Building Longitudinal *Google Trends* to Measure Dynamic Local-level Issue Attention*

Taeyong Park[†] Haewoon Kwak[‡] Jisun An[§]

Abstract

Google search indices can be useful for measuring dynamic local-level attention to public issues for which survey data is rarely available. However, there is a practical difficulty with constructing longitudinal Google Trends. Since Google Trends provides normalized counts from zero to 100 instead of absolute query counts, its search indices generated for different time windows are not on the same scale. Thus, merely pooling cross-sectional data sets does not produce a desirable longitudinal data set. To resolve this problem, we develop a method for rescaling Google Trends data and building longitudinal data sets at the media market level in the United States. We illustrate how to use longitudinal Google Trends to measure public attention to the issues of employment and the coronavirus.

*This paper is supported by the Qatar Foundation-CMUQ Seed Research Project Fund.
†Arts and Sciences, Carnegie Mellon University in Qatar.
‡Qatar Computing Research Institute, Hamad Bin Khalifa University.
§Qatar Computing Research Institute, Hamad Bin Khalifa University.

Google search data have become a popular source for measuring public attention to political issues. Internet searches reveal a strong interest in specific issues. The traces people leave during internet searching are naturally occurring, not affected by researcher intervention (Japec et al. 2015). Additionally, while people are honest when searching for information on the internet, they could lie or misreport in surveys to give socially desirable answers (Stephens-Davidowitz 2017; Tourangeau, Groves and Redline 2010). Hence, internet searches may be an accurate representation of the minds of the public.

In particular, the vast amount of real-time internet searches can track public interest in issues that tend to be dynamic and ephemeral when the static nature of "most important problem" (MIP) surveys fails to capture it (Ripberger 2011). Furthermore, MIP data are rarely available at local levels, such as counties and media markets, even though how issue attention changes over time may depend on local contexts. Thus, internet search indices, including Google Trends, can be especially useful to measure the local-level dynamic rise and fall of specific issues.

However, little research has explored longitudinal Google search data.[1] Although Mellon (2014) and Ripberger (2011) study time-varying Google search indices regarding such issues as employment, immigration, terrorism, and health care, they focus on national-level time-series data. The lack of research utilizing longitudinal Google searches is essentially due to the lack of longitudinal data per se. One of the major obstacles to constructing longitudinal Google searches is the practical difficulty with merging multiple regions' search data over time. Google Trends does not provide absolute query counts. It provides a relative pattern of search volumes, based on normalized counts from zero to 100. It is straightforward to compile either time-series data for a particular region or cross-sectional data for a specific period using the normalized counts. However, merely pooling multiple cross-sectional data

[1] We focus on Google searches. Google is the world's dominant search engine, fielding about 92 percent of the world's online queries (`https://gs.statcounter.com/search-engine-market-share`). The Google Trends website and API allow users to collect Google search data. We use the Google Trends API to collect the data used in this article. The Python code is available in the online appendix.

sets obtained in different time windows does not produce a desirable longitudinal data set.

To resolve this challenge, we provide practical guidance on building longitudinal Google Trends and explore potential applications to dynamic local-level issue attention. While the feasibility of longitudinal Google Trends as a measure of time-varying local issue attention should be tested in future research, it is imperative to create longitudinal search data. Indeed, testing for the feasibility is impossible unless longitudinal search data exist.

In the following section, we discuss why longitudinal search indices are useful as a measure of dynamic issue attention at the local level. After that, we suggest a novel method for rescaling Google Trends indices to construct longitudinal data sets with an illustrative example of the employment issue. Then, we apply the method to the issue of the COVID-19 pandemic. In the final section, we conclude with some caveats and future research.

Local Issue Attention and Google Search Data

We focus on longitudinal Google Trends data as a measure of issue attention at the media-market level in the United States.[2] Google Trends provides numerous countries' subregion data, including U.S. state searches. However, data on the 210 U.S. media markets, or Designated Market Areas (DMAs), are the only metro-level data accessible by Google Trends. Though city-level searches are also available, Google Trends does not cover all cities in any single country. Thus, currently, only the U.S. DMA data permit exhaustive local-level comparisons of Google searches.

Local-level comparisons are important in the study of political attitudes towards public issues. How people perceive public issues may vary by the extent to which people acquire information from their local experiences, such as conversations with family and co-workers and social communities. Previous research suggests local contexts matter for people's per-

[2] A media market is a geographic area where people receive the same television station offerings. We found that Google Trends media market data corresponded to Designated Market Areas as defined by Nielsen: `https://www.nielsen.com/us/en/intl-campaigns/dma-maps/`.

ceptions of the national economy (Ansolabehere, Meredith and Snowberg 2014; Reeves and Gimpel 2012) and racial groups (Wong 2007) as well as their political competence (Shaker 2012). Cramer (2016) argues that local communities function as a lens through which people interpret politics and that rural people's resentment toward large cities shapes their political views.

