Project_625

Soccer Regression

2022-11-11

Slide with Bullets

- Exploratory data analysis
- Model Building
- Evaluating and interpretation

Slide with R Output

```
data1=read.csv("C:/Users/USER/Desktop/SoccerRegression/Sample_data/Project_Data.csv")
head(data1)
```

```
##   leage        team Gls No.Pl no.sh SoT PK FK Att.3rd.T Succ.Drib Att.Drib
## 1   Epl      Arsenal  60    27   581 186  5 22      6399       319      590
## 2   Epl Aston Villa  50    31   461 159  3 20      4983       333      587
## 3   Epl    Brentford  46    29   436 141  6 11      4515       262      476
## 4   Epl     Brighton  40    26   482 141  4 15      6401       317      559
## 5   Epl      Burnley  32    23   405 119  1 14      4639       258      479
## 6   Epl      Chelsea  75    26   583 200  8 26      7764       364      624
##   Touches Prog.T GCA.Drib ShortAT_Pass MediumtAT_Pass LongAT_Pass
## 1   23628   1058        3         8210           8140        2584
## 2   20474    916        9         6803           6136        2571
## 3   20340    869        5         6312           6120        2852
## 4   24673   1043        3         8172           8207        2954
## 5   18247    708        5         5327           4472        3216
## 6   28767   1314        8        11617           9322        2514
```

Slide with Eda

```
## [1] 50.72449
```

```
## [1] 267.8718
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```
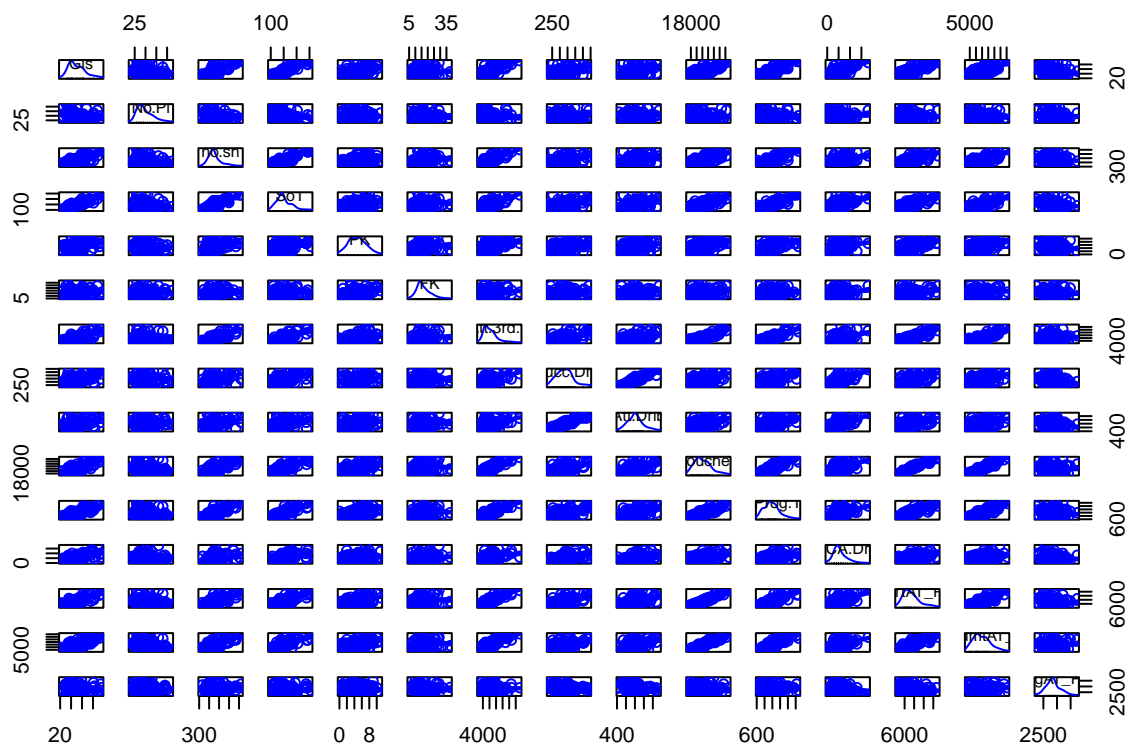
Normal linear regression

