

Author: Trey Chase

Driveline Open Biomechanics Project

Description

All data obtained in this report is from the Driveline Open Biomechanics project on Github.

The purpose of this project is to predict fastball velocity from the bio-mechanical data provided, as well as determine the five most important predictors for coaches to emphasize. The model uses a step-wise regression algorithm to determine the most relevant predictors, and then I modified the model to satisfy the multi-collinearity condition as well as removing any variables with coefficients that are not statistically significant via t-test. I then used a 95% bootstrap confidence interval to validate the R-squared statistic of the model. To determine the five most important predictors, I extracted the variables with coefficients that have lowest p-value. An alternative approach would be to use principal component analysis. Leveraging predictive modeling with biomechanical data has potential to increase efficiency when developing pitchers and add a few ticks to their fastball at every level.

```
## Rows: 411 Columns: 81
## -- Column specification -----
## Delimiter: ","
## chr (3): session_pitch, p_throws, pitch_type
## dbl (78): session, pitch_speed_mph, max_shoulder_internal_rotational_velo, m...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
dim(df)
```

```
## [1] 411 81
```

```
head(df)
```

```
## # A tibble: 6 x 81
##   session_pitch session p_throws pitch_type pitch_speed_mph
##   <chr>          <dbl> <chr>    <chr>          <dbl>
## 1 1031_2          1031 R      FF             90.4
## 2 1031_3          1031 R      FF             90.4
## 3 1097_1          1097 R      FF             77.6
## 4 1097_2          1097 R      FF             77
## 5 1097_3          1097 R      FF             76.1
## 6 1170_1          1170 R      FF             80.9
## # i 76 more variables: max_shoulder_internal_rotational_velo <dbl>,
## #   max_elbow_extension_velo <dbl>, max_torso_rotational_velo <dbl>,
## #   max_rotation_hip_shoulder_separation <dbl>, max_elbow_flexion <dbl>,
## #   max_shoulder_external_rotation <dbl>, elbow_flexion_fp <dbl>,
## #   elbow_pronation_fp <dbl>, rotation_hip_shoulder_separation_fp <dbl>,
## #   shoulder_horizontal_abduction_fp <dbl>, shoulder_abduction_fp <dbl>,
## #   shoulder_external_rotation_fp <dbl>, ...
```

```
summary(df)
```

```
## session_pitch      session      p_throws      pitch_type
## Length:411      Min.    :1031  Length:411      Length:411
## Class :character  1st Qu.:1830  Class :character  Class :character
## Mode  :character  Median :2861  Mode  :character  Mode  :character
##                Mean    :2601
##                3rd Qu.:2999
##                Max.    :3252
##
## pitch_speed_mph max_shoulder_internal_rotational_velo max_elbow_extension_velo
## Min.    :69.5      Min.    :2718                      Min.    :1772
## 1st Qu.:81.4      1st Qu.:4326                      1st Qu.:2308
## Median :85.3      Median :4533                      Median :2446
## Mean    :84.7      Mean    :4531                      Mean    :2465
## 3rd Qu.:87.9      3rd Qu.:4746                      3rd Qu.:2611
## Max.    :94.4      Max.    :5413                      Max.    :3100
##
## max_torso_rotational_velo max_rotation_hip_shoulder_separation
## Min.    : 848.4      Min.    :13.16
## 1st Qu.: 997.3      1st Qu.:27.57
## Median :1049.7      Median :32.32
## Mean    :1054.8      Mean    :32.25
## 3rd Qu.:1110.1      3rd Qu.:36.74
## Max.    :1383.2      Max.    :50.51
##
## max_elbow_flexion max_shoulder_external_rotation elbow_flexion_fp
## Min.    : 82.43      Min.    :143.3                      Min.    : 50.33
## 1st Qu.:110.75      1st Qu.:163.5                      1st Qu.: 92.09
## Median :119.27      Median :169.3                      Median :102.41
## Mean    :117.99      Mean    :169.1                      Mean    :103.24
## 3rd Qu.:124.14      3rd Qu.:176.4                      3rd Qu.:115.73
## Max.    :142.38      Max.    :191.5                      Max.    :137.98
##
## elbow_pronation_fp rotation_hip_shoulder_separation_fp
## Min.    : -38.3447    Min.    : 8.404
## 1st Qu.:  0.4375     1st Qu.:25.196
## Median : 12.2514     Median :29.991
## Mean    : 11.3472     Mean    :29.795
## 3rd Qu.: 23.2280     3rd Qu.:34.697
## Max.    : 64.7835     Max.    :48.764
##
## shoulder_horizontal_abduction_fp shoulder_abduction_fp
## Min.    : -6.529      Min.    : 63.37
## 1st Qu.:32.691      1st Qu.: 80.25
## Median :43.955      Median : 85.92
## Mean    :41.964      Mean    : 86.65
## 3rd Qu.:50.900      3rd Qu.: 93.26
## Max.    :71.291      Max.    :118.30
##
## shoulder_external_rotation_fp lead_knee_extension_angular_velo_fp
## Min.    : -17.45      Min.    : -480.273
## 1st Qu.: 33.68      1st Qu.:  5.865
```

```

## Median : 47.24          Median : 121.201
## Mean   : 47.40          Mean    : 127.209
## 3rd Qu.: 61.91          3rd Qu.: 245.734
## Max.   :114.83          Max.    : 632.028
##
## lead_knee_extension_angular_velo_br lead_knee_extension_angular_velo_max
## Min.    :-14.93          Min.     : 77.97
## 1st Qu.:178.42          1st Qu.:261.04
## Median  :267.92          Median   :347.84
## Mean    :288.24          Mean     :373.07
## 3rd Qu.:375.46          3rd Qu.:471.87
## Max.    :954.92          Max.     :955.15
##
## torso_anterior_tilt_fp torso_lateral_tilt_fp torso_rotation_fp
## Min.    :-35.8486        Min.     :-23.2480    Min.     :-34.422
## 1st Qu.: -14.6965        1st Qu.: -9.2243    1st Qu.: -2.703
## Median  : -6.4603        Median   : -4.3166    Median   :  5.535
## Mean    : -7.3104        Mean     : -4.2617    Mean     :  4.813
## 3rd Qu.: -0.2376        3rd Qu.: -0.0921    3rd Qu.: 12.295
## Max.    : 16.7567        Max.     : 16.0880    Max.     : 63.155
##
## pelvis_anterior_tilt_fp pelvis_lateral_tilt_fp pelvis_rotation_fp
## Min.    :-12.289         Min.     :-12.72950   Min.     : 1.877
## 1st Qu.:  2.548          1st Qu.: -3.34955    1st Qu.:27.095
## Median  :  7.563         Median   : -0.11740    Median   :34.314
## Mean    :  7.358         Mean     : -0.09267    Mean     :34.726
## 3rd Qu.: 12.577         3rd Qu.:  2.47725    3rd Qu.:42.092
## Max.    : 24.884         Max.     : 16.87230    Max.     :73.480
##
## max_cog_velo_x torso_rotation_min max_pelvis_rotational_velo
## Min.    :2.274          Min.     :-75.70     Min.     : 505.2
## 1st Qu.:2.911          1st Qu.: -43.63     1st Qu.: 687.6
## Median  :3.075          Median   : -37.54     Median   : 742.1
## Mean    :3.056          Mean     : -38.53     Mean     : 751.5
## 3rd Qu.:3.239          3rd Qu.: -31.14     3rd Qu.: 806.5
## Max.    :3.583          Max.     : -19.09     Max.     :1125.4
##
## glove_shoulders_horizontal_abduction_fp glove_shoulders_abduction_fp
## Min.    :-3.224          Min.     : 42.54
## 1st Qu.:31.587          1st Qu.: 68.15
## Median  :37.559          Median   : 75.00
## Mean    :38.002          Mean     : 74.46
## 3rd Qu.:44.933          3rd Qu.: 81.61
## Max.    :75.990          Max.     :109.83
##
## glove_shoulders_external_rotation_fp glove_shoulders_abduction_mer
## Min.    :-81.81          Min.     :12.28
## 1st Qu.: -52.50          1st Qu.:29.43
## Median  : -40.31          Median   :35.83
## Mean    : -39.41          Mean     :36.09
## 3rd Qu.: -29.93          3rd Qu.:42.61
## Max.    : 30.67          Max.     :72.33
##
## elbow_flexion_mer torso_anterior_tilt_mer torso_lateral_tilt_mer

