Politics of a Pandemic

Team Members: Cody Tong | Daniel Rossi | Fatima Carrillo | Geoff Farrell | Jackie Huang | William Pryor

Resources: [The COVID Tracking Project]

[Kaggle 2016 U.S. Presidential Election]

[US Census]

Overview

- How did different US states fare during the COVID19 pandemic in terms of cases and deaths?
- Is there a direct correlation between state political party preference and COVID cases and deaths by state?
- Why does the analysis trend more heavily on one side, if at all?

Three main data sources used for analysis

1. *The COVID Tracking Project: COVID19 test statistic data*

2. *Kaggle, 2016 U.S. Presidential Elections*

3. *U.S. Census, 2019 Data*

Pandas Code: Loading in .CSV resources

```
# variable to hold csv
#covid case data by state including testing
covidtrack_to_load = "Resources/all-states-history 10-26-2020.csv"
#data recording population of US states taken in 2019
populationfile = 'Resources/pop_est19.csv'
#matches state names to their abbreviations
keyfile = 'Resources/state_abbr list.csv'
#creates dataframe for state abbreviation keys
keydf = pd.read_csv(keyfile)
#creates dataframe for state population data
populationdf = pd.read_csv(populationfile)
#makes state name index of populationdf to support merging with other datasets
populationdf = populationdf.set_index(["NAME"])
# reads covidtrack data into dataframe
covidtrack_data = pd.read_csv(covidtrack_to_load)
#displays covidtrack data
covidtrack_data.head()
```

Pandas Code: Merging Dataframes

```
# renamed_df = final_data.rename(columns={"state_x":"States"})
final_data_lng_lat = pd.merge(final_data, state_coord_df, on="state", how="left"
final_data_lng_lat
```

```
#files to load in for dfs
geopolitics_file = "Resources/2016_US_County_Level_Presidential_Results.csv"
geopolitics_df = pd.read_csv(geopolitics_file)

merged_df_geo =pd.merge(geopolitics_df,keydf,on="state_abbr",how="left")
geopolitics_df_duplicates_removed = pd.DataFrame.drop_duplicates(merged_df_geo)
clean_politics_2016_df = pd.DataFrame(geopolitics_df_duplicates_removed)
clean_politics_2016_df
```

```
covid_state_politics = covid_state_politics.merge(abbr_df, on = 'States')
covid_state_politics.head()
```

Combined Dataframe with State Political Preference and COVID19 Data

States	Total positive	Total negative	Total test	Total Deaths	Population	Cases per 1 M	Deaths per 1 M	votes_dem	votes_gop	Party Score	Death Rate
Alabama	185322	1148993	1307694	2866	4903185.0	37796.248765	584.518022	718084.0	1306925.0	0.549445	0.015465
Alaska	14413	556423	570836	68	731545.0	19702.137257	92.953954	2697087.0	3781977.0	0.713142	0.004718
Arizona	238962	1481294	1714772	5875	7278717.0	32830.236428	807.147743	936250.0	1021154.0	0.916855	0.024585
Arkansas	106727	1195429	1294839	1833	3017804.0	35365.782536	607.395311	378729.0	677904.0	0.558676	0.017175
California	900957	17081357	17982314	17357	39512223.0	22801.982060	439.281789	7362490.0	3916209.0	1.880004	0.019265
Colorado	95087	1061126	1842254	2076	5758736.0	16511.783141	360.495775	1212209.0	1137455.0	1.065720	0.021833
Connecticut	68099	2139838	2207937	4589	3565287.0	19100.566097	1287.133406	884432.0	668266.0	1.323473	0.067387
Delaware	24168	317893	342061	685	973764.0	24819.155360	703.455868	235581.0	185103.0	1.272702	0.028343
Florida	782011	5202712	5984723	16652	21477737.0	36410.307101	775.314457	4485745.0	4605515.0	0.973994	0.021294
Georgia	351879	3109507	3461386	7827	10617423.0	33141.657820	737.184532	1837300.0	2068623.0	0.888175	0.022243
Hawaii	14877	295153	505253	212	1415872.0	10507.305745	149.731049	266827.0	128815.0	2.071397	0.014250

