Healthcare cost analysis across the u.s.

TEAM: SILVIO ZABALA, LYNN SOORS, JUSTIN KIM, CONNOR CASEY, CHANDARA IN, DANIEL SIMONSON

Project description

- Analyze and compare healthcare rates across different states in the U.S. and the types of insurance coverage
- Examine the relationship between types of insurance coverage and healthcare expenditure
- Uncover patterns and insights that may inform healthcare policy and economic planning

Our hypotheses

- Higher insurance coverage rates correlate with reduced total healthcare expenditure per state, as insured individuals are more likely to engage in preventive care, reducing long-term healthcare costs.
- Public insurance coverage rates correlate with higher total healthcare expenditure per state, as insured individuals are more likely to engage in preventive care, reducing long-term healthcare costs.
- States with a higher proportion of private insurance coverage may experience
  higher healthcare expenditures due to potentially higher costs associated with
  private healthcare services.
- Healthcare expenditures have risen consistently over the years.
- Health insurance coverage rates have shown an upward trend over time.

Data Cleaning and Preparation

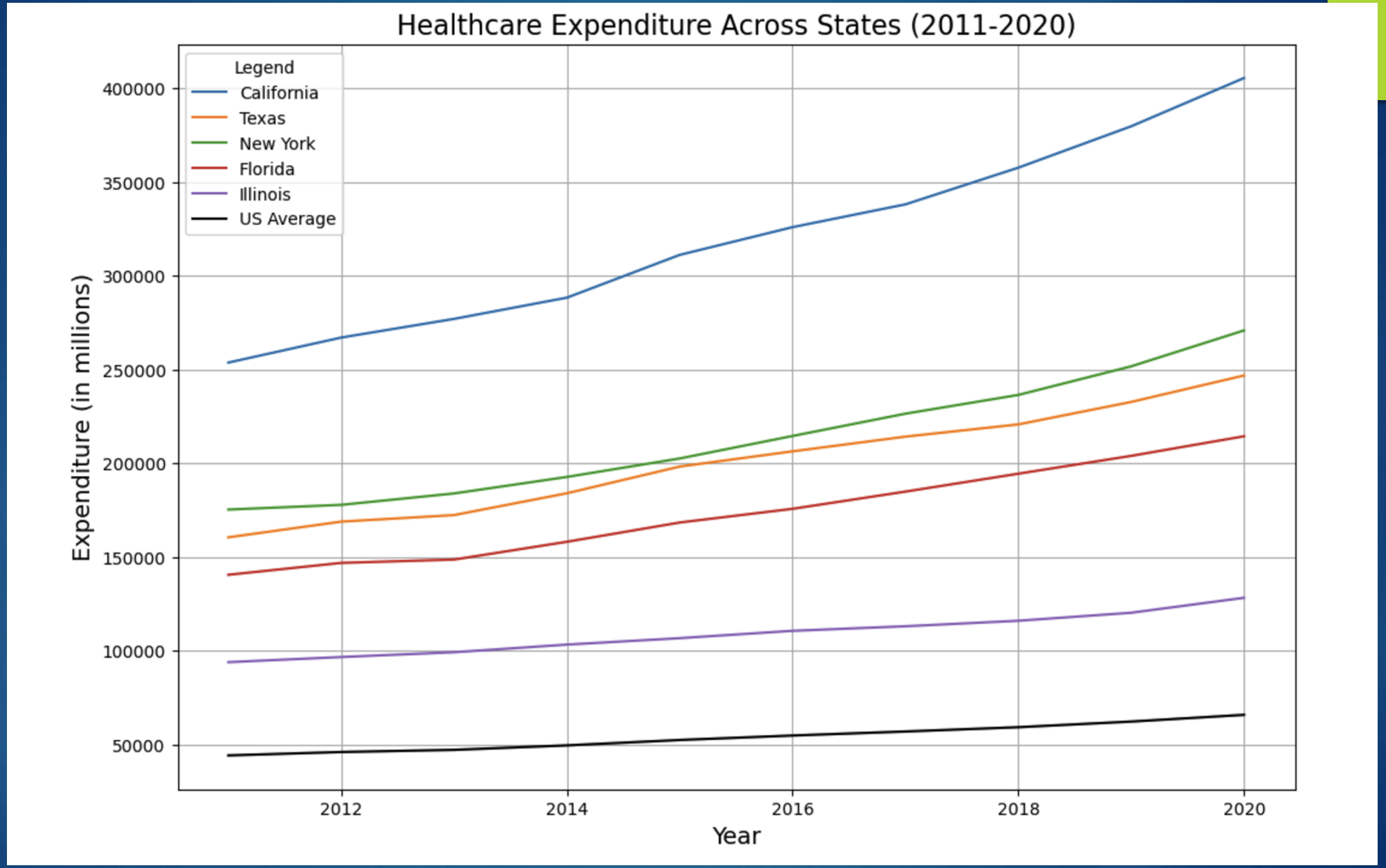
- **Removed Irrelevant or Problematic Rows**:
  - Rows with metrics unrelated to the analysis (e.g., "Percent," "MOError1," "MOError2") were excluded to focus on relevant coverage and expenditure data.
  - Data from the year 2020 was removed due to incomplete or unreliable values.
- **Data Transformation**:
  - Strings in the Value column were cleaned by removing commas (,) to prepare for numerical conversion.
  - Converted the Value column from string to float for numerical analysis.
- **Filtered Relevant Coverage Types**:
  - Focused on specific coverage types, such as "Private" and "Public," based on research questions.
  - Pivoted the table to organize data with Coverage as columns and Year as rows for easier analysis of trends.
- **Handling Missing Data**:
- Missing values in pivot tables were filled with zeros (.fillna(0)), ensuring that gaps did not disrupt visualizations or analysis.