```

```

## Min. : 66.76 Min. : -9.104 Min. : -2.248
## 1st Qu.: 87.71 1st Qu.: 9.265 1st Qu.: 17.041
## Median : 92.64 Median : 17.538 Median : 22.909
## Mean : 92.17 Mean : 17.262 Mean : 23.329
## 3rd Qu.: 97.02 3rd Qu.: 25.139 3rd Qu.: 30.258
## Max. : 116.95 Max. : 42.742 Max. : 45.517
##
## torso_rotation_mer elbow_varus_moment shoulder_internal_rotation_moment
## Min. : 76.98 Min. : 67.85 Min. : 63.24
## 1st Qu.: 98.55 1st Qu.: 96.90 1st Qu.: 91.47
## Median : 105.56 Median : 110.95 Median : 107.44
## Mean : 104.62 Mean : 111.27 Mean : 106.12
## 3rd Qu.: 112.65 3rd Qu.: 123.17 3rd Qu.: 116.85
## Max. : 132.07 Max. : 199.93 Max. : 181.46
##
## torso_anterior_tilt_br torso_lateral_tilt_br torso_rotation_br
## Min. : 6.018 Min. : -5.933 Min. : 93.69
## 1st Qu.: 27.204 1st Qu.: 11.489 1st Qu.: 113.23
## Median : 34.698 Median : 18.068 Median : 118.53
## Mean : 35.092 Mean : 17.115 Mean : 119.55
## 3rd Qu.: 44.105 3rd Qu.: 23.494 3rd Qu.: 126.31
## Max. : 60.478 Max. : 37.928 Max. : 144.43
##
## lead_knee_extension_from_fp_to_br cog_velo_pkh stride_length
## Min. : -17.71 Min. : 0.0539 Min. : 0.6980
## 1st Qu.: 1.88 1st Qu.: 0.2500 1st Qu.: 0.7986
## Median : 10.02 Median : 0.3214 Median : 0.8374
## Mean : 10.85 Mean : 0.3173 Mean : 0.8354
## 3rd Qu.: 17.52 3rd Qu.: 0.3750 3rd Qu.: 0.8766
## Max. : 55.19 Max. : 0.8571 Max. : 0.9956
##
## stride_angle arm_slot timing_peak_torso_to_peak_pelvis_rot_velo
## Min. : -12.286 Min. : 14.21 Min. : -0.02500
## 1st Qu.: -2.706 1st Qu.: 37.49 1st Qu.: 0.00000
## Median : 2.595 Median : 42.10 Median : 0.00830
## Mean : 1.919 Mean : 41.93 Mean : 0.01214
## 3rd Qu.: 5.928 3rd Qu.: 46.57 3rd Qu.: 0.01670
## Max. : 16.343 Max. : 60.38 Max. : 0.11670
##
## max_shoulder_horizontal_abduction shoulder_transfer_fp_br
## Min. : 15.75 Min. : 170.6
## 1st Qu.: 40.01 1st Qu.: 295.1
## Median : 50.68 Median : 329.8
## Mean : 48.24 Mean : 338.0
## 3rd Qu.: 56.05 3rd Qu.: 385.9
## Max. : 77.30 Max. : 503.3
##
## shoulder_generation_fp_br shoulder_absorption_fp_br elbow_transfer_fp_br
## Min. : 1.139 Min. : 0.00 Min. : 195.9
## 1st Qu.: 21.739 1st Qu.: 11.96 1st Qu.: 297.7
## Median : 32.111 Median : 20.26 Median : 336.6
## Mean : 33.209 Mean : 22.28 Mean : 341.4
## 3rd Qu.: 42.273 3rd Qu.: 30.30 3rd Qu.: 388.7
## Max. : 110.680 Max. : 78.32 Max. : 482.1

```

```

##
## elbow_generation_fp_br elbow_absorption_fp_br lead_hip_transfer_fp_br
## Min. : 0.000 Min. : 13.31 Min. : 5.881
## 1st Qu.: 2.938 1st Qu.: 40.52 1st Qu.: 27.837
## Median : 4.108 Median : 52.31 Median : 42.778
## Mean : 4.683 Mean : 51.90 Mean : 48.542
## 3rd Qu.: 5.764 3rd Qu.: 62.91 3rd Qu.: 67.810
## Max. :29.892 Max. :116.60 Max. :152.273
## NA's :8
## lead_hip_generation_fp_br lead_hip_absorption_fp_br lead_knee_transfer_fp_br
## Min. : 0.000 Min. : 0.8008 Min. : 1.876
## 1st Qu.: 9.179 1st Qu.: 70.5898 1st Qu.: 28.617
## Median : 17.434 Median : 96.8691 Median : 42.865
## Mean : 22.713 Mean : 99.1174 Mean : 44.547
## 3rd Qu.: 29.978 3rd Qu.:125.5535 3rd Qu.: 57.724
## Max. :138.125 Max. :233.1741 Max. :144.053
## NA's :8 NA's :8 NA's :8
## lead_knee_generation_fp_br lead_knee_absorption_fp_br rear_hip_transfer_pkh_fp
## Min. : 0.0981 Min. : 0.00 Min. : 1.677
## 1st Qu.:18.6322 1st Qu.: 11.14 1st Qu.: 74.847
## Median :32.4933 Median : 18.70 Median : 99.057
## Mean :33.4674 Mean : 20.73 Mean : 99.357
## 3rd Qu.:44.3273 3rd Qu.: 26.89 3rd Qu.:123.640
## Max. :92.5344 Max. :110.92 Max. :270.675
## NA's :8 NA's :8 NA's :8
## rear_hip_generation_pkh_fp rear_hip_absorption_pkh_fp
## Min. : 3.929 Min. : 0.2457
## 1st Qu.:117.425 1st Qu.: 32.3544
## Median :152.299 Median : 46.7506
## Mean :149.814 Mean : 51.5963
## 3rd Qu.:185.740 3rd Qu.: 67.3661
## Max. :300.149 Max. :175.7362
## NA's :8 NA's :8
## rear_knee_transfer_pkh_fp rear_knee_generation_pkh_fp
## Min. : 0.00 Min. : 0.0933
## 1st Qu.: 36.48 1st Qu.: 30.1995
## Median : 48.86 Median : 56.3184
## Mean : 52.66 Mean : 58.1526
## 3rd Qu.: 63.72 3rd Qu.: 78.5082
## Max. :161.49 Max. :210.7238
## NA's :8 NA's :8
## rear_knee_absorption_pkh_fp pelvis_lumbar_transfer_fp_br
## Min. : 0.189 Min. : 14.27
## 1st Qu.: 26.458 1st Qu.: 91.00
## Median : 42.712 Median :120.30
## Mean : 44.620 Mean :139.15
## 3rd Qu.: 60.303 3rd Qu.:170.34
## Max. :104.741 Max. :639.76
## NA's :8 NA's :8
## thorax_distal_transfer_fp_br rear_grf_x_max rear_grf_y_max
## Min. :207.9 Min. : 263.2 Min. : 3.239
## 1st Qu.:308.3 1st Qu.: 703.4 1st Qu.: 77.683
## Median :354.4 Median : 795.1 Median :111.171
## Mean :359.7 Mean : 801.8 Mean :121.749

