Question 1 – Comparing ANPP in Managed vs. Unmanaged Forests

Objective

This analysis aims to determine whether forest management has a statistically significant impact on forest productivity, as measured by Annual Net Primary Production (ANPP). To do this, we used a non-parametric statistical approach to compare log-transformed ANPP values between managed (M) and unmanaged (N) forest sites.

Data Preparation and Filtering

The research team imported the dataset from the 'annual_production' sheet and filtered it to retain only forest sites (i.e., BIOME_TYPE == 'F'). The research team excluded sites with missing ANPP values from the analysis. The ANPP column was converted from a character to a numeric format to ensure compatibility with statistical functions. The research team applied a natural logarithm transformation (log ANPP) to the ANPP variable to improve data normality and stabilise variance across groups.

Statistical Method

Given the non-normal distribution of ANPP, the research team applied a Kruskal-Wallis rank sum test to compare the two independent groups (Managed vs. Unmanaged). Unmanaged forests). This test is a robust, non-parametric alternative to one-way ANOVA.

kruskal_test <- kruskal.test(logANPP ~ MANAGEMENT, data = forest_data)

Results

- Managed forests (M): 62 sites

- Unmanaged forests (N): 98 sites

- Kruskal-Wallis statistic (χ^2): 3.81

- Degrees of freedom (df): 1

- p-value: 0.051

Interpretation

At the conventional significance level of $\alpha = 0.05$, the p-value of 0.051 is slightly above the threshold, indicating no statistically significant difference in logANPP between the two groups. However, the result is marginal and suggests a possible trend that may become significant with a larger dataset or additional control for covariates.

Descriptive Statistics

Management Type	Count	Median (logANPP)	IQR	Min	Max
Managed (M)	62	6.42	0.62	4.76	7.62
Unmanaged (N)	98	6.30	0.85	4.12	7.36

The median logANPP was slightly higher in managed forests, and the interquartile range (IQR) was narrower, indicating more consistent productivity among managed sites.

Boxplot Visualization

The research team generated a boxplot to visually compare the distribution of logANPP between the two forest management groups. The figure illustrates a modest difference in central tendency and spread between groups, consistent with the statistical results.

Conclusion

While the Kruskal-Wallis test did not yield a statistically significant result at $p < 0.05$, the analysis indicates a potential effect of forest management on productivity. Managed forests may exhibit slightly higher and more stable above-ground net primary productivity (ANPP) levels. To validate this trend, we recommend further investigations with larger datasets, the inclusion of environmental covariates (e.g., temperature, soil type), and the use of multivariate statistical approaches.

