



**INTRODUÇÃO À CIÊNCIA
DE DADOS**
Prof. Dr. Jodavid Ferreira



QUEM SOU

FORMAÇÃO

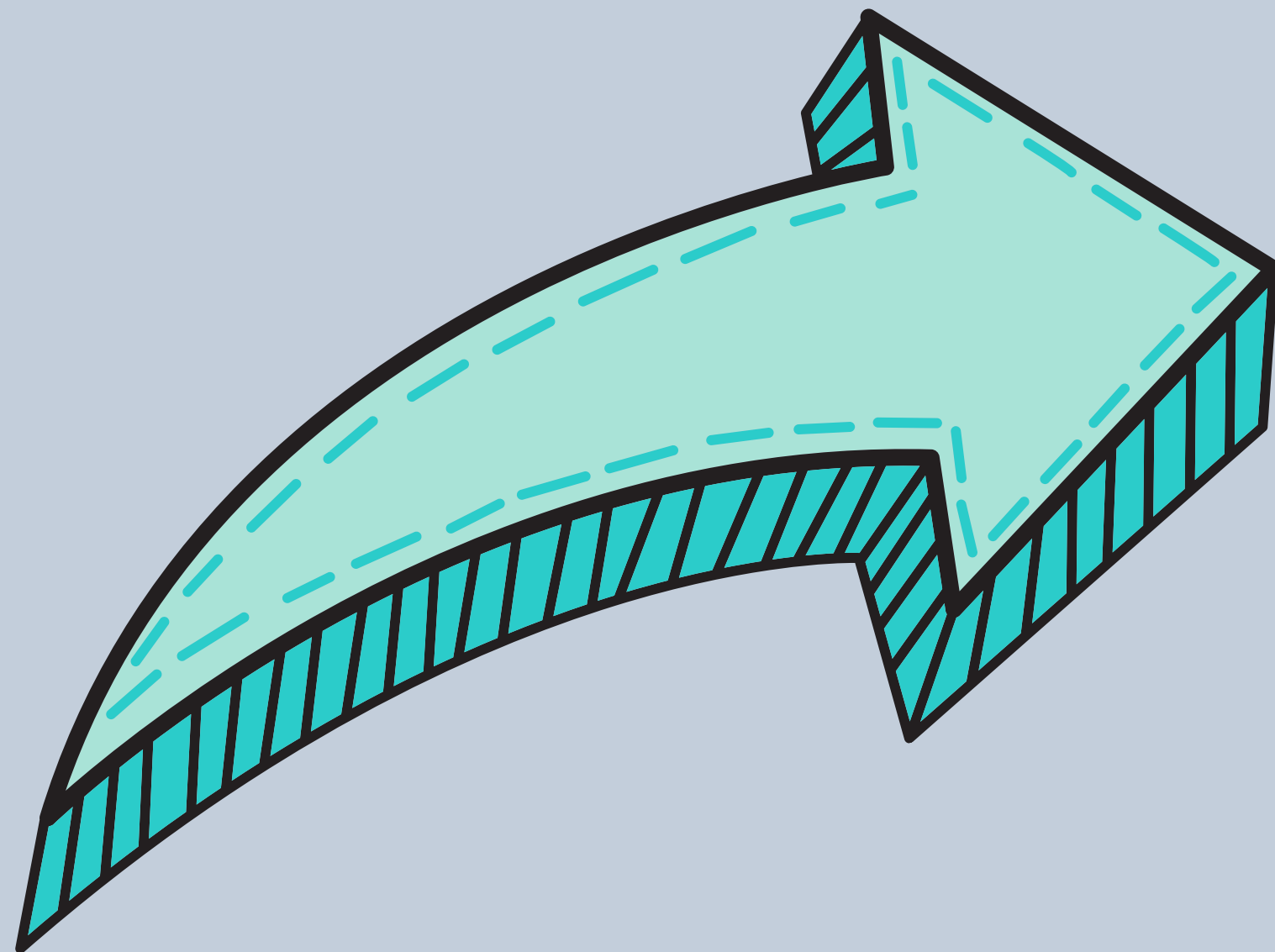
- GRADUAÇÃO EM ESTATÍSTICA - **UFPB** - 2015
- MESTRADO EM ESTATÍSTICA - **UFPE** - 2017
- DOUTORADO EM ESTATÍSTICA - **UFPE** - 2021
- PÓS-DOUTORADO - **MDS - UFPB** - 2024

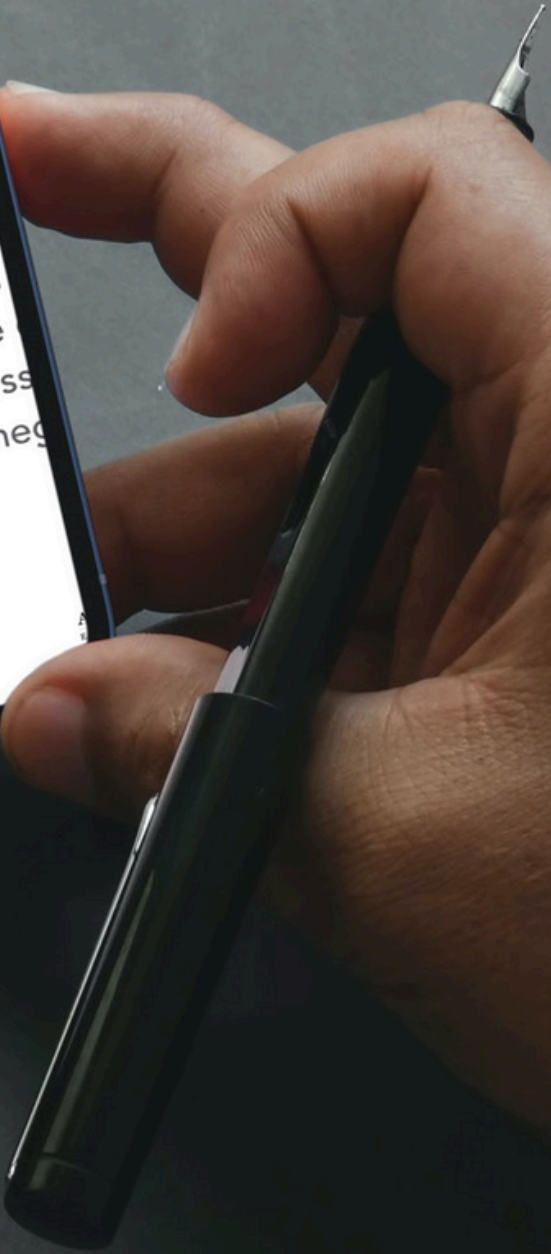
EXPERIÊNCIA

- CIENTISTA DE DADOS NA **HARTB GROUP**

ATUALMENTE

- PROFESSOR DEPARTAMENTO ESTATÍSTICA - **UFPE**





PRIMEIRO CONTATO

UM AMERICANO VÊ EM MÉDIA 16 MIL ANÚNCIOS, LOGOS E ETIQUETAS AO LONGO DE UM DIA¹;



DHARMA SINGH KHALSA, DOUTOR EM MEDICINA, COM CAMERON STAUTJ, BRAIN LONGEVITY: THE BREAKTHROUGH MEDICAL PROGRAM THAT IMPROVES YOUR MIND AND MEMORY. (NEW YORK, NY: GRAND CENTRAL PUBLISHING, 1999)

IMPORTÂNCIA DO CONTEXTO



- Quem é seu público?
- O que você precisa que ele saiba?

NÃO SE RESUME **AOS GRÁFICOS**, MAS TAMBÉM A
FORMA COMO **ALGO** É APRESENTADO

VOCÊ SABE A DIFERENÇA ENTRE
ANÁLISE EXPLORATÓRIA E
ANÁLISE EXPLANATÓRIA?

#STORYTELLING



ANÁLISE EXPLORATÓRIA

É compreender os dados e descobrir o que pode ser digno de nota ou interessante a destacar para outras pessoas.

ANÁLISE EXPLANATÓRIA

Quando pretendemos comunicar nossa análise para o público, significando que temos algo específico para explicar, uma história para contar.



CONTEXTO: APRESENTAÇÃO DOS RESULTADOS

- GRÁFICOS -



QUEM SOU

FORMAÇÃO

- GRADUAÇÃO EM ESTATÍSTICA - **UFPB** - 2015
- MESTRADO EM ESTATÍSTICA - **UFPE** - 2017
- DOUTORADO EM ESTATÍSTICA - **UFPE** - 2021
- PÓS-DOUTORADO - **MDS - UFPB** - 2024

EXPERIÊNCIA

- CIENTISTA DE DADOS NA **HARTB GROUP** - 2021 /2023

ATUALMENTE

- PROFESSOR DEPARTAMENTO ESTATÍSTICA - **UFPE**

JODAVID FERREIRA

QUEM SOU



2015

Graduação em
Estatística - UFPB



2017

Mestrado em
Estatística - UFPE



2021.1

Doutorado em
Estatística - UFPE



2021.1

Head de Ciência
de Dados




2024

Pós-doutorado no
MDS - UFPB



ATUAL


Professor UFPE



2015
Graduação em
Estatística
-UFPB-



2017
Mestrado em
Estatística
-UFPE-



2021.1
Doutorado em
Estatística
-UFPE-



2021.1
Head de Ciência
de Dados
-HARTB -

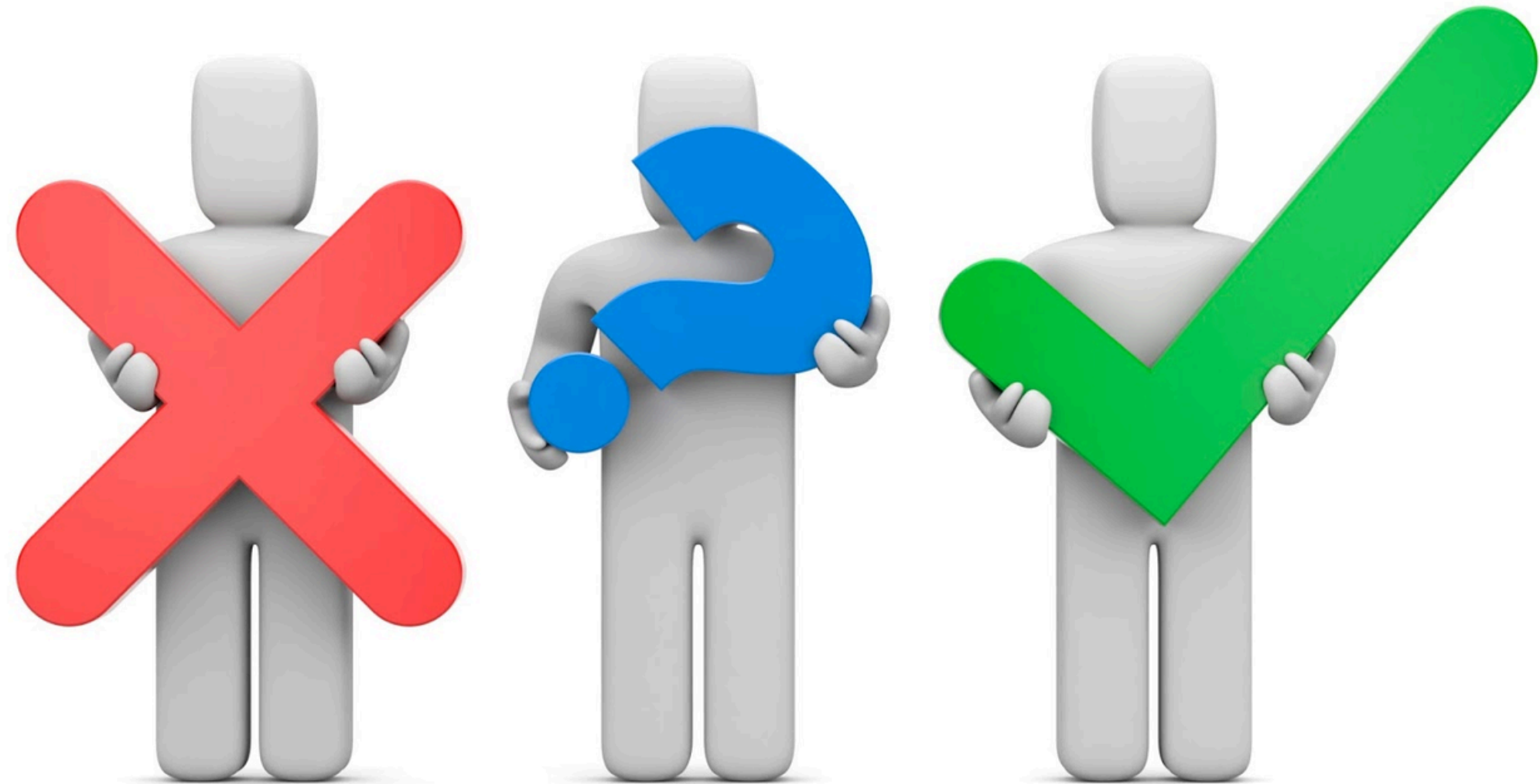


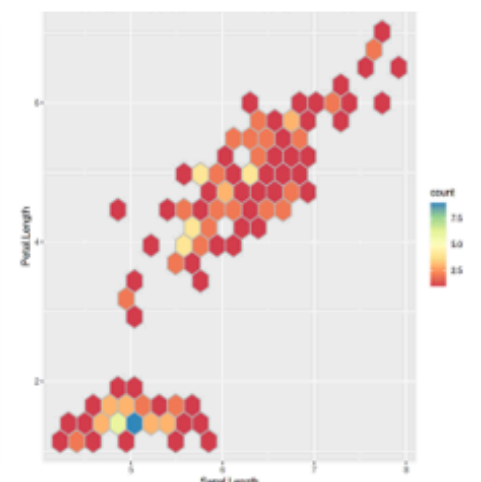
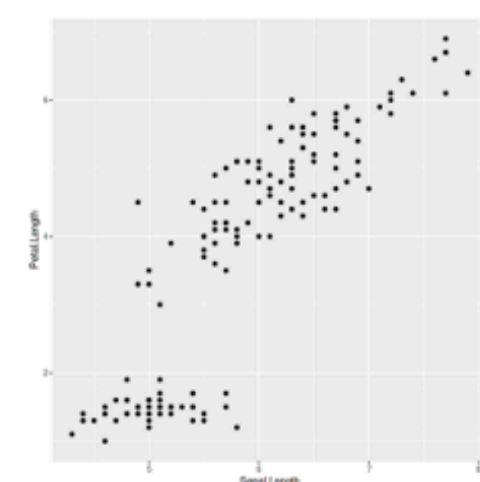
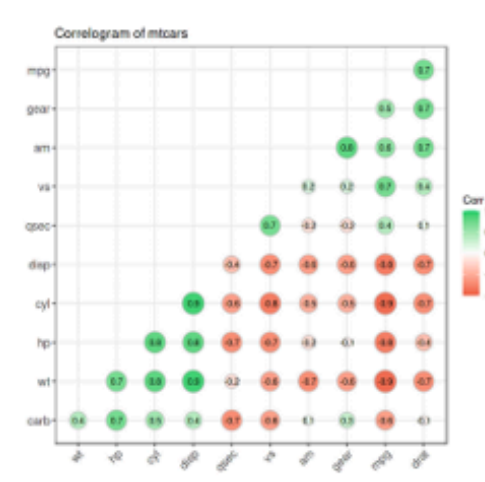
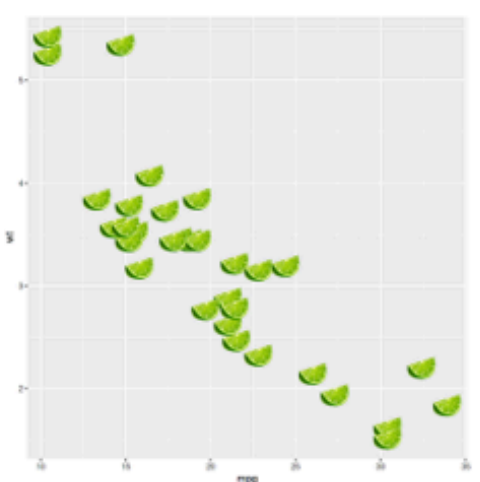
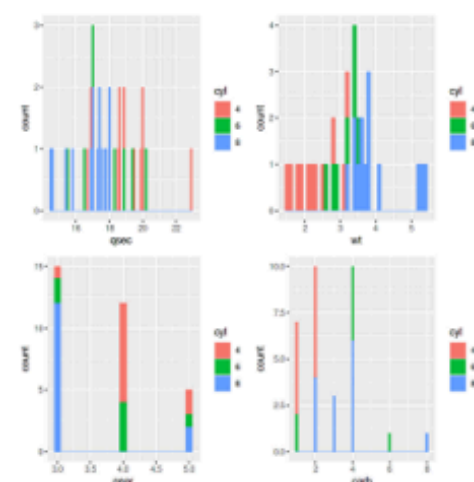
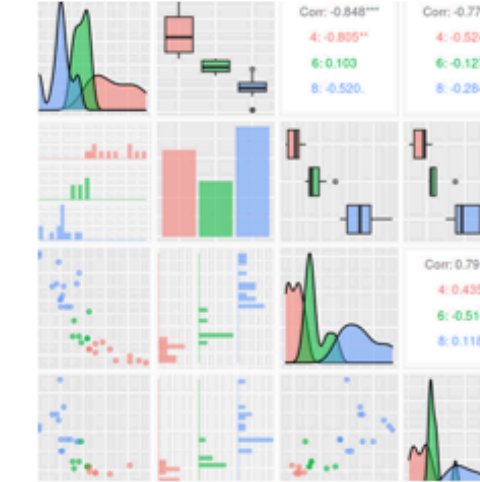
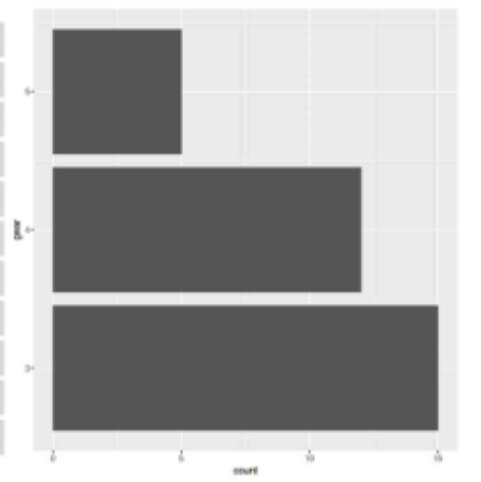
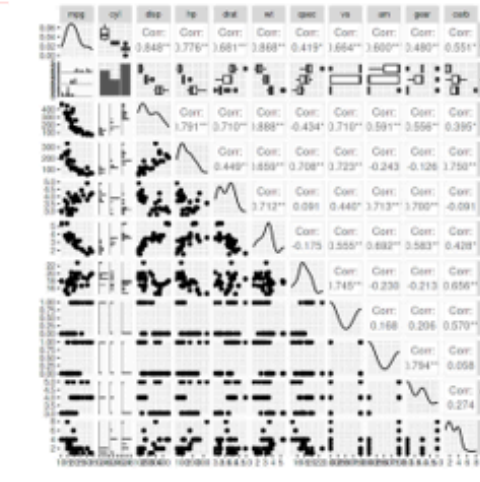
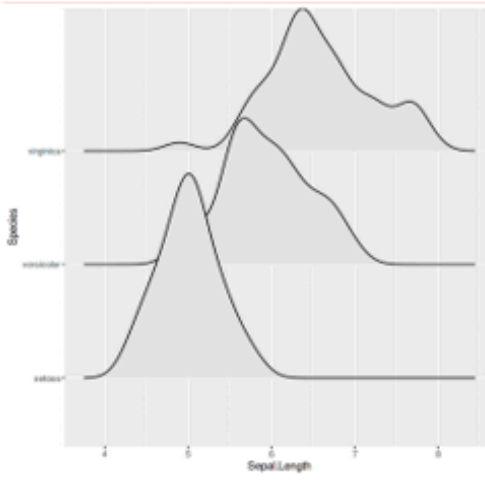
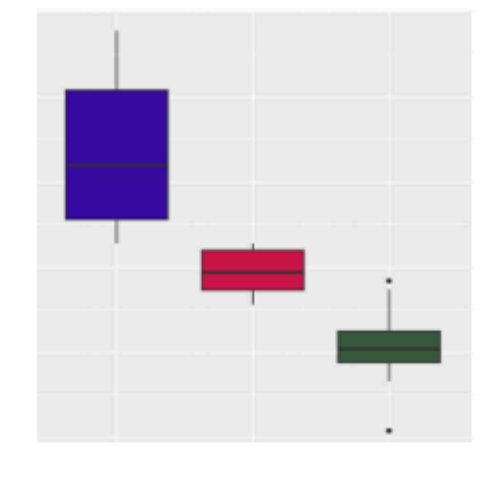
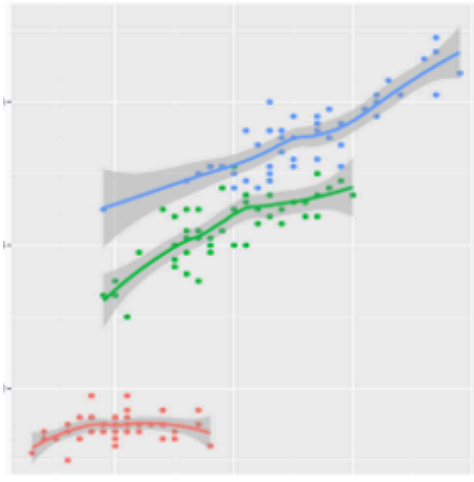
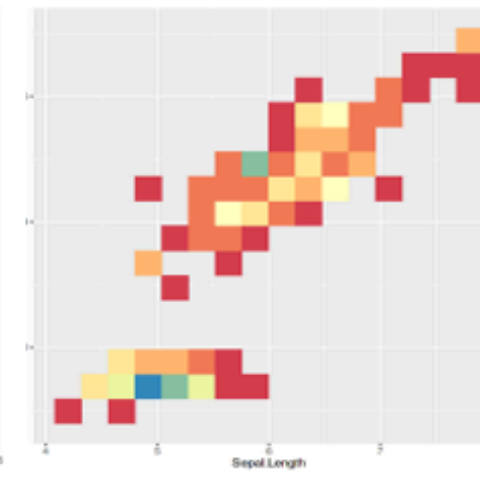
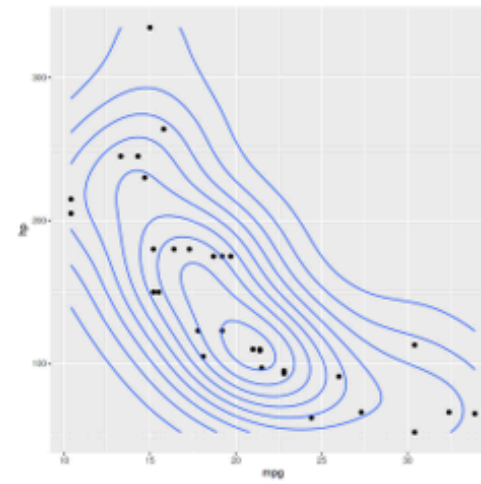
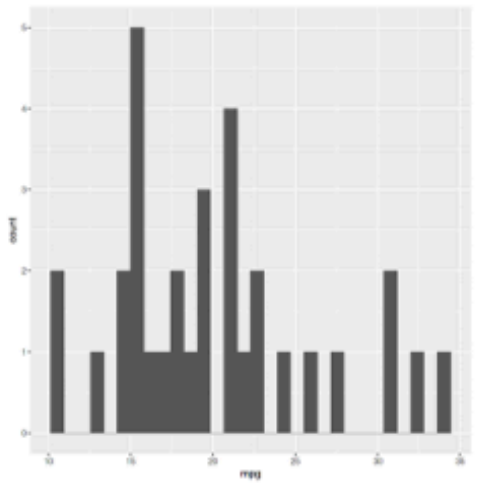
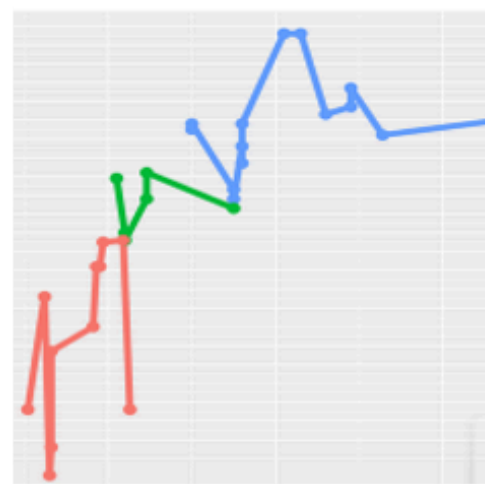
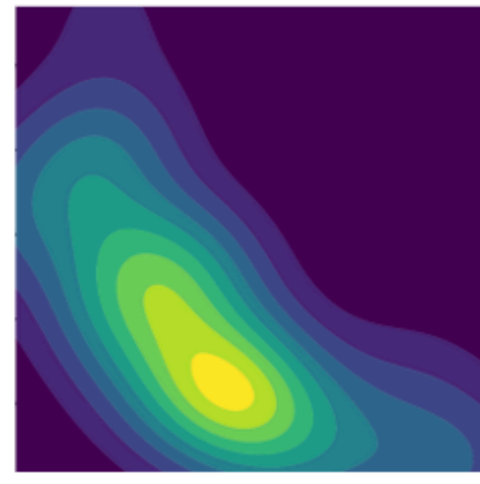
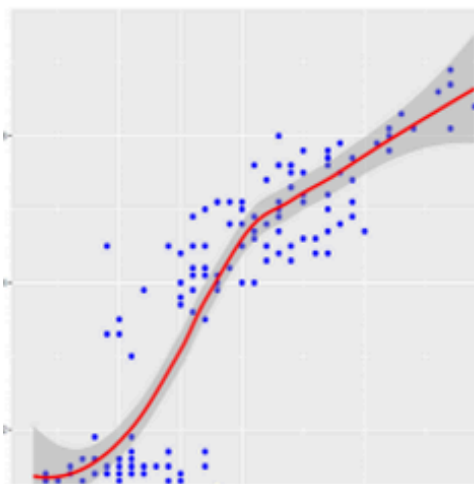
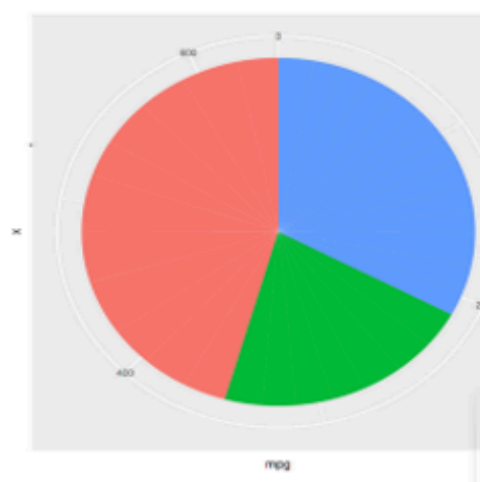
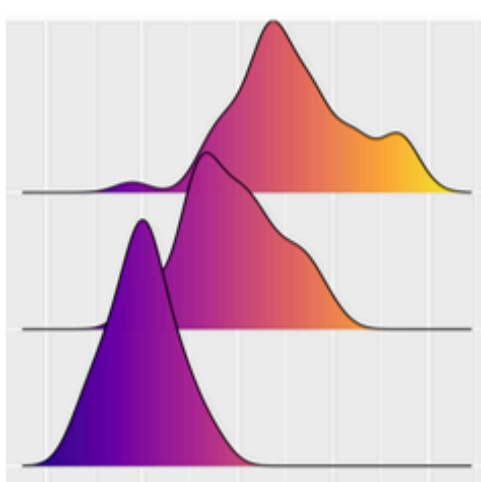
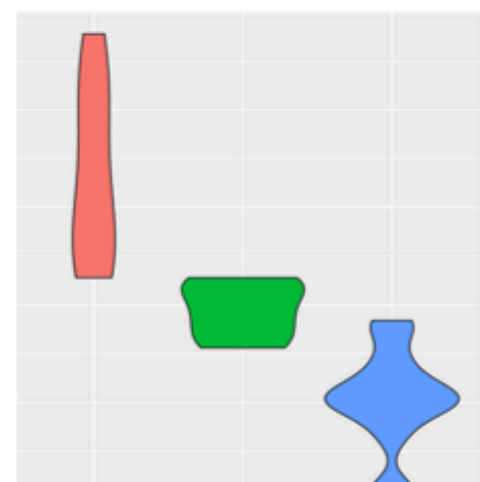
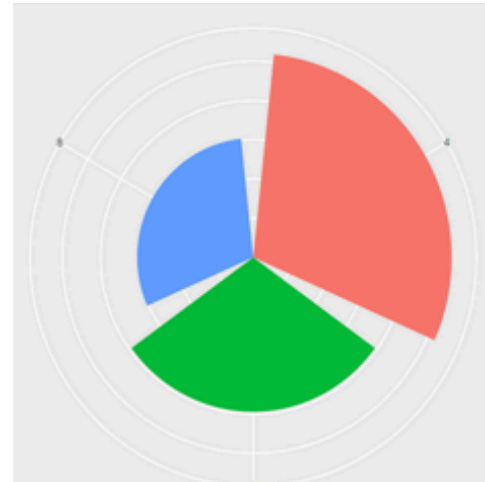
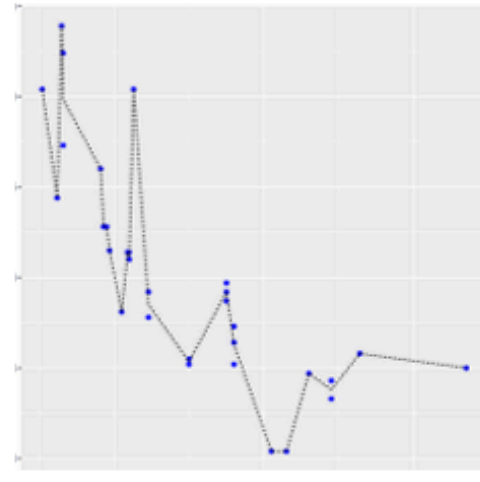
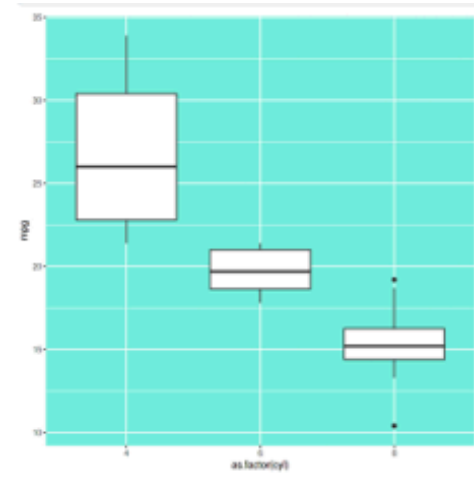
2024
Pós-doutorado no
MDS
-UFPB-



ATUAL
Professor UFPE
-UFPE -

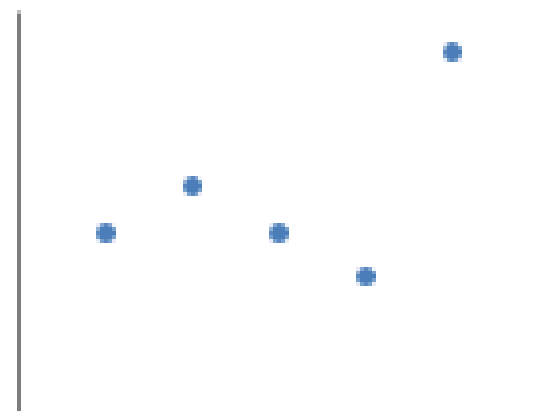
SERÁ QUE UTILIZAR O TIPO DE
GRÁFICO ADEQUADO É O
SUFICIENTE?!



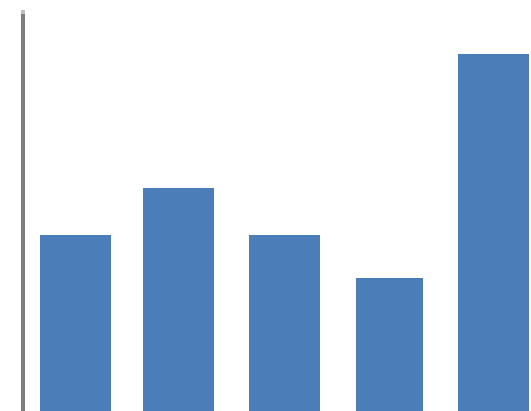


91

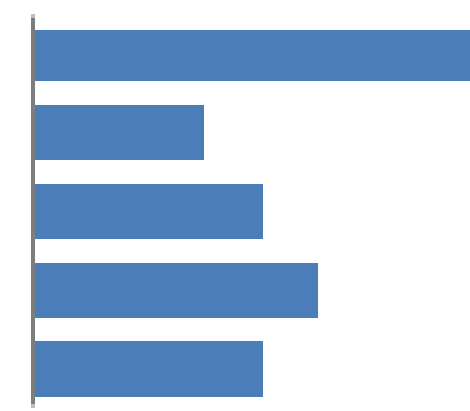
Simple text



Scatterplot



Vertical bar



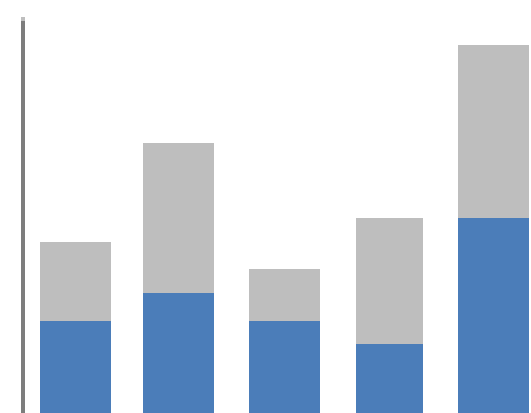
Horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

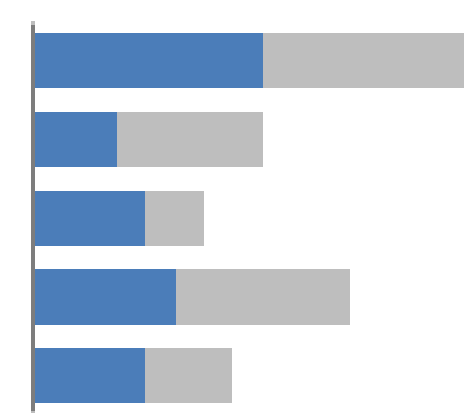
Table



Line



Stacked vertical bar

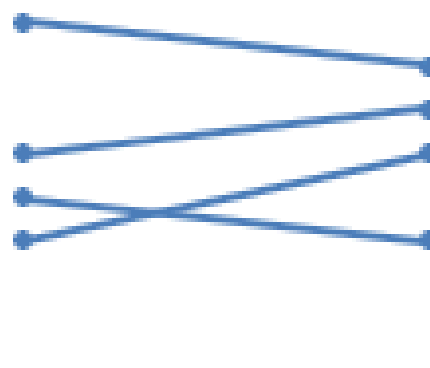


Stacked horizontal bar

BARRAS HORIZONTAIS EMPILHADAS

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap



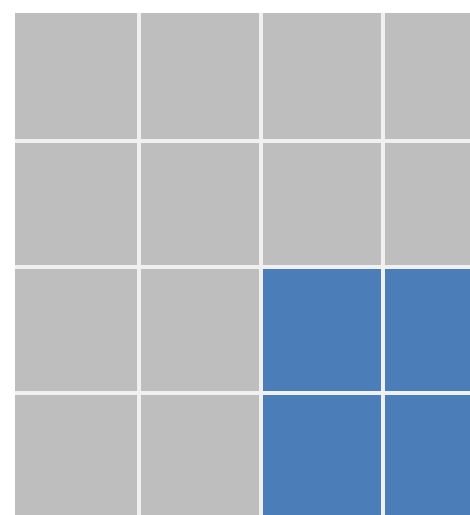
Slopegraph

GRÁFICO DE INCLINAÇÃO



Waterfall

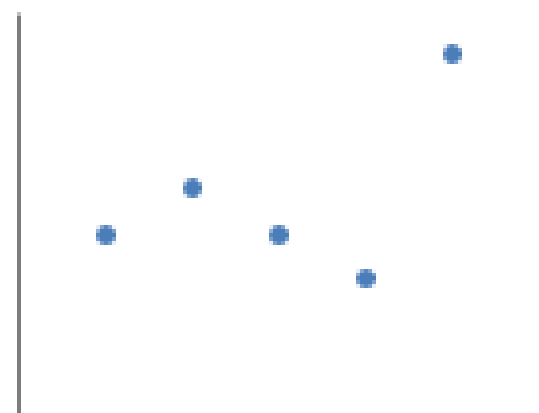
CASCATA



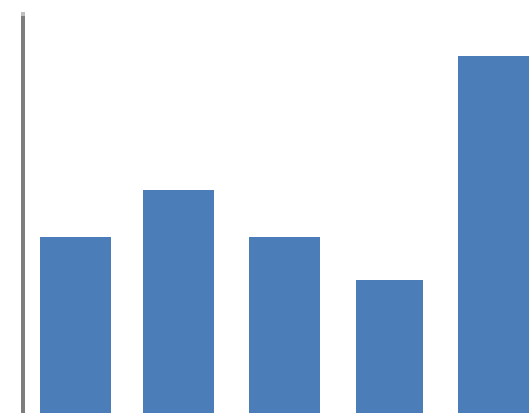
Square area

91

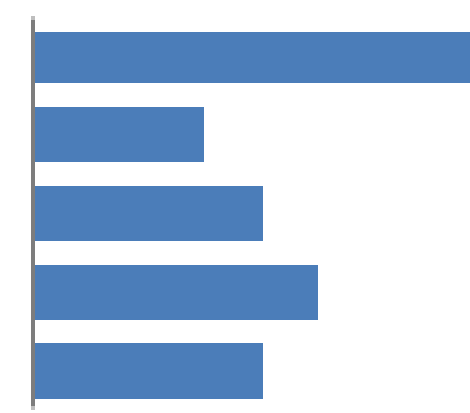
Simple text



Scatterplot



Vertical bar



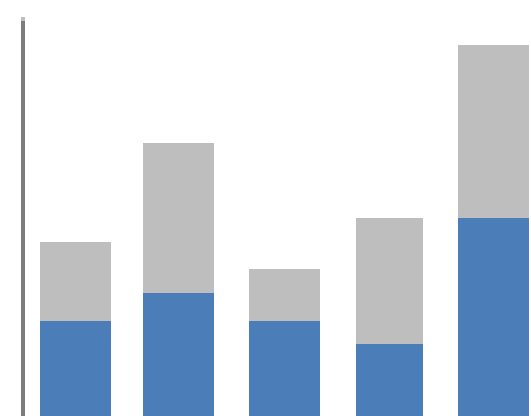
Horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

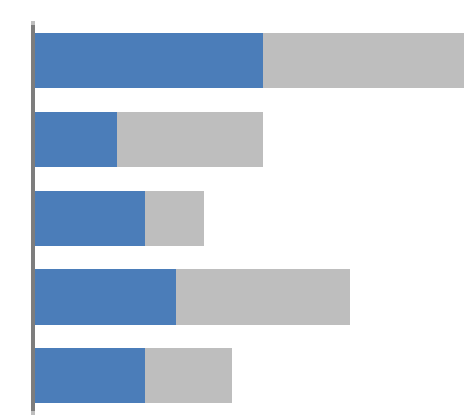
Table



Line



Stacked vertical bar

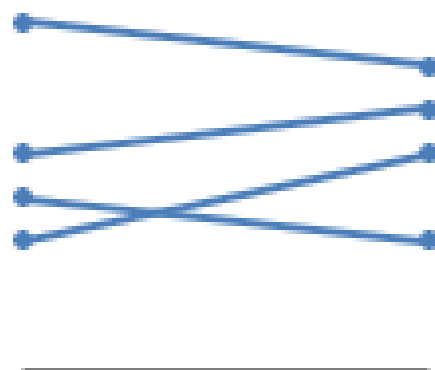


Stacked horizontal bar

BARRAS HORIZONTAIS EMPILHADAS

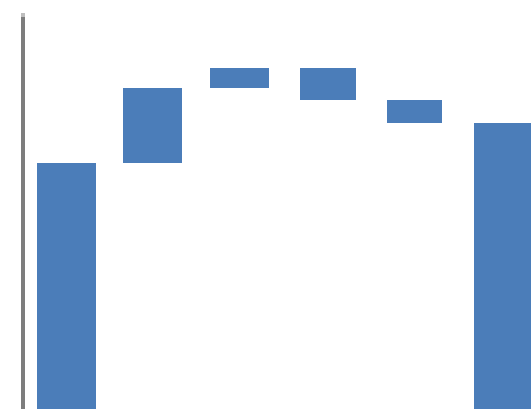
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap



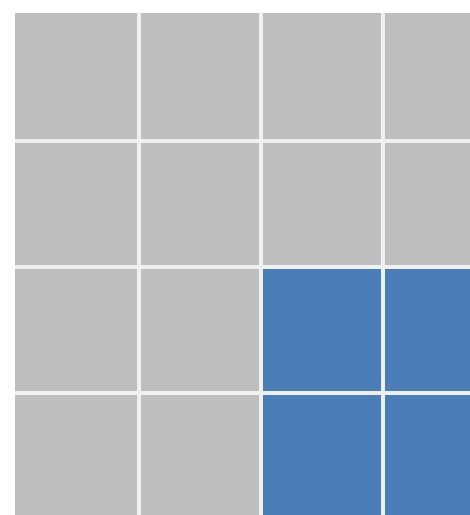
Slopegraph

GRÁFICO DE INCLINAÇÃO



Waterfall

CASCATA



Square area

TABELAS (TABLES)

Note como os dados se sobressaem mais que os componentes estruturais da tabela na segunda e terceiras iterações.

As bordas devem ser utilizadas para melhorar a legibilidade da tabela. São os dados que devem ser salientados, não as bordas.

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

MAPAS DE CALOR (HEATMAP)

Usar mapas de calor é um jeito de misturar detalhes em uma tabela usando pistas visuais.

O mapa de calor é um modo de visualiza dados em formato tabular, no qual, em vez dos números, você utiliza células coloridas que transmitem a magnitude relativa dos números.

Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

LOW-HIGH

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

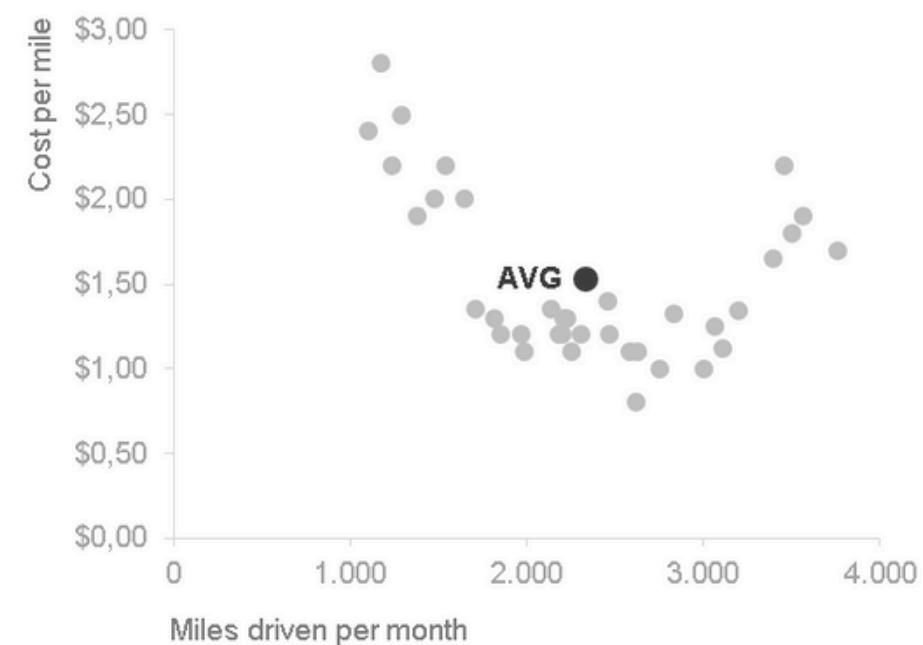
GRÁFICO DE DISPERSÃO (SCATTERPLOT)

O gráfico de dispersão podem ser úteis para mostrar a relação entre duas variáveis, pois eles permitem codificar dados simultaneamente em um eixo x horizontal e um eixo y vertical para ver se e qual relação existe.

Considere o exemplo de gerenciamento de **frota de ônibus** e desejamos entender a relação entre as **milhas dirigidas** e o **custo por milha**.

Se desejarmos focar nos casos nos quais o custo por milha está **acima da média**, faça um gráfico de dispersão ligeiramente modificado.

Cost per mile by miles driven



Cost per mile by miles driven

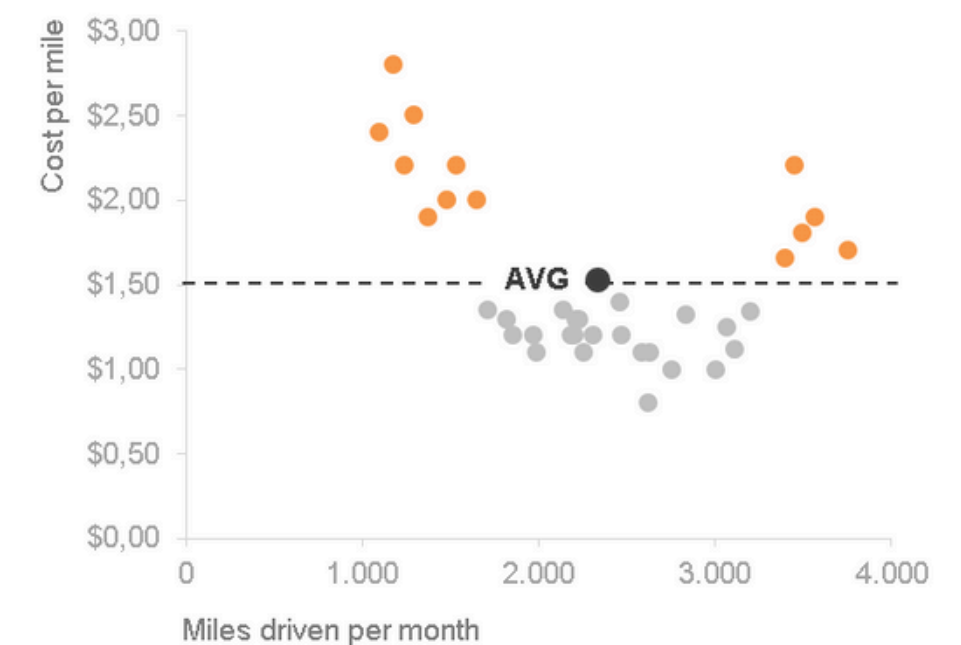
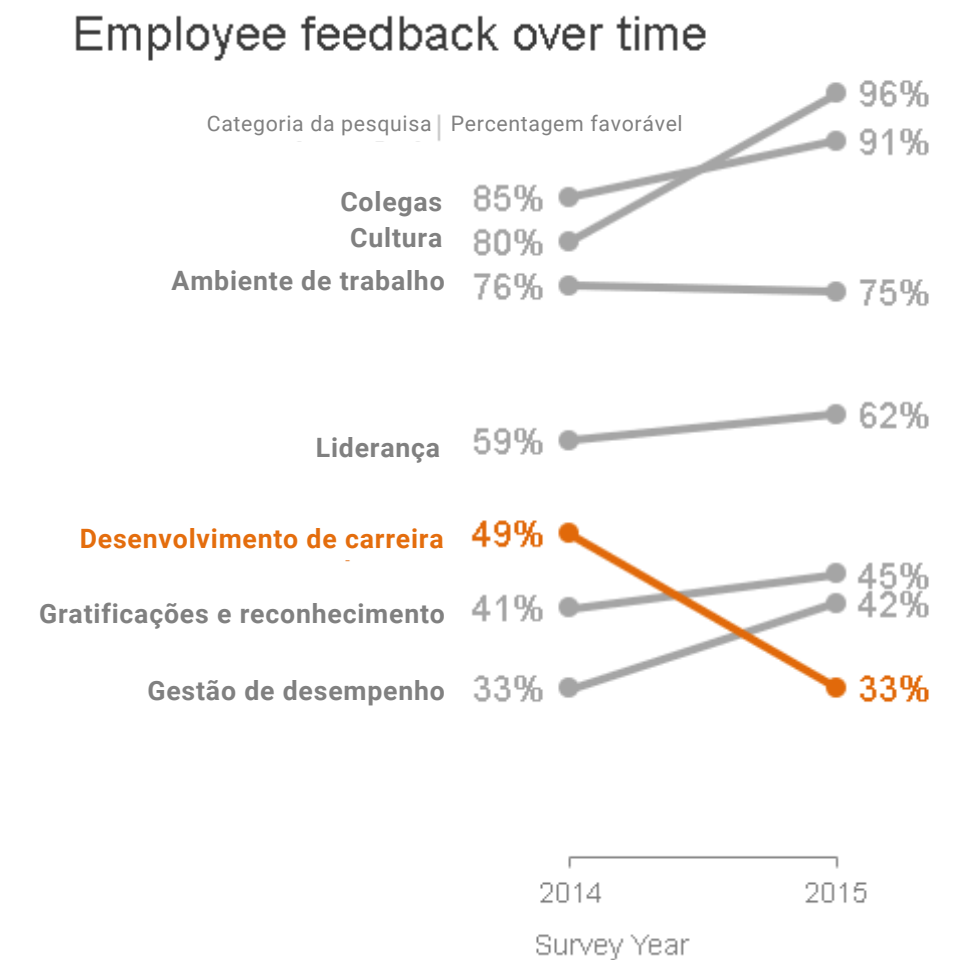
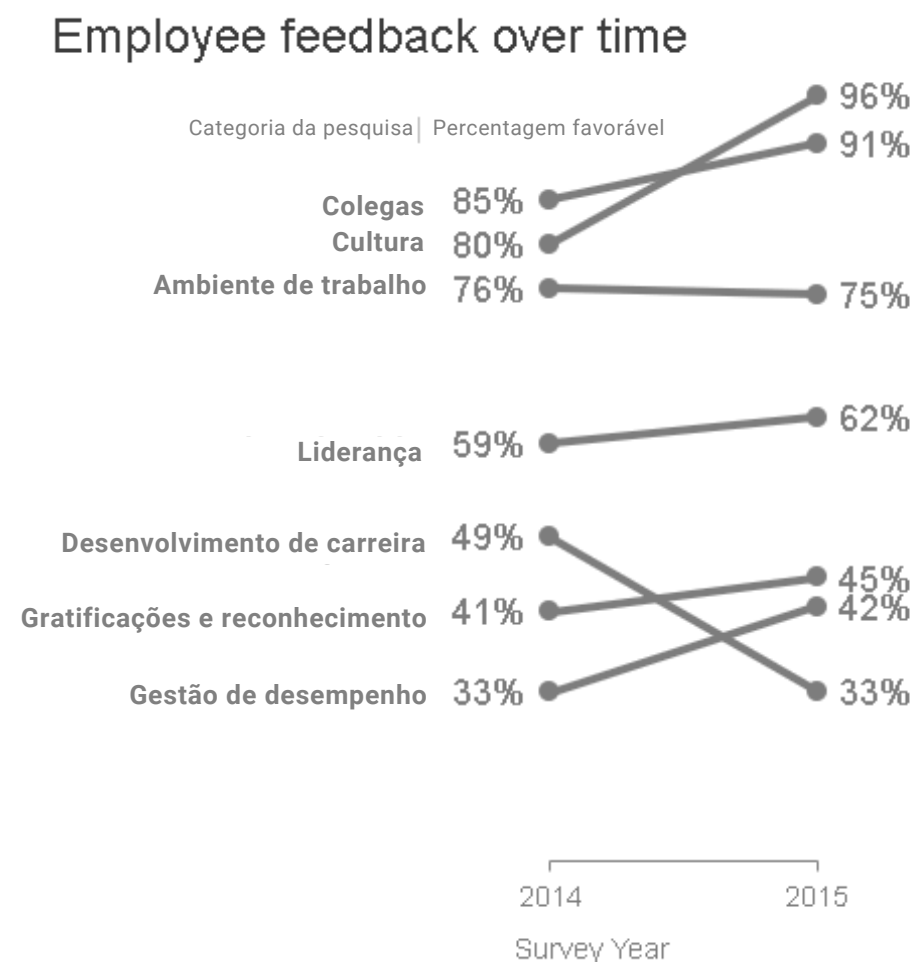


GRÁFICO DE INCLINAÇÃO (SLOPEGRAPH)

Gráficos de inclinação são úteis quando você tem dois períodos de tempo ou pontos de comparação e quer mostrar o comportamento de categorias nos dois tempos.

Se existem muitas linhas se sobrepondo, talvez o gráfico não se torne ideal, mas você pode enfatizar uma série por vez com sucesso.

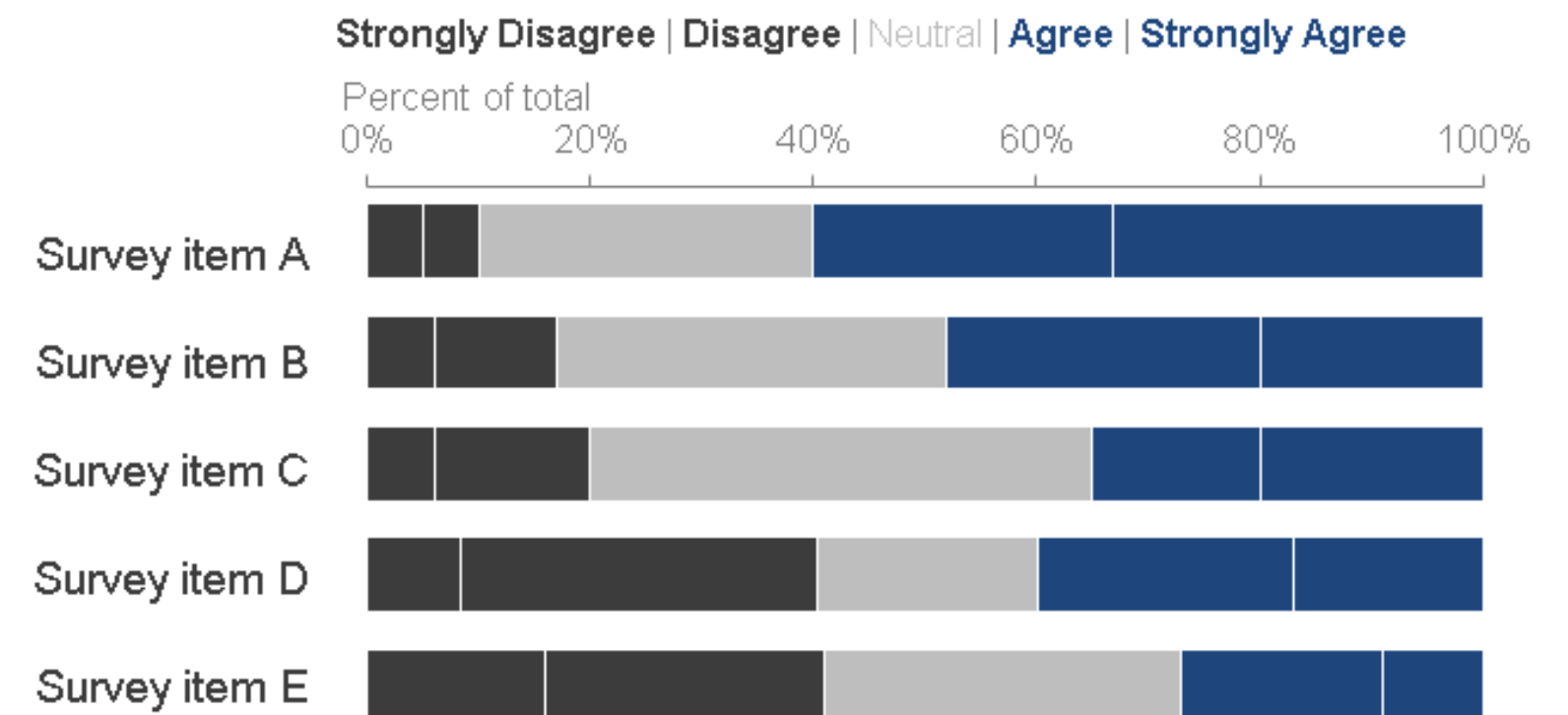


BARRAS HORIZONTAIS EMPILHADAS (STACKED HORIZONTAL BAR)

Esses gráficos podem ser usados para mostrar os totais em diferentes categorias, mas também dar uma ideia das partes subcomponentes. Eles podem ser estruturados de forma a mostrar valores absolutos ou a soma até 100%.

Essa estratégia pode funcionar bem para visualizar dados de pesquisa coletados ao longo de uma escala de Likert.

Survey results



A SER EVITADO

- Gráficos de pizza,
- Gráficos de rosca,
- Gráficos em 3D
- Eixos secundários



GRÁFICO DE SETOR, PIZZA...

O olho humano não consegue atribuir corretamente valores quantitativos no espaço bidimensional.

Supplier Market Share

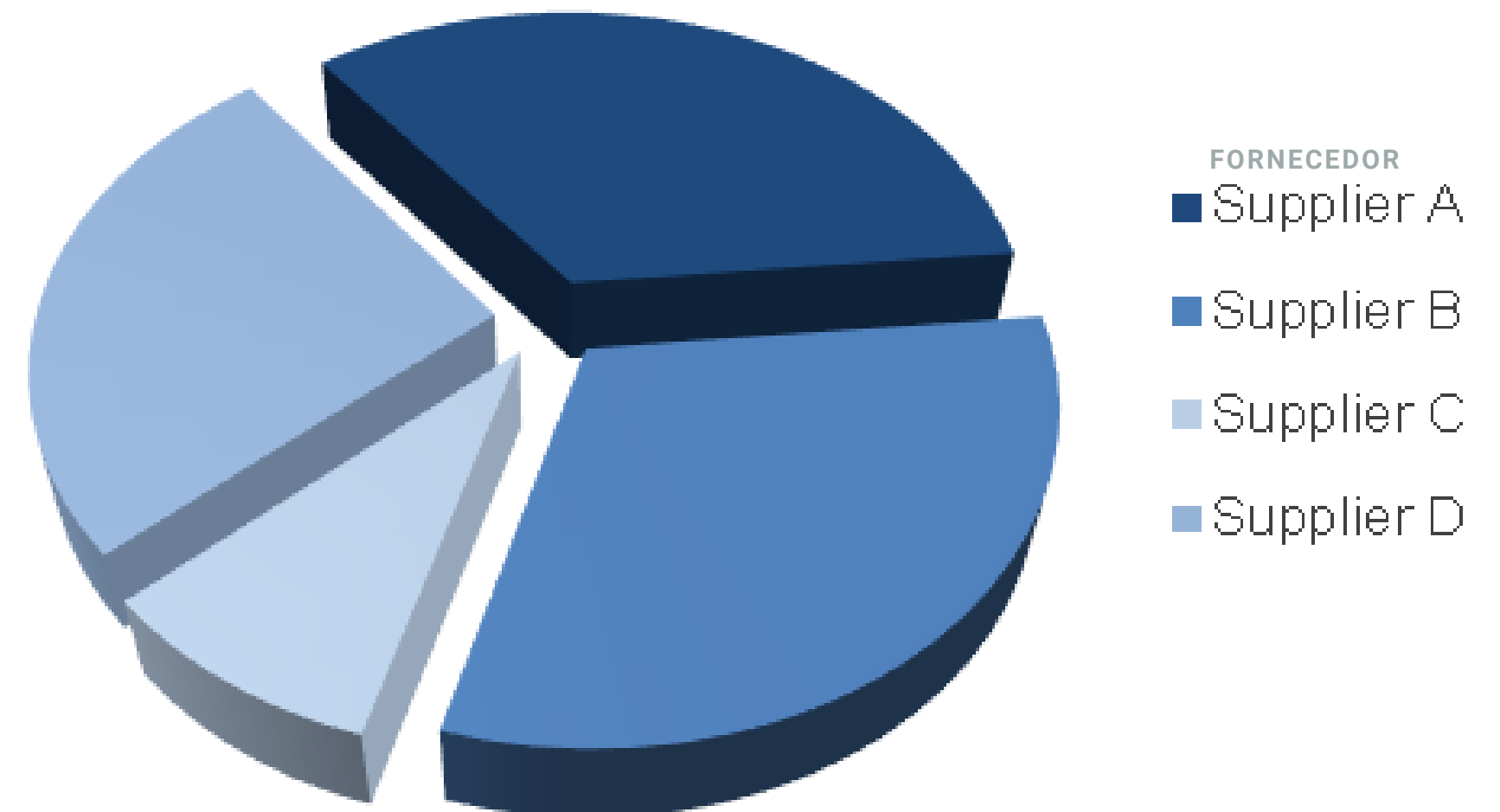


GRÁFICO DE SETOR, PIZZA...

O olho humano não consegue atribuir corretamente valores quantitativos no espaço bidimensional.

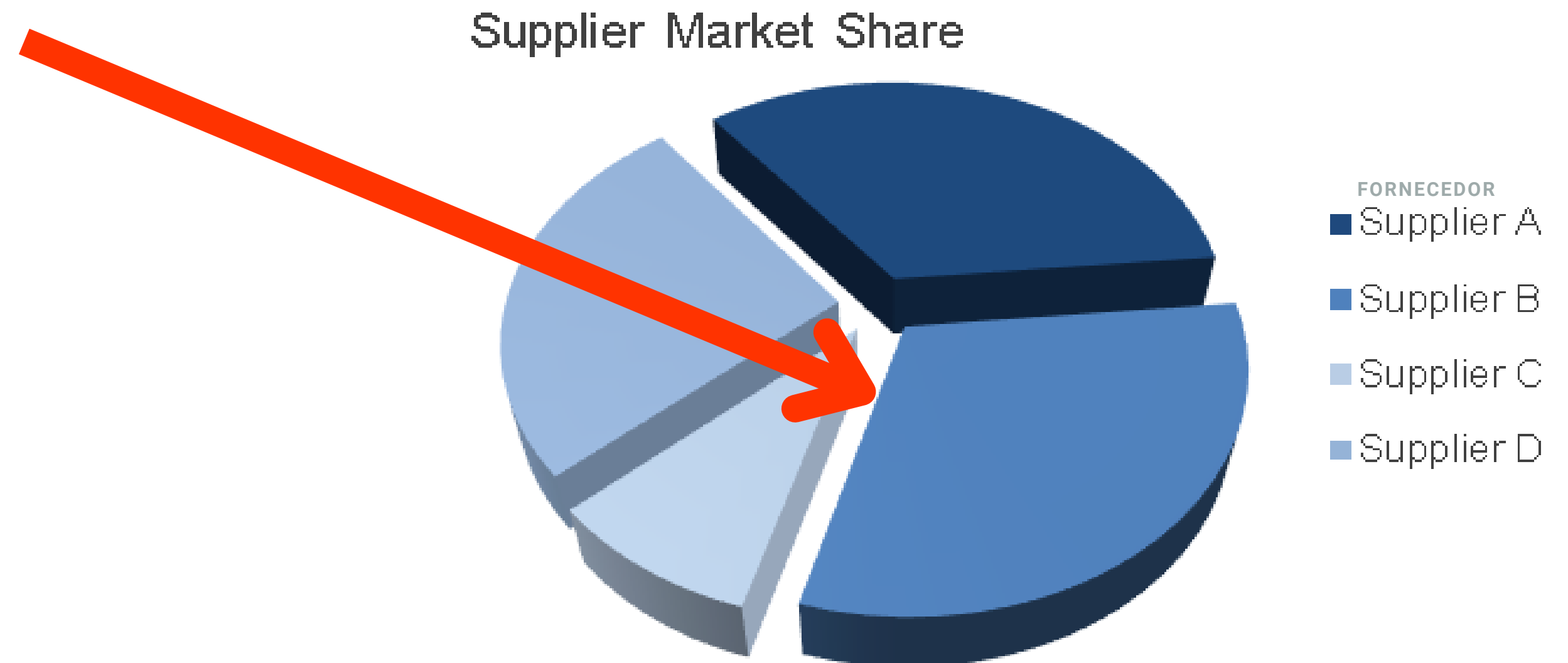


GRÁFICO DE SETOR, PIZZA...

O olho humano não consegue atribuir corretamente valores quantitativos no espaço bidimensional.

Quando segmentos possuem tamanhos parecidos, é difícil dizer qual o maior.

Supplier Market Share

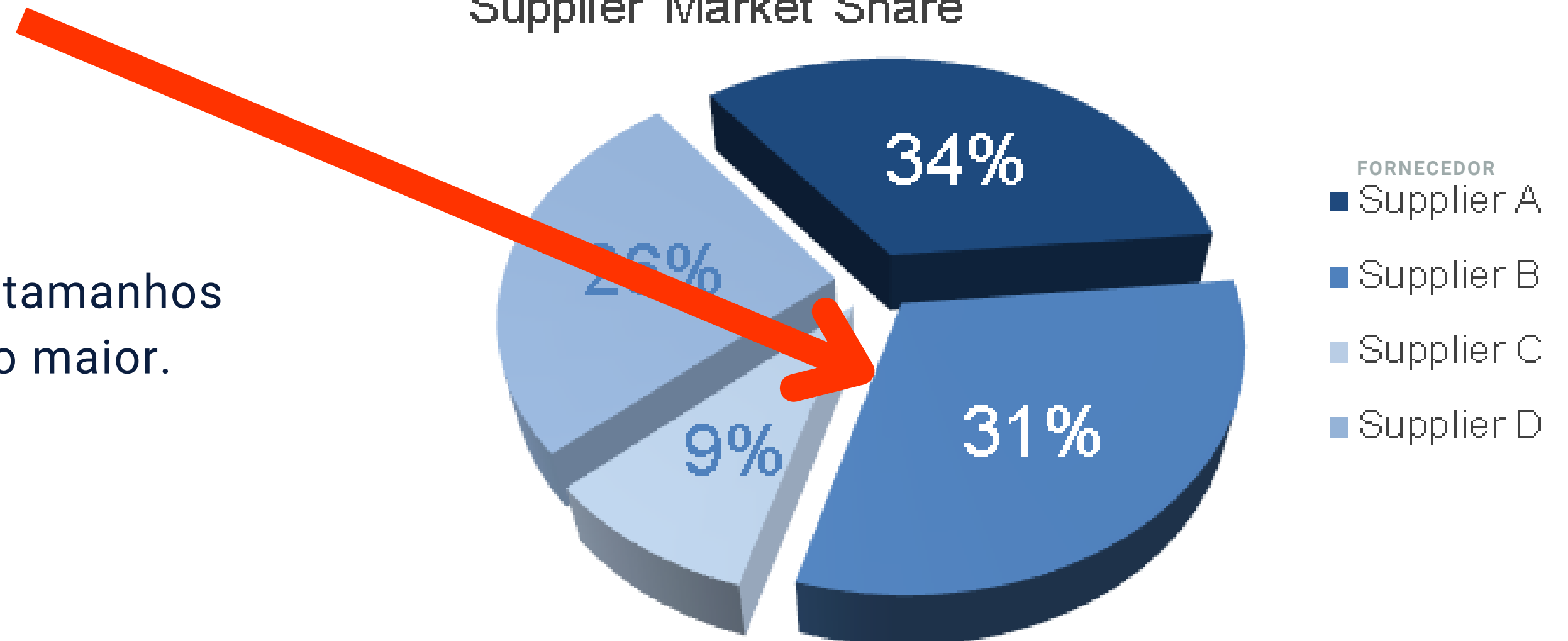
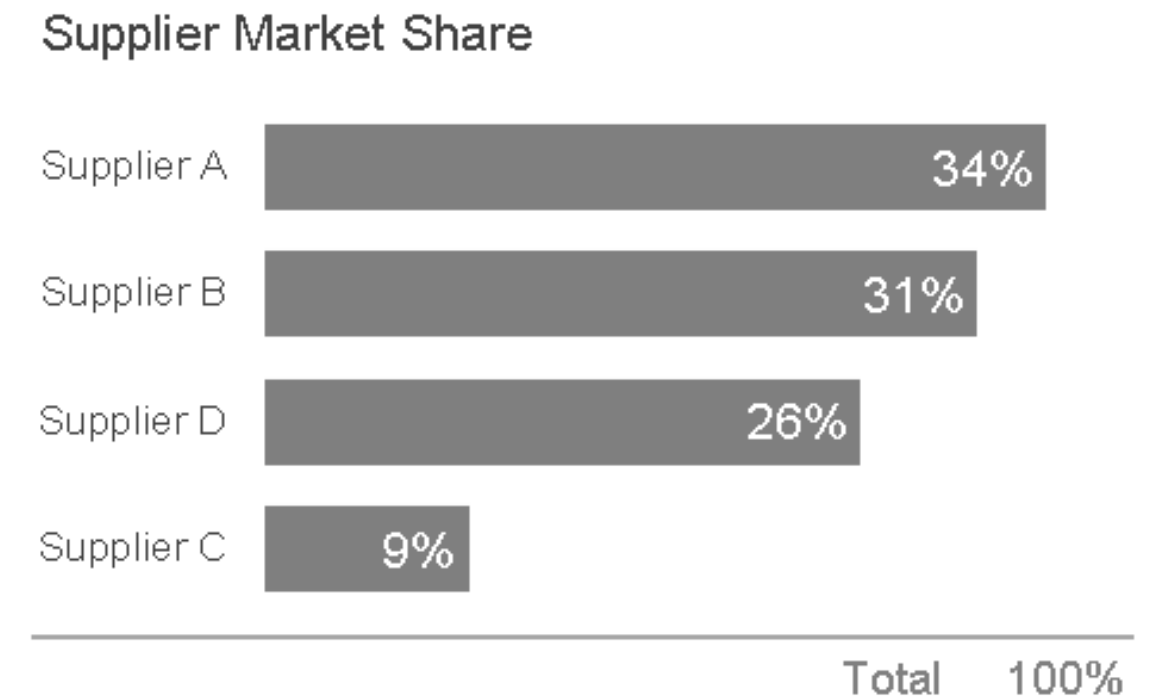


GRÁFICO DE SETOR, PIZZA...

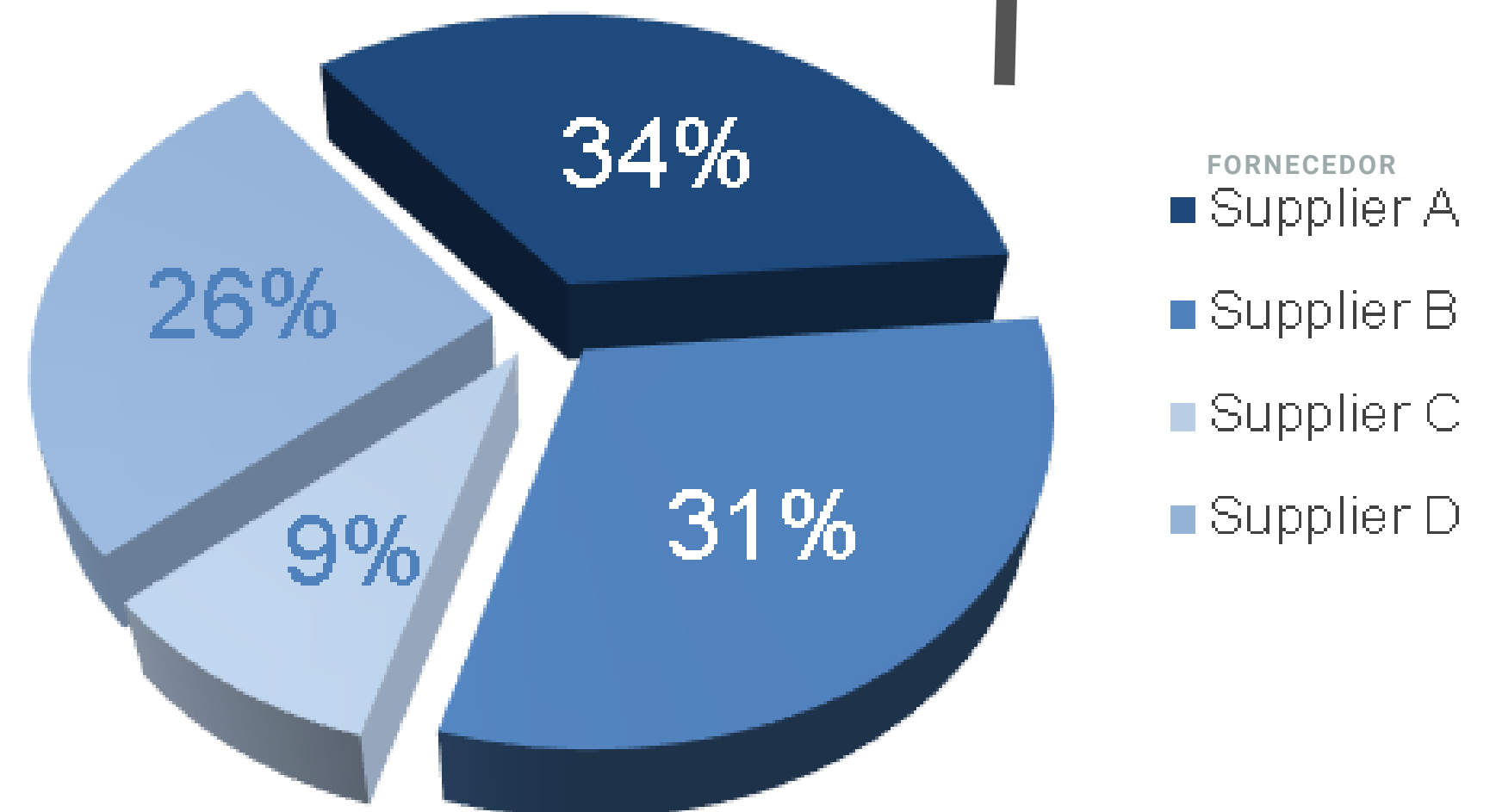
O olho humano não consegue atribuir corretamente valores quantitativos no espaço bidimensional.

Quando segmentos possuem tamanhos parecidos, é difícil dizer qual o maior.

#STORYTELLING



Supplier Market Share



NUNCA USE 3D*

Nunca utilize gráficos 3D quando tiver uma ou duas dimensões, ele introduz elementos desnecessários, como painéis laterais e chão.

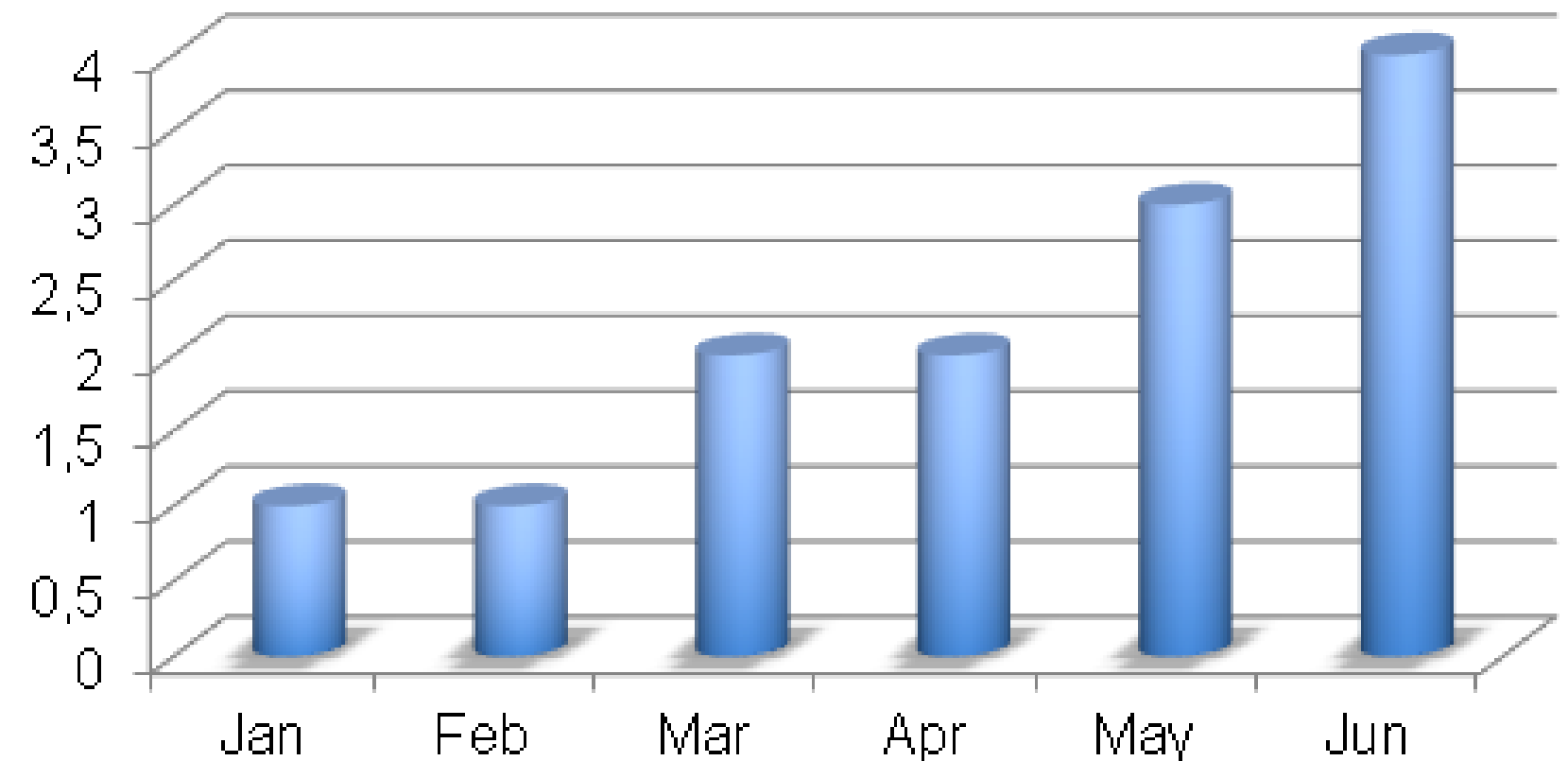
3D distorce os valores, dificultando a interpretação.

Quantas publicações havia em janeiro e fevereiro?

*única exceção é se você estiver representando uma terceira dimensão, de qualquer forma, cuidado!

#STORYTELLING

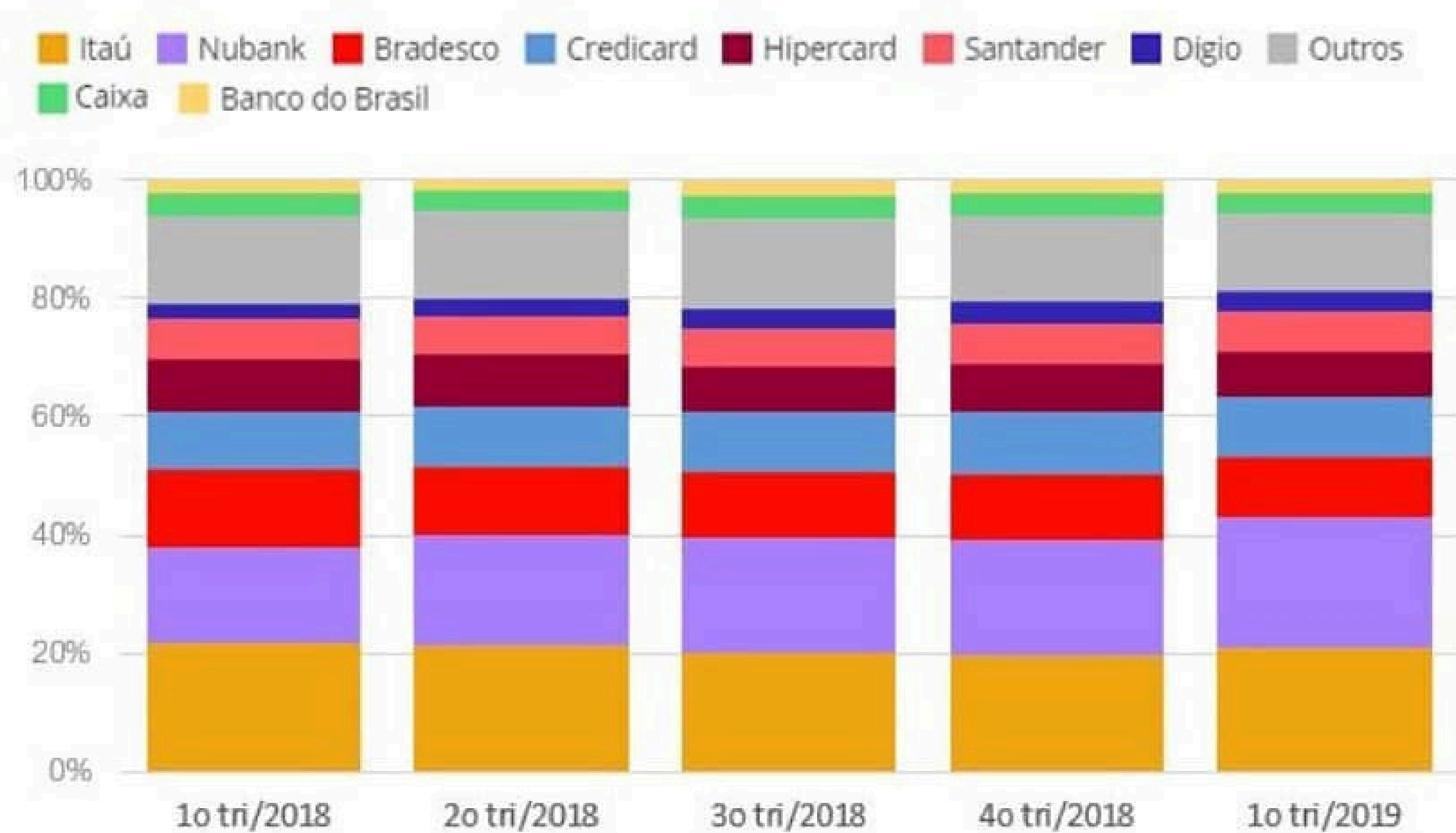
Number of issues



EXEMPLOS VERDADEIROS

Nubank é o grande destaque no Google em 2019

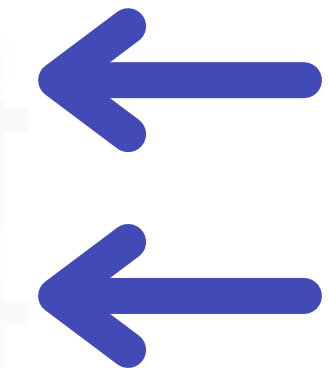
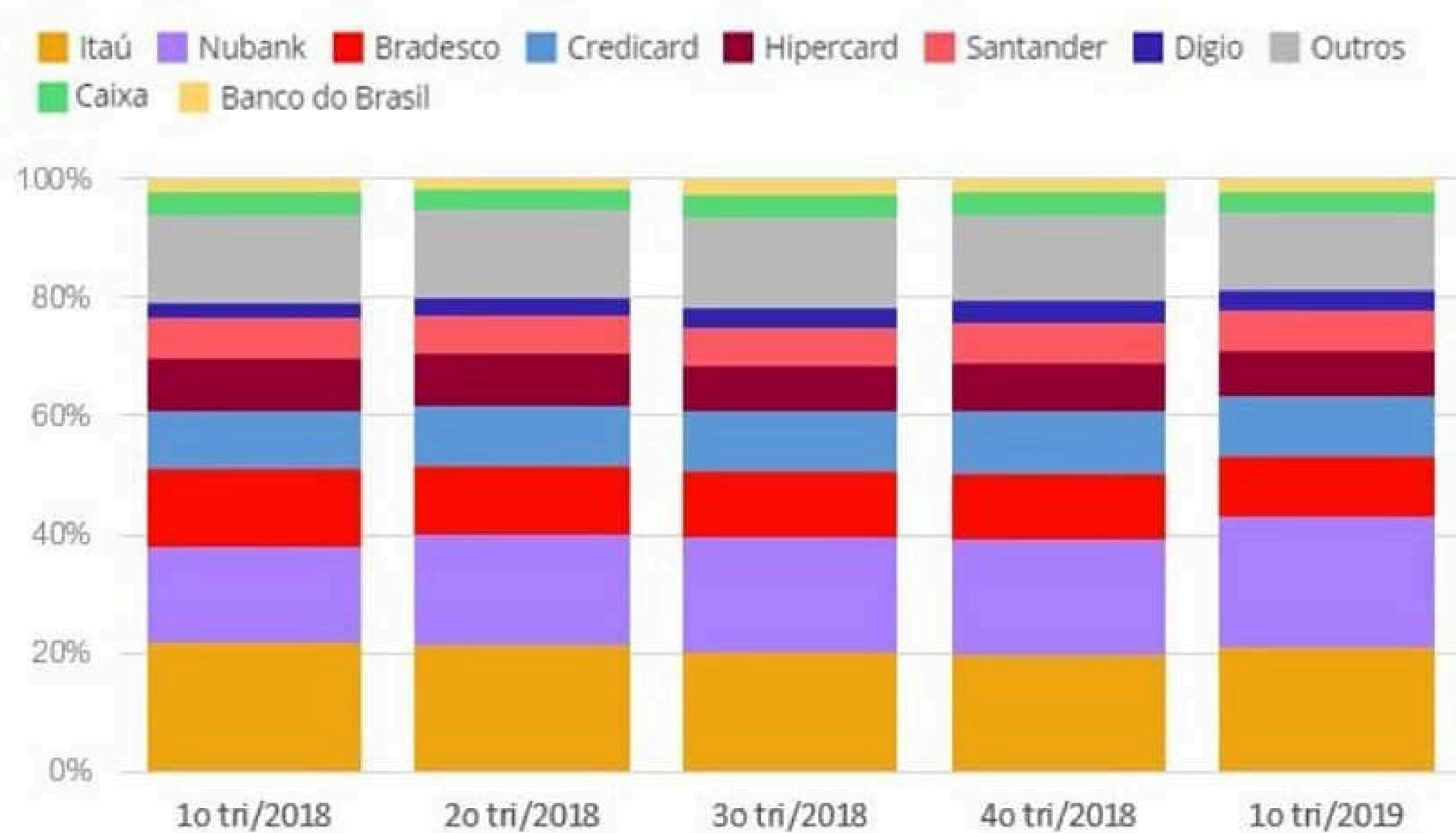
Fintech ultrapassou pela 1ª vez o Itaú em buscas por cartão (Participação do emissor em %)



Fonte: Google

Nubank é o grande destaque no Google em 2019

Fintech ultrapassou pela 1ª vez o Itaú em buscas por cartão (Participação do emissor em %)



Fonte: Google

Nubank ultrapassa Itaú em buscas por cartão de crédito

Fração do total de buscas por emissor de cartão de crédito no site do Google (%)

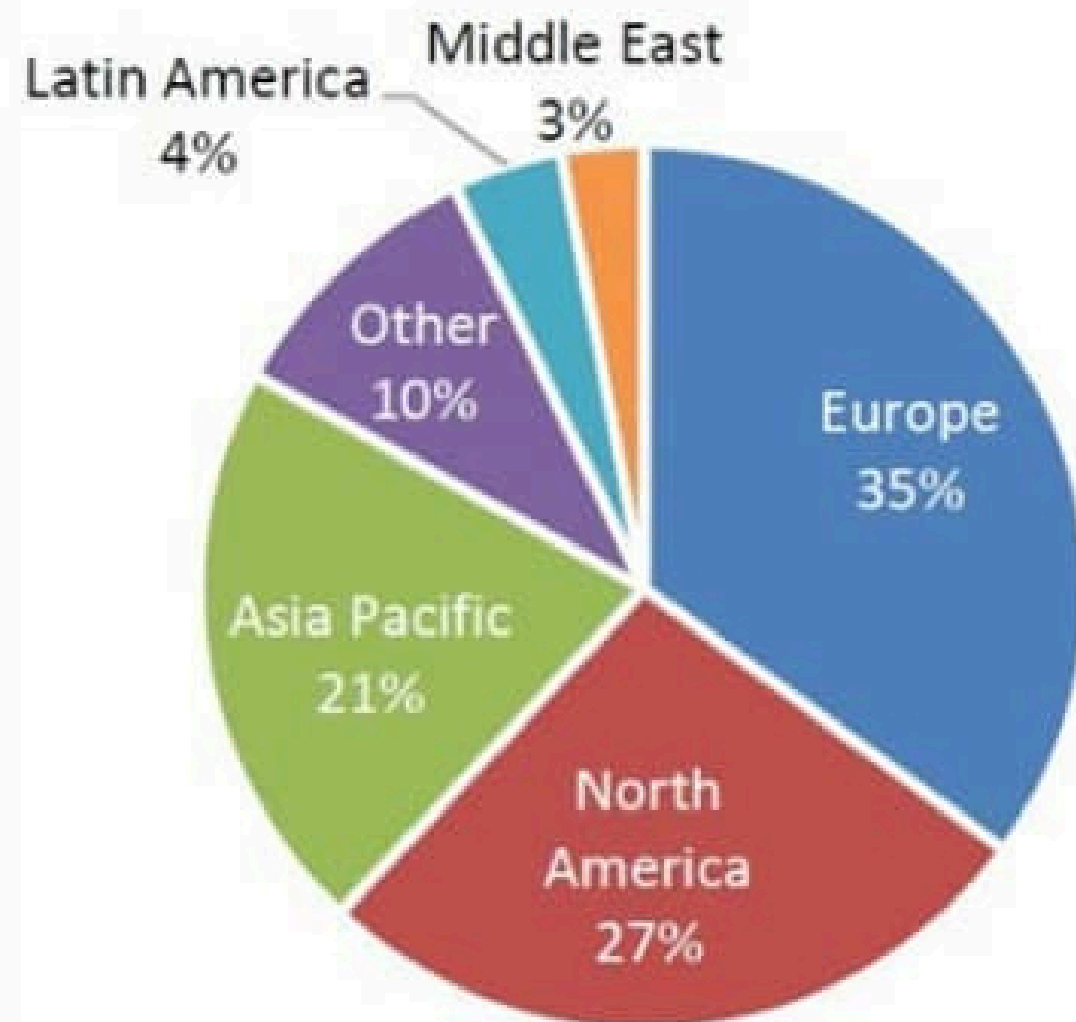


Fonte: Google

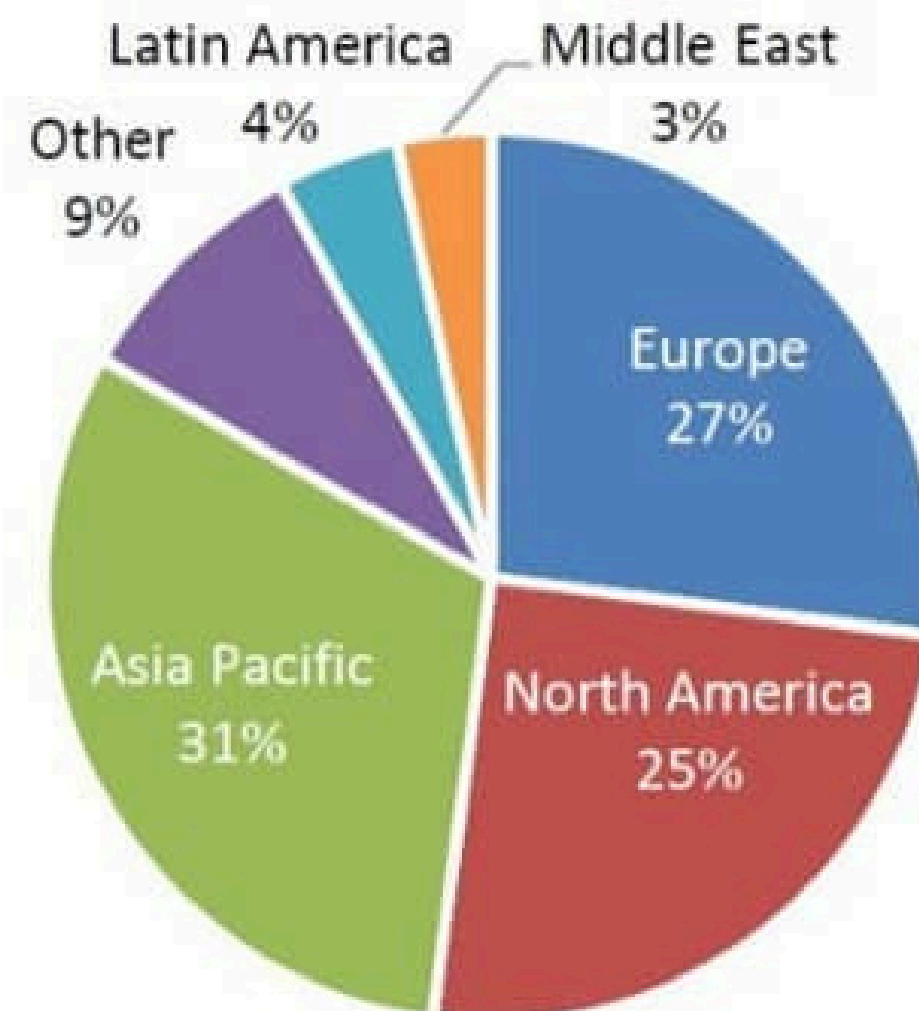
APAC Has Become the Largest Travel Market⁽¹⁾

Regional Share of Total Tourism Contribution to Global GDP

2000



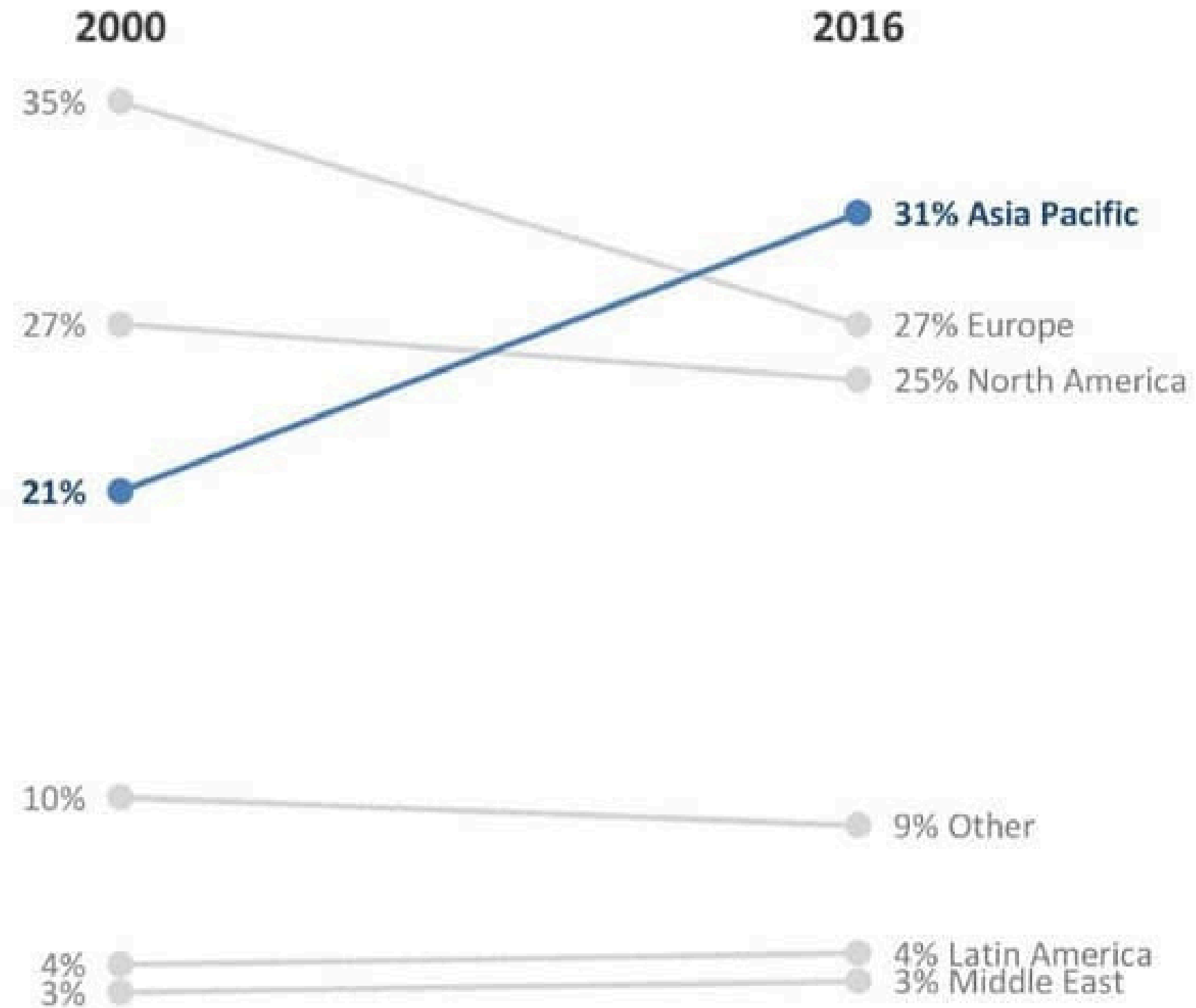
2016E



⁽¹⁾World Travel & Tourism Council

Asia Pacific has become the largest travel market in 2016

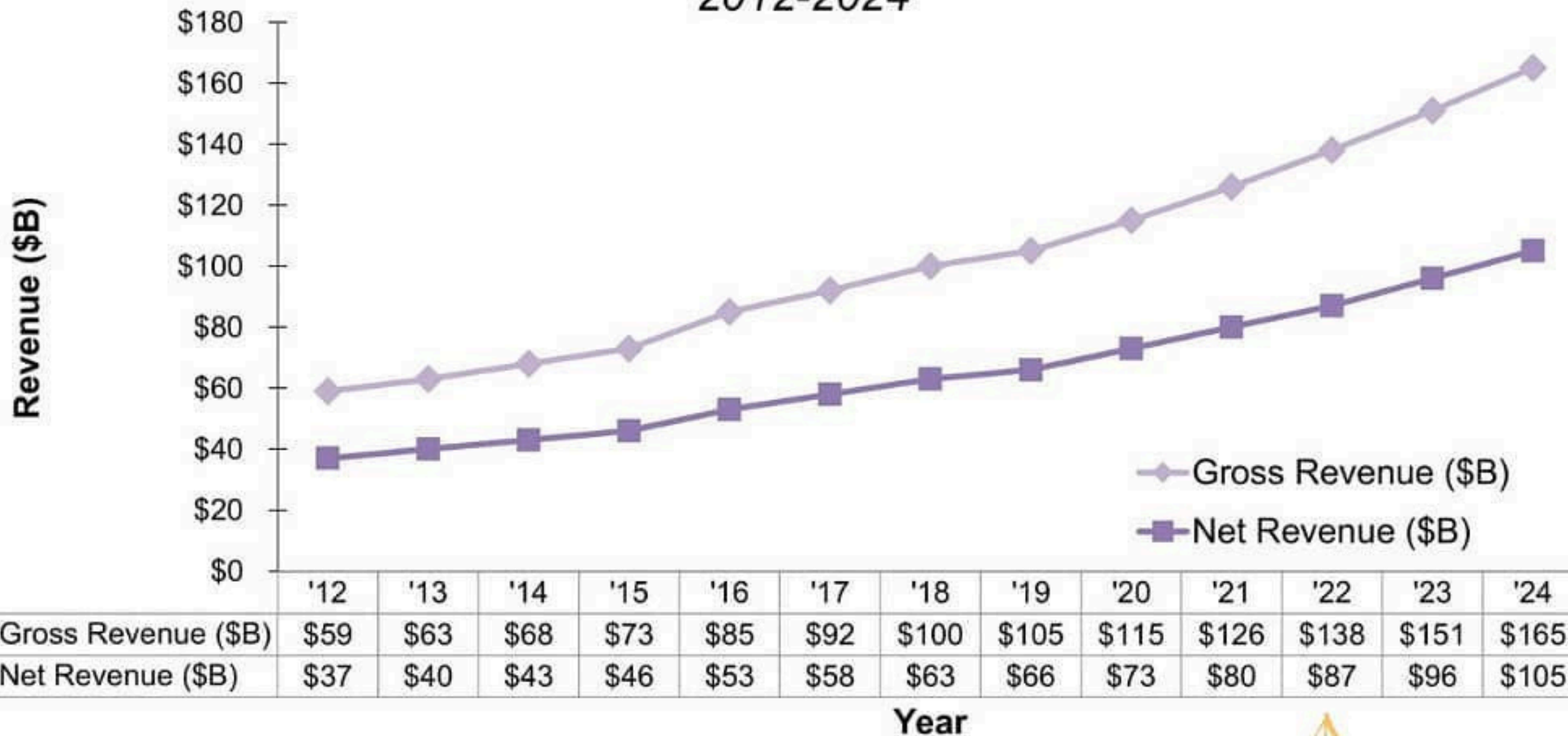
Regional share of total tourism contribution to global GDP⁽¹⁾



⁽¹⁾World Travel & Tourism Council

Net revenue is forecast to grow proportionately with gross revenue, as the gross-to-net differential is based on a historical average, not projected going forward.

Revenue 2012-2024

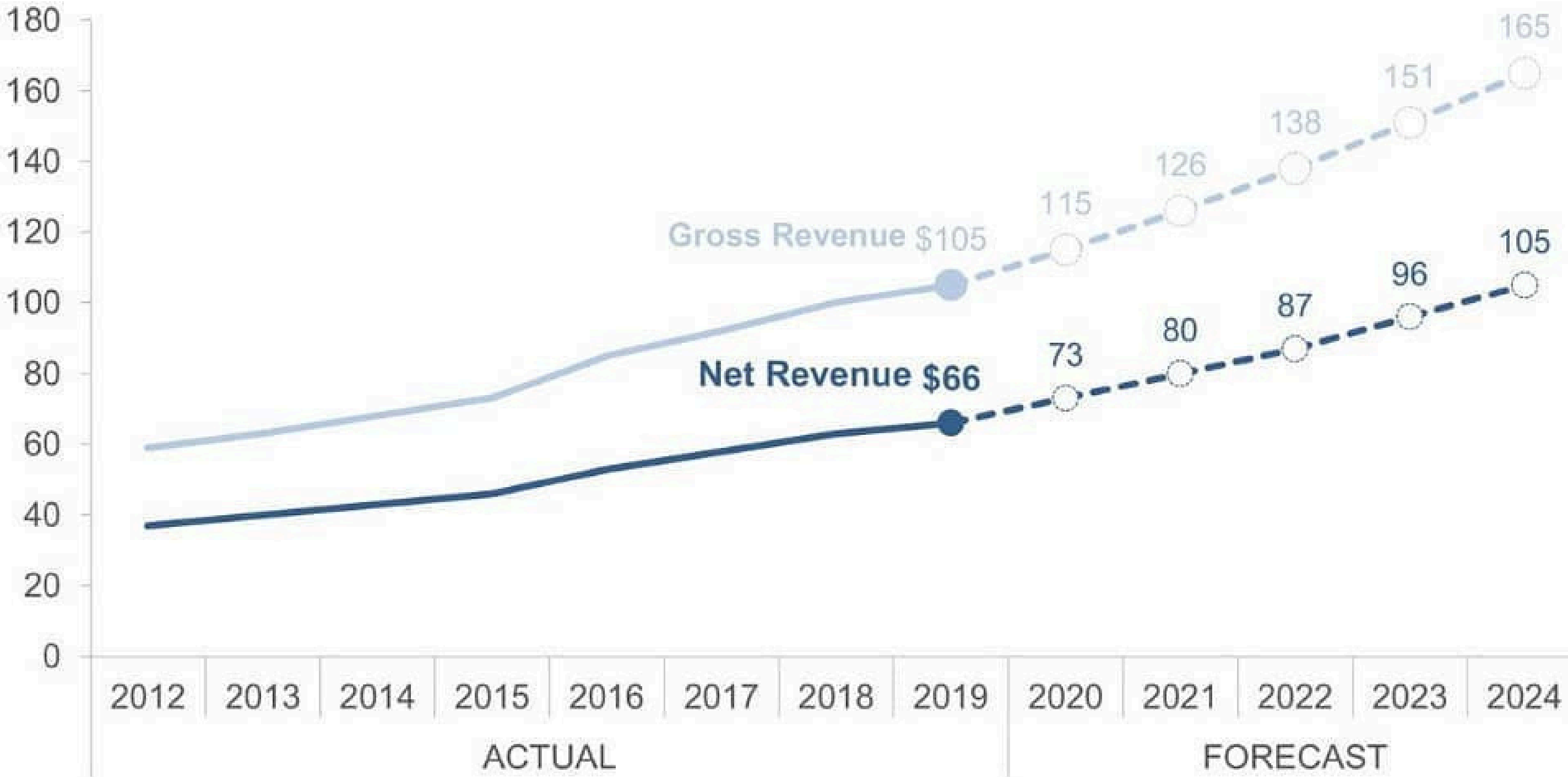


Data from 2020 forward is based on actuals through 12/31/19.



Based on historical average, **Net Revenue is forecast to continue to grow proportionately** to Gross Revenue

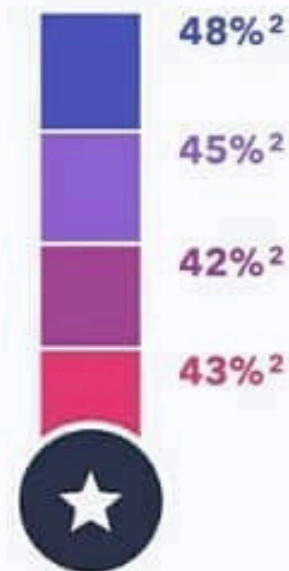
revenue in billions of dollars



VOCÊ SABE O QUE MOTIVA CADA GERAÇÃO A COMPARTILHAR SUA MARCA FAVORITA?



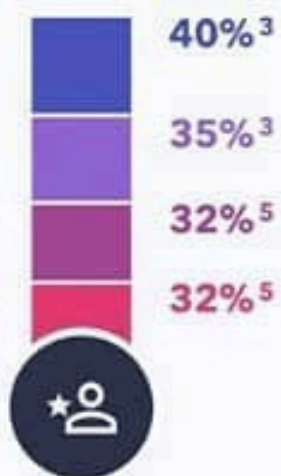
Produtos de qualidade



Recompensas



Relevância do conteúdo



Bom atendimento



Amor pela marca

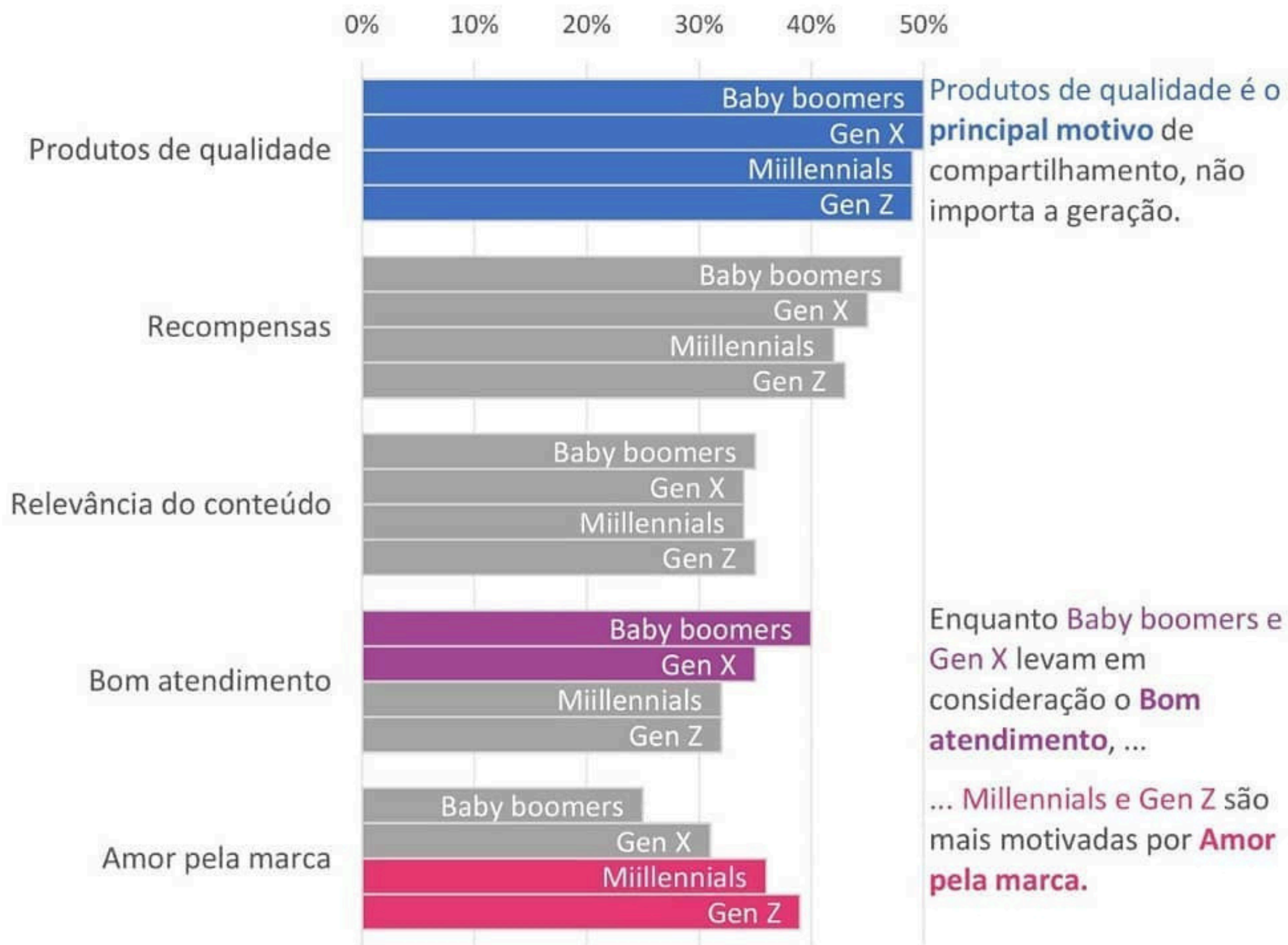
- B** Baby Boomers
- X** Gen X
- M** Millennials
- Z** Gen Z

COMPARTILHE
ESTE POST

@RKISO

Fonte: GlobalWebindex - base de 138.962 pessoas em 45 países

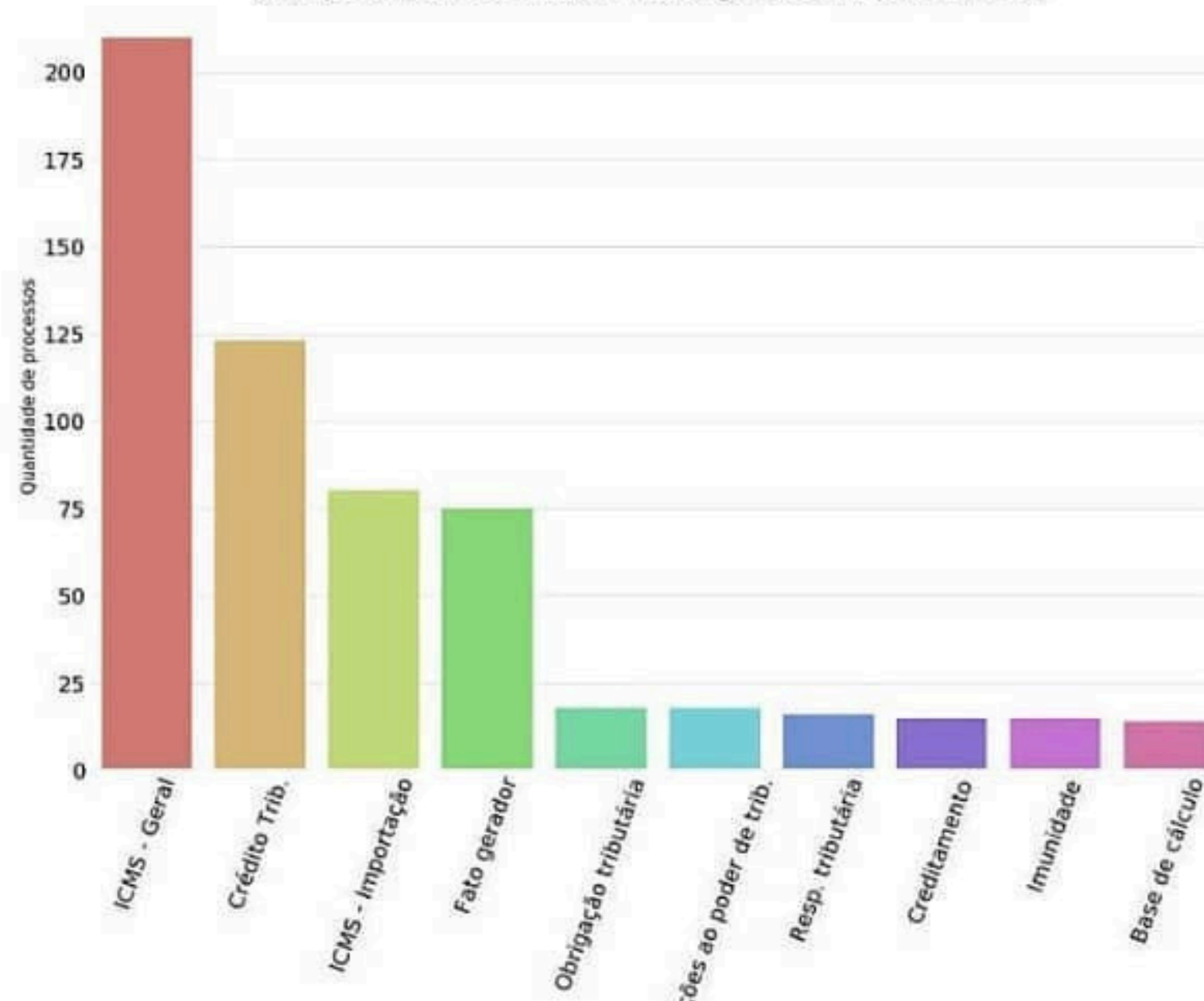
O que motiva cada geração a compartilhar sua marca favorita?



