

ANÁLISE DE DADOS EM R

Fernando de Souza, Me.



SUMÁRIO



01

Visualização de Dados

Porque isso importa?

02

O que evitar

Não repita isso em casa

03

O que agregar

Mudando suas visões sobre o assunto

04

ggplot2

Liberte o seu designer interior

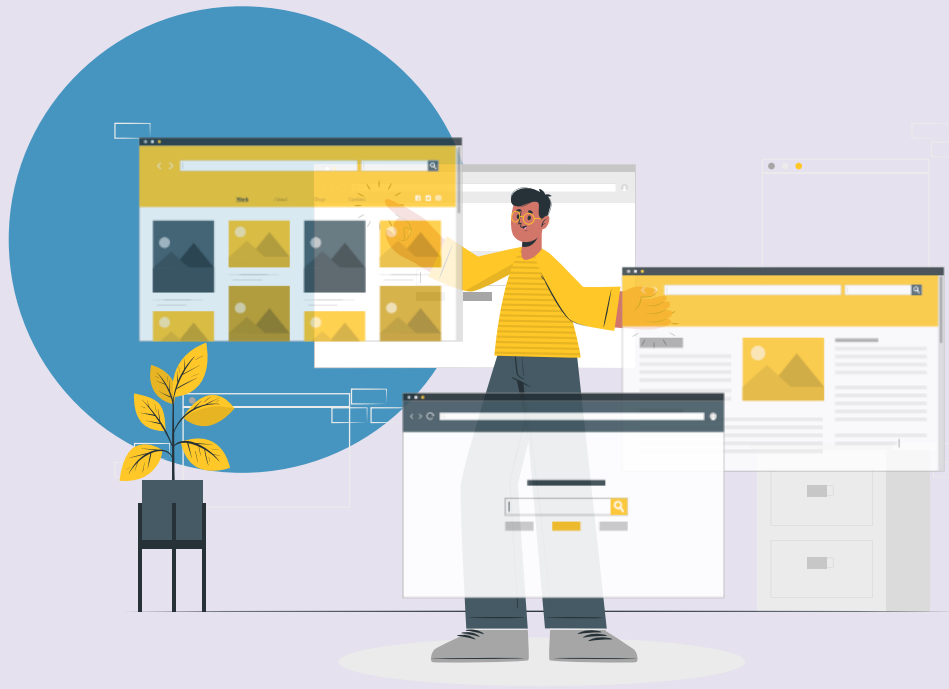
SOBRE O CURSO



Nosso curso irá percorrer 4 tópicos Principais:

1. Programação
2. Limpeza e Organização
3. Análise e Visualização
4. Comunicação





01

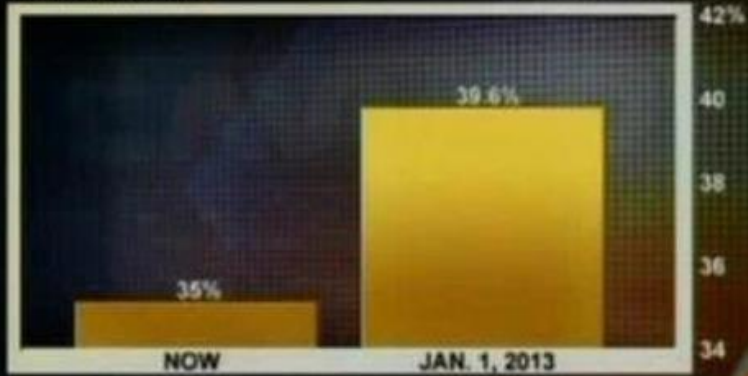
Visualização de Dados

Porque isso importa?



IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



8:01p ET

FOX
BUSINESS

TOP STORIES

TECHNOLOGY

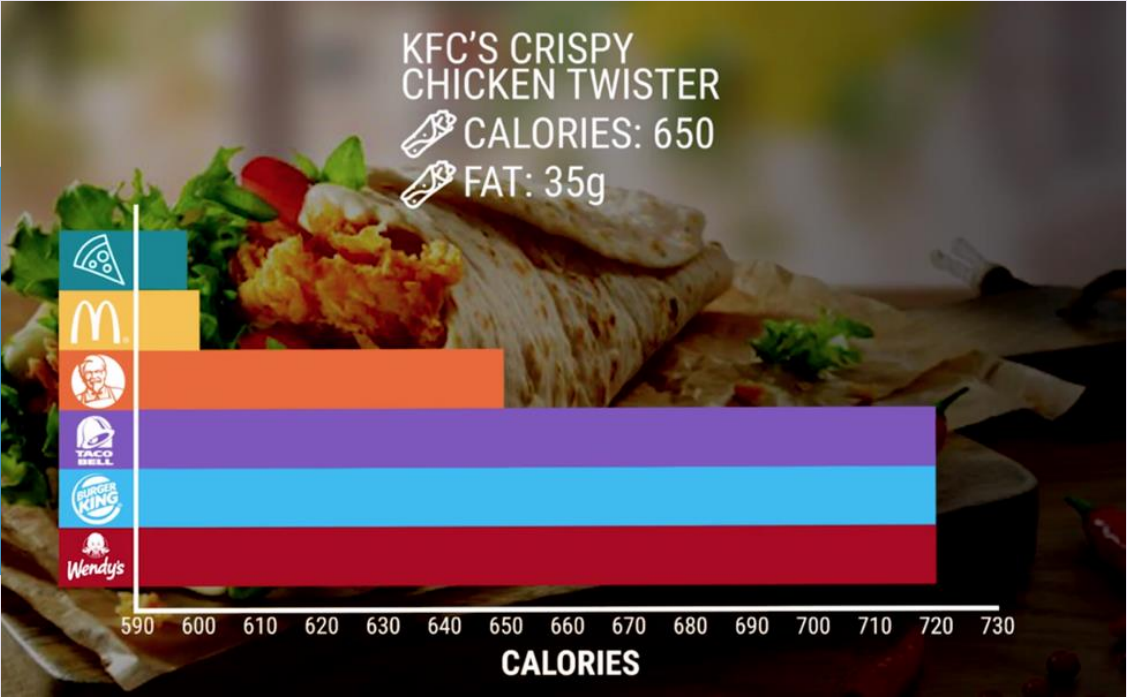
CONSUMER

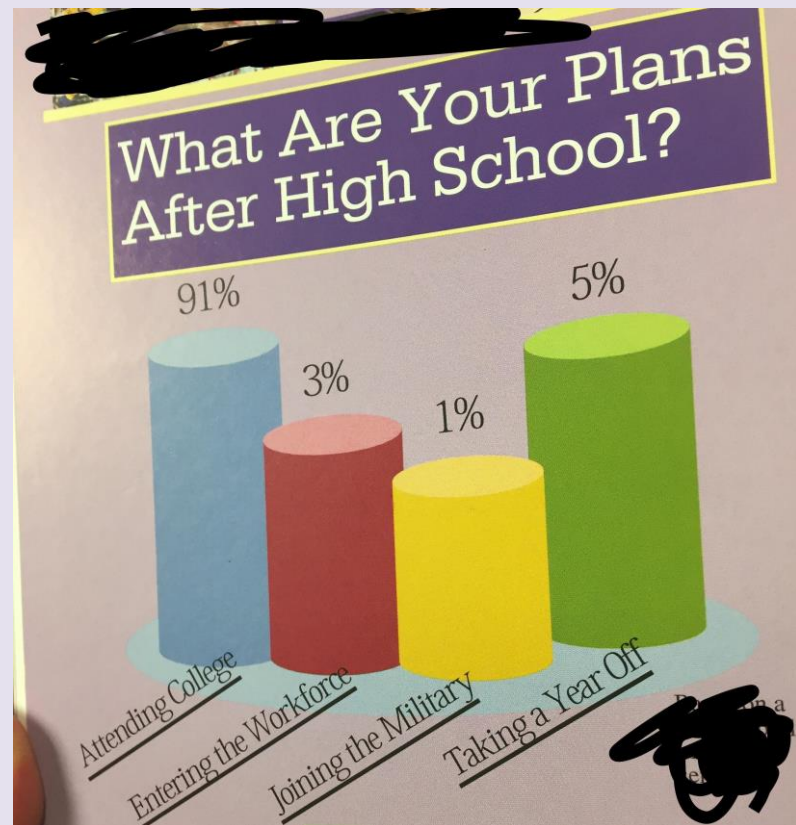
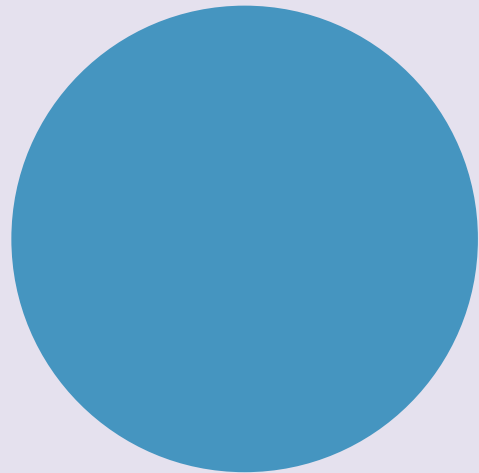
WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

