Assignment 2 Group 16(Xinyu Lin and Yujing Jiang)

Due at 11:59pm on October 1.

Github Link:https://github.com/Colin0817/assignment2_group16.git

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)
```

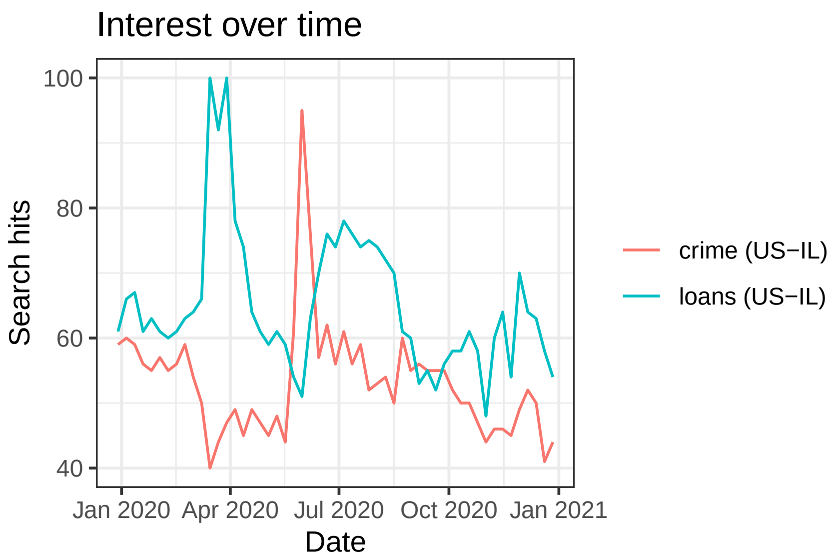
In this assignment, you will pull from APIs to get data from various data sources and use your data wrangling skills to use them all together. You should turn in a report in PDF or HTML format that addresses all of the questions in this assignment, and describes the data that you pulled and analyzed. 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)
```




```
saveRDS(res, "res.rds")
res <- readRDS("res.rds")
```

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

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

```
res_ <- res$interest_over_time
res_time <- as_tibble(res$interest_over_time)
res_time%>%
  group_by(keyword)%>%
  summarise(mean_hits = mean(hits, na.rm = TRUE),
            median_hits = median(hits, na.rm = TRUE),
            var_hits = var(hits, na.rm = TRUE))
```

```
# A tibble: 2 x 4
  keyword mean_hits median_hits var_hits
  <chr>       <dbl>       <int>    <dbl>
1 crime        53.2          53     76.9
2 loans        65.2          63    118.
```

Answer:

The mean of the `crime` is 53.2 the median is 53, and the variance is 76.9. The mean of the `loans` is 65.2, the median is 63, and the variance is 118.

- **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_location <- res$interest_by_city
head(res_location)
```

```
          location hits keyword   geo gprop
1             Anna  100   crime US-IL   web
2           Macomb   72   crime US-IL   web
3  North Riverside   70   crime US-IL   web
4       Streamwood   63   crime US-IL   web
5       Harrisburg   63   crime US-IL   web
6 Germantown Hills   63   crime US-IL   web
```

```
res_location_ <- as_tibble(res_location)
res_location_1  <- pivot_wider(res_location_,
                        names_from = keyword,
                        values_from = hits)
res_location_1 %>% arrange(desc(loans))
```

```
# A tibble: 345 x 5
   location          geo   gprop crime loans
   <chr>             <chr> <chr> <int> <int>
 1 Evergreen Park    US-IL web      NA   100
 2 Long Lake         US-IL web      NA    77
 3 Rosemont          US-IL web      NA    62
 4 Peotone           US-IL web      NA    61
 5 Channel Lake      US-IL web      NA    59
 6 Coal City         US-IL web      NA    57
 7 Dolton            US-IL web      NA    56
 8 East Saint Louis  US-IL web      NA    55
 9 Ford Heights      US-IL web      NA    54
10 Hazel Crest       US-IL web      NA    54
# i 335 more rows
```

Answer:

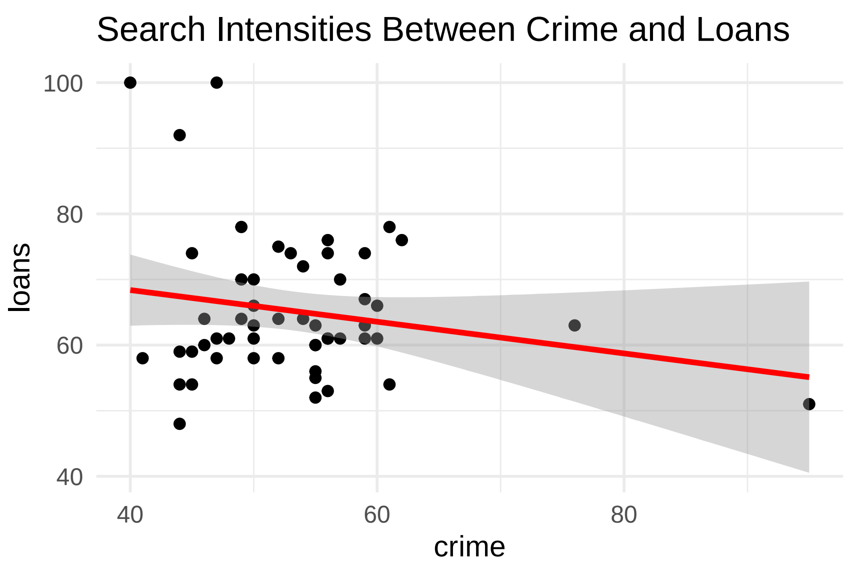
From the results, Evergreen Park has the highest search frequency for loans, which is 100.

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

