June 8th, 2023

Low-Income Jobs Lost To Covid-19

Final Presentation

By: Yarra Abozaed, Zach Johansen,
Janelle Acob, Tim Chen
DAT 4500

Table of contents

01 Original Data Set

ORIGINAL DATASET – WHAT IT IS

- Our main dataset focuses on the Estimated Low-Income Jobs Lost to COVID-19. The main variables in the dataset are the state names, proportion of low-income jobs lost, total job loss index, and the total of low-income workers employed per state.
- The dataset also includes low-income jobs lost per industry for 20 different industries. This allows us to compare the job loss index by state and by county across the country and see how they're different to each other.

X000	TotalJobLossIndex
X01	Agriculture
X02	Mining
X03	Utilities
X04	Construction
X05	Manufacturing
X06	Wholesale_Trade
X07	Retail_Trade
X08	Transportation_Warehousing
X09	Information
X10	Finance_Insurance
X11	RealEstate
X12	STEM
X13	CorpManagement
X14	WasteManagement
X15	Education
X16	Health_SocialAssistance
X17	Arts
X18	Accommodations_FoodService
X19	OtherServices
X20	PublicAdmin

ORIGINAL DATASET – COLLECTION PROCESS (pt. 1)

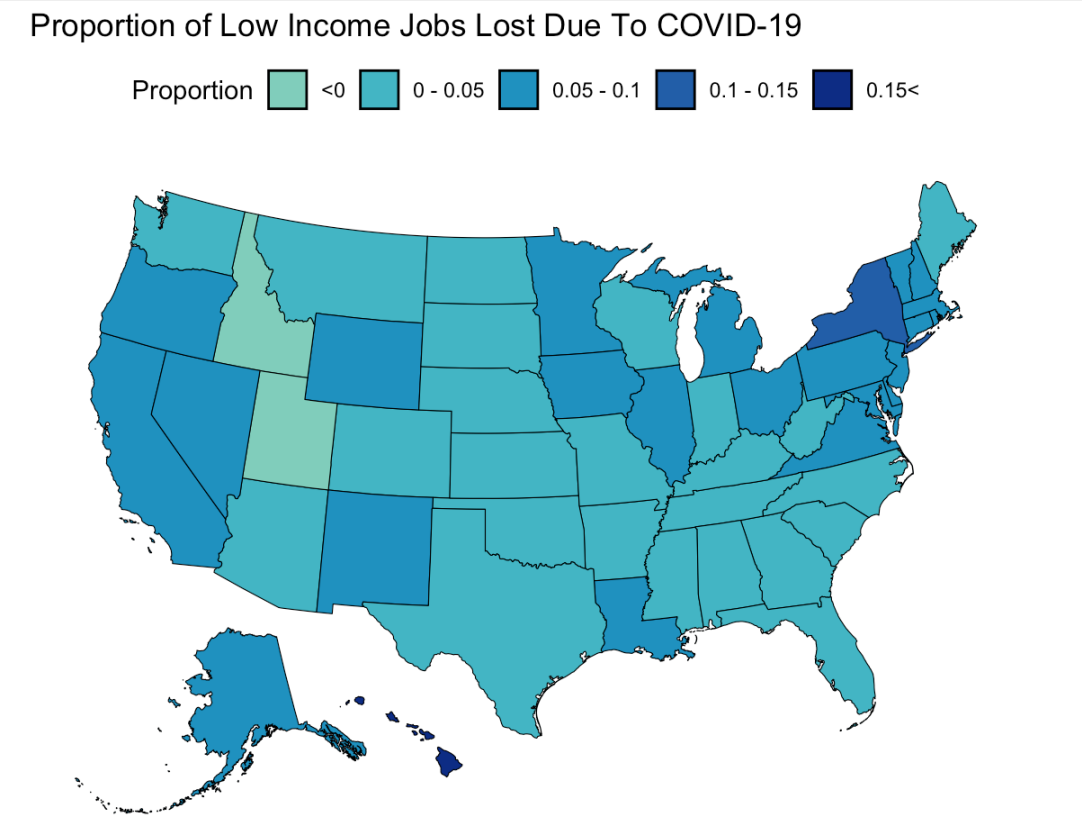
The Urban Institute data science team used data from the US Bureau of Labor Statistics, IPUMS 2012-2018 ACS microdata, and Urban's 2018 Census LODES data, in order to estimate the number of low-income jobs lost because of COVID-19 by industry for every census tract in the state. Low-income jobs are those with (<$40,000 salary).

