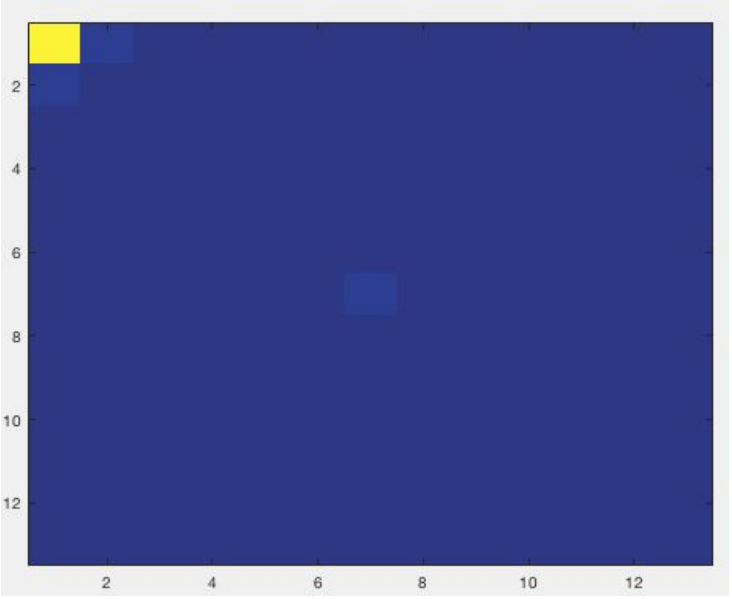
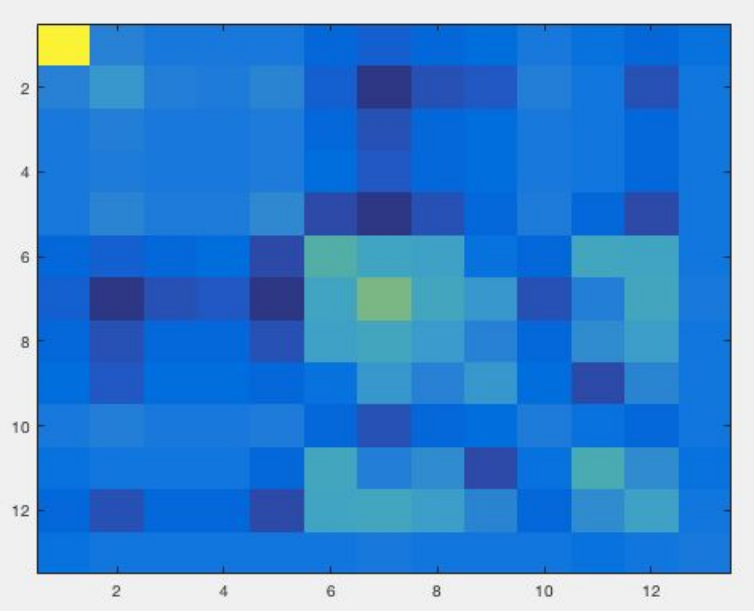
• Print the covariance matrix (the functions image() and imagesc() may be useful here). How should this matrix be interpreted? Specify the variables that have very high/low covariance.

Messure of relation.

Without normalization it's hard to make any conclusions at all. Because one cell in the matrix is so much bigger than the rest.



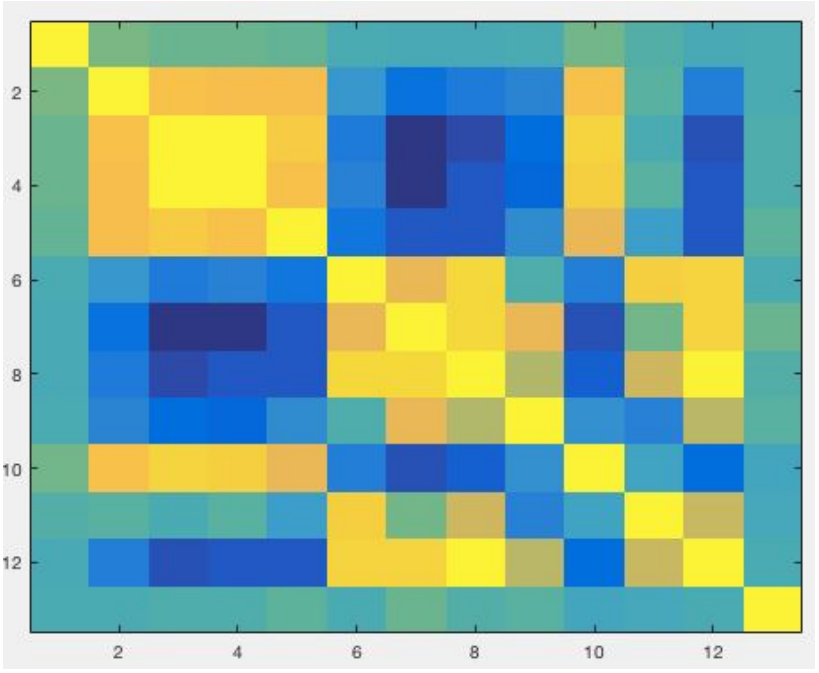
With normalization we can now start to be able to make some conclusions about the covariance between variables.



