

POLI 502 FA20: Homework 2

Due at the start of class on Wednesday, September 30, 2020

Directions. Answer each question completely. Submit your completed assignment as a L^AT_EX-typeset PDF file, along with R script file and data such that I can replicate all output exactly just by changing the working directory and running the script. You might consider using knitr to integrate R code into L^AT_EX-typeset answers. Please upload a zipped folder containing all files to Blackboard.

Exercise 1

Run the following code to install the World Development Indicators package and create a data frame

```
# Install and load World Development Indicators package
install.packages("WDI", repos="http://cran.us.r-project.org")
library(WDI)

# Load a data frame using 5 WDI variables over 10 years
WDI.data <- WDI(country="all", indicator=c("IQ.CPA.GNDR.XQ",
    "SP.DYN.CBRT.IN", "AG.LND.AGRI.K2", "AG.LND.AGRI.ZS"), start=1995, end=2005)
```

A. Aside from the ID variables (i.e., iso code, country, and year), explain what each variable is capturing, and its level of measurement (if numerical, explain whether it is discrete or continuous, and whether it is a ratio variable).

Hint: You will need the codebook. Googling the variable name takes you right to a useful webpage. Are there any limitations to using some of these variables over this time period?

B. Although there is a variable called “country,” you might notice that it includes regions as well as countries. Subset the data to exclude all regions, retaining only countries. **Hint:** `Filter()`, in conjunction with one other command you have learned, could be useful to accomplish this task easily.

Note: for parts C through E, ignore the presence of lakes/rivers/etc.

C. If you were to point to a random spot on a map of Honduras in 1998, what is the probability that you would be pointing to an area in which the land (in the actual country) is agricultural?

D. If you were to point to **three** random spots on a map of Albania in 2004, what is the probability that **all three** spots would represent agricultural land?

E. If you were to point to **six** random spots on a map of Chad in 2002, what is the probability that **at least one** spot would represent agricultural land?

Exercise 2

Go to <http://correlatesofwar.org> and download the appropriate data necessary to create data frames for:

1. inter-state war data from the COW war data 4.0
2. dyadic trade data from the COW trade data 4.0

A. Create new data (e.g., data frame or tibble) that is a subset of dyadic trade data including only dyads in which at least one state is located in Europe (ccode between 200 and 399).

B. Create new data to subset the inter-state war data (at the same unit of analysis) including only wars where at least one participant is a European state. **Hint:** *You will likely need to use `mutate()` with `group_by()`.*

C. Taking the subset data frames from parts A and B (and then summarizing as necessary), merge to create a new data frame where the unit of observation is the year. It should include the entire year range that is covered in both the dyadic trade and inter-state war data. It should include variables identifying 1) mean dyadic trade and 2) the count of wars (*not* states involved in war) that begin.

D. With the new data frame from part C., Use `ggplot()` to create a scatterplot where the x-axis represents *the (natural) log of* mean dyadic trade, while the y-axis represents the count of war onsets.

From the Text

Note: Do all math in R and submit the (annotated) code. Show all your work!

Diez et al.: Chapter 3 Exercises: 3.2, 3.6, 3.17, 3.35 (skip the standard deviation)