Final Project Status Update

Hannah Kost, Rose Kottwitz, Summer Liu

Table of Contents

01 Topic Overview

Our Topic Proposal

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

The Team and the Assigned Tasks

02

The Team: Who Will Lead

Summer Liu

Code writing and functions

Step 1

Hannah Kost

Data analyzation and organization

Step 2

Rose Kottwitz

Drawing conclusions and hand-ins

Step 3

The Team: Who Will Lead

Summer Liu

Code writing and functions

Step 1

TASKS/JOBS:

- Clean raw data and databases
- Build and define indicator framework
- Lead the creation of our code
- I/O Files
- Data Visualization

The Team: Who Will Lead

Hannah Kost

Data analyzation
and organization

Step 2

TASKS/JOBS:

- Lead the analysis and organization
- Create formulas
- Run tests
- Discover patterns
- Organize relevant results

The Team: Who Will Lead

TASKS/JOBS:

- Lead the Final Paper and hand-ins
- Create conclusions
- Construct and articulate processes
- Analyze results

Rose Kottwitz

Drawing conclusions and hand-ins

Step 3

03 A Look Into Our Databases

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

Database 1: Covid_Cases.csv

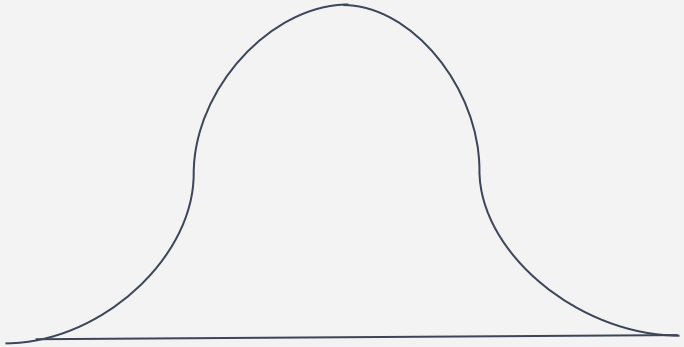
COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University

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

3,343 rows

Preliminary Research:

- National Numeral Distribution
- Regional Numeral Distribution
- Quarterly Numeral Distribution
- High profile/Low profile counties



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	Brantley County	13	12	1422	737	685

28,288 rows

Racial Distribution (M, F):

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

Database 2: Population_Est.csv

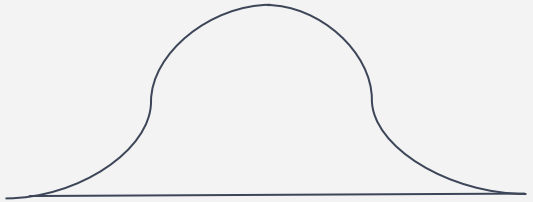
Annual County Resident 2020 Population Estimates by Age, Sex, Race, and Hispanic Origin based on 2010 Census

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

Preliminary Research:

- Age Group Distribution + Ratios
- Sex Depth Distribution + Ratios
- Race Depth Distribution + Ratios
- Population Numeral Distribution
- High profile/Low profile counties



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

Database 3: Unemployment_Rate.csv

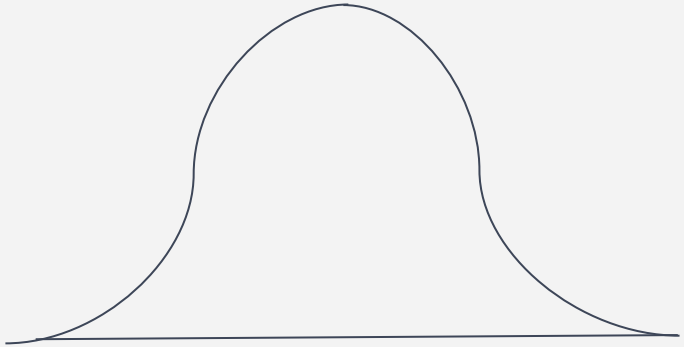
United States Department of Agriculture Economic Research Service

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

Preliminary Research:

- Unemployment Distribution + Ratios
- Establish defined range between distribution
- Race Depth Distribution + Ratios



Collection & Analyzation

04

Scores & Data Analysis

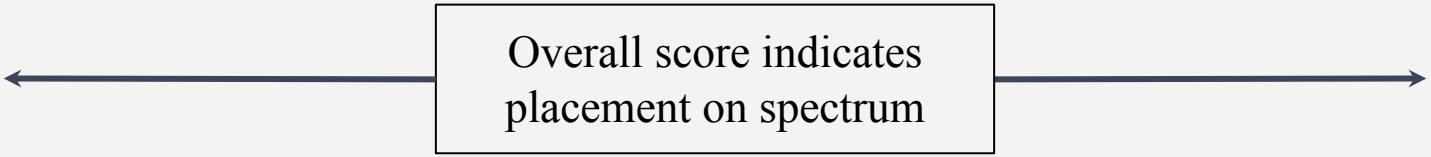
1. Calculate total socioeconomic index score per county
2. Create an overlapping linear visual for COVID cases and Socioeconomic status by county/state

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	—	—	—	—	—
Gender Distribution Data	—	—	—	—	—

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%

Range: 5.36



Scores & Data Analysis

Score	1	2	3	4	5
Racial Distribution Data	—	—	—	—	—

Calculate ratios to determine diversity, defined as:

$$\frac{\text{Population Proportion of WA}}{\text{Population Proportion of non-WA}}$$

Base scores on diversity range, split into 5 equally distributed categories

Scores & Data Analysis

Score	1	2	3	4	5
Gender Distribution Data	—	—	—	—	—

Calculations are TBD...

Scores & Data Analysis

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

05 Topic Relevance

Problem

- Pandemic essentials/preventatives out of reach
- Certain individuals have an overflowing access to resources
- Higher confirmed cases in areas where preventative resources ran thin

Solution

- Recognize the correlation
- Create an evenly distributed access to inelastic/essential resources

Newest Updates

06

What has changed...

Databases

- Updated to a new database which encompasses a holistic view on racial and sex distribution based on each US county.

Tasks

- all databases have been cleaned
- prepare for analysis

Process

- Fleshed out various tasks necessary to complete project

Topic Broadness

- Realistic application to all citizens
- Correlation with socioeconomic status and COVID-19 cases

07 Conclusion

Q&A