phone_No                              3.818e-01  1.583e-01   2.412 0.015878
## Phone_Yes                                   NA         NA      NA       NA
## Multiple_No                          -1.075e-01  2.421e-01  -0.444 0.656970
## Multiple_Yes                                NA         NA      NA       NA
## OnlineSecurity_No                    -9.717e-03  3.764e-01  -0.026 0.979404
## OnlineSecurity_Yes                          NA         NA      NA       NA
## OnlineBackup_No                       1.232e-01  2.194e-01   0.562 0.574365
## OnlineBackup_Yes                            NA         NA      NA       NA
## DeviceProtection_No                   2.520e-02  2.827e-01   0.089 0.928963
## DeviceProtection_Yes                        NA         NA      NA       NA
## TechSupport_No                        3.535e-01  2.080e-01   1.700 0.089134
## TechSupport_Yes                             NA         NA      NA       NA
## StreamingTV_No                       -1.327e+00  6.184e-01  -2.146 0.031913
## StreamingTV_Yes                             NA         NA      NA       NA
## StreamingMovies_No                   -1.320e+00  4.387e-01  -3.010 0.002617
## StreamingMovies_Yes                         NA         NA      NA       NA
## PaperlessBilling_No                  -2.609e-01  9.536e-02  -2.736 0.006216
## PaperlessBilling_Yes                        NA         NA      NA       NA
## PaymentMethod_Bank_Transfer_automatic_ -1.106e-01  1.391e-01  -0.795 0.426500
## PaymentMethod_Credit_Card__automatic_  2.480e-02  1.419e-01   0.175 0.861229
## PaymentMethod_Electronic_Check        4.497e-01  1.277e-01   3.521 0.000429
## PaymentMethod_Mailed_Check                  NA         NA      NA       NA
##
## (Intercept)                          **
## Population
## Children
## Age
## Income
## Outage_sec_perweek
## Email
## Contacts
## Yearly_equip_failure
## Tenure
## MonthlyCharge                        ***
## Bandwidth_GB_Year
## Response
## Fix
## Replacement
## Reliability
## Options
## Respectful
## Courteous
## Listening
## State_AK
## State_AL
## State_AR
## State_AZ
## State_CA
```

```
## State_CO
## State_CT
## State_DC
## State_DE
## State_FL
## State_GA
## State_HI
## State_IA
## State_ID
## State_IL
## State_IN
## State_KS
## State_KY
## State_LA
## State_MA
## State_MD
## State_ME
## State_MI
## State_MN
## State_MO
## State_MS
## State_MT
## State_NC
## State_ND
## State_NE
## State_NH
## State_NJ
## State_NM                                .
## State_NV
## State_NY
## State_OH
## State_OK
## State_OR
## State_PA
## State_PR
## State_RI                                **
## State_SC
## State_SD                                .
## State_TN
## State_TX
## State_UT
## State_VA
## State_VT
## State_WA
## State_WI
## State_WV
## State_WY
## Area_Rural
## Area_Suburban
## Area_Urban
## Marital_Divorced                        .
## Marital_Married
## Marital_Never_Married                   .
## Marital_Separated
```

```
## Marital_Widowed
## Gender_Female
## Gender_Male
## Gender_Nonbinary
## Techie_No                            ***
## Techie_Yes
## Contract_Month_to_month              ***
## Contract_One_year
## Contract_Two_Year
## Port_modem_No                        *
## Port_modem_Yes
## Tablet_No
## Tablet_Yes
## InternetService_DSL
## InternetService_Fiber_Optic          **
## InternetService_None
## Phone_No                             *
## Phone_Yes
## Multiple_No
## Multiple_Yes
## OnlineSecurity_No
## OnlineSecurity_Yes
## OnlineBackup_No
## OnlineBackup_Yes
## DeviceProtection_No
## DeviceProtection_Yes
## TechSupport_No                       .
## TechSupport_Yes
## StreamingTV_No                       *
## StreamingTV_Yes
## StreamingMovies_No                   **
## StreamingMovies_Yes
## PaperlessBilling_No                  **
## PaperlessBilling_Yes
## PaymentMethod_Bank_Transfer_automatic_
## PaymentMethod_Credit_Card__automatic_
## PaymentMethod_Electronic_Check       ***
## PaymentMethod_Mailed_Check
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8211.0  on 6999  degrees of freedom
## Residual deviance: 3025.5  on 6902  degrees of freedom
## AIC: 3221.5
##
## Number of Fisher Scoring iterations: 7
```

```r
par(mfrow = c(2,2))
plot(initialModel)
```




```r
# Solution: remove redundant columns & re-run the model.

initialModelDf <- tidy(initialModel)

# Several columns are labeled NA because of high correlation to another.
# They will be removed.
initialModelNACols <- initialModelDf$term[is.na(initialModelDf$p.value)]

# Recreate training dataframe & linear model without these column(s)
dfTrain2 <- dfTrain[, -which(names(dfTrain) %in% initialModelNACols)]

colNames2 <- colnames(dfTrain2[, -which(names(dfTrain2) %in% c("Churn_Yes", "Churn_No"))])

predictorVars2 <- paste(colNames2, collapse = " + ")

logRegForm2 <- paste("Churn_Yes ~ ", predictorVars2, sep = "")

initialModel2 <- glm(formula = logRegForm2, family = "binomial", data = dfTrain2)

vif(initialModel2) # This will run
```

```
##                          Population                            Children
##                            1.158813                           59.828465
##                                 Age                              Income
##                           60.877023                            1.038336
##                   Outage_sec_perweek                               Email
##                            1.029485                            1.036636
```

