Unit 3 Artifact

Data Visualization

Research Question

How does a book's presence on *The New York Times* Best Sellers list affect Google search trends around the time that the list is published?

Data Collection

Data* was collected using [Google Trends](). Google Trends allows you to enter a search term and see its search history over time, and the countries it is most searched in.

The books chosen were those whose first appearance on the *The New York Times* Best Seller list "Combined Print & E-Book Fiction" was during June 2017. Search trends were analyzed from the end of November 2016 to the beginning of November 2017. The month of June was chosen as a 'mid-way' point between the two beginning and end dates, as I wanted to analyze a full year's worth of search trends.

The search trends are recorded on a scale of 0 to 100, with 100 being the most the term has been searched for in the given range of dates (the peak search activity), as reported by Google. The intent is to analyze the dates that the Google Search activity peaked and compare it to the dates that the Best Sellers lists were released online.

If an individual book's search trends showed any abnormalities (no obvious peak, or a non-unique name), the author's name was appended to the search term to provide more accurate results. This was not done to every book, as it is optimal to have as large a sample size as possible, and using a simpler search term pulls more detailed results from Google Trends.

An important note on *The New York Times* Best Sellers lists: The date that each Best Sellers list is actually published is not the week that it reflects sales for. The lists are published two weeks after sales metrics are finalized. However, the lists are posted online far before their publication date, so this research is more concerned with the date the list is published online (when it is first publicly available and can start affecting search trends).

