Tidy Tuesday: NBER Publication Data

Kesava Asam

Report Last Run: 2021-10-03 20:51:36

Contents

1 Packages **3**

2 Data **3**

3 Function **7**

4 Subset **8**

4.1 Sex related 8

4.2 Health related 8

4.3 Retirement 8

4.4 Education 8

5 Defense **8**

6 Pollution **8**

6.1 Combine 9

7 Visualization **10**

8 References: **12**

1 Packages

```
# load required packages
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tidytuesdayR)
library(ggdark)
theme_set(theme_light())
```

2 Data

```
# read in the data manually
papers <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/20
```

```
## Rows: 29434 Columns: 4

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): paper, title
## dbl (2): year, month

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
programs <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/
```

```
## Rows: 21 Columns: 3

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): program, program_desc, program_category

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
paper_programs <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/
```

```
## Rows: 53996 Columns: 2
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): paper, program
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

NBR Publication data

Data set for graphs

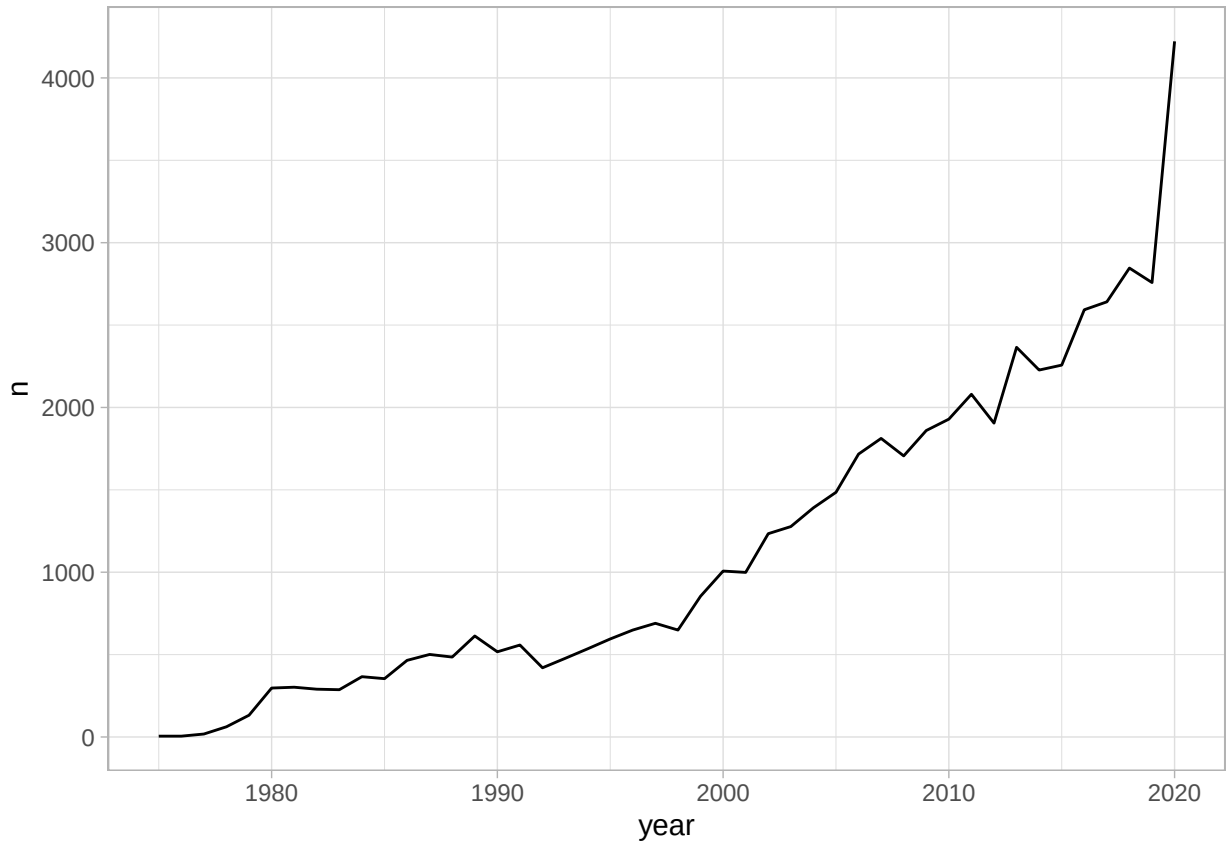
```
pub_data <-
  inner_join(papers, paper_programs) %>%
  inner_join(programs)
```

