Dimensionality Reduction

Lavinia Carabet

Dimensionality Reduction

Techniques that attempt to examine the underlying patterns or relationships for a large number of variables and determine whether the information can be better summarized in a few factors or components

Principal Component Analysis

Statistical method that projects a high-dimensional space into a much lower-dimensional subspace (2D or 3D)

Identifies principal components to reduce dimensionality while maintaining the inherent structure of the data

Principal components are uncorrelated linear combinations of the original variables with variances as large as possible, with each successive component explaining less and less variability

The first principal component can be defined as a direction that maximizes the variance of the projected data The i-th principal component can be taken as a direction orthogonal to the first i-1 principal components that maximizes the variance of the projected data

Principal components are eigenvectors of the data's covariance matrix often computed by eigen decomposition of the data covariance matrix or singular value decomposition of the data matrix

Eigenvector - of a linear transformation - is a non-zero vector that changes at most by a scalar factor when that linear transformation is applied to it

Eigenvalue - corresponding to the eigenvector - is the factor by which the eigenvector is scaled

Geometrically, an eigenvector, corresponding to a real non-zero eigenvalue, points in a direction in which it is stretched by the transformation and the eigenvalue is the factor by which it is stretched

Preparation

Load the GEO GSE2990 Sotiriou Breast Cancer data - Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2990

Dataset - microarray experiments (gene expression data) from primary breast tumors of tamoxifen-untreated patients

