

Saúde Psicológica e Bem-estar

Diagnóstico e conhecimento do panorama da saúde psicológica e bem-estar das crianças do 1º ciclo no Algarve

Licenciatura em Ciência de Dados 1º ano

UC Análise Exploratória de Dados – CDA1 e CDB1

Discentes: Camila Sousa 111017; Carolina Brunheta 110888; Marco Esperança 110451; Miguel Correia 110786

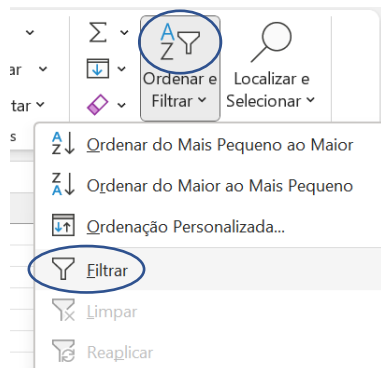
Docentes: Maria do Carmo Botelho; Luís Junqueira; Augusta Correia

Índice

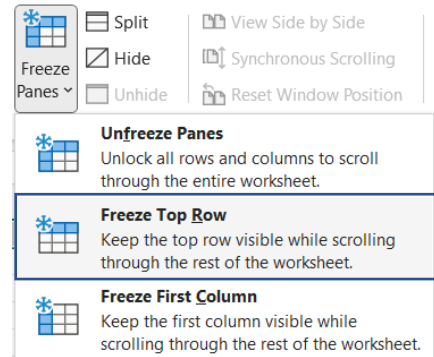
Preparação e organização do dataset em Excel	3
Jamovi	12

Preparação e organização do dataset em Excel

1º Ordenar e filtrar



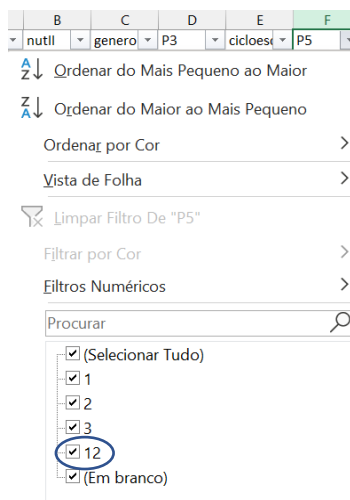
2º Fixar a primeira linha



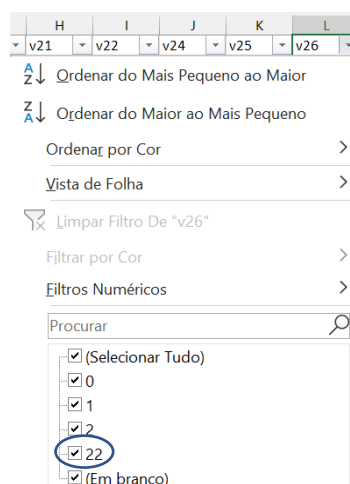
1	ID	nutII	genero	P3	cicloesq	P5	P6	v21	v22	v24	v25	v26	Pj	PI	Ecra_Q
14	13	105	1	7	1	2	2		1	2	0		0	0	1
15	14	105	1	7	1	3	2	1	2	1	2	0	2	2	1
16	15	105	1	7	1	2	2	2	0	1	1	2	0	1	1



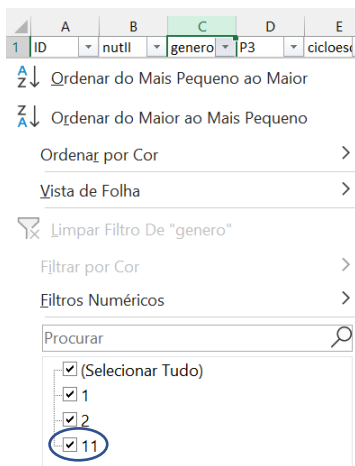
3º Corrigir possíveis erros



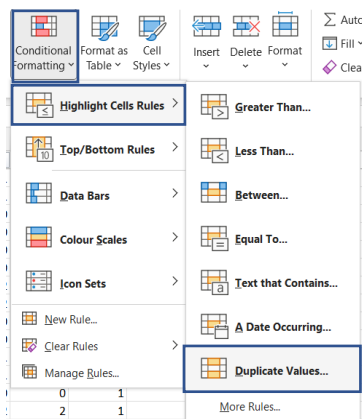
A variável P5, correspondente ao nível de escolaridade completo do pai ou pessoa cuidadora apresenta um valor, 12, que não está presente nas opções 1, 2 e 3.



