**Conceptual vs. Observed Structure of HRV indices**

Data Analaysis

# 1 Introduction

## 1.1 Introduction

The aim of this study is to explore the factor structure of HRV indices.

## 1.2 Databases

### 1.2.1 Glasgow University Database

The GUDB Database (**howell2018high?**) contains ECGs from 25 subjects. Each subject was recorded performing 5 different tasks for two minutes (sitting, doing a maths test on a tablet, walking on a treadmill, running on a treadmill, using a hand bike). The sampling rate is 250Hz for all the conditions.

The script to download and format the database using the [**ECG-GUDB**](https://github.com/berndporr/ECG-GUDB) Python package by Bernd Porr can be found [**here**](https://github.com/neuropsychology/NeuroKit/blob/dev/data/gudb/download_gudb.py).

### 1.2.2 MIT-BIH Arrhythmia Database

The MIT-BIH Arrhythmia Database [MIT-Arrhythmia; (**moody2001impact?**)] contains 48 excerpts of 30-min of two-channel ambulatory ECG recordings sampled at 360Hz and 25 additional recordings from the same participants including common but clinically significant arrhythmias (denoted as the MIT-Arrhythmia-x database).

The script to download and format the database using the can be found [**here**](https://github.com/neuropsychology/NeuroKit/blob/dev/data/mit_arrhythmia/download_mit_arrhythmia.py).

### 1.2.3 MIT-BIH Normal Sinus Rhythm Database

This database includes 18 clean long-term ECG recordings of subjects. Due to memory limits, we only kept the second hour of recording of each participant.

The script to download and format the database using the can be found [**here**](https://github.com/neuropsychology/NeuroKit/blob/dev/data/mit_normal/download_mit_normal.py).

### 1.2.4 Fantasia Database

The Fantasia database (**iyengar1996age?**) consists of twenty young and twenty elderly healthy subjects. All subjects remained in a resting state in sinus rhythm while watching the movie Fantasia (Disney, 1940) to help maintain wakefulness. The continuous ECG signals were digitized at 250 Hz. Each heartbeat was annotated using an automated arrhythmia detection algorithm, and each beat annotation was verified by visual inspection.

## 1.3 Procedure

## 1.4 Results

library(tidyverse)  
library(easystats)  
  
data <- read.csv("data/data.csv", stringsAsFactors = FALSE) %>%   
 select(-HRV\_ULF, -HRV\_VLF) %>% # Empty  
 filter(Database != "LUDB") # too short recordings, many indices didn't converge  
names(data) <- stringr::str\_remove(names(data), "HRV\_")

### 1.4.1 Redundant Indices

#### 1.4.1.1 Remove Equivalent (r higher than .995)

data %>%   
 correlation::correlation() %>%   
 filter(abs(r) > 0.995) %>%   
 arrange(Parameter1, desc(abs(r)))

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameter1 | Parameter2 | r | CI\_low | CI\_high | t | df | p | Method | n\_Obs |
| C1d | C1a | -1 | -1 | -1 | -1.1e+09 | 250 | 0 | Pearson | 252 |
| C2d | C2a | -1 | -1 | -1 | -Inf | 250 | 0 | Pearson | 252 |
| Cd | Ca | -1 | -1 | -1 | -Inf | 250 | 0 | Pearson | 252 |
| RMSSD | SDSD | 1 | 1 | 1 | 5.0e+04 | 250 | 0 | Pearson | 252 |
| RMSSD | SD1 | 1 | 1 | 1 | 5.0e+04 | 250 | 0 | Pearson | 252 |
| RMSSD | SD1d | 1 | 1 | 1 | 5.4e+02 | 250 | 0 | Pearson | 252 |
| RMSSD | SD1a | 1 | 1 | 1 | 4.7e+02 | 250 | 0 | Pearson | 252 |
| SD1 | SD1d | 1 | 1 | 1 | 5.4e+02 | 250 | 0 | Pearson | 252 |
| SD1 | SD1a | 1 | 1 | 1 | 4.7e+02 | 250 | 0 | Pearson | 252 |
| SD1d | SD1a | 1 | 1 | 1 | 2.5e+02 | 250 | 0 | Pearson | 252 |
| SD2 | SD2a | 1 | 1 | 1 | 2.9e+02 | 250 | 0 | Pearson | 252 |
| SD2 | SD2d | 1 | 1 | 1 | 2.0e+02 | 250 | 0 | Pearson | 252 |
| SDNN | SDNNa | 1 | 1 | 1 | 7.3e+02 | 250 | 0 | Pearson | 252 |
| SDNN | SDNNd | 1 | 1 | 1 | 5.8e+02 | 250 | 0 | Pearson | 252 |
| SDNNd | SDNNa | 1 | 1 | 1 | 3.2e+02 | 250 | 0 | Pearson | 252 |
| SDSD | SD1 | 1 | 1 | 1 | Inf | 250 | 0 | Pearson | 252 |
| SDSD | SD1d | 1 | 1 | 1 | 5.4e+02 | 250 | 0 | Pearson | 252 |
| SDSD | SD1a | 1 | 1 | 1 | 4.7e+02 | 250 | 0 | Pearson | 252 |

data <- data %>%   
 select(-SDSD, -SD1, -SD1d, -SD1a, -CVSD) %>% # Same as RMSSD   
 select(-SDNNd, -SDNNa) %>% # Same as SDNN  
 select(-SD2d, -SD2a) %>% # Same as SD2  
 select(-Cd) %>% # Same as Ca  
 select(-C1d, -C2d) # Same as C1a and C2a

