FP0

Sophia Manodori and Greta Anesko

2025-04-15

Data Overview

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
data <- read.csv("large_dataset.csv")
head(data)
```

```
##                             event_name         r_fighter       b_fighter winner
## 1 UFC Fight Night: Ribas vs. Namajunas      Amanda Ribas  Rose Namajunas   Blue
## 2 UFC Fight Night: Ribas vs. Namajunas     Karl Williams     Justin Tafa    Red
## 3 UFC Fight Night: Ribas vs. Namajunas  Edmen Shahbazyan       AJ Dobson    Red
## 4 UFC Fight Night: Ribas vs. Namajunas    Payton Talbott Cameron Saaiman    Red
## 5 UFC Fight Night: Ribas vs. Namajunas Billy Quarantillo   Youssef Zalal   Blue
## 6 UFC Fight Night: Ribas vs. Namajunas  Fernando Padilla    Luis Pajuelo    Red
##        weight_class is_title_bout gender               method finish_round
## 1 Women's Flyweight             0  Women Decision - Unanimous            5
## 2       Heavyweight             0    Men Decision - Unanimous            3
## 3      Middleweight             0    Men               KO/TKO            1
## 4      Bantamweight             0    Men               KO/TKO            2
## 5     Featherweight             0    Men           Submission            2
## 6     Featherweight             0    Men           Submission            1
##   total_rounds time_sec       referee r_kd r_sig_str r_sig_str_att
## 1            5      300  Jason Herzog    0        83           270
## 2            3      300     Herb Dean    0        40            61
## 3            3      273    Mark Smith    1        27            43
## 4            3       21 Chris Tognoni    1        79           129
## 5            3      110     Herb Dean    0        10            50
## 6            3      165  Jason Herzog    1        34            61
```

```
##   r_sig_str_acc r_str r_str_att r_str_acc r_td r_td_att r_td_acc r_sub_att
## 1          0.30   137       342      0.40    4        8     0.50         0
## 2          0.65   123       176      0.70    7       12     0.58         1
## 3          0.62    31        51      0.61    1        3     0.33         0
## 4          0.61    83       135      0.61    0        1     0.00         0
## 5          0.20    26        74      0.35    0        0     0.00         0
## 6          0.55    35        64      0.55    0        0     0.00         1
##   r_rev r_ctrl_sec r_wins_total r_losses_total r_age r_height r_weight r_reach
## 1     0         98           12              5    30   160.02    56.70  167.64
## 2     0        631           10              1    34   190.50   106.59  200.66
## 3     0         49           13              4    26   187.96    83.91  190.50
## 4     0         30            8              0    25   177.80    61.23  177.80
## 5     0         21           18              6    35   177.80    65.77  177.80
## 6     0         47           16              5    27   185.42    65.77  193.04
##   r_stance r_SLpM_total r_SApM_total r_sig_str_acc_total r_td_acc_total
## 1 Orthodox         4.63         3.40                0.40           0.51
## 2 Orthodox         2.87         1.70                0.52           0.50
## 3 Orthodox         3.60         4.09                0.52           0.38
## 4   Switch         8.05         3.58                0.54           0.00
## 5 Orthodox         7.36         5.57                0.56           0.23
## 6 Orthodox         6.48         5.25                0.42           0.00
##   r_str_def_total r_td_def_total r_sub_avg r_td_avg b_kd b_sig_str
## 1            0.61           0.85       0.7     2.07    0        93
## 2            0.60           1.00       0.2     4.75    0        21
## 3            0.45           0.63       0.6     2.24    0        15
## 4            0.51           0.90       0.5     0.00    0        31
## 5            0.43           0.61       1.1     1.24    0        33
## 6            0.53           1.00       0.8     0.00    0         5
##   b_sig_str_att b_sig_str_acc b_str b_str_att b_str_acc b_td b_td_att b_td_acc
## 1           188          0.49   169       281      0.60    1        3     0.33
## 2            43          0.48    27        50      0.54    0        0     0.00
## 3            46          0.32    20        54      0.37    0        0     0.00
## 4            64          0.48    33        66      0.50    0        3     0.00
## 5            56          0.58    37        62      0.60    2        2     1.00
## 6            13          0.38     7        15      0.47    0        0     0.00
##   b_sub_att b_rev b_ctrl_sec b_wins_total b_losses_total b_age b_height
## 1         0     1        419           13              6    31   165.10
## 2         0     0          0            7              4    30   182.88
## 3         0     1         75            7              3    32   185.42
## 4         0     0          0            9              2    23   172.72
## 5         2     0        175           14              5    27   177.80
## 6         0     0          0            8              2    29   177.80
##   b_weight b_reach b_stance b_SLpM_total b_SApM_total b_sig_str_acc_total
## 1    56.70  165.10 Orthodox         3.69         3.51                0.41
## 2   119.75  187.96 Southpaw         4.09         5.02                0.54
## 3    83.91  193.04 Orthodox         4.29         5.31                0.46
## 4    61.23  170.18 Southpaw         5.32         4.18                0.46
## 5    65.77  182.88   Switch         2.88         1.73                0.49
## 6    65.77  175.26 Orthodox         6.03         8.20                0.57
##   b_td_acc_total b_str_def_total b_td_def_total b_sub_avg b_td_avg kd_diff
## 1           0.47            0.63           0.59       0.5     1.38       0
## 2           0.00            0.47           0.50       0.0     0.00       0
## 3           0.75            0.46           0.65       0.3     1.67       1
## 4           0.28            0.51           0.47       0.7     0.91       1
```

```
## 5           0.34            0.65           0.60       1.3     2.28       0
## 6           0.00            0.52           0.66       0.0     0.00       1
##   sig_str_diff sig_str_att_diff sig_str_acc_diff str_diff str_att_diff
## 1          -10               82           -0.19      -32           61
## 2           19               18            0.17       96          126
## 3           12               -3            0.30       11           -3
## 4           48               65            0.13       50           69
## 5          -23               -6           -0.38      -11           12
## 6           29               48            0.17       28           49
##   str_acc_diff td_diff td_att_diff td_acc_diff sub_att_diff rev_diff
## 1        -0.20       3           5        0.17            0       -1
## 2         0.16       7          12        0.58            1        0
## 3         0.24       1           3        0.33            0       -1
## 4         0.11       0          -2        0.00            0        0
## 5        -0.25      -2          -2       -1.00           -2        0
## 6         0.08       0           0        0.00            1        0
##   ctrl_sec_diff wins_total_diff losses_total_diff age_diff height_diff
## 1          -321              -1                -1       -1       -5.08
## 2           631               3                -3        4        7.62
## 3           -26               6                 1       -6        2.54
## 4            30              -1                -2        2        5.08
## 5          -154               4                 1        8        0.00
## 6            47               8                 3       -2        7.62
##   weight_diff reach_diff SLpM_total_diff SApM_total_diff sig_str_acc_total_diff
## 1        0.00       2.54            0.94           -0.11                  -0.01
## 2      -13.16      12.70           -1.22           -3.32                  -0.02
## 3        0.00      -2.54           -0.69           -1.22                   0.06
## 4        0.00       7.62            2.73           -0.60                   0.08
## 5        0.00      -5.08            4.48            3.84                   0.07
## 6        0.00      17.78            0.45           -2.95                  -0.15
##   td_acc_total_diff str_def_total_diff td_def_total_diff sub_avg_diff
## 1              0.04              -0.02              0.26         0.2
## 2              0.50               0.13              0.50         0.2
## 3             -0.37              -0.01             -0.02         0.3
## 4             -0.28               0.00              0.43        -0.2
## 5             -0.11              -0.22              0.01        -0.2
## 6              0.00               0.01              0.34         0.8
##   td_avg_diff
## 1        0.69
## 2        4.75
## 3        0.57
## 4       -0.91
## 5       -1.04
## 6        0.00
```

```
sdata <- data[-c(1, 5, 8:25, 41:53)]
head(sdata)
```

```
##           r_fighter       b_fighter winner is_title_bout gender r_wins_total
## 1      Amanda Ribas  Rose Namajunas   Blue             0  Women           12
## 2     Karl Williams     Justin Tafa    Red             0    Men           10
## 3  Edmen Shahbazyan       AJ Dobson    Red             0    Men           13
## 4    Payton Talbott Cameron Saaiman    Red             0    Men            8
## 5 Billy Quarantillo   Youssef Zalal   Blue             0    Men           18
```

```
## 6  Fernando Padilla    Luis Pajuelo    Red             0    Men           16
##   r_losses_total r_age r_height r_weight r_reach r_stance r_SLpM_total
## 1              5    30   160.02    56.70  167.64 Orthodox         4.63
## 2              1    34   190.50   106.59  200.66 Orthodox         2.87
## 3              4    26   187.96    83.91  190.50 Orthodox         3.60
## 4              0    25   177.80    61.23  177.80   Switch         8.05
## 5              6    35   177.80    65.77  177.80 Orthodox         7.36
## 6              5    27   185.42    65.77  193.04 Orthodox         6.48
##   r_SApM_total r_sig_str_acc_total r_td_acc_total r_str_def_total
## 1         3.40                0.40          0.51            0.61
## 2         1.70                0.52          0.50            0.60
## 3         4.09                0.52          0.38            0.45
## 4         3.58                0.54          0.00            0.51
## 5         5.57                0.56          0.23            0.43
## 6         5.25                0.42          0.00            0.53
##   r_td_def_total r_sub_avg r_td_avg b_wins_total b_losses_total b_age b_height
## 1           0.85       0.7     2.07           13              6    31   165.10
## 2           1.00       0.2     4.75            7              4    30   182.88
## 3           0.63       0.6     2.24            7              3    32   185.42
## 4           0.90       0.5     0.00            9              2    23   172.72
## 5           0.61       1.1     1.24           14              5    27   177.80
## 6           1.00       0.8     0.00            8              2    29   177.80
##   b_weight b_reach b_stance b_SLpM_total b_SApM_total b_sig_str_acc_total
## 1    56.70  165.10 Orthodox         3.69         3.51                0.41
## 2   119.75  187.96 Southpaw         4.09         5.02                0.54
## 3    83.91  193.04 Orthodox         4.29         5.31                0.46
## 4    61.23  170.18 Southpaw         5.32         4.18                0.46
## 5    65.77  182.88   Switch         2.88         1.73                0.49
## 6    65.77  175.26 Orthodox         6.03         8.20                0.57
##   b_td_acc_total b_str_def_total b_td_def_total b_sub_avg b_td_avg kd_diff
## 1           0.47            0.63           0.59       0.5     1.38       0
## 2           0.00            0.47           0.50       0.0     0.00       0
## 3           0.75            0.46           0.65       0.3     1.67       1
## 4           0.28            0.51           0.47       0.7     0.91       1
## 5           0.34            0.65           0.60       1.3     2.28       0
## 6           0.00            0.52           0.66       0.0     0.00       1
##   sig_str_diff sig_str_att_diff sig_str_acc_diff str_diff str_att_diff
## 1          -10               82            -0.19      -32           61
## 2           19               18             0.17       96          126
## 3           12               -3             0.30       11           -3
## 4           48               65             0.13       50           69
## 5          -23               -6            -0.38      -11           12
## 6           29               48             0.17       28           49
##   str_acc_diff td_diff td_att_diff td_acc_diff sub_att_diff rev_diff
## 1        -0.20       3           5        0.17            0       -1
## 2         0.16       7          12        0.58            1        0
## 3         0.24       1           3        0.33            0       -1
## 4         0.11       0          -2        0.00            0        0
## 5        -0.25      -2          -2       -1.00           -2        0
## 6         0.08       0           0        0.00            1        0
##   ctrl_sec_diff wins_total_diff losses_total_diff age_diff height_diff
## 1          -321              -1               -1       -1       -5.08
## 2           631               3               -3        4        7.62
## 3           -26               6                1       -6        2.54
```