```

```
## 3rd Qu.:405.2          3rd Qu.: 877.7    3rd Qu.:161.773
## Max.      :537.0          Max.      :1238.3    Max.      :367.228
##          NA's      :8          NA's      :8
## rear_grf_z_max rear_grf_mag_max rear_grf_angle_at_max lead_grf_x_max
## Min.      : 712.8    Min.      : 831.9    Min.      :46.75    Min.      : 553.5
## 1st Qu.:1123.1    1st Qu.:1338.2    1st Qu.:56.63    1st Qu.: 919.2
## Median :1282.8    Median :1468.5    Median :58.87    Median :1038.2
## Mean     :1283.0    Mean     :1479.9    Mean     :59.51    Mean     :1055.7
## 3rd Qu.:1395.7    3rd Qu.:1587.7    3rd Qu.:62.63    3rd Qu.:1171.7
## Max.     :1892.6    Max.     :2183.7    Max.     :87.05    Max.     :1821.3
## NA's     :8        NA's     :8        NA's     :8        NA's     :8
## lead_grf_y_max lead_grf_z_max lead_grf_mag_max lead_grf_angle_at_max
## Min.      : 51.17    Min.      :1115    Min.      :1162    Min.      :48.72
## 1st Qu.:147.24    1st Qu.:1629    1st Qu.:1861    1st Qu.:57.62
## Median :205.27    Median :1772    Median :2067    Median :60.17
## Mean     :223.65    Mean     :1818    Mean     :2095    Mean     :60.45
## 3rd Qu.:271.23    3rd Qu.:1946    3rd Qu.:2273    3rd Qu.:63.03
## Max.     :820.48    Max.     :3273    Max.     :3633    Max.     :78.32
## NA's     :8        NA's     :8        NA's     :8        NA's     :8
## peak_rfd_rear peak_rfd_lead
## Min.      : 2.501    Min.      : 14.32
## 1st Qu.: 6.977    1st Qu.: 51.91
## Median : 9.716    Median : 73.93
## Mean     : 720.711    Mean     : 6742.60
## 3rd Qu.: 14.272    3rd Qu.: 116.01
## Max.     :16303.170    Max.     :144366.68
## NA's     :8        NA's     :8
```

```
#obtaining all quantitative variables, dropping all rows with na
q_df = df |>
  select_if(
    is.numeric,
  ) |>
  drop_na() |>
  select(
    -(session) #removing session column
  )
```

```
## [1] 0.6822101
```

```
## [1] "elbow_transfer_fp_br"
```

```
cor_plot = ggcorrplot(corr_matrix)
ggsave("dl_corr_plot.png", plot = cor_plot, width = 30, height = 30)
```

Saving the correlation plot to examine which variables are correlated with pitch speed.

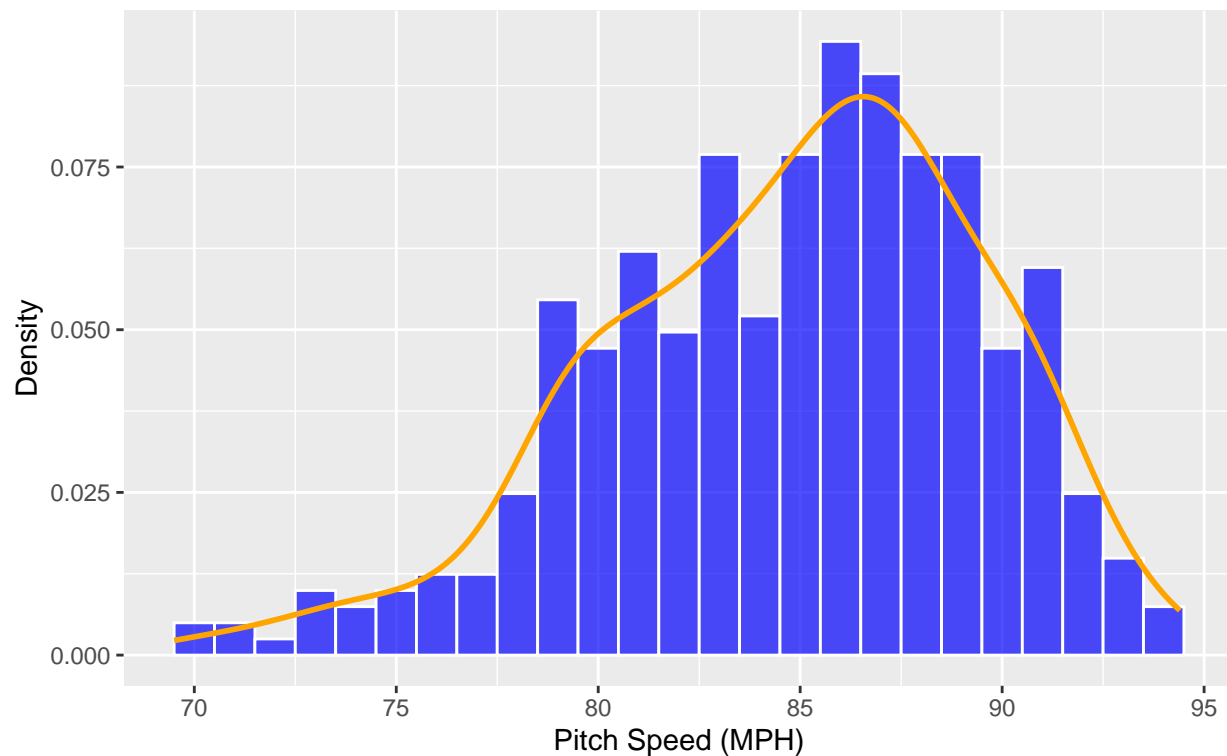
```
pitch_speed_plot = ggplot(q_df, aes(x = pitch_speed_mph)) +
  geom_histogram(aes(y = ..density..), binwidth = 1, color = "white", fill = "blue", alpha = 0.7) +
  geom_density(color = "orange", size = 1) +
  labs(x = "Pitch Speed (MPH)",
       y = "Density",
       title = "Histogram of Pitch Speed with Estimated Density Curve",
       subtitle = "Trey Chase")
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
pitch_speed_plot
```

```
## Warning: The dot-dot notation ('..density..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(density)' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Histogram of Pitch Speed with Estimated Density Curve
Trey Chase



