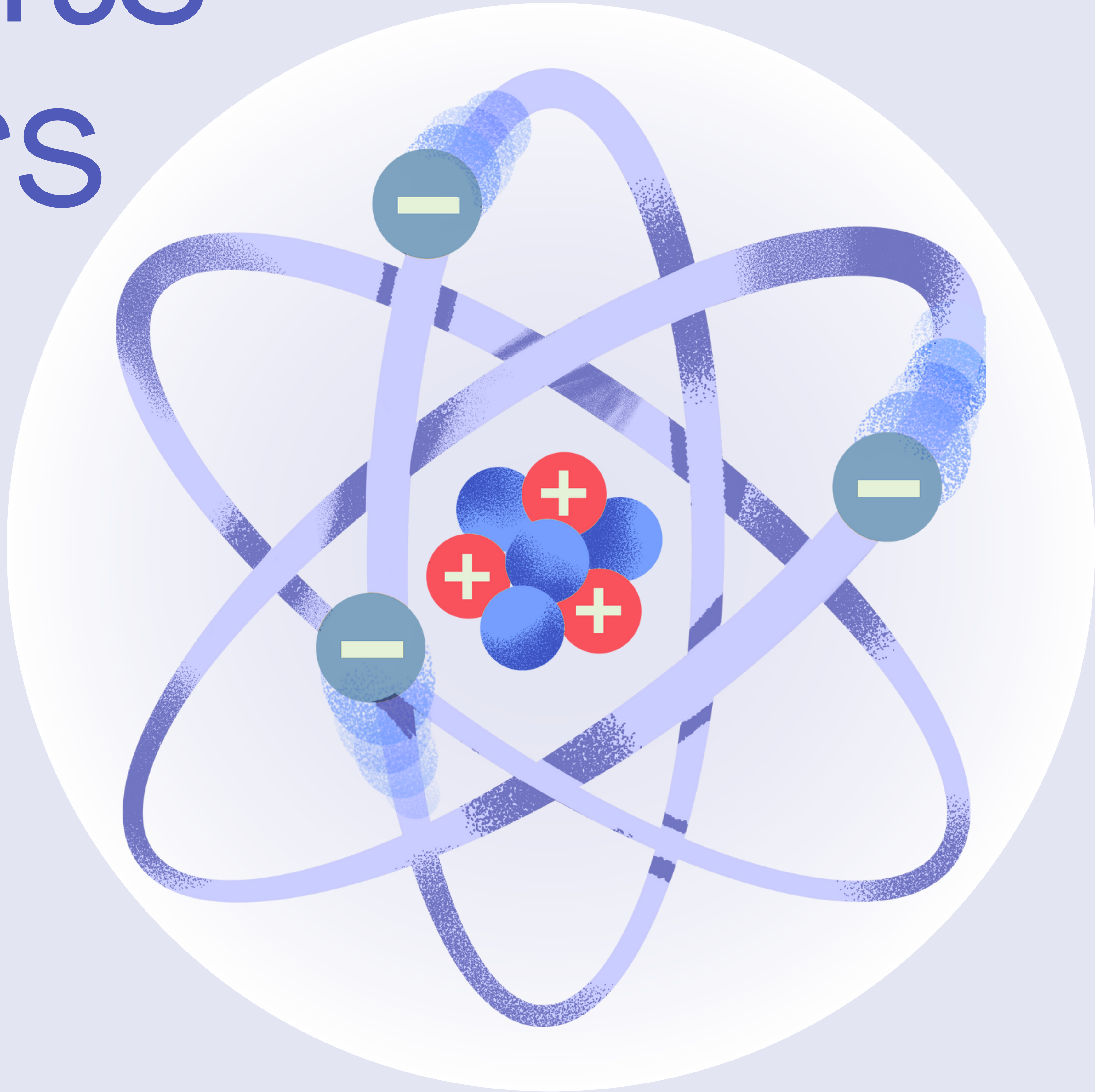
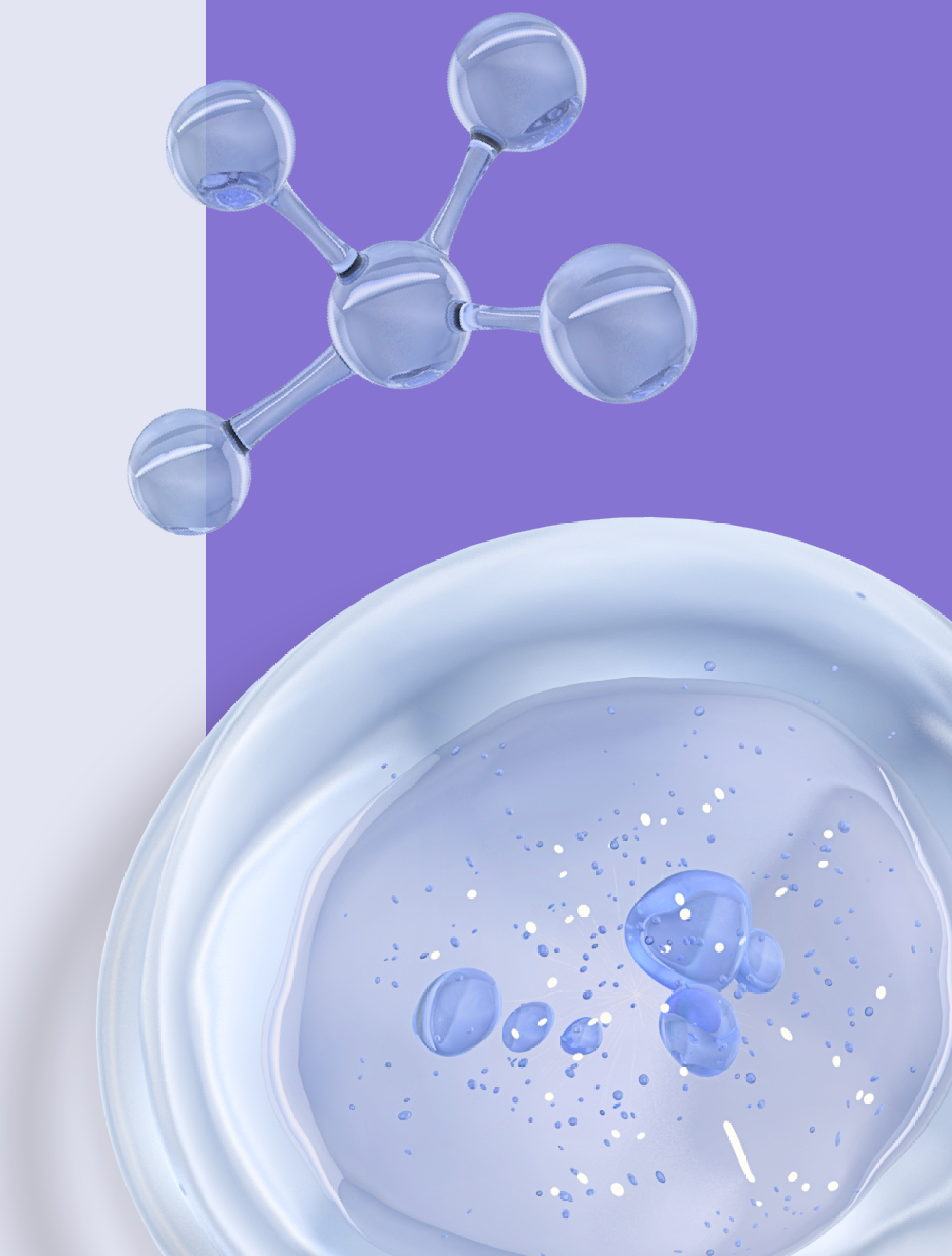
Personality traits and drug users



Project by Ait Taleb Naser Ikram

Unsupervised Learning



Objectives

The goal of the project is to examine and cluster different personalities based on the five-factor model (FFM).
In psychology,this theory describes human personalities based on five factors:

- Openness to experience
- Conscientiousness
- Extroversion
- Agreeableness
- Neuroticism

Given the complexities of human psychology, it may be difficult to capture all the nuances of an individual's personality,the goal of this part of the project is to compare different clusters on the 874.366 observations and analyze the five traits patterns within each cluster .

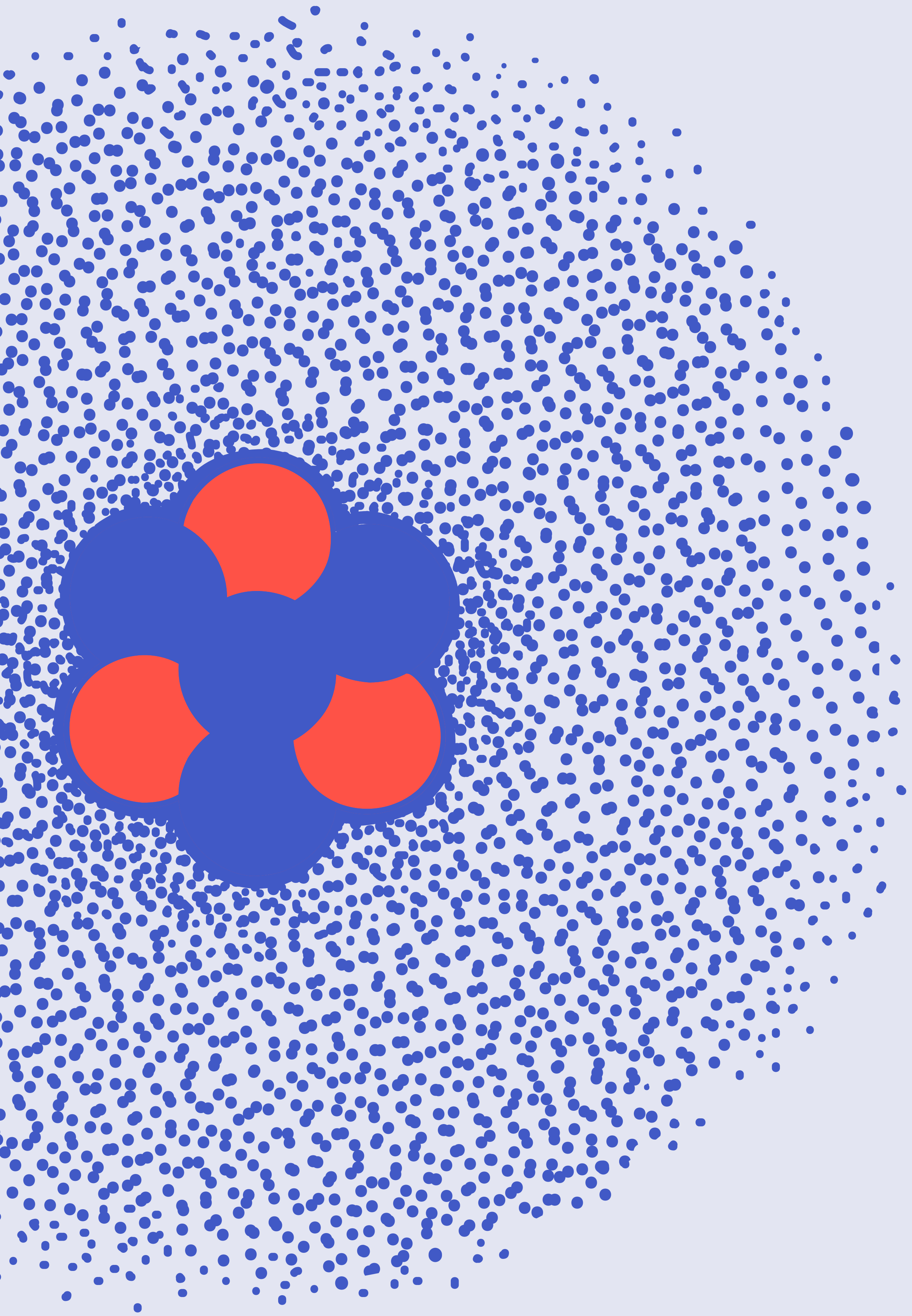


TABLE OF CONTENTS:

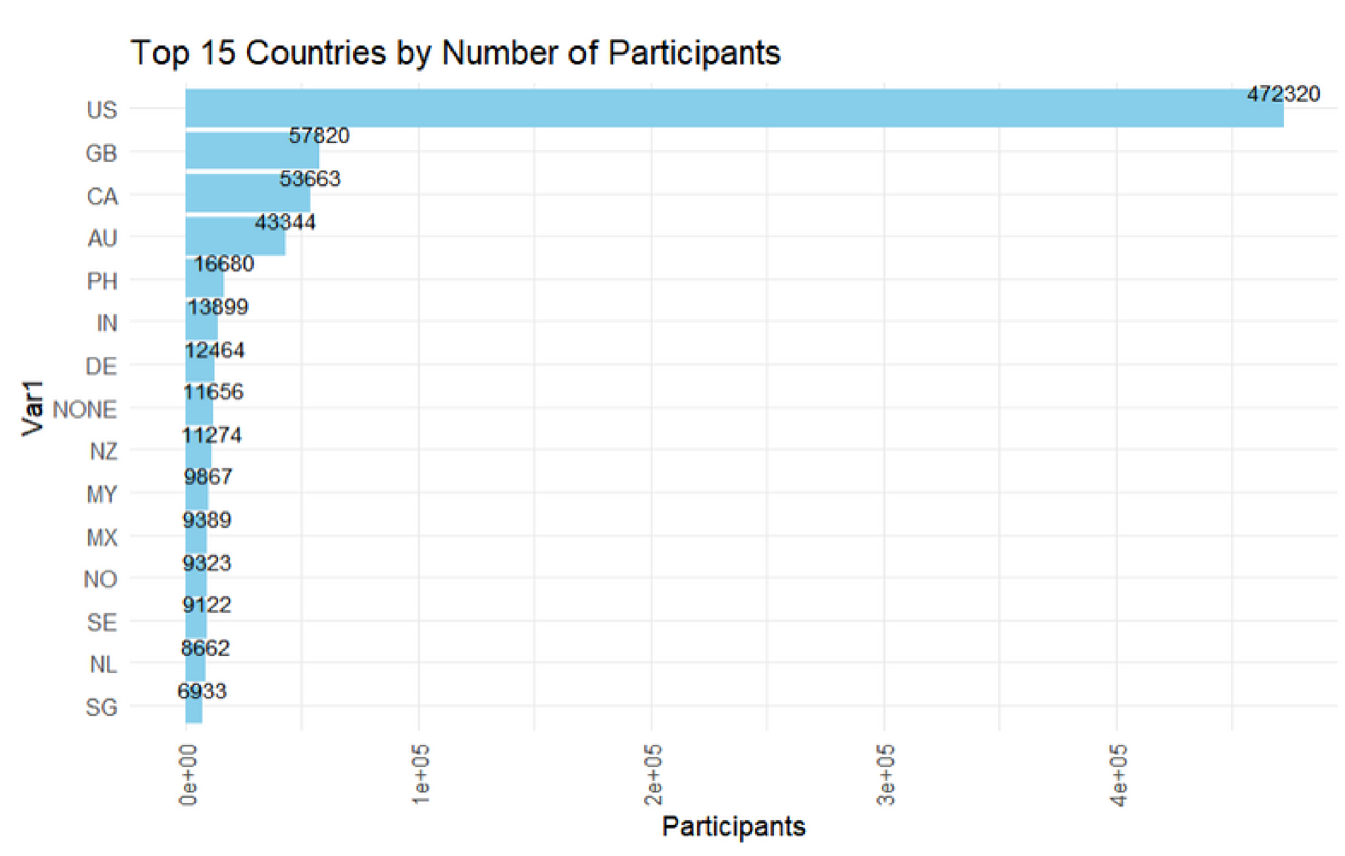
The Dataset

the questionnaire has 50 questions, each answer was rated on a five point scale.
The scale was labeled as following:

1=Disagree,
2= Slightly disagree,
3=Neutral,
4= Slightly agree,
5=Agree

questionnaire link:
https://openpsychometrics.org/tests/IPIP-BFFM/

```
'data.frame':   874366 obs. of  51 variables:
 $ EXT1    : num  4 3 2 2 3 3 4 3 1 3 ...
 $ EXT2    : num  1 5 3 2 3 3 3 1 5 3 ...
 $ EXT3    : num  5 3 4 2 3 4 4 5 3 2 ...
 $ EXT4    : num  2 4 4 3 3 2 3 2 5 3 ...
 $ EXT5    : num  5 3 3 4 5 4 3 5 2 3 ...
 $ EXT6    : num  1 3 2 2 3 2 3 2 3 2 ...
 $ EXT7    : num  5 2 1 2 3 2 5 5 2 4 ...
 $ EXT8    : num  2 5 3 4 5 3 3 2 4 3 ...
 $ EXT9    : num  4 1 2 1 3 3 4 3 5 3 ...
 $ EXT10   : num  1 5 5 4 4 4 3 2 4 5 ...
 $ EST1    : num  1 2 4 3 1 3 2 2 3 4 ...
 $ EST2    : num  4 3 4 3 5 4 4 4 3 3 ...
 $ EST3    : num  4 4 4 3 5 3 4 2 3 4 ...
 $ EST4    : num  2 1 2 2 3 2 2 4 3 1 ...
 $ EST5    : num  2 3 2 3 1 2 4 2 4 2 ...
 $ EST6    : num  2 1 2 2 1 1 2 1 3 3 ...
 $ EST7    : num  2 2 2 2 1 2 2 2 3 5 ...
 $ EST8    : num  2 1 2 2 1 1 2 1 3 4 ...
 $ EST9    : num  3 3 1 4 3 2 4 1 3 4 ...
 $ EST10   : num  2 1 3 3 2 2 4 1 3 5 ...
 $ AGR1    : num  2 1 1 2 1 2 1 2 5 2 ...
 $ AGR2    : num  5 4 4 4 5 3 2 5 3 5 ...
 $ AGR3    : num  2 1 1 3 1 1 1 2 5 3 ...
 $ AGR4    : num  4 5 4 4 5 4 5 4 1 4 ...
 $ AGR5    : num  2 1 2 2 1 2 3 3 5 2 ...
 $ AGR6    : num  3 5 4 4 3 3 5 2 3 3 ...
 $ AGR7    : num  2 3 1 2 1 2 3 2 4 1 ...
 $ AGR8    : num  4 4 4 4 5 3 4 4 2 3 ...
 $ AGR9    : num  3 5 4 3 5 4 4 4 3 4 ...
 $ AGR10   : num  4 3 3 4 3 4 5 4 2 2 ...
 $ CSN1    : num  3 3 4 2 5 3 3 5 2 1 ...
 $ CSN2    : num  4 2 2 4 1 2 2 1 5 5 ...
 $ CSN3    : num  3 5 2 4 5 4 4 5 1 5 ...
 $ CSN4    : num  2 3 2 4 1 1 2 1 5 5 ...
 $ CSN5    : num  2 3 3 1 3 3 1 4 1 1 ...
 $ CSN6    : num  4 1 3 2 1 2 4 2 4 5 ...
 $ CSN7    : num  4 3 4 2 5 4 4 3 3 3 ...
 $ CSN8    : num  2 3 2 3 1 3 2 2 4 1 ...
 $ CSN9    : num  4 5 4 1 5 4 2 5 2 1 ...
 $ CSN10   : num  4 3 2 4 5 3 5 5 2 5 ...
 $ OPN1    : num  5 1 5 4 5 5 5 4 3 5 ...
 $ OPN2    : num  1 2 1 2 1 1 2 1 1 1 ...
 $ OPN3    : num  4 4 2 5 5 5 4 3 3 5 ...
 $ OPN4    : num  1 2 1 2 1 1 3 1 1 1 ...
 $ OPN5    : num  4 3 4 3 5 3 4 5 3 5 ...
 $ OPN6    : num  1 1 2 1 1 1 1 1 3 1 ...
 $ OPN7    : num  5 4 5 4 5 5 5 4 4 5 ...
 $ OPN8    : num  3 2 3 4 3 4 5 3 3 5 ...
 $ OPN9    : num  4 5 4 3 5 5 4 4 3 5 ...
 $ OPN10   : num  5 3 4 3 5 2 4 5 3 5 ...
 $ country: chr  "GB" "MY" "GB" "GB" ...
```



This dataset is not balanced in countries, so a bias towards Western cultures may not accurately represent personality in other cultures.

