Dec

Isha Parikh

March 31, 2024

Data Understanding

This collection of data is part of the RNA-Seq (HiSeq) PANCAN data set, it is a random extraction of gene expressions of patients having different types of tumor: BRCA (Breast Cancer), KIRC (Kidney Renal Clear Cell Carcinoma), COAD (Colon Adenocarcinoma), LUAD(Lung Adenocarcinoma) and PRAD(Prostate Adenocarcinoma). Samples (instances) are stored row-wise. Variables (attributes) of each sample are RNA-Seq gene expression levels measured by illumina HiSeq platform. Some of the data points are removed(missing values) as per the the project requirement.

Data Acquisition

Loading and exploring the data and merging the actual data with the labels for better data understanding.

```r
# Load the data
data <- read.csv("https://docs.google.com/spreadsheets/d/e/2PACX-1vSx3gPbBa0-_s2mPqlXij2sTVdoF5hcp7zb5N
#str(data)

# Load the labels
label_data <- read.csv("https://docs.google.com/spreadsheets/d/e/2PACX-1vQvYplBm1JbPw-VSFB79N01SeR4lWTCI
#str(label_data)

# Merge the data with the labels
cancer_data <- merge(data, label_data, by = "X")

# Explore the data
head(cancer_data)
```

```
##            X gene_0    gene_1   gene_2    gene_3    gene_4 gene_5   gene_6
## 1   sample_0      0 2.0172093 3.265527 5.478487 10.431999      0 7.175175
## 2   sample_1      0 0.5927321 1.588421 7.586157  9.623011      0 6.816049
## 3  sample_10      0 3.4920711 3.553373 7.151707 10.253446      0 8.301258
## 4 sample_100      0 3.8828387 3.590722 5.892908  9.277211      0 7.450114
## 5 sample_101      0 4.8141659 3.928105 5.854978  8.846054      0 8.960601
## 6 sample_102      0 2.9904830 2.487358 6.938227 10.034826      0 6.910541
##      gene_7 gene_8 gene_9   gene_10   gene_11  gene_12   gene_13 gene_14
## 1 0.5918709      0      0 0.5918709 1.3342822 2.015391 0.5918709       0
## 2 0.0000000      0      0 0.0000000 0.5878450 2.466601 1.0043935       0
## 3 0.0000000      0      0 0.0000000 1.9995671 3.381962 0.0000000       0
## 4 1.0473290      0      0 0.0000000 1.7655347 2.834570 2.0703893       0
## 5 2.2638457      0      0 0.5610084 0.5610084 2.400019 0.0000000       0
```

```
## 6 0.3661403       0       0 0.6578229 0.0000000 1.108023 0.0000000       0
##   gene_15 gene_16 gene_17   gene_18   gene_19 gene_20 gene_21   gene_22
## 1       0       0       0 0.5918709  5.619994 1.334282 0.000000  9.796088
## 2       0       0       0 0.0000000 11.055208 3.562621 0.000000 10.070470
## 3       0       0       0 1.6516832  7.965224 2.146395 0.000000  8.115408
## 4       0       0       0 4.1099205  8.685334 5.008989 0.000000  9.061595
## 5       0       0       0 9.3814435 10.022507 4.544801 5.458789  8.077045
## 6       0       0       0 0.6578229 10.696203 3.039454 6.007274  8.067198
##   gene_23 gene_24    gene_25   gene_26   gene_27 gene_28  gene_29  gene_30
## 1       0       0 1.5986510 7.2151155 10.839070 6.620204 9.513538 0.000000
## 2       0       0 0.0000000 9.9498121  8.522476 1.174790 4.926991 0.000000
## 3       0       0 0.0000000 3.9200360  9.540533 7.617298 8.353231 0.000000
## 4       0       0 0.0000000 1.0473290  9.177088 5.112700 7.431957 0.000000
## 5       0       0 1.7556140 0.0000000  8.970207 4.216269 4.883645 8.310963
## 6       0       0 0.3661403 0.9004128  8.860860 5.432131 6.640087 0.000000
##     gene_31   gene_32   gene_33  gene_34   gene_35   gene_36   gene_37  gene_38
## 1 4.0636582  7.764805 4.7476559 13.71440 10.034496 0.0000000 0.0000000 9.833458
## 2 0.0000000  5.819832 1.3271700 13.28624  6.663316 0.5878450 0.0000000 9.533302
## 3 0.5144000  7.342919 2.1463945 13.92787  8.788738 0.8928184 0.0000000 9.975074
## 4 0.0000000  6.645893 0.3410749 13.55572  7.995106 0.0000000 0.0000000 9.578310
## 5 0.0000000 10.316904 1.2785795 14.46425 10.143243 5.0788916 0.0000000 9.343943
## 6 0.6578229  5.600659 7.9708192 14.11472  9.008154 0.9004128 0.3661403 9.255255
##     gene_39   gene_40 gene_41   gene_42   gene_43   gene_44 gene_45   gene_46
## 1 0.0000000 10.861265       0 0.0000000 0.5918709 1.0102786 6.962850 10.959705
## 2 0.8111422  9.728152       0 0.3236583 0.0000000 2.6512239 6.698691  9.830935
## 3 0.0000000  8.794315       0 0.0000000 0.0000000 0.0000000 4.279575  9.977538
## 4 0.0000000  9.091523       0 0.3410749 0.0000000 2.0703893 6.256760 10.891123
## 5 6.4116889  8.617218       0 0.0000000 0.0000000 0.9639919 6.766900  8.820495
## 6 0.3661403 10.556247       0 0.0000000 0.0000000 1.4505910 5.735677  9.968998
##     gene_47  gene_48   gene_49  gene_50   gene_51  gene_52   gene_53  gene_54
## 1  9.769509 12.24540 0.5918709 8.968828 10.175300 12.39906  8.479877 7.355765
## 2  8.818521 11.50562 0.3236583 9.314758  8.951270 12.35605 10.337444 8.010148
## 3  9.375407 11.52740 0.5144000 8.685758  9.354787 10.95176  9.879834 8.056269
## 4  9.917969 11.35223 1.0473290 9.113880  9.594325 11.98002  9.067346 8.268972
## 5 10.071127 10.64647 0.0000000 8.975624  9.026521 10.95498  9.827983 8.210365
## 6 10.168309 10.92434 0.3661403 8.334394  8.668403 11.37033  9.621810 8.304506
##     gene_55  gene_56  gene_57   gene_58  gene_59  gene_60  gene_61  gene_62
## 1 8.331051 10.64346 6.821225 10.637494 3.340391 6.792270 1.010279 0.000000
## 2 8.192377 11.12511 6.090792  6.803111 0.587845 6.340835 3.499846 3.866146
## 3 7.585473 10.52328 6.046956 10.016237 4.691947 6.737065 3.960299 2.279650
## 4 7.756223 10.73217 5.542258  8.974033 3.169925 7.181132 3.369229 0.616640
## 5 9.401863 10.13899 6.585989  8.498985 5.628914 6.343063 7.788444 0.000000
## 6 7.394514 10.38762 5.246997  8.538814 4.739281 7.004906 5.383041 1.450591
##     gene_63   gene_64   gene_65  gene_66  gene_67  gene_68  gene_69  gene_70
## 1 1.5986510  7.991964 10.427753 10.73430 3.478079 8.538115 3.926037 8.234750
## 2 0.3236583  9.926495  9.334065 10.21150 8.605872 6.266046 4.666637 9.107447
## 3 3.3214231  9.152401  9.978753 10.79083 5.140464 7.237841 7.040706 8.032861
## 4 3.4768094  9.918716 10.568270 11.73095 4.580947 8.351381 5.270529 9.153636
## 5 3.8860526 12.592429 11.131947 11.01917 2.843863 9.732059 5.994243 7.519880
## 6 3.7464740 11.226599 11.713498 11.76984 7.916656 8.365168 6.941388 7.916656
##     gene_71  gene_72  gene_73   gene_74  gene_75  gene_76   gene_77   gene_78
## 1 7.377931 4.511449 8.727812 0.0000000 7.377931 5.574656 3.1881635 10.685783
## 2 4.312113 4.516185 8.759459 0.5878450 6.361812 1.813607 0.3236583  9.698158
## 3 6.295374 7.021813 9.315555 0.8928184 7.136068 4.037602 2.2796496  8.572359
```

```
## 4 5.606836 3.640008 9.559568 0.3410749 6.136136 5.653441 1.7655347 10.104206
## 5 6.910625 5.683309 9.043517 0.0000000 9.310854 4.515113 0.0000000 10.183015
## 6 7.411757 6.260859 9.076906 0.6578229 7.610360 4.887442 3.3566932  9.000386
##     gene_79  gene_80  gene_81  gene_82   gene_83  gene_84   gene_85  gene_86
## 1 8.434069 10.825428 6.426765 8.691108 10.796510 4.267356 0.0000000 10.18790
## 2 9.144467 10.377004 8.771882 9.183683  2.176738 0.587845 0.3236583 11.34653
## 3 8.584936 10.016836 7.052764 8.630871 11.120879 4.871578 3.8786175 10.17650
## 4 8.792465 10.011129 7.099179 9.065909  7.309249 2.070389 0.0000000 10.36216
## 5 9.564662  8.673726 6.386909 8.051426  0.000000 0.000000 0.5610084 10.09890
## 6 8.664301  8.946137 9.234497 8.460526  7.328639 1.108023 5.3929940 10.91380
##      gene_87  gene_88   gene_89   gene_90  gene_91  gene_92  gene_93   gene_94
## 1  0.5918709  7.408398 13.230181  9.719182 4.228857 3.826406 5.997457 4.4458194
## 2  4.9273466 11.801559  8.885272  9.795850 3.137061 5.321437 7.716092 0.8111422
## 3  3.6515569  9.341510  6.612046  9.810008 2.279650 5.940282 6.197712 7.5463312
## 4  2.9386057  7.857564  6.354029 10.954342 1.047329 3.126527 5.290325 5.3291236
## 5 10.0460463 12.487860 11.642593  9.763169 5.961602 6.564413 9.153575 1.2785795
## 6  2.4149744  9.231188  7.033335 12.112707 2.491058 4.446170 5.470234 5.7199500
##     gene_95  gene_96  gene_97   gene_98  gene_99  gene_100  gene_101  gene_102
## 1 6.590142 8.647984 2.338339  9.713964 9.742431  9.067588 11.16992 10.071006
## 2 7.263748 8.518523 4.951005  9.139329 9.298777 11.463070 11.41457 11.816728
## 3 7.454176 8.299300 4.715663 10.483685 8.925943 10.625407 10.46849 10.601734
## 4 8.158862 8.653203 3.814212 12.508109 9.465838 10.786106 10.85399 10.443151
## 5 8.888637 8.731336 3.105259 10.649032 9.932032 10.350210 10.34283 10.222867
## 6 7.595354 9.470277 3.715081 11.258083 8.889264 10.422958 11.03216  9.944673
##     gene_103  gene_104 gene_105 gene_106  gene_107  gene_108  gene_109 gene_110
## 1 11.161000  9.822995 8.653447 2.824951 1.0102786 1.0102786 10.703047 11.406625
## 2 10.590905  9.256685 8.244759 1.174790 0.0000000 0.0000000  9.547434 10.940431
## 3 10.067824  9.643807 6.773284 0.000000 0.5144000 0.0000000  9.151359 11.245499
## 4 11.166872  9.638919 6.635174 1.975740 0.8479969 0.3410749  9.970681  8.777640
## 5  9.137834 11.315353 7.744349 2.843863 1.2785795 0.5610084  8.961979  8.491384
## 6 10.158496 10.843411 7.270445 2.414974 0.3661403 0.3661403  8.733493 11.644487
##    gene_111  gene_112 gene_113 gene_114  gene_115 gene_116  gene_117 gene_118
## 1 1.822037 11.940849 8.234750 9.416595 12.075780 7.282468 6.8147299 12.48070
## 2 2.328951  9.986397 7.765289 7.437078 10.462042 8.628052 0.8111422 12.16224
## 3 7.233744 10.724070 8.761937 8.036536 10.694915 8.201913 3.9200360 15.13147
## 4 0.000000 11.991593 7.973841 7.984515  9.587839 8.997555 4.1538053 12.01604
## 5 2.843863 10.226882 8.470268 6.984293 10.682802 9.059028 4.9293744 11.73461
## 6 1.289480 10.994658 9.330332 6.887452 10.481557 9.293614 3.1840263 12.30286
##    gene_119  gene_120 gene_121 gene_122 gene_123  gene_124  gene_125 gene_126
## 1 7.778629  9.577939 8.867514 7.463328 8.957096  9.764286  9.693785 6.692720
## 2 8.682945 10.651724 9.122499 7.066703 7.746823 10.615869  9.219430 7.543998
## 3 9.317494  9.344365 8.731726 6.318792 8.722975 10.628063 10.078725 9.401229
## 4 8.126188 11.212661 9.003566 6.785721 8.382840 10.386089  9.886611 7.854245
## 5 7.638204  9.696961 7.852155 9.567376 7.378676 10.127466 10.748453 9.835507
## 6 8.189296 10.537306 9.685984 7.769712 7.765886 10.711323 10.355219 8.295245
##    gene_127 gene_128 gene_129  gene_130 gene_131  gene_132 gene_133  gene_134
## 1 8.582691 10.27729  8.66791  9.439378 6.229680  8.821356 1.010279  6.412577
## 2 8.570038 10.70248 10.53478 11.380683 2.651224  9.496600 6.567101  5.815972
## 3 7.450650 11.17748 10.01026 10.903852 6.334195 10.163625 3.440062  3.549681
## 4 7.864186 11.47015 10.71281  9.955650 6.650574  9.169757 3.126527  2.887525
## 5 8.734557 11.15137 10.01656 11.891036 6.852461 12.101979 7.955231 12.042261
## 6 7.508761 11.45204 10.71739 11.890006 6.244463  7.379578 5.534585  3.039454
##      gene_135 gene_136 gene_137  gene_138 gene_139  gene_140  gene_141  gene_142
## 1 0.5918709 10.90122 9.431982 10.021896 7.559553 12.408112  8.896974 11.749480
```

```
## 2 0.0000000 12.46056 9.644480  8.410693 7.500579  9.969847  8.663387 11.327738
## 3 0.0000000 10.72954 9.304846  8.743310 6.326514 10.339449 10.130931 11.584061
## 4 1.0473290 10.54374 9.282780  9.606220 5.491853  9.762769  9.547280 11.474755
## 5 0.5610084 10.52560 8.721622  8.454715 8.740985 10.071127 12.640326  9.989834
## 6 0.3661403 11.06482 9.461436  8.687736 8.408737  9.779522  9.430969 11.670931
##    gene_143 gene_144 gene_145 gene_146 gene_147 gene_148 gene_149
## 1  9.107926  9.907700  8.126338 11.334508 10.473685  9.920593  9.798556
## 2  6.459884  9.401613  7.036613 10.169261 10.295941  4.256241  9.642665
## 3  9.351953  8.265554  7.601474  8.547431  9.684934  2.401630  9.332912
## 4  8.363114  9.676544  8.331920  8.424868 10.548032  2.536053  9.808964
## 5 10.210184  9.483642 12.390309  9.364896  9.159559 10.189775 10.873982
## 6  8.792016 10.618275  7.820134  8.430645 10.166226  2.934328  9.625013
##    gene_150 gene_151 gene_152 gene_153 gene_154  gene_155 gene_156  gene_157
## 1 11.134888 10.162593 13.84650 3.719370 5.962686  9.000600 9.747658 10.934126
## 2  7.717813  7.909605 12.35991 3.217277 8.401063 10.407778 8.996267 10.412316
## 3 10.377113  8.744750 11.41193 3.699918 7.549638  9.732936 8.520807  9.489856
## 4 11.646950  9.296304 11.44883 7.316508 4.996990 10.257305 9.671506  8.914684
## 5 10.731616 10.052514 13.37406 8.780507 6.669197  9.934838 9.851184 11.605220
## 6 10.787690  9.272205 11.47389 5.925912 6.112150 10.501817 8.925756  9.594610
##    gene_158 gene_159 gene_160  gene_161 gene_162 gene_163 gene_164 gene_165
## 1 9.406873 10.33676 8.940813  9.268395 9.399848 5.229403 1.822037 8.162537
## 2 9.378139 11.13289 8.097527  9.840466 8.710923 6.127588 1.004394 6.127588
## 3 8.486031 10.61559 8.137524  9.656116 9.260188 4.973460 7.274234 6.691925
## 4 9.252597 11.25515 8.850290 11.116649 9.480453 3.369229 0.616640 8.448187
## 5 7.539399 10.15531 9.872556 10.439332 8.442939 3.886053 1.945608 3.182819
## 6 8.220664 11.26217 7.666828  9.636577 9.279599 5.613314 2.062674 8.157539
##    gene_166  gene_167 gene_168 gene_169 gene_170 gene_171  gene_172  gene_173
## 1 3.542295 0.0000000  0.00000   0.0000 8.904629 8.996310  2.717803 1.3342822
## 2 3.562621 1.9133394  0.00000   0.0000 7.833693 9.032511  5.019729 0.0000000
## 3 2.892605 0.8928184  0.00000   0.5144 7.166495 8.469369  2.618427 1.6516832
## 4 6.820179 0.3410749  0.61664   0.0000 6.803059 9.043484  1.047329 0.3410749
## 5 3.105259 1.5367492  0.00000   0.0000 8.837085 8.308803 11.071730 0.0000000
## 6 2.159145 1.1080226  0.00000   0.0000 6.572465 7.775433  4.126453 0.6578229
##    gene_174 gene_175  gene_176 gene_177 gene_178 gene_179 gene_180 gene_181
## 1 4.189295 8.843824 10.463953 7.097168 11.33819 12.58226 0.000000 8.840636
## 2 4.694936 8.560459 10.933949 7.205930 11.29749 12.23375 1.813607 7.604916
## 3 5.396663 6.876836  8.119859 4.372019 11.40943 12.15836 2.146395 6.992984
## 4 7.330917 7.877333  9.229941 3.169925 11.99112 12.78315 2.600912 6.813358
## 5 7.462715 8.319577 10.095779 6.222611 11.56730 13.43719 9.387599 7.599831
## 6 6.260859 8.197841  9.356632 5.412697 11.40548 12.30459 0.000000 5.604777
##     gene_182 gene_183 gene_184  gene_185 gene_186 gene_187 gene_188 gene_189
## 1 10.660281 12.34402 6.771120 2.8249511 9.701785 9.051130 7.918261 5.305383
## 2 11.116805 11.67846 6.385749 0.0000000 9.850138 9.114666 8.601206 5.539965
## 3  9.920639 11.62534 5.050079 0.0000000 8.880689 9.812758 8.840888 8.693421
## 4  9.771159 11.55097 5.791814 0.0000000 9.866166 9.226732 8.704827 9.339701
## 5 10.704742 12.35966 7.823259 0.0000000 9.657697 9.363857 8.452760 8.567469
## 6  9.880686 11.06623 5.574574 0.6578229 8.992120 8.640480 8.342235 8.348640
##    gene_190 gene_191  gene_192 gene_193 gene_194  gene_195  gene_196  gene_197
## 1 2.185898 9.149747 10.365951 9.650743 11.16036  6.287023 10.081177 6.8658199
## 2 1.174790 8.719560  9.347012 9.306249 10.23886  7.213289  8.055863 1.5908176
## 3 2.973519 8.187144  8.570687 8.339921 10.03778  6.520762 10.350851 2.1463945
## 4 2.600912 8.214319  9.265781 8.313187 10.17330  6.934670 11.368905 1.2223718
## 5 3.687453 9.214868  7.854912 8.854968 10.73275  7.141862  8.287015 0.9639919
## 6 1.108023 8.979542  8.085786 8.543287 10.42067 10.255784  9.789446 1.4505910
```

```
##    gene_198  gene_199 gene_200  gene_201 gene_202  gene_203   gene_204 gene_205
## 1 11.22720 11.526059 12.00741 10.054591 8.384577 15.673080  1.5986510 5.149202
## 2 10.79442 10.693513 11.36865  9.226716 5.459786  9.184930  0.3236583 4.815816
## 3 10.68068  9.889700  9.44445  8.444746 9.362319  3.920036  0.0000000 4.372019
## 4 10.96996 10.784144 10.59439  9.005063 8.385142  5.056007  0.0000000 5.898854
## 5 10.82530  9.553752 10.92001  8.160905 7.794637  3.687453  0.0000000 6.127726
## 6 10.42114 10.450325 10.52746  8.110060 9.507398  4.844566  0.0000000 4.102389
##    gene_206  gene_207 gene_208 gene_209 gene_210  gene_211   gene_212  gene_213
## 1 3.773880 1.3342822 2.824951 4.304861 0.000000  8.743262 10.145244 11.772875
## 2 3.293282 2.0941685 0.587845 1.590818 1.004394 11.047172  6.722930 12.084586
## 3 3.960299 0.8928184 2.146395 2.514097 1.440208 10.496145  8.383143 11.318464
## 4 3.758442 0.3410749 2.834570 2.779617 1.378512  9.938208  6.449561  9.816664
## 5 4.762551 2.1135005 4.322916 1.755614 1.945608  6.311638  6.354172 12.974085
## 6 3.039454 0.3661403 1.959288 3.229173 1.450591  8.913667        NA 11.787752
##     gene_214  gene_215  gene_216 gene_217  gene_218   gene_219  gene_220
## 1 11.009885  8.067547 10.308771 4.106767 13.966992  0.5918709 0.5918709
## 2 10.904725 11.497742 10.597950 1.590818  2.176738  0.0000000 0.0000000
## 3 11.832043  9.197354  8.277552 3.122888  6.295374  0.0000000 0.0000000
## 4 14.439669  8.559442  7.377931 4.531899  3.944858  0.3410749 0.8479969
## 5  9.851927 10.124936  8.182310 1.536749  3.023273 11.8785223 11.7374290
## 6 11.558947  9.824340  9.623956 1.959288  3.184026  0.0000000 0.0000000
##     gene_221  gene_222 gene_223  gene_224  gene_225 gene_226 gene_227 gene_228
## 1 10.007966 0.0000000 3.188164 12.235880  9.838262 5.775011 3.926037 15.58264
## 2  6.375557 0.5878450 3.651924  9.299355 10.062694 4.680853 0.000000 10.71137
## 3  5.779680 0.0000000 3.495913 11.233224 10.035212 5.711726 3.381962 13.38295
## 4  3.094016 0.0000000 4.153805 11.436259  9.608685 5.955748 1.378512 12.60174
## 5  8.868338 0.5610084 9.445791 10.899334 10.249019 9.116988 2.113500 12.77973
## 6  5.345506 0.0000000 4.754802  9.934487  9.393661 7.987332 6.136050 13.81969
##    gene_229 gene_230  gene_231 gene_232  gene_233  gene_234  gene_235  gene_236
## 1 0.000000 16.27021 10.167694 15.79891 15.358229  8.979014 1.3342822 0.0000000
## 2 1.004394 16.65011  0.000000 16.74817  6.213225 10.565273 0.3236583 0.0000000
## 3 0.000000 16.94813  0.514400 16.54958  9.306800 10.276555 0.0000000 0.0000000
## 4 0.000000 16.38782  3.126527 16.30865  6.850499 11.393138 0.0000000 0.0000000
## 5 0.000000 15.32091  3.739729 14.46974  5.835174  8.409047 0.0000000 0.0000000
## 6 1.959288 17.00345  2.334625 16.00433 10.073633 10.122310 0.0000000 0.3661403
##      gene_237  gene_238 gene_239 gene_240 gene_241   gene_242 gene_243  gene_244
## 1 0.0000000 1.8220367        0 11.97760 4.411847 2.0153905 12.82509 10.308202
## 2 0.3236583 0.3236583        0 12.32073 2.006585 2.1767376 13.72191  9.788893
## 3 0.5144000 1.6516832        0 12.62217 2.618427 1.4402081 13.52004  9.980597
## 4 0.0000000 1.0473290        0 12.57249 2.242846 1.0473290 13.94446 10.386800
## 5 0.0000000 0.0000000        0 11.51835 3.457095 0.9639919 13.17563  9.865957
## 6 0.0000000 4.8874424        0 12.97619 1.108023 1.4505910 13.16059 10.028058
##    gene_245 gene_246 gene_247 gene_248 gene_249 gene_250 gene_251 gene_252
## 1 12.02261 11.11226 11.39064 8.177778 5.511901 11.07306 7.837395 9.003453
## 2 11.30941 10.11271 12.38889 7.411731 5.860369 12.71132 8.247149 9.363559
## 3 11.67287 10.12596 12.33467 8.072433 5.381917 11.70839 8.009067 8.428695
## 4 11.12322 10.59861 12.58682 7.407693 3.292782 11.77289 8.606590 8.790730
## 5 10.67485 10.70105 12.17631 7.861031 6.267811 11.57182 7.617394 8.750570
## 6 11.30595 10.19770 12.85368 7.980751 5.679863 12.15078 7.938903 9.216549
##    gene_253 gene_254 Class
## 1 8.576953        0  PRAD
## 2 8.173652        0  LUAD
## 3 8.245329        0  BRCA
## 4 8.177584        0  BRCA
```

```
## 5 8.353517          0  KIRC
## 6 9.054352          0  BRCA
```

