

Dataset Nutrition Label



Poverty and Drug Overdose

About

This dataset contains the percent of people in poverty by race and state. It also includes the total number of deaths and number of drug overdose deaths by state.

Data Creation Range: 2016

Created By: BC-6

Content: CSV files

Source: [BC-6 GitHub](#)

Alert Count: **5**

Completeness 1

Provenance 1

Misrepresentation 1

Collection 1

Socioeconomic Bias 1

Inaccurate Prediction 0

Description 0

Composition 2

Racial bias 1

Socioeconomic Bias 1

Use Cases:

Potential real-world applications of the dataset

1. Is poverty related to the number of drug overdose deaths in the US?
2. Is there a pattern regarding race and poverty?
3. Do certain states/regions have more poverty problems?
4. Where are the most drug overdose deaths occurring?

Badges:



About Humans

Dataset Information

Description

1. Tell us about this dataset

This dataset contains estimates on the number of total deaths and number of drug overdose related deaths for each state in the United States, as well as the percent of people who are in poverty by race for each state in 2016. The total number of deaths and number of drug overdose related deaths were provided by the CDC's Vital Statistics Rapid Release. The number of people in poverty by race for each state and the population numbers of each race were provided by the American Community Survey.

2. Is there an intended purpose for the dataset? What domain was it designed for?

The purpose of this dataset is to see if there is any correlation between poverty and drug overdose deaths.

3. Are there tasks for which the dataset would be cautioned against being used?

Since the first dataset combined the numbers of both Asians and Hawaiian/Pacific Islanders and the second dataset put them in separate columns, when the columns were combined, the lack of information on Hawaiian and Pacific Islanders in the second dataset caused many values in the combined column to be NA. Therefore, this dataset should not be used when finding the correlation between poverty and drug overdose deaths among these groups of people.

Composition

1. What data does each instance consist of?

There are 8 columns and 306 rows.

Row Information:

The rows include all the 50 states in the United States.

Column Information:

State: Abbreviation of each state

State.Name: Full name of each state

Year: The year the data was collected

Number of Deaths: Total number of deaths by state

Number of Drug Overdose Deaths: Total number of drug overdose deaths by state

race: Races within the dataset including white, black, hispanic, and multiple races

perc_in_pov: Shows the percentage of people in poverty

region: The region the state is in

2. Is any information missing from individual instances?

There are some NAs for the columns that calculate the percent of a certain population that is in poverty for each state because estimates with relative standard errors greater than 30% are not provided.

3. *Is your dataset a sample?*

Our dataset is not a sample. Due to rounding, these numbers will not reflect the exact number of deaths and percent of people in poverty.

4. *If the data was processed, was the raw data saved in addition to the preprocessed/cleaned/labeled data?*

[Drug Overdose Deaths](#)

[People in Poverty Based on Race for each State](#)

[Population by Race for each State](#)

5. *Is there confidential data included in this dataset?*

There is no confidential data included in this dataset.

6. *Does the dataset identify any subpopulations (e.g., by age, gender)?*

There are subpopulations within the dataset by race.

Provenance

1. *Who created the dataset and on behalf of which entity?*

Group BC-6 from INFO 201 created this dataset.

Collection

1. *Over what timeframe was the data collected?*

The data was collected from the year 2016.

2. *Has the data been reviewed for quality?*

N/A

3. *Were any ethical review processes conducted (e.g., by an institutional review board)?*

N/A

Management

1. *How can the data be accessed? If this is going to change over time, please indicate this.*

The data can be accessed through the GitHub repository.

2. *Will the dataset be distributed to any other individuals/third parties (e.g. companies, institutions, organizations)?*

This dataset will not be distributed.

3. *Will the dataset be updated?*

This dataset will not be updated.