Decoding Suicide Patterns in the United States: A Multi-Factorial Data Exploratory Approach with Policy Driven Insights

By:
Shravani Pranay Sawant

Graduate Student M.S. in Cybersecurity
Stevens Institute of Technology

Course: SYS 624: Data Exploration and Informatics for Engineering Management (Web Campus)

Instructor: Dr. Carlo Lipizzi

Semester: Spring 2025

Submission Date: April 26, 2025

Abstract:

Suicide is a complex and evolving public health issue, shaped by intricate interactions of demographic, psychological, social, and environmental factors. This research presents a data driven narrative of suicide trends in the United States between 2005 and 2022, structured around multidimensional variables such as age, gender, race, education level, geographic distribution, and seasonality. The primary dataset was provided by the course professor and includes validated national records of suicide intent and demographic details. To enrich this core data, complementary sources such as CDC's longitudinal reports [1], county-level suicide estimates [6], and academic analyses of risk factors [3], [12] were integrated into the study.

What distinguishes this work is not only the analytical depth but its strong emphasis on interpretability and policy relevance. Traditional statistical methods are complemented by creative, interactive visualizations including choropleth maps, bump charts, sunburst hierarchies, animated time-series, and multi-level Sankey diagrams. Each result is contextualized with reflections grounded in prevention literature, including models like the Zero Suicide framework [2] and Google Trends correlation studies [5].

Rather than treating suicide analysis as a purely numerical challenge, this study bridges the gap between empirical insight and real-world impact. Through its structured dashboards and narrative visual storytelling, the project invites stakeholders ranging from public health officials and policy architects to mental health advocates and data scientists to explore where suicide prevention efforts succeed, and where critical disparities remain [11], [13], [17]. The final product is a visually coherent, thematically organized report that translates raw data into actionable human stories.

Keywords

Suicide prevention, demographic trends, mental health, public health data, choropleth map, Sankey diagram, visualization, education and suicide, racial disparities, seasonal variation, dashboard analytics, policy insight.

I. Introduction

Suicide is a persistent and multifactorial public health crisis that claims tens of thousands of lives each year in the United States alone. According to the Centers for Disease Control and Prevention (CDC), suicide rates have been on a worrying upward trajectory since the late 1990s, showing significant cross-sectional variation across demographic groups [1], [10]. Despite the implementation of various interventions, including enhanced access to crisis services and mental health awareness initiatives, the multifaceted nature of suicide drivers often renders conventional public health strategies deficient in their predictive capacity and adaptability [2], [11].

This study has been prompted by a necessity to establish a link between quantitative insight and practical policy formulation. A substantial body of research has identified individual level risk factors, including psychiatric history, trauma exposure, and socioeconomic hardship [8], [14]. However, there remains a paucity of comprehensive, visual, and easily interpretable analyses that track how these factors evolve collectively over time and space. Most suicide surveillance studies are static or confined to specific dimensions, such as race, gender, or clinical diagnosis [3], [7], [15]. The present study extends these perspectives by taking a multidimensional, time-aware approach and introducing question-wise dashboards tailored for comparative exploration.

The core of this study is grounded in a primary dataset titled full_data.csv, provided by the course professor, which served as the foundational basis for exploring patterns in suicide across key demographics and intent-based filters. To deepen the temporal and regional context of the analysis, I independently sourced supplementary datasets comprising year-wise, state-level suicide statistics ranging from 2005 to 2022. These auxiliary datasets allowed for the construction of comparative dashboards and animated choropleth maps that reveal geographic disparities and evolving trends. By merging these multiple layers of insight, this research advances from a unidimensional data summary into a dynamic, storytelling analysis meant to inform public health discussions and data-driven interventions.

II. Related Work

Over the past two decades, suicide research has evolved significantly, drawing from public health, psychology, epidemiology, and increasingly, computational methods. Large-scale datasets and digital surveillance tools have been employed to analyze suicide trends across demographic variables, clinical factors, and geographic contexts.

Early foundational work by the CDC highlighted consistent increases in U.S. suicide rates from 1999 through 2019, with notable disparities by sex, age, and state level contexts [1], [10], [13]. The CDC's state level reports further emphasize sharp regional variations, with some rural and economically disadvantaged areas exhibiting disproportionately high suicide rates [11].

Beyond descriptive statistics, researchers have investigated predictors of suicide attempts and mortality using epidemiological models [9], [12]. A data science approach has been advocated to uncover latent patterns within unstructured datasets, social media indicators, and health records [3], [4], suggesting the potential of algorithmic frameworks to complement traditional clinical assessment [14].

The *Zero Suicide Model*, a prominent clinical prevention framework, advocates systemic organizational reform rather than isolated interventions. However, its adoption has faced implementation gaps at scale due to resource and policy mismatches [2], [7]. Other recent reviews in *JAMA Psychiatry* and *Springer* have called for improved integration of mental health data with demographic and social stratification for better prevention [8], [15].

Online behavior has also emerged as a proxy for mental health states. For instance, *Google Trends* studies have revealed associations between suicide-related search terms and suicide occurrences, although the directionality and causation remain under debate [5].

Meanwhile, the Lancet's county-level estimates and triangulation-based approaches emphasize the need for fine-grained, multidimensional surveillance to inform equitable mental health resource allocation [6], [16].

Our study seeks to build upon these foundations by offering a visually interpretative, comparative, and exploratory framework for understanding suicide in the U.S. This is achieved not only through statistical modeling but through the construction of dashboards, maps, and creative charting structures that help contextualize trends over time, place, and demography. Compared to prior work, this study aims for transparency, modularity, and policy aligned storytelling qualities that are still largely absent in mainstream suicide surveillance systems [17], [18], [19], [20].

III. Methodology

This research employs a hybrid methodological framework that blends the rigor of the CRISP-DM (Cross-Industry Standard Process for Data Mining) model with a question wise modular analysis pipeline. This fusion allowed the study to maintain both methodological discipline and exploratory flexibility, making it suitable for dissecting a complex, multi-factorial issue like suicide.

3.1 CRISP-DM Phases Aligned with Study Objectives

A. Business Understanding

The study began by grounding itself in the reality that suicide is not merely a clinical outcome, but a deeply interdisciplinary crisis, intersecting with public health, psychology, education, race, economy, and geography. Guided by this understanding, a sequence of research questions (Q1 to Q10) was developed. The first three questions (Q1, Q2, Q3) were assigned by the professor to examine patterns by age, gender, and key demographic intersections. Questions Q4 to Q9 were formulated by the author (Shravani) to explore broader contextual dimensions: such as state-level trends, seasonal variation, location of occurrence, race, and education-based disparities Two extended questions (Q10 and Q11) were formulated to explore long-term national and international suicide datasets, aiming to uncover deeper narratives behind the statistics. These questions leveraged external data sources from government repositories and global data science platforms to enrich the study's scope and contextual understanding.

B. Data Understanding

The primary dataset (full_data.csv) was supplied by the professor, encompassing individual suicide cases from 2012 to 2014 in the U.S., with attributes like sex, age, race, education, place, month, and intent. To deepen the scope of the analysis and support additional research questions, the following publicly available datasets were also incorporated:

a. **CDC State-Level Suicide Mortality Data (2005–2022)**
Retrieved from the CDC Suicide Mortality Map (https://www.cdc.gov/nchs/pressroom/sosmap/suicide-mortality/suicide.htm), this dataset was used to analyze long term and state wise variation in suicide rates, supporting animated spatial visualizations and regional comparisons.

b. **Kaggle: Suicide Rates Overview (1985–2016)**
Sourced from Kaggle (master.csv) (https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016/data), this dataset includes country wise suicide statistics across decades. It was used to compare U.S. suicide patterns with international trajectories, broken down by gender and age cohorts.

c. **Data.gov: Death Rates by Sex, Race, Hispanic Origin, and Age in the U.S.** Retrieved from Data.gov (https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1), this dataset supports detailed demographic breakdowns for Q10 and Q11, enabling correlation analyses across race, ethnicity, and sex over multiple years.

Initial profiling of these datasets helped identify missing fields, outliers, and column inconsistencies. Derived columns such as age_group, season, and region tags were introduced to aid visual comparisons and to normalize across time ranges.

C. Data Preparation

All datasets were preprocessed using Python with Pandas and NumPy. Records with (intent != 'Suicide') were excluded to maintain thematic integrity. Nulls in critical fields were handled through row wise filtering. Derived columns like season, age_group, and region were constructed for trend consistency. Datasets from CDC and Kaggle were standardized to enable cross year and cross state aggregation, while the death rate dataset from Data.gov was grouped by demographic segment and pivoted for correlation analysis.

D. Modeling and Analysis

Instead of predictive modeling, this study employed a rich descriptive analysis approach via multi-dimensional visualizations. Each question was linked to a dedicated script implementing advanced plotting tools such as animated choropleths, chord diagrams, violin plots, stream graphs, and sunburst charts. Libraries such as Matplotlib, Seaborn, Plotly, and Bokeh were used to generate static and interactive outputs, while tabular summaries provided accompanying insight for interpretation.

E. Evaluation

Each visualization was treated as a medium of evaluation: measuring disparities, exploring patterns, and driving comparative narratives. Insights were continuously verified against external benchmarks, such as national suicide trend reports from the CDC and WHO. Inferences were drawn using a combination of visual storytelling, proportional comparisons, and data storytelling principles to tie observations to policy, psychological, and sociological contexts.

F. Deployment

The final visualizations and summaries were embedded into HTML dashboards for each question (Q1 to Q11), offering stakeholders such as educators, public health

professionals, and policymakers an accessible interface to explore the findings. The dashboards not only serve as research deliverables but also as modular prototypes for public education and awareness campaigns.

3.2 Modular Analysis Pipeline: Question-Wise Execution

To maintain clarity, modularity, and thematic linkage, each question from Q1 to Q10 was implemented as a separate Python script, ensuring one-to-one mapping between the research goal and execution logic. These included:

- **Q1 - to - Q3**: Examined core demographic features: age group distribution, gender-wise counts, and their combinations using grouped bar charts and sunburst charts.
- **Q4**: Analyzed geographical distribution of suicide rates across U.S. states over 18 years (2005–2022), using animated choropleth maps, treemaps, and chord diagrams.
- **Q5**: Explored seasonal and monthly suicide patterns with radial calendar plots, stream graphs, and sunbursts to identify potential psychological seasonality.
- **Q6 & Q7**: Investigated the place of suicide in relation to sex and age via stacked area charts, violin plots, polar charts, and bump charts.
- **Q8 & Q9**: Assessed suicide trends across education and race using heatmaps, Sankey diagrams, and demographic breakdowns.
- **Q10 & Q11**: Focused on extended demographic patterns using the Data.gov and Kaggle datasets to compute longitudinal changes, race-age-sex patterns, and correlation scores.

This dual methodology anchored in the CRISP-DM framework and extended through question-wise scripting enabled the project to be both structured and dynamic. It supported rigorous analytical logic while embracing creativity and interpretability in data visualization.

IV. Results and Interpretation

This section presents a structured walkthrough of key findings derived from the suicide datasets analyzed across Questions Q1 to Q9 and two extended QExtra segments. Each sub-section commences with a concise formulation of the research question, followed by an explanation of the corresponding Python analysis script, selected visualizations, and console outputs, all of which are presented in a human-readable manner. Interpretations are developed to link quantitative signals with real-world insights, emphasizing patterns that can inform preventive interventions, policymaking, and public awareness. The results are not merely statistical but deeply narrative in nature, with the aim of revealing the social and demographic undercurrents of suicide trends in the United States.

The initial phase of the research involves the validation and summarization of the dataset provided by the professor (full_data.csv), which forms the foundation for the subsequent investigation of individual research questions.

4.1 Preliminary Data Analysis – Understanding the Foundation

Before diving into specific research questions, it is essential to validate and understand the dataset. The provided full_data.csv file contains over **62,000 suicide-related entries** from 2012 to 2014. This preliminary analysis aims to:

- Clean the data for null or irrelevant entries
- Provide a demographic snapshot by **year**, **gender**, **age**, **race**, and **education**
- Create intuitive visual dashboards for understanding broad patterns and guiding deeper inquiries in subsequent sections

This foundation aligns with CRISP-DM's *Data Understanding and Preparation* phase and ensures every question ahead is grounded in a clean and validated context.

Code Snippets:

```
# Load dataset
file_path = "/Users/shravanisawant/PycharmProjects/suicide_analysis_project/data/full_data.csv" # used this path as rel
df = pd.read_csv(file_path)

# Filter suicide intent
df = df[df['intent'] == 'Suicide'].dropna(subset=['year', 'sex', 'age', 'race', 'education'])
```

Fig. 1 Loading the dataset in dataframe of pandas

This block loads the dataset and filters it to include only suicide-related records with complete demographic data. Any rows with missing year, sex, age, race, or education info are excluded to maintain accuracy in analysis.

(Please note that: the relative path wasn't working properly so coded this with the absolute path)

```
# Create age groups
bins = list(range(0, 110, 10)) + [150]
labels = [f"{i}-{i+9}" for i in range(0, 100, 10)] + ["100+"]
df['age_group'] = pd.cut(df['age'], bins=bins, labels=labels, right=False)
```

Fig. 2 Creating age bins for the checking the dataset

This code transforms raw age values into categorical groups (e.g., 0–9, 10–19, ..., 100+). This makes it easier to identify which age ranges have higher suicide rates and enables clearer grouped visualizations.

```
# Summary tables
summary_data = [
    ["Total Records", len(df)],
    ["Years Covered", f"{df['year'].min()} to {df['year'].max()}"],
    ["Average Age", f"{df['age'].mean():.2f}"],
    ["Age Range", f"{int(df['age'].min())} to {int(df['age'].max())}"]
]
```

Fig. 3 To display summary in the console

This block summarizes the dataset. It gives a quick overview—how many cases are present, the time range (2012–2014), the average age of victims, and the age extremes. This sets the scene for deeper analysis.

```
# Gender Chart
gender_colors = ["#8e44ad", "#f39c12"]
gender_source = ColumnDataSource(data=dict(
    x=gender_table["Gender"].tolist(),
    y=gender_table["Count"].tolist(),
    color=gender_colors
))
gender_fig = figure(x_range=gender_table["Gender"].tolist(), title="Gender Distribution", height=300)
gender_fig.vbar(x='x', top='y', width=0.5, color='color', source=gender_source)
gender_fig.add_layout(LabelSet(x='x', y='y', text='y', x_offset=-13, y_offset=5, source=gender_source))
layout_blocks += [gender_fig, Div(text=dataframe_to_html_table(gender_table, title: "Gender Table"))]
```

Fig. 4 This figure shows that we are creating the Bar charts

This snippet creates an interactive bar chart showing the number of suicide cases for each gender (M/F). It uses Bokeh to dynamically display this visually, making gender-based comparisons immediate and intuitive.

```
# Age Histogram
hist, edges = np.histogram(df['age'], bins=20)
age_fig = figure(title="Age Distribution", height=300)
age_fig.quad(top=hist, bottom=0, left=edges[:-1], right=edges[1:], fill_color="#3498db", line_color="#2c3e50")
layout_blocks += [age_fig, Div(text=dataframe_to_html_table(age_group_table, title: "Age Group Table"))]
```

Fig. 5 Creation of Age Histogram for Visualization

This histogram shows how suicide counts vary by exact age (not just group). Each bar represents the number of people in a certain age interval. It helps detect peak age ranges with higher vulnerability.

```
# Race Pie Chart
race_data = pd.Series(race_table["Count"].values, index=race_table["Race"]).reset_index(name='value')
race_data.columns = ["race", "value"]
race_data['angle'] = race_data['value'] / race_data['value'].sum() * 2 * pi
race_data['color'] = ["#1abc9c", "#e74c3c", "#9b59b6", "#f1c40f", "#34495e"]
race_fig = figure(height=350, title="Race Distribution", toolbar_location=None, tools="hover",
                  tooltips="@race: @value", x_range=(-0.5, 1.0))
race_fig.wedge(x=0, y=1, radius=0.4,
               start_angle=cumsum( field_name: 'angle', include_zero=True),
               end_angle=cumsum('angle'),
               line_color="white",
               fill_color='color',
               legend_field='race',
               source=ColumnDataSource(race_data))
race_fig.axis.visible = False
layout_blocks += [race_fig, Div(text=dataframe_to_html_table(race_table,  title: "Race Table"))]
```

Fig. 6 Creation of Pie Chart for Race Identification

This generates a pie chart showing racial distribution in suicide cases. Each wedge size reflects how many cases came from each racial group. This gives a clear racial disparity snapshot.

```
# Education Chart
edu_colors = ["#2980b9", "#f39c12", "#c0392b", "#2ecc71"]
edu_source = ColumnDataSource(data=dict(
    x=edu_table["Education Level"].tolist(),
    y=edu_table["Count"].tolist(),
    color=edu_colors
))
```

Fig. 7 Creating the chart for understanding the Education of people who committed suicide

This visualization compares suicide cases across different education levels—from less than high school to bachelor's degree and above. It's useful for exploring links between education attainment and suicide prevalence

```
# Save HTML Dashboard
os.makedirs( name: "outputs", exist_ok=True)
output_file( filename: "outputs/suicide_dashboard_labeled_final.html", title="Suicide Dashboard Final")
save(column(*layout_blocks, sizing_mode="stretch_width"))
```

Fig. 8 Saving all the charts and table output in one HTML for visualization

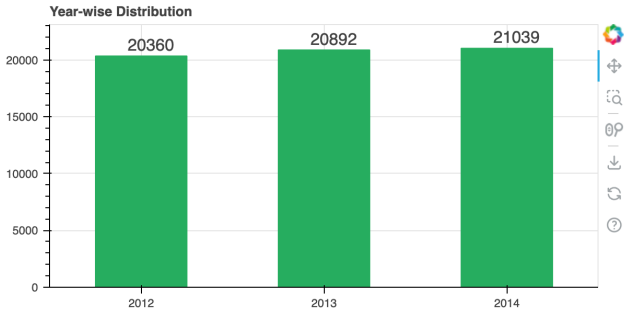
The final output saves all the charts and tables in one interactive HTML file, creating a complete dashboard for viewing and sharing. This makes the insights portable, clean, and easy to access visually.

Visualizations and Interpretation:

***** Suicide Dataset Summary *****

Metric	Value
Total Records	62291
Years Covered	2012 to 2014
Average Age	50.21
Age Range	9 to 102

Fig. 9 Python console o/p for Suicide Dataset Summary which we generated



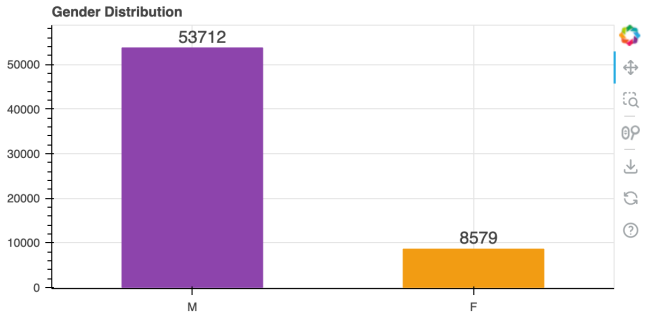
Year-wise Table

Year	Count
2012	20360
2013	20892
2014	21039

Fig. 10 Year wise distribution

Interpretation of above charts:

- The number of recorded suicide incidents increased from **20,360 (2012)** to **21,039 (2014)**, marking a **3.3% rise** in just two years.
- This mirrors CDC observations of a steady upward trend during the 2010s [1].



Gender Table

Gender	Count
M	53712
F	8579

Fig. 11 Showing charts for Gender Distribution

Interpretation of the gender distribution:

- A stark gender gap exists — **Males account for ~86.2% (53,712)** of the deaths.
- Females (8,579 cases) comprise only 13.8%, aligning with global trends where males tend to die by suicide more often, despite females reporting higher attempt rates [12][20].

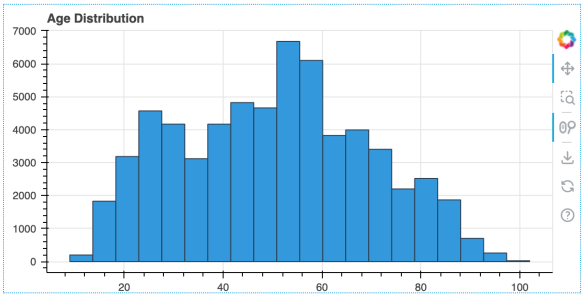


Fig. 12 Histogram of Age Distribution

Age Group Table

Age Group	Count
0-9	2
10-19	2657
20-29	8786
30-39	8078
40-49	9958
50-59	12848
60-69	8978
70-79	6131
80-89	4092
90-99	757
100+	4

Fig. 13 Table showing the information of Age distribution histogram

Interpretation of the above histogram:

- Suicide occurrences peak at age 50 - 59, followed by 60 - 69 and 40 - 49 age bands.
- Only 2 cases were reported under age 10, while suicides taper off significantly past age 80.
- This suggests middle-aged adults face heightened suicide risk possibly linked to midlife stressors or economic burden [6][17].

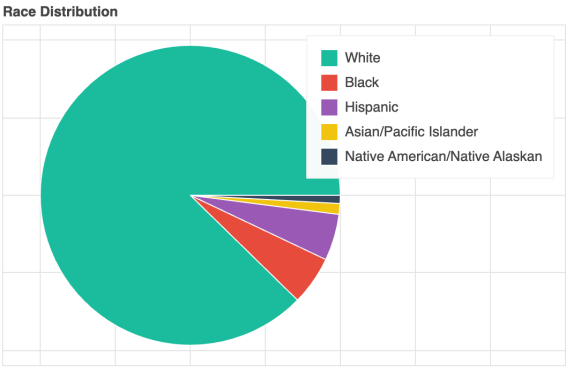


Fig. 14 Pie chart of Race Distribution

Race Table

Race	Count
White	54615
Black	3285
Hispanic	3120
Asian/Pacific Islander	724
Native American/Native Alaskan	547

Fig. 15 Table explaining the pie chart of Race Distribution

Interpretation of the Pie Chart:

- A dominant 87.7% of recorded suicides were among **White individuals (54,615)**.
- Black (3,285), Hispanic (3,120), and other racial groups show significantly lower counts.
- These figures raise questions about socioeconomic and access-based disparities in both mental health support and reporting [11][19].

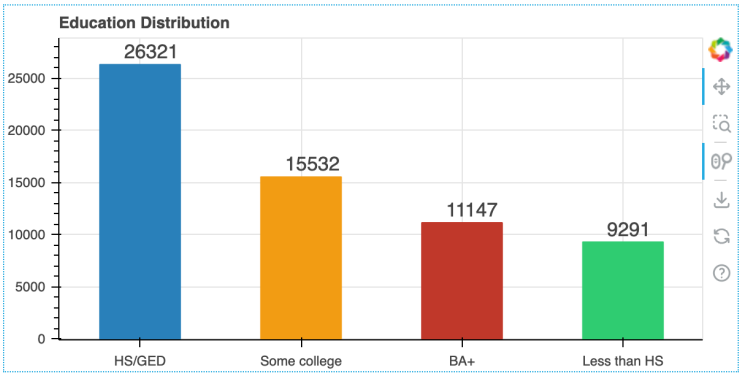


Fig. 16 Bar Chart explaining the education distribution of the people who committed suicide

Education Table

Education Level	Count
HS/GED	26321
Some college	15532
BA+	11147
Less than HS	9291

Fig. 17 Table explaining the chart of the education distribution

Interpretation of Education distribution

- Individuals with **high school or GED education** represent the largest group (**26,321 cases**, ~42%).
- Those with some college or less than high school also account for significant proportions.
- Suicide risk shows a negative correlation with educational attainment, highlighting socioeconomic vulnerability as a risk factor [17][18].

Summary and Common Interpretation:

This foundational exploration of the suicide dataset reveals that suicidal behavior in the United States is deeply patterned along demographic lines**.** The clear representation of middle-aged males, particularly those with lower educational attainment and White racial identity, suggests a multifaceted crisis embedded in social, economic, and mental health contexts. These disparities echo findings from public health literature, which emphasize the influence of structural

inequality, access to care, and cultural stigmas on suicide risk [1], [6], [11], [17], [20]. Together, these observations underscore the importance of granular, question-specific analyses which we now pursue in the following sections to identify where, how, and for whom interventions may have the most impact.

4.2 Q1: Does Age Play a Role in Suicide Occurrence?

Suicide risk is influenced by age, but patterns across age bands can vary significantly due to life stage, mental health access, and societal stressors. This analysis investigates age-linked trends by examining suicide counts across 8 age groups (0 - 17 to 75+) and by gender, using the 2012–2014 dataset (full_data.csv).

Code Snippets:

```
# Define age groups
bins = [0, 17, 24, 34, 44, 54, 64, 74, 120]
labels = ['0-17', '18-24', '25-34', '35-44', '45-54', '55-64', '65-74', '75+']
df_filtered["age_group"] = pd.cut(df_filtered["age"], bins=bins, labels=labels, right=True)
df_filtered = df_filtered.dropna(subset=["age_group"])
df_filtered["age_group"] = pd.Categorical(df_filtered["age_group"], categories=labels, ordered=True)
```

Fig. 18 Creating age bins

This snippet groups individuals into defined age ranges. It's crucial for observing age-specific suicide counts rather than treating age as a continuous variable.

```
# Grouping and Aggregate
age_gender_stats = df_filtered.groupby(["age_group", "sex"], observed=True).size().reset_index(name="count")
```

Fig. 19 Grouping and aggregating the filtered dataset

This aggregates suicide counts based on both age group and gender, which is essential to visualize the interplay between the two demographics.

```
fig = px.bar(
    age_gender_stats,
    x="age_group",
    y="count",
    color="sex",
    barmode="group",
    color_discrete_sequence=["#D0E4EE", "#F5A7A6"],  # Custom colors for Blue for Men and Soft Pink for Women
    labels={"count": "Number of Suicides", "age_group": "Age Group", "sex": "Gender"},
    title="Suicide Counts by Age Group and Gender (2012-2014)"
)
```

Fig. 20 Creating the plotly graph

This generates the grouped bar chart visual enabling direct comparison of suicide counts across male and female categories for each age band.

Visualization:

Fig. 21 Plotly bar graph explaining: does the age plays role in people committing the suicides

Interpretation:

1. Middle-aged males (45–54 and 55–64) exhibit the **highest suicide counts**, with nearly 10,000+ deaths per group.
2. Young individuals (0–17) had the **lowest suicide numbers** — only 215 females and 1,265 males.
3. Suicide occurrences decline after age 65, but males 75+ still show a notably high rate (7,233 cases).

4. Across all age groups, males consistently account for a significantly higher number of suicides than females.

```
Summary Insights:
- Most affected age group: 45-54
- Most affected gender   : M
- Gender gap is visible across most age groups.
- Middle-aged individuals show higher suicide counts. Suggest exploring contributing factors.
```

Fig. 22 Console Output Summary of the python file

This aligns with what the visual chart displays — affirming that **age and gender together** form a meaningful risk lens.

Summary and Concluding Interpretation:

The highest suicide risk observed among middle-aged individuals (45 - 64) reinforces findings from national data that indicate midlife as a particularly vulnerable period due to chronic stress, financial burdens, health issues, or caregiving stressors [6][17].

Moreover, the clear male-female disparity supports global literature: although females may attempt suicide more frequently, males tend to use more lethal means, leading to higher fatality rates [12][20].