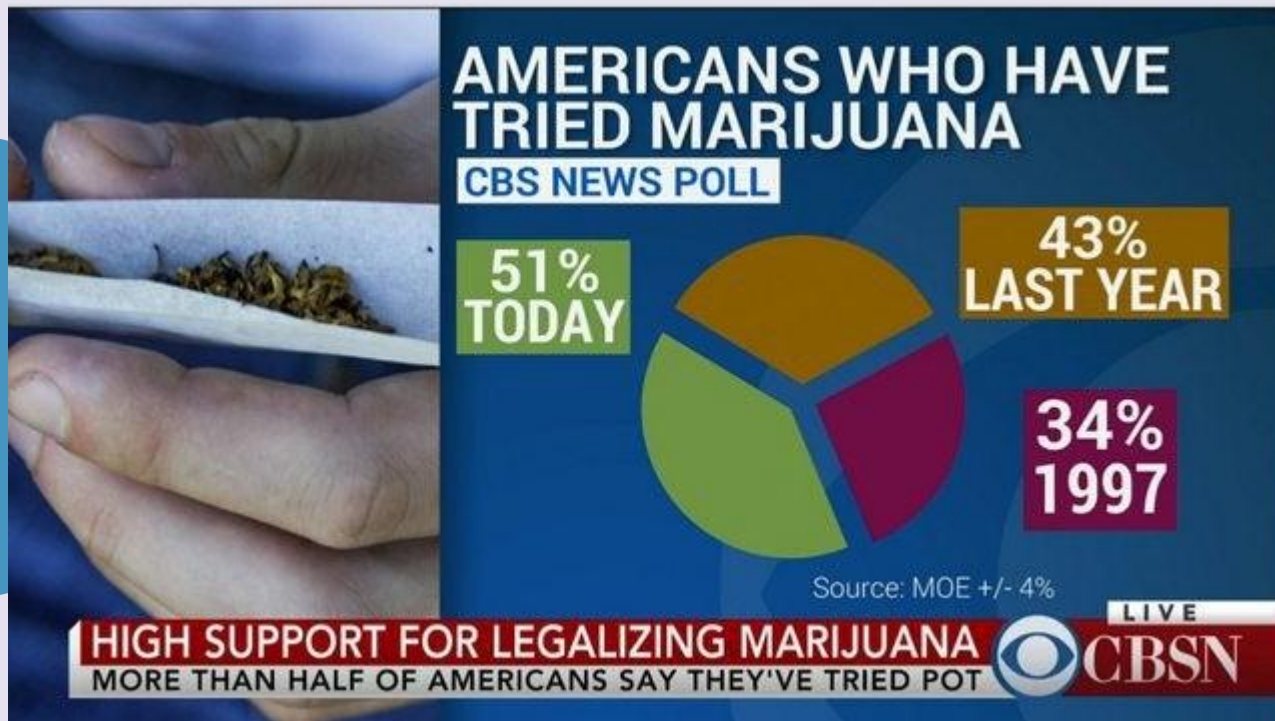
DOW 13008.68 ▼ 64.33

S&P 1379.32 ▼ 5.98

NASDAQ 2939.52 ▼ 6.32









Distribution of All TFBS Regions

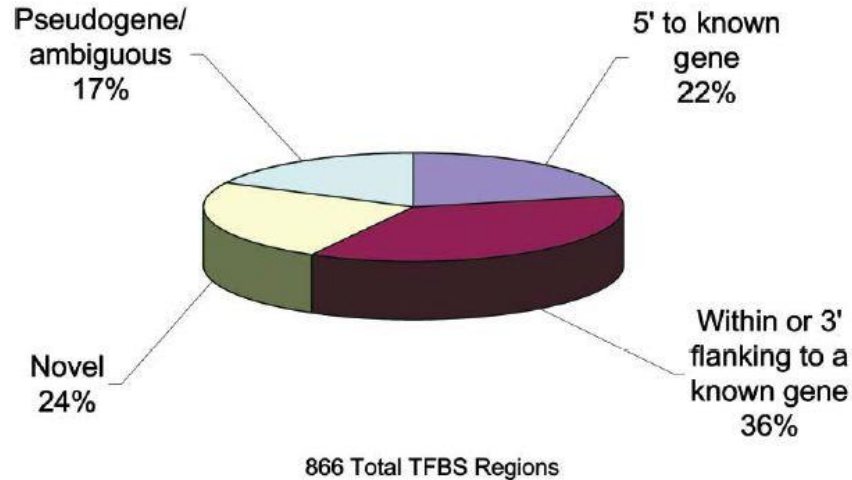


Figure 1. Classification of TFBS Regions

TFBS regions for Sp1, cMyc, and p53 were classified based upon proximity to annotations (RefSeq, Sanger hand-curated annotations, GenBank full-length mRNAs, and Ensembl predicted genes). The proximity was calculated from the center of each TFBS region. TFBS regions were classified as follows: within 5 kb of the 5' most exon of a gene, within 5 kb of the 3' terminal exon, or within a gene, novel or outside of