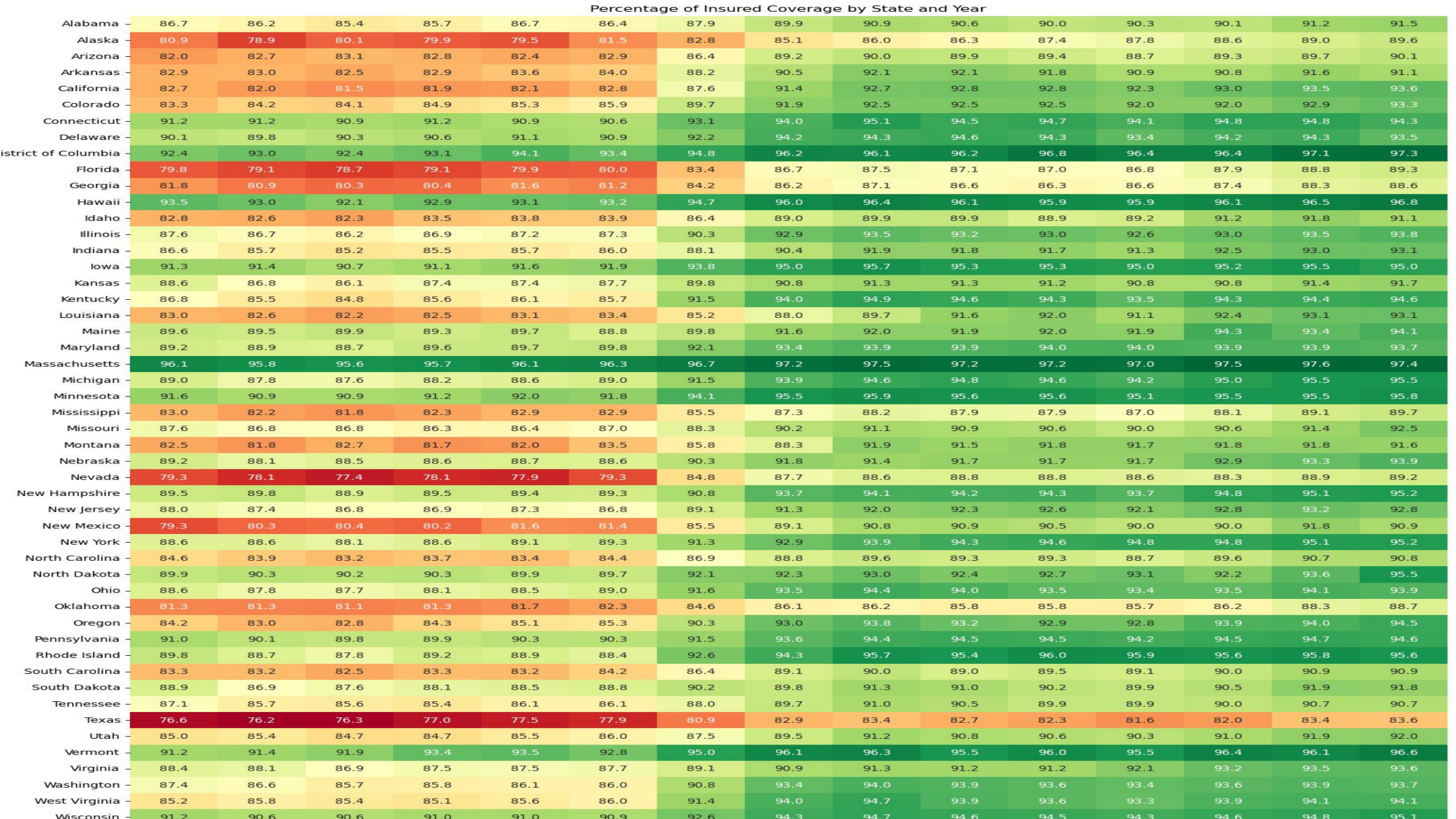
Data Cleaning and Preparation

```
#Load Data
df = insurance_data
df = df[df['Year'] != 2020]
df = df[df['Metric'] != 'Percent']
df = df[df['Metric'] != 'MOError1']
df = df[df['Metric'] != 'MOError2']

# Filter data to get rows where `Coverage` is "Any coverage" or "Total"
coverage_data = df[df['Coverage'].isin(['Any coverage', 'Uninsured'])]

coverage_data['Value'] =coverage_data['Value'].str.replace(',', '')
coverage_data['Value'] =coverage_data['Value'].astype(float)

# Pivot the table to have "Any coverage" and "Total" as columns for each state and year
coverage_pivot = coverage_data.pivot_table(
    index=['Nation', 'Year'],
    columns='Coverage',
    values='Value'
).reset_index()

# Calculate the percentage of insured people (Any coverage / Total * 100)
coverage_pivot['Percent Insured'] = (coverage_pivot['Any coverage'] / (coverage_pivot['Uninsured']+coverage_pivot['Any coverage'] ) )* 100

# Display the results
#print(coverage_pivot[['Nation', 'Year', 'Percent Insured']].head())

# Optionally, save the results to a new CSV file
coverage_pivot[['Nation', 'Year', 'Percent Insured']].to_csv('percent_insured_by_state_year.csv', index=False)


# Correcting the pivot statement
heatmap_data = coverage_pivot.pivot(index='Nation', columns='Year', values='Percent Insured')

# Plot the heatmap
plt.figure(figsize=(18, 20))
sns.heatmap(heatmap_data, annot=True, fmt=".1f", cmap="RdYlGn", cbar_kws={'label': 'Percent Insured'}, cbar=False)
plt.title('Percentage of Insured Coverage by State and Year')
plt.xlabel('Year')
plt.ylabel('')
plt.xticks(rotation=0)
plt.yticks(rotation=0)
```

Data Cleaning and Preparation

[44]:
```
# Filter only relevant coverage types
#insurance_data = insurance_data[insurance_data['Coverage'].isin(['Private', 'Public', 'Other'])]

# Plotting the line chart for "Private" insurance coverage over the years
plt.figure(figsize=(10, 6))
sns.lineplot(data=coverage_pivot, x='Year', y='Private', marker='o')

# Add labels and title
plt.title('Private Insurance Coverage Over the Years')
plt.xlabel('Year')
plt.ylabel('Coverage Rates')

# Format y-axis labels to show 0 decimal places
plt.gca().yaxis.set_major_formatter(plt.FuncFormatter(lambda x, _: f'{x:.0f}'))

# Remove grid lines
plt.grid(False)

# Add data labels for each point
for x, y in zip(coverage_pivot['Year'], coverage_pivot['Private']):
    plt.text(x, y, f'{y:.0f}', ha='center', va='bottom')

plt.show()
```