A variável v26 tem uma resposta 22, e as opções possíveis eram 0,1 e 2.



A variável género tem uma resposta 11. Optamos por eliminar as opções 3 e 4, "Outro" e "Prefiro não responder" respetivamente, por não serem utilizadas (todas as crianças responderam 1 ou 2) (nota: a professora disse que provavelmente foi a professora que escreveu)



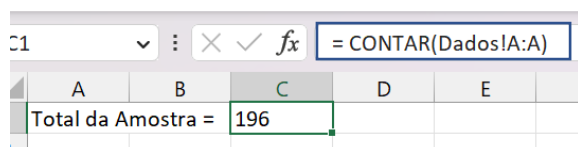
Verificamos se existem valores duplicados, e concluímos que não existem.



4º Atribuir código de não resposta a uma variável com valores omissos

Primeiramente calculamos a taxa de não resposta das variáveis a fim de escolher uma única variável para atribuir o código de não resposta

Numa nova folha de cálculo começamos por calcular o número total da amostra:



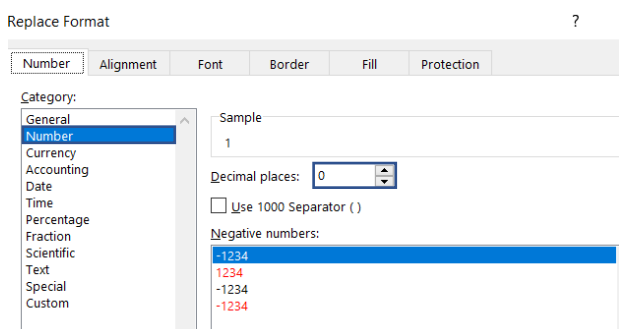
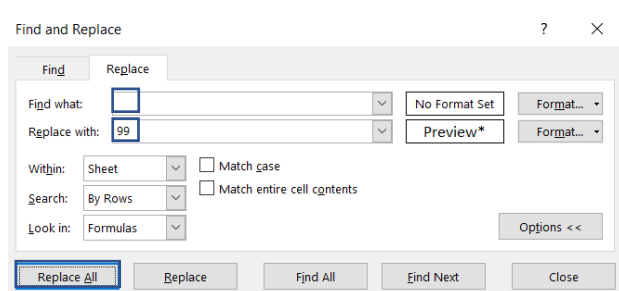
Com a funcionalidade filtro sabemos que as variáveis que possuem células vazias, isto é, em branco, são: ID, género, P5, P6, v21, v24, v26 e Ecra_Q558, as restantes apresentam uma taxa de não resposta (TNR%) de 0%.

Para calcular o número de células em branco destas variáveis recorreremos à função COUNTBLANK.

Para demonstrar a estrutura da função para todas as variáveis numa coluna ao lado usamos a função FORMULATEXTO.

1) Contar o total de registos		
Total da amostra =		196
2) Taxa de não resposta (TNR%)		
Variáveis	TNR%	Fórmula usada
Género:	0,51%	=COUNTBLANK(Dados!C2:C197)/\$C\$3
P5:	10,71%	=COUNTBLANK(Dados!F2:F197)/\$C\$3
P6:	9,18%	=COUNTBLANK(Dados!G2:G197)/\$C\$3
v21:	2,04%	=COUNTBLANK(Dados!H2:H197)/\$C\$3
v24:	1,53%	=COUNTBLANK(Dados!I2:I197)/\$C\$3
v26:	2,04%	=COUNTBLANK(Dados!L2:L197)/\$C\$3
Ecra_Q558:	5,61%	=COUNTBLANK(Dados!O2:O197)/\$C\$3

As variáveis P5 e P6 apresentam percentagens próximas, com uma diferença de 1,53 pontos percentuais, e suficientemente superiores às restantes para considerar ambas, uma vez que se trata do mesmo conceito, mas de géneros diferentes, tomando ambas como nível de escolaridade completo dos encarregados de educação. Ao selecionar as variáveis com Find and Replace substituímos as células em branco por 99 e formatamos a célula como número sem casas decimais. Adicionamos na folha de codificação o código 99 como “Sem resposta”



5º Atribuir texto às categorias de uma variável que tenha códigos

Optamos pela variável “Gênero” para atribuir texto em função da codificação. Para isso inseriu-se uma coluna “gênero decodificado” e usou-se a função HLOOKUP.