```
res_time1 <- res_time %>%
  select(date, keyword, hits) %>%
  pivot_wider(names_from = keyword, values_from = hits)
cor_time <- cor(res_time1$crime, res_time1$loans)
cor_time
```

```
[1] -0.1947519
```

```
library(ggplot2)
    ggplot(res_time1, aes(x = crime, y = loans)) +
      geom_point() +
      geom_smooth(method = "lm", color = "red") +
      labs(title = "Search Intensities Between Crime and Loans",
           x = "crime",
           y = "loans") +
      theme_minimal()
```




```
model <- lm(loans ~ crime, data = res_time1)
summary(model)
```

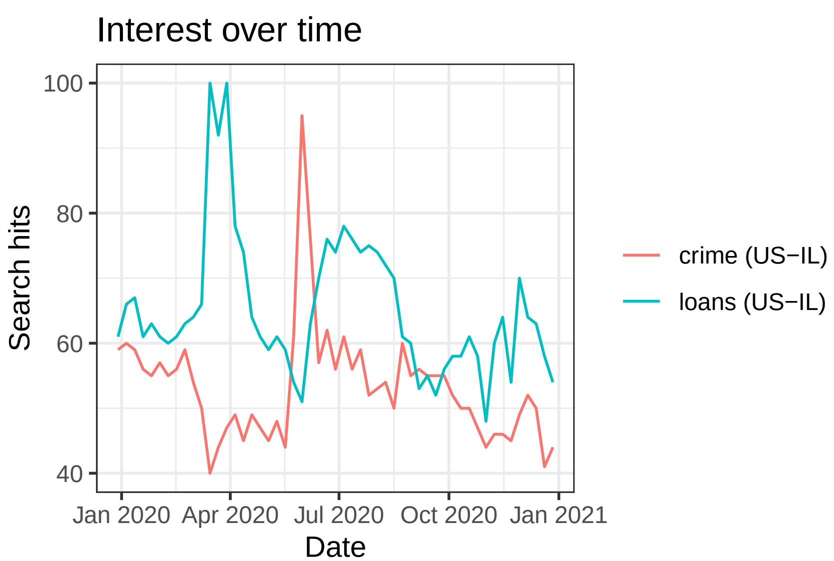
```
Call:
```

```
lm(formula = loans ~ crime, data = res_time1)

Residuals:
    Min      1Q  Median      3Q     Max 
-19.420  -7.489  -2.800   5.718  33.304 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  78.0407     9.1836   8.498 2.44e-11 ***
crime        -0.2414     0.1702  -1.418    0.162    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.77 on 51 degrees of freedom
Multiple R-squared:  0.03793,	Adjusted R-squared:  0.01906 
F-statistic: 2.011 on 1 and 51 DF,  p-value: 0.1623
```

