

Visual Analysis of Preferred Foot in European Soccer Leagues

Jacob Miller

Dataset(s)

Which dataset did you use of the following:

- **Soccer Dataset (<https://www.kaggle.com/hugomathien/soccer>)**
- ~~— IMDB Movie Dataset~~
- ~~— World Development Indicators Dataset~~

Motivation

I am interested in determining if there is a correlation between left/right footed soccer players and other foot skills. In other sports (e.g. baseball), left/right preference can have a significant impact on player performance (e.g. the strategic decision to put a left-handed baseball pitcher in against a batter who struggles against them). I am curious if this could be applied to the soccer world.

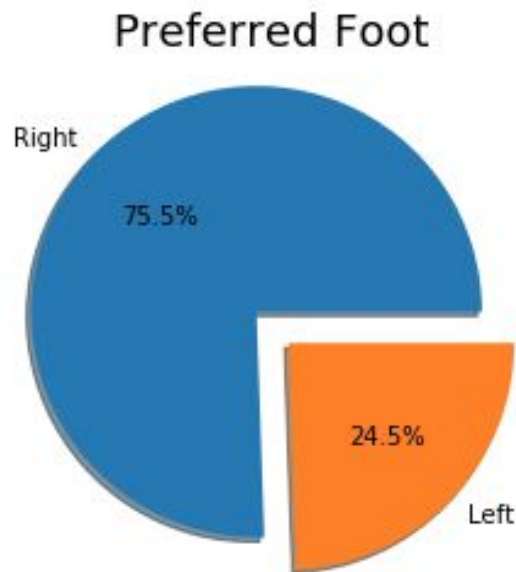
In the soccer world, this could be beneficial for coaches when deciding line ups or filling openings during a draft. It could also be useful for players when making key decisions on the field, for example who will be taking free kicks or penalty shots.

Research Question(s)

The question I am seeking to answer is:

Is there a positive, negative or neutral correlation between preferred foot (left/right) and other foot skills in the European soccer leagues?

Findings



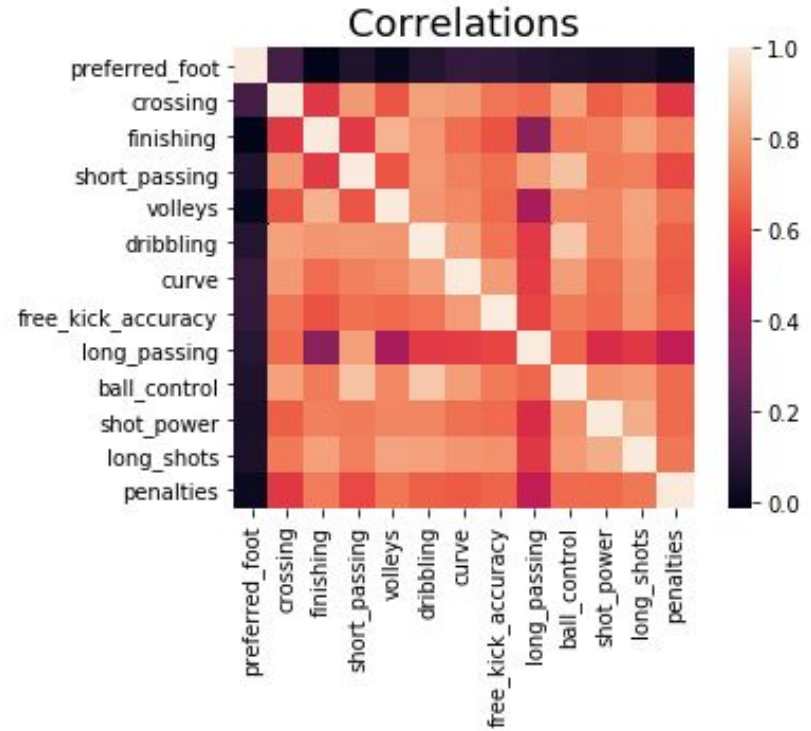
Out of 180,354 data points (after dropping NaN values), 75.5% of players are right-footed while 24.5% are left-footed

