A Different Way of Expected Wins

Reading Data:

```
hitting <- read.csv("FG_Custom_2015_2019.csv")
hitting[] <- lapply(hitting, gsub, pattern="%", replacement = "")

colnames(hitting)<- c("Season","Team","ExitVelocity_H","LaunchAngle_H","Barrel%_H","HardHit%_H","K%_H",

str(hitting)
```

```
## 'data.frame':    150 obs. of  19 variables:
##  $ Season                   : chr  "2019" "2017" "2019" "2016" ...
##  $ Team                     : chr  "HOU" "HOU" "MIN" "BOS" ...
##  $ ExitVelocity_H           : chr  "88.3" "88" "89.7" "89.1" ...
##  $ LaunchAngle_H            : chr  "13.4" "12.3" "14.9" "11.3" ...
##  $ Barrel%_H                : chr  "6.4" "5.3" "9.3" "5.6" ...
##  $ HardHit%_H               : chr  "35.5" "35.1" "39.6" "35.3" ...
##  $ K%_H                     : chr  "18.2" "17.3" "20.9" "18.4" ...
##  $ BB%_H                    : chr  "10.1" "8.1" "8.2" "8.8" ...
##  $ IsolatedPower_H          : chr  "0.221" "0.196" "0.224" "0.179" ...
##  $ LineDrive%_H             : chr  "21.5" "19.5" "22.1" "20.8" ...
##  $ Groundball%_H            : chr  "41.2" "43.7" "38.6" "44.4" ...
##  $ Flyball%_H               : chr  "37.3" "36.8" "39.3" "34.8" ...
##  $ WeightedRunsCreatedPlus_H: chr  "126" "121" "116" "113" ...
##  $ Runs_H                   : chr  "920" "896" "939" "878" ...
##  $ OPS_H                    : chr  "0.848" "0.823" "0.832" "0.81" ...
##  $ SLG_H                    : chr  "0.495" "0.478" "0.494" "0.461" ...
##  $ BABIP_H                  : chr  "0.296" "0.309" "0.3" "0.32" ...
##  $ HR_H                     : chr  "288" "238" "307" "208" ...
##  $ wOBA_H                   : chr  "0.355" "0.349" "0.347" "0.346" ...
```

```
pitching = read.csv("FG_Pitching_2015_2019.csv")
pitching[] <- lapply(pitching, gsub, pattern="%", replacement = "")

colnames(pitching) = c("Season","Team","Wins","BABIP_P","LOB%_P","HR/FB_P","ERA_P","FIP_P","xFIP_P","WA

str(pitching)
```

```
## 'data.frame':    150 obs. of  25 variables:
##  $ Season          : chr  "2015" "2018" "2016" "2015" ...
##  $ Team            : chr  "STL" "HOU" "CHC" "PIT" ...
##  $ Wins            : chr  "100" "103" "103" "98" ...
##  $ BABIP_P         : chr  "0.297" "0.283" "0.255" "0.302" ...
##  $ LOB%_P          : chr  "79.4" "77.9" "77.5" "75.0" ...
##  $ HR/FB_P         : chr  "9.4" "11.6" "13.1" "9.4" ...
```

```
##  $ ERA_P            : chr  "2.94" "3.11" "3.15" "3.23" ...
##  $ FIP_P            : chr  "3.48" "3.23" "3.77" "3.36" ...
##  $ xFIP_P           : chr  "3.71" "3.36" "3.74" "3.56" ...
##  $ WAR_P            : chr  "20.6" "28.6" "18.8" "19.5" ...
##  $ WHIP_P           : chr  "1.25" "1.1" "1.11" "1.24" ...
##  $ LineDrive%_P     : chr  "20.4" "20.5" "20.7" "22.3" ...
##  $ Groundball%_P    : chr  "48.0" "44.0" "46.9" "50.4" ...
##  $ Flyball%_P       : chr  "31.5" "35.5" "32.4" "27.4" ...
##  $ SwingingStrike%_P: chr  "9.8" "12.7" "11.0" "10.6" ...
##  $ K%_P             : chr  "21.7" "28.5" "24.3" "21.5" ...
##  $ BB%_P            : chr  "7.8" "7.4" "8.3" "7.3" ...
##  $ SIERA_P          : chr  "3.71" "3.27" "3.76" "3.56" ...
##  $ Soft%_P          : chr  "19.1" "20.6" "20.0" "21.6" ...
##  $ Med%_P           : chr  "52.3" "50.2" "50.3" "50.0" ...
##  $ Hard%_P          : chr  "28.6" "29.2" "29.6" "28.5" ...
##  $ ExitVelo_P       : chr  "87.8" "87.7" "88.3" "87.9" ...
##  $ LaunchAngle_P    : chr  "9.9" "11.3" "10.6" "7.8" ...
##  $ Barrel%_P        : chr  "4.3" "5.2" "4.8" "3.9" ...
##  $ HardHit%_P       : chr  "31.4" "32.8" "32.6" "33.3" ...
```

```
pitching_RunsAllowed = read.csv("FG_Pitching_RunsAllowed_2015_2019.csv")
pitching = cbind(pitching,pitching_RunsAllowed[,5])
colnames(pitching) = c("Season","Team","Wins","BABIP_P","LOB%_P","HR/FB_P","ERA_P","FIP_P","xFIP_P","WA
```

