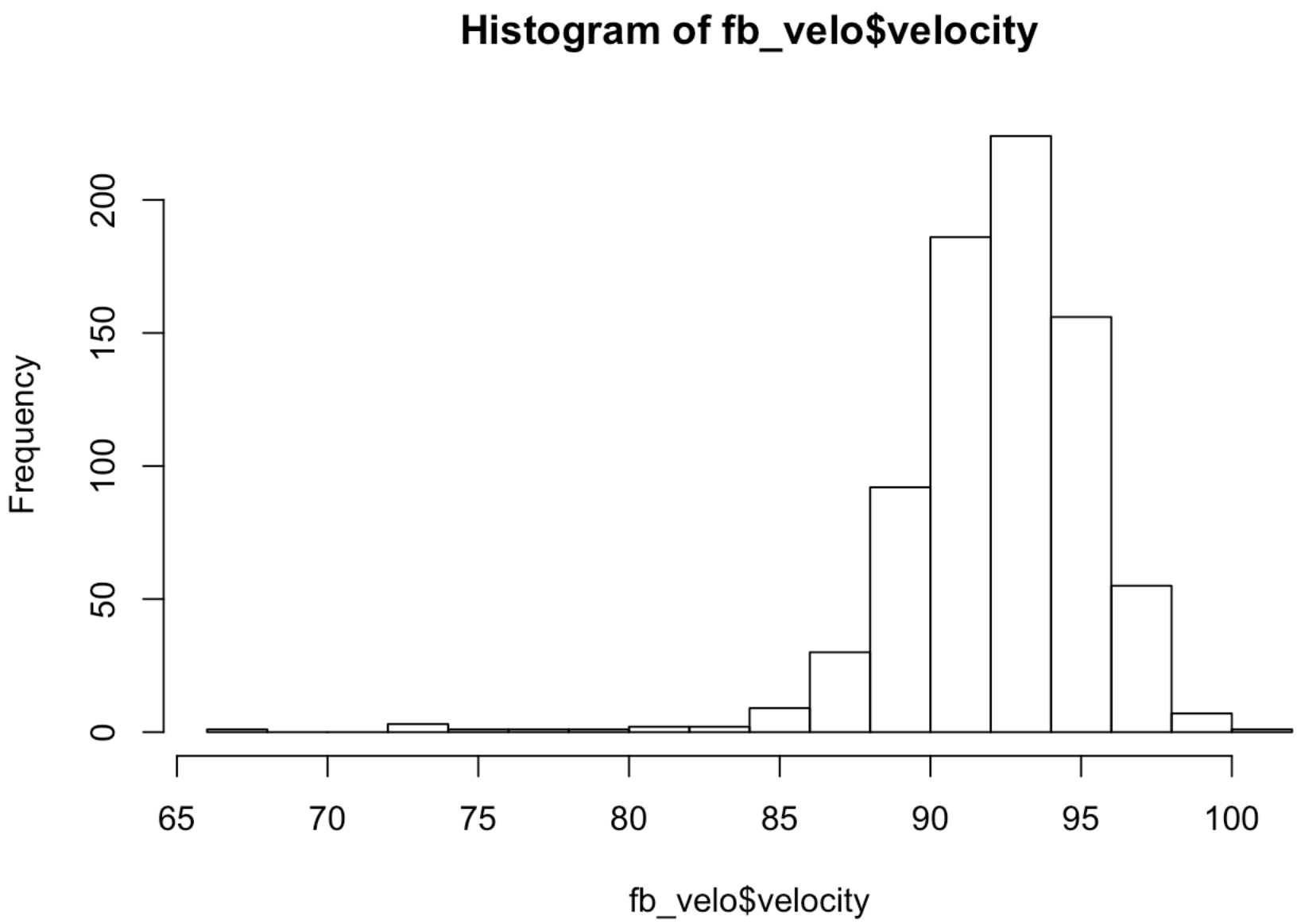
How Does Velocity Affect Fastball Success?

Luke Stageberg

March 25, 2019

```
fb_velo = read.csv('../data/Fastball_Data.csv')
#Naming the dataset "fb_velo" and importing it from my data folder in my Fastball_Vel
ocity folder
```

```
hist(fb_velo$velocity, breaks = 20)
```




```
#Creating a histogram to show the viewer what the average fastball velocity's were ac
ross the MLB in 2018
```

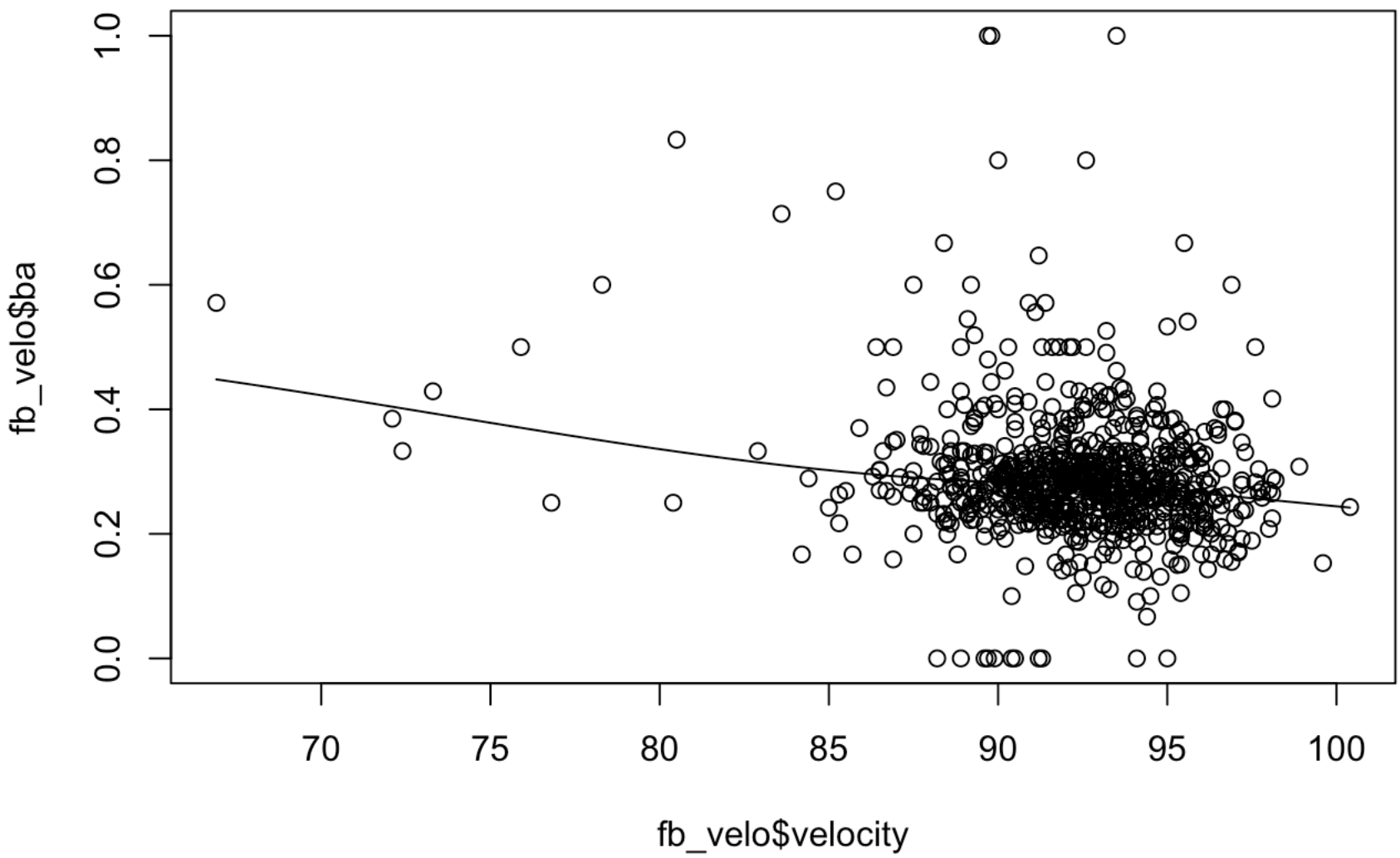
```
mod_velo_ba = lm(formula = ba ~ velocity, data = fb_velo)
#Creating a model using velocity as a predictor varibale for batting average
```

```
summary(mod_velo_ba)
```

```
##
## Call:
## lm(formula = ba ~ velocity, data = fb_velo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31658 -0.05239 -0.01085  0.02828  0.71671
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.870538   0.105564   8.247 7.04e-16 ***
## velocity    -0.006281   0.001143  -5.495 5.33e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1039 on 769 degrees of freedom
## Multiple R-squared:  0.03778,    Adjusted R-squared:  0.03653
## F-statistic: 30.19 on 1 and 769 DF,  p-value: 5.326e-08
```

```
#Producing a summary of the model from above (velocity predicting batting average)
```

```
scatter.smooth(x=fb_velo$velocity, y=fb_velo$ba, main="ba ~ velocity")
```




```
#Creating a scatter plot with a line of best fit as a visual alongside the model (sho
wing the relationship between batting average and average fastball velocity)
```

