

DATA 606 - Final Project

Joshua Sturm

December 15th, 2017

<https://www.kaggle.com/shoklan/which-year-produced-the-most/notebook> <https://github.com/fivethirtyeight/data/blob/master/congress-age/congress-terms.csv> http://online.wsj.com/public/resources/documents/info-CONGRESS_AGES_1009.html

Part 1: Introduction

Is there a relationship between the average age of congress (members) and the number of bills proposed?

The average age of congressional representatives has been steadily climbing since the second world war. The current (115th) one is among the oldest in its history. How has this affected the effectiveness of congress? Are older more representatives more or less active?

I plan to explore this via proxy, by taking a look at all the *constitutional amendments* proposed since the first congress through the 113th, and recording the age of each of the bill's sponsors. Additionally, I will seek any interesting tidbits in the data, such as the most active years, as well as which state representatives propose the most legislation.

Part 2: Data

Data collection

The amendment list was retrieved from Kaggle, while the members list was taken from FiveThirtyEight. Another source is from the Wall Street Journal.

The list of 11,000+ amendments was compiled by staff and volunteers of the National Archives and Records Administration. The list of representatives was compiled by The UnitedStates Project (House members), and The New York Times Congress API (senate).

Cases

Each case represents a constitutional amendment proposed by congress. There are a total of 11797 cases in this dataset.

Variables

The response variable is legislative activity and is numerical.

The explanatory variable is median age of congressional representatives and is numerical.

Type of study

This is an observational study.

Scope of inference - generalizability

Scope of inference - causality:

Part 3: Exploratory data analysis

```
library(tidyverse)
library(ggplot2)

#
# Load the files from the working directory
#
members_raw <- read.csv("congress_terms.csv")

#
# Tidy the datasets
#
# Keep only the relevant columns
amendments <- amendments_raw %>%
  select(5, 7:ncol(amendments_raw)-1, -6)

# Use regex to shift errant data to their appropriate columns
for (i in 1:(length(amendments$year))) {
  pat <- "\\D{3,}"
  if (grepl(pat, amendments[i, "month"]) == 1)
  {
    amendments[i, "year"] <- amendments[i, "month"]
    amendments[i, "month"] <- amendments[i, "day"]
    amendments[i, "day"] <- amendments[i, "congress"]
    amendments[i, "congress"] <- amendments[i, "congressional_session"]
    amendments[i, "congressional_session"] <- amendments[i, "joint_resolution_chamber"]
  }
}
amendments$year <- gsub("\\D{4}$", "", amendments$year)

members <- members_raw %>%
  select(-c(3,7))
```

Let's take a look at what the data has to say.

```
#
# Which years had the most bills?
#
ggplot(data=amendments, aes(x=year, fill=year)) +
  geom_bar()
```

1788	1809	1820	1840	1871	1890	1909	1920	1947	1960	1980
1789	1810	1829	1849	1872	1891	1910	1929	1948	1967	1986
1790	1811	1830	1850	1873	1892	1911	1930	1949	1968	1987
1791	1812	1831	1851	1874	1893	1912	1931	1950	1969	1988
1793	1813	1832	1852	1875	1894	1913	1932	1951	1970	1989
1794	1814	1833	1853	1876	1895	1914	1933	1952	1971	1990
1795	1815	1834	1854	1877	1896	1915	1934	1953	1972	1991
1797	1816	1835	1858	1878	1897	1916	1935	1954	1973	1992
1798	1817	1836	1860	1879	1898	1917	1936	1955	1974	1993
1799	1818	1837	1861	1880	1899	1918	1937	1956	1975	19931
1800	1819	1838	1862	1881	1900	1919	1938	1957	1976	1994
1801	1820	1839	1863	1882	1901	1920	1939	1958	1977	1995
1802	1821	1840	1864	1883	1902	1921	1940	1959	1978	1996
1803	1822	1841	1865	1884	1903	1922	1941	1960	1979	1997
1804	1823	1842	1866	1885	1904	1923	1942	1961	1980	1998
1805	1824	1843	1867	1886	1905	1924	1943	1962	1981	1999
1806	1825	1844	1868	1887	1906	1925	1944	1963	1982	2000
1807	1826	1845	1869	1888	1907	1926	1945	1964	1983	2001
1808	1827	1846	1870	1889	1908	1927	1946	1965	1984	2002