```
ggsave("pitch_speed_plot.png", pitch_speed_plot)
```

```
## Saving 6.5 x 4.5 in image
```

We can see that the distribution of pitch speeds are slightly skewed left but approximately normal.

```
set.seed(123) # For reproducibility
split <- initial_split(q_df, prop = 0.8) #splitting data into test and training data, looking for an 80%
train_df <- training(split)
test_df <- testing(split)
```

This step-wise regression will start by constructing regression models with all variables present, and will iterate and remove one variable based on the lowest AIC of the models.

```
library(car)
```

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

```
##
```

```
## Attaching package: 'car'
```

```
## The following object is masked from 'package:boot':
```

```
##
```

```
##      logit
```

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

```
##
```

```
##      some
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      recode
```

```
vifs = car::vif(selected_mod)
```

```
print(vifs)
```

```
## max_shoulder_internal_rotational_velo      max_torso_rotational_velo
##                2.285484                4.913453
##                max_elbow_flexion      max_shoulder_external_rotation
##                8.020336                4.750964
##                elbow_flexion_fp      elbow_pronation_fp
##                7.500229                2.112016
##                shoulder_abduction_fp      shoulder_external_rotation_fp
##                3.745980                5.228102
## lead_knee_extension_angular_velo_br lead_knee_extension_angular_velo_max
##                5.542250                5.259915
##                torso_anterior_tilt_fp      torso_lateral_tilt_fp
##                7.294458                4.063381
##                torso_rotation_fp      pelvis_anterior_tilt_fp
##                19.919216                2.320841
##                pelvis_rotation_fp      max_cog_velo_x
##                9.159052                2.842627
## glove_shoulder_horizontal_abduction_fp      glove_shoulder_abduction_fp
##                1.913885                2.270621
## glove_shoulder_external_rotation_fp      elbow_flexion_mer
##                2.635181                4.420009
##                torso_lateral_tilt_mer      torso_rotation_mer
##                63.322493                12.151858
##                elbow_varus_moment      shoulder_internal_rotation_moment
##                84.124006                88.902903
##                torso_lateral_tilt_br      torso_rotation_br
##                42.101716                19.335710
```



```
##      lead_knee_extension_from_fp_to_br      cog_velo_pkh
##                5.791639                2.422215
##      shoulder_absorption_fp_br      elbow_transfer_fp_br
##                4.264534                7.313654
##      elbow_generation_fp_br      lead_hip_transfer_fp_br
##                2.127481                5.031992
##      lead_hip_generation_fp_br      lead_hip_absorption_fp_br
##                3.662829                2.964820
##      lead_knee_generation_fp_br      rear_hip_transfer_pkh_fp
##                4.297034                5.275631
##      rear_knee_transfer_pkh_fp      rear_knee_absorption_pkh_fp
##                4.218693                2.661234
##      pelvis_lumbar_transfer_fp_br      rear_grf_z_max
##                2.603382                3.442771
##      lead_grf_y_max      lead_grf_z_max
##                2.434598                159.316201
##      lead_grf_mag_max      lead_grf_angle_at_max
##                148.874592                9.805557
```

Recursively remove VIFs > 4 to ensure the multi-collinearity check is confirmed.

```
signif_all <- names(vifs)

while(any(vifs >= 10)) { # while any of the vifs have a value greater than 10
  var_with_max_vif <- names(which.max(vifs)) # get the var with max vif
  signif_all <- signif_all[!signif_all %in% var_with_max_vif] # remove
  myForm <- as.formula(paste("pitch_speed_mph ~ ", paste(signif_all, collapse=" + "), sep="")) # new f
  selected_mod <- lm(myForm, data=train_df) # re-build model with new formula
  vifs <- car::vif(selected_mod) # update VIFs
}

summary(selected_mod)
```

```
##
## Call:
## lm(formula = myForm, data = train_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.1096 -1.0396  0.1078  1.0884  7.2857
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    12.0132319   7.3157384   1.642  0.101682
## max_shoulder_internal_rotational_velo  0.0008071  0.0004879   1.654  0.099187
## max_torso_rotational_velo    0.0205134  0.0022403   9.156 < 2e-16
## max_elbow_flexion    0.0616693  0.0291114   2.118  0.035017
## max_shoulder_external_rotation  0.1011347  0.0237768   4.254  2.87e-05
## elbow_flexion_fp    -0.0248327  0.0174797  -1.421  0.156522
## elbow_pronation_fp    0.0242014  0.0085995   2.814  0.005233
## shoulder_abduction_fp    0.0265147  0.0203532   1.303  0.193729
## shoulder_external_rotation_fp   -0.0372884  0.0085800  -4.346  1.94e-05
## lead_knee_extension_angular_velo_br  0.0061943  0.0016651   3.720  0.000240
```