Load also the annotation file for this dataframe

```r
dat <- read.table('./sotiriou.txt', header=T, row.names=1)
dim(dat)
```

```
## [1] 239 125
```

```r
#dat
dat[1:5,]
```

```
##              GSM65752  GSM65753 GSM65754 GSM65755 GSM65756  GSM65757  GSM65758
## 160020_at    7.127042  9.100651 8.203144 5.870628 6.812588  8.085146  8.293816
## 200616_s_at  9.900457  9.472446 9.991549 9.560798 9.453957 10.345307 10.947466
## 200702_s_at  9.927486  9.071813 9.621632 9.866459 9.845791  9.104895  9.434893
## 200769_s_at  5.492201  5.106272 5.189343 5.374762 5.210284  5.335828  5.329314
## 200998_s_at 11.135597 10.466324 9.779081 9.892760 8.966400 11.219400 10.655453
##             GSM65760  GSM65761  GSM65762  GSM65763  GSM65764  GSM65765
## 160020_at   5.909595  5.681155  9.480939  8.371547  9.516220  8.245924
## 200616_s_at 9.082748 10.787522  9.617904 10.421155  9.408520 10.121032
## 200702_s_at 9.994836  9.433560  9.426032  9.747579  8.554534  9.846077
## 200769_s_at 5.398720  2.500648  5.801551  5.338819  5.625358  2.950183
## 200998_s_at 9.392855  9.097819 10.404082 10.200129 10.817661  9.001823
##              GSM65766  GSM65767  GSM65768  GSM65769 GSM65770  GSM65771
## 160020_at    8.644868  8.120205  8.231659  8.033868 7.711178  9.022717
## 200616_s_at 10.547777 10.994460 10.493583  8.814009 8.574927 10.678708
## 200702_s_at  9.690883  9.640134  8.903683  9.698340 8.297697  9.932706
## 200769_s_at  3.664175  5.459398  5.975150  6.332186 6.192912  5.317168
## 200998_s_at 10.133112  9.691442 10.825326 10.061569 9.850459 10.047732
##              GSM65772 GSM65773  GSM65774  GSM65775  GSM65776 GSM65779  GSM65780
## 160020_at    7.119111 6.528758  6.857606  7.779015  8.830330 8.510281  8.256988
## 200616_s_at  9.810289 8.994529 10.454922  8.599527  8.584614 9.300683 11.086693
## 200702_s_at  9.967046 7.984709  8.739218  8.723353  7.463278 9.105808  9.529709
## 200769_s_at  5.374693 6.082383  6.595516  6.135871  5.778541 5.107219  5.345398
## 200998_s_at 10.732827 9.832214  9.427659 10.415562 10.422325 9.943425  9.546482
##             GSM65781  GSM65782 GSM65783  GSM65784  GSM65785 GSM65786 GSM65787
## 160020_at   6.801547  7.880265 7.917776  6.337465  9.020984 6.825657 7.818748
## 200616_s_at 8.018304 11.168007 9.660540 10.144745 10.917555 9.936215 9.781399
## 200702_s_at 8.501395  9.632202 9.617484  8.902115  9.996999 9.042858 8.971623
## 200769_s_at 6.134436  5.790030 5.341360  7.059119  6.571793 7.277563 6.644540
## 200998_s_at 8.824451  9.848869 8.630958  9.356876 10.230284 9.690731 9.496230
##              GSM65788 GSM65789  GSM65790 GSM65791  GSM65792  GSM65793  GSM65794
## 160020_at    8.068011 7.407714  8.988034 8.230169  7.635828  7.890231  5.979435
## 200616_s_at 10.789638 9.668978  8.181882 9.317535  9.719525 10.677062 10.252366
## 200702_s_at  9.916601 9.245667  8.798942 9.233922  8.259998  9.856308  9.371726
## 200769_s_at  4.795179 5.416566  5.602389 5.395962  6.238359  5.178281  4.852333
## 200998_s_at  9.753282 9.682318 10.066601 9.068825 10.020634  9.373232 10.301871
##              GSM65795  GSM65796  GSM65797  GSM65798  GSM65799  GSM65800
## 160020_at    7.984038  8.309032  9.380480  7.487749  7.529136  9.754835
## 200616_s_at 10.252106 10.565160 11.255753 11.165969 10.590101 10.080474
## 200702_s_at  8.526333 10.186376  9.419193  9.096370 10.133660  9.296562
## 200769_s_at  5.407855  5.252883  5.391743  6.109436  5.374098  5.873256
## 200998_s_at 10.446065  9.992373 10.694382  9.240809  9.797251 10.853646
##              GSM65801  GSM65802  GSM65803  GSM65804  GSM65805  GSM65806
## 160020_at    7.902266  6.115298  8.722334  7.736976  8.615979  6.387784
## 200616_s_at 10.853234  9.836458 11.210163 10.795851 10.578003  9.520421
## 200702_s_at  8.601716 11.137974 10.243359  9.488605  9.297246  8.246602
## 200769_s_at  6.286121  5.376139  5.154376  5.000554  5.358961  4.965700
## 200998_s_at 10.483693 10.078380 10.968063  9.949478 10.508184 10.341115
##             GSM65807 GSM65808  GSM65810  GSM65811 GSM65812  GSM65813  GSM65814
## 160020_at   7.752178 7.788342  8.051734  8.316980 7.246637  7.198145  8.467210
```

```
## 200616_s_at 8.633368 9.690017 10.462164 10.083944 8.280060 10.479837 10.339253
## 200702_s_at 9.093935 9.024625  9.267070  9.454722 8.480911 10.203286  8.726727
## 200769_s_at 5.875508 4.664785  5.312084  5.651056 5.238019  5.303193  5.871917
## 200998_s_at 9.264657 9.653221  9.314803  9.239377 9.638025  9.525043 10.353718
##              GSM65815  GSM65816 GSM65817  GSM65818  GSM65819 GSM65820 GSM65821
## 160020_at    7.177986  7.082744 6.882445  7.768723  6.167115 5.330163 5.583011
## 200616_s_at 10.714308 10.219222 8.791933  9.628977  9.210866 8.112195 8.118626
## 200702_s_at 10.188286  9.511245 8.934046  9.204855  9.602535 5.997627 6.653063
## 200769_s_at  4.961311  5.412523 6.277328  6.816162  5.392524 7.546326 6.425582
## 200998_s_at  9.378871 10.031773 9.336170 10.511327 10.356181 8.452654 8.062702
##             GSM65822 GSM65823 GSM65824 GSM65825 GSM65826 GSM65827 GSM65828
## 160020_at   5.294315 5.559391 5.446690 4.225444 6.009053 5.827058 5.349911
## 200616_s_at 6.738064 7.517329 7.567199 7.905552 7.538735 8.956796 8.172378
## 200702_s_at 7.672364 7.299689 7.349345 6.713698 5.455443 7.671101 7.368687
## 200769_s_at 9.696201 8.084531 7.120451 8.367263 6.354643 6.680537 6.183167
## 200998_s_at 6.733758 7.941445 8.403255 8.344021 7.928199 9.245684 8.823509
##             GSM65829 GSM65830 GSM65831 GSM65832 GSM65833 GSM65834 GSM65835
## 160020_at   4.858067 5.441667 5.263510 6.513543 5.312928 4.806524 5.264719
## 200616_s_at 6.969864 8.138992 9.420637 9.279276 7.304441 9.532450 7.673764
## 200702_s_at 6.191760 6.479657 5.780219 7.386742 7.147238 6.637015 6.842989
## 200769_s_at 5.900277 6.897877 5.836441 6.079820 7.561734 8.535315 5.973704
## 200998_s_at 8.729199 8.430658 9.094698 9.477792 7.959189 9.924686 9.568696
##             GSM65836 GSM65837 GSM65838 GSM65839 GSM65840 GSM65841 GSM65842
## 160020_at   6.452673 3.553217 5.308978 6.565343 6.688451 6.100403 6.131141
## 200616_s_at 6.799957 8.505827 7.906421 9.233981 9.922303 7.495615 7.817343
## 200702_s_at 6.064194 7.760058 7.233753 6.598780 8.639239 6.217803 7.393500
## 200769_s_at 7.869046 4.488966 8.189155 8.501925 5.838808 8.034822 8.088876
## 200998_s_at 7.774166 7.337992 9.162603 8.917767 9.313191 8.923974 9.322993
##             GSM65843 GSM65844  GSM65845 GSM65846 GSM65847 GSM65848 GSM65849
## 160020_at   6.085709 6.192932  6.225261 6.074410 5.867830 5.945478 5.238206
## 200616_s_at 8.990263 7.835177  9.618565 8.376150 8.702616 9.494268 7.608893
## 200702_s_at 7.269527 6.049418  7.906513 7.745725 6.415752 7.732984 7.236409
## 200769_s_at 7.366368 7.633503  7.443933 8.175065 7.641305 8.196143 5.752735
## 200998_s_at 9.091870 8.323871 10.222800 9.445476 9.117662 9.152535 8.141216
##             GSM65850 GSM65851 GSM65852 GSM65853 GSM65854 GSM65855 GSM65856
## 160020_at   5.346330 5.320685 4.607676 7.290498 5.897952 5.329988 4.716552
## 200616_s_at 7.758706 8.791167 8.582104 9.113793 7.690340 7.162172 7.440817
## 200702_s_at 7.265456 7.563192 7.054802 7.652907 6.813953 6.613734 7.370872
## 200769_s_at 6.840295 7.241069 7.131030 6.556301 7.752748 7.395102 8.917769
## 200998_s_at 8.365251 7.566129 8.166242 8.964722 7.838155 7.591849 8.602457
##             GSM65857 GSM65858 GSM65859 GSM65860 GSM65861 GSM65862 GSM65863
## 160020_at   6.228811 6.353189 6.319914 5.388501 6.803573 5.299090 5.403714
## 200616_s_at 8.139498 7.684448 7.838530 8.627343 7.661839 6.776158 8.299346
## 200702_s_at 7.664270 7.027039 5.985502 7.664480 7.183214 6.880971 7.023203
## 200769_s_at 8.150211 7.780264 8.299409 7.308597 7.794505 8.887082 7.843488
## 200998_s_at 9.505822 7.517558 8.455796 8.612803 8.932506 6.871481 7.989526
##             GSM65864 GSM65865 GSM65866 GSM65867 GSM65868 GSM65869 GSM65870
## 160020_at   5.705747 5.798020 5.112201 5.832906 5.293796 5.366202 5.479538
## 200616_s_at 7.191298 7.151177 8.257703 8.620128 8.232577 7.515288 7.064911
## 200702_s_at 7.635920 8.101510 7.722635 7.719481 7.600326 6.600945 6.979100
## 200769_s_at 5.346011 6.091211 7.676671 7.151661 7.458228 8.466384 8.840666
## 200998_s_at 8.151714 7.210733 9.757229 8.159228 7.850343 8.253440 8.822553
##             GSM65871 GSM65872 GSM65873 GSM65874 GSM65875 GSM65876 GSM65877
## 160020_at   5.302505 5.610217 5.334782 5.453561 3.772388 5.303982 5.700977
```

```
## 200616_s_at 4.983138 8.013414 6.621606 6.465911 6.808675 8.631659 6.425143
## 200702_s_at 6.690384 6.363681 7.123621 7.659471 7.556384 7.646633 6.850617
## 200769_s_at 5.418327 5.770689 5.859436 5.268106 6.627903 6.539166 5.232111
## 200998_s_at 6.575140 8.253160 7.767105 7.266600 6.125215 8.296543 8.194467
##             GSM65878 GSM65879 GSM65880
## 160020_at   6.151776 5.534651 5.311369
## 200616_s_at 6.915546 6.345979 9.133848
## 200702_s_at 7.660817 6.176146 6.391650
## 200769_s_at 8.577475 6.297069 7.325461
## 200998_s_at 8.986590 8.753346 8.562461
```

```
ann <- read.table('./sotiriouAnn.txt', header=T, row.names=1)
dim(ann)
```

```
## [1] 125  13
```

```
ann
```

```
##            site sample_name treatment dataset grade node size age er event.rfs
## GSM65752    KIU  KIU_101B88      none   KJ125     3    0  1.2  40  0         0
## GSM65753    KIU  KIU_105B13      none   KJ125     1    0  1.3  46  1         0
## GSM65754    KIU  KIU_106B55      none   KJ125     1    0  6.0  37  1         1
## GSM65755    KIU  KIU_111B51      none   KJ125     3    0  3.3  41  1         1
## GSM65756    KIU  KIU_113B11      none   KJ125     3    0  3.2  38  1         1
## GSM65757    KIU  KIU_120B73      none   KJ125     2    0  1.6  34  1         0
## GSM65758    KIU  KIU_124B25      none   KJ125     2    0  2.1  46  1         1
## GSM65760    KIU  KIU_127B00      none   KJ125     3    0  2.2  57  1         1
## GSM65761    KIU  KIU_134B33      none   KJ125     2    0  2.8  63  1         1
## GSM65762    KIU  KIU_136B04      none   KJ125     2    0  1.7  54  1         1
## GSM65763    KIU  KIU_140B91      none   KJ125     2    0  1.2  61  1         0
## GSM65764    KIU  KIU_144B49      none   KJ125     2    0  2.1  40  1         0
## GSM65765    KIU  KIU_151B84      none   KJ125     2    0  1.5  57  1         0
## GSM65766    KIU  KIU_155B52      none   KJ125     1    0  1.3  57  0         0
## GSM65767    KIU  KIU_163B27      none   KJ125     1    0  0.8  49  1         0
## GSM65768    KIU  KIU_164B81      none   KJ125     2    0  2.3  62  0         0
## GSM65769    KIU  KIU_172B19      none   KJ125     3    0  2.3  42  1         1
## GSM65770    KIU  KIU_177B67      none   KJ125     1    0  1.8  41  1         1
## GSM65771    KIU  KIU_184B38      none   KJ125     1    0  1.0  63  1         0
## GSM65772    KIU  KIU_188B13      none   KJ125     2    0  1.4  60  0         0
## GSM65773    KIU  KIU_196B81      none   KJ125     1    0  1.4  65  1         0
## GSM65774    KIU  KIU_197B95      none   KJ125     2    0  1.6  44  1         0
## GSM65775    KIU  KIU_199B55      none   KJ125     2    0  2.3  54  1         0
## GSM65776    KIU  KIU_205B99      none   KJ125     1    0  2.2  59  1         1
## GSM65779    KIU  KIU_220C70      none   KJ125     1    0  2.0  42  1         0
## GSM65780    KIU  KIU_227C50      none   KJ125     1    0  1.2  57  1         1
## GSM65781    KIU  KIU_229C44      none   KJ125     1    0  1.3  52  1         0
## GSM65782    KIU  KIU_231C80      none   KJ125     1    0  2.2  56  1         1
## GSM65783    KIU  KIU_233C91      none   KJ125     1    0  1.1  49  1         0
## GSM65784    KIU  KIU_242C21      none   KJ125     2    0  1.6  64  1         1
## GSM65785    KIU  KIU_243C70      none   KJ125     1    0  1.8  50  1         0
## GSM65786    KIU  KIU_247C76      none   KJ125     2    0  1.0  56  1         0
## GSM65787    KIU  KIU_248C91      none   KJ125     1    0  2.5  57  1         0
## GSM65788    KIU   KIU_24C30      none   KJ125     2    0  2.3  55  1         0
```

```
