

Visualização de Dados em R

Por Samuel Chagas de Assis

O que não é essa apresentação?

Regra

Alguns conceitos são subjetivos.

Tutorial Completo

Visualização de dados envolve o meio, objetivo, tipo de dados, viabilidade, etc. Não se limite por aqui.

Apenas um primeiro encontro

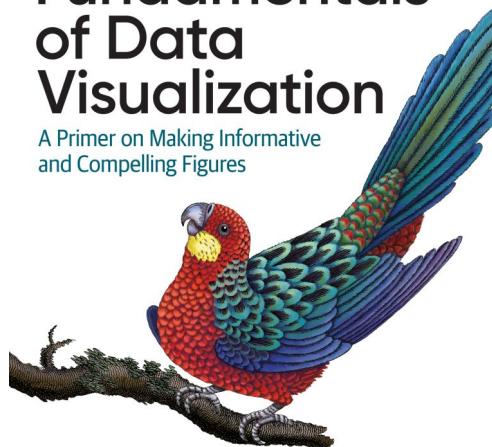
Poucos dados ainda? Aprenda agora, sofra menos depois. O R é uma ferramenta poderosa.

Principais Referências

O'REILLY®

Fundamentals of Data Visualization

A Primer on Making Informative
and Compelling Figures

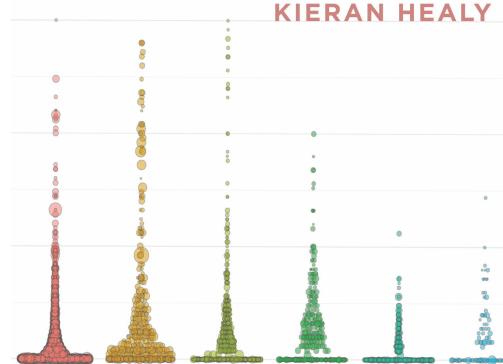


Claus O. Wilke

DATA VISUALIZATION

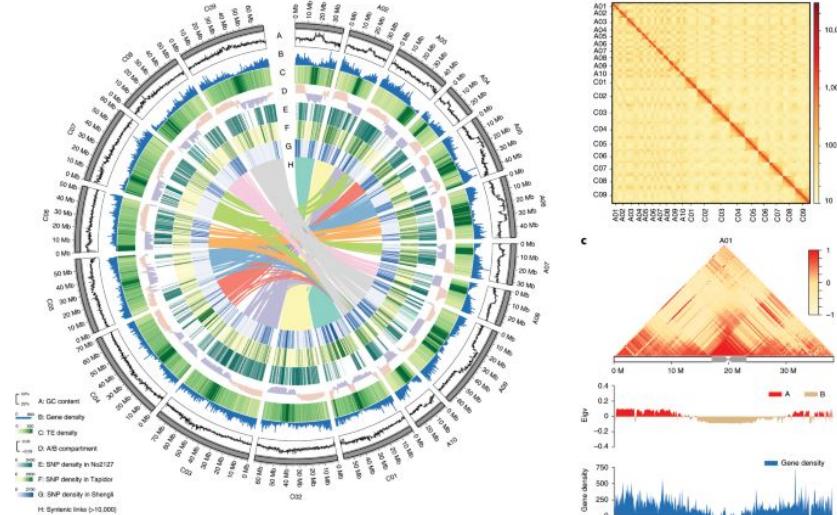
A PRACTICAL INTRODUCTION

KIERAN HEALY



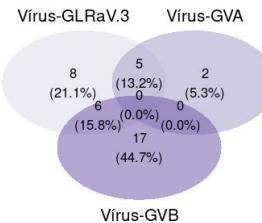
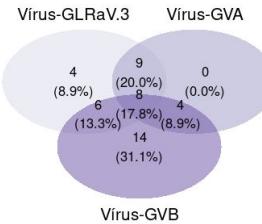
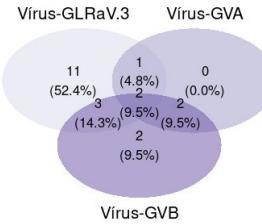
A visualização de dados é uma parte arte e outra parte ciência. O desafio é acertar na arte sem errar na ciência, e vice-versa.

Claus O. Wilke



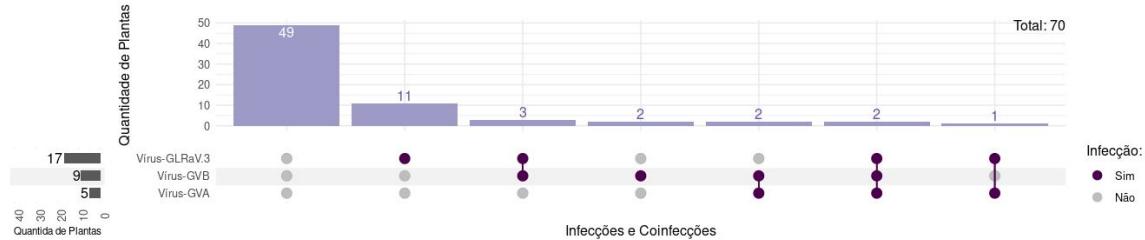
*Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nature Plants, 2020.*