```

## lead_knee_extension_angular_velo_max -0.0032685 0.0017547 -1.863 0.063543
## torso_anterior_tilt_fp 0.0318379 0.0218459 1.457 0.146123
## torso_lateral_tilt_fp -0.0670205 0.0283167 -2.367 0.018617
## pelvis_anterior_tilt_fp 0.0567593 0.0227372 2.496 0.013120
## pelvis_rotation_fp 0.1672631 0.0184424 9.069 < 2e-16
## max_cog_velo_x 4.4812459 0.7156129 6.262 1.41e-09
## glove_shoulder_horizontal_abduction_fp -0.0216585 0.0108643 -1.994 0.047164
## glove_shoulder_abduction_fp 0.0163906 0.0155897 1.051 0.293987
## glove_shoulder_external_rotation_fp -0.0278044 0.0093638 -2.969 0.003241
## elbow_flexion_mer -0.0785751 0.0215555 -3.645 0.000318
## torso_rotation_mer -0.1218707 0.0198981 -6.125 3.04e-09
## elbow_varus_moment 0.0384731 0.0126068 3.052 0.002492
## torso_lateral_tilt_br 0.1053713 0.0241265 4.367 1.77e-05
## lead_knee_extension_from_fp_to_br 0.0534942 0.0223974 2.388 0.017579
## cog_velo_pkh -2.2527959 1.7012471 -1.324 0.186508
## shoulder_absorption_fp_br -0.0442418 0.0138573 -3.193 0.001569
## elbow_transfer_fp_br 0.0469873 0.0038665 12.152 < 2e-16
## elbow_generation_fp_br 0.0531901 0.0480555 1.107 0.269303
## lead_hip_transfer_fp_br -0.0180542 0.0091067 -1.983 0.048392
## lead_hip_generation_fp_br -0.0402842 0.0099407 -4.052 6.56e-05
## lead_hip_absorption_fp_br -0.0132945 0.0044685 -2.975 0.003182
## lead_knee_generation_fp_br -0.0051314 0.0115253 -0.445 0.656493
## rear_hip_transfer_pkh_fp -0.0198730 0.0058429 -3.401 0.000768
## rear_knee_transfer_pkh_fp 0.0068513 0.0084418 0.812 0.417711
## rear_knee_absorption_pkh_fp -0.0206865 0.0074977 -2.759 0.006176
## pelvis_lumbar_transfer_fp_br 0.0086582 0.0024337 3.558 0.000439
## rear_grf_z_max -0.0004112 0.0009492 -0.433 0.665192
## lead_grf_y_max -0.0041456 0.0015010 -2.762 0.006124
## lead_grf_mag_max 0.0002301 0.0005435 0.423 0.672370
## lead_grf_angle_at_max 0.1172674 0.0361342 3.245 0.001315
##
## (Intercept)
## max_shoulder_internal_rotational_velo .
## max_torso_rotational_velo ***
## max_elbow_flexion *
## max_shoulder_external_rotation ***
## elbow_flexion_fp
## elbow_pronation_fp **
## shoulder_abduction_fp
## shoulder_external_rotation_fp ***
## lead_knee_extension_angular_velo_br ***
## lead_knee_extension_angular_velo_max .
## torso_anterior_tilt_fp
## torso_lateral_tilt_fp *
## pelvis_anterior_tilt_fp *
## pelvis_rotation_fp ***
## max_cog_velo_x ***
## glove_shoulder_horizontal_abduction_fp *
## glove_shoulder_abduction_fp
## glove_shoulder_external_rotation_fp **
## elbow_flexion_mer ***
## torso_rotation_mer ***
## elbow_varus_moment **
## torso_lateral_tilt_br ***

```

```
## lead_knee_extension_from_fp_to_br      *
## cog_velo_pkh                           **
## shoulder_absorption_fp_br              ***
## elbow_transfer_fp_br                   ***
## elbow_generation_fp_br                 *
## lead_hip_transfer_fp_br                *
## lead_hip_generation_fp_br              ***
## lead_hip_absorption_fp_br              **
## lead_knee_generation_fp_br             *
## rear_hip_transfer_pkh_fp               ***
## rear_knee_transfer_pkh_fp              **
## rear_knee_absorption_pkh_fp            **
## pelvis_lumbar_transfer_fp_br           ***
## rear_grf_z_max                         **
## lead_grf_y_max                         **
## lead_grf_mag_max                       **
## lead_grf_angle_at_max                  **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.003 on 282 degrees of freedom
## Multiple R-squared:  0.8441, Adjusted R-squared:  0.8225
## F-statistic: 39.15 on 39 and 282 DF,  p-value: < 2.2e-16
```

Multicollinearity has been removed. Now we need to remove the variables in the model that are not statistically significant from the t-test.

```
all_vars <- names(selected_mod[[1]])[-1] # names of all X variables
# Get the non-significant vars
summ <- summary(selected_mod) # model summary
pvals <- summ[[4]][, 4] # get all p values
not_significant <- character() # init variables that aren't statistically significant
not_significant <- names(which(pvals > 0.05))
not_significant <- not_significant[!not_significant %in% "(Intercept)"] # remove 'intercept'. Optional

# If there are any non-significant variables,
while(length(not_significant) > 0){
  all_vars <- all_vars[!all_vars %in% not_significant[1]]
  myForm <- as.formula(paste("pitch_speed_mph ~ ", paste (all_vars, collapse=" + "), sep="")) # new formula
  selected_mod <- lm(myForm, data=train_df) # re-build model with new formula

  # Get the non-significant vars.
  summ <- summary(selected_mod)
  pvals <- summ[[4]][, 4]
  not_significant <- character()
  not_significant <- names(which(pvals > 0.1))
  not_significant <- not_significant[!not_significant %in% "(Intercept)"]
}
summary(selected_mod)
```

