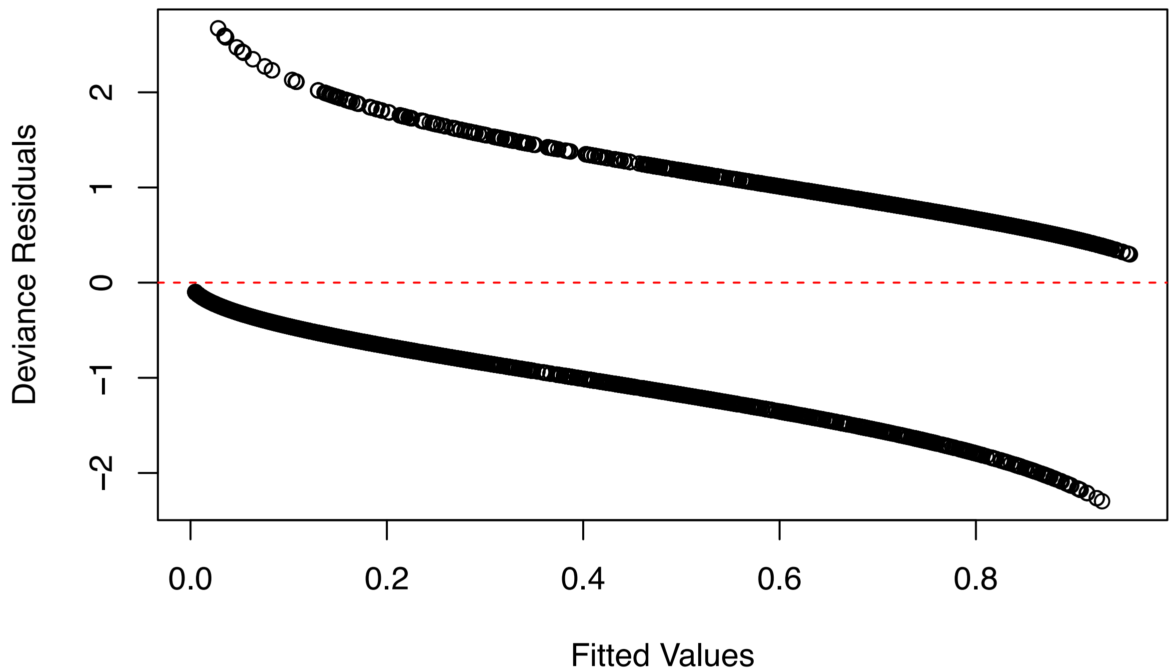
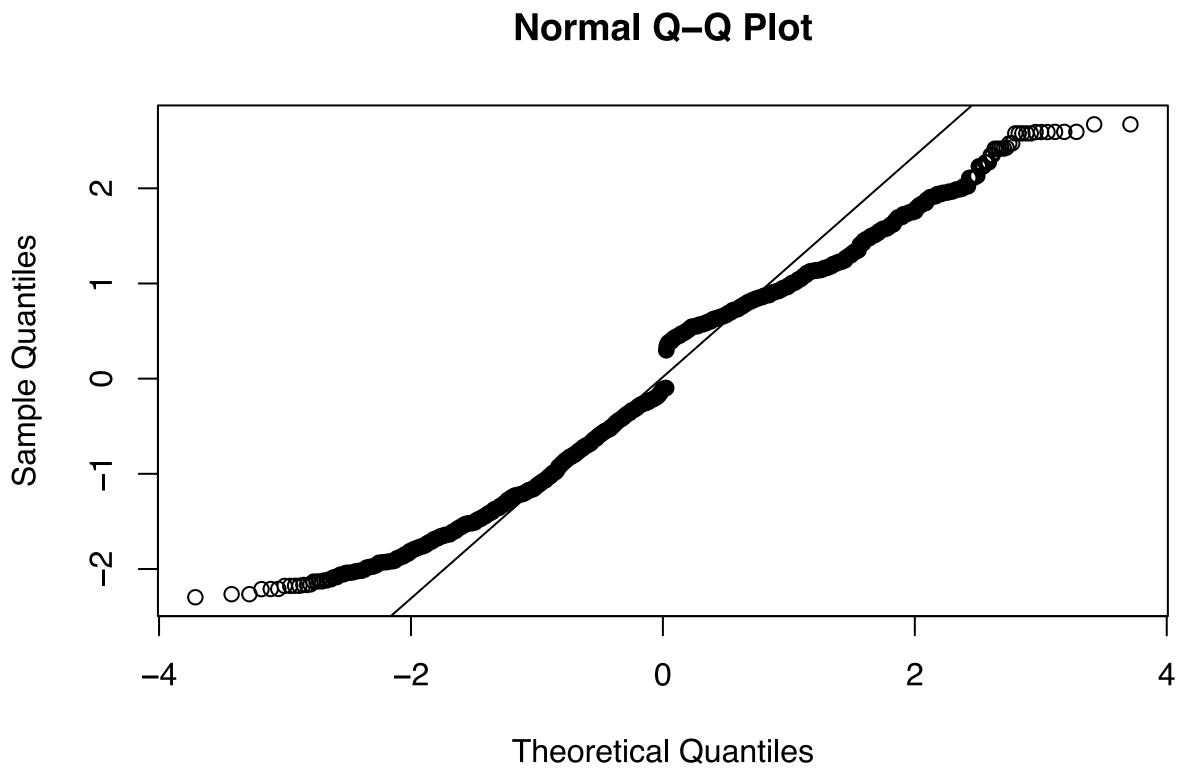
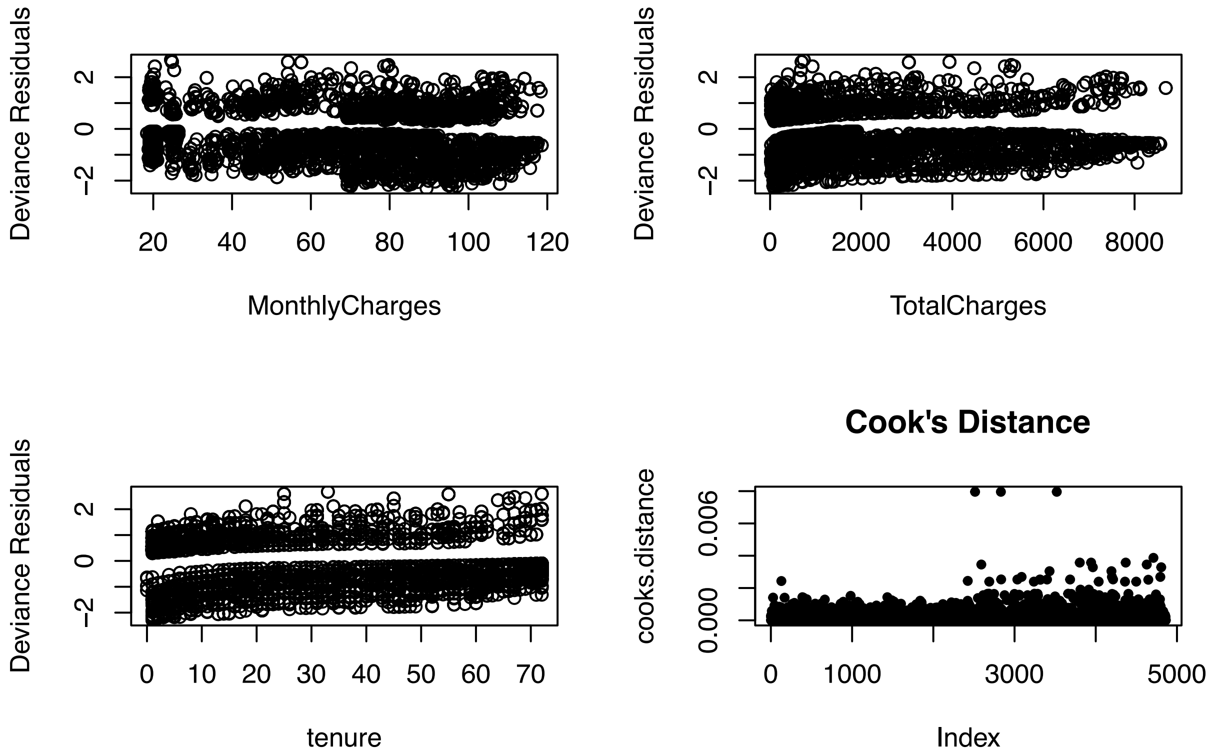


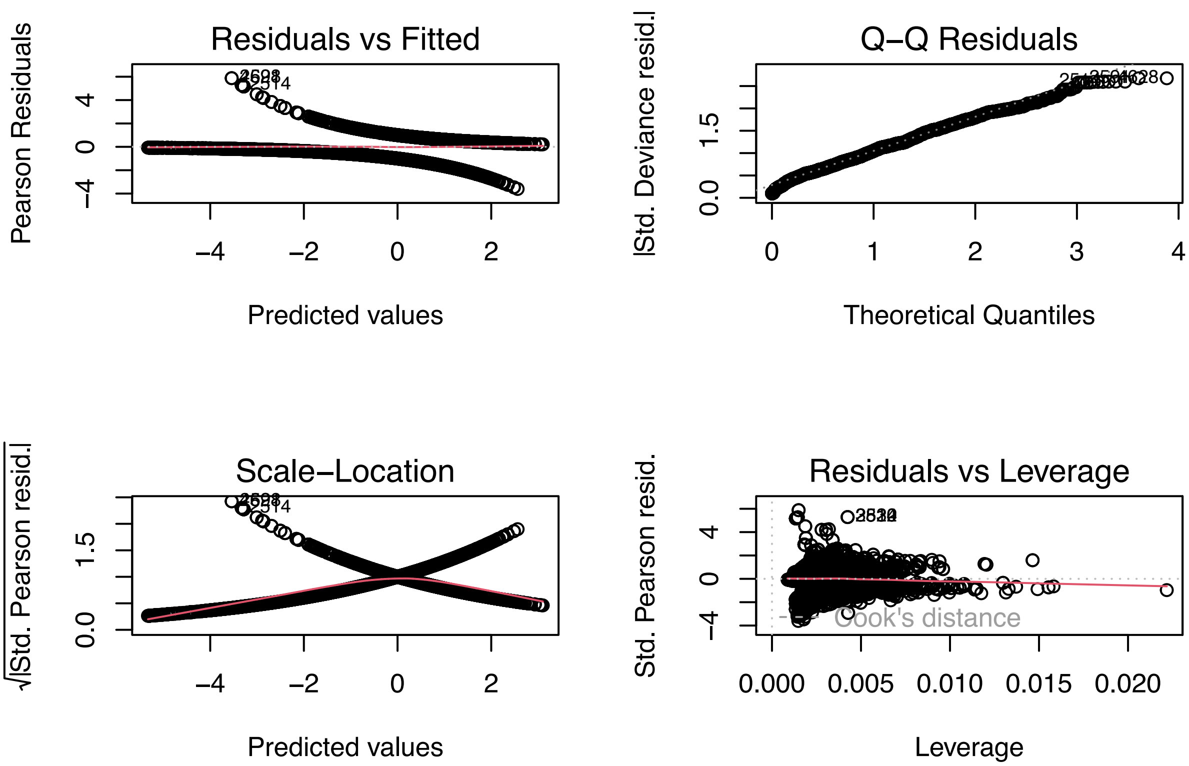

```r
qqnorm(residuals2)
qqline(residuals2)
```




```r
par(mfrow = c(2, 2))
plot(train$MonthlyCharges, residuals2, ylab = "Deviance Residuals", xlab = "MonthlyCharges")
plot(train$TotalCharges, residuals2, ylab = "Deviance Residuals", xlab = "TotalCharges")
plot(train$tenure, residuals2, ylab = "Deviance Residuals", xlab = "tenure")
cooks.distance <- cooks.distance(f_model)
plot(cooks.distance, pch = 20, main = "Cook's Distance")
```




```r
plot(f_model)
```




```
# Now the normal Q-Q Plot does not follow such a linear pattern
```