	preferred_foot	crossing	finishing	short_passing	volleys	dribbling	curve	free_kick_accuracy	long_passing	ball_control	shot_power	long_shots	penalties
preferred_foot	1.000000	0.174033	-0.012004	0.073302	0.011572	0.083831	0.128510	0.124475	0.094717	0.074678	0.057309	0.059462	0.016476
crossing	0.174033	1.000000	0.576896	0.790323	0.637527	0.809747	0.788924	0.708763	0.685649	0.807721	0.656740	0.716515	0.574208
finishing	-0.012004	0.576896	1.000000	0.580245	0.851482	0.784988	0.691082	0.633274	0.341121	0.720694	0.727835	0.806895	0.726234
short_passing	0.073302	0.790323	0.580245	1.000000	0.639995	0.788935	0.731948	0.693490	0.803073	0.890622	0.722320	0.729741	0.612511
volleys	0.011572	0.637527	0.851482	0.639995	1.000000	0.784247	0.752410	0.682909	0.414520	0.749459	0.746622	0.814894	0.713116
dribbling	0.083831	0.809747	0.784988	0.788935	0.784247	1.000000	0.810353	0.707322	0.579201	0.901730	0.744960	0.807175	0.663420
curve	0.128510	0.788924	0.691082	0.731948	0.752410	0.810353	1.000000	0.797842	0.586313	0.798598	0.694945	0.783732	0.649737
free_kick_accuracy	0.124475	0.708763	0.633274	0.693490	0.682909	0.707322	0.797842	1.000000	0.603286	0.720674	0.684191	0.773887	0.669018
long_passing	0.094717	0.685649	0.341121	0.803073	0.414520	0.579201	0.586313	0.603286	1.000000	0.675525	0.542244	0.576376	0.476750
ball_control	0.074678	0.807721	0.720694	0.890622	0.749459	0.901730	0.798598	0.720674	0.675525	1.000000	0.774856	0.793219	0.684410
shot_power	0.057309	0.656740	0.727835	0.722320	0.746622	0.744960	0.694945	0.684191	0.542244	0.774856	1.000000	0.838254	0.680887
long_shots	0.059462	0.716515	0.806895	0.729741	0.814894	0.807175	0.783732	0.773887	0.576376	0.793219	0.838254	1.000000	0.714596
penalties	0.016476	0.574208	0.726234	0.612511	0.713116	0.663420	0.649737	0.669018	0.476750	0.684410	0.680887	0.714596	1.000000

The values for correlating “preferred_foot” to other tracked foot skills

Findings

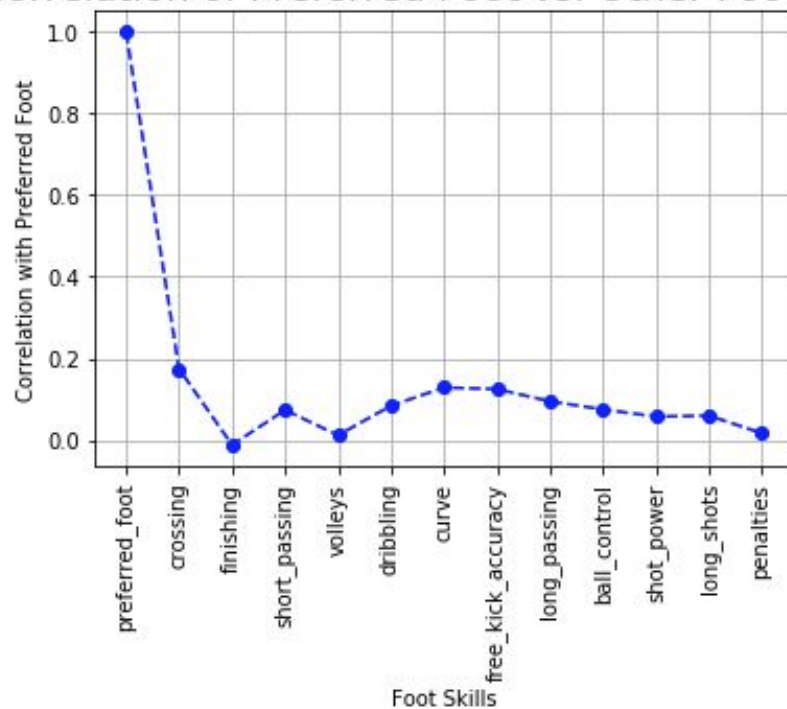
A heat map showing correlation of all foot skills. I found this to be very interesting because it shows very clearly that “preferred_foot” has almost no correlation whatsoever to any other foot skill. However, it does show that being good at one skill will likely translate into being good at another skill.