Merge Data frames:

```
baseball = merge(x = hitting,y = pitching,by = c("Team", "Season"))
```

Data Manipulation:

```
#Change all columns except team from characters to numeric
i = c(2:length(baseball))
baseball[ , i] <- apply(baseball[ , i], 2,
                    function(x) as.numeric(as.character(x)))
sapply(baseball, class)
```

```
##                      Team                    Season           ExitVelocity_H
##               "character"                 "numeric"                "numeric"
##             LaunchAngle_H                Barrel%_H              HardHit%_H
##                 "numeric"                 "numeric"                "numeric"
##                     K%_H                     BB%_H         IsolatedPower_H
##                 "numeric"                 "numeric"                "numeric"
##              LineDrive%_H             Groundball%_H              Flyball%_H
##                 "numeric"                 "numeric"                "numeric"
## WeightedRunsCreatedPlus_H                    Runs_H                   OPS_H
##                 "numeric"                 "numeric"                "numeric"
##                     SLG_H                   BABIP_H                    HR_H
##                 "numeric"                 "numeric"                "numeric"
##                    wOBA_H                      Wins                  BABIP_P
##                 "numeric"                 "numeric"                "numeric"
##                    LOB%_P                   HR/FB_P                   ERA_P
##                 "numeric"                 "numeric"                "numeric"
```

```
##                   FIP_P                  xFIP_P                   WAR_P
##               "numeric"              "numeric"               "numeric"
##                  WHIP_P            LineDrive%_P          Groundball%_P
##               "numeric"              "numeric"               "numeric"
##              Flyball%_P       SwingingStrike%_P                    K%_P
##               "numeric"              "numeric"               "numeric"
##                  BB%_P                SIERA_P                 Soft%_P
##               "numeric"              "numeric"               "numeric"
##                  Med%_P                Hard%_P              ExitVelo_P
##               "numeric"              "numeric"               "numeric"
##            LaunchAngle_P              Barrel%_P              HardHit%_P
##               "numeric"              "numeric"               "numeric"
##                   RA_P
##               "numeric"
```

Function to calculation expected wins:

```
PythagoreanWinningPercentage = function(RS,RA)
{
  (RS^1.83/ (RS^1.83 + RA^1.83))
}
```