str(cancer_data)

```
## 'data.frame':    801 obs. of   257 variables:
##  $ X       : chr  "sample_0" "sample_1" "sample_10" "sample_100" ...
##  $ gene_0  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_1  : num  2.017 0.593 3.492 3.883 4.814 ...
##  $ gene_2  : num  3.27 1.59 3.55 3.59 3.93 ...
##  $ gene_3  : num  5.48 7.59 7.15 5.89 5.85 ...
##  $ gene_4  : num  10.43 9.62 10.25 9.28 8.85 ...
##  $ gene_5  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_6  : num  7.18 6.82 8.3 7.45 8.96 ...
##  $ gene_7  : num  0.592 0 0 1.047 2.264 ...
##  $ gene_8  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_9  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_10 : num  0.592 0 0 0 0.561 ...
##  $ gene_11 : num  1.334 0.588 2 1.766 0.561 ...
##  $ gene_12 : num  2.02 2.47 3.38 2.83 2.4 ...
##  $ gene_13 : num  0.592 1.004 0 2.07 0 ...
##  $ gene_14 : num  0 0 0 0 0 ...
##  $ gene_15 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_16 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_17 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_18 : num  0.592 0 1.652 4.11 9.381 ...
##  $ gene_19 : num  5.62 11.06 7.97 8.69 10.02 ...
##  $ gene_20 : num  1.33 3.56 2.15 5.01 4.54 ...
##  $ gene_21 : num  0 0 0 0 5.46 ...
##  $ gene_22 : num  9.8 10.07 8.12 9.06 8.08 ...
##  $ gene_23 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_24 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_25 : num  1.6 0 0 0 1.76 ...
##  $ gene_26 : num  7.22 9.95 3.92 1.05 0 ...
##  $ gene_27 : num  10.84 8.52 9.54 9.18 8.97 ...
##  $ gene_28 : num  6.62 1.17 7.62 5.11 4.22 ...
##  $ gene_29 : num  9.51 4.93 8.35 7.43 4.88 ...
##  $ gene_30 : num  0 0 0 0 8.31 ...
##  $ gene_31 : num  4.064 0 0.514 0 0 ...
##  $ gene_32 : num  7.76 5.82 7.34 6.65 10.32 ...
##  $ gene_33 : num  4.748 1.327 2.146 0.341 1.279 ...
##  $ gene_34 : num  13.7 13.3 13.9 13.6 14.5 ...
##  $ gene_35 : num  10.03 6.66 8.79 8 10.14 ...
##  $ gene_36 : num  0 0.588 0.893 0 5.079 ...
##  $ gene_37 : num  0 0 0 0 0 ...
##  $ gene_38 : num  9.83 9.53 9.98 9.58 9.34 ...
##  $ gene_39 : num  0 0.811 0 0 6.412 ...
##  $ gene_40 : num  10.86 9.73 8.79 9.09 8.62 ...
##  $ gene_41 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_42 : num  0 0.324 0 0.341 0 ...
##  $ gene_43 : num  0.592 0 0 0 0 ...
##  $ gene_44 : num  1.01 2.651 0 2.07 0.964 ...
##  $ gene_45 : num  6.96 6.7 4.28 6.26 6.77 ...
##  $ gene_46 : num  10.96 9.83 9.98 10.89 8.82 ...
```

```
## $ gene_47 : num  9.77 8.82 9.38 9.92 10.07 ...
## $ gene_48 : num  12.2 11.5 11.5 11.4 10.6 ...
## $ gene_49 : num  0.592 0.324 0.514 1.047 0 ...
## $ gene_50 : num  8.97 9.31 8.69 9.11 8.98 ...
## $ gene_51 : num  10.18 8.95 9.35 9.59 9.03 ...
## $ gene_52 : num  12.4 12.4 11 12 11 ...
## $ gene_53 : num  8.48 10.34 9.88 9.07 9.83 ...
## $ gene_54 : num  7.36 8.01 8.06 8.27 8.21 ...
## $ gene_55 : num  8.33 8.19 7.59 7.76 9.4 ...
## $ gene_56 : num  10.6 11.1 10.5 10.7 10.1 ...
## $ gene_57 : num  6.82 6.09 6.05 5.54 6.59 ...
## $ gene_58 : num  10.64 6.8 10.02 8.97 8.5 ...
## $ gene_59 : num  3.34 0.588 4.692 3.17 5.629 ...
## $ gene_60 : num  6.79 6.34 6.74 7.18 6.34 ...
## $ gene_61 : num  1.01 3.5 3.96 3.37 7.79 ...
## $ gene_62 : num  0 3.866 2.28 0.617 0 ...
## $ gene_63 : num  1.599 0.324 3.321 3.477 3.886 ...
## $ gene_64 : num  7.99 9.93 9.15 9.92 12.59 ...
## $ gene_65 : num  10.43 9.33 9.98 10.57 11.13 ...
## $ gene_66 : num  10.7 10.2 10.8 11.7 11 ...
## $ gene_67 : num  3.48 8.61 5.14 4.58 2.84 ...
## $ gene_68 : num  8.54 6.27 7.24 8.35 9.73 ...
## $ gene_69 : num  3.93 4.67 7.04 5.27 5.99 ...
## $ gene_70 : num  8.23 9.11 8.03 9.15 7.52 ...
## $ gene_71 : num  7.38 4.31 6.3 5.61 6.91 ...
## $ gene_72 : num  4.51 4.52 7.02 3.64 5.68 ...
## $ gene_73 : num  8.73 8.76 9.32 9.56 9.04 ...
## $ gene_74 : num  0 0.588 0.893 0.341 0 ...
## $ gene_75 : num  7.38 6.36 7.14 6.14 9.31 ...
## $ gene_76 : num  5.57 1.81 4.04 5.65 4.52 ...
## $ gene_77 : num  3.188 0.324 2.28 1.766 0 ...
## $ gene_78 : num  10.69 9.7 8.57 10.1 10.18 ...
## $ gene_79 : num  8.43 9.14 8.58 8.79 9.56 ...
## $ gene_80 : num  10.83 10.38 10.02 10.01 8.67 ...
## $ gene_81 : num  6.43 8.77 7.05 7.1 6.39 ...
## $ gene_82 : num  8.69 9.18 8.63 9.07 8.05 ...
## $ gene_83 : num  10.8 2.18 11.12 7.31 0 ...
## $ gene_84 : num  4.267 0.588 4.872 2.07 0 ...
## $ gene_85 : num  0 0.324 3.879 0 0.561 ...
## $ gene_86 : num  10.2 11.3 10.2 10.4 10.1 ...
## $ gene_87 : num  0.592 4.927 3.652 2.939 10.046 ...
## $ gene_88 : num  7.41 11.8 9.34 7.86 12.49 ...
## $ gene_89 : num  13.23 8.89 6.61 6.35 11.64 ...
## $ gene_90 : num  9.72 9.8 9.81 10.95 9.76 ...
## $ gene_91 : num  4.23 3.14 2.28 1.05 5.96 ...
## $ gene_92 : num  3.83 5.32 5.94 3.13 6.56 ...
## $ gene_93 : num  6 7.72 6.2 5.29 9.15 ...
## $ gene_94 : num  4.446 0.811 7.546 5.329 1.279 ...
## $ gene_95 : num  6.59 7.26 7.45 8.16 8.89 ...
## $ gene_96 : num  8.65 8.52 8.3 8.65 8.73 ...
## $ gene_97 : num  2.34 4.95 4.72 3.81 3.11 ...
##   [list output truncated]
```

```r
summary(cancer_data)
```

```
##       X                gene_0            gene_1          gene_2
##  Length:801         Min.   :0.00000   Min.   :0.000   Min.   :0.000
##  Class :character   1st Qu.:0.00000   1st Qu.:2.299   1st Qu.:2.389
##  Mode  :character   Median :0.00000   Median :3.144   Median :3.121
##                     Mean   :0.02664   Mean   :3.011   Mean   :3.095
##                     3rd Qu.:0.00000   3rd Qu.:3.883   3rd Qu.:3.803
##                     Max.   :1.48233   Max.   :6.237   Max.   :6.063
##                                                       NA's   :1
##      gene_3           gene_4           gene_5       gene_6
##  Min.   : 5.009   Min.   : 8.436   Min.   :0   Min.   : 3.931
##  1st Qu.: 6.303   1st Qu.: 9.464   1st Qu.:0   1st Qu.: 6.676
##  Median : 6.656   Median : 9.792   Median :0   Median : 7.450
##  Mean   : 6.722   Mean   : 9.814   Mean   :0   Mean   : 7.406
##  3rd Qu.: 7.038   3rd Qu.:10.142   3rd Qu.:0   3rd Qu.: 8.122
##  Max.   :10.130   Max.   :11.356   Max.   :0   Max.   :10.718
##                                    NA's   :1
##      gene_7           gene_8            gene_9           gene_10
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   : 0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.: 0.0000
##  Median :0.4431   Median :0.00000   Median :0.00000   Median : 0.3602
##  Mean   :0.4999   Mean   :0.01674   Mean   :0.01343   Mean   : 0.6882
##  3rd Qu.:0.7894   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.: 0.9039
##  Max.   :2.7790   Max.   :1.78559   Max.   :4.06760   Max.   :12.2930
##
##      gene_11          gene_12          gene_13          gene_14
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.3958   1st Qu.:2.244   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.8011   Median :2.702   Median :0.0000   Median :0.0000
##  Mean   :0.9555   Mean   :2.667   Mean   :0.3181   Mean   :0.2146
##  3rd Qu.:1.3739   3rd Qu.:3.153   3rd Qu.:0.4276   3rd Qu.:0.3998
##  Max.   :7.8844   Max.   :4.784   Max.   :5.5479   Max.   :2.9918
##
##      gene_15            gene_16            gene_17          gene_18
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.0000   Min.   : 0.0000
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.0000   1st Qu.: 0.0000
##  Median :0.000000   Median :0.000000   Median :0.0000   Median : 0.6392
##  Mean   :0.005879   Mean   :0.003002   Mean   :0.0355   Mean   : 2.2044
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.0000   3rd Qu.: 2.7044
##  Max.   :1.000000   Max.   :0.581882   Max.   :2.5654   Max.   :10.8616
##
##      gene_19          gene_20          gene_21          gene_22          gene_23
##  Min.   : 5.437   Min.   :0.000   Min.   :0.000   Min.   : 4.190   Min.   :0
##  1st Qu.: 8.287   1st Qu.:2.589   1st Qu.:0.000   1st Qu.: 7.611   1st Qu.:0
##  Median : 9.158   Median :3.309   Median :2.435   Median : 8.263   Median :0
##  Mean   : 9.079   Mean   :3.286   Mean   :2.556   Mean   : 8.266   Mean   :0
##  3rd Qu.: 9.931   3rd Qu.:4.043   3rd Qu.:4.913   3rd Qu.: 9.015   3rd Qu.:0
##  Max.   :12.463   Max.   :7.165   Max.   :7.781   Max.   :13.135   Max.   :0
##
##      gene_24          gene_25          gene_26          gene_27
##  Min.   :0.0000   Min.   :0.0000   Min.   : 0.0000   Min.   : 6.201
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.4788   1st Qu.: 8.797
```

```
##    Median :0.0000     Median :0.0000     Median : 2.0419     Median : 9.247
##    Mean   :0.1055     Mean   :0.5265     Mean   : 2.7225     Mean   : 9.209
##    3rd Qu.:0.0000     3rd Qu.:0.8832     3rd Qu.: 4.2937     3rd Qu.: 9.673
##    Max.   :3.6025     Max.   :5.4982     Max.   :13.2648     Max.   :11.538
##
##      gene_28             gene_29             gene_30             gene_31
##    Min.   : 0.000     Min.   : 2.204     Min.   :0.000     Min.   :0.0000
##    1st Qu.: 3.053     1st Qu.: 4.995     1st Qu.:0.000     1st Qu.:0.0000
##    Median : 4.999     Median : 6.111     Median :0.000     Median :0.4789
##    Mean   : 4.674     Mean   : 6.183     Mean   :2.139     Mean   :1.1686
##    3rd Qu.: 6.290     3rd Qu.: 7.208     3rd Qu.:5.285     3rd Qu.:1.6077
##    Max.   :10.251     Max.   :13.105     Max.   :9.963     Max.   :9.5205
##
##      gene_32             gene_33             gene_34             gene_35
##    Min.   : 4.238     Min.   : 0.0000     Min.   : 9.378     Min.   : 4.095
##    1st Qu.: 6.253     1st Qu.: 0.3107     1st Qu.:12.894     1st Qu.: 7.680
##    Median : 6.757     Median : 1.2527     Median :13.692     Median : 8.574
##    Mean   : 6.790     Mean   : 2.2843     Mean   :13.711     Mean   : 8.545
##    3rd Qu.: 7.305     3rd Qu.: 3.2886     3rd Qu.:14.635     3rd Qu.: 9.601
##    Max.   :10.654     Max.   :13.3117     Max.   :17.366     Max.   :11.832
##
##      gene_36             gene_37             gene_38             gene_39
##    Min.   : 0.0000     Min.   :0.0000     Min.   : 7.191     Min.   :0.0000
##    1st Qu.: 0.0000     1st Qu.:0.0000     1st Qu.: 9.223     1st Qu.:0.0000
##    Median : 0.6599     Median :0.0000     Median : 9.513     Median :0.4609
##    Mean   : 1.1236     Mean   :0.2023     Mean   : 9.498     Mean   :1.3633
##    3rd Qu.: 1.6029     3rd Qu.:0.0000     3rd Qu.: 9.769     3rd Qu.:2.1434
##    Max.   :11.6154     Max.   :5.5165     Max.   :10.862     Max.   :8.5023
##                                           NA's   :1
##      gene_40             gene_41             gene_42             gene_43
##    Min.   : 7.596     Min.   :0.0000     Min.   :0.00000     Min.   :0.0000
##    1st Qu.: 9.276     1st Qu.:0.0000     1st Qu.:0.00000     1st Qu.:0.0000
##    Median : 9.760     Median :0.0000     Median :0.00000     Median :0.0000
##    Mean   : 9.746     Mean   :0.1081     Mean   :0.07864     Mean   :0.1505
##    3rd Qu.:10.208     3rd Qu.:0.0000     3rd Qu.:0.00000     3rd Qu.:0.0000
##    Max.   :14.155     Max.   :7.7602     Max.   :4.59928     Max.   :2.2297
##
##      gene_44             gene_45             gene_46             gene_47
##    Min.   : 0.0000     Min.   : 1.626     Min.   : 7.235     Min.   : 6.990
##    1st Qu.: 0.0000     1st Qu.: 5.377     1st Qu.: 9.702     1st Qu.: 9.466
##    Median : 0.6258     Median : 6.392     Median :10.154     Median : 9.895
##    Mean   : 1.6094     Mean   : 6.442     Mean   :10.141     Mean   : 9.885
##    3rd Qu.: 2.6412     3rd Qu.: 7.580     3rd Qu.:10.614     3rd Qu.:10.328
##    Max.   :11.6341     Max.   :11.275     Max.   :12.101     Max.   :12.804
##                                           NA's   :1
##      gene_48             gene_49             gene_50             gene_51
##    Min.   : 8.053     Min.   :0.0000     Min.   : 0.000     Min.   : 6.445
##    1st Qu.:10.987     1st Qu.:0.0000     1st Qu.: 8.589     1st Qu.: 8.722
##    Median :11.313     Median :0.4165     Median : 8.901     Median : 9.079
##    Mean   :11.287     Mean   :0.5068     Mean   : 8.858     Mean   : 9.098
##    3rd Qu.:11.617     3rd Qu.:0.8431     3rd Qu.: 9.187     3rd Qu.: 9.432
##    Max.   :12.738     Max.   :2.6146     Max.   :10.428     Max.   :11.314
##                                                             NA's   :1
##      gene_52             gene_53             gene_54             gene_55
```

```
##    Min.   : 9.673   Min.   : 7.640   Min.   : 5.792   Min.   : 3.519
##    1st Qu.:11.211   1st Qu.: 9.402   1st Qu.: 7.657   1st Qu.: 7.115
##    Median :11.575   Median : 9.713   Median : 7.997   Median : 8.153
##    Mean   :11.584   Mean   : 9.720   Mean   : 7.990   Mean   : 7.982
##    3rd Qu.:11.946   3rd Qu.:10.037   3rd Qu.: 8.340   3rd Qu.: 8.968
##    Max.   :13.505   Max.   :12.125   Max.   :10.812   Max.   :10.940
##
##     gene_56          gene_57          gene_58          gene_59
##    Min.   : 8.242   Min.   : 2.223   Min.   : 3.315   Min.   :0.000
##    1st Qu.: 9.962   1st Qu.: 5.343   1st Qu.: 7.982   1st Qu.:2.581
##    Median :10.237   Median : 6.218   Median : 9.169   Median :3.499
##    Mean   :10.278   Mean   : 6.316   Mean   : 9.259   Mean   :3.564
##    3rd Qu.:10.554   3rd Qu.: 7.116   3rd Qu.:10.720   3rd Qu.:4.537
##    Max.   :12.931   Max.   :11.383   Max.   :13.602   Max.   :8.212
##    NA's   :1
##     gene_60          gene_61          gene_62           gene_63
##    Min.   :1.939   Min.   : 0.000   Min.   : 0.0000   Min.   :0.000
##    1st Qu.:6.042   1st Qu.: 2.344   1st Qu.: 0.6845   1st Qu.:2.049
##    Median :6.561   Median : 5.864   Median : 1.8151   Median :2.987
##    Mean   :6.508   Mean   : 5.388   Mean   : 2.7100   Mean   :3.184
##    3rd Qu.:7.005   3rd Qu.: 8.027   3rd Qu.: 4.2324   3rd Qu.:4.105
##    Max.   :8.533   Max.   :13.563   Max.   :11.6473   Max.   :9.556
##
##     gene_64          gene_65          gene_66          gene_67
##    Min.   : 5.556   Min.   : 5.712   Min.   : 3.557   Min.   : 0.000
##    1st Qu.: 8.889   1st Qu.: 9.975   1st Qu.: 9.875   1st Qu.: 2.098
##    Median : 9.717   Median :10.465   Median :10.518   Median : 3.415
##    Mean   : 9.688   Mean   :10.386   Mean   :10.362   Mean   : 3.971
##    3rd Qu.:10.536   3rd Qu.:10.932   3rd Qu.:11.350   3rd Qu.: 5.306
##    Max.   :13.326   Max.   :12.518   Max.   :15.180   Max.   :12.787
##    NA's   :1
##     gene_68          gene_69          gene_70          gene_71
##    Min.   : 3.162   Min.   : 0.000   Min.   : 3.831   Min.   : 0.000
##    1st Qu.: 7.600   1st Qu.: 3.824   1st Qu.: 7.540   1st Qu.: 4.161
##    Median : 8.365   Median : 5.333   Median : 8.235   Median : 6.284
##    Mean   : 8.255   Mean   : 5.248   Mean   : 8.210   Mean   : 5.806
##    3rd Qu.: 9.057   3rd Qu.: 6.635   3rd Qu.: 8.908   3rd Qu.: 7.659
##    Max.   :11.182   Max.   :11.287   Max.   :11.534   Max.   :10.865
##    NA's   :1
##     gene_72          gene_73          gene_74           gene_75
##    Min.   :0.000   Min.   : 6.882   Min.   :0.0000   Min.   : 0.9781
##    1st Qu.:3.677   1st Qu.: 8.698   1st Qu.:0.0000   1st Qu.: 5.4547
##    Median :4.820   Median : 9.010   Median :0.5002   Median : 6.9143
##    Mean   :4.809   Mean   : 9.013   Mean   :0.7965   Mean   : 6.9580
##    3rd Qu.:6.034   3rd Qu.: 9.333   3rd Qu.:1.2490   3rd Qu.: 8.4612
##    Max.   :9.808   Max.   :10.878   Max.   :9.9009   Max.   :13.8611
##                                                       NA's   :3
##     gene_76          gene_77          gene_78          gene_79
##    Min.   : 0.000   Min.   : 0.0000   Min.   : 5.265   Min.   : 7.004
##    1st Qu.: 3.753   1st Qu.: 0.4718   1st Qu.: 8.670   1st Qu.: 8.494
##    Median : 4.662   Median : 1.3433   Median : 9.367   Median : 8.794
##    Mean   : 4.755   Mean   : 1.7510   Mean   : 9.332   Mean   : 8.776
##    3rd Qu.: 5.550   3rd Qu.: 2.4615   3rd Qu.:10.010   3rd Qu.: 9.074
##    Max.   :11.322   Max.   :12.2313   Max.   :12.501   Max.   :11.051
```

```
##
##     gene_80           gene_81           gene_82           gene_83
##  Min.   : 6.960   Min.   : 2.919   Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 9.005   1st Qu.: 6.337   1st Qu.: 8.348   1st Qu.: 1.444
##  Median : 9.425   Median : 6.982   Median : 8.776   Median : 3.013
##  Mean   : 9.406   Mean   : 7.055   Mean   : 8.796   Mean   : 4.579
##  3rd Qu.: 9.827   3rd Qu.: 7.696   3rd Qu.: 9.281   3rd Qu.: 7.799
##  Max.   :11.871   Max.   :10.700   Max.   :10.963   Max.   :14.402
##                   NA's   :1
##     gene_84           gene_85           gene_86           gene_87
##  Min.   : 0.000   Min.   : 0.0000   Min.   : 7.853   Min.   : 0.000
##  1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.:10.194   1st Qu.: 2.589
##  Median : 0.000   Median : 0.7031   Median :10.601   Median : 3.567
##  Mean   : 1.255   Mean   : 1.6241   Mean   :10.619   Mean   : 4.678
##  3rd Qu.: 1.733   3rd Qu.: 2.5032   3rd Qu.:11.071   3rd Qu.: 6.404
##  Max.   :10.657   Max.   :11.5759   Max.   :13.995   Max.   :13.066
##                   NA's   :1
##     gene_88           gene_89           gene_90           gene_91
##  Min.   : 2.780   Min.   : 1.243   Min.   : 7.432   Min.   : 0.000
##  1st Qu.: 7.960   1st Qu.: 7.414   1st Qu.: 9.157   1st Qu.: 1.289
##  Median : 9.574   Median : 9.230   Median : 9.693   Median : 3.224
##  Mean   : 9.726   Mean   : 9.341   Mean   : 9.840   Mean   : 3.431
##  3rd Qu.:11.705   3rd Qu.:10.767   3rd Qu.:10.498   3rd Qu.: 5.467
##  Max.   :14.205   Max.   :16.068   Max.   :13.310   Max.   :10.226
##
##     gene_92           gene_93           gene_94           gene_95
##  Min.   :0.000    Min.   : 0.000   Min.   : 0.000   Min.   : 1.675
##  1st Qu.:3.429    1st Qu.: 5.097   1st Qu.: 1.001   1st Qu.: 6.670
##  Median :4.881    Median : 7.109   Median : 3.392   Median : 7.768
##  Mean   :4.785    Mean   : 6.803   Mean   : 3.923   Mean   : 7.611
##  3rd Qu.:6.291    3rd Qu.: 8.534   3rd Qu.: 6.669   3rd Qu.: 8.684
##  Max.   :8.791    Max.   :12.765   Max.   :11.009   Max.   :12.120
##
##     gene_96           gene_97           gene_98           gene_99
##  Min.   : 6.414   Min.   :0.000    Min.   : 8.707   Min.   : 6.719
##  1st Qu.: 8.324   1st Qu.:2.694    1st Qu.:10.042   1st Qu.: 8.692
##  Median : 8.755   Median :3.738    Median :10.454   Median : 9.137
##  Mean   : 8.768   Mean   :3.758    Mean   :10.555   Mean   : 9.080
##  3rd Qu.: 9.173   3rd Qu.:4.821    3rd Qu.:11.040   3rd Qu.: 9.455
##  Max.   :12.717   Max.   :8.267    Max.   :13.319   Max.   :11.133
##                                                     NA's   :1
##     gene_100          gene_101          gene_102          gene_103
##  Min.   : 8.78    Min.   : 8.75    Min.   : 8.853   Min.   : 7.637
##  1st Qu.:10.16    1st Qu.:10.44    1st Qu.:10.034   1st Qu.: 9.917
##  Median :10.49    Median :10.70    Median :10.286   Median :10.147
##  Mean   :10.54    Mean   :10.71    Mean   :10.318   Mean   :10.150
##  3rd Qu.:10.85    3rd Qu.:11.00    3rd Qu.:10.571   3rd Qu.:10.396
##  Max.   :13.30    Max.   :12.37    Max.   :12.005   Max.   :11.492
##  NA's   :1
##     gene_104          gene_105          gene_106          gene_107
##  Min.   : 5.938   Min.   : 0.9057   Min.   :0.000    Min.   : 0.0000
##  1st Qu.: 9.158   1st Qu.: 5.2965   1st Qu.:1.392    1st Qu.: 0.7571
##  Median : 9.865   Median : 6.5569   Median :1.929    Median : 1.3025
##  Mean   : 9.778   Mean   : 6.4797   Mean   :2.036    Mean   : 1.5237
```

```
##    3rd Qu.:10.516   3rd Qu.: 7.6360   3rd Qu.:2.540   3rd Qu.: 2.0706
##    Max.   :12.129   Max.   :10.9234   Max.   :8.285   Max.   :10.3216
##    NA's   :1                                          NA's   :2
##      gene_108          gene_109          gene_110          gene_111
##    Min.   : 0.0000   Min.   : 7.441   Min.   : 7.087   Min.   : 0.0000
##    1st Qu.: 0.0000   1st Qu.: 9.095   1st Qu.: 9.236   1st Qu.: 0.6948
##    Median : 0.0000   Median : 9.465   Median : 9.918   Median : 1.5466
##    Mean   : 0.6283   Mean   : 9.509   Mean   : 9.874   Mean   : 2.1049
##    3rd Qu.: 0.7393   3rd Qu.: 9.936   3rd Qu.:10.505   3rd Qu.: 2.5586
##    Max.   :10.5636   Max.   :11.872   Max.   :12.739   Max.   :10.2552
##
##      gene_112          gene_113          gene_114          gene_115
##    Min.   : 9.051   Min.   : 6.797   Min.   : 4.979   Min.   : 7.43
##    1st Qu.:10.317   1st Qu.: 8.518   1st Qu.: 7.568   1st Qu.:10.25
##    Median :10.701   Median : 8.829   Median : 8.119   Median :10.69
##    Mean   :10.772   Mean   : 8.808   Mean   : 8.113   Mean   :10.65
##    3rd Qu.:11.137   3rd Qu.: 9.117   3rd Qu.: 8.620   3rd Qu.:11.11
##    Max.   :14.269   Max.   :10.510   Max.   :11.620   Max.   :13.22
##
##      gene_116          gene_117          gene_118          gene_119
##    Min.   : 5.678   Min.   :0.000   Min.   : 8.779   Min.   : 6.030
##    1st Qu.: 7.973   1st Qu.:2.682   1st Qu.:11.405   1st Qu.: 8.165
##    Median : 8.500   Median :3.889   Median :12.084   Median : 8.758
##    Mean   : 8.441   Mean   :3.861   Mean   :12.228   Mean   : 8.752
##    3rd Qu.: 8.915   3rd Qu.:5.006   3rd Qu.:12.983   3rd Qu.: 9.317
##    Max.   :11.383   Max.   :8.596   Max.   :16.263   Max.   :11.742
##                     NA's   :1
##      gene_120          gene_121          gene_122          gene_123
##    Min.   : 7.127   Min.   : 7.564   Min.   : 4.640   Min.   : 3.571
##    1st Qu.: 9.330   1st Qu.: 8.625   1st Qu.: 6.806   1st Qu.: 7.483
##    Median : 9.746   Median : 9.048   Median : 7.372   Median : 7.968
##    Mean   : 9.716   Mean   : 9.086   Mean   : 7.553   Mean   : 7.896
##    3rd Qu.:10.134   3rd Qu.: 9.549   3rd Qu.: 8.138   3rd Qu.: 8.456
##    Max.   :12.288   Max.   :10.959   Max.   :11.189   Max.   :10.282
##                                      NA's   :1
##      gene_124          gene_125          gene_126          gene_127
##    Min.   : 8.97   Min.   : 7.250   Min.   : 1.856   Min.   : 4.590
##    1st Qu.:10.22   1st Qu.: 9.914   1st Qu.: 6.783   1st Qu.: 6.903
##    Median :10.46   Median :10.287   Median : 8.300   Median : 7.695
##    Mean   :10.47   Mean   :10.223   Mean   : 8.130   Mean   : 7.718
##    3rd Qu.:10.71   3rd Qu.:10.594   3rd Qu.: 9.642   3rd Qu.: 8.542
##    Max.   :12.35   Max.   :12.554   Max.   :13.963   Max.   :10.137
##    NA's   :1
##      gene_128          gene_129          gene_130          gene_131
##    Min.   : 8.664   Min.   : 7.878   Min.   : 6.111   Min.   : 0.9854
##    1st Qu.:10.776   1st Qu.: 9.398   1st Qu.:10.548   1st Qu.: 5.4454
##    Median :11.083   Median : 9.904   Median :11.191   Median : 6.2297
##    Mean   :11.052   Mean   : 9.863   Mean   :11.093   Mean   : 6.3207
##    3rd Qu.:11.403   3rd Qu.:10.337   3rd Qu.:11.779   3rd Qu.: 7.1439
##    Max.   :12.579   Max.   :12.027   Max.   :13.913   Max.   :10.7518
##
##      gene_132          gene_133          gene_134          gene_135
##    Min.   : 3.361   Min.   : 0.000   Min.   : 0.000   Min.   :0.0000
##    1st Qu.: 7.635   1st Qu.: 2.444   1st Qu.: 3.370   1st Qu.:0.0000
```

```
## Median : 8.881   Median : 4.369   Median : 6.185   Median :0.6159
## Mean   : 8.811   Mean   : 4.522   Mean   : 6.985   Mean   :0.9689
## 3rd Qu.:10.185   3rd Qu.: 6.602   3rd Qu.:11.005   3rd Qu.:1.4125
## Max.   :14.992   Max.   :10.098   Max.   :16.301   Max.   :8.6436
##
##     gene_136          gene_137          gene_138          gene_139
## Min.   : 7.39   Min.   : 7.758   Min.   : 5.957   Min.   : 1.753
## 1st Qu.:10.45   1st Qu.: 8.982   1st Qu.: 8.423   1st Qu.: 6.425
## Median :10.88   Median : 9.242   Median : 8.922   Median : 7.437
## Mean   :10.87   Mean   : 9.301   Mean   : 8.868   Mean   : 7.303
## 3rd Qu.:11.26   3rd Qu.: 9.607   3rd Qu.: 9.373   3rd Qu.: 8.176
## Max.   :13.12   Max.   :11.046   Max.   :11.124   Max.   :11.968
##                                    NA's   :1
##     gene_140          gene_141          gene_142          gene_143
## Min.   : 7.285   Min.   : 6.301   Min.   : 7.984   Min.   : 4.193
## 1st Qu.: 9.800   1st Qu.: 9.383   1st Qu.:10.146   1st Qu.: 7.794
## Median :10.270   Median :10.061   Median :10.897   Median : 8.618
## Mean   :10.327   Mean   :10.217   Mean   :10.918   Mean   : 8.567
## 3rd Qu.:10.796   3rd Qu.:11.049   3rd Qu.:11.628   3rd Qu.: 9.463
## Max.   :12.814   Max.   :13.948   Max.   :14.365   Max.   :12.331
##                   NA's   :1
##     gene_144          gene_145          gene_146          gene_147
## Min.   : 6.121   Min.   : 5.022   Min.   : 7.281   Min.   : 7.297
## 1st Qu.: 8.837   1st Qu.: 7.919   1st Qu.: 8.611   1st Qu.: 9.563
## Median : 9.168   Median : 8.432   Median : 9.035   Median : 9.823
## Mean   : 9.174   Mean   : 8.913   Mean   : 9.248   Mean   : 9.812
## 3rd Qu.: 9.520   3rd Qu.: 9.182   3rd Qu.: 9.745   3rd Qu.:10.067
## Max.   :11.195   Max.   :14.088   Max.   :13.386   Max.   :11.265
##                   NA's   :1
##     gene_148          gene_149          gene_150          gene_151
## Min.   : 0.000   Min.   : 7.011   Min.   : 6.095   Min.   : 2.708
## 1st Qu.: 1.369   1st Qu.: 9.545   1st Qu.: 9.272   1st Qu.: 8.163
## Median : 4.256   Median : 9.939   Median :10.236   Median : 8.936
## Mean   : 4.791   Mean   : 9.967   Mean   :10.213   Mean   : 8.843
## 3rd Qu.: 8.265   3rd Qu.:10.410   3rd Qu.:11.112   3rd Qu.: 9.688
## Max.   :11.742   Max.   :13.165   Max.   :14.219   Max.   :11.596
##
##     gene_152          gene_153          gene_154          gene_155
## Min.   : 9.834   Min.   : 0.000   Min.   : 1.320   Min.   : 7.648
## 1st Qu.:11.644   1st Qu.: 4.734   1st Qu.: 5.654   1st Qu.: 9.727
## Median :12.176   Median : 6.226   Median : 6.720   Median :10.093
## Mean   :12.188   Mean   : 6.353   Mean   : 6.668   Mean   :10.062
## 3rd Qu.:12.687   3rd Qu.: 7.808   3rd Qu.: 7.808   3rd Qu.:10.418
## Max.   :15.045   Max.   :12.924   Max.   :11.179   Max.   :12.258
## NA's   :1
##     gene_156          gene_157          gene_158          gene_159
## Min.   : 5.925   Min.   : 8.120   Min.   : 0.000   Min.   : 8.87
## 1st Qu.: 8.648   1st Qu.: 9.673   1st Qu.: 8.256   1st Qu.:10.34
## Median : 9.217   Median :10.233   Median : 8.824   Median :10.69
## Mean   : 9.218   Mean   :10.287   Mean   : 8.916   Mean   :10.74
## 3rd Qu.: 9.737   3rd Qu.:10.793   3rd Qu.: 9.591   3rd Qu.:11.19
## Max.   :11.709   Max.   :14.007   Max.   :12.319   Max.   :12.52
## NA's   :1
##     gene_160          gene_161          gene_162          gene_163
```

```
##    Min.   : 2.887   Min.   : 7.830   Min.   : 7.572   Min.   : 0.000
##    1st Qu.: 7.485   1st Qu.: 9.493   1st Qu.: 8.642   1st Qu.: 3.888
##    Median : 8.106   Median : 9.909   Median : 8.933   Median : 4.970
##    Mean   : 8.059   Mean   : 9.875   Mean   : 8.971   Mean   : 5.005
##    3rd Qu.: 8.733   3rd Qu.:10.230   3rd Qu.: 9.289   3rd Qu.: 6.046
##    Max.   :10.781   Max.   :11.994   Max.   :10.776   Max.   :11.197
##                     NA's   :1
##      gene_164         gene_165         gene_166         gene_167
##    Min.   :0.0000   Min.   : 0.6268   Min.   :0.000   Min.   :0.0000
##    1st Qu.:0.8195   1st Qu.: 5.1029   1st Qu.:2.651   1st Qu.:0.6333
##    Median :1.5877   Median : 6.5499   Median :3.699   Median :1.3217
##    Mean   :2.1932   Mean   : 6.2698   Mean   :3.802   Mean   :1.4544
##    3rd Qu.:3.0342   3rd Qu.: 7.7203   3rd Qu.:4.932   3rd Qu.:2.0433
##    Max.   :9.9572   Max.   :10.5303   Max.   :8.840   Max.   :7.9839
##                     NA's   :1
##      gene_168         gene_169         gene_170         gene_171
##    Min.   :0.0000   Min.   :0.00000   Min.   : 4.103   Min.   : 5.731
##    1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.: 7.018   1st Qu.: 8.026
##    Median :0.0000   Median :0.00000   Median : 7.716   Median : 8.434
##    Mean   :0.1574   Mean   :0.06324   Mean   : 7.708   Mean   : 8.427
##    3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.: 8.488   3rd Qu.: 8.834
##    Max.   :4.2980   Max.   :2.56803   Max.   :11.027   Max.   :10.227
##
##      gene_172         gene_173         gene_174         gene_175
##    Min.   : 0.000   Min.   :0.0000   Min.   : 0.000   Min.   : 5.635
##    1st Qu.: 2.074   1st Qu.:0.0000   1st Qu.: 5.644   1st Qu.: 8.007
##    Median : 4.099   Median :0.4443   Median : 6.559   Median : 8.500
##    Mean   : 5.099   Mean   :0.9049   Mean   : 6.369   Mean   : 8.555
##    3rd Qu.: 7.988   3rd Qu.:1.2231   3rd Qu.: 7.308   3rd Qu.: 9.041
##    Max.   :13.252   Max.   :7.7172   Max.   :11.127   Max.   :12.775
##                                      NA's   :1
##      gene_176         gene_177         gene_178         gene_179
##    Min.   : 5.990   Min.   : 0.000   Min.   : 9.333   Min.   : 9.196
##    1st Qu.: 8.679   1st Qu.: 4.052   1st Qu.:11.152   1st Qu.:11.868
##    Median : 9.471   Median : 5.712   Median :11.442   Median :12.280
##    Mean   : 9.440   Mean   : 5.709   Mean   :11.418   Mean   :12.367
##    3rd Qu.:10.133   3rd Qu.: 7.098   3rd Qu.:11.705   3rd Qu.:12.903
##    Max.   :12.476   Max.   :12.929   Max.   :13.063   Max.   :14.355
##
##      gene_180         gene_181         gene_182         gene_183
##    Min.   : 0.0000   Min.   : 2.962   Min.   : 7.864   Min.   : 9.532
##    1st Qu.: 0.0000   1st Qu.: 6.977   1st Qu.: 9.744   1st Qu.:11.093
##    Median : 0.7959   Median : 7.630   Median :10.242   Median :11.534
##    Mean   : 2.8740   Mean   : 7.479   Mean   :10.205   Mean   :11.554
##    3rd Qu.: 4.8952   3rd Qu.: 8.263   3rd Qu.:10.660   3rd Qu.:11.979
##    Max.   :11.9186   Max.   :10.728   Max.   :13.745   Max.   :14.186
##
##      gene_184         gene_185         gene_186         gene_187
##    Min.   : 2.177   Min.   :0.0000   Min.   : 8.049   Min.   : 3.976
##    1st Qu.: 5.167   1st Qu.:0.0000   1st Qu.: 9.101   1st Qu.: 8.330
##    Median : 6.386   Median :0.0000   Median : 9.456   Median : 8.832
##    Mean   : 6.478   Mean   :0.3000   Mean   : 9.476   Mean   : 8.822
##    3rd Qu.: 7.659   3rd Qu.:0.4349   3rd Qu.: 9.790   3rd Qu.: 9.331
##    Max.   :11.464   Max.   :7.4639   Max.   :12.600   Max.   :11.986
```

```
##
##      gene_188         gene_189         gene_190         gene_191
##  Min.   : 3.677   Min.   : 3.085   Min.   :0.000   Min.   : 6.819
##  1st Qu.: 7.839   1st Qu.: 5.854   1st Qu.:0.762   1st Qu.: 8.563
##  Median : 8.261   Median : 6.973   Median :1.444   Median : 9.228
##  Mean   : 8.222   Mean   : 6.932   Mean   :1.723   Mean   : 9.284
##  3rd Qu.: 8.681   3rd Qu.: 8.066   3rd Qu.:2.532   3rd Qu.: 9.983
##  Max.   :11.132   Max.   :10.958   Max.   :8.790   Max.   :12.444
##  NA's   :1                         NA's   :1
##      gene_192         gene_193         gene_194         gene_195
##  Min.   : 6.104   Min.   : 4.061   Min.   : 9.224   Min.   : 0.000
##  1st Qu.: 8.165   1st Qu.: 8.553   1st Qu.:10.376   1st Qu.: 6.463
##  Median : 8.610   Median : 8.914   Median :10.721   Median : 7.559
##  Mean   : 8.663   Mean   : 8.888   Mean   :10.760   Mean   : 7.386
##  3rd Qu.: 9.129   3rd Qu.: 9.283   3rd Qu.:11.154   3rd Qu.: 8.386
##  Max.   :10.764   Max.   :11.165   Max.   :12.863   Max.   :12.538
##                   NA's   :1        NA's   :1
##      gene_196         gene_197         gene_198         gene_199
##  Min.   : 6.250   Min.   :0.000   Min.   : 9.77   Min.   : 8.50
##  1st Qu.: 8.310   1st Qu.:1.127   1st Qu.:10.54   1st Qu.: 9.83
##  Median : 8.805   Median :2.146   Median :10.77   Median :10.22
##  Mean   : 8.871   Mean   :2.885   Mean   :10.80   Mean   :10.21
##  3rd Qu.: 9.391   3rd Qu.:4.659   3rd Qu.:11.02   3rd Qu.:10.57
##  Max.   :11.398   Max.   :8.977   Max.   :12.35   Max.   :11.73
##                                   NA's   :3
##      gene_200         gene_201         gene_202         gene_203
##  Min.   : 5.225   Min.   : 4.797   Min.   : 5.103   Min.   : 0.000
##  1st Qu.: 9.214   1st Qu.: 7.883   1st Qu.: 8.219   1st Qu.: 4.007
##  Median :10.102   Median : 8.434   Median : 8.713   Median : 5.586
##  Mean   :10.053   Mean   : 8.397   Mean   : 8.659   Mean   : 6.899
##  3rd Qu.:10.994   3rd Qu.: 8.952   3rd Qu.: 9.142   3rd Qu.: 7.489
##  Max.   :15.814   Max.   :11.179   Max.   :12.223   Max.   :18.719
##  NA's   :1                                          NA's   :2
##      gene_204         gene_205         gene_206         gene_207
##  Min.   :0.0000   Min.   :2.063   Min.   :0.5474   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:4.216   1st Qu.:3.3930   1st Qu.:0.4876
##  Median :0.0000   Median :4.951   Median :3.9865   Median :1.0359
##  Mean   :0.4259   Mean   :4.991   Mean   :4.0190   Mean   :1.3002
##  3rd Qu.:0.5734   3rd Qu.:5.732   3rd Qu.:4.6110   3rd Qu.:1.8741
##  Max.   :4.5596   Max.   :8.479   Max.   :8.6800   Max.   :6.8003
##
##      gene_208         gene_209         gene_210         gene_211
##  Min.   :0.000   Min.   : 0.000   Min.   :0.0000   Min.   : 3.745
##  1st Qu.:1.275   1st Qu.: 2.029   1st Qu.:0.5336   1st Qu.: 7.894
##  Median :1.880   Median : 2.600   Median :1.0842   Median : 8.927
##  Mean   :2.041   Mean   : 2.721   Mean   :1.1388   Mean   : 8.841
##  3rd Qu.:2.747   3rd Qu.: 3.248   3rd Qu.:1.6459   3rd Qu.: 9.869
##  Max.   :5.944   Max.   :10.866   Max.   :4.6814   Max.   :12.740
##
##      gene_212         gene_213         gene_214         gene_215
##  Min.   : 4.728   Min.   : 6.031   Min.   : 8.652   Min.   : 4.608
##  1st Qu.: 8.120   1st Qu.:10.306   1st Qu.:10.668   1st Qu.: 9.059
##  Median : 8.562   Median :11.267   Median :11.430   Median : 9.926
##  Mean   : 8.514   Mean   :11.219   Mean   :11.449   Mean   : 9.850
```

```
## 3rd Qu.: 8.968   3rd Qu.:12.232   3rd Qu.:12.089   3rd Qu.:10.667
## Max.   :10.910   Max.   :15.145   Max.   :15.358   Max.   :13.658
## NA's   :3
##    gene_216         gene_217         gene_218         gene_219
## Min.   : 4.212   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
## 1st Qu.: 8.575   1st Qu.: 2.459   1st Qu.: 2.723   1st Qu.: 0.000
## Median : 9.597   Median : 3.311   Median : 4.231   Median : 0.000
## Mean   : 9.836   Mean   : 3.680   Mean   : 4.848   Mean   : 2.274
## 3rd Qu.:11.155   3rd Qu.: 4.426   3rd Qu.: 6.169   3rd Qu.: 1.001
## Max.   :14.523   Max.   :10.693   Max.   :14.200   Max.   :14.090
##                                                    NA's   :1
##    gene_220         gene_221         gene_222          gene_223
## Min.   : 0.0000   Min.   : 0.5117   Min.   :0.00000   Min.   : 0.000
## 1st Qu.: 0.0000   1st Qu.: 5.4151   1st Qu.:0.00000   1st Qu.: 2.462
## Median : 0.0000   Median : 7.7866   Median :0.00000   Median : 3.625
## Mean   : 2.0902   Mean   : 7.3825   Mean   :0.08157   Mean   : 4.307
## 3rd Qu.: 0.7028   3rd Qu.: 9.2343   3rd Qu.:0.00000   3rd Qu.: 5.383
## Max.   :13.9450   Max.   :12.9747   Max.   :2.29007   Max.   :11.830
##
##    gene_224         gene_225         gene_226         gene_227
## Min.   : 5.306   Min.   : 6.080   Min.   : 0.000   Min.   : 0.000
## 1st Qu.: 9.628   1st Qu.: 9.489   1st Qu.: 5.703   1st Qu.: 1.573
## Median :10.345   Median : 9.846   Median : 7.105   Median : 2.736
## Mean   :10.264   Mean   :10.017   Mean   : 6.977   Mean   : 3.280
## 3rd Qu.:11.071   3rd Qu.:10.423   3rd Qu.: 8.546   3rd Qu.: 4.352
## Max.   :13.737   Max.   :12.751   Max.   :13.937   Max.   :15.710
##
##    gene_228         gene_229         gene_230         gene_231
## Min.   : 6.362   Min.   :0.0000   Min.   :14.84   Min.   : 0.000
## 1st Qu.:11.850   1st Qu.:0.0000   1st Qu.:16.06   1st Qu.: 1.277
## Median :12.695   Median :0.9382   Median :16.43   Median : 2.687
## Mean   :12.724   Mean   :1.3212   Mean   :16.43   Mean   : 3.516
## 3rd Qu.:13.546   3rd Qu.:2.0268   3rd Qu.:16.81   3rd Qu.: 5.457
## Max.   :17.795   Max.   :7.3995   Max.   :17.98   Max.   :13.188
##                                                   NA's   :1
##    gene_232        gene_233         gene_234         gene_235
## Min.   :13.63   Min.   : 1.344   Min.   : 8.279   Min.   :0.0000
## 1st Qu.:15.60   1st Qu.: 7.273   1st Qu.: 9.546   1st Qu.:0.0000
## Median :15.98   Median : 8.725   Median :10.002   Median :0.0000
## Mean   :15.97   Mean   : 9.232   Mean   :10.030   Mean   :0.2499
## 3rd Qu.:16.34   3rd Qu.:10.986   3rd Qu.:10.475   3rd Qu.:0.0000
## Max.   :17.67   Max.   :17.985   Max.   :12.301   Max.   :9.6035
##                 NA's   :1
##    gene_236          gene_237         gene_238         gene_239
## Min.   :0.00000   Min.   :0.0000   Min.   : 0.0000   Min.   :0.00000
## 1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.:0.00000
## Median :0.00000   Median :0.0000   Median : 0.8596   Median :0.00000
## Mean   :0.06137   Mean   :0.1720   Mean   : 1.7449   Mean   :0.03468
## 3rd Qu.:0.00000   3rd Qu.:0.3576   3rd Qu.: 2.8069   3rd Qu.:0.00000
## Max.   :1.16060   Max.   :1.9874   Max.   :12.7301   Max.   :2.41281
##
##    gene_240         gene_241         gene_242         gene_243
## Min.   : 9.265   Min.   : 0.000   Min.   : 0.0000   Min.   :11.24
## 1st Qu.:11.792   1st Qu.: 1.235   1st Qu.: 0.5152   1st Qu.:12.96
```

```
##   Median :12.293    Median : 2.327    Median : 0.9532    Median :13.28
##   Mean   :12.253    Mean   : 2.910    Mean   : 1.0136    Mean   :13.35
##   3rd Qu.:12.748    3rd Qu.: 3.935    3rd Qu.: 1.4171    3rd Qu.:13.68
##   Max.   :14.554    Max.   :13.741    Max.   :11.6130    Max.   :16.14
##
##      gene_244          gene_245          gene_246          gene_247
##   Min.   : 8.451    Min.   : 9.315    Min.   : 8.345    Min.   :10.48
##   1st Qu.: 9.673    1st Qu.:11.072    1st Qu.:10.191    1st Qu.:12.24
##   Median : 9.933    Median :11.270    Median :10.539    Median :12.60
##   Mean   : 9.933    Mean   :11.298    Mean   :10.463    Mean   :12.56
##   3rd Qu.:10.174    3rd Qu.:11.535    3rd Qu.:10.782    3rd Qu.:12.91
##   Max.   :12.070    Max.   :12.636    Max.   :12.392    Max.   :14.20
##
##      gene_248          gene_249          gene_250          gene_251
##   Min.   : 5.528    Min.   :1.991     Min.   : 9.906    Min.   :6.314
##   1st Qu.: 6.950    1st Qu.:4.919     1st Qu.:11.645    1st Qu.:7.526
##   Median : 7.504    Median :5.514     Median :11.987    Median :7.803
##   Mean   : 7.503    Mean   :5.510     Mean   :11.997    Mean   :7.852
##   3rd Qu.: 7.941    3rd Qu.:6.112     3rd Qu.:12.348    3rd Qu.:8.134
##   Max.   :11.115    Max.   :9.967     Max.   :13.854    Max.   :9.539
##
##      gene_252          gene_253          gene_254            Class
##   Min.   : 7.403    Min.   : 7.432    Min.   :0.00000    Length:801
##   1st Qu.: 8.680    1st Qu.: 8.398    1st Qu.:0.00000    Class :character
##   Median : 8.956    Median : 8.698    Median :0.00000    Mode  :character
##   Mean   : 8.992    Mean   : 8.680    Mean   :0.08596
##   3rd Qu.: 9.260    3rd Qu.: 8.966    3rd Qu.:0.00000
##   Max.   :10.976    Max.   :10.002    Max.   :3.41099
##
```

