Data Analyst Nanodegree - Project 6: Data Visualization using Tableau

Stephan Teodorovich, June 2018

Contents:

1. Introduction

With the start of the 2018 FIFA World Cup, I decided to take a deeper look into the makeup of the 32 National Teams participating in the tournament focusing on the following questions:

1. Which professional football clubs, leagues, and FIFA confederations have the most players participating in the tournament?
2. Which teams have the greatest and least diversity in club, league, and FIFA confederation representation?
3. Is there a correlation between a team's geographic location (defined by continent) and its source of players?

Note: I have placed links to sources within this document, all of which are shown in section **6. Resouces/Links**

2. Data Review, Cleaning, and Preparation

To conduct my review, I first downloaded the wc2018-players.csv dataset from Kaggle.com. The original set contained the following variables:

- National Team
- Player Number
- Player Position
- Player Name
- Player Birthdate
- Player "Shirt" Name

- Player Club Team and Country
- Player Height
- Player Weight

From this set I removed all of the irrelevant information (Player Number, Position, Birthdate, Shirt Name, Height, and Weight). I then used Tableau's built-in Split, Alias, and Group functions to separate the Club and Country information, and create new League and FIFA Confederation variables, resulting in the following dataset:

- National Team
- Player Name
- Player Club
- League (Country)
- FIFA Confederation

3. Design

When considering the best way to "tell the story" I began by looking at what information I had. The data with which I was working did not lend itself to visualizations utilizing trendlines, scatterplots, boxplots, or other displays of statistical analysis results. Instead, my visualization would need to focus on how to highlight comparisons of relative proportions and amounts. Therefore, I determined that the most effective means of showing the data would by through bar graphs.

4. Feedback

I solicited the feedback on my visualizations from four individuals, representing a spectrum of expertise on interpreting data:

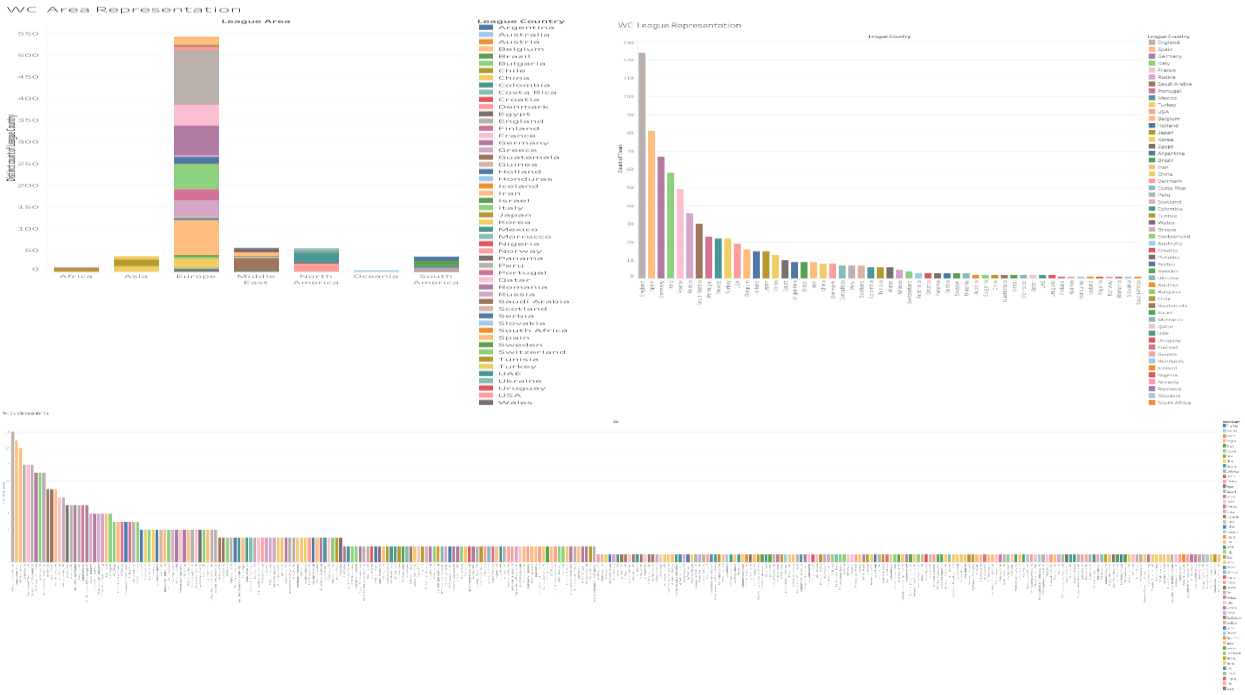
1. MH – A math teacher in the Los Angeles Unified School District with nearly 20 years of experience.
2. AM – A research scientist employed in the bio-pharmaceutical industry working on DNA-based tissue-typing products.
3. BI – A professor of literature at a Los Angeles-area college.
4. DV – A communications specialist at a Fortune 10 company

Before I showed my visualizations, I explained the three questions I wanted to answer, so they would have the proper context in which to provide actionable comments.