Ajudam a contar histórias

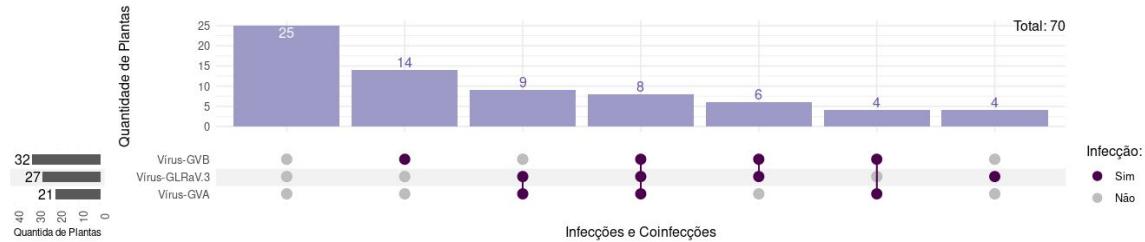


Ajudam a contar histórias

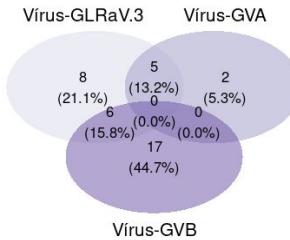
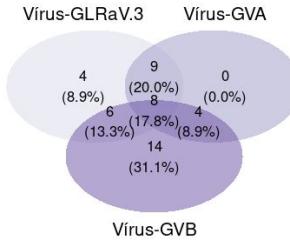
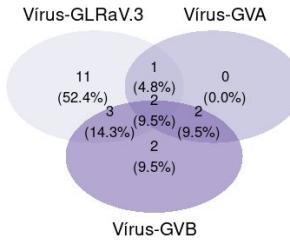
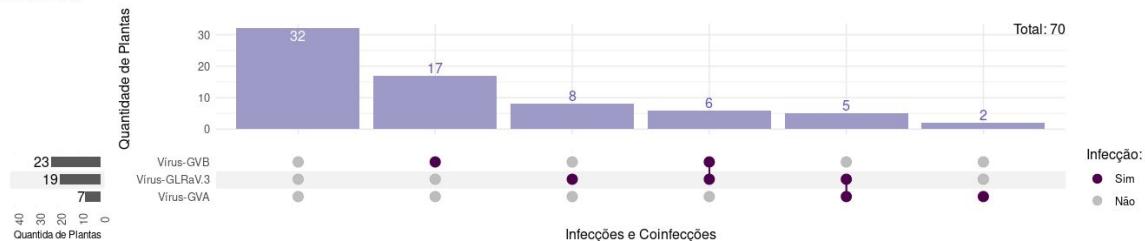
Vinícola A



Vinícola B



Vinícola C



Ajudam a contar histórias

	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Age	Sex	Residence	Residence_UF	Origin	Origin_UF	Nationality	Race	Smok_Bv	Alcoh_Bv	FirstSymptons_c	KitCovid	Comorbidities	Others_
2	59	M	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	BRANCO	2	2	14	3	DIABETES, HAS, DOEN	
3	43	M	FOZ DO IGUAC PR		CURITIBA PR	PR	BRASILEIRA	BRANCO	3	2	14	1	OBESIDADE, HAS	
4	65	M	SANTA TEREZII PR		IRAI - RS	RS	BRASILEIRA	BRANCO	3	9	14	1	OBESIDADE, DPOC, HA	
5	53	F	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	PARDO	2	9	10	2	OBESIDADE, DIABETES	
6	65	F	FOZ DO IGUAC PR		LAGOA VERME RS		BRASILEIRA	BRANCO	2	2	15	1	DPOC, DIABETE DPOC -	
7	67	M	MARECHAL CA PR		FOZ DO IGUAC PR		BRASILEIRA	PARDO	1	9	7	1	AVC	
8	54	F	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	BRANCO	2	2	7	1	OBESIDADE	
9	40	F	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	PARDO	2	2	8	2	HAS, OUTRAS LES	
10	39	F	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	PARDO	2	2	6	2	OBESIDADE, SEM MOR	
11	56	F	FOZ DO IGUAC PR		DOIS VIZINHOS PR		BRASILEIRA	PARDO	2	2	15	1	OBESIDADE, DIARTRO:	
12	37	M	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	PARDO	2	2	11	2	OBESIDADE	
13	46	F	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	PARDO	2	2	10	1	HAS	
14	55	M	FOZ DO IGUAC PR		TUCUNDUVA R PR		BRASILEIRA	BRANCO	3	2	11	1	OBESIDADE, DIABETES	
15	66	M	FOZ DO IGUAC PR		DOIS VIZINHOS PR		BRASILEIRA	PARDO	2	9	8	1	OBESIDADE, DIABETES	
16	78	F	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	PARDO	9	9	15	1	SEM MORBIDADES	
17	67	M	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	BRANCO	9	9	10	2	OBESIDADE, DIABETES	
18	45	F	FOZ DO IGUAC PR		TOLEDO PR		BRASILEIRA	BRANCO	9	9	0	3	OBESIDADE, H	
19	44	F	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	BRANCO	2	2	9	2	OBESIDADE, DIABETES	
20	54	M	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	BRANCO	9	9	15	1	SEM MORBIDADES	
21	42	M	FOZ DO IGUAC PR		SAO MIGUEL D PR		BRASILEIRA	PRETO	9	9	7	1	OBESIDADE, DIABETES	
22	58	F	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	PARDO	9	9	12	3	OBESIDADE, DPOC, DI	
23	51	F	FOZ DO IGUAC PR		SANTIAGO RS PR		BRASILEIRA	PARDO	2	2	8	3	DIABETES, HAS	
24	56	M	FOZ DO IGUAC PR		FOZ DO IGUAC PR		BRASILEIRA	BRANCO	3	2	6	3	DIABETES, HAS, DOEN	