```
cor(res_time1$crime, res_time1$loans, method = "spearman")
```

```
[1] 0.05104254
```

```
plot(res)
```

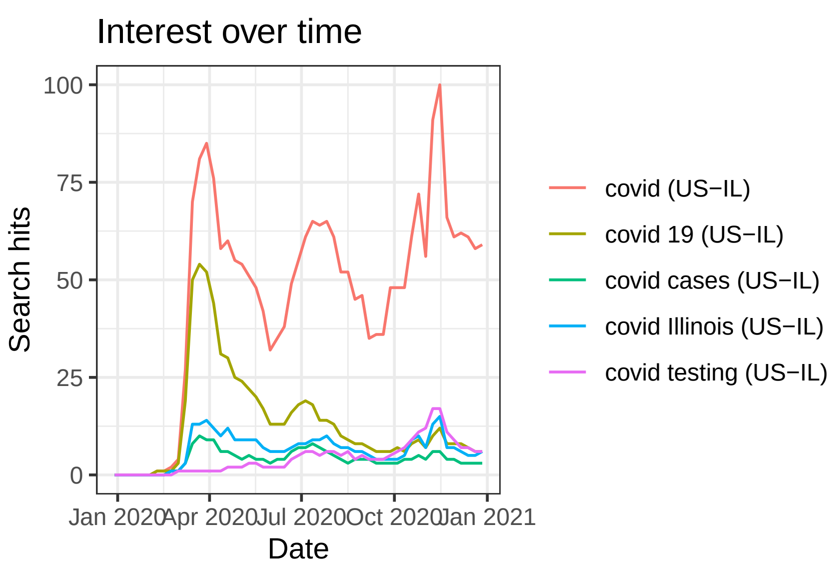


Answer:

The correlation coefficient is -0.1947519, indicating that there is an almost negligible relationship between the search frequency of these two keywords. Furthermore, the occasional negative correlation is very weak and insufficient to indicate a significant inverse relationship. The Spearman's correlation coefficient is 0.05104254, indicating that the correlation is not statistically significant when the normal distribution is not taken into account. From the graph "Search Intensities Between Crime and Loans", it is challenging to discern the evident linear correlation between the two variables. However, from the second one, the graph reveals a more pronounced inverse trend between crime and loans in general between March 2020 and October 2020. This indicates that as the search frequency for crime increases, the search frequency for loans decreases. However, before March and after October, the search rates for crime and loans exhibit a similar trend. So the relationship between search rates for loans and search rates for crime needs to be explored further.

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.

```r
res_covid <- gtrends(c("covid testing", "covid", "covid 19", "covid cases",
                       "covid Illinois"),
              geo = "US-IL",
              time = "2020-01-01 2020-12-31",
              low_search_volume = TRUE)
plot(res_covid)
```




```r
saveRDS(res_covid, "res_covid.rds")
res_covid <- readRDS("res_covid.rds")
```

Answer: Why we choose these keywords?

We use “covid testing”,“covid”,“covid 19”,“covid cases”, “covid Illinois” as keywords. The reasons for choosing these five keywords are as follows:

covid testing: This keyword reflects people’s interest in COVID-19 testing, especially during the pandemic when the availability of testing is directly related to infection rates. It shows the public’s demand and attitude toward testing.

covid: As the general name for the coronavirus pandemic, this keyword covers a wide range of information related to the virus. Using this keyword helps capture the overall attention and trends concerning COVID-19.

covid 19: This is the official name of the virus, commonly used in scientific and medical literature. Choosing this keyword helps gather more professional discussions and information.

covid cases: This keyword focuses on the search for confirmed cases, showing the public’s concern about the spread of the virus. Analyzing data from this keyword can reveal how people perceive the seriousness of the pandemic.

covid Illinois: This keyword is specific to COVID-19 information in Illinois, helping to analyze the pandemic situation and public reaction in that region. It provides a more localized perspective, suitable for studying specific regional policies and measures.

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

```r
res_covid_ <- res_covid$interest_over_time
res_covid_time <- as_tibble(res_covid_)
res_covid_time$hits <- as.numeric(res_covid_time$hits)
res_covid_time %>%
  group_by(keyword)%>%
  summarise(mean_hits1 = mean(hits, na.rm = TRUE),
            median_hits1 = median(hits, na.rm = TRUE),
            var_hits1 = var(hits, na.rm = TRUE))
```

```
# A tibble: 5 x 4
  keyword        mean_hits1 median_hits1 var_hits1
  <chr>               <dbl>        <dbl>     <dbl>
1 covid               47.7            52    622.
2 covid 19            13.8             9    171.
3 covid Illinois       6.65            7     15.8
4 covid cases          4.12            4      6.23
5 covid testing        4.41            4     16.5
```

Answer:

covid: The mean number of hits is 47.7, the median number of hits is 52, and the variance of the number of hits is 622.

covid 19: The average number of hits was 13.8, the median number of hits was 9, and the variance of the number of hits was 171.

covid Illinois: the mean number of hits is 6.65, the median number of hits is 7, and the variance of the number of hits is 15.8.

covid cases: the average number of hits was 4.12, the median number of hits was 4, and the variance of the number of hits was 6.23.

covid testing: mean number of hits is 4.41, median number of hits is 4, variance of hits is 16.5.

- **Which cities (locations) have the highest search frequency for `covid`?**

