Tidy Tuesday 4/5/23

CTB

2023-04-05

```r
tuesdata <- tidytuesdayR::tt_load('2019-02-19')
```

```
## --- Compiling #TidyTuesday Information for 2019-02-19 ----

## --- There is 1 file available ---

## --- Starting Download ---

##
##   Downloading file 1 of 1: 'phd_by_field.csv'

## --- Download complete ---
```

```r
phds <- tuesdata$phd_by_field
phds <- phds %>% mutate(broad_field = as.factor(broad_field),
                        major_field = as.factor(major_field))
phds
```

```
## # A tibble: 3,370 x 5
##    broad_field   major_field                             field  year n_phds
##    <fct>         <fct>                                   <chr> <dbl>  <dbl>
##  1 Life sciences Agricultural sciences and natural resources Agric~  2008    111
##  2 Life sciences Agricultural sciences and natural resources Agric~  2008     28
##  3 Life sciences Agricultural sciences and natural resources Agric~  2008      3
##  4 Life sciences Agricultural sciences and natural resources Agron~  2008     68
##  5 Life sciences Agricultural sciences and natural resources Anima~  2008     41
##  6 Life sciences Agricultural sciences and natural resources Anima~  2008     18
##  7 Life sciences Agricultural sciences and natural resources Anima~  2008     77
##  8 Life sciences Agricultural sciences and natural resources Envir~  2008    182
##  9 Life sciences Agricultural sciences and natural resources Fishi~  2008     52
## 10 Life sciences Agricultural sciences and natural resources Food ~  2008     96
## # ... with 3,360 more rows
```

```r
phds_groups <- phds %>% group_by(broad_field, year) %>%