```
##      leage       team Gls No.Pl no.sh SoT PK FK Att.3rd.T Succ.Drib Att.Drib
## 93 SerieA Sampdoria  42    34   386 118  2 12     4331       228      422
## 94 SerieA  Sassuolo  60    29   569 204  7 14     5259       403      678
## 95 SerieA    Spezia  38    30   381 127  3 17     4121       292      546
## 96 SerieA    Torino  43    31   470 142  6 16     5633       295      512
## 97 SerieA   Udinese  58    29   505 181  3 19     4780       363      680
## 98 SerieA   Venezia  34    39   349 113  3 22     3885       300      577
##     Touches Prog.T GCA.Drib ShortAT_Pass MediumtAT_Pass LongAT_Pass
## 93    20845    915        3         6565           6436        2904
## 94    24243   1137       10         8959           7443        2681
## 95    19562    725        3         6112           5591        2833
## 96    22075    879        6         7220           6933        2885
## 97    19463    844        8         6357           5570        2322
## 98    19554    768        1         5783           5800        2780
##
## Call:
## lm(formula = Gls ~ . - team, data = data1)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -12.7847  -3.1472   0.1909   3.4677  11.5182
##
## Coefficients:
```
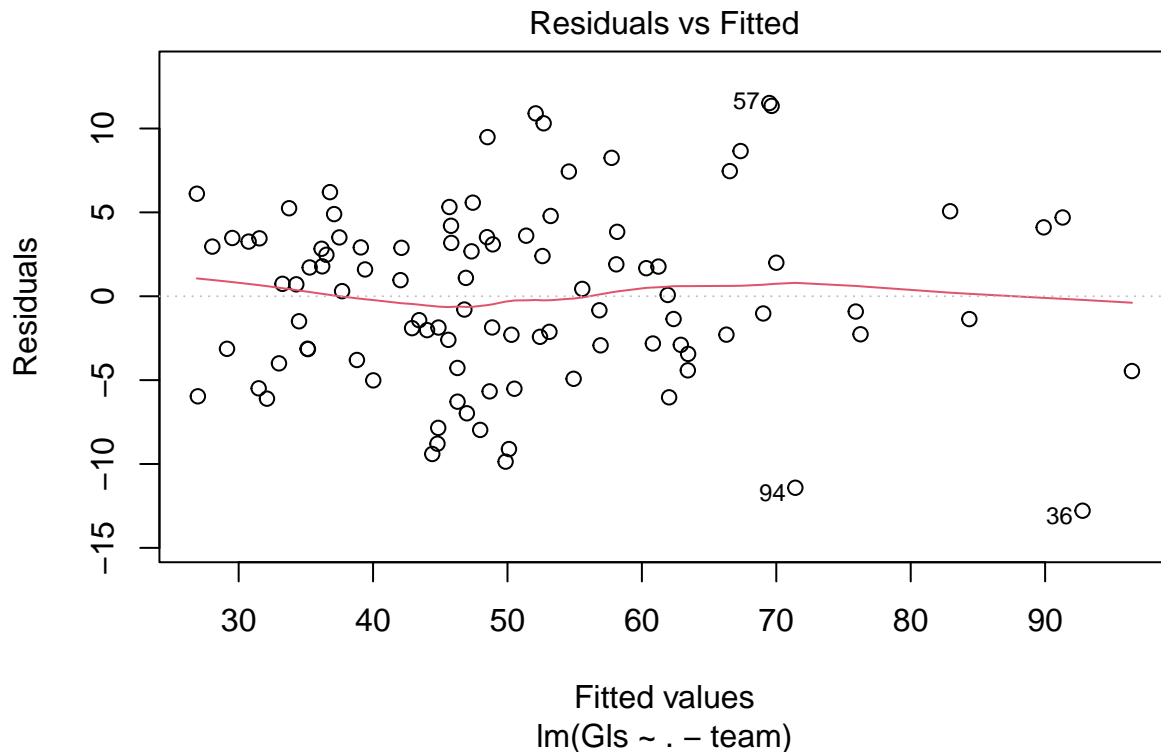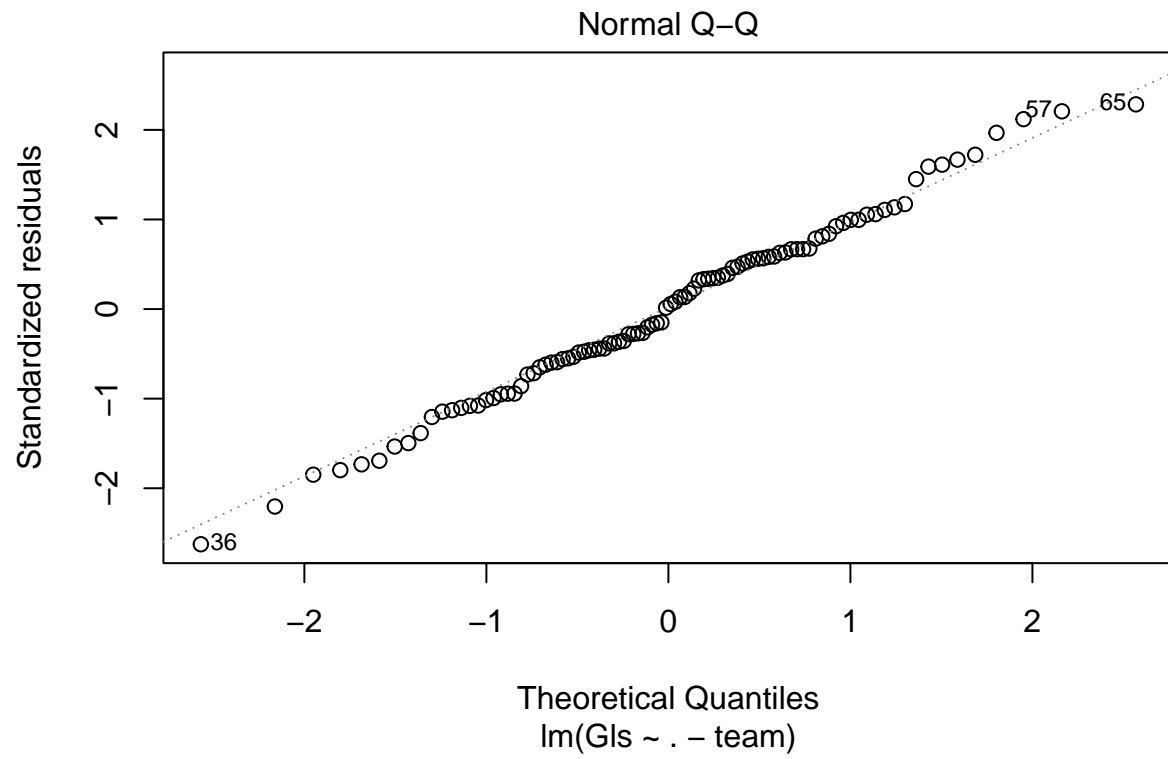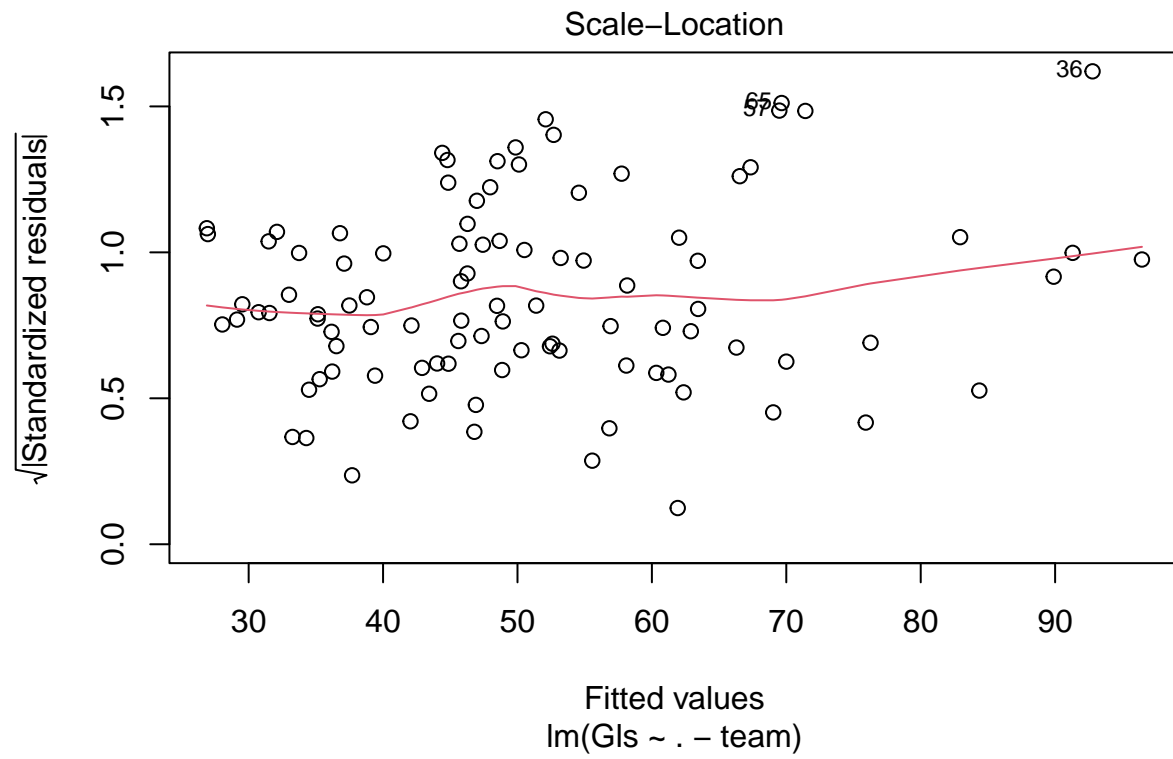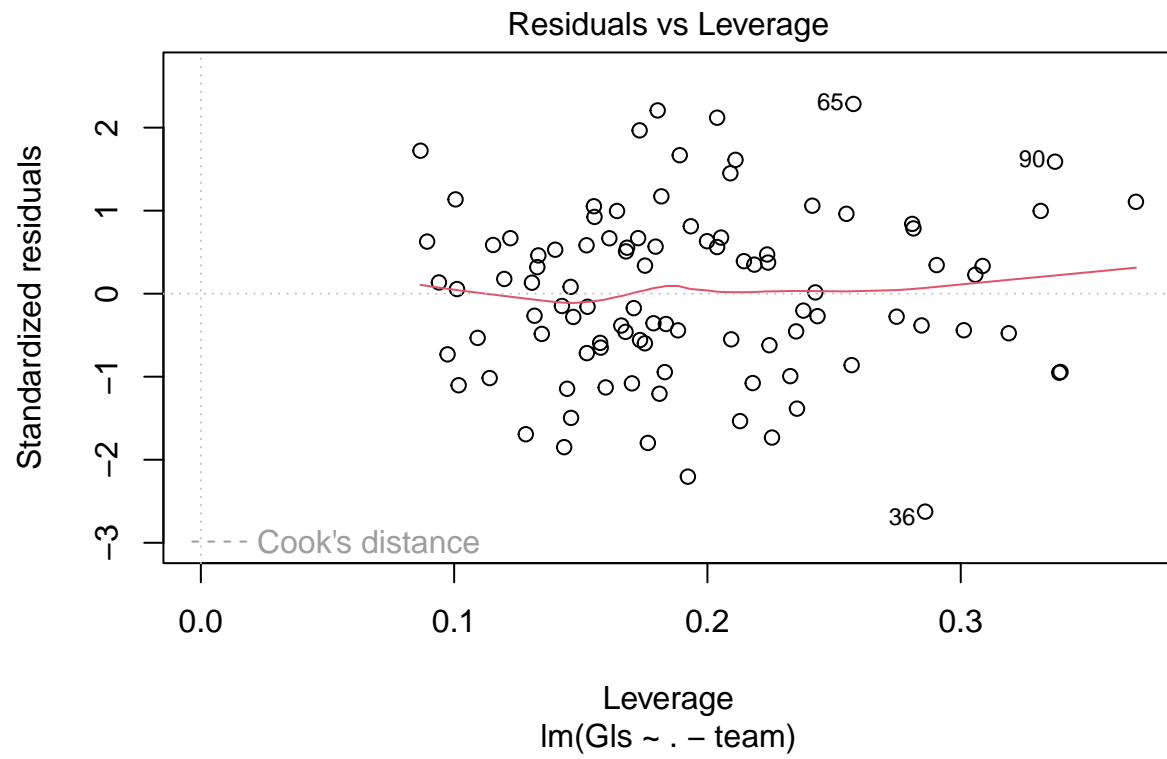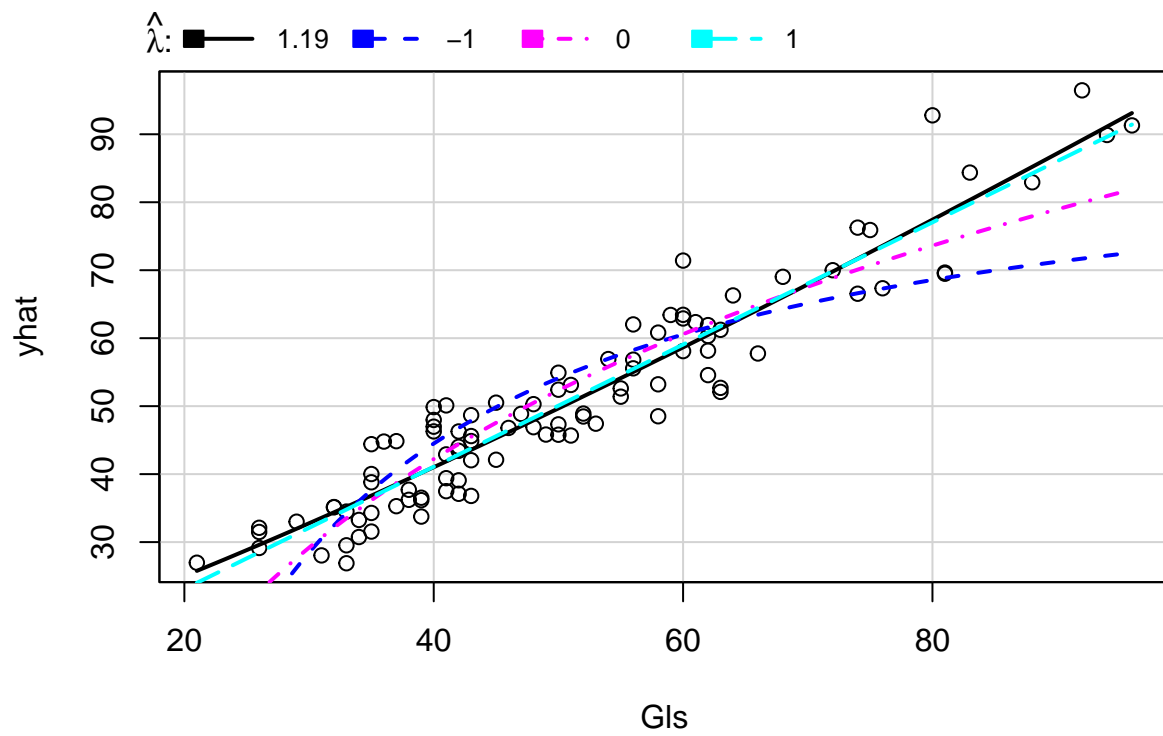
```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.697553  22.616813   0.031  0.97547
## leageEpl      -2.872250   2.678305  -1.072  0.28680
## leageLaliga   -5.091798   2.920440  -1.744  0.08514 .
## leageLigue1   -2.986266   2.450599  -1.219  0.22663
## leageSerieA   -2.472532   2.449365  -1.009  0.31584
## No.Pl          0.052577   0.179208   0.293  0.76999
## no.sh         -0.068939   0.027138  -2.540  0.01304 *
## SoT            0.411317   0.063921   6.435  8.8e-09 ***
## PK             0.606081   0.306578   1.977  0.05154 .
## FK            -0.113253   0.114452  -0.990  0.32543
## Att.3rd.T      0.002343   0.001819   1.288  0.20151
## Succ.Drib      0.005947   0.036716   0.162  0.87174
## Att.Drib      -0.018412   0.022177  -0.830  0.40892
## Touches       -0.003205   0.003533  -0.907  0.36715
## Prog.T         0.011743   0.009952   1.180  0.24156
## GCA.Drib       0.718889   0.266248   2.700  0.00848 **
## ShortAT_Pass   0.004504   0.003869   1.164  0.24781
## MediumtAT_Pass 0.002601   0.003232   0.805  0.42334
## LongAT_Pass    0.007285   0.004968   1.467  0.14648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.764 on 79 degrees of freedom
## Multiple R-squared:  0.899,  Adjusted R-squared:  0.876
## F-statistic: 39.06 on 18 and 79 DF,  p-value: < 2.2e-16
```
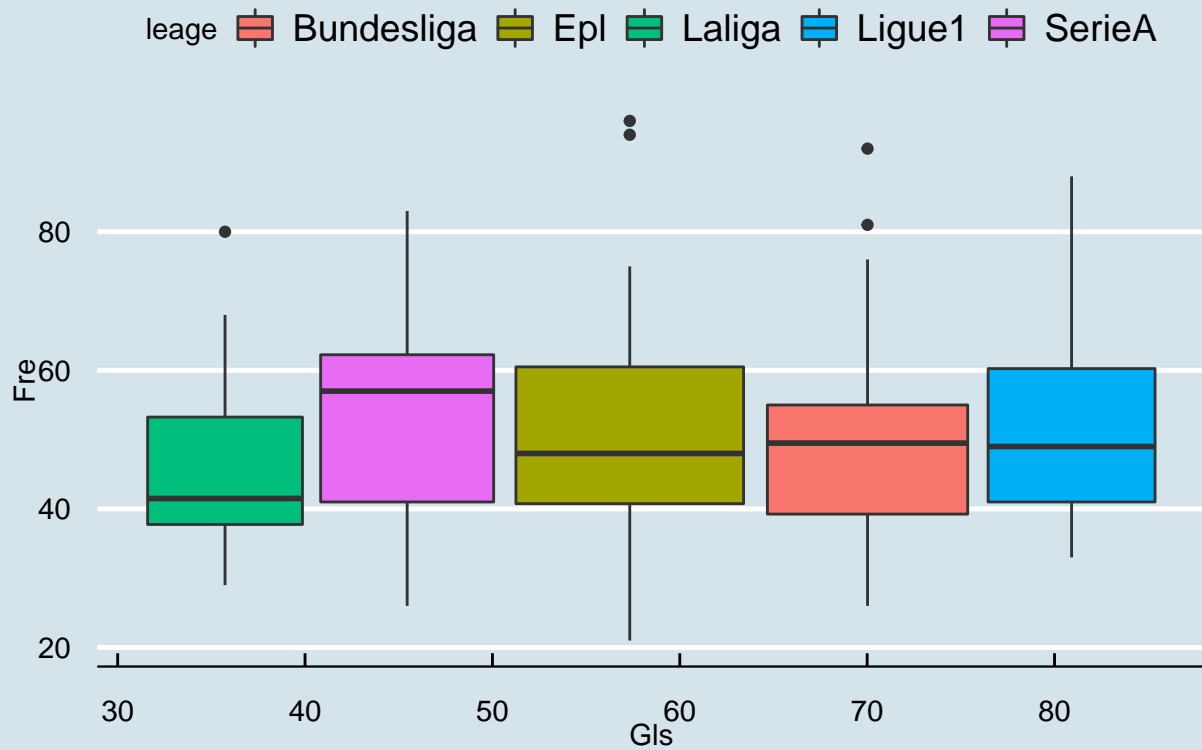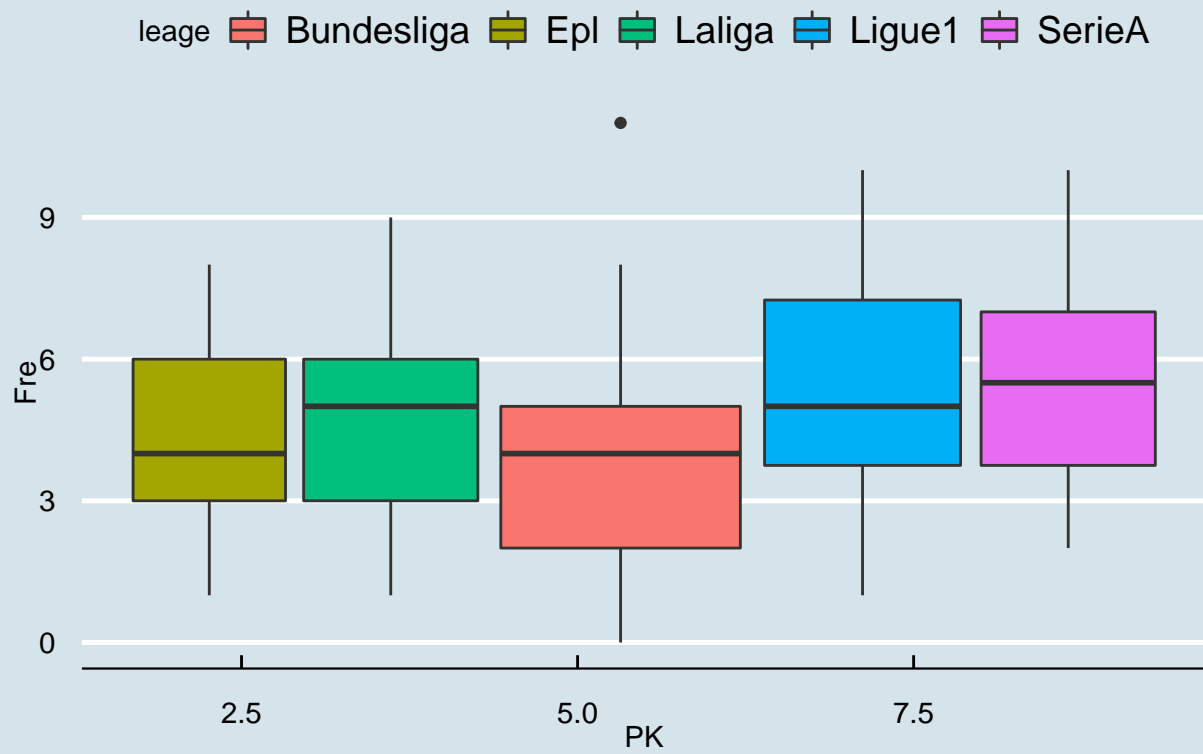

Residuals vs Fitted

lm(Gls ~ . – team)

Normal Q–Q

Theoretical Quantiles
lm(Gls ~ . − team)

4

Scale–Location

√|Standardized residuals|

Fitted values
lm(Gls ~ . – team)

Residuals vs Leverage

lm(Gls ~ . – team)

