Cancer Deep Dive: Mortality and Incidence rates across the US

Joe Uren, Kevin Yi, Akshay Sharathchandra, Emily Zhang

Final Project for DATASCI 200
August 7, 2024

Our dataset

Dataset: from the CDC encompassing information about mortality and incidence rates from several chronic diseases

- We chose to focus on cancer
- Big problem in US – in 2024, there will be an estimated 2 million new cancer diagnoses, and 600K deaths [1]

Data Cleaning:

- The original dataset had over 1 million rows and 37 columns
- Dropped 10 columns that contained only null values
- Narrowed dataset to only lung and colon cancer

[1]: National Cancer Institute, Cancer Statistics https://www.cancer.gov/about-cancer/understanding/statistics

Research Question

- We looked at common cancers
- In 2024, prostate, **lung**, and **colorectal** cancers will make up 48% of new cancer diagnoses for men [2]
- In 2024, breast, **lung**, and **colorectal** cancers will make up 51% of new cancer diagnoses for women [2]

- **Research question: how do the mortality and incidence rates for colon and lung cancer differ amongst location, racial groups, and sex within the US from 2008-19?**

[2]: National Cancer Institute, Cancer Statistics https://www.cancer.gov/about-cancer/understanding/statistics
*lung cancer = lung and bronchial cancer; colon cancer = colorectal cancer

Our Assumptions

Dataset is made up of data collected at the state level

- Data is collected in a similar manner across states
- Data is collected in a similar manner across regions (urban vs rural)
- Data collected regarding race/ethnicity is accurate (self-reported vs state-reported)

If the data provider (state) decides to suppress data for quality or confidentiality reasons, the CDC does not report these data

- These data have surpassed a QC-check at the state level

Our Assumptions

Dataset is made up of data collected at the state level

- Data is collected in a similar manner across states
- Data is collected in a similar manner across regions (urban vs rural)
- Data collected regarding race/ethnicity is accurate (self-reported vs state-reported)

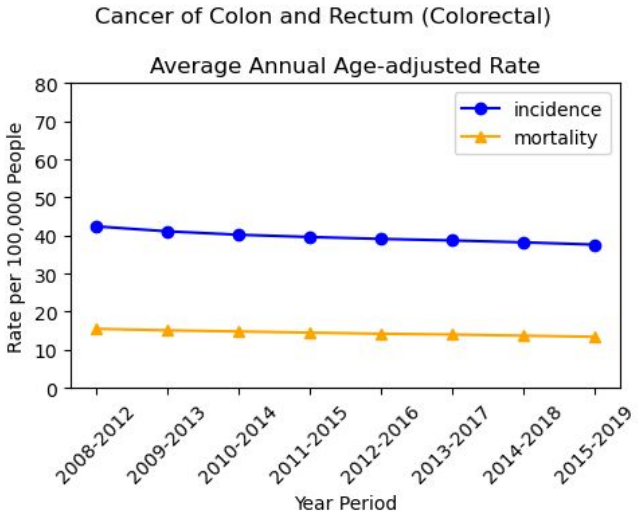
If the data provider (state) decides to suppress data for quality or confidentiality reasons, the CDC does not report these data

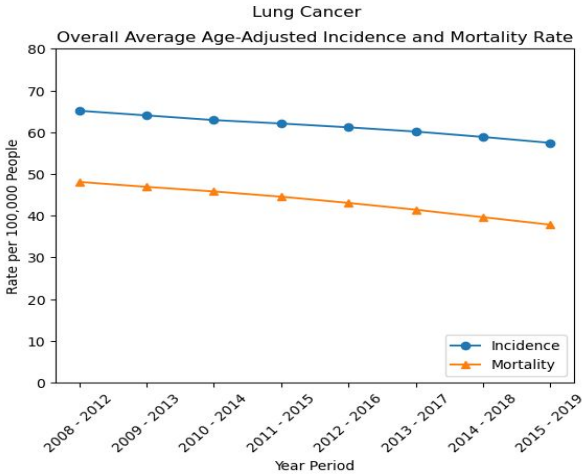
- These data have surpassed a QC-check at the state level

Disclaimer: for this analysis we are following the binary sex, and we are making assumptions based on provided racial/ethnic classifications

How have the incidence and mortality rates of colon and lung cancer changed over time?

