Final Project Presentation

Hannah Kost, Rose Kottwitz, Summer Liu

Table of Contents

01 Introduction

Issues of Access

Thomas R. Frieden

As reported by AMA

Murphy, Brendan. "Access, Not Hesitancy Now Biggest Barrier to COVID-19 Vaccination." *AMA*, 19 May 2021, www.ama-assn.org/delivering-care/public-health/access-not-hesitancy-now-biggest-barrier-covid-19-vaccination. Accessed 4 Dec. 2021.

Ann Lee and Sheila Davis

In the *Stanford Social Innovation Review*

Lee, Ann, and Sheila Davis. "Ensuring Equitable Access to Vaccines." *Stanford Social Innovation Review*, 29 Jun. 2021, https://ssir.org/articles/entry/ensuring_equitable_access_to_vaccines#. Accessed 4 Dec. 2021.

Potential Solutions

Thomas R. Frieden

As reported by AMA

Ann Lee and Sheila Davis

In the *Stanford Social Innovation Review*

Our informatics problem

COVID-19 Confirmed Cases

by United States county

- Employment
- Sex
- Socioeconomic Status
- Age
- Racial Identity

Our Goal:

To find a correlation between the number of confirmed cases in a U.S. county and socioeconomic status—calculated through analysis of the distribution of race, unemployment, sex, and age in said county—pointing towards a greater issue: the lack of access to preventative and essential resources.

Our Hypothesis: A Spectrum

Data Preparation

02

DATA EXPECTATIONS:

Based on our hypothesis we needed 2 datasets:

Dataset #1:

- COVID data (by county)
- Death count/Positive count
- Timeline
- Dependent/changing variable

Dataset #2:

- Brainstormed variables:
  - Race, age, sex, socioeconomic status, education, location, etc.
- Independent variables

Database 1: Covid_Cases.csv

COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University

UID	Province_State	County_Region	Combined_Key	Dates
84031109	Nebraska	US	Lancaster, Nebraska, US	3/13/2020 3/12/2021

3,343 rows

Dataset #1:

- COVID data by county
- Chose positive count, more numbers

Pro:
- Minimal cleaning, more accuracy

Con:
- massive data set

Dataset #1:

- COVID data (by county)
- Death count/Positive count
- Timeline
- Dependent/changing variable

Database 2: Population_Est.csv

Annual County Resident 2020 Population Estimates by Age, Sex, Race, and Hispanic Origin based on 2010 Census

28,288 rows

SUM LEV	State	County	STNAME	CITYNAME	YEAR	AGEGRP	TOT_POP	TOT_MALE	TOT_FEMALE
50	13	25	Georgia	Brantely County	13	12	1422	737	685

Dataset #2:

- Race, age, gender by county
- Independent variables
- But no socioeconomic indicators

Dataset #2:

- Brainstormed variables:
  - Race, age, sex, socioeconomic status, education, location, etc.
- Independent variables

Racial Distribution (M, F):

- WA 660, 648
- BA 61, 20
- IA 2, 3
- AA 1, 1
- NA 0, 0
- H 51, 3

Database 3: Unemployment_Rate.csv

United States Department of Agriculture Economic Research Service

State FIPS Code	County FIPS Code	County Name/State Abbrev	Period	Labor Force	Employed	Unemployed	Unemployment Rate
01	001	Autauga County, AL	July-20	25,811	24,190	1621	6.3%

45,066 rows

Dataset #3:

- Unemployment rate by county
- Independent variables
- Couldn't cover all variables with 1 database

Three databases

- All were consistent - sorted by county
- Missing minimal information
- Reliable
- Many diverse factors

Dataset #2:

- Brainstormed variables:
  - Race, age, sex, socioeconomic status, education, location, etc.
- Independent variables

03 Data Cleaning & Preprocessing

Cleaning and Preprocessing: Excel

COVID Data

EXTRACTION:
- March 13, 2020 -> March 12, 2021

DELETION:
- All other ID Properties
- Long./Lat., iso2, iso3, FIPS, Province_State

Step 1

Unemployment Data

EXTRACTION:
- Unemployment data
- March 2021 data

DELETION:
- July 2020 - August 2021 data, except March 2021
- specific metropolitan areas

