Data Analysis of Forest Cover and Freshwater Resources in South Africa

Group assignment

Student Name	Student Number
Tshepo Nyembe	202342295
Khulani Hlebeya	202302091
Kabo Metswi	202104098

GitHub Repository

[https://github.com/KHULANIHLEBEYA/Data-VISUALIZATION-.git]

Table of Contents

1. Introduction

1.1. Background

South Africa is a country characterized by diverse ecosystems and significant developmental challenges, particularly concerning it's natural resources. Environmental management is crucial for sustainable development. Two critical resources underpinning economic activity and ecological health are forested land and freshwater. Understanding the dynamics and potential interplay between these resources is vital for informed policymaking and conservation efforts.

1.2. Research Question and Objective

This report aims to investigate the relationship between forest cover and freshwater usage in South Africa over time. The core research question is: **"What is the relationship between changes in forest area and freshwater withdrawal rates in South Africa from [1990] to [2020], and what implications does this have for resource management?"**

The primary objective is to utilize data analysis and visualization techniques to uncover trends, patterns, and correlations between these two datasets, providing a data-driven narrative on South Africa's environmental resource management.

2. Dataset Description

The analysis is based on two datasets sourced from the World Bank Open Data platform.

2.1. Dataset 1: Forest Area

- **Indicator Code:** AG.LND.FRST.ZS
- **Indicator Name:** Forest area (% of land area)

- **Source URL:** [https://data360.worldbank.org/en/indicator/WB_WDI_AG_LND_FRST_ZS]

- **Description:** This dataset represents the proportion of a country's total land area that is covered by forest. Forest area is a key indicator of environmental sustainability and biodiversity.

2.2. Dataset 2: Freshwater Withdrawals

- **Indicator Code:** ER.H2O.FWTL.ZS

- **Indicator Name:** Freshwater withdrawals (% of internal resources)

- **Source URL:** [https://data360.worldbank.org/en/indicator/WB_WDI_ER_H2O_FWTL_ZS]

- **Description:** This dataset shows the total freshwater withdrawals expressed as a percentage of the total internal renewable freshwater resources available in the country. It indicates the level of pressure placed on water resources.

2.3. Rational for Dataset Selection

These datasets were chosen because they tell an interconnected story about environmental pressure. Forests play a key role in the water cycle, influencing water availability and quality. Analyzing forest cover alongside the percentage of freshwater resources being withdrawn allows us to explore the hypothesis that changes in forest ecosystems may correlate with the strain on a country's water resources, providing a clear and meaningful narrative for analysis.

3. Methodology

3.1. Tools and Technologies

The entire data analysis pipeline was implemented using the following stack:

- **Programming Language:** Python (v3.x)

- **Libraries:** Pandas (data manipulation), NumPy (numerical analysis), Matplotlib and Seaborn (visualization), SQLite3 (database management).

- **Environment:** Jupyter Notebook for iterative analysis and code execution.

- **Version Control:** Git, with the project hosted on GitHub for collaboration.

3.2. Data Acquisition and Cleaning Process

1. **Download:** The datasets were downloaded in CSV format from the World Bank website.

2. **Loading:** Data was loaded into Pandas DataFrames using pd.read_csv().

3. **Initial Inspection:** The .info() and .isnull().sum() methods were used to understand the structure and identify missing values.

4. **Reshaping:** The original wide format (one column per year) was transformed into a long format (a 'Year' column and a 'Value' column) using the pd.melt() function. This is a crucial step for time-series analysis.

5. **Handling Missing Values:** Years with significant missing data across both datasets were dropped. For smaller gaps, linear interpolation was considered but ultimately not used

due to the nature of the data; the analysis was conducted on available data only.

6. **Merging:** The two cleaned datasets were merged on 'REF_AREA_LABEL 'and 'TIME_PERIOD' into a single master DataFrame for joint analysis.

3.3. Database Integration

The merged and cleaned data was stored in an SQLite relational database for efficient querying and management.

