data\_exploration\_econ

Katie Yamabe

## Libraries

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

LOAD the libraries and the data’s

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.3 ✔ readr 2.1.4  
✔ forcats 1.0.0 ✔ stringr 1.5.0  
✔ ggplot2 3.4.3 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(fixest)  
library(rio)  
library(lubridate)

Load Google Trends Data

trends\_up\_to\_list <- list.files(pattern = "trends\_up\_to\_", full.names = TRUE)  
trends\_up\_to <-import\_list(trends\_up\_to\_list,rbind = TRUE, fill = TRUE)

Aggregating the Google Trends

trends\_up\_to <- trends\_up\_to %>%   
 mutate(temp = ymd(str\_sub(monthorweek,end = 10)))%>%   
 mutate(first\_of\_month = floor\_date(temp, unit = "month"))  
  
trends\_up\_to <- trends\_up\_to %>%   
 group\_by(schname, keyword)%>%   
 mutate((si = (index - mean(index))/sd(index)),na.rm = TRUE)

Loading the Score Card

score <- import("Most+Recent+Cohorts+(Scorecard+Elements).csv")  
score <- score %>%   
 mutate(opeid = OPEID)%>%  
 filter(score$PREDDEG == 3)  
  
id\_name\_link <- import("id\_name\_link.csv")   
  
id\_name\_link <- id\_name\_link %>%   
 group\_by(schname) %>%   
 mutate (n = n())%>%   
 filter(n == 1) %>%  
 ungroup()

Combine the Google Trends and Score Card

id\_name\_link <- inner\_join(trends\_up\_to, id\_name\_link, by = "schname")  
   
clean\_data <- inner\_join(id\_name\_link, score, by = c("unitid" = "UNITID", "opeid" = "OPEID"))