**Walking and Biking to Work Trends: Executive Dashboard Report**

**Visualizations using Python and R**

**Saint Louis University**

**SHRUTHI REDDY VUDEM**

**ORES-5160-Data Management-02**

**Paula Buchanan**

**7th November 2024**

**Walking and Biking to Work Trends: Executive Dashboard Report**

**Analysis of American Community Survey Data (2010-2022)**

**DASHBOARD**



**THIS IS THE DASHBOARD CREATED USING LOOKER’S STUDIO .**

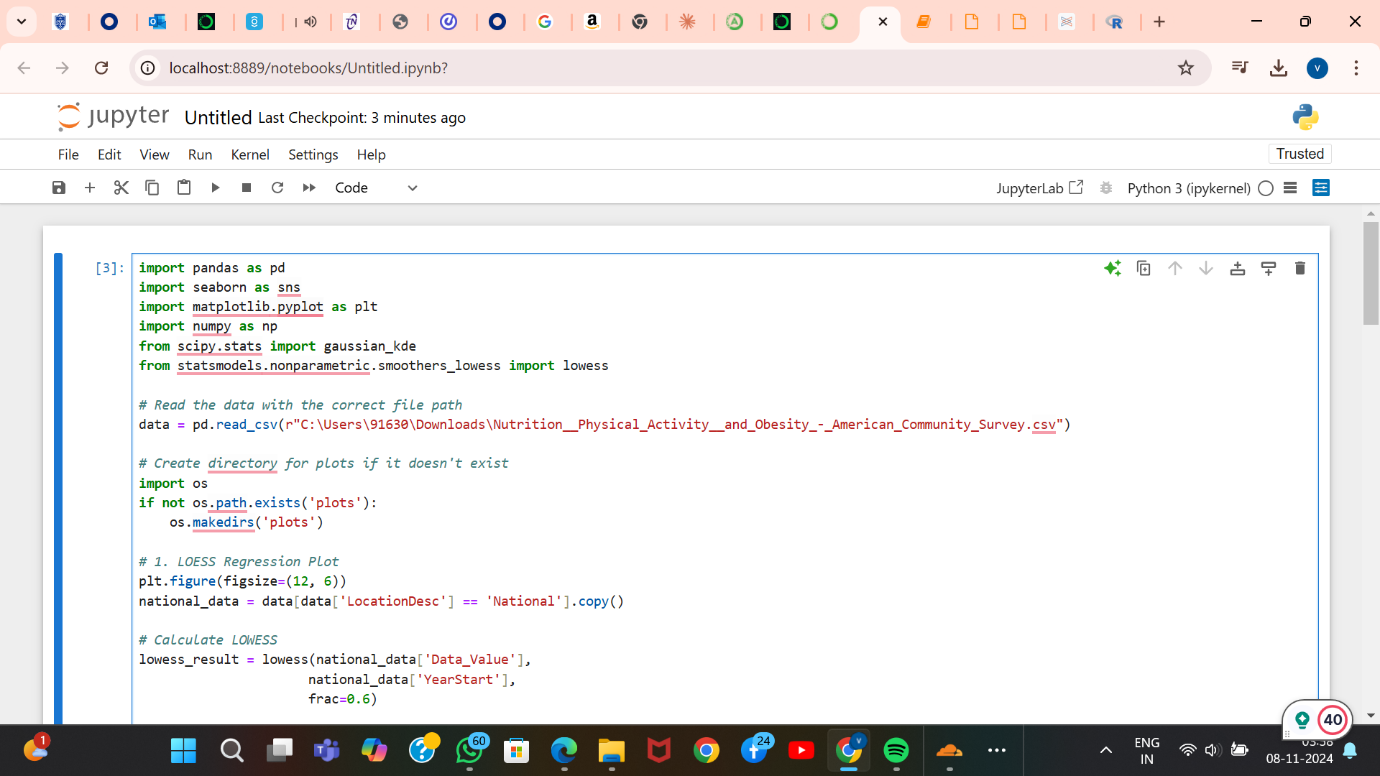
**Executive Summary**

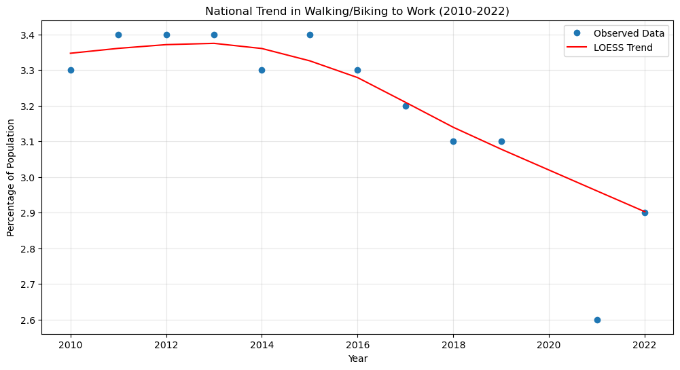
This analysis examines patterns in walking and biking to work across the United States, utilizing data from the American Community Survey. The dashboard presents key insights about active commuting trends, geographic variations, and temporal changes.