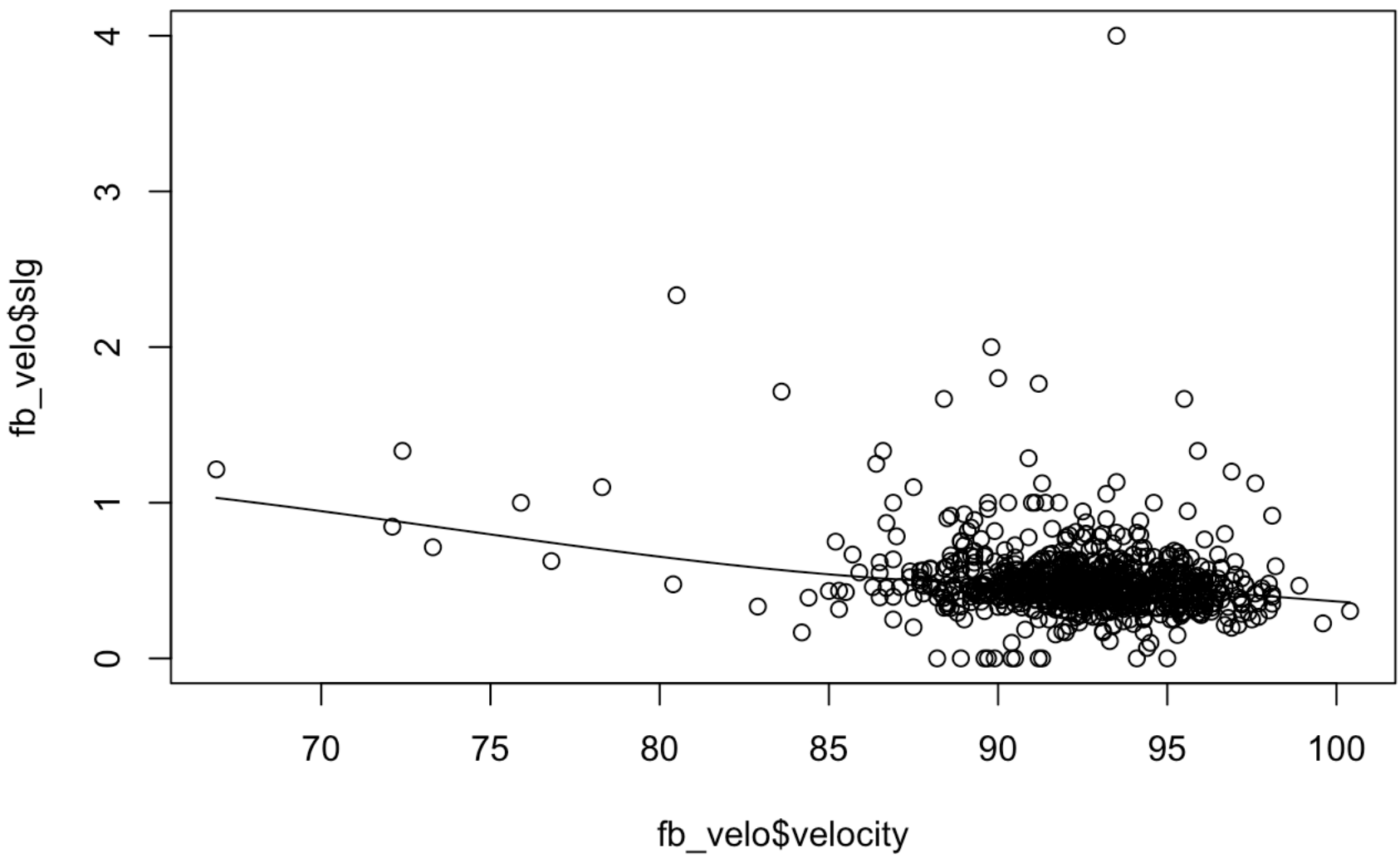
```
mod_velo_slg = lm(formula = slg ~ velocity, data = fb_velo)
#Creating a model using velocity as a predictor varibale for slugging percentage
```

```
summary(mod_velo_slg)
```

```
##
## Call:
## lm(formula = slg ~ velocity, data = fb_velo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5719 -0.1188 -0.0373  0.0641  3.5220
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.134433   0.257532   8.288 5.11e-16 ***
## velocity    -0.017716   0.002789  -6.353 3.61e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2535 on 769 degrees of freedom
## Multiple R-squared:  0.04987,    Adjusted R-squared:  0.04863
## F-statistic: 40.36 on 1 and 769 DF,  p-value: 3.614e-10
```

```
#Producing a summary of the model from above (velocity predicting slugging percentage
)
```

```
scatter.smooth(x=fb_velo$velocity, y=fb_velo$slg, main="slg ~ velocity")
```




```
#Creating a scatter plot with a line of best fit as a visual alongside the model (sho
wing the relationship between slugging percentage and average fastball velocity)
```

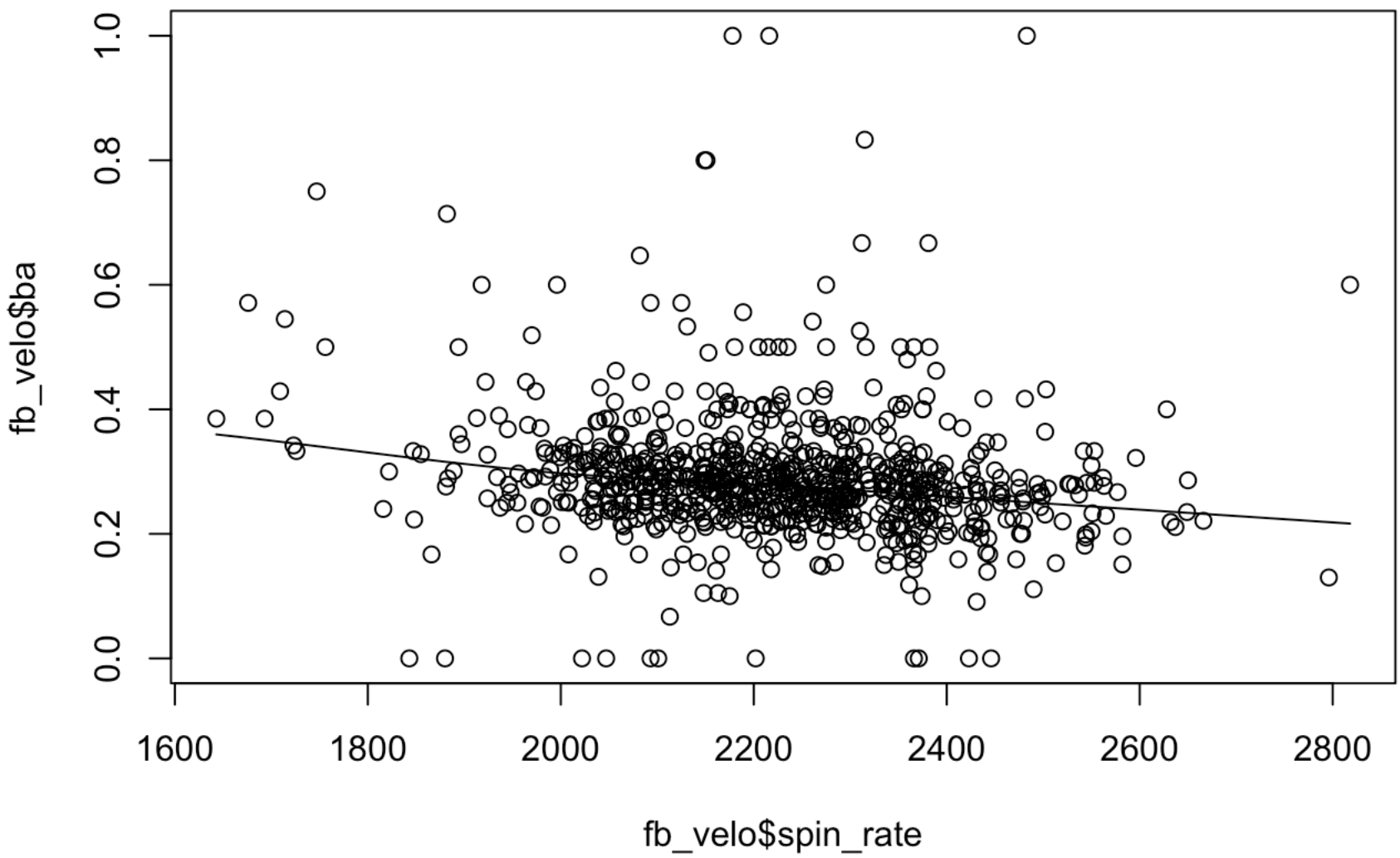
```
mod_spin_ba = lm(formula = ba ~ spin_rate, data = fb_velo)
#Creating a model using spin rate as a predictor varibale for batting average
```

```
summary(mod_spin_ba)
```

```
##
## Call:
## lm(formula = ba ~ spin_rate, data = fb_velo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32858 -0.05143 -0.01143  0.02646  0.73491
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.114e-01  5.041e-02  10.146  < 2e-16 ***
## spin_rate   -9.922e-05  2.261e-05  -4.388  1.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1046 on 769 degrees of freedom
## Multiple R-squared:  0.02443,    Adjusted R-squared:  0.02316
## F-statistic: 19.26 on 1 and 769 DF,  p-value: 1.302e-05
```

```
#Producing a summary of the model from above (spin rate predicting batting average)
```

```
scatter.smooth(x=fb_velo$spin_rate, y=fb_velo$ba, main="ba ~ spin_rate")
```




```
#Creating a scatter plot with a line of best fit as a visual alongside the model (sho
wing the relationship between batting average and average fastball spin rate)
```

