Group 3 Project Proposal (GitHub)
Peter Xenopoulos (pnx200), Stefan Cherubin (snc261)
https://github.com/NYU-VIS-FALL2018/laliga-viz

Competitiveness in La Liga

We will investigate how competitiveness in Spain's top soccer division, La Liga, has changed over the past 40 years. Specifically, we will guide the user through events in La Liga's history which have shaped its competitive landscape.

Data

1970-2017 La Liga data contains data with each row as a team's season total in a given La Liga season. We have the following attributes:

- Team (Categorical)
- Season (Quantitative)
- Team Record/Team Points (Quantitative)
- Goals For/Goals Against (Quantitative)
- Total Matches Played (Quantitative)

Questions and Tasks

- What are historical trends in competitiveness in La Liga?
- *TASK: Analyze the trends of competitiveness over our 40 year data set*
- *VALUES: Season, HomeTeam, AwayTeam, Standard deviation of points per match*

- Do trends in competitiveness correlate with happenings in the league?
- *TASK: Map events to the competitiveness timeline*
- *VALUES: Season, HomeTeam, AwayTeam, Standard deviation of points per match, Special Events*

- Has the league become uncompetitive?
- *TASK: Identify recent levels of our metric to compare to historical values*
- *VALUES: Season, HomeTeam, AwayTeam, Standard deviation of points per match*

Data Analysis

Although there are many ways to define competitiveness, we use the standard deviation of points per match. While we can present various ways to measure competitiveness, our analysis shows they are all highly correlated. Since our audience is casual soccer fans, we spare them the technical details up front.

Points per match is defined and calculated as follows:

- When a team wins a match, they earn 3 points
- Losing teams earn 0 points
- In the event of a draw, both teams earn 1 point
- There is no situation in which a team can lose points
- Points per match is simply the number of points a team has earned in a season divided by the number of matches they've played. Matches played should be equal for every team in a given season.

We use a per match basis to compare between seasons as different seasons had different amounts of teams (La Liga did not use a 20 team format until 1987.)

We then derive the standard deviation in points per match for each season. So that our metric is such that higher numbers imply more competition and lower numbers imply less competition, we take the reciprocal of the standard deviation of points per match. We are then left with our competition metric for that season.

An example is shown below for a fictional four team league in season 1:

Team	Points Per Match
One	2.5
Two	2.0
Three	1.5
Four	1.0

The standard deviation in points per match for season 1 is 0.65. Thus, the competitiveness would be 1/0.65 = 1.54. Now, consider a fictional season 2, where team *One* outperforms the rest.

Team	Points Per Match
One	2.5
Two	1.5
Three	1.5
Four	0.5

Now, the average points per match is 1.50 and the standard deviation is 0.82. Thus, the competitiveness would be 1.22, which is significantly lower than season 1. We see that the metric accurately portrays the relative change in competition.

For the rest of this analysis, we use simply the standard deviation of points per match in each season to reflect our initial analysis.