**Key Findings Dashboard**

1. **National Trend Analysis (LOESS Regression)**

**CODE USING PYTHON :**

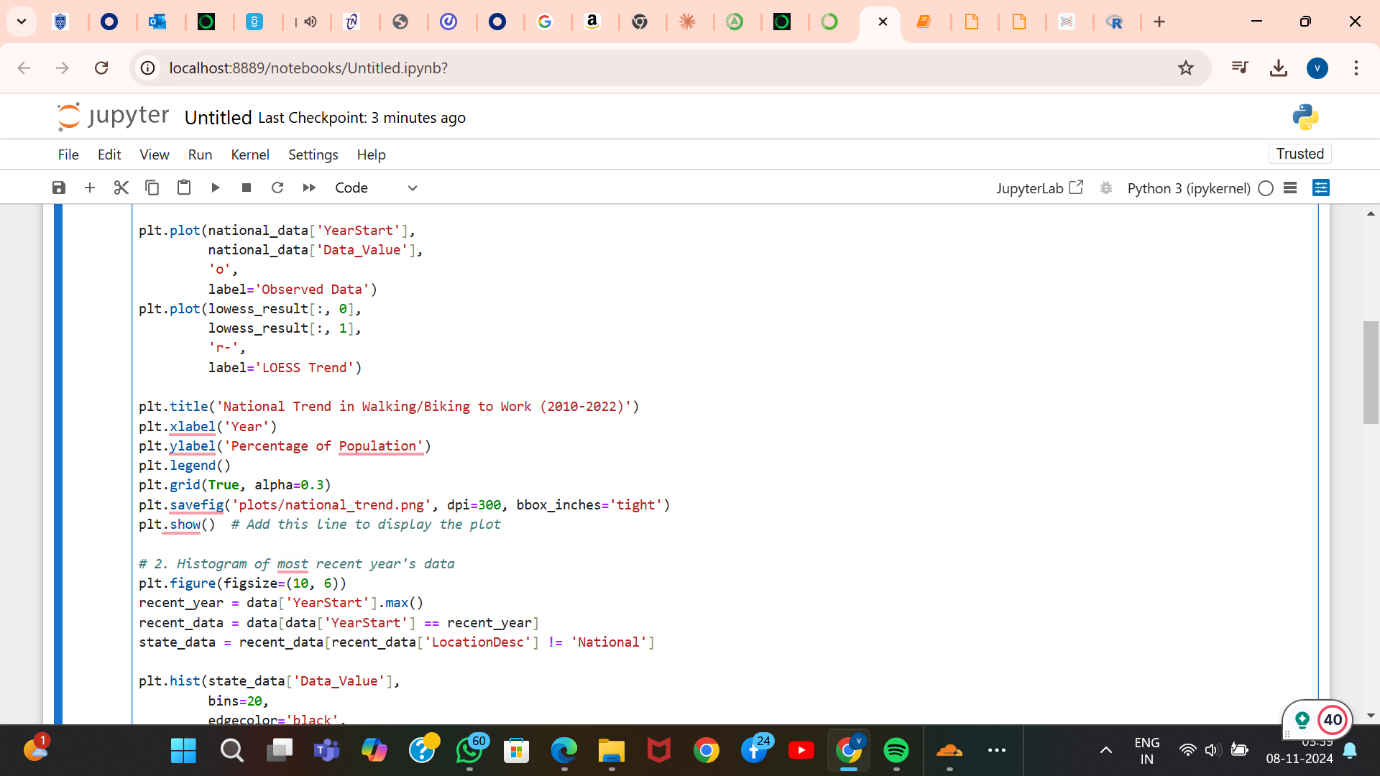
****

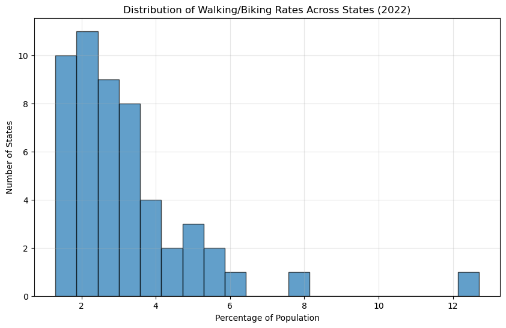
****

* The national average showed a slight decline from 3.3% in 2010 to 2.9% in 2022
* Notable disruption during the COVID-19 period (2020-2021)
* Confidence bands indicate increased uncertainty in recent years