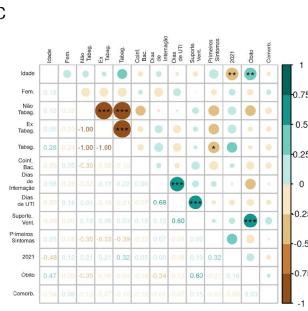
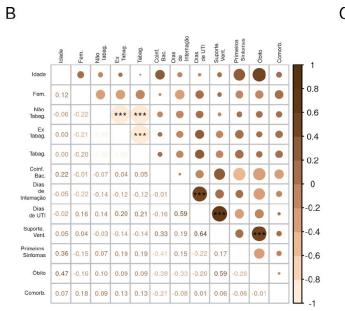
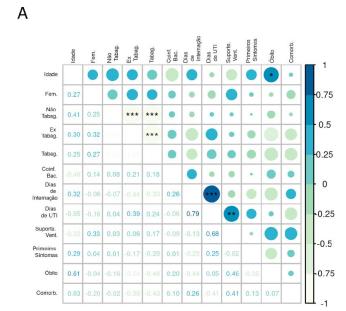
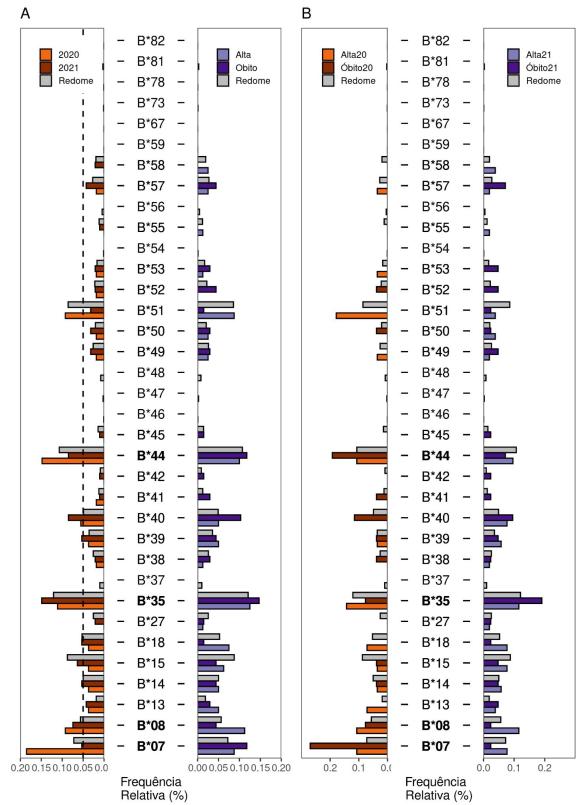
Ajudam a contar histórias

	Desfechos						p-valores		
	Alta20, N = 14 ¹	Alta21, N = 26 ¹	Obito20, N = 13 ¹	Obito21, N = 21 ¹	2020, N = 27 ¹	2021, N = 47 ¹	Alta20 vs. Obito20 ²	Alta21 vs. Obito21 ²	2020 vs. 2021 ²
Idade (em anos)	57,5 (52,8, 62,5)	44,0 (42,0, 50,0)	75,0 (65,0, 79,0)	58,0 (54,0, 65,0)	64,0 (55,5, 75,5)	52,0 (43,0, 58,5)	0,010	<0,001	<0,001
Faixa-etária (em anos)	0,0 (0,0%)	0,0 (0,0%)	1,0 (7,7%)	0,0 (0,0%)	1,0 (3,7%)	0,0 (0,0%)	0,029	<0,001	0,002
20-29	1,0 (7,1%)	6,0 (23,1%)	0,0 (0,0%)	1,0 (4,8%)	1,0 (3,7%)	7,0 (14,9%)			
30-39	2,0 (14,3%)	13,0 (50,0%)	0,0 (0,0%)	2,0 (9,5%)	2,0 (7,4%)	15,0 (31,9%)			
40-49	7,0 (50,0%)	5,0 (19,2%)	1,0 (7,7%)	11,0 (52,4%)	8,0 (29,6%)	16,0 (34,0%)			
50-59	2,0 (14,3%)	2,0 (7,7%)	4,0 (30,8%)	5,0 (23,8%)	6,0 (22,2%)	7,0 (14,9%)			
60-69	2,0 (14,3%)	0,0 (0,0%)	6,0 (46,2%)	2,0 (9,5%)	8,0 (29,6%)	2,0 (4,3%)			
70-79	0,0 (0,0%)	0,0 (0,0%)	1,0 (7,7%)	0,0 (0,0%)	1,0 (3,7%)	0,0 (0,0%)			
80+									
Sex									
<i>Feminino</i>	4,0 (28,6%)	14,0 (53,8%)	8,0 (61,5%)	10,0 (47,6%)	12,0 (44,4%)	24,0 (51,1%)	0,085	0,7	0,6
<i>Masculino</i>	10,0 (71,4%)	12,0 (46,2%)	5,0 (38,5%)	11,0 (52,4%)	15,0 (55,6%)	23,0 (48,9%)			
Raça Autodeclarada									
<i>Ásia/Ática</i>	1,0 (7,1%)	1,0 (3,8%)	0,0 (0,0%)	0,0 (0,0%)	1,0 (3,7%)	1,0 (2,1%)	>0,9	>0,9	0,8
<i>Branca</i>	7,0 (50,0%)	10,0 (38,5%)	7,0 (53,8%)	10,0 (47,6%)	14,0 (51,9%)	20,0 (42,6%)			
<i>Parda</i>	6,0 (42,9%)	14,0 (53,8%)	6,0 (46,2%)	11,0 (52,4%)	12,0 (44,4%)	25,0 (53,2%)			
<i>Preta</i>	0,0 (0,0%)	1,0 (3,8%)	0,0 (0,0%)	0,0 (0,0%)	0,0 (0,0%)	1,0 (2,1%)			
Número de Comorbidades									
<i>≥ 3 Comorbidades</i>	1,0 (7,1%)	0,0 (0,0%)	0,0 (0,0%)	2,0 (9,5%)	1,0 (3,7%)	2,0 (4,3%)			
<i>1 Comorbidade</i>	6,0 (42,9%)	10,0 (38,5%)	6,0 (46,2%)	6,0 (28,6%)	12,0 (44,4%)	16,0 (34,0%)			
<i>2 Comorbidades</i>	3,0 (21,4%)	8,0 (30,8%)	6,0 (46,2%)	8,0 (38,1%)	9,0 (33,3%)	16,0 (34,0%)			
<i>Sem Comorbidades</i>	4,0 (28,6%)	8,0 (30,8%)	1,0 (7,7%)	5,0 (23,8%)	5,0 (18,5%)	13,0 (27,7%)			
Tabagismo									
<i>Ex</i>	6,0 (42,9%)	2,0 (7,7%)	1,0 (7,7%)	4,0 (19,0%)	7,0 (25,9%)	6,0 (12,8%)			
<i>Ignorado</i>	1,0 (7,1%)	3,0 (11,5%)	5,0 (38,5%)	4,0 (19,0%)	6,0 (22,2%)	7,0 (14,9%)			
<i>Não</i>	6,0 (42,9%)	19,0 (73,1%)	4,0 (30,8%)	11,0 (52,4%)	10,0 (37,0%)	30,0 (63,8%)			
<i>Sim</i>	1,0 (7,1%)	2,0 (7,7%)	3,0 (23,1%)	2,0 (9,5%)	4,0 (14,8%)	4,0 (8,5%)			
Alcoolismo									
<i>Ex</i>	1,0 (7,1%)	0,0 (0,0%)	0,0 (0,0%)	0,0 (0,0%)	1,0 (3,7%)	0,0 (0,0%)	0,2	0,8	0,3
<i>Ignorado</i>	3,0 (21,4%)	8,0 (30,8%)	8,0 (61,5%)	9,0 (42,9%)	11,0 (40,7%)	17,0 (36,2%)			
<i>Não</i>	8,0 (57,1%)	17,0 (65,4%)	4,0 (30,8%)	11,0 (52,4%)	12,0 (44,4%)	28,0 (59,6%)			
<i>Sim</i>	2,0 (14,3%)	1,0 (3,8%)	1,0 (7,7%)	1,0 (4,8%)	3,0 (11,1%)	2,0 (4,3%)			
Coinfec. Bacteriana									
<i>Não</i>	6,0 (42,9%)	10,0 (38,5%)	8,0 (61,5%)	11,0 (52,4%)	14,0 (51,9%)	21,0 (44,7%)	0,3	0,3	0,6
<i>Sim</i>	8,0 (57,1%)	16,0 (61,5%)	5,0 (38,5%)	10,0 (47,6%)	13,0 (48,1%)	26,0 (55,3%)			
Dias de Internação									
31,5 (26,2)	29,5 (20,0)	22,0 (19,0)	28,0 (19,0)	30,0 (20,0)	29,0 (19,5)	0,3	0,6	>0,9	
41,5)	40,0)	32,0)	36,0)	39,5)	38,0)				
Dias de UTI									
25,5 (17,8)	24,0 (11,2)	20,0 (13,0)	24,0 (16,0)	25,0 (14,5)	24,0 (13,5)	0,6	0,8	0,8	
30,8)	35,0)	30,0)	30,0)	30,5)	35,0)				
Dias de Suporte Ventilatório									
17,5 (11,2)	20,5 (9,2)	22,0 (13,0)	25,0 (18,0)	20,0 (11,5)	24,0 (13,0)	0,3	0,057	0,6	
24,5)	29,2)	30,0)	36,0)	30,0)	32,5)				
Período Primeiros Sintomas									
9,2 (3,8)	9,7 (4,8)	6,2 (5,0)	9,5 (3,4)	7,7 (4,6)	9,6 (4,2)	0,075	0,7	0,058	