Scholars have also examined the role of local media markets in influencing political attitudes towards public issues. Gilliam and Iyengar (2000) show that the racial element of the crime script in local television news increases support for punitive approaches to crime and heightens negative attitudes about African-Americans among white viewers. Filla and Johnson (2010) and Althaus and Trautman (2008) suggest that people's exposure to local news matters for voter turnout.

Given that local contexts affect how people perceive and respond to public issues, it is worth measuring public attention to political issues at the local level. Particularly, capturing the dynamic rise and fall of issue attention is critical because it tends to vary across time according to the ebbs and flows of prominent events (Dennison 2019; Jones 1994; McCombs and Zhu 1995; Ripberger 2011). For instance, the coronavirus pandemic continues to evolve and affect different regions across the U.S. differently. Moreover, public attention to the virus might subside rather quickly.

Unfortunately, MIP surveys, the conventional method for measuring the salience of public issues, are rarely conducted at the local level. In particular, a longitudinal panel survey that permits local-level comparisons of dynamic issue attention is, to the best of our knowledge, nonexistent. Alternatively, longitudinal Google Trends can provide accessible and flexible measures. The method we develop below focuses on DMA-level Google searches though it is also applicable to state-level and cross-national data.

Building Longitudinal Google Trends

Problem

Google Trends provides normalized counts from zero to 100 instead of absolute query counts. Hence, cross-sectional data sets obtained in different time windows are not directly comparable. To illustrate how these features make it challenging to create longitudinal data sets, we take an example from the issue of employment.

Following Mellon (2014), suppose we use the Google search term "jobs" to measure DMA-level prominence of employment issues weekly from September 18 through November 26, 2016. Search results for the 10 weeks and 10 DMAs are displayed in Table 1.[3] The search index is given by a normalized 0–100 scale for each week, such that Rochester's search index is 50, and Spokane's index is 60 in week 9. These quantities indicate that Rochester's search volume is about half of the highest search volume in that week and about 5/6 of Spokane's search volume.[4]

Such inter-regional comparisons are possible in the same search window, in the same week in this example. However, normalized indices in different time windows are not on the same scale. For example, Austin's search indices 38 and 55 for the first and second weeks are proportional to the highest search volume for each of the two weeks. It is uncertain which index, 38 or 55, corresponds to a higher search volume. Therefore, simply combining the 10 columns in Table 1 does not generate a desirable longitudinal data set.

Rescaling search indices

We develop a method for rescaling Google Trends search indices to convert them into a common scale. We continue to use the example of the jobs index to describe this method. Specifically, we illustrate how to build a longitudinal data set for the weekly DMA-level jobs

[3]For simple discussion, we randomly chose the 10 DMAs.

[4]Google Trends returns only integers without providing information about how it rounds decimals. Therefore, rounding errors may occur.

Table 1: Weekly Google Searches for "jobs" for the Period of September 18 - November 26, 2016: 10 Designated Market Areas

	Week									
DMA	1	2	3	4	5	6	7	8	9	10
Austin TX	38	55	55	55	42	52	50	46	55	19
Baltimore MD	43	66	62	62	47	66	65	56	67	23
Baton Rouge LA	39	55	59	66	40	65	51	54	64	22
Honolulu HI	38	57	55	57	38	58	54	51	57	19
Lexington KY	44	60	61	62	44	72	71	53	67	21
Rochester NY	33	49	44	48	30	46	44	44	50	17
Spokane WA	40	50	49	50	39	56	57	51	60	19
Syracuse NY	32	49	50	53	34	49	47	49	49	18
Toledo OH	37	54	57	55	40	56	47	55	64	19
Wilmington NC	37	67	51	57	46	62	48	51	63	24

[*Notes*: Week 1 refers to September 18 - 24, Week 2 means September 25 - October 1, and so forth.]

index for the period between December 31, 2006, and December 31, 2016. We provide a three-step guideline.

In the first step, we select a reference that supplies a common scale. Time-series search data for individual DMAs are accessible by Google Trends and can be used as our reference because they are on the same scale across different time windows. We use the time-series data for Honolulu, Hawaii, as our reference without loss of generality. Though any DMA time series would work for the reference, we recommend that researchers select a DMA with a large population size to reduce the possibility that the reference index has a value of zero.[5]

Second, we may need to divide the reference time series into subset series and concatenate them after rescaling, depending on the length of the time under consideration. We call this the divide-and-concatenate process. For instance, our running example covers the period between December 31, 2006, and December 31, 2016, and this period is long enough to require the divide-and-concatenate process. We found that Google Trends returned a monthly time series, instead of a weekly series, for any period exceeding 269 weeks.[6] Hence, we divide the

[5]A DMA with a large population is less likely to have a search amount less than 1/100 of the highest search volume for a particular period, reducing the odds of having an index with a value of zero. Below, we discuss why it is advantageous to prevent a value of zero for the reference index.

