

XAVIER FERNÁNDEZ-I-MARÍN

FONTS D'INFORMACIÓ I INDICADORS PER ESTUDIS INTERNACIONALS

Contents

Prefaci 5

1 *Introducció* 7

1.1 *Organització del material* 8

1.2 *Pocs experts o molts ignorants?* 8

1.3 *Combinacions simples o matemàtiques complexes?* 9

1.4 *Gestió de dades* 10

1.5 *Un exemple: quants estats hi ha al món?* 12

1.6 *Glossari de conceptes* 13

2 *Context socioeconòmic* 19

2.1 *World Development Indicators* 19

2.2 *Desigualtat* 25

3 *Sistemes polítics* 29

3.1 *Democràcia* 29

3.2 *Limitacions al poder: vetos* 36

3.3 *Descentralització i federalisme* 38

3.4 *Regulació* 39

3.5 *Societat, partits polítics i eleccions* 42

4	<i>Relacions: Contagi, Imitació, Dependència</i>	43
4.1	<i>Comerç</i>	43
4.2	<i>Distància</i>	47
4.3	<i>Relacions històriques i diplomàtiques</i>	50
4.4	<i>Guia de lectura</i>	50
4.5	<i>Guia de lectura (Tena)</i>	50
4.6	<i>Guia de lectura (Skrede Gleditsch)</i>	51
4.7	<i>Guia de lectura (Barbieri, Keshk i Pollins)</i>	52
5	<i>Organitzacions Internacionals i tractats</i>	53
5.1	<i>Correlates of War: la base</i>	53
5.2	<i>Autoritat de les organitzacions internacionals</i>	56
5.3	<i>Llei internacional: tractats</i>	56
	<i>Referències</i>	59

Prefaci

El material que teniu a les mans és una barreja deliberada de recursos d'informació i de pistes per fer-los servir de manera pràctica. Si hagués optat per presentar només recursos d'informació, el material hagués estat un simple llistat, més o menys ordenat, més o menys actualitzat, però un llistat sobre el qual no se'n pot fer gaire res més que tenir-lo desat en algun lloc a mà per si en qualsevol moment convé consultar-lo. Però he optat per incloure també maneres de fer servir aquestes fonts d'informació per tal que puguin ser realment útils, que serveixin per a alguna cosa més que per a la consulta. D'aquesta manera, la intenció és que al llarg dels estudis, en qualsevol feina, treball, assaig que se us demani, pugueu afegir una figura o una taula amb dades tractades per vosaltres mateixos que doni el context necessari per presentar l'argument que voleu fer entendre, ja sigui en el context d'un conflicte, una situació, un esdeveniment, una guerra o un acord diplomàtic.

1

Introducció

La raó última de qualsevol disciplina científica és explicar per què passen les coses. Per saber *per què* passen les coses hem de saber abans *què* passa, ser capaços de descriure fenòmens. I per poder descriure necessitem poder mesurar i classificar els fenòmens que estudiem.

Per tant, a la base de qualsevol disciplina científica hi trobem la mesura i la classificació. De fet, la revolució científica iniciada a la il·lustració es basa fonamentalment en, dins les ciències naturals, recollir, mesurar i classificar. Es recullen i es mesuren mostres, que tant poden ser animals, vegetals com objectes celestes, fòssils d'Europa o plantes d'Amèrica i les illes del Pacífic. Darwin s'embarca al *Beagle* amb l'encàrrec de recollir mostres. En el procés, però, després d'observar, mesurar i classificar tot allò que va trobant, va elaborant una teoria de per què tot allò que ha trobat és al lloc on és, com hi ha pogut arribar i quins mecanismes s'hi han vist involucrats. Newton i Galileu també comencen les seves observacions amb una sistematització del que veuen, seguit per un procés de classificació (com dividir els cossos celestes, afegint categories i tipologies).

A les ciències socials és l'economia la que es fixa per primer cop en la mesura. De fet, el mot "estadística"¹ prové precisament de l'interès per poder mesurar el volum de recursos dels estats. L'objectiu, però, no era altre que el de poder fixar tributs amb més precisió, i no pas el gust pel coneixement o una neutra i asèptica voluntat de conèixer com de grans eren les incipients burocràcies. És, per tant, en un segon estadi quan arriba als *per quès*, als motius.

Tant la física com l'economia ho tenen relativament fàcil per obtenir mesures dels conceptes que estudien. La física només cal que es posi d'acord en l'escala de mesura de la distància o el temps, però ambdós són conceptes relativament simples de definir i mesurar². En canvi, tant per a la biologia com per a les ciències socials la dificultat per acordar una mesura pels conceptes que s'estudien és molt major. Mesurar un gen, definir una espècie, mesurar la intel·ligència o definir una democràcia són conceptes més ambigus que requereixen més que no pas acordar una escala de mesura.

És en aquest context on s'insereixen aquests materials docents que teniu

¹ En català és preferible fer servir "estatística", que seria la paraula que correspondria al manlleu de la llengua original, i no pas a "estadística", que és el concepte que prové no pas de la llengua original, sinó del castellà. Vegeu "Gabriel Bibiloni, Amb bones paraules: estatística": <https://bibiloni.cat/ambbonesparaules/estadistica.html>

² Tot i que és precisament la definició mateixa d'espai i temps la que Einstein més tard fa trontollar amb la seva teoria de la relativitat. No obstant això, la idea que les mesures en ciències físiques són conceptualment més simples que en ciències socials segueix essent vàlida.

a les mans. Cal entendre les fonts d'informació i els indicadors en relacions internacionals en un context més general on la ciència prova de definir, acotar, mesurar, i operacionalitzar idees que, posteriorment, seran emprades per entendre els *quès* i els *per quès*. Els següents materials docents estan pensats per introduir la operacionalització de conceptes d'ús habitual en relacions internacionals. En aquest procés és normal que hi hagi també espai per a la discussió de què és una democràcia, què és la desigualtat, etc. però l'objectiu no és pas aquest: l'objectiu és tenir eines per entendre com s'operacionalitza, es mesura i es classifica la democràcia, o com s'ordenen una sèrie de països pel seu nivell de desigualtat.

1.1 Organització del material

Els objectius dels materials docents són els següents:

1. Explicar quins són els conceptes i idees que la disciplina de les relacions internacionals i la política comparada està interessada a recollir, sistematitzar i mesurar.
2. Explicar com es recullen les dades per generar mesures de les idees i conceptes a estudiar.
3. Descriure com es tracten les dades per arribar a mesures i conceptes propis de la disciplina.

L'organització dels materials en quatre mòduls respon a la primera part: s'ha partit d'una categorització de conceptes i idees en quatre parts, que corresponen als quatre mòduls i que aquí només s'introduiran:

- Característiques socioeconòmiques
- Sistemes polítics
- Relacions
- Organitzacions internacionals

De la segona part (com es recullen les dades) en farem esment només en alguns casos, aquells que mereixen un tractament rellevant dins la disciplina.

Quant a la tercera part, la resta de la introducció s'ocupa de presentar els mètodes més habituals de tractament de les dades i de donar una introducció pràctica al programari que ens permetrà incorporar les dades a l'ordinador per tractar-les.

1.2 Pocs experts o molts ignorants?

El "Big Data" és un concepte que s'ha posat molt de moda actualment. Malgrat que no hi ha un consens general sobre el que és i el que no (i que tampoc no ens interessa discutir aquí), és cert que ha tornat a posar de relleu un aspecte que en el cas de les ciències socials és essencial: quina estratègia

seguim quan no disposem d'instruments de mesura directa dels fenòmens i conceptes que utilitzem.

Com pots saber si tens febre? La reacció normal és la de buscar un termòmetre i que sigui aquest instrument el qui ens doni la resposta. Però en cas de no tenir-ne un a mà, com ho pots fer? Molt probablement li demanis a la persona a qui tens al costat un veredict. Hi confiaràs? Hi confiaries més si sabessis que aquesta persona és personal mèdic o d'infermeria en contacte habitual amb temes de salut? O potser t'estimes més no deixar el veredict en mans d'una única persona i en canvi optes per demanar-ho a deu persones que tens al voltant i ponderes les seves respostes? En altres paraules, confies més en un únic expert sanitari o en molts ignorants?

Aquest és precisament el cas en relacions internacionals. D'entrada, com a ciències socials, malauradament, no tenim "termòmetres" per mesurar les constants dels països. Com podríem saber el nivell de democràcia d'un país? No hi ha màquines que puguin respondre a tal pregunta, de manera que el següent pas és el mateix que en el cas de no tenir un termòmetre a mà: és millor el que ens digui un expert o el que ens diguin molts ignorants?

Tradicionalment la recerca acadèmica s'ha basat en les opinions d'experts. Però darrerament, el "Big Data" i la seva promesa no de *molts* ignorants, sinó d'una quantitat enorme d'ells, ha entrat en joc. Si bé encara és d'hora per dir si finalment el Big Data triomfarà a l'hora de generar dades en relacions internacionals, el cert és que caldrà estar atents als esdeveniments dels propers anys i a aquesta tendència de generar mesures a través de molts ignorants, en contraposició a pocs experts.

1.3 *Combinacions simples o matemàtiques complexes?*

Un cop obtingudes les dades d'allò que ens interessa mesurar, el següent pas és el d'arribar a un número que resumeixi el resultat. Seguint l'exemple de la temperatura, el resultat de la medicció del termòmetre ha de ser en una escala coneguda i comparable. En ciències naturals no hi ha massa problemes i les mesures se solen definir sempre en base a constants universals. Però en el cas d'obtenir dades en escales sense una constant o una referència clara, què podem fer?

Les solucions passen per fer servir escales que siguin almenys matemàticament convenients. Per exemple, considerar que el zero és el valor més baix, i l'1 o el 100 el valor màxim. O considerar el zero com el valor mitjà i deixar un marge de números negatius i números positius. O estandarditzar el resultat a base de dividir per la desviació estàndard. En tot cas, s'acabi en una o altra escala, el procediment sempre compta. Per exemple, en el cas de la democràcia: si acordem que la democràcia té tres dimensions (diguem-ne llibertat política, selecció del legislatiu en base a eleccions i participació mínima), com combinem les tres?

En la majoria dels casos la solució per la qual opten les institucions que

s'han encarregat de generar aquestes dades és la de calcular mitjanes simples, o agregar (sumar) les diferents variables per generar un indicador global. És al procediment més habitual quan se sent a parlar de “indicador de globalització”, “índex de corrupció”, “nivell de democràcia”, etc. En realitat, però, aquesta solució basada en la combinació simple de variables sol assumir una sèrie de coses que no sempre són justificades o argumentades.

És aquí on entren els models estadístics, que permeten combinar les dades de diferents variables per crear indicadors d'una manera més sistemàtica i acurada. Mètodes com l'anàlisi factorial, els models d'ítem-resposta o l'anàlisi de components principals ajuden els investigadors a destriar millor quines són les ponderacions de cada variable en el conjunt de l'indicador final.

Un exemple d'indicador basat en l'agregació de diferents variables seria aquest:

$$\text{Indicador} = \text{Variable 1} + \text{Variable 2} + \text{Variable 3}$$

En canvi, un exemple d'indicador basat en models estadístics seria semblant a:

$$\text{Indicador} = P_1 * \text{Variable 1} + P_2 * \text{Variable 2} + P_3 * \text{Variable 3}$$

on P_n són els pesos o ponderacions corresponents a cada variable. Aquests pesos en calculen dins el mateix model en base a les mateixes dades. D'aquesta manera, per exemple, si la Variable 1 no discrimina bé entre aquests casos amb alt o baix valor de l'indicador, se li atorga un pes menor.

En resum, que no hi ha cap motiu científic per generar indicadors a través de combinacions simples de variables, més enllà de la conveniència i el fet que assignar pesos a vegades es pugui considerar un tema polític. Per exemple, l'índex de desenvolupament humà creat pel PNUD de les Nacions Unides ha optat per considerar que el desenvolupament es divideix en tres parts iguals: economia, educació i salut. Si es feien combinacions diferents i s'atorgaven pesos diferents, la posició dels Estats Units es veia compromesa i queia per sota de països com Cuba, de manera que es va optar per deixar l'índex com una mera agrupació, assumint que cada dimensió aporta el mateix, sense considerar què hi diuen les dades.

Per tant, sempre que sigui possible és preferible que facis servir indicadors que tinguin un model estadístic robust al darrera, i que no siguin una mera agrupació de valors creats amb una mitjana o una suma.

1.4 Gestió de dades

R és un programa d'anàlisi estadística que, entre altres avantatges, permet dos aspectes claus de qualsevol recerca: la gestió de dades i la visualització de resultats.

R no és estrictament un programa, sinó un llenguatge de programació amb el qual podem interactuar amb l'ordinador. Per instal·lar R cal [descar-](#)

regar la [versió corresponent](#), que es pot obtenir sense restriccions donat que és programari lliure (tant en el sentit de preu com en el sentit de llibertat per modificar-lo).

Per tal de treballar amb R és convenient fer servir una interfície gràfica que permeti de manera fàcil poder navegar pel sistema de fitxers de l'ordinador, tenir algunes plantilles per editar codi, autocompletar les comandes, o acolorir la sintaxi, entre altres. RStudio és una solució popular i raonable perquè té [versions per diferents plataformes](#) (Linux, Mac, Windows).

Una de les virtuts del programa és la de ser fàcilment extensible per qualsevol que tingui un nivell mitjà de programació amb el llenguatge, de manera que el nombre del que s'anomenen llibreries o paquets és avui en dia molt elevat i abasta un rang molt ample de disciplines, tècniques i usos. Abans de fer servir una llibreria per primer cop en el teu ordinador cal que la instal·lis, però només un cop. La resta de vegades que la vulguis fer servir en tindràs prou amb carregar-la en memòria amb `library()`. Els paquets que necessitaràs per seguir el material docent s'instal·len de la següent manera:

```
install.packages(c(
  "dplyr", "tidyr", "ggplot2",
  "countrycode", "WDI",
  "openxlsx", "readstata13", "readxl",
  "network", "GGally",
  "maps", "geosphere",
  "knitr"), dependencies = TRUE)
```

R és molt versàtil en quant a la capacitat d'incorporar dades de fonts molt diverses. Permet combinar fàcilment diferents matrius de dades fent sobrevir columnes coincidents (funcions `merge()` i `left_join()`), i també redimensionar les matrius de dades de manera ràpida (funcions `gather()` i `spread()`), que veurem més endavant. Els paquets per a la gestió del que s'anomena "dades netes" ("*tidy data*") i que caldrà carregar en cada sessió són els següents:

```
library(tidyr)
library(dplyr)
```

Per entendre com funciona la gestió de les *tidy data* vegeu els següents tutorials sobre `tidyr` en format [web](#), `dplyr` en format [web](#) i [vídeo](#) i un tutorial [conjunt](#).

Les capacitats gràfiques del programa també són destacables, en tant que de manera molt senzilla es poden obtenir figures de les dades a analitzar que permeten visualitzar els aspectes més rellevants. En aquest sentit, en aquests materials s'utilitzaran les capacitats del paquet `ggplot2`. Vegeu un tutorial complet de les [capacitats gràfiques de ggplot2](#) i en [vídeo](#).

```
library(ggplot2)
```

1.5 Un exemple: quants estats hi ha al món?

La primera pregunta que t'has de fer abans de començar a buscar dades en relacions internacionals és ¿de quantes unitats he de recollir dades? ¿quants estats hi ha al món?. En termes més generals i compatibles amb la resta de disciplines científiques, ¿de quantes unitats d'anàlisi estem parlant? I les notícies no són pas massa bones, ja que el nombre concret varia en funció de l'observador. Per a un xinès, Taiwan no és cap estat. Per a un iranià, Israel tampoc no ho és. No obstant aquestes discussions polítiques, el cert és que sol haver-hi cert consens acadèmic, manifestat a la [base de dades d'unitats del sistema internacional del projecte *Correlates of War*](#).

Els criteris per incorporar estats, però, són força restrictius: - Que sigui membre de les Nacions Unides. - Població d'almenys 500.000 habitants. - Reconeixement d'almenys dos potències majors. - Que el seu estatus estigui mantingut almenys durant un mes.