```
## Joining, by = "paper"
```

```
## Joining, by = "program"
```

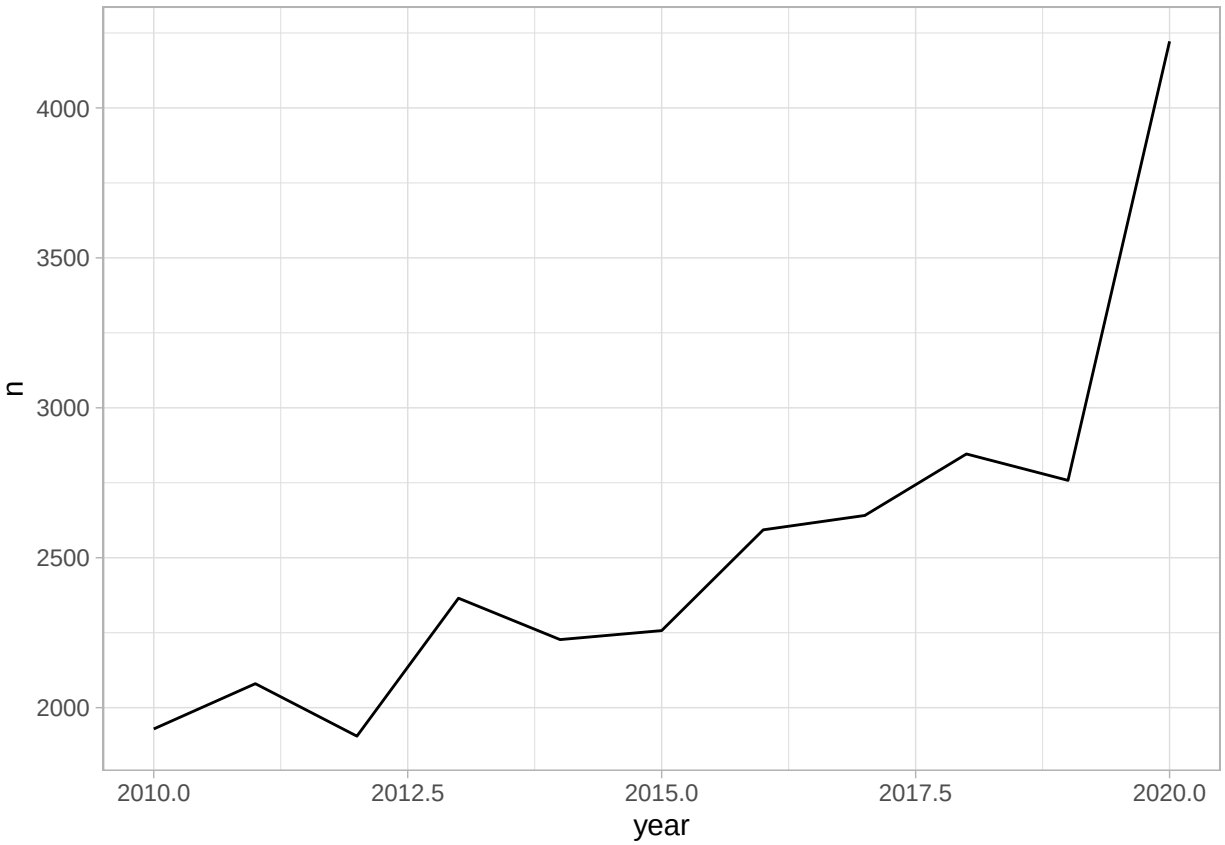
Check the data

```
pub_data %>%
  count(year) %>% filter(year != 2021) %>% # 2021 is not complete
  ggplot(aes(x = year, y = n)) +
  geom_line()
```



Check recent 10 years per month submissions?

