Lab 11 - Data, Aesthetics, & Geometries

Ignacio Cabezudo

November 9, 2017

Complete the following exercises below. Knit together the PDF document and commit both the Lab 11 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Which variables in your dataset are you interested in visualizing? Describe the level of measurement of these variables and what type of geography you think is appropriate to represent these variables. Give your reasoning for choosing the `geom_()` you selected.

The variables that will be visuallized will be nkill, iyear from the GTD dataset & year, country from the IMF dataset.

2. Is your data in the proper format to visualize the data in the way you want? Why or why not? *If you need/want to change the structure of your data, do it below.*

I don't know, I'm having trouble deciding on this. Will no more after our meeting today.

3. Create at least two different exploratory plots of the variables you chose using the skills we covered in class today. What types of mapping aesthetics did you choose and why? What do these plots tell you about your data?

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------- tidyverse 1.2.1 --
```

```
## v tibble  1.3.4     v purrr   0.2.4
## v tidyr   0.7.2     v dplyr   0.7.4
## v readr   1.1.1     v stringr 1.2.0
## v tibble  1.3.4     v forcats 0.2.0
```

```
## -- Conflicts -------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(hexbin)
```

```
## Warning: package 'hexbin' was built under R version 3.4.3
```

```
library(readxl)

gtd_filter <- c("country_txt", "nkill", "iyear", "imonth")
gtd_full <- read_csv("GTD FULL DB.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   eventid = col_double(),
##   iyear = col_integer(),
##   imonth = col_integer(),
##   iday = col_integer(),
##   extended = col_integer(),
##   country = col_integer(),
##   region = col_integer(),
```

```
##   latitude = col_double(),
##   longitude = col_double(),
##   specificity = col_integer(),
##   vicinity = col_integer(),
##   crit1 = col_integer(),
##   crit2 = col_integer(),
##   crit3 = col_integer(),
##   doubtterr = col_integer(),
##   alternative = col_integer(),
##   multiple = col_integer(),
##   success = col_integer(),
##   suicide = col_integer(),
##   attacktype1 = col_integer()
##   # ... with 44 more columns
## )
```

```
## See spec(...) for full column specifications.
```

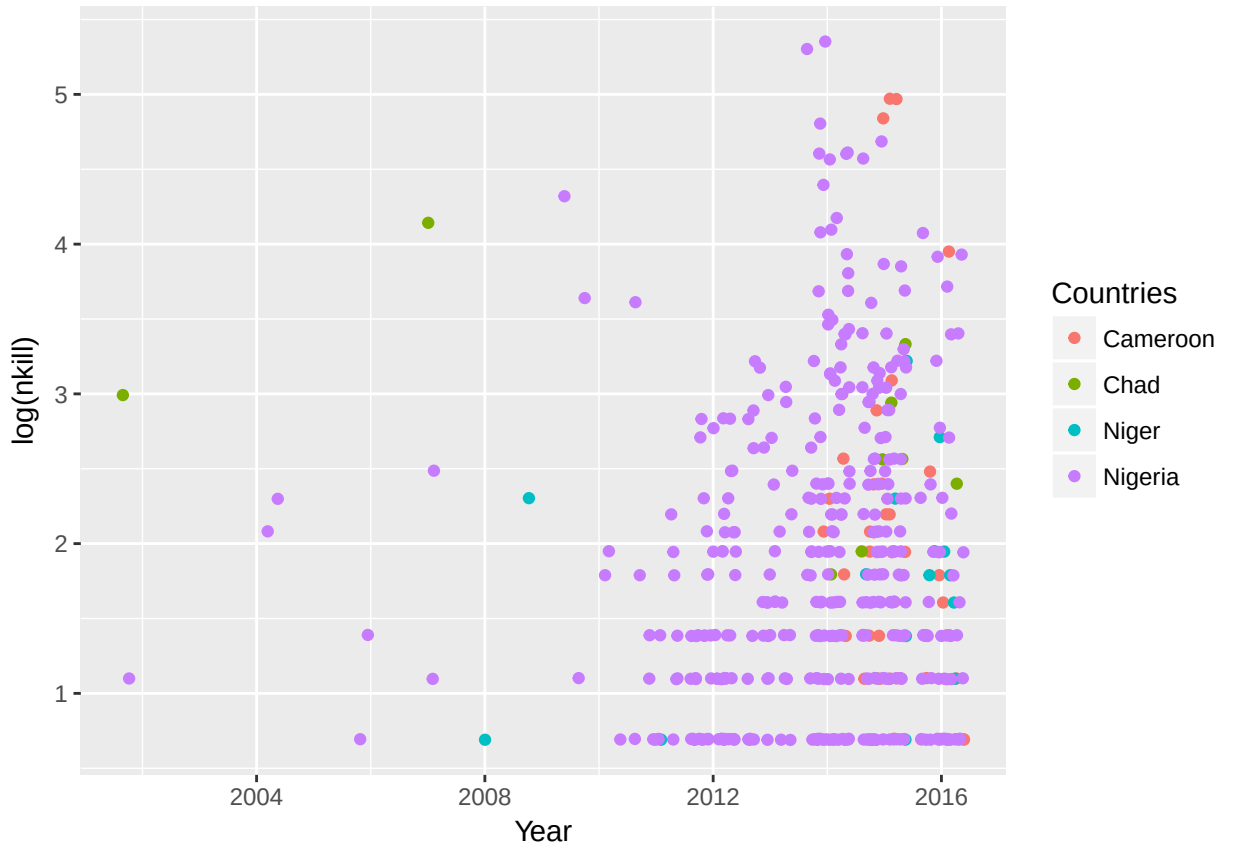
```
## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)
```

