**Problem 1)**

**A)**

As we know that there if X variable is correlated with each other is called multicollinearity occurs. There is way to check multicollinearity one is to check correlation matrix if two independent variables are highly correlated with each other equal or above 0.7 or depending on type of data it can also differ.

Another way To check for multicollinearity, we needed to verify the VIF statistic for each variable. Given that all the values were below 10, then there is not an issue. And if VIF is above 10 then variables are highly correlated to each other that cause multicollinearity .

We check multicollinearity because if variable is having higher VIF 10 or above then variable is remove from the model . We know that multicollinearity occurs when two or more of the independent variables used in regression are moderately or highly correlated, so checking for this is important. We could potentially see high correlations among the independent variables (i.e., extreme multicollinearity) which increases the likelihood of further errors later on - with the β estimates, standard errors, and so forth. Second, the regression results may be confusing and misleading, we might find that the p-values for both βˆ 1 and βˆ 2 (the least squares estimates) are nonsignificant.

**B)**

As we know most common form of overfitting is due to multicollinearity. Often we need to try to get information from sparse data point, but still have many variable that affect the result to get overfitting. overfitting is due to make model so precise to fit the training dataset. Model performs well on the training data but fails to perform better in test dataset. We solved the overfitting by solved using the PCA or the factor analysis and using this technique it can reduce shared variance or the multicollinearity into smaller groups.

As we know, the most popular methods out there for defining the best model include optimizing coefficient of determination r2 or Adj-R2, optimizing Mallows’ cp statistics, and using different selection methods (stepwise, backward and forward).

.

**C)**

As we know linear regression is best to determines fit model. Mainly used when there are large number of variables or having the smaller number of the observations in the dataset.

Techniques:

Ridge regression: is modify regression formula , Ridge regression is about replace the computation of beata, and Lasso regression: is to optimization technique is performs variable selection at the same time Both type of the regression techniques tries to minimize the MSE.

**Problem:3)**

**A)**

The aim of this study was to evaluate the structure of the MRS, they used exploratory factor analysis to determine the number and nature of underlying factors of the MRS. To discussed in this work. Music receptivity can be defined as a measure of the extent of internalization that an individual has, to a given piece of music, as measured at the point of listening. Through three studies, we demonstrate the psychometric properties of the construct— the Music Receptivity Scale (MRS).

B)

In the given study Principle component analysis is used with rotation. The type of the rotation used is Oblique. Which means the components are dependent.

C)

In the Given study the researchers come up with concentrated to 4 factors then later they come up 2 factors. They further refined the two-factor model by removing certain items that had relatively weaker loadings or that had considerable cross-loadings. It was obvious that we must take care of external environmental conditions before we attempt to measure music receptivity. Principal axis factoring was again performed on the reduced 20 items. Overall MSA was 0.89, and for individual items, it ranged from 0.74 to 0.94. Bartlett’s test of sphericity was also significant [χ 2 (190) = 2,397.69, p < 0.001].

D)

Talking about the concentrated factors of 4 factors

Emotion- Emotionally attached heart and good memories

Interest – is a what interest of music they like to listen .

Attention- is like what is intensity of focus was varying while listening it.

Hurdles- is like lyrics of given music or music sounded boring and etc.

They also Talking about the concentrated factors of 2 factors

Affect – both emotion and interest

Attention- focus was varying while listening it.

E)

As we talking about stability internal consistency measured using Cronbach’s alpha was 0.89. For the 20-item four-factor solution, the alpha values were 0.81 (emotion), 0.84 (interest), 0.68 (attention), and 0.59 (hurdle). For the 20-item two-factor solution, the alpha values were 0.87 (affect) and 0.75 (attention). And test–retest reliability was found to be very high, 0.87, p < 0.001. therefor consistency can mean that both the stability of the construct and the consistency of the musical piece influence the respondents.

F)

No, they do not use factor for the regression.

G)

As in conclusions I like to say that research was about confounding factors in clinical application of music therapy, societal and cultural bases of emotion induction through music, precision of delivery of music, instrumental music may be therapeutically superior to other modes, and necessity to account for a person’s internal state while assessing the degree of music receptivity.

**Problem:4)**

**A)**

**Chart, histogram

Description automatically generated**

* As we see that Total number of the components needed to explain 100% of the variance is 60. So total 60 components are needed to explain the 100% variance of the dataset.
* From the scree plot 5 components are determined which are above the 1 variance by eigenvalues which we setup between 0 to 1. And I will use same 5 components principle component analysis in a model .

