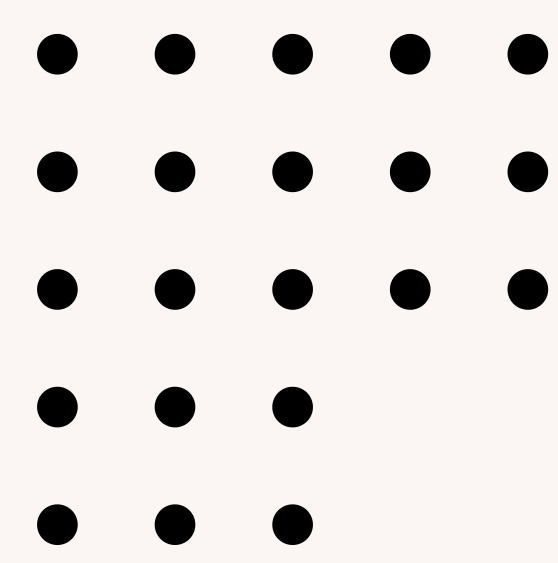
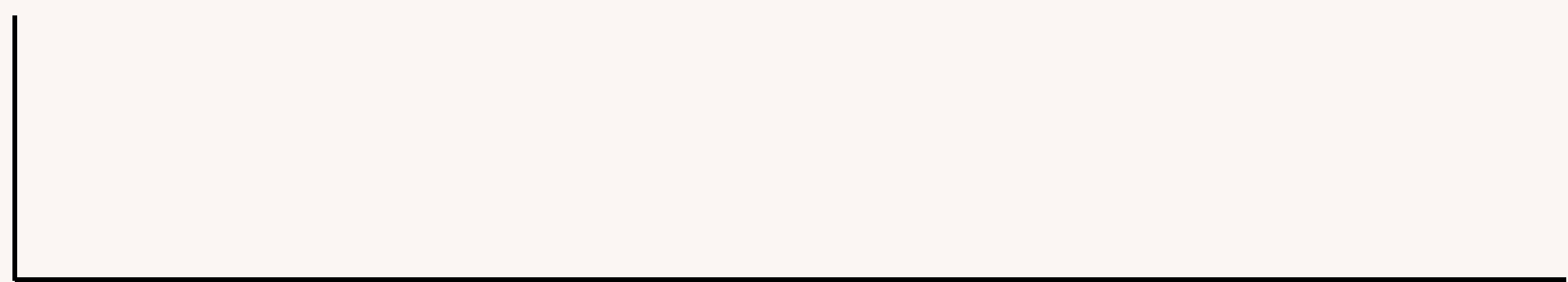


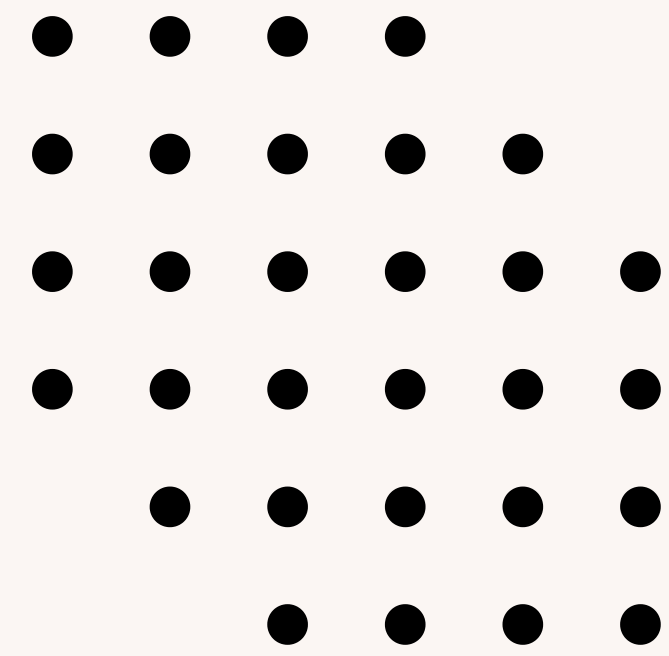
DISTÚRBIOS DE SONO

Análise e Modelagem dos Dados



TÓPICOS DE ABORDAGEM

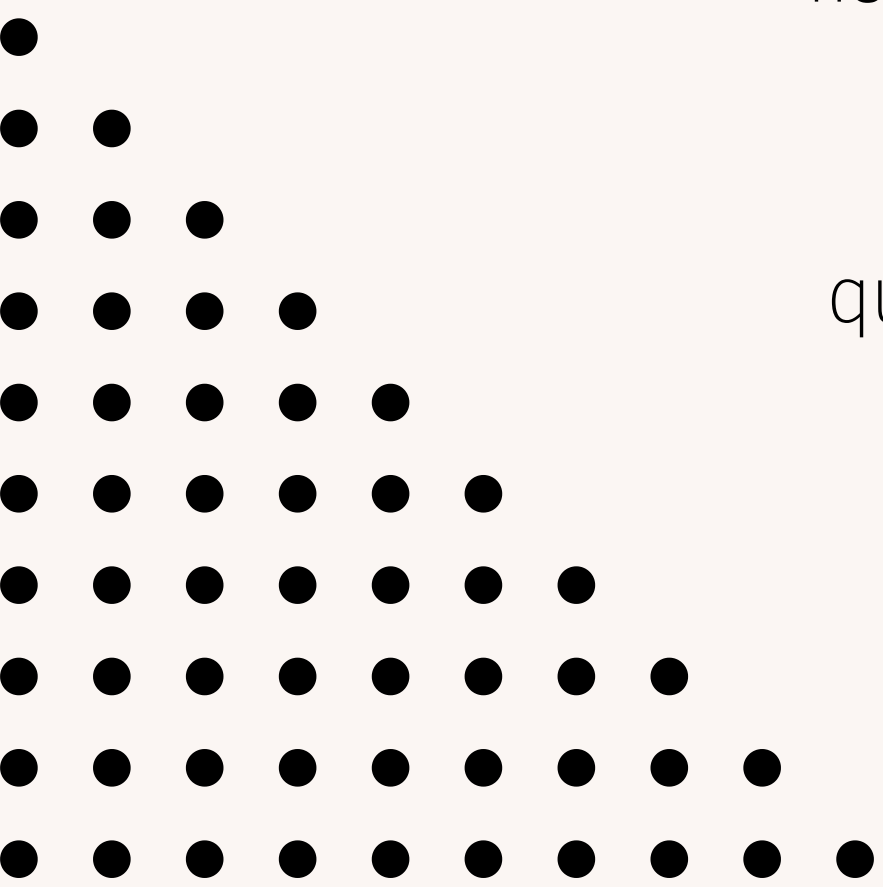
- Introdução
- Dados iniciais
- Modelos trabalhados
- Conclusões





INTRODUÇÃO

Os distúrbios do sono são problemas que interferem na qualidade do sono. Eles podem ser causados por uma variedade de fatores, incluindo estresse, ansiedade, depressão, doenças físicas, uso de medicamentos e hábitos de vida inadequados. Os distúrbios do sono podem ter um impacto significativo na qualidade de vida, afetando o trabalho, os relacionamentos e a saúde geral.



DATASET ESCOLHIDO

Sleep Health and Lifestyle Dataset

Unlock sleep insights with the Sleep Health Dataset



Data Card Code (64) Discussion (5)

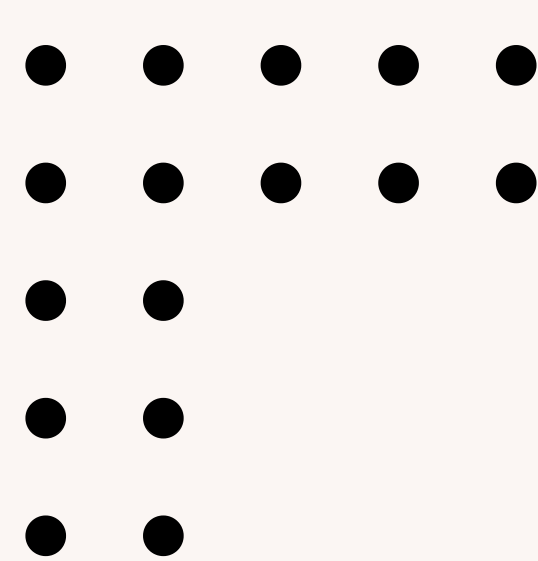
About Dataset

Usability ⓘ
10.00

DATASET ESCOLHIDO

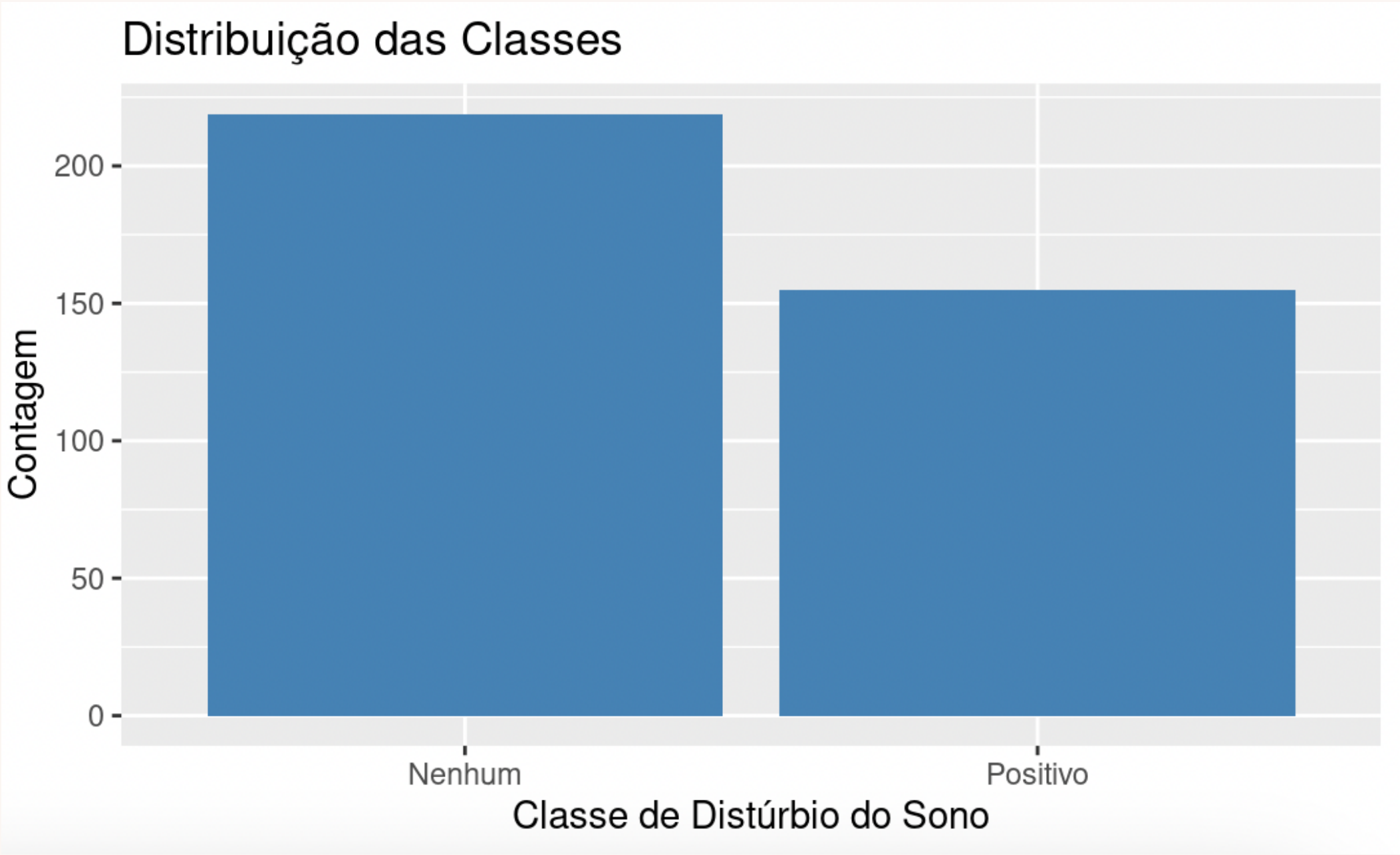
Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep
1	Male	27	Software Engine	6.1	6
2	Male	28	Doctor	6.2	6
3	Male	28	Doctor	6.2	6
4	Male	28	Sales Represent	5.9	4
5	Male	28	Sales Represent	5.9	4
6	Male	28	Software Engine	5.9	4