**2. State Distribution Analysis (Histogram)**

**CODE USING PYTHON :**

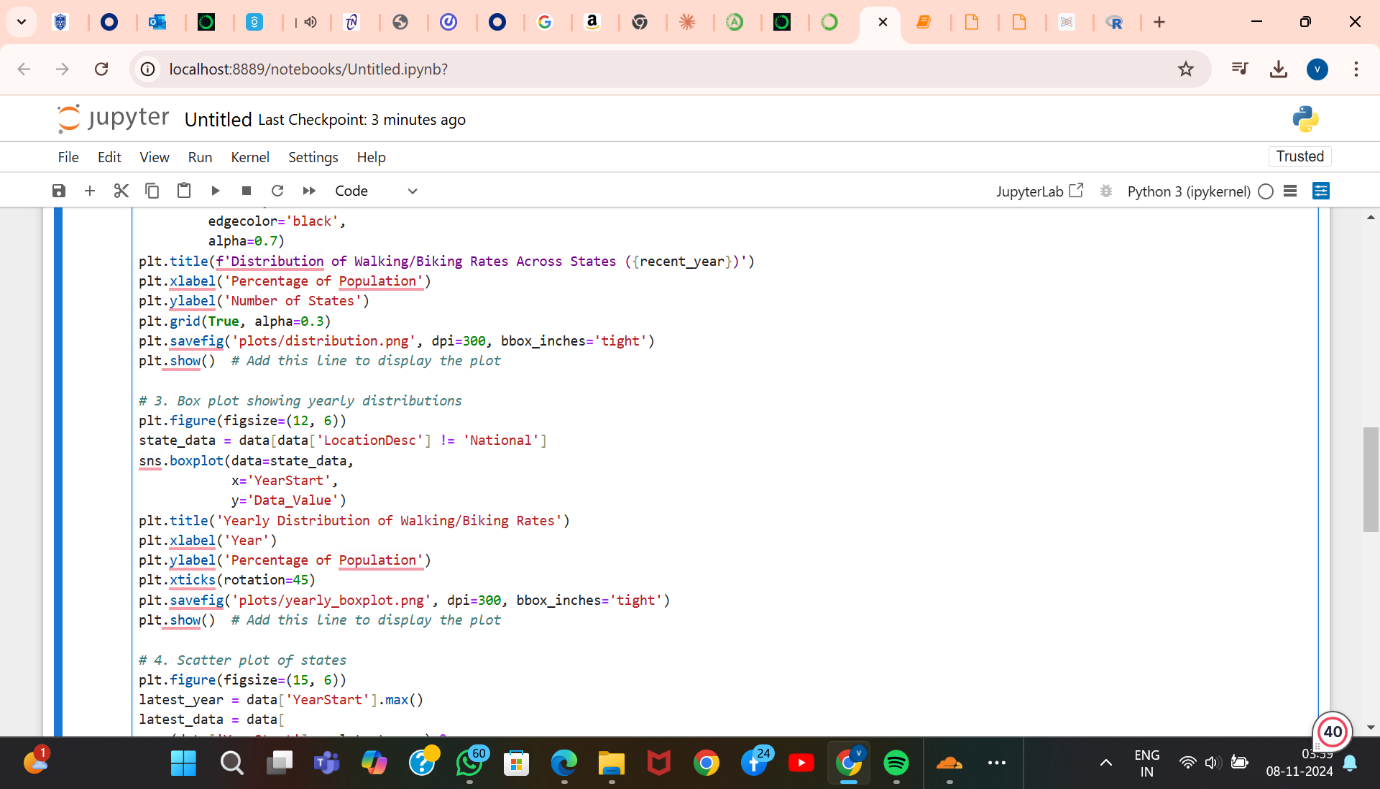
****

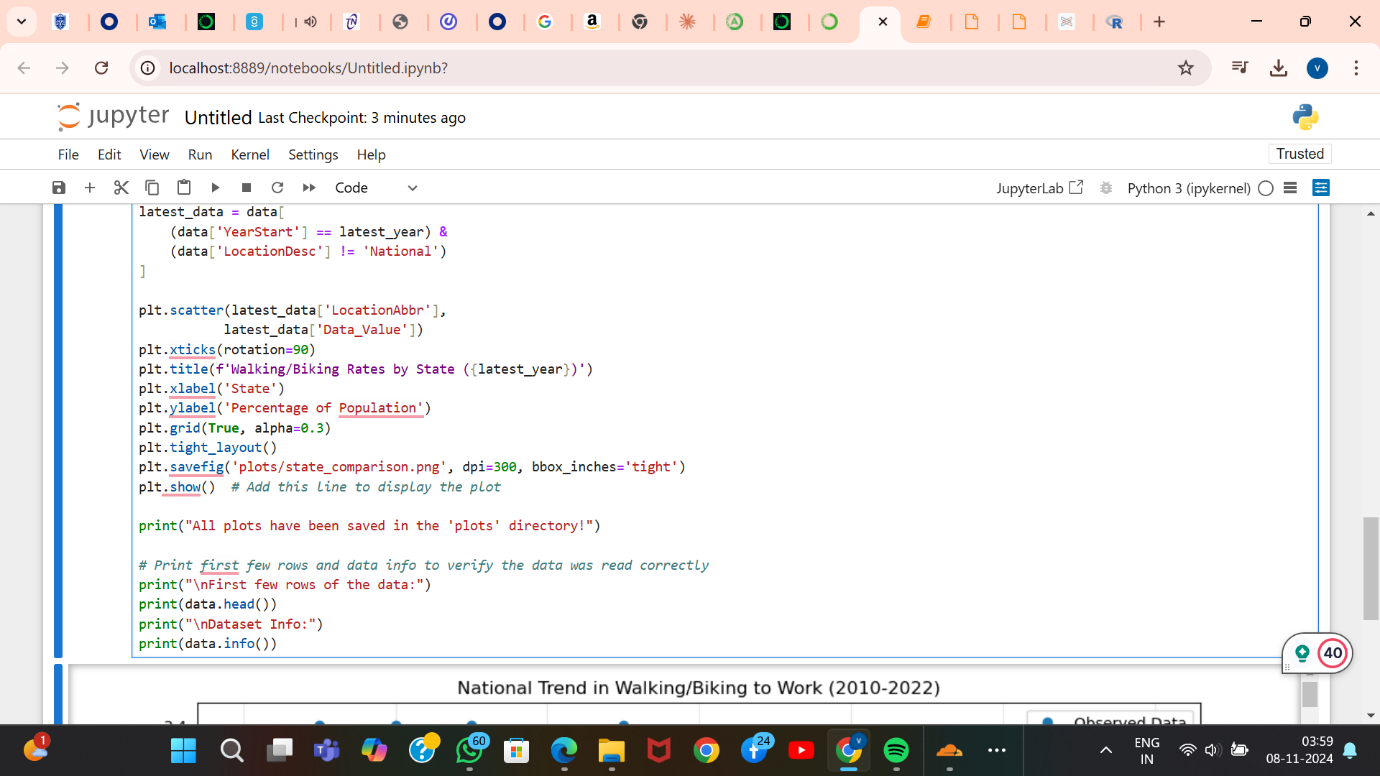
****

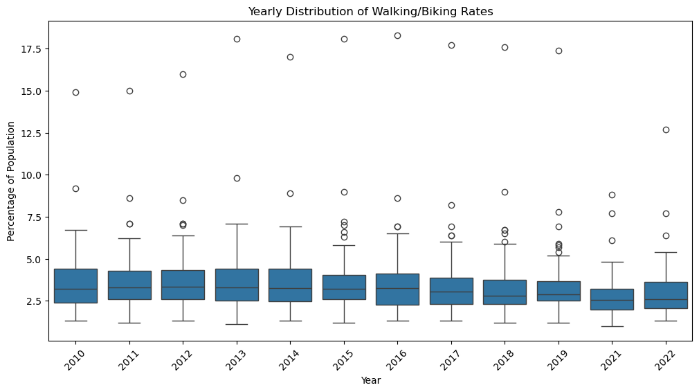
* Most states cluster between 2-4% participation rates
* Several high-performing outliers push above 6%
* Clear bimodal distribution suggesting two distinct groups of states