```
##       lambda      RSS
## 1  1.186499 2334.911
## 2 -1.000000 5964.748
## 3  0.000000 3397.275
## 4  1.000000 2359.527
```

Gls Boxplot
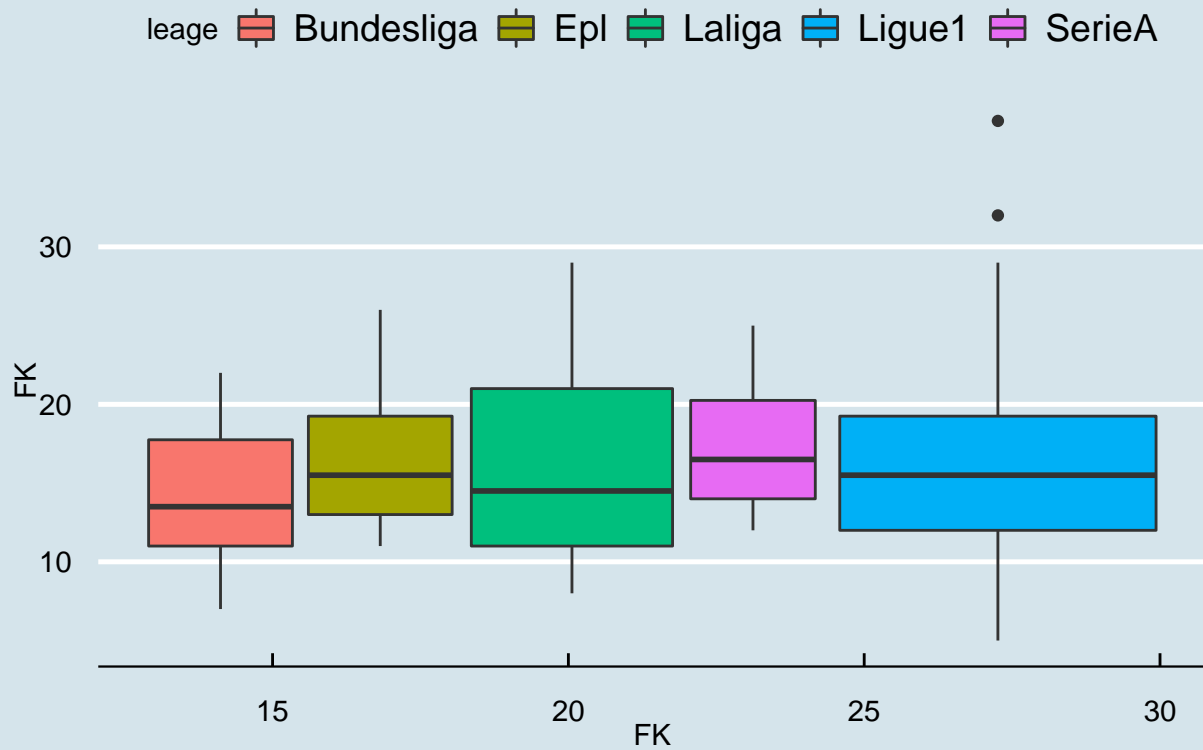
leage ☐ Bundesliga ☐ Epl ☐ Laliga ☐ Ligue1 ☐ SerieA

PK Boxplot

FK Boxplot

leage ▭ Bundesliga ▭ Epl ▭ Laliga ▭ Ligue1 ▭ SerieA
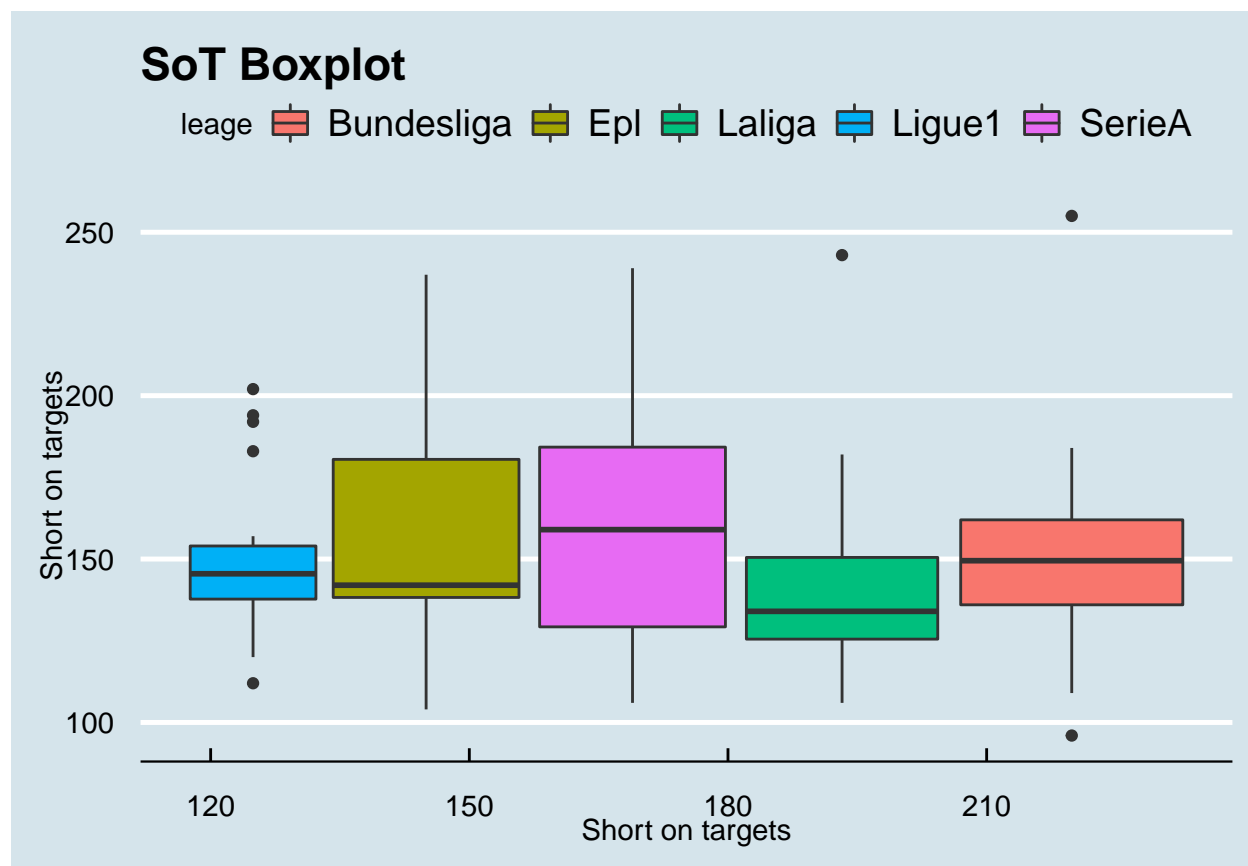
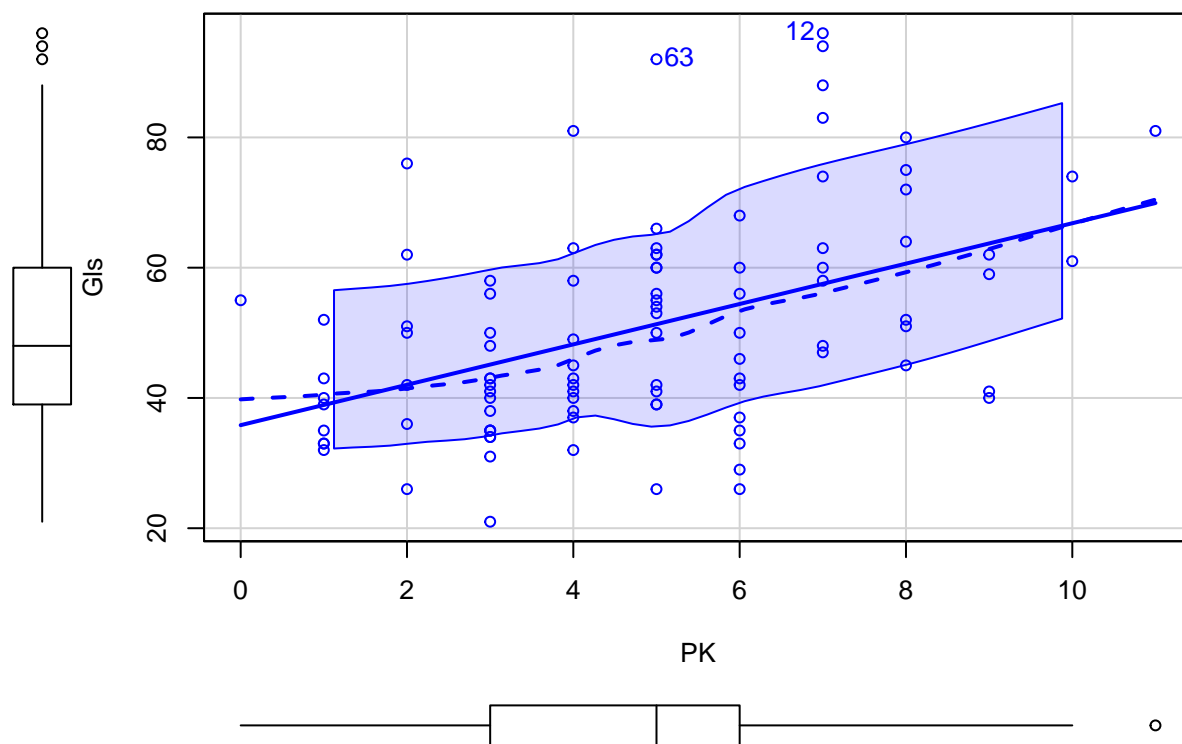**SoT Boxplot**

leage ⊟ Bundesliga ⊟ Epl ⊟ Laliga ⊟ Ligue1 ⊟ SerieA

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```

