adpricing

nwfried

2024-08-20

Load in Gentzkow dataset and filter for newspapers from 1960.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
data <- read_delim(file.path("data/30261-0005-Data.tsv"))
```

```
## Rows: 89815 Columns: 5
## -- Column specification -------------------------------------------------------
## Delimiter: "\t"
## chr (3): adprice, adDepth, adType
## dbl (2): recid, adpricetype
##
## 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.
```

```
head(data)
```

```
## # A tibble: 6 x 5
##         recid adpricetype adprice adDepth adType
##         <dbl>       <dbl> <chr>   <chr>   <chr>
## 1 1880010001           0 3.75    " "     " "
## 2 1880010002           0 3       " "     " "
## 3 1880010003           0 2.83    " "     " "
## 4 1880010004           0 3.34    " "     " "
## 5 1880010005           0 3.34    " "     " "
## 6 1880010006           0 3.12    " "     " "
```

```
# data <- data %>% filter(starts_with(recid))
ads <- read_delim(file.path("data/30261-0002-Data.tsv"))
```

```
## Rows: 58870 Columns: 24
## -- Column specification -------------------------------------------------------
## Delimiter: "\t"
## chr (21): when, papername, polaff, category, circ, circtype, circsworn, orig...
```

```
## dbl  (3): recid, permid, year
##
## 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.
ads <- ads %>% filter(year == 1960)
```

Cleaning data for relevant information.

```
ads <- ads %>% select(recid, permid, circ, adprice) %>%
  filter(adprice != " ")
```

```
ads <- ads %>% mutate(circ = as.numeric(circ)) %>%
  mutate(adprice = as.numeric(adprice))
corr <- ads %>% summarise(correlation = cor(circ, adprice, use = "complete.obs")) %>%
  pull(correlation)
print(corr)
```

```
## [1] 0.8218903
```

Correlation between circulation and adprice is 0.8218903 in 1960.