California’s suicide rate highest and lowest region 2017

Jax Skorich

Abstract

Introduction: Suicide is a critical public health issue influenced by a multitude of factors. Despite extensive research, the dynamics of suicide rates in a diverse state like California require continuous monitoring and analysis. Previous studies have highlighted demographic factors and regional variations in suicide rates, but there is a need for updated and localized data to address current trends and inform targeted interventions (NCBI; CDPH). This report explores the variation in suicide rates across different counties in California, highlighting significant gender disparities in 2017.

Objectives: The primary objective of this project is to create two data visualizations in Tableau after cleaning and tidying the dataset using Excel, Notepad++, and OpenRefine. A secondary objective is to identify and highlight the county with the highest and lowest suicide rates, as well as explore the differences in suicide rates between male and female genders in 2017.

Methods: Data for this study is sourced from the California Department of Public Health, covering annual suicide rates from 2000 to 2017 with demographic breakdowns. This dataset includes details on age, gender, and ethnicity. The data was analyzed using OpenRefine and Tableau, with visualizations in the form of a map and a bar chart used to present the findings.

Results: The analysis revealed that Humboldt County had the highest suicide rate in California in 2017, at 528.2 per 100,000 people, while Colusa County had the lowest rate at 26.6 per 100,000 people. Additionally, the suicide rate for men was significantly higher at 16.37 per 100,000 compared to the 4.44 per 100,000 for women.

Conclusions: These findings highlight the need for targeted mental health interventions in high-risk counties and among men who are disproportionately affected by suicide.

California's suicide rates highest and lowest region 2017

Introduction

Suicide is a critical public health issue influenced by a multitude of factors, including mental health disorders, socioeconomic conditions, and access to healthcare. Despite extensive research, the dynamics of suicide rates in a diverse state like California require continuous monitoring and analysis. According to the CDC, California had a suicide rate of 10.5 per 100,000 population, with 4,312 deaths in 2017. This high incidence underscores the urgent need for ongoing evaluation of trends and contributing factors to effectively address and reduce suicide rates.

Previous studies have highlighted significant demographic factors and regional variations in suicide rates within California. Research has shown that certain age groups, genders, and ethnicities are more vulnerable to suicide, necessitating targeted prevention efforts (NCBI; CDPH). For instance, men have been consistently found to have higher suicide rates compared to women, and regional disparities indicate that some counties have significantly higher rates than others. These variations suggest that localized factors, such as economic conditions and availability of mental health services, play a crucial role in influencing suicide rates.

The need for updated and localized data to address current trends and inform targeted interventions is paramount. The California Department of Public Health (CDPH) and other agencies have provided valuable datasets that allow for detailed analysis of suicide rates across different demographics and regions (NCBI; CDPH; CDC). By examining these datasets, researchers can identify emerging trends and at-risk populations, enabling the development of more effective prevention strategies. This continuous monitoring is essential to adapt to changing conditions and to implement timely interventions. Additionally, initiatives like "Suicide is Preventable" emphasize the importance of awareness and preventative measures.

This report explores the variation in suicide rates across different counties in California, with a specific focus on gender disparities in 2017. Using data visualization techniques, the study highlights significant differences in suicide rates between counties, identifying Humboldt County as having the highest suicide rate and Colusa County as having the lowest. Additionally, the report examines the stark difference in suicide rates between men and women. By addressing these disparities, the study aims to contribute to the development of more effective and focused suicide prevention strategies.

Objective(s)

The aim of this project is to identify and highlight the counties in California with the highest and lowest suicide rates. Additionally, this project aims to compare the suicide rates between men and women for the year 2017 to understand what (if any) difference or relationship these rates may have.

Methods

Data Source

Data.gov. (2024, May 14). State of California - Suicide deaths per 100,000 population (LGHC Indicator). [Data set]. https://catalog.data.gov/dataset/suicide-deaths-per-100000-population-lghc-indicator-9a8e1

The data source for this study is a dataset used for a Let's Get Healthy California indicator, specifically focusing on deaths by suicide per 100,000 population. The dataset is aggregated from statistical datasets compiled from California death records (Data.gov, 2024). This data provides a comprehensive overview of suicide rates across all counties in California, with demographic information, including age, gender, and ethnicity.

The Center for Health Statistics and Informatics, California Department Public Health organization is responsible for the original data collection for Let's Get Healthy California, an initiative that aims to improve the health and well-being of Californians. The data collected by this initiative are vital for understanding public health trends and developing strategies to address various health issues, including suicide. Detailed information on the goals and methods of this initiative can be found on their website (Slone, 2022).

The dataset spans from 2000 to 2017, providing an extensive period for analysis. However, for this particular study, the focus is solely on the year 2017. This allows for a detailed snapshot of the suicide rates throughout the counties. The variables included in this dataset are crucial for understanding the patterns and demographics of suicide in California. These variables include:

Year: The year of data collection (YYYY)

Year Ranges: Years of data ranges (YYYY-YYYY)

Region: The county in California where the death occurred (Alameda Amador, Butte, Calaveras, Colusa, Contra Costa, Del Norte, El Dorado, Fresno, Glenn, Humboldt, Imperial, Inyo, Kern, Kings, Lake, Lassen, Los Angeles, Madera, Marin Mariposa, Mendocino, Merced, Monterey, Napa, Nevada, Orange, Placer, Plumas, Riverside, Sacramento, San Benito, San Bernardino, San Diego, San Francisco, San Joaquin, San Luis Obispo, San Mateo, Santa Barbara, Santa Clara, Santa Cruz, Shasta, Siskiyou, Solano, Sonoma, Stanislaus, Sutter, Tehama, Trinity, Tulare, Tuolumne, Ventura, Yolo, Yuba) and all of California as a whole (CALIFORNIA).

Suicide Rate: The rate of suicides per 100,000 population

Numerator: The number count of individuals who died by suicide

Denominator: The population size

Gender: Gender of the individuals who died by suicide

Age Adjusted Rate: Adjusted age rate of the individuals who died by suicide.

Ethnicity: Ethnicity of the individuals who died by suicide

For the purposes of this project, the dataset was refined to focus exclusively on the year 2017, with ethnicity data excluded to streamline the analysis and concentrate on gender and county-level differences. This exclusion was necessary to simplify the dataset and focus on the primary variables of interest: suicide rates by gender and county (NCBI, CDPH). To further pursue this interest, the CALIFORNIA region was excluded to focus on the counties, and the total gender rate was excluded. The data were then cleaned and prepared using tools like Excel and OpenRefine to ensure accuracy and consistency before being visualized in Tableau.

Data Dictionary

Variable Name	Definition	Format	Values

Year	The year data collected	YYYY	2017
County	The county in California	Text	Alameda Amador, Butte, Calaveras, Colusa, Contra Costa, Del Norte, El Dorado, Fresno, Glenn, Humboldt, Imperial, Inyo, Kern, Kings, Lake, Lassen, Los Angeles, Madera, Marin, Mariposa, Mendocino, Merced, Monterey, Napa, Nevada, Orange, Placer, Plumas, Riverside, Sacramento, San Benito, San Bernardino, San Diego, San Francisco, San Joaquin, San Luis Obispo, San Mateo, Santa Barbara, Santa Clara, Santa Cruz, Shasta, Siskiyou, Solano, Sonoma, Stanislaus, Sutter, Tehama, Trinity, Tulare, Tuolumne, Ventura, Yolo, Yuba
Suicide Rate	Suicide rate per 100,000 population	Numeric	26.6-528.2, 4.4-10.6, continuous
Age Adjusted rate	Age adjusted rate per 100,000 population	Numeric	26.6-528.2, 4.4-10.6, continuous
Gender	Binary Gender of individuals who died by suicide	Text	Male, Female

Data Cleaning and Transformations

Data cleaning involved screening for missing values, correcting inconsistencies, and transforming variables where necessary. Steps included:

1. Screening the dataset for missing or inconsistent data entries.
2. Converting categorical variables such as gender and ethnicity into columns.
3. Validating data to ensure accuracy and consistency.
4. Documenting all changes and manipulations in JSON files for reproducibility (OpenRefine & Tableau).

The process in Excel included selecting all columns and rows, creating a table with the style tool. Then using the "Text to Columns" data tool to format the data properly. Detailed steps and transformations from Excel were documented using Notepad++ v8.6.9 and saved in a JSON file.

This cleaned dataset was then saved as CA_suicide_rates_v2 and further processed in OpenRefine v3.8.2 to ensure consistency and accuracy before visualization. Detailed steps and transformations in OpenRefine were documented by JSON file and included columnizing by key columns, splitting columns by separators, and performing text transformations to ensure data was in the correct format for analysis. Detailed steps and transformations in Tableau Desktop Public Edition 2024.2.0 (20242.24.0613.1930) were documented using Notepad++ v8.6.9 and saved to a final third JSON file for optimal reproducibility.