```
## [1] 12 63
```

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```
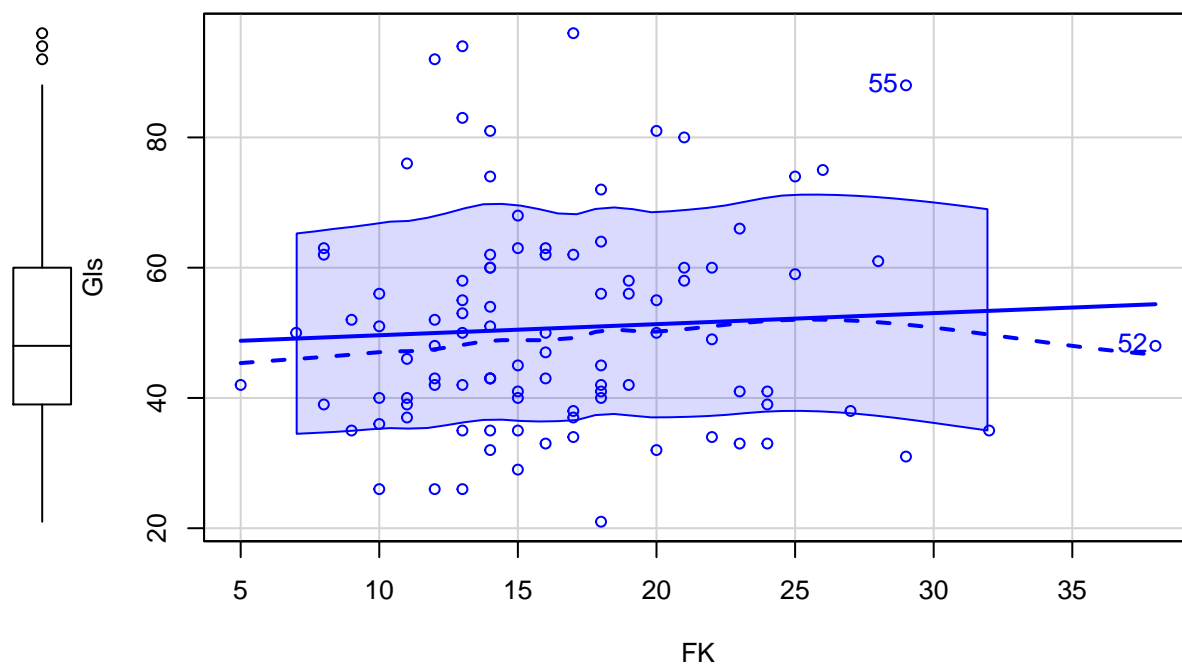
```
## [1] 52 55
```

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```

```
## [1] 11 65

## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```

```
## [1] 36 65

## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```
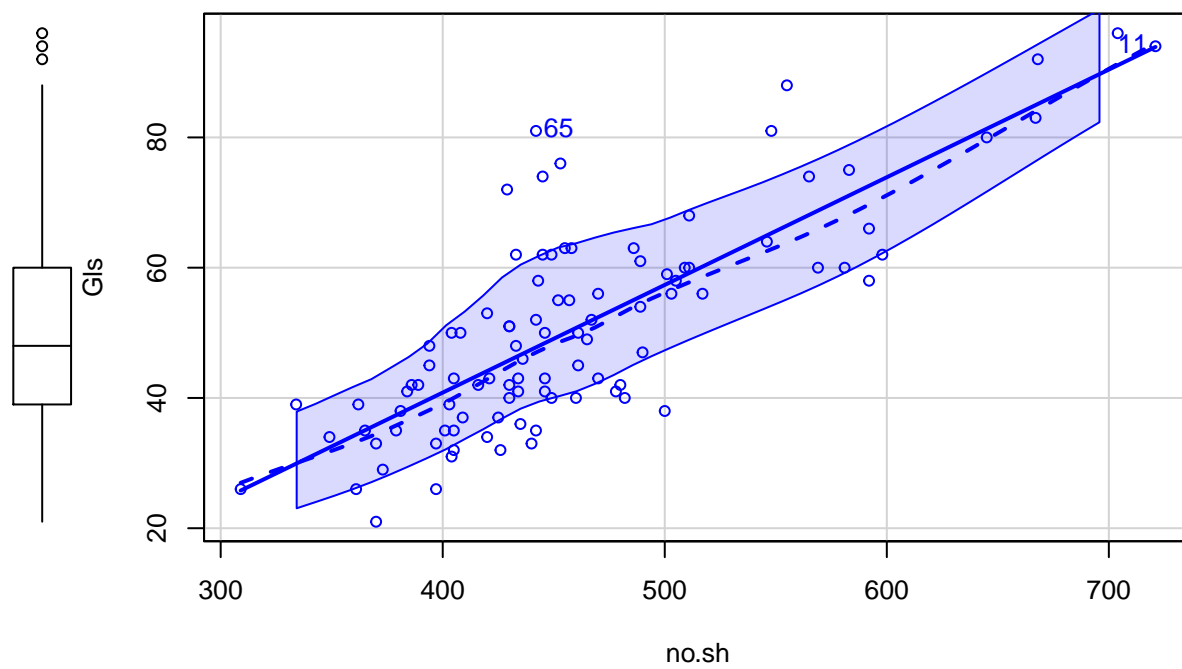
```
## [1] 12 72
```

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```
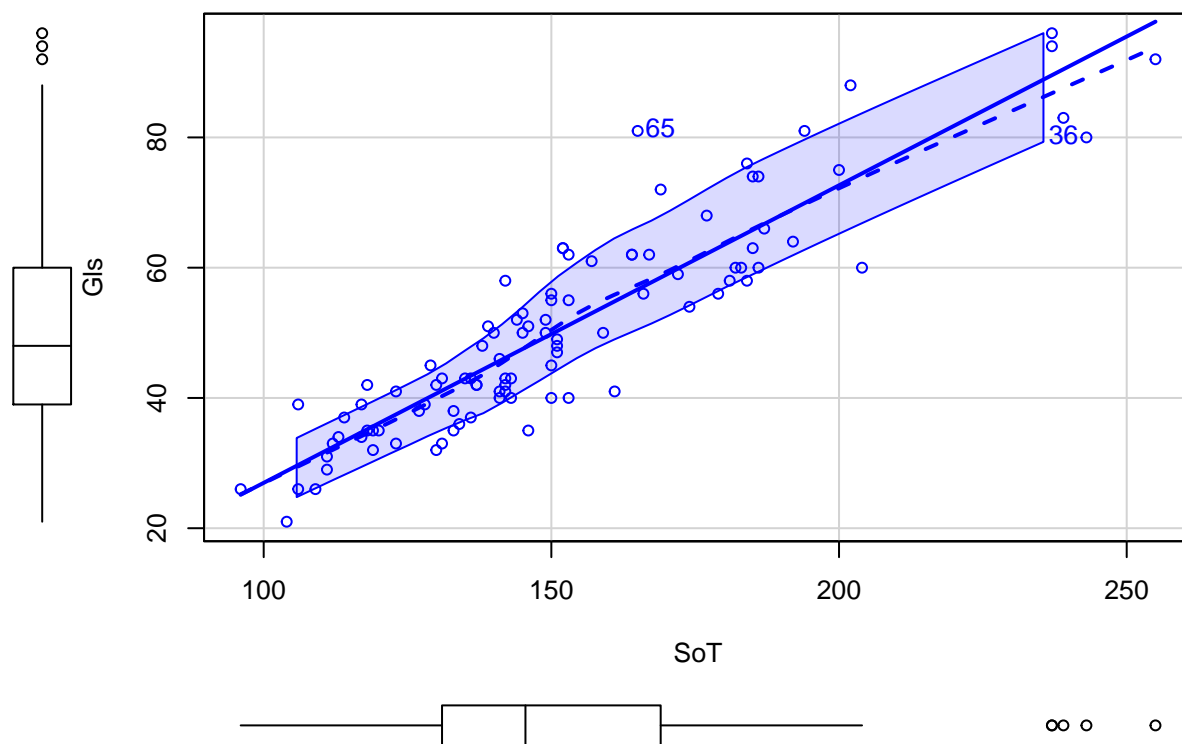
```
## [1] 55 86
```

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```