#### 1.4.1.2 Remove Strongly Correlated (r higher than .98)

data %>%   
 correlation::correlation() %>%   
 filter(abs(r) > 0.95) %>%  
 arrange(Parameter1, desc(abs(r)))

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameter1 | Parameter2 | r | CI\_low | CI\_high | t | df | p | Method | n\_Obs |
| CVNN | SD2 | 0.97 | 0.96 | 0.98 | 65 | 250 | 0 | Pearson | 252 |
| GI | AI | 0.99 | 0.99 | 0.99 | 138 | 250 | 0 | Pearson | 252 |
| GI | SI | 0.99 | 0.99 | 0.99 | 115 | 250 | 0 | Pearson | 252 |
| MeanNN | MedianNN | 0.99 | 0.98 | 0.99 | 99 | 250 | 0 | Pearson | 252 |
| PIP | IALS | 0.98 | 0.98 | 0.99 | 87 | 250 | 0 | Pearson | 252 |
| RMSSD | SDNN | 0.98 | 0.98 | 0.99 | 80 | 250 | 0 | Pearson | 252 |
| RMSSD | CVNN | 0.97 | 0.96 | 0.98 | 63 | 250 | 0 | Pearson | 252 |
| SDNN | SD2 | 0.99 | 0.99 | 0.99 | 120 | 250 | 0 | Pearson | 252 |
| SDNN | CVNN | 0.98 | 0.98 | 0.99 | 89 | 250 | 0 | Pearson | 252 |
| SI | AI | 0.97 | 0.96 | 0.98 | 63 | 250 | 0 | Pearson | 252 |
| TINN | S | 0.95 | 0.94 | 0.96 | 51 | 250 | 0 | Pearson | 252 |

data <- data %>%   
 select(-GI, -SI) %>% # Same as AI   
 select(-SD2) %>% # Same as SDNN  
 select(-MedianNN) %>% # Same as MeanNN  
 select(-IALS) %>% # Same as PIP  
 select(-SDNN, -CVNN) # Same as RMSSD

### 1.4.2 Recording Length

#### 1.4.2.1 Investigate effect

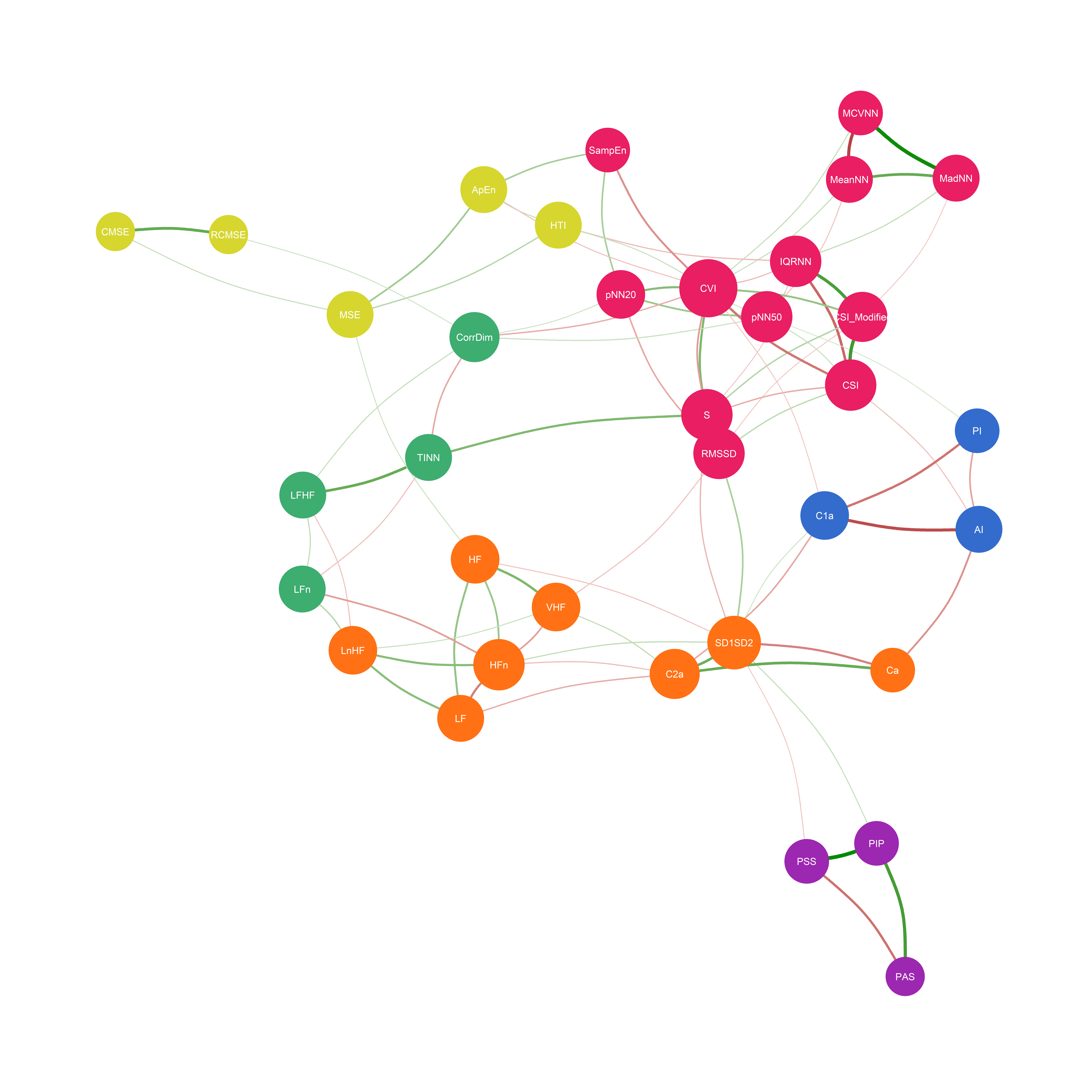
correlation(data) %>%   
 filter(Parameter2 == "Recording\_Length") %>%   
 arrange(desc(abs(r)))