ORIGINAL DATASET – COLLECTION PROCESS (pt. 2)

To estimate the number of low-income jobs lost:

- Independent contractors are removed from all datasets.
- The BLS current employment and job-loss data + ACS 2014-2018 microdata are used.
- The BLS data (which is on a state level) is adjusted to the ACS data (on a neighborhood level) to estimate job loss on a smaller (per neighborhood) scale.
- The new adjusted data is further adjusted to only low-income jobs lost in each census tract (areas smaller than a neighborhood).

ORIGINAL DATASET – GRAPH



Highlights:

- Hawaii has the highest proportion of low-income jobs lost, second is New York
- Utah and Idaho both gained low-income jobs

02

Research Question

OUR RESEARCH QUESTION

How do different states respond to the COVID-19 pandemic and how does that affect the workforce, specifically low-income employees?

RESEARCH GOALS

- **Create a mathematical model to predict the proportion of low-income jobs lost via a set of metrics chosen by our team.**

Approaches to Answering the Question:

- Seeking external datasets that may correlate with low-income job loss
- Create a metric that estimates low-income job loss using the correlated variables
- Interpret the metric and what it tells us about the relationship between low-income job loss and the external values

03 External Data Sets

EXTERNAL DATASETS – Purpose

We researched potential datasets that we thought could be used to model low-income job loss.

These variables can be broken down into three categories:

- **COVID-19 related metrics**
- **Economic metrics**
- **Political metrics**

EXTERNAL DATASETS – Length of Masking Requirement

The data was obtained from the Ballotpedia website, where researchers gathered information from each state's local website. The dataset was last updated on April 30th, 2022.

Source's Collection Process

The "Length of Masking Requirement" was generated by calculating the number of days between the end and start of each mandate.

Variable Used

EXTERNAL DATASETS – Gathering Restrictions

This variable was derived from the gathering bans dataset obtained from the Health Data Gov. website. This dataset provides information on gathering restrictions at the county level in each state.

Source's Collection Process

The "Gathering Restrictions" variable was generated by finding the county in each state with the longest gathering restriction length.

Variable Used

EXTERNAL DATASETS – Healthcare Rank

Data was obtained from the U.S. News 2021 rankings. This dataset provides a ranking from 1 to 50 for each state based on various factors, including health care access, healthcare quality, and public health.

Source's Collection Process

This variable provides a rank for each state relative to others. Note that a better healthcare rank is higher.

Variable Used

EXTERNAL DATASETS – CDC COVID DATA

This data was collected by the CDC in which it highlights the covid-19 vaccinations distribution per state. This dataset was last updated on May 10th, 2023.

The "Rate of Vaccinations" was generated by using the amount of Doses Admin by jurisdiction per 100k of the pop in each state.

The "Length of Vaccine Availability" was generated by looking at the Percent Total Pop with Completed Series (fully vaccinated).

Source's Collection Process

Variable Used

EXTERNAL DATASETS – Rate of Jobs That Went Remote

This data was collected by the U.S. Census Bureau's Household Pulse Survey in 2021. It includes the ranks of each state where most people worked remotely due to COVID-19.

Source's Collection Process

The variable "Rate of Jobs that Went Remote" was created by looking at the ranking of Proportion of Remote Workers in each state.

Variable Used

EXTERNAL DATASETS – Average Salary

Data was gathered by the Bureau of Labor Statistics, and was released in September 2022.

Average salary per state represents the mean wage of private industries per United States employee in 2021.

Source's Collection Process

Variable Used

EXTERNAL DATASETS – Proportion of Democratic Voters

Data was gathered by editors and reporters who work for the Cook Political Report with Amy Walter.

