

VISUALIZAÇÃO COMPUTACIONAL

Visualização de dados
em big data

TÓPICOS

- Importância da visualização de dados em big data
- Técnicas de visualização de dados

GRANDES VOLUMES DE DADOS



Virginia Tech Data Center
Fonte: Christopher Bowns, Wikimedia C

GRANDES VOLUMES DE DADOS

Produto A: 5 marcas distintas

GRANDES VOLUMES DE DADOS

Produto A: 5 marcas distintas



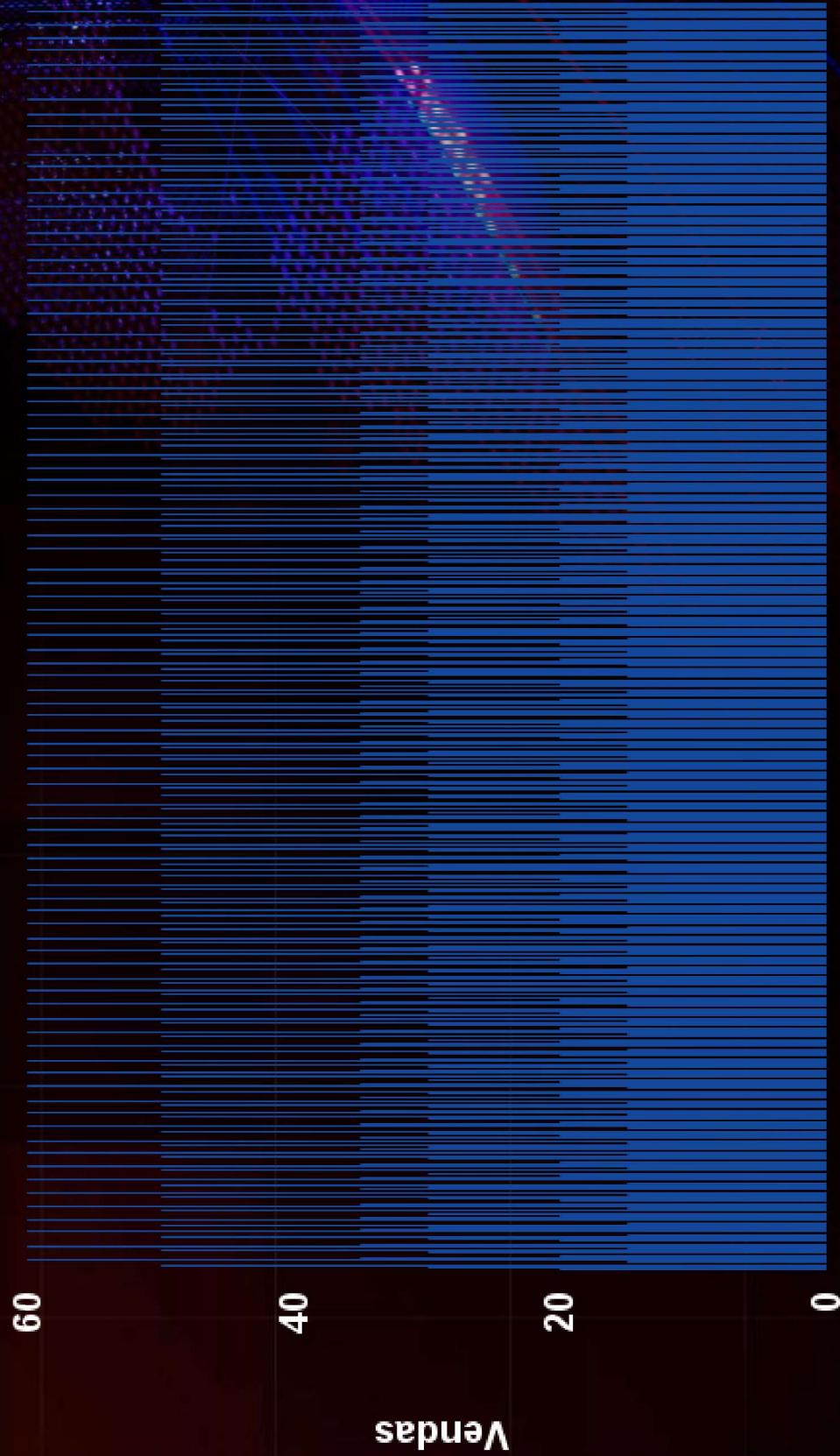
GRANDES VOLUMES DE DADOS

Produto B: 500 marcas distintas

?

GRANDES VOLUMES DE DADOS

Produto B: 500 marcas distintas



GRANDES VOLUMES DE DADOS

Produto B: 500 marcas distintas

60

40

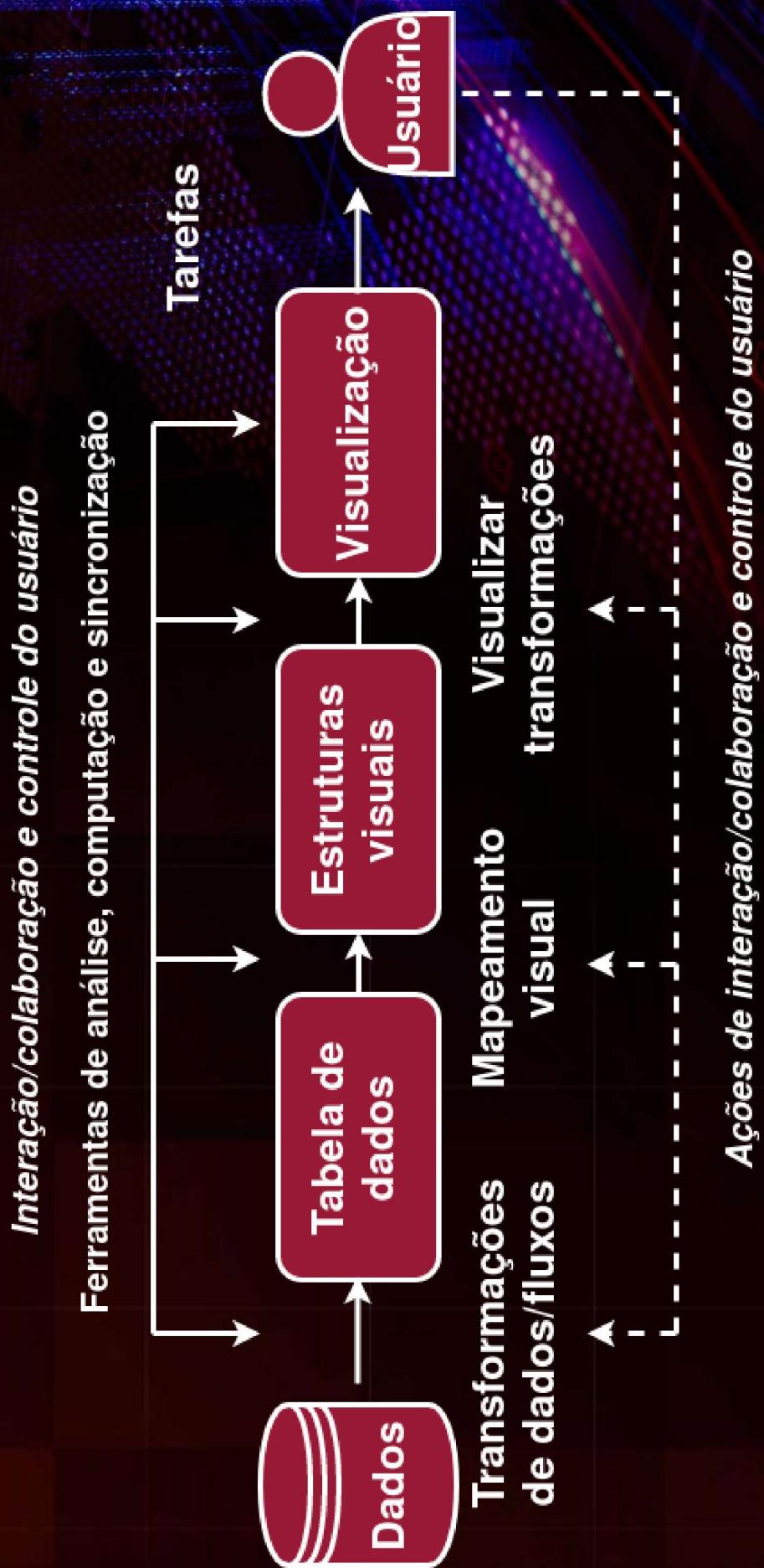
20

0

Vendas

Como interpretar esses dados?