#### 1.4.2.2 Adjust the data for recording length

data <- effectsize::adjust(data, effect="Recording\_Length") %>%   
 select(-Recording\_Length)

### 1.4.3 Gaussian Graphical Model

library(ggraph)  
  
data %>%   
 correlation::correlation(partial=FALSE) %>%   
 correlation::cor\_to\_pcor() %>%   
 filter(abs(r) > 0.2) %>%  
 tidygraph::as\_tbl\_graph(directed=FALSE) %>%   
 dplyr::mutate(closeness = tidygraph::centrality\_closeness(normalized = TRUE),  
 degree = tidygraph::centrality\_degree(normalized = TRUE),  
 betweeness = tidygraph::centrality\_betweenness(normalized = TRUE)) %>%  
 tidygraph::activate(nodes) %>%  
 dplyr::mutate(group1 = as.factor(tidygraph::group\_edge\_betweenness()),  
 # group2 = as.factor(tidygraph::group\_optimal()),  
 # group3 = as.factor(tidygraph::group\_walktrap()),  
 # group4 = as.factor(tidygraph::group\_spinglass()),  
 group5 = as.factor(tidygraph::group\_louvain())) %>%   
 ggraph::ggraph(layout = "fr") +  
 ggraph::geom\_edge\_arc(aes(colour = r, edge\_width = abs(r)), strength = 0.1, show.legend = FALSE) +  
 ggraph::geom\_node\_point(aes(size = degree, color = group5), show.legend = FALSE) +  
 ggraph::geom\_node\_text(aes(label = name), colour = "white") +  
 ggraph::scale\_edge\_color\_gradient2(low = "#a20025", high = "#008a00", name = "r") +  
 ggraph::theme\_graph() +  
 guides(edge\_width = FALSE) +  
 scale\_x\_continuous(expand = expansion(c(.10, .10))) +  
 scale\_y\_continuous(expand = expansion(c(.10, .10))) +  
 scale\_size\_continuous(range = c(20, 30)) +  
 scale\_edge\_width\_continuous(range = c(0.5, 2)) +  
 see::scale\_color\_material\_d(palette="rainbow", reverse=TRUE)



Groups were identified using the [tidygraph::group\_optimal](https://rdrr.io/cran/tidygraph/man/group_graph.html) algorithm.

