Chief Media Day 3

Skyler Hauser

2023-06-14

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
install.packages("readr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
```

```
## -- 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
```

```
library(readr)
library(ggplot2)
library(dplyr)
library(knitr)
library(dplyr)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

Sunsetter Product Interest Analysis

How has interest in the Sunsetter brand changed on a DMA-by-DMA basis from February to present?

Guidance: We're looking to see if geography is a major contributing factor to product interest. Search volume is a good proxy for product interest, so only use visits where medium = "cpc". Calculate and analyze changes over time by DMA, looking in particular for the largest changes, be it increases or reductions in search. Use a couple of visualizations to convey the main findings.

```
sunsetter_web_data <- read_csv("sunsetter_web_data.csv")
```

```
## Rows: 1368465 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): DMA, Region, Country, Medium
## dttm (1): Session Date / Time
##
## 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.
```

```
glimpse(sunsetter_web_data)
```

```
## Rows: 1,368,465
## Columns: 5
## $ `Session Date / Time` <dttm> 2023-06-01 00:00:54, 2023-06-01 00:00:46, 2023-~
## $ DMA                   <chr> "Rockford, IL", "Chicago, IL", "Dayton, OH", NA,~
## $ Region                <chr> "Illinois", "Illinois", "Ohio", NA, "Mississippi~
## $ Country               <chr> "USA", "USA", "USA", "USA", "USA", "USA", "USA",~
## $ Medium                <chr> "paid_social", "cpc", "paid_social", "cpc", "ref~
```

Filtered data for visits with medium = cpc

```
# Filter data for visits with medium = "cpc"
filtered_sunsetter <- sunsetter_web_data %>%
  filter(Medium == "cpc")
filtered_sunsetter
```

```
## # A tibble: 598,817 x 5
##    `Session Date / Time` DMA                   Region     Country Medium
##    <dttm>                <chr>                 <chr>      <chr>   <chr>
##  1 2023-06-01 00:00:46   Chicago, IL           Illinois   USA     cpc
##  2 2023-06-01 00:00:38   <NA>                  <NA>       USA     cpc
##  3 2023-06-01 00:00:29   Dallas-Fort Worth, TX Texas      USA     cpc
##  4 2023-06-01 00:00:28   <NA>                  <NA>       USA     cpc
##  5 2023-06-01 00:00:28   New York, NY          New Jersey USA     cpc
##  6 2023-06-01 00:00:23   New York, NY          New Jersey USA     cpc
##  7 2023-06-01 00:00:00   San Francisco, CA     California USA     cpc
##  8 2023-05-31 23:59:36   Detroit, MI           Michigan   USA     cpc
##  9 2023-05-31 23:59:27   Dallas-Fort Worth, TX Texas      USA     cpc
## 10 2023-05-31 23:59:09   Chattanooga, TN       Tennessee  USA     cpc
## # i 598,807 more rows
```

This code is filtering the data set to show only Medium column has the value "cpc".

Extraction of month and year from data set

```
# Extract the month and year from the Session Date / Time column
filtered_sunsetter$MonthYear <- format(filtered_sunsetter$`Session Date / Time`, "%Y-%m")
```

```r
print(head(filtered_sunsetter$MonthYear, 100))
```

```
##   [1] "2023-06" "2023-06" "2023-06" "2023-06" "2023-06" "2023-06" "2023-06"
##   [8] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [15] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [22] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [29] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [36] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [43] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [50] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [57] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [64] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [71] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [78] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [85] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [92] "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05" "2023-05"
##  [99] "2023-05" "2023-05"
```

Here, the code extracts the month and year from the Session Date / Time column in the filtered_sunsetter dataset. It is put into the form Year-Month. I cut the code to only show the first 100 because with all the rows it would make the report super long.

Grouping data by DMA and Month, Then calculated visit count

```r
# Group the data by DMA and MonthYear, and calculate the visit count
dma_visit_counts <- filtered_sunsetter %>%
  group_by(DMA, MonthYear) %>%
  summarise(VisitCount = n())
```

```
## `summarise()` has grouped output by 'DMA'. You can override using the `.groups`
## argument.
```

```r
dma_visit_counts
```

```
## # A tibble: 848 x 3
## # Groups:   DMA [211]
##    DMA                          MonthYear VisitCount
##    <chr>                        <chr>          <int>
##  1 Abilene-Sweetwater, TX       2023-02           24
##  2 Abilene-Sweetwater, TX       2023-03           33
##  3 Abilene-Sweetwater, TX       2023-04           46
##  4 Abilene-Sweetwater, TX       2023-05           53
##  5 Albany, GA                   2023-02           40
##  6 Albany, GA                   2023-03           85
##  7 Albany, GA                   2023-04          104
##  8 Albany, GA                   2023-05           97
##  9 Albany-Schenectady-Troy, NY 2023-02          333
## 10 Albany-Schenectady-Troy, NY 2023-03          589
## # i 838 more rows
```

This code groups the filtered_sunsetter dataset by two variables: 'DMA' (Designated Market Area) and 'MonthYear'. It then calculates the number of visits (visit count) for each combination of 'DMA' and 'MonthYear'. So for each place, and month we are looking at it calculates the change in that month.

