Assignment2

Leng Seong Che

2023-09-20

You may work in pairs or individually for this assignment. Make sure you join a group in Canvas if you are working in pairs. Turn in this assignment as an HTML or PDF file to ELMS. Make sure to include the R Markdown or Quarto file that was used to generate it.

```
library(tidyverse)
library(gtrendsR)
library(censusapi)
```

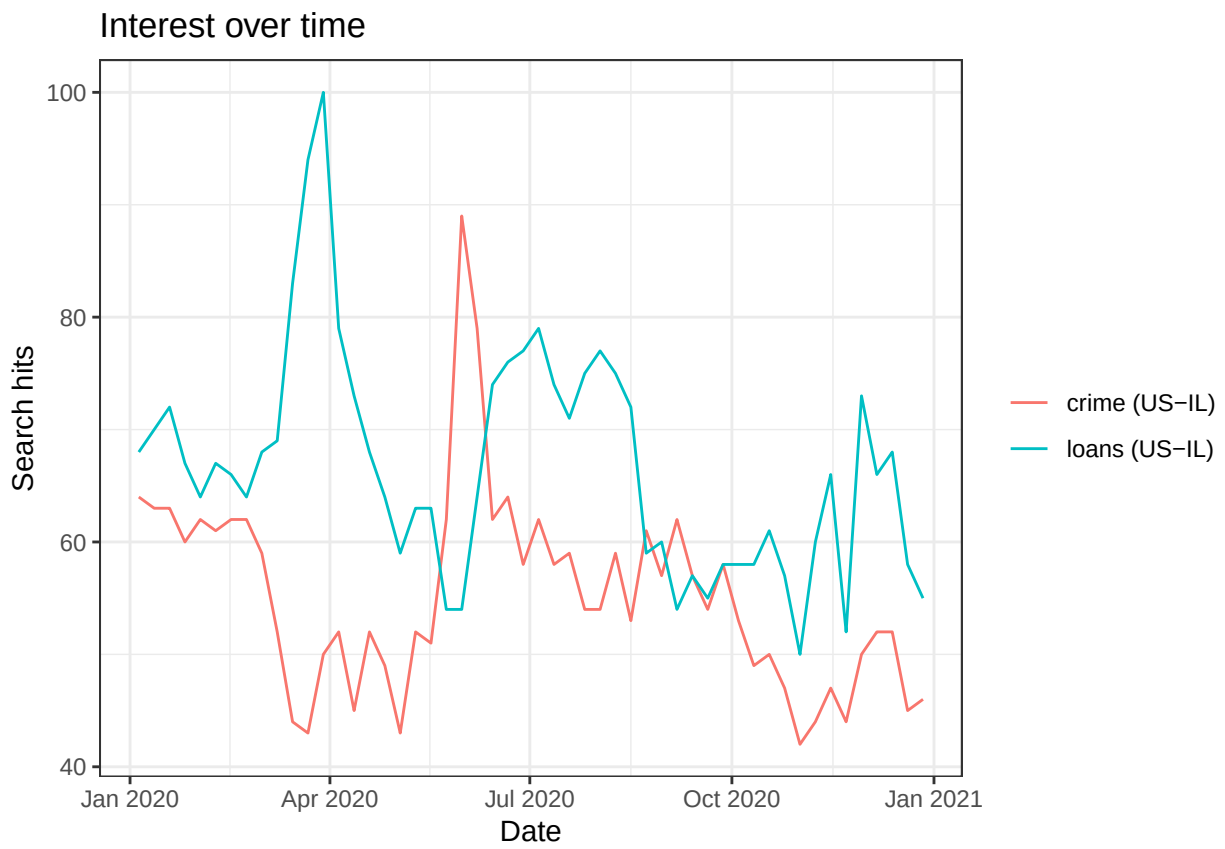
In this assignment, you will pull from APIs to get data from various data sources and use your data wrangling skills to use them all together. You should turn in a report in PDF or HTML format that addresses all of the questions in this assignment, and describes the data that you pulled and analyzed. You do not need to include full introduction and conclusion sections like a full report, but you should make sure to answer the questions in paragraph form, and include all relevant tables and graphics.

Whenever possible, use piping and `dplyr`. Avoid hard-coding any numbers within the report as much as possible.

Pulling from APIs

Our first data source is the Google Trends API. Suppose we are interested in the search trends for `crime` and `loans` in Illinois in the year 2020. We could find this using the following code:

```
res <- gtrends(c("crime", "loans"),
               geo = "US-IL",
               time = "2020-01-01 2020-12-31",
               low_search_volume = TRUE)
plot(res)
```



Answer the following questions for the keywords "crime" and "loans".

- Find the mean, median and variance of the search hits for the keywords.

```
res_time <- as_tibble(res$interest_over_time)
res_time %>%
  group_by(keyword)  %>%
  summarise(mean = mean(hits),
            median = median(hits),
            variance = sd(hits)**2)
```

```
## # A tibble: 2 x 4
##   keyword  mean median variance
##   <chr>   <dbl>  <dbl>    <dbl>
## 1 crime    55.2     54     78.4
## 2 loans    66.7     66    102.
```

The mean, median, and variance for "crime" are 55, 54, and 86.4 respectively and for "loans" are 66.5, 65, and 95.39 respectively.

