Assignment 6

Task 1 - Multivariate Data

(a) Yes, there is a correlation between x and y, since the correlation coefficient is very high, there is a positive linear correlation. Surprisingly, the mean, variance and persons correlation coefficient are the same for all four groups.

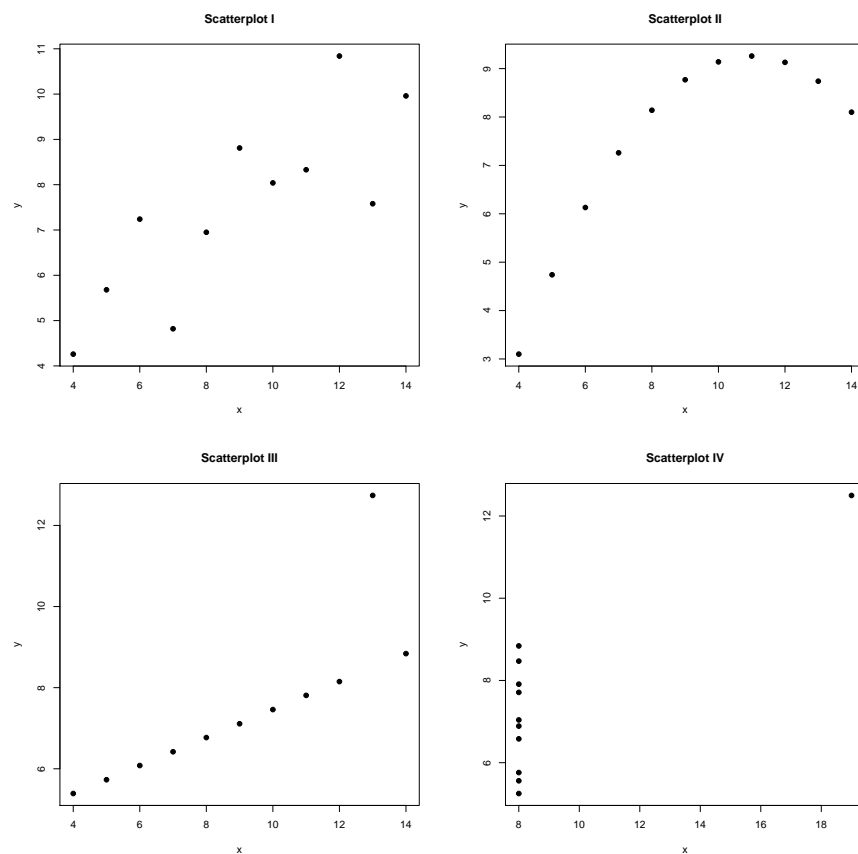|	I		II		III		IV	
	x	y	x	y	x	y	x	y
mean	9	7.50	9	7.50	9	7.5	9	7.50
variance	11.0	4.127	11.0	4.127	11.0	4.127	11.0	4.127
pearson	0.8164		0.8164		0.8164		0.8164	

(b) The plots are shown in Figure 1

Figure 1: Task 1 (b) - Scatter plots

It is not visible in the plots, that each of these has the same correlation and variance, as well as mean values for x and y.

(c) Scatterplot Matrices can be used to visualised this data. This way, the patterns between all variables are still visible and zooming into the plot can be used to see more details. The plot is shown in Figure 2.
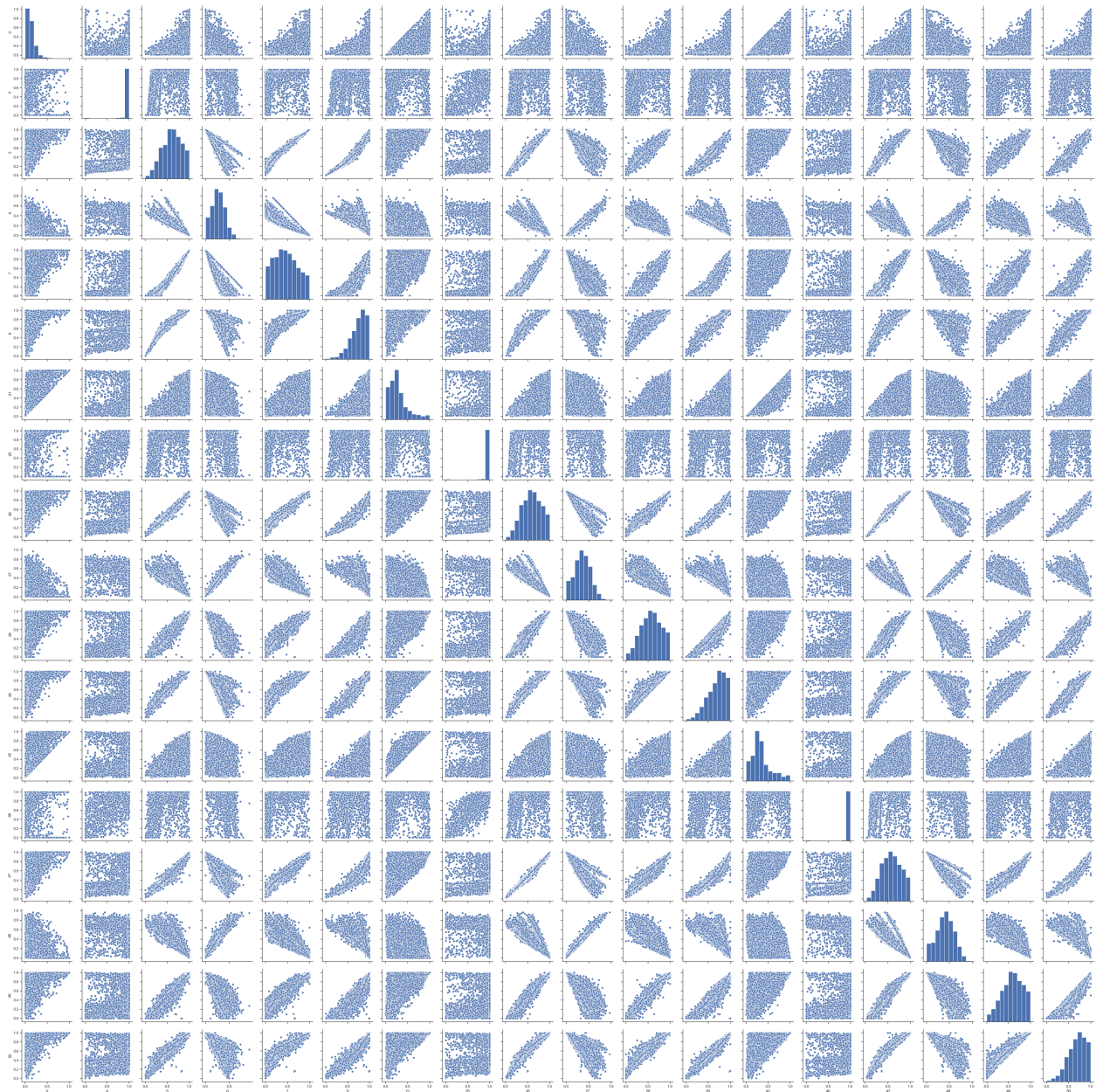