```
## [1] 63 86

## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```

```
## [1] 55 92
```

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```

```
## [1] 12 55
```

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```
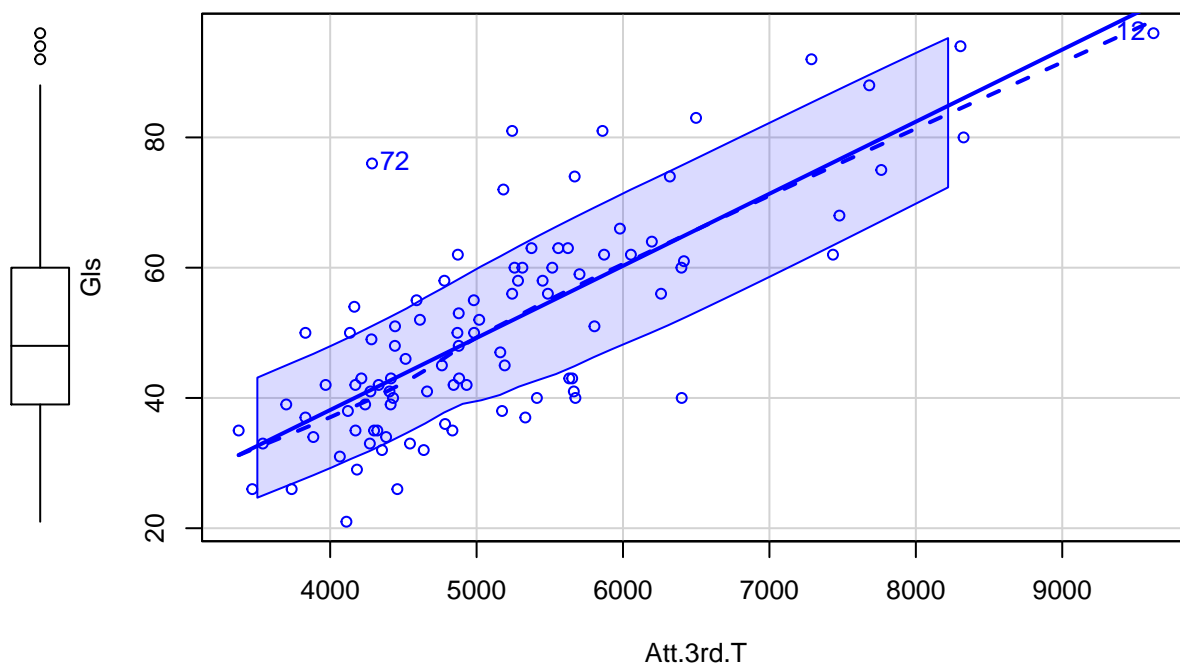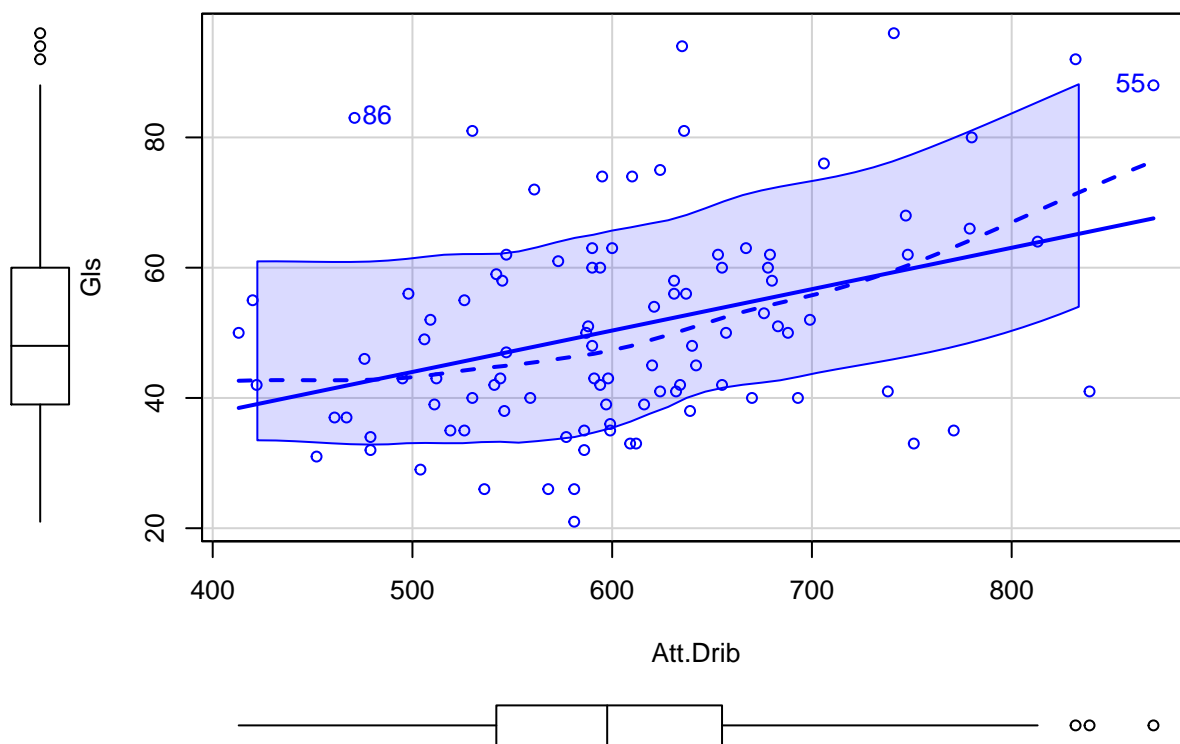
```
## [1] 11 12
```
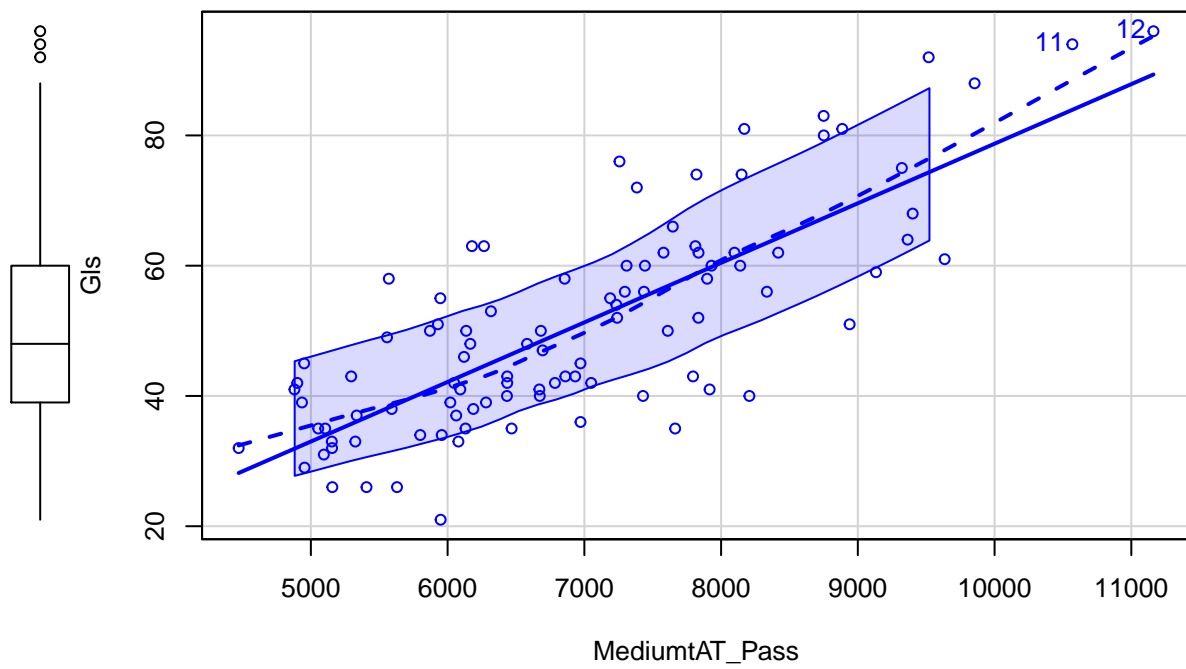
```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```

```
## [1] 11 38

## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```
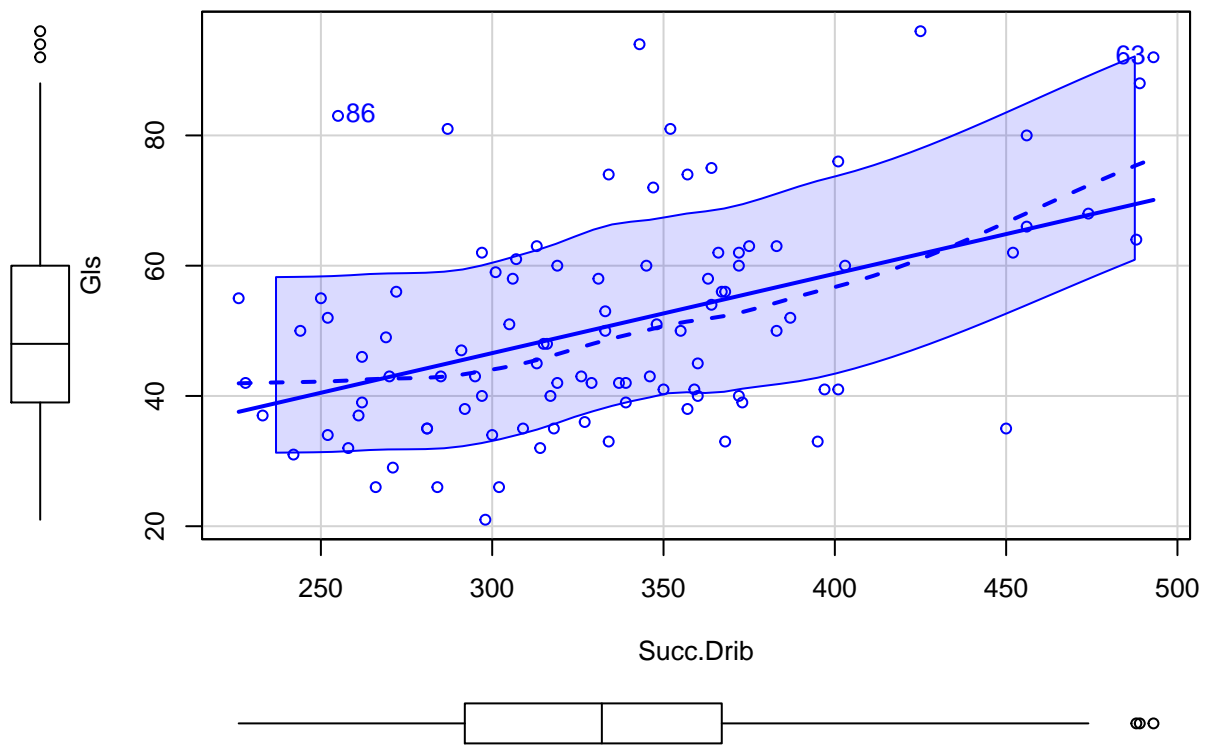
```
## [1] 11 12
```

