Causes of Death

Data Story Critique

Link to data story: Causes of Death

Prepared by

Hritvik Gaind

01 Introduction

About the Creator

- The article and visualization were created by **Nathan Yau**, a statistician and data visualization expert
- Yau runs FlowingData, a website dedicated to data visualization, statistical storytelling, and making data more understandable through visuals

Data Sources and Origins

- The dataset comes from the **Centers for Disease Control and Prevention (CDC)**, specifically from their Underlying Cause of Death database
- This database compiles mortality statistics in the U.S. based on death certificates filed across the country

Data Coverage

- **Time Range:** 2005–2014 (although CDC data is available from 1999 onward)
- The dataset documents mortality records in the U.S., detailing causes of death, demographic characteristics, and regional distribution

02 Data Processing and Tools

Data Processing

- The data was cleaned and organized to focus on causes of death, age groups, sex, and race.
- **Percentages** were calculated for each cause relative to the **total deaths** in each demographic, highlighting trends across different groups.

Tools Used

- **R** was used for data analysis and processing, allowing for manipulation and transformation of raw data into meaningful insights.
- **d3.js** was used for creating the interactive visualization, enabling users to explore the data by selecting different groups and viewing cause-of-death variations.

03 Main Ideas

Mortality trends

Differ significantly across age, sex, and race, with each group showing unique patterns in causes of death.

Age-related shifts

Prominent, with older individuals more likely to die from chronic diseases like cancer, while younger individuals face a higher risk of external causes like accidents.

Racial disparities

Exist where cause-of-death distributions are more variable in smaller racial groups, suggesting differences in health risks and healthcare access.

03 Key Takeaways

Age and cause of death

The likelihood of dying from cancer or chronic diseases increases with age, while younger people are more prone to external causes like accidents and violence

Sex Differences

Males are more likely to die from external causes, especially at younger ages, while females tend to have higher mortality from chronic diseases like cancer as they age

Racial Variations

Larger racial groups (e.g., whites and blacks) show more stable cause-of-death patterns, while smaller groups (e.g., American Indian and Asian) exhibit more variability due to smaller sample sizes

Evidence from the data

The interactive chart and percentages of deaths by cause across various demographics directly support these takeaways by clearly showing the differences in mortality trends

