Exploring American Cultural Influence through Analysis of Google Trends Data

Alec Nipp[a]
[a]Humanities, NCSSM, Durham, NC

January 11, 2022

Abstract

In addition to its powerful economic and diplomatic ties, the US also wields considerable cultural influence over other nations. Online platforms and social media facilitate the dissemination of US values and interests to the rest of the world. This investigation aimed to quantitatively explore the extent to which US internet activity influences the online activity of other nations. Using the PyTrends Python package, a year of time-indexed search frequency data for 17 search terms across 13 countries was obtained from the Google Trends API. To measure similarity to US search trends, the average absolute difference between the country's time series data and the US's time series data was then computed and placed in a Pandas dataframe. The countries with the most similar search trends to the US were Australia, Great Britain, France, India, and Germany. Nations with the least similar trends included China, Russia, Egypt, and Vietnam. Thus, Western countries with strong cultural or economic ties generally had the highest trend parity, while Eastern countries tended to have weaker trend similarity. The search terms that tracked closest to the US trends were 'Election Results' and 'Black Panther'. The former was easily predictable. The latter, however, demonstrates the global, multicultural appeal of US entertainment — Hollywood, in particular. These promising results support the validity of the described method as a means for investigating the global cultural network via Google Trends.

Keywords: Google Trends; search data; cultural influence; Americanization; Python

Author roles: Alec Nipp devised the methodology, collected data, and performed the analysis for this project.

1 Context and motivation

To many users of the World Wide Web, it may seem like the United States occupies an outsize share of online attention. News platforms, social media sites, and other internet spaces with a global audience all devote large portions of their spaces toward US-related topics. It's intuitive that the US would take up a greater fraction of international news than most countries. As the largest economy in the world, the US is of great importance to investors all across the globe. Furthermore, the US has a population of over 300 million, a majority of which have internet access. Global websites must appear relevant to this massive demographic of American internet users.

Numbers aside, however, the US still holds a great deal of cultural influence over the rest of the world. Young people in other nations look to the US as a cultural lodestar. According to Goldfarb, "young people of the world, aged 18 to 29, still regard the US in an overwhelmingly positive light" (Goldfarb, 2018). Goldfarb believes immigration is a major reason for the prevalence of American culture. Large numbers of immigrants move to America from around the world. These immigrants then share the American experience with their friends and family back in their home countries. As a result, US culture spreads over a global network of immigrants and their families, aided by digital communication tools such as email and social media.

Hollywood is another axis of US cultural influence. Movies are one of America's biggest and most profitable cultural exports. Even in countries that tightly regulate the showing of American-made movies, Hollywood films consistently outperform their global competitors (Mirrlees, 2020). According to Mirrlees, to maintain global dominance, Hollywood has embraced the "global blockbuster, which studios design to sell to viewers from many different countries, not only those in the US." These movies tend to downplay national elements and themes in favor of global ones. Star Wars, for example, doesn't endorse any patently American values or imagery. Global blockbusters may avoid overt expressions of American culture, but they ultimately buttress American cultural and economic influence around the world (Mirrlees, 2020).

US companies like Google, Facebook, Twitter, and Netflix dominate the global tech market. The ascendancy of American business prompts a passive form of cultural imperialism. Kooijman writes, "American pop-cultural consumer products that are marketed as universal embodiments of American values like freedom and democracy are passively perceived and consumed as such" Kooijman (2014). A global consumer cannot buy American products without buying into American culture.

The global adoption of American culture and institutions is called Americanization. Although Americanization is apparent, little research has been done to quantify the American presence in global information spaces. In this investigation, the author aims to document the extent to which Americanization affects global Google search traffic. Similar studies have been conducted in the past, but none of these focused quantitatively assessing pop culture. Furthermore, most studies use the Google Trends website, whereas this investigation fetches data directly from the Google API Stephens-Davidowitz and Varian (2014). The study will compare global search traffic to US search traffic for various current events and pop culture items. With this method, the paper explores the strength of US cultural influence on a country-by-country basis.

