# 

[**Timeline & Next Steps:**](#_775hpv283by4) **1**

[**Hackathon Resources:**](#_71h04k788yq0) **2**

[**MIT Datathon — Team C-006 Research Proposal**](#_68idywms210a) **3**

[**Potential analyses:**](#_rofkns2b5ql9) **3**

[**Hypotheses: Different social groups have different abilities to socially distance.**](#_y3rh11h531vc) **4**

[**Dataset Reviews:**](#_d6nyr2gx3egu) **5**

[**Interesting Research Links:**](#_ifsmuc50sp17) **10**

[**Methodologies:**](#_adkszeet5clv) **11**

[Colorado COVID-19 Social Distancing Index](#_1vsiqd7hl4o5) 11

[Unacast Social Distancing Index:](#_mf5kvt5ivhm3) 12

# 

# 

# Timeline & Next Steps:

1. ***Due Tuesday (5/12):***
   1. Each member reviews 3-4 datasets & literature
   2. Members add into google doc any other interesting questions/analyses
   3. Based on datasets & descriptions, members rank analyses
   4. Based on final research question, members start exploratory analysis in colab environment (either as R or Python notebook)
2. ***Due Wednesday (5/13) - Midpoint Presentation (6PM ET / 10PM GMT):***
   1. Based on exploratory analysis, members have each picked 1 interesting trend or relationship for presentation
   2. Members collaborate on presentation slide
   3. Members present
3. ***Due Thursday (5/14)***
4. ***Due Friday (5/15)***
   1. Members collaborate on final presentation slide
5. ***Due Saturday (5/16)*** ***- Final presentation (11AM ET / 3PM GMT)***

# 

# 

# 

# Hackathon Resources:

* [Setup Tutorial for BigQuery and Google Cloud](https://www.youtube.com/watch?v=ncVr0pyCb1M&feature=youtu.be) (please save any views under the team\_c006 dataset).
* [List of Datasets on Google Cloud](https://console.cloud.google.com/marketplace/details/bigquery-public-datasets/covid19-dataset-list)
* [Datasets Recommended by the Organisers](https://docs.google.com/spreadsheets/u/2/d/e/2PACX-1vSeiS8eXfuZaN9LWFF0beTHpDat0xKejZ9-Nn7AOKljTjVxK-ONPymyVAj7qe-P6NfdOKkmgT1hvuNd/pubhtml?gid=0&single=true)
* [Shared Google Colab](https://colab.research.google.com/drive/1mL8mmUU1GndIz3OL4-whFc7uTloYxIAl#scrollTo=GW5LcskRHn2g)
* [Conversion between all of our Time Zones](https://www.timeanddate.com/worldclock/converter.html?iso=20200512T020000&p1=tz_pt&p2=tz_et&p3=tz_gmt&p4=130&p5=tz_ist)
* [Research Questions Preference Form](https://tinyurl.com/ybaxj3nu)
* [Participant Guide](https://docs.google.com/presentation/d/e/2PACX-1vTvNgf-vLVPN0KE08WIuVJ15jwWhq2cab7zpXAYIXsGfXLywr4z9mbsD-4iUE3dGATSMYd1RzIYx0GU/pub?start=false&loop=false&delayms=3000&slide=id.g84740f8c38_4_0)
* [Presentation Proposal Document](https://docs.google.com/document/d/1AY0cnKu_IKjg3dlyjEjJR7G00bLCZEa-rpU7YeDNMw8/edit)

# MIT Datathon — Team C-006 Research Proposal

We propose exploring research questions related to both questions 2 and 3: racial and ethnic disparities in U.S. COVID-19 outcomes and pre-existing health disparities in the U.S.

Based on US Census Bureau data, African Americans account for 13.4% of the U.S population. Yet, according to provisional deaths data from the U.S. Centers of Disease Control and Prevention, African Americans account for approximately 22% of U.S. deaths from COVID-19. Provisional deaths data by age group – particularly for ages 35-44 -- show an even greater contrast between the COVID-19 related deaths by race/ethnicity that may be anticipated by Census data alone and what provisional death data show. Among COVID-19 deaths for which race and ethnicity data were available (Source: [NYC Health Date](https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-deaths-race-ethnicity-04162020-1.pdf)), identified death rates among Black/African American persons (92.3 deaths per 100,000 population) and Hispanic/Latino persons (74.3) that were substantially higher than that of white (45.2) or Asian (34.5) persons.

Theories for the overall differential in deaths include a higher prevalence of underlying conditions, disproportionate representation in “essential positions” that involve greater exposure to individuals who are COVID-19 positive, cultural norms that involve closer or prolonged exposure, and disparities in access to healthcare.

These may or may not be explanatory. Other possibilities include differences in the onset, location or modes of treatment; underlying health conditions that have not been identified as high risk factors for COVID-19; differences in exposure to environmental factors or in personal behaviors outside the healthcare arena.

Our project will use hospitalization datasets such as eICU and the CDC’s COVID-NET and, if available, data on socio-economic factors such as employment, to determine whether known health-associated risk factors and socio-economic factors largely explain these differentials, and if possible given the scope of data, to assess whether multiple underlying conditions or any of the underlying conditions individually is significantly predictive of COVID-19 related death. If time and available resources permit, we would also like to explore the relationship between sepsis related to COVID-19 and race/ethnicity, since racial/ethnic disparities in the rate of sepsis are a longstanding issue in health equity and sepsis is one of the COVID-19 related causes of death.

# Potential analyses:

* Difference in time to death from hospitalisation between different racial/ethnic groups.
* Categorization of death cause - Is there differences between different groups?
* Examine why 20-35 year old African Americans are dying at a higher rate?
* What about medicaid and medicare populations? Indicative of income differences? What about also digital adoption and leverage telemedicine?
* Prevalence of a particular social group among frontline workers/essential services vs COVID-19 related health outcomes?
* Geographic factors: density, infra, medical facilities and infection rates

# Hypotheses: Different social groups have different abilities to socially distance.

Disparities in ability to social distance relate to disparities in covid-related outcomes

* Part 1: Social distance index and relationship to infection/mortality (disparities) - assumed it is established in literature
* Part 2: Social distance index by race, occupation, gender, wealth, culture, household size?
  + How do we measure and quantify it
  + Look at how race correlate with the existing social distancing index
* Motivating Question: Is social distancing a privilege?

Additional areas of opportunities: improve the robustness of our model/ social distancing index

* Phases:
  + 1) Figure out equation/mathematical model for existing SDI’s
  + 2) If not available, can proxy using a list of features that we think reasonable
  + 3) Build out SDI/Compute it on a per county/state level.
  + 4) Connect the SDI to race (etc)
  + 5) Connect the SDI to morbidity/mortality rates (if time allows, to prove the robustness of the SDI)
* Next Steps:

|  |  |  |
| --- | --- | --- |
| Deadline | Modeling/EDA | Presentation |
| Wed | (Ian, Mikiko, Tashi)   * Research implementation of Social Distance Index * Compiling a list of features for the Social Distance Index * Figure out which datasets we want to use | (Eleanor, Shreyansh, Muhammad)   * Start a google doc presentation * Start templating out presentation * Finding & summarizing additional literature & citations for hypothesis & similar work   + Existing social distancing indexes, how did we choose the model   + What we know about disparities * Methodologies and approach |
| Thrs | * Finalize features from chosen datasets   + Figure out which features in the Colorado Index features in ACS should be used (BigQuery data) - Ian Ye & Shreynash   + Google mobility - breakout essential vs non-essential travel * Clean & process the data * Create regression model (linear regression model) or dummy index * [Start] Analyze against groups (racial + ethnic) | Check in meeting: 7pm pst/10 pm pst |
| Fri | * [Complete] Analyze against groups (racial + ethnic) * Prepare final notebook * Prepare presentation slide |  |
| Sat | 11Am ET - Presentation   * Rehearsing & final checks on the notebook |  |

* + - County-Level
    - Multiple ethnic/racial groups

# Dataset Reviews:

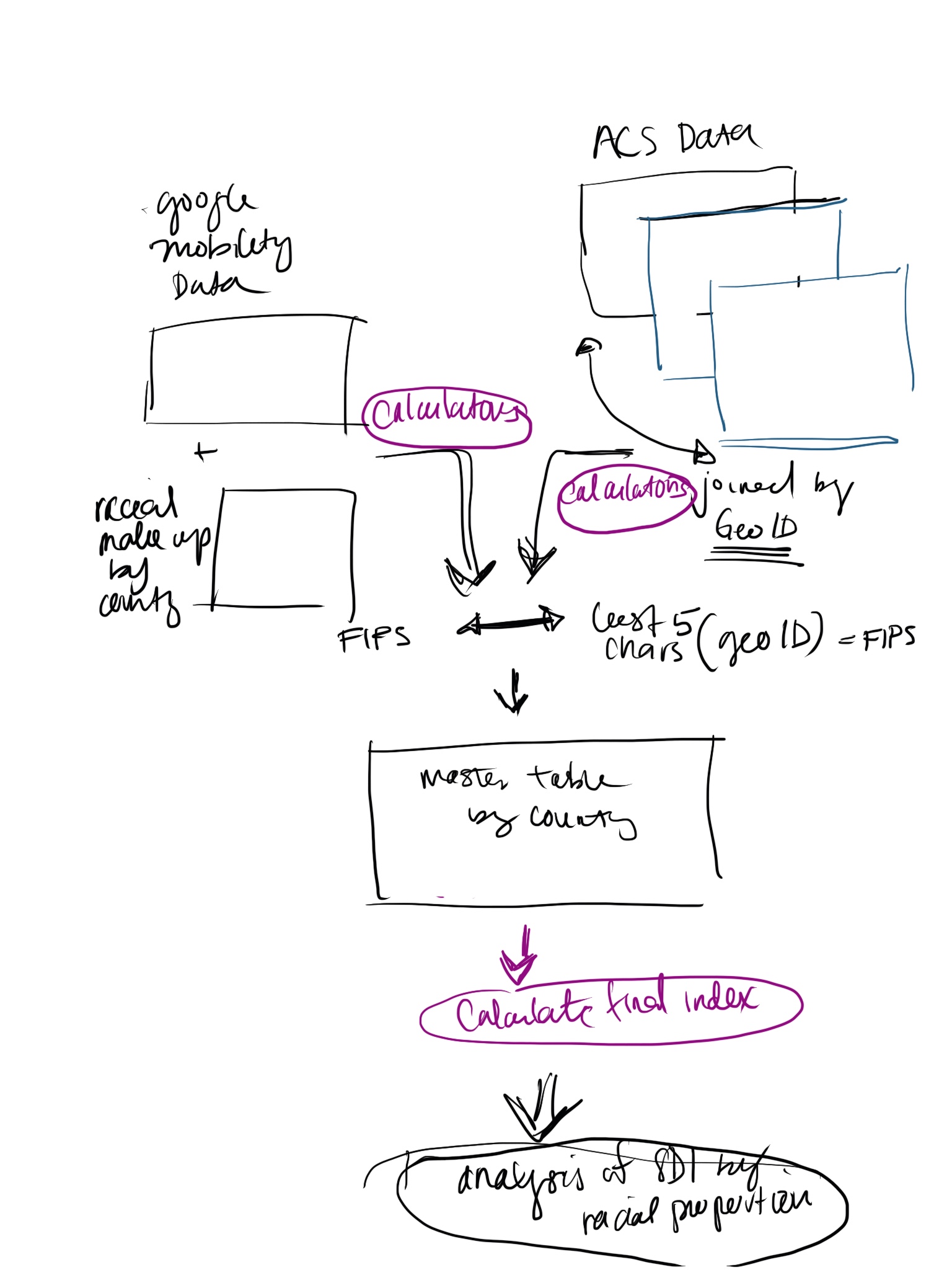
|  |  |  |
| --- | --- | --- |
| **Member** | **Dataset (with link)** | **Description** |
| Mikiko | [GDP and Income by County](https://console.cloud.google.com/marketplace/details/bureau-econonmic-analysis/gdp-income-by-county?filter=category:covid19&id=40a289fb-1a17-4d03-9b31-e6b3d0b5bccc) | GDP and Income by County  Bureau of Economic Analysis  US income and GDP at the county level   * This public dataset was created by the Bureau of Economic Analysis (BEA). It provides a county level view of income, wages, proprietors' income, dividends, interest, rents, and government benefits, including a number of federal and state-level subsidies. * Per capita income can be used to gauge the average financial health and associated social needs of an area. Analysis across regions offers a way to assess relative standard of living and quality of life of the population. Trends analysis of these data over time can also uncover specific regions of economic growth or decline across a variety of indicators. |
| [Health Professional Shortage Areas](https://console.cloud.google.com/marketplace/details/hhs/health-professional-shortage-areas?filter=category:covid19&id=23553ccd-11b9-44c9-bc0c-3c478e4d7ff4&project=example-project-277100&folder&organizationId) | * Health Professional Shortage Areas (HPSAs) are federal designations that **indicate health care provider shortages.** HRSA’s Bureau of Health Workforce (BHW) **develops shortage designation criteria and uses them to decide whether or not a geographic area or population group is a Health Professional Shortage Area (HPSA), Medically Underserved Area (MUA), or Medically Underserved Population (MUP).** * Data on State, Postal Code, Date Record Created |
| [Federal Reserve Economic Data](https://www.quandl.com/data/FRED-Federal-Reserve-Economic-Data) | * Time-series data for the US * Growth, employment, inflation, labor, manufacturing and other US economic statistics from the research department of the Federal Reserve Bank of St. Louis. * Daily Delivery |
| Ian | <https://jamanetwork.com/journals/jama/fullarticle/2766098> | * Racial and ethnic minority populations have a disproportionate burden of underlying comorbidities * Racial/ethnic minorities and poor people in urban settings live in more crowded conditions both by neighborhood and household assessments and are more likely to be employed in public-facing occupations (eg, services and transportation) that would prevent physical distancing. |
| <https://www.ncbi.nlm.nih.gov/pubmed/32390658> | * disproportionate representation of workers from low-income and racial/ethnic minority backgrounds * Lack of decent work for low-income workers who perform "essential" tasks * Work implication of discrimination * Role conflict and stress for women who are managing additional unpaid work, including caretaking responsibilities |
| <https://www.jacionline.org/article/S0091-6749(20)30632-1/pdf> | * air pollution is one of the factors that contributes to the disproportionate impact COVID-19 is having on inner-city racial minorities. |
|  |  |
| Eleanor | <https://covid19researchdatabase.org/> | * Appears to be a comprehensive database, but they need to approve individual registrants before it’s accessible. I’m in the queue, but they may only accept academic researchers. |
| <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/index.html> | * CDC data on provisional deaths by race/ethnicity and age group, hospitalizations, etc. Hospitalization data is pretty limited |
| <https://abc7ny.com/coronavirus-tracking-map-zip-code-nyc-update/6084283/> | Cases by poverty status, if you control for race/ethnicity |
| Tashi | SOCIAL DISTANCING INDEX |  |
| Muhammad | [Google Community Mobility Reports](https://console.cloud.google.com/marketplace/details/bigquery-public-datasets/covid19_google_mobility) | * **County-level** data on changes to different types of **mobility** (parks, grocery, recreational, transit, etc.) * If we can identify counties by race/ethnicity, would allow us to look at how different communities have reacted to the pandemic in terms of mobility and what effect that might have had on health outcomes. * If social distancing is a privilege, can mobility differences between social groups reveal this? Which communities are privileged in this manner?   Changes for each day are compared to a baseline value for that day of the week:  The baseline is the median value, for the corresponding day of the week, during the 5-week period Jan 3–Feb 6, 2020. |
| [American Community Survey](https://console.cloud.google.com/marketplace/details/united-states-census-bureau/acs?filter=solution-type:dataset&id=1282ab4c-78a4-4da5-8af8-cd693fe390ab) | * Detailed **demographic data** (income, race, ethnicities, population, etc.) at different geographic levels (**county, state, congressional district, school district, etc.**)   + Guide: [ACS Information Guide](https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf) * This would allow us to match counties by demographics (e.g., County X has a 70% African American population, and County Y has a 80% Caucasian population, how do their death rates compare?). |
| [7 Social Determinants of Health](https://console.cloud.google.com/marketplace/browse?filter=solution-type:dataset&q=sdoh) | * [Data on Low-Income Housing Developments](https://console.cloud.google.com/marketplace/details/housing-urban-development/lihtc-program?filter=solution-type%3Adataset&q=sdoh&id=689fd9ff-63c4-4fce-ba25-987bfd3ad28e) * [**County-Level** Nutritional Assistance Data](https://console.cloud.google.com/marketplace/details/us-dept-agriculture/snap-enrollment-by-county?filter=solution-type%3Adataset&q=sdoh&id=3c66f347-6858-4865-9684-5237b7735812) * [County-Level Medicare and Medicaid Dual Enrollment](https://console.cloud.google.com/marketplace/details/hhs/dual-enrollment?filter=solution-type:dataset&q=sdoh&id=90483cce-5bf2-4f6c-868a-b8de144a6d4a) |
|  |  |

Colorado (Home (0-10) + Neighborhood (0-10) + % Essential workers (0-10)) = 30 ⇔ Join (county) <=> (Travel 1 (1-5), Travel 2 (1-5), Density )(#/nation average)) = /15 ⇒ ()/45

**Dataset & Features:**

**Thursday**

* Review ACS - try to find relevant features or columns to be used for calculating Colorado SDI (Shreynash, Ian) + racial data (% blacks/whites/asians/non-white hispanics)
  + List here & queuing up queries
* Google Mobility data - Calculating essential vs non-essential (Tashi, Mikiko)
* Race by County - Import into notebook environment (Mikiko)
* Bringing in data & cleaning in the Colab notebook (includes the index & race by county) ⇒ write out a csv/dataset
* Calculate the total index (Muhammad, Mikiko)



# Interesting Research Links:

* [Black people four times more likely to die from Covid-19, ONS finds](https://amp.theguardian.com/world/2020/may/07/black-people-four-times-more-likely-to-die-from-covid-19-ons-finds)
* [Low-paid workers more likely to die from Covid-19 than higher earners](https://www.theguardian.com/world/2020/may/11/manual-workers-likelier-to-die-from-covid-19-than-professionals)
* [More people dying at home during Covid-19 pandemic – UK analysis](https://www.theguardian.com/society/2020/may/08/more-people-dying-at-home-during-covid-19-pandemic-uk-analysis)
* [Key Facts on Health and Health Care by Race and Ethnicity](https://www.kff.org/disparities-policy/report/key-facts-on-health-and-health-care-by-race-and-ethnicity/)
* [Growing Data Underscore that Communities of Color are Being Harder Hit by COVID-19](https://www.kff.org/coronavirus-policy-watch/growing-data-underscore-communities-color-harder-hit-covid-19/)
* [Factors Underlying Racial Disparities in Sepsis Management](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6315577/)
* [Racial Differences in Sepsis Mortality at United States Academic Medical Center-Affiliated Hospitals](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5953774/)
* [Black-white racial disparities in sepsis: a prospective analysis of the REasons for Geographic And Racial Differences in Stroke (REGARDS) cohort](https://ccforum.biomedcentral.com/articles/10.1186/s13054-015-0992-8)

# Methodologies:

## Colorado COVID-19 Social Distancing Index

* Relevant Documentation: [Colorado COVID-19 Social Distancing Index](https://www.coloradohealthinstitute.org/research/colorado-covid-19-social-distancing-index)

CHI’s Social Distancing Index rates every census tract in the state for obstacles to social distancing in our homes, neighborhoods, and workplaces. It is based on three measures:

* Homes: Overcrowded housing (percentage of residents living in homes with more than one person per room) - (percentile relative to state) - against Colorado counties (0-10)
* Neighborhoods: Population density (residents per square mile) (rank percentile) - against Colorado (0-10)
* Workplaces: Essential jobs (percentage of residents who work in one of 10 job categories. See box on the right side of the page.) - (rank percentile) - against Colorado (0-10)

Each census tract received a score on a scale of 1 to 10 based on its comparison to all other tracts in Colorado.

Higher numbers denote greater population densities, crowding, and proportions of essential jobs.

The overall index score for a tract is an average of its scores on the three measures.

Data for all three measures are from the U.S. Census Bureau’s Five-Year American Community Survey Estimates.

Essential Jobs

For this index, workers in the following 10 industries were deemed to have essential jobs:

* Agriculture, forestry, fishing and hunting, mining
* Construction (including electricians, mechanics, and plumbers)
* Manufacturing
* Wholesale trade
* Retail trade
* Transportation, warehousing, and utilities
* Waste management
* Education, health care and social assistance
* Food services
* Other services, including auto repair, child care, banks, and laundries

Additional Notes from Colorado Social Index team:

* The only one of these measures we had to apply estimates to was the essential labor force – for certain industries, we didn’t assume 100 percent of the workforce was “essential” but adjusted that based on occupational detail from BLS that pointed out which types of food service work was “essential" and what was not. Let me know if you want more detail on that.
* Finally, to index the values, we just created a scale of 0 to 10 for each, where 0 meant no challenge (e.g., zero percent of homes are overcrowded) and 10 meant the highest challenge in the state (e.g., the area with the greatest percentage of overcrowded homes). Index scores across the three indicators used to create the index (population density, overcrowding, and essential workforce) were used to come up with a final index score.

Tables & Features

|  |
| --- |
| * Overcrowding: defined as one or more occupant per room. Table ID B25014 * Population density: residents per square mile. This was actually coded in a GIS software. Total population can be taken from Table ID B01003, and square mileage will be included in most mapping software. This is just total population divided by square miles. * Low-income: population with incomes at or below 200 percent of federal poverty level. Table ID B05010 * Older adults: population age 65+. Table ID B01001 * People of color: population that does not identify as white (non-Hispanic / Latinx). Table ID B03002 * Essential labor force: Table ID C24050. The employment types we counted were:   + C24050\_002E   + C24050\_003E   + C24050\_004E   + C24050\_005E   + C24050\_006E (assumed 60% was essential workforce)   + C24050\_007M   + C24050\_011E   + C24050\_012M (assumed 64% was essential workforce)   + C24050\_013E (assumed 40% was essential workforce)   + C24050\_024E (assumed 18% was essential workforce) |

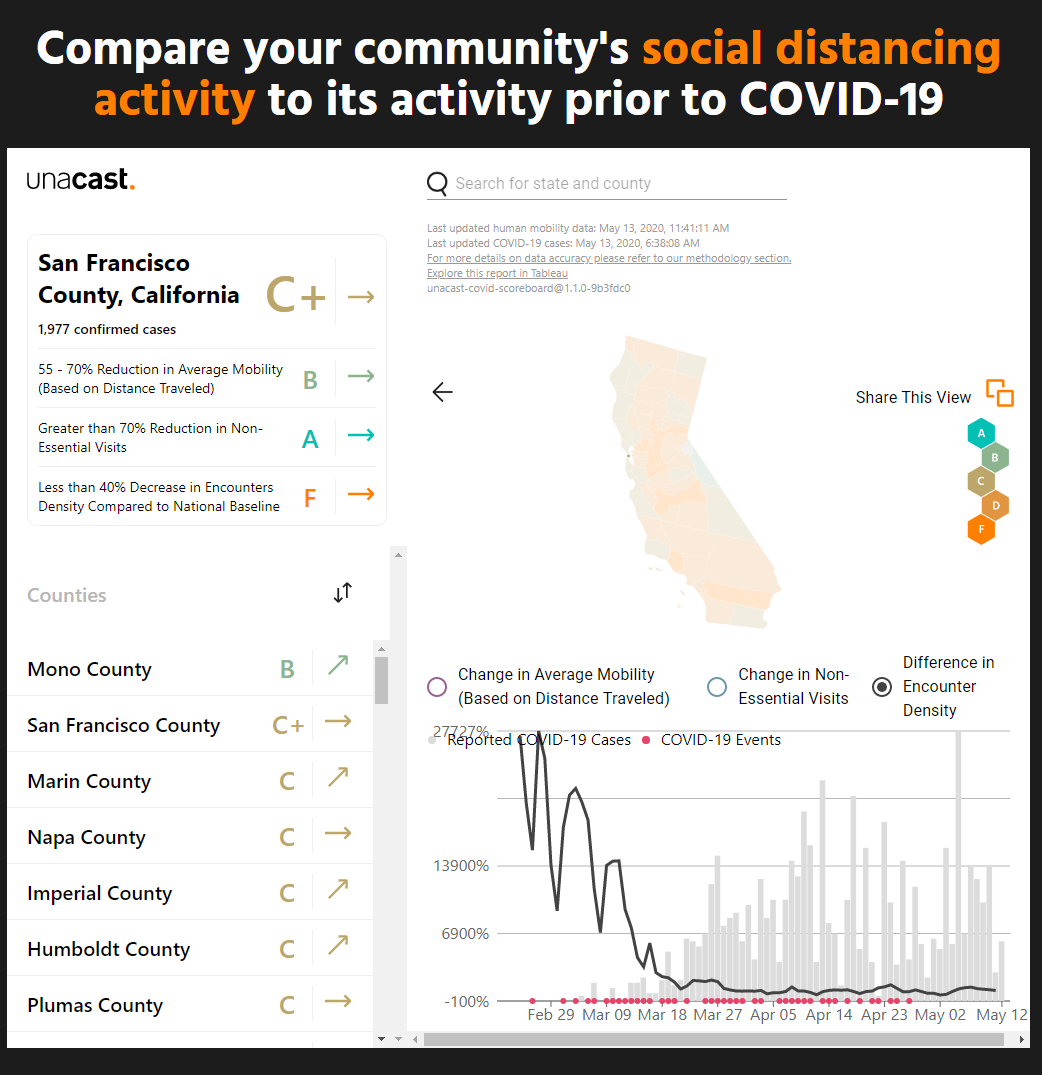
## Unacast Social Distancing Index:

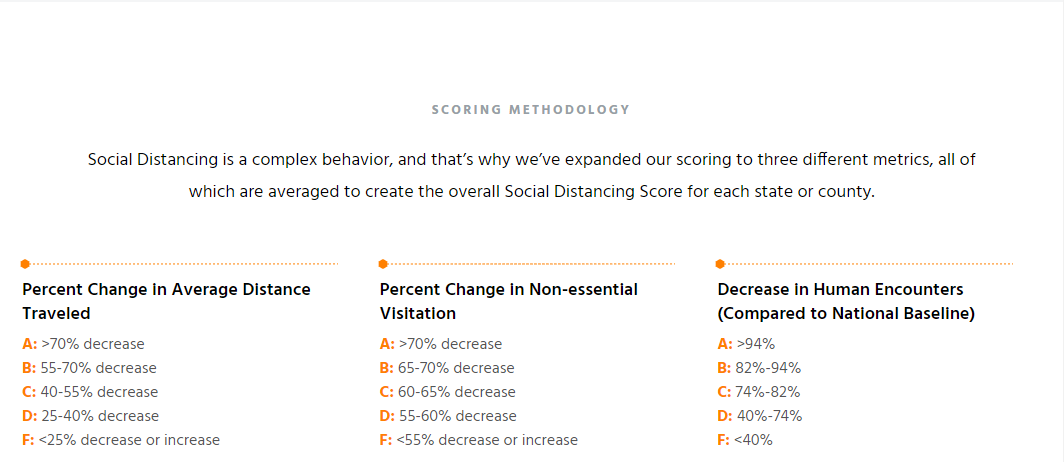
Relevant documentation:

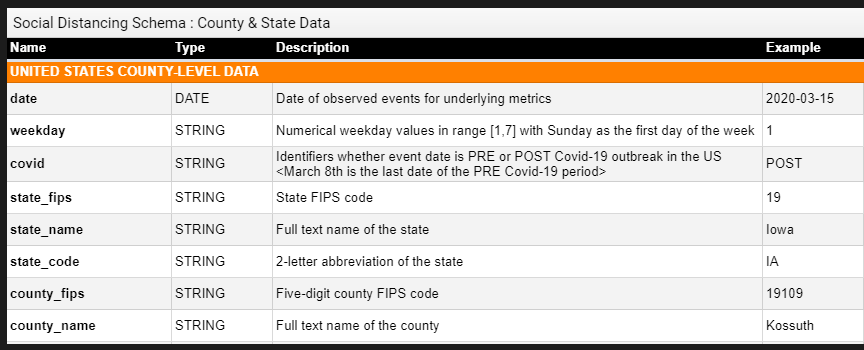
* Index dashboard: [Covid-19 Social Distancing Scoreboard — Unacast](https://www.unacast.com/covid19/social-distancing-scoreboard?view=state&fips=06)
* Index Methodology posts: [The Unacast Social Distancing Scoreboard — Unacast](https://www.unacast.com/post/the-unacast-social-distancing-scoreboard) & [Rounding out the Social Distancing Scoreboard — Unacast](https://www.unacast.com/post/rounding-out-the-social-distancing-scoreboard)
* Metric 1: Change in **average distance traveled** compared to a pre-COVID-19 period (% change) over time - A-F
* Metric 2: Change in **visitation to non-essential venues** compared to a pre-COVID-19 period (% change) over time - A-F
* Metric 3: Encounter Density (% of baseline aka ratio) - A-F

Encounter density - the probability that two devices that were in the same place at the same time

* To account for the likelihood that the people in a given community will contract COVID-19, we created this metric:
  + M= number of encountersarea (km²)/ baseline - 1
* We need to define: an encounter, normalization, the baseline, and a scoring range
  + **Encounter** - "proximity between any two users of the same province who were seen within a circle of radius R = 50m over a 1 hour period."
    - In other words: two devices within 50 meters of each other for 60 minutes or less.
  + **Normalization-** To hold every county and state to the same standard of measurement, we needed to normalize our metric.
    - encounters per square kilometer of land area
  + **Baseline** - intended to represent “business as usual”, is calculated as the national average encounter density during the 4 weeks that immediately precede COVID-19 outbreak (February 10th - March 8th). We use this baseline as a lead to define fixed ranges of encounter density for our grades, which are then expressed as a reduction from the fixed baseline.
    - We leaned on the human-encounter-reduction goals recommended by a variety of experts and studies (such as this one) and create this scoring range:
      * A: >94%
      * B: 82-94%
      * C: 74-82%
      * D: 40-74%
      * F: <40%







# Mentor Feedback:

## Midpoint Presentations

## 

|  |
| --- |
| **Arjumand** |
| 1. Behaviour vs ability vs both  * Rather than dichotmize by essential vs non-essential * To tackle the missing data issue by identifying whether they fit into one of these category (dichotomy score) * Whether we can impute the missing data using multi imputation strategies (we may not want to impute the outcome) * Compare who is social distancing for work who isnt  1. Should we combine the colorado index and the unacast index or only use one of them?   Multiple imputation strategy  Can do this if the index is an input to another equation.  Not if the index is the ultimate output  Everything is a behavior  Do you have the choice or not?  Think in theory whether the activity is essential or not  Different racial groups have different occupations.  -Arjumand Siddiqi (Expert Epidemiologist)  Disparities vs Social Determinatants of health  Disparties are when SDOH are misallocated?  Not about culture, not genetic, experiences of daily life based on racial groups   1. Kind of work you do (Occupation)    1. County level data on occupational data **by race** 2. Symptoms are so general    1. Syptooms dismissed for black people (especially women)    2. Inability to acces those minds 3. Rent vs own (county level) proxies of 4. BASIC regression 5. Decomposition regression : asks a slightly different question 6. - Normal regression: What is the relative cost of not social distancing when it comes to race? 7. - Decomposition: what variables account for the difference in the distribution (diff is it accounts for the absolute gaps between two distributions/ means)    1. You get causal variables    2. Ie. dichotomous variables- social distancing = you do it or dont       1. 70% of whites are able to social distance       2. 30% of blacks are able       3. Decomposition: absolute gap is 30%; how which of which are accounted for differences in density/ occupations       4. Absolute estimates add up to 30% (22% by all the known variables, 8% is unknown); residual becomes intuitive       5. If you give black the characteristics of white, how does their sd change? How much do these variables actually account for the diff?   **Dinardo Fortan,Tomas Lemieux**  Blider, Wahaka  Decomposing dichotomous outcomes  Blinder–Oaxaca decomposition |

|  |
| --- |
| [**Shahzad Amjad Khan**](https://app.slack.com/team/U013C0GJUE6) |
| Good brainstorming session. My apologies for having missed the presentation. Here are my two cents to the discussion:  With large portions (large being generally taken to be >40% ) of data missing, you need to answer the following questions:  1). Does the mechanism of "missingness" depend on observed data only (i.e. "Missing at Random" MAR)? OR  2). Does it depend on both the observed and the missing data ("Missing Completely at Random", MCAR)? OR  3). Is it Missing Not at Random (MNAR, mechanism for missingness only depends on missing data, e.g. only or mostly missing for people expected to have a certain outcome)?  Multiple imputation in large volume missing data scenario would only be valid in the MAR regime  4:37  In the MCAR case, just discard the missing data and analyse for people with complete data only. This will shrink the sample size and broaden confidence bars, but still avoid systematic bias  In the MNAR regime, you can look at both the "best-worst" and "worst-best" case sensitivity analysis and report those with your findings  4:40  Decomposition analysis suggested by @Arjumand Siddiqi [Mentor] [UToronto] sounds interesting, do give it a look. The regression coefficients from such a regression may be used as the  4:41  ...index you are interested in.  Also think about instrumental variables  Are there predictors that cannot be directly measured, can you look at them indirectly through instrumental variable regression  4:45  And finally, I think @Frank Kuchinski [Mentor] [Aerial Dat] makes a good point about "what is your story?" Formulate your question in terms of a story... There's too much analytics taking place around COVID that analysts often lose sight of the story telling aspect of impact!  4:46  Good luck going forward. Do reach out if you need any help clarifying any of the above suggestions. I believe I have already shared my email address with the group. |

|  |
| --- |
| From Michael Williams |
| Overall good presentation. One suggestion. If introducing the team add a slide with names, photographs and current role.  Clear problem statement and hypothesis. It's a little unclear whether the data being used (i.e. travel) has a direct relationship to social distancing. For example, people may not travel to work and may spend their time walking, in grocery stores and waiting for food pick up and choose not to social distance. The identification of data gaps was very clear  From Mentor #2: Looking good! Here are a few notes:  \* Please be careful about describing things as “racial” when they’re really “socioeconomic.” Of course they tend to be highly correlated, but it’s important not to use the shorthand. For instance on your slide about your logic, that first bubble should more accurately say that “Socioeconomic inequities leads to racial disparities in the ability to social distance.” It’s not that people with darker skin are innately less able to stand 6 feet away from one another!  But people with brown skin are disproportionately working in positions and living in areas where they can’t afford to social distance. It’s nitpicky, I know, but very important.  \* Before you use “SDI” as an acronym, you should spell out that it stands for social distancing index! :)  \* This mobility analysis is really fascinating. Have y'all seen the work the NYT did on this a few weeks back? [https://www.nytimes.com/interactive/2020/04/03/us/coronavirus-stay-home-rich-poor.html](https://slack-redir.net/link?url=https%3A%2F%2Fwww.nytimes.com%2Finteractive%2F2020%2F04%2F03%2Fus%2Fcoronavirus-stay-home-rich-poor.html) |

|  |
| --- |
| Laura Rosella |
| -Very interesting question with strong rationale  -Provide more clarity on what variables you will adjust for in the regression and why - some account for the association (mediators) and if you adjust can obscure the main exposure you are testing. Be careful to think through what you are adjusting for an why.  -Have you through through the different options for regression given the distribution of your outcome?  -I agree you can't create a new distancing measure so instead made sure you understand the biases - who does it capture - who is missing and in what direction will that impact the associations.  -Likewise with the mobility data - who is missing from this data and what selection bias will this introduce into your finding.  You may be able to use weighting to account for some of this but unless you know you are at risk of introducing selection biases that can introduce spurious findings, for example if the most deprived are missing form the data  -Good luck! |