- Both colon and lung cancers see declining trends in incidence and mortality rates

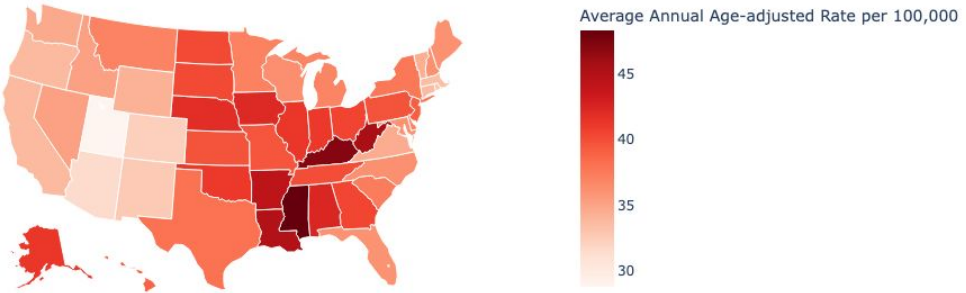




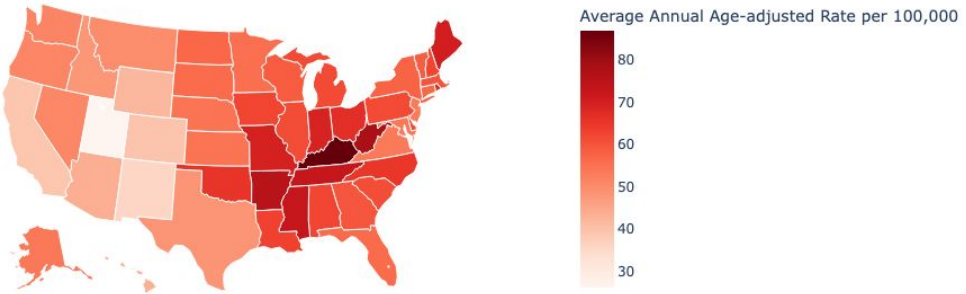
How do the incidence rates of colon and lung cancers differ by state?

- Higher incidence rates observed along the states in the Southeastern and Appalachian areas
- Lower incidence rates observed closer to the West coast
- Higher rates for lung cancer

Colon Cancer Incidence Rate [2015-19]



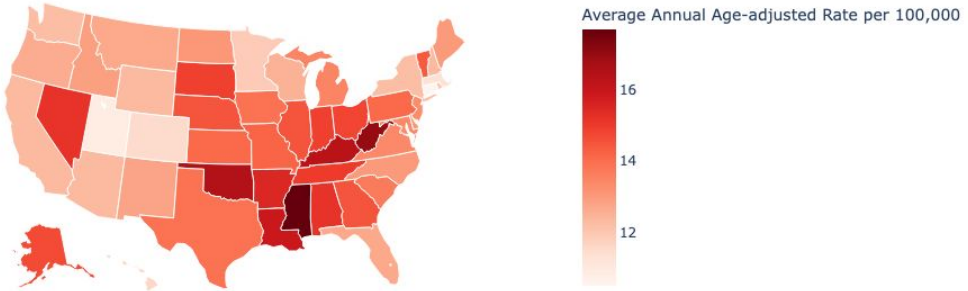
Lung Cancer Incidence Rate [2015-19]



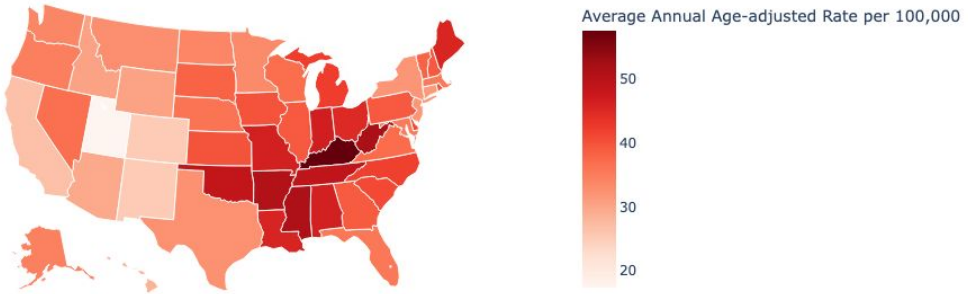
How do the mortality rates of colon and lung cancers differ by state?