2 Method

Since Google Trends allows queries for search data in specific countries, 13 nations not including the US were arbitrarily selected for analysis. All had large populations and developed or developing economies. At least one country was chosen from each continent, excluding Antarctica. The list of countries was saved in a CSV file.

Next, the search terms used to compare trends were chosen. Terms were chosen from the Google Trends "Year in Search" feature, which identifies the most frequently searched topics of a specific year(*Google Trends*, 2006). Terms that had specific relevance to the US were generally avoided. Searches for 'Joe Biden', for example, would obviously spike globally when they spike in the US. There's not much baseline traffic for 'Joe Biden' in other countries that doesn't directly pertain to the US President. Instead, greater emphasis was placed on topics that weren't inherently American. For example, one of the terms was 'hurricane'. This word is generic and likely has a lot of background search activity. If there are spikes in searches for 'hurricane', it could correlate with Hurricane Dorian or some other hurricane that affected the US. Searches for "hurricane" could also spike independent of events occurring in the US. By examining topics adjacent to the US and not explicitly relevant to it, the effect that US interests have on global search traffic could be accurately assessed. When completed, the list of search topics was saved in a CSV file.

For analysis, the files containing the country list and the topics list were read into a Python environment using the Pandas package. For each topic, the search trends for every country were queried and stored as a Pandas time series. A python package called PyTrends was used to access the Google API and send data requests. PyTrends returns a time-indexed table of the relative search frequency for the key word (*pytrends*, n.d.). Once the data for a certain country and topic combination were obtained, the time series was compared to the US trends for the same topic.

The average difference between the search intensities at each time stamp was computed and stored in a dataframe. That is, for a particular topic and country, the absolute value of the difference in search frequency between that country and the US for each entry was summed and then divided by the number of entries. Since Google Trends data is normalized relative to the peak intensity of the search query, the difference in population between countries was accounted for.

Since Google Trends places a limit on the volume of requests a particular client can make, the data requests would occasionally be rejected. Error handling was implemented in the code to preclude a fatal crash when the server rejects a request. When the server rebuffs a query, the program waits 15 seconds, then tries again. This works until the client exceeds its maximum daily request quota, at which point the Google API rejects all incoming requests for some unknown time period. When the author encountered this issue, he was able to resume data collection the next day. PyTrends does support the use of proxy servers to circumvent the Google API's limits, but the author had no access to such resources.

All of the average differences in search activity were stored in a Pandas dataframe where each row was a search term and each column was a country. This dataframe was then analyzed using Pandas and other tools included in the Python Anaconda distribution. MatPlotLib was used to create graphical representations of the data.

3 Results and discussion

A table of the closest matches to US trends is shown below.

Search Terms	Min Diff	Country
Election Results	0.078431	MX
Coronavirus	8.901961	CN
Stimulus	7.235294	DE
Unemployment	14.960784	GB
Iran	12.137255	MX
Elon Musk	6.901961	DE
Simp	7.058824	GB
Furlough	7.509804	IN
Among Us	2.666667	AU
Parasite	7.019608	IN
Black Panther	0.941176	AU
Tiger King	5.627451	AU
Ozark	7.372549	IN
Hurricane	8.039216	MX
Joker	6.764706	MX
Boomer	11.686275	IN
Brexit	5.470588	IN
Avg	10.765859	AU

Table 1: Table of closest matches to US search trends for each term. Country with minimum average difference is shown.

Australia, Mexico, and India all have the closest match to the US for several of the terms. Australia had the greatest overall similarity to the US with an average difference of 10.77 across all search terms. The highest trend parity of all country and search combinations was came from Mexico's search trends for 'Election Results'.

