Class 7: Machine Learning !

Woocheol Kim (PID: A16998418

Today, we are going to learn how to apply different machine learning methods, beginning with clustering:

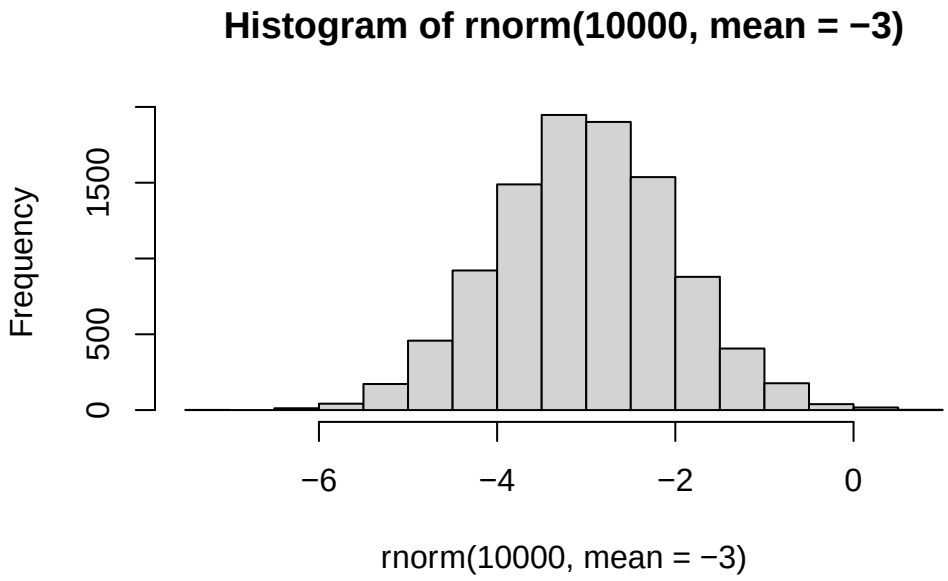
The goal here is to find groups/clusters in your input data.

First I will make up some data with clear groups. For this I will use the `rnorm()` function:

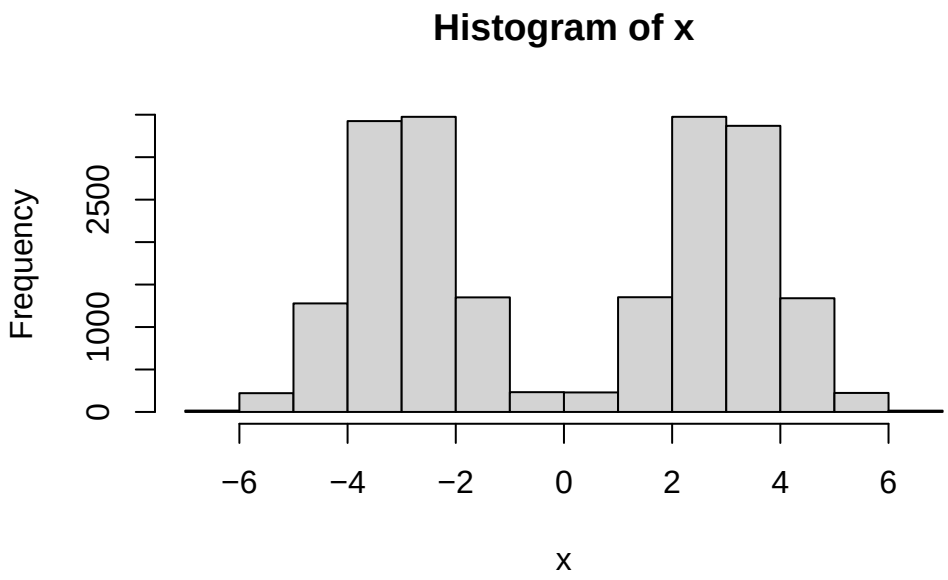
```
rnorm(10)
```

```
 [1]  1.2252965  1.8047302  0.7287003  0.2569677 -1.4240550  0.2942927
 [7] -1.4488933  0.2363693  0.7787932  0.9650687
```

```
hist(rnorm(10000, mean = -3) )
```



```r
n <- 10000
x <- c(rnorm(n,-3) , rnorm(n,+3))
hist(x)
```

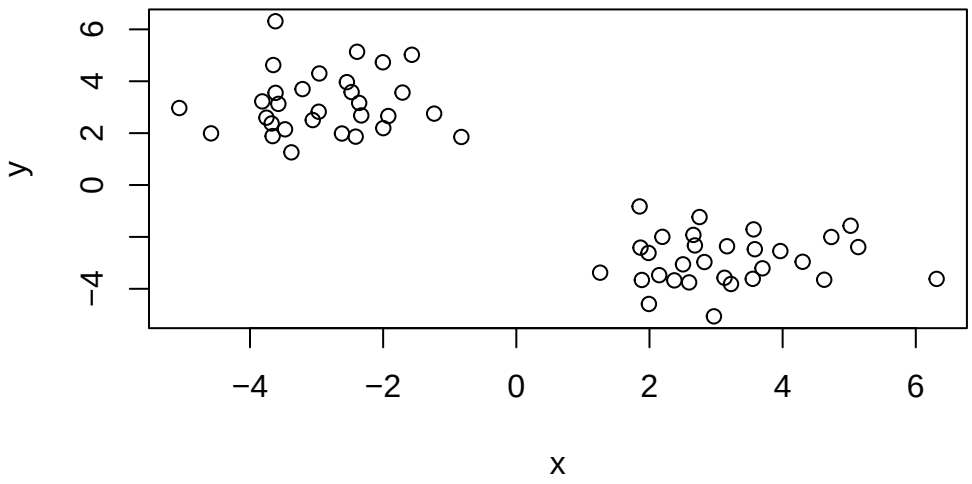



```r
n <- 30
x <- c(rnorm(n,-3) , rnorm(n,+3))
y <- rev(x)

z <- cbind(x,y)
head(z)
```

```
             x        y
[1,] -2.329109 2.682196
[2,] -3.378416 1.259114
[3,] -1.234163 2.751841
[4,] -2.958812 4.300482
[5,] -2.412750 1.865115
[6,] -3.659845 1.885551
```

```r
plot(z)
```



Use the `kmeans()` function setting k to 2 and nstart=20

Inspect/print the results

> Q. How many points are in each cluster?
>
> Q. What 'component' of your result object details - cluster size? - cluster assignment/membership? - cluster center?
>
> Q. Plot x colored by the kmeans cluster assignment and add cluster centers as blue points

```
km <- kmeans(z, centers = 2)
km
```

```
K-means clustering with 2 clusters of sizes 30, 30

Cluster means:
          x         y
1  3.151237 -2.881834
2 -2.881834  3.151237

Clustering vector:
 [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1
```

```
[39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Within cluster sum of squares by cluster:
[1] 67.47189 67.47189
 (between_SS / total_SS =  89.0 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

Results in kmeans object `km`

```
attributes(km)
```

```
$names
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"

$class
[1] "kmeans"
```

cluster size?

```
km$cluster
```

```
 [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1
[39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

cluster assignment/membership

```
km$cluster
```

```
 [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1
[39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

cluster center?

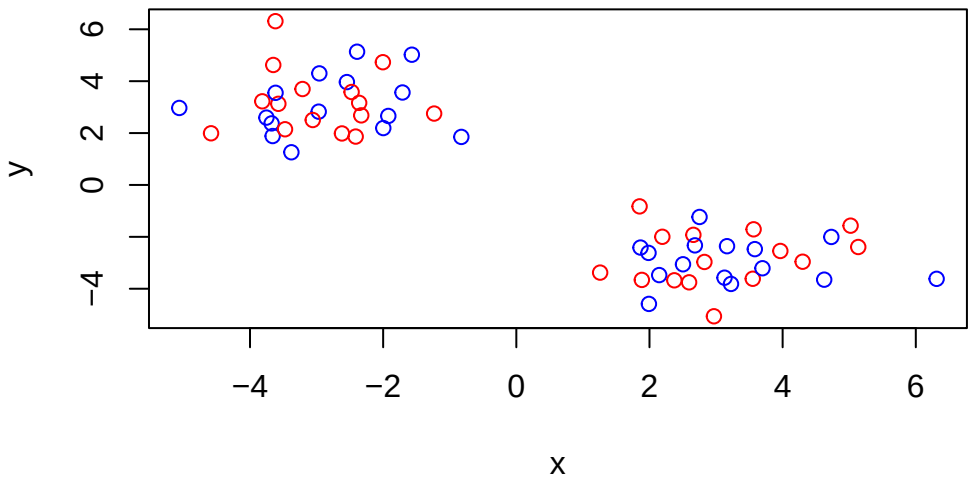
```
km$centers
```

```
          x         y
1  3.151237 -2.881834
2 -2.881834  3.151237
```

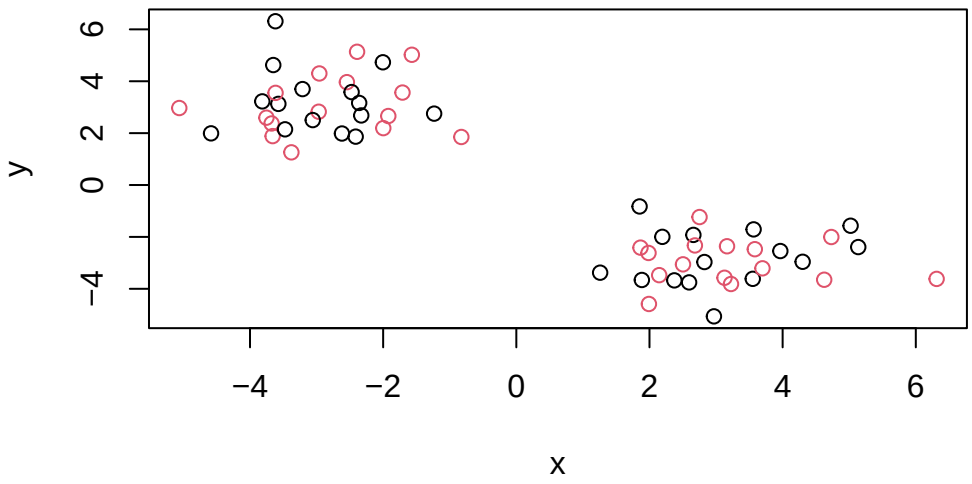
Q. Plot x colored by the kmeans cluster assignment and add cluster centers as blue points

R will re-cycle the shorter color vector to be the same length as the longer (number of data points) in z

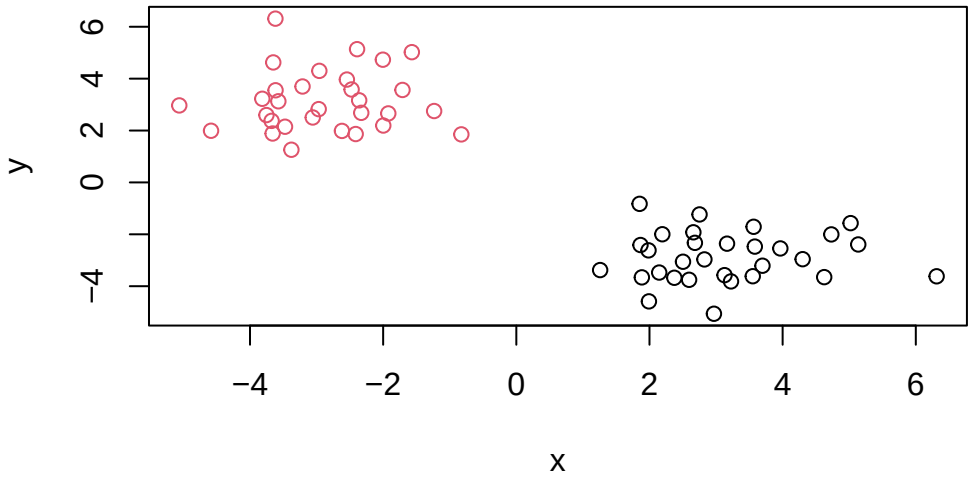
```
plot(z, col=c("red","blue") )
```