There are a total of 801 instances(patients)(rows/observations) and 20531 features (gene expression)(columns/variables) or 257 different genes. Here our target variable will be the cancer type (Class). Also, here we do not use X which is an additional unused feature. Thus, we drop it.

```r
if (!require("dplyr", character.only = TRUE)) {
  install.packages("dplyr", dependencies = TRUE)
}
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

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

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

```r
library(dplyr)

# Dropping column X (samplename)
cancer_data <- select(cancer_data, -X)
```

Data Exploration

Exploratory data plots

Plotting a histogram to understand the number of cases for each type of cancer

```
## Loading required package: ggplot2
```

```r
# Plot a histogram
ggplot(cancer_data, aes(x = Class)) +
  geom_bar(fill = "steelblue") +
  theme_minimal() +
  labs(title = "Number of cases for different types of Cancer", x = "Class", y = "Count") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
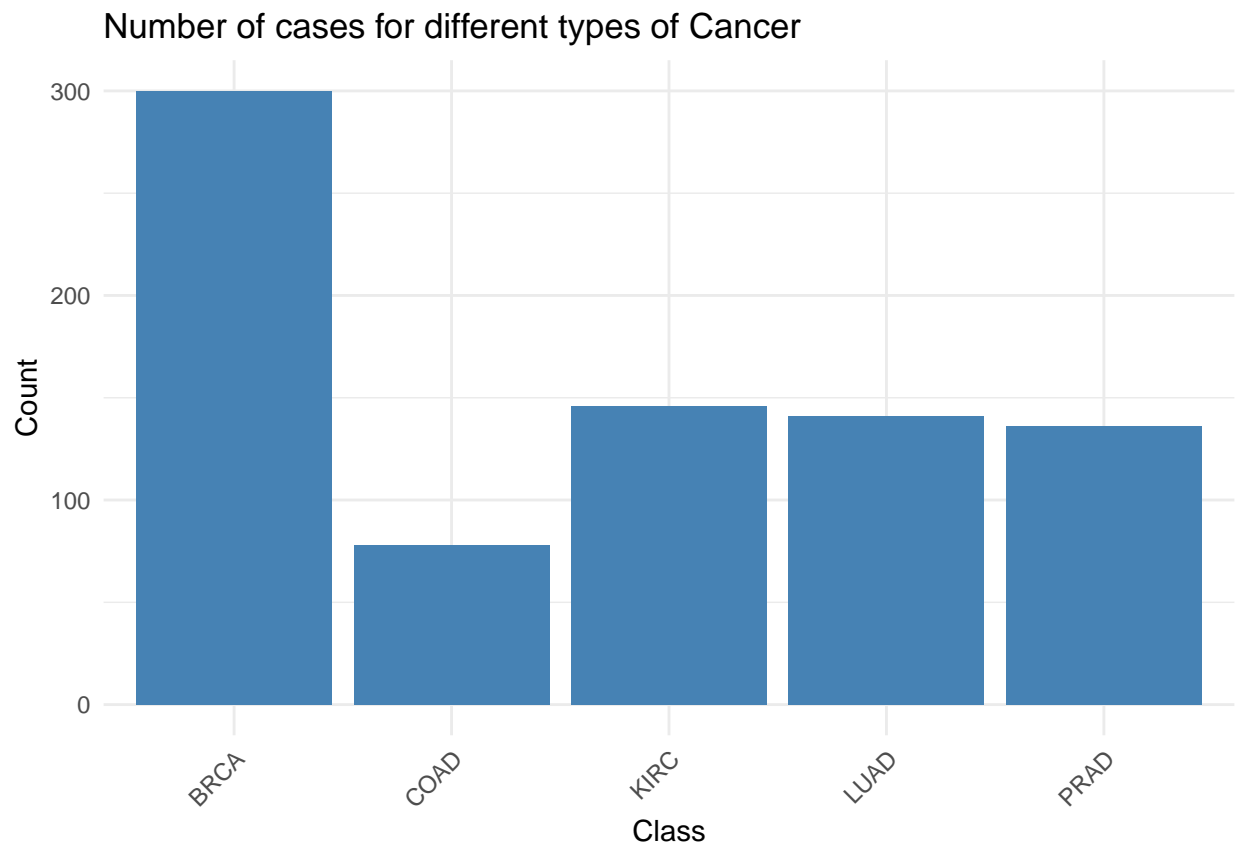

The graph indicates that there is significant class imbalance in the dataset. BRCA has the majority of samples. COAD has the least which may impact the training of the machine learning models as they might become biased towards overrepresented classes. KIRC, LUAD and PRAD cancers have similar amount of cases. For machine learning, such imbalances makes it important for us to use strategies like resampling or weighted class handling to ensure that the model generalizes well across all the cancer types.

Identification of missing values

I check if there are any missing values in the data, if so I impute the mean of each gene column in it's place which does not affect the data.

```
# Checking for missing values
sum(is.na(cancer_data))
```

```
## [1] 45
```

Data imputation of missing data

```
# Impute missing values with mean of each column
cancer_data[] <- lapply(cancer_data, function(x) {
  if(is.numeric(x)) {
    x[is.na(x)] <- mean(x, na.rm = TRUE)
    }
  return(x)
})

# Check again for missing values
sum(is.na(cancer_data))
```

```
## [1] 0
```

Detection of Outliers

It is important to identify anomalies that can skew data analysis and to study the underlying patterns in the data points. I perform identification of outliers using z scores. The threshold used is 2.5 which will help to detect significant deviations while minimizing false positives.

```
# Function for finding outliers
outliers <- function(data, column) {

  # Calculate mean and standard deviation
  mu <- mean(data[[column]], na.rm = TRUE)
  sd <- sd(data[[column]], na.rm = TRUE)

  # Calculate z-score
  z_scores <- (data[[column]] - mu) / sd

  # Identify z-scores above the threshold
  outlier_threshold <- which(abs(z_scores) > 2.5)

  return(outlier_threshold)
}

# Call the function for each gene
cols <- names(cancer_data)[-c(1, ncol(cancer_data))]
outliers_list <- list()
```

```
for(column in cols) {
  outliers_list[[column]] <- outliers(cancer_data, column)
}
#outliers_list

# Total number of outliers
total_outliers <- sum(lengths(outliers_list))
print(paste("The total number of outliers are:", total_outliers))
```

[1] "The total number of outliers are: 3611"

There are 3611 outliers in the data. Thus, the data needs to be normalized. Before, normalization we check the correlation between our features and the particular class of cancer.

```
# Visualize the outlier counts
outlier_counts <- sapply(outliers_list, length)
barplot(outlier_counts, main = "Outlier Counts per Gene", xlab = "Genes", ylab = "Number of Outliers", l
```

Outlier Counts per Gene
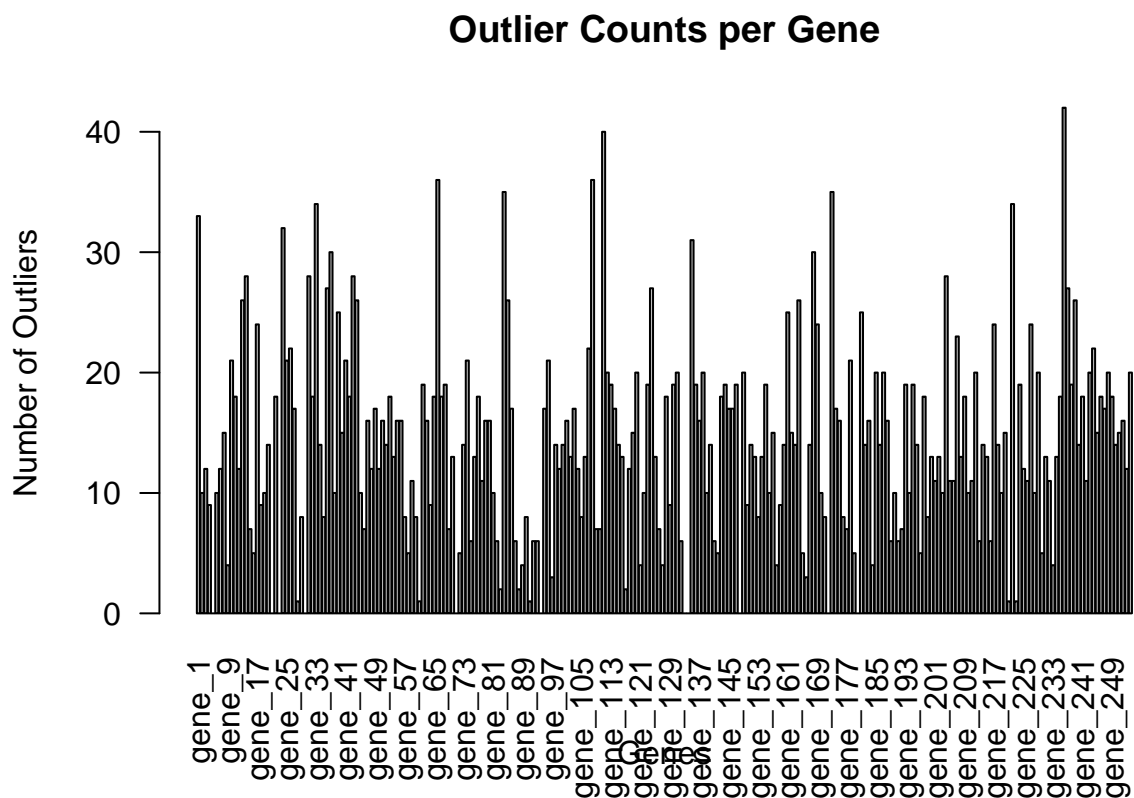

```
# Let's visualize genes with outlier counts above a certain threshold, e.g., greater than 10
high_outlier_genes <- names(which(outlier_counts > 20))
if (length(high_outlier_genes) > 0) {
  boxplot(cancer_data[, high_outlier_genes], las = 2, main = "Boxplots for Genes with High Outlier Coun
} else {
  print("No genes have more than 10 outliers.")
}
```

Boxplots for Genes with High Outlier Counts
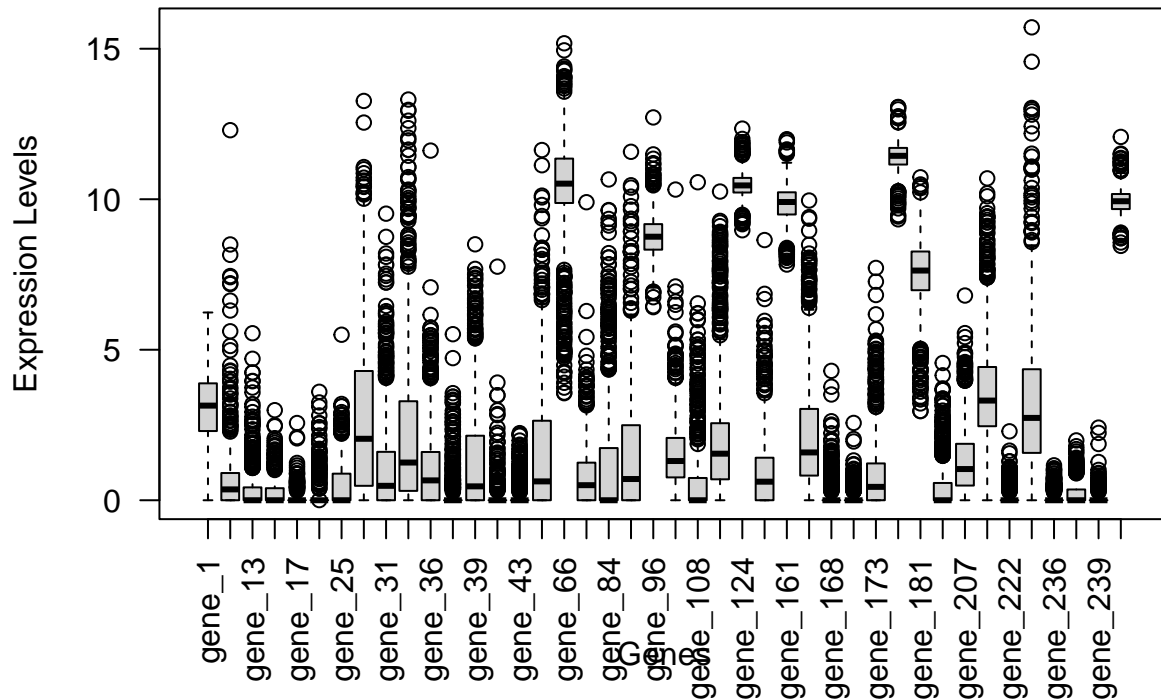
you can either keep them, especially for rare disease subtypes, or mitigate their influence using robust statistical methods or transformations (like log transformation)

Correlation Analysis

It is important to find the correlation between the genes and each unique class to select all the relevant features. I find the least correlated genes for each class to explore the data better. Here I use one hot encoding method to represent categorical variables as numerical values and use dummy variables.

```
## Loading required package: reshape2
```

Encode the categorical variable to factor using one hot encoding method

```r
# One hot encoding
cancer_data$Class <- as.factor(cancer_data$Class)
```

Finding the least correlated genes to each class

```r
# Create dummy variables
dummy_class <- model.matrix(~ Class - 1, data = cancer_data)

# Convert all data to numeric
gene_data <- select(cancer_data, -Class)
gene_data <- as.data.frame(lapply(gene_data, as.numeric))
```

```r
# Calculate correlations
cor_analysis <- (cor(cbind(gene_data, dummy_class), use = "complete.obs"))
cor_genes <- ncol(gene_data)
cor_class <- ncol(dummy_class)
final_cor <- cor_analysis[1:cor_genes, (cor_genes+1):(cor_genes+cor_class)]

# Melt and format the matrix
melt_cor <- melt(final_cor, varnames = c("Gene", "Class"))
melt_cor$Class <- gsub("Class", "", melt_cor$Class)

# Find the least 10 correlated genes for each class
order_data <- melt_cor[order(melt_cor$Class, melt_cor$value), ]
group_data <- group_by(order_data, Class)
least_cor_genes <- summarise(group_data, least_genes = Gene[1:5], least_val = value[1:5], .groups = 'dro

print(least_cor_genes)
```

```
## # A tibble: 25 x 3
##    Class least_genes least_val
##    <chr> <fct>           <dbl>
##  1 BRCA  gene_221       -0.776
##  2 BRCA  gene_89        -0.666
##  3 BRCA  gene_134       -0.634
##  4 BRCA  gene_216       -0.601
##  5 BRCA  gene_172       -0.588
##  6 COAD  gene_66        -0.656
##  7 COAD  gene_123       -0.555
##  8 COAD  gene_226       -0.533
##  9 COAD  gene_28        -0.523
## 10 COAD  gene_95        -0.489
## # i 15 more rows
```

```r
# Find common least correlated genes between the classes
genes_per_class <- by(least_cor_genes, least_cor_genes$Class, function(df) {
  as.character(df$least_genes)
})
common_genes <- Reduce(intersect, genes_per_class)
print(common_genes)
```

```
## character(0)
```

There are no common genes that show least correlation between the classes and that we can eliminate further during feature selection.