[6]We found that when Google Trends returns a time series with more than 270 data points, it appears to

Table 2: Concatenation for the Reference Time Series: Searches for "jobs" in Honolulu

Search Window (Weekly)		Before Rescaling	After Rescaling
	⋮	⋮	⋮
Window 1	2/5/12	40	114.2857
	2/12/12	34	97.14286
	2/19/12	35	100
	2/26/12	35	100
Window 2	2/26/12	79	100
	3/4/12	80	101.2658
	3/11/12	74	93.67089
	⋮	⋮	⋮

entire period into two: December 31, 2006 – February 26, 2012, and February 26, 2012 – December 31, 2016. In so doing, we ensure that neither of the two divided periods exceeds 269 weeks, and Google Trends generates weekly data. Contingent upon whether the time span under consideration exceeds 269 data points, researchers may need to decompose their reference time series into more than two windows or may not need to divide at all.[7]

When dividing the reference time series, it is crucial to have the preceding time window's end date identical to the following window's start date to employ it as a node connecting the two windows. Since each time window produces 0-100 normalized counts, the node belonging to both windows should be used to rescale the normalized counts, transforming them to the same scale. As Table 2 shows, the week of February 26 is the node, and we begin by converting the node value into 100 in both windows to obtain the scaling ratios: 100/35=2.857143 for the first window and 100/79=1.265823 for the second window.[8] Then, we multiply all the first-window values by 2.857143 and multiply all the second-window values by 1.265823 to rescale them. The rescaled time-series indices for Honolulu in the

change its time structure to limit the number of data points. For example, when Google Trends is asked to return a time series for 269 days, it returns daily data. If the time span becomes 270 days, it returns weekly data. Similarly, when the period spans 270 weeks, it returns monthly data instead of weekly data.

[7]We illustrate a single time-window case later in our second example – COVID-19.

[8]If the node has a value of zero, the scaling ratio is not computable. One solution is to find the first nonzero from the following weeks and make the second time window start from there. Another solution is to select a different reference DMA with a large population so that it is less likely to have a value of zero.

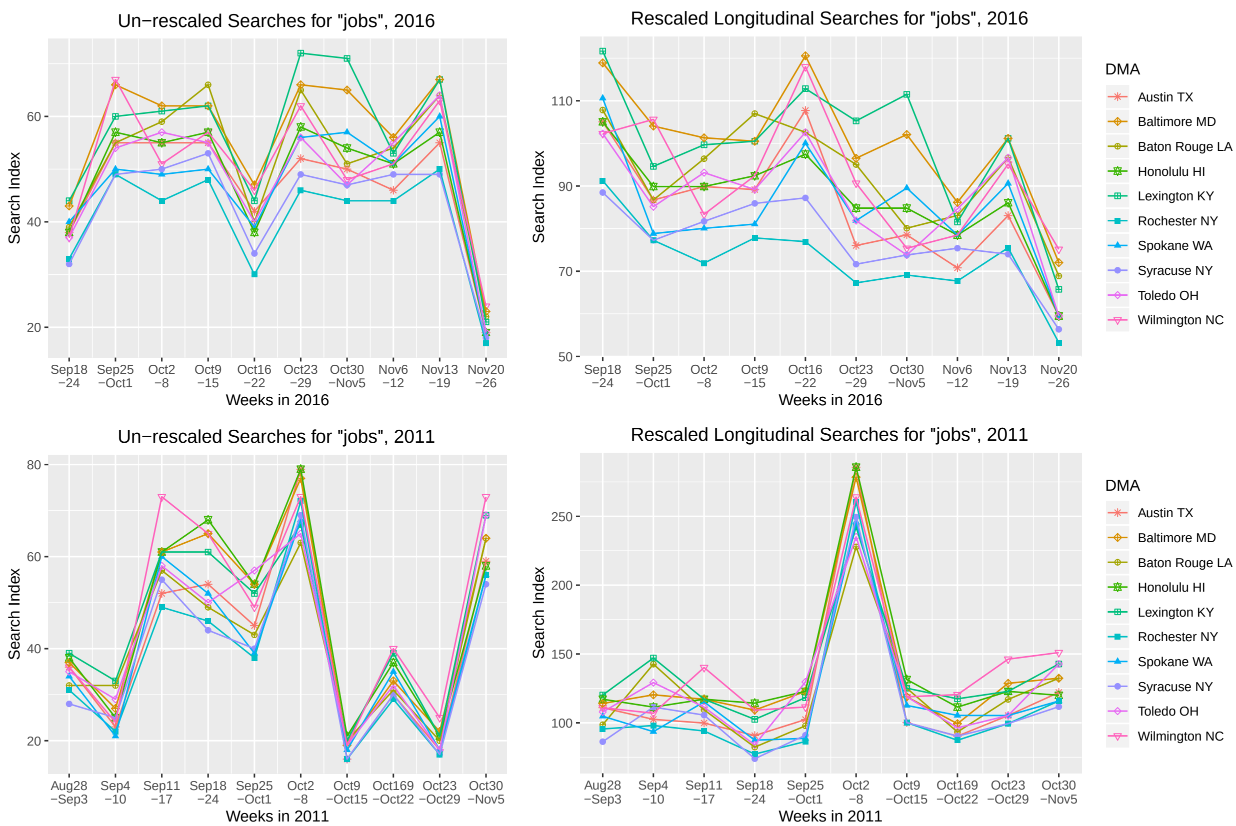
second column of the table are our reference time series.