Per incorporar les dades a R fem servir la funció `read.table()`, que és la que permet que arxius de dades separats per comes (amb l'extensió `.csv`) puguin ser llegits de manera sistemàtica pel programa, de manera que entre les comes (o les tabulacions, o els punts i coma, com definim nosaltres) hi hagi les variables de manera igualment ordenada a cada línia. La funció l'assignem a una cadena de text (en aquest cas `states`) que té el nom que li vulguem donar (tot i que no es permet que tingui espais) i amb aquest nom ens queda un objecte `data.frame` que després explorem amb la funció `str()`, per saber-ne la seva estructura.

```
url <- "http://www.correlatesofwar.org/data-sets/state-system-membership/system2016/at_download/file"
states <- read.table(url, header = TRUE, sep = ",", fill = TRUE)
str(states)

## 'data.frame': 15951 obs. of 4 variables:
## $ stateabb: Factor w/ 217 levels "AAB","AFG","ALB",...: 205 202 141 65 186 177 153 16 72 13 ...
## $ ccode : int 2 200 210 220 225 230 235 245 255 267 ...
## $ year : int 1816 1816 1816 1816 1816 1816 1816 1816 1816 1816 ...
## $ version : int 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
```

Tenim, doncs, 15,951 observacions i 4 variables (l'abreujatura `stateabb`, el codi intern de la base de dades del *Correlates of War* `ccode`, l'any `year` i la versió de la base de dades `version`). Cada observació correspon a un estat i un any.

Quants estats hi havia a l'inici del període cobert per la base de dades (1816) i quants n'hi ha ara? Per calcular-ho enviem a través d'una "canonada" (`%>%`) l'objecte `states` a la funció `'group_by()'` que agrupa les observacions per any. Després fem un resum per any del nombre (`n()`)

d'observacions i li assignem el nom N a la nova variable creada, i finalment ho filtrem pel valor mínim o (l) màxim de l'any.

```
states %>%
  group_by(year) %>%
  summarize(N = n()) %>%
  filter(year == min(year) | year == max(year))

## # A tibble: 2 x 2
##   year      N
##   <int> <int>
## 1  1816     23
## 2  2016    195
```

A nivell més pràctic, sempre es pot recórrer a la llista de la normativa [ISO-3166-1](#) que estandarditza els noms dels estats. Sembla una qüestió menor, però en realitat dona molts mals de cap tenir noms diferents per als països a les diferents bases de dades, de manera que ajuntar dades de diferents llocs pot portar a fer desaparèixer fàcilment el Regne Unit de la Gran Bretanya i Irlanda del Nord (que seria el nom "normalitzat" / "estàndard" / oficial), ja que en alguns llocs és "Regne Unit", en altres "Gran Bretanya", en altres simplement l'abreujatura "UK". Per sort, en R hi ha un paquet (`countrycode`) que permet passar entre diferents bases de dades minimitzant aquests problemes.

```
library(countrycode)
countrycode("United Kingdom", origin = "country.name", destination = "cowc")

## [1] "UKG"

countrycode("United Kingdom", origin = "country.name", destination = "cown")

## [1] 200

countrycode("United Kingdom", origin = "country.name", destination = "eurostat")

## [1] "UK"

countrycode("United Kingdom", origin = "country.name", destination = "wb")

## [1] "GBR"
```

1.6 Glossari de conceptes

La resta del capítol presenta un glossari de conceptes útils per familiaritzar-te amb la gestió de dades en general i amb la resta dels materials en particular.

1.6.1 Matriu de dades / Dataset

Un full de càlcul és un programa amb el qual estàs familiaritzat. Un arxiu de full de càlcul (el programa Calc de Libre/OpenOffice, o l'Excel de Microsoft Office) conté diferents “fulls”. Has d'imaginar-te una matriu de dades a un full de dades, encara que no és ben bé el mateix.

Una matriu de dades és, tècnicament, un recull de vectors de la mateixa mida encara que potencialment de tipus diferent. Imagina't el vector que conté els noms dels estats que hi ha al món, ordenat verticalment formant una columna. En aquest cas es tractaria d'un vector no numèric, perquè no conté números, sinó caràcters. Al costat, però, enganxat en una segona columna imagina't el valor de la població de cadascun dels estats corresponents, de manera que per al primer estat (primera observació del primer vector) li correspon el valor de la població (primera observació del segon vector).

Una matriu de dades es pot entendre també com una col·lecció de característiques (*variables*, a les columnes) per a un seguit de unitats d'anàlisi (*observacions*, a les files).

Evidentment, com que només hem parlat d'observacions/files en variables/columnes, una matriu de dades només té dues dimensions (files i columnes).

En aquest material les matrius de dades es refereixen a col·leccions de característiques (variables) de diferents unitats d'anàlisi: els països als Capítols 2 i 3, les relacions entre ells al Capítol 4 i les organitzacions internacionals al Capítol 5.

Un exemple simple de matriu de dades és el següent:

Country	Population
Andorra	76965
Canada	36708083
Senegal	15850567

Dos vectors (“Country” i “Population”) de mida igual (3), però cadascun d'un tipus diferent. El primer és no numèric, i el segon numèric. El primer fa referència a les unitats d'anàlisi (els estats), mentre que el segon és el que conté *dades* en sentit estricte. El primer només *marca* les observacions.

Un segon exemple conté una altra variable (Year), que també és numèrica (un any no deixa de ser quelcom que es pot tractar com un número), però en aquest cas no és una *dada*, sinó un *marcador* de la unitat d'observació.

Country	Year	Population
Andorra	2015	78014
Andorra	2016	77281
Andorra	2017	76965
Canada	2015	35832513
Canada	2016	36264604
Canada	2017	36708083
Senegal	2015	14976994
Senegal	2016	15411614
Senegal	2017	15850567

1.6.2 Dades mancants / Missing data

No sempre per a cada unitat d'anàlisi tenim una resposta en alguna de les variables. Especialment a les ciències socials el problema de no poder observar una característica en una unitat d'anàlisi és quelcom greu.

Per exemple, un físic teòric pot dissenyar un experiment que, per *alguna raó*, no surti bé, i per això en alguna de les seves observacions tindrà una dada mancants. El "per alguna raó", però, en ciències naturals en general no està relacionat amb cap característica de la unitat d'anàlisi. A les ciències socials, però, això no és generalment el cas, i comporta una sèrie de problemes que, tot i que estan fora de l'abast d'aquest material, convé si més no, esmentar.

Per exemple, la mateixa dada de població de l'exemple anterior. Pots pensar en algun cas en què *per alguna raó* no es pugui disposar de la població d'un estat? *Aquesta raó*, és aleatòria a la unitat d'anàlisi, o potser té alguna cosa a veure amb altres característiques de la mateixa? Efectivament, si a un estat li manca la dada de població el més probable és que sigui perquè es tracta d'un estat amb tan poca infraestructura que no és ni capaç de fer un cens en condicions per quelcom tan bàsic com comptar els ciutadans que en formen part. I, com pots imaginar, aquesta és una característica que també estarà associada a moltes altres dades mancants en altres variables. Per això, no tenir dades, en relacions internacionals i política comparada, normalment és un indicador *d'altres raons*, que convé, si més no, tenir al cap.

Les dades mancants es solen representar amb "NA" (*not available*). Un exemple d'una matriu de dades amb dades mancants és la següent:

Country	Year	Population
Andorra	2015	78014
Andorra	2016	77281
Andorra	2017	NA
Canada	2015	35832513
Canada	2016	36264604
Canada	2017	36708083
Senegal	2015	14976994
Senegal	2016	15411614
Senegal	2017	15850567

1.6.3 Unitats d'anàlisi

Una unitat d'anàlisi és un objecte sobre el qual en volem recollir unes dades i analitzar-lo. Com ja hem esmentat abans, en la major part d'aquest material aquests objectes seran els estats, però també les organitzacions internacionals, o els tractats, poden ser objecte d'anàlisi.

1.6.4 Observacions

Un estat pot ser observat en diferents moments del temps, de manera que la mateixa unitat d'anàlisi pot estar dividida en diferents observacions. Per tant, no confonguis mai una unitat d'anàlisi amb una observació. Una observació és tan sols una entrada (una fila) en una matriu de dades.

1.6.5 Variables

Cadascuna de les columnes que formen part d'una matriu de dades és una variable. N'hi ha de dos tipus, com ja hem comentat. Hi ha variables que contenen *marques*, que identifiquen les observacions. I hi ha variables que contenen *dades* en sí mateixes.

Les variables també es poden dividir en el tipus de format. Poden ser numèriques o no numèriques. En ciències naturals les numèriques solen ser més presents, mentre que a les ciències socials és molt habitual tenir variables no numèriques que ens informen de qualitats que tenen les nostres observacions.

1.6.6 Format de dades

Fixa't en la següent matriu de dades:

Country	Year	Population
Andorra	2015	78014
Andorra	2016	77281
Andorra	2017	76965
Canada	2015	35832513
Canada	2016	36264604
Canada	2017	36708083
Senegal	2015	14976994
Senegal	2016	15411614
Senegal	2017	15850567

Està perfectament organitzada per recollir altres variables i enganxar-les en més columnes. Per exemple, podríem enganxar el PIB, i d'aquesta manera molt fàcilment podríem crear una columna més que fos el PIB per càpita tot dividint el PIB per la població. Per tant, es tracta d'un format *convenient* per a comparar variables entre elles.

En canvi, però, es tracta d'un format que no és apte per gestionar les dades si el que volem és comparar en el temps les diferents observacions. Es fa difícil saber quina és la taxa de creixement de la població d'un any a l'altre. No podem calcular-la comparant observacions entre una fila i la fila anterior, ja que a vegades les files anteriors es corresponen a una altra unitat d'anàlisi (un altre estat). En aquest cas, ens convé *canviar el format de les dades* per un format que anomenem *ample* (en contrast amb el *llarg* que tenim fins ara). Així quedarien les dades en un format ample, preparat per a comparar observacions a través dels anys:

Country	2015	2016	2017
Andorra	78014	77281	76965
Canada	35832513	36264604	36708083
Senegal	14976994	15411614	15850567

Tenir la capacitat de passar d'un format a un altre ràpidament permet gestionar les dades de manera molt eficient i, d'aquesta manera, poder centrar-se en els aspectes de les dades que volem comparar (en aquest cas, el temps, o bé les variables entre elles).

2

Context socioeconòmic

Lectures:

- Milanovic (2010), capítol 2
- Milanovic (2010), capítol 2

2.1 World Development Indicators

La base de dades del Banc Mundial amb els “World Development Indicators” (WDI) és la font de dades sobre característiques socioeconòmiques més rellevant, tant en termes temporals (bona part de les sèries temporals comencen a mitjans del segle XX), d’unitats cobertes (nombre de països) i de variables. En realitat el Banc Mundial no és qui genera les dades de manera primària, sinó que són cadascun dels països els que s’encarreguen de proporcionar-les i el Banc Mundial les ordena, sistematitza i presenta de manera comprensiva.

El valor fonamental de la base de dades dels WDI, més enllà de l’abast que acabem de comentar, és que són indicadors oberts a l’accés i la consulta de qualsevol. D’aquesta manera, el primer lloc al qual has de recórrer a l’hora de fer un treball sobre qualsevol país, o per comparar països i regions, o, en definitiva, per encarar qualsevol investigació, és aquest. Tenir xifres al cap i poder-les contextualitzar en la geopolítica mundial és clau per entendre què passa, com passa i per què passa quelcom en relacions internacionals.

Hi ha dues qüestions a destacar per posar en perspectiva l’ús dels WDI com a font de dades. La primera és que les dades més noves sempre estan subjectes a canvis. És a dir, els instituts d’estadística estatals revisen la comptabilitat nacional regularment i no és estrany veure que hi ha canvis entre la primera xifra que es reporta un any i el que queda marcat com a dada històrica cinc anys després. Si bé no seran mai canvis exagerats, no és estrany trobar que la riquesa d’un país s’ha mogut en fins a un 5 per cent amunt o avall amb les successives revisions. Per tant, cal posar sempre les dades en perspectiva quan es tracta de dades recents.

La segona qüestió és que la qualitat i la disponibilitat de les dades estan clarament relacionades amb la capacitat organitzativa de l'estat. Dit d'una altra manera: un estat que funciona bé, amb capacitat per registrar transaccions, amb personal qualificat per comptabilitzar el que passa dins de les seves fronteres, per revisar-ho i per publicar-ho, ha de tenir certes capacitats organitzatives. I no és estrany que siguin els estats més rics els que puguin fer això. Si un estat té una baixa capacitat de funcionar, és més probable que les dades que genera siguin o de més baixa qualitat o, directament, que sigui incapaç de generar-les.

Si bé la disponibilitat de les dades en funció de la capacitat organitzativa i fortalesa de l'administració estatal és en certa manera lògica, no és normalment tant lògic considerar que la manca de dades pot alterar la validesa científica de la recerca. Quan analitzem dades esperem no trobar-hi forats, no trobar-hi dades que manquen, dades anomenades *perdues*¹. Si tenim la dissort que no n'hi ha en alguns casos, el que esperem almenys, des del punt de vista estadístic i matemàtic, és que el patró de dades perdudes sigui aleatori; no tingui a veure amb el que s'està analitzant.

A tall d'exemple: vols recollir dades sobre la riquesa d'un país. En aquest cas observaràs la magnitud de la seva economia tal i com la mesura el Producte Interior Brut. Ara bé, resulta que dels deu països que analitzes n'hi ha dos que no tenen dades. En sí mateix això no hauria de ser cap problema si els dos països que no reporten dades no responen a cap patró sistemàtic. Però resulta que en general en relacions internacionals aquest patró és esbiaixat i sistemàtic. És més probable que un país amb un nivell baix de riquesa no reporti el seu PIB que no pas que ho faci un país amb un nivell alt. I és així perquè en bona part el PIB connota també una certa capacitat d'un estat. I això és un problema a l'hora de treure conclusions perquè vol dir que les podries treure basant-te en un biaix original a les dades, pel fet que països de baixa riquesa són més propicis a no generar-ne. És més fàcil que els Estats Units donin un valor de PIB que no pas que ho faci el Sudan, i això està relacionat amb el fet que el PIB és més elevat en el primer cas que no pas en el segon. Per tant, obviar països que no tenen dades sense haver-se plantejat per què no les tenen no és una bona idea.

Cal anar amb compte, doncs, amb la disponibilitat de les dades socioeconòmiques, ja que tenir o no tenir una dada pot donar-nos molta informació sobre la dada mateixa, i en general aquesta informació sol perjudicar estats amb problemes de funcionament, que en general són aquells pels quals volem informació, precisament.

Les dades dels indicadors del Banc Mundial inclouen aspectes com agricultura, educació, ajut al desenvolupament, canvi climàtic, economia i creixement, mineria i recursos naturals, energia, medi ambient, deute extern, sector financer, gènere, salut, infraestructura, pobresa, sectors privat i públic, desenvolupament social, ciència i tecnologia, treball, protecció social, comerç, i desenvolupament urbà.

¹ Traducció del concepte "missing values" en anglès.