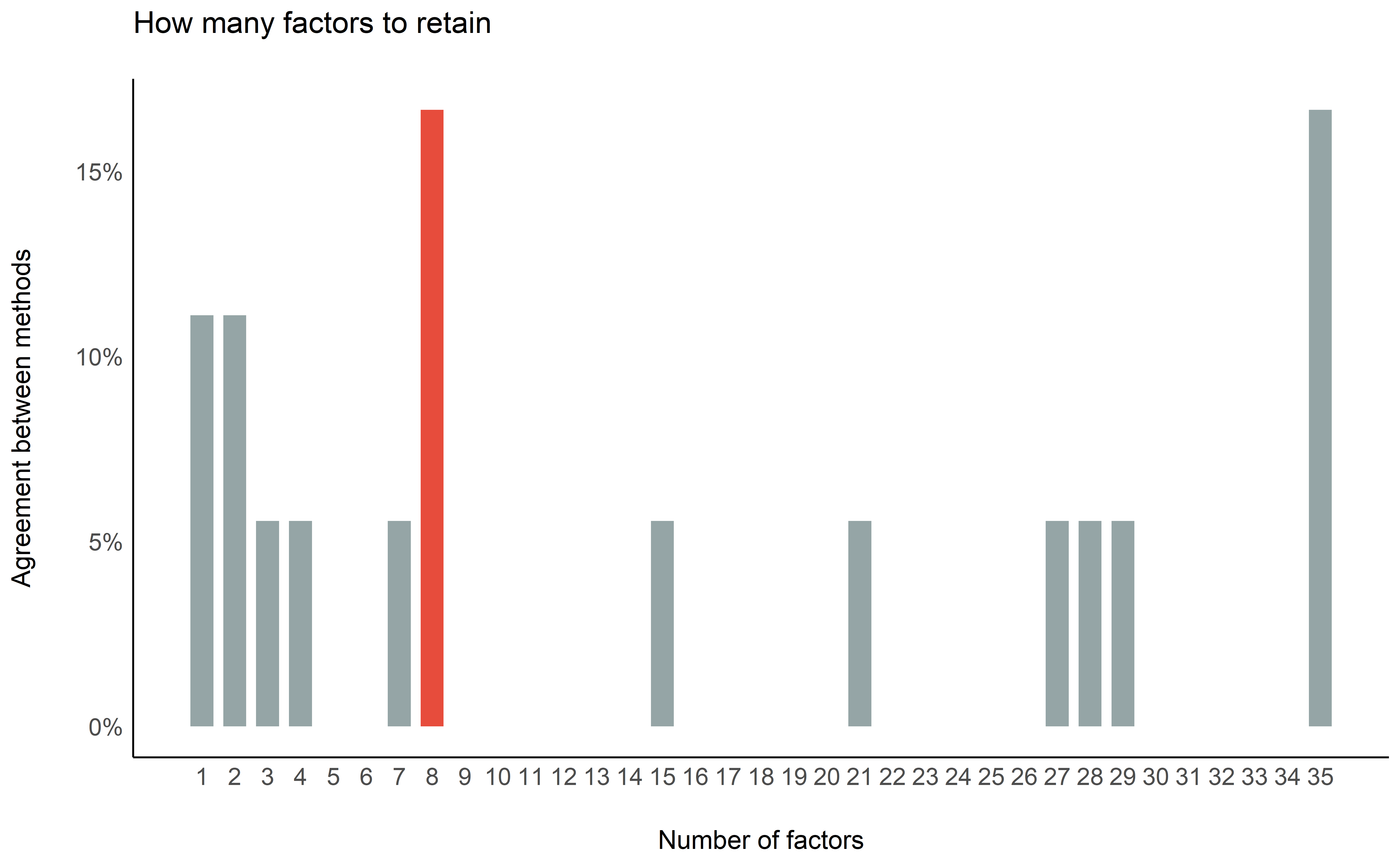
### 1.4.4 Factor Analysis

#### 1.4.4.1 How many factors

cor <- correlation::correlation(data[sapply(data, is.numeric)]) %>%   
 as.matrix()  
  
n <- parameters::n\_factors(data, cor=cor)  
  
n

|  |  |  |
| --- | --- | --- |
| n\_Factors | Method | Family |
| 1 | t | Multiple\_regression |
| 1 | p | Multiple\_regression |
| 2 | Acceleration factor | Scree |
| 2 | RMSEA | Fit |
| 3 | CNG | CNG |
| 4 | beta | Multiple\_regression |
| 7 | R2 | Scree\_SE |
| 8 | Optimal coordinates | Scree |