THE QUESTIONS

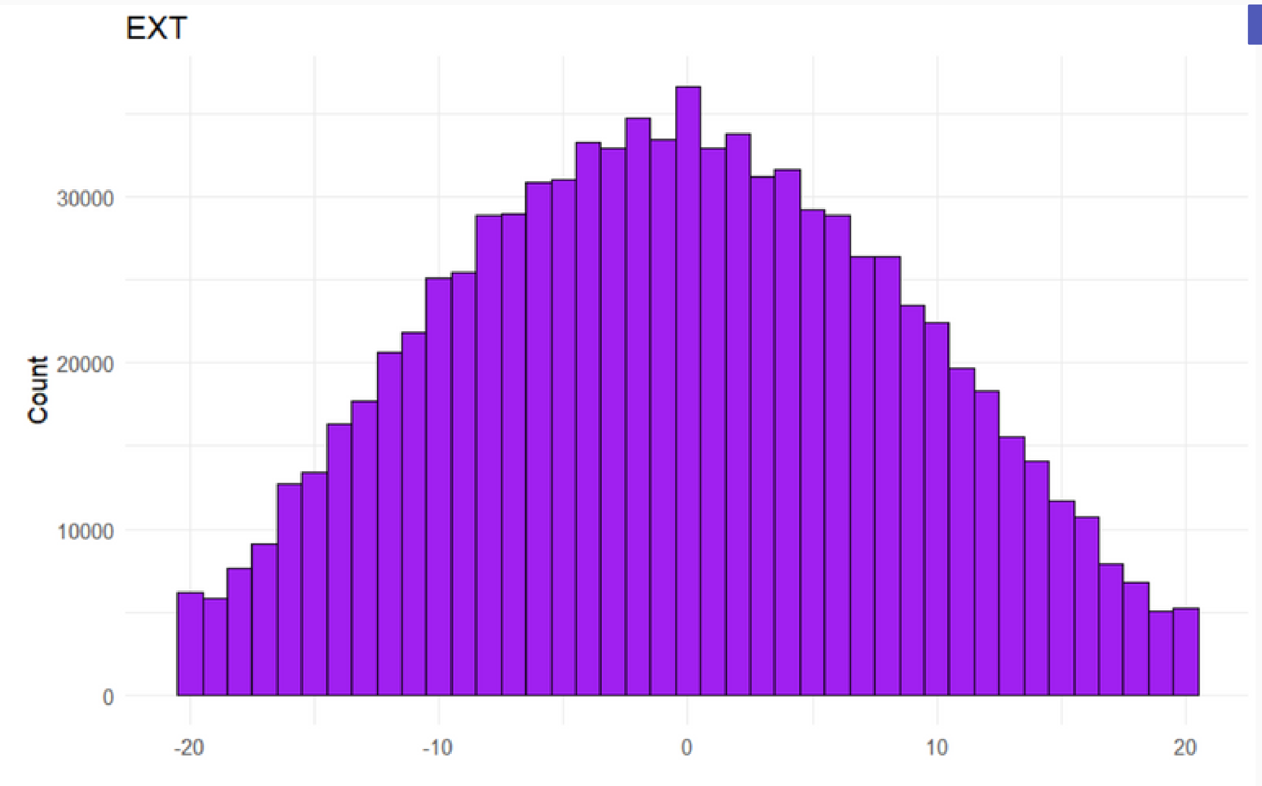
Code	Question	Key	Code	Question	Key
EXT1	I am the life of the party.	(+)	EST1	I get stressed out easily.	(+)
EXT2	I don't talk a lot.	(-)	EST2	I am relaxed most of the time.	(-)
EXT3	I feel comfortable around people.	(+)	EST3	I worry about things.	(+)
EXT4	I keep in the background.	(-)	EST4	I seldom feel blue.	(-)
EXT5	I start conversations.	(+)	EST5	I am easily disturbed.	(+)
EXT6	I have little to say.	(-)	EST6	I get upset easily.	(+)
EXT7	I talk to a lot of different people at parties.	(+)	EST7	I change my mood a lot.	(+)
EXT8	I don't like to draw attention to myself.	(-)	EST8	I have frequent mood swings.	(+)
EXT9	I don't mind being the center of attention.	(+)	EST9	I get irritated easily.	(+)
EXT10	I am quiet around strangers.	(-)	EST10	I often feel blue.	(+)

Code	Question	Key	Code	Question	Key
AGR1	I feel little concern for others.	(-)	CSN1	I am always prepared.	(+)
AGR2	I am interested in people.	(+)	CSN2	I leave my belongings around.	(-)
AGR3	I insult people.	(-)	CSN3	I pay attention to details.	(+)
AGR4	I sympathize with others' feelings.	(+)	CSN4	I make a mess of things.	(-)
AGR5	I am not interested in other people's problems.	(-)	CSN5	I get chores done right away.	(+)
AGR6	I have a soft heart.	(+)	CSN6	I often forget to put things back in their proper place.	(-)
AGR7	I am not really interested in others.	(-)	CSN7	I like order.	(+)
AGR8	I take time out for others.	(+)	CSN8	I shirk my duties.	(-)
AGR9	I feel others' emotions.	(+)	CSN9	I follow a schedule.	(+)
AGR10	I make people feel at ease.	(+)	CSN10	I am exacting in my work.	(+)

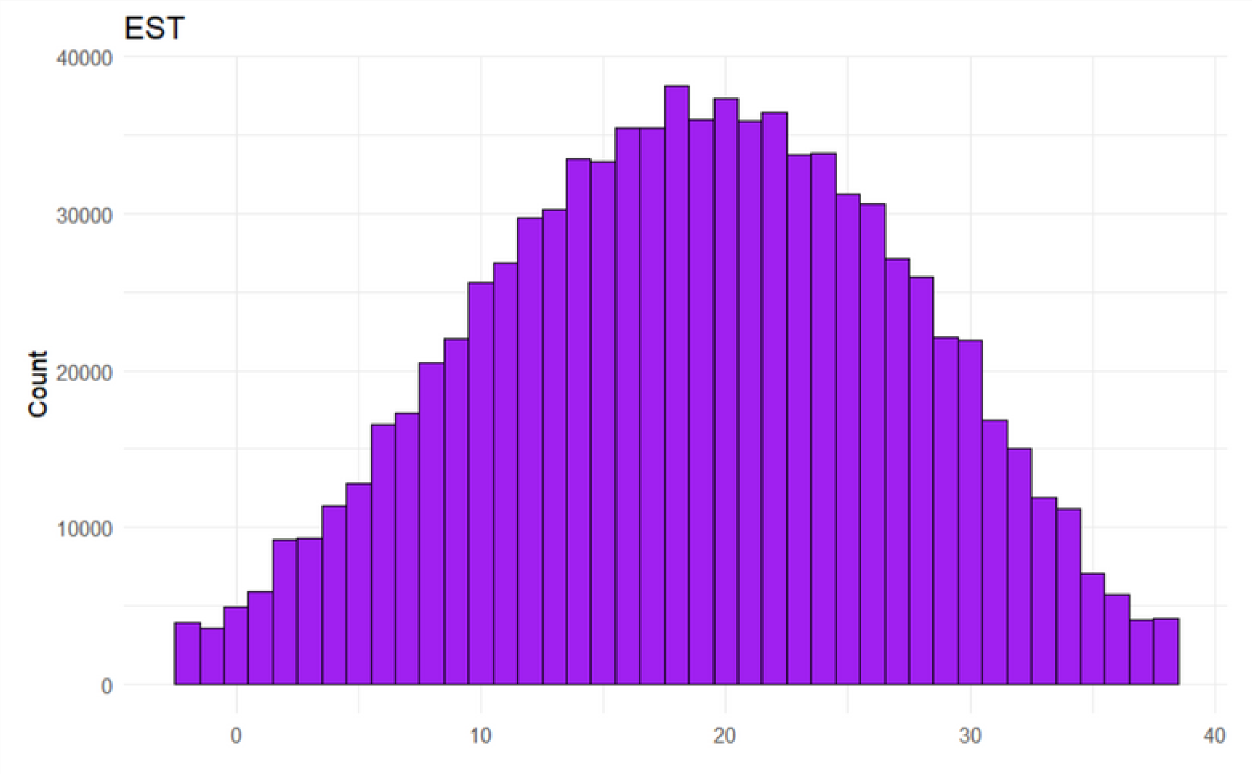
Code	Question	Key
OPN1	I have a rich vocabulary.	(+)
OPN2	I have difficulty understanding abstract ideas.	(-)
OPN3	I have a vivid imagination.	(+)
OPN4	I am not interested in abstract ideas.	(-)
OPN5	I have excellent ideas.	(+)
OPN6	I do not have a good imagination.	(-)
OPN7	I am quick to understand things.	(+)
OPN8	I use difficult words.	(-)
OPN9	I spend time reflecting on things.	(+)
OPN10	I am full of ideas.	(+)

According to the test documentation, questions can be positive keyed or negative keyed. Personality traits score are obtained by aggregating answers.