- Which cities (locations) have the highest search frequency for `loans`? Note that there might be multiple rows for each city if there were hits for both "crime" and "loans" in that city. It might be easier to answer this question if we had the search hits info for both search terms in two separate variables. That is, each row would represent a unique city.

```
#res_time_crime_loan <- spread(res_time, key = keyword, value = hits)
# cor(res_time_crime_loan$crime, res_time_crime_loan$loan)

res_city <- spread(res$interest_by_city, key = keyword, value = hits)

# find max
res_city$location[which.max(res_city$loans)]
```

```
## [1] "Alorton"
```

```
# sort res_city in descending order
arrange(res_city, desc(loans))
```

```
##               location   geo gprop crime loans
## 1              Alorton US-IL   web    NA   100
## 2            Roseville US-IL   web    NA    98
## 3                Henry US-IL   web    NA    77
## 4          Cerro Gordo US-IL   web    NA    74
## 5            Long Lake US-IL   web    NA    71
## 6             Rosemont US-IL   web    62    66
## 7                Warsaw US-IL   web    NA    58
## 8           Braceville US-IL   web    NA    57
## 9                Fulton US-IL   web    NA    56
## 10           Coal City US-IL   web    28    54
## 11          Carrollton US-IL   web    NA    52
## 12              Bement US-IL   web    NA    51
## 13              Dolton US-IL   web    55    51
## 14          Georgetown US-IL   web    NA    51
## 15        Channel Lake US-IL   web    NA    50
## 16           Nashville US-IL   web    NA    50
## 17             Peotone US-IL   web    NA    50
## 18                Witt US-IL   web    NA    50
## 19             Robbins US-IL   web    NA    47
## 20         Hazel Crest US-IL   web    NA    46
## 21        Calumet City US-IL   web    NA    45
## 22           Lewistown US-IL   web    NA    45
## 23           Riverdale US-IL   web    60    45
## 24      Olympia Fields US-IL   web    NA    44
## 25          Beach Park US-IL   web    NA    43
## 26            Sandoval US-IL   web    NA    43
## 27           Cambridge US-IL   web    NA    42
## 28    East Saint Louis US-IL   web    79    42
## 29             Danville US-IL   web    NA    41
## 30          New Athens US-IL   web    NA    41
## 31               Union US-IL   web    NA    41
## 32              Harvey US-IL   web    NA    40
## 33             Oakwood US-IL   web    NA    40
## 34              Canton US-IL   web    NA    39
## 35       South Holland US-IL   web    NA    39
## 36             Chester US-IL   web    NA    38
## 37             El Paso US-IL   web    NA    38
## 38             Lansing US-IL   web    NA    38
```

```
## 39        Sauk Village US-IL   web    NA    38
## 40         Bartonville US-IL   web    NA    37
## 41          New Boston US-IL   web    NA    37
## 42      West Frankfort US-IL   web    40    37
## 43          Bridgeport US-IL   web    NA    36
## 44          Brookfield US-IL   web    NA    36
## 45            Heyworth US-IL   web    NA    36
## 46      Kingston Mines US-IL   web    NA    36
## 47     University Park US-IL   web    NA    36
## 48            Hinckley US-IL   web    NA    35
## 49           Maryville US-IL   web    54    35
## 50             Savanna US-IL   web    NA    35
## 51            Smithton US-IL   web    NA    35
## 52     Stillman Valley US-IL   web    NA    35
## 53              Virden US-IL   web    NA    35
## 54               Flora US-IL   web    NA    34
## 55            Freeburg US-IL   web    NA    34
## 56          Harristown US-IL   web    NA    34
## 57               Orion US-IL   web    NA    34
## 58          Wood River US-IL   web    NA    34
## 59              Dwight US-IL   web    NA    33
## 60    Fairview Heights US-IL   web    NA    33
## 61          Belleville US-IL   web    66    32
## 62          Crainville US-IL   web    NA    32
## 63      Lake Summerset US-IL   web    NA    32
## 64        Mount Carmel US-IL   web    NA    32
## 65       North Chicago US-IL   web    NA    32
## 66    Oakbrook Terrace US-IL   web    NA    32
## 67            Robinson US-IL   web    NA    32
## 68              Sparta US-IL   web    NA    32
## 69            Abingdon US-IL   web    NA    31
## 70         Bourbonnais US-IL   web    NA    31
## 71       Johnston City US-IL   web    NA    31
## 72              McCook US-IL   web    NA    31
## 73          Metropolis US-IL   web    NA    31
## 74         Murphysboro US-IL   web    60    31
## 75           Okawville US-IL   web    NA    31
## 76            Streator US-IL   web    46    31
## 77              Arcola US-IL   web    NA    30
## 78             Bradley US-IL   web    64    30
## 79          Farmington US-IL   web    NA    30
## 80          Greenville US-IL   web    NA    30
## 81              Joliet US-IL   web    NA    30
## 82           Jonesboro US-IL   web    NA    30
## 83      Peoria Heights US-IL   web    NA    30
## 84         Bunker Hill US-IL   web    NA    29
## 85          Shullsburg US-IL   web    NA    29
## 86           Channahon US-IL   web    51    28
## 87               Lacon US-IL   web    NA    27
## 88         Rock Island US-IL   web    NA    27
## 89            Staunton US-IL   web    NA    27
## 90              Chenoa US-IL   web    NA    26
## 91         Christopher US-IL   web    NA    26
## 92              Delavan US-IL   web    NA    26
```

```
## 93              Eureka US-IL   web    NA    26
## 94           Fairfield US-IL   web    NA    26
## 95          Park Ridge US-IL   web    NA    26
## 96        Plato Center US-IL   web    NA    26
## 97          Caseyville US-IL   web    NA    25
## 98         Chillicothe US-IL   web    NA    25
## 99         Coal Valley US-IL   web    NA    25
## 100     Lawrenceville US-IL   web    NA    25
## 101             Olney US-IL   web    NA    25
## 102     Pinckneyville US-IL   web    NA    25
## 103       Shelbyville US-IL   web    68    25
## 104     Sleepy Hollow US-IL   web    NA    24
## 105     Carrier Mills US-IL   web    NA    23
## 106            Newton US-IL   web    NA    23
## 107        Palos Park US-IL   web    57    23
## 108             Posen US-IL   web    NA    23
## 109         Winnebago US-IL   web    NA    23
## 110            Auburn US-IL   web    NA    22
## 111       Boulder Hill US-IL   web    NA    22
## 112         Monticello US-IL   web    NA    22
## 113          New Baden US-IL   web    NA    22
## 114    Fox River Grove US-IL   web    NA    21
## 115        Willowbrook US-IL   web    55    21
## 116     Cherry Valley US-IL   web    NA    20
## 117          Fairbury US-IL   web    NA    20
## 118 Lake of the Woods US-IL   web    NA    20
## 119   Winthrop Harbor US-IL   web    NA    20
## 120    Davis Junction US-IL   web    NA    19
## 121         Hoopeston US-IL   web    NA    19
## 122          Metamora US-IL   web    NA    18
## 123      Morton Grove US-IL   web    NA    18
## 124          Gilberts US-IL   web    42    17
## 125           Oglesby US-IL   web    NA    17
## 126  South Barrington US-IL   web    NA    17
## 127           Cambria US-IL   web    NA    15
## 128        Lake Bluff US-IL   web    NA    15
## 129             Amboy US-IL   web    NA    14
## 130          Argenta US-IL   web    NA    14
## 131         Gillespie US-IL   web    NA    14
## 132            Mokena US-IL   web    NA    14
## 133             Wayne US-IL   web    NA    14
## 134              Anna US-IL   web   100    12
## 135             Niles US-IL   web    NA    10
## 136          Brighton US-IL   web    NA     8
## 137             Aledo US-IL   web    26    NA
## 138          Allerton US-IL   web    NA    NA
## 139          Altamont US-IL   web    NA    NA
## 140            Ancona US-IL   web    NA    NA
## 141           Ashland US-IL   web    NA    NA
## 142           Atlanta US-IL   web    NA    NA
## 143            Atwood US-IL   web    NA    NA
## 144       Bannockburn US-IL   web    NA    NA
## 145      Bedford Park US-IL   web    52    NA
## 146           Berwick US-IL   web    NA    NA
```

```
## 147           Big Rock US-IL   web    NA    NA
## 148      Blandinsville US-IL   web    NA    NA
## 149        Bloomington US-IL   web    58    NA
## 150        Blue Island US-IL   web    47    NA
## 151         Blue Mound US-IL   web    NA    NA
## 152             Bluffs US-IL   web    NA    NA
## 153            Bluford US-IL   web    NA    NA
## 154              Bowen US-IL   web    NA    NA
## 155          Broadview US-IL   web    NA    NA
## 156            Brocton US-IL   web    NA    NA
## 157            Buffalo US-IL   web    NA    NA
## 158           Buncombe US-IL   web    NA    NA
## 159         Burlington US-IL   web    NA    NA
## 160            Burnham US-IL   web    NA    NA
## 161           Bushnell US-IL   web    NA    NA
## 162              Butler US-IL   web    NA    NA
## 163               Byron US-IL   web    NA    NA
## 164         Camp Point US-IL   web    NA    NA
## 165      Campbell Hill US-IL   web    NA    NA
## 166             Capron US-IL   web    NA    NA
## 167       Carbon Cliff US-IL   web    NA    NA
## 168        Carlinville US-IL   web    74    NA
## 169        Carterville US-IL   web    48    NA
## 170              Casey US-IL   web    NA    NA
## 171             Catlin US-IL   web    NA    NA
## 172            Chatham US-IL   web    NA    NA
## 173      Chicago Ridge US-IL   web    52    NA
## 174            Chrisman US-IL   web    NA    NA
## 175              Cobden US-IL   web    NA    NA
## 176              Colusa US-IL   web    NA    NA
## 177             Cornell US-IL   web    NA    NA
## 178            Cortland US-IL   web    NA    NA
## 179      Crescent City US-IL   web    NA    NA
## 180             Curran US-IL   web    NA    NA
## 181            De Soto US-IL   web    NA    NA
## 182          Deer Park US-IL   web    NA    NA
## 183              DeKalb US-IL   web    57    NA
## 184               DePue US-IL   web    NA    NA
## 185           Divernon US-IL   web    NA    NA
## 186              Dixon US-IL   web    42    NA
## 187              Downs US-IL   web    NA    NA
## 188     East Galesburg US-IL   web    NA    NA
## 189       Edwardsville US-IL   web    56    NA
## 190             Elburn US-IL   web    59    NA
## 191          Elizabeth US-IL   web    NA    NA
## 192      Elizabethtown US-IL   web    NA    NA
## 193             Elkhart US-IL   web    NA    NA
## 194            Elmhurst US-IL   web    59    NA
## 195              Elwood US-IL   web    44    NA
## 196             Enfield US-IL   web    NA    NA
## 197            Fairmont US-IL   web    NA    NA
## 198       Farmersville US-IL   web    NA    NA
## 199       Ford Heights US-IL   web    47    NA
## 200      Franklin Park US-IL   web    38    NA
```

```
## 201           Freeport US-IL   web    64    NA
## 202         Gages Lake US-IL   web    NA    NA
## 203            Galatia US-IL   web    NA    NA
## 204             Galena US-IL   web    59    NA
## 205              Galva US-IL   web    NA    NA
## 206             Geneva US-IL   web    NA    NA
## 207             Girard US-IL   web    NA    NA
## 208         Glen Carbon US-IL   web    52    NA
## 209            Glenwood US-IL   web    52    NA
## 210          Goodfield US-IL   web    NA    NA
## 211         Grand Tower US-IL   web    NA    NA
## 212          Grant Park US-IL   web    NA    NA
## 213          Grayville US-IL   web    NA    NA
## 214         Green Oaks US-IL   web    52    NA
## 215         Greenfield US-IL   web    NA    NA
## 216          Gulf Port US-IL   web    NA    NA
## 217        Hainesville US-IL   web    NA    NA
## 218          Hampshire US-IL   web    NA    NA
## 219            Hampton US-IL   web    NA    NA
## 220         Harrisburg US-IL   web    NA    NA
## 221            Harvard US-IL   web    37    NA
## 222             Havana US-IL   web    NA    NA
## 223             Hebron US-IL   web    95    NA
## 224             Herrin US-IL   web    45    NA
## 225           Highland US-IL   web    47    NA
## 226            Hoffman US-IL   web    NA    NA
## 227           Hoyleton US-IL   web    NA    NA
## 228               Hume US-IL   web    NA    NA
## 229         Illiopolis US-IL   web    NA    NA
## 230           Iroquois US-IL   web    NA    NA
## 231             Itasca US-IL   web    32    NA
## 232       Jacksonville US-IL   web    50    NA
## 233               Kane US-IL   web    NA    NA
## 234            Kincaid US-IL   web    NA    NA
## 235           Kingston US-IL   web    NA    NA
## 236          Knoxville US-IL   web    NA    NA
## 237    Lake Barrington US-IL   web    45    NA
## 238     Lake Catherine US-IL   web    NA    NA
## 239         Lake Villa US-IL   web    42    NA
## 240           Lakemoor US-IL   web    NA    NA
## 241    Lakewood Shores US-IL   web    NA    NA
## 242            LaSalle US-IL   web    26    NA
## 243         Leaf River US-IL   web    NA    NA
## 244            Lebanon US-IL   web    15    NA
## 245             Leland US-IL   web    NA    NA
## 246       Libertyville US-IL   web    53    NA
## 247         Litchfield US-IL   web    NA    NA
## 248         Long Creek US-IL   web    NA    NA
## 249     Machesney Park US-IL   web    61    NA
## 250              Macomb US-IL   web    85    NA
## 251            Makanda US-IL   web    NA    NA
## 252          Manhattan US-IL   web    55    NA
## 253            Marissa US-IL   web    NA    NA
## 254            Markham US-IL   web    42    NA
```

