STATS101AFinalProject

Qianli Wu

3/3/2022

Contents

First Model - Full Model **2**

Data Inspection 2

Multivariate Regression Model of ALL Predictors 2

Data Cleaning 4

Second Model - Model without Incorrect Variables and Observations **6**

Multivariate Regression Model of ALL Predictors 6

Check Multi-Chollinearity 8

Check Variant IF 9

Third Model - Model without Collinearity Problems **14**

Multivariate Regression Model of independent Predictors 14

Check Variant Inflation Factor 16

Check Multi-Chollinearity 17

Forth Model - Best Sub-Model with Lowest BIC Score **19**

Identify the Best Submodel 19

Select Model with least BIC score 22

Make a Multivariate Regression Model of best Predictors 23

Check Variant Inflation Factor 25

Check Multi-Chollinearity 26

Residual Analysis **28**

Marginal Model Plot - Structure of Model 33

Added-Variable Plot 36

First Model - Full Model

Data Inspection

```r
# import data
lol.data <- read.csv("lol_games.csv")
# remove the column of game id and game duration
lol.data <- lol.data[,-1]
lol.data <- lol.data[,-1]
```

Multivariate Regression Model of ALL Predictors

```r
lol.model.first <- lm(goldDiff~., data = lol.data)
summary(lol.model.first)
```

```
##
## Call:
## lm(formula = goldDiff ~ ., data = lol.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12041.4  -1174.1      5.9   1172.9   8284.3
##
## Coefficients: (4 not defined because of singularities)
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -8.521e+02  4.638e+02  -1.837 0.066189 .
## expDiff                  2.450e-01  5.505e-03  44.508  < 2e-16 ***
## champLevelDiff           1.469e+03  3.794e+01  38.707  < 2e-16 ***
## isFirstTower             1.377e+03  4.853e+01  28.381  < 2e-16 ***
## isFirstBlood             3.136e+02  4.647e+02   0.675 0.499783
## killedFireDrake         -4.336e-01  2.523e+01  -0.017 0.986288
## killedWaterDrake         8.299e+00  2.515e+01   0.330 0.741400
## killedAirDrake           3.560e+01  2.532e+01   1.406 0.159691
## killedEarthDrake        -7.013e+00  2.523e+01  -0.278 0.781065
## killedElderDrake         2.673e+02  7.549e+01   3.542 0.000399 ***
## lostFireDrake           -4.536e+01  2.519e+01  -1.801 0.071747 .
## lostWaterDrake           1.837e+01  2.522e+01   0.728 0.466441
## lostAirDrake            -2.261e+01  2.524e+01  -0.896 0.370359
## lostEarthDrake           3.532e+00  2.510e+01   0.141 0.888115
## lostElderDrake          -2.529e+02  7.461e+01  -3.389 0.000702 ***
## killedBaronNashor        1.359e+02  3.003e+01   4.526 6.05e-06 ***
## lostBaronNashor         -2.533e+02  3.008e+01  -8.421  < 2e-16 ***
## killedRiftHerald         2.736e+02  2.595e+01  10.546  < 2e-16 ***
## lostRiftHerald          -2.556e+02  2.564e+01  -9.967  < 2e-16 ***
## destroyedTopInhibitor    2.576e+02  4.805e+01   5.361 8.37e-08 ***
## destroyedMidInhibitor    4.283e+02  3.558e+01  12.038  < 2e-16 ***
## destroyedBotInhibitor    1.373e+02  2.942e+01   4.667 3.07e-06 ***
## lostTopInhibitor        -2.109e+02  4.887e+01  -4.316 1.59e-05 ***
## lostMidInhibitor        -3.861e+02  3.550e+01 -10.877  < 2e-16 ***
## lostBotInhibitor        -5.222e+01  3.057e+01  -1.708 0.087658 .
```

```
## destroyedTopNexusTurret  3.399e+02  4.822e+01   7.048 1.86e-12 ***
## destroyedBotNexusTurret  3.126e+02  3.546e+01   8.816  < 2e-16 ***
## lostTopNexusTurret      -3.273e+02  4.911e+01  -6.665 2.69e-11 ***
## lostBotNexusTurret      -3.771e+02  3.592e+01 -10.498  < 2e-16 ***
## destroyedTopBaseTurret          NA         NA      NA       NA
## destroyedMidBaseTurret          NA         NA      NA       NA
## destroyedBotBaseTurret   2.735e+02  2.014e+01  13.583  < 2e-16 ***
## lostTopBaseTurret               NA         NA      NA       NA
## lostMidBaseTurret               NA         NA      NA       NA
## lostBotBaseTurret       -2.545e+02  2.019e+01 -12.607  < 2e-16 ***
## destroyedTopInnerTurret  6.902e+02  3.103e+01  22.242  < 2e-16 ***
## destroyedMidInnerTurret  7.386e+02  3.056e+01  24.171  < 2e-16 ***
## destroyedBotInnerTurret  4.628e+02  3.085e+01  14.998  < 2e-16 ***
## lostTopInnerTurret      -7.629e+02  3.175e+01 -24.026  < 2e-16 ***
## lostMidInnerTurret      -7.454e+02  3.071e+01 -24.269  < 2e-16 ***
## lostBotInnerTurret      -4.575e+02  3.124e+01 -14.645  < 2e-16 ***
## destroyedTopOuterTurret  3.845e+02  2.744e+01  14.012  < 2e-16 ***
## destroyedMidOuterTurret  3.987e+02  2.759e+01  14.448  < 2e-16 ***
## destroyedBotOuterTurret  3.523e+02  2.891e+01  12.185  < 2e-16 ***
## lostTopOuterTurret      -6.492e+02  2.613e+01 -24.848  < 2e-16 ***
## lostMidOuterTurret      -6.459e+02  2.616e+01 -24.687  < 2e-16 ***
## lostBotOuterTurret      -5.872e+02  2.792e+01 -21.032  < 2e-16 ***
## kills                    2.230e+02  3.147e+00  70.869  < 2e-16 ***
## deaths                  -2.007e+02  2.489e+00 -80.610  < 2e-16 ***
## assists                 -1.680e+01  1.424e+00 -11.801  < 2e-16 ***
## wardsPlaced             -1.784e-01  1.696e-01  -1.052 0.292831
## wardsDestroyed           1.121e+01  1.619e+00   6.923 4.53e-12 ***
## wardsLost               -1.436e+01  1.628e+00  -8.821  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1842 on 24863 degrees of freedom
## Multiple R-squared:  0.9623, Adjusted R-squared:  0.9622
## F-statistic: 1.323e+04 on 48 and 24863 DF,  p-value: < 2.2e-16
```

Data Cleaning

```r
# Since in summary of lol.model, we have all NA in these four predictors
# Thus, inspect the four columns
cat("Range of destroyedTopBaseTurret: ", range(lol.data$destroyedTopBaseTurret),
    "\nRange of destroyedMidBaseTurret: ", range(lol.data$destroyedMidBaseTurret),
    "\nRange of lostTopBaseTurret: ", range(lol.data$lostTopBaseTurret),
    "\nRange of lostMidBaseTurret: ", range(lol.data$lostMidBaseTurret))
```

```
## Range of destroyedTopBaseTurret:  0 0
## Range of destroyedMidBaseTurret:  0 0
## Range of lostTopBaseTurret:  0 0
## Range of lostMidBaseTurret:  0 0
```

```r
# Since all values in these four columns are 0, we can remove these predictors
```

```r
# import tidyverse for Data Cleaning
library(tidyverse)
# If the Base Turret is not destroyed, then the Inhibitor cannot be destroyed
# Thus, the observations with Top/Mid Inhibitor > 0 but Top/Mid BaseTurret are invalid
# Therefore, we remove these observations and then remove these columns

# Also, BaseTurrent cannot be destroyed twice
# Thus, the observations with BotBaseTurret > 2 are invalid
# We Also need to remove these
lol.data <- lol.data %>%
  # remove the observations with more than 1 Bottom Base Turrent being destroyed
  filter(destroyedBotBaseTurret < 2) %>%
  filter(lostBotBaseTurret < 2) %>%
  # remove the observations with destroyed Top inhibitor but not Top Base Turret
  filter(destroyedTopInhibitor == 0) %>%
  filter(lostTopInhibitor == 0) %>%
  # remove the observations with destroyed Mid inhibitor but not Mid Base Turret
  filter(destroyedMidInhibitor == 0) %>%
  filter(lostMidInhibitor == 0) %>%
  # remove the variables with all 0s
  select(!c(destroyedTopBaseTurret, destroyedMidBaseTurret,
            lostTopBaseTurret, lostMidBaseTurret, lostMidInhibitor,
            destroyedMidInhibitor, lostTopInhibitor, destroyedTopInhibitor))
```

```r
# Preview the dataframe
as.tibble(lol.data)
```

```
## Warning: `as.tibble()` was deprecated in tibble 2.0.0.
## Please use `as_tibble()` instead.
## The signature and semantics have changed, see `?as_tibble`.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
## # A tibble: 11,595 x 45
##    goldDiff expDiff champLevelDiff isFirstTower isFirstBlood killedFireDrake
##       <int>   <int>          <dbl>        <int>        <int>           <int>
##  1     9426   11980            1.8            1            1               2
##  2     9779   11356            2.2            1            1               0
##  3    -3706  -10219           -1.2            1            1               0
##  4    10237   12572            1.6            1            1               1
##  5     9417    9978            1.2            1            1               0
##  6   -13770   -9131           -1.2            0            1               0
##  7     -973   -1302            0              1            1               0
##  8   -11173   -7425           -1              0            1               0
##  9    12287   13710            1.8            1            1               0
## 10   -11571   -7112           -0.8            0            1               0
## # ... with 11,585 more rows, and 39 more variables: killedWaterDrake <int>,
## #   killedAirDrake <int>, killedEarthDrake <int>, killedElderDrake <int>,
## #   lostFireDrake <int>, lostWaterDrake <int>, lostAirDrake <int>,
## #   lostEarthDrake <int>, lostElderDrake <int>, killedBaronNashor <int>,
## #   lostBaronNashor <int>, killedRiftHerald <int>, lostRiftHerald <int>,
## #   destroyedBotInhibitor <int>, lostBotInhibitor <int>,
## #   destroyedTopNexusTurret <int>, destroyedBotNexusTurret <int>, ...
```

Second Model - Model without Incorrect Variables and Observations

Multivariate Regression Model of ALL Predictors

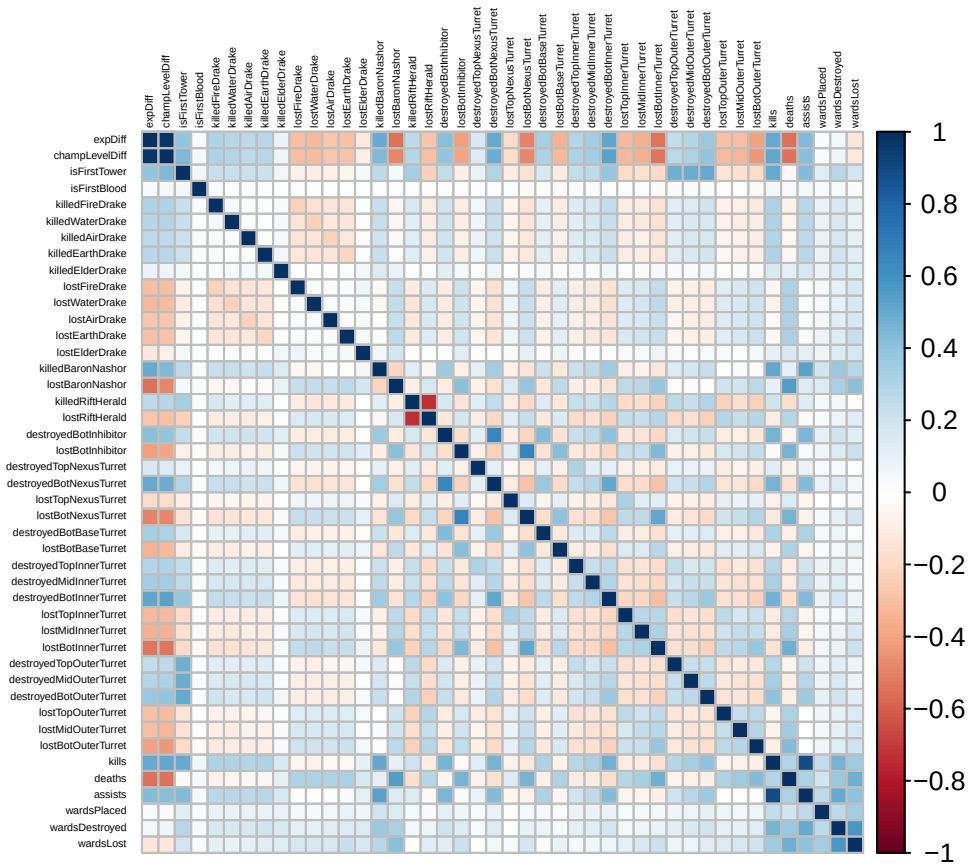
```
lol.model.second <- lm(goldDiff~., data = lol.data)
summary(lol.model.second)
```

```
##
## Call:
## lm(formula = goldDiff ~ ., data = lol.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10001.7  -1115.4     -3.6   1131.8   8437.9
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -8.484e+02  7.262e+02  -1.168 0.242694
## expDiff                  2.330e-01  9.566e-03  24.354  < 2e-16 ***
## champLevelDiff           1.497e+03  6.115e+01  24.477  < 2e-16 ***
## isFirstTower             1.360e+03  6.728e+01  20.213  < 2e-16 ***
## isFirstBlood             2.961e+02  7.269e+02   0.407 0.683736
## killedFireDrake         -9.434e+00  3.665e+01  -0.257 0.796893
## killedWaterDrake         3.900e+01  3.681e+01   1.060 0.289345
## killedAirDrake           6.630e+01  3.705e+01   1.790 0.073536 .
## killedEarthDrake         4.482e+01  3.619e+01   1.238 0.215623
## killedElderDrake         4.006e+02  1.268e+02   3.160 0.001582 **
## lostFireDrake            1.107e+01  3.643e+01   0.304 0.761272
## lostWaterDrake           5.330e+01  3.671e+01   1.452 0.146552
## lostAirDrake             1.513e+01  3.648e+01   0.415 0.678219
## lostEarthDrake           5.407e+01  3.624e+01   1.492 0.135735
## lostElderDrake          -2.591e+02  1.334e+02  -1.942 0.052126 .
## killedBaronNashor        3.634e+01  4.572e+01   0.795 0.426754
## lostBaronNashor         -2.147e+02  4.531e+01  -4.738 2.19e-06 ***
## killedRiftHerald         2.420e+02  3.597e+01   6.726 1.82e-11 ***
## lostRiftHerald          -2.574e+02  3.534e+01  -7.282 3.49e-13 ***
## destroyedBotInhibitor    1.547e+02  4.680e+01   3.304 0.000955 ***
## lostBotInhibitor        -1.056e+02  4.826e+01  -2.189 0.028615 *
## destroyedTopNexusTurret  4.375e+02  8.026e+01   5.452 5.08e-08 ***
## destroyedBotNexusTurret  2.604e+02  5.183e+01   5.023 5.17e-07 ***
## lostTopNexusTurret      -3.940e+02  8.084e+01  -4.873 1.11e-06 ***
## lostBotNexusTurret      -3.665e+02  5.315e+01  -6.896 5.62e-12 ***
## destroyedBotBaseTurret   4.878e+02  5.777e+01   8.444  < 2e-16 ***
## lostBotBaseTurret       -3.331e+02  5.815e+01  -5.729 1.04e-08 ***
## destroyedTopInnerTurret  7.293e+02  4.407e+01  16.550  < 2e-16 ***
## destroyedMidInnerTurret  8.723e+02  4.256e+01  20.496  < 2e-16 ***
## destroyedBotInnerTurret  4.717e+02  4.360e+01  10.821  < 2e-16 ***
## lostTopInnerTurret      -7.562e+02  4.563e+01 -16.574  < 2e-16 ***
## lostMidInnerTurret      -7.717e+02  4.300e+01 -17.945  < 2e-16 ***
## lostBotInnerTurret      -4.303e+02  4.432e+01  -9.708  < 2e-16 ***
## destroyedTopOuterTurret  3.836e+02  3.850e+01   9.963  < 2e-16 ***
```

```
## destroyedMidOuterTurret  4.112e+02  3.882e+01  10.590  < 2e-16 ***
## destroyedBotOuterTurret  4.056e+02  4.074e+01   9.956  < 2e-16 ***
## lostTopOuterTurret      -6.627e+02  3.676e+01 -18.025  < 2e-16 ***
## lostMidOuterTurret      -7.088e+02  3.703e+01 -19.140  < 2e-16 ***
## lostBotOuterTurret      -5.890e+02  3.953e+01 -14.901  < 2e-16 ***
## kills                    2.293e+02  4.621e+00  49.613  < 2e-16 ***
## deaths                  -2.129e+02  3.730e+00 -57.075  < 2e-16 ***
## assists                 -1.440e+01  2.126e+00  -6.775 1.30e-11 ***
## wardsPlaced             -3.277e-01  2.559e-01  -1.280 0.200506
## wardsDestroyed           1.099e+01  2.412e+00   4.555 5.29e-06 ***
## wardsLost               -1.608e+01  2.386e+00  -6.738 1.68e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1765 on 11550 degrees of freedom
## Multiple R-squared:  0.9608, Adjusted R-squared:  0.9606
## F-statistic:  6430 on 44 and 11550 DF,  p-value: < 2.2e-16
```