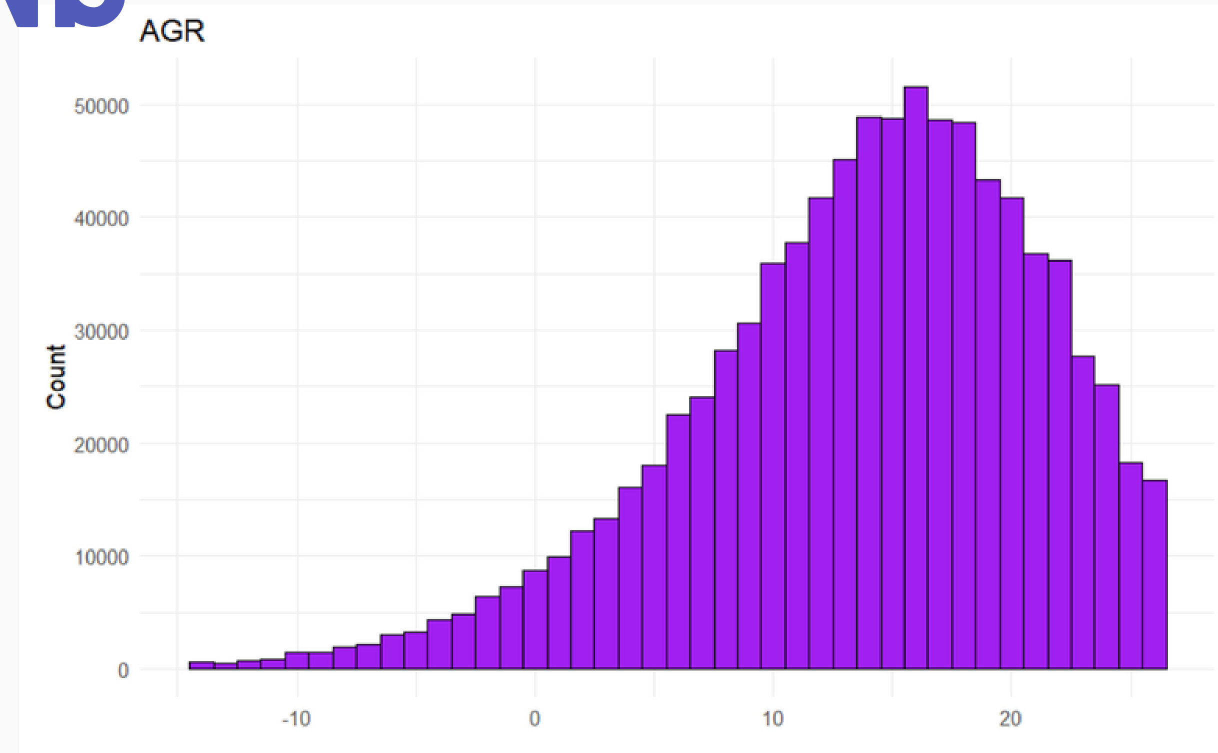
SCORE DISTRIBUTIONS



EXTROVERSION

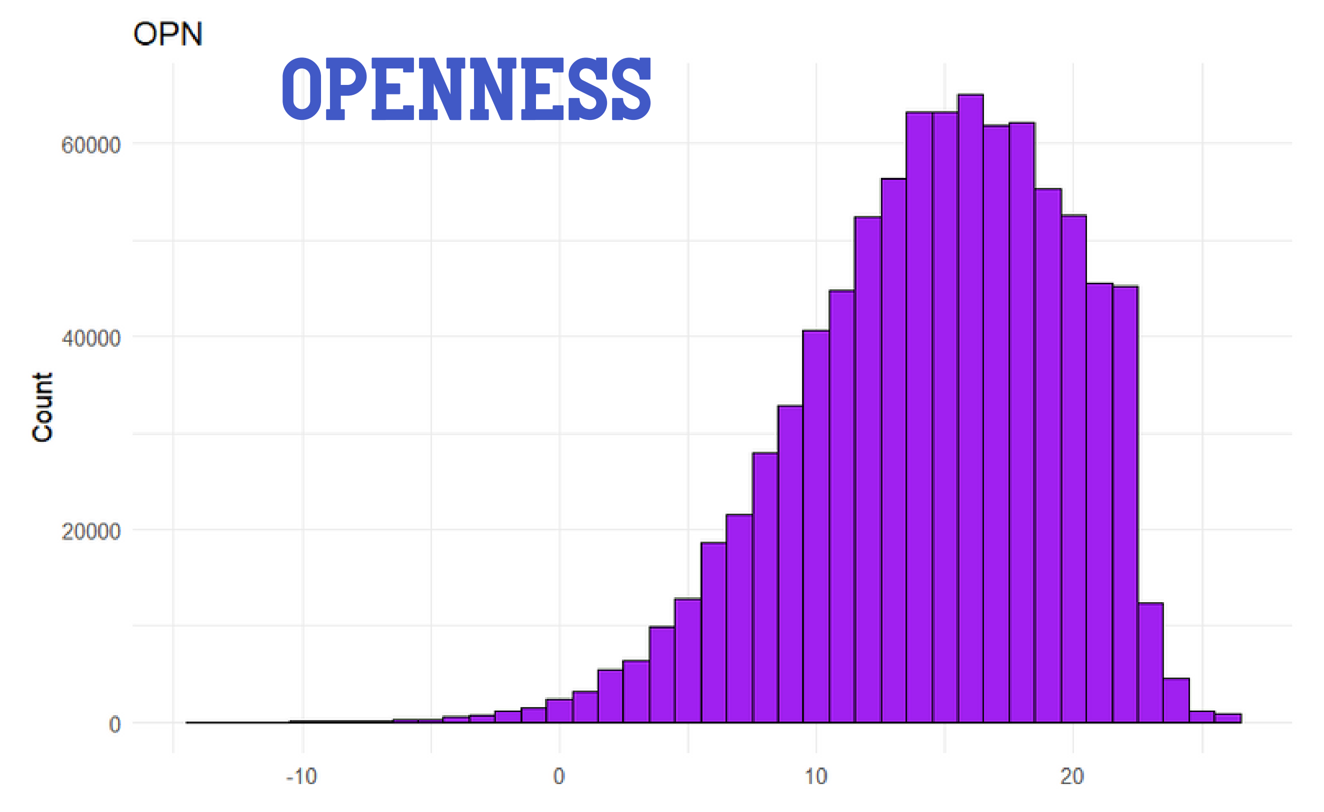


NEUROTICISM

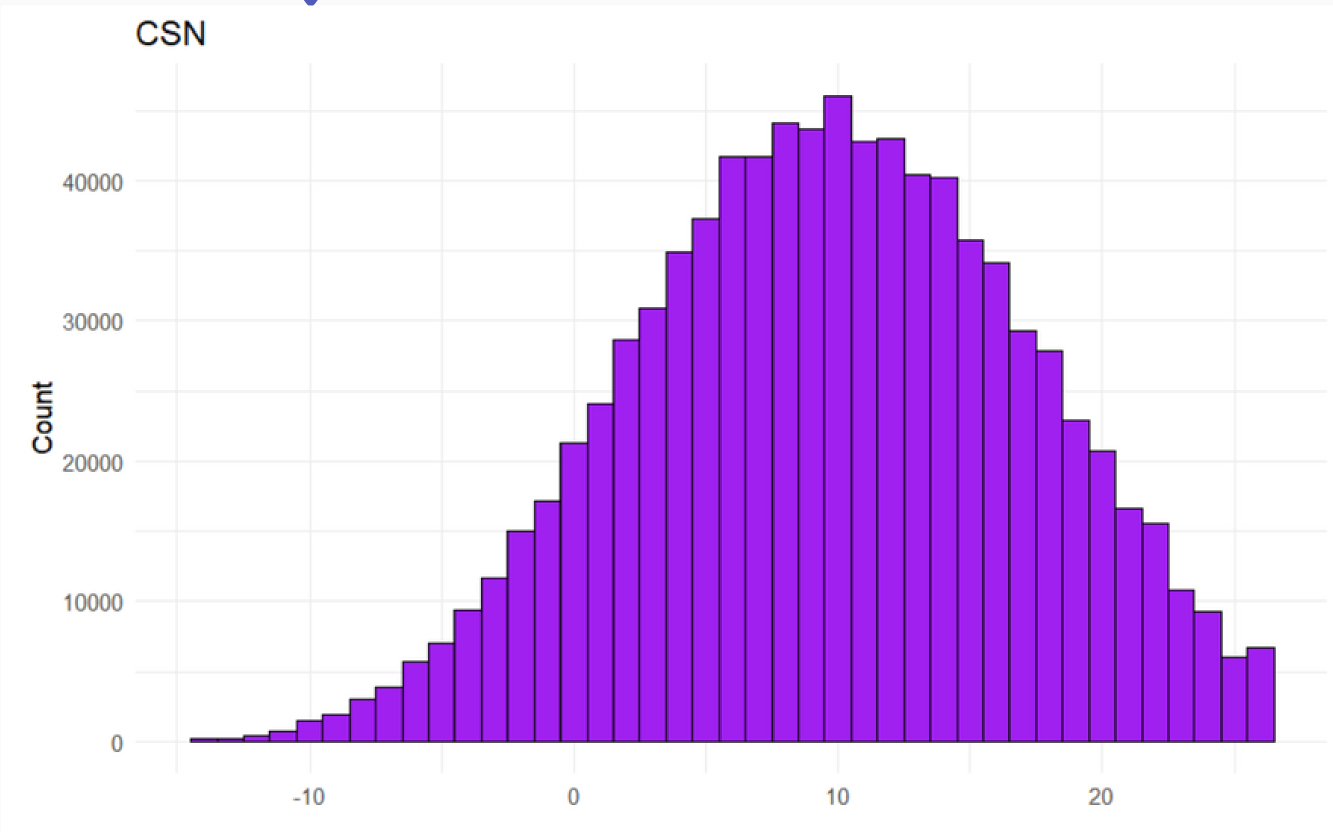


AGREEABLENESS

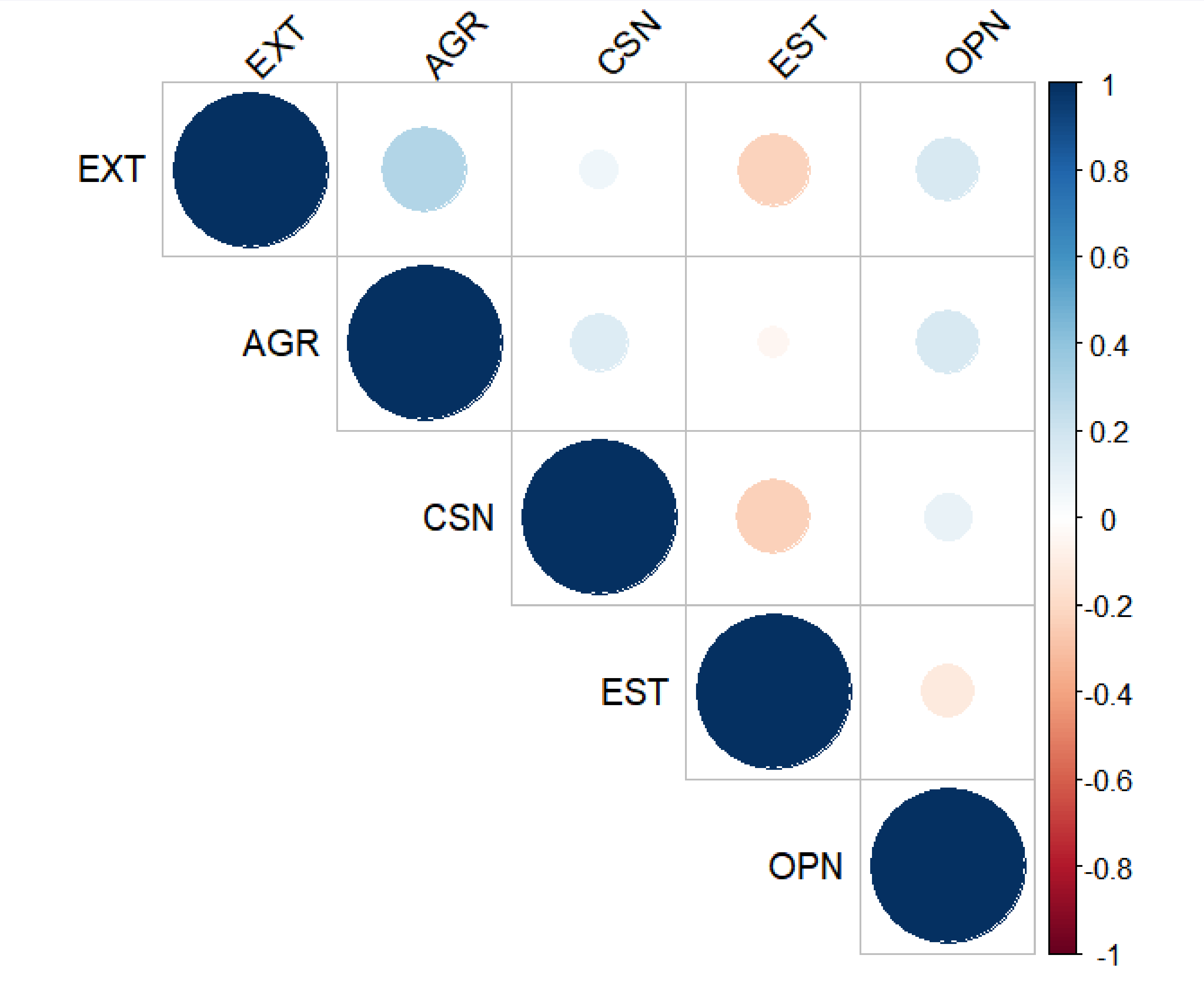
OPENNESS

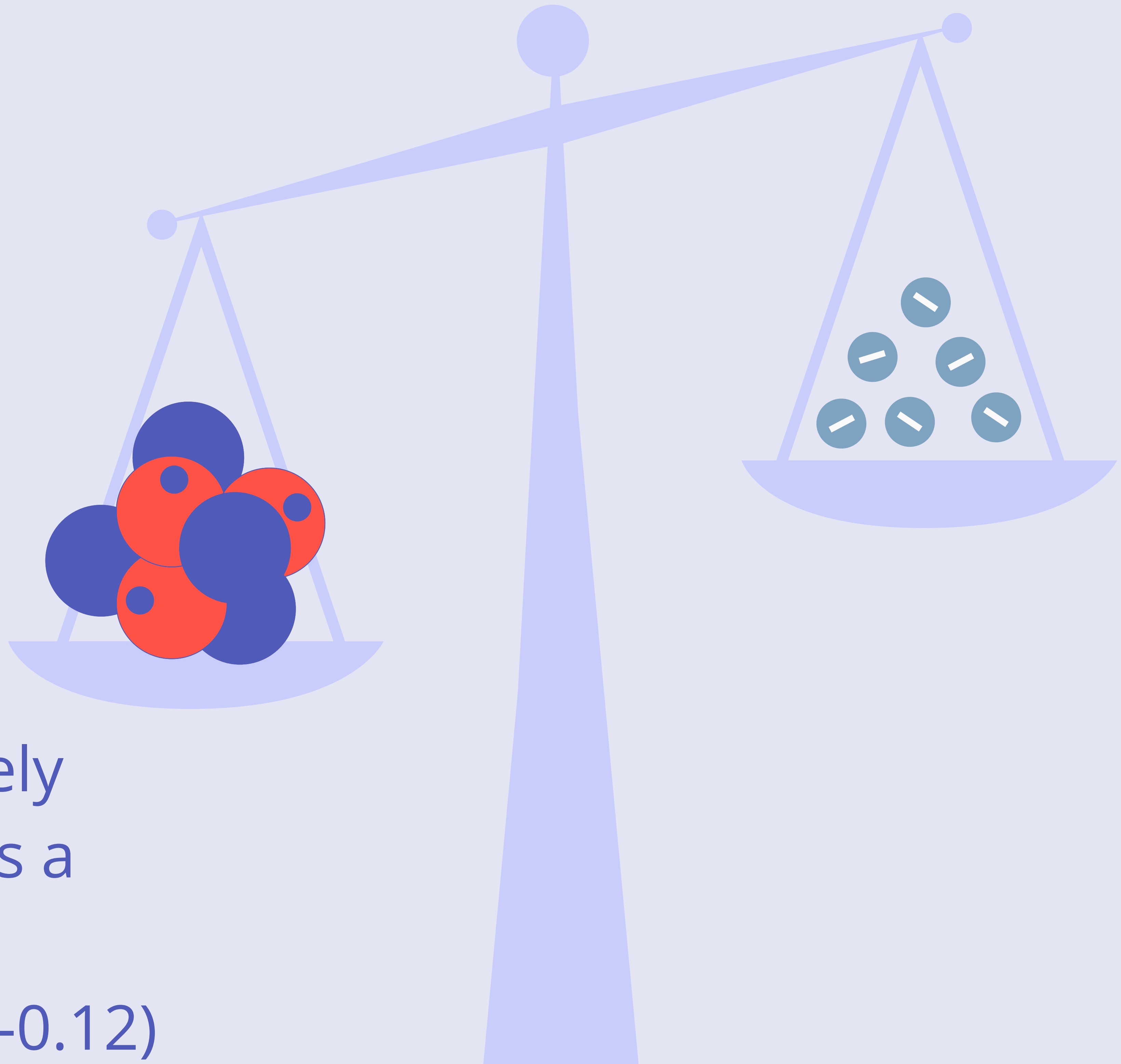


CONSCIENTIOUSNESS



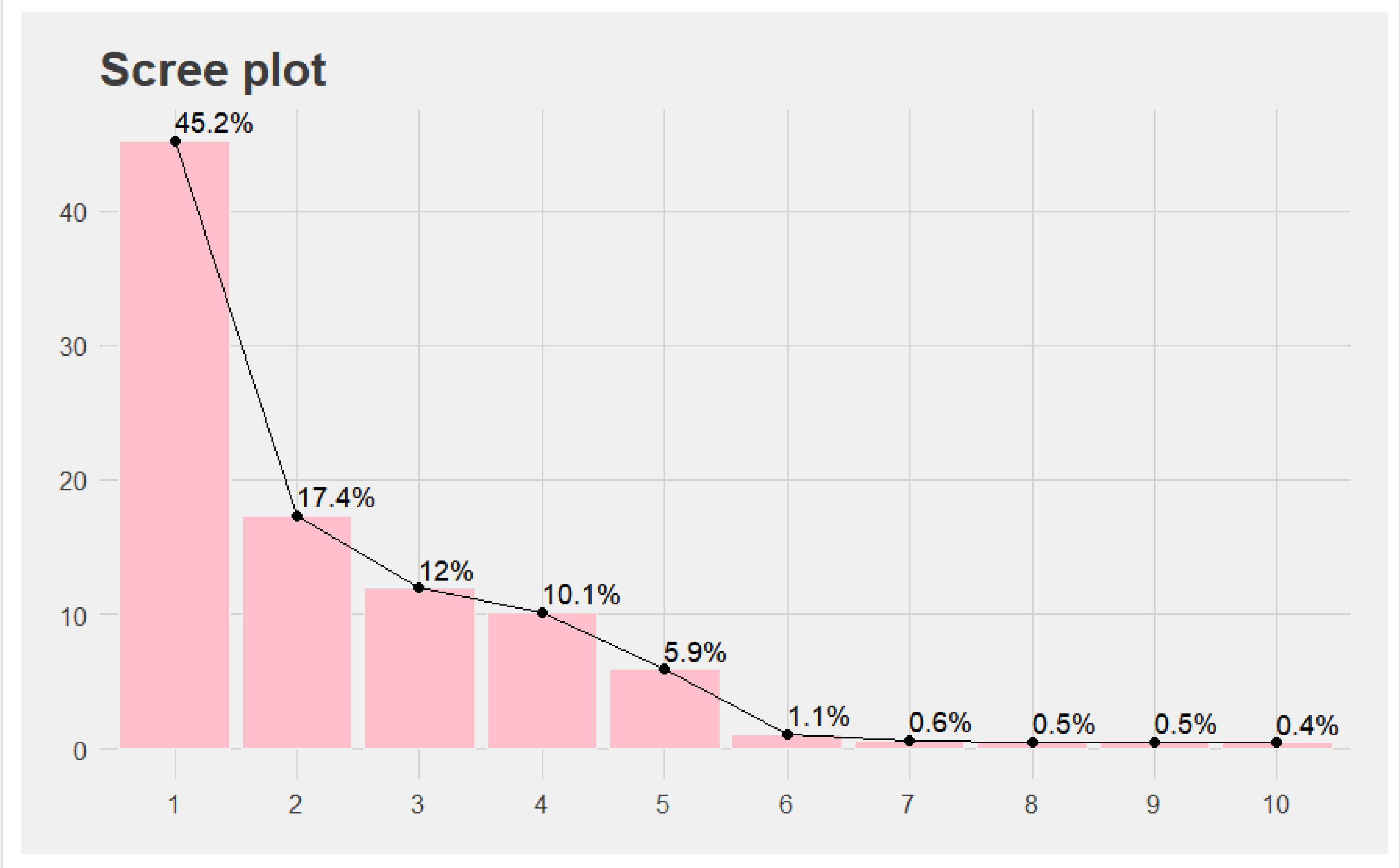
IS THERE ANY CORRELATION BETWEEN THE TRAITS?

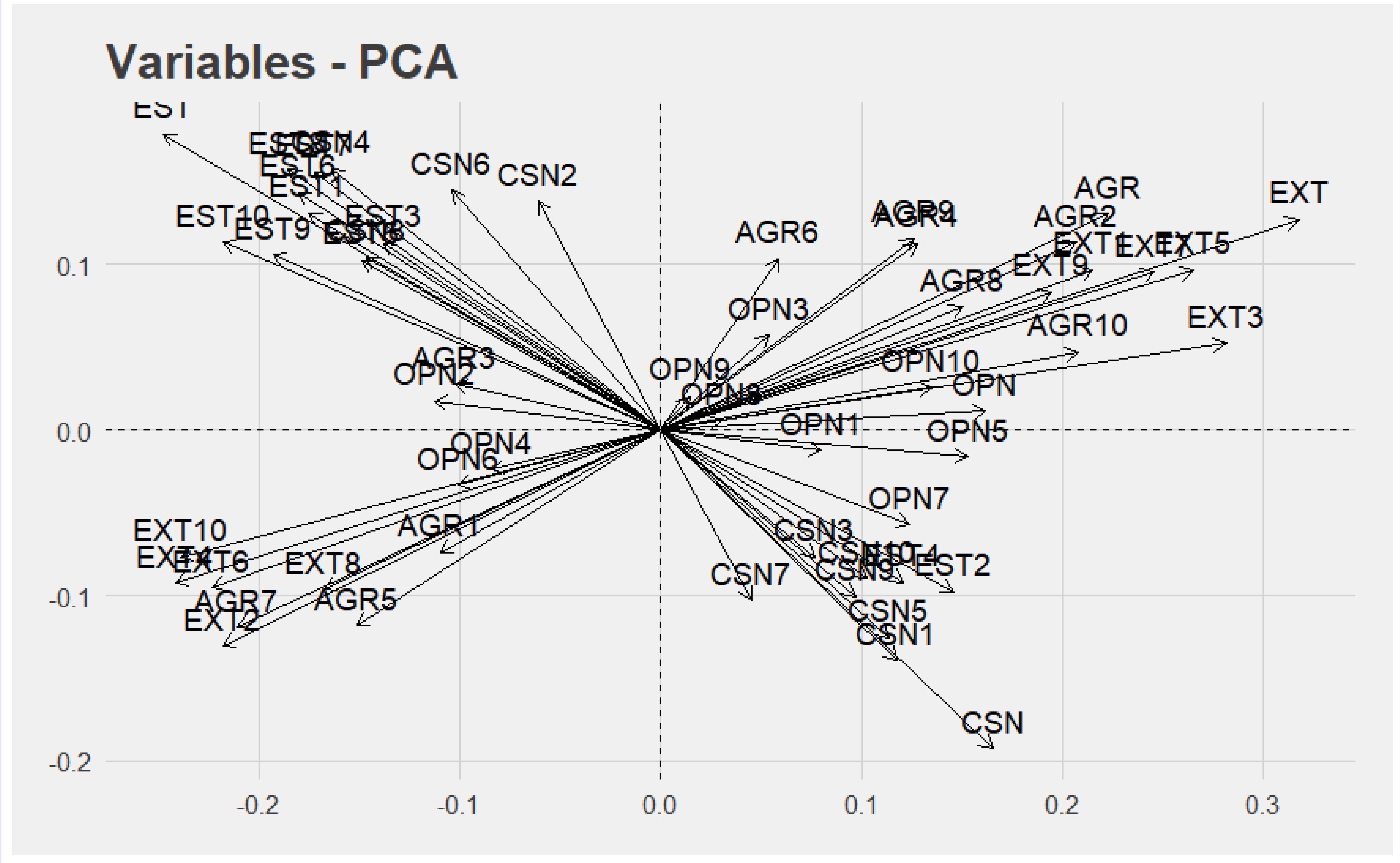


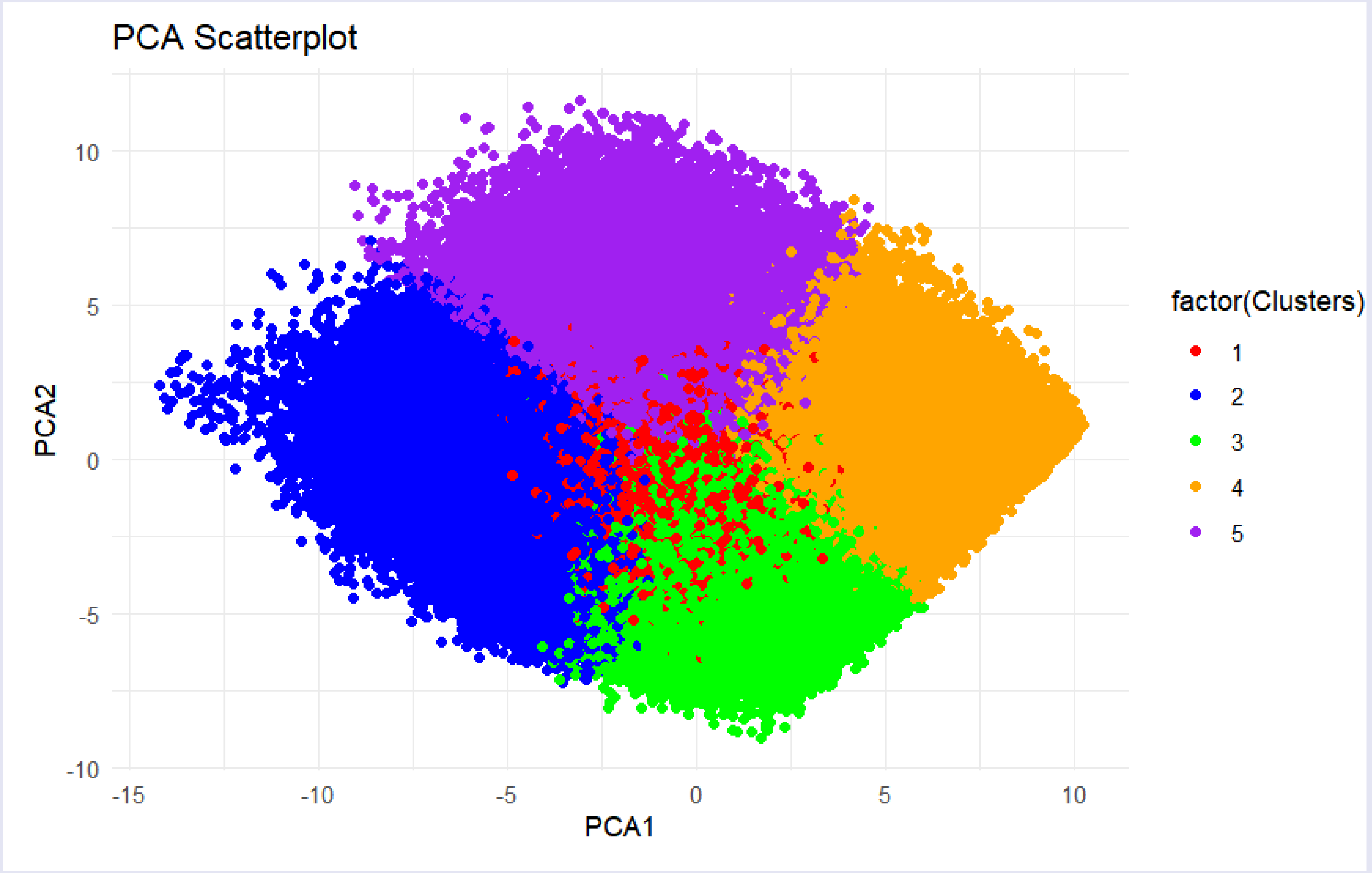


Extroversion and Agreeableness are positively correlated (+0.30). However, Neuroticism has a negative correlation with Extroversion, Conscientiousness and Openness (-0.22, -0.23, -0.12)

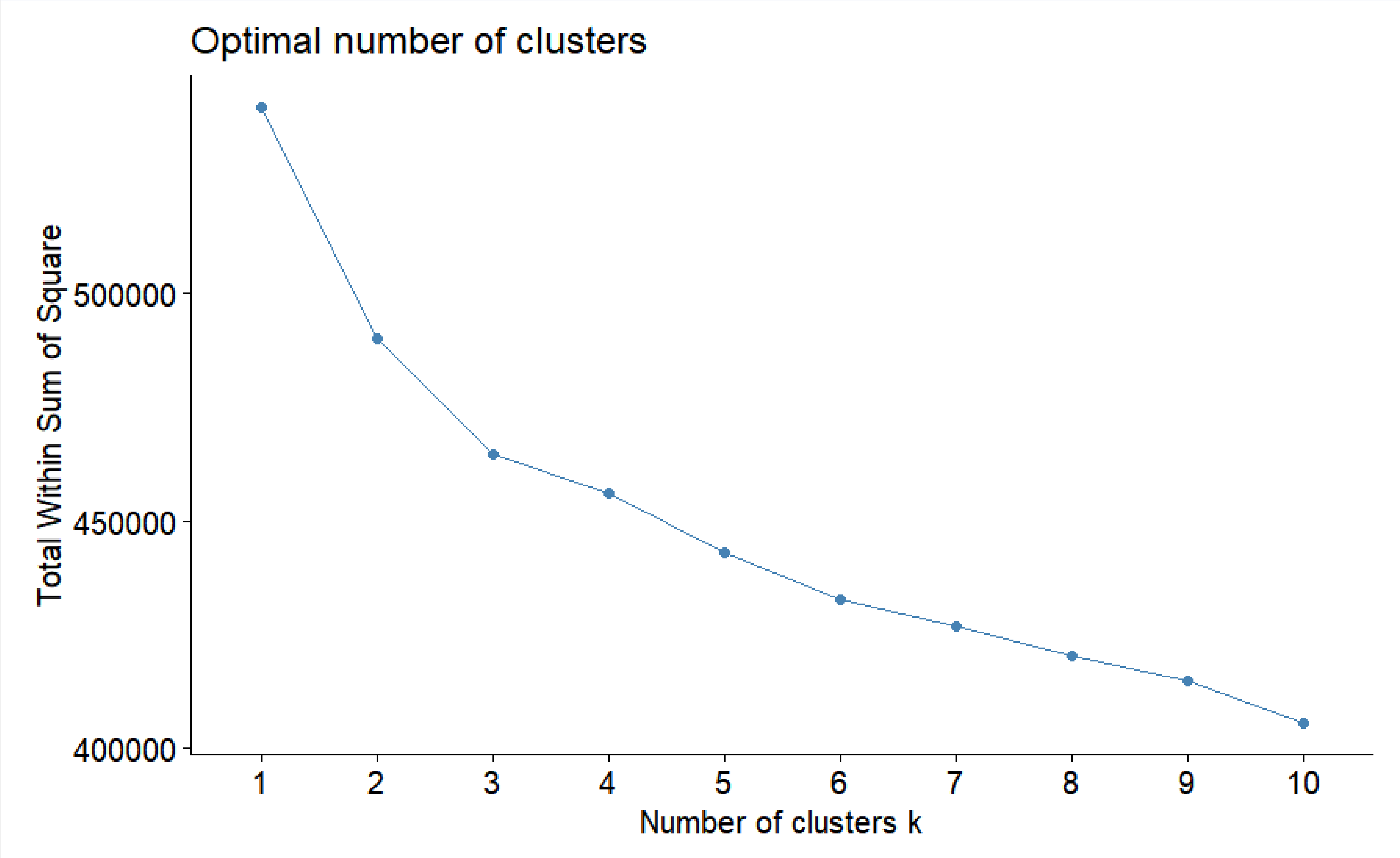
PCA

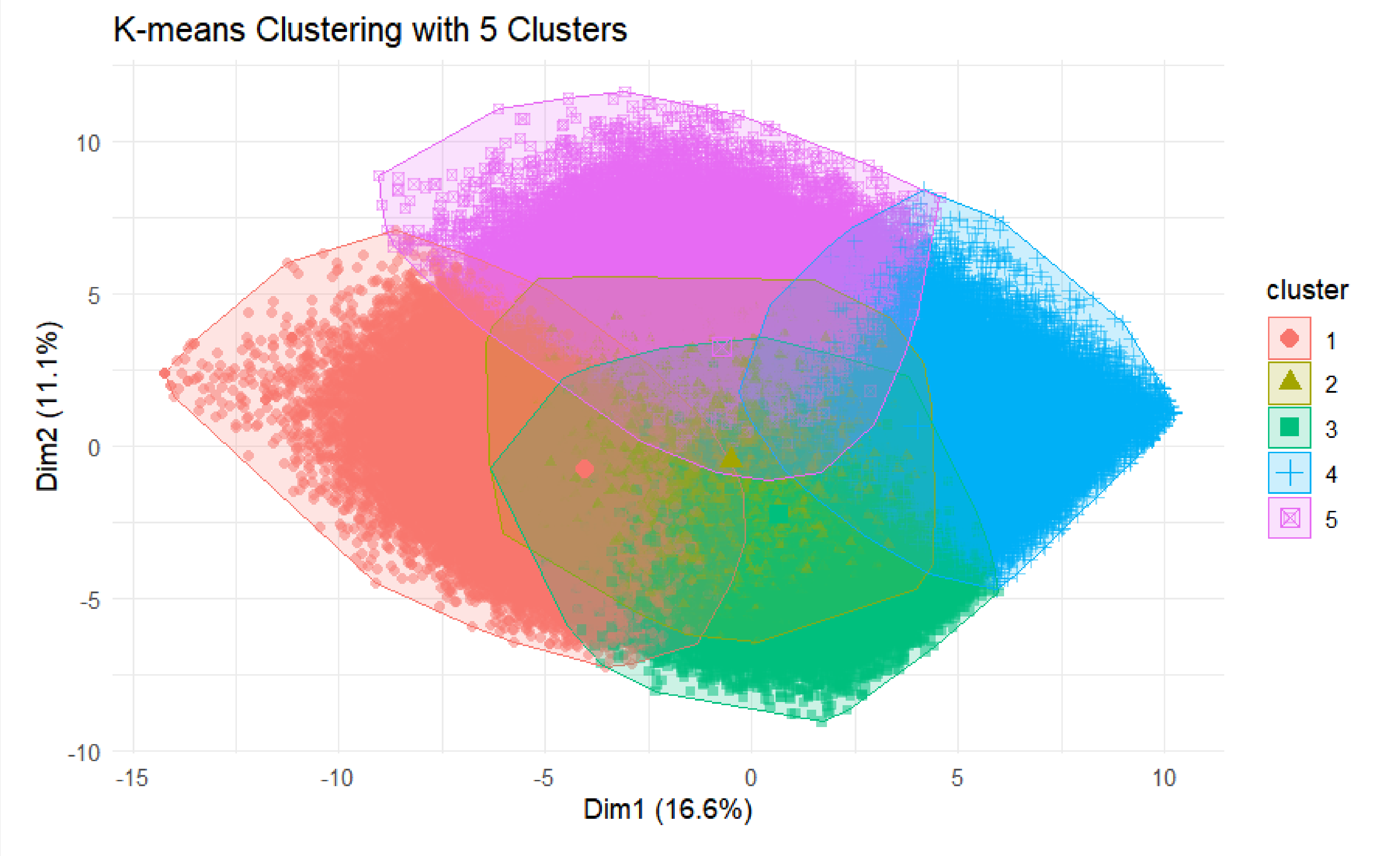

Scree plot
45.2%
17.4%
12%
10.1%
5.9%
1.1%
0.6%
0.5%
0.5%
0.4%
0
10
20
30
40
1 2 3 4 5 6 7 8 9 10


Variables - PCA
EXT
AGR
AGR2
AGR4
AGR6
AGR8
AGR10
EXT3
EXT5
EXT9
CSN6
CSN2
EST10
EST9
EST1
EST3
EST6
OPN3
OPN9
OPN10
OPN
OPN1
OPN5
OPN2
AGR3
OPN4
OPN6
OPN7
EXT10
EXT6
EXT8
AGR1
AGR7
AGR5
EXT2
CSN3
CSN7
EST2
CSN5
CSN1
CSN
0.1
0.0
-0.1
-0.2
-0.2 -0.1 0.0 0.1 0.2 0.3

