Stat 400 Project - Group 4

Megan Dunnahoo, Jasmine DeMeyer, Macey Dodd

Article

We chose the article "Using Monte Carlo Simulation to Calculate Match Importance: The Case of English Premier League" by Jiri Lahvicka. This article describes the process of using Monte Carlo simulations to predict the outcome of a match given the results of previous matches. It then goes further and uses Monte Carlo simulations to predict the final ranking of the teams in the English Premier League at the end of a season.

Soccer Information

The English Premier League is regarded as the most popular sports league in the world due to its massive audience views and impressive revenue. There are 20 teams in the English Premier League. Manchester United is considered to be the most popular football club with the Liverpool club in second. In football, a game can result in a tie as well as a win or a loss. Three points are awarded for a win, one for a draw and zero for a loss. At the end of each season, the lowest ranking three teams will be "relegated" or demoted to the lower football league, the English Football League (EFL). The highest three ranking clubs in the English Football League will be "promoted" into the Premier League.

Data and Code Setup

We wanted to use the specific variables FTR, FTAG, and FTHG, along with the identifier variables of Date, Away Team, and Home Team. There were no NA values in any of our selected variables. We used four seasons ranging from years 2011-2015. Each team played 19 away games and 19 home games.
We created functions to get the points scored and the outcome of the team for each team over the four years, specifying away or home games.

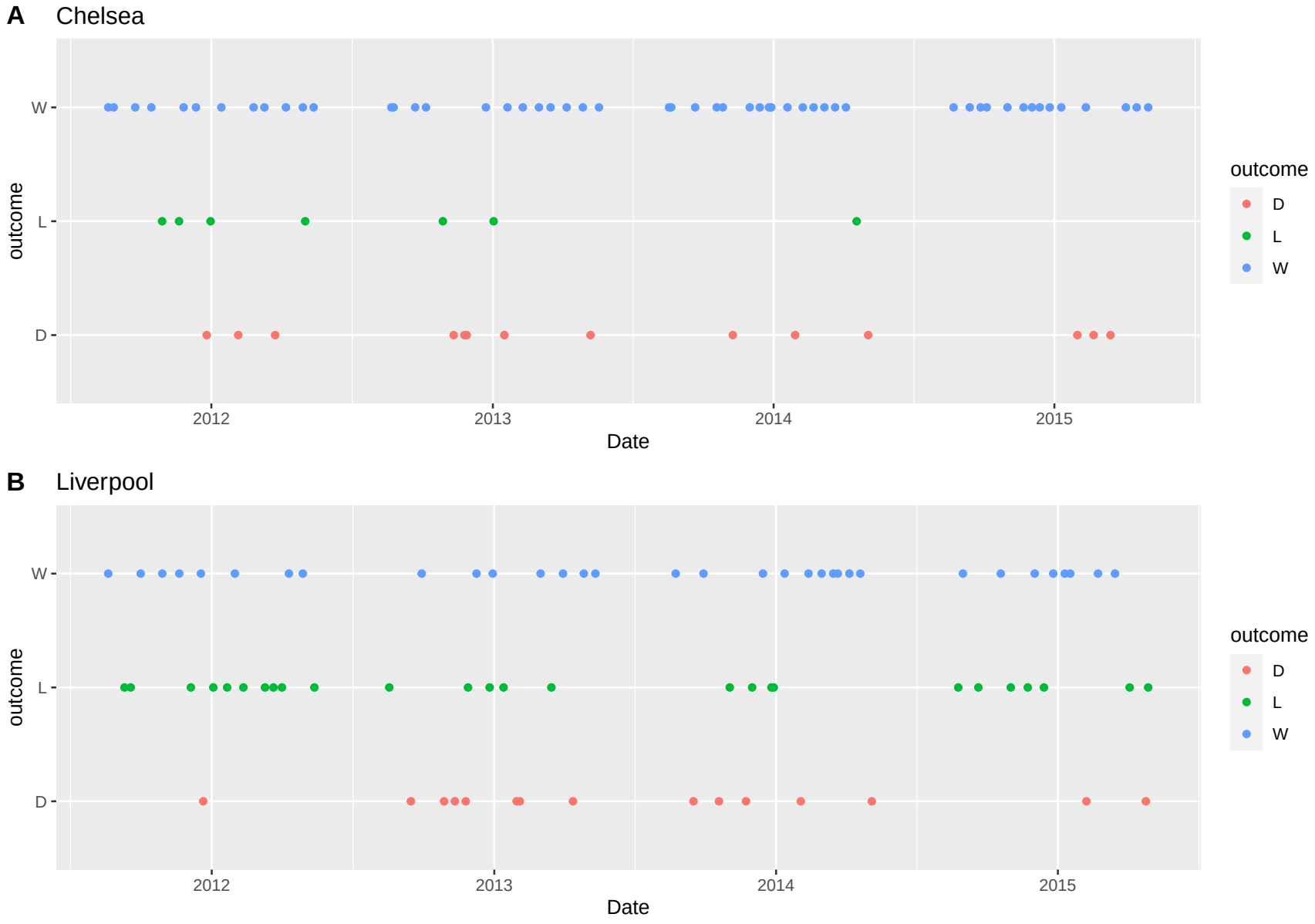
MC Estimation

We estimated the lambda home and lambda away values, which are the expected goals scored by the home and away team respectively, using Monte Carlo. These lambda values are assumed to be independent Poisson distributed variables and are calculated using the last 19 matches for each team.