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```
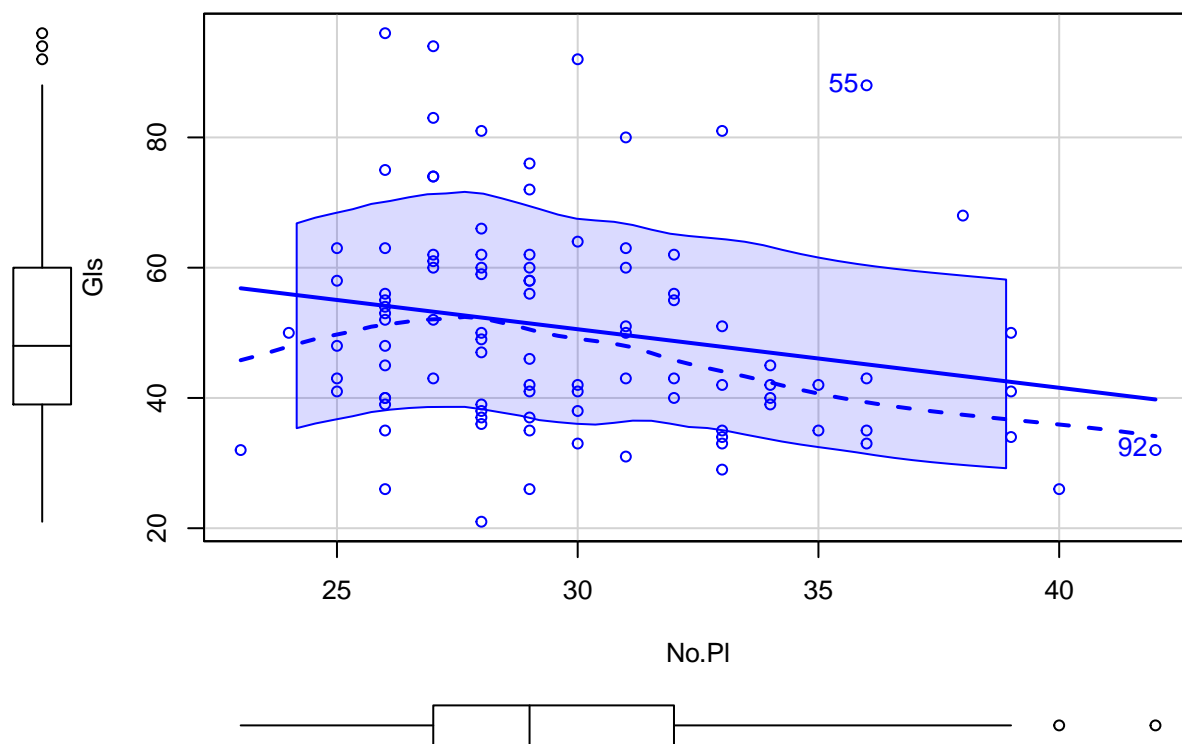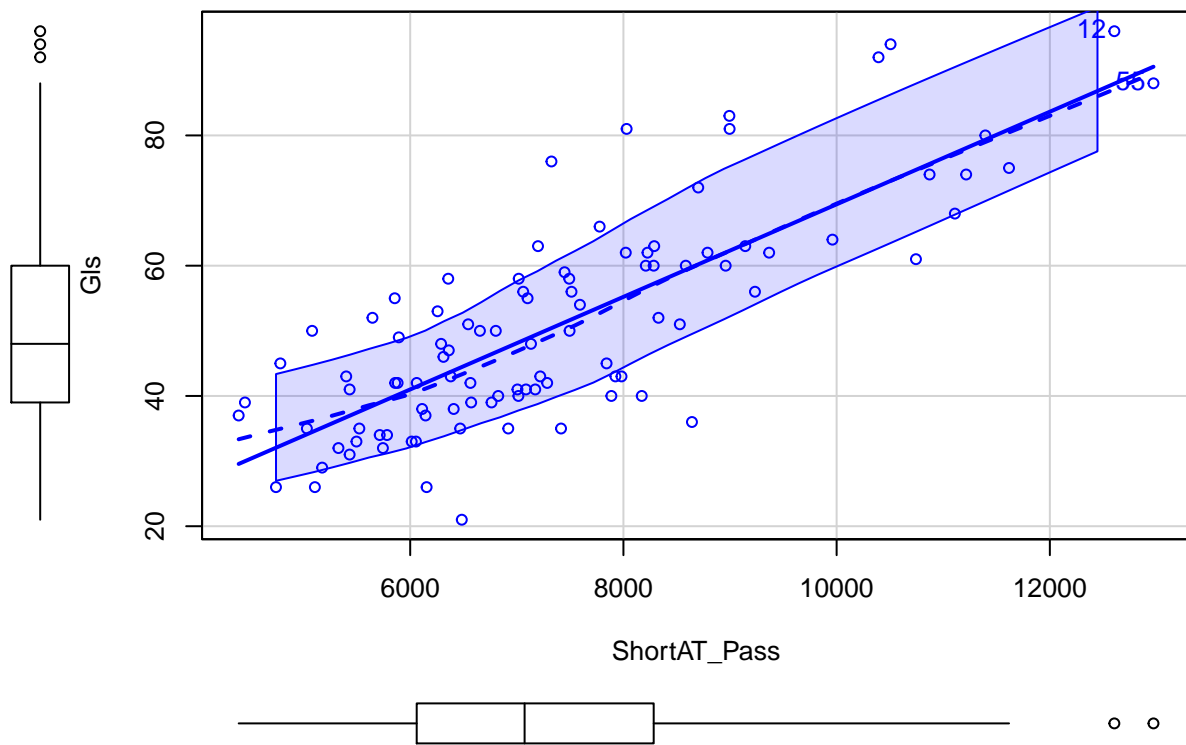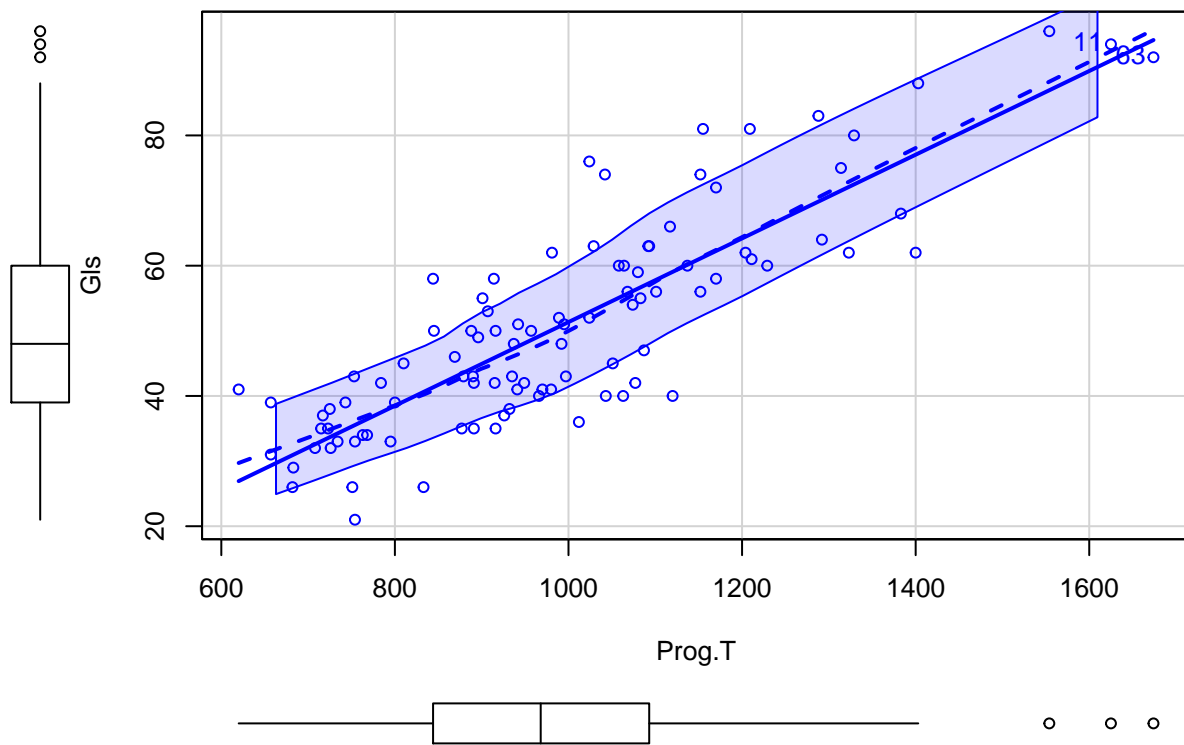
```
## [1] 11 63
```

```
## Warning in applyDefaults(id, defaults = list(method = "mahal", n = 2, cex = 1, :
## unnamed id arguments, will be ignored
```
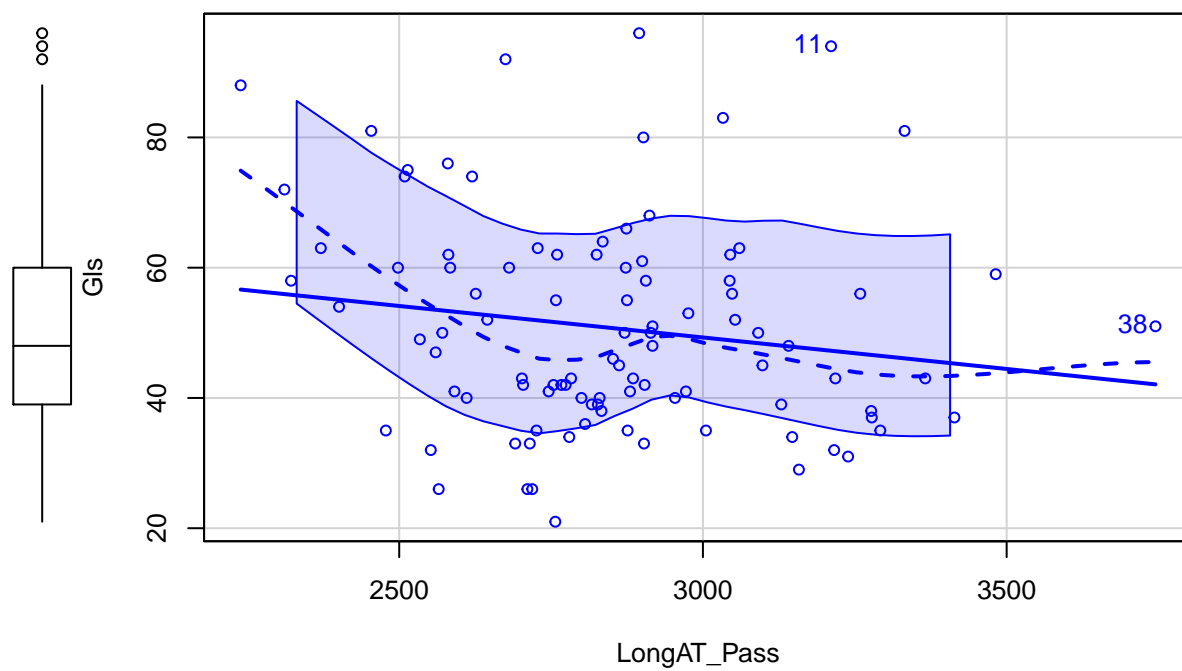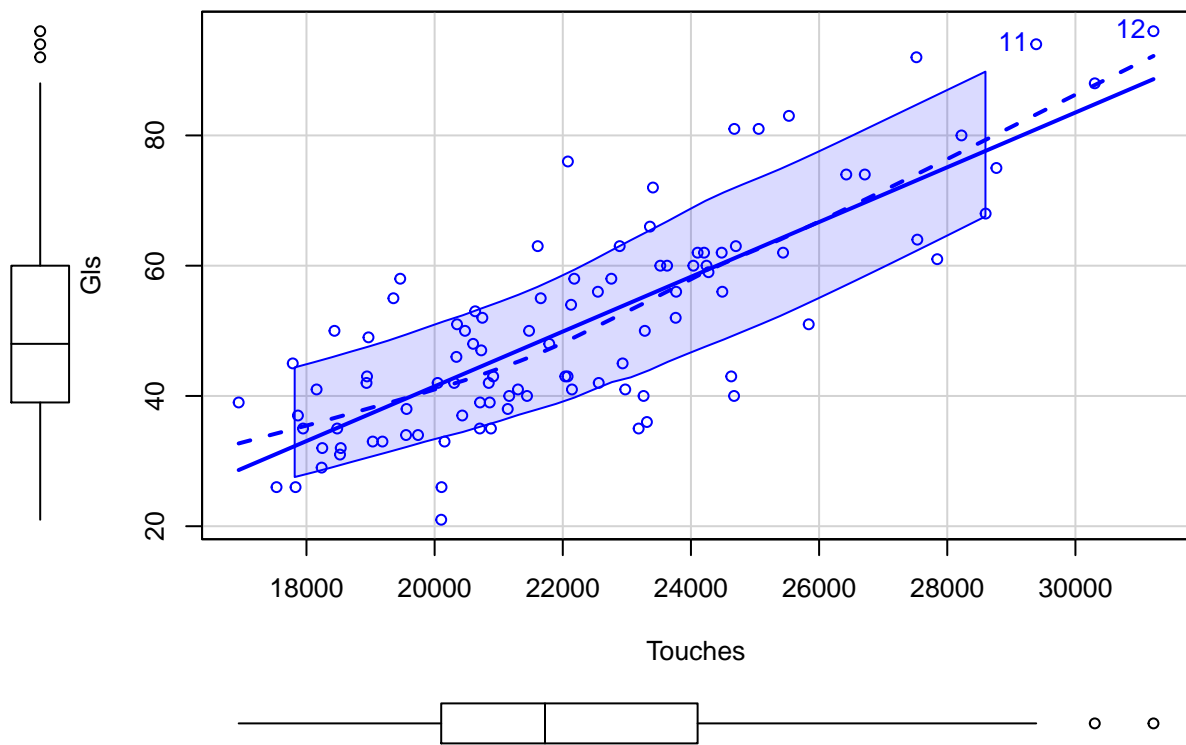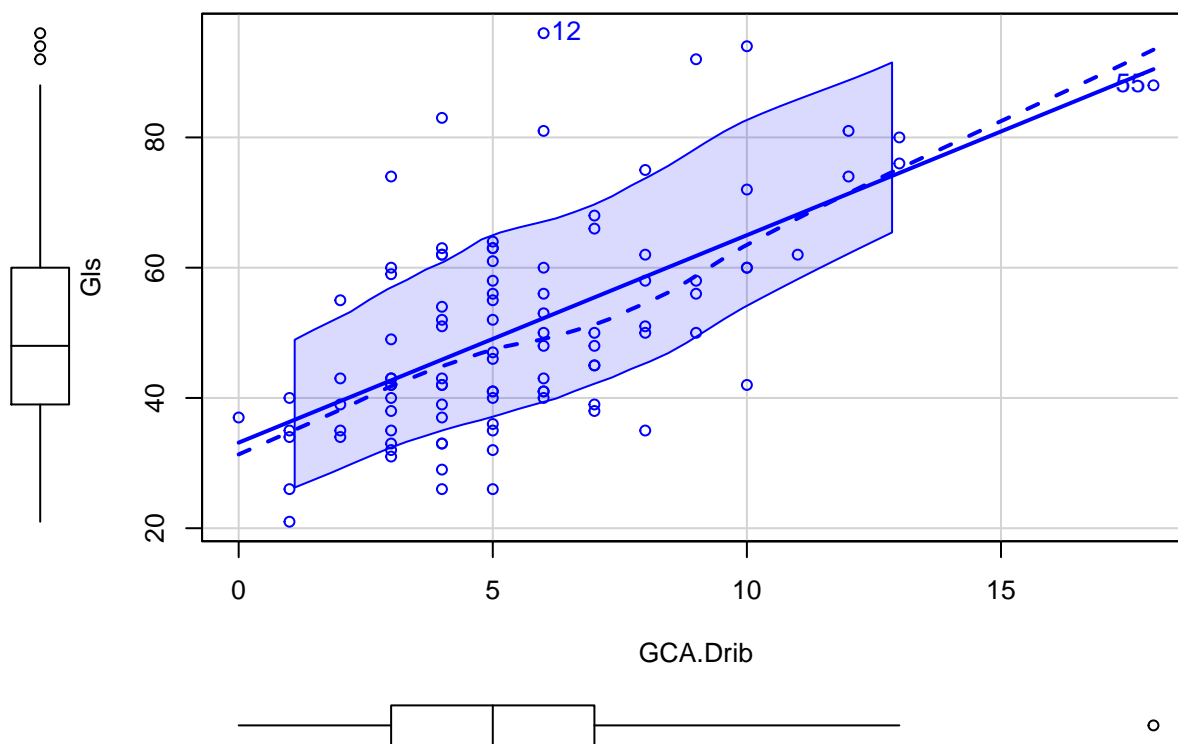
```
## [1] 12 55

## Start:  AIC=360.2
## Gls ~ (leage + team + No.Pl + no.sh + SoT + PK + FK + Att.3rd.T +
##     Succ.Drib + Att.Drib + Touches + Prog.T + GCA.Drib + ShortAT_Pass +
##     MediumtAT_Pass + LongAT_Pass) - team
##
##                   Df Sum of Sq     RSS    AIC
## - leage            4    103.31  2728.0 355.98
## - Succ.Drib        1      0.87  2625.5 358.23
## - No.Pl            1      2.86  2627.5 358.30
## - MediumtAT_Pass   1     21.52  2646.2 359.00
## - Att.Drib         1     22.90  2647.5 359.05
## - Touches          1     27.33  2652.0 359.21
## - FK               1     32.53  2657.2 359.41
## - ShortAT_Pass     1     45.04  2669.7 359.87
## - Prog.T           1     46.26  2670.9 359.91
## <none>                          2624.6 360.20
## - Att.3rd.T        1     55.12  2679.8 360.23
## - LongAT_Pass      1     71.45  2696.1 360.83
## - PK               1    129.84  2754.5 362.93
## - no.sh            1    214.40  2839.0 365.89
## - GCA.Drib         1    242.21  2866.9 366.85
## - SoT              1   1375.65  4000.3 399.50
##
## Step:  AIC=355.98
## Gls ~ No.Pl + no.sh + SoT + PK + FK + Att.3rd.T + Succ.Drib +
```

```
##      Att.Drib + Touches + Prog.T + GCA.Drib + ShortAT_Pass + MediumtAT_Pass +
##      LongAT_Pass
##
##                     Df Sum of Sq     RSS     AIC
## - No.Pl              1       0.77  2728.7  354.01
## - Succ.Drib          1       1.45  2729.4  354.03
## - Att.Drib           1      16.91  2744.9  354.59
## - FK                 1      41.26  2769.2  355.45
## - LongAT_Pass        1      43.92  2771.9  355.55
## - Att.3rd.T          1      45.27  2773.2  355.59
## <none>                            2728.0  355.98
## - Touches            1      63.97  2791.9  356.25
## - ShortAT_Pass       1      69.28  2797.2  356.44
## - MediumtAT_Pass     1      78.47  2806.4  356.76
## - Prog.T             1      86.77  2814.7  357.05
## - PK                 1     102.02  2830.0  357.58
## - no.sh              1     232.15  2960.1  361.99
## - GCA.Drib           1     261.47  2989.4  362.95
## - SoT                1    1461.41  4189.4  396.02
##
## Step:  AIC=354.01
## Gls ~ no.sh + SoT + PK + FK + Att.3rd.T + Succ.Drib + Att.Drib +
##      Touches + Prog.T + GCA.Drib + ShortAT_Pass + MediumtAT_Pass +
##      LongAT_Pass
##
##                     Df Sum of Sq     RSS     AIC
## - Succ.Drib          1       1.32  2730.0  352.06
## - Att.Drib           1      16.80  2745.5  352.61
## - FK                 1      40.94  2769.7  353.47
## - LongAT_Pass        1      44.11  2772.8  353.58
## - Att.3rd.T          1      44.51  2773.2  353.59
## <none>                            2728.7  354.01
## - Touches            1      64.48  2793.2  354.30
## - ShortAT_Pass       1      70.49  2799.2  354.51
## - MediumtAT_Pass     1      78.22  2807.0  354.78
## - Prog.T             1      88.15  2816.9  355.12
## - PK                 1     102.61  2831.3  355.63
## - no.sh              1     232.99  2961.7  360.04
## - GCA.Drib           1     265.66  2994.4  361.11
## - SoT                1    1520.04  4248.8  395.40
##
## Step:  AIC=352.06
## Gls ~ no.sh + SoT + PK + FK + Att.3rd.T + Att.Drib + Touches +
##      Prog.T + GCA.Drib + ShortAT_Pass + MediumtAT_Pass + LongAT_Pass
##
##                     Df Sum of Sq     RSS     AIC
## - FK                 1      40.37  2770.4  351.49
## - LongAT_Pass        1      43.84  2773.9  351.62
## - Att.3rd.T          1      43.92  2774.0  351.62
## <none>                            2730.0  352.06
## - Touches            1      65.43  2795.5  352.38
## - ShortAT_Pass       1      70.04  2800.1  352.54
## - MediumtAT_Pass     1      81.98  2812.0  352.96
## - Prog.T             1      88.11  2818.2  353.17
```

```
## - PK              1    103.48 2833.5 353.70
## - Att.Drib        1    145.06 2875.1 355.13
## - no.sh           1    231.84 2961.9 358.04
## - GCA.Drib        1    271.36 3001.4 359.34
## - SoT             1   1556.10 4286.1 394.26
##
## Step:  AIC=351.49
## Gls ~ no.sh + SoT + PK + Att.3rd.T + Att.Drib + Touches + Prog.T +
##     GCA.Drib + ShortAT_Pass + MediumtAT_Pass + LongAT_Pass
##
##                   Df Sum of Sq    RSS    AIC
## - LongAT_Pass      1     46.00 2816.4 351.11
## - Att.3rd.T        1     50.39 2820.8 351.26
## <none>                         2770.4 351.49
## - Touches          1     62.91 2833.3 351.70
## - ShortAT_Pass     1     63.86 2834.3 351.73
## - MediumtAT_Pass   1     78.50 2848.9 352.23
## - PK               1     81.33 2851.7 352.33
## - Prog.T           1    124.71 2895.1 353.81
## - Att.Drib         1    161.53 2931.9 355.05
## - GCA.Drib         1    257.59 3028.0 358.21
## - no.sh            1    319.63 3090.0 360.20
## - SoT              1   1692.30 4462.7 396.22
##
## Step:  AIC=351.11
## Gls ~ no.sh + SoT + PK + Att.3rd.T + Att.Drib + Touches + Prog.T +
##     GCA.Drib + ShortAT_Pass + MediumtAT_Pass
##
##                   Df Sum of Sq    RSS    AIC
## - ShortAT_Pass     1     18.42 2834.8 349.75
## - Touches          1     19.52 2835.9 349.79
## - MediumtAT_Pass   1     39.83 2856.2 350.48
## <none>                         2816.4 351.11
## - Att.3rd.T        1     78.21 2894.6 351.79
## - PK               1     88.07 2904.5 352.13
## - Prog.T           1     93.98 2910.4 352.33
## - GCA.Drib         1    263.13 3079.5 357.86
## - Att.Drib         1    294.77 3111.2 358.86
## - no.sh            1    335.53 3151.9 360.14
## - SoT              1   1738.74 4555.2 396.23
##
## Step:  AIC=349.75
## Gls ~ no.sh + SoT + PK + Att.3rd.T + Att.Drib + Touches + Prog.T +
##     GCA.Drib + MediumtAT_Pass
##
##                   Df Sum of Sq    RSS    AIC
## - Touches          1      1.49 2836.3 347.80
## - MediumtAT_Pass   1     21.84 2856.7 348.50
## <none>                         2834.8 349.75
## - Att.3rd.T        1     69.94 2904.8 350.14
## - Prog.T           1     92.03 2926.9 350.88
## - PK               1    103.36 2938.2 351.26
## - GCA.Drib         1    250.98 3085.8 356.06
## - Att.Drib         1    311.72 3146.6 357.97
```

