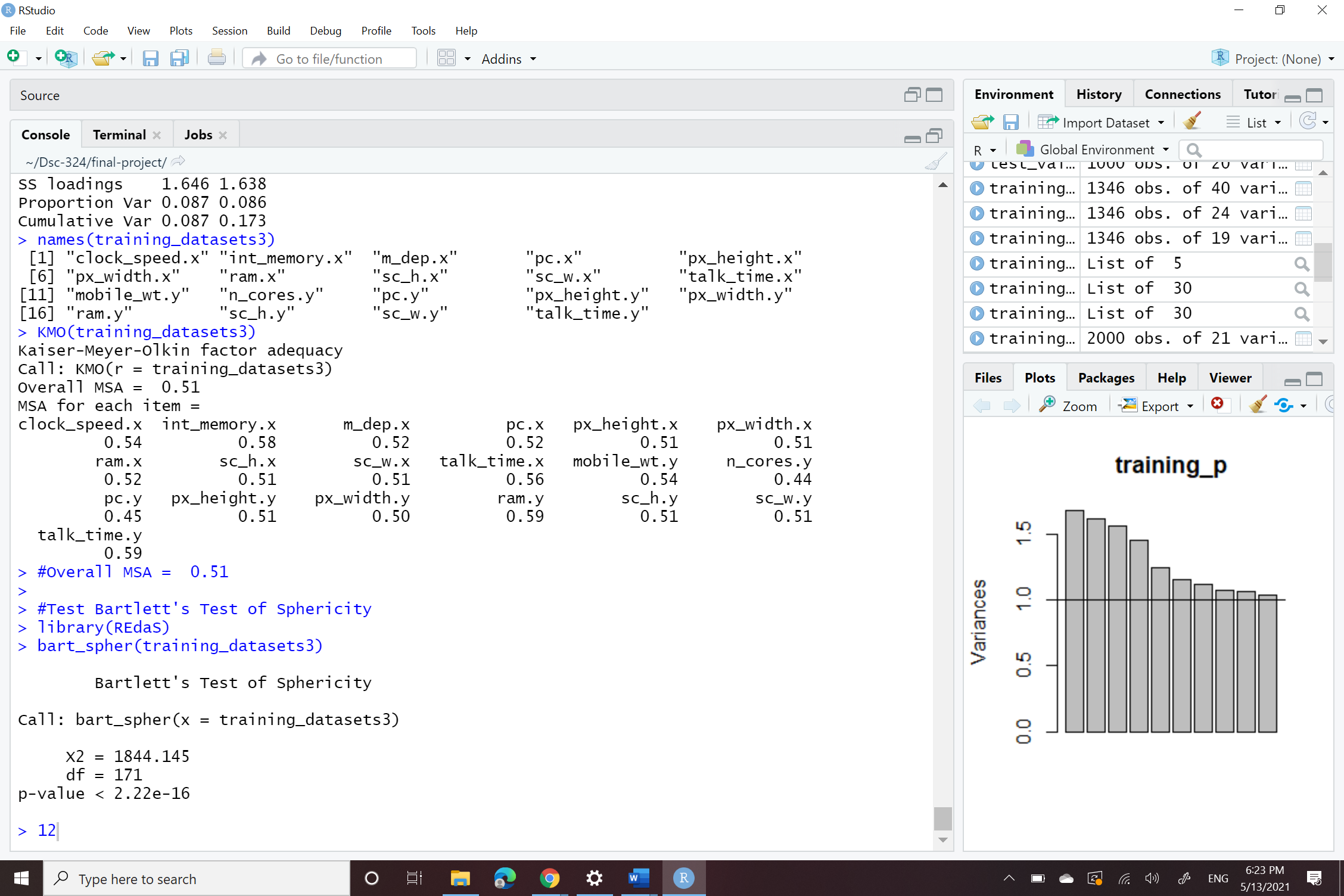
**Problem 1)**

Our dataset is for some relation between features of a mobile phone and its selling which after preprocessing consist of total of the 42 variables and total of the 1000 observations. Each observation is the features of mobile phone. Personally, I am applying the PCA analysis on the dataset to find out the insights of the data. Our focus or the research question is to predict the selling of mobile . So, we have 41 independent variables and the 1 dependent variables in the dataset.

In the our study we use Principle component analysis is used with rotation. The type of the rotation used is varimax.

In order to get insights for the data via Principle component analysis. I only use a variables which are continuous so I remove other variable because PAC analysis only continuous are use do analysis so after removing I come up 19 variables in total.



