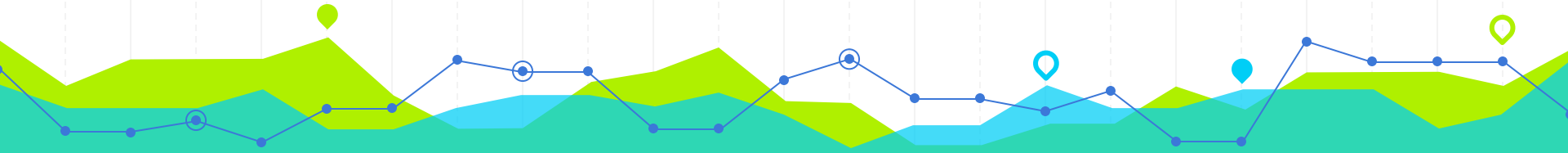


DSC 101 PROJECT

Aisha Mahmoud & Diego Mendez

3rd Sprint Review

ACHIEVEMENTS

- What did we want to achieve in the last sprint?

  After coming up with our plan, and method in the past sprints, it was time to actually execute it. We wanted to write all the code and find out the relevant stats for each player.

- What did we achieve in the last sprint?

We achieved loading the data set, refining the data, splitting it up into three separate data frames (attackers, midfielders, and defenders), and keeping only the relevant statistics for each category. We also found the results of who had the best stats in each category. After figuring out this “formula”, it will be easier to do it for the rest of the data set we haven’t finished yet.

ACHIEVEMENTS

- What are we especially proud of in the last sprint?

  We’re especially proud of getting to where we are, it seems daunting at first, and it’s one thing to do it on DataCamp but another to do it on your own with no prompts. It went better than expected!

PROBLEMS WE RAN INTO?

- What did we not achieve in the last sprint?

We didn’t get to plotting the EPL results, or finding all the stats for the American league. (Now that we figured out all the trials and errors with the EPL dataset, it should be much easier and faster to do the American league one.

- What are we going to do different in the next sprint?

In the next sprint, we’ll be more organized, and now that we know what to do, be more efficient.

CRITERIA FOR PLAYERS STATS

Defense:

- challenges won
- ball interceptions
- tackles
- ball recoveries

Midfield:

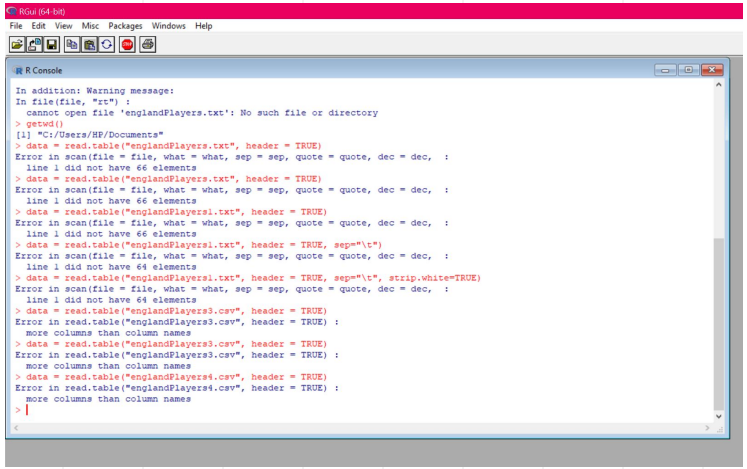
- assists
- accurate pass percentage
- key passes
- dribbles

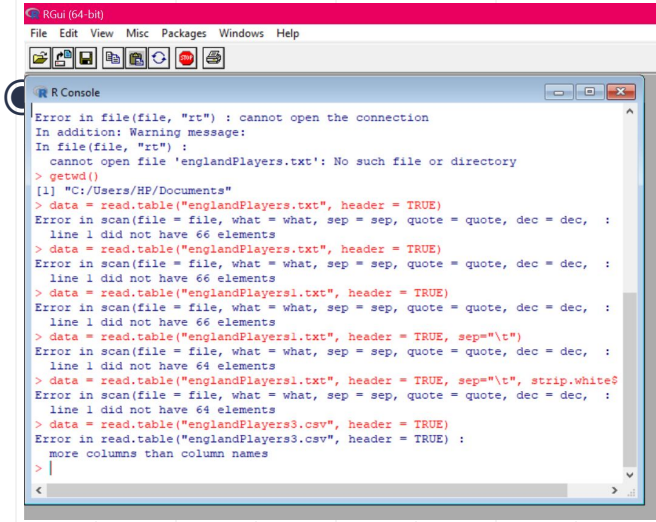
Forward:

- expected goals
- goals
- percentage of shots on target
- dribbles

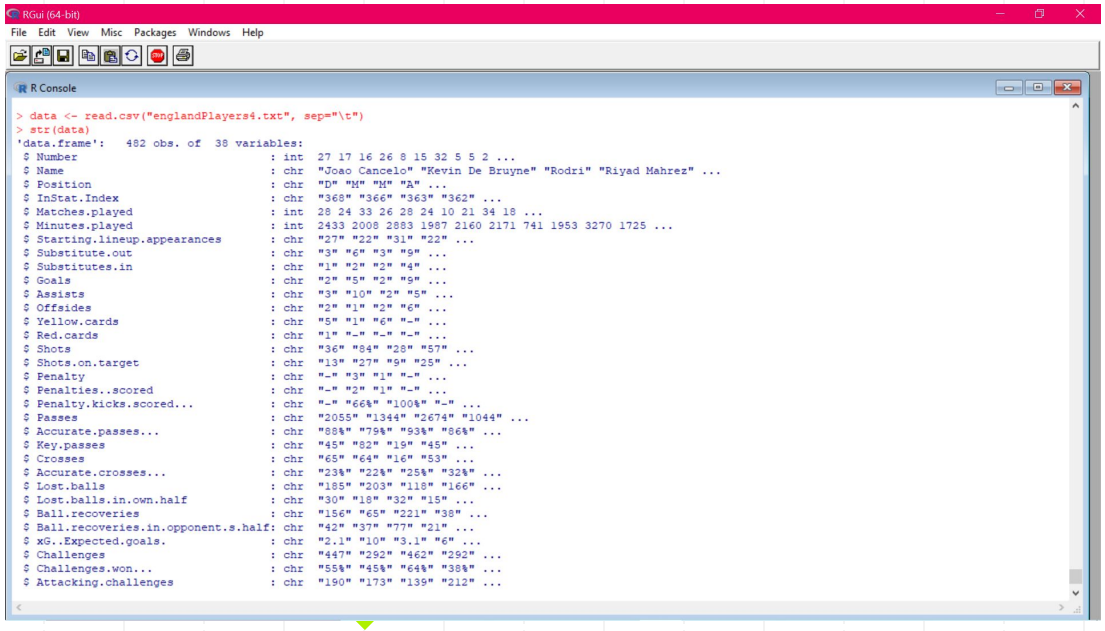
ERRORS WE RAN INTO

Started out in R. For some reason, had a lot of trouble trying to read in the file. It took a lot of trying different things to find out what worked.





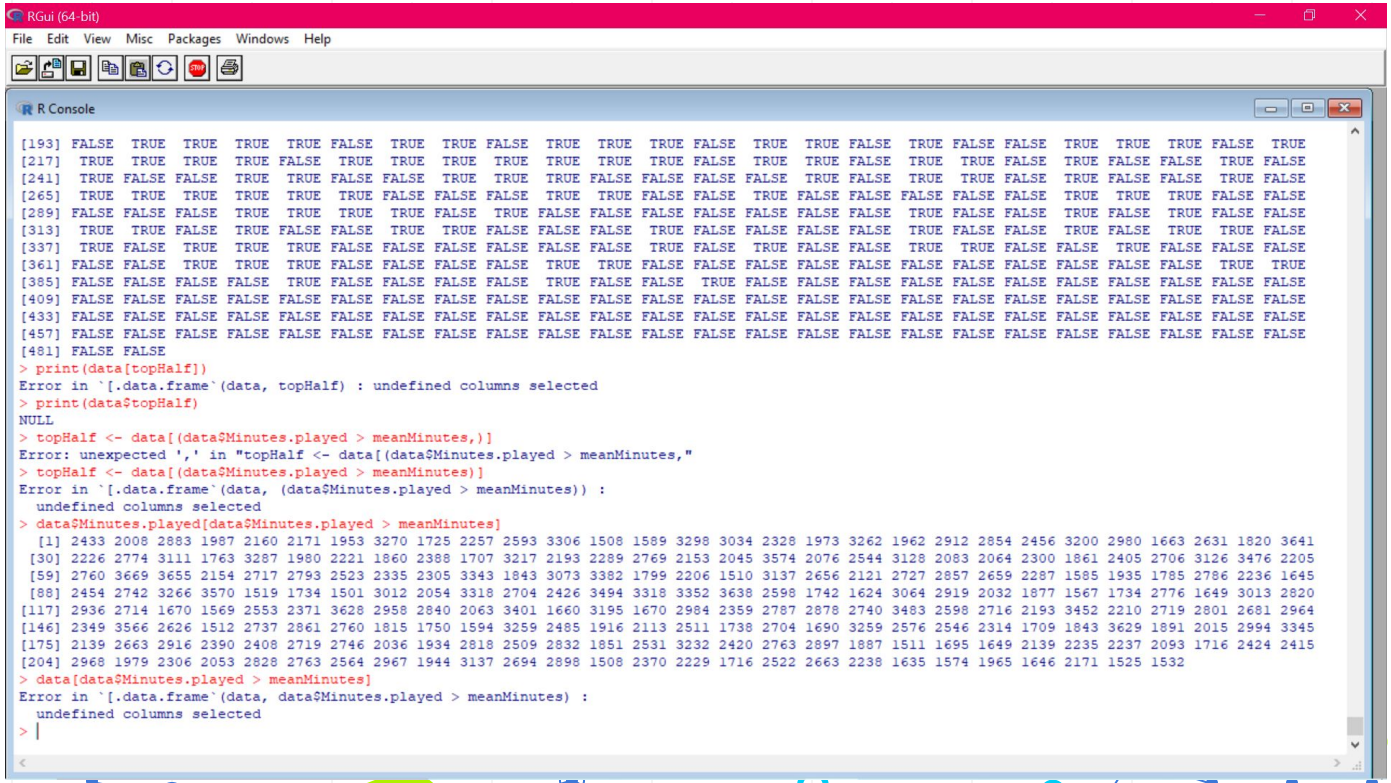
cont.

