Programa de Pós-Graduação em Computação Aplicada – PPGCAP
Universidade Federal do Pampa – Campus Bagé

Aplicação de Técnicas de Visualização em Dados de Desempenho e Frequência de Estudantes de uma Disciplina de Arquitetura de Computadores

Visualização de Dados

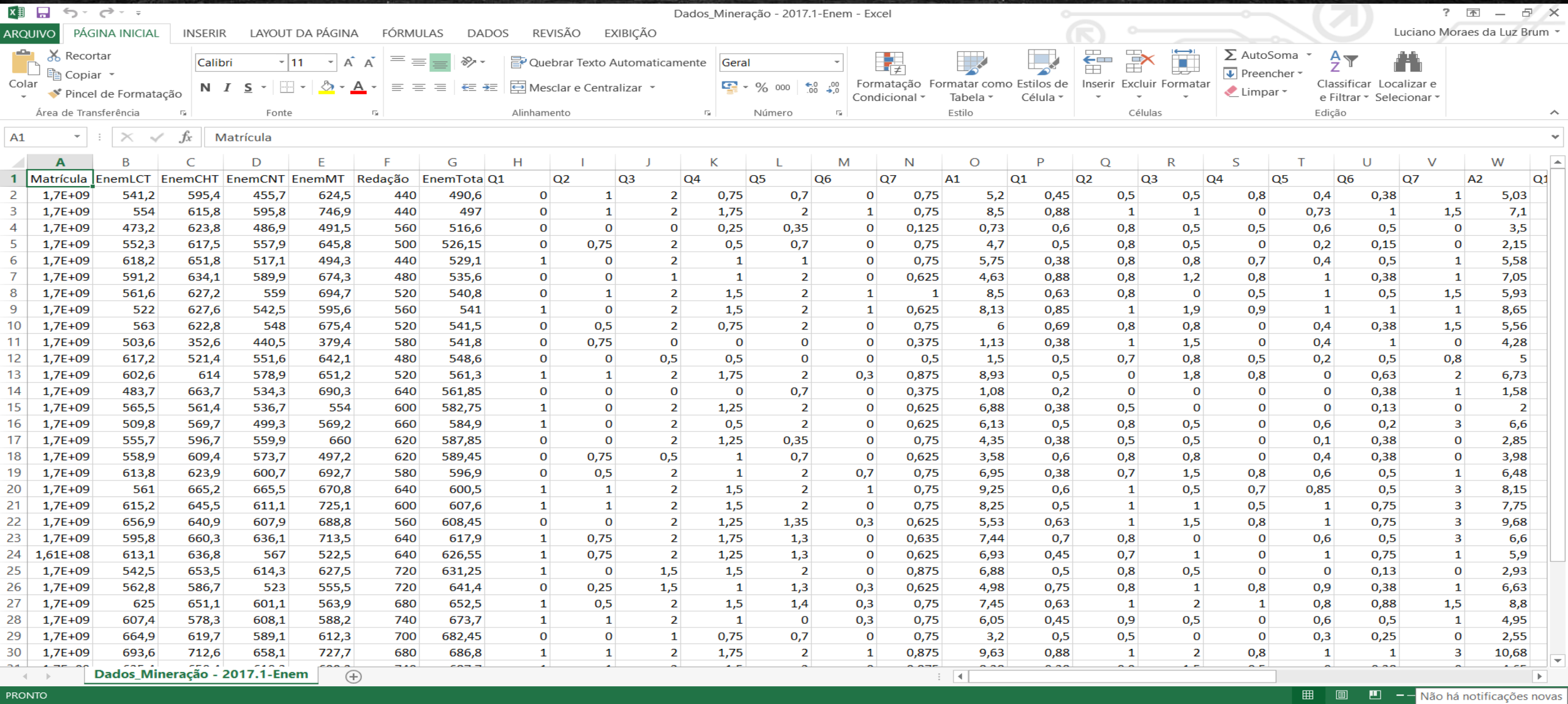
Discente: Luciano Moraes Da Luz Brum.

Docente: Dr. Sandro da Silva Camargo.

Objetivo

- Aplicar técnicas de **visualização de dados** utilizando uma base de dados de desempenho e frequência de estudantes de uma disciplina introdutória de Arquitetura de Computadores.
  - Diversos exemplos foram gerados.
  - Dinâmica de Apresentação: **Base de dados** – **Gráfico Resultante** - **Código R**.
  - Ferramenta Utilizada: **IDE RStudio (linguagem R).**

Base de Dados



Dados_Mineração - 2017.1-Enem - Excel

ARQUIVO | PÁGINA INICIAL | INSERIR | LAYOUT DA PÁGINA | FÓRMULAS | DADOS | REVISÃO | EXIBIÇÃO | Luciano Moraes da Luz Brum

