

```

#Scale Data
library(caret)
data("iris")
summary(iris[,1:4])
preprocessParams=preProcess(iris[,1:4], method=c("scale"))
print(preprocessParams)
transformed=predict(preprocessParams, iris[,1:4])
summary(transformed)

#Scale data #2
preprocessParams=scale(iris[,1:4])

#Center Data
library(caret)
data(iris)
summary(iris[,1:4])
preprocessParams=preProcess(iris[,1:4], method=c("center"))
print(preprocessParams)
summary(transformed)

#Standardize Data
library(caret)
data(iris)
summary(iris[,1:4])
preprocessParams=preProcess(iris[,1:4], method=c("center","scale"))
print(preprocessParams)
transformed=predict(preprocessParams, iris[,1:4])
summary(transformed)

#Min-Max Standardization for quantitative attributes
myd=read.table("census_small.csv", header = T, sep = ',')
myd$agemm=(myd$age-min(myd$age))/(max(myd$age)-min(myd$age))

#Normalize Data
library(caret)
data("iris")
summary(iris[,1:4])
preprocessParams=preProcess(iris[,1:4], method=c("range"))
print(preprocessParams)
transformed=predict(preprocessParams, iris[,1:4])
summary(transformed)
#Normalize Data 2
myd$average_monthly_hours=scale(myd$average_monthly_hours, center =
TRUE, scale=TRUE)

#Box-Cox Transform
library(mlbench)
library(caret)
data("PimaIndiansDiabetes")
summary(PimaIndiansDiabetes[,7:8])

```

```

preprocessParams=preProcess(PimaIndiansDiabetes[,7:8],
method="BoxCox")
transformed=predict(preprocessParams, PimaIndiansDiabetes[,7:8])
summary(transformed)

#Yeo-Johnson Transform
library(mlbench)
library(caret)
data("PimaIndiansDiabetes")
summary(PimaIndiansDiabetes[,7:8])
preprocessParams=preProcess(PimaIndiansDiabetes[,7:8],
method=c("YeoJohnson"))
print(preprocessParams)
transformed=predict(preprocessParams, PimaIndiansDiabetes[,7:8])
summary(transformed)

#Principal Component Anlaysis Transform
library(mlbench)
library(caret)
data(iris)
summary(iris)
preprocessParams=preProcess(iris, method=c("center","scale","pca"))
print(preprocessParams)
transformed=predict(preprocessParams, iris)
summary(transformed)

#Independent Component Analysis Transform
library(mlbench)
library(caret)
library(fastICA)
data("PimaIndiansDiabetes")
summary(PimaIndiansDiabetes[,1:8])
preprocessParams=preProcess(PimaIndiansDiabetes[,1:8], method =
c("center","scale","ica"), n.comp=5)
print(preprocessParams)
transformed=predict(preprocessParams, PimaIndiansDiabetes[,1:8])
summary(transformed)

#Replace data
myd=read.table("census_small.csv", header=T, sep=',')
myd[myd==98]=NA
myd$degree[myd$degree==8]=NA

#Label cateogircal variables
myd=read.table("census_small.csv", header = T, sep = ',')
myd$sex=factor(myd$sex, levels = c(1,2), labels = c("male","female"))

#Binning variables
par(mfrow=c(1,1))
bin=4

```

```
cutpoints=quantile(myd$age, (0:bin)/bin, na.rm = T)
myd$agebin=cut(myd$age, cutpoints, include.lowest = TRUE)
counts=table(myd$agebin)
barplot(counts, main="Bar Chart of binned age (equal size bins)")
counts=table(myd$agebinw)
barplot(counts, main = "Bar Chart of Binned age (equal width bins)")

#Remove Columns
myd$column_name=NULL
rm(myd$column_name)
```