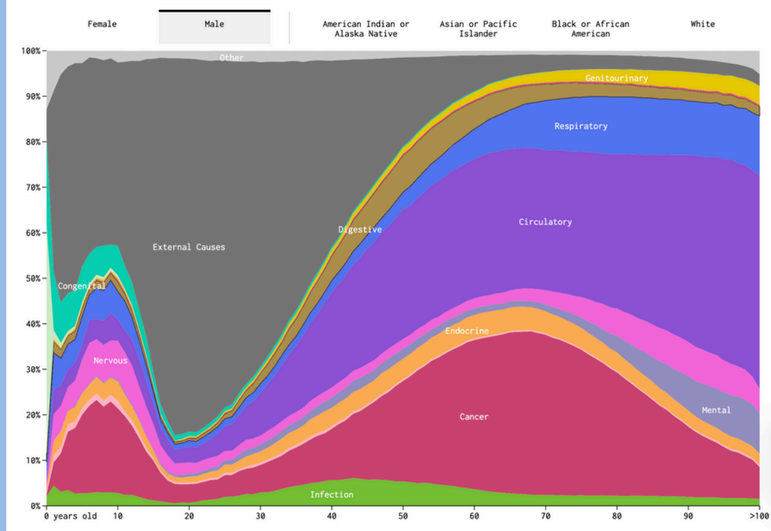
04 Strengths

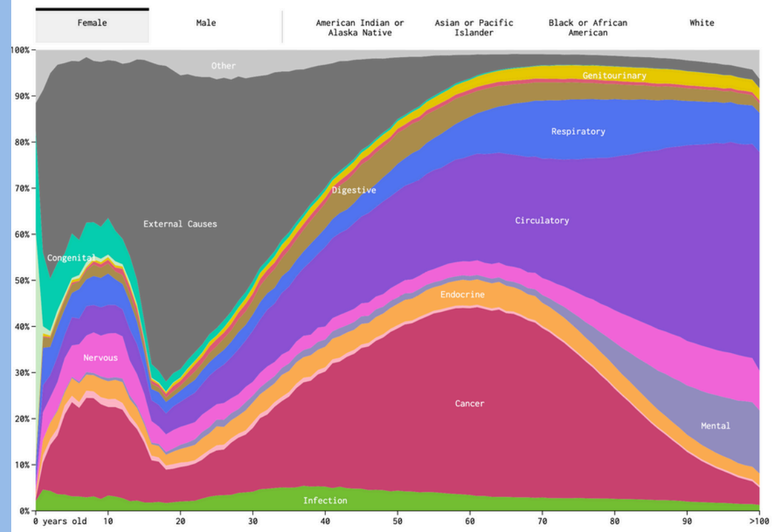
Interesting Insights

- **Adolescence and Early Adulthood (10–30 years)**
  - **Males:** External causes (accidents, violence) dominate, especially in late teens and early twenties, reflecting riskier behaviors.
  - **Females:** External causes contribute much less in this age range.
- **Midlife (30–50 years)**
  - Both genders experience a rise in cancer and endocrine-related causes, with males still showing higher proportions of external causes.
- **Old Age (50+ years)**
  - Circulatory diseases and cancer become the primary causes for both genders, with a slightly earlier onset in males.
  - Mental health issues are more prominent in females due to longer lifespans.

Key Takeaways

- Males have disproportionately higher mortality from external causes during youth.
- Females show a steadier shift from congenital and external causes to chronic diseases and mental health issues in older age.





04 Strengths

Interesting Insights

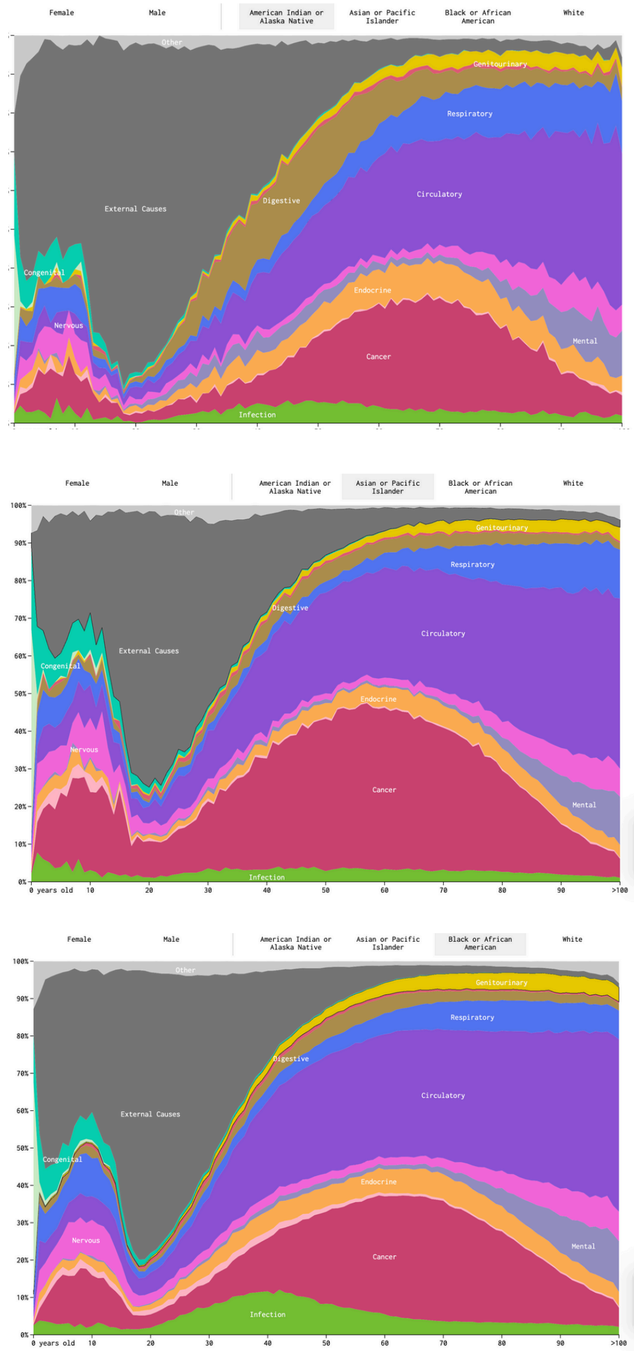
Variability in Mortality – Smaller groups (American Indian/Alaska Native, Asian/Pacific Islander) show more fluctuations, while larger groups (Black, White) have stable patterns.