Check Multi-Chollinearity

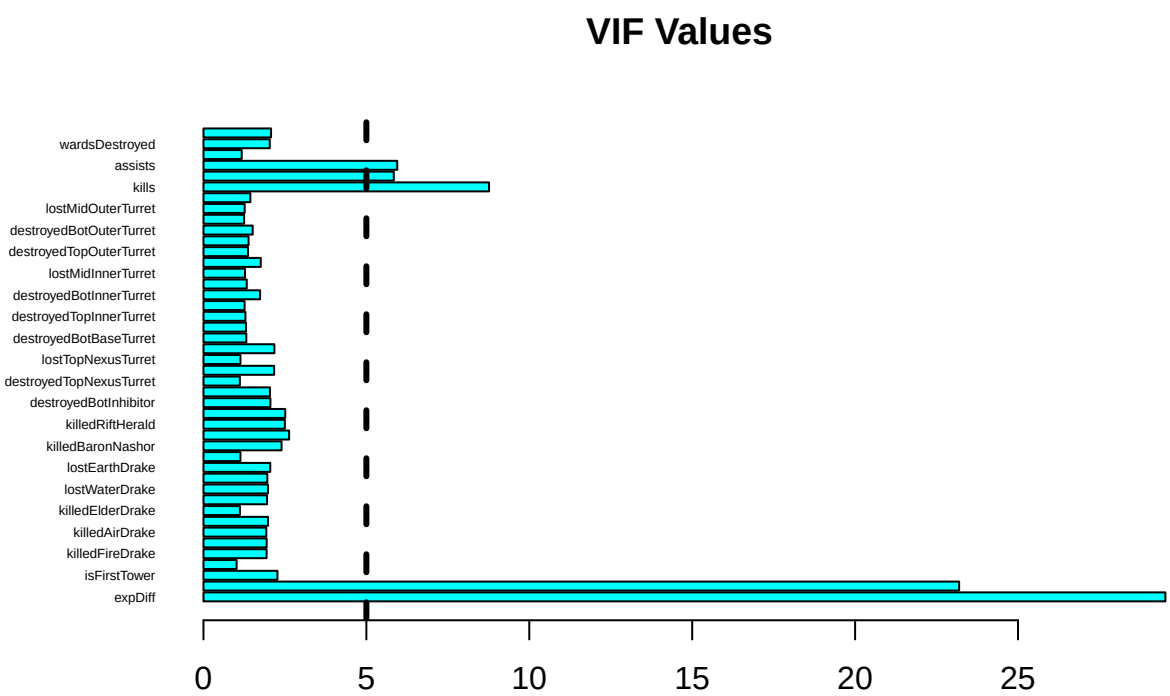
```r
# Check multi-chollinearity
library(corrplot)
# Visualize the aliasing in the model matrix, excluding the intercept.
X <- model.matrix(~.-1, data = lol.data[,-1])

# Create color map on pairwise correlations.
contrast.vectors.correlations <- cor(X)
corrplot(contrast.vectors.correlations, type = "full", addgrid.col = "gray",
         tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.3)
```



Check Variant IF

Predictors with VIF greater than 5 needs to be considered twice.

```
# Using 'vif' function in 'car' package
library(car)
vif_values <- vif(lol.model.second)
# Set the left margin of plot be 1
par(oma=c(0,1,0,0))
#create horizontal bar chart to display each VIF value
barplot(vif_values, main = "VIF Values", horiz = TRUE, col = "cyan", las=1, cex.names=0.4)

#add vertical line at 5
abline(v = 5, lwd = 3, lty = 2)
```



Now we look at Variance IF:
1. we need to remove **expDiff** since it highly correlates with **ChampLevelDiff**
2. we can calculate **KDA** by

$$KDA = \frac{kills + assists}{deaths}$$

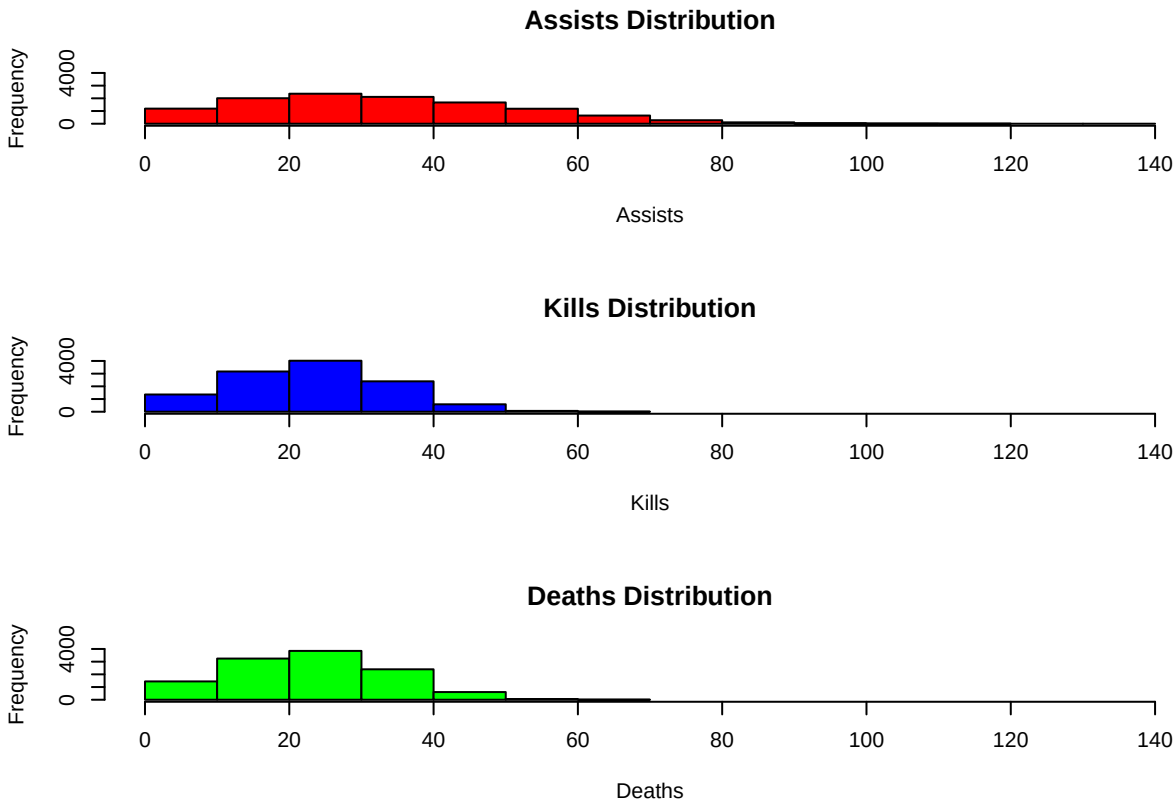
Then we **kills**, **deaths**, **assists**.

```
# remove expDiff
drop <- c("expDiff")
lol.data <- lol.data[,!(names(lol.data) %in% drop)]
as.tibble(lol.data)
```

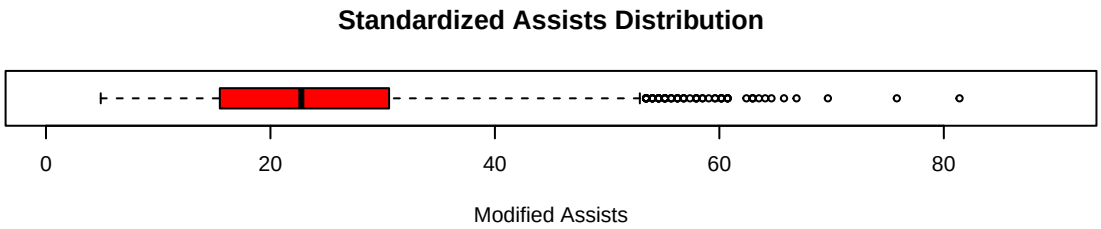
```
## # A tibble: 11,595 x 44
##    goldDiff champLevelDiff isFirstTower isFirstBlood killedFireDrake
##       <int>          <dbl>        <int>        <int>           <int>
```

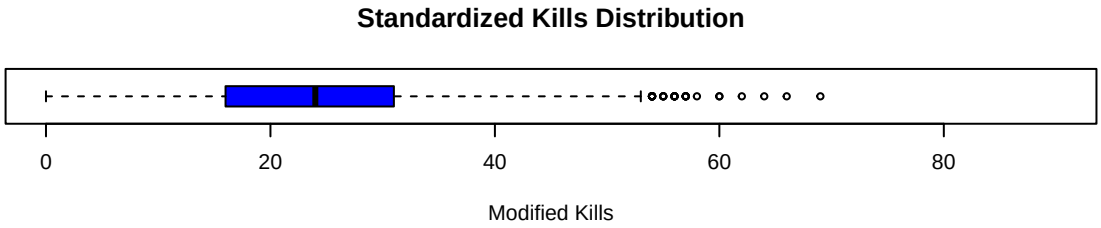
```
## 1     9426            1.8            1            1               2
## 2     9779            2.2            1            1               0
## 3    -3706           -1.2            1            1               0
## 4    10237            1.6            1            1               1
## 5     9417            1.2            1            1               0
## 6   -13770           -1.2            0            1               0
## 7     -973            0              1            1               0
## 8   -11173           -1              0            1               0
## 9    12287            1.8            1            1               0
## 10   -11571           -0.8            0            1               0
## # ... with 11,585 more rows, and 39 more variables: killedWaterDrake <int>,
## #   killedAirDrake <int>, killedEarthDrake <int>, killedElderDrake <int>,
## #   lostFireDrake <int>, lostWaterDrake <int>, lostAirDrake <int>,
## #   lostEarthDrake <int>, lostElderDrake <int>, killedBaronNashor <int>,
## #   lostBaronNashor <int>, killedRiftHerald <int>, lostRiftHerald <int>,
## #   destroyedBotInhibitor <int>, lostBotInhibitor <int>,
## #   destroyedTopNexusTurret <int>, destroyedBotNexusTurret <int>, ...
```