Pandas Code: Dataframe with State Coordinate Data for Creating Heatmaps

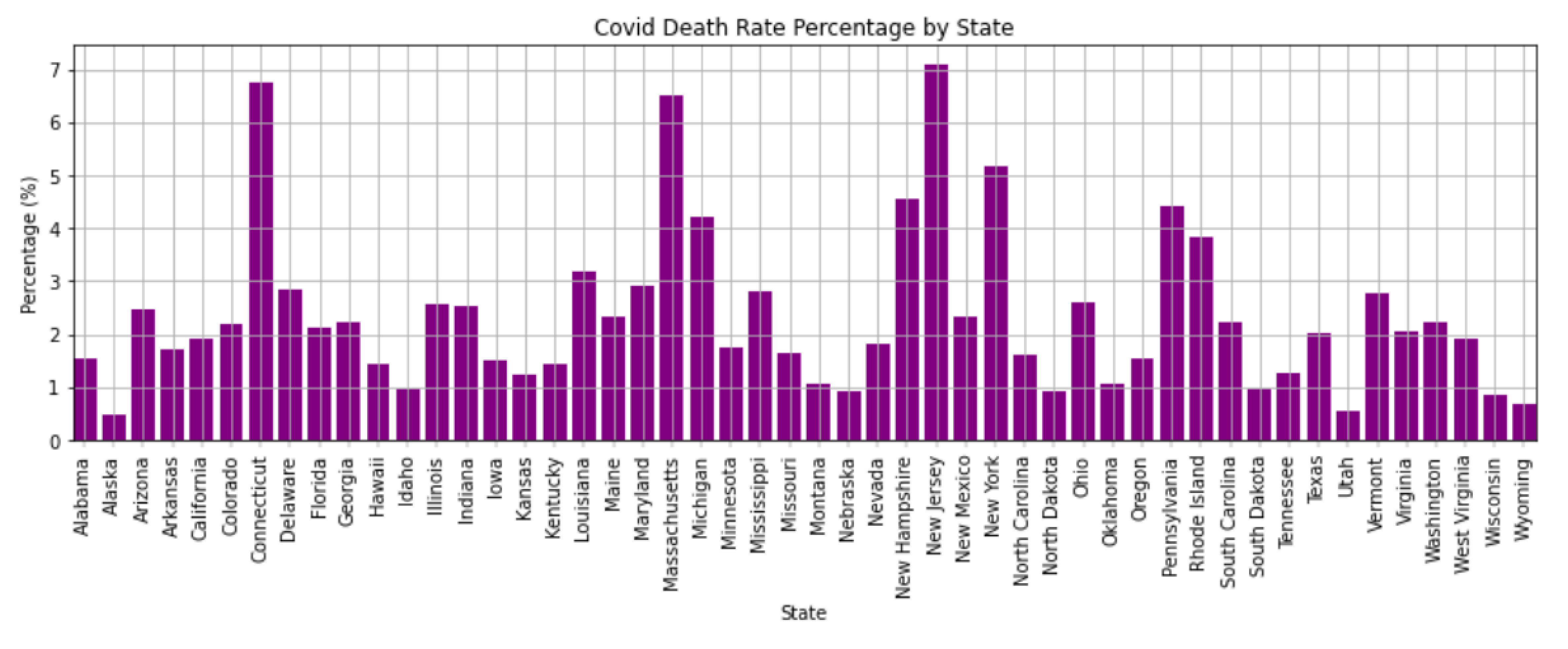
```
#loads csv containing state coordinate data for creating a heatmap plot
#loads data into df
state_coord_df = pd.read_csv('Resources/state_geocoord.csv', header=0)
state_coord_df.head()
state_coord_df.rename(columns = {'name':'States'}, inplace = True)
state_coord_df
```

Pandas Dataframe with State Coordinate Data for Creating Heatmaps

	state	latitude	longitude	States
0	AK	63.588753	-154.493062	Alaska
1	AL	32.318231	-86.902298	Alabama
2	AR	35.201050	-91.831833	Arkansas
3	AZ	34.048928	-111.093731	Arizona
4	CA	36.778261	-119.417932	California
5	CO	39.550051	-105.782067	Colorado
6	CT	41.603221	-73.087749	Connecticut
7	DC	38.905985	-77.033418	District of Columbia
8	DE	38.910832	-75.527670	Delaware
9	FL	27.664827	-81.515754	Florida
10	GA	32.157435	-82.907123	Georgia