Step 2

Population Data

EXTRACTION:
- Population by age, gender, race
- Created ratios

DELETION:
- No deletion necessary
- specific metropolitan areas

Step 3

Data Analysis + Programming

04

Scores & Data Analysis

Score	1	2	3	4	5
Unemployment Data	0.7-6.06%	6.07-11.42%	11.43-16.78%	16.79-22.14%	22.15-27.5%
Racial Distribution Data	.00823-.22725	.22726-.44625	.44526-.66525	.66526-.88426	.88427-1.10326
Gender Distribution Data	.30491-.54481	.54482-.78471	.78472-1.02460	1.02461-1.26450	1.26451-1.50439

Scores & Data Analysis

Score	1	2	3	4	5
Unemployment Data	0.7-6.06%	6.07-11.42%	11.43-16.78%	16.79-22.14%	22.15-27.5%

Unemployment % per county range: 0.7% - 27.5%
Point range: 5.36%

Scores & Data Analysis

Score	1	2	3	4	5
Racial Distribution Data	.00823-.22725	.22726-.44625	.44526-.66525	.66526-.88426	.88427-1.10326

$$\frac{\text{Total Minority Population}}{\text{Total Population}}$$

Avg ratio per county (by state) range: 0 - 1.277734
Point range: 0.25554

Database 2: Population_Est.csv

Annual County Resident 2020 Population Estimates by Age, Sex, Race, and Hispanic Origin based on 2010 Census

SUM LEV	State	County	STNAME	CITYNAME	YEAR	AGEGRP	TOT_POP	TOT_MALE	TOT_FEMALE
50	13	25	Georgia	Brantely County	13	12	1422	737	685

28,288 rows

https://uofnelincoln-my.sharepoint.com/:x:/g/personal/hkost2_unl_edu/EZBx_1CYQJ5Gp5PUqncN9rIBggqvMiMptDMQ2cdZJ-fsdg

Racial Distribution (M, F):

- WA 660, 648
- BA 61, 20
- IA 2, 3
- AA 1, 1
- NA 0, 0
- H 51, 3

Scores & Data Analysis

Score	1	2	3	4	5

$$\frac{\text{Total Minority Population}}{\text{Total Population}}$$

Gender Distribution Data	.30491-.54481	.54482-.78471	.78472-1.02460	1.02461-1.26450	1.26451-1.50439

https://uofnelincoln-my.sharepoint.com/:x:/g/personal/hkost2_unl_edu/EX38E2dQ0k5Fvf1UEHF70Z8BYO9J9-nZ31nj5Aqr9DvboQ

- Provided summer with an average # of cases per day list for correlation score

Sorting and Calculating through code

1. Variable.py

Separate variables file:
- Store score initial value
- Store score ranges

Will be imported into main function

3. Function: read_data_from_csv

- Read imported csv files (rates/ratios)
- Assigned score based on calculated range
- 3 variables: file name, column name, scope within function
- within main.py

2. Main.py

- Read in csv files
- Assign score based on calculated range
- Restrictions with decimals
- Imported Variable.py
- Used read_data_from_csv function
- Output into a new csv file for further analyzation

4. Correlation.py

- Calculate Pearson correlation coefficient
  - Index score
  - Avg. # of COVID cases per state
- Import numpy package, use correlation function

Scores & Data Analysis

A Perfect Score Distribution	3-6	7-10	11-15

Low Number of Confirmed Cases

Low Unemployment

Older Age

Low minority population

Average Number of Confirmed Cases

Average Unemployment

Variety of Ages

Even distribution of races

High Number of Confirmed Cases

Low Unemployment

Younger Age

Higher WA population

https://uofnelincoln-my.sharepoint.com/:x:/g/personal/hkost2_unl_edu/EZzikYZNQNhOkh0kWa9fFLsB311I9zf34SM8Qvw9xy7haw

