

# Partisan Conflict and Google Search Analysis

Ben Warren

2023-06-24

## Data Sources

**Partisan Conflict Index:** The Federal Reserve Bank of Philadelphia publishes the Partisan Conflict Index monthly on its public website. The index “tracks the degree of political disagreement among U.S. politicians at the federal level” as measured by the frequency of articles from major newspapers like the *Washington Post*, *New York Times*, and the *LA Times* that report political disagreement in a given month. Data is available dating back to 1981, but in this analysis, I utilize data since 2004.

**Google Trends:** Google publishes a data tool that allows users to view, compare, and download data on the “search interest” of various terms since 2004. For this analysis, I downloaded monthly data sets for the general political terms “Senate”, “Congress”, “Government”, and “Protest” and the partisan terms “Democrat” and “Republican,” confining all search data to the United States.

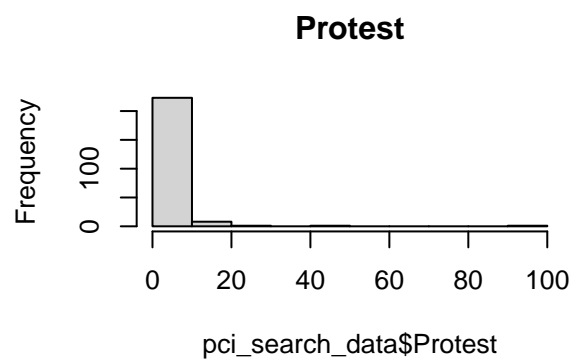
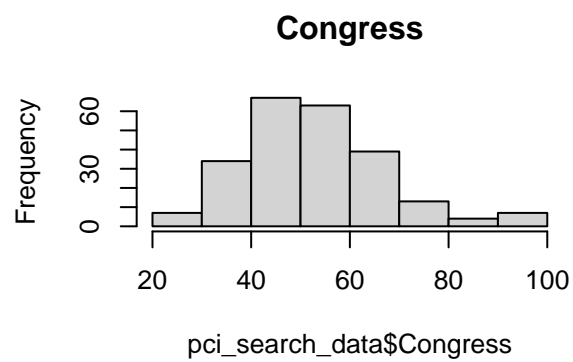
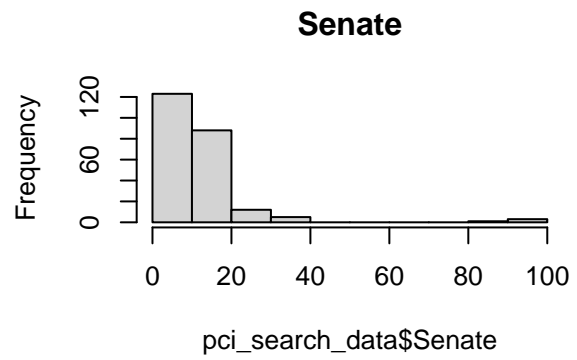
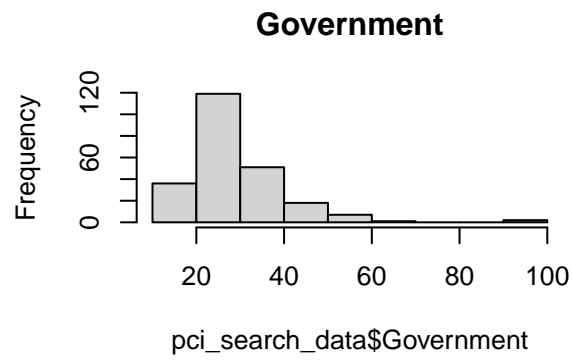
## Summary

The Partisan Conflict Index is a useful tool that measures the level of partisan disagreement within the federal government, but it is worth asking whether trends in the general populace are associated with trends in the PCI. For example, using Google search data, might we find a link between the public searching certain terms and higher levels of partisan conflict?

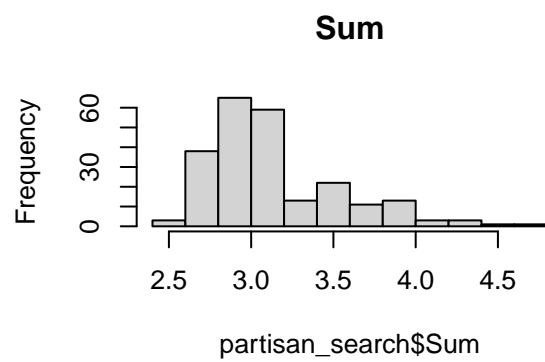
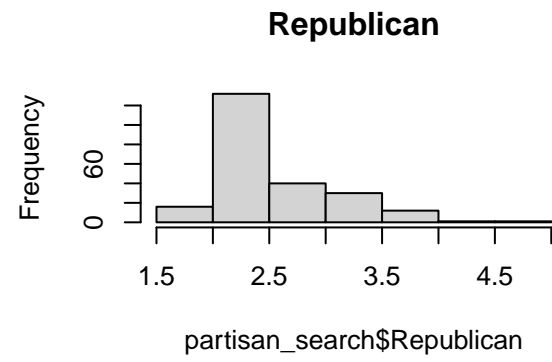
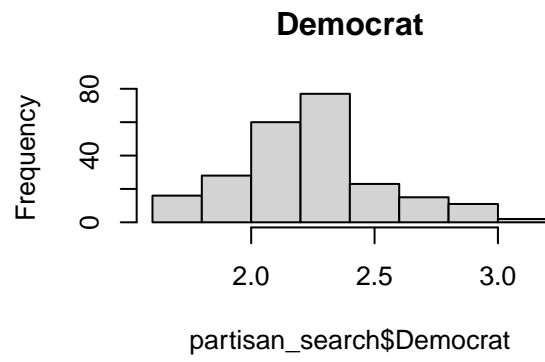
I approach this question using two different methods. First, I use simple regression analysis to identify potential relationships between the search trends and the PCI, finding a statistically significant (negative) correlation between searches for “Democrat” and the conflict index. Then, I use time-series analysis to better understand the temporal relationship between searches for political/partisan terms and PCI.

