Seminar 3

1. Load dataset *nama_10_a10_e_1_Data.csv* into RStudio. Your task is to study and illustrate the dependencies of people working market segment proportions across the EU. You should also study differences across states of the EU.

   (a) Transform data from long format to short format. Use column *GEO* (indicates state) as key variable for each row. Use values in column *NACE_R2_LABEL* (indicates market segment) to identify new columns. Use values from column *Value* (indicates percent of people working in the given segment).

   (b) Create a correlation matrix using build-in R function. Try different options for correlation matrix visualization using `corrplot`.

   (c) Create a single scatter plot for *Industry* and *Manufacturing*. Create a scatter plot matrix for all segments.

   (d) Use PCA to analyze the data. Use build-in R function `prcomp` and `autoplot` function from package `ggfortify`.

   (e) Use PCA to analyze the data. Do not use build-in R function `prcomp`.

2. **Using PCA for image compression:** Use R library `jpeg` to load figure 001.jpg from study materials.

   (a) Create an empty plot and place the image into the plot using `rasterImage` command.

   (b) The image was loaded as a data matrix with 258 rows (observations) and 186 columns (variables). Use PCA to reduce the dimensionality of the data.

   (c) Create an image to illustrate all eigenvectors (identifying principal components).

   (d) Figure out the minimum number of components so components explained at least 80%, 90%, and 99% variability of the data.

   (e) Use the number of components from the previous task to reconstruct the original image. How many numerical values did you need to obtain your reconstruction?

3. **Using PCA for image pattern recognition:** Use R library `jpeg` to load figures 001.jpg - 009.jpg from study materials. Transform data from matrix to vector and save them into a data frame.

   (a) Images were loaded as a data matrix with 9 rows (observations) and 47 988 columns (variables). Each variable defines a specific pixel of some image. Use PCA to identify common patterns of the data.

   (b) Create an image to illustrate all eigenvectors (identifying principal components).

   (c) Figure out the minimum number of components so components explained at least 80% variability of the data.

   (d) Use the number of components from the previous task to reconstruct the original images.