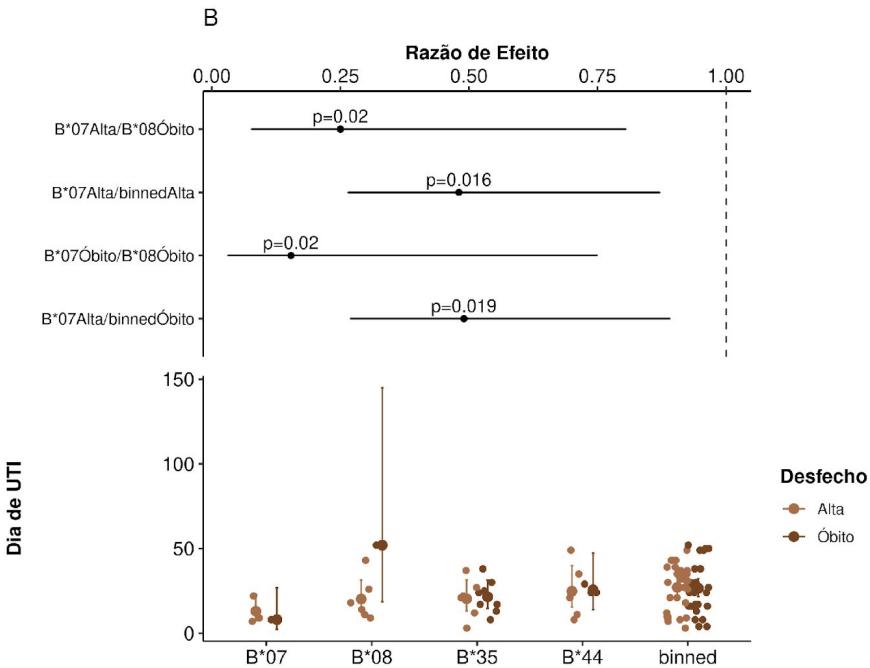
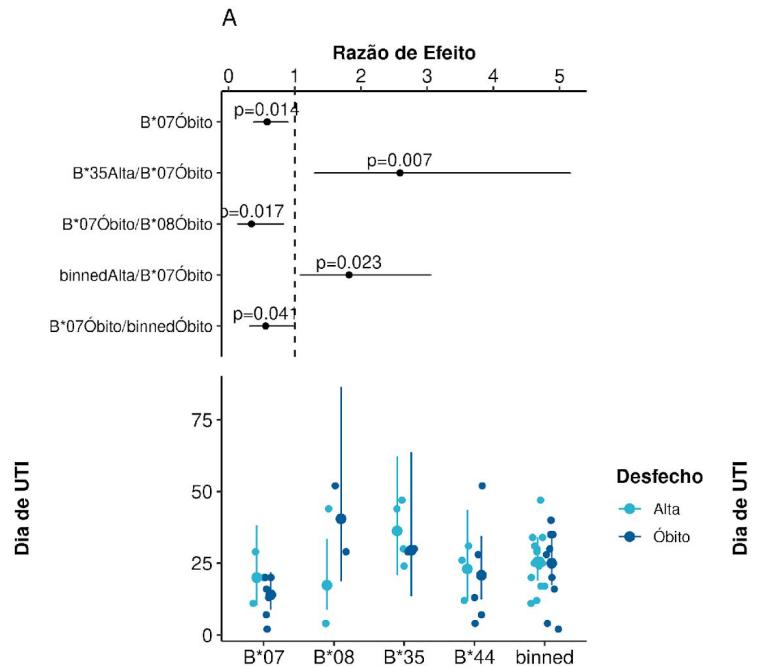
¹Median (IQR); n (%); Mean (SD)

²Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

Ajudam a contar histórias

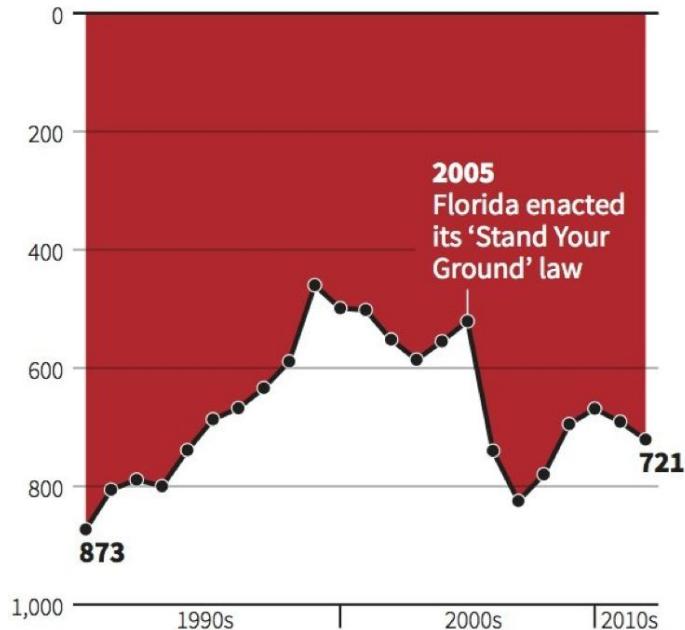


Ajudam a contar histórias



Gun deaths in Florida

Number of murders committed using firearms

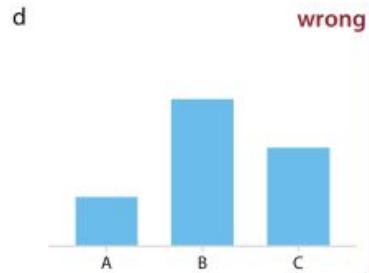
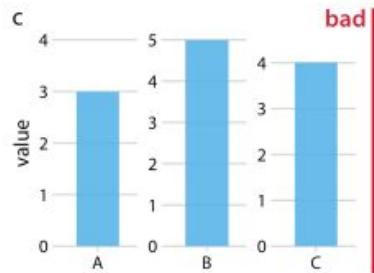
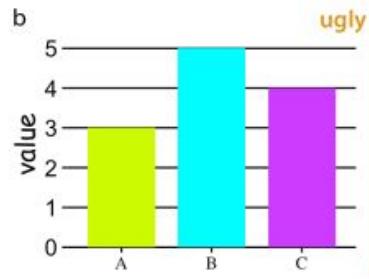
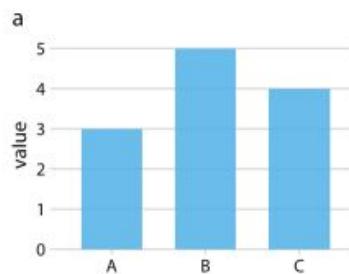


Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

O problema do Feio, Errado e Negativo



O problema do Feio, Errado e Negativo

2 MANIPULATING THE Y-AXIS

Expanding or compressing the scale on a graph can make changes in data seem more or less significant than they actually are.

MISLEADING

- The scale is disproportionate to the data, making the change over time seem small

VS

ACCURATE

- The scale is proportionate to the data, showing a greater change over time

4 USING THE WRONG GRAPH

The type of graph you use should depend on the type of data you want to visualize. Using the wrong type of graph can skew the data. Writers will sometimes use the wrong type of graph on purpose.

MISLEADING

- Pie charts are used to compare parts of a whole, not the difference between groups
- A different type of graph should be used to compare the three teams

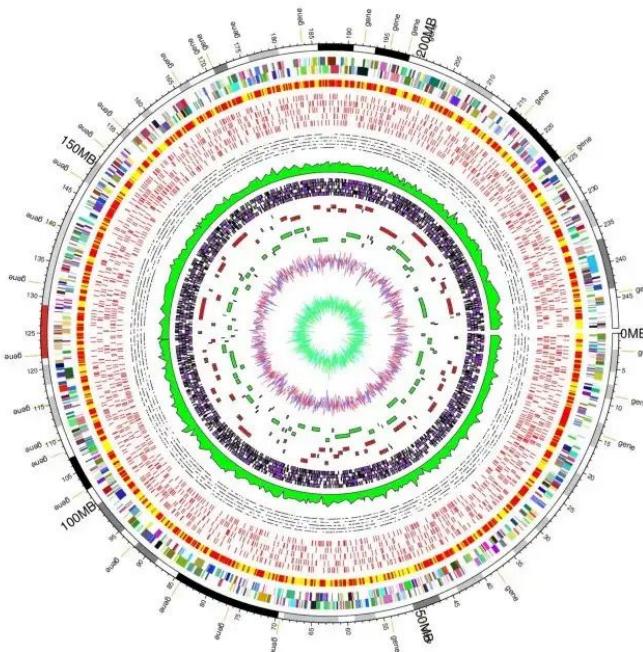
VS

ACCURATE

- Bar graphs are better for showing the differences between groups
- This chart is a better visualization of the data

