Bigfoot sightings Analysis

Chang Liu

About the Bigfoot Data

1. The bigfoot data is retrieved from Bigfoot Field Researchers Organization (BFRO), which collected and compiled the data from public reports and eyewitness accounts.

2. One row represents one record of Bigfoot sighting events.

3. Key Columns:

    -Date: The date on which the sighting occurred.

    -Location: Geographic location of the event, including state and county

4. My analysis will mainly focus on county-level data and find if some social factors will affect the number of bigfoot sightings.

Hypothesis

1. There is a relationship between the number of Bigfoot sightings in one county and its poverty rate as well as the number of people living in poverty.

2. There is a relationship between the number of Bigfoot sightings in a county and the level of education (proportion of population completing college-level education) in that county.

3. There is a relationship between the number of Bigfoot sightings in a county and the county's unemployment rate in the same year.

Data Restructure

1.In the original Bigfoot dataset, there is a column for the county, but an issue arises because many counties in the United States share the same name, such as "Washington County," making it difficult to accurately count sightings by county alone.

41]:

	county	state	season	latitude	longitude	date
0	Washakie County	Wyoming	Summer	NaN	NaN	NaT
1	Wyoming County	West Virginia	Winter	37.58135	-81.29745	2005-12-03
2	Windsor County	Vermont	Fall	43.46540	-72.70510	2005-10-08

Data Restructure

1.To address this, a new column, ‘county_state’, should be created that combines the names of the county and the state. This approach ensures each county is uniquely identified, allowing for precise recording and analysis of sightings data.

]:

	county_state	report_count
0	Pierce County, Washington	78
1	Snohomish County, Washington	51
2	Skamania County, Washington	49
3	Lewis County, Washington	47
4	Humboldt County, California	45

Data Restructure

2.Not every county has records of Bigfoot sightings at all times, but they might appear in other datasets such as unemployment rates or educational leve data. So, I choose to insert these counties that haven't appeared in bigfoot sightings and fill their missing values with zeros.

```
missing_counties = set(county_rate['county_state']) - set(county_counts['county_state'])

# Create a new DataFrame for these missing counties with zero sightings
year_columns = [col for col in county_counts.columns if col != 'county_state']
missing_data = {year: [0]*len(missing_counties) for year in year_columns}
missing_data['county_state'] = list(missing_counties)

missing_df = pd.DataFrame(missing_data)

# Append the new DataFrame to the existing sightings data
county_updated = pd.concat([county_counts, missing_df], ignore_index=True)

return county_updated
```

Hypothesis One: Poverty

- Use Correlation to analyze the relationship between bigfoot sightings and poverty population as well as poverty rate

	state_abbreviation	county_or_state_name	poverty_population	poverty_rate in 2021
2	AL	Autauga County	6296.0	10.7
3	AL	Baldwin County	25526.0	10.8
4	AL	Barbour County	5089.0	23.0
5	AL	Bibb County	4204.0	20.6
6	AL	Blount County	6992.0	12.0

Hypothesis One: County Level and State Level

1.State level
— The correlation between poverty rate and Bigfoot sightings is approximately 0.0133 which indicates almost no linear relationship

— The correlation between the impoverished population and Bigfoot sightings is approximately 0.5986, indicating a moderate to strong positive correlation.

— This may suggest that in states with larger populations, there are more Bigfoot sighting reports, possibly due to a wider geographical range and more resident activity.

2. County level
— The correlation coefficient between the poverty rate and Bigfoot sightings is -0.0586

— The correlation coefficient between the impoverished population and Bigfoot sightings is 0.1364. There is also a similar relationship between s the impoverished population and the number of Bigfoot sighting reports.

