> # examine the structure of the wbcd data frame

> str(wbcd)

'data.frame': 569 obs. of 32 variables:

$ id : int 87139402 8910251 905520 868871 9012568 906539 925291 87880 862989 89827 ...

$ diagnosis : chr "B" "B" "B" "B" ...

$ radius\_mean : num 12.3 10.6 11 11.3 15.2 ...

$ texture\_mean : num 12.4 18.9 16.8 13.4 13.2 ...

$ perimeter\_mean : num 78.8 69.3 70.9 73 97.7 ...

$ area\_mean : num 464 346 373 385 712 ...

$ smoothness\_mean : num 0.1028 0.0969 0.1077 0.1164 0.0796 ...

$ compactness\_mean : num 0.0698 0.1147 0.078 0.1136 0.0693 ...

$ concavity\_mean : num 0.0399 0.0639 0.0305 0.0464 0.0339 ...

$ points\_mean : num 0.037 0.0264 0.0248 0.048 0.0266 ...

$ symmetry\_mean : num 0.196 0.192 0.171 0.177 0.172 ...

$ dimension\_mean : num 0.0595 0.0649 0.0634 0.0607 0.0554 ...

$ radius\_se : num 0.236 0.451 0.197 0.338 0.178 ...

$ texture\_se : num 0.666 1.197 1.387 1.343 0.412 ...

$ perimeter\_se : num 1.67 3.43 1.34 1.85 1.34 ...

$ area\_se : num 17.4 27.1 13.5 26.3 17.7 ...

$ smoothness\_se : num 0.00805 0.00747 0.00516 0.01127 0.00501 ...

$ compactness\_se : num 0.0118 0.03581 0.00936 0.03498 0.01485 ...

$ concavity\_se : num 0.0168 0.0335 0.0106 0.0219 0.0155 ...

$ points\_se : num 0.01241 0.01365 0.00748 0.01965 0.00915 ...

$ symmetry\_se : num 0.0192 0.035 0.0172 0.0158 0.0165 ...

$ dimension\_se : num 0.00225 0.00332 0.0022 0.00344 0.00177 ...

$ radius\_worst : num 13.5 11.9 12.4 11.9 16.2 ...

$ texture\_worst : num 15.6 22.9 26.4 15.8 15.7 ...

$ perimeter\_worst : num 87 78.3 79.9 76.5 104.5 ...

$ area\_worst : num 549 425 471 434 819 ...

$ smoothness\_worst : num 0.139 0.121 0.137 0.137 0.113 ...

$ compactness\_worst: num 0.127 0.252 0.148 0.182 0.174 ...

$ concavity\_worst : num 0.1242 0.1916 0.1067 0.0867 0.1362 ...

$ points\_worst : num 0.0939 0.0793 0.0743 0.0861 0.0818 ...

$ symmetry\_worst : num 0.283 0.294 0.3 0.21 0.249 ...

$ dimension\_worst : num 0.0677 0.0759 0.0788 0.0678 0.0677 ...

>

> # drop the id feature

> wbcd <- wbcd[-1]

>

> # table of diagnosis

> table(wbcd$diagnosis)

B M

357 212

>

> # recode diagnosis as a factor

> wbcd$diagnosis <- factor(wbcd$diagnosis, levels = c("B", "M"),

+ labels = c("Benign", "Malignant"))

>

> # table or proportions with more informative labels

> round(prop.table(table(wbcd$diagnosis)) \* 100, digits = 1)

Benign Malignant

62.7 37.3

>

> # summarize three numeric features

> summary(wbcd[c("radius\_mean", "area\_mean", "smoothness\_mean")])

radius\_mean area\_mean smoothness\_mean

Min. : 6.981 Min. : 143.5 Min. :0.05263

1st Qu.:11.700 1st Qu.: 420.3 1st Qu.:0.08637

Median :13.370 Median : 551.1 Median :0.09587

Mean :14.127 Mean : 654.9 Mean :0.09636

3rd Qu.:15.780 3rd Qu.: 782.7 3rd Qu.:0.10530

Max. :28.110 Max. :2501.0 Max. :0.16340

>

> # create normalization function

> normalize <- function(x) {

+ return ((x - min(x)) / (max(x) - min(x)))