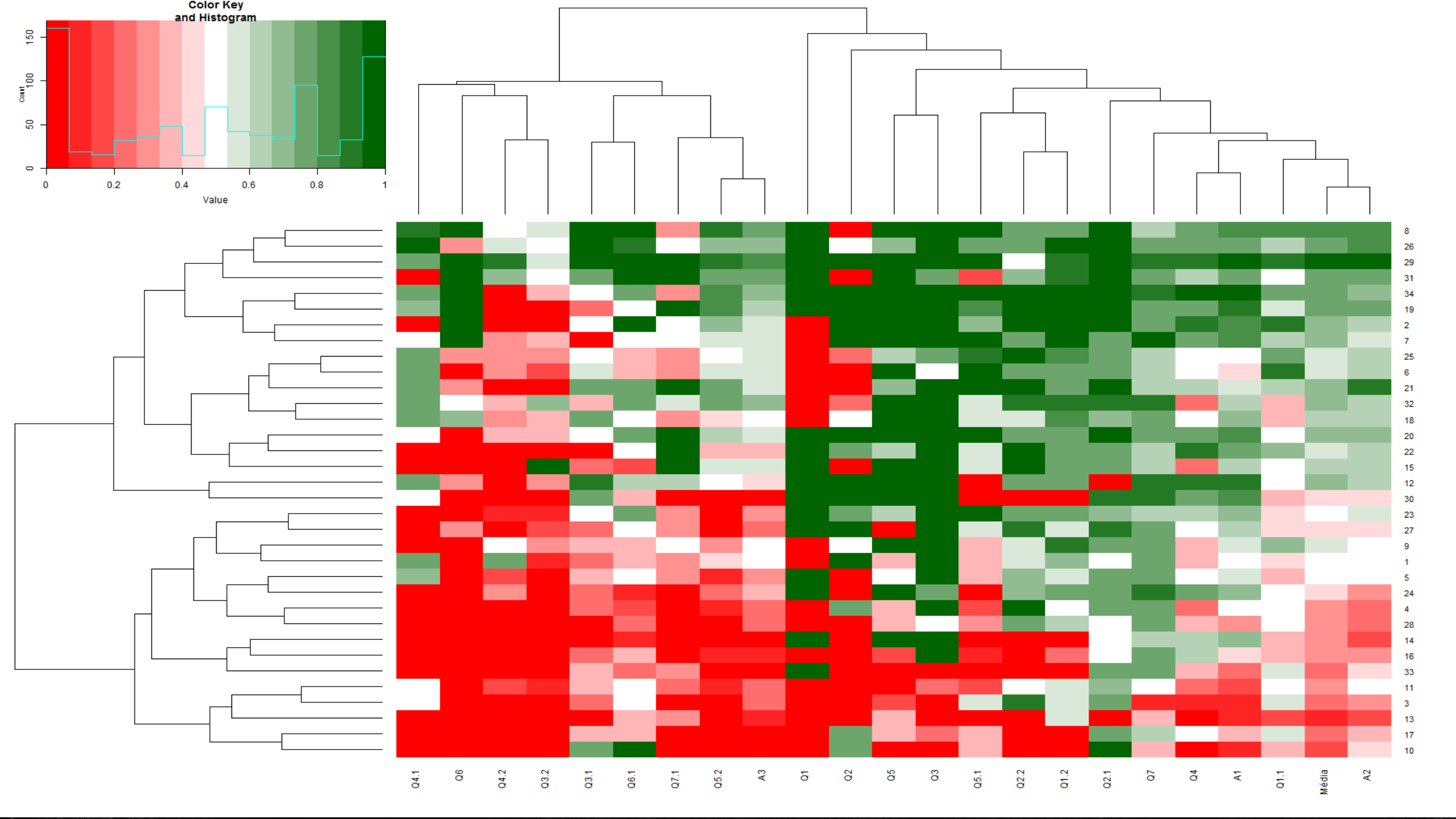
Recortar | Copiar | Pincel de Formatação | Colar | Área de Transferência | Calibri | 11 | Fonte | Quebrar Texto Automaticamente | Mesclar e Centralizar | Alinhamento | Geral | Número | Formatação Condicional | Formatar como Tabela | Estilos de Célula | Estilo | Inserir | Excluir | Formatar | Células | AutoSoma | Preencher | Limpar | Classificar e Filtrar | Localizar e Selecionar | Edição