## GSM65789  KIU  KIU_259C74      none   KJ125     1     0  1.0  65   1           0
## GSM65790  KIU  KIU_260C91      none   KJ125     2     0  2.1  58   1           0
## GSM65791  KIU  KIU_266C51      none   KJ125     1     0  2.2  58   1           0
## GSM65792  KIU  KIU_268C87      none   KJ125     2     0  1.5  32   1           0
## GSM65793  KIU  KIU_272C88      none   KJ125     2     0  1.7  45   1           0
## GSM65794  KIU  KIU_278C80      none   KJ125     2     0  1.1  56   1           0
## GSM65795  KIU  KIU_279C61      none   KJ125     3     0  1.9  50   1           0
## GSM65796  KIU  KIU_280C43      none   KJ125     2     0  0.9  45   1           1
## GSM65797  KIU  KIU_282C51      none   KJ125     1     0  1.1  55   1           0
## GSM65798  KIU  KIU_284C63      none   KJ125     1     0  1.0  48   1           0
## GSM65799  KIU  KIU_286C91      none   KJ125     2     0  1.8  62   1           0
## GSM65800  KIU   KIU_28C76      none   KJ125     1     0  2.0  56   1           0
## GSM65801  KIU  KIU_292C66      none   KJ125     2     0  2.0  51   1           0
## GSM65802  KIU  KIU_303C36      none   KJ125     3     0  2.3  37   0           0
## GSM65803  KIU  KIU_304C89      none   KJ125     3     0  1.5  54   0           1
## GSM65804  KIU  KIU_308C93      none   KJ125     2     0  2.1  38   0           1
## GSM65805  KIU  KIU_309C49      none   KJ125     1     0  1.2  44   1           0
## GSM65806  KIU  KIU_314B55      none   KJ125     3     0  3.0  38   0           1
## GSM65807  KIU  KIU_316C64      none   KJ125     1     0  1.3  51   1           0
## GSM65808  KIU   KIU_36C17      none   KJ125     2     0  2.2  46   1           0
## GSM65810  KIU   KIU_42C67      none   KJ125     1     0  2.6  59  NA           0
## GSM65811  KIU   KIU_43C47      none   KJ125     2     0  1.2  46   0           0
## GSM65812  KIU   KIU_52A90      none   KJ125     1     0  2.6  53   1           0
## GSM65813  KIU    KIU_5B97      none   KJ125     2     0  2.4  37   1           1
## GSM65814  KIU   KIU_65A68      none   KJ125     1     0  1.8  49   1           0
## GSM65815  KIU   KIU_74A63      none   KJ125     1     0  2.2  56   1           1
## GSM65816  KIU   KIU_86A40      none   KJ125     2     0  2.4  61   0           0
## GSM65817  KIU   KIU_87A79      none   KJ125     2     0  1.2  36   1           0
## GSM65818  KIU   KIU_88A67      none   KJ125     2     0  2.4  63   1           1
## GSM65819  KIU   KIU_89A64      none   KJ125     3     0  2.3  60   1           0
## GSM65820  OXF     OXFU_12      none   KJ125    NA     0  0.0  44   1           0
## GSM65821  OXF     OXFU_16      none   KJ125     2     0  2.6  46   1           0
## GSM65822  OXF     OXFU_37      none   KJ125     2     0  1.8  38   0           1
## GSM65823  OXF     OXFU_53      none   KJ125    NA     0  0.3  61   1           0
## GSM65824  OXF     OXFU_57      none   KJ125     2     0  2.0  43   0           1
## GSM65825  OXF     OXFU_88      none   KJ125     3     0  2.6  65  NA           0
## GSM65826  OXF     OXFU_90      none   KJ125    NA     0  1.4  61   1           1
## GSM65827  OXF     OXFU_93      none   KJ125    NA     0  0.9  58   1           0
## GSM65828  OXF    OXFU_104      none   KJ125    NA     0  3.1  60   1           1
## GSM65829  OXF    OXFU_126      none   KJ125    NA     0  1.0  45   0           1
## GSM65830  OXF    OXFU_127      none   KJ125     1     0  1.9  42   1           1
## GSM65831  OXF    OXFU_138      none   KJ125     3     0  3.0  55   1           0
## GSM65832  OXF    OXFU_145      none   KJ125     2     0  2.5  45   0           0
## GSM65833  OXF    OXFU_157      none   KJ125     3     0  2.0  42   0           1
## GSM65834  OXF    OXFU_181      none   KJ125     2     0  1.5  64   1           1
## GSM65835  OXF    OXFU_217      none   KJ125     2     0  1.0  53   0           1
## GSM65836  OXF    OXFU_220      none   KJ125     2     0  1.0  47   0           0
## GSM65837  OXF    OXFU_223      none   KJ125     3     0  2.1  64   1           1
## GSM65838  OXF    OXFU_245      none   KJ125    NA     0  1.0  54  NA           0
## GSM65839  OXF    OXFU_247      none   KJ125     3     0  4.5  73   0           1
## GSM65840  OXF    OXFU_254      none   KJ125    NA     0  2.0  48   1           1
## GSM65841  OXF    OXFU_281      none   KJ125     2     0  1.6  64   0           1
## GSM65842  OXF    OXFU_316      none   KJ125     3     0  2.2  47   0           0
## GSM65843  OXF    OXFU_320      none   KJ125     3     0  5.0  39   0           1
```

```
## GSM65844  OXF    OXFU_348      none   KJ125      2    0  4.5   65  1          1
## GSM65845  OXF    OXFU_360      none   KJ125      3    0  3.0   32  0          0
## GSM65846  OXF    OXFU_366      none   KJ125      3    0  2.5   57  0          1
## GSM65847  OXF    OXFU_373      none   KJ125     NA    0  1.8   64  1          1
## GSM65848  OXF    OXFU_382      none   KJ125      3    0  3.0   60  1          0
## GSM65849  OXF    OXFU_397      none   KJ125     NA    0  2.0   71  1          1
## GSM65850  OXF    OXFU_419      none   KJ125      3    0  0.7   42  0          0
## GSM65851  OXF    OXFU_449      none   KJ125     NA    0  3.0   57  1          0
## GSM65852  OXF    OXFU_476      none   KJ125     NA    0  2.5   53 NA          1
## GSM65853  OXF    OXFU_484      none   KJ125      2    0  1.3   64  1          0
## GSM65854  OXF    OXFU_491      none   KJ125     NA    0  2.0   66  1          0
## GSM65855  OXF    OXFU_513      none   KJ125      2    0  3.8   47  0          1
## GSM65856  OXF    OXFU_522      none   KJ125     NA    0  2.4   63  1          0
## GSM65857  OXF    OXFU_530      none   KJ125      2    0  1.8   54  1          1
## GSM65858  OXF    OXFU_531      none   KJ125      2    0  1.6   42 NA          0
## GSM65859  OXF    OXFU_533      none   KJ125      3    0  2.6   53  1          0
## GSM65860  OXF    OXFU_535      none   KJ125      3    0  1.5   59  1          0
## GSM65861  OXF    OXFU_543      none   KJ125      2    0  1.9   71  0          1
## GSM65862  OXF    OXFU_544      none   KJ125      2    0  1.8   54  1          0
## GSM65863  OXF    OXFU_547      none   KJ125      3    0  4.0   45  0          0
## GSM65864  OXF    OXFU_549      none   KJ125     NA    0  1.0   64  1          1
## GSM65865  OXF    OXFU_557      none   KJ125      3    0  3.0   43  0          0
## GSM65866  OXF    OXFU_559      none   KJ125      3    0  1.3   68  0          0
## GSM65867  OXF    OXFU_573      none   KJ125      3    0  2.0   63  1          0
## GSM65868  OXF    OXFU_598      none   KJ125      2    0  2.0   69  1          1
## GSM65869  OXF    OXFU_608      none   KJ125      2    0  3.0   62  1          1
## GSM65870  OXF    OXFU_662      none   KJ125      3    0  2.2   43  0          1
## GSM65871  OXF    OXFU_869      none   KJ125      1    0  1.0   48  1          0
## GSM65872  OXF   OXFU_1065      none   KJ125      2    0  2.0   43  0          1
## GSM65873  OXF   OXFU_1183      none   KJ125      1    0  0.8   50  1          0
## GSM65874  OXF   OXFU_1210      none   KJ125      1    0  0.8   43  1          0
## GSM65875  OXF   OXFU_1248      none   KJ125      1    0  2.0   70  1          0
## GSM65876  OXF   OXFU_1286      none   KJ125      1    0  2.0   52 NA          0
## GSM65877  OXF   OXFU_1328      none   KJ125     NA    0  1.3   49  1          0
## GSM65878  OXF   OXFU_1373      none   KJ125      2    0  2.0   38  0          1
## GSM65879  OXF   OXFU_1415      none   KJ125     NA    0  0.9   47  0          0
## GSM65880  OXF   OXFU_1605      none   KJ125      2    0  1.0   39  0          1
##             time.rfs event.dmfs  time.dmfs
## GSM65752  6.2465753          0  6.2465753
## GSM65753  7.3287671          0  7.3287671
## GSM65754  1.1671233          0  1.1671233
## GSM65755  0.4986301          1  0.4986301
## GSM65756  3.0821918          1  3.0821918
## GSM65757 10.8273973          0 10.8273973
## GSM65758  4.9972603          1  4.9972603
## GSM65760  1.9150685          1  1.9150685
## GSM65761  2.0000000          1  2.0000000
## GSM65762  2.4164384          0  2.4164384
## GSM65763  7.7452055          0  7.7452055
## GSM65764  5.4958904          0  5.4958904
## GSM65765  6.9123288          0  6.9123288
## GSM65766  9.9945205          0  9.9945205
## GSM65767  6.1643836          0  6.1643836
## GSM65768  9.8273973          0  9.8273973
```

```
## GSM65769  8.5780822          0  8.5780822
## GSM65770  6.8301370          1  6.8301370
## GSM65771  8.6630137          0  8.6630137
## GSM65772  9.5780822          0  9.5780822
## GSM65773  5.9123288          0  5.9123288
## GSM65774  5.1643836          0  5.1643836
## GSM65775 10.0767123          0 10.0767123
## GSM65776  4.4136986          1  4.4136986
## GSM65779  7.6630137          0  7.6630137
## GSM65780  9.0794521          0  9.0794521
## GSM65781  9.4958904          0  9.4958904
## GSM65782  6.4136986          1  6.4136986
## GSM65783  9.1616438          0  9.1616438
## GSM65784  2.1643836          0  2.1643836
## GSM65785  5.9972603          0  5.9972603
## GSM65786  4.1643836          0  4.1643836
## GSM65787  2.9150685          0  2.9150685
## GSM65788  5.9972603          0  5.9972603
## GSM65789  6.7452055          0  6.7452055
## GSM65790  4.4986301          0  4.4986301
## GSM65791  8.8273973          0  8.8273973
## GSM65792  8.9123288          0  8.9123288
## GSM65793  3.6657534          0  3.6657534
## GSM65794  8.6630137          0  8.6630137
## GSM65795  8.7452055          0  8.7452055
## GSM65796  1.0000000          0  1.0000000
## GSM65797  2.3315068          0  2.3315068
## GSM65798  9.4109589          0  9.4109589
## GSM65799  7.3287671          0  7.3287671
## GSM65800  6.2465753          0  6.2465753
## GSM65801  8.9945205          0  8.9945205
## GSM65802  6.7452055          0  6.7452055
## GSM65803  2.5808219          1  2.5808219
## GSM65804  2.2493151          1  2.2493151
## GSM65805  8.4109589          0  8.4109589
## GSM65806  0.1671233          0  0.1671233
## GSM65807  9.2438356          0  9.2438356
## GSM65808  8.2465753          0  8.2465753
## GSM65810  8.8273973          0  8.8273973
## GSM65811  0.5835616          0  0.5835616
## GSM65812 11.9095890          0 11.9095890
## GSM65813  0.7506849          0  0.7506849
## GSM65814  3.4986301          0  3.4986301
## GSM65815  5.9123288          1  5.9123288
## GSM65816  9.9945205          0  9.9945205
## GSM65817 10.1616438          0 10.1616438
## GSM65818  4.2465753          0  4.2465753
## GSM65819 11.4109589          0 11.4109589
## GSM65820 14.5342466          0 14.5342466
## GSM65821 11.1808219          0 11.1808219
## GSM65822  8.4767123          0  8.4767123
## GSM65823 14.1863014          0 14.1863014
## GSM65824 12.0493151          1 12.0493151
## GSM65825 13.9616438          0 13.9616438
```

```
## GSM65826  5.8219178          1  5.8219178
## GSM65827 13.7780822          0 13.7780822
## GSM65828  1.7753425          1  1.7753425
## GSM65829 11.4054794          0 11.4054794
## GSM65830 13.3397260          0 13.3397260
## GSM65831 13.7780822          0 13.7780822
## GSM65832 12.9150685          0 12.9150685
## GSM65833 13.4410959          0 13.4410959
## GSM65834 12.5452055          1 12.5452055
## GSM65835  1.5397260          1  1.5397260
## GSM65836 12.6164384          0 12.6164384
## GSM65837  5.1369863          1  5.1369863
## GSM65838 13.3123288          0 13.3123288
## GSM65839  0.6904110          0  1.5397260
## GSM65840 12.6410959          0 12.6410959
## GSM65841 10.5589041          0 10.5589041
## GSM65842 12.8301370          0 12.8301370
## GSM65843  7.0191781          0  7.0191781
## GSM65844  0.6054795          1  0.6054795
## GSM65845 12.5972603          0 12.5972603
## GSM65846  2.8438356          1  3.0465753
## GSM65847  2.9178082          1  2.9178082
## GSM65848 12.7890411          0 12.7890411
## GSM65849  2.8931507          1  2.8931507
## GSM65850 10.4520548          0 12.4219178
## GSM65851 12.4657534          0 12.4657534
## GSM65852  3.4712329          1  3.4712329
## GSM65853 10.7178082          0 10.7178082
## GSM65854 12.2657534          0 12.2657534
## GSM65855  3.1123288          0  3.1123288
## GSM65856  9.8164384          0  9.8164384
## GSM65857  2.7287671          1  4.7424658
## GSM65858 12.0410959          0 12.0410959
## GSM65859 12.3232877          0 12.3232877
## GSM65860 12.2547945          0 12.2547945
## GSM65861  2.6438356          1  2.6438356
## GSM65862 11.8602740          0 11.8602740
## GSM65863  9.2602740          0 11.9287671
## GSM65864 10.0438356          0 10.0438356
## GSM65865 12.0602740          0 12.0602740
## GSM65866 11.1726027          0 11.1726027
## GSM65867 11.5561644          0 11.5561644
## GSM65868  4.0849315          1  3.0082192
## GSM65869  2.6328767          1  2.6328767
## GSM65870  5.7068493          0 11.4602740
## GSM65871  8.7369863          0  8.7369863
## GSM65872  1.9972603          1  1.9972603
## GSM65873  4.3589041          0  4.3589041
## GSM65874  5.2602740          0  5.2602740
## GSM65875  8.9369863          0  8.9369863
## GSM65876  5.2410959          0  5.2410959
## GSM65877  7.2547945          0  7.2547945
## GSM65878  0.7342466          1  0.7342466
## GSM65879  3.5342466          0  3.5342466
```

```
## GSM65880  6.2931507          0  7.6958904
```

