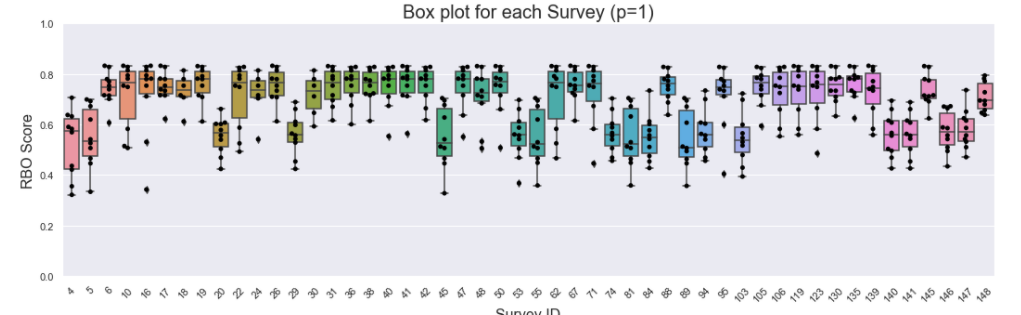
Explain that RBO we used is to compare the full URL

Box plot to represenst RBO score of each queries compare to the rest, grouped by survey ID

Each data point show the average RBO score when compare between result (of the same query) of the 1 survey compare to the rest.



Note: doing color for the column to differentiate countries.

Survey ID has low RBO score which mean they similarity to the rest is slow.

The people with survey 5 is from Brazil

For survey 4 and 5 then is no info from government in the top 3

Starting from people 6, there is 1 website government and that participant is from United States

…

Manually checking it seems to be that the reason there is low RBO because majority of pariticipant comes from the same country which is US, and other country there for showing different results from the majority.

However, interestingly, people with survey id 140 also comes from US (California) but having low RBO score compare to the rest

…

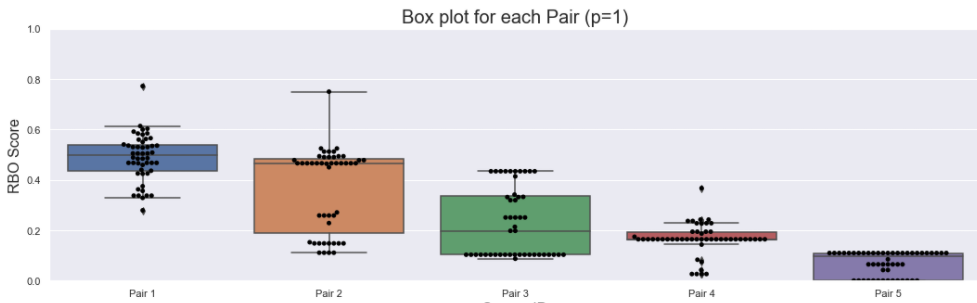
The graph generally showing that people comes from the same country would have search results similar to each other.

Using color scheema

jaccard similarity

Box plot for each participant in terms of governments , same for open rank

Below Box Plot is RBO score of each possitive and negative pair



**Shuffle and limit of the time**

Again the pair is at following

**Pair 1:**

Query 1: should i get tested for covid  
Query 2: should i not get tested for covid

**Pair 2:**

Query 3: should i get flu shot  
Query 4: should i not get flu shot

**Pair 3:**

Query 5: should i get vaccinated  
Query 6: should i avoid get vaccinated

**Pair 4:**

Query 7: should i wear facemask  
Query 8: should i not wear facemask

**Pair 5: the negate is in the word itself**

Query 9: is hydroxychloroquine effective for covid  
Query 10: is hydroxychloroquine ineffective for covid

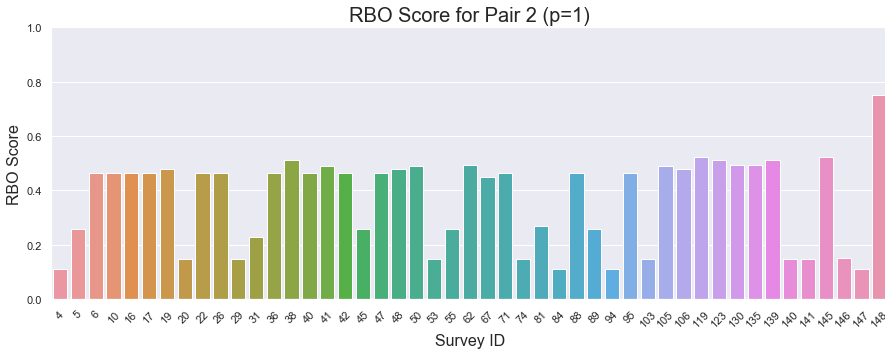
We might not have enough data, but can be consider for future research question

As we can see, the similarity is decreasing from pair 1 to pair 5 (these pair was not arranged in any particular order).

