Certificate Course
on
Psychoinformatics Using R

RAMAKRISHNA MISSION VIVEKANANDA CENTENARY COLLEGE, RAHARA

Submitted by: Nishanka Das
                                  (A01-1112-117-003-2022)

Supervised by: Mr. Joydeep Kumar Dam

Contents

ABSTRACT

To study human behavior one of the prime method is taking data from random people and asking their opinion about their behavior according to the question. Total number of participants N = 31. There are 25 questions , the 25 questions are dived inti 5 parts 5 in each . The parts have different aspects e.g. Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness . With using R programming language heatmap is made and observe the behavioral expressions.

INTRODUCTION

Psychoinformatics is an emerging interdisciplinary field that uses principles from computer science for the acquisition, organization, and synthesis of data collected from psychology to reveal information about psychological traits such as personality and mood. Psychoinformatics solves these problems by storing Big Data related to psychology (such as communications on smartphones or social media websites) and then data mining for relevant psychological information.

R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

The psych package (Revelle, 2021) has been developed at Northwestern University since 2005. To include functions most useful for personality, psychometric, and psychological research. A general purpose toolbox (payeh) for personality, psychometric theory and experimental psychology. Functions are primarily for multivariate analysis and scale construction using factor analysis, principal component analysis, cluster analysis and reliability analysis, although others provide basic descriptive statistics. Item Response Theory is done using factor analysis of tetrachoric and polychoric correlations. Functions for analyzing data at multiple levels include within and between group statistics, including correlations and factor analysis.

OBJECTIVE

The main objectives of the project is to predict , describe , explain or identify the psychological aspects. These goals are the foundation of most theories and studies in an attempt to understand the cognitive, emotional, and behavioral processes that people face in their daily lives. The main five parts of the observation is Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness.

MOTIVATION

In this project we have done survey and data handling of the collected data . By this methodology we have gathered some experience in R programming language which has a large aspect in future and also have improved social skills during survey.

METHODOLOGY

- Collecting questions for survey.
- Making a chart for taking survey of people.
- Asking a person for survey and explaining him/her why this survey is done .
- Asking them the 25 questions .
- Making the CSV format of the data.
- Plotting the data in R and making heatmap.