```
##                          Contacts                Yearly_equip_failure
##                          1.041108                            1.030770
##                            Tenure                      MonthlyCharge
##                      38310.144136                         223.798706
##                 Bandwidth_GB_Year                           Response
##                      44716.482087                           2.200078
##                               Fix                        Replacement
##                          1.963480                           1.603051
##                       Reliability                            Options
##                          1.293542                           1.419844
##                        Respectful                          Courteous
##                          1.507973                           1.332527
##                         Listening                           State_AK
##                          1.237404                           3.644054
##                          State_AL                           State_AR
##                          4.777113                           5.624182
##                          State_AZ                           State_CA
##                          3.503953                          13.562061
##                          State_CO                           State_CT
##                          4.754154                           3.167468
##                          State_DC                           State_DE
##                          1.535111                           1.492084
##                          State_FL                           State_GA
##                          8.801707                           7.072943
##                          State_HI                           State_IA
##                          2.025329                           7.473330
##                          State_ID                           State_IL
##                          3.255021                          10.916491
##                          State_IN                           State_KS
##                          6.768478                           4.913422
##                          State_KY                           State_LA
##                          6.865205                           5.104869
##                          State_MA                           State_MD
##                          5.036339                           4.338508
##                          State_ME                           State_MI
##                          4.123982                           8.135301
##                          State_MN                           State_MO
##                          8.112397                           9.165362
##                          State_MS                           State_MT
##                          5.192715                           3.643350
##                          State_NC                           State_ND
##                          8.423260                           5.293412
##                          State_NE                           State_NH
##                          5.966678                           3.151556
##                          State_NJ                           State_NM
##                          5.945482                           4.579563
##                          State_NV                           State_NY
##                          2.010766                          15.283579
##                          State_OH                           State_OK
##                          9.403462                           6.723220
##                          State_OR                           State_PA
##                          3.937764                          14.366615
##                          State_PR                           State_RI
##                          1.573397                           1.112185
```

```
##                                 State_SC                                 State_SD
##                                 3.866955                                 2.686724
##                                 State_TN                                 State_TX
##                                 5.217051                                16.447934
##                                 State_UT                                 State_VA
##                                 2.458230                                 7.519490
##                                 State_VT                                 State_WA
##                                 2.387527                                 6.308756
##                                 State_WI                                 State_WV
##                                 7.140430                                 7.268662
##                               Area_Rural                            Area_Suburban
##                                 1.356567                                 1.339358
##                         Marital_Divorced                          Marital_Married
##                                 1.667534                                 1.641546
##                   Marital_Never_Married                        Marital_Separated
##                                 1.613575                                 1.665308
##                            Gender_Female                            Gender_Male
##                               15.017940                               38.508376
##                                Techie_No                  Contract_Month_to_month
##                                 1.092151                                 2.441533
##                       Contract_One_year                           Port_modem_No
##                                 1.685699                                 1.029183
##                                Tablet_No                      InternetService_DSL
##                                 1.041397                               451.434531
##             InternetService_Fiber_Optic                                 Phone_No
##                                36.223074                                 1.048842
##                              Multiple_No                       OnlineSecurity_No
##                                 6.881755                                15.271264
##                         OnlineBackup_No                      DeviceProtection_No
##                                 5.647210                                 9.328813
##                          TechSupport_No                          StreamingTV_No
##                                 4.856047                                44.273127
##                      StreamingMovies_No                      PaperlessBilling_No
##                                22.157068                                 1.032294
## PaymentMethod_Bank_Transfer_automatic_  PaymentMethod_Credit_Card__automatic_
##                                 1.595173                                 1.553580
##          PaymentMethod_Electronic_Check
##                                 1.711518
```

```r
# Still too much multicollinearity. Review Correlation Matrix to determine which
# columns to remove

# Custom function found to improve readability of large correlation matrix
# source: http://www.sthda.com/english/wiki/correlation-matrix-a-quick-start-guide-to-analyze-format-an

# ++++++++++++++++++++++++++++
# flattenCorrMatrix
# ++++++++++++++++++++++++++++
# cormat : matrix of the correlation coefficients
# pmat : matrix of the correlation p-values
flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
```

```r
    column = rownames(cormat)[col(cormat)[ut]],
    cor  =(cormat)[ut],
    p = pmat[ut]
  )
}

dfTrain2RCor <- rcorr(as.matrix(dfTrain2))
flatCor <- flattenCorrMatrix(dfTrain2RCor$r, dfTrain2RCor$P)

# Create a subset of flatCor that have significant correlation values
highCorrDf <- flatCor[flatCor$cor > 0.75 | flatCor$cor < -0.75 , ]
highCorrDf
```

```
##                row            column        cor p
## 54          Tenure Bandwidth_GB_Year  0.9914758 0
## 3003 Gender_Female       Gender_Male -0.9552912 0
## 3160      Churn_No         Churn_Yes -1.0000000 0
```

```r
# Remove Gender_Female & Tenure

# Recreate traning dataframe and linear model without high correlation column(s)
dfTrain3 <- dfTrain2[, -which(names(dfTrain2) %in% c("Gender_Female", "Tenure"))]

colNames3 <- colnames(dfTrain3[, -which(names(dfTrain3) %in% c("Churn_Yes", "Churn_No"))])

predictorVars3 <- paste(colNames3, collapse = " + ")

logRegForm3 <- paste("Churn_Yes ~ ", predictorVars3, sep = "")

initialModel3 <- glm(formula = logRegForm3, family = "binomial", data = dfTrain3)

vifScores3 <- vif(initialModel3)
vifScores3
```

```
##                             Population                             Children
##                               1.157687                             1.032595
##                                    Age                               Income
##                               1.040654                             1.036962
##                     Outage_sec_perweek                                Email
##                               1.028976                             1.035882
##                               Contacts                  Yearly_equip_failure
##                               1.040542                             1.029360
##                          MonthlyCharge                     Bandwidth_GB_Year
##                              26.984630                             3.088986
##                               Response                                  Fix
##                               2.200687                             1.963199
##                            Replacement                           Reliability
##                               1.603502                             1.292881
##                                Options                          Respectful
##                               1.419682                             1.505535
##                              Courteous                           Listening
##                               1.329424                             1.236308
##                               State_AK                             State_AL
##                               3.643754                             4.760708
##                               State_AR                             State_AZ
```

