

# HW\_08

Preet Dabhi

2022-04-16

```
#####
# Company      : Stevens
# Project     : HW_08
# Purpose      : hclust and k-means
# First Name   : Preet
# Last Name    : Dabhi
# Id           : 10459151
# Date         : 04/16/2022

## Delete all the objects from your R- environment.
#####

#8.1
rm(list = ls())

dataSet<-read.csv("D:/SEM 3/CS 513/HW_08/wisc_bc_ContinuousVar.csv",na.strings = '?')
View(dataSet)
summary(dataSet)

##          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
##          radius_se      texture_se      perimeter_se      area_se
##  Min.   :0.1115  Min.   :0.3602  Min.   : 0.757  Min.   : 6.802
```

```

## 1st Qu.:0.2324 1st Qu.:0.8339 1st Qu.: 1.606 1st Qu.: 17.850
## Median :0.3242 Median :1.1080 Median : 2.287 Median : 24.530
## Mean   :0.4052 Mean  :1.2169 Mean  : 2.866 Mean  : 40.337
## 3rd Qu.:0.4789 3rd Qu.:1.4740 3rd Qu.: 3.357 3rd Qu.: 45.190
## Max.   :2.8730 Max.  :4.8850 Max.  :21.980 Max.  :542.200
## smoothness_se    compactness_se    concavity_se    concave.points_se
## Min.   :0.001713  Min.   :0.002252  Min.   :0.00000  Min.   :0.000000
## 1st Qu.:0.005169 1st Qu.:0.013080 1st Qu.:0.01509 1st Qu.:0.007638
## Median :0.006380  Median :0.020450  Median :0.02589  Median :0.010930
## Mean   :0.007041  Mean  :0.025478  Mean  :0.03189  Mean  :0.011796
## 3rd Qu.:0.008146 3rd Qu.:0.032450 3rd Qu.:0.04205 3rd Qu.:0.014710
## Max.   :0.031130  Max.  :0.135400  Max.  :0.39600  Max.  :0.052790
## symmetry_se     fractal_dimension_se radius_worst  texture_worst
## Min.   :0.007882  Min.   :0.0008948  Min.   : 7.93  Min.   :12.02
## 1st Qu.:0.015160 1st Qu.:0.0022480 1st Qu.:13.01 1st Qu.:21.08
## Median :0.018730  Median :0.0031870  Median :14.97  Median :25.41
## Mean   :0.020542  Mean  :0.0037949  Mean  :16.27  Mean  :25.68
## 3rd Qu.:0.023480 3rd Qu.:0.0045580 3rd Qu.:18.79 3rd Qu.:29.72
## Max.   :0.078950  Max.  :0.0298400  Max.  :36.04  Max.  :49.54
## perimeter_worst area_worst  smoothness_worst  compactness_worst
## Min.   : 50.41  Min.   :185.2  Min.   :0.07117  Min.   :0.02729
## 1st Qu.: 84.11  1st Qu.:515.3  1st Qu.:0.11660  1st Qu.:0.14720
## Median : 97.66  Median :686.5  Median :0.13130  Median :0.21190
## Mean   :107.26  Mean  :880.6  Mean  :0.13237  Mean  :0.25427
## 3rd Qu.:125.40  3rd Qu.:1084.0 3rd Qu.:0.14600  3rd Qu.:0.33910
## Max.   :251.20  Max.  :4254.0  Max.  :0.22260  Max.  :1.05800
## concavity_worst concave.points_worst symmetry_worst  fractal_dimension_worst
## Min.   :0.0000  Min.   :0.00000  Min.   :0.1565  Min.   :0.05504
## 1st Qu.:0.1145  1st Qu.:0.06493  1st Qu.:0.2504  1st Qu.:0.07146
## Median :0.2267  Median :0.09993  Median :0.2822  Median :0.08004
## Mean   :0.2722  Mean  :0.11461  Mean  :0.2901  Mean  :0.08395
## 3rd Qu.:0.3829  3rd Qu.:0.16140  3rd Qu.:0.3179  3rd Qu.:0.09208
## Max.   :1.2520  Max.  :0.29100  Max.  :0.6638  Max.  :0.20750