Once the reference time series is ready, the third step is to rescale all the cross-sectional indices. For example, consider the first week from Table 1. Baltimore has a value of 43, and Honolulu has a value of 38. For the same week, the reference data of the Honolulu time series give a value of 105.0632911. We begin by calculating the ratio between the Honolulu time series and the Honolulu cross-sectional index. This scaling ratio is computed as $105.0632911/38 = 2.764823$. Then, we multiply all the cross-sectional search indices in the corresponding week by this ratio. Accordingly, the value for Honolulu is converted to 105.0632911, and the value for Baltimore turns $43 \times 2.764823 = 118.8874$. For the second week from Table 1, the scaling ratio is $89.87341772/57 = 1.576727$, where the value of 89.87341772 is from the reference data of the Honolulu time series. The cross-sectional index for Honolulu is adjusted from 57 to 89.87341772, and Baltimore's index is rescaled to $66 \times 1.576727 = 104.064$. Then, all the converted values in the DMA data sets, such as 105.0632911, 118.8874, 89.87341772, and 104.064, are on the same scale and comparable with one another.

Based on these three steps, we built a longitudinal data set by merging the rescaled cross-sectional data. The upper panel in Figure 1 corresponds to the 10 DMAs and 10 weeks in Table 1. The lower panel shows another subset for the selected 10 weeks in 2011. For each panel, the right plot presents rescaled longitudinal searches while the left plot presents pooled cross-sectional searches readily accessible by Google Trends.

Comparing the pooled cross-sectional data and the longitudinal data derives several important implications. First, the longitudinal data track the actual trend over time that the pooled cross-sectional data fail to capture. For instance, according to the upper-panel left graph, the first week's search indices are overall lower than the second week's indices. However, once the time-series information is interwoven with the cross-sectional data through the rescaling process, it turns out that the second week has higher search volumes overall. Similarly, the pooled cross-sectional data in the lower panel do not capture the precise

Figure 1: Dynamic Issue Attention to “jobs” by Designated Market Area: Comparing Pooled Cross-Sectional Google Searches and Longitudinal Google Searches



amount of the ups and downs of the searches that occurred in the week of October 2, 2011. On the contrary, the rescaled search indices in the right graph track the sudden rises and falls accurately. It appears that the dramatic surge is due to the death of Steve Jobs, the co-founder of Apple Inc., on October 5, 2011. When using search indices as a measure of issue attention, it is critical to check for their content validity. For example, the job search index, as a measure of the prominence of employment issues, must not be confounded by searches for “Steve Jobs” (Mellon 2014). In this regard, tracking accurate time-varying searches can be used to detect the presence of confounders.

Second, the longitudinal search indices mitigate the problem of seemingly-flawed search results that the smallest DMAs by population have the highest search index. For instance, the job search index for Glendive, Montana, is 100 in 240 weeks of the 522 weeks in our analysis, despite its small population of about 4,000. A few other DMAs, such as Alpena in

Michigan, North Platte in Nebraska, and Zanesville in Ohio, have similar patterns. These cases make the pooled cross-sectional data even further inconsistent with the actual trend. Let us take an example of the deep valleys in the lower-left plot that appeared in the weeks of October 9 and 23 in 2011. In these weeks, Glendive's index is 100, and all other DMAs' indices range from 10 to 40, leading to the deep valleys in the plot. Although these search results seem irregular, it is hardly possible to discover whether they are valid or invalid until Google Trends provides relevant information. However, we suggest that using the longitudinal search indices can lessen related concerns. The time-series component in our rescaling scheme takes into account other weeks' search indices and reduces the influence of outliers. As a result, the extremely bumpy ups and downs revealed in the pooled cross-sectional data do not occur in the longitudinal data. On the other hand, the decline in the Thanksgiving week of November 20 in 2016 in the upper-left graph seems valid since the longitudinal data set shows a similar trend.