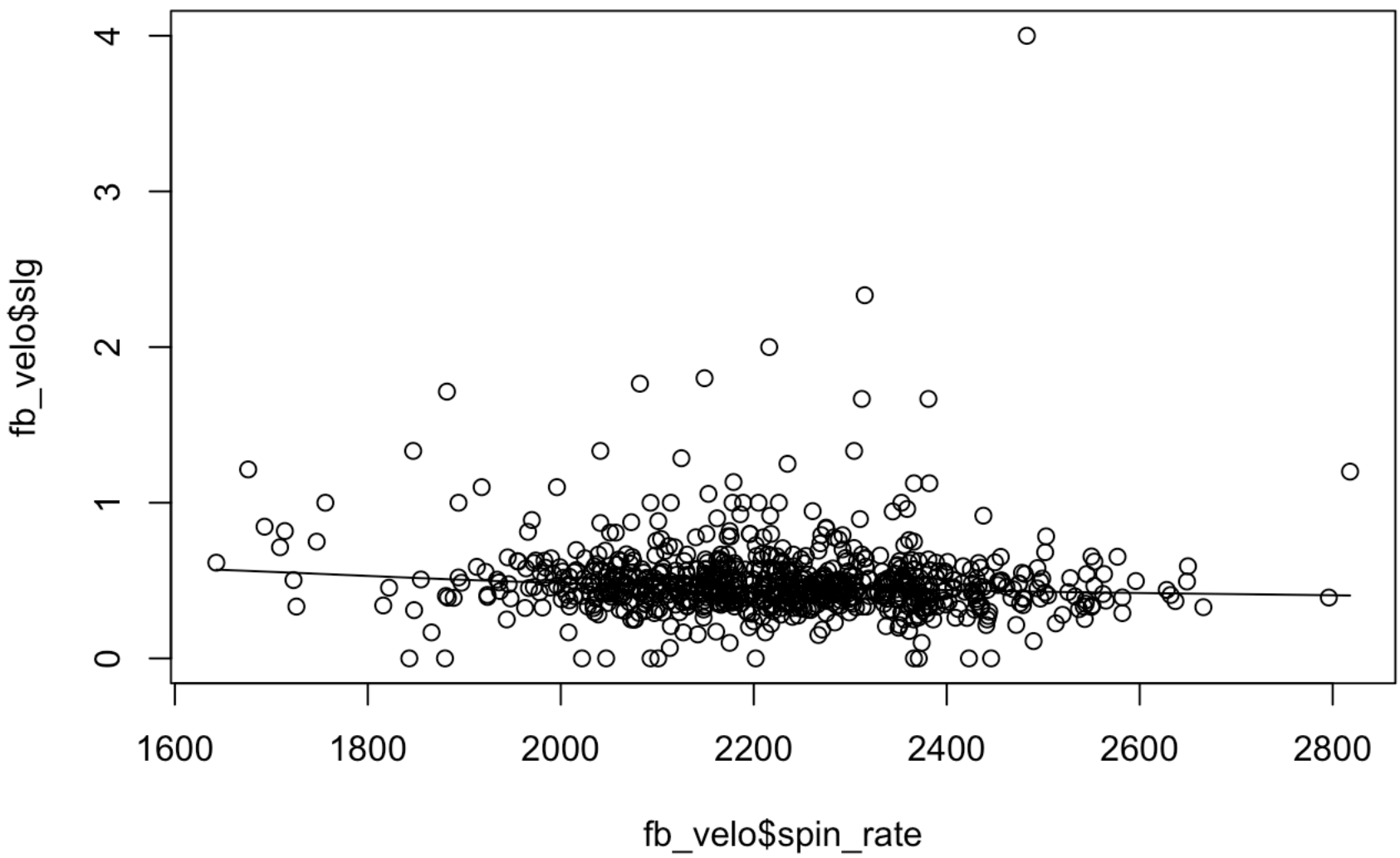
```
mod_spin_slg = lm(formula = slg ~ spin_rate, data = fb_velo)
#Creating a model using spin rate as a predictor varibale for sluggling percentage
```

```
summary(mod_spin_slg)
```

```
##
## Call:
## lm(formula = slg ~ spin_rate, data = fb_velo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5408 -0.1192 -0.0406  0.0610  3.5290
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.419e-01  1.250e-01   5.936 4.42e-09 ***
## spin_rate   -1.091e-04  5.606e-05  -1.946    0.052 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2594 on 769 degrees of freedom
## Multiple R-squared:  0.0049, Adjusted R-squared:  0.003606
## F-statistic: 3.787 on 1 and 769 DF,  p-value: 0.05202
```

```
#Producing a summary of the model from above (spin rate predicting sluggling percenta
ge)
```

```
scatter.smooth(x=fb_velo$spin_rate, y=fb_velo$slg, main="slg ~ spin_rate")
```




```
#Creating a scatter plot with a line of best fit as a visual alongside the model (sho
wing the relationship between slugging percentage and average fastball spin rate)
```

```
fb_velo$swing_and_miss_pct = with(fb_velo, whiffs / swings)
fb_velo[is.na(fb_velo)] <- 0
#Creating a new varible that is the percentage of a swing and miss (dividing whiffs b
y swings)
#Replacing any "na" in the data with "0"
```

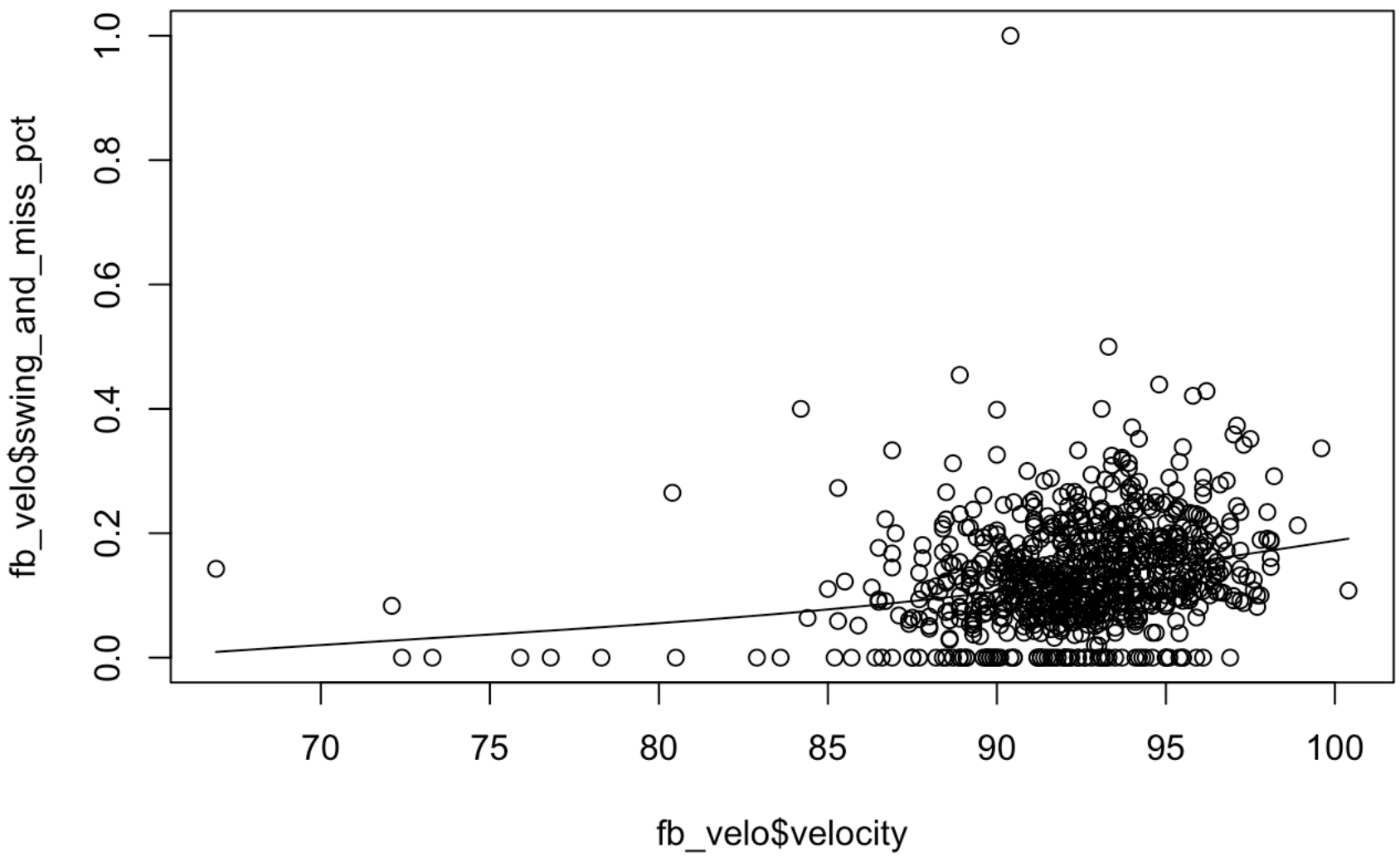
```
mod_velo_swing_and_miss =  lm(formula = swing_and_miss_pct ~ velocity, data = fb_velo
)
#Creating a model using velocity as a predictor varibale for swing and miss percentag
e
```

```
summary(mod_velo_swing_and_miss)
```

```
##
## Call:
## lm(formula = swing_and_miss_pct ~ velocity, data = fb_velo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16230 -0.05512 -0.00920  0.04267  0.87870
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.4488132  0.0889291  -5.047 5.61e-07 ***
## velocity     0.0063066  0.0009629   6.549 1.06e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08753 on 769 degrees of freedom
## Multiple R-squared:  0.05283,    Adjusted R-squared:  0.0516
## F-statistic: 42.89 on 1 and 769 DF,  p-value: 1.057e-10
```

```
#Producing a summary of the model from above (velocity predicting swing and miss perc
entage)
```

```
scatter.smooth(x=fb_velo$velocity, y=fb_velo$swing_and_miss_pct, main="swing_and_miss
_pct ~ velocity")
```




```
#Creating a scatter plot with a line of best fit as a visual alongside the model (sho
wing the relationship between swing and miss percentage and average fastball velocity
)
```

