

explore.R

nacnudus

Fri Nov 3 15:58:10 2017

```
# Explore data scraped from legislation.gov.uk
```

```
library(tidyverse)
library(stringr)
library(here)
```

```
leg <- read_tsv(here("lists", "scraped.tsv"))
```

```
## Parsed with column specification:
## cols(
##   name = col_character(),
##   `year-and-number` = col_character(),
##   type = col_character(),
##   slug = col_character()
## )
```

```
# Are all the years readsonable, and the distribution?
```

```
years <-
  leg %>%
  mutate(year = as.integer(str_sub(`year-and-number`, 1L, 4L)))
```

```
# The years look completely reasonable
```

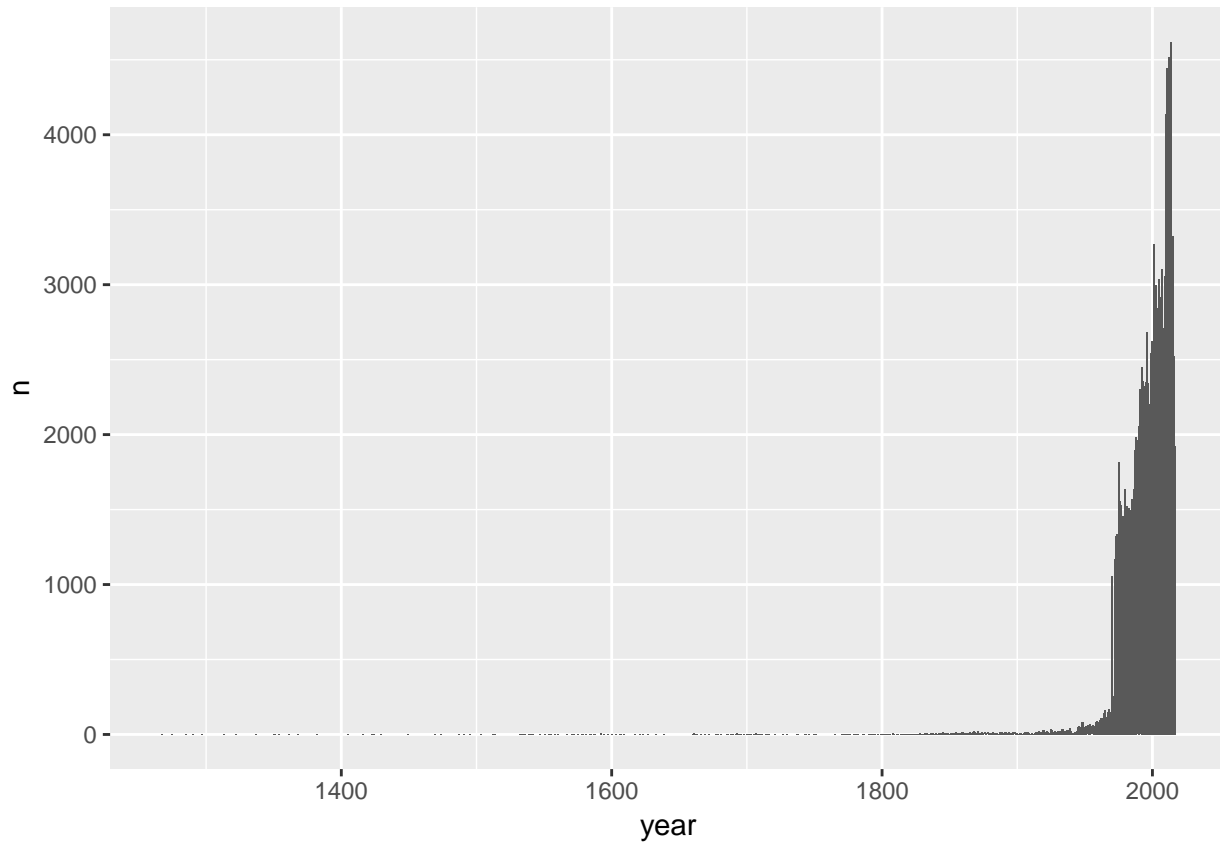
```
sort(unique(years$year))
```

```
## [1] 1267 1275 1285 1290 1297 1313 1322 1337 1350 1351 1354 1361 1368 1382
## [15] 1405 1416 1423 1424 1429 1449 1469 1474 1487 1491 1495 1503 1512 1514
## [29] 1532 1533 1534 1535 1536 1539 1540 1542 1547 1551 1554 1555 1558 1560
## [43] 1567 1571 1572 1575 1578 1579 1581 1584 1585 1587 1588 1592 1594 1597
## [57] 1600 1603 1606 1607 1609 1617 1621 1623 1627 1633 1639 1660 1661 1662
## [71] 1663 1666 1669 1672 1677 1679 1681 1685 1686 1688 1689 1690 1692 1693
## [85] 1694 1695 1696 1698 1700 1701 1702 1703 1705 1706 1707 1708 1709 1710
## [99] 1711 1714 1715 1716 1721 1727 1730 1737 1738 1743 1744 1746 1750 1751
## [113] 1765 1770 1771 1772 1773 1774 1775 1776 1779 1780 1781 1782 1786 1787
## [127] 1790 1792 1793 1795 1797 1799 1800 1801 1802 1803 1804 1805 1806 1808
## [141] 1809 1811 1812 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824
## [155] 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838
## [169] 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852
## [183] 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866
## [197] 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880
## [211] 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894
## [225] 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908
## [239] 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922
## [253] 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936
## [267] 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950
## [281] 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964
## [295] 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978
## [309] 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992
## [323] 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006
```

```
## [337] 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
```