Healthcare Expenditure Across States (2011-2020)
Legend
California
Texas
New York
Florida
Illinois
US Average
Expenditure (in millions)
400000
350000
300000
250000
200000
150000
100000
50000
2012
2014
2016
2018
2020
Year

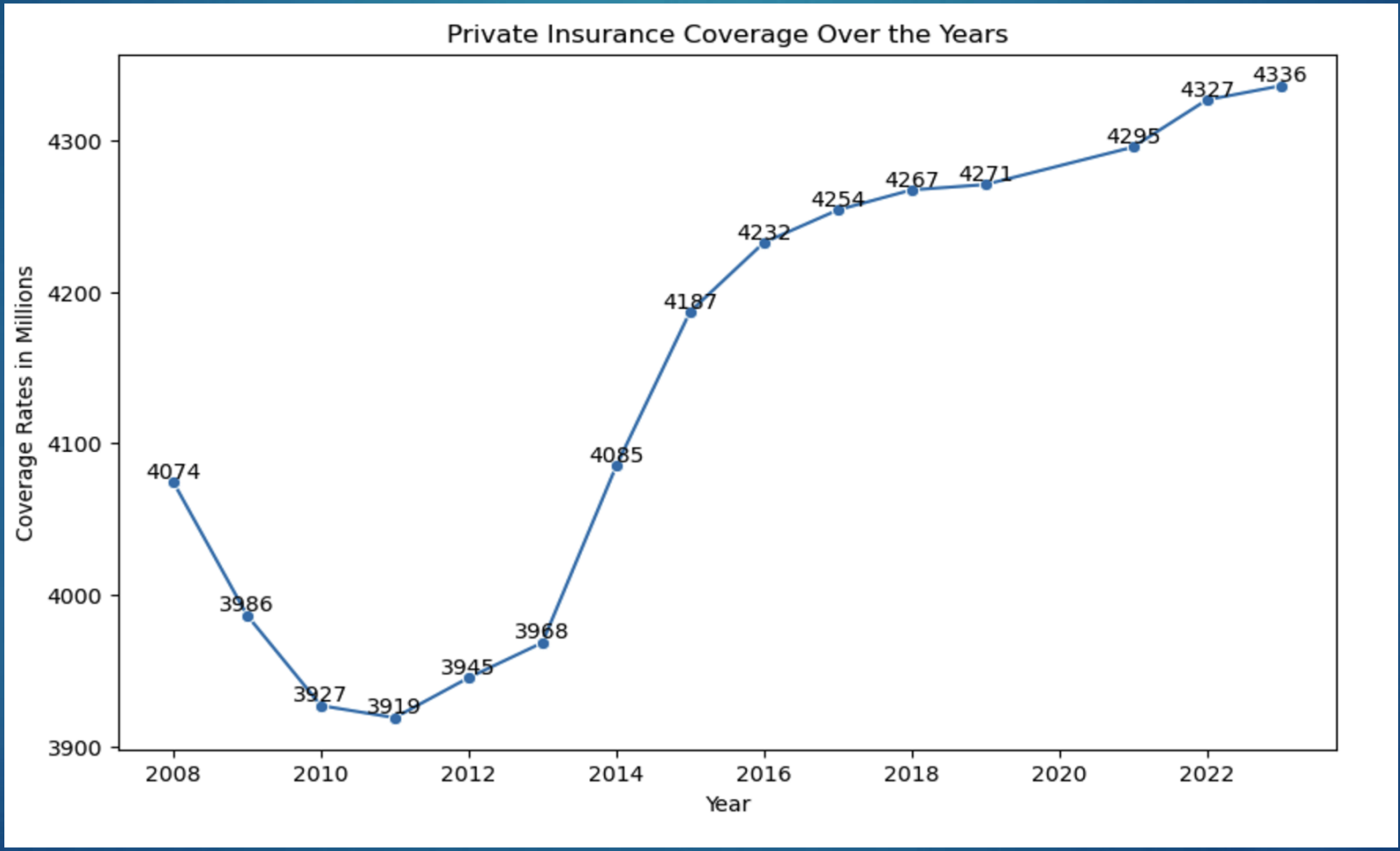


Alabama	86.7	86.2	85.4	85.7	86.7	86.4	87.9	89.9	90.9	90.6	90.0	90.3	90.1	91.2	91.5
Alaska	80.9	78.9	80.1	79.9	79.5	81.5	82.8	85.1	86.0	86.3	87.4	87.8	88.6	89.0	89.6
Arizona	82.0	82.7	83.1	82.8	82.4	82.9	86.4	89.2	90.0	89.9	89.4	88.7	89.3	89.7	90.1
Arkansas	82.9	83.0	82.5	82.9	83.6	84.0	88.2	90.5	92.1	92.1	91.8	90.9	90.8	91.6	91.1
California	82.7	82.0	81.5	81.9	82.1	82.8	87.6	91.4	92.7	92.8	92.8	92.3	93.0	93.5	93.6
Colorado	83.3	84.2	84.1	84.9	85.3	85.9	89.7	91.9	92.5	92.5	92.5	92.0	92.0	92.9	93.3
Connecticut	91.2	91.2	90.9	91.2	90.9	90.6	93.1	94.0	95.1	94.5	94.7	94.1	94.8	94.8	94.3
Delaware	90.1	89.8	90.3	90.6	91.1	90.9	92.2	94.2	94.3	94.6	94.3	93.4	94.2	94.3	93.5
istrict of Columbia	92.4	93.0	92.4	93.1	94.1	93.4	94.8	96.2	96.1	96.2	96.8	96.4	96.4	97.1	97.3
Florida	79.8	79.1	78.7	79.1	79.9	80.0	83.4	86.7	87.5	87.1	87.0	86.8	87.9	88.8	89.3
Georgia	81.8	80.9	80.3	80.4	81.6	81.2	84.2	86.2	87.1	86.6	86.3	86.6	87.4	88.3	88.6
Hawaii	93.5	93.0	92.1	92.9	93.1	93.2	94.7	96.0	96.4	96.1	95.9	95.9	96.1	96.5	96.8
Idaho	82.8	82.6	82.3	83.5	83.8	83.9	86.4	89.0	89.9	89.9	88.9	89.2	91.2	91.8	91.1
Illinois	87.6	86.7	86.2	86.9	87.2	87.3	90.3	92.9	93.5	93.2	93.0	92.6	93.0	93.5	93.8
Indiana	86.6	85.7	85.2	85.5	85.7	86.0	88.1	90.4	91.9	91.8	91.7	91.3	92.5	93.0	93.1
Iowa	91.3	91.4	90.7	91.1	91.6	91.9	93.8	95.0	95.7	95.3	95.3	95.0	95.2	95.5	95.0
Kansas	88.6	86.8	86.1	87.4	87.4	87.7	89.8	90.8	91.3	91.3	91.2	90.8	90.8	91.4	91.7
Kentucky	86.8	85.5	84.8	85.6	86.1	85.7	91.5	94.0	94.9	94.6	94.3	93.5	94.3	94.4	94.6
Louisiana	83.0	82.6	82.2	82.5	83.1	83.4	85.2	88.0	89.7	91.6	92.0	91.1	92.4	93.1	93.1
Maine	89.6	89.5	89.9	89.3	89.7	88.8	89.8	91.6	92.0	91.9	92.0	91.9	94.3	93.4	94.1
Maryland	89.2	88.9	88.7	89.6	89.7	89.8	92.1	93.4	93.9	93.9	94.0	94.0	93.9	93.9	93.7
Massachusetts	96.1	95.8	95.6	95.7	96.1	96.3	96.7	97.2	97.5	97.2	97.2	97.0	97.5	97.6	97.4
Michigan	89.0	87.8	87.6	88.2	88.6	89.0	91.5	93.9	94.6	94.8	94.6	94.2	95.0	95.5	95.5
Minnesota	91.6	90.9	90.9	91.2	92.0	91.8	94.1	95.5	95.9	95.6	95.6	95.1	95.5	95.5	95.8
Mississippi	83.0	82.2	81.8	82.3	82.9	82.9	85.5	87.3	88.2	87.9	87.9	87.0	88.1	89.1	89.7
Missouri	87.6	86.8	86.8	86.3	86.4	87.0	88.3	90.2	91.1	90.9	90.6	90.0	90.6	91.4	92.5
Montana	82.5	81.8	82.7	81.7	82.0	83.5	85.8	88.3	91.9	91.5	91.8	91.7	91.8	91.8	91.6