```
plot(z, col=c(1,2) )
```

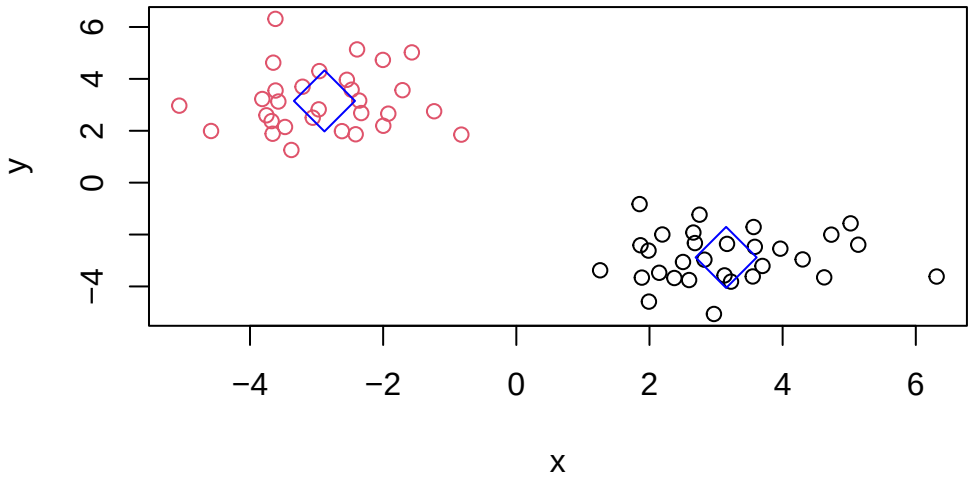



```
plot(z, col=km$cluster)
```



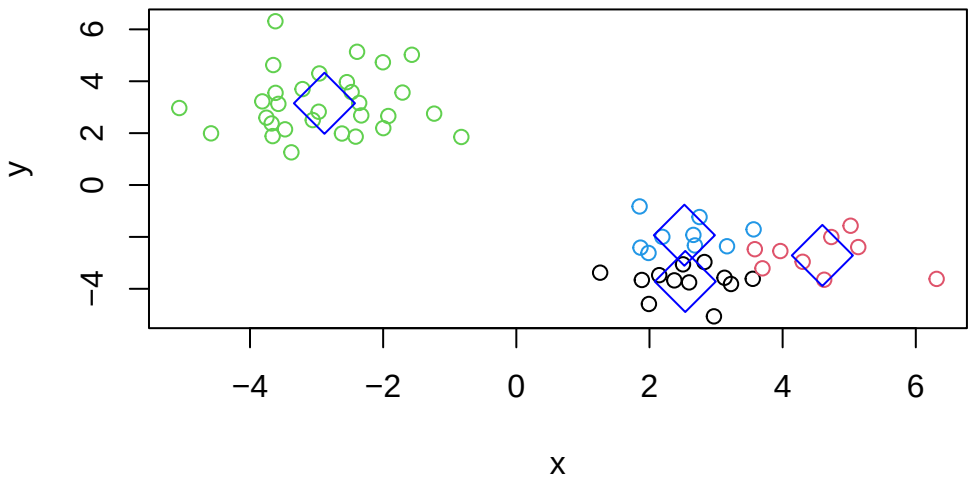
We can use the `points()` function to add new points to an existing plot… like the cluster centers.

```
plot(z, col=km$cluster)
points(km$centers, col="blue", pch=5, cex=3)
```



Q. Can you run kmeans and ask for 4 clusters please and plot the results like we have done above?

```
km4 <- kmeans(z, centers = 4)
plot(z, col=km4$cluster)
points(km4$centers, col="blue", pch=5, cex=3)
```



Hierarchical Clustering

Let's take our same made-up data **z** and see how hclust works.

First we need a distance matrix of our data to be clustered.

```
d <- dist(z)
hc <- hclust(d)
hc
```

```
Call:
hclust(d = d)

Cluster method   : complete
Distance         : euclidean
Number of objects: 60
```