Calculations of change in visit count from Feb to May 2023

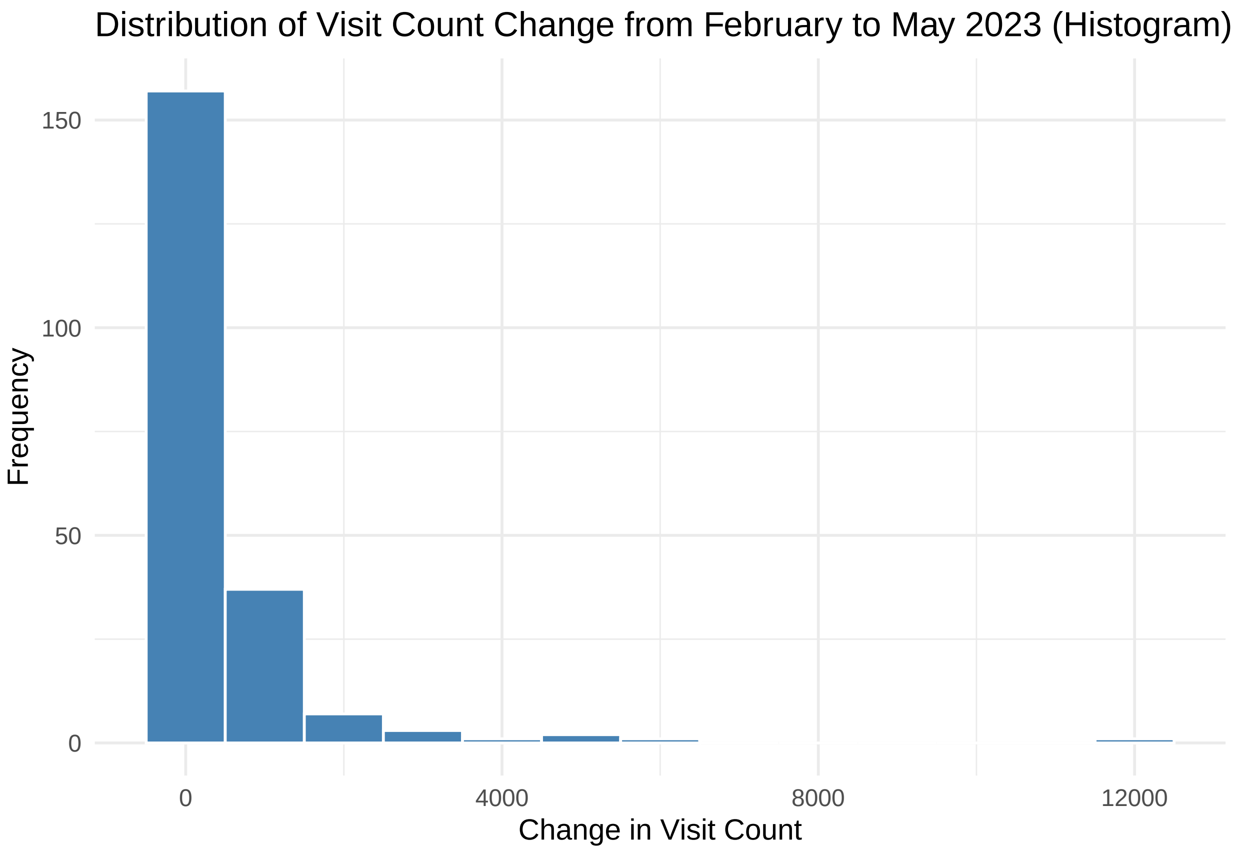
```
library(dplyr)
library(tidyr)
library(magrittr)
# Calculate the change in visit count from February to the present for each DMA
dma_change <- dma_visit_counts %>%
  pivot_wider(names_from = MonthYear, values_from = VisitCount) %>%
  select(-`2023-06`) %>%
  mutate(Change = `2023-05` - `2023-02`) %>%
  arrange(desc(Change))

dma_change <- na.omit(dma_change)

dma_change
```

```
## # A tibble: 209 x 6
## # Groups:   DMA [209]
##    DMA                     `2023-02` `2023-03` `2023-04` `2023-05` Change
##    <chr>                       <int>     <int>     <int>     <int>  <int>
##  1 New York, NY                 4595      9330     12511     16829  12234
##  2 Boston, MA                   1471      3415      5284      7021   5550
##  3 Philadelphia, PA             2213      4466      6085      7563   5350
##  4 Chicago, IL                  1892      3678      5034      6615   4723
##  5 Washington, DC               1852      3589      5388      5830   3978
##  6 Detroit, MI                   939      1947      2947      4098   3159
##  7 Minneapolis-St Paul, MN       661      1596      2224      3486   2825
##  8 Seattle-Tacoma, WA           1090      1944      2357      3875   2785
##  9 Los Angeles, CA              2270      3564      4320      4591   2321
## 10 Hartford, CT                  624      1348      2151      2928   2304
## # i 199 more rows
```

The pivot_wider function is making the data set wider by converting the MonthYear column values into individual columns.Each column represents a specific month and year combination, and the values in these columns correspond to the visit counts The mutate function is creating a new column named change by subtracting the visit count in May 2023 (represented by the column 2023-05) from the visit count in February 2023 (represented by the column 2023-02). This calculation represents the change in visit count from February to the present for each DMA. I had to do may because June had no data represented to calculate the difference. This calculation represents the total change in a specfic place over the months we are looking at. June we had no data so I chose to do Feb-May. The last column calculates the change for every city of every month we are looking at. For example, Atlanta GA change was 1871 while August GA was 97 over the 4 months. I ordered it so the largest change is seen first, New York.

Histogram of Distribution of Visit Change from Feb to May 2023

```
library(ggplot2)

histogram3 <- ggplot(dma_change, aes(x = Change)) +
  geom_histogram(binwidth = 1000, fill = "steelblue", color = "white") +
  labs(x = "Change in Visit Count", y = "Frequency",
       title = "Distribution of Visit Count Change from February to May 2023 (Histogram)") +
  theme_minimal()
histogram3
```



This histogram shows the change in Visit count per the amount of times it is seen. So this histogram makes it seem as though there were not a lot of large change with a few outliers, New York being the largest outlier.

DMA percent change calculations

```
dma_percentchange <- dma_visit_counts %>%
  pivot_wider(names_from = MonthYear, values_from = VisitCount) %>%
  select(-`2023-06`) %>%
  mutate(Change = `2023-05` - `2023-02`) %>%
  mutate(Percentage_Change = (Change / `2023-02`) * 100) %>%
  arrange(desc(Change))
dma_percentchange <- na.omit(dma_percentchange)
dma_percentchange
```

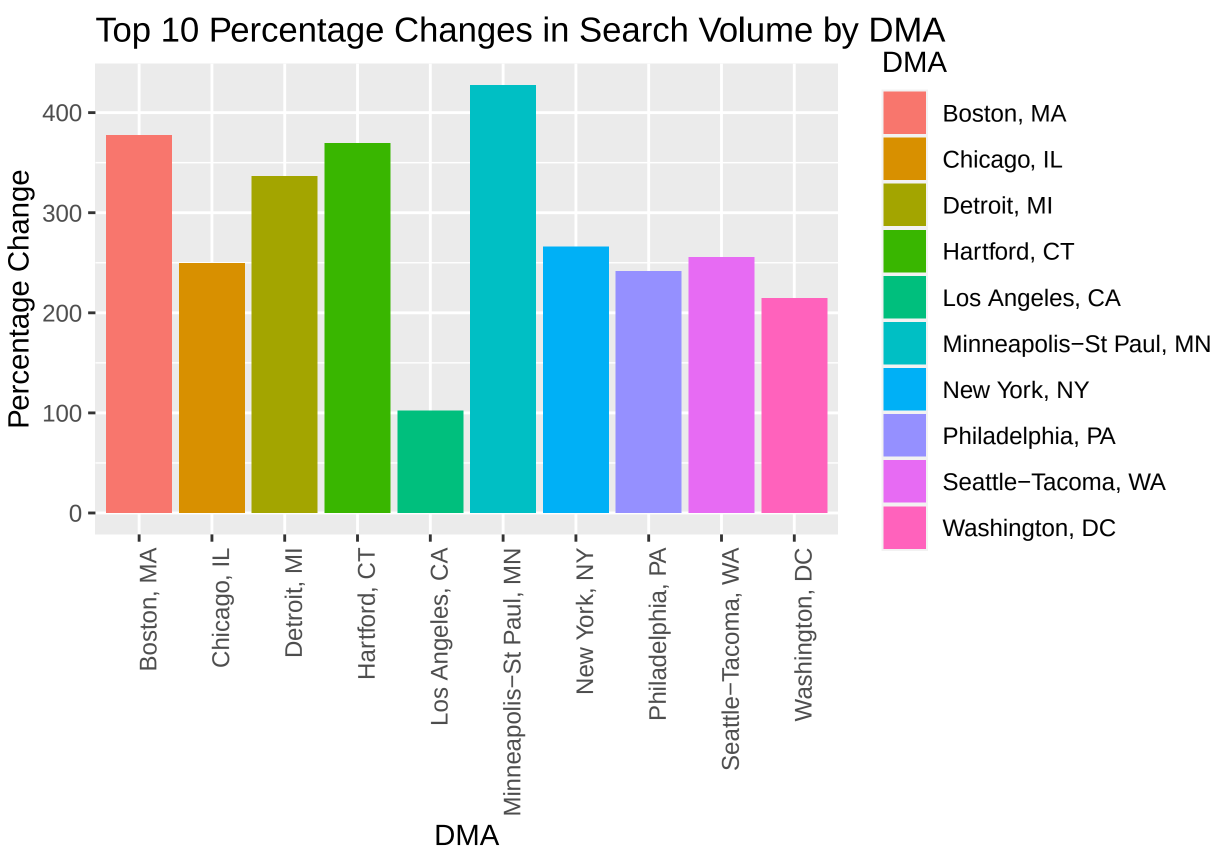
```
## # A tibble: 209 x 7
## # Groups:   DMA [209]
##    DMA           `2023-02` `2023-03` `2023-04` `2023-05` Change Percentage_Change
##    <chr>             <int>     <int>     <int>     <int>  <int>             <dbl>
##  1 New York, NY       4595      9330     12511     16829  12234              266.
##  2 Boston, MA         1471      3415      5284      7021   5550              377.
##  3 Philadelphi~       2213      4466      6085      7563   5350              242.
##  4 Chicago, IL        1892      3678      5034      6615   4723              250.
##  5 Washington,~       1852      3589      5388      5830   3978              215.
##  6 Detroit, MI         939      1947      2947      4098   3159              336.
##  7 Minneapolis~        661      1596      2224      3486   2825              427.
##  8 Seattle-Tac~       1090      1944      2357      3875   2785              256.
```

```
##  9 Los Angeles~      2270       3564       4320       4591    2321                  102.
## 10 Hartford, CT       624       1348       2151       2928    2304                  369.
## # i 199 more rows
```