## Data Manipulation

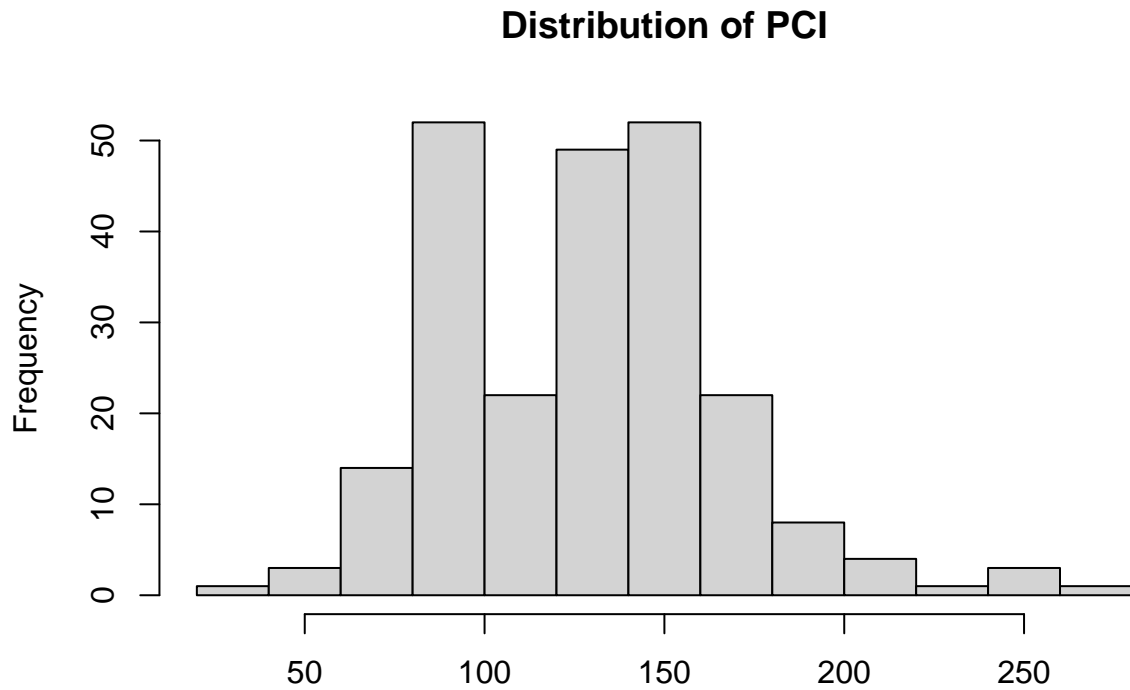
The distributions of the data are variable, with “senate”, “government”, and “protest” right-skewed and “Congress” close to normal. I perform transformations on the data to normalize them, then calculate z-scores from the transformed data.



The data for partisan search terms (Democrat, Republican) are also skewed, so I applied a log transformation to those values as well.



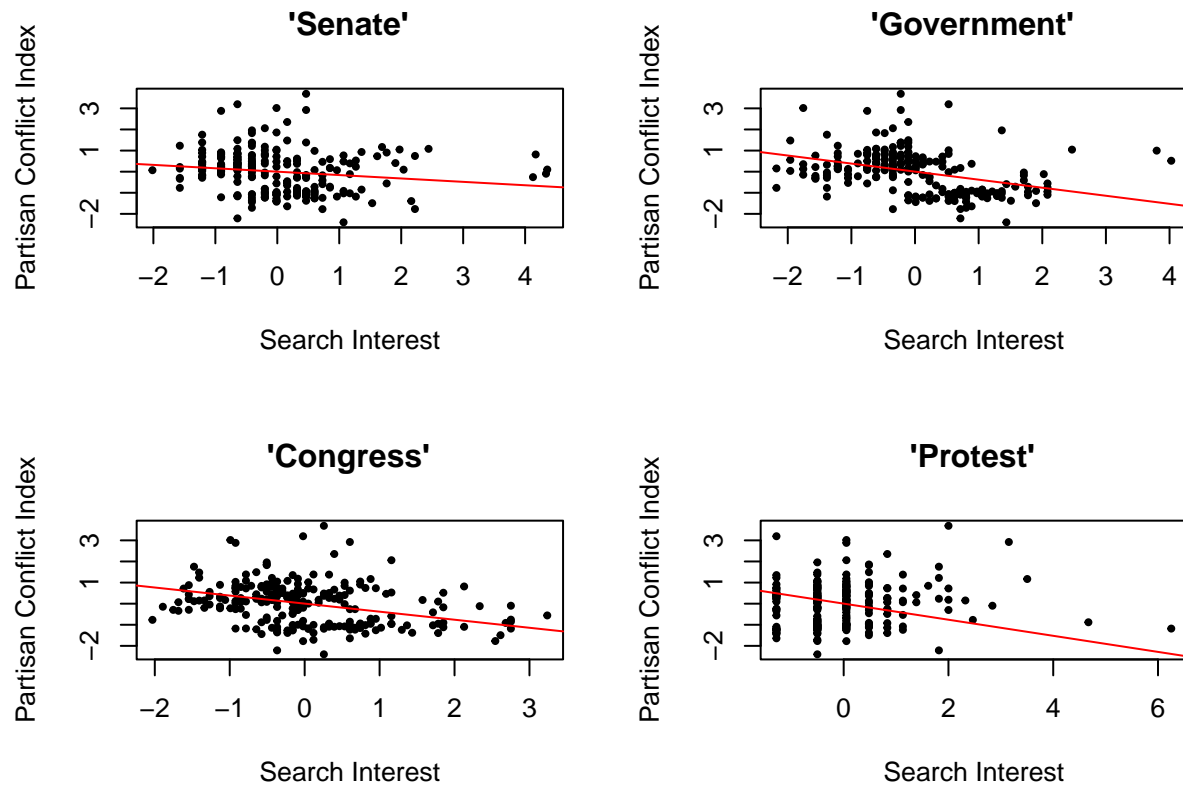
The PCI variable is already distributed close to normally.



I scale all of the variables to z-score values to better compare them.

#### Regression Analysis

Now, for the regression analysis. Below, I regress PCI against each of the search term variables to find any correlation between the two. I plot the variables against each other and plot the linear model as well.