Remove data points with a Cook's Distance greater than .001

```
cooksd2 <- cooks.distance(f_model)
outliers2 <- which(cooksd2 > 0.002)
train_clean2 <- train[-outliers2, ]


# Create model without  outliers
# f_model_cleaned <- glm(Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure, 3) +

f_model_cleaned <- glm(Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure, 3) + S


#Test it only the train_clean2:
print(f_model_cleaned)
```

```
##
## Call:  glm(formula = Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges,
##     2) + poly(tenure, 3) + SeniorCitizen + InternetService +
##     StreamingTV + StreamingMovies + Contract, family = "binomial",
##     data = train_clean2)
##
## Coefficients:
```

```
##                     (Intercept)            poly(MonthlyCharges, 2)1
##                         -1.0135                            -41.4221
##        poly(MonthlyCharges, 2)2              poly(TotalCharges, 2)1
##                          3.6818                             44.6777
##          poly(TotalCharges, 2)2                    poly(tenure, 3)1
##                         10.8576                           -101.2892
##                poly(tenure, 3)2                    poly(tenure, 3)3
##                         -3.8141                            -16.5357
##                  SeniorCitizen1         InternetServiceFiber optic
##                          0.4880                              1.6222
##               InternetServiceNo      StreamingTVNo internet service
##                         -1.3513                                  NA
##                  StreamingTVYes  StreamingMoviesNo internet service
##                          0.5719                                  NA
##              StreamingMoviesYes                    ContractOne year
##                          0.5015                             -0.8105
##                ContractTwo year
##                         -2.0572
##
## Degrees of Freedom: 4830 Total (i.e. Null);  4816 Residual
## Null Deviance:       6694
## Residual Deviance: 4540  AIC: 4570
```

```r
evaluation(f_model_cleaned, train_clean2)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

## Training Set Metrics:
## AUC: 0.8511899
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1768  712
##    Yes  405 1946
## Recall: 0.8277329
## F1-score: 0.7770014

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.8454064
## [1] "Confusion Matrix:"
##       Predicted
```

```
## Actual   No  Yes
##    No  1127  457
##    Yes   87  499
## Recall: 0.8515358
## F1-score: 0.6472114
```

```r
#Trying the step function
step_model <- step( f_model_cleaned, k= log(nrow(df)))
```

```
## Start:  AIC=4672.74
## Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure,
##     3) + SeniorCitizen + InternetService + StreamingTV + StreamingMovies +
##     Contract
##
##                           Df Deviance    AIC
## - poly(TotalCharges, 2)    2   4552.5 4667.7
## - poly(MonthlyCharges, 2)  2   4553.4 4668.6
## <none>                         4539.8 4672.7
## - StreamingMovies          1   4561.0 4685.0
## - SeniorCitizen            1   4565.6 4689.7
## - StreamingTV              1   4566.9 4690.9
## - poly(tenure, 3)          3   4633.9 4740.2
## - InternetService          1   4618.3 4742.3
## - Contract                 2   4678.1 4793.3
##
## Step:  AIC=4667.66
## Churn ~ poly(MonthlyCharges, 2) + poly(tenure, 3) + SeniorCitizen +
##     InternetService + StreamingTV + StreamingMovies + Contract
##
##                           Df Deviance    AIC
## <none>                         4552.5 4667.7
## - poly(MonthlyCharges, 2)  2   4570.8 4668.3
## - StreamingMovies          1   4569.5 4675.8
## - StreamingTV              1   4574.3 4680.6
## - SeniorCitizen            1   4577.9 4684.2
## - InternetService          1   4622.6 4728.9
## - Contract                 2   4698.6 4796.1
## - poly(tenure, 3)          3   4862.2 4950.8
```

```r
summary(step_model)
```

```
##
## Call:
## glm(formula = Churn ~ poly(MonthlyCharges, 2) + poly(tenure,
##     3) + SeniorCitizen + InternetService + StreamingTV + StreamingMovies +
##     Contract, family = "binomial", data = train_clean2)
##
## Coefficients: (2 not defined because of singularities)
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -0.78140    0.14733  -5.304 1.13e-07 ***
## poly(MonthlyCharges, 2)1         -26.67568    9.88197  -2.699 0.006946 **
## poly(MonthlyCharges, 2)2          15.37508    4.42251   3.477 0.000508 ***
## poly(tenure, 3)1                 -62.98780    4.46284 -14.114  < 2e-16 ***
```

```
## poly(tenure, 3)2                     11.02733    3.13907   3.513 0.000443 ***
## poly(tenure, 3)3                    -14.49748    2.81922  -5.142 2.71e-07 ***
## SeniorCitizen1                        0.48386    0.09713   4.982 6.30e-07 ***
## InternetServiceFiber optic            1.49500    0.18075   8.271  < 2e-16 ***
## InternetServiceNo                    -1.50771    0.23938  -6.298 3.01e-10 ***
## StreamingTVNo internet service             NA         NA      NA       NA
## StreamingTVYes                        0.50482    0.10870   4.644 3.41e-06 ***
## StreamingMoviesNo internet service         NA         NA      NA       NA
## StreamingMoviesYes                    0.44339    0.10802   4.105 4.05e-05 ***
## ContractOne year                     -0.79932    0.11091  -7.207 5.73e-13 ***
## ContractTwo year                     -2.08310    0.20009 -10.411  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6693.7  on 4830  degrees of freedom
## Residual deviance: 4552.5  on 4818  degrees of freedom
## AIC: 4578.5
##
## Number of Fisher Scoring iterations: 5
```

```r
evaluation(step_model, train_clean2)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Training Set Metrics:
## AUC: 0.8510832
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1806  674
##    Yes  458 1893
## Recall: 0.8051893
## F1-score: 0.7698251

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.8474828
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1148  441
```

```
##    Yes   98  488
## Recall: 0.8327645
## F1-score: 0.6442244
```

```r
#This lowers the AIC, but also lowers the recall
#So we'll leave the previous model as best
```

Idea: what if we have a separate model for categorizing points that have a high cooks distance? Maybe there is something unique to outliers that gives away whether or not they will churn.

