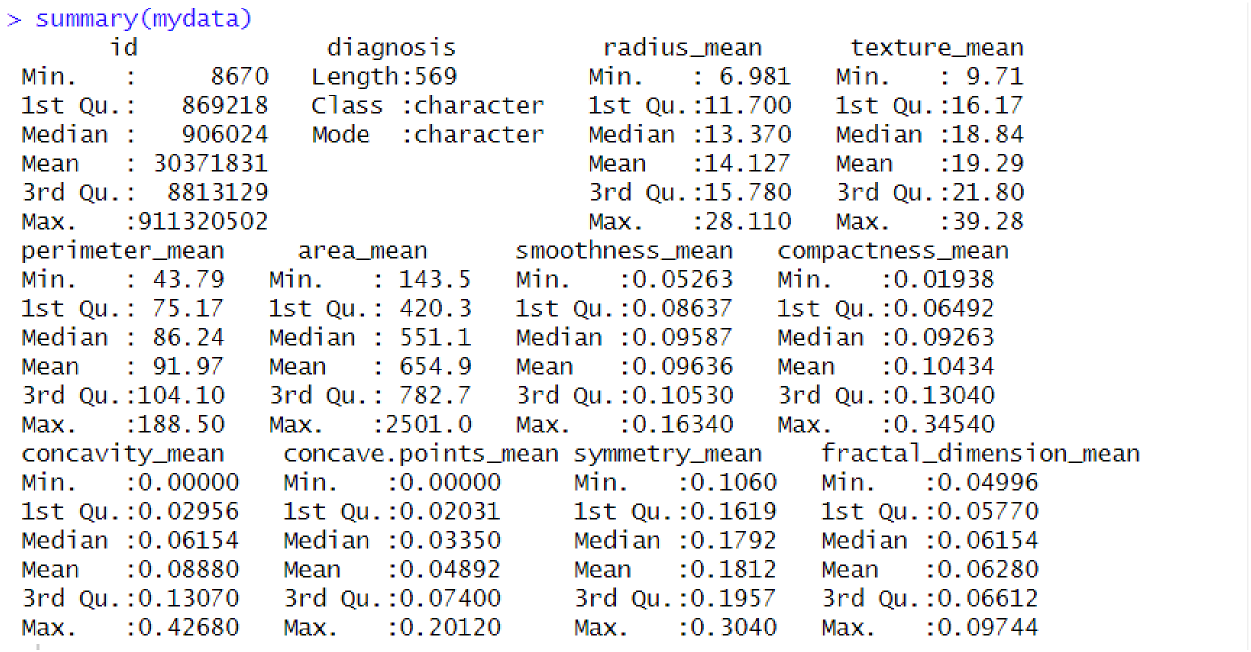
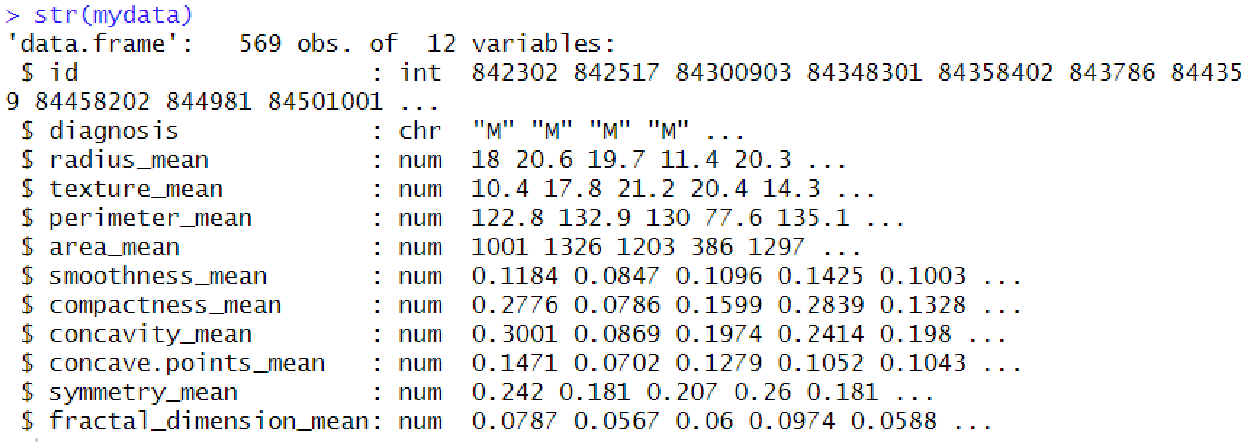
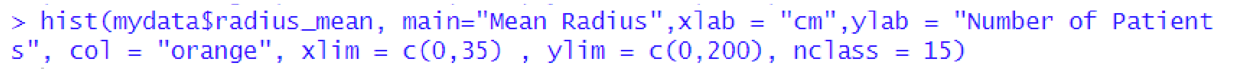


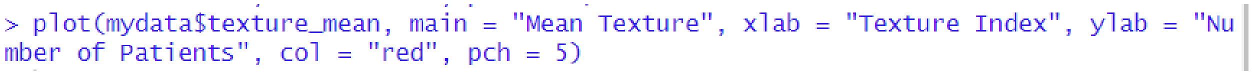
```
> summary(mydata)
       id               diagnosis          radius_mean      texture_mean  
 Min.   :     8670   Length:569         Min.   : 6.981   Min.   : 9.71  
 1st Qu.:   869218   Class :character   1st Qu.:11.700   1st Qu.:16.17  
 Median :   906024   Mode  :character   Median :13.370   Median :18.84  
 Mean   : 30371831                      Mean   :14.127   Mean   :19.29  
 3rd Qu.:  8813129                      3rd Qu.:15.780   3rd Qu.:21.80  
 Max.   :911320502                      Max.   :28.110   Max.   :39.28  
 perimeter_mean     area_mean      smoothness_mean   compactness_mean 
 Min.   : 43.79   Min.   : 143.5   Min.   :0.05263   Min.   :0.01938  
 1st Qu.: 75.17   1st Qu.: 420.3   1st Qu.:0.08637   1st Qu.:0.06492  
 Median : 86.24   Median : 551.1   Median :0.09587   Median :0.09263  
 Mean   : 91.97   Mean   : 654.9   Mean   :0.09636   Mean   :0.10434  
 3rd Qu.:104.10   3rd Qu.: 782.7   3rd Qu.:0.10530   3rd Qu.:0.13040  
 Max.   :188.50   Max.   :2501.0   Max.   :0.16340   Max.   :0.34540  
 concavity_mean    concave.points_mean symmetry_mean    fractal_dimension_mean
 Min.   :0.00000   Min.   :0.00000     Min.   :0.1060   Min.   :0.04996       
 1st Qu.:0.02956   1st Qu.:0.02031     1st Qu.:0.1619   1st Qu.:0.05770       
 Median :0.06154   Median :0.03350     Median :0.1792   Median :0.06154       
 Mean   :0.08880   Mean   :0.04892     Mean   :0.1812   Mean   :0.06280       
 3rd Qu.:0.13070   3rd Qu.:0.07400     3rd Qu.:0.1957   3rd Qu.:0.06612       
 Max.   :0.42680   Max.   :0.20120     Max.   :0.3040   Max.   :0.09744       
```