Now, I plot a correlation graph using UMAPs used for visualizing high-dimensional data in a lower dimensional space to see the correlation between the genes for each class.

```
## Loading required package: umap
```

```r
set.seed(3434)

# UMAP dimensionality reduction
```

```
umap_plot <- umap(gene_data)

# Convert for ggplot
umap_df <- as.data.frame(umap_plot$layout)
umap_df$Class <- cancer_data$Class

# Plot a UMAP
ggplot(umap_df, aes(x = V1, y = V2, color = Class)) +
  geom_point(alpha = 0.7) +
  labs(title = "UMAP for Correlation between Genes of each Class", x = "Genes", y = "Genes") +
  theme_minimal() +
  theme(legend.position = "right")
```
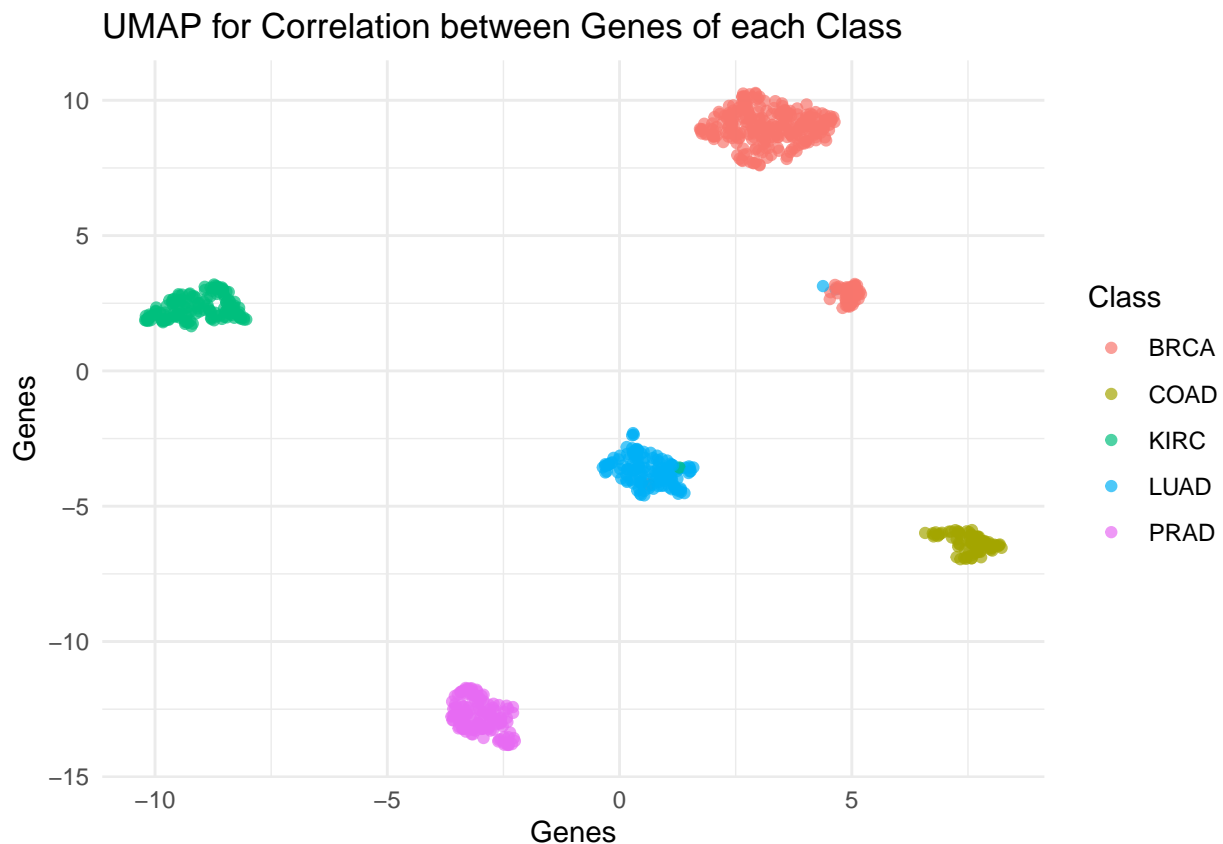


```
# Check correlation between genes
gene_cor_matrix <- cor(gene_data, use = "complete.obs")
gene_cor_matrix <- melt(gene_cor_matrix)


ggplot(gene_cor_matrix, aes(x = value)) +
  geom_histogram(binwidth = 0.1, fill = "steelblue", color = "black") +
  labs(title = "Distribution of Correlation Values", x = "Correlation Value", y = "Frequency") +
  theme_minimal()
```

```
## Warning: Removed 1014 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

Distribution of Correlation Values
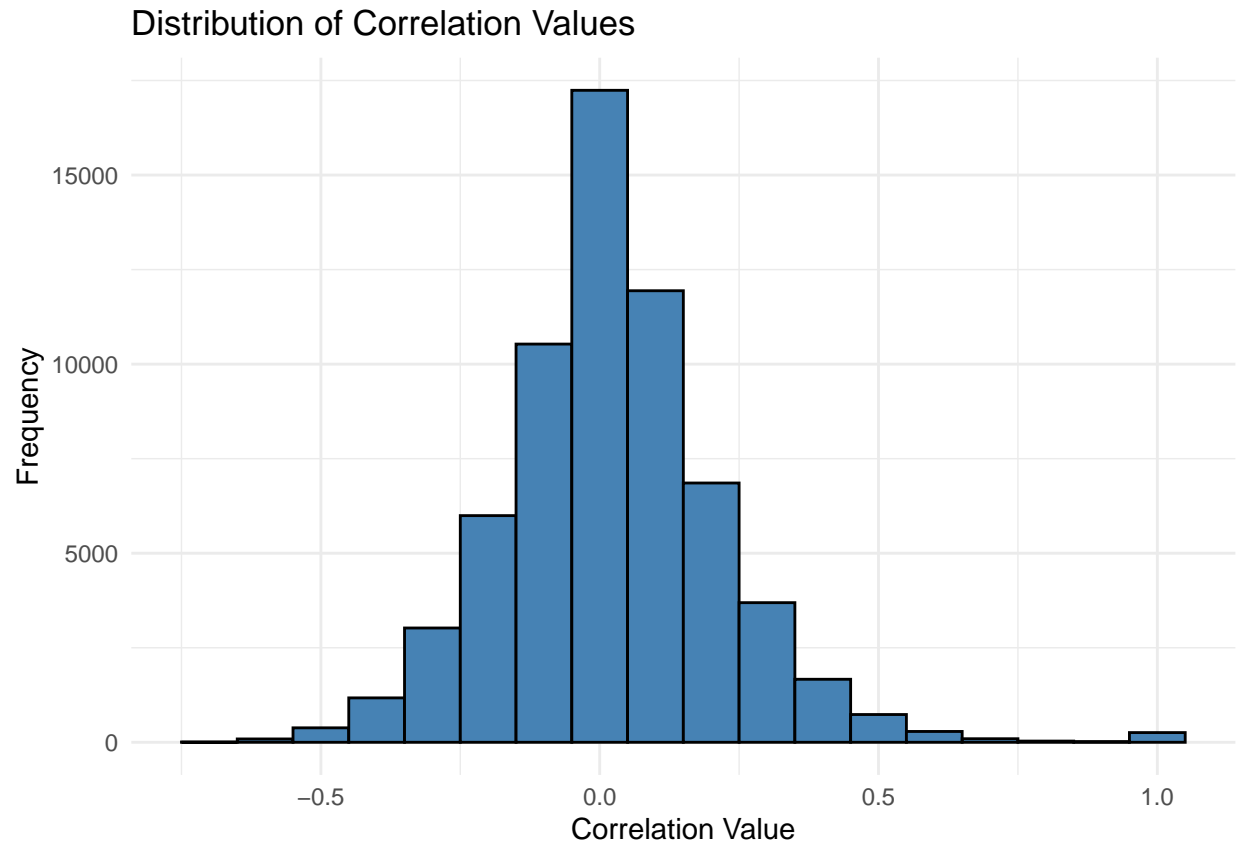
The histogram shows a normal-like distribution of correlation values. It indicates that most gene expression relationships are near zero, suggesting little to no linear correlation between them. Linear correlation may not capture all types of relationships. Look for non-linear patterns or interactions between genes using methods like decision trees, random forests, or neural networks that can model complex relationships.

The UMAP plot shows distinct clusters corresponding to different classes of cancer which indicates patterns of gene correlation unique to each class. It can be seen that some genes of BRCA form clusters which might possibly be outliers. Hence, we need to normalize the data. Also, the separation between the clusters suggests that the gene expression profiles for these classes are quite different.

Data Preparation

Data Cleaning & Shaping

There are a number of outliers detected. Thus, the data needs to be normalized for better results. Before, normalization I log2 transform the data to reduce the skewness introduced by outliers.

Transformation of features to adjust distribution

```
# Perform log transform
log_tranform_data <- select(cancer_data, -Class)
log_tranform_data <- log2(log_tranform_data + 1)
log_tranform_data$Class <- cancer_data$Class
```

```r
summary(log_tranform_data)
```

```
##      gene_0              gene_1           gene_2           gene_3
##  Min.   :0.00000   Min.   :0.000   Min.   :0.000   Min.   :2.587
##  1st Qu.:0.00000   1st Qu.:1.722   1st Qu.:1.761   1st Qu.:2.869
##  Median :0.00000   Median :2.051   Median :2.041   Median :2.937
##  Mean   :0.02922   Mean   :1.916   Mean   :1.977   Mean   :2.944
##  3rd Qu.:0.00000   3rd Qu.:2.288   3rd Qu.:2.264   3rd Qu.:3.007
##  Max.   :1.31170   Max.   :2.855   Max.   :2.820   Max.   :3.476
##      gene_4          gene_5      gene_6           gene_7            gene_8
##  Min.   :3.238   Min.   :0   Min.   :2.302   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:3.387   1st Qu.:0   1st Qu.:2.940   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :3.432   Median :0   Median :3.079   Median :0.5291   Median :0.00000
##  Mean   :3.433   Mean   :0   Mean   :3.058   Mean   :0.5077   Mean   :0.01673
##  3rd Qu.:3.478   3rd Qu.:0   3rd Qu.:3.189   3rd Qu.:0.8394   3rd Qu.:0.00000
##  Max.   :3.627   Max.   :0   Max.   :3.551   Max.   :1.9180   Max.   :1.47798
##      gene_9             gene_10          gene_11          gene_12
##  Min.   :0.000000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.000000   1st Qu.:0.0000   1st Qu.:0.4811   1st Qu.:1.698
##  Median :0.000000   Median :0.4438   Median :0.8489   Median :1.888
##  Mean   :0.009065   Mean   :0.5580   Mean   :0.8419   Mean   :1.845
##  3rd Qu.:0.000000   3rd Qu.:0.9289   3rd Qu.:1.2473   3rd Qu.:2.054
##  Max.   :2.341304   Max.   :3.7326   Max.   :3.1513   Max.   :2.532
##     gene_13          gene_14          gene_15            gene_16
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.0000   Median :0.0000   Median :0.000000   Median :0.000000
##  Mean   :0.2865   Mean   :0.2241   Mean   :0.006396   Mean   :0.003521
##  3rd Qu.:0.5136   3rd Qu.:0.4852   3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :2.7110   Max.   :1.9970   Max.   :1.000000   Max.   :0.661641
##     gene_17          gene_18          gene_19         gene_20
##  Min.   :0.00000   Min.   :0.000   Min.   :2.686   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:3.215   1st Qu.:1.844
##  Median :0.00000   Median :0.713   Median :3.345   Median :2.107
##  Mean   :0.03616   Mean   :1.106   Mean   :3.323   Mean   :2.037
##  3rd Qu.:0.00000   3rd Qu.:1.889   3rd Qu.:3.450   3rd Qu.:2.334
##  Max.   :1.83407   Max.   :3.568   Max.   :3.751   Max.   :3.029
##     gene_21         gene_22          gene_23     gene_24          gene_25
##  Min.   :0.000   Min.   :2.376   Min.   :0   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:3.106   1st Qu.:0   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.780   Median :3.212   Median :0   Median :0.0000   Median :0.0000
##  Mean   :1.378   Mean   :3.199   Mean   :0   Mean   :0.0922   Mean   :0.4866
##  3rd Qu.:2.564   3rd Qu.:3.324   3rd Qu.:0   3rd Qu.:0.0000   3rd Qu.:0.9131
##  Max.   :3.134   Max.   :3.821   Max.   :0   Max.   :2.2024   Max.   :2.7000
##     gene_26          gene_27          gene_28         gene_29
##  Min.   :0.0000   Min.   :2.848   Min.   :0.000   Min.   :1.680
##  1st Qu.:0.5644   1st Qu.:3.292   1st Qu.:2.019   1st Qu.:2.584
##  Median :1.6050   Median :3.357   Median :2.585   Median :2.830
##  Mean   :1.5133   Mean   :3.348   Mean   :2.380   Mean   :2.809
##  3rd Qu.:2.4043   3rd Qu.:3.416   3rd Qu.:2.866   3rd Qu.:3.037
##  Max.   :3.8344   Max.   :3.648   Max.   :3.492   Max.   :3.818
##     gene_30          gene_31          gene_32          gene_33
##  Min.   :0.0000   Min.   :0.0000   Min.   :2.389   Min.   :0.0000
```

25

```
##    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:2.859    1st Qu.:0.3903
##    Median :0.0000    Median :0.5645    Median :2.955    Median :1.1717
##    Mean   :0.9913    Mean   :0.8035    Mean   :2.953    Mean   :1.3000
##    3rd Qu.:2.6519    3rd Qu.:1.3828    3rd Qu.:3.054    3rd Qu.:2.1005
##    Max.   :3.4545    Max.   :3.3951    Max.   :3.543    Max.   :3.8391
##      gene_34          gene_35          gene_36          gene_37
##    Min.   :3.375    Min.   :2.349    Min.   :0.0000    Min.   :0.0000
##    1st Qu.:3.796    1st Qu.:3.118    1st Qu.:0.0000    1st Qu.:0.0000
##    Median :3.877    Median :3.259    Median :0.7311    Median :0.0000
##    Mean   :3.873    Mean   :3.239    Mean   :0.8553    Mean   :0.1712
##    3rd Qu.:3.967    3rd Qu.:3.406    3rd Qu.:1.3801    3rd Qu.:0.0000
##    Max.   :4.199    Max.   :3.682    Max.   :3.6571    Max.   :2.7041
##      gene_38          gene_39          gene_40          gene_41
##    Min.   :3.034    Min.   :0.0000    Min.   :3.104    Min.   :0.00000
##    1st Qu.:3.354    1st Qu.:0.0000    1st Qu.:3.361    1st Qu.:0.00000
##    Median :3.394    Median :0.5469    Median :3.428    Median :0.00000
##    Mean   :3.391    Mean   :0.8751    Mean   :3.423    Mean   :0.09246
##    3rd Qu.:3.429    3rd Qu.:1.6523    3rd Qu.:3.486    3rd Qu.:0.00000
##    Max.   :3.568    Max.   :3.2483    Max.   :3.922    Max.   :3.13097
##      gene_42          gene_43          gene_44          gene_45
##    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :1.393
##    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:2.673
##    Median :0.0000    Median :0.0000    Median :0.7012    Median :2.886
##    Mean   :0.0782    Mean   :0.1547    Mean   :0.9957    Mean   :2.857
##    3rd Qu.:0.0000    3rd Qu.:0.0000    3rd Qu.:1.8644    3rd Qu.:3.101
##    Max.   :2.4852    Max.   :1.6914    Max.   :3.6593    Max.   :3.618
##      gene_46          gene_47          gene_48          gene_49
##    Min.   :3.042    Min.   :2.998    Min.   :3.178    Min.   :0.0000
##    1st Qu.:3.420    1st Qu.:3.388    1st Qu.:3.583    1st Qu.:0.0000
##    Median :3.479    Median :3.446    Median :3.622    Median :0.5023
##    Mean   :3.475    Mean   :3.441    Mean   :3.618    Mean   :0.5004
##    3rd Qu.:3.538    3rd Qu.:3.502    3rd Qu.:3.657    3rd Qu.:0.8821
##    Max.   :3.712    Max.   :3.787    Max.   :3.780    Max.   :1.8538
##      gene_50          gene_51          gene_52          gene_53
##    Min.   :0.000    Min.   :2.896    Min.   :3.416    Min.   :3.111
##    1st Qu.:3.261    1st Qu.:3.281    1st Qu.:3.610    1st Qu.:3.379
##    Median :3.308    Median :3.333    Median :3.652    Median :3.421
##    Mean   :3.297    Mean   :3.334    Mean   :3.652    Mean   :3.420
##    3rd Qu.:3.349    3rd Qu.:3.383    3rd Qu.:3.694    3rd Qu.:3.464
##    Max.   :3.514    Max.   :3.622    Max.   :3.858    Max.   :3.714
##      gene_54          gene_55          gene_56          gene_57
##    Min.   :2.764    Min.   :2.176    Min.   :3.208    Min.   :1.688
##    1st Qu.:3.114    1st Qu.:3.021    1st Qu.:3.455    1st Qu.:2.665
##    Median :3.169    Median :3.194    Median :3.490    Median :2.852
##    Mean   :3.165    Mean   :3.150    Mean   :3.494    Mean   :2.843
##    3rd Qu.:3.223    3rd Qu.:3.317    3rd Qu.:3.530    3rd Qu.:3.021
##    Max.   :3.562    Max.   :3.578    Max.   :3.800    Max.   :3.630
##      gene_58          gene_59          gene_60          gene_61
##    Min.   :2.109    Min.   :0.000    Min.   :1.555    Min.   :0.000
##    1st Qu.:3.167    1st Qu.:1.840    1st Qu.:2.816    1st Qu.:1.741
##    Median :3.346    Median :2.170    Median :2.919    Median :2.779
##    Mean   :3.336    Mean   :2.106    Mean   :2.901    Mean   :2.404
##    3rd Qu.:3.551    3rd Qu.:2.469    3rd Qu.:3.001    3rd Qu.:3.174
##    Max.   :3.868    Max.   :3.204    Max.   :3.253    Max.   :3.864
```

```
##    gene_62         gene_63         gene_64         gene_65
## Min.   :0.0000   Min.   :0.000   Min.   :2.713   Min.   :2.747
## 1st Qu.:0.7524   1st Qu.:1.608   1st Qu.:3.306   1st Qu.:3.456
## Median :1.4932   Median :1.995   Median :3.422   Median :3.519
## Mean   :1.5449   Mean   :1.943   Mean   :3.408   Mean   :3.505
## 3rd Qu.:2.3875   3rd Qu.:2.352   3rd Qu.:3.528   3rd Qu.:3.577
## Max.   :3.6608   Max.   :3.400   Max.   :3.841   Max.   :3.757
##    gene_66         gene_67         gene_68         gene_69
## Min.   :2.188   Min.   :0.000   Min.   :2.057   Min.   :0.000
## 1st Qu.:3.443   1st Qu.:1.631   1st Qu.:3.104   1st Qu.:2.270
## Median :3.526   Median :2.142   Median :3.227   Median :2.663
## Mean   :3.484   Mean   :2.135   Mean   :3.199   Mean   :2.564
## 3rd Qu.:3.626   3rd Qu.:2.657   3rd Qu.:3.330   3rd Qu.:2.933
## Max.   :4.016   Max.   :3.785   Max.   :3.607   Max.   :3.619
##    gene_70         gene_71         gene_72         gene_73
## Min.   :2.272   Min.   :0.000   Min.   :0.000   Min.   :2.979
## 1st Qu.:3.094   1st Qu.:2.368   1st Qu.:2.226   1st Qu.:3.278
## Median :3.207   Median :2.865   Median :2.541   Median :3.323
## Mean   :3.193   Mean   :2.630   Mean   :2.458   Mean   :3.322
## 3rd Qu.:3.309   3rd Qu.:3.114   3rd Qu.:2.814   3rd Qu.:3.369
## Max.   :3.648   Max.   :3.569   Max.   :3.434   Max.   :3.570
##    gene_74          gene_75          gene_76          gene_77
## Min.   :0.0000   Min.   :0.9841   Min.   :0.000   Min.   :0.0000
## 1st Qu.:0.0000   1st Qu.:2.6914   1st Qu.:2.249   1st Qu.:0.5575
## Median :0.5851   Median :2.9870   Median :2.501   Median :1.2285
## Mean   :0.6657   Mean   :2.9320   Mean   :2.474   Mean   :1.2085
## 3rd Qu.:1.1693   3rd Qu.:3.2415   3rd Qu.:2.712   3rd Qu.:1.7914
## Max.   :3.4464   Max.   :3.8935   Max.   :3.623   Max.   :3.7259
##    gene_78         gene_79         gene_80         gene_81
## Min.   :2.647   Min.   :3.001   Min.   :2.993   Min.   :1.970
## 1st Qu.:3.274   1st Qu.:3.247   1st Qu.:3.323   1st Qu.:2.875
## Median :3.374   Median :3.292   Median :3.382   Median :2.997
## Mean   :3.363   Mean   :3.288   Mean   :3.377   Mean   :2.999
## 3rd Qu.:3.461   3rd Qu.:3.333   3rd Qu.:3.437   3rd Qu.:3.120
## Max.   :3.755   Max.   :3.591   Max.   :3.686   Max.   :3.548
##    gene_82         gene_83         gene_84          gene_85
## Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:3.225   1st Qu.:1.289   1st Qu.:0.0000   1st Qu.:0.0000
## Median :3.289   Median :2.005   Median :0.0000   Median :0.7697
## Mean   :3.286   Mean   :2.130   Mean   :0.7813   Mean   :0.9977
## 3rd Qu.:3.362   3rd Qu.:3.137   3rd Qu.:1.4506   3rd Qu.:1.8042
## Max.   :3.581   Max.   :3.945   Max.   :3.5431   Max.   :3.6526
##    gene_86         gene_87         gene_88         gene_89
## Min.   :3.146   Min.   :0.000   Min.   :1.918   Min.   :1.166
## 1st Qu.:3.485   1st Qu.:1.844   1st Qu.:3.164   1st Qu.:3.073
## Median :3.536   Median :2.191   Median :3.402   Median :3.355
## Mean   :3.536   Mean   :2.327   Mean   :3.388   Mean   :3.324
## 3rd Qu.:3.593   3rd Qu.:2.888   3rd Qu.:3.667   3rd Qu.:3.557
## Max.   :3.906   Max.   :3.814   Max.   :3.927   Max.   :4.093
##    gene_90         gene_91         gene_92         gene_93
## Min.   :3.076   Min.   :0.000   Min.   :0.000   Min.   :0.000
## 1st Qu.:3.344   1st Qu.:1.195   1st Qu.:2.147   1st Qu.:2.608
## Median :3.419   Median :2.079   Median :2.556   Median :3.020
## Mean   :3.432   Mean   :1.879   Mean   :2.440   Mean   :2.890
```

```
##   3rd Qu.:3.523   3rd Qu.:2.693   3rd Qu.:2.866   3rd Qu.:3.253
##   Max.   :3.839   Max.   :3.489   Max.   :3.292   Max.   :3.783
##      gene_94         gene_95         gene_96         gene_97
##   Min.   :0.000   Min.   :1.420   Min.   :2.890   Min.   :0.000
##   1st Qu.:1.001   1st Qu.:2.939   1st Qu.:3.221   1st Qu.:1.885
##   Median :2.135   Median :3.132   Median :3.286   Median :2.244
##   Mean   :1.934   Mean   :3.078   Mean   :3.284   Mean   :2.161
##   3rd Qu.:2.939   3rd Qu.:3.276   3rd Qu.:3.347   3rd Qu.:2.541
##   Max.   :3.586   Max.   :3.714   Max.   :3.778   Max.   :3.212
##      gene_98         gene_99        gene_100        gene_101
##   Min.   :3.279   Min.   :2.948   Min.   :3.290   Min.   :3.285
##   1st Qu.:3.465   1st Qu.:3.277   1st Qu.:3.481   1st Qu.:3.516
##   Median :3.518   Median :3.341   Median :3.522   Median :3.549
##   Mean   :3.528   Mean   :3.331   Mean   :3.526   Mean   :3.549
##   3rd Qu.:3.590   3rd Qu.:3.386   3rd Qu.:3.567   3rd Qu.:3.585
##   Max.   :3.840   Max.   :3.601   Max.   :3.838   Max.   :3.741
##      gene_102        gene_103        gene_104        gene_105
##   Min.   :3.301   Min.   :3.111   Min.   :2.795   Min.   :0.9303
##   1st Qu.:3.464   1st Qu.:3.448   1st Qu.:3.345   1st Qu.:2.6546
##   Median :3.496   Median :3.479   Median :3.442   Median :2.9178
##   Mean   :3.500   Mean   :3.478   Mean   :3.424   Mean   :2.8588
##   3rd Qu.:3.532   3rd Qu.:3.510   3rd Qu.:3.525   3rd Qu.:3.1104
##   Max.   :3.701   Max.   :3.643   Max.   :3.715   Max.   :3.5757
##      gene_106        gene_107         gene_108        gene_109
##   Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :3.077
##   1st Qu.:1.258   1st Qu.:0.8149   1st Qu.:0.0000   1st Qu.:3.336
##   Median :1.550   Median :1.2034   Median :0.0000   Median :3.387
##   Mean   :1.520   Mean   :1.2001   Mean   :0.4880   Mean   :3.391
##   3rd Qu.:1.824   3rd Qu.:1.6183   3rd Qu.:0.7985   3rd Qu.:3.451
##   Max.   :3.215   Max.   :3.5010   Max.   :3.5315   Max.   :3.686
##      gene_110        gene_111         gene_112        gene_113
##   Min.   :3.016   Min.   :0.0000   Min.   :3.329   Min.   :2.963
##   1st Qu.:3.356   1st Qu.:0.7611   1st Qu.:3.500   1st Qu.:3.251
##   Median :3.449   Median :1.3485   Median :3.549   Median :3.297
##   Mean   :3.437   Mean   :1.3646   Mean   :3.555   Mean   :3.292
##   3rd Qu.:3.524   3rd Qu.:1.8313   3rd Qu.:3.601   3rd Qu.:3.339
##   Max.   :3.780   Max.   :3.4925   Max.   :3.933   Max.   :3.525
##      gene_114        gene_115        gene_116        gene_117
##   Min.   :2.580   Min.   :3.076   Min.   :2.739   Min.   :0.000
##   1st Qu.:3.099   1st Qu.:3.491   1st Qu.:3.166   1st Qu.:1.880
##   Median :3.189   Median :3.547   Median :3.248   Median :2.289
##   Mean   :3.181   Mean   :3.540   Mean   :3.235   Mean   :2.194
##   3rd Qu.:3.266   3rd Qu.:3.598   3rd Qu.:3.310   3rd Qu.:2.586
##   Max.   :3.658   Max.   :3.830   Max.   :3.630   Max.   :3.262
##      gene_118        gene_119        gene_120        gene_121
##   Min.   :3.290   Min.   :2.814   Min.   :3.023   Min.   :3.098
##   1st Qu.:3.633   1st Qu.:3.196   1st Qu.:3.369   1st Qu.:3.267
##   Median :3.710   Median :3.287   Median :3.426   Median :3.329
##   Mean   :3.720   Mean   :3.280   Mean   :3.419   Mean   :3.331
##   3rd Qu.:3.806   3rd Qu.:3.367   3rd Qu.:3.477   3rd Qu.:3.399
##   Max.   :4.110   Max.   :3.671   Max.   :3.732   Max.   :3.580
##      gene_122        gene_123        gene_124        gene_125
##   Min.   :2.496   Min.   :2.192   Min.   :3.318   Min.   :3.044
##   1st Qu.:2.965   1st Qu.:3.085   1st Qu.:3.488   1st Qu.:3.448
```