Figure 2: Task 1 (d) - Scatterplot Matrices

Task 2 - Dimensionality Reduction (DR)

a) Dimensionality reduction is specially useful for reducing computation time and space (storage and memory). Sometime redundant dimensions are causing an overhead, or dimensions can be combined, without loosing information. Also better for visualization, because the user can understand the data.
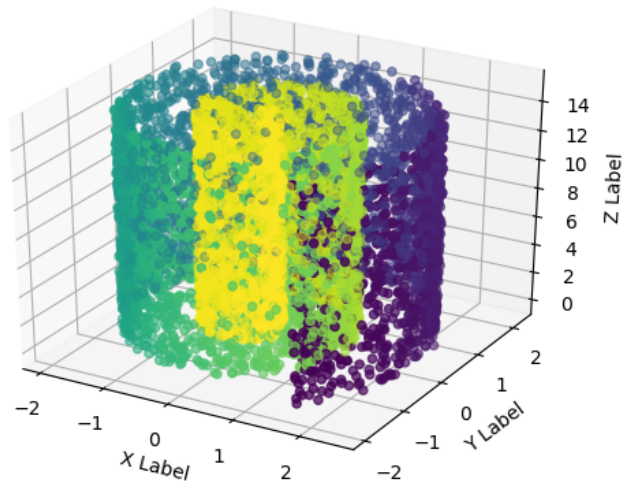
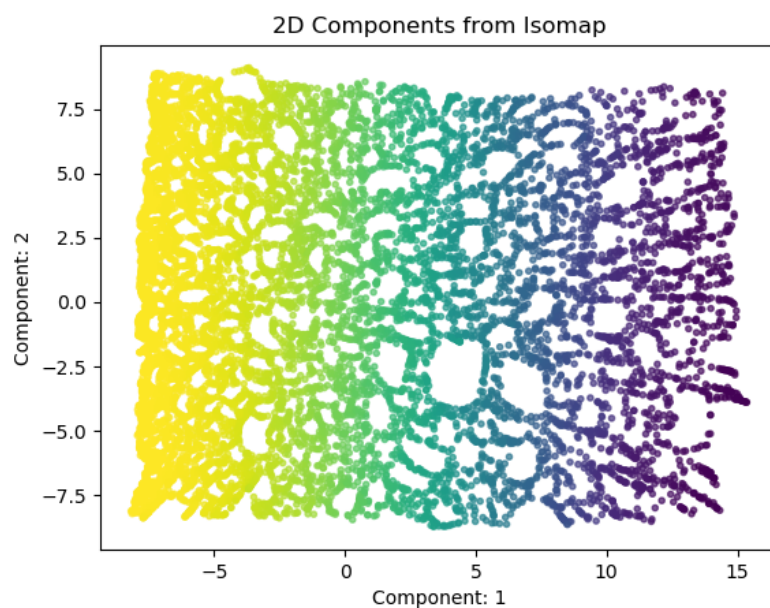b) See Figure Figure 3.

Figure 3: raw data

c) • See Figure 4

Figure 4: after pca

- The given spiral is unrolled, so the spiral information is gone missing, but the mat is was made of still preserves the 2d-information.

-