Findings

A line plot focusing in on “preferred_foot” versus other foot skills. This shows the actual values from the heatmap on the previous slide, showing just how un-correlated this statistic is.

Correlation of Preferred Foot vs. Other Foot Skills



Acknowledgements

I showed these plots to my girlfriend, who despises all things number-related, and she drew appropriate conclusions from the visualizations.

References

I did not use any outside references besides the Python and Seaborn documentation.

Visual Analysis of Preferred Foot in European Soccer Leagues

January 30, 2019

1 Visual Analysis of Preferred Foot in European Soccer Leagues

1.0.1 Jacob Miller

Using the European Soccer Database from Kaggle, I am going to determine if there is any correlation (positive, negative, neutral) between foot preference and other foot skills.

```
In [1]: # Package imports
        %matplotlib inline
        import sqlite3
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns

In [2]: # Load Player_Attributes as df
        cnx = sqlite3.connect('../Week 5/Week-5-Exercises-2/database.sqlite')
        df = pd.read_sql_query("SELECT * FROM Player_Attributes", cnx)

In [3]: # Count NaN, drop NaN values
        print('NaN values -before- dropping:')
        print(df.isnull().sum().sum())
        df.dropna(inplace = True)
        print('\nNaN values -after- dropping:')
        print(df.isnull().sum().sum())
```

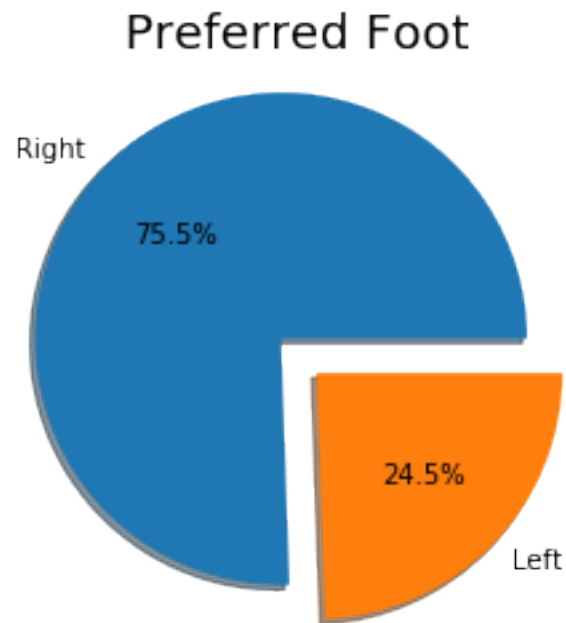
```
NaN values -before- dropping:
47301
```

```
NaN values -after- dropping:
0
```

```
In [4]: # Count how many prefer left/right
        preferred_foot = df['preferred_foot'].value_counts()
        print('Preferred foot:\n' + preferred_foot.to_string())
```

```
Preferred foot:
right    136247
left     44107
```

```
In [5]: # Display preferred foot as pie chart
fig1, ax1 = plt.subplots()
ax1.pie(preferred_foot,
        labels = ['Right', 'Left'],
        autopct = '%1.1f%%',
        explode = [0.1, 0.1],
        shadow = True)
ax1.axis('equal')
plt.title('Preferred Foot',
        fontdict = {'fontsize' : 18})
plt.show()
```



```
In [6]: # Turn preferred foot into integer for analysis
print('Foot preference -before- changing to integer:')
print(df['preferred_foot'].head().to_string())

df.loc[df['preferred_foot'] == 'right', 'preferred_foot'] = 0
df.loc[df['preferred_foot'] == 'left', 'preferred_foot'] = 1

print('\nFoot preference -after- changing to integer:')
print(df['preferred_foot'].head().to_string())
```

```
Foot preference -before- changing to integer:
0    right
1    right
```

```

2    right
3    right
4    right