```
## Median :3.066    Median :3.165    Median :3.518    Median :3.497
## Mean   :3.084    Mean   :3.146    Mean   :3.519    Mean   :3.486
## 3rd Qu.:3.192    3rd Qu.:3.241    3rd Qu.:3.549    3rd Qu.:3.535
## Max.   :3.608    Max.   :3.496    Max.   :3.738    Max.   :3.761
##    gene_126         gene_127         gene_128         gene_129
## Min.   :1.514    Min.   :2.483    Min.   :3.273    Min.   :3.150
## 1st Qu.:2.960    1st Qu.:2.982    1st Qu.:3.558    1st Qu.:3.378
## Median :3.217    Median :3.120    Median :3.595    Median :3.447
## Mean   :3.147    Mean   :3.112    Mean   :3.590    Mean   :3.439
## 3rd Qu.:3.412    3rd Qu.:3.254    3rd Qu.:3.633    3rd Qu.:3.503
## Max.   :3.903    Max.   :3.477    Max.   :3.763    Max.   :3.703
##    gene_130         gene_131         gene_132         gene_133
## Min.   :2.830    Min.   :0.9895   Min.   :2.125    Min.   :0.000
## 1st Qu.:3.530    1st Qu.:2.6883   1st Qu.:3.110    1st Qu.:1.784
## Median :3.608    Median :2.8539   Median :3.305    Median :2.425
## Mean   :3.590    Mean   :2.8442   Mean   :3.265    Mean   :2.280
## 3rd Qu.:3.676    3rd Qu.:3.0257   3rd Qu.:3.483    3rd Qu.:2.926
## Max.   :3.899    Max.   :3.5548   Max.   :3.999    Max.   :3.472
##    gene_134         gene_135         gene_136         gene_137
## Min.   :0.000    Min.   :0.0000   Min.   :3.069    Min.   :3.131
## 1st Qu.:2.127    1st Qu.:0.0000   1st Qu.:3.518    1st Qu.:3.319
## Median :2.845    Median :0.6923   Median :3.570    Median :3.356
## Mean   :2.763    Mean   :0.7794   Mean   :3.567    Mean   :3.363
## 3rd Qu.:3.586    3rd Qu.:1.2705   3rd Qu.:3.616    3rd Qu.:3.407
## Max.   :4.113    Max.   :3.2696   Max.   :3.819    Max.   :3.590
##    gene_138         gene_139         gene_140         gene_141
## Min.   :2.799    Min.   :1.461    Min.   :3.051    Min.   :2.868
## 1st Qu.:3.237    1st Qu.:2.892    1st Qu.:3.433    1st Qu.:3.376
## Median :3.310    Median :3.077    Median :3.494    Median :3.468
## Mean   :3.298    Mean   :3.033    Mean   :3.498    Mean   :3.477
## 3rd Qu.:3.375    3rd Qu.:3.198    3rd Qu.:3.560    3rd Qu.:3.591
## Max.   :3.600    Max.   :3.697    Max.   :3.788    Max.   :3.902
##    gene_142         gene_143         gene_144         gene_145
## Min.   :3.167    Min.   :2.376    Min.   :2.832    Min.   :2.590
## 1st Qu.:3.478    1st Qu.:3.136    1st Qu.:3.298    1st Qu.:3.157
## Median :3.573    Median :3.266    Median :3.346    Median :3.238
## Mean   :3.568    Mean   :3.245    Mean   :3.344    Mean   :3.291
## 3rd Qu.:3.659    3rd Qu.:3.387    3rd Qu.:3.395    3rd Qu.:3.347
## Max.   :3.942    Max.   :3.737    Max.   :3.608    Max.   :3.915
##    gene_146         gene_147         gene_148         gene_149
## Min.   :3.050    Min.   :3.053    Min.   :0.000    Min.   :3.002
## 1st Qu.:3.265    1st Qu.:3.401    1st Qu.:1.244    1st Qu.:3.398
## Median :3.327    Median :3.436    Median :2.394    Median :3.451
## Mean   :3.351    Mean   :3.434    Mean   :2.144    Mean   :3.452
## 3rd Qu.:3.426    3rd Qu.:3.468    3rd Qu.:3.212    3rd Qu.:3.512
## Max.   :3.847    Max.   :3.616    Max.   :3.671    Max.   :3.824
##    gene_150         gene_151         gene_152         gene_153
## Min.   :2.827    Min.   :1.891    Min.   :3.437    Min.   :0.000
## 1st Qu.:3.361    1st Qu.:3.196    1st Qu.:3.660    1st Qu.:2.519
## Median :3.490    Median :3.313    Median :3.720    Median :2.853
## Mean   :3.478    Mean   :3.286    Mean   :3.719    Mean   :2.806
## 3rd Qu.:3.598    3rd Qu.:3.418    3rd Qu.:3.775    3rd Qu.:3.139
## Max.   :3.928    Max.   :3.655    Max.   :4.004    Max.   :3.799
##    gene_154         gene_155         gene_156         gene_157
```

```
##  Min.   :1.214   Min.   :3.112   Min.   :2.792   Min.   :3.189
##  1st Qu.:2.734   1st Qu.:3.423   1st Qu.:3.270   1st Qu.:3.416
##  Median :2.949   Median :3.472   Median :3.353   Median :3.490
##  Mean   :2.902   Mean   :3.466   Mean   :3.349   Mean   :3.492
##  3rd Qu.:3.139   3rd Qu.:3.513   3rd Qu.:3.424   3rd Qu.:3.560
##  Max.   :3.606   Max.   :3.729   Max.   :3.668   Max.   :3.908
##     gene_158        gene_159        gene_160        gene_161
##  Min.   :0.000   Min.   :3.303   Min.   :1.959   Min.   :3.142
##  1st Qu.:3.210   1st Qu.:3.503   1st Qu.:3.085   1st Qu.:3.391
##  Median :3.296   Median :3.548   Median :3.187   Median :3.447
##  Mean   :3.301   Mean   :3.552   Mean   :3.169   Mean   :3.441
##  3rd Qu.:3.405   3rd Qu.:3.607   3rd Qu.:3.283   3rd Qu.:3.489
##  Max.   :3.735   Max.   :3.757   Max.   :3.558   Max.   :3.700
##     gene_162        gene_163        gene_164        gene_165
##  Min.   :3.100   Min.   :0.000   Min.   :0.0000   Min.   :0.702
##  1st Qu.:3.269   1st Qu.:2.289   1st Qu.:0.8635   1st Qu.:2.612
##  Median :3.312   Median :2.578   Median :1.3717   Median :2.916
##  Mean   :3.316   Mean   :2.532   Mean   :1.4291   Mean   :2.805
##  3rd Qu.:3.363   3rd Qu.:2.817   3rd Qu.:2.0123   3rd Qu.:3.124
##  Max.   :3.558   Max.   :3.608   Max.   :3.4538   Max.   :3.527
##     gene_166        gene_167        gene_168        gene_169
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:1.868   1st Qu.:0.7078   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :2.232   Median :1.2152   Median :0.0000   Median :0.00000
##  Mean   :2.167   Mean   :1.1510   Mean   :0.1506   Mean   :0.06802
##  3rd Qu.:2.568   3rd Qu.:1.6056   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :3.299   Max.   :3.1673   Max.   :2.4055   Max.   :1.83513
##     gene_170        gene_171        gene_172        gene_173
##  Min.   :2.351   Min.   :2.751   Min.   :0.000   Min.   :0.0000
##  1st Qu.:3.003   1st Qu.:3.174   1st Qu.:1.620   1st Qu.:0.0000
##  Median :3.124   Median :3.238   Median :2.350   Median :0.5304
##  Mean   :3.111   Mean   :3.234   Mean   :2.297   Mean   :0.6949
##  3rd Qu.:3.246   3rd Qu.:3.298   3rd Qu.:3.168   3rd Qu.:1.1525
##  Max.   :3.588   Max.   :3.489   Max.   :3.833   Max.   :3.1239
##     gene_174        gene_175        gene_176        gene_177
##  Min.   :0.000   Min.   :2.730   Min.   :2.805   Min.   :0.000
##  1st Qu.:2.733   1st Qu.:3.171   1st Qu.:3.275   1st Qu.:2.337
##  Median :2.917   Median :3.248   Median :3.388   Median :2.747
##  Mean   :2.853   Mean   :3.251   Mean   :3.376   Mean   :2.664
##  3rd Qu.:3.054   3rd Qu.:3.328   3rd Qu.:3.477   3rd Qu.:3.018
##  Max.   :3.600   Max.   :3.784   Max.   :3.752   Max.   :3.800
##     gene_178        gene_179        gene_180        gene_181
##  Min.   :3.369   Min.   :3.350   Min.   :0.0000   Min.   :1.986
##  1st Qu.:3.603   1st Qu.:3.686   1st Qu.:0.0000   1st Qu.:2.996
##  Median :3.637   Median :3.731   Median :0.8447   Median :3.109
##  Mean   :3.633   Mean   :3.738   Mean   :1.2976   Mean   :3.067
##  3rd Qu.:3.667   3rd Qu.:3.797   3rd Qu.:2.5595   3rd Qu.:3.211
##  Max.   :3.814   Max.   :3.941   Max.   :3.6914   Max.   :3.552
##     gene_182        gene_183        gene_184        gene_185
##  Min.   :3.148   Min.   :3.397   Min.   :1.668   Min.   :0.0000
##  1st Qu.:3.426   1st Qu.:3.596   1st Qu.:2.624   1st Qu.:0.0000
##  Median :3.491   Median :3.648   Median :2.885   Median :0.0000
##  Mean   :3.483   Mean   :3.648   Mean   :2.863   Mean   :0.2752
##  3rd Qu.:3.543   3rd Qu.:3.698   3rd Qu.:3.114   3rd Qu.:0.5209
```

```
##  Max.   :3.882   Max.   :3.925   Max.   :3.640   Max.   :3.0813
##    gene_186        gene_187        gene_188        gene_189
##  Min.   :3.178   Min.   :2.315   Min.   :2.226   Min.   :2.030
##  1st Qu.:3.336   1st Qu.:3.222   1st Qu.:3.144   1st Qu.:2.777
##  Median :3.386   Median :3.298   Median :3.211   Median :2.995
##  Mean   :3.387   Mean   :3.290   Mean   :3.200   Mean   :2.964
##  3rd Qu.:3.432   3rd Qu.:3.369   3rd Qu.:3.275   3rd Qu.:3.180
##  Max.   :3.766   Max.   :3.699   Max.   :3.601   Max.   :3.580
##    gene_190        gene_191        gene_192        gene_193
##  Min.   :0.0000   Min.   :2.967   Min.   :2.829   Min.   :2.339
##  1st Qu.:0.8188   1st Qu.:3.257   1st Qu.:3.196   1st Qu.:3.256
##  Median :1.2913   Median :3.354   Median :3.265   Median :3.309
##  Mean   :1.2872   Mean   :3.356   Mean   :3.269   Mean   :3.302
##  3rd Qu.:1.8187   3rd Qu.:3.457   3rd Qu.:3.340   3rd Qu.:3.362
##  Max.   :3.2914   Max.   :3.749   Max.   :3.556   Max.   :3.605
##    gene_194        gene_195        gene_196        gene_197
##  Min.   :3.354   Min.   :0.000   Min.   :2.858   Min.   :0.000
##  1st Qu.:3.508   1st Qu.:2.900   1st Qu.:3.219   1st Qu.:1.089
##  Median :3.551   Median :3.098   Median :3.294   Median :1.654
##  Mean   :3.554   Mean   :3.033   Mean   :3.299   Mean   :1.710
##  3rd Qu.:3.603   3rd Qu.:3.230   3rd Qu.:3.377   3rd Qu.:2.501
##  Max.   :3.793   Max.   :3.759   Max.   :3.632   Max.   :3.319
##    gene_198        gene_199        gene_200        gene_201
##  Min.   :3.429   Min.   :3.248   Min.   :2.638   Min.   :2.535
##  1st Qu.:3.529   1st Qu.:3.437   1st Qu.:3.353   1st Qu.:3.151
##  Median :3.557   Median :3.488   Median :3.473   Median :3.238
##  Mean   :3.560   Mean   :3.485   Mean   :3.455   Mean   :3.227
##  3rd Qu.:3.587   3rd Qu.:3.533   3rd Qu.:3.584   3rd Qu.:3.315
##  Max.   :3.738   Max.   :3.670   Max.   :4.072   Max.   :3.606
##    gene_202        gene_203        gene_204        gene_205
##  Min.   :2.610   Min.   :0.000   Min.   :0.0000   Min.   :1.615
##  1st Qu.:3.205   1st Qu.:2.325   1st Qu.:0.0000   1st Qu.:2.383
##  Median :3.280   Median :2.723   Median :0.0000   Median :2.573
##  Mean   :3.268   Mean   :2.791   Mean   :0.3786   Mean   :2.559
##  3rd Qu.:3.342   3rd Qu.:3.085   3rd Qu.:0.6539   3rd Qu.:2.751
##  Max.   :3.725   Max.   :4.302   Max.   :2.4750   Max.   :3.245
##    gene_206        gene_207        gene_208        gene_209
##  Min.   :0.6299   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:2.1352   1st Qu.:0.573   1st Qu.:1.186   1st Qu.:1.599
##  Median :2.3180   Median :1.026   Median :1.526   Median :1.848
##  Mean   :2.2976   Mean   :1.038   Mean   :1.492   Mean   :1.832
##  3rd Qu.:2.4883   3rd Qu.:1.523   3rd Qu.:1.906   3rd Qu.:2.087
##  Max.   :3.2750   Max.   :2.964   Max.   :2.796   Max.   :3.569
##    gene_210        gene_211        gene_212        gene_213
##  Min.   :0.0000   Min.   :2.246   Min.   :2.518   Min.   :2.814
##  1st Qu.:0.6169   1st Qu.:3.153   1st Qu.:3.189   1st Qu.:3.499
##  Median :1.0595   Median :3.311   Median :3.257   Median :3.617
##  Mean   :0.9842   Mean   :3.284   Mean   :3.246   Mean   :3.602
##  3rd Qu.:1.4038   3rd Qu.:3.442   3rd Qu.:3.317   3rd Qu.:3.726
##  Max.   :2.5063   Max.   :3.780   Max.   :3.574   Max.   :4.013
##    gene_214        gene_215        gene_216        gene_217
##  Min.   :3.271   Min.   :2.487   Min.   :2.382   Min.   :0.000
##  1st Qu.:3.544   1st Qu.:3.330   1st Qu.:3.259   1st Qu.:1.790
##  Median :3.636   Median :3.450   Median :3.406   Median :2.108
```

```
## Mean   :3.633    Mean   :3.430    Mean   :3.417    Mean   :2.111
## 3rd Qu.:3.710    3rd Qu.:3.544    3rd Qu.:3.603    3rd Qu.:2.440
## Max.   :4.032    Max.   :3.874    Max.   :3.956    Max.   :3.548
##    gene_218         gene_219         gene_220         gene_221
## Min.   :0.000    Min.   :0.0000   Min.   :0.0000   Min.   :0.5961
## 1st Qu.:1.897    1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:2.6815
## Median :2.387    Median :0.0000   Median :0.0000   Median :3.1353
## Mean   :2.371    Mean   :0.8635   Mean   :0.7738   Mean   :2.9994
## 3rd Qu.:2.841    3rd Qu.:1.0010   3rd Qu.:0.7679   3rd Qu.:3.3553
## Max.   :3.926    Max.   :3.9155   Max.   :3.9016   Max.   :3.8047
##    gene_222         gene_223         gene_224         gene_225
## Min.   :0.00000  Min.   :0.000    Min.   :2.657    Min.   :2.824
## 1st Qu.:0.00000  1st Qu.:1.791    1st Qu.:3.410    1st Qu.:3.391
## Median :0.00000  Median :2.209    Median :3.504    Median :3.439
## Mean   :0.08817  Mean   :2.191    Mean   :3.485    Mean   :3.458
## 3rd Qu.:0.00000  3rd Qu.:2.674    3rd Qu.:3.593    3rd Qu.:3.514
## Max.   :1.71812  Max.   :3.681    Max.   :3.881    Max.   :3.781
##    gene_226         gene_227         gene_228         gene_229
## Min.   :0.000    Min.   :0.000    Min.   :2.880    Min.   :0.0000
## 1st Qu.:2.745    1st Qu.:1.363    1st Qu.:3.684    1st Qu.:0.0000
## Median :3.019    Median :1.901    Median :3.776    Median :0.9548
## Mean   :2.933    Mean   :1.891    Mean   :3.771    Mean   :0.9910
## 3rd Qu.:3.255    3rd Qu.:2.420    3rd Qu.:3.863    3rd Qu.:1.5978
## Max.   :3.901    Max.   :4.063    Max.   :4.232    Max.   :3.0703
##    gene_230         gene_231         gene_232         gene_233
## Min.   :3.986    Min.   :0.000    Min.   :3.870    Min.   :1.229
## 1st Qu.:4.092    1st Qu.:1.187    1st Qu.:4.053    1st Qu.:3.048
## Median :4.124    Median :1.882    Median :4.085    Median :3.282
## Mean   :4.123    Mean   :1.863    Mean   :4.084    Mean   :3.296
## 3rd Qu.:4.154    3rd Qu.:2.691    3rd Qu.:4.116    3rd Qu.:3.583
## Max.   :4.246    Max.   :3.827    Max.   :4.223    Max.   :4.247
##    gene_234         gene_235         gene_236         gene_237
## Min.   :3.214    Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:3.399    1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :3.460    Median :0.0000   Median :0.0000   Median :0.0000
## Mean   :3.461    Mean   :0.1625   Mean   :0.0697   Mean   :0.1859
## 3rd Qu.:3.520    3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.4411
## Max.   :3.733    Max.   :3.4065   Max.   :1.1114   Max.   :1.5789
##    gene_238         gene_239         gene_240         gene_241
## Min.   :0.000    Min.   :0.00000  Min.   :3.360    Min.   :0.000
## 1st Qu.:0.000    1st Qu.:0.00000  1st Qu.:3.677    1st Qu.:1.160
## Median :0.895    Median :0.00000  Median :3.733    Median :1.734
## Mean   :1.059    Mean   :0.03583  Mean   :3.726    Mean   :1.716
## 3rd Qu.:1.929    3rd Qu.:0.00000  3rd Qu.:3.781    3rd Qu.:2.303
## Max.   :3.779    Max.   :1.77096  Max.   :3.959    Max.   :3.882
##    gene_242         gene_243         gene_244         gene_245
## Min.   :0.0000   Min.   :3.614    Min.   :3.240    Min.   :3.367
## 1st Qu.:0.5995   1st Qu.:3.803    1st Qu.:3.416    1st Qu.:3.594
## Median :0.9658   Median :3.836    Median :3.451    Median :3.617
## Mean   :0.9117   Mean   :3.842    Mean   :3.450    Mean   :3.620
## 3rd Qu.:1.2733   3rd Qu.:3.876    3rd Qu.:3.482    3rd Qu.:3.648
## Max.   :3.6568   Max.   :4.099    Max.   :3.708    Max.   :3.769
##    gene_246         gene_247         gene_248         gene_249
## Min.   :3.224    Min.   :3.521    Min.   :2.707    Min.   :1.580
```

```
##    1st Qu.:3.484    1st Qu.:3.727    1st Qu.:2.991    1st Qu.:2.565
##    Median :3.528    Median :3.766    Median :3.088    Median :2.704
##    Mean   :3.518    Mean   :3.760    Mean   :3.082    Mean   :2.686
##    3rd Qu.:3.559    3rd Qu.:3.798    3rd Qu.:3.160    3rd Qu.:2.830
##    Max.   :3.743    Max.   :3.926    Max.   :3.599    Max.   :3.455
##       gene_250         gene_251         gene_252         gene_253
##    Min.   :3.447    Min.   :2.871    Min.   :3.071    Min.   :3.076
##    1st Qu.:3.661    1st Qu.:3.092    1st Qu.:3.275    1st Qu.:3.232
##    Median :3.699    Median :3.138    Median :3.316    Median :3.278
##    Mean   :3.699    Mean   :3.144    Mean   :3.319    Mean   :3.273
##    3rd Qu.:3.739    3rd Qu.:3.191    3rd Qu.:3.359    3rd Qu.:3.317
##    Max.   :3.893    Max.   :3.398    Max.   :3.582    Max.   :3.460
##       gene_254         Class
##    Min.   :0.00000   BRCA:300
##    1st Qu.:0.00000   COAD: 78
##    Median :0.00000   KIRC:146
##    Mean   :0.08832   LUAD:141
##    3rd Qu.:0.00000   PRAD:136
##    Max.   :2.14110
```

Normalization of feature values

Here, I normalize using Z score normalization method.

```
# Function for normalization
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x))) }

# Call the function
norm_data <- select(log_tranform_data, -Class)
norm_data <- as.data.frame(lapply(norm_data, normalize))
norm_data$Class <- log_tranform_data$Class

summary(norm_data)
```

```
##       gene_0            gene_1           gene_2           gene_3
##    Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##    1st Qu.:0.00000   1st Qu.:0.6031   1st Qu.:0.6245   1st Qu.:0.3165
##    Median :0.00000   Median :0.7183   Median :0.7235   Median :0.3929
##    Mean   :0.02228   Mean   :0.6710   Mean   :0.7011   Mean   :0.4016
##    3rd Qu.:0.00000   3rd Qu.:0.8013   3rd Qu.:0.8027   3rd Qu.:0.4721
##    Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##       gene_4           gene_5          gene_6           gene_7
##    Min.   :0.0000   Min.   : NA    Min.   :0.0000   Min.   :0.0000
##    1st Qu.:0.3838   1st Qu.: NA    1st Qu.:0.5113   1st Qu.:0.0000
##    Median :0.4979   Median : NA    Median :0.6223   Median :0.2759
##    Mean   :0.5015   Mean   :NaN    Mean   :0.6058   Mean   :0.2647
##    3rd Qu.:0.6166   3rd Qu.: NA    3rd Qu.:0.7107   3rd Qu.:0.4377
##    Max.   :1.0000   Max.   : NA    Max.   :1.0000   Max.   :1.0000
##                     NA's   :801
##       gene_8            gene_9           gene_10          gene_11
##    Min.   :0.00000   Min.   :0.000000   Min.   :0.0000   Min.   :0.0000
```

```
##   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.0000   1st Qu.:0.1527
##   Median :0.00000   Median :0.000000   Median :0.1189   Median :0.2694
##   Mean   :0.01132   Mean   :0.003872   Mean   :0.1495   Mean   :0.2672
##   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.2489   3rd Qu.:0.3958
##   Max.   :1.00000   Max.   :1.000000   Max.   :1.0000   Max.   :1.0000
##
##     gene_12           gene_13           gene_14           gene_15
##  Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.000000
##  1st Qu.:0.6705    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.000000
##  Median :0.7458    Median :0.0000    Median :0.0000    Median :0.000000
##  Mean   :0.7287    Mean   :0.1057    Mean   :0.1122    Mean   :0.006396
##  3rd Qu.:0.8113    3rd Qu.:0.1894    3rd Qu.:0.2430    3rd Qu.:0.000000
##  Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.000000
##
##     gene_16           gene_17           gene_18           gene_19
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.0000    Min.   :0.0000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.0000    1st Qu.:0.4967
##  Median :0.000000   Median :0.00000   Median :0.1998    Median :0.6182
##  Mean   :0.005322   Mean   :0.01972   Mean   :0.3098    Mean   :0.5976
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.5295    3rd Qu.:0.7177
##  Max.   :1.000000   Max.   :1.00000   Max.   :1.0000    Max.   :1.0000
##
##     gene_20           gene_21           gene_22           gene_23
##  Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   : NA
##  1st Qu.:0.6086    1st Qu.:0.0000    1st Qu.:0.5054    1st Qu.: NA
##  Median :0.6957    Median :0.5680    Median :0.5782    Median : NA
##  Mean   :0.6726    Mean   :0.4398    Mean   :0.5696    Mean   :NaN
##  3rd Qu.:0.7706    3rd Qu.:0.8180    3rd Qu.:0.6561    3rd Qu.: NA
##  Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   : NA
##                                                        NA's   :801
##     gene_24           gene_25           gene_26           gene_27
##  Min.   :0.00000   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000    1st Qu.:0.1472    1st Qu.:0.5552
##  Median :0.00000   Median :0.0000    Median :0.4186    Median :0.6361
##  Mean   :0.04186   Mean   :0.1802    Mean   :0.3947    Mean   :0.6246
##  3rd Qu.:0.00000   3rd Qu.:0.3382    3rd Qu.:0.6270    3rd Qu.:0.7095
##  Max.   :1.00000   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_28           gene_29           gene_30           gene_31
##  Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##  1st Qu.:0.5781    1st Qu.:0.4227    1st Qu.:0.0000    1st Qu.:0.0000
##  Median :0.7402    Median :0.5379    Median :0.0000    Median :0.1663
##  Mean   :0.6815    Mean   :0.5282    Mean   :0.2870    Mean   :0.2367
##  3rd Qu.:0.8207    3rd Qu.:0.6347    3rd Qu.:0.7677    3rd Qu.:0.4073
##  Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_32           gene_33           gene_34           gene_35
##  Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##  1st Qu.:0.4069    1st Qu.:0.1017    1st Qu.:0.5112    1st Qu.:0.5768
##  Median :0.4909    Median :0.3052    Median :0.6090    Median :0.6829
##  Mean   :0.4887    Mean   :0.3386    Mean   :0.6046    Mean   :0.6677
##  3rd Qu.:0.5763    3rd Qu.:0.5471    3rd Qu.:0.7180    3rd Qu.:0.7932
##  Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
```

```
##     gene_36          gene_37          gene_38          gene_39
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.5984   1st Qu.:0.0000
##  Median :0.1999   Median :0.0000   Median :0.6739   Median :0.1684
##  Mean   :0.2339   Mean   :0.0633   Mean   :0.6678   Mean   :0.2694
##  3rd Qu.:0.3774   3rd Qu.:0.0000   3rd Qu.:0.7384   3rd Qu.:0.5087
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##     gene_40          gene_41           gene_42           gene_43
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.3149   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.3959   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.3898   Mean   :0.02953   Mean   :0.03147   Mean   :0.09146
##  3rd Qu.:0.4679   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##
##     gene_44          gene_45          gene_46          gene_47
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.5754   1st Qu.:0.5645   1st Qu.:0.4938
##  Median :0.1916   Median :0.6712   Median :0.6534   Median :0.5672
##  Mean   :0.2721   Mean   :0.6583   Mean   :0.6467   Mean   :0.5615
##  3rd Qu.:0.5095   3rd Qu.:0.7678   3rd Qu.:0.7405   3rd Qu.:0.6385
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##     gene_48          gene_49          gene_50          gene_51
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.6730   1st Qu.:0.0000   1st Qu.:0.9280   1st Qu.:0.5304
##  Median :0.7375   Median :0.2710   Median :0.9411   Median :0.6021
##  Mean   :0.7304   Mean   :0.2699   Mean   :0.9381   Mean   :0.6023
##  3rd Qu.:0.7959   3rd Qu.:0.4758   3rd Qu.:0.9528   3rd Qu.:0.6703
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##     gene_52          gene_53          gene_54          gene_55
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.4389   1st Qu.:0.4440   1st Qu.:0.4384   1st Qu.:0.6026
##  Median :0.5346   Median :0.5144   Median :0.5080   Median :0.7264
##  Mean   :0.5335   Mean   :0.5129   Mean   :0.5031   Mean   :0.6948
##  3rd Qu.:0.6293   3rd Qu.:0.5856   3rd Qu.:0.5757   3rd Qu.:0.8142
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##     gene_56          gene_57          gene_58          gene_59
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.4161   1st Qu.:0.5030   1st Qu.:0.6014   1st Qu.:0.5744
##  Median :0.4765   Median :0.5990   Median :0.7032   Median :0.6773
##  Mean   :0.4828   Mean   :0.5945   Mean   :0.6972   Mean   :0.6573
##  3rd Qu.:0.5442   3rd Qu.:0.6861   3rd Qu.:0.8196   3rd Qu.:0.7707
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##     gene_60          gene_61          gene_62          gene_63
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.7426   1st Qu.:0.4506   1st Qu.:0.2055   1st Qu.:0.4730
##  Median :0.8031   Median :0.7191   Median :0.4079   Median :0.5869
##  Mean   :0.7927   Mean   :0.6220   Mean   :0.4220   Mean   :0.5713
##  3rd Qu.:0.8516   3rd Qu.:0.8214   3rd Qu.:0.6522   3rd Qu.:0.6917
```

```
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_64           gene_65           gene_66           gene_67
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.5259   1st Qu.:0.7024   1st Qu.:0.6865   1st Qu.:0.4310
## Median :0.6285   Median :0.7647   Median :0.7317   Median :0.5660
## Mean    :0.6167   Mean    :0.7508   Mean    :0.7091   Mean    :0.5641
## 3rd Qu.:0.7228   3rd Qu.:0.8218   3rd Qu.:0.7868   3rd Qu.:0.7019
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_68           gene_69           gene_70           gene_71
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.6758   1st Qu.:0.6273   1st Qu.:0.5975   1st Qu.:0.6634
## Median :0.7551   Median :0.7358   Median :0.6796   Median :0.8027
## Mean    :0.7365   Mean    :0.7084   Mean    :0.6692   Mean    :0.7371
## 3rd Qu.:0.8213   3rd Qu.:0.8103   3rd Qu.:0.7534   3rd Qu.:0.8727
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_72           gene_73           gene_74           gene_75
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.6481   1st Qu.:0.5055   1st Qu.:0.0000   1st Qu.:0.5868
## Median :0.7400   Median :0.5828   Median :0.1698   Median :0.6884
## Mean    :0.7159   Mean    :0.5798   Mean    :0.1932   Mean    :0.6695
## 3rd Qu.:0.8196   3rd Qu.:0.6603   3rd Qu.:0.3393   3rd Qu.:0.7759
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_76           gene_77           gene_78           gene_79
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.6207   1st Qu.:0.1496   1st Qu.:0.5654   1st Qu.:0.4172
## Median :0.6903   Median :0.3297   Median :0.6561   Median :0.4934
## Mean    :0.6830   Mean    :0.3243   Mean    :0.6458   Mean    :0.4860
## 3rd Qu.:0.7484   3rd Qu.:0.4808   3rd Qu.:0.7344   3rd Qu.:0.5623
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_80           gene_81           gene_82           gene_83
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.4757   1st Qu.:0.5734   1st Qu.:0.9006   1st Qu.:0.3268
## Median :0.5614   Median :0.6506   Median :0.9186   Median :0.5082
## Mean    :0.5536   Mean    :0.6517   Mean    :0.9177   Mean    :0.5399
## 3rd Qu.:0.6401   3rd Qu.:0.7287   3rd Qu.:0.9389   3rd Qu.:0.7953
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_84           gene_85           gene_86           gene_87
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.4451   1st Qu.:0.4834
## Median :0.0000   Median :0.2107   Median :0.5130   Median :0.5745
## Mean    :0.2205   Mean    :0.2732   Mean    :0.5122   Mean    :0.6101
## 3rd Qu.:0.4094   3rd Qu.:0.4939   3rd Qu.:0.5884   3rd Qu.:0.7573
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_88           gene_89           gene_90           gene_91
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.6201   1st Qu.:0.6514   1st Qu.:0.3519   1st Qu.:0.3425
## Median :0.7390   Median :0.7478   Median :0.4492   Median :0.5958
```

