Renee Fischer, Max Ercolani, Jackson Howard

FP04 - Alumni Map Data Analysis

Overview

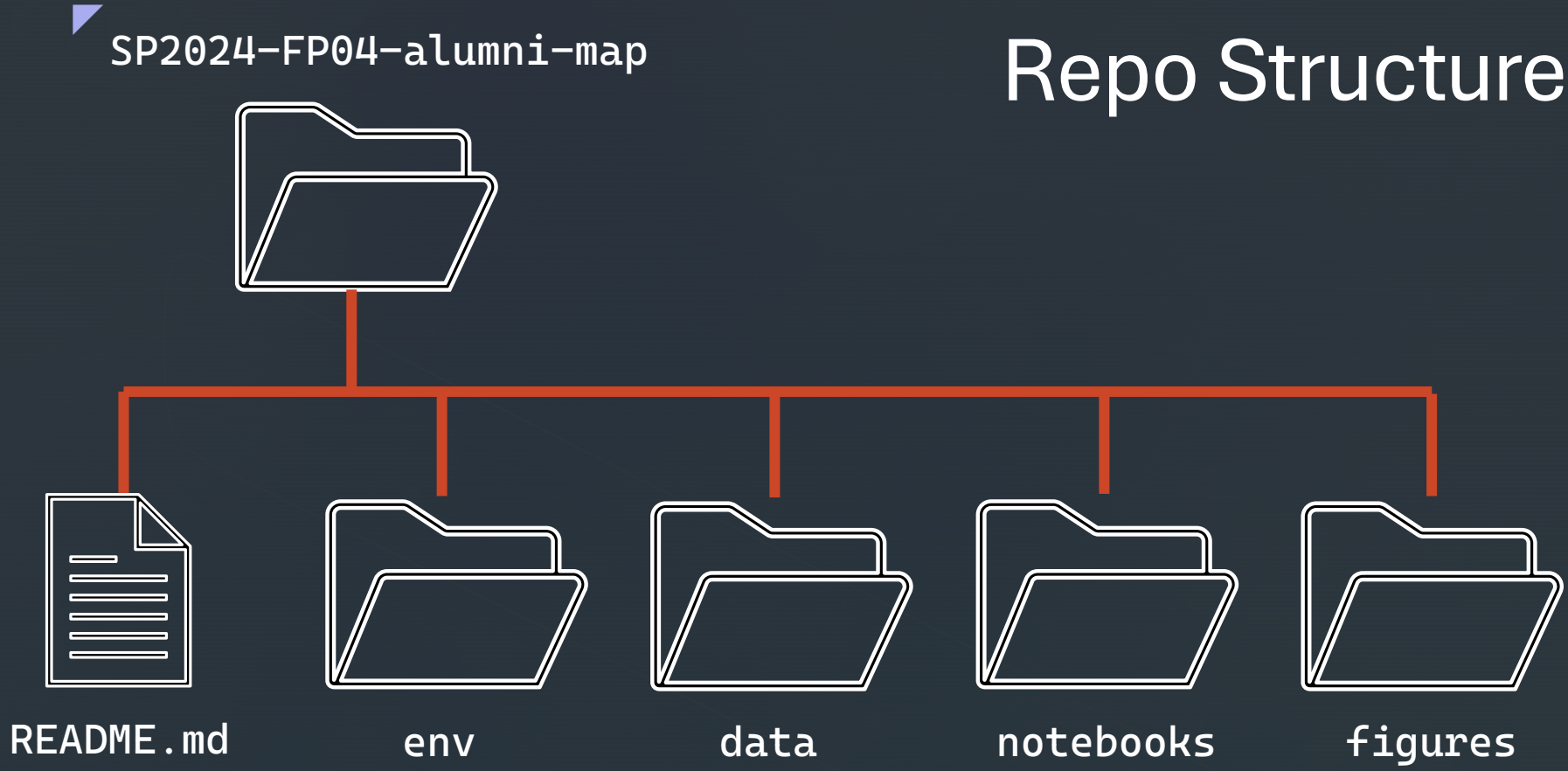
- **Objective:** Create a comprehensive map of the current residence of Mines Geophysics alumni.
- **Research Question:** Are there patterns in where alumni take residence following graduation from the Mines Department of Geophysics?

