Analyzing the Relationship Between Deforestation Rates and Temperature Anomalies

Rebekah Northrup
January 2025

Table of contents

01

Why This Dataset?

Global Warming Dataset: 195 Countries (1900–2023)

Motivation:

- Interest in understanding the environmental impact of deforestation on global temperature anomalies

Questions I hoped to explore:

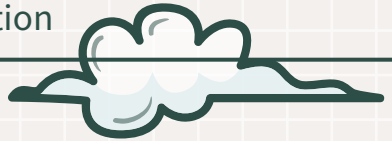
- Is there a correlation between deforestation rates and temperature anomalies?
- Can deforestation rates predict temperature anomalies?

02

Approach to Analysis

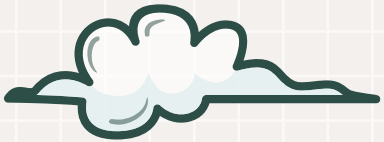
Approach to Analysis

Initial Steps	
Data Cleaning:	1. Handled missing values and checked for outliers 2. Calculated annual deforestation rates from forest area data
Exploratory Data Analysis (EDA):	1. Visualized the distribution of variables 2. Investigated correlations between deforestation rates and temperature anomalies
Statistical Modeling:	1. Performed linear regression to test the predictive relationship between variables
Challenges Faced:	1. Extreme skewness in deforestation rates led to clustered data 2. Near-zero correlation indicated weak relationships, requiring careful interpretation



Approach to Analysis

Lessons Learned	
What Worked:	1. Visualizations provided clear insights into data distribution and trends 2. Statistical tests validated the absence of significant relationships
What Didn't Work:	1. Linear regression failed to provide meaningful results due to weak correlations 2. Extreme outliers in deforestation rates distorted visualizations



03

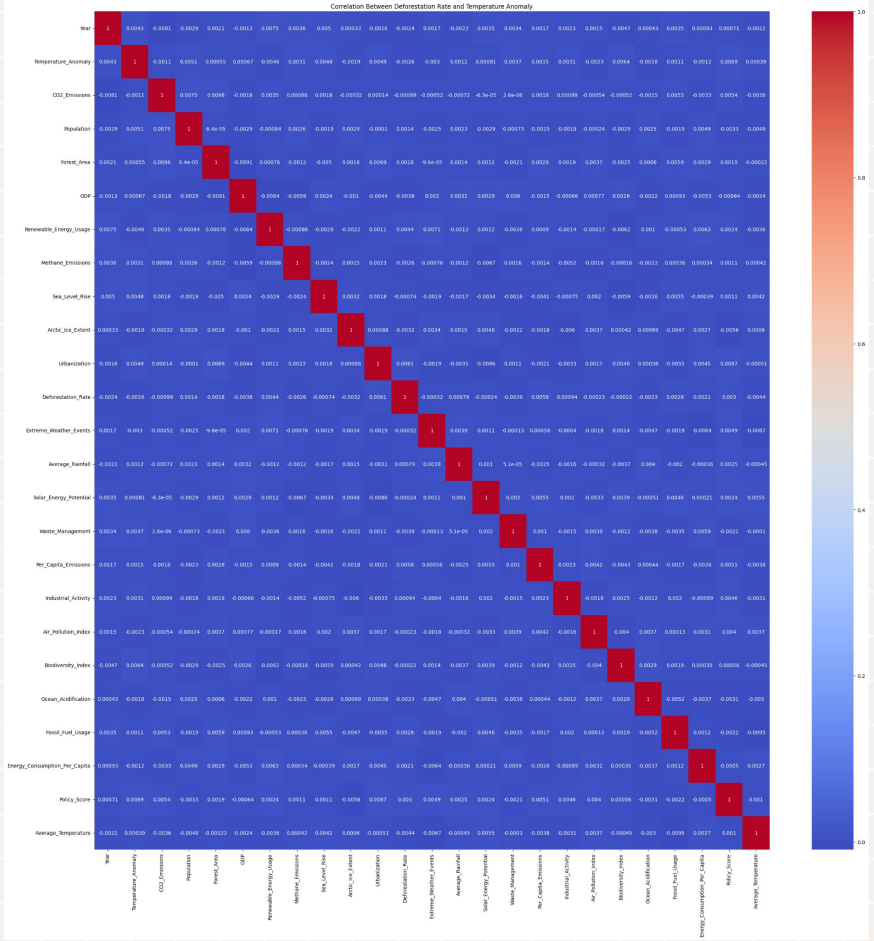
Key Findings

Correlation Analysis:

Heatmap Revealed:

- Correlation between deforestation rates and temperature anomalies: -0.0026
- No significant linear relationship detected

	Deforestation_Rate	Temperature_Anomaly
Deforestation_Rate	1.000000	−0.002633
Temperature_Anomaly	−0.002633	1.000000