Building the best model with the interactions

```r
# Builds a model using a formula and data, returns the model and its F1-score on the training data
build_model_formula <- function(formula, data) {
  # Fit the logistic regression model
  model <- glm(formula, data = data, family = "binomial")

  # Make predictions on the training set
  predictions <- predict(model, newdata = data, type = "response")

  # Convert probabilities to binary predictions
  binary_predictions <- ifelse(predictions > 0.5, 1, 0)

  # Calculate confusion matrix
  confusion_matrix <- table(data$Churn, binary_predictions)

  # Calculate precision, recall, and F-score
  precision <- confusion_matrix[2, 2] / sum(confusion_matrix[, 2])
  recall <- confusion_matrix[2, 2] / sum(confusion_matrix[2, ])
  f_score <- 2 * (precision * recall) / (precision + recall)

  return(list(model = model, f_score = f_score))
}

#Finds the best interactions to fit into the model
find_best_model2 <- function(data) {
  # List of factors to consider
  covariates <- c("poly(MonthlyCharges,2)", "poly(TotalCharges,2)", "poly(tenure,3)")
  selected_factors <- c("SeniorCitizen", "InternetService", "StreamingTV", "StreamingMovies", "Contr
  factors <- c("SeniorCitizen", "Partner", "MultipleLines", "InternetService", "OnlineBackup", "Stre

  # Initialize variables to store the best model and F-score
  best_global_model <- NULL
  best_global_f_score <- 0

  # Choose best interactions starting at 1 interaction
  flen <- length(factors)
  chosen_interactions <- c()
  while(TRUE) {

    best_model <- NULL
    best_f_score <- 0
    # Add a dummy interaction to increase the number of interactions for the next for loop
    chosen_interactions <- c(chosen_interactions, "Contract")
```

```r
    for (i in 1:flen^2) {

      # Remove the previously added interaction
      chosen_interactions <- chosen_interactions[-length(chosen_interactions)]

       # Create the forumula with the chosen interactions
      if(i %% flen == (i-1) %/% flen) next
      chosen_interactions <- c(chosen_interactions, paste(factors[i %% flen + 1], factors[(i-1) %/%
      formula <- reformulate(c(covariates, selected_factors, chosen_interactions), response = "Churn
      # print(formula)

      # Build the model
      result <- build_model_formula(formula, data)
      model <- result$model
      f_score <- result$f_score

      # Update the best model if the F-score is higher
      if (f_score > best_f_score) {
        best_model <- model
        best_f_score <- f_score
      }
    }
    if (best_f_score > best_global_f_score) {
        best_global_model <- best_model
        best_global_f_score <- best_f_score
    } else break
  }

# Return the best model and its F-score
  return(list(model = best_model, f_score = best_f_score))
}
```

```r
# Call the function with your training data
result <- find_best_model2(train_clean2)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```r
best_model_Winte <- result$model
best_f_score <- result$f_score

# Print the best model and its F-score
best_model_Winte$formula
```

```
## Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure,
##     3) + SeniorCitizen + InternetService + StreamingTV + StreamingMovies +
##     Contract + SeniorCitizen * InternetService
## <environment: 0x13feb2958>
```

```r
print(best_model_Winte)
```

```
##
## Call:  glm(formula = formula, family = "binomial", data = data)
##
## Coefficients:
```

```
##                               (Intercept)
##                                   -1.0659
##                  poly(MonthlyCharges, 2)1
##                                  -40.5207
##                  poly(MonthlyCharges, 2)2
##                                    2.4183
##                    poly(TotalCharges, 2)1
##                                   49.4840
##                    poly(TotalCharges, 2)2
##                                   11.0555
##                          poly(tenure, 3)1
##                                 -105.8805
##                          poly(tenure, 3)2
##                                   -4.5028
##                          poly(tenure, 3)3
##                                  -16.4412
##                            SeniorCitizen1
##                                    0.9283
##                InternetServiceFiber optic
##                                    1.6829
##                         InternetServiceNo
##                                   -1.2524
##            StreamingTVNo internet service
##                                        NA
##                            StreamingTVYes
##                                    0.5654
##        StreamingMoviesNo internet service
##                                        NA
##                        StreamingMoviesYes
##                                    0.5030
##                          ContractOne year
##                                   -0.8128
##                          ContractTwo year
##                                   -2.0624
## SeniorCitizen1:InternetServiceFiber optic
##                                   -0.5666
##          SeniorCitizen1:InternetServiceNo
##                                   -0.2769
##
## Degrees of Freedom: 4830 Total (i.e. Null);  4814 Residual
## Null Deviance:       6694
## Residual Deviance: 4534  AIC: 4568
```

```
evaluation(best_model_Winte, train_clean2)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Setting levels: control = No, case = Yes
```

```
## Setting direction: controls < cases
```

```
## Training Set Metrics:
```

```
## AUC: 0.8517806
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1774  706
##    Yes  395 1956
## Recall: 0.8319864
## F1-score: 0.780371

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.845624
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1117  467
##    Yes   84  502
## Recall: 0.8566553
## F1-score: 0.6456592
```

```
#Best model with interactions

f_model <- glm(Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure, 3) + InternetS
print(f_model)
```

```
##
## Call:  glm(formula = Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges,
##     2) + poly(tenure, 3) + InternetService + StreamingTV + StreamingMovies +
##     Contract + SeniorCitizen * PaymentMethod + Contract * SeniorCitizen,
##     family = "binomial", data = train_clean2)
##
## Coefficients:
##                                      (Intercept)
##                                         -1.02420
##                          poly(MonthlyCharges, 2)1
##                                        -40.36063
##                          poly(MonthlyCharges, 2)2
##                                          3.83958
##                            poly(TotalCharges, 2)1
##                                         51.61468
##                            poly(TotalCharges, 2)2
##                                          9.17601
##                                  poly(tenure, 3)1
##                                       -107.61070
##                                  poly(tenure, 3)2
##                                         -2.68595
##                                  poly(tenure, 3)3
##                                        -16.36771
```

```
##                          InternetServiceFiber optic
##                                             1.51021
##                                   InternetServiceNo
##                                            -1.24303
##                      StreamingTVNo internet service
##                                                  NA
##                                       StreamingTVYes
##                                             0.55438
##                  StreamingMoviesNo internet service
##                                                  NA
##                                  StreamingMoviesYes
##                                             0.44684
##                                     ContractOne year
##                                            -0.72956
##                                     ContractTwo year
##                                            -1.91322
##                                       SeniorCitizen1
##                                             0.17070
##                PaymentMethodCredit card (automatic)
##                                            -0.02012
##                       PaymentMethodElectronic check
##                                             0.22597
##                           PaymentMethodMailed check
##                                            -0.16058
## SeniorCitizen1:PaymentMethodCredit card (automatic)
##                                             0.77175
##        SeniorCitizen1:PaymentMethodElectronic check
##                                             0.30561
##            SeniorCitizen1:PaymentMethodMailed check
##                                             0.49719
##                     ContractOne year:SeniorCitizen1
##                                            -0.27793
##                     ContractTwo year:SeniorCitizen1
##                                            -0.42190
##
## Degrees of Freedom: 4830 Total (i.e. Null);  4808 Residual
## Null Deviance:       6694
## Residual Deviance: 4519  AIC: 4565
```

```r
evaluation(f_model, train_clean2)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

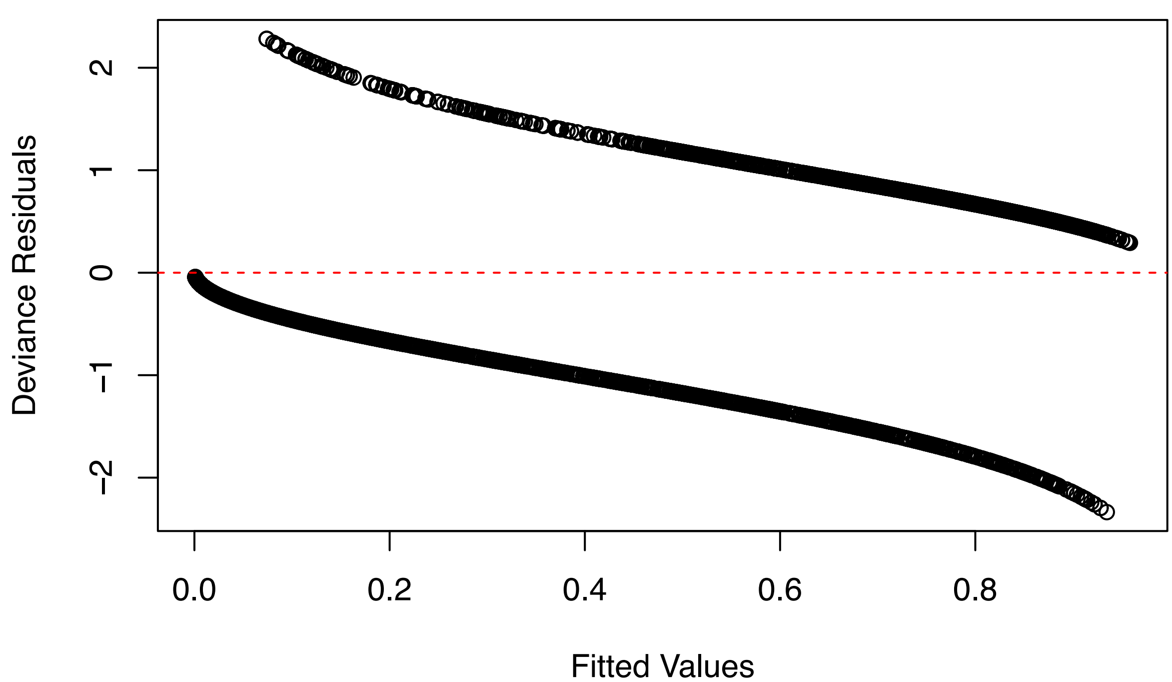
```
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
```

```
## Training Set Metrics:
## AUC: 0.8527371
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1808  672
```

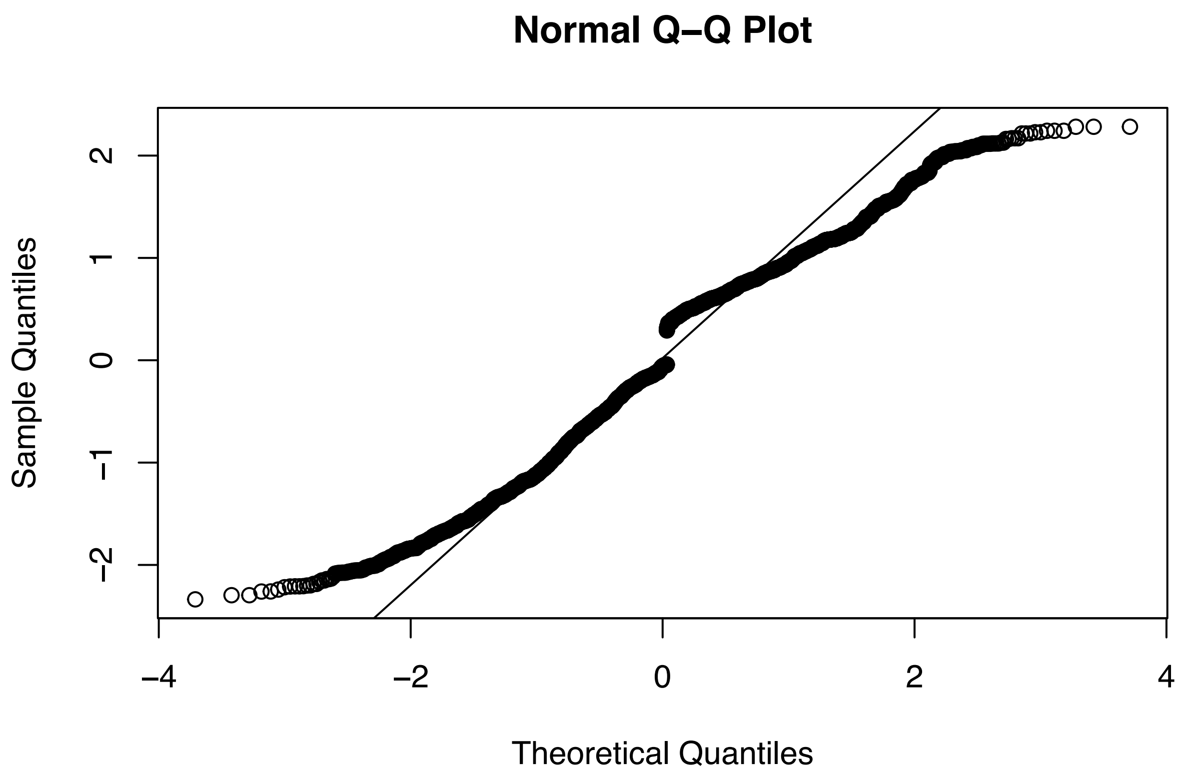
```
##    Yes  455 1896
## Recall: 0.8064653
## F1-score: 0.7708884

