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)

Warning in (function (input = "", file = NULL, text = NULL, cmd = NULL, :  
Stopped early on line 1562. Expected 6 fields but found 5. Consider fill=TRUE  
and comment.char=. First discarded non-empty line: <<11,yti career institute -  
york,yti.edu,2,>>

Warning in (function (input = "", file = NULL, text = NULL, cmd = NULL, :  
Stopped early on line 1095. Expected 6 fields but found 5. Consider fill=TRUE  
and comment.char=. First discarded non-empty line: <<9,heidelberg  
university,heidelberg.edu,2,>>

Warning in (function (input = "", file = NULL, text = NULL, cmd = NULL, :  
Stopped early on line 1094. Expected 6 fields but found 5. Consider fill=TRUE  
and comment.char=. First discarded non-empty line: <<8,mount vernon nazarene  
university,mvnu.edu,2,>>

Warning in (function (input = "", file = NULL, text = NULL, cmd = NULL, :  
Stopped early on line 3280. Expected 6 fields but found 5. Consider fill=TRUE  
and comment.char=. First discarded non-empty line: <<41,potomac state college  
of west virginia university,potomac state college of west virginia  
university,1,>>

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))

Loading the Score Card

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

Combine the Google Trends and Score Card

trends\_up\_to <- trends\_up\_to %>%   
 inner\_join(id\_name\_link, join\_by(schname))

Warning in inner\_join(., id\_name\_link, join\_by(schname)): Detected an unexpected many-to-many relationship between `x` and `y`.  
ℹ Row 1561 of `x` matches multiple rows in `y`.  
ℹ Row 3591 of `y` matches multiple rows in `x`.  
ℹ If a many-to-many relationship is expected, set `relationship =  
 "many-to-many"` to silence this warning.

score <- score %>%   
 mutate(opeid = OPEID)  
  
clean\_data <- trends\_up\_to %>%   
 inner\_join(score, join\_by(opeid))

Warning in inner\_join(., score, join\_by(opeid)): Detected an unexpected many-to-many relationship between `x` and `y`.  
ℹ Row 47385 of `x` matches multiple rows in `y`.  
ℹ Row 1073 of `y` matches multiple rows in `x`.  
ℹ If a many-to-many relationship is expected, set `relationship =  
 "many-to-many"` to silence this warning.