

PSC 400

SYRACUSE UNIVERSITY

# **DATA ANALYTICS FOR POLITICAL SCIENCE**

**FINDING AND CLEANING DATA**

# ASSIGNMENTS

- Review Exercise 3 due today
- Data Analysis Memo due on Friday
- Will post Problem Set 2 soon
  - due next week Friday

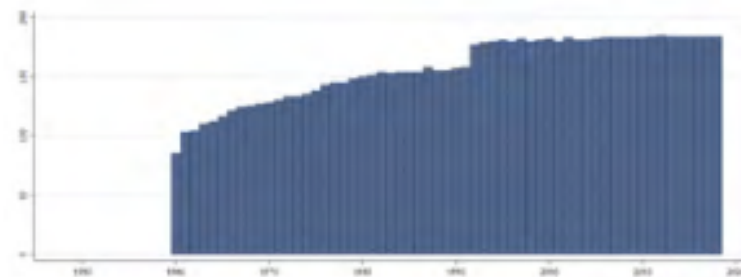
# EXERCISE: QOG

## 4.104.178 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:2017 Max. Year: 2017  
N: 184



Min. Year:1960 Max. Year: 2018  
N: 196 n: 9276  $\bar{N}$ : 157  $\bar{T}$ : 47

- Explore this variable through summary statistics and graphs

## EXERCISE: QOG

- `wdi_lifexpm`: male life expectancy
- `wdi_lifexpf`: female life expectancy
- Create a new variable that measures the difference between male and female life expectancy
- Plot the new variable

## EXERCISE: QOG

- **wdi\_lifexpm**: male life expectancy
- **wdi\_lifexpf**: female life expectancy
- Create a new variable that is the average of **wdi\_lifexpm** and **wdi\_lifexpf**
- Plot the new variable

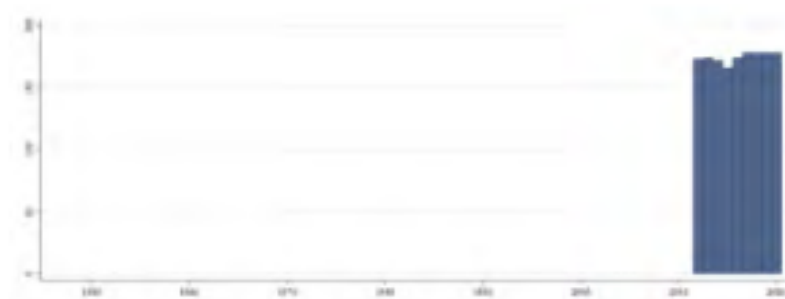
# EXERCISE: QOG

## 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  $\bar{N}$ : 175  $\bar{T}$ : 9

- Create a new variable **ti\_cpi\_max10** where 0 is the highest level of corruption and 10 is the lowest

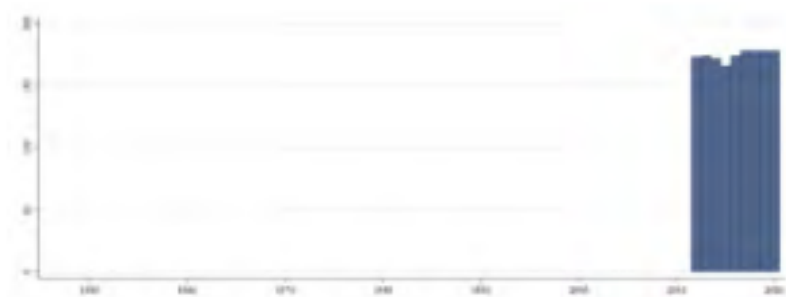
# EXERCISE: QOG

## 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  
N: 178



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

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

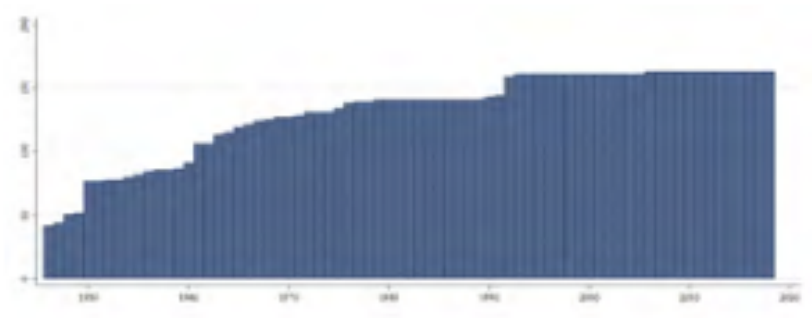
# EXERCISE: QOG

## 4.70.1 Real GDP per Capita (mad\_gdpppc)

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  $\bar{N}$ : 131  $\bar{T}$ : 55

- Plot the density of mad\_gdpppc
- Create a new variable mad\_gdpppc\_log that is the logged value of mad\_gdpppc
- Plot the density of mad\_gdpppc\_log



# EXERCISE: QOG

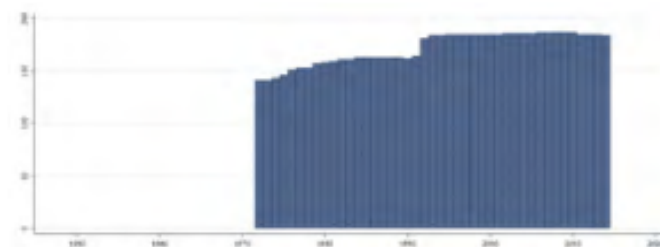
## 4.54.5 Regime Type (simplified) (ht\_regtype1)

A simplified, collapsed version of ht\_regtype, where all monarchical regimes with amalgams [ht\_regtype = 16, 17, 23 or 24] are treated as monarchies, all military regimes with sub-types and amalgams [ht\_regtype = 4, 5, 6, 7 or 18] are treated as military regimes, and multiparty regimes with sub-types are treated as multiparty regimes [ht\_regtype = 1 or 2]. Only pure noparty [ht\_regtype = 3] and one-party [ht\_regtype = 8] regimes are treated as no-party and one-party regimes, respectively. The minor types [ht\_regtype = 9, 19, 20, 21, 22 or 25] are treated as other.

1. Monarchy
2. Military
3. One party
4. Multi-party
9. No-party
99. Other
100. Democracy



Min. Year: 2014 Max. Year: 2014  
N: 184



Min. Year: 1972 Max. Year: 2014  
N: 199 n: 7390  $\bar{N}$ : 172  $\bar{T}$ : 37

# EXERCISE: QOG

## 4.54.5 Regime Type (simplified) (ht\_regtype1)

A simplified, collapsed version of ht\_regtype, where all monarchical regimes with amalgams [ht\_regtype = 16, 17, 23 or 24] are treated as monarchies, all military regimes with sub-types and amalgams [ht\_regtype = 4, 5, 6, 7 or 18] are treated as military regimes, and multiparty regimes with sub-types are treated as multiparty regimes [ht\_regtype = 1 or 2]. Only pure noparty [ht\_regtype = 3] and one-party [ht\_regtype = 8] regimes are treated as no-party and one-party regimes, respectively. The minor types [ht\_regtype = 9, 19, 20, 21, 22 or 25] are treated as other.

1. Monarchy
2. Military
3. One party
4. Multi-party
9. No-party
99. Other
100. Democracy

- Create a new factor variable called dem that is:
  - "Democracy" if a country is a democracy
  - "Not a democracy" otherwise
- Create a new dummy variable called dem.bin that is:
  - 1 if a country is a democracy
  - 0 otherwise
- Describe the variable using a graph