```
d
```

```
          1          2          3          4          5          6          7
```

2	1.7681087						
3	1.0971589	2.6126719					
4	1.7364833	3.0701776	2.3179091				
5	0.8213516	1.1400643	1.4749074	2.4958363			
6	1.5509683	0.6867509	2.5757318	2.5146253	1.2472627		
7	0.4834178	2.1603997	1.1994550	1.2838871	1.3005441	1.8233585	
8	1.0775499	2.8454813	0.9411252	1.4500436	1.8385550	2.5729231	0.7634011
9	2.3492511	3.3762938	3.0572085	0.7640314	3.0243003	2.7388677	1.9483838
10	2.7480296	2.3991484	3.8341655	2.4898856	2.8699008	1.7716151	2.7089984
11	2.0744130	3.7339373	2.1237570	1.0470897	2.8948683	3.2920291	1.6063752
12	0.4070812	2.0196539	0.6947460	1.9410872	0.9330135	1.9015891	0.6686786
13	1.3447013	2.4438241	2.1913808	0.6539199	1.9984954	1.8663952	1.0037386
14	0.6548548	1.6188980	1.7355531	1.4753439	1.1091391	1.1667476	0.6961642
15	1.3221134	1.8783474	2.3698254	1.3247741	1.7151162	1.2446520	1.2142806
16	1.3815166	1.1529505	2.4706435	2.0563513	1.3611957	0.4878071	1.5350373
17	3.8528399	5.0594850	4.2855441	2.1177657	4.6083057	4.4275450	3.3903584
18	1.3012107	2.8317570	1.7862201	0.5325956	2.1044349	2.3598783	0.8217705
19	1.2642287	0.8923200	2.3203776	2.2149758	1.0979913	0.3201131	1.5091240
20	1.4268594	1.3877596	2.5240163	1.8812168	1.5265176	0.7158517	1.5048153
21	0.9116357	2.4928863	1.4923572	0.8664794	1.7182011	2.0699073	0.4328436
22	1.7174199	2.6202819	0.9885762	3.2475621	1.5868404	2.8340819	2.0195703
23	0.7544781	1.0510026	1.5826240	2.3396971	0.2388046	1.0457132	1.2070201
24	2.4589031	4.1741062	2.2934373	1.5652644	3.2664118	3.7683422	2.0177186
25	1.5820865	2.0132022	2.6241369	1.3754660	1.9531753	1.3477720	1.4562799
26	1.5530422	2.3045521	2.5125910	0.9966258	2.0716255	1.6663924	1.3140014
27	2.3593719	1.4115631	3.4361336	2.8242319	2.1760393	0.9313165	2.5150987
28	2.4541645	4.0000351	2.6499444	1.0093937	3.2705130	3.4886300	1.9712079
29	0.7498871	1.2850476	1.8400567	1.7997710	0.9071720	0.8629752	0.9607286
30	0.5907189	1.6664945	0.9465870	2.3161957	0.5285569	1.6899129	1.0370114
31	6.5080009	6.4536277	5.8572179	8.1371181	6.0113847	7.0240817	6.8837514
32	7.5030599	7.2957485	6.9077076	9.1637278	6.9569769	7.9005212	7.8974206
33	9.0258834	9.2637381	8.1872008	10.5026325	8.6659207	9.7803254	9.3307160
34	8.4546710	7.9366853	8.0147251	10.1714998	7.8104330	8.5911116	8.8880144
35	8.6167987	8.4729063	7.9657173	10.2496611	8.1003092	9.0703012	8.9959536
36	8.5474834	8.3274780	7.9376871	10.2028840	8.0028070	8.9382704	8.9391971
37	8.4907560	8.8624880	7.6022088	9.9057686	8.1883920	9.3427092	8.7685628
38	6.8351069	6.6190577	6.2619547	8.5044404	6.2810964	7.2222881	7.2344075
39	5.4575177	5.6302146	4.7246035	7.0298896	5.0423744	6.1422297	5.8020866
40	7.8442237	7.8985938	7.1069444	9.4170540	7.4006072	8.4530825	8.1920652
41	8.1034191	7.7980410	7.5486216	9.7832292	7.5264266	8.4216015	8.5097560
42	7.6111443	7.2750204	7.0845009	9.3008934	7.0208507	7.9017305	8.0238790
43	8.1818511	8.2705286	7.4222783	9.7366859	7.7542533	8.8188820	8.5213206
44	10.6949642	10.8495198	9.8761607	12.1931888	10.3056430	11.3907643	11.0106775

45	7.9077454	7.5783519	7.3705108	9.5941274	7.3214730	8.2050551	8.3183001
46	8.3013615	8.1045189	7.6836817	9.9523151	7.7637759	8.7104533	8.6904229
47	7.6480829	7.5069532	7.0140325	9.2891158	7.1271588	8.1002453	8.0312241
48	8.4289904	8.3694434	7.7379282	10.0362855	7.9435457	8.9499978	8.7945548
49	6.7887248	6.8245861	6.0834527	8.3838128	6.3301425	7.3777539	7.1466844
50	8.4740252	8.7414624	7.6292007	9.9440605	8.1244199	9.2486736	8.7752887
51	9.3827864	8.9575568	8.8723611	11.0810100	8.7716465	9.6021615	9.8017028
52	9.4052070	9.3889384	8.6784991	10.9876162	8.9412873	9.9639219	9.7589116
53	7.3496549	7.5505433	6.5519272	8.8695774	6.9642441	8.0690475	7.6718220
54	7.4570938	7.4774114	6.7441635	9.0478188	6.9972498	8.0374660	7.8134776
55	7.6147789	7.2045476	7.1303815	9.3185122	7.0000414	7.8423753	8.0374660
56	6.5992419	6.4013431	6.0231657	8.2666653	6.0498145	7.0000414	6.9972498
57	8.7047374	8.7610711	7.9526003	10.2661918	8.2666653	9.3185122	9.0478188
58	6.4151325	6.6178905	5.6370609	7.9526003	6.0231657	7.1303815	6.7441635
59	7.0431783	6.5584568	6.6178905	8.7610711	6.4013431	7.2045476	7.4774114
60	7.0870563	7.0431783	6.4151325	8.7047374	6.5992419	7.6147789	7.4570938

	8	9	10	11	12	13	14
2							
3							
4							
5							
6							
7							
8							
9	2.2118193						
10	3.4050135	2.1761092					
11	1.2036738	1.6498937	3.5296645				
12	0.9296142	2.6170359	3.1547130	2.0735747			
13	1.5072923	1.0261476	1.9894924	1.5890770	1.6546469		
14	1.4592562	1.9243725	2.0989310	2.1357139	1.0585937	0.9052547	
15	1.9147849	1.4991242	1.4966409	2.2441437	1.7164644	0.6757018	0.6770257
16	2.2984187	2.2514371	1.5088927	2.8901473	1.7760899	1.4032019	0.8395422
17	3.3466039	1.6888085	3.6430905	2.2600073	4.0281639	2.6472162	3.5478352
18	0.9265648	1.2874012	2.7077338	0.9371897	1.4469240	0.7182561	1.2162815
19	2.2633701	2.4844065	1.7881422	2.9734147	1.6339887	1.5730605	0.8466885
20	2.2615330	2.0317191	1.3609727	2.7605056	1.8317436	1.2280932	0.8181170
21	0.7646348	1.5718056	2.6599181	1.2411710	1.0753357	0.7458911	0.9039700
22	1.9270884	3.9583698	4.3809537	3.1113299	1.3619786	3.0159185	2.3531024
23	1.8214876	2.8333688	2.6334433	2.8133755	0.9686591	1.8111584	0.9093044
24	1.4635088	2.1190823	4.0515546	0.5228292	2.3879565	2.1090209	2.6034140
25	2.1328838	1.4097359	1.2728576	2.3558862	1.9753318	0.7673175	0.9365622
26	1.9064444	1.0735918	1.5592651	1.9973843	1.9141680	0.4304677	0.9732698
27	3.2772526	2.7937773	1.0869345	3.7637047	2.7448460	2.1902651	1.8192133

28	1.7132147	1.3588041	3.4393376	0.5601466	2.5205360	1.6553566	2.3812033
29	1.7148909	2.2024391	2.0580047	2.4639727	1.1453119	1.2036563	0.3338931
30	1.4010410	2.9399682	3.1606444	2.5381190	0.4722089	1.9324345	1.1578999
31	6.7946678	8.8321342	8.7920759	7.9308474	6.2151040	7.8511498	7.0640556
32	7.8481044	9.8429542	9.6718872	8.9988304	7.2296579	8.8476224	8.0338315
33	9.0735398	11.2439608	11.5205632	10.0852638	8.6792789	10.3316599	9.6377891
34	8.9502310	10.8004874	10.3347154	10.1369226	8.2340104	9.7808116	8.9168696
35	8.9013114	10.9442998	10.8418925	10.0270258	8.3272793	9.9607593	9.1648759
36	8.8769441	10.8856537	10.7090181	10.0190047	8.2710616	9.8921455	9.0798770
37	8.4644505	10.6561561	11.0567347	9.4362262	8.1294332	9.7722476	9.1176358
38	7.2030367	9.1779459	8.9937200	8.3631023	6.5674894	8.1792242	7.3605733
39	5.6527663	7.7447627	7.8863381	6.7636874	5.1370923	6.7889000	6.0464646
40	8.0289376	10.1343927	10.2149318	9.1172562	7.5268833	9.1779908	8.4255836
41	8.4897446	10.4498521	10.1889651	9.6509340	7.8441428	9.4458854	8.6155878
42	8.0251914	9.9595288	9.6677400	9.1952306	7.3602016	8.9515398	8.1141098
43	8.3381442	10.4596436	10.5778377	9.4114760	7.8581216	9.5113496	8.7699064
44	10.7679910	12.9313285	13.1444338	11.7851546	10.3552794	12.0094088	11.2985925
45	8.3114955	10.2555985	9.9711169	9.4780499	7.6538747	9.2488974	8.4136323
46	8.6225220	10.6375041	10.4817734	9.7625647	8.0220734	9.6460609	8.8376590
47	7.9519887	9.9792289	9.8716992	9.0885467	7.3625951	8.9924520	8.1930120
48	8.6681662	10.7419093	10.7196481	9.7767876	8.1266498	9.7698754	8.9924520