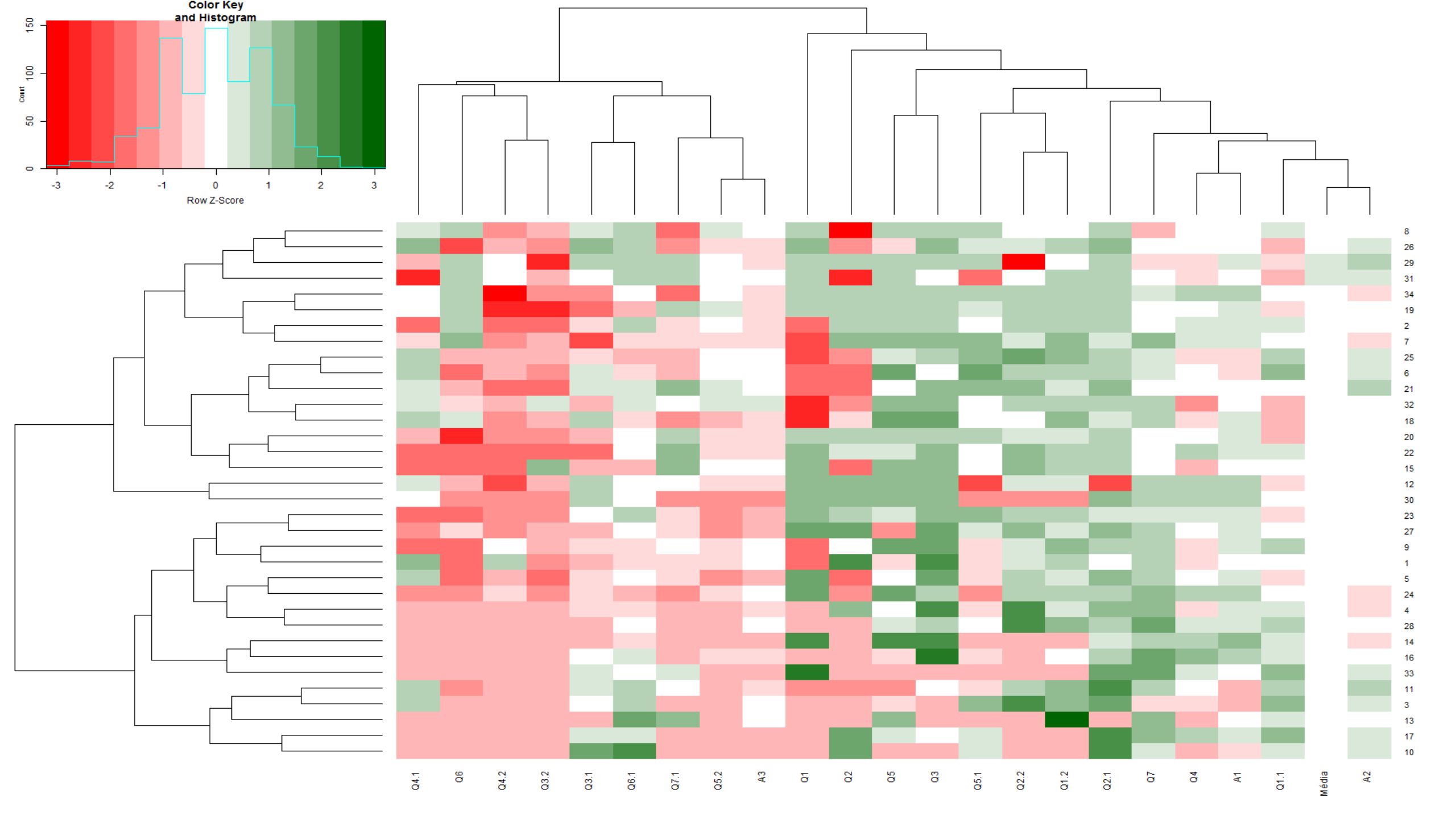
A1 | Matrícula

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Matrícula	EnemLCT	EnemCHT	EnemCNT	EnemMT	Redação	EnemTota	Q1	Q2	Q3	Q4	Q5	Q6	Q7	A1	Q1	Q2	Q3	Q4	Q5	Q6	Q7	A2	Q1
2	1,7E+09	541,2	595,4	455,7	624,5	440	490,6	0	1	2	0,75	0,7	0	0,75	5,2	0,45	0,5	0,5	0,8	0,4	0,38	1	5,03	
3	1,7E+09	554	615,8	595,8	746,9	440	497	0	1	2	1,75	2	1	0,75	8,5	0,88	1	1	0	0,73	1	1,5	7,1	
4	1,7E+09	473,2	623,8	486,9	491,5	560	516,6	0	0	0	0,25	0,35	0	0,125	0,73	0,6	0,8	0,5	0,5	0,6	0,5	0	3,5	
5	1,7E+09	552,3	617,5	557,9	645,8	500	526,15	0	0,75	2	0,5	0,7	0	0,75	4,7	0,5	0,8	0,5	0	0,2	0,15	0	2,15	
6	1,7E+09	618,2	651,8	517,1	494,3	440	529,1	1	0	2	1	1	0	0,75	5,75	0,38	0,8	0,8	0,7	0,4	0,5	1	5,58	
7	1,7E+09	591,2	634,1	589,9	674,3	480	535,6	0	0	1	1	2	0	0,625	4,63	0,88	0,8	1,2	0,8	1	0,38	1	7,05	
8	1,7E+09	561,6	627,2	559	694,7	520	540,8	0	1	2	1,5	2	1	1	8,5	0,63	0,8	0	0,5	1	0,5	1,5	5,93	
9	1,7E+09	522	627,6	542,5	595,6	560	541	1	0	2	1,5	2	1	0,625	8,13	0,85	1	1,9	0,9	1	1	1	8,65	
10	1,7E+09	563	622,8	548	675,4	520	541,5	0	0,5	2	0,75	2	0	0,75	6	0,69	0,8	0,8	0	0,4	0,38	1,5	5,56	
11	1,7E+09	503,6	352,6	440,5	379,4	580	541,8	0	0,75	0	0	0	0	0,375	1,13	0,38	1	1,5	0	0,4	1	0	4,28	
12	1,7E+09	617,2	521,4	551,6	642,1	480	548,6	0	0	0,5	0,5	0	0	0,5	1,5	0,5	0,7	0,8	0,5	0,2	0,5	0,8	5	
13	1,7E+09	602,6	614	578,9	651,2	520	561,3	1	1	2	1,75	2	0,3	0,875	8,93	0,5	0	1,8	0,8	0	0,63	2	6,73	
14	1,7E+09	483,7	663,7	534,3	690,3	640	561,85	0	0	0	0	0,7	0	0,375	1,08	0,2	0	0	0	0	0,38	1	1,58	
15	1,7E+09	565,5	561,4	536,7	554	600	582,75	1	0	2	1,25	2	0	0,625	6,88	0,38	0,5	0	0	0	0,13	0	2	
16	1,7E+09	509,8	569,7	499,3	569,2	660	584,9	1	0	2	0,5	2	0	0,625	6,13	0,5	0,8	0,5	0	0,6	0,2	3	6,6	
17	1,7E+09	555,7	596,7	559,9	660	620	587,85	0	0	2	1,25	0,35	0	0,75	4,35	0,38	0,5	0,5	0	0,1	0,38	0	2,85	
18	1,7E+09	558,9	609,4	573,7	497,2	620	589,45	0	0,75	0,5	1	0,7	0	0,625	3,58	0,6	0,8	0,8	0	0,4	0,38	0	3,98	
19	1,7E+09	613,8	623,9	600,7	692,7	580	596,9	0	0,5	2	1	2	0,7	0,75	6,95	0,38	0,7	1,5	0,8	0,6	0,5	1	6,48	
20	1,7E+09	561	665,2	665,5	670,8	640	600,5	1	1	2	1,5	2	1	0,75	9,25	0,6	1	0,5	0,7	0,85	0,5	3	8,15	
21	1,7E+09	615,2	645,5	611,1	725,1	600	607,6	1	1	2	1,5	2	0	0,75	8,25	0,5	1	1	0,5	1	0,75	3	7,75	
22	1,7E+09	656,9	640,9	607,9	688,8	560	608,45	0	0	2	1,25	1,35	0,3	0,625	5,53	0,63	1	1,5	0,8	1	0,75	3	9,68	
23	1,7E+09	595,8	660,3	636,1	713,5	640	617,9	1	0,75	2	1,75	1,3	0	0,635	7,44	0,7	0,8	0	0	0,6	0,5	3	6,6	
24	1,61E+08	613,1	636,8	567	522,5	640	626,55	1	0,75	2	1,25	1,3	0	0,625	6,93	0,45	0,7	1	0	1	0,75	1	5,9	
25	1,7E+09	542,5	653,5	614,3	627,5	720	631,25	1	0	1,5	1,5	2	0	0,875	6,88	0,5	0,8	0,5	0	0	0,13	0	2,93	
26	1,7E+09	562,8	586,7	523	555,5	720	641,4	0	0,25	1,5	1	1,3	0,3	0,625	4,98	0,75	0,8	1	0,8	0,9	0,38	1	6,63	
27	1,7E+09	625	651,1	601,1	563,9	680	652,5	1	0,5	2	1,5	1,4	0,3	0,75	7,45	0,63	1	2	1	0,8	0,88	1,5	8,8	
28	1,7E+09	607,4	578,3	608,1	588,2	740	673,7	1	1	2	1	0	0,3	0,75	6,05	0,45	0,9	0,5	0	0,6	0,5	1	4,95	
29	1,7E+09	664,9	619,7	589,1	612,3	700	682,45	0	0	1	0,75	0,7	0	0,75	3,2	0,5	0,5	0	0	0,3	0,25	0	2,55	
30	1,7E+09	693,6	712,6	658,1	727,7	680	686,8	1	1	2	1,75	2	1	0,875	9,63	0,88	1	2	0,8	1	1	3	10,68	