```
##                          5.617083                          3.495241
##                          State_CA                          State_CO
##                         13.545746                          4.785820
##                          State_CT                          State_DC
##                          3.176472                          1.533743
##                          State_DE                          State_FL
##                          1.491932                          8.800033
##                          State_GA                          State_HI
##                          7.032459                          2.022672
##                          State_IA                          State_ID
##                          7.455605                          3.245224
##                          State_IL                          State_IN
##                         10.858681                          6.748232
##                          State_KS                          State_KY
##                          4.891250                          6.840905
##                          State_LA                          State_MA
##                          5.094795                          5.015020
##                          State_MD                          State_ME
##                          4.353219                          4.117854
##                          State_MI                          State_MN
##                          8.095272                          8.080175
##                          State_MO                          State_MS
##                          9.172972                          5.176303
##                          State_MT                          State_NC
##                          3.622725                          8.390932
##                          State_ND                          State_NE
##                          5.270418                          5.955732
##                          State_NH                          State_NJ
##                          3.145758                          5.936869
##                          State_NM                          State_NV
##                          4.590277                          2.005627
##                          State_NY                          State_OH
##                         15.257302                          9.413122
##                          State_OK                          State_OR
##                          6.685231                          3.895252
##                          State_PA                          State_PR
##                         14.346126                          1.573819
##                          State_RI                          State_SC
##                          1.111060                          3.864074
##                          State_SD                          State_TN
##                          2.669693                          5.207114
##                          State_TX                          State_UT
##                         16.426083                          2.456808
##                          State_VA                          State_VT
##                          7.524211                          2.382903
##                          State_WA                          State_WI
##                          6.278175                          7.144376
##                          State_WV                       Area_Rural
##                          7.230626                          1.352108
##                     Area_Suburban                  Marital_Divorced
##                          1.339063                          1.666716
##                   Marital_Married             Marital_Never_Married
##                          1.641120                          1.612251
##                 Marital_Separated                      Gender_Male
```

```
##                                1.663673                                1.035141
##                               Techie_No                 Contract_Month_to_month
##                                1.091842                                2.441099
##                       Contract_One_year                           Port_modem_No
##                                1.685807                                1.028936
##                               Tablet_No                     InternetService_DSL
##                                1.040765                                2.604568
##             InternetService_Fiber_Optic                                Phone_No
##                                5.919464                                1.048313
##                             Multiple_No                       OnlineSecurity_No
##                                5.076896                                1.056707
##                         OnlineBackup_No                     DeviceProtection_No
##                                2.904468                                1.628242
##                          TechSupport_No                          StreamingTV_No
##                                1.600083                                8.266480
##                      StreamingMovies_No                     PaperlessBilling_No
##                               11.735742                                1.031437
## PaymentMethod_Bank_Transfer_automatic_  PaymentMethod_Credit_Card__automatic_
##                                1.592327                                1.552805
##          PaymentMethod_Electronic_Check
##                                1.709706
```

```r
# Now that correlation has been reduced, I will analyze VIF scores. Those with a
# value above 5 will be removed.

# Convert to vifScores to a dataframe for easier manipulation
vifDf3 <- as.data.frame(vifScores3)
vifDf3 <- cbind(dv = rownames(vifDf3), vifDf3)
rownames(vifDf3) <- 1:nrow(vifDf3)
vifDf3
```

```
##                                       dv vifScores3
## 1                             Population   1.157687
## 2                               Children   1.032595
## 3                                    Age   1.040654
## 4                                 Income   1.036962
## 5                     Outage_sec_perweek   1.028976
## 6                                  Email   1.035882
## 7                               Contacts   1.040542
## 8                    Yearly_equip_failure   1.029360
## 9                          MonthlyCharge  26.984630
## 10                      Bandwidth_GB_Year   3.088986
## 11                               Response   2.200687
## 12                                    Fix   1.963199
## 13                            Replacement   1.603502
## 14                            Reliability   1.292881
## 15                                Options   1.419682
## 16                             Respectful   1.505535
## 17                              Courteous   1.329424
## 18                              Listening   1.236308
## 19                               State_AK   3.643754
## 20                               State_AL   4.760708
## 21                               State_AR   5.617083
## 22                               State_AZ   3.495241
## 23                               State_CA  13.545746
```

```
## 24                              State_CO   4.785820
## 25                              State_CT   3.176472
## 26                              State_DC   1.533743
## 27                              State_DE   1.491932
## 28                              State_FL   8.800033
## 29                              State_GA   7.032459
## 30                              State_HI   2.022672
## 31                              State_IA   7.455605
## 32                              State_ID   3.245224
## 33                              State_IL  10.858681
## 34                              State_IN   6.748232
## 35                              State_KS   4.891250
## 36                              State_KY   6.840905
## 37                              State_LA   5.094795
## 38                              State_MA   5.015020
## 39                              State_MD   4.353219
## 40                              State_ME   4.117854
## 41                              State_MI   8.095272
## 42                              State_MN   8.080175
## 43                              State_MO   9.172972
## 44                              State_MS   5.176303
## 45                              State_MT   3.622725
## 46                              State_NC   8.390932
## 47                              State_ND   5.270418
## 48                              State_NE   5.955732
## 49                              State_NH   3.145758
## 50                              State_NJ   5.936869
## 51                              State_NM   4.590277
## 52                              State_NV   2.005627
## 53                              State_NY  15.257302
## 54                              State_OH   9.413122
## 55                              State_OK   6.685231
## 56                              State_OR   3.895252
## 57                              State_PA  14.346126
## 58                              State_PR   1.573819
## 59                              State_RI   1.111060
## 60                              State_SC   3.864074
## 61                              State_SD   2.669693
## 62                              State_TN   5.207114
## 63                              State_TX  16.426083
## 64                              State_UT   2.456808
## 65                              State_VA   7.524211
## 66                              State_VT   2.382903
## 67                              State_WA   6.278175
## 68                              State_WI   7.144376
## 69                              State_WV   7.230626
## 70                           Area_Rural   1.352108
## 71                       Area_Suburban   1.339063
## 72                     Marital_Divorced   1.666716
## 73                      Marital_Married   1.641120
## 74                Marital_Never_Married   1.612251
## 75                    Marital_Separated   1.663673
## 76                         Gender_Male   1.035141
## 77                           Techie_No   1.091842
```