✓ fx `=HLOOKUP(Dados!C2:C197;'Variáveis e códigos'!C$4:F$5;2;FALSE)`

D	E	F	G	H	I
genero decodificado	P3	cicloesc	P5	P6	v21
Masculino	6	1	1	1	
Masculino	6	1	3	3	
Masculino	6	1	99	3	
Masculino	6	1	1	1	
Masculino	6	1	2	1	
Masculino	6	1	2	2	
Masculino	6	1	2	1	
Masculino	6	1	3	3	
Masculino	7	1	3	2	
Masculino	7	1	2	1	
Masculino	7	1	2	2	
Masculino	7	1	2	2	
Masculino	7	1	2	2	
Masculino	7	1	3	2	
Masculino	7	1	2	3	
Masculino	7	1	1	1	

1	2	3	4
Masculino	Feminino	Outro	Prefiro não responder

131	105	1	Masculino
132	105	1	Masculino
133	105		#N/A
134	105	1	Masculino

Nas células cujo campo de leitura do código está vazio, surge #N/A, que indica a impossibilidade de executar a correspondência e como resolução, selecionamos a coluna, copiamos e colamos como valores a fim de ficar apenas o resultado da fórmula e não a fórmula, para finalmente fazermos Find and Replace dos valores #N/A para uma célula vazia

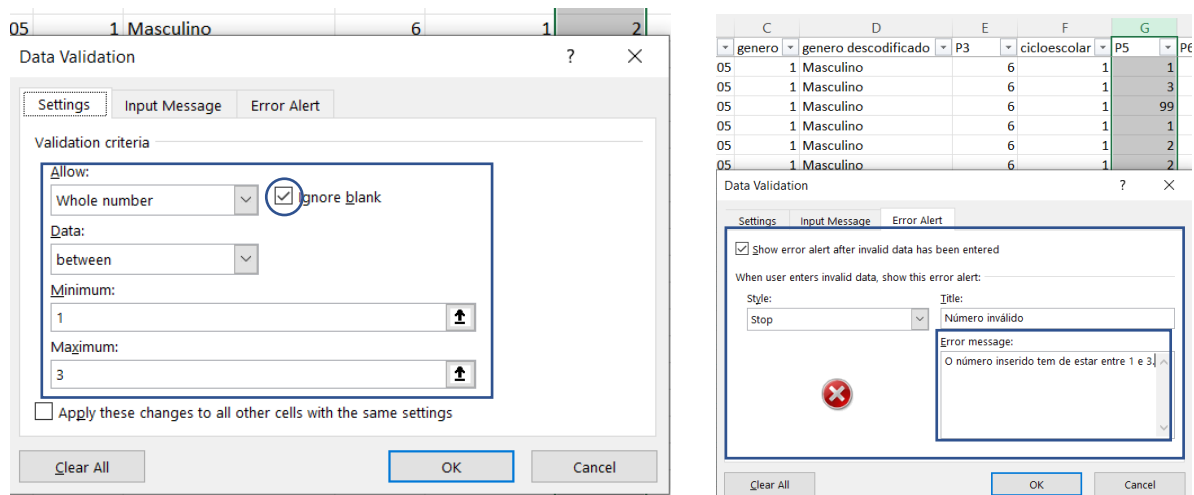


6º Criar regra de validação para duas variáveis:

Qualitativa (Ordinal):