| 8 | Parallel analysis | Scree |
| 8 | Kaiser criterion | Scree |
| 15 | SE Scree | Scree\_SE |
| 21 | BIC | Fit |
| 27 | CRMS | Fit |
| 28 | TLI | Fit |
| 29 | Bentler | Bentler |
| 35 | Bartlett | Barlett |
| 35 | Anderson | Barlett |
| 35 | Lawley | Barlett |

plot(n) +  
 see::theme\_modern()

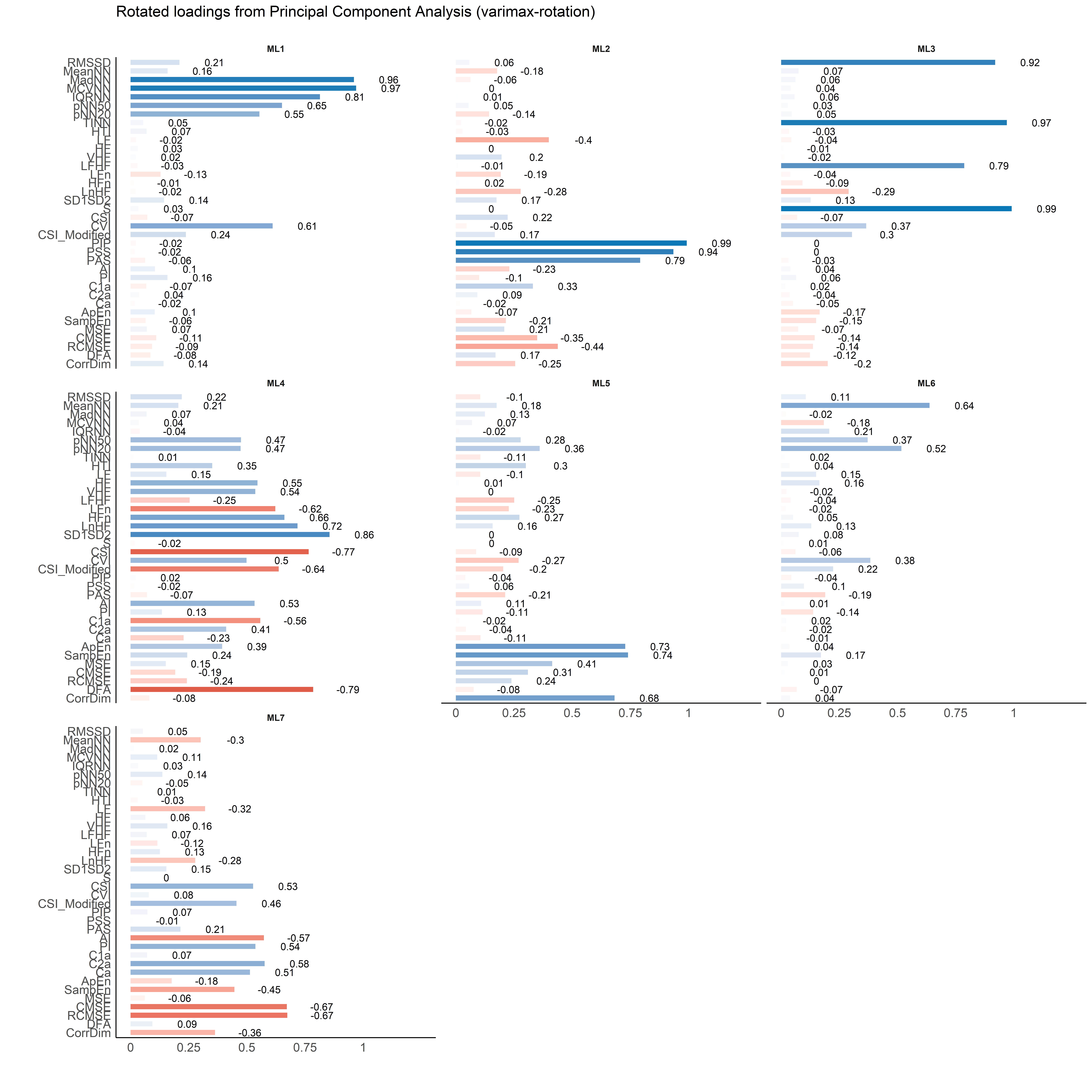


#### 1.4.4.2 Exploratory Factor Analysis (EFA)

efa <- parameters::factor\_analysis(data, cor=cor, n=7, rotation="varimax", fm="ml")  
  
print(efa, threshold="max", sort=TRUE)