+ }

>

> # test normalization function - result should be identical

> normalize(c(1, 2, 3, 4, 5))

[1] 0.00 0.25 0.50 0.75 1.00

> normalize(c(10, 20, 30, 40, 50))

[1] 0.00 0.25 0.50 0.75 1.00

>

> # normalize the wbcd data

> wbcd\_n <- as.data.frame(lapply(wbcd[2:31], normalize))

>

> # confirm that normalization worked

> summary(wbcd\_n$area\_mean)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.0000 0.1174 0.1729 0.2169 0.2711 1.0000

>

> # create training and test data

> wbcd\_train <- wbcd\_n[1:469, ]

> wbcd\_test <- wbcd\_n[470:569, ]

>

> # create labels for training and test data

>

> wbcd\_train\_labels <- wbcd[1:469, 1]

> wbcd\_test\_labels <- wbcd[470:569, 1]

>

> ## Step 3: Training a model on the data ----

>

> # load the "class" library

> library(class)

>

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test,

+ cl = wbcd\_train\_labels, k=21)

>

> ## Step 4: Evaluating model performance ----

>

> # load the "gmodels" library

> library(gmodels)

>

> # Create the cross tabulation of predicted vs. actual

> CrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred,

+ prop.chisq=FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 100

| wbcd\_test\_pred

wbcd\_test\_labels | Benign | Malignant | Row Total |

-----------------|-----------|-----------|-----------|

Benign | 61 | 0 | 61 |

| 1.000 | 0.000 | 0.610 |

| 0.968 | 0.000 | |

| 0.610 | 0.000 | |

-----------------|-----------|-----------|-----------|

Malignant | 2 | 37 | 39 |

| 0.051 | 0.949 | 0.390 |

| 0.032 | 1.000 | |

| 0.020 | 0.370 | |

-----------------|-----------|-----------|-----------|

Column Total | 63 | 37 | 100 |

| 0.630 | 0.370 | |

-----------------|-----------|-----------|-----------|

>

> ## Step 5: Improving model performance ----

>

> # use the scale() function to z-score standardize a data frame

> wbcd\_z <- as.data.frame(scale(wbcd[-1]))

>

> # confirm that the transformation was applied correctly

> summary(wbcd\_z$area\_mean)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-1.4532 -0.6666 -0.2949 0.0000 0.3632 5.2459

>

> # create training and test datasets

> wbcd\_train <- wbcd\_z[1:469, ]

> wbcd\_test <- wbcd\_z[470:569, ]

>

> # re-classify test cases

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test,

+ cl = wbcd\_train\_labels, k=21)

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test, cl = wbcd\_train\_labels, k=1)

> wbcd\_test\_pred

[1] Benign Malignant Benign Benign Malignant Benign Malignant Benign Malignant Benign Malignant Benign Malignant Malignant Benign Benign Malignant Benign Malignant

[20] Benign Malignant Malignant Malignant Malignant Benign Benign Benign Benign Benign Malignant Malignant Benign Malignant Malignant Benign Benign Benign Benign

[39] Benign Malignant Malignant Benign Malignant Malignant Benign Malignant Malignant Malignant Malignant Malignant Malignant Benign Benign Benign Benign Benign Benign

[58] Benign Malignant Benign Benign Benign Benign Benign Malignant Malignant Benign Benign Benign Benign Benign Malignant Benign Benign Malignant Malignant

[77] Benign Benign Benign Benign Benign Benign Benign Malignant Benign Benign Malignant Benign Benign Benign Benign Malignant Benign Benign Benign

[96] Benign Benign Malignant Benign Malignant

Levels: Benign Malignant

> View(knn\_model)