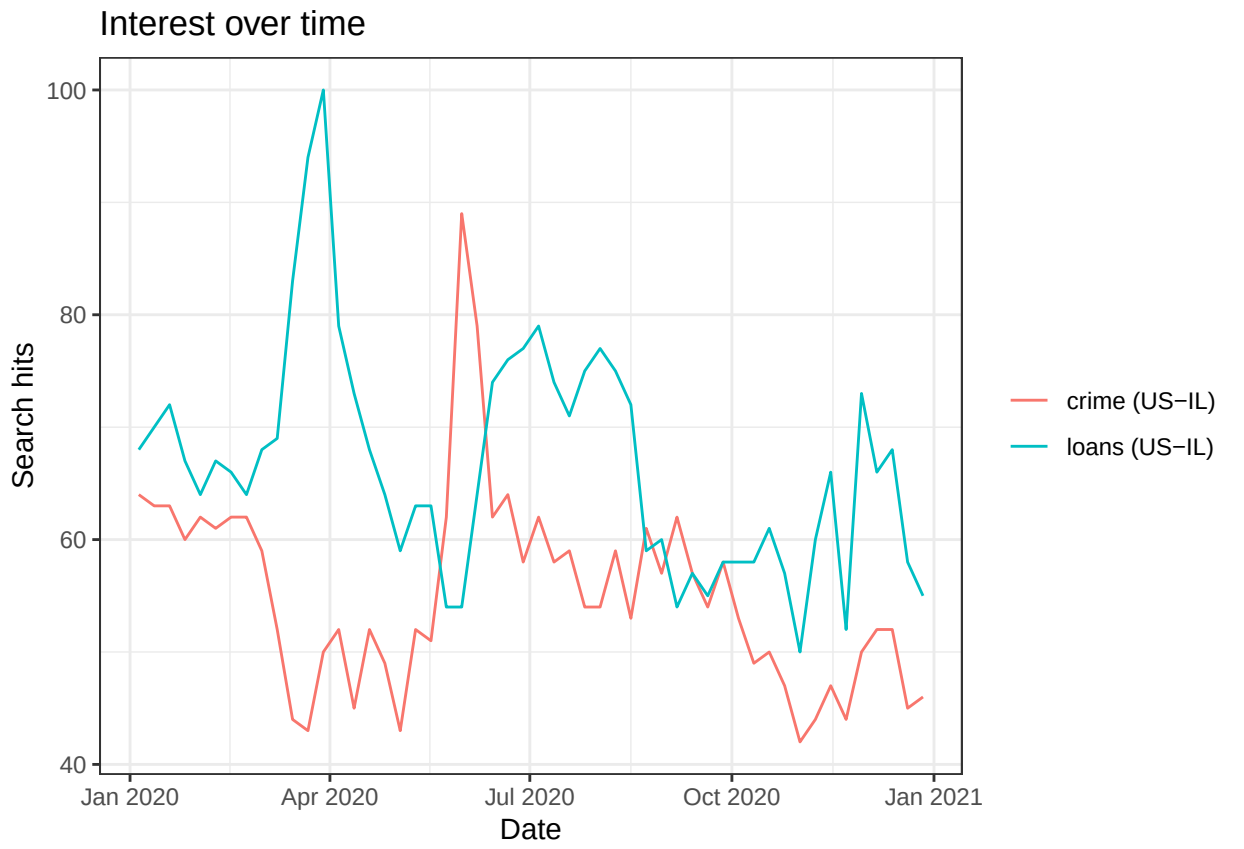
```
## 255              Maroa US-IL   web    NA    NA
## 256       Martinsville US-IL   web    NA    NA
## 257          Mascoutah US-IL   web    44    NA
## 258         Mason City US-IL   web    NA    NA
## 259              Mazon US-IL   web    NA    NA
## 260            Mendota US-IL   web    NA    NA
## 261            Mettawa US-IL   web    NA    NA
## 262           Monmouth US-IL   web    51    NA
## 263      Monroe Center US-IL   web    NA    NA
## 264     Mount Prospect US-IL   web    37    NA
## 265         Mount Zion US-IL   web    58    NA
## 266           Moweaqua US-IL   web    NA    NA
## 267               Nebo US-IL   web    NA    NA
## 268          New Lenox US-IL   web    51    NA
## 269            Niantic US-IL   web    NA    NA
## 270             Normal US-IL   web    63    NA
## 271        Norris City US-IL   web    NA    NA
## 272       North Aurora US-IL   web    NA    NA
## 273    North Riverside US-IL   web    82    NA
## 274        North Utica US-IL   web    NA    NA
## 275          Oak Grove US-IL   web    39    NA
## 276           Oak Park US-IL   web    62    NA
## 277            Oquawka US-IL   web    NA    NA
## 278        Park Forest US-IL   web    52    NA
## 279         Pearl City US-IL   web    NA    NA
## 280         Pittsfield US-IL   web    NA    NA
## 281            Pontiac US-IL   web    57    NA
## 282         Port Byron US-IL   web    NA    NA
## 283  Prairie du Rocher US-IL   web    NA    NA
## 284             Quincy US-IL   web    49    NA
## 285             Ramsey US-IL   web    NA    NA
## 286           Richview US-IL   web    NA    NA
## 287            Ridgway US-IL   web    NA    NA
## 288          Riverside US-IL   web    46    NA
## 289            Roanoke US-IL   web    NA    NA
## 290           Rockford US-IL   web    70    NA
## 291         Round Lake US-IL   web    36    NA
## 292       Saint Joseph US-IL   web    NA    NA
## 293       Saint Libory US-IL   web    NA    NA
## 294              Salem US-IL   web    41    NA
## 295              Savoy US-IL   web    58    NA
## 296           Saybrook US-IL   web    NA    NA
## 297      Schiller Park US-IL   web    61    NA
## 298             Seneca US-IL   web    NA    NA
## 299             Sesser US-IL   web    NA    NA
## 300           Shabbona US-IL   web    NA    NA
## 301          Sheffield US-IL   web    NA    NA
## 302            Sheldon US-IL   web    NA    NA
## 303          Shorewood US-IL   web    45    NA
## 304             Sidell US-IL   web    NA    NA
## 305            Simpson US-IL   web    NA    NA
## 306           Somonauk US-IL   web    NA    NA
## 307            Sorento US-IL   web    NA    NA
## 308        South Elgin US-IL   web    59    NA
```

```
## 309 South Jacksonville US-IL   web    63    NA
## 310        South Pekin US-IL   web    NA    NA
## 311       South Roxana US-IL   web    NA    NA
## 312      Spring Valley US-IL   web    NA    NA
## 313        Stronghurst US-IL   web    NA    NA
## 314        Sugar Grove US-IL   web    43    NA
## 315           Sullivan US-IL   web    NA    NA
## 316        Summerfield US-IL   web    NA    NA
## 317            Tamaroa US-IL   web    NA    NA
## 318           Thornton US-IL   web    NA    NA
## 319             Tilton US-IL   web    68    NA
## 320             Tolono US-IL   web    NA    NA
## 321             Toulon US-IL   web    NA    NA
## 322            Trenton US-IL   web    NA    NA
## 323            Trivoli US-IL   web    NA    NA
## 324               Troy US-IL   web    58    NA
## 325         Troy Grove US-IL   web    NA    NA
## 326              Ullin US-IL   web    NA    NA
## 327   Venetian Village US-IL   web    54    NA
## 328             Venice US-IL   web    NA    NA
## 329          Vermilion US-IL   web    NA    NA
## 330         Versailles US-IL   web    NA    NA
## 331           Victoria US-IL   web    NA    NA
## 332             Vienna US-IL   web    NA    NA
## 333              Viola US-IL   web    NA    NA
## 334             Walnut US-IL   web    NA    NA
## 335           Washburn US-IL   web    NA    NA
## 336           Waterloo US-IL   web    41    NA
## 337             Wenona US-IL   web    NA    NA
## 338       West Chicago US-IL   web    52    NA
## 339    Western Springs US-IL   web    60    NA
## 340            Wheaton US-IL   web    54    NA
## 341         White City US-IL   web    NA    NA
## 342         White Hall US-IL   web    NA    NA
## 343      Williamsville US-IL   web    NA    NA
## 344        Willisville US-IL   web    NA    NA
## 345         Wilmington US-IL   web    NA    NA
## 346         Winchester US-IL   web    NA    NA
## 347        Wonder Lake US-IL   web    45    NA
## 348              Worth US-IL   web    49    NA
## 349            Wyoming US-IL   web    NA    NA
## 350              Xenia US-IL   web    NA    NA
```

