````
---
title: "Class07 Lab"
author: "Nicole Alfonso"
format: html
editor: visual
---

## Hands on with Principle Component Analysis (PCA)

First, we need to import our data.

> **Q1**. How many rows and columns are in your new data frame named `x`? What R functions could you use to answer this questions?

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

nrow(x)
ncol(x)
```

Checking your data. It is always a good idea to examine your imported data to make sure it meets your expectations.

```{r}
#To view the entire data frame
View(x)
#To view the first 6 rows of the data frame
head(x)
#To view the last 6 rows of the data frame
tail(x)
```

It appears that the data is not set properly, as the first column is labeled as 'X', giving us 5 variables not 4. To fix this we use the function rownames().

```{r}
#To class for the first column
rownames(x) <- x[,1]
#To remove the first column
x <- x[,-1]
head(x)
```

Another way to do it is by calling read.csv()

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

```{r}
#To find out the dimensions (x, y) of the data frame: dim()
dim(x)
```

> **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?

I think the solution for the 'row-names problem' that I prefer is the: 'x \<- read.csv(url, row.names=1)' approach, as you have more control as to which column that you are changing. I think using the x\[,-1\] method would work if you only had to adjust the first column, because you might continue to erase more variables.
````

> Spotting major differences and trends

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