São Paulo, o estado com mais demandas no STF.

jurisintel.com.br

Temas de Direito Tributário nos processos*
em que o Estado de São Paulo figura como parte no STF

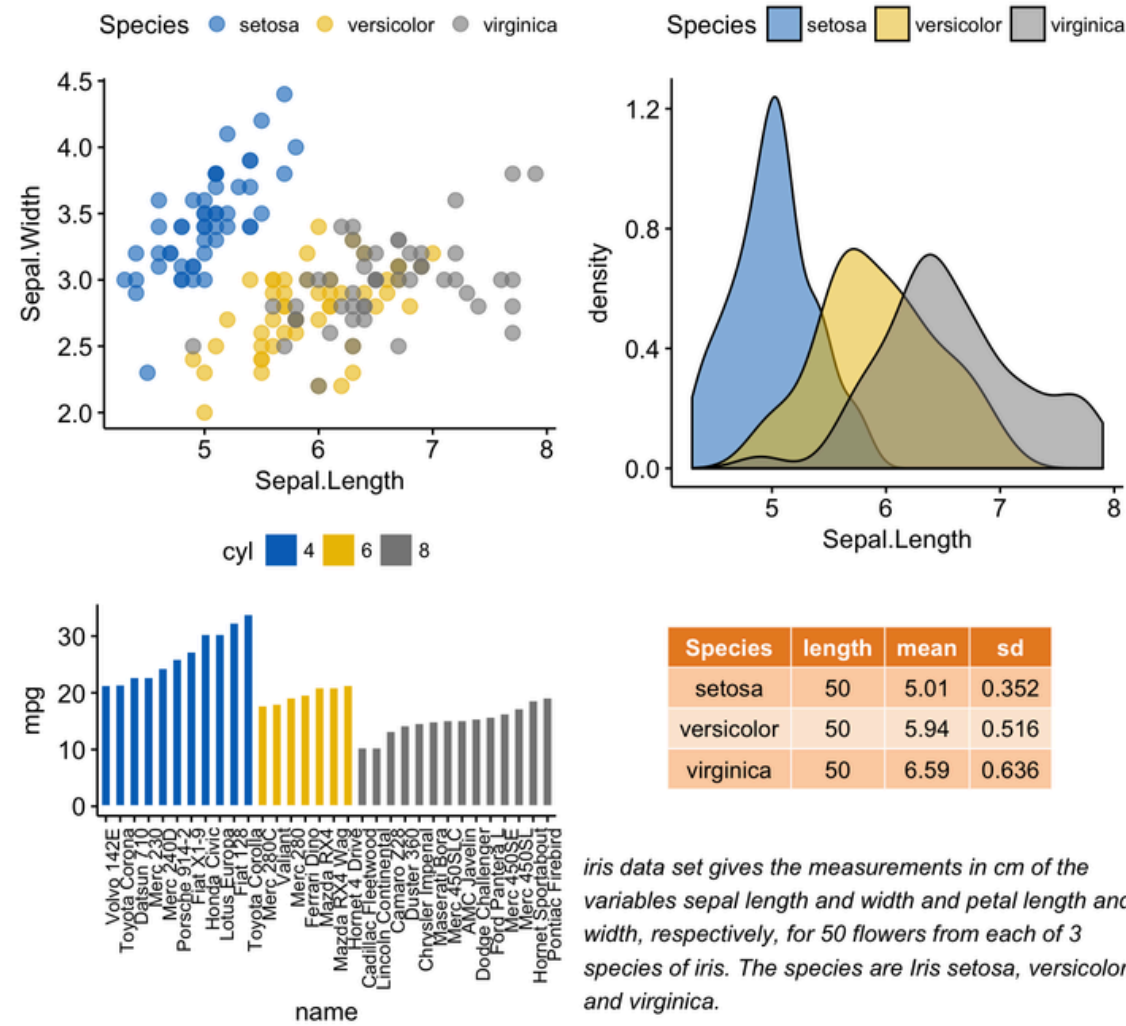


Processos tributários do Estado de SP no STF estão **concentrados em 4 temas**



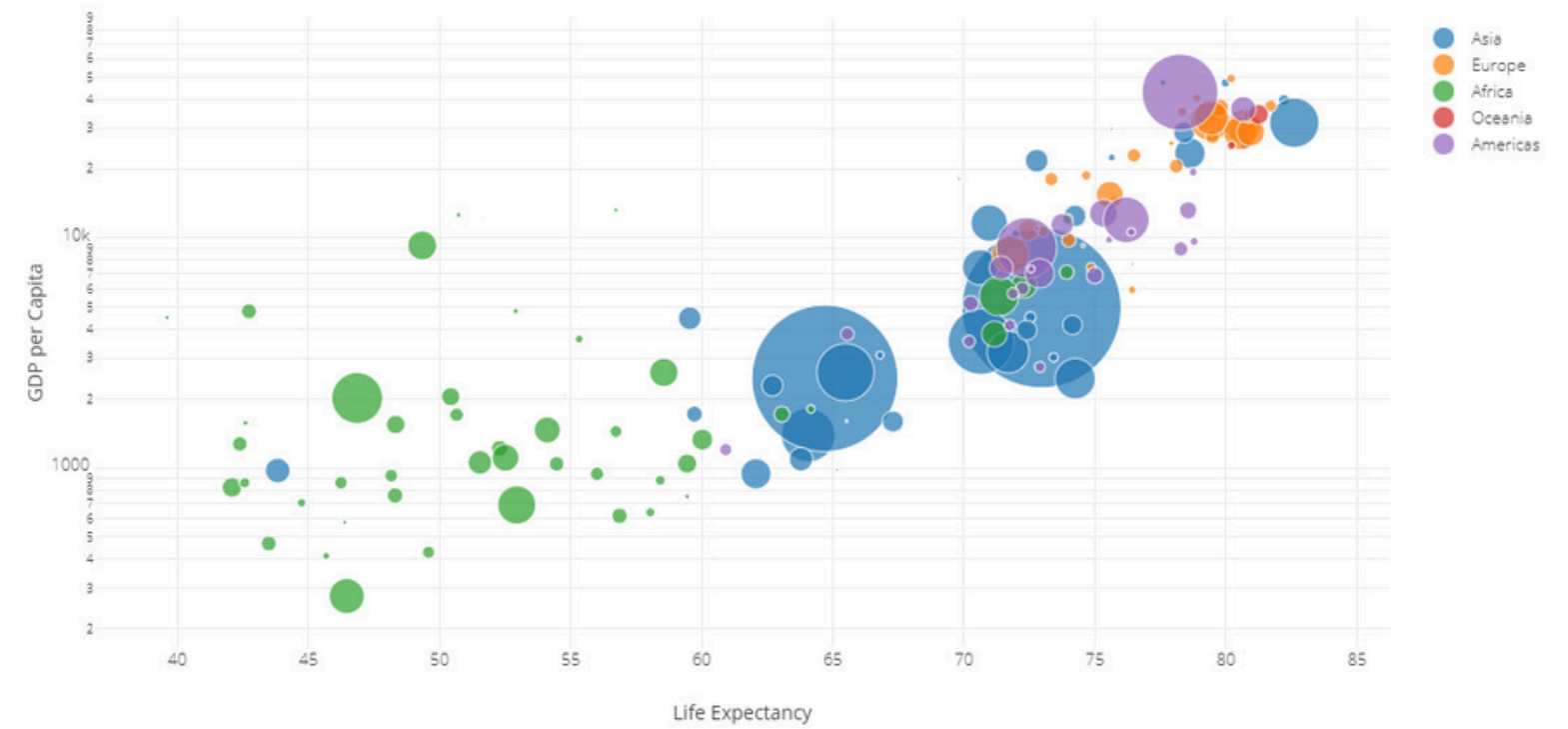
dicas de pacotes no R

ggplot2



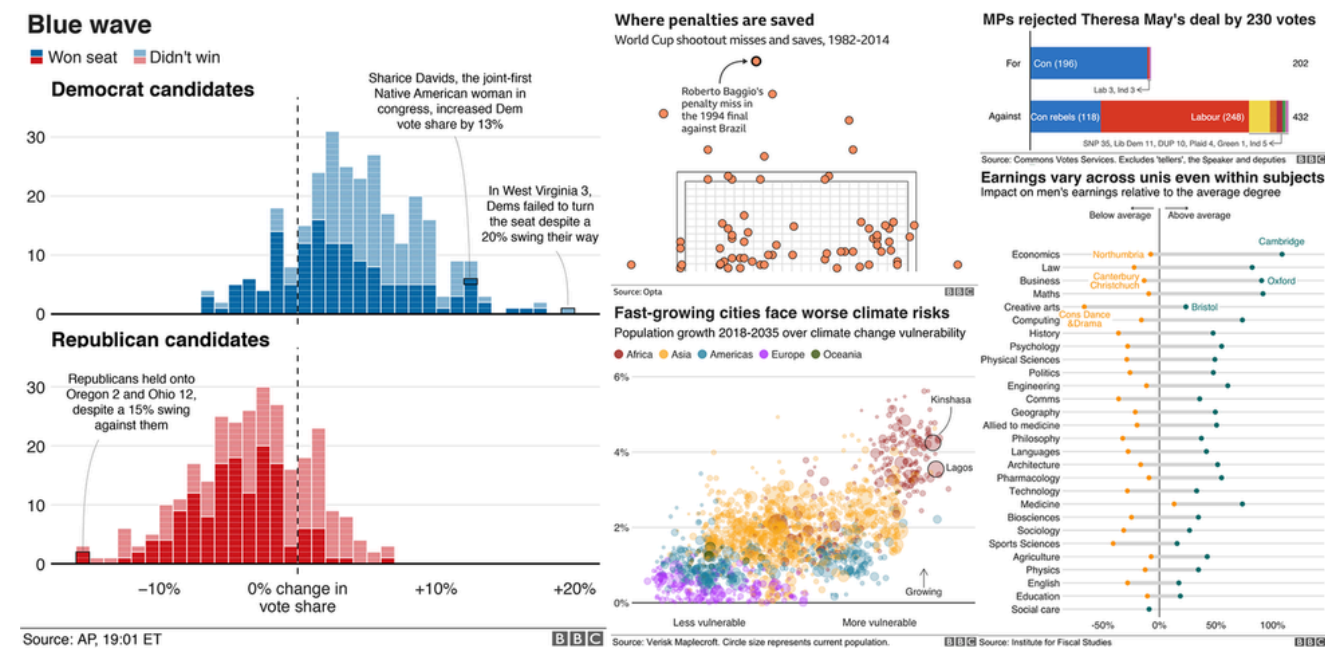
iris data set gives the measurements in cm of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*.

plotly



bbplot

<https://bbc.github.io/rcookbook/>



Source: AP, 19:01 ET

BBC

Source: Venk Maplecraft. Circle size represents current population.

Source: Institute for Fiscal Studies

BBC

PRÓXIMOS PASSOS

1. ENTENDA O CONTEXTO:

- Saiba claramente para quem você está comunicando, o que precisa que saibam ou façam, como vai se comunicar e quais dados tem para apoiar seu argumento.
- Empregue conceitos como a história de 3 minutos, a grande ideia e o storyboard para enunciar sua história e planejar o conteúdo e o fluco desejados.



PRÓXIMOS PASSOS

2. ESCOLHA UM VISUAL APROPRIADO:

- Para realçar um ou dois números, texto simples é o melhor.
- Gráficos de linhas normalmente são melhores para dados contínuos.
- Gráficos de barras funcionam bem para dados categóricos e se possível devem ter uma linha de base zero.
- Deixe a relação que você quer mostrar guiar o tipo de gráfico escolhido.
- Evite pizzas, roscas, 3d e eixos y secundários, devido à dificuldade de interpretação visual.



PRÓXIMOS PASSOS

3. ELIMINE A SATURAÇÃO:

- Identifique os elementos que não acrescentam valor informativo e remova-os de seus visuais.
- Use o contraste estrategicamente.
- Empregue alinhamento de elementos e mantenha espaço em branco para ajudar a tornar a interpretação de seus visuais uma experiência confortável para seu público.



PRÓXIMOS PASSOS

4. FOCALIZE A ATENÇÃO PARA ONDE VOCÊ DESEJA

- Empregue o poder dos atributos pré-atentivos, como cor, tamanho e posição, para sinalizar o que é importante.
- Use atributos estratégicos para chamar a atenção para onde você deseja que seu público olhe e para guiá-lo em seu visual.
- Avalie a eficácia dos atributos pré-atentivos em seu visual, aplicando o teste "para onde seus olhos são atraídos?".



PRÓXIMOS PASSOS

5. PENSE COMO UM DESIGNER

- Ofereça ao seu público possibilidades visuais como indícios de como interagir com sua comunicação: realce o que for importante, elimine as distrações e crie uma hierarquia visual de informações.
- Torne seus designs acessíveis, não complicando demais e usando texto para legendar e explicar.
- Aumente a tolerância de seu público a problemas de design tornando seus visuais esteticamente atraentes.
- Trabalhe para ganhar a aceitação do público para seus designs visuais.



PRÓXIMOS PASSOS

6. CONTE UMA HISTÓRIA

- Elabore uma história com início (trama), meio (reviravoltas) e fim (chamada para ação) claros.
- Use conflito e tensão para chamar e manter a atenção do público.
- Considere a ordem e a maneira de sua narrativa.
- Utilize o poder da repetição para ajudar a fixar suas histórias.
- Empregue táticas como lógica vertical e horizontal, storyboard inverso e busque uma nova perspectiva para garantir que sua história seja claramente compreendida em sua comunicação.

#STORYTELLING



REFERÊNCIAS E FONTES

1. LIVRO: STORYTELLING COM DADOS: UM GUIA SOBRE VISUALIZAÇÃO DE DADOS PARA
PROFISSIONAIS DE NEGÓCIOS

2. @STORYTELLINGCOMDADOS