This includes the percent change since february for each DMA.

Bar chart of Top 10 Percent Changes in Search Volume by DMA

```
library(ggplot2)
graph2 <- ggplot(head(dma_percentchange, 10), aes(x = DMA, y = Percentage_Change, fill = DMA)) +
  geom_bar(stat = "identity") +
  labs(x = "DMA", y = "Percentage Change", title = "Top 10 Percentage Changes in Search Volume by DMA")
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

graph2
```



This visual shows the measures of percent change since february. As we can see Minneapolis St. Paul had the highest change and Los Angeles CA has the smallest percent change. Even though New York had the largest change, it did not have the largest percent change. That means, the visit amount change was not as large as the numbers show us.

Largest changes in search volume by DMA

```
# Find the largest changes (increases or reductions) in search volume by DMA
largest_changes <- dma_change %>%
  arrange(desc(Change))
```

```
  head(10)
```

```
## [1] 10
```

```
largest_changes <- largest_changes[-5, ]  # Remove the row corresponding to June
largest_changes <- largest_changes[, -6]  # Remove the column corresponding to June


largest_changes
```

```
## # A tibble: 208 x 5
## # Groups:   DMA [208]
##    DMA                     `2023-02` `2023-03` `2023-04` `2023-05`
##    <chr>                       <int>     <int>     <int>     <int>
##  1 New York, NY                 4595      9330     12511     16829
##  2 Boston, MA                   1471      3415      5284      7021
##  3 Philadelphia, PA             2213      4466      6085      7563
##  4 Chicago, IL                  1892      3678      5034      6615
##  5 Detroit, MI                   939      1947      2947      4098
##  6 Minneapolis-St Paul, MN       661      1596      2224      3486
##  7 Seattle-Tacoma, WA           1090      1944      2357      3875
##  8 Los Angeles, CA              2270      3564      4320      4591
##  9 Hartford, CT                  624      1348      2151      2928
## 10 Pittsburgh, PA                990      1594      2365      2986
## # i 198 more rows
```

The arrange function is sorting the data in descending order so the top value will be first, the DMA entries with the largest changes will appear at the top of the dataset.

The head function is selecting just the first 10 rows from the dataset, showing the 10 largest changes in search volume. The largest_changes table displays the top 10 DMAs with the largest changes in search volume. The rows are sorted in descending order based on the "Change" column.

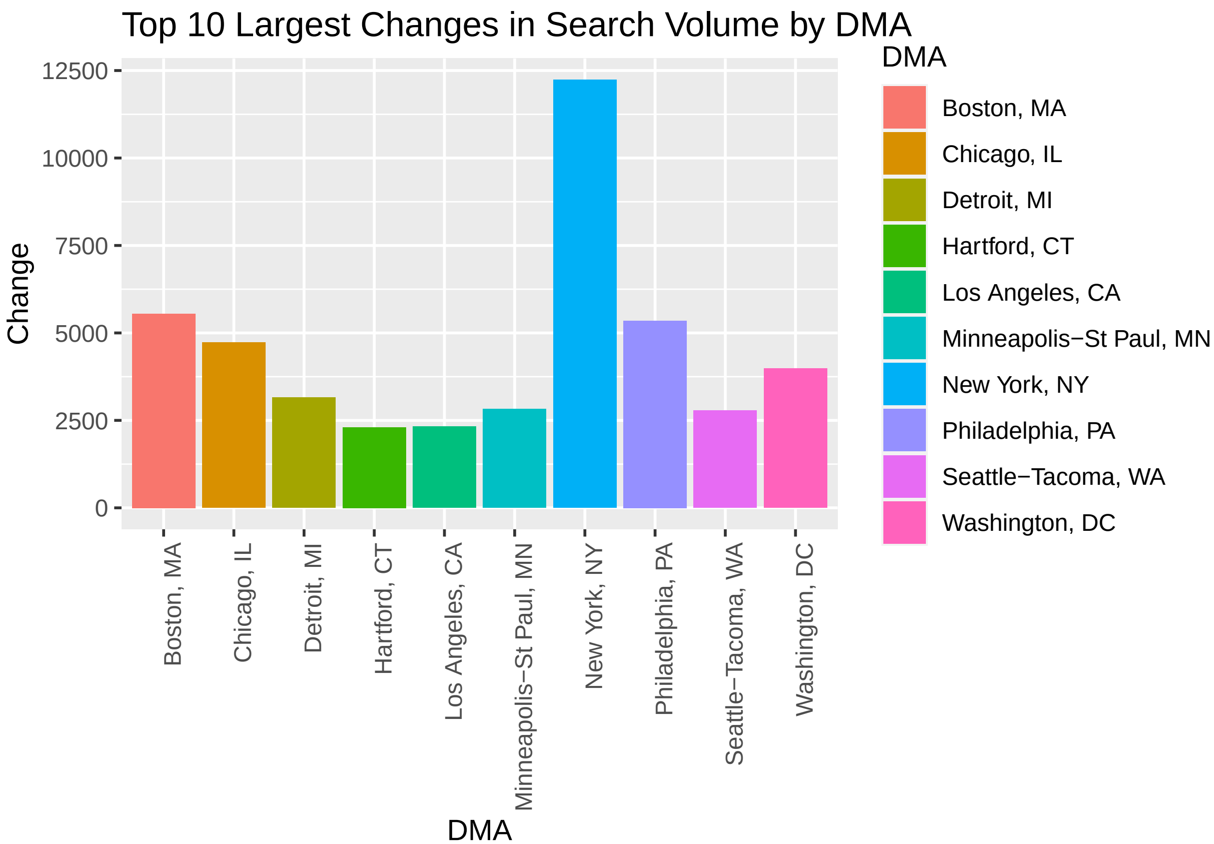
For example, in the first row, the DMA has no specific label (NA), and the search volume in February was 14,510, which increased to 49,815 in June. The change in search volume from February to June was 35,305.