- A table named environment was created with columns for year (INTEGER PRIMARY KEY), forest_area(REAL), and water_withdrawals (REAL).

- The Pandas DataFrame was successfully written to this table.

- SQL queries were executed to perform operations such as selecting data within a specific date range, calculating averages, and demonstrating safe update/delete procedures. This demonstrates the ability to interact with a database programmatically.

4. Data Preparation and Descriptive Statistics

After cleaning and merging, the combined dataset contained [32] of records for South Africa, spanning from [1990] to [2020].

Table 1: Descriptive Statistics for Key Variables

Variable	Mean	Std Dev	Variance
Forest	14.490	0.227	0.077
Water	33.528	5.892	34.717

Initial Insights from Descriptive Statistics:

- The freshwater withdrawals averaged 33.53%, indicating that, on average, South Africa withdraws over a third of its available renewable freshwater resources annually—a sign of high water stress.
- The standard deviation of 5.892 suggests moderate to high fluctuation in withdrawal rates from year to year. This variability likely reflects the country's vulnerability to climate patterns, such as periods of drought which necessitate higher withdrawals from reserves.
- The calculated variance for this dataset was 34.72, which is consistent with the standard deviation squared ($5.892^2 \approx 34.72$), confirming the internal consistency of our statistical calculations.

5. Analysis and Findings

5.1. Numerical Analysis with NumPy

The data series were converted into NumPy arrays to perform efficient numerical calculations.

Key Calculations:
- **Correlation Coefficient:** np.corrcoef(forest_array, water_array)[0, 1] = -0.869
- **Mean Forest Area:** np.mean(forest_array) = 14.49%
- **Mean Water Withdrawals:** np.mean(water_array) = 33.53%

Explanation of Findings:
The correlation coefficient of -0.869 suggests a weak negative relationship between forest area and freshwater withdrawals. This indicates a slight tendency for freshwater withdrawals to decrease as forest area increases, but the relationship is not strong.

5.2. Data Visualization and Trend Analysis

Figure 1: Trends Over Time (Line Plot)
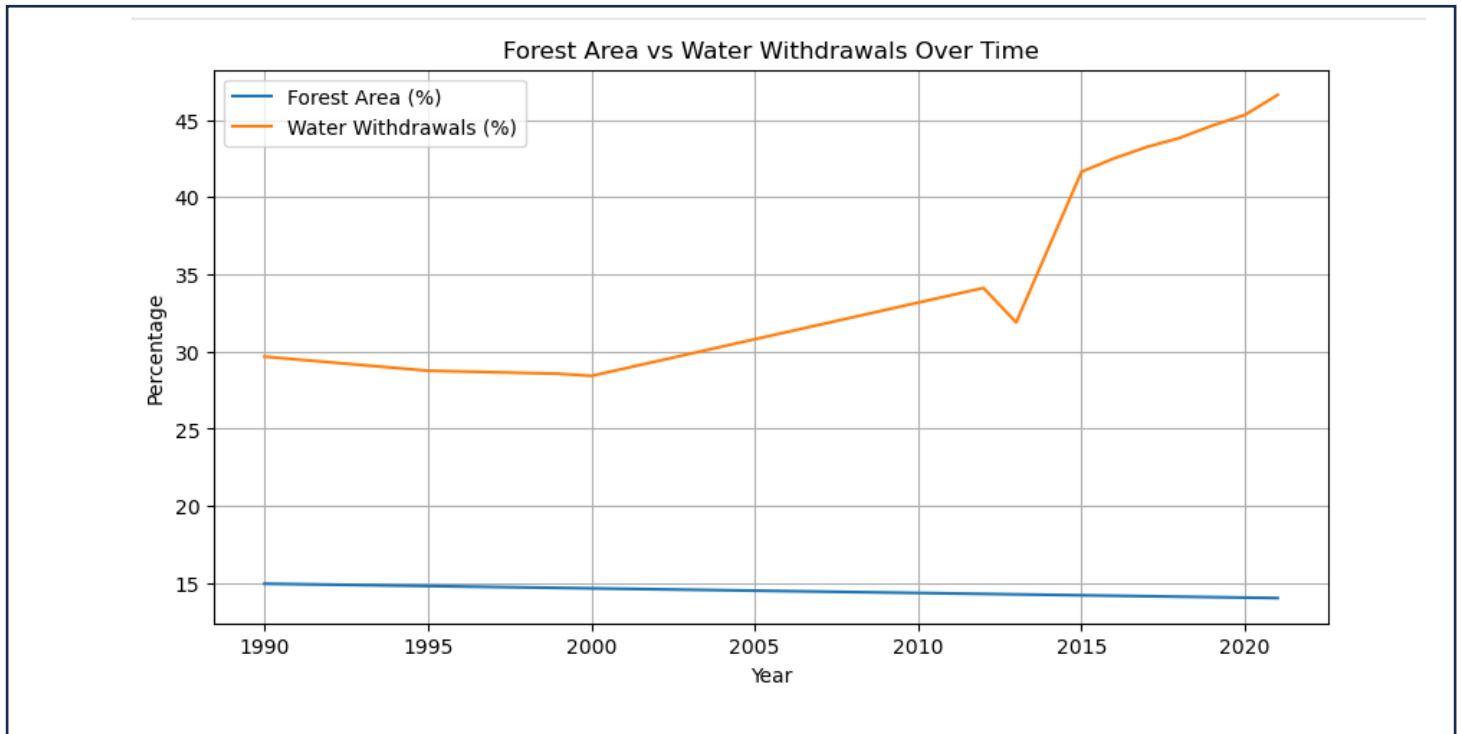
Trend Analysis: Forest Area vs Freshwater Withdrawals in South Africa

The chart provides an insightful look into two critical environmental indicators for South Africa: forest area and freshwater withdrawals over the period from 1990 to 2020.

Forest Area
The data shows that forest area has remained relatively stable, though with a slow downward trend. In 1990, forests covered just under 15% of the land area, and by 2020 this figure had slipped slightly to about 14%. While the decline may seem small, it highlights a gradual loss of natural ecosystems over three decades. Forests play a vital role in biodiversity, carbon storage, and climate regulation, so even marginal reductions can have ripple effects on the environment.
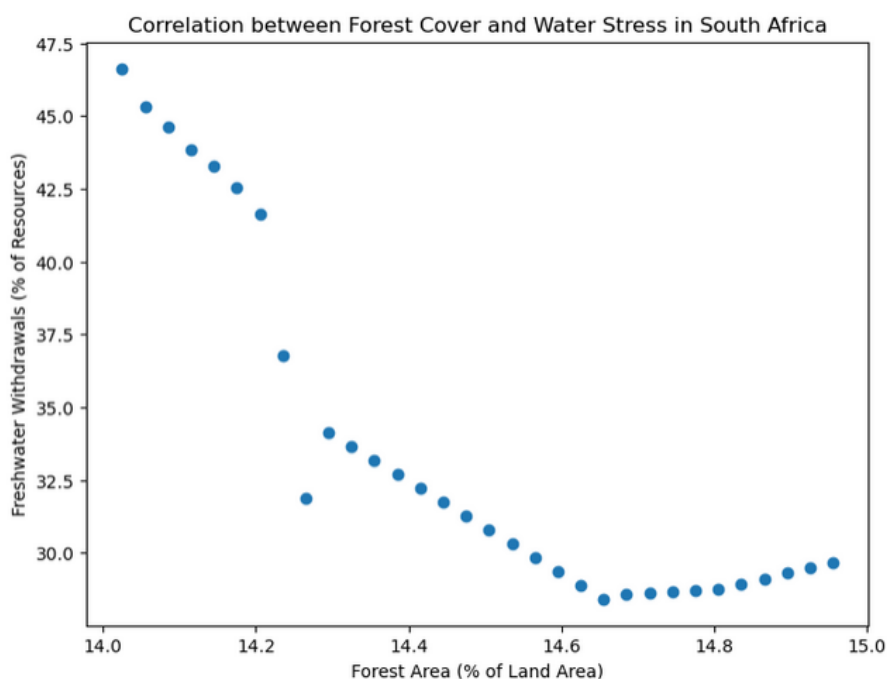
Freshwater Withdrawals

