

POLS 7012: Problem Set 5 (Answer Key)

Joe Ornstein

Due September 30, 2020

Instructions

Do democracies or autocracies devote a greater share of national resources to the military? In this problem set, you will test your new skills importing, tidying, transforming, and summarizing data to answer this question. You will need two datasets:

- The [Polity 5 Time Series](#) for the “Polity Score” of each country over time (-10 is strongly autocratic, +10 is strongly democratic).
- The [National Material Capabilities](#) dataset, for information on each country’s military capabilities.

As you will see, there can be more than one way to answer this question. But our approach will be to compute the number of military personnel by country as a share of the total population, and compare that value for democracies and autocracies over time. To do so, complete the following steps:

1. Download the two datasets and read them into R.
2. Keep only the variables you need, and merge the two datasets.¹
3. Create a new variable equal to the number of military personnel per 1,000 people in the country’s total population.
4. Create a new `regime` variable, coding countries as a “Democracy” if their Polity2 score is greater than five and an “Autocracy” otherwise.
5. Compute the average share of the population in the military for autocracies and democracies over time (`group_by` regime type and year), as well as the number of countries in each category (`summarize` function). Assign this new grouped dataframe to an object.
6. Plot your results (`geom_line` would be nice).

Load the Data

If you download the files to your `data/` folder, then this code will read them into R.

```
# load tidyverse and readxl (library to read Excel documents)
library(tidyverse)
library(readxl)

# read Polity 5 data into memory
polity5 <- read_xls(path = 'data/p5v2018.xls')

# read NMC dataset
nmc <- read_csv(file = 'data/NMC_5_0.csv')
```

¹Consult the codebooks (included at the links above) to determine which variables to use as keys for the merge, and which variables represent your outcomes of interest.

Merge the Data

Next, keep the variables you need and merge by `ccode` and `year`.

```
# keep only the variables we want:
# country, year, military personnel, population, and polity2 score
nmc <- nmc %>%
  select(ccode, year, milper, tpop)

polity5 <- polity5 %>%
  select(ccode, country, year, polity2)

# merge by ccode and year
data <- left_join(polity5, nmc,
  by = c('ccode', 'year'))
```

Clean Up The Data

Notice that not all the country-years have entries in NMC, and there are some missing values in `polity2` and `milper` coded as `-88`, `-66`, and `-9`. Let's filter that out and create our two new variables.

```
data <- data %>%
  filter(!is.na(milper),
    polity2 != -88,
    polity2 != -66,
    milper != -9) %>%
  mutate(regime = case_when(polity2 > 5 ~ 'Democracy',
    TRUE ~ 'Autocracy'),
    military_personnel_per_thousand = milper / tpop * 1000)
```

Summarize The Data

Finally, summarize and plot the results by regime type and year.

```
summary_data <- data %>%
  group_by(regime, year) %>%
  summarize(avg_military_strength = mean(military_personnel_per_thousand),
    num_countries = n())

head(summary_data, 5)

## # A tibble: 5 x 4
## # Groups:   regime [1]
##   regime    year avg_military_strength num_countries
##   <chr>    <dbl>           <dbl>          <int>
## 1 Autocracy 1816             8.46             13
## 2 Autocracy 1817             7.96             13
## 3 Autocracy 1818             7.66             13
## 4 Autocracy 1819             8.29             15
## 5 Autocracy 1820             8.16             15

figure1 <- ggplot(summary_data) +
  geom_line(mapping = aes(x=year, y=avg_military_strength, color = regime)) +
  theme_classic() +
  labs(x = 'Year', y = 'Military Personnel Per Thousand People',
    color = 'Regime Type')
```

figure1