The table provides information about specific DMAs, their search volume over the months, and the changes in search volume. However, without further context or additional data, it is challenging to draw broader conclusions or analyze the overall trend in interest in the Sunsetter brand.

New york has the largest change of 12234.

Bar chart of 10 Largest Changes in Search Volume by DMA

```
largest_changes <- dma_change %>%
  arrange(desc(`2023-05` - `2023-02`)) %>%
  head(10)


graph3 <- ggplot(largest_changes, aes(x = DMA, y = `2023-05` - `2023-02`, fill = DMA)) +
  geom_bar(stat = "identity") +
  labs(x = "DMA", y = "Change", title = "Top 10 Largest Changes in Search Volume by DMA") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

graph3
```



This visual Shows the greatest to least changes, of the highest represented changes. New York is one and then Boston is to follow.

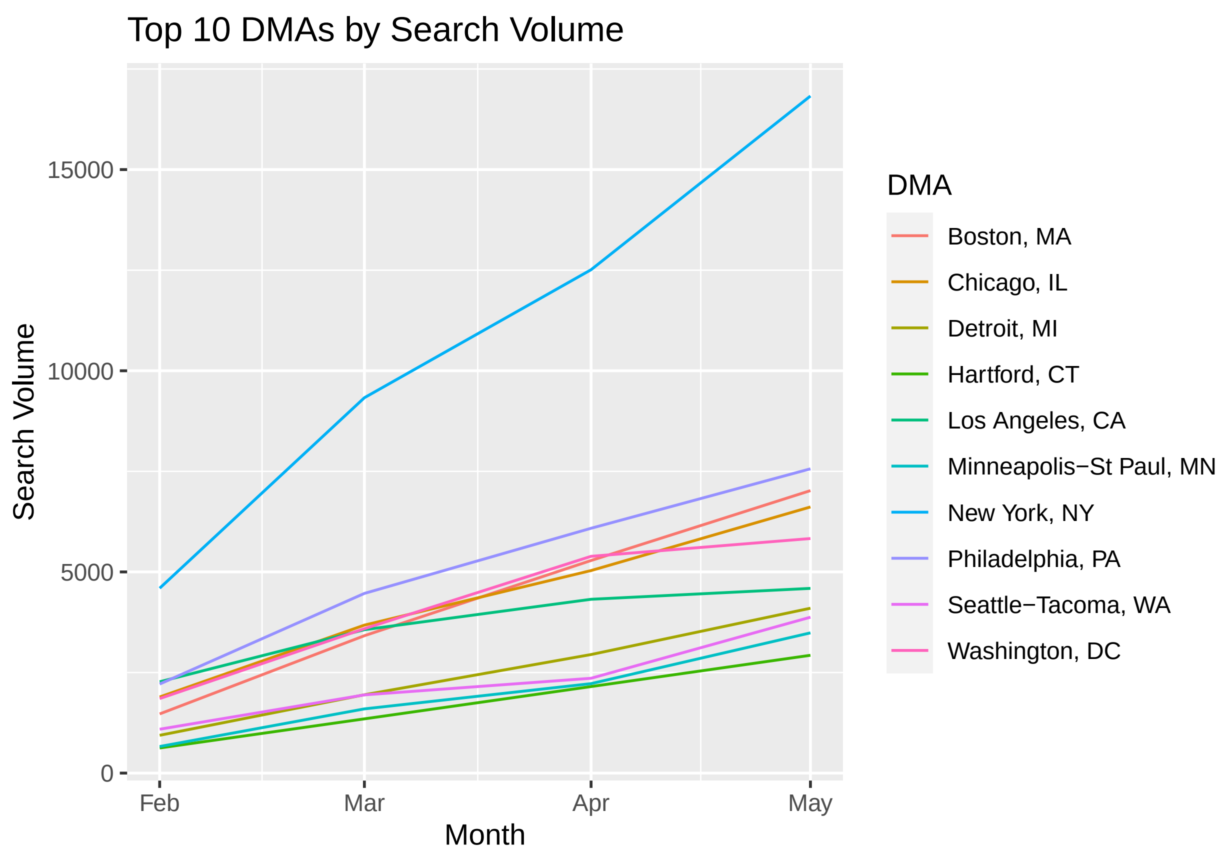
Line graph of 10 DMAs by Search Volume

```
library(ggplot2)

# Assuming you have a dataframe called `largest_changes` containing the data

# Convert the month columns to long format
largest_changes_long <- tidyr::gather(largest_changes, Month, SearchVolume, -DMA)

# Convert the Month column to a date format
largest_changes_long$Month <- lubridate::ymd(paste0(largest_changes_long$Month, "-01"))
```

```
## Warning: 10 failed to parse.
```

```
# Plot the line chart
ggplot(largest_changes_long, aes(x = Month, y = SearchVolume, group = DMA, color = DMA)) +
  geom_line() +
  labs(title = "Top 10 DMAs by Search Volume", x = "Month", y = "Search Volume")
```

```
## Warning: Removed 10 rows containing missing values (`geom_line()`).
```



This is a line graph showing the top 10 DMA changes by search Volume. As you can see New Yorks line is way above the rest. Hartford CT has the smallest changes.

Similar analysis to the percentage changes with the line chart, we'll do this for four geographical regions to compare differences.

Regions: New England (Boston...), Southeast (Miami...), Southwest (LA...), Midwest (Chicago...)

Calculate the change in visit count from February to the present for each DMA

```
dma_change <- dma_visit_counts %>%
  pivot_wider(names_from = MonthYear, values_from = VisitCount) %>%
  select(-`2023-06`) %>%
  mutate(Change = `2023-05` - `2023-02`) %>%
  arrange(desc(Change))

dma_change <- na.omit(dma_change)

dma_change
```

```
## # A tibble: 209 x 6
## # Groups:   DMA [209]
##    DMA                         `2023-02` `2023-03` `2023-04` `2023-05` Change
##    <chr>                           <int>     <int>     <int>     <int>  <int>
```

```
##   1 New York, NY                  4595      9330     12511     16829  12234
##   2 Boston, MA                    1471      3415      5284      7021   5550
##   3 Philadelphia, PA              2213      4466      6085      7563   5350
##   4 Chicago, IL                   1892      3678      5034      6615   4723
##   5 Washington, DC                1852      3589      5388      5830   3978
##   6 Detroit, MI                    939      1947      2947      4098   3159
##   7 Minneapolis-St Paul, MN        661      1596      2224      3486   2825
##   8 Seattle-Tacoma, WA            1090      1944      2357      3875   2785
##   9 Los Angeles, CA               2270      3564      4320      4591   2321
##  10 Hartford, CT                   624      1348      2151      2928   2304
## # i 199 more rows
```

```r
# New England DMA codes
new_england_dma <- c("Boston, MA", "Hartford, CT", "Providence, RI","Portland-Auburn, ME",
"Springfield-Holyoke, MA")
new_england_dma
```

```
## [1] "Boston, MA"              "Hartford, CT"
## [3] "Providence, RI"          "Portland-Auburn, ME"
## [5] "Springfield-Holyoke, MA"
```

```r
# Southeast DMA codes
southeast_dma <- c("New York, NY", "Philadelphia, PA", "Pittsburgh, PA", "Atlanta, GA","Baltimore, MD" )
southeast_dma
```

```
## [1] "New York, NY"     "Philadelphia, PA" "Pittsburgh, PA"   "Atlanta, GA"
## [5] "Baltimore, MD"
```

```r
# Southwest DMA codes
southwest_dma <- c("Los Angeles, CA", "San Francisco, CA","Sacramento, CA", "Phoenix, AZ", "Albuquerque
southwest_dma
```

```
## [1] "Los Angeles, CA"   "San Francisco, CA" "Sacramento, CA"
## [4] "Phoenix, AZ"       "Albuquerque, NM"
```

```r
# Midwest DMA codes
midwest_dma <- c("Chicago, IL", "Detroit, MI", "Minneapolis-St Paul, MN","Cleveland, OH", "Indianapolis
midwest_dma
```

```
## [1] "Chicago, IL"             "Detroit, MI"
## [3] "Minneapolis-St Paul, MN" "Cleveland, OH"
## [5] "Indianapolis, IN"
```

