PSC 400
SYRACUSE UNIVERSITY

DATA ANALYTICS FOR POLITICAL SCIENCE

FINDING AND CLEANING DATA

CONGRESS.CSV

Variable	Description
name	name of the congressional representative
state	state of the congressional representative
district	district number of the congressional representative
party	party of the congressional representative
congress	congressional session number
dwnom1	DW-NOMINATE score (first dimension)
dwnom2	DW-NOMINATE score (second dimension)

DATA

- **How do I find data to answer a question I have?**

DATA

- **What's the thing I'm interested in?**
  - **People: Behavior of voters, politicians, judges, etc.**
  - **Institutions: democracy, voting systems, constitutions**
  - **Policies**
  - **Events: Wars, coups, etc.**
  - **…**

DATA

- **What's my unit of analysis?**
  - **Individuals**
    - Voters, legislators, judges, etc.
  - **Aggregate units (of individuals)**
    - States, countries, etc.

DATA

- **What's the geographic area I'm interested in?**
  - Single country (or part of country)
  - Comparing several countries
  - Between countries (e.g. trade, wars)

DATA

- **What's the time frame I'm interested in?**
  - Present (or recent past)
  - More distant past
  - Development over time

SOME DATA SOURCES

- **Voters/Citizens**
  - ANES
  - Comparative Study of Electoral Systems
  - World Values Survey
  - Afro/Americas/Arab/Asian/Caucasus/Euro/Latinobarometer
- **Politicians**
  - DW-NOMINATE, DIME
  - Comparative Manifestos Data, Chapel Hill Expert Survey

SOME DATA SOURCES

- **Institutions**
  - Polity IV
  - Varieties of Democracy
  - Quality of Government
- **International Relations**
  - Correlates of War
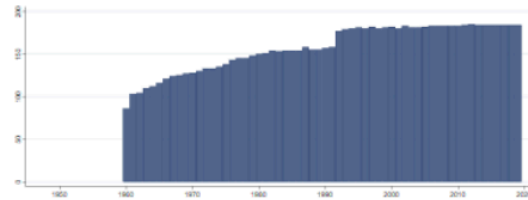 - UCDP/PRIO Armed Conflict Dataset

SOME DATA SOURCES

- https://www.gu.se/en/quality-government/qog-data

EXERCISE: QOG

4.75.67   Life expectancy at birth, total (years) (wdi_lifexp)

Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.

Min. Year:2018 Max. Year: 2018
N: 184

Min. Year:1960 Max. Year: 2019
N: 196 n: 9460 $\overline{N}$: 158 $\overline{T}$: 48

- **Explore this variable through summary statistics and graphs**

EXERCISE: QOG

- Create a new variable that measures the difference between male and female life expectancy
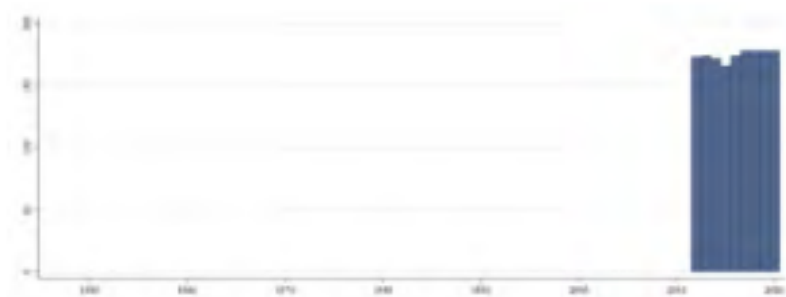- Summarize the new variable in a plot

EXERCISE: QOG

4.94.1 Corruption Perceptions Index (ti_cpi)

Corruption Perceptions Index. Scale of 0-100 where a 0 equals the highest level of perceived corruption and 100 equals the lowest level of perceived corruption.

Min. Year:2017 Max. Year: 2017
N: 178

Min. Year:2012 Max. Year: 2020
N: 178 n: 1571 $\overline{N}$: 175 $\overline{T}$: 9

- Create a new variable ti_cpi_max10 where 0 is the highest level of corruption and 10 is the lowest
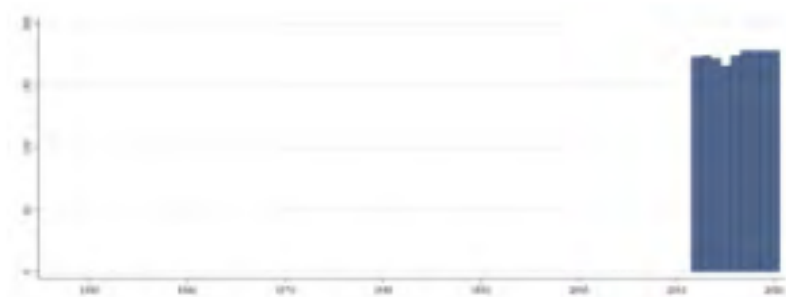
EXERCISE: QOG

4.94.1 Corruption Perceptions Index (ti_cpi)

Corruption Perceptions Index. Scale of 0-100 where a 0 equals the highest level of perceived corruption and 100 equals the lowest level of perceived corruption.