Conduct Principal Component Analysis (PCA) and plot PCA results

`prcomp` performs a PCA by a singular value decomposition of the given (centered and possibly scaled) data matrix and returns a list with class `prcomp` containing the following components:

`sdev` the standard deviations of the principal components (i.e., the square roots of the eigenvalues of the covariance/correlation matrix, though the calculation is actually done with the singular values of the data matrix)

`rotation` the matrix of variable loadings (i.e., a matrix whose columns contain the eigenvectors); (the coordinates of the variables -genes- in the projected principal components' space)

`X` the value of the rotated data (the centered (and scaled if requested) data multiplied by the rotation matrix); (the coordinates of the observations -samples- in the projected principal components' space

`center`,`scale` the centering and scaling used, or FALSE

```
dat.pca <- prcomp(t(dat))
# unclass(dat.pca)

dat.loadings <- dat.pca$x[,1:2]      #dim(dat.loadings) [1] 125   2
dat.loadings
```

```
##                   PC1          PC2
## GSM65752 -14.92456344 -14.67002303
## GSM65753 -17.99926227   5.01873475
## GSM65754 -21.97354514   2.42028906
## GSM65755 -10.79713986  -9.38130114
## GSM65756 -14.26983862  -5.74194199
## GSM65757 -17.29355494  -2.04883659
## GSM65758 -26.56613784   3.44526064
## GSM65760  -9.21772822 -11.77618373
## GSM65761 -12.51569635 -12.98409816
## GSM65762 -17.47428838   4.96624609
## GSM65763 -21.85924545   8.26062150
## GSM65764 -15.38425743  11.68214464
## GSM65765 -22.66576725   0.41360007
## GSM65766 -31.48160123   5.11562410
## GSM65767 -16.21433595   2.22536762
## GSM65768  -9.45464513  -7.18468035
## GSM65769 -11.58828600  -4.80960579
## GSM65770 -12.20507465  12.79583223
## GSM65771 -27.69494347   3.70043820
## GSM65772 -14.55425062  -5.02265929
## GSM65773  -5.40054185   5.69097189
## GSM65774  -5.03615728  -5.41812221
## GSM65775 -12.31335216   4.33201319
## GSM65776  -8.77553098  15.05716233
## GSM65779 -19.50729077   6.80129585
## GSM65780 -21.77700786   4.80715349
## GSM65781  -0.01069766   1.31370499
## GSM65782 -19.96577691  -0.77671354
## GSM65783 -25.11545179   3.44101158
```

```
## GSM65784 -10.96091939  -8.73117495
## GSM65785 -20.63259971   5.52791805
## GSM65786  -4.29565962  -6.67219848
## GSM65787 -15.07279600   6.87366553
## GSM65788 -29.82268762  -0.03909213
## GSM65789 -21.89897379   5.74635714
## GSM65790 -14.80896133  10.72691999
## GSM65791 -18.61568268   8.91547616
## GSM65792  -6.86098554 -21.13922323
## GSM65793 -27.31436691  -4.02842422
## GSM65794 -16.33480540 -10.35144687
## GSM65795  -9.90799098 -13.50493489
## GSM65796 -23.07141055   1.79146773
## GSM65797 -22.27313784   0.72940181
## GSM65798 -18.70801260  -2.62774575
## GSM65799 -23.11984635   1.95582074
## GSM65800 -21.96043686   6.31659018
## GSM65801 -15.38463909   0.29130114
## GSM65802 -14.29494391 -23.60889788
## GSM65803 -30.50295653  -0.70055336
## GSM65804 -20.51461424 -15.32063741
## GSM65805 -19.53294963 -12.24287372
## GSM65806  -9.66339271 -15.78209465
## GSM65807 -17.18738377  11.75299208
## GSM65808 -19.81072184   4.69409794
## GSM65810 -18.89960112   2.17736254
## GSM65811 -20.58980251  -3.49349442
## GSM65812 -17.74320325   9.74555188
## GSM65813 -16.08828565  -8.91633835
## GSM65814 -18.84969391   3.76522235
## GSM65815 -21.16041336  -2.73914263
## GSM65816 -12.44757273 -19.90100126
## GSM65817  -7.39666072   4.58860859
## GSM65818  -9.71204477   2.55591870
## GSM65819  -7.43068102 -12.16808389
## GSM65820  26.01033936  -7.39553293
## GSM65821  17.55212213   0.53342840
## GSM65822  19.60074519  -0.02582156
## GSM65823  20.50316908  -0.61660715
## GSM65824  16.43941525   4.15736409
## GSM65825  27.91878694 -10.27511560
## GSM65826  16.60789712   8.80256920
## GSM65827   6.79815116   7.49509897
## GSM65828  19.40277838  -1.10023345
## GSM65829  22.00679539   2.93884612
## GSM65830  17.17696404   5.44050236
## GSM65831  17.36110928  -3.47945020
## GSM65832  15.60362430   2.96287547
## GSM65833  22.51259616  -2.19205561
## GSM65834  20.97595256   2.15829222
## GSM65835  14.11306695  -0.93594897
## GSM65836  16.91298530   6.66563420
## GSM65837  18.31212618   0.17423229
## GSM65838  23.64417815  -7.79432966
```

```
## GSM65839  22.46051893  -4.06322390
## GSM65840   8.26406360   5.11189394
## GSM65841  20.12558156  -8.58953102
## GSM65842  21.31103441 -12.29672605
## GSM65843  19.16516480  -4.96181843
## GSM65844  20.04899954  -3.81862113
## GSM65845  19.53675376  -7.55081487
## GSM65846  14.05133677   1.51419980
## GSM65847  10.88387567   5.99719531
## GSM65848   8.62477485   7.24712472
## GSM65849  11.23017172   5.23467093
## GSM65850  18.33797768  -8.55260828
## GSM65851  24.80714485  -8.12601649
## GSM65852  23.84059351 -17.66702329
## GSM65853   7.72899716  11.28159305
## GSM65854  18.92649366   7.58352217
## GSM65855  18.56765568   4.07973868
## GSM65856  23.95380870  -4.23833833
## GSM65857  18.79430520  -1.04011605
## GSM65858  15.68275648   9.85833578
## GSM65859  13.30679918  14.56060555
## GSM65860  14.63411555   6.63436037
## GSM65861  14.18446365   4.71652853
## GSM65862  18.00774958   7.64068954
## GSM65863  19.49506395  -4.94521072
## GSM65864  13.57363957   6.51800381
## GSM65865  13.27992355   9.56874878
## GSM65866  19.10410101  -7.33473093
## GSM65867  16.53715140   9.59718320
## GSM65868  21.84608096  -4.37235215
## GSM65869  18.48233354   0.42937267
## GSM65870  22.13656982  -0.58727567
## GSM65871  19.47220390   6.52517234
## GSM65872  11.38475447   7.69104965
## GSM65873  12.37195278   7.88811262
## GSM65874  16.83611299   0.46793638
## GSM65875  23.78123403  -7.58771744
## GSM65876  15.51956153   2.92473424
## GSM65877  12.49204486   7.14178923
## GSM65878  15.88521378   2.19498646
## GSM65879  14.94831903   7.91479510
## GSM65880  13.83060289  10.03541283
```

```
levels(as.factor(ann$site))
```

```
## [1] "KIU" "OXF"
```

```
dat.loadings[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]]
```

```
##     GSM65752     GSM65753     GSM65754     GSM65755     GSM65756     GSM65757
## -14.92456344 -17.99926227 -21.97354514 -10.79713986 -14.26983862 -17.29355494
##     GSM65758     GSM65760     GSM65761     GSM65762     GSM65763     GSM65764
## -26.56613784  -9.21772822 -12.51569635 -17.47428838 -21.85924545 -15.38425743
```

```
##     GSM65765     GSM65766     GSM65767     GSM65768     GSM65769     GSM65770
## -22.66576725 -31.48160123 -16.21433595  -9.45464513 -11.58828600 -12.20507465
##     GSM65771     GSM65772     GSM65773     GSM65774     GSM65775     GSM65776
## -27.69494347 -14.55425062  -5.40054185  -5.03615728 -12.31335216  -8.77553098
##     GSM65779     GSM65780     GSM65781     GSM65782     GSM65783     GSM65784
## -19.50729077 -21.77700786  -0.01069766 -19.96577691 -25.11545179 -10.96091939
##     GSM65785     GSM65786     GSM65787     GSM65788     GSM65789     GSM65790
## -20.63259971  -4.29565962 -15.07279600 -29.82268762 -21.89897379 -14.80896133
##     GSM65791     GSM65792     GSM65793     GSM65794     GSM65795     GSM65796
## -18.61568268  -6.86098554 -27.31436691 -16.33480540  -9.90799098 -23.07141055
##     GSM65797     GSM65798     GSM65799     GSM65800     GSM65801     GSM65802
## -22.27313784 -18.70801260 -23.11984635 -21.96043686 -15.38463909 -14.29494391
##     GSM65803     GSM65804     GSM65805     GSM65806     GSM65807     GSM65808
## -30.50295653 -20.51461424 -19.53294963  -9.66339271 -17.18738377 -19.81072184
##     GSM65810     GSM65811     GSM65812     GSM65813     GSM65814     GSM65815
## -18.89960112 -20.58980251 -17.74320325 -16.08828565 -18.84969391 -21.16041336
##     GSM65816     GSM65817     GSM65818     GSM65819
## -12.44757273  -7.39666072  -9.71204477  -7.43068102
```

```r
length(dat.loadings[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]])
```