Comparison of logANPP Between Managed and Unmanaged Forests
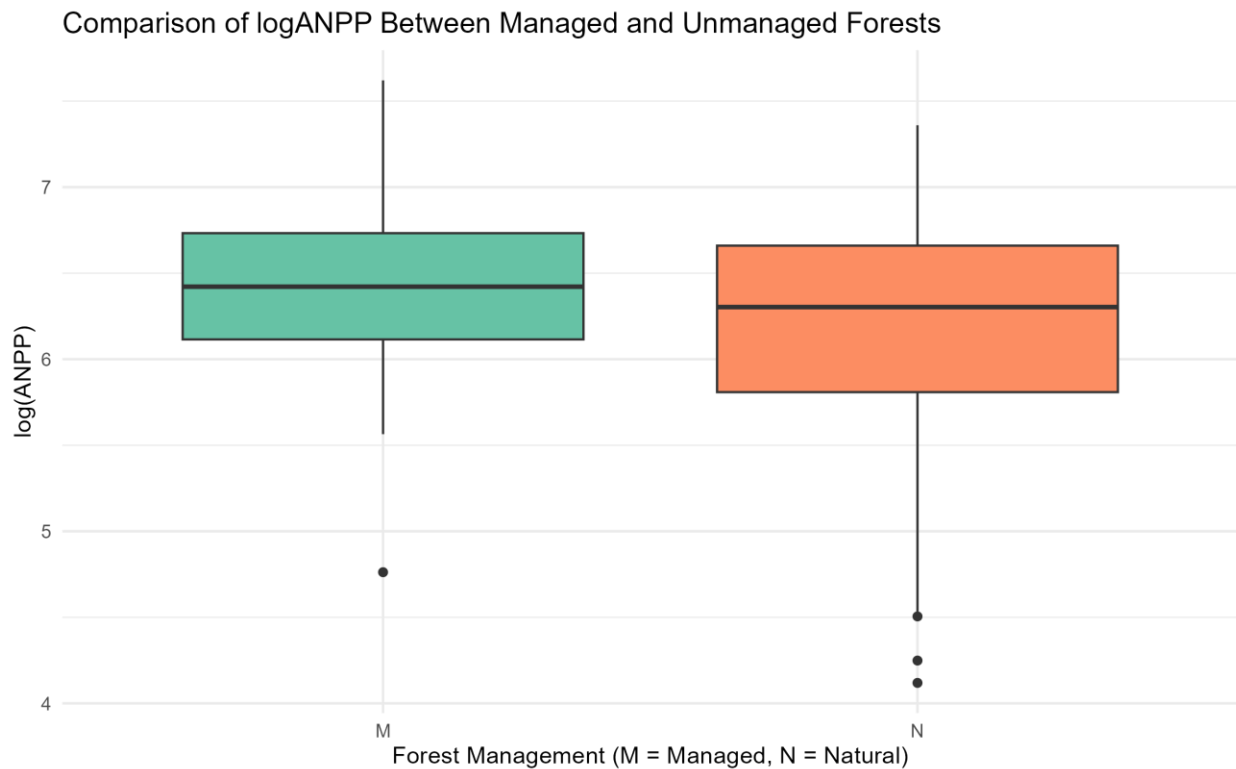
Question 2 – Mapping European Forest Sites: Managed vs. Unmanaged

Objective

The goal of this analysis was to visualize the spatial distribution of forest research sites across Europe and to distinguish visually between managed and unmanaged forests using map-based representation. This task aimed to combine spatial filtering, geospatial data transformation, and professional-level cartographic visualization.

Data Preparation and Filtering

The research team first filtered the dataset from the 'annual_production' sheet to retain only forest sites (BIOME_TYPE == 'F'). To focus the analysis on Europe, the team used geographic coordinates to extract the relevant subset.

- Latitude between 35° and 70° N

- Longitude between −25° and 40° E

Example R code:

```
europe_data <- forest_data %>%

  filter(LATITUDE >= 35, LATITUDE <= 70,

      LONGITUDE >= -25, LONGITUDE <= 40)
```

Visualization Approaches

1. Basic Scatter Map (Exploratory Visualization)

The research team created an initial exploratory plot using ggplot2, mapping longitude versus latitude and applying colour codes to differentiate sites by management status.

- Green: Managed forests (M)

- Blue: Unmanaged forests (N)

The map allowed us to efficiently validate the spatial filtering and confirm the correct geographic location of all sites within Europe.

2. Final Map with European Basemap

The research team employed geospatial tools to generate the final map:

- SF: for transforming site data into spatial objects (SF objects)

- rnaturalearth: to add a country-level base map of Europe

- ggplot2: for layered plotting with colour distinction

The research team plotted each site on the European base map, applying management type as a categorical visual indicator.