49	7.0135416	9.0927700	9.1399248	8.1252177	6.4795689	8.1266498	7.3625951
50	8.5140754	10.6861991	10.9833090	9.5252402	8.1252177	9.7767876	9.0885467
51	9.8129240	11.7320321	11.3566213	10.9833090	9.1399248	10.7196481	9.8716992
52	9.6007251	11.7028757	11.7320321	10.6861991	9.0927700	10.7419093	9.9792289
53	7.4581478	9.6007251	9.8129240	8.5140754	7.0135416	8.6681662	7.9519887
54	7.6718220	9.7589116	9.8017028	8.7752887	7.1466844	8.7945548	8.0312241
55	8.0690475	9.9639219	9.6021615	9.2486736	7.3777539	8.9499978	8.1002453
56	6.9642441	8.9412873	8.7716465	8.1244199	6.3301425	7.9435457	7.1271588
57	8.8695774	10.9876162	11.0810100	9.9440605	8.3838128	10.0362855	9.2891158
58	6.5519272	8.6784991	8.8723611	7.6292007	6.0834527	7.7379282	7.0140325
59	7.5505433	9.3889384	8.9575568	8.7414624	6.8245861	8.3694434	7.5069532
60	7.3496549	9.4052070	9.3827864	8.4740252	6.7887248	8.4289904	7.6480829

	15	16	17	18	19	20	21
2							
3							
4							
5							
6							
7							
8							
9							
10							

11								
12								
13								
14								
15								
16	0.7609451							
17	3.1859677	3.9402137						
18	1.3273381	1.9529601	2.5813517					
19	0.9860356	0.3033782	4.1691151	2.0427366				
20	0.5613585	0.2353702	3.7200749	1.8269328	0.5287123			
21	1.1902872	1.7045277	2.9607402	0.3896446	1.7498095	1.6156400		
22	3.0297898	2.8969851	5.2632863	2.7245511	2.6645216	3.0203669	2.3899840	
23	1.4884981	1.1254856	4.4412291	1.9811828	0.8701648	1.2878415	1.6031635	
24	2.7577363	3.3833659	2.4227337	1.4372655	3.4483947	3.2639769	1.6995247	
25	0.2601391	0.8626164	3.0948633	1.4709137	1.1307617	0.6320981	1.3881814	
26	0.4262098	1.1797661	2.7615328	1.1485392	1.4122444	0.9658861	1.1424026	
27	1.5205745	0.9863247	4.4282094	2.8388475	1.1217172	1.0278306	2.6432013	
28	2.3304775	3.0462045	1.7002335	1.1801265	3.1791883	2.8821573	1.5555256	
29	0.8100236	0.6318161	3.8505849	1.5492276	0.5488415	0.7019829	1.2263575	
30	1.8320952	1.6871344	4.4271537	1.8548840	1.4767470	1.8007260	1.4685969	
31	7.7154025	7.3175907	10.1412445	7.6164193	7.0207917	7.5137858	7.2747381	
32	8.6709405	8.2261967	11.1928100	8.6489370	7.9252171	8.4320274	8.2982181	
33	10.3110260	10.0170248	12.3451982	9.9700672	9.7324639	10.1922970	9.6745344	
34	9.5108346	8.9740099	12.2571646	9.6793397	8.6709193	9.1957000	9.3084624	
35	9.8104645	9.3861781	12.2471278	9.7284951	9.0864370	9.5885495	9.3874164	
36	9.7169176	9.2685840	12.2231736	9.6862870	8.9672203	9.4754578	9.3379677	
37	9.7941476	9.5481094	11.6944471	9.3737748	9.2717828	9.7121117	9.0944355	
38	7.9955259	7.5478260	10.5443265	7.9923877	7.2467906	7.7539119	7.6383091	
39	6.7130002	6.3861306	8.9935161	6.5026713	6.0988819	6.5664757	6.1760229	
40	9.0878034	8.7276569	11.3605404	8.8886866	8.4344893	8.9167988	8.5650182	
41	9.2413684	8.7668105	11.8295783	9.2735819	8.4642187	8.9782657	8.9167988	
42	8.7350191	8.2513362	11.3586072	8.7950675	7.9484684	8.4642187	8.4344893	
43	9.4348293	9.0868618	11.6607874	9.2068604	8.7950675	9.2735819	8.8886866	
44	11.9689705	11.6476928	14.0450251	11.6607874	11.3586072	11.8295783	11.3605404	
45	9.0359044	8.5541652	11.6476928	9.0868618	8.2513362	8.7668105	8.7276569	
46	9.4770827	9.0359044	11.9689705	9.4348293	8.7350191	9.2413684	9.0878034	
47	8.8376590	8.4136323	11.2985925	8.7699064	8.1141098	8.6155878	8.4255836	
48	9.6460609	9.2488974	12.0094088	9.5113496	8.9515398	9.4458854	9.1779908	
49	8.0220734	7.6538747	10.3552794	7.8581216	7.3602016	7.8441428	7.5268833	
50	9.7625647	9.4780499	11.7851546	9.4114760	9.1952306	9.6509340	9.1172562	
51	10.4817734	9.9711169	13.1444338	10.5778377	9.6677400	10.1889651	10.2149318	
52	10.6375041	10.2555985	12.9313285	10.4596436	9.9595288	10.4498521	10.1343927	
53	8.6225220	8.3114955	10.7679910	8.3381442	8.0251914	8.4897446	8.0289376	

54	8.6904229	8.3183001	11.0106775	8.5213206	8.0238790	8.5097560	8.1920652
55	8.7104533	8.2050551	11.3907643	8.8188820	7.9017305	8.4216015	8.4530825
56	7.7637759	7.3214730	10.3056430	7.7542533	7.0208507	7.5264266	7.4006072
57	9.9523151	9.5941274	12.1931888	9.7366859	9.3008934	9.7832292	9.4170540
58	7.6836817	7.3705108	9.8761607	7.4222783	7.0845009	7.5486216	7.1069444
59	8.1045189	7.5783519	10.8495198	8.2705286	7.2750204	7.7980410	7.8985938
60	8.3013615	7.9077454	10.6949642	8.1818511	7.6111443	8.1034191	7.8442237

	22	23	24	25	26	27	28
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23	1.7979713						
24	3.2551738	3.2115328					
25	3.2896021	1.7222027	2.8762803				
26	3.2672258	1.8563609	2.5198014	0.3829778			
27	3.7617548	1.9661955	4.2746343	1.4533692	1.8363289		
28	3.6382689	3.1581359	0.8306835	2.3831882	2.0021815	3.8335164	
29	2.3246903	0.6784962	2.9243304	1.0461022	1.1876240	1.6112296	2.7149977
30	1.2208692	0.6545102	2.8601379	2.0895800	2.1126812	2.5948259	2.9688593
31	4.8921408	6.2469126	7.9634380	7.9607707	8.0338102	7.8650342	8.4799101
32	5.9314036	7.1881841	9.0465640	8.9099476	9.0069157	8.7027756	9.5439557
33	7.3174497	8.9046978	9.9948498	10.5669547	10.5789196	10.6632407	10.6452729
34	7.0262227	8.0268541	10.2444747	9.7317742	9.8817064	9.3007342	10.6632407
35	7.0040005	8.3338023	10.0406215	10.0529349	10.1364060	9.8817064	10.5789196
36	6.9655172	8.2337917	10.0525337	9.9556814	10.0529349	9.7317742	10.5669547

37	6.7737463	8.4261263	9.3193465	10.0525337	10.0406215	10.2444747	9.9948498
38	5.2810464	6.5117937	8.4261263	8.2337917	8.3338023	8.0268541	8.9046978
39	3.7864760	5.2810464	6.7737463	6.9655172	7.0040005	7.0262227	7.3174497
40	6.1760229	7.6383091	9.0944355	9.3379677	9.3874164	9.3084624	9.6745344
41	6.5664757	7.7539119	9.7121117	9.4754578	9.5885495	9.1957000	10.1922970
42	6.0988819	7.2467906	9.2717828	8.9672203	9.0864370	8.6709193	9.7324639
43	6.5026713	7.9923877	9.3737748	9.6862870	9.7284951	9.6793397	9.9700672
44	8.9935161	10.5443265	11.6944471	12.2231736	12.2471278	12.2571646	12.3451982
45	6.3861306	7.5478260	9.5481094	9.2685840	9.3861781	8.9740099	10.0170248
46	6.7130002	7.9955259	9.7941476	9.7169176	9.8104645	9.5108346	10.3110260
47	6.0464646	7.3605733	9.1176358	9.0798770	9.1648759	8.9168696	9.6377891
48	6.7889000	8.1792242	9.7722476	9.8921455	9.9607593	9.7808116	10.3316599
49	5.1370923	6.5674894	8.1294332	8.2710616	8.3272793	8.2340104	8.6792789
50	6.7636874	8.3631023	9.4362262	10.0190047	10.0270258	10.1369226	10.0852638
51	7.8863381	8.9937200	11.0567347	10.7090181	10.8418925	10.3347154	11.5205632
52	7.7447627	9.1779459	10.6561561	10.8856537	10.9442998	10.8004874	11.2439608
53	5.6527663	7.2030367	8.4644505	8.8769441	8.9013114	8.9502310	9.0735398
54	5.8020866	7.2344075	8.7685628	8.9391971	8.9959536	8.8880144	9.3307160
55	6.1422297	7.2222881	9.3427092	8.9382704	9.0703012	8.5911116	9.7803254
56	5.0423744	6.2810964	8.1883920	8.0028070	8.1003092	7.8104330	8.6659207
57	7.0298896	8.5044404	9.9057686	10.2028840	10.2496611	10.1714998	10.5026325
58	4.7246035	6.2619547	7.6022088	7.9376871	7.9657173	8.0147251	8.1872008
59	5.6302146	6.6190577	8.8624880	8.3274780	8.4729063	7.9366853	9.2637381
60	5.4575177	6.8351069	8.4907560	8.5474834	8.6167987	8.4546710	9.0258834

	29	30	31	32	33	34	35
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							