Application to COVID-19

We now turn to two examples to illustrate the potential applications of longitudinal Google Trends. The first analysis uses daily searches for "coronavirus" over the period between February 1 - May 9, 2020. We track how public attention to the pandemic varies over time in different DMAs. After that, we analyze daily searches for the Centers for Disease Control and Prevention (CDC), CNN, and Fox News over the same period. We examine public attentiveness to the three different information sources during the pandemic to see if the type of information sources people seek varies as their attention to the issue evolves over time.

Analysis 1: Dynamic public attention to the pandemic

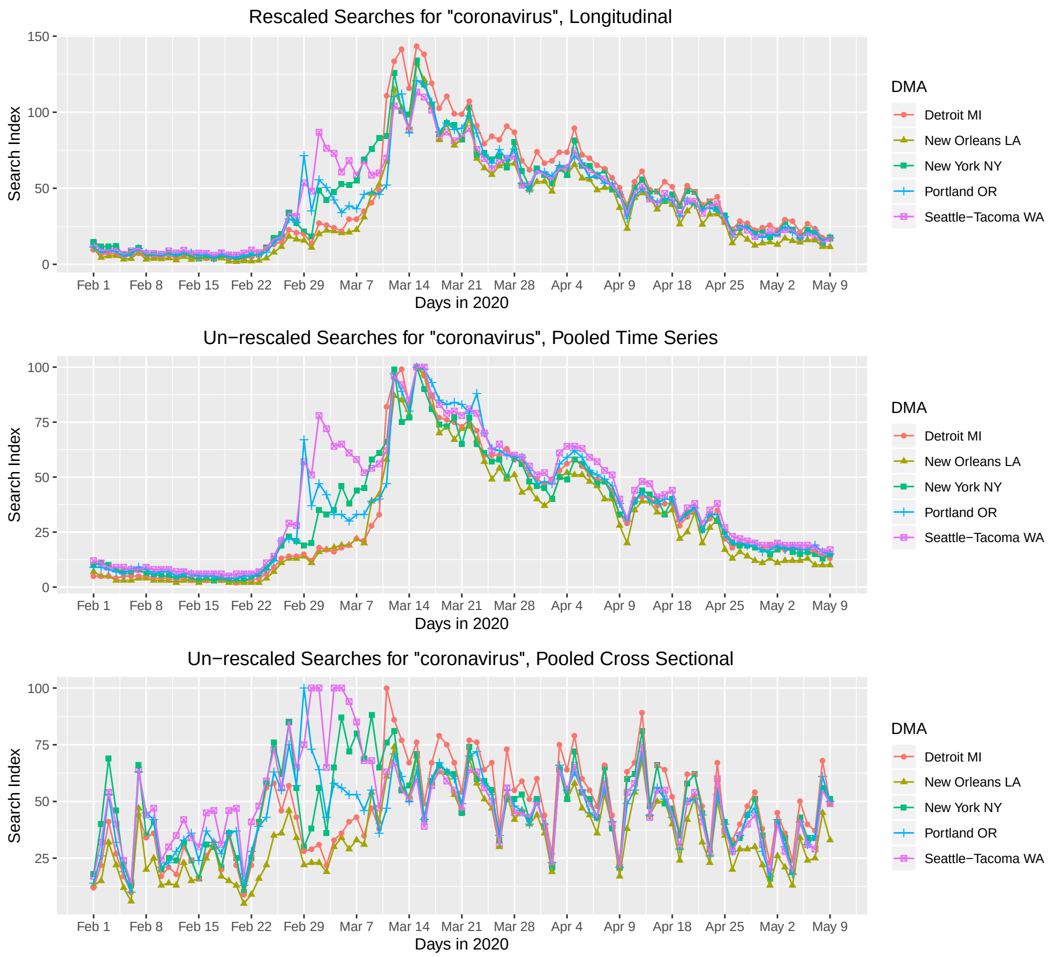
We use the Honolulu time-series data as our reference to rescale daily searches for "coronavirus." The time period in our analysis spans 99 days. Google Trends returns a daily time series for the entire period, and the divide-and-concatenate process is unnecessary.