Nebraska	89.2	88.1	88.5	88.6	88.7	88.6	90.3	91.8	91.4	91.7	91.7	91.7	92.9	93.3	93.9
Nevada	79.3	78.1	77.4	78.1	77.9	79.3	84.8	87.7	88.6	88.8	88.8	88.6	88.3	88.9	89.2
New Hampshire	89.5	89.8	88.9	89.5	89.4	89.3	90.8	93.7	94.1	94.2	94.3	93.7	94.8	95.1	95.2
New Jersey	88.0	87.4	86.8	86.9	87.3	86.8	89.1	91.3	92.0	92.3	92.6	92.1	92.8	93.2	92.8
New Mexico	79.3	80.3	80.4	80.2	81.6	81.4	85.5	89.1	90.8	90.9	90.5	90.0	90.0	91.8	90.9
New York	88.6	88.6	88.1	88.6	89.1	89.3	91.3	92.9	93.9	94.3	94.6	94.8	94.8	95.1	95.2
North Carolina	84.6	83.9	83.2	83.7	83.4	84.4	86.9	88.8	89.6	89.3	89.3	88.7	89.6	90.7	90.8
North Dakota	89.9	90.3	90.2	90.3	89.9	89.7	92.1	92.3	93.0	92.4	92.7	93.1	92.2	93.6	95.5
Ohio	88.6	87.8	87.7	88.1	88.5	89.0	91.6	93.5	94.4	94.0	93.5	93.4	93.5	94.1	93.9
Oklahoma	81.3	81.3	81.1	81.3	81.7	82.3	84.6	86.1	86.2	85.8	85.8	85.7	86.2	88.3	88.7
Oregon	84.2	83.0	82.8	84.3	85.1	85.3	90.3	93.0	93.8	93.2	92.9	92.8	93.9	94.0	94.5
Pennsylvania	91.0	90.1	89.8	89.9	90.3	90.3	91.5	93.6	94.4	94.5	94.5	94.2	94.5	94.7	94.6
Rhode Island	89.8	88.7	87.8	89.2	88.9	88.4	92.6	94.3	95.7	95.4	96.0	95.9	95.6	95.8	95.6
South Carolina	83.3	83.2	82.5	83.3	83.2	84.2	86.4	89.1	90.0	89.0	89.5	89.1	90.0	90.9	90.9
South Dakota	88.9	86.9	87.6	88.1	88.5	88.8	90.2	89.8	91.3	91.0	90.2	89.9	90.5	91.9	91.8
Tennessee	87.1	85.7	85.6	85.4	86.1	86.1	88.0	89.7	91.0	90.5	89.9	89.9	90.0	90.7	90.7
Texas	76.6	76.2	76.3	77.0	77.5	77.9	80.9	82.9	83.4	82.7	82.3	81.6	82.0	83.4	83.6
Utah	85.0	85.4	84.7	84.7	85.5	86.0	87.5	89.5	91.2	90.8	90.6	90.3	91.0	91.9	92.0
Vermont	91.2	91.4	91.9	93.4	93.5	92.8	95.0	96.1	96.3	95.5	96.0	95.5	96.4	96.1	96.6
Virginia	88.4	88.1	86.9	87.5	87.5	87.7	89.1	90.9	91.3	91.2	91.2	92.1	93.2	93.5	93.6
Washington	87.4	86.6	85.7	85.8	86.1	86.0	90.8	93.4	94.0	93.9	93.6	93.4	93.6	93.9	93.7
West Virginia	85.2	85.8	85.4	85.1	85.6	86.0	91.4	94.0	94.7	93.9	93.6	93.3	93.9	94.1	94.1
Wisconsin	91.2	90.6	90.6	91.0	91.0	90.9	92.6	94.3	94.7	94.6	94.5	94.3	94.6	94.8	95.1

Total insured analysis

- The data for percentage insured has been shown state by state. As a general trend the amount of people insured has gone up from 2010 to 2020 with no state having a lower percentage of people covered in 2010 compared to 2020.
- One of the most telling statistics from this is that in 2010 only 10 states had a insurance coverage rate of above 90 percent. By 2020 only seven states were below a 90 percent insurance coverage rate.
- Nevada's 11.8 percent increase was the greatest increase seen from any state. Delaware had the smallest increase with the percentage only going up 3.5 percent.