```
## 4            30             -1               -2        2        5.08
## 5          -154              4                1        8        0.00
## 6            47              8                3       -2        7.62
##   weight_diff reach_diff SLpM_total_diff SApM_total_diff sig_str_acc_total_diff
## 1        0.00       2.54            0.94           -0.11                  -0.01
## 2      -13.16      12.70           -1.22           -3.32                  -0.02
## 3        0.00      -2.54           -0.69           -1.22                   0.06
## 4        0.00       7.62            2.73           -0.60                   0.08
## 5        0.00      -5.08            4.48            3.84                   0.07
## 6        0.00      17.78            0.45           -2.95                  -0.15
##   td_acc_total_diff str_def_total_diff td_def_total_diff sub_avg_diff
## 1              0.04              -0.02              0.26         0.2
## 2              0.50               0.13              0.50         0.2
## 3             -0.37              -0.01             -0.02         0.3
## 4             -0.28               0.00              0.43        -0.2
## 5             -0.11              -0.22              0.01        -0.2
## 6              0.00               0.01              0.34         0.8
##   td_avg_diff
## 1        0.69
## 2        4.75
## 3        0.57
## 4       -0.91
## 5       -1.04
## 6        0.00
```

```
str(data)
```

```
## 'data.frame':    7439 obs. of  95 variables:
##  $ event_name            : chr  "UFC Fight Night: Ribas vs. Namajunas" "UFC Fight Night: Ribas vs. Na
##  $ r_fighter             : chr  "Amanda Ribas" "Karl Williams" "Edmen Shahbazyan" "Payton Talbott" .
##  $ b_fighter             : chr  "Rose Namajunas" "Justin Tafa" "AJ Dobson" "Cameron Saaiman" ...
##  $ winner                : chr  "Blue" "Red" "Red" "Red" ...
##  $ weight_class          : chr  "Women's Flyweight" "Heavyweight" "Middleweight" "Bantamweight" ...
##  $ is_title_bout         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ gender                : chr  "Women" "Men" "Men" "Men" ...
##  $ method                : chr  "Decision - Unanimous" "Decision - Unanimous" "KO/TKO" "KO/TKO" ...
##  $ finish_round          : int  5 3 1 2 2 1 3 1 3 3 ...
##  $ total_rounds          : num  5 3 3 3 3 3 3 3 3 3 ...
##  $ time_sec              : int  300 300 273 21 110 165 300 135 300 300 ...
##  $ referee               : chr  "Jason Herzog" "Herb Dean" "Mark Smith" "Chris Tognoni" ...
##  $ r_kd                  : int  0 0 1 1 0 1 0 0 0 1 ...
##  $ r_sig_str             : int  83 40 27 79 10 34 29 7 25 105 ...
##  $ r_sig_str_att         : int  270 61 43 129 50 61 68 17 79 207 ...
##  $ r_sig_str_acc         : num  0.3 0.65 0.62 0.61 0.2 0.55 0.42 0.41 0.31 0.5 ...
##  $ r_str                 : int  137 123 31 83 26 35 129 12 64 106 ...
##  $ r_str_att             : int  342 176 51 135 74 64 188 23 137 208 ...
##  $ r_str_acc             : num  0.4 0.7 0.61 0.61 0.35 0.55 0.69 0.52 0.47 0.51 ...
##  $ r_td                  : int  4 7 1 0 0 0 0 2 2 0 ...
##  $ r_td_att              : int  8 12 3 1 0 0 0 3 8 1 ...
##  $ r_td_acc              : num  0.5 0.58 0.33 0 0 0 0 0.66 0.25 0 ...
##  $ r_sub_att             : int  0 1 0 0 0 1 0 0 0 0 ...
##  $ r_rev                 : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ r_ctrl_sec            : int  98 631 49 30 21 47 10 87 389 12 ...
##  $ r_wins_total          : int  12 10 13 8 18 16 21 16 14 14 ...
```

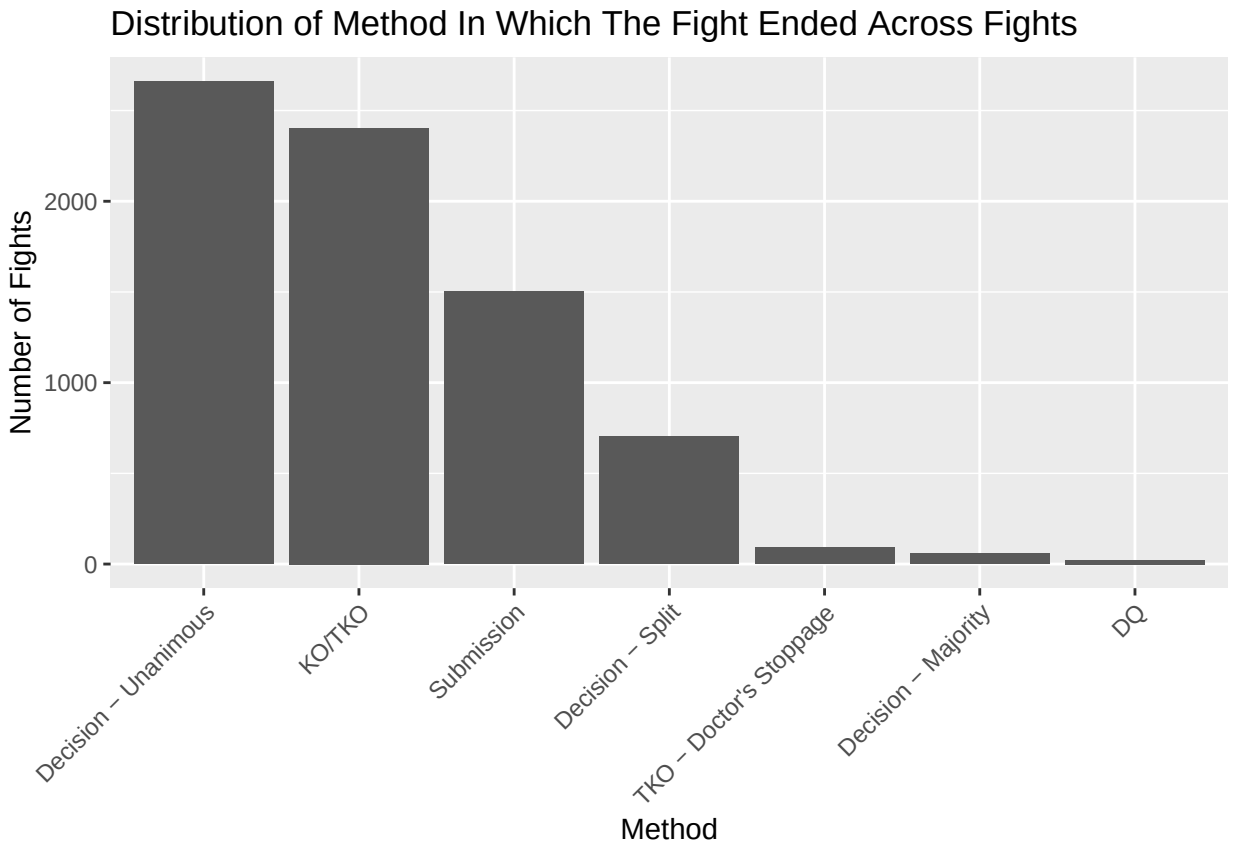
```
##  $ r_losses_total       : int  5 1 4 0 6 5 8 6 2 5 ...
##  $ r_age                : num  30 34 26 25 35 27 37 28 29 29 ...
##  $ r_height             : num  160 190 188 178 178 ...
##  $ r_weight             : num  56.7 106.6 83.9 61.2 65.8 ...
##  $ r_reach              : num  168 201 190 178 178 ...
##  $ r_stance             : chr  "Orthodox" "Orthodox" "Orthodox" "Switch" ...
##  $ r_SLpM_total         : num  4.63 2.87 3.6 8.05 7.36 6.48 4.37 3.21 3.28 3.44 ...
##  $ r_SApM_total         : num  3.4 1.7 4.09 3.58 5.57 5.25 4.07 3.74 2.55 3.36 ...
##  $ r_sig_str_acc_total  : num  0.4 0.52 0.52 0.54 0.56 0.42 0.43 0.37 0.47 0.46 ...
##  $ r_td_acc_total       : num  0.51 0.5 0.38 0 0.23 0 0.36 0.6 0.22 0 ...
##  $ r_str_def_total      : num  0.61 0.6 0.45 0.51 0.43 0.53 0.48 0.57 0.68 0.63 ...
##  $ r_td_def_total       : num  0.85 1 0.63 0.9 0.61 1 0.46 0.72 0.85 0.28 ...
##  $ r_sub_avg            : num  0.7 0.2 0.6 0.5 1.1 0.8 0.9 0.7 0 0.3 ...
##  $ r_td_avg             : num  2.07 4.75 2.24 0 1.24 0 0.71 3.02 1.15 0 ...
##  $ b_kd                 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ b_sig_str            : int  93 21 15 31 33 5 35 4 18 63 ...
##  $ b_sig_str_att        : int  188 43 46 64 56 13 66 11 62 226 ...
##  $ b_sig_str_acc        : num  0.49 0.48 0.32 0.48 0.58 0.38 0.53 0.36 0.29 0.27 ...
##  $ b_str                : int  169 27 20 33 37 7 225 11 52 63 ...
##  $ b_str_att            : int  281 50 54 66 62 15 320 21 108 227 ...
##  $ b_str_acc            : num  0.6 0.54 0.37 0.5 0.6 0.47 0.7 0.52 0.48 0.28 ...
##  $ b_td                 : int  1 0 0 0 2 0 3 0 1 1 ...
##  $ b_td_att             : int  3 0 0 3 2 0 4 0 7 1 ...
##  $ b_td_acc             : num  0.33 0 0 0 1 0 0.75 0 0.14 1 ...
##  $ b_sub_att            : int  0 0 0 0 2 0 0 1 0 0 ...
##  $ b_rev                : int  1 0 1 0 0 0 0 0 0 0 ...
##  $ b_ctrl_sec           : int  419 0 75 0 175 0 669 0 115 23 ...
##  $ b_wins_total         : int  13 7 7 9 14 8 17 29 20 9 ...
##  $ b_losses_total       : int  6 4 3 2 5 2 6 12 10 2 ...
##  $ b_age                : num  31 30 32 23 27 29 34 34 36 30 ...
##  $ b_height             : num  165 183 185 173 178 ...
##  $ b_weight             : num  56.7 119.8 83.9 61.2 65.8 ...
##  $ b_reach              : num  165 188 193 170 183 ...
##  $ b_stance             : chr  "Orthodox" "Southpaw" "Orthodox" "Southpaw" ...
##  $ b_SLpM_total         : num  3.69 4.09 4.29 5.32 2.88 6.03 3.38 6.14 4 7.15 ...
##  $ b_SApM_total         : num  3.51 5.02 5.31 4.18 1.73 8.2 3.59 6.29 4.08 7.36 ...
##  $ b_sig_str_acc_total  : num  0.41 0.54 0.46 0.46 0.49 0.57 0.47 0.48 0.43 0.4 ...
##  $ b_td_acc_total       : num  0.47 0 0.75 0.28 0.34 0 0.23 0.46 0.28 1 ...
##  $ b_str_def_total      : num  0.63 0.47 0.46 0.51 0.65 0.52 0.58 0.47 0.57 0.6 ...
##  $ b_td_def_total       : num  0.59 0.5 0.65 0.47 0.6 0.66 1 0.58 0.68 0.84 ...
##  $ b_sub_avg            : num  0.5 0 0.3 0.7 1.3 0 0.6 0.8 0.2 0 ...
##  $ b_td_avg             : num  1.38 0 1.67 0.91 2.28 0 1.64 1.63 1.28 0.28 ...
##  $ kd_diff              : int  0 0 1 1 0 1 0 0 0 1 ...
##  $ sig_str_diff         : int  -10 19 12 48 -23 29 -6 3 7 42 ...
##  $ sig_str_att_diff     : int  82 18 -3 65 -6 48 2 6 17 -19 ...
##  $ sig_str_acc_diff     : num  -0.19 0.17 0.3 0.13 -0.38 0.17 -0.11 0.05 0.02 0.23 ...
##  $ str_diff             : int  -32 96 11 50 -11 28 -96 1 12 43 ...
##  $ str_att_diff         : int  61 126 -3 69 12 49 -132 2 29 -19 ...
##  $ str_acc_diff         : num  -0.2 0.16 0.24 0.11 -0.25 0.08 -0.01 0 -0.01 0.23 ...
##  $ td_diff              : int  3 7 1 0 -2 0 -3 2 1 -1 ...
##  $ td_att_diff          : int  5 12 3 -2 -2 0 -4 3 1 0 ...
##  $ td_acc_diff          : num  0.17 0.58 0.33 0 -1 0 -0.75 0.66 0.11 -1 ...
##  $ sub_att_diff         : int  0 1 0 0 -2 1 0 -1 0 0 ...
##  $ rev_diff             : int  -1 0 -1 0 0 0 0 0 1 0 ...
```