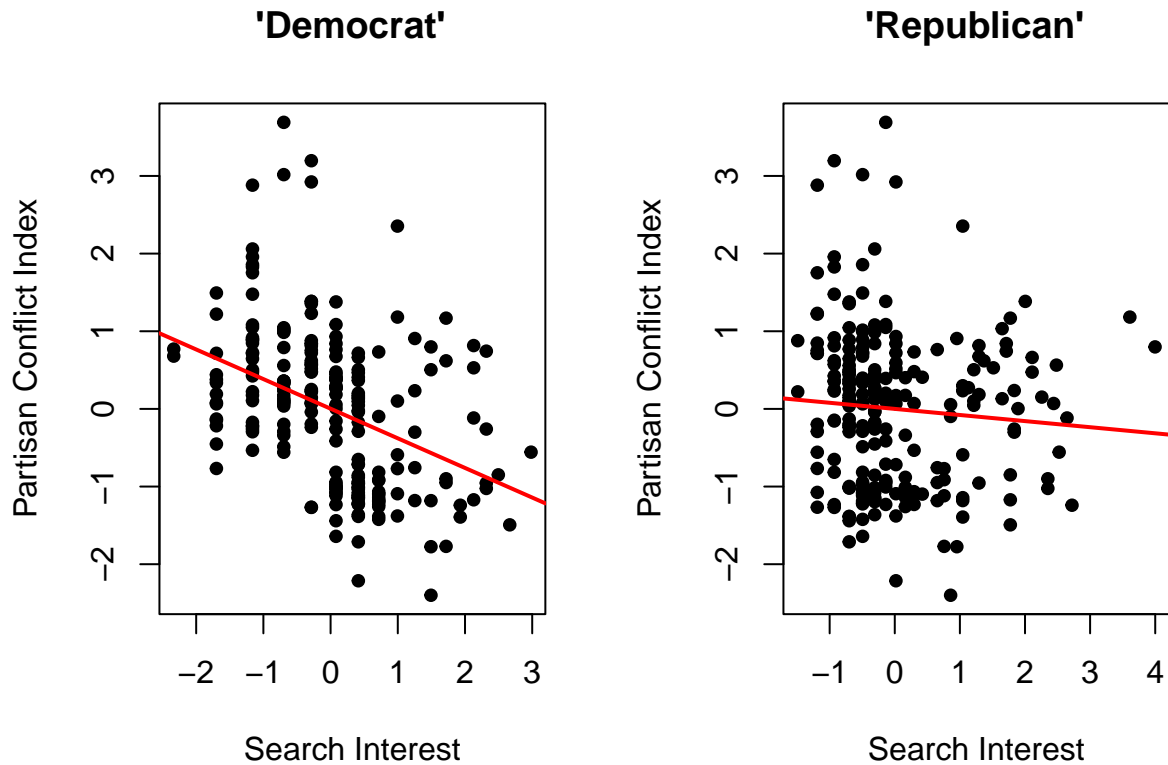


The only set of variables that passes the model utility test for a linear relationship is ‘Government,’ with an R-squared value of 0.08652. Interestingly, the direction of this relationship is negative, implying that when people search ‘Government’ *more* there is less partisan conflict. The model estimates that for each increase of 1 unit of search interest in the term, the conflict index should decrease by 0.3 units.

As a statistical note, search interest is measured by a scale of the topic’s proportion to all searches on all topics. So, a 1 unit increase in this metric equates to the topic increasing 1 percentile in the ranking of all topics.

In addition to general government-related terms, I wondered if there might be a correlation between searches for each of the major American political parties and the conflict index. One could imagine, for example, heightened partisan conflict increasing the level of awareness of each of the parties among the general public (particularly because of the way conflict is measured - through media coverage).

So, I performed regression analysis using PCI and trends for the search terms “Democrat” and “Republican” and for a scaled sum of search interest for both.

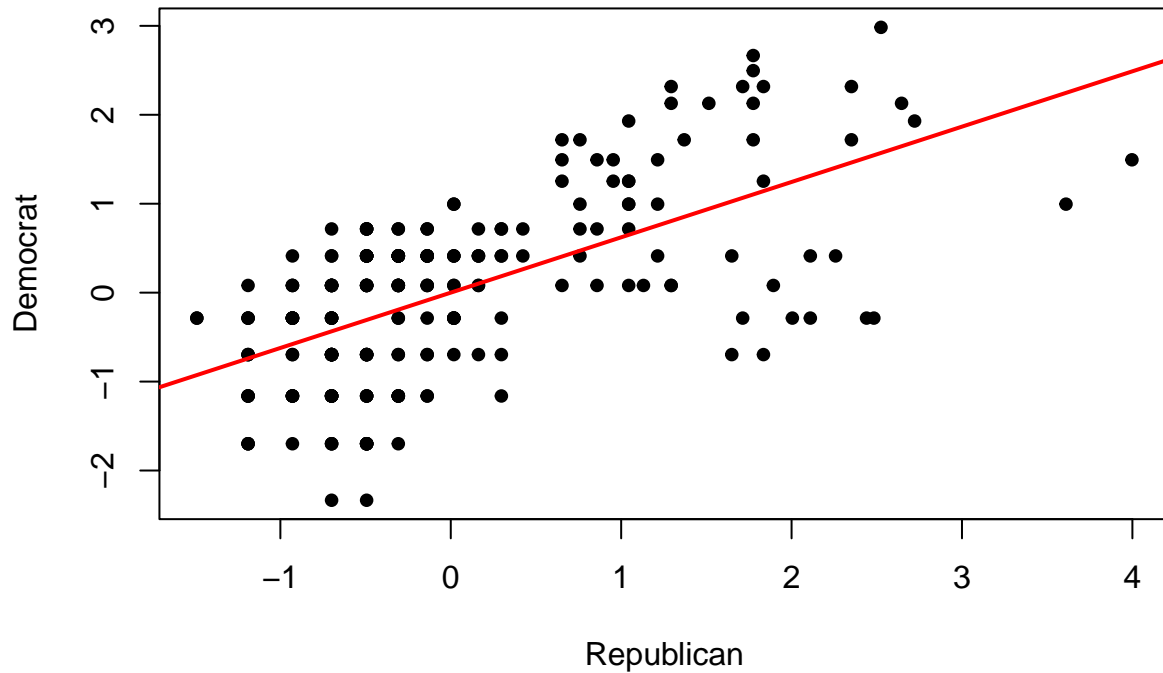


The search interest variable for ‘Democrat’ passed the model utility test for a linear relationship with the conflict index and returned an R-squared value of 0.17, stronger than the largest R-squared value for one of the generic government terms. The associated decrease in conflict index was 0.0422 units for every 1 unit of increased search interest. Similar to the previous findings that searches for “Government” were associated with lower partisan conflict, this is intriguing.