Example R code:

```
world <- ne_countries(scale = "medium", returnclass = "sf")

Europe <- world %>% filter(region_un == "Europe")

europe_sites_sf <- st_as_sf(europe_data, coords = c("LONGITUDE", "LATITUDE"), crs =
4326)
```
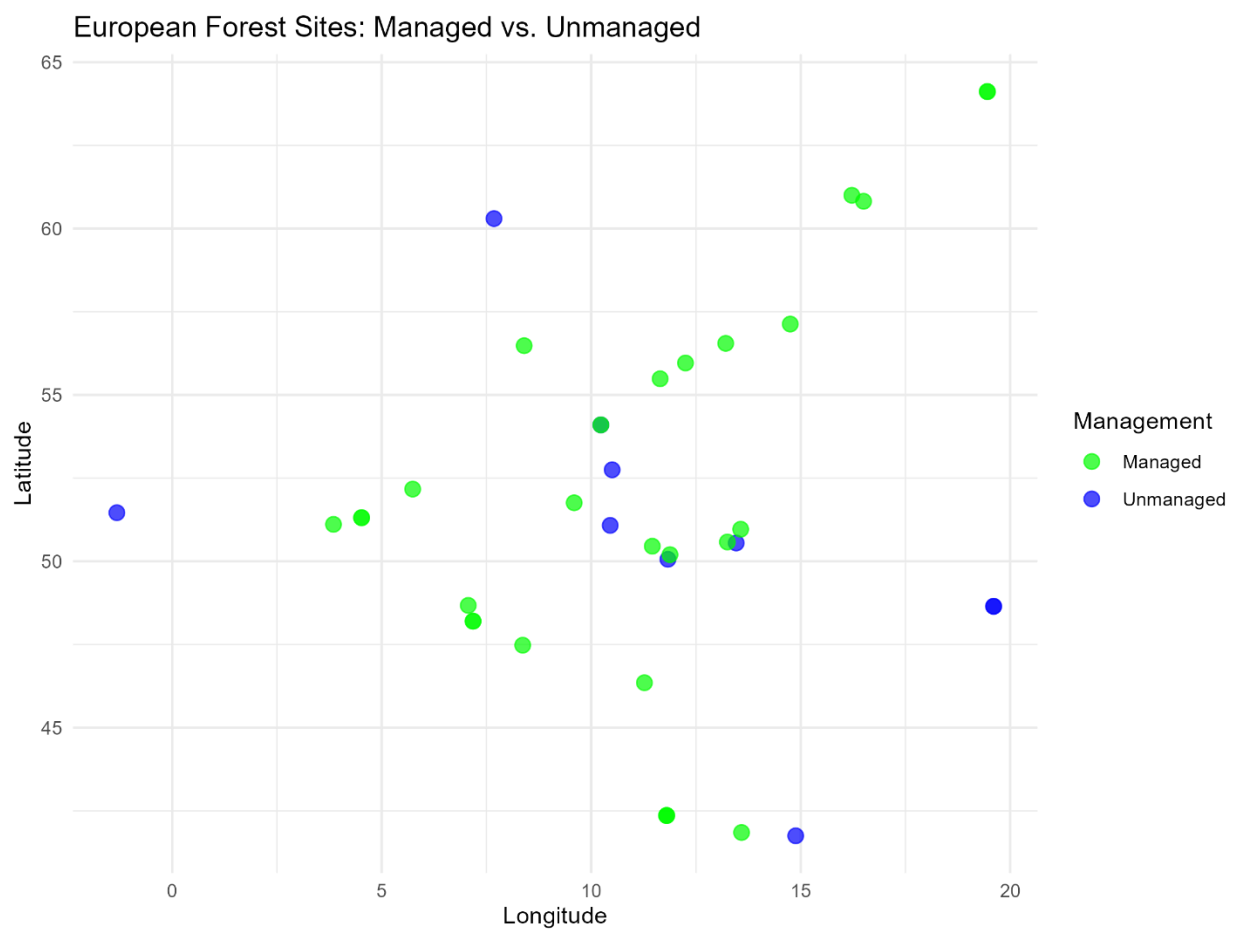
The final map offers several insights:

The study revealed a wide geographic spread of both managed and unmanaged sites across Europe, from west to east.

- Certain regions (e.g., Central Europe) appear to have a higher density of managed forest sites, possibly due to intensified forest policy or long-standing silvicultural practices.

- In contrast, some northern or eastern regions show a relatively higher proportion of natural/unmanaged sites, indicating potential conservation areas or low-intensity use.