```
##  $ ctrl_sec_diff         : int  -321 631 -26 30 -154 47 -659 87 274 -11 ...
##  $ wins_total_diff       : int  -1 3 6 -1 4 8 4 -13 -6 5 ...
##  $ losses_total_diff     : int  -1 -3 1 -2 1 3 2 -6 -8 3 ...
##  $ age_diff              : num  -1 4 -6 2 8 -2 3 -6 -7 -1 ...
##  $ height_diff           : num  -5.08 7.62 2.54 5.08 0 ...
##  $ weight_diff           : num  0 -13.2 0 0 0 ...
##  $ reach_diff            : num  2.54 12.7 -2.54 7.62 -5.08 ...
##  $ SLpM_total_diff       : num  0.94 -1.22 -0.69 2.73 4.48 0.45 0.99 -2.93 -0.72 -3.71 ...
##  $ SApM_total_diff       : num  -0.11 -3.32 -1.22 -0.6 3.84 ...
##  $ sig_str_acc_total_diff: num  -0.01 -0.02 0.06 0.08 0.07 ...
##  $ td_acc_total_diff     : num  0.04 0.5 -0.37 -0.28 -0.11 ...
##  $ str_def_total_diff    : num  -0.02 0.13 -0.01 0 -0.22 ...
##  $ td_def_total_diff     : num  0.26 0.5 -0.02 0.43 0.01 0.34 -0.54 0.14 0.17 -0.56 ...
##  $ sub_avg_diff          : num  0.2 0.2 0.3 -0.2 -0.2 0.8 0.3 -0.1 -0.2 0.3 ...
##  $ td_avg_diff           : num  0.69 4.75 0.57 -0.91 -1.04 0 -0.93 1.39 -0.13 -0.28 ...
```

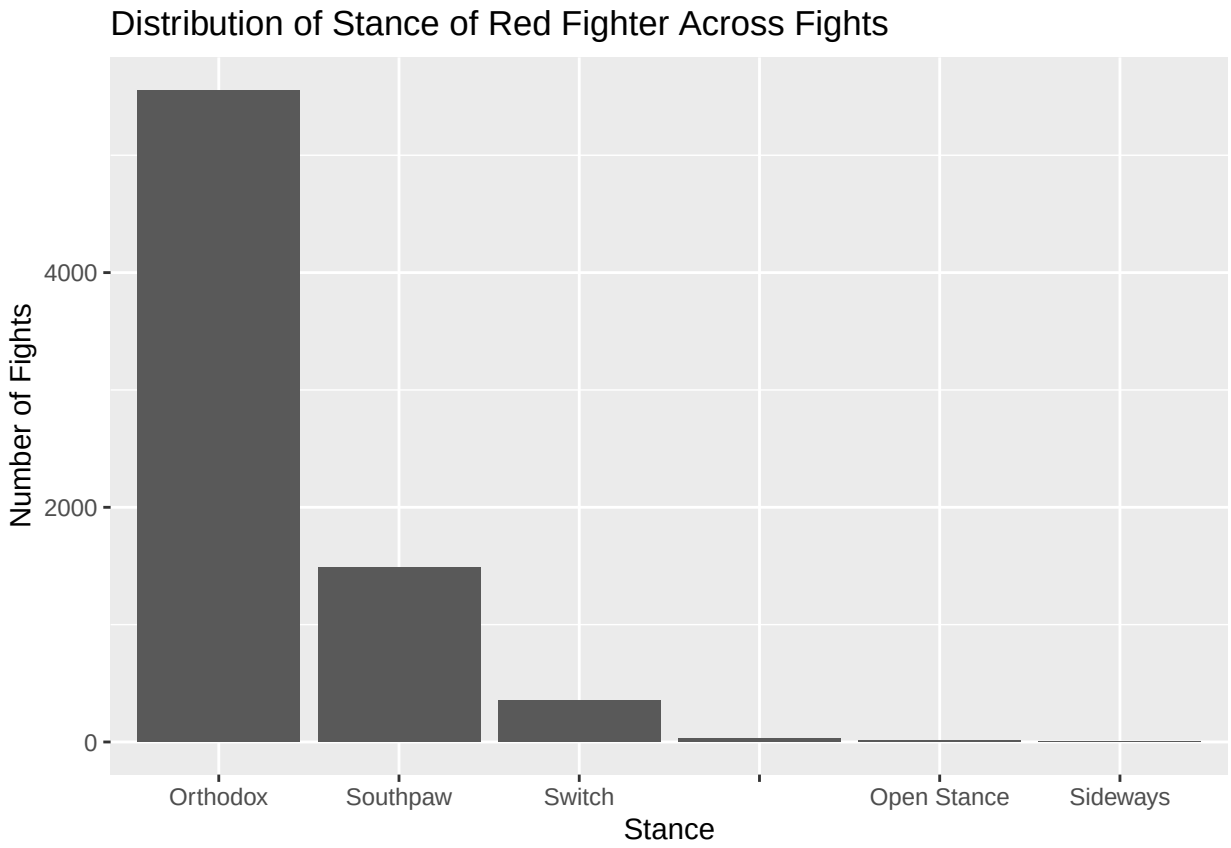
The data is from the (UFC Complete Dataset on Kaggle)[https://www.kaggle.com/datasets/maksbasher/ufc-complete-dataset-all-events-1996-2024] that is parsed from the (UFCStats website)[http://www.ufcstats.com/statistics/events/completed]. It has detailed information on all UFC fights from 1996-2024, and contains all individual fighter statistics as well as the in fight statistics. The dataset contains 95 variables. For the purposes of this project, we will be primarily focusing on the fighter statistics as predictor variables for fight outcomes. Fighter statistics include physical statistics such as height, weight, and stance, and career statistics such as strikes accuracy and takedown accuracy. This dataset also includes many differences in fighter statistics between the Red and Blue fighters, all of which are calculated as Red minus Blue.

Data Variability

```
# histogram for method in which the fight ended
ggplot(data = data, mapping = aes(x=fct_infreq(method))) +
  geom_bar() +
  labs(title = "Distribution of Method In Which The Fight Ended Across Fights", x = "Method", y = "Numbe
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

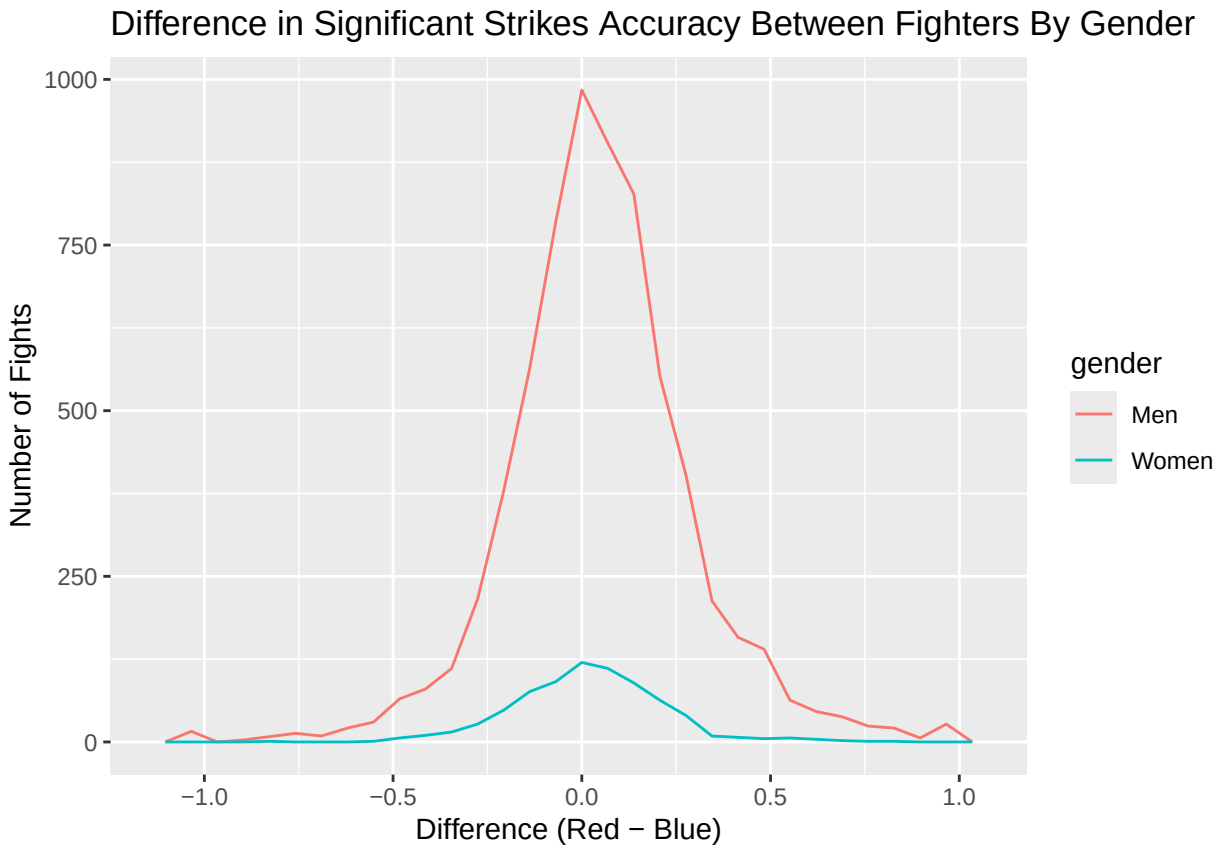



```
# histogram for red stance
ggplot(data = data, mapping = aes(x=fct_infreq(r_stance))) +
  geom_bar() +
  labs(title = "Distribution of Stance of Red Fighter Across Fights", x = "Stance", y = "Number of Figh
```




```
# lineplot of difference in significant strikes accuracy between fighters by gender
ggplot(data = data, mapping = aes(x=sig_str_acc_diff, color = gender)) +
  geom_freqpoly() +
  labs(title = "Difference in Significant Strikes Accuracy Between Fighters By Gender", x = "Difference
```

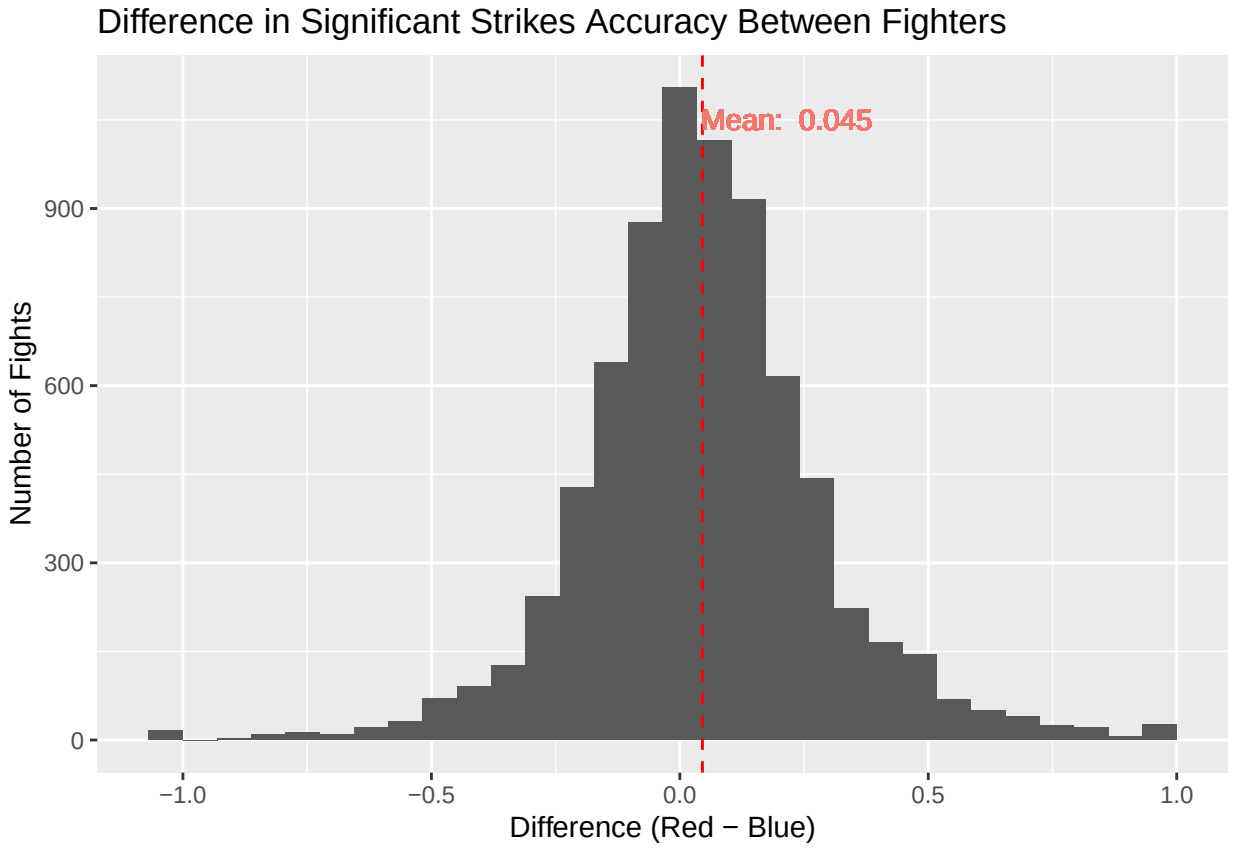
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# histogram of difference in significant strikes accuracy between fighters
ggplot(data = data) +
  geom_histogram(mapping = aes(x = sig_str_acc_diff)) +
  geom_vline(aes(xintercept = mean(sig_str_acc_diff)),
             color = "red", linetype="dashed") +
  geom_text(aes(x = mean(sig_str_acc_diff) + 0.17, y = 1050, label = paste("Mean: ", as.character(round
  theme(legend.position = "none") +
  labs(title = "Difference in Significant Strikes Accuracy Between Fighters", x = "Difference (Red - Blu
```

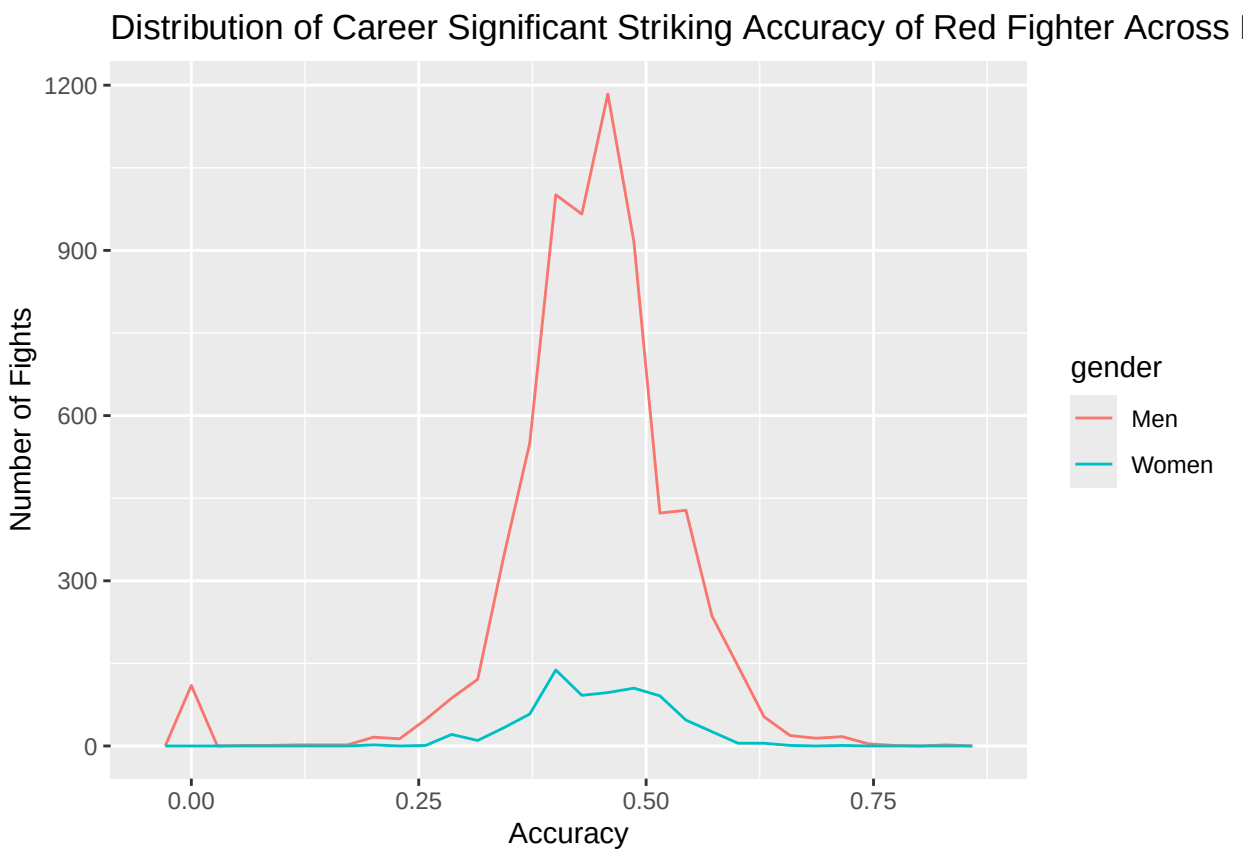
```
## Warning in geom_text(aes(x = mean(sig_str_acc_diff) + 0.17, y = 1050, label = paste("Mean: ", : All a
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



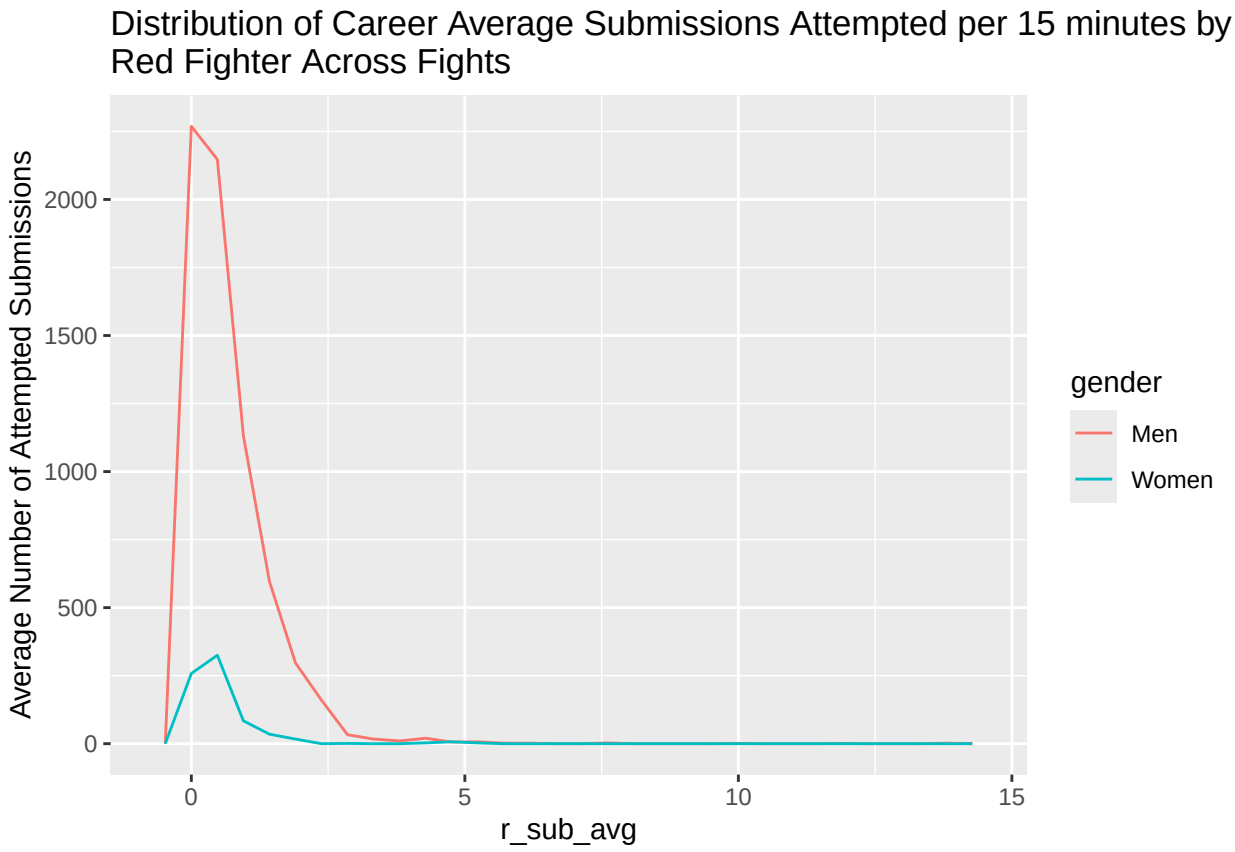

```
# red total average striking
ggplot(data = data, mapping = aes(x=r_sig_str_acc_total, color = gender)) +
  geom_freqpoly() +
  labs(title = "Distribution of Career Significant Striking Accuracy of Red Fighter Across Fights", x =
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



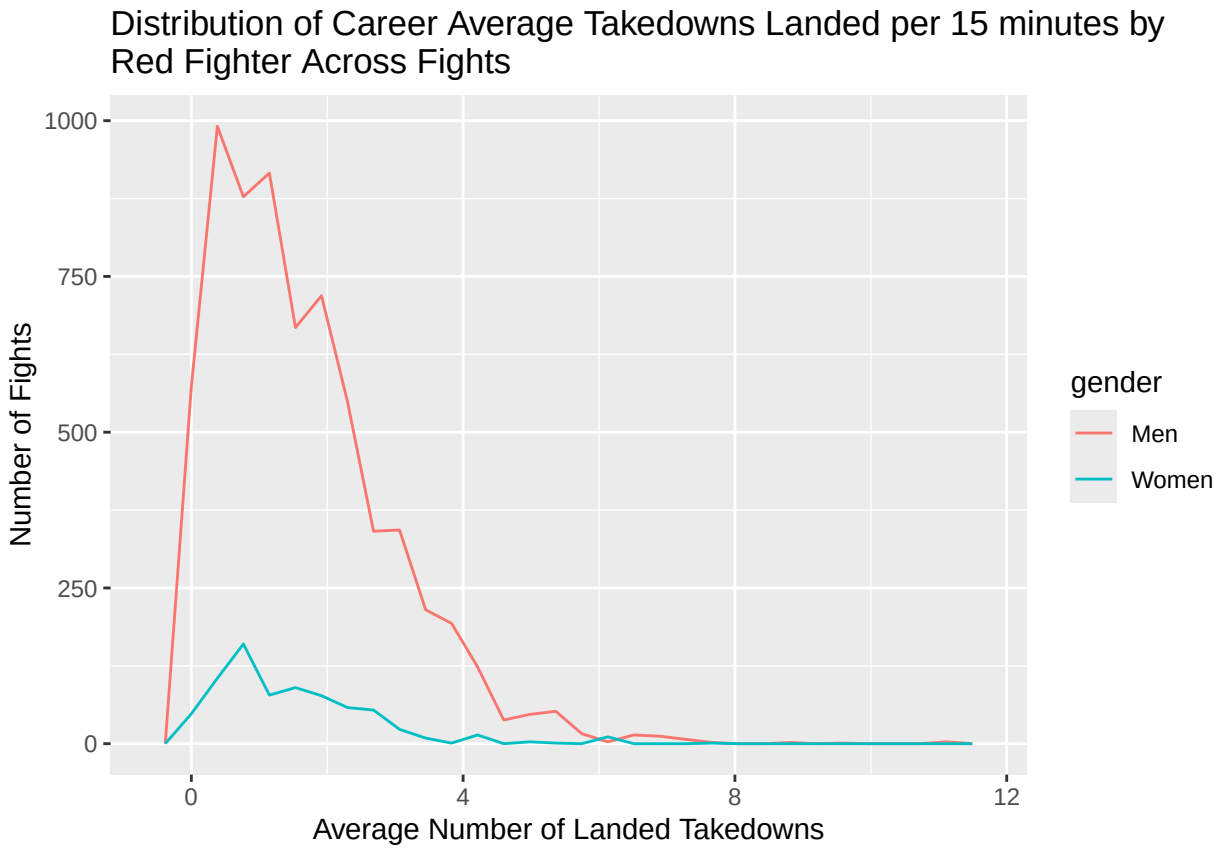

```
# red submission average
ggplot(data = data, mapping = aes(x=r_sub_avg, color = gender)) +
  geom_freqpoly() +
  labs(title = "Distribution of Career Average Submissions Attempted per 15 minutes by\nRed Fighter Acr
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# red takedown average
ggplot(data = data, mapping = aes(x=r_td_avg, color = gender)) +
  geom_freqpoly() +
  labs(title = "Distribution of Career Average Takedowns Landed per 15 minutes by\nRed Fighter Across F
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



In the histogram of the differences in significant strikes accuracy between fighters, we can see that the differences follow an approximately normal distribution centered around 0.045, meaning that the red fighter has only a very slightly higher significant strike accuracy rate compared to the blue fighter; the accuracy rates do not appear to be significantly affected by whether the fighter is the red or blue fighter.