- Higher mortality rates observed along the states in the Southeastern and Appalachian areas
- Lower mortality rates observed closer to the West coast
- Higher rates for lung cancer

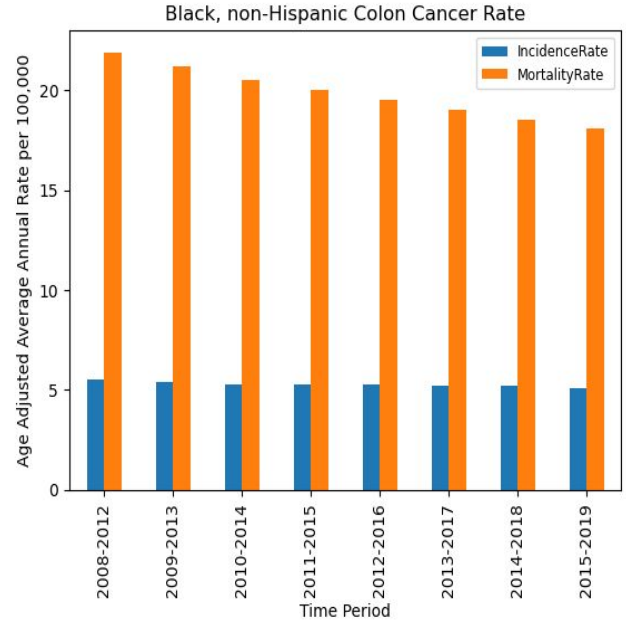
Colon Cancer Mortality Rate [2015-19]



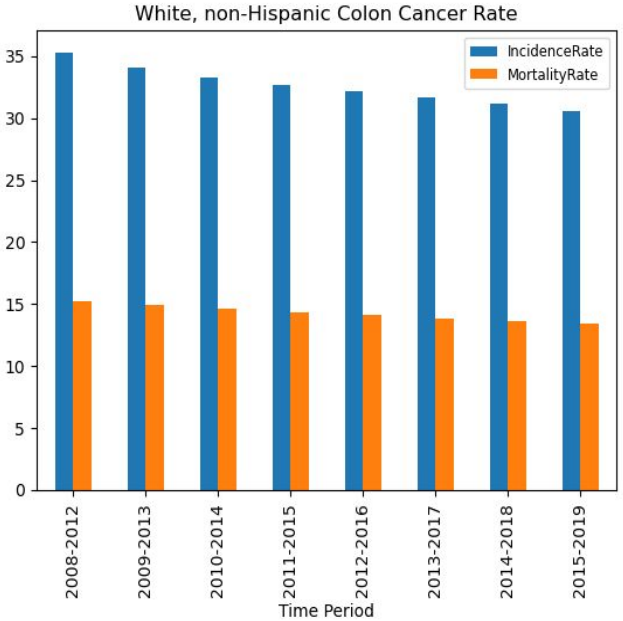
Lung Cancer Mortality Rate [2015-19]



