

# MIT COVID-19 Datathon

// Is Social Distancing a Privilege? //



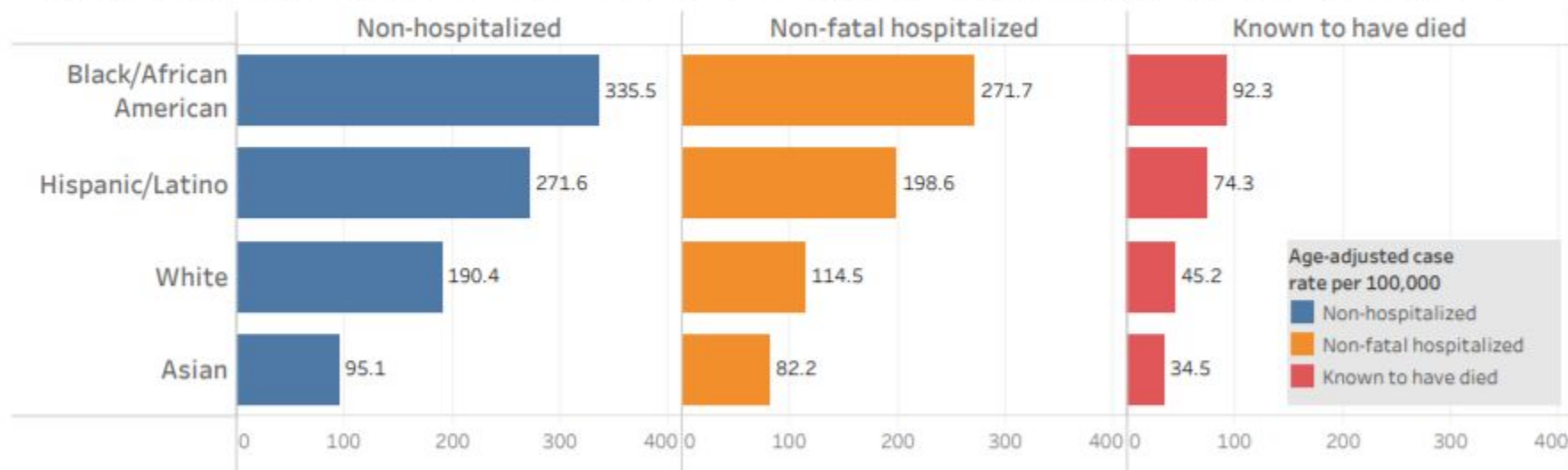
Track C: Disparities in COVID-19 Health Outcomes  
Team 006

## Research Question

**Exploring the Relationship between  
Socioeconomic Status and Social  
Distancing in the United States**

# Disparities in Health Outcomes

Age-adjusted rates of lab confirmed COVID-19 non hospitalized cases, estimated non-fatal hospitalized cases, and patients known to have died 100,000 by race/ethnicity group as of April 16, 2020



[Source: NYC Health](#)