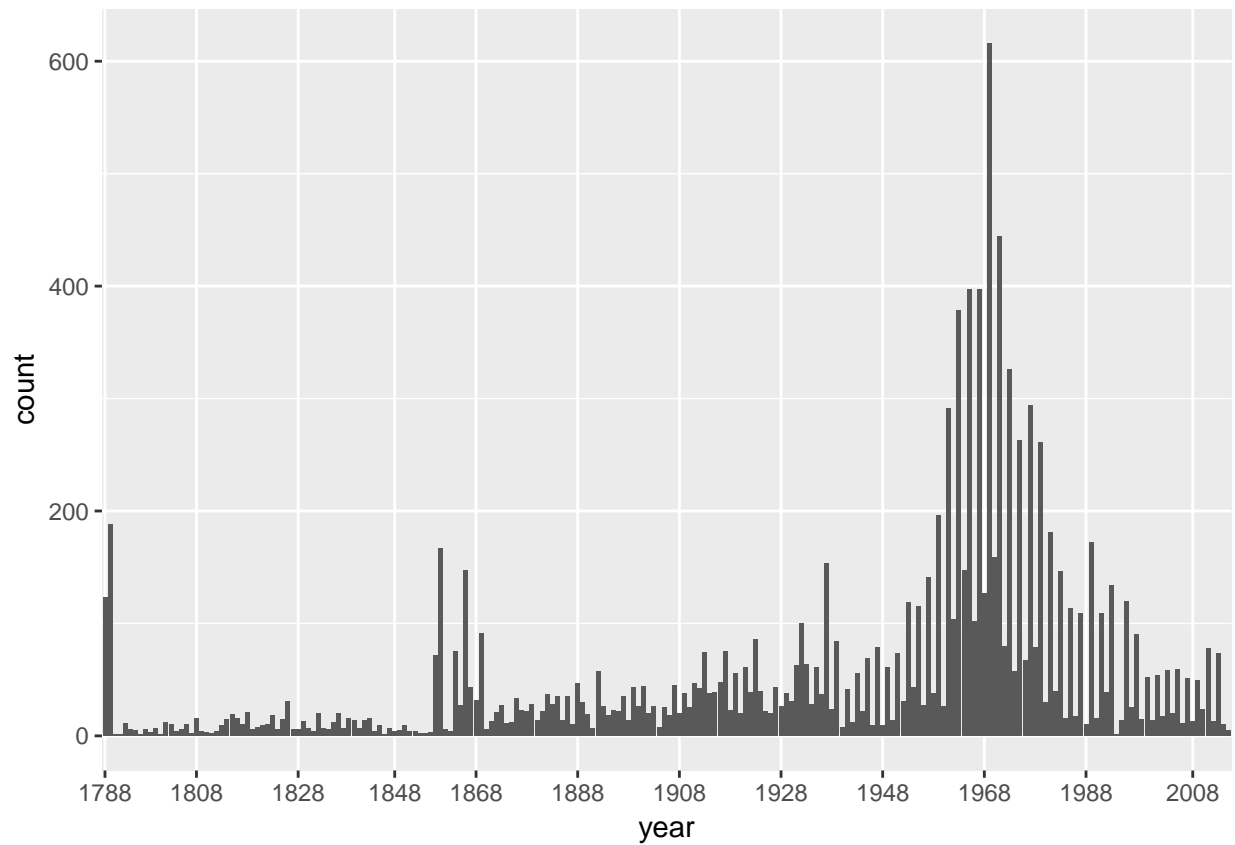
Part 4: Inference

Conclusion

Relevant summary statistics

```
lvls <- levels(amendments$year)

ggplot(amendments, aes(year)) +
  geom_bar() +
  scale_x_discrete(breaks=seq(1788, 2014, 20))
```



There is a noticeable spike in amendmendt proposals in the latter half of the 20th century. My guess would be that it's related to the civil rights movement, and the sweeping changes that came about.