The top plot in Figure 2 presents five examples of longitudinal daily searches. It appears that the American public had been relatively less concerned about the coronavirus until late February when Italy started seeing significant increases in confirmed cases. Since then, the coronavirus had been a more salient issue in Seattle-Tacoma and Portland until early March than any other DMAs. These results make sense because Seattle was the first U.S. epicenter of COVID-19. As New York City and New Orleans emerged as next epicenters of the pandemic, the issue became more salient in those regions than Seattle-Tacoma and Portland in mid-March. At the same time, Detroit started seeing higher search volumes than New York. Though we do not present here, we also found that the trends of the Lansing and Grand Rapids-Kalamazoo-Battle Creek DMAs were similar to Detroit's. Perhaps, the coronavirus outbreak was particularly salient to Michigan residents. Consistent with this inference, a survey using geotagged twitter data shows that Michigan residents complained about depression and anxiety during the virus crisis more than people in any other state.[9]

These dynamic rises and falls of public attention to COVID-19 in different DMAs are not accurately captured by pooling time-series searches across DMAs or pooling cross-sectional searches across time. As the middle plot in Figure 2 displays, the pooled time-series searches provide a trend that resembles the longitudinal data. Yet, in each time point, the searches fail to deliver the correct ranks among different DMAs. For example, the pooled time-series data do not detect Detroit's surge at the peak and afterward. In the bottom plot, while the pooled cross-sectional searches present the correct ranks among different DMAs in each day, they fail to capture the accurate trend over time.

[9]Retrieved from MLive.com on June 17, 2020: `https://www.mlive.com/coronavirus/2020/05/michigan-is-no-1-state-for-tweets-about-depression-anxiety-during-coronavirus-pandemic.html`.

Figure 2: Comparing Un-rescaled and Rescaled Longitudinal Google Searches for "coronavirus": 5 DMAs



Analysis 2: Information seeking in the pandemic

In our second analysis, we track longitudinal daily searches for the CDC, CNN, and Fox News over the period between February 1 - May 9, 2020. We utilize daily search indices generated to compare the search terms "cdc," "cnn," and "fox news." The Google Trends website and API allow users to compare up to five distinct sets of search terms simultaneously.

This so-called comparison-search index is given by the ratio of each query's search volume to the sum of the search volumes of the queries under comparison. For example, Philadelphia's comparison-search indices for CDC, CNN, and Fox News for February 1 are 7, 48, and 45. These values indicate that the search volume for CDC is approximately 7/48 and

7/45 of those for CNN and Fox News, respectively. These values are on the same scale only for February 1, and we need a reference time series to convert them into a common scale for the whole 99 days.

We use Atlanta's time-series search indices for CNN as our reference because the Atlanta indices do not have a value of zero over the entire period. This time series data comes from the search for "cnn," not from the comparison search. Since the comparison-search indices for CDC, CNN, and Fox News are on the same scale for each day, time-series indices for a single search term, whether CDC, CNN, or Fox News, allow us to rescale the cross-sectional comparison-search indices.[10]