És molt simple obtenir les dades dels WDI per analitzar, a través del paquet WDI del programa R.

```
library(WDI)
```

El primer pas és buscar la variable que t'interessi fent una cerca a la base de dades amb la funció `WDIsearch()` i posar com a primer argument de la funció la paraula clau que vols buscar (en anglès). En aquest cas, buscarem dades que tinguin a veure amb Internet:

```
WDIsearch("Internet")
```

```
##      indicator
## [1,] "WP14910.9"
## [2,] "WP14910.8"
## [3,] "WP14910.7"
## [4,] "WP14910.6"
## [5,] "WP14910.5"
## [6,] "WP14910.4"
## [7,] "WP14910.3"
## [8,] "WP14910.2"
## [9,] "WP14910.10"
## [10,] "WP14910.1"
## [11,] "SF.CMN.INET.USER.10K"
## [12,] "IT.BBD.USEC.CD"
## [13,] "IT.NET.USER.ZS"
## [14,] "IT.NET.USER.P3"
## [15,] "IT.NET.USER.P2"
## [16,] "IT.NET.USER"
## [17,] "IT.NET.USEC.ZS"
## [18,] "IT.NET.USEC.CD"
## [19,] "IT.NET.TELC.CD"
## [20,] "IT.NET.SUB.CN"
## [21,] "IT.NET.SUB.CD"
## [22,] "IT.NET.SECR.P6"
## [23,] "IT.NET.SECR"
## [24,] "IT.NET.ISPC.CD"
## [25,] "IT.NET.HOST.P4"
## [26,] "IT.NET.EDUC.ZS"
## [27,] "IT.NET.CONN.CN"
## [28,] "IT.NET.CONN.CD"
## [29,] "IT.NET.BNDW.PC"
## [30,] "IT.NET.BNDW"
## [31,] "gwp1.6"
## [32,] "gwp1.5"
## [33,] "gwp1.4"
```

```

## [34,] "gwp1.3"
## [35,] "gwp1.2"
## [36,] "gwp1.1"
## [37,] "gwp1"
## [38,] "2.0.cov.Int"
## [39,] "2.0.hoi.Int"
## [40,] "gf5.6"
## [41,] "gf5.5"
## [42,] "gf5.4"
## [43,] "gf5.3"
## [44,] "gf5.2"
## [45,] "gf5.1"
## [46,] "gf5"
##      name
## [1,] "Used the Internet to pay bills or buy things, income, richest 60% (% ages 15+) [w2]"
## [2,] "Used the Internet to pay bills or buy things, income, poorest 40% (% ages 15+) [w2]"
## [3,] "Used the Internet to pay bills or buy things, secondary education or more (% ages 15+) [w2]"
## [4,] "Used the Internet to pay bills or buy things, primary education or less (% ages 15+) [w2]"
## [5,] "Used the Internet to pay bills or buy things, older adults (% ages 25+) [w2]"
## [6,] "Used the Internet to pay bills or buy things, young adults (% ages 15-24) [w2]"
## [7,] "Used the Internet to pay bills or buy things, female (% age 15+) [w2]"
## [8,] "Used the Internet to pay bills or buy things, male (% age 15+) [w2]"
## [9,] "Used the Internet to pay bills or buy things, rural (% age 15+) [w2]"
## [10,] "Used the Internet to pay bills or buy things (% age 15+) [w2]"
## [11,] "Internet users (per 10,000 people)"
## [12,] "Fixed broadband Internet access tariff (US$ per month)"
## [13,] "Individuals using the Internet (% of population)"
## [14,] "Internet users (per 1,000 people)"
## [15,] "Internet users (per 100 people)"
## [16,] "Internet users"
## [17,] "Internet total monthly price (% of monthly GNI per capita)"
## [18,] "Price basket for Internet (US$ per month)"
## [19,] "Internet telephone access charges ($ per 30 off-peak hours)"
## [20,] "Fixed broadband Internet monthly subscription (current LCU)"
## [21,] "Fixed broadband Internet monthly subscription (current US$)"
## [22,] "Secure Internet servers (per 1 million people)"
## [23,] "Secure Internet servers"
## [24,] "Internet service provider access charges ($ per 30 off-peak hours)"
## [25,] "Internet hosts (per 10,000 people)"
## [26,] "Schools connected to the Internet (%)"
## [27,] "Fixed broadband Internet connection charge (current LCU)"
## [28,] "Fixed broadband Internet connection charge (current US$)"
## [29,] "International Internet bandwidth (bits per person)"
## [30,] "International Internet bandwidth (Mbps)"

```

```
## [31,] "Access to a mobile phone or internet at home, older adults (% age 35+)"
## [32,] "Access to a mobile phone or internet at home, young adults (% ages 15-34)"
## [33,] "Access to a mobile phone or internet at home, income, richest 60% (% age 15+)"
## [34,] "Access to a mobile phone or internet at home, income, poorest 40% (% age 15+)"
## [35,] "Access to a mobile phone or internet at home, male (% age 15+)"
## [36,] "Access to a mobile phone or internet at home, female (% age 15+)"
## [37,] "Access to a mobile phone or internet at home (% age 15+)"
## [38,] "Coverage: Internet "
## [39,] "H0I: Internet "
## [40,] "Made payment using the internet, older adults (% age 35+)"
## [41,] "Made payment using the internet, young adults (% ages 15-34)"
## [42,] "Made payment using the internet, income, richest 60% (% age 15+)"
## [43,] "Made payment using the internet, income, poorest 40% (% age 15+)"
## [44,] "Made payment using the internet, male (% age 15+)"
## [45,] "Made payment using the internet, female (% age 15+)"
## [46,] "Made payment using the internet (% age 15+)"
```

El resultat ens retorna dues columnes. La primera (indicator) és la paraula clau de l'indicador, i la segona (name) és el nom complet de l'indicador, amb un petit detall. Per saber més detalls dels indicadors vés a la pàgina web del Banc Mundial i hi trobaràs una descripció més detallada de la font originària de les dades.

Per tenir el percentatge d'usuaris d'Internet amb banda ample només has d'especificar l'indicador a la funció `WDI()` (en aquest cas, `IT.NET.BBND.P2`) i desar-ho en un objecte que creïs amb el nom que vulguis (li direm en aquest cas `int`)²:

```
int <- WDI(indicator = "IT.NET.BBND.P2")
```

Ara en memòria hi tens un objecte anomenat `int` amb el qual pots operar. Per veure la seva estructura fes servir la funció `str()`:

```
str(int)
## 'data.frame': 1848 obs. of 4 variables:
## $ iso2c : chr "1A" "1A" "1A" "1A" ...
## $ country : chr "Arab World" "Arab World" "Arab World" "Arab World" ...
## $ IT.NET.BBND.P2: num 2.67 2.091 1.954 1.304 0.893 ...
## $ year : num 2011 2010 2009 2008 2007 ...
```

Es tracta d'una matriu de dades (*data frame*) amb 1848 observacions i 4 variables, la primera de les quals és el codi ISO del país en dos dígitos, la segona el nom del país (o de la regió, en cas que la variable també tingui variants regionals), la tercera el valor de la variable que busques i la quarta l'any d'observació.

Per tenir una idea dels valors que conté la matriu de dades no cal observar els 1848 casos, sinó que amb els primers ja te'n pots fer una idea. En aquest cas, la funció `head()` mostra els primers valors.

² Els arguments de la funció `WDI` també permeten especificar el país (`country =`), l'any d'inici (`start =`) i l'any de finalització (`end =`)

```
head(int)

##   iso2c   country IT.NET.BBND.P2 year
## 1    1A Arab World      2.66989 2011
## 2    1A Arab World      2.09127 2010
## 3    1A Arab World      1.95441 2009
## 4    1A Arab World      1.30378 2008
## 5    1A Arab World      0.89296 2007
## 6    1A Arab World      0.50375 2006
```

Per realitzar una figura d'un estat concret has de filtrar el nom que vulguis a la variable `country` (o, si el saps, fer servir el seu codi ISO). En aquest cas, començant per l'objecte `int`, a través d'una *canonada* (representada per `%>%`) l'envies a la funció `filter()` a la qual li concretes que vols aquelles observacions el `country` de les quals sigui estrictament igual a "Germany".

```
int.de <- int %>%
  filter(country == "Germany")
```

Ara l'objecte `int.de` el tens desat a la memòria, i pots fer-lo servir per dibuixar una figura amb l'evolució temporal, a través d'especificar els elements estètics (`aes`) `x` i `y` corresponents a l'eix horitzontal i vertical, respectivament. Finalment, l'element geomètric que vols que aparegui a la figura és una línia, i l'afegeixes a l'objecte definit per `ggplot()` amb la funció `geom_line()`:

```
ggplot(int.de, aes(x = year, y = IT.NET.BBND.P2)) +
  geom_line()
```

Finalment, cal que la figura realitzada tingui uns eixos adequats tant en el seu rang com en els noms, de manera que has d'afegir altres elements. En aquest cas, `expand_limits()` et permet dir-li que com que l'eix vertical es correspon a un percentatge, necessites que el valor mínim sigui el zero, i el valor màxim el 100. Només així pots tenir una perspectiva real de fins a quin punt canvia el valor anualment. Per especificar un text millor per als eixos necessites les funcions `xlab()` i `ylab()`:

```
ggplot(int.de, aes(x = year, y = IT.NET.BBND.P2)) +
  geom_line() +
  expand_limits(y = c(0, 100)) +
  xlab("Any") + ylab("% usuaris d'Internet\namb banda ampla")
```

Hi ha també un altre paquet en R que permet interactuar de manera similar a com ho fa el WDI, però amb les dades de països europeus provinents d'*Eurostat*, la oficina d'informació estadística europea. Es tracta del paquet `eurostat`.

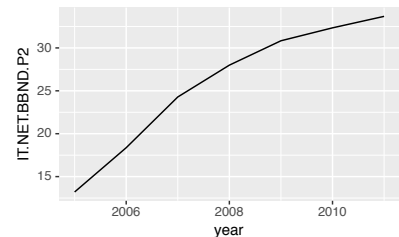


Figure 2.1: Resultat inicial d'una figura sense interès per aspectes formals.

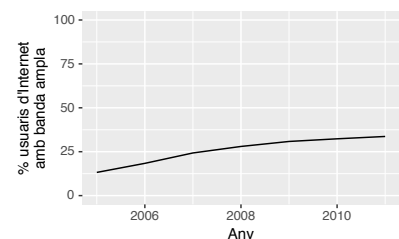


Figure 2.2: Evolució temporal del percentatge d'usuaris d'Internet amb banda ampla a Alemanya (2005-2011). Font: World Development Indicators.

2.2 Desigualtat

A banda de les mesures socioeconòmiques generals que podem trobar als *World Development Indicators*, dins mateix del Banc Mundial un dels seus economistes, Branko Milanovic, manté una [acurada base de dades sobre desigualtat](#), juntament amb una referència bibliogràfica imprescindible per entendre el procés de desenvolupament econòmic i social en els últims cent cinquanta anys, i de retruc les relacions internacionals. Es tracta de Milanovic (2010).

El primer capítol explica què entenem per desigualtat, fa un repàs a les teories per les quals tradicionalment hem entès que tenia un efecte sobre el comportament econòmic, i —especialment rellevant per a fonts d'informació—, presenta breument com es mesura la desigualtat.

El coeficient de Gini és la mesura estàndard de la *desigualtat* econòmica dins d'un grup social. El més habitual és considerar els estats com els grups socials de referència, però una de les virtuts del coeficient de Gini és que és aplicable a qualsevol grup social format per més de dos unitats (ja siguin persones o llars, vegeu el capítol). Ho calculem sumant la diferència d'ingressos (a vegades també es fa servir renda) entre totes les combinacions possibles d'unitats d'un grup social. En un grup social de 3 unitats (A, B i C), la operació és igual a $(A - B) + (A - C) + (B - C)$. A aquesta suma de diferències d'ingressos li dividim el nombre d'unitats (tres en l'exemple anterior) i la mitjana dels ingressos del grup. Aquesta operació aritmètica ens dona un valor que està en un rang teòric entre 0 i 1. Com que ja hem comentat que és una mesura de desigualtat, el zero es correspon a la menor desigualtat, o quan tothom en el grup té exactament el mateix nivell d'ingressos (fixa't que les diferències d'ingressos serien totes zero, i la seva suma també seria zero, de manera que al numerador només hi quedaria zero, que dividit per qualsevol cosa dona zero). A l'altre extrem, hi ha l'1, que es correspon a casos de grups amb extrema desigualtat, quan un únic individu té tots els ingressos i la resta no en tenen gens. Com s'indica a la lectura, el rang efectivament observable del coeficient de Gini, va de 0.25 a 0.60, o més habitualment expressat en una escala de zero a cent: de 25 a 60, que és el que coneixem com a índex de Gini.

El segon capítol es centra fonamentalment a explicar la desigualtat en termes comparats entre països i al llarg del temps. Milanovic explica quines estratègies es fan servir per comparar ingressos en diferents moments del temps, introduint el concepte de paritat de poder adquisitiu (PPP, per les sigles en anglès). També explica per què és millor que ens refiem de les dades d'ingressos obtingudes a partir de les enquestes, que no pas de les dades obtingudes a partir de registres burocràtics (declaracions d'impostos, bàsicament), tot i que les primeres tenen un recorregut temporal més breu (des de meitats del segle XX per a economies avançades, i des de finals del XX per a economies emergents, i encara molt problemàtiques per a economies

no desenvolupades).

El nivell de renda dels països presentat en aquest segon capítol no deixa de ser el que hem introduït a la secció anterior (el PIB), tot i que en no en valor absolut (que és una mesura de la *mida d'una economia*), sinó en valor relatiu, que esdevé llavors una mesura de *riquesa d'una societat*.

Per tant, a l'hora de voler obtenir dades de riquesa o benestar material, i sobretot vols fer comparacions en el temps, has de fer servir sèries temporals en del PIB per càpita en PPP³:

```
gdp.pc.ppp <- WDI(indicator = "NY.GDP.PCAP.PP.KD",
                 start = 1990, end = 2015)
str(gdp.pc.ppp)

## 'data.frame':   6864 obs. of  4 variables:
## $ iso2c          : chr  "1A" "1A" "1A" "1A" ...
## $ country        : chr  "Arab World" "Arab World" "Arab World" "Arab World" ...
## $ NY.GDP.PCAP.PP.KD: num  15343 15199 15174 15056 14445 ...
## $ year           : num  2015 2014 2013 2012 2011 ...
```

Per calcular com ha anat divergint el PIB per càpita en els estats pels quals hi ha dades fem servir la desviació estàndard, que és una mesura de diferència respecte a la mitjana en un grup d'observacions⁴. El primer que fem és agafar l'objecte `gdp.pc.ppp` que conté les quatre variables que hem obtingut abans (codi ISO en dos caràcters, nom del país, valor de la variable, i any) i enviar-lo a una canonada (`%>%`) per agrupar-lo per anys, de tal manera la última línia, que fa un resum de la variable que ens interessa, ens farà el resum anualment. En aquest cas, el resum és l'aplicació de la variable `sd()`, que calcula la desviació estàndard (*standard deviation*), i que obvia els valors perduts (`na.rm = TRUE`).

```
gdp.pc.de <- gdp.pc.ppp %>%
  group_by(year) %>%
  summarize(Desviació = sd(NY.GDP.PCAP.PP.KD, na.rm = TRUE))
```

Per generar la Figura hem de passar l'objecte acabat de crear a la funció `ggplot()` com anteriorment, i especificar els valors que aniran a cada eix, així com el tipus de figura que volem (línies, per marcar una evolució temporal), i fixar un valor de referència per a l'eix vertical.

```
ggplot(gdp.pc.de, aes(x = year, y = Desviació)) +
  geom_line() +
  expand_limits(y = 0) +
  xlab("Any")
```

Per obtenir les dades dels índexs de Gini, preparades per Branko Milanovic, hem de fer servir les seves matrius, que estan en format Stata (extensió `.dta`), però que a través del paquet `foreign` es poden llegir fàcilment en R:

³ Recorda que necessites la llibreria WDI

⁴ La desviació estàndard s'interpreta com el valor esperable que té la diferència de dues observacions agafades a l'atzar en un grup de valors. Com a promig, quan divergiran dues observacions?