SOFTWARE USED

- R Gui
- Psych package of R
- Readr package of R
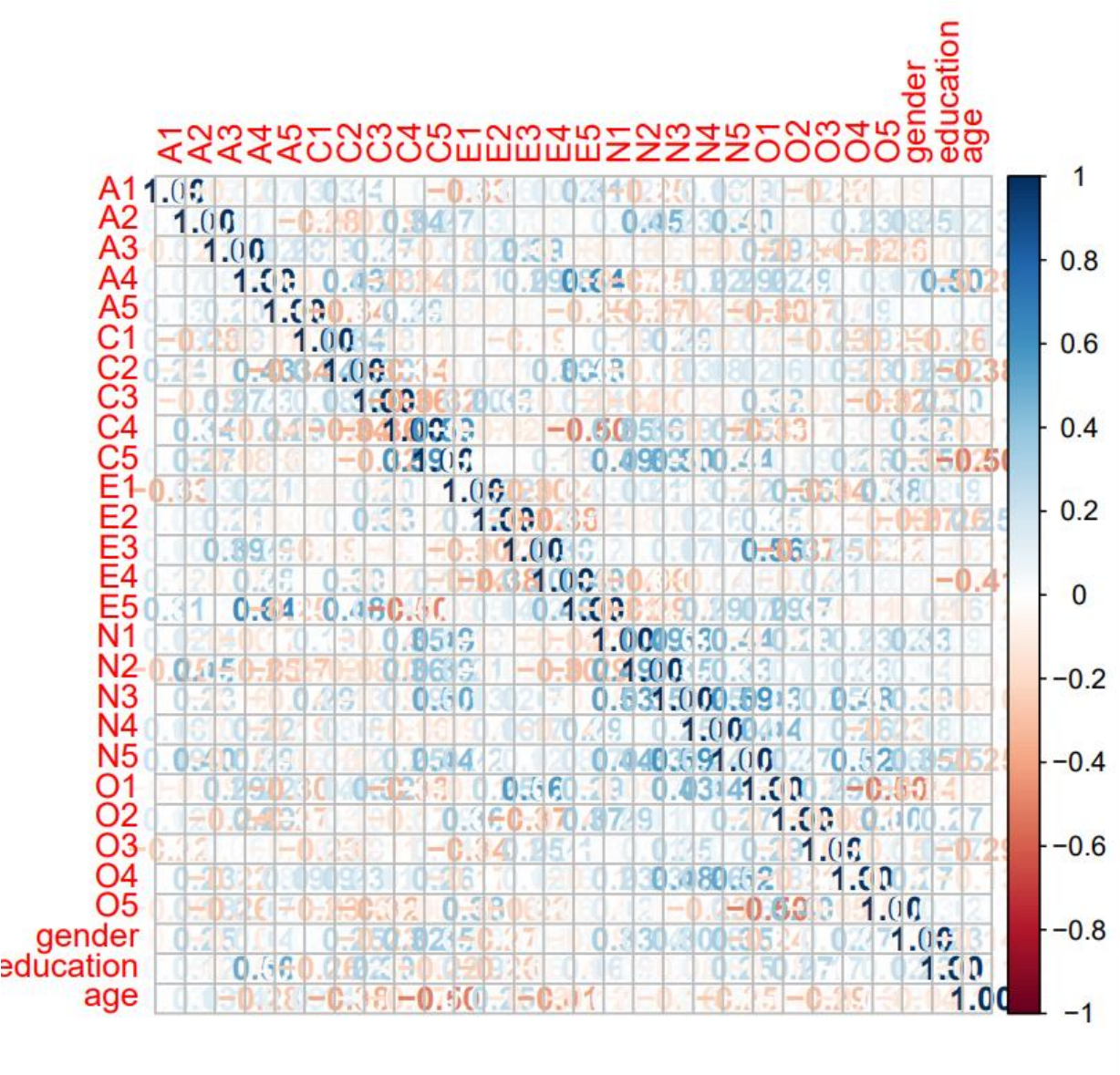- Corrplot