PROCESSO DE VISUALIZAÇÃO DE DADOS



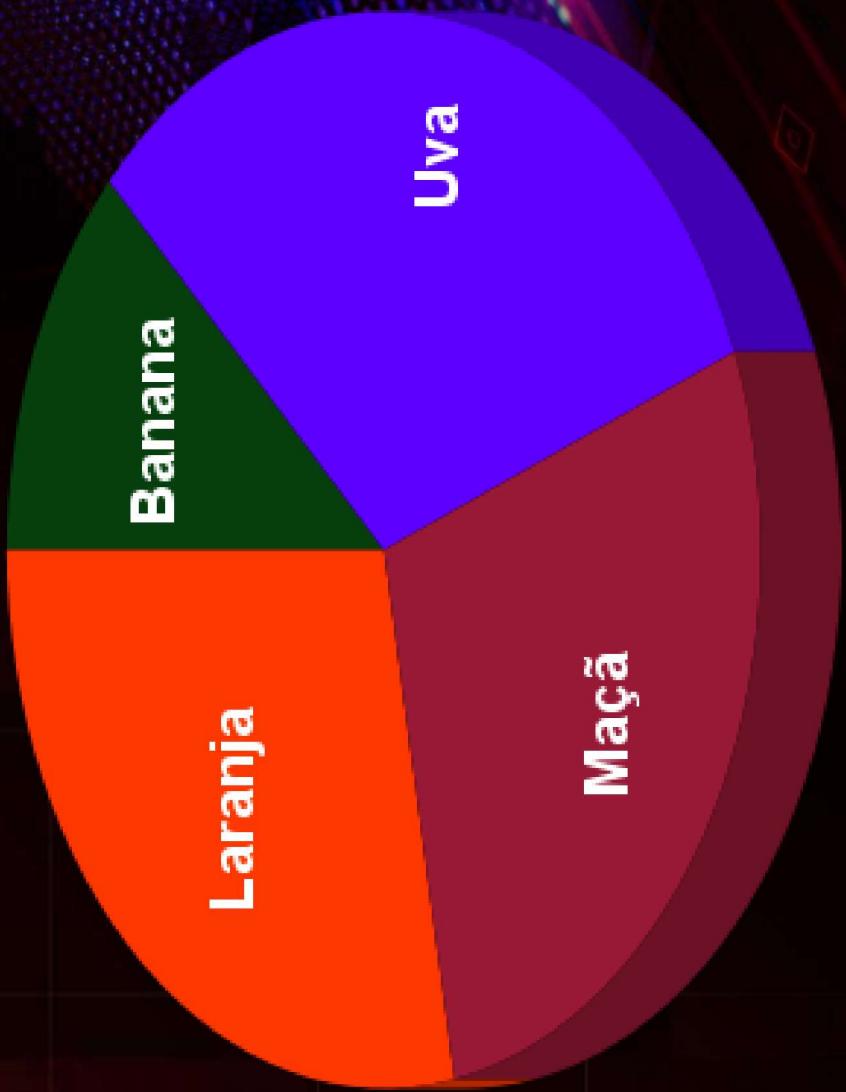
IMPORTÂNCIA DA VISUALIZAÇÃO DE DADOS

- Apresentar os resultados de uma análise
- Recurso diferencial quando não há questões, ou hipóteses prévias sobre os dados
- Facilitar o processo de compreensão humana

IMPORTÂNCIA DA VISUALIZAÇÃO DE DADOS

| Frutas | Frutas | Frutas | Frutas | Frutas | Frutas |
|---------|---------|---------|---------|---------|---------|
| Banana | Uva | Laranja | Banana | Uva | Banana |
| Uva | Banana | Maçã | Uva | Banana | Maçã |
| Maçã | Laranja | Uva | Maçã | Laranja | Uva |
| Laranja | Maçã | Banana | Laranja | Maçã | Laranja |
| ... | ... | ... | ... | ... | ... |
| Laranja | Laranja | Uva | Banana | Laranja | Laranja |
| Uva | Banana | Laranja | Uva | Banana | Uva |
| Banana | Uva | Maçã | Laranja | Uva | Banana |
| Maçã | Laranja | Laranja | Maçã | Uva | Maçã |
| Laranja | Uva | Banana | Uva | Laranja | Laranja |
| Uva | Maçã | Uva | Laranja | Maçã | Uva |

IMPORTÂNCIA DA VISUALIZAÇÃO DE DADOS



BENEFÍCIOS DA VISUALIZAÇÃO DE DADOS

- Interpretação mais rápida
- Independente do idioma
- Identificação de problemas nos próprios dados

BENEFÍCIOS DA VISUALIZAÇÃO DE DADOS

| Benefícios | % |
|---|----|
| Melhor tomada de decisão | 77 |
| Melhor análise <i>ad hoc</i> de dados | 43 |
| Aprimoramento da colaboração / compartilhamento de informações | 41 |
| Fornecimento de recursos de autoatendimento aos usuários finais | 36 |
| Maior retorno do investimento (ROI) | 34 |
| Economia de tempo | 20 |
| Redução de carga em TI | 15 |

Fonte: Milani et al., 2020 apud Ward et al., 2015

TÉCNICAS DE VISUALIZAÇÃO DE DADOS

- Existem diversas formas de visualização de dados
- Algumas técnicas são mais frequentes em big data
- Podem variar de acordo com os dados:
 - Quantidade de atributos
 - Tipo de dados
 - Localização
 - Hierarquia
 - Tempo

TÓPICOS

1. Importância da visualização de dados em big data
2. Técnicas de visualização de dados

TÉCNICAS DE VISUALIZAÇÃO DE DADOS

| Técnica | Uso |
|----------------------------------|--|
| Histograma | Distribuição de frequências |
| Gráfico de dispersão | Relacionamento entre variáveis, correlação, dispersão |
| Gráfico de coordenadas paralelas | Relação entre muitas variáveis, posicionamento, volume + variedade |
| Mapa de árvore | Dados hierárquicos, entendimento do todo |
| Mapa coroplético | Informações geolocalizadas, mapas |

Fonte: Milani et al., 2020

TÉCNICAS DE VISUALIZAÇÃO DE DADOS

| Técnica | Uso |
|----------------------------------|--|
| Histograma | Distribuição de frequências |
| Gráfico de dispersão | Relacionamento entre variáveis, correlação, dispersão |
| Gráfico de coordenadas paralelas | Relação entre muitas variáveis, posicionamento, volume + variedade |
| Mapa de árvore | Dados hierárquicos, entendimento do todo |
| Mapa coroplético | Informações geolocalizadas, mapas |
| Existem muitas outras | |

Fonte: Milani et al., 2020

HISTOGRAMA

- Avaliar distribuição de valores
- Variável única
- Intervalos:
 - Distribuição de frequência
 - Caixas (bins)
- Compreender melhor os dados

HISTOGRAMA



GRÁFICO DE DISPERSÃO

- Comparar duas variáveis numéricas:
- Baseado no sistema de coordenadas cartesiano
- Visualizar correlações
- Padrão de correlação:
 - Positiva, negativa, nula, linear, exponencial
- Força da correlação:
 - Forte, fraca ou nula