```r
res_covid_location <- res_covid$interest_by_city
res_covid_location_ <- as_tibble(res_covid_location)
res_covid_location_$hits <- as.numeric(res_covid_location_$hits)
res_covid_location_1  <- pivot_wider(res_covid_location_,
                          names_from = keyword,
                          values_from = hits)
res_covid_location_1 <- res_covid_location_1 %>%
  unnest(cols = everything())
res_covid_location_1 %>% arrange(desc(covid))
```

```
# A tibble: 557 x 8
   location       geo   gprop `covid testing` covid `covid 19` `covid cases`
   <chr>          <chr> <chr>           <dbl> <dbl>      <dbl>         <dbl>
 1 Evergreen Park US-IL web               100   100         NA            NA
 2 Elizabeth      US-IL web                NA    94         NA            NA
 3 Wilmette       US-IL web                NA    94         NA            NA
 4 Highland Park  US-IL web                NA    91         NA            NA
 5 River Forest   US-IL web                55    89         NA            NA
 6 Westfield      US-IL web                NA    89         NA            NA
 7 Andover        US-IL web                NA    87         NA            NA
 8 New Lenox      US-IL web                49    85         NA            NA
 9 Evanston       US-IL web                NA    85         NA            NA
10 Sleepy Hollow  US-IL web                NA    83         NA            NA
# i 547 more rows
# i 1 more variable: `covid Illinois` <dbl>
```

Answer:

From the results, Evergreen Park has the highest search frequency for `covid`, which is 100.

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

```
res_covid_time1 <- res_covid_time %>%
  select(date, keyword, hits) %>%
  pivot_wider(names_from = keyword, values_from = hits)
corr_matrix <- cor(res_covid_time1[, c("covid", "covid 19", "covid testing",
                                       "covid cases", "covid Illinois")],
                   use = "complete.obs")
corr_matrix <- data.frame(corr_matrix)
corr_matrix
```

```
                   covid   covid.19 covid.testing covid.cases covid.Illinois
covid          1.0000000  0.5613752     0.6089544   0.8224879      0.9091744
covid 19       0.5613752  1.0000000    -0.2513807   0.8229000      0.7561185
covid testing  0.6089544 -0.2513807     1.0000000   0.1851953      0.3744597
covid cases    0.8224879  0.8229000     0.1851953   1.0000000      0.8901142
covid Illinois 0.9091744  0.7561185     0.3744597   0.8901142      1.0000000
```

Answer:

A correlation coefficient of 0.822 was observed between the keyword "covid" and both "covid cases". The resulting correlation coefficient of "covid Illinois" and "covid" is 0.909. They are indicating a strong positive correlation. The correlation coefficients with "covid-19" vs "covid" and "covid testing" vs "covid" are comparatively weaker. The keyword "covid 19" exhibits a relatively robust positive correlation with "covid cases" and "covid Illinois," yet displays a comparatively weaker correlation with "covid testing." The keyword "covid testing" demonstrates a relatively weak correlation with all other keywords. The correlation coefficient between the keywords "covid cases" and "covid Illinois" is 0.890, indicating a strong positive correlation.

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, save it as a text file, then read this key in the `cs_key` object. We will use this object in all following API queries. Note that I called my text file `census-key.txt` – yours might be different!

```
cs_key <- read_file("census-key.txt")
```

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. Documentation for the 5-year ACS API can be found here: https://www.census.gov/data/developers/data-sets/acs-5year.html. The information about the variables used here can be found here: https://api.census.gov/data/2022/acs/acs5/variables.html.

```
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)
saveRDS(acs_il, "acs_il.rds")
acs_il <- readRDS("acs_il.rds")
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.

```r
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.

```r
acs_il <-
  acs_il %>%
  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.

```r
acs_clean <- acs_il %>%
  mutate(location = sub(",.*", "", NAME))
acs_clean <- acs_clean %>%
  mutate(location = str_replace(location, " village", "")) %>%
  mutate(location = str_replace(location, " city", "")) %>%
  mutate(location = str_replace(location, " CDP", "")) %>%
  mutate(location = str_replace(location, " town", "")) %>%
  mutate(location = trimws(location))
head(acs_clean)
```

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

```r
# library(stringr)
# acs_clean <- acs_il %>%
#  mutate(location = str_extract(NAME, "^[^,]+"),
#        location = str_replace(location, " city|village|town|CDP", ""),
#         location = str_trim(location, side="right"))
#head(acs_clean)
```

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.**

```
location_google <- res_location_1$location
location_acs <- acs_clean$location
google_in <- setdiff(location_google, location_acs)
acs_in <- setdiff(location_acs, location_google)
ummatched <- length(google_in) + length(acs_in)
ummatched
```

```
[1] 1133
```

```
merge_ <- inner_join(res_location_1, acs_clean, by = "location")
```

Answer:

There are 1133 cities does not appear in both data sets.