```

Foot preference -after- changing to integer:

```

0    0
1    0
2    0
3    0
4    0

```

```

In [7]: # Pull out foot-related skills
        foot_skills = df[['preferred_foot',
                           'crossing',
                           'finishing',
                           'short_passing',
                           'volleys',
                           'dribbling',
                           'curve',
                           'free_kick_accuracy',
                           'long_passing',
                           'ball_control',
                           'shot_power',
                           'long_shots',
                           'penalties']]

```

```

In [8]: # Compute and show correlations
        correlations = foot_skills.corr()
        print(correlations)

```

	preferred_foot	crossing	finishing	short_passing \
preferred_foot	1.000000	0.174033	-0.012004	0.073302
crossing	0.174033	1.000000	0.576896	0.790323
finishing	-0.012004	0.576896	1.000000	0.580245
short_passing	0.073302	0.790323	0.580245	1.000000
volleys	0.011572	0.637527	0.851482	0.639995
dribbling	0.083831	0.809747	0.784988	0.788935
curve	0.128510	0.788924	0.691082	0.731948
free_kick_accuracy	0.124475	0.708763	0.633274	0.693490
long_passing	0.094717	0.685649	0.341121	0.803073
ball_control	0.074678	0.807721	0.720694	0.890622
shot_power	0.057309	0.656740	0.727835	0.722320
long_shots	0.059462	0.716515	0.806895	0.729741
penalties	0.016476	0.574208	0.726234	0.612511

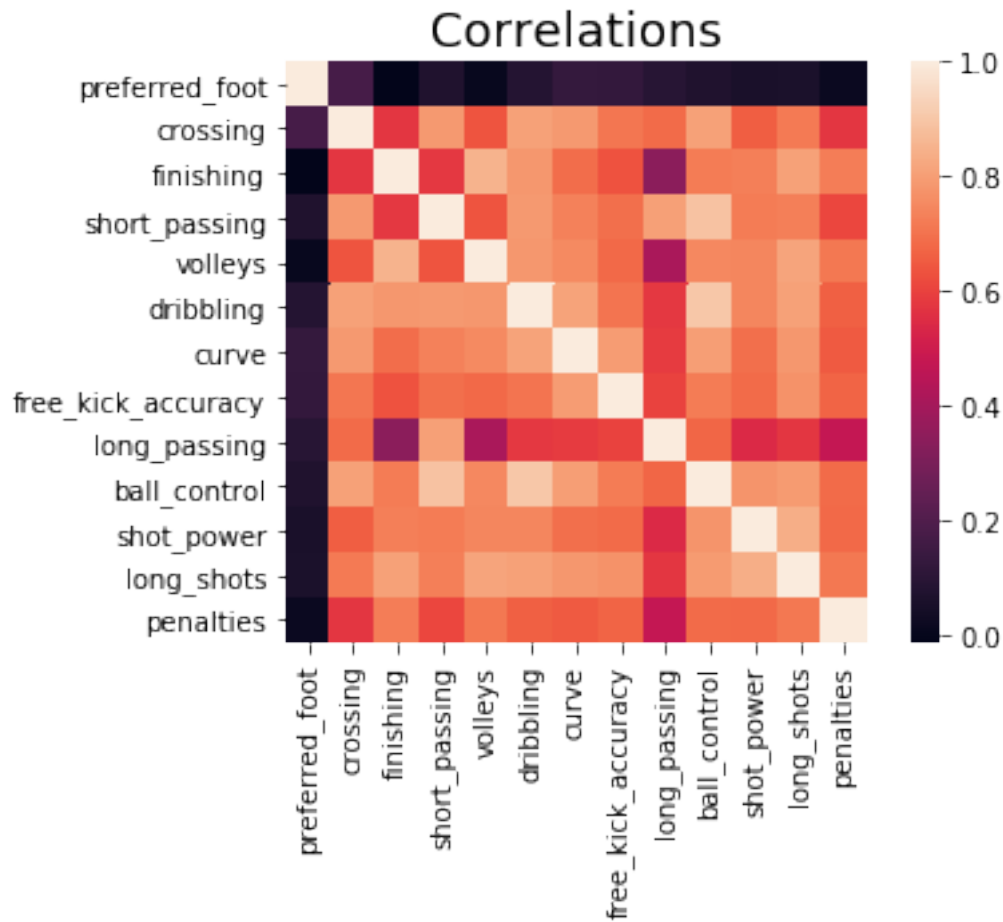
	volleys	dribbling	curve	free_kick_accuracy \
preferred_foot	0.011572	0.083831	0.128510	0.124475