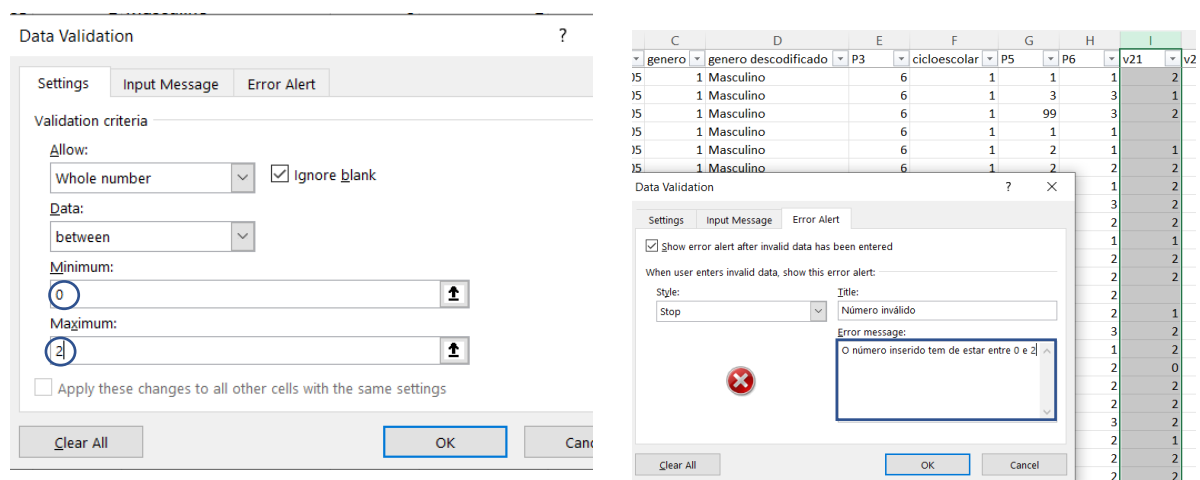
A variável P5 possui uma codificação de 3 valores: 1, 2 e 3 e por isso qualquer outro valor inserido deve ser rejeitado para evitar incoerências na inserção dos dados.

Dito isto, ao selecionar a coluna através de “Data Validation” em “Settings” definimos o intervalo de números permitidos entre 1 e 3 e, em caso de erro, em “Erros Alert” criamos uma mensagem de erro que será apresentada ao inserir um valor errado.



Tendo em conta que a variável P6 tem a mesma codificação achamos lógico fazer esta regra para a mesma.

No que diz respeito às restantes variáveis qualitativas ordinais, v21, v22, v24, v25, v26, pj, pl, que variam entre 0, 1 e 2 aplicamos o mesmo processo anteriormente, mas com valores diferentes.



Quantitativa (Contínua):

A variável Ecra_Q558 corresponde às horas passadas em cada dia à frente de um ecrã e por isso, tendo em conta a duração de um dia em horas, o valor máximo que pode adquirir é 24. Pela mesma razão e pelo mesmo processo referidos anteriormente prosseguimos da seguinte forma:

Data Validation

Settings

Input Message

Error Alert

Validation criteria

Allow:

Whole number

☒ Ignore blank

Data:

between

Minimum:

0

Maximum:

24

☐ Apply these changes to all other cells with the same settings

Clear All

OK

Cancel

Data Validation

Settings

Input Message

Error Alert

☒ Show error alert after invalid data has been entered

When user enters invalid data, show this error alert:

Style:

Stop

Title:

Número de horas diárias inválido

Error message:

O número inserido tem de estar entre 0 e 24

Clear All

OK

Cancel



7º Construir uma tabela de frequências absolutas com funções de contagem e cálculo de percentagens:

Perceber a distribuição por género da amostra total:

Género	n	%
Masculino	100	51,3%
Feminino	95	48,7%
Total	195	100,0%

Mostrar Fórmulas: ☒ Mostrar Fórmulas

Género	n	%
Masculino	=COUNTIF(Dados!C2:C197;1)	=B2/\$B\$4
Feminino	=COUNTIF(Dados!C3:C197;2)	=B3/\$B\$4
Total	=SUM(B2:B3)	=B4/\$B\$4

Distinguir a faixa etária da população em estudo e o respetivo peso na amostra:

P3	n	%
6	47	24,5%
7	32	16,7%
8	60	31,3%
9	45	23,4%
10	8	4,2%
Total	192	100,0%

P3	n	%
6	=COUNTIF(Dados!\$E\$2:\$E\$197;=B7/\$B\$12	
7	=COUNTIF(Dados!\$E\$2:\$E\$197;=B8/\$B\$12	
8	=COUNTIF(Dados!\$E\$2:\$E\$197;=B9/\$B\$12	
9	=COUNTIF(Dados!\$E\$2:\$E\$197;=B10/\$B\$12	
10	=COUNTIF(Dados!\$E\$2:\$E\$197;=B11/\$B\$12	
Total	=SUM(B7:B11)	=B12/\$B\$12