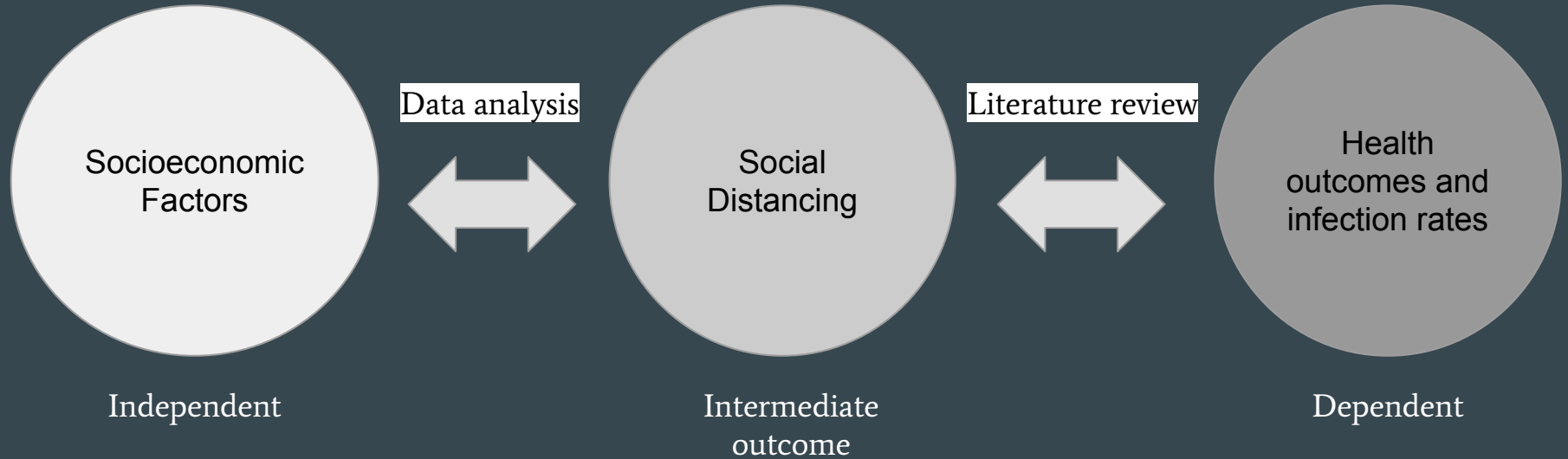
# Public Policy Significance

If social distancing is failing to protect higher risk populations, then public policy solutions are called for.

**Hypothesis: The capacity for social distancing varies by socioeconomic status and demographic factors.**

- Where it is established in the literature that social distancing affects infection rates.

# Understanding Why - the Logic



# Methodology

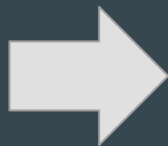
## Social Distancing Index

- A means of quantifying social distancing

Socioeconomic status data  
(ability)

+

Mobility data (behavior)



## Variance in SDI

- Use multivariate regression:

The importance of different socioeconomic factors with regard to one's capacity to socially distance.

# Social Distancing Indices (SDI)

Mobility (behavior)

The logo for unacast, featuring the word "unacast" in a lowercase, sans-serif font. The "u" and "a" are white, while "necast" is black. A small orange dot is positioned to the right of the "t".

+

Socioeconomic (ability)

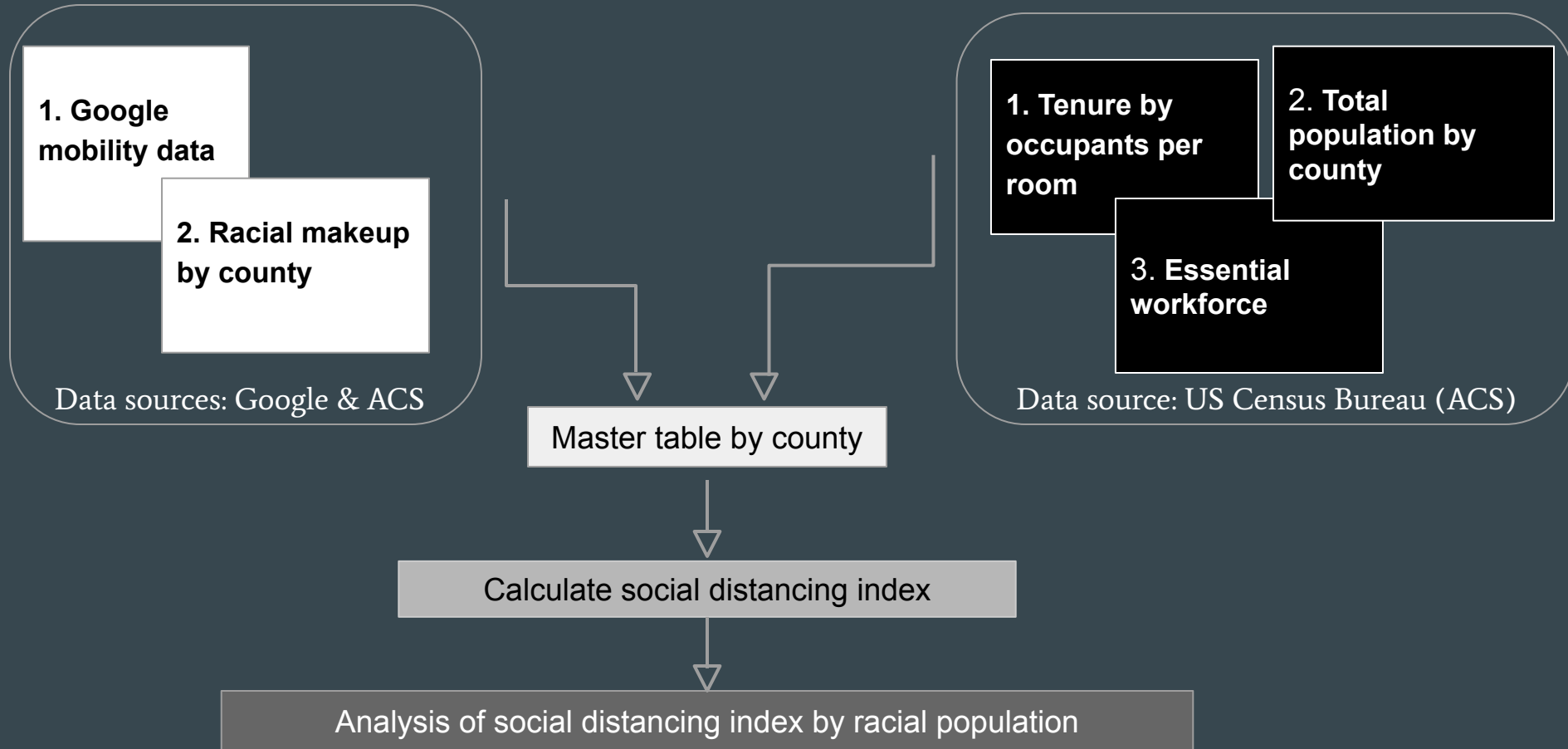


1. **Change in average distance traveled**  
compared to a pre-COVID-19 period (% change over time)
2. **Change in visitation to non-essential venues**  
compared to a pre-COVID-19 period (% change over time )
3. **Encounter Density**  
(% of baseline aka ratio)

1. **Homes: Overcrowded housing**  
(% of residents living in homes with more than one person per room)
2. **Neighborhoods: Population density**  
(residents per square mile)
3. **Workplaces: Essential jobs**  
(% of residents who work in one of 10 job categories).



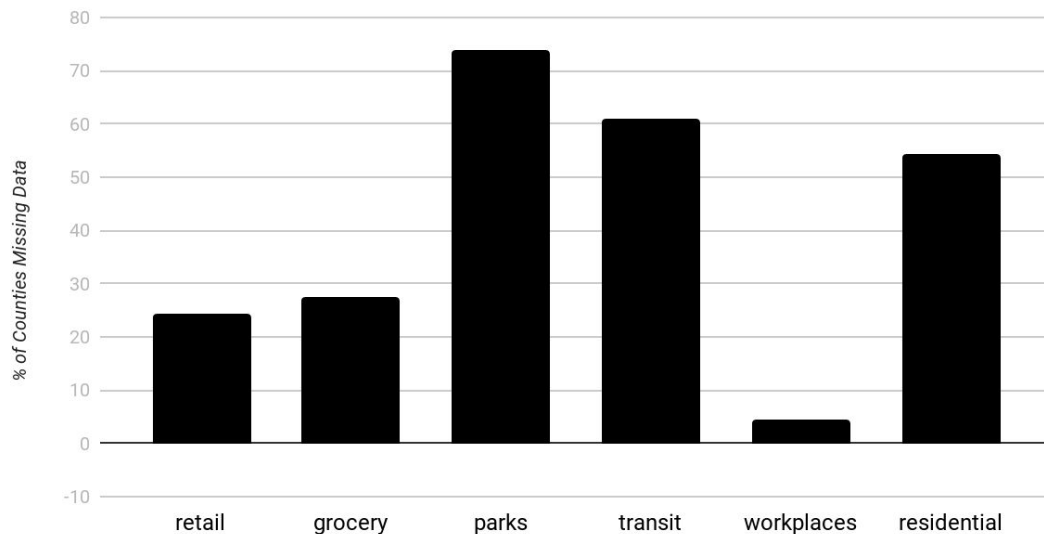
# Data Processing - Reproducing A New SDI



# Data Cleaning

- “Parks” and “transit” were removed from our analysis
- Data was missing at random (MAR)
- Missing data in the mobility report was dropped
- 85% of the original data was kept in the final analysis

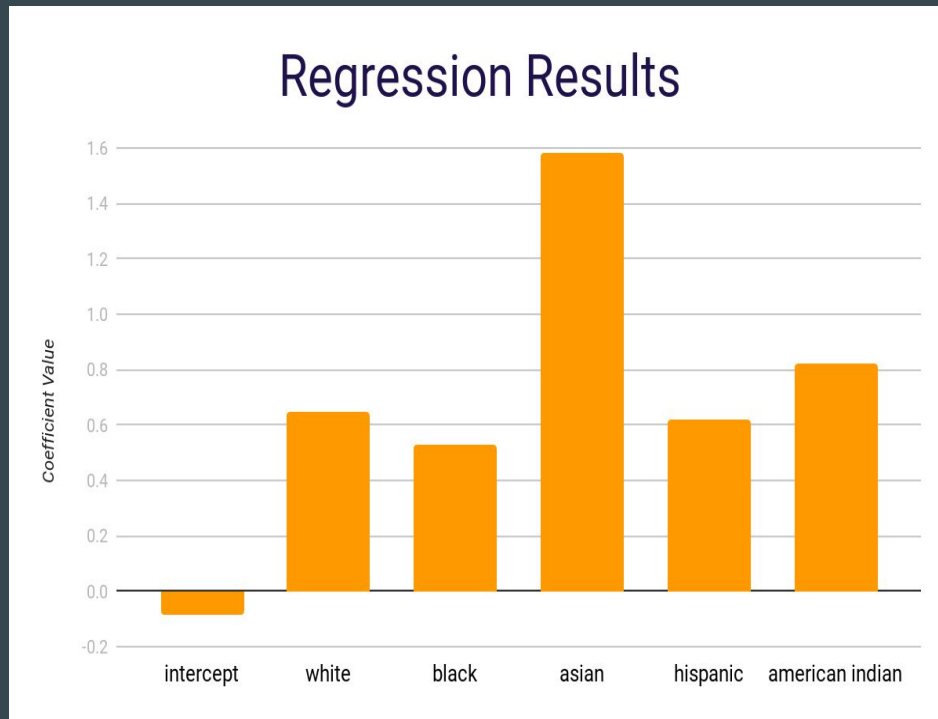
## Missing Data by Feature



# Result - Multivariate Regression

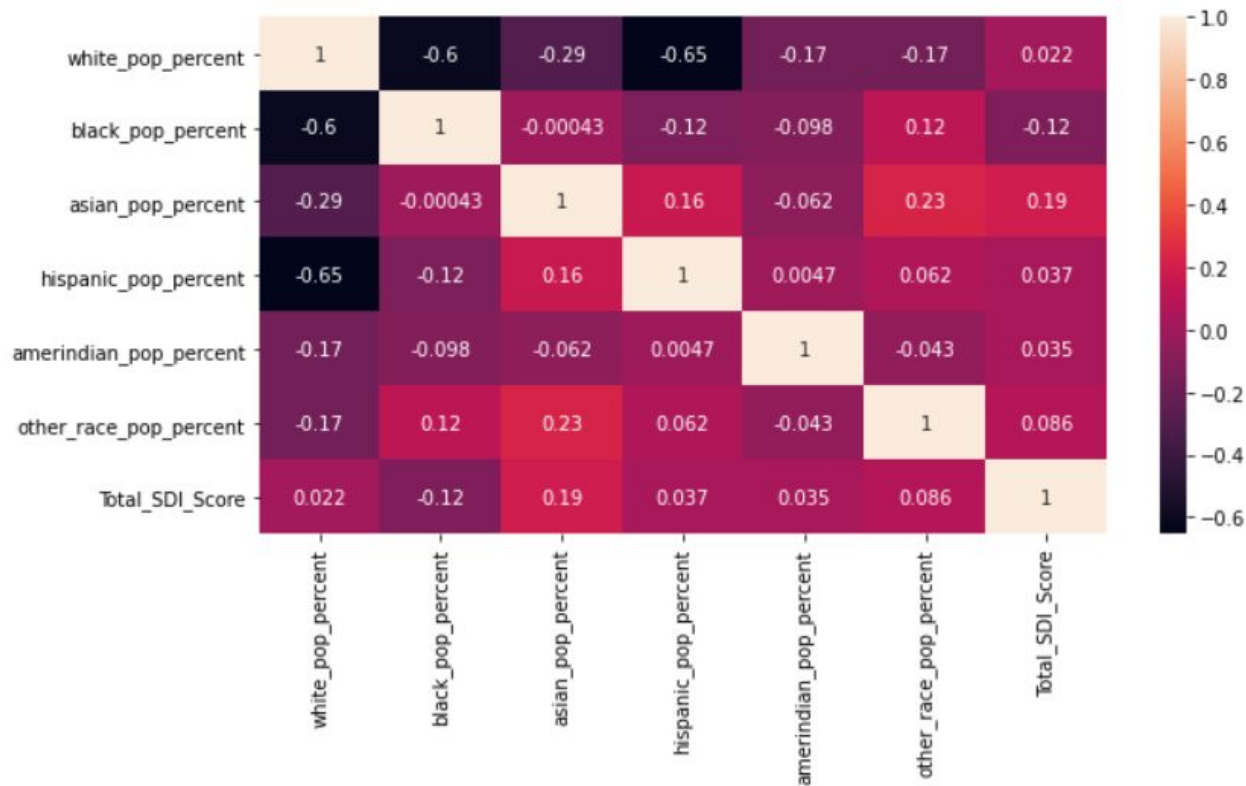
What our analysis showed:

- Black population at greatest risk
- Hispanic
- White
- Asian



# Result - Correlation Matrix

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f65a30a2d30>



# Future Research

## **Data Collection:**

Additional data  
enrichment around  
mobile location data

## **Statistical tools:**

Decomposition  
Regression Analysis

# Challenges Facing Different Subpopulations

- Nearly 25% of currently employed Hispanic and Black or African American workers are employed in service industry jobs, compared to 16% of non-Hispanic whites.
- Black or African Americans make up 12% of all employed workers, but account for 30% of licensed practical and licensed vocational nurses.
- Hispanic workers have lower rates of access to paid leave than white non-Hispanic workers.
- Multi-generational households, which may be more common among some racial and ethnic families, may confront space and caregiving challenges that impede social distancing.
- Those in the lowest income brackets report a significantly lower perceived risk of exposure than those in the highest income groups