**B)**

**Formula for each component:**

**RC1=0.50\*Q1+(-0.57\*Q3) +0.791\*Q5+0.80\*Q7+0.63\*Q9+0.73\*Q11+0.81\*Q13+0.74\*Q17+0.58\*Q21+0.71\*Q23+(-0.44\*Q19)**

**RC2 = 0.80\*Q2+0.67\*Q4+0.63\*Q6+0.74\*Q8+0.71\*Q14+0.53\*Q18(-0.77\*Q22)+0.49\*Q10+0.43\*Q28+0.48\*Q34**

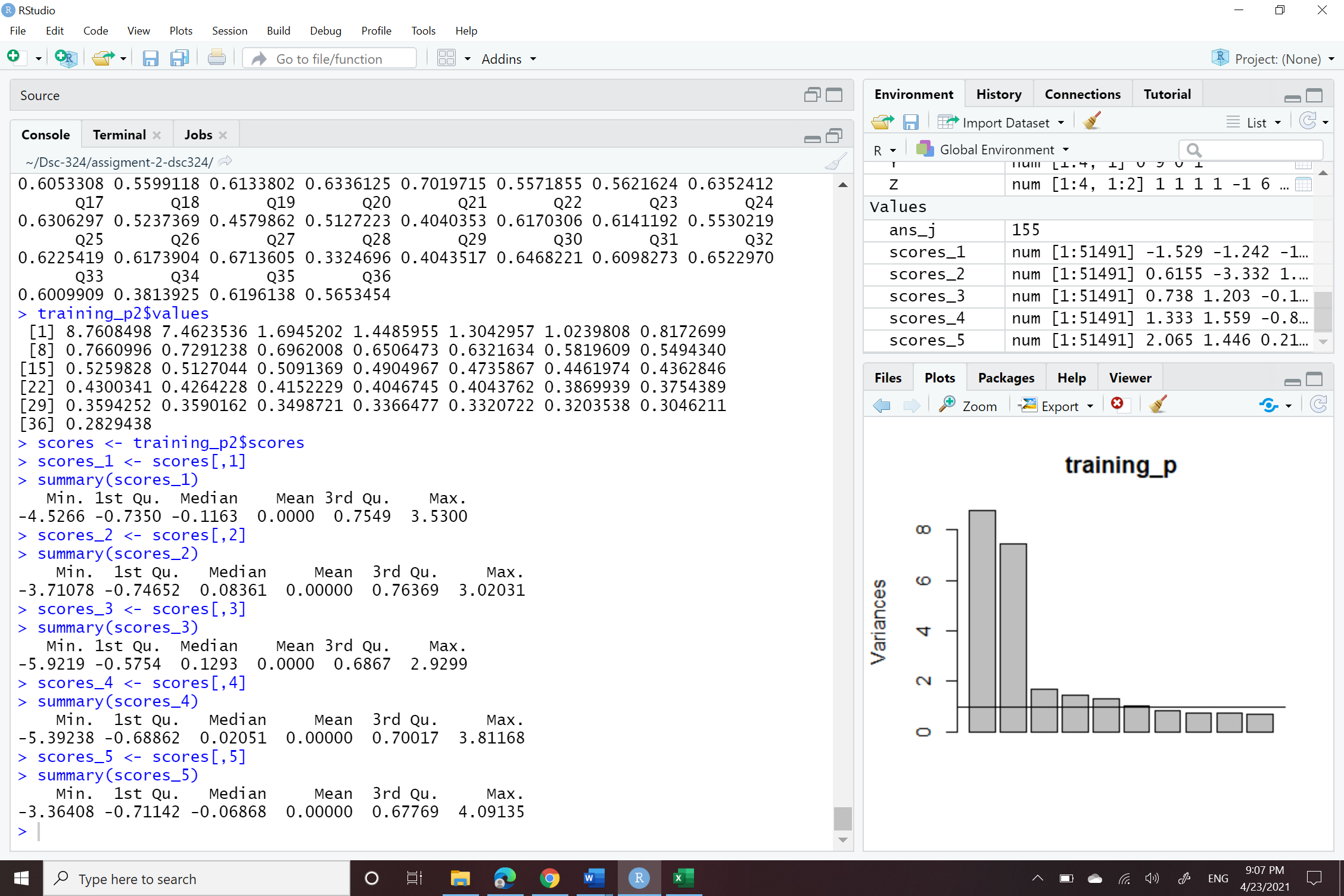
**RC3 =-0.44\*Q1(-0.44\*Q9)+0.66\*Q15+0.71\*Q25+0.77\*Q27+0.73\*Q31+0.71\*Q33+0.69\*Q35+ 0.49\*Q19+ 0.47\*Q29**