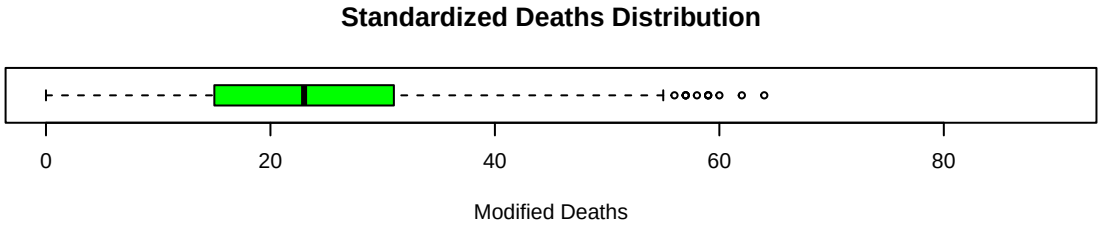
```r
# To avoid those variables with larger ranges dominating over those with small ranges
# We standardize assist and kills by z-score = (value - mean) / standard deviation
par(mfrow=c(3,1))
hist(lol.data$assists, col = "red", breaks = 14,
     xlim = c(0, 140), ylim = c(0, 4000),
     xlab = "Assists", main = "Assists Distribution")
hist(lol.data$kills, col="blue", breaks = 7,
     xlim = c(0, 140),  ylim = c(0, 4000),
     xlab = "Kills", main = "Kills Distribution")
hist(lol.data$deaths, col="green", breaks = 7,
     xlim = c(0, 140),  ylim = c(0, 4000),
     xlab = "Deaths", main = "Deaths Distribution")
```




```r
# We can find after modification, the mean and variance of kills and assists are the same,
# as well as their ranges
par(mfrow=c(3,1))
# lol.data$kills = (lol.data$kills - mean(lol.data$kills)) / sd(lol.data$kills)
lol.data$assists = ((lol.data$assists - mean(lol.data$assists)) / sd(lol.data$assists)
                    * sd(lol.data$kills) + mean(lol.data$kills))
# lol.data$deaths = (lol.data$deaths - mean(lol.data$deaths)) / sd(lol.data$deaths)
boxplot(lol.data$assists, col="red", horizontal = T,
        ylim = c(0, 90),
     xlab = "Modified Assists", main = "Standardized Assists Distribution")
boxplot(lol.data$kills, col="blue", horizontal = T, ylim = c(0, 90),
     xlab = "Modified Kills", main = "Standardized Kills Distribution")
boxplot(lol.data$deaths, col="green", horizontal = T, ylim = c(0, 90),
     xlab = "Modified Deaths", main = "Standardized Deaths Distribution")
```








```r
# Then, we can start to calculate KDA
# lapply(lol.data, function(x){ length(which(x==0))}) # Check the number of zeros in each column
# Set those death = 0 to death = 1, since that's how KDA is calculated
lol.data$deaths[lol.data$deaths == 0] <- 1
lol.data$KDA <- (lol.data$kills + lol.data$assists) / lol.data$deaths
# Remove Kills, Assists, and Deaths
lol.data <- lol.data %>%
  select(!c(kills, assists, deaths))
# Perview the data frame
as.tibble(lol.data)
```

```
## # A tibble: 11,595 x 42
##    goldDiff champLevelDiff isFirstTower isFirstBlood killedFireDrake
##       <int>          <dbl>        <int>        <int>           <int>
##  1     9426            1.8            1            1               2
##  2     9779            2.2            1            1               0
##  3    -3706           -1.2            1            1               0
##  4    10237            1.6            1            1               1
##  5     9417            1.2            1            1               0
##  6   -13770           -1.2            0            1               0
##  7     -973            0              1            1               0
##  8   -11173           -1              0            1               0
##  9    12287            1.8            1            1               0
## 10   -11571           -0.8            0            1               0
## # ... with 11,585 more rows, and 37 more variables: killedWaterDrake <int>,
## #   killedAirDrake <int>, killedEarthDrake <int>, killedElderDrake <int>,
## #   lostFireDrake <int>, lostWaterDrake <int>, lostAirDrake <int>,
## #   lostEarthDrake <int>, lostElderDrake <int>, killedBaronNashor <int>,
```

```
## #   lostBaronNashor <int>, killedRiftHerald <int>, lostRiftHerald <int>,
## #   destroyedBotInhibitor <int>, lostBotInhibitor <int>,
## #   destroyedTopNexusTurret <int>, destroyedBotNexusTurret <int>, ...
```

Third Model - Model without Collinearity Problems

Multivariate Regression Model of independent Predictors

```
lol.model.third <- lm(goldDiff~., data=lol.data)
summary(lol.model.third)
```

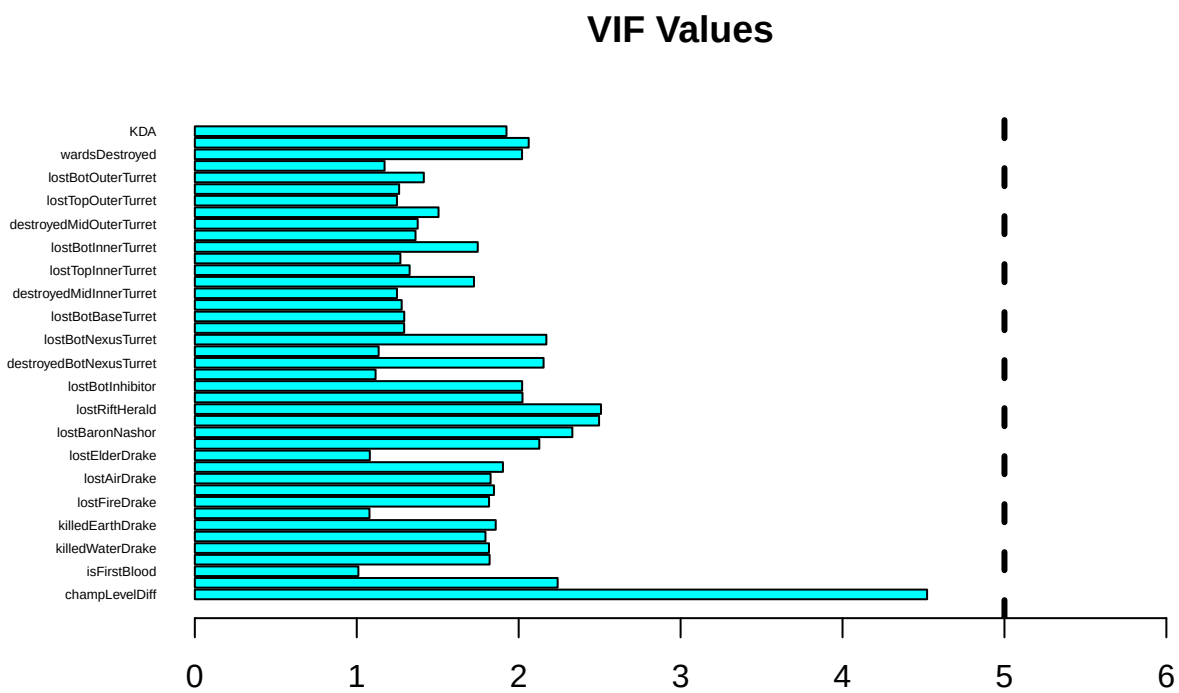
```
##
## Call:
## lm(formula = goldDiff ~ ., data = lol.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13655.1  -1348.4     14.5   1383.8  11119.2
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1807.2257   877.6866   2.059 0.039509 *
## champLevelDiff           4043.8378    32.6843 123.724  < 2e-16 ***
## isFirstTower             1176.6193    80.9105  14.542  < 2e-16 ***
## isFirstBlood            -3383.6611   876.6454  -3.860 0.000114 ***
## killedFireDrake           196.2550    43.0004   4.564 5.07e-06 ***
## killedWaterDrake          246.9319    43.1105   5.728 1.04e-08 ***
## killedAirDrake            268.8132    43.2590   6.214 5.34e-10 ***
## killedEarthDrake          250.1415    42.4114   5.898 3.78e-09 ***
## killedElderDrake          521.7862   150.7069   3.462 0.000538 ***
## lostFireDrake             -50.1654    42.5556  -1.179 0.238494
## lostWaterDrake            -10.2383    42.9087  -0.239 0.811416
## lostAirDrake              -58.6469    42.6204  -1.376 0.168839
## lostEarthDrake            -56.8105    42.2628  -1.344 0.178903
## lostElderDrake           -706.6476   157.5361  -4.486 7.34e-06 ***
## killedBaronNashor          81.6730    52.1368   1.567 0.117256
## lostBaronNashor          -403.8815    51.6465  -7.820 5.74e-15 ***
## killedRiftHerald          254.1866    43.5265   5.840 5.37e-09 ***
## lostRiftHerald           -194.6079    42.7673  -4.550 5.41e-06 ***
## destroyedBotInhibitor     382.7348    56.2154   6.808 1.04e-11 ***
## lostBotInhibitor         -345.1334    58.1355  -5.937 2.99e-09 ***
## destroyedTopNexusTurret   543.8298    97.1026   5.601 2.19e-08 ***
## destroyedBotNexusTurret   526.2495    62.5097   8.419  < 2e-16 ***
## lostTopNexusTurret       -427.3345    97.8445  -4.367 1.27e-05 ***
## lostBotNexusTurret       -534.6657    64.2477  -8.322  < 2e-16 ***
## destroyedBotBaseTurret    629.1211    69.3334   9.074  < 2e-16 ***
## lostBotBaseTurret        -542.2556    70.0204  -7.744 1.04e-14 ***
## destroyedTopInnerTurret   915.2063    53.1241  17.228  < 2e-16 ***
## destroyedMidInnerTurret  1106.1597    51.2210  21.596  < 2e-16 ***
## destroyedBotInnerTurret   679.8774    52.5518  12.937  < 2e-16 ***
## lostTopInnerTurret       -861.0352    55.2024 -15.598  < 2e-16 ***
## lostMidInnerTurret       -999.8983    51.9426 -19.250  < 2e-16 ***
## lostBotInnerTurret       -662.9870    53.4710 -12.399  < 2e-16 ***
## destroyedTopOuterTurret   452.3235    46.5240   9.722  < 2e-16 ***
## destroyedMidOuterTurret   583.9646    46.8636  12.461  < 2e-16 ***
## destroyedBotOuterTurret   527.1938    49.2215  10.711  < 2e-16 ***
```

```
## lostTopOuterTurret       -623.3690    44.5062 -14.006  < 2e-16 ***
## lostMidOuterTurret       -734.1133    44.7568 -16.402  < 2e-16 ***
## lostBotOuterTurret       -692.8396    47.4200 -14.611  < 2e-16 ***
## wardsPlaced                 0.1776     0.3096   0.574 0.566245
## wardsDestroyed             20.8927     2.9099   7.180 7.41e-13 ***
## wardsLost                 -26.1810     2.8794  -9.093  < 2e-16 ***
## KDA                       247.8466    11.2524  22.026  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2136 on 11553 degrees of freedom
## Multiple R-squared:  0.9425, Adjusted R-squared:  0.9423
## F-statistic:  4620 on 41 and 11553 DF,  p-value: < 2.2e-16
```

Check Variant Inflation Factor

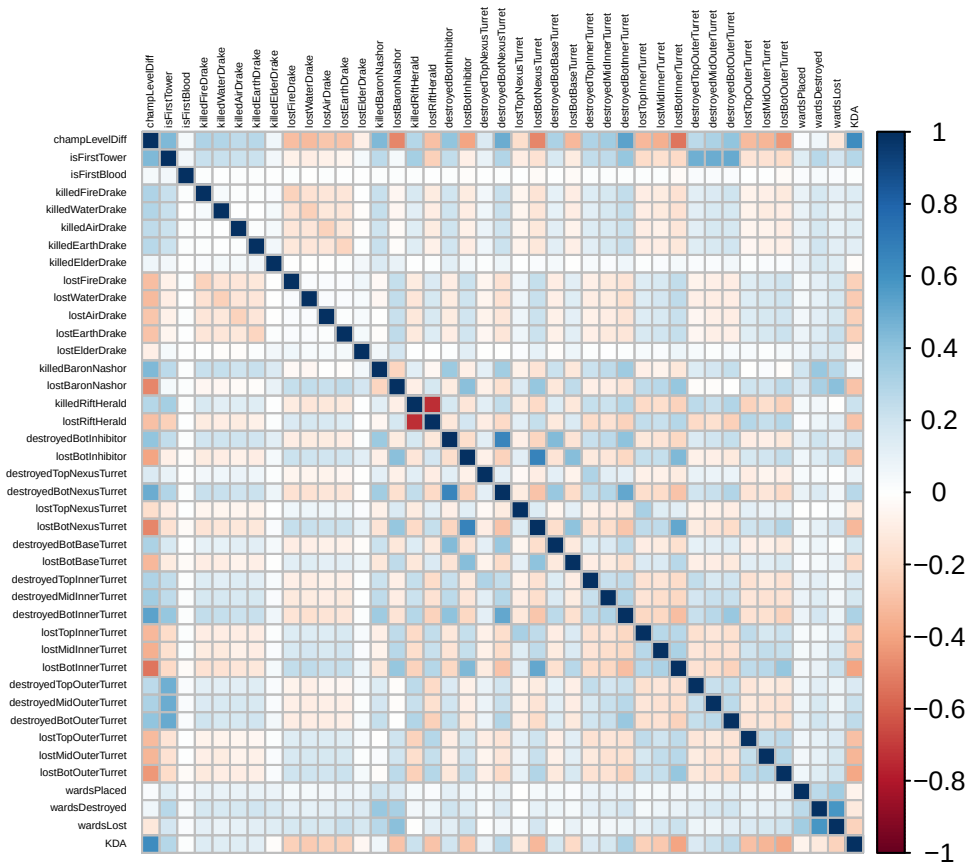
Predictors with VIF greater than 5 needs to be considered twice.

```
# Using 'vif' function in 'car' package
library(car)
vif_values <- vif(lol.model.third)
# Set the left margin of plot be 1
par(oma=c(0,1,0,0))
#create horizontal bar chart to display each VIF value
barplot(vif_values, main = "VIF Values",
        horiz = TRUE, col = "cyan", las=1, cex.names=0.4, xlim = c(0, 6))

#add vertical line at 5
abline(v = 5, lwd = 3, lty = 2)
```



