STA531 Final Project-Preliminary Report

Eric Su

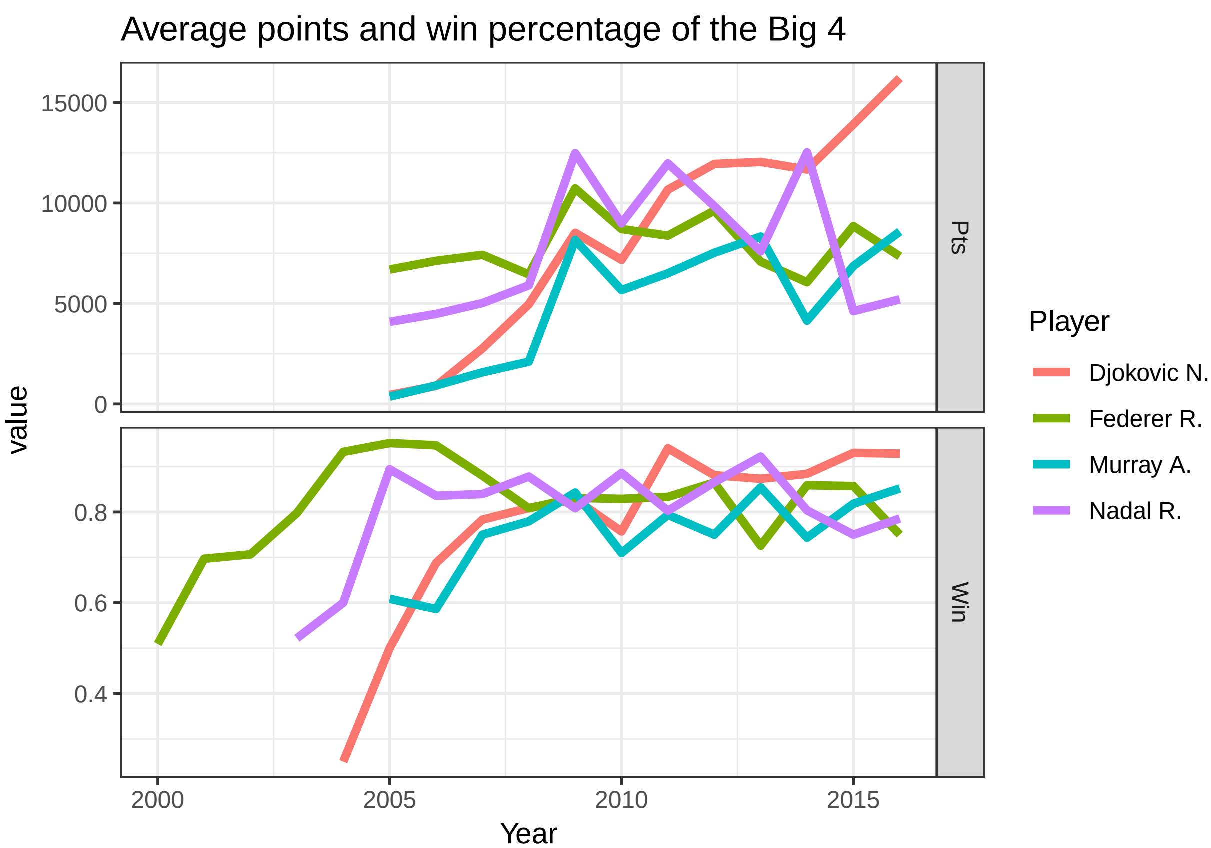
2019-04-08

R Markdown

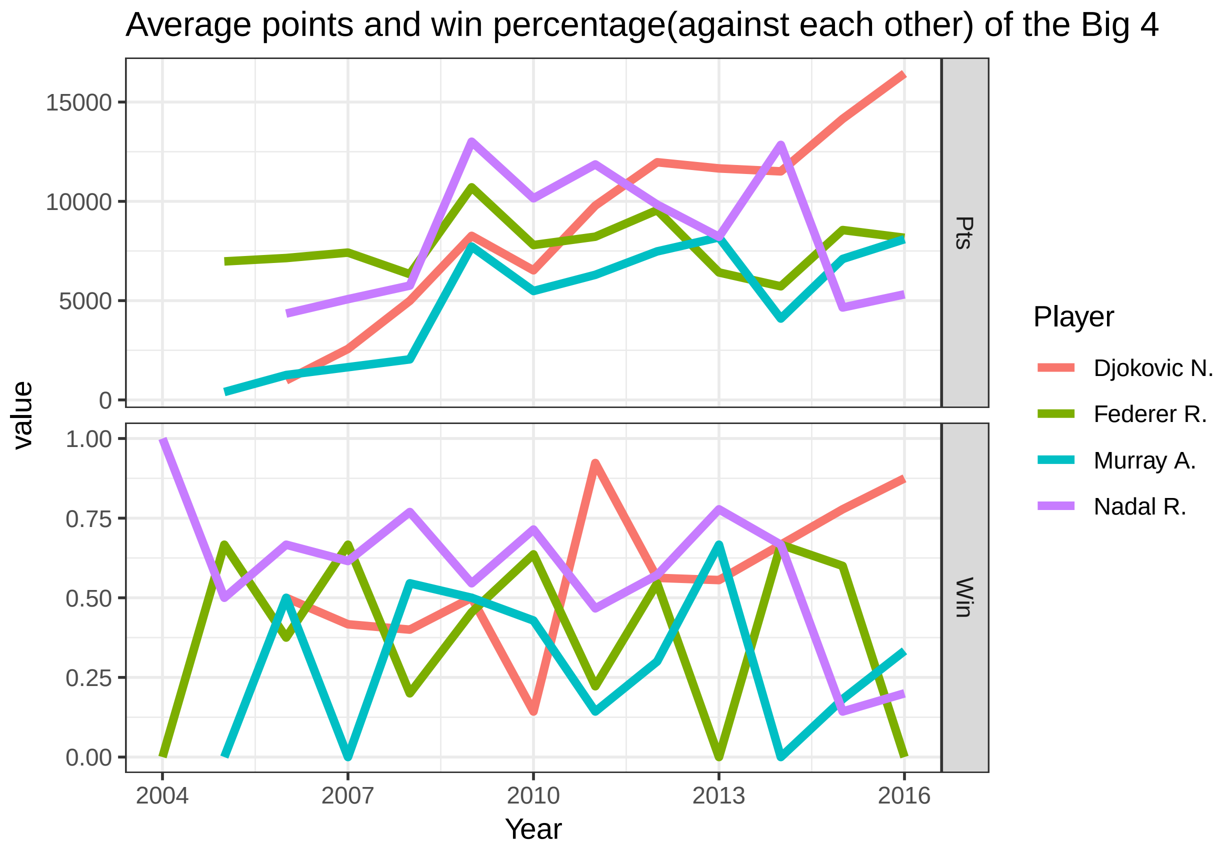
```
## # A tibble: 5 x 3
##   Series         Pts  Rank
##   <chr>        <dbl> <dbl>
## 1 ATP250        960.  93.4
## 2 ATP500       1419.  75.9
## 3 Grand Slam   1773.  68.9
## 4 Masters 1000 2060.  46.5
## 5 Masters Cup  5201.   4.95
```