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.8472944
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1156  428
##    Yes  104  482
## Recall: 0.8225256
## F1-score: 0.644385
```

Residual Analysis of Numeric + Factor + Interactions Model

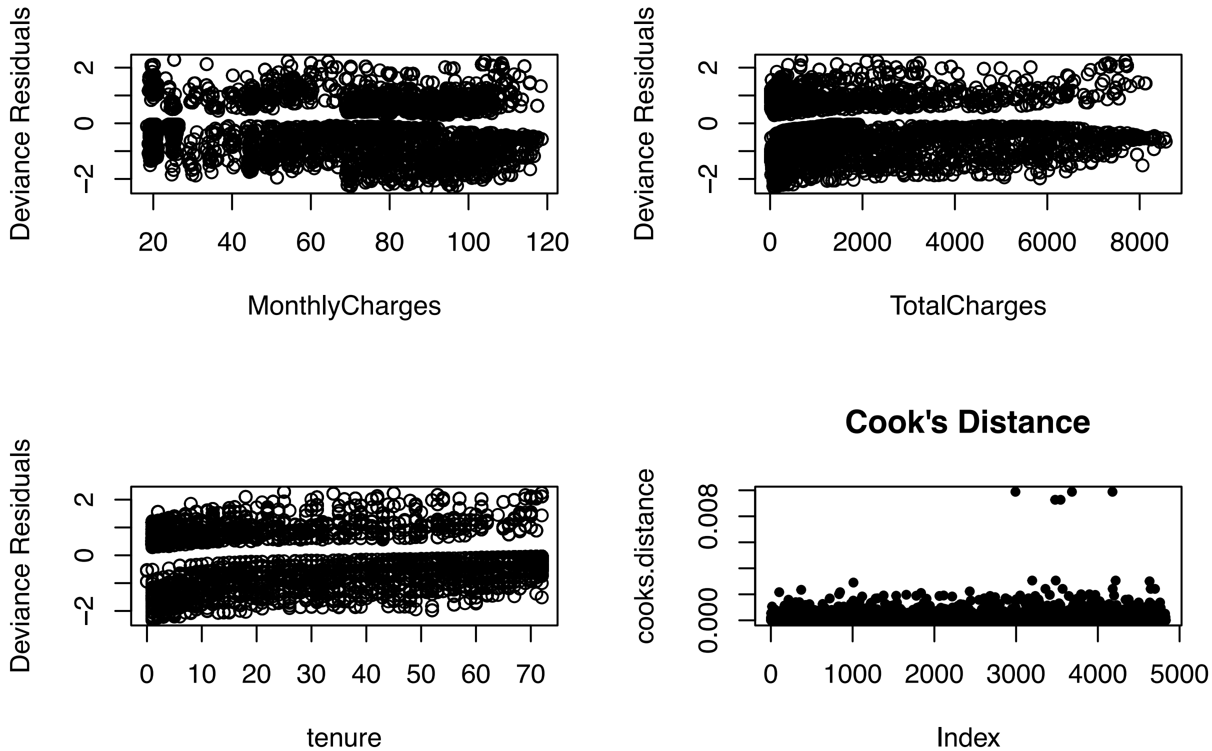
```r
residuals2 <- residuals(f_model, type = "deviance")
#For checking heteroscedasticity
plot(fitted(f_model), residuals2, ylab = "Deviance Residuals", xlab = "Fitted Values")
abline(h = 0, col = "red", lty = 2)
```



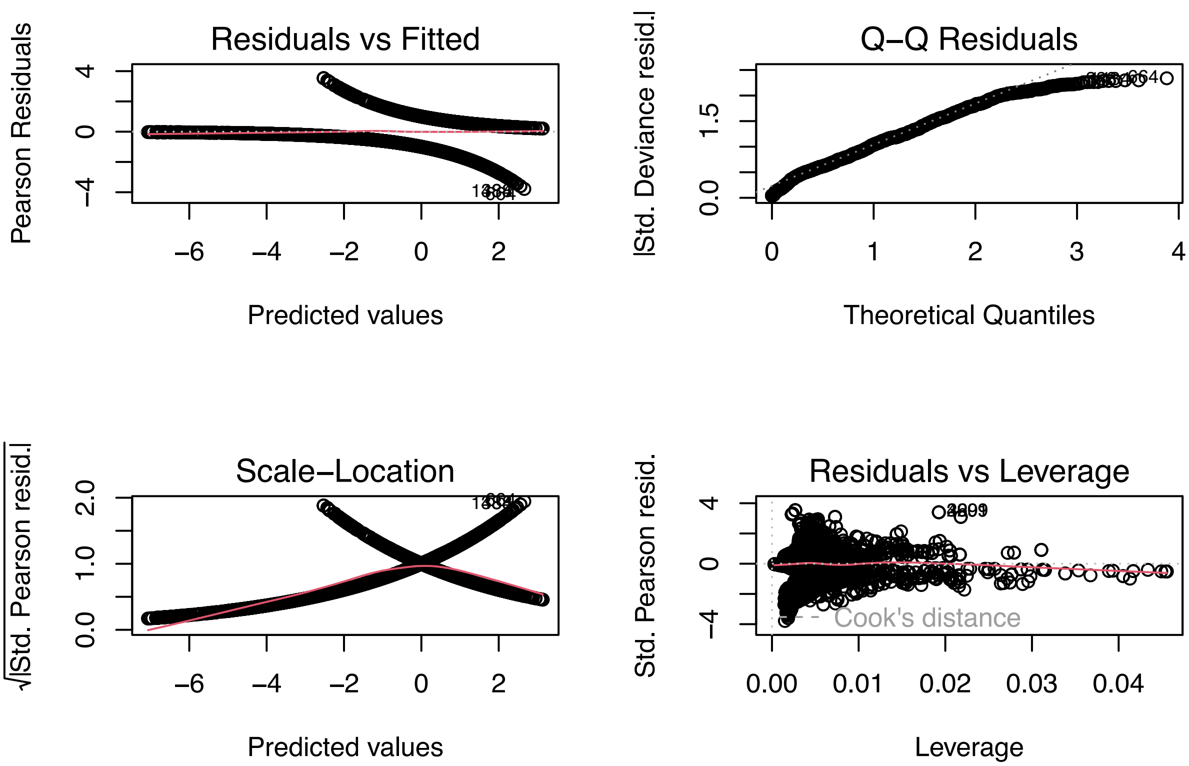
```
qqnorm(residuals2)
qqline(residuals2)
```



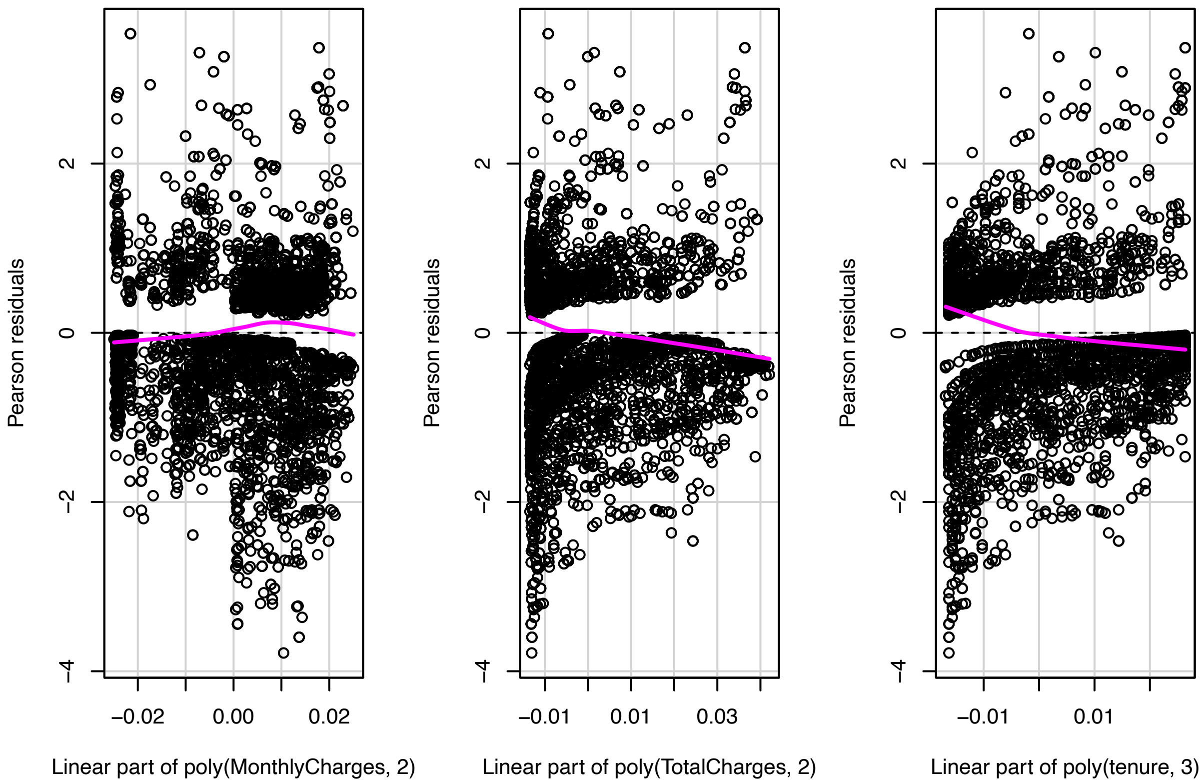