> View(knn\_test\_data)

> View(knn\_test\_model)

> knn\_train\_mode0\_lbl <- knn\_model[1:840, 3]

> knn\_test\_model\_lbl <- knn\_model[841:1200, 3]

> #load class library

> library(class)

>

>

>

> #5. Create 100 random distributed test data points

> x2 <- runif(100, 0, 1)

> y2 <- runif(100, 0, 1)

> #coords2 <- data.frame(x=x2, y=y2)

>

> #create knn-test-data df

> knn\_test\_data <- assign\_Quad(x2, y2, 100)

> #6 first attempt #or knn\_test\_data

> knn\_pred1 <- knn(train = knn\_train\_model, test = knn\_test\_model, cl= knn\_train\_mode0\_lbl, k=34)

> View(knn\_pred1)

> #Evaluate model performance

> library(gmodels)

>

> #Create cross-table on pred. vs actual

> CrossTable(x = knn\_test\_model\_lbl, y = knn\_pred1, prop.chisq = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 360

| knn\_pred1

knn\_test\_model\_lbl | 1 | 2 | 3 | 4 | Row Total |

-------------------|-----------|-----------|-----------|-----------|-----------|

1 | 103 | 0 | 0 | 0 | 103 |

| 1.000 | 0.000 | 0.000 | 0.000 | 0.286 |

| 1.000 | 0.000 | 0.000 | 0.000 | |

| 0.286 | 0.000 | 0.000 | 0.000 | |

-------------------|-----------|-----------|-----------|-----------|-----------|

2 | 0 | 112 | 0 | 0 | 112 |

| 0.000 | 1.000 | 0.000 | 0.000 | 0.311 |

| 0.000 | 1.000 | 0.000 | 0.000 | |

| 0.000 | 0.311 | 0.000 | 0.000 | |

-------------------|-----------|-----------|-----------|-----------|-----------|

3 | 0 | 0 | 45 | 0 | 45 |

| 0.000 | 0.000 | 1.000 | 0.000 | 0.125 |

| 0.000 | 0.000 | 1.000 | 0.000 | |

| 0.000 | 0.000 | 0.125 | 0.000 | |

-------------------|-----------|-----------|-----------|-----------|-----------|

4 | 0 | 0 | 0 | 100 | 100 |

| 0.000 | 0.000 | 0.000 | 1.000 | 0.278 |

| 0.000 | 0.000 | 0.000 | 1.000 | |

| 0.000 | 0.000 | 0.000 | 0.278 | |

-------------------|-----------|-----------|-----------|-----------|-----------|

Column Total | 103 | 112 | 45 | 100 | 360 |

| 0.286 | 0.311 | 0.125 | 0.278 | |

-------------------|-----------|-----------|-----------|-----------|-----------|

> ##### Chapter 3: Classification using Nearest Neighbors --------------------

>

> ## Example: Classifying Cancer Samples ----

> ## Step 2: Exploring and preparing the data ----

>

> # import the CSV file

> wbcd <- read.csv("wisc\_bc\_data.csv", stringsAsFactors = FALSE)

Error in file(file, "rt") : cannot open the connection

In addition: Warning message:

In file(file, "rt") :

cannot open file 'wisc\_bc\_data.csv': No such file or directory

> wbcd <- read.csv("C:\\Users\\Kevin Oane\\Documents\\DB2\\db\_last\\wisc\_bc\_data.csv", stringsAsFactors = FALSE)

>

> # examine the structure of the wbcd data frame

> str(wbcd)

'data.frame': 569 obs. of 32 variables:

$ id : int 87139402 8910251 905520 868871 9012568 906539 925291 87880 862989 89827 ...

$ diagnosis : chr "B" "B" "B" "B" ...

$ radius\_mean : num 12.3 10.6 11 11.3 15.2 ...

$ texture\_mean : num 12.4 18.9 16.8 13.4 13.2 ...