  summarise(tot_phds = sum(n_phds, na.rm=TRUE)) %>%
  ungroup()
```
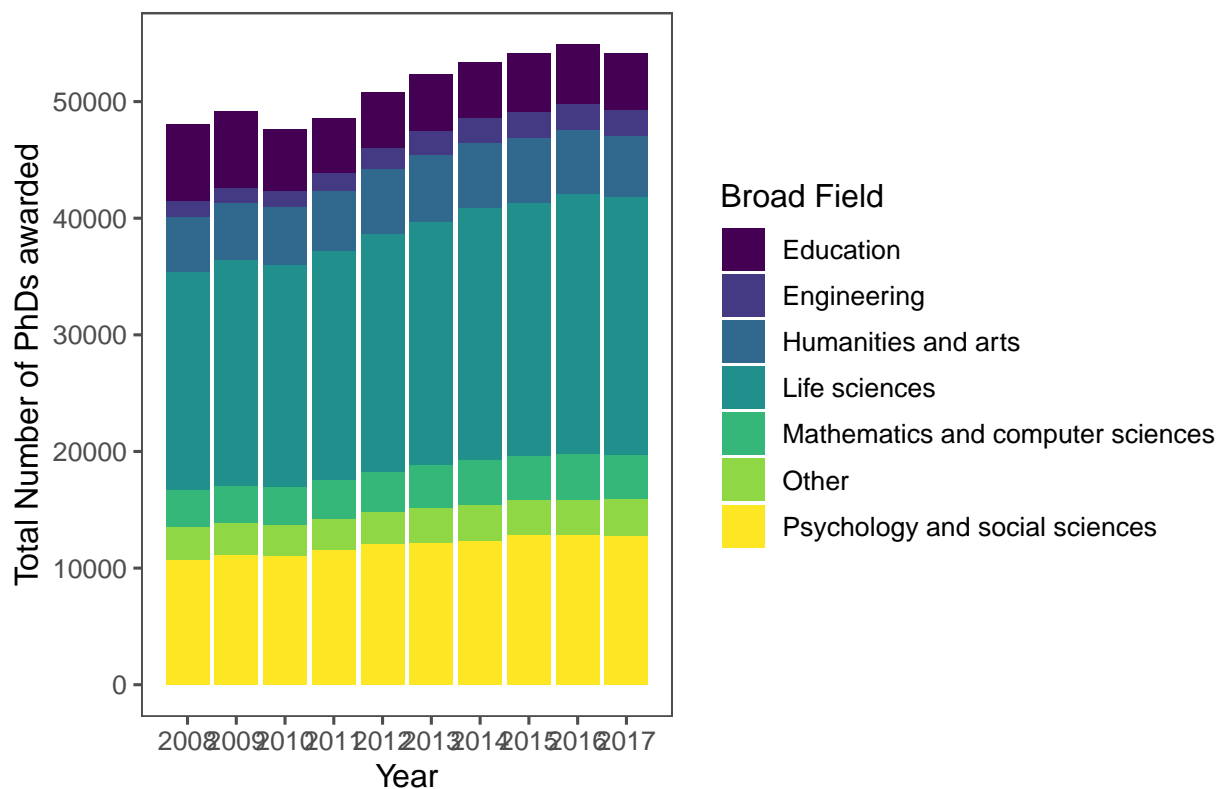
```
## 'summarise()' has grouped output by 'broad_field'. You can override using the
## '.groups' argument.
```

```
phds_groups
```

```
## # A tibble: 70 x 3
##    broad_field  year tot_phds
##    <fct>       <dbl>    <dbl>
##  1 Education    2008     6561
##  2 Education    2009     6528
##  3 Education    2010     5287
##  4 Education    2011     4670
##  5 Education    2012     4803
##  6 Education    2013     4934
##  7 Education    2014     4789
##  8 Education    2015     5098
##  9 Education    2016     5146
## 10 Education    2017     4823
## # ... with 60 more rows
```

```r
broad_field_plot <- ggplot(data=phds_groups, aes(x=year, y=tot_phds, fill=broad_field)) +
  geom_bar(position="stack", stat="identity") +
  scale_x_continuous(breaks=seq(2008, 2017, 1)) +
  ggtitle("Number of PhDs awarded by field, 2008-2017") +
  ylab("Total Number of PhDs awarded") +
  xlab("Year") +
  scale_y_continuous(breaks=seq(0, 60000, 10000)) +
  scale_fill_viridis(discrete=TRUE, name="Broad Field") +
  theme_few()
broad_field_plot
```

Number of PhDs awarded by field, 2008–2017


```r
ggsave("C:/Users/ctber/Documents/R/TidyTuesdays_qCMB/broad_field_plot.png",
       plot=broad_field_plot, width=9, height=5)
```

```r
lifesci <- phds %>% filter(broad_field == "Life sciences")
```

```r
lifesci_groups <- lifesci %>% group_by(major_field, year) %>%
  summarise(tot_phds = sum(n_phds, na.rm=TRUE)) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'major_field'. You can override using the