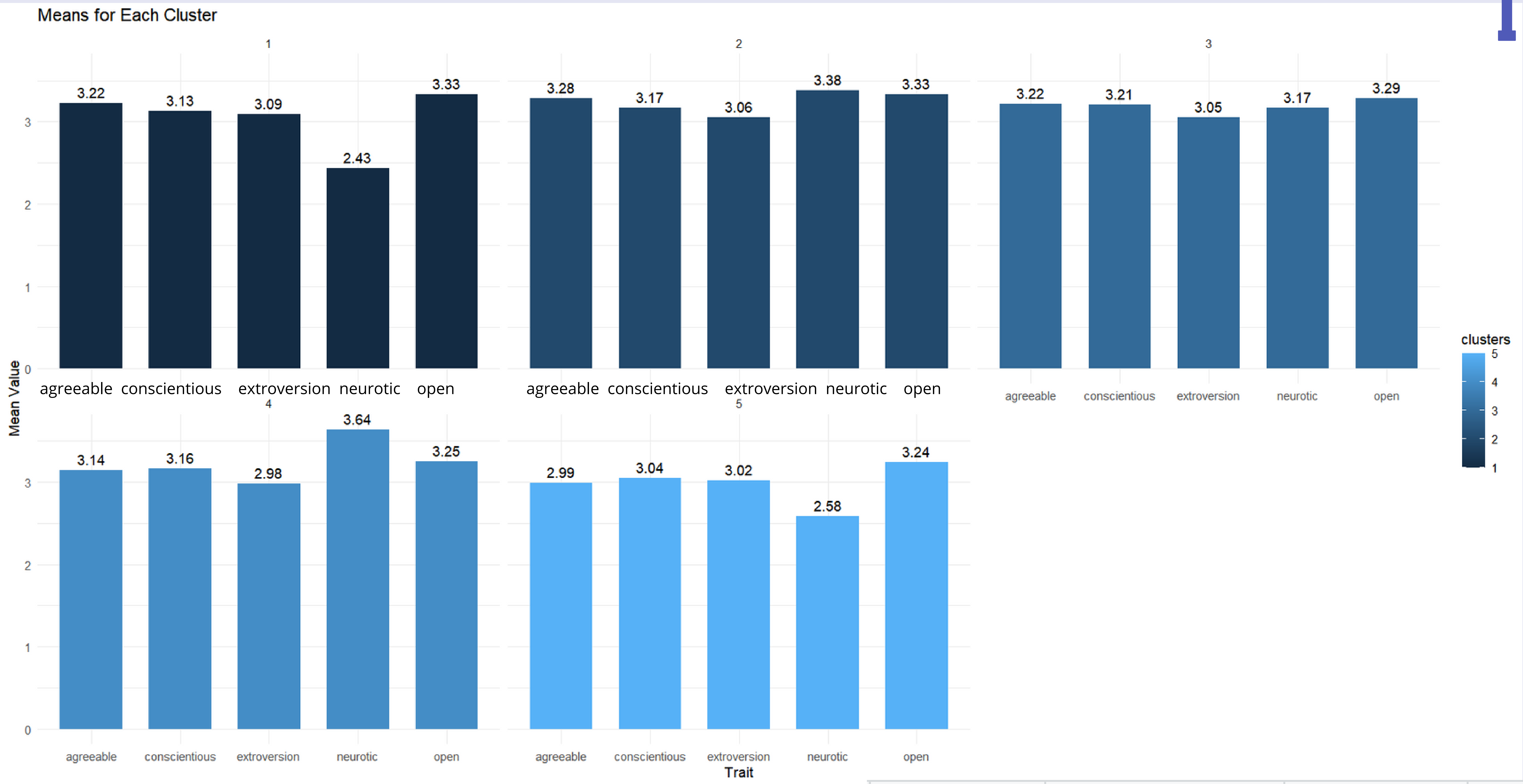

PCA Scatterplot
PCA2
PCA1
10
5
0
-5
-10
-15 -10 -5 0 5 10
factor(Clusters)
1
2
3
4
5

K-MEANS CLUSTERING





TRAITS OF EACH CLUSTER



clusters	extroversion	neurotic	agreeable	conscientious	open
1	3.091157	2.434367	3.223059	3.131907	3.333914
2	3.055680	3.382191	3.280862	3.172741	3.334443
3	3.049193	3.172176	3.217602	3.211748	3.288881
4	2.979527	3.639560	3.140786	3.162522	3.252047
5	3.019535	2.583860	2.986129	3.044277	3.239013

Conclusions

While comprehensively capturing the intricacies of an individual's personality is complex, these five clusters serve as an initial framework for understanding variations in personality traits across different groups. To further explore this, I've requested two individuals to provide their results for clustering analysis.

Results summary

Your results from the IPIP Big Five Factor Markers are in the table below. The table contains a raw score and also a percentile, what percent of other people who have taken this test that you score higher than.

Factor	Factor label	Raw score	Score percentile
I	Extroversion		37
II	Emotional stability		19
III	Agreeableness		40
IV	Conscientiousness		57
V	Intellect/Imagination		28

Big five personality trait scores calculated by openpsychometrics.org

cluster 1

Results summary

Your results from the IPIP Big Five Factor Markers are in the table below. The table contains a raw score and also a percentile, what percent of other people who have taken this test that you score higher than.

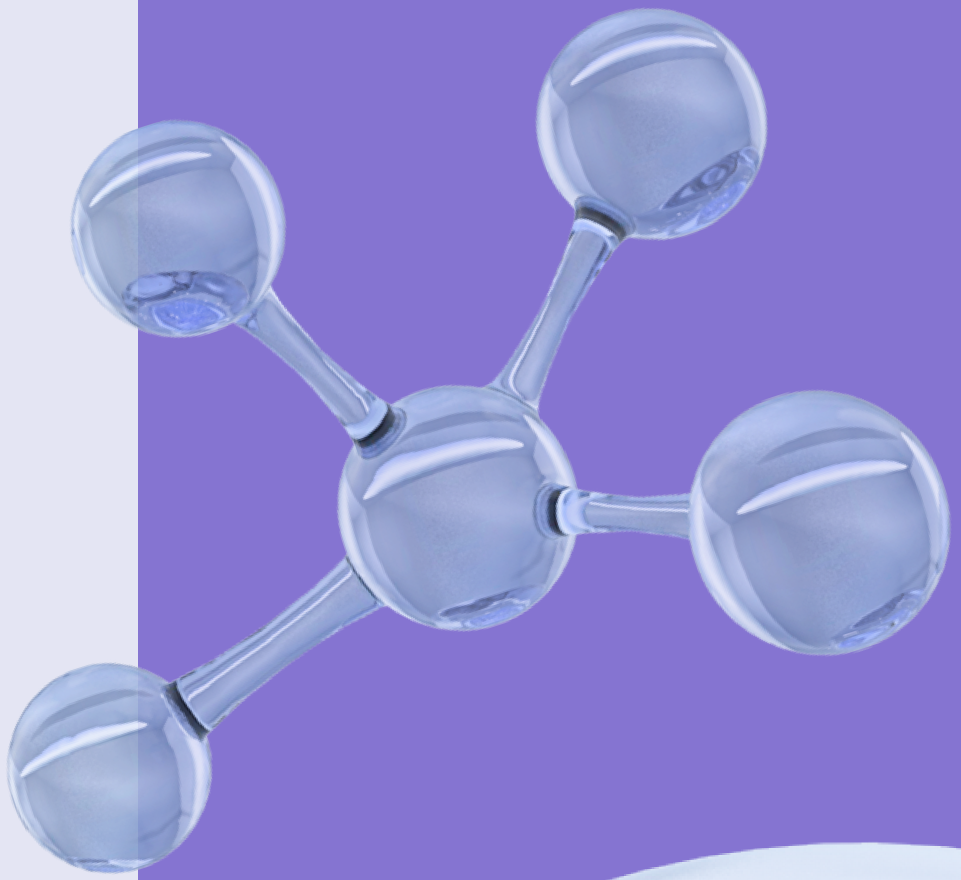
Factor	Factor label	Raw score	Score percentile
I	Extroversion		86
II	Emotional stability		81
III	Agreeableness		62
IV	Conscientiousness		15
V	Intellect/Imagination		6

Big five personality trait scores calculated by openpsychometrics.org

cluster 3

*Emotional stability = Neuroticism
Intellect/immagination = Openess

Supervised Learning



Research Questions

- What are the personality traits of illegal drug users and legal drug users?
- Is there any correlation between specific drugs consumption
- Compare classification models to classify DRUG USER/NOT USER
- Which personality trait best classify each drug user

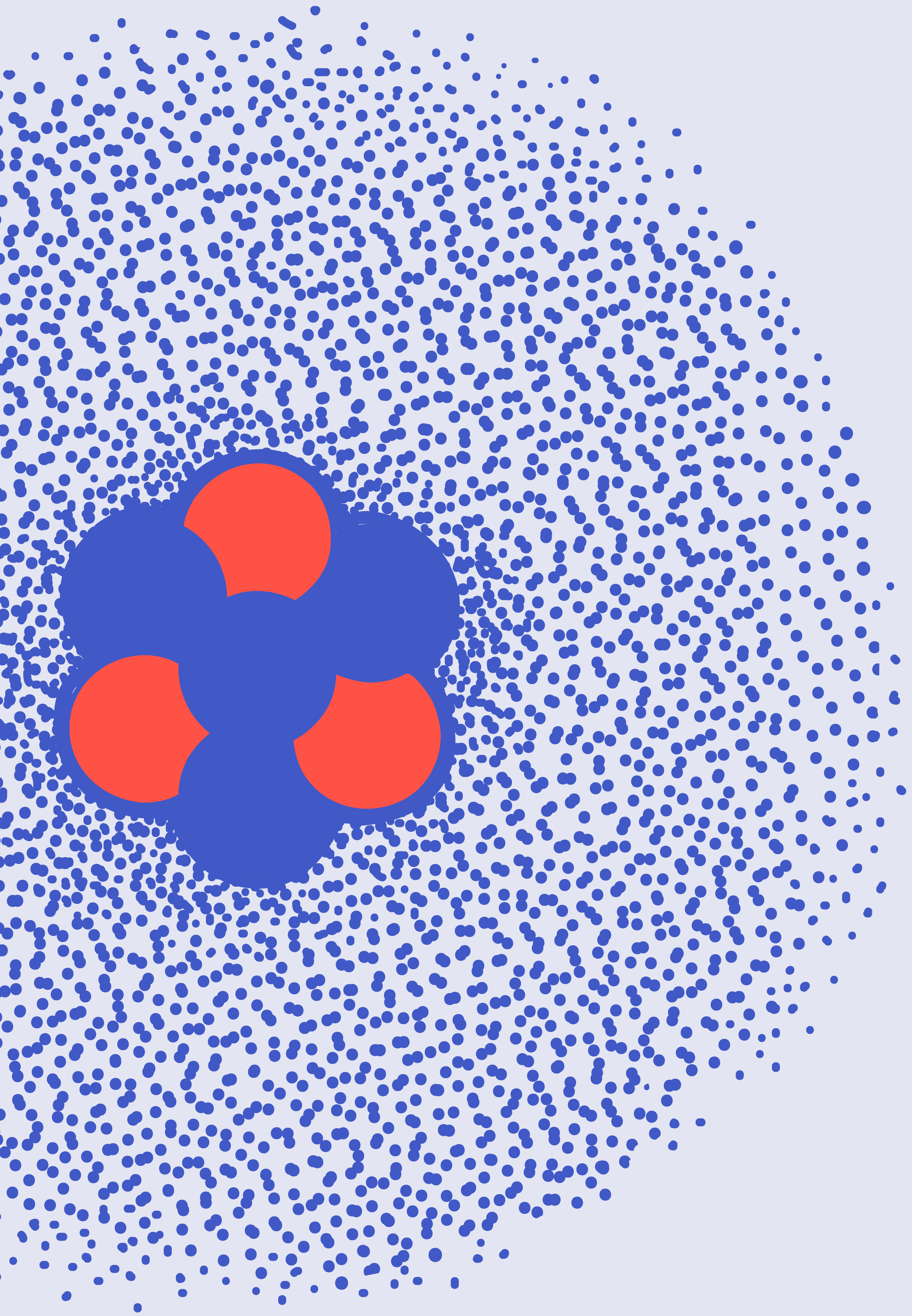


TABLE OF CONTENTS:

The Dataset

Feature	Description
ID	Identification
Age	Age range of participant
Gender	Male or Female
Education	Level of education
Country	Country of orgin
Ethnicity	Ethnicity/Race of participant
Nscore	Quantified NEO Five-Factor Inventory Neuroticism score
Escore	Quantified NEO Five-Factor Inventory Extraversion score
Oscore	Quantified NEO Five-Factor Inventory Openness to experience score
Ascore	Quantified NEO Five-Factor Inventory Agreeableness score
Cscore	Quantified NEO Five-Factor Inventory Conscientiousness score
Impulsive	Quantified BIS-11 impulsiveness score
SS	Quantified Impulsive Sensation Seeking score
Drug	Various drugs were examined and measured in terms of frequency of use

Alcohol
Amphet
Amyl: (nitrite)
Benzos
Caff: (caffeine)
Cannabis
Choc: (chocolate)
Coke
Crack
Ecstasy
Heroin
Ketamine
Legalh
LSD
Meth: (methadone)
Mushroom
Nicotine
Semer: fictitious drug (i.e. control)
VSA: (volatile substance abuse consumption)

Source: Fehrman,Elaine, Egan,Vincent, and Mirkes,Evgeny. (2016). Drug Consumption (Quantified). UCI Machine Learning Repository. https://doi.org/10.24432/C5TC7S.

Encoding the variables

Education:

0 = Left school before 16 years

1 = Left school at 16 years

2 = Left school at 17 years

3 = Left school at 18 years

4 = Some college or university, no certificate or degree

5 = Professional certificate/ diploma

6 = University degree

7 = Masters degree

8 = Doctorate degree

Country:

0 = Australia

1 = Canada

2 = New Zealand

3 = Other

4 = Republic of Ireland

5 = UK

6 = USA

Age:

0 = 18-24

1 = 25-34

2 = 35-44

3 = 45-54

4 = 55-64

5 = 65+

Gender:

0 = F

1 = M

Ethincity:

0 = Asian

1 = Black

2 = Mixed-Black/Asian

3 = Mixed-White/Asian

4 = Mixed-White/Black

5 = Other

6 = White

Drug Use:

0 = never used the drug

1 = used it over a decade ago

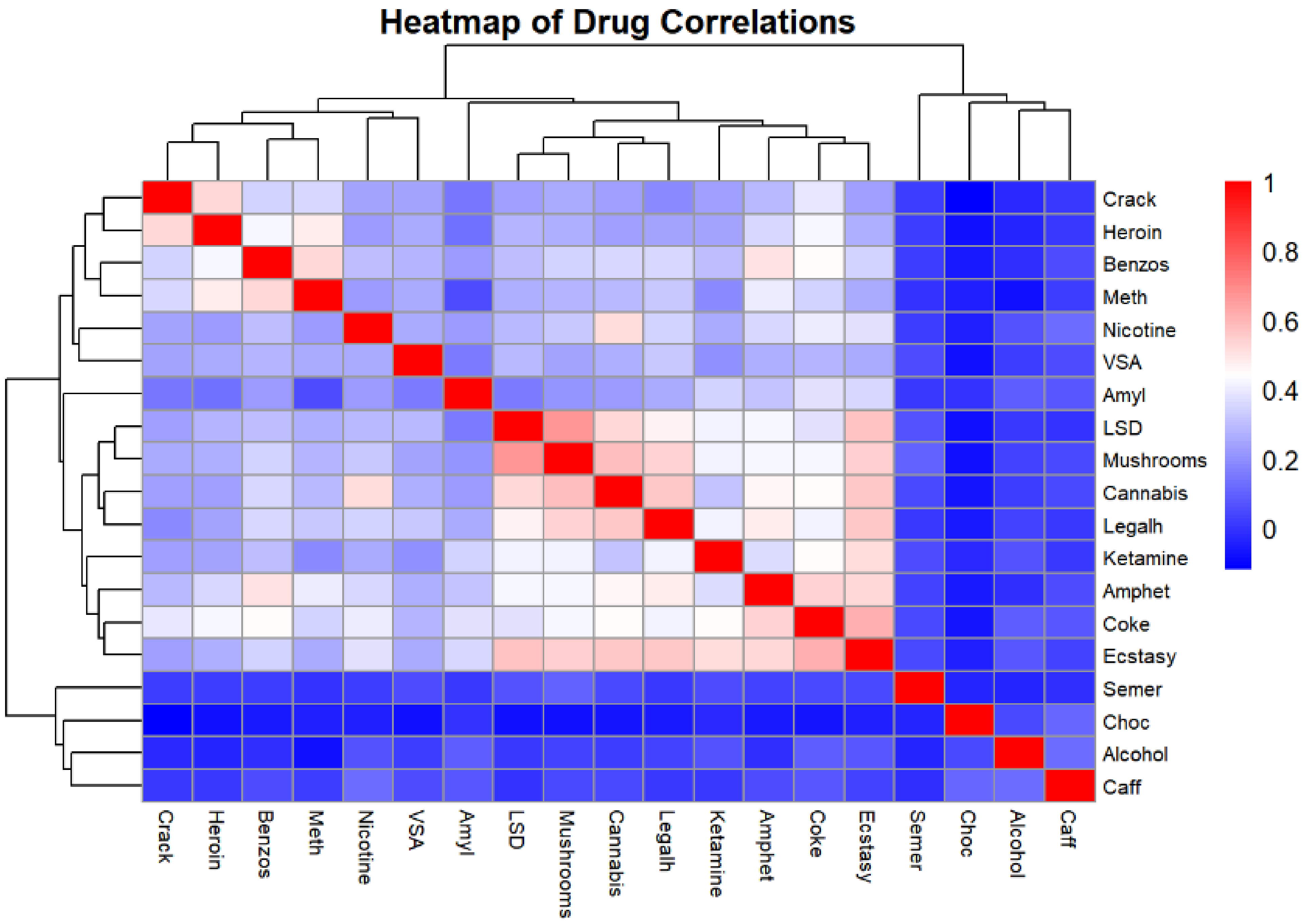
2 = in the last decade

3 = used in the last year

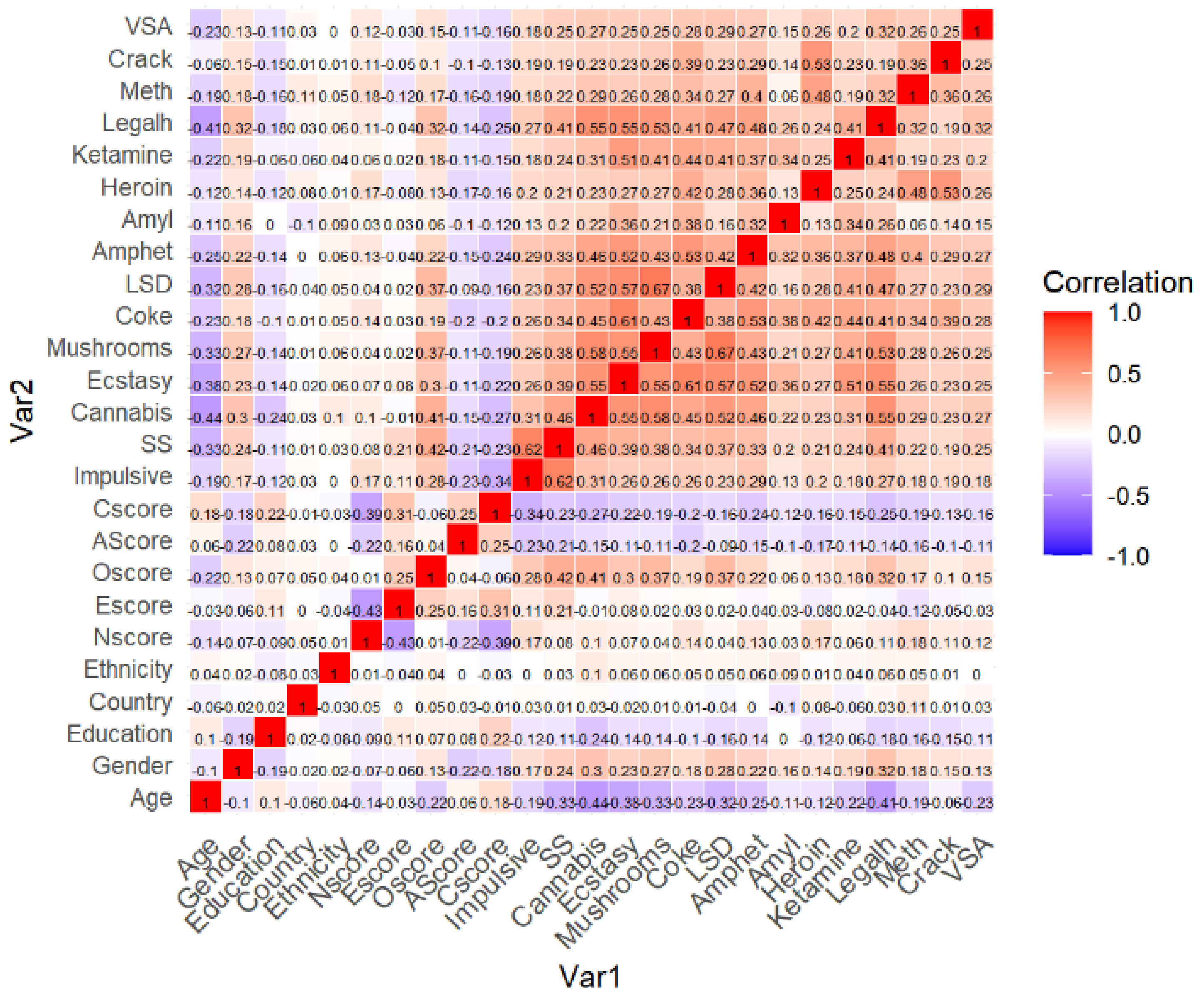
4 = used in the last month

5 = used in the last week

6 = used in the last day



Correlations



Mushrooms-LSD (0.67)