This code arranged the DMAs into regions, New England, Southeast, Southwest, and Midwest. I've included the top 5 DMAs from each location.

```r
# Filter the data based on DMA codes for each region
new_england_data <- dma_change[dma_change$DMA %in% new_england_dma, ]
southeast_data <- dma_change[dma_change$DMA %in% southeast_dma, ]
southwest_data <- dma_change[dma_change$DMA %in% southwest_dma, ]
midwest_data <- dma_change[dma_change$DMA %in% midwest_dma, ]
```

```r
# Rank the data by absolute search volume for each region
new_england_ranked <- new_england_data %>%
  arrange(desc(`2023-05`)) %>%
  head(5) %>%
  select(-Change)

new_england_ranked
```

```
## # A tibble: 5 x 5
## # Groups:   DMA [5]
##   DMA                     `2023-02` `2023-03` `2023-04` `2023-05`
##   <chr>                       <int>     <int>     <int>     <int>
## 1 Boston, MA                   1471      3415      5284      7021
## 2 Hartford, CT                  624      1348      2151      2928
## 3 Providence, RI                385       829      1165      1598
## 4 Portland-Auburn, ME           191       435       631       845
## 5 Springfield-Holyoke, MA       196       366       575       685
```

```
southeast_ranked <- southeast_data %>%
  arrange(desc(`2023-05`)) %>%
  head(5) %>%
  select(-Change)
southeast_ranked
```

```
## # A tibble: 5 x 5
## # Groups:   DMA [5]
##   DMA              `2023-02` `2023-03` `2023-04` `2023-05`
##   <chr>                <int>     <int>     <int>     <int>
## 1 New York, NY          4595      9330     12511     16829
## 2 Philadelphia, PA      2213      4466      6085      7563
## 3 Atlanta, GA           1543      2865      2728      3414
## 4 Pittsburgh, PA         990      1594      2365      2986
## 5 Baltimore, MD          615      1206      1566      1848
```

```
southwest_ranked <- southwest_data %>%
  arrange(desc(`2023-05`)) %>%
  head(5)%>%
  select(-Change)
southwest_ranked
```

```
## # A tibble: 5 x 5
## # Groups:   DMA [5]
##   DMA               `2023-02` `2023-03` `2023-04` `2023-05`
##   <chr>                 <int>     <int>     <int>     <int>
## 1 Los Angeles, CA        2270      3564      4320      4591
## 2 San Francisco, CA      1126      1836      2387      3035
## 3 Sacramento, CA          632      1012      1348      1735
## 4 Phoenix, AZ             789      1259      1490      1727
## 5 Albuquerque, NM         138       200       277       356
```

```
midwest_ranked <- midwest_data %>%
  arrange(desc(`2023-05`)) %>%
  head(5)%>%
  select(-Change)
midwest_ranked
```

```
## # A tibble: 5 x 5
## # Groups:   DMA [5]
##   DMA                     `2023-02` `2023-03` `2023-04` `2023-05`
##   <chr>                       <int>     <int>     <int>     <int>
## 1 Chicago, IL                  1892      3678      5034      6615
## 2 Detroit, MI                   939      1947      2947      4098
## 3 Minneapolis-St Paul, MN       661      1596      2224      3486
```

```
## 4 Cleveland, OH                  945       1598       2365       2826
## 5 Indianapolis, IN               559       1011       1287       1663
```

New England Region:

Boston, MA: The search volume for Boston shows an increasing trend over the months, reaching a peak in May with a search volume of 7,021.

Hartford, CT: Hartford also shows a steady increase in search volume, reaching 2,928 in May.

Providence, RI: Providence follows a similar pattern with a search volume of 1,598 in May.

Portland-Auburn, ME: Portland-Auburn experiences a moderate increase in search volume, reaching 845 in May.

Springfield-Holyoke, MA: Springfield-Holyoke shows a gradual increase in search volume, reaching 685 in May.

Southeast Region:

New York, NY: New York has the highest search volume among the Southeast DMAs, peaking at 16,829 in May.

Philadelphia, PA: Philadelphia follows a similar pattern, with a search volume of 7,563 in May.

Atlanta, GA: Atlanta exhibits a moderate increase in search volume, reaching 3,414 in May.

Pittsburgh, PA: Pittsburgh experiences a gradual increase, reaching 2,986 in May.

Baltimore, MD: Baltimore shows a steady increase in search volume, reaching 1,848 in May.

Southwest Region:

Los Angeles, CA: Los Angeles has the highest search volume among the Southwest DMAs, peaking at 4,591 in May.

San Francisco, CA: San Francisco follows a similar pattern, with a search volume of 3,035 in May.

Sacramento, CA: Sacramento exhibits a moderate increase in search volume, reaching 1,735 in May.

Phoenix, AZ: Phoenix experiences a gradual increase, reaching 1,727 in May.

Albuquerque, NM: Albuquerque shows a steady increase in search volume, reaching 356 in May.

Midwest Region:

Chicago, IL: Chicago has the highest search volume among the Midwest DMAs, peaking at 6,615 in May.

Detroit, MI: Detroit follows a similar pattern, with a search volume of 4,098 in May.

Minneapolis-St Paul, MN: Minneapolis-St Paul exhibits a moderate increase in search volume, reaching 3,486 in May.

Cleveland, OH: Cleveland experiences a gradual increase, reaching 2,826 in May.

Indianapolis, IN: Indianapolis shows a steady increase in search volume, reaching 1,663 in May.