• Print the correlation matrix (the functions image() and imagesc() may be useful here). How should this matrix be interpreted? What is the difference to the covariance matrix? Specify the variables that are strongly correlated and the variables that are uncorrelated.

In this matrix we can see which variables correlate with each other. The more dark blue => less correlation. Yellow => more correlation.

The difference between covariance matrix and correlation matrix is normalized by dividing the covariance matrix by the following $\sqrt{\sigma_X * \sigma_Y}$. That way we get a much “stronger” normalization than just normalize ever variable by it self like we did before.



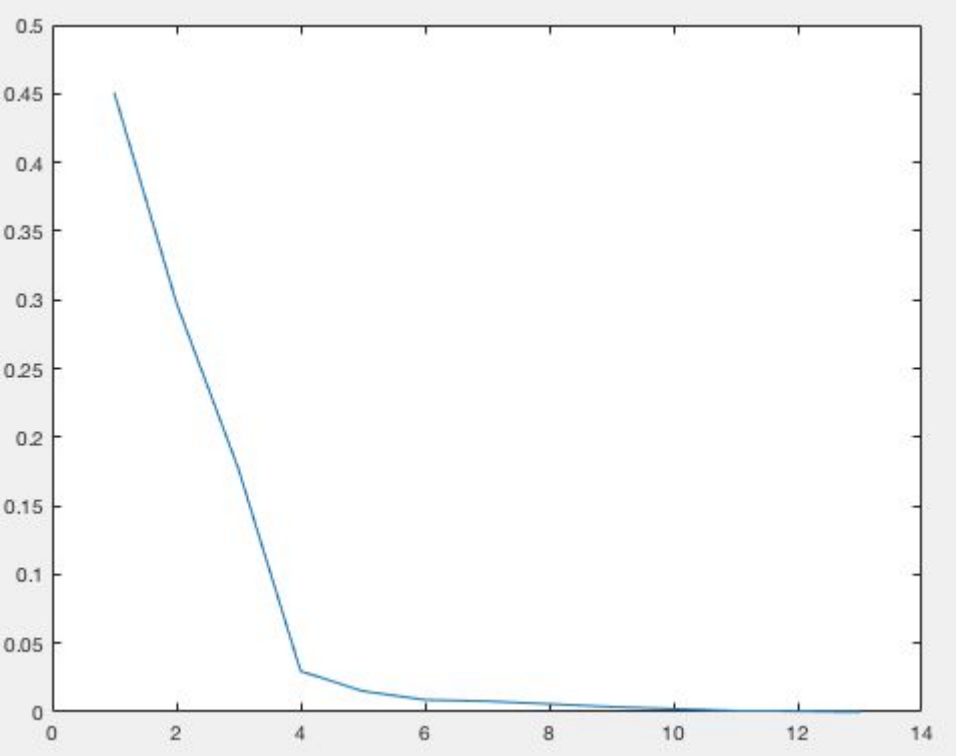
For instance, we can see that variable three (average lifespan of women) and four (average lifespan of men) have a strong correlation.

• Since the different variables are measured in different units it can be wise to normalize
each variable on its own before performing PCA. Why is that?

Because all variables have different scales and variance.

• When you do PCA you get eigenvalues. What do they mean? How can you tell from the eigenvalues how many dimensions you need in order to represent the data?

The eigenvalues represent how much of the variance the eigenvector dimentions represent of the data. By plotting the percentage of the eigenvalues we can easily see that over 90% of all variations in the data is represented with only 3 dimensions.



• Print the first two principal components against each other for every country. Colourize it according to the given classification. Is there a pattern? To what class should e.g. Georgia belong to according to your analysis?

Using the first two principal components, there seem to be a pattern. PC1 covers most of the variance and splits the data decently.

Georgia belongs to the *developing* category.

• If there are only two classes it is easy to use FLD to build a classifier, how?

Because you only need one line to separate the classes. With more classes you need a plane or hyper planes.

Maximize distance between classes and minimize distance within classes.

• If you only consider the industrialized and the developing world, how well does FLD separate them? Also, look at the discriminant function and see if you can extract which attributes that are important.

FLD manages to separate them quite well.

The most important attributes can be find by max(abs(w)).