

The data come from the Survey of Earned Doctorates, a very useful resource for tracking trends in PhDs awarded in the United States. The plot is made with geom\_line() and geom\_label\_repel(). The trick, if it can be dignified with that term, is to use geom\_label\_repel() on a subset of the data that contains the last year of observations only. That way we can label the endpoints in a nice way, which I think is often preferable to a key or legend that the reader has to refer to in order to decode the graph. The gghighlight package will do this for you in a single step. But this works, too.

Here’s the code for the plot shown here.

library(tidyverse)

library(janitor)

library(socviz)

library(ggrepel)

library(gghighlight)

library(earned\_doctorates)

library(showtext)

showtext\_auto()

library(myriad)

import\_myriad\_semi()

theme\_set(theme\_myriad\_semi())

phds <- read\_csv("data/earned\_doctorates\_table.csv")

phds <- clean\_names(gather(phds, year, n, `2006`:`2016`))

phds$year <- as.numeric(phds$year)

phds\_all <- phds %>% group\_by(discipline, year) %>%

tally(n)

p <- ggplot(phds\_all, aes(x = int\_to\_year(year), y = n, color = discipline)) +

geom\_line(size = 1.1) +

geom\_label\_repel(data = subset(phds\_all, year == 2016),

aes(x = int\_to\_year(year), y = n,

label = discipline,

color = discipline),

size = rel(2.1),

nudge\_x = 1,

label.padding = 0.2,

box.padding = 0.1,

segment.color = NA,

inherit.aes = FALSE) +

scale\_y\_continuous(labels = scales::comma) +

scale\_x\_date(breaks = int\_to\_year(seq(2006, 2016, by = 2)),

date\_labels = "%Y") +

coord\_cartesian(xlim = c(int\_to\_year(2006), int\_to\_year(2018))) +

guides(color = FALSE, label = FALSE) +

labs(x = "Year", y = "Count",

title = "Doctorates Awarded in the U.S., 2006-2016",

subtitle = "Selected Disciplines",

caption = "Source: Survey of Earned Doctorates")

ggsave("figures/socsci\_phd\_trends.png", p, width = 8, height = 6)