Matplotlib Code: COVID19 Deaths Per States Barchart

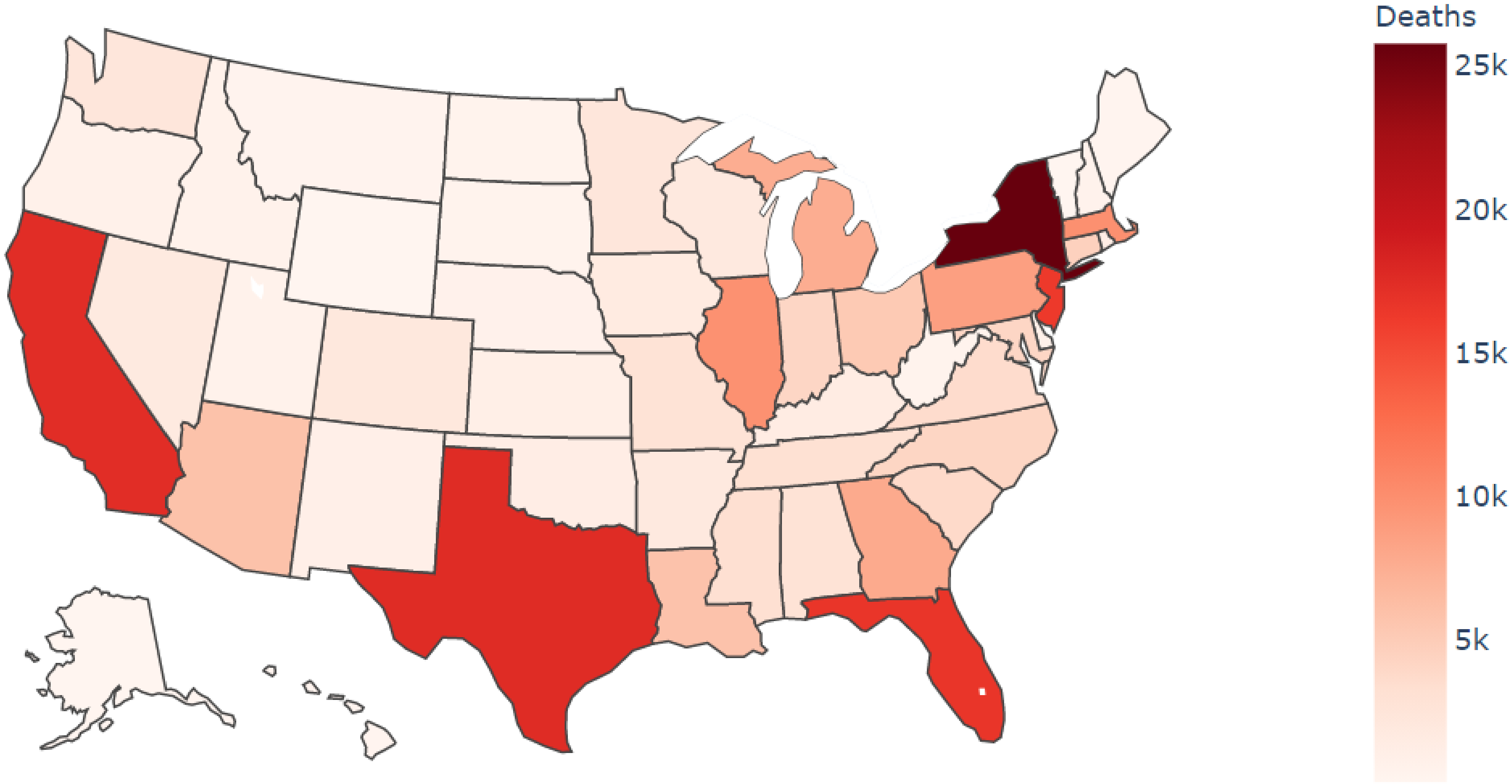
```
#Creates barchart for US states based on death rate
states_list = covidpolitics_df['States'].tolist()
x_axis = np.arange(len(covidpolitics_df))
final_percentage = covidpolitics_df['Death Rate']*100

tick_mark = [tick for tick in x_axis]

plt.xticks(tick_mark, covidpolitics_df['States'], rotation = 'vertical')
plt.bar(x_axis, final_percentage, color="purple")

plt.tick_params(axis='x', which='major', labelsize=10, width=0.25)

plt.gca().margins(x=0)
plt.gcf().canvas.draw()
tl = plt.gca().get_xticklabels()
maxsize = max([t.get_window_extent().width for t in tl])
m = 0.2 # inch margin
s = maxsize/plt.gcf().dpi*55+2*m
margin = m/plt.gcf().get_size_inches()[0]

plt.gcf().subplots_adjust(left=margin, right=1.-margin)
plt.gcf().set_size_inches(12, plt.gcf().get_size_inches()[1])

plt.title("Covid Death Rate Percentage by State")
plt.ylabel("Percentage (%)")
plt.xlabel("State")
plt.grid()

plt.savefig("percentages_vs_states", dpi=300)
```


Covid Death Rate Percentage by State
Percentage (%)
0
1
2
3
4
5
6
7
State
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming

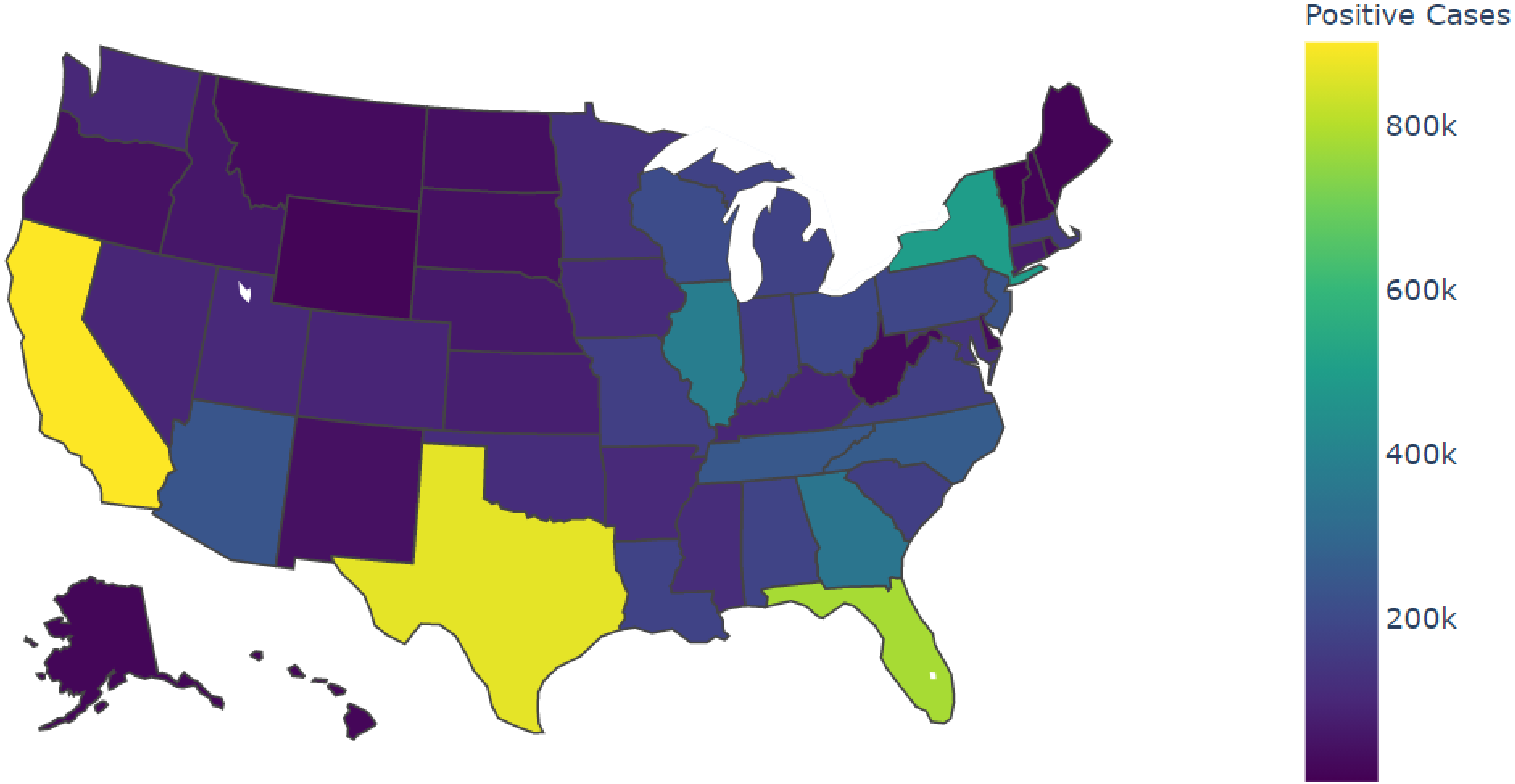
Choropleth: Heatmap Code Example

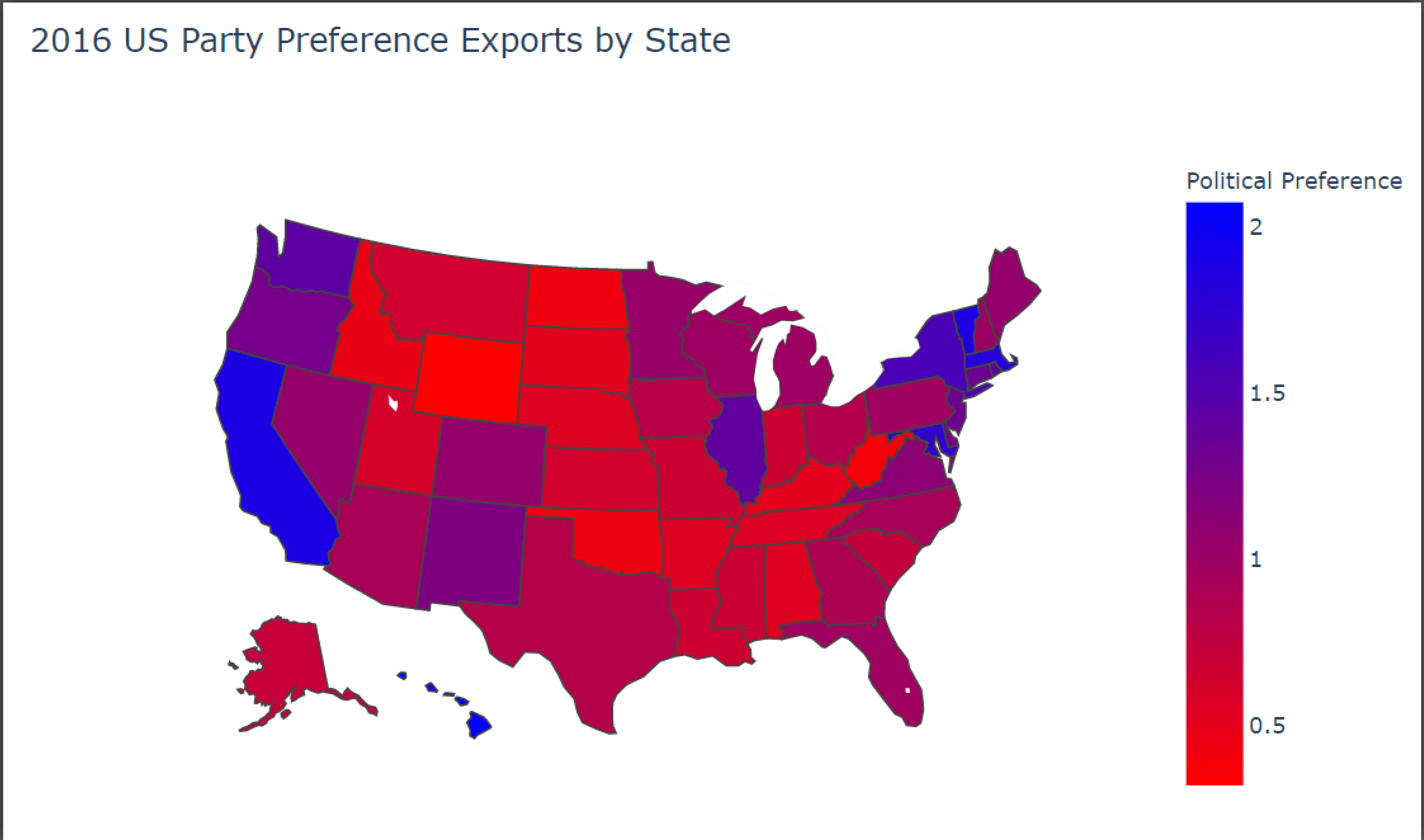
```
# Create choropleth for deaths by state

fig = go.Figure(data=go.Choropleth(
    locations=covidpolitics_df['state'], # Spatial coordinates
    z = covidpolitics_df['Total Deaths'].astype(float), # Data to be color-coded
    locationmode = 'USA-states', # set of locations match entries in `locations`
    colorscale = 'Reds',
    colorbar_title = "Deaths",
))
fig.update_layout(
    title_text = 'State Deaths',
    geo_scope='usa', # limit map scope to USA
)
fig.show()
```