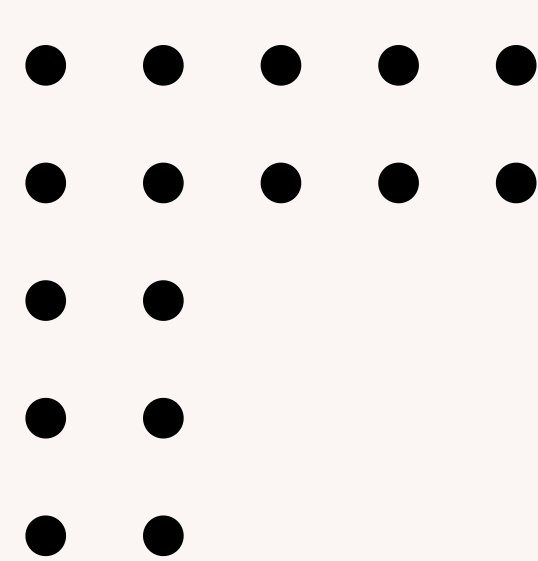
Physical Activity	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disor
42	6	Overweight	126/83	77	4200	None
60	8	Normal	125/80	75	10000	None
60	8	Normal	125/80	75	10000	None
30	8	Obese	140/90	85	3000	Sleep Apnea
30	8	Obese	140/90	85	3000	Sleep Apnea
30	8	Obese	140/90	85	3000	Insomnia



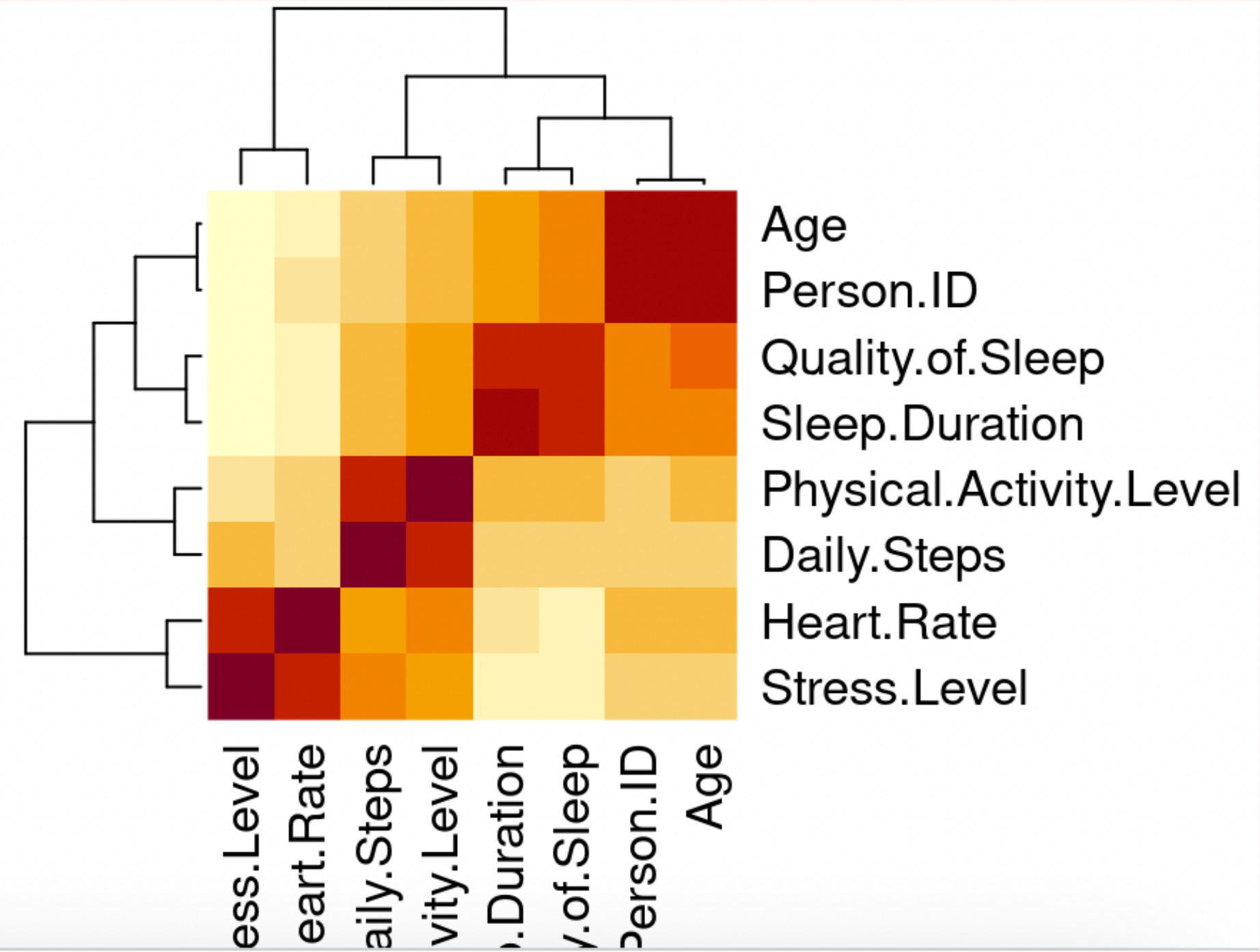
DISTRIBUIÇÃO DOS DADOS

Após reunirmos os dois distúrbios do sono distintos em uma única variável ("Positive"), esta é a distribuição da quantidade de cada classe:

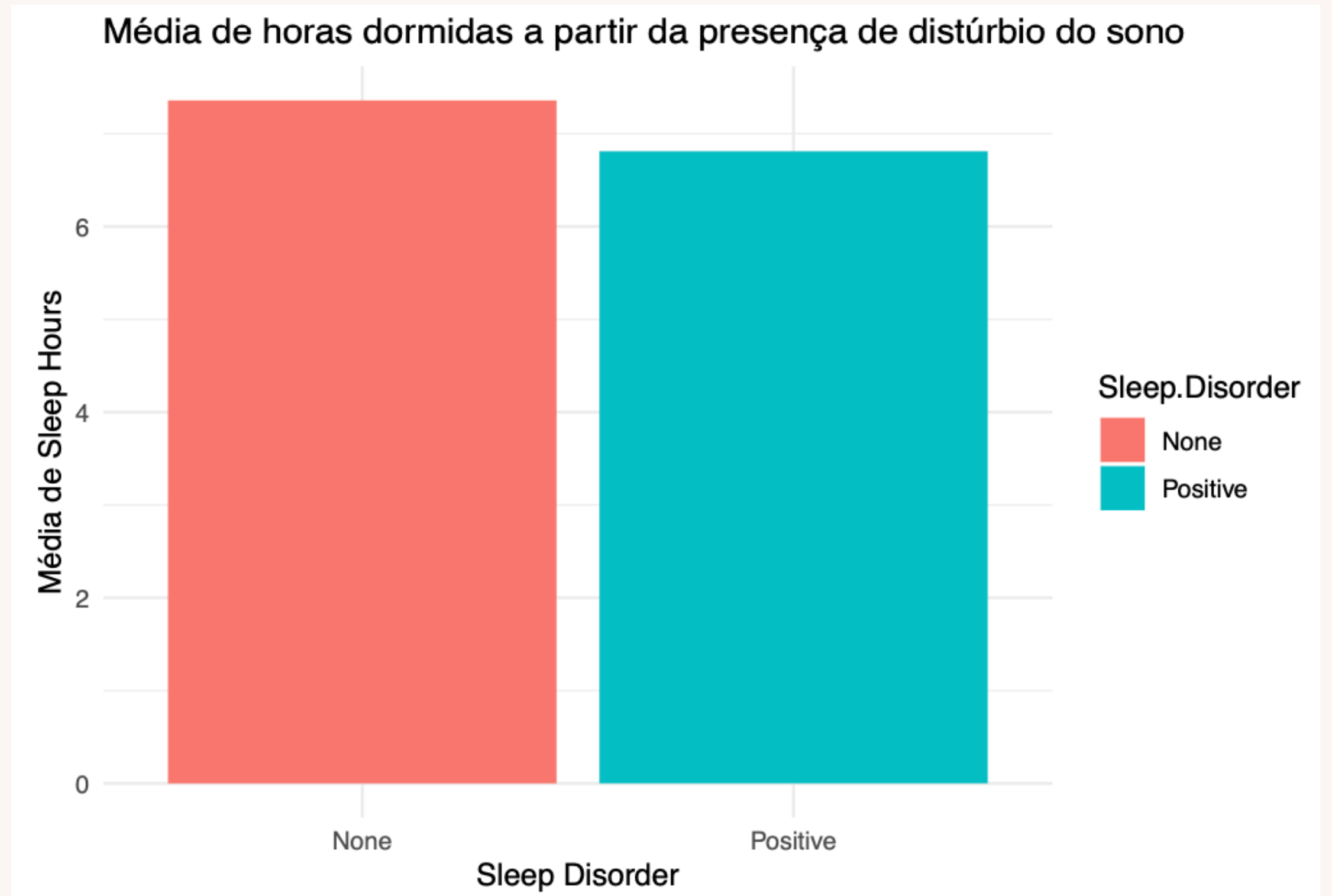


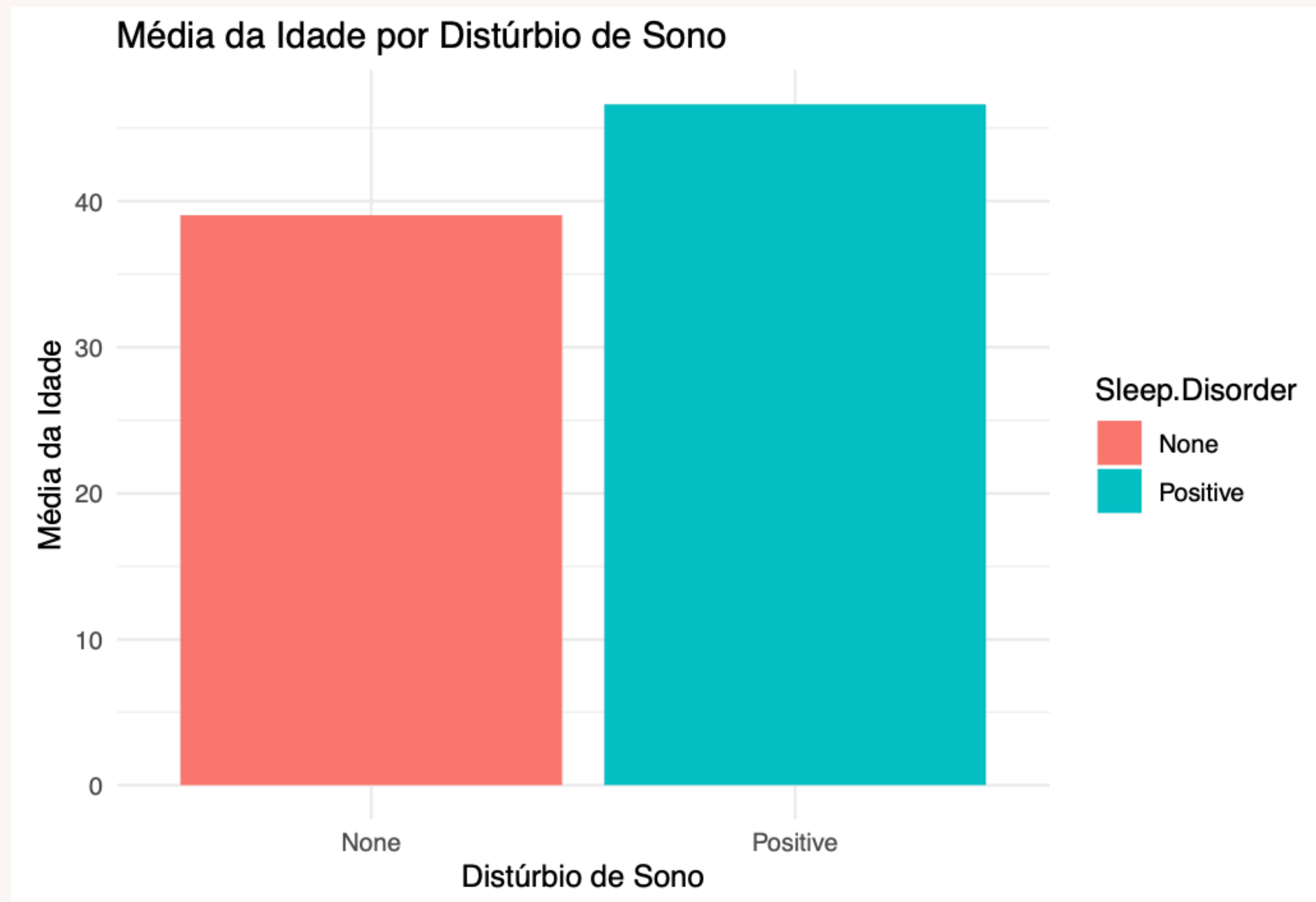
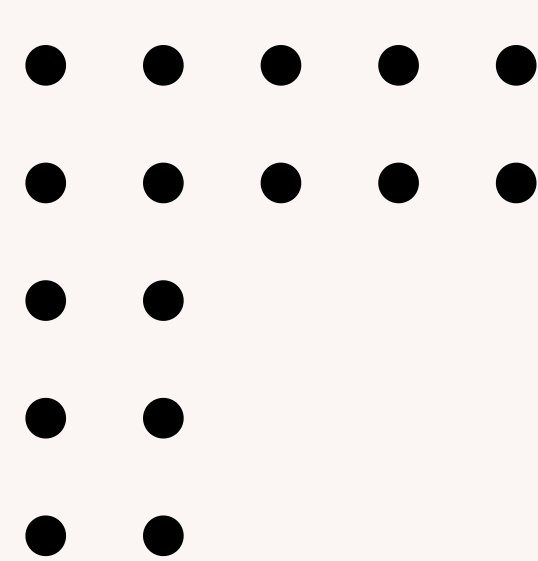


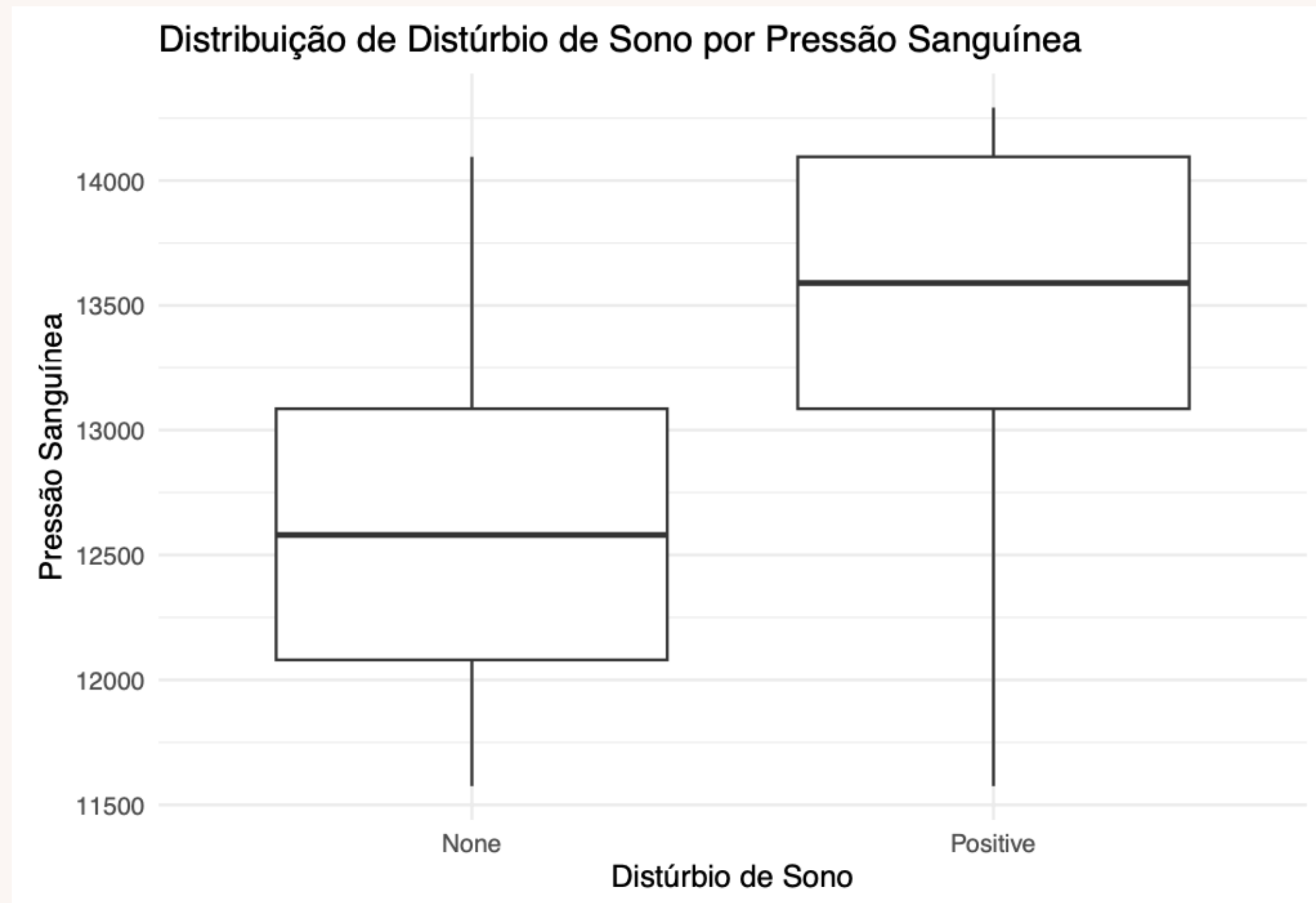
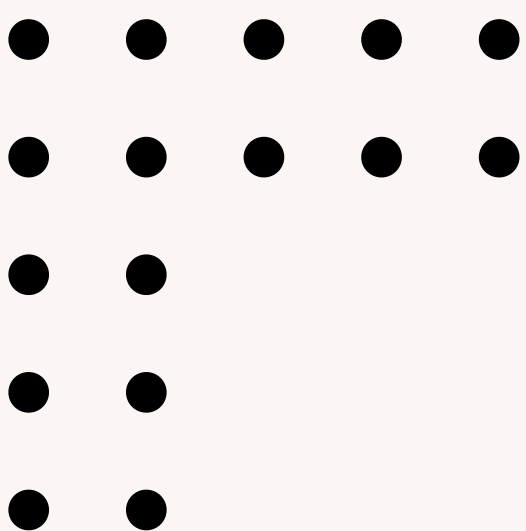
CORRELAÇÃO ENTRE VARIÁVEIS