> # Rotated loadings from Factor Analysis (varimax-rotation)  
>   
> Variable | ML4 | ML1 | ML3 | ML2 | ML7 | ML5 | ML6 | Complexity | Uniqueness  
> ------------------------------------------------------------------------------------------  
> SD1SD2 | 0.86 | | | | | | | 1.28 | 0.17   
> DFA | -0.79 | | | | | | | 1.24 | 0.31   
> CSI | -0.77 | | | | | | | 2.06 | 0.06   
> LnHF | 0.72 | | | | | | | 2.22 | 0.20   
> HFn | 0.66 | | | | | | | 1.47 | 0.46   
> CSI\_Modified | -0.64 | | | | | | | 3.47 | 0.12   
> LFn | -0.62 | | | | | | | 1.67 | 0.49   
> C1a | -0.56 | | | | | | | 1.71 | 0.57   
> HF | 0.55 | | | | | | | 1.22 | 0.67   
> VHF | 0.54 | | | | | | | 1.47 | 0.65   
> HTI | 0.35 | | | | | | | 2.12 | 0.78   
> MCVNN | | 0.97 | | | | | | 1.12 | 4.99e-03   
> MadNN | | 0.96 | | | | | | 1.06 | 0.05   
> IQRNN | | 0.81 | | | | | | 1.15 | 0.29   
> pNN50 | | 0.65 | | | | | | 3.08 | 0.11   
> CVI | | 0.61 | | | | | | 3.93 | 0.02   
> pNN20 | | 0.55 | | | | | | 3.90 | 0.05   
> S | | | 0.99 | | | | | 1.00 | 0.02   
> TINN | | | 0.97 | | | | | 1.03 | 0.04   
> RMSSD | | | 0.92 | | | | | 1.30 | 0.03   
> LFHF | | | 0.79 | | | | | 1.45 | 0.25   
> PIP | | | | 0.99 | | | | 1.02 | 4.87e-03   
> PSS | | | | 0.94 | | | | 1.03 | 0.11   
> PAS | | | | 0.79 | | | | 1.46 | 0.24   
> LF | | | | -0.40 | | | | 2.79 | 0.68   
> RCMSE | | | | | -0.67 | | | 2.49 | 0.21   
> CMSE | | | | | -0.67 | | | 2.38 | 0.26   
> C2a | | | | | 0.58 | | | 1.91 | 0.49   
> AI | | | | | -0.57 | | | 2.48 | 0.31   
> PI | | | | | 0.54 | | | 1.68 | 0.62   
> Ca | | | | | 0.51 | | | 1.51 | 0.67   
> SampEn | | | | | | 0.74 | | 2.38 | 0.09   
> ApEn | | | | | | 0.73 | | 1.88 | 0.24   
> CorrDim | | | | | | 0.68 | | 2.22 | 0.27   
> MSE | | | | | | 0.41 | | 2.00 | 0.75   
> MeanNN | | | | | | | 0.64 | 2.25 | 0.37   
>   
> The 7 latent factors (varimax rotation) accounted for 70.43% of the total variance of the original data (ML4 = 17.93%, ML1 = 10.90%, ML3 = 10.87%, ML2 = 10.17%, ML7 = 9.55%, ML5 = 7.45%, ML6 = 3.57%).

plot(efa) +  
 see::theme\_modern()



#### 1.4.4.3 Confirmatory Factor Analysis (CFA)

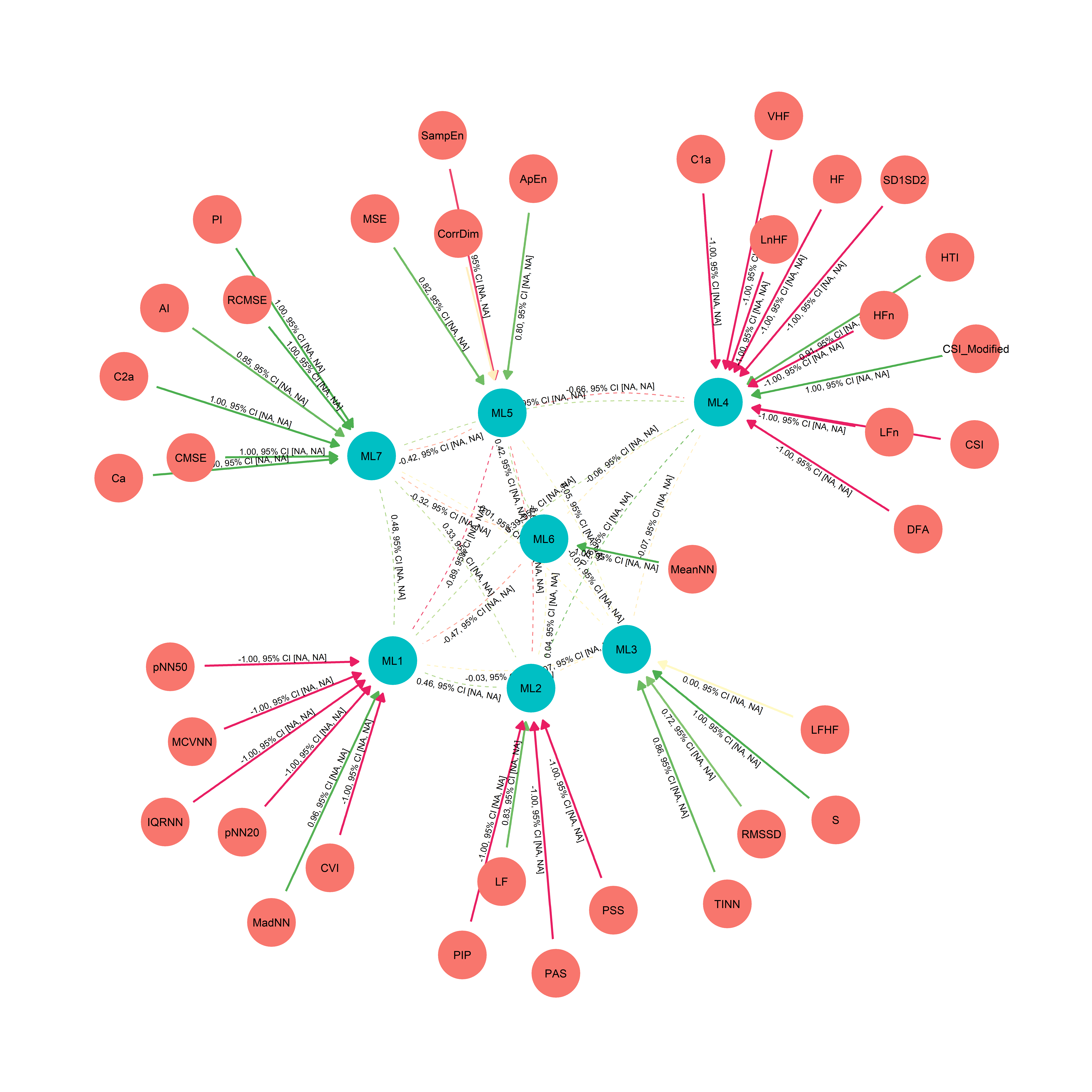
library(lavaan)  
  
model <- parameters::efa\_to\_cfa(efa, threshold = "max")  
cfa <- lavaan::cfa(model, data=data) %>%  
 parameters::parameters(standardize=TRUE)

> Error in if (ncol(S) == 1L) { : argument is of length zero

cfa

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | To | Operator | From | Coefficient | SE | CI\_low | CI\_high | p | Type |
| 1 | ML4 | =~ | HTI | 0.91 |  |  |  | 0 | Loading |
| 2 | ML4 | =~ | HF | -1.00 |  |  |  | 0 | Loading |
| 3 | ML4 | =~ | VHF | -1.00 |  |  |  | 0 | Loading |
| 4 | ML4 | =~ | LFn | -1.00 |  |  |  | 0 | Loading |
| 5 | ML4 | =~ | HFn | -1.00 |  |  |  | 0 | Loading |
| 6 | ML4 | =~ | LnHF | -1.00 |  |  |  | 0 | Loading |
| 7 | ML4 | =~ | SD1SD2 | -1.00 |  |  |  | 0 | Loading |
| 8 | ML4 | =~ | CSI | -1.00 |  |  |  | 0 | Loading |
| 9 | ML4 | =~ | CSI\_Modified | 1.00 |  |  |  | 0 | Loading |
| 10 | ML4 | =~ | C1a | -1.00 |  |  |  | 0 | Loading |
| 11 | ML4 | =~ | DFA | -1.00 |  |  |  | 0 | Loading |
| 12 | ML1 | =~ | MadNN | 0.96 |  |  |  | 0 | Loading |
| 13 | ML1 | =~ | MCVNN | -1.00 |  |  |  | 0 | Loading |
| 14 | ML1 | =~ | IQRNN | -1.00 |  |  |  | 0 | Loading |
| 15 | ML1 | =~ | pNN50 | -1.00 |  |  |  | 0 | Loading |
| 16 | ML1 | =~ | pNN20 | -1.00 |  |  |  | 0 | Loading |
| 17 | ML1 | =~ | CVI | -1.00 |  |  |  | 0 | Loading |
| 18 | ML3 | =~ | RMSSD | 0.72 |  |  |  | 0 | Loading |
| 19 | ML3 | =~ | TINN | 0.86 |  |  |  | 0 | Loading |
| 20 | ML3 | =~ | LFHF | 0.00 |  |  |  | 0 | Loading |
| 21 | ML3 | =~ | S | 1.00 |  |  |  | 0 | Loading |
| 22 | ML2 | =~ | LF | 0.83 |  |  |  | 0 | Loading |
| 23 | ML2 | =~ | PIP | -1.00 |  |  |  | 0 | Loading |
| 24 | ML2 | =~ | PSS | -1.00 |  |  |  | 0 | Loading |
| 25 | ML2 | =~ | PAS | -1.00 |  |  |  | 0 | Loading |
| 26 | ML7 | =~ | AI | 0.85 |  |  |  | 0 | Loading |
| 27 | ML7 | =~ | PI | 1.00 |  |  |  | 0 | Loading |
| 28 | ML7 | =~ | C2a | 1.00 |  |  |  | 0 | Loading |
| 29 | ML7 | =~ | Ca | 1.00 |  |  |  | 0 | Loading |
| 30 | ML7 | =~ | CMSE | 1.00 |  |  |  | 0 | Loading |
| 31 | ML7 | =~ | RCMSE | 1.00 |  |  |  | 0 | Loading |
| 32 | ML5 | =~ | ApEn | 0.80 |  |  |  | 0 | Loading |
| 33 | ML5 | =~ | SampEn | -0.90 |  |  |  | 0 | Loading |
| 34 | ML5 | =~ | MSE | 0.82 |  |  |  | 0 | Loading |
| 35 | ML5 | =~ | CorrDim | -0.06 |  |  |  | 0 | Loading |
| 36 | ML6 | =~ | MeanNN | 1.00 |  |  |  | 0 | Loading |
| 80 | ML4 | ~~ | ML1 | 0.39 |  |  |  | 0 | Correlation |
| 81 | ML4 | ~~ | ML3 | -0.07 |  |  |  | 0 | Correlation |
| 82 | ML4 | ~~ | ML2 | 0.76 |  |  |  | 0 | Correlation |
| 83 | ML4 | ~~ | ML7 | 0.39 |  |  |  | 0 | Correlation |
| 84 | ML4 | ~~ | ML5 | -0.66 |  |  |  | 0 | Correlation |
| 85 | ML4 | ~~ | ML6 | -0.06 |  |  |  | 0 | Correlation |
| 86 | ML1 | ~~ | ML3 | -0.03 |  |  |  | 0 | Correlation |
| 87 | ML1 | ~~ | ML2 | 0.46 |  |  |  | 0 | Correlation |
| 88 | ML1 | ~~ | ML7 | 0.48 |  |  |  | 0 | Correlation |
| 89 | ML1 | ~~ | ML5 | -0.89 |  |  |  | 0 | Correlation |
| 90 | ML1 | ~~ | ML6 | -0.47 |  |  |  | 0 | Correlation |
| 91 | ML3 | ~~ | ML2 | -0.07 |  |  |  | 0 | Correlation |
| 92 | ML3 | ~~ | ML7 | -0.01 |  |  |  | 0 | Correlation |
| 93 | ML3 | ~~ | ML5 | 0.05 |  |  |  | 0 | Correlation |
| 94 | ML3 | ~~ | ML6 | -0.01 |  |  |  | 0 | Correlation |
| 95 | ML2 | ~~ | ML7 | 0.33 |  |  |  | 0 | Correlation |
| 96 | ML2 | ~~ | ML5 | -0.64 |  |  |  | 0 | Correlation |
| 97 | ML2 | ~~ | ML6 | 0.04 |  |  |  | 0 | Correlation |
| 98 | ML7 | ~~ | ML5 | -0.42 |  |  |  | 0 | Correlation |
| 99 | ML7 | ~~ | ML6 | -0.32 |  |  |  | 0 | Correlation |
| 100 | ML5 | ~~ | ML6 | 0.42 |  |  |  | 0 | Correlation |

