Equity in Healthcare

Team 7: Morgan Shepherd, David Baker, Addison Thalhamer, J Isaac Torres M, Nolan Sverdrup

A. Identifying the Vulnerability Factors

After the team analyzed literature related to general socioeconomic factors that can lead to vulnerability against disease outbreaks and pandemics, the top five vulnerability factors chosen are the following:

1. **Age -** Both the elderly and children are particularly vulnerable. In the event of a disaster, "older residents as a group are more likely to lack the physical and economic resources necessary for effective response and recovery and are more likely to suffer health-related consequences." School closures can also lead to developmental issues in children. [1]
2. **Neighborhood & Environment (aka Crowded Conditions) –** One's environment and housing situation is another social determinant of health, with living space and commuting habits being highly correlated with risk. Counties with a high volume of bi-directional commuters were found to have higher rates of viral infection, with public transportation and carpooling leading to the most vulnerability. Housing environment was also found to be correlated with risk, and susceptibility to viral infection was higher with increased housing insecurity and housing density or crowded conditions. A denser population, especially within the same structure, raised potential of exposure. [2][3]
3. **Access to Healthcare –** Whether a person has health insurance, wealth, transportation to health facilities, or any other attributes that prevent people from obtaining medical aid, are the main reasons to change a person's vulnerability to outbreaks and pandemics. In addition to the status of the person, a region's public health infrastructure is an essential attribute to the person's vulnerabilities. If a poor public health infrastructure is present, this will increase the mortality and morbidity of any person regardless of wealth or medical insurance ownership [4].
4. **Socioeconomic Status –** Socioeconomic status has been identified as a contributing factor to communities' vulnerability against disease outbreaks and pandemics. [5] It has been identified that these factors have the potential to exacerbate the effects of an outbreak, decreasing a community's ability to withstand the emergency.
5. **Education –** A person's education level was found to be a major contributor to a community's vulnerability against disease outbreak. Knowing about hygiene methods, protocols to follow during a pandemic (social distancing, quarantine, etc.), or even knowing how viruses affect human health are key roles that affect a person's vulnerability. [4]

B. Consulting US Census Data

This assessment will be looking at 176 tracts, all Denver County tracts except for two tracts that have no residents: tracts 9800.01 and 9801. Then, the team filtered these tracts out in our pre-processing. Denver is a consolidated city *and* county. A summarized table with all the data has been attached in a separate file. However, the following are the attributes that correspond to each of the five factors.

1. **Age**
    a. Population below the age of 5 years old
    b. Population below the age of 18 years old
    c. Population over 65 years old
    d. Median age per tract
2. **Neighborhood & Environment, aka Crowded Conditions**:
    a. Occupied housing units
    b. 1 or less occupants per room
    c. 1.01 to 1.50 occupants per year
    d. 1.51 and more occupants per room
    e. 1-Person household
    f. 2-Person household
    g. 3-Person household
    h. 4-Person household
    i. 5-Person household
    j. 6-Person household
    k. 7-Person household
    l. Total commuters
    m. Private vehicle drove alone
    n. Private vehicle carpooled
    o. Public transportation
    p. No workers in household
    q. 1 Worker
    r. 2 Workers
    s. 3 or more workers
3. **Access to Healthcare**
    a. Population with private health insurance
    b. Population with public health insurance
    c. Percent of households with broadband internet
    d. Population with a disability
4. **Socioeconomic Status**
    a. Count of us citizen data
    b. Not a us citizen
    c. Population 16 years and up
    d. Unemployment rate population 16 years and up
    e. Total housing units

f. Occupied units paying rent
g. Individuals whose poverty status is determined
h. Individuals with income 500% of poverty level or lower
i. Total population (white)

5. **Education**
    a. Enrolled k-12
    b. Enrolled college/ graduate school
    c. Population 25 yr. and higher
    d. 25 Yr. and older no high school diploma
    e. 25 Yr. and older bachelor's degree or higher
    f. 25 Yr. and older high school graduate no bachelor's degree
    g. 25 Yr. and older high school graduate or higher

C. Data Pre-Processing

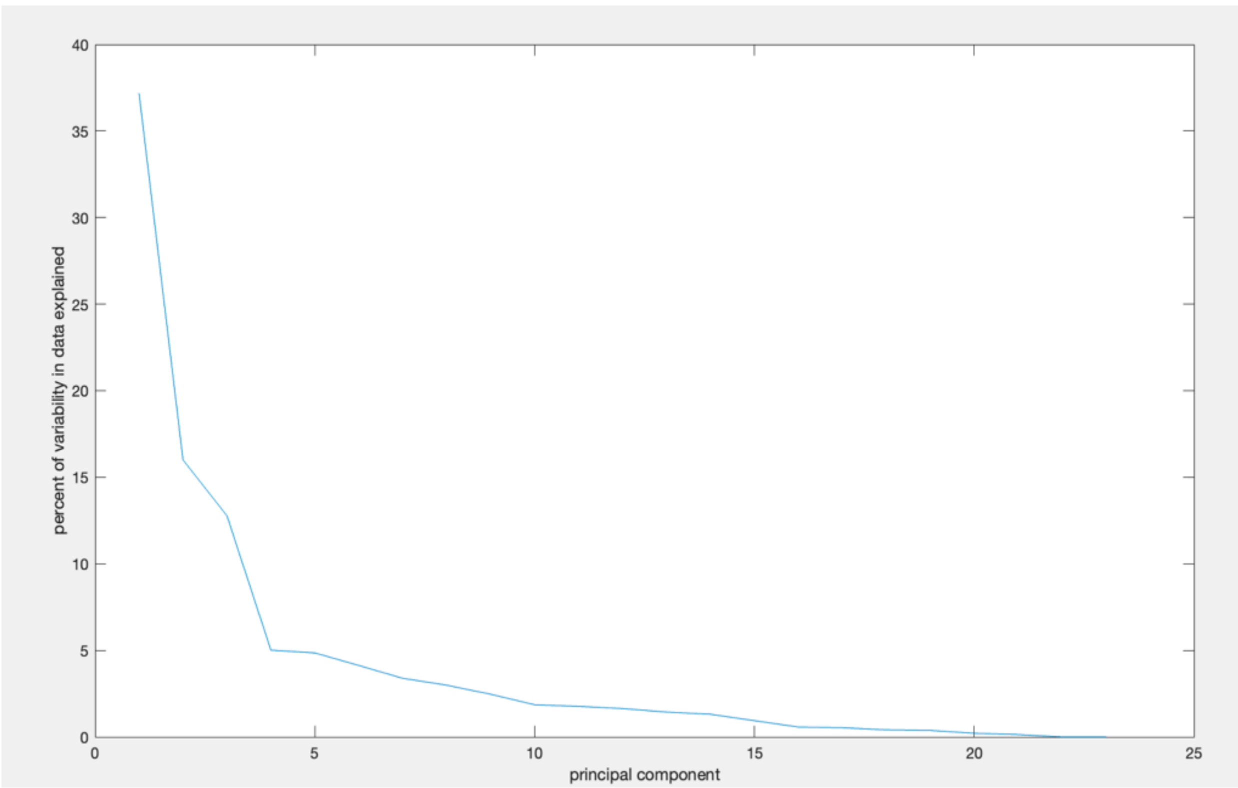
Principal component analysis is sensitive to skewness and scale, so we had to take some steps to prepare our data.
The data prep followed this general procedure:

1. The team converted any per-tract counts to percentages of the total population, so that the size of the tract would not influence analysis.
2. If the data is such that a higher value indicates higher vulnerability, then is left alone. If a higher value indicates lower vulnerability, then multiply by negative 1. This approach was adopted from Cutter [6]. With this transformation, across all the data, a higher (more positive) value now indicates more vulnerability. This was applied to the following attributes:
    - Private vehicle drove alone
    - With private health insurance
    - With public health insurance
    - Households with internet
    - Total population (white)
    - 25 Yr. and older bachelor’s degree or higher
    - 25 Yr. and older high school graduate no bachelor’s degree
    - 25 Yr. and older high school graduate
3. The data was normalized (z-score normalization) to have zero mean and standard deviation of 1.

A few attributes were handled differently as they presented data differently. Specifically, some datasets categorized data with which vulnerability increased across categories of measurement (e.g. 1-7 person household, number of workers in household, etc.). For these categories, first a weighted index was established – assuming, for instance, house crowding scaled linearly with vulnerability, a one person house was given a factor of 1, a 2 person house a factor of 2, 3 person factor of 3, and so on. The percentages of households in each tract were then multiplied by their corresponding factors and summed across the tracts, yielding our weighted indices. For attributes identified as decreasing vulnerability, the team assigned a negative value before adding to the weighted index (private commuters, houses with no workers). then scaled these indices as a percentage of the maximum weighted index across all tracts, and finally re-normalized using a z-score normalization to match the rest of our data.

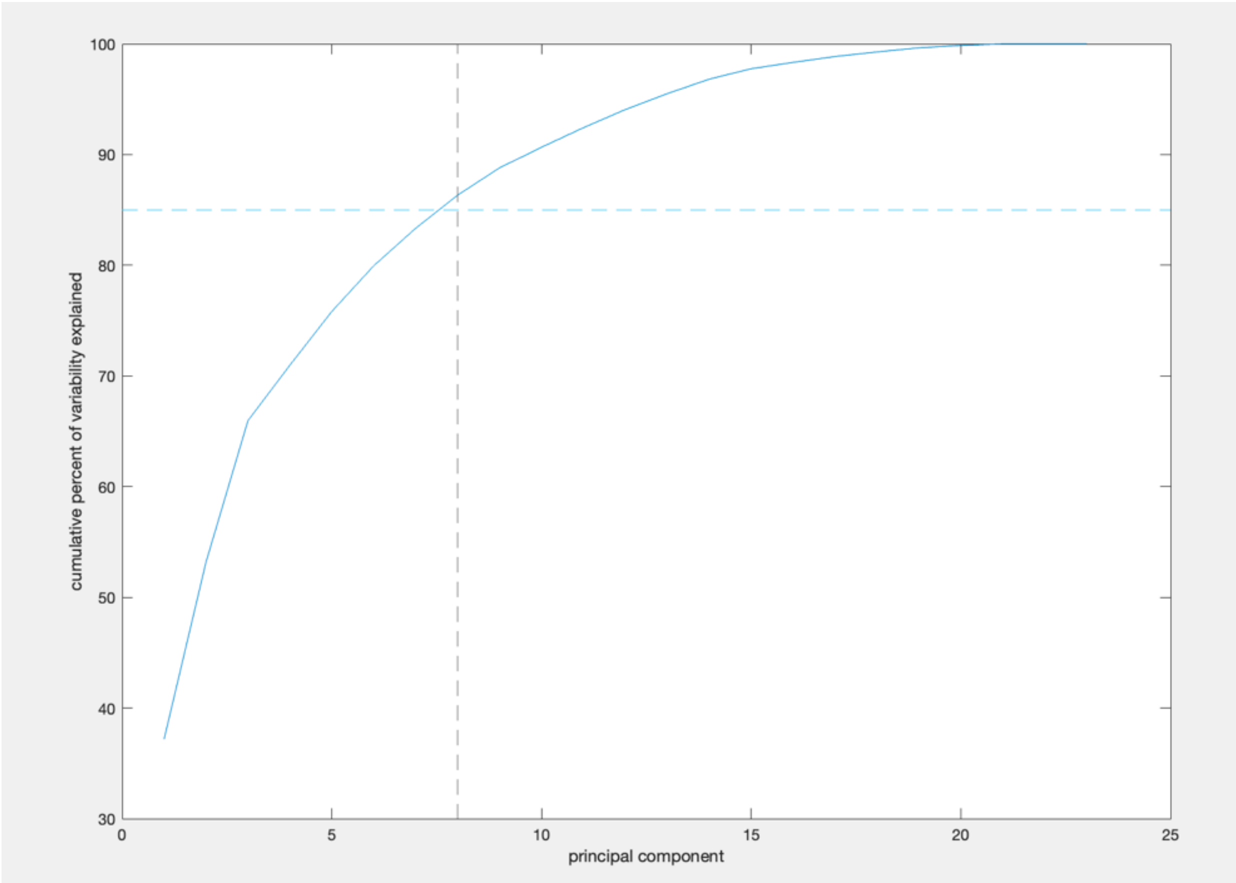
D. Dimensionality Reduction

The following plot shows the principal components sorted by the percentage of variability in data explained (PVE) by each one:



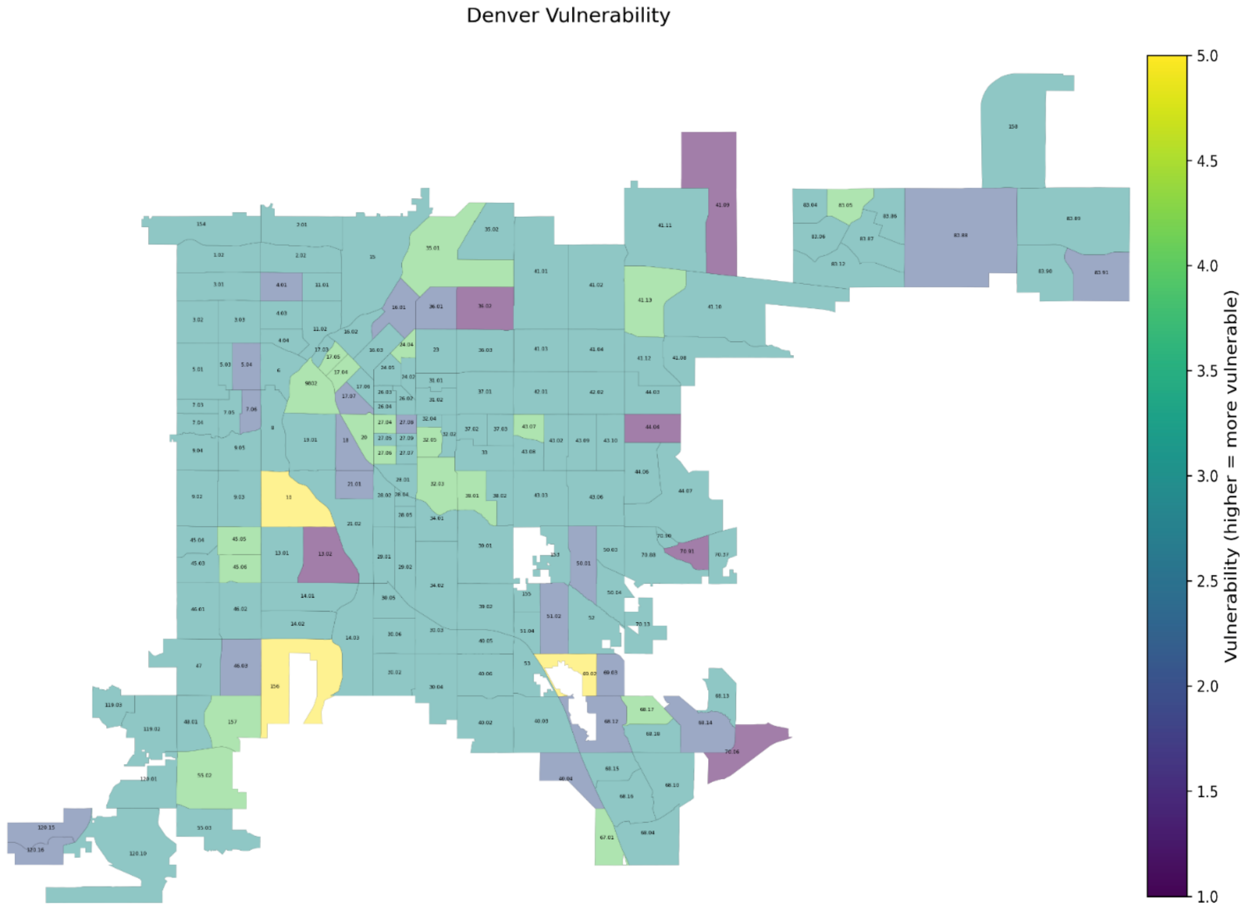
Graph 1: PVE by each Principal Component

A more useful way to look at this data is to look at the cumulative PVE (Graph 2). This shows that **8 principal components** explain more than 85% of the variability in data.



Graph 2: Total number of 8 principal components

E. Computing Vulnerability Level



Five Most Vulnerable Tracts:

Census Tract	Vulnerability Score (Ranked 1-5)	Vulnerability Score (Before ranking)
Census Tract 41.13	4	1.328504112
Census Tract 20	4	1.335600282
Census Tract 156	5	2.139131051
Census Tract 10	5	2.438855708
Census Tract 69.02	5	2.536779244

The table above shows the top five most vulnerable census tracts. Census tracts 69.02, 10, and 156 are ranked as 5 when categorized based on the vulnerability categorization strategies specified in the assignment. There were around 20 census tracts belonging to the next bin of vulnerability rankings. Prior to ranking the tracts in the 1 through 5 vulnerability categories, the census tracts that were the fourth and fifth most vulnerable tracts were 20 and 41.13 respectively.

These tracts were likely ranked among the top five most vulnerable census tracts because they had some of the highest levels in each of the named attributes. Tracts 10, 41.13, and 156 had some of the highest percentages of households with large numbers of occupants per household. Tracts 10, 20, 69.02, and 156 had high percentages of households with commuters. Tracts 10 and 69.02 were among the tracts with higher percentages of households with disabilities. Census tracts 69.02 and 156 were among the highest unemployment rates. Census tracts 20 and 69.02 had higher percentages of households paying rent rather than owning their homes. Census tracts 10 and 156 had higher percentages of low-income households. Tracts 20, 41.13, and 69.02 had higher percentages of people of color. Tracts 10, 69.02 and 156 ranked high in percentage of their population not earning a bachelor's degree.

As illustrated above by the recurring appearance of the top-ranked census tracts in many of the included attributes, it is clear why these tracts are considered the most vulnerable. They are consistently ranked among the highest in each attribute, proving their increased vulnerability against disease outbreaks.

Sources

[1] Heinz Center for Science, Economics, and the Environment. 2002. *Human Links to Coastal Disasters*. Washington, D.C.: The H. John Heinz III Center for Science, Economics and the Environment.

[2] Brakefield, Whitney S et al. "Social Determinants and Indicators of COVID-19 Among Marginalized Communities: A Scientific Review and Call to Action for Pandemic Response and Recovery." *Disaster medicine and public health preparedness* vol. 17

[3] Rogers, Christopher J et al. "Preparing for the next outbreak: A review of indices measuring outbreak preparedness, vulnerability, and resilience." *Preventive medicine reports* vol. 35 (2023)

[4] Fallah-Aliabadi, Saeed, et al. "Social vulnerability indicators in pandemics focusing on COVID-19: A systematic literature review." *Public Health Nursing*, vol. 39, no. 5, 7 Apr. 2022, pp. 1142–1155, https://doi.org/10.1111/phn.13075.

[5] Hammer CC, Brainard J, Hunter PR. Risk factors and risk factor cascades for communicable disease outbreaks in complex humanitarian emergencies: a qualitative systematic review. BMJ Glob Health. 2018 Jul 6;3(4):e000647. doi: 10.1136/bmjgh-2017-000647. PMID: 30002920; PMCID: PMC6038842.

[6] S.L. Cutter, B.J. Boruff and W.L. Shirley, "Social Vulnerability to Environmental Hazards," Social Science Quarterly, vol. 84, no. 2, Jun. 2003, pp. 242–261.