State Deaths



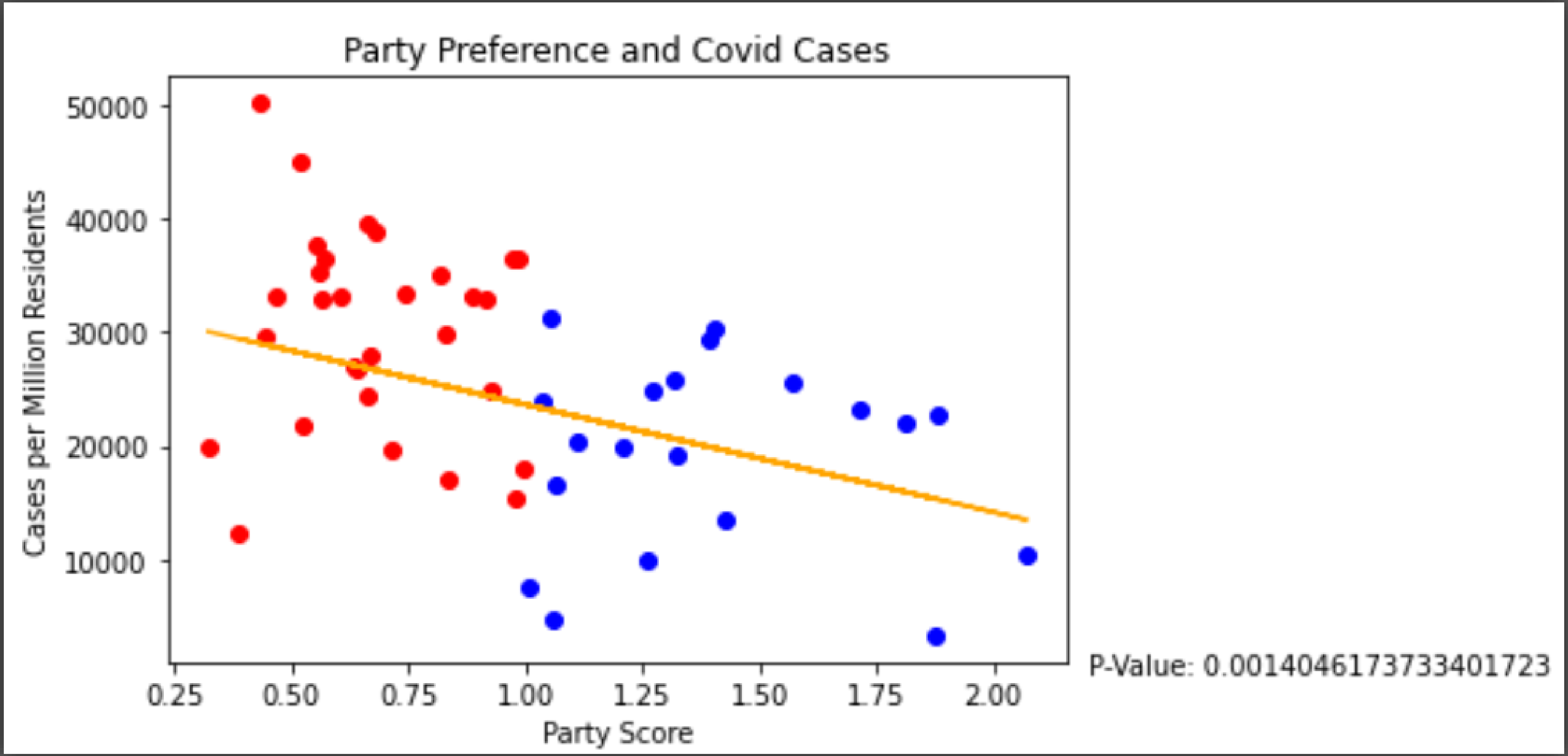
State Positive Cases

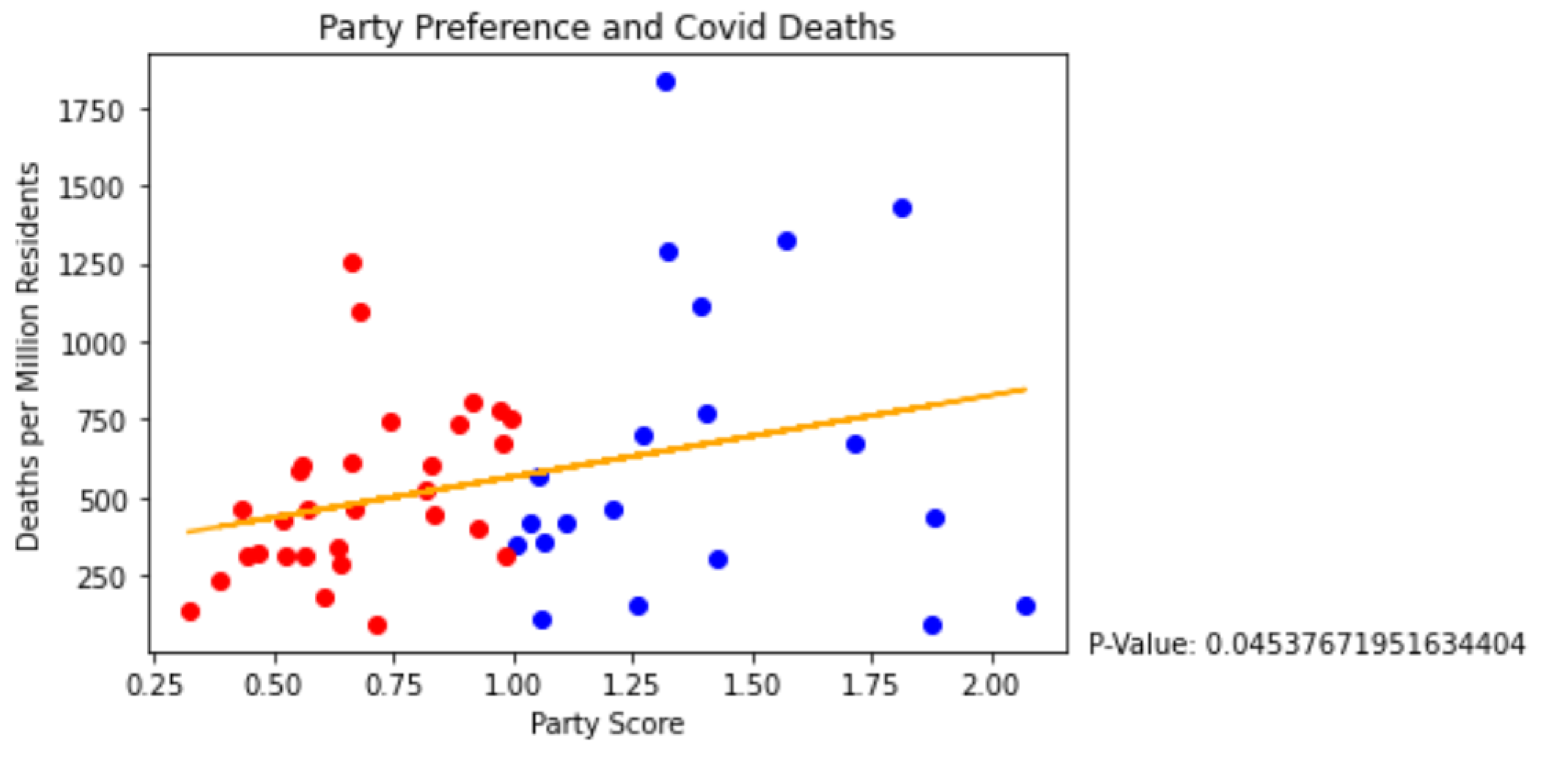



2016 US Party Preference Exports by State
Political Preference
2
1.5
1
0.5

Linear Regression Scatterplot Code Example: COVID Cases per 1 M Residents of US states

```
import matplotlib.pyplot as plt
import scipy.stats as st

# Calculate the correlation coefficient and linear regression model
# for mouse weight and average tumor volume for the Capomulin regimen
st.linregress(covid_state_politics["Party Score"], covid_state_politics["Cases per capita"])
# linregresult(slope=(-9466.724941731432), intercept=33108.252994614144)
slope = -9466.724941731432
intercept = 33108.252994614144
# rvalue = -0.43954453811755007
pvalue = 0.0014046173733401723
pvaluemarker = "P-Value: {}".format(pvalue)
#stderr= 2792.2825243710254
linreg = (covid_state_politics["Party Score"]*slope+intercept)
partyline = 1
plt.scatter(covid_democrat["Party Score"], covid_democrat["Cases per capita"], c='blue')
plt.scatter(covid_gop["Party Score"], covid_gop["Cases per capita"], c='red')
plt.plot(covid_state_politics["Party Score"],linreg,color="orange")
plt.text(2.2,0,pvaluemarker)
plt.ylabel("Cases per Million Residents")
plt.xlabel("Party Score")
plt.title("Party Preference and Covid Cases")
```


Party Preference and Covid Cases
Cases per Million Residents
10000
20000
30000
40000
50000