```
pub_data %>%
  count(year) %>% filter(year > 2009 & year != 2021) %>%
  ggplot(aes(x = year, y = n)) +
  geom_line()
```


4000
3500
3000
2500
2000
n
2010.0
2012.5
2015.0
2017.5
2020.0
year

3 Function

Make a function to pick keywords

```
pick_title <- function(term_int) {

  pub_data %>%
    filter(year != 2021 & year > 1989) %>%
    distinct(title, .keep_all = T) %>%
    filter(grepl({{term_int}}, title)) %>%
    group_by(year) %>%
    mutate(count = n()) %>%
    ungroup()
}
```

4 Subset

4.1 Sex related

```
sex_df <-
  pick_title('Gender|Sex |Male|Female|Girls|Boys|Women|Men ') %>%
  mutate('category' = "Gender")
```

4.2 Health related

```
health_df <-
  pick_title('Health|fitness') %>%
  mutate('category' = "Health")
```

4.3 Retirement

```
retire_df <-
  pick_title('Retirement') %>%
  mutate('category' = "Retirement")
```

4.4 Education

```
edu_df <-
  pick_title('Education|University|School') %>%
  mutate('category' = "Education")
```

5 Defense

```
def_df <-
  pick_title('Defense|Army|Military|Navy') %>%
  mutate('category' = "Defense")
```

6 Pollution

```
poll_df <-
  pick_title('Pollution|Climate|Recycling|Hazardous') %>%
  mutate('category' = "Pollution/Climate")
```

6.1 Combine

```
df_cat_int <- rbind(health_df, edu_df, retire_df,
                    sex_df, def_df, poll_df)
```

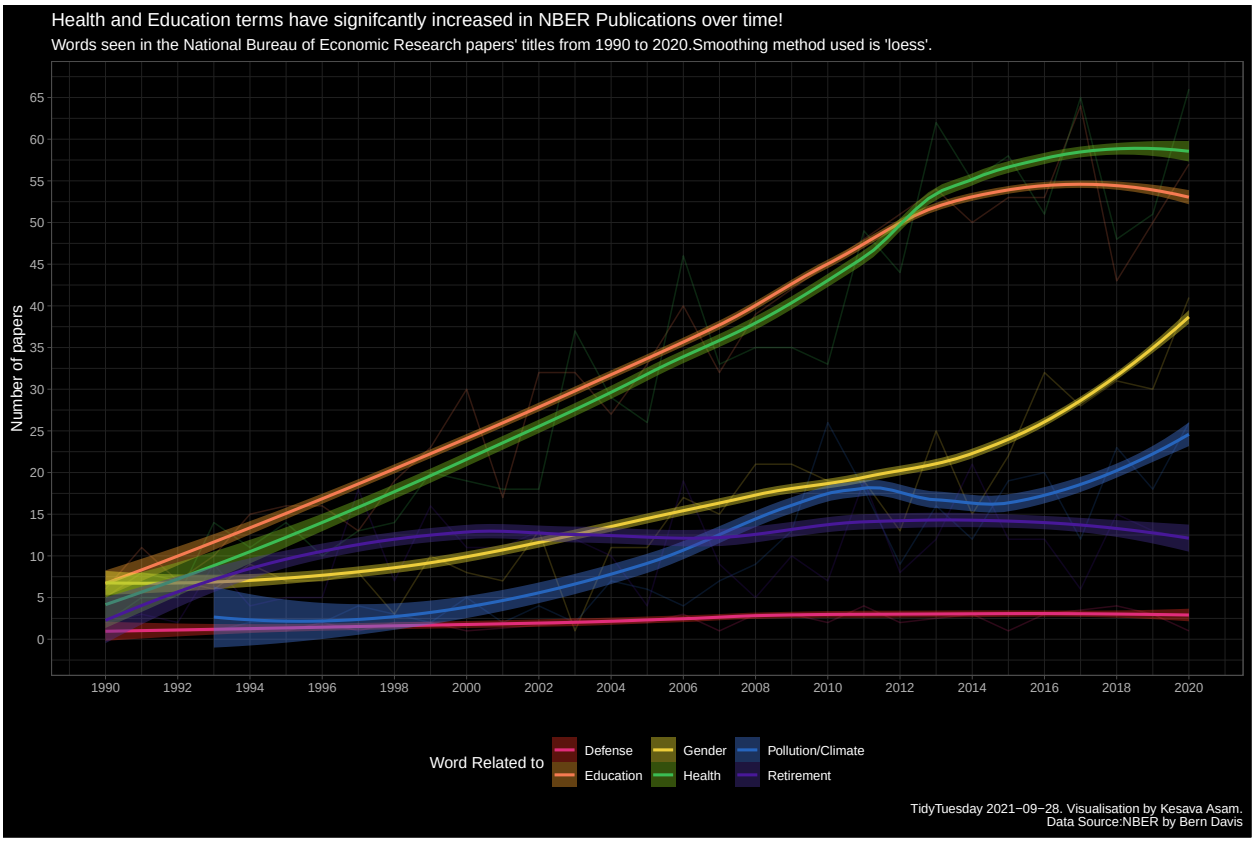
7 Visualization