Represents the proportion of democratic votes in each state relative to the total votes in each state.

Source's Collection Process

Variable Used

EXTERNAL DATASETS – Summary

Variable	Source
Rate of Vaccination	CDC
Length of Vaccine Availability	CDC
Length of Masking Requirement	Ballotpedia Org.
Gathering Restrictions (Max Length)	Health Data Gov.
Average Salary (per state, 2021)	Statista
Healthcare Rank (2021)	U.S. News
Proportion of Democratic Voters (2020)	Cook Political
Rate of Jobs that went Remote	Team Flow HQ

04 Hypothesis Testing

HYPOTHESIS TESTING – What Is It?

Hypothesis testing is a method which confirms a statistical correlation exists between two variables.

In all the tests conducted, we utilized an alpha value of 0.05 as the threshold for determining statistical significance, assuming that a correlation exists between the variables.

HYPOTHESIS TESTING – Why Did We Use It?

In our analysis, hypothesis testing was employed to assess the existence of a correlation between the proportion of low-income jobs lost and each variable under consideration.

We then later selected the highest correlated variables, and used them in our principle component analysis and linear regression models.

HYPOTHESIS TESTING – Results

Variable	P-Value	Correlation Coefficient	Relationship With PLIJs Lost
Rate of Vaccination	1.497×10^{-6}	0.57	Moderate & Positive (PLIJs Lost ↑, Rate of Vaccination ↑)
Length of Vaccine Availability	6.24×10^{-6}	0.60	Moderate & Positive (PLIJs Lost ↑, Length of Vaccine Availability ↑)
Length of Masking Requirement	1.929×10^{-7}	0.70	Strong & Positive (PLIJs Lost ↑, Length of Requirements ↑)
Gathering Restrictions (Max Length)	0.0009606	0.45	Moderate & Positive (PLIJs Lost ↑, Length of Gathering Restrictions ↑)
Healthcare Rank (2021)	0.001125	-0.50	Moderate & Negative (PLIJs Lost ↑, Healthcare rank ↑)
Average Salary (per state, 2021)	0.00113	0.45	Moderate & Positive (PLIJs Lost ↑, Average Salary ↑)
Rate of Jobs that went Remote	0.04745	0.29	Weak & Positive (PLIJs Lost ↑, Rate of Jobs that went Remote ↑)
Proportion of Democratic Voters (2020)	2.97×10^{-8}	0.68	Strong & Negative (PLIJs Lost ↑, Prop. of Democratic Voters ↑)

05

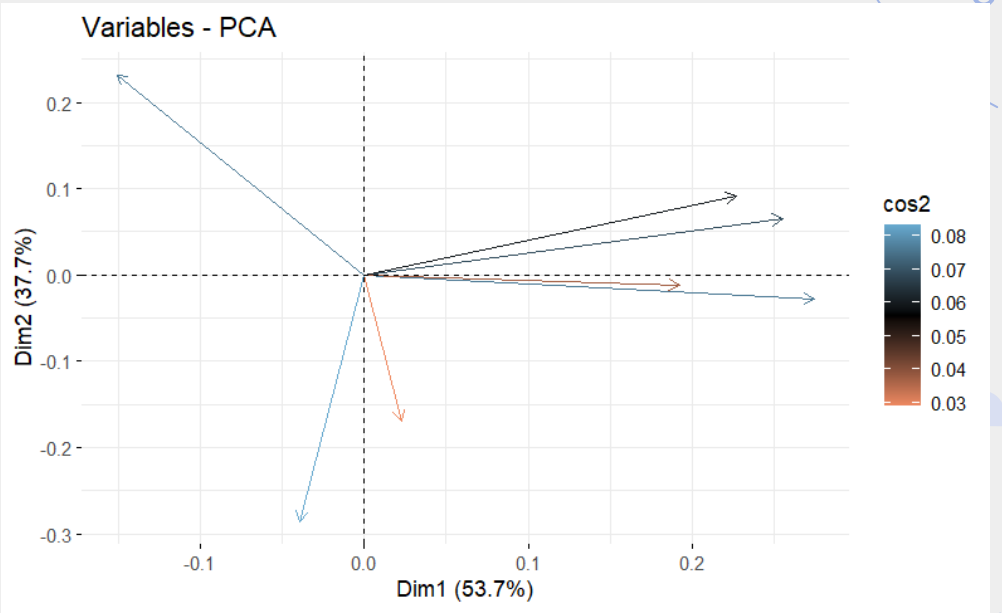
PCA ANALYSIS

PCA – What is PCA?