```
## Warning: 246 parsing failures.
## row # A tibble: 5 x 5 col     row       col               expected actual               file expected
## ... ................. ... ................................................................ ........
## See problems(...) for more details.
```

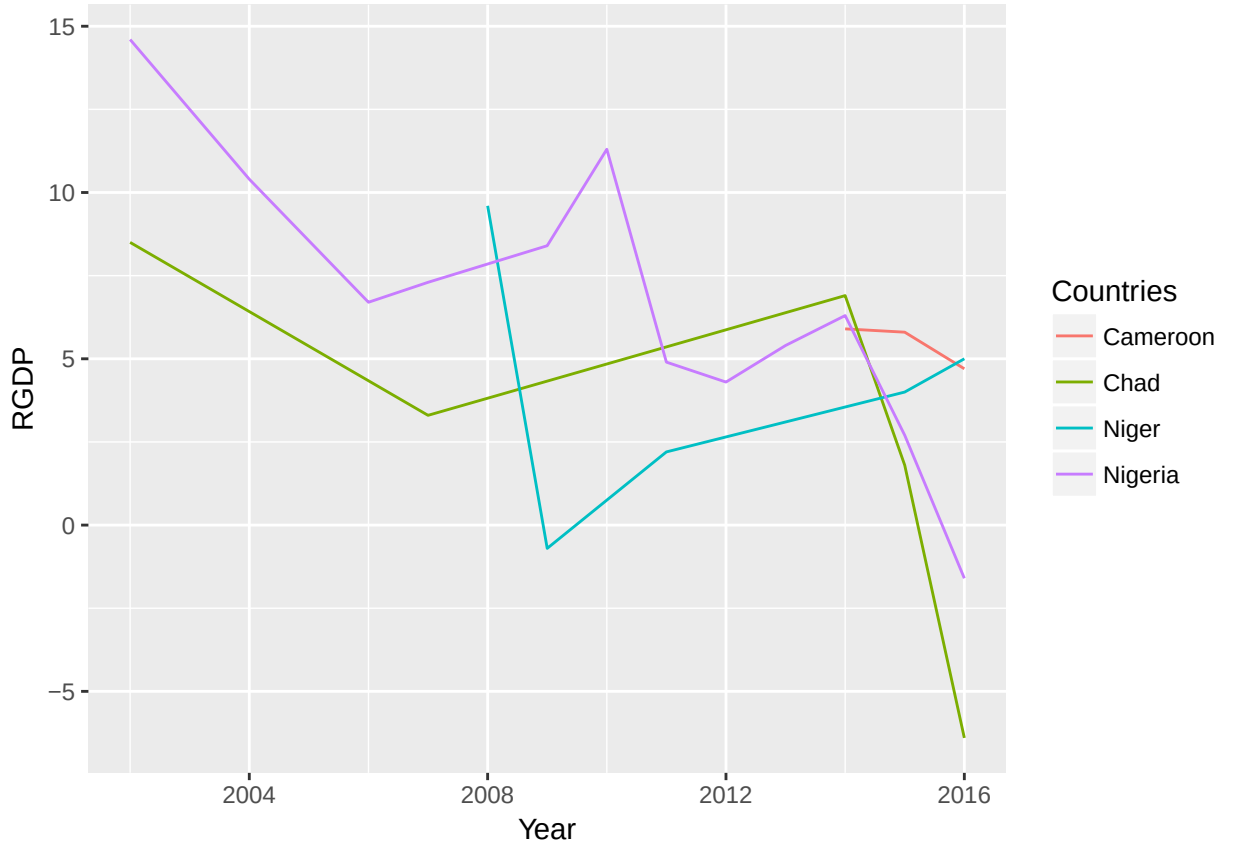
```
IMF_Real_GDP_Growth <- read_excel("IMF Real GDP Growth.xls")