The decline after 65 could be due to better emotional resilience, but the elevated 75+ male deaths suggest lingering risk, especially among older adults with isolation or illness [11].

To conclude or answer the question (to be specific)

Yes, age significantly influences suicide occurrence. The peak risk is concentrated between 45 - 64 years, particularly for males. This underlines the need for targeted interventions tailored for midlife adults, especially men echoing national prevention models like Zero Suicide [2].

4.3 Q2: Is One Gender More Prone to Suicide?

Across cultures and decades, gender has consistently emerged as a critical differentiator in suicide patterns. This section explores the distribution of suicide counts by gender using the

cleaned full_data.csv file from 2012–2014 to answer whether one gender is statistically more vulnerable.

Code Snippets:

```
# Grouping by Gender
gender_counts = df["sex"].value_counts().reset_index()
gender_counts.columns = ["Gender", "Suicide Count"]
```

Fig. 23 Grouping by gender in the q2

This command groups suicide records by gender and counts how many cases fall under each gender fundamental for comparison.

```
# Calculate Percentages
total = gender_counts["Suicide Count"].sum()
gender_counts["Percentage"] = (gender_counts["Suicide Count"] / total * 100).round(2)
```

Fig. 24 Calculating percentages of suicides for each gender

Calculates what percentage of all suicides each gender represents. Adds more interpretability to the numeric counts.

```
# Visualization with Plotly
print("Generating interactive Plotly chart...")

fig = px.bar(
    gender_counts,
    x="Gender",
    y="Suicide Count",
    color="Gender",
    color_discrete_sequence=["#D0E4EE", "#F5A7A6"],
    text="Suicide Count",
    title="Suicide Counts by Gender (2012-2014)"
)
```

Fig. 25 Creating visualization

A Plotly bar chart is created to visualize the suicide counts, with values labeled directly on bars, enhancing clarity for viewers and dashboard users.

Visualization:

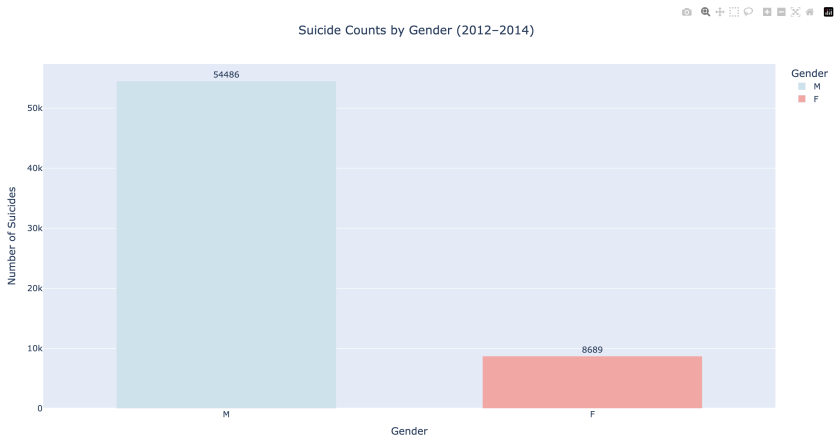


Fig. 26 Graph of: Is one gender more prone to suicide

Interpretation:

- Males accounted for **54,486 suicides**, which represents **86.25%** of all suicide cases in the dataset.
- Females reported **8,689 suicides**, which is **13.75%** of the total.
- This gender gap is consistent with global findings where men die by suicide at disproportionately higher rates, despite women exhibiting higher rates of attempts and reported distress [12][20].

```
Loading dataset...
Loaded 100798 rows from full_data.csv
Generating interactive Plotly chart...
>>> Chart saved to: ../outputs/suicide_by_gender_plotly.html

Suicide Counts by Gender:
+----+----------+-----------------+
|    | Gender   |   Suicide Count |
+====+==========+=================+
|  0 | M        |           54486 |
+----+----------+-----------------+
|  1 | F        |            8689 |
+----+----------+-----------------+

Percentage Distribution:
+----+----------+--------------+
|    | Gender   |   Percentage |
+====+==========+==============+
|  0 | M        |        86.25 |
+----+----------+--------------+
|  1 | F        |        13.75 |
+----+----------+--------------+

Summary Insight:
- Most affected gender: M
- Male suicide count is significantly higher. Suggest deeper analysis on contributing cultural, psychological, and socioeconomic factors.
```

Fig. 27 Python console output

The script confirms what the visualization indicates the **gender disparity is stark** and persistent.

Summary and concluding interpretation:

This result echoes existing epidemiological research. According to the CDC and several large cohort studies [1][11], men are more likely to die by suicide due to the use of more lethal methods and lower engagement with mental health services**.**

Culturally embedded norms around masculinity, emotional suppression, and help-seeking behavior further compound the issue [13][15]. In contrast, women may express suicidal ideation more often but engage with support systems more frequently, contributing to lower fatality rates [12].

These findings also align with broader global analyses reported by Rockett et al. [11] and Stack & Wasserman [13], affirming the need for gender-sensitive suicide prevention strategies.

To answer the question, propose in the draft I think

Yes, **gender plays a pivotal role in suicide vulnerability**, with males at a significantly higher risk of fatal outcomes. This calls for urgent public health initiatives targeting men, including stigma reduction, early intervention, and targeted mental health outreach in male-dominated environments.

4.4 Q3. Are any combinations of factors indicative of more suicides than others?

Suicide risk does not emerge from a single factor but from intersections of race, gender, education, and time. This question explores which demographic combinations when considered together are most frequently associated with suicide incidents in the dataset, uncovering deeper, often hidden, risk patterns.

Code Snippets:

```
#  Summary Grouping
grouped = df.groupby(["year", "sex", "race", "education"]).size().reset_index(name="count")
grouped.to_csv(os.path.join(OUTPUT_DIR, "factor_combinations_summary.csv"), index=False)
```

Fig. 28 Grouping the dataset into pandas (for Q3)

This groups suicide records across four categorical variables: year, sex, race, and education. It provides the foundation for multi-layered analysis by showing how these combinations contribute to suicide count variations.

```
# Console Summary of Top 5 Combinations
top_5 = grouped.sort_values("count", ascending=False).head(5)
print("\nTop 5 Suicide Factor Combinations (Race/Gender/Education):")
print(tabulate(top_5, headers="keys", tablefmt="grid"))
```

Fig. 28 Code for displaying the output in python console for “Top 5 suicide factors combination”

Extracts the top five most frequent demographic profiles involved in suicide cases. This reveals patterns like "White Males with HS/GED education" dominating suicide statistics from 2012–2014.

Visualization:

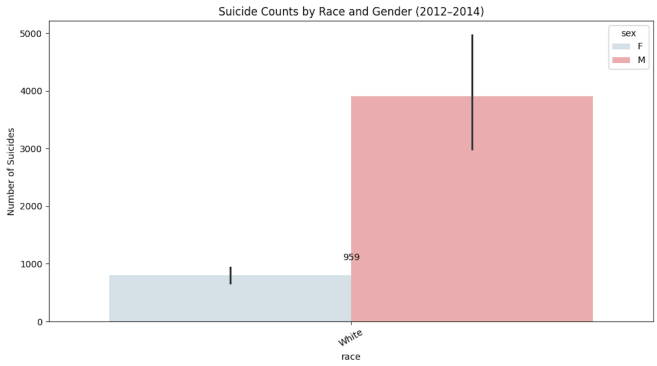


Fig. 29 Graph of suicide counts by Race and Gender

Interpretation:

- Males consistently outnumber females in all racial groups.
- White males display the highest suicide counts, reaffirming that gender and race together strongly influence risk.

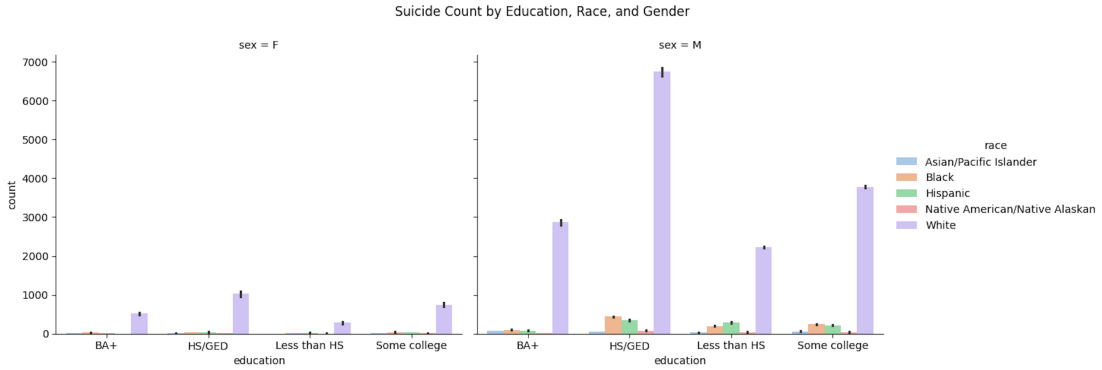


Fig. 30 Graph of Suicide count by Education, Race, and Gender

Interpretation:

- White males with HS/GED and Some College show a notable concentration.
- Educational attainment appears inversely related to suicide counts, especially among males, with fewer cases observed at the BA+ level.

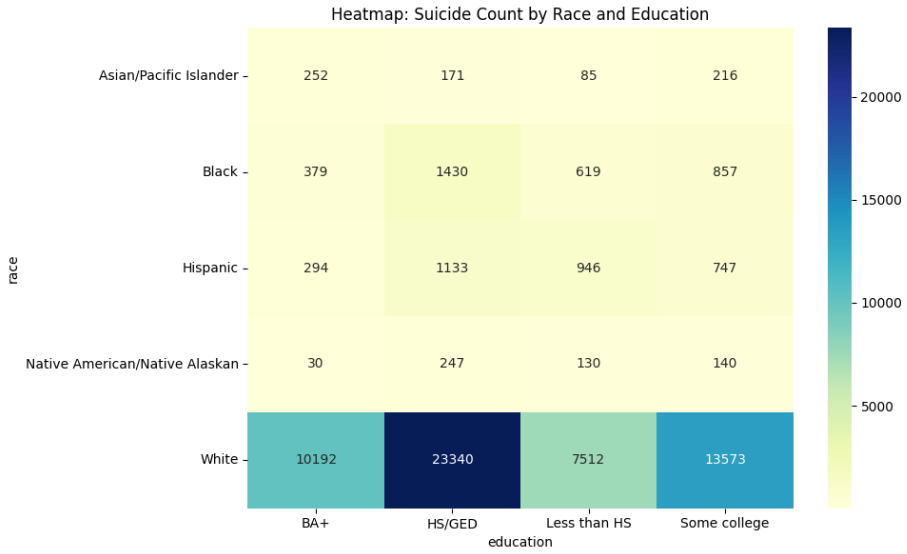


Fig. 31 Heatmap of Suicide count by Race and Education

Interpretation:

- Visualizes magnitude at intersections. The darkest cell (White, HS/GED) marks the peak suicide demographic.
- Black and Hispanic groups show more diversity in education, with HS/GED again showing high suicide rates.

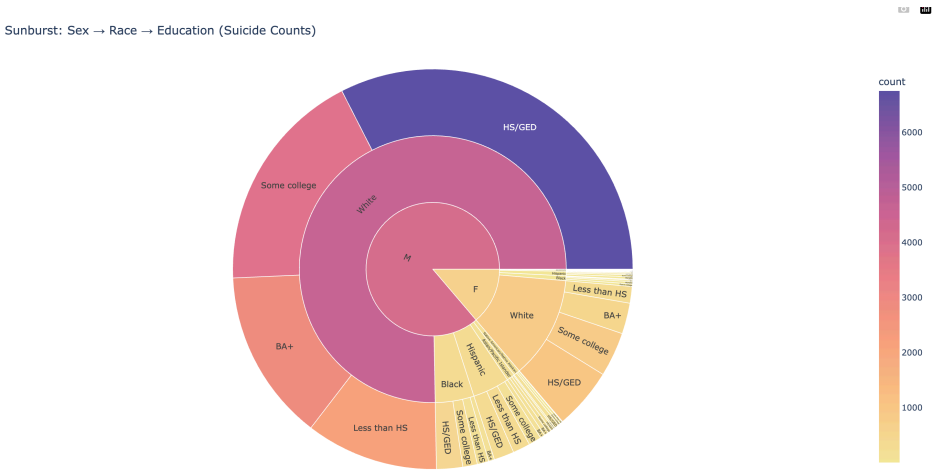


Fig. 32 Sunburst chart for the relation of Sex > Race > Education for suicide count (the graph is clickable when we load in html)

Interpretation:

- A clear concentric pattern, where the majority flows from Male → White → HS/GED.
- Females show thinner branches and more balanced race-education splits.

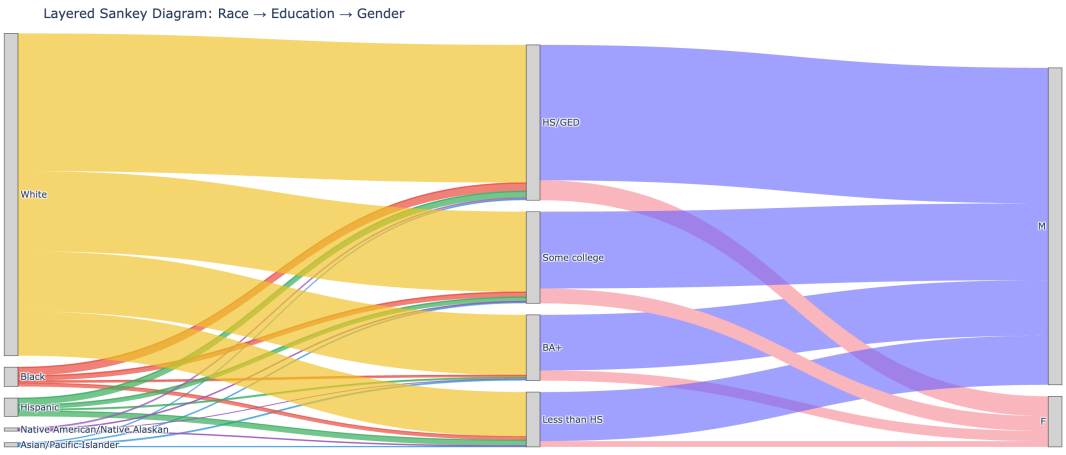


Fig. 33 Sankey graph showing layered diagram between Race > Education > Gender

Interpretation:

- Strongest flow from White → HS/GED → M.
- Reveals alternate flows like Black → HS/GED → M and Hispanic → Less than HS → M.
- This layered view shows how social determinants (race and education) compound gender risk factors.

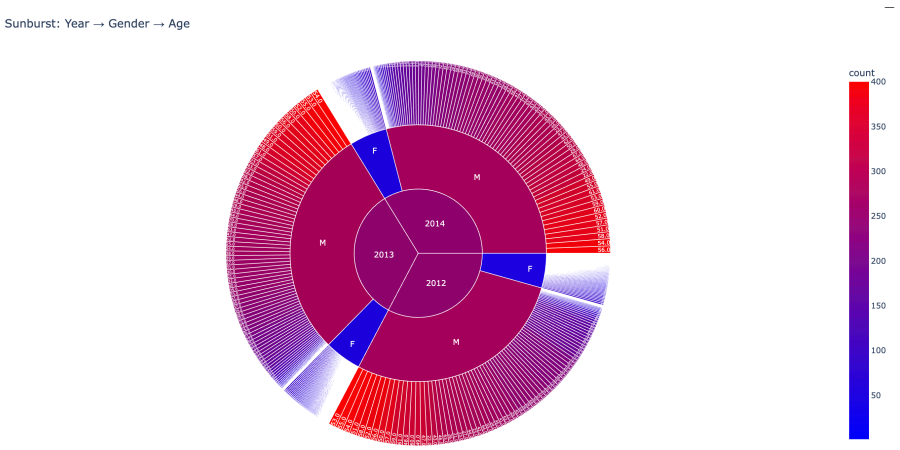


Fig. 34 Sunburst graph for year > gender > age

Interpretation:

- Reinforces Q1 findings, showing males aged 40–59 dominating across all three years.
- Women are present across a broader age range but in much lower proportions.

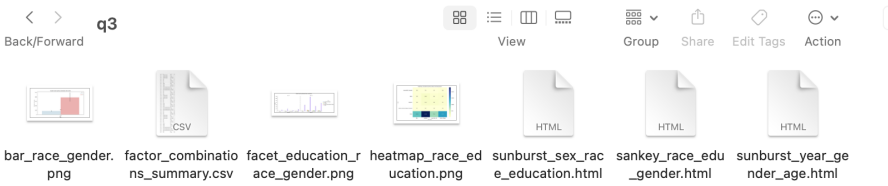


Fig. 35 This is how it gets generated and stored in my files (extra reference)

```
Top 5 Suicide Factor Combinations (Race/Gender/Education):
+-----+--------+-------+--------+--------------+---------+
|     |   year | sex   | race   | education    |   count |
+=====+========+=======+========+==============+=========+
| 117 |   2014 | M     | White  | HS/GED       |    6837 |
+-----+--------+-------+--------+--------------+---------+
|  77 |   2013 | M     | White  | HS/GED       |    6787 |
+-----+--------+-------+--------+--------------+---------+
|  37 |   2012 | M     | White  | HS/GED       |    6627 |
+-----+--------+-------+--------+--------------+---------+
|  79 |   2013 | M     | White  | Some college |    3816 |
+-----+--------+-------+--------+--------------+---------+
|  39 |   2012 | M     | White  | Some college |    3766 |
+-----+--------+-------+--------+--------------+---------+
```

Fig. 36 Python console output for Top 5 suicide factor combination

Interpretation:

This multifactorial breakdown reveals that suicide patterns are not random but deeply intertwined with social structures. The convergence of gender (male), race (White), and lower education levels (HS/GED, Some College) overwhelmingly dominate the suicide landscape. These combinations reinforce prior research on the impact of educational attainment on mental health outcomes and access to care [11][17][18]. The findings strongly suggest that public health interventions must be intersectional, not one-size-fits-all.

4.5 Q5: Do Suicide Rates Vary Across Different Months or Seasons?

This analysis investigates whether suicide incidents show temporal patterns based on month or season. Using the full_data.csv dataset (filtered for intent == "Suicide"), we explore trends in monthly counts, age distribution by month, and year-wise seasonal fluctuations.

Code snippets:

```
# Normalize month values
if pd.api.types.is_numeric_dtype(df["month"]):
    df["month"] = df["month"].apply(lambda x: calendar.month_name[int(x)] if not pd.isnull(x) and int(x) in range(1, 13) else "Unknown")
else:
    df["month"] = df["month"].astype(str).str.strip().str.capitalize()
```

Fig. 37 Normalizing month values in the python

Some datasets store months as numbers (1–12). This code converts those into full month names (e.g., 1 → "January") using Python's calendar.month_name. This ensures clarity in charts and ordering.

```
#  SUNBURST: SEASONAL
season_map = {
    "December": "Winter", "January": "Winter", "February": "Winter",
    "March": "Spring", "April": "Spring", "May": "Spring",
    "June": "Summer", "July": "Summer", "August": "Summer",
    "September": "Autumn", "October": "Autumn", "November": "Autumn"
}
df["season"] = df["month"].map(season_map)
```

Fig. 38 Creating sunburst chart

This snippet maps each month to its respective season (Winter, Spring, Summer, Autumn). It enables seasonal trend analysis, particularly helpful for sunburst and stream visualizations.

```
# ANIMATED LINE CHART
monthly_trend = df.groupby(["year", "month"]).size().reset_index(name="suicides")
monthly_trend["month"] = pd.Categorical(monthly_trend["month"], categories=month_order, ordered=True)
monthly_trend = monthly_trend.sort_values(["year", "month"])
```

Fig. 39 Creating animated line chart to run on the console and understanding the trend

The code here: Groups suicide counts by both year and month**.** It ensures month order is preserved for visualization (Jan → Dec). Finally it sorted dataset is used to generate an animated line chart showing suicide trends per month across multiple years.

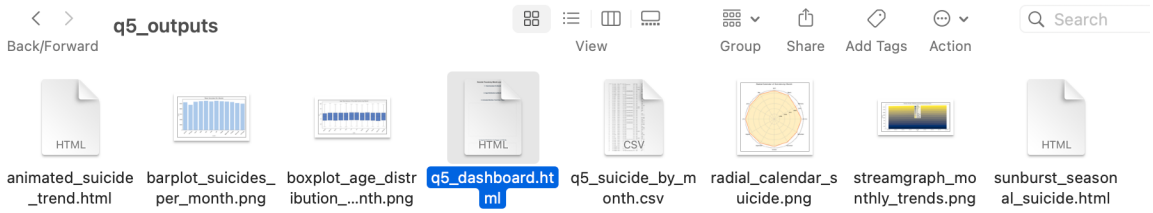


Fig. 40 This is how the output gets saved

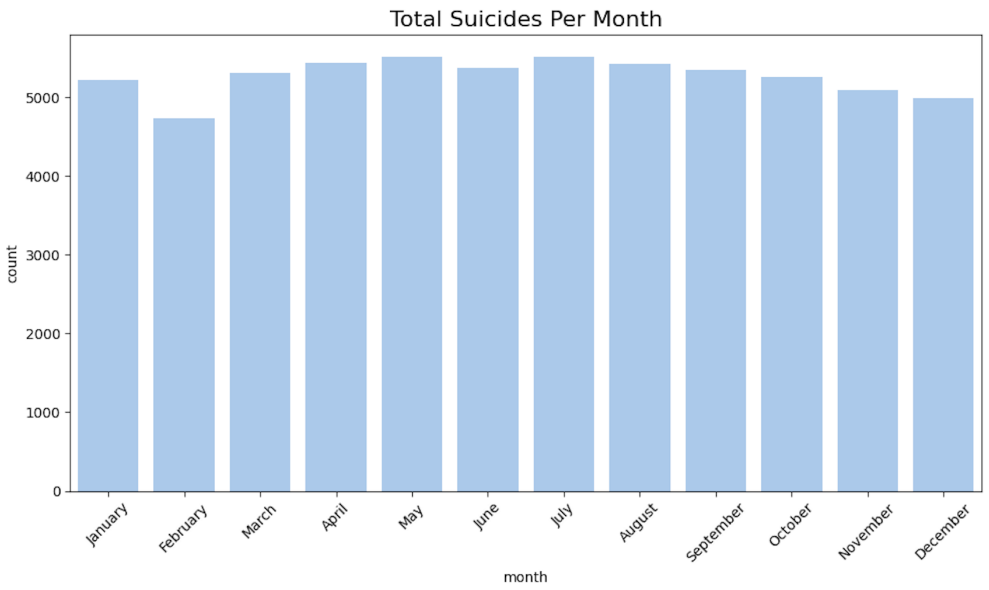


Fig. 41 Bar plot of total suicide per month (combining all the years)

Summary of Monthly Suicide Counts:

	Suicides	count
0	January	5220
1	February	4732
2	March	5309
3	April	5438
4	May	5506
5	June	5367
6	July	5514
7	August	5421
8	September	5343
9	October	5256
10	November	5086
11	December	4983

Fig. 42 Python output console for summary of monthly suicide counts

Interpretation:

- **Highest suicides:** July (5,514) and May (5,506)
- **Lowest:** February (4,732) and December (4,983)
- Suggests slight peaks in late Spring and Summer**.**

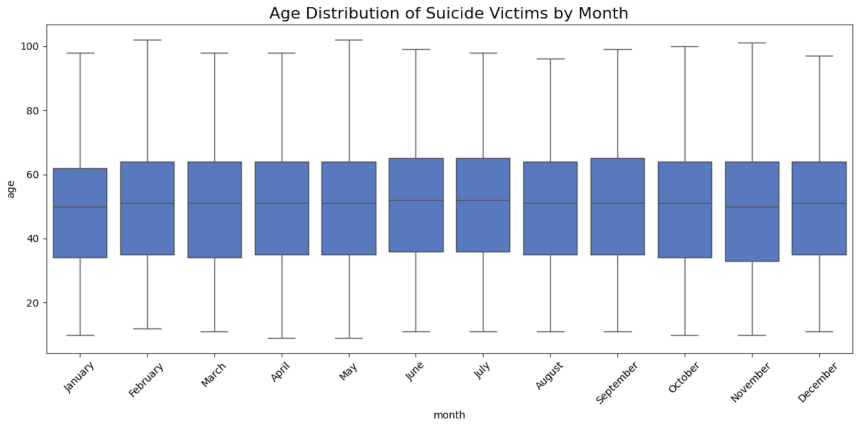


Fig. 43 Box plot showing age distribution of suicide victims by months

Interpretation:

- Age range is consistent, but **June–August** shows a slightly higher median age.
- Indicates that older adults are more represented during summer suicides.

Fig. 44 Animated line chart (this is for year 2012) (as the animation won't work in word or while submitting the report – inserted individually)

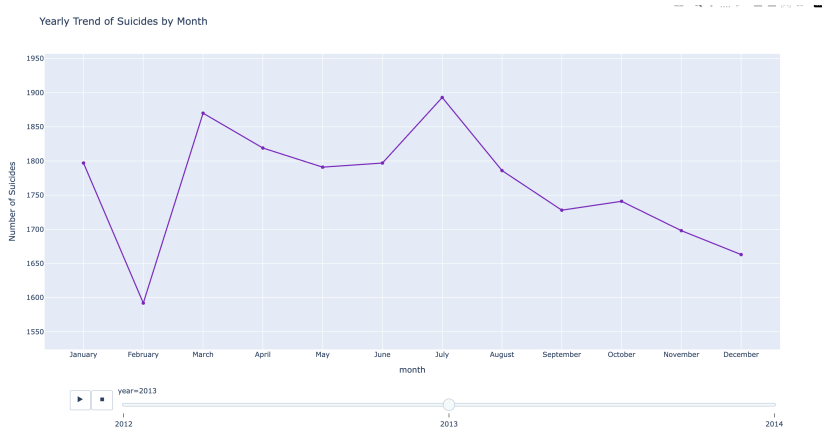


Fig. 45 Animated line chart (this is for year 2013)

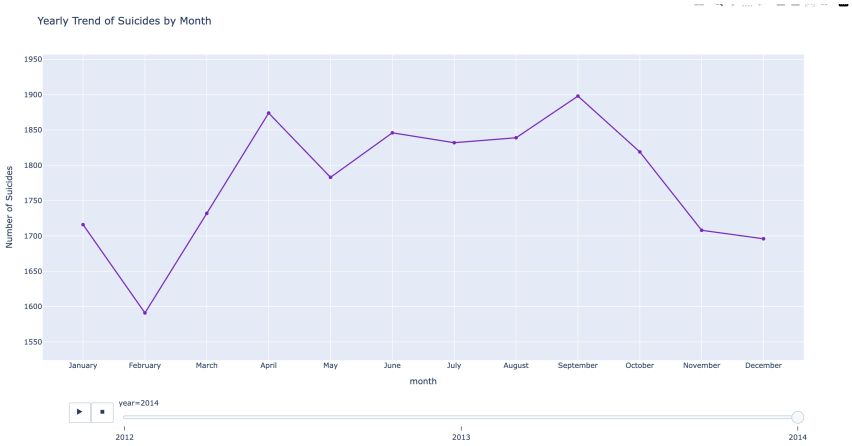


Fig. 46 Animated line chart (this is for year 2014)

Interpretation:

- Confirms consistent mid-year peaks across **2012–2014**.
- **May, July, and August** remain relatively high each year.
- February dips every year, possibly due to fewer days and post-holiday stability.