```
##
## Call:
## lm(formula = myForm, data = train_df)
```

```

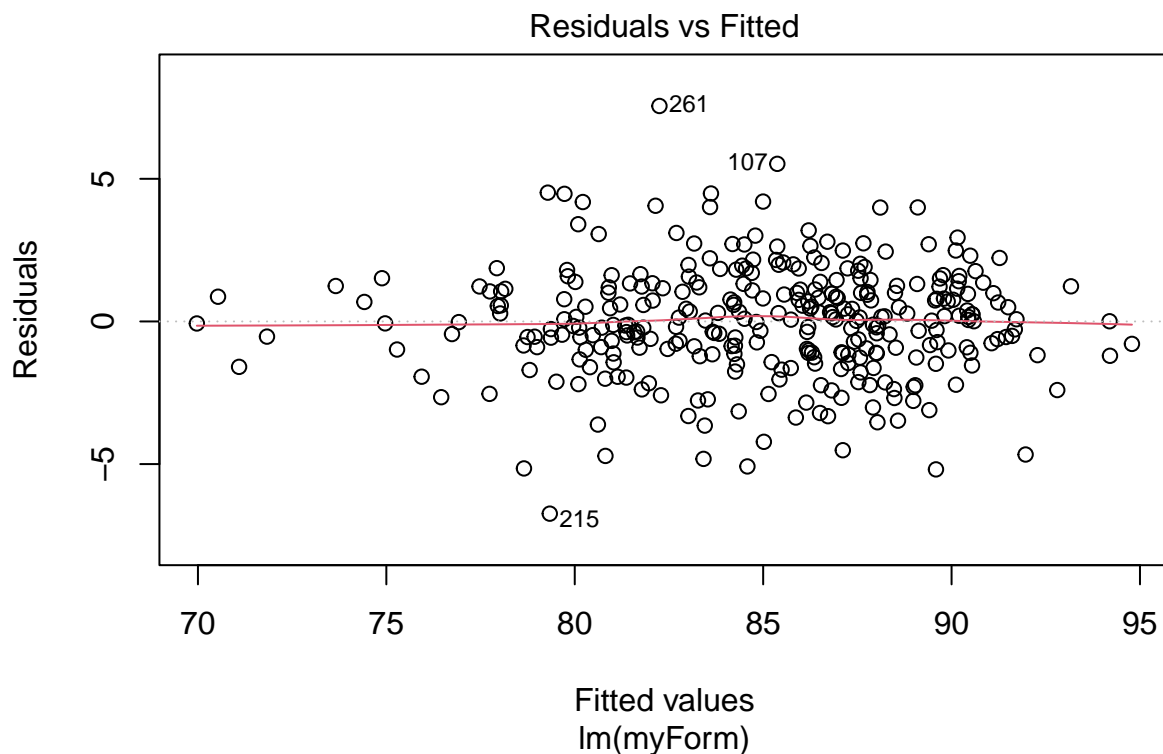
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.7371 -1.0973  0.0581  1.1605  7.5529
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   16.280539    6.134699   2.654 0.008394
## max_torso_rotational_velo      0.021577    0.001832  11.779 < 2e-16
## max_elbow_flexion              0.080054    0.027607   2.900 0.004018
## max_shoulder_external_rotation 0.113679    0.019702   5.770 2.02e-08
## elbow_flexion_fp              -0.041197    0.015181  -2.714 0.007048
## elbow_pronation_fp            0.027486    0.008048   3.415 0.000728
## shoulder_external_rotation_fp -0.038954    0.007089  -5.495 8.54e-08
## lead_knee_extension_angular_velo_br 0.005459    0.001513   3.607 0.000364
## lead_knee_extension_angular_velo_max -0.003993    0.001648  -2.423 0.015980
## torso_lateral_tilt_fp         -0.091676    0.027036  -3.391 0.000793
## pelvis_anterior_tilt_fp        0.062023    0.020879   2.971 0.003218
## pelvis_rotation_fp            0.164132    0.015582  10.534 < 2e-16
## max_cog_velo_x                 4.202985    0.633500   6.635 1.58e-10
## glove_shoulder_horizontal_abduction_fp -0.020555    0.010334  -1.989 0.047622
## glove_shoulder_external_rotation_fp -0.025859    0.008446  -3.062 0.002405
## elbow_flexion_mer              -0.076919    0.019464  -3.952 9.74e-05
## torso_rotation_mer             -0.120593    0.018019  -6.693 1.12e-10
## elbow_varus_moment             0.052714    0.011049   4.771 2.90e-06
## torso_lateral_tilt_br          0.084919    0.016643   5.102 6.06e-07
## lead_knee_extension_from_fp_to_br 0.062547    0.021158   2.956 0.003369
## shoulder_absorption_fp_br      -0.038536    0.010559  -3.649 0.000311
## elbow_transfer_fp_br           0.044617    0.003291  13.559 < 2e-16
## lead_hip_transfer_fp_br        -0.015724    0.007560  -2.080 0.038409
## lead_hip_generation_fp_br      -0.041649    0.008247  -5.050 7.78e-07
## lead_hip_absorption_fp_br      -0.015213    0.004065  -3.743 0.000219
## rear_hip_transfer_pkh_fp       -0.018309    0.003977  -4.604 6.20e-06
## rear_knee_absorption_pkh_fp    -0.017192    0.006802  -2.528 0.012009
## pelvis_lumbar_transfer_fp_br    0.009474    0.002199   4.308 2.25e-05
## lead_grf_y_max                 -0.004798    0.001325  -3.621 0.000345
## lead_grf_angle_at_max          0.099055    0.034547   2.867 0.004442
##
## (Intercept)                   **
## max_torso_rotational_velo      ***
## max_elbow_flexion              **
## max_shoulder_external_rotation ***
## elbow_flexion_fp              **
## elbow_pronation_fp            ***
## shoulder_external_rotation_fp ***
## lead_knee_extension_angular_velo_br ***
## lead_knee_extension_angular_velo_max *
## torso_lateral_tilt_fp         ***
## pelvis_anterior_tilt_fp        **
## pelvis_rotation_fp            ***
## max_cog_velo_x                 ***
## glove_shoulder_horizontal_abduction_fp *
## glove_shoulder_external_rotation_fp **
## elbow_flexion_mer              ***