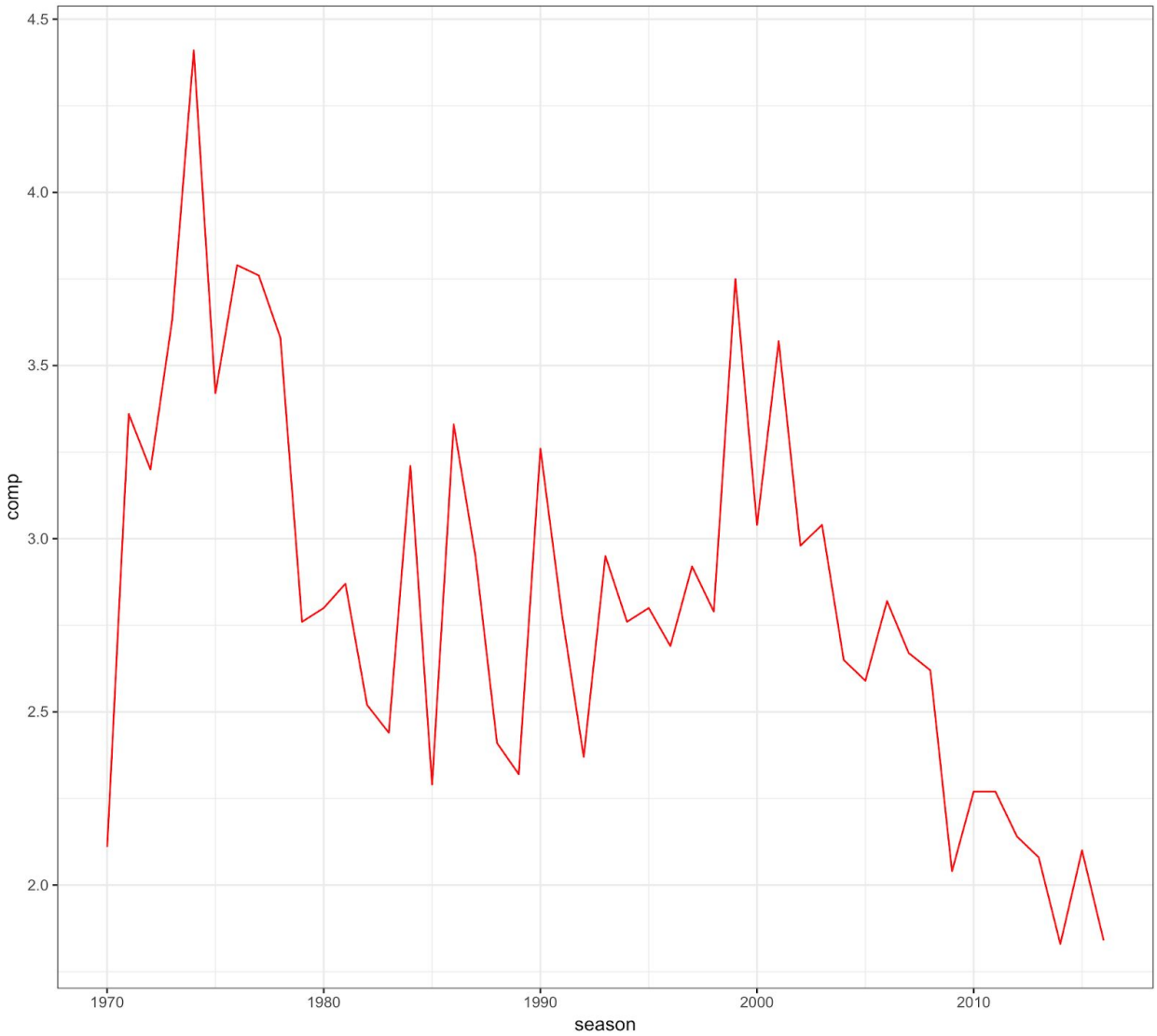


Figure 1: Measure of competitiveness over time

In general, we may first notice that our measure of competitiveness has dropped substantially in the last decade. This means the variance in points per match has risen substantially. We see that prior to this decade, although competitiveness varied, and was even rising.