Data Acquisition

Alumni Mail List from GP Dept.

Anonymous Data

Repo Structure



```
# import modules
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import requests
```

Data Cleaning

- Separate Degrees
- Add Country entry to States
- Removed Canadians
- Added Latitude and Longitudes

	Affiliation	PrefClassYear	Degrees	State	Country	Latitude	Longitude
0	Alumni	1983.0	BSc	CO	United States	39.783730	-100.445882
1	Alumni	1995.0	BSc	CO	United States	39.783730	-100.445882
2	Alumni	2013.0	MSc	NaN	Bahrain	26.155125	50.534461
3	Alumni	1999.0	MSc	NaN	United Kingdom	54.702354	-3.276575
4	Alumni	1982.0	BSc	OK	United States	39.783730	-100.445882
...	...	...	...	...	...	...	...
2225	Alumni	2012.0	BSc	CO	United States	39.783730	-100.445882
2226	Alumni	1982.0	BSc	TX	United States	39.783730	-100.445882
2227	Alumni	1981.0	BSc	UT	United States	39.783730	-100.445882
2228	Alumni	2019.0	BSc	CO	United States	39.783730	-100.445882
2229	Alumni	2012.0	MSc	CO	United States	39.783730	-100.445882

Clean Degrees, States, and Countries

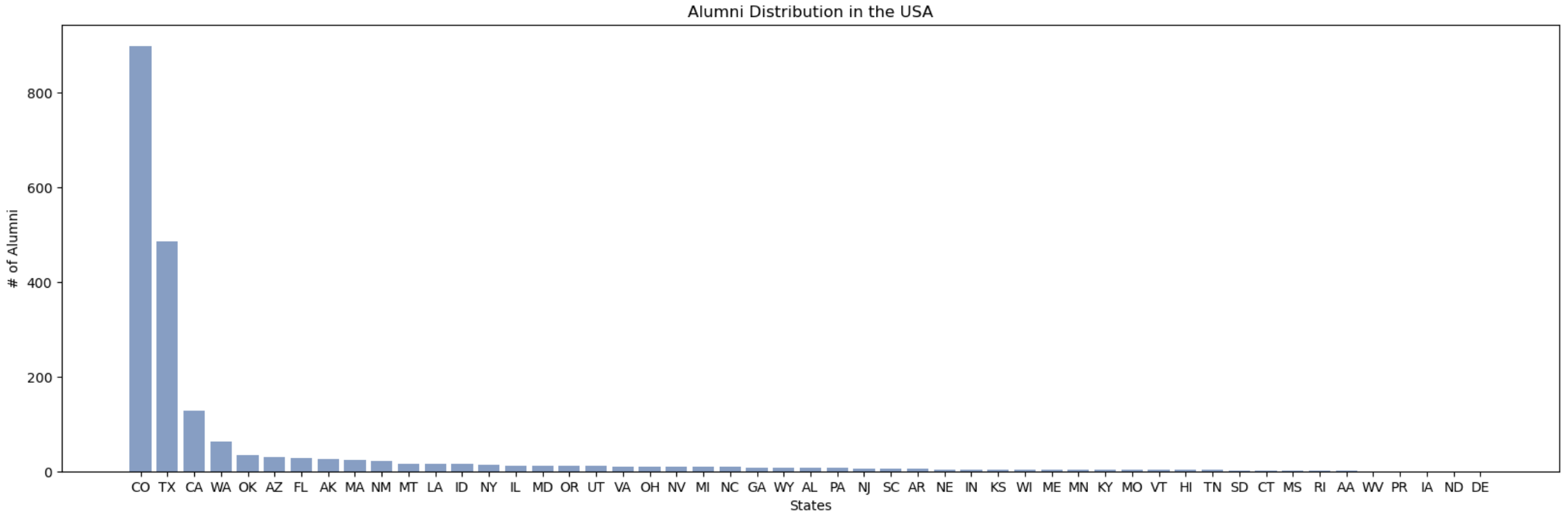
```
# clean the dataframe

#rename StateOrProvince to State
df.rename(columns={'StateOrProvince': 'State'}, inplace=True)

#delete rows if Affiliation is not alumni
df = df[df['Affiliation'] == 'Alumni']

#clean Degree column to include only degree level
validDegrees = ['BSc','MSc','PhD']
df['Degrees'] = df['Degrees'].fillna('').apply(lambda x: [deg.strip() for deg in x.split() if deg.strip() in validDegrees])

#split people with multiple degrees into separate rows
df = df.explode('Degrees')

#replace NaN in country column with United States or Canada based on StateorProvince
usStates = ['AL', 'AK', 'AZ', 'AR', 'CA', 'CO', 'CT', 'DE',
                                'FL', 'GA', 'HI', 'ID', 'IL', 'IN', 'IA', 'KS',
                                'KY', 'LA', 'ME', 'MD', 'MA', 'MI', 'MN', 'MS',
                                'MO', 'MT', 'NE', 'NV', 'NH', 'NJ', 'NM', 'NY',
                                'NC', 'ND', 'OH', 'OK', 'OR', 'PA', 'RI', 'SC',
                                'SD', 'TN', 'TX', 'UT', 'VT', 'VA', 'WA', 'WV',
                                'WI', 'WY','PR']
df['Country'] = df.apply(lambda row: 'United States' if row['State'] in usStates else row['Country'], axis=1)

#remove province from Canada
df['State'] = np.where(df['Country'] == 'Canada', '', df['State'])

df.reset_index(drop=True, inplace=True)
df
```

