Final Report: Analysis of Childhood Obesity Trends Across the United States

Introduction

Childhood obesity remains a pressing public health concern in the United States. To better understand its trends, this report analyzes state-wise and year-wise obesity rates among children using the *Youth Risk Behavior Surveillance System* dataset from the website Data.gov. Statistical tests and visualizations were used to investigate differences across states and years, as well as highlight overall trends.

This analysis focuses on the following questions:

1. Are there statistically significant differences in obesity rates across years?
2. Are there statistically significant differences in obesity rates across states?
3. What insights can be derived from visualizing these trends using bar and pie charts?

Methods

The dataset includes state-level percentages of childhood obesity rates recorded over multiple years. Data preparation involved:

- Removing rows with missing obesity rate values (Data_Value).
- Ensuring numeric formatting for calculations and statistical tests.
- Selecting data that is refers to the question "Percent of students in grades 9-12 who have obesity"

The following analyses and visualizations were performed:

1. **Pie Chart**: Displays the proportion of childhood obesity rates by state for the for the year of 2021, chosen for its completeness of data and the most recent year complete data was available.
2. **Bar Charts**: Show childhood obesity rates across states for the four most complete years, highlighting trends over time. They are fortunately are spaced out between the years 2011 - 2021.
3. **Statistical Tests**:
    - **ANOVA** and **Tukey's HSD** tests assessed differences in obesity rates across years and states.
    - **Normality (Shapiro-Wilk)** and **Variance Homogeneity (Levene's Test)** checks were performed to validate assumptions for ANOVA.

Hypotheses

For the hypotheses of the tests, they will be assigned in blocks for tests that are as follows:

1) Shapiro-Wilk Tests on the Years:

H_0 : The data for that specific year follows a normal distribution.

H_1: The data for that specific year does not follow a normal distribution.

2) Levene's Tests on the Years:

H_0 : The variance across all years is equal.
H_1: The variance across all years is not equal.

3) ANOVA by Year

H_0 : The mean obesity rate across years is the same.
H_1: The mean obesity rate across years is the not same.

4) Tukey's by Year

H_0 : The mean obesity rate between each pair of years is the same.
H_1: The mean obesity rate between each pair of years is the not same.

