A&S 500 Final Project

Jacob Saindon

Loading Packages

```
install.packages("gt")
install.packages("htmltools")
```

```
library(tidyverse)
library(lubridate)
library(gt)
```

Data

```
snap <- read_csv("data/snap_political_20.csv")
```

Tidying

```
snap |>
  filter(
    CountryCode == "united states",
    !is.na(Interests),
    !is.na(`Regions (Included)`)
  ) |>
  mutate(
    STATE_INC = str_split(`Regions (Included)`, ",")
    ) |>
  unnest(cols = STATE_INC) -> snap_states
```

```
snap_states |>
  mutate(
    S_MONTH = month(StartDate, label = TRUE, abbr = TRUE)) -> snap_sm

snap_sm |>
  mutate(
    TARGET = str_split(`Interests`, ",")
    ) |>
    unnest(cols = TARGET) -> snap_unnest
```

```
snap_unnest |>
  summarise(
      .by = c(STATE_INC,TARGET,S_MONTH),
      TOTAL = n()) |>
  mutate(
    id = row_number()) -> snap_tidy # all states, interests, and months in separate rows

snap_tidy
```

```
# A tibble: 2,065 x 5
   STATE_INC  TARGET                                S_MONTH TOTAL    id
   <chr>      <chr>                                 <ord>   <int> <int>
 1 California Advocates & Activists                 Sep        20     1
 2 California Bookworms & Avid Readers              Sep        20     2
 3 California Collegiates                           Sep        20     3
 4 California Investors & Entrepreneurs             Sep        20     4
 5 California Money Minders                         Sep        20     5
 6 California News Watchers                         Sep        20     6
 7 California Philanthropists                       Sep        20     7
 8 California TV Network Viewers (CNN)              Sep        20     8
 9 California TV Network Viewers (FOX News Channel) Sep        20     9
10 California TV Network Viewers (MSNBC)            Sep        20    10
# i 2,055 more rows
```

Data Summaries

```
snap_sm
```

```
# A tibble: 2,230 x 40
   ADID          CreativeUrl `Currency Code` Spend Impressions StartDate EndDate
   <chr>         <chr>       <chr>           <dbl>       <dbl> <chr>     <chr>
 1 7652fe9f1e45~ https://ww~ USD             56423    18413015 2020/09/~ 2020/1~
 2 c68426091abf~ https://ww~ USD             50478    16525973 2020/09/~ 2020/1~
 3 ef900766a7e9~ https://ww~ USD             42543    14095083 2020/09/~ 2020/1~
 4 35a5ec450278~ https://ww~ USD             33743     9883322 2020/10/~ 2020/1~
 5 35a5ec450278~ https://ww~ USD             33743     9883322 2020/10/~ 2020/1~
 6 35a5ec450278~ https://ww~ USD             33743     9883322 2020/10/~ 2020/1~
 7 35a5ec450278~ https://ww~ USD             33743     9883322 2020/10/~ 2020/1~
 8 35a5ec450278~ https://ww~ USD             33743     9883322 2020/10/~ 2020/1~
 9 35a5ec450278~ https://ww~ USD             33743     9883322 2020/10/~ 2020/1~
10 35a5ec450278~ https://ww~ USD             33743     9883322 2020/10/~ 2020/1~
# i 2,220 more rows
# i 33 more variables: OrganizationName <chr>, BillingAddress <chr>,
#   CandidateBallotInformation <chr>, PayingAdvertiserName <chr>,
#   CommitteeName <lgl>, CommitteeIdentificationNumber <lgl>,
#   DisclosureNameOfCommittee <lgl>, AdvertisingJurisdiction <lgl>,
#   Gender <chr>, AgeBracket <chr>, CountryCode <chr>,
#   `Regions (Included)` <chr>, `Regions (Excluded)` <chr>, ...
```

```
snap_sm |>
  group_by(STATE_INC) |>
  summarise(
    COUNT = n()) |>
  arrange(-COUNT) -> state_count

 # count of ads run by state
```