#Set the ranges for the data using filter
plotTerror <- filter(gtd_full, country_txt == c("Nigeria", "Niger", "Chad", "Cameroon") & iyear > 2000 &
```

```
## Warning in country_txt == c("Nigeria", "Niger", "Chad", "Cameroon"): longer
## object length is not a multiple of shorter object length
```

```
plotTerror <- select(plotTerror,gtd_filter)

#Fix the GDP growth data
temp <- gather(IMF_Real_GDP_Growth, key = Year, value = RGDP, ...=-'X__1')

#Rename columns to setup keys for join
names(temp)[names(temp) == 'X__1'] <- 'Countries'
names(plotTerror)[names(plotTerror) == 'country_txt'] <- 'Countries'
names(plotTerror)[names(plotTerror) == 'iyear'] <- "Year"

#Join the two datasets

joinT <- merge(plotTerror, temp, by = c("Countries", "Year"))

#Make some plots

ggplot(joinT, aes(x = Year, y = log(nkill), col = Countries )) +
  geom_jitter()
```

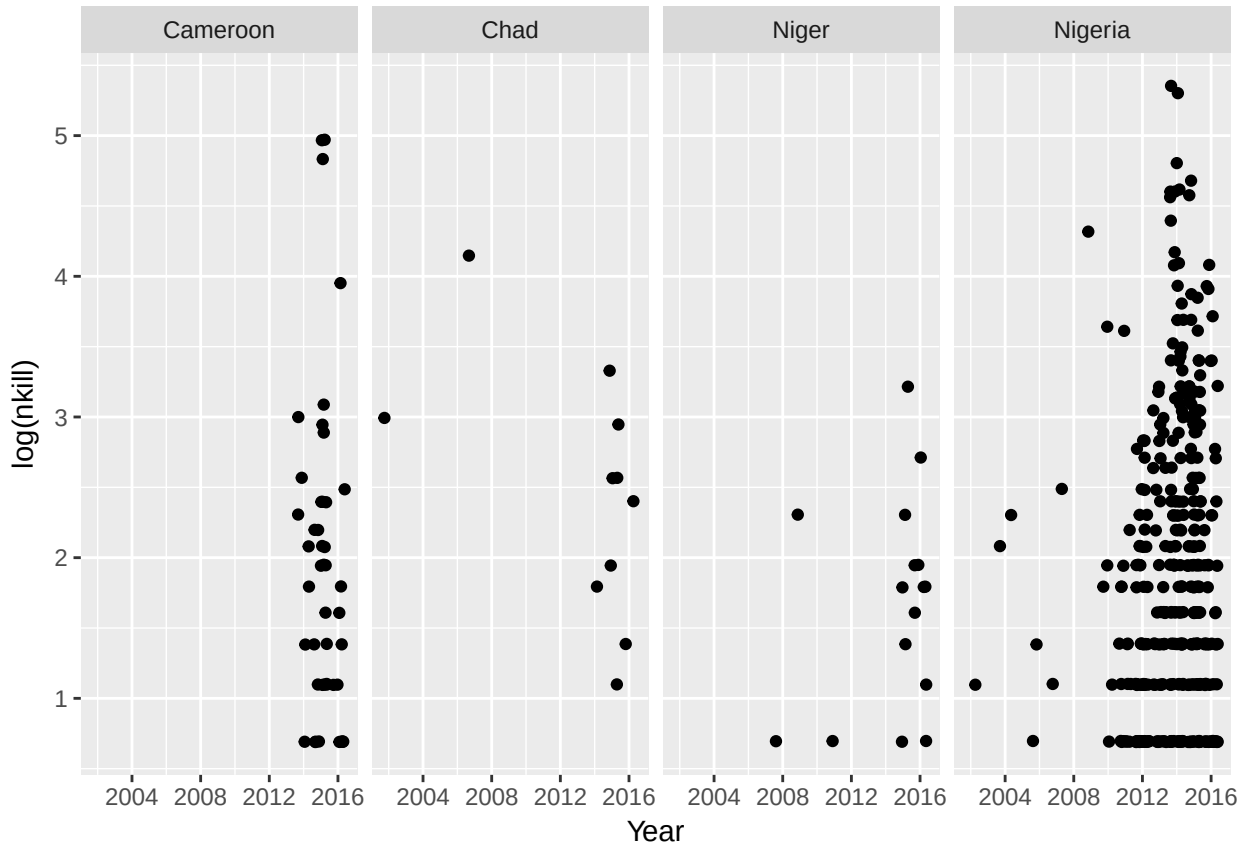