5) ANOVA by State

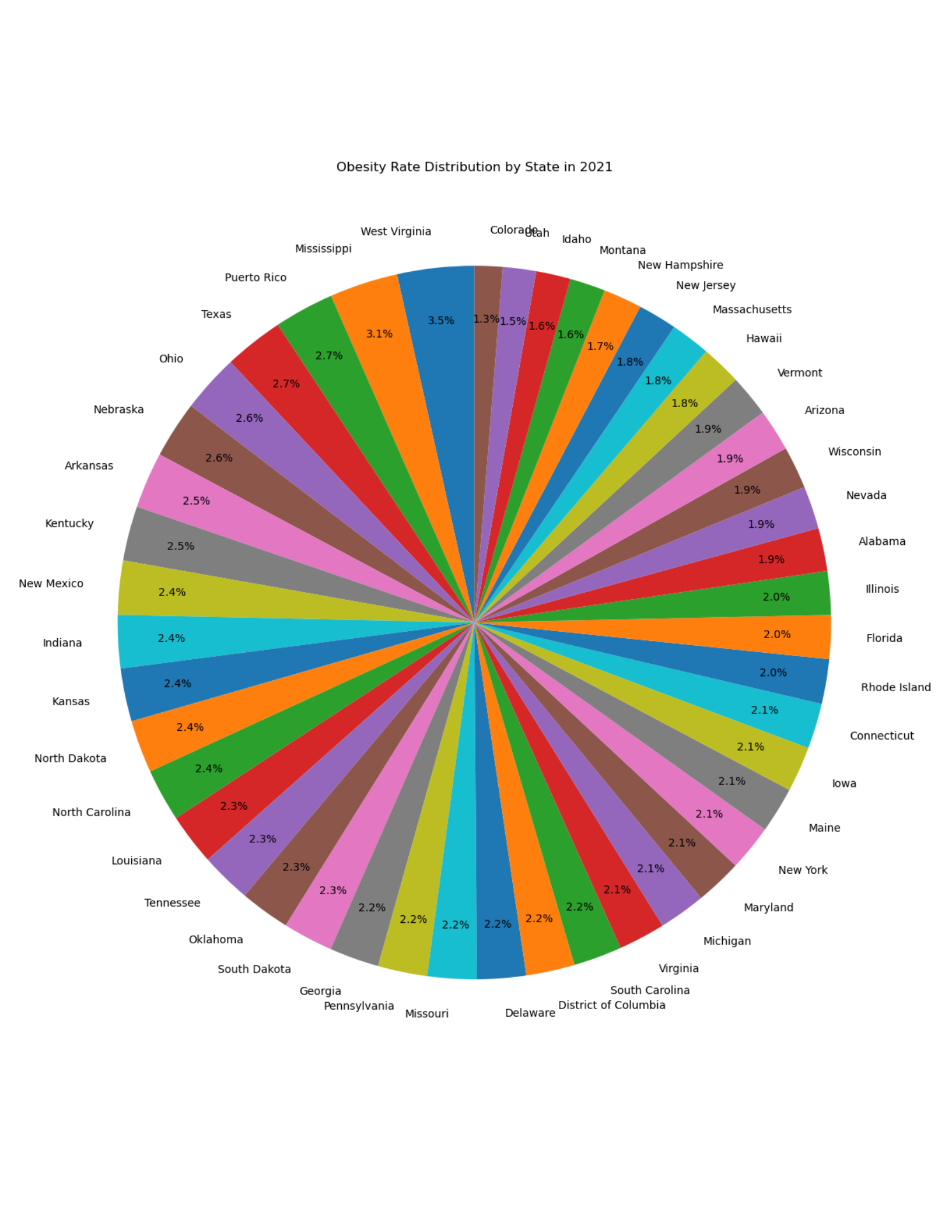
H_0 : The mean obesity rate across states is the same.
H_1: The mean obesity rate across states is the not same.