Min. Year:2017 Max. Year: 2017	Min. Year:2012 Max. Year: 2020
N: 178	N: 178 n: 1571 $\overline{N}$: 175 $\overline{T}$: 9

- Create a new variable ti_cpi_reverse where 100 is the *highest* level of corruption and 0 is the lowest
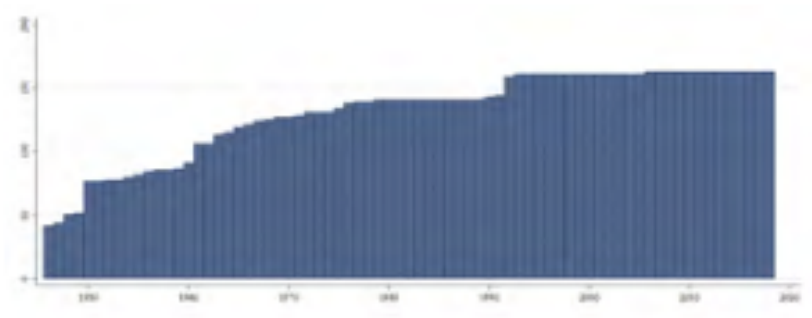
EXERCISE: QOG

4.70.1   Real GDP per Capita (mad_gdppc)

Real GDP per capita in 2011 US dollars, multiple benchmarks.

Min. Year:2017 Max. Year: 2017
N: 163

Min. Year:1946 Max. Year: 2018
N: 175 n: 9559 $\overline{N}$: 131 $\overline{T}$: 55

- Plot the density of mad_gdppc
- Create a new variable mad_gdppc_log that is the logged value of mad_gdppc (function: log)
- Plot the density of mad_gdppc_logrr

EXERCISE: QOG

4.41.1 Colonial Origin (ht_colonial)

This is a tenfold classification of the former colonial ruler of the country. Following Bernard et al. (2004), we have excluded the British settler colonies (the US, Canada, Australia, Israel and New Zealand), and exclusively focused on "Western overseas" colonialism. This implies that only Western colonizers (e.g. excluding Japanese colonialism), and only countries located in the non-Western hemisphere "overseas" (e.g. excluding Ireland & Malta), have been coded. Each country that has been colonized since 1700 is coded. In cases of several colonial powers, the last one is counted, if it lasted for 10 years or longer. The categories are the following:

0. Never colonized by a Western overseas colonial power
1. Dutch
2. Spanish
3. Italian
4. US
5. British
6. French
7. Portuguese
8. Belgian
9. British-French
10. Australian

EXERCISE: QOG

4.41.1 Colonial Origin (ht_colonial)

This is a tenfold classification of the former colonial ruler of the country. Following Bernard et al. (2004), we have excluded the British settler colonies (the US, Canada, Australia, Israel and New Zealand), and exclusively focused on "Western overseas" colonialism. This implies that only Western colonizers (e.g. excluding Japanese colonialism), and only countries located in the non-Western hemisphere "overseas" (e.g. excluding Ireland & Malta), have been coded. Each country that has been colonized since 1700 is coded. In cases of several colonial powers, the last one is counted, if it lasted for 10 years or longer. The categories are the following:

0. Never colonized by a Western overseas colonial power
1. Dutch
2. Spanish
3. Italian
4. US
5. British
6. French
7. Portuguese
8. Belgian
9. British-French
10. Australian

- Create a variable "colonized" that is 1 if the country was every colonized by a Western colonial power, and 0 if not

EXERCISE: QOG

4.41.1 Colonial Origin (ht_colonial)

This is a tenfold classification of the former colonial ruler of the country. Following Bernard et al. (2004), we have excluded the British settler colonies (the US, Canada, Australia, Israel and New Zealand), and exclusively focused on "Western overseas" colonialism. This implies that only Western colonizers (e.g. excluding Japanese colonialism), and only countries located in the non-Western hemisphere "overseas" (e.g. excluding Ireland & Malta), have been coded. Each country that has been colonized since 1700 is coded. In cases of several colonial powers, the last one is counted, if it lasted for 10 years or longer. The categories are the following:

0. Never colonized by a Western overseas colonial power
1. Dutch
2. Spanish
3. Italian
4. US
5. British
6. French
7. Portuguese
8. Belgian
9. British-French
10. Australian

- Create a variable "colonized2" that is "colonized" if the country was every colonized by a Western colonial power, and "not colonized" if not