```
ggplot(joinT, aes(x = Year, y = RGDP, col = Countries)) +
  geom_line()
```

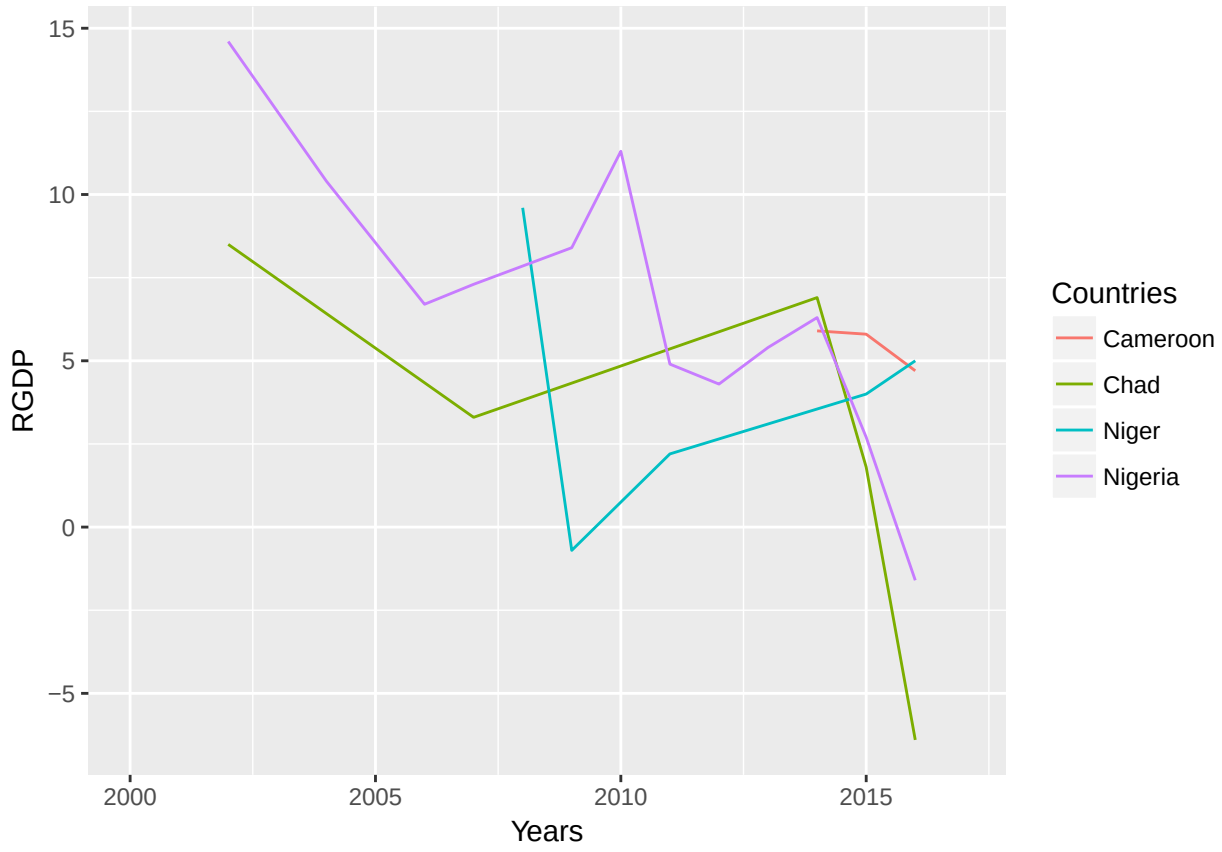


4. Create at least three variations of the plots you've already made by modifying some of the arguments we covered in class (i.e. `position`, `scale`, `size`, `linetype` etc.). Do any of these modifications help you understand your data better? Why or why not? Do any of them create a misleading interpretation of the relationships between your variables? If yes, how so?

```
ggplot(joinT, aes(x = Year, y = log(nkill))) +
  geom_jitter()+
  facet_grid(. ~ Countries)
```




```
ggplot(joinT, aes(x = Year, y = RGDP, col = Countries)) +
  geom_line() +
  scale_x_continuous(
    name = "Years",
    limits = c(2000, 2017))
```



5. From the plots you've created thus far, do any of them seem appropriate for a general audience? Why or why not? If so, what do you think you'd still need to do to make them more suitable as explanatory visualizations?

Not really, nothing looks good or make sense yet for anyone but me. I'm going to try and do some stats manipulation of the data to see if I can find a better solution.