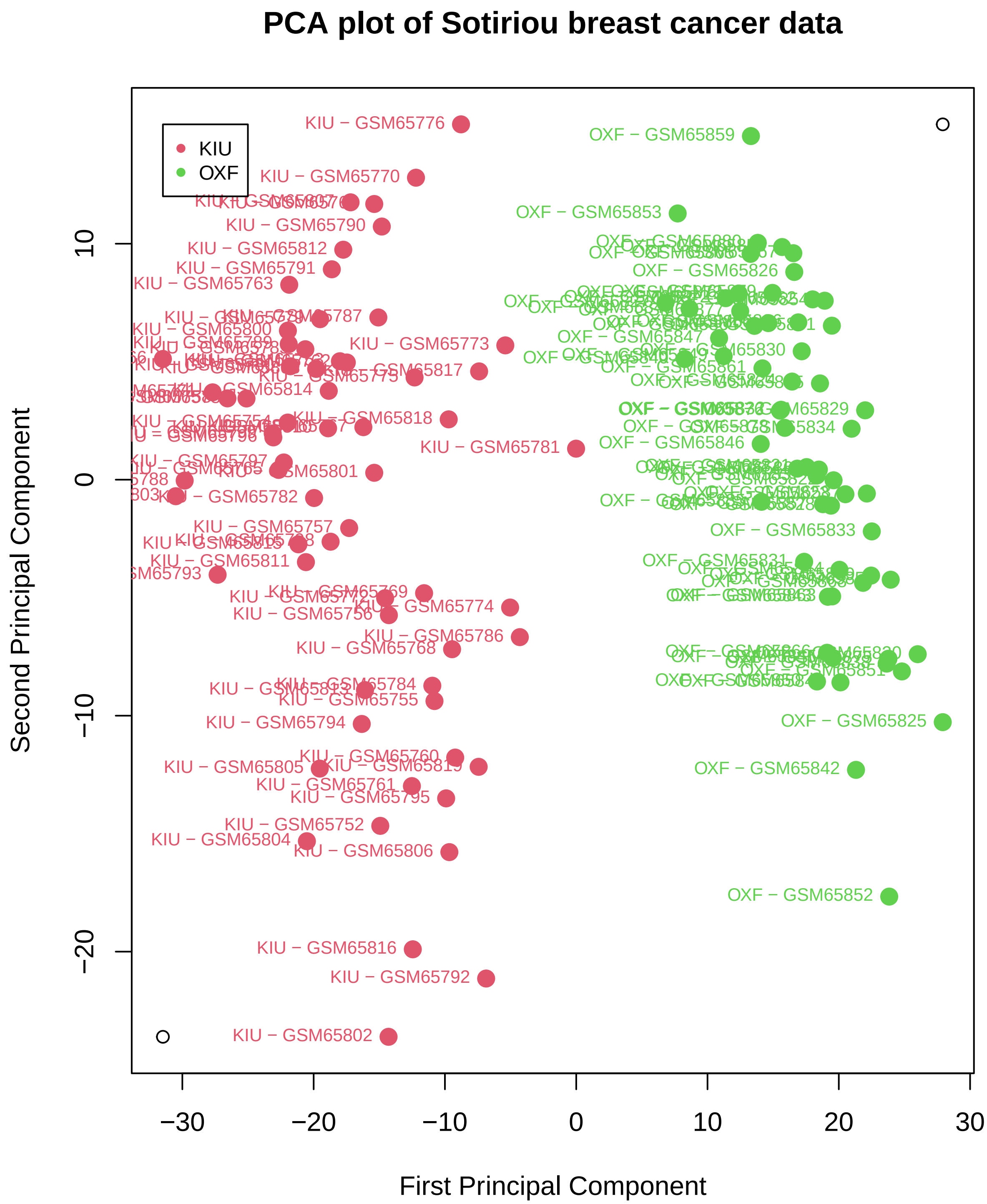
```
## [1] 64
```

```r
length(dat.loadings[,1][as.character(ann$site)==levels(as.factor(ann$site))[2]])
```

```
## [1] 61
```

```r
col <- as.numeric(as.factor(unique(ann$site))) +1

plot(range(dat.loadings[,1]), range(dat.loadings[,2]),
     xlab='First Principal Component',ylab='Second Principal Component',
     main='PCA plot of Sotiriou breast cancer data')

points(dat.loadings[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]],
       dat.loadings[,2][ as.character(ann$site)==levels(as.factor(ann$site))[1]],
       col=col[1],pch=16,cex=1.5)

text(dat.loadings[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]],
     dat.loadings[,2][ as.character(ann$site)==levels(as.factor(ann$site))[1]],
     col=col[1] ,cex=0.7,
     labels= paste(levels(as.factor(ann$site))[1], '-',
                   row.names(ann[as.character(ann$site)==levels(as.factor(ann$site))[1],]), sep= ' '),
     pos=2)

points(dat.loadings[,1][as.character(ann$site)==levels(as.factor(ann$site))[2]],
       dat.loadings[,2][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
       col=col[2],pch=16,cex=1.5)

text(dat.loadings[,1][as.character(ann$site)==levels(as.factor(ann$site))[2]],
     dat.loadings[,2][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
     col=col[2],cex=0.7,
     labels= paste(levels(as.factor(ann$site))[2], '-',
```

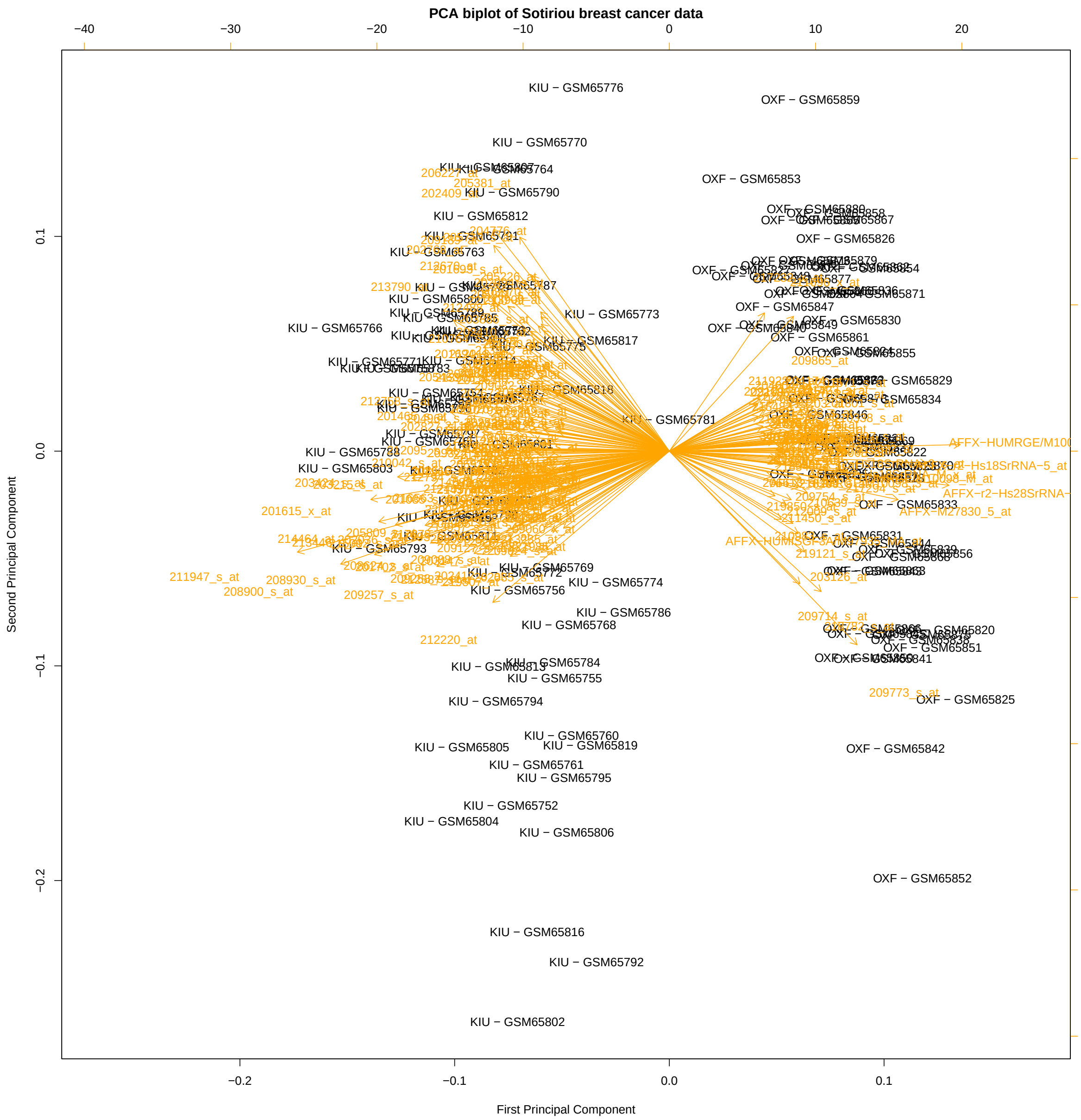
```
                    row.names(ann[as.character(ann$site)==levels(as.factor(ann$site))[2],]), sep= ' '),
     pos=2)

legend(min(range(dat.loadings[,1])), max(range(dat.loadings[,2]) ),
       levels(as.factor(ann$site)),
       col=col,pch=16,cex=.75)
```



Biplot

Visualize both the observations (samples) and the variables (genes) of a data matrix on the same plot

```r
names(dat) <- paste(as.character(ann$site), '-', row.names(ann), sep= ' ')
dat.pca <- prcomp(t(dat))
col <- c("black", "orange")
biplot(dat.pca,scale=TRUE,col=col,
       xlab='First Principal Component', ylab='Second Principal Component',
       main='PCA biplot of Sotiriou breast cancer data')
```



Scree plot corresponding to the PCA above

```r
# standard deviation of the principal components
# (i.e. the square roots of the eigenvalues of the covariance/correlation matrix)
```

```
print("Standard deviation of the principal components")
```

```
## [1] "Standard deviation of the principal components"
```

```
dat.pca$sdev
```

```
##   [1] 1.808960e+01 7.950493e+00 6.759851e+00 5.425392e+00 4.861711e+00
##   [6] 4.261595e+00 3.853985e+00 3.335201e+00 3.058479e+00 2.669527e+00
##  [11] 2.594844e+00 2.470256e+00 2.301769e+00 2.265594e+00 2.218159e+00
##  [16] 2.114738e+00 2.042543e+00 1.968784e+00 1.952570e+00 1.832840e+00
##  [21] 1.784867e+00 1.773668e+00 1.668423e+00 1.628210e+00 1.619220e+00
##  [26] 1.576649e+00 1.539699e+00 1.496120e+00 1.474833e+00 1.456386e+00
##  [31] 1.420215e+00 1.409495e+00 1.396450e+00 1.373183e+00 1.335322e+00
##  [36] 1.315715e+00 1.286264e+00 1.277919e+00 1.214822e+00 1.192632e+00
##  [41] 1.179874e+00 1.165884e+00 1.154381e+00 1.134524e+00 1.108172e+00
##  [46] 1.102001e+00 1.089750e+00 1.065017e+00 1.052027e+00 1.030344e+00
##  [51] 1.009403e+00 9.920168e-01 9.770316e-01 9.577450e-01 9.426928e-01
##  [56] 9.292646e-01 9.279851e-01 9.161050e-01 8.983570e-01 8.877442e-01
##  [61] 8.759392e-01 8.579782e-01 8.469782e-01 8.395788e-01 8.273752e-01
##  [66] 7.952735e-01 7.913936e-01 7.750593e-01 7.664755e-01 7.594623e-01
##  [71] 7.317864e-01 7.242785e-01 7.161824e-01 7.037500e-01 7.006396e-01
##  [76] 6.962052e-01 6.749395e-01 6.679053e-01 6.548958e-01 6.403585e-01
##  [81] 6.327154e-01 6.187611e-01 6.106445e-01 5.974590e-01 5.897933e-01
##  [86] 5.889002e-01 5.759384e-01 5.665320e-01 5.596018e-01 5.489625e-01
##  [91] 5.392246e-01 5.278660e-01 5.173781e-01 5.135641e-01 4.965416e-01
##  [96] 4.903519e-01 4.873866e-01 4.668871e-01 4.610425e-01 4.505757e-01
## [101] 4.395313e-01 4.355703e-01 4.280842e-01 4.189822e-01 4.095283e-01
## [106] 4.021540e-01 3.878831e-01 3.841837e-01 3.738866e-01 3.654198e-01
## [111] 3.582208e-01 3.509715e-01 3.384415e-01 3.241306e-01 3.185773e-01
## [116] 3.119209e-01 3.054884e-01 2.848054e-01 2.714743e-01 2.632768e-01
## [121] 2.502334e-01 2.432711e-01 2.226430e-01 2.136655e-01 8.033395e-15
```