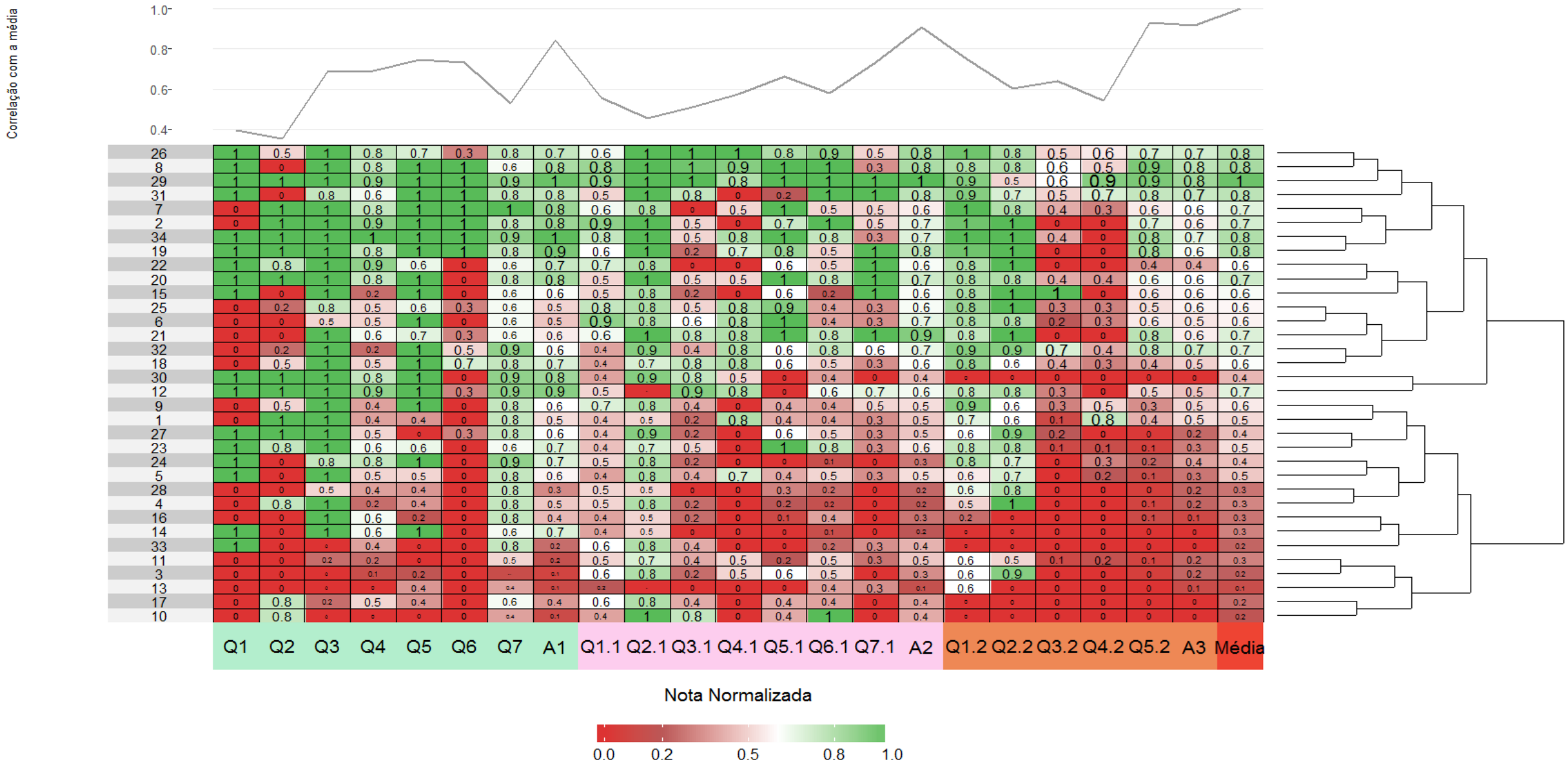
Dados_Mineração - 2017.1-Enem

PRONTO | Não há notificações novas

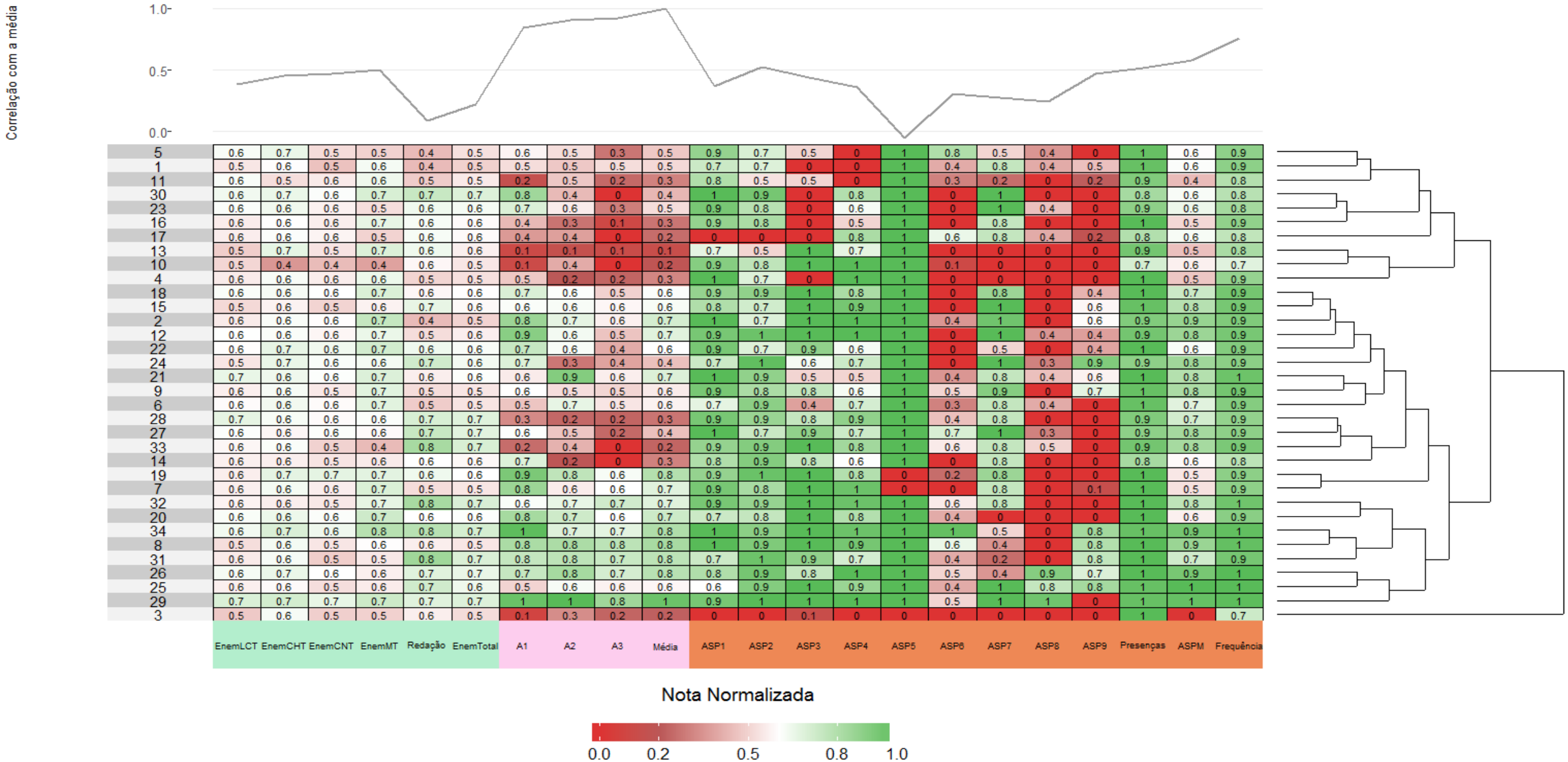

Color Key
and Histogram
Count
Value
0
50
100
150
0
0.2
0.4
0.6
0.8
1
8
26
29
31
34
19
2
7
25
6
21
32
18
20
22
15
12
30
23
27
9
1
5
24
4
28
14
16
33
11
3
13
17
10
Q4.1
Q6
Q4.2
Q3.2
Q3.1
Q6.1
Q7.1
Q5.2
A3
Q1
Q2
Q5
Q3
Q5.1
Q2.2
Q1.2