```
## 78                Contract_Month_to_month   2.441099
## 79                      Contract_One_year   1.685807
## 80                          Port_modem_No   1.028936
## 81                              Tablet_No   1.040765
## 82                    InternetService_DSL   2.604568
## 83            InternetService_Fiber_Optic   5.919464
## 84                               Phone_No   1.048313
## 85                           Multiple_No   5.076896
## 86                      OnlineSecurity_No   1.056707
## 87                        OnlineBackup_No   2.904468
## 88                    DeviceProtection_No   1.628242
## 89                         TechSupport_No   1.600083
## 90                         StreamingTV_No   8.266480
## 91                     StreamingMovies_No  11.735742
## 92                    PaperlessBilling_No   1.031437
## 93 PaymentMethod_Bank_Transfer_automatic_   1.592327
## 94  PaymentMethod_Credit_Card__automatic_   1.552805
## 95         PaymentMethod_Electronic_Check   1.709706
```

```
highVif3 <- vifDf3[vifDf3$vifScores3 > 5, ]
highVif3
```

```
##                            dv vifScores3
## 9                MonthlyCharge  26.984630
## 21                    State_AR   5.617083
## 23                    State_CA  13.545746
## 28                    State_FL   8.800033
## 29                    State_GA   7.032459
## 31                    State_IA   7.455605
## 33                    State_IL  10.858681
## 34                    State_IN   6.748232
## 36                    State_KY   6.840905
## 37                    State_LA   5.094795
## 38                    State_MA   5.015020
## 41                    State_MI   8.095272
## 42                    State_MN   8.080175
## 43                    State_MO   9.172972
## 44                    State_MS   5.176303
## 46                    State_NC   8.390932
## 47                    State_ND   5.270418
## 48                    State_NE   5.955732
## 50                    State_NJ   5.936869
## 53                    State_NY  15.257302
## 54                    State_OH   9.413122
## 55                    State_OK   6.685231
## 57                    State_PA  14.346126
## 62                    State_TN   5.207114
## 63                    State_TX  16.426083
## 65                    State_VA   7.524211
## 67                    State_WA   6.278175
## 68                    State_WI   7.144376
## 69                    State_WV   7.230626
## 83 InternetService_Fiber_Optic   5.919464
## 85                  Multiple_No   5.076896
## 90              StreamingTV_No   8.266480
```

```
## 91          StreamingMovies_No  11.735742
```

```r
highVifColNames = highVif3$dv

# Remove columns with high VIF values to improve the model
dfTrain4 <- dfTrain3[, !(names(dfTrain3)) %in% highVifColNames]

colNames4 <- colnames(dfTrain4[, -which(names(dfTrain4) %in% c("Churn_Yes", "Churn_No"))])

predictorVars4 <- paste(colNames4, collapse = " + ")

logRegForm4 <- paste("Churn_Yes ~ ", predictorVars4, sep = "")

initialModel4 <- glm(formula = logRegForm4, family = "binomial", data = dfTrain4)

vifScores4 <- vif(initialModel4)

vifDf4 <- as.data.frame(vifScores4)
vifDf4 <- cbind(dv = rownames(vifDf4), vifDf4)
rownames(vifDf4) <- 1:nrow(vifDf4)
vifDf4
```

```
##                                        dv vifScores4
## 1                              Population   1.043509
## 2                                Children   1.011322
## 3                                     Age   1.016409
## 4                                  Income   1.014383
## 5                      Outage_sec_perweek   1.010017
## 6                                   Email   1.009129
## 7                                Contacts   1.011952
## 8                     Yearly_equip_failure   1.011681
## 9                        Bandwidth_GB_Year   1.169911
## 10                               Response   2.251477
## 11                                    Fix   1.955916
## 12                            Replacement   1.599852
## 13                            Reliability   1.277703
## 14                                Options   1.388031
## 15                             Respectful   1.515552
## 16                              Courteous   1.325045
## 17                              Listening   1.220991
## 18                                State_AK   1.009096
## 19                                State_AL   1.010724
## 20                                State_AZ   1.013382
## 21                                State_CO   1.012128
## 22                                State_CT   1.007842
## 23                                State_DC   1.008335
## 24                                State_DE   1.009998
## 25                                State_HI   1.007140
## 26                                State_ID   1.009721
## 27                                State_KS   1.015741
## 28                                State_MD   1.009516
## 29                                State_ME   1.013685
## 30                                State_MT   1.015798
## 31                                State_NH   1.008038
## 32                                State_NM   1.012204
```

```
## 33                              State_NV   1.008958
## 34                              State_OR   1.011030
## 35                              State_PR   1.022129
## 36                              State_RI   1.008787
## 37                              State_SC   1.011242
## 38                              State_SD   1.013036
## 39                              State_UT   1.007195
## 40                              State_VT   1.010227
## 41                            Area_Rural   1.351662
## 42                         Area_Suburban   1.346828
## 43                      Marital_Divorced   1.637378
## 44                       Marital_Married   1.578545
## 45                  Marital_Never_Married   1.580992
## 46                      Marital_Separated   1.615795
## 47                            Gender_Male   1.016486
## 48                              Techie_No   1.016636
## 49                Contract_Month_to_month   1.725459
## 50                      Contract_One_year   1.628795
## 51                          Port_modem_No   1.011216
## 52                              Tablet_No   1.018093
## 53                    InternetService_DSL   1.096724
## 54                               Phone_No   1.017788
## 55                      OnlineSecurity_No   1.014177
## 56                        OnlineBackup_No   1.026956
## 57                    DeviceProtection_No   1.016952
## 58                         TechSupport_No   1.013858
## 59                    PaperlessBilling_No   1.012997
## 60 PaymentMethod_Bank_Transfer_automatic_   1.534794
## 61  PaymentMethod_Credit_Card__automatic_   1.501451
## 62         PaymentMethod_Electronic_Check   1.650736
```

