Our lab is international. People born all over the world have come to work in my group. I’m proud of this fact, especially in the current political climate. I’ve previously used the GoogleMaps API to display a heat map on our lab webpage. It shows where in the world people in the lab come from. This was OK, but I wanted to get an R based solution to make this graphic to make it easier to automate updates.

This post is an explainer and “how to” guide.

The idea is to create graphics to describe the origins of a group of people. For my use-case it is my research group, but it could be any group of people. All you need is a list of countries that the people come from.

In the example for this post, I took the [Top 100 Footballers voted for by Guardian readers in 2016](https://docs.google.com/spreadsheets/d/12n0Nu9Uu7CCG_w4sTbhFzijeAKhMbXgkzxkRKlg7vb4/edit#gid=885154705). In my dataset, I store the countries of origin in ISO2 format. This means Spain is ES, Germany is DE and so on. I converted the Guardian dataset to ISO2 format using a lookup and then I was ready to put it into the R script.

if (!require("rworldmap")) {

install.packages("rworldmap")

library(rworldmap)

}

# ggplot2, ggFlags, dplyr are needed for the bar charts

library(ggplot2)

library(dplyr)

if (!require("ggflags")) {

library(ggflags)

}

# csv file with each person as a row and containing a column with the header Origin and

# countries in 2-letter ISO format (change joinCode for other formats)

file\_name <- file.choose()

df1 <- read.csv(file\_name, header = TRUE, stringsAsFactors = FALSE)

countries\_lab <- as.data.frame(table(df1$Origin))

colnames(countries\_lab) <- c("country", "value")

matched <- joinCountryData2Map(countries\_lab, joinCode="ISO2", nameJoinColumn="country")

This part of the script sets up the libraries that are needed. We’ll use the rworldmap package first. This was very useful for guidance. We load in the csv file which contains the countries of origin for the people we want to map out. It doesn’t need anything more than one column with the ISO2 codes. If it does it’s OK. As long as the header for the countries column is called “Origin”, all will be OK.

This column is extracted and a new dataframe is made from it which has each country as a row and the number of instances of that country as a second column. These are labelled “country” and “value” for convenience. Now rworldmap does its thing with the joinCountryData2Map line. Next we make the map!

# make png of the map

png(file = "labWorldMap.png",

width = 1024, height = 768)

par(mai=c(0,0,0.2,0))

mapCountryData(matched,

nameColumnToPlot="value",

mapTitle= "",

catMethod = "logFixedWidth",

colourPalette = "heat",

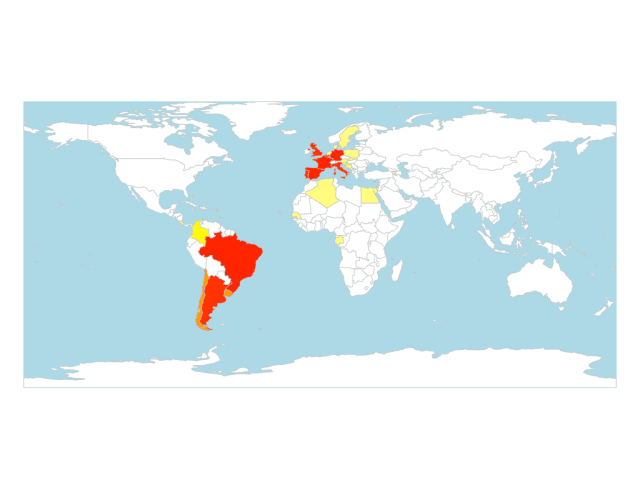
oceanCol="lightblue",

missingCountryCol="white",

addLegend = FALSE,

lwd = 1)

dev.off()

Where in the world…. heat map showing country of origin for the people in the dataset

This makes a nice map. I’ve hidden the legend which shows what the colours mean. The map can be customised in lots of ways. I liked the way this map looked and my other aim was to make a chart to show the relative numbers of people in each country. Speaking of which…

# make bar chart of lab members

countries\_lab %>%

mutate(code = tolower(country)) %>%

ggplot(aes(x = reorder(country, value), y = value)) +

geom\_bar(stat = "identity") +

geom\_flag(y = -1, aes(country = code), size = 4) +

scale\_y\_continuous(expand = c(0.1, 1)) +

xlab("Country") +

ylab("Members") +

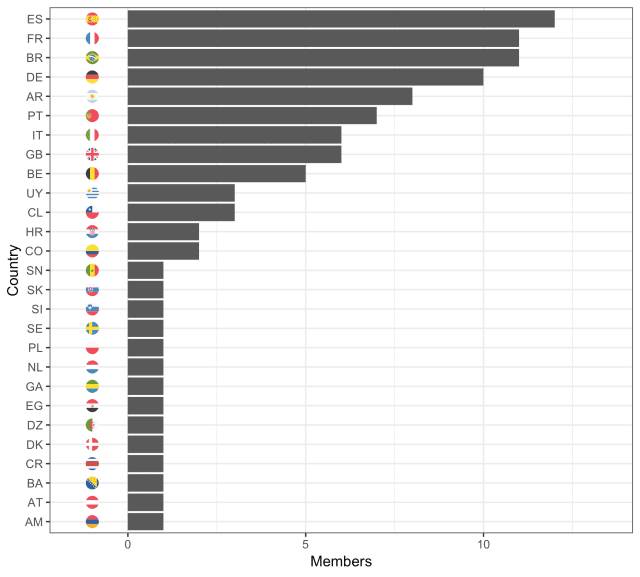
theme\_bw() +

theme(legend.title = element\_blank()) +

coord\_flip()

ggsave("plot.png", plot = last\_plot())

Using the data frame we made previously, we can pipe it to ggplot2 via the wonders of dplyr. I am using geom\_flag here from the ggflags library. Note that this is a fork of ggflags which gives circular flags which look great on the graph. The geom\_flag needs a lowercase entry for each ISO2 country code so the first step is to mutate the country column to make a new lowercase column called code.

Bar chart of the same dataset using flag emojis for the tick labels

That’s it! With a csv file and a few lines of R code you can generate some nice looking graphics.

The dataset shows that the country that produced the biggest fraction of the world’s best footballers (as voted for by Guardian readers) was Spain. There are no surprises in this dataset. The most prominent European and South American countries giving a strong showing.

—

The post title is taken from “All Around The World” by The Dead Milkmen. Many songs in my library with this title, but this paranoid extraterrestrial tune gets the pick this time.