Q2.1
Q7
Q4
A1
Q1.1
Média
A2


Color Key
and Histogram
Count
0
50
100
150
-3
-2
-1
0
1
2
3
Row Z-Score
8
26
29
31
34
19
2
7
25
6
21
32
18
20
22
15
12
30
23
27
9
1
5
24
4
28
14
16
33
11
3
13
17
10
Q4.1
Q6
Q4.2
Q3.2
Q3.1
Q6.1
Q7.1
Q5.2
A3
Q1
Q2
Q5
Q3
Q5.1
Q2.2
Q1.2
Q2.1
Q7
Q4
A1
Q1.1
Média
A2

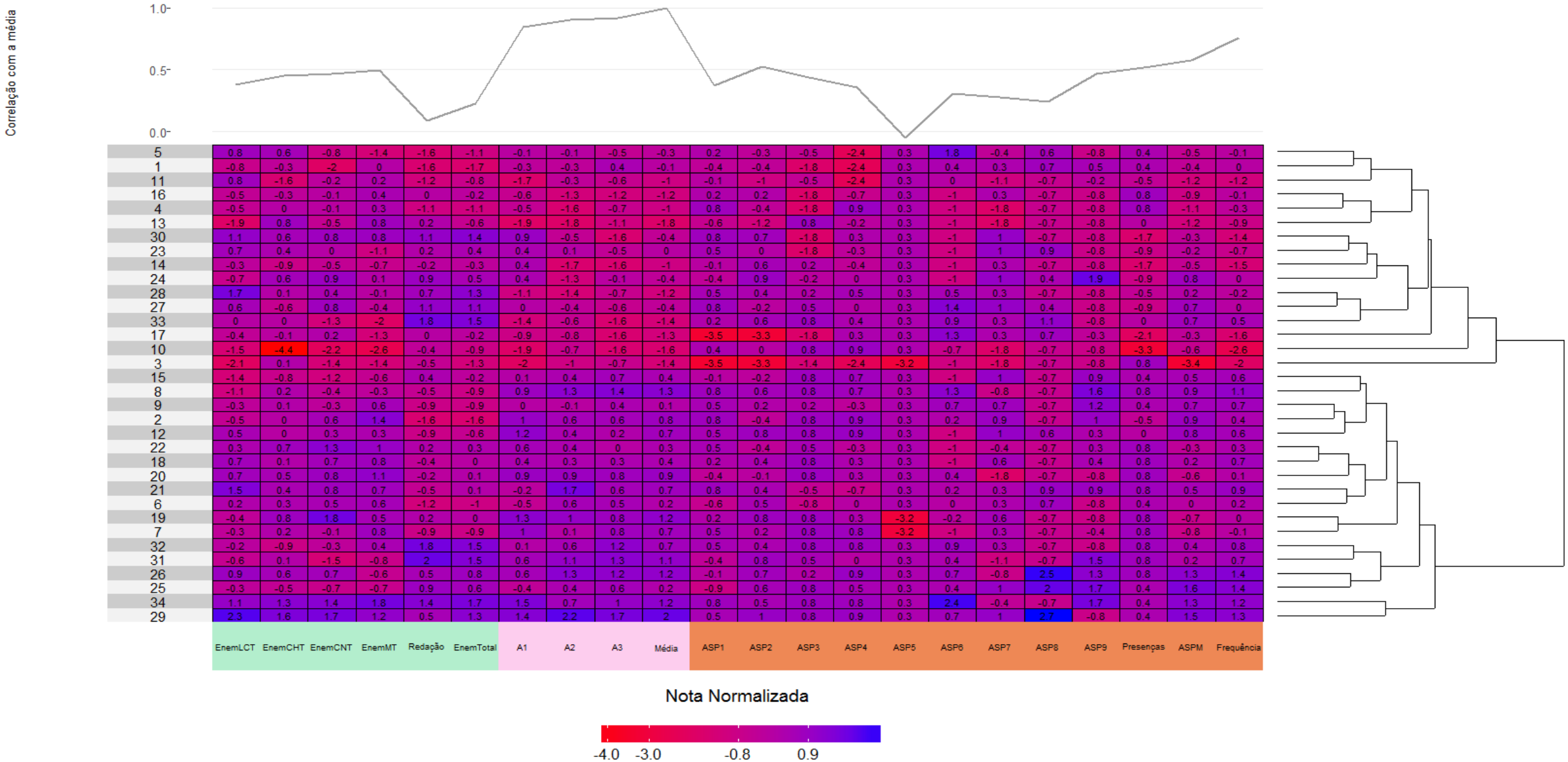
Mapa de Calor de desempenho de alunos de IAC