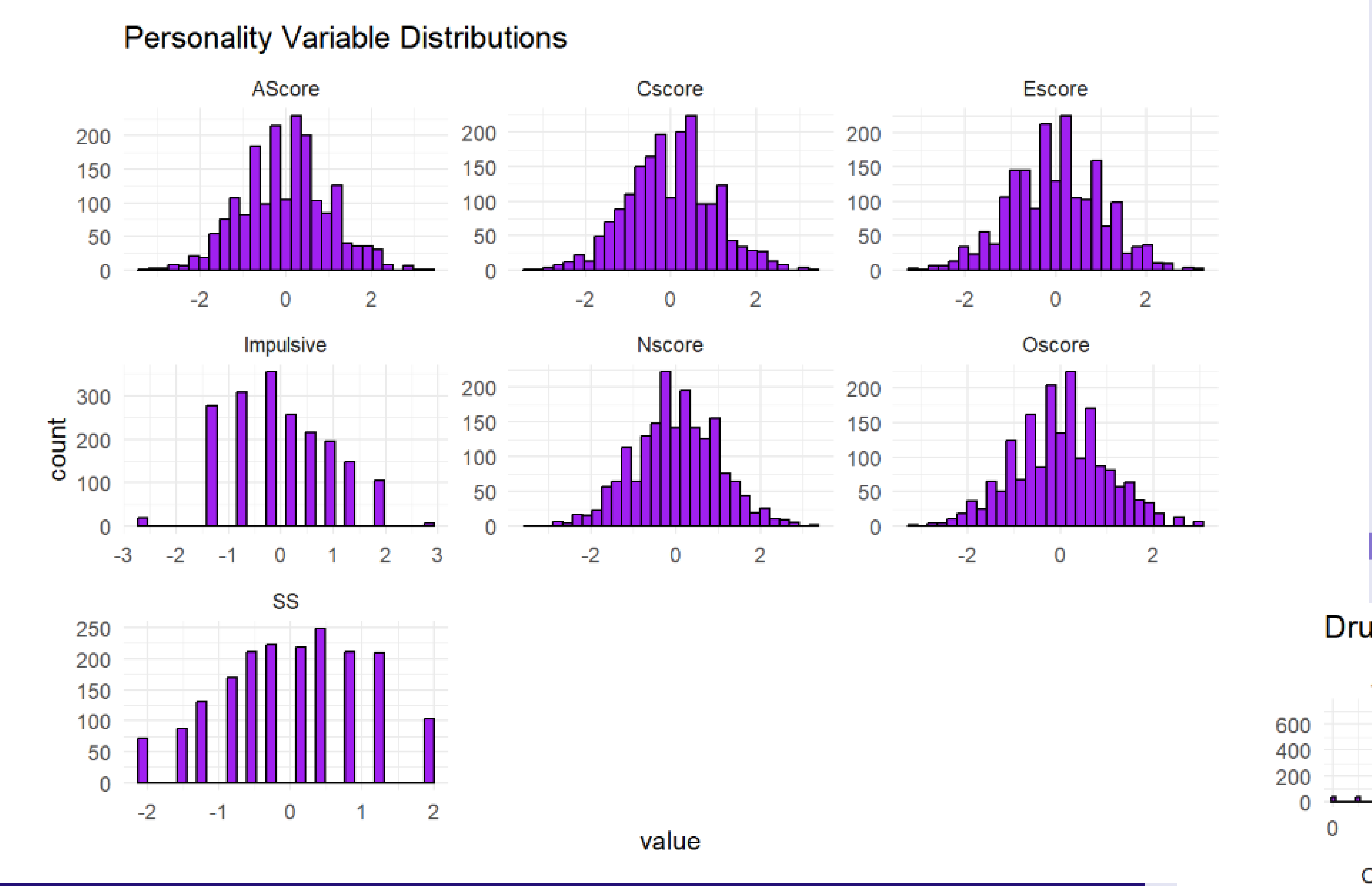
Ecstasy-Coke (0.61)

SS - Impulsive (0.62)

Mushrooms-Cannabis (0.58)

Ecstasy- LSD (0.57)

Variables distibution



Since not all drugs have balanced frequencies of use, we will only consider:

-Alcohol
-Benzos
-Nicotine
-Coke
-Ecstasy
-LSD
-Mushrooms
-Cannabis

DATA PREPROCESSING

Creating 2 different datasets: LEGAL DRUGS and ILLEGAL DRUGS

Transform the frequency of use into a binary variable:
0 for NON USER
1 for USER

NON-USER:
-never used
-used over a decade ago

USER:
-used in last day
-used in last week
-used in last month
-used in last year
-used in last decade

methodology

Logistic Regression

to predict whether an observation belongs to drug user or non-user. The coefficients represent the change in the log-odds of the outcome for a one-unit change in the predictor.

LDA

to find a linear combination of features that best separates different classes.

Decision Tree

The splits in the tree will show us the most informative features for classifying the drug user.

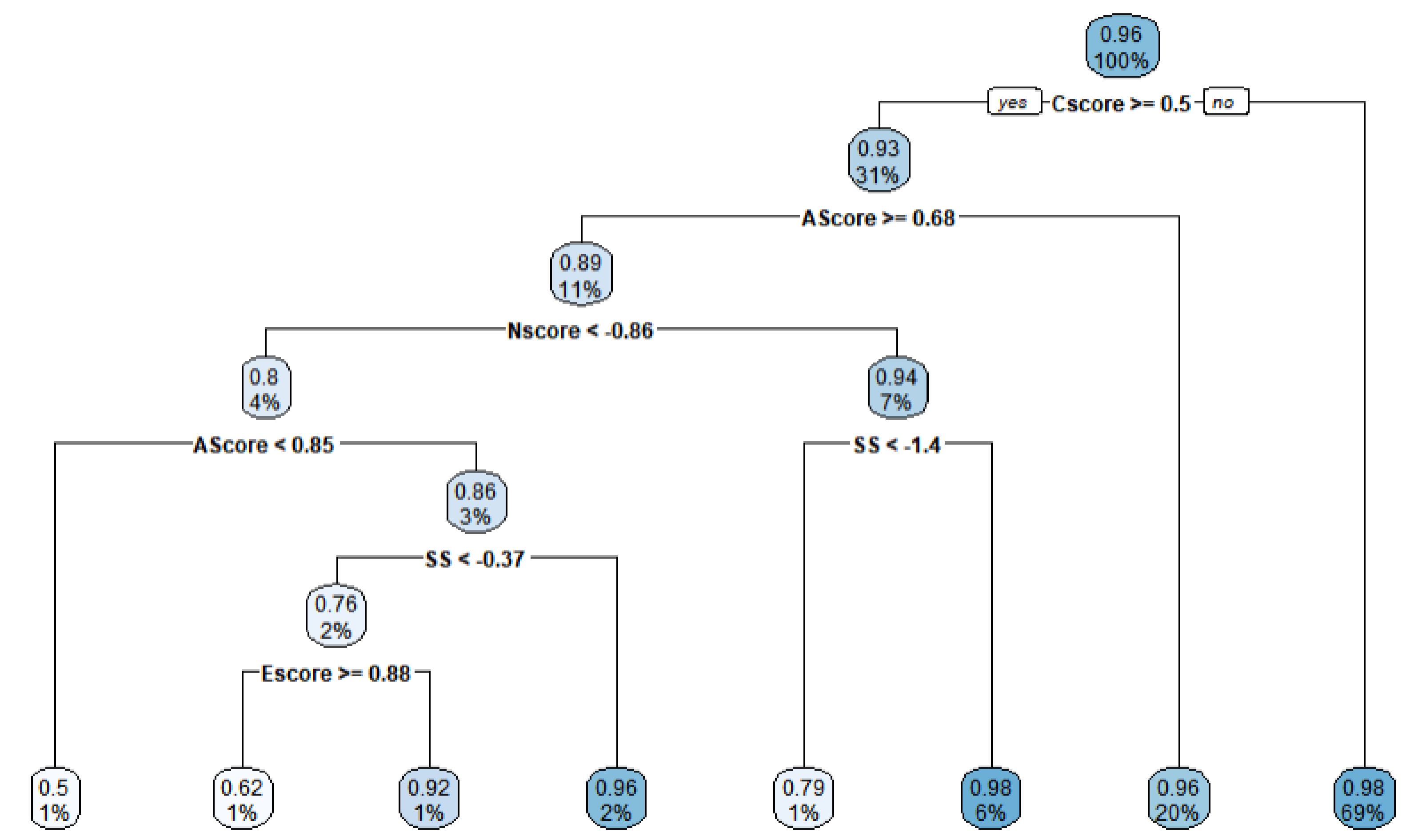
*fitting the model only on the personality traits

ALCOHOL

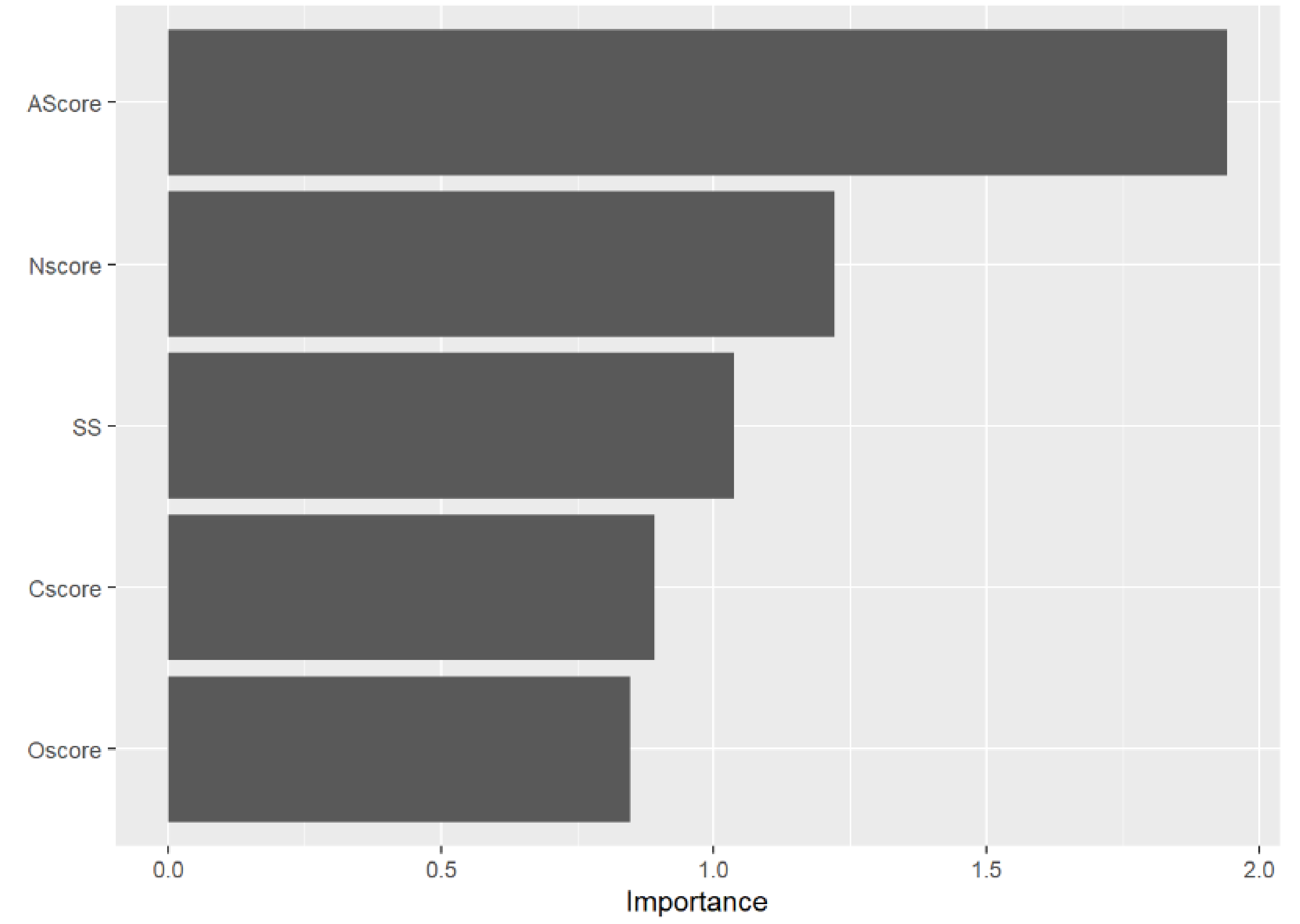
```
Call:
glm(formula = Alcohol ~ Nscore + Escore + Oscore + AScore + Cscore +
    Impulsive + SS, family = "binomial", data = dtrain_alc)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.47512    0.16340  21.268   <2e-16 ***
Nscore       0.16092    0.16617   0.968   0.3329
Escore       0.09438    0.17553   0.538   0.5908
Oscore      -0.17927    0.16018  -1.119   0.2631
AScore      -0.12251    0.15085  -0.812   0.4167
Cscore      -0.30495    0.16175  -1.885   0.0594 .
Impulsive    0.02639    0.20274   0.130   0.8965
SS           0.45805    0.20416   2.244   0.0249 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 472.20  on 1507  degrees of freedom
Residual deviance: 449.95  on 1500  degrees of freedom
AIC: 465.95

Number of Fisher Scoring iterations: 6
```

```
## Call:
## lda(Alcohol ~ Nscore + Escore + Oscore + AScore + Cscore + Impulsive +
##     SS, data = dtrain_alc)
##
## Prior probabilities of groups:
##          0          1
## 0.03647215 0.96352785
##
## Group means:
##        Nscore       Escore       Oscore      AScore      Cscore   Impulsive
## 0 -0.24697073  0.037018182 -0.056070909  0.27669709  0.43814145 -0.35080091
## 1  0.03355943 -0.006089635 -0.004114976 -0.02022479 -0.02060255  0.02099461
##            SS
## 0 -0.45651509
## 1  0.01234409
##
## Coefficients of linear discriminants:
##                   LD1
## Nscore     0.22293971
## Escore     0.13261222
## Oscore    -0.25293549
## AScore    -0.17667571
## Cscore    -0.47054991
## Impulsive  0.01634932
## SS         0.68352594
```

ALCOHOL



0.96
100%
yes
Cscore >= 0.5
no
0.93
31%
AScore >= 0.68
0.89
11%
Nscore < -0.86
0.8
4%
0.94
7%
AScore < 0.85
SS < -1.4
0.86
3%
SS < -0.37
0.76
2%
Escore >= 0.88
0.5
1%
0.62
1%
0.92
1%
0.96
2%
0.79
1%
0.98
6%
0.96
20%
0.98
69%



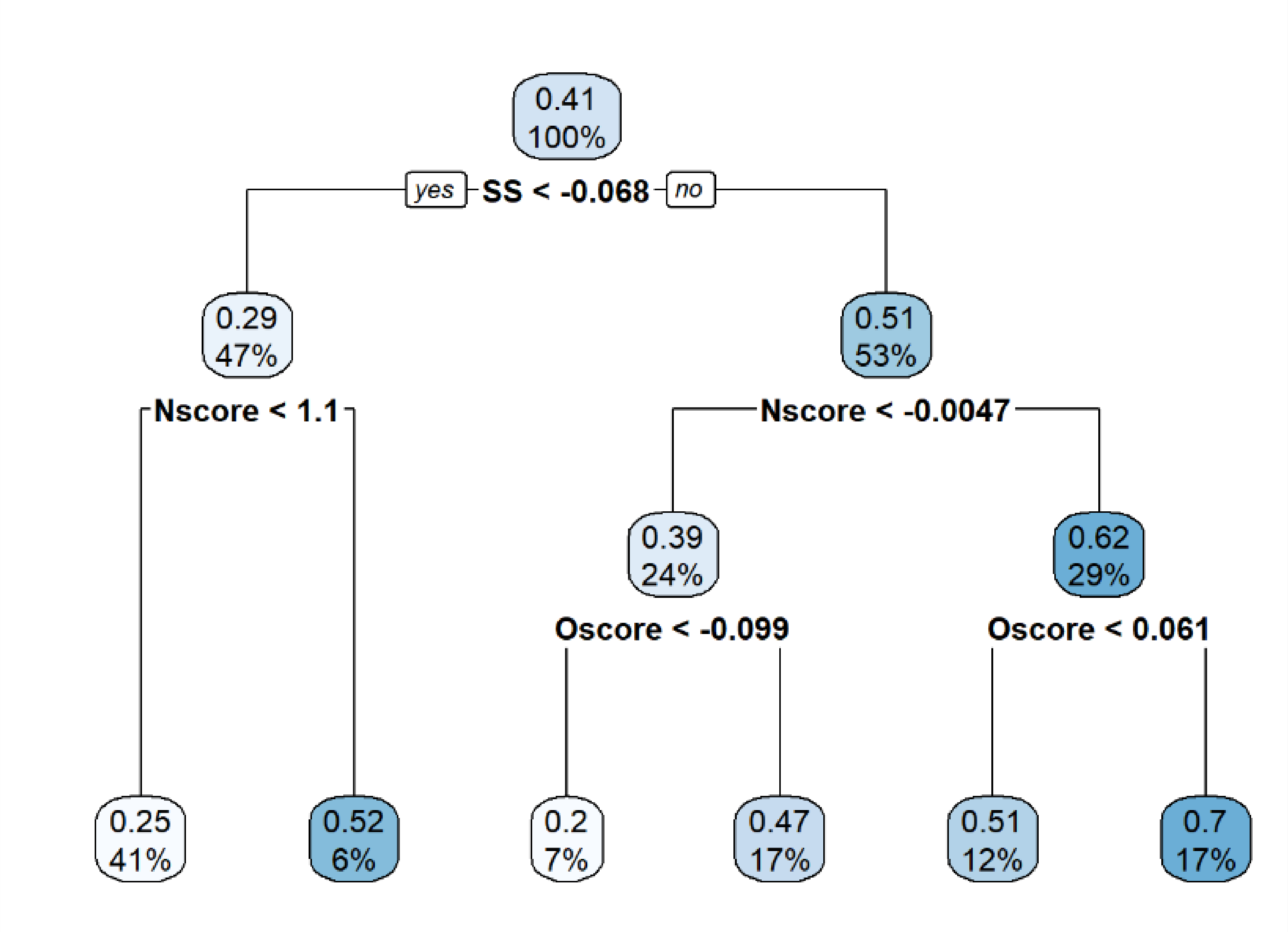
AScore
Nscore
SS
Cscore
Oscore
0.0
0.5
1.0
1.5
2.0
Importance

BENZOS

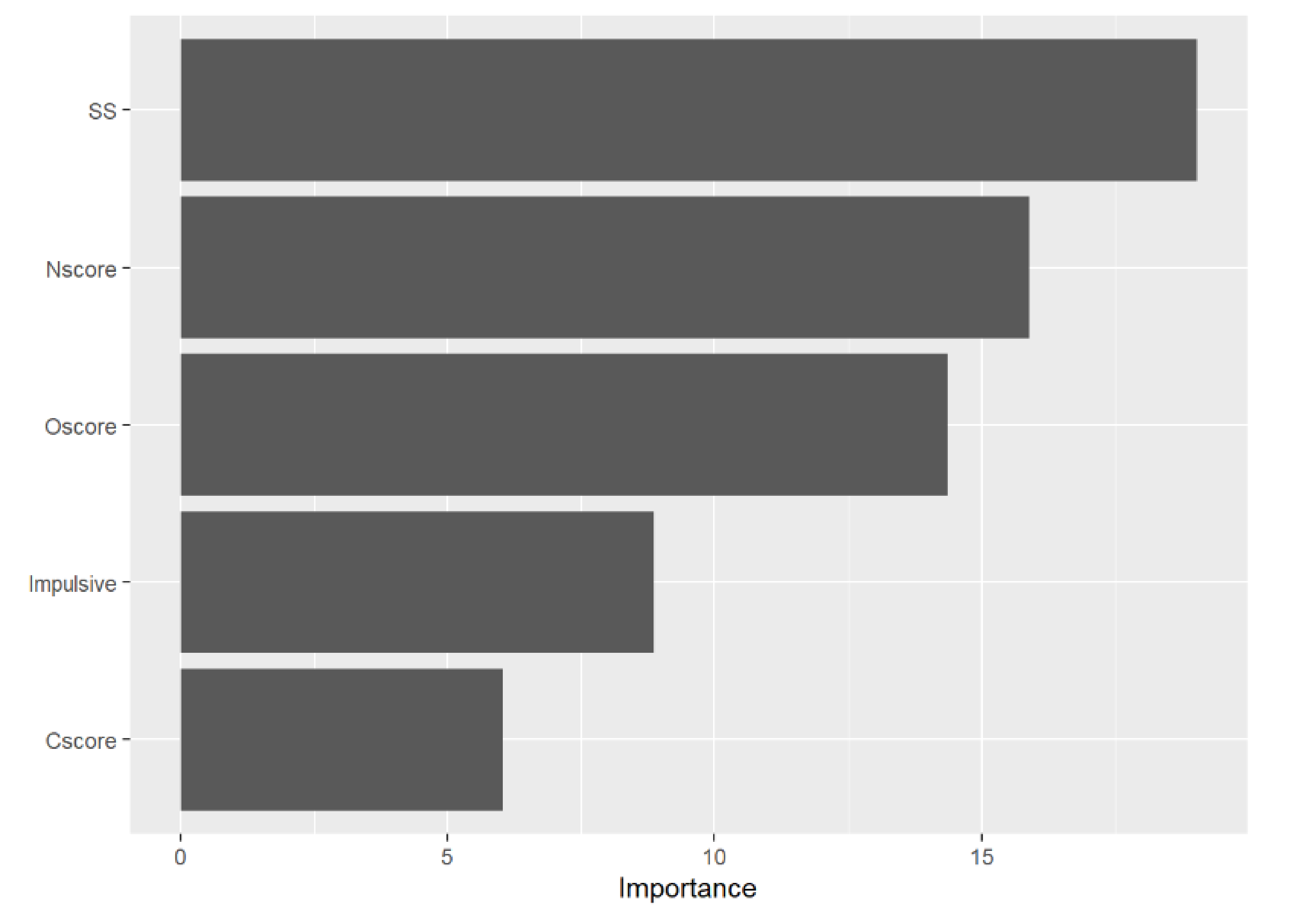
```
Call:
glm(formula = Benzos ~ Nscore + Escore + Oscore + AScore + Cscore +
    Impulsive + SS, family = "binomial", data = dtrain_benzos)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.43548    0.05682  -7.664 1.81e-14 ***
Nscore       0.34214    0.06772   5.052 4.37e-07 ***
Escore      -0.13832    0.06924  -1.998   0.0458 *
Oscore       0.37284    0.06546   5.696 1.23e-08 ***
AScore      -0.14221    0.06036  -2.356   0.0185 *
Cscore      -0.10636    0.06650  -1.599   0.1097
Impulsive    0.05712    0.07761   0.736   0.4617
SS           0.33654    0.08001   4.206 2.59e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2039.0  on 1507  degrees of freedom
Residual deviance: 1830.7  on 1500  degrees of freedom
AIC: 1846.7

Number of Fisher Scoring iterations: 4
```

```
## Call:
## lda(Benzos ~ Nscore + Escore + Oscore + AScore + Cscore + Impulsive +
##     SS, data = dtrain_benzos)
##
## Prior probabilities of groups:
##         0         1
## 0.5921751 0.4078249
##
## Group means:
##       Nscore      Escore     Oscore     AScore     Cscore  Impulsive         SS
## 0 -0.1627828  0.06170409 -0.1811388  0.1085532  0.1544612 -0.1554794 -0.2063871
## 1  0.2935666 -0.10067315  0.2482829 -0.1806607 -0.2337749  0.2439906  0.2880184
##
## Coefficients of linear discriminants:
##                   LD1
## Nscore     0.44529841
## Escore    -0.17696682
## Oscore     0.47263521
## AScore    -0.17620837
## Cscore    -0.14250948
## Impulsive  0.07588566
## SS         0.43348271
```