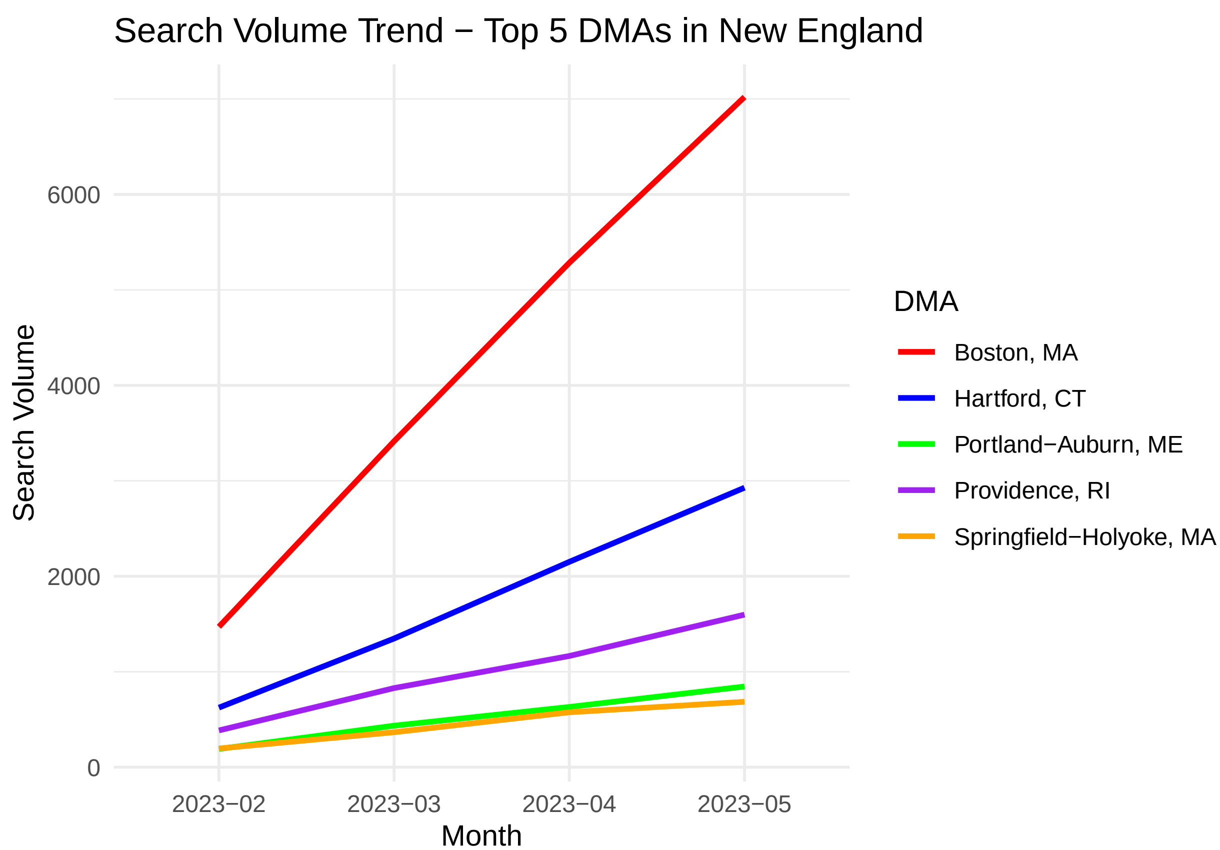
Overall, the data sets provide insights into the search volume trends and the top-performing DMAs in each region. These trends can be further analyzed to understand the patterns and dynamics of user interest and activity within specific geographic areas.

Line chart for New England

```r
new_england_ranked_long <- new_england_ranked %>%
  tidyr::pivot_longer(cols = -DMA, names_to = "Month", values_to = "Search_Volume")
```

```
# Convert Month column to character if necessary
new_england_ranked_long$Month <- as.character(new_england_ranked_long$Month)

# Ensure Search_Volume column is numeric
new_england_ranked_long$Search_Volume <- as.numeric(new_england_ranked_long$Search_Volume)

# Define custom colors for the DMAs
DMA_colors <- c("Boston, MA" = "red", "Hartford, CT" = "blue", "Portland-Auburn, ME" = "green", "Provid

ggplot(new_england_ranked_long, aes(x = Month, y = Search_Volume, color = DMA, group = DMA)) +
  geom_line(size = 1) +
  labs(title = "Search Volume Trend - Top 5 DMAs in New England",
       x = "Month",
       y = "Search Volume") +
  scale_color_manual(values = DMA_colors) +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



X-axis: The X-axis represents the different months and percent change (February, March, April, May and the change) within the New England region.

Y-axis: The Y-axis represents the search volume, which indicates the level of interest or activity related to a specific topic or keyword.

Lines: Each line represents the search volume trend over time for a particular month. The lines represent the DMAs (Boston, Hartford, Portland-Auburn, Providence, and Springfield-Holyoke) within the New England region.

Legend: The legend displays the color-coded representation of the DMAs (Boston, Hartford, Portland-Auburn, Providence, and Springfield-Holyoke) within the New England region.

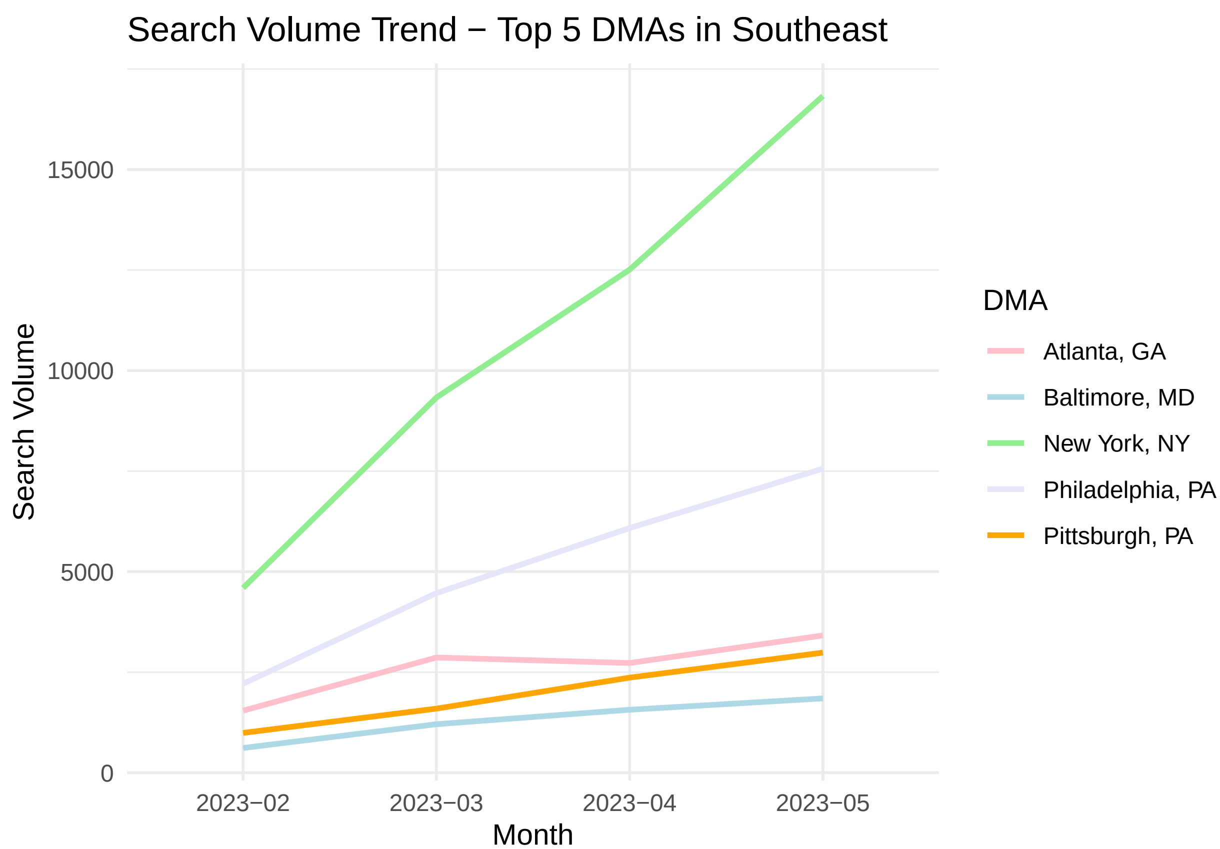
Title: The title of the graph is "Search Volume Trend - Top 5 DMAs in New England"

Axes labels: The X-axis is labeled as "Month," and the Y-axis is labeled as "Search Volume."

Color scheme: The color scheme used for the lines representing different DMAs is customized. In the provided code, the colors used are red for Boston, blue for Hartford, green for Portland-Auburn, and purple for Providence and orange for Springfield-Holyoke.

Line chart for Southeast