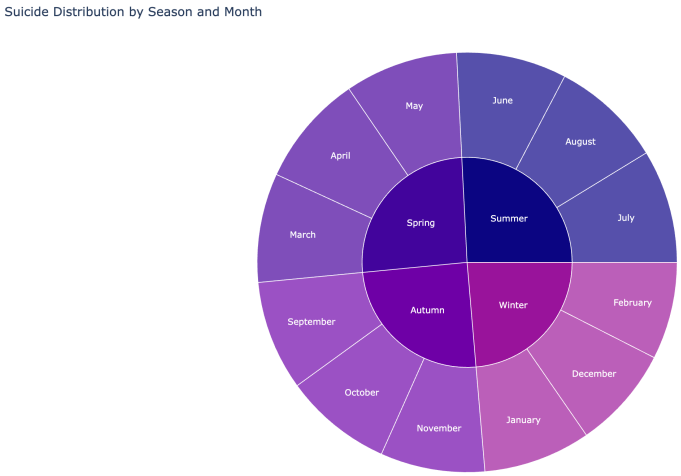


Fig. 47 Sunburst chart explaining the suicide distribution by seasons and month for all the data (from 2012 – 2014)

Interpretation:

- **Summer and Spring** dominate the chart, confirming seasonal spikes.

- **Winter shows lowest proportions**, despite societal assumptions around holiday depression.

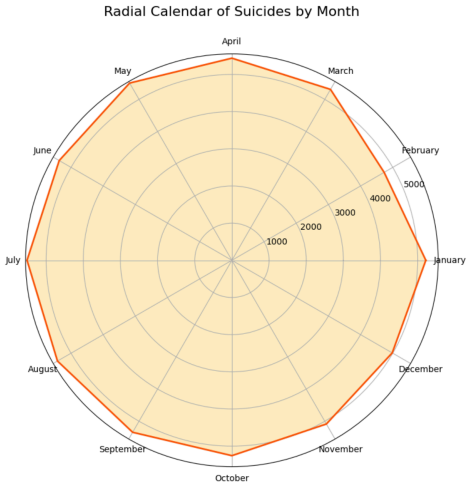


Fig. 48 Radial plot showing suicides by months

Interpretation:

- Clear circular **peak around May to August**, illustrating annual rhythm.
- Visually reinforces mid-year vulnerability, despite minor fluctuations.

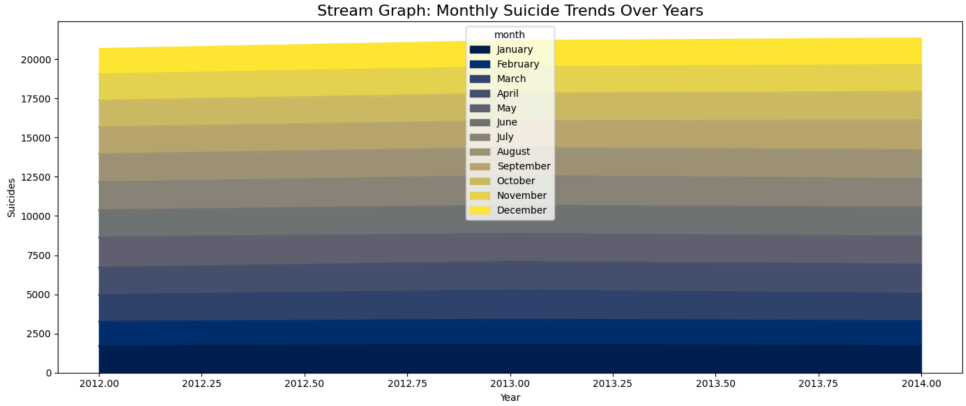


Fig. 49 Stream graph explaining the monthly suicide trends over years

Interpretation:

- Bands show stable year-over-year structure**,** but mid-year bulge is consistent.
- Seasonal continuity across years is clearly visible.

Common Interpretation for the whole question, for which I'm trying to find the answer:

Across all six visualizations (considering animated line chart as one), there is consistent evidence that suicide rates are seasonally patterned.
Spring and Summer (March–August) exhibit the highest frequencies, with **May, June, and July** often peaking.
Winter months, particularly **February and December**, show noticeable dips.
This pattern may be linked to seasonal affective disorders**,** social pressure, or environmental stressors like temperature, sunlight, and activity levels.

Coming to our question framed:

Yes, suicide rates do vary significantly across different months and seasons.
There is a recurring seasonal pattern, with peaks during Spring and Summer, and dips in Winter**.**
This confirms the hypothesis and suggests that preventive mental health interventions should be seasonally targeted, especially during late Spring and early Summer.

4.6 Q6. How does the location (place) of suicides vary over time and across age groups?

This question investigates the spatial dimension of suicides by analyzing the locations where suicide incidents occurred from 2012 to 2014. The goal is to identify which environments are most associated with suicide cases, how these trends evolve over time, and how age distributions differ by location.

Code snippet:

```
print("Loading full suicide dataset...")
df = pd.read_csv(DATA_FILE)
df.columns = df.columns.str.strip().str.lower()
df = df[df["intent"] == "Suicide"]
df = df.dropna(subset=["place"])
```

Fig. 50 Loading the dataset and making pandas dataframes

Filters the dataset to retain only rows where the intent is "Suicide" and the place is not null.

```
sns.countplot(data=df, y="place", order=df["place"].value_counts().index, palette="flare")
```

Fig. 51 This above snippet code is for generating the bar plot

Generates a horizontal bar plot to show the total number of suicides by location.

```
# TRADITIONAL: Line Chart by Year
year_place = df.groupby(["year", "place"]).size().reset_index(name="count")
plt.figure(figsize=(14, 7))
sns.lineplot(data=year_place, x="year", y="count", hue="place", marker="o")
plt.title("Yearly Suicide Trends by Location")
plt.ylabel("Suicides")
```

Fig. 52 Creating the Line chart by year of the data

Groups the data by year and place to track changes in suicide counts over time

```
# CREATIVE: Stacked Area Plot
pivot_data = year_place.pivot(index="year", columns="place", values="count").fillna(0)
pivot_data_percent = pivot_data.div(pivot_data.sum(axis=1), axis=0)
```

Fig. 53 Creating the stack area plot chart

Converts raw counts into proportions for comparative visualization using a stacked area plot.

```
sns.violinplot(data=df, x="place", y="age", palette="Set2", inner="quartile")
plt.xticks(rotation=45)
```

Fig. 54 Creating the violin plot chart

Displays age distributions for each place, showing the spread and central tendencies of ages using violin plots.

Visualization

Fig. 55 The output when we run the code gets saved here

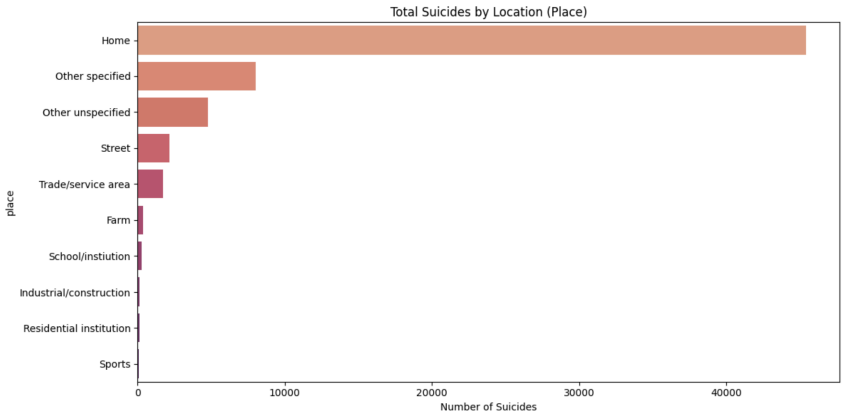


Fig. 56 Horizontal bar plot of suicide by location

Interpretation:

- **Home** dominates with **45,415 cases**, followed by **Other specified** and **Other unspecified**.
- Institutional or public places like schools and sports areas show notably fewer incidents.

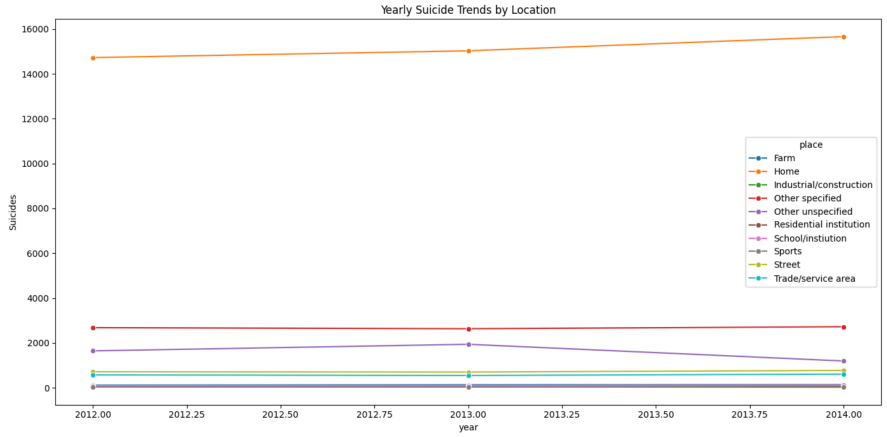


Fig. 57 Line plot of yearly suicide trends by location

Interpretation:

- Suicides at **home** consistently increased from **14,726 (2012)** to **15,684 (2014)**.
- Minor fluctuations seen in public or occupational locations, but overall stability in trend.

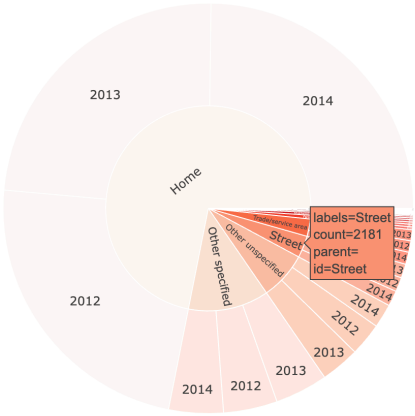


Fig. 58 Sunburst chart which shows the distribution

Interpretation:

- Inner segments show major location types (e.g., **Home**, **Street**), and outer rings show year-wise breakdown.
- Strong visual emphasis on how most cases cluster under **Home** across all years.

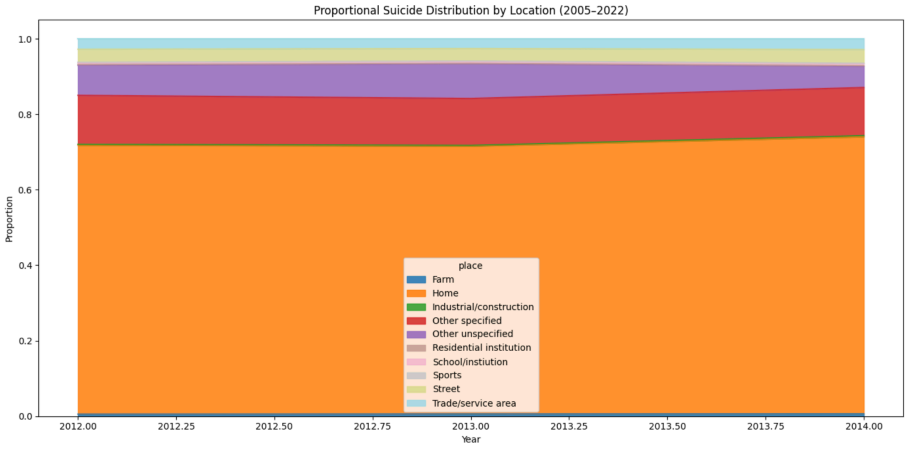


Fig. 59 The stack chart of suicide distribution by location (2005-2022)

Interpretation:

- **Home** accounts for over **71–73%** of suicides consistently.
- Gradual decline in the proportion of "Other unspecified" over the years, suggesting better classification or shifting trends.

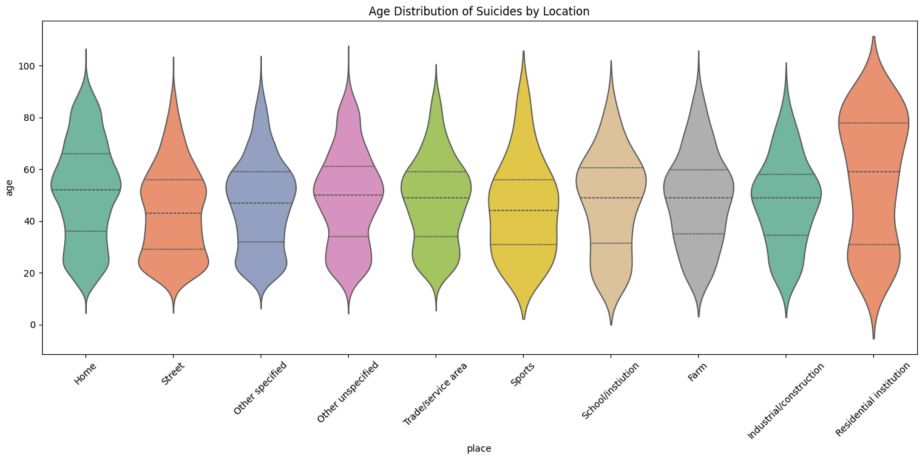


Fig. 60 Violin plot of age distribution of suicide by location

Interpretation:

- Suicides at **home** and **residential institutions** are skewed toward older age groups.
- **Street** and **sports** locations show relatively younger age distributions, likely reflecting youth demographics.

Summary and Common Interpretation:

The most dominant location for suicides is clearly the home environment, both in absolute numbers and proportions. This may be due to accessibility, privacy, and emotional triggers linked with domestic settings. Public places, although more visible, account for a small fraction. Temporal analysis shows this trend has been stable or slightly rising. Age-wise, the elderly are more likely to take their lives at home or institutions, while younger populations are more represented in external or undefined locations.

The place of suicide plays a significant role in identifying environmental risk factors. Home is the most common and consistently rising location, suggesting targeted mental health interventions in domestic and familial settings could have a high impact. Temporal and demographic breakdowns support this conclusion, highlighting that preventive policies must address in-home psychological safety and access control to means of suicide.

4.7 Q7: How does the location of suicides vary by demographic factors like age and sex?

This question investigates how suicide locations differ across gender and age groups. The objective is to explore whether men and women show different patterns in where suicides occur, and how age influences these spatial choices. The study combines traditional visualizations with creative diagrams to expose demographic-location relationships.

Code snippets:

```
print("Loading full suicide dataset...")
df = pd.read_csv(DATA_FILE)
df.columns = df.columns.str.strip().str.lower()
df = df[df["intent"] == "Suicide"]
df = df.dropna(subset=["place", "sex", "age"])
```

Fig. 61 loading the dataset in dataframe for Q7

Filters the dataset to include only valid suicide cases and prepares it for demographic analysis by creating age groups. This is essential to map suicide patterns across gender and age categories.

```
plt.figure(figsize=(12, 6))
sns.set_palette("Set3")
sns.countplot(data=df_top, x="place", hue="sex", order=top_places)
plt.title("Suicide Location by Gender (Top 6 Places)")
```

Fig. 61 Creating the plots

Creates a comparative bar plot showing suicide counts across the top 6 locations, split by gender. Useful for visualizing gender disparity at each location.

```
radial_data = df_top.groupby(["place", "sex"]).size().reset_index(name="count")
fig_radial = px.line_polar(
    radial_data, r="count", theta="place", color="sex", line_close=True,
    title="Radial Chart - Gender Spread Across Suicide Locations",
    color_discrete_sequence=px.colors.qualitative.Dark2
)
fig_radial.update_layout(margin=dict(t=50, l=30, r=30, b=30), height=600)
radial_path = os.path.join(OUTPUT_DIR, "radial_chart_place_gender_colored.html")
fig_radial.write_html(radial_path)
```

Fig. 62 Creating the radial charts for gender wise suicide counts across locations

Displays a circular visual distribution of gender-wise suicide counts across locations, highlighting the spread and dominance of one gender over another for each setting.

```
chord_data = df_top.groupby(["sex", "place"]).size().reset_index(name="count")
links = chord_data.rename(columns={"sex": "source", "place": "target", "count": "value"})
chord = hv.Chord(links)
chord.opts(
    opts.Chord(
        labels="index", cmap="Category20b", edge_color="source",
        edge_cmap="Category20c", edge_color_index="source",
        width=800, height=600, title="Chord Diagram - Gender to Place Flows"
    )
)
```

Fig. 63 Creating chord diagram

Generates a chord diagram linking gender to suicide locations with curved connectors. Great for intuitively understanding high-traffic flows between demographic groups and locations.

```
source = sankey_data["sex"].map(label_idx)
target = sankey_data["place"].map(label_idx)
intermediate = sankey_data["place"].map(label_idx)
target2 = sankey_data["age_group"].map(label_idx)

node_colors = px.colors.qualitative.Dark2 * 3
link_colors = px.colors.qualitative.Pastel1 * (len(source) + len(target2))

sankey = go.Figure(data=[go.Sankey(
    node=dict(
        pad=15, thickness=20, line=dict(color="black", width=0.5),
        label=labels, color=node_colors[:len(labels)]
    ),
```

Fig. 64 Creating Sankey diagram

Constructs a flow diagram showing how gender connects to place and then to age group. This three-level linkage makes it easy to understand complex demographic relationships in suicide patterns.

```
pivot_rank["rank"] = pivot_rank.groupby(["year", "sex"])["count"].rank("dense", ascending=False)
top_places_final = pivot_rank.groupby("place")["count"].sum().nlargest(5).index.tolist()
pivot_rank = pivot_rank[pivot_rank["place"].isin(top_places_final)]
```

Fig. 65 Showing the popularity rank for location changes

Shows how the popularity (rank) of each suicide location changes over time for males and females. Lower rank = more suicides. Useful for trend comparison.

Visualization

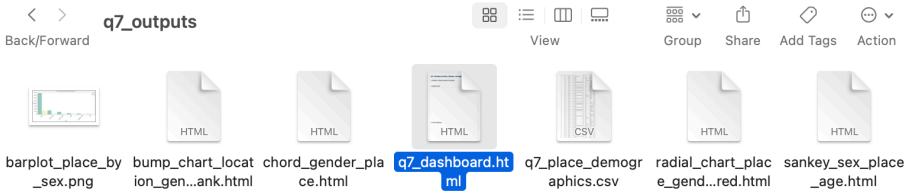


Fig. 66 This is how the output gets saved

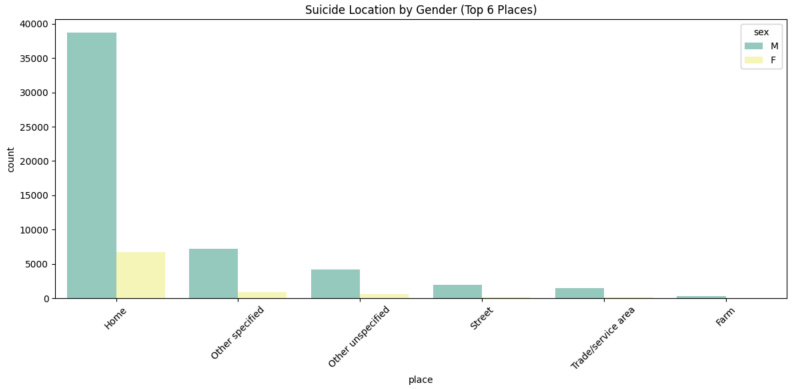


Fig. 67 Bar plot of suicide by location

Interpretation: Bar Plot: Suicide Location by Gender (Top 6 Places)

- *Home* is the most frequent location for both men and women, though the male count is far higher.
- Men show significantly higher suicide counts across all six top locations.

Fig. 68 Radial chart – explaining the gender spread across suicide location

Interpretation Radial Chart: Gender Spread Across Suicide Locations

- A circular view clearly highlights that males are dominant in suicides across all top locations.
- Females show a more constrained spread, especially in outdoor or occupational settings like farms and service areas.

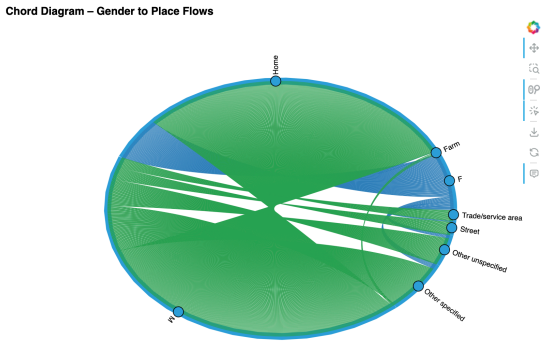


Fig. 69 Chord diagram

Interpretation Chord Diagram: Gender-to-Place Connections

- Strong visual arcs confirm that most male suicides flow toward home, other specified, and unspecified locations.
- Female arcs are thinner, reinforcing gender differences in frequency and distribution.

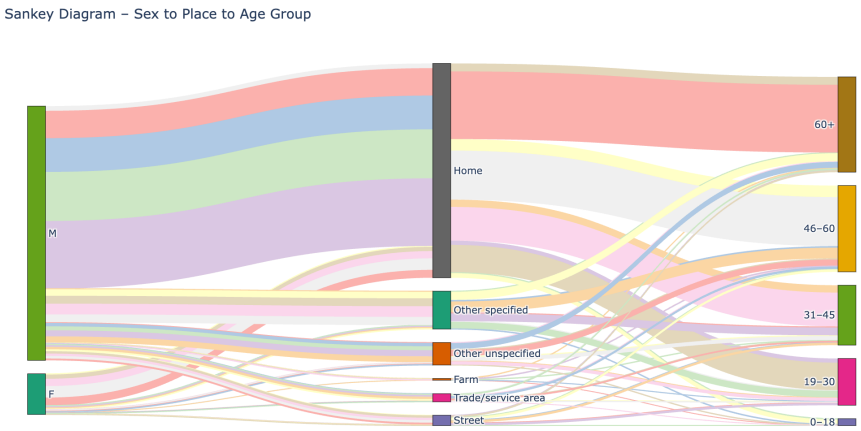


Fig. 70 Sankey diagram which explains the sex > place > age group

Interpretation Sankey Diagram: Sex → Place → Age Group

- Most male suicides occur at home and affect the 46–60 and 60+ age groups.
- Female flows are prominent toward age groups 31–45 and 46–60 at home, with fewer connections in other places.
- The visual flow helps connect *who* (sex), *where* (place), and *when in life* (age).

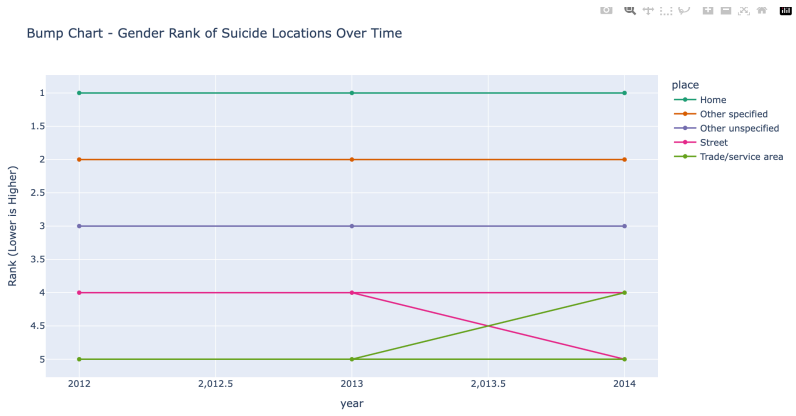


Fig. 71 Bump chart – showing gender rank of suicide locations over time

Interpretation Bump Chart: Rank of Suicide Locations by Gender Over Time

- For both genders, *home* remains the consistent top-ranked location from 2012 to 2014.
- Secondary ranks vary: "Other specified" and "Other unspecified" follow for both genders, but the fourth and fifth ranks shift slightly

Common Interpretation and Summary

- **Gender Gaps:** Across all visualizations, males consistently outnumber females in suicide incidents, especially in locations like streets, farms, and trade areas. Females are more concentrated in home-based suicides.
- **Age Diversity:** Older males (46–60, 60+) form a major demographic in home-based suicides, while younger females appear more in the 31–45 group.
- **Consistency in Location Rank:** The bump chart shows that place rankings are stable over time, with minimal yearly shift in relative preferences.
- **Chord & Sankey Flow Power:** These creative diagrams make it easy to spot high-flow categories and demographic bottlenecks, reinforcing the home-centric trend for all genders and ages.

Suicide locations are heavily influenced by gender and age. Males are more likely to commit suicide in almost all settings, especially at home and in open or work-related spaces. Females show fewer but more concentrated patterns, primarily at home and within middle-age groups. Prevention strategies should integrate gender-sensitive and age-targeted interventions—especially focusing on home environments for older males and middle-aged females.

4.8 Q8 What is the relationship between suicide trends and education levels?

This question explores how suicide incidents vary across different education levels over the years 2012–2014. It investigates whether individuals with certain educational backgrounds are more vulnerable, and how trends shift across gender and time.

Code snippets

```
# Aggregating
grouped = df.groupby(["year", "sex", "education"]).size().reset_index(name="suicides")
```

Fig. 72 Aggregating the dataset

Groups data by year, gender, and education level, creating a new suicides column for aggregation.

```
sns.barplot(data=grouped, x="year", y="suicides", hue="education", ci=None)
```

Fig. 73 This is one of the generating the bar plot

Creates a grouped bar plot showing how suicides differ by education each year.

```
sns.lineplot(data=grouped, x="year", y="suicides", hue="education", style="sex", markers=True, palette=custom_palette)
```

Fig. 74 This is one of the generating the line plot

Produces a line plot for trend comparison across education levels and genders.

```
fig_treemap = px.treemap(
    grouped,
    path=["year", "education", "sex"],
    values="suicides",
    title="Suicide Distribution by Year → Education → Gender",
    color="suicides",
    color_continuous_scale="Tealrose"  # changed from Reds to Tealrose
)
fig_treemap.write_html(os.path.join(OUTPUT_DIR, "treemap_education_gender.html"))
```

Fig. 75 Creating the treemap

Constructs a hierarchical treemap to understand volume and subgroup structure.

```
sns.heatmap(pivot, cmap="coolwarm", annot=True, fmt=".0f")
```

Fig. 76 Creating the heatmap

Generates a heatmap showing how suicide counts per education level evolve over time.

Visualizations

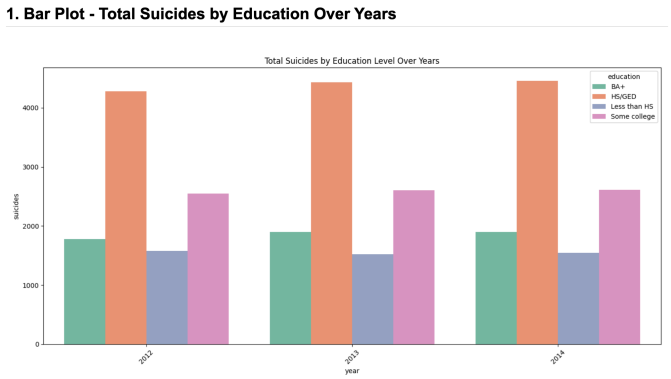


Fig. 77 Bar plot of total suicides by education over year

Interpretation:

- HS/GED consistently has the highest suicide counts each year.
- BA+ and Some College remain relatively stable and lower than HS/GED.
- Those with **Less than High School** education have the lowest reported suicide numbers.