GRÁFICO DE DISPERSÃO



GRÁFICO DE DISPERSÃO

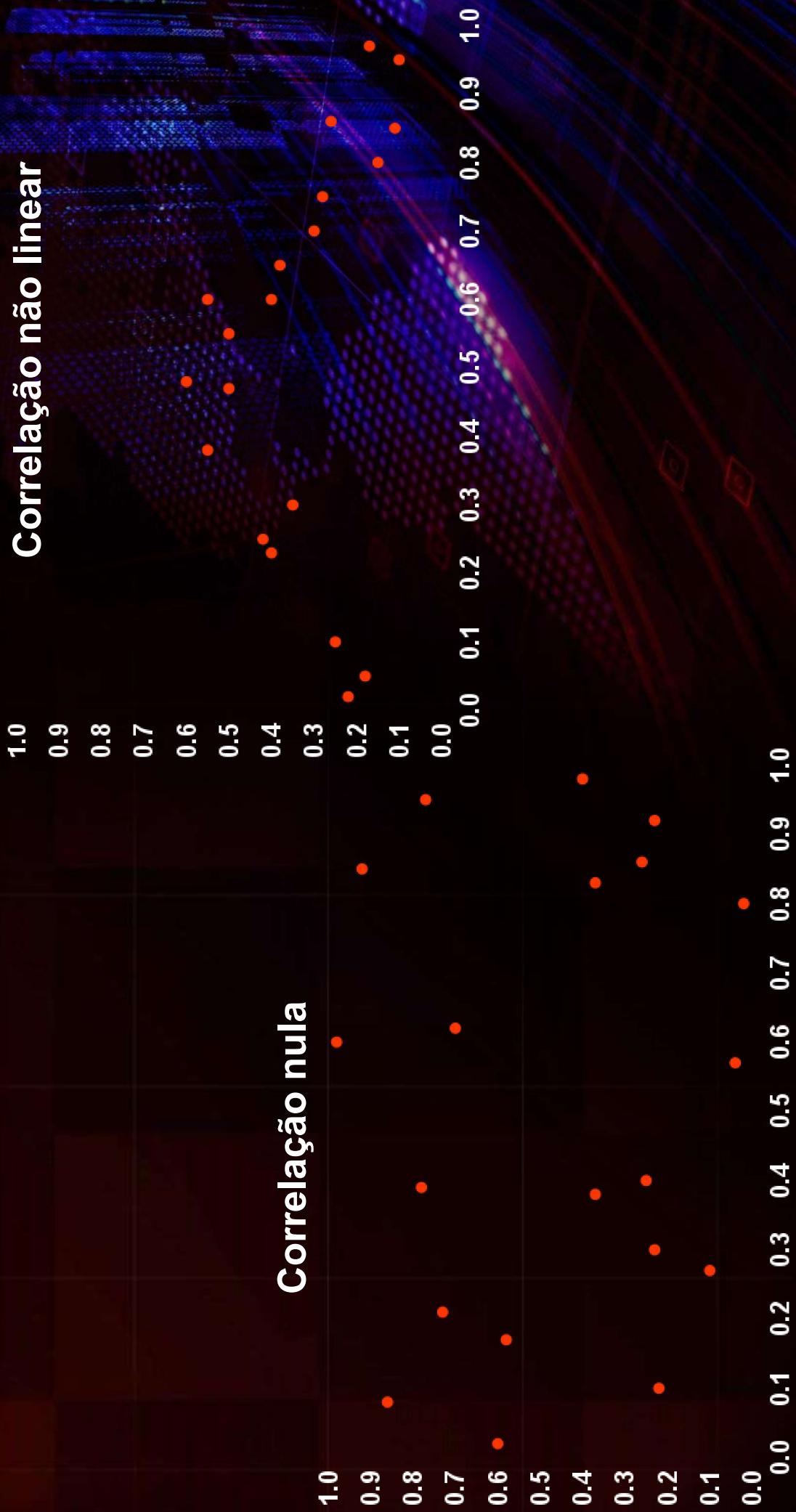


GRÁFICO DE DISPERSÃO



GRÁFICO DE DISPERSÃO

Correlação nula



Correlação nula

Correlação não linear

GRÁFICO DE COORDENADAS PARALELAS

- Visualizar múltiplas variáveis simultaneamente
- Identificar padrões nos dados
- Variáveis com valores distintos:
 - Normalizar
- Mostrar correlações multivariadas:
 - Difícil de visualizar
- Poucas ferramentas automáticas

GRÁFICO DE COORDENADAS PARALELAS

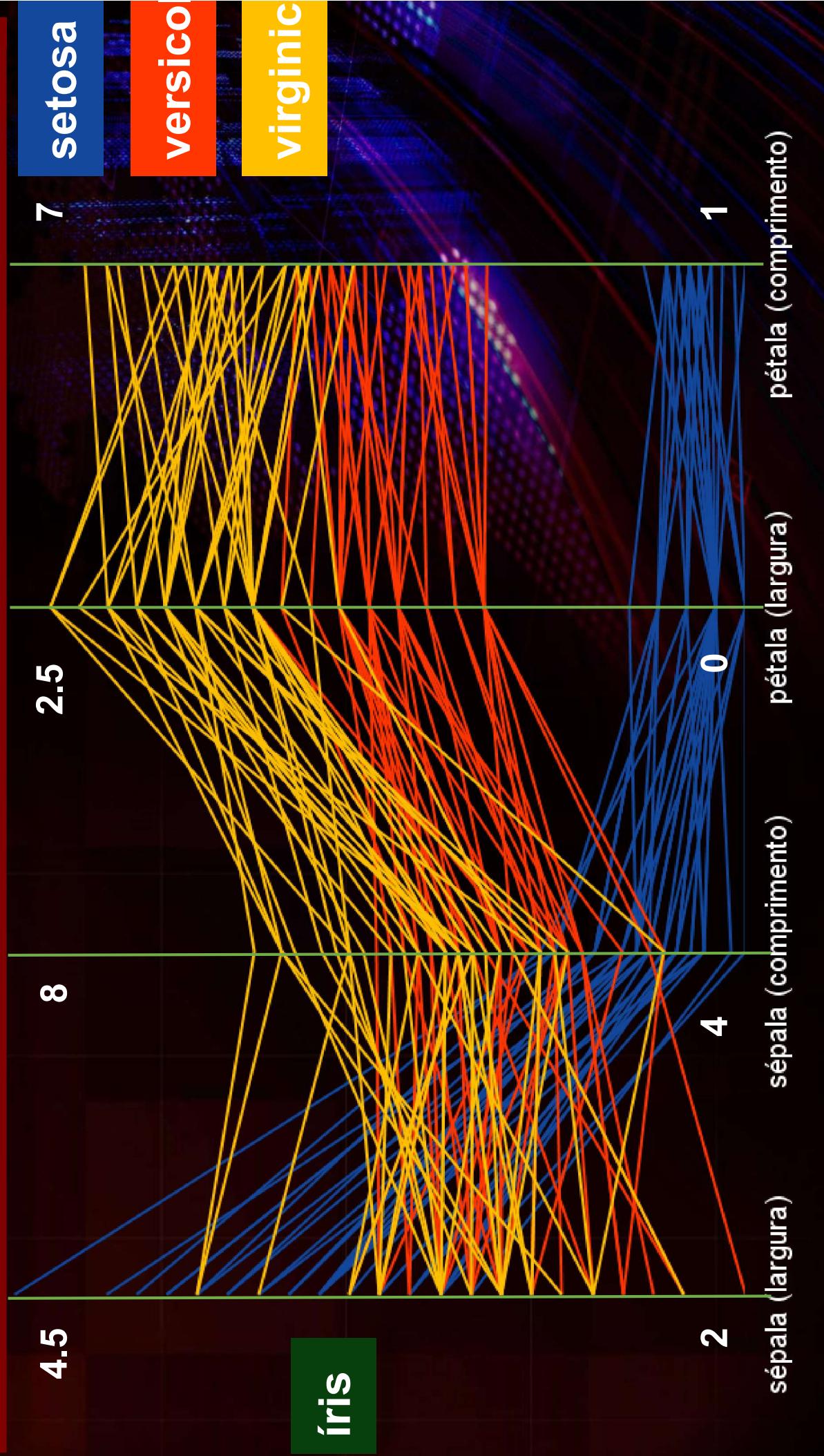


GRÁFICO DE RADAR

- Variação radial do gráfico de coordenadas paralelas
- 3 ou mais variáveis numéricas
- Permite ver a distância entre características
- Sinônimos:
 - Teia de aranha
 - Polar
 - Polígono irregular

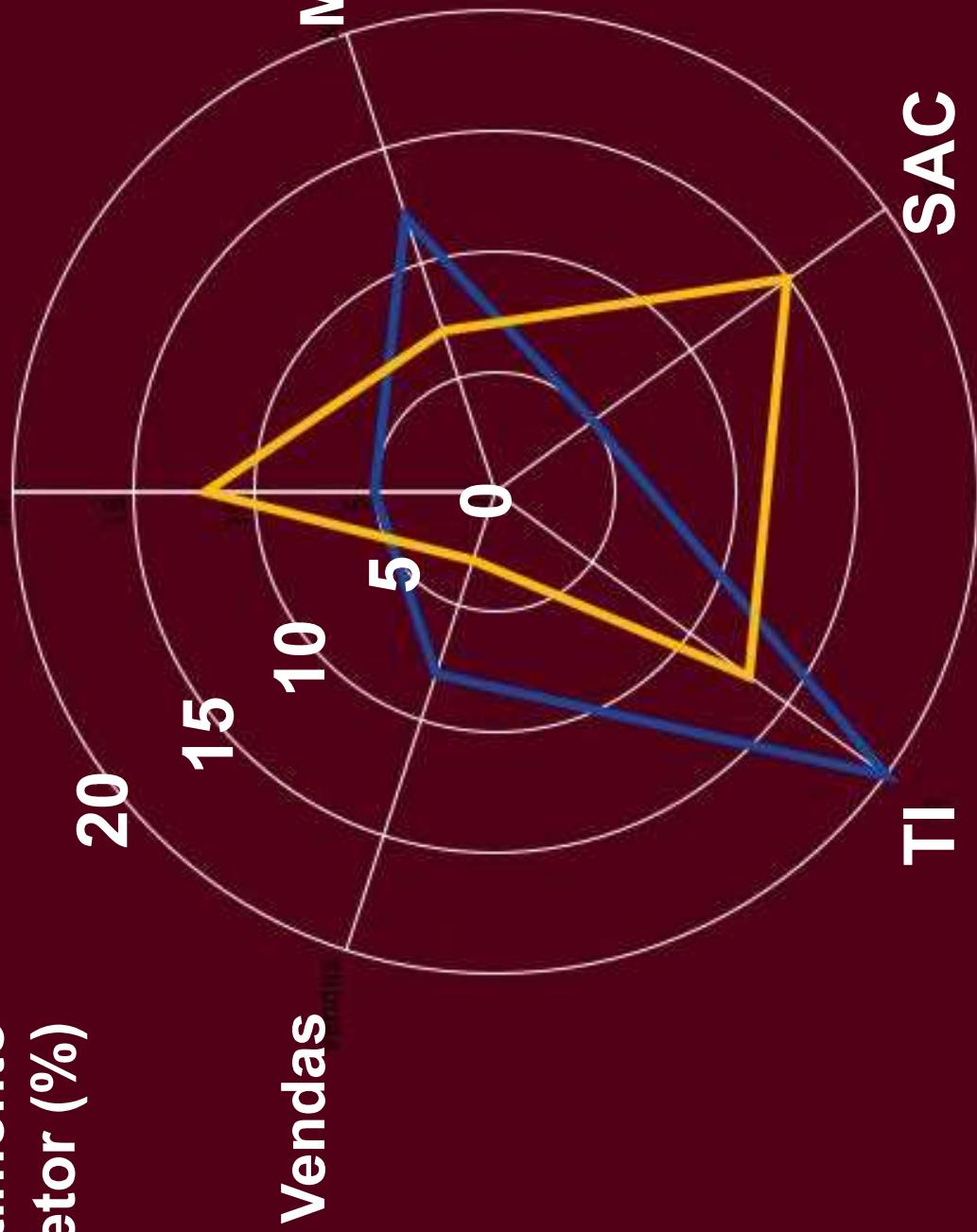
GRÁFICO DE RADAR

Investimento por setor (%)

Administração

Atual

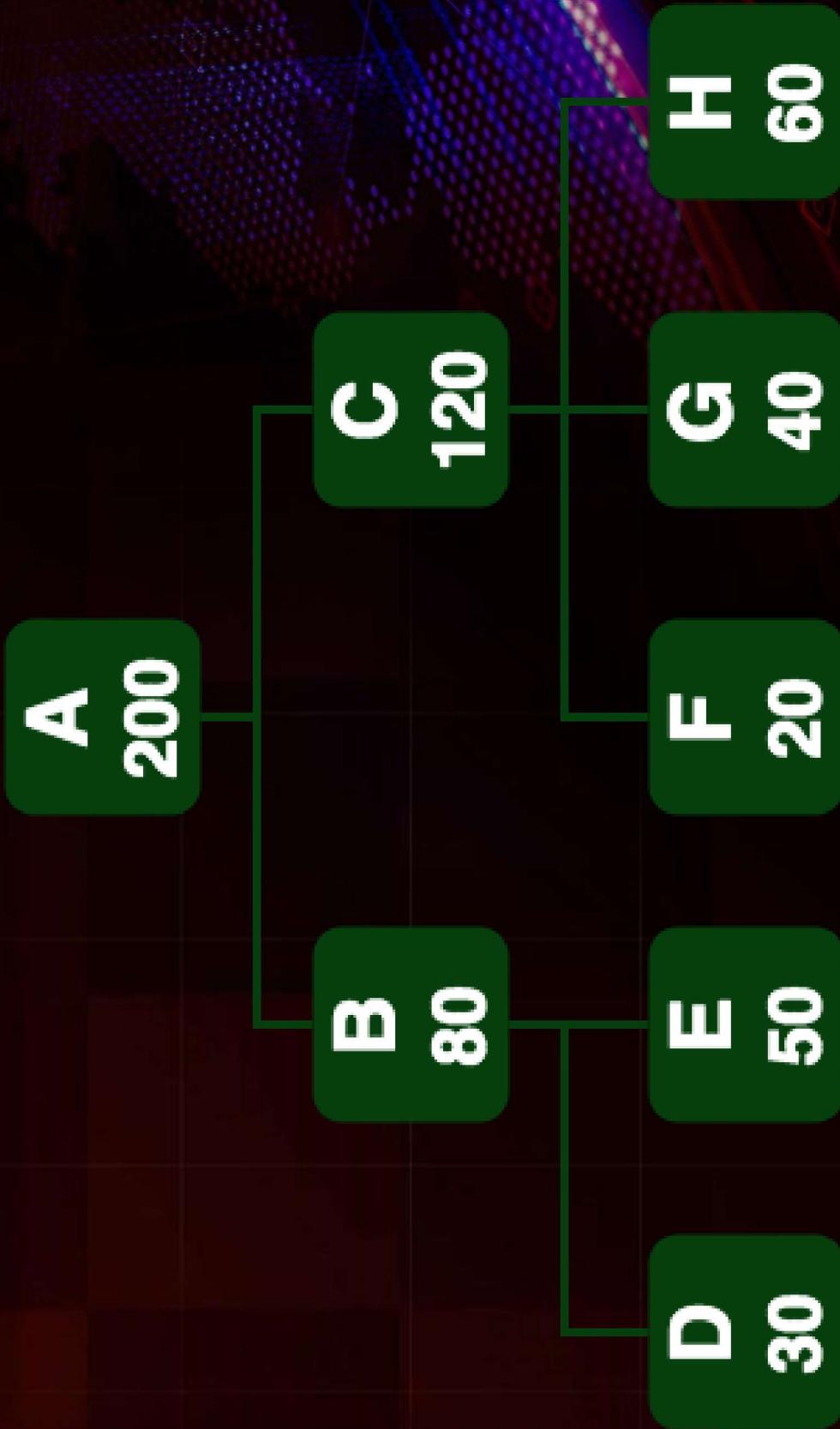
Previsto



MAPA DE ÁRVORE

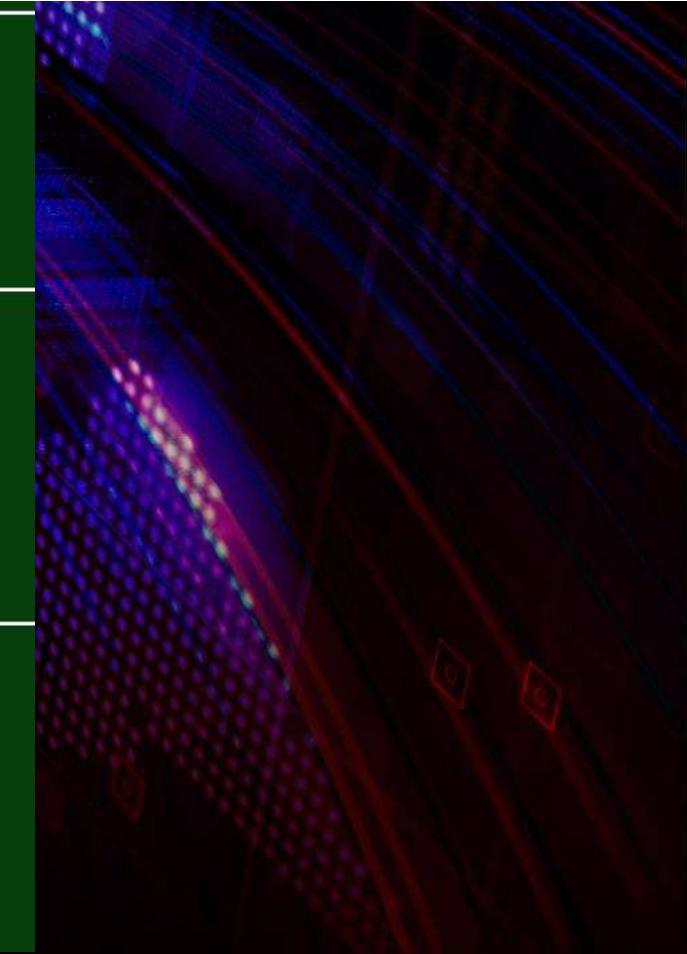
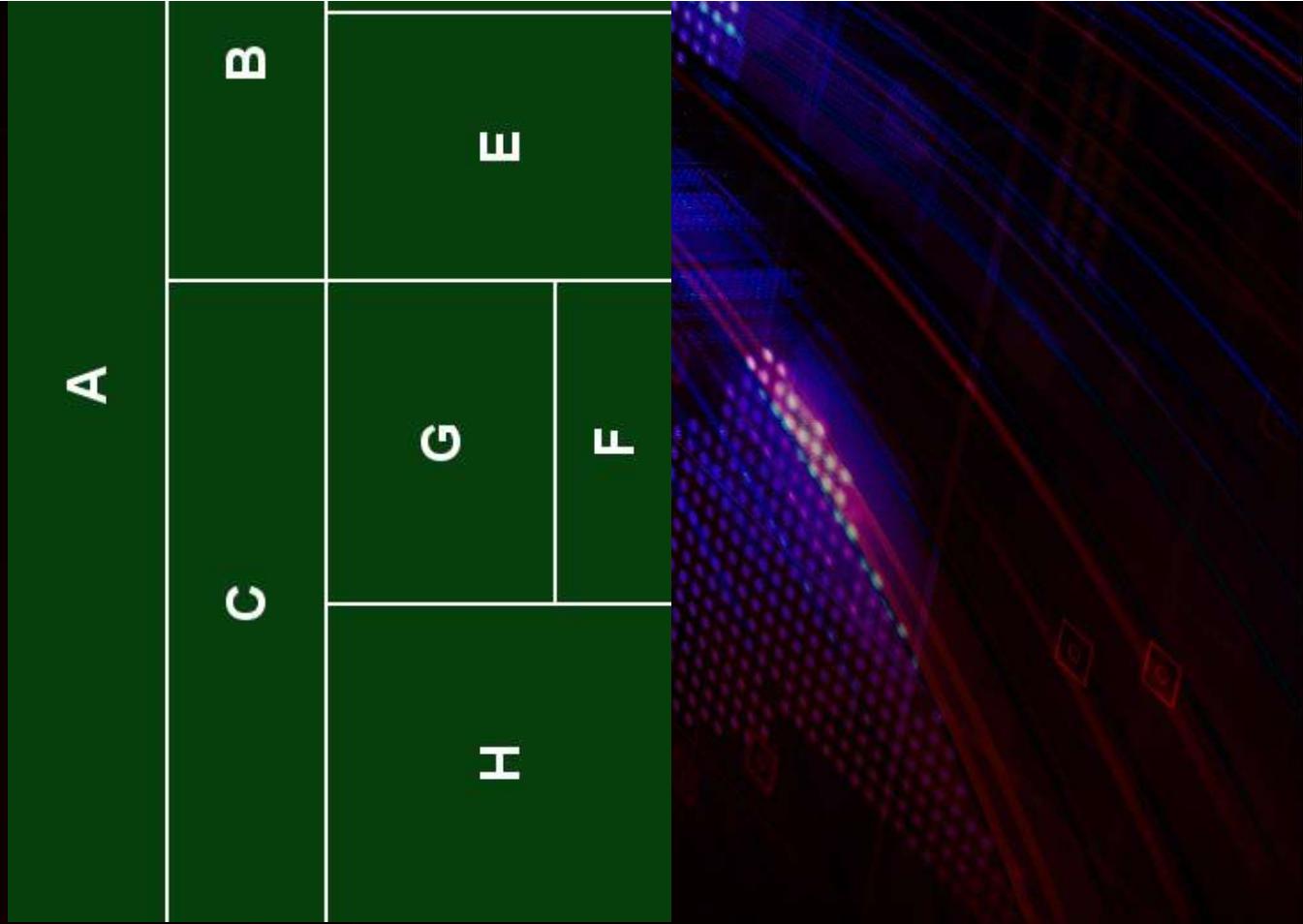
- Apresentar informações relacionais / hierárquicas
- Conjunto de retângulos aninhados
- Cada retângulo pode ser subdividido:
 - Área proporcional ao tamanho
 - Uso eficiente do espaço
- Representação de valores nominais ou categóricos

MAPA DE ÁRVORE



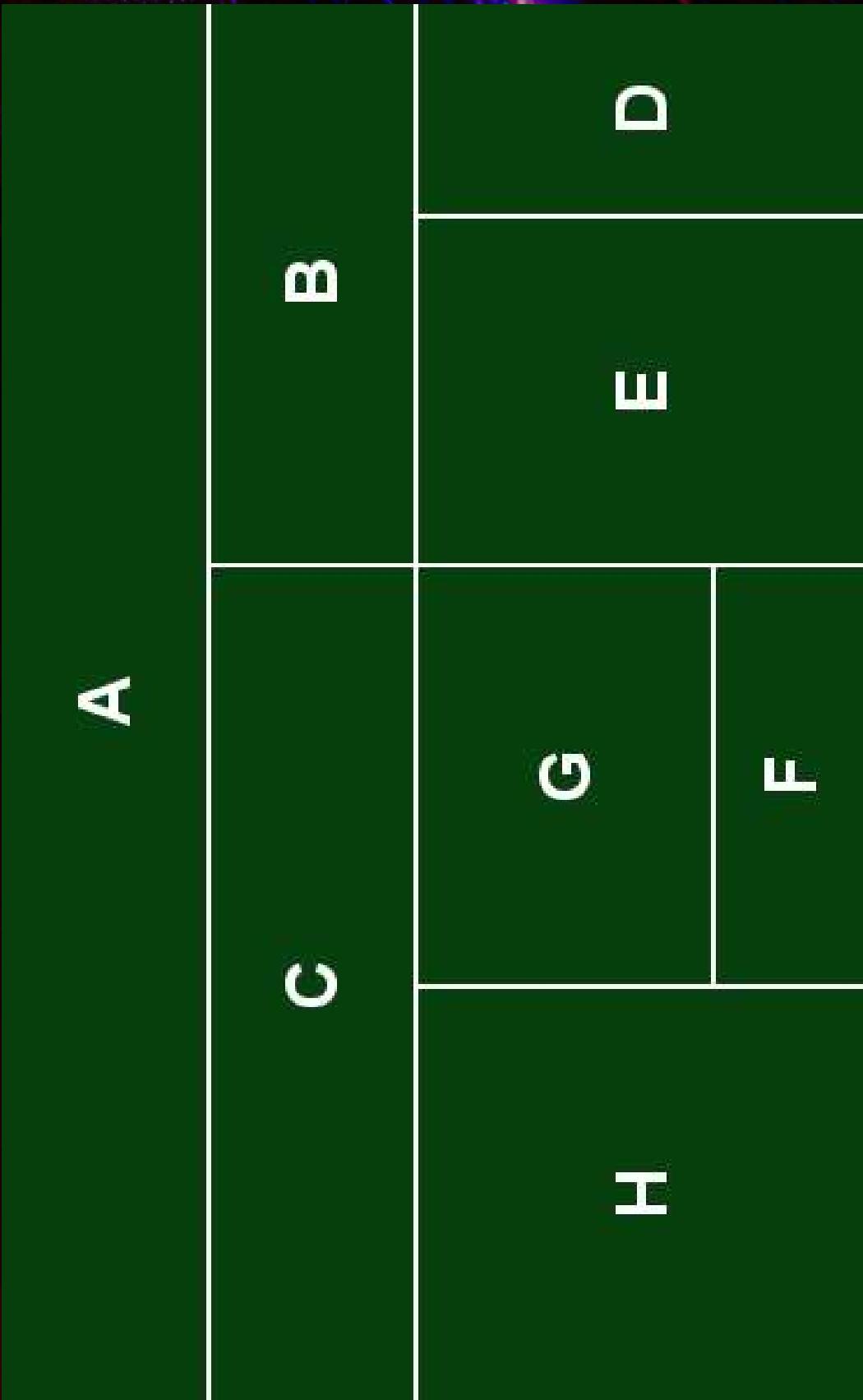
Fonte: Milani et al., 2020

MAPA DE ÁRVORE



Fonte: Milani et al., 2020

MAPA DE ÁRVORE



Fonte: Milani et al., 2020

MAPA DE ÁRVORE- NEWSMAP.JS

Ifix fecha sessão com ganhos de 0,10%; FII é destaque de alta

| | | | | | | | |
|--|---|---|--|---|---|--|---|
| Netflix e Amazon são usadas por golpistas: veja como fugir dessa cilada | Crescimento da inflação argentina deixa iPhone com preços inacreditáveis | Teste de estresse leva Intel Core i9-13900K a 100°C com consumo de 420W | Crescimento da inflação argentina deixa iPhone com preços inacreditáveis | Peso da aprovação da PEC dos Benefícios na disputa. eleitoral é tema de debate na CNN VÍSAO | Congresso derruba vetos e retoma compensações a estados pela redução do ICMS dos combustíveis | Morre Ivana Trump, ex-esposa do ex-presidente dos EUA Donald | BA: 2.75; entenda a subvariação da ômicron que já chegou a 10 países e tem 'apelido' criado na internet |
| Petróleo fecha em queda, pressionado por dólar e expectativas de postura agressiva do BC americano | Ibovespa: Sem commodities, qual o fundo do poço? | Novidade! O Nubank e como elas vão funcionar? | Novidade! O Bill Gates que são as caixinhas do Nubank e como elas vão cresceram tanto nos últimos 10 anos? | Por que as telas celulares cresceram tanto a fortuna para sua fundação | Visava indigenista morio no Amazonas diz esperar narratação de Bolsonaro, Mourão e da Fimapi | Mulher de 7 universidades brasileiras entre as 10 melhores da América Latina | Bill Gates planeja doar quase toda a fortuna para sua fundação |
| Apresentado no Fenerbahçe, Arão posta 'carta de despedida' ao Flamengo: "Estou na história" | Invasão ao Maracanã pode causar perda de mando de campo para o Flamengo | Corinthians se reapresenta; time pode ter novidades contra o Ceará | "Desculpa o palavrão": ator da Globo sobre o tom e expõe que Caio Castro 'arrregou' em luta contra ele | "Batimento cardíaco" de rádio é detectado a bilhões de anos-luz da Terra | Peixe-remo gigante de quase 6 metros de comprimento é encontrado no Chile; assista | O videogame usado por médicos para tratar hiperatividade | Alimentos que podem ser consumidos no café da manhã de diabéticos |
| Cássio pede contato de Marcos Leonardo para agradecer apoio após agressão em Santos x Corinthians | Corinthians acerta liberação definitiva de Vitiinho com o FC Arouca; veja como ficou a operação | Após briga, Otaviano Costa parcela conta de água de R\$ 374 mil em 60 vezes | Globo vende Simaria bloquia Paramount no Instagram e ameaça chamar polícia contra saiba o valor | Podemos ter encontrado outro mamífero – além de ônibus – que pratica agricultura. | Covid: dor de garganta agora é o principal sintoma; confira os outros | Covid: dor de garganta agora é o principal sintoma; confira os outros | União da Saúde da Família do bairro Alto informa que não está entregando a população |

World

Technology

Business

Sports

Science

Health

Fonte: newsmap.js

NUVEM DE PALAVRAS

- Quantificar a ocorrência de termos
- Valores nominais
- Identificar palavras-chave:
 - Textos
 - Mídias sociais
- Pesquisas de opinião
- Buscas na internet

NUVEM DE PALAVRAS



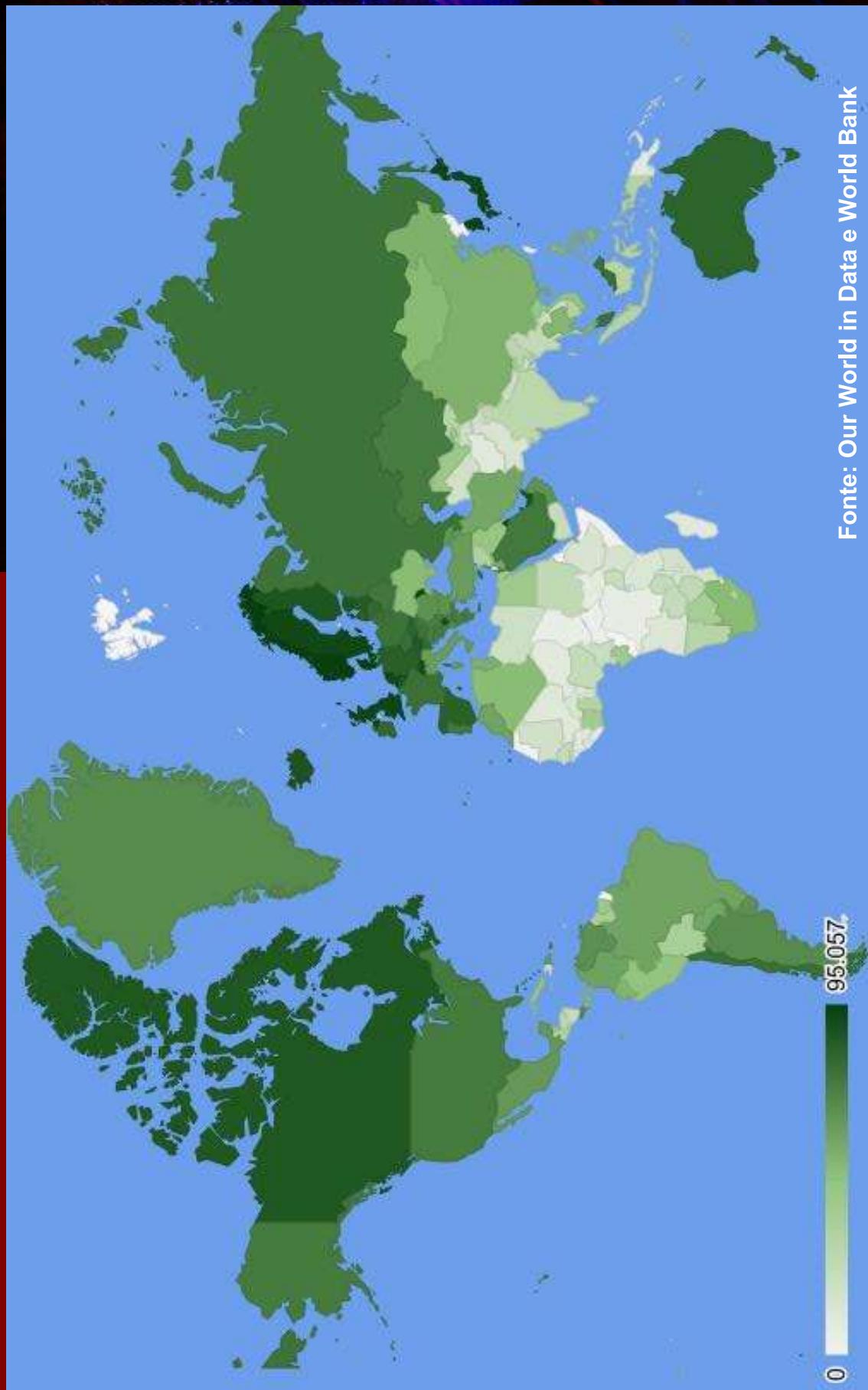
Fonte: wordclouds.com

MAPA COROPLÉTICO

- Apresentar informações geolocalizadas
- Diferentes divisões geográficas:
 - Países, continentes, estados, cidades, bairros...
- Representação de proporções:
 - Valores numéricos / nominais
 - Dados normalizados (magnitude)
- Interatividade

MAPA COROPLÉTICO

Popula
com ac
à Intern
por paí



MAPA DE CALOR OU DENSIDADE

- Apresentar informações geolocalizadas
- Diferentes divisões geográficas:
 - Países, continentes, estados, cidades, bairros...
 - Concentração de algum fenômeno