Check Multi-Chollinearity

```r
# Check multi-chollinearity
library(corrplot)
# Visualize the aliasing in the model matrix, excluding the intercept.
X <- model.matrix(~.-1, data = lol.data[,-1])

# Create color map on pairwise correlations.
contrast.vectors.correlations <- cor(X)
corrplot(contrast.vectors.correlations, type = "full", addgrid.col = "gray",
         tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.3)
```




```r
contrast.vectors.correlations[c(15, 16), c(15, 16)]
```

```
##                  lostBaronNashor killedRiftHerald
## lostBaronNashor       1.00000000      -0.08792164
## killedRiftHerald     -0.08792164       1.00000000
```

Unsure Since the absolute value of covariance between killed rift herald and lost rift herald are 0.73836 > 0.6, we can delete one predictor since they're nearly dependent. According to Lecture 8_2, we can only use one "representative" variable from among the set of correlated variables. Thus, we choose **killedRiftHerald** as "representative" variable and remove lostRiftHerald from our dataset.

```r
# There're other correlated predictors such as KDA and champLDiff,
# destroyedBotNexusTurret and destroyedBotInhibitor,
# lostBotNexusTurret and lostBotInhibitor
contrast.vectors.correlations[abs(contrast.vectors.correlations) > 0.6]
```

```
##  [1]  1.0000000  0.6208897  1.0000000  1.0000000  1.0000000  1.0000000
##  [7]  1.0000000  1.0000000  1.0000000  1.0000000  1.0000000  1.0000000
## [13]  1.0000000  1.0000000  1.0000000  1.0000000  1.0000000 -0.7383634
## [19] -0.7383634  1.0000000  1.0000000  0.6563267  1.0000000  0.6645360
## [25]  1.0000000  0.6563267  1.0000000  1.0000000  0.6645360  1.0000000
## [31]  1.0000000  1.0000000  1.0000000  1.0000000  1.0000000  1.0000000
## [37]  1.0000000  1.0000000  1.0000000  1.0000000  1.0000000  1.0000000
## [43]  1.0000000  1.0000000  1.0000000  1.0000000  1.0000000  0.6208897
## [49]  1.0000000
```

Problem: Should we also remove one in each pair of KDA and champLDiff, destroyedBotNexusTurret and destroyedBotInhibitor, lostBotNexusTurret and lostBotInhibitor?

```r
# Remove lostRiftHerald
lol.data <- lol.data %>%
  select(!c(lostRiftHerald))
# Preview the Data Set
as.tibble(lol.data)
```

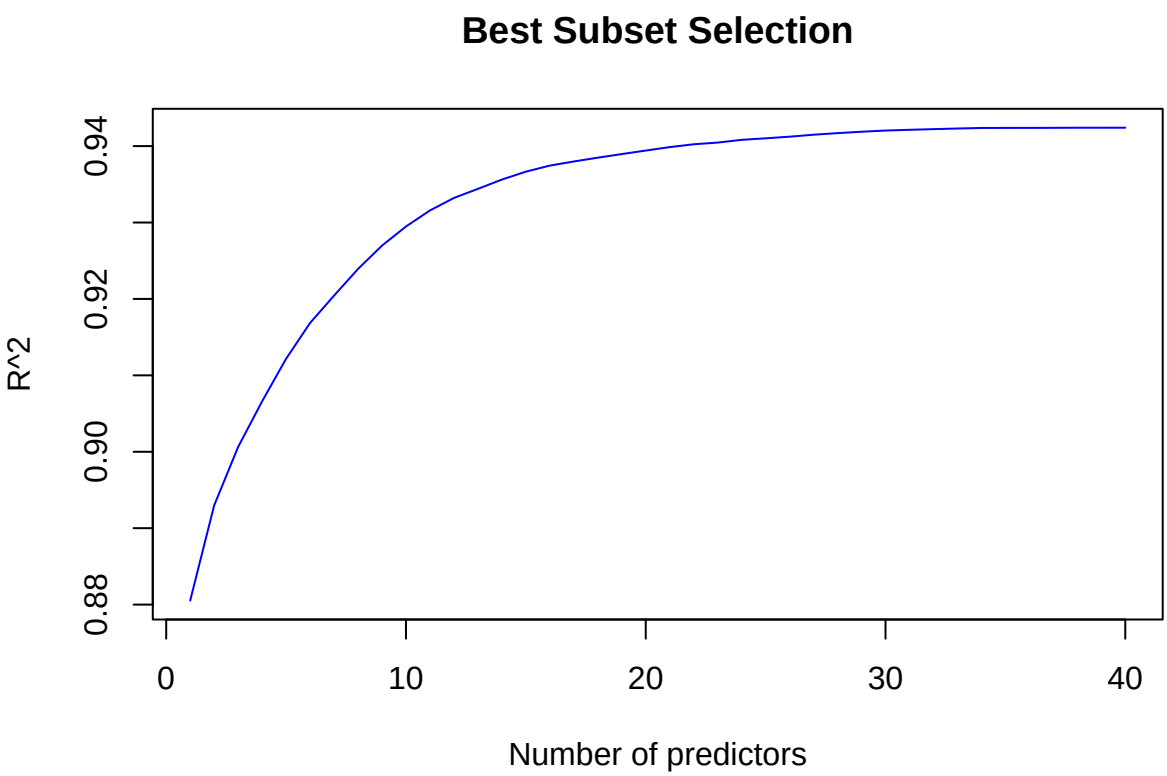
```
## # A tibble: 11,595 x 41
##    goldDiff champLevelDiff isFirstTower isFirstBlood killedFireDrake
##       <int>          <dbl>        <int>        <int>           <int>
##  1     9426            1.8            1            1               2
##  2     9779            2.2            1            1               0
##  3    -3706           -1.2            1            1               0
##  4    10237            1.6            1            1               1
##  5     9417            1.2            1            1               0
##  6   -13770           -1.2            0            1               0
##  7     -973            0              1            1               0
##  8   -11173           -1              0            1               0
##  9    12287            1.8            1            1               0
## 10   -11571           -0.8            0            1               0
## # ... with 11,585 more rows, and 36 more variables: killedWaterDrake <int>,
## #   killedAirDrake <int>, killedEarthDrake <int>, killedElderDrake <int>,
## #   lostFireDrake <int>, lostWaterDrake <int>, lostAirDrake <int>,
## #   lostEarthDrake <int>, lostElderDrake <int>, killedBaronNashor <int>,
## #   lostBaronNashor <int>, killedRiftHerald <int>, destroyedBotInhibitor <int>,
## #   lostBotInhibitor <int>, destroyedTopNexusTurret <int>,
## #   destroyedBotNexusTurret <int>, lostTopNexusTurret <int>, ...
```

Forth Model - Best Sub-Model with Lowest BIC Score

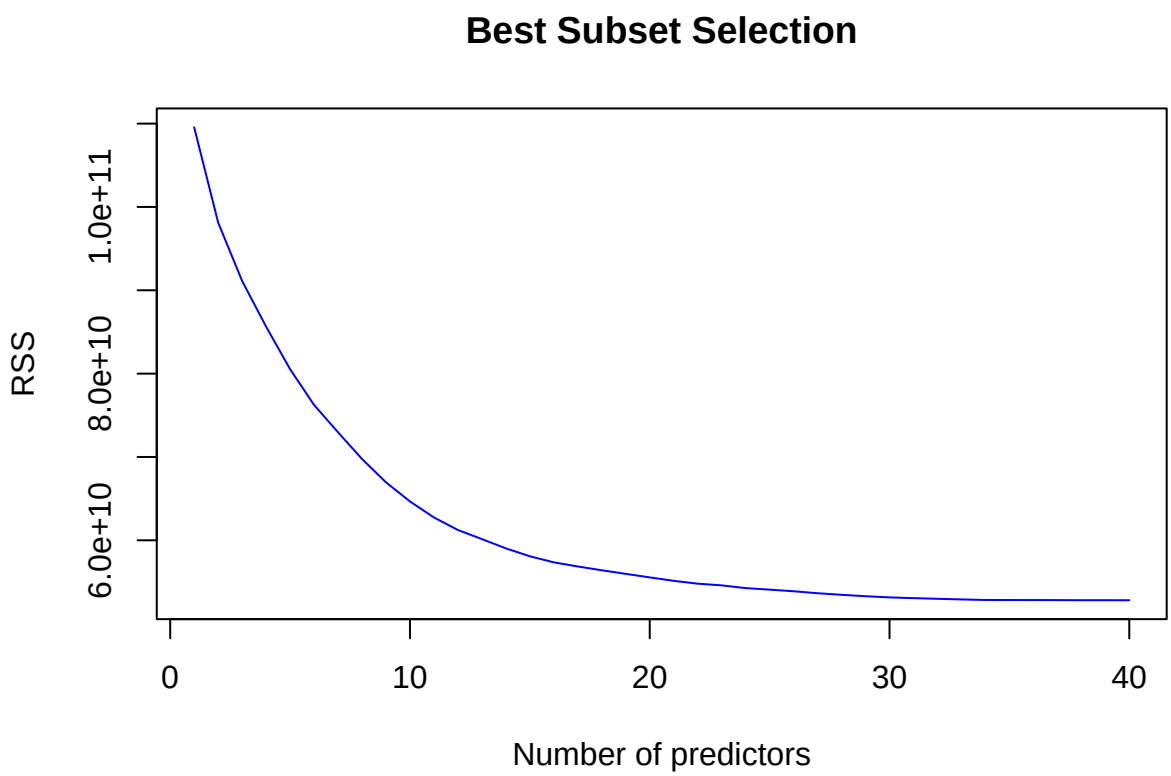
Identify the Best Submodel

```
# eliminate predictors
library(leaps)
Best_Subset <- regsubsets(goldDiff ~., data = lol.data,
                          nbest = 1, nvmax=NULL,
                          force.in = NULL, force.out=NULL,
                          method = "exhaustive",
                          really.big=F)
summary_best_subset <- summary(Best_Subset)
# as.data.frame(summary_best_subset$outmat)
```

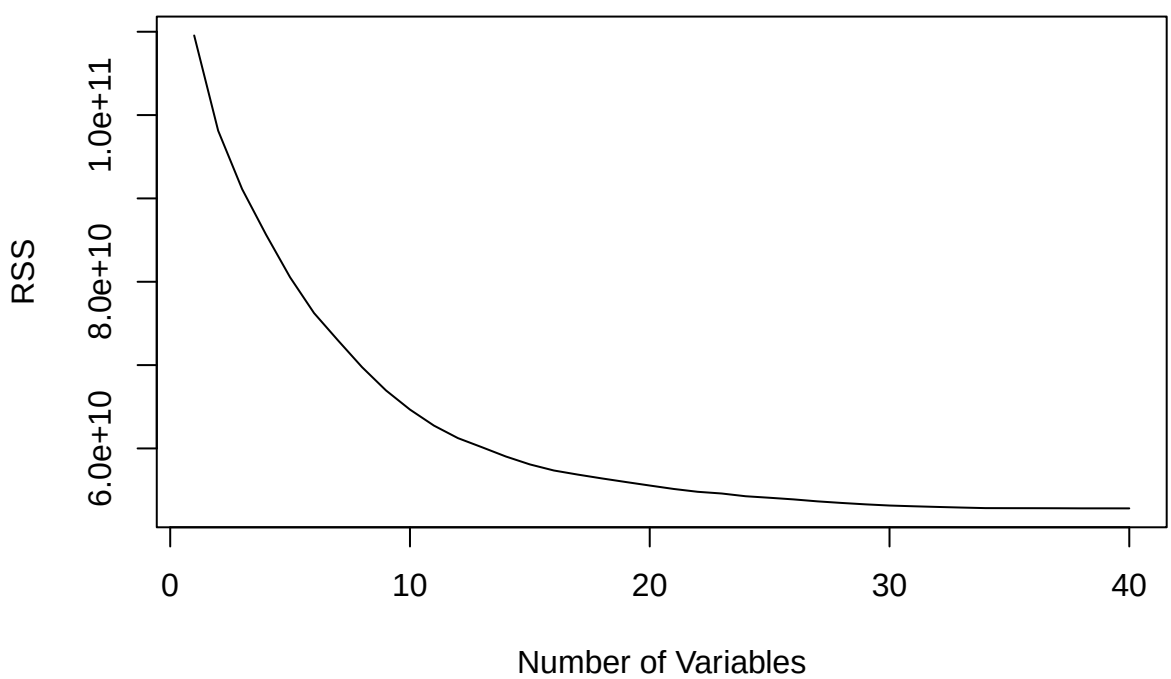
```
# R square will always increase if one predictor is added
plot(summary_best_subset$rsq, xlab = "Number of predictors", ylab = "R^2",
     type = "l", col = "blue", main = 'Best Subset Selection')
```



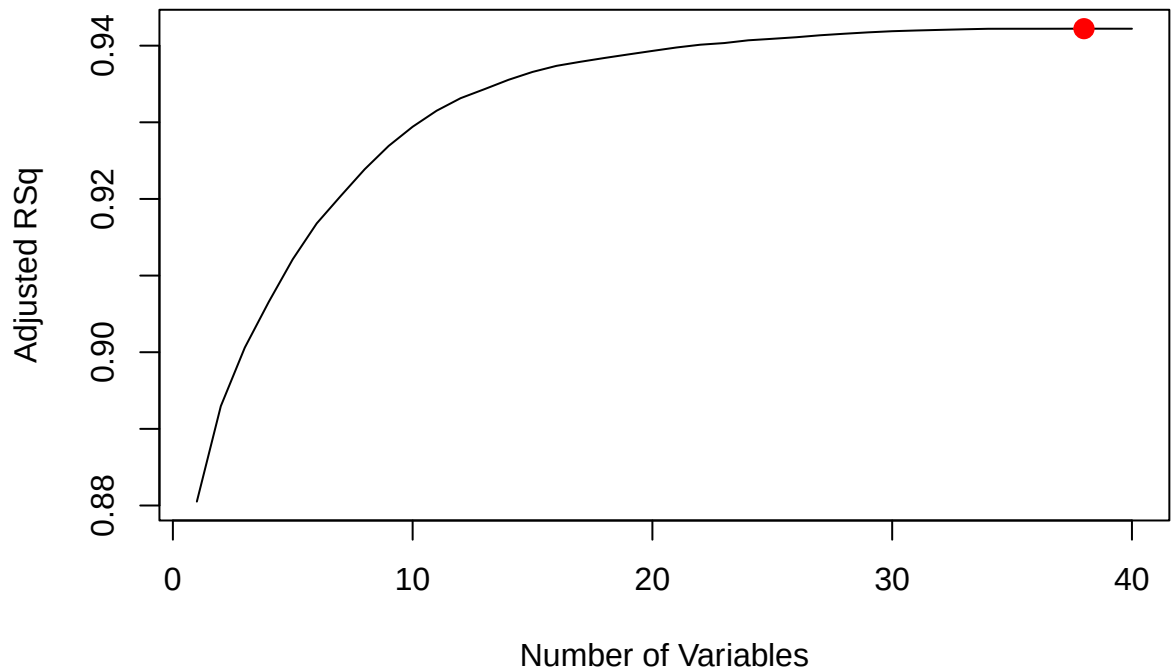

```
plot(summary_best_subset$rss, xlab = "Number of predictors", ylab = "RSS",
     type = "l", col = "blue", main = 'Best Subset Selection')
```



