About the transformation - outliers\_boxplot: identifies outliers by the boxplot rule (Q1 - 1.5·IQR, Q3 + 1.5·IQR) and can remove them from numeric attributes.

# NA and Outlier analysis  
  
# installation   
#install.packages("daltoolbox")  
  
# loading DAL  
library(daltoolbox)

Remove outliers via boxplot and inspect the result.

# Outlier removal using boxplot rule  
  
# The class uses the boxplot rule to define outliers.  
  
# An outlier is a value smaller than $Q\_1 - 1.5\cdot IQR$ or larger than $Q\_3 + 1.5\cdot IQR$.  
   
# The class removes outliers in numeric attributes.  
  
# Removing outliers from a data frame  
  
# Example outlier removal code  
out\_obj <- outliers\_boxplot() # outlier analysis class  
out\_obj <- fit(out\_obj, iris) # computes limits via quartiles and IQR  
iris.clean <- transform(out\_obj, iris) # returns cleaned dataset  
  
# inspection of cleaned dataset  
head(iris.clean)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

nrow(iris.clean)

## [1] 146

Visualize which rows were flagged as outliers.

# Visualizing the actual outliers  
  
idx <- attr(iris.clean, "idx")  
print(table(idx))

## idx  
## FALSE TRUE   
## 146 4

iris.outliers <- iris[idx,]  
head(iris.outliers)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 16 5.7 4.4 1.5 0.4 setosa  
## 33 5.2 4.1 1.5 0.1 setosa  
## 34 5.5 4.2 1.4 0.2 setosa  
## 61 5.0 2.0 3.5 1.0 versicolor

References - Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley.