Project

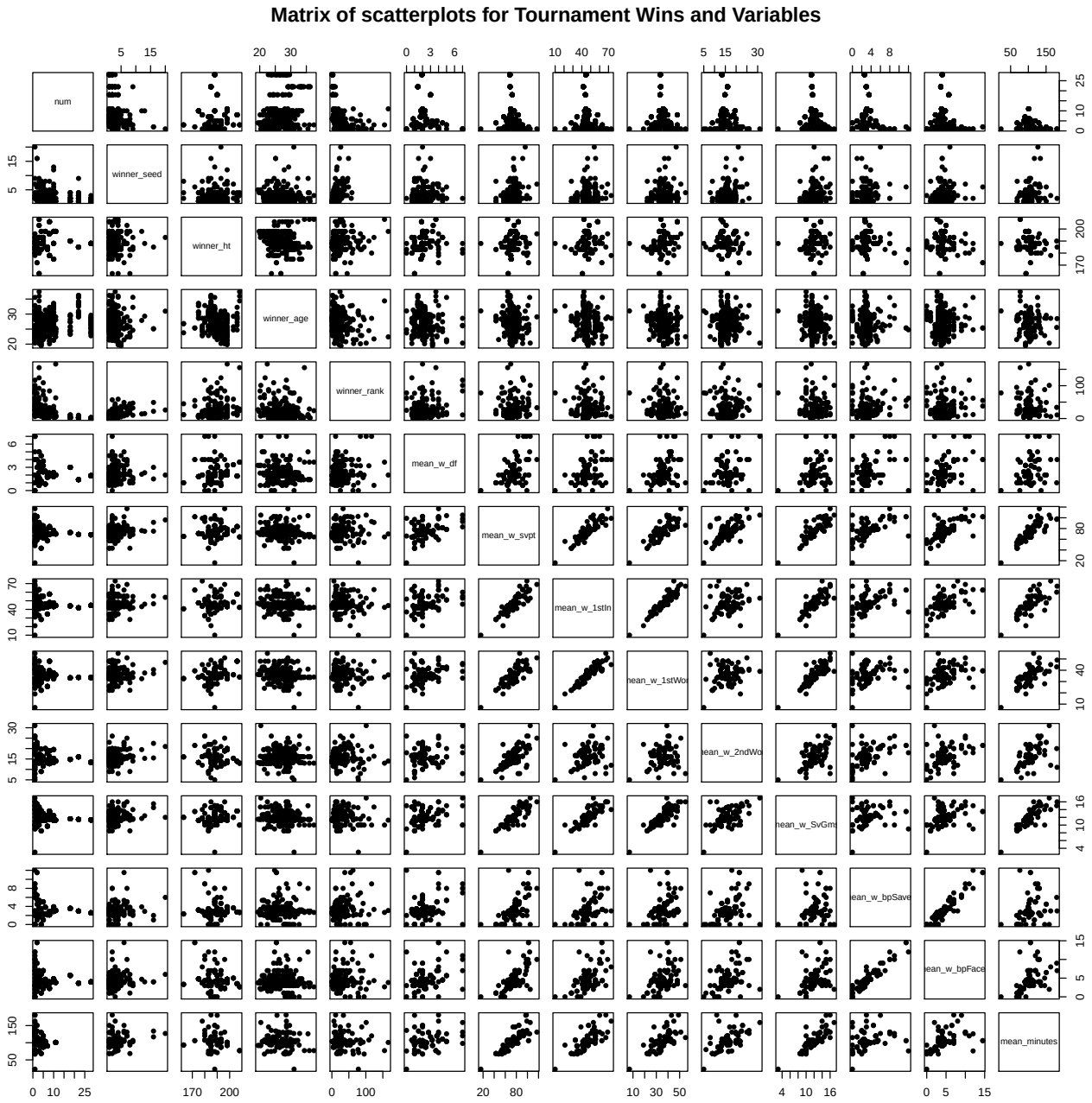
Steven Herrera and Ethan Shen

11/09/2018

```
library(tidyverse)
library(olsrr)
library(cowplot)
library(car)
```

Exploratory Data Analysis

```
pairs(num ~ winner_seed + winner_ht + winner_age + winner_rank +
        mean_w_df + mean_w_svpt + mean_w_1stIn + mean_w_1stWon +
        mean_w_2ndWon + mean_w_SvGms + mean_w_bpSaved + mean_w_bpFaced +
        mean_minutes, data=hard, pch = 16,
      main = "Matrix of scatterplots for Tournament Wins and Variables")
```



Linear Regression

To begin our regression models, we will

```
#backward <- ols_step_backward_aic(testmodel)
```

Linear Regression Assumptions

Hypothesis Tests + Confidence Intervals

Conclusion

```
hard
```

```
## # A tibble: 270 x 41
## # Groups:   winner_name [65]
##    tourney_id tourney_name surface draw_size tourney_level tourney_date
##    <chr>      <chr>        <chr>       <int> <chr>                <int>
##  1 2017-M020  Brisbane     Hard           32 A                 20170102
##  2 2017-0891  Chennai      Hard           32 A                 20170102
##  3 2017-0451  Doha         Hard           32 A                 20170102
##  4 2017-0301  Auckland     Hard           32 A                 20170109
##  5 2017-0375  Montpellier  Hard           32 A                 20170206
##  6 2017-7434  Sofia        Hard           32 A                 20170206
##  7 2017-0402  Memphis      Hard           32 A                 20170213
##  8 2017-0407  Rotterdam    Hard           32 A                 20170213
##  9 2017-0496  Marseille    Hard           32 A                 20170220
## 10 2017-M004  Acapulco     Hard           32 A                 20170227
## # ... with 260 more rows, and 35 more variables: match_num <int>,
## #   winner_id <int>, winner_seed <int>, winner_entry <chr>,
## #   winner_name <chr>, winner_hand <chr>, winner_ht <int>,
## #   winner_ioc <chr>, winner_age <dbl>, winner_rank <int>,
## #   winner_rank_points <int>, score <chr>, best_of <int>, round <chr>,
## #   minutes <int>, w_ace <int>, w_df <int>, w_svpt <int>, w_1stIn <int>,
## #   w_1stWon <int>, w_2ndWon <int>, w_SvGms <int>, w_bpSaved <int>,
## #   w_bpFaced <int>, mean_w_ace <dbl>, mean_w_df <dbl>, mean_w_svpt <dbl>,
## #   mean_w_1stIn <dbl>, mean_w_1stWon <dbl>, mean_w_2ndWon <dbl>,
## #   mean_w_SvGms <dbl>, mean_w_bpSaved <dbl>, mean_w_bpFaced <dbl>,
## #   mean_minutes <dbl>, num <int>
```

```
testhard <- lm(num ~ winner_seed + winner_ht + winner_age + winner_rank + mean_w_df + mean_w_svpt + mea
```

```
backward <- ols_step_backward_aic(testhard)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . winner_seed
## 2 . winner_ht
## 3 . winner_age
## 4 . winner_rank
## 5 . mean_w_df
## 6 . mean_w_svpt
## 7 . mean_w_1stIn
## 8 . mean_w_1stWon
## 9 . mean_w_2ndWon
## 10 . mean_w_SvGms
```

```
## 11 . mean_w_bpSaved
## 12 . mean_w_bpFaced
## 13 . mean_minutes
##
##
## Variables Removed:
##
## - mean_w_bpSaved
## - mean_w_1stIn
## - mean_w_df
##
## No more variables to be removed.
```