External Causes – Higher among American Indian/Alaska Native and Black populations, while lower for Asians.

Chronic Diseases – Cancer is more prominent in White and Asian groups, circulatory diseases in Black populations, and diabetes-related deaths in American Indian/Alaska Native groups.

Mental & Nervous Disorders – Mental health-related deaths rise in older White populations, while nervous system-related deaths are higher in younger American Indian/Alaska Native groups.

Infections – Slightly more prominent in American Indian/Alaska Native populations.



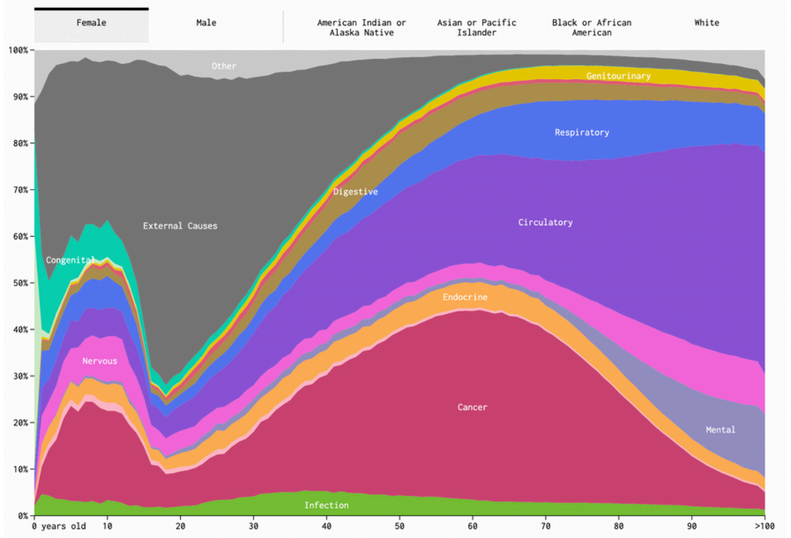
04 Strengths

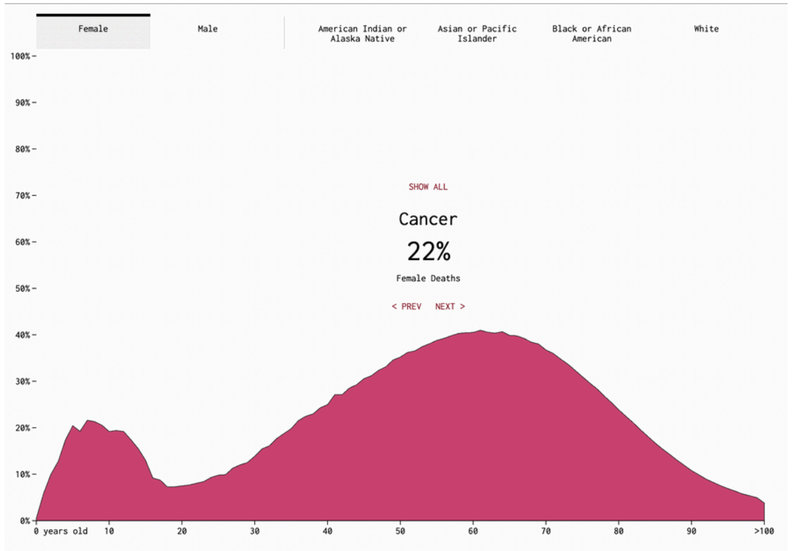
Interactive Filters

- One of the biggest strengths of this visualisation is how interactive it is. Being able to click on different causes of death and see how they play out across different age groups makes the data so much more engaging. Instead of just looking at a static chart, this lets you deep dive into specific causes like cancer and really understand when and how they impact people the most.

- The filtering makes it easy to spot trends that might otherwise get lost in the bigger picture. It gives a clear, focused view while still letting you step back and see the broader patterns when needed.

- This kind of interactivity isn’t just a nice touch - it completely changes how the data is understood, making it more intuitive and meaningful.