```
##   Mean   :0.7316    Mean   :0.7373    Mean   :0.4670    Mean   :0.5385
##   3rd Qu.:0.8710    3rd Qu.:0.8167    3rd Qu.:0.5865    3rd Qu.:0.7719
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_92           gene_93           gene_94           gene_95
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##   1st Qu.:0.6523    1st Qu.:0.6894    1st Qu.:0.2791    1st Qu.:0.6624
##   Median :0.7765    Median :0.7982    Median :0.5953    Median :0.7465
##   Mean   :0.7412    Mean   :0.7640    Mean   :0.5392    Mean   :0.7228
##   3rd Qu.:0.8708    3rd Qu.:0.8599    3rd Qu.:0.8196    3rd Qu.:0.8090
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_96           gene_97           gene_98           gene_99
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##   1st Qu.:0.3726    1st Qu.:0.5869    1st Qu.:0.3315    1st Qu.:0.5033
##   Median :0.4460    Median :0.6987    Median :0.4257    Median :0.6019
##   Mean   :0.4437    Mean   :0.6729    Mean   :0.4430    Mean   :0.5859
##   3rd Qu.:0.5141    3rd Qu.:0.7912    3rd Qu.:0.5540    3rd Qu.:0.6708
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_100          gene_101          gene_102          gene_103
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##   1st Qu.:0.3480    1st Qu.:0.5062    1st Qu.:0.4077    1st Qu.:0.6347
##   Median :0.4237    Median :0.5776    Median :0.4892    Median :0.6912
##   Mean   :0.4313    Mean   :0.5775    Mean   :0.4969    Mean   :0.6902
##   3rd Qu.:0.5046    3rd Qu.:0.6577    3rd Qu.:0.5792    3rd Qu.:0.7511
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_104          gene_105          gene_106          gene_107
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##   1st Qu.:0.5978    1st Qu.:0.6518    1st Qu.:0.3914    1st Qu.:0.2328
##   Median :0.7031    Median :0.7513    Median :0.4823    Median :0.3437
##   Mean   :0.6839    Mean   :0.7290    Mean   :0.4728    Mean   :0.3428
##   3rd Qu.:0.7943    3rd Qu.:0.8241    3rd Qu.:0.5672    3rd Qu.:0.4622
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_108          gene_109          gene_110          gene_111
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##   1st Qu.:0.0000    1st Qu.:0.4239    1st Qu.:0.4448    1st Qu.:0.2179
##   Median :0.0000    Median :0.5092    Median :0.5663    Median :0.3861
##   Mean   :0.1382    Mean   :0.5155    Mean   :0.5515    Mean   :0.3907
##   3rd Qu.:0.2261    3rd Qu.:0.6136    3rd Qu.:0.6651    3rd Qu.:0.5244
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_112          gene_113          gene_114          gene_115
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##   1st Qu.:0.2837    1st Qu.:0.5122    1st Qu.:0.4816    1st Qu.:0.5513
##   Median :0.3636    Median :0.5946    Median :0.5651    Median :0.6250
##   Mean   :0.3742    Mean   :0.5858    Mean   :0.5579    Mean   :0.6157
##   3rd Qu.:0.4511    3rd Qu.:0.6687    3rd Qu.:0.6367    3rd Qu.:0.6935
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##
##     gene_116          gene_117          gene_118          gene_119
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
```

```
## 1st Qu.:0.4784   1st Qu.:0.5764   1st Qu.:0.4185   1st Qu.:0.4459
## Median :0.5708   Median :0.7017   Median :0.5123   Median :0.5514
## Mean   :0.5562   Mean   :0.6726   Mean   :0.5244   Mean   :0.5432
## 3rd Qu.:0.6401   3rd Qu.:0.7926   3rd Qu.:0.6293   3rd Qu.:0.6451
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_120          gene_121          gene_122          gene_123
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.4878   1st Qu.:0.3498   1st Qu.:0.4219   1st Qu.:0.6844
## Median :0.5681   Median :0.4784   Median :0.5127   Median :0.7460
## Mean   :0.5582   Mean   :0.4840   Mean   :0.5294   Mean   :0.7314
## 3rd Qu.:0.6403   3rd Qu.:0.6243   3rd Qu.:0.6260   3rd Qu.:0.8047
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_124          gene_125          gene_126          gene_127
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.4059   1st Qu.:0.5637   1st Qu.:0.6053   1st Qu.:0.5024
## Median :0.4772   Median :0.6313   Median :0.7129   Median :0.6409
## Mean   :0.4796   Mean   :0.6173   Mean   :0.6833   Mean   :0.6329
## 3rd Qu.:0.5513   3rd Qu.:0.6854   3rd Qu.:0.7942   3rd Qu.:0.7758
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_128          gene_129          gene_130          gene_131
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.5810   1st Qu.:0.4124   1st Qu.:0.6547   1st Qu.:0.6622
## Median :0.6568   Median :0.5361   Median :0.7279   Median :0.7268
## Mean   :0.6465   Mean   :0.5215   Mean   :0.7117   Mean   :0.7230
## 3rd Qu.:0.7336   3rd Qu.:0.6377   3rd Qu.:0.7914   3rd Qu.:0.7938
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_132          gene_133          gene_134          gene_135
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.5257   1st Qu.:0.5138   1st Qu.:0.5173   1st Qu.:0.0000
## Median :0.6295   Median :0.6983   Median :0.6917   Median :0.2117
## Mean   :0.6082   Mean   :0.6565   Mean   :0.6719   Mean   :0.2384
## 3rd Qu.:0.7248   3rd Qu.:0.8428   3rd Qu.:0.8718   3rd Qu.:0.3886
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_136          gene_137          gene_138          gene_139
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.5981   1st Qu.:0.4105   1st Qu.:0.5466   1st Qu.:0.6402
## Median :0.6683   Median :0.4911   Median :0.6388   Median :0.7227
## Mean   :0.6633   Mean   :0.5057   Mean   :0.6236   Mean   :0.7033
## 3rd Qu.:0.7295   3rd Qu.:0.6009   3rd Qu.:0.7188   3rd Qu.:0.7768
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_140          gene_141          gene_142          gene_143
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.5186   1st Qu.:0.4916   1st Qu.:0.4018   1st Qu.:0.5587
## Median :0.6018   Median :0.5799   Median :0.5233   Median :0.6538
## Mean   :0.6073   Mean   :0.5889   Mean   :0.5169   Mean   :0.6384
## 3rd Qu.:0.6910   3rd Qu.:0.6990   3rd Qu.:0.6343   3rd Qu.:0.7431
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
```

```
##    gene_144          gene_145          gene_146          gene_147
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.6006   1st Qu.:0.4276   1st Qu.:0.2697   1st Qu.:0.6178
## Median :0.6621   Median :0.4887   Median :0.3478   Median :0.6799
## Mean   :0.6601   Mean   :0.5290   Mean   :0.3779   Mean   :0.6755
## 3rd Qu.:0.7254   3rd Qu.:0.5715   3rd Qu.:0.4717   3rd Qu.:0.7371
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_148          gene_149          gene_150          gene_151
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.3388   1st Qu.:0.4821   1st Qu.:0.4849   1st Qu.:0.7398
## Median :0.6521   Median :0.5465   Median :0.6024   Median :0.8060
## Mean   :0.5839   Mean   :0.5468   Mean   :0.5910   Mean   :0.7911
## 3rd Qu.:0.8748   3rd Qu.:0.6205   3rd Qu.:0.7007   3rd Qu.:0.8657
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_152          gene_153          gene_154          gene_155
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.3936   1st Qu.:0.6631   1st Qu.:0.6354   1st Qu.:0.5042
## Median :0.4986   Median :0.7509   Median :0.7250   Median :0.5827
## Mean   :0.4961   Mean   :0.7386   Mean   :0.7057   Mean   :0.5730
## 3rd Qu.:0.5952   3rd Qu.:0.8261   3rd Qu.:0.8046   3rd Qu.:0.6504
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_156          gene_157          gene_158          gene_159
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.5462   1st Qu.:0.3156   1st Qu.:0.8595   1st Qu.:0.4410
## Median :0.6405   Median :0.4183   Median :0.8825   Median :0.5394
## Mean   :0.6355   Mean   :0.4221   Mean   :0.8837   Mean   :0.5479
## 3rd Qu.:0.7220   3rd Qu.:0.5161   3rd Qu.:0.9115   3rd Qu.:0.6702
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_160          gene_161          gene_162          gene_163
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.7041   1st Qu.:0.4467   1st Qu.:0.3705   1st Qu.:0.6344
## Median :0.7677   Median :0.5469   Median :0.4642   Median :0.7144
## Mean   :0.7564   Mean   :0.5355   Mean   :0.4727   Mean   :0.7017
## 3rd Qu.:0.8278   3rd Qu.:0.6224   3rd Qu.:0.5750   3rd Qu.:0.7806
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_164          gene_165          gene_166          gene_167
## Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.2500   1st Qu.:0.6760   1st Qu.:0.5663   1st Qu.:0.2235
## Median :0.3971   Median :0.7838   Median :0.6768   Median :0.3837
## Mean   :0.4138   Mean   :0.7442   Mean   :0.6571   Mean   :0.3634
## 3rd Qu.:0.5826   3rd Qu.:0.8574   3rd Qu.:0.7786   3rd Qu.:0.5069
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##    gene_168          gene_169          gene_170          gene_171
## Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.5270   1st Qu.:0.5736
## Median :0.00000   Median :0.00000   Median :0.6244   Median :0.6599
## Mean   :0.06263   Mean   :0.03707   Mean   :0.6140   Mean   :0.6544
## 3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.7235   3rd Qu.:0.7411
```

```
##  Max.   :1.00000   Max.    :1.00000   Max.    :1.0000   Max.    :1.0000
##
##     gene_172          gene_173          gene_174          gene_175
##  Min.   :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##  1st Qu.:0.4227   1st Qu.:0.0000   1st Qu.:0.7590   1st Qu.:0.4184
##  Median :0.6132   Median :0.1698   Median :0.8103   Median :0.4915
##  Mean   :0.5993   Mean   :0.2225   Mean   :0.7925   Mean   :0.4941
##  3rd Qu.:0.8265   3rd Qu.:0.3689   3rd Qu.:0.8484   3rd Qu.:0.5672
##  Max.   :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_176          gene_177          gene_178          gene_179
##  Min.   :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##  1st Qu.:0.4959   1st Qu.:0.6150   1st Qu.:0.5261   1st Qu.:0.5685
##  Median :0.6157   Median :0.7228   Median :0.6027   Median :0.6454
##  Mean   :0.6032   Mean   :0.7010   Mean   :0.5942   Mean   :0.6576
##  3rd Qu.:0.7091   3rd Qu.:0.7941   3rd Qu.:0.6705   3rd Qu.:0.7575
##  Max.   :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_180          gene_181          gene_182          gene_183
##  Min.   :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##  1st Qu.:0.0000   1st Qu.:0.6449   1st Qu.:0.3779   1st Qu.:0.3776
##  Median :0.2288   Median :0.7174   Median :0.4670   Median :0.4756
##  Mean   :0.3515   Mean   :0.6905   Mean   :0.4562   Mean   :0.4761
##  3rd Qu.:0.6934   3rd Qu.:0.7825   3rd Qu.:0.5386   3rd Qu.:0.5708
##  Max.   :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_184          gene_185          gene_186          gene_187
##  Min.   :0.0000   Min.    :0.00000   Min.    :0.0000   Min.    :0.0000
##  1st Qu.:0.4852   1st Qu.:0.00000   1st Qu.:0.2700   1st Qu.:0.6553
##  Median :0.6172   Median :0.00000   Median :0.3548   Median :0.7100
##  Mean   :0.6059   Mean   :0.08931   Mean   :0.3560   Mean   :0.7048
##  3rd Qu.:0.7335   3rd Qu.:0.16906   3rd Qu.:0.4320   3rd Qu.:0.7615
##  Max.   :1.0000   Max.    :1.00000   Max.    :1.0000   Max.    :1.0000
##
##     gene_188          gene_189          gene_190          gene_191
##  Min.   :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##  1st Qu.:0.6679   1st Qu.:0.4818   1st Qu.:0.2488   1st Qu.:0.3715
##  Median :0.7166   Median :0.6226   Median :0.3923   Median :0.4956
##  Mean   :0.7086   Mean   :0.6027   Mean   :0.3911   Mean   :0.4975
##  3rd Qu.:0.7631   3rd Qu.:0.7422   3rd Qu.:0.5526   3rd Qu.:0.6269
##  Max.   :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_192          gene_193          gene_194          gene_195
##  Min.   :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##  1st Qu.:0.5051   1st Qu.:0.7244   1st Qu.:0.3506   1st Qu.:0.7714
##  Median :0.5990   Median :0.7667   Median :0.4487   Median :0.8240
##  Mean   :0.6051   Mean   :0.7607   Mean   :0.4561   Mean   :0.8070
##  3rd Qu.:0.7033   3rd Qu.:0.8083   3rd Qu.:0.5678   3rd Qu.:0.8594
##  Max.   :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##     gene_196          gene_197          gene_198          gene_199
##  Min.   :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##  1st Qu.:0.4662   1st Qu.:0.3281   1st Qu.:0.3221   1st Qu.:0.4480
##  Median :0.5628   Median :0.4983   Median :0.4156   Median :0.5683
```

```
## Mean    :0.5692    Mean    :0.5154    Mean    :0.4250    Mean    :0.5611
## 3rd Qu.:0.6708    3rd Qu.:0.7535    3rd Qu.:0.5121    3rd Qu.:0.6748
## Max.    :1.0000    Max.    :1.0000    Max.    :1.0000    Max.    :1.0000
##
##     gene_200           gene_201           gene_202           gene_203
## Min.    :0.0000    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
## 1st Qu.:0.4985    1st Qu.:0.5748    1st Qu.:0.5334    1st Qu.:0.5406
## Median :0.5822    Median :0.6560    Median :0.6010    Median :0.6329
## Mean    :0.5699    Mean    :0.6456    Mean    :0.5900    Mean    :0.6488
## 3rd Qu.:0.6600    3rd Qu.:0.7279    3rd Qu.:0.6568    3rd Qu.:0.7172
## Max.    :1.0000    Max.    :1.0000    Max.    :1.0000    Max.    :1.0000
##
##     gene_204           gene_205           gene_206           gene_207
## Min.    :0.0000    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
## 1st Qu.:0.0000    1st Qu.:0.4712    1st Qu.:0.5691    1st Qu.:0.1933
## Median :0.0000    Median :0.5879    Median :0.6382    Median :0.3461
## Mean    :0.1530    Mean    :0.5793    Mean    :0.6305    Mean    :0.3501
## 3rd Qu.:0.2642    3rd Qu.:0.6970    3rd Qu.:0.7026    3rd Qu.:0.5140
## Max.    :1.0000    Max.    :1.0000    Max.    :1.0000    Max.    :1.0000
##
##     gene_208           gene_209           gene_210           gene_211
## Min.    :0.0000    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
## 1st Qu.:0.4242    1st Qu.:0.4480    1st Qu.:0.2461    1st Qu.:0.5909
## Median :0.5458    Median :0.5179    Median :0.4227    Median :0.6943
## Mean    :0.5338    Mean    :0.5133    Mean    :0.3927    Mean    :0.6767
## 3rd Qu.:0.6817    3rd Qu.:0.5847    3rd Qu.:0.5601    3rd Qu.:0.7795
## Max.    :1.0000    Max.    :1.0000    Max.    :1.0000    Max.    :1.0000
##
##     gene_212           gene_213           gene_214           gene_215
## Min.    :0.0000    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
## 1st Qu.:0.6355    1st Qu.:0.5714    1st Qu.:0.3596    1st Qu.:0.6081
## Median :0.6999    Median :0.6696    Median :0.4795    Median :0.6942
## Mean    :0.6890    Mean    :0.6576    Mean    :0.4758    Mean    :0.6802
## 3rd Qu.:0.7565    3rd Qu.:0.7607    3rd Qu.:0.5775    3rd Qu.:0.7624
## Max.    :1.0000    Max.    :1.0000    Max.    :1.0000    Max.    :1.0000
##
##     gene_216           gene_217           gene_218           gene_219
## Min.    :0.0000    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
## 1st Qu.:0.5572    1st Qu.:0.5047    1st Qu.:0.4833    1st Qu.:0.0000
## Median :0.6501    Median :0.5942    Median :0.6081    Median :0.0000
## Mean    :0.6571    Mean    :0.5951    Mean    :0.6038    Mean    :0.2205
## 3rd Qu.:0.7758    3rd Qu.:0.6878    3rd Qu.:0.7237    3rd Qu.:0.2557
## Max.    :1.0000    Max.    :1.0000    Max.    :1.0000    Max.    :1.0000
##
##     gene_220           gene_221           gene_222            gene_223
## Min.    :0.0000    Min.    :0.0000    Min.    :0.00000    Min.    :0.0000
## 1st Qu.:0.0000    1st Qu.:0.6499    1st Qu.:0.00000    1st Qu.:0.4866
## Median :0.0000    Median :0.7914    Median :0.00000    Median :0.6001
## Mean    :0.1983    Mean    :0.7490    Mean    :0.05132    Mean    :0.5951
## 3rd Qu.:0.1968    3rd Qu.:0.8599    3rd Qu.:0.00000    3rd Qu.:0.7264
## Max.    :1.0000    Max.    :1.0000    Max.    :1.00000    Max.    :1.0000
##
##     gene_224           gene_225           gene_226           gene_227
## Min.    :0.0000    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
```

```
##    1st Qu.:0.6150   1st Qu.:0.5921   1st Qu.:0.7036   1st Qu.:0.3356
##    Median :0.6918   Median :0.6425   Median :0.7739   Median :0.4680
##    Mean   :0.6766   Mean   :0.6620   Mean   :0.7518   Mean   :0.4655
##    3rd Qu.:0.7649   3rd Qu.:0.7206   3rd Qu.:0.8344   3rd Qu.:0.5957
##    Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##       gene_228         gene_229         gene_230         gene_231
##    Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##    1st Qu.:0.5943   1st Qu.:0.0000   1st Qu.:0.4087   1st Qu.:0.3102
##    Median :0.6622   Median :0.3110   Median :0.5298   Median :0.4919
##    Mean   :0.6587   Mean   :0.3228   Mean   :0.5265   Mean   :0.4869
##    3rd Qu.:0.7265   3rd Qu.:0.5204   3rd Qu.:0.6469   3rd Qu.:0.7032
##    Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##       gene_232         gene_233         gene_234         gene_235
##    Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##    1st Qu.:0.5188   1st Qu.:0.6029   1st Qu.:0.3555   1st Qu.:0.0000
##    Median :0.6105   Median :0.6802   Median :0.4731   Median :0.0000
##    Mean   :0.6062   Mean   :0.6848   Mean   :0.4749   Mean   :0.0477
##    3rd Qu.:0.6972   3rd Qu.:0.7801   3rd Qu.:0.5899   3rd Qu.:0.0000
##    Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##       gene_236         gene_237         gene_238         gene_239
##    Min.   :0.00000  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##    1st Qu.:0.00000  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##    Median :0.00000  Median :0.0000   Median :0.2368   Median :0.00000
##    Mean   :0.06271  Mean   :0.1177   Mean   :0.2802   Mean   :0.02023
##    3rd Qu.:0.00000  3rd Qu.:0.2794   3rd Qu.:0.5103   3rd Qu.:0.00000
##    Max.   :1.00000  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##
##       gene_240         gene_241         gene_242         gene_243
##    Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##    1st Qu.:0.5295   1st Qu.:0.2989   1st Qu.:0.1639   1st Qu.:0.3897
##    Median :0.6220   Median :0.4468   Median :0.2641   Median :0.4585
##    Mean   :0.6110   Mean   :0.4422   Mean   :0.2493   Mean   :0.4705
##    3rd Qu.:0.7031   3rd Qu.:0.5933   3rd Qu.:0.3482   3rd Qu.:0.5405
##    Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##       gene_244         gene_245         gene_246         gene_247
##    Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##    1st Qu.:0.3751   1st Qu.:0.5635   1st Qu.:0.5008   1st Qu.:0.5075
##    Median :0.4494   Median :0.6217   Median :0.5861   Median :0.6036
##    Mean   :0.4471   Mean   :0.6284   Mean   :0.5651   Mean   :0.5898
##    3rd Qu.:0.5165   3rd Qu.:0.6985   3rd Qu.:0.6440   3rd Qu.:0.6831
##    Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##       gene_248         gene_249         gene_250         gene_251
##    Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##    1st Qu.:0.3186   1st Qu.:0.5254   1st Qu.:0.4790   1st Qu.:0.4199
##    Median :0.4277   Median :0.5991   Median :0.5654   Median :0.5073
##    Mean   :0.4206   Mean   :0.5897   Mean   :0.5650   Mean   :0.5188
##    3rd Qu.:0.5086   3rd Qu.:0.6667   3rd Qu.:0.6541   3rd Qu.:0.6084
##    Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
```

```
##       gene_252           gene_253           gene_254        Class
##   Min.    :0.0000   Min.    :0.0000   Min.    :0.00000   BRCA:300
##   1st Qu.:0.3992    1st Qu.:0.4077    1st Qu.:0.00000    COAD: 78
##   Median :0.4787    Median :0.5259    Median :0.00000    KIRC:146
##   Mean    :0.4856   Mean    :0.5147   Mean    :0.04125   LUAD:141
##   3rd Qu.:0.5634    3rd Qu.:0.6282    3rd Qu.:0.00000    PRAD:136
##   Max.    :1.0000   Max.    :1.0000   Max.    :1.00000
##
```

After normalization, we need to check if there are still any outliers.

```r
# Function for finding outliers
outliers <- function(data, column) {

  # Calculate mean and standard deviation
  mu <- mean(data[[column]], na.rm = TRUE)
  sd <- sd(data[[column]], na.rm = TRUE)

  # Calculate z-score
  z_scores <- (data[[column]] - mu) / sd

  # Identify z-scores above the threshold
  outlier_threshold <- which(abs(z_scores) > 2.5)

  return(outlier_threshold)
}

# Call the function for each gene
cols <- names(norm_data)[-c(1, ncol(norm_data))]
outliers_list <- list()

for(column in cols) {
  outliers_list[[column]] <- outliers(norm_data, column)
}
#outliers_list

# Total number of outliers
total_outliers <- sum(lengths(outliers_list))
print(paste("The total number of outliers are:", total_outliers))
```

```
## [1] "The total number of outliers are: 3530"
```

There was a decrease in the number of outliers to 3530. There still are outliers but these will be kept in the dataset since these outliers might show some extreme true positive values which might help during training the model.

Filtering out columns that only have Zero values

```r
# Remove columns with only 0 values
normalized_data <- select(norm_data, -Class)
zero_cols <- sapply(normalized_data, function(x) length(unique(x)) == 1)
constant_column_names <- names(normalized_data)[zero_cols]
normalized_data <- normalized_data[, !zero_cols]
```

```r
print(paste(constant_column_names, "removed because they only have zero values"))
```

```
## [1] "gene_5 removed because they only have zero values"
## [2] "gene_23 removed because they only have zero values"
```

Identification of principal components (PCA)