27

```
## - no.sh           1    385.08 3219.9 360.23
## - SoT             1   2226.57 5061.4 404.55
##
## Step:  AIC=347.8
## Gls ~ no.sh + SoT + PK + Att.3rd.T + Att.Drib + Prog.T + GCA.Drib +
##     MediumtAT_Pass
##
##                   Df Sum of Sq    RSS    AIC
## - MediumtAT_Pass  1     39.84 2876.2 347.17
## <none>                        2836.3 347.80
## - Prog.T          1     92.57 2928.9 348.95
## - Att.3rd.T       1     93.03 2929.4 348.96
## - PK              1    103.55 2939.9 349.31
## - GCA.Drib        1    249.91 3086.2 354.07
## - Att.Drib        1    336.02 3172.3 356.77
## - no.sh           1    406.79 3243.1 358.93
## - SoT             1   2278.73 5115.1 403.59
##
## Step:  AIC=347.17
## Gls ~ no.sh + SoT + PK + Att.3rd.T + Att.Drib + Prog.T + GCA.Drib
##
##             Df Sum of Sq    RSS    AIC
## <none>                    2876.2 347.17
## - PK         1     99.19 2975.4 348.49
## - Att.3rd.T  1    142.40 3018.6 349.90
## - GCA.Drib   1    245.66 3121.8 353.20
## - Prog.T     1    261.81 3138.0 353.70
## - Att.Drib   1    320.75 3196.9 355.53
## - no.sh      1    468.80 3345.0 359.96
## - SoT        1   2381.42 5257.6 404.28
##
## Call:
## lm(formula = Gls ~ no.sh + SoT + PK + Att.3rd.T + Att.Drib +
##     Prog.T + GCA.Drib, data = data1)
##
## Residuals:
##     Min      1Q   Median      3Q      Max
## -12.9952  -3.2206   0.2331   4.0150  12.8571
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.786480   4.969815  -0.359 0.720088
## no.sh       -0.083715   0.021857  -3.830 0.000237 ***
## SoT          0.455415   0.052756   8.632    2e-13 ***
## PK           0.482263   0.273743   1.762 0.081510 .
## Att.3rd.T    0.002396   0.001135   2.111 0.037550 *
## Att.Drib    -0.023802   0.007513  -3.168 0.002096 **
## Prog.T       0.017892   0.006251   2.862 0.005232 **
## GCA.Drib     0.699625   0.252339   2.773 0.006760 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.653 on 90 degrees of freedom
```

```
## Multiple R-squared:  0.8893, Adjusted R-squared:  0.8807
## F-statistic: 103.3 on 7 and 90 DF,  p-value: < 2.2e-16
```