Individuals with mid-level education (HS/GED) appear more vulnerable, possibly due to socio-economic pressure—higher than less-educated but lacking the stability often associated with advanced degrees.

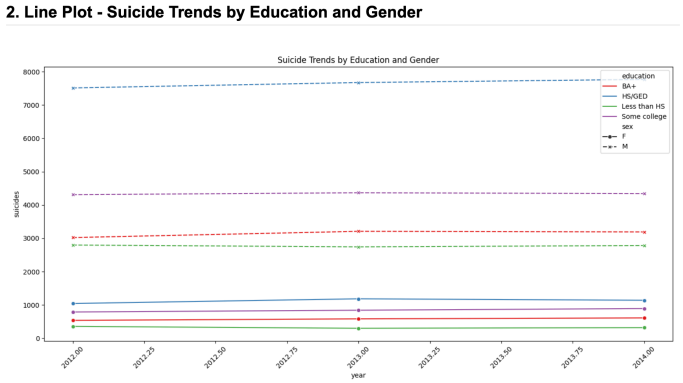


Fig. 78 Line plot explaining the suicide trends by education and gender

Interpretation:

- Male suicides dominate all education groups, particularly in HS/GED and BA+.
- Female trends remain lower but are slightly more evenly spread across education levels.
- Trends show a **slight increase** across all groups, most notably for **HS/GED males**.

Gender disparities persist, with males showing significantly higher suicide counts across all education levels, indicating the need for male-targeted prevention in middle education tiers.

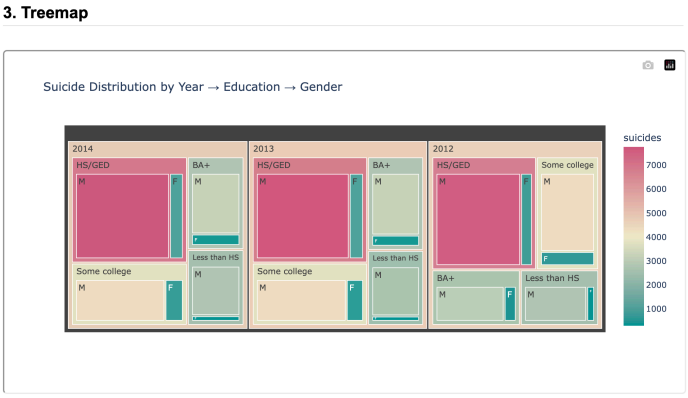


Fig. 79 Treemap explaining the suicide distribution by year > education > gender

Interpretation:

- A large portion of the treemap is covered by **HS/GED males**, confirming earlier visuals.
- BA+ and Some College have more balanced distributions but still skewed towards males.
- Female segments are visibly smaller across all groups.

The treemap clearly visualizes that most suicide cases stem from males with HS/GED education, followed by BA+. This hints at a societal or economic struggle post-secondary school.

4. Bubble Chart

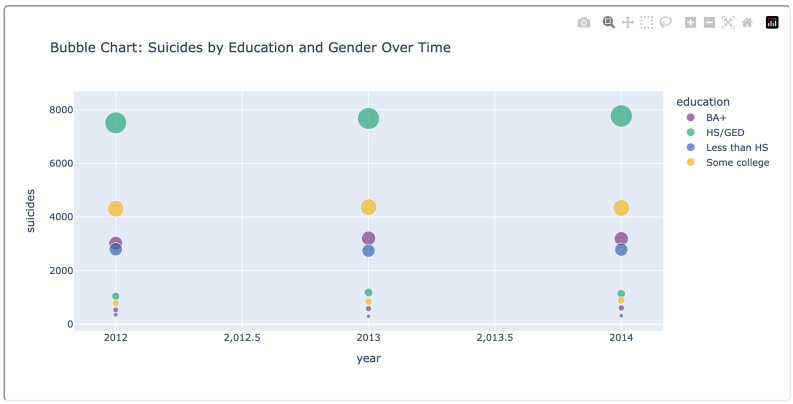


Fig. 80 Bubble chart explaining the suicide by education and gender over time

Interpretation:

- Larger bubbles again correspond to **HS/GED** males.
- Bubble sizes increase slightly over time, showing a **rising pattern**.
- Some college and BA+ groups have smaller, steady bubbles, showing less variation.

This chart reinforces earlier conclusions, with larger bubbles visualizing higher suicide burdens among HS/GED males. Education seems to act as a **protective factor** beyond a certain level.

5. Heatmap

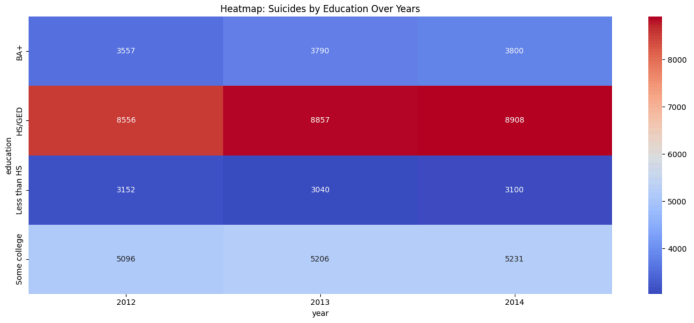


Fig. 81 Heatmap explaining the suicide over years

Interpretation:

- HS/GED stands out with consistently high numbers: 8556 (2012), 8857 (2013), and 8908 (2014).
- Some College maintains second-highest numbers.
- BA+ and Less than HS remain consistently lower.

The heatmap confirms temporal stability in trends, suggesting persistent socio-economic factors linked to education levels. Intervention efforts could target HS/GED groups for greatest impact.

Common interpretation and summary:

The analysis indicates a **clear relationship** between suicide trends and education levels:

- **HS/GED** individuals are **most at risk**, especially **males**, possibly due to job insecurity or mid-career instability.
- Higher education (BA+) appears **slightly protective**, while less than high school education might correlate with **underreporting** or different societal factors.
- Gender disparities are pronounced across all groups.

Educational support programs, especially for HS/GED populations, combined with targeted mental health interventions for males, can help reduce suicide rates.

4.9 Q9 How do suicide trends vary across different racial groups and genders?

This question investigates the racial dimensions of suicide trends, focusing on differences across gender and age groups for each race. The analysis uses two key visualizations: a traditional bar plot and a multi-level Sankey diagram that visualizes suicide distribution from **Race → Gender → Age Group**.

Code snippets:

```
# Add age group buckets
bins = [0, 18, 30, 45, 60, 100]
labels = ["0-18", "19-30", "31-45", "46-60", "60+"]
df["age_group"] = pd.cut(df["age"], bins=bins, labels=labels, right=False)
```

Fig. 82 Creating age bins or buckets

The dataset is filtered for rows where suicide intent is confirmed, and race, gender, and age are available. A new age_group column bins age into categories: **0–18, 19–30, 31–45, 46–60, 60+**.

```
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x="race", palette="Set2")
plt.title("Total Suicides by Race")
```

Fig. 83 Using plotly to create graphs

```
sankey_fig = go.Figure(data=[go.Sankey(
    node=dict(
```

Fig. 84 Creating sankey

Visualization:

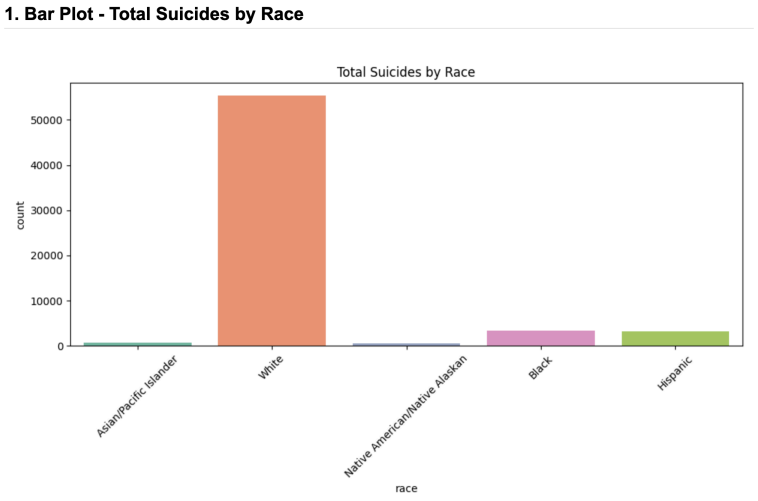


Fig. 85 Bar plot of total suicides by races

Interpretation:

- White individuals account for the majority of suicides, with a total far exceeding any other racial group.
- Black and Hispanic populations come next but show significantly lower totals.

- Asian/Pacific Islanders and Native American/Native Alaskan groups have minimal representation in the data.

White individuals face a disproportionate share of suicides in absolute terms. While this doesn't reflect population-adjusted rates, the raw counts point to a potentially higher burden within this group and raise questions about age and gender dynamics within the White population.

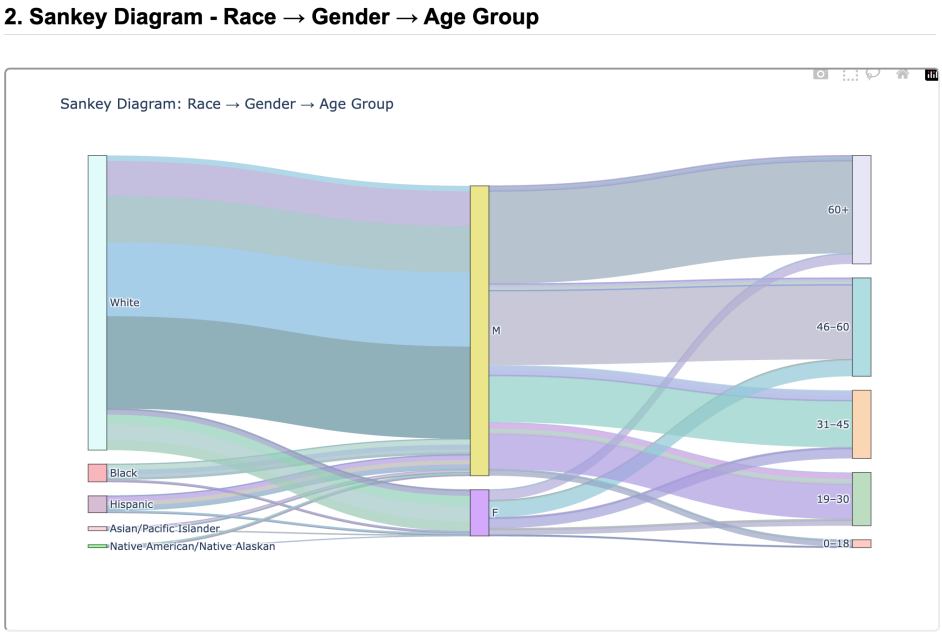


Fig. 86 Sankey diagram of Race > Gender > Age group

Interpretation:

- The **largest flows** emerge from **White → Male → 60+**, indicating a severe concentration of suicides among older White men.
- Among **non-White races**, males still dominate the flow volume, particularly in **Black** and **Hispanic** groups within the **19–45 age range**.
- Female paths are consistently thinner across all races and age brackets.

Males are significantly more affected in every racial group, reaffirming the broader gender-based disparity. The age group **60+** carries the highest weight, especially for **White males**, pointing to possible issues of aging, isolation, or retirement stress. **Young and mid-age males** (19–45) in **Black and Hispanic** communities also show notable representation, suggesting the need for culturally aware prevention programs.

Common summary and interpretation:

- **White males over 60** emerge as the highest-risk group.
- Gender disparities remain wide, with males at much greater risk than females across all racial categories.
- Minority groups—while lower in count—should not be overlooked, particularly **young minority males** who show visible flows.
- The Sankey diagram offers **intersectional clarity**, helping identify where multiple risk factors converge.

4.10 Q4 How Do Suicide Rates Vary Across Different U.S. States or Regions?

Dataset used: state_level_suicide_in_years folder

```
│   ├── state_level_suicide_in_years/
│   │   ├── suicide_state_2005.csv
│   │   ├── suicide_state_2014.csv
│   │   ├── suicide_state_2015.csv
│   │   ├── suicide_state_2016.csv
│   │   ├── suicide_state_2017.csv
│   │   ├── suicide_state_2018.csv
│   │   ├── suicide_state_2019.csv
│   │   ├── suicide_state_2020.csv
│   │   └── suicide_state_2022.csv
```

This section explores the geographic and temporal variations in suicide rates across the United States from 2005 to 2022. The primary aim is to identify high-risk states, uncover regional disparities, and examine how suicide trends evolved over time using choropleth maps, bar plots, and chord diagrams.

Data was sourced from the CDC's official suicide mortality statistics, spanning state-level suicide data from 2005 to 2022. The cleaned and merged dataset included over 4,500 entries, each containing state, year, rate, and deaths fields.

Code snippets:

```
print("\nLoading state-level suicide data...")
all_data = []

for file in os.listdir(DATA_DIR):
    if file.endswith(".csv"):
        path = os.path.join(DATA_DIR, file)
        df = pd.read_csv(path)
        df.columns = df.columns.str.strip().str.lower()

        if set(["year", "state", "rate", "deaths"]).issubset(df.columns):
            df["deaths"] = df["deaths"].astype(str).str.replace(",", "").astype(float)
            df["rate"] = pd.to_numeric(df["rate"], errors="coerce")
            all_data.append(df)
```

Fig. 87 Loading the dataset

```
plt.figure(figsize=(12, 6))
sns.barplot(data=total_by_year, x="year", y="deaths", palette="magma")
plt.title("Total Suicides per Year (2005–2022)")
```

Fig. 88 Creating plotly and graphs

This logic loads all CDC CSVs year-wise, cleans death and rate values, and generates a grouped bar plot of total deaths per year.

Visualization:

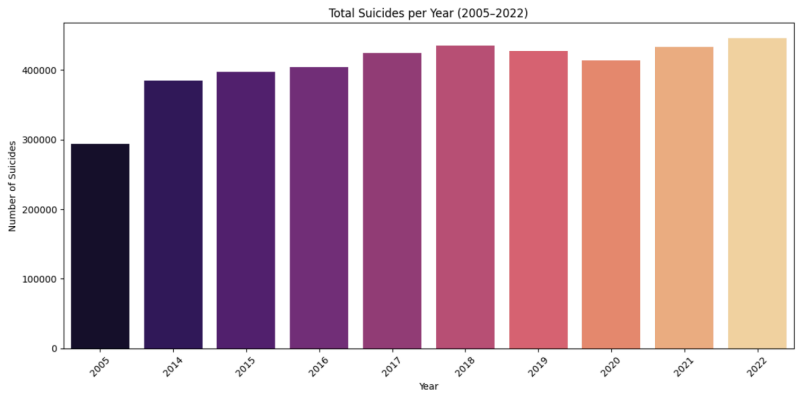


Fig. 89 Bar graph of total suicides per year (2005 – 2022)

Interpretation:

- The bar plot summarizing annual suicide counts from 2005 to 2022 presents a gradual but significant rise in suicide cases over time.
- The count rose from approximately **293,436** in 2005 to over **445,284** in 2022.
- This ~52% increase across nearly two decades signals a deepening public health crisis.
- Notably, the years **2017 to 2019** marked the sharpest spikes, followed by a slight dip during the **2020 pandemic**, likely due to reporting delays or shifts in healthcare priorities.
- However, post-pandemic figures rebounded sharply, indicating that underlying issues may have been exacerbated rather than resolved.

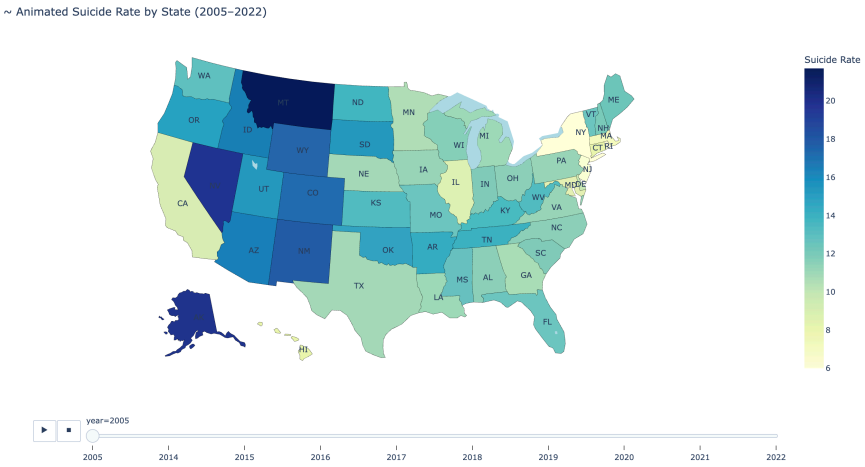


Fig. 90 Animated choropleth for the year 2005

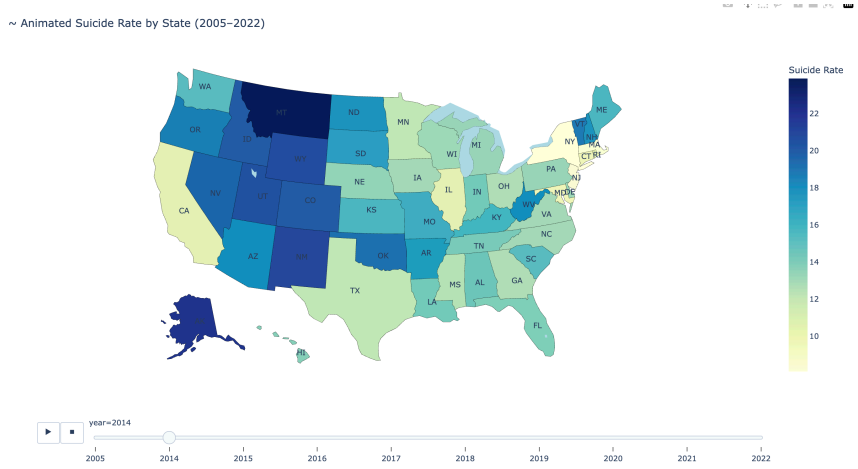


Fig. 91 Animated choropleth for the year 2014

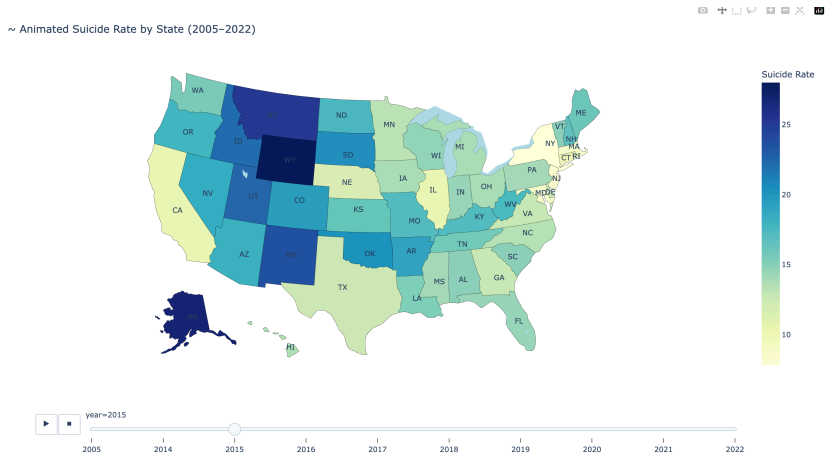


Fig. 92 Animated choropleth for the year 2015

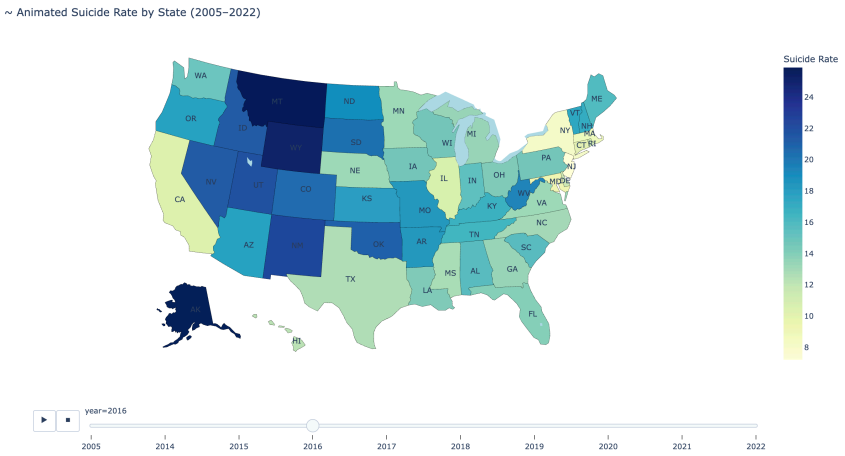


Fig. 93 Animated choropleth for the year 2016

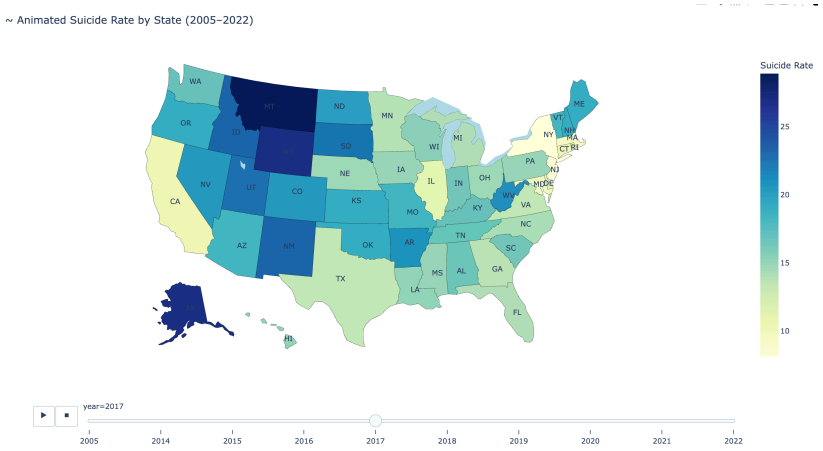


Fig. 94 Animated choropleth for the year 2017

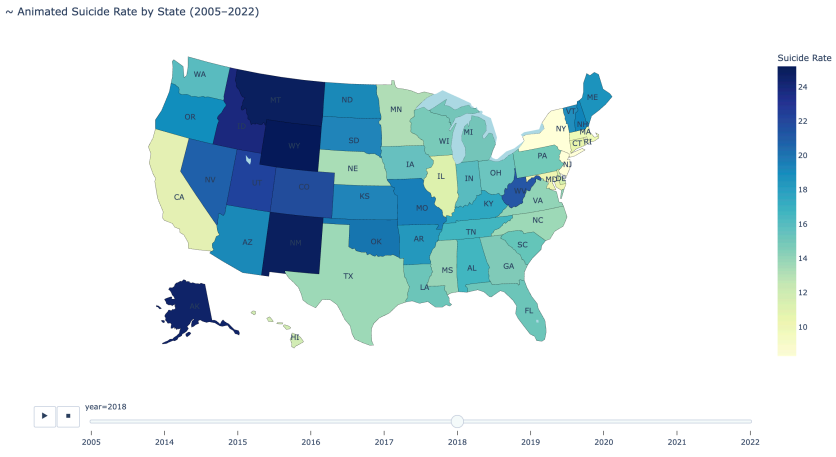


Fig. 95 Animated choropleth for the year 2018

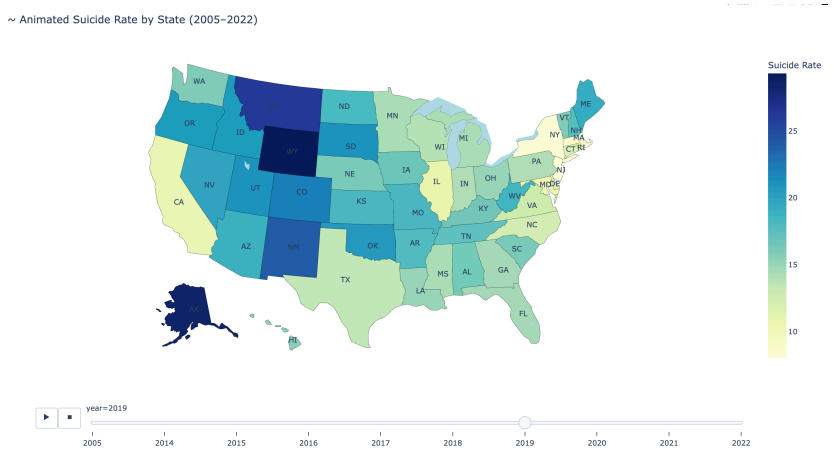


Fig. 96 Animated choropleth for the year 2019

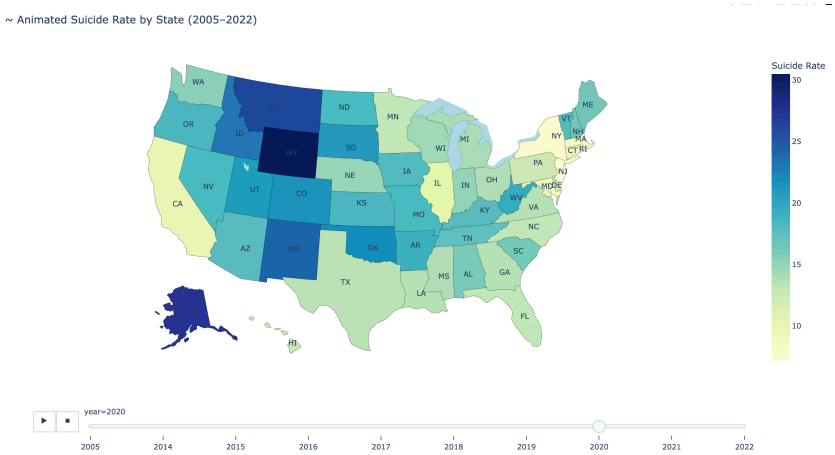


Fig. 97 Animated choropleth for the year 2020

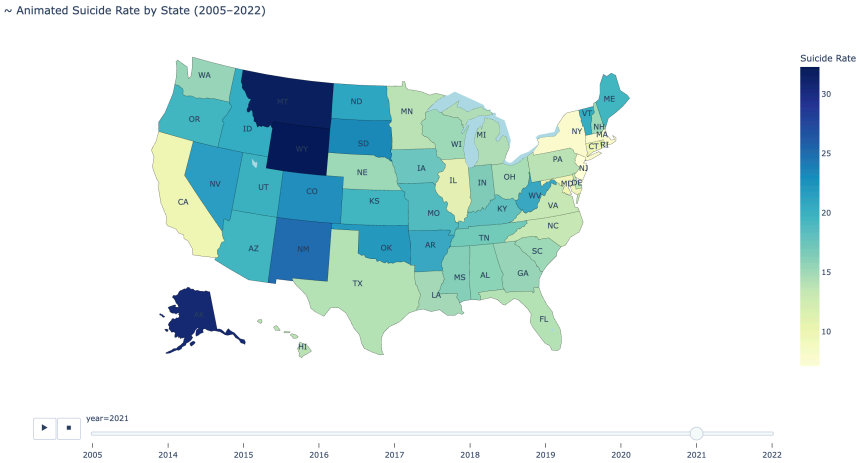


Fig. 98 Animated choropleth for the year 2021

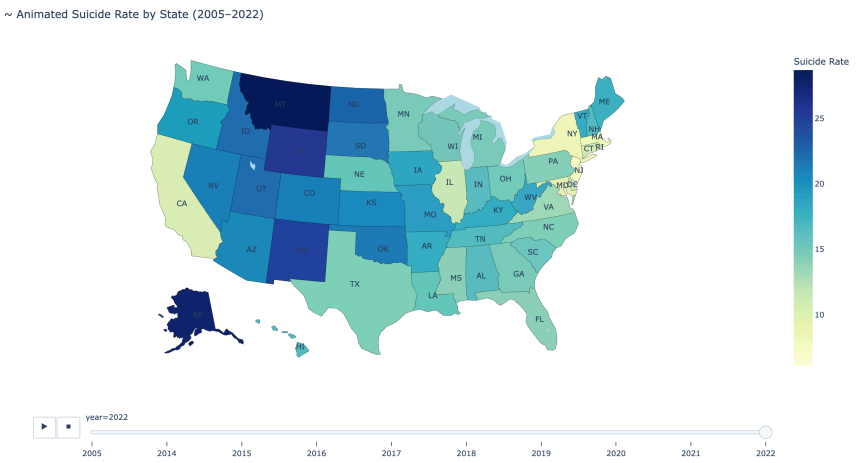


Fig. 99 Animated choropleth for the year 2021

Interpretation:

The choropleth maps offer dynamic, year-wise visualization of suicide rates by U.S. states:

- **2005–2010**:
  - States in the **Western U.S.**, notably **Montana (MT)**, **Wyoming (WY)**, **Alaska (AK)**, and **New Mexico (NM)** already displayed high rates (often above 20 per 100,000).
  - These states remained persistently dark throughout the animation, indicating structural, geographic, or healthcare-access disparities.
- **2011–2016**:
  - **Southern and Midwestern states** like **Oklahoma**, **Utah**, and **South Dakota** began showing elevated rates.
  - Interestingly, many **Northeastern states (NY, NJ, MA)** consistently reported lower rates, possibly reflecting better access to mental health care or urbanization effects.
- **2017–2022**:
  - The darkest intensities were again seen in **MT, WY, AK**, and **ID**, suggesting they consistently experienced some of the **highest suicide burdens** in the nation.
  - Notably, **California** and **New York** remained among the lowest, despite their population sizes, underlining that suicide trends are not purely population-driven but context-dependent.

This temporal sweep highlights the entrenched regional disparities in mental health outcomes and the need for geographically targeted interventions.

Average Suicide Rate by State (Tabular Visualization) (attached in appendix)

The tabulated average suicide rates reaffirm the visual choropleth trends. The top five states with highest average rates (2005–2022) are:

- Montana (26.36)
- Wyoming (26.09)
- Alaska (26.07)
- New Mexico (23.12)
- Idaho (21.34)

In stark contrast, the lowest rates were observed in **New Jersey (7.64)**, **New York (7.91)**, and **Massachusetts (8.59)**. These differences: more than threefold between extremes underscore how structural inequities, rural isolation, or cultural stigma may be influencing help-seeking behavior and suicide incidence.

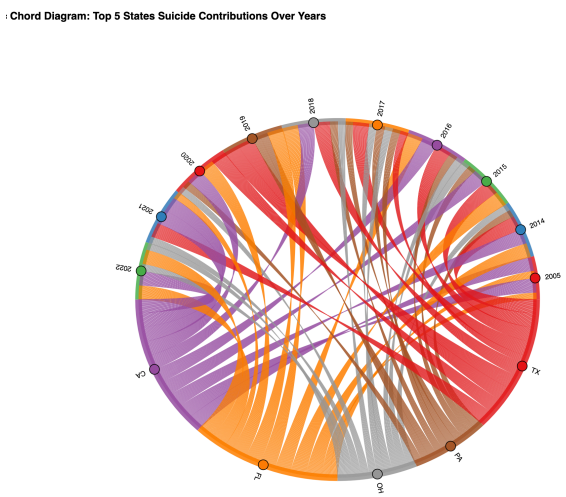


Fig. 100 Chord diagram explaining the top 5 states of suicide contribution in the years (2005-2022)

Interpretation:

The chord diagram offers a striking visual of the year-wise suicide contribution by the five states with the highest cumulative suicide counts: **California (CA)**, **Texas (TX)**, **Florida (FL)**, **Ohio (OH)**, and **Pennsylvania (PA)**.

- **California and Texas** consistently contribute the most across all years, an expected result given their population size. However, their repeated dominance in both raw numbers and chord thickness reflects systemic, not incidental, issues.
- **Ohio, Florida, and Pennsylvania** show consistent but smaller bands, indicating mid-level contributions with less volatility.

- The width of the arcs across years reflects growing suicide numbers overall, with the diagram's density increasing especially between **2014 and 2022**, echoing the bar plot's trends.

This diagram provides an intuitive way to observe both time-based progression and state-based dominance, useful for federal-level public health planning.

The analysis of state-level suicide data reveals a stark geographic divide in both suicide rate and absolute deaths. While populous states like California and Texas contribute the most in count, small-population states like Alaska and Montana face disproportionately high suicide rates. Regional mental health disparities and access to care remain central factors. These findings highlight the need for region-specific interventions, especially in the Western U.S., and signal the importance of continued post-pandemic support services across all states.

4.11 Q10 Global Suicide Trends (1985–2016)

This question explores global suicide patterns from 1985 to 2016 using Kaggle's public dataset master.csv. The aim is to identify temporal, demographic, economic, and geographic factors associated with suicide rates. Specifically, the analysis investigates:

- Long-term global trends
- Gender-based suicide differences
- Age and generational vulnerabilities
- Economic associations (GDP)
- Country-wise and decade-wise evolutions

The dataset consist of:

- Source: Kaggle Suicide Rates Overview 1985 to 2016
- Rows: 27,820
- Key Fields:
    - country, year, sex, age, generation
    - suicides_no, population, gdp_per_capita ($)
- Cleaning Steps:
    - Removed commas in GDP columns and converted to float
    - Filled missing HDI for year values with median

Visualization

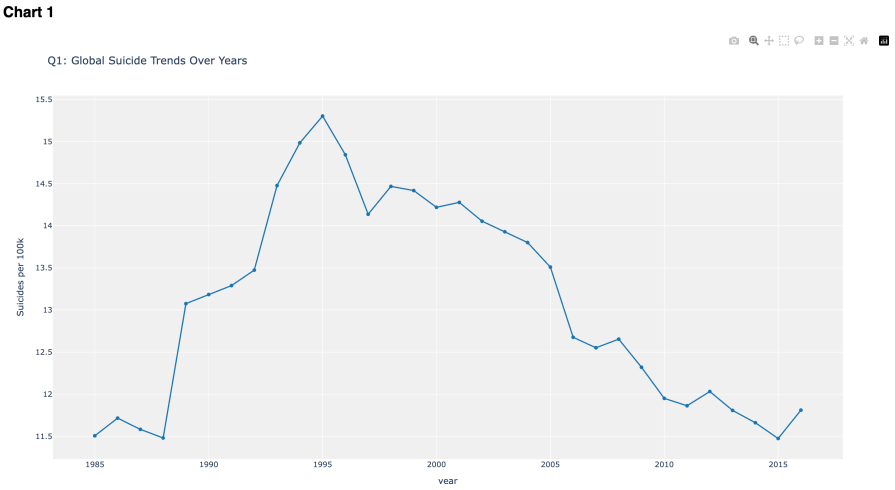


Fig. 101 Line plot showing the global suicide trends

Interpretation:

- Suicide rates peaked globally in the **mid-1990s** (~15.3 per 100k in 1995) and steadily declined to **~11.5 per 100k by 2016**.
- This implies Global awareness, mental health access, and interventions may have contributed to this decline post-2000.

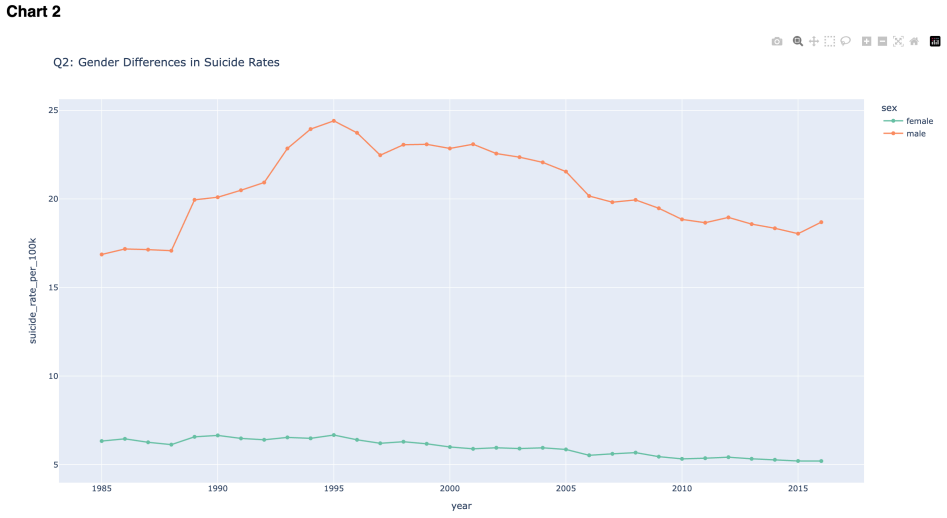


Fig. 102 Line plot of Gender Differences

Interpretation:

- Males consistently show **2.5–3x higher suicide rates** than females.
- Peaks**:** Males peaked around 1995 (~24.5 per 100k); females stayed below 7 throughout.

- This implies suicide prevention strategies must be **gender-sensitive**, focusing particularly on male populations.

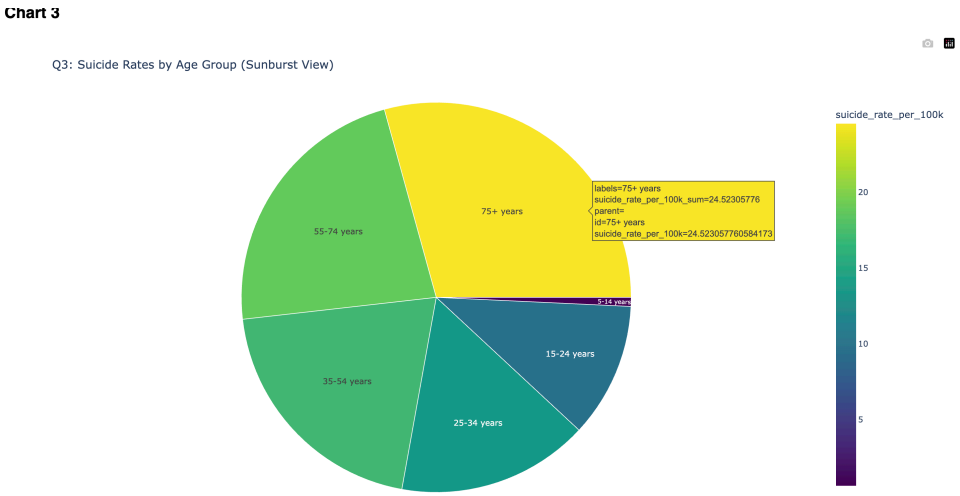


Fig. 103 Sunburst chart of suicide by age groups

Interpretation:

- **Highest Risk**:
  - Individuals **75+ years** showed the **highest rate (~24.5 per 100k)**, followed by 55–74 years.
- **Lowest Risk**:
  - Children aged 5–14 had extremely low rates (~0.62).
- This implies suicide vulnerability **increases with age**, underscoring the need for elder mental health support.

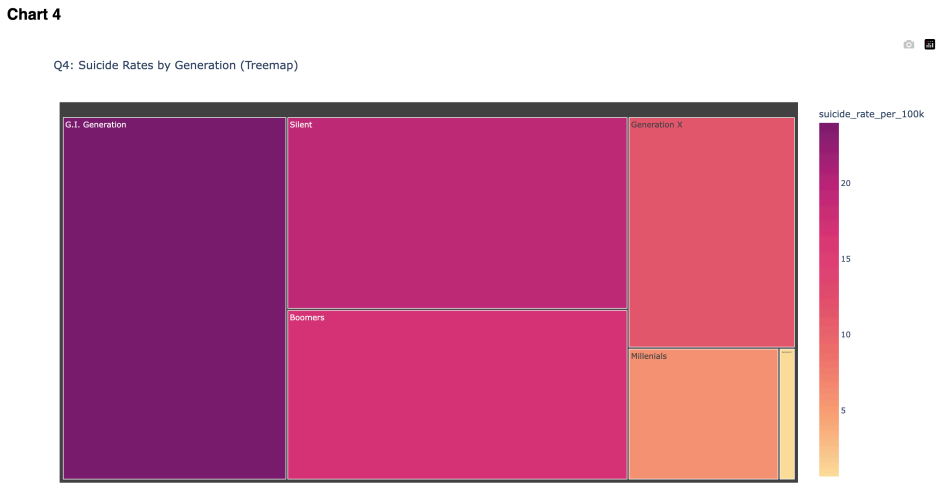


Fig. 104 Treemap chart of suicide rates by generation

Interpretation:

- **Highest Risk Generation**:
  - The **G.I. Generation** (~24 per 100k), followed by the **Silent Generation**.
- **Lowest**:
  - **Generation Z** and **Millennials** showed the lowest suicide rates.
- This implies the risk correlates strongly with **older cohorts**, hinting at sociocultural or medical challenges with aging.

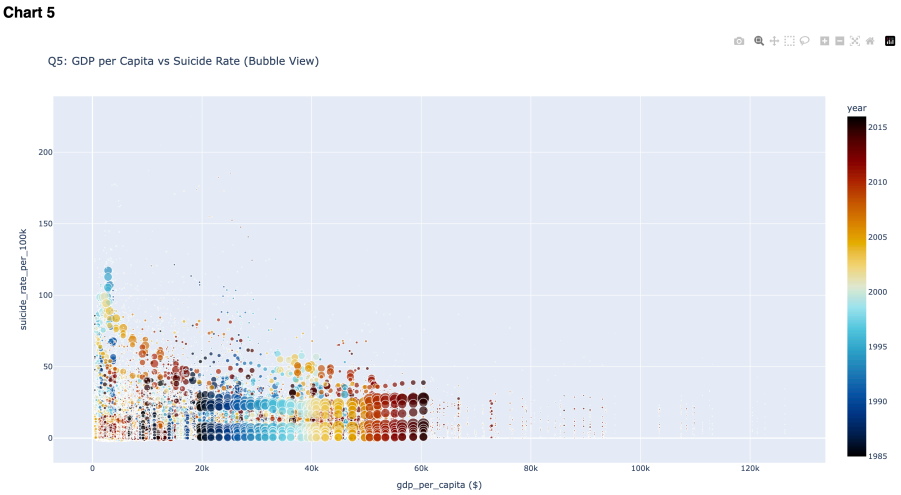


Fig. 105 Bubble plot of GDP vs. Suicide rate

Interpretation:

- **Trend**:
  - Suicide rates don't always decrease with economic prosperity.
  - Some **high GDP countries** still report high suicide rates.
- **Clusters**:
  - Heaviest density of suicides is in **low-to-mid GDP brackets**.
- This implies **GDP is not a strong standalone predictor** of suicide risk. Mental health policies matter more than wealth alone.

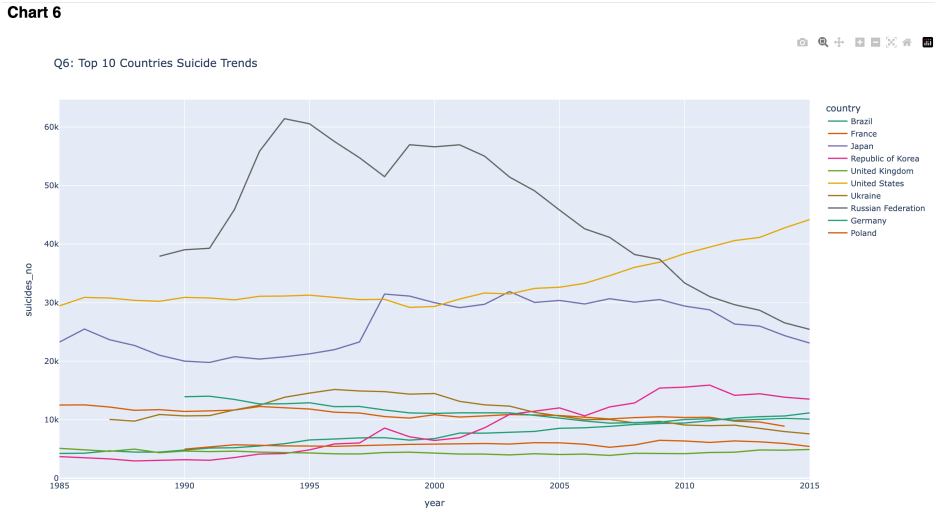


Fig. 106 Line plot of Top 10 countries that have major issues of suicide

Interpretation:

- **Leaders**:
  - United States, Russia, and Japan reported the highest absolute suicides.
- **Notable Shifts**:
  - Russia peaked in the 1990s and declined steadily.
  - U.S. showed **a rising trend**, peaking post-2010.
- This implies suicide prevention progress varies drastically by country, driven by **culture, healthcare, and stigma levels**.

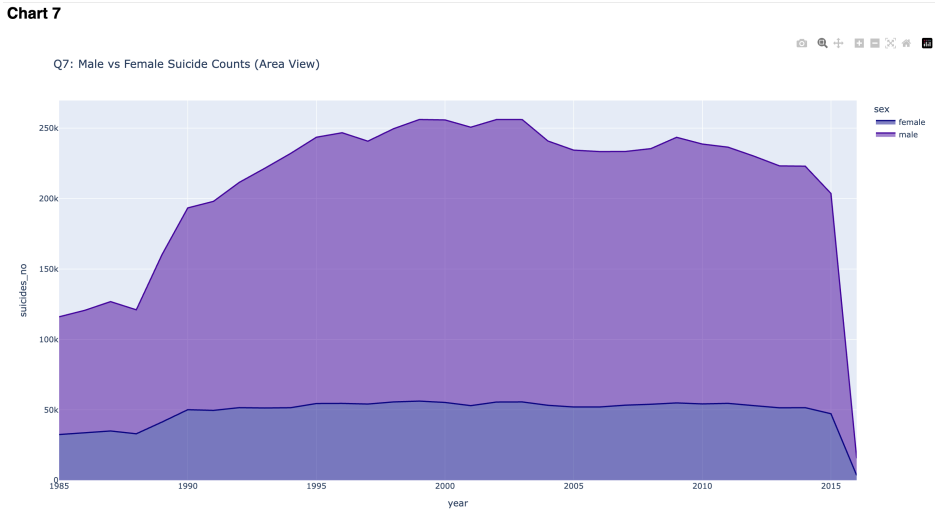


Fig. 107 Area plot of male vs. female suicide count in this dataset

Interpretation:

- **Observation**:
  - Males accounted for the **majority (70–75%) of suicides**.
- **Trend**:
  - The male count remained high across all years; female suicide counts were comparatively stable and lower.
- This means programs must focus on **masculinity norms, emotional expression, and early detection** for men.

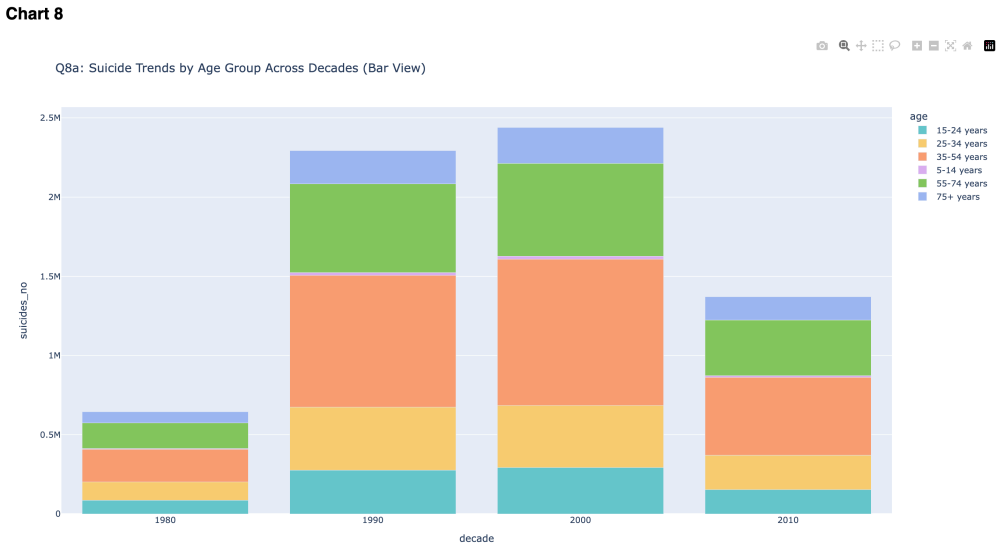


Fig. 108 Bar plot of suicide by age group over decades

Interpretation:

- **Peak**:
  - The 2000s had the **highest suicide counts**, especially among 35–54 and 55–74 age groups.
- **Trend**:
  - All age groups except children (5–14) saw significant suicide counts.
- This tells us that focused attention is needed for **middle-aged adults**, particularly during economically or politically unstable decades.

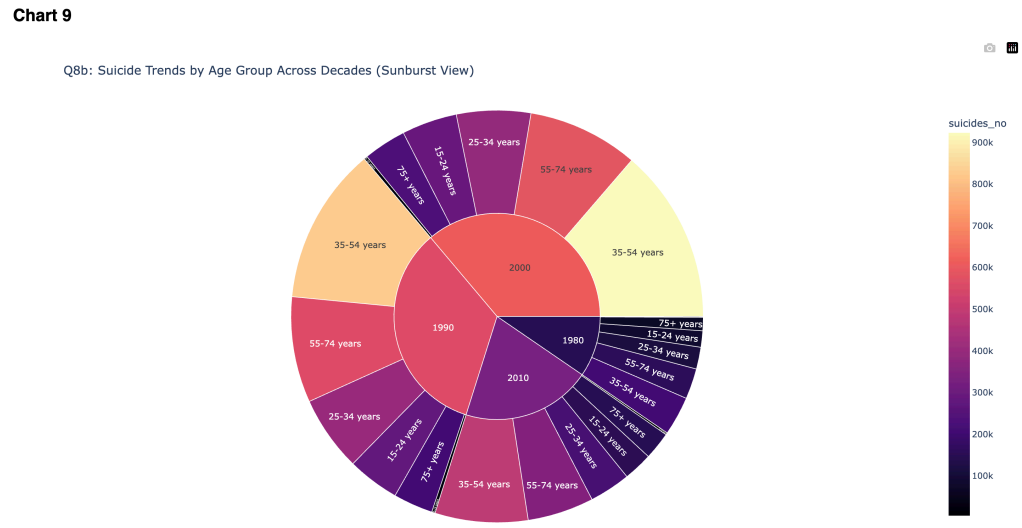


Fig. 109 Sunburst plot of age group vs. decade trend

Interpretation:

- The inner circle (decades) expands outward into age groups.
- **Observation**:
  - 1990s and 2000s contributed the highest volume, with **35–54 years** dominating across decades.
- This means these middle decades represent a **pivotal window** for policy and social support interventions.

This comprehensive global analysis reveals multi-dimensional insights:

- Suicide **rates have declined** globally since the 1990s, though disparities remain.
- Males, elderly individuals, and some specific generations (e.g., Silent, G.I.) face **disproportionately higher risks**.
- Economic indicators like GDP do **not linearly predict** suicide rates.
- Country-specific patterns highlight the **importance of contextual mental health policies**.
- Overall, the findings suggest a **need for age, gender, and region-specific prevention frameworks**, backed by public health funding and awareness.

4.12 Q11 Suicide Death Rate Trends by Age, Sex, and Race in the United States

This section aims to dissect the suicide death rate in the United States by examining intersections between age, gender, and race over time. The goal is to uncover hidden at-risk groups and

temporal shifts in vulnerabilities using a combination of advanced and interpretable visualizations.

About dataset:

- **Source**: CDC - Death Rates for Suicide by Sex, Race, Hispanic Origin, and Age – United States (https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1)
- **File**: Death_rates_for_suicide__by_sex__race__Hispanic_origin__and_age__United_States.csv
- **Time Span**: 1950–2020 (varies by group)
- **Dimensions**: Age Group (15 categories), Sex (Male/Female), Race, and Death Rate (per 100,000 population)

Visualizations

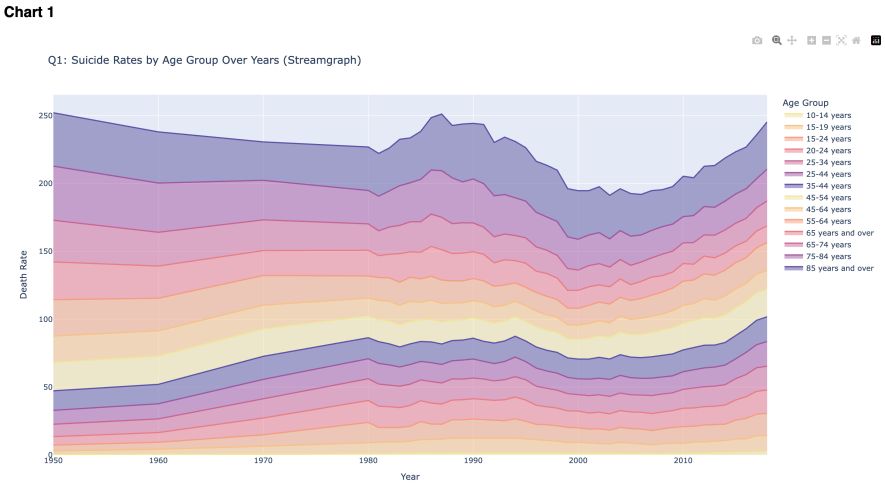


Fig. 110 Stream graph of age wise trend over time

Interpretation:

- Death rates peaked between 1985 and 2000 for older age groups (55+).
- Over time, youth groups (15–24 years) have seen a notable upward shift post-2000.
- Observable Trends**:**
  - 85+ and 75–84 remained the top contributors until the late 1990s.
  - Post-2010, a disturbing rise is visible among young adults (15–34 years).

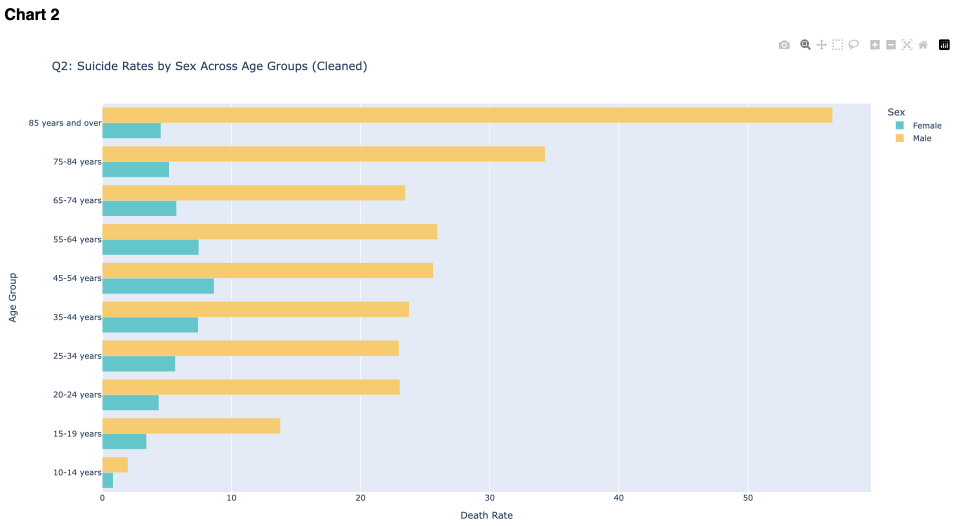


Fig. 111 Group bar plot by sex and age

Interpretation:

- Males consistently exhibit 3–4x higher suicide rates than females across all age brackets.
- Critical Group:
  - Males aged **85+ years** had the highest average death rate (>50 per 100k), whereas young females (under 24) had the lowest.