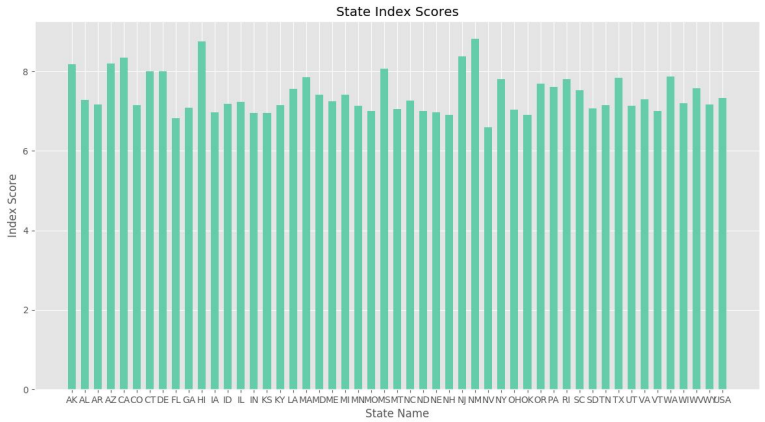
05 Data Visualization + MATPLOTLIB

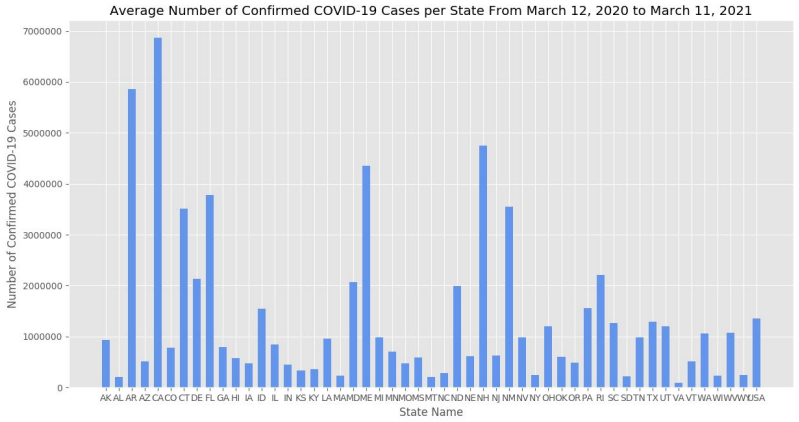
COVID_rate_chart.py

- Matplotlib package - bar graphs
- Average number of confirmed COVID-19 cases
  - Per state
  - March 13, 2020 - March 12, 2021
- Read in COVID_SCORE_STATEAVG.csv

Index_score_chart.py

- Matplotlib package - bar graphs
- Calculated sum index score across all counties > averaged to find state
  - Read in SCORE_OVERALL.csv