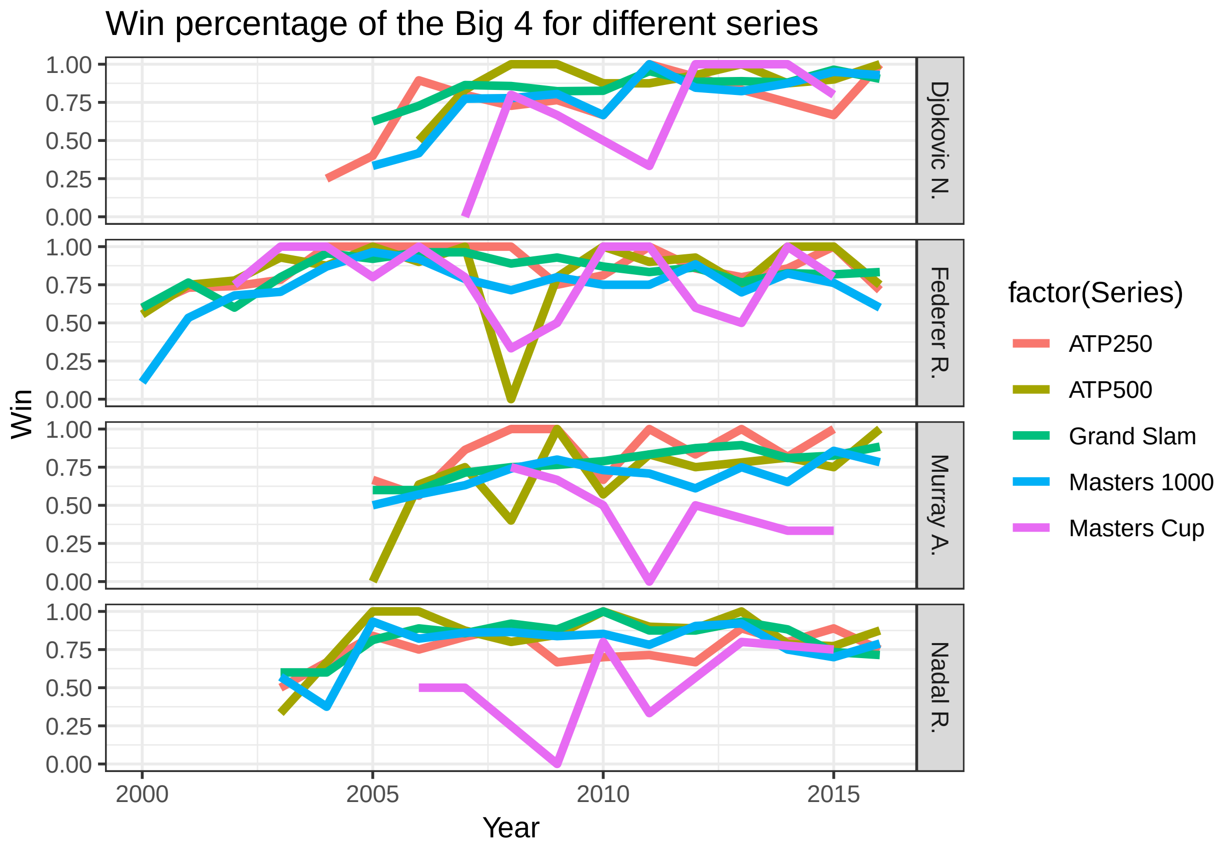
```
##                 Pts     logPts       Rank         WL
## Pts       1.0000000  0.7740350 -0.3450371  0.2156367
## logPts    0.7740350  1.0000000 -0.8023401  0.2519962
## Rank     -0.3450371 -0.8023401  1.0000000 -0.1754293
## WL        0.2156367  0.2519962 -0.1754293  1.0000000
```

```
##           Win       Pts
## Win 1.0000000 0.6515231
## Pts 0.6515231 1.0000000
```



