

# Assignment 4: Becoming an Independent Data Scientist

## 1 Region and domain

European football league  
Sports and Athletics

note: original topic and region was weather data and copenhagen, but data sets was very sparse (or very bad), so i visited the forums and read that it was allowed to define your own topic and region.

## 2 Research Question

Which European football leagues have the highest and lowest average goals per season, ranging for each football season between 2008 to 2016?

## 3 Links

The .sqlite file containing all data can be downloaded at kaggle, containing 6 datasets i dismantled into 6 .csv files. Of those 6 .csv files i use 'matches' and 'leagues' which i have uploaded to drop box since notebook will not allow files larger than 25mb.

- SQLite database from kaggle: <https://www.kaggle.com/hugomathien/soccer>
- dismantled .csv files: <https://www.dropbox.com/sh/mlgk66b2lv36keo/AABvunoXF-CuyusOtasOF1wFa?dl=0>

## 4 Image

See image below.

## 5 Discussion

This visualization was concerned on highlighting the european football leagues with the maximum and minimum goals scored per season for the eight seasons played between 2008 and 2016. Looking at the graph, it can be observed that the

The plot tried to find a balance between simplicity and detailed info in mind, and it could be argued that some of the leagues could be removed from the plot. Plotlines for most and least scoring leagues were highlighted by increasing width and alpha while also recieving a distinct linestyle of 'dashed'.

Colors were tried to be as distinct and contrastful as possible, while also being somewhat pleasing to the eye. It should be noted however, that finding 11 different colors and setting the alpha to half of what is normal, it requires extra energy to map the legends by using the colors.