Missing Data

```
summary(sdata)
```

```
##   r_fighter          b_fighter           winner          is_title_bout
##  Length:7439        Length:7439        Length:7439        Min.   :0.00000
##  Class :character   Class :character   Class :character   1st Qu.:0.00000
##  Mode  :character   Mode  :character   Mode  :character   Median :0.00000
##                                                           Mean   :0.05579
##                                                           3rd Qu.:0.00000
##                                                           Max.   :1.00000
##
##     gender           r_wins_total    r_losses_total      r_age
##  Length:7439        Min.   :  0.00   Min.   : 0.00   Min.   :19.00
##  Class :character   1st Qu.: 13.00   1st Qu.: 5.00   1st Qu.:34.00
##  Mode  :character   Median : 18.00   Median : 7.00   Median :38.00
##                     Mean   : 19.19   Mean   : 7.94   Mean   :38.32
##                     3rd Qu.: 24.00   3rd Qu.:11.00   3rd Qu.:42.00
##                     Max.   :253.00   Max.   :53.00   Max.   :65.00
```

```
##                                                NA's   :76
##     r_height        r_weight         r_reach        r_stance
##  Min.   :152.4   Min.   : 52.16   Min.   :147.3   Length:7439
##  1st Qu.:172.7   1st Qu.: 65.77   1st Qu.:177.8   Class :character
##  Median :177.8   Median : 77.11   Median :182.9   Mode  :character
##  Mean   :178.6   Mean   : 76.49   Mean   :183.2
##  3rd Qu.:185.4   3rd Qu.: 83.91   3rd Qu.:190.5
##  Max.   :210.8   Max.   :156.49   Max.   :213.4
##                                   NA's   :412
##   r_SLpM_total     r_SApM_total    r_sig_str_acc_total r_td_acc_total
##  Min.   : 0.000   Min.   : 0.000   Min.   :0.0000      Min.   :0.0000
##  1st Qu.: 2.570   1st Qu.: 2.510   1st Qu.:0.4000      1st Qu.:0.2900
##  Median : 3.330   Median : 3.180   Median :0.4500      Median :0.3900
##  Mean   : 3.412   Mean   : 3.283   Mean   :0.4414      Mean   :0.3883
##  3rd Qu.: 4.205   3rd Qu.: 3.945   3rd Qu.:0.5000      3rd Qu.:0.5000
##  Max.   :23.330   Max.   :15.480   Max.   :0.8300      Max.   :1.0000
##
##  r_str_def_total  r_td_def_total     r_sub_avg          r_td_avg
##  Min.   :0.0000   Min.   :0.0000   Min.   : 0.0000   Min.   : 0.000
##  1st Qu.:0.5100   1st Qu.:0.5000   1st Qu.: 0.1000   1st Qu.: 0.620
##  Median :0.5600   Median :0.6300   Median : 0.5000   Median : 1.330
##  Mean   :0.5431   Mean   :0.6028   Mean   : 0.6452   Mean   : 1.598
##  3rd Qu.:0.6000   3rd Qu.:0.7400   3rd Qu.: 0.9000   3rd Qu.: 2.290
##  Max.   :0.8400   Max.   :1.0000   Max.   :13.8000   Max.   :11.110
##
##   b_wins_total    b_losses_total       b_age          b_height
##  Min.   :  0.00   Min.   : 0.000   Min.   :21.00   Min.   :152.4
##  1st Qu.: 11.00   1st Qu.: 4.000   1st Qu.:33.00   1st Qu.:172.7
##  Median : 16.00   Median : 7.000   Median :37.00   Median :177.8
##  Mean   : 17.17   Mean   : 7.399   Mean   :37.72   Mean   :178.6
##  3rd Qu.: 21.00   3rd Qu.:10.000   3rd Qu.:42.00   3rd Qu.:185.4
##  Max.   :253.00   Max.   :53.000   Max.   :81.00   Max.   :210.8
##                                    NA's   :190
##     b_weight         b_reach        b_stance          b_SLpM_total
##  Min.   : 52.16   Min.   :147.3   Length:7439        Min.   : 0.000
##  1st Qu.: 65.77   1st Qu.:175.3   Class :character   1st Qu.: 2.340
##  Median : 70.31   Median :182.9   Mode  :character   Median : 3.250
##  Mean   : 76.32   Mean   :182.8                      Mean   : 3.269
##  3rd Qu.: 83.91   3rd Qu.:190.5                      3rd Qu.: 4.110
##  Max.   :349.27   Max.   :213.4                      Max.   :11.030
##                   NA's   :888
##   b_SApM_total    b_sig_str_acc_total b_td_acc_total   b_str_def_total
##  Min.   : 0.000   Min.   :0.0000      Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 2.600   1st Qu.:0.3900      1st Qu.:0.2500   1st Qu.:0.4900
##  Median : 3.290   Median :0.4400      Median :0.3600   Median :0.5400
##  Mean   : 3.455   Mean   :0.4293      Mean   :0.3602   Mean   :0.5222
##  3rd Qu.: 4.180   3rd Qu.:0.4900      3rd Qu.:0.4800   3rd Qu.:0.5900
##  Max.   :42.000   Max.   :1.0000      Max.   :1.0000   Max.   :1.0000
##
##  b_td_def_total     b_sub_avg          b_td_avg         kd_diff
##  Min.   :0.0000   Min.   : 0.0000   Min.   : 0.000   Min.   :-4.00000
##  1st Qu.:0.4500   1st Qu.: 0.0000   1st Qu.: 0.490   1st Qu.: 0.00000
##  Median :0.6100   Median : 0.4000   Median : 1.160   Median : 0.00000
##  Mean   :0.5653   Mean   : 0.5995   Mean   : 1.464   Mean   : 0.06789
```

```
##  3rd Qu.:0.7200   3rd Qu.: 0.8000   3rd Qu.: 2.120   3rd Qu.: 0.00000
##  Max.   :1.0000   Max.   :16.4000   Max.   :13.950   Max.   : 5.00000
##
##   sig_str_diff      sig_str_att_diff   sig_str_acc_diff      str_diff
##  Min.   :-157.000   Min.   :-318.000   Min.   :-1.00000   Min.   :-276.000
##  1st Qu.:  -8.000   1st Qu.: -14.000   1st Qu.:-0.09000   1st Qu.: -12.000
##  Median :   4.000   Median :   4.000   Median : 0.04000   Median :   6.000
##  Mean   :   4.861   Mean   :   5.564   Mean   : 0.04549   Mean   :   8.783
##  3rd Qu.:  18.000   3rd Qu.:  26.000   3rd Qu.: 0.17000   3rd Qu.:  29.000
##  Max.   : 312.000   Max.   : 461.000   Max.   : 1.00000   Max.   : 315.000
##
##   str_att_diff       str_acc_diff         td_diff          td_att_diff
##  Min.   :-339.000   Min.   :-1.00000   Min.   :-12.0000   Min.   :-44.0000
##  1st Qu.: -19.000   1st Qu.:-0.09000   1st Qu.: -1.0000   1st Qu.: -2.0000
##  Median :   6.000   Median : 0.03000   Median :  0.0000   Median :  0.0000
##  Mean   :   9.993   Mean   : 0.04029   Mean   :  0.3335   Mean   :  0.2883
##  3rd Qu.:  38.000   3rd Qu.: 0.17000   3rd Qu.:  1.0000   3rd Qu.:  3.0000
##  Max.   : 462.000   Max.   : 1.00000   Max.   : 20.0000   Max.   : 26.0000
##
##   td_acc_diff        sub_att_diff        rev_diff          ctrl_sec_diff
##  Min.   :-1.00000   Min.   :-7.0000   Min.   :-3.000000   Min.   :-1164.00
##  1st Qu.:-0.20000   1st Qu.: 0.0000   1st Qu.: 0.000000   1st Qu.:  -53.00
##  Median : 0.00000   Median : 0.0000   Median : 0.000000   Median :    5.00
##  Mean   : 0.08486   Mean   : 0.1303   Mean   : 0.001344   Mean   :   42.92
##  3rd Qu.: 0.50000   3rd Qu.: 0.0000   3rd Qu.: 0.000000   3rd Qu.:  144.00
##  Max.   : 1.00000   Max.   :10.0000   Max.   : 5.000000   Max.   : 1301.00
##
##  wins_total_diff    losses_total_diff     age_diff        height_diff
##  Min.   :-241.000   Min.   :-47.0000   Min.   :-24.000   Min.   :-30.48000
##  1st Qu.:  -4.000   1st Qu.: -3.0000   1st Qu.: -3.000   1st Qu.: -5.08000
##  Median :   2.000   Median :  0.0000   Median :  0.000   Median :  0.00000
##  Mean   :   2.019   Mean   :  0.5412   Mean   :  0.388   Mean   :  0.04609
##  3rd Qu.:   8.000   3rd Qu.:  4.0000   3rd Qu.:  4.000   3rd Qu.:  5.08000
##  Max.   : 249.000   Max.   : 46.0000   Max.   : 17.000   Max.   : 33.02000
##                                        NA's   :213
##   weight_diff         reach_diff        SLpM_total_diff   SApM_total_diff
##  Min.   :-258.550   Min.   :-27.9400   Min.   :-8.9900   Min.   :-39.4900
##  1st Qu.:   0.000   1st Qu.: -5.0800   1st Qu.:-0.8600   1st Qu.: -1.0400
##  Median :   0.000   Median :  0.0000   Median : 0.1300   Median : -0.1200
##  Mean   :   0.171   Mean   :  0.1901   Mean   : 0.1426   Mean   : -0.1716
##  3rd Qu.:   0.000   3rd Qu.:  5.0800   3rd Qu.: 1.1600   3rd Qu.:  0.7900
##  Max.   :  52.160   Max.   : 33.0200   Max.   :18.7800   Max.   : 12.6400
##                     NA's   :1038
##  sig_str_acc_total_diff td_acc_total_diff  str_def_total_diff
##  Min.   :-0.70000       Min.   :-1.00000   Min.   :-0.58000
##  1st Qu.:-0.06000       1st Qu.:-0.13000   1st Qu.:-0.04000
##  Median : 0.01000       Median : 0.02000   Median : 0.01000
##  Mean   : 0.01211       Mean   : 0.02816   Mean   : 0.02091
##  3rd Qu.: 0.08000       3rd Qu.: 0.19000   3rd Qu.: 0.08000
##  Max.   : 0.83000       Max.   : 1.00000   Max.   : 0.72000
##
##  td_def_total_diff   sub_avg_diff        td_avg_diff
##  Min.   :-1.00000   Min.   :-15.10000   Min.   :-11.7700
##  1st Qu.:-0.14000   1st Qu.: -0.40000   1st Qu.: -0.8700
```

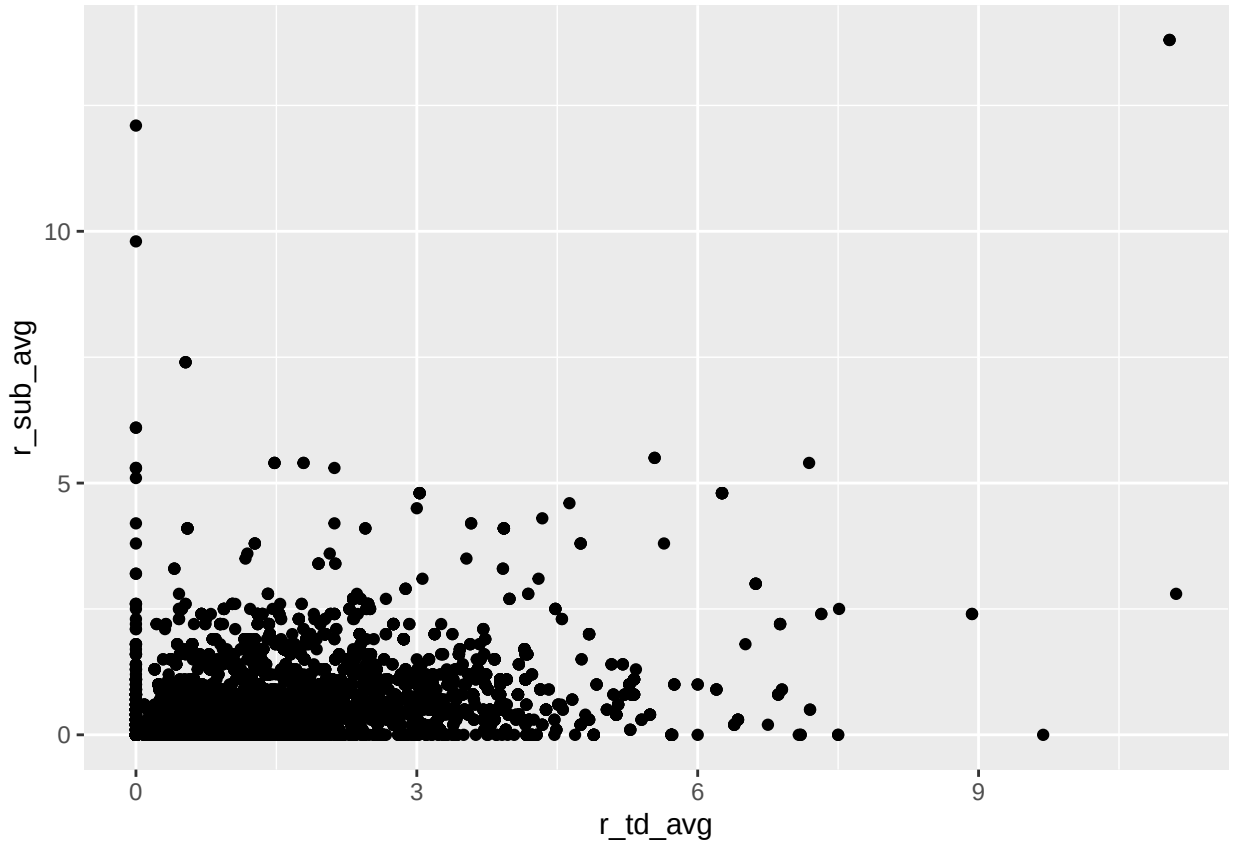
```
##  Median : 0.02000   Median :  0.00000   Median :  0.0900
##  Mean   : 0.03751   Mean   :  0.04565   Mean   :  0.1345
##  3rd Qu.: 0.21000   3rd Qu.:  0.50000   3rd Qu.:  1.1600
##  Max.   : 1.00000   Max.   : 13.80000   Max.   : 11.1100
##
```

```
names(which(colSums(is.na(data)) > 0))
```

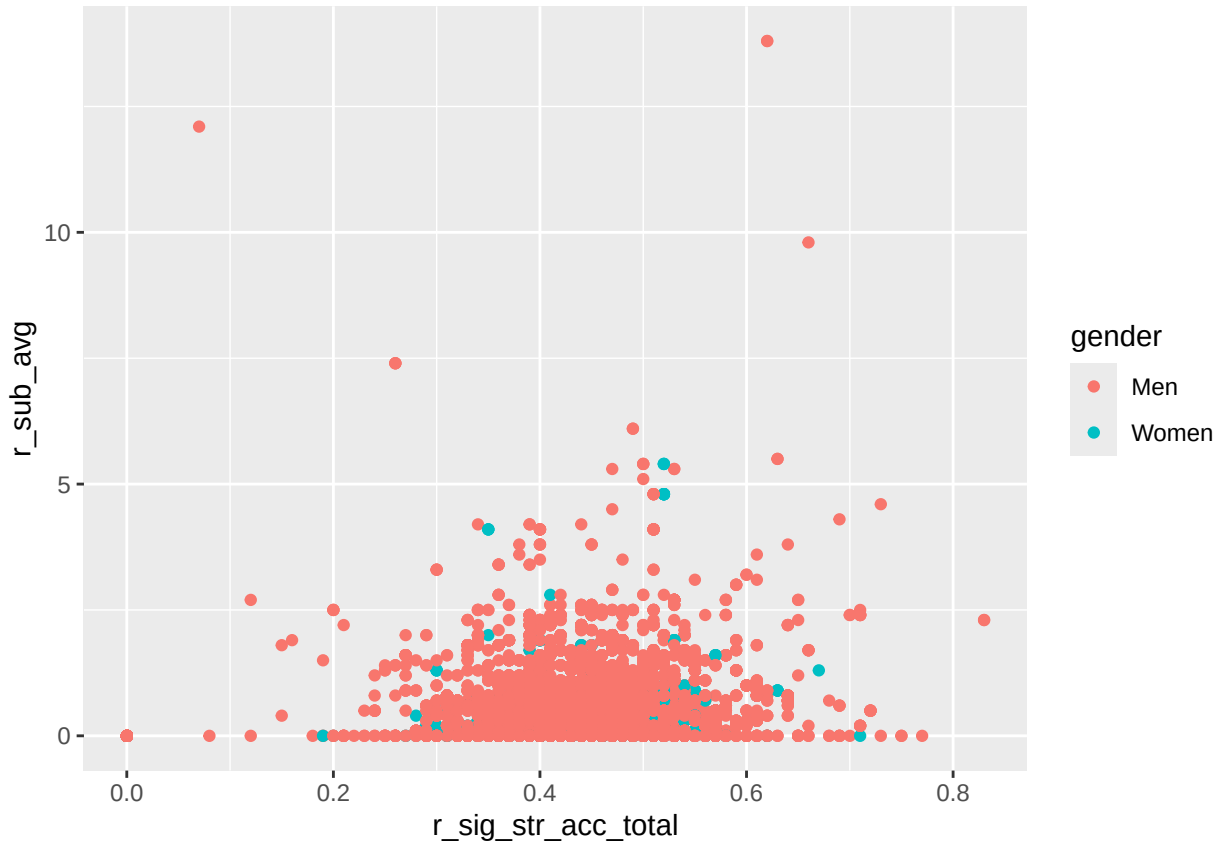
```
## [1] "total_rounds" "r_age"        "r_reach"      "b_age"        "b_reach"
## [6] "age_diff"     "reach_diff"
```

There is missing data for total rounds in a fight, ages of Red and Blue fighters and differences in age, and reaches of Red and Blue fighters and differences in reaches. We will focus on using other fighter and fight attributes in our model.

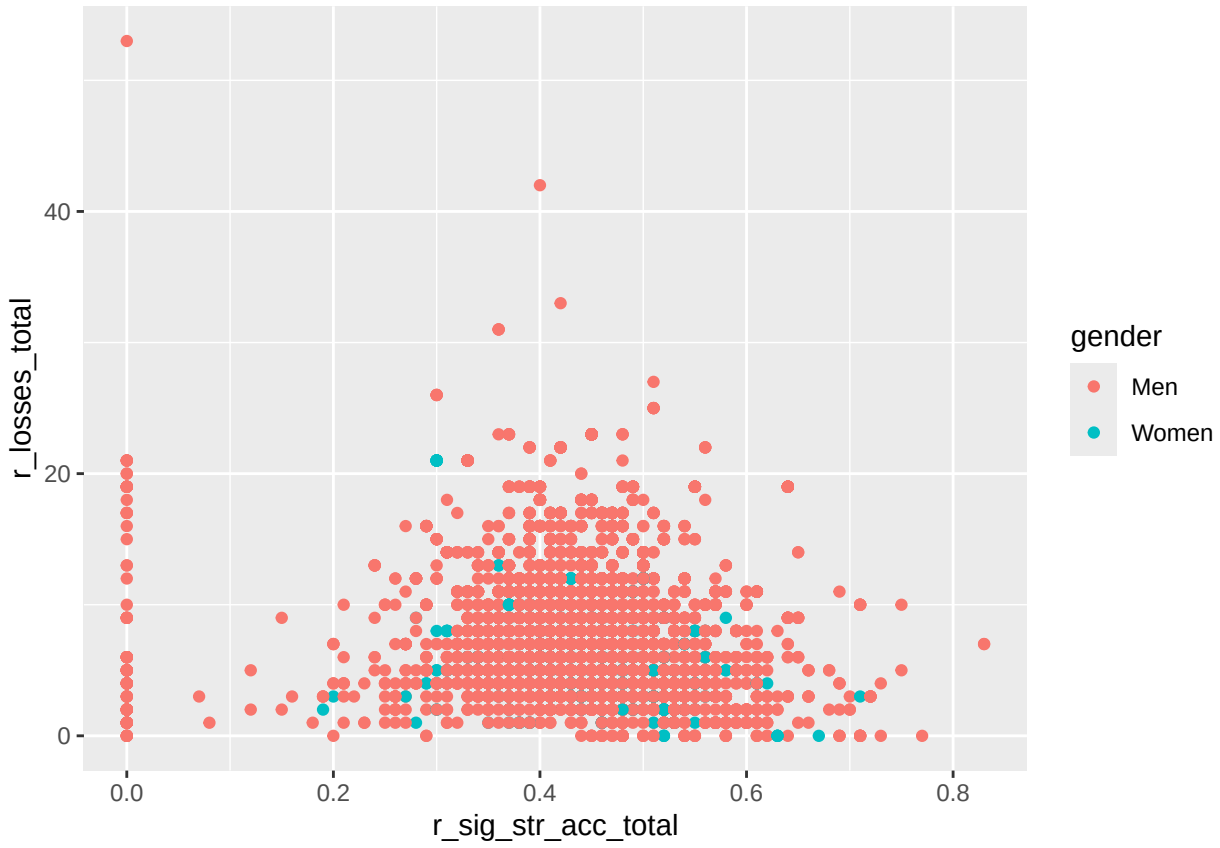
```
ggplot(data = data, mapping = aes(x=r_td_avg, y = r_sub_avg)) +geom_point()
```



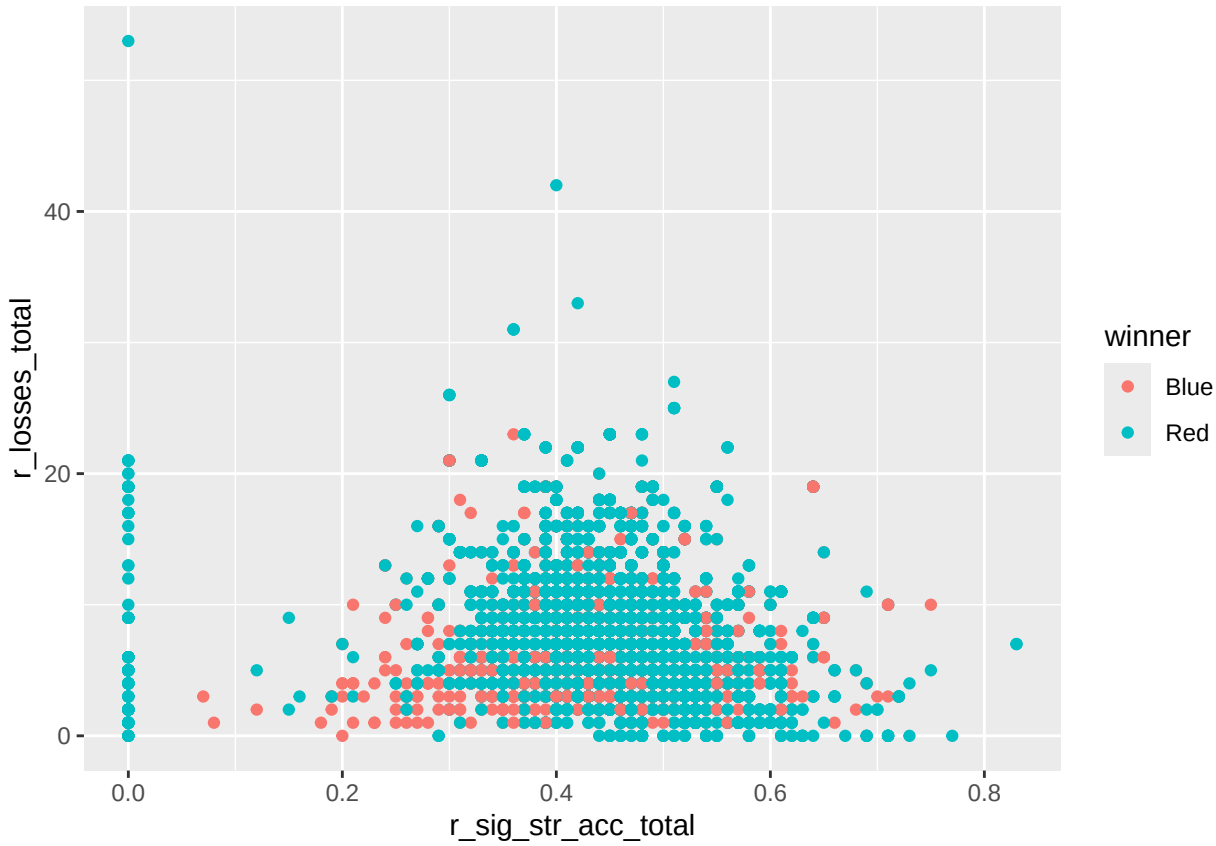

```
ggplot(data = data, mapping = aes(x=r_sig_str_acc_total, y = r_sub_avg, colour = gender)) +geom_point()
```