The summed search interest for each of the partisan terms also passed the model utility test and presented a modest R-squared, but this likely occurs mostly because the democrat term is associated with lower conflict, so it does not reveal new information.

As one might expect, search interest for ‘Democrat’ and ‘Republican’ are often both high during the same periods, as the regression plot below demonstrates. This makes the previous finding more interesting: though there exists a moderately strong correlation ( $R\text{-squared} = 0.38$ ) between the two searches, only searches for ‘Democrat’ are associated with any change in the conflict index.

## Search Interest for 'Democrat' vs. 'Republican'

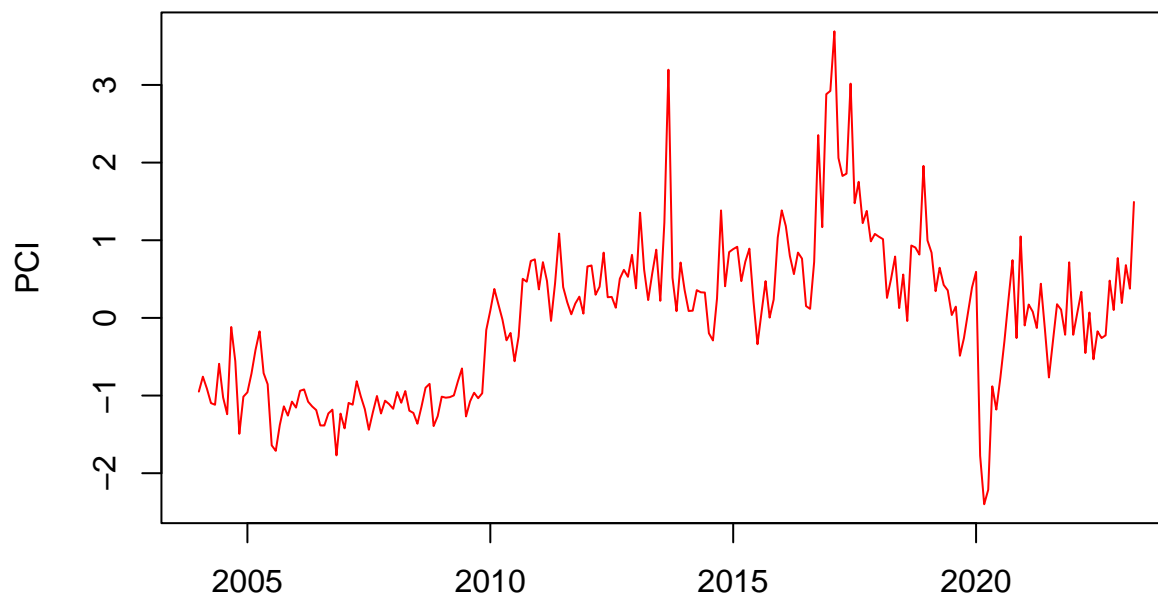


### Time Series Analysis

Since the data from both the Partisan Conflict Index and Google Trends are associated with time - each were collected on a monthly basis for the analysis period - a basic time series analysis could be fruitful.

A quick look at the time series for the Partisan Conflict Index shows a general upward trend over time, an observation backed up by research from numerous political scientists (not to mention the lived experience of those of us involved in the American political system).

## Partisan Conflict Index, 2004–2023

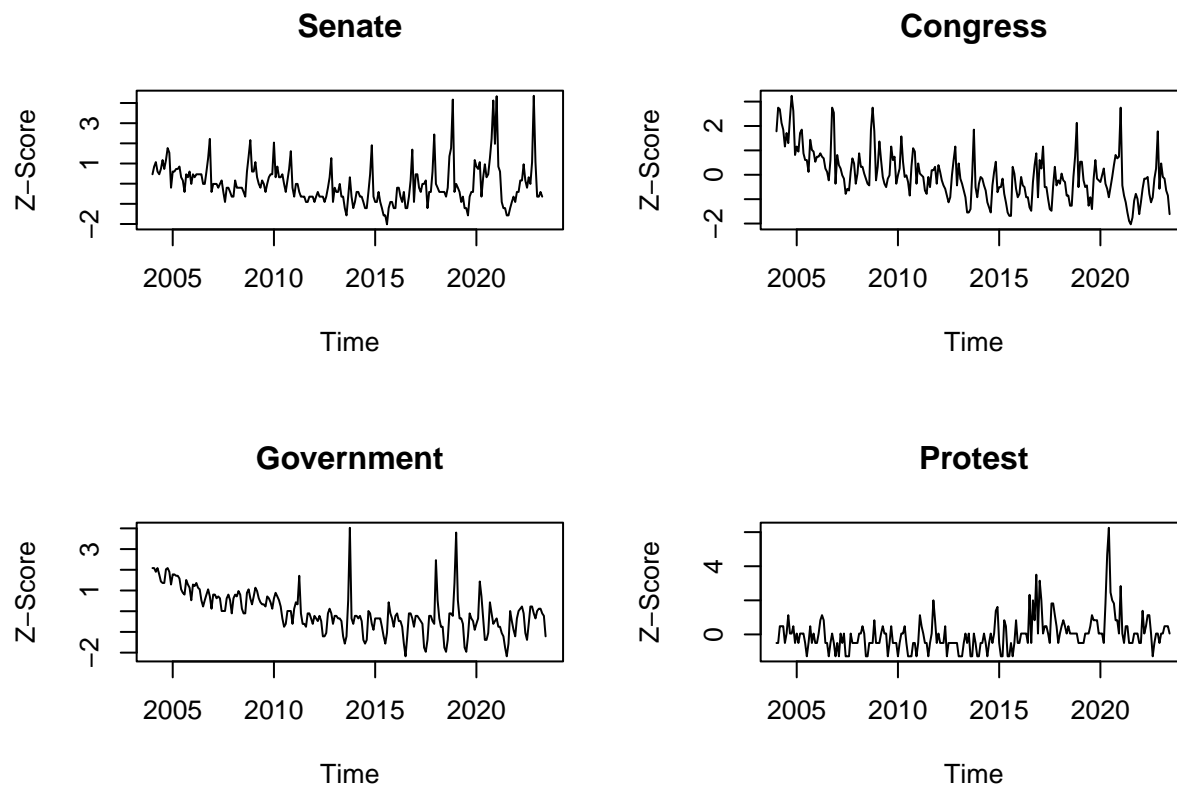


Within this overall trend, several noticeable peaks and valleys are evident. In September 2013, the month leading up to the U.S. government shutdown that year resulted in one of the highest conflict scores. The highest score, however, was in February of 2017, the first full month of Donald Trump’s controversial presidency.