6) Tukey's by State

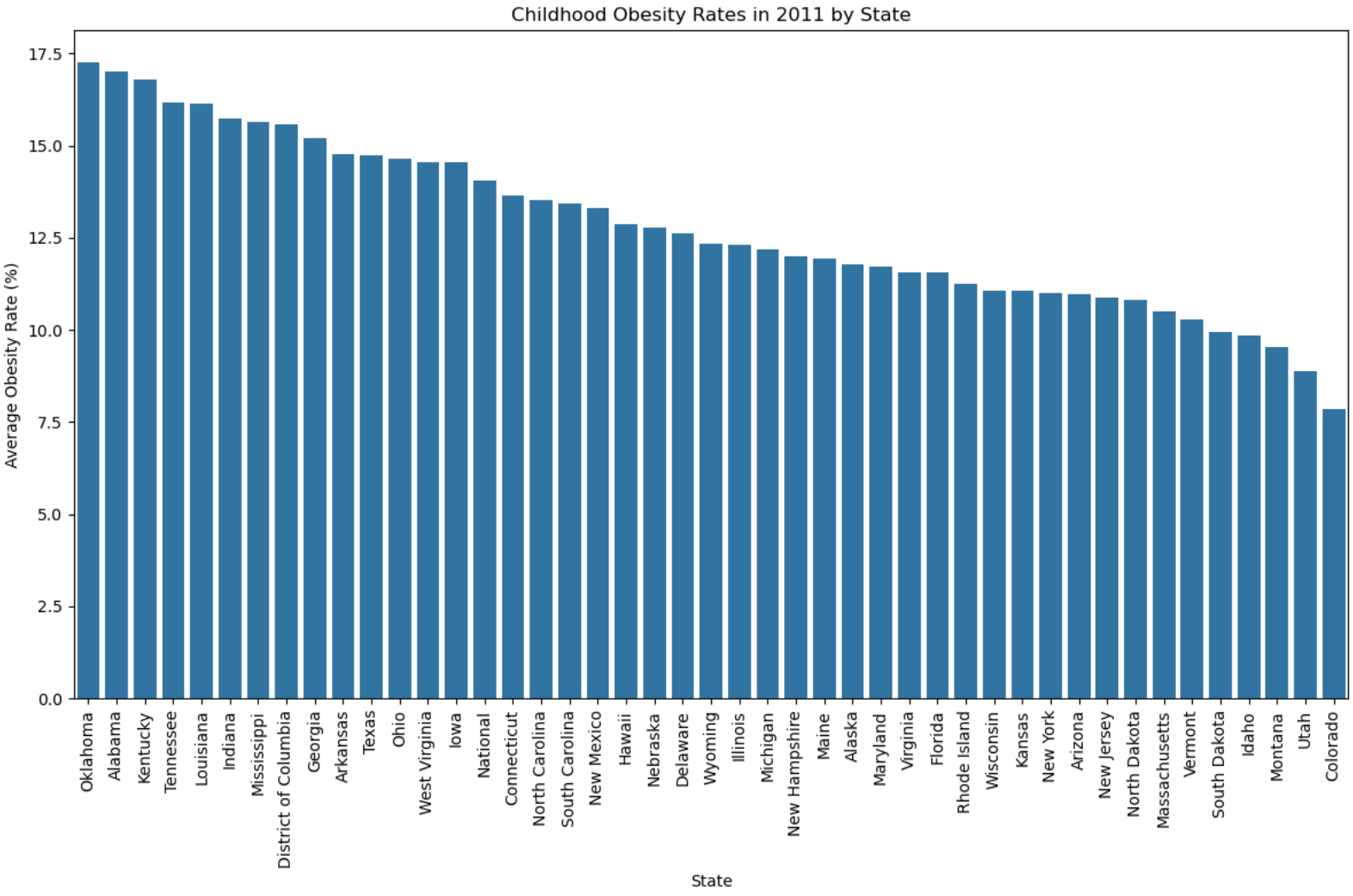
H_0 : The mean obesity rate between each pair of states is the same.
H_1: The mean obesity rate between each pair of states is the not same.