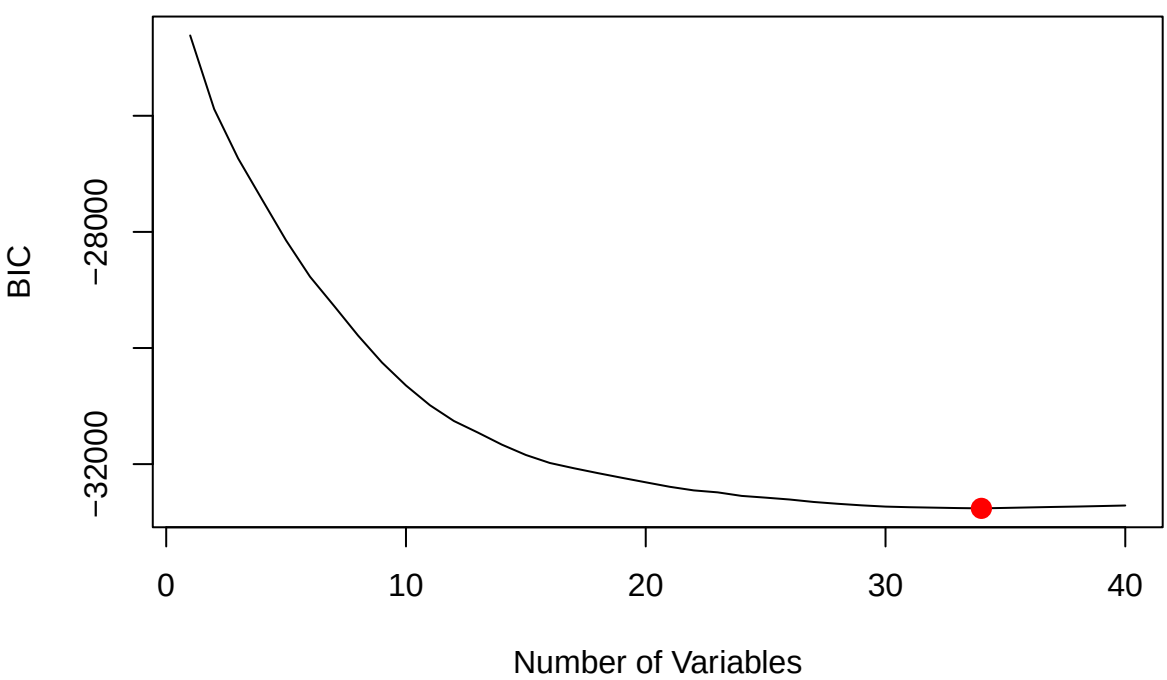

```r
# Plot RSS
plot(summary_best_subset$rss, xlab = "Number of Variables", ylab = "RSS", type = "l")
```




```r
# Plot Adjusted R^2, highlight max value
plot(summary_best_subset$adjr2, xlab = "Number of Variables", ylab = "Adjusted RSq", type = "l")
max = which.max(summary_best_subset$adjr2)
points(max, summary_best_subset$adjr2[max], col = "red", cex = 2, pch = 20)
```




```
# Plot BIC, highlight min value
plot(summary_best_subset$bic, xlab = "Number of Variables", ylab = "BIC", type = "l")
min = which.min(summary_best_subset$bic)
points(min, summary_best_subset$bic[min], col = "red", cex = 2, pch = 20)
```



BIC shows that 34 variables are the best!!!!!

Select Model with least BIC score

```r
# We first select the best model according to BIC.
modelwith.minimum.BIC <- which.min(summary_best_subset$bic)
best.model <- summary_best_subset$which[modelwith.minimum.BIC,][-1]
# Only keep the predictors are indicated by 'TRUE' and our response 'goldDiff'
keep <- names(best.model[best.model==T])
keep <- append(keep, c("goldDiff"))
# Using 'keep' as a mask to select best predictors in data set
lol.data <- lol.data[,(names(lol.data) %in% keep)]
# Preview the selected data set with only best predictors
as.tibble(lol.data)
```

```
## # A tibble: 11,595 x 35
##    goldDiff champLevelDiff isFirstTower isFirstBlood killedFireDrake
##       <int>          <dbl>        <int>        <int>           <int>
##  1     9426            1.8            1            1               2
##  2     9779            2.2            1            1               0
##  3    -3706           -1.2            1            1               0
##  4    10237            1.6            1            1               1
##  5     9417            1.2            1            1               0
##  6   -13770           -1.2            0            1               0
##  7     -973            0              1            1               0
##  8   -11173           -1              0            1               0
##  9    12287            1.8            1            1               0
## 10   -11571           -0.8            0            1               0
## # ... with 11,585 more rows, and 30 more variables: killedWaterDrake <int>,
## #   killedAirDrake <int>, killedEarthDrake <int>, killedElderDrake <int>,
## #   lostElderDrake <int>, lostBaronNashor <int>, killedRiftHerald <int>,
## #   destroyedBotInhibitor <int>, lostBotInhibitor <int>,
## #   destroyedTopNexusTurret <int>, destroyedBotNexusTurret <int>,
## #   lostTopNexusTurret <int>, lostBotNexusTurret <int>,
## #   destroyedBotBaseTurret <int>, lostBotBaseTurret <int>, ...
```

Make a Multivariate Regression Model of best Predictors

```r
lol.model.forth <- lm(goldDiff~., data=lol.data)
# T-test or Partial F-test
summary(lol.model.forth)
```

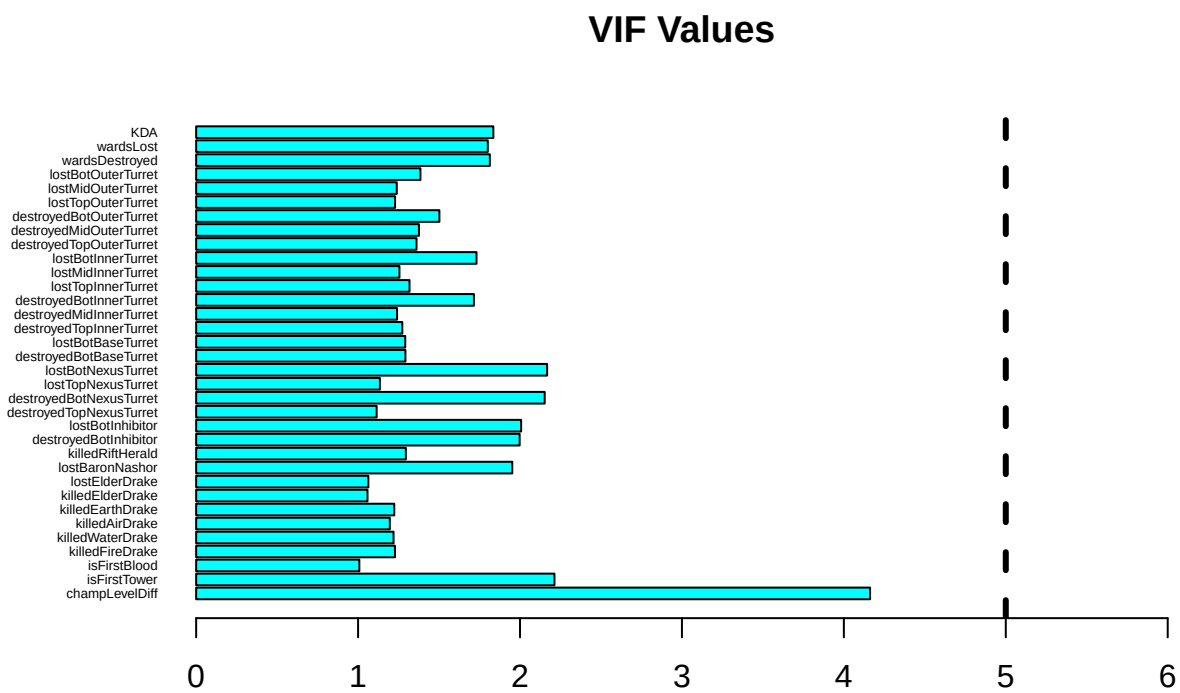
```
##
## Call:
## lm(formula = goldDiff ~ ., data = lol.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13732.3  -1353.0     15.3   1381.9  10937.3
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1695.173    877.950   1.931 0.053528 .
## champLevelDiff           4053.567     31.384 129.162  < 2e-16 ***
## isFirstTower             1157.007     80.481  14.376  < 2e-16 ***
## isFirstBlood            -3546.504    876.412  -4.047 5.23e-05 ***
## killedFireDrake           228.014     35.343   6.452 1.15e-10 ***
## killedWaterDrake          279.743     35.347   7.914 2.72e-15 ***
## killedAirDrake            307.186     35.342   8.692  < 2e-16 ***
## killedEarthDrake          282.802     34.442   8.211 2.43e-16 ***
## killedElderDrake          562.987    149.404   3.768 0.000165 ***
## lostElderDrake           -656.906    156.411  -4.200 2.69e-05 ***
## lostBaronNashor          -442.994     47.300  -9.366  < 2e-16 ***
## killedRiftHerald          391.966     31.375  12.493  < 2e-16 ***
## destroyedBotInhibitor     388.289     55.904   6.946 3.97e-12 ***
## lostBotInhibitor         -344.340     57.980  -5.939 2.95e-09 ***
## destroyedTopNexusTurret   540.034     97.120   5.561 2.75e-08 ***
## destroyedBotNexusTurret   523.226     62.549   8.365  < 2e-16 ***
## lostTopNexusTurret       -431.785     97.909  -4.410 1.04e-05 ***
## lostBotNexusTurret       -544.031     64.254  -8.467  < 2e-16 ***
## destroyedBotBaseTurret    625.708     69.381   9.018  < 2e-16 ***
## lostBotBaseTurret        -554.073     70.027  -7.912 2.76e-15 ***
## destroyedTopInnerTurret   920.396     53.079  17.340  < 2e-16 ***
## destroyedMidInnerTurret  1103.373     51.108  21.589  < 2e-16 ***
## destroyedBotInnerTurret   679.652     52.469  12.953  < 2e-16 ***
## lostTopInnerTurret       -876.853     55.066 -15.924  < 2e-16 ***
## lostMidInnerTurret      -1014.826     51.700 -19.629  < 2e-16 ***
## lostBotInnerTurret       -672.725     53.276 -12.627  < 2e-16 ***
## destroyedTopOuterTurret   448.192     46.530   9.632  < 2e-16 ***
## destroyedMidOuterTurret   581.536     46.898  12.400  < 2e-16 ***
## destroyedBotOuterTurret   532.642     49.217  10.822  < 2e-16 ***
## lostTopOuterTurret       -641.507     44.193 -14.516  < 2e-16 ***
## lostMidOuterTurret       -745.565     44.393 -16.794  < 2e-16 ***
## lostBotOuterTurret       -702.474     46.971 -14.955  < 2e-16 ***
## wardsDestroyed             19.876      2.759   7.204 6.23e-13 ***
## wardsLost                 -26.819      2.694  -9.957  < 2e-16 ***
## KDA                       251.228     10.998  22.844  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2138 on 11560 degrees of freedom
## Multiple R-squared:  0.9424, Adjusted R-squared:  0.9422
## F-statistic:  5560 on 34 and 11560 DF,  p-value: < 2.2e-16
```

Check Variant Inflation Factor

Predictors with VIF greater than 5 needs to be considered twice.

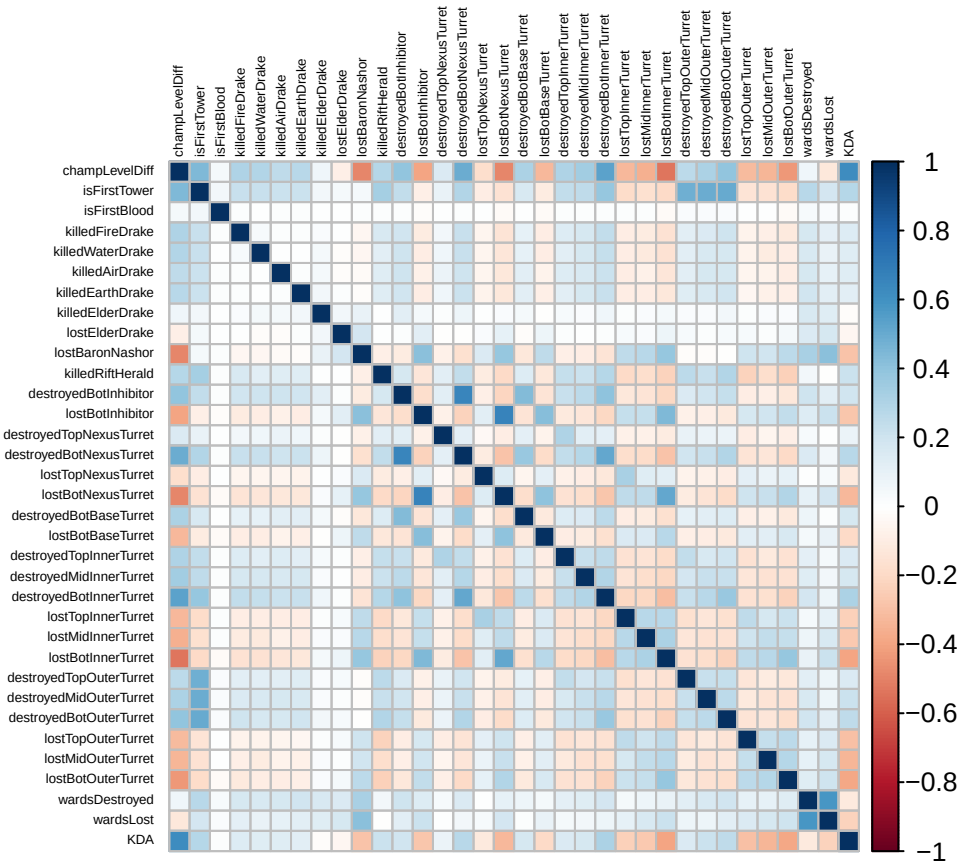
```
# Using 'vif' function in 'car' package
library(car)
vif_values <- vif(lol.model.forth)
# Set the left margin of plot be 1
par(oma=c(0,1,0,0))
#create horizontal bar chart to display each VIF value
barplot(vif_values, main = "VIF Values",
        horiz = TRUE, col = "cyan", las=1, cex.names=0.4, xlim = c(0, 6))

#add vertical line at 5
abline(v = 5, lwd = 3, lty = 2)
```



Since no VIF is greater than 5, there're **no serious** collinearity problem with predictors.

