Untitled

RN7

June 22, 2018

Introduction ## - hi, how’s it going? my name is ryo nakagawara and i’m going to be talking about visualizing the world cup with R

## 1

* just a little bit about me, my background is mainly in the social sciences, specifically in psychology and economics
* and currently i work for ACDI/VOCA, which is an international development NGO
* my main intersts are soccer or football - i lived in both england and america so i swap between both terms which well, annoys everyone equally

## 2

* anyways, the agenda for today is…
* that i’m going to be talking about these 4 different types of visualizations
* this talk isn’t going to be as serious as previous talks, so sit back, relax because…

## 3

“Now this is where the fun begins!” gif

## 4

* so quick summary: the world cup is divided into 2 stages, the group stage and then the knock-out rounds
* there are 8 groups of 4 teams each, and each team plays the other once, and the top 2 teams based on 3 points for a win, 1 point for a draw, and 0 points for a loss go through!
* As you can see this is how Group D unfolded throughout the last 90 minutes of gameplay.
* last important bit, the last 2 games of the group happen simultaneously to avoid teams colluding with eachother to go through to the next round.
* On the left, you see how the teams were ranked at the start then on the right, how the teams were at the final whistle.
* Argentina, bottom of the group at the start but still with a fighting chance to finish 2nd, absolutely needed to win
* and many Argentine prayers were answered when Messi scored in the 14th minute.
* However, they crashed down to last place when Moses equalized but eventually made it through to the knock-out stages from a goal by Marcos Rojo!
* These graphs really capture the drama of the group stages with the hopes and dreams of the teams and their respective nations on the line!
* So let me explain in some of the most interesting bits of the code used to create these!

## 5

* the geom\_flag() function that shows the wonderful flags require two-digit ISO codes as input!
* now, i can google the codes everytime and look through a list but that’s annoying!
* so what’s the solution? countrycode package!
* all you need to do is: input the name of the countries and specify that in origin argument, country.name
* specify that you want the output to be ISO2C or other types in the destination argument
* et VOILA!
* the countrycode() function can take a lot of different inputs, “country names in german”, “3 digit ISO code”, “eurostat”, “imf”
* so basically you can do the reverse of what I just did >>> it’s a very useful package!
* Next, I want to show you…

## 6

* how I displayed the text in the official world cup font, DUSHA!
* I thought having the text be in regular font was kinda boring,
* and having seen a lot of the marketed graphics and merchandise I found and downloaded the .TFF files
* and installed them!
* Then, with the extrafont package, I was able to use them in R!
* Admittedly, some letters look a bit squished and are hard to see. But, overall it brings a little flavor into the titles and annotations in the plots!

## 7

* To create these I had to make a lot of labels in a lot of dataframes…

## 8

* But soon I was able to create a template of sorts, a set of label dataframes with the same general structure which I could just pass different country names, goal scorers, minute times into…

## 9

* If i had more time, i really want to brainstorm HOW to turn this into a package, but for now this “template” or sorts will have to do…!
* the most difficult part of making these was that it was tough to keep track of how teams were ranked both beforean event happened and how things changed
* i did these AFTER the matches finished, so i had to do a lot of backtracking and figuring out which teams were ranked where, at a number of specific times, across 2 different games.
* Going back to the actual plots…

## 10

we can see that some groups are full of excitement!! - Iran just needed one more goal after scoring in the 93rd minute, to leapfrog Portugal and take 2nd place! It was a really exciting end to the game!

## 11

* others… like Group C… were not! after matchday 2, France was 100% through, Denmark just needed to draw their game.
* But even here, Australia could still have qualified if Denmark lost and they had won!

That’s part of the fun, you never really know what’s going to happen!

Now let’s go on to…

## 12

* Recreating the goals of the World Cup!
* For doing these, I primarily used the ggsoccer package by Ben Torvaney which is used with ggplot2 to draw a soccer field / football pitch in R!
* Basically, annotate\_pitch() creates the markings for the soccer field such as the center circle, 18-yard box, penalty spot, etc. while theme\_pitch() erases the extraneous axes and background stuff from the default ggplot style.
* By using the limits arguments in the different coord\_\*() functions, like coord\_fixed , coord\_cartesian … here I used coord\_flip()
* so that I can focus on a certain area of the pitch and orient it in a way that I want.