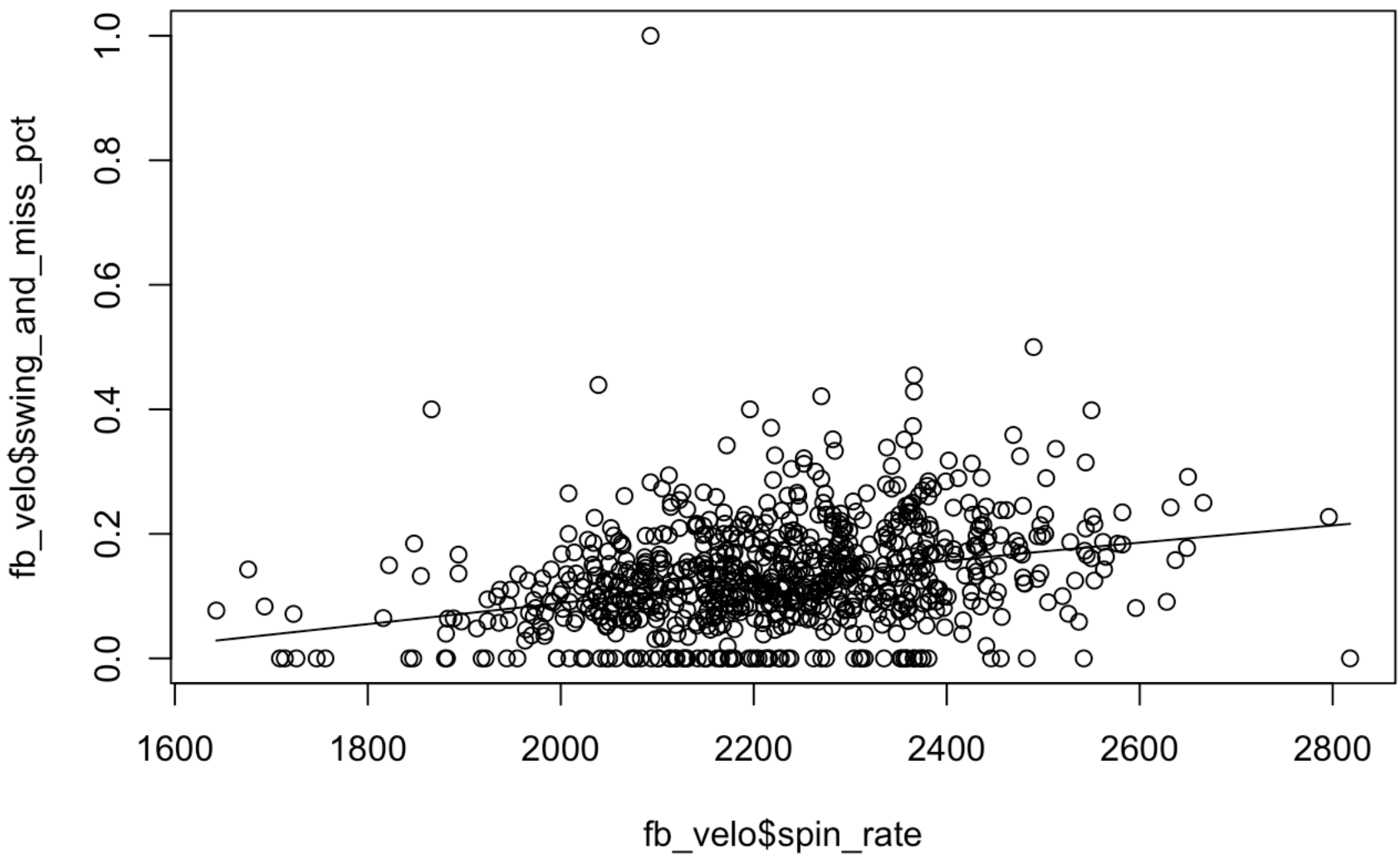
```
mod_spin_swing_and_miss =  lm(formula = swing_and_miss_pct ~ spin_rate, data = fb_vel
o)
#Creating a model using spin rate as a predictor varibale for swing and miss percenta
ge
```

```
summary(mod_spin_swing_and_miss)
```

```
##
## Call:
## lm(formula = swing_and_miss_pct ~ spin_rate, data = fb_velo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22513 -0.05061 -0.00761  0.04293  0.88686
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.102e-01  4.151e-02  -5.062 5.19e-07 ***
## spin_rate    1.545e-04  1.862e-05   8.295 4.84e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08616 on 769 degrees of freedom
## Multiple R-squared:  0.08213,    Adjusted R-squared:  0.08094
## F-statistic: 68.81 on 1 and 769 DF,  p-value: 4.835e-16
```

```
#Producing a summary of the model from above (spin rate predicting swing and miss per
centage)
```

```
scatter.smooth(x=fb_velo$spin_rate, y=fb_velo$swing_and_miss_pct, main="swing_and_mis
s_pct ~ spin_rate")
```



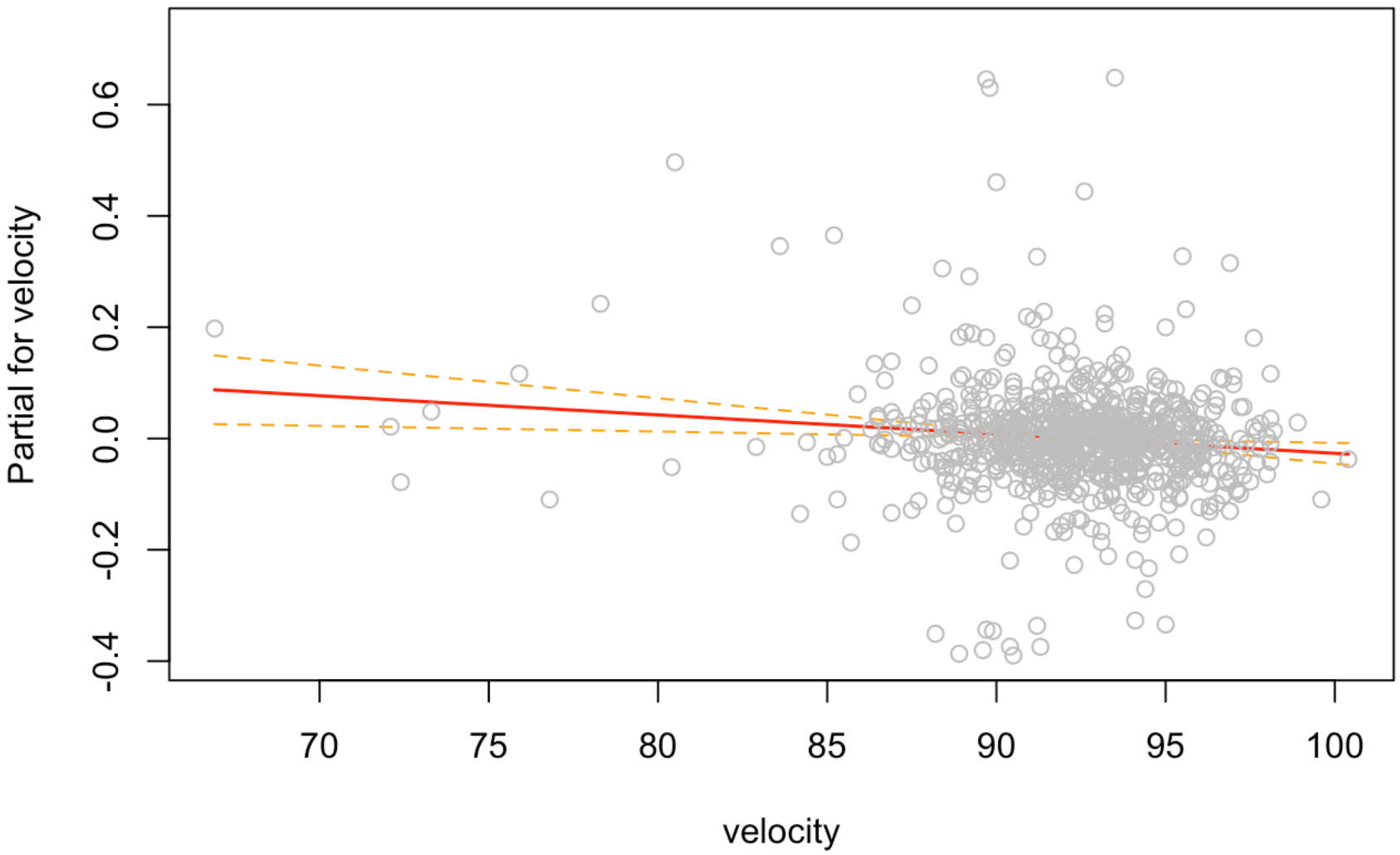

```
#Creating a scatter plot with a line of best fit as a visual alongside the model (sho
wing the relationship between swing and miss percentage and average fastball spin rat
e)
```