Check Multi-Chollinearity

```
# Visualize the aliasing in the model matrix, excluding the intercept.
X <- model.matrix(~.-1, data = lol.data[,-1])

# Create color map on pairwise correlations.
contrast.vectors.correlations <- cor(X)
corrplot(contrast.vectors.correlations, type = "full", addgrid.col = "gray",
         tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.4)
```



From the multi collinearity color graph above, we have grids other than the diagonal do not contain deeper color, which means in general multicollinearity is not problematic

---

- Why we drop **expDiff** instead of **ChampLevelDiff**?

- A:
- Why we can drop predictors with high p-value?

- A: As mentioned by lecture 7-1

- Why observations with **Death** == 0 can be dropped?

- A: In Calculating KDA, death = 0 is same with death = 1
- Why **Assist** is a good predictor? It's Estimated Coefficient is negative

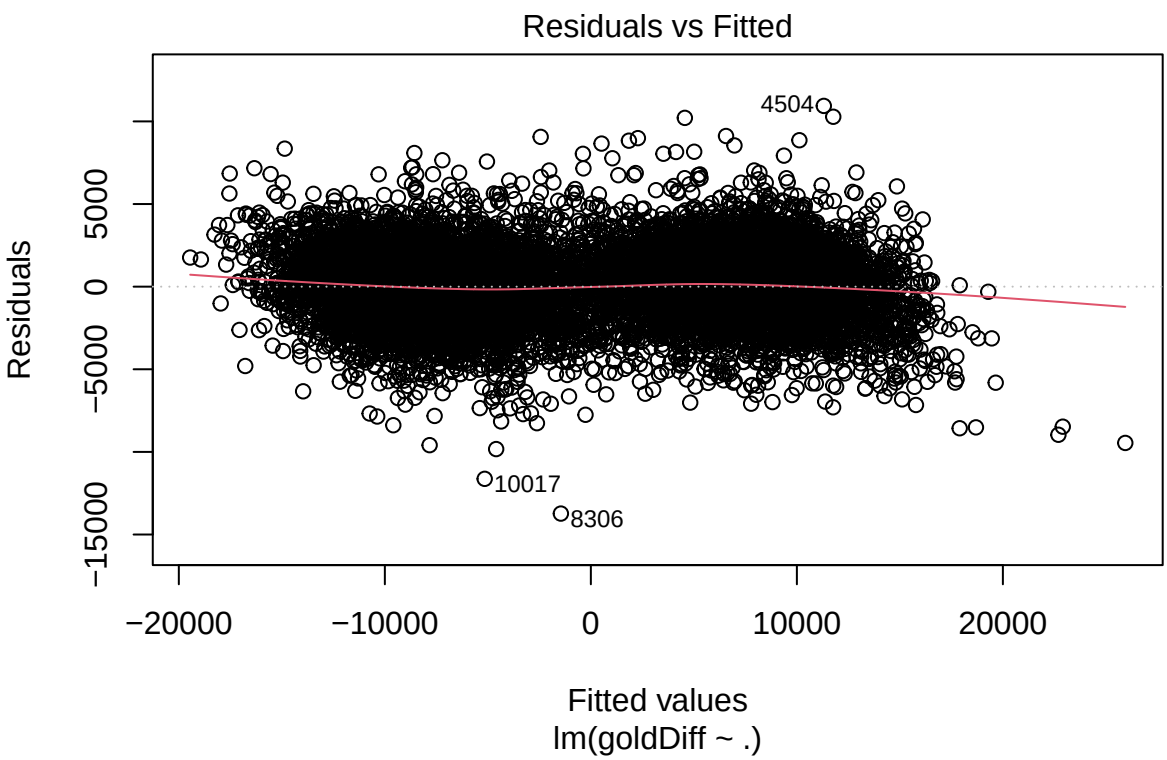
- A: it doesn't matter
- Why the weight of **Assist** == the weight of **Kill**
- A: some scarifies in accuracy is necessary to simplify the model. We adjust the range of it so it won't dominate with Kills.

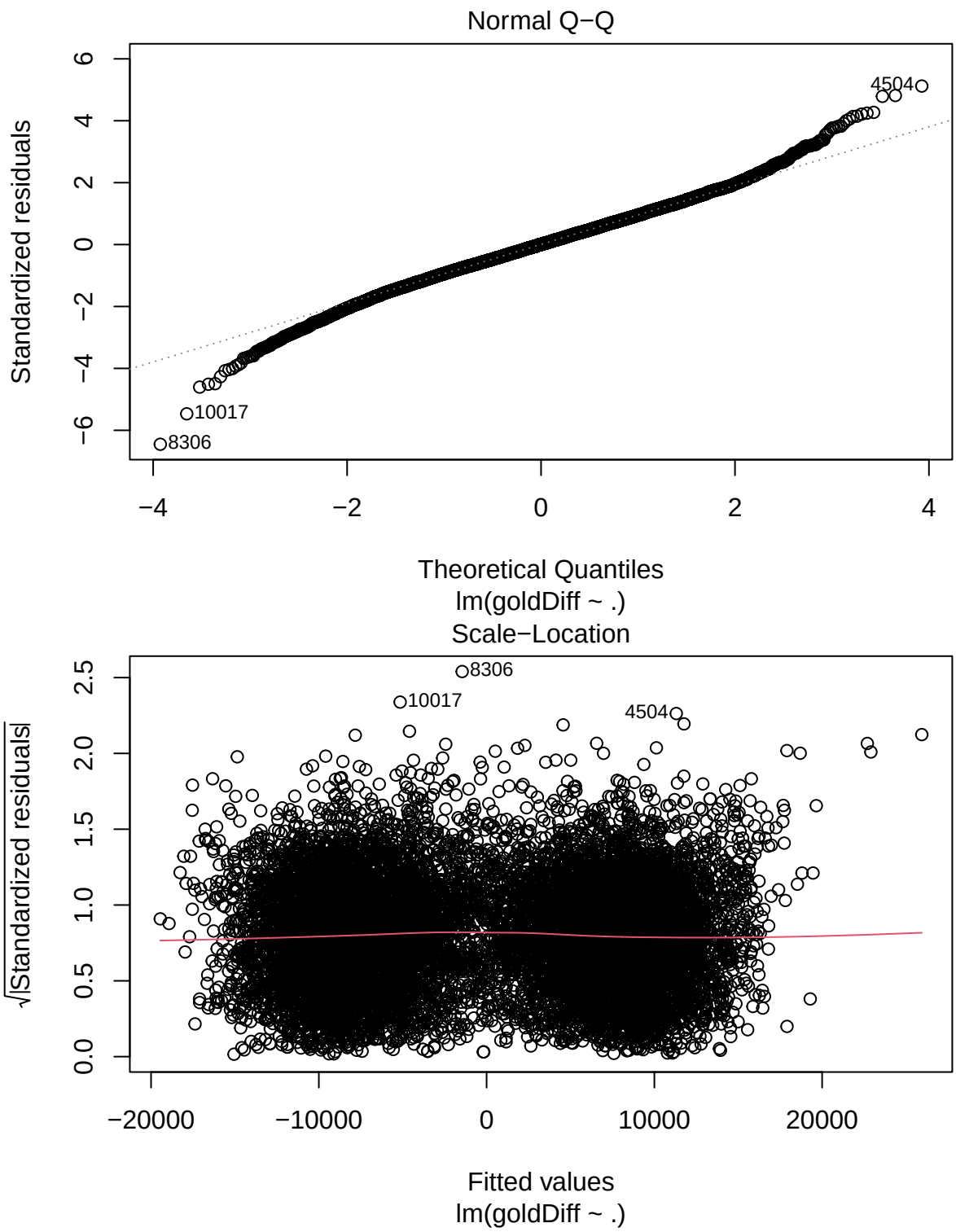
Residual Analysis

Calculate Pesky Points: High leverage Points and Outliers

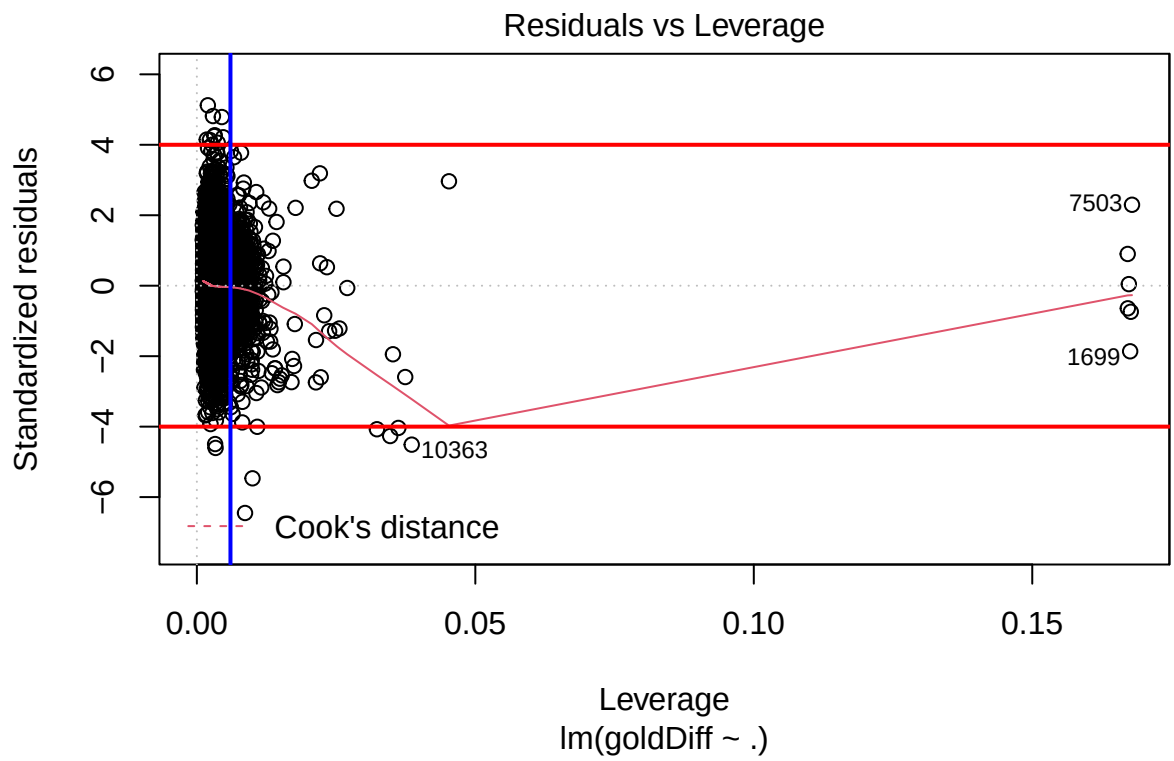
```
# number of coefficient in our model, including the intercept
# The sum of diagonal of leverage matrix
p <- length(lol.data)
n <- length(lol.data$goldDiff) # Number of observations in our model
# a high leverage point is the point with more than twice of the average leverage score
high_leverage <- 2*p/n
```

```
# Display four plot at a time
# par(mfrow=c(2,2 ))
plot(lol.model.forth)
```






```
# Since our dataset can be considered as 'large', according to Campuswire post
# High leverage points with standardized residual outside of (-4, 4) are considered bad
abline(a =4, b=0, col = "red", lwd = 2)
abline(a =-4, b=0, col = "red", lwd = 2)
abline(v = high_leverage, col = "blue", lwd = 2)
```



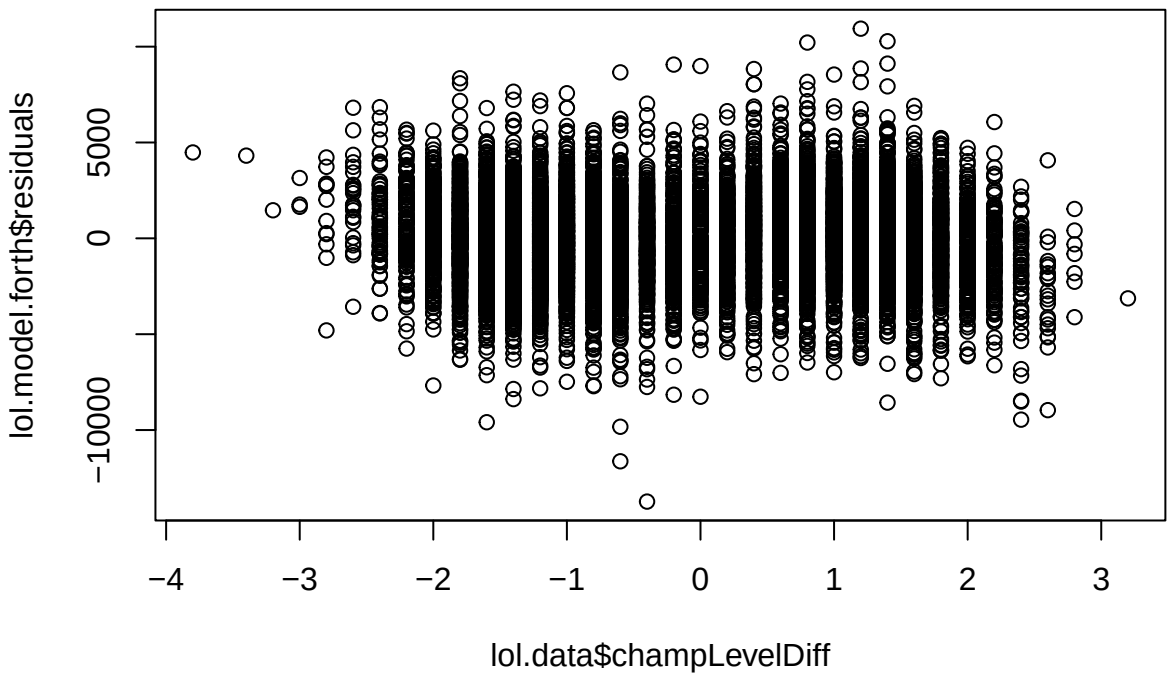
Constant variance of the residuals: Since we have high leverage points, we should use Scale-Location plot to check the constant of variance assumption. Since the residuals are scattered uniformly and there's no evident trend shown in the Scale-Location plot, we can say that the assumption of constant of variance is satisfied.