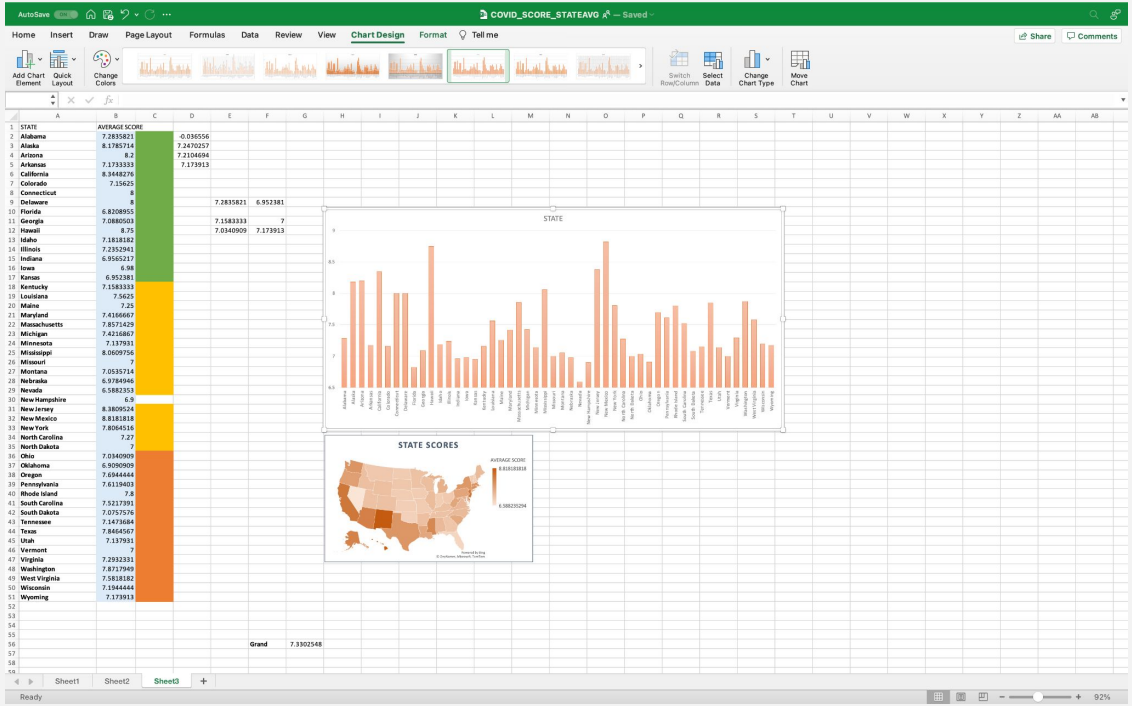




Matplotlib in Python

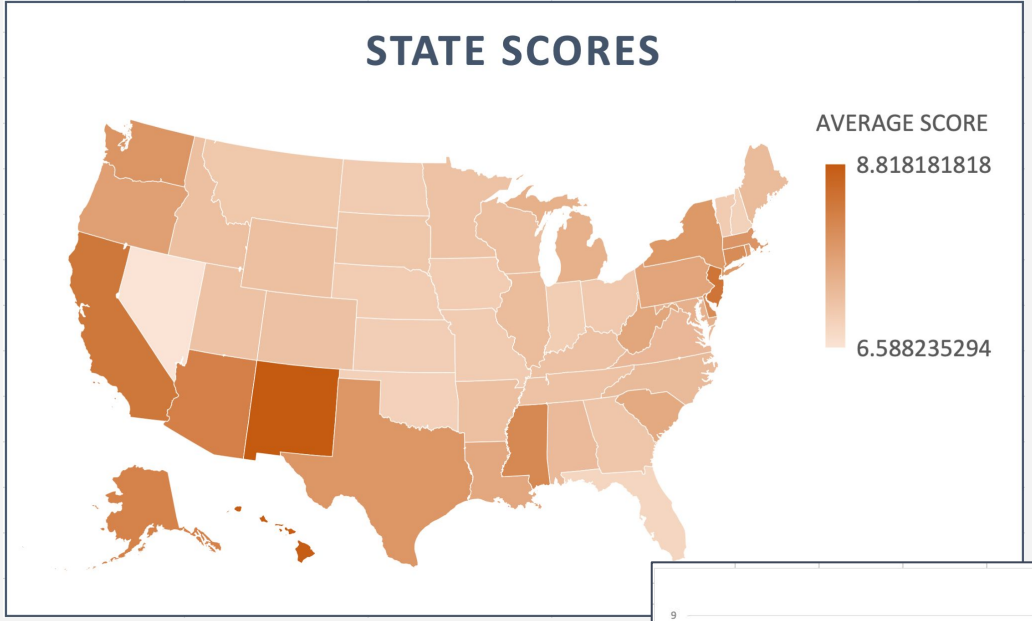
- Create initial graphs for index scores AND positive case count by state

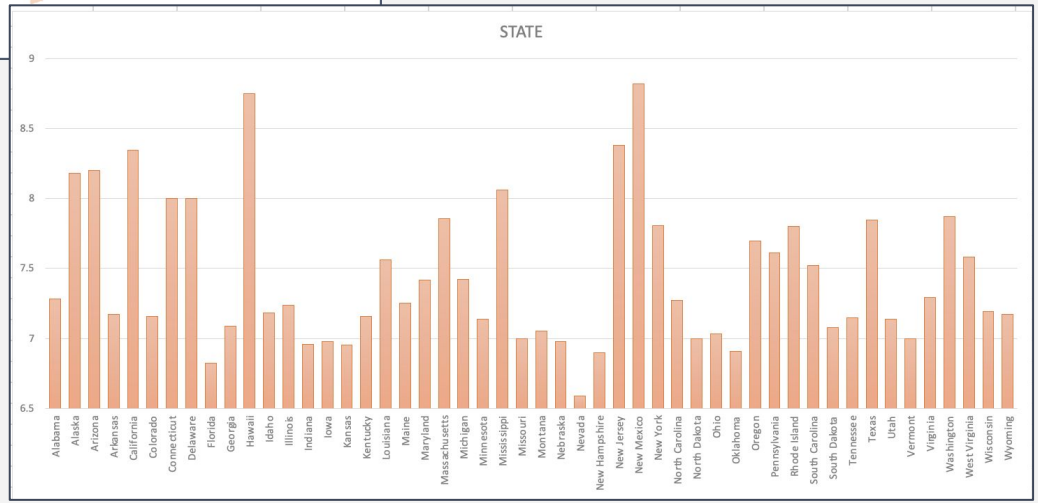
CREATING COVID NUMBERS DEPTH CHART:



STATE	AVERAGE SCORE				
Alabama	7.2835821		-0.036556		
Alaska	8.1785714		7.2470257		
Arizona	8.2		7.2104694		
Arkansas	7.1733333		7.173913		
California	8.3448276				
Colorado	7.15625				
Connecticut	8				
Delaware	8			7.2835821	6.952381
Florida	6.8208955				
Georgia	7.0880503			7.1583333	7
Hawaii	8.75			7.0340909	7.173913
Idaho	7.1818182				
Illinois	7.2352941				
Indiana	6.9565217				
Iowa	6.98				
Kansas	6.952381				
Kentucky	7.1583333				
Louisiana	7.5625				
Maine	7.25				
Maryland	7.4166667				
Massachusetts	7.8571429				
Michigan	7.4216867				
Minnesota	7.137931				
Mississippi	8.0609756				
Missouri	7				
Montana	7.0535714				
Nebraska	6.9784946				
Nevada	6.5882353				
New Hampshire	6.9				
New Jersey	8.3809524				
New Mexico	8.8181818				
New York	7.8064516				
North Carolina	7.27				
North Dakota	7				
Ohio	7.0340909				
Oklahoma	6.9090909				
Oregon	7.6944444				
Pennsylvania	7.6119403				
Rhode Island	7.8				
South Carolina	7.5217391				
South Dakota	7.0757576				
Tennessee	7.1473684				
Texas	7.8464567				
Utah	7.137931				
Vermont	7				
Virginia	7.2932331				
Washington	7.8717949				
West Virginia	7.5818182				
Wisconsin	7.1944444				
Wyoming	7.173913				
				Grand	7.3302548