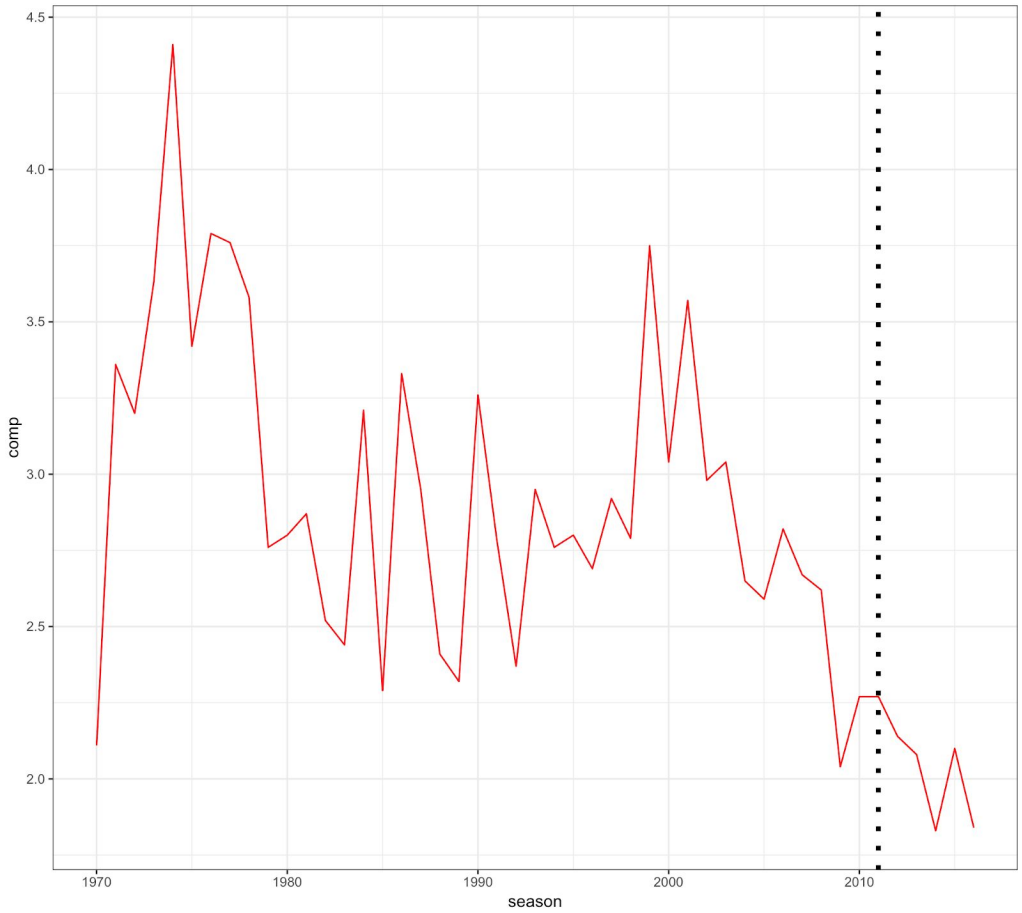
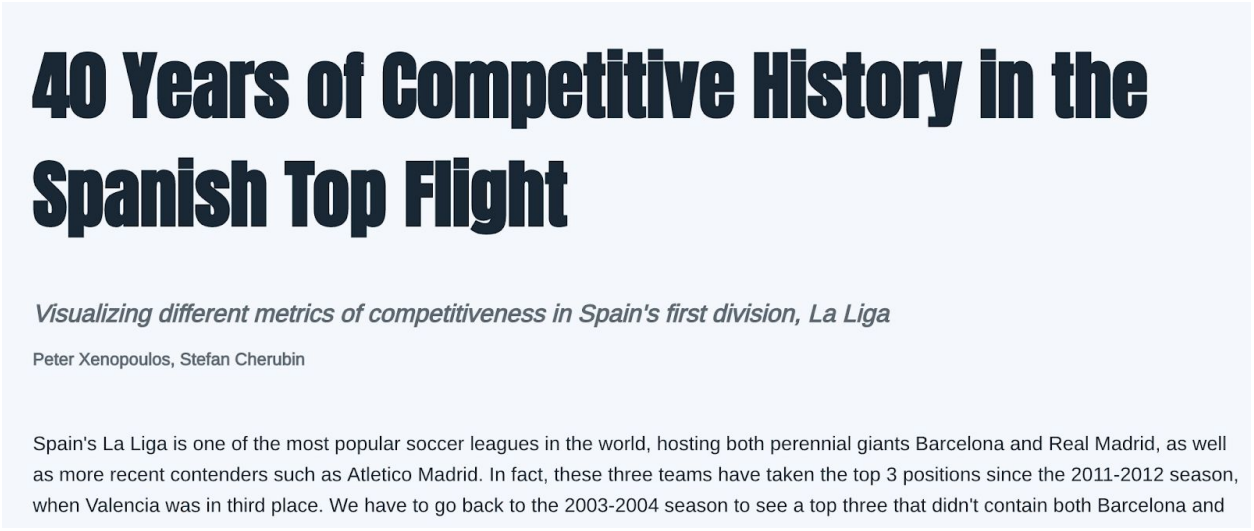


Figure 2: Introduction of UEFA Financial Fair Play Rules

UEFA introduced financial rules for member clubs in 2011 in order to curb excessive spending and unhealthy financial practices. We see that while competition seems to have dropped before the introduction of the rules, competition seems to have declined at a similar pace post implementation.

Storyboarding

We first will introduce our visualization in a web page, which will have a title that reads like a news story. We provide an example below. The title is vague enough to have the user interested enough to keep scrolling, but informative to actually convey what the visualization investigates.



40 Years of Competitive History in the Spanish Top Flight

Visualizing different metrics of competitiveness in Spain's first division, La Liga

Peter Xenopoulos, Stefan Cherubin

Spain's La Liga is one of the most popular soccer leagues in the world, hosting both perennial giants Barcelona and Real Madrid, as well as more recent contenders such as Atletico Madrid. In fact, these three teams have taken the top 3 positions since the 2011-2012 season, when Valencia was in third place. We have to go back to the 2003-2004 season to see a top three that didn't contain both Barcelona and

After a brief introduction, we will present a simple scatterplot visualization to set up our “competition” metric. We hope that the audience will see that Barcelona and Real Madrid start to drift away consistently from the pack.