Add Latitudes and Longitudes

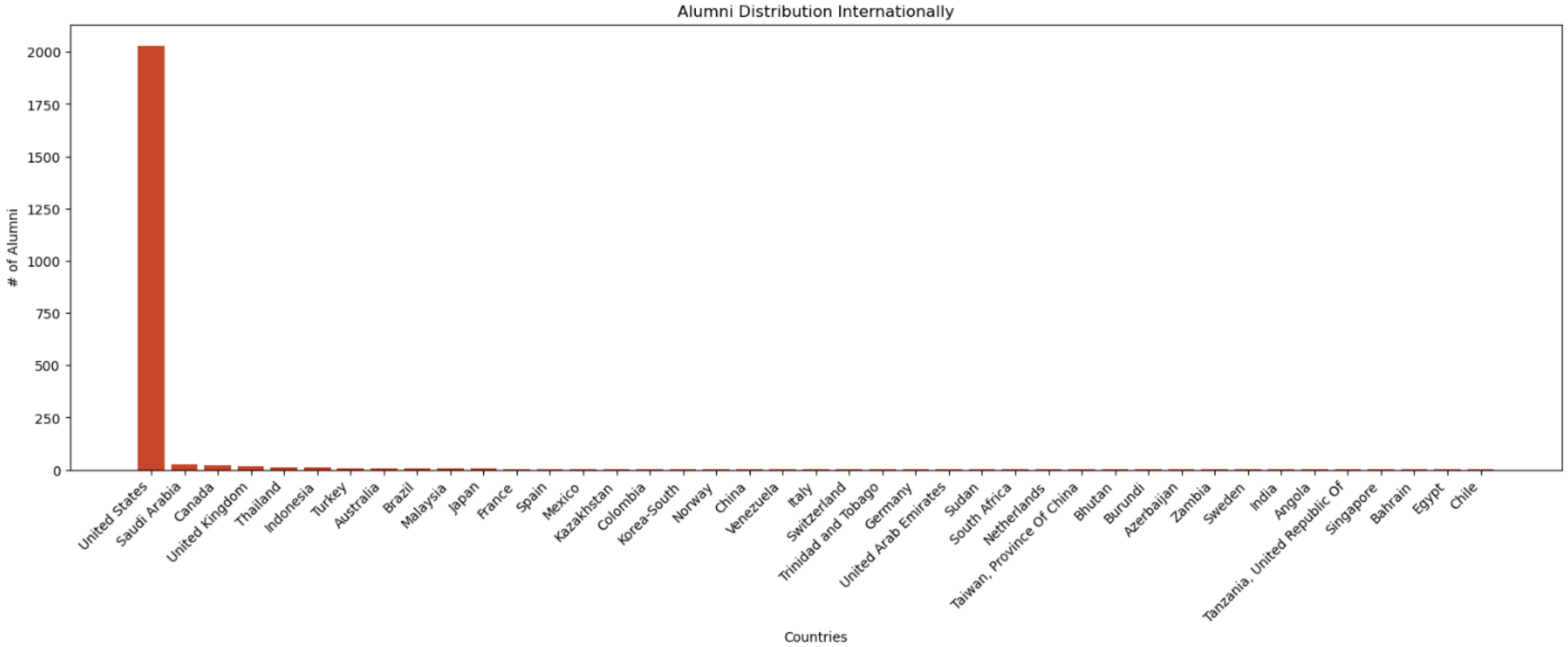
```
# count unique states, provinces, and countries
uniqueStates = df['State'].unique()
uniqueCountries = df['Country'].unique()

# create a dictionary to store country centers
countryCenters = {}

# get latitude and longitude centers for each country
for country in uniqueCountries:
    center = get_boundingbox_country(country, output_as='center')
    if center:
        countryCenters[country] = center

#manually add some latitudes and longtitudes
countryCenters['Taiwan, Province of China'] = ['23.6978','120.9605']
countryCenters['Tanzania, United Republic Of'] = ['-6.3690','34.8888']

# add latitude and longitude centers to the df
df['Latitude'] = df['Country'].map(lambda x: countryCenters.get(x, [None, None])[0])
df['Longitude'] = df['Country'].map(lambda x: countryCenters.get(x, [None, None])[1])

# save updated df to GP_Alumni_List_Cleaned.csv
df.to_csv(path, index=False)
```