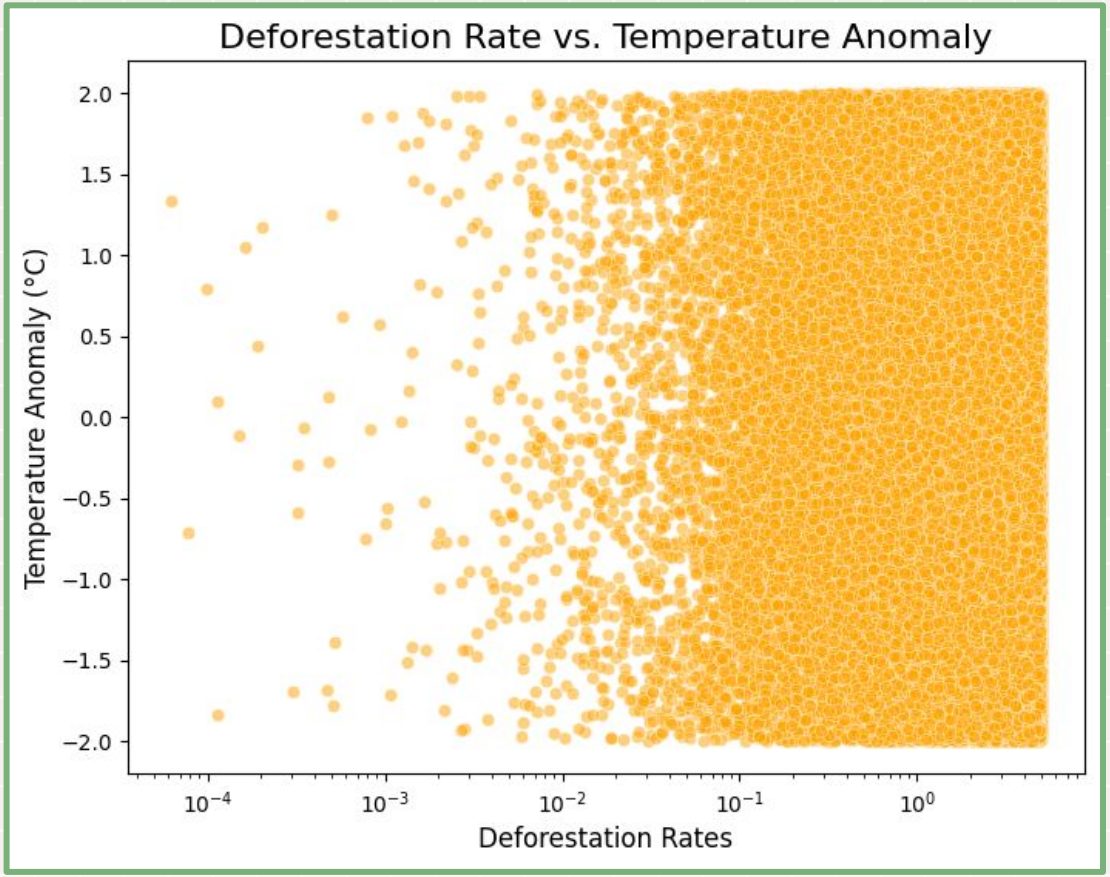
Linear Regression Results:

- **R^2** Score: ~4.67e-06
  - Indicates the model explains no variance in temperature anomalies
- Mean Squared Error: ~1.336
  - Highlights minimal predictive power of deforestation rates

```python
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, mean_squared_error

# Prepare data for regression
regression_data = data.dropna(subset=['Deforestation_Rate', 'Temperature_Anomaly'])
X = regression_data[['Deforestation_Rate']].values
y = regression_data['Temperature_Anomaly'].values

# Split into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Fit linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate model
print(f"R^2 Score: {r2_score(y_test, y_pred)}")
print(f"Mean Squared Error: {mean_squared_error(y_test, y_pred)}")
```

Visualizations:

- **Scatter Plot: Deforestation vs. Temperature Anomaly**
  - Plotted on a logarithmic scale for clarity
  - Showed no discernible trend between the variables



04

Conclusion

Next Steps:

- **Future Investigations:**
  - Feature engineering:
    - Apply non-linear transformations to deforestation rates
    - Incorporate additional predictors (e.g., CO_2 emissions, industrial activity)
  - Advanced modeling:
    - Use machine learning models like Random Forests or Gradient Boosting
  - Geospatial Analysis:
    - Explore regional patterns in deforestation and temperature anomalies
- **Limitations Encountered:**
  - Lack of other environmental factors in the dataset
  - Insufficient granularity in deforestation data for regional analysis

Conclusion:

- **Weak Relationship:**
  - Both the statistical results (R^2 and correlation) and visual analysis confirm that deforestation rates are not a significant linear driver of temperature anomalies in this dataset.
- **Outliers and Skewness:**
  - The extreme skewness in the distribution of deforestation rates might obscure any underlying patterns or relationships
  - Potential outliers could distort the regression and scatter plot results
- **Other Influences:**
  - Temperature anomalies are likely influenced by other variables, such as greenhouse gas emissions, industrialization, or renewable energy usage
- **Takeaways:**
  - Deforestation impact on climate requires a broader context
  - Encourages integration of additional variable for more robust analyses

Thank You!

rebekah1247@gmail.com
(910)203-9389