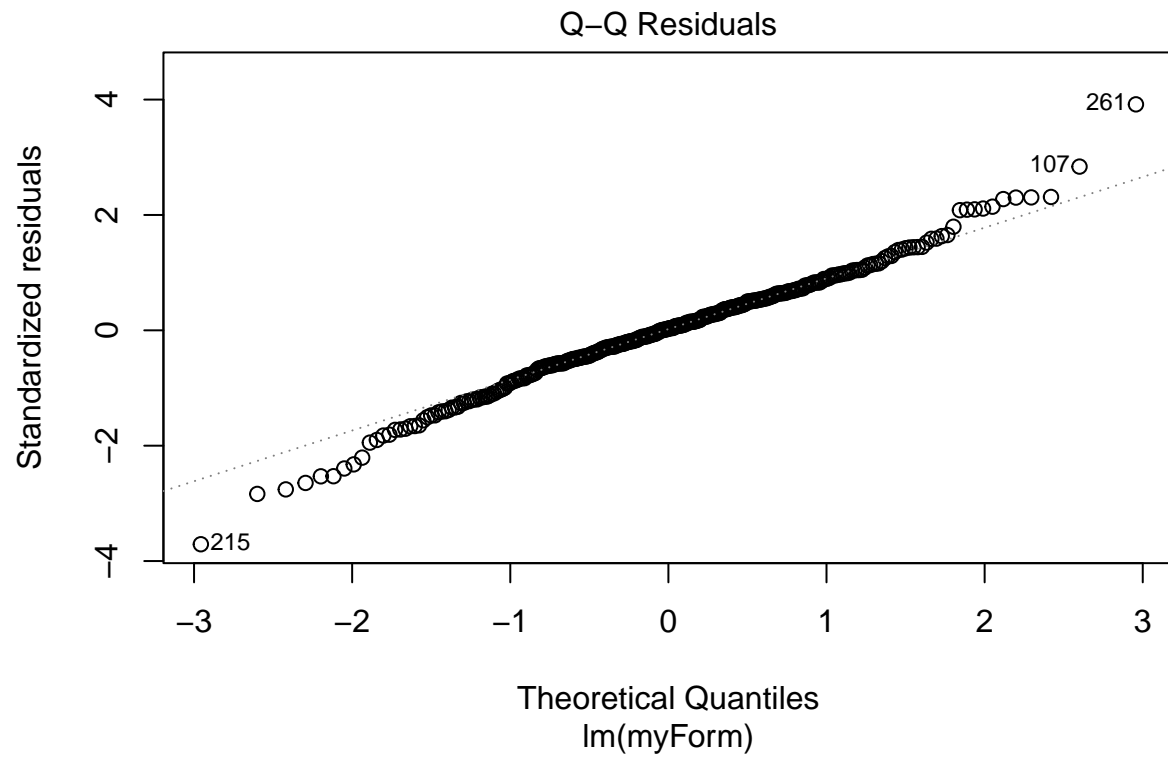
```

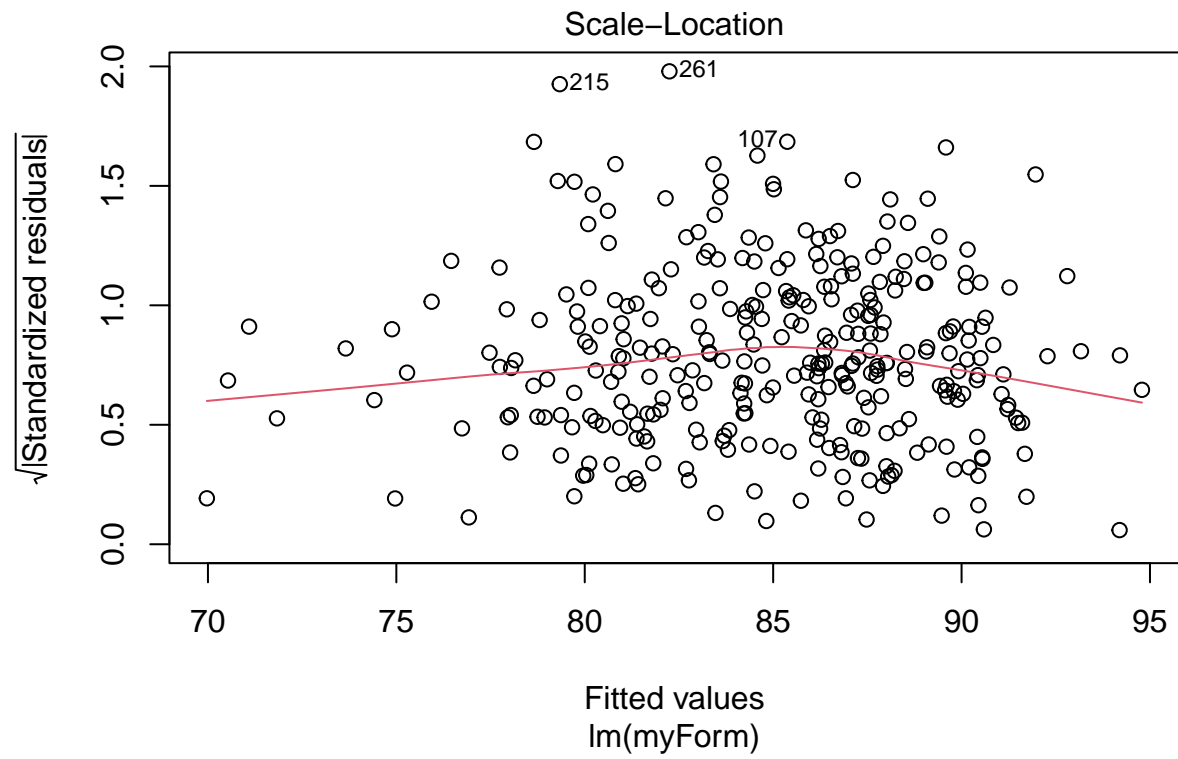
```
## torso_rotation_mer          ***
## elbow_varus_moment          ***
## torso_lateral_tilt_br       ***
## lead_knee_extension_from_fp_to_br **
## shoulder_absorption_fp_br   ***
## elbow_transfer_fp_br        ***
## lead_hip_transfer_fp_br     *
## lead_hip_generation_fp_br   ***
## lead_hip_absorption_fp_br   ***
## rear_hip_transfer_pkh_fp     ***
## rear_knee_absorption_pkh_fp  *
## pelvis_lumbar_transfer_fp_br ***
## lead_grf_y_max               ***
## lead_grf_angle_at_max        **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.008 on 292 degrees of freedom
## Multiple R-squared:  0.8377, Adjusted R-squared:  0.8216
## F-statistic: 51.97 on 29 and 292 DF,  p-value: < 2.2e-16
```

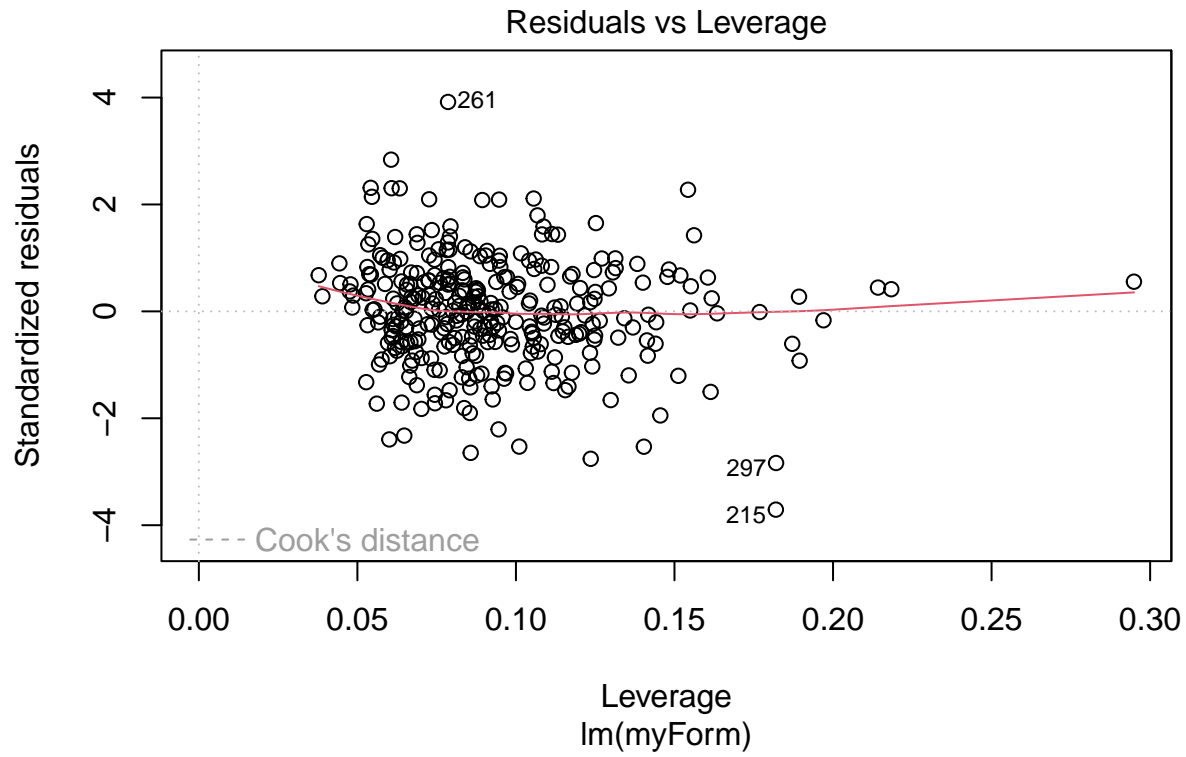
Multi-collinearity and statistical significance ($p < 0.05$) have been confirmed.

