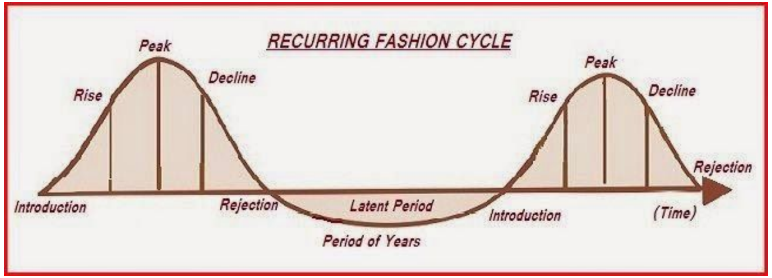
# Fashion Trends Geographically Mapped

Bea Giacalone, Lizzy Qian, Anjana Begur

# Context/Background

- We know fashion trends are cyclical[1] and ebb and flow through countries.
- The fashion industry as a whole is a large part of the global economy[2], and insights into how it moves could be critical for making smart business decisions.




Source

1 Business of Fashion 2 Joint Economic Committee

# Question

How do fashion trends move geographically on a global scale?

# Abstract

**GAP**: Fashion trends move and are cyclical over time. However, the same collections are launched at different times in different countries. We should then be able to find a general pattern of how trends move in terms of geography rather than only time.

**HOW**: Here we will use Google Trends and the Pinterest API to study how trends for different items moved over time and we will study patterns. We add timestamps to different trends to find the difference in peaks of interest over geography.

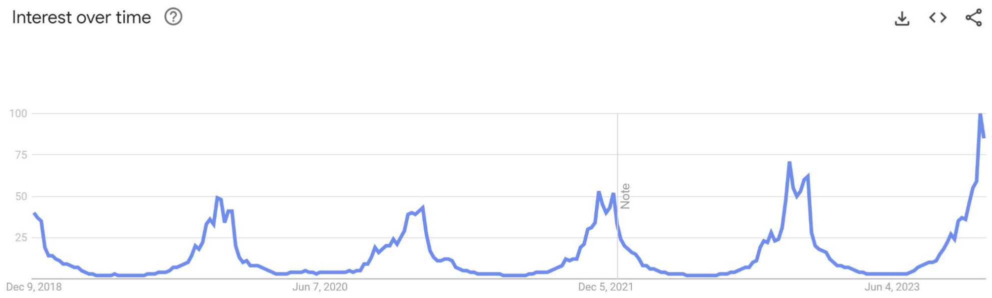
**Results:** We found United States had the least delay of the five countries investigated, leading us to conclude that cyclic trends start in the US. Australia and European countries, have increasingly greater delays. Consumers and fashion stakeholders should look to the US and Australia to observe the beginning of trend cycles, to and Italy and Europe to follow these trends.

**Audience**: Fashion brands and magazine readers

# Data: Why Google Trends?

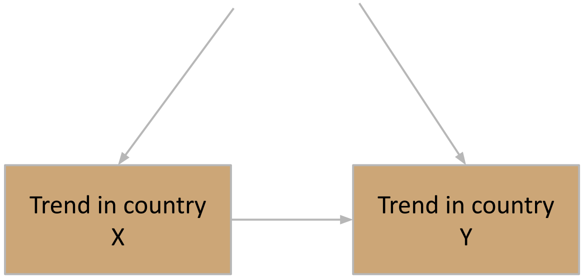
- Wealth of data
- Reliably consistent across countries
- General accessibility to public
    - Avoids focusing only on certain demographics that have access to

Trends Page for Uggs in the USA



# Limitations of Data: Confounders

1. Weather
2. Manufacturing country
3. Cultural Attire Norms




How we approached them?

1. Choosing items that have shown a consistent interest over time
2. Differentiating the countries in a way that the place where the item was produced is differentiated

# Methodology: Data Collection

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Country | Fashion Bra | Week | Interest |
| 1302 | United States | Montcler | 12/9/18 | 69 |
| 1303 | United States | Montcler | 12/16/18 | 57 |
| 1304 | United States | Montcler | 12/23/18 | 61 |
| 1305 | United States | Montcler | 12/30/18 | 50 |
| 1306 | United States | Montcler | 1/6/19 | 48 |
| 1307 | United States | Montcler | 1/13/19 | 63 |
| 1308 | United States | Montcler | 1/20/19 | 47 |
| 1309 | United States | Montcler | 1/27/19 | 46 |
| 1310 | United States | Montcler | 2/3/19 | 38 |
| 1311 | United States | Montcler | 2/10/19 | 55 |
| 1312 | United States | Montcler | 2/17/19 | 43 |
| 1313 | United States | Montcler | 2/24/19 | 30 |
| 1314 | United States | Montcler | 3/3/19 | 28 |
| 1315 | United States | Montcler | 3/10/19 | 23 |
| 1316 | United States | Montcler | 3/17/19 | 19 |
| 1317 | United States | Montcler | 3/24/19 | 17 |