Distribution of Goals

Boxplot of Goals
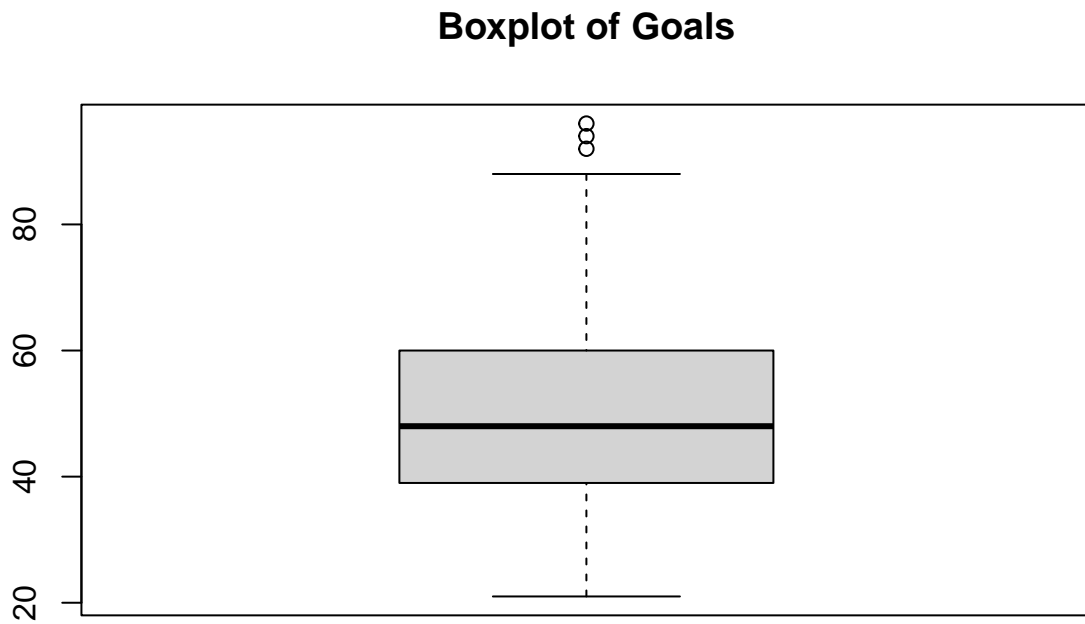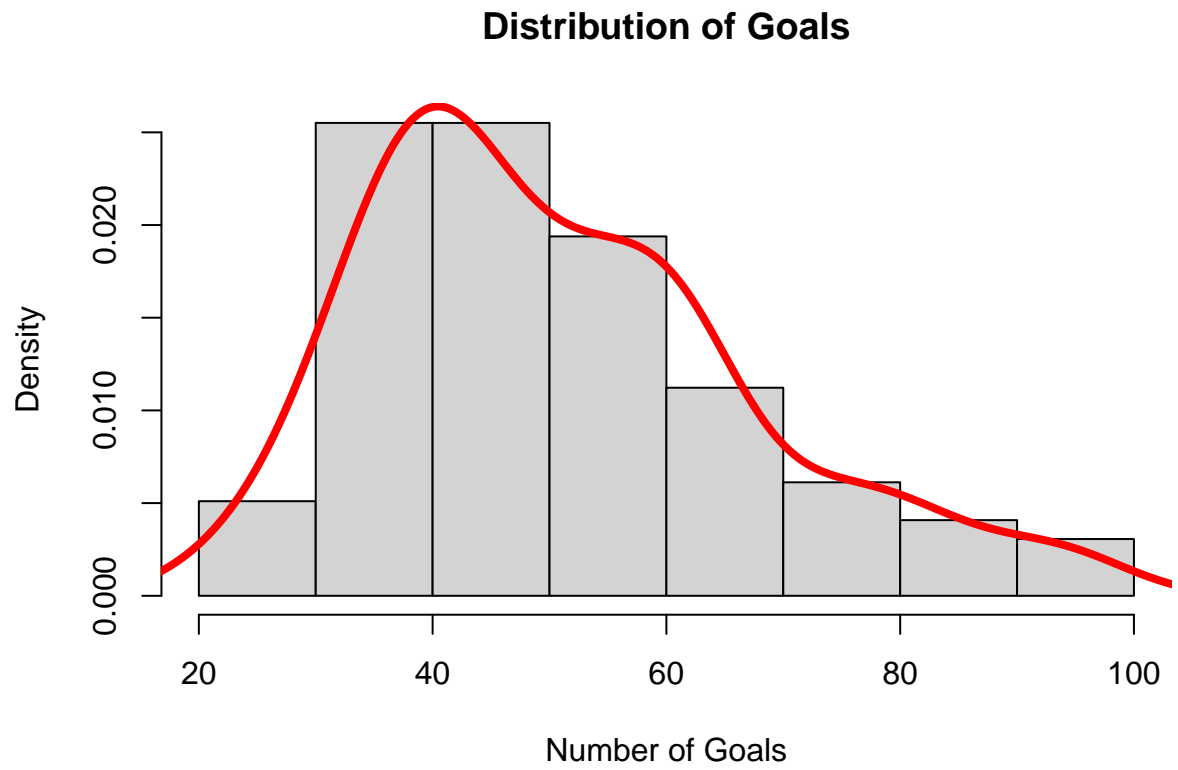
```
## 
## Call:
## glm(formula = Gls ~ . - team, family = poisson(), data = data1)
## 
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.01244  -0.55563   0.00422   0.40505   1.76659
## 
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     3.296e+00  5.452e-01   6.046 1.49e-09 ***
## leageEpl       -3.036e-02  6.677e-02  -0.455  0.64935
## leageLaliga    -5.656e-02  7.197e-02  -0.786  0.43197
## leageLigue1    -9.408e-03  5.897e-02  -0.160  0.87324
## leageSerieA     2.046e-03  5.936e-02   0.034  0.97251
## No.Pl          -3.579e-03  4.606e-03  -0.777  0.43715
## no.sh          -1.366e-03  6.565e-04  -2.081  0.03742 *
## SoT             7.118e-03  1.511e-03   4.710 2.48e-06 ***
## PK              1.150e-02  7.538e-03   1.525  0.12727
## FK             -1.911e-03  2.897e-03  -0.660  0.50945
## Att.3rd.T       2.836e-05  4.267e-05   0.665  0.50629
## Succ.Drib       5.631e-04  8.920e-04   0.631  0.52789
## Att.Drib       -7.153e-04  5.508e-04  -1.299  0.19407
## Touches        -4.415e-05  8.517e-05  -0.518  0.60420
## Prog.T          2.681e-04  2.341e-04   1.145  0.25204
## GCA.Drib        1.620e-02  6.155e-03   2.632  0.00848 **
## ShortAT_Pass    5.289e-05  9.428e-05   0.561  0.57484
## MediumtAT_Pass  5.678e-05  7.699e-05   0.738  0.46081
## LongAT_Pass     6.281e-05  1.213e-04   0.518  0.60471
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for poisson family taken to be 1)
## 
##     Null deviance: 492.772  on 97  degrees of freedom
## Residual deviance:  63.397  on 79  degrees of freedom
## AIC: 661.78
## 
## Number of Fisher Scoring iterations: 4
```

Overdispersion check

```
## Warning: package 'qcc' was built under R version 4.2.2
```

```
## Package 'qcc' version 2.7
```

```
## Type 'citation("qcc")' for citing this R package in publications.
```

```
## 
## Overdispersion test Obs.Var/Theor.Var Statistic p-value
##       poisson data          5.280916  512.2488       0
```

Overdispersion is present, so we try fitting quasipoisson

```
## 
## Call:
## glm(formula = Gls ~ . - team, family = quasipoisson(), data = data1)
```

```
##
## Deviance Residuals:
##      Min       1Q    Median        3Q       Max
## -2.01244  -0.55563   0.00422   0.40505   1.76659
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.296e+00  4.865e-01   6.774 2.01e-09 ***
## leageEpl       -3.036e-02  5.959e-02  -0.509  0.61187
## leageLaliga    -5.656e-02  6.423e-02  -0.880  0.38126
## leageLigue1    -9.408e-03  5.263e-02  -0.179  0.85858
## leageSerieA     2.046e-03  5.298e-02   0.039  0.96930
## No.Pl          -3.579e-03  4.110e-03  -0.871  0.38660
## no.sh          -1.366e-03  5.859e-04  -2.332  0.02226 *
## SoT             7.118e-03  1.349e-03   5.277 1.12e-06 ***
## PK              1.150e-02  6.728e-03   1.709  0.09144 .
## FK             -1.911e-03  2.585e-03  -0.739  0.46200
## Att.3rd.T       2.836e-05  3.808e-05   0.745  0.45867
## Succ.Drib       5.631e-04  7.961e-04   0.707  0.48148
## Att.Drib       -7.153e-04  4.916e-04  -1.455  0.14961
## Touches        -4.415e-05  7.602e-05  -0.581  0.56302
## Prog.T          2.681e-04  2.089e-04   1.283  0.20311
## GCA.Drib        1.620e-02  5.494e-03   2.949  0.00419 **
## ShortAT_Pass    5.289e-05  8.415e-05   0.629  0.53148
## MediumtAT_Pass  5.678e-05  6.872e-05   0.826  0.41109
## LongAT_Pass     6.281e-05  1.083e-04   0.580  0.56356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 0.7965294)
##
##     Null deviance: 492.772  on 97  degrees of freedom
## Residual deviance:  63.397  on 79  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```