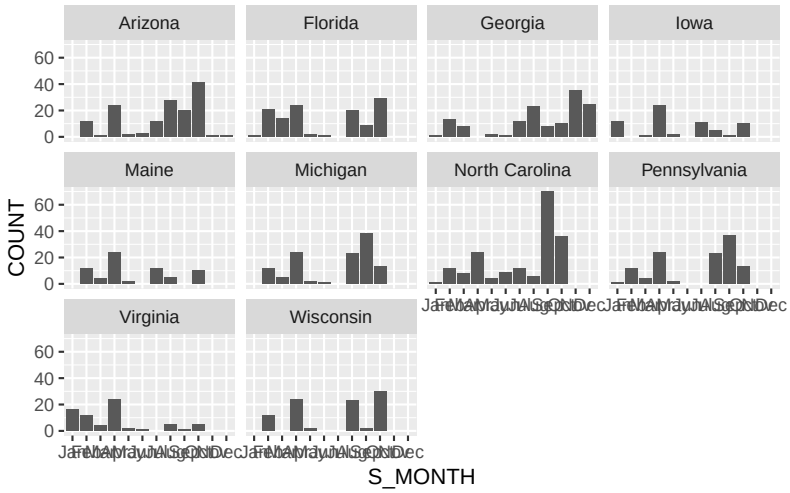
```
snap_sm |>
    filter(
    STATE_INC == "North Carolina" | STATE_INC == "Arizona" | STATE_INC == "Georgia" | STATE_
  group_by(STATE_INC, S_MONTH)|>
  summarise(
```

```
    COUNT = n()) |>
  arrange(-COUNT) -> sm_count

sm_count # count of ads run by state, per month
```

```
# A tibble: 87 x 3
# Groups:   STATE_INC [10]
   STATE_INC      S_MONTH COUNT
   <chr>          <ord>   <int>
 1 North Carolina Sep        70
 2 Arizona        Oct        41
 3 Michigan       Sep        38
 4 Pennsylvania   Sep        37
 5 North Carolina Oct        36
 6 Georgia        Nov        35
 7 Wisconsin      Oct        30
 8 Florida        Oct        29
 9 Arizona        Aug        28
10 Georgia        Dec        25
# i 77 more rows
```

```
sm_count |>
ggplot(aes(x=S_MONTH, y=COUNT))+
  geom_col()+
  facet_wrap(~STATE_INC) # volume of ads launching per state
```




```
snap_tidy |>
  group_by(STATE_INC, S_MONTH) |>
  summarise(TOTAL = sum(TOTAL)) |>
  arrange(-TOTAL) -> sm_sum # total number of audiences targeted by all ads, per state, per

sm_sum
```

```
# A tibble: 297 x 3
# Groups:   STATE_INC [51]
   STATE_INC      S_MONTH TOTAL
   <chr>          <ord>   <int>
 1 Arizona        Oct       736
 2 Pennsylvania   Sep       434
 3 Wisconsin      Oct       386
 4 Michigan       Sep       380
 5 Florida        Mar       258
 6 Florida        Feb       237
 7 Arizona        Sep       231
 8 North Carolina Oct       218
 9 California     Sep       204
10 Georgia        Nov       159
# i 287 more rows
```

```r
snap_tidy |>
  group_by(TARGET, S_MONTH) |>
  summarise(
    SUM = sum(TOTAL)) |>
  arrange(-SUM) -> tm_sum # total number of ads including a given parameter, per month

tm_sum
```

```
# A tibble: 451 x 3
# Groups:   TARGET [189]
   TARGET                        S_MONTH   SUM
   <chr>                         <ord>   <int>
 1 Political News Watchers       Feb       420
 2 Green Living Enthusiasts      Feb       414
 3 Bookworms & Avid Readers      Feb       408
 4 Outdoor & Nature Enthusiasts Feb        408
 5 Political News Watchers       Aug       293
 6 Advocates & Activists         Apr       176
 7 Political News Watchers       Apr       176
 8 Political News Watchers       Jul       151
 9 Advocates & Activists         Jul       150
10 TV Network Viewers (CNN)      Jul       143
# i 441 more rows
```

```r
snap_tidy |>
  group_by(TARGET, STATE_INC) |>
  summarise(
    SUM = sum(TOTAL)) |>
  arrange(-SUM) -> ts_sum # total number of ads including a given parameter, per state

ts_sum
```

```
# A tibble: 1,336 x 3
# Groups:   TARGET [189]
   TARGET                   STATE_INC         SUM
   <chr>                    <chr>           <int>
 1 Political News Watchers  Arizona            71
 2 Advocates & Activists    Arizona            66
```

```
 3 Political News Watchers Pennsylvania      63
 4 Political News Watchers Michigan          61
 5 Political News Watchers North Carolina    61
 6 Political News Watchers Georgia           60
 7 Advocates & Activists   Michigan          59
 8 Advocates & Activists   North Carolina    57
 9 Political News Watchers Maine             57
10 Advocates & Activists   Georgia           55
# i 1,326 more rows
```