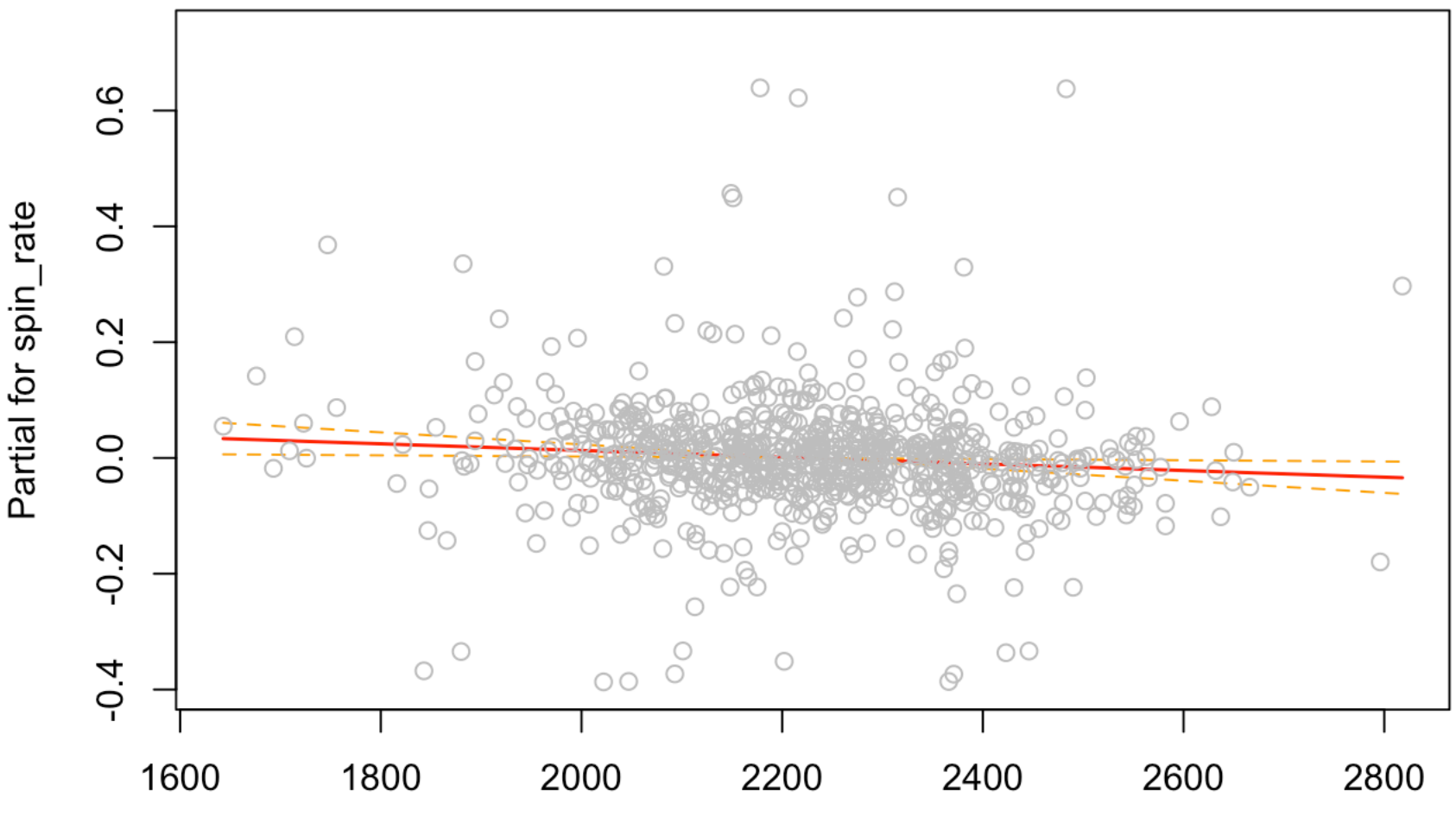
```
mod_velo_spin_ba = lm(formula = ba ~ velocity + spin_rate + I(log(total_pitches)), da
ta = fb_velo)
#Creating a model using velocity, spin rate, and total pitches as a predictor varibal
e for batting average
```

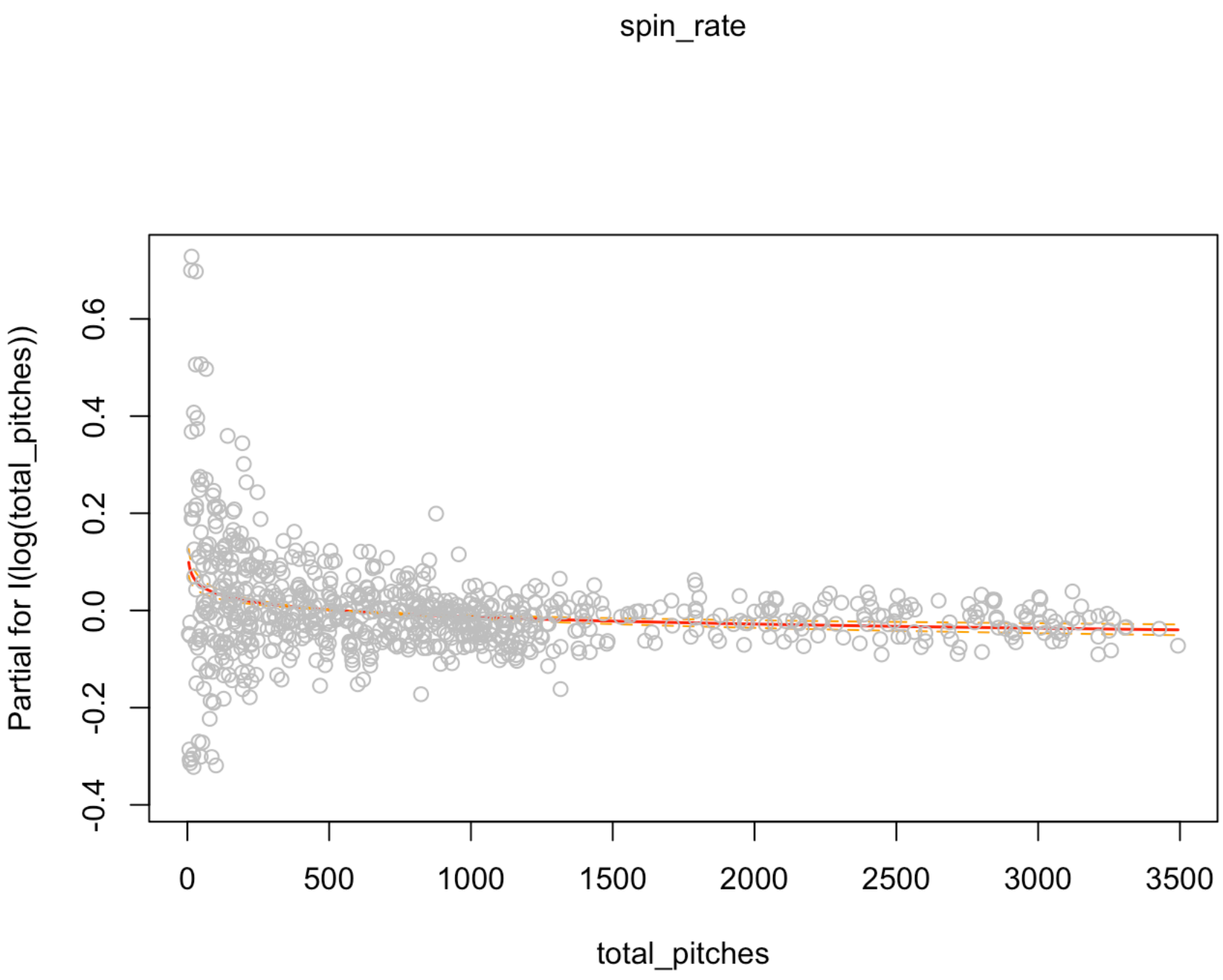
```
summary(mod_velo_spin_ba)
```

```
##
## Call:
## lm(formula = ba ~ velocity + spin_rate + I(log(total_pitches)),
##     data = fb_velo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39841 -0.04228  0.00004  0.03505  0.65259
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.703e-01  1.025e-01   8.491  < 2e-16 ***
## velocity               -3.449e-03  1.215e-03  -2.838  0.00466 **
## spin_rate              -5.758e-05  2.345e-05  -2.456  0.01428 *
## I(log(total_pitches)) -2.118e-02  2.959e-03  -7.157 1.93e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1003 on 767 degrees of freedom
## Multiple R-squared:  0.1054, Adjusted R-squared:  0.1019
## F-statistic: 30.14 on 3 and 767 DF,  p-value: < 2.2e-16
```

```
termplot(mod_velo_spin_ba,partial.resid = T, se = T)
```