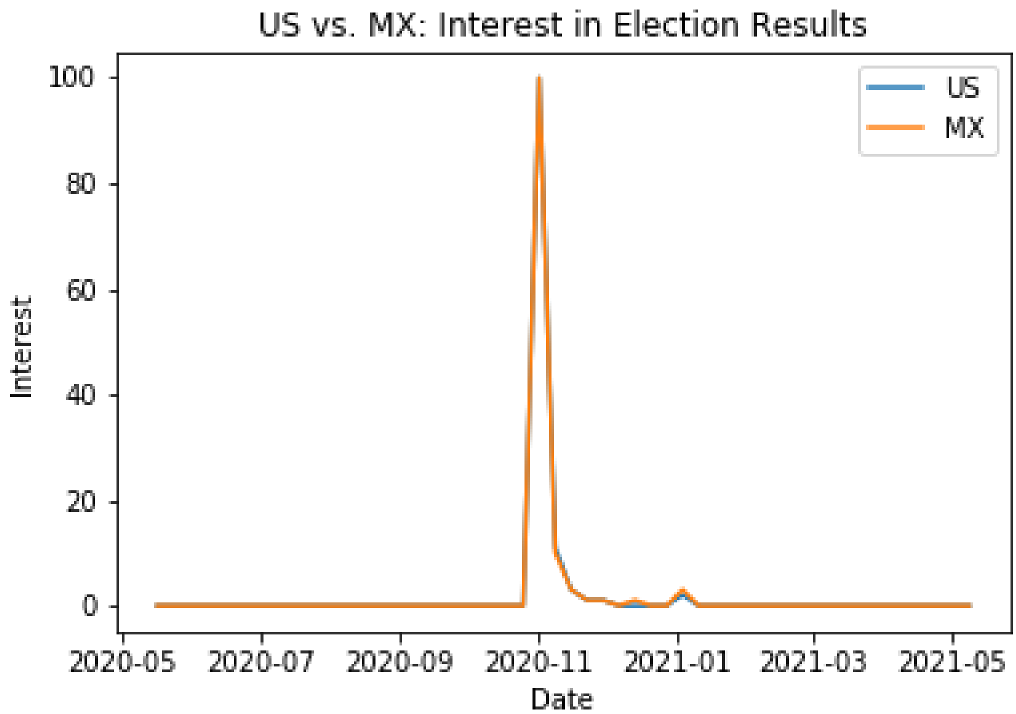

Figure 1: Search trends for US and Mexico plotted against one another.

As shown in Figure 1, Mexico's search trends almost perfectly match the trends of the US. Since that comparison yielded an average difference of 0.078, we would expect a plot like that. Below is a more typical trend comparison.

