I started this project with a modest scope, but had very little idea how to execute on my vision with the lectures we had been given. My original plans for this project were to use the Bing Web Search API to pull two separate queries, the first getting the count for new news results each day for the year for a search of “Wagner PMC Africa,” and create a heatmap for the results. The second search would be for additional search queries, for example “United States + Russia World News,” and create a second heatmap, and potentially overlay these two maps on each other. My ultimate goal was to be to determine if events were occurring involving the Wagner PMC in Africa as a reaction to other world events by Russia’s rivals on the world stage. If there was a concentration of results in my Wagner PMC chart graphic shortly after my USA + R search, it would indicate that a potential causation was there and worth looking into.

What I had initially envisioned as being the most difficult part of this project actually ended being simpler than I had imagined. Microsoft and Azure have extremely helpful, descriptive documentation on how to set up a free Azure account, utilize the API key provided to you, and be able to call on it without issue. The only struggle I had here was finding where in my account my API key was housed. After I found the location and plugged it into the provided API call code, I was able to pull results for anything I wanted.

My next issue was where I struggled for days was trying to modify that initial search template Bing provided to get only the count of results, not the results themselves, or even the names of the results. The lecture slides on API calls had different motives, and much of the research I was able to find was for pulling data including headers, dates, authors, etc. After several days of searching, I decided to pivot, but I didn’t know where to yet. I decided to delve a little further into the functionality and capabilities of the Bing search API. While I was browsing one evening, I found a list of requests that can be used, as well as an in-depth look at how to set search parameters. In this documentation I found the “response” is what is used to get the count of results, and the “freshness” parameter is what I would use to get the date of the article. Using this info, I was finally able to put together the most important piece of my code:



With this, I was able to put together a table of results for each day in a calendar year, after setting the correct parameters. What I was missing however, was ensuring each day had only new articles. Without that portion of the code, it simply returned the amount of articles, adding new ones each day. But by adding a line as simple as:



It now provided me with only the new ones. Now as I had mentioned, I was originally intending on grabbing results for the full year. After testing my finally working code twice, I decided to try it a third time, this time creating the data frame with which to build my models. Right on schedule, I ran out of API calls with my free account. A short day of despair and failing to create new free accounts, I went back to the Azure documentation and found that I had $200 worth of free credit to use. After upgrading to the next tier, I decided for a less lofty goal of the previous month’s worth of data, the scale of my project being an unfortunate casualty of my initial ambition.

Lastly, as I mentioned, my next plan was to use the data gathered into these two data frames to create heatmaps for the each day of the month. Unfortunately, despite my best efforts I struggled to take the information carried in my data frame and transform it to a format that would be conducive to a proper daily heatmap graph. After trying a few different options, I settled on a scatterplot, which was still able to show my main premise, albeit in a less impressive format.

If I were to redo this project, I would try to utilize the documentation up front to its fullest extent. If I had spent more time initially looking into the user doc for the application I was working with, I have had a significantly larger portion of time to work on formatting the final version of my data, and perhaps the extra time would have allowed me to perfect the visual representation in a manner worthy of the source material and my vision for this dataset.