- **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?**

```
average_income <- mean(merge_$hh_income, na.rm = TRUE)
merge_fil <- merge_ %>%
  mutate(group = ifelse(hh_income > average_income, "above_ave", "below_ave"))
mean_keyword <- merge_fil%>%
  group_by(group)%>%
  summarise(mean_crime_ = mean(crime, na.rm = TRUE),
            mean_loans_ = mean(loans, na.rm = TRUE))
mean_keyword
```

```
# A tibble: 3 x 3
  group     mean_crime_ mean_loans_
  <chr>           <dbl>       <dbl>
1 above_ave        58          54.6
2 below_ave        64.6        48.5
3 <NA>            NaN         NaN
```

Answer:

The lower popularity of searches related to "crime" in cities with higher incomes may be attributed to the lower incidence of criminal activity in these cities. The greater prevalence of searches for "loans" may indicate that residents of these cities are more inclined to view loans as a means of investment and are able to improve their quality of life through borrowing.

The higher prevalence of searches for "crime" in cities with lower household incomes may be indicative of a greater focus on crime in these cities. This could reflect the fact that areas with poor socioeconomic conditions may have higher crime rates or residents may be more sensitive to crime perceptions. The search prevalence for "loans" is lower than that observed in higher income cities. The higher search prevalence for loans could imply that residents in these cities are more reliant on loans to make ends meet, reflecting greater economic stress.

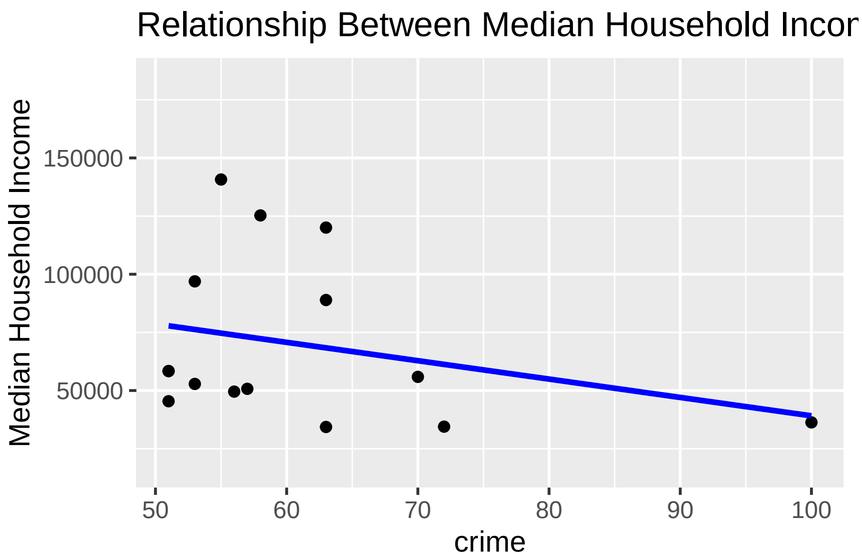
- **Is there a relationship between the median household income and the search popularity of the Google trends terms? Describe the relationship and use a scatterplot with `qplot()`.**

```
cor_crime <- cor(merge_fil$hh_income, merge_fil$crime, use = "complete.obs")
cor_loans <- cor(merge_fil$hh_income, merge_fil$loans, use = "complete.obs")
cor_crime; cor_loans
```

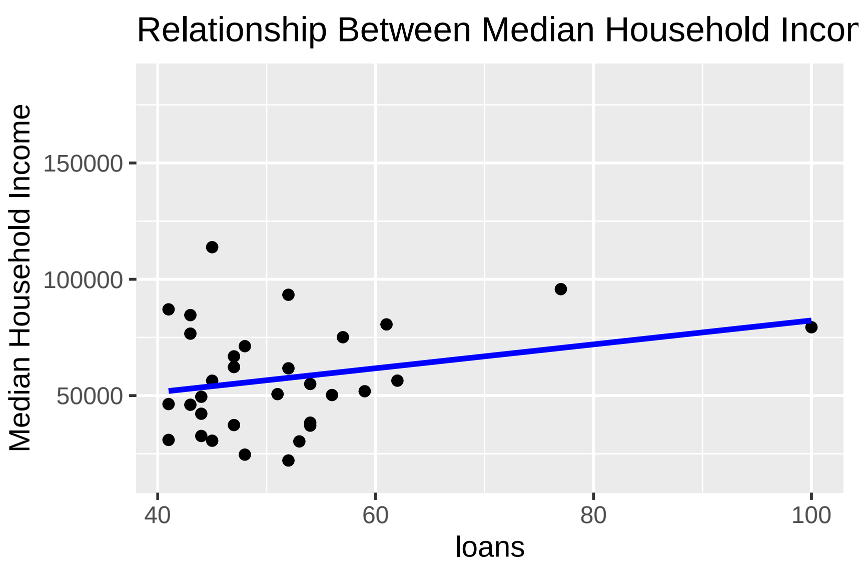
```
[1] -0.283292
```

```
[1] 0.2607699
```

```
qplot(x=crime,y=hh_income,data = merge_fil,geom = "point",
      xlab = "crime",
      ylab = "Median Household Income",
      main = "Relationship Between Median Household Income and Search Popularity for Crime")+
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```




