Data Analytics for Social Science
Project 2: Country-level data

Johan A. Elkink
jos.elkink@ucd.ie

due 11 April 2017

This assignment is about country-level data, where we have observations on each country, typically over a number of different years. In this case, however, we just select data from 2001 to avoid statistical complications that arise from using time-series data. The outcome variable we will investigate is membership of the International Criminal Court (ICC). In 2001 the distribution is depicted in Table 1. In other words, not all countries signed up to the ICC treaty, and not all of those who signed ratified subsequently. The question then is, how do we explain this variation and how do we predict, for a given country, whether that country is likely or not likely to sign up?

The model specifications are already in labs 7 and 8 and can be used directly from that. You might also want to expand models by adding an additional explanatory variable, and you can choose whether you want to investigate ratifying or signing the ICC treaty, but it will suffice to just use the lab output. The analysis should include at least one regression model and one tree-based model and you might want to include some graphical descriptions of the data or key relationships, based on what we did in earlier parts of the course.[1] The focus is of course on the interpretation and contextualisation, following the outline in the syllabus: "Each essay should consist of a short introduction, a description and motivation of the data and methods used (approximately 25% of the essay), the analysis including necessary graphs and tables (approximately 35%), and an interpretation and conclusion (approximately 40%). Everything needs to be properly referenced." The total length, excluding tables and bibliography, should be between 2,000 and 2,500 words.

While we use the data from Ross and Voeten (2016), who investigate joining international organizations more generally, and while we are inspired by their work to select

[1]Note that it will often be easier to work with the "design matrix" derived from the output from a regression—as in the labs—than with the original data set, as all missing data will have been removed.

	Not ratified	Ratified	Sum
Not signed	88	1	89
Signed	60	45	105
Sum	148	46	194

Table 1: Ratification and signature numbers in 2001.

Variable	Description
iccRatified	Dummy variable whether the country has ratified the ICC treaty in that year or earlier.
iccSigned	Dummary variable whether the country has signed the ICC treaty in that year or earlier.
polity2	Regime type scale, from -10 = dictatorship to +10 = democracy.
democracy	Binary democracy variable, 0 = non-democracy, 1 = democracy.
logoil	The log of total oil exports.
lngdp	The log of Gross Domestic Product (GDP).
tradegdp	The proportion of international trade (imports and exports) over GDP.
lnpop	The log of the total population.

Table 2: Overview of variables in replication data of Ross and Voeten (2016), merged with data on ICC ratification.

some of the explanatory variables, it might be helpful to also look at works that are directly related to the ICC. On Blackboard you can find the introduction to the PhD thesis by Kevin Coffey (2015), including an extensive bibliography, discussing the various motivations to sign or ratify the ICC treaty, in particular for Subsaharan African states. Of course, this assignment is about all countries, not just that particular region.

Variables

We will make use of the replication data of a published paper: Ross and Voeten (2016). This data is so-called panel data, which means that it has a set of units, observed over a range of different time periods. In this case countries over a number of years.

Ross and Voeten (2016) refer to the concept of "structured international organisations", which are international organisations that have a reasonable level of organisational structure, without being a full-blown supranational organisation. I.e. more structured than a trade agreement, but less structured than the World Bank. We will look at one specific organisation, namely the International Criminal Court, and data for this particular organisation has been downloaded from Wikipedia[2] and then merged with the Ross and Voeten (2016) data. The code on how this was done is below in case you are curious how to download Wikipedia data directly into R.

There are quite a lot of variables in the replication data—it looks like they merged a lot of different data sources and then distributed the whole file as "replication data", without removing first the unused variables. This gives us a lot to experiment with, but in the absence of a codebook, it is at times unclear what different variables represent. Some are obvious from the name, though, or are mentioned in the paper. Based on my interpretation of the names, key variables are described in Table 2. Generally, variable names that start with "log" or "ln" are logged versions of the original variables, while variable names ending with "sq" are squared versions of original variables.[3]

[2] https://en.wikipedia.org/wiki/States_parties_to_the_Rome_Statute_of_the_International_Criminal_Court

[3] See `names(ross2001)` for a list of all variables names.

References

Coffey, Kevin. 2015. Why States Commit to International Criminal Justice: Support for the International Criminal Court in Sub-Saharan Africa PhD thesis University College Dublin.

Ross, Michael L. and Erik Voeten. 2016. "Oil and international cooperation." *International Studies Quarterly* 60(1):85–97.
URL: *https://www.sscnet.ucla.edu/polisci/faculty/ross/publications.html*

Downloading the ICC data

```
library(rvest)
library(rio)

ross <- import("ReplicationdataRossVoeten.dta")

icc_page <- read_html("https://en.wikipedia.org/wiki/States_parties_to_the_Rome_Statu

members <- icc_page %>% html_nodes("table") %>% .[[1]] %>% html_table()

cname <- substr(trimws(members$`State party[1]`), 1, 100)
cname <- gsub("[[].[]]", "", cname)

names <- data.frame(cname = cname)

rossCode <- unique(ross[, c("ccode", "ctryname")])
rossCode <- rossCode[rossCode$ctryname != "",]

names <- merge(names, rossCode, by.x = "cname", by.y = "ctryname", all.x = TRUE)

names$ccode[names$cname == " Switzerland"] <- 225
names$ccode[names$cname == "Andorra"] <- NA
names$ccode[names$cname == "Antigua and Barbuda"] <- 58
names$ccode[names$cname == "Congo, Democratic Republic of the"] <- 490
names$ccode[names$cname == "Congo, Republic of the"] <- 484
names$ccode[names$cname == "Cook Islands"] <- NA
names$ccode[names$cname == "Gambia, The"] <- 420
names$ccode[names$cname == "Korea, South"] <- 732
names$ccode[names$cname == "Macedonia, Republic of"] <- 343
names$ccode[names$cname == "Nauru"] <- NA
names$ccode[names$cname == "ote d'Ivoire ! Cte d'Ivoire"] <- 437
names$ccode[names$cname == "Palestine"] <- NA
names$ccode[names$cname == "Saint Kitts and Nevis"] <- 60
```

```
names$ccode[names$cname == "Saint Lucia"] <- 56
names$ccode[names$cname == "Saint Vincent and the Grenadines"] <- 57
names$ccode[names$cname == "Serbia"] <- 345

members$ccode <- names$ccode
members$signed <- as.integer(substr(members$Signed, nchar(members$Signed) - 4, 100))
members$ratified <- as.integer(substr(members$`Ratified or acceded`,
        nchar(members$`Ratified or acceded`) - 4, 100))

export(members[, c("ccode", "signed", "ratified")], file = "icc.dta", version = 10)
```