```
par(mfrow = c(2, 2))
plot(train_clean2$MonthlyCharges, residuals2, ylab = "Deviance Residuals", xlab = "MonthlyCharges")
plot(train_clean2$TotalCharges, residuals2, ylab = "Deviance Residuals", xlab = "TotalCharges")
plot(train_clean2$tenure, residuals2, ylab = "Deviance Residuals", xlab = "tenure")
cooks.distance <- cooks.distance(f_model)
plot(cooks.distance, pch = 20, main = "Cook's Distance")
```



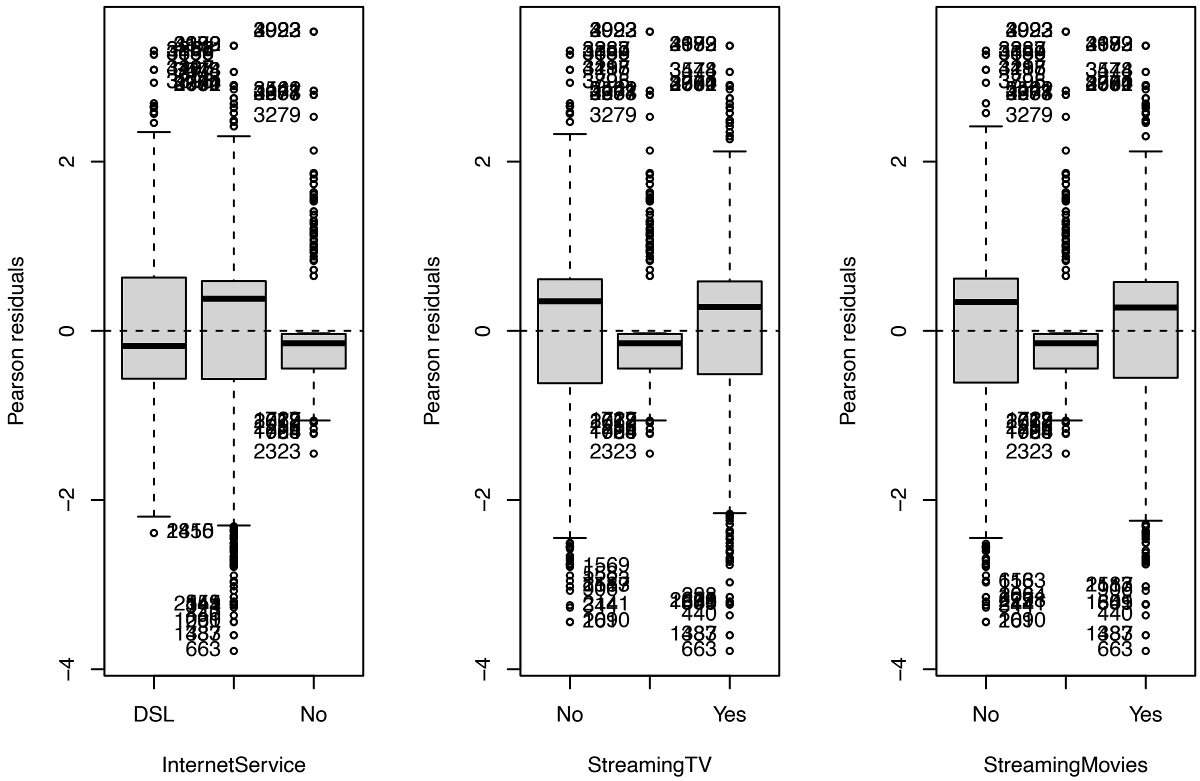

```r
plot(f_model)
```

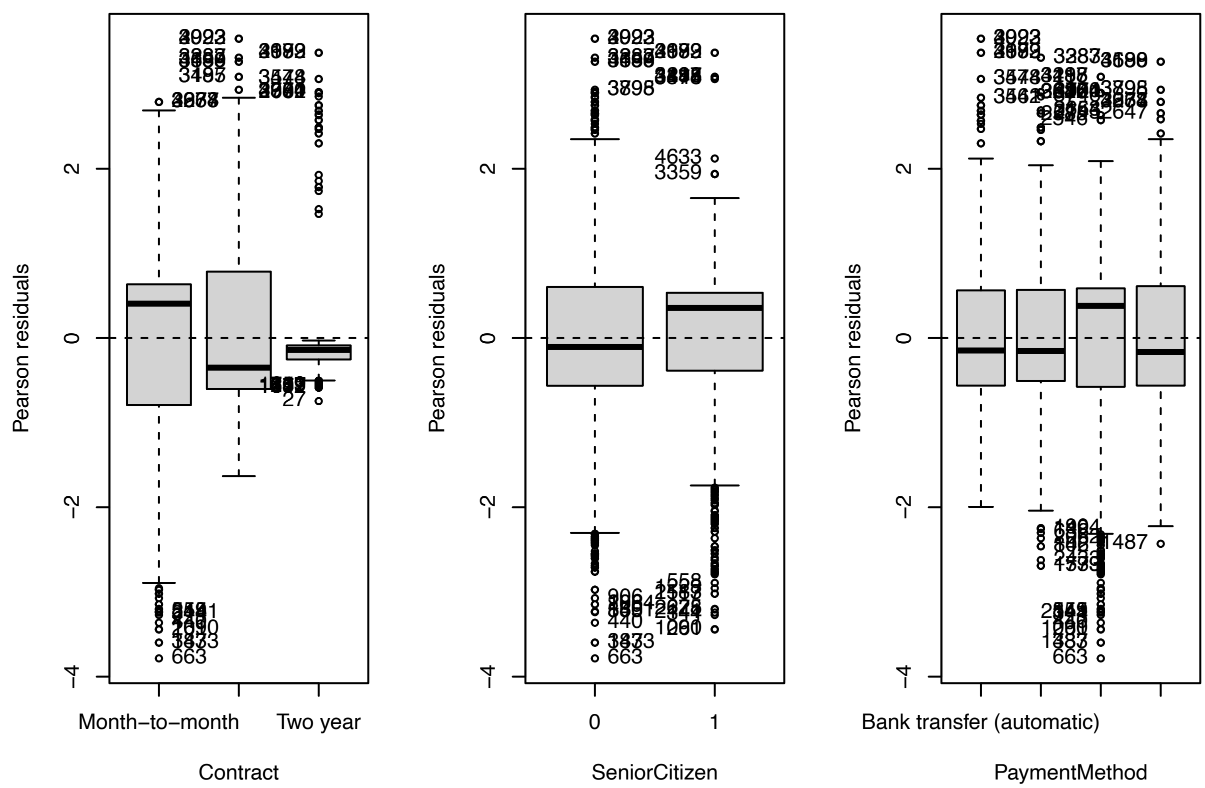



```r
#Added
residualPlots(f_model, layout=c(1, 3))
```

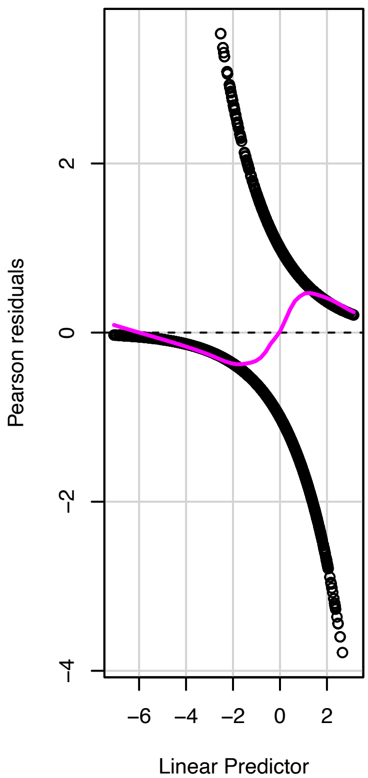


Pearson residuals
2
0
−2
−4
−0.02
0.00
0.02
Linear part of poly(MonthlyCharges, 2)
Pearson residuals
2
0
−2
−4
−0.01
0.01
0.03
Linear part of poly(TotalCharges, 2)
Pearson residuals
2
0
−2
−4
−0.01
0.01
Linear part of poly(tenure, 3)

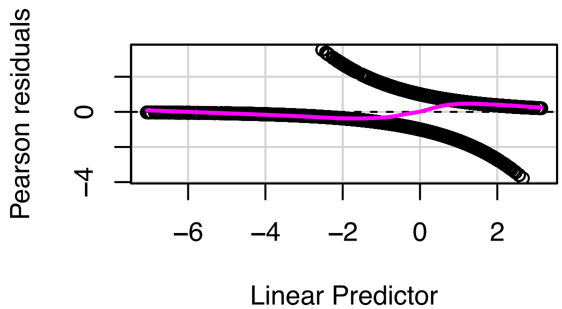

Pearson residuals
InternetService
DSL
No
StreamingTV
No
Yes
StreamingMovies
No
Yes

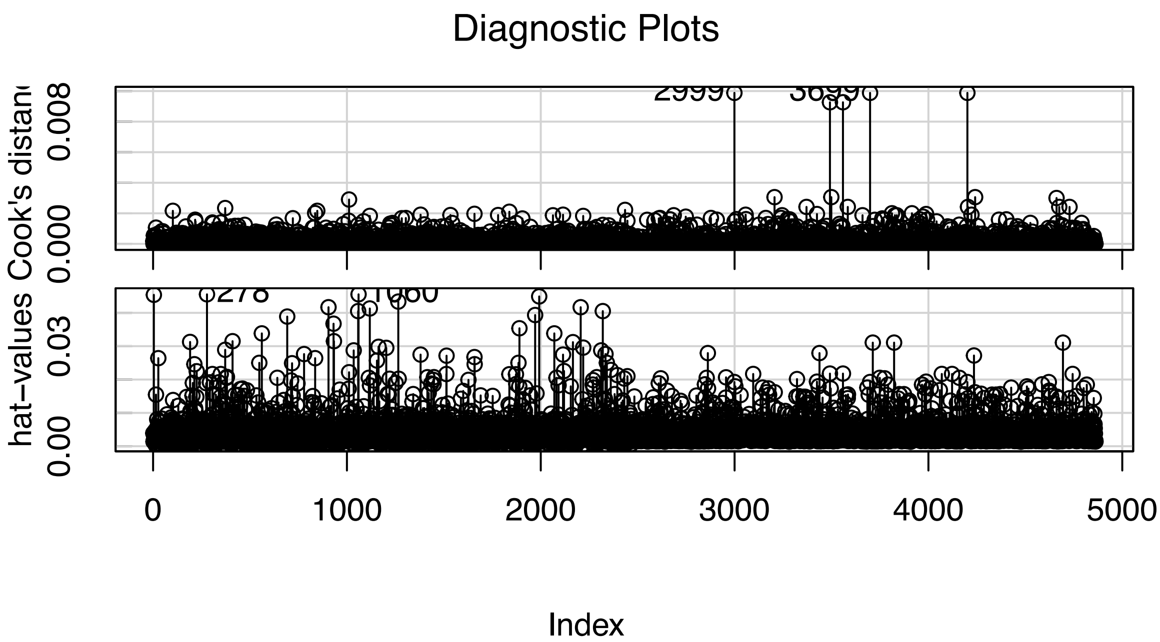


Warning in residualPlots.default(model, ...): No possible lack-of-fit tests




```
residualPlot(f_model)
?residualPlots
influenceIndexPlot(f_model, vars=c("Cook", "hat"))
```


Pearson residuals
0
−4
−6
−4
−2
0
2
Linear Predictor