- **Europe_forest_map.png:**

European Forest Sites: Managed vs. Unmanaged
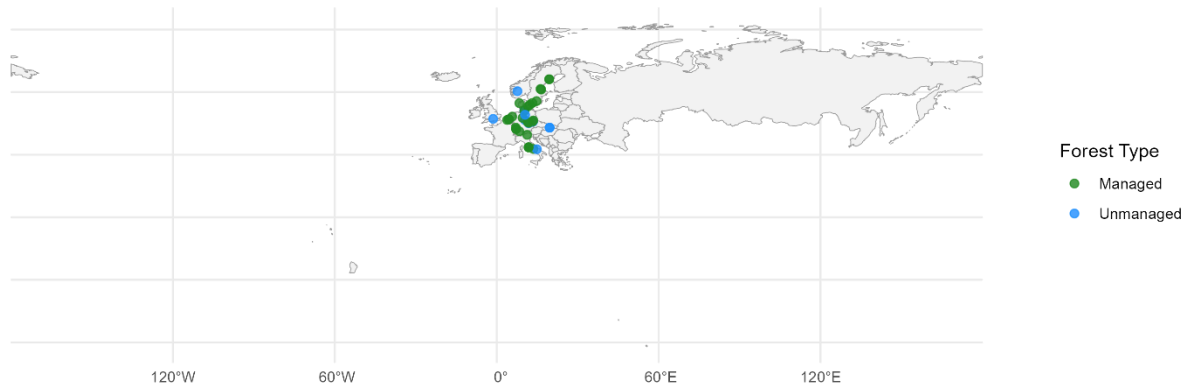
Question 3 Report: Estimating Budburst Timing Based on Growing Degree-Days (GDD)

Objective

The goal of this analysis is to evaluate the heat accumulation required for tree budburst using the concept of Growing degree days (GDD). Based on observed data from the BEL.11 site for the years 2018 and 2019, we aim to:

- Calculate the GDD required for budburst in 2018 (known budburst date: April 16).
- Estimate the expected date of budburst in 2019, assuming a constant GDD threshold.
- Compare GDD accumulation in both years through visualization.

Data Description

The dataset includes daily average temperatures recorded for the site BEL.11 over two years (2018 and 2019), with each day represented as Day of Year (DOY) 1 to 366.

Methodology

3.1. Threshold and Concept

- The research team adopted a temperature threshold of 5°C.
- The research team calculates daily heat for each day using the following formula:

Daily Heat = max(Temperature - 5, 0)

- GDD is the cumulative sum of daily heat values from January 1 up to the budburst date.

3.2. GDD Calculation for 2018

- Budburst in 2018 occurred on April 16 (day of year = 106).
- All temperature values up to Day of Year (DOY) 106 were processed.
- After removing invalid values (e.g., NA or >50°C), GDD was calculated as follows:

GDD_2018 = 204.76 °C-days

3.3. Budburst Date Estimation for 2019

- The same method was consistently applied to the 2019 data as well.
- After cleaning outliers (>50°C) and applying the GDD threshold, cumulative GDD was computed.
- The first day in 2019 when the cumulative GDD surpassed the threshold of 204.76 was identified:

Estimated Budburst Date in 2019: March 31 (DOY = 90)

3.4. Interpretation

- The tree in 2019 reached the required heat accumulation approximately **16 days earlier** than in 2018.
- The advancement of GDD accumulation in 2019 indicates a warmer late winter and early spring, which likely contributed to accelerated phenological development.
- This interannual variation is important for understanding climate sensitivity in temperate forest ecosystems.

4. Visualization

The research team plotted a cumulative Growing Degree Day (GDD) graph for both years to illustrate the differences in heat accumulation patterns over time.

- **2018** line rises steadily, reaching 204.76 by mid-April.
- **2019** accumulates GDD faster, surpassing the threshold by the end of March.
- A dashed horizontal line indicates the GDD threshold for visual reference.

Conclusion

This GDD-based approach enables effective prediction of phenological events such as budburst. The results illustrate how relatively small shifts in daily temperatures can significantly alter plant development timelines. Such findings are valuable for climate impact assessments, forest management, and ecological forecasting.

- **Cumulative_GDD_comparison_2018_2019.png:**

Cumulative Growing Degree Days (GDD) for 2018 and 2019
Threshold Temperature = 5°C