BENZOS



0.41
100%
yes
SS < -0.068
no
0.29
47%
0.51
53%
Nscore < 1.1
Nscore < -0.0047
0.39
24%
0.62
29%
Oscore < -0.099
Oscore < 0.061
0.25
41%
0.52
6%
0.2
7%
0.47
17%
0.51
12%
0.7
17%



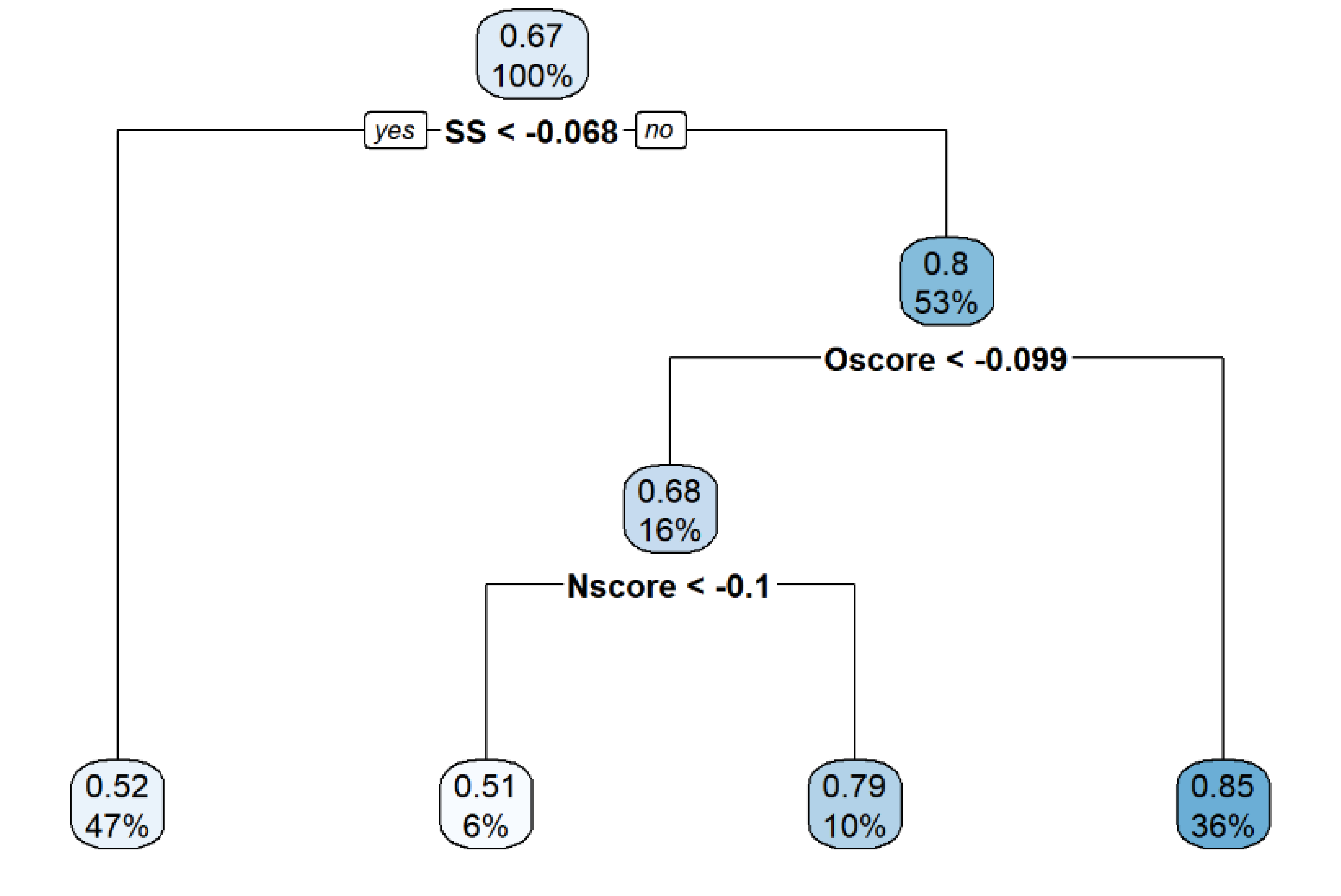
SS
Nscore
Oscore
Impulsive
Cscore
0
5
10
15
Importance

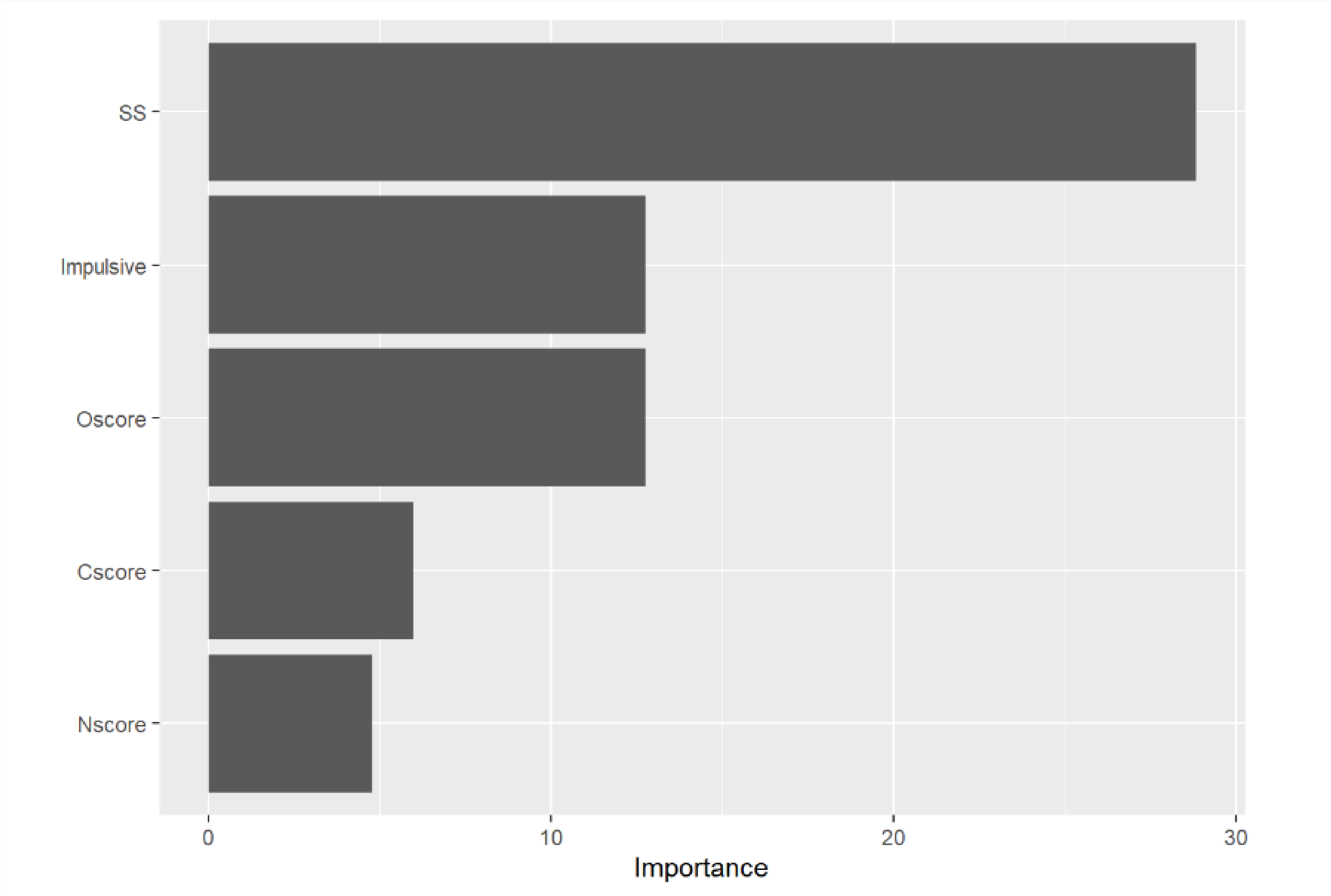
NICOTINE

```
Call:
glm(formula = Nicotine ~ Nscore + Escore + Oscore + AScore +
    Cscore + Impulsive + SS, family = "binomial", data = dtrain_benzos)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.81927    0.06091  13.451  < 2e-16 ***
Nscore       0.06768    0.07052   0.960    0.337
Escore      -0.08424    0.07447  -1.131    0.258
Oscore       0.27787    0.06782   4.097 4.18e-05 ***
AScore      -0.02309    0.06388  -0.362    0.718
Cscore      -0.28616    0.07018  -4.078 4.55e-05 ***
Impulsive   -0.01619    0.08204  -0.197    0.844
SS           0.60495    0.08655   6.990 2.76e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1914.6  on 1507  degrees of freedom
Residual deviance: 1707.0  on 1500  degrees of freedom
AIC: 1723

Number of Fisher Scoring iterations: 4
```

```
## Call:
## lda(Nicotine ~ Nscore + Escore + Oscore + AScore + Cscore + Impulsive +
##     SS, data = dtrain_nic)
##
## Prior probabilities of groups:
##         0         1
## 0.3309019 0.6690981
##
## Group means:
##       Nscore       Escore     Oscore      AScore     Cscore  Impulsive
## 0 -0.1329990  0.003937615 -0.3101229  0.13102924  0.2862047 -0.2954788
## 1  0.1006392 -0.008698821  0.1443889 -0.07884229 -0.1473279  0.1572399
##           SS
## 0 -0.4449180
## 1  0.2129254
##
## Coefficients of linear discriminants:
##                   LD1
## Nscore     0.07477321
## Escore    -0.10245864
## Oscore     0.34543442
## AScore    -0.03557461
## Cscore    -0.34918448
## Impulsive -0.02354387
## SS         0.75409855
```

NICOTINE


0.67
100%
yes
SS < -0.068
no
0.8
53%
Oscore < -0.099
0.68
16%
Nscore < -0.1
0.52
47%
0.51
6%
0.79
10%
0.85
36%

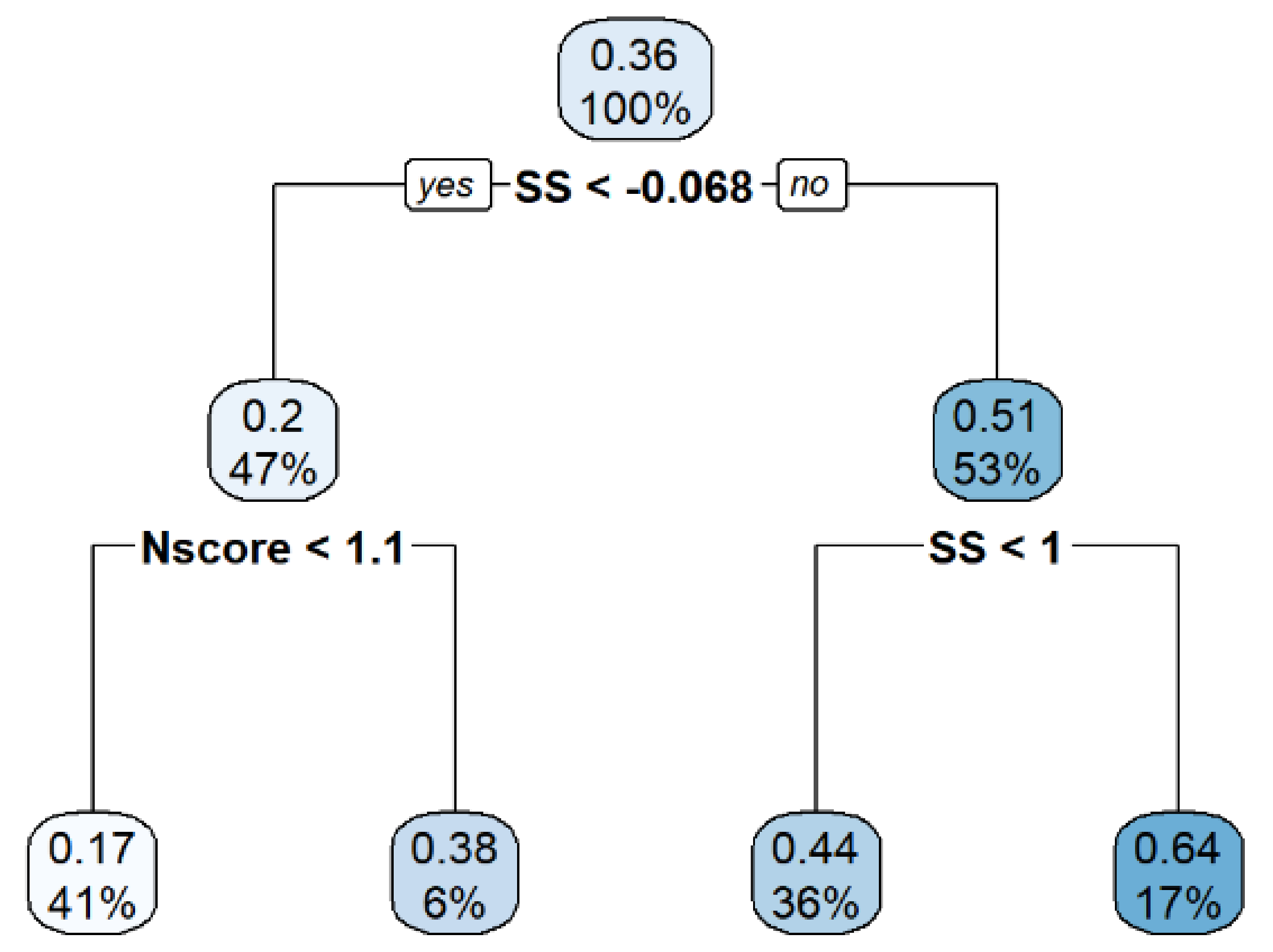

SS
Impulsive
Oscore
Cscore
Nscore
0
10
20
30
Importance

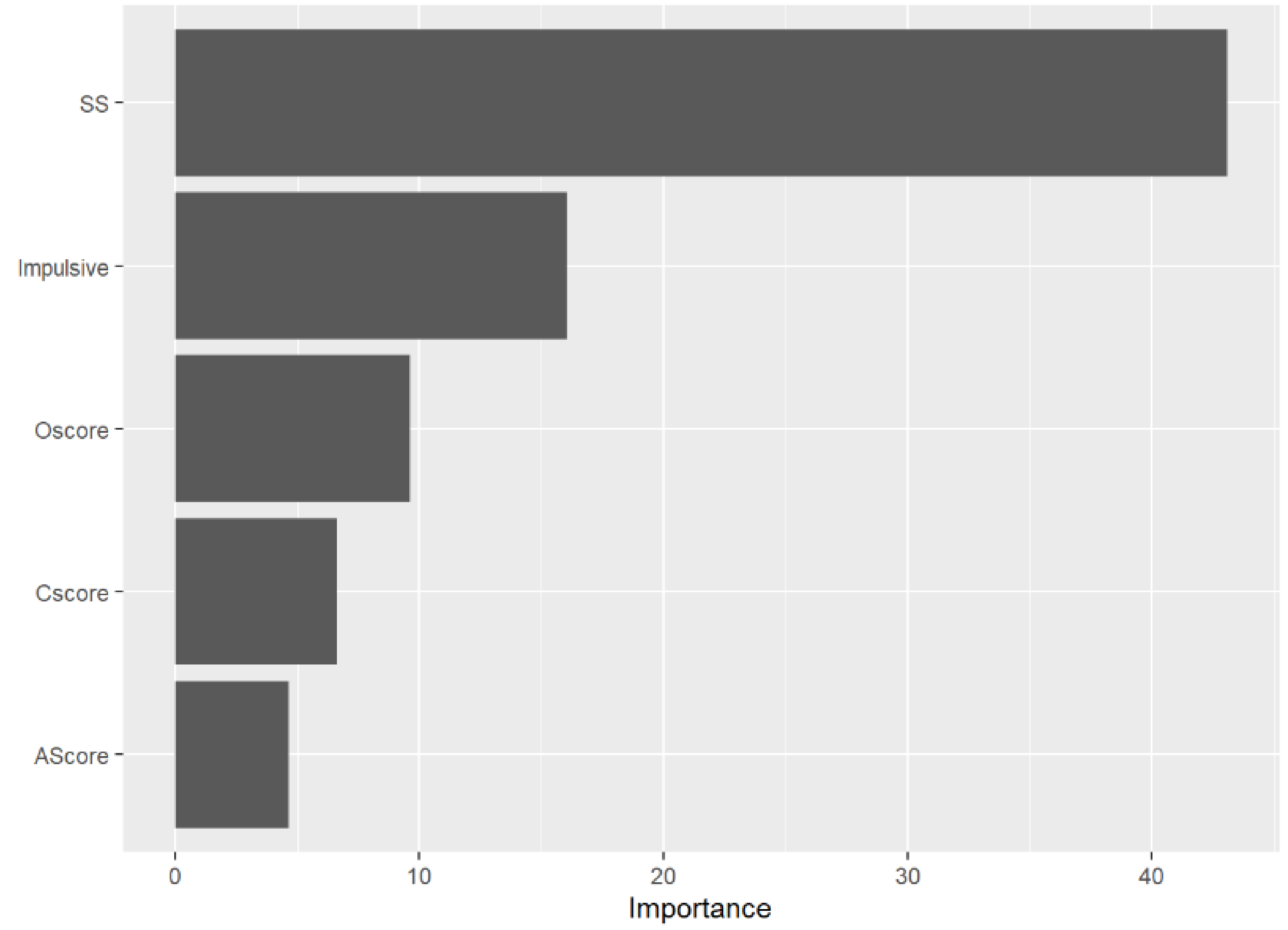
COKE

```
Call:
glm(formula = Coke ~ Nscore + Escore + Oscore + AScore + Cscore +
    Impulsive + SS, family = "binomial", data = dtrain_coke)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept) -0.673993   0.059775 -11.276  < 2e-16 ***
Nscore       0.197768   0.069074   2.863 0.004195 ** 
Escore       0.103853   0.070863   1.466 0.142774    
Oscore       0.197740   0.066379   2.979 0.002892 ** 
AScore      -0.208287   0.062285  -3.344 0.000825 ***
Cscore      -0.204245   0.069167  -2.953 0.003148 ** 
Impulsive    0.009922   0.079456   0.125 0.900626    
SS           0.617895   0.084229   7.336  2.2e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1974.3  on 1507  degrees of freedom
Residual deviance: 1737.3  on 1500  degrees of freedom
AIC: 1753.3

Number of Fisher Scoring iterations: 4
```

```
## Call:
## lda(Coke ~ Nscore + Escore + Oscore + AScore + Cscore + Impulsive +
##     SS, data = dtrain_coke)
##
## Prior probabilities of groups:
##        0        1
## 0.637931 0.362069
##
## Group means:
##        Nscore      Escore    Oscore     AScore     Cscore  Impulsive         SS
## 0 -0.08075761 -0.03212476 -0.154519  0.1169344  0.1428587 -0.1739371 -0.2552191
## 1  0.20671663  0.04412414  0.255649 -0.2319765 -0.2623952  0.3269938  0.4365356
##
## Coefficients of linear discriminants:
##                   LD1
## Nscore     0.23695945
## Escore     0.12581467
## Oscore     0.21883931
## AScore    -0.24047606
## Cscore    -0.24382264
## Impulsive  0.01617947
## SS         0.72722059
```

COKE


0.36
100%
yes
SS < -0.068
no
0.2
47%
0.51
53%
Nscore < 1.1
SS < 1
0.17
41%
0.38
6%
0.44
36%
0.64
17%