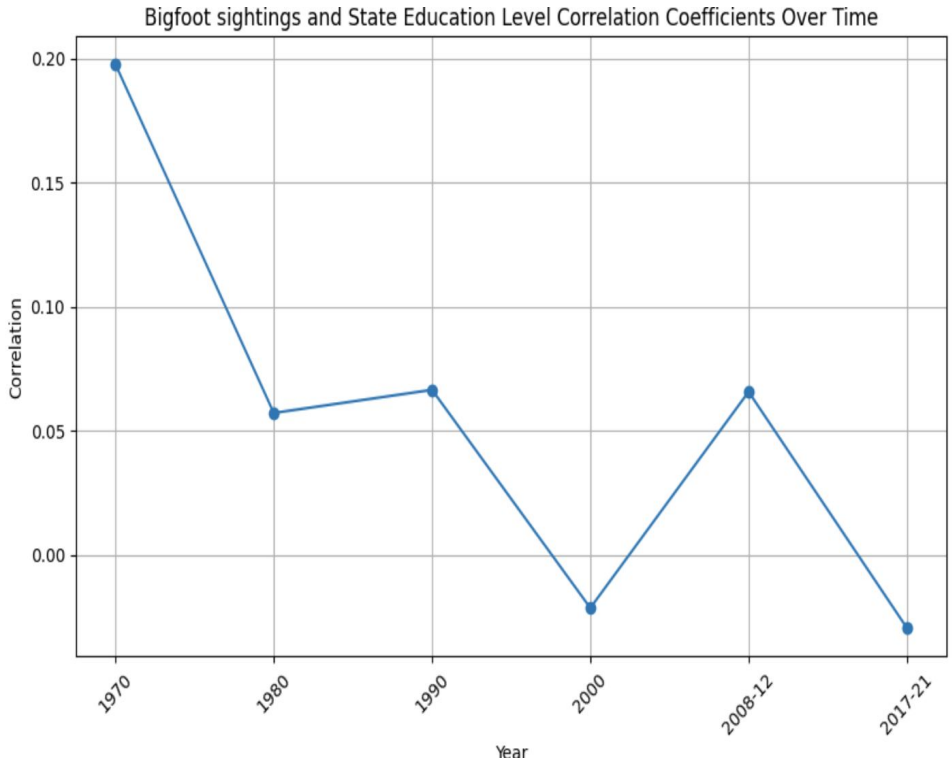
— This may imply that in areas with larger populations, especially in economically disadvantaged areas, there may be more opportunities to report Bigfoot sightings due to higher population density or more frequent outdoor activities by people, which proves the conclusion when we analyze the state-level data

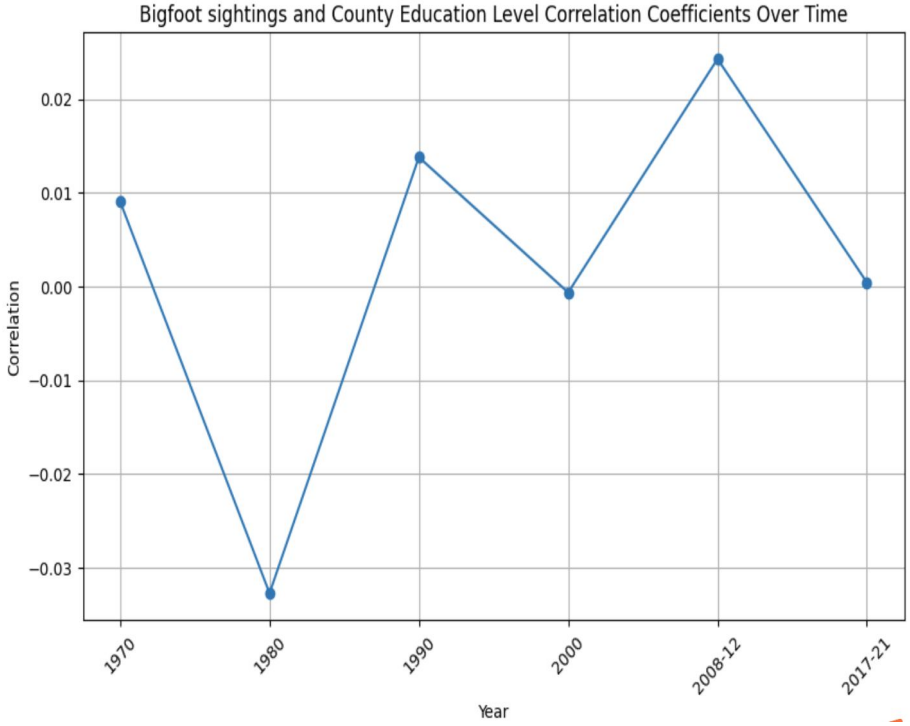
Hypothesis Two: Educational Level

- Use Correlation to analyze the relationship between bigfoot sightings and educational level(the percentage of population who finish the college-level education (The time range is different)

	county_state	1970	1980	1990	2000	2008-12	2017-21
2	Autauga County, Alabama	6.4	12.1	14.5	18.0	21.707831	28.131469
3	Baldwin County, Alabama	6.5	12.1	16.8	23.1	27.741591	32.450286
4	Barbour County, Alabama	7.3	9.2	11.8	10.9	14.524286	11.153098
5	Bibb County, Alabama	4.2	4.9	4.7	7.1	8.996005	11.913807
6	Blount County, Alabama	2.7	5.3	7.0	9.6	12.381469	14.903610

Correlation Change Over time period





Hypothesis Two: Educational Level

1. Overall, these data suggest that the correlation between Bigfoot sightings and education levels has varied across different time periods, from a noticeable positive correlation in the 1970s to a slight negative correlation in the 21st century. This could be related to changes in social and cultural attitudes, reporting behaviors, and public interest in mystical phenomena.