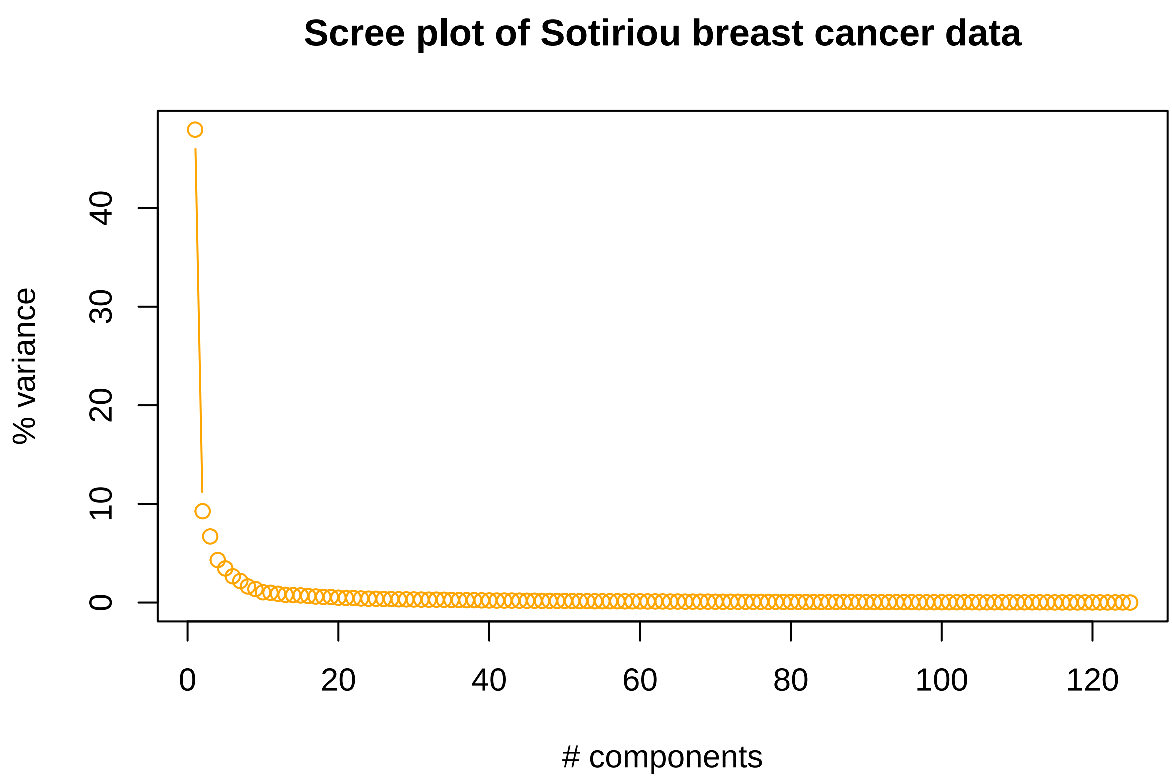
```
# percent variability of the principal components
print("Percent variability of the principal components")
```

```
## [1] "Percent variability of the principal components"
```

```
dat.pca.var <- round(dat.pca$sdev^2 / sum(dat.pca$sdev^2)*100,2)
dat.pca.var
```

```
##   [1] 47.95  9.26  6.70  4.31  3.46  2.66  2.18  1.63  1.37  1.04  0.99  0.89
##  [13]  0.78  0.75  0.72  0.66  0.61  0.57  0.56  0.49  0.47  0.46  0.41  0.39
##  [25]  0.38  0.36  0.35  0.33  0.32  0.31  0.30  0.29  0.29  0.28  0.26  0.25
##  [37]  0.24  0.24  0.22  0.21  0.20  0.20  0.20  0.19  0.18  0.18  0.17  0.17
##  [49]  0.16  0.16  0.15  0.14  0.14  0.13  0.13  0.13  0.13  0.12  0.12  0.12
##  [61]  0.11  0.11  0.11  0.10  0.10  0.09  0.09  0.09  0.09  0.08  0.08  0.08
##  [73]  0.08  0.07  0.07  0.07  0.07  0.07  0.06  0.06  0.06  0.06  0.05  0.05
##  [85]  0.05  0.05  0.05  0.05  0.05  0.04  0.04  0.04  0.04  0.04  0.04  0.04
##  [97]  0.03  0.03  0.03  0.03  0.03  0.03  0.03  0.03  0.02  0.02  0.02  0.02
## [109]  0.02  0.02  0.02  0.02  0.02  0.02  0.01  0.01  0.01  0.01  0.01  0.01
## [121]  0.01  0.01  0.01  0.01  0.00
```

```r
plot(c(1:length(dat.pca.var)),dat.pca.var,type='b',
     xlab='# components',ylab='% variance',
     main='Scree plot of Sotiriou breast cancer data', col='orange')
```



How much variability in the data is explained using only the first two eigenvalues?

```r
#summary(dat.pca)
summary(dat.pca)$importance[, 1:2]
```

```
##                             PC1      PC2
## Standard deviation     18.0896 7.950493
## Proportion of Variance  0.4795 0.092620
## Cumulative Proportion   0.4795 0.572120
```

```r
variability <- round((summary(dat.pca)$importance[3,2])*100,2)
variability
```

```
## [1] 57.21
```

```r
# or
```

```r
variability <- round((summary(dat.pca)$importance[2,1] +
                        summary(dat.pca)$importance[2,2])*100,2)
variability
```

```
## [1] 57.21
```

```r
#or

variability <- dat.pca.var[1] + dat.pca.var[2]
variability
```

```
## [1] 57.21
```

Multidimensional scaling (MDS)

Dimensionality reduction technique that fits the original data into a low-dimensional coordinate system, such that any distortion caused by dimension reduction is minimized

MDS uses the distances or similarities between instances (genes or samples) in representing proximities, while preserving (nearly matching) the original distances or similarities

`Stress` is the measure used to determine how close the low-dimensional space matches the high-dimensional space

Metric (classical) MDS

Determine the distance or similarity values between all pairs of genes/samples

Arranges the N items in low-dimensional space using the actual magnitudes of the distances/similarities

Also known as `principal coordinate analysis`

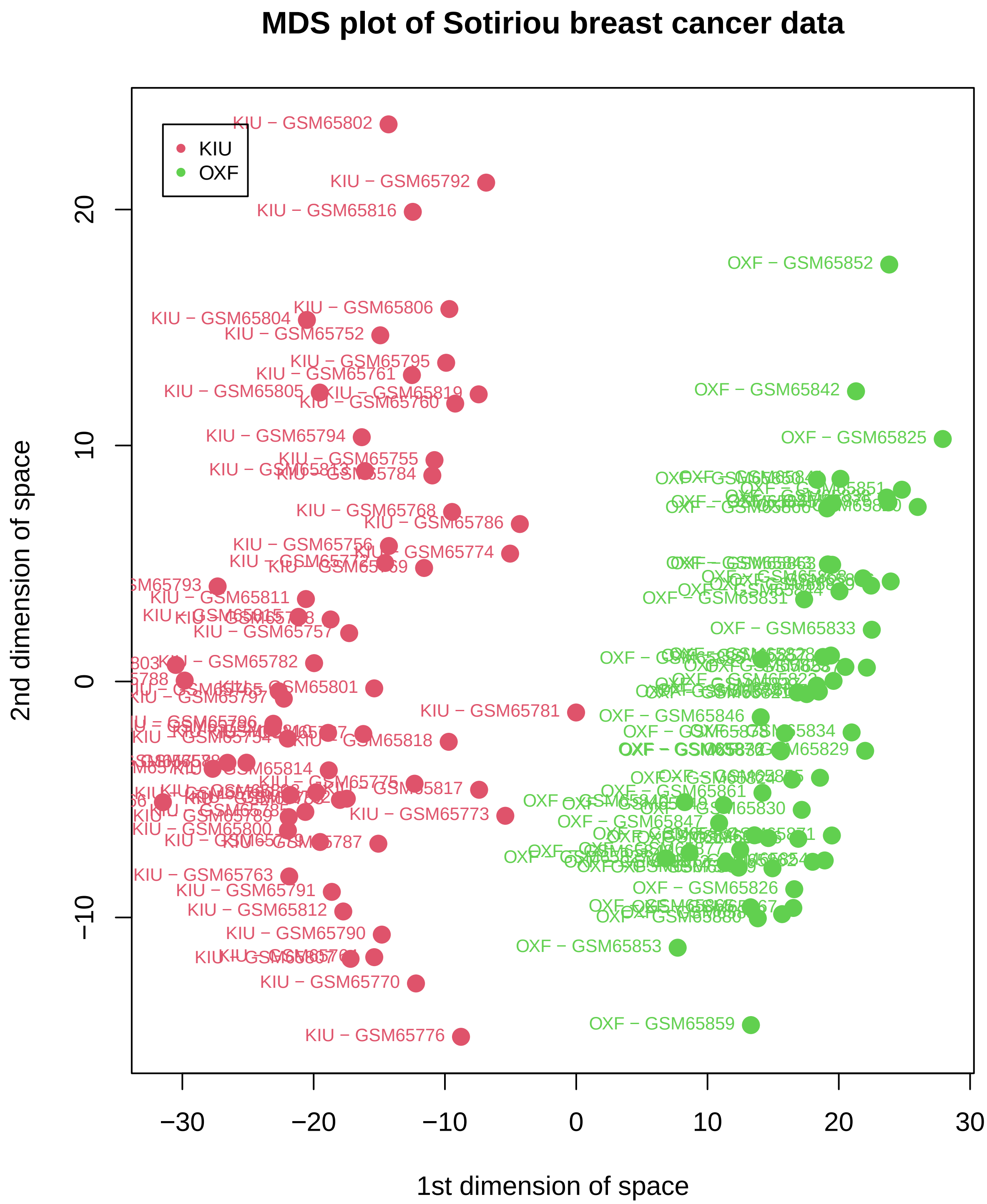
`dist` this function computes and returns the distance matrix computed by using the specified distance measure to compute the distances between the rows of a data matrix