Midlothian has the highest search frequency (100) on “loans”, followed by Alorton (78) and Long Lake (62).

- Is there a relationship between the search intensities between the two keywords we used?

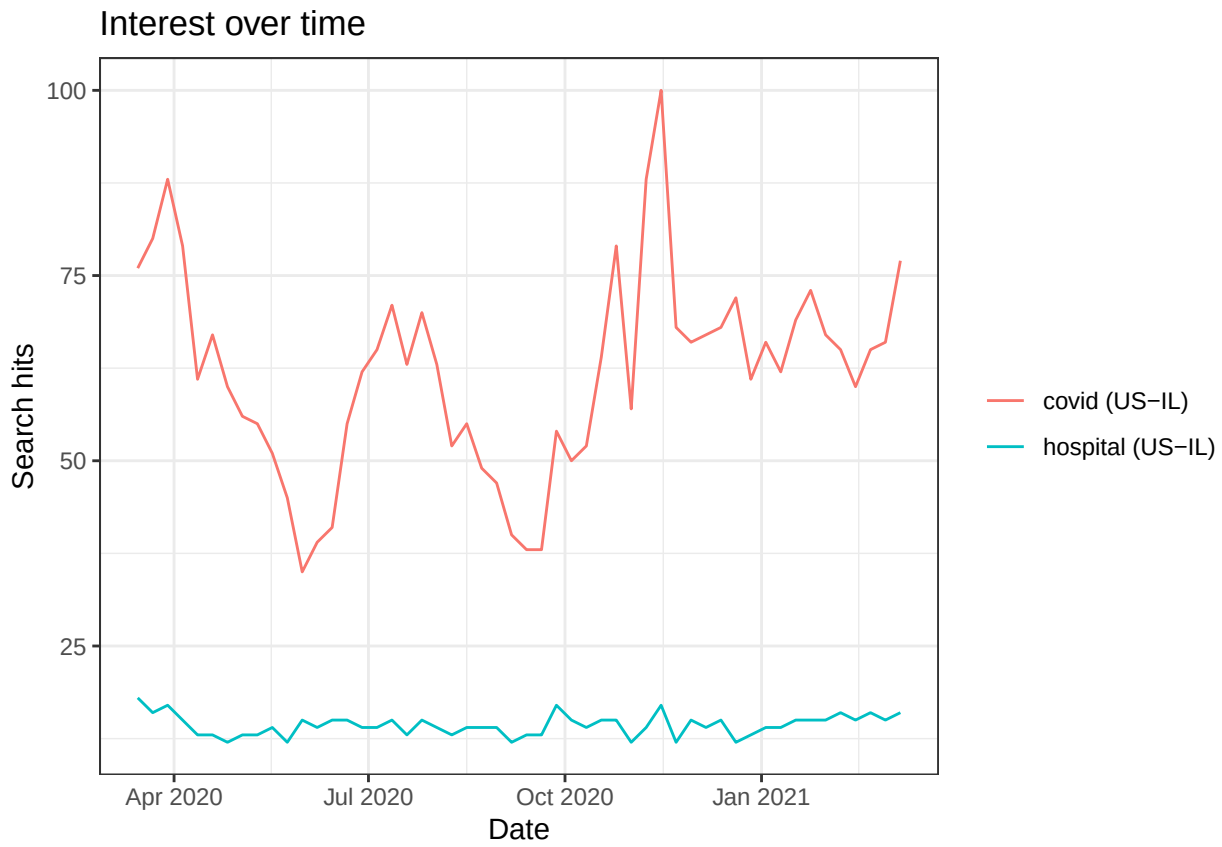
```
plot(res)
```



According to the graph above, search frequencies for "crime" and "loans" have similar trends at the beginning of 2020, where they both went up and down from January to around February 2020. From March to April, search frequency for "loans" increased drastically from approximately 65 to 100, while search frequency for "crime" decreased before it increased again. In other words, the two keywords have a similar trend between January and February and most time between July 2020 and January 2021. However, from March to June 2020, they seem to have a inverse relationship.

