Exploring GTrends results

Em Maloney

Google Trends

(github link: https://github.com/ekmaloney/gtrends_occs)

For each occupation search term, I send 37 requests to the Google trends API, each spanning 1 year in length, from January 1 2018-January 1 2019 through January 1 2021-January 1 2022. The benefit of getting such a large range of dates is that then we will have both:

- multiple observations per week, which may help with idiosyncrasies in the google trends returns
- uncovering seasonality in the trends (does the search spike in a certain month even before the pandemic?)

```r
library(tidyverse)
library(purrr)

d <- readRDS("request_dfs.RDS")
accountant_request <- d[[1]]
accountant_request %>% select(-together) %>%
knitr::kable()
```

b	e	keyword	geo
2018-01-01	2019-01-01	accountant	US
2018-02-01	2019-02-01	accountant	US
2018-03-01	2019-03-01	accountant	US
2018-04-01	2019-04-01	accountant	US
2018-05-01	2019-05-01	accountant	US
2018-06-01	2019-06-01	accountant	US
2018-07-01	2019-07-01	accountant	US
2018-08-01	2019-08-01	accountant	US
2018-09-01	2019-09-01	accountant	US

b	e	keyword	geo
2018-10-01	2019-10-01	accountant	US
2018-11-01	2019-11-01	accountant	US
2018-12-01	2019-12-01	accountant	US
2019-01-01	2020-01-01	accountant	US
2019-02-01	2020-02-01	accountant	US
2019-03-01	2020-03-01	accountant	US
2019-04-01	2020-04-01	accountant	US
2019-05-01	2020-05-01	accountant	US
2019-06-01	2020-06-01	accountant	US
2019-07-01	2020-07-01	accountant	US
2019-08-01	2020-08-01	accountant	US
2019-09-01	2020-09-01	accountant	US
2019-10-01	2020-10-01	accountant	US
2019-11-01	2020-11-01	accountant	US
2019-12-01	2020-12-01	accountant	US
2020-01-01	2021-01-01	accountant	US
2020-02-01	2021-02-01	accountant	US
2020-03-01	2021-03-01	accountant	US
2020-04-01	2021-04-01	accountant	US
2020-05-01	2021-05-01	accountant	US
2020-06-01	2021-06-01	accountant	US
2020-07-01	2021-07-01	accountant	US
2020-08-01	2021-08-01	accountant	US
2020-09-01	2021-09-01	accountant	US
2020-10-01	2021-10-01	accountant	US
2020-11-01	2021-11-01	accountant	US
2020-12-01	2021-12-01	accountant	US
2021-01-01	2022-01-01	accountant	US

What does the API return?

For each request (row in the previous data frame), the API returns a dataframe split by week. I have used map to iterate over the df, so it's returned as a list of 37 dataframes. Thus, we can easily "pluck" the values of interest from the list.

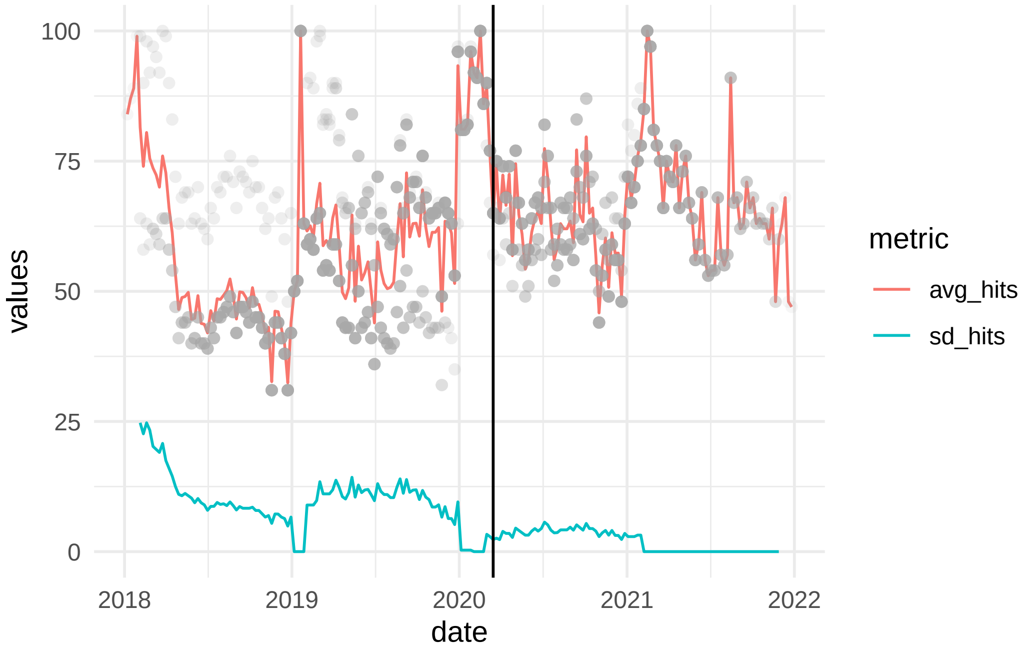
```
accountant <- readRDS("example_return.RDS")

#is a list with 37 dfs in it
length(accountant)
```

```
[1] 37
```

```
#get the interest over time
hits_df <- accountant %>% map_df("interest_over_time")
hits_df <- hits_df %>%
           mutate(date = lubridate::ymd(date)) %>%
           group_by(date) %>%
           mutate(avg_hits = mean(hits),
                  sd_hits = sd(hits)) %>%
           pivot_longer(avg_hits:sd_hits,
                        names_to = "metric",
                        values_to = "values") %>%
           filter(!is.na(values))

ggplot(data = hits_df, mapping = aes(x = date,
                                     y = values,
                                     color = metric)) +
  geom_line() + geom_point(aes(x = date, y = hits), color = "dark grey",
                           alpha = 0.1) + theme_minimal() +
  geom_vline(aes(xintercept = lubridate::ymd("2020-03-15")))
```