```
# Custom colors
colrs_1 <- c("#E83122", "#FAA42E", "#FAEB36", "#82CA20", "#487DE7", "#4B369D")
colrs_2 <- c('#E12D7B','#F67B52', "#EDCD3B", "#3BBC54", "#2665BD", "#481899")

# Plot
p <- df_cat_int %>%
  ggplot(aes(x = year, y = count, color = category, fill = category)) +
  geom_line(alpha = 0.20) +
  geom_smooth(method = 'loess', formula = 'y ~ x')+
  scale_color_manual(values=colrs_2)+
  scale_fill_manual(values=colrs_1)+
  scale_x_continuous(breaks = seq(1990, 2020, 2)) +
  scale_y_continuous(breaks = seq(0, 70, 5)) +
  labs(title = "Health and Education terms have signifcantly increased in NBER Publications over time!"
    subtitle= "Words seen in the National Bureau of Economic Research papers' titles from 1990 to 2020.S
    caption = "TidyTuesday 2021-09-28. Visualisation by Kesava Asam.\n Data Source:NBER by Bern Davis",
       x= "", y = "Number of papers",
       fill = "Word Related to",
       color = "Word Related to")
```

```
# Using the dark_mode from ggdark by nsgrantham
p + dark_mode() +
  theme(legend.position = "bottom")
```

```
## Inverted geom defaults of fill and color/colour.
## To change them back, use invert_geom_defaults().
```


Health and Education terms have signifcantly increased in NBER Publications over time!
Words seen in the National Bureau of Economic Research papers' titles from 1990 to 2020.Smoothing method used is 'loess'.
Number of papers
0
5
10
15
20
25
30
35
40
45
50
55
60
65
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
Word Related to
Defense
Education
Gender
Health
Pollution/Climate
Retirement
TidyTuesday 2021−09−28. Visualisation by Kesava Asam.
Data Source:NBER by Bern Davis

8 References:

- [Color palletes](https://www.schemecolor.com/kaleidoscope-rainbow.php)
- [GGDark](https://github.com/nsgrantham/ggdark)

```
sessionInfo()
```

```
## R version 4.1.1 (2021-08-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] ggdark_0.2.1       tidytuesdayR_1.0.1 forcats_0.5.1      stringr_1.4.0
##  [5] dplyr_1.0.7        purrr_0.3.4        readr_2.0.2        tidyr_1.1.4
##  [9] tibble_3.1.5       ggplot2_3.3.5      tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.7       lattice_0.20-44  lubridate_1.7.10 assertthat_0.2.1
##  [5] digest_0.6.28    utf8_1.2.2       R6_2.5.1         cellranger_1.1.0
##  [9] backports_1.2.1  reprex_2.0.1     evaluate_0.14    highr_0.9
## [13] httr_1.4.2       pillar_1.6.3     rlang_0.4.11     curl_4.3.2
## [17] readxl_1.3.1     rstudioapi_0.13  Matrix_1.3-4     rmarkdown_2.11
## [21] splines_4.1.1    labeling_0.4.2   bit_4.0.4        munsell_0.5.0
## [25] broom_0.7.9      compiler_4.1.1   modelr_0.1.8     xfun_0.26
## [29] pkgconfig_2.0.3  mgcv_1.8-36      htmltools_0.5.2  tidyselect_1.1.1
## [33] fansi_0.5.0      crayon_1.4.1     tzdb_0.1.2       dbplyr_2.1.1
## [37] withr_2.4.2      grid_4.1.1       nlme_3.1-152     jsonlite_1.7.2
## [41] gtable_0.3.0     lifecycle_1.0.1  DBI_1.1.1        magrittr_2.0.1
## [45] scales_1.1.1     cli_3.0.1        stringi_1.7.4    vroom_1.5.5
## [49] farver_2.1.0     fs_1.5.0         xml2_1.3.2       ellipsis_0.3.2
## [53] generics_0.1.0   vctrs_0.3.8      tools_4.1.1      bit64_4.0.5
## [57] glue_1.4.2       hms_1.1.1        parallel_4.1.1   fastmap_1.1.0
## [61] yaml_2.2.1       colorspace_2.0-2 rvest_1.0.1      knitr_1.36
## [65] haven_2.4.3      usethis_2.0.1
```