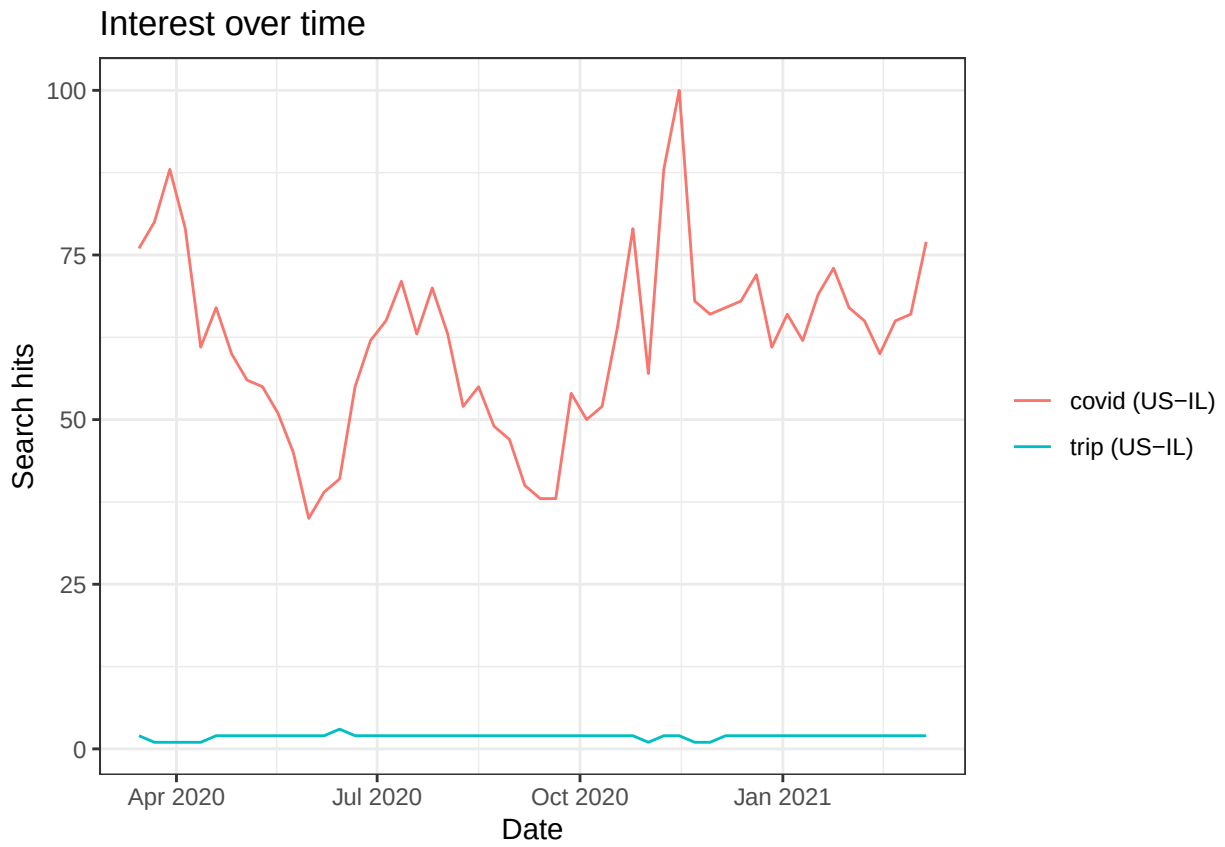
Repeat the above for keywords related to covid. Make sure you use multiple keywords like we did above. Try several different combinations and think carefully about words that might make sense within this context.

```
covid_hos <- gtrends(c("covid", "hospital"),
                     geo = "US-IL",
                     time = "2020-03-11 2021-3-11",
                     low_search_volume = TRUE)
plot(covid_hos)
```



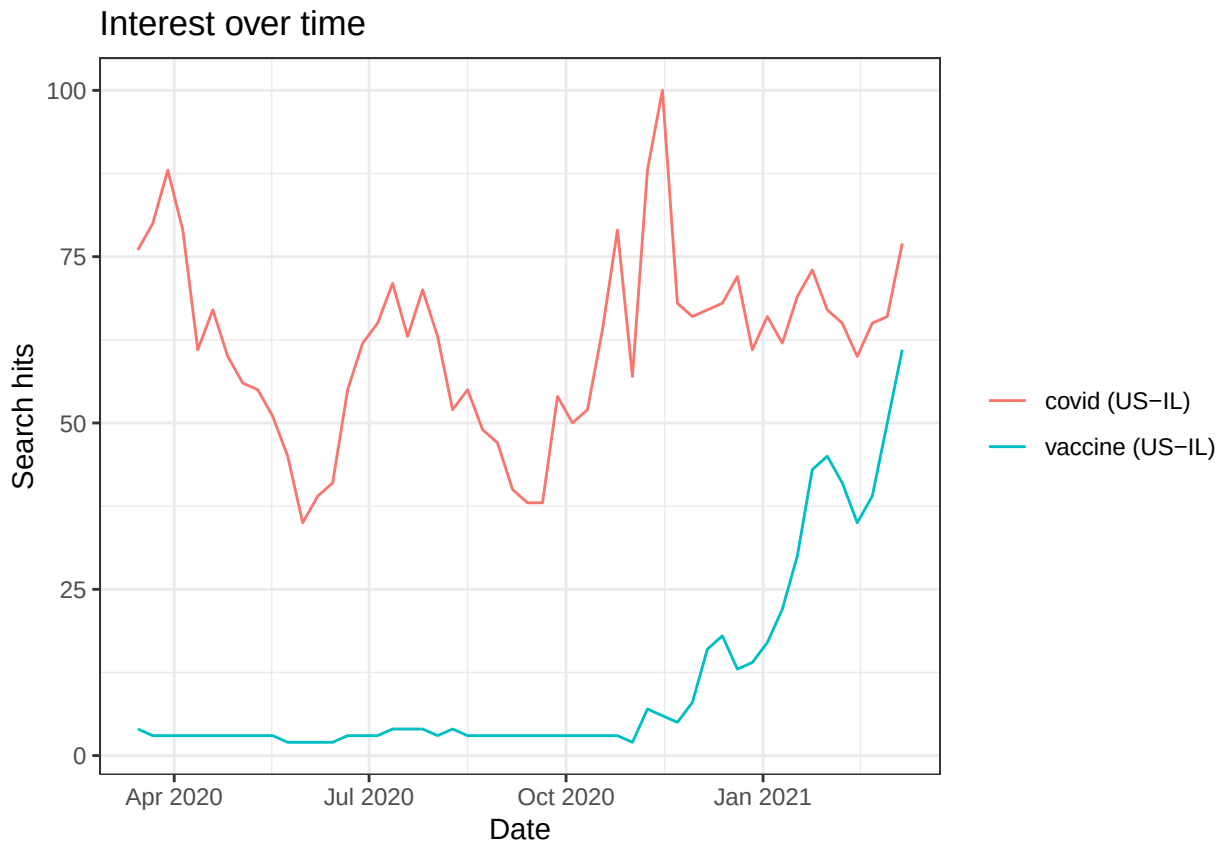
We chose the period from 2020-03-11 (the start of declaring COVID as pandemic by WHO) to 2021-3-11 to investigate the relationship for search keywords covid and hospital. No clear relationship is suggested in the plot, and it seems like people do not search for "hospital" very often during the period.

```
covid_trp <- gtrends(c("covid", "trip"),
                     geo = "US-IL",
                     time = "2020-03-11 2021-3-11",
                     low_search_volume = TRUE)
plot(covid_trp)
```



Again, no clear relationship is found between keywords covid and trip. While search popularity for covid went back and forth from 2020-03-11 to 2021-3-11, very few searches of trip were catched. The result for trip is expected since people rarely traveled during this period to avoid transmission.

```
covid_vac <- gtrends(c("covid", "vaccine"),
                     geo = "US-IL",
                     time = "2020-03-11 2021-3-11",
                     low_search_volume = TRUE)
plot(covid_vac)
```



The search popularities for covid and vaccine seem to have a positive relationship starting from November 2020. As the vaccine became available around this time, the search for vaccine increases drastically and follows a similar pattern of covid.