any annotation, and pseudogene/ambiguous (TFBS overlapping or flanking pseudogene annotations, limited to chromosome 22, or TFBS regions falling into more than one of the above categories).

Cawley S, et al. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 116:499-509

VISUALIZAÇÃO



1. Uma má visualização gera ruído no processo de entendimento
2. Uma má visualização pode sugerir manipulação de dados e omissão de evidência
3. Uma boa visualização ajuda o pesquisador/profissional a **contar histórias com dados**





Gráficos contam histórias

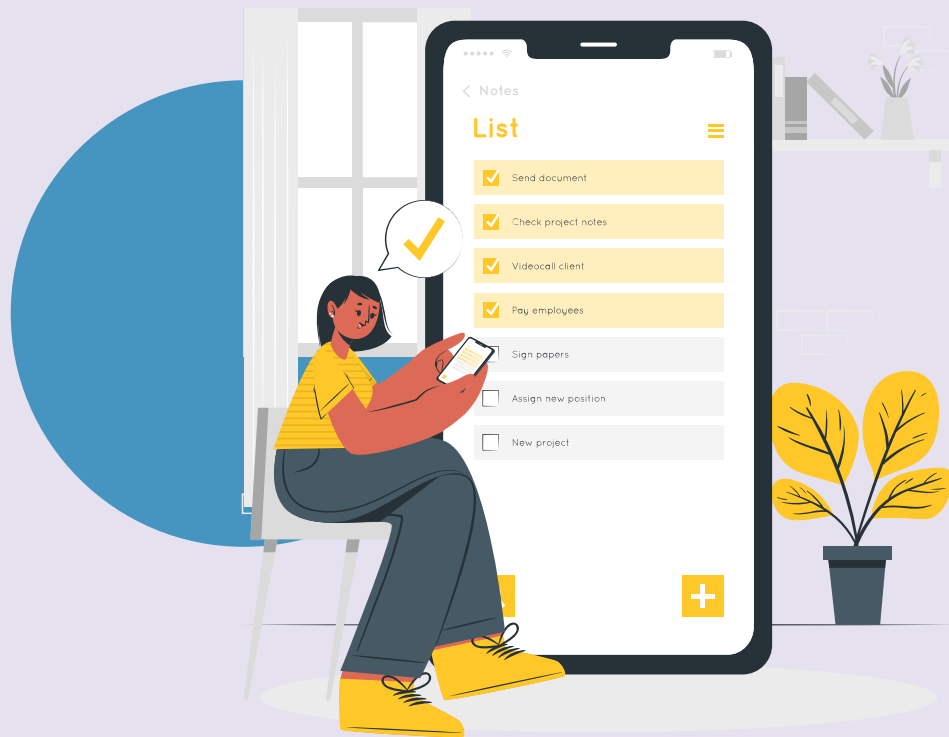


→

Gráficos contam a sua história



Dúvidas?



02

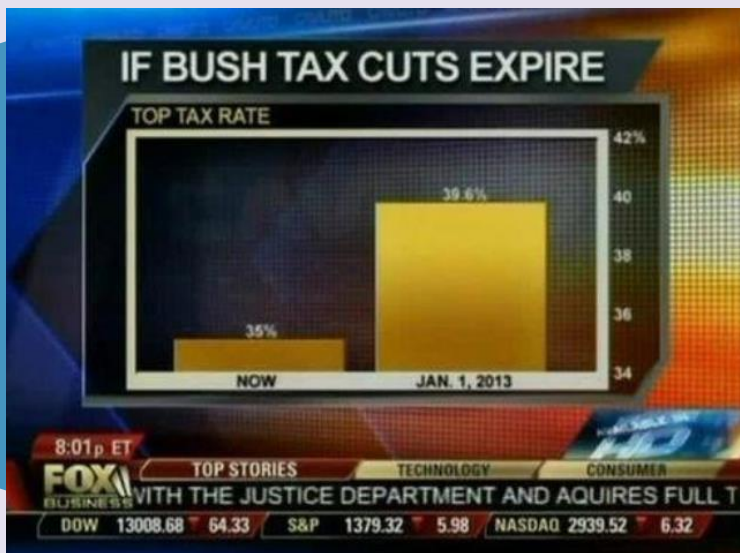
O que evitar

Não repita isso em casa



REGRA #1

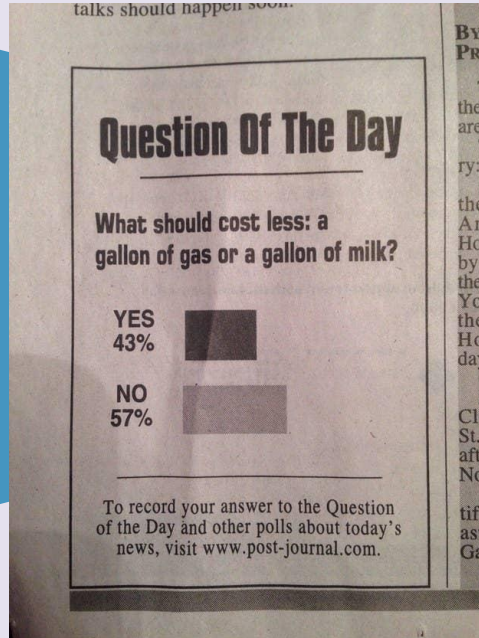
SEMPRE O DESIGN CERTO



1. Nunca omita a baseline
2. Nunca manipule o Eixo y
3. Nunca suprima evidência
4. Nunca use o tipo de gráfico errado
5. Nunca quebre convenções

