Air Temperature Exploration

Laura Keen

May 6, 2020

```
library(haven)
library(tidyverse)
library(foreign)
library(broom)
library(knitr)
library(dplyr)
library(readr)
```

Introduction

This document shows data preparation and visualization of annual air temperature data for 24 African countries from 2000-2017. These countries were selected based on research project scope that focuses on data available via Afrobarometer dataset. All data was downloaded courtesy of AidData, a Research Lab at William & Mary.

Test with Uganda

First, I downloaded air temperature from Uganda to examine the structure of the dataset. I clean it and exported a new csv file.

```
uga_air_temp <- read_csv("data/data_raw/UGA_air_temp.csv")
uga_air_temp <- Filter(function(x)!all(is.na(x)), uga_air_temp)
uga_air_temp <- uga_air_temp[,c(58, 59, 56, 57, 60:69,1:55)]
write_csv(uga_air_temp, "data/data_intermediate/uga_air_temp.csv")
```

I downloaded air temperature data from AidData for all Afrobarometer countries for which air temperature data was available (all but Capo Verde). I read in all the files and merged them into one dataframe.

All Afrobarometer Countries

```
getwd
```

```
## function ()
## .Internal(getwd())
## <bytecode: 0x7fd44f21b290>
## <environment: namespace:base>
```

```
mydir = ("~/GitHub/Climate_and_Conflict/data/data_raw/")
myfiles = list.files(path=mydir, pattern ="air_temp", full.names=TRUE)
air_temp <-list.files(path = mydir, pattern = "*.csv", full.names = T) %>%
  map_df(~read_csv(., col_types = cols(.default = "c")))

# delete empty columns
air_temp <- Filter(function(x)!all(is.na(x)), air_temp)

# reorder columns
air_temp <- air_temp[,c(1, 57,58,56,59,60:81,2:56)]

# map(air_temp, ~sum(is.na(.)))

# delete columns with mainly NA values
air_temp <- air_temp[,-c(5,6,8,10,12,16,18:27)]

# delete odd string in column names
names(air_temp) = gsub(pattern = "^udel_air_temp_v501_mean.", replacement = "", x = names(air_temp))
air_temp <- air_temp[, -c(1,66)]
```

```
df <- air_temp %>%
  gather(key = "year", value = "estimate", -c(1:10))

df_2 <- df %>%
  separate(year, into=c("year", "stat"))

df_3 <- df_2 %>%
  spread(key = "stat", value = "estimate")

# sapply(df_3, class)
# Convert character values to numeric
cols.num <- c("Shape_Area", "Shape_Length", "year", "max", "mean", "min")
df_3[cols.num] <- sapply(df_3[cols.num],as.numeric)

# round numeric values
df_3 <- df_3 %>%
  mutate_if(is.numeric, round, digits=4)

# reorder columns again
df_3 <- df_3[,c(1:3,10,4,5,7,8,6,9,11:14)]
df_3 <- df_3[order(df_3$NAME_0,df_3$year),]
write_csv(df_3, "data/data_intermediate/all_air_temp.csv")
```

Visualizations

I created a new dataset with the mean annual mean temperature by country and plotted the trends over time.

```
df_plot <- df_3 %>%
  group_by(NAME_0, year) %>%
  summarise(annual_mean = mean(mean, na.rm = TRUE)
```

```
            )

df_plot %>% ggplot(aes(x=year, y=annual_mean)) +
  geom_line(aes(color=NAME_0)) +
  labs(title = "Mean Temperature Over Time", x="year", y="mean temperature", color="Country") +
  theme_minimal()
```