```
qplot(x=loans,y=hh_income,data = merge_fil,geom = "point",
      xlab = "loans",
      ylab = "Median Household Income",
      main = "Relationship Between Median Household Income and Search Popularity for Loans")+
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```



Answer:

The correlation coefficient between family income and crime search degree is -0.283, indicating that there is a low correlation between the two variables and that the relationship is negative. The correlation coefficient between household income and loan search degree is 0.261, indicating that the correlation between household income and crime search degree is low and positive. An examination of the scatterplot of CRIME and HOUSEHOLD INCOME reveals that the data points are concentrated in the lower left of the plot, indicating that household income is higher in areas with lower crime search degrees. In general, there seems to be a negative correlation between lower crime search rates and higher household incomes. The concentration of points in the scatterplot LOANS and HOUSEHOLD INCOME suggests that as household income increases, loan search rates increase, and that income growth levels off or increases at a slower rate after a certain point is reached.

Repeat the above steps using the covid 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.**

```
loc_google <- res_covid_location_1$location
loc_acs <- acs_clean$location
google_in1 <- setdiff(loc_google, loc_acs)
acs_in1 <- setdiff(loc_acs, loc_google)
ummatched <- length(google_in1) + length(acs_in1)
ummatched
```

```
[1] 947
```

```
merge_1 <- inner_join(res_covid_location_1, acs_clean, by = "location",
                      relationship = "many-to-many")
merge_1 <- merge_1 %>%
  distinct(location, .keep_all = TRUE)
```

- **Compute the mean of the search popularity for 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?**

```
ave_income1 <- mean(merge_1$hh_income, na.rm = TRUE)
merge_1 <- merge_1 %>%
  mutate(group1 = ifelse(hh_income > ave_income1, "above_average",
                         "below_average"))
merge_2 <- merge_1 %>%
  group_by(group1) %>%
  summarise(mean_covid = mean(covid, na.rm = TRUE),
```

```r
            mena_testing = mean(`covid testing`, na.rm = TRUE),
            mean_covid19 = mean(`covid 19`, na.rm = TRUE),
            mean_cases = mean(`covid cases`, na.rm = TRUE),
            mean_Illinois = mean(`covid Illinois`, na.rm = TRUE))
merge_2
```

```
# A tibble: 3 x 6
  group1        mean_covid mena_testing mean_covid19 mean_cases mean_Illinois
  <chr>              <dbl>        <dbl>        <dbl>      <dbl>         <dbl>
1 above_average       83.3         50.1           98         69            96
2 below_average       81.2         50.3           93         88            97
3 <NA>               NaN          NaN            NaN        NaN           NaN
```

Answer:

The search rate for the "covid" keyword is slightly higher for high-income households (83.3) than for low-income households (81.2), but the difference is not significant.

The search rate is slightly higher for low-income households (50.3) than for high-income households (50.1), suggesting that low-income households may be searching for information related to COVID testing more frequently.

The search rate for low-income households (93) is significantly lower than that for high-income households (98), indicating that high-income households are investing more search activity in COVID-19-related information.

The search rate for low-income households (88) is significantly higher than that for high-income households (69), which may indicate that low-income households are more likely to pay attention to information about covid cases.

The search rate in covid Illinois is slightly lower for higher-income households (96) than for lower-income households (97), a relatively small difference.

