Individual Assignment Report

Mathilda Ahr

January 15th, 2020

1 Introduction

What was the total number of fatalities in Indonesia in year 2004 ?

2 PLAN

Make the JSON file into csv for R:

- should be a table with observations of violence (the 315 cases) in rows and characteristics in columns. Each dictionary would be a row and each key a column. USE python
- filter out any other country than Indonesia and any other year than 2004 USE R (or Python)
- take out some useless keys. Keys that we want to keep: id, year, conflict_name, country, data_start, date_end, best, high, low USE Python

Start working with R around 14:30: data should be csv format
Make plot with ggplot (start max 15:00) USE R
Be done coding at 15:45

3 Methods

3.1 Exploring data

Data is a list of dictionaries indicated by curly brackets within square brackets. There are 377 lines in the sample data (bash: 'wc UCDP_conflict_sample.json').

Looked at first few lines:
"country": events took place in Turkey
"best": the best estimate number of fatalities is 33
"type_of_violence": type of violence 1 which is state-based conflict
"year": took place in 1997

3.2 Choosing country and year

I used the terminal to effectuate the first step of data exploration. I mainly used the command 'grep' and 'wc' in order to pick a country and a year with a sufficient data amount. The manual helped checking what the number given by 'wc' represented.

Find a country: grep '"country": "Indonesia"' conflict_data_full_lined.json — wc -¿ 1615.
There are 1615 cases of violence for Indonesia. 1615 entries for one country seems enough to answer the question.
Find a year: grep '"country": "Indonesia"' conflict_data_full_lined.json — grep '"year": 2004' — wc -¿ 315.
There are 315 entries for Indonesia for the year 2004. 315 cases of violence in a single year for the country Indonesia is more than enough to answer the research question.