Linear Alg. Homework 2 - Aaron Miller 17322856

Q1

In the SVD of A, we define

$$A = Q_1 \Sigma Q_2^T$$

where $Q_1 (m \times m)$ $Q_2 (n \times n)$ are matrices with columns made of eigenvectors of AA^T & A^TA respectively. $\Sigma (m \times n)$ has 'r' singular values on the diagonal, labeled σ_i, $i = 1..r$ & they are the roots of the non-zero eigenvalues of both AA^T & A^TA. Mathematically,

$$A = Q_1 \Sigma Q_2^T = (\vec{u}_1, \vec{u}_2 .. \vec{u}_r ... \vec{u}_m) \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix} (\vec{v}_1, \vec{v}_2 .. \vec{v}_r .. \vec{v}_n)^T$$

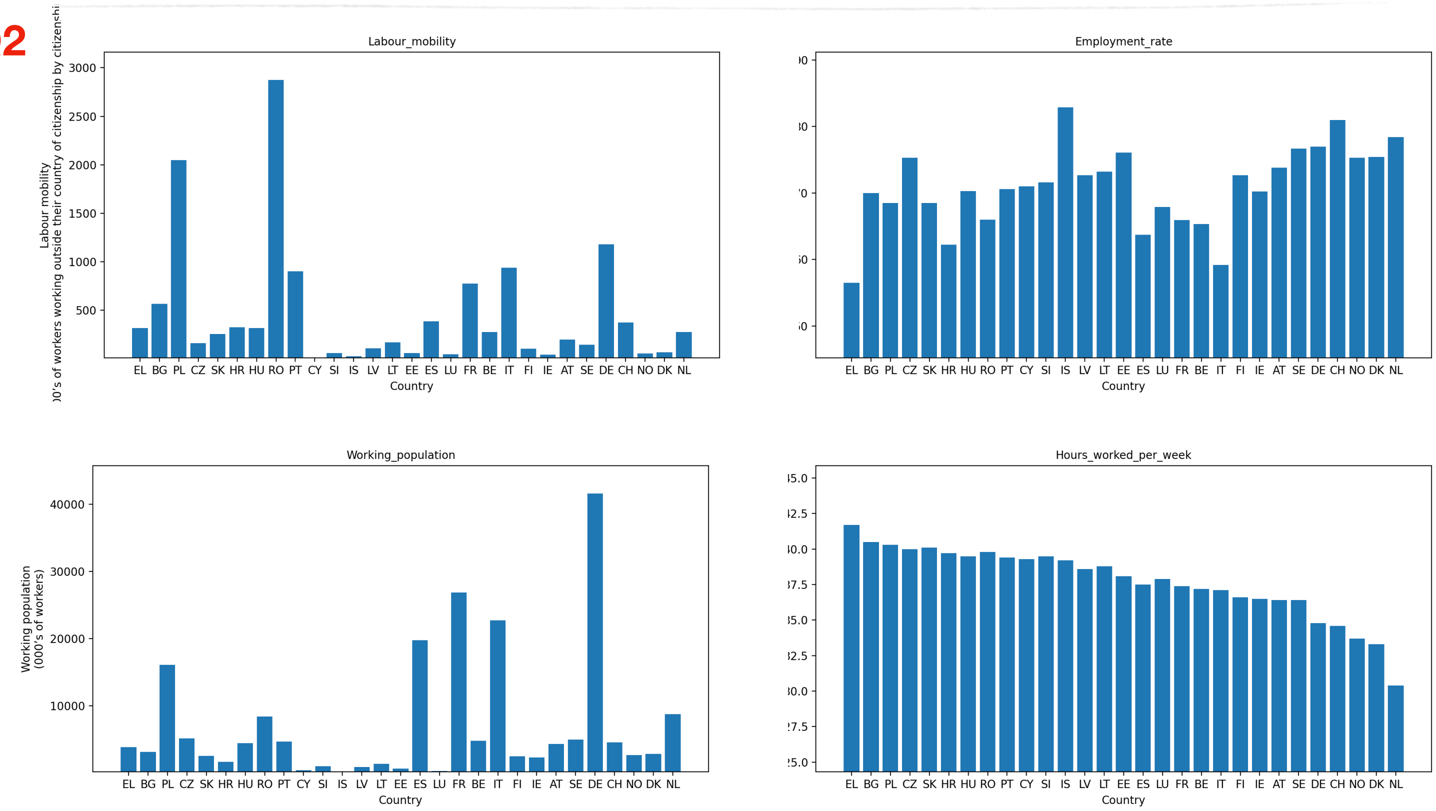
(*) $$= \vec{u}_1 \sigma_1 \vec{v}_1^T + \vec{u}_2 \sigma_2 \vec{v}_2^T + .. + \vec{u}_r \sigma_r \vec{v}_r^T$$

where u_i, v_i are eigenvectors ~~created~~ which create the colums of Q_1 & Q_2. These values of σ_i are arranged in descending order of magnitude with 'i'. So, the sum (*) can be cut of at the k-th order matrix approximation