```
highVif4 <- vifDf4[vifDf4$vifScores4 > 5, ]
highVif4
```

```
## [1] dv         vifScores4
## <0 rows> (or 0-length row.names)
```

```
# There are no VIF values above 5 remaining. This means there is confidence in
# the p-values from initialModel4

summary(initialModel4)
```

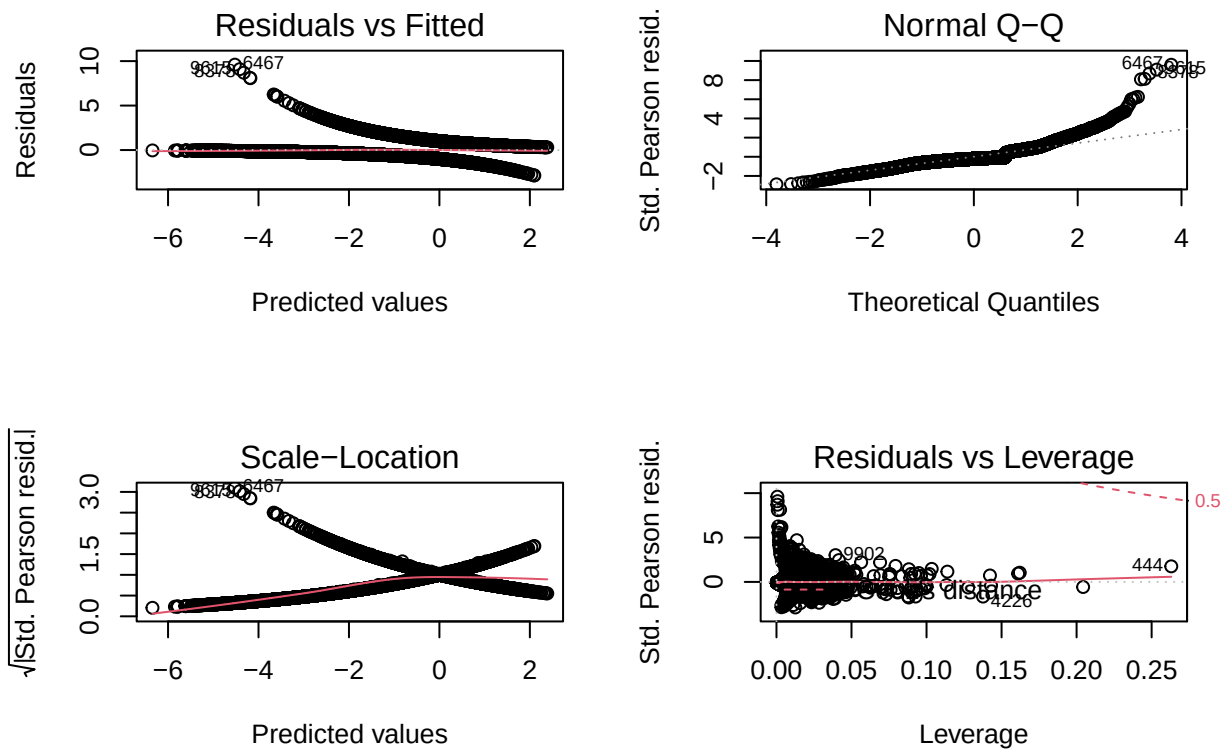
```
##
## Call:
## glm(formula = logRegForm4, family = "binomial", data = dfTrain4)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1046  -0.6239  -0.2959   0.6532   3.0114
##
## Coefficients:
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             2.185e-01  4.027e-01   0.542  0.58748
## Population                             -6.771e-07  2.349e-06  -0.288  0.77317
## Children                                1.520e-02  1.569e-02   0.968  0.33286
```

```
## Age                                  9.639e-05  1.616e-03   0.060  0.95244
## Income                              -5.680e-07  1.168e-06  -0.486  0.62674
## Outage_sec_perweek                  -2.597e-03  1.114e-02  -0.233  0.81571
## Email                                9.833e-03  1.104e-02   0.891  0.37292
## Contacts                             2.556e-02  3.338e-02   0.766  0.44376
## Yearly_equip_failure                -2.003e-02  5.309e-02  -0.377  0.70593
## Bandwidth_GB_Year                   -6.847e-04  1.986e-05 -34.469  < 2e-16
## Response                             5.517e-03  4.746e-02   0.116  0.90746
## Fix                                 -4.308e-02  4.461e-02  -0.966  0.33419
## Replacement                         -1.854e-02  4.055e-02  -0.457  0.64752
## Reliability                         -1.403e-02  3.649e-02  -0.385  0.70054
## Options                              6.852e-03  3.782e-02   0.181  0.85622
## Respectful                           2.824e-02  3.926e-02   0.719  0.47200
## Courteous                           -2.220e-02  3.735e-02  -0.595  0.55215
## Listening                            2.676e-02  3.537e-02   0.757  0.44930
## State_AK                             5.475e-02  3.526e-01   0.155  0.87660
## State_AL                            -4.882e-01  2.684e-01  -1.819  0.06891
## State_AZ                             2.483e-01  3.038e-01   0.817  0.41378
## State_CO                             1.811e-01  2.650e-01   0.684  0.49427
## State_CT                            -7.774e-02  3.621e-01  -0.215  0.82999
## State_DC                             2.652e-01  7.924e-01   0.335  0.73790
## State_DE                             2.165e-01  6.548e-01   0.331  0.74093
## State_HI                             1.114e-01  6.249e-01   0.178  0.85856
## State_ID                             1.150e-01  3.716e-01   0.309  0.75696
## State_KS                             2.135e-01  2.356e-01   0.906  0.36486
## State_MD                             2.516e-01  2.837e-01   0.887  0.37513
## State_ME                             9.995e-02  2.994e-01   0.334  0.73848
## State_MT                             1.721e-01  3.382e-01   0.509  0.61096
## State_NH                             6.136e-02  3.748e-01   0.164  0.86996
## State_NM                            -2.423e-01  2.988e-01  -0.811  0.41735
## State_NV                            -8.389e-01  5.271e-01  -1.591  0.11152
## State_OR                             1.384e-01  3.313e-01   0.418  0.67625
## State_PR                            -1.014e-01  5.901e-01  -0.172  0.86356
## State_RI                            -2.277e+00  1.107e+00  -2.057  0.03972
## State_SC                            -1.860e-01  3.216e-01  -0.578  0.56306
## State_SD                            -6.156e-01  3.743e-01  -1.645  0.10006
## State_UT                             1.190e-01  3.938e-01   0.302  0.76256
## State_VT                            -1.562e-01  4.175e-01  -0.374  0.70825
## Area_Rural                           8.136e-04  8.171e-02   0.010  0.99206
## Area_Suburban                        4.414e-02  8.133e-02   0.543  0.58732
## Marital_Divorced                    -1.075e-01  1.027e-01  -1.047  0.29523
## Marital_Married                     -1.609e-01  1.063e-01  -1.514  0.13002
## Marital_Never_Married               -1.785e-01  1.065e-01  -1.677  0.09355
## Marital_Separated                    3.611e-02  1.044e-01   0.346  0.72932
## Gender_Male                          1.831e-01  6.671e-02   2.745  0.00605
## Techie_No                           -4.886e-01  8.655e-02  -5.645 1.65e-08
## Contract_Month_to_month              1.852e+00  9.121e-02  20.304  < 2e-16
## Contract_One_year                    1.765e-01  1.129e-01   1.563  0.11815
## Port_modem_No                       -9.053e-02  6.655e-02  -1.360  0.17373
## Tablet_No                            7.767e-02  7.368e-02   1.054  0.29185
## InternetService_DSL                  9.938e-01  7.123e-02  13.953  < 2e-16
## Phone_No                             1.468e-01  1.120e-01   1.311  0.18976
## OnlineSecurity_No                   -5.154e-03  6.993e-02  -0.074  0.94125
## OnlineBackup_No                     -4.565e-01  6.712e-02  -6.801 1.04e-11
```