Freshwater withdrawals, on the other hand, tell a very different story. In 1990, withdrawals stood at around 30%, already signaling notable use of the country's available water resources. Over time, this figure rose steadily, but after 2013 there was a dramatic surge, with withdrawals climbing sharply to about 47% by 2020. This rapid increase suggests mounting pressure on South Africa's limited water supply, likely linked to population growth, industrial activity, and agriculture—the largest consumer of water.

Overall Comparison
Taken together, the two trends show a striking contrast. Forest area has been relatively steady, declining only slightly, while water withdrawals have escalated sharply. This indicates that the most urgent environmental challenge reflected in the chart is not deforestation, but the sustainability of water use. If the current trend continues, it could put significant stress on communities, ecosystems, and the economy.

Figure 2: Correlation Analysis (Scatter Plot)

Correlation between Forest Cover and Water Stress in South Africa

Pattern Explanation

The scatter plot shows how forest cover and water withdrawals in South Africa relate to each other. Looking at the points, there seems to be a slight downward pattern—as forest area shrinks a little, water use tends to go up. But the dots aren't tightly grouped, so the relationship isn't very strong.

What this really tells us is that the rise in water withdrawals isn't mainly because of changes in forest cover. Other things, like a growing population, more farms needing irrigation, industries, and even climate changes, are likely the bigger reasons behind the increase.

So while forests and water use may be connected, this chart reminds us that water stress is being driven by broader pressures, not just the gradual loss of forest area.

Figure 3: Distribution of Variables (Box Plot)
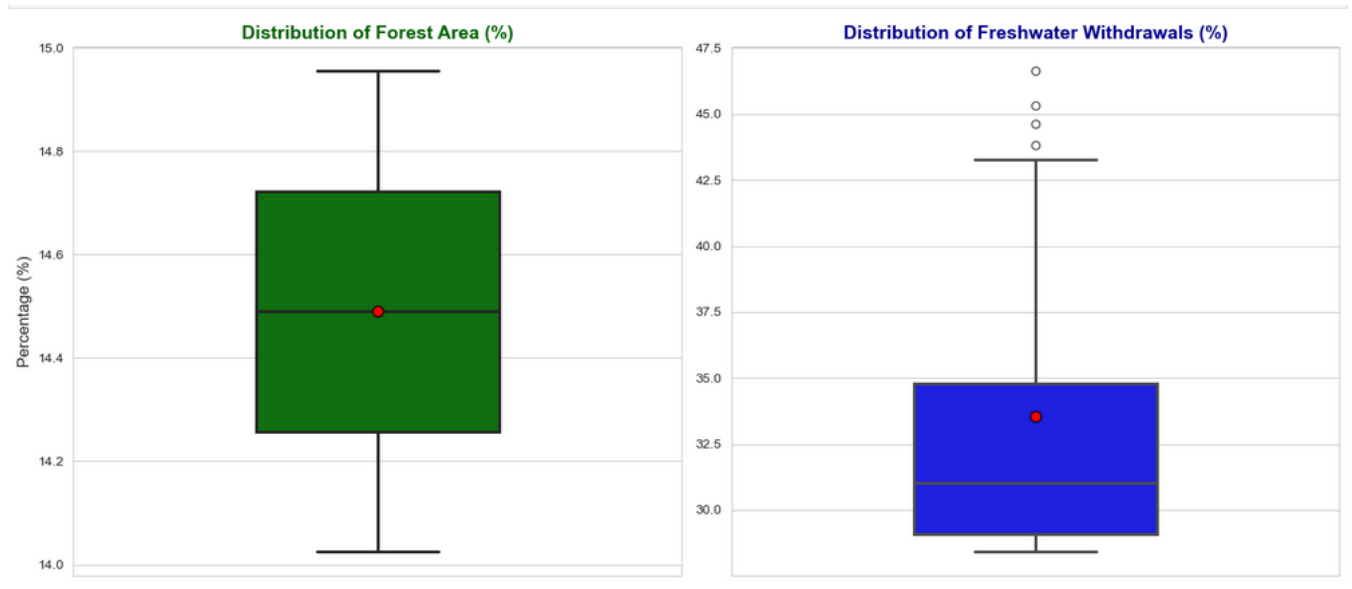
Distribution Insights