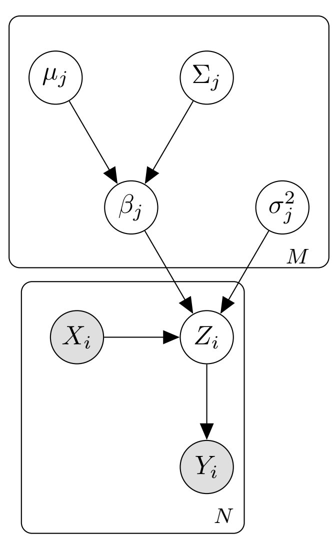
```
##             Win       Pts
## Win 1.0000000 0.4765591
## Pts 0.4765591 1.0000000
```





Model

For each match, the model can be expressed using the graph below.



where matches are represented using the index $i = 1, ..., N$ and players are represented using the index

$j = 1, ..., M$ with variables

$$\begin{aligned}
X_i &: \text{Vector of external conditions (Series, Court, Surface, Round, Best of 3/4, Opponent rank)} \\
Y_i &: \text{Match outcome} \\
Z_{i,1:2} &: \text{Performance level of the two players} \\
\beta_j &: \text{Vector of regression coefficients} \\
\sigma_j^2 &: \text{Variance parameter for performance} \\
\mu_j &: \text{Mean hyperparameter for } \beta_j \\
\Sigma_j &: \text{Variance hyperparameter for } \beta_j
\end{aligned}$$

The outcome of each match will be modelled using the distribution:

$$Y_i = \begin{cases} 1 & \text{if } Z_{i,1} \geq Z_{i,2} \\ 0 & \text{if } Z_{i,1} < Z_{i,2} \end{cases}$$

For player j, his performance level in match i will be a linear combination of X_i(external factors) as shown below.

$$Z_{i,1 \text{ or } 2}^{(j)} = X_i^T \beta_j + \epsilon_{i,j} \quad \epsilon_{i,j} \sim N(0, \sigma_j^2)$$

with

$$\beta_j \sim N(\mu_j, \Sigma_j)$$

The prior distributions for this model will be

$$\mu_j \sim$$

The parameters of our model would be estimated using a Gibbs sampler with full conditionals as follows.

$$\begin{aligned}
p(Z_{i,1}^{(j)} \mid X_i, Y_i, Z_{i,2}, \beta_j, \sigma_j^2) &= \begin{cases} \text{Truncated Normal}(X_i^T \beta_j, \sigma_j^2, Z_{i,2}, \infty) & \text{if } Y_i = 1 \\ \text{Truncated Normal}(X_i^T \beta_j, \sigma_j^2, -\infty, Z_{i,2}) & \text{if } Y_i = 0 \end{cases} \\
p(Z_{i,2}^{(j)} \mid X_i, Y_i, Z_{i,1}, \beta_j, \sigma_j^2) &= \begin{cases} \text{Truncated Normal}(X_i^T \beta_j, \sigma_j^2, -\infty, Z_{i,1}) & \text{if } Y_i = 1 \\ \text{Truncated Normal}(X_i^T \beta_j, \sigma_j^2, Z_{i,1}, \infty) & \text{if } Y_i = 0 \end{cases} \\
p(\beta_j \mid) & \\
p(\mu_j \mid) &
\end{aligned}$$