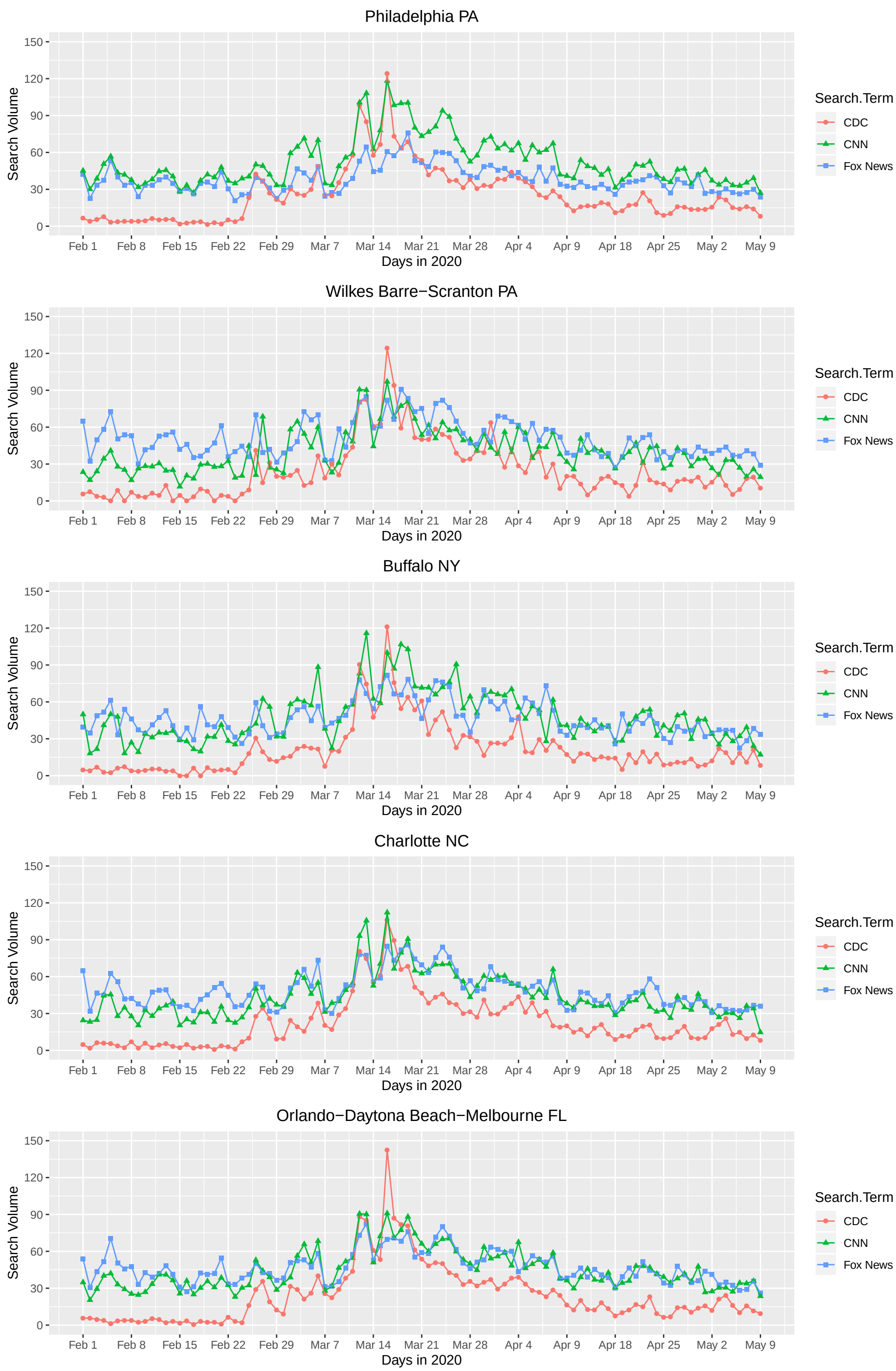
Figure 3 shows five examples of the rescaled longitudinal comparison-search indices for CDC, CNN, and Fox News. From the five examples and many other DMAs, though not presented here, we find that the search for CDC relative to CNN and Fox News starts increasing around February 24 when Italy reported a surge in their coronavirus cases. Then, it hits its peak in mid-March and gradually subsides. This pattern is similar to the overall longitudinal trend for the "coronavirus" search displayed in Figure 2. We deem that the comparison-search indices for CDC are a proxy for public attention to COVID-19.

We also notice a nuanced information-seeking pattern that would not be identified without the longitudinal data. When public attention to the coronavirus was relatively low until February 23, the search volumes for Fox News were higher than those for CNN among the five DMAs except for Philadelphia. As the public attention increased, the concern about the virus outbreak appeared to push up the search for CNN more than Fox News. One hypothetical explanation is that people tend to seek out politically less biased information when searching for updates on the public health crisis, in that Fox News has been alleged of having Republican Party bias in their news coverage.[11] This analysis is more or less

[10] In theory, the individual CDC, CNN, and Fox News time series must produce the same rescaled data. In practice, however, differences could occur because of rounding errors.

[11] In accordance with this hypothetical explanation, monthly U.S. traffic data show that people's visit to partisan websites, such as The Daily Caller on the right and Truthdig on the left, was stagnant or falling between February and March. Fox News also saw only modestly increasing traffic, compared to other large

Figure 3: Rescaled Searches for Information Sources in the COVID-19 Pandemic



outlets' big jumps. Retrieved from the New York Times: `https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html`.

anecdotal though we found similar patterns from multiple other DMAs, where the search for Fox News was higher than CNN before the virus outbreak. Future research should conduct statistical analyses to delve into further the pattern described here, while the present study focuses on illustrating several avenues for the potential application of longitudinal local-level Google Trends.

Additionally, we explore one aspect of the fault line between the rural and urban regions across the country. According to Figure 3, the residents of the rural rust-belt Wilkes Barre-Scranton media market in Pennsylvania tend to search for Fox News, a conservative channel, more than CNN. On the contrary, the residents of the Philadelphia DMA, about 120 miles away from Wilkes Barre-Scranton, overall search for CNN more than Fox News. This rural-urban gap characterized by the different preferred news channels echoes previous research referring to America's rural-urban political divide as the main dimension of political conflict (Cramer 2016; Rodden 2019). The gap is also in line with the 2016 presidential election results. Trump performed well in Wilkes Barre-Scranton as a presidential candidate, receiving 69.7% of the votes in the Lycoming county and 69.4% in the Schuylkill county, among others. By contrast, his vote share was 15.3% in the Philadelphia county within the Philadelphia DMA.[12]

Conclusion

In this article, we develop a new method for constructing DMA-level longitudinal Google Trends. This method is also applicable to the U.S. state-level or cross-national data. We also discuss several avenues for the application of longitudinal Google searches. Measuring the dynamic rises and falls of local-level issue attention can open the door to new research questions.