```

```
table(dataSet$diagnosis)
```

```

##
##      B      M
## 357 212

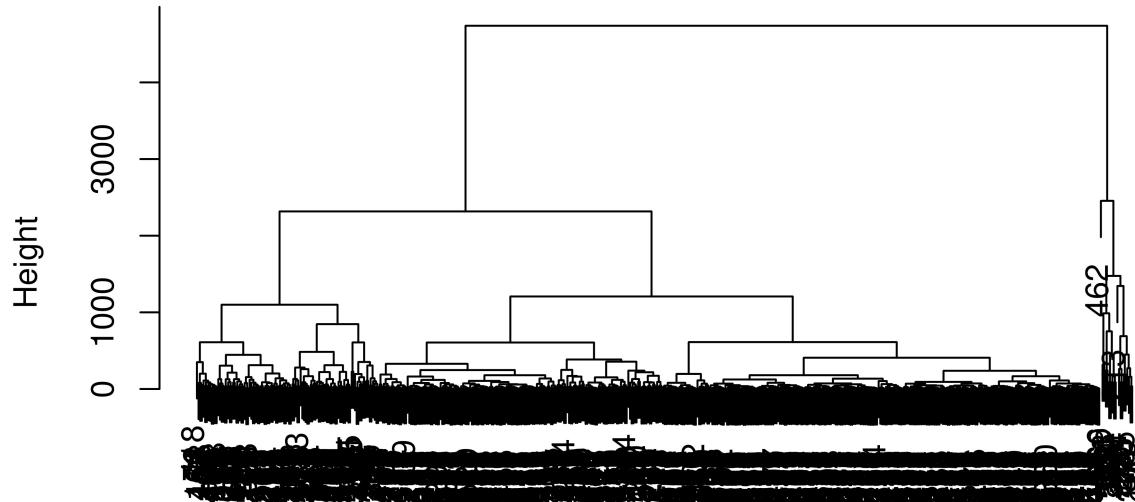
```

```

#factoring
dataSet<-na.omit(dataSet)
dataSet<-dataSet[-1]
cancerdist<-dist(dataSet[,-1])
hclustresult<-hclust(cancerdist)
plot(hclustresult)

```

## Cluster Dendrogram



```
cancerdist  
hclust (*, "complete")
```

```

hclust2<-cutree(hclustresult,2)
table(hclust2,dataSet[,1])

##
## hclust2    B    M
##      1 357 192
##      2    0   20

#8.2
rm(list=ls())

#load the file
dataSet<-read.csv("D:/SEM 3/CS 513/HW_08/wisc_bc_ContinuousVar.csv",na.strings = '?')
View(dataSet)
summary(dataSet)

##          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
## radius_se      texture_se     perimeter_se    area_se
## Min.   :0.1115  Min.   :0.3602  Min.   : 0.757  Min.   : 6.802
## 1st Qu.:0.2324 1st Qu.:0.8339 1st Qu.: 1.606 1st Qu.: 17.850
## Median :0.3242  Median :1.1080  Median : 2.287  Median : 24.530
## Mean   :0.4052  Mean   :1.2169  Mean   : 2.866  Mean   : 40.337
## 3rd Qu.:0.4789 3rd Qu.:1.4740 3rd Qu.: 3.357 3rd Qu.: 45.190
## Max.   :2.8730  Max.   :4.8850  Max.   :21.980 Max.   :542.200
## smoothness_se  compactness_se  concavity_se   concave.points_se
## Min.   :0.001713 Min.   :0.002252 Min.   :0.00000 Min.   :0.000000
## 1st Qu.:0.005169 1st Qu.:0.013080 1st Qu.:0.01509 1st Qu.:0.007638
## Median :0.006380 Median :0.020450 Median :0.02589 Median :0.010930
## Mean   :0.007041 Mean   :0.025478 Mean   :0.03189 Mean   :0.011796
## 3rd Qu.:0.008146 3rd Qu.:0.032450 3rd Qu.:0.04205 3rd Qu.:0.014710
## Max.   :0.031130 Max.   :0.135400 Max.   :0.39600 Max.   :0.052790
## symmetry_se    fractal_dimension_se radius_worst   texture_worst
## Min.   :0.007882 Min.   :0.0008948 Min.   : 7.93  Min.   :12.02
## 1st Qu.:0.015160 1st Qu.:0.0022480 1st Qu.:13.01 1st Qu.:21.08
## Median :0.018730 Median :0.0031870 Median :14.97  Median :25.41
## Mean   :0.020542 Mean   :0.0037949 Mean   :16.27  Mean   :25.68
## 3rd Qu.:0.023480 3rd Qu.:0.0045580 3rd Qu.:18.79 3rd Qu.:29.72
## Max.   :0.078950 Max.   :0.0298400 Max.   :36.04  Max.   :49.54
## perimeter_worst area_worst     smoothness_worst compactness_worst
## Min.   : 50.41  Min.   :185.2  Min.   :0.07117 Min.   :0.02729
## 1st Qu.: 84.11  1st Qu.:515.3  1st Qu.:0.11660 1st Qu.:0.14720
## Median : 97.66  Median :686.5  Median :0.13130 Median :0.21190
## Mean   :107.26  Mean   :880.6  Mean   :0.13237 Mean   :0.25427
## 3rd Qu.:125.40 3rd Qu.:1084.0 3rd Qu.:0.14600 3rd Qu.:0.33910
## Max.   :251.20  Max.   :4254.0 Max.   :0.22260 Max.   :1.05800
## concavity_worst concave.points_worst symmetry_worst fractal_dimension_worst
## Min.   :0.0000  Min.   :0.00000 Min.   :0.1565 Min.   :0.05504
## 1st Qu.:0.1145  1st Qu.:0.06493 1st Qu.:0.2504 1st Qu.:0.07146
## Median :0.2267  Median :0.09993 Median :0.2822 Median :0.08004
## Mean   :0.2722  Mean   :0.11461 Mean   :0.2901 Mean   :0.08395
## 3rd Qu.:0.3829  3rd Qu.:0.16140 3rd Qu.:0.3179 3rd Qu.:0.09208
## Max.   :1.2520  Max.   :0.29100 Max.   :0.6638 Max.   :0.20750