```
# Now the normal Q-Q Plot does not follow such a linear pattern
```

Remove influential points

```
cooksd2 <- cooks.distance(f_model)
outliers3 <- which(cooksd2 > 0.002)
train_clean3 <- train_clean2[-outliers3, ]


# Create model without  outliers
f_model_cleaned3 <- glm(Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure, 3) +
print(f_model_cleaned3)
```

```
##
## Call:  glm(formula = Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges,
##     2) + poly(tenure, 3) + InternetService + StreamingTV + StreamingMovies +
##     Contract + SeniorCitizen * PaymentMethod + Contract * SeniorCitizen,
##     family = "binomial", data = train_clean3)
##
## Coefficients:
##                                         (Intercept)
##                                            -1.03445
##                             poly(MonthlyCharges, 2)1
##                                           -40.53821
##                             poly(MonthlyCharges, 2)2
```

```
##                                               4.46690
##                              poly(TotalCharges, 2)1
##                                              61.76584
##                              poly(TotalCharges, 2)2
##                                              10.08875
##                                    poly(tenure, 3)1
##                                            -117.90566
##                                    poly(tenure, 3)2
##                                              -4.41398
##                                    poly(tenure, 3)3
##                                             -16.29033
##                          InternetServiceFiber optic
##                                               1.50846
##                                   InternetServiceNo
##                                              -1.22906
##                      StreamingTVNo internet service
##                                                    NA
##                                       StreamingTVYes
##                                               0.58171
##                  StreamingMoviesNo internet service
##                                                    NA
##                                   StreamingMoviesYes
##                                               0.41474
##                                     ContractOne year
##                                              -0.75806
##                                     ContractTwo year
##                                              -1.94601
##                                       SeniorCitizen1
##                                               0.08858
##               PaymentMethodCredit card (automatic)
##                                              -0.02116
##                      PaymentMethodElectronic check
##                                               0.22261
##                          PaymentMethodMailed check
##                                              -0.17435
## SeniorCitizen1:PaymentMethodCredit card (automatic)
##                                               1.06508
##        SeniorCitizen1:PaymentMethodElectronic check
##                                               0.40381
##            SeniorCitizen1:PaymentMethodMailed check
##                                               0.92579
##                     ContractOne year:SeniorCitizen1
##                                              -0.76121
##                     ContractTwo year:SeniorCitizen1
##                                             -14.42896
##
## Degrees of Freedom: 4807 Total (i.e. Null);  4785 Residual
## Null Deviance:       6661
## Residual Deviance: 4415  AIC: 4461
```

```
evaluation(f_model_cleaned3, train_clean3)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

## Training Set Metrics:
## AUC: 0.8571402
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1796  677
##    Yes  418 1917
## Recall: 0.820985
## F1-score: 0.7778454

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.8467934
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1149  435
##    Yes   95  491
## Recall: 0.837884
## F1-score: 0.6494709
```

```r
# Model has better AIC and F1score than the previous, so we reiterate to get it even better

result <- find_best_model2(train_clean3)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```r
best_model_Winte2 <- result$model
best_f_score2 <- result$f_score
best_model_Winte2$formula
```

```
## Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure,
##     3) + SeniorCitizen + InternetService + StreamingTV + StreamingMovies +
##     Contract + PaymentMethod * InternetService
## <environment: 0x10ec31780>
```

```r
print(best_model_Winte2)
```

```
##
## Call:  glm(formula = formula, family = "binomial", data = data)
##
## Coefficients:
##                                                      (Intercept)
##                                                          -1.2993
```

```
##                                              poly(MonthlyCharges, 2)1
##                                                              -45.8149
##                                              poly(MonthlyCharges, 2)2
##                                                                4.0706
##                                                poly(TotalCharges, 2)1
##                                                               59.4328
##                                                poly(TotalCharges, 2)2
##                                                               11.6779
##                                                      poly(tenure, 3)1
##                                                             -115.0405
##                                                      poly(tenure, 3)2
##                                                               -5.5501
##                                                      poly(tenure, 3)3
##                                                              -16.1588
##                                                        SeniorCitizen1
##                                                                0.4043
##                                           InternetServiceFiber optic
##                                                                1.7680
##                                                     InternetServiceNo
##                                                               -0.6624
##                                       StreamingTVNo internet service
##                                                                    NA
##                                                        StreamingTVYes
##                                                                0.6050
##                                   StreamingMoviesNo internet service
##                                                                    NA
##                                                    StreamingMoviesYes
##                                                                0.4485
##                                                      ContractOne year
##                                                               -0.8888
##                                                      ContractTwo year
##                                                               -2.1547
##                                 PaymentMethodCredit card (automatic)
##                                                                0.6361
##                                        PaymentMethodElectronic check
##                                                                0.2607
##                                            PaymentMethodMailed check
##                                                                0.2112
## InternetServiceFiber optic:PaymentMethodCredit card (automatic)
##                                                               -0.5829
##         InternetServiceNo:PaymentMethodCredit card (automatic)
##                                                               -1.4639
##       InternetServiceFiber optic:PaymentMethodElectronic check
##                                                                0.1092
##                InternetServiceNo:PaymentMethodElectronic check
##                                                               -0.7566
##           InternetServiceFiber optic:PaymentMethodMailed check
##                                                               -0.5876
##                    InternetServiceNo:PaymentMethodMailed check
##                                                               -0.7180
##
## Degrees of Freedom: 4807 Total (i.e. Null);  4784 Residual
## Null Deviance:       6661
## Residual Deviance: 4424  AIC: 4472
```

```r
evaluation(best_model_Winte2, train_clean3)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Training Set Metrics:
## AUC: 0.8574414
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1807  666
##    Yes  400 1935
## Recall: 0.8286938
## F1-score: 0.7840357

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.8454629
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1146  438
##    Yes   93  493
## Recall: 0.8412969
## F1-score: 0.649967
```

```r
Anova(best_model_Winte2)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Churn
##                              LR Chisq Df Pr(>Chisq)
## poly(MonthlyCharges, 2)        15.927  2  0.0003480 ***
## poly(TotalCharges, 2)          17.923  2  0.0001283 ***
## poly(tenure, 3)                95.360  3  < 2.2e-16 ***
## SeniorCitizen                  16.779  1  4.201e-05 ***
## InternetService                69.715  1  < 2.2e-16 ***
## StreamingTV                    28.855  1  7.799e-08 ***
## StreamingMovies                16.171  1  5.787e-05 ***
## Contract                      142.970  2  < 2.2e-16 ***
## PaymentMethod                  13.257  3  0.0041131 **
## InternetService:PaymentMethod  20.813  6  0.0019821 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# This reiteration has given us the same model as before

# Print the best model and its F-score
print(best_model_Winte2)
```

```
##
## Call:  glm(formula = formula, family = "binomial", data = data)
##
## Coefficients:
##                                                 (Intercept)
##                                                     -1.2993
##                                       poly(MonthlyCharges, 2)1
##                                                    -45.8149
##                                       poly(MonthlyCharges, 2)2
##                                                      4.0706
##                                         poly(TotalCharges, 2)1
##                                                     59.4328
##                                         poly(TotalCharges, 2)2
##                                                     11.6779
##                                               poly(tenure, 3)1
##                                                   -115.0405
##                                               poly(tenure, 3)2
##                                                     -5.5501
##                                               poly(tenure, 3)3
##                                                    -16.1588
##                                                 SeniorCitizen1
##                                                      0.4043
##                                     InternetServiceFiber optic
##                                                      1.7680
##                                              InternetServiceNo
##                                                     -0.6624
##                                 StreamingTVNo internet service
##                                                          NA
##                                                 StreamingTVYes
##                                                      0.6050
##                             StreamingMoviesNo internet service
##                                                          NA
##                                             StreamingMoviesYes
##                                                      0.4485
##                                               ContractOne year
##                                                     -0.8888
##                                               ContractTwo year
##                                                     -2.1547
##                         PaymentMethodCredit card (automatic)
##                                                      0.6361
##                                PaymentMethodElectronic check
##                                                      0.2607
##                                    PaymentMethodMailed check
##                                                      0.2112
## InternetServiceFiber optic:PaymentMethodCredit card (automatic)
##                                                     -0.5829
##          InternetServiceNo:PaymentMethodCredit card (automatic)
##                                                     -1.4639
```

```
##        InternetServiceFiber optic:PaymentMethodElectronic check
##                                                          0.1092
##                 InternetServiceNo:PaymentMethodElectronic check
##                                                         -0.7566
##            InternetServiceFiber optic:PaymentMethodMailed check
##                                                         -0.5876
##                     InternetServiceNo:PaymentMethodMailed check
##                                                         -0.7180
##
## Degrees of Freedom: 4807 Total (i.e. Null);  4784 Residual
## Null Deviance:       6661
## Residual Deviance: 4424  AIC: 4472
```

```
evaluation(best_model_Winte2, train_clean3)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Training Set Metrics:
## AUC: 0.8574414
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1807  666
##    Yes  400 1935
## Recall: 0.8286938
## F1-score: 0.7840357

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.8454629
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1146  438
##    Yes   93  493
## Recall: 0.8412969
## F1-score: 0.649967
```

```
print(paste("Best F-score:", best_f_score2))
```

```
## [1] "Best F-score: 0.784035656401945"
```

```r
Anova(best_model_Winte2)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Churn
##                               LR Chisq Df Pr(>Chisq)
## poly(MonthlyCharges, 2)         15.927  2  0.0003480 ***
## poly(TotalCharges, 2)           17.923  2  0.0001283 ***
## poly(tenure, 3)                 95.360  3  < 2.2e-16 ***
## SeniorCitizen                   16.779  1  4.201e-05 ***
## InternetService                 69.715  1  < 2.2e-16 ***
## StreamingTV                     28.855  1  7.799e-08 ***
## StreamingMovies                 16.171  1  5.787e-05 ***
## Contract                       142.970  2  < 2.2e-16 ***
## PaymentMethod                   13.257  3  0.0041131 **
## InternetService:PaymentMethod   20.813  6  0.0019821 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
step(best_model_Winte2)
```