Once we set up the background for the competition metric, we will then show the complete history of the competition metric. This chart will be interactive, as if the user hovers over a season, they will get the breakdown of the points per match for each team in that season. This can also further elucidate the competitiveness in that season. Finally, we also include bars that indicate events. These lines can be hovered over to receive a description of that event and its effect on competition in the league. We show a sketch below.

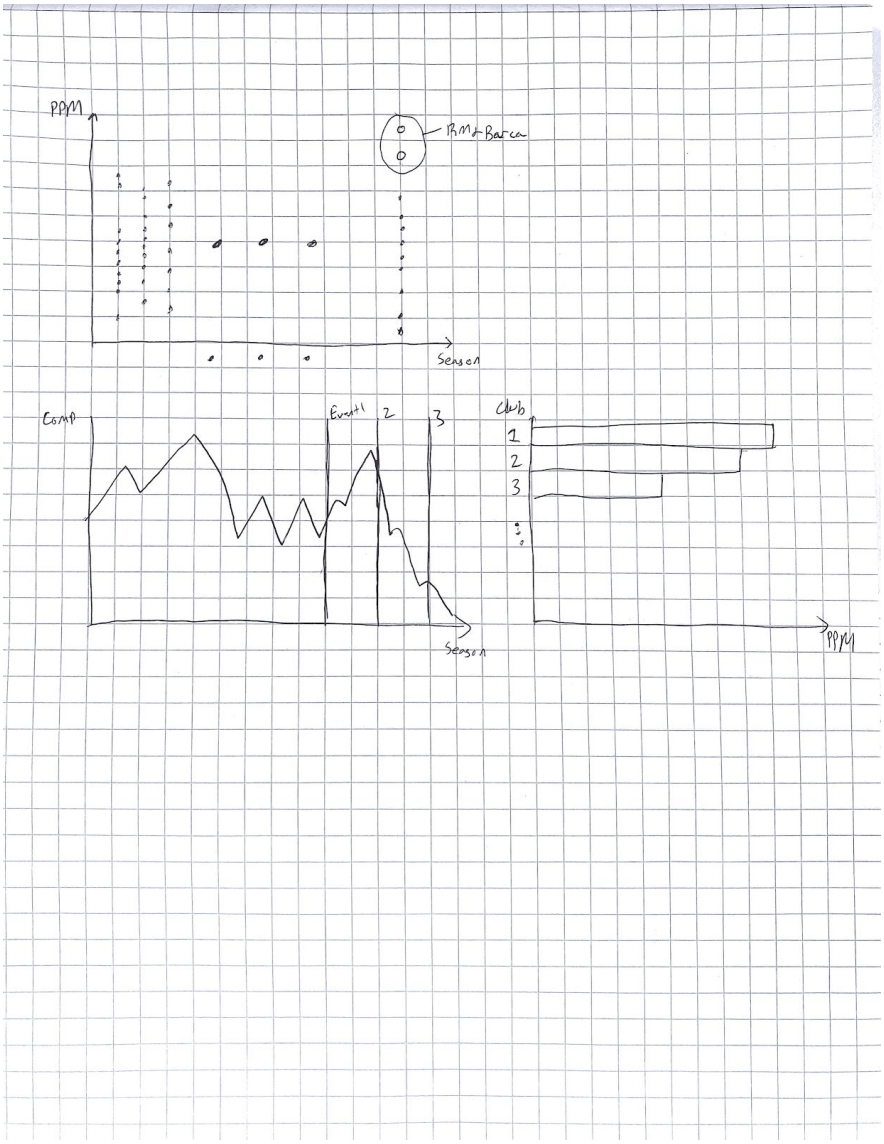


Figure 3: Sketch of Proposed Charts

Implementation

The GitHub repository is found here and the demo page is found here.

Feedback

Our first feedback from Cristian was to make it so that our competition metric was such that higher numbers indicated more competition and smaller numbers indicated less. Then, our second set of feedback from Gromit was to include a chart, like a scatterplot, which shows the distribution/variation in points per match for each season, so that understanding our competition metric is easier.