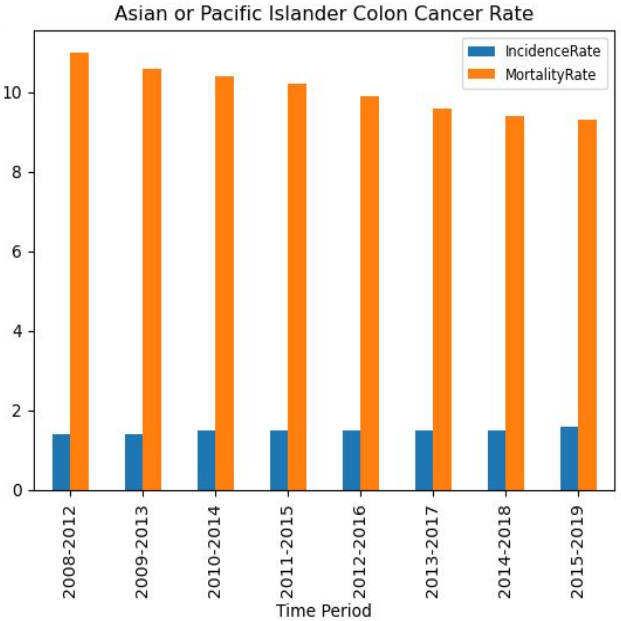
How have the incidence and mortality rates of colon cancer by racial groups changed over time?



Low incidence rate
High mortality rate

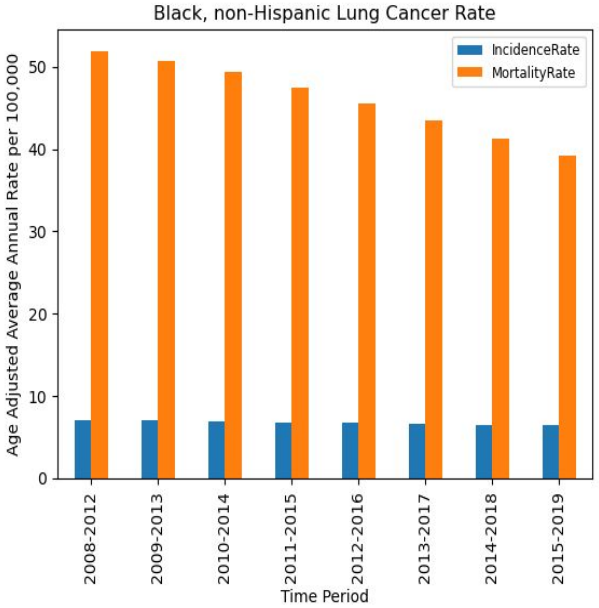


High incidence rate
Low mortality rate

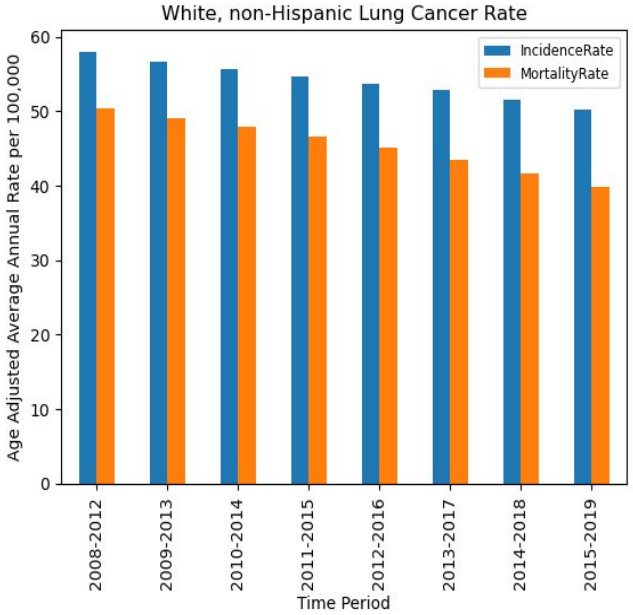


Low incidence rate
Mortality rate higher than Incidence rate
Lowest rates overall

