

Datasheet for the Dataset: Death rates for suicide, by sex, race, Hispanic origin, and age: United States

Julian Marrero

April 18, 2024

1 Motivation

Why was the dataset created?

The dataset was created to monitor and analyze trends in suicide rates across different demographics in the United States. This helps in understanding the impact of socio-economic, racial, and gender factors on suicide rates.

Who funded the creation of the dataset?

The dataset was published by Centers for Disease Control and Prevention. The maintainer of the dataset is the National Center for Health Statistics.

2 Composition

What does the dataset represent?

The dataset represents suicide rates in the United States per 100,000 people, broken down by sex, race, Hispanic origin, and age.

What and who is included in the dataset?

It includes individuals of different sexes, ages, racial and Hispanic origin categories.

How many instances are in the dataset?

There are 6,390 records in the dataset.

Are there any missing values?

Yes, there are missing values in the ‘ESTIMATE’ column and many instances where flags indicate data quality or specific notes on the data.

3 Collection Process

How was the data collected?

Data collected from death certificates and processed by National Center for Health Statistics.

Who was involved in the data collection process?

NCHS, National Vital Statistics System (NVSS); Grove RD, Hetzel AM. Vital statistics rates in the United States, 1940–1960. National Center for Health Statistics. 1968; numerator data from NVSS annual public-use Mortality Files; denominator data from U.S. Census Bureau national population estimates; and Murphy SL, Xu JQ, Kochanek KD, Arias E, Tejada-Vera B. Deaths: Final data for 2018. National Vital Statistics Reports; vol 69 no 13. Hyattsville, MD: National Center for Health Statistics. 2021. Available from: <https://www.cdc.gov/nchs/products/nvsr.htm>. For more information on the National Vital Statistics System, see the corresponding Appendix entry at <https://www.cdc.gov/nchs/data/hus/hus19-appendix-508.pdf>.

Over what timeframe was the data collected?

The dataset spans from 1950 to recent years.

4 Preprocessing/Cleaning

Was any preprocessing, cleaning, or labeling of the data done?

No.

5 Recommended Uses

For what tasks is the dataset suitable?

Suitable for studying trends in mental health, assessing the impact of interventions, public health policy making, and sociological research.

What tasks is the dataset not suitable for?

Not suitable for individual-level predictions or assessments due to its aggregated nature.

6 Ethical Considerations

What ethical considerations were taken into account?

Maintaining anonymity and ensuring data privacy are crucial. Also, careful interpretation to avoid stigmatizing communities.

Are there any legal concerns?

Data must comply with health data regulations like HIPAA in the US, particularly in how data is anonymized and shared.

7 Access

How can one access the dataset?

This dataset is intended for public access and use. Dataset can be accessed from <https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1>

8 Citation

How should the dataset be cited?

U.S. Department of Health & Human Services. 2021. Death rates for suicide, by sex, race, Hispanic origin, and age: United States. Available at: <https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1>

9 Usage Liscence

Under what license is the dataset distributed?

Copyright and Use: Public Domain

Restrictions: The use of U.S. government materials should not imply government endorsement of products, services, or entities. Additionally, trademarks, logos, and other identifiers of government agencies cannot be used without permission.