```
## DeviceProtection_No                    -3.437e-01  6.692e-02  -5.136 2.81e-07
## TechSupport_No                         -1.340e-01  6.839e-02  -1.960  0.04999
## PaperlessBilling_No                    -8.015e-02  6.762e-02  -1.185  0.23590
## PaymentMethod_Bank_Transfer_automatic_ -1.508e-01  9.934e-02  -1.518  0.12905
## PaymentMethod_Credit_Card__automatic_  -4.148e-03  1.016e-01  -0.041  0.96742
## PaymentMethod_Electronic_Check          1.776e-01  8.895e-02   1.997  0.04580
##
## (Intercept)
## Population
## Children
## Age
## Income
## Outage_sec_perweek
## Email
## Contacts
## Yearly_equip_failure
## Bandwidth_GB_Year                      ***
## Response
## Fix
## Replacement
## Reliability
## Options
## Respectful
## Courteous
## Listening
## State_AK
## State_AL                               .
## State_AZ
## State_CO
## State_CT
## State_DC
## State_DE
## State_HI
## State_ID
## State_KS
## State_MD
## State_ME
## State_MT
## State_NH
## State_NM
## State_NV
## State_OR
## State_PR
## State_RI                               *
## State_SC
## State_SD
## State_UT
## State_VT
## Area_Rural
## Area_Suburban
## Marital_Divorced
## Marital_Married
## Marital_Never_Married                  .
## Marital_Separated
```

```
## Gender_Male                            **
## Techie_No                              ***
## Contract_Month_to_month                ***
## Contract_One_year
## Port_modem_No
## Tablet_No
## InternetService_DSL                    ***
## Phone_No
## OnlineSecurity_No
## OnlineBackup_No                        ***
## DeviceProtection_No                    ***
## TechSupport_No                         *
## PaperlessBilling_No
## PaymentMethod_Bank_Transfer_automatic_
## PaymentMethod_Credit_Card__automatic_
## PaymentMethod_Electronic_Check         *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8211.0  on 6999  degrees of freedom
## Residual deviance: 5675.2  on 6937  degrees of freedom
## AIC: 5801.2
##
## Number of Fisher Scoring iterations: 5
```

```r
par(mfrow = c(2,2))
plot(initialModel4)
```




```r
# Those that are noteworthy have p-values <0.05
initModel4Df <- tidy(initialModel4)
options(scipen = 999)

# Selecting useful predictor variables
reducedModelCols <- initModel4Df$term[initModel4Df$p.value < 0.05 ]

# Removing "Intercept"
reducedModelCols <- reducedModelCols[reducedModelCols != "(Intercept)"]
reducedModelCols
```

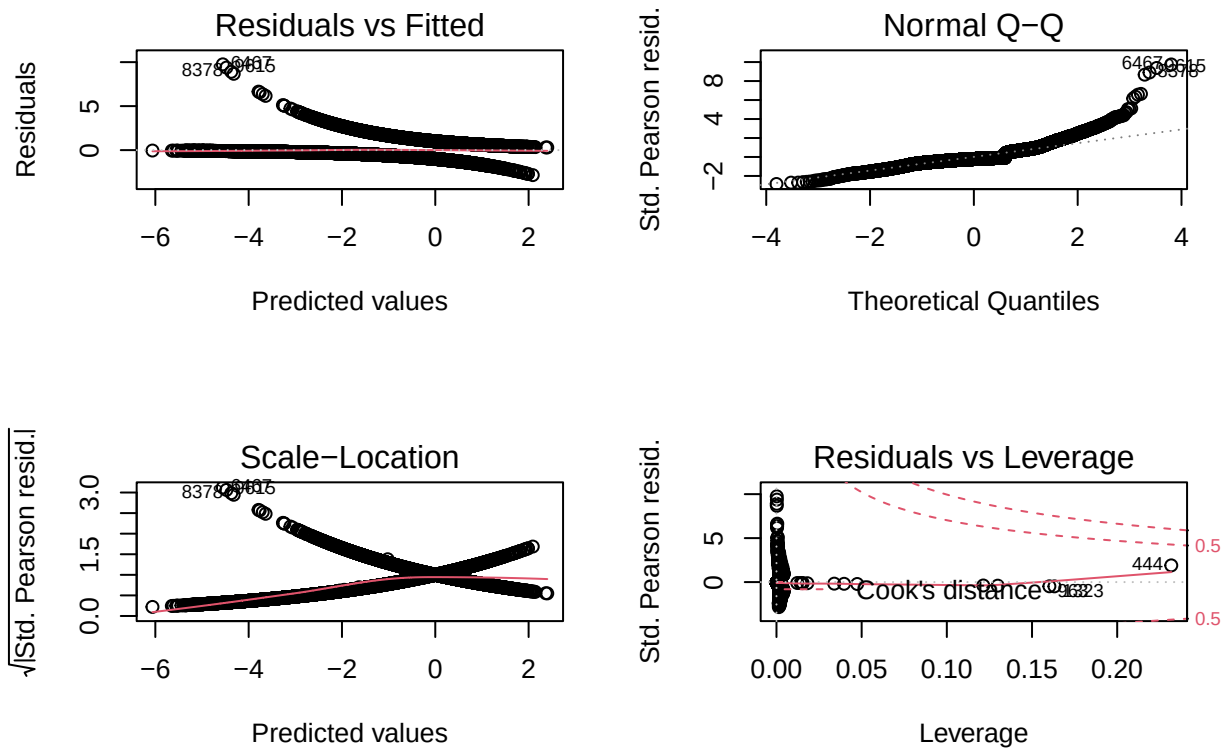
```
##  [1] "Bandwidth_GB_Year"              "State_RI"
##  [3] "Gender_Male"                    "Techie_No"
##  [5] "Contract_Month_to_month"        "InternetService_DSL"
##  [7] "OnlineBackup_No"                "DeviceProtection_No"
##  [9] "TechSupport_No"                 "PaymentMethod_Electronic_Check"
```