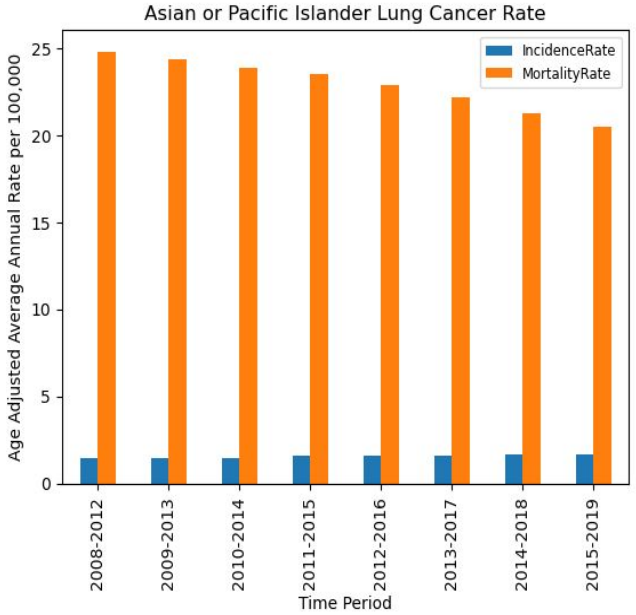
How have the incidence and mortality rates of lung cancer by racial groups changed over time?



Low incidence rate
High mortality rate

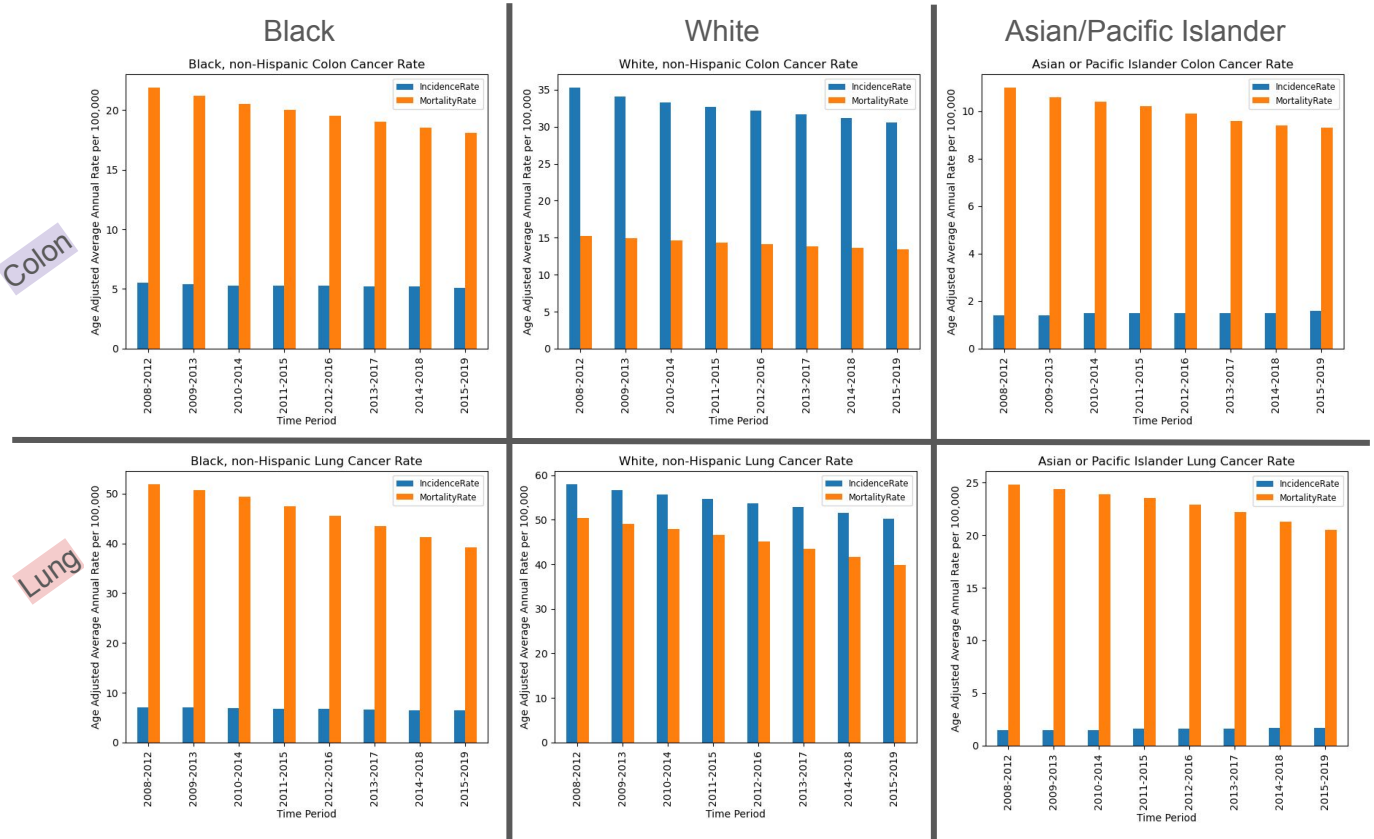


High incidence rate
Low mortality rate



Low incidence rate
High mortality rate
Lowest rates overall

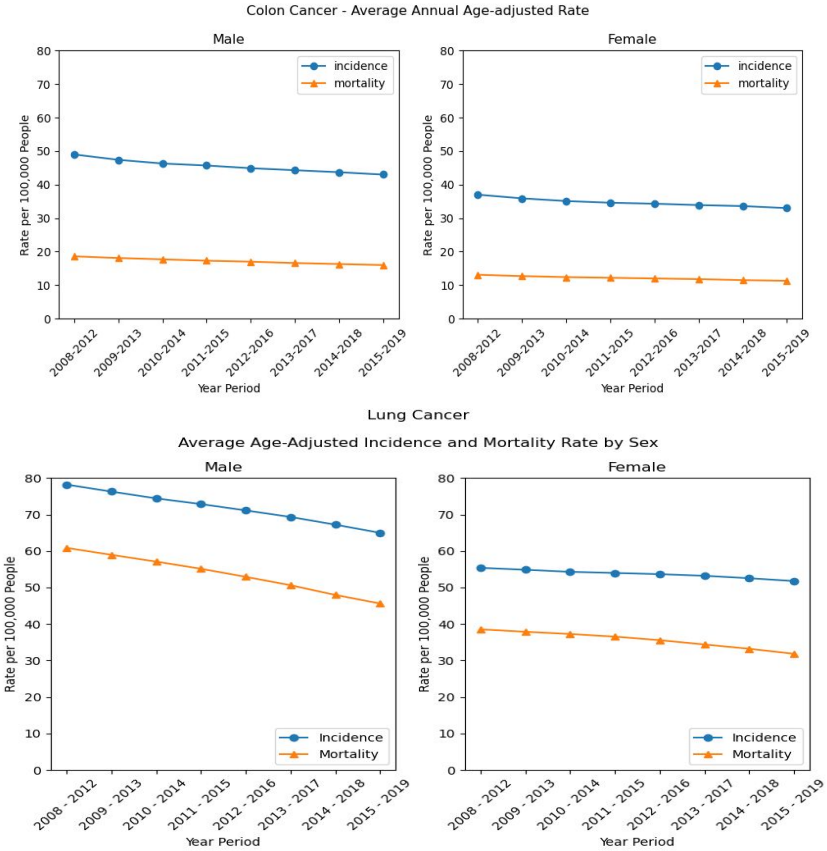
How have the incidence and mortality rates of colon & lung cancers by racial group changed over time?



- Asian/PI have the lowest incidence and mortality rates across both cancers
- Lung cancer seems more dangerous to all groups than colon (consistent with heatmaps)
- Incidence > Mortality suggest earlier detection, access to quality healthcare

How have the incidence and mortality rates of colon & lung cancers by sex changed over time?

- Mortality and incidence rates for both sexes have decreased over time
- Decreasing mortality rates are likely due to (in part) better treatment availability
- Smoking habits play a big role in sex differences for lung cancer



Key Takeaways

- Declining age-adjusted incidence and mortality rates across sexes and racial groups
- Males show higher rates, with steeper declines compared to females
- Regionally, the Southeastern and Appalachian areas exhibit highest lung cancer incidence rates, necessitating targeted public health interventions
- Asian populations have the lowest colon and lung cancer mortality rates, while Black non-Hispanics face the highest colon and lung cancer mortality rates

Future Directions

- Enhanced screening tools, improving healthcare access across the country
- Tailored public health campaigns to address disparities among minority communities
- Further education about the risk factors for cancer and prevention strategies