## 13

* So, I’m sure you are all wondering, where did I find the data points for the players, the ball, etc.?
* Well, those that are soccer fans know about sports analytics companies, most notably Opta, that generate a huge amount of data for every player for every match, but the thing is it’s not easy for a regular guy like me to buy it!
* If you’ve been following the World Cup on Twitter you might have seen some great viz by the Financial Times, FiveThirtyEight, the BBC, etc. who got data from Opta, there was also this other guy that scraped a data dashboard from a soccer data website (that in fact, create their viz from purchased data from Opta) with **RSelenium** or some other JS scrapers.
* but I could do neither of those things so I resorted to creating the coordinate positions by hand.

## 14

* Thankfully, due to the plotting system in ggsoccer and ggplot2, it’s very easy to figure out the positions on the soccer field plot and with a little bit of practice it doesn’t take too much time.

Here is a small part of what some of the dataframes look like with all the coordinate points stored for different parts of the plot.

## 15

and here’s an example of the ggplot2 code: - after creating the soccer field layout i then plot a lot of segments and curves using the respective geoms to plot out the movement of the players and the ball!

## 16

* so here was my first try, Gazinsky scoring the first goal of the World Cup!

## 17

* and here is another, Cristiano scoring a hat trick in what was probably one of the games of the tournament!

## 18

* So, I wanted to talk about how I got the flags into the title for this previous plot.
* I basically put a lot of spaces between the country name and the score in the annotation code, and then I specified the coordinates for the flags to appear in that empty space in the title!
* Yes! It’s okay, you can laugh. It is a terrible hack but it works!

## 19

* and for all the Japanese people in the audience, I also recreated Osako’s winner against Colombia!

OK… now these are cool but soccer is a moving - flowing game! It’s not enough to just show static images, it just doesn’t capture the feel of the sport! So, the next step was to try and animate these images using gganimate and tweenr!

## 21

here is the gganimate version of Gazinsky’s goal…

* basically you have to add some kind of time variable in your data that keeps track of the WHEN of your data points.
* then you have to specify that time variable in the frame argument of the geom you want animated
* but as you saw, the animation was very choppy and it didn’t really look like soccer at all
* we can do better than this by making everything run… smooother!
* and this is where tweenr comes in.

## 23

For those of you that aren’t familiar, tweenr allows you to interpolate data between different states and specify other aspects like the easing of the transition.

## 24

Now let’s check out Osako’s goal vs. colombia!

and here’s another one ## 25 Japan’s offside trap against Senegal!

against the height and physicality of Senegal, Japan’s strategy was basically …

## 26

* you can’t lose the aerial battle, if you set an offside trap!

and yes, this internet meme was created in R too! …

Anyways, let’s get back on topic…

the LAST thing I wanted to show you was …

## 27

* animating national team uniforms!
* about a month ago, Guy Abel created this really cool blog post on animating soccer uniforms with R
* what he did was that he downloaded uniforms from a website and then used the magick package to animate the transitions
* the thing that got to me was, that to use all these uniform graphics Guy used this website called historicalkits.co.uk which is a website I’ve been using since I was a little kid, so when I saw the blog post I was like
* HEY - WHY DIDNT I THINK OF DOING THIS??

## 28

* so with the help of Guy’s awesome code I went to work, and obviously I didn’t want to copy exactly what he did but instead with Japan’s uniforms
* so I went ahead and combined Japan’s uniforms with those of their opponents in each of the games, in each of the world cups
* to create a historical slideshow showing all of Japan’s World Cup games!

## 29

20~30 seconds?

So I want to finish up by some comments on making these…

## 30

* There is definintly room for improvement for these, but I already put a lot of work during my free time,
* free time which was already limited by the fact that most of it was spent actually watching and enjoying the games!
* even still it was a a great way to challenge my R skills in so many different ways
* and i think if you want to get better at R, you should test yourself by working on topics and subjects that you truly love

## 31

all right, that’s all from me so thank you for listening! and i hope that maybe i’ve got some of you interested in soccer!

unfortunately, after the final tonight, you’re going to have to wait another 4 years for the next one!!