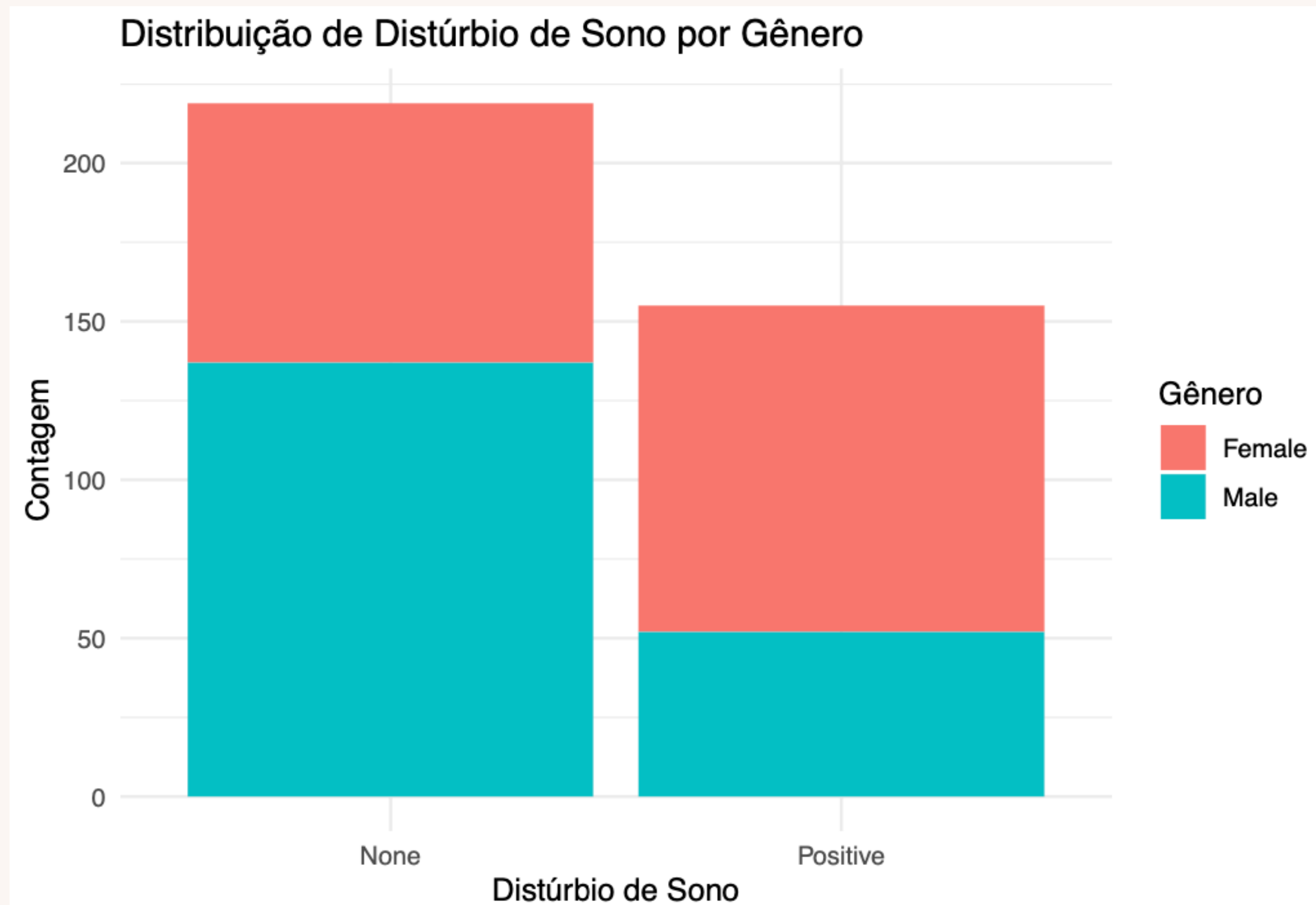
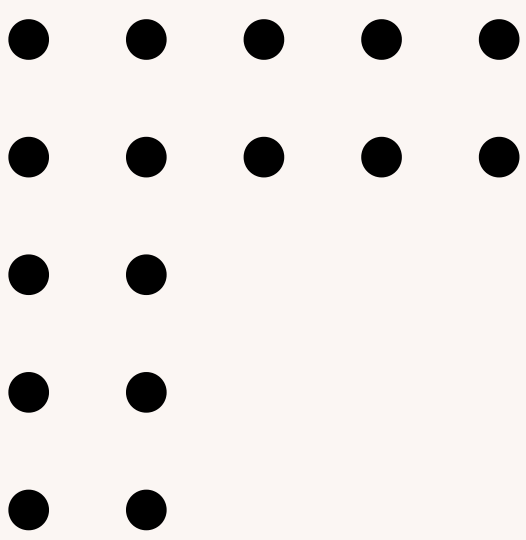


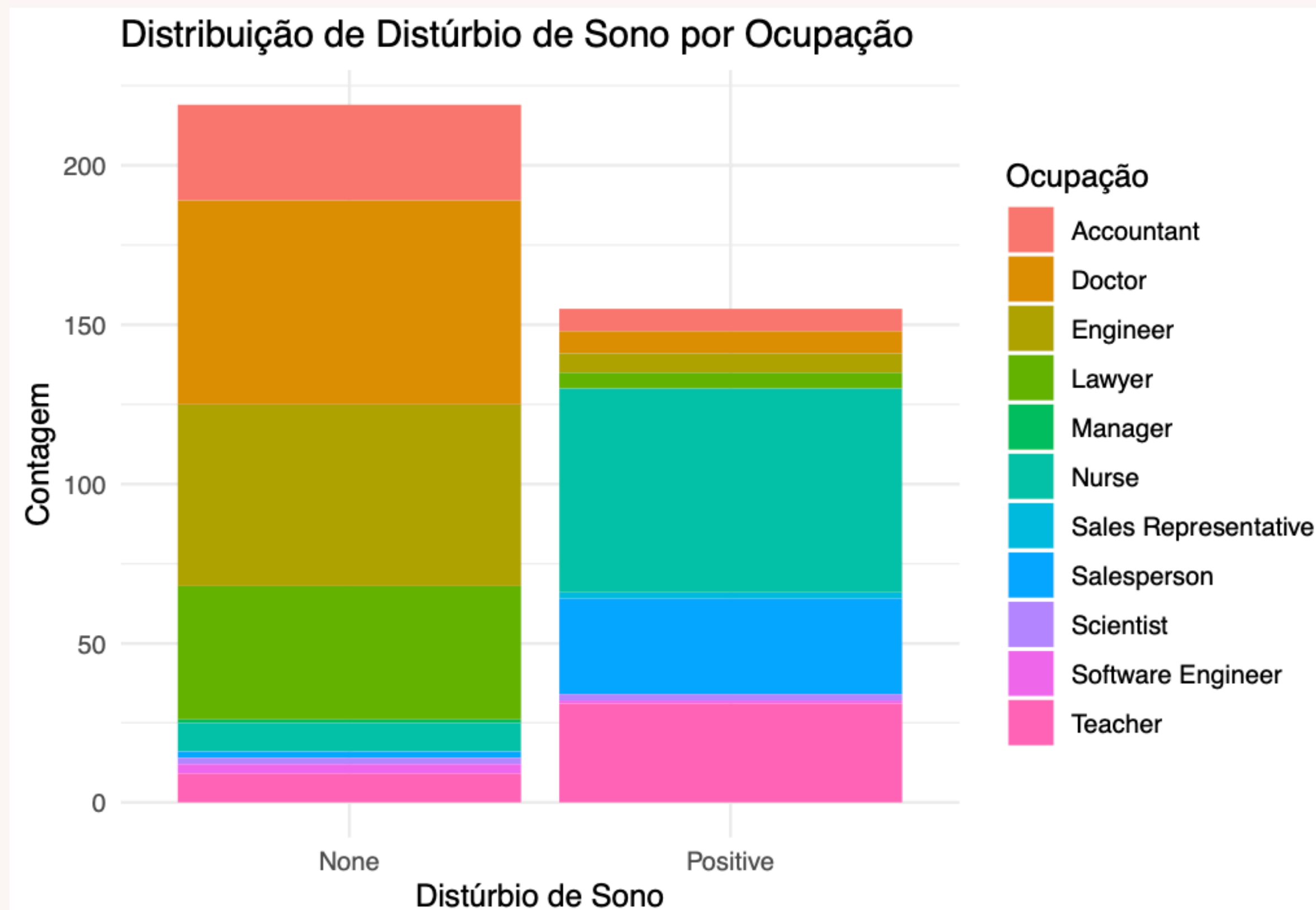
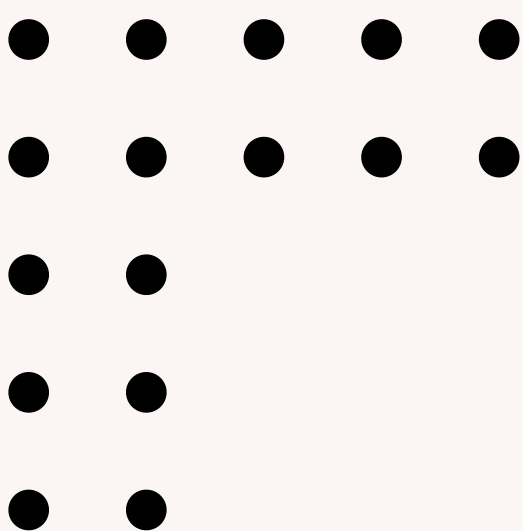
ANÁLISES INICIAIS