Google Trends + ACS

Now lets add another data set. The `censusapi` package provides a nice R interface for communicating with this API. However, before running queries we need an access key. This (easy) process can be completed here:

https://api.census.gov/data/key_signup.html

Once you have an access key, store this key in the `cs_key` object. We will use this object in all following API queries.

```
cs_key <- "c007f74dde577f3e0344ae0a2a9721ed20e27142"
```

In the following, we request basic socio-demographic information (population, median age, median household income, income per capita) for cities and villages in the state of Illinois.

```
acs_il <- getCensus(name = "acs/acs5",
                    vintage = 2020,
                    vars = c("NAME",
                             "B01001_001E",
                             "B06002_001E",
                             "B19013_001E",
                             "B19301_001E"),
```

```
                    region = "place:*",
                    regionin = "state:17",
                    key = cs_key)
head(acs_il)
```

```
##   state place                        NAME B01001_001E B06002_001E B19013_001E
## 1    17 15261 Coatsburg village, Illinois         180        35.6       55714
## 2    17 15300    Cobden village, Illinois        1018        44.2       38750
## 3    17 15352      Coffeen city, Illinois         640        33.4       35781
## 4    17 15378   Colchester city, Illinois        1347        42.2       43942
## 5    17 15469    Coleta village, Illinois         230        27.7       56875
## 6    17 15495    Colfax village, Illinois        1088        32.5       58889
##   B19301_001E
## 1       27821
## 2       19979
## 3       26697
## 4       24095
## 5       23749
## 6       24861
```

Convert values that represent missings to NAs.

```
acs_il[acs_il == -666666666] <- NA
```

Now, it might be useful to rename the socio-demographic variables (`B01001_001E` etc.) in our data set and assign more meaningful names.

```
acs_il <-
  acs_il %>%
  dplyr::rename(pop = B01001_001E,
         age = B06002_001E,
         hh_income = B19013_001E,
         income = B19301_001E)
```

It seems like we could try to use this location information listed above to merge this data set with the Google Trends data. However, we first have to clean `NAME` so that it has the same structure as `location` in the search interest by city data. Add a new variable `location` to the ACS data that only includes city names.

```
acs_il$location <- str_remove_all(acs_il$NAME, ", Illinois")
acs_il$location <- str_remove_all(acs_il$location , " city")
acs_il$location <- str_remove_all(acs_il$location , " village")
acs_il$location <- str_remove_all(acs_il$location , " town")
acs_il$location[which(acs_il$location =="St. Anne")] <- "Saint Anne"
acs_il$location[which(acs_il$location =="East St. Louis")] <- "East Saint Louis"
```

Answer the following questions with the "crime" and "loans" Google trends data and the ACS data.

- First, check how many cities don't appear in both data sets, i.e. cannot be matched. Then, create a new data set by joining the Google Trends and the ACS data. Keep only cities that appear in both data sets.

```
merged <- left_join(x=res_city,y=acs_il, by="location")
merged <- merged %>% drop_na(state)
nrow(merged)
```

```
## [1] 331
```

336 cities appear in both Google trends data and the ACS data, and thus 1142 cities don't appear in both datasets.

- Compute the mean of the search popularity for both keywords for cities that have an above average median household income and for those that have an below average median household income. When building your pipe, start with creating the grouping variable and then proceed with the remaining tasks. What conclusions might you draw from this?

```
merged %>%
  mutate(mean = mean(hh_income, na.rm = TRUE))%>%
  mutate(group = ifelse(hh_income > mean, "above average", "below average"))%>%
  group_by(group)%>%
  summarise(crime = mean(crime, na.rm = TRUE),
            loan = mean(loans, na.rm = TRUE))
```

```
## # A tibble: 2 x 3
##   group         crime  loan
##   <chr>         <dbl> <dbl>
## 1 above average  50.9  29.2
## 2 below average  55.1  35.8
```

The mean search popularity of "crime" for cities that have an above average median household income is 25.82979 and for those that have an below average median household income is 27.75009. For the keyword "loans", the mean search popularity are 26.82000 and 32.18681, respectively. For both keywords, those with an below average median household income have a higher mean search popularity. The reason for higher mean search popularity of "crime" can be that those with lower average median household income live in some neighborhoods with a relatively higher number of crimes. Houses in areas with more crimes can be more affordable. The reason for higher mean search popularity of "loans" can be these households need more loans for various living expenses such as education. Also, the low search popularity might be due to less access to internet for lower-income.

- Is there a relationship between the median household income and the search popularity of the Google trends terms? Describe the relationship and use a scatterplot with `qplot()`.

```
cor.test(merged$hh_income, merged$crime, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  merged$hh_income and merged$crime
## t = -1.2801, df = 80, p-value = 0.2042
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.34798920  0.07771525
## sample estimates:
##        cor
## -0.1416813
```

```r
cor.test(merged$hh_income, merged$loans, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  merged$hh_income and merged$loans
## t = -3.6064, df = 127, p-value = 0.0004447
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4537407 -0.1392747
## sample estimates:
##       cor
## -0.3047914
```