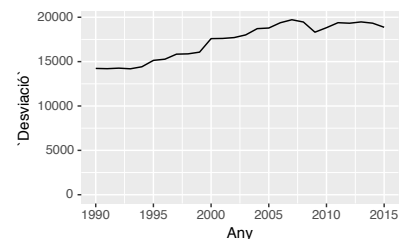


Figure 2.3: Evolució temporal de la desigualtat de renda entre països, mesurada per la desviació estàndard anual de la renda per càpita als països del món. Font: World Development Indicators. Fixa't que la crisi financera del 2008 en realitat va significar un equipament entre els països, un fre a la creixent desigualtat entre estats.

```
# previous read
library(foreign)
gini <- read.dta("https://wfs.gc.cuny.edu/njohnson/www/BrankoData/allginis_2016.dta")

library(foreign)
gini <- read.dta("https://www.gc.cuny.edu/getmedia/92e1b3ac-eae5-47
```

La distribució dels índexs de Gini (variable `Giniall`) l'obtenim amb un histograma, on hi marquem els rangs teòrics.

```
ggplot(gini, aes(x = Giniall)) +
  geom_histogram() +
  xlab("Coeficients de Gini") +
  expand_limits(x = c(0, 100))
```



Figure 2.4: Distribució dels índexs de Gini estandarditzats (1980-2015)

L'evolució de la mitjana dels índexs de Gini es pot obtenir agrupant les observacions anuals, i fent-ne la mitjana⁵.

```
gini.any <- gini %>%
  group_by(year) %>%
  summarize(Mitjana = mean(Giniall, na.rm = TRUE))

ggplot(gini.any, aes(x = year, y = Mitjana)) +
  geom_line() +
  xlab("Any") + ylab("Mitjana dels índexs de Gini") +
  expand_limits(y = c(0, 100))
```

⁵ Fixa't, però, que aquesta és la mitjana dels índexs de Gini de cada país, de manera que els països tenen el mateix pes, sense ponderar pel seu nivell de població. Per tant, no es pot llegir com una evolució de la desigualtat al món, sinó només com una evolució de la desigualtat entre Estats.

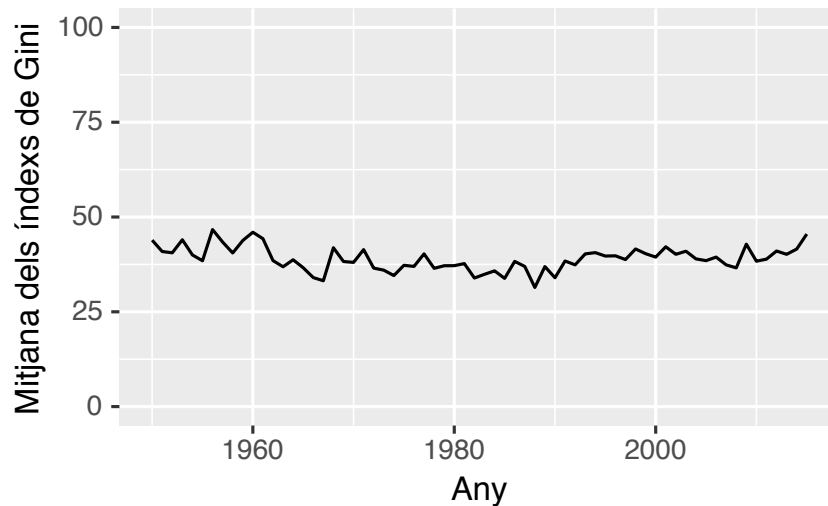


Figure 2.5: Evolució temporal de la mitjana d'índexs de Gini dels diferents països (1980-2015). Mesura la desigualtat entre Estats, però no la desigualtat global.

3

Sistemes polítics

Lectures:

- Straface and Page (2009)
- Przeworski, Ortega, and Rapoport (1997)
- Szmolka Vida (2010)

Aquest mòdul està centrat en com mesurar la democràcia, en tant que aspecte fonamental que defineix qualsevol sistema polític. Però també comentarem breument altres tipologies relacionades, com les limitacions al poder, la descentralització, la governança i característiques socials i del sistema de partits. Straface and Page (2009) presenta els problemes, limitacions i possibilitats de totes aquestes mesures de qualitat institucional de manera general.

3.1 Democràcia

És democràtic el règim polític d'un estat, diguem-ne, X? L'atorgament de credencials democràtiques als països és un afer que transcendeix els marges acadèmics tradicionals i s'endinsa de ple en el món de la diplomàcia i les relacions internacionals. Per això, des de l'acadèmia el tractament de les democràcies ha resultat sempre molt curós. Deixant de banda els debats teòrics sobre els components de les democràcies, quan és l'hora de classificar règims dels països segons un suposat component democràtic hi ha bàsicament dues aproximacions, que es corresponen amb les dues parts que es presentaran a continuació.

3.1.1 Przeworski, Ortega & Rapoport (1997): simple, eficient, clar

Przeworski, Ortega, and Rapoport (1997) presenta estratègia per respondre clarament a la pregunta de si és democràtic el règim polític d'un país concret. L'estratègia dels autors es basa en els següents principis.

- Simplificar al màxim la definició de democràcia. En lloc de definir una democràcia com l'agregació de diferents principis teòrics molt elaborats,

els autors opten per rebaixar-la a aspectes, si es vol, procedimentals.

- Buscar una operacionalització que sigui clara i eficient. Un cop una democràcia s'ha definit d'una manera molt simple, cal buscar una manera de mesurar-lo que no deixi lloc a l'ambigüitat, de manera que classificar els països sigui una feina quasi-mecànica, sense lloc per a matisos ni malentesos. Sense lloc perquè sigui qui sigui qui faci la classificació pugui arribar a les mateixes conclusions
- Clarificar quin és el tipus d'error que és preferible cometre. En un test d'embaràs, és millor que el resultat sigui positiu tot i que la dona no estigui embarassada o és millor que el resultat sigui negatiu i que la dona estigui embarassada? Què és preferible que passi? A l'hora de dissenyar un test com aquest, quin error és preferible cometre? Doncs en el cas de voler saber si un règim polític és una democràcia, el principi és el mateix: és millor per als autors classificar com a no democràtic un règim que no ho és que no pas fer que règims no democràtics siguin mal classificats com a democràtics. En realitat, doncs, els autors volen estar segurs que només els països classificats com a democràtics ho són de veritat.

D'aquesta manera s'aconsegueix una classificació difícilment controvertida, almenys pel que fa al procés de generació de la mesura. Es pot estar més o menys d'acord en si la definició de democràcia ho comprèn tot, però no es pot estar en desacord en com s'ha arribat a dir que el país X és una democràcia.

Una altre avantatge d'aquesta aproximació és que permet comparar règims contemporanis amb règims de fa dos-cents anys. Si és tan simple obtenir dades i arribar a una conclusió sobre el règim d'X el 2015, també ho és per aplicar els mateixos procediments sobre el règim d'X el 1820. El component temporal i històric es pot mantenir.

D'aquesta manera, aquestes dades ens permeten als politòlegs poder fer estudis comparatius sobre els efectes de les democràcies en altres aspectes (creixement econòmic, sobretot, però també altres resultats) fent servir mecanismes causals, ja que per tal d'establir mecanismes causals tenir dades d'evolució temporal és clau.

Una altra versió més moderna dels mateixos principis es pot trobar a Boix, Miller, and Rosato (2013), i les dades actualitzades es poden aconseguir de la següent manera:

```
d <- read.table("https://sites.google.com/site/mkmtwo/democracy-v2.0.csv?attredirects=0",
  header = TRUE, sep = ",", quote = "\"")
```

L'objecte d en aquest cas és una matriu amb el següent nombre de files i columnes

```
dim(d)
```

```
## [1] 17632 10
```

I amb la següent estructura:

```
str(d)

## 'data.frame': 17632 obs. of 10 variables:
## $ country : Factor w/ 219 levels "AFGHANISTAN",...: 204 204 204 204 204 204 204 204 204 204
## $ ccode : int 2 2 2 2 2 2 2 2 2 2 ...
## $ abbreviation : Factor w/ 218 levels "ADO","AFG","ALB",...: 202 202 202 202 202 202 202 202 202
## $ year : int 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 ...
## $ democracy : int 1 1 1 1 1 1 1 1 1 1 ...
## $ sovereign : int 1 1 1 1 1 1 1 1 1 1 ...
## $ democracy_trans : int 0 0 0 0 0 0 0 0 0 0 ...
## $ democracy_breakdowns : int 0 0 0 0 0 0 0 0 0 0 ...
## $ democracy_duration : int 1 2 3 4 5 6 7 8 9 10 ...
## $ democracy_omitteddata: int 1 1 1 1 1 1 1 1 1 1 ...
```

Aquesta llista ens mostra tant el nom de les variables de la matriu de dades com el seu tipus (si contenen números -num-, valors repetits -factor- o text -char-).

Les següents línies generen valors amb el nombre de països (nC, per “number of countries”, nY per “number of years”, i n0 per “number of observations”) que necessitem a continuació. La manera de fer-ho és referint-nos a una variable de l'objecte (d\$country), i aplicant una funció que torna els possibles valors que té la variable (unique()) i finalment aplicant una funció que es limita a calcular quina és la llargada (length()) del vector de possibles valors.

```
nC <- length(unique(d$country))
```

```
nC
```

```
## [1] 219
```

```
nY <- length(unique(d$year))
```

```
nY
```

```
## [1] 211
```

```
n0 <- length(d$country)
```

```
n0
```

```
## [1] 17632
```

Per calcular la proporció de democràcies en tots els països considerats des del 1800 al 2007 simplement cal obtenir els resultats de la variable democracy mitjançant una taula i després dividir-ho pel total d'observacions:

```
table(d$democracy)
```

```
##
##      0      1
## 11566 5422

table(d$democracy) / n0

##
##      0      1
## 0.65597 0.30751
```

Per saber com ha evolucionat temporalment la proporció de democràcies primer cal agrupar les dades per la variable temporal (`year`) i després obtenir el nombre de casos que són efectivament una democràcia (`length(which(democracy == 1))`) i dividir-ho pel nombre total de règims (`length(democracy)`). Com que les dades ja han estat agrupades per any, totes aquestes operacions es fan prenent les dades anuals com a referència.

```
pd <- d %>%
  group_by(year) %>%
  summarize(pD = length(which(democracy == 1)) / length(democracy))
```

Amb el nou objecte acabat de crear (`pd`, per “proportion of democracies”) fem una figura que mostri les línies (`geom_line()`) de l’evolució temporal en l’eix horitzontal (`x = year`) i la proporció de democràcies calculada anteriorment en l’eix vertical (`y = pD`).

```
ggplot(pd, aes(x = year, y = pD)) +
  geom_line()
```

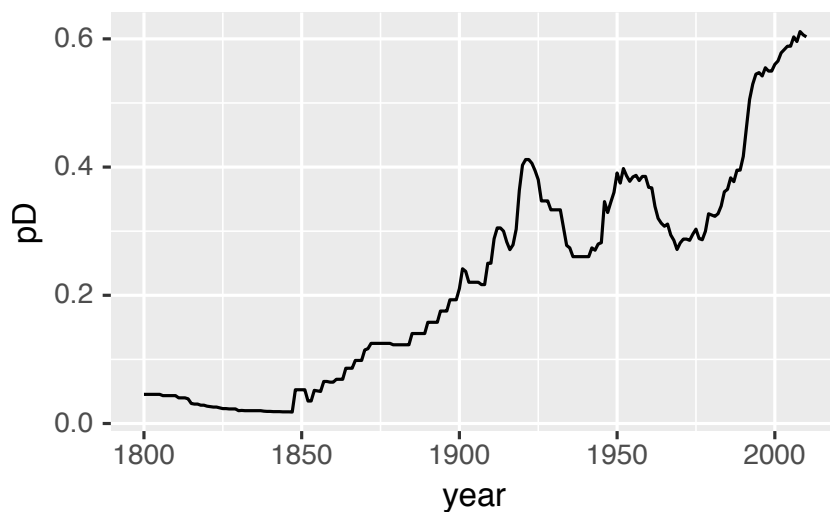


Figure 3.1: Evolució temporal de la proporció de democràcies al món.

També es poden agrupar les observacions de la matriu de dades per països, i calcular per a cada país quina és la proporció de temps que el seu règim ha estat democràtic:

```
pyd <- d %>%
  group_by(country) %>%
  summarize(pyD = length(which(democracy == 1)) / length(democracy))
```

Finalment tenim un objecte que conté, per a cada país, la proporció de temps en què el seu règim ha estat democràtic. Podem obtenir les cinc primeres observacions seleccionant-les amb els claudàtors¹:

```
pyd[1:5,]

## # A tibble: 5 x 2
##   country      pyD
##   <fct>      <dbl>
## 1 AFGHANISTAN  0
## 2 ALBANIA      0.186
## 3 ALGERIA      0
## 4 ANDORRA      1
## 5 ANGOLA       0

ggplot(pyd, aes(x = pyD)) +
  geom_histogram()
```

3.1.2 Szmolka (2010): Refinat, complex

En contrast amb una aproximació mecànica, simple, clara i eficient com la que acabem de presentar, hi ha les aproximacions basades en principis ben diferents. Szmolka Vida (2010) presenta una detallada descripció dels procediments per redimensionar la democràcia no com un concepte híbrid, sinó com un concepte gradual, amb diferents nivells. D'aquesta manera l'autora caracteritza i classifica els règims polítics aclarint els supòsits de partida i creant dos pols (democràcies plenes en contrast als autoritarismes tancats) entre els quals hi situa els règims híbrids.

3.1.3 Polity IV: l'estàndard

El Polity IV és, segurament, "La mesura" de democràcia en relació a la qual la resta es comparen. Un dels avantatges és que a banda de les dades crues també hi ha accessibles per a qualsevol qui ho desitgi tant els manuals metodològics com els informes de cadascun dels països. Obtenir les dades en R es pot fer fàcilment a través d'un paquet específic:

```
library(readxl)
tmp <- tempfile()
```

¹ [1:5,]: abans de la coma hi ha les files que desitgem obtenir (de la 1 a la 5) i després de la coma les columnes que volem obtenir (totes), en aquest cas per omissió, ja que en no explicitar-ne cap assumim que les volem totes.

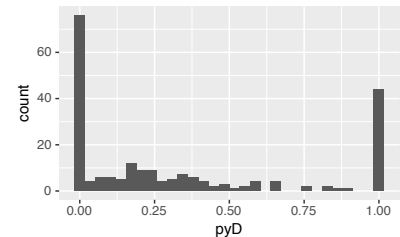


Figure 3.2: Distribució de la proporció de temps que els països han estat en règim democràtic.

```

download.file("http://www.systemicpeace.org/inscr/p4v2017.xls", tmp)
polity <- read_excel(tmp)
str(polity)

## Classes 'tbl_df', 'tbl' and 'data.frame': 17395 obs. of 36 variables:
## $ cyear : num 21800 21801 21802 21803 21804 ...
## $ ccode : num 2 2 2 2 2 2 2 2 2 ...
## $ scode : chr "USA" "USA" "USA" "USA" ...
## $ country : chr "United States" "United States" "United States" "United States" ...
## $ year : num 1800 1801 1802 1803 1804 ...
## $ flag : num 0 0 0 0 0 0 0 0 0 ...
## $ fragment: num NA NA NA NA NA NA NA NA NA ...
## $ democ : num 7 7 7 7 7 7 7 7 9 ...
## $ autoc : num 3 3 3 3 3 3 3 3 0 ...
## $ polity : num 4 4 4 4 4 4 4 4 9 ...
## $ polity2 : num 4 4 4 4 4 4 4 4 9 ...
## $ durable : num NA NA NA NA NA NA NA NA NA 0 ...
## $ xrreg : num 3 3 3 3 3 3 3 3 3 ...
## $ xrcomp : num 3 3 3 3 3 3 3 3 3 ...
## $ xropen : num 4 4 4 4 4 4 4 4 4 ...
## $ xconst : num 7 7 7 7 7 7 7 7 7 ...
## $ parreg : num 4 4 4 4 4 4 4 4 2 ...
## $ parcomp : num 2 2 2 2 2 2 2 2 4 ...
## $ exrec : num 8 8 8 8 8 8 8 8 8 ...
## $ exconst : num 7 7 7 7 7 7 7 7 7 ...
## $ polcomp : num 2 2 2 2 2 2 2 2 9 ...
## $ prior : num NA NA NA NA NA NA NA NA NA 4 ...
## $ emonth : num NA NA NA NA NA NA NA NA NA 3 ...
## $ eday : num NA NA NA NA NA NA NA NA NA 4 ...
## $ eyear : num NA NA NA NA NA ...
## $ eprec : num NA NA NA NA NA NA NA NA NA 1 ...
## $ interim : num NA NA NA NA NA NA NA NA NA NA ...
## $ bmonth : num 1 NA NA NA NA NA NA NA NA 3 ...
## $ bday : num 1 NA NA NA NA NA NA NA NA 5 ...
## $ byear : num 1800 NA NA NA NA ...
## $ bprec : num 1 NA NA NA NA NA NA NA NA 1 ...
## $ post : num 4 NA NA NA NA NA NA NA NA 9 ...
## $ change : num 88 NA NA NA NA NA NA NA NA 5 ...
## $ d4 : num 1 NA NA NA NA NA NA NA NA 1 ...
## $ sf : num NA NA NA NA NA NA NA NA NA NA ...
## $ regtrans: num NA NA NA NA NA NA NA NA NA 2 ...