Top Interests by State and Month

```
snap_tidy |>
  slice_max(TOTAL, by = c(STATE_INC, S_MONTH)) -> max_list # all top interests per state, pe

max_list
```

```
# A tibble: 752 x 5
   STATE_INC  TARGET                                S_MONTH TOTAL    id
   <chr>      <chr>                                 <ord>   <int> <int>
 1 California Advocates & Activists                 Sep        20     1
 2 California Bookworms & Avid Readers              Sep        20     2
 3 California Collegiates                           Sep        20     3
 4 California Investors & Entrepreneurs             Sep        20     4
 5 California Money Minders                         Sep        20     5
 6 California News Watchers                         Sep        20     6
 7 California Philanthropists                       Sep        20     7
 8 California TV Network Viewers (CNN)              Sep        20     8
 9 California TV Network Viewers (FOX News Channel) Sep        20     9
10 California TV Network Viewers (MSNBC)            Sep        20    10
# i 742 more rows
```

```
max_list |>
  group_by(STATE_INC, S_MONTH) |>
  subset(select = c(STATE_INC, S_MONTH, TARGET)) |>
    nest(.key = "TOP_TARGET") |>
  arrange(S_MONTH, STATE_INC) -> top_snap
```

```
top_snap # column with list (tibble) of top interests by per state, per month
```

```
# A tibble: 297 x 3
# Groups:   STATE_INC, S_MONTH [297]
   STATE_INC      S_MONTH TOP_TARGET
   <chr>          <ord>   <list>
 1 California     Jan     <tibble [2 x 1]>
 2 Florida        Jan     <tibble [2 x 1]>
 3 Georgia        Jan     <tibble [2 x 1]>
 4 Illinois       Jan     <tibble [2 x 1]>
 5 Iowa           Jan     <tibble [4 x 1]>
 6 Nevada         Jan     <tibble [7 x 1]>
 7 New Hampshire  Jan     <tibble [4 x 1]>
 8 New York       Jan     <tibble [2 x 1]>
 9 North Carolina Jan     <tibble [2 x 1]>
10 Ohio           Jan     <tibble [2 x 1]>
# i 287 more rows
```

Plotting

```
# month_max |>
#   ggplot(aes(S_MONTH, MAX, fill = STATE_INC))+
#   geom_col()+
#   theme(legend.position="none")
```

```
#data |>
#group by (month, state, interest) |>
#summarise (n == n()) |>
#filter (n==max(n)) |>
#select(-month)
#maxes |>
#leftjoin(data)
```

Text analysis of targeting categories, by state.

Unnest both “Regions” and “Interest” fields to get tidy.

#dlxs are oracle datalogix audiences? https://businesshelp.snapchat.com/s/article/custom-audiences?language=en_US

targetsmart and i360: https://www.axios.com/2022/09/08/snap-voter-data-republican-democrats

Geographic Scales:

1. Region (State)
    - Inclusions
    - Exclusions
    - Total Ads
    - Total Spend
    - Total Impressions
    - Interests (text analysis)
2. Metros (City)
    - Inclusions
    - Exclusions
    - Total Ads
    - Total Spend
    - Total Impressions
    - Interests (text analysis)
3. Postal Codes (Zip)
    - Inclusions
    - Exclusions
    - Total Ads
    - Total Spend
    - Total Impressions
    - Interests (text analysis)
4. Location Categories
    - Inclusions
    - Exclusions
    - Total Ads

- Total Spend
- Total Impressions
- Interests (text analysis)

5. Electoral Districts
    - Inclusions
    - Exclusions
    - Total Ads
    - Total Spend
    - Total Impressions
    - Interests (text analysis)