**RC5 = 0.40\*q18+0.64\*Q24+0.71\*Q30+0.76\*Q32+0.68\*Q36+0.42\*Q20**

**RC4=0.53\*Q10+0.74\*Q12+0.73\*Q16+0.52\*Q20+0.71\*Q26**

1. RC1 component is **feeling**
2. RC2 component is **Abandoned**
3. RC3 component is **Openness**
4. RC4 component is **Conscientiousness**
5. RC5 component is **reassurance**

C)



As we see that component 1 the minimum score is -0.7350 and the maximum score is 3.5300.

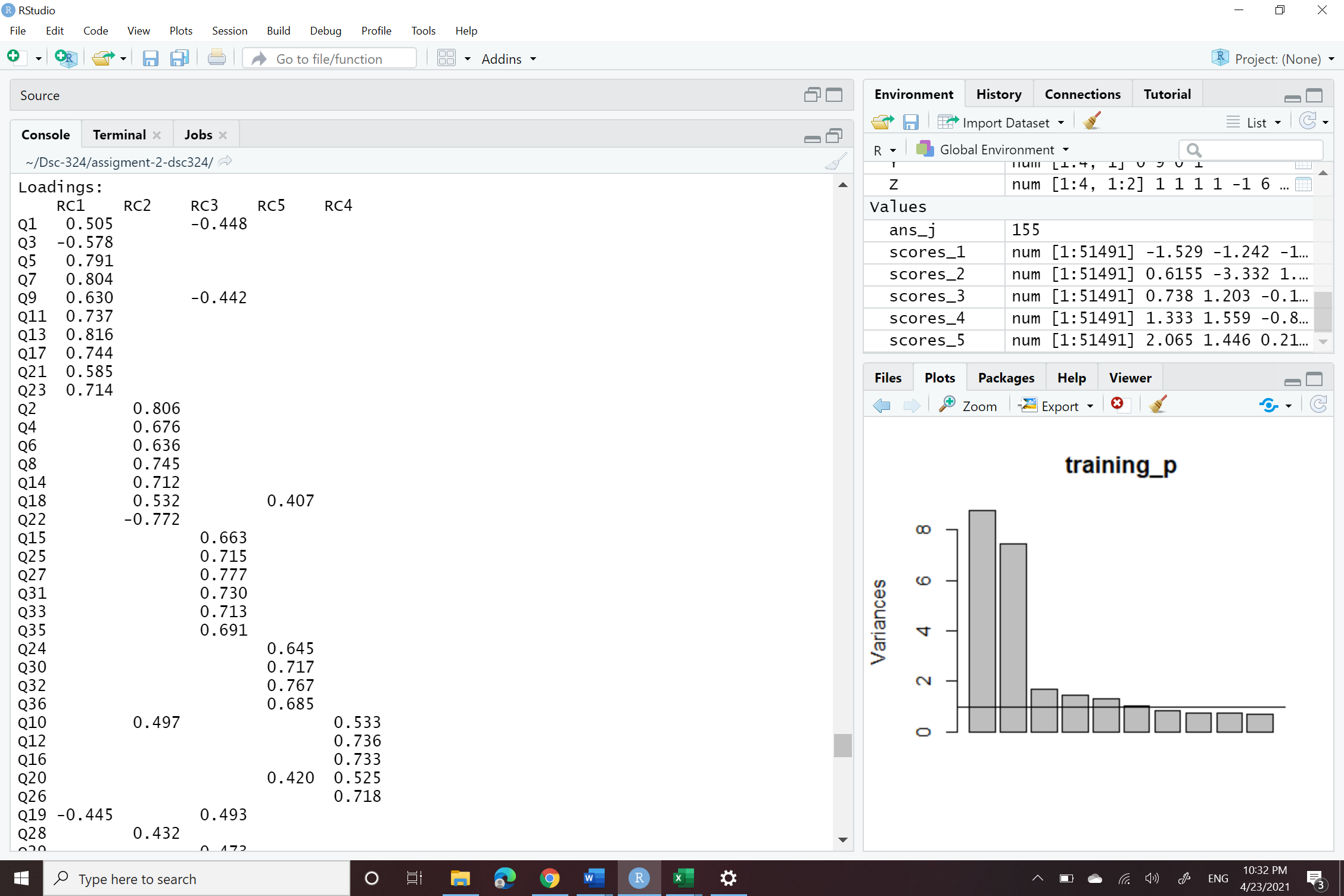
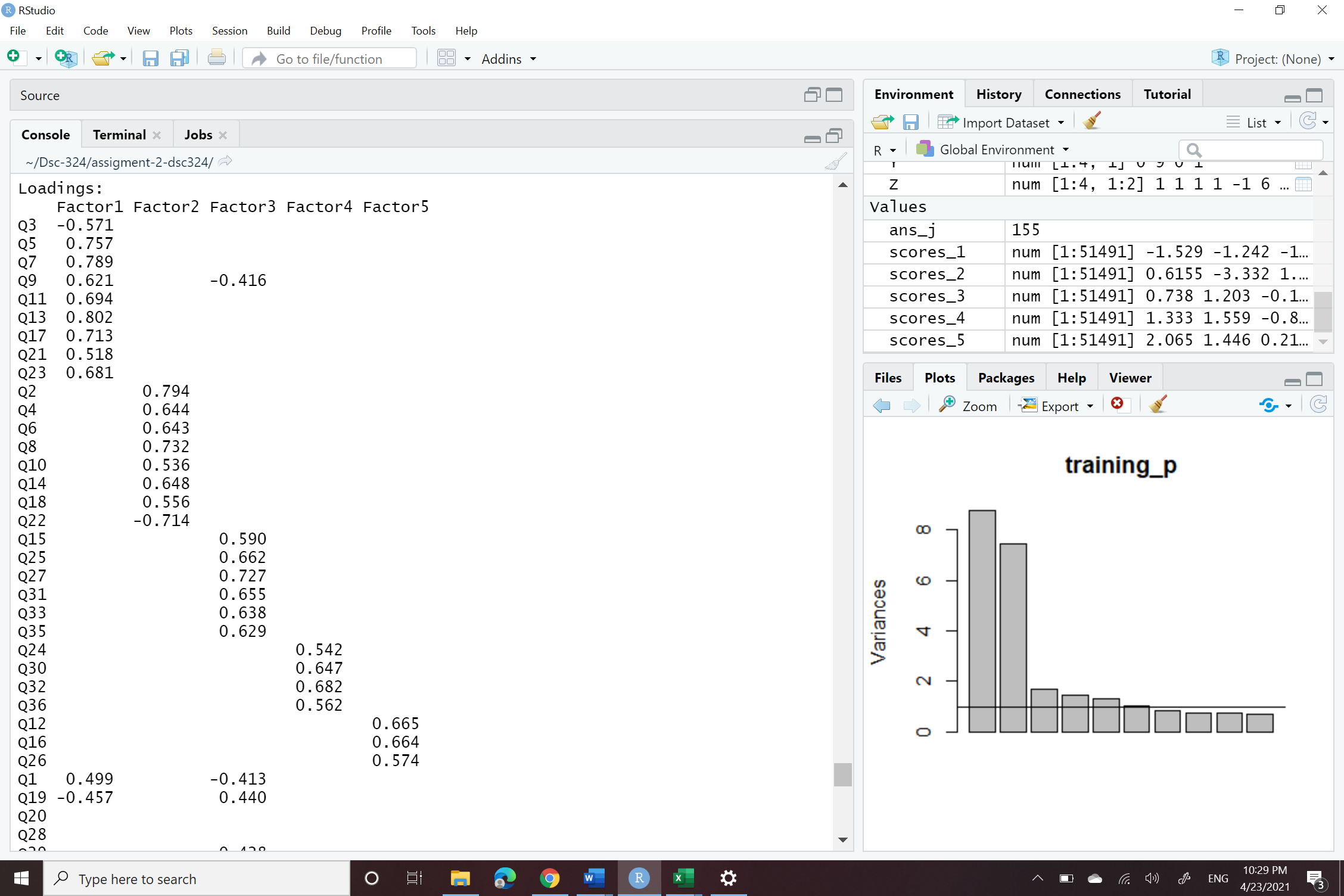
As we see for component 2 the minimum score is -3.71078 and the maximum score is 3.02031.