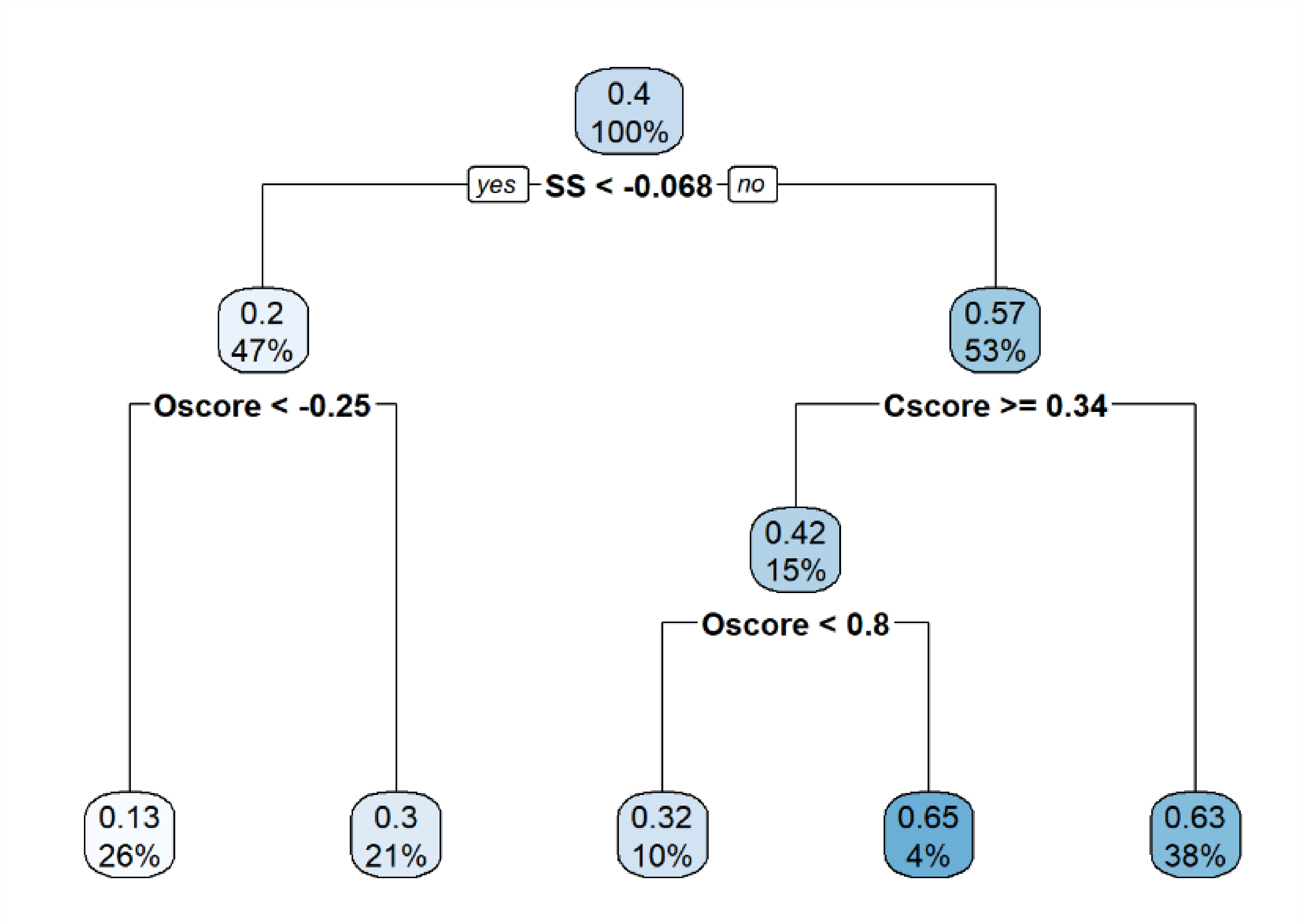

SS
Impulsive
Oscore
Cscore
AScore
0
10
20
30
40
Importance

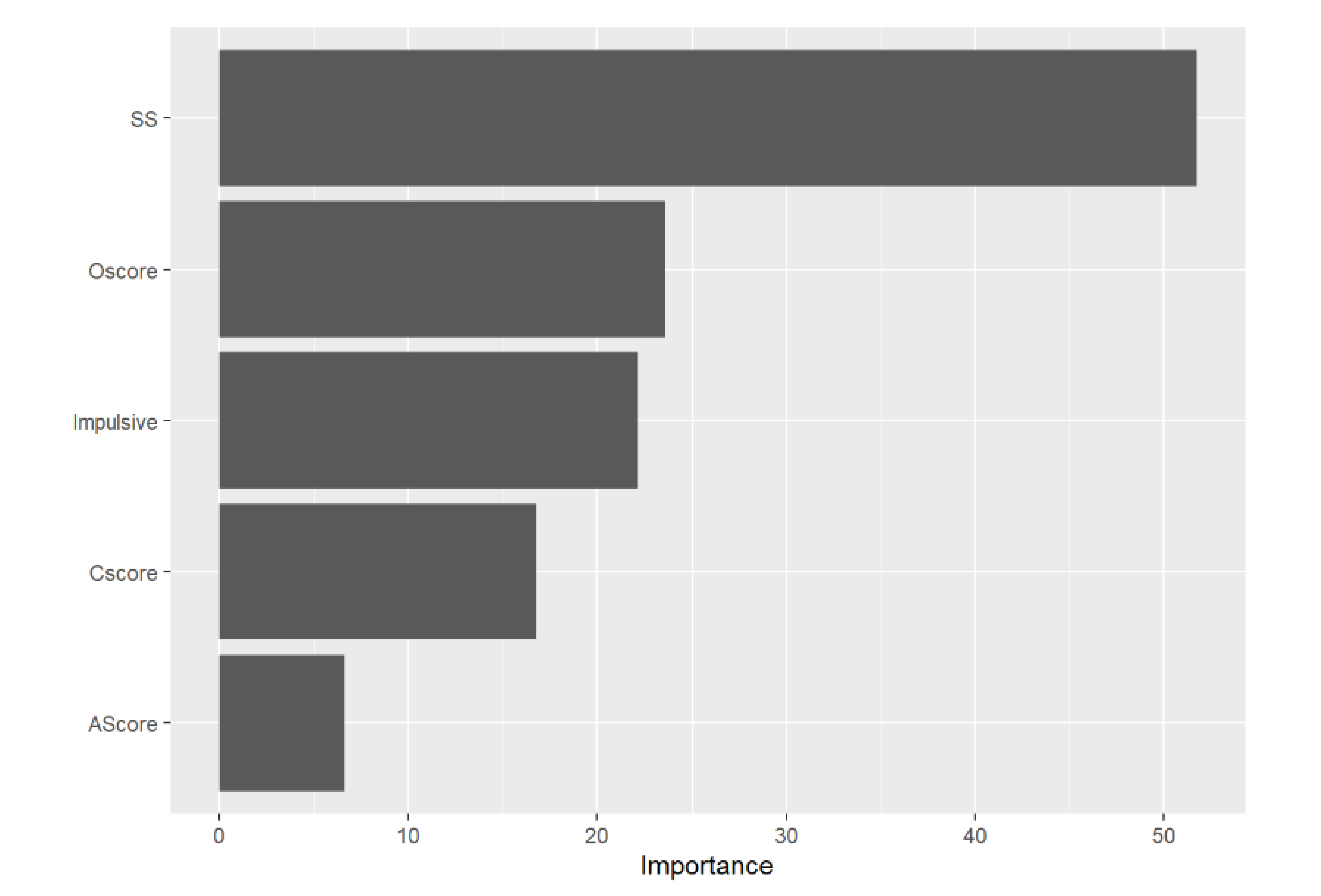
ECSTASY

```
Call:
glm(formula = Ecstasy ~ Nscore + Escore + Oscore + AScore + Cscore +
    Impulsive + SS, family = "binomial", data = dtrain_ecst)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -0.51568    0.06019  -8.568  < 2e-16 ***
Nscore       0.01188    0.06935   0.171   0.8640    
Escore      -0.03850    0.07223  -0.533   0.5940    
Oscore       0.41851    0.06860   6.101 1.06e-09 ***
AScore      -0.08460    0.06301  -1.343   0.1794    
Cscore      -0.32941    0.07076  -4.655 3.24e-06 ***
Impulsive   -0.15145    0.08075  -1.875   0.0607 .  
SS           0.85024    0.08806   9.655  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2028.8  on 1507  degrees of freedom
Residual deviance: 1699.4  on 1500  degrees of freedom
AIC: 1715.4

Number of Fisher Scoring iterations: 4
```

```
## Call:
## lda(Ecstasy ~ Nscore + Escore + Oscore + AScore + Cscore + Impulsive +
##     SS, data = dtrain_ecst)
##
## Prior probabilities of groups:
##         0         1
## 0.6007958 0.3992042
##
## Group means:
##        Nscore      Escore     Oscore      AScore     Cscore  Impulsive
## 0 -0.05237461 -0.03472177 -0.2389952  0.08953788  0.1722041 -0.1855476
## 1  0.13725890  0.04093967  0.3446290 -0.15828837 -0.2688615  0.2978692
##           SS
## 0 -0.3229251
## 1  0.4740827
##
## Coefficients of linear discriminants:
##                   LD1
## Nscore     0.01990236
## Escore    -0.03410211
## Oscore     0.40502216
## AScore    -0.08291762
## Cscore    -0.33147214
## Impulsive -0.14904415
## SS         0.85292562
```

ECSTASY


0.4
100%
yes
SS < -0.068
no
0.2
47%
Oscore < -0.25
0.57
53%
Cscore >= 0.34
0.42
15%
Oscore < 0.8
0.13
26%
0.3
21%
0.32
10%
0.65
4%
0.63
38%

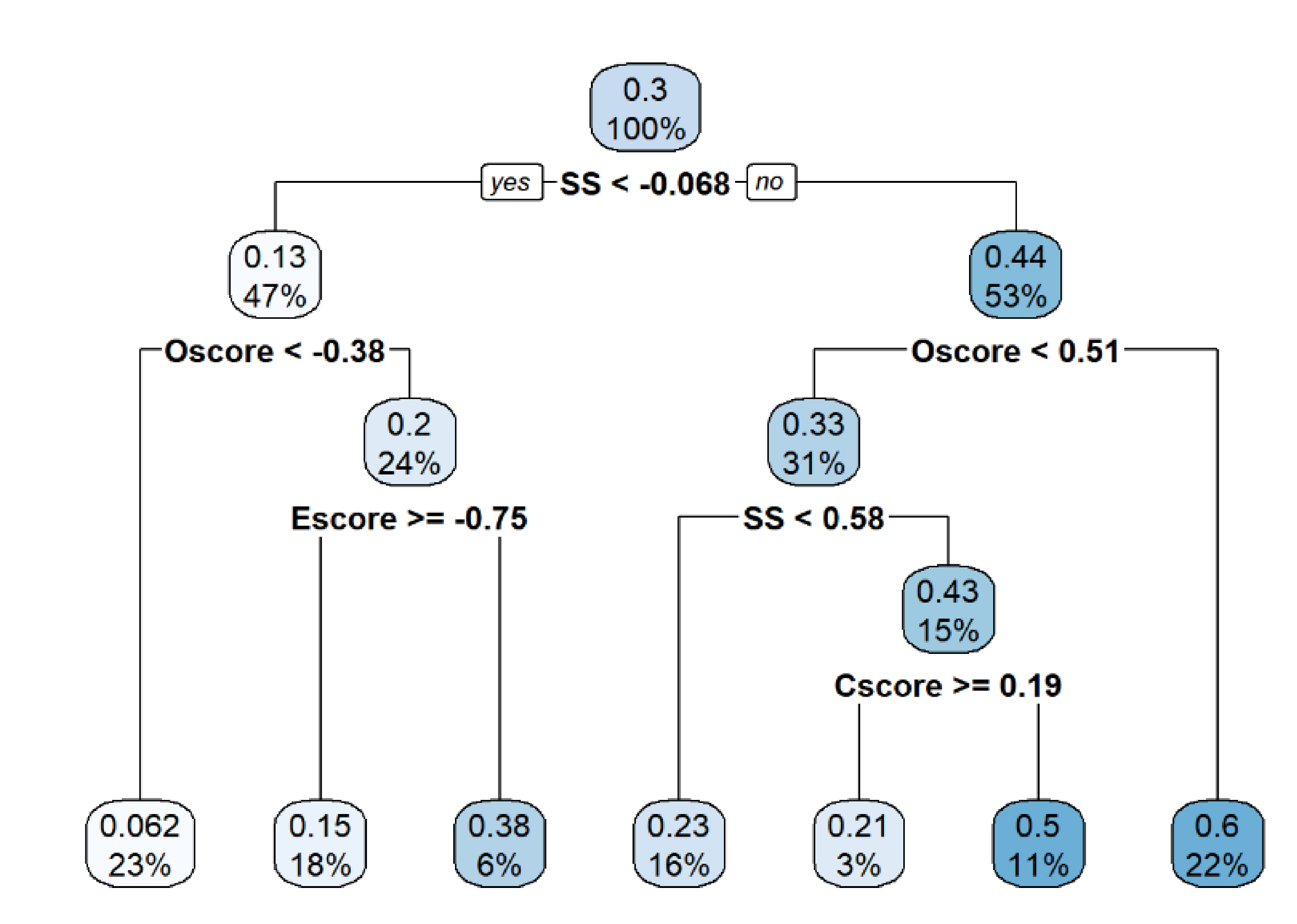

SS
Oscore
Impulsive
Cscore
AScore
0
10
20
30
40
50
Importance

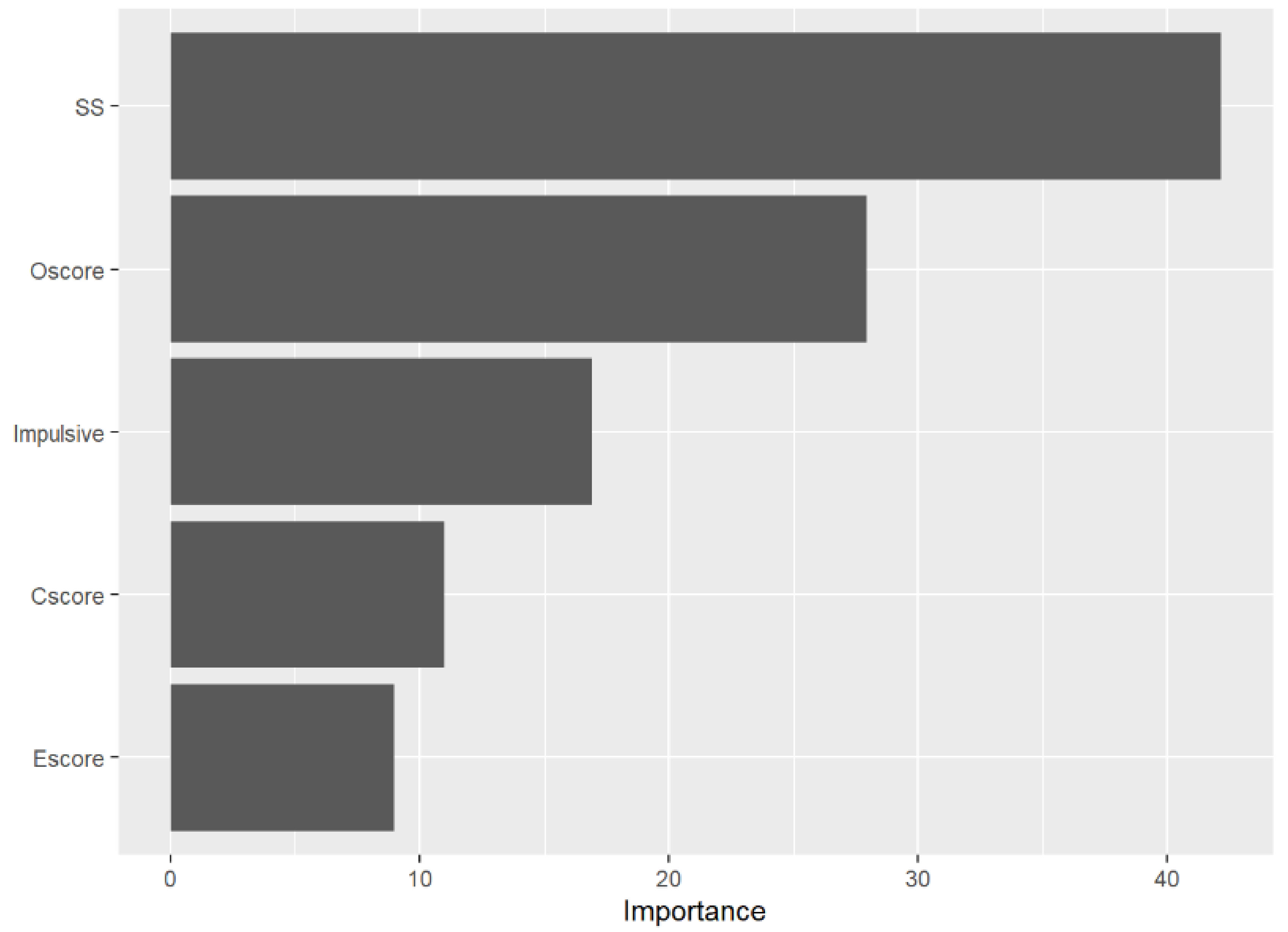
LSD

```
Call:
glm(formula = LSD ~ Nscore + Escore + Oscore + AScore + Cscore +
    Impulsive + SS, family = "binomial", data = dtrain_lsd)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.11656    0.06983 -15.990  < 2e-16 ***
Nscore      -0.14066    0.07430  -1.893  0.05836 .
Escore      -0.29084    0.07736  -3.759  0.00017 ***
Oscore       0.75977    0.07860   9.667  < 2e-16 ***
AScore      -0.10750    0.06748  -1.593  0.11114
Cscore      -0.12795    0.07548  -1.695  0.09006 .
Impulsive   -0.08902    0.08741  -1.018  0.30852
SS           0.79748    0.09425   8.461  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1834.8  on 1507  degrees of freedom
Residual deviance: 1501.9  on 1500  degrees of freedom
AIC: 1517.9

Number of Fisher Scoring iterations: 5
```

```
## Call:
## lda(LSD ~ Nscore + Escore + Oscore + AScore + Cscore + Impulsive +
##     SS, data = dtrain_lsd)
##
## Prior probabilities of groups:
##         0         1
## 0.7029178 0.2970822
##
## Group means:
##         Nscore      Escore     Oscore      AScore      Cscore  Impulsive
## 0 -0.005503425 -0.01530563 -0.2336484  0.05294982  0.08702271 -0.1312925
## 1  0.091544844  0.02100833  0.5325990 -0.15690868 -0.21893259  0.3356723
##           SS
## 0 -0.2371471
## 1  0.5450972
##
## Coefficients of linear discriminants:
##                   LD1
## Nscore    -0.12268406
## Escore    -0.25957931
## Oscore     0.67314117
## AScore    -0.08424437
## Cscore    -0.12972513
## Impulsive -0.08045348
## SS         0.71964962
```

LSD


0.3
100%
yes
SS < -0.068
no
0.13
47%
0.44
53%
Oscore < -0.38
Oscore < 0.51
0.2
24%
0.33
31%
Escore >= -0.75
SS < 0.58
0.43
15%
Cscore >= 0.19
0.062
23%
0.15
18%
0.38
6%
0.23
16%
0.21
3%
0.5
11%
0.6
22%