```
# There's precious little until the C21st, especially before the C19th.
```

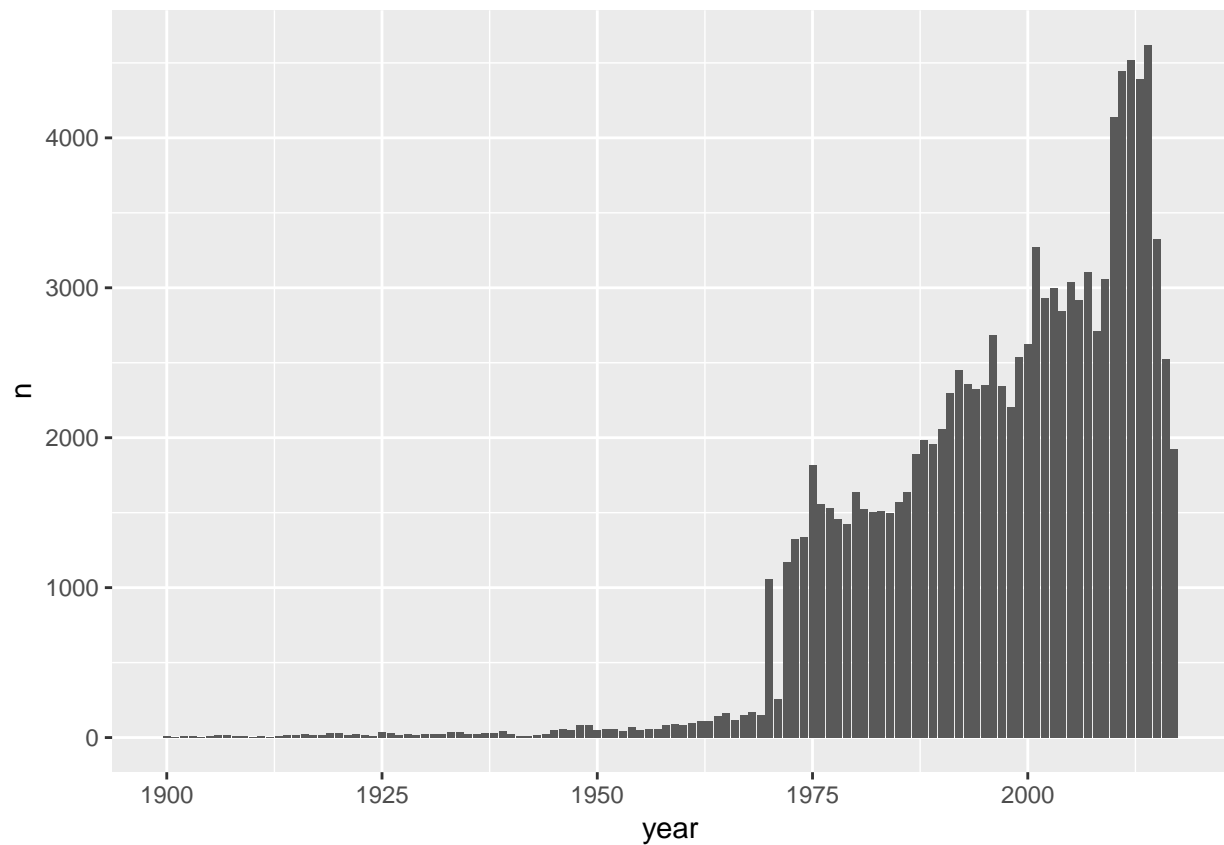
```
years %>%  
  count(year, sort = TRUE) %>%  
  ggplot(aes(year, n)) +  
  geom_bar(stat = "identity")
```



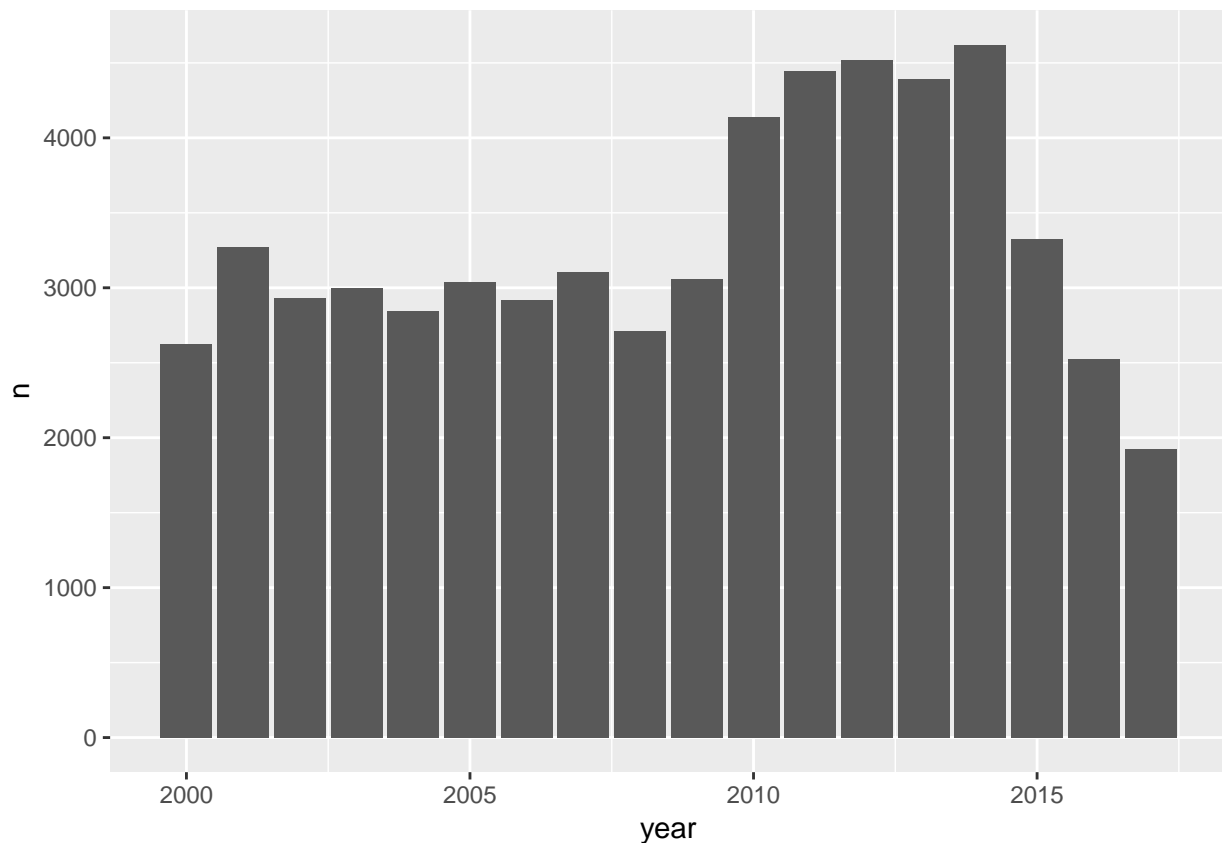
```
# It really only kicks off in 1972, with another big year in 1970.
```

```
# There's a drop in the last three years -- is there a lag? Is legislation  
# dated by the year it was written, rather than the year it was passed?
```

```
years %>%  
  count(year, sort = TRUE) %>%  
  filter(year >= 1900) %>%  
  ggplot(aes(year, n)) +  
  geom_bar(stat = "identity")
```