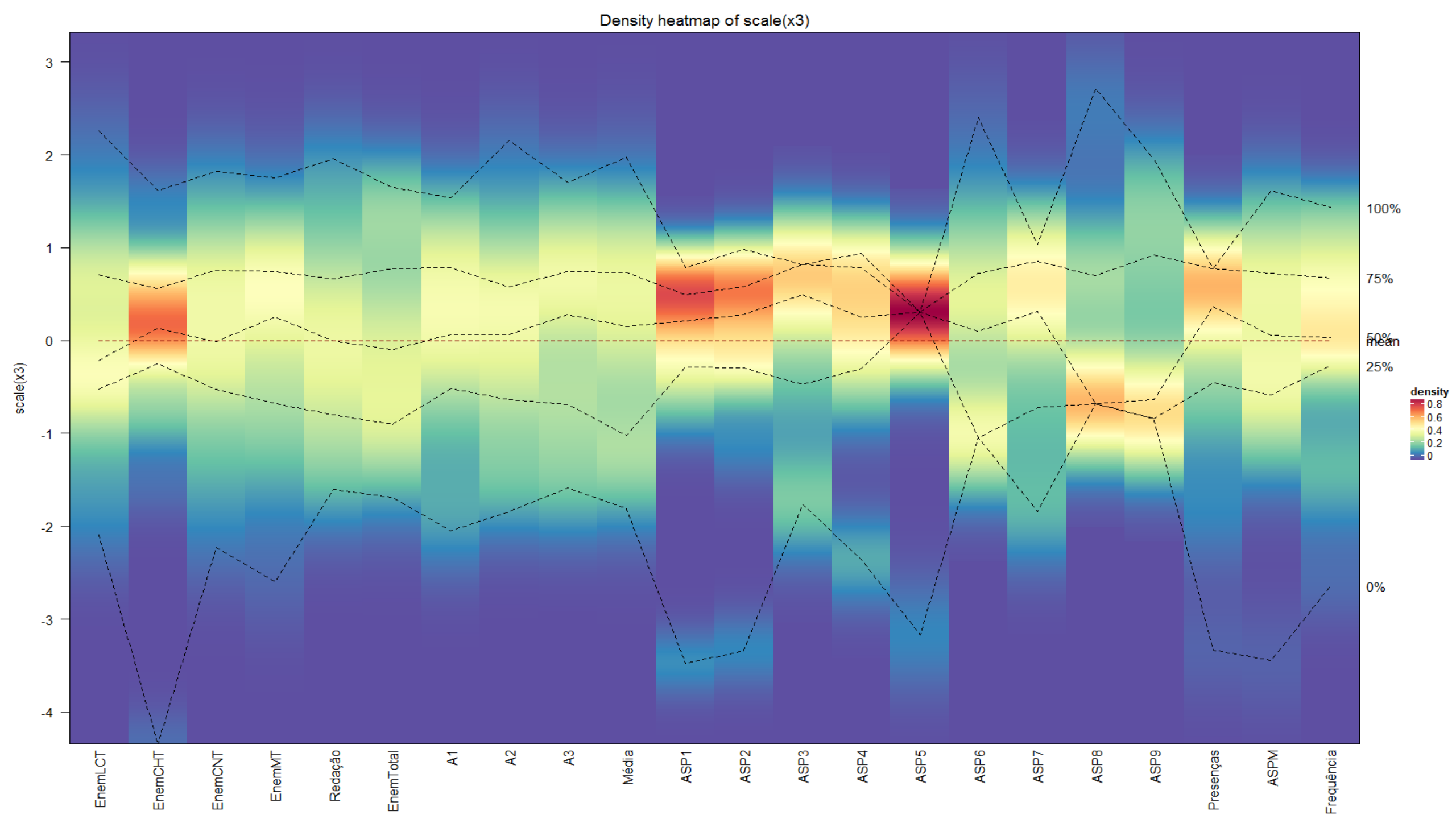


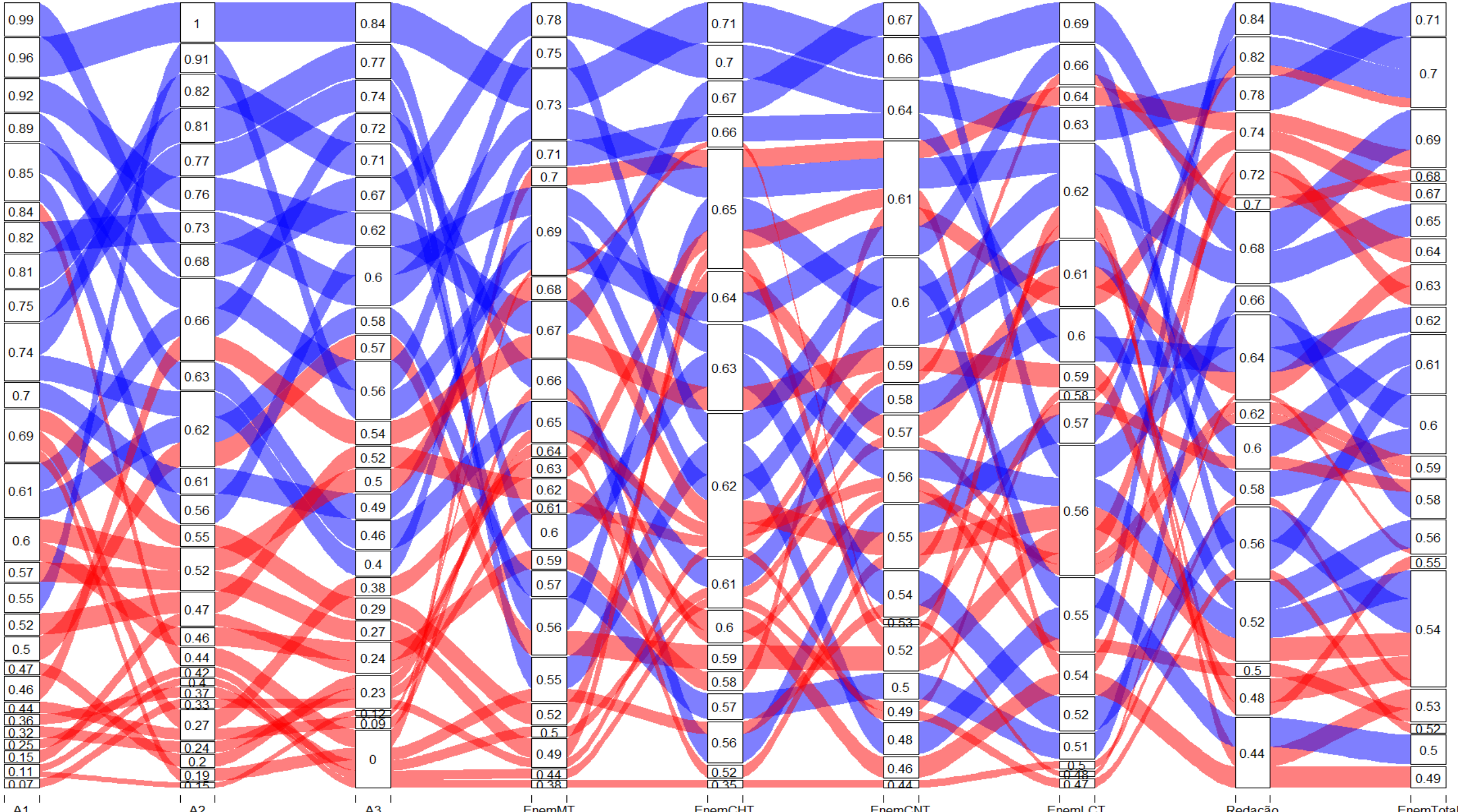
Mapa de Calor de desempenho de alunos de IAC