$$A_k = A_1 + A_2 .. + A_k$$

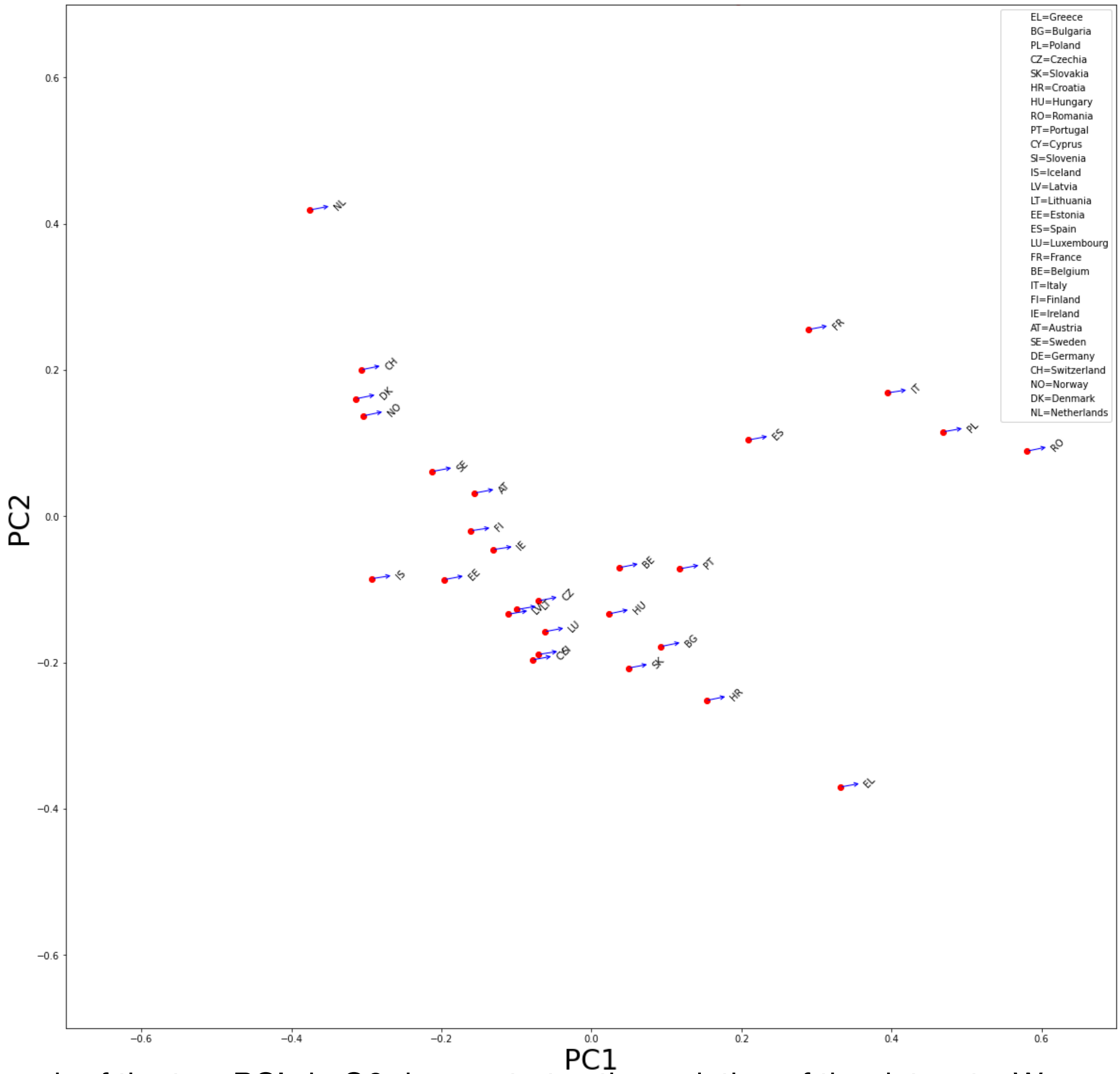
$$A_1 = \vec{u}_1 \sigma_1 \vec{v}_1^T$$ @ First order in σ_i

Q2



Q3. Q4. Q5 - See Script

Q6



Q7

The graph of the two PC's in Q6 demonstrates the variation of the datasets. We can also analyse the correlation matrix to see relationships between variables:

C = [[1.   -0.48 0.2   -0.22]
 [-0.48 1.   -0.28 -0.17]
 [0.2 -0.28 1.   0.48]
 [-0.22 -0.17 0.48   1.]]

From each coefficient C_ij we can tell the relationship by the sign and magnitude:
• If both variables tend to increase or decrease together, the coefficient is positive.
• If one variable tends to increase as the other decreases, the coefficient is negative.

This means there that for C_12 = C_21 = -.48 < 0 is anti correlated implying that as hours worked goes up, employment rate goes down. This is a strong anti correlation at -.48 and tells us that the poorest countries work the hardest. The relationship between labour mobility (working population) and hours worked is weakly (anti-)correlated with C_13(4) ~ (-)0.2.

Similar relationships can be extracted from the correlation matrix to show how the variables are related, however I am conscious of the page limit and am unable to discuss this further.

From the plot of the PC's (Q6) we observe some clustering of geographically similar countries such as Sweden, Norway, and Denmark however this is not a hard rule as economically similar countries such as Switzerland are peppered near them. This shows that similar countries vary in the same way, as expected. Focusing on Ireland, we see that we lie near (0,0) showing we don't vary to wildly from an average EU country, in terms of these economic parameters anyway.