



POLS 285

LAB 2

15 January 2021

TA: Mathieu Dubeau

Agenda



Assignment One



Data Cleaning Best Practices



**Application to World Development Indicators
(WDI) Dataset**



Introduction of Polity V Dataset



Principles of Merging/Joining Data



ASSIGNMENT 1

Data Cleaning Best Practices

The objective of data cleaning is **to fix any data that is incorrect, inaccurate, incomplete, incorrectly formatted, duplicated, or even irrelevant to the objective of the data set.**

1. Data should be **accurate**
2. Data needs to be **consistent**
3. Data should be **valid**
4. Data should be **complete**
5. Data should be **uniform**

Things to watch for:

Duplicate data: There are 2 or more identical records. This may cause misrepresentation.

Conflicting Data: When there are same records with different attributes, it means data is conflicting.

Incomplete Data: The data that has missing attributes. Because the social world is complex, this occurs frequently. Coding missing values correctly is vital to obtaining accurate results.

Invalid Data: Data attributes are not conforming to standardization. For example, the difference between absolute values and percentages. Each variable must use a uniform unit of measurement

Let's dive in!