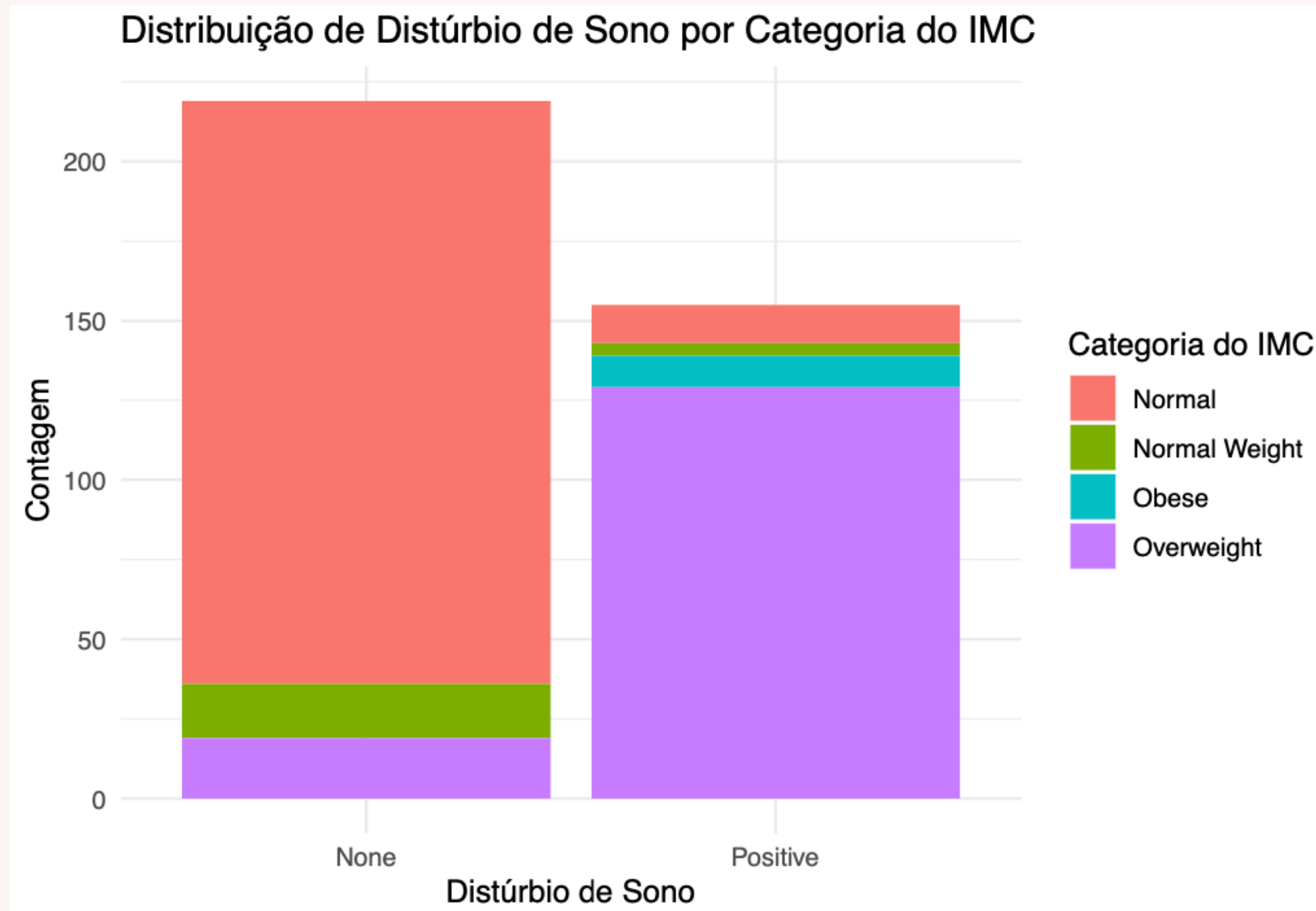
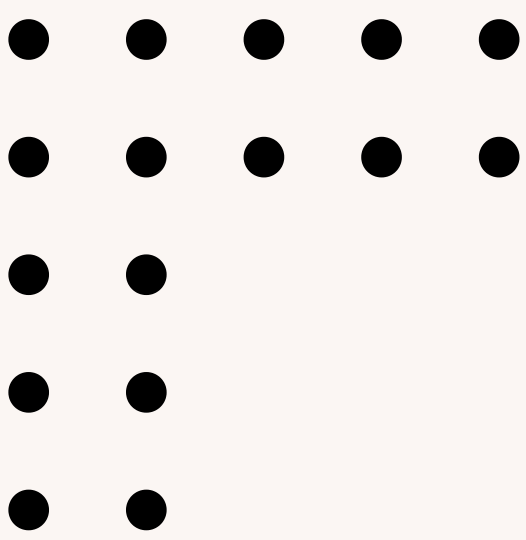


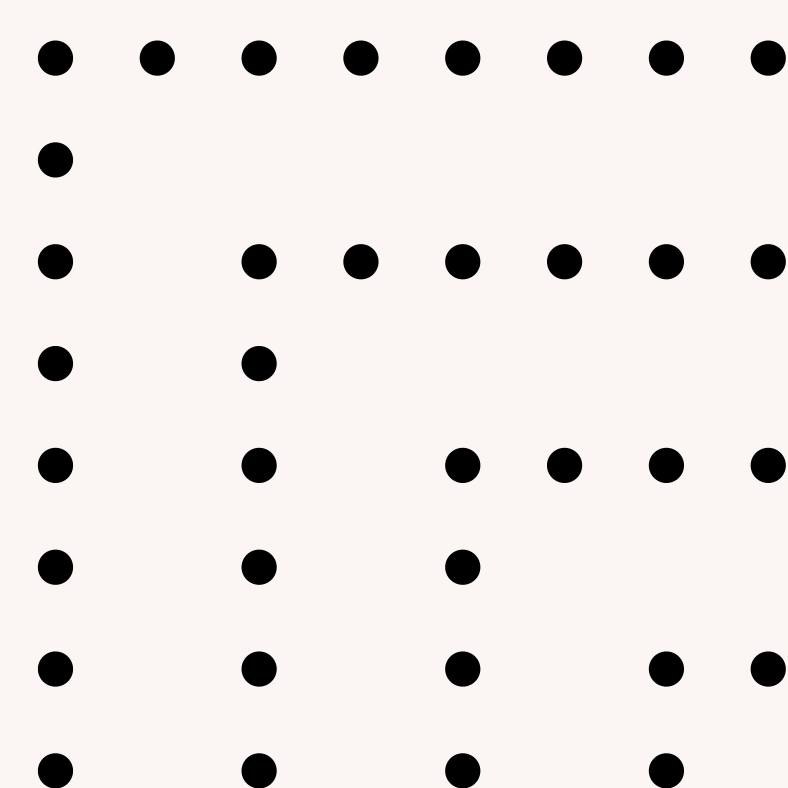






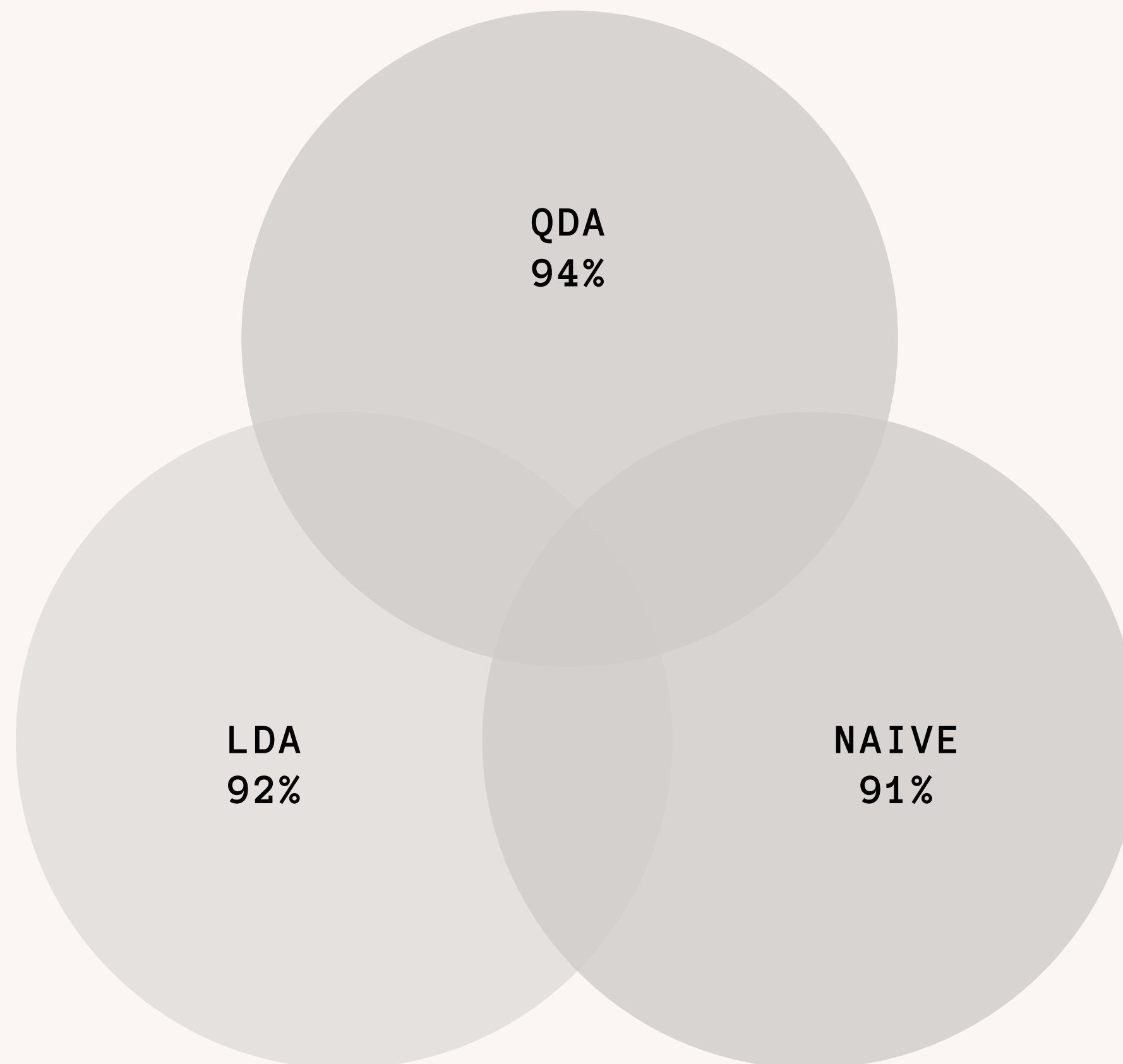


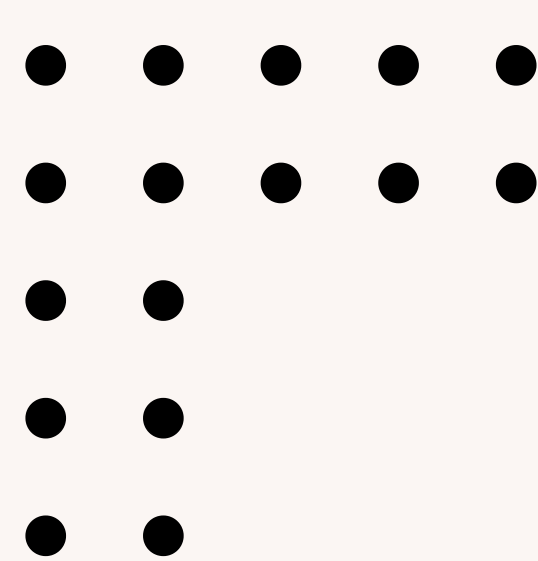




LDA, QDA E NAIVE BAYES

Acurácias de cada modelo aplicado:





REGRESSÃO LOGÍSTICA

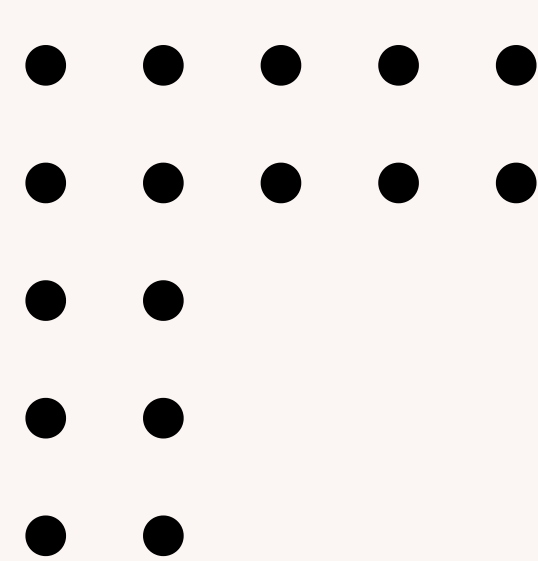
Recall: 87,2%

Precisão: 97,6%

Especificidade: 98,6%

Prediction	Truth	
	None	Positive
None	60	6
Positive	1	41

Matriz de confusão



PERCEPTRON

Acurácia: 45%

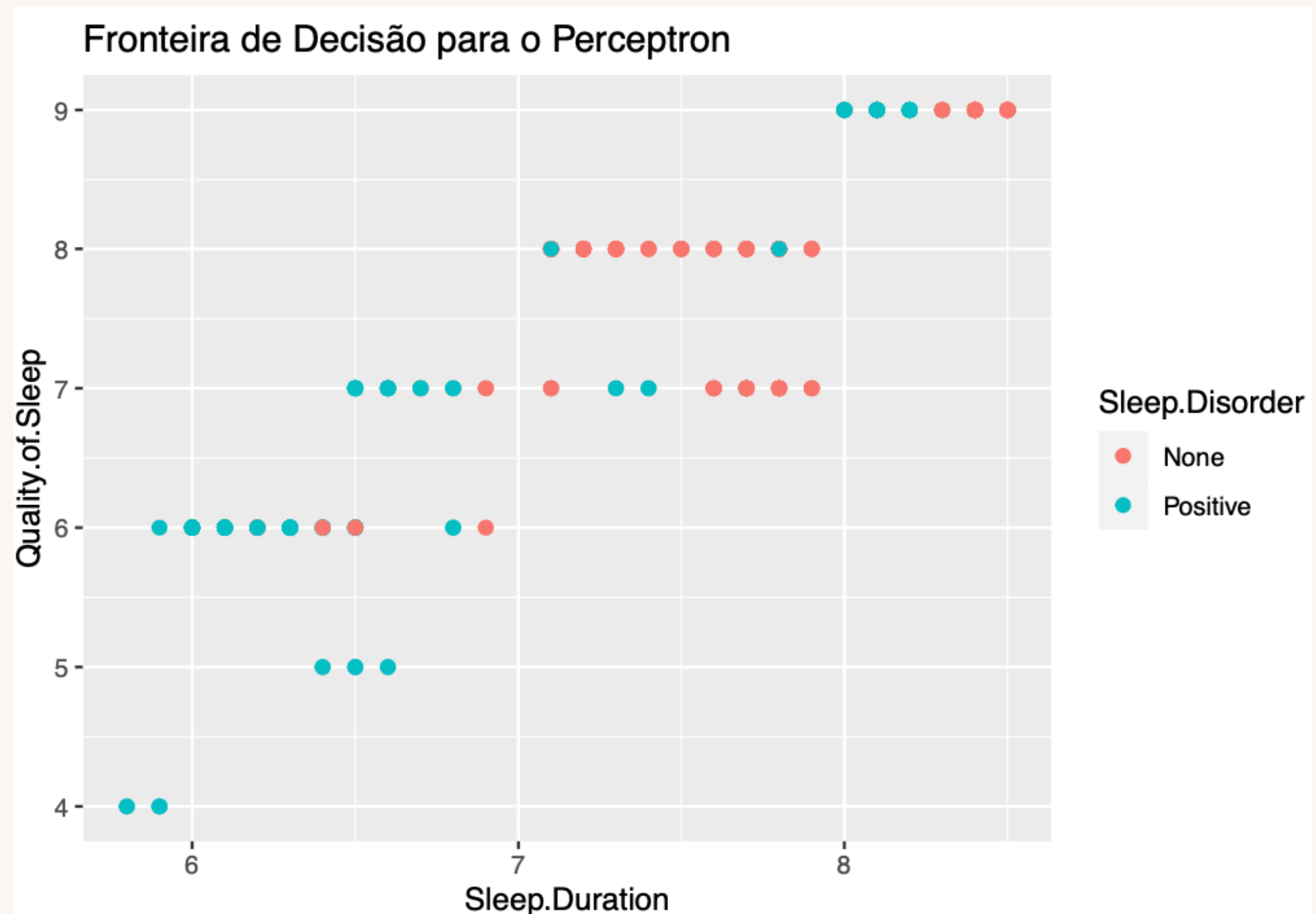
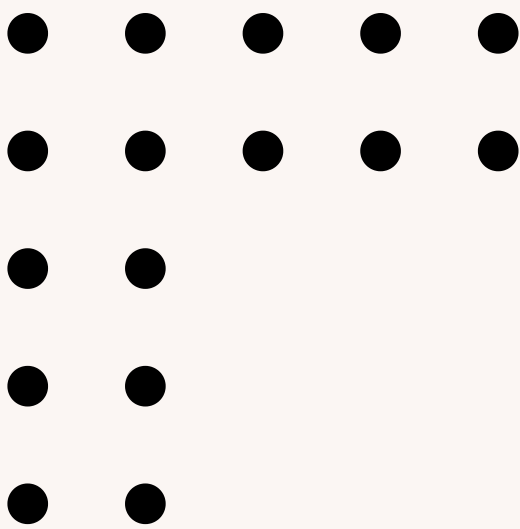


Gráfico perceptron



SVM

Reference		
Prediction	None	Positive
None	44	22
Positive	20	22

Matriz de confusão

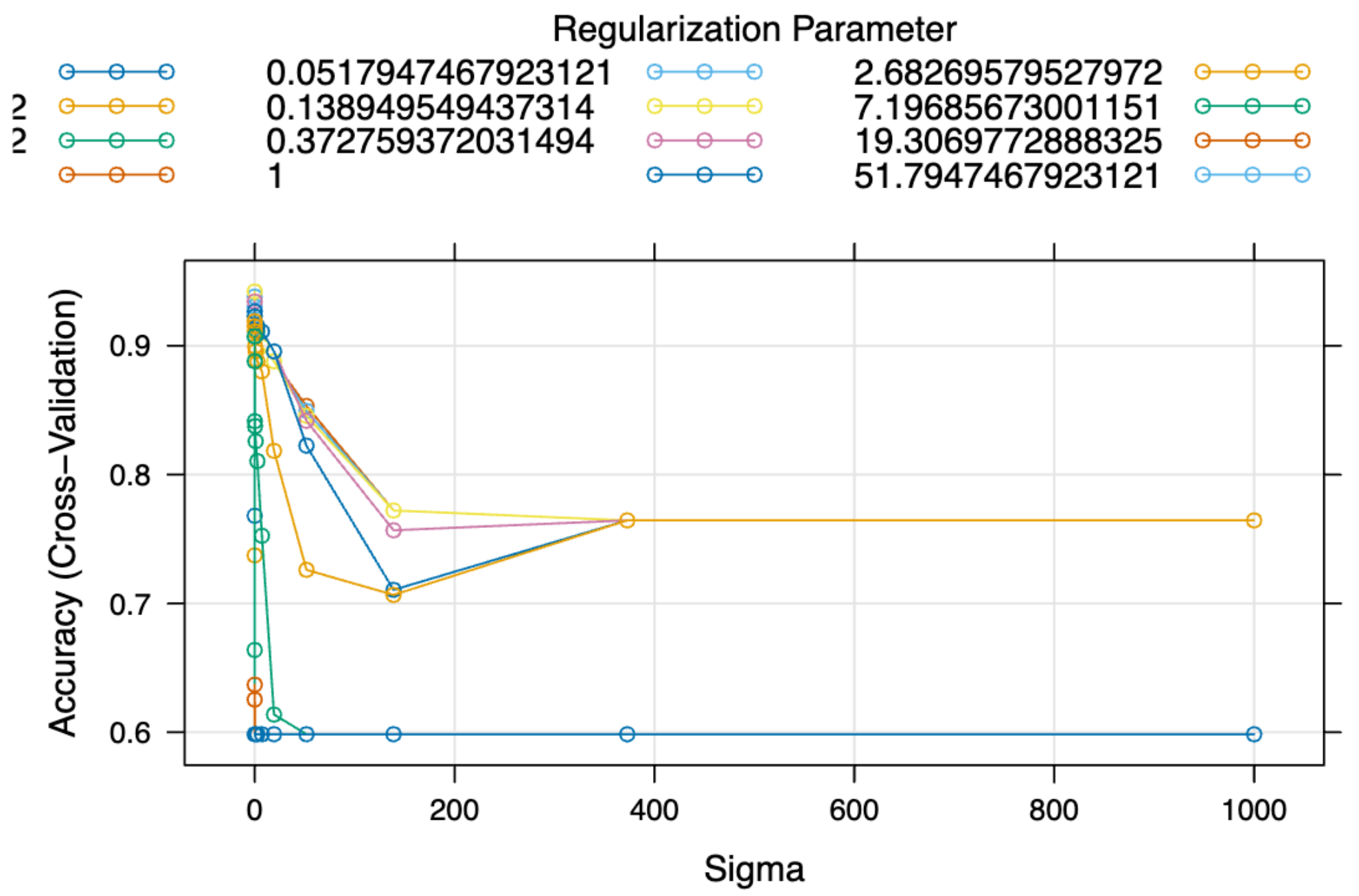
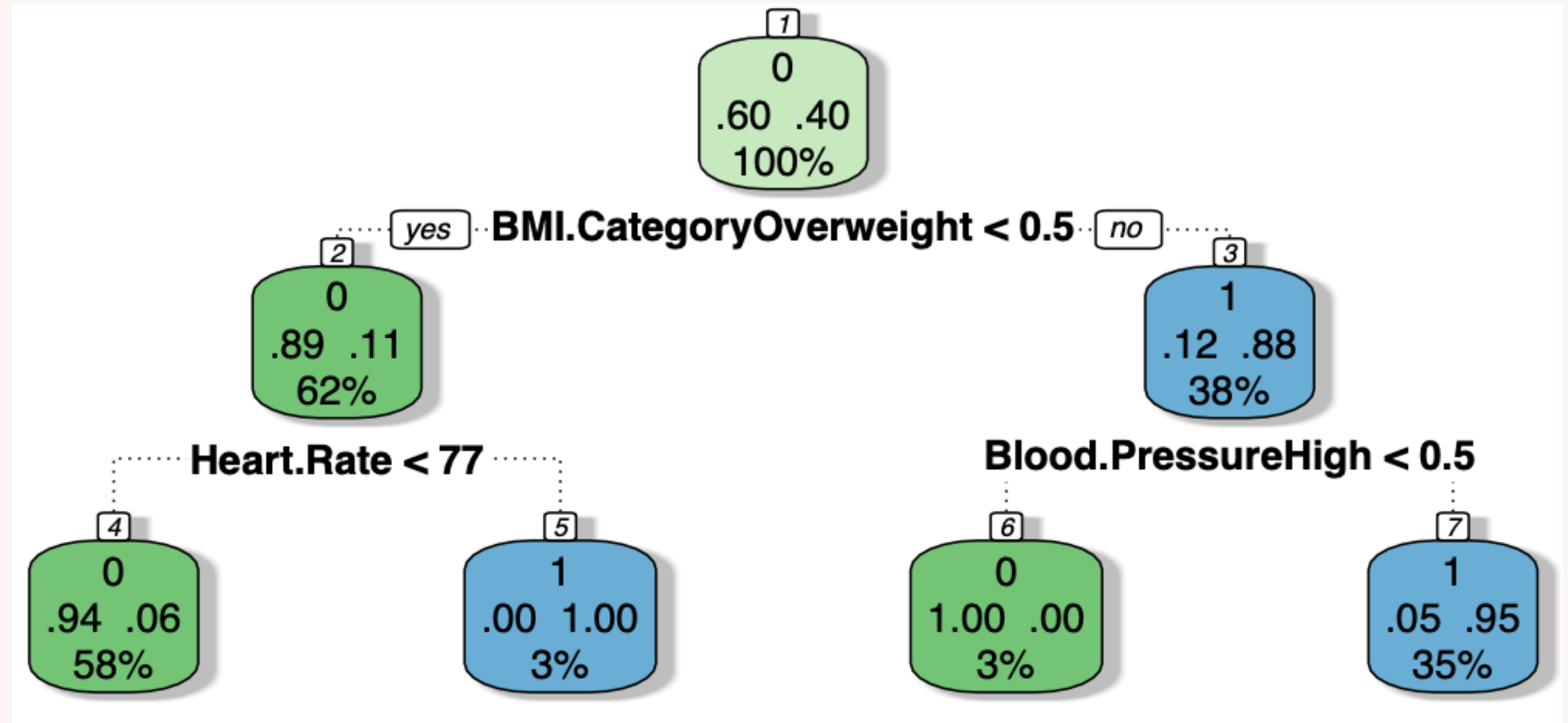
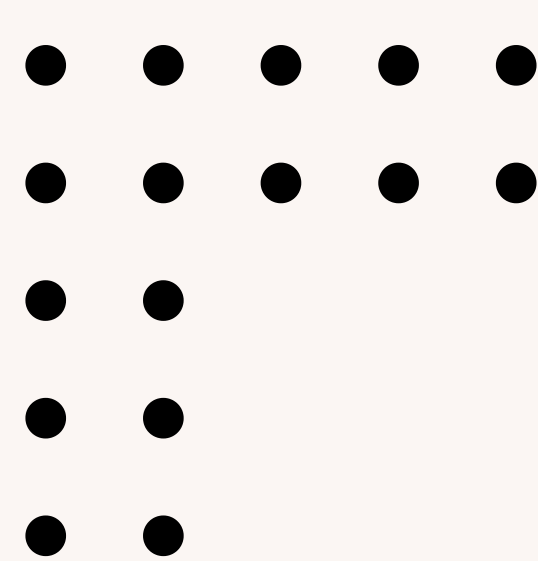