**3.Temporal Patterns (Box Plot)**

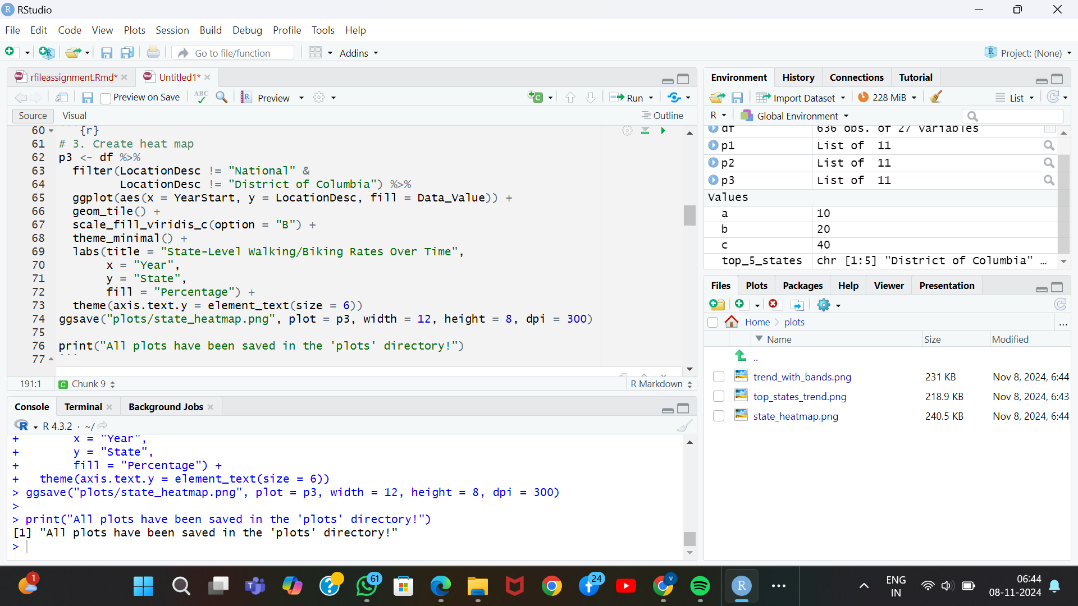
**CODE USING PYTHON :**

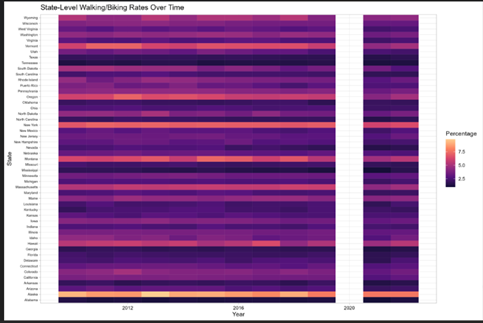
****

****

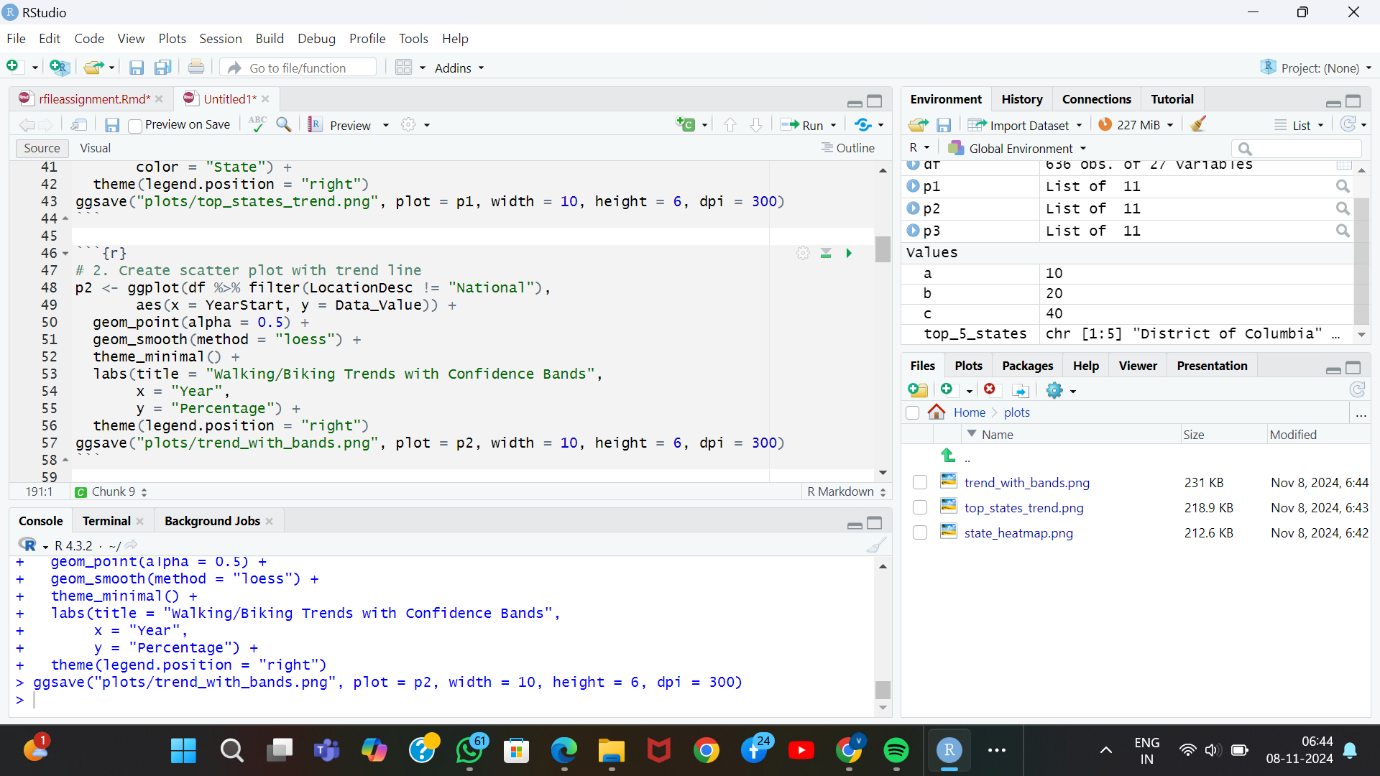