Partial for velocity
0.6
0.4
0.2
0.0
-0.2
-0.4
70
75
80
85
90
95
100
velocity


Partial for spin_rate
0.6
0.4
0.2
0.0
-0.2
-0.4
1600
1800
2000
2200
2400
2600
2800



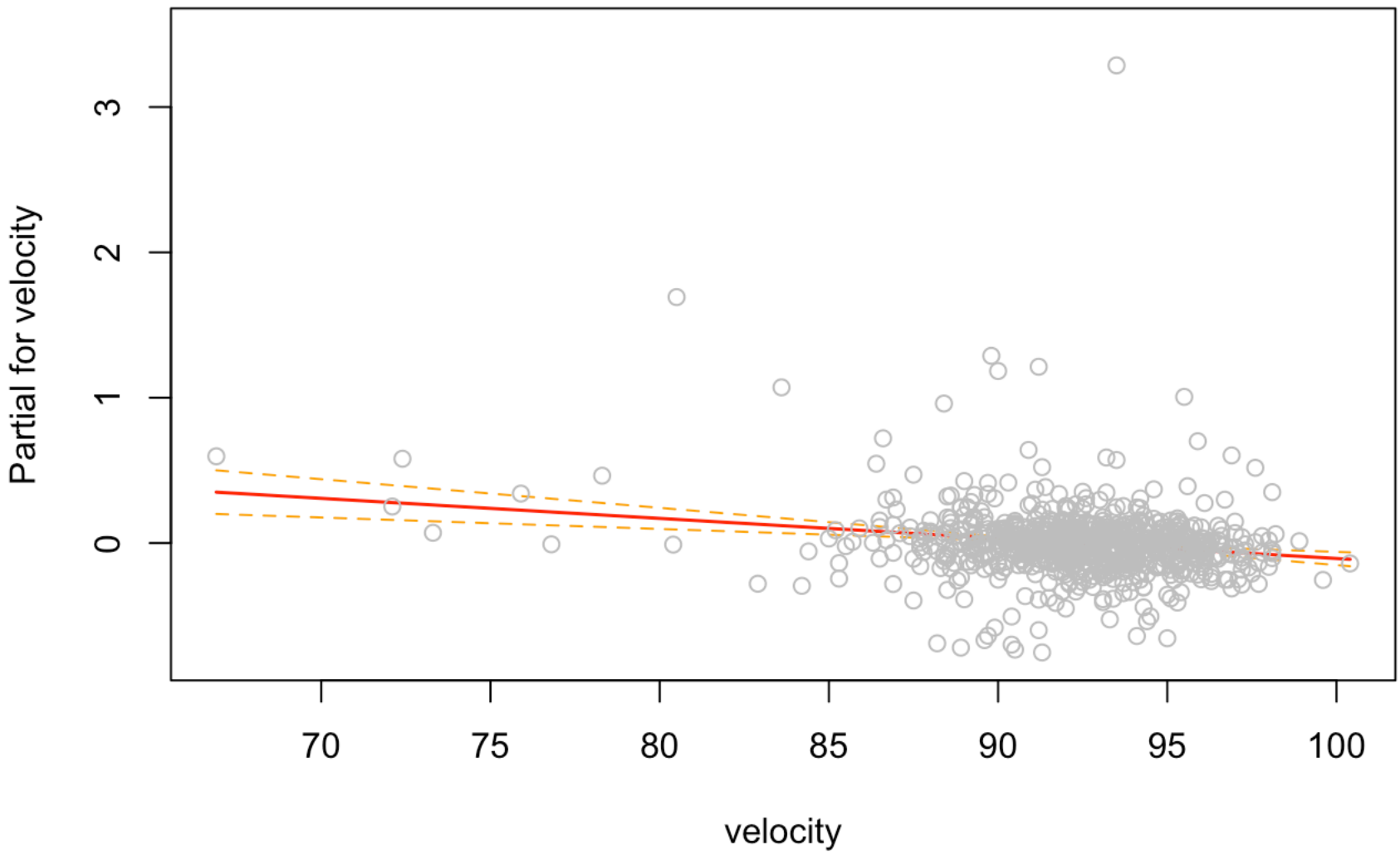

```
#Producing a summary of the model from above (velocity,spin rate, and total pitches p
redicting batting average)
#Plotting regression term agaisnt the predictors (batting average agaisnt velocity, s
pin rate, and total pitches)
```

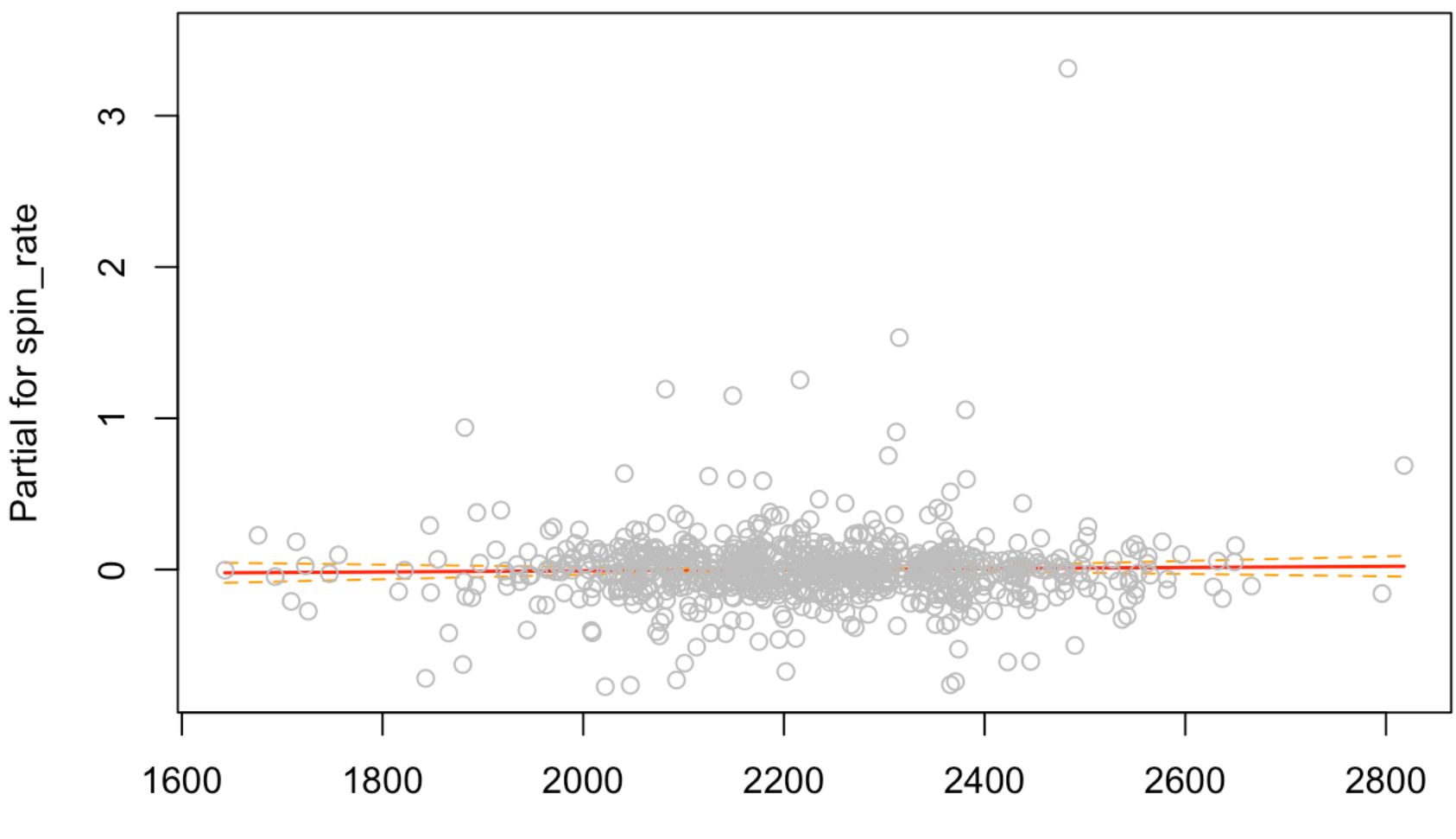
```
mod_velo_spin_slg = lm(formula = slg ~ velocity + spin_rate + I(log(total_pitches)),
data = fb_velo)
#Creating a model using velocity, spin rate, and total pitches as a predictor varibal
e for sluggling percentage
```