We also get the related queries and search terms from these results:

```
#get the related queries
queries <- accountant %>% map_df("related_queries")

queries %>%
  count(value) %>% arrange(desc(n))
```

```
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
```

an important application of regression analysis in accounting is in the estimation of cos

```
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
     n
1   47
2   38
3   37
4   37
5   37
6   37
7   37
8   37
```

```
9    37
10   37
11   37
12   37
13   37
14   37
15   37
16   36
17   35
18   34
19   31
20   29
21   28
22   27
23   26
24   21
25   20
26   19
27   19
28   19
29   19
30   19
31   16
32   15
33   15
34   15
35   14
36   14
37   14
38   13
39   13
40   13
41   13
42   12
43   12
44   12
45   12
46   12
47   11
48   11
49   11
50   11
51   11
```

```
52  11
53  11
54  11
55  11
56  11
57  10
58  10
59  10
60  10
61   9
62   9
63   9
64   9
65   9
66   9
67   9
68   9
69   9
70   9
71   8
72   8
73   8
74   8
75   7
76   7
77   7
78   7
79   7
80   7
81   7
82   7
83   7
84   7
85   6
86   6
87   6
88   6
89   6
90   6
91   6
92   5
93   5
94   5
```

```
95   5
96   5
97   5
98   5
99   5
100  5
101  4
102  4
103  4
104  4
105  4
106  4
107  4
108  4
109  4
110  4
111  4
112  4
113  4
114  3
115  3
116  3
117  3
118  3
119  3
120  3
121  3
122  3
123  3
124  3
125  3
126  3
127  3
128  3
129  3
130  3
131  2
132  2
133  2
134  2
135  2
136  2
137  2
```

```
138  2
139  2
140  2
141  2
142  2
143  2
144  2
145  2
146  2
147  2
148  2
149  2
150  2
151  2
152  2
153  2
154  2
155  2
156  2
157  2
158  2
159  2
160  2
161  2
162  2
163  2
164  1
165  1
166  1
167  1
168  1
169  1
170  1
171  1
172  1
173  1
174  1
175  1
176  1
177  1
178  1
179  1
180  1
```

```
181  1
182  1
183  1
184  1
185  1
186  1
187  1
188  1
189  1
190  1
191  1
192  1
193  1
194  1
195  1
196  1
197  1
198  1
```

```
#get the interest over time
topics <- accountant %>% map_df("related_topics")

topics %>%
  count(value) %>% arrange(desc(n))
```

```
                              value  n
1                        QuickBooks 42
2                        Accountant 38
3                  Bachelor's degree 38
4                           Finance 38
5               Forensic accountant 38
6               Forensic accounting 38
7                  Forensic science 38
8                            Salary 38
9                 senior accountant 38
10                          Account 37
11                       Accounting 37
12                          Average 37
13                         Business 37
14                    Certification 37
15       Certified Public Accountant 37
16                              Job 37
```

