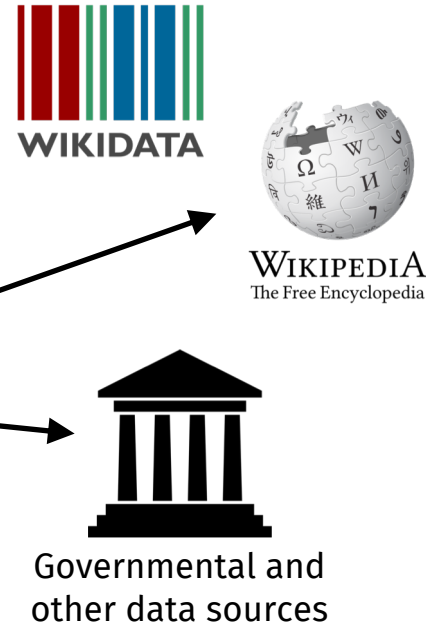


Step 1 Entity identification



Step 2 Data collection



Step 3 Cleaning, verification

name	sex	ethnicity	religion	birth
Ben Cardin	male	white	judaism	1943-10-05
Amy Klobuchar	female	white	protestantism reformed	1960-05-25
Roger Wicker	male	white	protestantism baptist	1951-07-05
Claire McCaskill	female	white	catholicism	1953-07-24
Jon Tester	male	white	protestantism evangelical	1956-08-21
Sherrod Brown	male	white	protestantism lutheran	1952-11-09
Bob Casey, Jr.	male	white	catholicism	1960-04-13
Sheldon Whitehouse	male	white	NA	1955-10-20

```
str_replace(france_political$party,  
"a\\.C)?|\\(app\\.\\| \\|Les
```

```
usas_political$party == "Anti-J", "AJ"  
ifelse(usas_political$party == "Anti-M  
ifelse(usas_political$pa
```

Step 4 Database integration



- A** Identify and scrape parliamentary members' index sites
- B** Extract person-specific Wikipedia and Wikidata IDs

- A** Scrape basic information from members' index sites
- B** Automatically retrieve bulk of information using IDs with Wikidata and Wikipedia APIs
- C** Scrape external sources for additional and missing data

- A** Remove noise in variable values
- B** Harmonize variables within and across legislatures
- C** Perform checks against alternative data sources

- A** Assemble database via Wikipedia and Wikidata IDs
- B** Provide convenient access via R package legislatoR