```
southeast_ranked_long <- southeast_ranked %>%
  tidyr::pivot_longer(cols = -DMA, names_to = "Month", values_to = "Search_Volume")

# Convert Month column to character if necessary
southeast_ranked_long$Month <- as.character(southeast_ranked_long$Month)

# Ensure Search_Volume column is numeric
southeast_ranked_long$Search_Volume <- as.numeric(southeast_ranked_long$Search_Volume)


# Define custom colors for the DMAs
DMA_colors <- c("Atlanta, GA" = "pink", "Baltimore, MD" = "lightblue", "New York, NY" = "lightgreen", "

ggplot(southeast_ranked_long, aes(x = Month, y = Search_Volume, color = DMA, group = DMA)) +
  geom_line(size = 1) +
  labs(title = "Search Volume Trend - Top 5 DMAs in Southeast",
       x = "Month",
       y = "Search Volume") +
  scale_color_manual(values = DMA_colors) +
  theme_minimal()
```



X-axis: The X-axis represents the different months and percent change (February, March, April, May and the change)

Y-axis: The Y-axis represents the search volume, which indicates the level of interest or activity related to a specific topic or keyword.

Lines: Each line represents the search volume trend over time for a particular month. The lines represent the DMAS (Atlanta, Baltimore, New York, Philadelphia, and Pittsburg) within the Southeast region.

Legend: The legend displays the color-coded representation of the DMAs (Atlanta, Baltimore, New York, Philadelphia, and Pittsburg) within the Southeast region. Title: The title of the graph is “Search Volume Trend - Top 5 DMAs in Southeast”

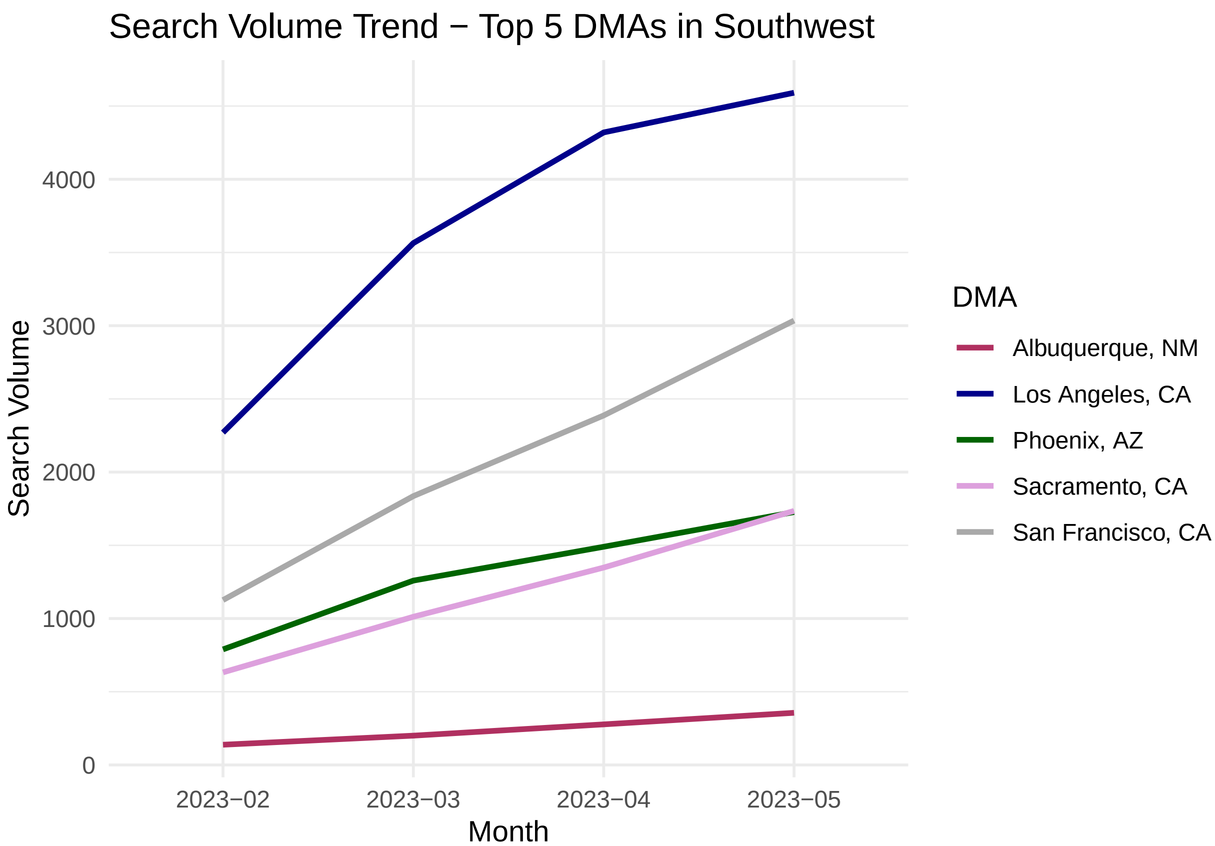
Axes labels: The X-axis is labeled as “Month,” and the Y-axis is labeled as “Search Volume.”

Color scheme: The color scheme used for the lines representing different DMAs is customized. In the provided code, the colors used are pink for Atlanta, lightblue for Baltimore, lightgreen for New York, and lavender for Philadelphia and orange for Pittsburgh.

Bar chart for Southwest

```
southwest_ranked_long <- southwest_ranked %>%
  tidyr::pivot_longer(cols = -DMA, names_to = "Month", values_to = "Search_Volume")

# Convert Month column to character if necessary
southwest_ranked_long$Month <- as.character(southwest_ranked_long$Month)

# Ensure Search_Volume column is numeric
```

```
southeast_ranked_long$Search_Volume <- as.numeric(southeast_ranked_long$Search_Volume)

# Define custom colors for the DMAs
DMA_colors <- c("Albuquerque, NM" = "maroon", "Los Angeles, CA" = "darkblue", "Phoenix, AZ" = "darkgreen