By using KMO test Overall MSA 0.51 and for individual items, it ranged from 0.45 to 0.59. and by using Test Bartlett's Test of Sphericity p-value < 2.22e-16 (Very Small Number)

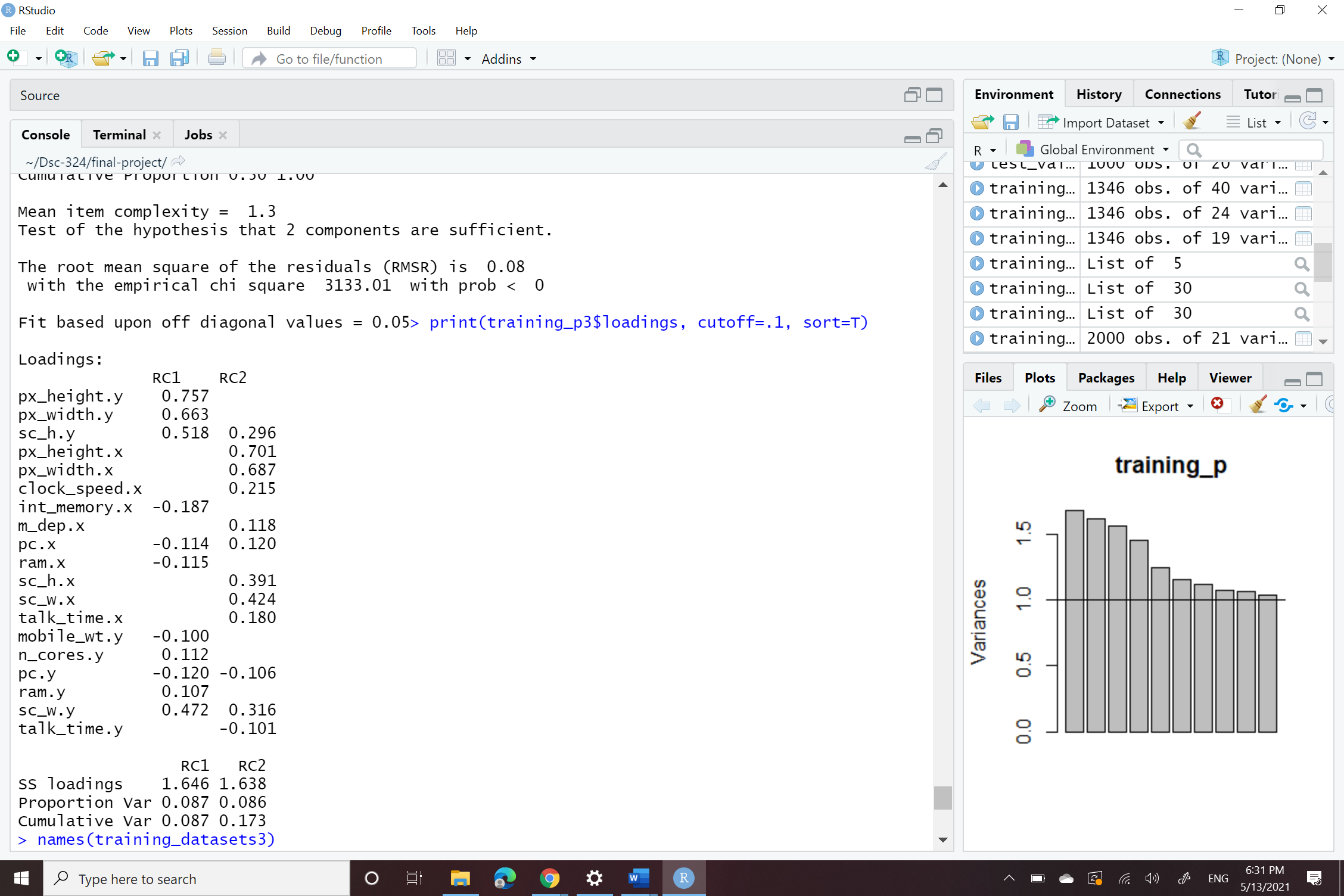
Chart, histogram

Description automatically generated

From the scree plot all components are determined which are above the 1 variance by eigenvalues which we setup between 0 to 1. And I will use same 2 components principle component analysis in a model .

As a result, after applying Principle component analysis the varimax rotations on the 2 factors. We set to the cutoff to 0.1 because our variables are smaller in number there were present so for that reason, we need to less the cutoff to 0.1. and there are some variables are not contributing the variance or loading. So I just remove that re-run PAC analysis.