package of R

RESULT

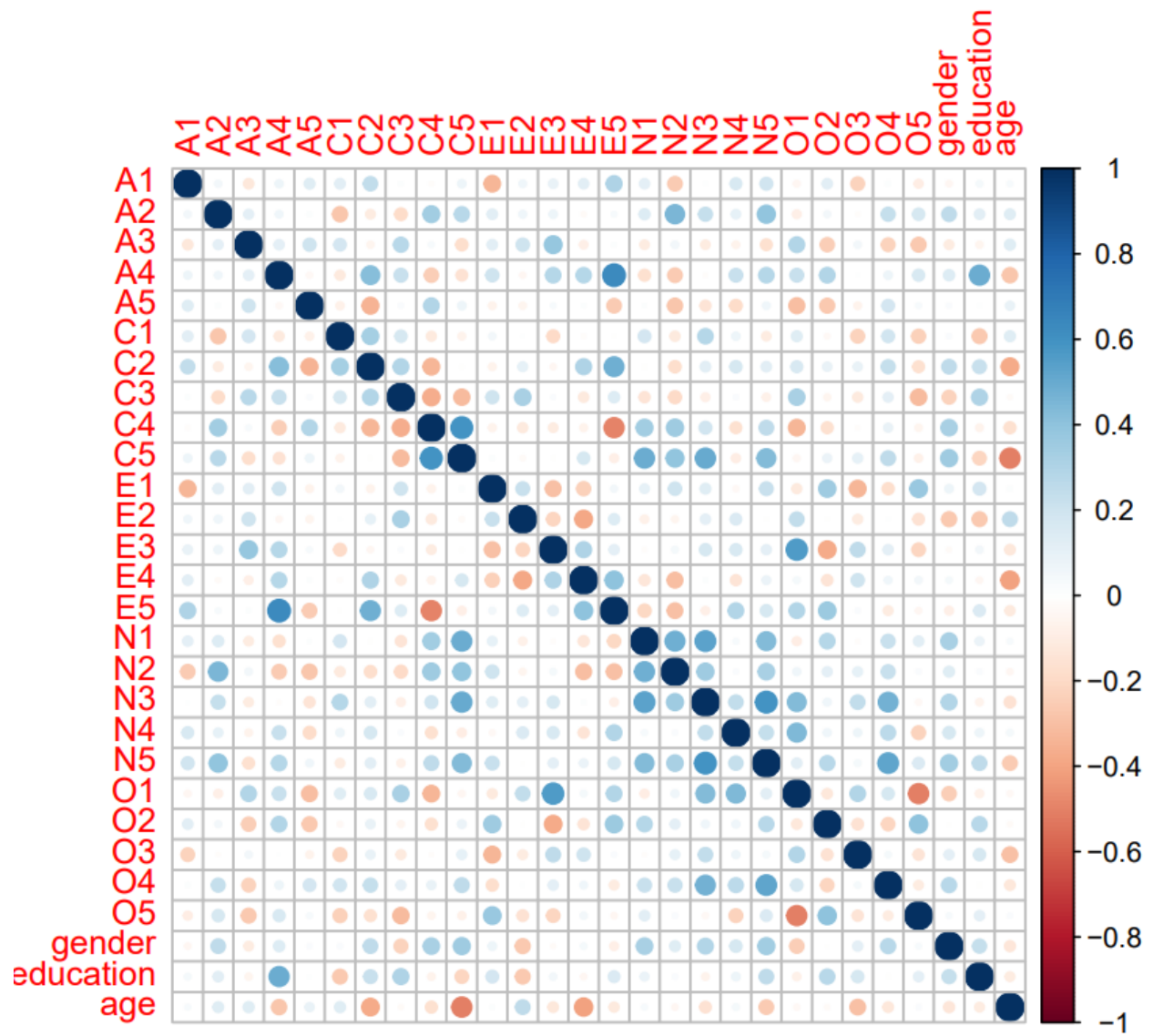Fig 1 : corrplot(M, method = 'number')