Analysis

Peak Search Date With Respect To List Publication Date
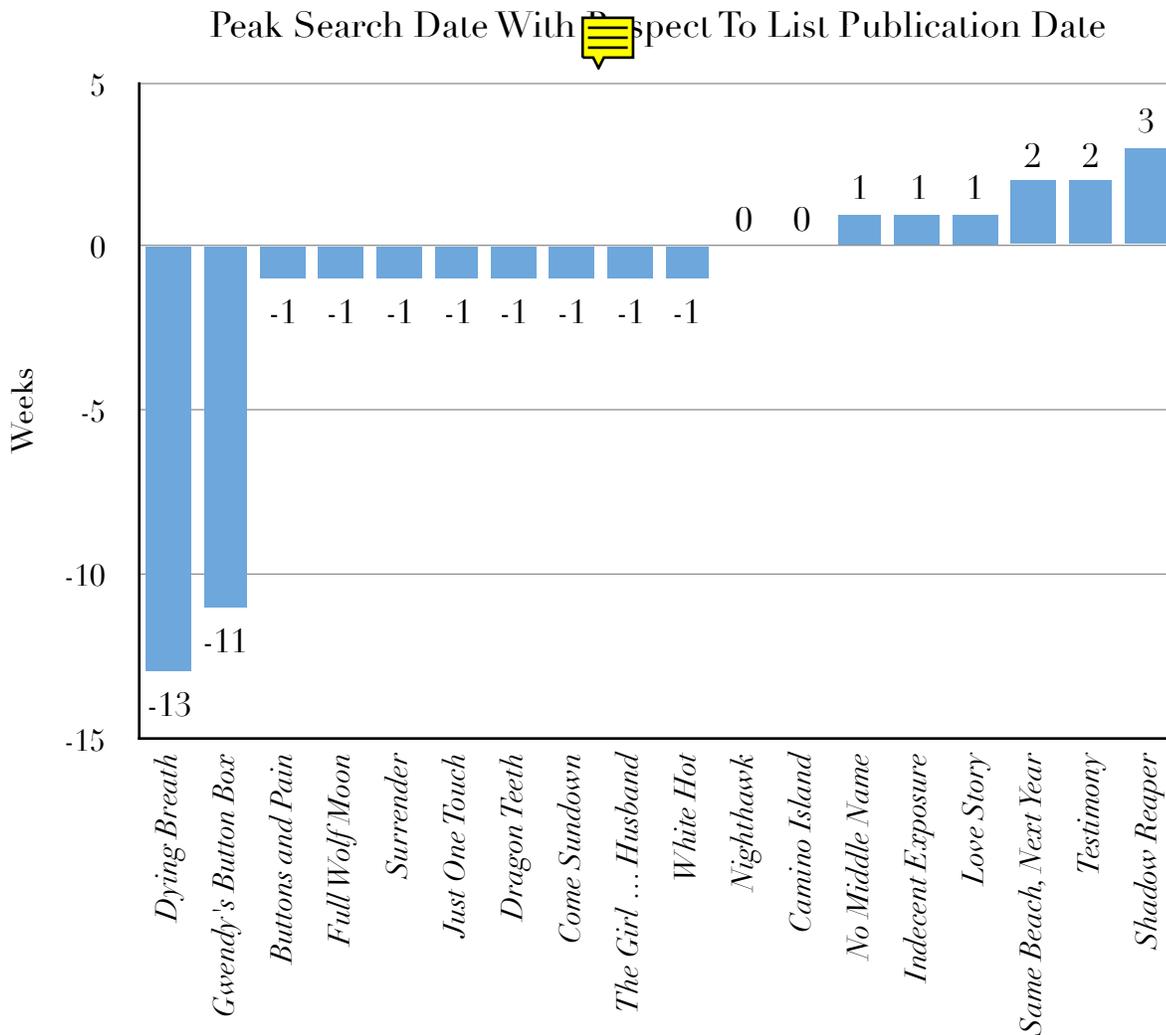
When looking at the peak search date with respect to the Best Sellers list publication date, one can analyze the difference between the two dates. The mode number of weeks before or after the list's publication date that the search trends peak is -1, meaning that most of the books' search interest peaks sometime within a week before the Best Sellers list is posted online. Extrapolating from this, it appears that more times than not, the Best Sellers list almost seems to occur in response to Google Search trends for said books. In other edge cases, the book's peak interest date is long before it's appearance on the Best Sellers list. This may be attributed to the book's announcement, or other unaccounted for publicity surrounding the book and/or author.
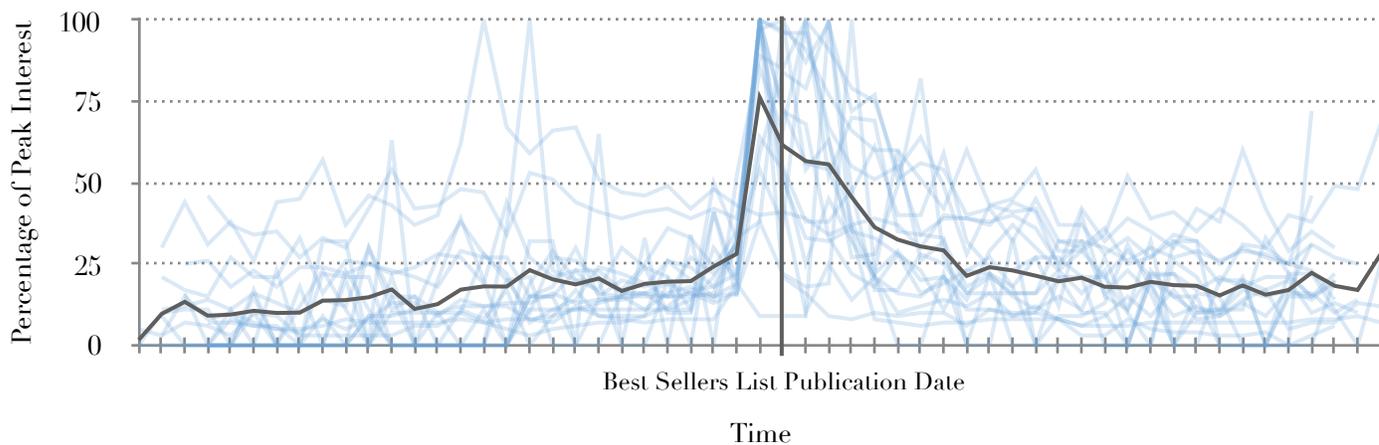
Interest Over Time (All Books)
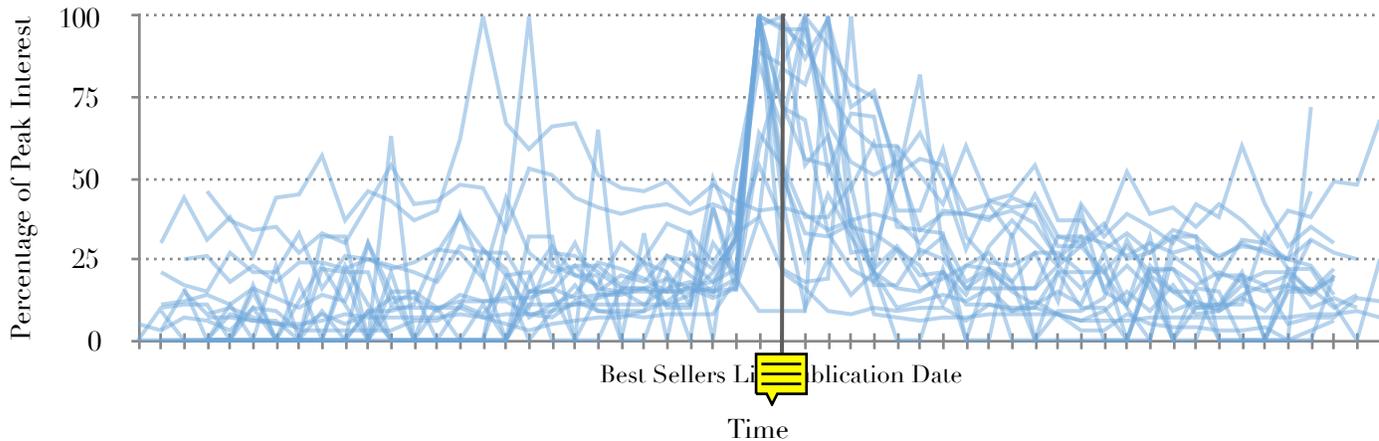
Average Interest Over Time
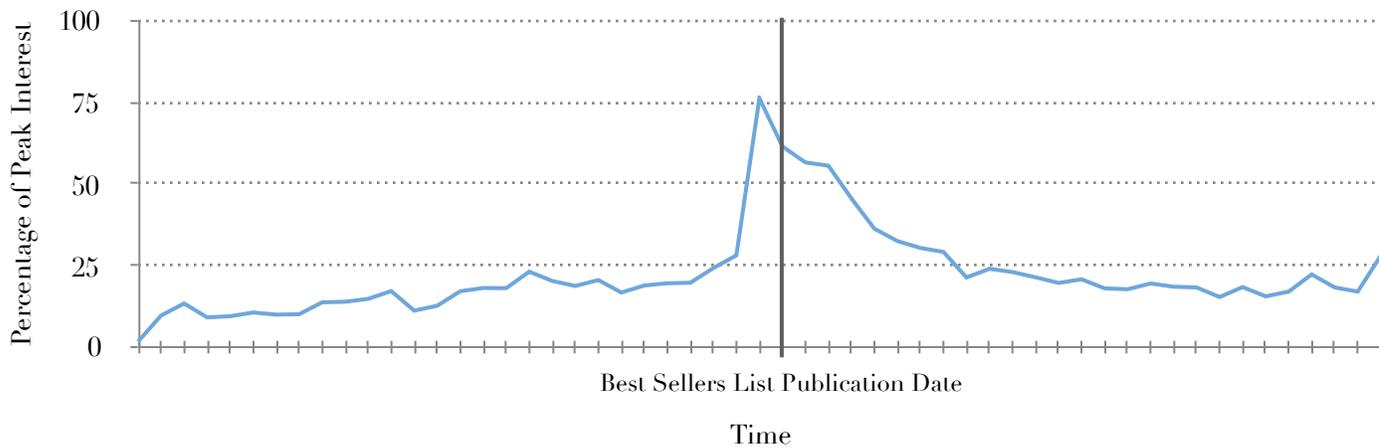
Combined

Looking deeper into the data using the charts on page 3, it is clear that the majority of the books analyzed reach their peak interest before *The New York Times* Best Sellers list data is finalized. Furthermore, a number of the books peak after the list is published online. However, as those books don't reach peak interest levels at a consistent distance from the list's publication date (as most books do before the list is published), it can't be assumed that these interest levels are directly caused by the books' appearance on *The New York Times* Best Sellers list.

The average of all the books' search interests show a clear peak in interest around a week before the list is published online, reaffirming the analysis made on page 2.

Reflection

Given the data that was collected, it is difficult to say to what extent a book's presence on *The New York Times* Best Sellers list affects its Google Search trends. A majority of the books' peak in interest occurs before the Best Sellers list's metrics are finalized, and those books whose peak in interest occurs after the list's publication do not happen any consistent length of time afterwards, leaving us unable to draw conclusions about the exact effects of a book's appearance on the Best Sellers list has on its Google Search trends. While it is clear that there is some correlation, it cannot be definitively said whether or not *The New York Times* Best Sellers list greatly affects Google Search trends, or if the peaks in interest come from an external source, and the result is the books being bought more and appearing on the Best Sellers list.

––––––––––––––––––

* Raw data is available here.