If we consider total of 2 variance the cumulative variance is 0.173 means 17.3% which is very lower cumulative variance.



**Problem 2)**

A)

This‎ study aimed to determine the role of orientation to happiness and life goals in predicting subjective ‎wellbeing. .The study population included all students at Tehran University of Medical Sciences in the academic year 2016-2017 (N=2100). In the canonical analysis, if the reliability of the variables is high (0.8 or more), the ratio of cases to independent variables should be 10 to 1. According to the considerations related to test power, the sample size in this study was estimated to be 285 people. We know that this data is suitable for the canonical correlation analysis because only data contains less than 300 observations.

B)

In the paper Canonical correlation is being applied to the between two groups of the data the one group is H1: There is a relationship between the components of happiness orientation, including pleasantness, meaning, and engagement with the components of subjective wellbeing such as positive affections, negative affections, and life satisfaction. has the 3 variables and the other group H2: There is a relationship between the components of life goals, including wealth, fame, image, growth, relationship, and community with the components of subjective wellbeing such as positive affections, negative affections, and life satisfaction . has the 6 variables.. All the data type is metric. In the given study first they took data from 20016 to 2017 and SPSS 16 software was used for data analysis. Using the canonical correlation model, the relations between variables of orientations to happiness, life aspirations, and mental wellbeing were analyzed.

‎c)

The samples were collected by the cluster sampling method. Among the first ten faculties of the university, three faculties (medicine, dentistry, and pharmacy) were randomly selected. Next, from each faculty, three classes of first- and second-year students were randomly selected. The selection of first- and second-year students compared to older students was based on the notion that the mental wellbeing of this group of students was less dependent on intra-university factors. In the given study they did not evaluate the stability of the components.

D)

In the given study they concentrate on the 2 canonical variates and yes , interpret the variates in the form of the original variables and also Redundancy and adequacy analysis is applied to the canonical variates in terms of the both the groups.

E)

To conclude, A important prerequisite for wellbeing is the orientation to happiness. People who experience a higher level of satisfaction in life have an orientation to meaning and engagement and pursue intrinsic goals such as personal growth, relationship, and community. I like say This result was consistent with previous research findings in the field of improving life satisfaction that found the meaning and engagement stronger than the pleasure.

**Problem 3)**

**Part-1)**

A)

As we see wilk’s lambda test there are total 7 variables in which only 5 variable is significant and out of all 5 passes the hypothesis test as the p-value till the 9 variate is 0 but for the 6th  and above variates they fails to reject the null hypothesis as their p-value is > 0.05 or 5%. To conclude only 5 variable are to be considered for further analysis.

B) According to Wilk's Lambda Test there are 5 significant variates to be considered. As they manages to pass the test as those 5 variates has p-value <0.05 or 5%.

c) The first two concor are displayed below. The first concor is having the value of 0.4621615 and the value of the second concor is 0.3453192.

D)

As we see in above analysis with the spending and music variable by applying the Wilk's Lambda Test we get to know that how many important or usable vitiates are present of the 7 total variates. So as per the output of the test displayed in the top table there are only 5 out of 19 significant variates are present as all the 5 variates exceeds the p-value greater than 0.05 or 5%.

As CV1 has 0.4621615 that means using the first variate 46.21% of the overlapping correlation is exists in both the x and y variates or between spending and music and so 53.79% is error or it can be explained by adding some other variable of music.

**Part-2**

A)

CV1 (for music)=

(0.047906430 ) \* Music + (0.003073999 )\* Slow.songs.or.fast.songs + (0.232657577 )\* Dance + (0.011726873 )

\* Folk + (0.10941444 )\* Country + (0.058060463)\* Classical.music + (0.105037478 )\* Musical + (0.518424962 )\*