Gráfico SVM

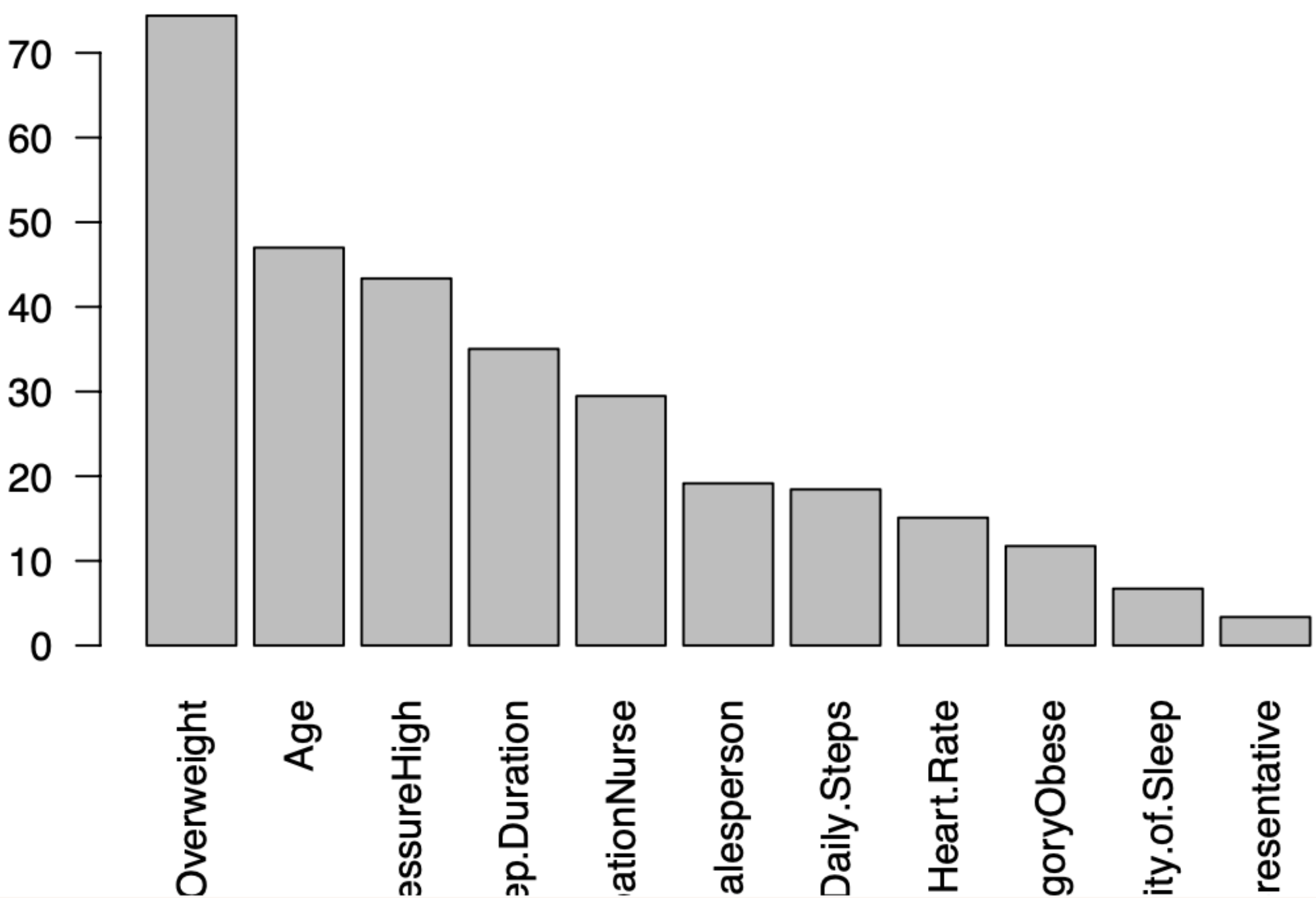
ÁRVORE DE DECISÃO

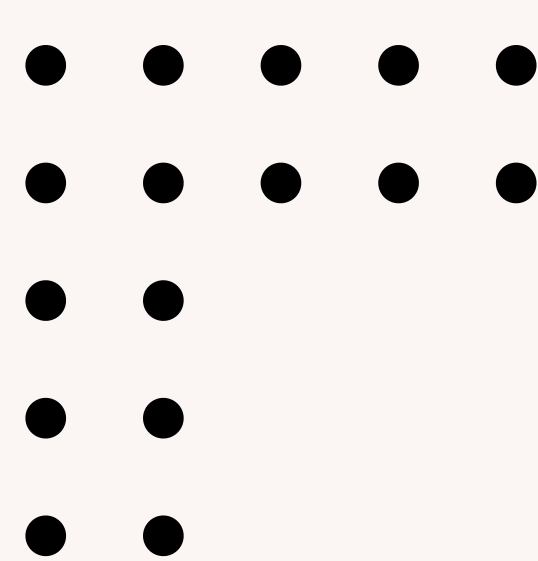
Acurácia: 91,74%



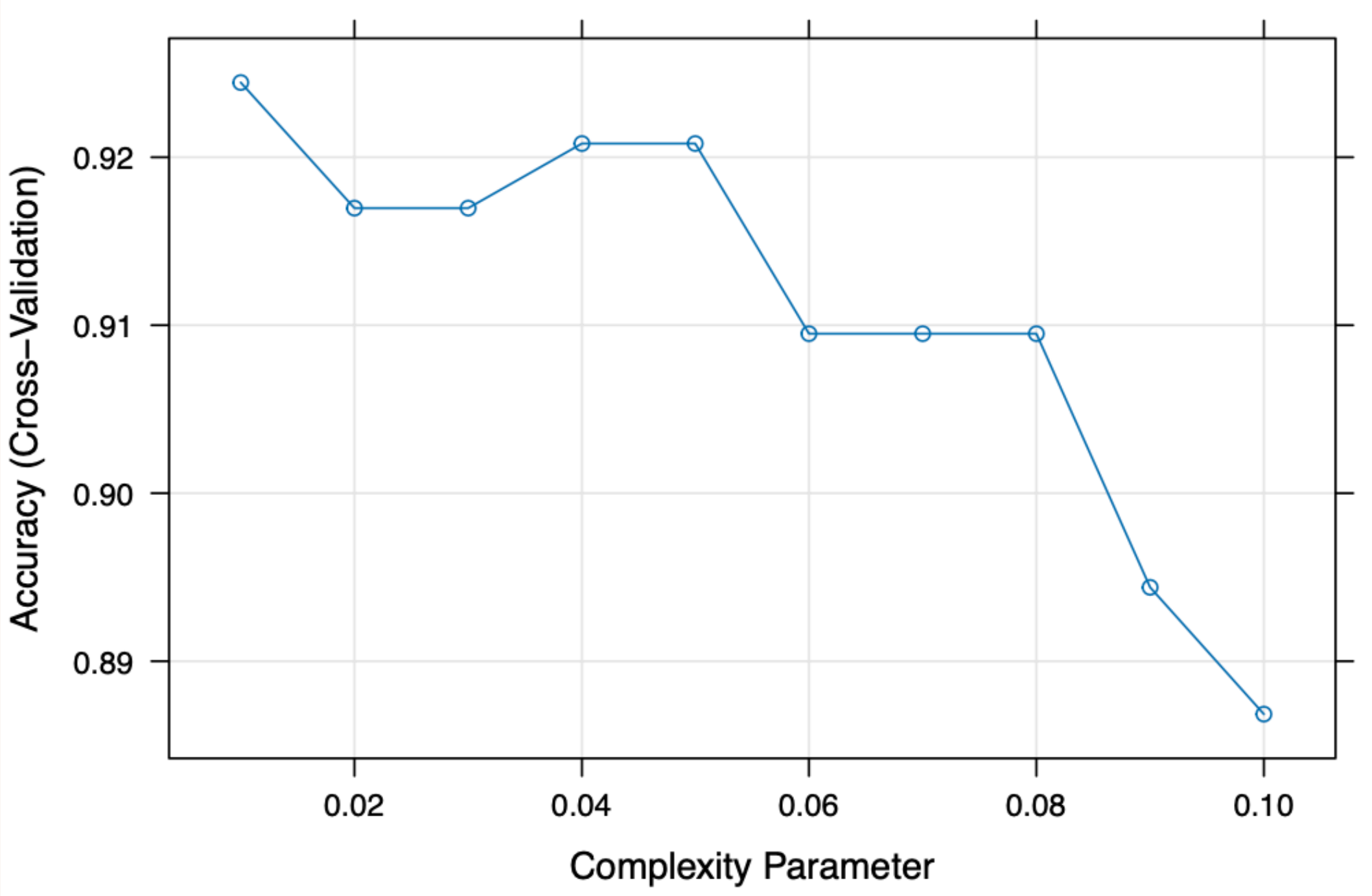


VARIÁVEIS + IMPORTANTES

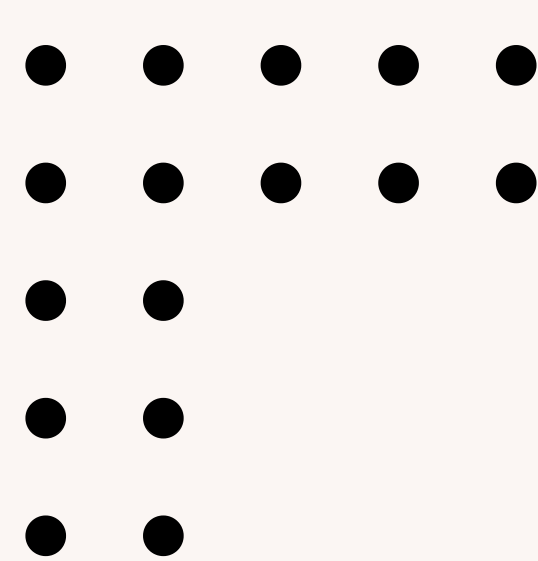




PLOT ACURÁCIA CROSS VALIDATION



Acurácia x Complexidade



RANDOM FOREST

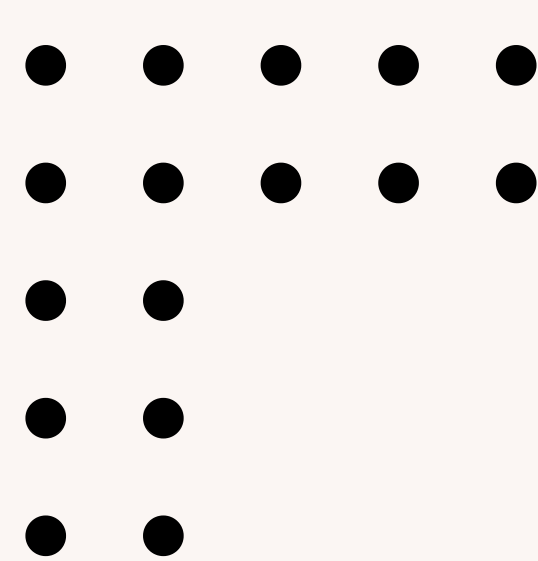
Acurácia: 94%

	Reference	
Prediction	0	1
0	57	5
1	4	43

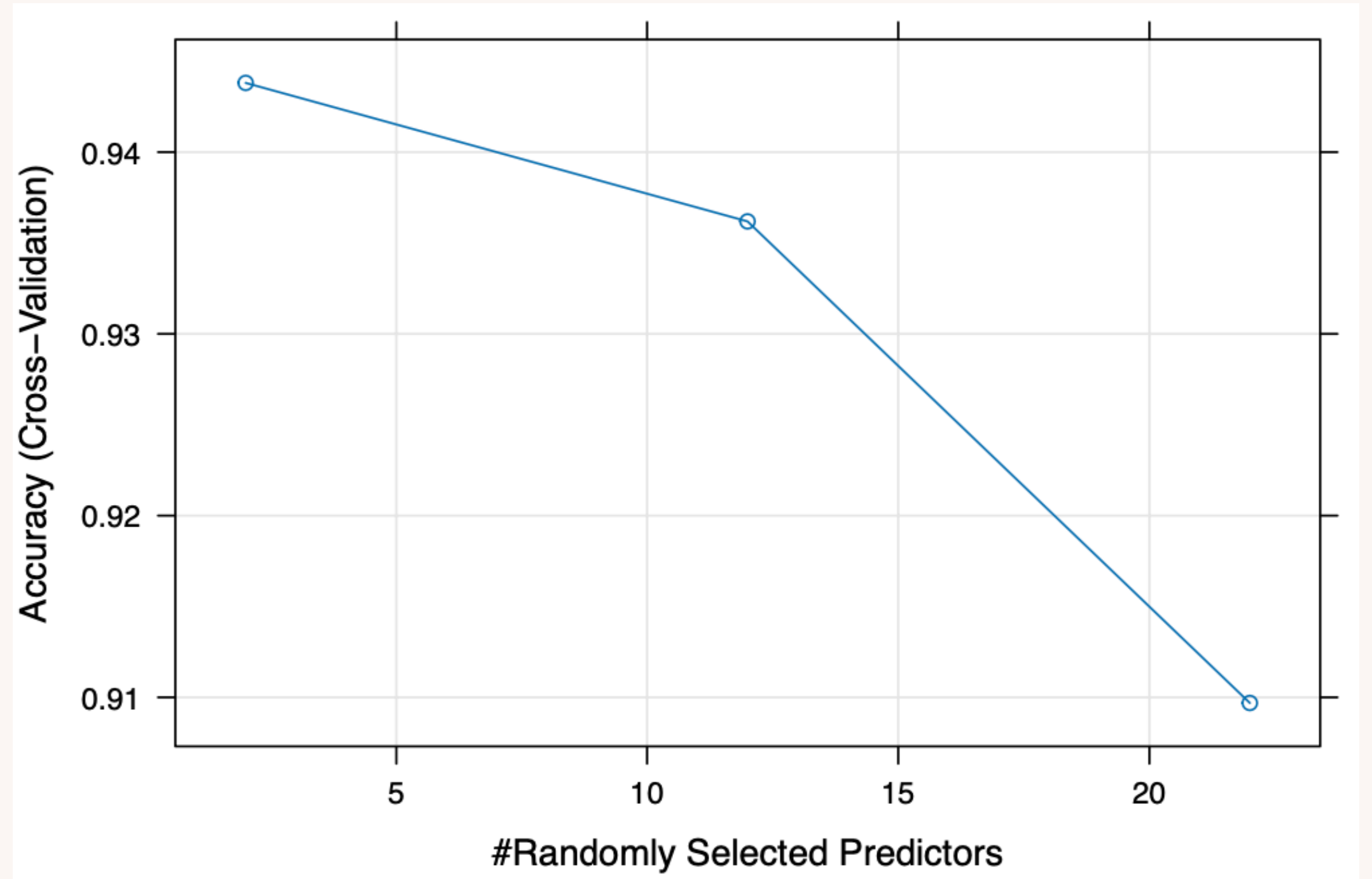
Matriz de confusão

	MeanDecreaseGini
GenderMale	2.33999865
Age	11.95556305
OccupationDoctor	2.77424266
OccupationEngineer	1.97958341
OccupationLawyer	1.44038284
OccupationManager	0.11477089
OccupationNurse	4.95946672
OccupationSales Representative	0.17659797