17	Job description	37
18	Public	37
19	Tax	37
20	The Accountant	37
21	1-800Accountant	29
22	Chartered accountant	21
23	Small business	14
24	Statutory auditor	14
25	ADP, LLC	12
26	Chartered	12
27	Estate	12
28	KPMG	12
29	LinkedIn	12
30	Adviser	11
31	Soul	11
32	QuickBooks Desktop	10
33	Financial adviser	9
34	Corporate tax	8
35	The Accountant Of Auschwitz	8
36	Balance	7
37	Bookkeeper	7
38	Entry-level job	7
39	Federal Bureau of Investigation	7
40	Project accounting	7
41	Bookkeeping	6
42	Revenue	6
43	Accounts payable	5
44	Career	5
45	General ledger	5
46	Intuit	5
47	Anna Kendrick	4
48	Associate degree	4
49	Ben Affleck	4
50	Duty	4
51	Financial transaction	4
52	Nurse	4
53	Bachelor of Science	3
54	Expense	3
55	Goaltender	3
56	Skill	3
57	Auditor	2
58	Corporation	2
59	Engineer	2

```
60                               Equity  2
61                      Fund accounting  2
62                            H&R Block  2
63                              Payroll  2
64                     Registered nurse  2
65                                 Wage  2
66                                Asset  1
67                                Audit  1
68              Chief Financial Officer  1
69                          Comptroller  1
70                             Director  1
71                           Experience  1
72                           Income tax  1
73                               Ledger  1
74               Occupational Therapist  1
75                        Part-time job  1
76 Tax preparation in the United States  1
77                           Tax report  1
```

Is essential in either of these lists? No.

```
queries %>% filter(str_detect(value, "essential"))
```

```
[1] subject         related_queries value           geo
[5] keyword         category
<0 rows> (or 0-length row.names)
```

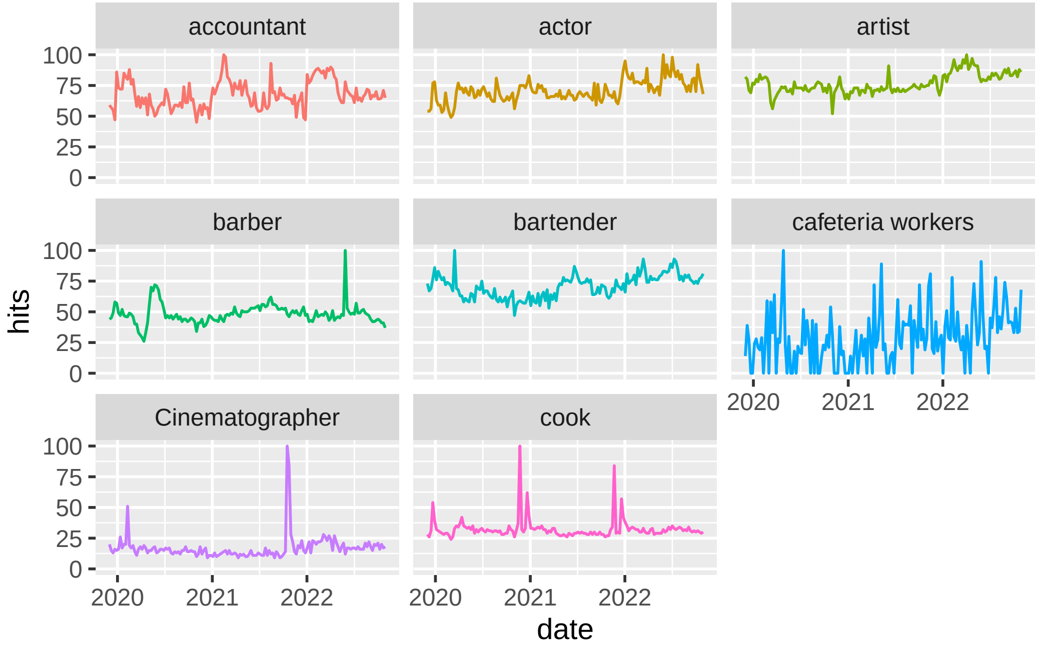
```
topics %>% filter(str_detect(value, "essential"))
```

```
[1] subject        related_topics value          geo            keyword
[6] category
<0 rows> (or 0-length row.names)
```

```
d <- readRDS("eight_results.RDS")


#merge into one df
all_terms <- d %>% map_df("interest_over_time")
```

```
ggplot(all_terms, mapping = aes(x = date, y = hits, color = keyword)) +
       geom_line() + facet_wrap(~keyword) +
  theme(legend.position = "none")
```



```
d %>%
  map_df("related_queries") %>%
  filter(str_detect(value, "essential"))
```

```
[1] subject         related_queries value           geo
[5] keyword         category
<0 rows> (or 0-length row.names)
```

```
d %>%
  map_df("related_topics") %>%
  filter(str_detect(value, "essential"))
```

```
[1] subject        related_topics value           geo             keyword
[6] category
<0 rows> (or 0-length row.names)
```