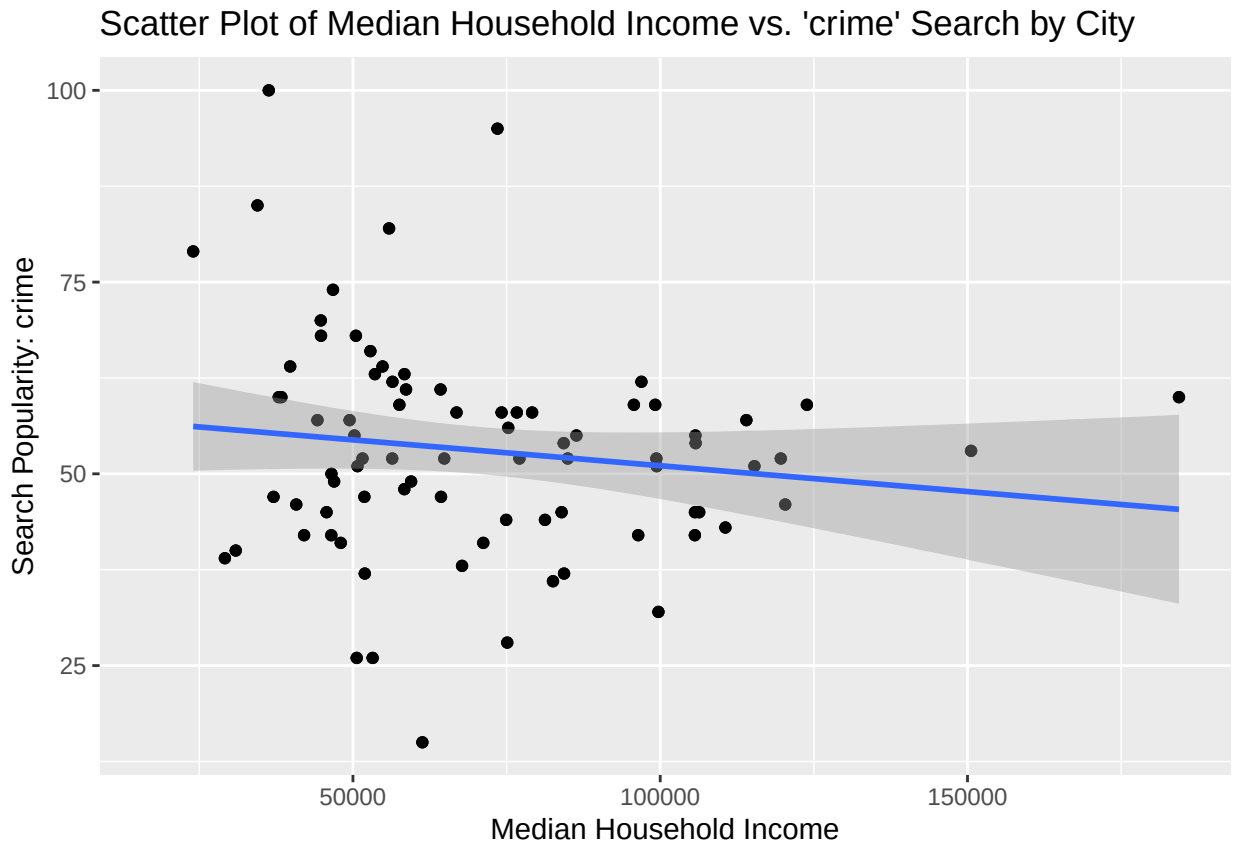
```r
qplot(hh_income, crime, data = merged)+
  geom_point() +
  geom_smooth(method = lm)+
  labs(
    title = "Scatter Plot of Median Household Income vs. 'crime' Search by City",
    x = "Median Household Income",
    y = "Search Popularity: crime"
  )
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

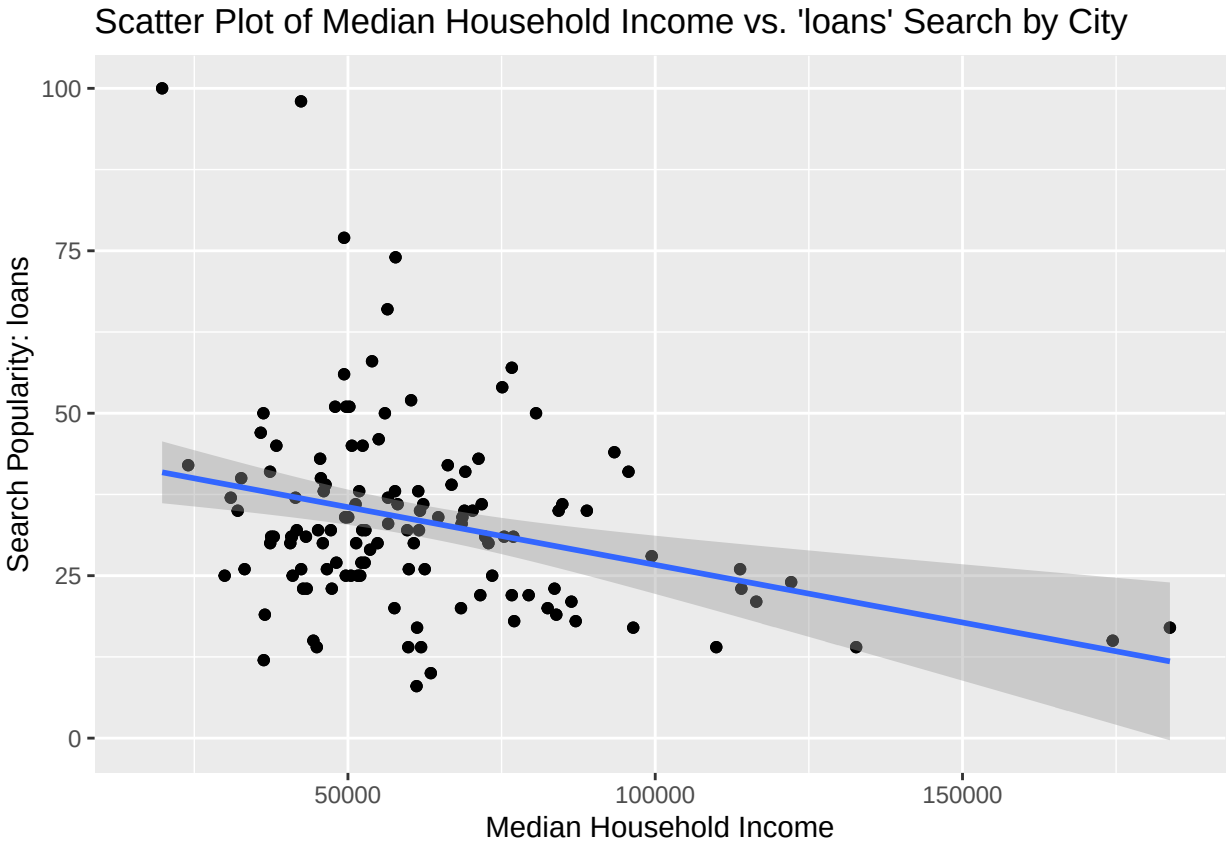
```
## Warning: Removed 249 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 249 rows containing missing values (`geom_point()`).
## Removed 249 rows containing missing values (`geom_point()`).
```




```
#qplot(hh_income, crime, data = merged, geom = c("point", "smooth"))

qplot(hh_income, loans, data = merged)+
  geom_point() +
  geom_smooth(method = lm)+
  labs(
    title = "Scatter Plot of Median Household Income vs. 'loans' Search by City",
    x = "Median Household Income",
    y = "Search Popularity: loans"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 202 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 202 rows containing missing values (`geom_point()`).
## Removed 202 rows containing missing values (`geom_point()`).
```



The results from the Pearson correlation test suggest a negative statistically significant correlation between the median household income and the search popularity for "loans" and a statistically non-significant correlation between the median household income and "crime". These can also be observed in the scatter plots. For "crime", the majority of the cities have search popularity below 40 regardless of median household income. A high outlier (100) is found for North Aurora with median household income $90,315. A slightly decreasing trend according to the the regression line, however, the correlation test suggests an absence of statistically significant relationship between them. For "loans", a decreasing trend is suggested based on the regression line. As the median household income increases, the search popularity for "loans" decrease. For those with median household income higher than $100,000, the searches are mostly lower than 25. Those median household income higher than lower than $100,000 have a wider range of search numbers.

Repeat the above steps using the covid data and the ACS data.

```
cov_vac_byCity <- covid_vac$interest_by_city
cov_vac_city <- spread(cov_vac_byCity, key = keyword, value = hits)
acs_il_edited <- acs_il
acs_il_edited$location[which(acs_il_edited$location =="Lakewood")] <- "Village of Lakewood"

merged_cov <- left_join(x=cov_vac_city,y=acs_il_edited, by="location")
merged_cov <- merged_cov %>% drop_na(state)
nrow(merged_cov)
```

```
## [1] 329
```

333 cities appear in both the covid data and the ACS data and 1133 cities do not.

```
merged_cov %>%
  mutate(mean = mean(hh_income, na.rm = TRUE))%>%
  mutate(group = ifelse(hh_income > mean, "above average", "below average"))%>%
  group_by(group)%>%
  summarise(covid = mean(covid, na.rm = TRUE),
            vaccine = mean(vaccine, na.rm = TRUE))
```

```
## # A tibble: 2 x 3
##   group         covid vaccine
##   <chr>         <dbl>   <dbl>
## 1 above average  69.2    54.3
## 2 below average  58.7    35.9
```

The mean search popularity of "covid" for cities that have an above average median household income is 70.25000 and for those that have an below average median household income is 59.07258. For the keyword "vaccine", the mean search popularity are 65.73333 and 43.24719, respectively. Similar to the results for "crime" and "loans", those with an below average median household income have a higher mean search popularity for both "covid" and "vaccine". Again, the general reason can be that they have less access to internet. The results suggest that households with higher income are more concerned with COVID and vaccination. They are more aware of the pandemic because they might have more resources to access the information about COVID in daily life. In cities that have an below average median household income, there can be less awareness of the pandemic due to the poor resources of public health.