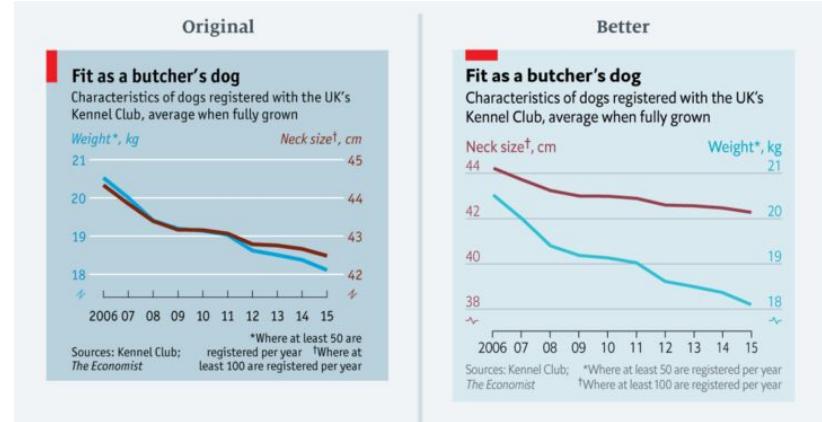
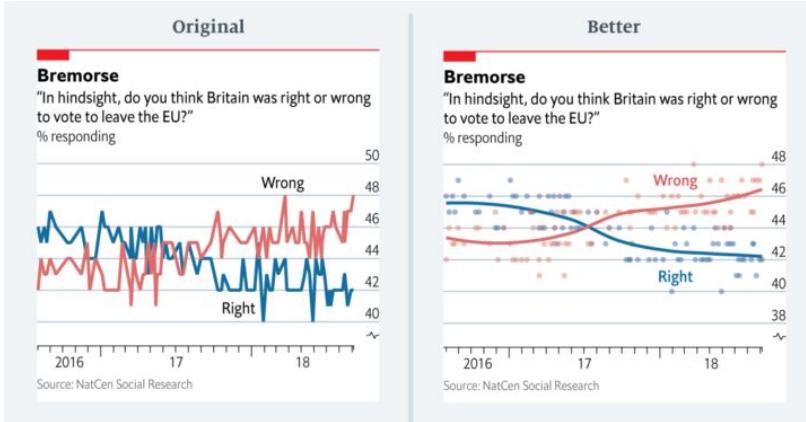
5 Ways Writers Use Misleading Graphs To Manipulate You [INFOGRAPHIC] - Venngage

O problema do Feio, Errado e Negativo



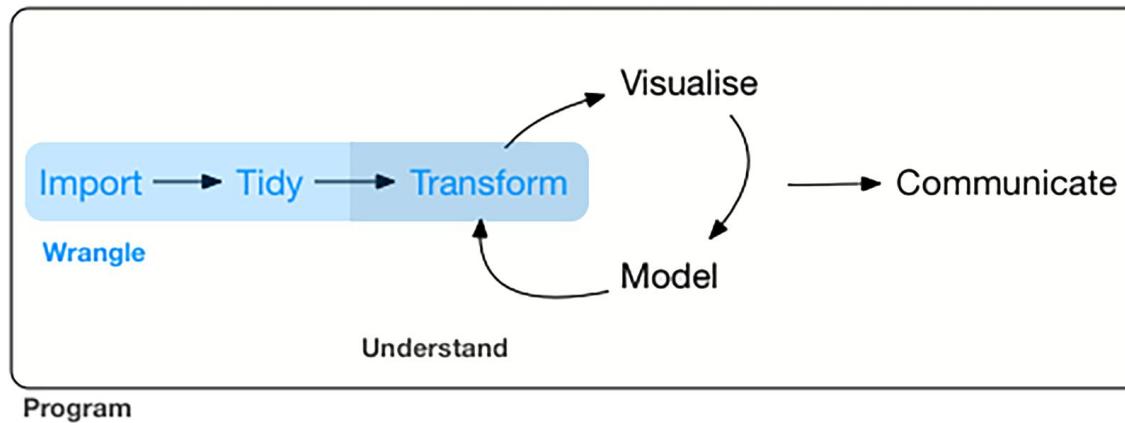
[Making genomic data come alive with circos plots](#)
| by Maria Nattestad | Medium.

O problema do Feio, Errado e Negativo



Mistakes, we've drawn a few. Learning from our errors in data... | by Sarah Leo | The Economist

Análise e Visualização de Dados



[Chapter 2 Basic data wrangling | Data
Visualisation using COVID-19 Statistics](#)