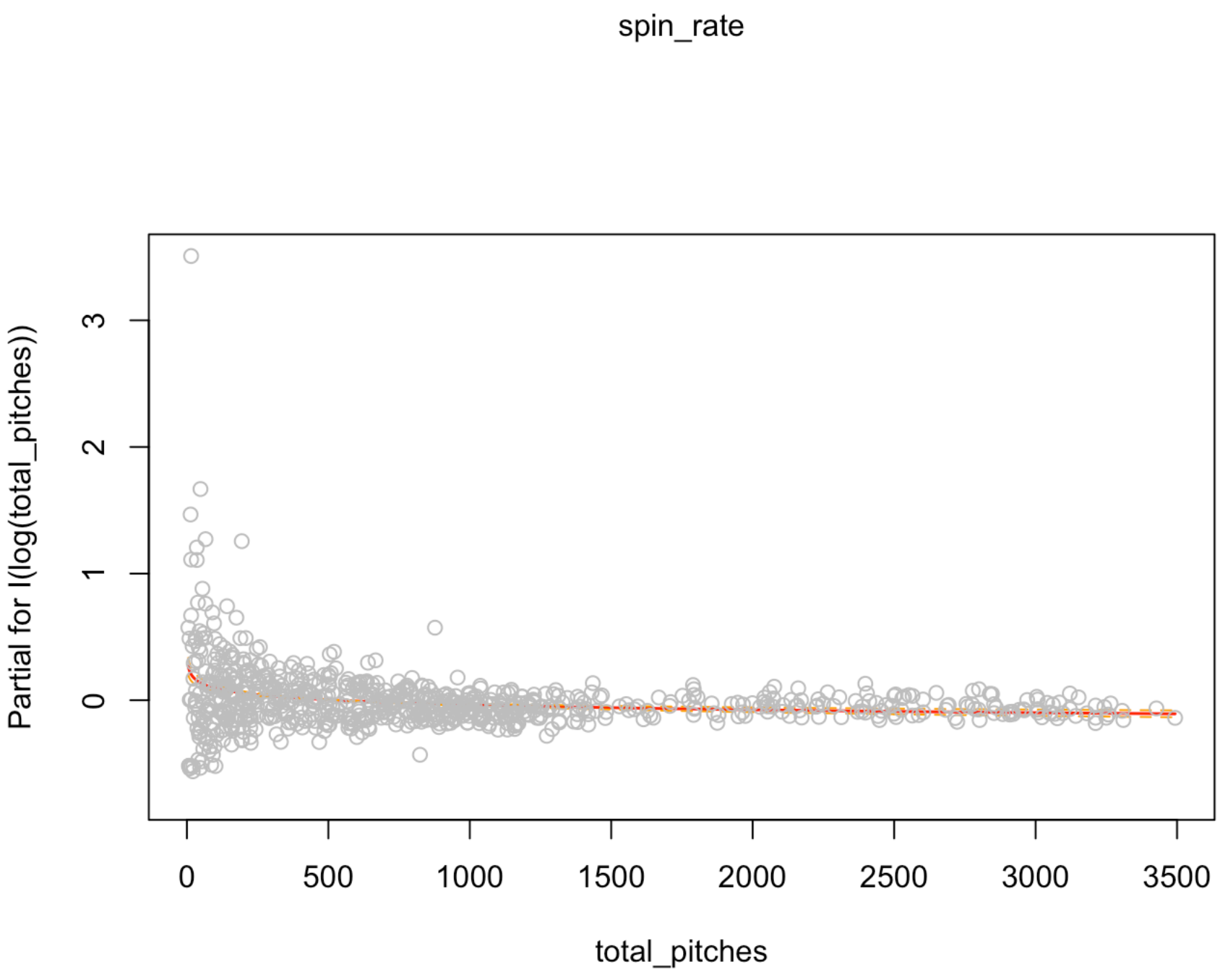
```
summary(mod_velo_spin_slg)
```

```
##
## Call:
## lm(formula = slg ~ velocity + spin_rate + I(log(total_pitches)),
##     data = fb_velo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7665 -0.0982 -0.0034  0.0812  3.3032
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             2.050e+00  2.492e-01   8.225 8.31e-16 ***
## velocity               -1.379e-02  2.955e-03  -4.667 3.60e-06 ***
## spin_rate               3.685e-05  5.701e-05   0.646    0.518
## I(log(total_pitches)) -5.724e-02  7.196e-03  -7.955 6.42e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2439 on 767 degrees of freedom
## Multiple R-squared:  0.1225, Adjusted R-squared:  0.1191
## F-statistic:  35.7 on 3 and 767 DF,  p-value: < 2.2e-16
```

```
termplot(mod_velo_spin_slg,partial.resid = T, se = T)
```


Partial for velocity
0
1
2
3
70
75
80
85
90
95
100
velocity


Partial for spin_rate
0
1
2
3
1600
1800
2000
2200
2400
2600
2800



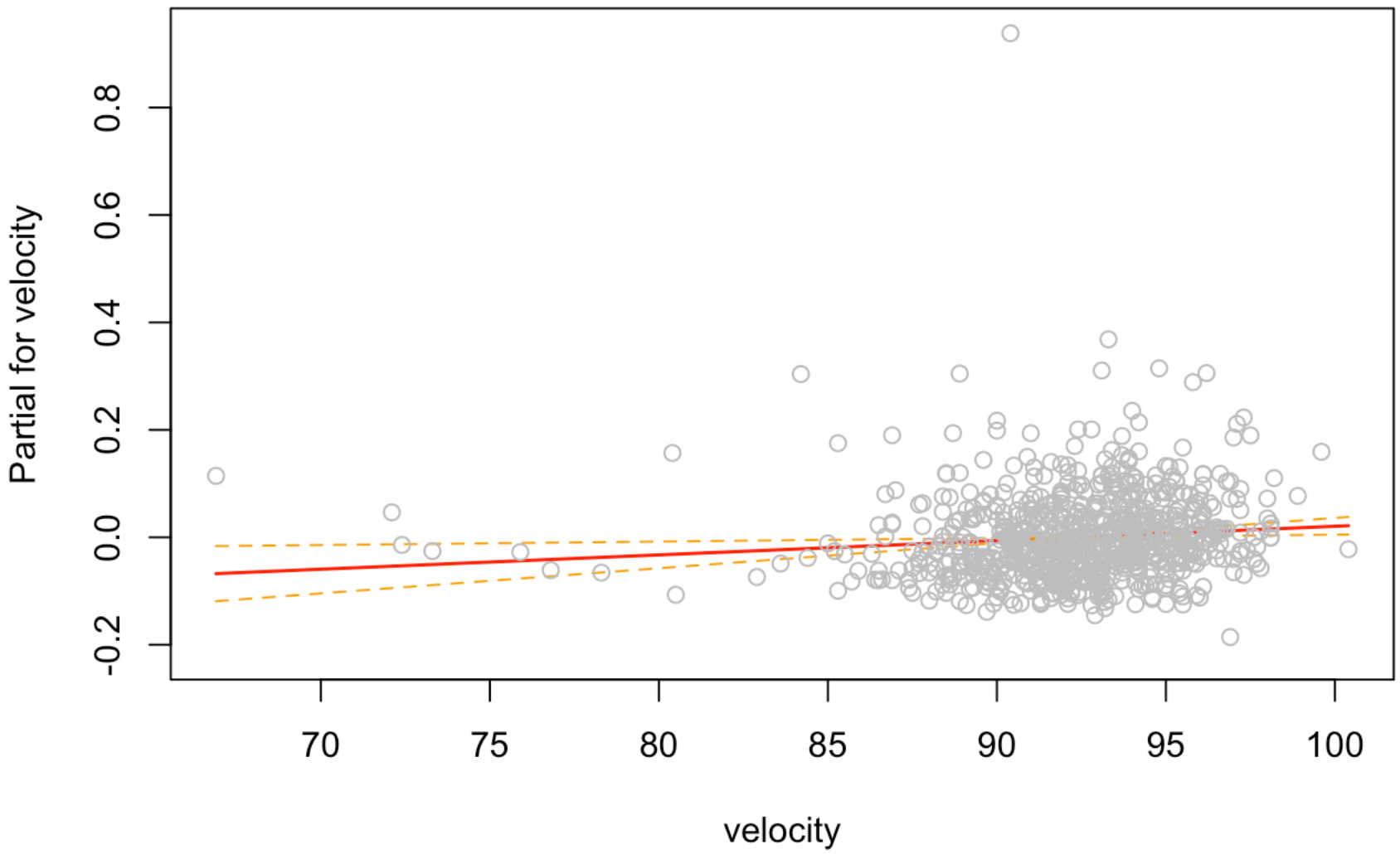

```
#Producing a summary of the model from above (velocity,spin rate, and total pitches p
redicting slugging percentage)
#Plotting regression term agaisnt the predictors (slugging percentage agaisnt velocit
y, spin rate, and total pitches)
```