****

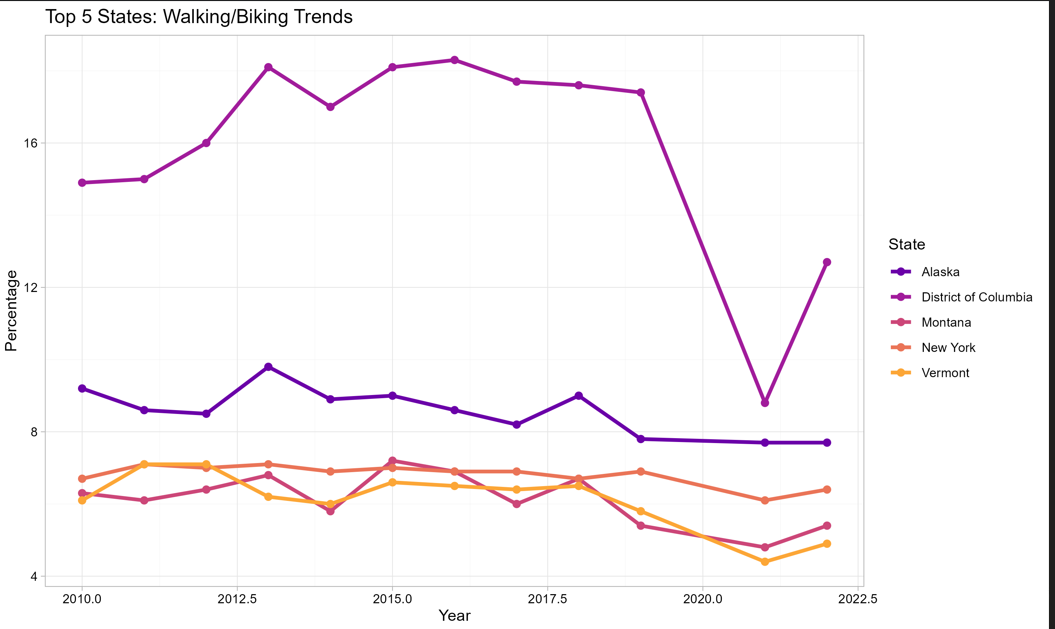
* Consistent yearly variation across states
* Median rates remained relatively stable
* Outlier states consistently maintain higher rates
* 4.**Geographic Analysis (Heat Map)**