- We could not scrape google trends, so we collected data manually and inputted it into a CSV
    - Manually downloaded 200 separate datasets and pieced into 1 large dataset
- Manually collected data on 40 fashion brands in 5 different countries
    - Decided US, UK, Italy, Australia, and France were largest epicenters of modern Western fashion industry

# Methodology: Data Cleaning

- Used pandas library to create dataframes to store data
- Cleaned dataframes
- Cleaned missing values
- Changed values denoted < 1 to be 0
- Added title to interests column (manual)

```
import pandas as pd
from datetime import datetime, timedelta
from itertools import combinations
import matplotlib.pyplot as plt
import numpy as np
import matplotlib.pyplot as plt
from scipy.signal import correlate, find_peaks
```

```
df = pd.read_csv('fashion_master_list.csv')
df = df[['Country', 'Fashion Brand', 'Week','Interest']]
countries = df['Country'].to_numpy()
countries = np.unique(countries).tolist()
brands = df['Fashion Brand'].to_numpy()
brands = np.unique(brands).tolist()
#item = df[(df['Fashion Brand']=='Adidas') & (df['Country']=='Australia')]

items = []
for b in brands:
    item = []
    for c in countries:
        item.append(df[(df['Fashion Brand']==b) & (df['Country']==c)])
    items.append(item)
```

# Methodology: Data Smoothing

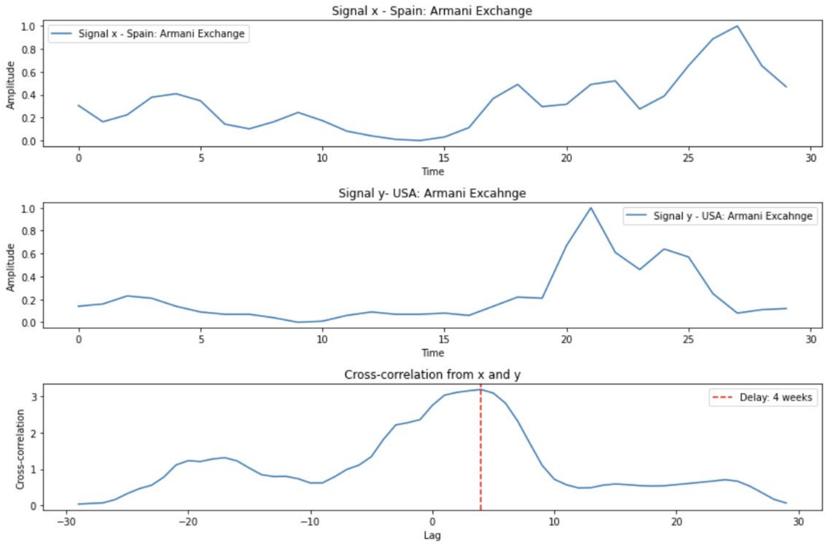
- We used a moving average to smooth the data, with a window size of 2

```
# --------------------------------------------------------------------------------
#                            1) Smooth the data
# --------------------------------------------------------------------------------
def smooth_data(df, window_size=2):
    df = df.copy()

    # b) Convert 'Week' column to datetime
    df['Week'] = pd.to_datetime(df['Week'])

    # c) Apply a moving average to smooth the data ---> larger the window size, the smoother the
    df['Smoothed Interest'] = df['Interest'].rolling(window=window_size, min_periods=1).mean()

    return df
```

# Methodology: find delay

Example with one of the Items: Armani Exchange

```
def find_delay(a, b):
    x = moving_average(a,1)
    y = moving_average(b,1)

    if np.max(x) > np.min(x):
        x = (x - np.min(x)) / (np.max(x) - np.min(x))

    if np.max(y) > np.min(y):
        y = (y - np.min(y)) / (np.max(y) - np.min(y))

    # Compute cross-correlation
    corr = correlate(x, y, mode='full')

    # Find the index of the peak in the cross-correlation
    peak_idx = np.argmax(corr)

    # Compute the delay
    delay = peak_idx - (len(x) - 1)
```