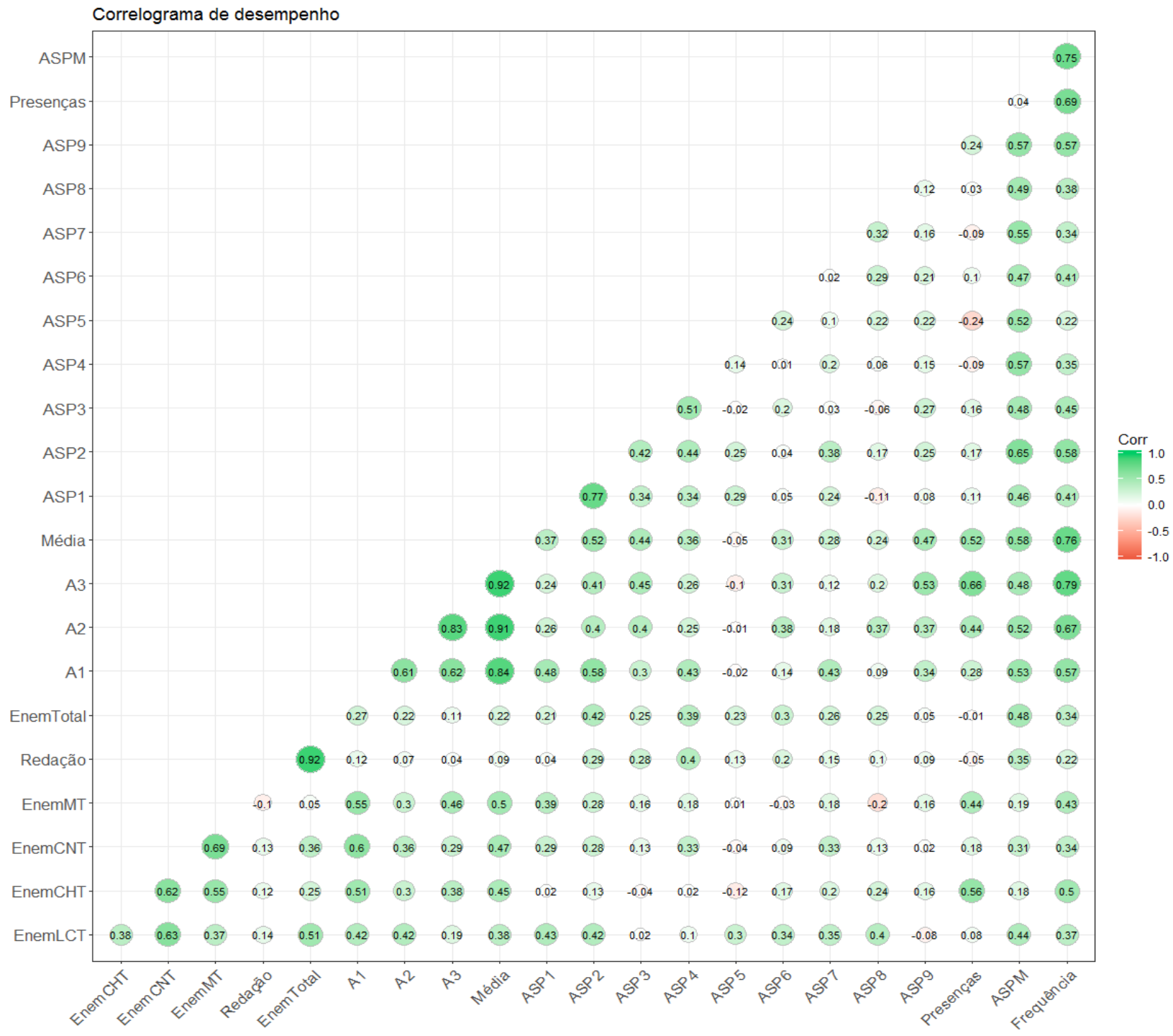


Mapa de Calor de desempenho de alunos de IAC




Density heatmap of scale(x3)
scale(x3)
3
2
1
0
-1
-2
-3
-4
100%
75%
50%
mean
25%
0%
density
0.8
0.6
0.4
0.2
0
EnemLCT
EnemCHT
EnemCNT
EnemMT
Redação
EnemTotal
A1
A2
A3
Média
ASP1
ASP2
ASP3
ASP4
ASP5
ASP6
ASP7
ASP8
ASP9
Presenças
ASPM
Frequência


A1
A2
A3
EnemMT
EnemCHT
EnemCNT
EnemLCT
Redação
EnemTotal


Correlograma de desempenho
ASPM
Presenças
ASP9
ASP8
ASP7
ASP6
ASP5
ASP4
ASP3
ASP2
ASP1
Média
A3
A2
A1
EnemTotal
Redação
EnemMT
EnemCNT
EnemCHT
EnemLCT
EnemCHT
EnemCNT
EnemMT
Redação
EnemTotal
A1
A2
A3
Média
ASP1
ASP2
ASP3
ASP4
ASP5
ASP6
ASP7
ASP8
ASP9
Presenças
ASPM
Frequência
Corr
1.0
0.5
0.0
-0.5
-1.0

Conclusões

- Técnicas de visualização podem facilitar o processo analítico.
- Gráficos podem possuir grandes quantidades de informação e ainda assim podem possuir uma excelente interpretabilidade.
- Foi possível gerar gráficos expressivos e efetivos nesta análise.
- Verificou-se a importância das notas do ENEM dos estudantes no desempenho final, apesar de não superar a importância das notas nas avaliações da disciplina de IAC.

Referências Bibliográficas

LARSON, R.; FARBER, B. Elementary Statistics: Picturing the World, Prentice-Hall, Englewood Cliffs, NJ, 2003.

HAN, J.; KAMBER, M. Data Mining: Concepts and Techniques. 2° ed. Morgan Kaufmann Publishers, p. 5–7, 2006.