Analysis

The analysis of the dataset was conducted using a combination of tools to ensure thorough data cleaning, transformation, and visualization. The primary tools used were Notepad++ v8.6.9, OpenRefine v3.8.2, Microsoft Excel, and Tableau Desktop Public Edition 2024.2.0 (20242.24.0613.1930). Tableau Desktop Public Edition 2024.2.0 was the primary tool used for data visualization. Two key visualizations were created:

1. **Map Visualization:** This map highlighted the counties with the highest and lowest suicide rates in California for the year 2017.
2. **Bar Chart:** This bar chart illustrated the differences in suicide rates between genders for the year 2017.

The analysis focused exclusively on data from the year 2017, excluding data from other years and ethnicity data to streamline the analysis. This decision was made to concentrate on the primary variables of interest: county-level and gender-specific suicide rates. Outliers were carefully examined, and any inconsistencies were addressed during the data cleaning process to ensure accurate and reliable results.

Results

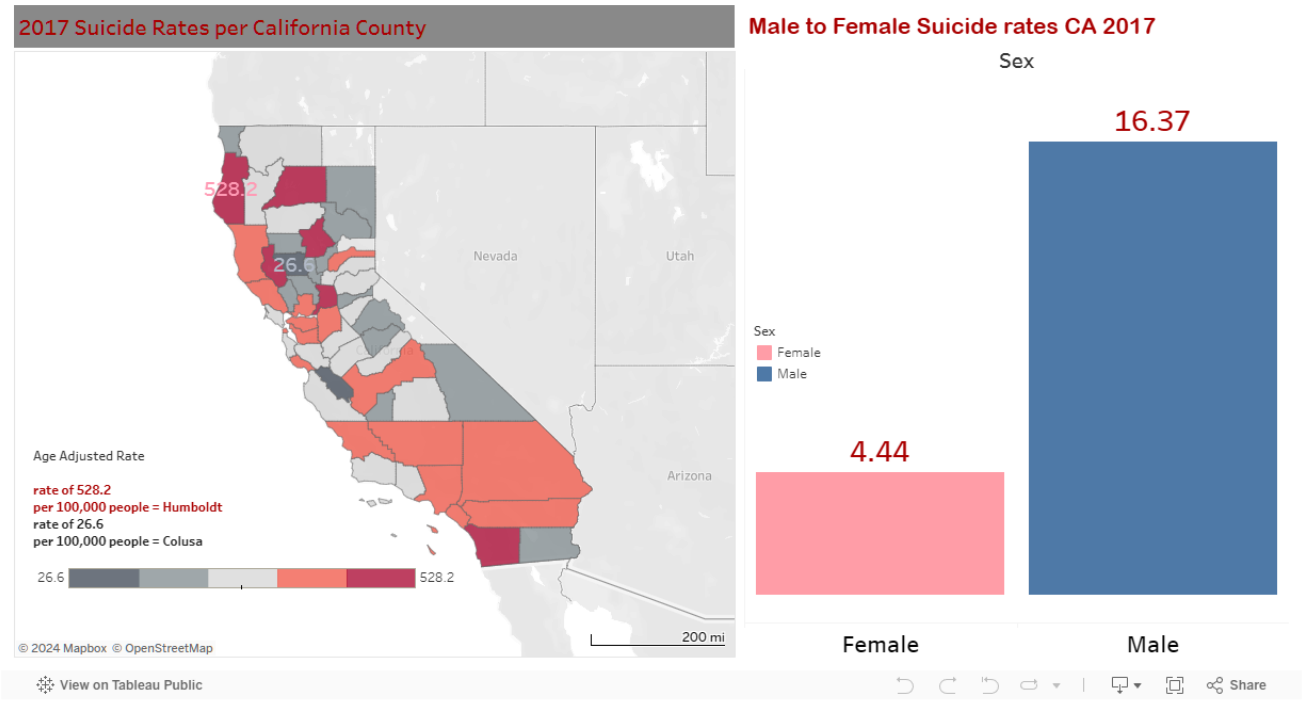
The analysis revealed significant variations in suicide rates across different counties in California for the year 2017. Humboldt County was identified as having the highest suicide rate, with a staggering 528.2 deaths per 100,000 people. In stark contrast, Colusa County had the lowest suicide rate at 26.6 deaths per 100,000 people. These findings highlight the substantial regional disparities in suicide rates within the state.