Summary Statistics:

```
#summary statistics
summary(baseball)
```

```
##      Team               Season     ExitVelocity_H  LaunchAngle_H
##  Length:150         Min.   :2015   Min.   :85.50   Min.   : 6.20
##  Class :character   1st Qu.:2016   1st Qu.:87.80   1st Qu.:10.80
##  Mode  :character   Median :2017   Median :88.30   Median :11.80
##                     Mean   :2017   Mean   :88.24   Mean   :11.87
##                     3rd Qu.:2018   3rd Qu.:88.80   3rd Qu.:13.18
##                     Max.   :2019   Max.   :90.00   Max.   :15.30
##    Barrel%_H       HardHit%_H         K%_H           BB%_H
##  Min.   :3.100   Min.   :27.10   Min.   :15.90   Min.   : 6.300
##  1st Qu.:4.900   1st Qu.:32.73   1st Qu.:20.30   1st Qu.: 7.500
##  Median :5.750   Median :34.40   Median :21.60   Median : 8.200
##  Mean   :5.824   Mean   :34.45   Mean   :21.68   Mean   : 8.267
##  3rd Qu.:6.500   3rd Qu.:36.20   3rd Qu.:23.07   3rd Qu.: 9.000
##  Max.   :9.300   Max.   :40.00   Max.   :26.40   Max.   :10.500
##  IsolatedPower_H   LineDrive%_H   Groundball%_H     Flyball%_H
##  Min.   :0.1080   Min.   :18.70   Min.   :38.10   Min.   :27.60
##  1st Qu.:0.1502   1st Qu.:20.30   1st Qu.:42.50   1st Qu.:33.38
##  Median :0.1645   Median :21.00   Median :43.90   Median :35.00
##  Mean   :0.1654   Mean   :20.97   Mean   :44.03   Mean   :35.00
##  3rd Qu.:0.1817   3rd Qu.:21.57   3rd Qu.:45.80   3rd Qu.:36.60
##  Max.   :0.2240   Max.   :24.60   Max.   :51.90   Max.   :41.10
##  WeightedRunsCreatedPlus_H     Runs_H          OPS_H            SLG_H
##  Min.   : 76.00            Min.   :573.0   Min.   :0.6590   Min.   :0.3570
##  1st Qu.: 90.25            1st Qu.:680.5   1st Qu.:0.7190   1st Qu.:0.4010
##  Median : 96.00            Median :729.0   Median :0.7380   Median :0.4170
##  Mean   : 96.57            Mean   :733.8   Mean   :0.7389   Mean   :0.4183
```

```
##  3rd Qu.:101.75            3rd Qu.:773.2   3rd Qu.:0.7570   3rd Qu.:0.4338
##  Max.   :126.00            Max.   :943.0   Max.   :0.8480   Max.   :0.4950
##     BABIP_H            HR_H           wOBA_H            Wins
##  Min.   :0.2760   Min.   :100.0   Min.   :0.2880   Min.   : 47.00
##  1st Qu.:0.2920   1st Qu.:165.2   1st Qu.:0.3093   1st Qu.: 71.00
##  Median :0.2980   Median :193.5   Median :0.3165   Median : 81.00
##  Mean   :0.2985   Mean   :193.2   Mean   :0.3173   Mean   : 80.97
##  3rd Qu.:0.3050   3rd Qu.:220.0   3rd Qu.:0.3250   3rd Qu.: 91.00
##  Max.   :0.3320   Max.   :307.0   Max.   :0.3550   Max.   :108.00
##     BABIP_P           LOB%_P        HR/FB_P          ERA_P
##  Min.   :0.2550   Min.   :68.0   Min.   : 9.40   Min.   :2.940
##  1st Qu.:0.2890   1st Qu.:71.1   1st Qu.:12.00   1st Qu.:3.820
##  Median :0.2965   Median :72.9   Median :13.00   Median :4.180
##  Mean   :0.2959   Mean   :72.8   Mean   :13.17   Mean   :4.235
##  3rd Qu.:0.3030   3rd Qu.:74.2   3rd Qu.:14.10   3rd Qu.:4.630
##  Max.   :0.3200   Max.   :79.4   Max.   :19.00   Max.   :5.670
##      FIP_P           xFIP_P          WAR_P           WHIP_P
##  Min.   :3.230   Min.   :3.330   Min.   : 1.00   Min.   :1.100
##  1st Qu.:3.970   1st Qu.:3.982   1st Qu.:10.03   1st Qu.:1.252
##  Median :4.220   Median :4.260   Median :14.15   Median :1.320
##  Mean   :4.235   Mean   :4.235   Mean   :14.33   Mean   :1.320
##  3rd Qu.:4.525   3rd Qu.:4.465   3rd Qu.:18.48   3rd Qu.:1.380
##  Max.   :5.560   Max.   :5.230   Max.   :30.40   Max.   :1.510
##   LineDrive%_P   Groundball%_P    Flyball%_P    SwingingStrike%_P
##  Min.   :18.70   Min.   :38.30   Min.   :27.40   Min.   : 8.40
##  1st Qu.:20.30   1st Qu.:42.52   1st Qu.:33.60   1st Qu.: 9.80
##  Median :21.00   Median :44.00   Median :35.00   Median :10.35
##  Mean   :20.96   Mean   :44.05   Mean   :34.98   Mean   :10.47
##  3rd Qu.:21.60   3rd Qu.:45.38   3rd Qu.:36.40   3rd Qu.:11.00
##  Max.   :23.10   Max.   :50.40   Max.   :40.50   Max.   :13.00
##       K%_P           BB%_P           SIERA_P         Soft%_P
##  Min.   :17.00   Min.   : 6.100   Min.   :3.270   Min.   :15.00
##  1st Qu.:19.73   1st Qu.: 7.700   1st Qu.:3.905   1st Qu.:17.32
##  Median :21.55   Median : 8.300   Median :4.190   Median :18.50
##  Mean   :21.70   Mean   : 8.267   Mean   :4.152   Mean   :18.31
##  3rd Qu.:23.27   3rd Qu.: 8.800   3rd Qu.:4.378   3rd Qu.:19.25
##  Max.   :28.50   Max.   :10.300   Max.   :4.890   Max.   :21.60
##      Med%_P         Hard%_P        ExitVelo_P    LaunchAngle_P
##  Min.   :41.90   Min.   :25.60   Min.   :86.20   Min.   : 7.80
##  1st Qu.:46.45   1st Qu.:29.95   1st Qu.:87.83   1st Qu.:10.90
##  Median :49.15   Median :31.95   Median :88.40   Median :11.90
##  Mean   :48.66   Mean   :33.04   Mean   :88.23   Mean   :11.86
##  3rd Qu.:50.70   3rd Qu.:35.80   3rd Qu.:88.60   3rd Qu.:12.90
##  Max.   :54.30   Max.   :42.80   Max.   :89.90   Max.   :15.70
##    Barrel%_P       HardHit%_P         RA_P
##  Min.   :3.800   Min.   :28.60   Min.   :525.0
##  1st Qu.:5.200   1st Qu.:33.10   1st Qu.:670.5
##  Median :5.800   Median :34.20   Median :723.0
##  Mean   :5.799   Mean   :34.41   Mean   :733.8
##  3rd Qu.:6.500   3rd Qu.:35.77   3rd Qu.:795.5
##  Max.   :8.300   Max.   :40.50   Max.   :981.0
```