Apêndice

```
install.packages("rpart")

install.packages("caret")

install.packages("rpart.plot")

install.packages("cvTools")

install.packages("rattle")

install.packages("RColorBrewer")

install.packages("party")

install.packages("partykit")

install.packages("C50")

install.packages("e1071")

install.packages("plotly")

install.packages("ComplexHeatmap")

install.packages("circlize")
```

Apêndice

```
install.packages("gplots")

install.packages("devtools")

install.packages("ggcorrplot")

install.packages("cdparcoord")

install.packages("andrews")

install.packages("igraph")

install.packages("network")

install.packages("sna")

install.packages("visNetwork")

install.packages("threejs")

install.packages("networkD3")

install.packages("ndtv")

install.packages("treemap")
```

Apêndice

```
install.packages("alluvial")

library(alluvial)

library(ggcorrplot)

library(MASS)

library(cdparcoord)

library("andrews")

library(treemap)

library("igraph")

library("network")

library("sna")

library("visNetwork")

library("threejs")

library("networkD3")

library("ndtv")
```

Apêndice

```
library(e1071)

library(rpart)

library(caret)

library(rpart.plot)

library(cvTools)

library(rattle)

library(RColorBrewer)

library(party)

library(partykit)

library(C50)

library(plotly)

library(circlize)

library(superheat)

library(gplots)
```

Apêndice

```
library(devtools)

install_github("jokergoo/ComplexHeatmap")

devtools::install_github("rlbarter/superheat")

library("ComplexHeatmap")

aaa<-read.csv('Dados_Mineração - 2017.1-Enem.csv',sep=";",stringsAsFactors=FALSE)

for(i in 1:nrow(aaa)){

 for(j in 1:ncol(aaa)){

   aaa[i,j] = as.numeric(gsub("[,]",".",aaa[i,j]))

  }}

for(j in 1:ncol(aaa)){

    aaa[,j] = as.numeric(aaa[,j])

}

aaa<-aaa[,-1]

doit <- function(x) {(x - min(x, na.rm=TRUE))/(max(x,na.rm=TRUE) -min(x, na.rm=TRUE))}
```

Apêndice

```
aaa <- as.data.frame(lapply(aaa, doit))

aaa<-aaa[-35,] ##Apagar amostras extras

aaa<-aaa[-35,] ##Apagar amostras extras

x1<-data.matrix(x1)

plot_ly(z = x1, type = "heatmap")

heatmap.2(x1)

heatmap(x1,Colv=F, Rowv=F, scale='none')

heatmap(x1,Colv=NA, Rowv=NA, scale='none')

x2<-aaa[,(7:29)]

x2<-data.matrix(x2)

heatmap.2(x2)

#x1.size <- scale(x1)

superheat(x2,scale = F,pretty.order.rows = FALSE,pretty.order.cols = FALSE,yr.axis.name = "Desempenho normalizado",yt = cor(x2)[, "Média"],yt.plot.type = "line",yt.axis.name = "Correlação com a média",title = "Mapa de Calor de desempenho de alunos de IAC",column.title = "Nota Normalizada",left.label.size = 0.45)
```

Apêndice

```
superheat(x2,scale = F,pretty.order.rows = FALSE,pretty.order.cols = FALSE,yr.axis.name = "Desempenho normalizado",yt = cor(x2)[, "Média"],yt.plot.type = "line",yt.axis.name = "Correlação com a média",title = "Mapa de Calor de desempenho de alunos de IAC",column.title = "Nota Normalizada",left.label.size = 0.20, X.text = round(as.matrix(x2), 1),left.label.text.size = 4,heat.pal = c("#dd3333","#bb5858", "white", "#58bb58"),heat.pal.values = c(0, 0.2, 0.6, 1),bottom.label.col =c("#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee4433"),X.text.size = x2.size)

superheat(x2,scale = F,pretty.order.rows = FALSE,pretty.order.cols = FALSE,yr.axis.name = "Desempenho normalizado",yt = cor(x2)[, "Média"],yt.plot.type = "line",yt.axis.name = "Correlação com a média",title = "Mapa de Calor de desempenho de alunos de IAC",column.title = "Nota Normalizada",left.label.size = 0.20, X.text = round(as.matrix(x2), 1),left.label.text.size = 4,heat.pal = c("#dd3333","#bb5858", "white", "#58bb58"),heat.pal.values = c(0, 0.2, 0.6, 1),bottom.label.col =c("#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee4433"),X.text.size = x2.size, row.dendrogram = TRUE)

x3<-aaa

x3<-x3[,-(7:13)]

x3<-x3[,-(8:14)]

x3<-x3[,-(9:13)]

x3<-data.matrix(x3)

#x3.size <- scale(x3)

superheat(x3,scale = F,pretty.order.rows = FALSE,pretty.order.cols = FALSE,yr.axis.name = "Desempenho normalizado",yt = cor(x3)[, "Média"],yt.plot.type = "line",yt.axis.name = "Correlação com a média",title = "Mapa de Calor de desempenho de alunos de IAC",column.title = "Nota Normalizada",left.label.size = 0.20, X.text = round(as.matrix(x3), 1),left.label.text.size = 4,heat.pal = c("#dd3333","#bb5858", "white", "#58bb58"),heat.pal.values = c(0, 0.2, 0.6, 1),bottom.label.col =c("#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855"),row.dendrogram = TRUE,X.text.size = 3,bottom.label.text.size = 4)
```

Apêndice

```
superheat(x3,scale = T,pretty.order.rows = FALSE,pretty.order.cols = FALSE,yr.axis.name = "Desempenho normalizado",yt = cor(x3)[, "Média"],yt.plot.type = "line",yt.axis.name = "Correlação com a média",title = "Mapa de Calor de desempenho de alunos de IAC",column.title = "Nota Normalizada",left.label.size = 0.20, X.text = round(scale(x3), 1),left.label.text.size = 4,heat.pal = c("black", "red", "blue"),heat.pal.values = c(-1, 0, 1),bottom.label.col
=c("#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#b3eecd","#fdcdee","#fdcdee","#fdcdee","#fdcdee","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855","#ee8855"),row.dendrogram = TRUE,X.text.size = 3,bottom.label.text.size = 2.5)

densityHeatmap(scale(x3))

aaa1<-aaa

x3$Média<-cut(as.numeric(x3$Média), c(-1,0.599,1.0), labels=c("REP","AP"))

x3[] <- lapply(x3, function(x) type.convert(as.character(x)))

tit2d <- aggregate( Média ~ A1 + A2 + A3 + EnemMT + EnemCHT + EnemCNT + EnemLCT + Redação + EnemTotal, data=x3, sum)

alluvial(tit2d[,1:9], freq= tit2d$Média, border=NA, col=ifelse( tit2d$Média < 0.6, "red", "blue") )

ggcorrplot(round(cor(x3),2), type = "lower",  lab = TRUE,  lab_size = 2,  method="circle",  colors = c("tomato2", "white", "springgreen3"),  title="Correlograma de desempenho",  ggtheme=theme_bw)
```