The lowest conflict score came in March 2020, when the nation was first gripped by the COVID-19 pandemic. This was a period when lawmakers were focused on combatting the pandemic and its impacts on American life and the nation’s economy, something that brought together both sides of the aisle on many initiatives.

Interest in the various government-related search terms was quite variable over the analysis period,

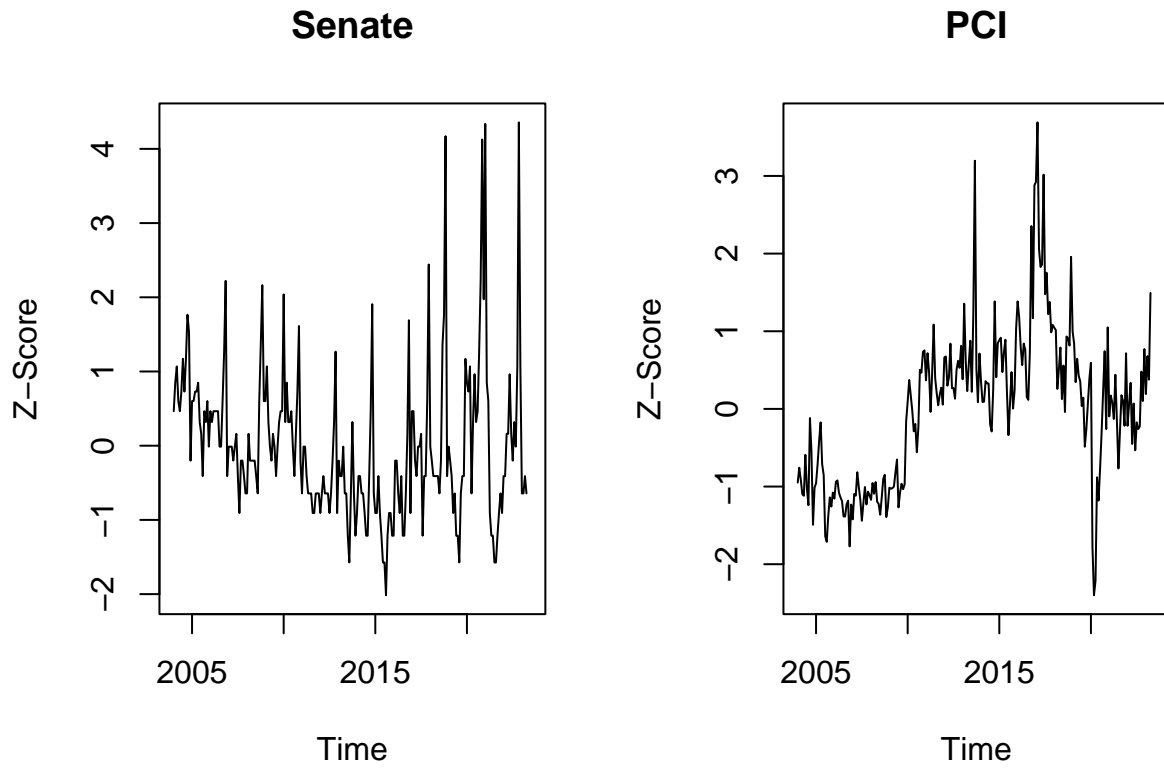




Each of the search terms has peaks that correspond to real-life events: searches for “Senate” and “Congress” increase around election years. Searches for “Government,” similar to the Partisan Conflict Index, peaked in 2013, around conversations about the government shutdown. Searches for “Protest” peaked in 2020, when Black Lives Matter protests were held in most major cities in the U.S.

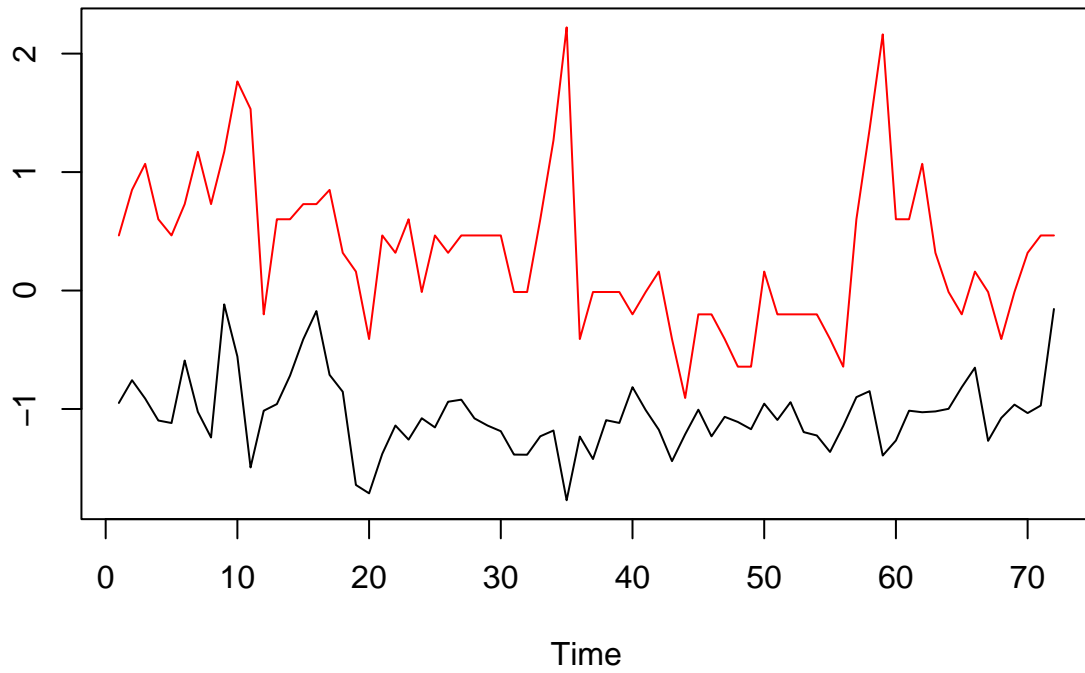
When we consider the side-by-side plots of PCI with each of the terms, we see certain similarities in when each of them rise and trough, but no single term seems to trend too similarly to PCI.