We conclude with some caveats. First, Google Trends returns only integers without

[12] The vote shares are retrieved from the New York Times website: `https://www.nytimes.com/elections/2016/results/president`.

clarifying how it rounds decimals. Consequently, there could be measurement errors such that two regions with the same search index, in fact, have different search volumes. For the same reason, a value of zero for the search index may range from zero to values like 0.49. This problem might produce an error when rescaling normalized counts of queries. For instance, the coronavirus search index for March 1 is zero for Alpena, Michigan, and 39 for Las Vegas, Nevada. After rescaling, those values become zero and 18.75. If the search volume for Alpena were 0.45, instead of zero, its rescaled value would have been 0.2163461, instead of zero. Unfortunately, researchers cannot avoid this problem until Google Trends provides more precise data with decimals. Depending on search terms and periods, the impact of rounding errors could be large or small. We recommend researchers be aware that such errors might exist and analyze their data with caution.

Second, there is a seemingly-flawed pattern in the smallest DMAs by population, including Alpena, MI, Glendive, MT, North Platte, NE, Zanesville, OH, and a few others. In both job and COVID-19 search indices, we found many instances that those DMAs have a value of 100. These observations seem abnormal considering the population size, but we cannot assess whether they are valid without knowing how Google Trends generates data. Nonetheless, we suggest that the time-series component in our rescaling scheme reduces the impact of outliers, making longitudinal Google Trends less susceptible to the flaw that might exist in Google Trends' data generating process.

With these caveats in mind, we hope that the guidance we provide helps researchers and practitioners build and analyze longitudinal Google Trends. Also, we hope that the present study promotes further discussions about the feasibility of internet search indices as a measure of dynamic issue attention.

References

Althaus, Scott L and Todd C Trautman. 2008. "The impact of television market size on voter turnout in American elections." *American Politics Research* 36(6):824–856.

Ansolabehere, Stephen, Marc Meredith and Erik Snowberg. 2014. "Mecro-Economic Voting: Local Information and Micro-Perceptions of the Macro-Economy." *Economics and Politics* 26(3):380–410.

Cramer, Katherine J. 2016. *The politics of resentment: Rural consciousness in Wisconsin and the rise of Scott Walker*. University of Chicago Press.

Dennison, James. 2019. "A review of public issue salience: Concepts, determinants and effects on voting." *Political Studies Review* 17(4):436–446.

Filla, Jackie and Martin Johnson. 2010. "Local news outlets and political participation." *Urban Affairs Review* 45(5):679–692.

Gilliam, Franklin D and Shanto Iyengar. 2000. "Prime suspects: The influence of local television news on the viewing public." *American Journal of Political Science* 44(3):560–573.

Japec, Lilli, Frauke Kreuter, Marcus Berg, Paul Biemer, Paul Decker, Cliff Lampe, Julia Lane, Cathy O'Neil and Abe Usher. 2015. "Big data in survey research: AAPOR task force report." *Public Opinion Quarterly* 79(4):839–880.

Jones, Bryan D. 1994. *Reconceiving decision-making in democratic politics: Attention, choice, and public policy.* University of Chicago Press.

McCombs, Maxwell and Jian-Hua Zhu. 1995. "Capacity, diversity, and volatility of the public agenda: Trends from 1954 to 1994." *Public Opinion Quarterly* 59(4):495–525.

Mellon, Jonathan. 2014. “Internet search data and issue salience: The properties of Google Trends as a measure of issue salience.” *Journal of Elections, Public Opinion & Parties* 24(1):45–72.

Reeves, Andrew and James G Gimpel. 2012. “Ecologies of Unease: Geographic Context and National Economic Evaluations.” *Political Behavior* 34(3):507–534.

Ripberger, Joseph T. 2011. “Capturing curiosity: Using internet search trends to measure public attentiveness.” *Policy Studies Journal* 39(2):239–259.

Rodden, Jonathan A. 2019. *Why cities lose: The deep roots of the urban-rural political divide*. Hachette UK.

Shaker, Lee. 2012. “Local political knowledge and assessments of citizen competence.” *Public Opinion Quarterly* 76(3):525–537.

Stephens-Davidowitz, Seth. 2017. *Everybody lies: Big data, new data, and what the internet can tell us about who we really are*. HarperCollins New York, NY.

Tourangeau, Roger, Robert M Groves and Cleo D Redline. 2010. “Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error.” *Public Opinion Quarterly* 74(3):413–432.

Wong, Cara J. 2007. ““Little” and “big” pictures in our heads: Race, local context, and innumeracy about racial groups in the United States.” *Public Opinion Quarterly* 71(3):392–412.