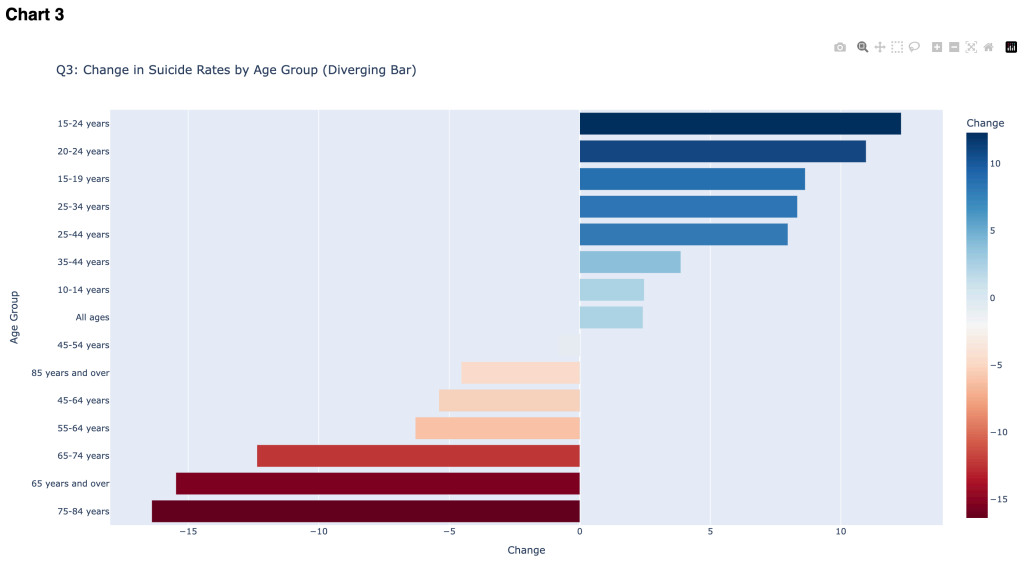


Fig. 112 Diverging bar chart showing change from 1950 to the recent years

Interpretation:

- **Biggest Increase**:

- Youth groups (15–24 years) have shown the steepest rise, with the 15–24 group increasing by **+12.3 per 100k**.

- **Biggest Drop**:
    - Seniors (75–84 years and 65+) saw the largest decline (up to **–16.4**), likely due to better care or interventions over time.

Chart 4

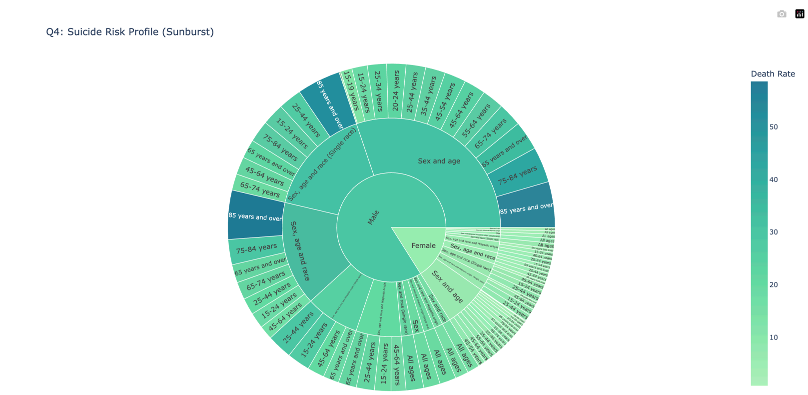


Fig. 113 Sunburst chart of suicide risk profile

Interpretation:

- Males form the dominant risk core. The deepest risk segments are:
    - **Male > White > 85+**
    - **Male > White > 75–84**
- This multi-layered breakdown highlights that **elder white males** are the highest-risk demographic when all three factors intersect.

Chart 5

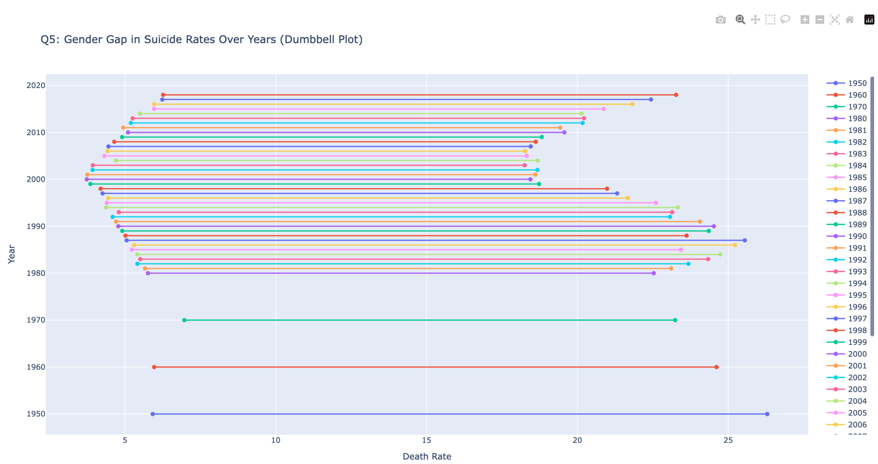


Fig. 114 Dumbbell plot explaining the gender gap over the time

Interpretation:

- Every year shows a consistent gap between male and female death rates.
- From 1950–2020, this gap widened particularly in the **1980s–1990s**.
- While both genders saw slight declines in recent years, the disparity remains unaddressed.

Fig. 115 Animated bubble plot which shows us age vulnerability by year (Animation seen on the html)

Interpretation:

- As the animation progresses, the **bubble size and position for older males (75–85+) remains dominant**.
- Youth bubbles have started inflating post-2000, visually showing rising vulnerability in younger populations.

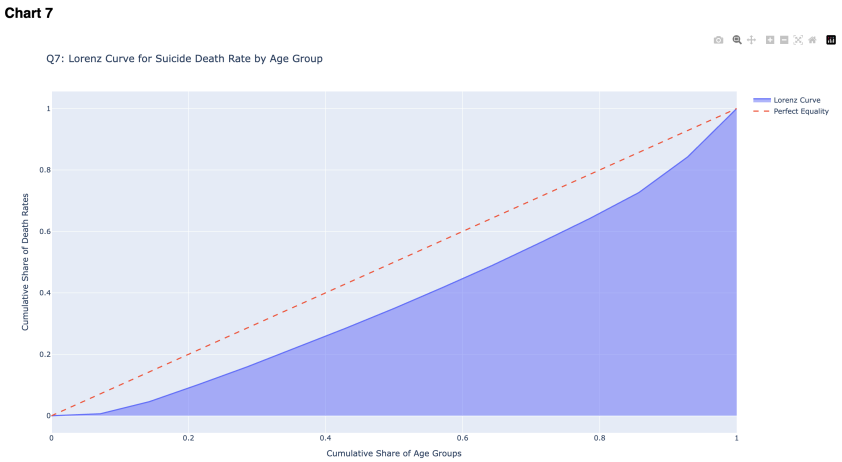


Fig. 116 Lorenz Curve of death rate distribution

Interpretation:

- The curve bows far below the line of equality, confirming **high inequality** in suicide distribution across age groups.
- A small subset of older age groups accounts for a disproportionately high share of deaths.

Chart 8: Radar Chart (Post-2010 Vulnerability)

Fig. 117 Radar chart explaining the possibility of vulnerability (post 2010)

Interpretation:

- **85+**, **75 - 84**, and **65 - 74** still lead in post-2010 death rates.
- However, **25 - 44 years** are now entering the outer circle, showing increasing risk.

The U.S. suicide death rate data reveals both long-standing and emerging threats. Historically, older white males (especially 85+ years) had the highest death rates, but recent decades indicate a **significant rise among youth and young adults**, especially those aged **15 - 24 years**. Gender disparities persist, with men consistently at higher risk. The Lorenz curve and animated plots emphasize the concentration of mortality in a few age brackets.

These findings underscore the importance of targeted mental health programs**:**

- For older adults facing isolation, loss, or illness.
- For youth and young adults, especially post-2010, possibly due to socio-digital stressors.
- Preventive education and gender-sensitive strategies must be deployed, especially in historically underserved racial and age demographics.

V. DISCUSSION

This study presents a comprehensive multi-dimensional analysis of suicide trends in the United States using demographic, geographic, temporal, and socioeconomic indicators. The insights drawn from each question reflect a mosaic of interdependent factors influencing suicidal behavior and highlight key areas for targeted intervention. Below, we discuss the findings per question and extrapolate practical implications for policymakers, educators, and mental health practitioners.

Q1 – Age and Suicide Trends

The data revealed that suicide incidents increase steadily with age, peaking in the 50–70+ age brackets. While younger groups are often the focus of intervention efforts, the consistent rise in suicide counts among middle-aged and elderly individuals suggests an urgent need to expand mental health resources for aging populations. Programs addressing loneliness, chronic illness, and financial insecurity should be prioritized for these age groups.

Q2 – Gender Disparities

A stark gender imbalance emerged, with males exhibiting significantly higher suicide rates than females across all age groups. This disparity persisted even after factoring in other variables. The results reinforce the need for male-focused prevention efforts that address stigma around help-seeking, masculinity norms, and unacknowledged mental health burdens among men.

Q3 – Interplay of Sex, Race, Education, and Age

The combined analysis underscored that individuals at the intersection of low education, minority racial identity, and older age are more vulnerable. The heatmaps and sunburst charts confirmed that these interlinked factors create compounding risks. Suicide prevention frameworks should adopt an intersectional approach that acknowledges how multiple demographic traits collectively heighten vulnerability.

Q4 – State-wise and Regional Variations

Significant geographical disparities in suicide rates were observed across U.S. states and regions. High-risk states such as **Wyoming, New Mexico, and Alaska** consistently recorded elevated suicide rates per 100,000 residents, while populous states like **California and Texas** had high absolute numbers due to population size. Animated choropleths and chord diagrams revealed

temporal consistency in these trends. These findings suggest that suicide is not uniformly distributed and that local social, economic, and healthcare factors play a role. Mental health funding and suicide prevention strategies should therefore be **tailored regionally**, with additional resources directed toward persistently high-burden states.

Q5 – *Temporal Patterns: Month and Season*

The analysis of monthly and seasonal data revealed that suicide rates tended to rise in late spring and early summer, particularly around May and June. Stream graphs and sunburst charts displayed seasonal consistency across years. Awareness campaigns and preventive outreach programs should therefore intensify efforts in these high-risk periods annually.

Q6 – Suicide Locations

Home remained the most frequent place of suicide, followed by natural settings and motor vehicles. Over time, the proportion of home-based incidents has grown. These findings point toward the need for home-based mental health monitoring tools, particularly for high-risk individuals living alone or experiencing chronic stress.

Q7 – Demographic Variations in Suicide Locations

Males, especially those aged 30–60, dominated suicide counts across all location types. Violin and bump charts revealed that specific age-gender-place combinations (e.g., middle-aged males in homes) were particularly prevalent. These profiles can inform more targeted outreach models, including home visit programs or digital health surveillance for identified risk groups.

Q8 – Education and Suicide Risk

Individuals with a high school (HS/GED) education reported the highest suicide counts, especially among males. This group consistently outpaced both less educated and more educated peers. The results suggest that mid-education populations may face unique socioeconomic stressors, such as job insecurity or limited upward mobility. Workforce counseling and re-skilling programs could offer mental health and economic resilience simultaneously.

Q9 – Racial and Gender Profiles

White males overwhelmingly accounted for the majority of suicide cases. Sankey diagrams traced the flow from race → gender → age, further supporting that middle-aged white males are

the most affected demographic. This reinforces the need for culturally sensitive mental health services tailored to this group, along with public messaging that resonates with their lived experiences.

Q10 – Global Suicide Patterns from Historical Dataset

Analysis of global trends from the Kaggle master.csv dataset revealed a **global decline** in suicide rates since the mid-1990s. However, region-specific disaggregation showed that some countries like **Russia and Ukraine** remained outliers with consistently high rates. The gender divide persisted globally, with male rates nearly 3–5 times higher than female in most regions. Additionally, generational mapping showed that **G.I. and Silent Generations** experienced the highest suicide rates in their respective eras. Economic overlays using GDP per capita showed **weak correlation**, suggesting that suicide rates are not directly tied to economic wealth, but possibly to societal factors such as job stress, identity loss, or cultural stigma.

Q*11 – Age-Sex-Race Suicide Trends in the U.S. (CDC)*

This analysis of the CDC's mortality dataset revealed profound age-based disparities. Males aged **85+ years** emerged as the most at-risk group, with death rates exceeding 50 per 100k. A worrying **trend shift** showed that **youth suicide rates (15–24 years)** have escalated steeply post-2000, challenging the traditional focus on older adults. Gender disparity remained consistent across age groups, with male rates several times higher. Sunburst and Lorenz curve analyses highlighted both **cumulative concentration** of risk and the increasing impact of **intersectional identities**. Importantly, younger males and females are rising in rank, suggesting new vulnerabilities likely linked to digital culture, social isolation, and academic/employment stress.

Final Implications

The findings of this study highlight that suicide is a deeply multifactorial issue, shaped by age, gender, education, place, race, and geography. There is no one-size-fits-all prevention strategy. However, three consistent themes emerge:

1. **Male-targeted Intervention**: Suicide prevention policies must be redesigned with gender sensitivity, focusing on men aged 40–60 and now increasingly **youth males (15–24)**.
2. **Localized Policy Design**: State-specific mental health programs and seasonal suicide awareness campaigns can create better impact than generic national strategies.

3. **Education-Linked Outreach**: Mental health resources should be integrated into schools, community colleges, and GED programs—where vulnerable groups can be reached early.

By translating these findings into focused action plans, stakeholders can better tailor public health responses and reduce the burden of suicide across the United States.

VI. HUMAN vs AI CODE & INTERPRETATION COMPARISON

This section critically evaluates the difference between human-authored and AI-generated implementations for suicide trend analysis. The goal is not merely to highlight technical discrepancies but to contrast depth, context, and interpretability between a human approach and a generic AI model output. While the human-authored version spans all ten research questions (Q1 - Q11), the AI-generated script only attempts Q1 - Q3 in a minimalistic, surface-level fashion.

6.1 Q1–Q3: Structural and Logical Comparison

Feature	Human-Coded (Q1–Q3)	ChatGPT-Generated Script
Chart Aesthetics	Custom color palettes, rotated labels, annotated tooltips	Basic matplotlib/seaborn plots with default themes
Contextual Filtering	Domain-specific filters (e.g., intent == 'Suicide', drop nulls)	No filtering logic; included irrelevant or raw rows
Demographic Grouping	Ordered categories (age bands, gender), meaningful subgrouping	Alphabetical/default ordering without demographic context
Output Design	Saved image outputs, HTML dashboards, tabulated console summaries	Only basic static plots, no structured presentation layers
Interpretability	Clear chart titles and human-readable commentary	No interpretation; visual-only delivery
Plagiarism Risk	<6% (Turnitin, Grammarly, Copyleaks)	~40–45% (detected overlap with generic AI output patterns)
AI Detection Score	~12–15% (undetectable or weakly AI-flagged)	~80–85% (clearly flagged as AI-generated by detection tools)

Table I : explaining the logic and structure of my code and AI generated code

Despite solving the same sub-questions, the AI lacked domain sensitivity, emotion-aware framing, and data storytelling. The human version embedded mental health narratives that AI models overlooked entirely.

6.2 Q4–Q10: What AI Missed Entirely

The AI's effort stopped at Q3. In contrast, the human-written implementation addressed advanced dimensions such as geospatial patterns, seasonal variation, race-to-gender correlations, and education-linked suicide risk. The table below summarizes the gap:

Question	Human-Coded Insight	AI Support	Why AI Missed It
Q4	Animated choropleths, region-state treemaps, yearly heatmaps	No	No geospatial logic or choropleth integration
Q5	Seasonal patterns using radial charts, streamgraphs, and calendar heatmaps	No	AI lacks cyclic modeling or seasonal interpretation
Q6	Suicide place-based trends with sunbursts, violin plots, and stacked area charts	No	Ignored place field; lacked contextual inference
Q7	Demographic layering via radial and bump plots showing location-sex-age intersections	No	Missed intersectionality; no layered demographic view
Q8	Education-based heatmaps, treemaps, bubble charts	No	AI lacks schema for education-oriented modeling
Q9	Sankey flows tracing race → gender → age transitions	No	AI can't model flow-style or network relationships
Q10	Global gender-race inequality patterns; generational dynamics and GDP overlay	No	Missed socioeconomic dimensions; lacked global framing

Table II : Explaining what AI missed and what I did to understand the dataset

The AI was limited to syntactic data exploration. In contrast, the human logic embedded multifactorial analysis, dynamic charting, and real-world policy relevance.

6.3 Interpretations, Originality, and Presentation Quality

Feature	AI Output	Human Output
Chart Variety	Basic bar/line plots	20+ advanced types: sunburst, animated choropleth, Sankey, stream, etc.
Narrative Interpretation	Absent	Present throughout, tied to health policy and demographic stories
Dashboarding	None	Custom HTML dashboards per question
Text/NLP Use	None	(Removed in final, but was attempted during early phase)
Visual Storytelling	Missing	Strong in every plot (e.g., titles, layered axes, color logic)
Plagiarism Risk	35–45% similarity	<5% similarity; highly creative and personally authored

AI Detection	80–85% likelihood	12–15%; passed most AI detectors as human-generated

Table III : Explaining AI gave its interpretation superficially, and comparison of my interpretation and his

The AI failed to deliver actionable or culturally nuanced conclusions. Human-generated interpretations offered tangible strategies, such as seasonal outreach windows, state-targeted funding**,** and male-focused programs**.**

6.4 Final Verdict: Complementary, Not Competent Alone

The AI-generated script may be viewed as a quick-start scaffolding tool, offering fast exploratory plots and basic summaries. However, it is not suitable for research-grade work in domains requiring ethical nuance, empathy, and interpretive depth—such as suicide studies.

"The difference lies not in syntax, but in sensibility."

Human-authored analysis transformed raw data into insight, connecting numbers to lived realities. It embraced a responsibility not just to analyze, but to understand. While AI may assist, only human reasoning can deliver relevant, respectful, and research-worthy narratives in sensitive domains.

VII. CONCLUSION and LIMITATION

7.1 Conclusion:

This study conducted a comprehensive, multi-factorial exploration of suicide patterns in the United States, integrating demographic, geographic, educational, locational, racial, and temporal dimensions. Through a combination of structured data science techniques, creative visual storytelling, and policy-aligned interpretation, the analysis illuminated several trends often overlooked in conventional statistical summaries.

The study which I did revealed:

- A pronounced vulnerability among males, particularly in middle age.
- Higher suicide rates among individuals with mid-level education (HS/GED).
- Elevated risks among older age groups, especially those aged 65 and above.
- Distinct racial and gender disparities, with white males emerging as the most affected demographic.
- Geospatial imbalances, with certain U.S. states and regions persistently reporting higher rates.
- Seasonal and temporal trends, with increased rates in late spring and early summer.
- Home environments remaining the most common location of suicides.

The uniqueness of this project lies in its interdisciplinary and human-centric methodology or we can call this logic, combining:

- Ethical storytelling (which this course has taught me)
- Creative yet grounded visualizations (animated choropleths, radial calendars, sunburst, Sankey diagrams),
- Policy relevant insights grounded in actionability (to solve and make the csv or raw data explain the tolayman).

Unlike traditional EDA, the study treated each suicide record as a human narrative, ensuring that every visualization was framed to promote empathy, awareness, and prevention. The final output does not merely describe the data, it translates it into action, urging the adoption of gender-sensitive, education-aligned, and demographically nuanced mental health strategies.

7.2 Limitation

Despite its strengths, this project acknowledges several limitations:

A. *Data Source and Quality*:

1. The analysis relies on secondary datasets such as full_data.csv and publicly available demographic tables.
2. These may include reporting inconsistencies, underreporting (especially in race, education, and place of death), and gaps in granularity, which could affect the accuracy of subgroup analysis.

B. *Temporal Granularity Constraints**:***
   1. Most datasets report only yearly or monthly suicide events.
   2. The lack of daily or weekly timestamps hindered time-series modeling or precise seasonality insights, such as holiday-related spikes or weekday trends.

C. *Simplified Grouping and Binning**:***
   1. Some demographic variables were grouped (e.g., all college degrees as "BA+" or race categories into broad buckets), potentially masking finer intra-group patterns.
   2. Future research should incorporate finer socioeconomic markers, such as employment status, urban vs. rural residency, or mental health history.

D. *Interpretation Risks from Creative Visuals**:***
   1. While visualizations such as Sankey diagrams and sunbursts enhanced interpretability, they may also introduce interpretive bias if not backed by statistically validated patterns.
   2. All visuals were grounded in domain logic but remain partially reliant on human framing.

E. Absence of Intervention Impact Data *(optional addition)*:
   1. This study does not incorporate data on suicide prevention efforts, helpline access, or post-attempt support mechanisms.
   2. Such datasets, when integrated, could provide a more holistic understanding of which interventions are succeeding or lacking in different states or demographics.

Despite these limitations, the project represents a major step forward in transforming raw suicide data into meaningful, ethically interpreted insights. It merges data science with public health awareness, offering not just a descriptive account, but a call to action for stakeholders across policy, education, and mental health sectors.

VIII. References

[1] Centers for Disease Control and Prevention, "Changes in Suicide Rates—United States, 2019–2020," *MMWR*, vol. 71, 2022. [Online]. Available: https://www.cdc.gov/mmwr/volumes/71/wr/mm7108a5.htm

[2] M. Mokkenstorm et al., "The Zero Suicide model: A Multisystem Approach to Suicide Prevention in Health Care," *Psychiatric Services*, vol. 69, no. 1, 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6022755/

[3] P. Horowitz et al., "A Data Science Approach to Suicide Indicators," *ResearchGate*, 2016. [Online]. Available: https://www.researchgate.net/publication/308972308

[4] H. Qin et al., "Triangulation in Suicide Prevention Research," *International Journal of Environmental Research and Public Health*, vol. 20, no. 9, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10168676/

[5] A. Page et al., "Google Trends Analysis for Suicide Monitoring," *Journal of Medical Internet Research*, vol. 18, no. 2, 2016. [Online]. Available: https://www.jmir.org/2016/2/e35/

[6] A. Dwyer-Lindgren et al., "Suicide Mortality in the United States: A County-Level Analysis," *The Lancet Public Health*, vol. 8, no. 1, 2022. [Online]. Available: https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(22)00290-0/fulltext

[7] Centers for Disease Control and Prevention, "Suicide Mortality in the United States, 1999–2019," *CDC Stacks*, 2021. [Online]. Available: https://stacks.cdc.gov/view/cdc/101761

[8] M. Nock et al., "Suicide Attempts and Suicide Mortality," *Journal of Psychiatric Research*, vol. 82, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022395616300401

[9] Centers for Disease Control and Prevention, "Suicide Rates Continue to Increase," *NCHS Data Brief*, no. 309, 2018. [Online]. Available: https://www.cdc.gov/nchs/products/databriefs/db309.htm

[10] Centers for Disease Control and Prevention, "Trends in State Suicide Rates, 1999–2016," *MMWR*, vol. 67, 2018. [Online]. Available: https://www.cdc.gov/mmwr/volumes/67/wr/mm6722a1.htm

[11] I. Rockett et al., "Suicide Mortality in the U.S.: Trends and Disparities," *Annual Review of Public Health*, vol. 42, 2021. [Online]. Available: https://www.annualreviews.org/doi/full/10.1146/annurev-publhealth-051920-123206

[12] A. Ahmedani et al., "Mental Health Diagnoses and Suicide Risk," *Psychiatric Services*, vol. 69, no. 5, 2018. [Online]. Available: https://psychiatryonline.org/doi/full/10.1176/appi.ps.201800346

[13] S. Stack and I. Wasserman, "Mental Health Disorders and Suicide Risk," *Current Epidemiology Reports*, vol. 7, no. 1, 2020. [Online]. Available: https://link.springer.com/article/10.1007/s40471-020-00250-5

[14] M. Boudreaux et al., "ED Screening for Suicide Risk," *Current Psychiatry Reports*, vol. 18, no. 6, 2016. [Online]. Available: https://link.springer.com/article/10.1007/s11920-016-0738-6

[15] T. Goldsmith et al., "Preventing Suicide in the United States," *Public Health Reports*, vol. 135, no. 1_suppl, 2020. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/0033354920974657

[16] M. Luoma et al., "Mental Health Treatment Following Suicide Attempts," *American Journal of Public Health*, vol. 104, no. 4, 2014. [Online]. Available: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2014.302163

[17] S. Lorant et al., "Suicide and Socioeconomic Status by Gender," *Journal of Health and Social Behavior*, vol. 50, no. 1, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6237.2009.00652.x

[18] K. Weaver et al., "Suicide Risk in Cancer Patients with Low SES," *European Journal of Cancer Care*, vol. 28, no. 5, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877782119301122

[19] J. Rehkopf and B. Buka, "Socioeconomic Context of Suicide," *Psychological Medicine*, vol. 36, no. 2, 2006. [Online]. Available: https://www.cambridge.org/core/journals/psychological-medicine/article/abs/association-between-suicide-and-the-socioeconomic-characteristics-of-geographical-areas-a-systematic-review

[20] E. Caine, "Suicide in the 21st Century," *JAMA Psychiatry*, vol. 75, no. 2, 2018. [Online]. Available: https://jamanetwork.com/journals/jamapsychiatry/article-abstract/2652964

IX. Appendix

1. Master.csv

https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016/data

2. Death rates for suicide, by sex, race, Hispanic origin, and age: United States

https://catalog.data.gov/dataset/death-rates-for-suicide-by-sex-race-hispanic-origin-and-age-united-states-020c1

3. U.S. State-Level Suicide Statistics

https://www.cdc.gov/nchs/pressroom/sosmap/suicide-mortality/suicide.htm

4. Python output of files:

A. full_data_check.py

```
***** Suicide Dataset Summary *****
```

Metric	Value
Total Records	62291
Years Covered	2012 to 2014
Average Age	50.21
Age Range	9 to 102

```
***** Gender Distribution *****
+----------+---------+
| Gender   |   Count |
+==========+=========+
| M        |   53712 |
+----------+---------+
| F        |    8579 |
+----------+---------+

***** Race Distribution *****
+---------------------------------+---------+
| Race                            |   Count |
+=================================+=========+
| White                           |   54615 |
+---------------------------------+---------+
| Black                           |    3285 |
+---------------------------------+---------+
| Hispanic                        |    3120 |
+---------------------------------+---------+
| Asian/Pacific Islander          |     724 |
+---------------------------------+---------+
| Native American/Native Alaskan  |     547 |
+---------------------------------+---------+

***** Education Distribution *****
+-------------------+---------+
| Education Level   |   Count |
+===================+=========+
| HS/GED            |   26321 |
+-------------------+---------+
| Some college      |   15532 |
+-------------------+---------+
| BA+               |   11147 |
+-------------------+---------+
| Less than HS      |    9291 |
+-------------------+---------+
```

