

Biography of Congress
1917 – 1980

Web scrape, collect:
Name, Party, DOB

Propublica API
1980 – 2019

API call, collect:
Name, Party, Gender, DOB

Raw dataframe
~ 28,000 rows
10,000 rows with gender
assigned

Use DOB and first name to
determine gender

R package **gender**

Dataframe
~ 27,500 rows with gender
assigned

Drop the ~500 rows with
N/A gender

Plot!