crossing	0.637527	0.809747	0.788924	0.708763
finishing	0.851482	0.784988	0.691082	0.633274
short_passing	0.639995	0.788935	0.731948	0.693490
volleys	1.000000	0.784247	0.752410	0.682909
dribbling	0.784247	1.000000	0.810353	0.707322
curve	0.752410	0.810353	1.000000	0.797842
free_kick_accuracy	0.682909	0.707322	0.797842	1.000000
long_passing	0.414520	0.579201	0.586313	0.603286
ball_control	0.749459	0.901730	0.798598	0.720674
shot_power	0.746622	0.744960	0.694945	0.684191
long_shots	0.814894	0.807175	0.783732	0.773887
penalties	0.713116	0.663420	0.649737	0.669018

	long_passing	ball_control	shot_power	long_shots	\
preferred_foot	0.094717	0.074678	0.057309	0.059462	
crossing	0.685649	0.807721	0.656740	0.716515	
finishing	0.341121	0.720694	0.727835	0.806895	
short_passing	0.803073	0.890622	0.722320	0.729741	
volleys	0.414520	0.749459	0.746622	0.814894	
dribbling	0.579201	0.901730	0.744960	0.807175	
curve	0.586313	0.798598	0.694945	0.783732	
free_kick_accuracy	0.603286	0.720674	0.684191	0.773887	
long_passing	1.000000	0.675525	0.542244	0.576376	
ball_control	0.675525	1.000000	0.774856	0.793219	
shot_power	0.542244	0.774856	1.000000	0.838254	
long_shots	0.576376	0.793219	0.838254	1.000000	
penalties	0.476750	0.684410	0.680887	0.714596	

	penalties
preferred_foot	0.016476
crossing	0.574208
finishing	0.726234
short_passing	0.612511
volleys	0.713116
dribbling	0.663420
curve	0.649737
free_kick_accuracy	0.669018
long_passing	0.476750
ball_control	0.684410
shot_power	0.680887
long_shots	0.714596
penalties	1.000000

```
In [9]: # Plot correlations
        ax2 = sns.heatmap(correlations, square = True)
        plt.title('Correlations', fontdict = {'fontsize' : 18})
        plt.show()
```



```
In [10]: # Plot correlations with preferred foot
fig3, ax3 = plt.subplots()
ax3.plot(correlations['preferred_foot'], 'bo--')
plt.grid(True)
plt.xlabel('Foot Skills')
plt.ylabel('Correlation with Preferred Foot')
plt.xticks(rotation = 90)
plt.title('Correlation of Preferred Foot vs. Other Foot Skills',
          fontdict = {'fontsize' : 18})
plt.show()
```

Correlation of Preferred Foot vs. Other Foot Skills

