First Project R Markdown

2024-11-17

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
## package 'mlbench' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\trive\AppData\Local\Temp\RtmpSOTQfa\downloaded_packages
```

```
## Warning: package 'mlbench' was built under R version 4.4.2
```

```
##          Id Cl.thickness Cell.size Cell.shape Marg.adhesion Epith.c.size
## 1 1000025            5         1          1             1            2
## 2 1002945            5         4          4             5            7
## 3 1015425            3         1          1             1            2
## 4 1016277            6         8          8             1            3
## 5 1017023            4         1          1             3            2
## 6 1017122            8        10         10             8            7
##   Bare.nuclei Bl.cromatin Normal.nucleoli Mitoses     Class
## 1           1           3               1       1    benign
## 2          10           3               2       1    benign
## 3           2           3               1       1    benign
## 4           4           3               7       1    benign
## 5           1           3               1       1    benign
## 6          10           9               7       1 malignant
```

```
## 'data.frame':    699 obs. of  11 variables:
##  $ Id             : chr  "1000025" "1002945" "1015425" "1016277" ...
##  $ Cl.thickness   : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 5 5 3 6 4 8 1 2 2 4 ...
##  $ Cell.size      : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 4 1 8 1 10 1 1 1 2 ...
##  $ Cell.shape     : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 4 1 8 1 10 1 2 1 1 ...
##  $ Marg.adhesion  : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 1 5 1 1 3 8 1 1 1 1 ...
##  $ Epith.c.size   : Ord.factor w/ 10 levels "1"<"2"<"3"<"4"<..: 2 7 2 3 2 7 2 2 2 2 ...
##  $ Bare.nuclei    : Factor w/ 10 levels "1","2","3","4",..: 1 10 2 4 1 10 10 1 1 1 ...
##  $ Bl.cromatin    : Factor w/ 10 levels "1","2","3","4",..: 3 3 3 3 3 9 3 3 1 2 ...
##  $ Normal.nucleoli: Factor w/ 10 levels "1","2","3","4",..: 1 2 1 7 1 7 1 1 1 1 ...
##  $ Mitoses        : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 5 1 ...
##  $ Class          : Factor w/ 2 levels "benign","malignant": 1 1 1 1 1 2 1 1 1 1 ...
```

```
##        Id              Cl.thickness   Cell.size      Cell.shape   Marg.adhesion
## Length:699          1      :145    1      :384    1      :353    1      :407
## Class :character    5      :130    10     : 67    2      : 59    2      : 58
## Mode  :character    3      :108    3      : 52    10     : 58    3      : 58
##                     4      : 80    2      : 45    3      : 56    10     : 55
##                     10     : 69    4      : 40    4      : 44    4      : 33
##                     2      : 50    5      : 30    5      : 34    8      : 25
##                     (Other):117   (Other): 81    (Other): 95    (Other): 63
##   Epith.c.size   Bare.nuclei     Bl.cromatin    Normal.nucleoli    Mitoses
## 2      :386    1      :402    2      :166    1      :443    1      :579
## 3      : 72    10     :132    3      :165    10     : 61    2      : 35
## 4      : 48    2      : 30    1      :152    3      : 44    3      : 33
## 1      : 47    5      : 30    7      : 73    2      : 36    10     : 14
## 6      : 41    3      : 28    4      : 40    8      : 24    4      : 12
## 5      : 39    (Other): 61    5      : 34    6      : 22    7      :  9
## (Other): 66    NA's   : 16    (Other): 69    (Other): 69    (Other): 17
##       Class
## benign    :458
## malignant:241
##
##
##
##
##
```
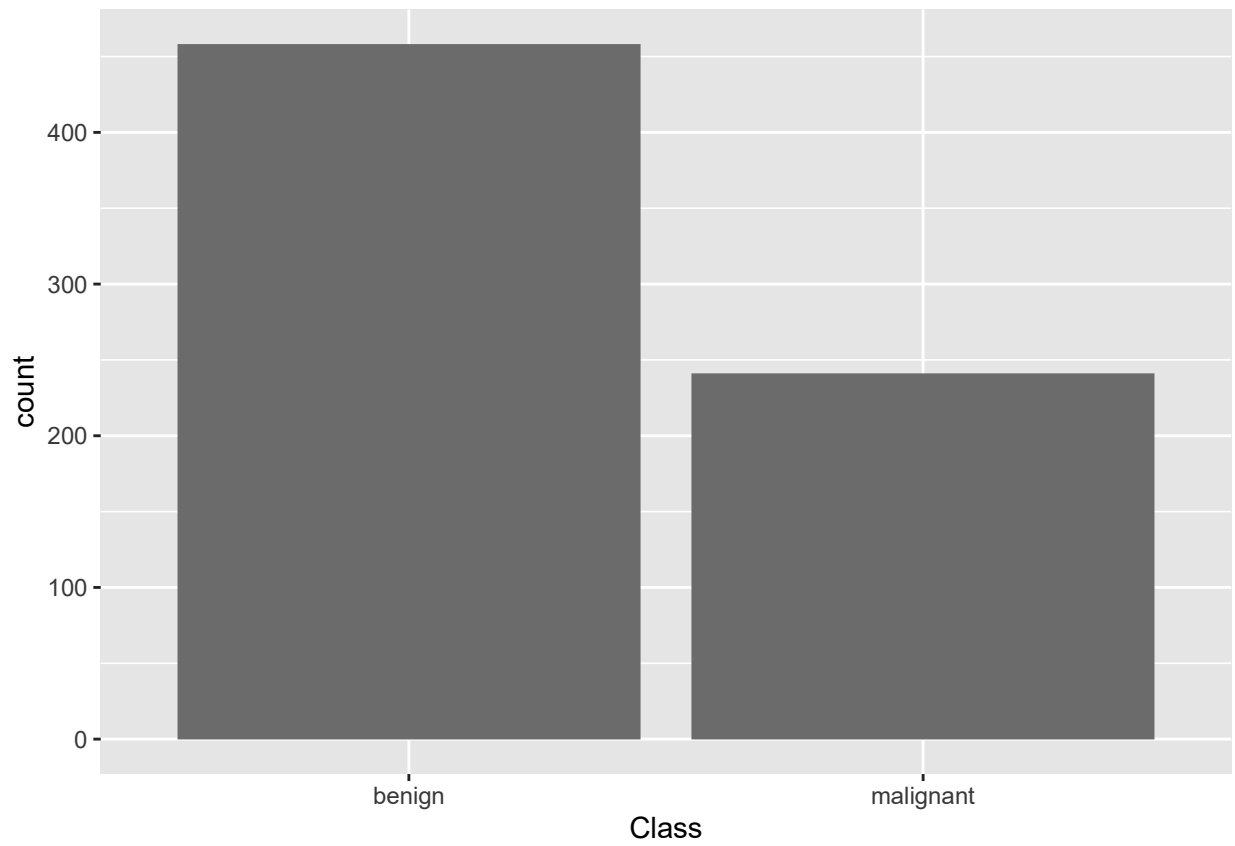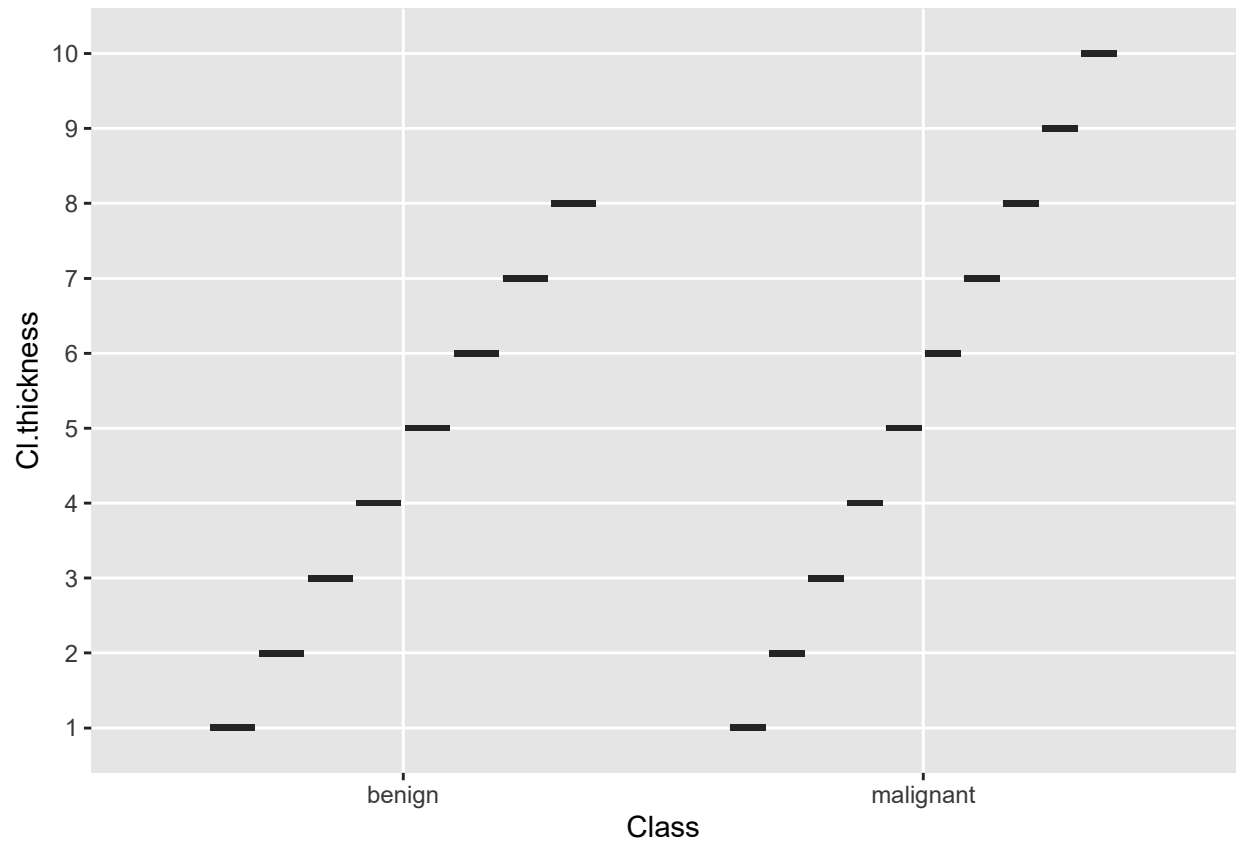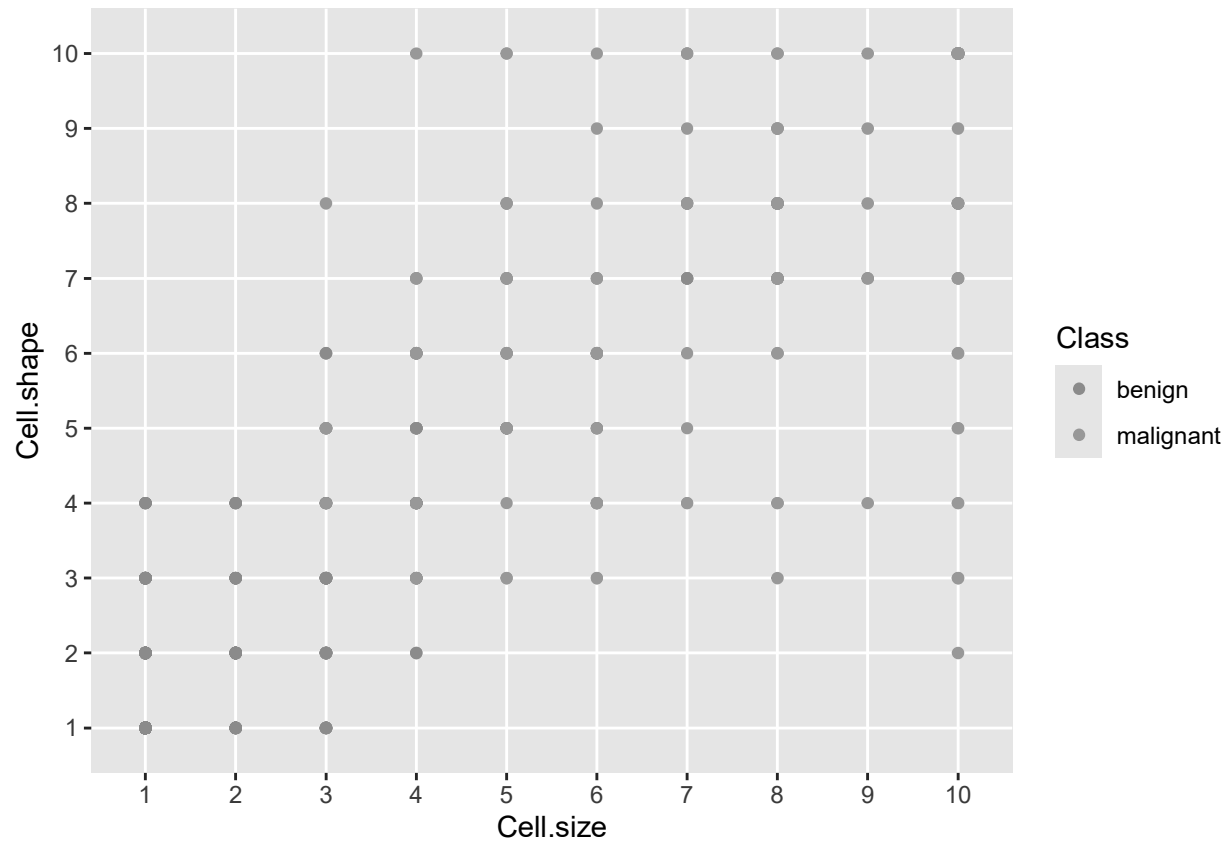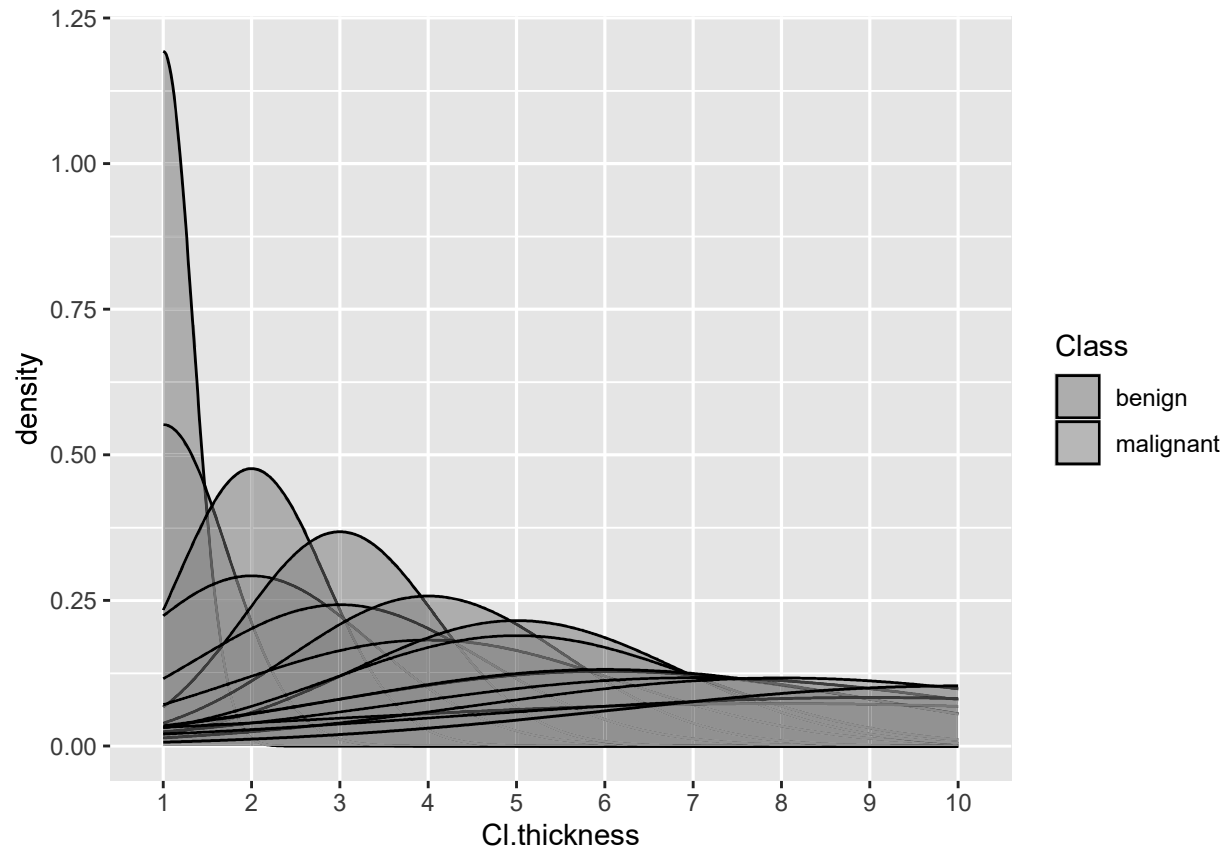
```
## Warning: Groups with fewer than two data points have been dropped.

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```

```
## package 'rpart' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'rpart'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\trive\Downloads\R-4.4.1\library\00LOCK\rpart\libs\x64\rpart.dll to
## C:\Users\trive\Downloads\R-4.4.1\library\rpart\libs\x64\rpart.dll: Permission
## denied

## Warning: restored 'rpart'

##
## The downloaded binary packages are in
##   C:\Users\trive\AppData\Local\Temp\RtmpSOTQfa\downloaded_packages

## Warning: package 'rpart' was built under R version 4.4.2
```
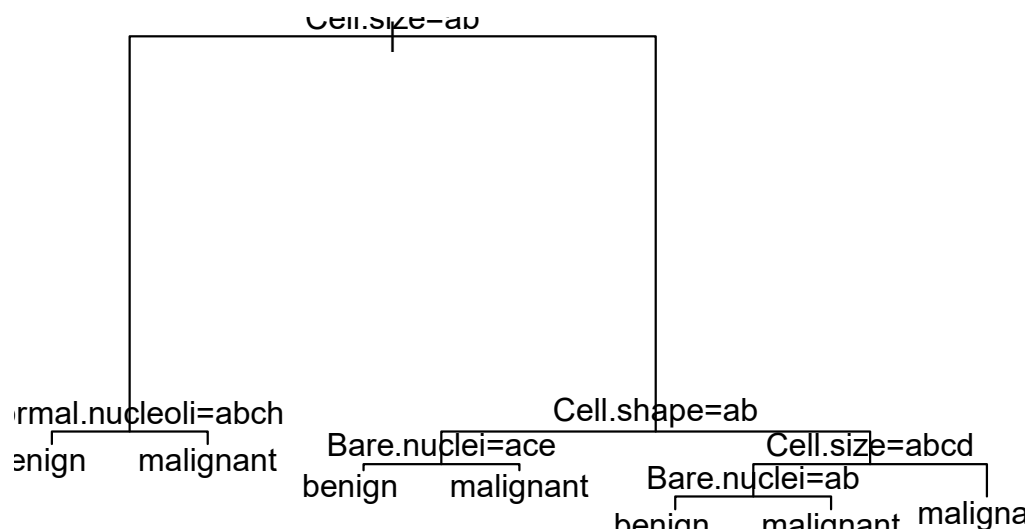
Cell.size=ab

rmal.nucleoli=abch
benign   malignant

Cell.shape=ab
Bare.nuclei=ace
benign   malignant

Cell.size=abcd
Bare.nuclei=ab
benign   malignant   maligna

```
## Warning: package 'caret' was built under R version 4.4.2

## Loading required package: lattice

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  benign malignant
##   benign        442         9
##   malignant      16       232
##
##                Accuracy : 0.9642
##                  95% CI : (0.9477, 0.9767)
##     No Information Rate : 0.6552
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9214
##
##  Mcnemar's Test P-Value : 0.2301
##
##             Sensitivity : 0.9651
##             Specificity : 0.9627
##          Pos Pred Value : 0.9800
##          Neg Pred Value : 0.9355
##              Prevalence : 0.6552
```

```
##          Detection Rate : 0.6323
##    Detection Prevalence : 0.6452
##       Balanced Accuracy : 0.9639
##
##        'Positive' Class : benign
##


## Confusion Matrix and Statistics
##
##            Reference
## Prediction  benign malignant
##   benign       443         3
##   malignant     15       238
##
##                Accuracy : 0.9742
##                  95% CI : (0.9596, 0.9847)
##     No Information Rate : 0.6552
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9437
##
##  Mcnemar's Test P-Value : 0.009522
##
##             Sensitivity : 0.9672
##             Specificity : 0.9876
##          Pos Pred Value : 0.9933
##          Neg Pred Value : 0.9407
##              Prevalence : 0.6552
##          Detection Rate : 0.6338
##    Detection Prevalence : 0.6381
##       Balanced Accuracy : 0.9774
##
##        'Positive' Class : benign
##


## CART
##
## 683 samples
##  10 predictor
##   2 classes: 'benign', 'malignant'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 546, 546, 547, 547, 546
## Resampling results across tuning parameters:
##
##   cp          Accuracy   Kappa
##   0.02092050  0.9385144  0.8645561
##   0.05439331  0.9267497  0.8383238
##   0.79079498  0.8552168  0.6349432
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was cp = 0.0209205.
```

```
## Naive Bayes
##
## 683 samples
##  10 predictor
##   2 classes: 'benign', 'malignant'
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 546, 547, 546, 547, 546
## Resampling results across tuning parameters:
##
##   usekernel  Accuracy   Kappa
##   FALSE      0.3499249  0
##    TRUE      0.6500751  0
##
## Tuning parameter 'laplace' was held constant at a value of 0
## Tuning
##  parameter 'adjust' was held constant at a value of 1
## Accuracy was used to select the optimal model using the largest value.
## The final values used for the model were laplace = 0, usekernel = TRUE
##  and adjust = 1.


##
## Call:
## summary.resamples(object = results)
##
## Models: rpart, naive_bayes
## Number of resamples: 5
##
## Accuracy
##                  Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## rpart       0.9197080 0.919708 0.9264706 0.9385144 0.9558824 0.9708029    0
## naive_bayes 0.6470588 0.649635 0.6496350 0.6500751 0.6496350 0.6544118    0
##
## Kappa
##                  Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## rpart       0.8210426 0.8277517 0.8341059 0.8645561 0.9034091 0.9364711    0
## naive_bayes 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000    0


## Warning: package 'boot' was built under R version 4.4.2


##
## Attaching package: 'boot'


## The following object is masked from 'package:lattice':
##
##     melanoma


##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
```
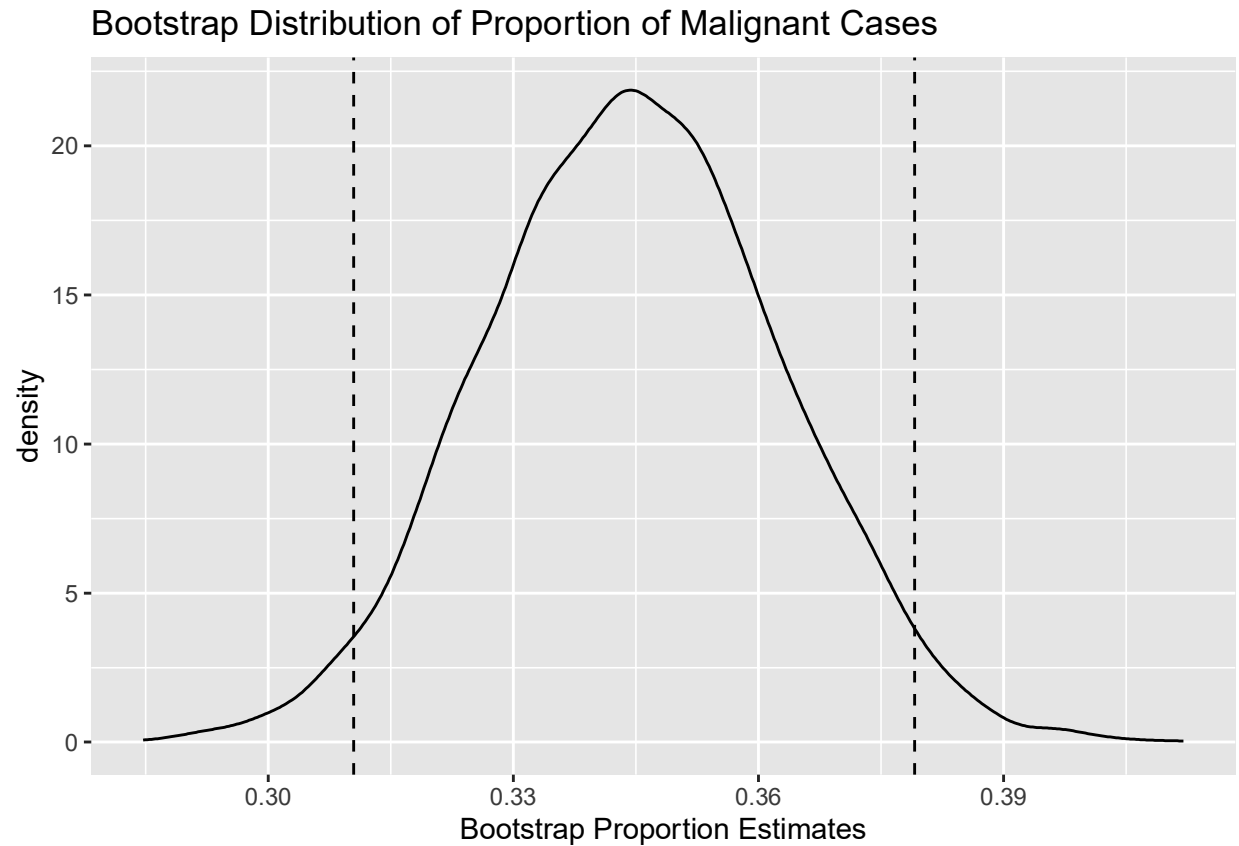
```
## boot(data = data$Class, statistic = boot.fn_class, R = 5000)
##
##
## Bootstrap Statistics :
##      original        bias     std. error
## t1* 0.3447783 -4.978541e-05  0.01795806


##      2.5%      97.5%
## 0.3104435 0.3791130
```

Bootstrap Distribution of Proportion of Malignant Cases


```
## Warning: package 'xgboost' was built under R version 4.4.2

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:xgboost':
##
##     slice

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 109   3
##          1   1  58
##
##                Accuracy : 0.9766
##                  95% CI : (0.9412, 0.9936)
##     No Information Rate : 0.6433
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9487
##
##  Mcnemar's Test P-Value : 0.6171
##
##             Sensitivity : 0.9909
##             Specificity : 0.9508
##          Pos Pred Value : 0.9732
##          Neg Pred Value : 0.9831
##              Prevalence : 0.6433
##          Detection Rate : 0.6374
##    Detection Prevalence : 0.6550
##       Balanced Accuracy : 0.9709
##
##        'Positive' Class : 0
##


## Warning: package 'randomForest' was built under R version 4.4.2


## randomForest 4.7-1.2


## Type rfNews() to see new features/changes/bug fixes.


##
## Attaching package: 'randomForest'


## The following object is masked from 'package:dplyr':
##
##     combine


## The following object is masked from 'package:ggplot2':
##
##     margin


## Warning: package 'randomForestSRC' was built under R version 4.4.2


##
##   randomForestSRC 3.3.1
##
##   Type rfsrc.news() to see new features, changes, and bug fixes.
##
```

```
##
## Attaching package: 'randomForestSRC'


## The following objects are masked from 'package:e1071':
##
##     impute, tune


##
## Call:
##  randomForest(formula = Class ~ ., data = train_data, nodesize = 10,      ntrees = 500, importance =
##               Type of random forest: classification
##                     Number of trees: 500
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 2.73%
## Confusion matrix:
##           benign malignant class.error
## benign       301         8  0.02588997
## malignant      5       162  0.02994012


## Confusion Matrix and Statistics
##
##            Reference
## Prediction  benign malignant
##    benign      132         2
##    malignant     3        70
##
##                Accuracy : 0.9758
##                  95% CI : (0.9445, 0.9921)
##     No Information Rate : 0.6522
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9469
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9778
##             Specificity : 0.9722
##          Pos Pred Value : 0.9851
##          Neg Pred Value : 0.9589
##              Prevalence : 0.6522
##          Detection Rate : 0.6377
##    Detection Prevalence : 0.6473
##       Balanced Accuracy : 0.9750
##
##        'Positive' Class : benign
##
```