Exploring Wins compared to the other variable in the data set:

```r
#Remove Team from data so dataframe and be used of modeling
baseball_m = baseball[,-c(1,2,14,24,25,26,27,35)]
```

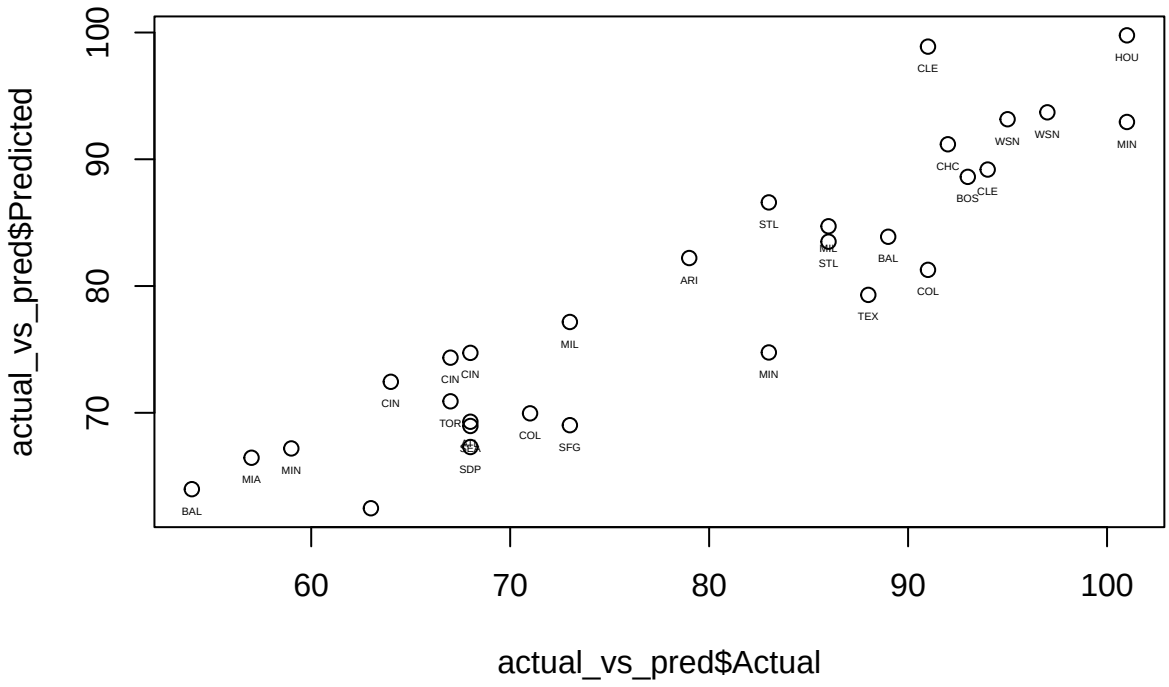
Modeling:

```r
model = glm(formula = Wins ~ WHIP_P + wOBA_H + `LOB%_P`,
    family = poisson(link = "log"), data = train)

predict(model, type="response",test)
```

```
##        1        7       12       15       18       23       31       33
## 82.20959 69.29579 83.88855 63.97649 88.61239 91.18369 72.44613 74.72877
##       34       37       39       44       45       53       74       75
## 74.35045 89.18811 98.88520 81.27706 69.95791 99.77530 62.47241 66.45661
##       77       78       81       82       85      112      120      124
## 77.16532 84.71546 74.76029 67.18932 92.94156 67.30138 68.96185 69.02326
##      127      128      136      145      147      148
## 83.50118 86.59634 79.30267 70.90639 93.15746 93.70301
```

```r
actual_vs_pred = data.frame( predict(model, type="response",test),baseball_m$Wins[-index])
colnames(actual_vs_pred) = c("Predicted","Actual")
actual_vs_pred$team = baseball$Team[-index]
actual_vs_pred$season = baseball$Season[-index]
actual_vs_pred$RS =  baseball$Runs_H[-index]
actual_vs_pred$RA =  baseball$RA_P[-index]
actual_vs_pred$PythagoreanW = PythagoreanWinningPercentage(actual_vs_pred$RS,actual_vs_pred$RA) * 162
actual_vs_pred$PWL_Diff = actual_vs_pred$Actual - actual_vs_pred$PythagoreanW
actual_vs_pred$diff = actual_vs_pred$Actual - actual_vs_pred$Predicted
actual_vs_pred$change = abs(actual_vs_pred$PWL_Diff) - abs(actual_vs_pred$diff)
```