Search interest in “Senate” and PCI follow a similar trend between 2005 and 2010

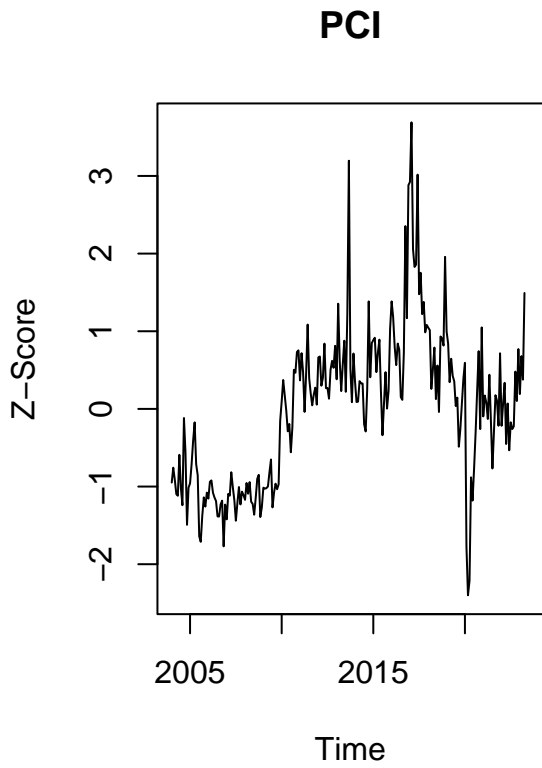
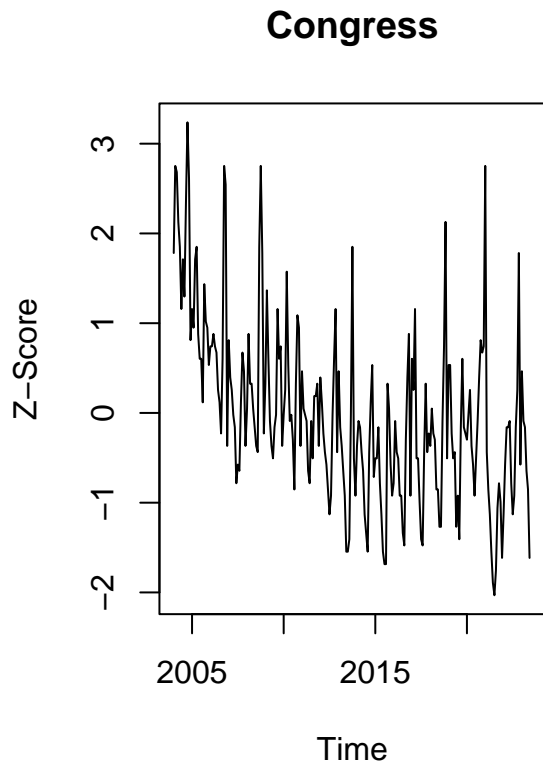


If we zoom in to the 2000-2010 period, save for a few instances where the two series diverge, they appear to share similar patterns. It seems likely that increased partisan conflict occurred at similar times to when Google users were most likely to search “Senate,” perhaps to keep tabs on a particularly contentious issue.

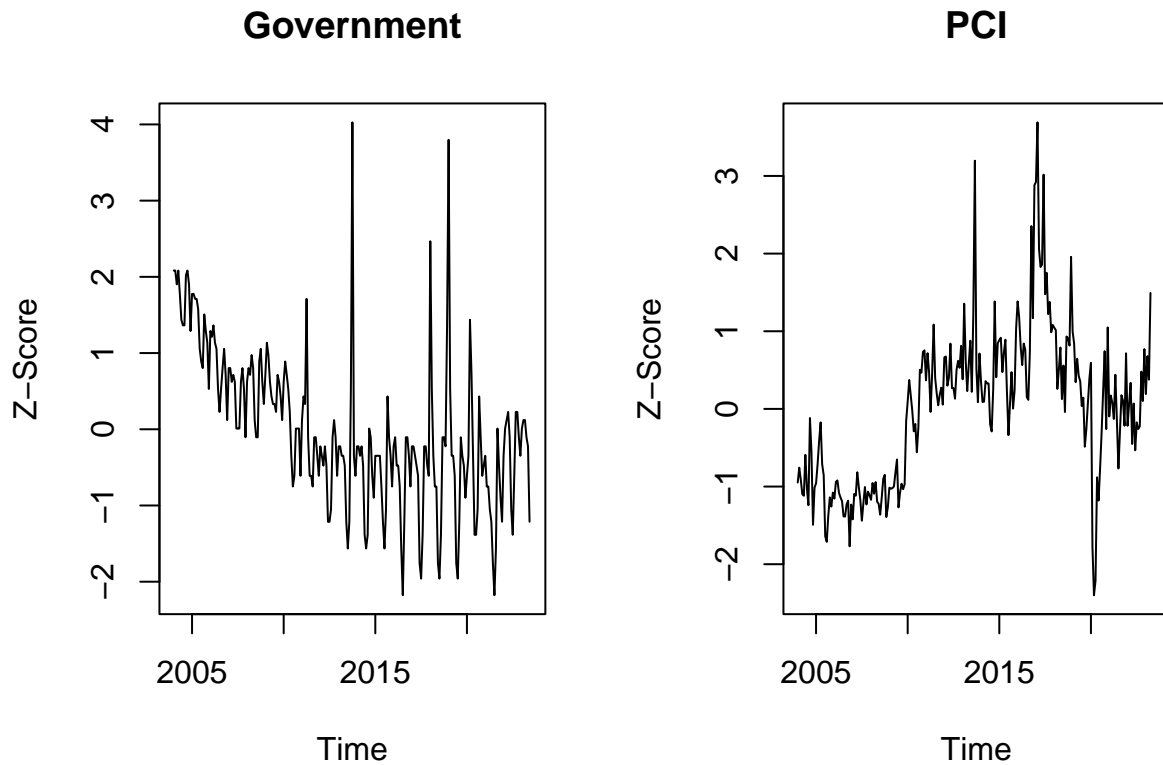
### 'Senate' Search Interest (Red) and Partisan Conflict (Black), 2000–2010



Searches for “Congress” are so variable, likely due to upward trends every 2 years for elections, that any patterns similar to PCI trends are difficult to parse.