```
cor.test(merged_cov$hh_income, merged_cov$covid, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  merged_cov$hh_income and merged_cov$covid
## t = 8.9653, df = 186, p-value = 3.274e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4408565 0.6419551
## sample estimates:
##       cor
## 0.5493102
```

```
cor.test(merged_cov$hh_income, merged_cov$vaccine, method = "pearson")
```

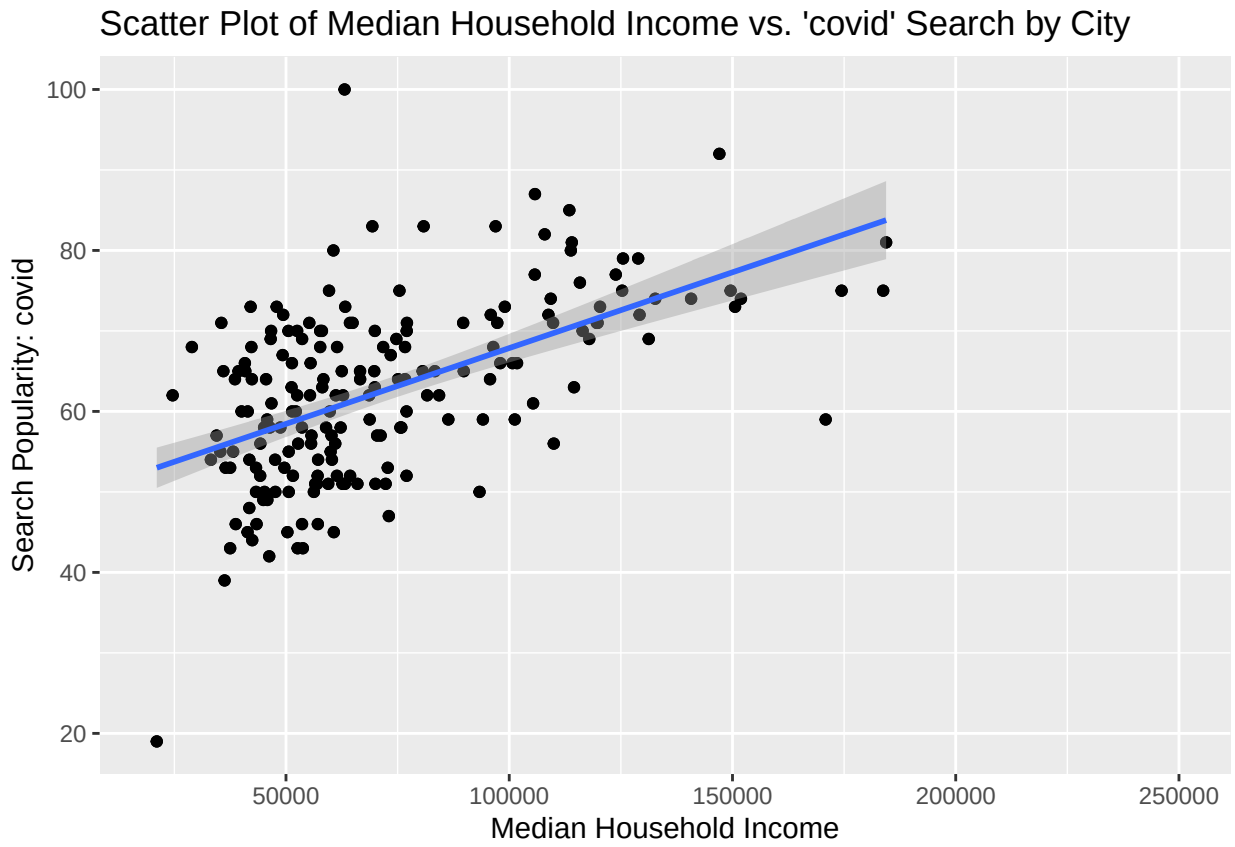
```
##
##  Pearson's product-moment correlation
##
## data:  merged_cov$hh_income and merged_cov$vaccine
## t = 10.641, df = 166, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5371786 0.7188916
## sample estimates:
##       cor
## 0.6367952
```

```
qplot(hh_income, covid, data = merged_cov)+
  geom_point() +
  geom_smooth(method = lm)+
  labs(
    title = "Scatter Plot of Median Household Income vs. 'covid' Search by City",
    x = "Median Household Income",
    y = "Search Popularity: covid"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 141 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 141 rows containing missing values (`geom_point()`).
## Removed 141 rows containing missing values (`geom_point()`).
```



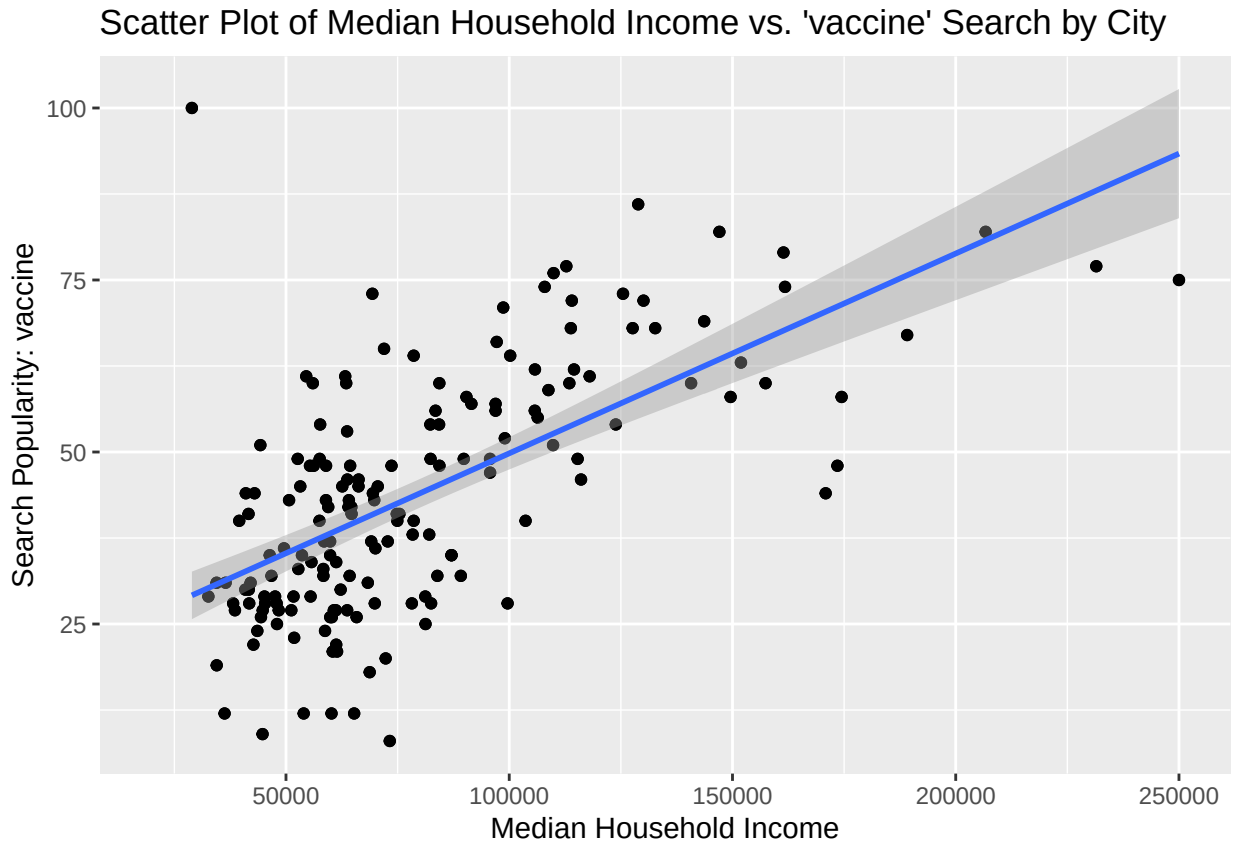

```
#qplot(hh_income, crime, data = merged, geom = c("point", "smooth"))

qplot(hh_income, vaccine, data = merged_cov)+
  geom_point() +
  geom_smooth(method = lm)+
  labs(
    title = "Scatter Plot of Median Household Income vs. 'vaccine' Search by City",
    x = "Median Household Income",
```

```
    y = "Search Popularity: vaccine"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 161 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 161 rows containing missing values (`geom_point()`).
## Removed 161 rows containing missing values (`geom_point()`).
```



The results from the Pearson correlation test suggest a positive statistically significant correlation between the median household income and both keywords "covid" and "vaccine". The scatter plot results are consistent with the correlation tests. For "covid", the majority of the cities with median household income lower than $10,000 have search popularity centered around 40. They generally have a wider range of search popularity than those with median household income higher than $10,000. The latter mostly have over 70 searches for "covid". Based on the plot of median household income and "loans", as the median household income increases, the search popularity for "loans" seems to increase as well. About half of the cities with median household income lower than $125,000 have search popularity below 60, while the majority of those with median household income higher than $125,000 have search popularity above 60.