In addition to the geographic differences, the analysis also highlighted significant gender disparities in suicide rates for the same year. The suicide rate for men was markedly higher, at 16.37 per 100,000, compared to 4.44 per 100,000 for women. This gender disparity underscores the need for targeted interventions to address the specific needs of different demographic groups.

Link to live visualization: https://public.tableau.com/app/profile/jax.skorich/viz/CA_suicide_rates-tableauwbk/2017SuicideRatesCalifornia?publish=yes

In California, Humboldt County faces a staggering suicide rate of 528.2 per 100,000 people, while Colusa County reports the lowest at 26.6.

Additionally, in 2017, the suicide rate for men was significantly higher at 16.37 per 100,000 compared to 4.44 for women.



Details

These visualizations provide a clear and impactful representation of the data, facilitating a better understanding of the patterns and trends in suicide rates across California. The map visualization effectively highlights the counties with the highest and lowest rates, while the bar chart underscores the significant gender differences in suicide rates. These findings are crucial for informing targeted mental health interventions and policies aimed at reducing suicide rates across the state.

Conclusions

The findings of this study highlight significant regional and gender disparities in suicide rates across California in 2017. Humboldt County's exceptionally high suicide rate suggests that localized factors, such as economic conditions, access to mental health services, and cultural influences, may play a crucial role in elevating suicide risks in specific areas. Conversely, the low rate in Colusa County indicates potential protective factors that warrant further investigation. The gender disparity, with men experiencing significantly higher suicide rates than women, underscores the necessity for gender-specific mental health interventions. Potential causes for this phenomenon could include societal expectations, differences in help-seeking behaviors, and variations in mental health conditions between genders.

Ethically, it is essential to approach these findings with sensitivity and care, avoiding stigmatization of specific groups or regions. The data should be used to inform targeted interventions that address the unique needs of high-risk populations without perpetuating stereotypes or biases. The implications for

future research include the need for more granular data to explore the underlying causes of these disparities, the effectiveness of current intervention strategies, and the development of new, more targeted approaches to suicide prevention.

Limitations

Several limitations affect the findings of this study. First, the dataset only includes information up to 2017, which may not reflect more current trends or emerging factors influencing suicide rates. Additionally, the exclusion of ethnicity data and the focus on a single year may limit the comprehensiveness of the analysis, potentially overlooking important demographic and temporal variations. Biases or potential errors in how data were collected and recorded could also affect the accuracy of the results. For example, differences in reporting practices across counties might lead to inconsistencies in the data. Time restrictions and resource limitations constrained the geographic scope of the analysis, preventing a more extensive examination of trends over a longer period.

Future research should aim to include more recent data and consider a broader range of demographic factors, including ethnicity, to provide a more comprehensive understanding of suicide trends. Additionally, efforts should be made to standardize data collection practices across regions to ensure consistency and accuracy. Further analysis is needed to identify and address the specific needs of different population groups, enabling the development of more effective, evidence-based suicide prevention strategies.

References

Data.gov. (2024). State of California - Suicide deaths per 100,000 population (LGHC Indicator). [Data set]. https://catalog.data.gov/dataset/suicide-deaths-per-100000-population-lghc-indicator-9a8e1

CDPH, Injury And Violence Prevention Branch. (2021). Violent deaths involving multiple victims in California, 2017-2018. In CalVDRS. https://www.cdph.ca.gov/Programs/CCDPHP/DCDIC/SACB/CDPH%20Document%20Library/CA%20Violent%20Death%20Reporting%20System%20(CalVDRS)/CalVDRS_Multi-Vtms_2018_ADA.pdf

Lund, J. J., Tomsich, E., Schleimer, J. P., & Pear, V. A. (2023). Changes in suicide in California from 2017 to 2021: a population-based study. Injury Epidemiology., 10(1). CDC. https://doi.org/10.1186/s40621-023-00429-6

Slone, L. (2022). *Living well / reducing suicide - let's get healthy California*. Let's Get Healthy California. https://letsgethealthy.ca.gov/goals/living-well/reducing-suicide/

Suicide Prevention Awareness. (2024). Know the warning signs of suicidal behavior. https://www.suicideispreventable.org/