```r
# Creating a reduced model
reducedPredictorVars <- paste(reducedModelCols, collapse = " + ")

reducedLogRegForm <- paste("Churn_Yes ~ ", reducedPredictorVars, sep = "")

reducedModel <- glm(formula = reducedLogRegForm, family = "binomial", data = dfTrain)

# Check for VIF
vif(reducedModel)
```

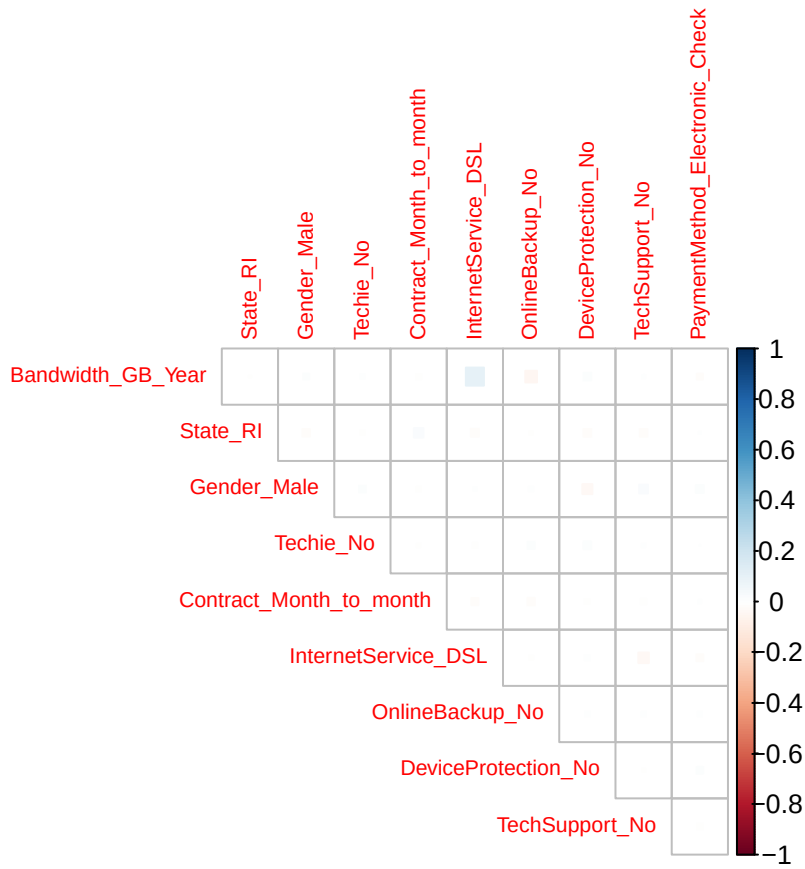
```
##              Bandwidth_GB_Year                        State_RI
```

```
##                        1.156037                        1.002069
##                     Gender_Male                       Techie_No
##                        1.004136                        1.006974
##         Contract_Month_to_month            InternetService_DSL
##                        1.088392                        1.085749
##                 OnlineBackup_No            DeviceProtection_No
##                        1.016972                        1.004996
##                  TechSupport_No PaymentMethod_Electronic_Check
##                        1.001948                        1.004421
```