```r
# Apply PCA
pca_results <- prcomp(normalized_data)
str(pca_results)
```

```
## List of 5
##  $ sdev    : num [1:253] 1.118 0.807 0.756 0.587 0.573 ...
##  $ rotation: num [1:253, 1:253] 0.00916 -0.02814 -0.0306 -0.02991 -0.06632 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:253] "gene_0" "gene_1" "gene_2" "gene_3" ...
##   .. ..$ : chr [1:253] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : Named num [1:253] 0.0223 0.671 0.7011 0.4016 0.5015 ...
##   ..- attr(*, "names")= chr [1:253] "gene_0" "gene_1" "gene_2" "gene_3" ...
##  $ scale   : logi FALSE
##  $ x       : num [1:801, 1:253] -0.589 -0.449 -0.719 -0.381 2.529 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : NULL
##   .. ..$ : chr [1:253] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```

```r
summary(pca_results)
```

```
## Importance of components:
##                            PC1     PC2     PC3     PC4     PC5     PC6    PC7
## Standard deviation      1.1184 0.80659 0.75624 0.58725 0.57276 0.47273 0.4221
## Proportion of Variance  0.1692 0.08798 0.07734 0.04664 0.04437 0.03022 0.0241
## Cumulative Proportion   0.1692 0.25714 0.33448 0.38112 0.42548 0.45570 0.4798
##                            PC8     PC9    PC10    PC11    PC12    PC13   PC14
## Standard deviation      0.38890 0.35382 0.33115 0.29553 0.28477 0.28158 0.2665
## Proportion of Variance  0.02045 0.01693 0.01483 0.01181 0.01097 0.01072 0.0096
## Cumulative Proportion   0.50025 0.51718 0.53201 0.54383 0.55479 0.56552 0.5751
##                           PC15    PC16   PC17    PC18    PC19    PC20    PC21
## Standard deviation      0.26286 0.25236 0.2507 0.24640 0.24342 0.24085 0.23326
## Proportion of Variance  0.00934 0.00861 0.0085 0.00821 0.00801 0.00785 0.00736
## Cumulative Proportion   0.58446 0.59307 0.6016 0.60979 0.61780 0.62564 0.63300
##                           PC22    PC23    PC24    PC25    PC26    PC27    PC28
## Standard deviation      0.22929 0.22286 0.22012 0.21785 0.21530 0.21478 0.21121
## Proportion of Variance  0.00711 0.00672 0.00655 0.00642 0.00627 0.00624 0.00603
## Cumulative Proportion   0.64011 0.64683 0.65338 0.65980 0.66607 0.67231 0.67834
##                           PC29    PC30    PC31    PC32    PC33    PC34    PC35
## Standard deviation      0.20852 0.20581 0.20416 0.20070 0.19847 0.19735 0.19672
## Proportion of Variance  0.00588 0.00573 0.00564 0.00545 0.00533 0.00527 0.00523
## Cumulative Proportion   0.68422 0.68995 0.69559 0.70103 0.70636 0.71163 0.71686
##                           PC36    PC37    PC38    PC39    PC40    PC41    PC42
```

```
## Standard deviation      0.1922 0.19078 0.18872 0.18745 0.18518 0.18146 0.17969
## Proportion of Variance 0.0050 0.00492 0.00482 0.00475 0.00464 0.00445 0.00437
## Cumulative Proportion  0.7219 0.72678 0.73159 0.73635 0.74098 0.74544 0.74980
##                            PC43    PC44    PC45    PC46    PC47    PC48    PC49
## Standard deviation      0.17877 0.17801 0.17760 0.17384 0.17176 0.17062 0.16778
## Proportion of Variance 0.00432 0.00429 0.00427 0.00409 0.00399 0.00394 0.00381
## Cumulative Proportion  0.75412 0.75841 0.76268 0.76676 0.77075 0.77469 0.77850
##                            PC50    PC51    PC52    PC53    PC54    PC55    PC56
## Standard deviation      0.16653 0.16466 0.16385 0.16050 0.15970 0.15788 0.15710
## Proportion of Variance 0.00375 0.00367 0.00363 0.00348 0.00345 0.00337 0.00334
## Cumulative Proportion  0.78225 0.78591 0.78954 0.79303 0.79648 0.79985 0.80319
##                            PC57    PC58    PC59    PC60    PC61    PC62    PC63
## Standard deviation      0.15609 0.15447 0.15294 0.15256 0.15079 0.15021 0.14879
## Proportion of Variance 0.00329 0.00323 0.00316 0.00315 0.00307 0.00305 0.00299
## Cumulative Proportion  0.80648 0.80971 0.81287 0.81602 0.81909 0.82214 0.82514
##                            PC64    PC65    PC66    PC67    PC68    PC69    PC70
## Standard deviation      0.14693 0.14554 0.14485 0.14356 0.14221 0.14182 0.14071
## Proportion of Variance 0.00292 0.00286 0.00284 0.00279 0.00273 0.00272 0.00268
## Cumulative Proportion  0.82806 0.83092 0.83376 0.83655 0.83928 0.84200 0.84468
##                            PC71    PC72    PC73    PC74    PC75    PC76    PC77
## Standard deviation      0.13894 0.13735 0.13646 0.1360 0.13487 0.13387 0.13227
## Proportion of Variance 0.00261 0.00255 0.00252 0.0025 0.00246 0.00242 0.00237
## Cumulative Proportion  0.84729 0.84984 0.85236 0.8549 0.85732 0.85974 0.86211
##                            PC78    PC79    PC80    PC81    PC82    PC83    PC84
## Standard deviation      0.13138 0.1304 0.12950 0.12937 0.12793 0.12694 0.12502
## Proportion of Variance 0.00233 0.0023 0.00227 0.00226 0.00221 0.00218 0.00211
## Cumulative Proportion  0.86444 0.8667 0.86901 0.87127 0.87349 0.87567 0.87778
##                            PC85    PC86    PC87    PC88    PC89    PC90    PC91
## Standard deviation      0.12411 0.1217 0.12135 0.12025 0.11987 0.11886 0.11836
## Proportion of Variance 0.00208 0.0020 0.00199 0.00196 0.00194 0.00191 0.00189
## Cumulative Proportion  0.87986 0.8819 0.88386 0.88581 0.88776 0.88967 0.89156
##                            PC92    PC93    PC94    PC95    PC96    PC97    PC98
## Standard deviation      0.11660 0.11634 0.11466 0.11408 0.11251 0.11230 0.1120
## Proportion of Variance 0.00184 0.00183 0.00178 0.00176 0.00171 0.00171 0.0017
## Cumulative Proportion  0.89340 0.89523 0.89701 0.89877 0.90048 0.90219 0.9039
##                            PC99   PC100   PC101   PC102   PC103   PC104   PC105
## Standard deviation      0.11086 0.10956 0.1087 0.10736 0.10660 0.10583 0.10496
## Proportion of Variance 0.00166 0.00162 0.0016 0.00156 0.00154 0.00151 0.00149
## Cumulative Proportion  0.90555 0.90717 0.9088 0.91033 0.91186 0.91338 0.91487
##                           PC106   PC107   PC108   PC109   PC110   PC111   PC112
## Standard deviation      0.10421 0.10271 0.10248 0.10222 0.10081 0.10018 0.09955
## Proportion of Variance 0.00147 0.00143 0.00142 0.00141 0.00137 0.00136 0.00134
## Cumulative Proportion  0.91634 0.91776 0.91918 0.92060 0.92197 0.92333 0.92467
##                           PC113   PC114   PC115   PC116   PC117   PC118   PC119
## Standard deviation      0.09843 0.09781 0.09712 0.09638 0.09617 0.09480 0.09415
## Proportion of Variance 0.00131 0.00129 0.00128 0.00126 0.00125 0.00122 0.00120
## Cumulative Proportion  0.92598 0.92727 0.92855 0.92980 0.93105 0.93227 0.93347
##                           PC120   PC121   PC122   PC123   PC124   PC125   PC126
## Standard deviation      0.09296 0.09211 0.09189 0.09136 0.09027 0.09004 0.08905
## Proportion of Variance 0.00117 0.00115 0.00114 0.00113 0.00110 0.00110 0.00107
## Cumulative Proportion  0.93464 0.93578 0.93693 0.93806 0.93916 0.94025 0.94133
##                           PC127   PC128   PC129   PC130   PC131   PC132   PC133
## Standard deviation      0.08863 0.08748 0.08682 0.08655 0.08604 0.08545 0.08510
## Proportion of Variance 0.00106 0.00103 0.00102 0.00101 0.00100 0.00099 0.00098
```

```
## Cumulative Proportion  0.94239 0.94342 0.94444 0.94546 0.94646 0.94744 0.94842
##                           PC134   PC135   PC136   PC137   PC138   PC139   PC140
## Standard deviation      0.08481 0.08411 0.08305 0.08255 0.08194 0.08156 0.08104
## Proportion of Variance  0.00097 0.00096 0.00093 0.00092 0.00091 0.00090 0.00089
## Cumulative Proportion   0.94940 0.95035 0.95129 0.95221 0.95312 0.95402 0.95490
##                           PC141   PC142   PC143   PC144   PC145   PC146   PC147
## Standard deviation      0.08006 0.07888 0.07810 0.07761 0.07718 0.07654 0.07558
## Proportion of Variance  0.00087 0.00084 0.00082 0.00081 0.00081 0.00079 0.00077
## Cumulative Proportion   0.95577 0.95661 0.95744 0.95825 0.95906 0.95985 0.96062
##                           PC148   PC149   PC150   PC151   PC152   PC153   PC154
## Standard deviation      0.07495 0.07463 0.07413 0.07372 0.07347 0.07290 0.07221
## Proportion of Variance  0.00076 0.00075 0.00074 0.00073 0.00073 0.00072 0.00071
## Cumulative Proportion   0.96138 0.96213 0.96288 0.96361 0.96434 0.96506 0.96577
##                           PC155   PC156   PC157   PC158   PC159   PC160   PC161
## Standard deviation       0.0721 0.07140 0.07115 0.07033 0.06945 0.06929 0.06876
## Proportion of Variance   0.0007 0.00069 0.00068 0.00067 0.00065 0.00065 0.00064
## Cumulative Proportion    0.9665 0.96716 0.96784 0.96851 0.96916 0.96981 0.97045
##                           PC162   PC163   PC164   PC165   PC166   PC167   PC168
## Standard deviation      0.06832 0.06772 0.06696 0.06614 0.06569 0.06511 0.06486
## Proportion of Variance  0.00063 0.00062 0.00061 0.00059 0.00058 0.00057 0.00057
## Cumulative Proportion   0.97108 0.97170 0.97231 0.97290 0.97349 0.97406 0.97463
##                           PC169   PC170   PC171   PC172   PC173   PC174   PC175
## Standard deviation      0.06449 0.06420 0.06323 0.06307 0.06253 0.06204 0.06153
## Proportion of Variance  0.00056 0.00056 0.00054 0.00054 0.00053 0.00052 0.00051
## Cumulative Proportion   0.97519 0.97575 0.97629 0.97683 0.97736 0.97788 0.97839
##                           PC176   PC177   PC178   PC179   PC180   PC181   PC182
## Standard deviation      0.06113 0.06096 0.06036 0.06014 0.05953 0.05897 0.05892
## Proportion of Variance  0.00051 0.00050 0.00049 0.00049 0.00048 0.00047 0.00047
## Cumulative Proportion   0.97889 0.97940 0.97989 0.98038 0.98086 0.98133 0.98180
##                           PC183   PC184   PC185   PC186   PC187   PC188   PC189
## Standard deviation      0.05800 0.05794 0.05675 0.05656 0.05632 0.05587 0.05544
## Proportion of Variance  0.00045 0.00045 0.00044 0.00043 0.00043 0.00042 0.00042
## Cumulative Proportion   0.98225 0.98271 0.98314 0.98357 0.98400 0.98442 0.98484
##                           PC190   PC191   PC192   PC193   PC194   PC195   PC196
## Standard deviation      0.05485 0.05466 0.05441 0.05407 0.05324 0.05259 0.05250
## Proportion of Variance  0.00041 0.00040 0.00040 0.00040 0.00038 0.00037 0.00037
## Cumulative Proportion   0.98525 0.98565 0.98605 0.98645 0.98683 0.98720 0.98758
##                           PC197   PC198   PC199   PC200   PC201   PC202   PC203
## Standard deviation      0.05220 0.05158 0.05143 0.05046 0.05027 0.05012 0.04948
## Proportion of Variance  0.00037 0.00036 0.00036 0.00034 0.00034 0.00034 0.00033
## Cumulative Proportion   0.98795 0.98831 0.98866 0.98901 0.98935 0.98969 0.99002
##                           PC204   PC205   PC206   PC207   PC208   PC209   PC210
## Standard deviation      0.04906 0.04865 0.04800 0.04772 0.04745 0.04687 0.04657
## Proportion of Variance  0.00033 0.00032 0.00031 0.00031 0.00030 0.00030 0.00029
## Cumulative Proportion   0.99035 0.99067 0.99098 0.99128 0.99159 0.99189 0.99218
##                           PC211   PC212   PC213   PC214   PC215   PC216   PC217
## Standard deviation      0.04583 0.04572 0.04528 0.04500 0.04416 0.04359 0.04328
## Proportion of Variance  0.00028 0.00028 0.00028 0.00027 0.00026 0.00026 0.00025
## Cumulative Proportion   0.99246 0.99275 0.99302 0.99330 0.99356 0.99382 0.99407
##                           PC218   PC219   PC220   PC221   PC222   PC223   PC224
## Standard deviation      0.04307 0.04255 0.04196 0.04139 0.04100 0.04074 0.04006
## Proportion of Variance  0.00025 0.00024 0.00024 0.00023 0.00023 0.00022 0.00022
## Cumulative Proportion   0.99432 0.99457 0.99481 0.99504 0.99526 0.99549 0.99571
##                           PC225   PC226   PC227   PC228   PC229   PC230   PC231
```

```
## Standard deviation      0.03988 0.03934 0.03908 0.03876 0.03853 0.03813 0.03772
## Proportion of Variance  0.00022 0.00021 0.00021 0.00020 0.00020 0.00020 0.00019
## Cumulative Proportion   0.99592 0.99613 0.99634 0.99654 0.99674 0.99694 0.99713
##                           PC232   PC233   PC234   PC235   PC236   PC237   PC238
## Standard deviation      0.03716 0.03688 0.03671 0.03638 0.03534 0.03519 0.03447
## Proportion of Variance  0.00019 0.00018 0.00018 0.00018 0.00017 0.00017 0.00016
## Cumulative Proportion   0.99732 0.99750 0.99768 0.99786 0.99803 0.99820 0.99836
##                           PC239   PC240   PC241   PC242   PC243   PC244   PC245
## Standard deviation      0.03396 0.03369 0.03313 0.03273 0.03179 0.03147 0.03082
## Proportion of Variance  0.00016 0.00015 0.00015 0.00014 0.00014 0.00013 0.00013
## Cumulative Proportion   0.99851 0.99867 0.99882 0.99896 0.99910 0.99923 0.99936
##                           PC246   PC247   PC248   PC249   PC250   PC251   PC252
## Standard deviation      0.03073 0.02950 0.02920 0.02765 0.02544 0.01841 0.01546
## Proportion of Variance  0.00013 0.00012 0.00012 0.00010 0.00009 0.00005 0.00003
## Cumulative Proportion   0.99949 0.99961 0.99972 0.99982 0.99991 0.99996 0.99999
##                           PC253
## Standard deviation      0.008393
## Proportion of Variance  0.000010
## Cumulative Proportion   1.000000
```

```r
# Plot PCA graph
plot(pca_results, type = "l")
```

pca_results
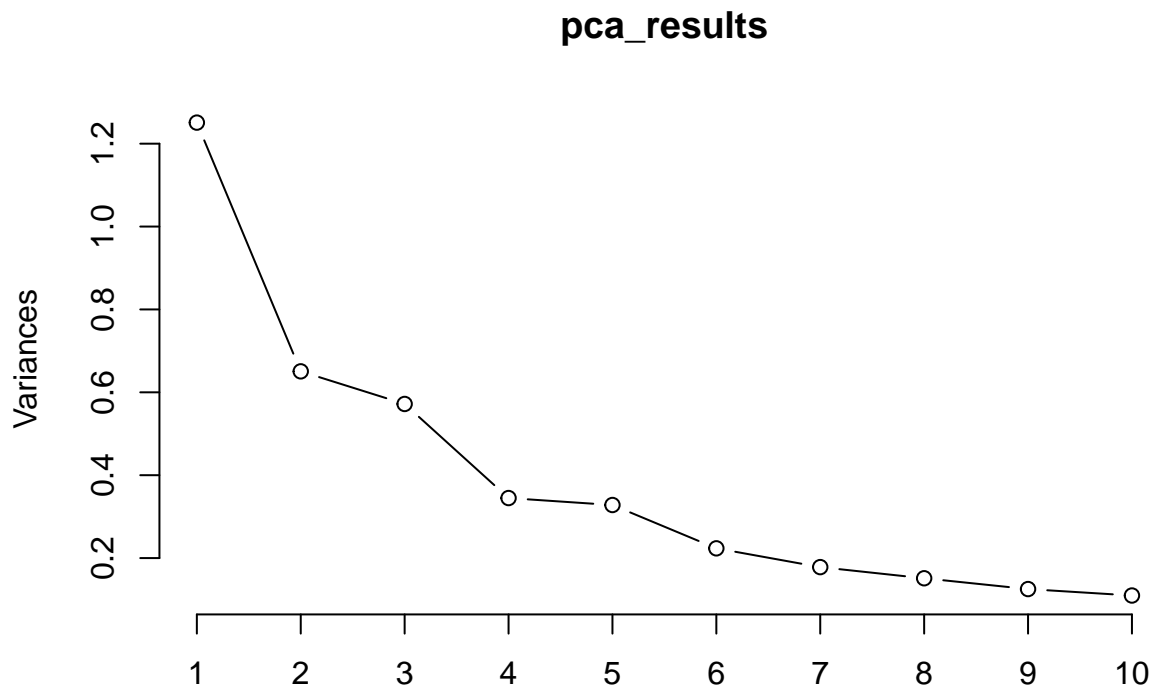
The exact choice depends on the level of variance you want to capture and the trade-off you're willing to make between dimensionality reduction and information retention.

Now, I make a new dataset with PCA features.

```
# Making a new dataset
pc_scores <- as.data.frame(pca_results$x)
pc_scores$Class <- norm_data$Class

# Components that reach 90% variance
pc_comp <- cumsum(pca_results$sdev^2 / sum(pca_results$sdev^2))
num_components <- which(pc_comp >= 0.90)[1]
pca_dataset <- pc_scores[, 1:num_components]
pca_dataset$Class <- norm_data$Class
```

By comparing both the results of pca_dataset and normalized_data we can try and get a better model.

```
# Normalized Dataset
normalized_data$Class <- norm_data$Class
str(normalized_data)
```

```
## 'data.frame':    801 obs. of  254 variables:
##  $ gene_0  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_1  : num  0.558 0.235 0.759 0.801 0.889 ...
##  $ gene_2  : num  0.742 0.486 0.775 0.78 0.816 ...
##  $ gene_3  : num  0.122 0.579 0.495 0.223 0.214 ...
##  $ gene_4  : num  0.712 0.44 0.653 0.317 0.158 ...
##  $ gene_6  : num  0.584 0.532 0.733 0.622 0.812 ...
##  $ gene_7  : num  0.35 0 0 0.539 0.89 ...
##  $ gene_8  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_9  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_10 : num  0.18 0 0 0 0.172 ...
##  $ gene_11 : num  0.388 0.212 0.503 0.466 0.204 ...
##  $ gene_12 : num  0.629 0.708 0.842 0.766 0.697 ...
##  $ gene_13 : num  0.247 0.37 0 0.597 0 ...
##  $ gene_14 : num  0 0 0 0 0 ...
##  $ gene_15 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_16 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_17 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_18 : num  0.188 0 0.394 0.66 0.946 ...
##  $ gene_19 : num  0.0379 0.8503 0.449 0.5537 0.729 ...
##  $ gene_20 : num  0.404 0.723 0.546 0.854 0.816 ...
##  $ gene_21 : num  0 0 0 0 0.859 ...
##  $ gene_22 : num  0.731 0.756 0.562 0.661 0.558 ...
##  $ gene_24 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gene_25 : num  0.51 0 0 0 0.542 ...
##  $ gene_26 : num  0.792 0.9 0.599 0.27 0 ...
##  $ gene_27 : num  0.897 0.504 0.687 0.624 0.587 ...
##  $ gene_28 : num  0.839 0.321 0.89 0.748 0.682 ...
##  $ gene_29 : num  0.802 0.415 0.723 0.653 0.41 ...
##  $ gene_30 : num  0 0 0 0 0.932 ...
##  $ gene_31 : num  0.689 0 0.176 0 0 ...
##  $ gene_32 : num  0.644 0.33 0.582 0.473 0.963 ...
##  $ gene_33 : num  0.657 0.317 0.431 0.11 0.309 ...
##  $ gene_34 : num  0.612 0.56 0.637 0.593 0.699 ...
##  $ gene_35 : num  0.837 0.442 0.707 0.615 0.847 ...
##  $ gene_36 : num  0 0.182 0.252 0 0.712 ...
##  $ gene_37 : num  0 0 0 0 0 ...
```

```
## $ gene_38 : num  0.755 0.679 0.79 0.691 0.63 ...
## $ gene_39 : num  0 0.264 0 0 0.89 ...
## $ gene_40 : num  0.568 0.391 0.23 0.283 0.198 ...
## $ gene_41 : num  0 0 0 0 0 0 0 0 0 ...
## $ gene_42 : num  0 0.163 0 0.17 0 ...
## $ gene_43 : num  0.397 0 0 0 0 ...
## $ gene_44 : num  0.275 0.511 0 0.442 0.266 ...
## $ gene_45 : num  0.719 0.698 0.453 0.659 0.703 ...
## $ gene_46 : num  0.804 0.59 0.619 0.791 0.379 ...
## $ gene_47 : num  0.546 0.377 0.478 0.571 0.596 ...
## $ gene_48 : num  0.912 0.775 0.779 0.745 0.604 ...
## $ gene_49 : num  0.362 0.218 0.323 0.558 0 ...
## $ gene_50 : num  0.944 0.958 0.932 0.95 0.944 ...
## $ gene_51 : num  0.807 0.577 0.656 0.701 0.592 ...
## $ gene_52 : num  0.742 0.731 0.369 0.638 0.37 ...
## $ gene_53 : num  0.222 0.65 0.551 0.366 0.54 ...
## $ gene_54 : num  0.374 0.511 0.52 0.562 0.55 ...
## $ gene_55 : num  0.746 0.731 0.661 0.681 0.858 ...
## $ gene_56 : num  0.563 0.662 0.538 0.581 0.455 ...
## $ gene_57 : num  0.659 0.586 0.581 0.526 0.636 ...
## $ gene_58 : num  0.814 0.486 0.769 0.687 0.647 ...
## $ gene_59 : num  0.661 0.208 0.783 0.643 0.852 ...
## $ gene_60 : num  0.829 0.778 0.823 0.87 0.778 ...
## $ gene_61 : num  0.261 0.562 0.598 0.551 0.811 ...
## $ gene_62 : num  0 0.624 0.468 0.189 0 ...
## $ gene_63 : num  0.405 0.119 0.621 0.636 0.673 ...
## $ gene_64 : num  0.404 0.653 0.559 0.653 0.933 ...
## $ gene_65 : num  0.76 0.616 0.703 0.778 0.846 ...
## $ gene_66 : num  0.746 0.71 0.75 0.811 0.765 ...
## $ gene_67 : num  0.571 0.862 0.692 0.655 0.513 ...
## $ gene_68 : num  0.772 0.519 0.636 0.754 0.882 ...
## $ gene_69 : num  0.636 0.691 0.831 0.732 0.775 ...
## $ gene_70 : num  0.68 0.774 0.656 0.779 0.595 ...
## $ gene_71 : num  0.859 0.675 0.803 0.763 0.836 ...
## $ gene_72 : num  0.717 0.717 0.875 0.645 0.798 ...
## $ gene_73 : num  0.513 0.521 0.656 0.713 0.591 ...
## $ gene_74 : num  0 0.194 0.267 0.123 0 ...
## $ gene_75 : num  0.716 0.652 0.701 0.636 0.819 ...
## $ gene_76 : num  0.75 0.412 0.644 0.755 0.68 ...
## $ gene_77 : num  0.555 0.109 0.46 0.394 0 ...
## $ gene_78 : num  0.812 0.697 0.552 0.745 0.755 ...
## $ gene_79 : num  0.402 0.579 0.441 0.493 0.678 ...
## $ gene_80 : num  0.824 0.743 0.676 0.675 0.406 ...
## $ gene_81 : num  0.585 0.835 0.659 0.664 0.58 ...
## $ gene_82 : num  0.915 0.935 0.913 0.93 0.888 ...
## $ gene_83 : num  0.902 0.423 0.912 0.774 0 ...
## $ gene_84 : num  0.677 0.188 0.721 0.457 0 ...
## $ gene_85 : num  0 0.111 0.626 0 0.176 ...
## $ gene_86 : num  0.444 0.631 0.442 0.473 0.429 ...
## $ gene_87 : num  0.176 0.673 0.581 0.519 0.909 ...
## $ gene_88 : num  0.574 0.876 0.723 0.612 0.914 ...
## $ gene_89 : num  0.91 0.731 0.602 0.585 0.852 ...
## $ gene_90 : num  0.454 0.467 0.47 0.66 0.462 ...
## $ gene_91 : num  0.684 0.587 0.491 0.296 0.802 ...
```

```
##   $ gene_92 : num  0.69 0.808 0.849 0.621 0.887 ...
##   $ gene_93 : num  0.742 0.826 0.753 0.701 0.884 ...
##   $ gene_94 : num  0.682 0.239 0.863 0.742 0.331 ...
##   $ gene_95 : num  0.656 0.709 0.724 0.774 0.822 ...
##   $ gene_96 : num  0.428 0.406 0.368 0.429 0.442 ...
##   $ gene_97 : num  0.541 0.801 0.783 0.706 0.634 ...
##   $ gene_98 : num  0.254 0.112 0.432 0.85 0.469 ...
##   $ gene_99 : num  0.731 0.638 0.556 0.673 0.77 ...
##   $ gene_100: num  0.0761 0.6376 0.4546 0.4907 0.3916 ...
##    [list output truncated]
```

PCA Dataset
```r
str(pca_dataset)
```

```
## 'data.frame':    801 obs. of  97 variables:
##   $ PC1  : num  -0.589 -0.449 -0.719 -0.381 2.529 ...
##   $ PC2  : num  -1.721 0.83 -0.168 -0.173 -0.218 ...
##   $ PC3  : num  1.073 0.216 -0.539 -0.743 -0.294 ...
##   $ PC4  : num  0.865 -0.213 0.65 0.659 -0.215 ...
##   $ PC5  : num  0.2693 0.8228 0.2523 -0.0369 -0.5031 ...
##   $ PC6  : num  0.17243 0.52843 -0.37922 -0.00356 -0.28121 ...
##   $ PC7  : num  -0.557 -0.542 -0.209 0.137 0.588 ...
##   $ PC8  : num  0.0154 0.552 -0.2753 0.1211 0.474 ...
##   $ PC9  : num  0.0552 -0.01599 0.43876 0.09318 -0.00508 ...
##   $ PC10 : num  0.4417 -0.1603 -0.0512 -0.2793 -0.1591 ...
##   $ PC11 : num  -0.1268 0.3798 -0.0123 0.4011 -0.202 ...
##   $ PC12 : num  0.179 0.405 0.105 0.322 0.206 ...
##   $ PC13 : num  -0.1695 -0.0369 0.1849 -0.3319 0.2469 ...
##   $ PC14 : num  0.3683 -0.2008 -0.2603 0.0549 -0.3363 ...
##   $ PC15 : num  -0.224 0.1129 -0.1645 -0.0309 0.1492 ...
##   $ PC16 : num  0.078 0.2954 0.0542 0.2489 -0.0208 ...
##   $ PC17 : num  -0.3744 0.1538 -0.1025 -0.071 -0.0968 ...
##   $ PC18 : num  -0.06916 -0.09388 0.00589 0.43028 -0.10765 ...
##   $ PC19 : num  -0.0637 -0.3527 0.4175 -0.5365 -0.1913 ...
##   $ PC20 : num  0.1397 -0.5106 -0.5255 0.4572 0.0904 ...
##   $ PC21 : num  -0.1785 0.1246 -0.3446 0.0794 0.4233 ...
##   $ PC22 : num  0.0864 -0.1101 -0.0122 0.1825 -0.097 ...
##   $ PC23 : num  -0.17911 -0.06451 0.00595 0.23942 -0.26843 ...
##   $ PC24 : num  0.4659 -0.0275 -0.2385 0.0263 0.1292 ...
##   $ PC25 : num  0.1456 -0.0471 -0.1374 -0.2344 -0.1132 ...
##   $ PC26 : num  -0.2085 0.0957 0.2273 0.1657 -0.0662 ...
##   $ PC27 : num  0.1386 0.0165 0.0752 0.0996 0.0821 ...
##   $ PC28 : num  0.2434 0.0687 -0.1563 -0.0708 0.1868 ...
##   $ PC29 : num  0.1049 0.3921 -0.1 -0.4124 -0.0327 ...
##   $ PC30 : num  -0.173 -0.259 -0.093 -0.23 0.125 ...
##   $ PC31 : num  0.247 0.197 -0.115 -0.396 0.124 ...
##   $ PC32 : num  0.306 -0.107 0.145 0.183 0.248 ...
##   $ PC33 : num  0.0299 0.4543 -0.1695 0.1127 -0.1124 ...
##   $ PC34 : num  -0.0105 -0.1304 0.0331 0.1638 -0.061 ...
##   $ PC35 : num  0.0193 0.276 0.5896 0.1451 -0.0228 ...
##   $ PC36 : num  0.0393 0.2431 -0.1663 -0.1834 -0.0775 ...
##   $ PC37 : num  -0.213 0.4529 0.0667 0.1341 0.0983 ...
##   $ PC38 : num  0.0375 -0.0827 0.1341 0.1571 -0.2988 ...
##   $ PC39 : num  -0.11746 0.00087 -0.27447 0.24548 -0.11395 ...
```

```
## $ PC40 : num  0.015512 -0.056263 -0.000736 0.128993 -0.036467 ...
## $ PC41 : num  -0.02985 0.28946 -0.34232 -0.00609 -0.139 ...
## $ PC42 : num  -0.3727 -0.162 -0.0572 -0.2823 0.0306 ...
## $ PC43 : num  0.177 0.0186 -0.032 0.0516 0.2825 ...
## $ PC44 : num  0.2032 0.1457 0.0809 0.2309 -0.0305 ...
## $ PC45 : num  0.15794 -0.13045 0.1239 0.00508 0.00694 ...
## $ PC46 : num  -0.0362 -0.0356 0.2561 0.2256 0.0106 ...
## $ PC47 : num  -0.0132 0.0466 -0.0898 -0.0438 -0.1046 ...
## $ PC48 : num  0.051 0.2582 0.0625 0.118 -0.1289 ...
## $ PC49 : num  0.1185 -0.0823 -0.2636 0.0434 -0.1319 ...
## $ PC50 : num  0.00823 0.26396 0.17417 0.08613 0.01066 ...
## $ PC51 : num  0.1572 -0.1406 -0.1438 0.2681 -0.0134 ...
## $ PC52 : num  -0.1789 0.2017 -0.0068 -0.1235 0.2526 ...
## $ PC53 : num  -0.3536 -0.1472 -0.1148 -0.2313 -0.0258 ...
## $ PC54 : num  -0.1068 0.0292 -0.045 -0.2078 -0.2214 ...
## $ PC55 : num  -0.204 0.1266 0.0473 0.029 0.0596 ...
## $ PC56 : num  0.0194 0.3419 -0.062 -0.0297 -0.0176 ...
## $ PC57 : num  0.3806 -0.098 0.0292 -0.072 0.134 ...
## $ PC58 : num  -0.0576 0.0196 0.042 0.195 0.1514 ...
## $ PC59 : num  0.0878 0.2817 0.0914 0.0909 -0.0612 ...
## $ PC60 : num  0.07442 -0.18565 0.00253 0.04425 0.19227 ...
## $ PC61 : num  0.2111 0.0383 0.0174 0.3582 -0.3702 ...
## $ PC62 : num  -0.04492 -0.00798 0.01317 -0.03779 0.02321 ...
## $ PC63 : num  -0.1158 -0.045 0.1946 0.079 -0.0494 ...
## $ PC64 : num  -0.0137 -0.1734 0.0768 -0.2336 -0.0197 ...
## $ PC65 : num  0.1983 0.0619 -0.0239 -0.1185 -0.22 ...
## $ PC66 : num  -0.1361 -0.0799 -0.0768 -0.0509 -0.2281 ...
## $ PC67 : num  -0.0199 0.1163 0.0193 -0.1093 -0.0763 ...
## $ PC68 : num  0.0683 -0.2143 0.1792 -0.0129 0.0382 ...
## $ PC69 : num  -0.0631 0.1658 -0.0278 0.1504 -0.0864 ...
## $ PC70 : num  -0.0378 0.4273 0.119 -0.1479 0.2318 ...
## $ PC71 : num  0.0476 0.3345 0.0414 0.2547 0.2457 ...
## $ PC72 : num  0.01592 -0.00874 -0.06384 -0.00419 0.10865 ...
## $ PC73 : num  0.0359 0.0169 0.0148 -0.1581 0.165 ...
## $ PC74 : num  0.1973 -0.1902 -0.0882 0.3188 -0.0524 ...
## $ PC75 : num  0.0372 -0.2095 -0.0018 0.122 -0.0324 ...
## $ PC76 : num  0.1217 0.1299 0.0095 0.3094 -0.1639 ...
## $ PC77 : num  -0.04769 -0.00393 0.14994 0.16123 0.13297 ...
## $ PC78 : num  0.146 -0.1074 0.0178 0.0414 0.1922 ...
## $ PC79 : num  -0.0278 -0.1857 -0.1205 -0.0483 -0.0289 ...
## $ PC80 : num  -0.0225 -0.0171 -0.0274 -0.0421 0.1162 ...
## $ PC81 : num  -0.0183 -0.088 0.0103 -0.1418 -0.1661 ...
## $ PC82 : num  0.0198 -0.1397 -0.211 -0.2193 0.0252 ...
## $ PC83 : num  -0.07155 -0.05263 -0.00407 -0.03725 -0.11106 ...
## $ PC84 : num  -0.2204 -0.0553 0.149 -0.0713 -0.0365 ...
## $ PC85 : num  -0.0388 0.0179 -0.0115 -0.1076 0.0479 ...
## $ PC86 : num  0.0698 0.102 0.1852 -0.0799 0.2056 ...
## $ PC87 : num  0.1277 -0.0728 -0.0872 -0.0192 -0.1235 ...
## $ PC88 : num  -0.1291 0.0745 -0.0298 0.0338 0.0458 ...
## $ PC89 : num  0.0408 0.0407 -0.1564 0.1556 0.1068 ...
## $ PC90 : num  0.09327 0.03245 0.02407 0.00961 -0.02265 ...
## $ PC91 : num  -0.0608 -0.2627 0.1664 0.1612 0.0171 ...
## $ PC92 : num  -0.144 -0.1557 -0.0821 0.1176 -0.1422 ...
## $ PC93 : num  -0.21347 0.00036 -0.05136 -0.03996 0.05096 ...
```

```
##  $ PC94 : num  0.0319 0.1321 0.0242 -0.1295 -0.0905 ...
##  $ PC95 : num  0.2796 0.0687 0.0513 -0.2606 0.089 ...
##  $ PC96 : num  -0.12616 -0.22815 -0.00604 -0.12835 0.02608 ...
##  $ Class: Factor w/ 5 levels "BRCA","COAD",..: 5 4 1 1 3 1 3 4 3 4 ...
```