# Methodology: Find delay matrices

```python
def create_delay_matrix(input_list):
    """
    Creates an n x n matrix of delays.
    :param input_list: A list of floats.
    :return: A numpy matrix of size n x n, where each entry is the result of find_delay(a, b).
    """
    n = len(input_list)
    delay_matrix = np.zeros((n, n))  # Initialize an n x n matrix

    for i in range(n):
        for j in range(n):
            delay_matrix[i, j] = find_delay(input_list[i], input_list[j])

    return delay_matrix
```

```python
delay_matrices = []
for brand in items:
    brand_list = []
    for country in brand:
        smoothed_item = smooth_data(country)
        country_interest_array = smoothed_item['Smoothed Interest'].to_numpy()[30:-200]
        brand_list.append(country_interest_array)

    matrix = create_delay_matrix(brand_list)
    delay_matrices.append(matrix)
```

```python
delay_matrices
```

```
[array([[ 0.,  0.,  0.,  0.,  0.],
        [ 0.,  0.,  0.,  1.,  0.],
        [ 0.,  0.,  0.,  1.,  0.],
        [ 0., -1., -1.,  0., -1.],
        [ 0.,  0.,  0.,  1.,  0.]]),
 array([[ 0.,  0., -1.,  0.,  0.],
        [ 0.,  0., -2.,  0.,  0.],
        [ 1.,  2.,  0.,  1.,  1.],
        [ 0.,  0., -1.,  0.,  0.],
        [ 0.,  0., -1.,  0.,  0.]]),
 array([[ 0., -1., -4.,  0.,  0.],
        [ 1.,  0.,  0.,  0.,  0.],
        [ 4.,  0.,  0.,  0.,  4.],
        [ 0.,  0.,  0.,  0.,  0.],
        [ 0.,  0., -4.,  0.,  0.]]),
 array([[ 0.,  0.,  0.,  0.,  0.],
        [ 0.,  0.,  0.,  0.,  0.],
        [ 0.,  0.,  0.,  1.,  0.],
        [ 0.,  0., -1.,  0.,  0.],
        [ 0.,  0.,  0.,  0.,  0.]]),
 array([[  0., -18., -19., -18., -17.],
        [ 18.,   0.,  -1.,   0.,   0.],
        [ 19.,   1.,   0.,   2.,   2.],
        [ 18.,   0.,  -2.,   0.,   0.],
        [ 17.,   0.,  -2.,   0.,   0.]]),
```

# Methodology: Find Average Delay Matrix

- We used a delay matrix to map out how late a country was to a trend.
- Our heatmap represents the average delay each of the 5 countries have

```python
# AVERAGE MATRIX
stacked_matrices = np.stack(delay_matrices, axis=0)
    # Compute the mean along the new dimension
mean_matrix = np.mean(stacked_matrices, axis=0)
print('AVERAGE DELAY MATRIX')
print(np.unique(countries).tolist())
print(mean_matrix)
```

```
AVERAGE DELAY MATRIX
['Australia', 'France', 'Italy', 'United Kingdom', 'United States']
[[ 0.         -3.45454545 -4.27272727 -3.09090909 -0.90909091]
 [ 3.45454545  0.         -0.72727273  0.18181818  0.36363636]
 [ 4.27272727  0.72727273  0.          1.          2.81818182]
 [ 3.09090909 -0.18181818 -1.          0.          2.09090909]
 [ 0.90909091 -0.36363636 -2.81818182 -2.09090909  0.        ]]
```

Delay France -> Australia = 3.45

Delay Australia -> France = -3.45

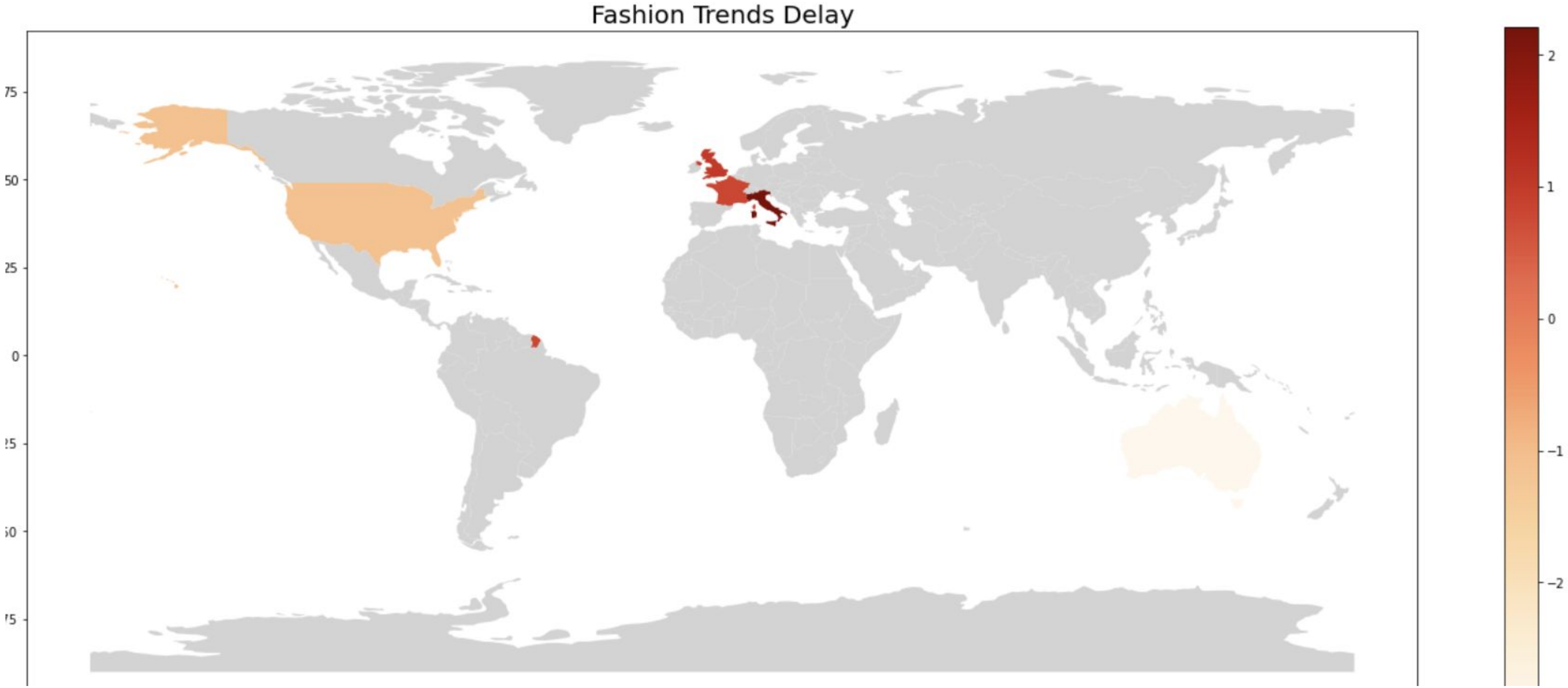
# Methodology: Data Visualization

- Creating a heatmap - good for showing relative relationships
- Matplotlib and geopandas were the main libraries we used
- Overlaid on a map for a more easily interpretable result

```python
australia = sum([-3.45454545, -4.27272727, -3.09090909, -0.90909091]) / 4
france = sum([3.45454545, -0.72727273, 0.18181818, 0.36363636]) / 4
italy = sum([4.27272727, 0.72727273, 1., 2.81818182]) / 4
UK = sum([3.09090909, -0.18181818, -1., 2.09090909]) / 4
USA = sum([0.90909091, -0.36363636, -2.81818182, -2.09090909]) / 4
```

```python
# ---------------------------------------------------------------------------
#                         6) Add heat map
# ---------------------------------------------------------------------------
import geopandas as gpd
# Load world map
world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))

# Assuming you have a dictionary like {'CountryName': average_delay, ...}
delays = {'Australia': australia, 'France': france, 'Italy': italy, 'United Kingdom': UK, 'United States of America': USA}

# Add a delay column to the world dataframe
world['delay'] = world['name'].map(delays)

# Plotting
fig, ax = plt.subplots(1, 1, figsize=(25,10))
world.plot(column='delay', ax=ax, legend=True, cmap='OrRd', missing_kwds={'color': 'lightgrey'})

plt.title('Fashion Trends Delay',fontsize=20)
plt.show()
```

# Final Result:




Here, darker colors represent more delay

# Conclusions + Results

- We found that the ranking of the countries from least to most delayed was:
  1. Australia
  2. US
  3. France
  4. UK
  5. Italy
- Most amount of delay in UK and Italy shows they tend to be influenced by the trends of geographically larger western countries such as Australia and the US
- American and Australian consumers are more likely to buy rapidly and lead trends

# What does this mean for brands?

- A stakeholder in the fashion market now knows that the US leads worldwide fashion trends cycles
- Look to the Australia first and the US second for upcoming fashion trends
- Fashion brands should deemphasize marketing in Italy and European countries, and emphasize marketing in the US and Australia

# Issues We Faced

- Obtaining Data: Google does not want people to scrape their resources, and are very good at detecting when a computer is sending in too many requests. We could not find an API to scrape Google Trends

- Compilation issues - since a lot of the code was written separately, debugging and making sure everything was compatible was also a struggle
- Finding the best way to represent data - we were hoping to use arrows and show more movement, but within the scope of our project the best way to summarize things was a heatmap

# Further Improvements

- If there were a better API to scrape google trends, more data would be beneficial to have a more comprehensive global outlook
- A possible expansion of this project would be to implement some sort of predicting algorithm to show where trends are headed
- To include more detail on how exactly the trends move, an interactive dashboard where users can map trends from country to country would be ideal

# Thank You!

Questions?