MAPA DE CALOR OU DENSIDADE



Mapa e Filtros

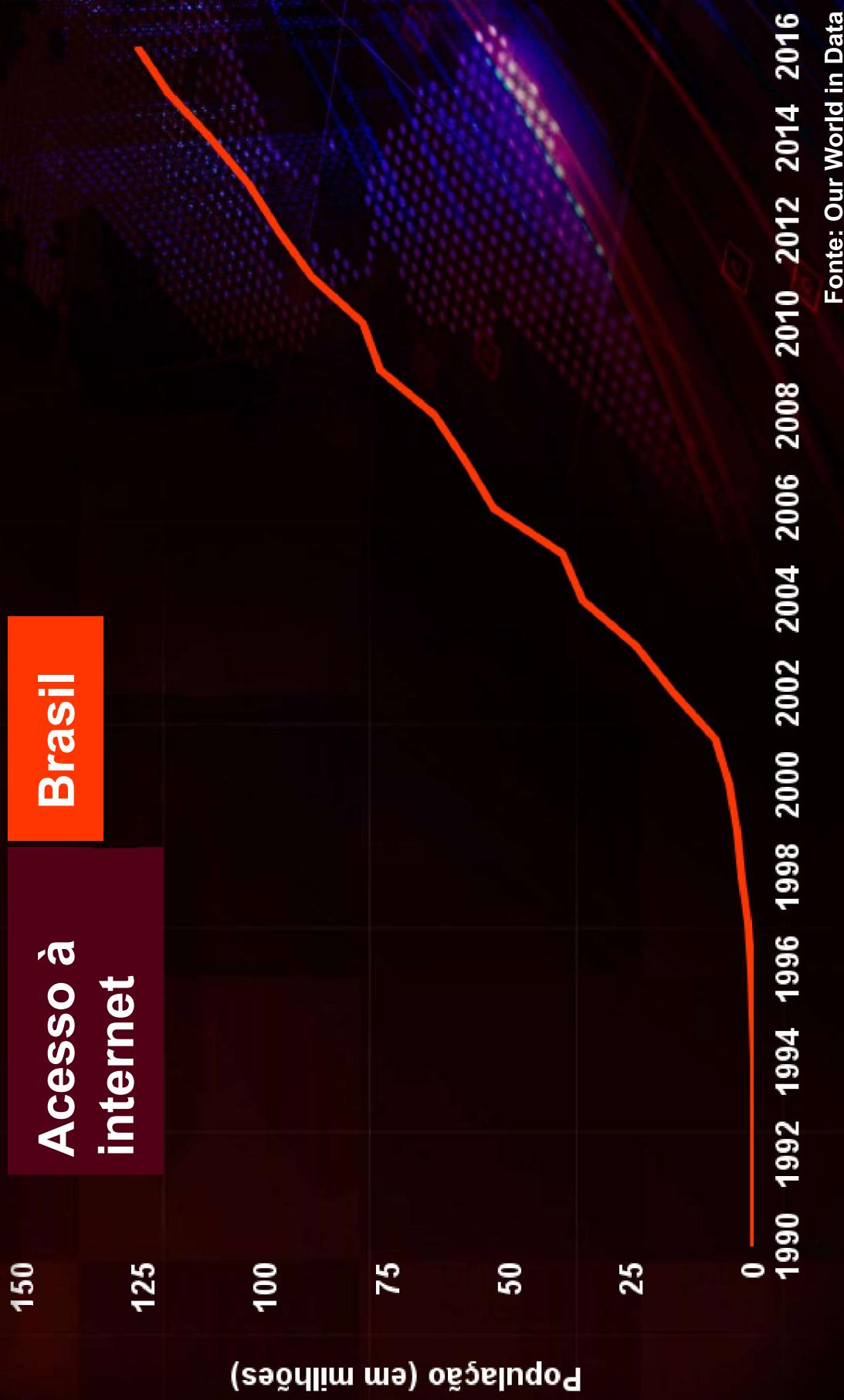


Fonte: <https://interscity.org/health/>

SÉRIE TEMPORAL

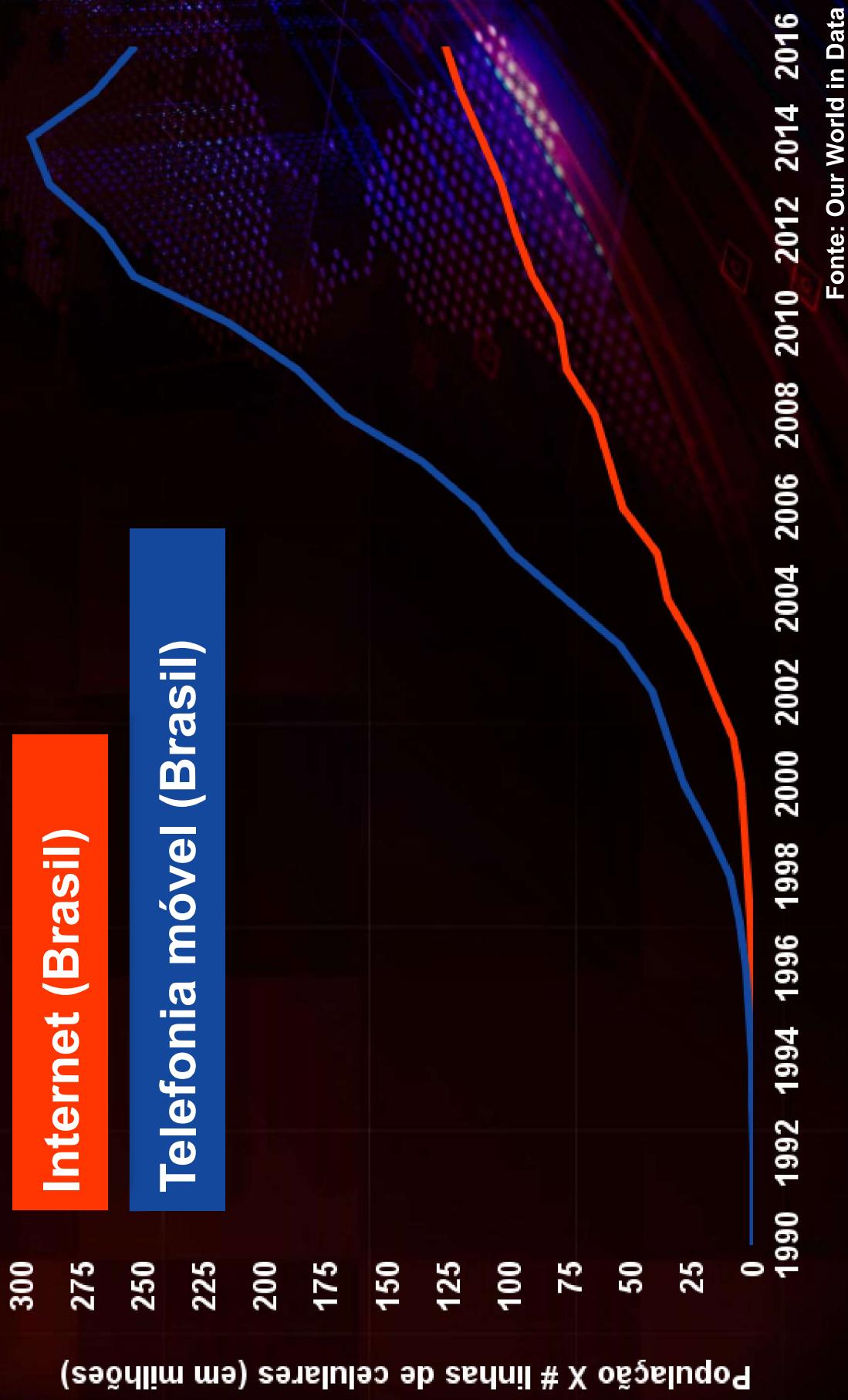
- Apresentar informações ao longo do tempo:
 - Padrões, sazonalidades, evolução histórica
- Diferentes períodos:
 - Ano, mês, dia, hora...
 - Valores numéricos

SÉRIE TEMPORAL



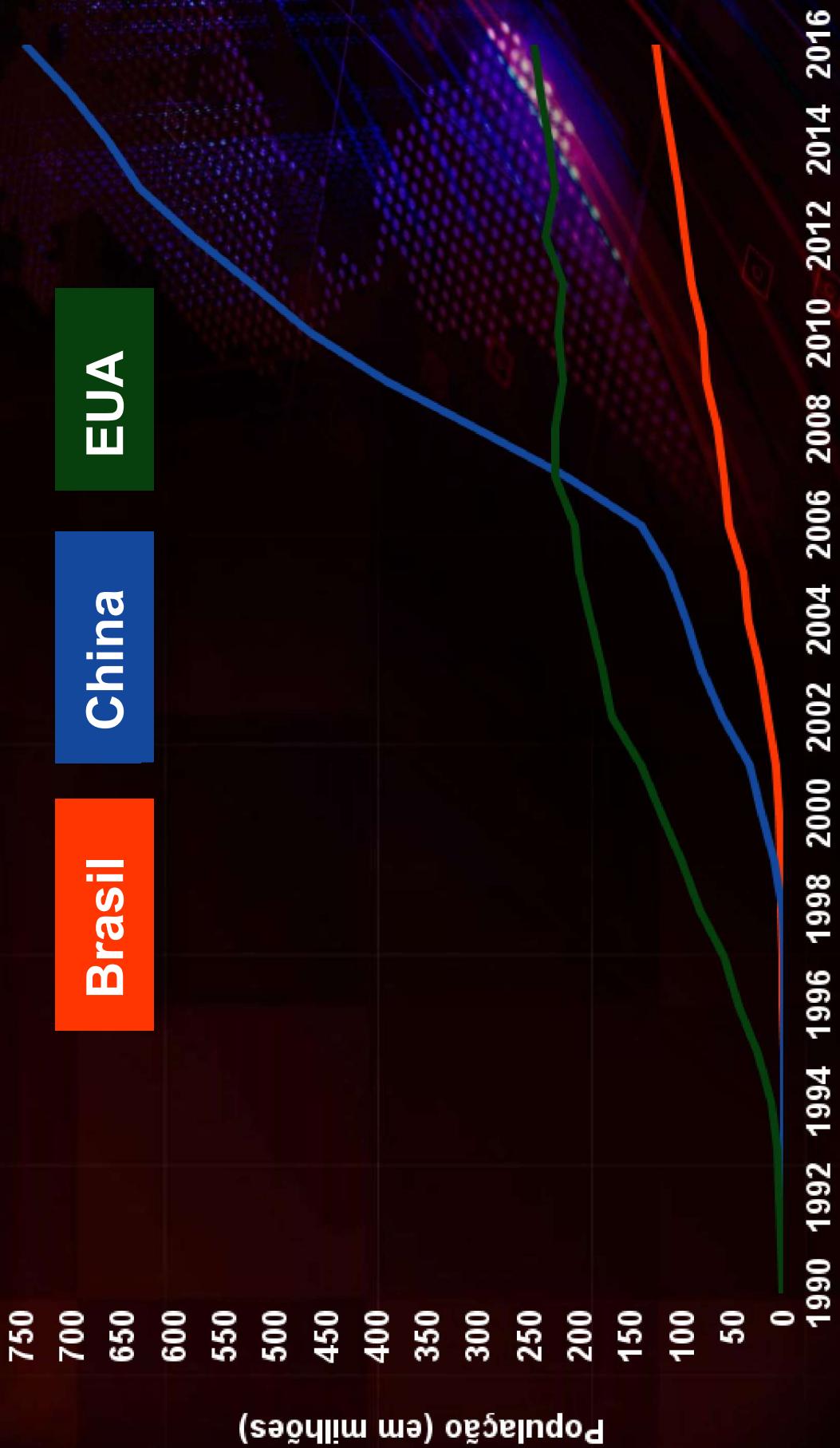
Fonte: Our World in Data

SÉRIE TEMPORAL



Fonte: Our World in Data

SÉRIE TEMPORAL



Fonte: Our World in Data

DIAGRAMA DE CORDA

- Mostrar a relação entre objetos
- Similar a grafos:
 - Nós e arestas
- Valores numéricos / nominais
- Entendimento do todo