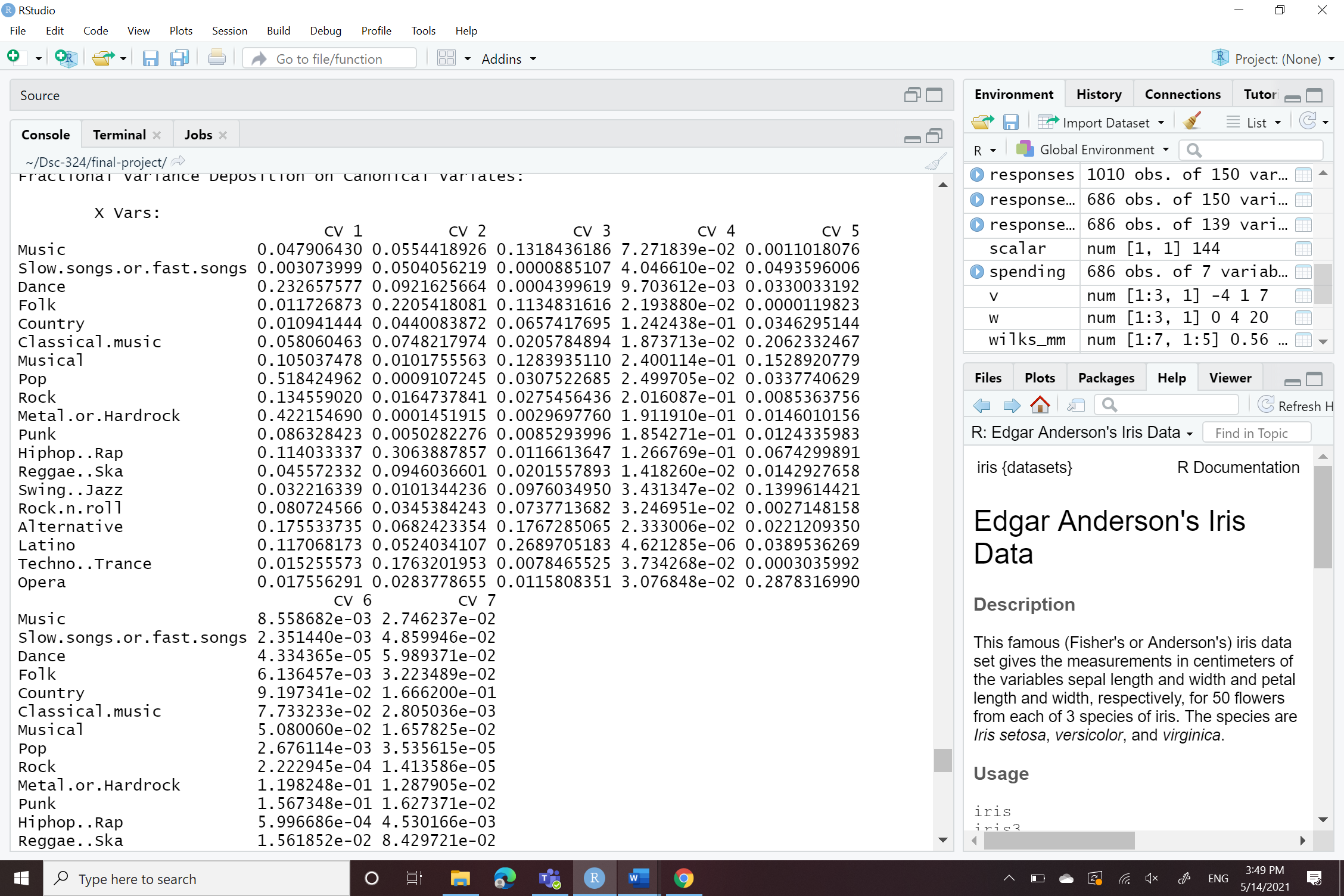
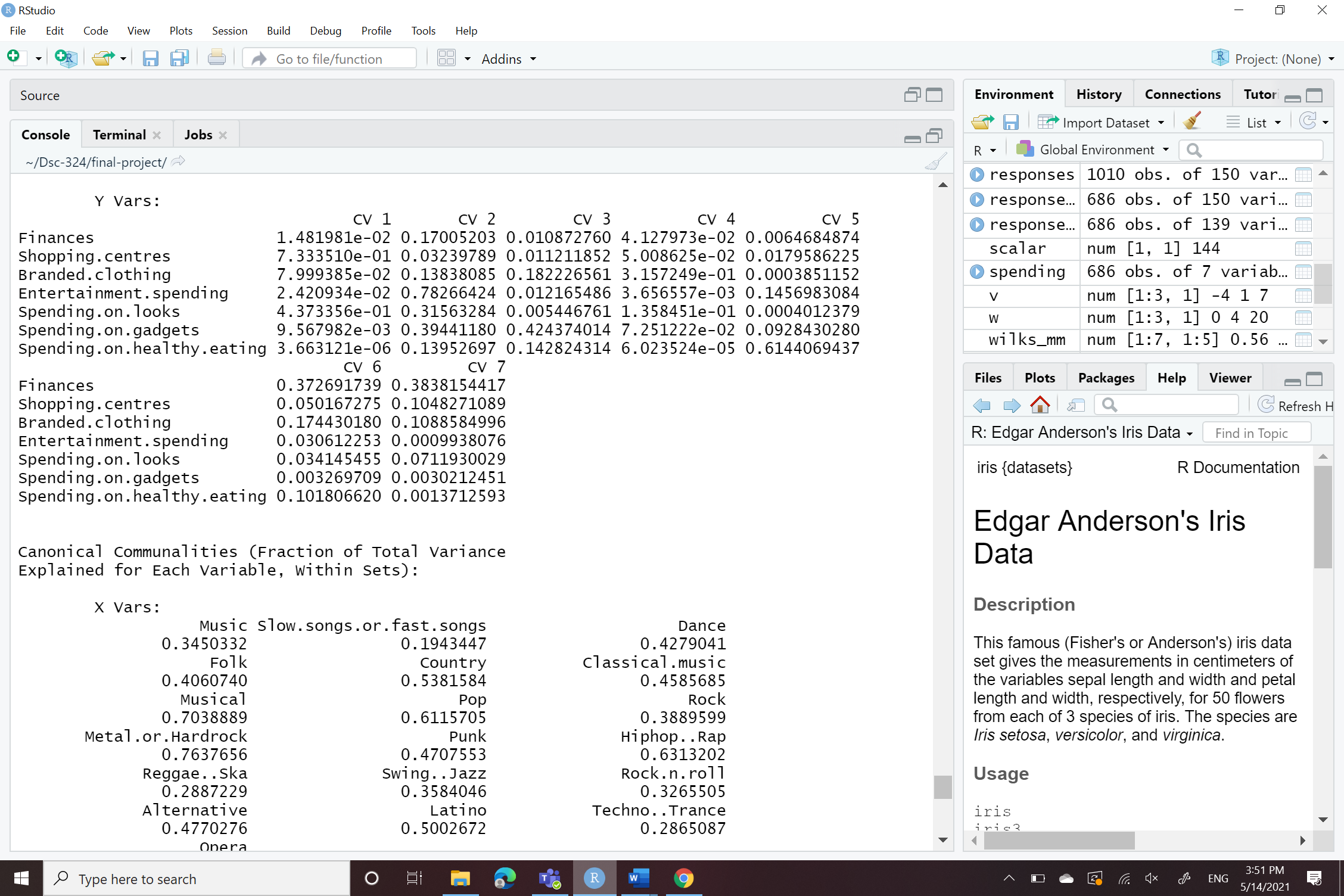
Pop +(0.134559020 )\* Rock + (0.422154690 )\* Metal.or.Hardrock + (0.086328423 )\* Punk + (0.114033337 )\* Hiphop..Rap + (0.045572332 )\* Reggae..Ska + (0.032216339)\* Swing..Jazz + (0.080724566 )\* Rock.n.roll +