(Question Correlation)

Fig 2 : corrplot(M)

(Question Correlation)
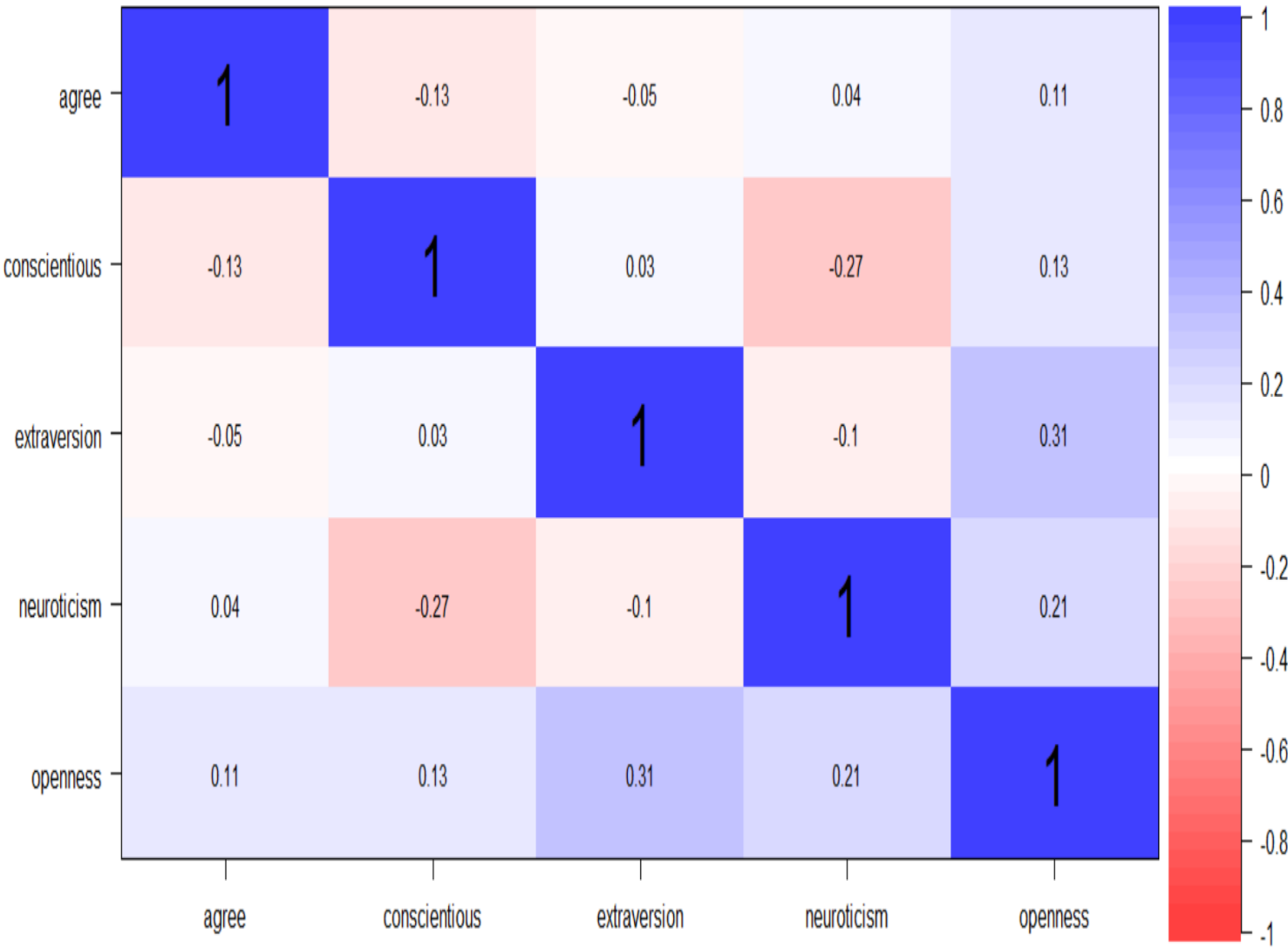
Correlation plot

Fig : 3 cor.plot(scores$scores)

(Human Qualities Correlation)

DISCUSSION

By working this project we have gathered knowledge about R programming language . We learnt about psychology and about graphs, plots . We learnt bi.bar , bar plot, plot , histogram , density . From the graph we can understand the correlation of the entities.

CONCLUSION

Psychoinformatics is an emerging interdisciplinary field that uses principles from computer science for the acquisition, organization, and synthesis of data collected from psychology to reveal information about psychological traits such as personality and mood. In future we are thinking to take 3 years of data from same people and see their mental difference .

REFERENCE LITERATURE

- How To: Install R and the psych package William Revelle Department of Psychology Northwestern University September 8, 2020
- An introduction to the psych package: Part I: data entry and data description William Revelle Department of Psychology Northwestern University September 24, 2022
- Recorded Behavior as a Valuable Resource for Diagnostics in Mobile Phone Addiction: Evidence from Psychoinformatics Christian Montag 1,*, Konrad Błaszkiewicz 2 , Bernd Lachmann 1 , Rayna Sariyska 1 , Ionut Andone 2 , Boris Trendafilov 2 and Alexander Markowetz 2, 9 October 2015
- Psycho-Informatics: Big Data shaping modern psychometrics Alexander Markowetz a , Konrad Błaszkiewicz a , Christian Montag b , Christina Switala c , Thomas E. Schlaepfer c 19 November 2013

Appendix A: R Source Code

>install.packages("psych")

#installing psych package

> install.packages("readr")

#installing for getting CSV file in R

>data2 <- read.csv("E:\\EDUCATION\\1st SEM\\R\\ND_PSYI_Data_Sem1_2022.CSV")

got csv file in data2 variable

> head(data2)

  A1 A2 A3 A4 A5 C1 C2 C3 C4 C5 E1 E2 E3 E4 E5 N1 N2 N3 N4 N5 O1 O2 O3 O4 O5 gender education age

1 6 3 2 6 2 4 6 2 4 6 1 4 2 6 6 6 1 6 4 3 4 6 6 2 4   2   3 17

2 1 2 6 6 2 5 5 6 1 1 3 2 4 5 4 2 1 4 2 1 6 5 3 1 2   1   3 18

3 5 2 5 6 4 5 6 5 1 4 2 3 5 5 6 4 1 1 1 4 4 2 3 1   1   3 19

4 1 1 4 6 4 6 6 2 1 4 4 1 3 6 4 6 3 6 2 5 5 5 6 5 5   2   3 18

5 6 6 3 6 5 1 6 4 1 5 1 5 5 6 6 2 1 5 6 6 6 3 5 6 1   2   3 18

6 6 1 3 6 4 6 6 6 1 1 3 6 2 2 6 6 3 6 6 6 6 6 2 6 1   1   4 53

> r.data2 <- cos(data2)