04 Strengths

Step-by-Step Explanation

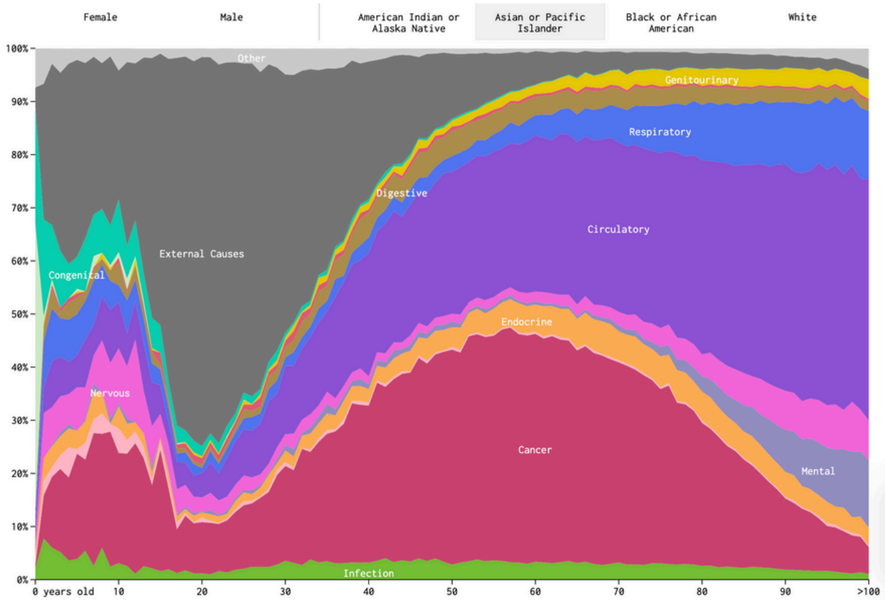
- The analysis is clearly structured, with examples provided to help break down mortality trends in a way that is easy to understand. This approach ensures that even those without a background in data analysis can grasp the key takeaways. By progressively introducing insights, the visualisation makes it simple to follow how causes of death shift across different ages and demographics.

Real-World Relevance

- Since the data is sourced from the CDC’s Underlying Cause of Death database, it reflects real-world health trends and demographic disparities. This makes the analysis highly relevant to public health discussions, helping to highlight patterns in disease prevalence, external causes of death, and racial or gender-based health inequalities.

05 Weaknesses

- **Stacked format makes comparison difficult:** Stacked format distorts proportions, making it hard to assess each cause’s true percentage.
- **Color and labeling issues:** Similar colours reduce clarity, labels are hard to read, and "Other" is too vague.
- **Lack of Multiple Filters:** The inability to apply multiple filters simultaneously (e.g., comparing black females to black males) limits the depth of comparison. Including this option could allow for more nuanced insights into how both sex and race interact with mortality trends.
- **Oversimplification:** This interactive chart, while engaging, could oversimplify the complex nature of mortality data by focusing primarily on percentages. Raw numbers or more detailed statistics could offer a clearer understanding of actual death counts, especially for smaller populations or rare causes of death.



05 Weaknesses

- **Data Limitations:** The dataset only covers deaths from 2005–2014, which might not fully reflect more recent trends or account for changes in mortality causes over time. Additionally, smaller racial groups (e.g., American Indian, Asian) may have sample size issues, leading to higher variability and less reliable insights in those populations.

- **Missing Processing Details:** The data story does not explain how the raw mortality data from the CDC's Underlying Cause of Death database was processed before visualization. Key steps such as data cleaning, handling missing values, applying filtering criteria, and performing data transformations are not detailed. Without this information, it is difficult to assess the accuracy and reliability of the analysis or understand any potential biases introduced during processing.

- **Lack of Geographic Breakdown:** The data story presents mortality trends at a broad national level, but regional variations are not explored. Different states or cities may have unique health risks, and adding a geographic filter could enhance the analysis.

06 Improvements

- **Better Comparison with Grouped Visuals:** Replacing the stacked format with grouped bar charts or treemaps improves clarity and avoids distortion
- **Enhanced Color & Labeling:** A distinct color palette and clearer labels, especially for ambiguous categories like "Other," improve readability
- **Updated Data Inclusion:** Using post-2014 data and addressing sample size issues in smaller racial groups enhances accuracy
- **Multi-filter Functionality:** Allowing simultaneous filters (e.g., black females vs. black males) enables deeper demographic insights
- **Raw Data Visibility:** Displaying both raw numbers and percentages provides a clearer view of mortality trends
- **Data Transparency:** Detailing data processing steps (cleaning, normalization, statistical methods) improves user understanding

Thank you