8º Construir uma tabela de frequências dinâmica:



Ao selecionar os dados criamos uma tabela de frequência dinâmica e formatamos da seguinte maneira:

The sequence of screenshots illustrates the configuration of the 'Ecra_Q558' field in a PivotTable:

- Value Field Settings (First):** Source Name: Ecra_Q558. Summarise value field by: Count.
- Value Field Settings (Second):** Source Name: Ecra_Q558. Custom Name: %. Show values as: % of Grand Total. Base field: Ecra_Q558.
- Value Field Settings (Third):** Source Name: Ecra_Q558. Custom Name: %Acum. Summarise value field by: Count.
- Value Field Settings (Fourth):** Source Name: Ecra_Q558. Custom Name: %Acum. Show values as: % Running Total In. Base field: Ecra_Q558.

The **PivotTable Fields** task pane on the right shows the configuration for the 'Ecra_Q558' field:

- Filters:** (Empty)
- Columns:** Σ Values
- Rows:** Ecra_Q558
- Σ Values:** n, %, %Acum

Resultado:

Ecra_Q558	n	%	%Acum
0	4	2,2%	2,2%
1	117	63,2%	65,4%
2	54	29,2%	94,6%
3	9	4,9%	99,5%
4	1	0,5%	100,0%
Total	185	100,0%	



9º Construir uma tabela de cruzamentos dinâmica:

Value Field Settings

Source Name: v26

Custom Name:

Summarise Values By: Show Values As

Summarise value field by

Choose the type of calculation that you want to use to summarise data from the selected field

Sum
Count
Average
Max.
Min.
Product

Number Format

OK Cancel

Value Field Settings

Source Name: v26

Custom Name: %L

Summarise Values By: Show Values As

Summarise value field by

Choose the type of calculation that you want to use to summarise data from the selected field

Sum
Count
Average
Max.
Min.
Product

Number Format

OK Cancel

Value Field Settings

Source Name: v26

Custom Name: %L

Summarise Values By: Show Values As

Show values as

% of Row Total

Base field:

ID
nutri
genero
genero decodificado
P3
cicloescolar

Base item:

Number Format

OK Cancel

PivotTable Fields

Choose fields to add to report:

Search

☐ v24
☐ v25
☒ v26
☐ Pj
☐ PI
☐ Ecra_Q558

More Tables...

Drag fields between areas below:

Filters

Columns

Rows

Σ Values

Σ Values

genero decodificado

n
%L
%C

Value Field Settings

Source Name: v26

Custom Name: %C

Summarise Values By Show Values As

Summarise value field by

Choose the type of calculation that you want to use to summarise data from the selected field

Sum
Count
Average
Max.
Min.
Product

Number Format OK Cancel

Value Field Settings

Source Name: v26

Custom Name: %C

Summarise Values By Show Values As

Show values as

% of Column Total

Base field: ID
nutri
genero
P3
cicloescolar


Base item:

Number Format OK Cancel

v26											Total		
Género	n	0			1			2			n	%L	%C
		%L	%C	n	%L	%C	n	%L	%C				
Feminino	84	89,4%	56,0%	9	9,6%	27,3%	1	1,1%	12,5%	94	100,0%	49,2%	
Masculino	66	68,0%	44,0%	24	24,7%	72,7%	7	7,2%	87,5%	97	100,0%	50,8%	
Total	150	78,5%	100,0%	33	17,3%	100,0%	8	4,2%	100,0%	191	100,0%	100,0%	

Jamovi

Importando o ficheiro excel para o jamovi, realizamos algumas análises de estatística descritiva.

Começamos por escolher as variáveis a relacionar. Depois da seleção começamos por alterar o nome das variáveis “transformar” em  para o significado destas.