```

La variable `polity2` és la que resumeix el règim polític del país, i que es calcula de manera molt simple a través de sumes i mitjanes combinades de les dades observades (que és precisament el que critiquen Treier and

Jackman (2008) amb la seva proposta). El rang de la variable va de -10 fins a 10, essent -10 un règim autoritari i 10 un règim democràtic. Normalment se situa el punt de tall de les democràcies en el 6, a partir del qual un règim es considera democràtic. Cal destacar també que es tracta d'una base de dades molt completa en termes temporals (des del 1800) com en termes d'abast de països.

La següent figura presenta l'evolució temporal de la mitjana dels països. Convé eixamplar els límits de l'eix vertical per incloure tots els valors entre -10 i 10, ja que aquesta és l'escala original de la mesura, i d'aquesta manera s'obté una evolució temporal en el context adequat del rang de valors que pot prendre.

```
pmy <- polity %>%
  group_by(year) %>%
  summarize(Mitjana = mean(polity2, na.rm = TRUE))
ggplot(pmy, aes(x = year, y = Mitjana)) +
  geom_line() +
  expand_limits(y = c(-10, 10))
```

Tenir accés fàcil a aquestes dades et pot resultar molt útil a l'hora de fer treballs o analitzar l'evolució temporal i històrica del país o països que vulguis considerar. En aquest cas, per exemple, les dades es filtren per 3 països i es mostren les seves trajectòries.

```
p3c <- polity %>%
  filter(country %in% c("Argentina", "Mexico", "Spain"))
ggplot(p3c, aes(x = year, y = polity2, color = country)) +
  geom_line()
```

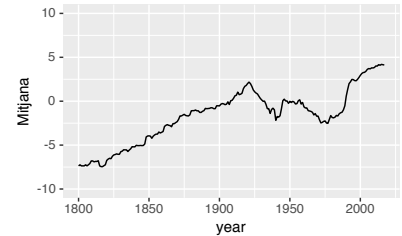
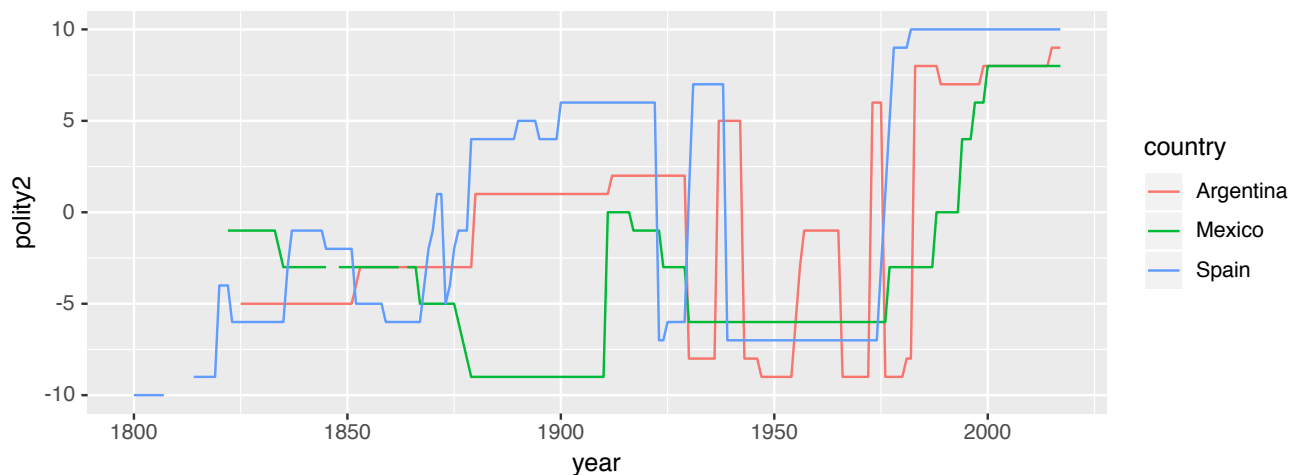


Figure 3.3: Evolució temporal de la mitjana i la mediana de l'indicador agregat polity2. Per calcular la mitjana i la mediana cal no fer cas dels valors perduts, i per això es fa servir `na.rm = TRUE`, que es llegeix com na (valors perduts) rm (remove, esborrar).

Figure 3.4: Evolució temporal de tres països seleccionats.

3.1.4 *Altres mesures de democràcia*

Més enllà del Polity IV en aquest apartat mencionem breument altres fonts d'informació des d'on obtenir dades de democràcia i llibertats polítiques.

La Freedom House és una institució nord-americana que genera anualment des de 1973 una mesura de llibertats civils i polítiques. Per tant, està més pensada cap a qüestions menys procedimentals i més de drets polítics. Es pot trobar a la [web de la Freedom House](#).

Cheibub, Gandhi, and Vreeland (2010) és una altra referència on es classifiquen els règims polítics en democràcies / dictadures, i també dins de les democràcies com a parlamentàries, semi-presidencials i presidencials. El rang temporal va des de 1946 (o l'any de la independència) fins al 2008. És especialment detallat el [document tècnic "codebook"](#).

Cal també destacar el projecte V-Dem (Varieties of Democracy) que està generant un índex democràtic basat en 11 components secundaris i 5 de principals (electoral, liberal, deliberativa, igualitària i participativa) i basat en codificacions d'experts nacionals des del 1900 fins a data d'avui (Coppedge et al. 2017).

Finalment, i tot i que no suposen un pas en la recollida de dades, sinó una contribució a com les dades s'agreguen en índexs, cal mencionar Treier and Jackman (2008), que representen segurament l'expressió més moderna i més complexa en termes metodològics. En aquest cas s'opta per mirar d'incloure el màxim nombre de components en el càlcul del nombre final assignat a cada règim. I en lloc de decantar-se per un tipus d'error preferible a l'hora de donar un veredict binari (democràcia-sí / democràcia-no) el mètode matemàtic utilitzat permet donar amb molta precisió els marges d'error de cadascun dels països, classificar-los en una dimensió contínua (no binària) i poder calcular amb precisió quina és la probabilitat que el país X sigui més democràtic que el país Y i per quin marge.

3.2 *Limitacions al poder: vetos*

Una altra variable clau per entendre la presa de decisions dels països és la capacitat que té el sistema polític i institucional de limitar o constrenyir les decisions que es prenguin. Es pot entendre com el nombre de filtres que ha de passar una decisió per a poder ser efectivament adoptada. Per exemple, una dictadura no té cap mena de veto: hi ha un poder únic i centralitzat que ho decideix tot. Un sistema basat en majories fortes, parlamentarisme, divisió de poders horitzontal (legislatiu, executiu i judicial) i vertical (federalisme) posa més punts de veto a les decisions. Al capdavall l'índex mesura la dificultat d'un sistema polític d'adaptar-se a les circumstàncies i a les preferències dels seus actors, i es fa servir sovint per predir la variabilitat de les polítiques i el risc de canvis polítics (Henisz 2002, Tsebelis (2002)).

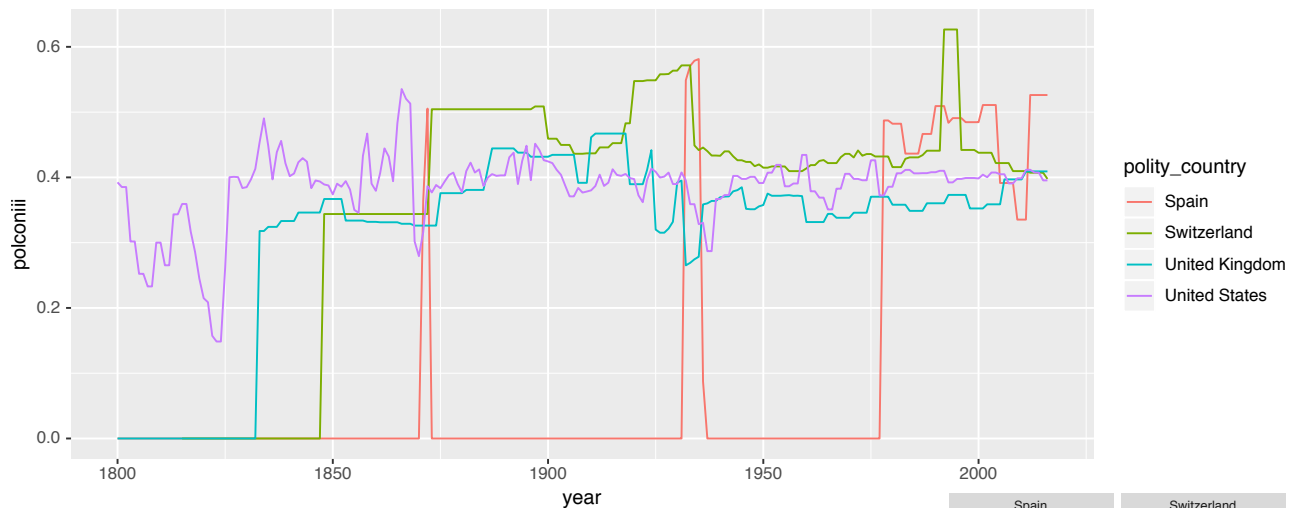
Les dades del que s'anomenen *political constraints* les podem carregar de

la següent manera:

```
library(openxlsx)
url <- "https://mgmt.wharton.upenn.edu/wp-content/uploads/2017/02/POLCON_2017.xlsx"
polcon <- read.xlsx(url, sheet = 1, cols = 1:8)
```

La variable `polconiii` conté els valors de l'índex, que podem representar per a una sèrie d'estats candidats que triem prèviament, de manera sobreposada en una única figura:

```
candidats <- c("United States", "Switzerland", "United Kingdom", "Spain")
polcon.candidats <- polcon %>%
  filter(polity_country %in% candidats)
ggplot(polcon.candidats, aes(x = year, y = polconiii, color = polity_country)) +
  geom_line()
```



O bé de manera independent en figuretes separades a través de la funció `facet_wrap()`.

```
ggplot(polcon.candidats, aes(x = year, y = polconiii)) +
  geom_line() +
  facet_wrap(~ polity_country)
```

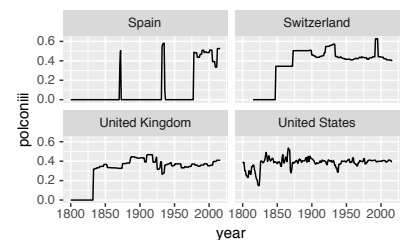


Figure 3.6: Evolució temporal de l'índex de limitacions del sistema polític (PolCon) per a quatre estats, en diferents gràfiques.

Una altra mesura de la tipologia del sistema polític en quant a limitacions de poder és la primera dimensió (executiu-partits) de Lijphart (2012), que classifica els règims en un continuum que va entre els majoritaris i els consociacionals. Podeu veure altres especificacions al [Grup de política comparada de la Universitat de Greifswald](#).

3.3 Descentralització i federalisme

La divisió de poders vertical en forma de federalisme i descentralització ha estat abordada també de manera comparativa a través d'índexs que calculen el grau de descentralització. L'aproximació més global i sistemàtica és la de mesura d'autoritat regional descrita a Hooghe et al. (2016)², tot i que tradicionalment per mesurar el nivell de descentralització s'ha fet servir la dada del Banc Mundial sobre despeses d'unitats subestats.

El primer capítol del llibre és un exemple excel·lent a com es passa en política comparada de conceptes a operacionalitzacions i s'acaba amb índexs, amb un enorme interès en els detalls procedimentals i les decisions que cal prendre.

L'índex d'autoritat regional cobreix 81 estats entre 1950 i 2010, i comprèn dues dimensions: *self rule* i *shared rule*. La primera fa referència a la capacitat que té un territori subnacional per exercir el govern en el seu propi territori. La segona fa referència a la capacitat que té un territori subnacional de co-exercir el govern de l'estat.

Les dades en aquesta base estan disponibles per a les unitats regionals, i també com a agregats per països. En aquest últim cas els índexs agregats els podem obtenir de la següent manera:

```
library(readstata13)
tmp <- tempfile()
download.file("http://garymarks.web.unc.edu/files/2017/04/RAI-country-data-online-version.11.01.15-2-1.zip",
  tmp)
file <- unzip(tmp, files = "RAI country data online version.11.01.15.dta")
rai <- read.dta13(file)
dim(rai)

## [1] 3827  30

names(rai)

## [1] "cowcode"      "iso3166"
## [3] "country_id"   "country_name"
## [5] "abbr_country" "year"
## [7] "n_instdepth" "n_policy"
## [9] "n_fiscauto"   "n_borrowauto"
## [11] "n_rep"        "n_lawmaking"
## [13] "n_execcon"    "n_fiscon"
## [15] "n_borrowcon"  "n_constit"
## [17] "n_selfrule"   "n_sharedrule"
## [19] "n_RAI"        "components"
## [21] "n_rep0"       "n_assembly"
## [23] "n_executive"  "n_lawmaking0"
## [25] "n_law_a"      "n_law_b"
```

² Més endavant veurem una aproximació a l'autoritat internacional dels mateixos autors.

```
## [27] "n_law_c"      "n_law_d"
## [29] "n_law_e"      "n_law_f"
```

Les variables `n_selfrule` i `n_sharedrule` corresponen a les dues dimensions, respectivament, i la variable `n_RAI` és la seva suma.

Per representar els valors en dos eixos per a un any concret (el darrer valor) hem de filtrar les dades per any, i reanomenar les variables:

```
rai.last <- filter(rai, year == max(year))
ggplot(rai.last, aes(x = n_selfrule, y = n_sharedrule, label = country_name)) +
  geom_text() +
  xlab("Self rule") + ylab("Shared rule")
```

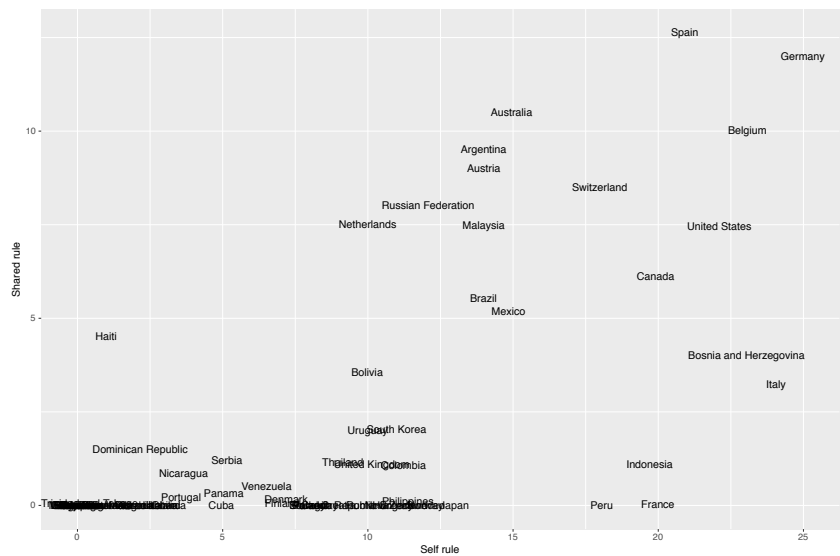


Figure 3.7: Distribució de les dues dimensions (self-rule i shared rule) de l'índex d'autoritat regional en l'últim any disponible.

3.4 Regulació

El Banc Mundial manté una base de dades d'indicadors de governança, els **Worldwide Governance Indicators**, que tot i el poc abast temporal (1996-2006), representen un esforç considerable per sintetitzar dades de diferents fonts.