OccupationSalesperson	2.47682586
OccupationScientist	0.13671391
OccupationSoftware Engineer	0.01952514
OccupationTeacher	1.32092499
Sleep.Duration	7.96856133
Quality.of.Sleep	4.58227247
Physical.Activity.Level	3.66397138
Stress.Level	4.50391342
BMI.CategoryNormal Weight	0.46159706
BMI.CategoryObese	0.98036395
BMI.CategoryOverweight	15.99155492
Blood.PressureHigh	14.61405392
Heart.Rate	5.28131015
Daily.Steps	5.90845729

Variáveis e suas importâncias



PLOT ACURÁCIA CROSS VALIDATION

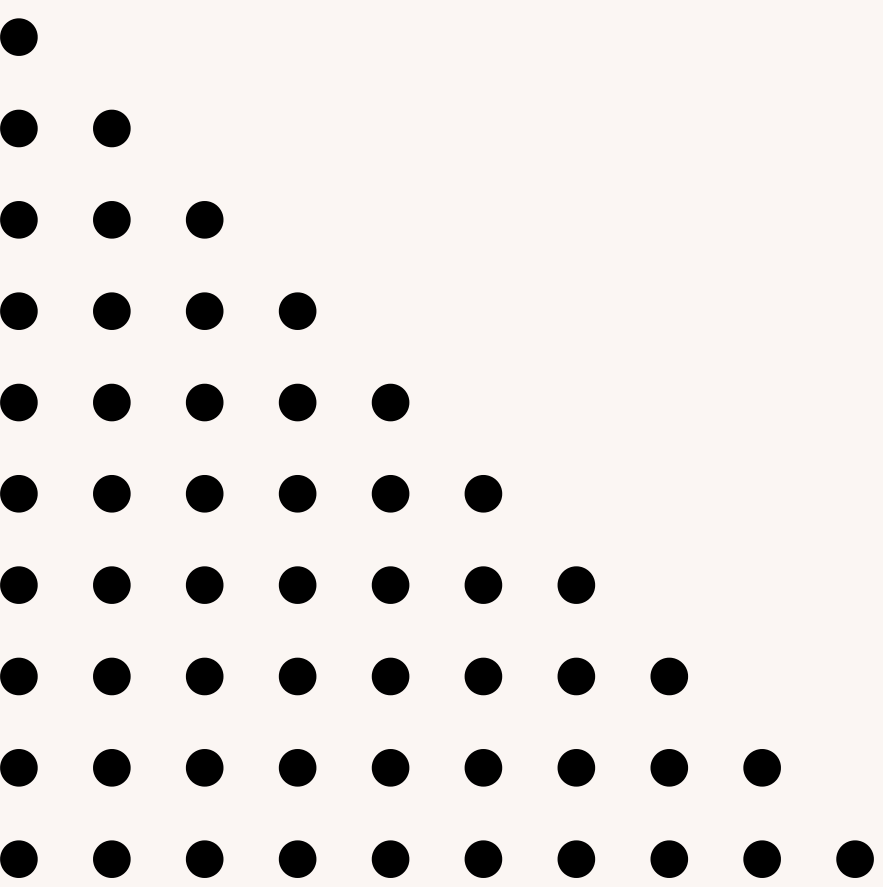


Acurácia x Preditores



CONCLUSÕES

Regressão Logística x Random Forest

- 
- 1 Acurácia de cada um
 - 2 Vantagens e desvantagens de cada
 - 3 Melhor modelo