

## Assignment 5: Ethnic Diversity

Created: 2020-01-18

Due Date:

### Instructions

Prepare a presentation about the effect of ethno-linguistic fractionalization on contemporary outcomes.

- Slide 1: Calculate your country's ethno-linguistic fractionalization (ELF) score. (Although you will have this score in your dataset, please show the calculations from the raw data).
- Slide 2: Discuss whether you think this score accurately depicts the degree of heterogeneity in your country, and whether a different measure might be more descriptively valid.
- Slide 3: Use a scatter plot of Sub-Saharan countries to illustrate the general relationship between ethno-linguistic fractionalization and the percent change in per-capita GDP from 1960 to 1992, highlighting your country.
- Slide 4: Interpret your results to answer whether and how ethno-linguistic fractionalization in your country can account for its relative growth rate.

---

### 1. Preparation

If you completed the R tutorial or the previous problem sets, the tidyverse package should already be installed on your computer. If not, run this code chunk.

```
install.packages("tidyverse")
```

Similarly, if you haven't installed the gghighlight package, do so by running this code chunk.

```
install.packages("gghighlight")
```

Now, load both packages.

```
library(tidyverse)
library(gghighlight)
```

Finally, load the data set from the URL provided.

```
africa_df <- read_csv('https://african-states-book.info/img/portfolio/problem-sets/data/05-data.csv')
```

As usual, you'll need to assign your country name and your country's ISO 3-character code to the following variables. The `my_country` variable will be used to label plots, while the `my_iso_code` variable will be used to uniquely identify your country in the dataset.

```
my_country <- "Kenya"
my_iso_code <- "KEN"
```

## 2. Computing Ethno-Linguistic Fractionalization

Ethno-linguistic fractionalization can be measured as an index that takes values ranging from 0 to 1. Values closer to 0 indicate ethnic homogeneity, while values close to 1 indicate ethnic diversity. More precisely, we can interpret  $1 - ELF$  as the probability that two randomly selected individuals share a common ethnicity. Fearon (2003) gives the following definition of ethno-linguistic fractionalization:

$$ELF_j \equiv 1 - \sum_{i=1}^N p_{ij}^2$$

Where  $p$  is the proportion of the population from ethnic group  $i$ , and  $ELF$  is evaluated for each country  $j$ .

First, let's use `head` to take a look at the raw data. You can also look at the whole dataset in RStudio by typing `View(africa_df)` in the console, or by clicking on `africa_df` under the Data heading in the Environment pane (which, by default, will be in the upper-right of the screen).

```
head(africa_df)
```

country	iso3c	group	gpro	ef	avg_gdp_grwth92
Guinea-Bissau	GNB	balanta	0.27	0.8177	2.019133
Guinea-Bissau	GNB	fulani	0.22	0.8177	2.019133
Guinea-Bissau	GNB	mandjako-papel	0.20	0.8177	2.019133
Guinea-Bissau	GNB	mandinka	0.13	0.8177	2.019133
Guinea-Bissau	GNB	cape verdean	0.05	0.8177	2.019133
Guinea-Bissau	GNB	brame	0.04	0.8177	2.019133

As you can see, the data is already in a neat, organized format. Each row contains data on one ethnic group within a country. The

gpro column represents the proportion of the country's population accounted for by that ethnic group. To calculate *ELF* from the raw data, we'll simply use `dplyr` to apply the ELF function to each country.

```
africa_df <- africa_df %>%
  group_by(country) %>%
  mutate(my_ef = 1 - sum(gpro^2))
```

Before moving on, let's verify that our calculations for your country return the same value calculated by Fearon (2003).

```
africa_df %>%
  filter(iso3c == my_iso_code) %>%
  summarise(
    "My EF" = unique(my_ef),
    "Fearon's EF" = unique(ef)
  )
```

country	My EF	Fearon's EF
Kenya	0.852044	0.852044

Based on this table, our calculations have accurately replicated the data presented in Fearon (2003).

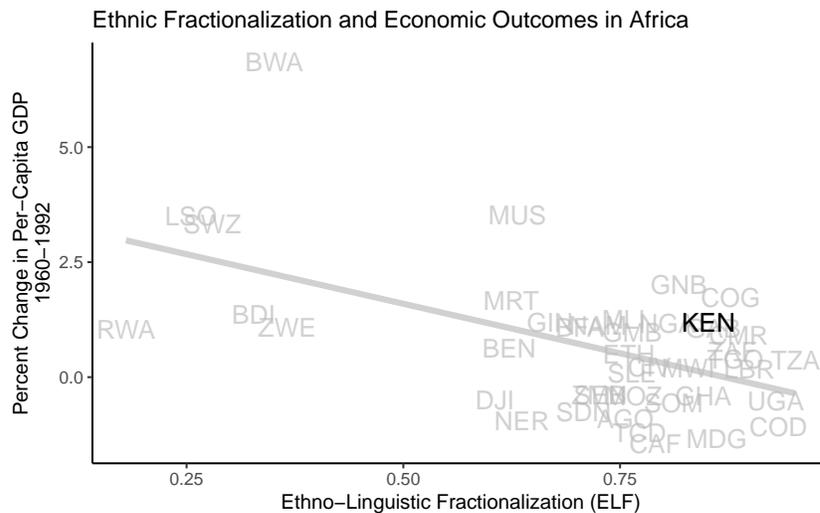
### 3. Fractionalization and Average Percent Change in Per-Capita GDP (1960 to 1992)

Now we'll do some data manipulation with `dplyr` to get the data set ready to plot. To reformat the data contains a single row for each country, we'll select the variables we wish to plot, then use `unique` to remove duplicate rows.

```
df <- africa_df %>%
  select(country, iso3c, avg_gdp_grwth92, my_ef) %>%
  unique()
```

Now that the data is prepared, we'll make a scatter plot of the relationship between EF and annual percent change in per-capita GDP. We'll also highlight the position of your country relative to the rest of Sub-Saharan Africa.

```
ggplot(df, aes(x = my_ef, y = avg_gdp_grwth92, label = iso3c)) +
  geom_text(size = 3) +
  geom_smooth(method = "lm", se = FALSE) +
  gghighlight(iso3c == my_iso_code) +
  theme_classic(7) +
  labs(
    x = "Ethno-Linguistic Fractionalization (ELF)",
    y = "Percent Change in Per-Capita GDP\n1960-1992",
    title = "Ethnic Fractionalization and Economic Outcomes in Africa"
  )
)
```



As in the previous problem set, you can save this plot as an image or PDF using the `ggsave` function.

```
ggsave('figure-05-1.png')
```

In R Studio, you can also save the plot by clicking the “Export” button at the top of the plots window.

What does this plot tell you about the general trend of ethno-linguistic fractionalization and percent change in per-capita GDP? Is your country above or below the trend line? What other factors might explain average growth in per-capita GDP for your country?

### Works Cited

Fearon, James D. 2003. “Ethnic and Cultural Diversity by Country.” *Journal of Economic Growth* 8 (2): 195-222.