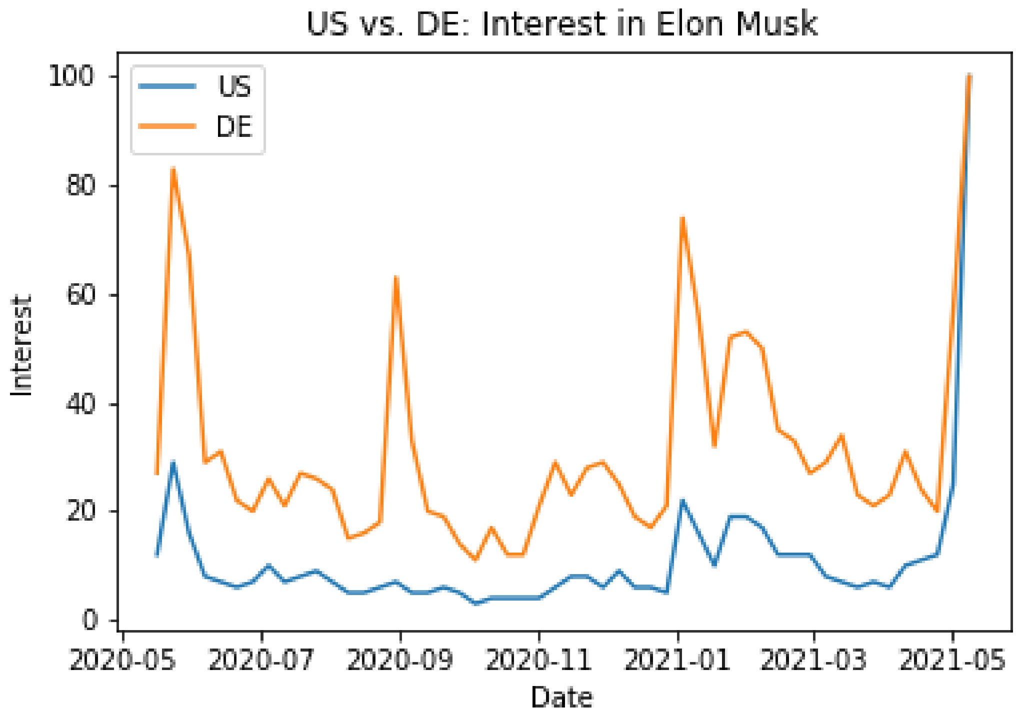

Figure 2: Searches for 'Elon Musk' in the US and Germany plotted against one another. The average difference is 6.90.

Figure 2 shows the average matching trend. The lines have the same shape, but the search intensity in Germany is much more evenly distributed.

Search Terms	Max Diff	Country
Election Results	1.333333	IN
Coronavirus	22.882353	DE
Stimulus	23.333333	VN
Unemployment	48.313725	EG
Iran	40.274510	BR
Elon Musk	20.450980	CN
Simp	47.901961	CN
Furlough	27.372549	TR
Among Us	32.294118	JP
Parasite	50.333333	CN
Black Panther	7.980392	RU
Tiger King	30.333333	VN
Ozark	43.901961	CN
Hurricane	50.294118	VN
Joker	34.960784	CN
Boomer	64.450980	CN
Brexit	24.529412	CN
Avg	28.159170	CN

Table 2: Table of least optimal fits to US search trends for each term. Country with maximum average difference is shown.

For eight of the search terms, China had the least similarity to the US. As a result, China also had the highest average difference from the US across all search terms. Below is an example of US and China search trends plotted against one another.

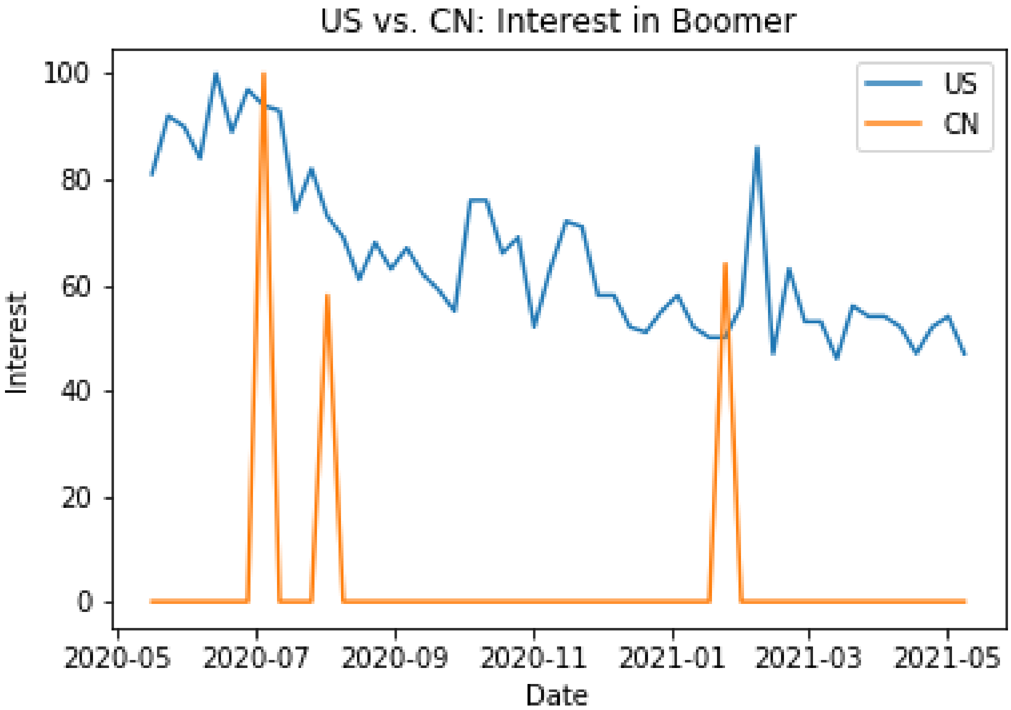

Figure 3: Searches for ‘Boomer’ in US and China. There is no discernible correlation.