$ perimeter\_mean : num 78.8 69.3 70.9 73 97.7 ...

$ area\_mean : num 464 346 373 385 712 ...

$ smoothness\_mean : num 0.1028 0.0969 0.1077 0.1164 0.0796 ...

$ compactness\_mean : num 0.0698 0.1147 0.078 0.1136 0.0693 ...

$ concavity\_mean : num 0.0399 0.0639 0.0305 0.0464 0.0339 ...

$ points\_mean : num 0.037 0.0264 0.0248 0.048 0.0266 ...

$ symmetry\_mean : num 0.196 0.192 0.171 0.177 0.172 ...

$ dimension\_mean : num 0.0595 0.0649 0.0634 0.0607 0.0554 ...

$ radius\_se : num 0.236 0.451 0.197 0.338 0.178 ...

$ texture\_se : num 0.666 1.197 1.387 1.343 0.412 ...

$ perimeter\_se : num 1.67 3.43 1.34 1.85 1.34 ...

$ area\_se : num 17.4 27.1 13.5 26.3 17.7 ...

$ smoothness\_se : num 0.00805 0.00747 0.00516 0.01127 0.00501 ...

$ compactness\_se : num 0.0118 0.03581 0.00936 0.03498 0.01485 ...

$ concavity\_se : num 0.0168 0.0335 0.0106 0.0219 0.0155 ...

$ points\_se : num 0.01241 0.01365 0.00748 0.01965 0.00915 ...

$ symmetry\_se : num 0.0192 0.035 0.0172 0.0158 0.0165 ...

$ dimension\_se : num 0.00225 0.00332 0.0022 0.00344 0.00177 ...

$ radius\_worst : num 13.5 11.9 12.4 11.9 16.2 ...

$ texture\_worst : num 15.6 22.9 26.4 15.8 15.7 ...

$ perimeter\_worst : num 87 78.3 79.9 76.5 104.5 ...

$ area\_worst : num 549 425 471 434 819 ...

$ smoothness\_worst : num 0.139 0.121 0.137 0.137 0.113 ...

$ compactness\_worst: num 0.127 0.252 0.148 0.182 0.174 ...

$ concavity\_worst : num 0.1242 0.1916 0.1067 0.0867 0.1362 ...

$ points\_worst : num 0.0939 0.0793 0.0743 0.0861 0.0818 ...

$ symmetry\_worst : num 0.283 0.294 0.3 0.21 0.249 ...

$ dimension\_worst : num 0.0677 0.0759 0.0788 0.0678 0.0677 ...

>

> # drop the id feature

> wbcd <- wbcd[-1]

>

> # table of diagnosis

> table(wbcd$diagnosis)

B M

357 212

>

> # recode diagnosis as a factor

> wbcd$diagnosis <- factor(wbcd$diagnosis, levels = c("B", "M"),

+ labels = c("Benign", "Malignant"))

>

> # table or proportions with more informative labels

> round(prop.table(table(wbcd$diagnosis)) \* 100, digits = 1)

Benign Malignant

62.7 37.3

>

> # summarize three numeric features

> summary(wbcd[c("radius\_mean", "area\_mean", "smoothness\_mean")])

radius\_mean area\_mean smoothness\_mean

Min. : 6.981 Min. : 143.5 Min. :0.05263

1st Qu.:11.700 1st Qu.: 420.3 1st Qu.:0.08637

Median :13.370 Median : 551.1 Median :0.09587

Mean :14.127 Mean : 654.9 Mean :0.09636

3rd Qu.:15.780 3rd Qu.: 782.7 3rd Qu.:0.10530

Max. :28.110 Max. :2501.0 Max. :0.16340

>

> # create normalization function

> normalize <- function(x) {

+ return ((x - min(x)) / (max(x) - min(x)))

+ }

>

> # test normalization function - result should be identical

> normalize(c(1, 2, 3, 4, 5))

[1] 0.00 0.25 0.50 0.75 1.00

> normalize(c(10, 20, 30, 40, 50))

[1] 0.00 0.25 0.50 0.75 1.00

>

> # normalize the wbcd data

> wbcd\_n <- as.data.frame(lapply(wbcd[2:31], normalize))

>

> # confirm that normalization worked

> summary(wbcd\_n$area\_mean)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.0000 0.1174 0.1729 0.2169 0.2711 1.0000

>

> # create training and test data

> wbcd\_train <- wbcd\_n[1:469, ]

> wbcd\_test <- wbcd\_n[470:569, ]

>

> # create labels for training and test data

>

> wbcd\_train\_labels <- wbcd[1:469, 1]

> wbcd\_test\_labels <- wbcd[470:569, 1]

>

> ## Step 3: Training a model on the data ----

>

> # load the "class" library

> library(class)

>

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test,

+ cl = wbcd\_train\_labels, k=21)

>

> ## Step 4: Evaluating model performance ----

>

> # load the "gmodels" library

> library(gmodels)

>

> # Create the cross tabulation of predicted vs. actual

> CrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred,

+ prop.chisq=FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 100

| wbcd\_test\_pred

wbcd\_test\_labels | Benign | Malignant | Row Total |

-----------------|-----------|-----------|-----------|

Benign | 61 | 0 | 61 |

| 1.000 | 0.000 | 0.610 |

| 0.968 | 0.000 | |

| 0.610 | 0.000 | |

-----------------|-----------|-----------|-----------|

Malignant | 2 | 37 | 39 |

| 0.051 | 0.949 | 0.390 |

| 0.032 | 1.000 | |

| 0.020 | 0.370 | |

-----------------|-----------|-----------|-----------|

Column Total | 63 | 37 | 100 |

| 0.630 | 0.370 | |

-----------------|-----------|-----------|-----------|

>

> ## Step 5: Improving model performance ----

>

> # use the scale() function to z-score standardize a data frame

> wbcd\_z <- as.data.frame(scale(wbcd[-1]))

>

> # confirm that the transformation was applied correctly

> summary(wbcd\_z$area\_mean)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-1.4532 -0.6666 -0.2949 0.0000 0.3632 5.2459

>

> # create training and test datasets

> wbcd\_train <- wbcd\_z[1:469, ]

> wbcd\_test <- wbcd\_z[470:569, ]

>

> # re-classify test cases

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test,

+ cl = wbcd\_train\_labels, k=21)

>

> # Create the cross tabulation of predicted vs. actual

> CrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred,

+ prop.chisq=FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 100

| wbcd\_test\_pred

wbcd\_test\_labels | Benign | Malignant | Row Total |

-----------------|-----------|-----------|-----------|

Benign | 61 | 0 | 61 |

| 1.000 | 0.000 | 0.610 |

| 0.924 | 0.000 | |

| 0.610 | 0.000 | |

-----------------|-----------|-----------|-----------|

Malignant | 5 | 34 | 39 |

| 0.128 | 0.872 | 0.390 |

| 0.076 | 1.000 | |

| 0.050 | 0.340 | |

-----------------|-----------|-----------|-----------|

Column Total | 66 | 34 | 100 |

| 0.660 | 0.340 | |

-----------------|-----------|-----------|-----------|

>

> # try several different values of k

> wbcd\_train <- wbcd\_n[1:469, ]

> wbcd\_test <- wbcd\_n[470:569, ]

>

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test, cl = wbcd\_train\_labels, k=1)

> wbCrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred, prop.chisq=FALSE)

Error in wbCrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred, prop.chisq = FALSE) :

could not find function "wbCrossTable"

>

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test, cl = wbcd\_train\_labels, k=5)

> CrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred, prop.chisq=FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 100

| wbcd\_test\_pred

wbcd\_test\_labels | Benign | Malignant | Row Total |

-----------------|-----------|-----------|-----------|

Benign | 61 | 0 | 61 |

| 1.000 | 0.000 | 0.610 |

| 0.968 | 0.000 | |

| 0.610 | 0.000 | |

-----------------|-----------|-----------|-----------|

Malignant | 2 | 37 | 39 |

| 0.051 | 0.949 | 0.390 |

| 0.032 | 1.000 | |

| 0.020 | 0.370 | |

-----------------|-----------|-----------|-----------|

Column Total | 63 | 37 | 100 |

| 0.630 | 0.370 | |

-----------------|-----------|-----------|-----------|

>

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test, cl = wbcd\_train\_labels, k=11)

> CrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred, prop.chisq=FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 100

| wbcd\_test\_pred

wbcd\_test\_labels | Benign | Malignant | Row Total |

-----------------|-----------|-----------|-----------|

Benign | 61 | 0 | 61 |

| 1.000 | 0.000 | 0.610 |

| 0.953 | 0.000 | |

| 0.610 | 0.000 | |

-----------------|-----------|-----------|-----------|

Malignant | 3 | 36 | 39 |

| 0.077 | 0.923 | 0.390 |

| 0.047 | 1.000 | |

| 0.030 | 0.360 | |

-----------------|-----------|-----------|-----------|

Column Total | 64 | 36 | 100 |

| 0.640 | 0.360 | |

-----------------|-----------|-----------|-----------|

>

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test, cl = wbcd\_train\_labels, k=15)

> CrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred, prop.chisq=FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 100

| wbcd\_test\_pred

wbcd\_test\_labels | Benign | Malignant | Row Total |

-----------------|-----------|-----------|-----------|

Benign | 61 | 0 | 61 |

| 1.000 | 0.000 | 0.610 |

| 0.953 | 0.000 | |

| 0.610 | 0.000 | |

-----------------|-----------|-----------|-----------|

Malignant | 3 | 36 | 39 |

| 0.077 | 0.923 | 0.390 |

| 0.047 | 1.000 | |

| 0.030 | 0.360 | |

-----------------|-----------|-----------|-----------|

Column Total | 64 | 36 | 100 |

| 0.640 | 0.360 | |

-----------------|-----------|-----------|-----------|

>

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test, cl = wbcd\_train\_labels, k=21)

> CrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred, prop.chisq=FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 100

| wbcd\_test\_pred

wbcd\_test\_labels | Benign | Malignant | Row Total |

-----------------|-----------|-----------|-----------|

Benign | 61 | 0 | 61 |

| 1.000 | 0.000 | 0.610 |

| 0.968 | 0.000 | |

| 0.610 | 0.000 | |

-----------------|-----------|-----------|-----------|

Malignant | 2 | 37 | 39 |

| 0.051 | 0.949 | 0.390 |

| 0.032 | 1.000 | |

| 0.020 | 0.370 | |

-----------------|-----------|-----------|-----------|

Column Total | 63 | 37 | 100 |

| 0.630 | 0.370 | |

-----------------|-----------|-----------|-----------|

>

> wbcd\_test\_pred <- knn(train = wbcd\_train, test = wbcd\_test, cl = wbcd\_train\_labels, k=27)

> CrossTable(x = wbcd\_test\_labels, y = wbcd\_test\_pred, prop.chisq=FALSE)

Cell Contents

|-------------------------|

| N |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 100

| wbcd\_test\_pred

wbcd\_test\_labels | Benign | Malignant | Row Total |

-----------------|-----------|-----------|-----------|

Benign | 61 | 0 | 61 |

| 1.000 | 0.000 | 0.610 |

| 0.938 | 0.000 | |

| 0.610 | 0.000 | |

-----------------|-----------|-----------|-----------|

Malignant | 4 | 35 | 39 |

| 0.103 | 0.897 | 0.390 |

| 0.062 | 1.000 | |

| 0.040 | 0.350 | |

-----------------|-----------|-----------|-----------|

Column Total | 65 | 35 | 100 |

| 0.650 | 0.350 | |

-----------------|-----------|-----------|-----------|