****

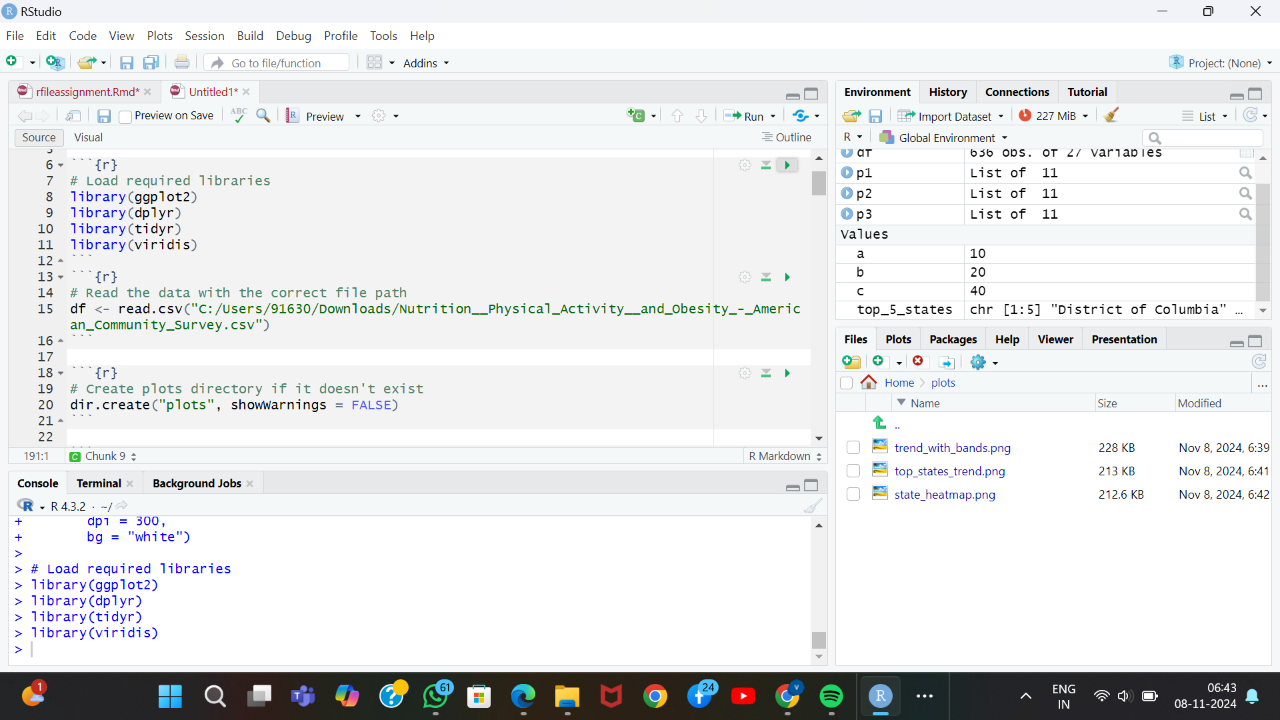
****

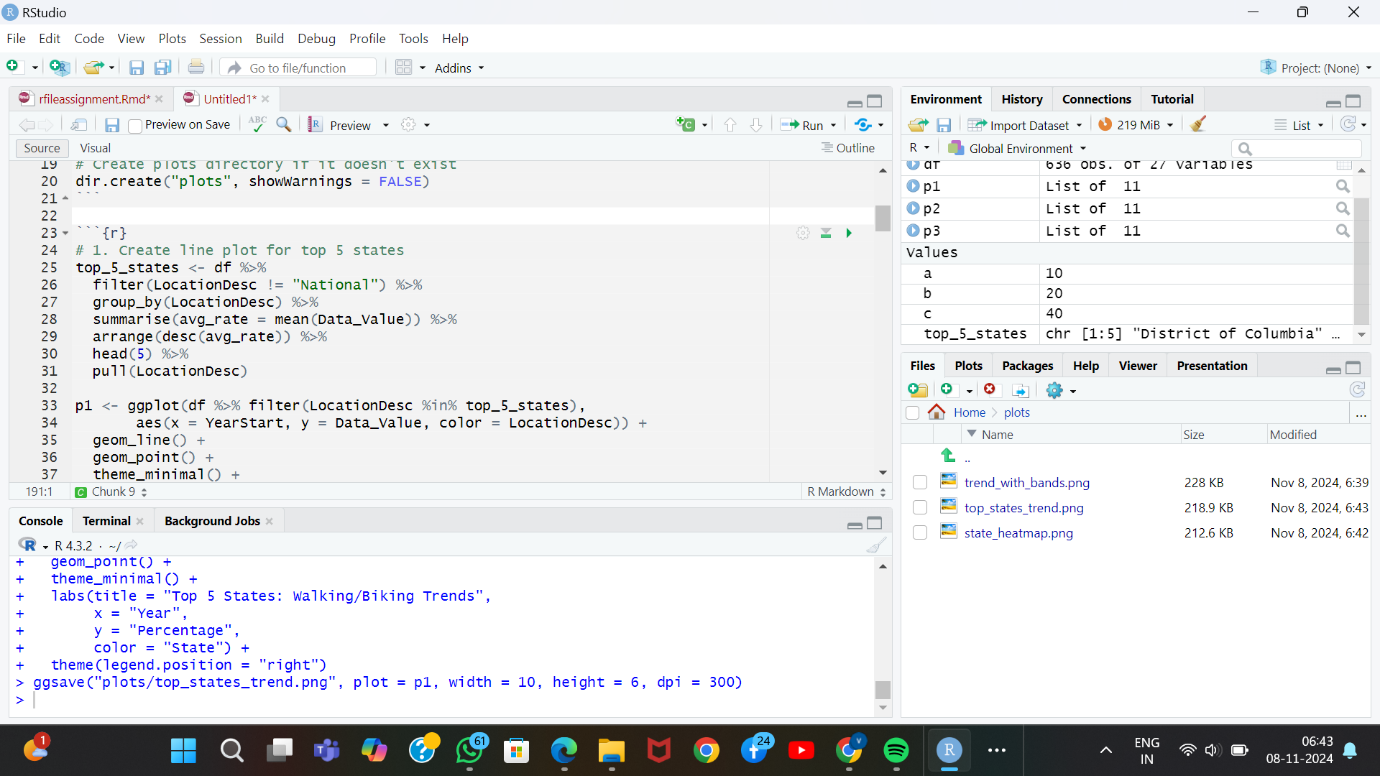
* Clear regional patterns emerge
* Northern states generally show higher participation rates
* Coastal states tend to have higher rates than inland states