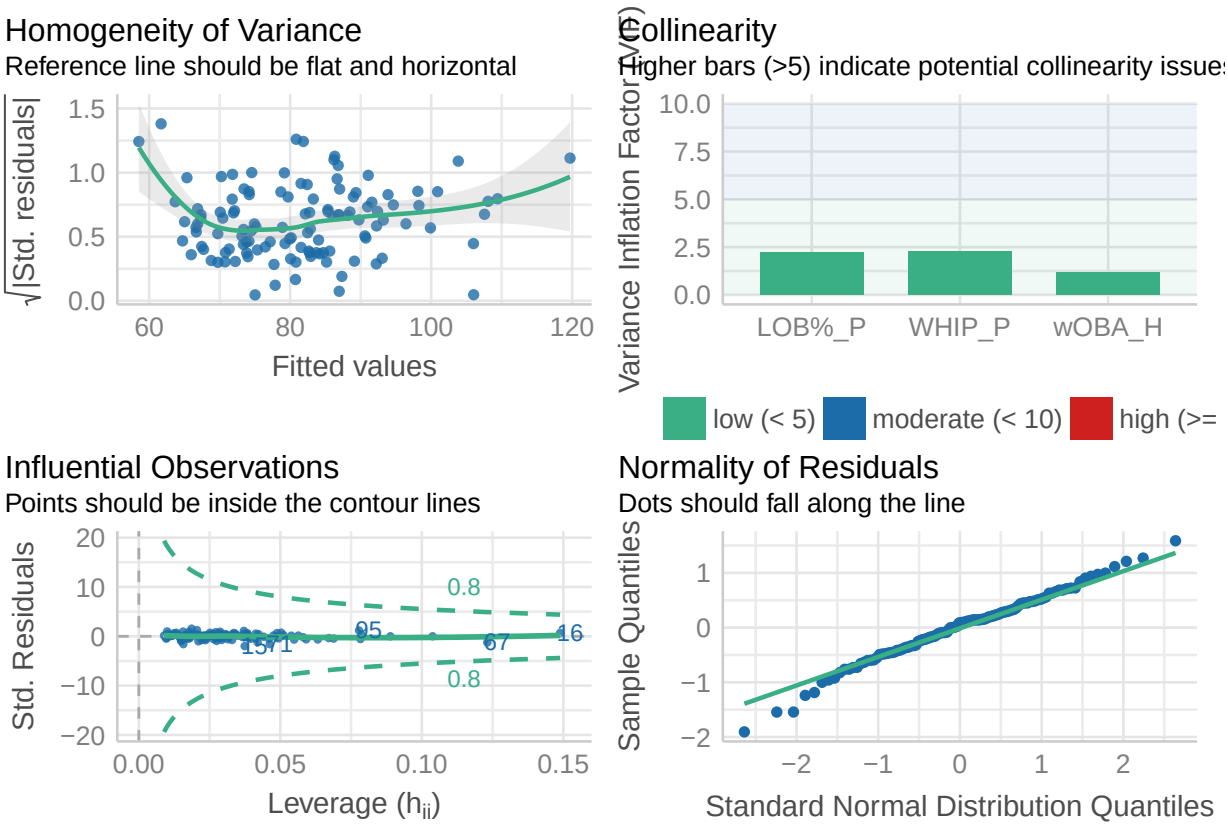



```
model2 = glm(formula = Wins ~ WHIP_P + wOBA_H + `LOB%_P` + `K%_H`,
    family = poisson(link = "log"), data = train)

model3 = lm(Wins ~ Runs_H, data = baseball)
model4 = lm(Wins ~ RA_P, data = baseball)
```

