gomez2

Andres Gomez

2023-01-23

Breast Cancer Wisconsin Diagnostic Dataset

from UCI Machine Learning Repository

In this report, we provide summary statistics and visualizations for the brca data set that is located within the R base Package. The brca data set represents biopsy features for the classification between malignant (M) and benign (B) of breast masses. It includes data for the mean, standard error (se), and worst value of 10 nuclear measurements.

The following sample presents the first 6 rows of the data (x) along with its predictors (y). Said predictors are divided into benign (B) and malignant (M):

Data from X (): This data contains variables such as radius, texture, perimeter, area, smoothness, compactness, concavity, number of concave portions (concave_pts), symmetry, and the fractal dimension of the nucleus (fractal_dim):

```
Br <- as.data.frame(brca)
summX <- Br %>% select(x.radius_mean, x.texture_mean, x.perimeter_mean, x.area_mean, x.smoothness_mean,
head(summX)
```

```
##   x.radius_mean x.texture_mean x.perimeter_mean x.area_mean x.smoothness_mean
## 1        13.540          14.36            87.46       566.3           0.09779
## 2        13.080          15.71            85.63       520.0           0.10750
## 3         9.504          12.44            60.34       273.9           0.10240
## 4        13.030          18.42            82.61       523.8           0.08983
## 5         8.196          16.84            51.71       201.9           0.08600
## 6        12.050          14.63            78.04       449.3           0.10310
##   x.compactness_mean x.concavity_mean x.concave_pts_mean x.symmetry_mean
## 1            0.08129          0.06664           0.047810          0.1885
## 2            0.12700          0.04568           0.031100          0.1967
## 3            0.06492          0.02956           0.020760          0.1815
## 4            0.03766          0.02562           0.029230          0.1467
## 5            0.05943          0.01588           0.005917          0.1769
## 6            0.09092          0.06592           0.027490          0.1675
##   x.fractal_dim_mean
## 1            0.05766
## 2            0.06811
## 3            0.06905
## 4            0.05863
## 5            0.06503
## 6            0.06043
```

```
#head(Br$y)
```

Note that the previous summary only presents the mean of the 10 nuclear measurements as a summary for the data you may encounter in the brca data set.

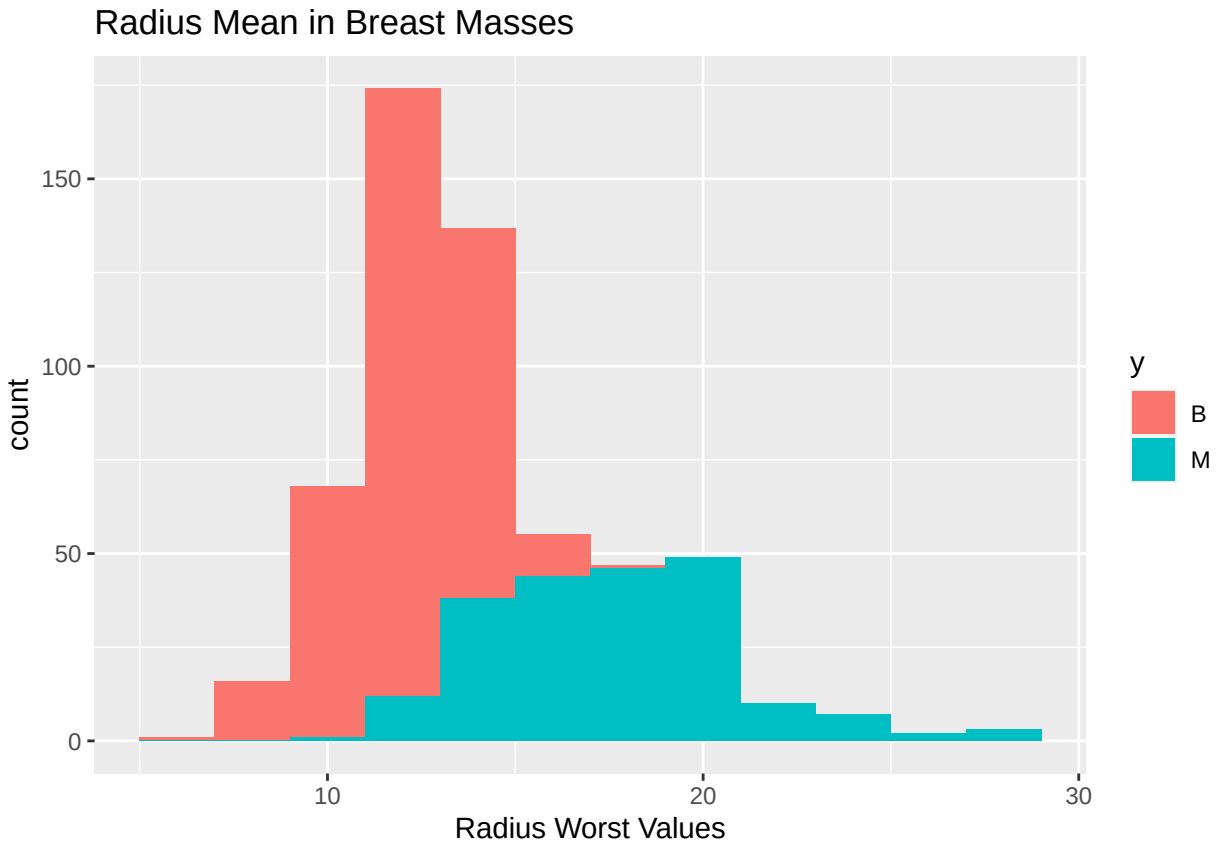
####Data from y The y data, inside the brca data set, contains the outcomes. It contains whether the mass is malignant (M) or benign (B):

```
## [1] B B B B B B
## Levels: B M
```

Including Plots

The variables that will be considered, from the data set, are the radius mean, compactness mean, texture_mean, and the concavity mean. These variables are though to be the most important predictors

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.