plot(cfa)



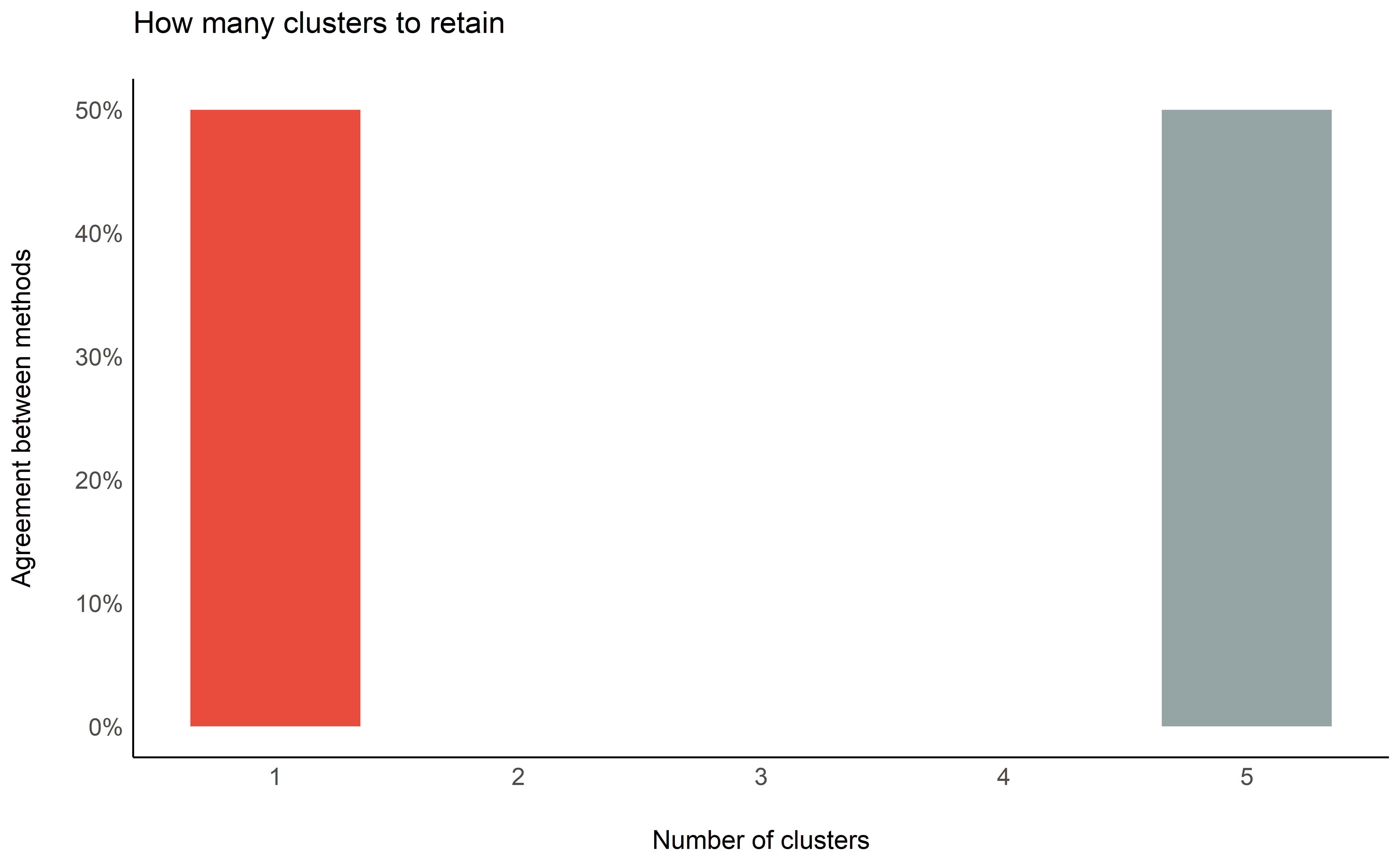
### 1.4.5 Cluster Analysis

#### 1.4.5.1 How many clusters

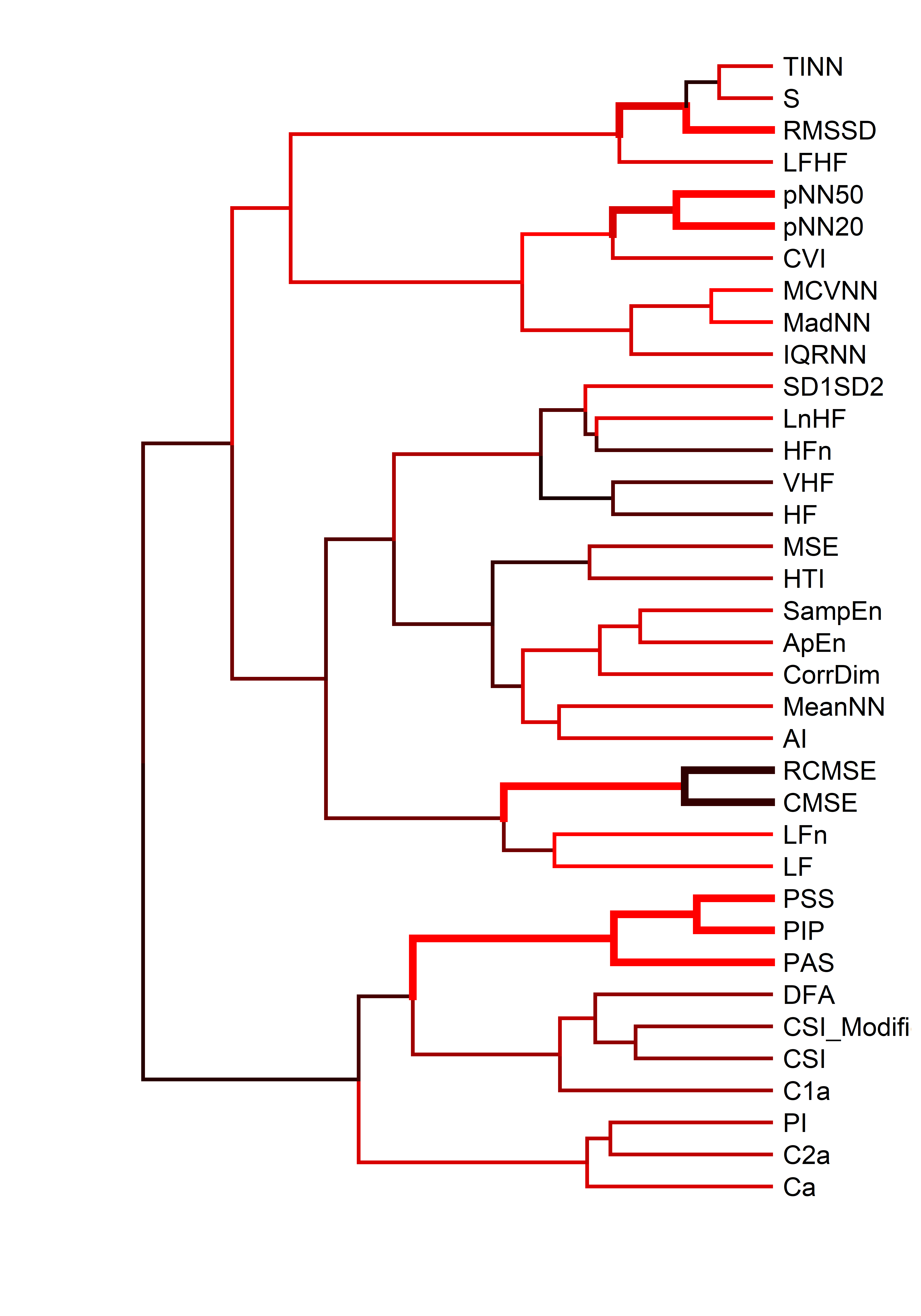
dat <- effectsize::standardize(data[sapply(data, is.numeric)])  
  
n <- parameters::n\_clusters(t(dat), package = c("mclust", "cluster"))  
  
n

|  |  |  |
| --- | --- | --- |
| n\_Clusters | Method | Package |
| 1 | Tibs2001SEmax | cluster |
| 5 | Mixture | mclust |

plot(n) +  
 theme\_modern()



library(dendextend)  
  
dat <- effectsize::standardize(data[sapply(data, is.numeric)])  
  
result <- pvclust::pvclust(dat, method.dist="euclidean", method.hclust="ward.D2", nboot=10, quiet=TRUE)  
  
result %>%   
 as.dendrogram() %>%   
 sort() %>%   
 dendextend::pvclust\_show\_signif\_gradient(result, signif\_col\_fun = grDevices::colorRampPalette(c("black", "red"))) %>%   
 dendextend::pvclust\_show\_signif(result, signif\_value = c(2, 1)) %>%  
 dendextend::as.ggdend() %>%   
 ggplot2::ggplot(horiz=TRUE, offset\_labels = -1)



## 1.5 References