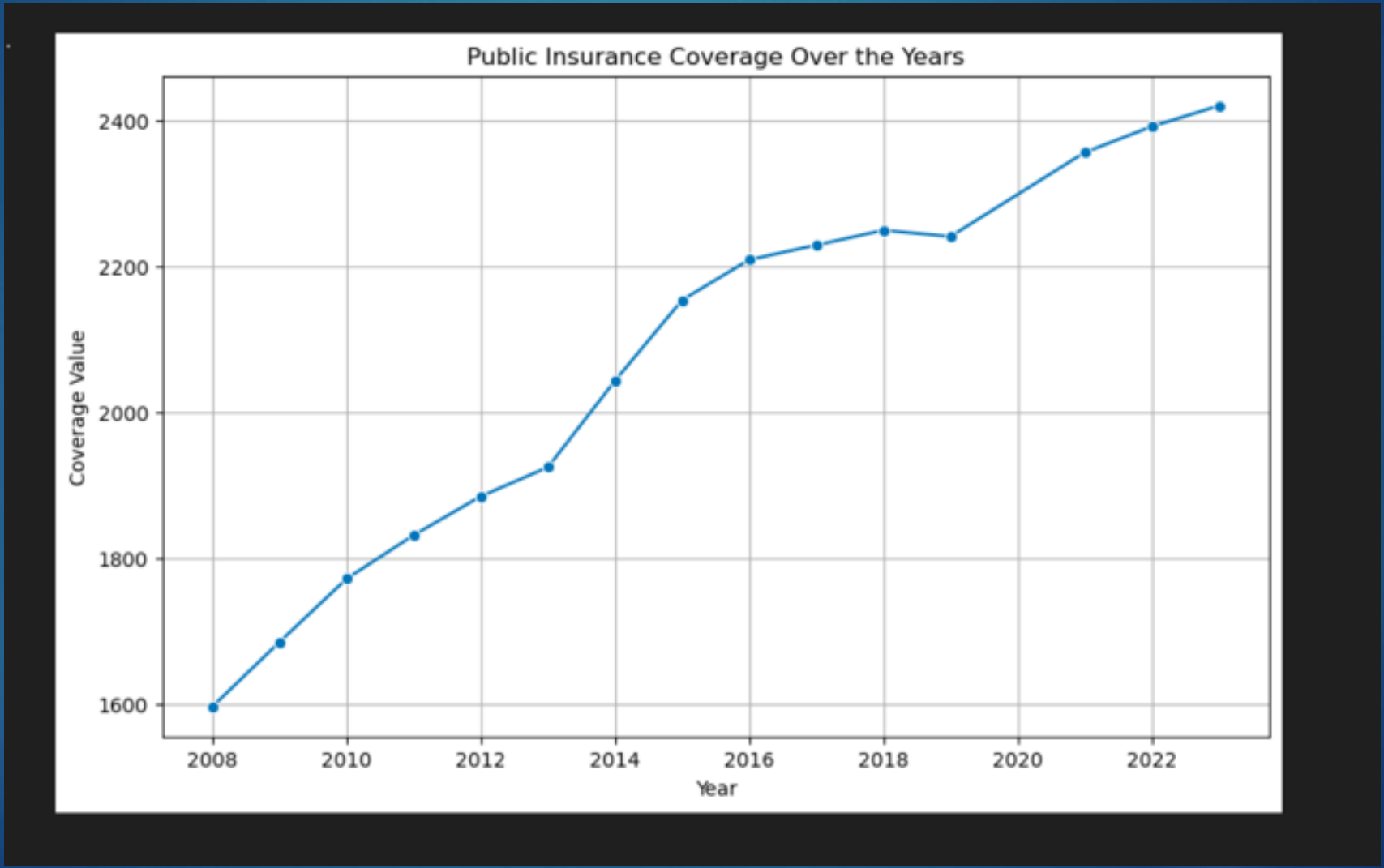
How has the yearly trend in private insurance coverage varied by state from 2011 to 2020?



Private Insurance over the years Analysis

- We looked at the number of people covered by private insurance and
public insurance over the years of 2010 to 2020.
- In terms of private insurance, the number of people only went down in 2011. From there the
number of people covered went up each year.
- In particular huge growth occurred from 2013 to 2016. With the number going up 246 thousand people in that period alone.
- Before that the number from 2010 to 2012
went up 41,000 people. After the number from 2017 to 2020 went up 10,000 people.
- That three year rate increase clearly the most significant change with regards to rates of private insurance

What are the trends in public insurance from 2011 to 2020?



