Exploring the CO2 Dataset

Ken Wood

2024-10-06

Objective:

In this assignment, you will apply basic statistical analysis techniques using the R programming language on the CO2 dataset (make sure you use the CO2 dataset and not the co2 dataset, they are different). This dataset details CO2 uptake in grass plants under different environmental conditions. Your tasks will include data exploration, visualization, hypothesis testing with a t-test, and examining correlations.

Data Overview:

The CO2 dataset contains observations from an experiment on the cold tolerance of the grass species Echinochloa crus-galli. Variables include treatment types, CO2 uptake, concentration, temperature, and more. This rich dataset allows for comprehensive statistical analysis.

Instructions:

In your own R script file, please complete the following tasks:

- Data Exploration: Familiarize yourself with the dataset using the head() and summary() functions in R.

```
# Load the datasets package (usually not necessary as it's loaded by default)
library(datasets)

# Import the CO2 dataset
data(CO2)

# Display the first few rows of the dataset
head(CO2)
```

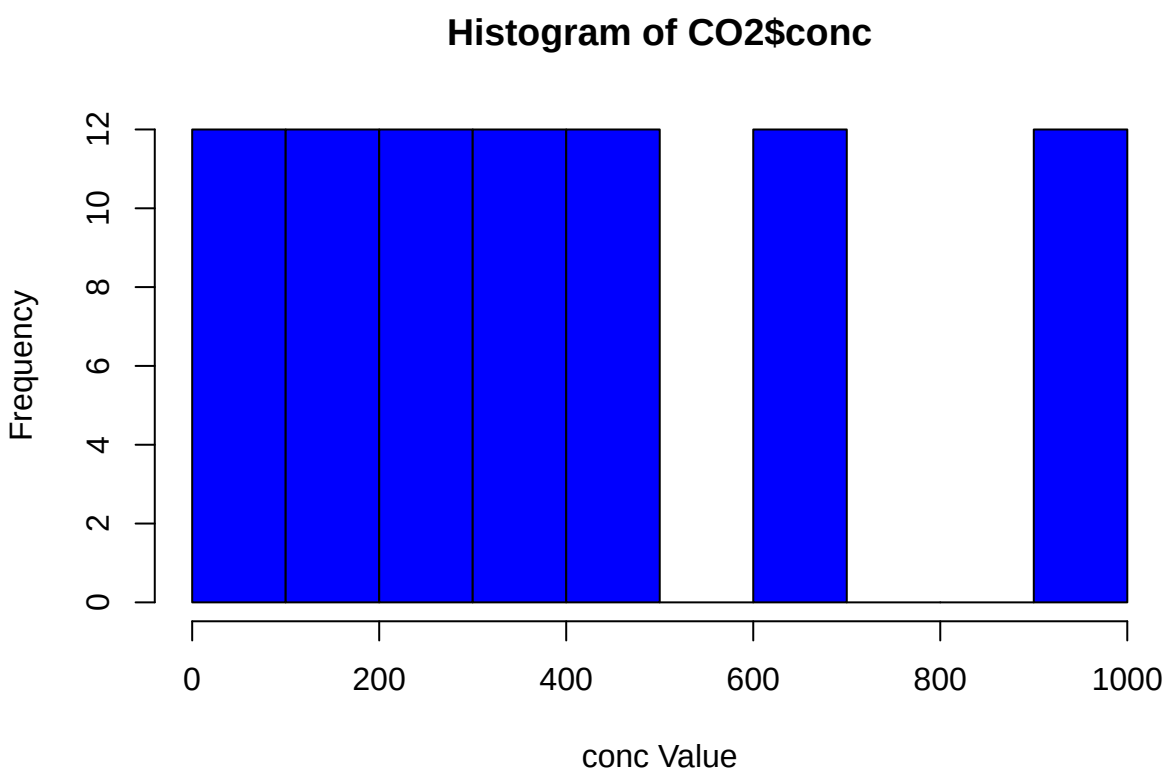
```
##   Plant   Type  Treatment conc uptake
## 1   Qn1 Quebec nonchilled   95   16.0
## 2   Qn1 Quebec nonchilled  175   30.4
## 3   Qn1 Quebec nonchilled  250   34.8
## 4   Qn1 Quebec nonchilled  350   37.2
## 5   Qn1 Quebec nonchilled  500   35.3
## 6   Qn1 Quebec nonchilled  675   39.2
```

```
# Get a set of summary stats for the dataset
summary(CO2)
```