## `.groups` argument.
```

```r
min(lifesci_groups$year)
```

```
## [1] 2008
```

```r
max(lifesci_groups$year)
```

```
## [1] 2017
```

```
lifesci_groups
```

```
## # A tibble: 60 x 3
##    major_field                              year tot_phds
##    <fct>                                   <dbl>    <dbl>
##  1 Agricultural sciences and natural resources  2008     1198
##  2 Agricultural sciences and natural resources  2009     1283
##  3 Agricultural sciences and natural resources  2010     1100
##  4 Agricultural sciences and natural resources  2011     1206
##  5 Agricultural sciences and natural resources  2012     1255
##  6 Agricultural sciences and natural resources  2013     1324
##  7 Agricultural sciences and natural resources  2014     1338
##  8 Agricultural sciences and natural resources  2015     1434
##  9 Agricultural sciences and natural resources  2016     1381
## 10 Agricultural sciences and natural resources  2017     1606
## # ... with 50 more rows
```

```
lifesci_plot <- ggplot(data=lifesci_groups, aes(x=year, y=tot_phds, fill=major_field)) +
  geom_bar(position="stack", stat="identity") +
  scale_x_continuous(breaks=seq(2008, 2017, 1)) +
  ggtitle("Number of Life Science PhDs by field, 2008-2017") +
  ylab("Total Number of PhDs awarded") +
  xlab("Year") +
  scale_y_continuous(breaks=seq(0, 25000, 5000)) +
  scale_fill_viridis(discrete=TRUE, name="Major Field") +
  theme_few()
lifesci_plot
```

Number of Life Science PhDs by field, 2008–2017
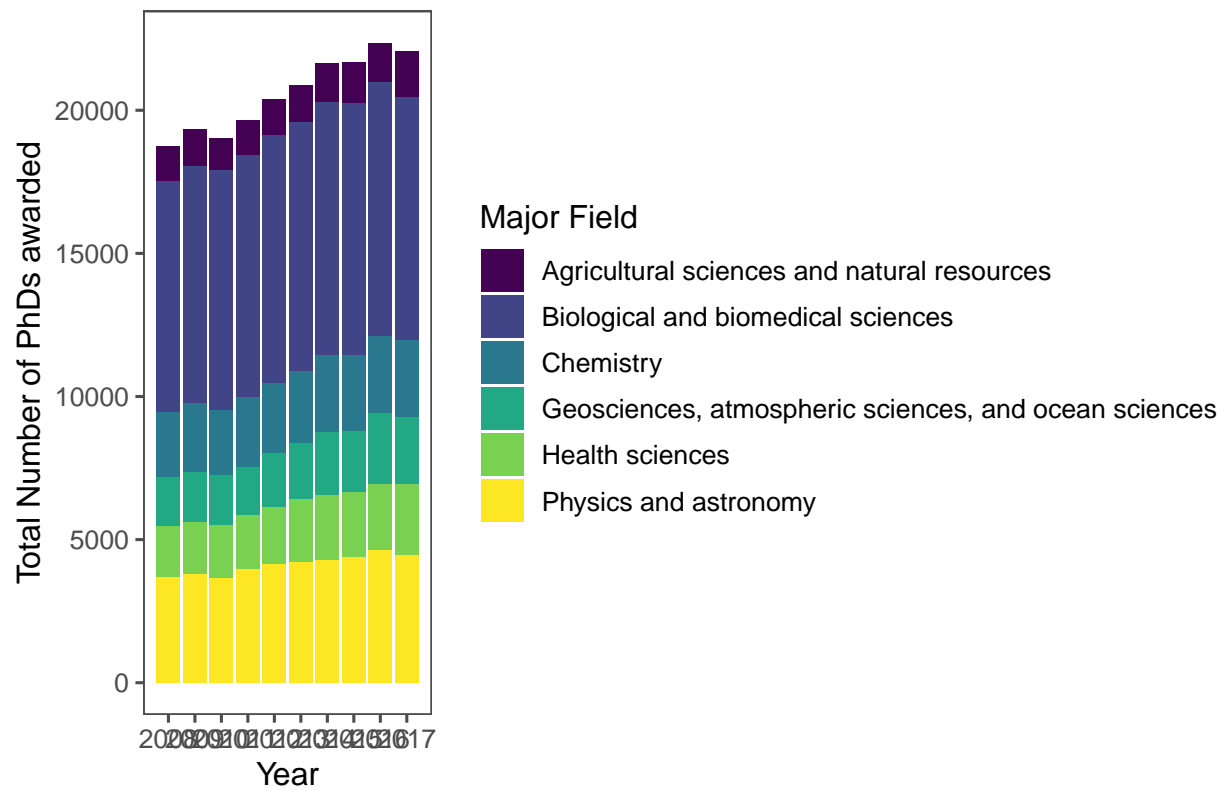

```
ggsave("C:/Users/ctber/Documents/R/TidyTuesdays_qCMB/lifesci_plot.png",
       plot=lifesci_plot, width=9, height=5)
```