```
# View standard plots
par(mfrow = c(2,2))
plot(reducedModel)
```




```
# Check for correlation
reducedCorMatrix <- cor(dfTrain[, names(dfTrain) %in% reducedModelCols])
par(mfrow = c(1,1))
corrplot(reducedCorMatrix,
         method = 'square',
         type = 'upper',
         tl.cex = .7,
         diag = FALSE,
         order = 'original')
```

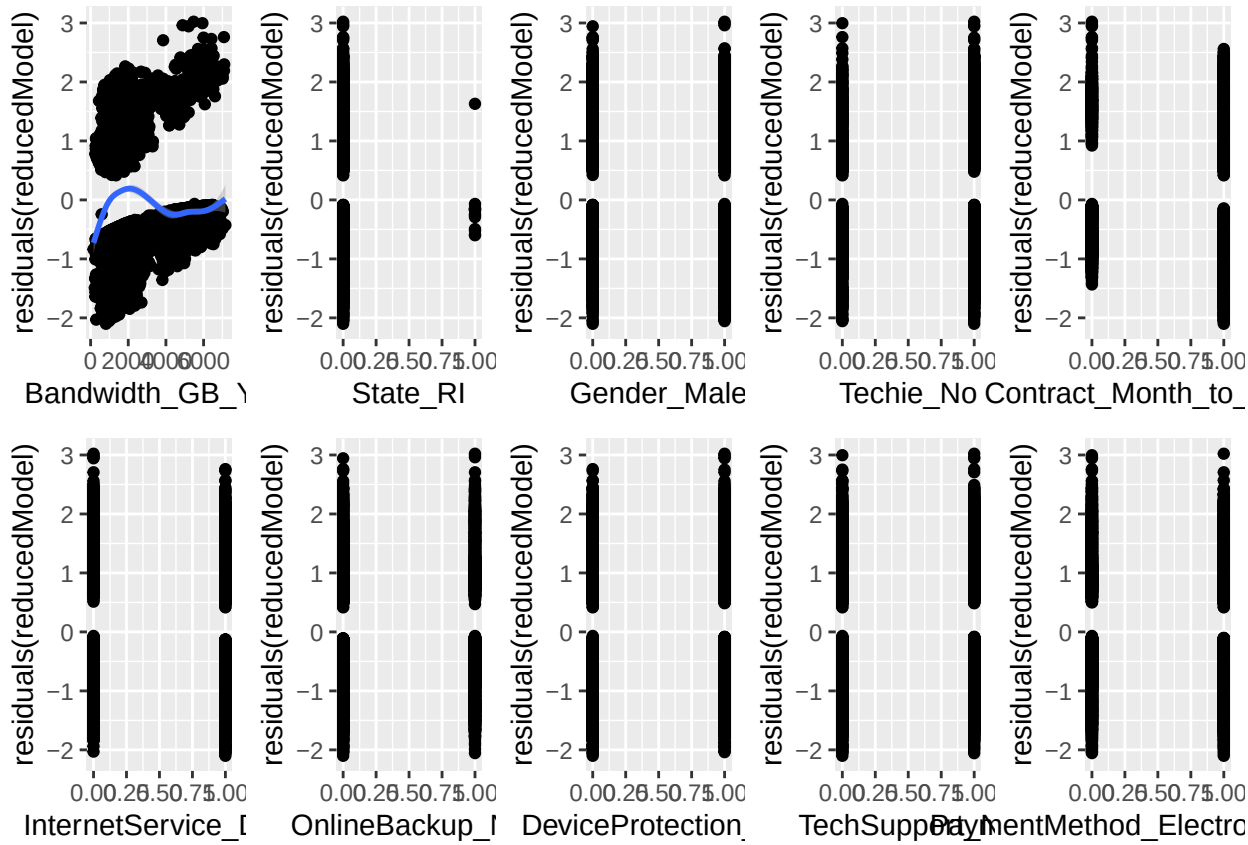



```r
# View residuals by individual predictor variables
plot_grid(
  # plot 1: Bandwidth
  ggplot(dfTrain, aes(x=Bandwidth_GB_Year, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
  # plot 2: State
  ggplot(dfTrain, aes(x=State_RI, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
  # plot 3: Gender
  ggplot(dfTrain, aes(x=Gender_Male, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
  # Plot 4: Techie
  ggplot(dfTrain, aes(x=Techie_No, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
  # Plot 5: Contract
  ggplot(dfTrain, aes(x=Contract_Month_to_month, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
  # Plot 6: Internet
  ggplot(dfTrain, aes(x=InternetService_DSL, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
```

```r
  # Plot 7: Online Backup
  ggplot(dfTrain, aes(x=OnlineBackup_No, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
  # Plot 8: Device Protection
  ggplot(dfTrain, aes(x=DeviceProtection_No, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
  # Plot 9: Device Protection
  ggplot(dfTrain, aes(x=TechSupport_No, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),
  # Plot 10: Payment Method
  ggplot(dfTrain, aes(x=PaymentMethod_Electronic_Check, y=residuals(reducedModel))) +
    geom_point() +
    geom_smooth(),

  ncol = 5, nrow = 2
)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
# Checking Prediction Data accuracy
summary(reducedModel) # All predictor variables are significant
```

```
##
## Call:
## glm(formula = reducedLogRegForm, family = "binomial", data = dfTrain)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1004  -0.6353  -0.3009   0.6677   3.0215
##
## Coefficients:
##                                  Estimate  Std. Error z value
## (Intercept)                    0.18232395  0.12440694   1.466
## Bandwidth_GB_Year             -0.00068130  0.00001967 -34.628
## State_RI                      -2.21766612  1.09018742  -2.034
## Gender_Male                    0.18743504  0.06606727   2.837
## Techie_No                     -0.48922746  0.08566468  -5.711
## Contract_Month_to_month        1.75606124  0.07216314  24.335
## InternetService_DSL            0.97731054  0.07061959  13.839
## OnlineBackup_No               -0.45462064  0.06655163  -6.831
## DeviceProtection_No           -0.34571302  0.06627877  -5.216
## TechSupport_No                -0.12317466  0.06775621  -1.818
## PaymentMethod_Electronic_Check  0.21732126  0.06914537   3.143
##                                          Pr(>|z|)
## (Intercept)                               0.14277
```

```
## Bandwidth_GB_Year              < 0.0000000000000002 ***
## State_RI                                   0.04193 *
## Gender_Male                                0.00455 **
## Techie_No                         0.00000001123413 ***
## Contract_Month_to_month        < 0.0000000000000002 ***
## InternetService_DSL            < 0.0000000000000002 ***
## OnlineBackup_No                   0.00000000000843 ***
## DeviceProtection_No               0.00000018278381 ***
## TechSupport_No                             0.06908 .
## PaymentMethod_Electronic_Check             0.00167 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8211.0  on 6999  degrees of freedom
## Residual deviance: 5713.1  on 6989  degrees of freedom
## AIC: 5735.1
##
## Number of Fisher Scoring iterations: 5
```

```r
dfTest$predictionData <- predict(reducedModel, newdata = dfTest, type = "response")

actual_response <- dfTest$Churn_Yes
predicted_response <- round(dfTest$predictionData)
outcomes <- table(predicted_response, actual_response)

confusion <- confusionMatrix(outcomes)
confusion
```

```
## Confusion Matrix and Statistics
##
##                   actual_response
## predicted_response    0    1
##                  0 1998  292
##                  1  265  445
##
##                Accuracy : 0.8143
##                  95% CI : (0.7999, 0.8281)
##     No Information Rate : 0.7543
##     P-Value [Acc > NIR] : 0.000000000000002116
##
##                   Kappa : 0.4928
##
##  Mcnemar's Test P-Value : 0.2706
##
##             Sensitivity : 0.8829
##             Specificity : 0.6038
##          Pos Pred Value : 0.8725
##          Neg Pred Value : 0.6268
##              Prevalence : 0.7543
##          Detection Rate : 0.6660
##    Detection Prevalence : 0.7633
##       Balanced Accuracy : 0.7433
##
```

```
##        'Positive' Class : 0
##
```