Normality: Since the sample size is large, according to Central Limit Theorem, we can still get approximately unbiased estimates of the coefficients.

Residuals are independent: Generally its hard to check. Usually there're two classes of dependencies:
1. Residuals correlate with another variable (reponse and predictors)
2. Residuals correlate with other (close) residuals (autocorrelation)

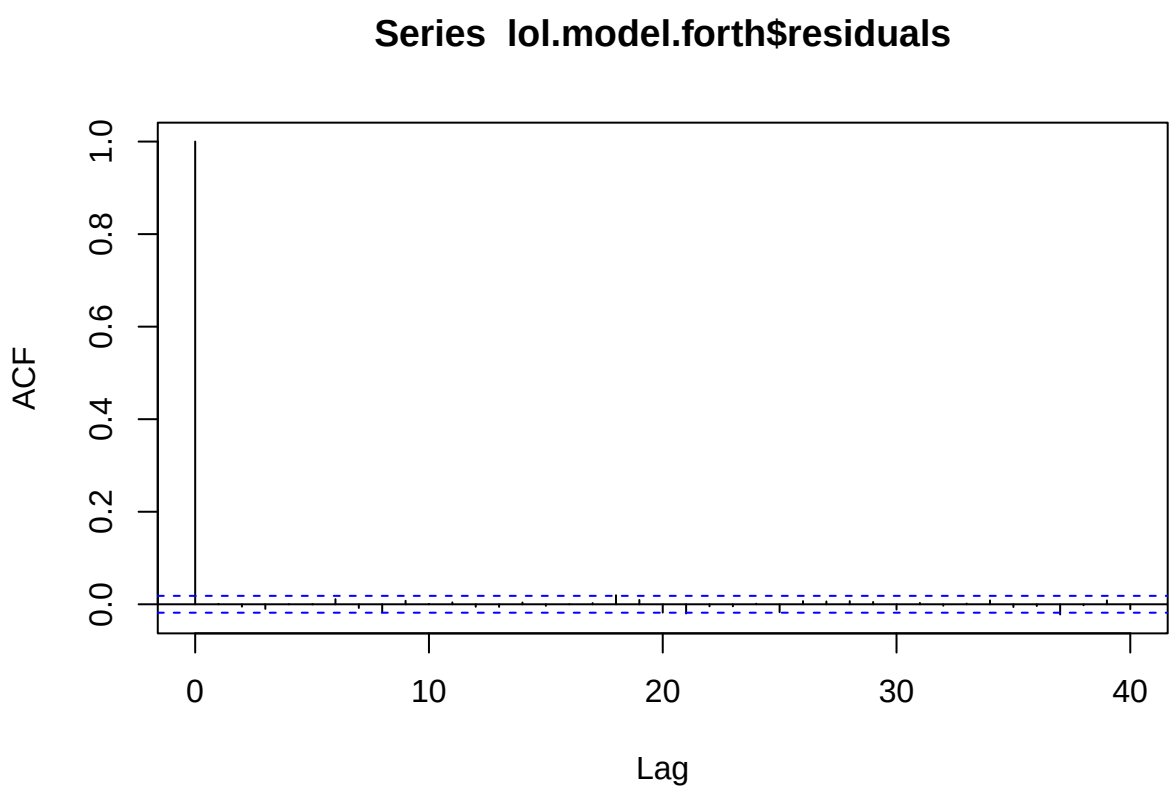
1 Residuals correlate with another variable (reponse and predictors) From the Residuals vs Fitted plot, since the dots are randomly scattered, we get the residuals are independent of y.
For correlation with other variables, we choose the first predictor to represent (I don't want to show too many plots)

```
plot(lol.data$champLevelDiff, lol.model.forth$residuals)
```




```
acf(lol.model.forth$residuals, type = "correlation")
```

2 autocorrelation

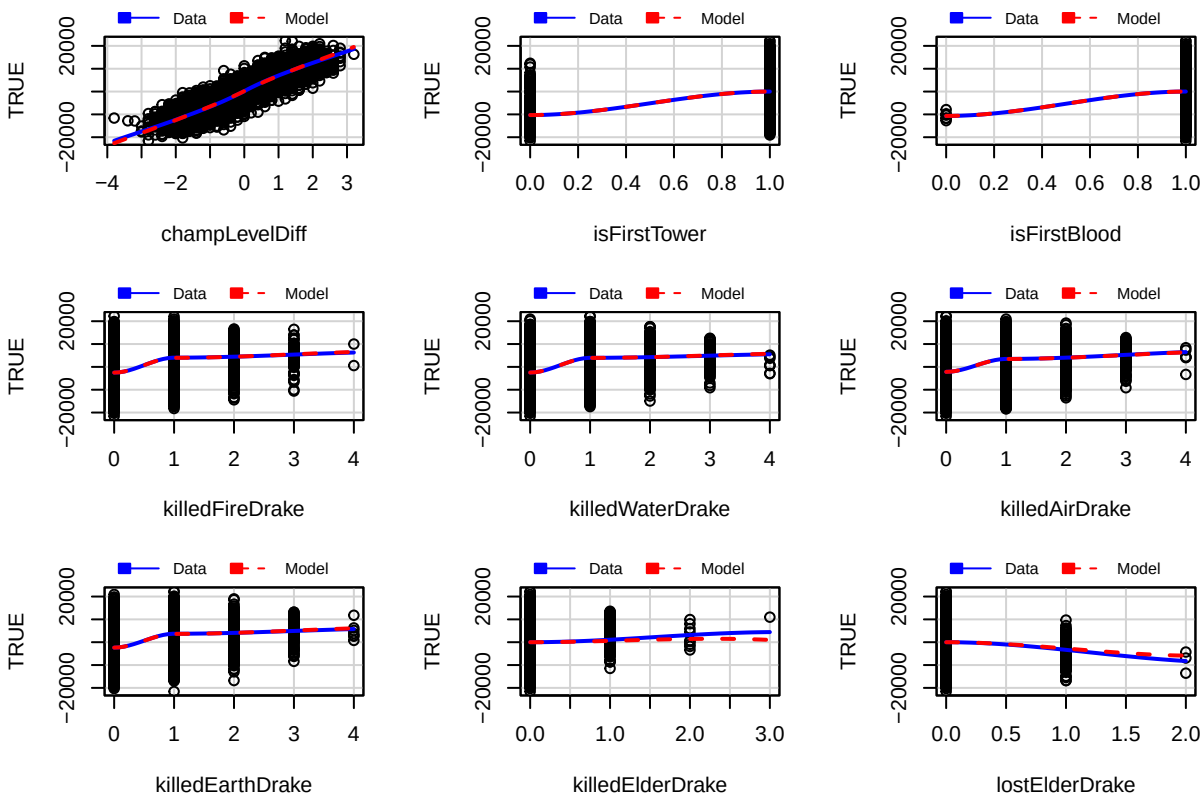


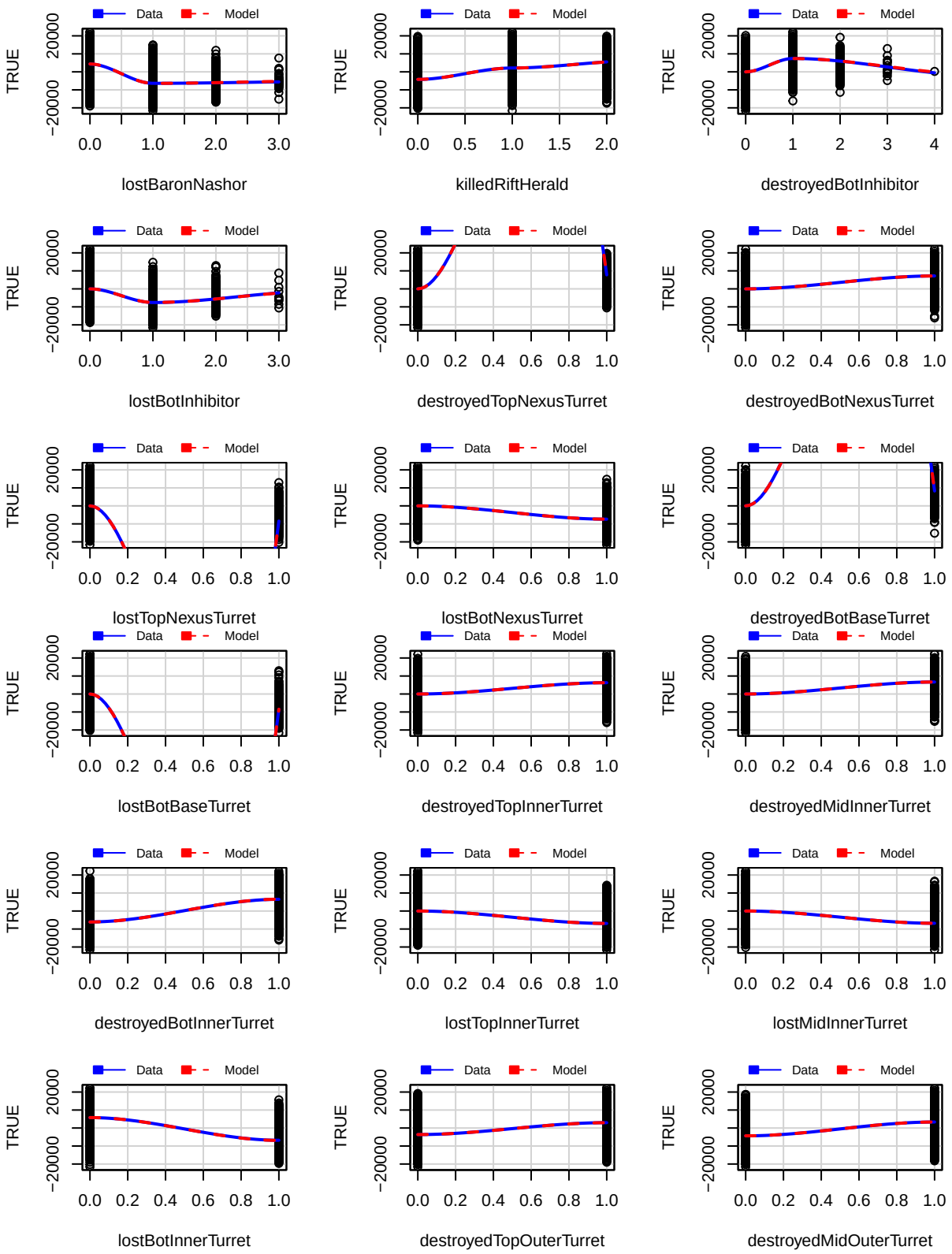
From the acf plot above, since the lags are almost 0 except for itself, we get the residuals are not autocorrelated.

