

adpricing

nwfried

2024-08-27

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    " "     " "
```

```
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, circatype, 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)
glimpse(ads)
```

```
## Rows: 1,747
## Columns: 24
## $ recid      <dbl> 1960200001, 1960200002, 1960200003, 1960200004, 1960~
## $ permid     <dbl> 4862, 6893, 2396, 5499, 4866, 4440, 5500, 4435, 4868~
## $ year       <dbl> 1960, 1960, 1960, 1960, 1960, 1960, 1960, 1960, 1960~
## $ when       <chr> "e", "m", "e", "e", "e", "e", "e", "e", "e", "m", "e~
## $ papername   <chr> "Star", "Post-Herald", "News", "Decatur Daily", "Eag~
## $ polaff      <chr> "ID", "ID", "ID", "D", "I", "ID", "ID", "I", "ID", "~
## $ category    <chr> " ", " ", " ", " ", " ", " ", " ", " ", " ", " ", " ~
## $ circ        <chr> "20006", "96300", "187516", "13700", "23381", "20335~
## $ circatype   <chr> "EP Number", "EP Number", "EP Number", "EP Number", ~
## $ circsworn   <chr> " ", " ", " ", " ", " ", " ", " ", " ", " ", " ", " ~
## $ origpublisher <chr> "Consolidated Pub. Co., 216 W. 10th St., Anniston, A~
## $ subpricetype <chr> "5", " ", " ", " ", "5", "5", " ", " ", "5", "5", " ~
## $ subprice     <chr> "16.631196504", " ", " ", "14.927628996", "14.927628~
## $ adprice      <chr> "2.101455212", "10.01927662", "14.19397545", "1.8629~
## $ adpricetype  <chr> "5", "5", "5", "5", "5", "1", "1", "5", "5", "5", "5~
## $ adDepth      <chr> "LINE", "LINE", "LINE", "LINE", "LINE", "LINE", "LIN~
## $ adType       <chr> "FLAT", "FLAT", "FLAT", "FLAT", "FLAT", " ", " ", " ~
## $ origpages    <chr> " ", " ", " ", " ", " ", " ", " ", " ", " ", " ", " ~
## $ minpage      <chr> " ", " ", " ", " ", " ", " ", " ", " ", " ", " ", " ~
## $ maxpage      <chr> " ", " ", " ", " ", " ", " ", " ", " ", " ", " ", " ~
## $ editors      <chr> " ", " ", " ", " ", " ", " ", " ", " ", " ", " ", " ~
## $ publishers   <chr> " ", " ", " ", " ", " ", " ", " ", " ", " ", " ", " ~
## $ party_endorsement <chr> " ", "1", "1", "-1", "1", "0", "0", " ", "-1", "0", ~
## $ ownerpermid  <chr> " ", "575", "116", " ", " ", "93", "93", " ", "116",~
```

Cleaning data for relevant information.

```
ads <- ads %>% 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.

Wrangle data to get an average ad price per city:

Reading in location data and merging into dataframe;

```
location <- read_delim(file.path("data/30261-0001-Data.tsv"))
```

```
## Rows: 7774 Columns: 10
## -- Column specification -----
## Delimiter: "\t"
## chr (5): nextpermid, papername_constant, dateestab, polaff, when
```

```
## dbl (5): permid, citypermid, nonpol, min_year, max_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.
```

```
bypermid = join_by(permid)
ads <- left_join(ads, location, by = bypermid)
```

Group by citypermid and calculate average adprice;

```
ads <- ads %>% group_by(citypermid) %>%
  summarise(adprice = mean(adprice))
```

Export to read into main file.

```
write_csv(ads, "data/ads.csv")
```