`cmdscale` classical multidimensional scaling of a data matrix. takes a set of distances/dissimilarities and returns a set of points such that the distances between the points are approximatively equal to the dissimilarities

'points' a matrix with k=2 columns whose rows give the coordinates of the points chosen to represent the dissimilarities k the dimension of the space which the data are to be represented in

```r
dat.dist <- dist(t(dat), method = "euclidean")
dat.loc <- cmdscale(dat.dist)

col <- as.numeric(as.factor(unique(ann$site))) +1

#xlab='1st dimension of space coordinates of the points representing dissimilarities between all pairs
#ylab='2nd dimension of space coordinates of the points representing dissimilarities between all pairs

plot(dat.loc, type = "n",xlab='1st dimension of space', ylab='2nd dimension of space')

points(dat.loc[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]],
       dat.loc[,2][as.character(ann$site)==levels(as.factor(ann$site))[1]],
```

```
       col=col[1],pch=16,cex=1.5)

text(dat.loc[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]],
     dat.loc[,2][ as.character(ann$site)==levels(as.factor(ann$site))[1]],
     col=col[1] ,cex=0.7,
     labels= paste(levels(as.factor(ann$site))[1], '-',
                   row.names(ann[as.character(ann$site)==levels(as.factor(ann$site))[1],]), sep= ' '),
     pos=2)

points(dat.loc[,1][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
       dat.loc[,2][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
       col=col[2],pch=16,cex=1.5)

text(dat.loc[,1][as.character(ann$site)==levels(as.factor(ann$site))[2]],
     dat.loc[,2][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
     col=col[2],cex=0.7,
     labels= paste(levels(as.factor(ann$site))[2], '-',
                   row.names(ann[as.character(ann$site)==levels(as.factor(ann$site))[2],]), sep= ' '),
     pos=2)

title(main='MDS plot of Sotiriou breast cancer data')

legend(min(range(dat.loc[,1])), max(range(dat.loc[,2]) ), levels(as.factor(ann$site)),
       col=col,pch=16,cex=.75)
```



Non-Metric MDS

Determine the distance or similarity values between all pairs of genes/samples

Arrange the N items in low-dimensional space using only the rank orders of the distances/similarities

`isoMDS` Kruskal's non-metric MDS chooses a k-dimensional (default k=2) configuration to minimize the stress, which is the square root of the ratio of the sum of squared differences between the input distances and those of the configuration to the sum of configuration distances squared

```
Arguments:

`d` distance structure of the form returned by dist, or a full, symmetric matrix.
    Data are assumed to be dissimilarities or relative distances,
    but must be positive except for self-distance.
    Both missing and infinite values are allowed

`y` an initial configuration.
    If none is supplied, cmdscale is used to provide the classical solution,
    unless there are missing or infinite dissimilarities.

`k` the desired dimension for the solution, passed to cmdscale

`trace` logical for tracing optimization (default TRUE).
       If TRUE, the initial stress and the current stress are printed out every 5 iterations

Returns:

`points` a k-column vector of the fitted configuration
`stress` the final stress achieved (in percent)
         Kruskal's guidelines for stress values:

         Stress     Goodness of fit
         20%        Poor
         10%        Fair
         5%         Good
         2.5%       Excellent
         0%         Perfect
```

```
library(MASS)
dat.dist <- dist(t(dat))
dat.mds <- isoMDS(dat.dist)
```

```
## initial  value 16.338038
## iter   5 value 12.929006
## iter   5 value 12.920081
## iter   5 value 12.911624
## final  value 12.911624
## converged
```

```
dat.mds$stress
```

```
## [1] 12.91162
```

```
dat.mds$points
```

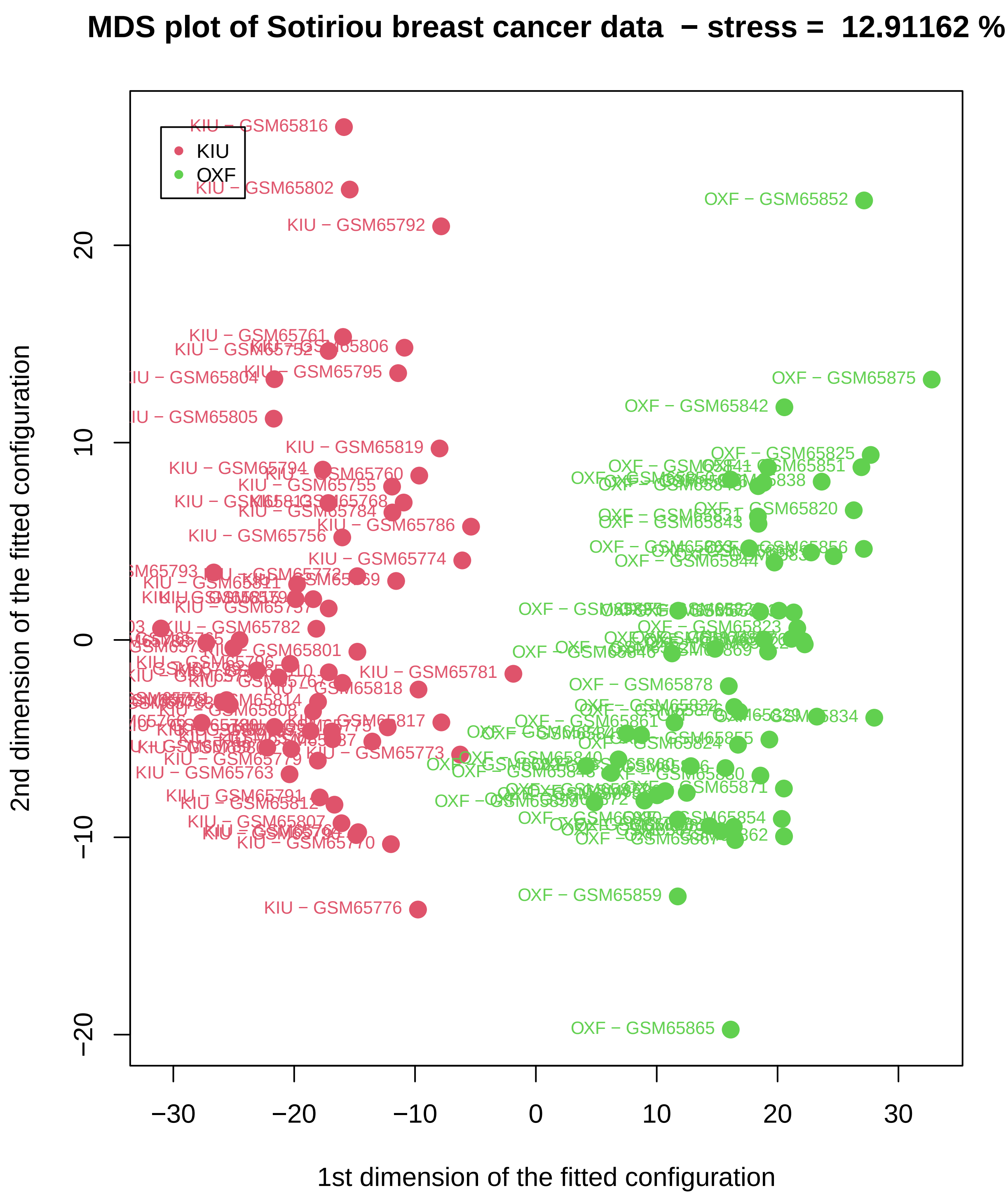
```
##                        [,1]          [,2]
## KIU - GSM65752 -17.151027  14.641093982
## KIU - GSM65753 -16.883410  -4.637055785
## KIU - GSM65754 -21.287060  -1.895686705
## KIU - GSM65755 -11.899699   7.770232937
## KIU - GSM65756 -16.022333   5.195935020
## KIU - GSM65757 -17.145175   1.595395055
## KIU - GSM65758 -25.926887  -3.142196581
## KIU - GSM65760  -9.651152   8.328254579
## KIU - GSM65761 -15.954336  15.353882743
## KIU - GSM65762 -16.821964  -5.028819276
## KIU - GSM65763 -20.384621  -6.806571394
## KIU - GSM65764 -14.706683  -9.742171976
## KIU - GSM65765 -24.518610  -0.006284012
## KIU - GSM65766 -27.650678  -4.197785312
## KIU - GSM65767 -16.008319  -2.175667431
## KIU - GSM65768 -10.938730   6.963334809
## KIU - GSM65769 -11.565692   2.988331559
## KIU - GSM65770 -11.993451 -10.347841968
## KIU - GSM65771 -25.593893  -3.047530644
## KIU - GSM65772 -14.776140   3.231414442
## KIU - GSM65773  -6.269454  -5.807296550
## KIU - GSM65774  -6.090261   4.030232976
## KIU - GSM65775 -12.262017  -4.429950163
## KIU - GSM65776  -9.754503 -13.661446432
## KIU - GSM65779 -18.044675  -6.114094477
## KIU - GSM65780 -21.631204  -4.406496593
## KIU - GSM65781  -1.862114  -1.714468235
## KIU - GSM65782 -18.168993   0.567211841
## KIU - GSM65783 -25.328102  -3.282019741
## KIU - GSM65784 -11.871661   6.459008856
## KIU - GSM65785 -18.643626  -4.622697531
## KIU - GSM65786  -5.365098   5.740490991
## KIU - GSM65787 -13.535170  -5.147597109
## KIU - GSM65788 -27.266132  -0.133209234
## KIU - GSM65789 -22.221519  -5.460472963
## KIU - GSM65790 -14.861950  -9.886741736
## KIU - GSM65791 -17.881775  -7.980270553
## KIU - GSM65792  -7.837928  20.959090232
## KIU - GSM65793 -26.654949   3.418750605
## KIU - GSM65794 -17.629829   8.631692124
## KIU - GSM65795 -11.397825  13.520460099
## KIU - GSM65796 -20.337572  -1.209592993
## KIU - GSM65797 -25.037526  -0.408396023
## KIU - GSM65798 -18.421368   2.067289882
## KIU - GSM65799 -23.076445  -1.543285039
## KIU - GSM65800 -20.237424  -5.529577013
## KIU - GSM65801 -14.770628  -0.597225403
## KIU - GSM65802 -15.404161  22.823974290
## KIU - GSM65803 -31.019931   0.582009327
## KIU - GSM65804 -21.637043  13.215207683
## KIU - GSM65805 -21.695918  11.212533556
## KIU - GSM65806 -10.873126  14.809000678
## KIU - GSM65807 -16.083831  -9.288840842
```

```
## KIU - GSM65808 -18.441697  -3.620804169
## KIU - GSM65810 -17.131395  -1.640564049
## KIU - GSM65811 -19.758147   2.822279648
## KIU - GSM65812 -16.665574  -8.348639541
## KIU - GSM65813 -17.174748   6.937611790
## KIU - GSM65814 -18.026697  -3.134713461
## KIU - GSM65815 -19.881315   2.068410381
## KIU - GSM65816 -15.883076  25.989102864
## KIU - GSM65817  -7.822488  -4.176508388
## KIU - GSM65818  -9.711061  -2.512556659
## KIU - GSM65819  -7.970089   9.703869008
## OXF - GSM65820  26.301652   6.572513909
## OXF - GSM65821  14.817114  -0.449889967
## OXF - GSM65822  22.247129  -0.223754255
## OXF - GSM65823  21.631298   0.597635028
## OXF - GSM65824  16.725348  -5.309647811
## OXF - GSM65825  27.705806   9.374931837
## OXF - GSM65826  16.378895  -9.467242710
## OXF - GSM65827   4.160650  -6.382864233
## OXF - GSM65828  20.111154   1.481693966
## OXF - GSM65829  23.250079  -3.885700136
## OXF - GSM65830  18.581051  -6.872847638
## OXF - GSM65831  18.371559   6.254751645
## OXF - GSM65832  16.408230  -3.396329917
## OXF - GSM65833  21.334099   1.394280995
## OXF - GSM65834  28.000274  -3.941378608
## OXF - GSM65835  11.779884   1.482996680
## OXF - GSM65836  15.685625  -6.491082073
## OXF - GSM65837  21.146913   0.055545053
## OXF - GSM65838  23.647233   8.020667820
## OXF - GSM65839  24.639877   4.247008685
## OXF - GSM65840   6.833662  -6.051474929
## OXF - GSM65841  19.209951   8.739261080
## OXF - GSM65842  20.563981  11.791177257
## OXF - GSM65843  18.418868   5.887395120
## OXF - GSM65844  19.738720   3.922813470
## OXF - GSM65845  18.390451   7.794848771
## OXF - GSM65846  11.264766  -0.681867057
## OXF - GSM65847   7.504335  -4.736806799
## OXF - GSM65848   6.252025  -6.743715442
## OXF - GSM65849   8.700927  -4.808869087
## OXF - GSM65850  16.121194   8.127041486
## OXF - GSM65851  26.936038   8.753281312
## OXF - GSM65852  27.157881  22.271310537
## OXF - GSM65853   4.851631  -8.236269914
## OXF - GSM65854  20.347172  -9.069325266
## OXF - GSM65855  19.321568  -5.047125307
## OXF - GSM65856  27.136160   4.613976265
## OXF - GSM65857  18.558166   1.423730807
## OXF - GSM65858  14.361319  -9.429736487
## OXF - GSM65859  11.739784 -12.995563014
## OXF - GSM65860  12.826588  -6.394592941
## OXF - GSM65861  11.467749  -4.186009486
## OXF - GSM65862  20.535134  -9.954669243
```

```
## OXF - GSM65863  17.645194   4.645146128
## OXF - GSM65864  12.485211  -7.747147721
## OXF - GSM65865  16.121976 -19.745240698
## OXF - GSM65866  18.839104   7.958324192
## OXF - GSM65867  16.487401 -10.142281024
## OXF - GSM65868  22.776003   4.433617303
## OXF - GSM65869  19.209990  -0.589129127
## OXF - GSM65870  22.132595  -0.062026796
## OXF - GSM65871  20.521898  -7.528778400
## OXF - GSM65872   8.972325  -8.137810114
## OXF - GSM65873  10.032451  -7.870412779
## OXF - GSM65874  18.864495   0.024373613
## OXF - GSM65875  32.754722  13.194626880
## OXF - GSM65876  16.823131  -3.632139765
## OXF - GSM65877  10.698055  -7.660935646
## OXF - GSM65878  15.961626  -2.338682247
## OXF - GSM65879  15.290518  -9.689594984
## OXF - GSM65880  11.741269  -9.103034191
```

```r
col <- as.numeric(as.factor(unique(ann$site))) +1

# xlab='1st dimension of the fitted configuration coordinates of the points
# representing dissimilarities between all pairs of samples'
# ylab='2nd dimension of the fitted configuration coordinates of the points
# representing dissimilarities between all pairs of samples'

plot(dat.mds$points, type = "n",
     xlab='1st dimension of the fitted configuration',
     ylab='2nd dimension of the fitted configuration')

points(dat.mds$points[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]],
       dat.mds$points[,2][as.character(ann$site)==levels(as.factor(ann$site))[1]],
       col=col[1],pch=16,cex=1.5)

text(dat.mds$points[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]],
     dat.mds$points[,2][ as.character(ann$site)==levels(as.factor(ann$site))[1]],
     col=col[1], cex=0.7,
     labels= paste(levels(as.factor(ann$site))[1], '-',
                   row.names(ann[as.character(ann$site)==levels(as.factor(ann$site))[1],]), sep= ' '),
     pos=2)

points(dat.mds$points[,1][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
       dat.mds$points[,2][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
       col=col[2],pch=16,cex=1.5)

text(dat.mds$points[,1][as.character(ann$site)==levels(as.factor(ann$site))[2]],
     dat.mds$points[,2][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
     col=col[2],cex=0.7,
     labels= paste(levels(as.factor(ann$site))[2], '-',
                   row.names(ann[as.character(ann$site)==levels(as.factor(ann$site))[2],]), sep= ' '),
     pos=2)

title(main=paste('MDS plot of Sotiriou breast cancer data', ' - stress = ',
                 round(dat.mds$stress,5), '%'))
```

```
legend(min(range(dat.mds$points[,1])), max(range(dat.mds$points[,2]) ),
       levels(as.factor(ann$site)),
       col=col,pch=16,cex=.75)
```