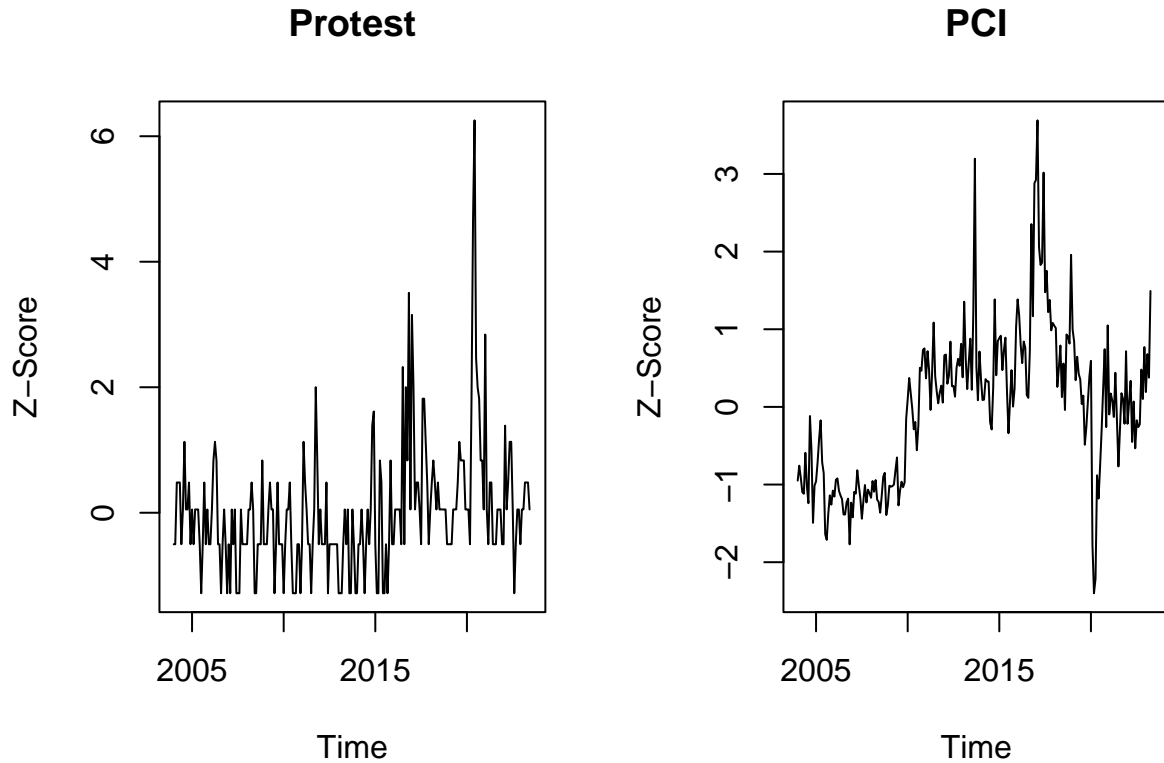


Overall trends for “Government” searches and PCI don’t seem to line up, but the aforementioned spikes (the 2013 government shutdown, e.g.) appear in both.



Intuitively, the search term “Protest” might be linked with periods of greater social unrest, the same periods which one might assume lead to greater partisan conflict. However, with the exception of a spike in the months leading up to and the first months of the Trump presidency, no similar patterns appear in the data.

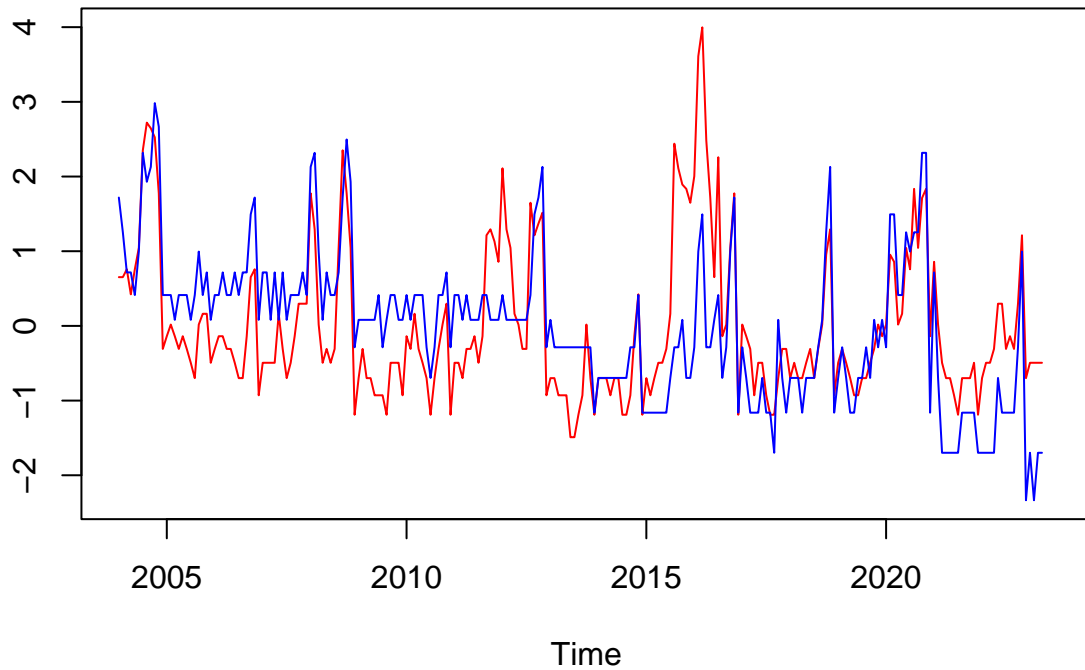
In fact, the most prominent spike in ‘protest’ searches occurred in the same month as the most prominent trough in PCI, March 2020, when Black Lives Matter protests met the impending COVID-19 emergency.



Partisan searches for “Democrat” and “Republican” may be linked to PCI as well. It follows logically that when U.S. politics are more partisan (periods with a high Partisan Conflict Index), more people will search terms related to the parties themselves.