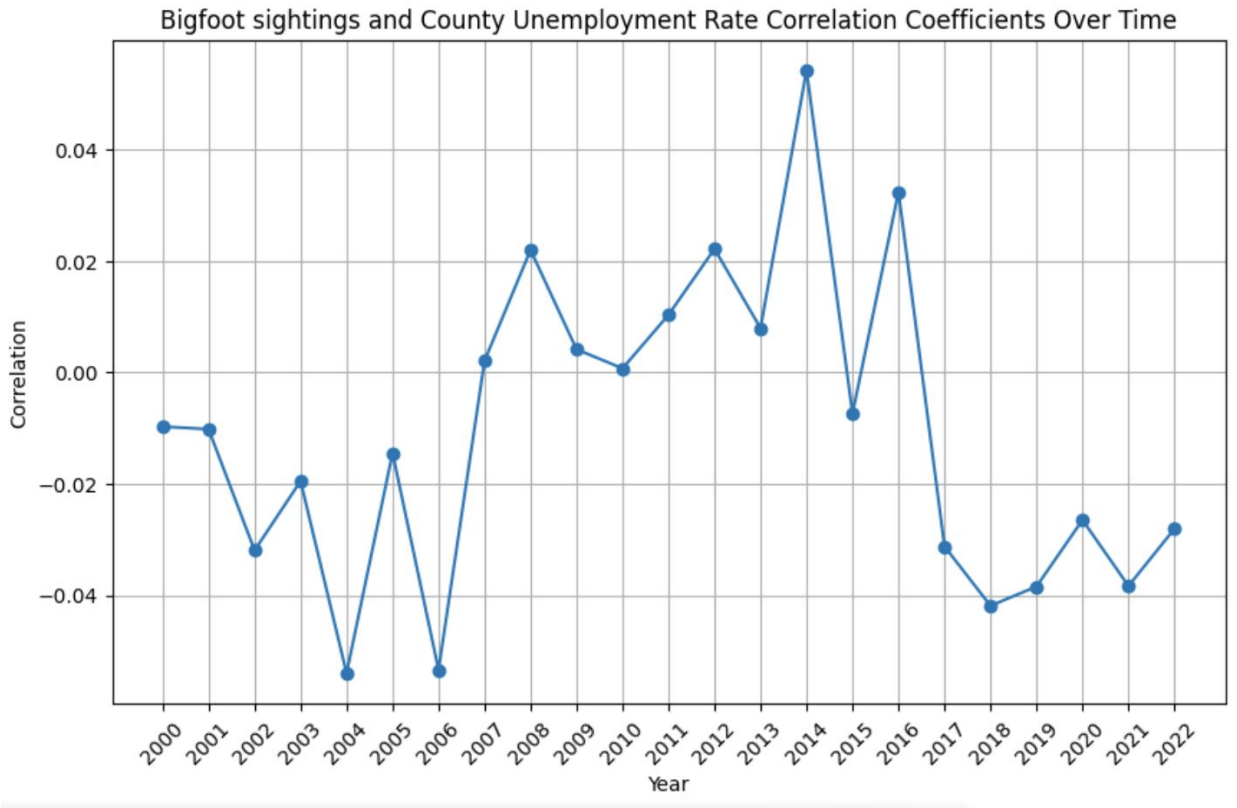
2. Comparing these county-level results with the state-level data, we observe that the overall patterns are similar. At the county level, the correlations are generally weaker and show varied trends across the decades, which may indicate that local factors at the county level dilute or modify the influence that educational attainment has on reporting Bigfoot sightings.

Hypothesis Three: Unemployment Rate

- Use Correlation to analyze the relationship between bigfoot sightings and unemployment rate

	Area_Name	Unemployment_rate_2000	Unemployment_rate_2001	Unemployment_rate_2002	Unemployment_rate_2003
2	Autauga County, AL	4.1	4.1	4.8	5.0
3	Baldwin County, AL	3.7	4.3	5.0	4.9
4	Barbour County, AL	5.6	7.6	7.7	7.1
5	Bibb County, AL	5.4	6.8	7.0	5.9
6	Blount County, AL	3.5	3.7	5.4	4.6

Correlation Graph



Hypothesis Three Analysis: County Level

1.From the graph above, it is evident that most years exhibit negative correlation coefficients, suggesting a slight negative correlation between Bigfoot sightings and unemployment rates in most years.

2.Overall, these data reveal that while there are annual fluctuations, there is generally no strong significant correlation between Bigfoot sightings and unemployment rates. The variation in the strength and direction of this relationship could be influenced by multiple factors such as changes in the economic environment, public attention, and variations in reporting behaviors.density of educational institutions, or even regional differences in how educational attainment influences perceptions of folklore and unexplained phenomena might play a role in these varying correlations.

Future Plan

1. Add comments in Notebook
2. Add state level for unemployment data
3. Try to analyze the data from another dimension. For example, I will attempt to discover if the counties with the fewest sightings in each state have significantly different average education levels and unemployment rates compared to other counties.

Thank you!