Findings + Figures

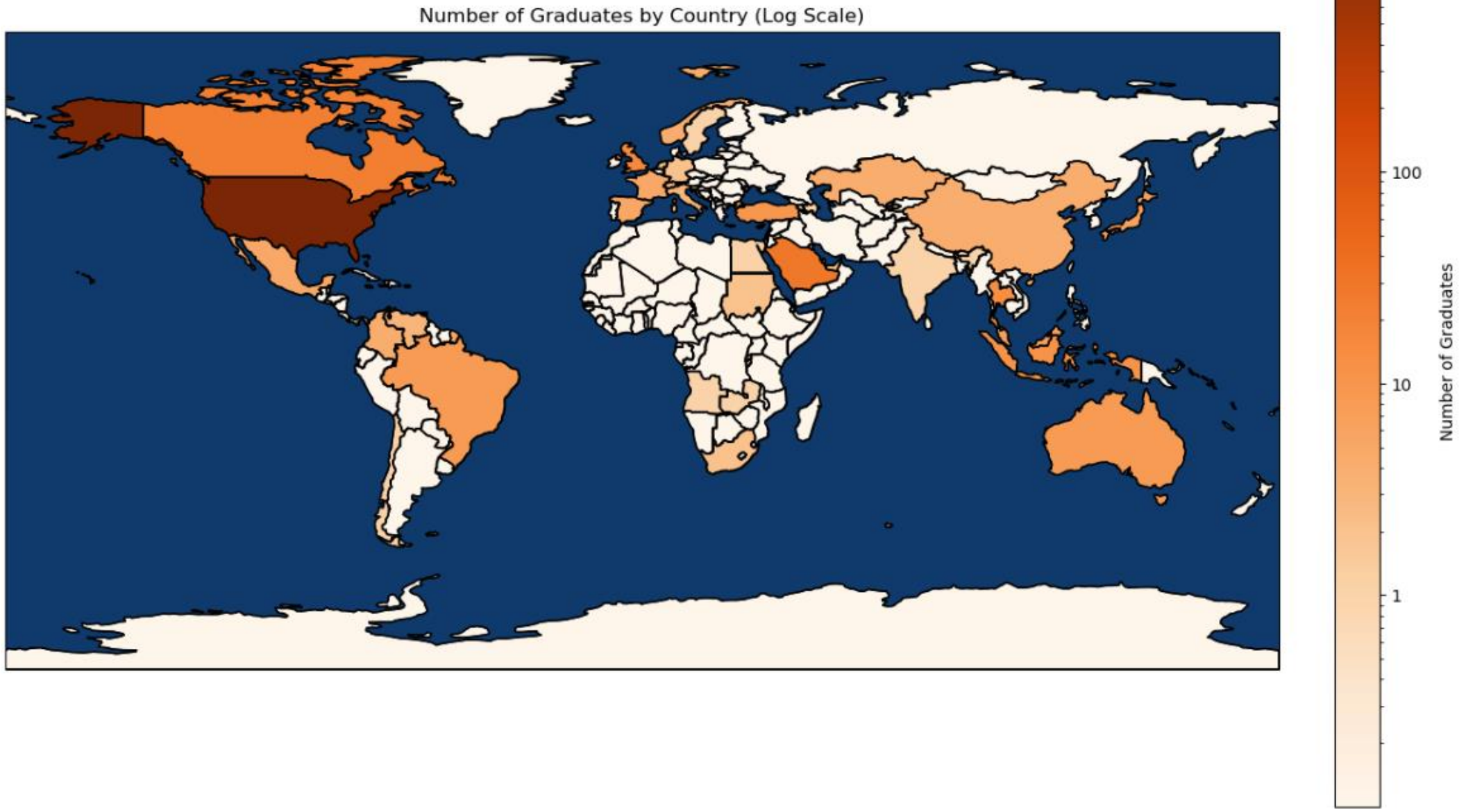
Pandas value count Bar Graphs



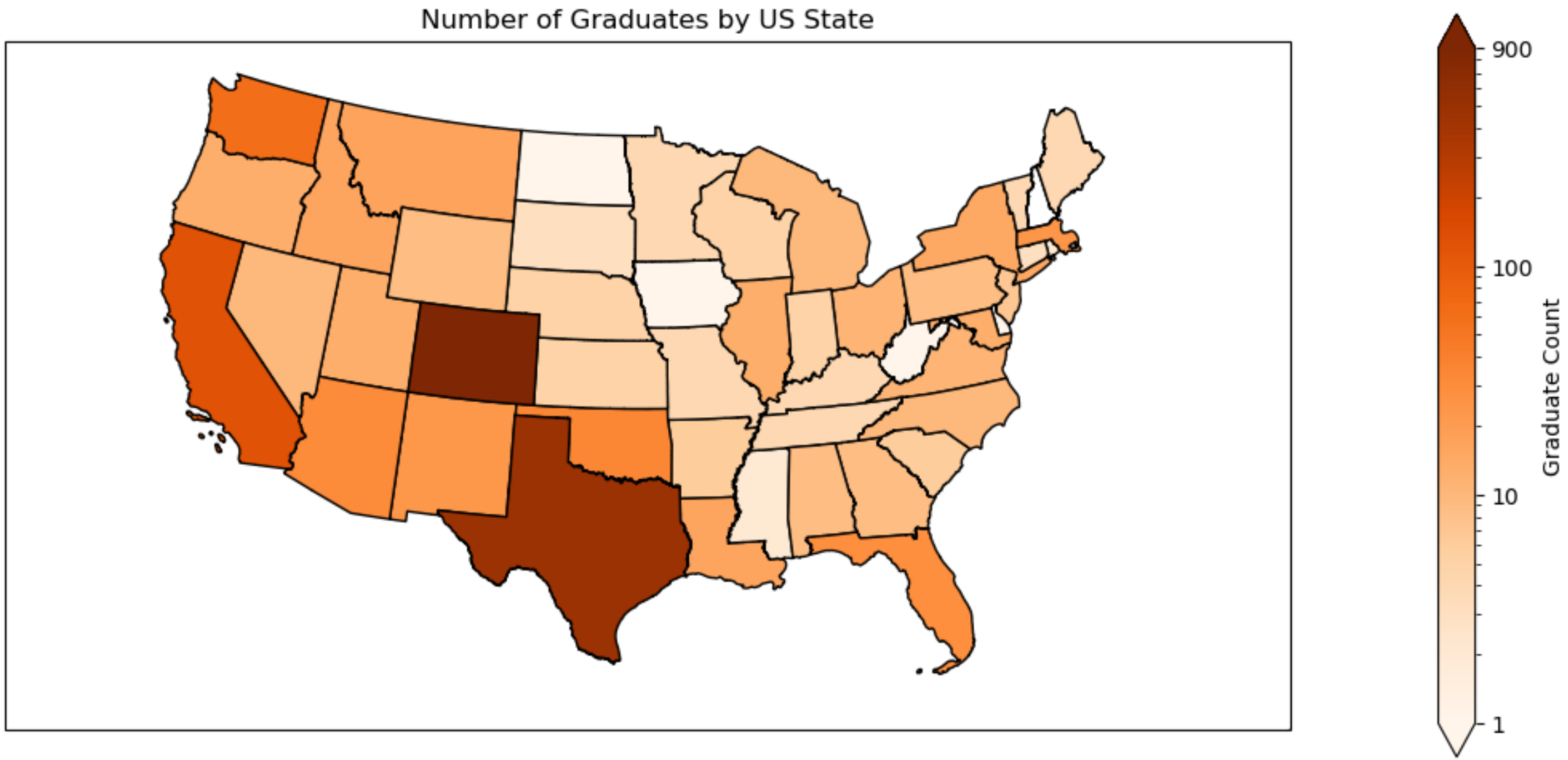
Pandas value count Bar Graphs

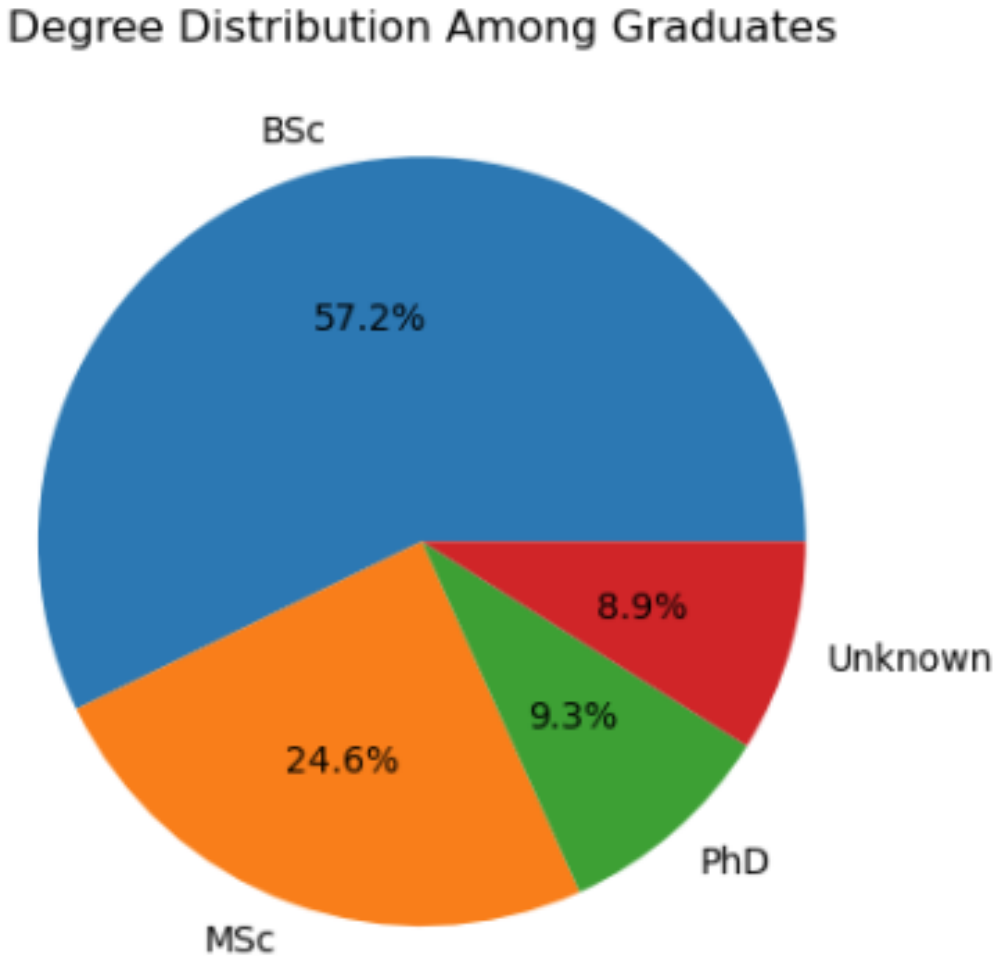


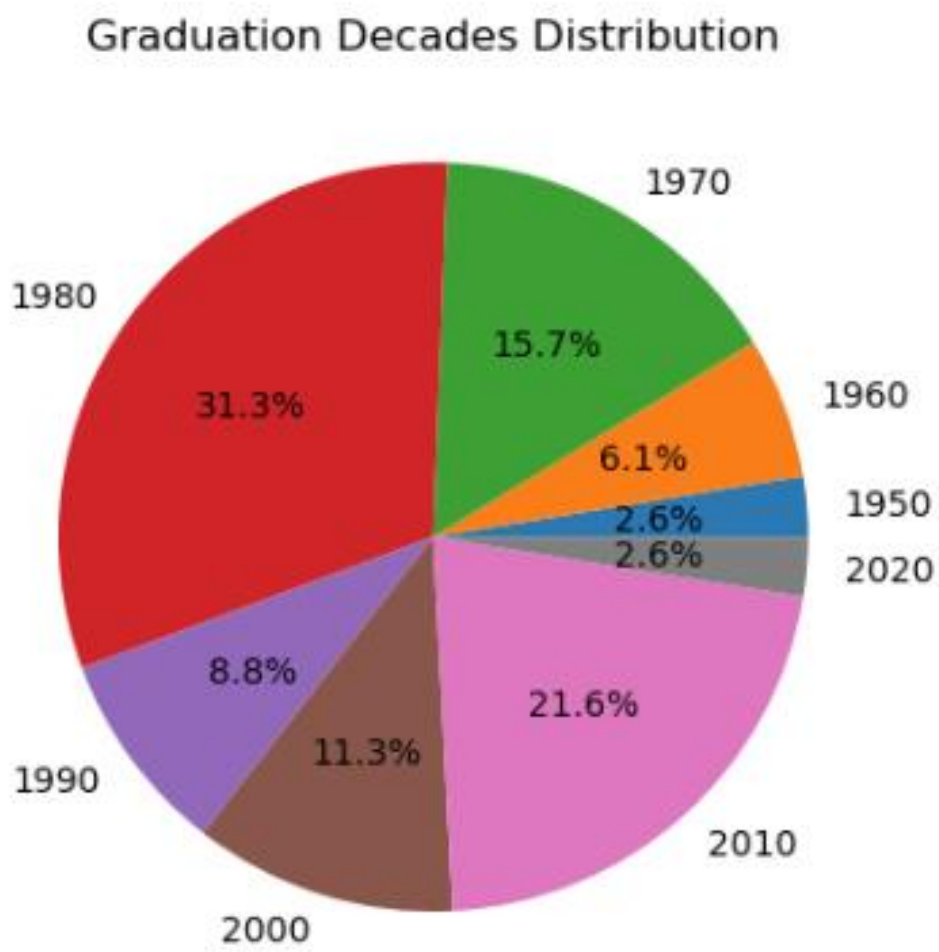
GIS Maps: (cartopy, geopandas, geopy)



GIS Maps: (cartopy, geopandas, geopy)




Degree Distribution Among Graduates
BSc
57.2%
8.9%
Unknown
9.3%
24.6%
PhD
MSc


Graduation Decades Distribution
1970
15.7%
1980
31.3%
1960
6.1%
2.6%
1950
2.6%
2020
8.8%
21.6%
11.3%
1990
2010
2000

Main Findings and Significance

- High Concentrations of Alumni in CO, USA
- Many Unique Locations (51 States, 42 Countries)
- # of Degrees Scale with Level of Degree

Interpretation

- Why are there so many alumni in TX and CA
- Why are there more alumni from the 1980s

Growth Opportunities

- Optimize process for finding coordinates for US States
- Interactive Maps
- Saving Figures Optimally
- Port to GP100 Website

Acknowledgments

People

- Melinda Gale
- GP Alumni
- Bia

Libraries

- NumPy
- Pandas
- Matplotlib
- CartoPy
- GeoPy
- GeoPandas