Figure 3 shows how the trends for ’Boomer’ in the US and China are starkly different. With an average difference of 64.45 between for this entry, we would expect the plot to look like this.

Search Term	Average Diff
Election Results	0.343891
Black Panther	3.351433
Among Us	10.695324
Elon Musk	11.977376
Brexit	12.197587
Coronavirus	12.886878
Joker	14.312217
Stimulus	15.466063
Avg	16.910389
Ozark	17.641026
Furlough	19.390649
Tiger King	19.476621
Iran	20.852187
Hurricane	23.778281
Parasite	23.889894
Simp	24.352941
Boomer	28.060332
Unemployment	28.803922

Table 3: Table of differences for each search term, averaged across all countries, sorted by magnitude. The average difference for all terms and countries is 16.91.

In Table 3, we see that the US and the world have the most similar trends for ‘Election Results’, ‘Black Panther’, ‘Among Us’, and ‘Elon Musk’. Interest in Election Results is an obvious outlier, with an average difference nearly ten times smaller than the next smallest entry.

Search Term	Average Diff
AU	10.765859
GB	11.678201
FR	12.905421
IN	13.715110
DE	13.940023
MX	14.725490
BR	15.816609
Avg	16.910389
TR	17.671280
JP	19.552480
VN	19.963091
RU	20.377163
EG	20.565167
CN	28.159170

Table 4: Table of average differences across all terms for each country, sorted by magnitude.

According to Table 3, Australia, Great Britain, France, India, and Germany trends have the greatest similarity to US trends. China, Egypt, Russia, Vietnam, and Japan all have the least similarity. Interestingly, however, China has the highest search parity for one term: 'Coronavirus'. The plot for this term is shown below.

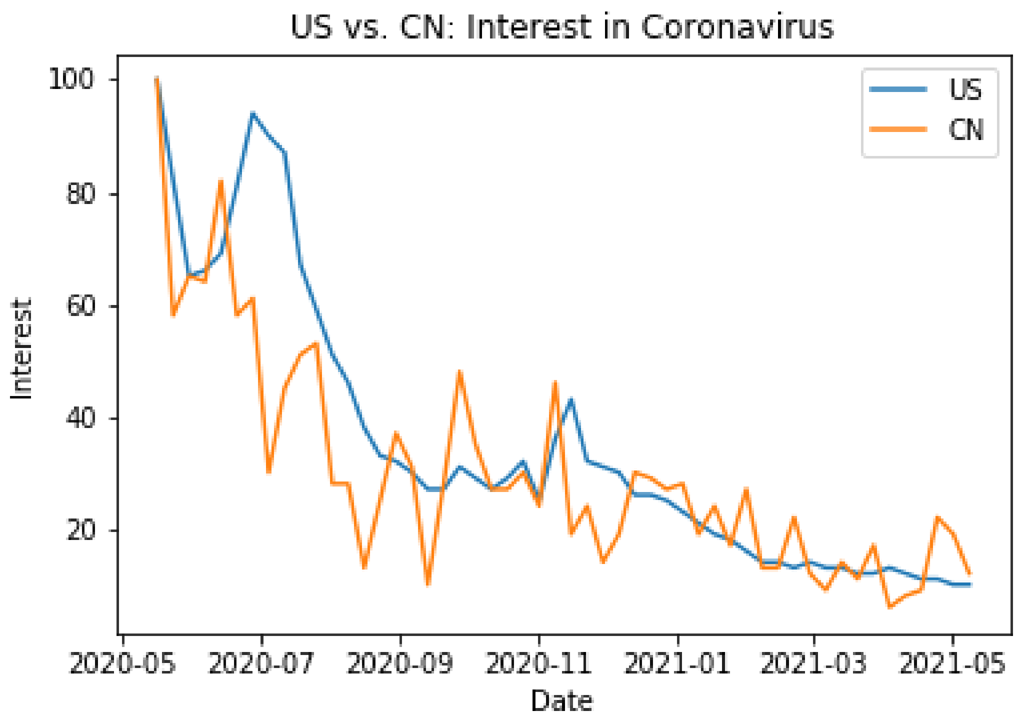

Figure 4: Searches for 'Coronavirus' in US and China.