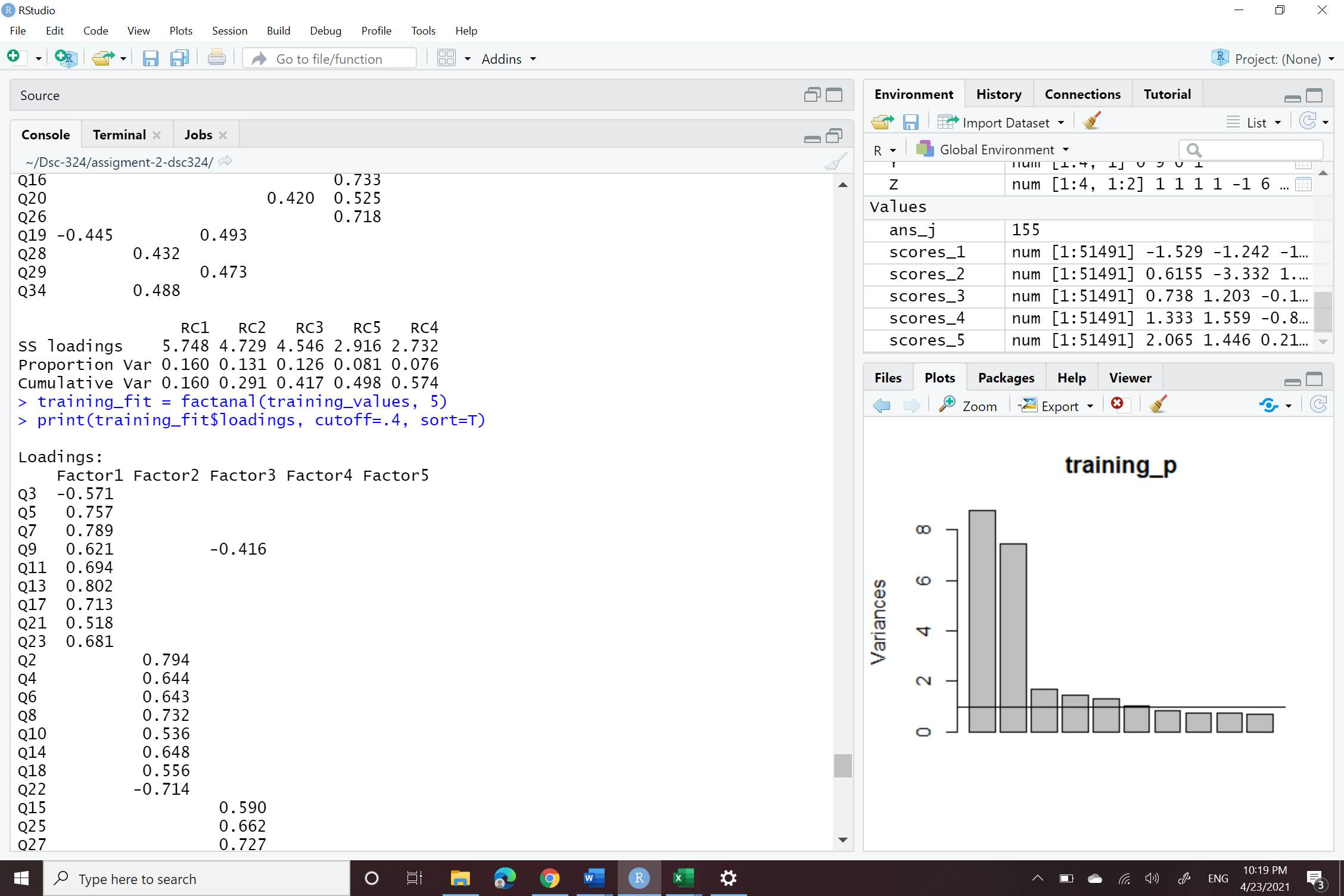
Then we see component 3 the minimum score is -5.9219 and the maximum score is 2.9299.