Regression Diagnostics

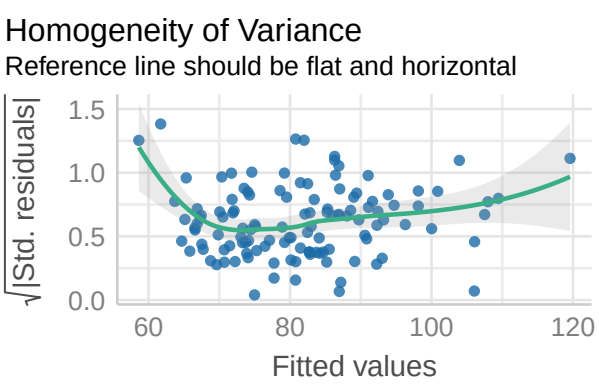
```
library(performance)
check_model(model)
```

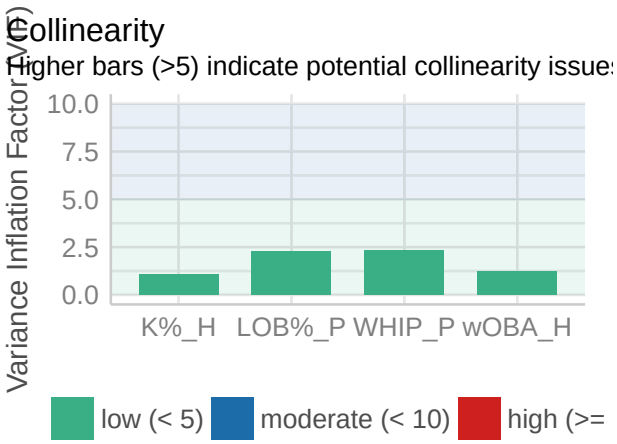


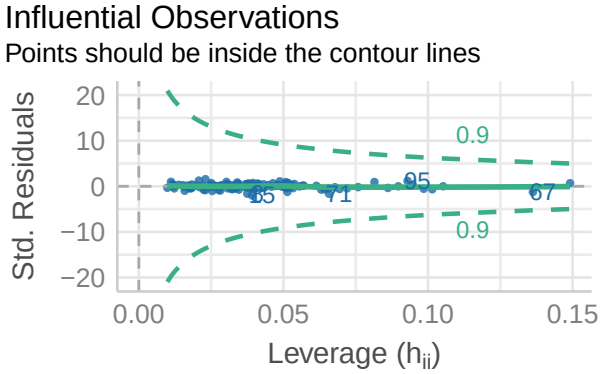

```
model_performance(model)
```

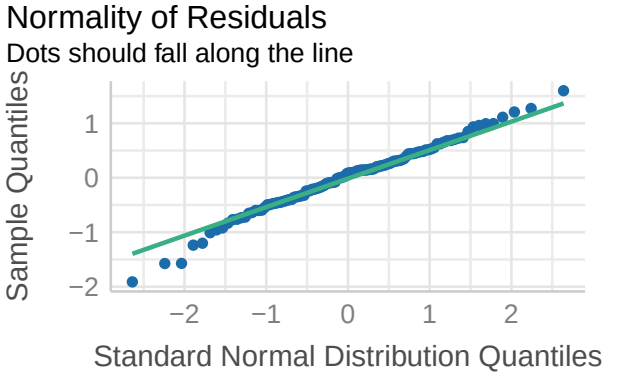
```
## # Indices of model performance
##
## AIC     |     BIC | Nagelkerke's R2 |  RMSE | Sigma | Score_log | Score_spherical
## ------------------------------------------------------------------------------------
## 795.994 | 807.144 |           0.928 | 5.214 | 0.592 |    -3.283 |           0.089
```

```
check_model(model2)
```










```
model_performance(model2)
```

```
## # Indices of model performance
##
## AIC     |     BIC | Nagelkerke's R2 |  RMSE | Sigma | Score_log | Score_spherical
## ------------------------------------------------------------------------------------
## 797.986 | 811.923 |           0.928 | 5.212 | 0.594 |    -3.283 |           0.089
```