- Gives completely independent variables
- Grants smaller amounts of variables
- Allows for neater interpretation and understanding

PCA – Why Did We Use It?

- Allows for a simpler understanding of our predictors
- Most of the variability from the predictors is present
- Created more interpretable data than a six figure model



PCA – Results (C1 and C2)

Interpretations

	Component 1	Component 2
Health Care Rank 2021	0.50628620	0.15467635
Doses of the COVID-19 Vaccine Administered	0.54527833	-0.06696825
Proportion of Remote Workers	0.45113703	0.21801794
Total Days of Mask Mandate	0.04482315	-0.40083484
Annual Income 2021	-0.29881730	0.55074145
Proportion of Democratic Voters in the 2020 Election	0.38128874	-0.02809592
Gathering Size Ban	-0.07835304	-0.67770537

PCA – Linear Regression vs PCA R^2 (and why PCA is better)

	Linear Model with all possible predictors	Linear model with only the statistically significant predictors	Principal Component Analysis
Statistical Significance	>0.5	<0.5	<0.5
R^2_adj	0.5257	0.5466	0.3871
Correlation Coefficients	0.3493	0.6309	0

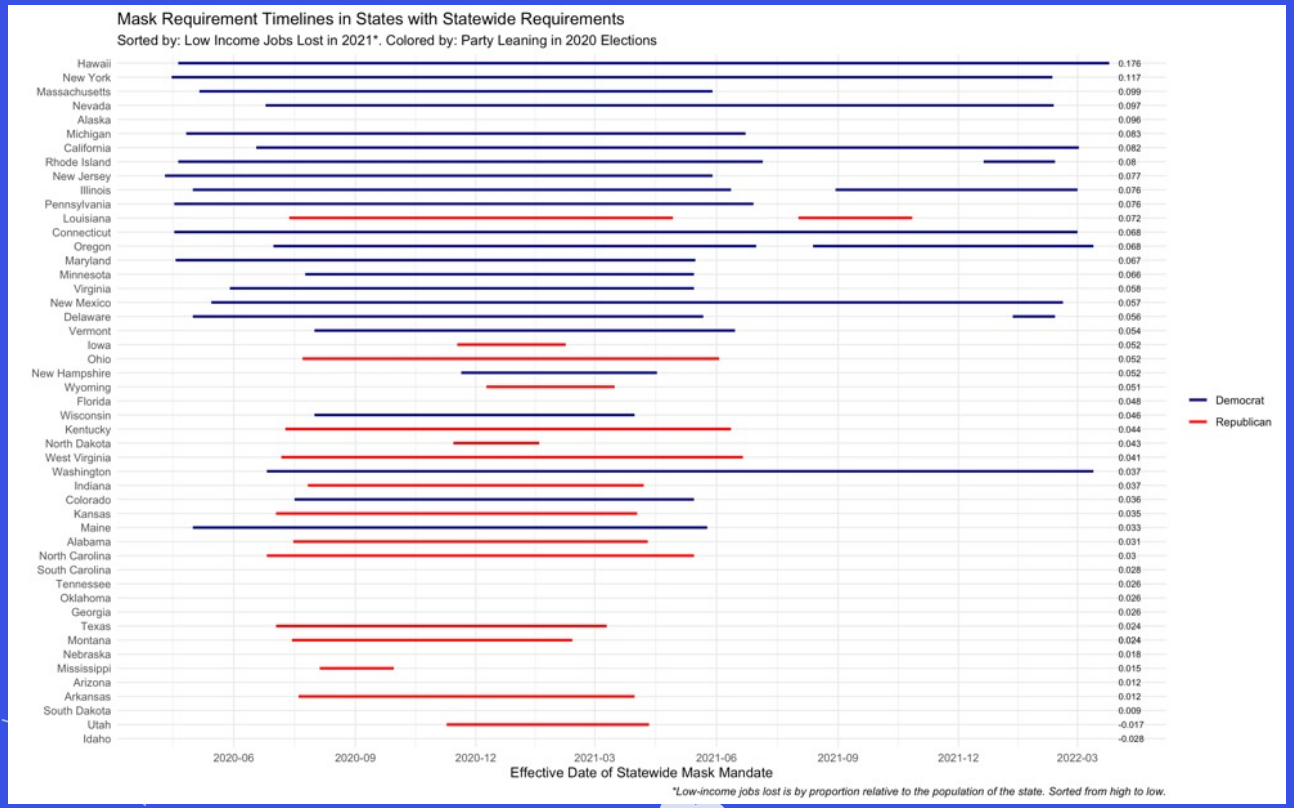
06 Final Discussion

DISCUSSION – Final Model

LowIncomeJobsLost = 0.049508 + 0.009850 × Component1 + 0.007802 × Component 2 + *error*

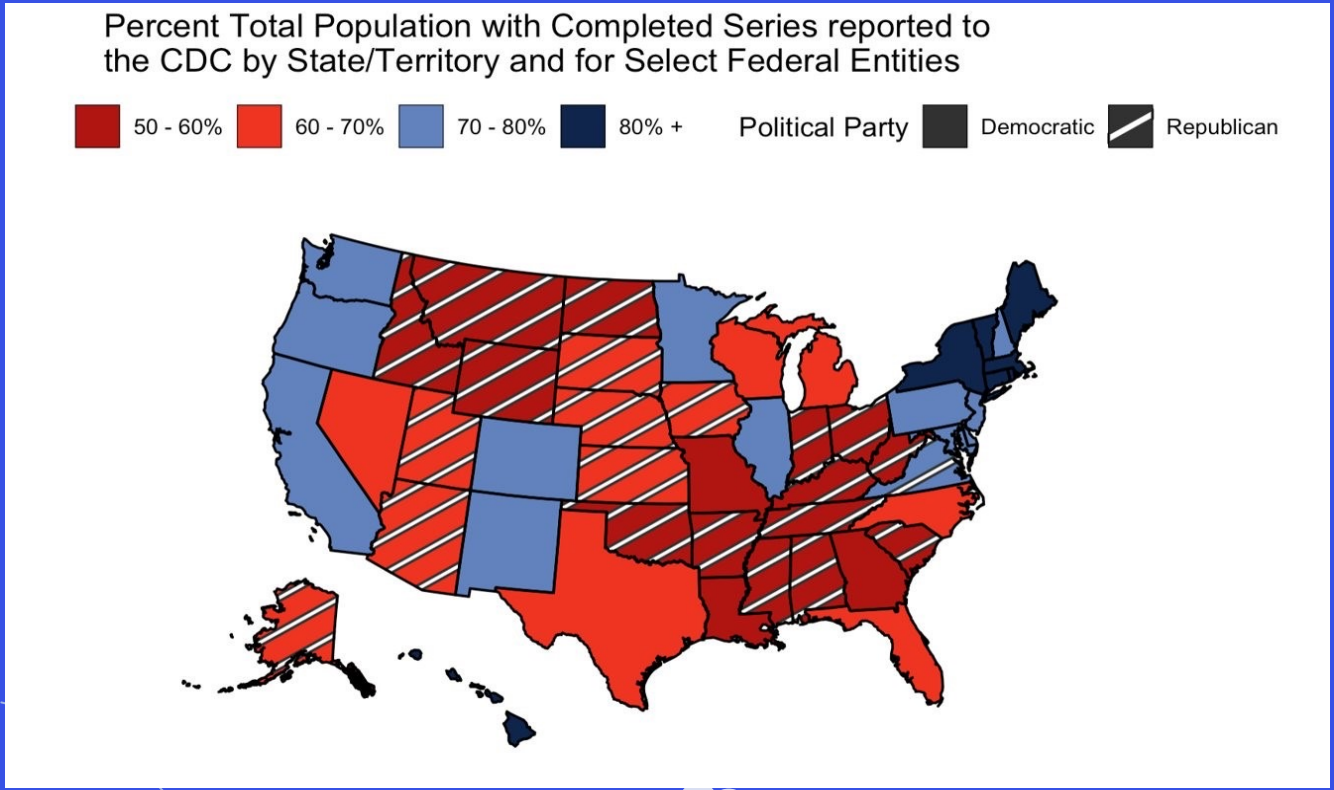
Health Care Ranking	0.50628620 × 0.009850 + 0.15467635 × 0.007802 = **0.006193704**
Doses of the COVID-19 Vaccine Administered	0.54527833 × 0.009850 - 0.06696825 × 0.007802 = **0.004848505**
Proportion of Workers that went Remote	0.45113703 × 0.009850 + 0.21801794 × 0.007802 = **0.006144676**
Mask Mandate Length	0.04482315 × 0.009850 - 0.40083484 × 0.007802 = **-0.002685805**
Average Annual Income	0.38128874 × 0.009850 - 0.02809592 × 0.007802 = **0.00353649**
Proportion of Democratic Voter	-0.29881730 × 0.009850 + 0.55074145 × 0.007802 = **0.001353534**
Gathering Size Ban	-0.29881730 × 0.009850 + 0.55074145 × 0.007802 = **-0.006059235**