L'aproximació en aquest cas no es basa en fonaments teòrics sòlids dels indicadors, sinó que la idea és combinar dades de moltes fonts diferents de tal manera que puguem classificar els països anualment segons sis dimensions institucionals i de govern:

- Veü i rendició de comptes
- Estabilitat política i absència de violència
- Efectivitat del govern

- Qualitat reguladora
- Imperi de la llei
- Control de la corrupció

Els indicadors han rebut força crítiques, que tot i ser en part contestades pels responsables, representen en part un llast almenys des del punt de vista acadèmic. No obstant això, es tracta d'una base de dades d'un gran abast i que incorpora idees que malgrat ser molt esteses, no sempre s'han pogut mesurar fàcilment (com per exemple l'efectivitat dels governs).

Hi ha una crítica fonamental i una limitació. D'una banda, la crítica és per la part de l'enllaç entre les dades i la teoria. Els indicadors resultants són un compendi de variables agafades de moltes fonts diferents i que no sempre està clar que tinguin clara relació amb el concepte que pretenen mesurar (validesa interna). La limitació és deguda al fet que els índexs estan generats de tal manera que estan centrats en zero anualment. D'aquesta manera, tots els països estan més o menys lluny del a mitjana centrada a zero. Però no hi ha manera de dir si d'any en any un país ha millorat en termes absoluts, ja que precisament per estar centrada la mesura en el zero, sempre serà relativa a la resta de països. Per tant, es tracta d'un índex que només es pot interpretar d'any en any, i no pas com una sèrie temporal de valors connectats.

La introducció de dades no és automàtica, sinó que o bé cal que **descarreguis les dades** des d'un full de càlcul i el netegis manualment, o bé pots accedir a través d'una **API del Banc Mundial** i et generis un arxiu amb els països, les variables i els anys que desitgis.

Manualment és tediós, però per a fer-ho en un únic indicador i per a un únic any és possible d'aquesta manera:

```
library(openxlsx)
url <- "http://info.worldbank.org/governance/wgi/pdf/wgidataset.xlsx"
indicador <- "Government effectiveness"
indicador.llibre <- "GovernmentEffectiveness"
any <- 2016
any.columna <- 105 # "DA" Estimador, any 2016
wgi <- read.xlsx(url, sheet = indicador.llibre,
                startRow = 15, cols = c(1, 2, any.columna),
                na.strings = "#N/A")
str(wgi)

## 'data.frame': 214 obs. of 3 variables:
## $ Country/Territory: chr "Aruba" "Andorra" "Afghanistan" "Angola" ...
## $ WBCode : chr "ABW" "ADO" "AFG" "AGO" ...
## $ Estimate : num 0.896 1.86 -1.222 -1.039 1.329 ...
```

Els deu estats amb governs menys efectius són els següents:


```
wgi %>%
  arrange(Estimate) %>%
  slice(1:10)
```

##	Country/Territory	WBCode	Estimate
## 1	South Sudan	SSD	-2.2599
## 2	Somalia	SOM	-2.1804
## 3	Haiti	HTI	-2.0583
## 4	Libya	LBY	-1.8915
## 5	Syrian Arab Republic	SYR	-1.8250
## 6	Yemen, Rep.	YEM	-1.8195
## 7	Central African Republic	CAF	-1.7681
## 8	Eritrea	ERI	-1.6850
## 9	Korea, Dem. Rep.	PRK	-1.6475
## 10	Guinea-Bissau	GNB	-1.6354

La primera línia pren l'objecte creat anteriorment i l'envia a una canonada, que serà utilitzada per les funcions que vénen després. La segona línia ha arranjat (ordenat, `arrange()`) les dades per ordre de la variable `Estimate`, i la tercera línia fa un tall de les deu primeres observacions (`slice()`).

Els deu estats amb governs més efectius són els següents (igual que abans, però amb valor descendent aplicant la funció `desc()`):

```
wgi %>%
  arrange(desc(Estimate)) %>%
  slice(1:10)
```

##	Country/Territory	WBCode	Estimate
## 1	Singapore	SGP	2.2086
## 2	Switzerland	CHE	2.0256
## 3	Denmark	DNK	1.8876
## 4	Norway	NOR	1.8829
## 5	Hong Kong SAR, China	HKG	1.8609
## 6	Andorra	ADO	1.8602
## 7	New Zealand	NZL	1.8576
## 8	Finland	FIN	1.8485
## 9	Netherlands	NLD	1.8413
## 10	Japan	JPN	1.8324

La distribució dels valors la podem obtenir amb un histograma³, i afegir-hi una línia vertical al lloc del país candidat que vulguem.

```
candidat <- "Germany"
ggplot(wgi, aes(x = Estimate)) +
  geom_histogram() +
  geom_vline(xintercept = wgi$Estimate[wgi$'Country/Territory' == candidat])
```

³ Fixa't que la variable està centrada en el zero, per definició, com ja hem comentat.

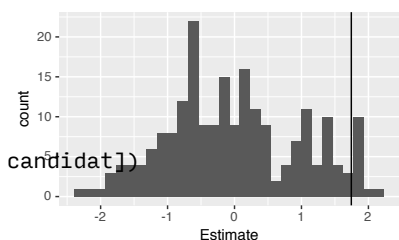


Figure 3.8: Distribució dels valors de l'eficàcia del govern, amb una línia vertical assenyalant un país candidat, en aquest cas Alemanya.

En aquest cas, i degut a què la variable de l'estat conté un caràcter estrany, hem hagut de referir-nos a la variable afegint-hi un accent obert al principi i al final.

3.5 Societat, partits polítics i eleccions

Com a estudiant de relacions internacionals, cal que no oblidis aspectes electorals i de valors socials, encara que no siguin centrals a la disciplina. En aquest apartat hi ha breument mencionats els projectes més rellevants a nivell comparat.

En primer lloc, les posicions dels partits en diferents eixos i com compararlos es poden trobar al [Comparative Party Manifesto](#), que a través de l'anàlisi de programes electorals (textos) genera indicadors de posicions ideològiques dels partits. Té un abast de 56 estats des del final de la segona guerra mundial fins a l'actualitat, amb 715 eleccions, més de mil partits i més de quatre mil programes electorals analitzats (Volkens 2017).

En segon lloc, les dades de resultats electorals a nivell comparat es poden obtenir del [Global Elections Database](#)⁴, i estan sistematitzats a Brancati (2014).

En tercer lloc, el [World Values Survey](#) és una enquesta administrada a diferents estats en successives onades (fins el 2015 n'hi ha hagut 6), que pretén, extreure valors socials a diferents països d'una manera comparada. Les microdades (totes i cadascuna de les persones enquestades) són fàcilment accessibles, però la mateixa pàgina web té una [interfície online](#) que permet fer consultes simples. En la mateixa línia es troba l'enquesta social europea ([European Social Survey](#)), que genera dades comparades per a 36 països europeus cada dos anys.

⁴ Cal registre previ, però permet construir-se un mateix les matrius de dades.

4

Relacions: Contagi, Imitació, Dependència

Lectures:

- Tena Junguito (1992), capítol 1
- Tena Junguito (1992), capítol 2

La idea és simple, però molt poderosa: com més relacions hi hagi entre dos estats, més interessos tindran en comú, més es coneixeran i menors seran les desconfiances, de manera que la probabilitat de resoldre conflictes de manera violenta tendirà a reduir-se. El resum és clar: més contacte, menys guerra.

És aquesta idea la que viu en el projecte *Correlates of War*, que es dedica a compilar matrius de dades que mesurin el conflicte entre nacions a través del temps, els seus llaços històrics i colonials, el flux de comerç, l'engranatge diplomàtic, les fronteres comuns, o les aliances.

En contrast amb el que hem vist fins ara on les matrius de dades tenen una unitat d'anàlisi única (un estat), que es pot observar a través del temps (dimensió temporal) i a les quals els podem assignar diferents variables, a partir d'ara les unitats d'anàlisi seran unitats diàdiques (Estat A - Estat B), de manera que per cada any no tindrem uns dos-cents països a observar, sinó uns dos-cents països per dos-cents països relacionats. Com pots veure, la complexitat per gestionar aquest tipus de dades s'incrementa exponencialment. Per exemple, comptant una mitjana de 100 Estats al món des de 1950 fins a 2010 (60 anys), tenim $100 * 60 = 6000$ valors. En canvi, amb dades diàdiques tenim 60 anys igualment, però a cada any hi tenim $100 * 100 = 10.000$ relacions, fent un total de 600.000 valors a gestionar.

4.1 Comerç

Les dades més complexes però alhora més refinades que es poden trobar en relació al contacte entre països fan referència al **comerç bilateral** que s'estableix entre ells. En aquest sentit, Gleditsch (2002) elabora una primera proposta metodològica centrada a resoldre el problema de falta de registres en segons quins països. Darrerament, aquesta proposta s'ha vist en certa

manera superada per Barbieri, Keshk, and Pollins (2009), que té un abast més general.

Preparar les dades per a què R les pugui manipular no és pas complicat, tot i que el següent fragment d'instruccions sembli enrevessat. Però pensa que en quatre línies escasses estas realitzant, de manera automatitzada, la descàrrega d'un arxiu comprimit des d'una URL, la descompressió d'un arxiu contingut en el .zip inicial, la lectura per part d'R de la matriu de dades com si es tractés d'una taula i la creació d'un objecte que pugui ser processat posteriorment.

Per obtenir les dades cal fer servir la funció `read.table()`, que permet llegir arxius separats per comes (o tabulacions, o punt-i-coma). No obstant això, com que l'arxiu està comprimit, abans cal descarregar-lo per tenir-lo en memòria d'R. La penúltima última línia ens dóna el nom de les variables i l'última el nombre d'observacions i de variables.

```
tmp <- tempfile()
download.file("http://www.correlatesofwar.org/data-sets/bilateral-trade/cow_trade_4.0/at_download/file",
             tmp, method = "libcurl", mode = "wb")
file <- unzip(tmp, files = "COW_Trade_4.0/Dyadic_COW_4.0.csv")
trade <- read.table(file, header = TRUE, sep = ",", na = "-9", fill = TRUE)
names(trade)

## [1] "ccode1"          "ccode2"
## [3] "year"           "importer1"
## [5] "importer2"      "flow1"
## [7] "flow2"          "smoothflow1"
## [9] "smoothflow2"    "smoothtotrade"
## [11] "spike1"         "spike2"
## [13] "dip1"           "dip2"
## [15] "trdspike"       "tradedip"
## [17] "bel_lux_alt_flow1" "bel_lux_alt_flow2"
## [19] "china_alt_flow1" "china_alt_flow2"
## [21] "source1"        "source2"
## [23] "version"

dim(trade)

## [1] 826893    23
```

Per tal de simplificar la gestió de l'exemple concret, farem servir només dades de pocs països. Especifiquem una llista de països candidats i després enviem l'objecte `trade` original a una canonada que primer redueix el nombre de casos a aquells que són dels països candidats (funció `filter()`) i després es queda només amb 5 variables (funció `select()`), per acabar reanomenant la columna del comerç total (`rename()`):

```

candidats <- c("United States of America", "France",
              "Spain", "United Kingdom", "Italy",
              "Argentina", "Mexico")
trade.candidats <- trade %>%
  filter(importer1 %in% candidats & importer2 %in% candidats) %>%
  select(year, importer1, importer2, smoothtotrade) %>%
  rename(trade = smoothtotrade)
dim(trade.candidats)

## [1] 2894    4

```

Com que les dades només són unidireccionals (només hi ha una parella $A - B$, però no la mateixa parella $B - A$), cal repetir-les per tal de poder trobar totes les possibles combinacions. Aquest és un cas específic d'aquesta matriu de dades, que no es troba en forma de xarxa, però en altres casos aquest pas no és necessari:

```

trade.candidats.repeticio <- trade.candidats %>%
  select(importer1 = importer2, importer2 = importer1, year, trade)
trade.candidats <- bind_rows(trade.candidats, trade.candidats.repeticio) %>%
  unique()

```

La Figura total de les importacions des del primer Estat candidat (en aquest cas, els Estats Units) en relació amb la resta de països es pot obtenir amb les següents instruccions:

```

imp.candidat.1 <- filter(trade.candidats, importer1 == candidats[1])
ggplot(imp.candidat.1, aes(x = year, y = trade, color = importer2)) +
  geom_line()

```

També de manera simple podem obtenir quina és la proporció que cada país representa sobre el total d'importacions dels Estats Units, assumint que només la resta dels quatre Estats conformen el total de socis comercials. En aquest cas, hem d'agrupar les dades per anys i calcular-ne la proporció que cada país representa sobre el total anual. En aquest cas, però, en lloc d'obtenir un resum anual (com hem fet fins ara amb la funció `summarize()`), el que fem és obtenir un valor per a cada any i parella d'importacions / exportacions (amb la funció `mutate()`), de manera que la proporció és el valor d'un país concret `smoothflow1` dividit per la suma dels demés països en aquell any concret (`sum(smoothflow1)`), ja que hem agrupat per any.

```

imp.prop <- imp.candidat.1 %>%
  group_by(year) %>%
  mutate(prop.imp = trade / sum(trade))

```

En aquest cas la figura amb proporcions és la següent:

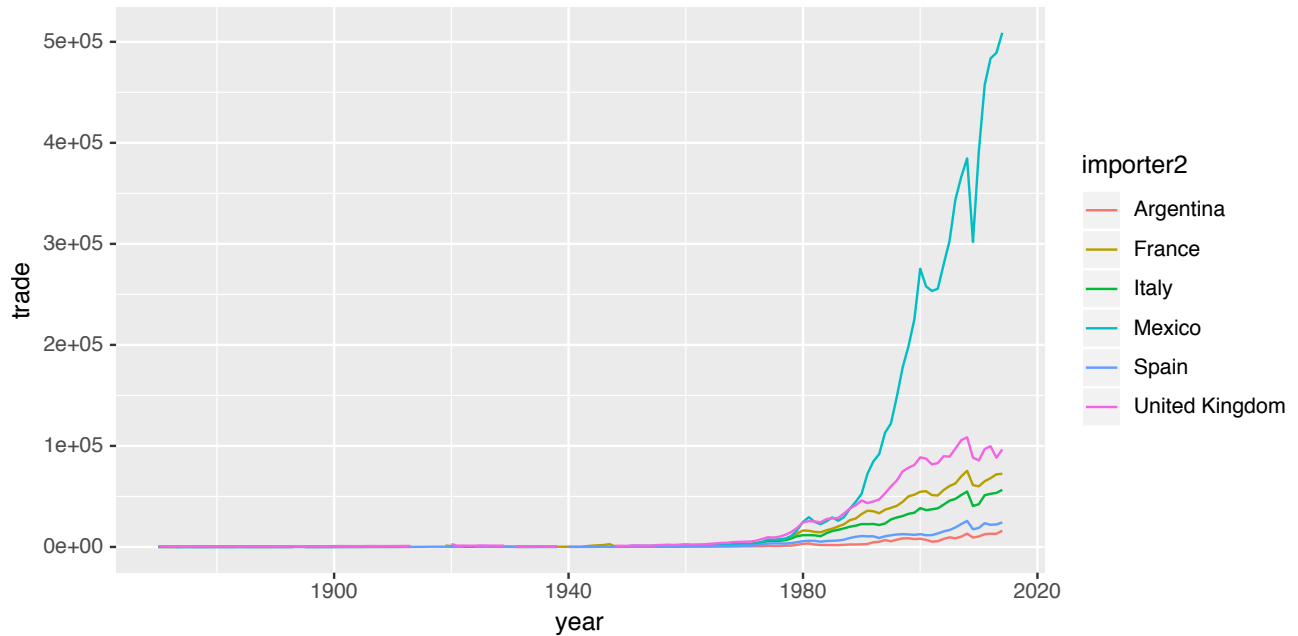


Figure 4.1: Evolució del flux d'importacions des dels Estats Units, en termes absoluts

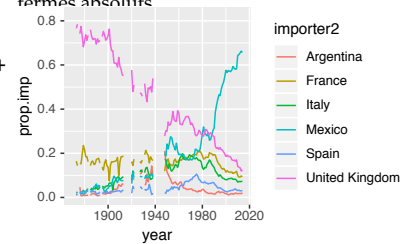


Figure 4.2: Evolució del flux d'importacions des dels Estats Units, en proporció

```
ggplot(imp.prop, aes(x = year, y = prop.imp, color = importer2)) +  
  geom_line()
```

Una altra manera de visualitzar dades de matrius complexes és a través de figures de mosaic. En aquest cas, però, s'ha de sacrificar la dimensió temporal (l'any), per tal de guanyar la dimensió de relació diàdica (un a un, amb tots els membres del grup)¹

```
trade.candidats.2010 <- filter(trade.candidats, year == 2010)  
ggplot(trade.candidats.2010,  
  aes(x = importer1, y = importer2)) +  
  geom_tile(aes(fill = trade))
```

Finalment, les matrius diàdiques també es poden mostrar amb figures de xarxa. En aquest cas, cada unitat d'anàlisi és un node que té un punt a l'espai, i les relacions s'expressen a través de línies que connecten les unitats. La mida de les línies sol respectar la proporció o el flux total d'intercanvi.

```
library(network)  
trade.candidats.2010.t <- reshape2::acast(trade.candidats.2010,  
  importer1 ~ importer2, value.var = "trade", fill = 0)  
trade.candidats.2010.t.n <- (trade.candidats.2010.t /  
  max(trade.candidats.2010.t, na.rm = TRUE)) + 0.001
```

Fixa't que hi ha valors a l'origen de importer1, ja que les dades no estan repetides, i el que són importacions en aquest cas per Espanya estan calculades com exportacions des d'altres països. Però gestionar això requereix més complexitat, fora de l'abast i el propòsit d'aquests materials docents.