```
20
21
22
23
24
25
26
27
28
29
30  1.1038219
31  6.9158744  5.9270204
32  7.8641364  6.9158744  1.1038219
33  9.5439557  8.4799101  2.9688593  2.7149977
34  8.7027756  7.8650342  2.5948259  1.6112296  3.8335164
35  9.0069157  8.0338102  2.1126812  1.1876240  2.0021815  1.8363289
36  8.9099476  7.9607707  2.0895800  1.0461022  2.3831882  1.4533692  0.3829778
37  9.0465640  7.9634380  2.8601379  2.9243304  0.8306835  4.2746343  2.5198014
38  7.1881841  6.2469126  0.6545102  0.6784962  3.1581359  1.9661955  1.8563609
39  5.9314036  4.8921408  1.2208692  2.3246903  3.6382689  3.7617548  3.2672258
40  8.2982181  7.2747381  1.4685969  1.2263575  1.5555256  2.6432013  1.1424026
41  8.4320274  7.5137858  1.8007260  0.7019829  2.8821573  1.0278306  0.9658861
42  7.9252171  7.0207917  1.4767470  0.5488415  3.1791883  1.1217172  1.4122444
43  8.6489370  7.6164193  1.8548840  1.5492276  1.1801265  2.8388475  1.1485392
44 11.1928100 10.1412445  4.4271537  3.8505849  1.7002335  4.4282094  2.7615328
45  8.2261967  7.3175907  1.6871344  0.6318161  3.0462045  0.9863247  1.1797661
46  8.6709405  7.7154025  1.8320952  0.8100236  2.3304775  1.5205745  0.4262098
47  8.0338315  7.0640556  1.1578999  0.3338931  2.3812033  1.8192133  0.9732698
48  8.8476224  7.8511498  1.9324345  1.2036563  1.6553566  2.1902651  0.4304677
49  7.2296579  6.2151040  0.4722089  1.1453119  2.5205360  2.7448460  1.9141680
50  8.9988304  7.9308474  2.5381190  2.4639727  0.5601466  3.7637047  1.9973843
51  9.6718872  8.7920759  3.1606444  2.0580047  3.4393376  1.0869345  1.5592651
52  9.8429542  8.8321342  2.9399682  2.2024391  1.3588041  2.7937773  1.0735918
53  7.8481044  6.7946678  1.4010410  1.7148909  1.7132147  3.2772526  1.9064444
54  7.8974206  6.8837514  1.0370114  0.9607286  1.9712079  2.5150987  1.3140014
55  7.9005212  7.0240817  1.6899129  0.8629752  3.4886300  0.9313165  1.6663924
56  6.9569769  6.0113847  0.5285569  0.9071720  3.2705130  2.1760393  2.0716255
57  9.1637278  8.1371181  2.3161957  1.7997710  1.0093937  2.8242319  0.9966258
58  6.9077076  5.8572179  0.9465870  1.8400567  2.6499444  3.4361336  2.5125910
59  7.2957485  6.4536277  1.6664945  1.2850476  4.0000351  1.4115631  2.3045521
60  7.5030599  6.5080009  0.5907189  0.7498871  2.4541645  2.3593719  1.5530422
          36         37         38         39         40         41         42
2
```

3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37 2.8762803
38 1.7222027 3.2115328
39 3.2896021 3.2551738 1.7979713
40 1.3881814 1.6995247 1.6031635 2.3899840
41 0.6320981 3.2639769 1.2878415 3.0203669 1.6156400
42 1.1307617 3.4483947 0.8701648 2.6645216 1.7498095 0.5287123
43 1.4709137 1.4372655 1.9811828 2.7245511 0.3896446 1.8269328 2.0427366
44 3.0948633 2.4227337 4.4412291 5.2632863 2.9607402 3.7200749 4.1691151
45 0.8626164 3.3833659 1.1254856 2.8969851 1.7045277 0.2353702 0.3033782

46	0.2601391	2.7577363	1.4884981	3.0297898	1.1902872	0.5613585	0.9860356
47	0.9365622	2.6034140	0.9093044	2.3531024	0.9039700	0.8181170	0.8466885
48	0.7673175	2.1090209	1.8111584	3.0159185	0.7458911	1.2280932	1.5730605
49	1.9753318	2.3879565	0.9686591	1.3619786	1.0753357	1.8317436	1.6339887
50	2.3558862	0.5228292	2.8133755	3.1113299	1.2411710	2.7605056	2.9734147
51	1.2728576	4.0515546	2.6334433	4.3809537	2.6599181	1.3609727	1.7881422
52	1.4097359	2.1190823	2.8333688	3.9583698	1.5718056	2.0317191	2.4844065
53	2.1328838	1.4635088	1.8214876	1.9270884	0.7646348	2.2615330	2.2633701
54	1.4562799	2.0177186	1.2070201	2.0195703	0.4328436	1.5048153	1.5091240
55	1.3477720	3.7683422	1.0457132	2.8340819	2.0699073	0.7158517	0.3201131
56	1.9531753	3.2664118	0.2388046	1.5868404	1.7182011	1.5265176	1.0979913
57	1.3754660	1.5652644	2.3396971	3.2475621	0.8664794	1.8812168	2.2149758
58	2.6241369	2.2934373	1.5826240	0.9885762	1.4923572	2.5240163	2.3203776
59	2.0132022	4.1741062	1.0510026	2.6202819	2.4928863	1.3877596	0.8923200
60	1.5820865	2.4589031	0.7544781	1.7174199	0.9116357	1.4268594	1.2642287

	43	44	45	46	47	48	49
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							

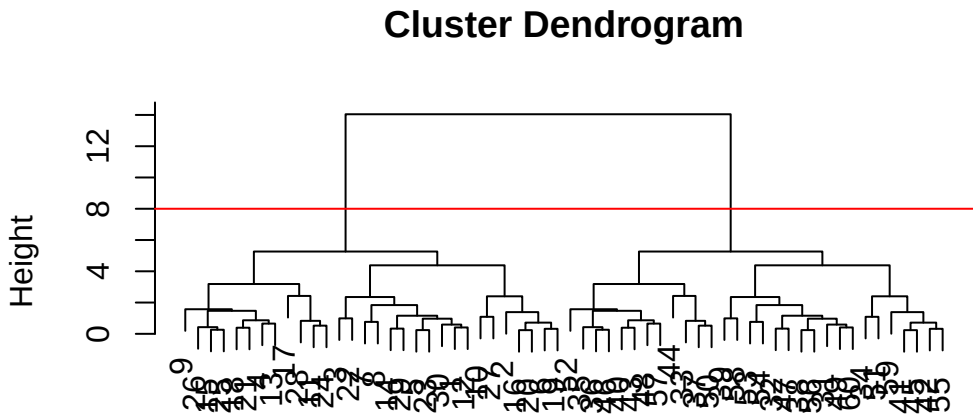
```
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44  2.5813517
45  1.9529601  3.9402137
46  1.3273381  3.1859677  0.7609451
47  1.2162815  3.5478352  0.8395422  0.6770257
48  0.7182561  2.6472162  1.4032019  0.6757018  0.9052547
49  1.4469240  4.0281639  1.7760899  1.7164644  1.0585937  1.6546469
50  0.9371897  2.2600073  2.8901473  2.2441437  2.1357139  1.5890770  2.0735747
51  2.7077338  3.6430905  1.5088927  1.4966409  2.0989310  1.9894924  3.1547130
52  1.2874012  1.6888085  2.2514371  1.4991242  1.9243725  1.0261476  2.6170359
53  0.9265648  3.3466039  2.2984187  1.9147849  1.4592562  1.5072923  0.9296142
54  0.8217705  3.3903584  1.5350373  1.2142806  0.6961642  1.0037386  0.6686786
55  2.3598783  4.4275450  0.4878071  1.2446520  1.1667476  1.8663952  1.9015891
56  2.1044349  4.6083057  1.3611957  1.7151162  1.1091391  1.9984954  0.9330135
57  0.5325956  2.1177657  2.0563513  1.3247741  1.4753439  0.6539199  1.9410872
58  1.7862201  4.2855441  2.4706435  2.3698254  1.7355531  2.1913808  0.6947460
59  2.8317570  5.0594850  1.1529505  1.8783474  1.6188980  2.4438241  2.0196539
60  1.3012107  3.8528399  1.3815166  1.3221134  0.6548548  1.3447013  0.4070812
          50         51         52         53         54         55         56
2
3
4
5
6
7
8
9
10
11
```