As we can see in Figure 4, the US trends track with the China trends remarkably well.

4 Implications/Applications

This investigation was primarily exploratory. In order to draw any definitive conclusions about the correlations, a more thorough analysis with better controls would be required. Nevertheless, there were several noteworthy observations.

4.1 Observations

Western countries tended to be more similar to the US. Australia, Great Britain, and France had the three lowest average differences. At the fifth spot, Germany was not far behind.

All of these countries speak English or have a large English-speaking population. If residents of these nations can communicate in English over the internet, they're more likely to interact with US citizens, discuss US-related topics, and pick up US slang such as 'Simp'. Furthermore, Western countries have a shared cultural ancestry. They're more likely to have similar values and interests to the US, which may lead to a stronger pop-cultural link.

Nations with an average level of US-similarity included Mexico, Brazil, and Turkey. These all have average differences between 14 and 18, slightly higher than the 10-14 range of the Western countries. Once again, the cultural similarity to the US seems to track how closely each country's search data resembles that of the US. Latin America and the Middle East are culturally distinct from the US, and English isn't the first language in either region. Despite these barriers, the US has considerable diplomatic and economic ties to Mexico, Brazil, and Turkey. Financial and political relationships with the US likely catalyzed Americanization in these countries. Moderate levels of Americanization could explain the appreciable similarity in search trends.

The countries with the weakest trend similarity were Japan, Vietnam, Russia, Egypt, and China. Japan had the lowest average difference with 19.55; China had the highest with 28.16. It's unsurprising that China, Egypt, Vietnam, and Russia would be so different from the US. All have had limited interaction with the West for geographic, social, or political reasons. China, Russia, and Vietnam were all opposed to the US during the Cold War. Weak or hostile relationships with the US likely impeded cultural exchange, reducing US soft power in those regions.

Japan is the most US-similar of the Eastern countries. The US and Japan share close diplomatic and economic ties, which facilitate cultural exchange and Americanization. However, Japan's cultural heritage is starkly different from that of the US. That, in combination with the geographic distance between the countries, could account for the large disparities in search trend patterns.

For the search terms themselves, there was a wide range of average differences. The closest-matching trend on average was searches for 'Election Results', followed by 'Black Panther', 'Among Us', and 'Elon Musk'. The US election was the largest election to take place in the last year, so it follows that global interest in elections would be dictated by the progress of US elections. The movie Black Panther also had a notably low average difference of 3.35. This supports the view that Hollywood dominates the global movie industry. US television programs like Tiger King and Ozark had far less search parity than movies like Black Panther and Joker, indicating that Hollywood's cultural influence is much stronger.

The close parallels in search data suggest Black Panther is a topic of international discussion. Citizens of every country may congregate on internet forums to talk about Hollywood movies like Black Panther, which often are released both domestically and globally. Future studies should focus specifically on pop culture icons like Black Panther to further document the wide reach of Hollywood and US entertainment.

For reasons similar to Black Panther, the popular video game Among Us also had high search parity. Among Us exploded in popularity on internet platforms such as Twitch and YouTube. On these platforms, internet celebrities have international audiences that number in the millions. Video games like Among Us are broadcasted to viewers in the US and around the world. The cultural exchange mediated by video games and streaming platforms is another phenomenon for future research.

4.2 Method

The procedures used in this investigation were effective and easily reproducible. Several improvements can be made, however.

For projects that require a large number of queries to the Google API, researchers should consider the use of proxy servers. Routing the API requests through proxies would allow data collection unrestricted by the Google API's limits.