1. Sorting & Filtering
2. Conditional Formatting
3. Use excel to create 2 types of visualization

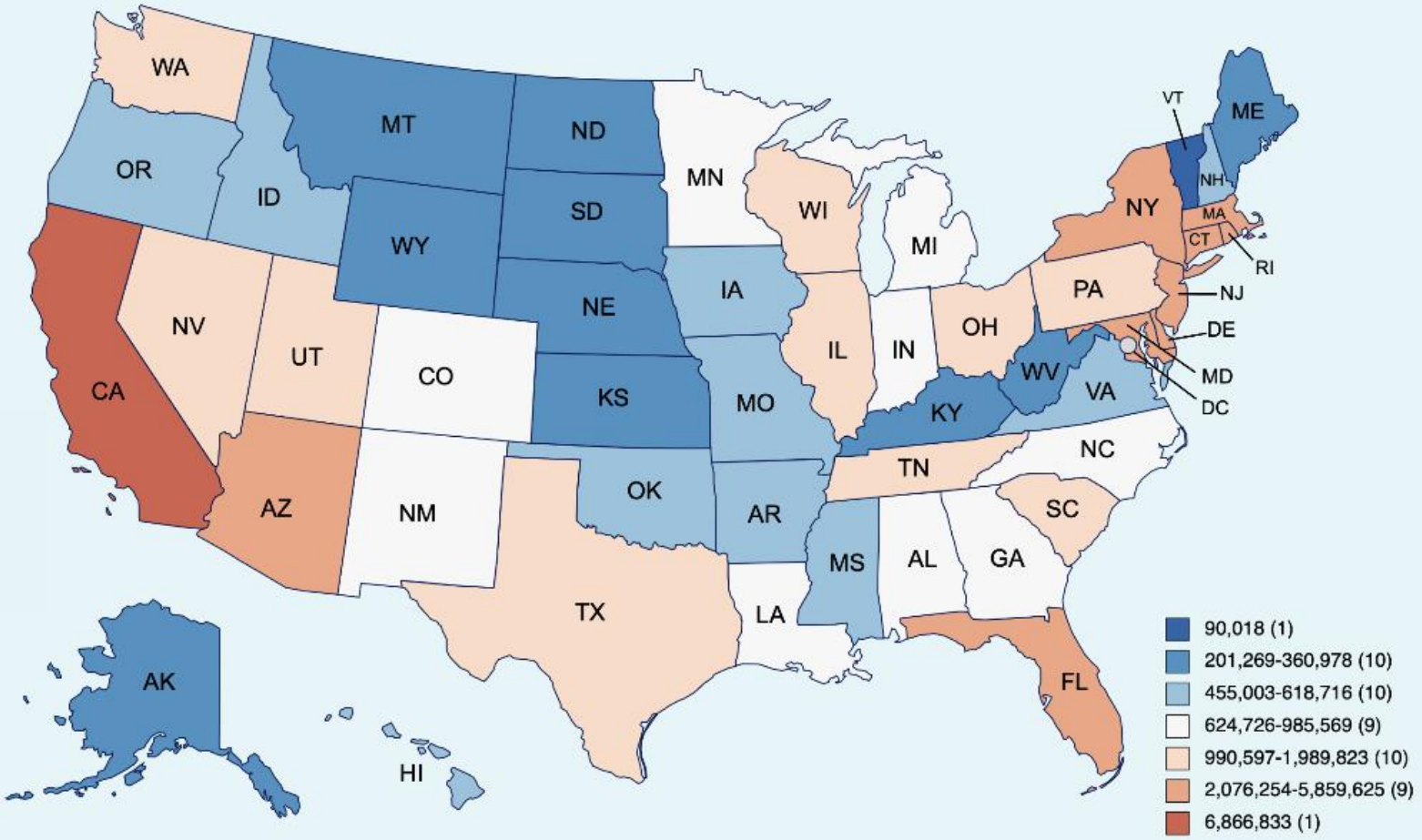

STATE SCORES
AVERAGE SCORE
8.818181818
6.588235294


STATE
9
8.5
8
7.5
7
6.5
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

CREATING COVID NUMBERS DEPTH CHART:

	A	B	C
1	Province_State	COVID_AVG	
2	Vermont	90018	DARKEST
3	Montana	201269.983	
4	Alaska	205015.788	
5	South Dakota	219333.25	
6	West Virginia	230585.86	
7	Maine	238699.333	
8	North Dakota	247405.164	
9	Wyoming	249946.08	
10	Nebraska	280486.779	
11	Kansas	331969.364	
12	Kentucky	360978.844	MEDIUM
13	Iowa	455003.347	
14	Mississippi	473692.179	
15	Idaho	480107.217	
16	Oregon	486210.605	
17	Virginia	507660.97	
18	Arkansas	509871.195	
19	Hawaii	575973.714	
20	Missouri	591266.297	
21	Oklahoma	607111.608	
22	New Hampshire	618176.583	MIDDLE
23	New Mexico	624726.743	
24	Minnesota	704725.022	
25	Colorado	781959.848	
26	Georgia	799076.814	
27	Indiana	846920.979	
28	Alabama	936812.841	
29	Louisiana	958097.455	
30	North Carolina	985240.235	
31	Michigan	985569.828	MIDDLE 2
32	Tennessee	990597.093	
33	Washington	1065837.59	
34	Wisconsin	1070393.53	
35	Ohio	1200553.18	
36	Utah	1200677.51	
37	South Carolina	1269975.81	
38	Texas	1286757.61	
39	Illinois	1552897.21	
40	Pennsylvania	1556906.25	
41	Nevada	1989823.79	MIDDLE 3
42	Maryland	2076254.35	
43	Delaware	2133250	
44	Rhode Island	2214268	
45	Connecticut	3506685.4	
46	New York	3552649.8	
47	Florida	3774085.55	
48	Massachusetts	4356430.12	
49	New Jersey	4752105.57	
50	Arizona	5859625.41	MEDIUM
51	California	6866833.52	DARKEST

1. Sorting & Filtering
2. Conditional Formatting
3. Split into groups
4. MapChart.net


WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MI
IN
KY
TN
MS
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
VT
NH
ME
MA
CT
RI
NJ
DE
MD
DC
AK
HI
90,018 (1)
201,269-360,978 (10)
455,003-618,716 (10)
624,726-985,569 (9)
990,597-1,989,823 (10)
2,076,254-5,859,625 (9)
6,866,833 (1)

06

Conclusions

No Correlation Found

- Correlation Score: 0.1668
- Can't say for certain that there is no correlation
- More avenues for research:
  - Looking only at counties, not consolidating to states
  - Comparing COVID-19 mitigation strategies in NY and CA
    - Are issues of access tied to population density?

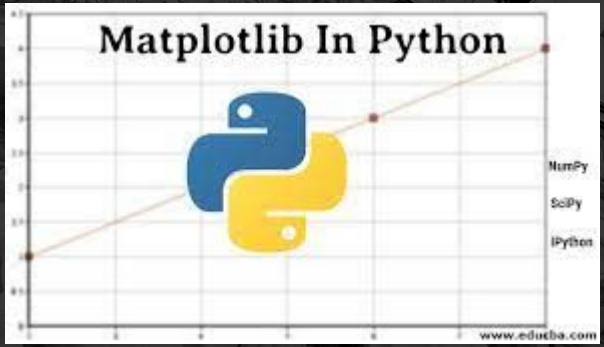
Improved Skills

- Importance of data visualization
  - Affects people's reaction to info. and can affect findings
  - Using variety of approaches and perspectives
- Better prepared to handle data-heavy projects in the future

07 Appendix

Python Purpose:

- Calculate and export
- Matplotlib
- Numpy



NumPy

Q&A