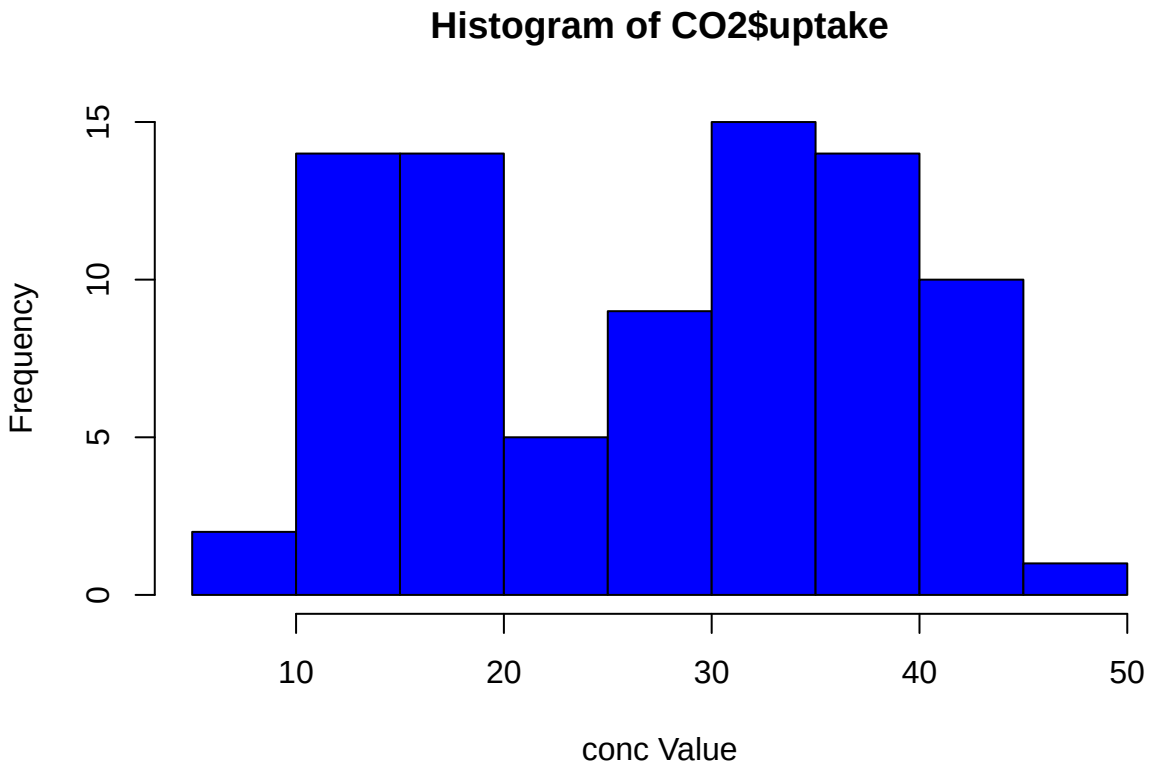
```
##      Plant             Type         Treatment       conc          uptake
##  Qn1    : 7   Quebec     :42   nonchilled:42   Min.   :  95   Min.   : 7.70
##  Qn2    : 7   Mississippi:42   chilled   :42   1st Qu.: 175   1st Qu.:17.90
##  Qn3    : 7                                    Median : 350   Median :28.30
##  Qc1    : 7                                    Mean   : 435   Mean   :27.21
##  Qc3    : 7                                    3rd Qu.: 675   3rd Qu.:37.12
##  Qc2    : 7                                    Max.   :1000   Max.   :45.50
##  (Other):42
```

Data Visualization: Create visualizations to understand the distributions and relationships in the data:

```
# Create a histogram
hist(CO2$conc,
     main="Histogram of CO2$conc",
     xlab="conc Value",
     ylab="Frequency",
     col="blue",
     border="black")
```




```
# Create a histogram
hist(CO2$uptake,
     main="Histogram of CO2$uptake",
     xlab="conc Value",
     ylab="Frequency",
     col="blue",
     border="black")
```



t-Test: Conduct a t-test to compare the mean CO2 uptake between two treatment groups. Clearly state your hypothesis, perform the test, and interpret the results. Correlation Analysis: Calculate and interpret the correlation coefficients between CO2 uptake and other numeric variables in the dataset.

```
t_test_results <- t.test(uptake ~ Treatment, data = CO2,
                         var.equal = FALSE)
print(t_test_results)
```

```
##
##  Welch Two Sample t-test
##
## data:  uptake by Treatment
## t = 3.0485, df = 80.945, p-value = 0.003107
## alternative hypothesis: true difference in means between group nonchilled and group chilled is not e
## 95 percent confidence interval:
##   2.382366 11.336682
## sample estimates:
## mean in group nonchilled    mean in group chilled
##                 30.64286                 23.78333
```

```
cor(CO2$uptake,CO2$conc)
```

```
## [1] 0.4851774
```