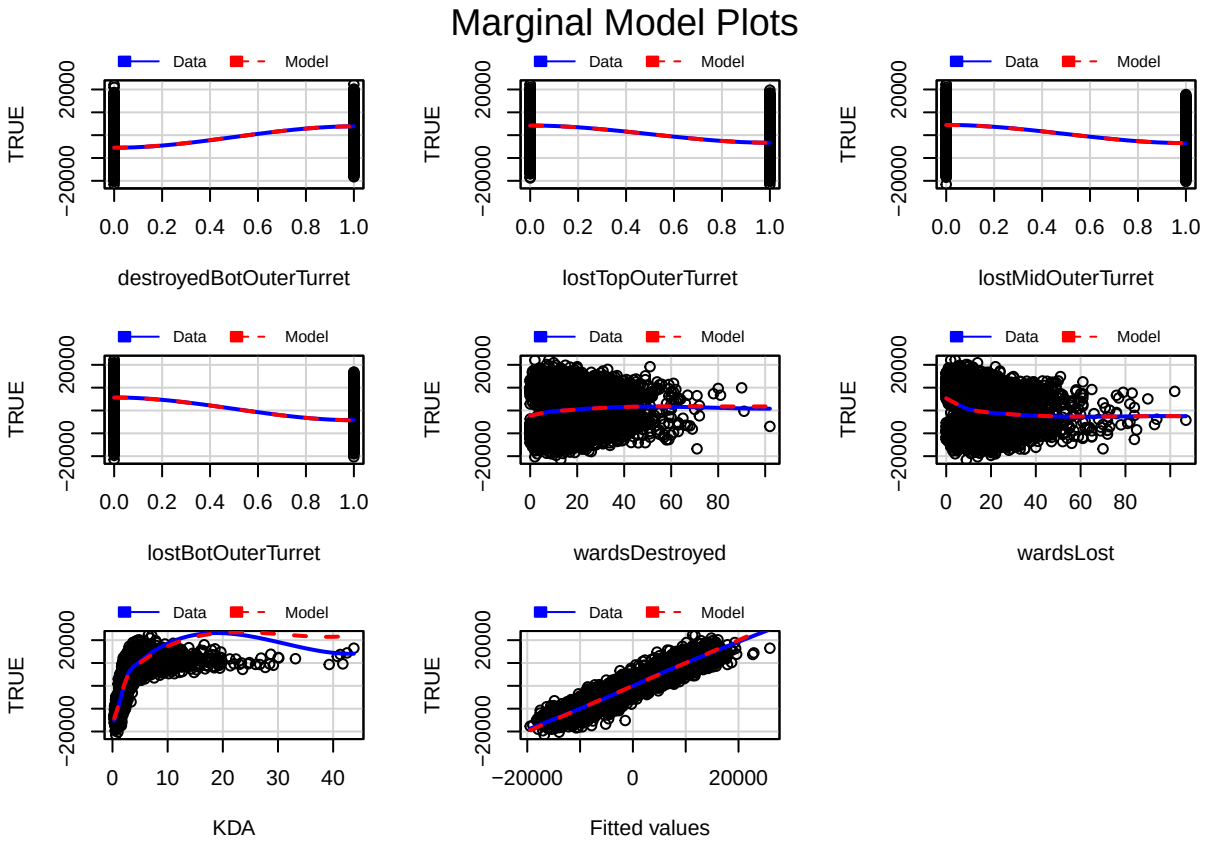
Marginal Model Plot - Structure of Model

Marginal plots allow us to check if the structure of the model captures the relationship between the response and a single predictor. When the two functions are similar in each of the graphs, there is evidence that the model fits well.
When the two functions differ in at least one of the graphs, there is evidence that the model does not fit well.

```
mmps(lol.model.forth, ask = F)
```



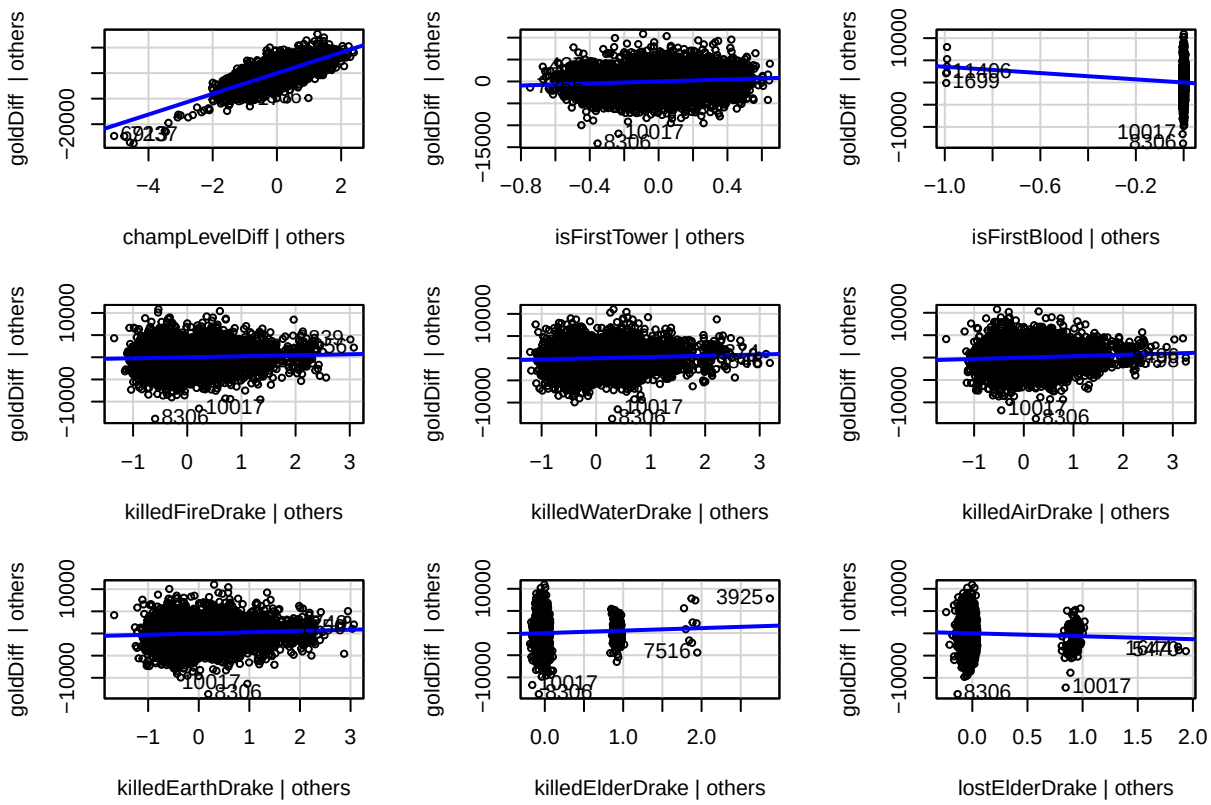

Data
Model
TRUE
20000
−20000
0.0
1.0
2.0
3.0
lostBaronNashor
0.0
0.5
1.0
1.5
2.0
killedRiftHerald
0
1
2
3
4
destroyedBotInhibitor
lostBotInhibitor
0.0
0.2
0.4
0.6
0.8
1.0
destroyedTopNexusTurret
destroyedBotNexusTurret
lostTopNexusTurret
lostBotNexusTurret
destroyedBotBaseTurret
lostBotBaseTurret
destroyedTopInnerTurret
destroyedMidInnerTurret
destroyedBotInnerTurret
lostTopInnerTurret
lostMidInnerTurret
lostBotInnerTurret
destroyedTopOuterTurret
destroyedMidOuterTurret

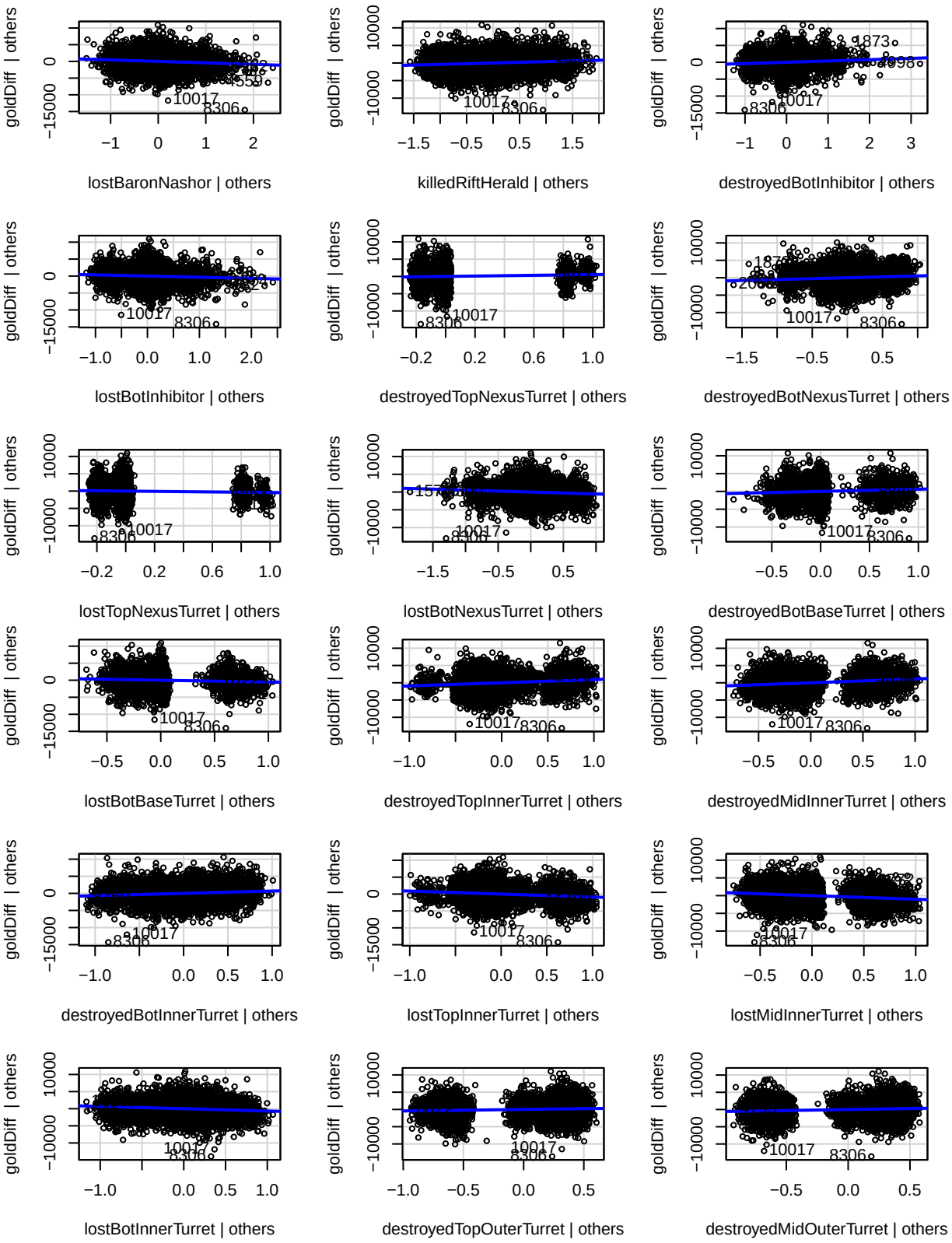


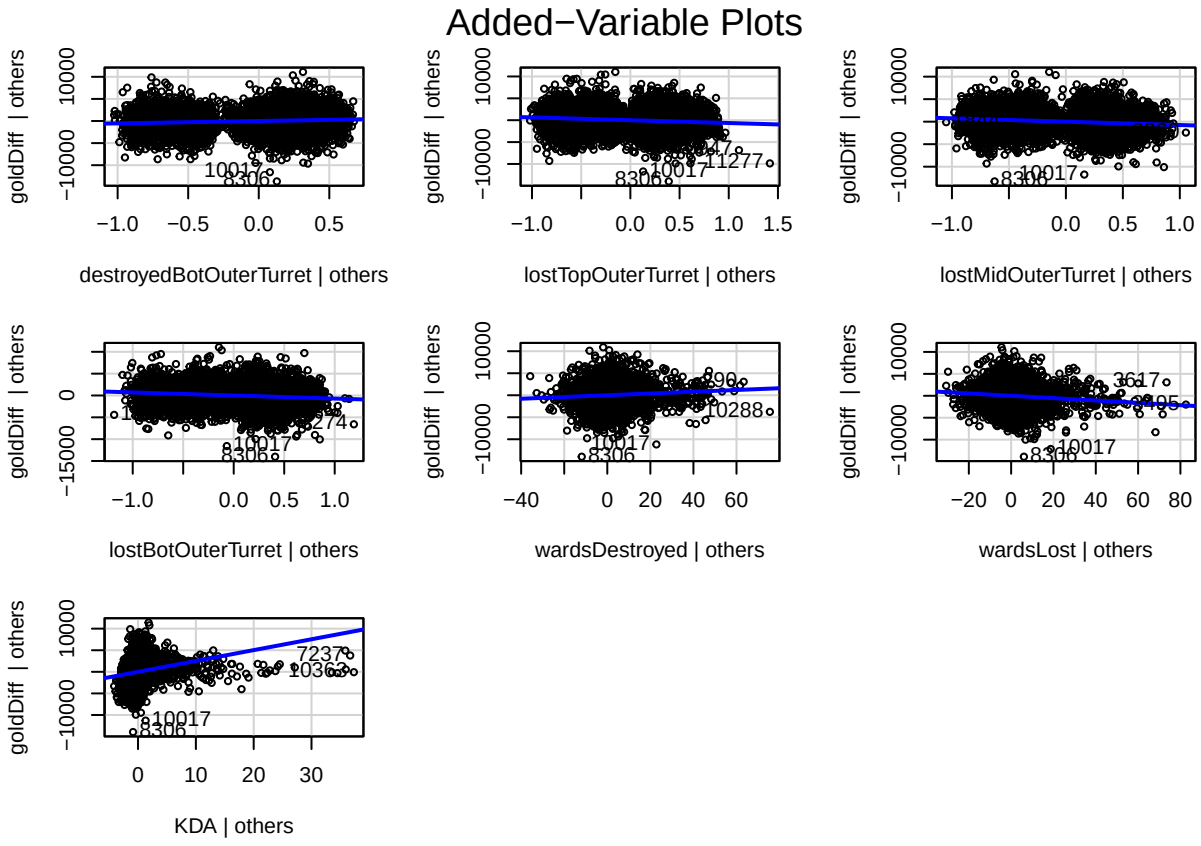
Since for all plots (except for upper tail in KDA), two functions perfectly match with each other, there is evidence that the model fits well.

Added-Variable Plot

```
# Plots for some 'significant' predictors
par(mfrow=c(2,2))
#avPlot(lol.model.forth, "champLevelDiff")
#avPlot(lol.model.forth, "destroyedBotBaseTurret")
#avPlot(lol.model.forth, "lostTopOuterTurret")
#avPlot(lol.model.forth, "KDA")
avPlots(lol.model.forth, ask =F)
```




goldDiff | others
lostBaronNashor | others
killedRiftHerald | others
destroyedBotInhibitor | others
lostBotInhibitor | others
destroyedTopNexusTurret | others
destroyedBotNexusTurret | others
lostTopNexusTurret | others
lostBotNexusTurret | others
destroyedBotBaseTurret | others
lostBotBaseTurret | others
destroyedTopInnerTurret | others
destroyedMidInnerTurret | others
destroyedBotInnerTurret | others
lostTopInnerTurret | others
lostMidInnerTurret | others
lostBotInnerTurret | others
destroyedTopOuterTurret | others
destroyedMidOuterTurret | others


Added−Variable Plots
goldDiff | others
destroyedBotOuterTurret | others
lostTopOuterTurret | others
lostMidOuterTurret | others
lostBotOuterTurret | others
wardsDestroyed | others
wardsLost | others
KDA | others