```
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51  3.5296645
52  1.6498937  2.1761092
53  1.2036738  3.4050135  2.2118193
54  1.6063752  2.7089984  1.9483838  0.7634011
```

```
55  3.2920291  1.7716151  2.7388677  2.5729231  1.8233585
56  2.8948683  2.8699008  3.0243003  1.8385550  1.3005441  1.2472627
57  1.0470897  2.4898856  0.7640314  1.4500436  1.2838871  2.5146253  2.4958363
58  2.1237570  3.8341655  3.0572085  0.9411252  1.1994550  2.5757318  1.4749074
59  3.7339373  2.3991484  3.3762938  2.8454813  2.1603997  0.6867509  1.1400643
60  2.0744130  2.7480296  2.3492511  1.0775499  0.4834178  1.5509683  0.8213516
           57         58         59
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
```

```
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58  2.3179091
59  3.0701776  2.6126719
60  1.7364833  1.0971589  1.7681087
```

```r
plot(hc)
abline(h=8, col="red")
```



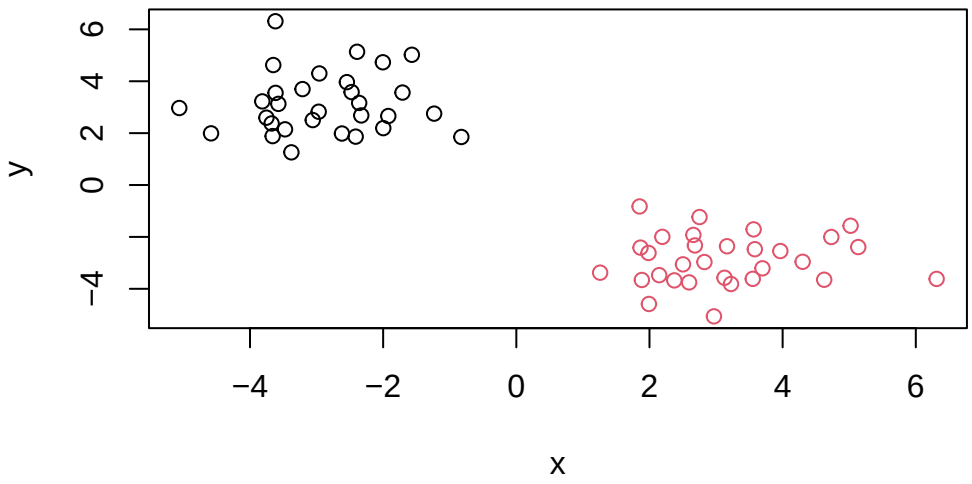
I can get my cluster membership vector by "cutting the tree" with the `cutree()` function like so:

```
grps <- cutree(hc, h=8)
grps
```

```
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
[39] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
```

Can you plot `z` colored by our hclust results:

```
plot(z, col=grps)
```



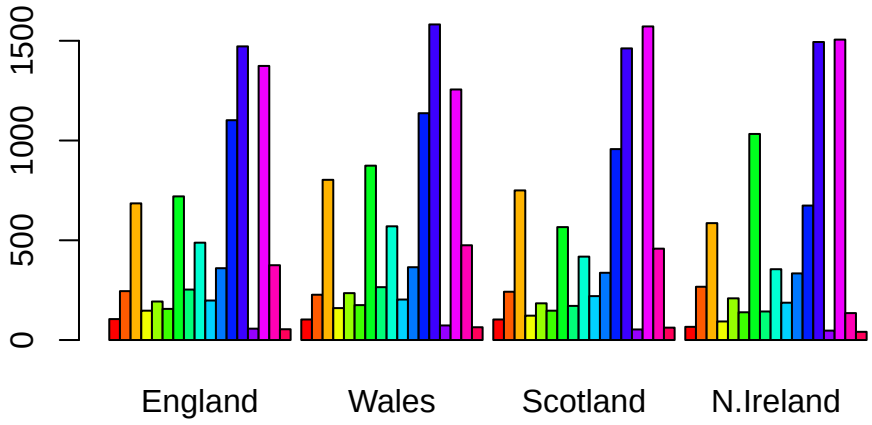
PCA of UK food data

Read data from the UK food

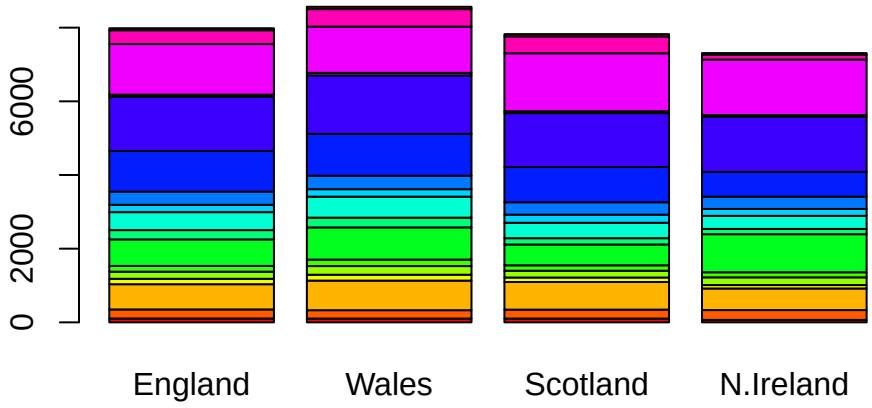
```
url <- "https://tinyurl.com/UK-foods"
x <- read.csv(url, row.names=1)
head(x)
```

```
               England Wales Scotland N.Ireland
Cheese             105   103      103        66
Carcass_meat       245   227      242       267
Other_meat         685   803      750       586
Fish               147   160      122        93
Fats_and_oils      193   235      184       209
Sugars             156   175      147       139
```

```
barplot(as.matrix(x), beside=T, col=rainbow(nrow(x)))
```

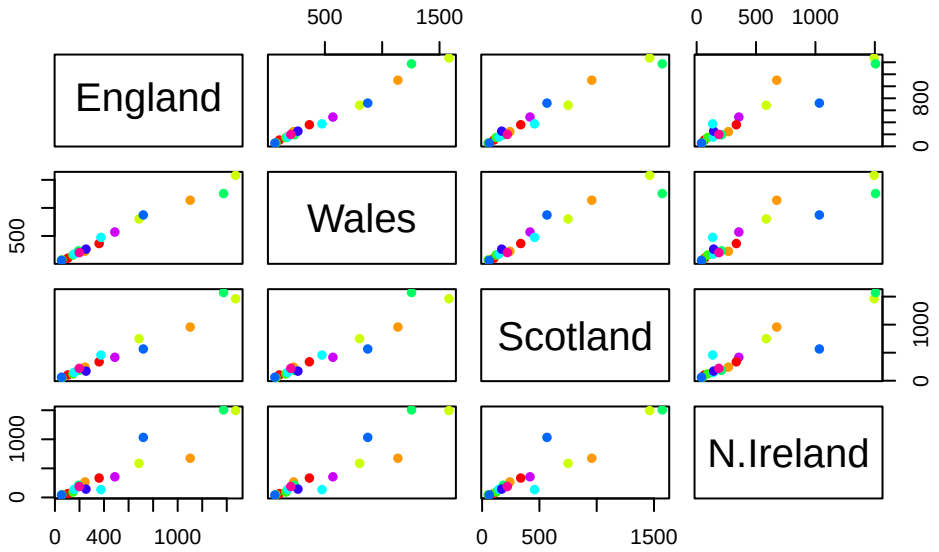



```
barplot(as.matrix(x), beside=F, col=rainbow(nrow(x)))
```



A so-called "Pairs" plot can be useful for small datasets like this

```
pairs(x, col=rainbow(10), pch=16)
```



It is hard to see structure and trends in even this small data-set. How will we ever do this when we have big datasets with 1,000s or 10s of thousands of things we are measuring…

PCA to the rescue

Let's see how PCA deals with this dataset. So main function in base R to do PCA is called `prcomp()`

```
pca <- prcomp( t(x) )
summary(pca)
```

```
Importance of components:
                            PC1      PC2      PC3        PC4
Standard deviation     324.1502 212.7478 73.87622 2.921e-14
Proportion of Variance   0.6744   0.2905  0.03503 0.000e+00
Cumulative Proportion    0.6744   0.9650  1.00000 1.000e+00
```

Let's see what is inside this `pca` object taht we created from running `prcomp()`

```
attributes(pca)
```

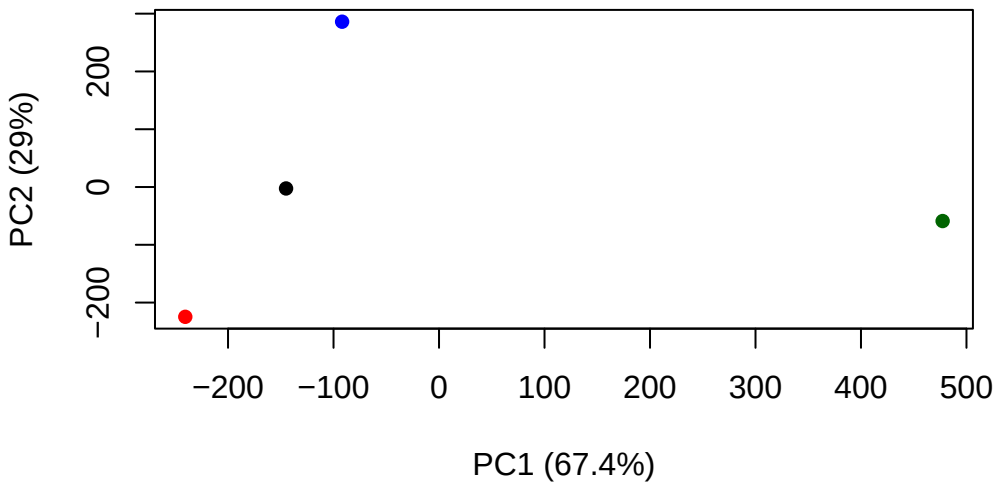
```
$names
[1] "sdev"     "rotation" "center"   "scale"    "x"

$class
[1] "prcomp"
```

```
pca$x
```

```
                PC1         PC2        PC3           PC4
England   -144.99315   -2.532999 105.768945 -9.152022e-15
Wales     -240.52915 -224.646925 -56.475555  5.560040e-13
Scotland   -91.86934  286.081786 -44.415495 -6.638419e-13
N.Ireland  477.39164  -58.901862  -4.877895  1.329771e-13
```

```
plot(pca$x[,1], pca$x[,2],
     col=c("black","red","blue","darkgreen"), pch=16,
     xlab="PC1 (67.4%)", ylab="PC2 (29%)")
```



```r
## Lets focus on PC1 as it accounts for > 90% of variance
par(mar=c(10, 3, 0.35, 0))
barplot( pca$rotation[,1], las=2 )
```