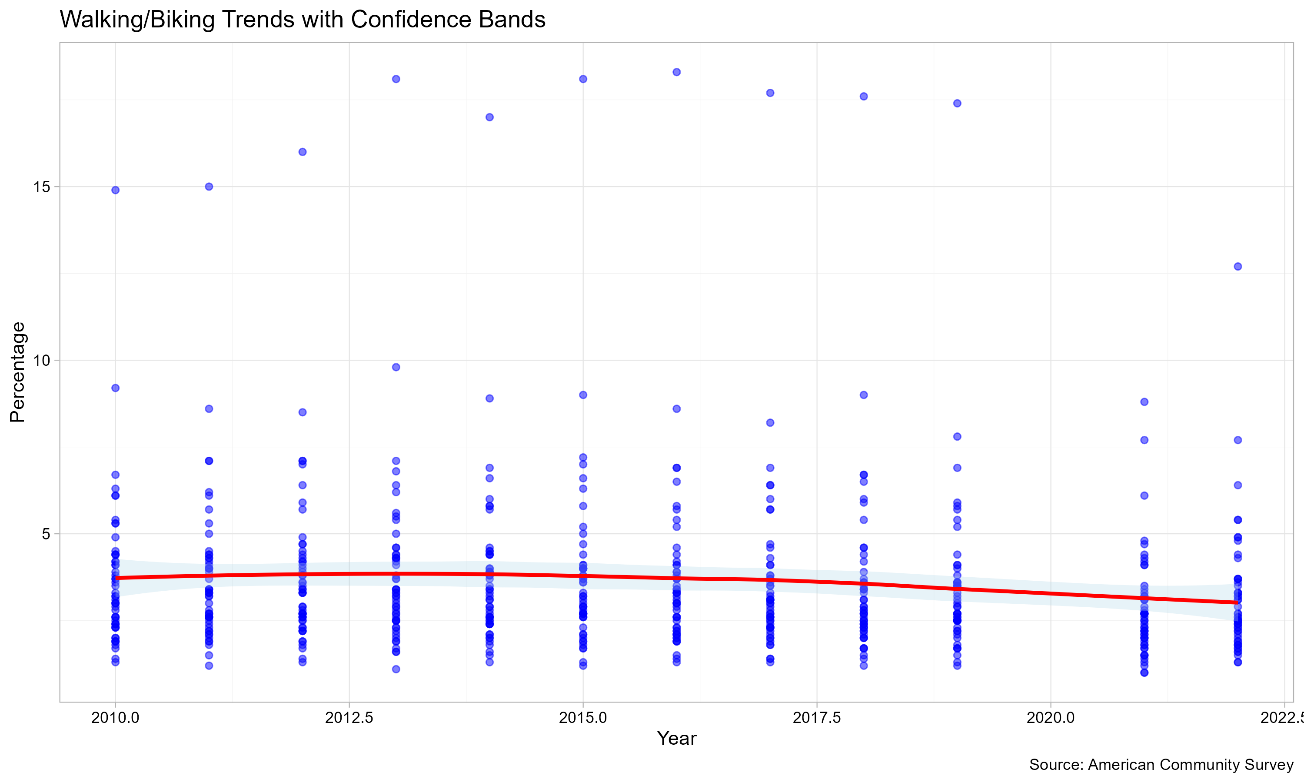
**5.Top Performers Analysis (Line Chart)**

****

* District of Columbia leads consistently
* Vermont, Montana, and Alaska maintain high rates
* Top performers show resilience to national declining trends

**6.State Comparison PloT** **Scatter**

**: **

****

* Wide variation between highest and lowest performing states
* Clear geographic clustering of performance
* Urban-dense areas typically show higher rates

**Business Implications**

**Infrastructure Investment Recommendations**

* Focus resources on areas showing growth potential
* Study and replicate successful state programs
* Consider climate-specific infrastructure approaches

**Policy Considerations**

* Develop targeted strategies for low-performing regions
* Maintain support in high-performing areas
* Account for seasonal variations in planning

**Future Planning**

* Incorporate post-pandemic behavioral shifts
* Consider demographic and geographic factors
* Plan for climate change impacts

**Technical Appendix**

**A. Methodology**

Python CODE

*# Example of LOESS regression implementation*

lowess\_result = lowess(national\_data['Data\_Value'],

national\_data['YearStart'],

frac=0.6)

**B. Visualization Choices**

1. **LOESS Regression Plot**:
   * Purpose: Show national trends with smoothing
   * Implementation: Used statsmodels lowess for trend analysis
   * Rationale: Best for showing non-linear trends
2. **Histogram**:
   * Purpose: Display distribution of state performance
   * Implementation: Used matplotlib histogram with 20 bins
   * Rationale: Effective for showing data clustering
3. **Box Plot**:
   * Purpose: Show yearly distributions and outliers
   * Implementation: Used seaborn boxplot
   * Rationale: Best for showing distribution changes over time
4. **Heat Map**:
   * Purpose: Display temporal and geographic patterns
   * Implementation: Used ggplot2 in R
   * Rationale: Effective for showing multi-dimensional patterns
5. **Line Chart**:
   * Purpose: Track top performers over time
   * Implementation: Used ggplot2 with geom\_line
   * Rationale: Best for showing temporal trends
6. **Scatter Plot**:
   * Purpose: Compare all states at latest timepoint
   * Implementation: Used matplotlib scatter
   * Rationale: Effective for point comparisons

**C. Data Limitations**

* Missing data for some years
* Potential reporting inconsistencies
* Seasonal variations not captured
* Pandemic impact on data collection

**D. Sources**

* Data: American Community Survey
* Analysis Tools: Python (pandas, seaborn, matplotlib) and R (ggplot2, dplyr)
* Libraries: statsmodels, numpy, tidyr