

class7 lab (Hands on with Principal Component Analysis (PCA))

Josef Urrete

1. PCA of UK Food data

```
url <- "https://tinyurl.com/UK-foods"
x <- read.csv(url)
```

Q1. How many rows and columns are in dataframe x? What R functions could you use to answer these questions?

```
#use nrow and ncol or dim()
dim(x)
```

```
[1] 17  5
```

```
#to preview x
##View(x)
```

```
#how minus indexing works
rownames(x) <- x[,1]
x <- x[,-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

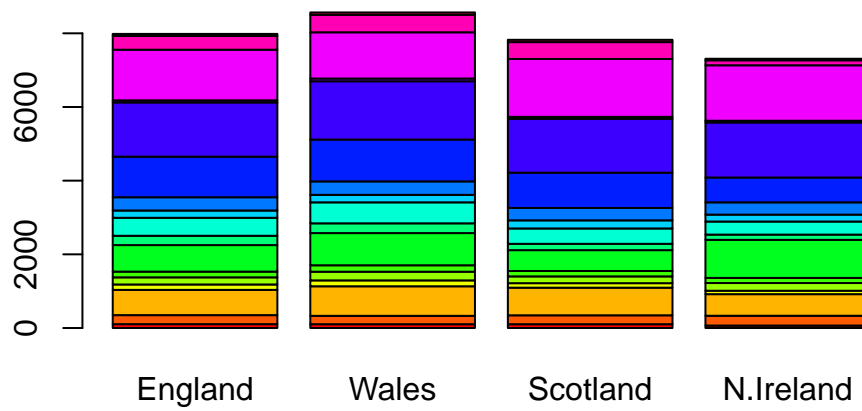
```
x <- read.csv(url, row.names = 1)
head(x)
```

	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

Q2: Which approach to solving the “row-names problem” mentioned above do you prefer and why? Is one approach more robust than another under certain circumstances?

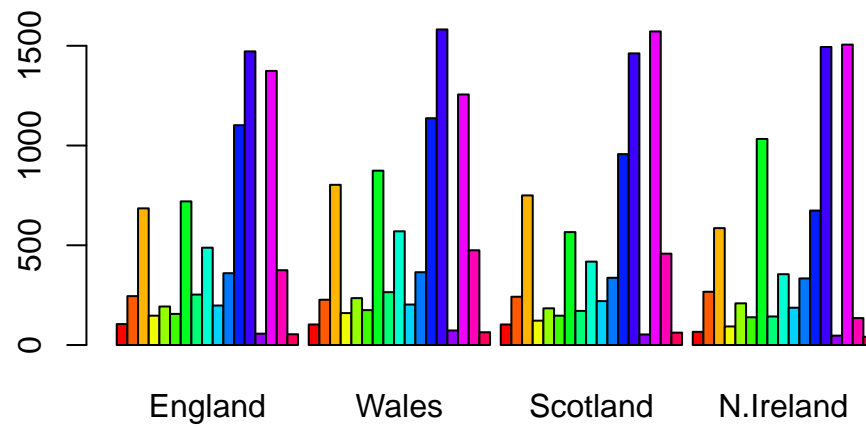
Making basic visualizations

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

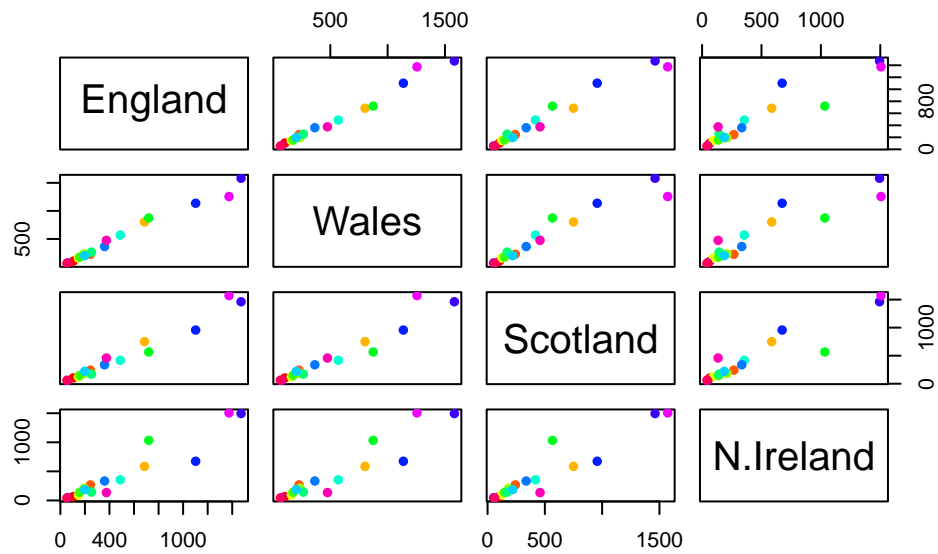


Q3: Changing optional argument in barplot

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



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



Applying PCA

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

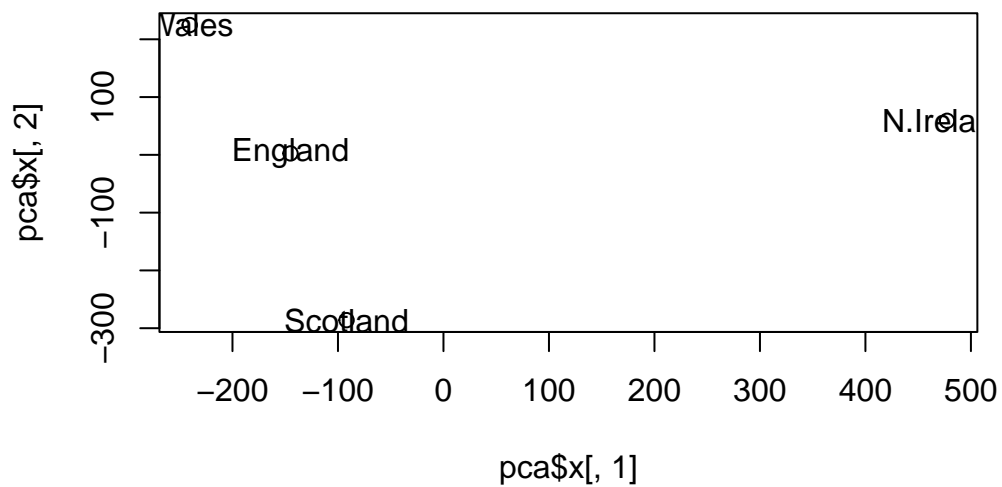
```
$names  
[1] "sdev"      "rotation" "center"    "scale"     "x"
```

```
$class  
[1] "prcomp"
```

Q7: complete the code below to generate a plot of PC1 vs PC2. Second line adds text labels over the data points.

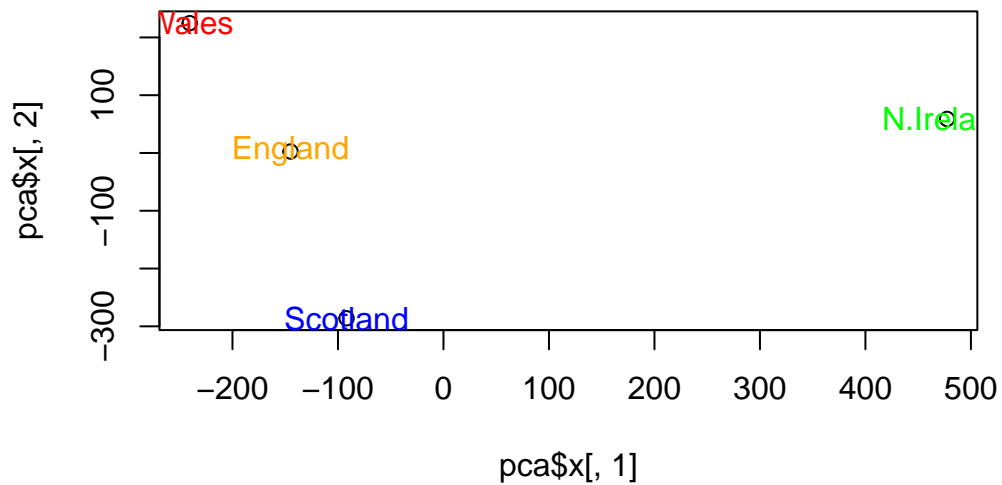
using `pca$x`

```
plot(pca$x[,1], pca$x[,2])  
text(pca$x[,1], pca$x[,2], colnames(x))
```



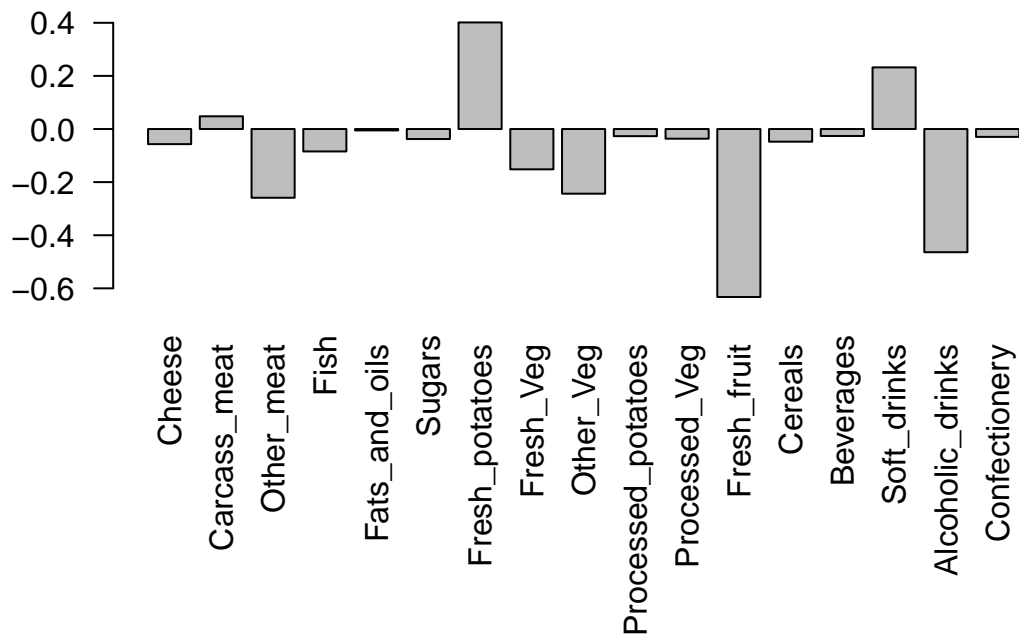
Q8: Customize plot so colors of country names match colors in original map and table

```
country_cols <- c("orange", "red", "blue", "green")
plot(pca$x[,1],pca$x[,2])
text(pca$x[,1], pca$x[,2], colnames(x), col=country_cols)
```



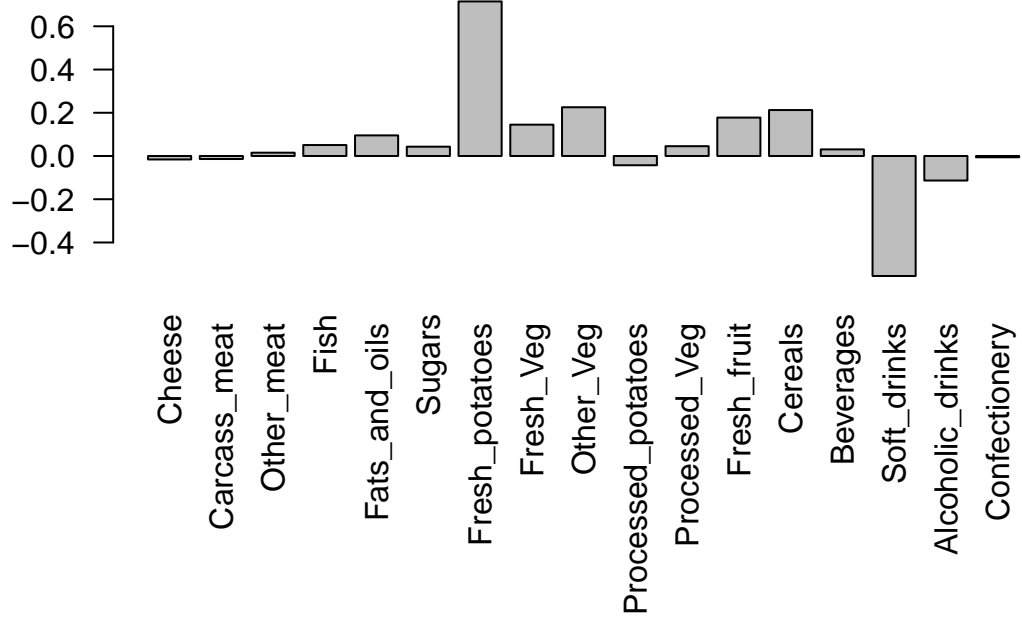
PC1

```
par(mar=c(10,3,0.35,0))
barplot(pca$rotation[,1], las=2)
```

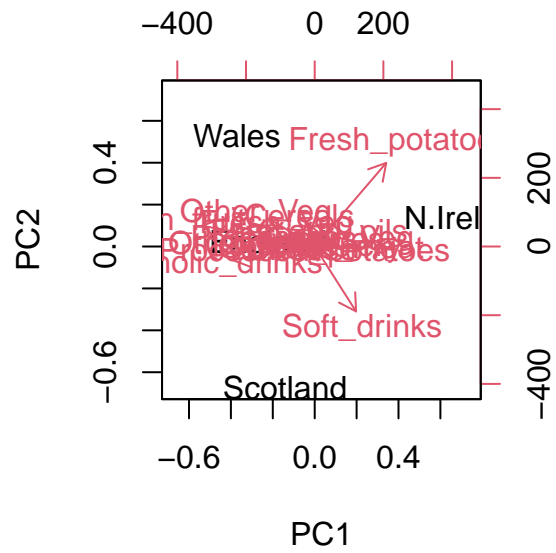


Q9: Generate similar loadings plot for PC2. What two food groups feature prominently and what does PC2 mainly tell us about?

```
par(mar=c(10, 3, 0.35, 0))
barplot(pca$rotation[,2], las=2)
```



```
biplot(pca)
```




```
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
```

Q10: how many genes and samples are in this ddata set?

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

```
[1] 100 10
```

100 genes and 10 samples

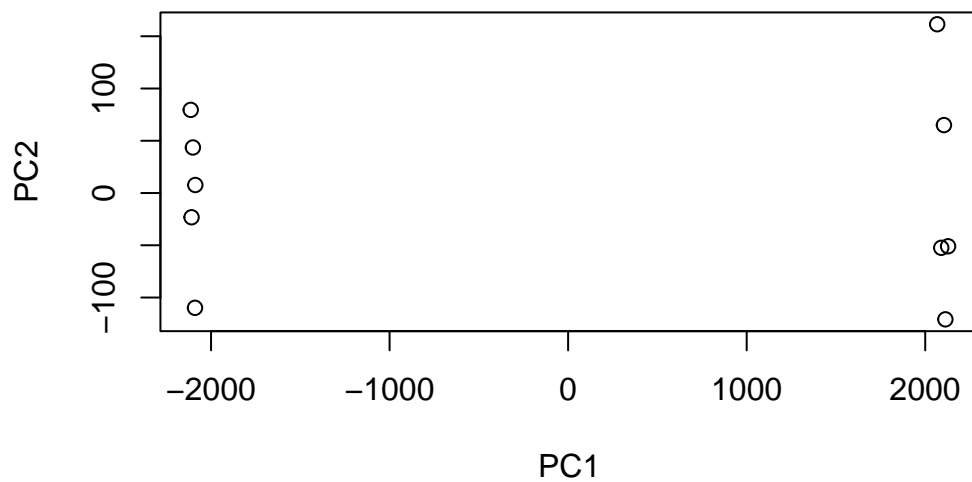
```
pca <- prcomp(t(rna.data))
summary(pca)
```

Importance of components:

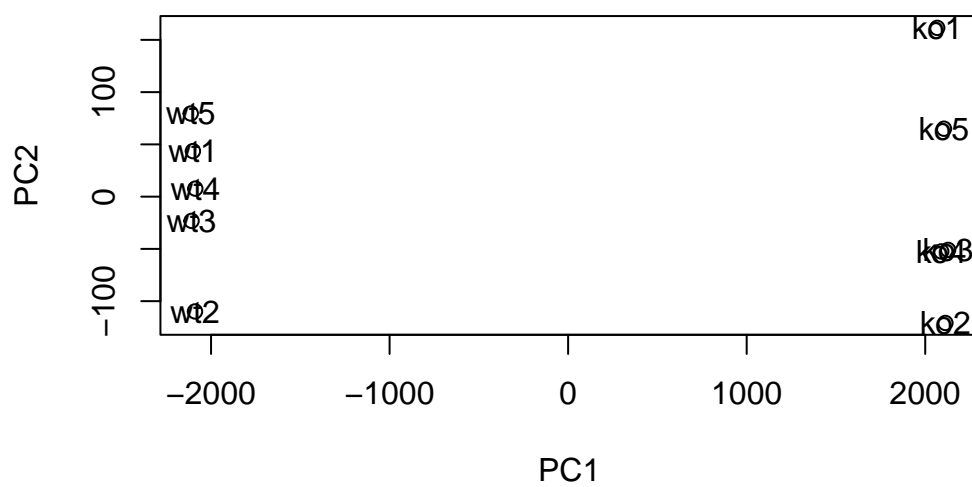
	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	2214.2633	88.9209	84.33908	77.74094	69.66341	67.78516
Proportion of Variance	0.9917	0.0016	0.00144	0.00122	0.00098	0.00093
Cumulative Proportion	0.9917	0.9933	0.99471	0.99593	0.99691	0.99784

	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

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



```
plot(pca$x[,1], pca$x[,2], xlab="PC1", ylab="PC2")
text(pca$x[,1], pca$x[,2], colnames(rna.data))
```



```
sort(pca$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		