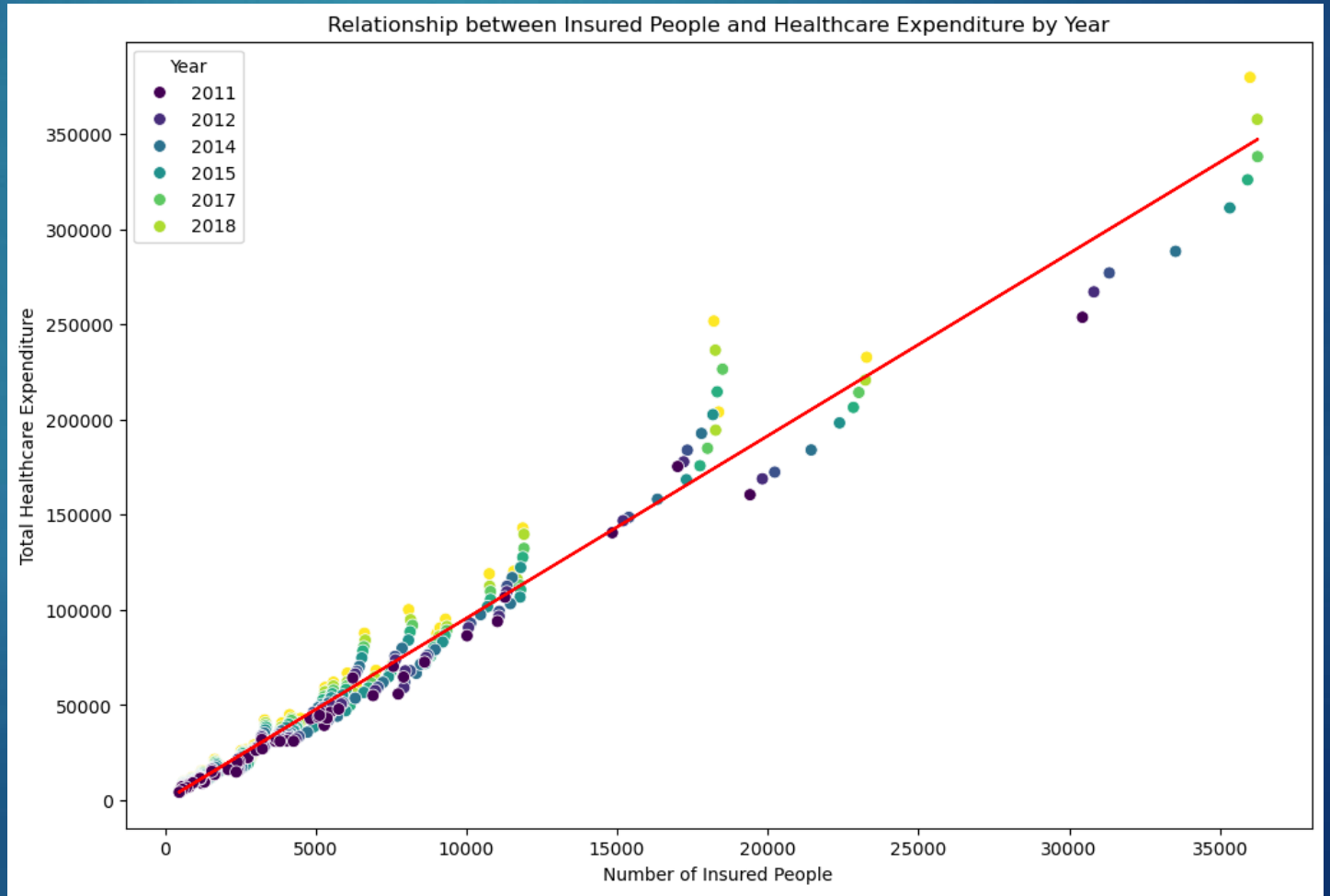
Public Insurance over the years analysis

- When looking at public insurance its coverage rate has grown every year from
  2010 to 2020. It has been a fairly steady rate of growth in coverage from year to year.
- The biggest growth for public insurance was from the years of 2013 to 2015

Comparison of public and private insurance coverage growth over the years

- Public and private insurance have had pretty similar growth of number of people in the US from the years of 2010 to 2020.
- The major similarity of the new is that for the most part the number grew each year the study looked at.
- The major difference is that though both had their highest growth rates from 2013 to 2016 private insurance had a higher growth rate during that time.

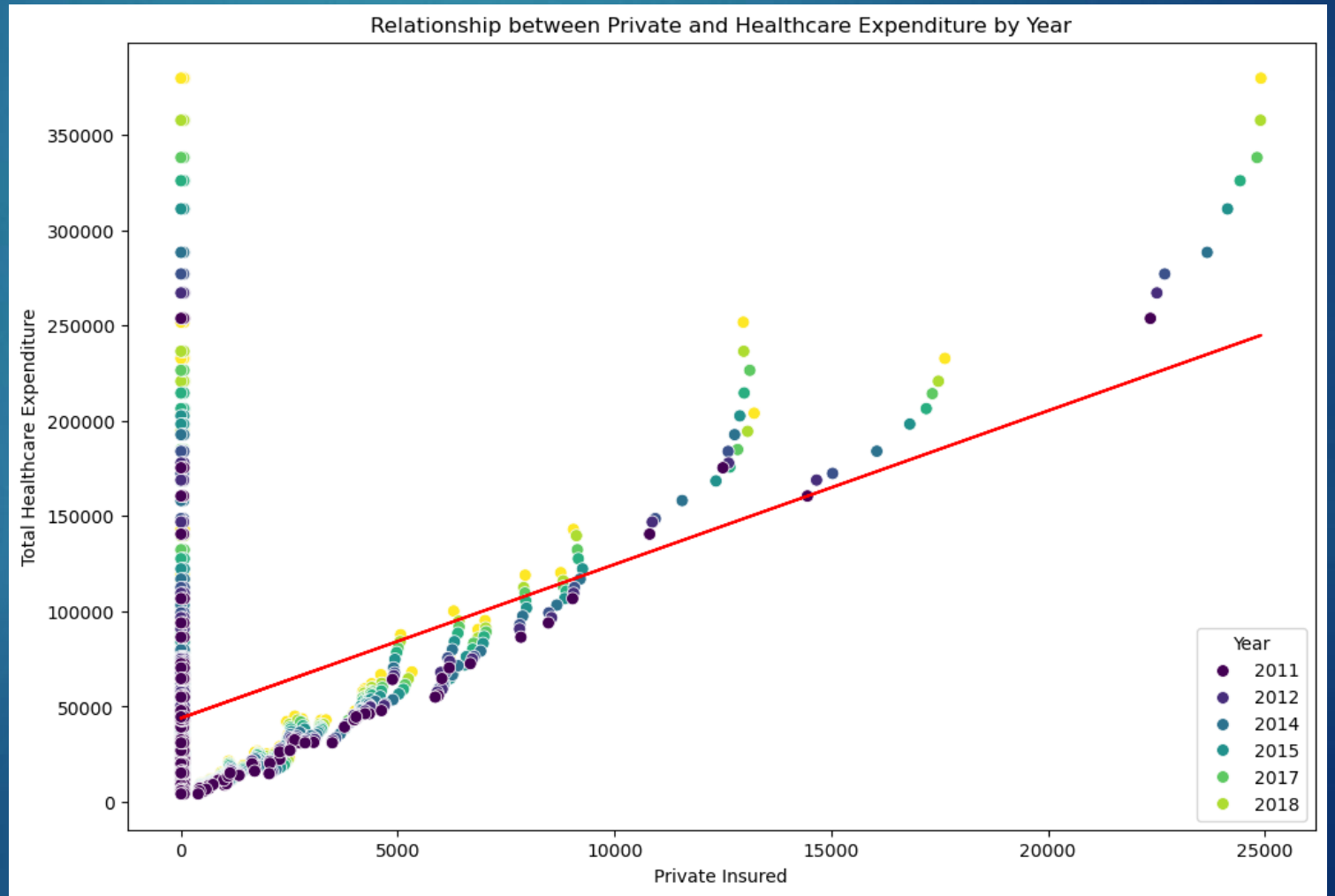
What is the relationship between the number of insured people and the total healthcare expenditure for each year from 2011 to 2020?