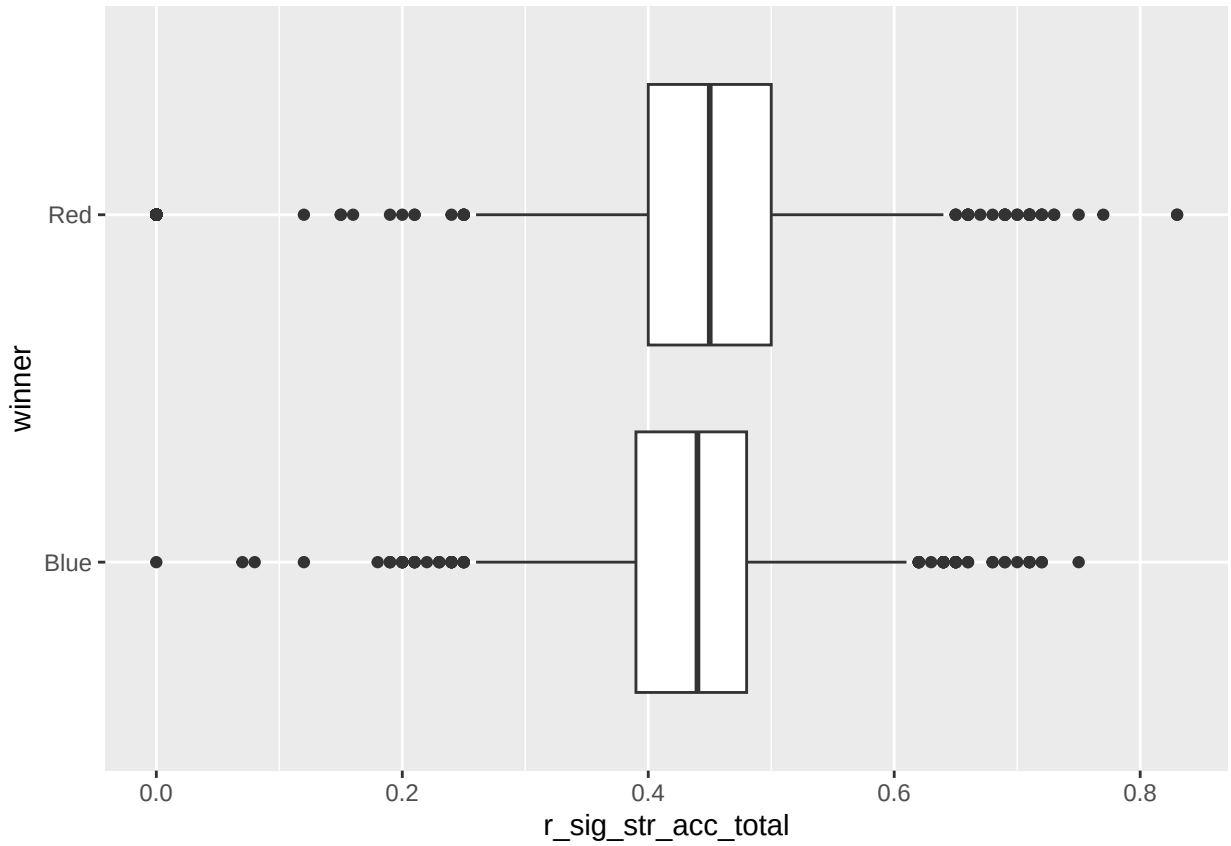

```
ggplot(data = data, mapping = aes(x=r_sig_str_acc_total, y = r_losses_total, colour = gender)) +geom_po
```



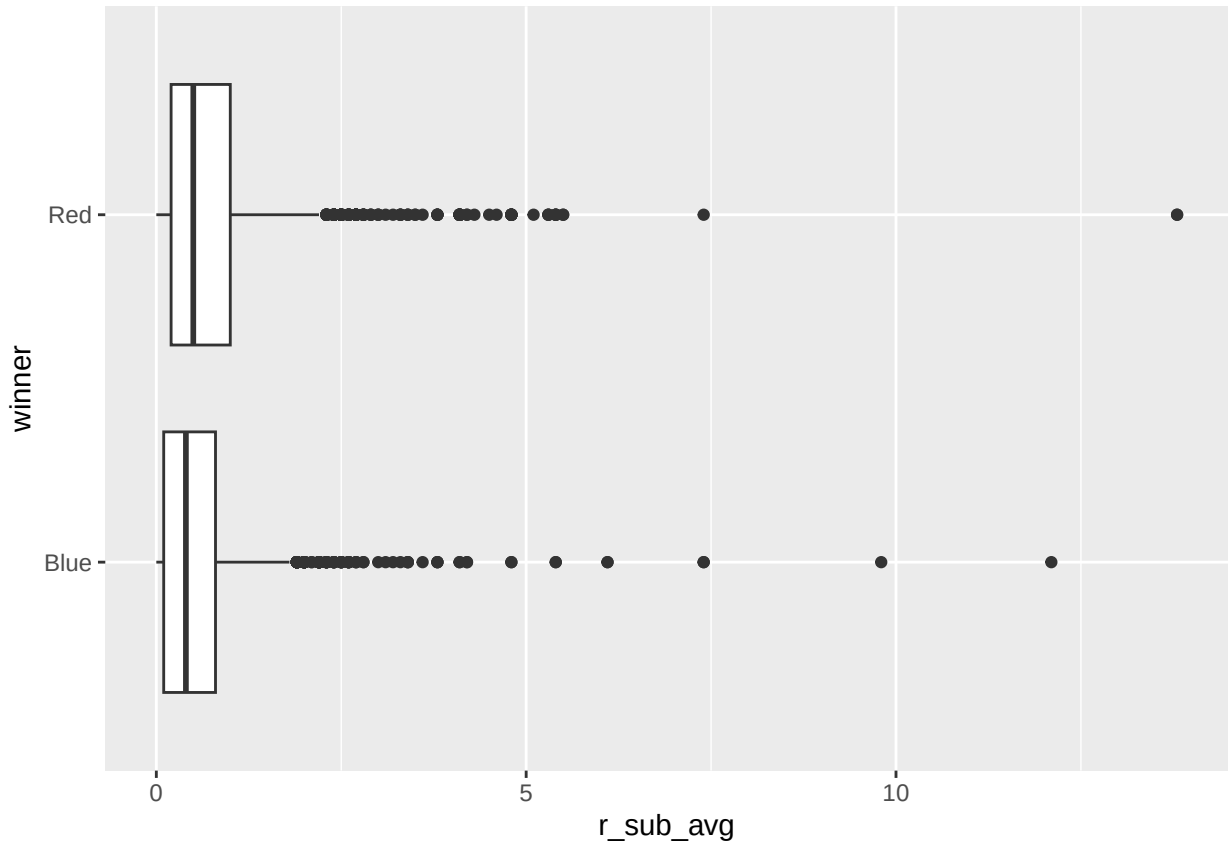

```
ggplot(data = data, mapping = aes(x=r_sig_str_acc_total, y = r_losses_total, colour = winner)) +geom_poi
```



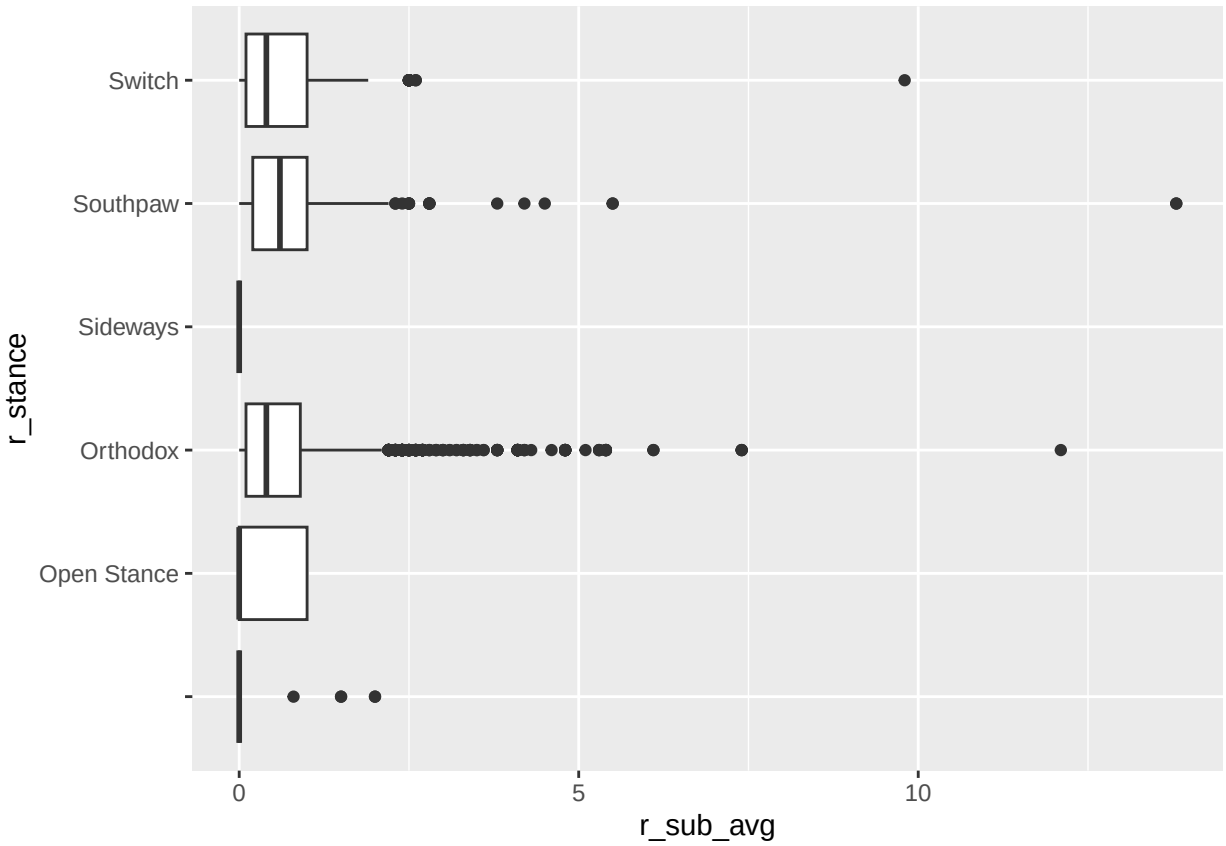

```
ggplot(data = data, mapping = aes(x=r_sig_str_acc_total, y = winner)) + geom_boxplot()
```




```
ggplot(data = data, mapping = aes(x=r_sub_avg, y = winner)) + geom_boxplot()
```




```
ggplot(data = data, mapping = aes(x=r_sub_avg, y = r_stance)) + geom_boxplot()
```



We do not see significant covariance present in important variables in our model.