>head(r.data2)

       A1   A2   A3   A4   A5   C1   C2   C3   C4   C5   E1   E2
E3   E4

1 0.9601703 -0.9899925 -0.4161468 0.9601703 -0.4161468 -0.6536436 0.9601703 -0.4161468 -0.6536436 0.9601703 0.5403023 -0.6536436 -0.4161468 0.9601703

2 0.5403023 -0.4161468 0.9601703 0.9601703 -0.4161468 0.2836622 0.2836622 0.9601703 0.5403023 0.5403023 -0.9899925 -0.4161468 -0.6536436 0.2836622

3 0.2836622 -0.4161468 0.2836622 0.9601703 -0.6536436 0.2836622 0.9601703 0.2836622 0.5403023 -0.6536436 -0.4161468 -0.9899925 0.2836622 0.2836622

4 0.5403023 0.5403023 -0.6536436 0.9601703 -0.6536436 0.9601703 0.9601703 -0.4161468 0.5403023 -0.6536436 -0.6536436 0.5403023 -0.9899925 0.9601703

5 0.9601703 0.9601703 -0.9899925 0.9601703 0.2836622 0.5403023 0.9601703 -0.6536436 0.5403023 0.2836622 0.5403023 0.2836622 0.2836622 0.9601703

6 0.9601703 0.5403023 -0.9899925 0.9601703 -0.6536436 0.9601703 0.9601703 0.9601703 0.5403023 0.5403023 -0.9899925 0.9601703 -0.4161468 -0.4161468

     E5   N1   N2   N3   N4   N5   O1   O2   O3   O4   O5   gender education

1 0.9601703 0.9601703 0.5403023 0.9601703 -0.6536436 -0.9899925 -0.6536436 0.9601703 0.9601703 -0.4161468 -0.6536436 -0.4161468 -0.9899925

2 -0.6536436 -0.4161468 0.5403023 -0.6536436 -0.4161468 0.5403023 0.9601703 0.2836622 -0.9899925 0.5403023 -0.4161468 0.5403023 -0.9899925

3 0.9601703 -0.6536436 0.5403023 0.5403023 0.5403023 0.5403023 -0.6536436 -0.6536436 -0.4161468 -0.9899925 0.5403023 0.5403023 -0.9899925

4 -0.6536436 0.9601703 -0.9899925 0.9601703 -0.4161468 0.2836622 0.2836622 0.2836622 0.9601703 0.2836622 0.2836622 -0.4161468 -0.9899925

5 0.9601703 -0.4161468 0.5403023 0.2836622 0.9601703 0.9601703 0.9601703 -0.9899925 0.2836622 0.9601703 0.5403023 -0.4161468 -0.9899925

6 0.9601703 0.9601703 -0.9899925 0.9601703 0.9601703 0.9601703 0.9601703 0.9601703 -0.4161468 0.9601703 0.5403023 0.5403023 -0.6536436

    age

1 -0.2751633

2 0.6603167

3 0.9887046

4 0.6603167

5 0.6603167

6 -0.9182828

>library(corrplot)

> m<- r.data2

>corrplot(m)

#getting result of Fig 2 plot

>corrplot(m , method = 'numbers') #getting the Fig 1 plot

>keys.list <- list(agree=c("-A1","A2","A3","A4","A5"),

conscientious=c("C1","C2","C3","-C4","-C5"),

extraversion=c("-E1","-E2","E3","E4","E5"),

neuroticism=c("N1","N2","N3","N4","N5"),

openness = c("O1","-O2","O3","O4","-O5"))

> keys <- make.keys(M,keys.list)

 >scores <- scoreItems(keys,M,totals="TRUE")

>cor(scores$scores)

>cor.plot(scores$scores) #plotting the main heatmap