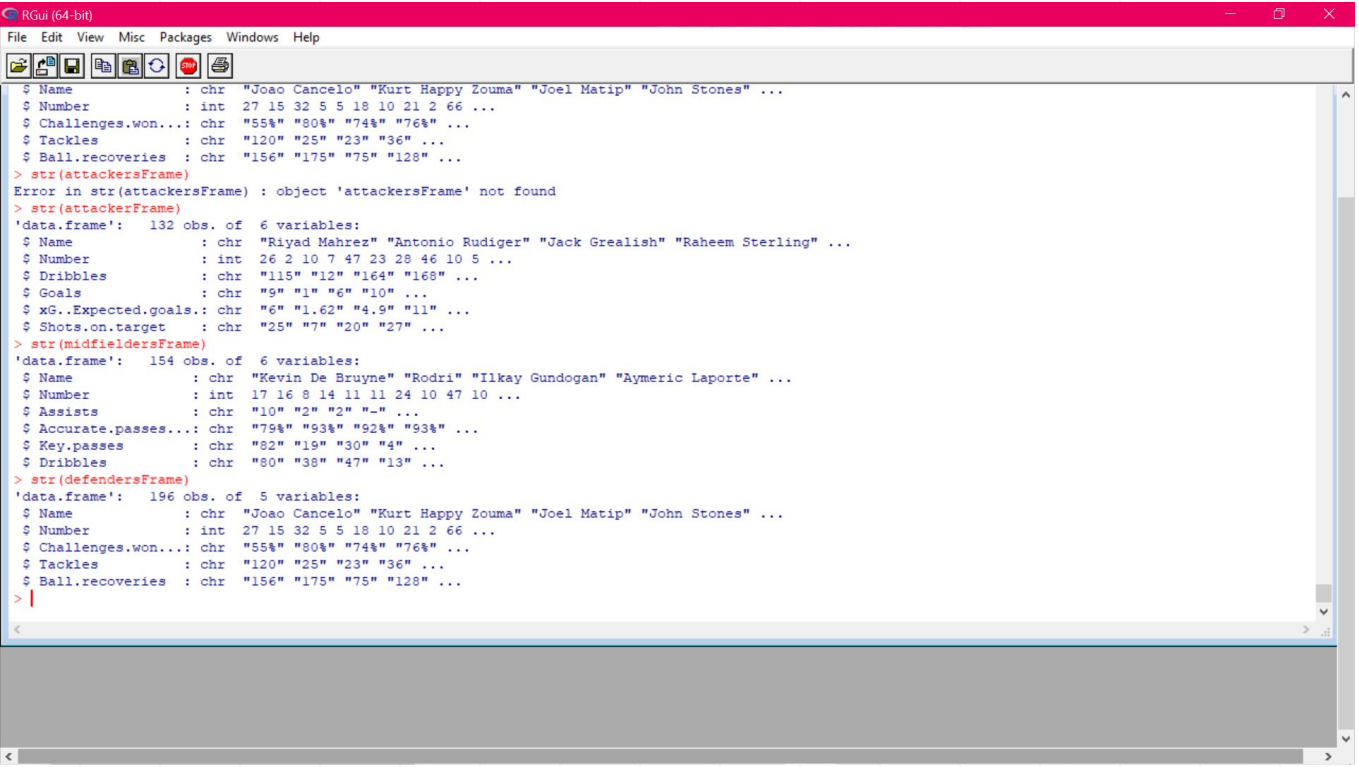


Finally got it to work!