Análise e Visualização de Dados em R



Análise e Visualização de Dados em R

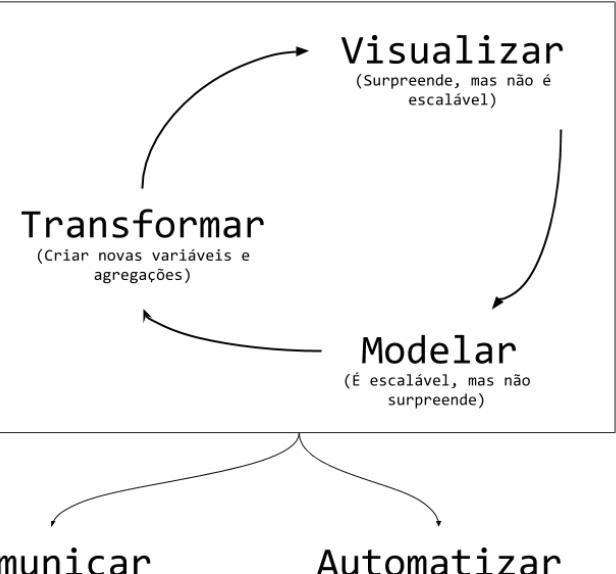


Importar

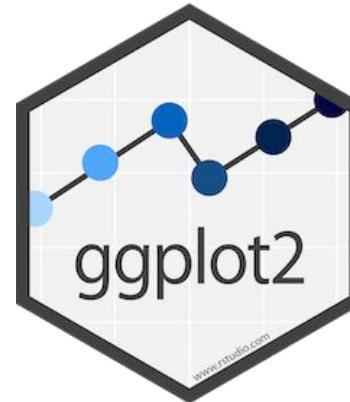


Arrumar

(Armazenar os dados
consistentemente)



Análise e Visualização de Dados em R



Análise e Visualização de Dados em R



Importar

readr
xml2
haven
readxl

Arrumar →

tidyverse
janitor
tibble

%>%
magrittr

Transformar

dplyr
stringr
forcats
lubridate
data.table

Visualizar

ggplot2

Modelar

caret
mlr
lme4
keras

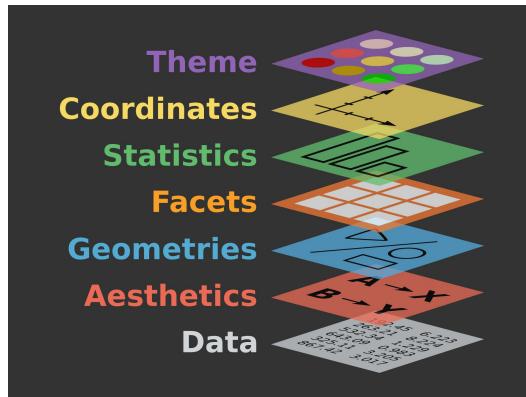
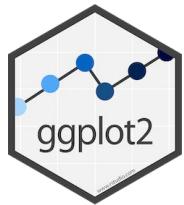
Comunicar

rmarkdown shiny

Automatizar

plumber openCPU

Análise e Visualização de Dados em R



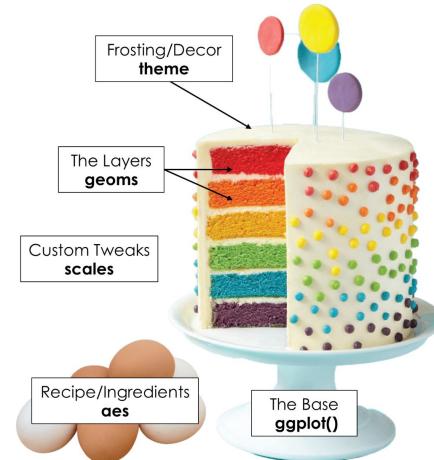
**ggplot is a little bit
like cake...**

We always start by setting up the foundation with `ggplot()`

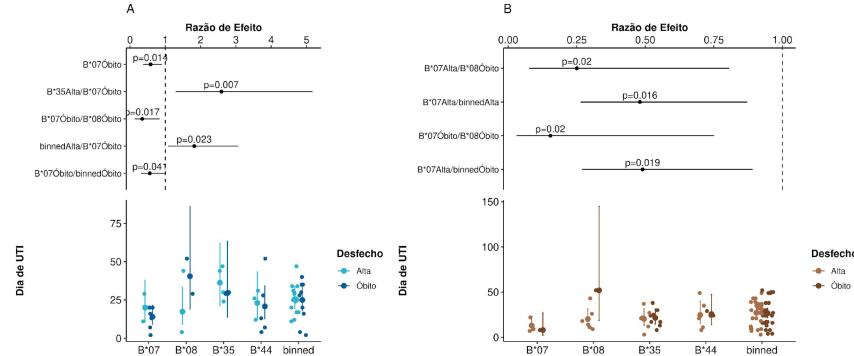
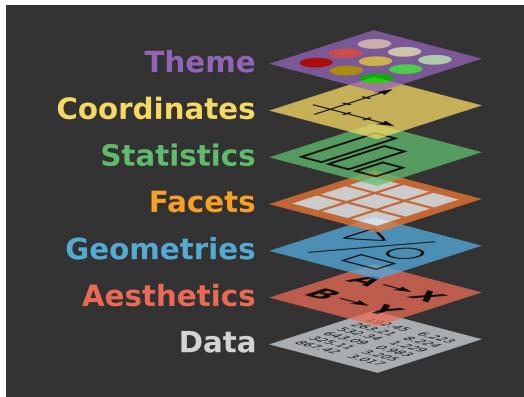
We specify our ingredients (data variables) with an `aes` mapping

We can create layers to our plot with `geoms`

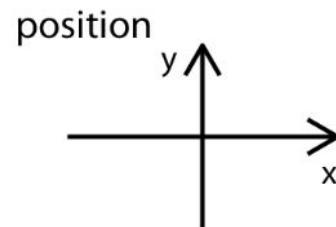
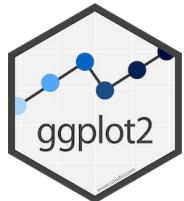
We can style our `cake` ggplot with `themes`. We have out-of-the-box options, or we can go totally custom!



