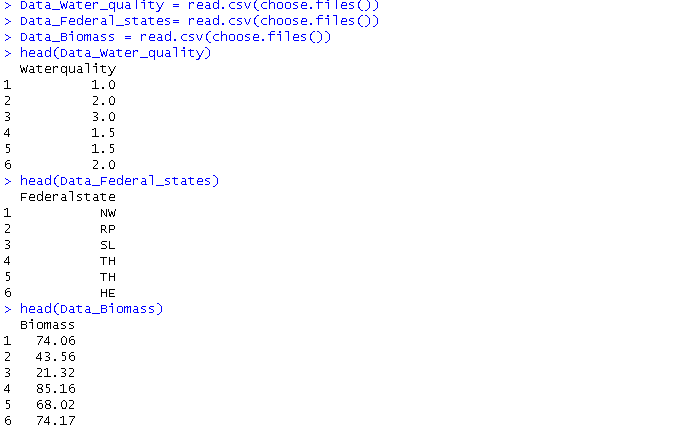
**WATER QUALITY ANALYSIS IN GERMAN RUNNING WATERS**

**STUDENT NAME: [YOUR NAME]  
MATRICULATION NUMBER: [YOUR MATRICULATION NUMBER]  
COURSE CODE: AB1 4005 + SA1 4805  
SEMESTER: WS 24/25  
INSTRUCTOR: RUDOLF  
SUBMISSION DATE: 28/02/2025**

### ****1.0 Introduction****

This report analyzes the water quality, federal state distribution, and biomass of indicator species larvae in 60 German monitoring sites. The objective is to examine statistical analysis and , visualize the datasets provided. The analysis was started with importation of the dataset directly in .txt format by using “*read.csv(choose.files())*” for each datasets and the datasets was preview with “*head.()*”



### ****2.0 Water Quality Analysis****

#### **2.1 Descriptive Statistics**

**Mean Water Quality:** [1.858333]

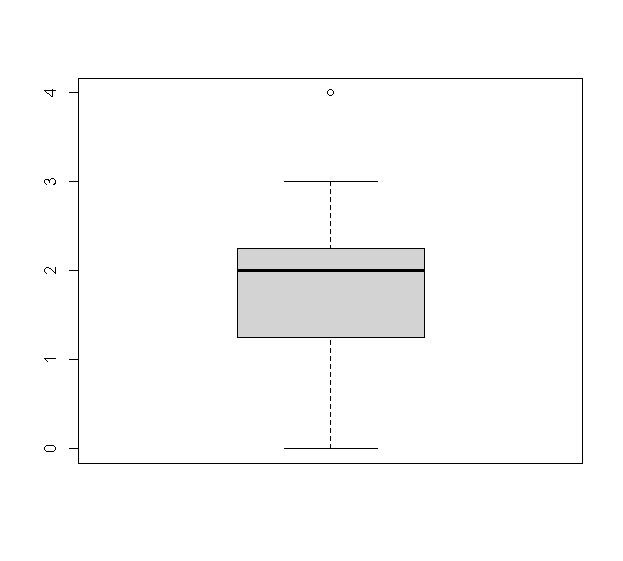
**Median Water Quality:** [2]

**Standard Deviation:** [0.6650472]

**Interquartile Range (IQR):** [0.75]

#### **2.2 Graphical Representation**

#### Figure 1: Boxplot for Water Quality



Comparison with descriptive statistics :

The boxplot shows that the water quality values range from approximately 0 to an outlier at 4, with a median of about 2. The interquartile range (IQR) is 0.75, reflecting a moderate spread around the median, and the standard deviation (approximately 0.67) supports this moderate variability. The boxplot’s center and spread align well with the descriptive statistics: the median (2) sits within the middle 50% of the data (the box), while the mean (about 1.86) is influenced by the right-side tail and the outlier at 4.

### ****3.0 Federal States Analysis****

#### **3.1 Descriptive Statistics**

BW BY HE NW RP SL TH

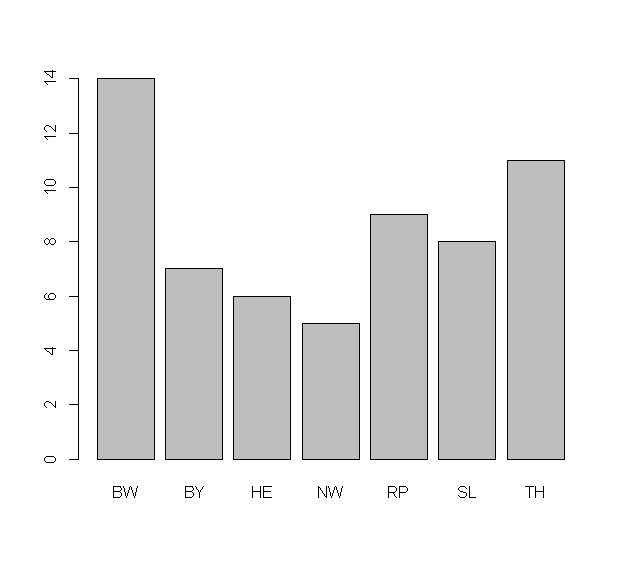
14 7 6 5 9 8 11

**Most Frequent Federal State:** [BW]

**Least Frequent Federal State:** [NW]

#### **3.2 Graphical Representation**

**Figure 2: Bar Chart of Monitoring Sites by Federal State**



**Comparison with descriptive statistics**

The Bar chat further illustrate the Mode and Least occurring number and show how federal state contribute to the variables. This distribution indicates that the monitoring efforts are not evenly spread across the federal states.