DISCUSSION – The Bigger Picture (pt. 1)



States that did more to mitigate the COVID-19 pandemic lost the most low-income jobs.

DISCUSSION – The Bigger Picture (pt. 2)



Democratic-leaning states were the most proactive during the pandemic thus losing the most jobs.

DISCUSSION – So What?

First Component-Mainly used to show the effects of health care ranking, doses of the COVID-19 vaccine administered, proportion of remote workers, and proportion of Democratic voters.

Second Component - Mainly used as a way to pull the data back from over predicting our variables.

PCA

Democratic-leaning states did the most to mitigate the pandemic.

Political-leaning and COVID-19 mitigation are strongly correlated with each other and with low-income jobs lost.

Democratic-leaning states did the most to mitigate the pandemic and lost the most low-income jobs.

Hypothesis Testing

GitHub Pages

https://yarrabozaed.github.io
/Data-and-Society/

GitHub Repo

https://github.com/Yarrabozaed/Data-and-Society

List of References

Variable	Dataset
Length of masking requirement	https://ballotpedia.org/State-level_mask_requirements_in_response_to_the_coronavirus_(COVID-19)_pandemic,_2020-2022
Healthcare rank (2021)	https://www.usnews.com/news/best-states/rankings/health-care
Gathering Restrictions (Max Length)	https://healthdata.gov/dataset/U-S-State-and-Territorial-Gathering-Bans-March-11-/8tfm-md2h/data?no_mobile=true
Rate of Vaccination	https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-total-admin-count-total
Length of Vaccine Availability	https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-total-admin-count-total
Rate of Jobs that went Remote	https://www.teamflowhq.com/blog/states-where-the-most-people-worked-remote-because-of-covid-19
Political Standing	https://ballotpedia.org/Election_results,_2020:_State_trifectas_and_the_2020_presidential_vote
Average Salary (per state, 2021)	https://www.statista.com/statistics/243850/private-industry-wages-per-employee-in-the-us-by-state/
Proportion of Democratic Voters (2020)	https://www.cookpolitical.com/2020-national-popular-vote-tracker
Low-Income Jobs Lost Due to COVID-19	https://datacatalog.urban.org/dataset/estimated-low-income-jobs-lost-covid-19https://datacatalog.urban.org/dataset/estimated-low-income-jobs-lost-covid-19

QUESTIONS?

Template Credits

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution