Scenario Options

Scenario 1: Infectious disease outbreak (simulated) in California

You have been given three datasets: dataset one contains weekly data about a novel infectious respiratory disease outbreak containing the number of cases and case severity by demographic categories (age category, race, sex) and geographic categories (county) for California counties (except for Los Angeles county); dataset two contains morbidity and population data by similar demographic and geographic categories for Los Angeles county, and dataset three populations estimates for 2023 by county and demographic categories.

Description

You are an employee of the California Department of Public Health working on infectious disease surveillance.

Research question

Your supervisors are interested to know the course of the outbreak, if it is disproportionately affecting certain demographic or geographic populations and how prevention and treatment resources should be allocated.

Data

You have been given three datasets:

1. Dataset one (sim_novelid_CA.csv) contains weekly data about cases and case severity by demographic categories (age category, race, sex) and geographic categories (county) for California counties (except for Los Angeles county);
2. Dataset two (sim_novelid_LACounty.csv) contains population data by similar demographic and geographic categories for Los Angeles county.
3. Dataset three (ca_pop_2023.csv) population estimates data by demographic category and county for 2023.

Data Merging, Summarization, and Reporting: Milestones 4-6

For milestones 4-6, you will join these datasets together to create a single dataset with aggregated summary morbidity and/or severity metrics and calculate rates. This dataset should be used to generate print-worthy visualizations (at least one table and one graph) to be included in the final report.

Scenario 2: Mortality and Medicare services in the United States

This scenario will utilize state-level mortality data for the United States (via CDC Wonder), population data by state and age group, and state-level Medicare enrollment and spending (via Center for Medicare Services and Kaiser Family Foundation). Teams will be asked to aggregate mortality data for a chosen category of cause of death and age group, calculate mortality rate for the specified cause and age group, and compare with a population-adjusted measure of Medicare utilization/enrollment.

Description

You work for a non-profit organization that is focused on making recommendations for public health related policy changes. Your organization is working on defending a policy proposal that would increase funding to Medicare and Medicaid programs.

Research Question

You have been asked to explore existing public data sources to determine if there is evidence of a correlation between mortality rates and health care expenditures in the United States. Additionally, you have been asked to highlight a few states that have higher mortality rates and lower public healthcare spending than other states.

Data

You have the following data available to you:

1. Dataset one
    - Option A: (kff_cms_medicare_state_summary.csv) contains state-level data on enrollment, utilization, and spending for Medicare parts A and B. Medicare eligibility typically is for adults age 65 and over (original dataset from Kaiser Family Foundation - Medicare spendingLinks to an external site.)
    - Option B: (kff_cms_medicaid_state_summary.csv) contains state-level data on Medicaid enrollment and spending by age group (original dataset compiled from Kaiser Family Foundation Medicaid spendingLinks to an external site. and enrollmentLinks to an external site. data)
2. Dataset two (cdc_wonder_mortality.csv) contains mortality counts for the year 2021 by state, age group, and underlying cause of death (original dataset from CDC WONDER Underlying Cause of DeathLinks to an external site.)
3. Dataset three (cdc_wonder_population.csv) contains population denominators for each state and age group (original dataset from CDC WONDER Underlying Cause of DeathLinks to an external site.)

Data Merging, Summarization, and Reporting: Milestones 4-6

For milestones 4-6, you will join these datasets together to create a single dataset with aggregated summary mortality and adjusted Medicaid/Medicare spending by state. You will also create a mortality rate variable using the data from the mortality and population datasets. This dataset should be used to generate print-worthy visualizations (at least one table and one graph) to be included in the final report.

Scenario 3: Environment and asthma in California

This scenario will utilize census tract level data on environmental measures (via Cal EnviroScreen), a mapping of census tract to county, and rates of asthma emergency department visits by county, age group, and race/ethnicity (via California Department of Health Care Access and Information). Teams will be asked to summarize tract level environmental measures into county level measures and compare with asthma ED rates overall or by a specific strata.

Description

You work for a public health agency that has been tasked with identifying counties to target environmental interventions with the goal of decreasing acute asthma.

Research Question

Your task is to utilize these datasets to create county-level measures to accomplish two goals. The first is to compare county asthma ED rates with a county CES measure to assess if there appears to be a correlation. The second is to compare asthma ED rates with county-level summaries for specific environmental measures to determine if those specific measures may be worth further investigation.

Data

You are limited to using only publicly-available datasets and have been provided with the following:

- Dataset one (calenviroscreen_measures_2021.csv) contains many environmental measures by California census tract (original sourceLinks to an external site. and data dictionaryLinks to an external site.).
- Dataset two (calenviroscreen_scores_demog_2021.csv) contains tract level Cal Enviro Screen (CES) scores and demographics. The Cal Enviro Screen score is a metric compiled from many measures, all of which are included in the first dataset (original sourceLinks to an external site. and data dictionaryLinks to an external site.).
- Dataset three (chhs_asthma_ed.csv) is from the California Health and Human Services department and contains age-adjusted asthma emergency department (ED) visits/rates by county, year, age group, and race/ethnicity. (original sourceLinks to an external site., data dictionaryLinks to an external site.)

Data Merging, Summarization, and Reporting: Milestones 4-6

For milestones 4-6, the three datasets will need to be joined together to create a single dataset that contains CES summary, environmental measure(s), and asthma rates by county. This dataset should be used to generate print-worthy visualizations (at least one table and one graph) to be included in the final report.