# OUR DEEP APPRECIATION FOR

Team MIT

Our Mentors:

Shahzad Amjad Khan

Frank Kuchinski

Andrew Marshall

Arjumand Siddiqi

Michael Williams

The Datathon Sponsors

# Team



Eleanor  
Dehoney



Ian Yen



Mikiko  
Bazely



Tashi T.  
Gurung



Shreyansh  
Bardia



Muhammad  
Jarir Kanji



# Google Colab

<https://colab.research.google.com/drive/11I-IQAxkhCS4pCam3u2YMQ3DuDN-eteY?usp=sharing#scrollTo=bUwkeIR3hU7I>

# Github Repository

[github.com/TashiNyangmi/MIT-Challenge-2020](https://github.com/TashiNyangmi/MIT-Challenge-2020)

# References

1. Kumar, S., Piper, K., Galloway, D. D., Hadler, J. L., & Grefenstette, J. J. (2015). Is population structure sufficient to generate area-level inequalities in influenza rates? An examination using agent-based models. *BMC Public Health*, 15(1), 947.
2. Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual review of sociology*, 35, 553-572.
3. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly*, 2(8), 113-122.
4. Wagner, A. B., Hill, E. L., Ryan, S. E., Sun, Z., Deng, G., Bhadane, S., ... & Acharya, J. (2020). Social Distancing Has Merely Stabilized COVID-19 in the US. *medRxiv*.
5. Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., ... & Ajelli, M. (2020). Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*.

*Special credits to Unacast and Colorado Health Institute for the information about their SDIs*