Non-linear Dimensionality Reduction

Weighted Graph Laplacian

Determines the subspace that best preserves local distances and minimizes large distances Does not calculate linear projections of the data (e.g. MDS & PCA)

Builds a graph from neighborhood information of the data set

Each data point serves as a vertex (node) on the graph and connectivity between vertices is governed by the proximity of neighboring points (edge weights)

The graph thus generated can be considered as a discrete approximation of the low-dimensional manifold in the high-dimensional space

Minimization of a cost function based on the graph ensures that points close to each other on the manifold are mapped close to each other in the low-dimensional space, preserving local distances

Distances are calculated between each pair of genes/samples

Each pair of vertices is assigned a weight specific to the distance between them

A kernel is implemented to transform the distances to a predefined function (cells in adjacency matrix)

The Laplacian operator decomposes the adjacency matrix

```
k.speClust2 <- function (X, qnt=NULL) {

    dist2full <- function(dis) {
              n <- attr(dis, "Size")
            full <- matrix(0, n, n)
            full[lower.tri(full)] <- dis
            full + t(full)
    }

    #squared Euclidean distances between all pairs of samples
    dat.dis <- dist(t(X),"euc")^2

    if(!is.null(qnt)) {eps <- as.numeric(quantile(dat.dis,qnt))}
    if(is.null(qnt)) {eps <- min(dat.dis[dat.dis!=0])}

    # a radial basis function (RBF) kernel to transform the distances
    # the RBF kernel decreases with distance, ranges from 0 to 1 (identity), and
    # is readily interpreted as a similarity measure
    kernel <- exp(-1 * dat.dis/(eps))

    # calculate the adjacency matrix K1 - square matrix with elements indicating
    # whether pairs of vertices are adjacent or not in the graph
    K1 <- dist2full(kernel)
    diag(K1) <- 0

    # calculate the degree matrix D - diagonal matrix calculated from the row sums of K1
    # contains information about the degree of each vertex
    # (i.e. the number of edges attached to each vertex)
    D = matrix(0,ncol=ncol(K1),nrow=ncol(K1))
    tmpe <- apply(K1,1,sum)
    tmpe[tmpe>0] <- 1/sqrt(tmpe[tmpe>0])
    tmpe[tmpe<0] <- 0
```

```
    diag(D) <- tmpe

    # calculate the normalized Laplacian
    L <- D%*% K1 %*% D

    # calculate eigenvectors by single value decomposition of the Laplacian and
    # place as columns of matrix X
    X <- svd(L)$u

    # scale the rows of matrix X to unit length and place in matrix Y
    # can then create n-dimensional embedding of data utilizing the first n columns of the matrix Y
    Y <- X / sqrt(apply(X^2,1,sum))
}
```

Plot a two-dimensional embedding of the weighted graph Laplacian

```
# center and scale the rows of the data matrix
dat.t.c.s <- t(dat)
dat.t.c.s <- scale(dat.t.c.s, center=T, scale=T)

# conduct spectral graph dimensionality reduction
phi <- k.speClust2(t(dat.t.c.s), qnt=NULL)
#phi

#plot
col <- as.numeric(as.factor(unique(ann$site))) +1

plot(range(phi[,1]),range(phi[,2]),
     xlab="phi1",ylab="phi2",
     main="Weighted Graph Laplacian plot of Sotiriou breast cancer data")

points(phi[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]],
       phi[,2][as.character(ann$site)==levels(as.factor(ann$site))[1]],
       col=col[1],pch=16,cex=1.5)

text(phi[,1][as.character(ann$site)==levels(as.factor(ann$site))[1]],
     phi[,2][ as.character(ann$site)==levels(as.factor(ann$site))[1]],
     col=col[1],cex=0.7,
     labels= paste(levels(as.factor(ann$site))[1], '-',
                   row.names(ann[as.character(ann$site)==levels(as.factor(ann$site))[1],]), sep= ' '),
     pos=2)

points(phi[,1][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
       phi[,2][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
       col=col[2],pch=16,cex=1.5)

text(phi[,1][as.character(ann$site)==levels(as.factor(ann$site))[2]],
     phi[,2][ as.character(ann$site)==levels(as.factor(ann$site))[2]],
     col=col[2],cex=0.7,
     labels= paste(levels(as.factor(ann$site))[2], '-',
                   row.names(ann[as.character(ann$site)==levels(as.factor(ann$site))[2],]), sep= ' '),
     pos=2)
```

```
legend(min(range(phi[,1])), max(range(phi[,2])-0.5),
       levels(as.factor(ann$site)),col=col,pch=16,cex=.75)
```