REGRA #2

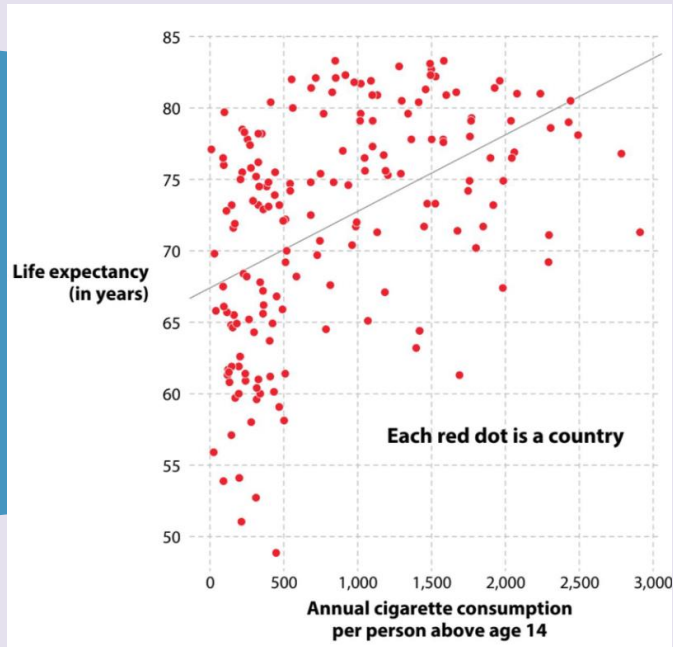
NÃO SEJA DÚBIO



1. Dados que não condizem com o que é plotado

REGRA #3

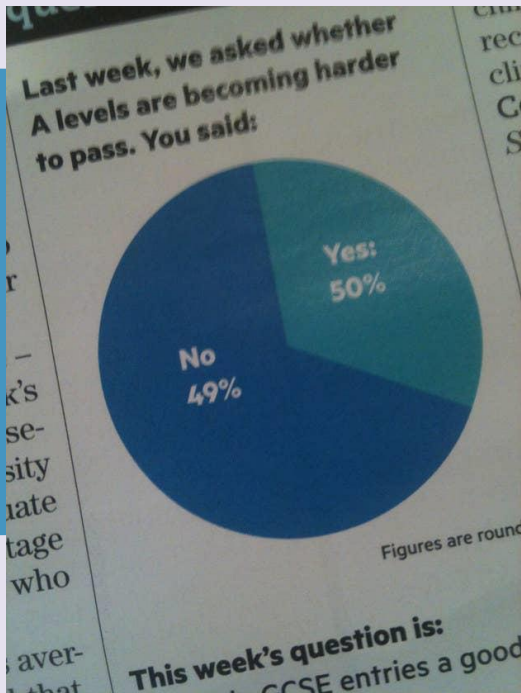
NÃO AFIRME O QUE VOCÊ NÃO PODE PROVAR



1. Nem sempre o padrão que se encontra é pela causa que se imputa

REGRA #4

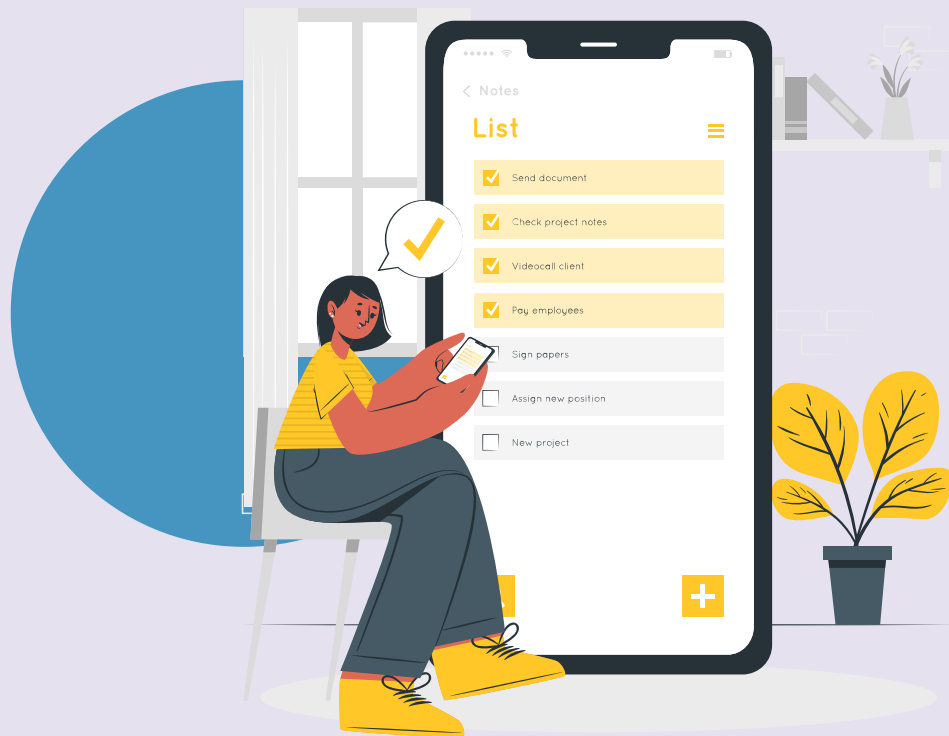
NÃO USE GRÁFICOS DE ÁREA



1. A mente humana foi criada para entender apenas 2 DIMENSÕES.



Dúvidas?