With Pair 2, especially, there is a big gap in between, either the pair is having around 0.5 RBO or below.

Doing Jaccquad

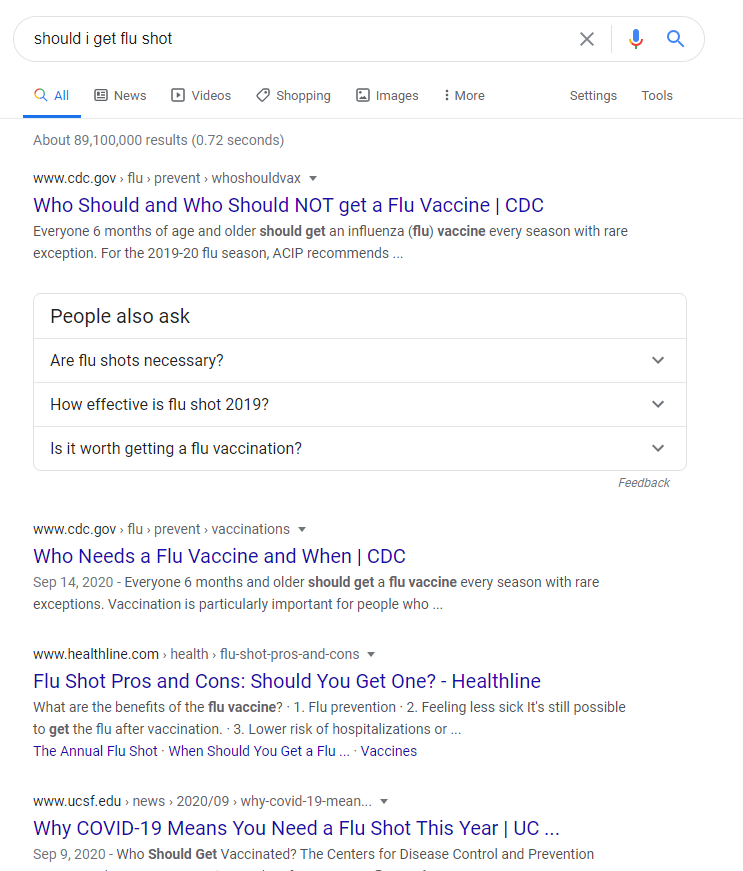
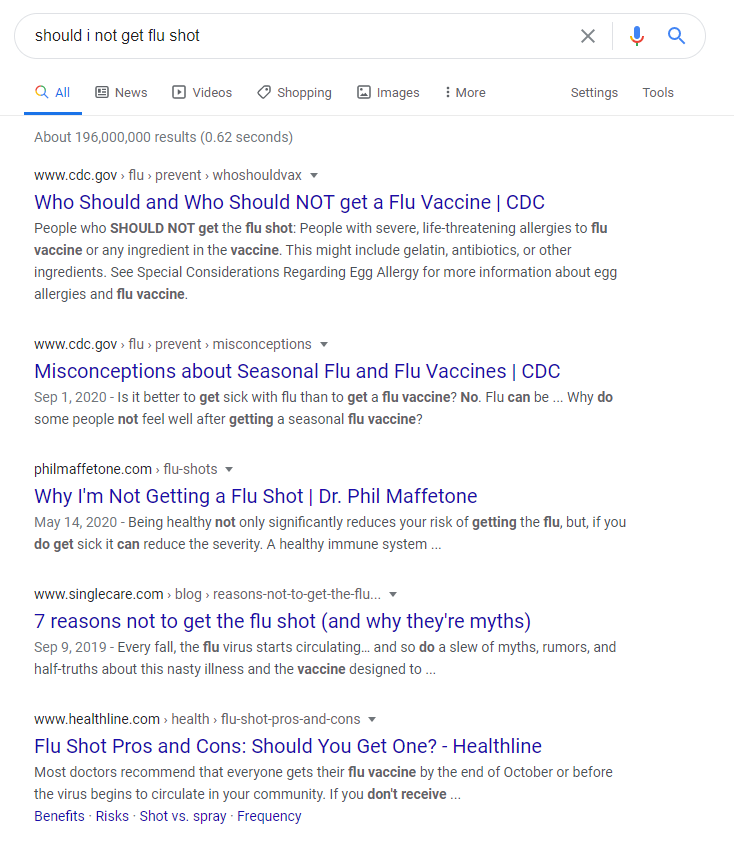
Now, let’s take a closer look into Pair 2:



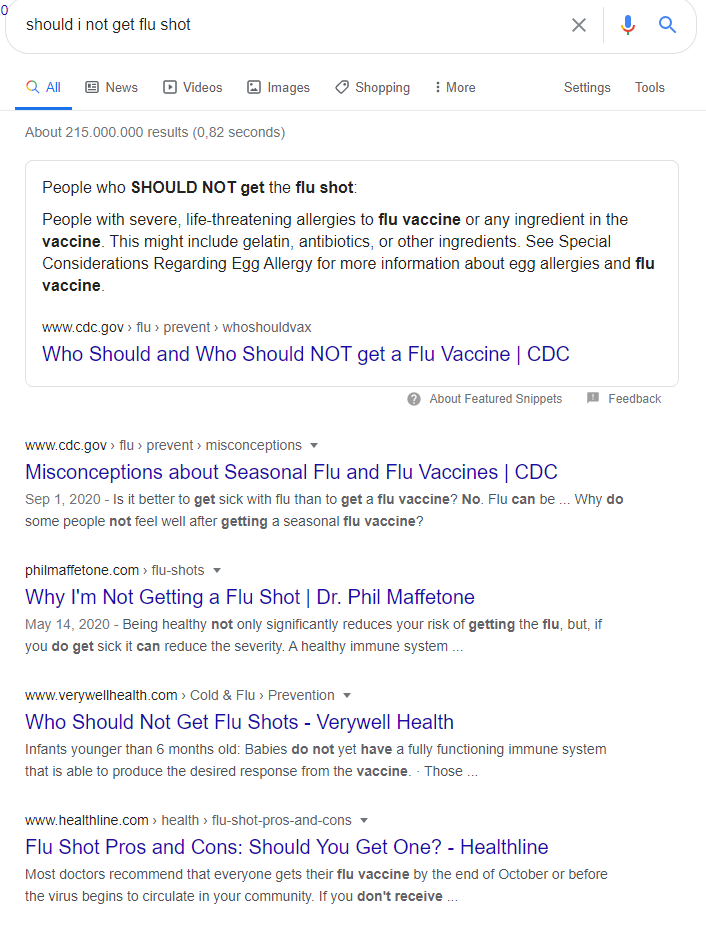
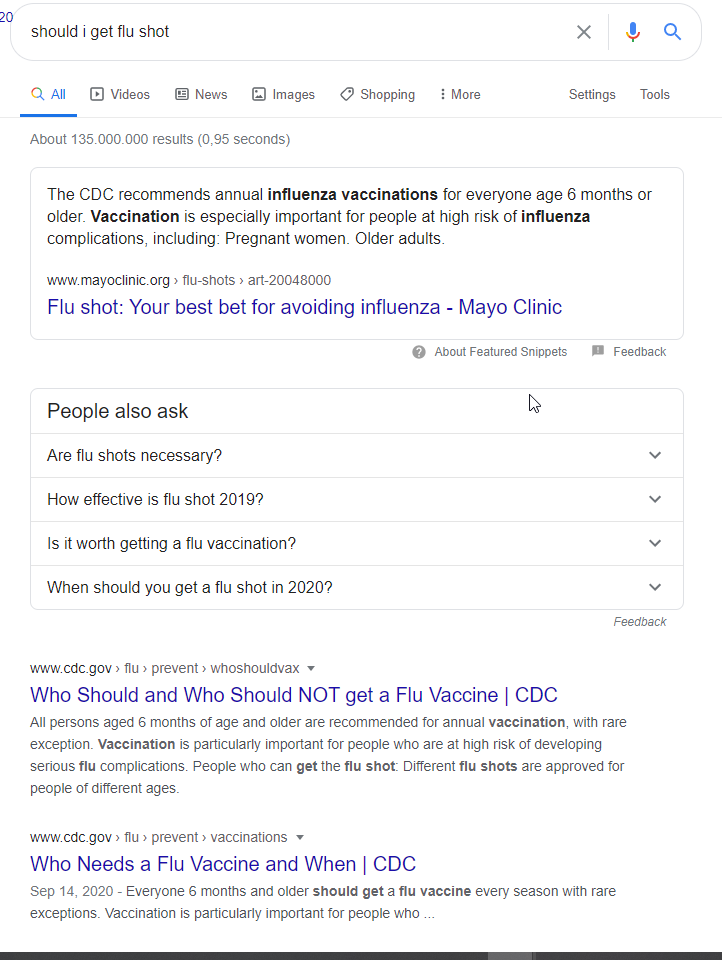
Doing color scheme.