```
## Start:  AIC=4472.49
## Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure,
##     3) + SeniorCitizen + InternetService + StreamingTV + StreamingMovies +
##     Contract + PaymentMethod * InternetService
##
##                                 Df Deviance    AIC
## <none>                               4424.5 4472.5
## - InternetService:PaymentMethod  6   4445.3 4481.3
## - poly(MonthlyCharges, 2)        2   4440.4 4484.4
## - poly(TotalCharges, 2)          2   4442.4 4486.4
## - StreamingMovies                1   4440.7 4486.7
## - SeniorCitizen                  1   4441.3 4487.3
## - StreamingTV                    1   4453.3 4499.3
## - poly(tenure, 3)                3   4519.9 4561.9
## - Contract                       2   4567.5 4611.5
```

```
##
## Call:  glm(formula = Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges,
##     2) + poly(tenure, 3) + SeniorCitizen + InternetService +
##     StreamingTV + StreamingMovies + Contract + PaymentMethod *
##     InternetService, family = "binomial", data = data)
##
## Coefficients:
##                                                    (Intercept)
##                                                        -1.2993
##                                        poly(MonthlyCharges, 2)1
##                                                       -45.8149
##                                        poly(MonthlyCharges, 2)2
##                                                         4.0706
##                                          poly(TotalCharges, 2)1
##                                                        59.4328
```

```
##                                            poly(TotalCharges, 2)2
##                                                           11.6779
##                                                 poly(tenure, 3)1
##                                                         -115.0405
##                                                 poly(tenure, 3)2
##                                                           -5.5501
##                                                 poly(tenure, 3)3
##                                                          -16.1588
##                                                   SeniorCitizen1
##                                                            0.4043
##                                       InternetServiceFiber optic
##                                                            1.7680
##                                                InternetServiceNo
##                                                           -0.6624
##                                   StreamingTVNo internet service
##                                                                NA
##                                                   StreamingTVYes
##                                                            0.6050
##                               StreamingMoviesNo internet service
##                                                                NA
##                                               StreamingMoviesYes
##                                                            0.4485
##                                                 ContractOne year
##                                                           -0.8888
##                                                 ContractTwo year
##                                                           -2.1547
##                             PaymentMethodCredit card (automatic)
##                                                            0.6361
##                                     PaymentMethodElectronic check
##                                                            0.2607
##                                        PaymentMethodMailed check
##                                                            0.2112
## InternetServiceFiber optic:PaymentMethodCredit card (automatic)
##                                                           -0.5829
##          InternetServiceNo:PaymentMethodCredit card (automatic)
##                                                           -1.4639
##        InternetServiceFiber optic:PaymentMethodElectronic check
##                                                            0.1092
##                 InternetServiceNo:PaymentMethodElectronic check
##                                                           -0.7566
##            InternetServiceFiber optic:PaymentMethodMailed check
##                                                           -0.5876
##                    InternetServiceNo:PaymentMethodMailed check
##                                                           -0.7180
##
## Degrees of Freedom: 4807 Total (i.e. Null);  4784 Residual
## Null Deviance:       6661
## Residual Deviance: 4424  AIC: 4472
```

```r
evaluation(best_model_Winte2, train_clean3)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Training Set Metrics:
## AUC: 0.8574414
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1807  666
##    Yes  400 1935
## Recall: 0.8286938
## F1-score: 0.7840357

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.8454629
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1146  438
##    Yes   93  493
## Recall: 0.8412969
## F1-score: 0.649967
```

```r
f_model <- glm(Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure, 3) + InternetS
evaluation(f_model, train_clean3)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Training Set Metrics:
## AUC: 0.8555873
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1777  696
##    Yes  384 1951
## Recall: 0.835546
## F1-score: 0.7832196

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases

## Test Set Metrics:
## AUC: 0.8451845
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1126  458
##    Yes   88  498
## Recall: 0.8498294
## F1-score: 0.6459144
```

```r
print(f_model)
```

```
##
## Call:  glm(formula = Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges,
##     2) + poly(tenure, 3) + InternetService + StreamingTV + StreamingMovies +
##     Contract + Contract * SeniorCitizen, family = "binomial",
##     data = train_clean2)
##
## Coefficients:
##                         (Intercept)            poly(MonthlyCharges, 2)1
##                             -1.0189                            -41.1922
##            poly(MonthlyCharges, 2)2              poly(TotalCharges, 2)1
##                              3.6265                             45.3095
##              poly(TotalCharges, 2)2                   poly(tenure, 3)1
##                             11.0865                           -101.5759
##                    poly(tenure, 3)2                   poly(tenure, 3)3
##                             -3.5985                            -16.5038
##          InternetServiceFiber optic                  InternetServiceNo
##                              1.6155                             -1.3464
##     StreamingTVNo internet service                      StreamingTVYes
##                                  NA                              0.5705
## StreamingMoviesNo internet service                  StreamingMoviesYes
##                                  NA                              0.4973
##                    ContractOne year                   ContractTwo year
##                             -0.7731                             -2.0024
##                      SeniorCitizen1     ContractOne year:SeniorCitizen1
##                              0.5471                             -0.2474
##    ContractTwo year:SeniorCitizen1
##                             -0.4968
##
## Degrees of Freedom: 4830 Total (i.e. Null);  4814 Residual
## Null Deviance:       6694
## Residual Deviance: 4538  AIC: 4572
```

```r
# We get the same model as before, meaning we cannot get a better model with our method
# Let us try step
step(f_model_cleaned3)
```

```
## Start:  AIC=4460.96
```

```
## Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure,
##     3) + InternetService + StreamingTV + StreamingMovies + Contract +
##     SeniorCitizen * PaymentMethod + Contract * SeniorCitizen
##
##                               Df Deviance    AIC
## <none>                             4415.0 4461.0
## - SeniorCitizen:PaymentMethod  3   4426.5 4466.5
## - poly(MonthlyCharges, 2)      2   4427.9 4469.9
## - StreamingMovies              1   4428.7 4472.7
## - poly(TotalCharges, 2)        2   4433.1 4475.1
## - Contract:SeniorCitizen       2   4434.3 4476.3
## - StreamingTV                  1   4441.8 4485.8
## - InternetService              1   4478.5 4522.5
## - poly(tenure, 3)              3   4517.0 4557.0
```

```
##
## Call:  glm(formula = Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges,
##     2) + poly(tenure, 3) + InternetService + StreamingTV + StreamingMovies +
##     Contract + SeniorCitizen * PaymentMethod + Contract * SeniorCitizen,
##     family = "binomial", data = train_clean3)
##
## Coefficients:
##                                   (Intercept)
##                                      -1.03445
##                       poly(MonthlyCharges, 2)1
##                                     -40.53821
##                       poly(MonthlyCharges, 2)2
##                                       4.46690
##                         poly(TotalCharges, 2)1
##                                      61.76584
##                         poly(TotalCharges, 2)2
##                                      10.08875
##                               poly(tenure, 3)1
##                                    -117.90566
##                               poly(tenure, 3)2
##                                      -4.41398
##                               poly(tenure, 3)3
##                                     -16.29033
##                      InternetServiceFiber optic
##                                       1.50846
##                              InternetServiceNo
##                                      -1.22906
##                   StreamingTVNo internet service
##                                            NA
##                                  StreamingTVYes
##                                       0.58171
##               StreamingMoviesNo internet service
##                                            NA
##                              StreamingMoviesYes
##                                       0.41474
##                               ContractOne year
##                                      -0.75806
##                               ContractTwo year
##                                      -1.94601
```

```
##                                    SeniorCitizen1
##                                           0.08858
##                  PaymentMethodCredit card (automatic)
##                                          -0.02116
##                         PaymentMethodElectronic check
##                                           0.22261
##                             PaymentMethodMailed check
##                                          -0.17435
## SeniorCitizen1:PaymentMethodCredit card (automatic)
##                                           1.06508
##        SeniorCitizen1:PaymentMethodElectronic check
##                                           0.40381
##            SeniorCitizen1:PaymentMethodMailed check
##                                           0.92579
##                       ContractOne year:SeniorCitizen1
##                                          -0.76121
##                       ContractTwo year:SeniorCitizen1
##                                         -14.42896
##
## Degrees of Freedom: 4807 Total (i.e. Null);  4785 Residual
## Null Deviance:       6661
## Residual Deviance: 4415  AIC: 4461
```

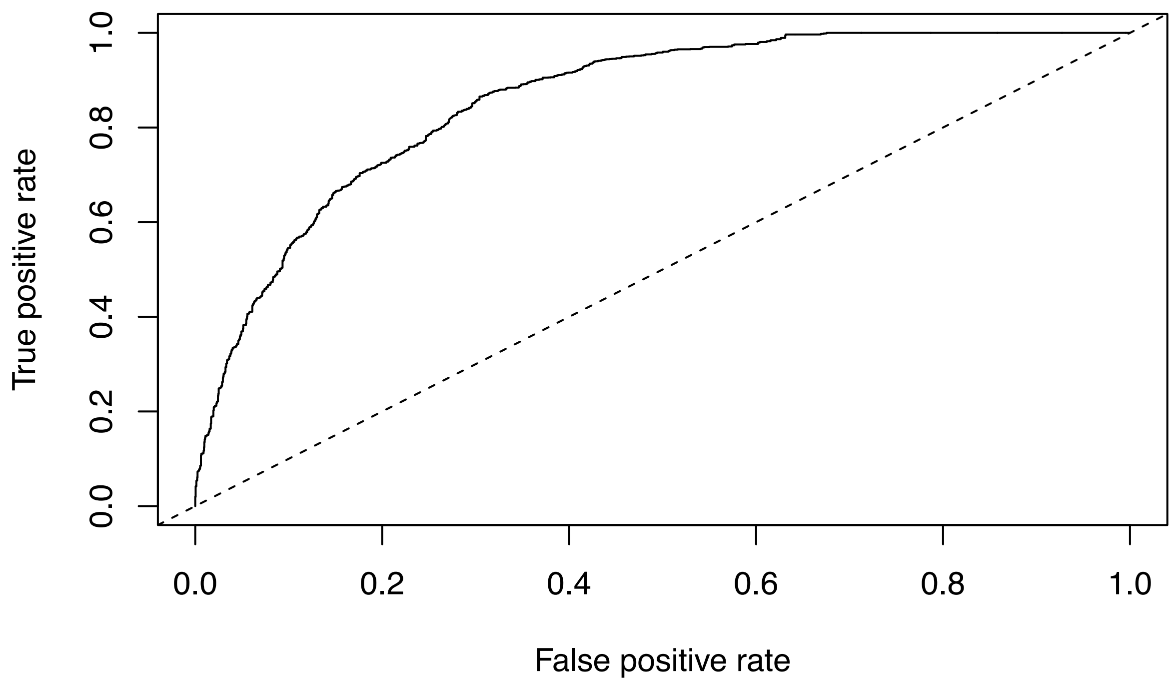
```
# It shows that this is indeed the best model
```

Final analysis Goodness of fit and Model Interpretation. Train and Test datasets.