Total number of people insured all insurance analysis

- When looking the data of total health care expenditure to number of insured people our data used the r squared to calculate how much our dependent variable of total health care expenditure by people was affected by our independent variable in this case number of people insured.
- The points in the data accumulate all 50 states citizens total health care expenditure of each year between 2010 to 2020.
- One thing to point out is there seems to be three clusters of data with very few outliers. With this it is no surprise that for this test we have a very high r squared value of .976.
- This combined with looking at the direction of line of best fit indicates that it is likely the more people a state has insured the higher the total health expenditure for a state is going to be

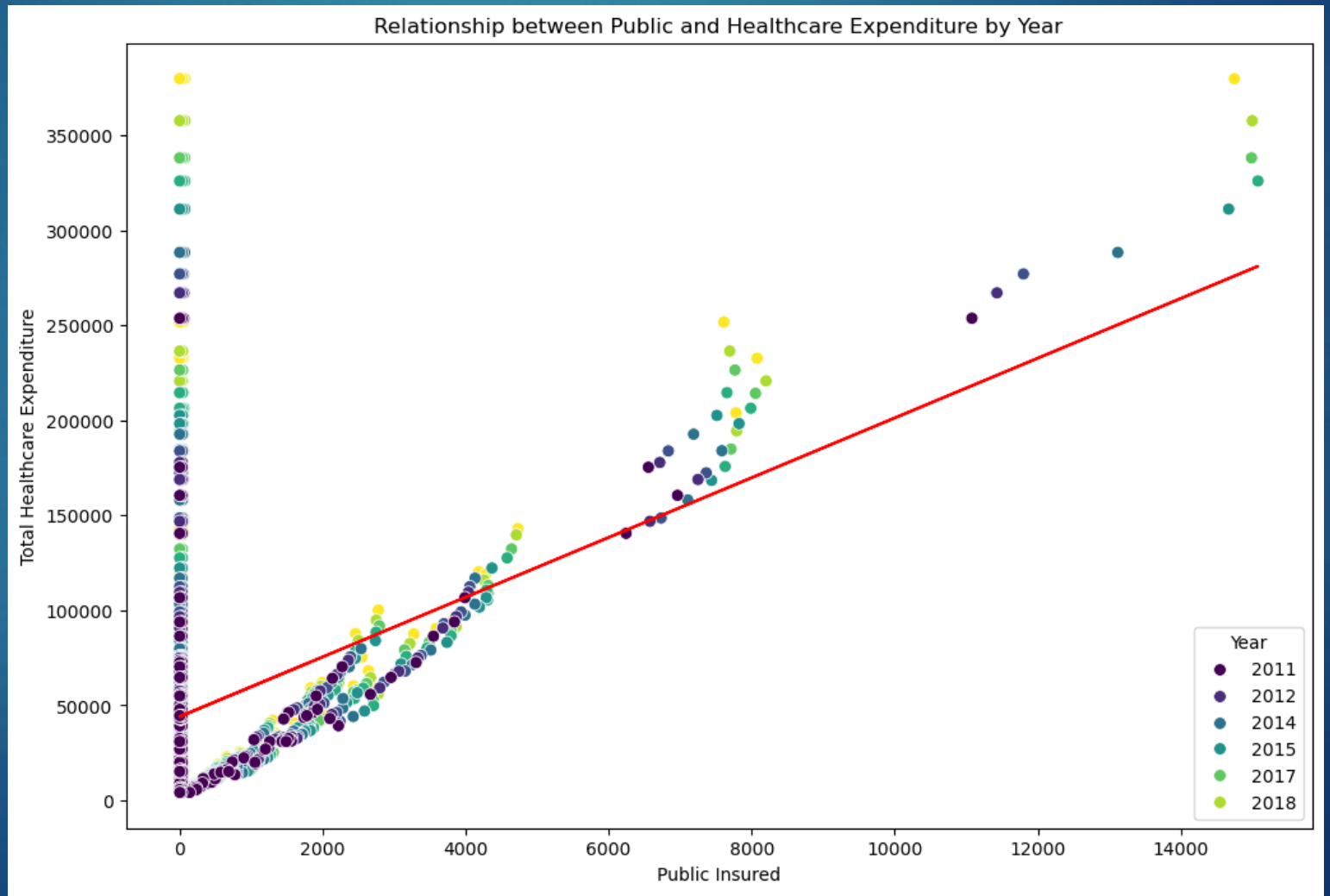
What is the relationship between the number of insured people on public vs private insurance and the total healthcare expenditure for each year from 2011 to 2020?



Total number of people on private insurance analysis

- When looking at the data for total health care expenditure compared to the
amount of people covered by private healthcare and the total health care expenditure a r squared analysis was run. Like the total amount of people insured the data clusters around three major spots.
- The difference to the total people insured is that the r squared value is .148. This value indicates that the amount of people covered on private insurance is not likely to affect the total health expenditure citizens spend for a given state. Certainly not as likely to affect as total people insured.
- However similarly to total people the line of best indicates if there is an affect it is the effect of the more people on private insurance the higher the health care expenditure.

How many people insured in each state via public insurance vs how much total expenditure per state 2010 to 2020

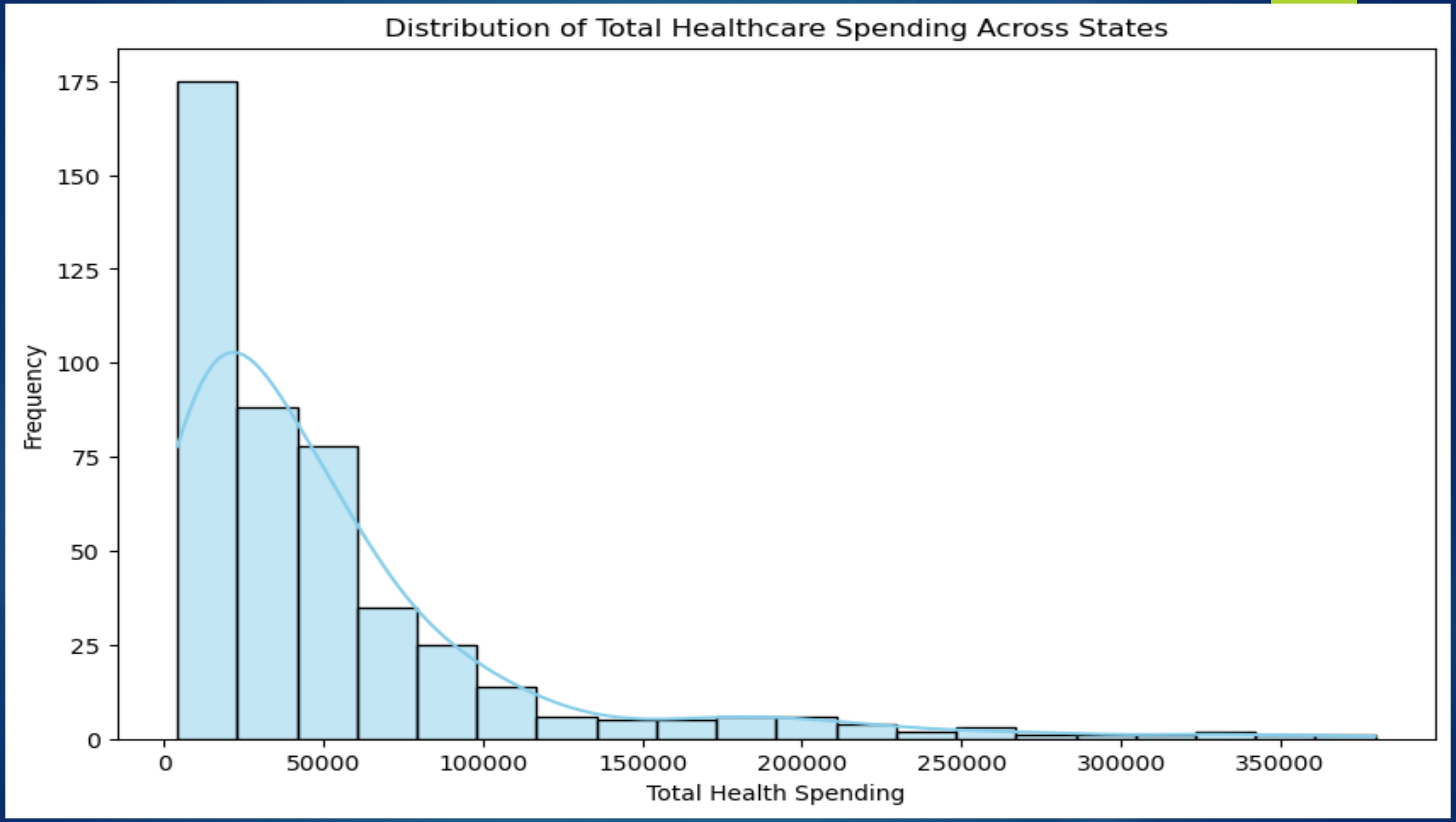


Total number of people on private insurance analysis

- When looking at the data for total health care expenditure compared to the amount of people covered by public healthcare an r square analysis was run.
- A similar three data clusters is found to both number of total people insured and total people
  insured by private health care. Its r squared value is .159 which is higher than private
  insurance, but far less then total people insured.
- We can therefore say is has slightly more likelihood of having affected total health care expenditure then private insurance. Very slight though probably not enough to be too significant.
- However similarly to total people the line of best indicates if there is an affect it is the effect of the more people on private insurance the higher the health care expenditure. Whereas it is much less likely to have affected total health care expenditure then total people covered by all insurance

Total health expenditure compared to type and amount of people insured analysis

- The amount of total healthcare expenditure the citizens of a state spend does not seem to be affected that much by either the number of people who have private health care or the number of people who have public health care.
- However, the amount of health care expenditure does seem to be very affected by the amount of people covered by any type of health care.
- So, it is not the type that seems to matter but the raw amount


Distribution of Total Healthcare Spending Across States
Frequency
0
25
50
75
100
125
150
175
0
50000
100000
150000
200000
250000
300000
350000
Total Health Spending

Analysis

- The data is right-skewed, indicating that states have relatively lower total healthcare spending, with the frequency of states decreasing as spending levels increase, and only a few states exhibiting

Summary of Hypotheses

- How does the amount of coverage of public vs private insurance impact how all citizens in a state spend on healthcare in a year?
- How has the yearly trend in insurance coverage varied by state from 2011 to 2020?
- What are the trends in private insurance enrollment by state from 2011 to 2020?
- What are the trends in public insurance enrollment by state from 2011 to 2020?
- What is the relationship between the number of insured people and the total healthcare expenditure for each year from 2011 to 2020?
- What is the relationship between the number of insured people on public vs private insurance and the total healthcare expenditure for each year from 2011 to 2020?

Our Conclusion/ Further Studies

- health care expenditure per capita or person.
- diving more into that huge jump
  of coverage from 2013 to 2016
- One other thing could be expanded on is comparing individual states on these factors. This study mostly used states as points in an overall larger look as expenditure overall. Comparing
  things by region or highest spending to lower spending states would be an interesting
  continuation.

Sources

Census.gov Health insurance Coverage in the United States

https://www.census.gov/library/publications/2023/demo/p60-281.html#:~:text=HIC%2D4_ACS.%20Health%20Insurance%20Coverage%20Status%20and%20Type%20of%20Coverage%20by%20State%2D%2DAll%20Persons%3A%202008%20to%202023%20%5B%3C1.0%20MB%5D

KFF.org Health Care expenditures by State

https://www.kff.org/other/state-indicator/health-care-expenditures-by-state-of-residence-in-millions/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D