These score is just to compare between possitive and negative query of the same person, (not comparying between different survey ID). But interestingly, people in US still get the higher RBO score compare to the rest of the world

Here is example of survey 6 comes from US

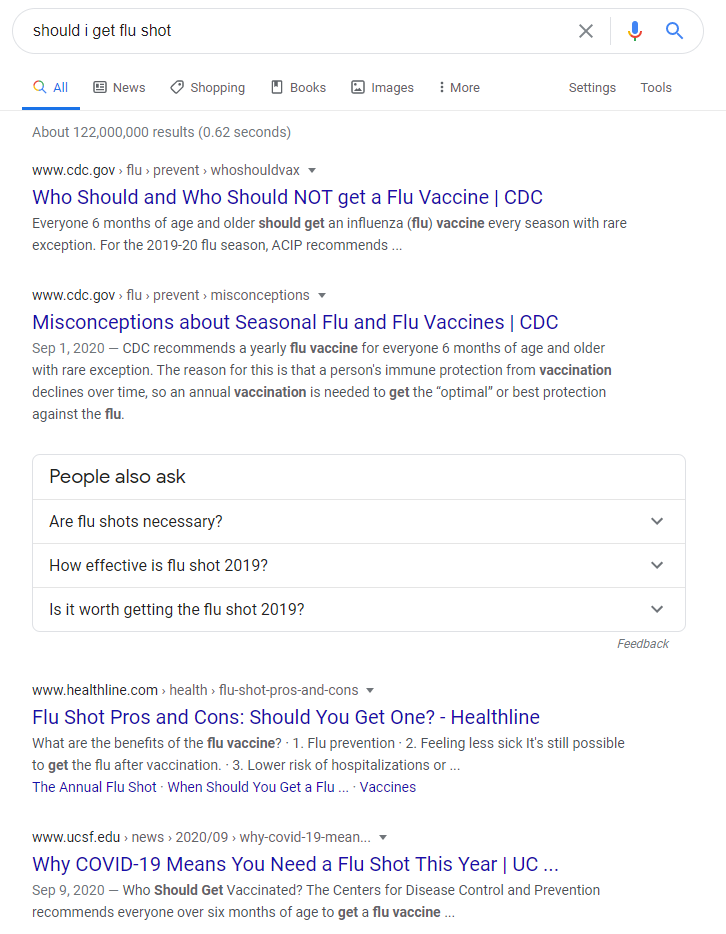
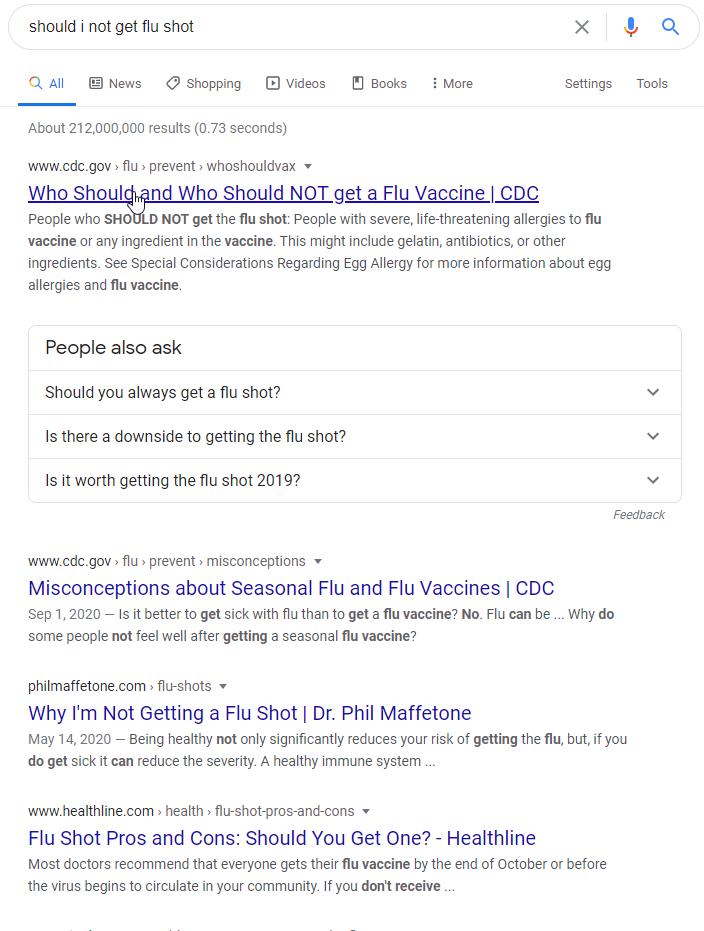
Here is example of survey 4 comes from germany  
Positive: <http://fairness.datasciencevn.com/uploads/html/4-3-1601904205.html>  
Negative: <http://fairness.datasciencevn.com/uploads/html/4-4-1601904071.html>