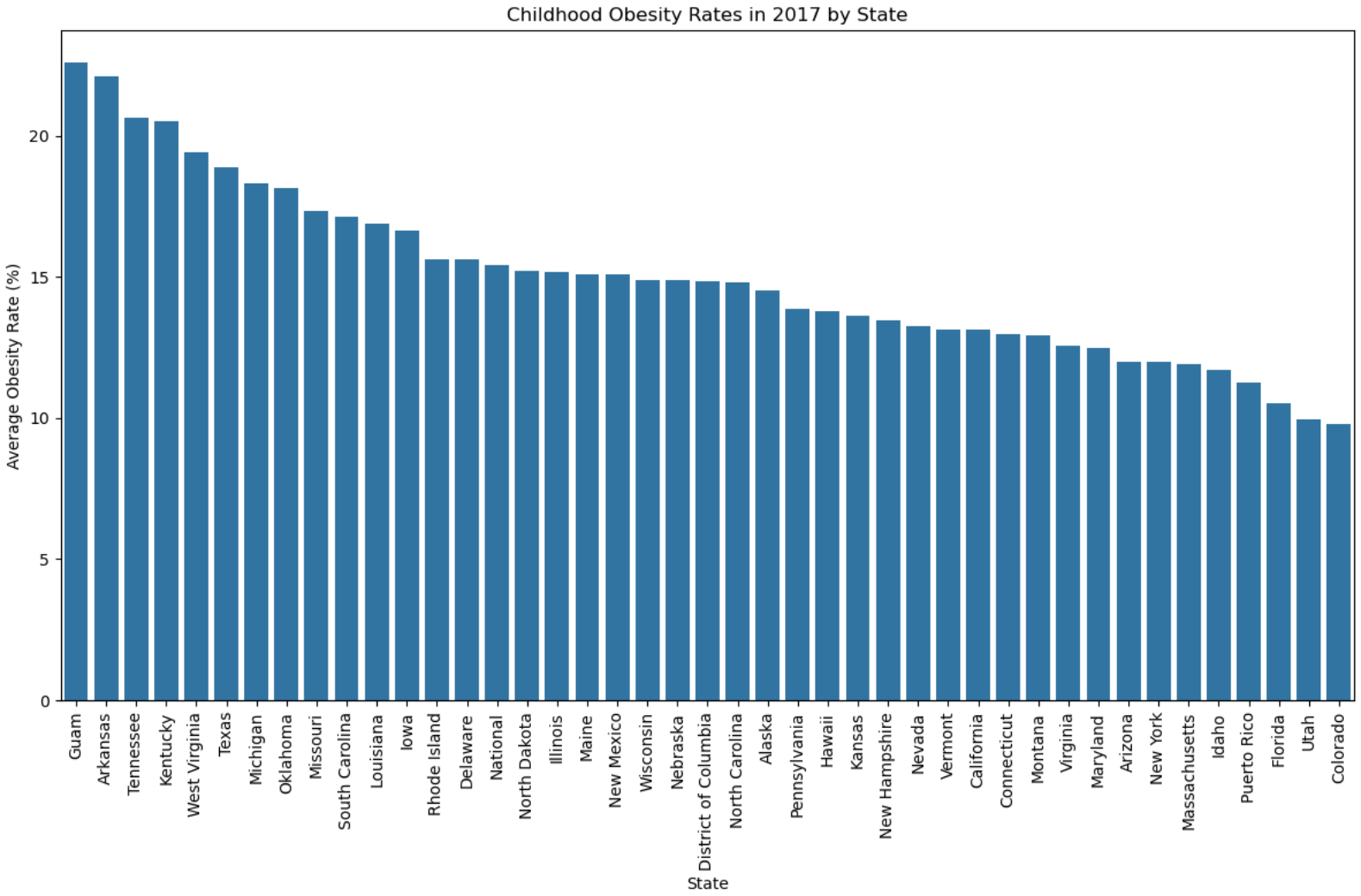
Visualizations

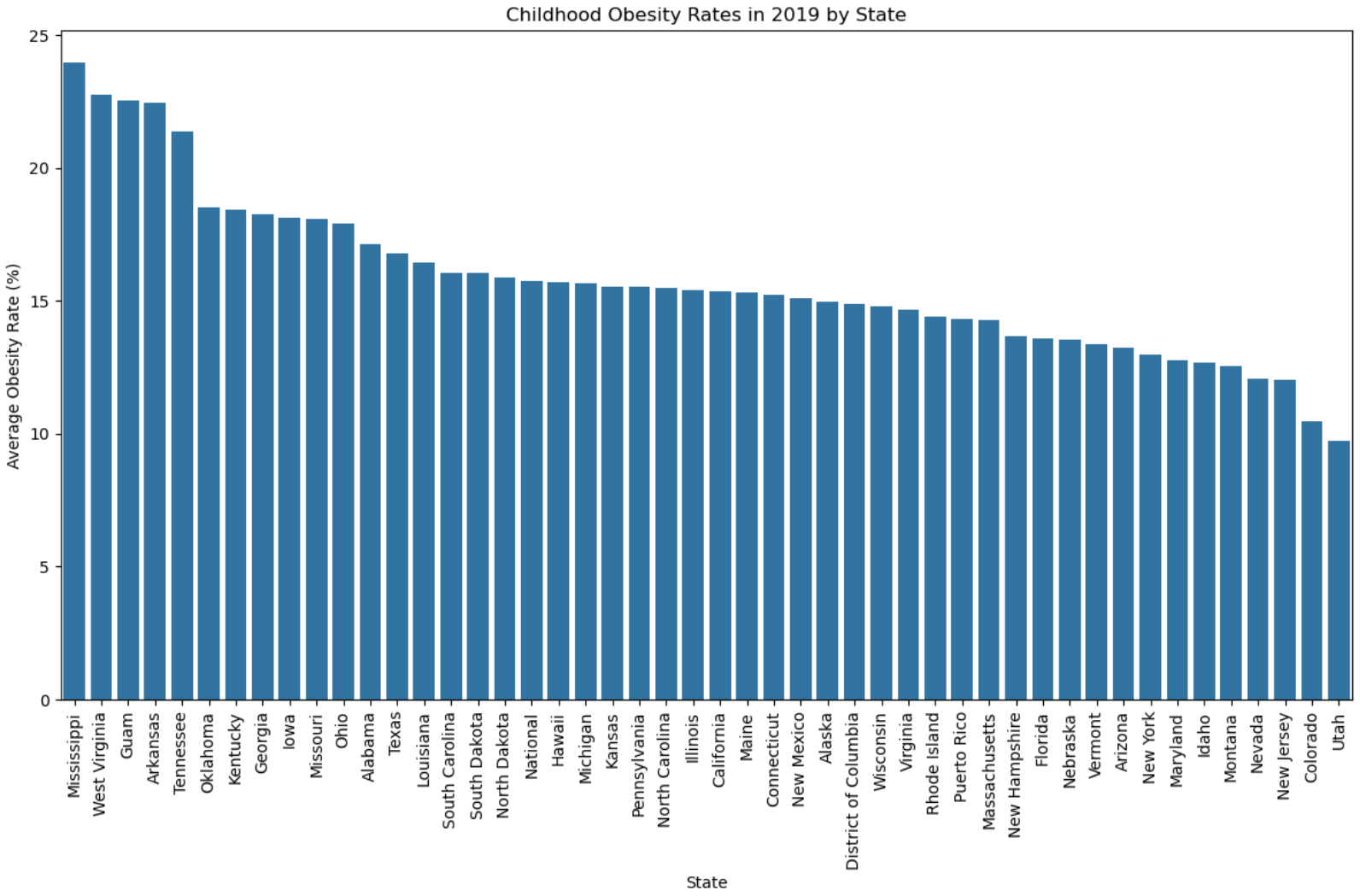
1. **Pie Chart**:
   The pie chart, based on data from the most recent complete year (**2021**), illustrates the proportion of childhood obesity rates attributed to each state. The National average was excluded to focus on state contributions to the national total. States with higher obesity rates contributed a more significant proportion, while smaller states collectively represented a minor share.

