

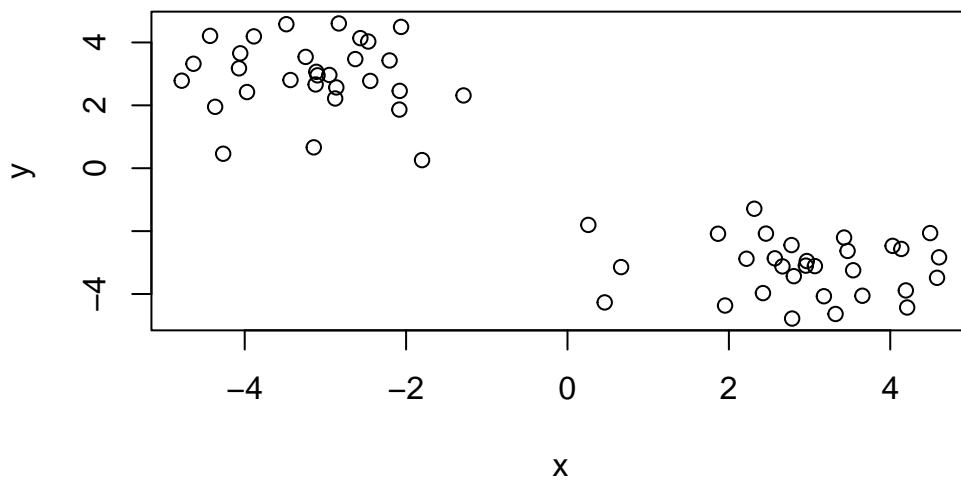
Class 7: Machine Learning

Mahsa Naeimi

Example of k-means clustering

First step is to make up some data with a known structure, so we know what the answer should be.

```
tmp <- c(rnorm(30, mean= -3), rnorm(30, mean=3))  
x <- cbind(x=tmp, y= rev(tmp))  
plot(x)
```



Now we have some structured data in `x`. Let's see if k-mean is able to identify the two groups.

```
k <- kmeans(x, centers= 2, nstart = 20)
k
```

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

Cluster means:

	x	y
1	2.934783	-3.142104
2	-3.142104	2.934783

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 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
```

Within cluster sum of squares by cluster:

```
[1] 61.45519 61.45519
(between_SS / total_SS = 90.0 %)
```

Available components:

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

Let's explore k:

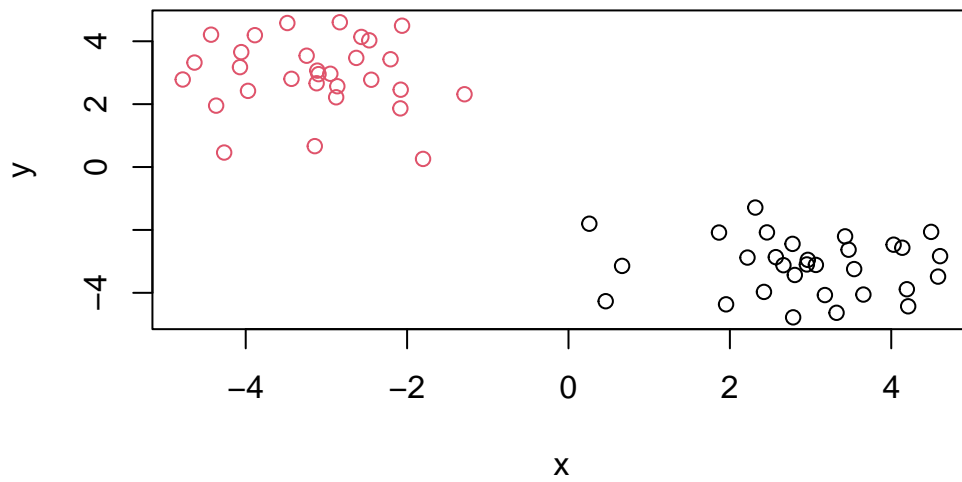
```
k$size
```

[1] 30 30

```
k$centers
```

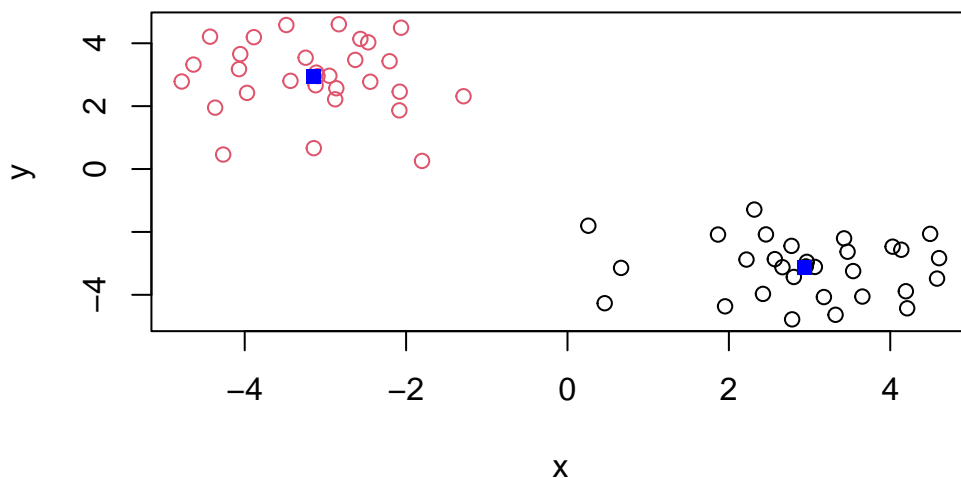
	x	y
1	2.934783	-3.142104
2	-3.142104	2.934783

```
plot(x, col= k$cluster)
```



Now we can add the clusters centers:

```
plot(x, col= k$cluster)
points(k$centers, col= 'blue', pch= 15)
```



Example of Hierarchical Clustering

Let's use the same data as before, which we stored in 'x'. We will use the 'hclust()' function.

```
dist(x)
```

	1	2	3	4	5	6	7
2	0.5388040						
3	2.7626156	2.2256934					
4	0.7790038	0.6177529	2.3126622				
5	0.6813359	0.1452447	2.0816574	0.6174304			
6	1.6429408	1.2323936	1.7861007	1.7676931	1.1675051		
7	1.5479148	1.5649991	3.0741154	2.1624766	1.6300241	1.3007743	
8	1.9623455	1.5741324	1.8657051	2.1144623	1.5133386	0.3468048	1.3616007
9	1.6440675	1.8643567	3.6684278	2.3839810	1.9687861	1.9315897	0.6697024
10	0.6511228	1.0189078	3.1545640	1.4242917	1.1548227	1.6991725	1.0873411
11	1.1415501	0.9035729	2.3084519	1.5192976	0.9173073	0.6444822	0.8184429
12	2.6658125	2.5947241	3.5236750	3.2113236	2.6257584	1.8382394	1.1329707
13	1.5641852	1.2023045	1.9728033	1.7727960	1.1587487	0.1894812	1.1112948
14	2.8406916	2.3226905	0.9136363	2.6291747	2.1983737	1.4020829	2.6626045

15	1.6718779	1.2966574	1.9187678	1.8555991	1.2455823	0.1465967	1.1865027
16	1.8694426	1.3659520	1.2445468	1.7595066	1.2531066	0.5416464	1.8348099
17	1.8891468	1.7844467	2.9134398	2.4016520	1.8145226	1.1394547	0.4770888
18	4.3848654	4.0502511	3.5085212	4.5963338	3.9958015	2.8296640	3.2004145
19	2.2742573	1.7468956	0.8041821	2.0348538	1.6176175	1.0086202	2.3091226
20	1.8973159	1.5891219	2.1997187	2.1752871	1.5584495	0.5077451	1.0532664
21	3.9529734	3.5204913	2.4836821	3.9797539	3.4322506	2.3100982	3.1259741
22	2.3872107	1.9440794	1.5935438	2.4179907	1.8570388	0.7515774	1.8564887
23	2.0357790	1.5957195	1.5970006	2.0853113	1.5127190	0.4061515	1.6106351
24	4.4673393	3.9532875	2.1199808	4.2438965	3.8295979	2.9416690	4.0754223
25	1.1305840	1.3212552	3.2061519	1.8487194	1.4270700	1.5428167	0.5647217
26	1.1521306	0.6689769	1.7708333	1.1698691	0.5810717	0.5985665	1.4372971
27	2.2115871	2.2233178	3.4951544	2.8270305	2.2787231	1.7209645	0.6683118
28	2.4613424	2.2166934	2.6854465	2.8171787	2.2000158	1.1560367	1.2347449
29	1.3344541	0.7960053	1.4417939	1.0757755	0.6569336	0.8777023	1.8641580
30	3.0631760	2.8302209	3.0995197	3.4302339	2.8129247	1.7405297	1.7296542
31	10.1687883	9.6302152	7.4264470	9.7283875	9.4894837	8.8219758	10.0219356
32	9.2511939	8.7267817	6.5325668	8.6617661	8.5815685	8.2121748	9.5043595
33	10.0592033	9.5206766	7.3002412	9.5806210	9.3779487	8.7758658	10.0060294
34	11.1030807	10.5656272	8.3405300	10.5992736	10.4219361	9.8561537	11.0972133
35	9.5964995	9.0688656	6.8632199	9.0207415	8.9236257	8.5185657	9.8050423
36	11.0222025	10.4918216	8.2771400	10.4583718	10.3466749	9.9001494	11.1763280
37	7.1145278	6.5771108	4.3984228	6.7108405	6.4381774	5.7484122	6.9559851
38	9.2192817	8.6837865	6.4587590	8.6956567	8.5393357	8.0302632	9.2946884
39	9.0295202	8.4923340	6.2669047	8.5258490	8.3484952	7.8034330	9.0577868
40	8.5019923	7.9657592	5.7978001	8.1104586	7.8279730	7.1000189	8.2761277
41	9.8183729	9.2820384	7.0563869	9.3022051	9.1378455	8.6072414	9.8636611
42	8.5264367	7.9941875	5.7756461	7.9789156	7.8491805	7.3985520	8.6787525
43	9.4640661	8.9328346	6.8105843	9.1175008	8.7987150	7.9979131	9.1185257
44	10.5552163	10.0185746	7.7928831	10.0412550	9.8744981	9.3323392	10.5830678
45	9.0092472	8.4771892	6.2589205	8.4592920	8.3321651	7.8793712	9.1579871
46	9.6347718	9.1004764	6.8771461	9.0996392	8.9557616	8.4644572	9.7325438
47	8.1741366	7.6378683	5.4122575	7.6612728	7.4936458	6.9786476	8.2448332
48	9.7143458	9.1806590	6.9584542	9.1742757	9.0358413	8.5535712	9.8237768
49	10.9385151	10.3998330	8.1813403	10.4649427	10.2574131	9.6391409	10.8595340
50	10.1117407	9.5805130	7.3639400	9.5543458	9.4354207	8.9844684	10.2612171
51	10.9725665	10.4478117	8.2503258	10.3816696	10.3025875	9.9161960	11.2039092
52	11.4557918	10.9228044	8.7017258	10.9077088	10.7778935	10.2948433	11.5613418
53	9.4824680	8.9465076	6.7210517	8.9629348	8.8021905	8.2818900	9.5423661
54	10.8211849	10.2867596	8.0630840	10.2849606	10.1420737	9.6406569	10.9032987
55	9.5258135	8.9922839	6.7704342	8.9849466	8.8474414	8.3690500	9.6406569
56	9.8562505	9.3353830	7.1531146	9.2520870	9.1903086	8.8474414	10.1420737
57	9.9073764	9.3963798	7.2527604	9.2728000	9.2520870	8.9849466	10.2849606

58	7.8256992	7.2983582	5.0959934	7.2527604	7.1531146	6.7704342	8.0630840
59	10.0012027	9.4804520	7.2983582	9.3963798	9.3353830	8.9922839	10.2867596
60	10.5199357	10.0012027	7.8256992	9.9073764	9.8562505	9.5258135	10.8211849
	8	9	10	11	12	13	14
2							
3							
4							
5							
6							
7							
8							
9	2.0253304						
10	1.9500426	1.0150466					
11	0.8849645	1.3608377	1.0668674				
12	1.6638584	1.4683490	2.2147304	1.7117577			
13	0.4021596	1.7443394	1.5588104	0.4922940	1.6916400		
14	1.3094040	3.3168087	3.0547146	2.0381971	2.8493659	1.5814255	
15	0.2912104	1.8305843	1.6710456	0.6042555	1.6916629	0.1128316	1.4879616
16	0.6863472	2.4486675	2.0816721	1.1064254	2.3365401	0.7284426	0.9801896
17	1.0800378	1.0926648	1.5200201	0.9022798	0.8113239	0.9649896	2.3818457
18	2.4830111	3.7523267	4.1924512	3.2442891	2.3791669	2.8491936	2.5959687
19	1.0615627	2.9303093	2.5434871	1.5893679	2.7206466	1.1980432	0.5943214
20	0.3420840	1.7226512	1.7757229	0.7589544	1.3479719	0.4131601	1.6442479
21	1.9997287	3.7714380	3.9302855	2.8776354	2.6781133	2.4018726	1.6021298
22	0.5071798	2.5240465	2.4378928	1.3716002	1.9841147	0.8793075	0.8702479
23	0.2730339	2.2668756	2.1035431	1.0423542	1.9276994	0.5548081	1.0519791
24	2.7432663	4.7451099	4.6371918	3.5848034	3.9050333	3.0982601	1.6312285
25	1.7104049	0.5431308	0.5578393	0.9157531	1.6689397	1.3675721	2.9448021
26	0.9450248	1.9454880	1.3998006	0.6189960	2.2579340	0.6303452	1.6962706
27	1.6510348	0.9111142	1.7047018	1.4072886	0.5574616	1.5439930	2.9384028
28	0.8840564	1.8450273	2.2116017	1.3330876	0.8861373	1.0732094	1.9686485
29	1.1929859	2.3579705	1.7214521	1.0459604	2.6549352	0.9772156	1.5674927
30	1.4340562	2.2576792	2.7709749	1.9434843	0.9284690	1.6766012	2.2848002
31	8.6704112	10.6908534	10.4691974	9.4623655	9.8222523	8.9947271	7.4249606
32	8.1606832	10.1435040	9.6857010	8.8053230	9.6058693	8.4001489	6.8514974
33	8.6465526	10.6718499	10.3931852	9.4099485	9.8755639	8.9535657	7.3738425
34	9.7362581	11.7612234	11.4562817	10.4864022	10.9865534	10.0357625	8.4544840
35	8.4546182	10.4499197	10.0177693	9.1199054	9.8692383	8.7055833	7.1455416
36	9.8189801	11.8283138	11.4292744	10.5112315	11.1818290	10.0855056	8.5137475
37	5.6008088	7.6239231	7.3986850	6.3887441	6.8138282	5.9215654	4.3515485
38	7.9341988	9.9518669	9.5954659	8.6505847	9.2702190	8.2135586	6.6351379
39	7.6962513	9.7182610	9.3889445	8.4291750	9.0030729	7.9850198	6.4040901
40	6.9326238	8.9457700	8.7674614	7.7433798	8.0492451	7.2683621	5.7133870

41	8.5021942	10.5236483	10.1860590	9.2315723	9.8076683	8.7892559	7.2087002
42	7.3241234	9.3281459	8.9260535	8.0074997	8.7220519	7.5844604	6.0165907
43	7.8018810	9.7864868	9.6866881	8.6420921	8.7860125	8.1574284	6.6391845
44	9.2214773	11.2447656	10.9188302	9.9589765	10.5037893	9.5134809	7.9322371
45	7.8022897	9.8084011	9.4093105	8.4894408	9.1895289	8.0650387	6.4955079
46	8.3727669	10.3883145	10.0197498	9.0822463	9.7170559	8.6483979	7.0715645
47	6.8848803	8.9011565	8.5450137	7.5986131	8.2378062	7.1621377	5.5843478
48	8.4645486	10.4786316	10.1034556	9.1698796	9.8153596	8.7378708	7.1621377
49	9.5022834	11.5267728	11.2654992	10.2754551	10.6982170	9.8153596	8.2378062
50	8.9043023	10.9128137	10.5151787	9.5955114	10.2754551	9.1698796	7.5986131
51	9.8541803	11.8477227	11.4043716	10.5151787	11.2654992	10.1034556	8.5450137
52	10.2009996	12.2179651	11.8477227	10.9128137	11.5267728	10.4786316	8.9011565
53	8.1812666	10.2009996	9.8541803	8.9043023	9.5022834	8.4645486	6.8848803
54	9.5423661	11.5613418	11.2039092	10.2612171	10.8595340	9.8237768	8.2448332
55	8.2818900	10.2948433	9.9161960	8.9844684	9.6391409	8.5535712	6.9786476
56	8.8021905	10.7778935	10.3025875	9.4354207	10.2574131	9.0358413	7.4936458
57	8.9629348	10.9077088	10.3816696	9.5543458	10.4649427	9.1742757	7.6612728
58	6.7210517	8.7017258	8.2503258	7.3639400	8.1813403	6.9584542	5.4122575
59	8.9465076	10.9228044	10.4478117	9.5805130	10.3998330	9.1806590	7.6378683
60	9.4824680	11.4557918	10.9725665	10.1117407	10.9385151	9.7143458	8.1741366
	15	16	17	18	19	20	21

2
3
4
5
6
7
8
9
10
11
12
13
14
15

16	0.6771059						
17	0.9989302	1.6757203					
18	2.7537169	2.9469883	2.7236306				
19	1.1319072	0.4832951	2.1172306	2.9629547			
20	0.3670246	0.9891170	0.7392164	2.4877783	1.3967604		
21	2.2903225	2.2252939	2.6789167	1.1412090	2.0871356	2.1599474	
22	0.7676494	0.7077023	1.5302920	2.2392990	0.8350943	0.8067603	1.5764370
23	0.4494915	0.4690266	1.3508498	2.5320551	0.7931611	0.6147265	1.9255644

24	2.9907374	2.5987491	3.6972082	2.4728887	2.2184948	3.0236957	1.4016592
25	1.4691828	2.0219103	1.0323918	3.7516806	2.5044893	1.4608608	3.6064253
26	0.6972016	0.7188473	1.4709620	3.4264927	1.1516104	1.0434810	2.8535474
27	1.5811229	2.2580122	0.5823108	2.8848101	2.6974788	1.3089963	3.0477255
28	1.0230193	1.5680059	0.7616109	1.9836142	1.8925480	0.6606210	1.9433245
29	1.0105596	0.6938474	1.8829003	3.6130958	0.9765756	1.3769416	2.9188714
30	1.6163895	2.0914958	1.2597250	1.4953258	2.3425967	1.2639181	1.7764510
31	8.8939693	8.3882055	9.6544184	7.8737600	7.9282710	8.9687354	7.1617499
32	8.3185911	7.7005268	9.2309079	8.1437316	7.2172704	8.4952332	7.1710406
33	8.8565272	8.3200977	9.6581989	8.0249094	7.8503270	8.9558016	7.2442200
34	9.9404573	9.3908087	10.7565791	9.1416279	8.9170494	10.0491877	8.3604973
35	8.6213302	8.0136963	9.5190297	8.3325596	7.5305580	8.7863215	7.3899107
36	9.9977340	9.4051344	10.8723334	9.5121672	8.9231425	10.1457307	8.6341829
37	5.8212656	5.3170868	6.5989403	5.0598769	4.8610099	5.9035511	4.2024942
38	8.1226434	7.5474568	8.9783890	7.6026464	7.0684157	8.2567936	6.7156510
39	7.8919458	7.3296406	8.7315096	7.2948189	6.8534175	8.0150109	6.4271829
40	7.1651239	6.6870449	7.8986086	6.1085378	6.2391949	7.2232525	5.3857624
41	8.6966488	8.1307655	9.5382346	8.0730008	7.6534891	8.8214073	7.2232525
42	7.4978540	6.9012455	8.3837095	7.1902695	6.4190389	7.6534891	6.2391949
43	8.0503760	7.6192562	8.7157260	6.6864955	7.1902695	8.0730008	6.1085378
44	9.4198192	8.8598439	10.2513664	8.7157260	8.3837095	9.5382346	7.8986086
45	7.9779378	7.3832885	8.8598439	7.6192562	6.9012455	8.1307655	6.6870449
46	8.5583842	7.9779378	9.4198192	8.0503760	7.4978540	8.6966488	7.1651239
47	7.0715645	6.4955079	7.9322371	6.6391845	6.0165907	7.2087002	5.7133870
48	8.6483979	8.0650387	9.5134809	8.1574284	7.5844604	8.7892559	7.2683621
49	9.7170559	9.1895289	10.5037893	8.7860125	8.7220519	9.8076683	8.0492451
50	9.0822463	8.4894408	9.9589765	8.6420921	8.0074997	9.2315723	7.7433798
51	10.0197498	9.4093105	10.9188302	9.6866881	8.9260535	10.1860590	8.7674614
52	10.3883145	9.8084011	11.2447656	9.7864868	9.3281459	10.5236483	8.9457700
53	8.3727669	7.8022897	9.2214773	7.8018810	7.3241234	8.5021942	6.9326238
54	9.7325438	9.1579871	10.5830678	9.1185257	8.6787525	9.8636611	8.2761277
55	8.4644572	7.8793712	9.3323392	7.9979131	7.3985520	8.6072414	7.1000189
56	8.9557616	8.3321651	9.8744981	8.7987150	7.8491805	9.1378455	7.8279730
57	9.0996392	8.4592920	10.0412550	9.1175008	7.9789156	9.3022051	8.1104586
58	6.8771461	6.2589205	7.7928831	6.8105843	5.7756461	7.0563869	5.7978001
59	9.1004764	8.4771892	10.0185746	8.9328346	7.9941875	9.2820384	7.9657592
60	9.6347718	9.0092472	10.5552163	9.4640661	8.5264367	9.8183729	8.5019923
	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 0.3515215
24 2.2360905 2.5436841
25 2.2173607 1.9185828 4.4529330
26 1.2773747 0.9317604 3.3175579 1.4496506
27 2.0744559 1.9236843 4.1728039 1.1485836 2.0109702
28 1.1116275 1.1164063 3.0641833 1.7728880 1.7029492 1.1179591
29 1.3847781 1.0828243 3.1950657 1.8476857 0.4269785 2.4325469 2.0327135
30 1.5111592 1.6233073 3.0736349 2.2911924 2.3067699 1.3953036 0.6135677
31 8.1675889 8.4453290 5.9577307 10.3622925 9.0792529 10.1254762 9.0108607
32 7.7033942 7.8991465 5.7814511 9.7181878 8.3014547 9.7757720 8.7274490
33 8.1499055 8.4109689 5.9785904 10.3186421 8.9955844 10.1495651 9.0427706
34 9.2425956 9.4968405 7.0858953 11.3979674 10.0568147 11.2536128 10.1495651
35 7.9891565 8.1962222 6.0126146 10.0345658 8.6279423 10.0568147 8.9955844
36 9.3428245 9.5656950 7.2891284 11.4269452 10.0345658 11.3979674 10.3186421
37 5.0998775 5.3729837 2.9144080 7.2891284 6.0126146 7.0858953 5.9785904
38 7.4514911 7.6854803 5.3729837 9.5656950 8.1962222 9.4968405 8.4109689
39 7.2085479 7.4514911 5.0998775 9.3428245 7.9891565 9.2425956 8.1499055
40 6.4271829 6.7156510 4.2024942 8.6341829 7.3899107 8.3604973 7.2442200
41 8.0150109 8.2567936 5.9035511 10.1457307 8.7863215 10.0491877 8.9558016
42 6.8534175 7.0684157 4.8610099 8.9231425 7.5305580 8.9170494 7.8503270
43 7.2948189 7.6026464 5.0598769 9.5121672 8.3325596 9.1416279 8.0249094
44 8.7315096 8.9783890 6.5989403 10.8723334 9.5190297 10.7565791 9.6581989
45 7.3296406 7.5474568 5.3170868 9.4051344 8.0136963 9.3908087 8.3200977
46 7.8919458 8.1226434 5.8212656 9.9977340 8.6213302 9.9404573 8.8565272
47 6.4040901 6.6351379 4.3515485 8.5137475 7.1455416 8.4544840 7.3738425
48 7.9850198 8.2135586 5.9215654 10.0855056 8.7055833 10.0357625 8.9535657
49 9.0030729 9.2702190 6.8138282 11.1818290 9.8692383 10.9865534 9.8755639

50	8.4291750	8.6505847	6.3887441	10.5112315	9.1199054	10.4864022	9.4099485
51	9.3889445	9.5954659	7.3986850	11.4292744	10.0177693	11.4562817	10.3931852
52	9.7182610	9.9518669	7.6239231	11.8283138	10.4499197	11.7612234	10.6718499
53	7.6962513	7.9341988	5.6008088	9.8189801	8.4546182	9.7362581	8.6465526
54	9.0577868	9.2946884	6.9559851	11.1763280	9.8050423	11.0972133	10.0060294
55	7.8034330	8.0302632	5.7484122	9.9001494	8.5185657	9.8561537	8.7758658
56	8.3484952	8.5393357	6.4381774	10.3466749	8.9236257	10.4219361	9.3779487
57	8.5258490	8.6956567	6.7108405	10.4583718	9.0207415	10.5992736	9.5806210
58	6.2669047	6.4587590	4.3984228	8.2771400	6.8632199	8.3405300	7.3002412
59	8.4923340	8.6837865	6.5771108	10.4918216	9.0688656	10.5656272	9.5206766
60	9.0295202	9.2192817	7.1145278	11.0222025	9.5964995	11.1030807	10.0592033
	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	2.6169011		
31	8.8343344	8.9374318	
32	7.9661914	8.8343344	2.6169011

33	8.7274490	9.0108607	0.6135677	2.0327135			
34	9.7757720	10.1254762	1.3953036	2.4325469	1.1179591		
35	8.3014547	9.0792529	2.3067699	0.4269785	1.7029492	2.0109702	
36	9.7181878	10.3622925	2.2911924	1.8476857	1.7728880	1.1485836	1.4496506
37	5.7814511	5.9577307	3.0736349	3.1950657	3.0641833	4.1728039	3.3175579
38	7.8991465	8.4453290	1.6233073	1.0828243	1.1164063	1.9236843	0.9317604
39	7.7033942	8.1675889	1.5111592	1.3847781	1.1116275	2.0744559	1.2773747
40	7.1710406	7.1617499	1.7764510	2.9188714	1.9433245	3.0477255	2.8535474
41	8.4952332	8.9687354	1.2639181	1.3769416	0.6606210	1.3089963	1.0434810
42	7.2172704	7.9282710	2.3425967	0.9765756	1.8925480	2.6974788	1.1516104
43	8.1437316	7.8737600	1.4953258	3.6130958	1.9836142	2.8848101	3.4264927
44	9.2309079	9.6544184	1.2597250	1.8829003	0.7616109	0.5823108	1.4709620
45	7.7005268	8.3882055	2.0914958	0.6938474	1.5680059	2.2580122	0.7188473
46	8.3185911	8.8939693	1.6163895	1.0105596	1.0230193	1.5811229	0.6972016
47	6.8514974	7.4249606	2.2848002	1.5674927	1.9686485	2.9384028	1.6962706
48	8.4001489	8.9947271	1.6766012	0.9772156	1.0732094	1.5439930	0.6303452
49	9.6058693	9.8222523	0.9284690	2.6549352	0.8861373	0.5574616	2.2579340
50	8.8053230	9.4623655	1.9434843	1.0459604	1.3330876	1.4072886	0.6189960
51	9.6857010	10.4691974	2.7709749	1.7214521	2.2116017	1.7047018	1.3998006
52	10.1435040	10.6908534	2.2576792	2.3579705	1.8450273	0.9111142	1.9454880
53	8.1606832	8.6704112	1.4340562	1.1929859	0.8840564	1.6510348	0.9450248
54	9.5043595	10.0219356	1.7296542	1.8641580	1.2347449	0.6683118	1.4372971
55	8.2121748	8.8219758	1.7405297	0.8777023	1.1560367	1.7209645	0.5985665
56	8.5815685	9.4894837	2.8129247	0.6569336	2.2000158	2.2787231	0.5810717
57	8.6617661	9.7283875	3.4302339	1.0757755	2.8171787	2.8270305	1.1698691
58	6.5325668	7.4264470	3.0995197	1.4417939	2.6854465	3.4951544	1.7708333
59	8.7267817	9.6302152	2.8302209	0.7960053	2.2166934	2.2233178	0.6689769
60	9.2511939	10.1687883	3.0631760	1.3344541	2.4613424	2.2115871	1.1521306
	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 4.4529330
 38 1.9185828 2.5436841
 39 2.2173607 2.2360905 0.3515215
 40 3.6064253 1.4016592 1.9255644 1.5764370
 41 1.4608608 3.0236957 0.6147265 0.8067603 2.1599474
 42 2.5044893 2.2184948 0.7931611 0.8350943 2.0871356 1.3967604
 43 3.7516806 2.4728887 2.5320551 2.2392990 1.1412090 2.4877783 2.9629547
 44 1.0323918 3.6972082 1.3508498 1.5302920 2.6789167 0.7392164 2.1172306
 45 2.0219103 2.5987491 0.4690266 0.7077023 2.2252939 0.9891170 0.4832951
 46 1.4691828 2.9907374 0.4494915 0.7676494 2.2903225 0.3670246 1.1319072
 47 2.9448021 1.6312285 1.0519791 0.8702479 1.6021298 1.6442479 0.5943214
 48 1.3675721 3.0982601 0.5548081 0.8793075 2.4018726 0.4131601 1.1980432
 49 1.6689397 3.9050333 1.9276994 1.9841147 2.6781133 1.3479719 2.7206466
 50 0.9157531 3.5848034 1.0423542 1.3716002 2.8776354 0.7589544 1.5893679
 51 0.5578393 4.6371918 2.1035431 2.4378928 3.9302855 1.7757229 2.5434871
 52 0.5431308 4.7451099 2.2668756 2.5240465 3.7714380 1.7226512 2.9303093
 53 1.7104049 2.7432663 0.2730339 0.5071798 1.9997287 0.3420840 1.0615627
 54 0.5647217 4.0754223 1.6106351 1.8564887 3.1259741 1.0532664 2.3091226
 55 1.5428167 2.9416690 0.4061515 0.7515774 2.3100982 0.5077451 1.0086202
 56 1.4270700 3.8295979 1.5127190 1.8570388 3.4322506 1.5584495 1.6176175
 57 1.8487194 4.2438965 2.0853113 2.4179907 3.9797539 2.1752871 2.0348538
 58 3.2061519 2.1199808 1.5970006 1.5935438 2.4836821 2.1997187 0.8041821

59	1.3212552	3.9532875	1.5957195	1.9440794	3.5204913	1.5891219	1.7468956
60	1.1305840	4.4673393	2.0357790	2.3872107	3.9529734	1.8973159	2.2742573
	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	2.7236306							
45	2.9469883	1.6757203						
46	2.7537169	0.9989302	0.6771059					
47	2.5959687	2.3818457	0.9801896	1.4879616				
48	2.8491936	0.9649896	0.7284426	0.1128316	1.5814255			
49	2.3791669	0.8113239	2.3365401	1.6916629	2.8493659	1.6916400		
50	3.2442891	0.9022798	1.1064254	0.6042555	2.0381971	0.4922940	1.7117577	
51	4.1924512	1.5200201	2.0816721	1.6710456	3.0547146	1.5588104	2.2147304	
52	3.7523267	1.0926648	2.4486675	1.8305843	3.3168087	1.7443394	1.4683490	
53	2.4830111	1.0800378	0.6863472	0.2912104	1.3094040	0.4021596	1.6638584	
54	3.2004145	0.4770888	1.8348099	1.1865027	2.6626045	1.1112948	1.1329707	
55	2.8296640	1.1394547	0.5416464	0.1465967	1.4020829	0.1894812	1.8382394	
56	3.9958015	1.8145226	1.2531066	1.2455823	2.1983737	1.1587487	2.6257584	
57	4.5963338	2.4016520	1.7595066	1.8555991	2.6291747	1.7727960	3.2113236	
58	3.5085212	2.9134398	1.2445468	1.9187678	0.9136363	1.9728033	3.5236750	
59	4.0502511	1.7844467	1.3659520	1.2966574	2.3226905	1.2023045	2.5947241	
60	4.3848654	1.8891468	1.8694426	1.6718779	2.8406916	1.5641852	2.6658125	
	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
 52
 53
 54
 55
 56
 57
 58
 59
 60
 2
 3
 4
 5
 6
 7

1.0668674							
1.3608377	1.0150466						
0.8849645	1.9500426	2.0253304					
0.8184429	1.0873411	0.6697024	1.3616007				
0.6444822	1.6991725	1.9315897	0.3468048	1.3007743			
0.9173073	1.1548227	1.9687861	1.5133386	1.6300241	1.1675051		
1.5192976	1.4242917	2.3839810	2.1144623	2.1624766	1.7676931	0.6174304	
2.3084519	3.1545640	3.6684278	1.8657051	3.0741154	1.7861007	2.0816574	
0.9035729	1.0189078	1.8643567	1.5741324	1.5649991	1.2323936	0.1452447	
1.1415501	0.6511228	1.6440675	1.9623455	1.5479148	1.6429408	0.6813359	
	57	58	59				

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 2.3126622
59 0.6177529 2.2256934
60 0.7790038 2.7626156 0.5388040
```

```
clustering <- hclust( dist(x))
clustering
```

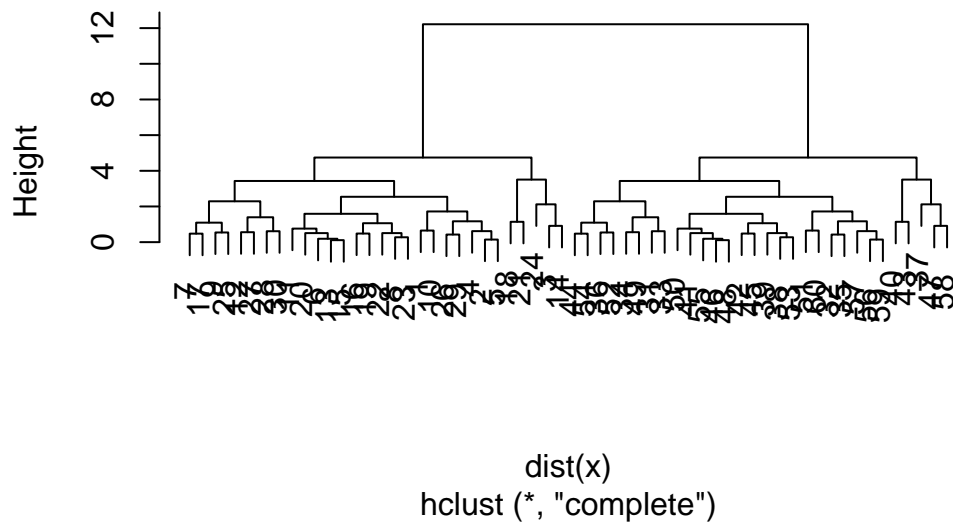
Call:

```
hclust(d = dist(x))
```

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

```
plot(clustering)
```

Cluster Dendrogram

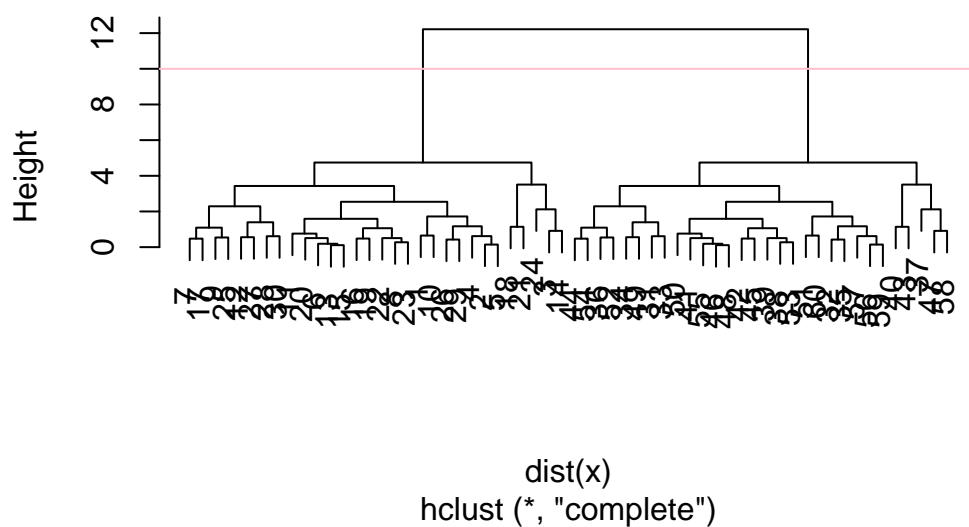


group in the left from 1-30 and group to the right from 30-60

Let's add horizontal line:

```
plot(clustering)
abline(h = 10, col = 'pink')
```

Cluster Dendrogram



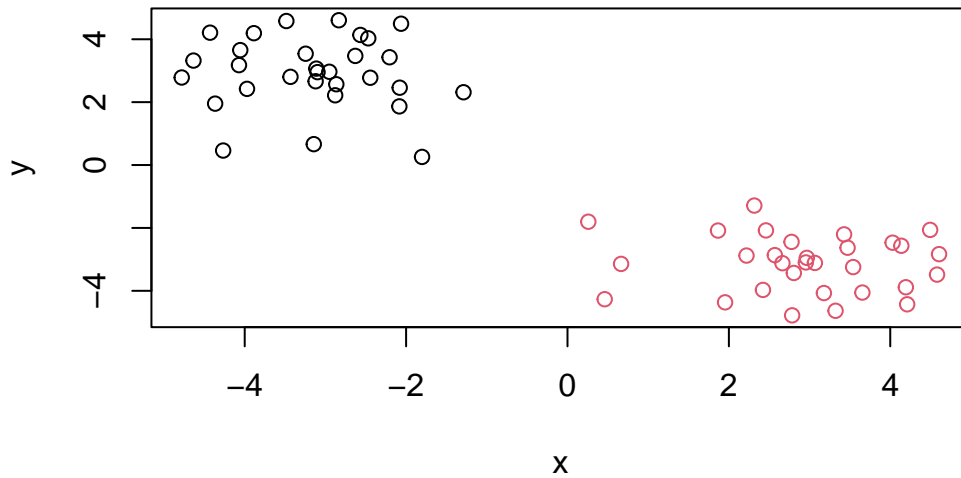
to get our results (i.e., membership vector) we need to “cut” the tree. The function for doing that is `cutree()`

```
subgroups <- cutree(clustering, h = 10)
subgroups
```

[illegible]

plotting this..

```
plot(x, col = subgroups)
```



You can also “cut” your tree with the number of clusters you want:

```
cutree(clustering, h =2)
```

```
[1] 1 1 2 1 1 3 4 3 4 1 3 5 3 2 3 3 4 6 3 3 6 3 3 7 4
[26] 1 5 5 1 5 8 9 8 8 9 10 11 12 12 13 12 12 13 10 12 12 14 12 8 12
[51] 9 10 12 10 12 9 9 14 9 9
```

Principle Component Analysis (PCA)

PCA of UK food

First we read the data:

```
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

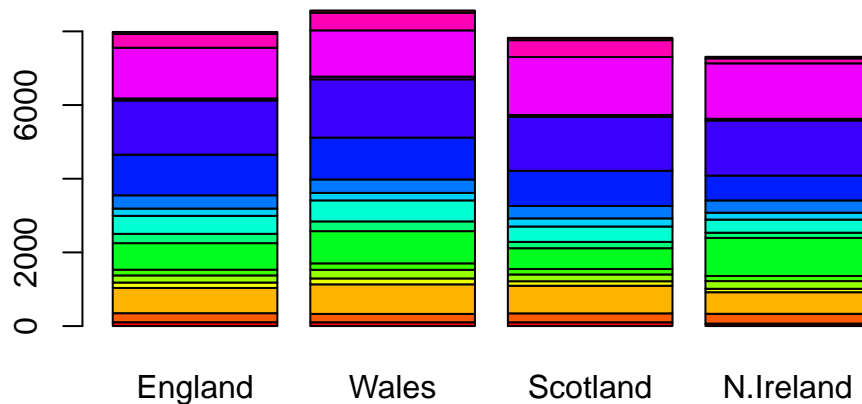
Carcass_meat	245	227	242	267
Other_meat	685	803	750	586
Fish	147	160	122	93
Fats_and_oils	193	235	184	209
Sugars	156	175	147	139

Now we can generate some basic visualization

```
rainbow (nrow(x))
```

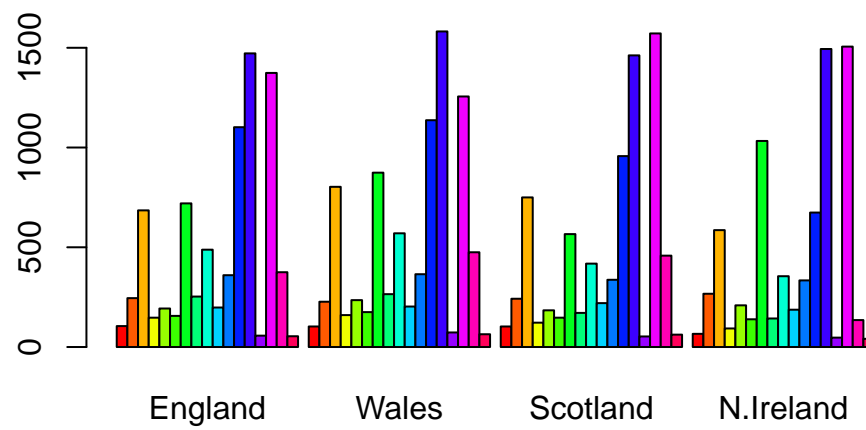
```
[1] "#FF0000" "#FF5A00" "#FFB400" "#F0FF00" "#96FF00" "#3CFF00" "#00FF1E"
[8] "#00FF78" "#00FFD2" "#00D2FF" "#0078FF" "#001EFF" "#3C00FF" "#9600FF"
[15] "#F000FF" "#FF00B4" "#FF005A"
```

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



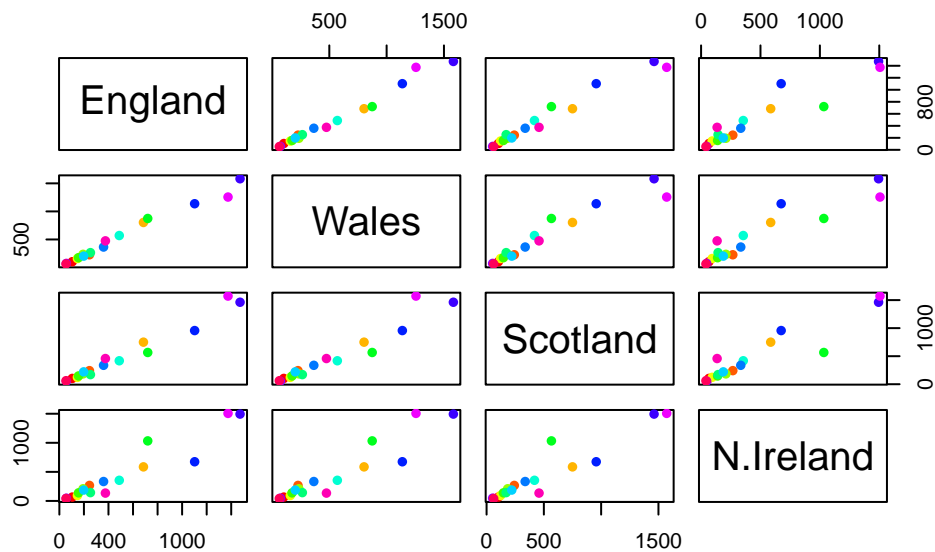
Let's refine our barplot

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



Other useful visualization:

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



Let's apply PCA. For that we need to use the command `prcomp()`. This function expects the transpose of our data

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

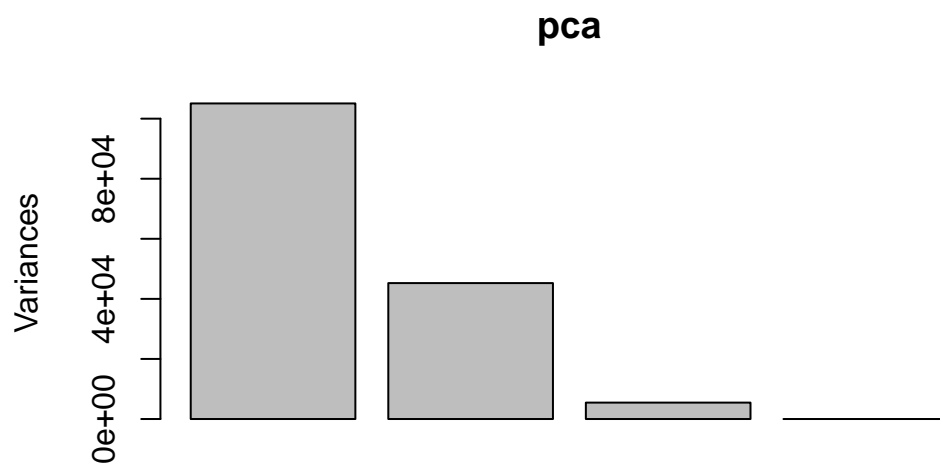
pca <- prcomp ( t(x))
summary(pca)
```

Importance of components:

	PC1	PC2	PC3	PC4
Standard deviation	324.1502	212.7478	73.87622	4.189e-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 plot the PCA results:

```
plot(pca)
```



We need to access the results of PCA analysis

```
attributes(pca)
```

\$names

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

\$class

```
[1] "prcomp"
```

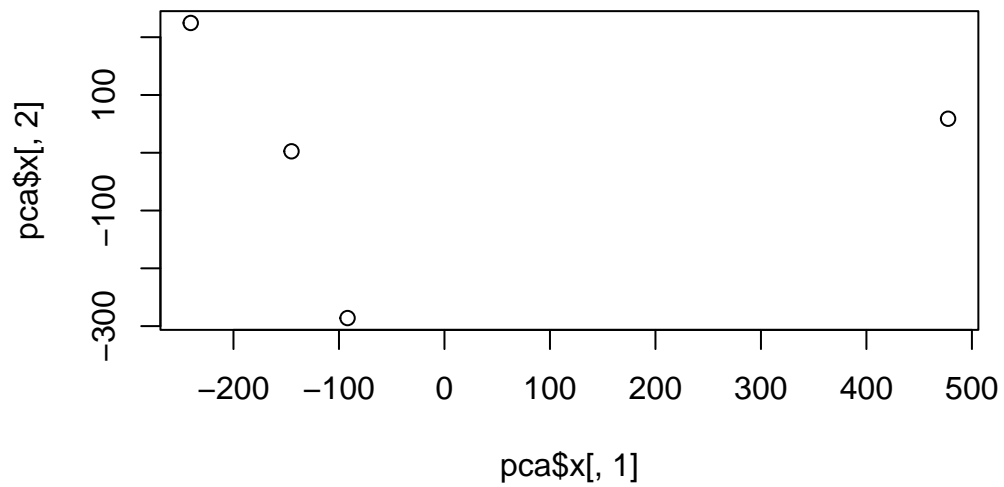
We can explore the `pca$x` dataframe:

```
pca$x
```

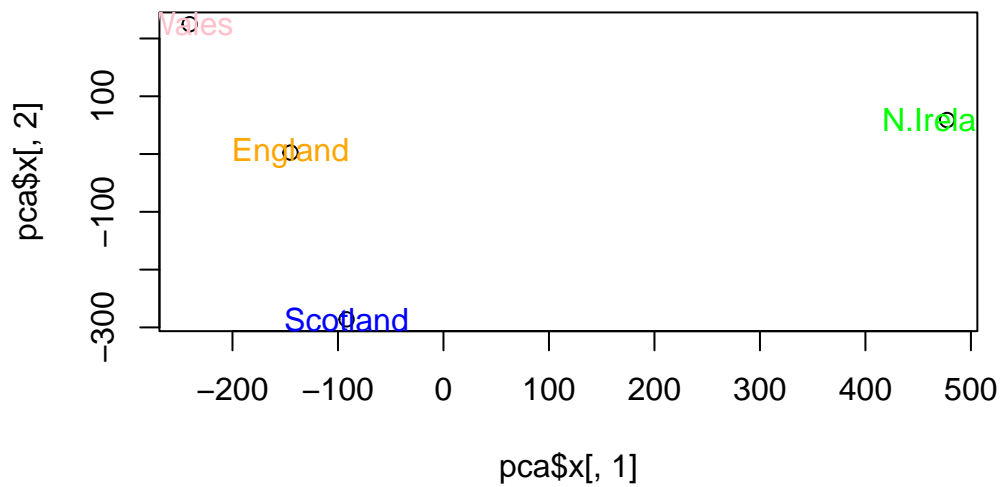
	PC1	PC2	PC3	PC4
England	-144.99315	2.532999	-105.768945	2.842865e-14
Wales	-240.52915	224.646925	56.475555	7.804382e-13
Scotland	-91.86934	-286.081786	44.415495	-9.614462e-13
N.Ireland	477.39164	58.901862	4.877895	1.448078e-13

Plotting:


```
plot( x=pca$x[,1], y=pca$x[,2])
```



```
plot( x=pca$x[,1], y=pca$x[,2])  
color_countries <- c('orange', 'pink', 'blue', 'green')  
text( x=pca$x[,1], y=pca$x[,2], colnames(x), col = color_countries)
```



PCA of a RNA-seq data set

First step as always to load the data:

```
url2 <- "https://tinyurl.com/expression-CSV"
rna.data <- read.csv(url2, row.names=1)
head(rna.data)
```

	wt1	wt2	wt3	wt4	wt5	ko1	ko2	ko3	ko4	ko5
gene1	439	458	408	429	420	90	88	86	90	93
gene2	219	200	204	210	187	427	423	434	433	426
gene3	1006	989	1030	1017	973	252	237	238	226	210
gene4	783	792	829	856	760	849	856	835	885	894
gene5	181	249	204	244	225	277	305	272	270	279
gene6	460	502	491	491	493	612	594	577	618	638

Q. How many genes and samples are in this data set?

```
dim(rna.data)
```

```
[1] 100 10
```

100 genes and 10 samples

Let's apply PCA:

```
pca_rna = prcomp (t(rna.data))  
summary(pca_rna)
```

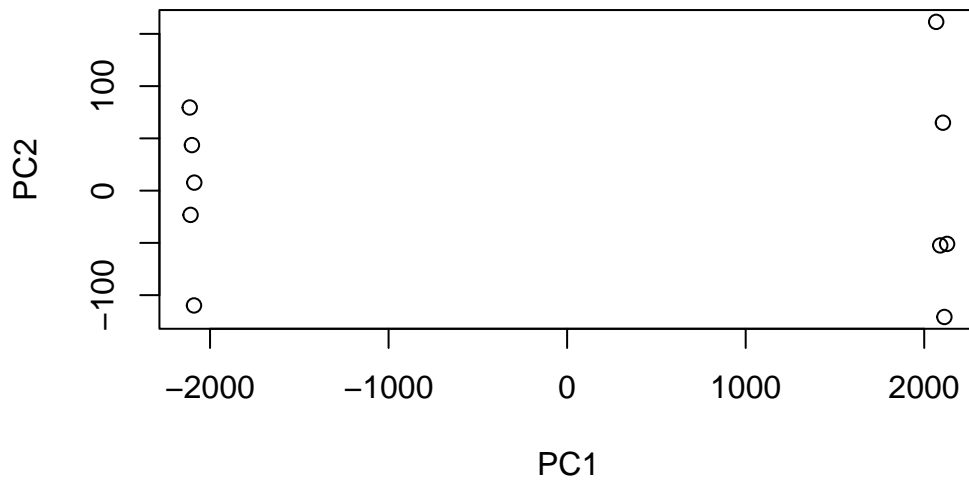
Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	2214.2633	88.9209	84.33908	77.74094	69.66341	67.78516
Proportion of Variance	0.9917	0.0016	0.00144	0.00122	0.00098	0.00093
Cumulative Proportion	0.9917	0.9933	0.99471	0.99593	0.99691	0.99784

	PC7	PC8	PC9	PC10
Standard deviation	65.29428	59.90981	53.20803	3.142e-13
Proportion of Variance	0.00086	0.00073	0.00057	0.000e+00
Cumulative Proportion	0.99870	0.99943	1.00000	1.000e+00

Let's plot the principle component 1 and 2 :

```
plot(pca_rna$x[,1], pca_rna$x[,2], xlab='PC1', ylab='PC2')
```



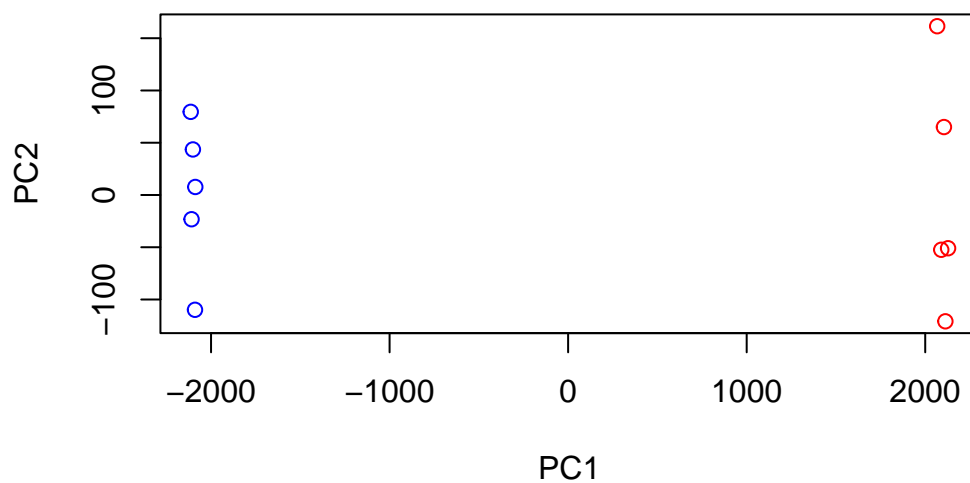
```
colnames(rna.data)
```

```
[1] "wt1" "wt2" "wt3" "wt4" "wt5" "ko1" "ko2" "ko3" "ko4" "ko5"
```

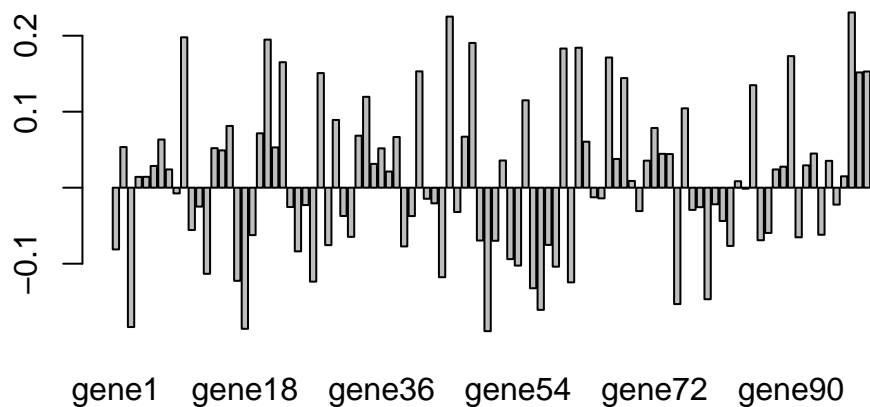
```
cols_samples <- c(rep('blue', 5), rep('red',5))  
cols_samples
```

```
[1] "blue" "blue" "blue" "blue" "blue" "red" "red" "red" "red" "red"
```

```
plot(pca_rna$x[,1], pca_rna$x[,2], xlab='PC1', ylab='PC2', col = cols_samples)
```



```
barplot(pca_rna$rotation[,1])
```



```
sort(pca_rna$rotation[,1])
```

gene50	gene18	gene3	gene57	gene75	gene79
-0.188796985	-0.185668500	-0.183374164	-0.160771014	-0.153164404	-0.146803635
gene56	gene61	gene27	gene17	gene44	gene13
-0.132330117	-0.124572881	-0.123615228	-0.122536548	-0.117808971	-0.113357525
gene59	gene54	gene53	gene25	gene1	gene39
-0.103935563	-0.102503320	-0.093979884	-0.083761992	-0.081247810	-0.077306742
gene82	gene29	gene58	gene51	gene49	gene86
-0.076658760	-0.075605635	-0.075274651	-0.069855142	-0.069530208	-0.069165267
gene91	gene32	gene19	gene94	gene87	gene11
-0.065288752	-0.064721235	-0.062411218	-0.061938300	-0.059547317	-0.055698801
gene81	gene40	gene31	gene46	gene70	gene77
-0.043780416	-0.037323670	-0.037219970	-0.031990529	-0.030784982	-0.029225446
gene78	gene24	gene12	gene26	gene96	gene80
-0.025639741	-0.025407507	-0.024870802	-0.022868107	-0.022293151	-0.021824860
gene43	gene42	gene65	gene64	gene9	gene84
-0.020617052	-0.014550791	-0.014052839	-0.012639567	-0.007495075	-0.001289937
gene83	gene69	gene4	gene5	gene97	gene37
0.008504287	0.008871890	0.014242602	0.014303808	0.014994546	0.021280555
gene88	gene8	gene89	gene6	gene92	gene35
0.024015925	0.024026657	0.027652967	0.028634131	0.029394259	0.031349942

gene95	gene71	gene52	gene67	gene74	gene73
0.035342407	0.035589259	0.035802086	0.037840851	0.044286948	0.044581700
gene93	gene15	gene36	gene14	gene22	gene2
0.044940861	0.049090676	0.051765605	0.052004194	0.053013523	0.053465569
gene63	gene7	gene38	gene47	gene33	gene20
0.060529157	0.063389255	0.066665407	0.067141911	0.068437703	0.071571203
gene72	gene16	gene30	gene76	gene55	gene34
0.078551648	0.081254592	0.089150461	0.104435777	0.114988217	0.119604059
gene85	gene68	gene28	gene99	gene100	gene41
0.134907896	0.144227333	0.150812015	0.151678253	0.152877246	0.153077075
gene23	gene66	gene90	gene60	gene62	gene48
0.165155192	0.171311307	0.173156806	0.183139926	0.184203008	0.190495289
gene21	gene10	gene45	gene98		
0.194884023	0.197905454	0.225149201	0.230633225		