```

```
table(dataSet$diagnosis)
```

```
##
##   B   M
## 357 212
```

```

#To factor the data set
dataSet<-na.omit(dataSet)
dataSet<-dataSet[-1]
k_means<- kmeans(dataSet[,-1],2,nstart = 10)
k_means$cluster

## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
## 1 1 1 2 1 2 1 2 2 2 2 1 1 2 2 2 2 1 1 2
## 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
## 2 2 2 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 2
## 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
## 2 2 1 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 2
## 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
## 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 1 2 1 2 1
## 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
## 2 2 1 1 2 1 2 1 2 2 2 2 2 2 2 1 2 2 2 2
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
## 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
## 2 1 1 2 2 2 2 1 2 1 2 2 2 2 2 1 2 2 2 2
## 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
## 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
## 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
## 2 1 1 2 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
## 1 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2
## 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
## 2 1 1 2 2 2 2 1 2 2 1 2 1 2 2 2 2 2 1 1
## 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
## 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 1 1 2
## 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
## 2 2 2 2 1 2 2 2 2 2 1 2 1 1 1 2 1 2 1 2
## 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
## 1 1 1 2 1 1 2 2 2 2 2 2 1 2 1 2 2 1 2 2
## 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
## 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
## 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
## 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
## 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2
## 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
## 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2
## 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
## 2 2 2 2 2 1 1 2 1 1 2 2 1 1 2 2 2 2 2 2
## 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
## 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 2 2 2
## 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
## 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2
## 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
## 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2
## 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460
## 2 1 2 2 1 2 1 2 2 1 2 1 2 2 2 2 2 2 2 2
## 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480

```

```
##   1   1   2   2   2   2   2   2   2   1   2   2   2   2   2   2   2   2   2   2   2   2   2  
## 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500  
##   2   2   2   2   2   2   2   1   2   2   2   1   1   2   2   2   2   2   2   2   2   1   1  
## 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520  
##   2   2   2   1   2   2   2   2   2   2   2   2   2   2   2   2   2   2   1   1   2   2  
## 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540  
##   2   1   2   2   2   2   2   2   2   2   2   2   2   2   2   1   2   1   2   2   2   2  
## 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560  
##   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2   2  
## 561 562 563 564 565 566 567 568 569  
##   2   2   2   1   1   1   2   1   2
```

```
table(k_means$cluster,dataSet[,1])
```

```
##  
##          B     M  
##   1   1 130  
##   2 356  82
```