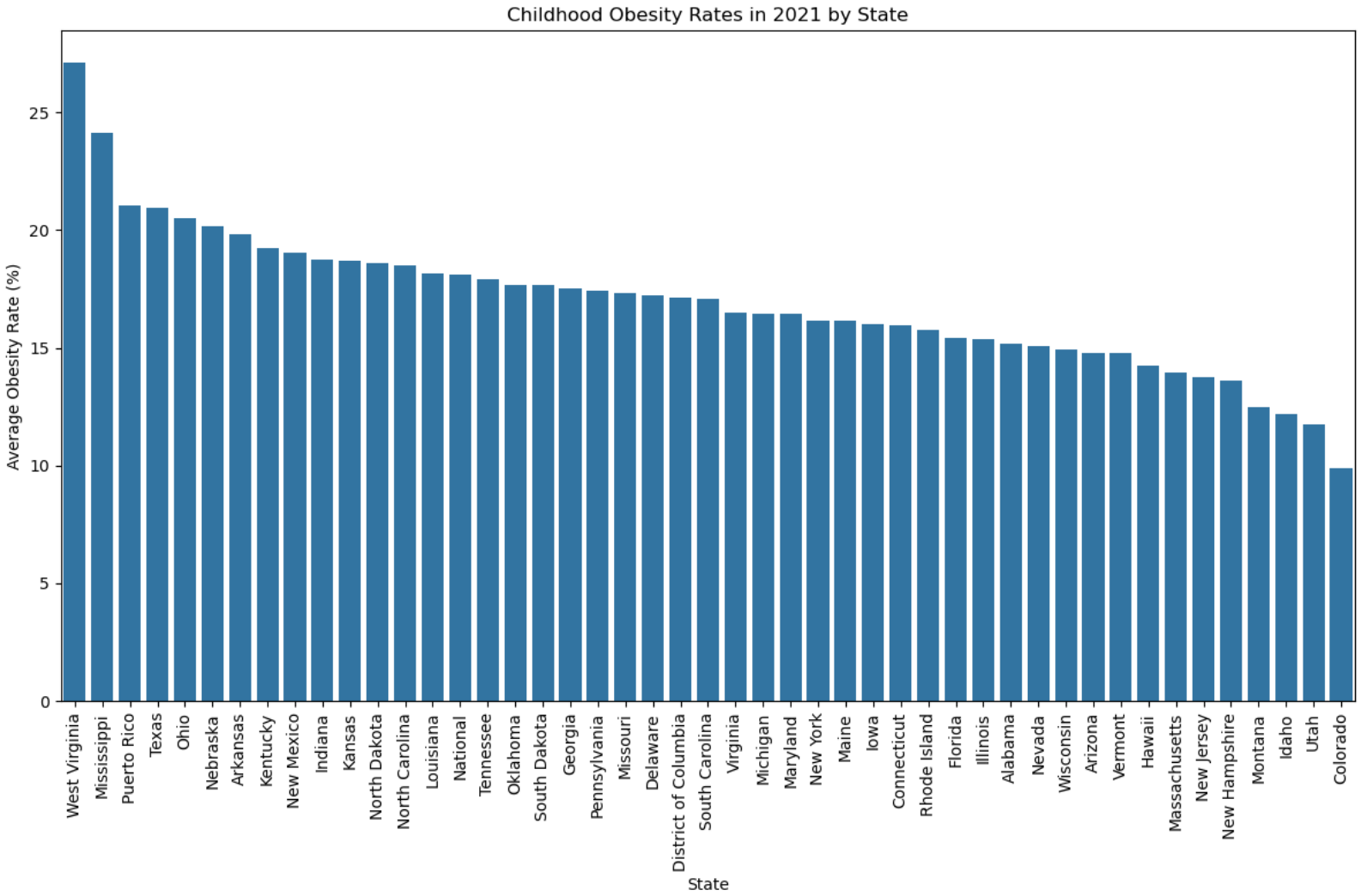

Obesity Rate Distribution by State in 2021
West Virginia
Colorado
Utah
Idaho
Montana
New Hampshire
New Jersey
Massachusetts
Hawaii
Vermont
Arizona
Wisconsin
Nevada
Alabama
Illinois
Florida
Rhode Island
Connecticut
Iowa
Maine
New York
Maryland
Michigan
Virginia
South Carolina
District of Columbia
Delaware
Missouri
Pennsylvania
Georgia
South Dakota
Oklahoma
Tennessee
Louisiana
North Carolina
North Dakota
Kansas
Indiana
New Mexico
Kentucky
Arkansas
Nebraska
Ohio
Texas
Puerto Rico
Mississippi
3.5%
1.3%
1.5%
1.6%
1.6%
1.7%
1.8%
1.8%
1.8%
1.9%
1.9%
1.9%
1.9%
1.9%
2.0%
2.0%
2.0%
2.1%
2.1%
2.1%
2.1%
2.1%
2.1%
2.2%
2.2%
2.2%
2.2%
2.2%
2.3%
2.3%
2.3%
2.3%
2.4%
2.4%
2.4%
2.4%
2.4%
2.5%
2.5%
2.6%
2.6%
2.7%
2.7%
3.1%

2. **Bar Charts**:
   Four bar charts were generated for the years with the most complete data: **2011**, **2017**, **2019**, and **2021**. These visualizations reveal fluctuations in obesity rates by state, indicating that some states consistently rank higher than others, while some rankings rise and fall.




Childhood Obesity Rates in 2017 by State
Average Obesity Rate (%)
0
5
10
15
20
State
Guam
Arkansas
Tennessee
Kentucky
West Virginia
Texas
Michigan
Oklahoma
Missouri
South Carolina
Louisiana
Iowa
Rhode Island
Delaware
National
North Dakota
Illinois
Maine
New Mexico
Wisconsin
Nebraska
District of Columbia
North Carolina
Alaska
Pennsylvania
Hawaii
Kansas
New Hampshire
Nevada
Vermont
California
Connecticut
Montana
Virginia
Maryland
Arizona
New York
Massachusetts
Idaho
Puerto Rico
Florida
Utah
Colorado


Childhood Obesity Rates in 2019 by State
Average Obesity Rate (%)
0
5
10
15
20
25
Mississippi
West Virginia
Guam
Arkansas
Tennessee
Oklahoma
Kentucky
Georgia
Iowa
Missouri
Ohio
Alabama
Texas
Louisiana
South Carolina
South Dakota
North Dakota
National
Hawaii
Michigan
Kansas
Pennsylvania
North Carolina
Illinois
California
Maine
Connecticut
New Mexico
Alaska
District of Columbia
Wisconsin
Virginia
Rhode Island
Puerto Rico
Massachusetts
New Hampshire
Florida
Nebraska
Vermont
Arizona
New York
Maryland
Idaho
Montana
Nevada
New Jersey
Colorado
Utah
State



Statistical Analyses

1. **Normality and Homogeneity of Variance**
    - **Shapiro-Wilk Test**: Data for obesity rates across years and states deviated from normality (p-values < 0.05).
    - **Levene's Test**: Variance across groups was heterogeneous (p-value < 0.05).

These results suggest caution when interpreting ANOVA, though it remains robust for moderately non-normal data with unequal variances in large datasets.

2. **ANOVA and Tukey's HSD**
    - **Across Years**:
        - ANOVA showed significant differences in obesity rates across years (**F-statistic = 92.19633443700475, p-value = 1.3487981553505422e-174**).
        - Tukey's HSD identified a majority of specific year-pairs with significant differences, indicating an upward trend in some states and a leveling off in others.

- **Across States**:
    - ANOVA confirmed statistically significant differences between states (**F-statistic = 31.36966774213451, p-value = 6.902862496880047e-256**).
    - Tukey's HSD revealed that a substantial amount, enough to reject the null hypothesis in most cases, of State-State pairs demonstrated a statistically significant difference. Since there is so much data, more analysis can be performed to find differences between the individual states as well.

Discussion

The analysis underscores significant disparities in childhood obesity rates both temporally and geographically.

1. **Temporal Trends**:
    - The bar charts highlight noticeable changes in obesity rates over the selected years. Some states experienced gradual changes, while others showed stabilization or minor changes.
    - Statistical tests confirmed these differences, indicating the need for further analysis.
    - Some states show a clear trends of being the heaviest or least-heaviest states, with these trends lasting upwards of a decade.
2. **Geographical Patterns**:
    - Both the pie chart and bar graphs emphasize the disproportionate burden of childhood obesity in certain states, particularly in the South and Midwest regions. These areas consistently recorded the highest rates over time. States and territories that are economically depressed also shared this characteristic.
3. **Policy Implications**:
    - This data highlights the importance of state-specific strategies to address childhood obesity. Investments in education, access to nutritious foods, and physical activity programs may yield significant public health benefits.
    - It is also useful to examine which states have successful child-health policies, and which ones could improve or mitigate the effect of societal and environmental effects.