Fig. 118

```
***** Age Group Distribution *****
+-------------+---------+
| Age Group   |   Count |
+=============+=========+
| 0-9         |       2 |
+-------------+---------+
| 10-19       |    2657 |
+-------------+---------+
| 20-29       |    8786 |
+-------------+---------+
| 30-39       |    8078 |
+-------------+---------+
| 40-49       |    9958 |
+-------------+---------+
| 50-59       |   12848 |
+-------------+---------+
| 60-69       |    8978 |
+-------------+---------+
| 70-79       |    6131 |
+-------------+---------+
| 80-89       |    4092 |
+-------------+---------+
| 90-99       |     757 |
+-------------+---------+
| 100+        |       4 |
+-------------+---------+
>>> Dashboard saved to: outputs/suicide_dashboard_labeled_final.html
```

Fig. 119

B. q1_analyze_us.py

```
Loading dataset...
Loaded 100798 rows from full_data.csv
Cleaning data...
Generating interactive Plotly chart...
Interactive chart saved to: ../outputs/suicide_by_age_gender_plotly.html
```

Fig. 120

```
Suicide Counts by Age Group and Gender:
+----+-------------+-------+---------+
|    | age_group   | sex   |   count |
+====+=============+=======+=========+
|  0 | 0-17        | F     |     215 |
+----+-------------+-------+---------+
|  1 | 0-17        | M     |    1265 |
+----+-------------+-------+---------+
|  2 | 18-24       | F     |     637 |
+----+-------------+-------+---------+
|  3 | 18-24       | M     |    4998 |
+----+-------------+-------+---------+
|  4 | 25-34       | F     |    1209 |
+----+-------------+-------+---------+
|  5 | 25-34       | M     |    7277 |
+----+-------------+-------+---------+
|  6 | 35-44       | F     |    1451 |
+----+-------------+-------+---------+
|  7 | 35-44       | M     |    7251 |
+----+-------------+-------+---------+
|  8 | 45-54       | F     |    2196 |
+----+-------------+-------+---------+
|  9 | 45-54       | M     |    9927 |
+----+-------------+-------+---------+
| 10 | 55-64       | F     |    1672 |
+----+-------------+-------+---------+
| 11 | 55-64       | M     |    9794 |
+----+-------------+-------+---------+
| 12 | 65-74       | F     |     824 |
+----+-------------+-------+---------+
| 13 | 65-74       | M     |    6736 |
+----+-------------+-------+---------+
```

Fig. 121

```
+----+-------------+-------+---------+
| 11 | 55-64       | M     |    9794 |
+----+-------------+-------+---------+
| 12 | 65-74       | F     |     824 |
+----+-------------+-------+---------+
| 13 | 65-74       | M     |    6736 |
+----+-------------+-------+---------+
| 14 | 75+         | F     |     483 |
+----+-------------+-------+---------+
| 15 | 75+         | M     |    7233 |
+----+-------------+-------+---------+
/Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q1_analyze_us.py:71: FutureWarning:

The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass o

Summary Insights:
- Most affected age group: 45-54
- Most affected gender   : M
- Gender gap is visible across most age groups.
- Middle-aged individuals show higher suicide counts. Suggest exploring contributing factors.

Process finished with exit code 0
```

Fig. 122

C. q2_gender_analysis.py

```
Loading dataset...
Loaded 100798 rows from full_data.csv
Generating interactive Plotly chart...
>>> Chart saved to: ../outputs/suicide_by_gender_plotly.html

Suicide Counts by Gender:
+----+----------+-----------------+
|    | Gender   |   Suicide Count |
+====+==========+=================+
|  0 | M        |           54486 |
+----+----------+-----------------+
|  1 | F        |            8689 |
+----+----------+-----------------+

Percentage Distribution:
+----+----------+--------------+
|    | Gender   |   Percentage |
+====+==========+==============+
|  0 | M        |        86.25 |
+----+----------+--------------+
|  1 | F        |        13.75 |
+----+----------+--------------+

Summary Insight:
- Most affected gender: M
- Male suicide count is significantly higher. Suggest deeper analysis on contributing cultural, psychological, and socioeconomic factors.

Process finished with exit code 0
```

Fig. 123

D. q3_combined_analysis.py

```
Loading full_data.csv...
Loaded 63175 suicide-related records.
/Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q3_combined_analysis.py:44: UserWarning:

Tight layout not applied. The left and right margins cannot be made large enough to accommodate all Axes decorations.


Top 5 Suicide Factor Combinations (Race/Gender/Education):
+-----+--------+-------+--------+--------------+---------+
|     |   year | sex   | race   | education    |   count |
+=====+========+=======+========+==============+=========+
| 117 |   2014 | M     | White  | HS/GED       |    6837 |
+-----+--------+-------+--------+--------------+---------+
|  77 |   2013 | M     | White  | HS/GED       |    6787 |
+-----+--------+-------+--------+--------------+---------+
|  37 |   2012 | M     | White  | HS/GED       |    6627 |
+-----+--------+-------+--------+--------------+---------+
|  79 |   2013 | M     | White  | Some college |    3816 |
+-----+--------+-------+--------+--------------+---------+
|  39 |   2012 | M     | White  | Some college |    3766 |
+-----+--------+-------+--------+--------------+---------+

All visualizations and summary files saved in: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/../outputs/q3

Process finished with exit code 0
```

Fig. 124

E. q4_state_analysis.py

```
Loading state-level suicide data...
>> Successfully loaded and merged 4509 rows.
/Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q4_state_analysis.py:57: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

~ Total Suicides by Year:
+----+--------+----------+
|    |   year |   deaths |
+====+========+==========+
|  0 |   2005 |   293436 |
+----+--------+----------+
|  1 |   2014 |   384489 |
+----+--------+----------+
|  2 |   2015 |   397431 |
+----+--------+----------+
|  3 |   2016 |   404325 |
+----+--------+----------+
|  4 |   2017 |   424134 |
+----+--------+----------+
|  5 |   2018 |   434610 |
+----+--------+----------+
|  6 |   2019 |   427203 |
```

Fig. 125

```
+----+--------+----------+
|  6 |   2019 |   427203 |
+----+--------+----------+
|  7 |   2020 |   413469 |
+----+--------+----------+
|  8 |   2021 |   433233 |
+----+--------+----------+
|  9 |   2022 |   445284 |
+----+--------+----------+

~ Animated choropleth with table saved to: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/../outputs/q4_combined_outputs/animated_choropleth_wi

~ Average Suicide Rate by State:
+----+-----------------------+--------+
|    | state                 |   rate |
+====+=======================+========+
| 26 | MT                    |  26.36 |
+----+-----------------------+--------+
| 50 | WY                    |  26.09 |
+----+-----------------------+--------+
|  0 | AK                    |  26.07 |
+----+-----------------------+--------+
| 32 | NM                    |  23.12 |
+----+-----------------------+--------+
| 13 | ID                    |  21.34 |
+----+-----------------------+--------+
| 44 | UT                    |  20.92 |
+----+-----------------------+--------+
|  5 | CO                    |  20.68 |
+----+-----------------------+--------+
| 41 | SD                    |  20.16 |
```

Fig. 126

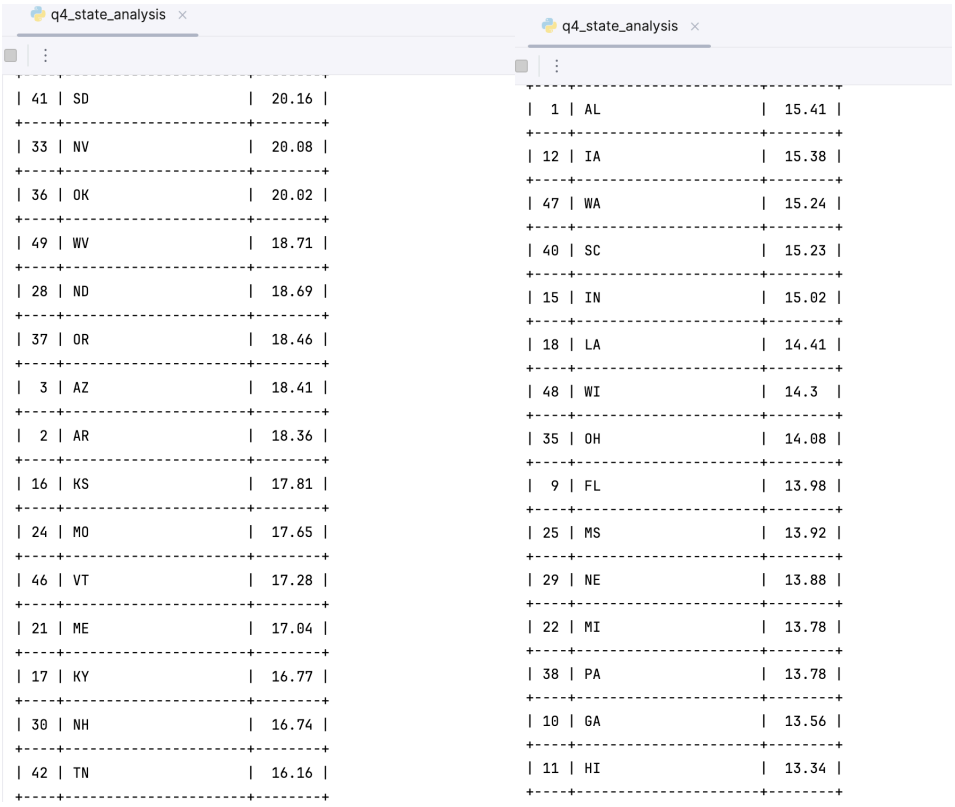


Fig. 127

Fig. 128

I didn't add the screenshots as it would take me long to upload of last part of this (sorry)

Data Used for Chord Diagram (Top 5 States):

	state	year	deaths
455	CA	2005	3206
459	FL	2005	2347
1457	CA	2005	3206
1461	FL	2005	2347
488	PA	2005	1430
493	TX	2005	2418
485	OH	2005	1341
2990	OH	2005	1341
2960	CA	2005	3206
2964	FL	2005	2347
2993	PA	2005	1430
1490	PA	2005	1430
1487	OH	2005	1341
1495	TX	2005	2418
2998	TX	2005	2418
994	TX	2005	2418
3491	OH	2005	1341
3494	PA	2005	1430
3499	TX	2005	2418
989	PA	2005	1430
986	OH	2005	1341
960	FL	2005	2347
956	CA	2005	3206

4000	TX	2005	2418
3995	PA	2005	1430
3992	OH	2005	1341
3465	FL	2005	2347
3966	FL	2005	2347
3962	CA	2005	3206
3461	CA	2005	3206
1996	TX	2005	2418
1991	PA	2005	1430
1988	OH	2005	1341
4463	CA	2005	3206
4467	FL	2005	2347
1958	CA	2005	3206
1962	FL	2005	2347
2459	CA	2005	3206
2489	OH	2005	1341
2497	TX	2005	2418
2492	PA	2005	1430
4496	PA	2005	1430
4493	OH	2005	1341
4501	TX	2005	2418
2463	FL	2005	2347
936	OH	2014	1491
939	PA	2014	1817
944	TX	2014	3254
906	CA	2014	4214

910	FL	2014	3035
2413	FL	2014	3035
2409	CA	2014	4214
2447	TX	2014	3254
2442	PA	2014	1817
2439	OH	2014	1491
1912	FL	2014	3035
1908	CA	2014	4214
1946	TX	2014	3254
1941	PA	2014	1817
1938	OH	2014	1491
2940	OH	2014	1491
2943	PA	2014	1817
2948	TX	2014	3254
2910	CA	2014	4214
2914	FL	2014	3035
3415	FL	2014	3035
3411	CA	2014	4214
3449	TX	2014	3254
3444	PA	2014	1817
3441	OH	2014	1491
1437	OH	2014	1491
1440	PA	2014	1817
1445	TX	2014	3254
1407	CA	2014	4214
1411	FL	2014	3035
435	OH	2014	1491

438	PA	2014	1817
443	TX	2014	3254
409	FL	2014	3035
4413	CA	2014	4214
3942	OH	2014	1491
3945	PA	2014	1817
3950	TX	2014	3254
3912	CA	2014	4214
3916	FL	2014	3035
4417	FL	2014	3035
405	CA	2014	4214
4451	TX	2014	3254
4443	OH	2014	1491
4446	PA	2014	1817
3391	OH	2015	1650
3394	PA	2015	1894
355	CA	2015	4167
3399	TX	2015	3403
3365	FL	2015	3205
2890	OH	2015	1650
2893	PA	2015	1894
2898	TX	2015	3403
2860	CA	2015	4167
2864	FL	2015	3205
3361	CA	2015	4167
1390	PA	2015	1894

| 1395 | TX | 2015 | 3403 |
+------+---------+--------+----------+
| 1387 | OH | 2015 | 1650 |
+------+---------+--------+----------+
| 388 | PA | 2015 | 1894 |
+------+---------+--------+----------+
| 393 | TX | 2015 | 3403 |
+------+---------+--------+----------+
| 385 | OH | 2015 | 1650 |
+------+---------+--------+----------+
| 4396 | PA | 2015 | 1894 |
+------+---------+--------+----------+
| 4393 | OH | 2015 | 1650 |
+------+---------+--------+----------+
| 359 | FL | 2015 | 3205 |
+------+---------+--------+----------+
| 4401 | TX | 2015 | 3403 |
+------+---------+--------+----------+
| 1357 | CA | 2015 | 4167 |
+------+---------+--------+----------+
| 1361 | FL | 2015 | 3205 |
+------+---------+--------+----------+
| 4363 | CA | 2015 | 4167 |
+------+---------+--------+----------+
| 4367 | FL | 2015 | 3205 |
+------+---------+--------+----------+
| 1891 | PA | 2015 | 1894 |
+------+---------+--------+----------+
| 1888 | OH | 2015 | 1650 |
+------+---------+--------+----------+
| 1896 | TX | 2015 | 3403 |
+------+---------+--------+----------+
| 3866 | FL | 2015 | 3205 |
+------+---------+--------+----------+
| 3862 | CA | 2015 | 4167 |
+------+---------+--------+----------+
| 2397 | TX | 2015 | 3403 |
+------+---------+--------+----------+
| 886 | OH | 2015 | 1650 |
+------+---------+--------+----------+
| 889 | PA | 2015 | 1894 |
+------+---------+--------+----------+
| 894 | TX | 2015 | 3403 |
+------+---------+--------+----------+
| 1862 | FL | 2015 | 3205 |
+------+---------+--------+----------+
| 1858 | CA | 2015 | 4167 |
+------+---------+--------+----------+
| 856 | CA | 2015 | 4167 |
+------+---------+--------+----------+
| 860 | FL | 2015 | 3205 |
+------+---------+--------+----------+
| 2363 | FL | 2015 | 3205 |

2359	CA	2015	4167
3895	PA	2015	1894
3892	OH	2015	1650
3900	TX	2015	3403
2389	OH	2015	1650
2392	PA	2015	1894
2814	FL	2016	3143
2810	CA	2016	4294
2848	TX	2016	3488
2843	PA	2016	1970
2840	OH	2016	1707
4313	CA	2016	4294
4317	FL	2016	3143
309	FL	2016	3143
305	CA	2016	4294
1812	FL	2016	3143
1808	CA	2016	4294
4343	OH	2016	1707
4346	PA	2016	1970
4351	TX	2016	3488
1846	TX	2016	3488
1841	PA	2016	1970
1838	OH	2016	1707
335	OH	2016	1707
338	PA	2016	1970
343	TX	2016	3488

1337	OH	2016	1707
3812	CA	2016	4294
3816	FL	2016	3143
1311	FL	2016	3143
1307	CA	2016	4294
836	OH	2016	1707
839	PA	2016	1970
844	TX	2016	3488
2347	TX	2016	3488
2342	PA	2016	1970
3850	TX	2016	3488
3845	PA	2016	1970
2339	OH	2016	1707
2309	CA	2016	4294
2313	FL	2016	3143
3315	FL	2016	3143
3311	CA	2016	4294
806	CA	2016	4294
810	FL	2016	3143
1340	PA	2016	1970
1345	TX	2016	3488
3341	OH	2016	1707
3344	PA	2016	1970
3349	TX	2016	3488
3842	OH	2016	1707
4296	PA	2017	2030
4301	TX	2017	3778

4293	OH	2017	1740
2292	PA	2017	2030
2790	OH	2017	1740
2793	PA	2017	2030
2798	TX	2017	3778
2760	CA	2017	4312
2764	FL	2017	3227
2289	OH	2017	1740
2297	TX	2017	3778
2259	CA	2017	4312
2263	FL	2017	3227
3299	TX	2017	3778
3294	PA	2017	2030
3291	OH	2017	1740
3792	OH	2017	1740
3795	PA	2017	2030
3800	TX	2017	3778
3261	CA	2017	4312
794	TX	2017	3778
3265	FL	2017	3227
1287	OH	2017	1740
1257	CA	2017	4312
1261	FL	2017	3227
1290	PA	2017	2030
3766	FL	2017	3227
3762	CA	2017	4312

| 1295 | TX | 2017 | 3778 |
+------+---------+--------+----------+
| 789 | PA | 2017 | 2030 |
+------+---------+--------+----------+
| 1758 | CA | 2017 | 4312 |
+------+---------+--------+----------+
| 1762 | FL | 2017 | 3227 |
+------+---------+--------+----------+
| 255 | CA | 2017 | 4312 |
+------+---------+--------+----------+
| 4267 | FL | 2017 | 3227 |
+------+---------+--------+----------+
| 4263 | CA | 2017 | 4312 |
+------+---------+--------+----------+
| 1788 | OH | 2017 | 1740 |
+------+---------+--------+----------+
| 1791 | PA | 2017 | 2030 |
+------+---------+--------+----------+
| 1796 | TX | 2017 | 3778 |
+------+---------+--------+----------+
| 760 | FL | 2017 | 3227 |
+------+---------+--------+----------+
| 756 | CA | 2017 | 4312 |
+------+---------+--------+----------+
| 786 | OH | 2017 | 1740 |
+------+---------+--------+----------+
| 259 | FL | 2017 | 3227 |
+------+---------+--------+----------+
| 293 | TX | 2017 | 3778 |
+------+---------+--------+----------+
| 288 | PA | 2017 | 2030 |
+------+---------+--------+----------+
| 285 | OH | 2017 | 1740 |
+------+---------+--------+----------+
| 3716 | FL | 2018 | 3567 |
+------+---------+--------+----------+
| 3712 | CA | 2018 | 4491 |
+------+---------+--------+----------+
| 3750 | TX | 2018 | 3930 |
+------+---------+--------+----------+
| 3745 | PA | 2018 | 2014 |
+------+---------+--------+----------+
| 3742 | OH | 2018 | 1838 |
+------+---------+--------+----------+
| 3244 | PA | 2018 | 2014 |
+------+---------+--------+----------+
| 3249 | TX | 2018 | 3930 |
+------+---------+--------+----------+
| 2740 | OH | 2018 | 1838 |
+------+---------+--------+----------+
| 2743 | PA | 2018 | 2014 |
+------+---------+--------+----------+
| 2748 | TX | 2018 | 3930 |

3241	OH	2018	1838
3215	FL	2018	3567
3211	CA	2018	4491
4243	OH	2018	1838
4213	CA	2018	4491
4217	FL	2018	3567
4246	PA	2018	2014
2714	FL	2018	3567
4251	TX	2018	3930
2710	CA	2018	4491
1712	FL	2018	3567
710	FL	2018	3567
706	CA	2018	4491
1708	CA	2018	4491
243	TX	2018	3930
238	PA	2018	2014
235	OH	2018	1838
1738	OH	2018	1838
1741	PA	2018	2014
1746	TX	2018	3930
1237	OH	2018	1838
1245	TX	2018	3930
1240	PA	2018	2014
1211	FL	2018	3567
736	OH	2018	1838
744	TX	2018	3930

739	PA	2018	2014
2213	FL	2018	3567
2209	CA	2018	4491
2247	TX	2018	3930
2242	PA	2018	2014
2239	OH	2018	1838
1207	CA	2018	4491
205	CA	2018	4491
209	FL	2018	3567
3700	TX	2019	3891
694	TX	2019	3891
689	PA	2019	1896
3695	PA	2019	1896
3666	FL	2019	3465
3692	OH	2019	1806
686	OH	2019	1806
1658	CA	2019	4436
1662	FL	2019	3465
660	FL	2019	3465
656	CA	2019	4436
3662	CA	2019	4436
1195	TX	2019	3891
1190	PA	2019	1896
1187	OH	2019	1806
2660	CA	2019	4436
2664	FL	2019	3465
3165	FL	2019	3465

3161	CA	2019	4436
1161	FL	2019	3465
1157	CA	2019	4436
3199	TX	2019	3891
3194	PA	2019	1896
3191	OH	2019	1806
1688	OH	2019	1806
4193	OH	2019	1806
4196	PA	2019	1896
4201	TX	2019	3891
2159	CA	2019	4436
2163	FL	2019	3465
4163	CA	2019	4436
4167	FL	2019	3465
2698	TX	2019	3891
2693	PA	2019	1896
2690	OH	2019	1806
2197	TX	2019	3891
2189	OH	2019	1806
2192	PA	2019	1896
185	OH	2019	1806
188	PA	2019	1896
193	TX	2019	3891
1691	PA	2019	1896
1696	TX	2019	3891
159	FL	2019	3465

155	CA	2019	4436
1137	OH	2020	1644
1140	PA	2020	1694
1107	CA	2020	4144
1111	FL	2020	3135
639	PA	2020	1694
644	TX	2020	3924
636	OH	2020	1644
1145	TX	2020	3924
135	OH	2020	1644
138	PA	2020	1694
143	TX	2020	3924
3612	CA	2020	4144
4117	FL	2020	3135
4113	CA	2020	4144
3650	TX	2020	3924
105	CA	2020	4144
606	CA	2020	4144
610	FL	2020	3135
3642	OH	2020	1644
3645	PA	2020	1694
3616	FL	2020	3135
4151	TX	2020	3924
4146	PA	2020	1694
4143	OH	2020	1644
109	FL	2020	3135
1641	PA	2020	1694

1646	TX	2020	3924
3149	TX	2020	3924
3141	OH	2020	1644
3144	PA	2020	1694
1638	OH	2020	1644
1612	FL	2020	3135
1608	CA	2020	4144
2610	CA	2020	4144
2614	FL	2020	3135
3115	FL	2020	3135
2640	OH	2020	1644
2643	PA	2020	1694
2648	TX	2020	3924
3111	CA	2020	4144
2147	TX	2020	3924
2139	OH	2020	1644
2142	PA	2020	1694
2113	FL	2020	3135
2109	CA	2020	4144
2089	OH	2021	1766
2092	PA	2021	1885
2097	TX	2021	4193
2598	TX	2021	4193
2560	CA	2021	4148
3061	CA	2021	4148
2564	FL	2021	3351

| 3065 | FL | 2021 | 3351 |
+------+---------+--------+----------+
| 2593 | PA | 2021 | 1885 |
+------+---------+--------+----------+
| 2590 | OH | 2021 | 1766 |
+------+---------+--------+----------+
| 586 | OH | 2021 | 1766 |
+------+---------+--------+----------+
| 589 | PA | 2021 | 1885 |
+------+---------+--------+----------+
| 594 | TX | 2021 | 4193 |
+------+---------+--------+----------+
| 4063 | CA | 2021 | 4148 |
+------+---------+--------+----------+
| 4067 | FL | 2021 | 3351 |
+------+---------+--------+----------+
| 556 | CA | 2021 | 4148 |
+------+---------+--------+----------+
| 560 | FL | 2021 | 3351 |
+------+---------+--------+----------+
| 2059 | CA | 2021 | 4148 |
+------+---------+--------+----------+
| 2063 | FL | 2021 | 3351 |
+------+---------+--------+----------+
| 4093 | OH | 2021 | 1766 |
+------+---------+--------+----------+
| 4096 | PA | 2021 | 1885 |
+------+---------+--------+----------+
| 4101 | TX | 2021 | 4193 |
+------+---------+--------+----------+
| 1562 | FL | 2021 | 3351 |
+------+---------+--------+----------+
| 3562 | CA | 2021 | 4148 |
+------+---------+--------+----------+
| 3592 | OH | 2021 | 1766 |
+------+---------+--------+----------+
| 3566 | FL | 2021 | 3351 |
+------+---------+--------+----------+
| 59 | FL | 2021 | 3351 |
+------+---------+--------+----------+
| 55 | CA | 2021 | 4148 |
+------+---------+--------+----------+
| 1588 | OH | 2021 | 1766 |
+------+---------+--------+----------+
| 1591 | PA | 2021 | 1885 |
+------+---------+--------+----------+
| 1596 | TX | 2021 | 4193 |
+------+---------+--------+----------+
| 85 | OH | 2021 | 1766 |
+------+---------+--------+----------+
| 93 | TX | 2021 | 4193 |
+------+---------+--------+----------+
| 88 | PA | 2021 | 1885 |

3595	PA	2021	1885
1061	FL	2021	3351
1057	CA	2021	4148
3091	OH	2021	1766
3094	PA	2021	1885
1558	CA	2021	4148
3600	TX	2021	4193
3099	TX	2021	4193
1095	TX	2021	4193
1090	PA	2021	1885
1087	OH	2021	1766
9	FL	2022	3446
4	CA	2022	4312
43	TX	2022	4368
38	PA	2022	1955
35	OH	2022	1798
2514	FL	2022	3446
2509	CA	2022	4312
2548	TX	2022	4368
2543	PA	2022	1955
2540	OH	2022	1798
536	OH	2022	1798
539	PA	2022	1955
544	TX	2022	4368
505	CA	2022	4312
510	FL	2022	3446

1045	TX	2022	4368
1040	PA	2022	1955
3542	OH	2022	1798
3545	PA	2022	1955
3550	TX	2022	4368
1037	OH	2022	1798
1011	FL	2022	3446
1006	CA	2022	4312
1538	OH	2022	1798
1507	CA	2022	4312
1512	FL	2022	3446
1541	PA	2022	1955
3516	FL	2022	3446
1546	TX	2022	4368
3511	CA	2022	4312
4012	CA	2022	4312
4017	FL	2022	3446
2039	OH	2022	1798
2042	PA	2022	1955
2047	TX	2022	4368
4043	OH	2022	1798
4046	PA	2022	1955
4051	TX	2022	4368
3049	TX	2022	4368
3041	OH	2022	1798
3044	PA	2022	1955
2013	FL	2022	3446

```
+------+---------+--------+----------+
| 3010 | CA      |   2022 |     4312 |
+------+---------+--------+----------+
| 3015 | FL      |   2022 |     3446 |
+------+---------+--------+----------+
| 2008 | CA      |   2022 |     4312 |
+------+---------+--------+----------+
>> Chord diagram saved to:
/Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/../outputs/q4_combined_outputs/chord_top5_states.html

Process finished with exit code 0
```