```
#Assessment metric: area under the ROC curve score and confusion table prediction capability analysi

#Comparing to the previous model using anova
f_model_cleaned2 <- glm(Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure, 3) +
anova(f_model_cleaned2, f_model_cleaned3)
```

```
## Analysis of Deviance Table
##
## Model 1: Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure,
##     3) + InternetService + StreamingTV + StreamingMovies + Contract
## Model 2: Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure,
##     3) + InternetService + StreamingTV + StreamingMovies + Contract +
##     SeniorCitizen * PaymentMethod + Contract * SeniorCitizen
##   Resid. Df Resid. Dev Df Deviance
## 1      4794     4478.2
## 2      4785     4415.0  9   63.233
```

```
# Compare residual deviance, AIC and F1-score with the model without interactions

# ROC curve
library(ROCR)
roc_data<-prediction(predict(f_model_cleaned3, type="response"), train_clean3$Churn)
par(mfrow=c(1,1))
plot(performance(roc_data,"tpr","fpr"))
abline(0,1,lty=2)
```




```r
# Area under ROC curve, Recall, F1-score, Anova
evaluation(f_model_cleaned3, train_clean3)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes

## Setting direction: controls < cases

## Training Set Metrics:
## AUC: 0.8571402
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1796  677
##    Yes  418 1917
## Recall: 0.820985
## F1-score: 0.7778454

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases

## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
```

```
## Test Set Metrics:
## AUC: 0.8467934
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1149  435
##    Yes   95  491
## Recall: 0.837884
## F1-score: 0.6494709
```

```
f_model_cleaned3$formula
```

```
## Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenure,
##     3) + InternetService + StreamingTV + StreamingMovies + Contract +
##     SeniorCitizen * PaymentMethod + Contract * SeniorCitizen
```

```
Anova(f_model_cleaned3)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Churn
##                              LR Chisq Df Pr(>Chisq)
## poly(MonthlyCharges, 2)        12.901  2  0.0015796 **
## poly(TotalCharges, 2)          18.110  2  0.0001168 ***
## poly(tenure, 3)               102.002  3  < 2.2e-16 ***
## InternetService                63.491  1  1.611e-15 ***
## StreamingTV                    26.822  1  2.231e-07 ***
## StreamingMovies                13.724  1  0.0002117 ***
## Contract                      136.125  2  < 2.2e-16 ***
## SeniorCitizen                  16.513  1  4.831e-05 ***
## PaymentMethod                  12.472  3  0.0059287 **
## SeniorCitizen:PaymentMethod    11.494  3  0.0093320 **
## Contract:SeniorCitizen         19.359  2  6.254e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Anova says that StreamingMovies is not significant, but if we take it out all metrics worsen somewh
f_model_cleaned_no_movies <- glm(Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges, 2) + poly(tenu
print(f_model_cleaned_no_movies)
```

```
##
## Call:  glm(formula = Churn ~ poly(MonthlyCharges, 2) + poly(TotalCharges,
##     2) + poly(tenure, 3) + InternetService + StreamingTV + Contract +
##     SeniorCitizen * PaymentMethod + Contract * SeniorCitizen,
##     family = "binomial", data = train_clean3)
##
## Coefficients:
##                                     (Intercept)
##                                        -0.70653
##                         poly(MonthlyCharges, 2)1
##                                       -21.81133
##                         poly(MonthlyCharges, 2)2
```

```
##                                               10.60518
##                               poly(TotalCharges, 2)1
##                                               54.81777
##                               poly(TotalCharges, 2)2
##                                                8.14292
##                                     poly(tenure, 3)1
##                                            -113.21962
##                                     poly(tenure, 3)2
##                                              -2.56972
##                                     poly(tenure, 3)3
##                                             -16.00634
##                           InternetServiceFiber optic
##                                                1.21369
##                                    InternetServiceNo
##                                              -1.23604
##                       StreamingTVNo internet service
##                                                     NA
##                                       StreamingTVYes
##                                                0.55936
##                                     ContractOne year
##                                              -0.74382
##                                     ContractTwo year
##                                              -1.94845
##                                       SeniorCitizen1
##                                                0.09644
##                  PaymentMethodCredit card (automatic)
##                                              -0.01766
##                        PaymentMethodElectronic check
##                                                0.24971
##                            PaymentMethodMailed check
##                                              -0.18303
## SeniorCitizen1:PaymentMethodCredit card (automatic)
##                                                1.12830
##        SeniorCitizen1:PaymentMethodElectronic check
##                                                0.41546
##            SeniorCitizen1:PaymentMethodMailed check
##                                                0.94635
##                     ContractOne year:SeniorCitizen1
##                                              -0.79758
##                     ContractTwo year:SeniorCitizen1
##                                             -14.49383
##
## Degrees of Freedom: 4807 Total (i.e. Null);  4786 Residual
## Null Deviance:       6661
## Residual Deviance: 4429  AIC: 4473
```

```
evaluation(f_model_cleaned_no_movies, train_clean3)
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
```

```
## Training Set Metrics:
## AUC: 0.8563685
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1795  678
##    Yes  437 1898
## Recall: 0.812848
## F1-score: 0.7729587
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
```

```
## Test Set Metrics:
## AUC: 0.8452162
## [1] "Confusion Matrix:"
##       Predicted
## Actual   No  Yes
##    No  1152  432
##    Yes   98  488
## Recall: 0.8327645
## F1-score: 0.6480744
```

```r
Anova(f_model_cleaned_no_movies)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Churn
##                            LR Chisq Df Pr(>Chisq)
## poly(MonthlyCharges, 2)       7.872  2   0.019527 *
## poly(TotalCharges, 2)        14.279  2   0.000793 ***
## poly(tenure, 3)              98.913  3  < 2.2e-16 ***
## InternetService              49.912  1  1.608e-12 ***
## StreamingTV                  24.840  1  6.228e-07 ***
## Contract                    136.301  2  < 2.2e-16 ***
## SeniorCitizen                18.506  1  1.693e-05 ***
## PaymentMethod                15.408  3   0.001499 **
## SeniorCitizen:PaymentMethod  12.881  3   0.004900 **
## Contract:SeniorCitizen       20.662  2  3.261e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#get the coef() and coef(exp())
coef(f_model_cleaned3)
```

```
##                                          (Intercept)
##                                          -1.03444516
##                             poly(MonthlyCharges, 2)1
##                                         -40.53820853
```

```
##                              poly(MonthlyCharges, 2)2
##                                            4.46690344
##                                poly(TotalCharges, 2)1
##                                           61.76584281
##                                poly(TotalCharges, 2)2
##                                           10.08875029
##                                      poly(tenure, 3)1
##                                         -117.90566228
##                                      poly(tenure, 3)2
##                                           -4.41398136
##                                      poly(tenure, 3)3
##                                          -16.29033171
##                            InternetServiceFiber optic
##                                            1.50846066
##                                     InternetServiceNo
##                                           -1.22906404
##                          StreamingTVNo internet service
##                                                    NA
##                                        StreamingTVYes
##                                            0.58170643
##                      StreamingMoviesNo internet service
##                                                    NA
##                                    StreamingMoviesYes
##                                            0.41473813
##                                      ContractOne year
##                                           -0.75806323
##                                      ContractTwo year
##                                           -1.94600764
##                                        SeniorCitizen1
##                                            0.08858450
##                    PaymentMethodCredit card (automatic)
##                                           -0.02115969
##                           PaymentMethodElectronic check
##                                            0.22260509
##                               PaymentMethodMailed check
##                                           -0.17435030
## SeniorCitizen1:PaymentMethodCredit card (automatic)
##                                            1.06507626
##        SeniorCitizen1:PaymentMethodElectronic check
##                                            0.40381404
##            SeniorCitizen1:PaymentMethodMailed check
##                                            0.92578765
##                       ContractOne year:SeniorCitizen1
##                                           -0.76121425
##                       ContractTwo year:SeniorCitizen1
##                                          -14.42895675
```

```
#!!!! Use this one to interpret how each variable predicts whether the customer leaves
exp(coef(f_model_cleaned3))
```

```
##                                           (Intercept)
##                                          3.554235e-01
##                              poly(MonthlyCharges, 2)1
##                                          2.480160e-18
```

```
##                             poly(MonthlyCharges, 2)2 
##                                         8.708664e+01 
##                               poly(TotalCharges, 2)1 
##                                         6.676744e+26 
##                               poly(TotalCharges, 2)2 
##                                         2.407069e+04 
##                                     poly(tenure, 3)1 
##                                         6.226177e-52 
##                                     poly(tenure, 3)2 
##                                         1.210688e-02 
##                                     poly(tenure, 3)3 
##                                         8.417804e-08 
##                           InternetServiceFiber optic 
##                                         4.519768e+00 
##                                    InternetServiceNo 
##                                         2.925663e-01 
##                       StreamingTVNo internet service 
##                                                   NA 
##                                       StreamingTVYes 
##                                         1.789089e+00 
##                   StreamingMoviesNo internet service 
##                                                   NA 
##                                   StreamingMoviesYes 
##                                         1.513974e+00 
##                                     ContractOne year 
##                                         4.685731e-01 
##                                     ContractTwo year 
##                                         1.428432e-01 
##                                       SeniorCitizen1 
##                                         1.092627e+00 
##                 PaymentMethodCredit card (automatic) 
##                                         9.790626e-01 
##                       PaymentMethodElectronic check 
##                                         1.249327e+00 
##                           PaymentMethodMailed check 
##                                         8.400026e-01 
## SeniorCitizen1:PaymentMethodCredit card (automatic) 
##                                         2.901060e+00 
##        SeniorCitizen1:PaymentMethodElectronic check 
##                                         1.497525e+00 
##            SeniorCitizen1:PaymentMethodMailed check 
##                                         2.523855e+00 
##                     ContractOne year:SeniorCitizen1 
##                                         4.670989e-01 
##                     ContractTwo year:SeniorCitizen1 
##                                         5.414816e-07
```

```
#0.5*(1-0.5)*(coef(f_model_cleaned3))
```