### ****4.**0 **Biomass Analysis****

#### **4.1 Descriptive Statistics**

**Mean Biomass:** [46.214]

**Median Biomass:** [43.045]

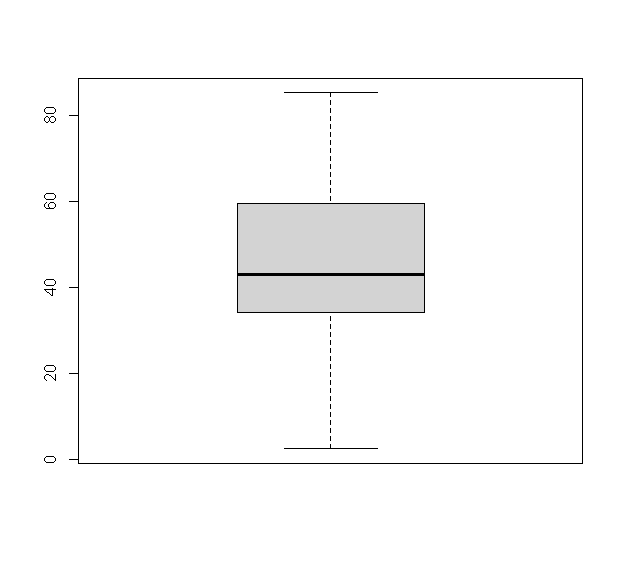
**Range:** [2.44 to 85.16]

**IQR:** [24.5175]

**Standard Deviation:** [18.88334]

#### **4.2 Boxplot of Biomass**

Figure 3: Boxplot of Biomass



Comparison with descriptive statistics:

The boxplot in Figure 3 visually represents the distribution of biomass values, aligning with the descriptive statistics provided. The median biomass value (43.045) is positioned near the center of the interquartile range (IQR), which is 24.5175, indicating moderate variability.The boxplot provides a clear visual summary of the biomass distribution, complementing the numerical statistics.

#### **4.3 Histogram and Cumulative Distribution**

Figure 4: Histogram of Biomass Distribution

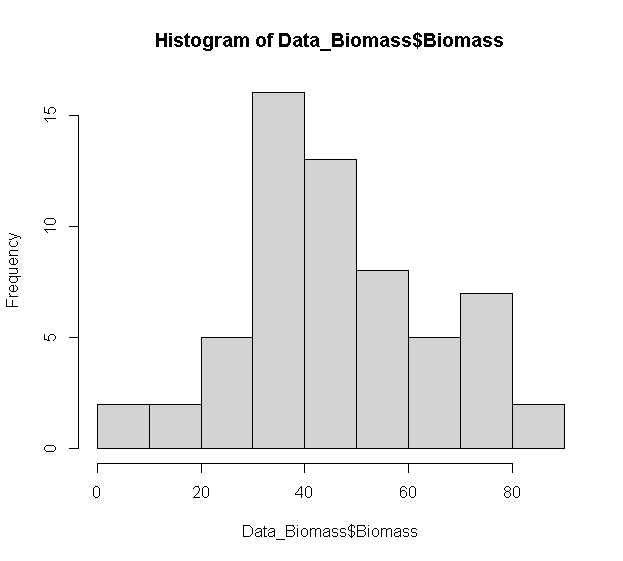
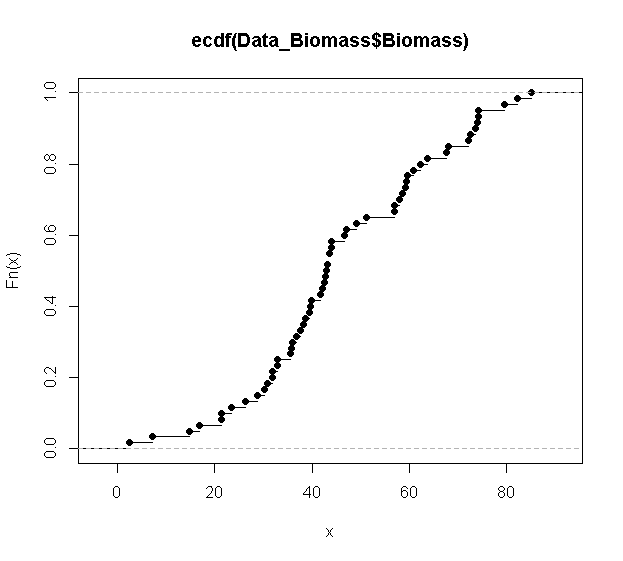


Figure 5: Cumulative Biomass Distribution  


Analysis of skewness

The datasets tends to be normal not really skewed as the length f lower and upper whisk of the boxplot tends to be equal, which can be confirmed by performing a normal test on the datasets

#### **4.4 Additional Biomass Analysis**

**95th Percentile Biomass:** [74.4915]

**Percentage of Values Below 48g:** [61.66667]

### ****5. 0 Analysis Strategy****

#### **5.1 Checking if Federal States Manipulate Water Quality Data**

To really get if the federal state influences the water quality classification, a **Chi-Square test** will be used.

**Checking if Biomass is Linked to Water Quality**

To determine the relationship between the Biomass and Water Quality, a correlational analysis will be used, regression model will established the relationship and can be well be visualized by scatter plots to observe trends.

**Analyzing Biomass Changes Over Time**

To assess biomass variation over three years, normal test will be performed on the biomass and either **Paired t-test** (for normal distribution) or **Wilcoxon signed-rank test** (for non-normal data) will be used and **line charts** to visualize it.