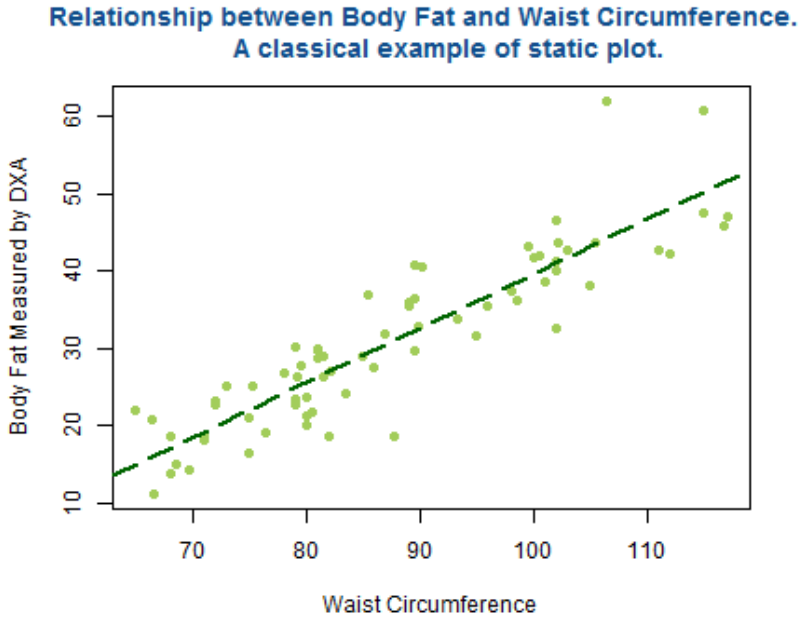
Body Fat Clustering

Project for the Course in "Developing Data Products"

Sofia Cividini - BSc, MSc, MSc, MD
Molecular Biologist specialized in Biostatistics and Bioinformatics

Introduction.

The goal of this application was to create an interactive graph. In fact, through this application, an user can group, in a certain number of clusters, the observations in the data set called bodyfat according to the variables used on the y-axis and on the x-axis of the graph itself.



Source: "Garcia et al. (2005), Improved prediction of body fat by measuring skinfold thickness, circumferences, and bone breadths. Obesity Research, 13(3), 626-634".

Main Characteristics of the data set.

The body fat dataset (bodyfat) is included in the R library called TH.data. It contains body fat measurements and anthropometric measurements which were obtained for 71 healthy German women. The 10 collected variables are the following ones:

VARIABLES	EXPLANATION
DEXfat	Body fat measured by DXA, response variable
age	Age in years
waistcirc	Waist circumference
hipcirc	Hip circumference
elbowbreadth	Breadth of the elbow
kneebreadth	Breadth of the knee.
anthro3a	Sum of logarithm of three anthropometric measurements.
anthro3b	Sum of logarithm of three anthropometric measurements.
anthro3c	Sum of logarithm of three anthropometric measurements.
anthro4	Sum of logarithm of three anthropometric measurements.

My application.

Body Fat Clustering

Y Variable

DEXfat

X Variable

waistcm

Cluster count

4

An Interactive Graph for the Body Fat Clustering.

Project for the course in 'Developing Data Products' - Data Science Specialization, JHU.

Student: Sofia Cividini

This application is able to produce an interactive graph that shows different types of clusters based on the selection of the following variables on the y-axis and on the x-axis. It is also possible to change the number of clusters from a minimum of 2 to a maximum of 7.

DEXfat: body fat measured by DXA (Dual Energy X-Ray Absorptiometry)

age: age in years

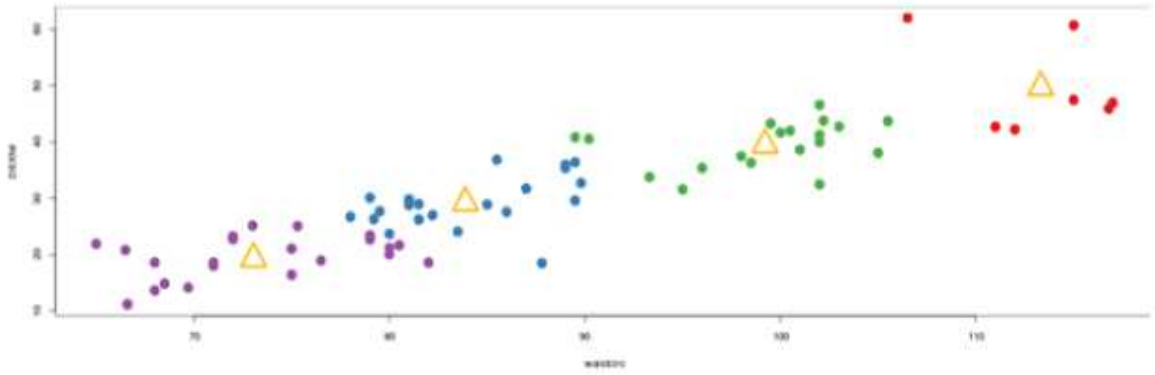
waistcm: waist circumference

hipcm: hip circumference

elbowbreadth: breadth of the elbow

kneebreadth: breadth of the knee

These data come from the data set called bodyfat which is available in the R library called TH.data. This data set contains the data collected from 71 healthy female subjects on 10 variables, both body fat measurements and anthropometric measurements.



Conclusion.

This is a funny application which allows to see how the several observations are grouped in different clusters according to the combined use of 6 variables out of 10 from the original data set.
For example, it is possible to keep fixed the variable corresponding to the body fat (DEXfat) on the y-axis and see how the observations are put in different clusters according to the values of the waist circumference or to the values of the hip circumference on the x-axis.

In the figure, there are the main three different phenotypes of body fat: low, medium and 'high'.