(0.175533735 )\* Alternative + (0.117068173 )\* Latino + (0.015255573)\* Techno..Trance + (0.017556291)\* Opera

CV1(for spending )= (1.481981e-02)\*Finances+(7.333510e-01)\*Shopping.centers+(7.999385e-02)\*Branded.clothing+(2.420934e-02)\*Entertainment.spending+(4.373356e-01)\*Spending.on.looks+(9.567982e-03)\*Spending.on.gadgets+(3.663121e-06)\*Spending.on.helthy.eating.

B)

As we see in table given blow x vars: is music variable and y vars: is spending it view for Fractional Variance Deposition on Canonical Variates.

C )

As we see in above analysis from music section most contributing variables are pop with 0.51, Metal.or.Hardrock with 0.42, Dance with 0.23 and other are also contributing to the correlation but they are lower than then this variable. On the other side with we talk about spending in that we can see that most contributing variables are Spending.on.gadgets is 9.56e-03 than after Branded.clothing is 7.999e-02 than after Shopping.centers is 7.333e-01 and other are also contributing to the correlation but they are lower than then this variable. So from we can say that people have interest in pop music , dance and Metal.or.Hardrock and that people are interesting in Spending.on.gadgets , Branded.clothing and Shopping.centers . Also, there are all positive correlations on both the side x and y so there could not have be inverse relationships between Spending and the music.

To conclude for the first canonical variate overlapping correlation is 46.21% and the remaining 53.79% can be achieved by adding more genres of music with little more type of Spending also that is how we can get more about of different how different types Spending tend to listen to what different type of music.