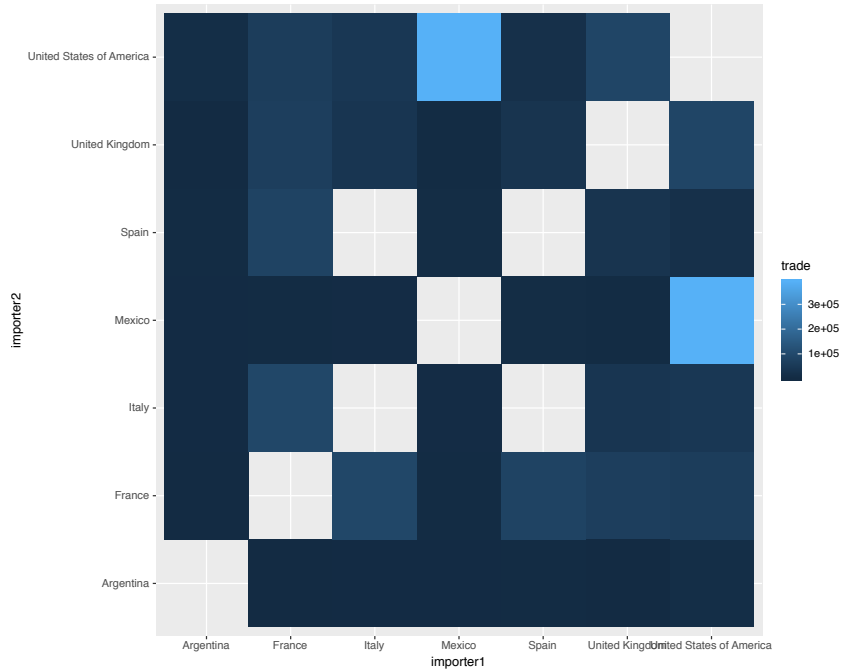


Figure 4.3: Valor de les importacions des del país i al país 2, en valors absoluts.

```
xarxa <- network(trade.candidats.2010.t.n, matrix.type = "adjacency",
  ignore.eval = FALSE, names.eval = "trade")
```

Un cop calculat l'objecte de relacions binàries, només cal passar-lo a la funció `ggnet2()` de la llibreria `GGally`, que és la que s'encarrega de la visualització:

```
library(GGally)
ggnet2(xarxa, label = TRUE, edge.size = "trade")
```

4.2 Distància

Considerar la distància geogràfica entre dos països com una variable més de relació entre ells no és senzill, al contrari del que podria semblar a simple vista. Si més no, els mapes són una manera molt simple de veure si un país és a prop d'un altre o no. Però hi ha diferents aspectes que compliquen molt un càlcul simple:

- Com tractar insularitats. Per calcular la distància entre Itàlia i Espanya, val més que fem servir les penínsules o les illes?
- Com tractar territoris d'ultramar o fora de l'abast geogràfic del gruix de la població. Per exemple, les illes Canàries. Cenyint-nos a qualsevol territori de l'estat espanyol, aquest és més a prop de Mauritània que d'Alemanya, per la proximitat de les Canàries al primer.

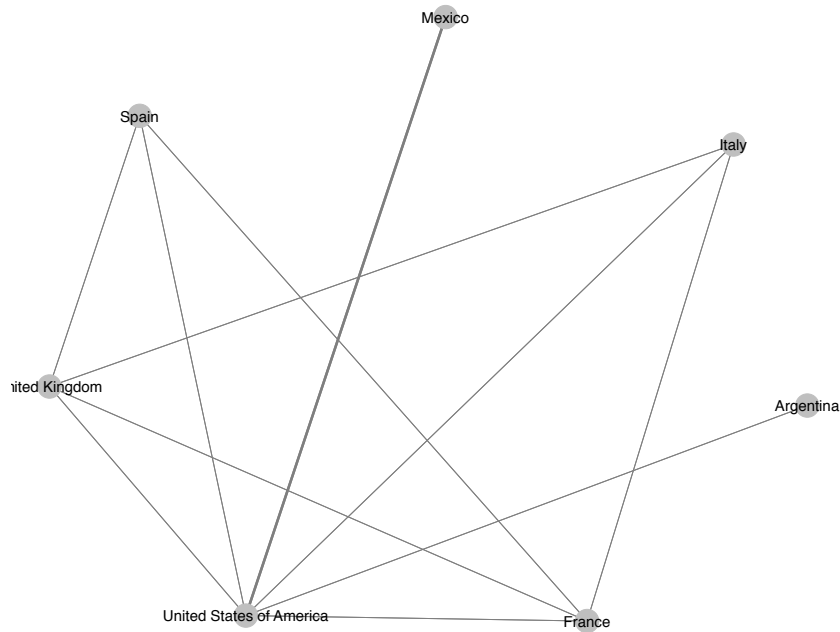


Figure 4.4: Xarxa de comerç.

- Si hi ha una frontera comuna, és automàtic que parlem de distància zero?

De manera que al capdavant cal sempre fer alguna mena de tria per simplificar el resultat. I normalment la tria és entre o bé considerar matrius binàries de si dos estats comparteixen frontera o no (Stinnett et al. 2002), o bé calcular la distància entre les capitals (i potser ponderant per fronteres comunes d'alguna manera)².

Les **dades de contigüitat** les pots obtenir de la següent manera³:

```
tmp <- tempfile()
download.file("http://www.correlatesofwar.org/data-sets/direct-contiguity/direct-contiguity-v3-2/at_download/file", file = "DirectContiguity320/contdird.csv")
cont <- read.table(file, header = TRUE, sep = ",")
head(cont)
```

```
##   dyad state1no state1ab state2no state2ab
## 1 2020         2     USA         20     CAN
## 2 2020         2     USA         20     CAN
## 3 2020         2     USA         20     CAN
## 4 2020         2     USA         20     CAN
## 5 2020         2     USA         20     CAN
## 6 2020         2     USA         20     CAN
##   year conttype version
## 1 1920         1     3.2
## 2 1921         1     3.2
## 3 1922         1     3.2
```

² Per una discussió tècnica complexa del fet que la distància entre països és més que no pas la geografia vegeu Beck, Gleditsch, and Beardsley (2006)

³ Consulta el llibre de codis per més detalls sobre la variable de contigüitat, un valor d'1 és contigüitat terrestre, i els valors de 2 a 5 són diferents distàncies de contigüitat marítima.


```
## 4 1923      1      3.2
## 5 1924      1      3.2
## 6 1925      1      3.2
```

Creem un diagrama de xarxa amb les contigüitats d'uns quants països candidats, i per a aquelles parelles per les quals no tenim dades els afegim un zero, indicant que no tenen cap mena de contigüitat, ni tan sols marítima.

```
candidats <- c("USA", "CAN", "MEX", "AND", "FRN", "SPN", "ITA", "UKG")
cont.candidats <- cont %>%
  filter(year == 2010) %>%
  filter(state1ab %in% candidats & state2ab %in% candidats) %>%
  select(state1ab, state2ab, conttype)
cont.matriu <- reshape2::acast(cont.candidats,
                               state1ab ~ state2ab, value.var = "conttype",
                               fill = 0)
xarxa <- network(cont.matriu,
                 matrix.type = "adjacency", ignore.eval = FALSE,
                 names.eval = "conttype")
ggnet2(xarxa, label = TRUE, edge.lty = "conttype")
```

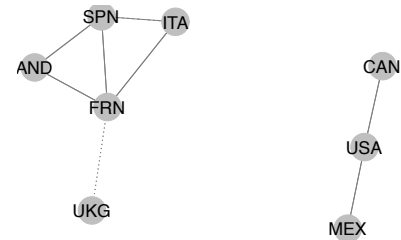


Figure 4.5: Representació de la xarxa de contigüitats de diferents països seleccionats. El tipus de línia representa el tipus de contigüitat.

Les dades de distància entre capitals són simples d'obtenir amb el paquet geosphere:

```
library(maps)      # Permet importar latituds i longituds
data(world.cities)
library(geosphere) # Permet calcular distàncies a través de punts en l'espai geogràfic
capitals <- world.cities %>%
  filter(capital == 1) %>%
  select(country.etc, lat, long)
dist.capitals <- distm(cbind(capitals$long, capitals$lat))
str(dist.capitals)
```

```
## num [1:230, 1:230] 0 1986720 3892960 4758970 18458482 ...
```

El resultat, una matriu quadrada de 230x230 casos, es pot també estudiar com una xarxa de relacions

```
dimnames(dist.capitals)[[1]] <- capitals$country.etc
dimnames(dist.capitals)[[2]] <- capitals$country.etc
dist.capitals <- dist.capitals / max(dist.capitals, na.rm = TRUE)
xarxa <- network(dist.capitals[1:10,1:10],
                 matrix.type = "adjacency", ignore.eval = FALSE,
                 names.eval = "distance")
ggnet2(xarxa, label = TRUE, edge.size = "distance")
```

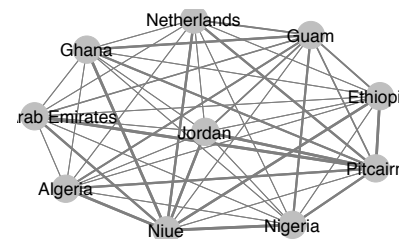


Figure 4.6: Representació de la xarxa de distàncies entre 10 capitals.

4.3 *Relacions històriques i diplomàtiques*

Més enllà de fluxos directament observables com són els relacionats amb el comerç, o de mesures de relació geogràfica, trobem també associacions entre estats basades en història comuna, en colonitzacions passades que deixen empremta. La *base de dades de contigüitat i descolonització* té una estructura molt semblant a la de contigüitat. La *base de dades de religions* detallada a Maoz and Henderson (2013) conté també valors rellevants per a establir lligams històrics. Val la pena destacar també la *base de dades de disputes militars entre estats* detallada a Palmer et al. (2015) i la de *base d'aliances* detallada a Gibler (2008).

En una altra dimensió se situa l'estudi de la semblança entre països en el nivell diplomàtic. En aquest cas, el que es busca és veure comportaments semblants a nivell de *vot a l'assemblea general de les Nacions Unides*.

Voeten (2013) presenta les dades de vot compartit entre països, i també hi afegeix el càlcul dels anomenats "punts ideals", o preferències estatals, que no deixa de ser una aplicació semblant al mètode de Treier and Jackman (2008) per calcular nivells de democràcia.

4.4 *Guia de lectura*

Les lectures del mòdul de relacions interestatals estan plantejades en un ordre concret que et convé seguir:

- Tena
- Skrede Gleditsch
- Barbieri, Keshk i Pollins

4.5 *Guia de lectura (Tena)*

La lectura de Tena s'ha d'entendre en el context en què es publica: el Banc d'Espanya. Per a aquest organisme públic és essencial poder elaborar registres precisos i fiables dels fluxos comercials que manté l'estat espanyol. No hi ha un interès acadèmic, sinó pràctic: poder preveure millor l'estat de l'economia i, per tant, fer previsions més fiables que ajudin els polítics a prendre decisions sobre política econòmica.

L'objectiu de la lectura no és pas el de fer-te recordar sistemàticament totes les possibles fonts de biaixos que poden haver-hi quan es compilen dades de comerç internacional, sino fer-te conscient que una activitat aparentment tant rutinària i tan poc donada a la interpretació com és registrar i donar valor a totes les mercaderies que travessen una frontera de fet conté

Quan estiguis llegint fixa't en:

- Quin nombre d'estats té en compte aquesta lectura? Quina part del món està representada amb aquests problemes?

- Creus que són estats amb infraestructura per poder realitzar acuradament un inventari d'importacions o exportacions o estats on la infraestructura és precària i, per tant, els problemes podrien ser encara més greus?
- Quina solució proposaries per millorar la fiabilitat de les mesures de comerç internacional?

4.6 Guia de lectura (Skrede Gleditsch)

Com a complement a la lectura de Tena, la lectura de Skrede Gleditsch no es centra tant en les dificultats tècniques que representa poder mesurar correctament els fluxes de comerç internacional, sinó les implicacions empíriques que això té. És a dir, no li interessa tant identificar d'on poden provenir els biaixos, sinó *quin efecte* poden tenir aquests biaixos a l'hora d'establir relacions causals entre el fluxe de comerç internacional i els conflictes interestatals. Has d'entendre la lectura, doncs, com una part secundària de la feina principal que interessa a aquesta línia acadèmica: respondre la pregunta de si les relacions comercials redueixen o no els conflictes polítics i militars entre països.

D'aquesta manera s'entén molt millor per què l'autor fa tant d'èmfasi en el problema de les dades que manquen (*missing data*). A algú del Banc Central d'un país el podria preocupar més aviat poc tenir una dada que falta en la relació amb un país concret per quadrar els comptes, per fer prediccions econòmiques, etc... (recorda que aquest és l'interès inicial que tenen els estats per registrar aquestes transaccions, i que queden molt ben explicades a la lectura de Tena). Però a un investigador la inexistència d'una dada pot significar una manera de registrar un conflicte. De fet, si sóc governant d'un estat que té un conflicte polític obert amb un altre, potser de cara endins ja em convé dir-li a la meua població que l'altre és l'enemic i que per tant no hi tenim tractes comercials (encara que sí que n'hi hagi). Per tant, faré que sigui una decisió política que el registre d'intercanvi comercial sigui zero o no conegut (*missing*).

Fixa't en la lectura, doncs, com una lluita dels investigadors per tal de desvetllar els motius reals de no disposar de dades de comerç bilateral. No és que els preocupin les dades en sí mateixes (a millors dades, millors resultats), sinó que els preocupen els (si vols veure-ho així) "missatges subliminals" que les dades o la seva inexistència els donen sobre el grau de conflicte interestatal.

Centra els teus esforços en identificar quina és la solució fonamental que presenta Skrede Gleditsch per solucionar el problema de la disposició de dades de comerç

4.7 Guia de lectura (Barbieri, Keshk i Pollins)

El que de manera implícita en la lectura anterior de Skrede Gleditsch (relació entre disposició de dades i quina informació això ens dóna per avaluar el nivell de conflicte entre dos estats) es converteix en aquesta lectura en l'objectiu explícit.

Mira de relacionar la primera part d'aquesta lectura (*Common Problems with Official Trade Statistics, Inaccurate Trade Reports, Discrepant Partner Reports*) amb el que has après de la lectura anterior de Tena. I la secció de *Missing Trade Reports* amb la de Skrede Gleditsch.