$$\lambda_{home} = \frac{\text{Average goals scored by home team} + \text{Average goals conceded by away team}}{2}$$

$$\lambda_{away} = \frac{\text{Average goals scored by away team} + \text{Average goals conceded by home team}}{2}$$

Exploratory Plot of Outcomes Before Match (Chelsea vs Liverpool)

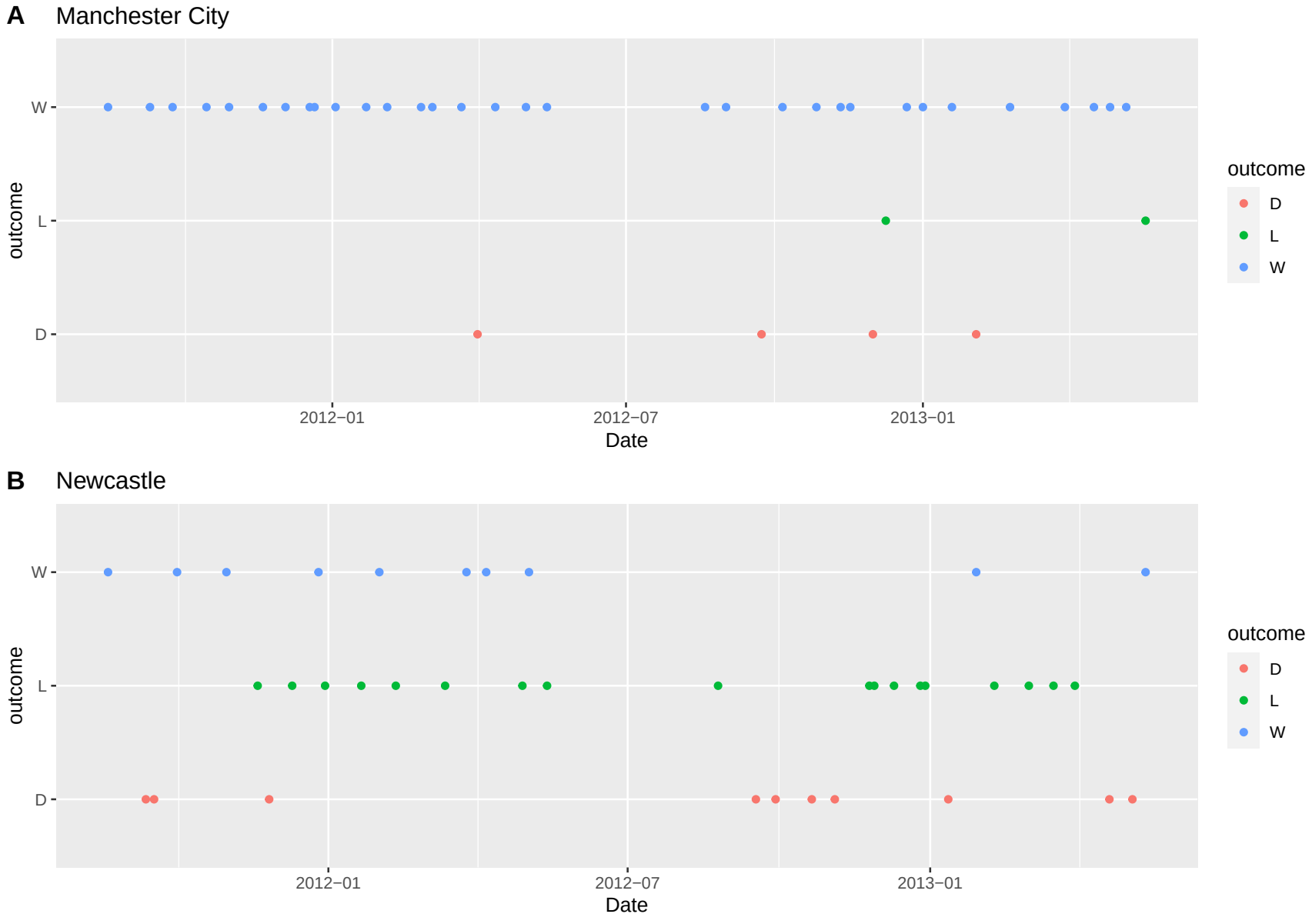


Prediction for Match 2014-2015 Chelsea (home) vs Liverpool (away)

Predictions are in terms of the home team
Actual result was a draw

```
##   Match Result Occurances Percent
## 1           Win      51999  51.999
## 2          Loss      29180  29.180
## 3          Draw      18821  18.821
```

Exploratory Plot of Outcomes Before Match (Man City vs Newcastle)



Prediction for Match 2013-2014 Manchester City (home) vs Newcastle (away)

Predictions are in terms of the home team
Actual result was a home win

	Match Result	Occurances	Percent
1	Win	60084	60.084
2	Loss	23188	23.188
3	Draw	16728	16.728

Going Forward

In the next week we plan to continue to expand on our exploration of this data. We plan to integrate code into our Monte Carlo function to predict the final rank of any two teams we have chosen at the time of a specific match. We also hope to add to our final contingency table a variable that consists of the probability of the ranks for these two teams to add to the number of wins, losses, and draws they had in our simulations.