De seguida criámos uma tabela de contingência com as variáveis idade e tempo diário de exposição de ecrã:

Tabelas de Contingência

Idade	Tempo diário de exposição a ecrãs					Total
	0	1	2	3	4	
5	4	0	0	0	0	4
6	0	35	8	1	0	44
7	0	17	10	1	1	29
8	0	37	16	3	0	56
9	0	24	20	0	0	44
10	0	4	0	4	0	8
Total	4	117	54	9	1	185

Optamos por criar ainda uma só relativa às percentagens em coluna separadamente para ter uma melhor leitura:

Tabelas de Contingência

Idade		Tempo diário de exposição a ecrãs					Total
		0	1	2	3	4	
5	% em linha	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %
6	% em linha	0.0 %	79.5 %	18.2 %	2.3 %	0.0 %	100.0 %
7	% em linha	0.0 %	58.6 %	34.5 %	3.4 %	3.4 %	100.0 %
8	% em linha	0.0 %	66.1 %	28.6 %	5.4 %	0.0 %	100.0 %
9	% em linha	0.0 %	54.5 %	45.5 %	0.0 %	0.0 %	100.0 %
10	% em linha	0.0 %	50.0 %	0.0 %	50.0 %	0.0 %	100.0 %
Total	% em linha	2.2 %	63.2 %	29.2 %	4.9 %	0.5 %	100.0 %

Em R:

```
Jmv::contTables(  
  formula = ~ Idade: 'Tempo diário de exposição a ecrãs',  
  data = data,  
  ci = FALSE,  
  obs = FALSE,  
  pcRow = TRUE)  
jmv::descriptives(  
  data = data,  
  vars = Idade)
```

De seguida obtemos uma tabela referente, somente, à idade:

Estatística Descritiva

	Idade
N	196
Omisso	0
Média	7.61
Mediana	8.00
Desvio-padrão	1.25
Mínimo	5
Máximo	10

No que diz respeito à irritabilidade e birras e partilha com outras crianças obtemos a seguinte estatística descritiva:

Estatística Descritiva

Estatística Descritiva

	compartilhamento com outras crianças	irritabilidade e birras
N	193	196
Omisso	3	0
Média	1.56	0.372
Mediana	2	0.00
Desvio-padrão	0.651	0.581
Mínimo	0	0
Máximo	2	2

Frequências

Frequências de compartilhamento com outras crianças

	compartilhamento com outras crianças	Contagens	% do Total	% acumulada
0		17	8.8 %	8.8 %
1		50	25.9 %	34.7 %
2		126	65.3 %	100.0 %

Frequências de irritabilidade e birras

irritabilidade e birras	Contagens	% do Total	% acumulada
0	133	67.9 %	67.9 %
1	53	27.0 %	94.9 %
2	10	5.1 %	100.0 %

Em R:

```
jmv::descriptives(
  data = data,
  vars = vars(compartilhamento com outras crianças, irritabilidade e birras),
  freq = TRUE)
```

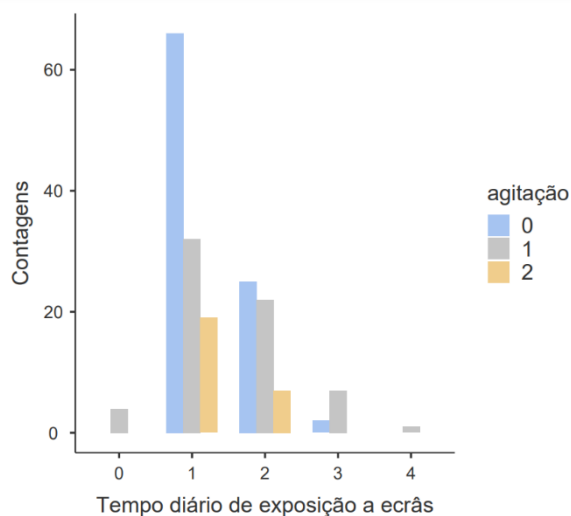
Nesta tabela acabamos por deixar as variáveis com o nome predefinido para que a tabela ficasse simples, para analisarmos alguns dos comportamentos.

Estatística Descritiva

Estatística Descritiva

	v26	v25	v21	v24	v22
N	192	196	192	193	196
Omisso	4	0	4	3	0
Média	0.255	0.372	1.73	1.56	0.602
Mediana	0.00	0.00	2.00	2	0.00
Desvio-padrão	0.524	0.581	0.548	0.651	0.712
Mínimo	0	0	0	0	0
Máximo	2	2	2	2	2

A fim de relacionar alguns comportamentos com o tempo de exposição a ecrãs, construímos os seguintes gráficos:



```
jmv::contTables(
  formula = ~ Tempo diário de
exposição de ecrãs:agitação,
  data = data,
  barplot = TRUE)
```