I have edited the comments from my reviewers for each iteration of my visualization, to focus only on the relevant portion.

4.1 First Iteration:

My first visualization was a skeletal outline of what I intended, and I expected questions and severe criticism noting incompleteness and ambiguity. The first version consisted of these three bar graphs:



4.1.1 Review Comments

MH: *I'm not sure I understand what each chart/graph is supposed to represent.*

AM: *The comparisons (such as club representation) are interesting, but there seems to be far too many bars to make meaningful sense of the information. Perhaps simpler graphs, or more detailed drill-down graphs would help*

BI: *Wow, there are a lot of colors on here!*

DV: *I don't get what you're trying to say with these. Are you focusing on the world cup teams or the players?*

4.2 Second Iteration

For my second version, I made the following changes & additions:

1. Clarified the graph titles to make them more explanatory
2. Separated the visualization into three distinct sections: Clubs, Leagues, and Confederations
3. Added two new graphs per section (for a total of 9 images): A global map, and a Team-specific breakdown.

4.2.1 Review Comments

MH: *These are much more clear about what you're trying to represent. However, it seems that adding some additional information into each graph/chart would make it even more useful. Also, I'm not really able to easily follow how this answers your third question.*

AM: *It looks like you're on the right path for giving a thorough review. I'd look into including some of the actual data to each image to make them more clear and to highlight differences. I*

also think you might want to add one more level of exploration, because as it stands one is made to try and infer any pattern between what leagues and a team's geographical location.

BI: *I like how you separated these into three sections! It makes it easier to begin to get what you're trying to say However, it's still a little hard to follow because of the number of colors on some of these graphs (like Club Representation). Maybe putting in some numbers to help make it easier to figure out would be a good idea.*

DV: *I see a story developing, focusing on the teams and not the players. But I'm still not entirely sure what you want to say.*

4.3 Third Iteration

For my third (and final) version I made the following changes:

1. Minor text edits to try and further improve clarity
2. Added number totals to all of the bar graphs to enhance understanding
3. Added six new graphs to more clearly address my third question

4.3.1 Review Comments

MH: *The numbers in the graphs are a big help in getting to see the overall picture. And the new graphs (for the confederations) do a good job of answering the third question.*

AM: *Adding data was a considerable improvement. These illustrations do a good job of showing the breakdowns of where the talent comes from, and where it is going. It's too bad you didn't have data to do some statistical analysis, or perhaps use historical data to try and see changes over time. Still, I think you've managed to put together a coherent collection that answers all three of your questions.*

BI: *Oh, the numbers are so helpful! Even though I know I keep mentioning the colors I think that the way you have it now makes them less confusing. The new graphs are also very interesting!*

DV: *You've got a story here that does a good job of showing where the teams get their players, and which leagues & confederations are the most important in providing the talent. There are still some things that could be tweaked, but I think you've done a good job of communicating the answers to your questions.*

5. Conclusion

Because my final project used data that severely restricted the types of visualizations I could use (no trendlines, no scatterplots, no representations of statistical analysis, etc) I was forced to find ways to convey information through a limited range of options.

Although I had some initial idea regarding the best way to present the data, the feedback I obtained proved very helpful in fine tuning each individual image, as well as providing guidance on additional graphs to help clarify the story I was trying to tell.

I was also able to take feedback from people with different relationships to data: two people comfortable with quantitative analysis and representation of data for an academic purpose, and two people who are more qualitative and for whom data as an illustration is more beneficial.

Using feedback from a diverse group permitted me to strike as much a middle ground as possible, providing a clear data narrative.

6. Resources/Links

- Dataset: www.kaggle.com
- FIFA Confederations: https://www.fifa.com/associations/index.html
- Project (v1): https://public.tableau.com/profile/stephan.teodorovich#!/vizhome/WC2018_v1/Dashboard1
- Project (v2): https://public.tableau.com/profile/stephan.teodorovich#!/vizhome/WC2018_02/Dashboard1
- Project (Final): https://public.tableau.com/profile/stephan.teodorovich#!/vizhome/WC2018_v3/Dashboard1