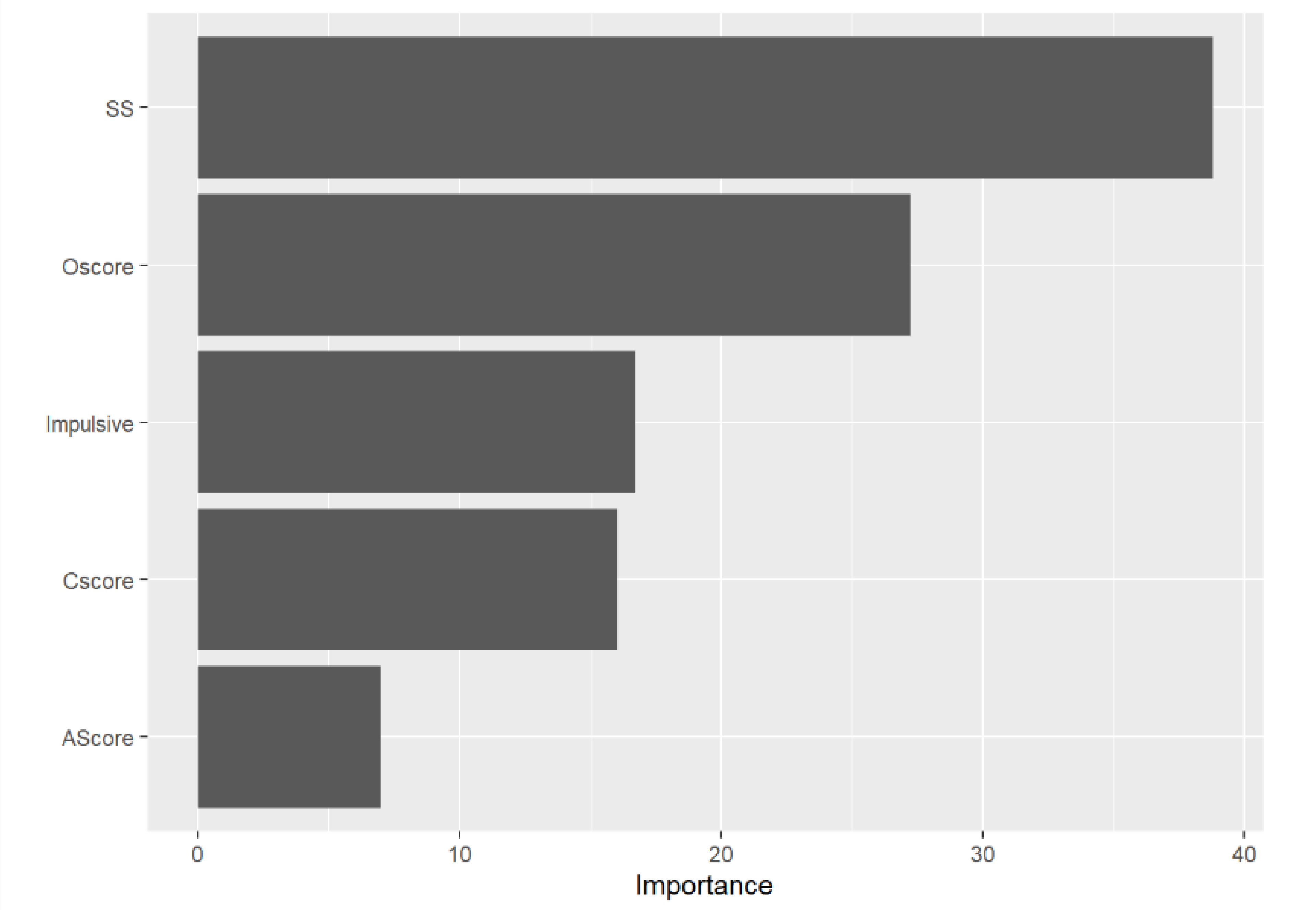

SS
Oscore
Impulsive
Cscore
Escore
0
10
20
30
40
Importance

MUSHROOM

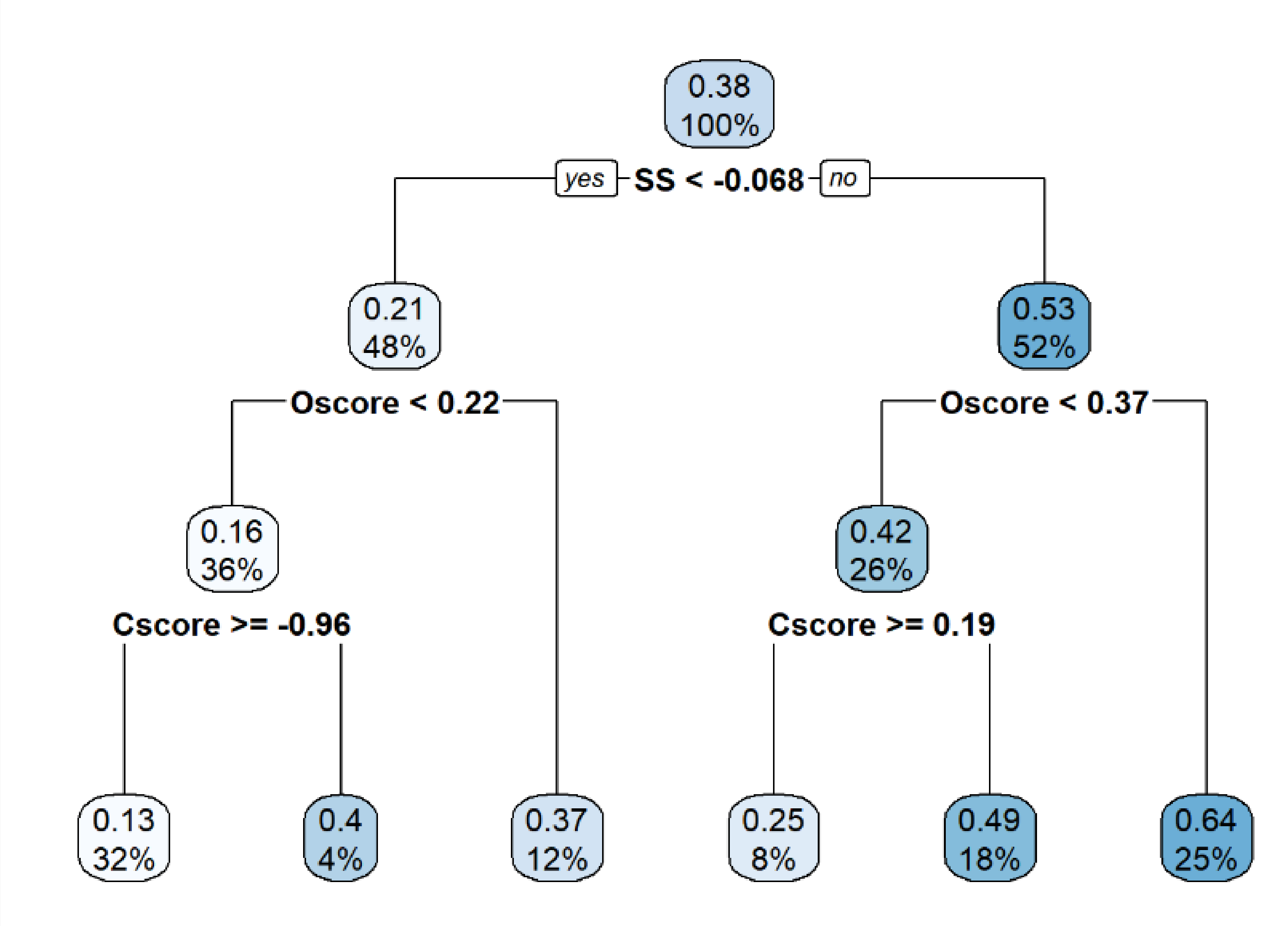
```
Call:
glm(formula = Mushrooms ~ Nscore + Escore + Oscore + AScore +
    Cscore + Impulsive + SS, family = "binomial", data = dtrain_mush)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.62295    0.06090 -10.229  < 2e-16 ***
Nscore      -0.14659    0.07028  -2.086 0.036998 *
Escore      -0.22570    0.07288  -3.097 0.001954 **
Oscore       0.63197    0.07147   8.842  < 2e-16 ***
AScore      -0.11361    0.06344  -1.791 0.073310 .
Cscore      -0.23386    0.07068  -3.309 0.000937 ***
Impulsive    0.04616    0.08285   0.557 0.577424
SS           0.60640    0.08614   7.039 1.93e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1995.8  on 1507  degrees of freedom
Residual deviance: 1682.3  on 1500  degrees of freedom
AIC: 1698.3

Number of Fisher Scoring iterations: 4
```

```
## Call:
## lda(Mushrooms ~ Nscore + Escore + Oscore + AScore + Cscore +
##     Impulsive + SS, data = dtrain_mush)
##
## Prior probabilities of groups:
##         0         1
## 0.6246684 0.3753316
##
## Group means:
##        Nscore      Escore     Oscore      AScore     Cscore  Impulsive
## 0 -0.02384056 -0.02826404 -0.2800829  0.08490221  0.1443214 -0.1891101
## 1  0.08045947  0.00643318  0.4040378 -0.13672055 -0.2456964  0.3082904
##           SS
## 0 -0.2804036
## 1  0.4413136
##
## Coefficients of linear discriminants:
##                   LD1
## Nscore    -0.14660400
## Escore    -0.22430899
## Oscore     0.62655042
## AScore    -0.11337630
## Cscore    -0.24421024
## Impulsive  0.04968021
## SS         0.61184814
```

MUSHROOM



SS
Oscore
Impulsive
Cscore
AScore
0
10
20
30
40
Importance



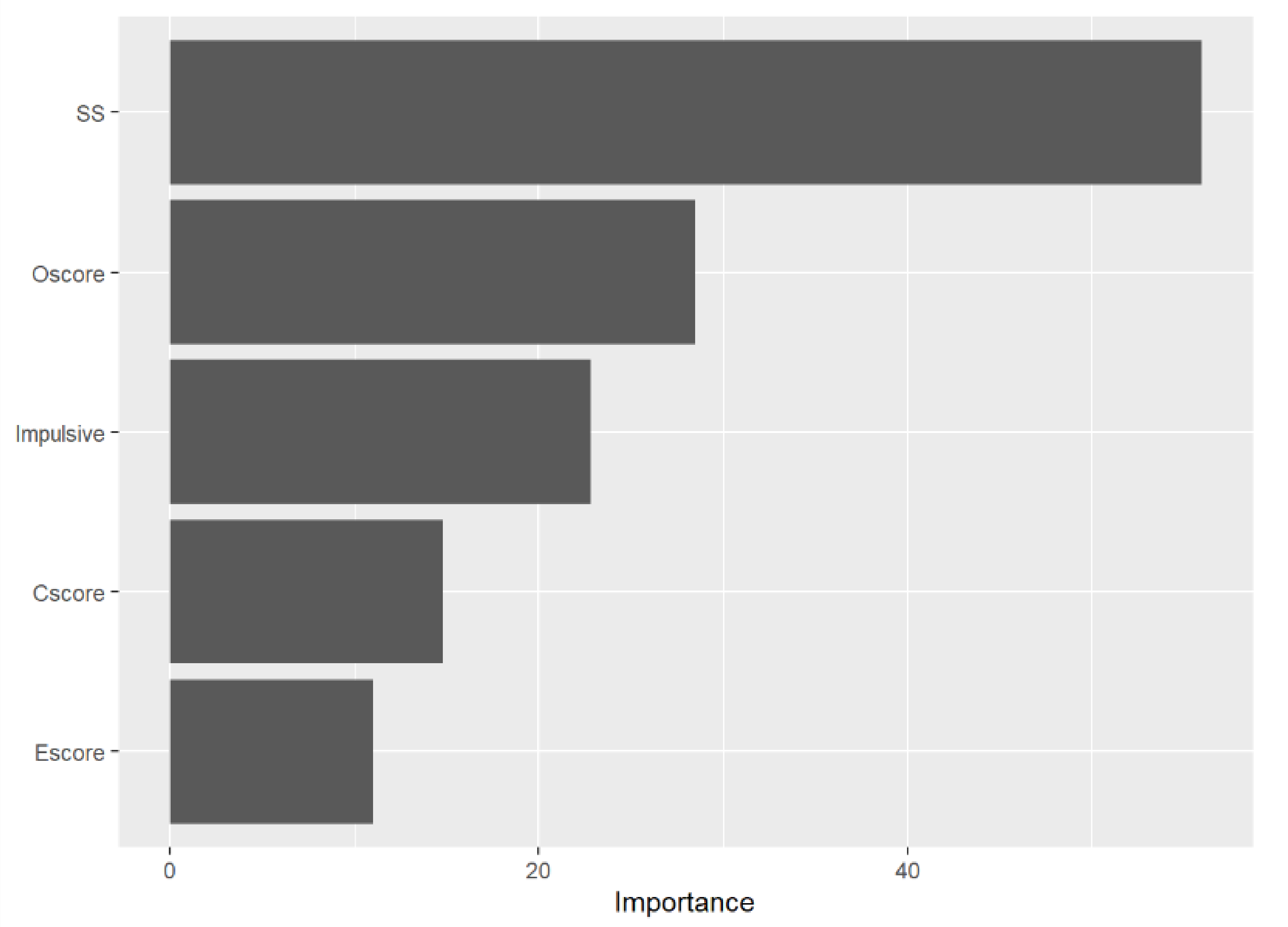
0.38
100%
yes
SS < -0.068
no
0.21
48%
0.53
52%
Oscore < 0.22
Oscore < 0.37
0.16
36%
0.42
26%
Cscore >= -0.96
Cscore >= 0.19
0.13
32%
0.4
4%
0.37
12%
0.25
8%
0.49
18%
0.64
25%

CANNABIS

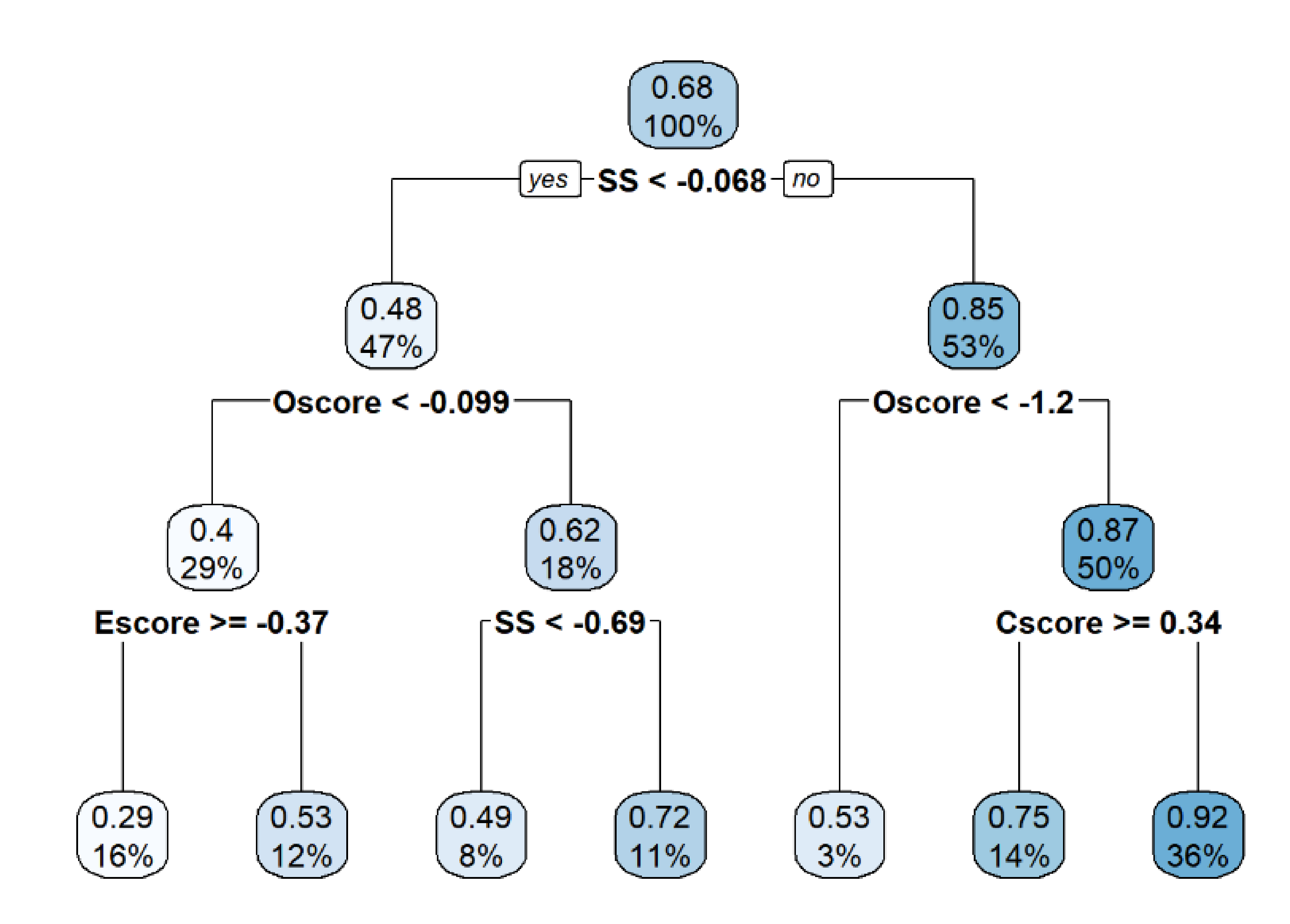
```
Call:
glm(formula = Cannabis ~ Nscore + Escore + Oscore + AScore + 
    Cscore + Impulsive + SS, family = "binomial", data = dtrain_cannab)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  1.02852    0.07011  14.670  < 2e-16 ***
Nscore       0.02529    0.07838   0.323 0.746912    
Escore      -0.30131    0.08421  -3.578 0.000346 ***
Oscore       0.68831    0.07785   8.842  < 2e-16 ***
AScore      -0.20274    0.07105  -2.853 0.004324 ** 
Cscore      -0.38280    0.07799  -4.908 9.18e-07 ***
Impulsive    0.01217    0.09012   0.135 0.892570    
SS           0.79708    0.09645   8.264  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1894.3  on 1507  degrees of freedom
Residual deviance: 1463.6  on 1500  degrees of freedom
AIC: 1479.6

Number of Fisher Scoring iterations: 5
```

```
## Call:
## lda(Cannabis ~ Nscore + Escore + Oscore + AScore + Cscore + Impulsive + 
##     SS, data = dtrain_cannab)
## 
## Prior probabilities of groups:
##        0        1 
## 0.321618 0.678382 
## 
## Group means:
##       Nscore      Escore     Oscore     AScore     Cscore  Impulsive         SS
## 0 -0.1853797  0.05035103 -0.4982536  0.2486108  0.3783960 -0.4033014 -0.5887662
## 1  0.1222753 -0.03053029  0.2273607 -0.1317151 -0.1851024  0.2021626  0.2721205
## 
## Coefficients of linear discriminants:
##                    LD1
## Nscore     0.016476869
## Escore    -0.228794529
## Oscore     0.558614816
## AScore    -0.172055558
## Cscore    -0.302448452
## Impulsive -0.001940207
## SS         0.668686637
```

CANNABIS



SS
Oscore
Impulsive
Cscore
Escore
0
20
40
Importance



0.68
100%
yes
SS < -0.068
no
0.48
47%
0.85
53%
Oscore < -0.099
Oscore < -1.2
0.4
29%
0.62
18%
0.87
50%
Escore >= -0.37
SS < -0.69
Cscore >= 0.34
0.29
16%
0.53
12%
0.49
8%
0.72
11%
0.53
3%
0.75
14%
0.92
36%

ROC CURVE

Comparing AUC

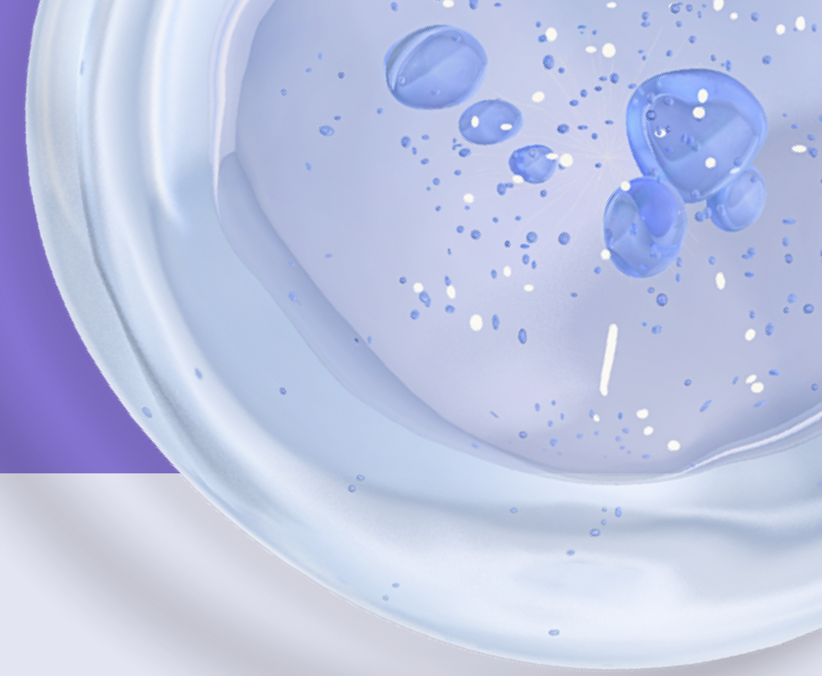
ALCOHOL: 0.5992795	ECSTASY: 0.7774887
BENZOS: 0.7265746	LSD: 0.7744562
NICOTINE: 0.6924293	MUSHROOMS: 0.7754536
COKE: 0.7038781	CANNABIS: 0.8305785

LDA ACCURACY

Comparing LDA predictions

ALCOHOL: 0.9654255	ECSTASY: 0.5851064
BENZOS: 0.6728723	LSD: 0.7446809
NICOTINE: 0.6914894	MUSHROOMS: 0.7287234
COKE: 0.6702128	CANNABIS: 0.7473404

Conclusions (1)

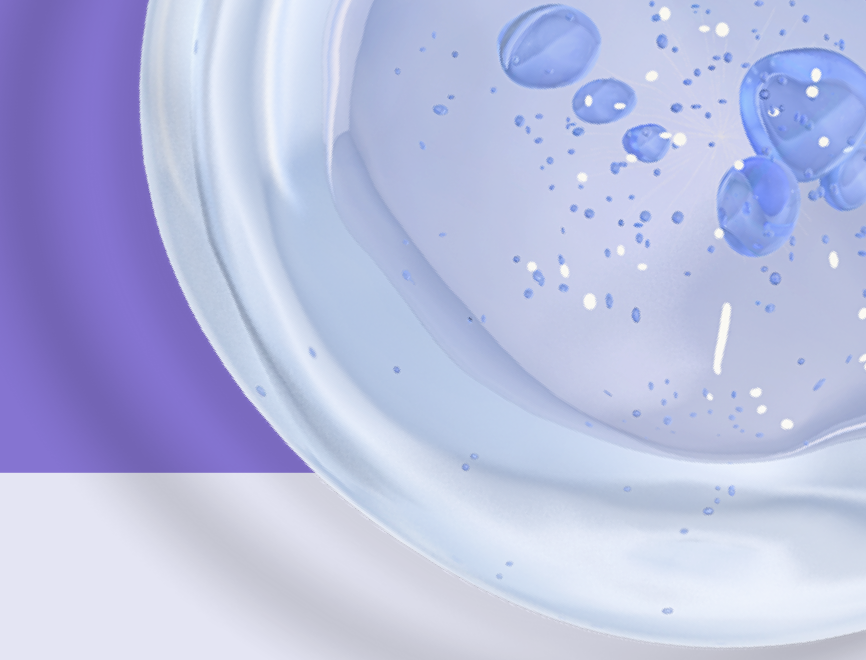


1. **Sensation Seeking** Score (SS) emerges as a crucial predictor for **both illegal and legal** drug use, with Cannabis users exhibiting the highest scores. Illegal drug users, in particular, are influenced by a combination of Impulsiveness and Openness traits, alongside SS.

2. **Legal drug** consumption analysis requires consideration of factors such as accessibility and societal perceptions, which can influence results significantly. **Neuroticism** (Nscore) emerges as a prominent predictor following SS, indicating its importance in identifying drug users.

3. Despite the high accuracy in predicting **Alcohol** user, AUC: 0.60 shows that the model has **no discrimination capacity** to distinguish between users and not-user, because of the imbalanced datasets.

Conclusions (2)



4. Certain drug consumptions exhibit strong correlations, indicating potential patterns in user behavior. Notably, **Mushrooms and LSD** display the strongest correlation **(0.67)**, with similar trait importance rankings. **Ecstasy and Coke** also exhibit a positive correlation **(0.61)**, but users differ as Coke show less impulsivity and openness.

5. Positive correlations are observed between **Mushrooms and Cannabis (0.58)**, with users of both drugs exhibiting high SS scores and similar levels of openness. Additionally, both **Ecstasy and Cannabis** users demonstrate **higher impulsivity** compared to other drug users.