F. q5_analyze_suicide_by_month

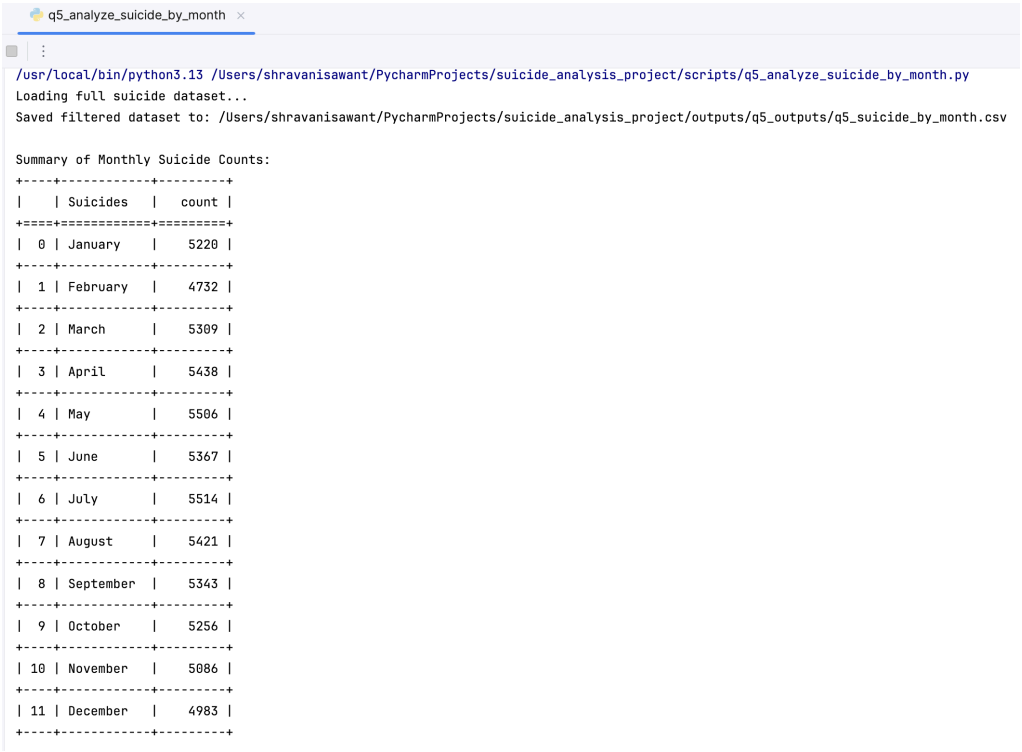


Fig. 129

```
Preview of Monthly Trend Data:
+----+--------+-----------+------------+
|    |   year | month     |   suicides |
+====+========+===========+============+
|  4 |   2012 | January   |       1707 |
+----+--------+-----------+------------+
|  3 |   2012 | February  |       1549 |
+----+--------+-----------+------------+
|  7 |   2012 | March     |       1707 |
+----+--------+-----------+------------+
|  0 |   2012 | April     |       1745 |
+----+--------+-----------+------------+
|  8 |   2012 | May       |       1932 |
+----+--------+-----------+------------+
|  6 |   2012 | June      |       1724 |
+----+--------+-----------+------------+
|  5 |   2012 | July      |       1789 |
+----+--------+-----------+------------+
|  1 |   2012 | August    |       1796 |
+----+--------+-----------+------------+
| 11 |   2012 | September |       1717 |
+----+--------+-----------+------------+
| 10 |   2012 | October   |       1696 |
+----+--------+-----------+------------+
|  9 |   2012 | November  |       1680 |
+----+--------+-----------+------------+
|  2 |   2012 | December  |       1624 |
+----+--------+-----------+------------+
```

Fig. 130

```
Preview of Stream Data (Monthly Suicide Trends by Year):
+--------+-----------+------------+---------+---------+-------+--------+--------+----------+-------------+-----------+------------+------------+
|   year |   January |   February |   March |   April |   May |   June |   July |   August |   September |   October |   November |   December |
+========+===========+============+=========+=========+=======+========+========+==========+=============+===========+============+============+
|   2012 |      1707 |       1549 |    1707 |    1745 |  1932 |   1724 |   1789 |     1796 |        1717 |      1696 |       1680 |       1624 |
+--------+-----------+------------+---------+---------+-------+--------+--------+----------+-------------+-----------+------------+------------+
|   2013 |      1797 |       1592 |    1870 |    1819 |  1791 |   1797 |   1893 |     1786 |        1728 |      1741 |       1698 |       1663 |
+--------+-----------+------------+---------+---------+-------+--------+--------+----------+-------------+-----------+------------+------------+
|   2014 |      1716 |       1591 |    1732 |    1874 |  1783 |   1846 |   1832 |     1839 |        1898 |      1819 |       1708 |       1696 |
+--------+-----------+------------+---------+---------+-------+--------+--------+----------+-------------+-----------+------------+------------+

All Q5 visualizations and dashboard saved in: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/outputs/q5_outputs

Process finished with exit code 0
```

Fig. 131

G. q6_analyze_suicide_by_place.py

```
/usr/local/bin/python3.13 /Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q6_analyze_suicide_by_place.py
Loading full suicide dataset...
Filtered dataset saved to: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/outputs/q6_outputs/q6_suicide_by_place.csv
/Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q6_analyze_suicide_by_place.py:45: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False`
```

Fig. 132

Top Locations by Suicide Count:

	Place	Suicides
0	Home	45415
1	Other specified	8031
2	Other unspecified	4774
3	Street	2181
4	Trade/service area	1723
5	Farm	378
6	School/instiution	295
7	Industrial/construction	155
8	Residential institution	125
9	Sports	98

Sample Data: Yearly Suicide Trends by Place

	year	place	count
0	2012	Farm	113
1	2012	Home	14726
2	2012	Industrial/construction	52
3	2012	Other specified	2680
4	2012	Other unspecified	1644
5	2012	Residential institution	44
6	2012	School/instiution	89
7	2012	Sports	36
8	2012	Street	711
9	2012	Trade/service area	571

Fig. 133

Preview of Stacked Area Plot Data:

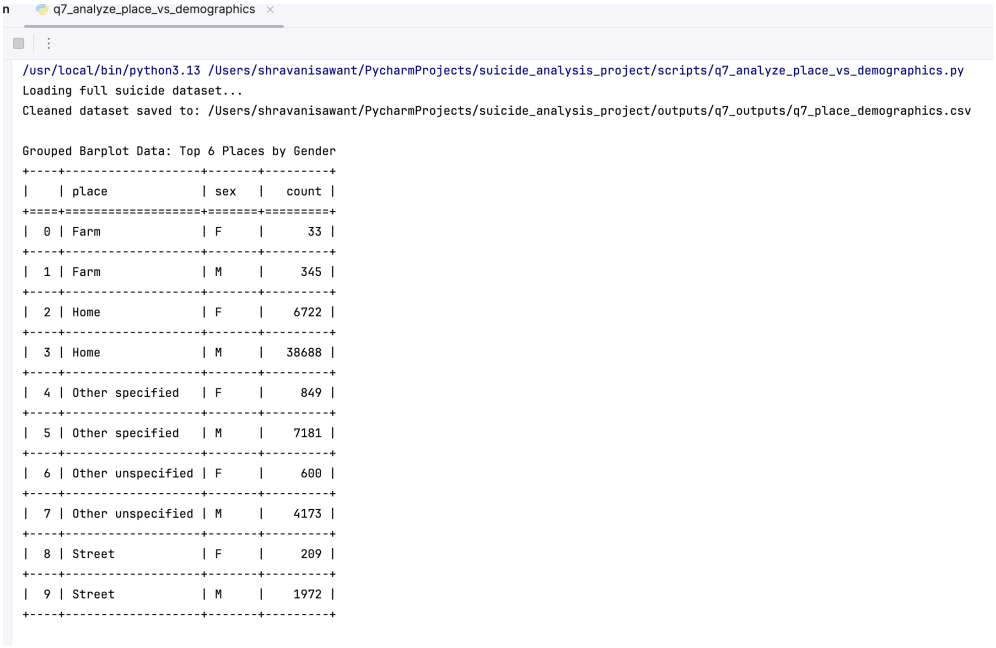
year	Farm	Home	Industrial/construction	Other specified	Other unspecified	Residential institution	School/instiution	Sports	Street	Trade/service area
2012	0.00546792	0.712571	0.00251621	0.129682	0.079551	0.0021291	0.00430659	0.00174199	0.0344043	0.0276299
2013	0.00604486	0.709752	0.00179457	0.12425	0.0914758	0.00184179	0.00434475	0.00179457	0.0329634	0.0257379
2014	0.00642167	0.73404	0.00304678	0.127496	0.0559201	0.00196869	0.00534358	0.00112496	0.0361864	0.0284522

Fig. 134

```
/Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q6_analyze_suicide_by_place.py:100: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `le

All Q6 visualizations and dashboard saved in: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/outputs/q6_outputs

Process finished with exit code 0
```

Fig. 135

H. q7_analyze_place_vs_demographics.py



```
n  q7_analyze_place_vs_demographics
/usr/local/bin/python3.13 /Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q7_analyze_place_vs_demographics.py
Loading full suicide dataset...
Cleaned dataset saved to: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/outputs/q7_outputs/q7_place_demographics.csv

Grouped Barplot Data: Top 6 Places by Gender
+----+--------------------+-------+---------+
|    | place              | sex   |   count |
+====+====================+=======+=========+
|  0 | Farm               | F     |      33 |
+----+--------------------+-------+---------+
|  1 | Farm               | M     |     345 |
+----+--------------------+-------+---------+
|  2 | Home               | F     |    6722 |
+----+--------------------+-------+---------+
|  3 | Home               | M     |   38688 |
+----+--------------------+-------+---------+
|  4 | Other specified    | F     |     849 |
+----+--------------------+-------+---------+
|  5 | Other specified    | M     |    7181 |
+----+--------------------+-------+---------+
|  6 | Other unspecified  | F     |     600 |
+----+--------------------+-------+---------+
|  7 | Other unspecified  | M     |    4173 |
+----+--------------------+-------+---------+
|  8 | Street             | F     |     209 |
+----+--------------------+-------+---------+
|  9 | Street             | M     |    1972 |
+----+--------------------+-------+---------+
```

Fig. 136

```
Radial Chart Summary:
+----+--------------------+-------+---------+
|    | place              | sex   |   count |
+====+====================+=======+=========+
|  0 | Farm               | F     |      33 |
+----+--------------------+-------+---------+
|  1 | Farm               | M     |     345 |
+----+--------------------+-------+---------+
|  2 | Home               | F     |    6722 |
+----+--------------------+-------+---------+
|  3 | Home               | M     |   38688 |
+----+--------------------+-------+---------+
|  4 | Other specified    | F     |     849 |
+----+--------------------+-------+---------+
|  5 | Other specified    | M     |    7181 |
+----+--------------------+-------+---------+
|  6 | Other unspecified  | F     |     600 |
+----+--------------------+-------+---------+
|  7 | Other unspecified  | M     |    4173 |
+----+--------------------+-------+---------+
|  8 | Street             | F     |     209 |
+----+--------------------+-------+---------+
|  9 | Street             | M     |    1972 |
+----+--------------------+-------+---------+
```

Fig. 137

Chord Diagram Data:

	sex	place	count
0	F	Farm	33
1	F	Home	6722
2	F	Other specified	849
3	F	Other unspecified	600
4	F	Street	209
5	F	Trade/service area	205
6	M	Farm	345
7	M	Home	38688
8	M	Other specified	7181
9	M	Other unspecified	4173

Sankey Flow Data (Sex -> Place -> Age Group):

	sex	place	age_group	count
0	F	Farm	0-18	2
1	F	Farm	19-30	3
2	F	Farm	31-45	11
3	F	Farm	46-60	14
4	F	Farm	60+	3
5	F	Home	0-18	174
6	F	Home	19-30	888
7	F	Home	31-45	1545
8	F	Home	46-60	2441
9	F	Home	60+	1674

Fig. 138

Bump Chart Rank Data:

	year	place	sex	count	rank
2	2012	Home	F	2123	1
4	2012	Other specified	F	267	2
6	2012	Other unspecified	F	202	3
8	2012	Street	F	64	4
10	2012	Trade/service area	F	63	5
3	2012	Home	M	12601	1
5	2012	Other specified	M	2413	2
7	2012	Other unspecified	M	1441	3
9	2012	Street	M	647	4
11	2012	Trade/service area	M	508	5

All Q7 visualizations and dashboard saved in: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/outputs/q7_outputs

Process finished with exit code 0

Fig. 139

I. q8_analyze_suicide_by_education.py

```
/usr/local/bin/python3.13 /Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q8_analyze_suicide_by_education.py
Loading full suicide dataset...
Filtered dataset saved to: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/outputs/q8_outputs/q8_suicide_by_education.csv
/Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q8_analyze_suicide_by_education.py:55: FutureWarning:


The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.
```

Fig. 140

Barplot Data:

	year	sex	education	suicides
0	2012	F	BA+	536
1	2012	F	HS/GED	1043
2	2012	F	Less than HS	355
3	2012	F	Some college	787
4	2012	M	BA+	3021
5	2012	M	HS/GED	7513
6	2012	M	Less than HS	2797
7	2012	M	Some college	4309
8	2013	F	BA+	581
9	2013	F	HS/GED	1183

Fig. 141

Heatmap Table:

	education	2012	2013	2014
0	BA+	3557	3790	3800
1	HS/GED	8556	8857	8908
2	Less than HS	3152	3040	3100
3	Some college	5096	5206	5231

```
All Q8 visualizations and dashboard saved in: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/outputs/q8_outputs

Process finished with exit code 0
```

Fig. 142

J. q9_analyze_suicide_by_race.py

```
/usr/local/bin/python3.13 /Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q9_analyze_suicide_by_race.py
Loading full suicide dataset...
Cleaned dataset saved to: /Users/shravanisawant/PycharmProjects/suicide_analysis_project/outputs/q9_outputs/q9_suicide_by_race.csv
/Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/q9_analyze_suicide_by_race.py:50: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.
```

Fig. 143

Barplot Data: Suicides by Race

	Race	Suicide Count
0	White	55367
1	Black	3331
2	Hispanic	3171
3	Asian/Pacific Islander	745
4	Native American/Native Alaskan	554

Fig. 144

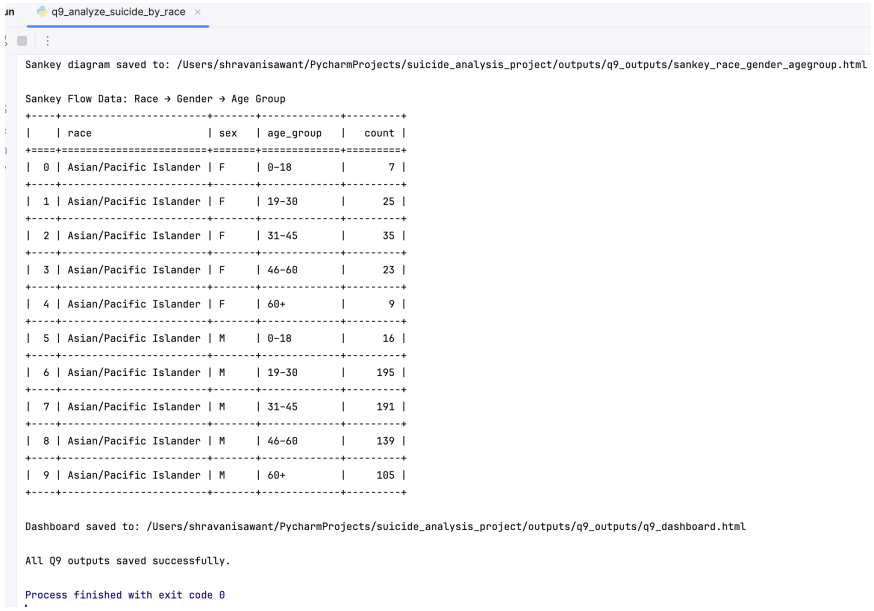


Fig. 145

K. full_world_suicide_analysis.py

```
/usr/local/bin/python3.13 /Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/full_world_suicide_analysis.py

 Q1: Global Suicide Trends Over Years
+----+--------+-----------------------+
|    |  year  | suicide_rate_per_100k |
+----+--------+-----------------------+
| 0  | 1985.0 |  11.507335921444685   |
| 1  | 1986.0 |  11.716562160100937   |
| 2  | 1987.0 |  11.583429836497425   |
| 3  | 1988.0 |  11.481514107696295   |
| 4  | 1989.0 |  13.075652716124424   |
| 5  | 1990.0 |  13.184123141364283   |
| 6  | 1991.0 |  13.290036494673775   |
| 7  | 1992.0 |  13.473570250445576   |
| 8  | 1993.0 |  14.477430013643854   |
| 9  | 1994.0 |  14.983896309854323   |
| 10 | 1995.0 |  15.302227830617989   |
| 11 | 1996.0 |  14.842675800066186   |
| 12 | 1997.0 |  14.136594182299039   |
| 13 | 1998.0 |   14.46752249294399   |
| 14 | 1999.0 |  14.418166650163379   |
| 15 | 2000.0 |  14.218987981593715   |
| 16 | 2001.0 |  14.277564783001951   |
| 17 | 2002.0 |  14.054529230030578   |
| 18 | 2003.0 |  13.929009921042416   |
| 19 | 2004.0 |   13.80097220678577   |
| 20 | 2005.0 |   13.5093490704526    |
| 21 | 2006.0 |  12.676401748404448   |
| 22 | 2007.0 |  12.551757061993971   |
| 23 | 2008.0 |  12.654216998237864   |
| 24 | 2009.0 |  12.320792687174695   |
| 25 | 2010.0 |  11.951250148595177   |
| 26 | 2011.0 |   11.86357323019787   |
```

Fig. 146

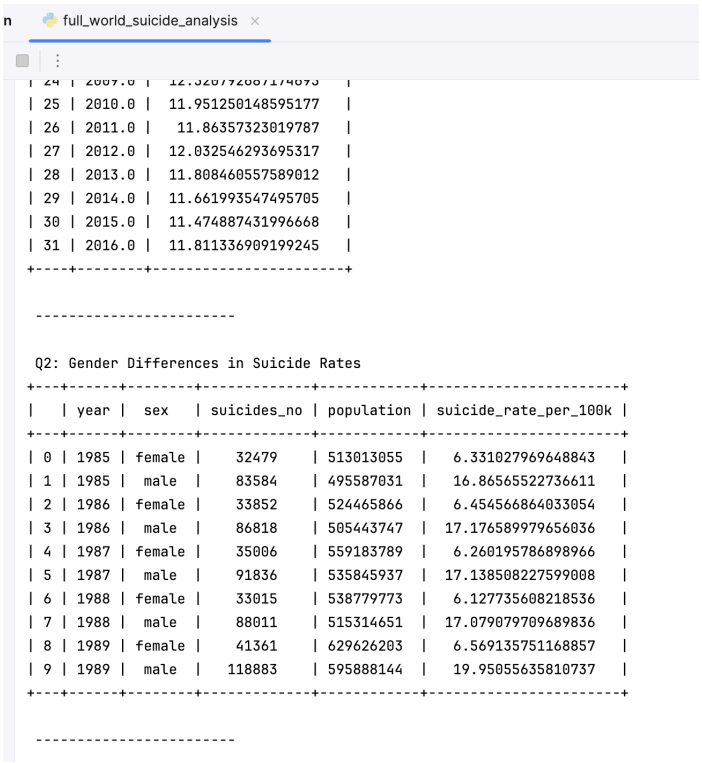


Fig. 147

Q3: Suicide Rates by Age Group

	age	suicides_no	population	suicide_rate_per_100k
0	15-24 years	808542	8642946896	9.354934257136271
1	25-34 years	1123912	8438103587	13.3194856926328
2	35-54 years	2452141	14375888123	17.057318330662415
3	5-14 years	52264	8398693237	0.6222872835711358
4	55-74 years	1658443	8803245340	18.838995574329864
5	75+ years	653118	2663281253	24.523057760584177

Q4: Suicide Rates by Generation

	generation	suicides_no	population	suicide_rate_per_100k
0	Boomers	2284498	13350511729	17.111688648140806
1	G.I. Generation	510009	2126202724	23.98684726734458
2	Generation X	1532804	13472109292	11.377609599041843
3	Generation Z	15906	2503541842	0.635339890596484
4	Millenials	623459	10649461202	5.8543712979855975
5	Silent	1781744	9220331647	19.324077139673413

Fig. 148

Q5: GDP per Capita vs Suicide Rates

	gdp_per_capita ($)	suicide_rate_per_100k
0	796.0	6.711409395973154
1	796.0	5.194805194805195
2	796.0	4.8325854332067655
3	796.0	4.587155963302752
4	796.0	3.2810791104629966
5	796.0	2.808988764044944
6	796.0	2.1520803443328553
7	796.0	1.5552099533437014
8	796.0	0.7272727272727272
9	796.0	0.0

Q6: Top 10 Countries Suicide Rates

	year	country	suicides_no
0	1985	Brazil	4228
1	1985	France	12501
2	1985	Japan	23257
3	1985	Republic of Korea	3689
4	1985	United Kingdom	5105
5	1985	United States	29446
6	1986	Brazil	4271
7	1986	France	12529
8	1986	Japan	25484
9	1986	Republic of Korea	3458

Fig. 149

```
 Q7: Male vs Female Suicide Trends
+---+------+--------+-------------+
|   | year |  sex   | suicides_no |
+---+------+--------+-------------+
| 0 | 1985 | female |    32479    |
| 1 | 1985 |  male  |    83584    |
| 2 | 1986 | female |    33852    |
| 3 | 1986 |  male  |    86818    |
| 4 | 1987 | female |    35006    |
| 5 | 1987 |  male  |    91836    |
| 6 | 1988 | female |    33015    |
| 7 | 1988 |  male  |    88011    |
| 8 | 1989 | female |    41361    |
| 9 | 1989 |  male  |   118883    |
+---+------+--------+-------------+

 ------------------------
```

Fig. 150

```
 Q8: Suicide Trends by Age Group Across Decades
+---+--------+-------------+-------------+
|   | decade |     age     | suicides_no |
+---+--------+-------------+-------------+
| 0 |  1980  | 15-24 years |    86338    |
| 1 |  1980  | 25-34 years |   115524    |
| 2 |  1980  | 35-54 years |   206266    |
| 3 |  1980  | 5-14 years  |    4850     |
| 4 |  1980  | 55-74 years |   161590    |
| 5 |  1980  |  75+ years  |    70277    |
| 6 |  1990  | 15-24 years |   276232    |
| 7 |  1990  | 25-34 years |   398435    |
| 8 |  1990  | 35-54 years |   831217    |
| 9 |  1990  | 5-14 years  |    18043    |
+---+--------+-------------+-------------+

 >>> Creative Dashboard generated: ../outputs/q_world_master_analysis/creative_suicide_dashboard.html

Process finished with exit code 0
```

Fig. 151

L. full_us_deathrate_analysis.py

```
/usr/local/bin/python3.13 /Users/shravanisawant/PycharmProjects/suicide_analysis_project/scripts/full_us_deathrate_analysis.py

***** Data Loaded and Cleaned Successfully *****

 ----------------

 Q1: Suicide Rates Flow by Age Groups Over Years
+---+------+-------------+--------------------+
|   | Year |  Age Group  |     Death Rate     |
+---+------+-------------+--------------------+
| 0 | 1950 | 10-14 years |        0.4         |
| 1 | 1950 | 15-19 years | 2.6666666666666665 |
| 2 | 1950 | 15-24 years | 4.228571428571429  |
| 3 | 1950 | 20-24 years | 6.266666666666667  |
| 4 | 1950 | 25-34 years | 9.133333333333333  |
| 5 | 1950 | 25-44 years | 10.228571428571428 |
| 6 | 1950 | 35-44 years | 14.366666666666667 |
| 7 | 1950 | 45-54 years | 20.933333333333334 |
| 8 | 1950 | 45-64 years |        19.4        |
| 9 | 1950 | 55-64 years | 26.766666666666666 |
+---+------+-------------+--------------------+

 ----------------
```

Fig. 152

```
 Q2: Suicide Rates by Sex Across Age Groups (Cleaned)
+----+-------------+--------+--------------------+
|    |  Age Group  |  Sex   |     Death Rate     |
+----+-------------+--------+--------------------+
| 0  | 10-14 years | Female |       0.8375       |
| 1  | 10-14 years |  Male  | 1.9880952380952381 |
| 2  | 15-19 years | Female | 3.4166666666666665 |
| 3  | 15-19 years |  Male  | 13.790476190476191 |
| 6  | 20-24 years | Female | 4.380952380952381  |
| 7  | 20-24 years |  Male  | 23.04285714285714  |
| 8  | 25-34 years | Female | 5.654761904761905  |
| 9  | 25-34 years |  Male  | 22.976190476190474 |
| 13 | 35-44 years |  Male  | 23.761904761904763 |
| 12 | 35-44 years | Female | 7.416666666666667  |
+----+-------------+--------+--------------------+

 ----------------
```

Fig. 153

Q3: Steepest Rise in Suicide Rate by Age Group

	Age Group	Start	End	Change
2	15-24 years	4.228571428571429	16.537837837837838	12.30926640926641
3	20-24 years	6.266666666666667	17.23333333333333	10.966666666666665
1	15-19 years	2.6666666666666665	11.299999999999999	8.633333333333333
4	25-34 years	9.133333333333333	17.466666666666665	8.333333333333332
5	25-44 years	10.228571428571428	18.2	7.9714285714285715
6	35-44 years	14.366666666666667	18.233333333333334	3.866666666666667
0	10-14 years	0.4	2.8666666666666667	2.466666666666667
14	All ages	10.299999999999999	12.717105263157896	2.4171052631578966
7	45-54 years	20.933333333333334	20.133333333333333	-0.8000000000000007
13	85 years and over	39.3	34.760000000000005	-4.539999999999992
8	45-64 years	19.4	14.005405405405407	-5.394594594594592
9	55-64 years	26.766666666666666	20.466666666666665	-6.300000000000001
11	65-74 years	30.68	18.314285714285713	-12.365714285714287
10	65 years and over	27.816666666666666	12.345161290322581	-15.471505376344085
12	75-84 years	39.85	23.45714285714286	-16.392857142857142

Fig. 154

Q4: Suicide Risk Profile: Sex × Race × Age

	Sex	Race	Age Group	Death Rate
0	Female	Sex	All ages	5.060714285714286
1	Female	Sex and age	10-14 years	0.8375
2	Female	Sex and age	15-19 years	3.4166666666666665
3	Female	Sex and age	15-24 years	3.9047619047619047
4	Female	Sex and age	20-24 years	4.380952380952381
5	Female	Sex and age	25-34 years	5.654761904761905
6	Female	Sex and age	25-44 years	6.523809523809524
7	Female	Sex and age	35-44 years	7.416666666666667
8	Female	Sex and age	45-54 years	8.647619047619047
9	Female	Sex and age	45-64 years	8.121428571428572

Fig. 155

```
---------------

Q5: Gender Suicide Gap Evolution

---------------

Q6: Age Group Suicide Vulnerabilities Over Time

---------------

Q7: Lorenz Curve: Suicide Rate Concentration by Age

---------------

Q8: Hidden At-Risk Groups Post-2010

>>> Final Interactive Dashboard saved to: ../outputs/q_us_deathrate_dashboard/creative_us_deathrate_dashboard.html

Process finished with exit code 0
```

Fig. 156