Now time for working with the actual data...

Next came trying to extract values. We only wanted the players that have played more than the avg amount of minutes played, a.k.a. the top half. We remembered we did something like this in class (with hotdogs…?)





Separating the values. One logic error we had here is that we realized we took players from our big original dataset and put them into the data frames for each position, instead of the refined top half group of players. This was later fixed.

RESULTS

-moved everything into collaboratory

-here are some results we've gotten:

DEFENDERS

```
str(defendersFrame)
```

```
'data.frame':   92 obs. of  5 variables:
 $ Name           : chr  "Joao Cancelo" "Kurt Happy Zouma" "John Stones" "Harry Maguire" ...
 $ Number         : int  27 15 5 5 2 21 2 66 6 3 ...
 $ Challenges.won...: chr  "55%" "80%" "76%" "72%" ...
 $ Tackles        : int  120 25 36 57 41 72 56 99 34 46 ...
 $ ballRecoveries : int  156 175 128 305 143 104 132 195 197 217 ...
```

Who made the most tackles?

```
[9] defendersFrame[which.max(defendersFrame$Tackles),]
```

A data.frame: 1 × 5

	Name	Number	Challenges.won...	Tackles	ballRecoveries
	<chr>	<int>	<chr>	<int>	<int>
175	Luke Ayling	2	60%	186	308

Which defender had the most ball recoveries?

```
[10] defendersFrame[which.max(defendersFrame$ballRecoveries),]
```

A data.frame: 1 × 5

	Name	Number	Challenges.won...	Tackles	ballRecoveries
	<chr>	<int>	<chr>	<int>	<int>
139	Jan Bednarek	35	67%	75	329

MIDFIELDERS

0s completed

ATTACKERS

```
[ ] #finding who had the most:("Dribbles","Goals", "expectedGoals", "Shots.on.target")
```

```
print("The attacker with the most amount of goals:")
attackersFrame[which.max(attackersFrame$Goals),]

print("\nThe attacker with the most amount of expected goals:")
attackersFrame[which.max(attackersFrame$expectedGoals),]

print("\n\nThe attacker with the most amount of dribbles:")
attackersFrame[which.max(attackersFrame$Dribbles),]

print("\nThe attacker with the most amount of shots on target:")
attackersFrame[which.max(attackersFrame$Shots.on.target),]

#why does \n not work for making a new line?
```

[1] "The attacker with the most amount of goals:"

A data.frame: 1 × 6

	Name	Number	Dribbles	Goals	expectedGoals	Shots.on.target
	<chr>	<int>	<chr>	<int>	<chr>	<chr>
14	Mohamed Salah	11	134	22	20	52

[1] "\nThe attacker with the most amount of expected goals:"

A data.frame: 1 × 6

	Name	Number	Dribbles	Goals	expectedGoals	Shots.on.target
	<chr>	<int>	<chr>	<int>	<chr>	<chr>
14	Mohamed Salah	11	134	22	20	52

[1] "\n\nThe attacker with the most amount of dribbles:"

A data.frame: 1 × 6

	Name	Number	Dribbles	Goals	expectedGoals	Shots.on.target
	<chr>	<int>	<chr>	<int>	<chr>	<chr>
190	Adama Traore	37	329	2	3.2	13

[1] "\nThe attacker with the most amount of shots on target:"

A data.frame: 1 × 6

	Name	Number	Dribbles	Goals	expectedGoals	Shots.on.target
	<chr>	<int>	<chr>	<int>	<chr>	<chr>
26	Harry Kane	10	119	22	20	53

Here we see that Mo Salah had both the highest number of expected goals, and the highest number of goals. So he really lived up to his expectations.

Questions for the product owner...

Are these results so far what you expected?

What's your favorite color so we can make that the color of one of our histograms? :)

Questions for other team members...

What do you think was the hardest part about all the coding done so far in this project?

How long did it take for you to load in the data set? Did it go smoother for you?

Stay tuned for the next sprint review!