Modeling

Model Construction

Creation of training & validation subsets for normalized_data

```
## Loading required package: caret
```

```
## Loading required package: lattice
```

```r
set.seed(3434)
# Subsets based on index and classes
norm_train_index <- createDataPartition(normalized_data$Class, p = 0.8, list = FALSE)

# Training set
norm_train_subset <- normalized_data[norm_train_index, ]

# Testing set
norm_test_subset <- normalized_data[-norm_train_index, ]
```

Creation of training & validation subsets for pca_dataset

```r
set.seed(3434)
# Subsets based on index and classes
pca_train_index <- createDataPartition(pca_dataset$Class, p = 0.8, list = FALSE)

# Training set
pca_train_subset <- pca_dataset[pca_train_index, ]

# Testing set
pca_test_subset <- pca_dataset[-pca_train_index, ]
```

Creation of model A: SVM

My model A is SVM. I train my model on both normalized_data and pca_dataset.

```
## Loading required package: kernlab
```

```
## Warning: package 'kernlab' was built under R version 4.3.3
```

```
##
## Attaching package: 'kernlab'
```

```
## The following object is masked from 'package:ggplot2':
##
##      alpha
```

```r
# SVM model for norm_train_subset
set.seed(3434)
svm_model_norm <- ksvm(Class ~ ., data = norm_train_subset, kernel = "rbfdot")
```

```r
# SVM model for pca_train_subset
set.seed(3434)
svm_model_pca <- ksvm(Class ~ ., data = pca_train_subset, kernel = "rbfdot")
```

Creation of model B: Gradient Boosting

My model B is Gradient Boosting. I train my model on both normalized_data and pca_dataset.

```
## Loading required package: gbm
```

```
## Warning: package 'gbm' was built under R version 4.3.3
```

```
## Loaded gbm 2.2.2
```

```
## This version of gbm is no longer under development. Consider transitioning to gbm3, https://github.co
```

```r
# Gradient Boosting model for normalized_data
set.seed(3434)
gbm_model_norm <- train(Class ~ ., data = norm_train_subset, method = "gbm", verbose = FALSE)
```

```r
# Gradient Boosting model for normalized_data
set.seed(3434)
gbm_model_pca <- train(Class ~ ., data = pca_train_subset, method = "gbm", verbose = FALSE)
```

Creation of model C: Random Forest

My model C is Random Forest. I train my model on both normalized_data and pca_dataset.

```r
if (!require("randomForest", character.only = TRUE)) {
  install.packages("randomForest", dependencies = TRUE)
}
```

```
## Loading required package: randomForest
```

```
## Warning: package 'randomForest' was built under R version 4.3.3
```

```
## randomForest 4.7-1.2
```

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

```
##
## Attaching package: 'randomForest'

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

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

```r
library(randomForest)
```

```r
# Random Forest model for normalized_data
set.seed(3434)
rf_model_norm <- randomForest(Class ~ ., data = norm_train_subset, verbose = FALSE)
```

```r
# Random Forest model for pca_train_subset
set.seed(3434)
rf_model_pca <- randomForest(Class ~ ., data = pca_train_subset, verbose = FALSE)
```

```r
importance(rf_model_norm)
```

```
##          MeanDecreaseGini
## gene_0        0.017873789
## gene_1        0.309557988
## gene_2        0.390139843
## gene_3        3.044044147
## gene_4        0.192644469
## gene_6        0.181528430
## gene_7        0.182836118
## gene_8        0.032891385
## gene_9        0.003933333
## gene_10       0.192089172
## gene_11       0.171633046
## gene_12       0.159830439
## gene_13       0.057072486
## gene_14       0.056459881
## gene_15       0.006000000
## gene_16       0.007058824
## gene_17       0.027030885
## gene_18       8.653397187
## gene_19       0.240455700
## gene_20       0.202346539
## gene_21       0.120755821
## gene_22       0.319306225
## gene_24       0.010431373
## gene_25       0.148593366
## gene_26       3.913090377
## gene_27       0.546843023
## gene_28       3.812953891
## gene_29       4.266874612
```

```
## gene_30      10.790092527
## gene_31       1.091797539
## gene_32       0.514987916
## gene_33       0.654128834
## gene_34       1.301820150
## gene_35       2.076826985
## gene_36       2.750301633
## gene_37       0.347061094
## gene_38       0.224941286
## gene_39       2.277543735
## gene_40       0.269958528
## gene_41       0.033322364
## gene_42       0.019031452
## gene_43       0.308080798
## gene_44       3.089119257
## gene_45       4.375171086
## gene_46       3.015318232
## gene_47       0.960115611
## gene_48       0.591575772
## gene_49       0.189506726
## gene_50       0.262174407
## gene_51       0.172189436
## gene_52       0.251833250
## gene_53       0.257244052
## gene_54       0.347986010
## gene_55       2.530369185
## gene_56       0.232405102
## gene_57       0.383803192
## gene_58       0.683803582
## gene_59       0.274211378
## gene_60       0.222660326
## gene_61       3.907935815
## gene_62       4.130431538
## gene_63       0.478845220
## gene_64       2.030397018
## gene_65       0.328246367
## gene_66       3.453397497
## gene_67       1.999902769
## gene_68       0.599706467
## gene_69       1.334597231
## gene_70       0.935033957
## gene_71       0.810278652
## gene_72       0.480090788
## gene_73       0.241988604
## gene_74       0.148772553
## gene_75       4.022204803
## gene_76       0.411221571
## gene_77       0.368278100
## gene_78       1.655027777
## gene_79       0.139515176
## gene_80       0.188807178
## gene_81       0.840355859
## gene_82       2.635395835
## gene_83       9.597623251
```

```
## gene_84      0.686795104
## gene_85      0.263875326
## gene_86      0.950163870
## gene_87      2.369170864
## gene_88     15.055396348
## gene_89     19.731405774
## gene_90      0.385104563
## gene_91      8.390584860
## gene_92      3.103981435
## gene_93      2.594571150
## gene_94      3.180280547
## gene_95      1.175797362
## gene_96      0.204997436
## gene_97      0.292278322
## gene_98      1.231746355
## gene_99      0.569536328
## gene_100     0.343783488
## gene_101     0.800605035
## gene_102     0.217369697
## gene_103     0.201752872
## gene_104     0.289769902
## gene_105     1.298053181
## gene_106     0.343822376
## gene_107     0.762103751
## gene_108     0.507925248
## gene_109     1.402004879
## gene_110     0.432179109
## gene_111     0.219314381
## gene_112     0.219072159
## gene_113     0.172934407
## gene_114     0.581658543
## gene_115     0.504432513
## gene_116     0.630970009
## gene_117     1.343793679
## gene_118     0.981765633
## gene_119     1.231862223
## gene_120     0.695379822
## gene_121     2.869610156
## gene_122     5.110875959
## gene_123     3.420685791
## gene_124     0.310455040
## gene_125     0.360866644
## gene_126     1.091799123
## gene_127     1.106869621
## gene_128     0.306517196
## gene_129     2.001586445
## gene_130     0.447215483
## gene_131     1.978386175
## gene_132     2.866629523
## gene_133     3.994281678
## gene_134     7.345719263
## gene_135     0.985096894
## gene_136     1.395391612
## gene_137     0.420757764
```

```
gene_138   0.149874603
gene_139   0.434879552
gene_140   1.076980631
gene_141   5.930685043
gene_142   7.773887301
gene_143   0.909810859
gene_144   0.243607431
gene_145   10.296663075
gene_146   4.443217216
gene_147   0.551241630
gene_148   13.413845113
gene_149   0.161963548
gene_150   1.392330383
gene_151   1.450695108
gene_152   0.285345204
gene_153   1.164807130
gene_154   0.583368459
gene_155   0.258271209
gene_156   0.267431073
gene_157   0.441562080
gene_158   0.884286386
gene_159   0.512035254
gene_160   1.088460355
gene_161   0.237411846
gene_162   0.371188284
gene_163   0.851934960
gene_164   0.211108685
gene_165   2.872639904
gene_166   0.199751952
gene_167   0.079799607
gene_168   0.085434390
gene_169   0.021021067
gene_170   0.457791116
gene_171   0.435769180
gene_172   7.605730602
gene_173   0.421466894
gene_174   0.196489663
gene_175   0.241097880
gene_176   0.369219622
gene_177   3.403615381
gene_178   0.299088216
gene_179   1.205942952
gene_180   10.676082597
gene_181   2.283019549
gene_182   1.001245141
gene_183   0.543311717
gene_184   9.327350164
gene_185   0.113308404
gene_186   0.239871297
gene_187   0.173168357
gene_188   0.227950757
gene_189   4.846322713
gene_190   0.964628509
gene_191   1.255123201
```

```
gene_192   0.878727214
gene_193   0.757579041
gene_194   0.629029779
gene_195   0.380009446
gene_196   0.780214680
gene_197   6.329066574
gene_198   0.235398517
gene_199   1.167819799
gene_200   1.314384874
gene_201   0.542678139
gene_202   2.104085687
gene_203   19.026449955
gene_204   3.046428372
gene_205   0.222497689
gene_206   0.173610666
gene_207   0.602268935
gene_208   0.241129198
gene_209   0.256944533
gene_210   0.252243739
gene_211   0.965695063
gene_212   0.886211877
gene_213   3.692869274
gene_214   3.133529581
gene_215   5.267067628
gene_216   12.090232214
gene_217   0.703646529
gene_218   6.689253710
gene_219   9.815136109
gene_220   13.159993250
gene_221   19.841468688
gene_222   0.088654135
gene_223   9.611731066
gene_224   2.489945941
gene_225   2.397676259
gene_226   2.481018883
gene_227   0.743083362
gene_228   1.270987270
gene_229   0.302217656
gene_230   0.483450890
gene_231   7.825290404
gene_232   0.400296057
gene_233   8.412534863
gene_234   1.756547801
gene_235   0.098178939
gene_236   0.007000000
gene_237   0.049765320
gene_238   0.813379115
gene_239   0.049783911
gene_240   0.697392078
gene_241   0.996678139
gene_242   0.102301166
gene_243   1.373887592
gene_244   0.240822831
gene_245   0.194051416
```

```
## gene_246      0.569072476
## gene_247      1.015776626
## gene_248      0.711448643
## gene_249      0.365639835
## gene_250      1.187603529
## gene_251      0.268376663
## gene_252      0.483857672
## gene_253      0.287618703
## gene_254      0.141403617
```

```r
# Plot variable importance
varImpPlot(rf_model_norm)
```


rf_model_norm
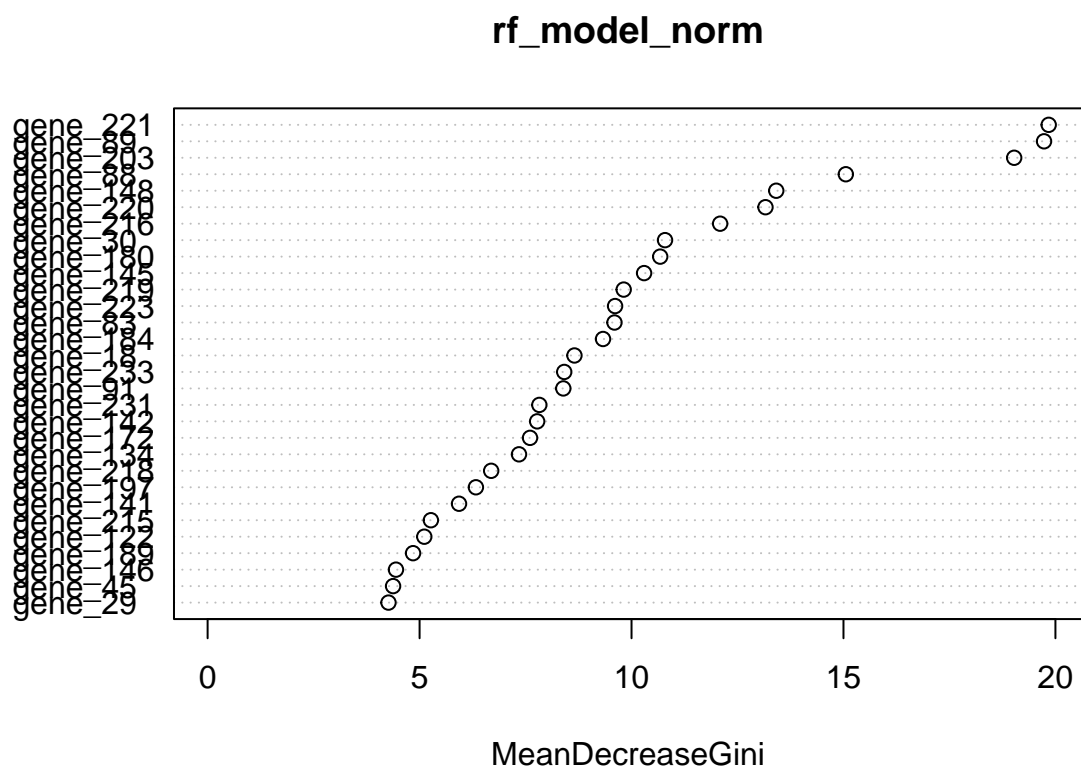
```r
varImpPlot(rf_model_pca)
```

rf_model_pca
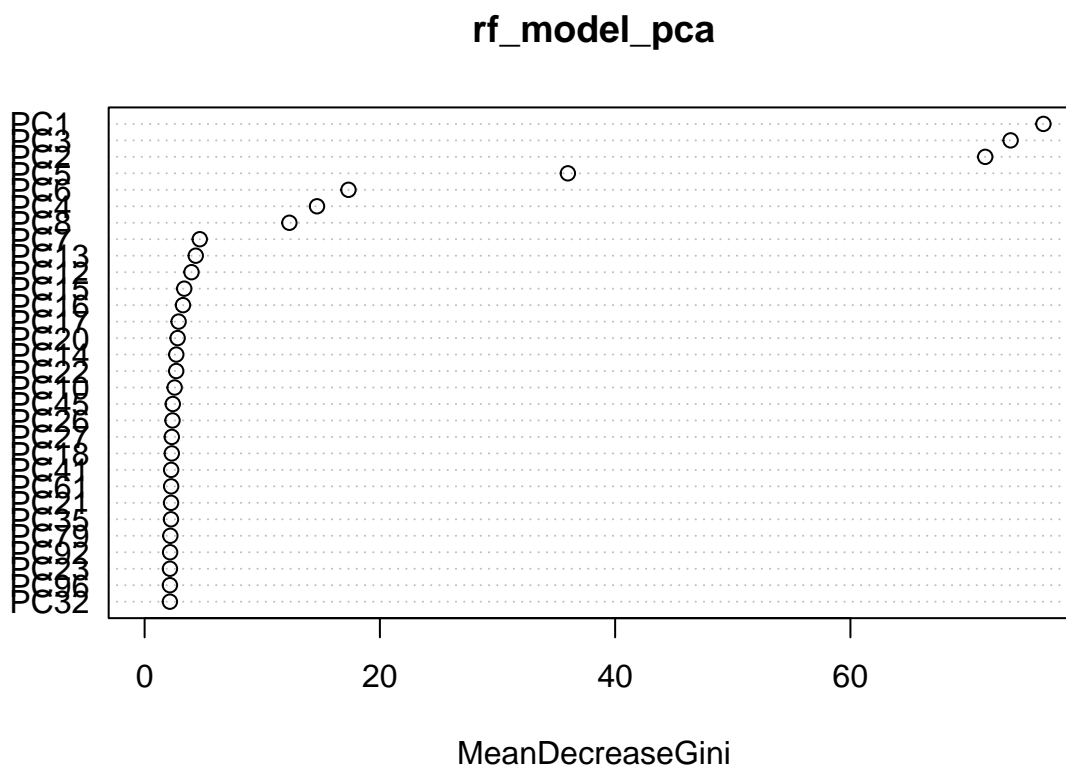
This plot displays feature importance scores from a Random Forest model, where variables are possibly principal components (PC1, PC2, …, PC19) from PCA. The x-axis shows the Mean Decrease Gini score, indicating the importance of each component in the model. # Evaluation ## Model Evaluation ### Evaluation of fit of models with holdout method For SVM model

```r
# Evaluate SVM model for norm_test_subset
set.seed(3434)
svm_predict_norm_test <- predict(svm_model_norm, norm_test_subset)
confusionMatrix(svm_predict_norm_test, norm_test_subset$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction BRCA COAD KIRC LUAD PRAD
##       BRCA   60    0    0    0    0
##       COAD    0   15    0    0    0
##       KIRC    0    0   29    0    0
##       LUAD    0    0    0   28    0
##       PRAD    0    0    0    0   27
##
## Overall Statistics
##
##                  Accuracy : 1
##                    95% CI : (0.9771, 1)
##       No Information Rate : 0.3774
##       P-Value [Acc > NIR] : < 2.2e-16
```

```
##
##                    Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: BRCA Class: COAD Class: KIRC Class: LUAD
## Sensitivity              1.0000     1.00000      1.0000      1.0000
## Specificity              1.0000     1.00000      1.0000      1.0000
## Pos Pred Value           1.0000     1.00000      1.0000      1.0000
## Neg Pred Value           1.0000     1.00000      1.0000      1.0000
## Prevalence               0.3774     0.09434      0.1824      0.1761
## Detection Rate           0.3774     0.09434      0.1824      0.1761
## Detection Prevalence     0.3774     0.09434      0.1824      0.1761
## Balanced Accuracy        1.0000     1.00000      1.0000      1.0000
##                      Class: PRAD
## Sensitivity              1.0000
## Specificity              1.0000
## Pos Pred Value           1.0000
## Neg Pred Value           1.0000
## Prevalence               0.1698
## Detection Rate           0.1698
## Detection Prevalence     0.1698
## Balanced Accuracy        1.0000
```

```r
# Evaluate SVM model for pca_test_subset
set.seed(3434)
svm_predict_pca_test <- predict(svm_model_pca, pca_test_subset)
confusionMatrix(svm_predict_pca_test, pca_test_subset$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction BRCA COAD KIRC LUAD PRAD
##       BRCA   60    0    0    1    0
##       COAD    0   15    0    0    0
##       KIRC    0    0   29    0    0
##       LUAD    0    0    0   27    0
##       PRAD    0    0    0    0   27
##
## Overall Statistics
##
##                Accuracy : 0.9937
##                  95% CI : (0.9655, 0.9998)
##     No Information Rate : 0.3774
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9917
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
```

```
##                      Class: BRCA Class: COAD Class: KIRC Class: LUAD
## Sensitivity                 1.0000     1.00000      1.0000      0.9643
## Specificity                 0.9899     1.00000      1.0000      1.0000
## Pos Pred Value              0.9836     1.00000      1.0000      1.0000
## Neg Pred Value              1.0000     1.00000      1.0000      0.9924
## Prevalence                  0.3774     0.09434      0.1824      0.1761
## Detection Rate              0.3774     0.09434      0.1824      0.1698
## Detection Prevalence        0.3836     0.09434      0.1824      0.1698
## Balanced Accuracy           0.9949     1.00000      1.0000      0.9821
##                      Class: PRAD
## Sensitivity                 1.0000
## Specificity                 1.0000
## Pos Pred Value              1.0000
## Neg Pred Value              1.0000
## Prevalence                  0.1698
## Detection Rate              0.1698
## Detection Prevalence        0.1698
## Balanced Accuracy           1.0000
```

For Gradient Boosting model

```r
# Evaluate Gradient Boosting model for norm_test_subset
set.seed(3434)
gb_predict_norm_test <- predict(gbm_model_norm, norm_test_subset)
confusionMatrix(gb_predict_norm_test, norm_test_subset$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction BRCA COAD KIRC LUAD PRAD
##       BRCA   60    0    0    1    0
##       COAD    0   15    0    0    0
##       KIRC    0    0   28    0    0
##       LUAD    0    0    1   27    0
##       PRAD    0    0    0    0   27
##
## Overall Statistics
##
##                Accuracy : 0.9874
##                  95% CI : (0.9553, 0.9985)
##     No Information Rate : 0.3774
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9833
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: BRCA Class: COAD Class: KIRC Class: LUAD
## Sensitivity                 1.0000     1.00000      0.9655      0.9643
## Specificity                 0.9899     1.00000      1.0000      0.9924
## Pos Pred Value              0.9836     1.00000      1.0000      0.9643
```

```
## Neg Pred Value              1.0000    1.00000    0.9924    0.9924
## Prevalence                  0.3774    0.09434    0.1824    0.1761
## Detection Rate              0.3774    0.09434    0.1761    0.1698
## Detection Prevalence        0.3836    0.09434    0.1761    0.1761
## Balanced Accuracy           0.9949    1.00000    0.9828    0.9783
##                         Class: PRAD
## Sensitivity                 1.0000
## Specificity                 1.0000
## Pos Pred Value              1.0000
## Neg Pred Value              1.0000
## Prevalence                  0.1698
## Detection Rate              0.1698
## Detection Prevalence        0.1698
## Balanced Accuracy           1.0000
```

```r
# Evaluate Gradient Boosting model for pca_test_subset
set.seed(3434)
gb_predict_pca_test <- predict(gbm_model_pca, pca_test_subset)
confusionMatrix(gb_predict_pca_test, pca_test_subset$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction BRCA COAD KIRC LUAD PRAD
##       BRCA   59    0    0    0    1
##       COAD    0   15    0    0    0
##       KIRC    0    0   29    0    0
##       LUAD    1    0    0   28    0
##       PRAD    0    0    0    0   26
##
## Overall Statistics
##
##                Accuracy : 0.9874
##                  95% CI : (0.9553, 0.9985)
##     No Information Rate : 0.3774
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9834
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: BRCA Class: COAD Class: KIRC Class: LUAD
## Sensitivity               0.9833     1.00000      1.0000      1.0000
## Specificity               0.9899     1.00000      1.0000      0.9924
## Pos Pred Value            0.9833     1.00000      1.0000      0.9655
## Neg Pred Value            0.9899     1.00000      1.0000      1.0000
## Prevalence                0.3774     0.09434      0.1824      0.1761
## Detection Rate            0.3711     0.09434      0.1824      0.1761
## Detection Prevalence      0.3774     0.09434      0.1824      0.1824
## Balanced Accuracy         0.9866     1.00000      1.0000      0.9962
##                      Class: PRAD
## Sensitivity               0.9630
```

```
## Specificity              1.0000
## Pos Pred Value            1.0000
## Neg Pred Value            0.9925
## Prevalence               0.1698
## Detection Rate           0.1635
## Detection Prevalence     0.1635
## Balanced Accuracy        0.9815
```

For Random Forest model

```
# Evaluate Random Forest model for norm_test_subset
set.seed(3434)
rf_predict_norm_test <- predict(rf_model_norm, norm_test_subset)
confusionMatrix(rf_predict_norm_test, norm_test_subset$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction BRCA COAD KIRC LUAD PRAD
##       BRCA   60    0    0    1    0
##       COAD    0   15    0    0    0
##       KIRC    0    0   29    0    0
##       LUAD    0    0    0   27    0
##       PRAD    0    0    0    0   27
##
## Overall Statistics
##
##                Accuracy : 0.9937
##                  95% CI : (0.9655, 0.9998)
##     No Information Rate : 0.3774
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9917
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: BRCA Class: COAD Class: KIRC Class: LUAD
## Sensitivity               1.0000     1.00000      1.0000      0.9643
## Specificity               0.9899     1.00000      1.0000      1.0000
## Pos Pred Value            0.9836     1.00000      1.0000      1.0000
## Neg Pred Value            1.0000     1.00000      1.0000      0.9924
## Prevalence                0.3774     0.09434      0.1824      0.1761
## Detection Rate            0.3774     0.09434      0.1824      0.1698
## Detection Prevalence      0.3836     0.09434      0.1824      0.1698
## Balanced Accuracy         0.9949     1.00000      1.0000      0.9821
##                      Class: PRAD
## Sensitivity               1.0000
## Specificity               1.0000
## Pos Pred Value            1.0000
## Neg Pred Value            1.0000
## Prevalence                0.1698
```

```
## Detection Rate            0.1698
## Detection Prevalence      0.1698
## Balanced Accuracy         1.0000
```

```r
# Evaluate Random Forest model for pca_test_subset
set.seed(3434)
rf_predict_pca_test <- predict(rf_model_pca, pca_test_subset)
confusionMatrix(rf_predict_pca_test, pca_test_subset$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction BRCA COAD KIRC LUAD PRAD
##       BRCA   60    1    0    1    1
##       COAD    0   14    0    0    0
##       KIRC    0    0   29    0    0
##       LUAD    0    0    0   27    0
##       PRAD    0    0    0    0   26
##
## Overall Statistics
##
##                Accuracy : 0.9811
##                  95% CI : (0.9459, 0.9961)
##     No Information Rate : 0.3774
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9749
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: BRCA Class: COAD Class: KIRC Class: LUAD
## Sensitivity               1.0000     0.93333      1.0000      0.9643
## Specificity               0.9697     1.00000      1.0000      1.0000
## Pos Pred Value            0.9524     1.00000      1.0000      1.0000
## Neg Pred Value            1.0000     0.99310      1.0000      0.9924
## Prevalence                0.3774     0.09434      0.1824      0.1761
## Detection Rate            0.3774     0.08805      0.1824      0.1698
## Detection Prevalence      0.3962     0.08805      0.1824      0.1698
## Balanced Accuracy         0.9848     0.96667      1.0000      0.9821
##                      Class: PRAD
## Sensitivity               0.9630
## Specificity               1.0000
## Pos Pred Value            1.0000
## Neg Pred Value            0.9925
## Prevalence                0.1698
## Detection Rate            0.1635
## Detection Prevalence      0.1635
## Balanced Accuracy         0.9815
```

Evaluation with k-fold cross-validation and tuning of the model

Using cross validation with 5 folds so that the data trains well for all the different classes.

```r
# Creating train control
set.seed(3434)
train_control <- trainControl(method = "cv", number = 5, verboseIter = FALSE, savePredictions = "final")


if (!require("caret", character.only = TRUE)) {
  install.packages("caret", dependencies = TRUE)
}
library(caret)


# SVM grid
svm_grid <- expand.grid(C = c(0.1, 1, 10), sigma = c(0.001, 0.01, 0.05))

# k-fold cross-validation SVM model for normalized_data
set.seed(3434)
cross_svm_model_norm <- train(Class ~ ., data = normalized_data, method = "svmRadial", trControl =
                              train_control, metric = "Accuracy", tuneGrid = svm_grid)
```

```
## maximum number of iterations reached 0.002729239 0.002654863maximum number of iterations reached 0.00
```

```r
# k-fold cross-validation SVM model for pca_dataset
set.seed(3434)
cross_svm_model_pca <- train(Class ~ ., data = pca_dataset, method = "svmRadial", trControl =
                             train_control, metric = "Accuracy", tuneGrid = svm_grid)
```

```
## maximum number of iterations reached 0.007704758 0.007561212maximum number of iterations reached 0.00
```

```r
# Gradient Boosting grid
gbm_grid <- expand.grid(n.trees = c(50, 100), interaction.depth = c(1, 3), shrinkage = c(0.1),  n.minob

# k-fold cross-validation Gradient Boosting model for normalized_data
set.seed(3434)
cross_gbm_model_norm <- train(Class ~ ., data = normalized_data, method = "gbm", trControl =
                                  train_control, metric = "Accuracy", verbose = FALSE, tuneGrid = gbm_grid

# k-fold cross-validation Gradient Boosting model for pca_dataset
set.seed(3434)
cross_gbm_model_pca <- train(Class ~ ., data = pca_dataset, method = "gbm", trControl = train_control,
                      metric = "Accuracy", verbose = FALSE, tuneGrid = gbm_grid)


# Random forest grid
rf_grid <- expand.grid(mtry = c(2, 4, 6))

# k-fold cross-validation Gradient Boosting model for normalized_data
set.seed(3434)
cross_rf_model_norm <- train(Class ~ ., data = normalized_data, method = "rf", trControl = train_control
                              metric = "Accuracy", verbose = FALSE, tuneGrid = rf_grid)
```

```r
# k-fold cross-validation Gradient Boosting model for pca_dataset
set.seed(3434)
cross_rf_model_pca <- train(Class ~ ., data = pca_dataset, method = "rf", trControl = train_control,
                            metric = "Accuracy", verbose = FALSE, tuneGrid = rf_grid)
```

```r
create_ensemble <- function(rf_model, gbm_model, svm_model, test_data) {

  # Predictions for Random Forest
  rf_predictions <- predict(rf_model, test_data)

  # Predictions for Gradient Boosting
  gbm_predictions <- predict(gbm_model, newdata = test_data)

  # Predictions for SVM
  svm_predictions <- predict(svm_model, test_data)

  # Combine predictions into a matrix
  combined_predictions <- cbind(as.character(rf_predictions),
                                as.character(gbm_predictions),
                                as.character(svm_predictions))

  # Perform majority vote with random selection in case of a tie
  ensemble_predictions <- apply(combined_predictions, 1, function(row) {
    freq <- table(row)
    return(names(which.max(freq)))
  })
  return(ensemble_predictions)
  }
```

```r
norm_ensemble <- create_ensemble(cross_rf_model_norm, cross_gbm_model_norm, cross_svm_model_norm, norm_
```

```r
pca_ensemble <- create_ensemble(cross_rf_model_pca, cross_gbm_model_pca, cross_svm_model_pca, pca_test_s
```

```r
norm_ensemble_accuracy <- sum(norm_ensemble == norm_test_subset$Class) / length(norm_test_subset$Class)
print(paste("Ensemble Accuracy:", norm_ensemble_accuracy))
```

```
## [1] "Ensemble Accuracy: 1"
```

```r
pca_ensemble_accuracy <- sum(pca_ensemble == pca_test_subset$Class) / length(pca_test_subset$Class)
print(paste("Ensemble Accuracy:", pca_ensemble_accuracy))
```

```
## [1] "Ensemble Accuracy: 1"
```