About the transformation - outliers\_gaussian: flags as outliers values beyond mean ± 3 standard deviations, assuming approximately normal distribution.

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

Remove outliers using the 3-sigma rule and inspect.

# Outlier removal using Gaussian rule (3σ)  
# An outlier is a value smaller than $\overline{x} - 3\,\sigma\_x$ or larger than $\overline{x} + 3\,\sigma\_x$.  
  
# The class removes outliers in numeric attributes.  
  
# Removing outliers from a data frame  
  
# Example outlier removal code  
out\_obj <- outliers\_gaussian() # outlier analysis class  
out\_obj <- fit(out\_obj, iris) # computes limits based on mean and std dev  
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] 149

List observations identified as outliers.

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

## idx  
## FALSE TRUE   
## 149 1

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

References - Pukelsheim, F. (1994). The Three Sigma Rule. The American Statistician 48(2):88–91.