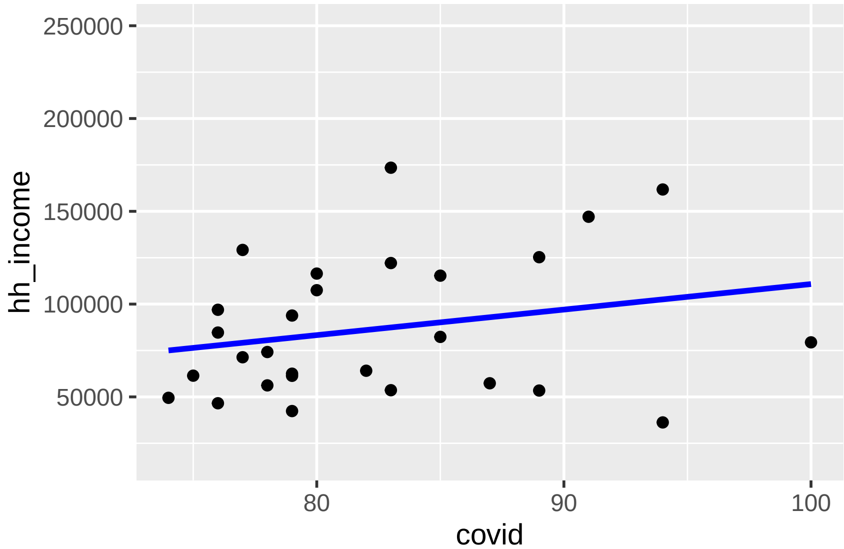
- **Is there a relationship between the median household income and the search popularity of the Google trends terms? Describe the relationship and use a scatterplot with `qplot()`.**

```r
cor_covid <- cor(merge_1$hh_income, merge_1$covid, use = "complete.obs")
cor_covid_testing <- cor(merge_1$hh_income, merge_1$`covid testing`,
                         use = "complete.obs")
cor_covid_19 <- cor(merge_1$hh_income, merge_1$`covid 19`,
                    use = "complete.obs")
cor_covid_cases <- cor(merge_1$hh_income, merge_1$`covid cases`,
                       use = "complete.obs")
```

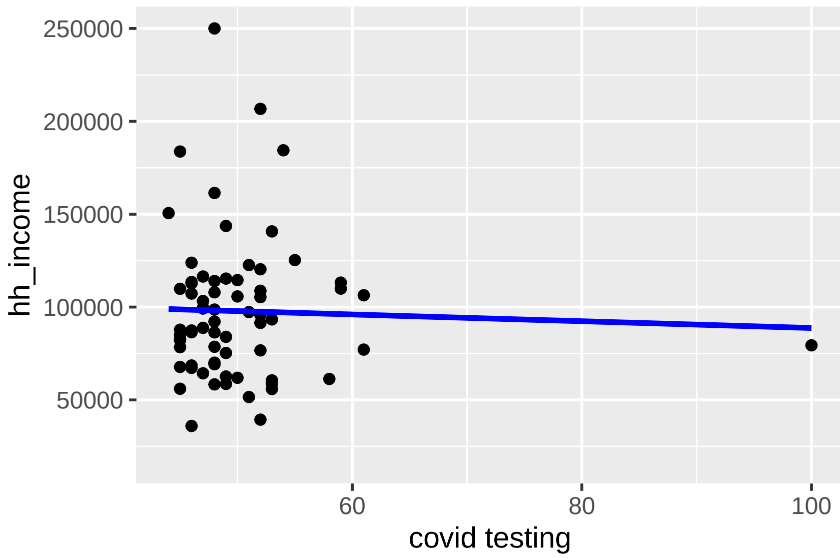
```
cor_covid_Illinois <- cor(merge_1$hh_income, merge_1$`covid Illinois`,
                          use = "complete.obs")
cor_matr <- data.frame(c(cor_covid,cor_covid_testing,cor_covid_19,
                         cor_covid_cases,cor_covid_Illinois))
cor_matr <- rename(cor_matr, cor_hhincome=c.cor_covid..cor_covid_testing..cor_covid_19..cor_c
cor_matr <- cor_matr%>%
  mutate(values = c("covid", "covid testing", "covid 19", "covid cases",
                    "covid Illinois"))%>%
  select(values, everything())
cor_matr
```

```
          values cor_hhincome
1          covid   0.24226565
2  covid testing  -0.03484257
3       covid 19   0.37753183
4    covid cases  -0.69145630
5 covid Illinois   0.04946039
```