Análise e Visualização de Dados em R



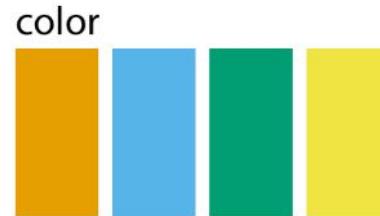
Análise e Visualização de Dados em R



shape



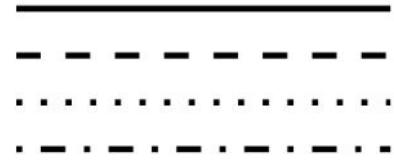
size



line width



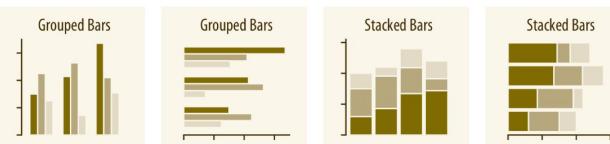
line type



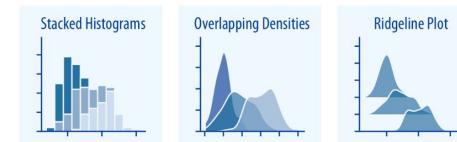
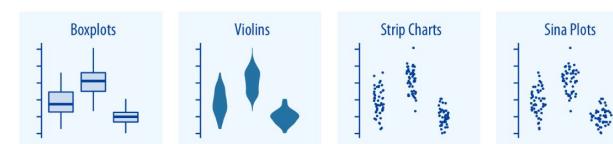
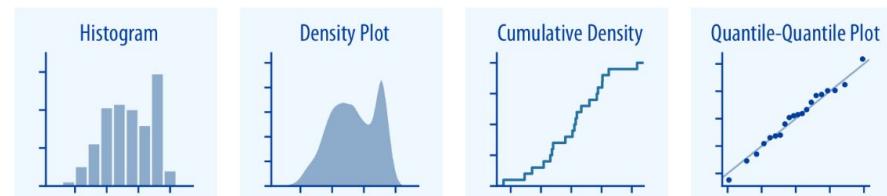
Análise e Visualização de Dados em R



Quantidades



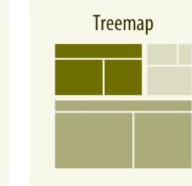
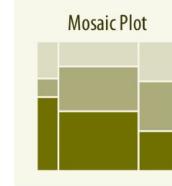
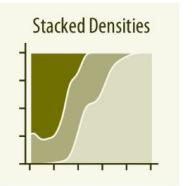
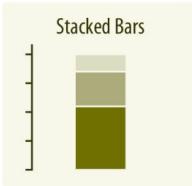
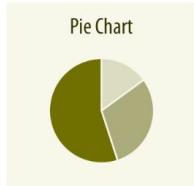
Distribuições



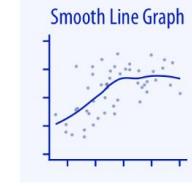
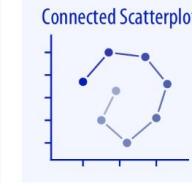
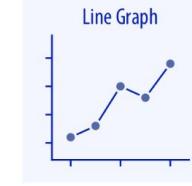
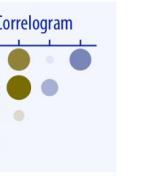
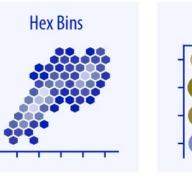
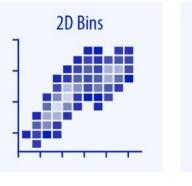
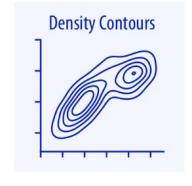
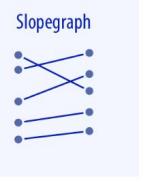
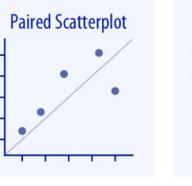
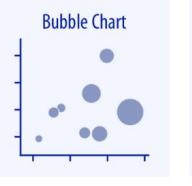
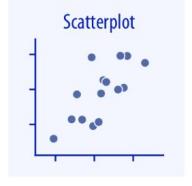
Análise e Visualização de Dados em R



Proporções



Relações x-y



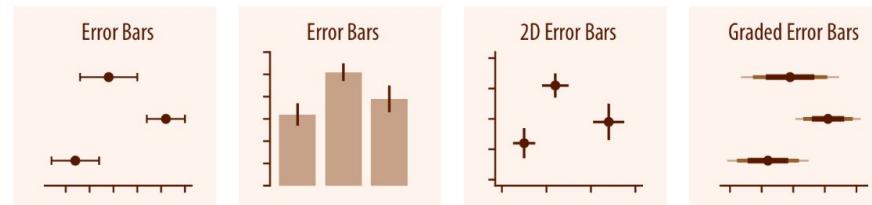
Análise e Visualização de Dados em R



Dados Geoespaciais



Incerteza



Algumas dicas

Curso-R - Livro (Português)

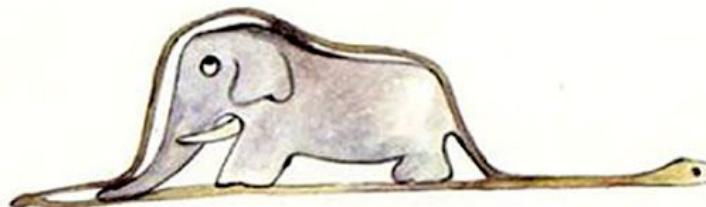
The R Graph Gallery Vários exemplos de gráficos com códigos)

R for Data Science (Hadley Wickham, criador do ggplot2, reshape2, etc)

Style guide · Advanced R. (Wickham style - making things easy to read)

Chapter 4 Complete protocols | DataViz protocols

As pessoas grandes sempre têm a necessidade de explicações detalhadas...



“É um chapéu ou uma jibóia
digerindo um elefante?”

–O Pequeno Príncipe