```
# There was noticeably more legislation from 2010 to 2014.  
years %>%  
  count(year, sort = TRUE) %>%  
  filter(year >= 2000) %>%  
  ggplot(aes(year, n)) +  
  geom_bar(stat = "identity")
```



```
# What formats of year and number are there? A lot
# This looks for the strings other than the year and the number
split_year_number <- function(x) {
  unlist(str_split(x, "\\s", n = Inf))
}

x <-
  leg %>%
  mutate(substring = map(`year-and-number`, split_year_number)) %>%
  unnest() %>%
  group_by(slug) %>%
  mutate(id = row_number()) %>%
  ungroup()

x %>%
  filter(!(id %in% c(1, 3))) %>%
  count(id, substring, sort = TRUE)
```

```
## # A tibble: 684 x 3
##       id substring      n
##   <int> <chr>      <int>
## 1     2 No.      111173
## 2     4 (C.       5862
## 3     4 (W.       5815
## 4     2 c.       4960
## 5     4 (S.       4921
## 6     4 (L.      1272
```

```
## 7      4 (N.I.      624
## 8      6 (W.      460
## 9      2 Chapter    289
## 10     2 asp       265
## # ... with 674 more rows
```

```
x %>%
  group_by(slug) %>%
  summarise(n_substrings = max(id)) %>%
  count(n_substrings, sort = TRUE)
```

```
## # A tibble: 8 x 2
##   n_substrings     n
##         <dbl> <int>
## 1           3.00 98235
## 2           5.00 17606
## 3           7.00   937
## 4           1.00    10
## 5           6.00     6
## 6           8.00     6
## 7           9.00     4
## 8          10.0     3
```

```
x %>%
  group_by(slug) %>%
  filter(max(id) == 5) %>%
  ungroup() %>%
  slice(1:10) %>%
  select(`year-and-number`, slug)
```

```
## # A tibble: 10 x 2
##   `year-and-number` slug
##   <chr>             <chr>
## 1 2017 No. 1055 (C. 99) /uksi/2017/1055/contents/made
## 2 2017 No. 1055 (C. 99) /uksi/2017/1055/contents/made
## 3 2017 No. 1055 (C. 99) /uksi/2017/1055/contents/made
## 4 2017 No. 1055 (C. 99) /uksi/2017/1055/contents/made
## 5 2017 No. 1055 (C. 99) /uksi/2017/1055/contents/made
## 6 2017 No. 1052 (C. 98) /uksi/2017/1052/contents/made
## 7 2017 No. 1052 (C. 98) /uksi/2017/1052/contents/made
## 8 2017 No. 1052 (C. 98) /uksi/2017/1052/contents/made
## 9 2017 No. 1052 (C. 98) /uksi/2017/1052/contents/made
## 10 2017 No. 1052 (C. 98) /uksi/2017/1052/contents/made
```

```
# How many titles are SHOUTY CAPS? Only 32 after all
leg %>%
  filter(name == toupper(name)) %>%
  print(n = Inf)
```

```
## # A tibble: 42 x 4
##   name                                `year-and- type      slug
##   <chr>                                <chr>      <chr>    <chr>
## 1 THE ELECTRICITY SAFETY, QUALITY AN~ 2012 No. ~ Northern Ir~ /nistr/2012~
## 2 THE CARDIFF TO GLAN CONWY TRUNK RO~ 2012 No. ~ Wales Statu~ /wsi/2012/~
## 3 GORCHYMYN CEFNFFORDD CAERDYDD I LA~ 2012 No. ~ Wales Statu~ /wsi/2012/~
## 4 THE AGENCY WORKERS REGULATIONS (NO~ 2011 No. ~ Northern Ir~ /nistr/2011~
```

```
## 5 THE A46 TRUNK ROAD (M40 JUNCTION 1~ 2010 No. ~ UK Statutor~ /uksi/2010~
## 6 THE A35 TRUNK ROAD (WINTERBOURNE A~ 2010 No. ~ UK Statutor~ /uksi/2010~
## 7 THE LONDON TO FISHGUARD TRUNK ROAD~ 2009 No. ~ Wales Statu~ /wsi/2009/~
## 8 GORCHYMYN CEFNFFORDD LLUNDAIN I AB~ 2009 No. ~ Wales Statu~ /wsi/2009/~
## 9 THE POLICE PENSION (NORTHERN IRELA~ 2009 No. ~ Northern Ir~ /nistr/2009~
## 10 THE CHILDCARE ACT 2006 (COMMENCEME~ 2008 No. ~ Wales Statu~ /wsi/2008/~
## 11 GORCHYMYN DEDDF GOFAL PLANT 2006 (~ 2008 No. ~ Wales Statu~ /wsi/2008/~
## 12 EQUALITY ACT (SEXUAL ORIENTATION) ~ 2006 No. ~ Northern Ir~ /nistr/2006~
## 13 POLICE SERVICE OF NORTHERN IRELAND~ 2006 No. ~ Northern Ir~ /nistr/2006~
## 14 THE RAILWAYS (SAFETY MANAGEMENT) R~ 2006 No. ~ Northern Ir~ /nistr/2006~
## 15 THE COMMONHOLD (LAND REGISTRATION)~ 2004 No. ~ UK Statutor~ /uksi/2004~
## 16 THE STATISTICS OF TRADE (CUSTOMS A~ 2003 No. ~ UK Statutor~ /uksi/2003~
## 17 THE FRIENDLY SOCIETIES (GENERAL CH~ 1999 No. ~ UK Statutor~ /uksi/1999~
## 18 THE A12 TRUNK ROAD (COLCHESTER ROA~ 1995 No. ~ UK Statutor~ /uksi/1995~
## 19 THE DARTFORD-THURROCK CROSSING (AM~ 1995 No. ~ UK Statutor~ /uksi/1995~
## 20 THE INDUSTRIAL TRIBUNALS (ENFORCEM~ 1995 No. ~ UK Statutor~ /uksi/1995~
## 21 THE LEGAL AID ADVISORY COMMITTEE(D~ 1995 No. ~ UK Statutor~ /uksi/1995~
## 22 THE COURT OF PROTECTION RULES 1994 1994 No. ~ UK Statutor~ /uksi/1994~
## 23 THE INDUSTRIAL ASSURANCE (PERMIUM ~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 24 THE ISLES OF SCILLY (LOCAL GOVERN~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 25 THE NATIONAL INSURANCE AND CIVIL S~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 26 THE EXTINGUISHMENT OR MODIFICATION~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 27 THE TOWN AND COUNTRY PLANNING (ENF~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 28 THE NATIONAL ASSISTANCE (COMPENSAT~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 29 THE NATIONAL ASSISTANCE (POWERS OF~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 30 THE CIVIL AVIATION (BIRTHS, DEATHS~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 31 THE STOPPING UP OF HIGHWAYS (CONCU~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 32 THE TOWN AND COUNTRY PLANNING (TRA~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 33 THE LOCAL GOVERNMENT SUPERANNUATIO~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 34 THE TOWN AND COUNTRY PLANNING (ENF~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 35 THE FACTORIES ACT, 1937 (EXTENSION~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 36 THE FIRE SERVICES (PENSIONABLE EMP~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 37 THE COMPENSATION (DEFENCE) NOTICE ~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 38 THE TRADING WITH THE ENEMY (ENEMY ~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 39 THE TRADING WITH THE ENEMY (ENEMY ~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 40 THE TRADING WITH THE ENEMY (ENEMY ~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 41 THE TRADING WITH THE ENEMY (ENEMY ~ 1948 No. ~ UK Statutor~ /uksi/1948~
## 42 THE TRADING WITH THE ENEMY (ENEMY ~ 1948 No. ~ UK Statutor~ /uksi/1948~
```

```
# How many titles begin with "the"? Most of them
```

```
leg %>%
  mutate(name = tolower(name),
         the = str_detect(name, "^the")) %>%
  count(the, sort = TRUE)
```

```
## # A tibble: 3 x 2
```

```
##   the      n
##   <lgl> <int>
## 1 T     102246
## 2 F      14567
## 3 NA         2
```

```
# Are the legislation types reasonable?
```

```
count(leg, type, sort = TRUE) %>%
  print(n = Inf)
```

```
## # A tibble: 22 x 2
##   type                n
##   <chr>              <int>
## 1 UK Statutory Instruments 77727
## 2 Northern Ireland Statutory Rules 15991
## 3 Wales Statutory Instruments 8578
## 4 Scottish Statutory Instruments 7705
## 5 UK Public General Acts 4270
## 6 Northern Ireland Orders in Council 624
## 7 UK Statutory Rules and Orders 307
## 8 Acts of the Northern Ireland Parliament 287
## 9 Acts of the Scottish Parliament 265
## 10 UK Local Acts 260
## 11 Acts of the Northern Ireland Assembly 173
## 12 Church Measures 133
## 13 Acts of the Old Scottish Parliament 109
## 14 Acts of the English Parliament 79
## 15 Acts of the National Assembly for Wales 64
## 16 UK Ministerial Orders 62
## 17 Church Instruments 60
## 18 Acts of the Parliament of Great Britain 58
## 19 Measures of the National Assembly for Wales 44
## 20 Acts of the Old Irish Parliament 11
## 21 Northern Ireland Statutory Rules and Orders 6
## 22 Measures of the Northern Ireland Assembly 2
```

```
# What's the distribution of types of legislation? Not much from Scotland
years %>%
  filter(year >= 1970) %>%
  ggplot(aes(year, fill = type)) +
  geom_bar(stat = "count", position = "stack")
```