03

O que agregar

Mudando suas visões sobre o assunto

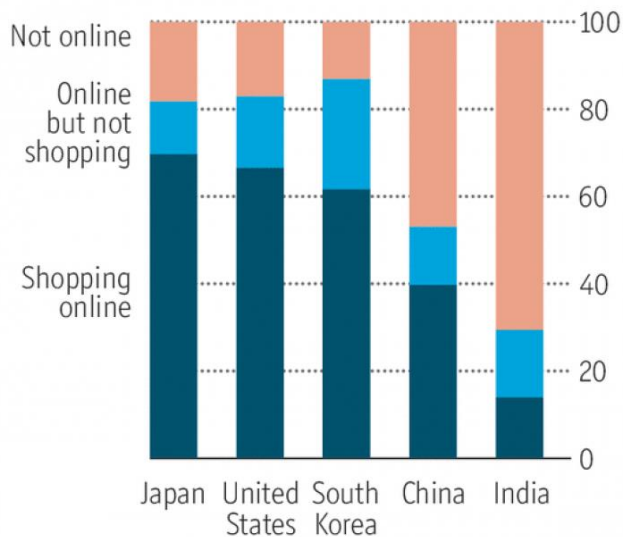


TIPS DE VISUALIZAÇÃO



Room for growth

Internet use, % of population, 2016



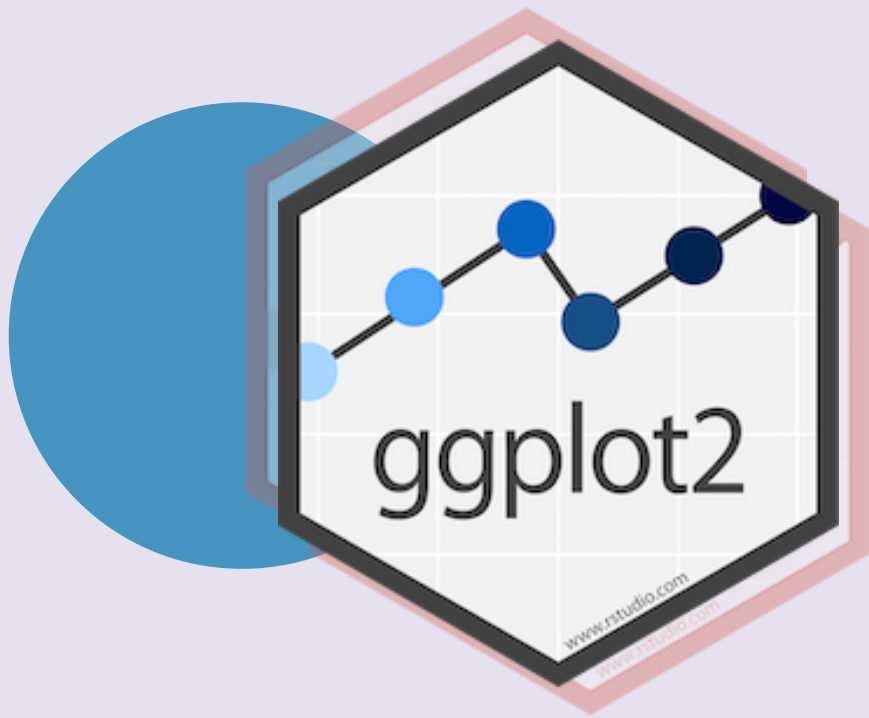
Source: eMarketer

Economist.com

1. A mente da pessoa ocidental procura padrões em um padrão de Z.
2. Cores são suas amigas. Use elas a seu favor para dar destaque ao que quer chamar a atenção
3. Favoreça gráficos de barra e pontos a outros tipos.
4. As vezes palavras fazem um trabalho melhor que um gráfico



Dúvidas?



04

ggplot2

Liberte o seu designer interior



ggplot2



1. Pacote para visualizações dentro do universo Tidyverse
2. Funciona num esquema similar ao de uma cebola / boneca russa
3. Tem suporte para mais de 50 tipos de gráficos

ggplot2



1. Barras (Crescimento e Padrões Categóricos)
2. Jitter (Crescimento e Padrões Numéricos)
3. Linhas (Crescimento e Padrões Numéricos)
4. Pontos (Crescimento e Padrões Numéricos)
5. Boxplot (Médias)



Dúvidas?

HORA DE PROGRAMAR