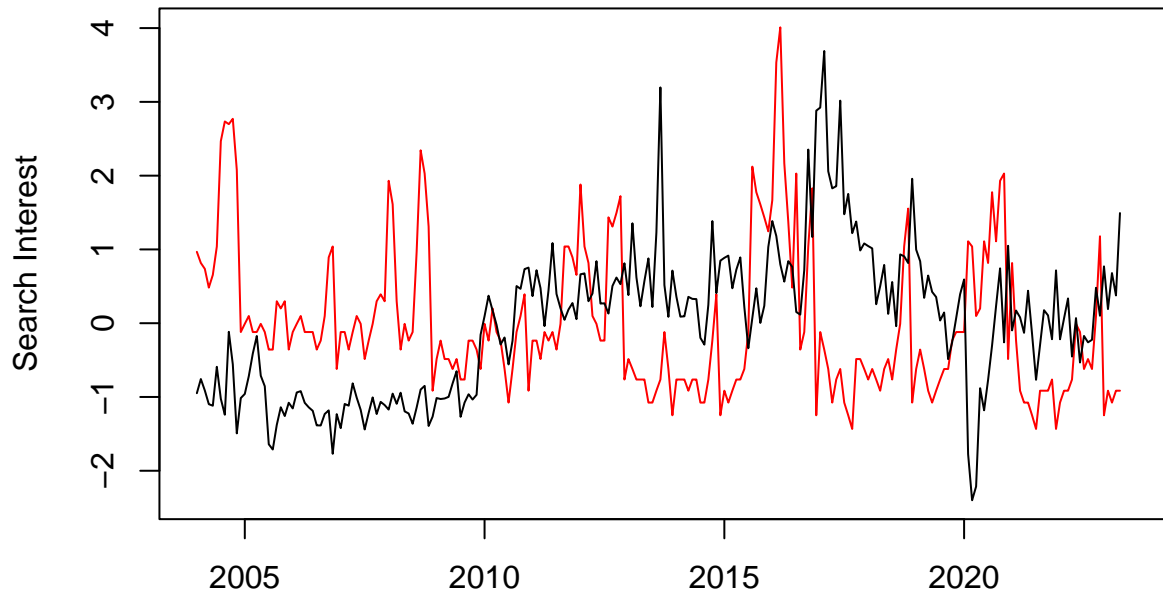
Search trends for partisan terms are very similar over the analysis period. As a general rule, when searches for “Democrat” were high, searches for “Republican” were as well, and vice versa. One outlier occurred in 2016, when the Republican candidate (and eventual winner of that presidential election) Donald Trump rose to political fame.

### Search Interest in 'Democrat' (Blue) and 'Republican' (Red), 2004–2023



When we plot the sum of search interest for partisan search terms versus the Partisan Conflict Interest, the temporal patterns do not appear to mirror each other, though they are not wildly different.

### Searches Interest for Partisan Terms (Red) and Partisan Conflict Index (Black), 2004–2023



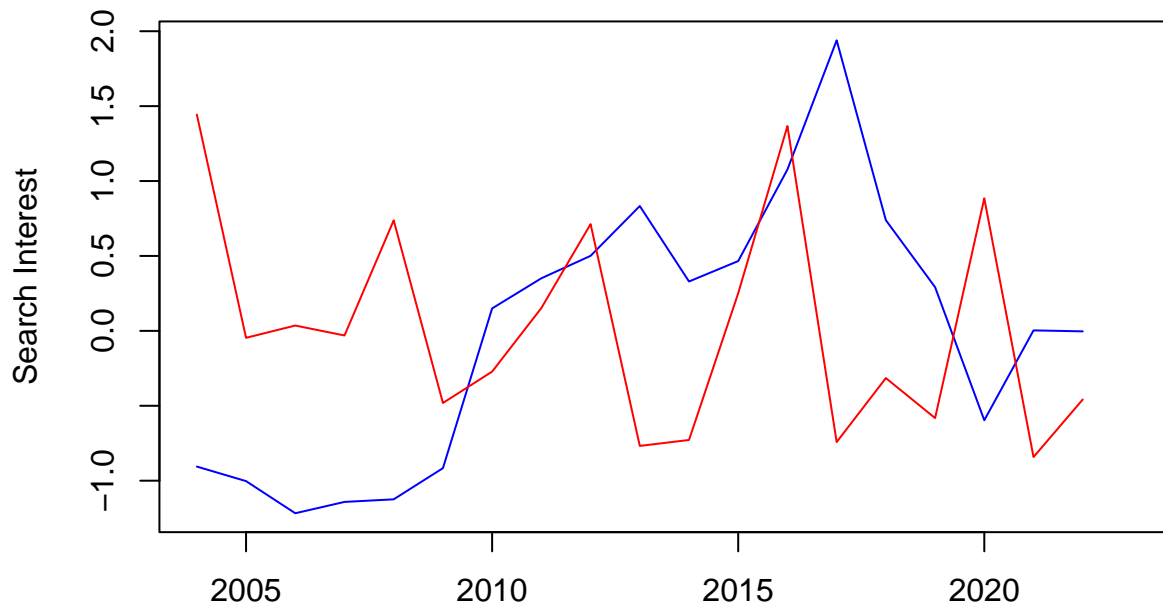
When we break down the comparison to the average of each variable by year, the finding holds.

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduced by coercion
```

```
## Warning in xy.coords(x, y): NAs introduced by coercion
```



### Yearly Avg PCI (blue) and Avg Partisan Search Interest (red), 2004–2023



#### Conclusions:

1. Increased Google searches for the term “Democrat” and the term “Government” were associated with lower partisan conflict as measured by the Partisan Conflict Index.
2. Search interest in terms related to both the Democratic and Republican parties is correlated, and surges in interest occur at roughly the same time, with several key exceptions (e.g., the 2016 presidential primary and campaign).
3. None of the search terms I analyzed showed temporal patterns that closely resembled the conflict index, though the summed search interest in “Democrat” and “Republican” was closest.
4. Search interest remains a largely untapped tool to understand the link between what is on the minds of the general populace and sociopolitical questions.

#### Ideas for further analysis:

1. Using similar Google Trends data for political candidates, along with those candidates’ polling results during key points in the election cycle, identify correlations and predictive power of search results on polling results and election outcomes.
2. Using Google Trends data for products made by different companies, along with data on stock prices, analyze the relationship between searches for, say, the latest Iphone, on Google, and Apple’s stock.
3. Expand from regression and time-series analysis into predictive analytics, applying supervised and unsupervised learning techniques to situations like those described above.