```
model_plot = plot(selected_mod)
```









```
model_plot
```

```
## NULL
```

```
ggsave("model_plot.png", model_plot)
```

```
## Saving 6.5 x 4.5 in image
```

```
predict_df = as.data.frame(predict(selected_mod, test_df, level = 0.95, interval = "predict"))
```

```
head(test_df |>
  select(pitch_speed_mph))
```

```
## # A tibble: 6 x 1
##   pitch_speed_mph
##         <dbl>
## 1          90.4
## 2          77.6
## 3          80.9
## 4          84.8
## 5          86.9
## 6          87.4
```



```
head(predict_df)
```

```
##           fit           lwr           upr
## 1 92.84310 88.72892 96.95729
## 2 81.46993 77.26061 85.67925
## 3 84.80432 80.63517 88.97347
## 4 82.43143 78.20383 86.65903
## 5 87.81619 83.66195 91.97043
## 6 88.92070 84.81302 93.02838
```

```
combined_df = bind_cols(predict_df, test_df) #combining the two data sets together by observation number
```

```
combined_df = combined_df |>
  select(c(pitch_speed_mph, fit, upr, lwr)) |>
  rename(
    predicted_mph = fit,
    upper_bound = upr,
    lower_bound = lwr
  )
```

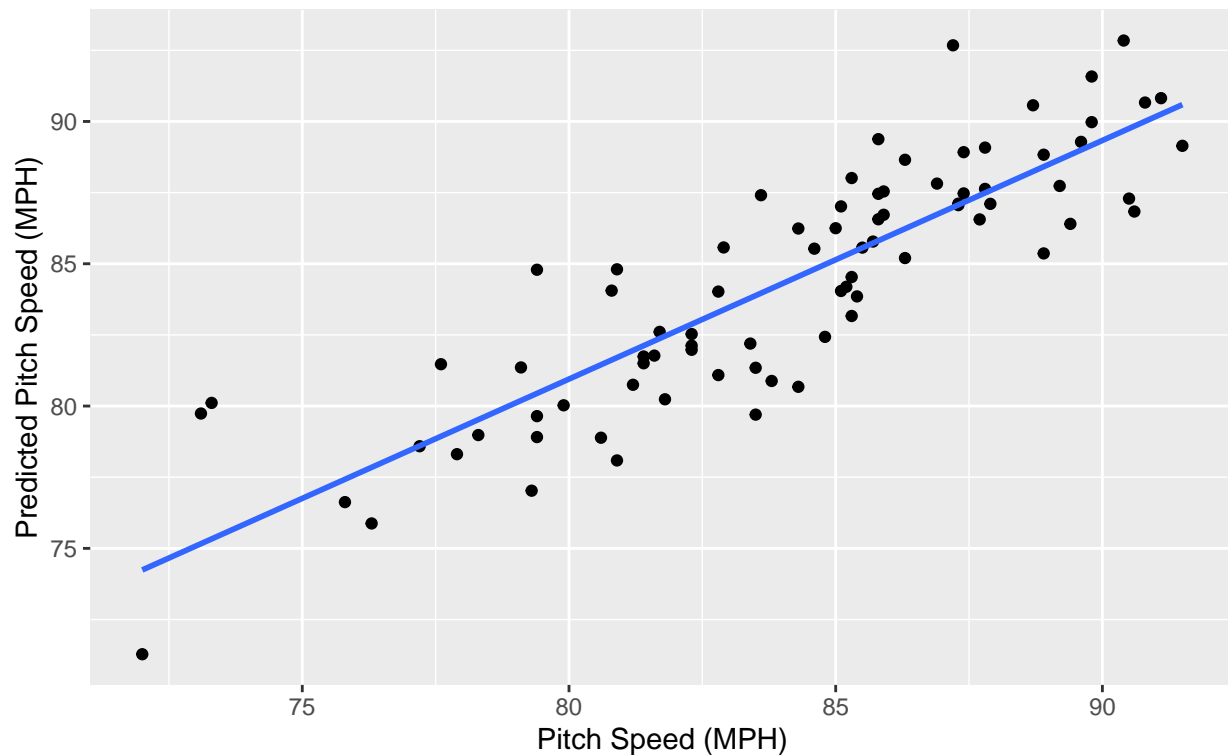
```
write.csv(combined_df, "prediction_data.csv", row.names = FALSE)
```

```
predicted_plot = ggplot(combined_df) +
  geom_point(aes(x = pitch_speed_mph, y = predicted_mph)) +
  geom_smooth(aes(x = pitch_speed_mph, y = predicted_mph), method = "lm", formula = y ~ x, se = FALSE) +
  labs(
    x = "Pitch Speed (MPH)",
    y = "Predicted Pitch Speed (MPH)",
    title = "Pitch Speed Vs. Predicted Pitch Speed",
    subtitle = "Trey Chase"
  )
```

```
predicted_plot
```

Pitch Speed Vs. Predicted Pitch Speed

Trey Chase



```
ggsave("prediced_plot.png", predicted_plot)
```

```
## Saving 6.5 x 4.5 in image
```

```
rsquared_bootstrap <- function(data, indices) {  
  # Re-sample the data  
  resampled_data <- data[indices, ]  
  
  # Fit the model on the resampled data using the original formula  
  boot_model <- lm(formula(selected_mod), data = resampled_data)  
  
  # Return the R-squared value  
  return(summary(boot_model)$r.squared)  
}
```

```
bootstrap = boot(data = q_df, statistic = rsquared_bootstrap, R = 3000) #generating the R^2 bootstrap s
```

```
bootstrap
```

```
##  
## ORDINARY NONPARAMETRIC BOOTSTRAP  
##  
##  
## Call:
```

```
## boot(data = q_df, statistic = rsquared_bootstrap, R = 3000)
##
##
## Bootstrap Statistics :
##      original      bias    std. error
## t1* 0.8248275 0.01283282 0.01715126
```

```
boot.ci(bootstrap, type = "perc")
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 3000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = bootstrap, type = "perc")
##
## Intervals :
## Level      Percentile
## 95%      ( 0.8025,  0.8690 )
## Calculations and Intervals on Original Scale
```

The results of the bootstrapping indicates that the regression model is sound, with an R-squared confidence interval between .8016 and .8619, and a mean R-squared of 0.8248.

```
p_values = summary(selected_mod)$coefficients[, "Pr(>|t|)"]
p_values = p_values[-1] #removing the intercept

top_vars = names(sort(p_values)[1:5]) #obtaining the top 5 significant variables
print(top_vars)
```

```
## [1] "elbow_transfer_fp_br"      "max_torso_rotational_velo"
## [3] "pelvis_rotation_fp"       "torso_rotation_mer"
## [5] "max_cog_velo_x"
```

These are the top 5 variables that are used to predict pitch speed in the selected model. If a pitching coach were to suggest mechanical changes to increase fastball velocity, a good place to start would be considering these five movements. The variables are:

1. Elbow transfer
2. Max torso rotational velocity
3. Pelvis rotation
4. Torso rotation
5. Max cog velocity

– Interpretation of Variables–

– From Driveline open bio data key –

1. Energy transfer across throwing elbow between foot plant and ball release
2. Max trunk axial rotation angle at ball release
3. Pelvis axial rotation at foot plant
4. Trunk axial rotation at maximum external rotation
5. Peak center of gravity velocity towards home plate