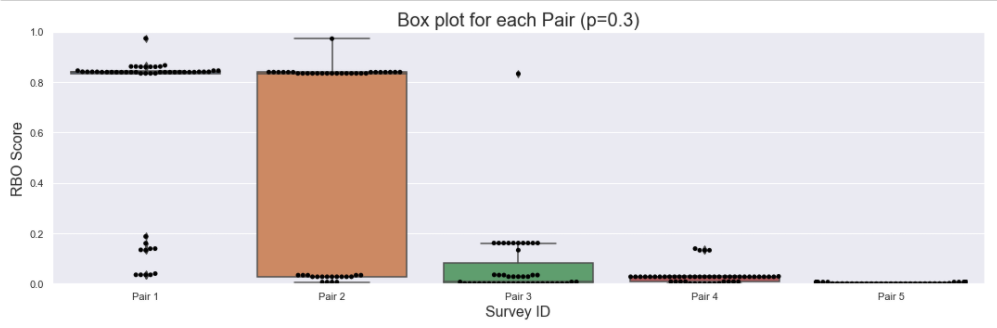
Even that person comes from Germany, the participant seems got results from US websites.

Here is results of survey ID 148, who has an outstanding RBO score comparing to the rest.

Possitive: <http://fairness.datasciencevn.com/uploads/html/148-3-1602072414.html>  
Negative: <http://fairness.datasciencevn.com/uploads/html/148-4-1602072766.html>

If we change to p=0.3. it’s meaning putting more weight to the top results, put not exactly the top 3. (when I read more, that weighting is becoming more complex, so I don’t think we need to clearly define it but maybe rather using it as an indicator and to be signal before looking at any individual results)



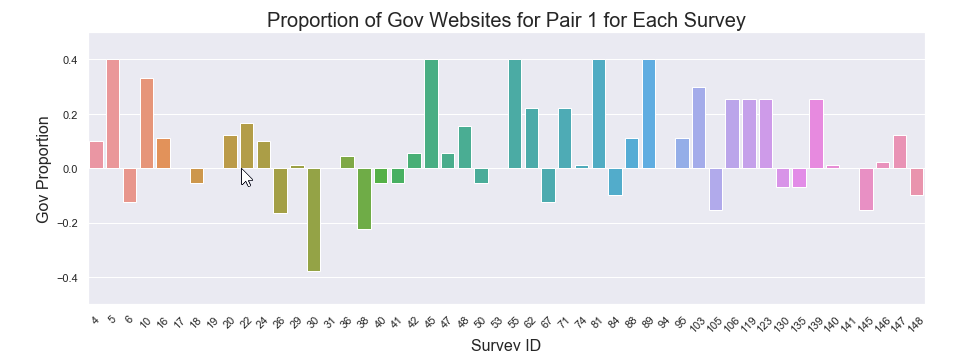
As shown above, with Pair 1, majority of participant got very high similarity.

While Pair also have not much similiarity between possitive and negative.

Show couple of example.

For looking at government.

The graph is showing the gap. And then I have the talbe to show all the number



The result is calculate by

Percentage of governmnet in positive querry’s result - Percentage of governmnet in negative querry’s result.

Therefore, it the final results is possitve nunber, we know the gov website got more than negative.

The survey 5 which having significantly high proportion of Gov Website.

Other survey would have high proportion of government website also comes from Brazil, except 81 is come from Argentina