More stringent controls will also be necessary to draw sound conclusions. Some of the data similarity could arise from random chance. To establish a null mean, researchers could generate a test set of random trend graphs, then take the mean of the average difference between the random graphs and the actual data. This could serve as a reference for the amount of correlation that arises randomly. Statistical tests may then be used to verify the existence of significant correlation.

Acknowledgements

Appreciation is extended to the NCSSM Interdisciplinary Course Group, the Department of Humanities (Elizabeth Moose, Dean), and the Department of Science (Dr. Amy Sheck, Dean). Additional thanks are given to the instructors of the NCSSM Digital Humanities course (Michelle Brenner and Robert Gotwals).

Notes

References

Goldfarb, M. (2018, Jul). *Soft power is the secret to america's influence - from culture to foreign aid.* Retrieved from http://www.raconteur.net/american-culture-rules-the-world/.

Google trends. (2006, May). https://trends.google.com/trends/. ([Online; accessed 10-May-2021])

Kooijman, J. (2014). 1. we are the world. america's dominance in global pop culture. In *Fabricating the absolute fake-revised edition* (pp. 23–42). Amsterdam University Press.

Mirrlees, T. (2020, Jun). *Global hollywood: An entertainment imperium, by integration.* Retrieved from https://cineaction.ca/issue-99/global-hollywood-an-entertainment-imperium-by-integration/

pytrends. (n.d.). Retrieved from https://pypi.org/project/pytrends/

Stephens-Davidowitz, S., & Varian, H. (2014). A hands-on guide to google data. *further details on the construction can be found on the Google Trends page*.

Supplementary Information

Below is the code used to fetch and organize data from the Google API.

```
import pandas as pd
import numpy as np
import pytrends
from pytrends.request import TrendReq
import matplotlib.pyplot as plt
import warnings
import time as timer
warnings.filterwarnings('ignore')

# Establish connection to API
pytrends = TrendReq(hl='en-US', tz=360)

# Import search terms and country codes for queries

search_terms = pd.read_csv("search_terms.csv")
country_codes = pd.read_csv("country_code_data.csv")
print(search_terms)
print(country_codes)

countries = list(country_codes["country_code"])
terms = list(search_terms["search_terms"])
print(countries)
print(terms)

def total_diffs(list1, list2):
    """
    Sum of absolute value of differences
    """
    if list2 == 0 or list1 == 0:
        return None
    total = 0
    for i in range(len(list1)):
        total += abs(list1[i]-list2[i])
    return total

def avg_total_diffs(list1, list2):
    """
    Average of abs value differences between lists
    """
    return total_diffs(list1, list2) / len(list1)


def retrieve_data(country, term, time="today 12-m"):
    """
    Retrieves the Google search data on the term starting from time
    """
    try:
        pytrends.build_payload([term], cat=0, timeframe = time, geo=country, gprop='')
        df = pytrends.interest_over_time()
    except KeyError: # No data available for this term
        df = 0
    except: # Have maxed out request frequency, so wait to try again
        print("Received response error or some other error. Waiting to try again.")
        timer.sleep(15)
        return(retrieve_data(country, term, time))
    return df
def get_activity_list(country, term, time="today 12-m"):
    """
    Returns the search trend for a country/term as a list
    """
```

```
    df = retrieve_data(country, term, time)
    return list(df.iloc[:,0])

def wait_for(seconds):
    print(f"Now waiting for {seconds} seconds")
    timer.sleep(seconds)

# Fetch data

diff_df = []
for term in terms:
    diffs = []
    US_DATA = get_activity_list('US', term)
    for country in countries[1:]:
        print(f"Getting {term} for {country}")
        this_data = get_activity_list(country, term)
        diffs.append(avg_total_diffs(US_DATA, this_data))
    diff_df.append(diffs)

print(diff_df)

# Convert to Pandas DataFrame

processed_df = pd.DataFrame(diff_df, columns=countries[1:])
processed_df.insert(0, "Search Term", terms)
print(processed_df)

# Save dataframe as CSV
processed_df.to_csv('/Users/alecnipp/Documents/NCSSM_senior/sem2/dig_hum/Python/search_results_terms_full.csv')
```