ggplot(southwest_ranked_long, aes(x = Month, y = Search_Volume, color = DMA, group = DMA)) +
  geom_line(size = 1) +
  labs(title = "Search Volume Trend - Top 5 DMAs in Southwest",
       x = "Month",
       y = "Search Volume") +
  scale_color_manual(values = DMA_colors) +
  theme_minimal()
```



X-axis: The X-axis represents the different months and percent change (February, March, April, May and the change)

Y-axis: The Y-axis represents the search volume, which indicates the level of interest or activity related to a specific topic or keyword.

Lines: Each line represents the search volume trend over time for a particular month. The lines represents the DMAs (Albuquerque, Los Angeles, Phoenix, Sacramento, and San Francisco) within the Midwest region.

Legend: The legend displays the color-coded representation of the DMAs (Albuquerque, Los Angeles, Phoenix, Sacramento, and San Francisco) within the Midwest region.

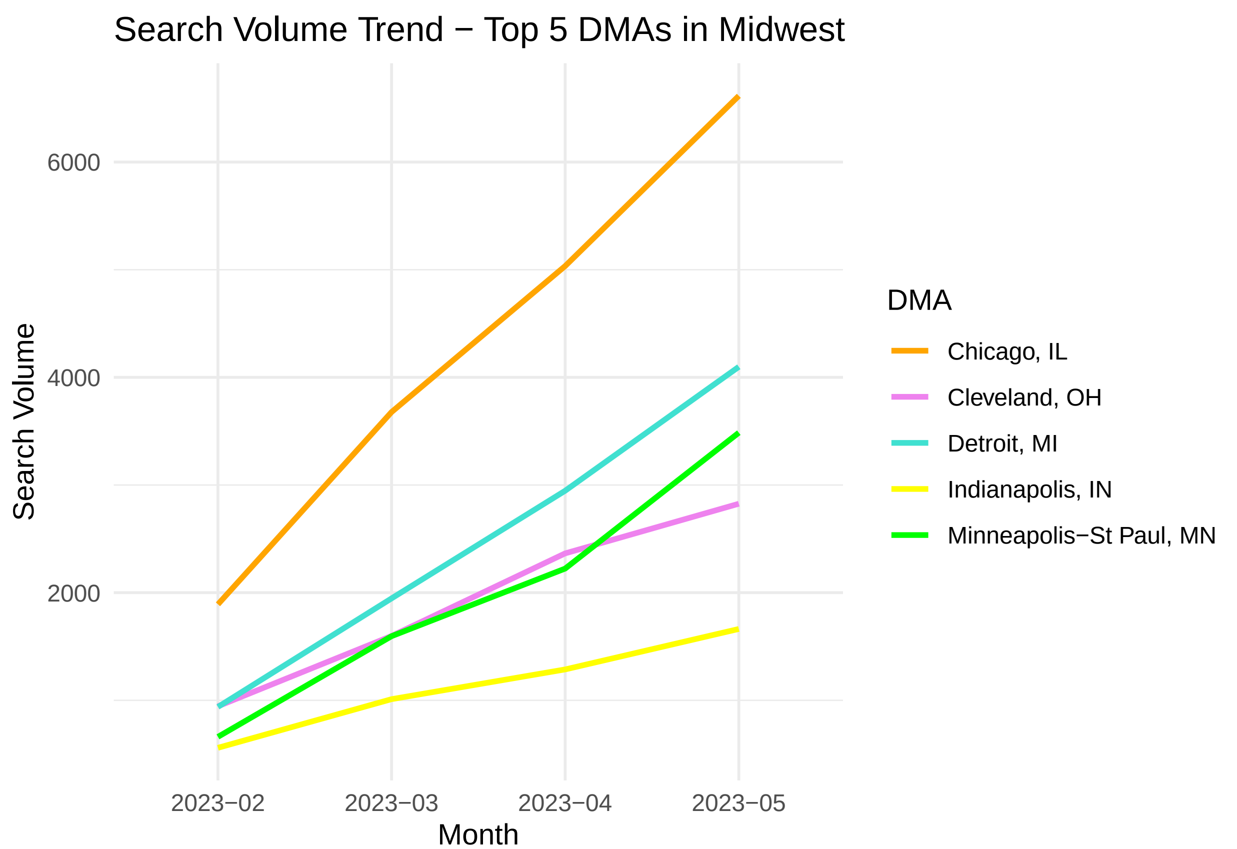
Title: The title of the graph is “Search Volume Trend - Top 5 DMAs in Southwest”

Axes labels: The X-axis is labeled as “Month,” and the Y-axis is labeled as “Search Volume.”

Color scheme: The color scheme used for the lines representing different DMAs is customized. In the provided code, the colors used are maroon for Albuquerque, NM, darkblue for Los Angeles, CA, darkgreen for Phoenix, AZ, and plum for Sacramento, CA and darkgrey for San Francisco, CA.

Line chart for Midwest

```
midwest_ranked_long <- midwest_ranked %>%
  tidyr::pivot_longer(cols = -DMA, names_to = "Month", values_to = "Search_Volume")

# Convert Month column to character if necessary
midwest_ranked_long$Month <- as.character(midwest_ranked_long$Month)

# Ensure Search_Volume column is numeric
midwest_ranked_long$Search_Volume <- as.numeric(midwest_ranked_long$Search_Volume)

# Define custom colors for the DMAs
DMA_colors <- c("Chicago, IL" = "orange", "Cleveland, OH" = "violet", "Detroit, MI" = "turquoise", "Indi

ggplot(midwest_ranked_long, aes(x = Month, y = Search_Volume, color = DMA, group = DMA)) +
  geom_line(size = 1) +
  labs(title = "Search Volume Trend - Top 5 DMAs in Midwest",
       x = "Month",
       y = "Search Volume") +
  scale_color_manual(values = DMA_colors) +
  theme_minimal()
```



X-axis: The X-axis represents different months and percent change(February, March, April, May and the change)

Y-axis: The Y-axis represents the search volume, which indicates the level of interest or activity related to a specific topic or keyword.

Lines: Each line represents the search volume trend over time for a particular month. The lines represents the DMAs (Chicago, Detroit, Minneapolis-St Paul, Cleveland, and Indianapolis) within the Midwest region.

Legend: The legend displays the color-coded representation of the DMAs (Chicago, Detroit, Minneapolis-St Paul, Cleveland, and Indianapolis) within the Midwest region

Title: The title of the graph is "Search Volume Trend - Top 5 DMAs in Midwest."

Axes labels: The X-axis is labeled as "Month," and the Y-axis is labeled as "Search Volume."

Color scheme: The color scheme used for the lines representing different DMAs is customized with the scale_color_manual function. In the provided code, the colors used are orange for Chicago, violet for Cleveland, turquoise for Detroit, and yellow for Indianapolis, and Green for Minneapolis-St Paul.

Brief Blurb

New England:

The top 5 DMAs in New England based on search volume in May 2023 are Boston, MA; Hartford, CT; Providence, RI; Portland-Auburn, ME; and Springfield-Holyoke, MA.

Boston, MA has the highest search volume throughout all the months, with a significant increase in search volume from February to May 2023.

Southeast:

The top 5 DMAs in the Southeast region based on search volume in May 2023 are New York, NY; Philadelphia, PA; Atlanta, GA; Pittsburgh, PA; and Baltimore, MD.

New York, NY has the highest search volume among all the DMAs, with a substantial increase in search volume from February to May 2023.

Southwest:

The top 5 DMAs in the Southwest region based on search volume in May 2023 are Los Angeles, CA; San Francisco, CA; Sacramento, CA; Phoenix, AZ; and Albuquerque, NM.

Los Angeles, CA has the highest search volume among the DMAs, with a noticeable decrease in search volume from April to May 2023.

Midwest:

The top 5 DMAs in the Midwest region based on search volume in May 2023 are Chicago, IL; Detroit, MI; Minneapolis-St Paul, MN; Cleveland, OH; and Indianapolis, IN.

Chicago, IL has the highest search volume throughout all the months, with a substantial increase in search volume from February to May 2023.

Overall, the findings suggest varying levels of interest or activity in different regions and DMAs. Certain areas, such as Boston, MA in New England and New York, NY in the Southeast, exhibit consistently high search volumes, while other regions experience fluctuations in search volume over time. These insights can help understand the geographic trends and preferences related to the analyzed topic or keyword.