The box plots highlight how differently forest area and freshwater withdrawals behave over time.

Forest Area:

The distribution is tight and uniform, with values clustered between 14% and 15%. This confirms that forest cover has been relatively stable, showing little variation across the years.

Freshwater Withdrawals:

The distribution here is much wider, stretching from around 29% to well over 40%. The presence of outliers at the higher end reflects years where withdrawals spiked dramatically. This spread shows that water use has been far less consistent, with big fluctuations compared to the steady forest data.

In short, the box plots confirm the earlier trend: forest area has remained steady, while water withdrawals have been volatile and increasingly stressed over time.6. Conclusion and Recommendations

6. Recommendations and Further Research

6.1Conclusion

The combined analysis of forest cover and freshwater withdrawals in South Africa highlights two contrasting dynamics.

Forest Area: Forest cover has remained remarkably stable, showing only a slow decline from about 15% in 1990 to roughly 14% in 2020. This consistency is further supported by the box plot, which reveals a tight distribution with minimal variability.

Freshwater Withdrawals: In contrast, freshwater withdrawals have experienced a sharp and uneven increase, particularly after 2013, reaching almost 47% by 2020. The box plot shows a wide

distribution with outliers, confirming high variability and significant spikes in certain years.

Correlation Patterns: The scatter plot suggests a weak negative relationship between forest area and water withdrawals. While declines in forest cover coincide slightly with rising withdrawals, the weak correlation points to other drivers—such as agriculture, industrial demand, population growth, and climate change—as the dominant factors shaping water stress.

Together, these findings show that while forests remain crucial for ecosystem health, South Africa's freshwater challenges are primarily linked to human demand and economic pressures rather than forest loss alone.

6.2 Recommendations

Based on these findings, the following steps are recommended:

Policy Focus:

Implement water conservation strategies that directly address overuse.

Invest in efficient irrigation systems for agriculture, which is the largest water consumer.

Strengthen urban water management through recycling, leak reduction, and public awareness campaigns.

Holistic Analysis:

Future research should build a more comprehensive model by including variables such as:

Population growth

Agricultural output and irrigated land area

GDP and industrial activity

Climate data (precipitation, drought frequency, temperature trends)

Spatial Analysis:

Conduct provincial- or regional-level studies to uncover local variations in both forest cover and water withdrawals that national-level data may hide.

6.3 Final Reflection

This study makes it clear that South Africa's forests are stable but shrinking slowly, while water withdrawals are rising sharply and unpredictably. The country's biggest challenge lies not in deforestation but in ensuring sustainable water use. Tackling this issue will require evidence-driven policies, investment in efficiency, and deeper research into the broader factors driving demand. By doing so, South Africa can safeguard both its ecosystems and its people for the future.

7.References

1. The World Bank Group. (2024). *Forest area (% of land area) - South Africa*. Data360. Retrieved from [https://data360.worldbank.org/en/indicator/WB_WDI_AG_LND_FRST_ZS]

2. The World Bank Group. (2024). *Freshwater withdrawals (% of internal resources) - South Africa*. Data360. Retrieved from [https://data360.worldbank.org/en/indicator/WB_WDI_ER_H2O_FWTL_ZS]

3. McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference.

4. Hunter, J. D. (2007). *Matplotlib: A 2D graphics environment*. Computing in Science & Engineering, 9(3), 90–95.