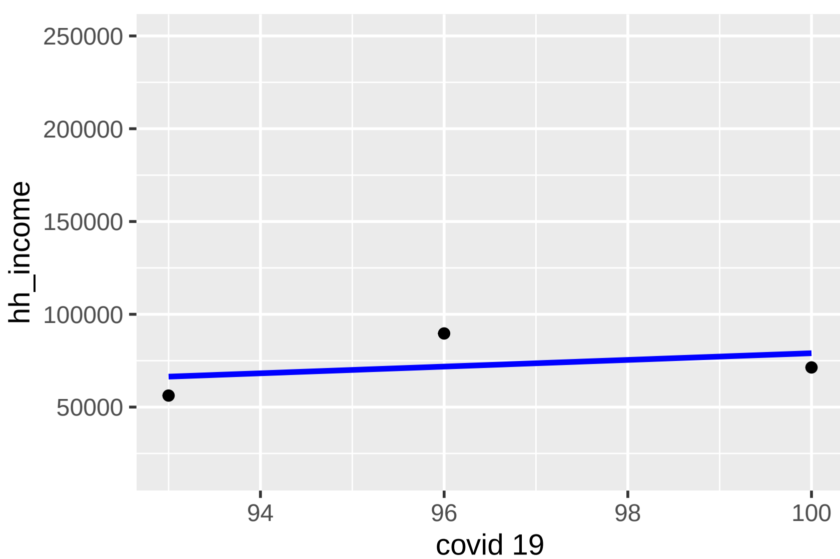
```
qplot(x=covid,y=hh_income,data = merge_1,geom = "point")+
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```



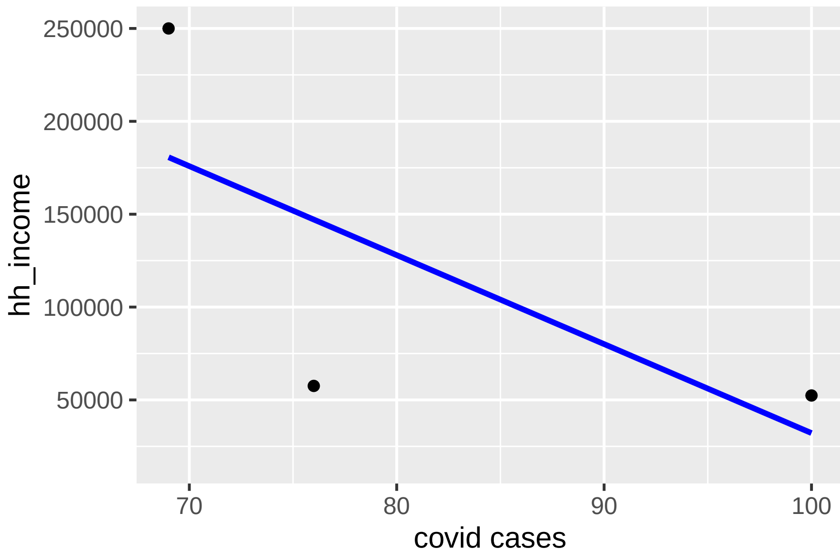

```
qplot(x=`covid testing`,y=hh_income,data = merge_1,geom = "point")+
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```



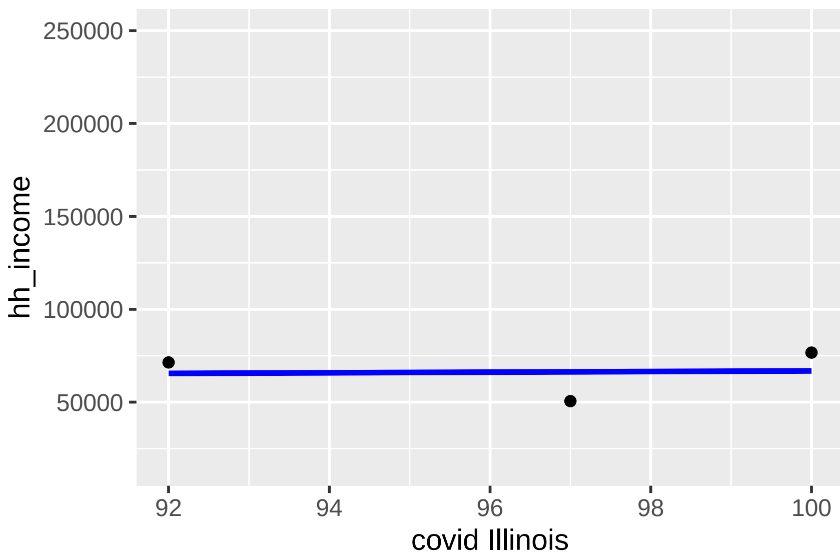

```
qplot(x=`covid 19`,y=hh_income,data = merge_1,geom = "point")+
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```




```
qplot(x=`covid cases`,y=hh_income,data = merge_1,geom = "point")+
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```




```r
qplot(x=`covid Illinois`,y=hh_income,data = merge_1,geom = "point")+
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```



Answer:

The correlation between the keywords “covid”, “covid-19”, and “covid Illinois” vs household income is not statistically significant, yet variables demonstrate a positive relationship. The correlation between “covid testing” and household income is negative and not statistically significant. The keyword “covid cases” and household income have a relatively significant correlation and it is negative. This suggests that individuals may demonstrate greater risk resistance and lower concern for the keywords “covid cases” as income levels rise.

The sample size is relatively limited, necessitating an expansion to obtain more robust scientific results.