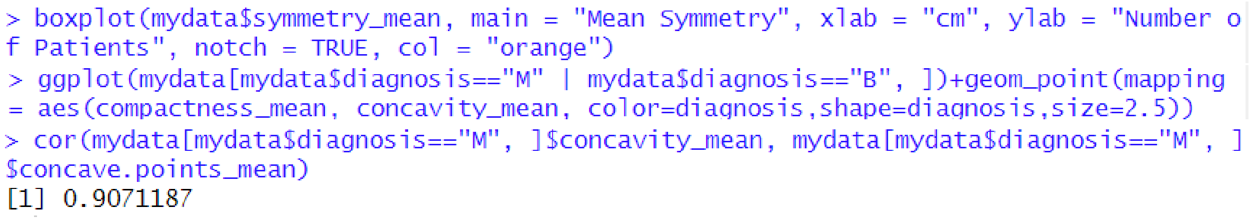
```
> str(mydata)
'data.frame':	569 obs. of  12 variables:
 $ id                    : int  842302 842517 84300903 84348301 84358402 843786 84435
9 84458202 844981 84501001 ...
 $ diagnosis             : chr  "M" "M" "M" "M" ...
 $ radius_mean           : num  18 20.6 19.7 11.4 20.3 ...
 $ texture_mean          : num  10.4 17.8 21.2 20.4 14.3 ...
 $ perimeter_mean        : num  122.8 132.9 130 77.6 135.1 ...
 $ area_mean             : num  1001 1326 1203 386 1297 ...
 $ smoothness_mean       : num  0.1184 0.0847 0.1096 0.1425 0.1003 ...
 $ compactness_mean      : num  0.2776 0.0786 0.1599 0.2839 0.1328 ...
 $ concavity_mean        : num  0.3001 0.0869 0.1974 0.2414 0.198 ...
 $ concave.points_mean   : num  0.1471 0.0702 0.1279 0.1052 0.1043 ...
 $ symmetry_mean         : num  0.242 0.181 0.207 0.26 0.181 ...
 $ fractal_dimension_mean: num  0.0787 0.0567 0.06 0.0974 0.0588 ...
```



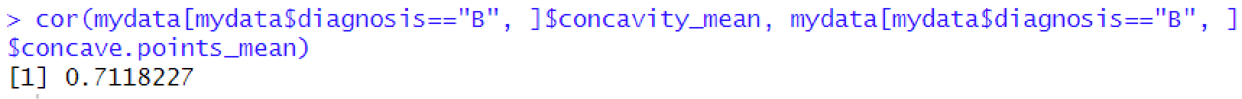
```
> hist(mydata$radius_mean, main="Mean Radius",xlab = "cm",ylab = "Number of Patient
s", col = "orange", xlim = c(0,35) , ylim = c(0,200), nclass = 15)
```



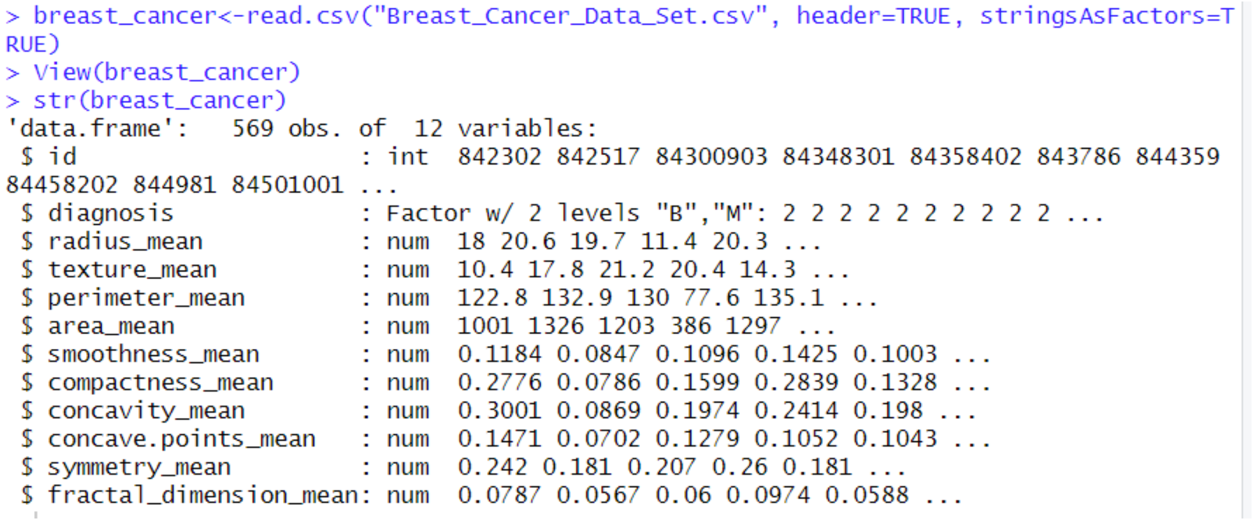
```
> plot(mydata$texture_mean, main = "Mean Texture", xlab = "Texture Index", ylab = "Nu
mber of Patients", col = "red", pch = 5)
```



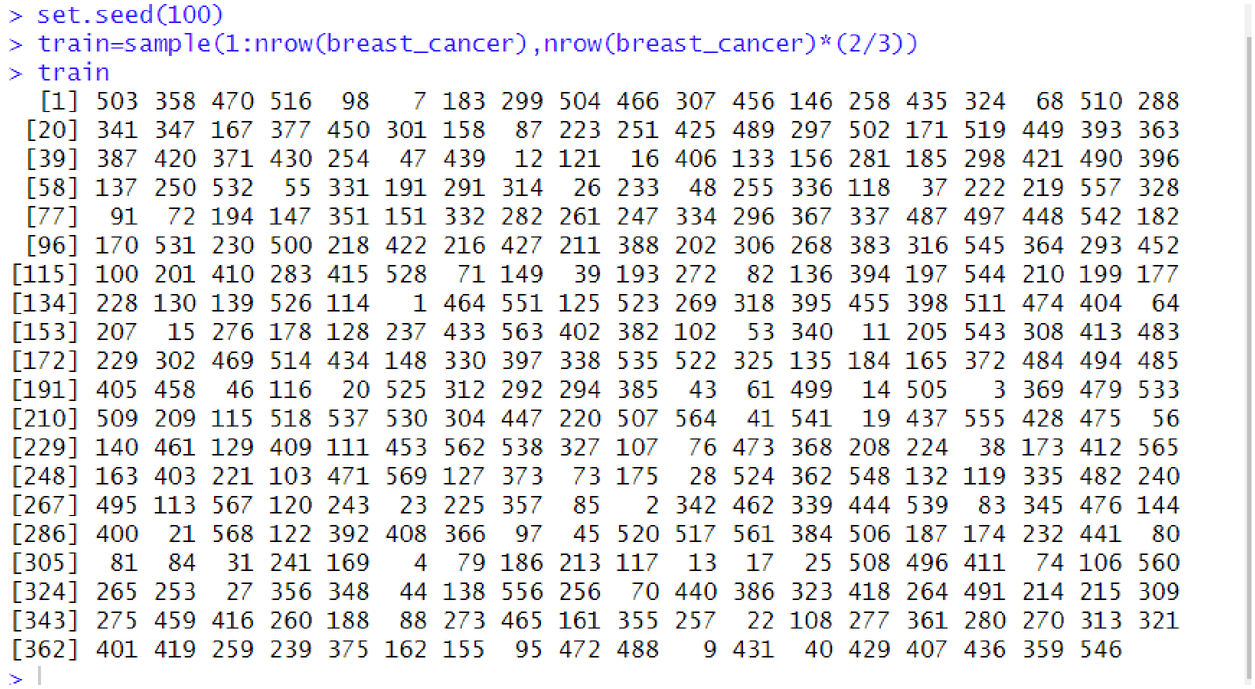
```
> boxplot(mydata$symmetry_mean, main = "Mean Symmetry", xlab = "cm", ylab = "Number o
f Patients", notch = TRUE, col = "orange")
> ggplot(mydata[mydata$diagnosis=="M" | mydata$diagnosis=="B", ])+geom_point(mapping
= aes(compactness_mean, concavity_mean, color=diagnosis,shape=diagnosis,size=2.5))
> cor(mydata[mydata$diagnosis=="M", ]$concavity_mean, mydata[mydata$diagnosis=="M", ]
$concave.points_mean)
[1] 0.9071187
```



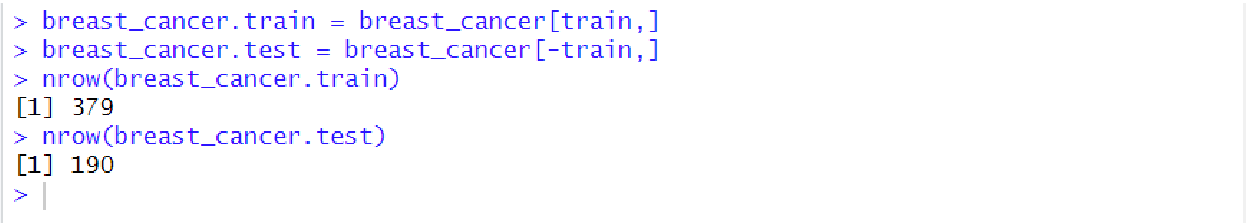
```
> cor(mydata[mydata$diagnosis=="B", ]$concavity_mean, mydata[mydata$diagnosis=="B", ]
$concave.points_mean)
[1] 0.7118227
```



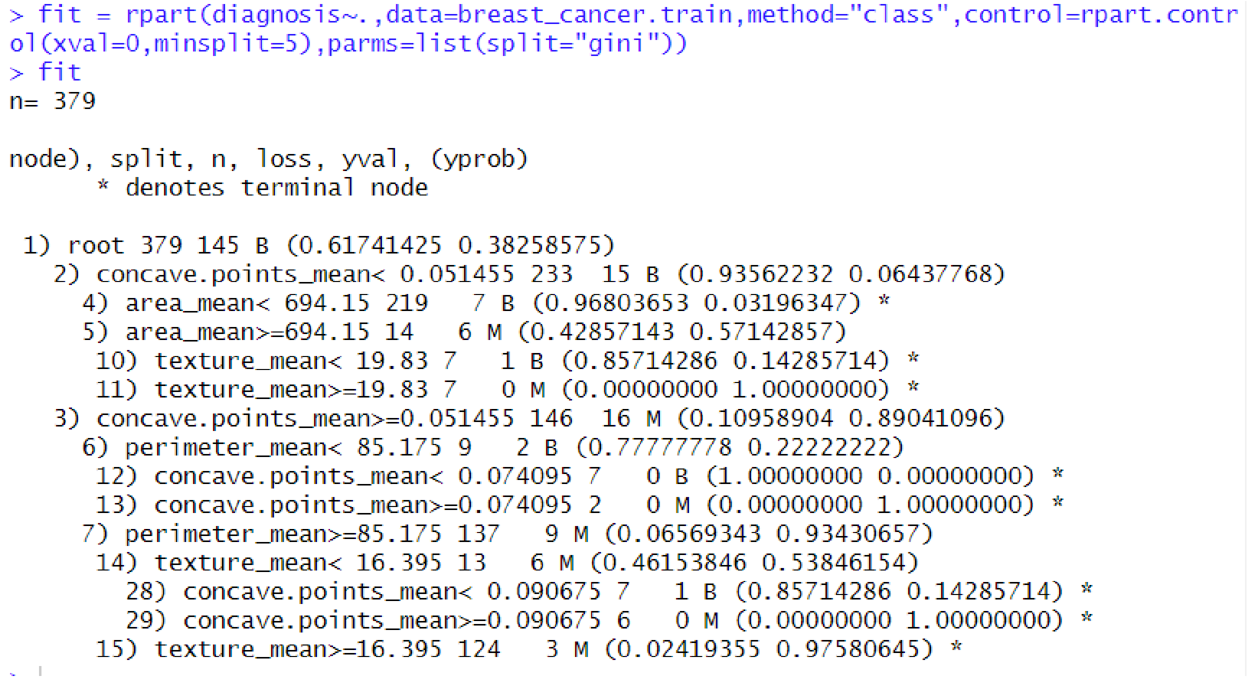
```
> breast_cancer<-read.csv("Breast_Cancer_Data_Set.csv", header=TRUE, stringsAsFactors=T
RUE)
> View(breast_cancer)
> str(breast_cancer)
'data.frame':	569 obs. of  12 variables:
 $ id                    : int  842302 842517 84300903 84348301 84358402 843786 844359
84458202 844981 84501001 ...
 $ diagnosis             : Factor w/ 2 levels "B","M": 2 2 2 2 2 2 2 2 2 2 ...
 $ radius_mean           : num  18 20.6 19.7 11.4 20.3 ...
 $ texture_mean          : num  10.4 17.8 21.2 20.4 14.3 ...
 $ perimeter_mean        : num  122.8 132.9 130 77.6 135.1 ...
 $ area_mean             : num  1001 1326 1203 386 1297 ...
 $ smoothness_mean       : num  0.1184 0.0847 0.1096 0.1425 0.1003 ...
 $ compactness_mean      : num  0.2776 0.0786 0.1599 0.2839 0.1328 ...
 $ concavity_mean        : num  0.3001 0.0869 0.1974 0.2414 0.198 ...
 $ concave.points_mean   : num  0.1471 0.0702 0.1279 0.1052 0.1043 ...
 $ symmetry_mean         : num  0.242 0.181 0.207 0.26 0.181 ...
 $ fractal_dimension_mean: num  0.0787 0.0567 0.06 0.0974 0.0588 ...
```



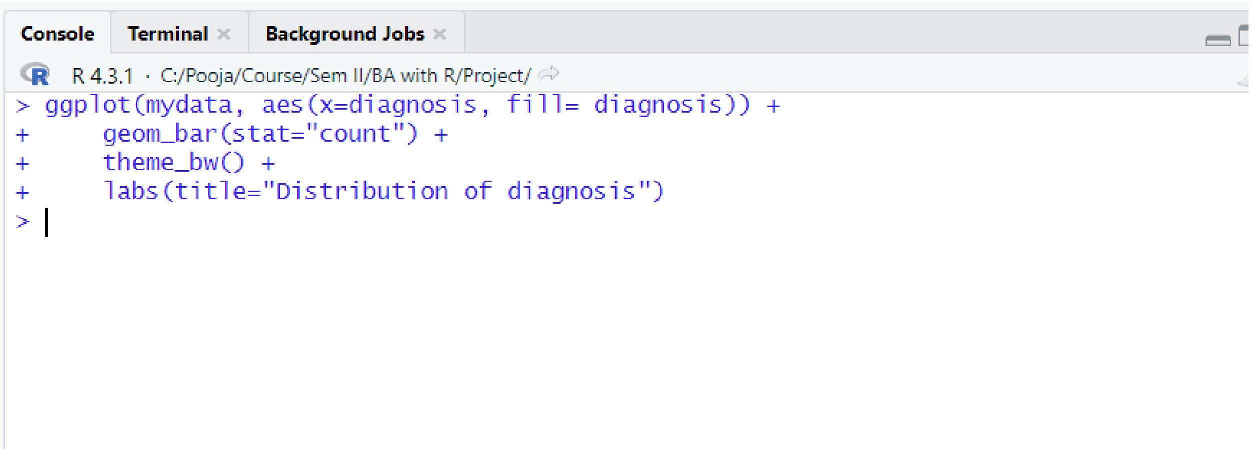
```
> set.seed(100)
> train=sample(1:nrow(breast_cancer),nrow(breast_cancer)*(2/3))
> train
  [1] 503 358 470 516  98   7 183 299 504 466 307 456 146 258 435 324  68 510 288
 [20] 341 347 167 377 450 301 158  87 223 251 425 489 297 502 171 519 449 393 363
 [39] 387 420 371 430 254  47 439  12 121  16 406 133 156 281 185 298 421 490 396
 [58] 137 250 532  55 331 191 291 314  26 233  48 255 336 118  37 222 219 557 328
 [77]  91  72 194 147 351 151 332 282 261 247 334 296 367 337 487 497 448 542 182
 [96] 170 531 230 500 218 422 216 427 211 388 202 306 268 383 316 545 364 293 452
[115] 100 201 410 283 415 528  71 149  39 193 272  82 136 394 197 544 210 199 177
[134] 228 130 139 526 114   1 464 551 125 523 269 318 395 455 398 511 474 404  64
[153] 207  15 276 178 128 237 433 563 402 382 102  53 340  11 205 543 308 413 483
[172] 229 302 469 514 434 148 330 397 338 535 522 325 135 184 165 372 484 494 485
[191] 405 458  46 116  20 525 312 292 294 385  43  61 499  14 505   3 369 479 533
[210] 509 209 115 518 537 530 304 447 220 507 564  41 541  19 437 555 428 475  56
[229] 140 461 129 409 111 453 562 538 327 107  76 473 368 208 224  38 173 412 565
[248] 163 403 221 103 471 569 127 373  73 175  28 524 362 548 132 119 335 482 240
[267] 495 113 567 120 243  23 225 357  85   2 342 462 339 444 539  83 345 476 144
[286] 400  21 568 122 392 408 366  97  45 520 517 561 384 506 187 174 232 441  80
[305]  81  84  31 241 169   4  79 186 213 117  13  17  25 508 496 411  74 106 560
[324] 265 253  27 356 348  44 138 556 256  70 440 386 323 418 264 491 214 215 309
[343] 275 459 416 260 188  88 273 465 161 355 257  22 108 277 361 280 270 313 321
[362] 401 419 259 239 375 162 155  95 472 488   9 431  40 429 407 436 359 546
> 
```



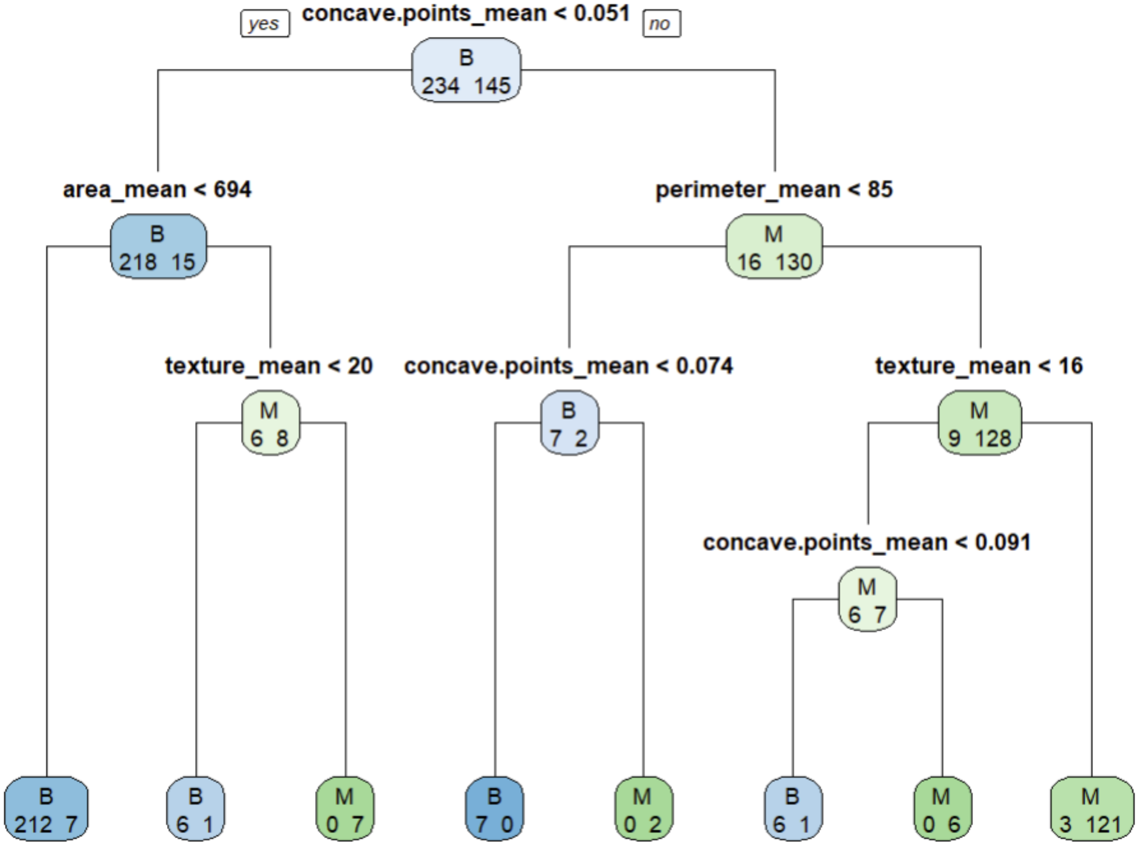
```
> breast_cancer.train = breast_cancer[train,]
> breast_cancer.test = breast_cancer[-train,]
> nrow(breast_cancer.train)
[1] 379
> nrow(breast_cancer.test)
[1] 190
> 
```

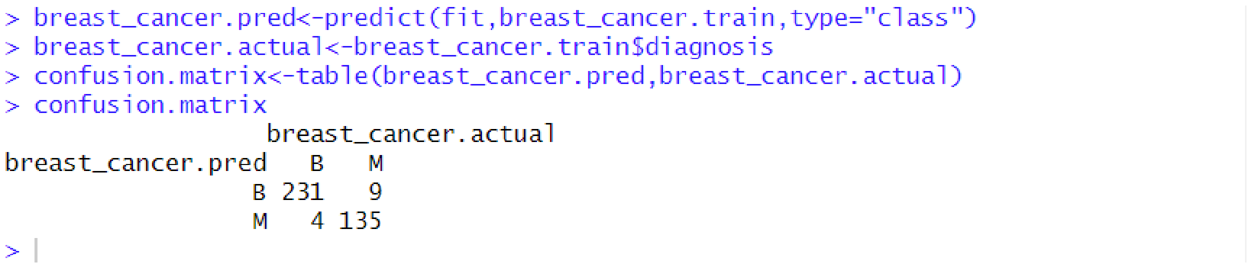


```
> fit = rpart(diagnosis~.,data=breast_cancer.train,method="class",control=rpart.contr
ol(xval=0,minsplit=5),parms=list(split="gini"))
> fit
n= 379

node), split, n, loss, yval, (yprob)
      * denotes terminal node

 1) root 379 145 B (0.61741425 0.38258575)
   2) concave.points_mean< 0.051455 233  15 B (0.93562232 0.06437768)
     4) area_mean< 694.15 219   7 B (0.96803653 0.03196347) *
     5) area_mean>=694.15 14   6 M (0.42857143 0.57142857)
      10) texture_mean< 19.83 7   1 B (0.85714286 0.14285714) *
      11) texture_mean>=19.83 7   0 M (0.00000000 1.00000000) *
   3) concave.points_mean>=0.051455 146  16 M (0.10958904 0.89041096)
     6) perimeter_mean< 85.175 9   2 B (0.77777778 0.22222222)
      12) concave.points_mean< 0.074095 7   0 B (1.00000000 0.00000000) *
      13) concave.points_mean>=0.074095 2   0 M (0.00000000 1.00000000) *
     7) perimeter_mean>=85.175 137   9 M (0.06569343 0.93430657)
      14) texture_mean< 16.395 13   6 M (0.46153846 0.53846154)
        28) concave.points_mean< 0.090675 7   1 B (0.85714286 0.14285714) *
        29) concave.points_mean>=0.090675 6   0 M (0.00000000 1.00000000) *
      15) texture_mean>=16.395 124   3 M (0.02419355 0.97580645) *
```

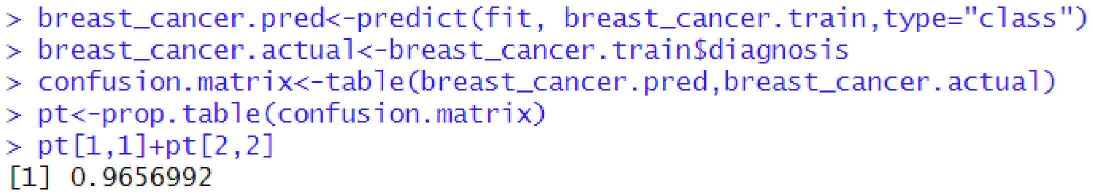


```
> ggplot(mydata, aes(x=diagnosis, fill= diagnosis)) +
+     geom_bar(stat="count") +
+     theme_bw() +
+     labs(title="Distribution of diagnosis")
> 
```

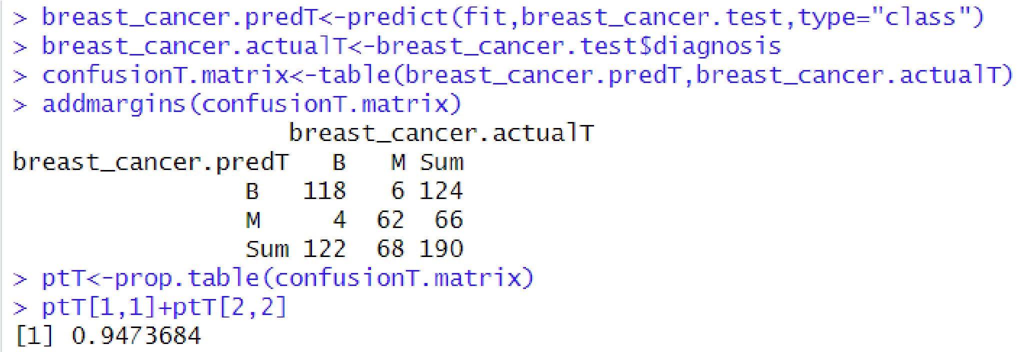




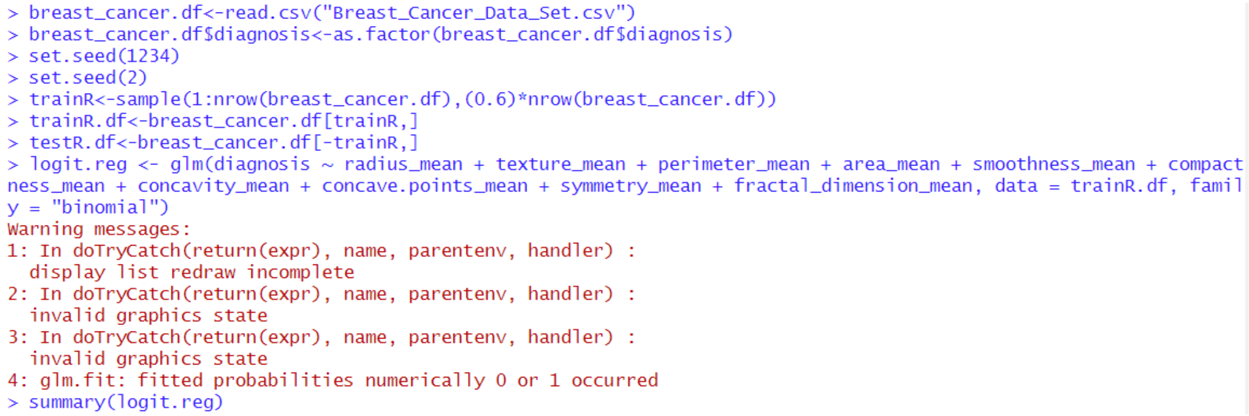
```
> breast_cancer.pred<-predict(fit,breast_cancer.train,type="class")
> breast_cancer.actual<-breast_cancer.train$diagnosis
> confusion.matrix<-table(breast_cancer.pred,breast_cancer.actual)
> confusion.matrix
                  breast_cancer.actual
breast_cancer.pred   B   M
                 B 231   9
                 M   4 135
>
```



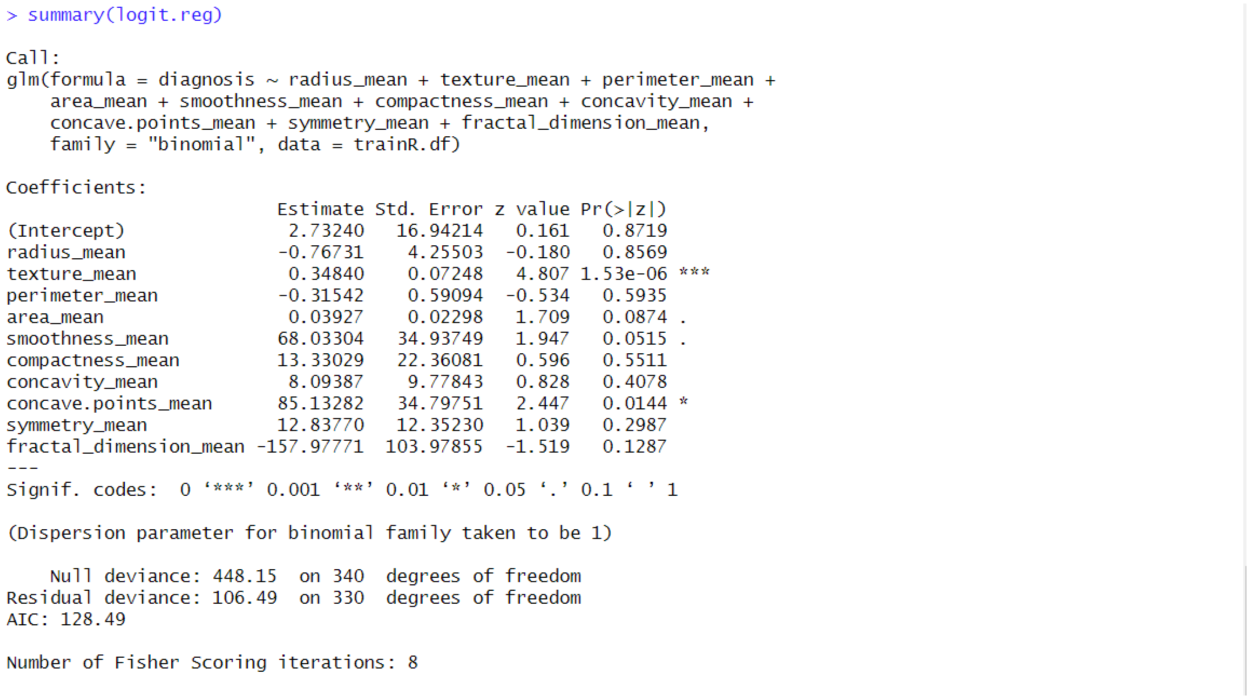
```
> breast_cancer.pred<-predict(fit, breast_cancer.train,type="class")
> breast_cancer.actual<-breast_cancer.train$diagnosis
> confusion.matrix<-table(breast_cancer.pred,breast_cancer.actual)
> pt<-prop.table(confusion.matrix)
> pt[1,1]+pt[2,2]
[1] 0.9656992
```



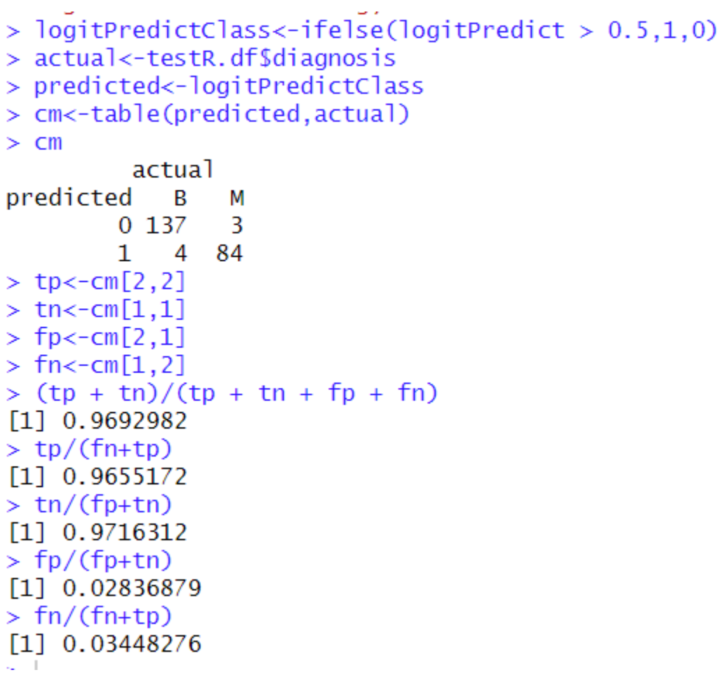
```
> breast_cancer.predT<-predict(fit,breast_cancer.test,type="class")
> breast_cancer.actualT<-breast_cancer.test$diagnosis
> confusionT.matrix<-table(breast_cancer.predT,breast_cancer.actualT)
> addmargins(confusionT.matrix)
                   breast_cancer.actualT
breast_cancer.predT   B   M Sum
                B   118   6 124
                M     4  62  66
                Sum 122  68 190
> ptT<-prop.table(confusionT.matrix)
> ptT[1,1]+ptT[2,2]
[1] 0.9473684
```



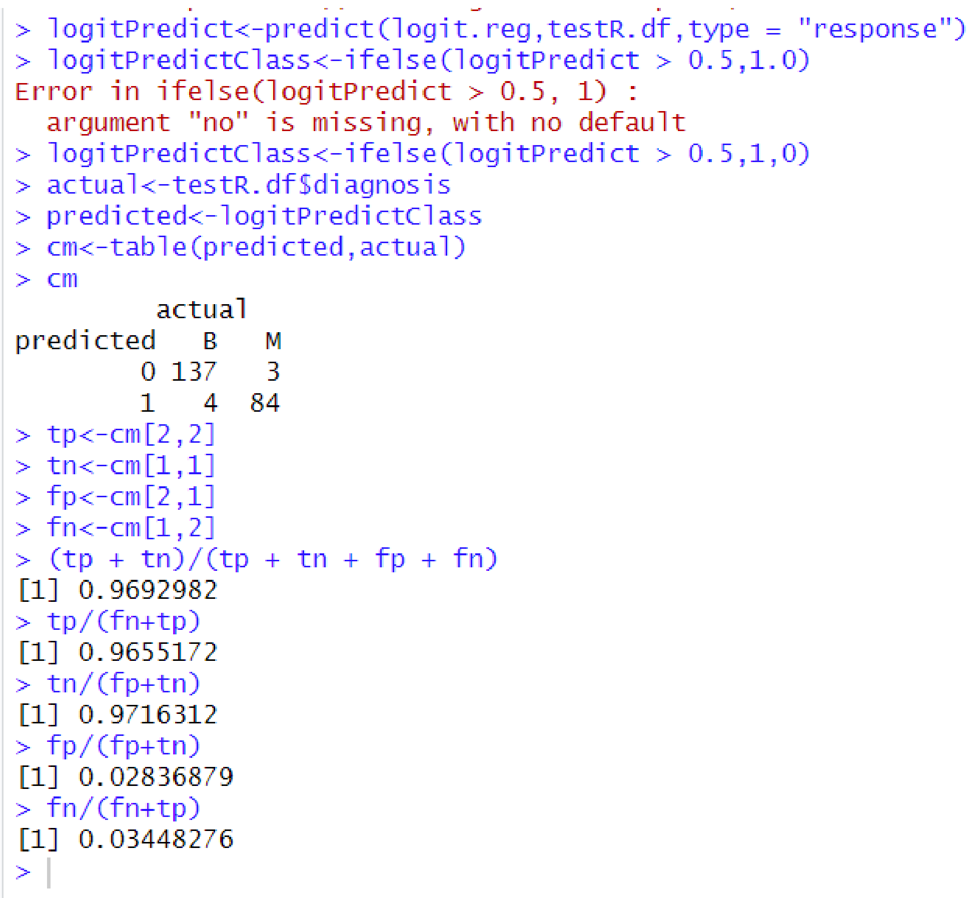
```
> breast_cancer.df<-read.csv("Breast_Cancer_Data_Set.csv")
> breast_cancer.df$diagnosis<-as.factor(breast_cancer.df$diagnosis)
> set.seed(1234)
> set.seed(2)
> trainR<-sample(1:nrow(breast_cancer.df),(0.6)*nrow(breast_cancer.df))
> trainR.df<-breast_cancer.df[trainR,]
> testR.df<-breast_cancer.df[-trainR,]
> logit.reg <- glm(diagnosis ~ radius_mean + texture_mean + perimeter_mean + area_mean + smoothness_mean + compact
ness_mean + concavity_mean + concave.points_mean + symmetry_mean + fractal_dimension_mean, data = trainR.df, famil
y = "binomial")
Warning messages:
1: In doTryCatch(return(expr), name, parentenv, handler) :
  display list redraw incomplete
2: In doTryCatch(return(expr), name, parentenv, handler) :
  invalid graphics state
3: In doTryCatch(return(expr), name, parentenv, handler) :
  invalid graphics state
4: glm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(logit.reg)
```



```
> summary(logit.reg)

Call:
glm(formula = diagnosis ~ radius_mean + texture_mean + perimeter_mean + 
    area_mean + smoothness_mean + compactness_mean + concavity_mean + 
    concave.points_mean + symmetry_mean + fractal_dimension_mean, 
    family = "binomial", data = trainR.df)

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)    
(Intercept)               2.73240   16.94214   0.161   0.8719    
radius_mean              -0.76731    4.25503  -0.180   0.8569    
texture_mean              0.34840    0.07248   4.807 1.53e-06 ***
perimeter_mean           -0.31542    0.59094  -0.534   0.5935    
area_mean                 0.03927    0.02298   1.709   0.0874 .  
smoothness_mean          68.03304   34.93749   1.947   0.0515 .  
compactness_mean         13.33029   22.36081   0.596   0.5511    
concavity_mean            8.09387    9.77843   0.828   0.4078    
concave.points_mean      85.13282   34.79751   2.447   0.0144 *  
symmetry_mean            12.83770   12.35230   1.039   0.2987    
fractal_dimension_mean -157.97771  103.97855  -1.519   0.1287    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 448.15  on 340  degrees of freedom
Residual deviance: 106.49  on 330  degrees of freedom
AIC: 128.49

Number of Fisher Scoring iterations: 8
```



```
> logitPredictClass<-ifelse(logitPredict > 0.5,1,0)
> actual<-testR.df$diagnosis
> predicted<-logitPredictClass
> cm<-table(predicted,actual)
> cm
         actual
predicted   B   M
        0 137   3
        1   4  84
> tp<-cm[2,2]
> tn<-cm[1,1]
> fp<-cm[2,1]
> fn<-cm[1,2]
> (tp + tn)/(tp + tn + fp + fn)
[1] 0.9692982
> tp/(fn+tp)
[1] 0.9655172
> tn/(fp+tn)
[1] 0.9716312
> fp/(fp+tn)
[1] 0.02836879
> fn/(fn+tp)
[1] 0.03448276
```



```
> logitPredict<-predict(logit.reg,testR.df,type = "response")
> logitPredictClass<-ifelse(logitPredict > 0.5,1.0)
Error in ifelse(logitPredict > 0.5, 1) : 
  argument "no" is missing, with no default
> logitPredictClass<-ifelse(logitPredict > 0.5,1,0)
> actual<-testR.df$diagnosis
> predicted<-logitPredictClass
> cm<-table(predicted,actual)
> cm
         actual
predicted   B   M
        0 137   3
        1   4  84
> tp<-cm[2,2]
> tn<-cm[1,1]
> fp<-cm[2,1]
> fn<-cm[1,2]
> (tp + tn)/(tp + tn + fp + fn)
[1] 0.9692982
> tp/(fn+tp)
[1] 0.9655172
> tn/(fp+tn)
[1] 0.9716312
> fp/(fp+tn)
[1] 0.02836879
> fn/(fn+tp)
[1] 0.03448276
> 
```

```
                Variables
1                      id
2               diagnosis
3             radius_mean
4            texture_mean
5          perimeter_mean
6               area_mean
7         smoothness_mean
8        compactness_mean
9          concavity_mean
10    concave points_mean
11          symmetry_mean
12 fractal_dimension_mean

Definition
1
Unique Patient id representing patient
2  This variable indicates whether a breast tumor is benign or malignant. Ben
ign tumors are non-cancerous and typically pose no threat to health. Malignan
t tumors are cancerous and have the potential to spread to other parts of the
body, posing a significant health risk. (B=Benign, M=Malignant)
3
This feature represents the average radius of the tumor cells, which is the d
istance from the center to the outer edge of the tumor.
4
Mean texture refers to the average variation in grayscale intensities of the
pixels within the tumor region as observed in medical images such as mammogra
ms or MRI scans.
```

```
5
The mean perimeter of the tumor represents the average length of the boundary
of the tumor.
6
Mean area refers to the average size of the tumor region, measured in square
units.
7
Mean smoothness characterizes the smoothness of the contour of the tumor boun
dary.
8
Mean compactness is a measure of how closely the tumor cells are packed toget
her relative to their perimeter.
9
Mean concavity refers to the average severity of concavities or inward depres
sions along the boundary of the tumor.
10
Mean concave points represent the average number of concavities or inward cur
vatures along the boundary of the tumor.
11
Mean symmetry measures the symmetry of the tumor shape, comparing the left an
d right sides of the tumor boundary.
12
Fractal dimension quantifies the degree of irregularity and self-similarity i
n the shape of the tumor, with higher values indicating greater complexity.
     Mean std.dev   Min   Max
1
2
3   14.27   3.524 6.981 28.11
4   19.29   4.301  9.71 39.28
5   19.97  24.298 43.79 188.5
6   654.9 351.914 143.5  2501
7   0.096   0.052 0.052 0.163
8  0.1043   0.014 0.019 0.345
9   0.088   0.079     0 0.426
10  0.048   0.038     0 0.201
11  0.181   0.027 0.106 0.304
12  0.628   0.007 0.049 0.097
```

| Variables | Definition | Mean | Std.dev | Min | Max |
|---|---|---|---|---|---|
| **id** | Unique Patient id representing patient | --- | --- | --- | --- |
| **diagnosis (B=Benign, M=Malignant)** | This variable indicates whether a breast tumor is benign or malignant. Benign tumors are non-cancerous and typically pose no threat to health. Malignant tumors are cancerous and have the potential to spread to other parts of the body, posing a significant health risk. | --- | --- | --- | --- |
| **radius_Mean** | This feature represents the average radius of the tumor cells, which is the distance from the center to the outer edge of the tumor. | 14.27 | 3.524 | 6.981 | 28.11 |
| **texture_Mean** | Mean texture refers to the average variation in grayscale intensities of th | 19.29 | 4.301 | 9.71 | 39.28 |

| | | | | | |
|---|---|---|---|---|---|
| | e pixels within the tumor region as observed in medical images such as mammograms or MRI scans. | | | | |
| **perimeter_Mean** | The mean perimeter of the tumor represents the average length of the boundary of the tumor. | 19.97 | 24.298 | 43.79 | 188.5 |
| **area_Mean** | Mean area refers to the average size of the tumor region, measured in square units. | 654.9 | 351.914 | 143.5 | 2501 |
| **smoothness_Mean** | Mean smoothness characterizes the smoothness of the contour of the tumor boundary. | 0.096 | 0.052 | 0.052 | 0.163 |
| **compactness_Mean** | Mean compactness is a measure of how closely the tumor cells are packed together relative to their perimeter. | 0.1043 | 0.014 | 0.019 | 0.345 |
| **concavity_Mean** | Mean concavity refers to the average severity of concavities or inward depressions along the boundary of the tumor. | 0.088 | 0.079 | 0 | 0.426 |
| **concave Points_Mean** | Mean concave points represent the average number of concavities or inward curvatures along the boundary of the tumor. | 0.048 | 0.038 | 0 | 0.201 |
| **symmetry_Mean** | Mean symmetry measures the symmetry of the tumor shape, comparing the left and right sides of the tumor boundary. | 0.181 | 0.027 | 0.106 | 0.304 |
| **fractal_dimension_Mean** | Fractal dimension quantifies the degree of irregularity and self-similarity in the shape of the tumor, with higher values indicating greater complexity. | 0.628 | 0.007 | 0.049 | 0.097 |

| Variables | Definition | Mean | Std.dev | Min | Max |
|---|---|---|---|---|---|
| id | Unique Patient id representing patient | --- | --- | --- | --- |

| | | | | | |
|---|---|---|---|---|---|
| **diagnosis (B=Benign,M=Malignant)** | This variable indicates whether a breast tumor is benign or malignant. Benign tumors are non-cancerous and typically pose no threat to health. Malignant tumors are cancerous and have the potential to spread to other parts of the body, posing a significant health risk. | --- | --- | --- | --- |
| **radius_Mean** | This feature represents the average radius of the tumor cells, which is the distance from the center to the outer edge of the tumor. | 14.27 | 3.524 | 6.981 | 28.11 |
| **texture_Mean** | Mean texture refers to the average variation in grayscale intensities of the pixels within the tumor region as observed in medical images such as mammograms or MRI scans. | 19.29 | 4.301 | 9.71 | 39.28 |
| **perimeter_Mean** | The mean perimeter of the tumor represents the average length of the boundary of the tumor. | 19.97 | 24.298 | 43.79 | 188.5 |
| **area_Mean** | Mean area refers to the average size of the tumor region, measured in square units. | 654.9 | 351.914 | 143.5 | 2501 |
| **smoothness_Mean** | Mean smoothness characterizes the smoothness of the contour of the tumor boundary. | 0.096 | 0.052 | 0.052 | 0.163 |
| **compactness_Mean** | Mean compactness is a measure of how closely the tumor cells are packed together relative to their perimeter. | 0.1043 | 0.014 | 0.019 | 0.345 |
| **concavity_Mean** | Mean concavity refers to the average severity of concavities or inward depressions along the boundary of the tumor. | 0.088 | 0.079 | 0 | 0.426 |
| **concave Points_Mean** | Mean concave points represent the average number of concavities or inward curvatures along the boundary of the tumor. | 0.048 | 0.038 | 0 | 0.201 |
| **symmetry_Mean** | Mean symmetry measures the symmetry of the tumor shape, comparing the left and right sides of the tumor boundary. | 0.181 | 0.027 | 0.106 | 0.304 |
| **fractal_dimension_Mean** | Fractal dimension quantifies the degree of irregularity and self-similarity in the shape of the tumor, with higher values indicating greater complexity. | 0.628 | 0.007 | 0.049 | 0.097 |