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Party Score
P-Value: 0.0014046173733401723


Party Preference and Covid Deaths
Deaths per Million Residents
1750
1500
1250
1000
750
500
250
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Party Score
P-Value: 0.04537671951634404

Statistics Code: Independent Sample T-Test of Death Rate by State Based on Party Preference

```
Republican_preference_df = clean_df[clean_df['Party Score']<1]
Democrat_Preference_df = clean_df[clean_df['Party Score']>1]

p1_mean = Republican_preference_df["Death Rate"].mean()
p2_mean = Democrat_Preference_df["Death Rate"].mean()
```

```
Republican_preference_df_reduced = clean_df[clean_df['Party Score']<low_range]
Democrat_Preference_df_reduced = clean_df[clean_df['Party Score']>high_range]

p1_mean_reduced = Republican_preference_df_reduced["Death Rate"].mean()
p2_mean_reduced = Democrat_Preference_df_reduced["Death Rate"].mean()
```

```
p1_mean_reduced
```

```
0.014542982775021127
```

```
p2_mean_reduced
```

```
0.03561365321861592
```

```
stats.ttest_ind(Republican_preference_df_reduced["Death Rate"], Democrat_Preference_df_reduced["Death Rate"],
                equal_var=False)
```

```
Ttest_indResult(statistic=-3.7872858878325055, pvalue=0.001673935775959739)
```

Hypothesis Testing with Independent Sample T-Test

Null Hypothesis: There is no difference in COVID19 death rates in US states based on political preference

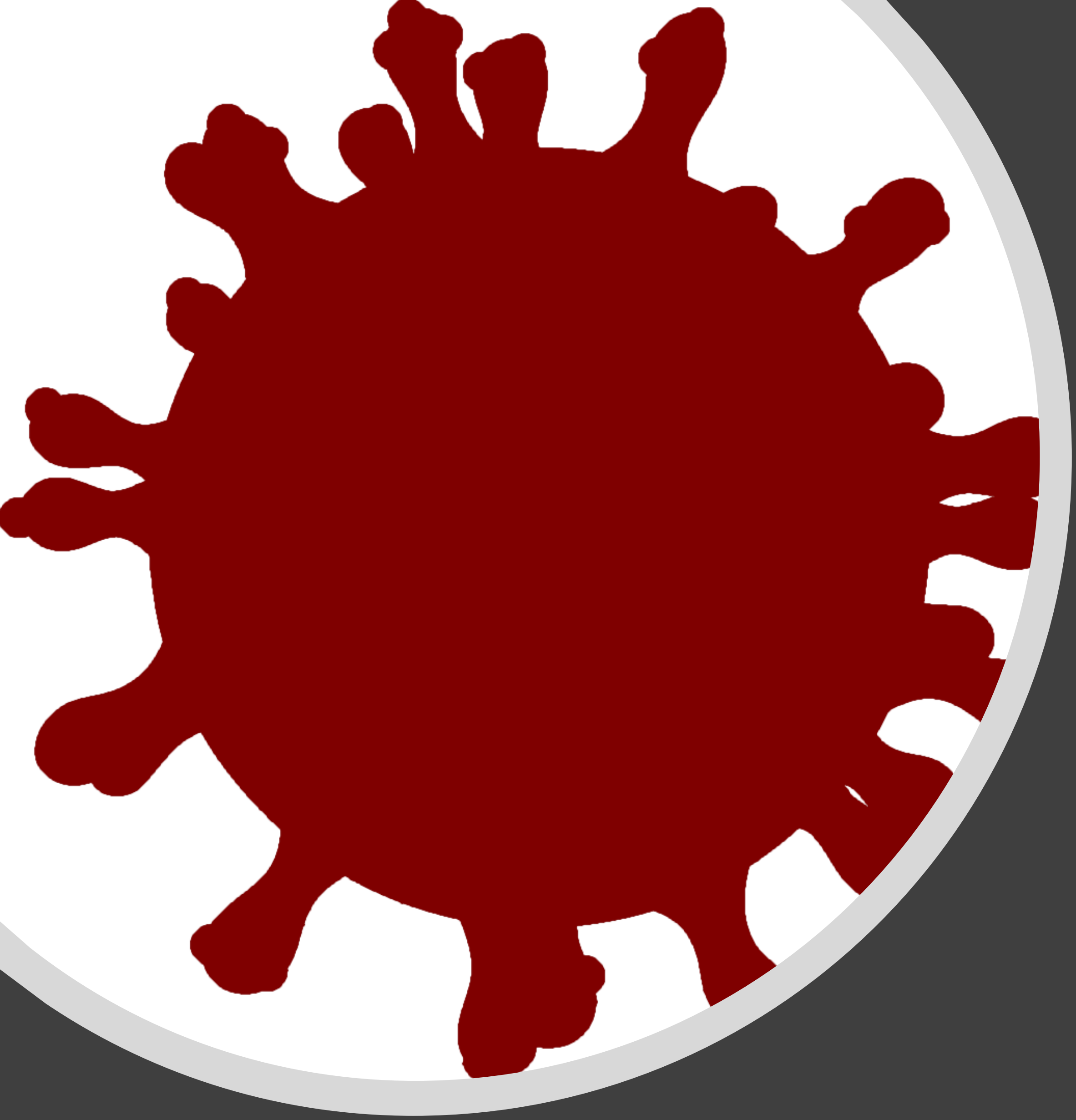
Alternative Hypothesis: Political preference is correlated to a difference in COVID19 death rates in US states

	Political Preference	N (States/Party)	Mean (Deaths/Case)	Test Statistic	P-Value
COVID19 Death Rate	Republican	30	0.01454	-3.78729	0.00167
	Democrat	20	0.03561		

Conclusions of Analysis

1. Geographically, most COVID19 death hotspots located in Northeastern United States
   - Potentially due to influence of climate
   - Higher population densities on average

2. Data suggests contracting COVID19 more likely in red states, but surviving more likely in blue states
   - Political Variables potentially at play:
     - Higher mask wearing compliance in blue states results in fewer cases
   - Non-political variables potentially at play:
     - Population densities
     - Climate
     - Geographic distribution of different COVID19 strains



Final Thoughts

- Correlation =/= Causation
- While we discovered a trend relating state political preference to COVID19 Outcomes, its dangerous to draw conclusions based on such a limited analysis of variables!