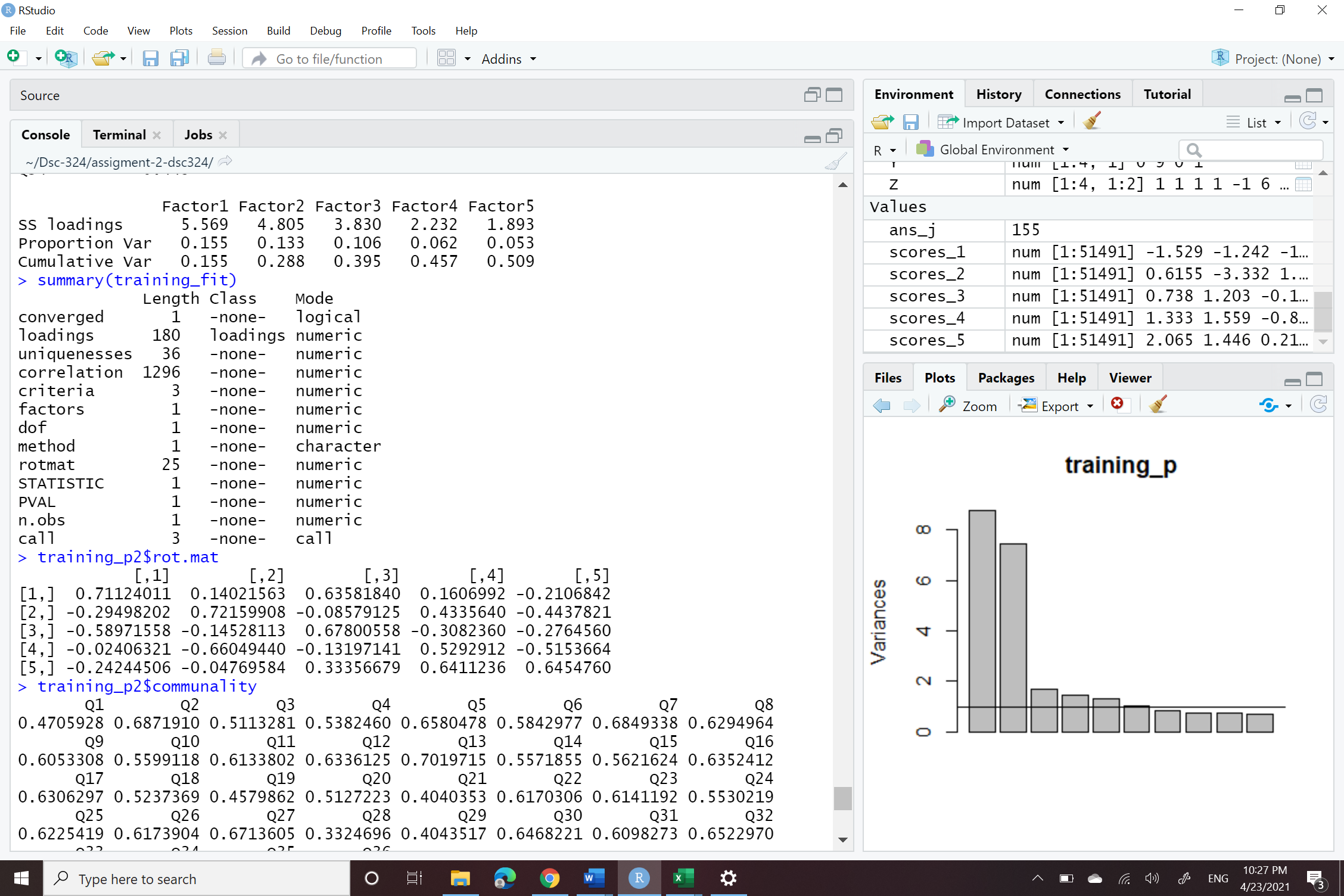
Then we see component 4 the minimum score is -5.39238 and the maximum score is3.81168.

Least we see component 5 the minimum score is -3.36408 and the maximum score is 4.09135.

D)







As we see after applying the factor analysis it did not provide the better result then the Principle component analysis it was some what the same result given by the Factor analysis. The factor analysis did not give the better but rather it provided little worse result because by using the PCA we can gather max cumulative variance up to 57% but the factor analysis only gives cumulative variance of 50% so in total for particular this dataset PCA works better then the factor analysis. And yes factor analysis do change the ability to interpret results because using it, It provided confidence that the results provided by the PCA are accurate.