Literature Review

Morrill and Chinn examined the relationship between behavioral and environmental factors influencing childhood obesity in their paper, *The Obesity Epidemic in the United States*. They identified key contributors such as the amount of time spent watching television, whether schools mandated physical education, and the availability of vending machines or fast-food restaurants. These factors, they argued, function at a societal level where state policies play a critical role in shaping outcomes (Morrill, 2004). Their findings suggest that states addressing these factors through policy interventions are more likely to see improved health outcomes, while others that neglect these issues continue to face significant challenges.

In *The Social and Emotional Lives of Overweight, Obese, and Severely Obese Children*, Harrist et al. explored the emotional impact of obesity on children. The study highlighted the role of environmental and demographic factors, noting, “Second, the sample was gathered from

rural elementary schools with a relatively large proportion of Native American students, both risk factors for obesity" (Harrist, 2016). This aligns with my statistical analysis, which indicates that rural states, such as West Virginia, Mississippi, Texas, and Ohio, rank among the highest in childhood obesity rates. West Virginia, in particular, exemplifies this trend, with its rural demographics contributing significantly to its high obesity rates.

A related global perspective is provided in an article from *The Economist*, titled "Why the War on Childhood Obesity Is Failing." The author reports that "[s]ince 1990 global rates have doubled among adults and quadrupled among children" (The Economist, 2024). This finding underscores the rapid and widespread nature of the obesity epidemic, particularly in developing countries. My analysis supports this claim, demonstrating a consistent upward trend in U.S. childhood obesity rates over the past decade. While the data primarily reflects trends in the United States, it aligns with global patterns, highlighting the broader implications of the obesity crisis.

Conclusion

This report underscores the importance of data-driven insights in addressing childhood obesity. Through visualizations and statistical analyses, it provides a clearer understanding of trends across states and years, emphasizing the need for targeted interventions. Future research should investigate behavioral and demographic factors to further identify the root causes of these disparities.

The analysis reveals statistically significant differences between states and years, indicating that childhood obesity is not only evolving over time but also shaped by state-level policies and environmental factors. Over the past decade, childhood obesity has steadily increased, as evidenced by the growing range of values in bar graph visualizations.

For example, West Virginia's progression highlights the severity of this issue: the state ranked 13th in childhood obesity in 2011, 5th in 2017, 2nd in 2019, and 1st in 2021. By 2021, West Virginia had the highest percentage of obese children in the United States at 3.5%. These trends are further supported by detailed statistical analyses and visual representations, such as those found in the *analysis_results.txt* file, which provides deeper insights into these patterns.

Although many of the Shapiro-Wilk normality tests produced p-values indicating non-normal distributions, the ANOVA tests remained appropriate due to the dataset's size. While the validity of these tests may have some limitations, the visualizations effectively highlight significant trends, providing actionable insights despite potential imperfections.

Bibliography and Appendix

Centers for Disease Control and Prevention. (n.d.). *Nutrition, Physical Activity, and Obesity - Youth Risk Behavior Surveillance System*. Data.gov. Retrieved from https://catalog.data.gov/dataset/nutrition-physical-activity-and-obesity-youth-risk-behavior-surveillance-system

Harrist, A. W., Swindle, T. M., Hubbs-Tait, L., Topham, G. L., Shriver, L. H., & Page, M. C. (2016). The social and emotional lives of overweight, obese, and severely obese children. *Child Development, 87*(5), 1564–1580. https://www.jstor.org/stable/24809555

Morrill, A. C., & Chinn, C. D. (2004). The obesity epidemic in the United States. *Journal of Public Health Policy, 25*(3/4), 353–366. https://doi.org/10.1057/palgrave.jphp.3190030

The Economist. (2024, August 8). Why the war on childhood obesity is failing. *The Economist*.