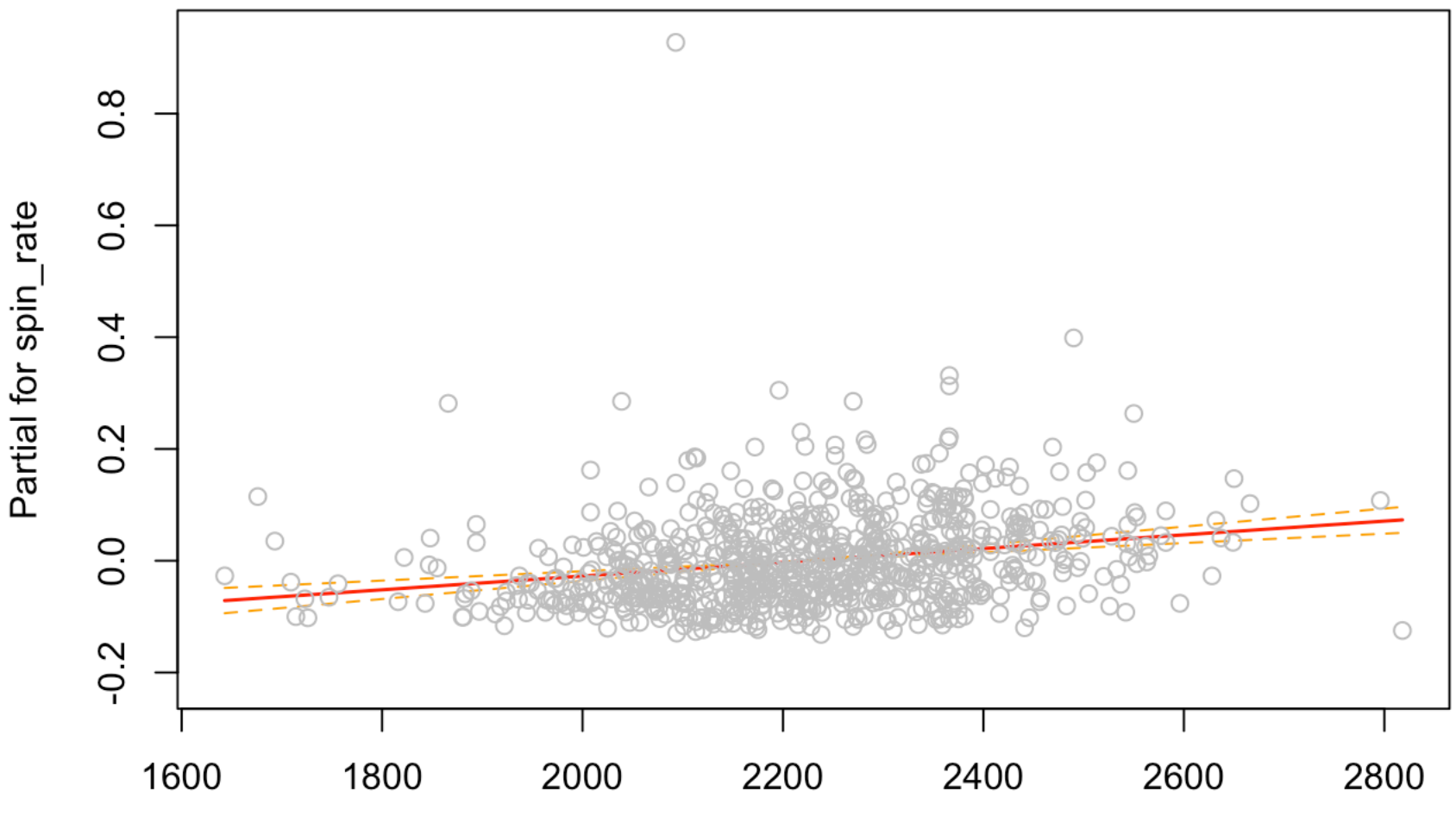
```
mod_velo_spin_swing_and_miss = lm(formula = swing_and_miss_pct ~ velocity + spin_rate
+ I(log(total_pitches)), data = fb_velo)
#Creating a model using velocity, spin rate, and total pitches as a predictor varibal
e for swing and miss percentage
```

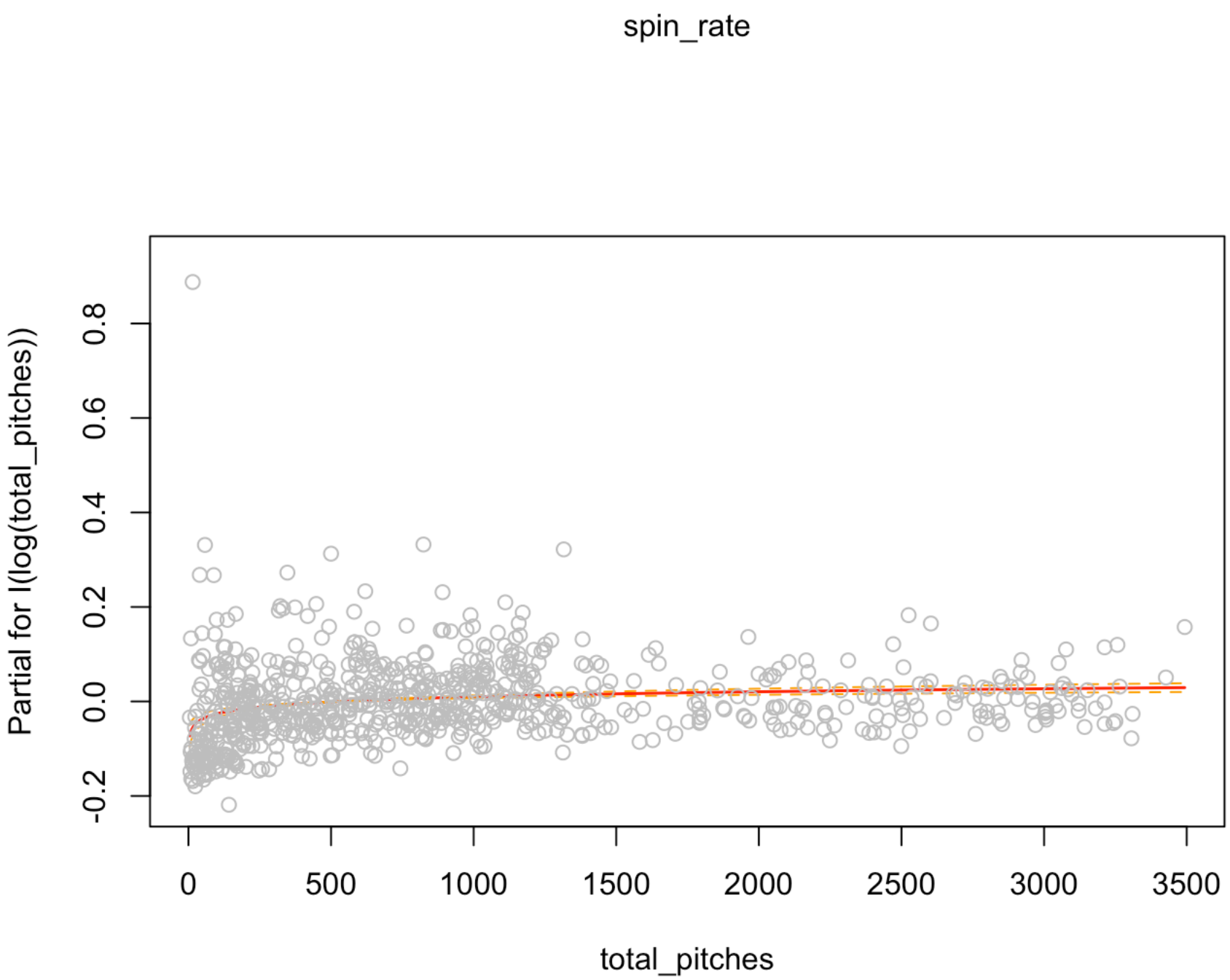
```
summary(mod_velo_spin_swing_and_miss)
```

```
##
## Call:
## lm(formula = swing_and_miss_pct ~ velocity + spin_rate + I(log(total_pitches)),
##     data = fb_velo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.19798 -0.05239 -0.01186  0.03870  0.94346
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -4.837e-01  8.513e-02  -5.681 1.90e-08 ***
## velocity               2.664e-03  1.009e-03   2.639  0.00849 **
## spin_rate              1.229e-04  1.947e-05   6.310 4.70e-10 ***
## I(log(total_pitches))  1.558e-02  2.458e-03   6.339 3.94e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08331 on 767 degrees of freedom
## Multiple R-squared:  0.144,  Adjusted R-squared:  0.1407
## F-statistic: 43.02 on 3 and 767 DF,  p-value: < 2.2e-16
```

```
termplot(mod_velo_spin_swing_and_miss,partial.resid = T, se = T)
```


Partial for velocity
-0.2
0.0
0.2
0.4
0.6
0.8
70
75
80
85
90
95
100
velocity


Partial for spin_rate
-0.2
0.0
0.2
0.4
0.6
0.8
1600
1800
2000
2200
2400
2600
2800




```
#Producing a summary of the model from above (velocity,spin rate, and total pitches p
redicting swing and miss percentage)
#Plotting regression term agaisnt the predictors (swing and miss percentage agaisnt v
elocity, spin rate, and total pitches)
```