Identifica clarament les diferències entre Barbieri, Keshk i Pollin (BKP) d'una banda i Skrede Gleditsch per l'altra. Has de tenir en compte, però, que Skrede Gleditsch és anterior:

- Quines són les innovacions tècniques que plantegen BKP?
- Com milloren la manera de Skrede Gleditsch?
- Creus que val la pena l'esforç i dedicació de BKP per obtenir la millora que argumenten?

Més enllà de les lectures, ets capaç d'imaginar-te en un món connectat a la xarxa, si hi ha alternatives a les sèries de comerç internacional per avaluar la relació econòmica entre dos estats? Si haguessis de començar a estudiar-ho a data d'avui, creus que faries servir les sèries de comerç internacional o buscaries una altra cosa?

5

Organitzacions Internacionals i tractats

Lectures:

- Ahrne et al. (2015), capítol 8

En els darrers anys hi ha hagut un gran increment en l'interès per estudiar organitzacions internacionals de manera comparada. Tradicionalment, les Relacions Internacionals o els estudis des del punt de vista del Dret o l'economia s'han enfocat a analitzar organitzacions internacionals amb aproximacions quasi-antropològiques. És a dir, amb estudis de cas en profunditat, però sense relacionar-los amb altres organitzacions. Es considerava que eren organitzacions prou diferents com per a merèixer tractaments diferents, i no pas considerar-se part d'una mateixa història. Però la ciència política, i especialment en els darrers anys, ha estat ocupada en estudiar de manera comparada les organitzacions internacionals. En aquest mòdul introduïrem les compilacions de dades sobre organitzacions que existeixen.

5.1 *Correlates of War: la base*

La referència en organitzacions internacionals des d'una perspectiva comparada és la [base de dades de *Correlates of War* sobre organitzacions internacionals](#), activa des dels anys setanta (Wallace and Singer 1970). La idea primigènia era la d'estudiar fins a quin punt el contacte dels països en diferents organitzacions internacionals aplacava els afanys d'agressió i de guerra mútua. No deixa de ser una extensió del que hem vist al mòdul anterior sobre relacions bilaterals, contagi i contacte. Però en aquest cas, formalitzat en organitzacions internacionals.

La base de dades actual, en la seva versió 2.3, va ser publicada el 2015 i conté dades d'organitzacions internacionals entre el 1815 i el 2010. Les característiques per definir una organització com a organització internacional passen per complir els tres criteris següents:

- *Membres*: Com a mínim hi ha d'haver tres membres del sistema internacional d'estats.

- *Sessions*: Com a mínim tenir una sessió plenària cada 10 anys.
- *Secretariat*: Té un secretariat permanent amb la seva corresponent seu central.

Les dades s'extreuen del [Yearbook of International Organizations](#), que publica anualment la Unió d'Organitzacions Internacionals, i que conté més de 65,000 entrades. Però amb els criteris abans esmentats el nombre d'organitzacions el 2010 es queda en poc més de dues-centes.

Per obtenir les dades d'estats membres a organitzacions internacionals es fa de manera similar a les demés base de dades del projecte COW que hem vist anteriorment:

```
tmp <- tempfile()
download.file("http://www.correlatesofwar.org/data-sets/IGOs/IGO_stateunit_v2.3.zip", tmp)
file <- unzip(tmp, files = "IGO_stateunit_v2.3.csv")
igos <- read.table(file, header = TRUE, sep = ",", na = "-9", fill = TRUE)
names(igos)[1:15] # Només els primers 15 noms de variables

## [1] "ccode" "country" "state"
## [4] "year" "AAAID" "AAB"
## [7] "AACB" "AACarib" "AALCO"
## [10] "AARO" "AATA" "AATPO"
## [13] "ABEDA" "ABEPSEAC" "ACC"

dim(igos)

## [1] 8102 533
```

En aquest cas, la matriu consta de més de vuit mil entrades, corresponents a països * anys, i les columnes, a partir de la cinquena (AAAID) corresponen al tipus de relació entre aquest estat concret i l'organització internacional (des de cap relació de pertinença fins a pertinença plena).

Cal que et refiïs del [llibre de codis](#) per poder obtenir el nom complet de l'organització, i no només l'acrònim tal i com surt a la base de dades. Per exemple, per traçar el nombre d'estats membres de ple dret a les Nacions Unides has d'escollir només la variable "UN" i deixar córrer anys anteriors a la seva creació. Després, crees una variable que té valor 1 *si* el valor de "UN" és 1 (pertenença completa a l'organització) i és sabut (no és un valor perdut o *missing*), i valor 0 quan el valor de "UN" és qualsevol altra cosa, a través de la funció `ifelse()`. Un cop tens la variable de pertinença completa `full.membership` agrupes les dades per any i simplement sumes el nombre d'uns.

```
igo.un <- igos %>%
  select(country, year, UN) %>%
  filter(year >= 1945) %>%
```

```
mutate(full.membership = ifelse(UN == 1 & !is.na(UN), 1, 0)) %>%
group_by(year) %>%
summarize(N = sum(full.membership))
```

I fas la figura d'evolució temporal.

```
ggplot(igo.un, aes(x = year, y = N)) +
geom_line() +
expand_limits(y = 0)
```

En el cas de voler el nombre d'organitzacions al qual un estat pertany cal primer de tot filtrar les observacions per estat membre (`filter()`) i després realitzar una transformació de les dades per tal que les organitzacions internacionals, que són a les columnes, quedin com una variable més¹ a través de la funció `gather()`. Després només cal que apliquis altre cop la transformació per quedar-te amb els casos on l'estat és membre de ple dret i descartis els altres.

```
candidat <- "Argentina"
igo.candidat <- igos %>%
  filter(country == candidat) %>%
  gather(IGO, membership, -ccode, -country, -state, -year) %>%
  mutate(full.membership = ifelse(membership == 1 & !is.na(membership), 1, 0))
head(igo.candidat)
```

```
##   ccode  country    state year  IGO
## 1   160 Argentina argentina 1845 AAAID
## 2   160 Argentina argentina 1850 AAAID
## 3   160 Argentina argentina 1855 AAAID
## 4   160 Argentina argentina 1860 AAAID
## 5   160 Argentina argentina 1865 AAAID
## 6   160 Argentina argentina 1870 AAAID
## membership full.membership
## 1          -1              0
## 2          -1              0
## 3          -1              0
## 4          -1              0
## 5          -1              0
## 6          -1              0
```

Ara agrupa les pertinences per any i calcula la suma d'organitzacions anuals.

```
igo.candidat.anual <- igo.candidat %>%
group_by(year) %>%
summarize(N = sum(full.membership))
```

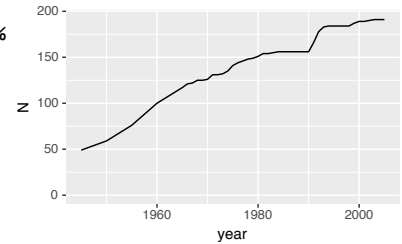


Figure 5.1: Evolució del nombre absolut d'estats membres de ple dret de les Nacions Unides.

¹ Això es coneix com un *reshape*, donar-li forma a les dades. En aquest cas, de versió ampla ("wide", perquè els noms de les variables contenen informació) a versió llarga ("long", que és on s'han multiplicat el nombre d'observacions).

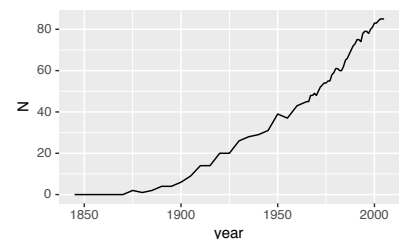


Figure 5.2: Evolució del nombre d'organitzacions internacionals del qual és membre de ple dret un estat candidat, en aquest cas l'Argentina, en termes absoluts.

```
ggplot(igo.candidat.anual, aes(x = year, y = N)) +
  geom_line() +
  expand_limits(y = 0)
```

5.2 Autoritat de les organitzacions internacionals

Recentment s'han començat a estudiar característiques administratives i polítiques de les organitzacions internacionals, més enllà de la xarxa de pertinença. La referència més destacada, tant per abast (quantitat d'organitzacions) com per sistematització (tractament acurat de les dades i interès profund per quantificar i tractar les característiques recollides) és la de Hooghe and Marks (2015).

Basant-se en la base de dades del COW els autors afegeixen tres criteris més:

- *Multifuncionals*: que tinguin parts legislatives, executives i administratives.
- *Personal*: que tinguin almenys 50 persones treballant.
- *Web*: que tinguin una adreça i una web.

En aquest cas els autors busquen establir una mesura de *delegació d'autoritat* dels estats membres cap a les organitzacions internacionals. És a dir, analitzen la capacitat que té l'organització internacional per a fer complir la seva autoritat als estats membres, condicionada a dos factors: el nombre d'estats membres i l'abast del seu *policy portfolio*².

Per a mesurar la delegació d'autoritat estableixen dues mesures diferents: *delegació* ("delegation") entesa com fins a quin punt els estats membres cedeixen l'autoritat de prendre decisions a una part de l'organització (normalment un consell executiu; i *compartició* ("pooling"), que és el fet de prendre decisions de manera conjunta a través dels estats membres i es troba en com els estats membres regulen la presa de decisions, els processos de ratificació i fins a quin punt són vinculants.

A data de finals de 2017, les dades de delegació en organitzacions internacionals s'estan revisant i no estan prou polides per a fer-les servir sense un procediment substancial de neteja, fora de l'abast d'aquests materials docents.

² Que pot ser traduït com la "cartera de polítiques", entès com el tipus de coses que una organització internacional vol fer: es vol centrar en una única cosa o vol fer-ne moltes?

5.3 Llei internacional: tractats

Barbara Koremenos (2013) presenta una base de dades basada en una obra feta uns anys abans (B. Koremenos et al. 2001) que s'anomena *The Continent of International Law*, i que conté tractats internacionals i té una aproximació basada en el dret. També en aquest sentit val la pena destacar la bases de dades de tractats multilaterals de Gamble (2005)

Quant a aplicacions sectorials als tractats, podem trobar dades d'acords internacionals amb la [feina de Robert Mitchell](#), i de compliment de tractats en drets humans a [Simmons \(2009\)](#).

Referències

- Ahrne, Góran, Nils Brunsson, Dieter Kerwer, Jarle Trondal, Frode Vegge-land, Ole Jacob Sending, Ian Hurd, et al. 2015. *Puentes, Fronteras Y Murallas Disciplinarias En Torno a Las Organizaciones Internacionales*. Vol. 9. CIDE.
- Barbieri, Katherine, Omar MG Keshk, and Brian M Pollins. 2009. "Trading Data: Evaluating Our Assumptions and Coding Rules." *Conflict Management and Peace Science* 26 (5). Sage Publications Sage UK: London, England: 471–91.
- Beck, Nathaniel, Kristian Skrede Gleditsch, and Kyle Beardsley. 2006. "Space Is More Than Geography: Using Spatial Econometrics in the Study of Political Economy." *International Studies Quarterly* 50 (1). Wiley Online Library: 27–44.
- Boix, Carles, Michael Miller, and Sebastian Rosato. 2013. "A Complete Data Set of Political Regimes, 1800–2007." *Comparative Political Studies* 46 (12). SAGE Publications Sage CA: Los Angeles, CA: 1523–54.
- Brancati, Dawn. 2014. "Global Elections Database [Computer File]." *New York: Global Elections Database [Distributor], Date Accessed 1: 20*.
- Cheibub, José Antonio, Jennifer Gandhi, and James Raymond Vreeland. 2010. "Democracy and Dictatorship Revisited." *Public Choice* 143 (1-2). Springer: 67–101.
- Coppedge, Michael, John Gerring, Staffan I Lindberg, Svend-Erik Skaaning, Jan Teorell, Joshua Krusell, Kyle L Marquardt, et al. 2017. "V-Dem Methodology V7."
- Gamble, John. 2005. "Comprehensive Statistical Database of Multilateral Treaties." In *American Political Science Association Annual Meeting*.
- Gibler, Douglas M. 2008. *International Military Alliances, 1648-2008*. CQ Press.
- Gleditsch, Kristian Skrede. 2002. "Expanded Trade and Gdp Data." *Journal of Conflict Resolution* 46 (5). Sage Publications Thousand Oaks: 712–24.
- Henisz, Witold J. 2002. "The Institutional Environment for Infrastructure Investment." *Industrial and Corporate Change* 11 (2). Oxford Univ Press: 355–89.
- Hooghe, Liesbet, and Gary Marks. 2015. "Delegation and Pooling in International Organizations." *The Review of International Organizations* 10 (3).

Springer: 305–28.

Hooghe, Liesbet, Gary Marks, Arjan H Schakel, Sara Niedzwiecki, Sandra Chapman Osterkatz, and Sarah Shair-Rosenfield. 2016. *Measuring Regional Authority*. Oxford University Press Oxford, UK.

Koremenos, B., C. Lipson, D. Snidal, and A. Kydd. 2001. "The Rational Design of International Institutions." *International Organization* 55. Cambridge Univ Press: 761–99.

Koremenos, Barbara. 2013. "The Continent of International Law." *Journal of Conflict Resolution* 57 (4). SAGE Publications: 653–81. doi:[doi:10.1177/0022002712448904](https://doi.org/10.1177/0022002712448904).

Lijphart, Arend. 2012. *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. Yale University Press.

Maoz, Zeev, and Errol A Henderson. 2013. "The World Religion Dataset, 1945–2010: Logic, Estimates, and Trends." *International Interactions* 39 (3). Taylor & Francis: 265–91.

Milanovic, Branko. 2010. *The Haves and the Have-Nots: A Brief and Idiosyncratic History of Global Inequality*. Basic books.

Palmer, Glenn, Vito d'Orazio, Michael Kenwick, and Matthew Lane. 2015. "The Mid4 Dataset, 2002–2010: Procedures, Coding Rules and Description." *Conflict Management and Peace Science* 32 (2). SAGE Publications Sage UK: London, England: 222–42.

Przeworski, Adam, Juan Manuel Ortega, and Sara Gordon Rapoport. 1997. "Una Defensa de La Concepción Minimalista de La Democracia." *Revista Mexicana de Sociología*. JSTOR, 3–36.

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge University Press.

Stinnett, Douglas M, Jaroslav Tir, Paul F Diehl, Philip Schafer, and Charles Gochman. 2002. "The Correlates of War (Cow) Project Direct Contiguity Data, Version 3.0." *Conflict Management and Peace Science* 19 (2). Sage Publications Sage CA: Thousand Oaks, CA: 59–67.

Straface, Fernando, and María Page. 2009. "Hacia Una Economía Política de Los Indicadores de Calidad Institucional." *Revista Del CLAD Reforma Y Democracia*, no. 43. Centro Latinoamericano de Administración para el Desarrollo.

Szmolka Vida, Inmaculada. 2010. "Los Regímenes Políticos Híbridos: Democracias Y Autoritarismos Con Adjetivos. Su Conceptualización, Categorización Y Operalización Dentro de La Tipología de Régimenes Políticos." *Revista de Estudios Políticos*, no. 147. Centro de Estudios Políticos y Constitucionales: 103–35.

Tena Junguito, Antonio. 1992. "Las Estadísticas Históricas Del Comercio Internacional: Fiabilidad Y Comparabilidad (1890-1960)." Banco de España. Servicios de estudios.

Treier, Shawn, and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52 (1). Wiley Online Library: 201–

17.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work*. Princeton University Press.

Voeten, Erik. 2013. "Data and Analyses of Voting in the Un General Assembly." In *Handbook of International Organization*, edited by Bob Reinalda. Routledge.

Volkens, Pola AND Matthieß, Andrea AND Lehmann. 2017. "The Manifesto Data Collection. Manifesto Project (Mrg/Cmp/Marpor). Version 2017b." Berlin: Wissenschaftszentrum Berlin für Sozialforschung. doi:[10.25522/manifesto.mpds.2017b](https://doi.org/10.25522/manifesto.mpds.2017b).

Wallace, Michael, and J. David Singer. 1970. "International Governmental Organization in the Global System, 1815-1964." *International Organization* 24 (2): 239–87.