DIAGRAMA DE CORDA

Migração
global

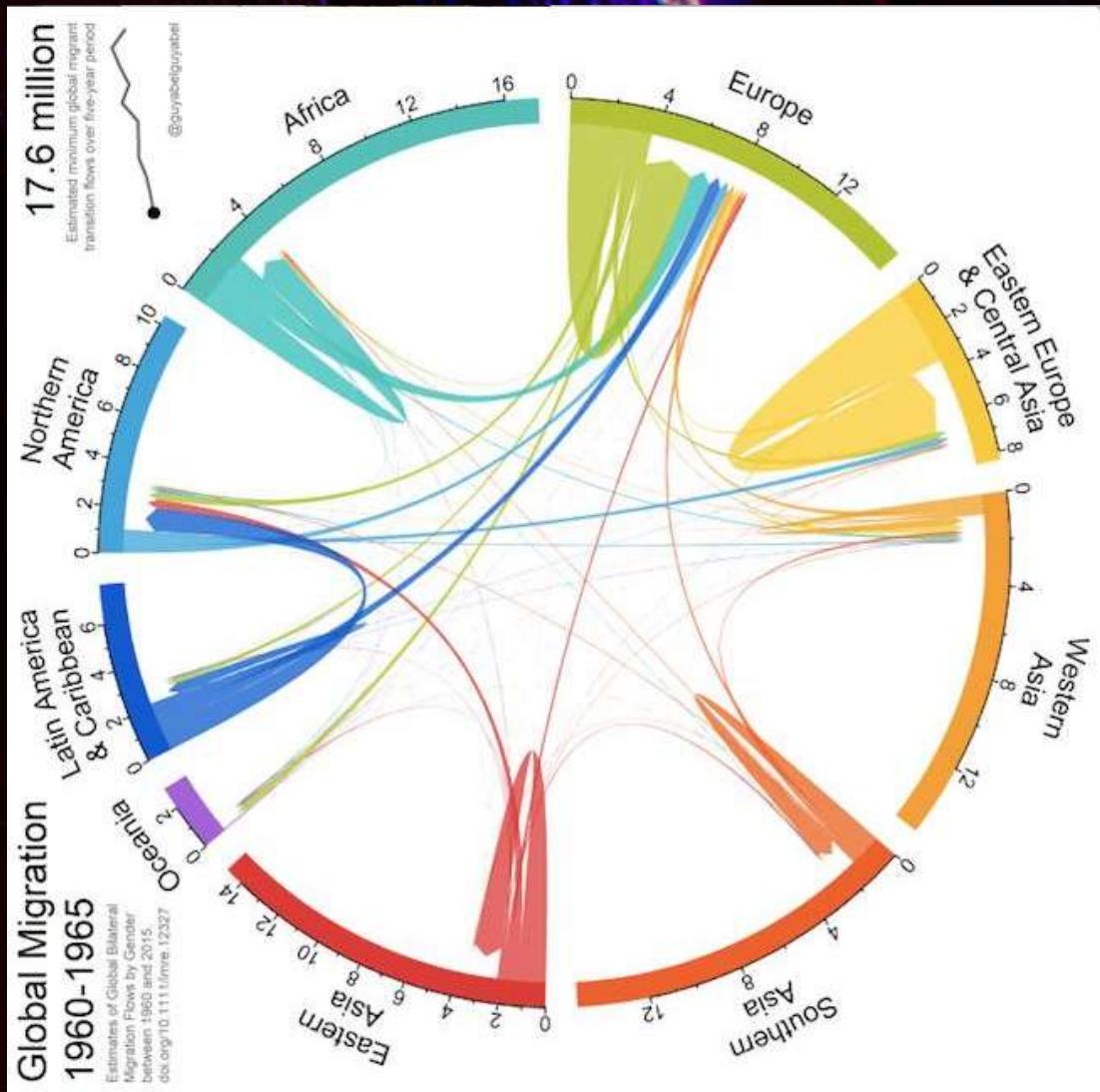
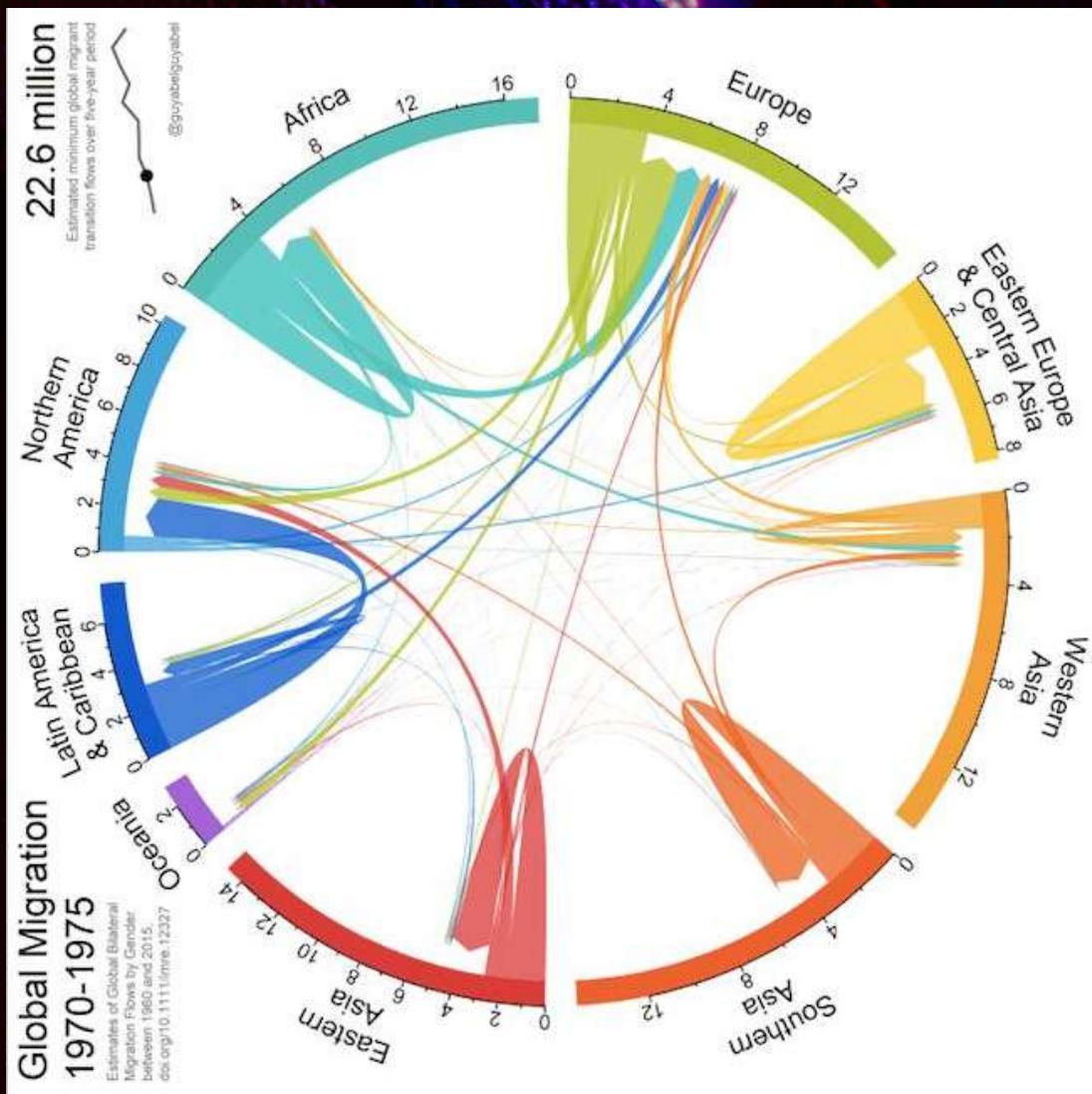


DIAGRAMA DE CORDA

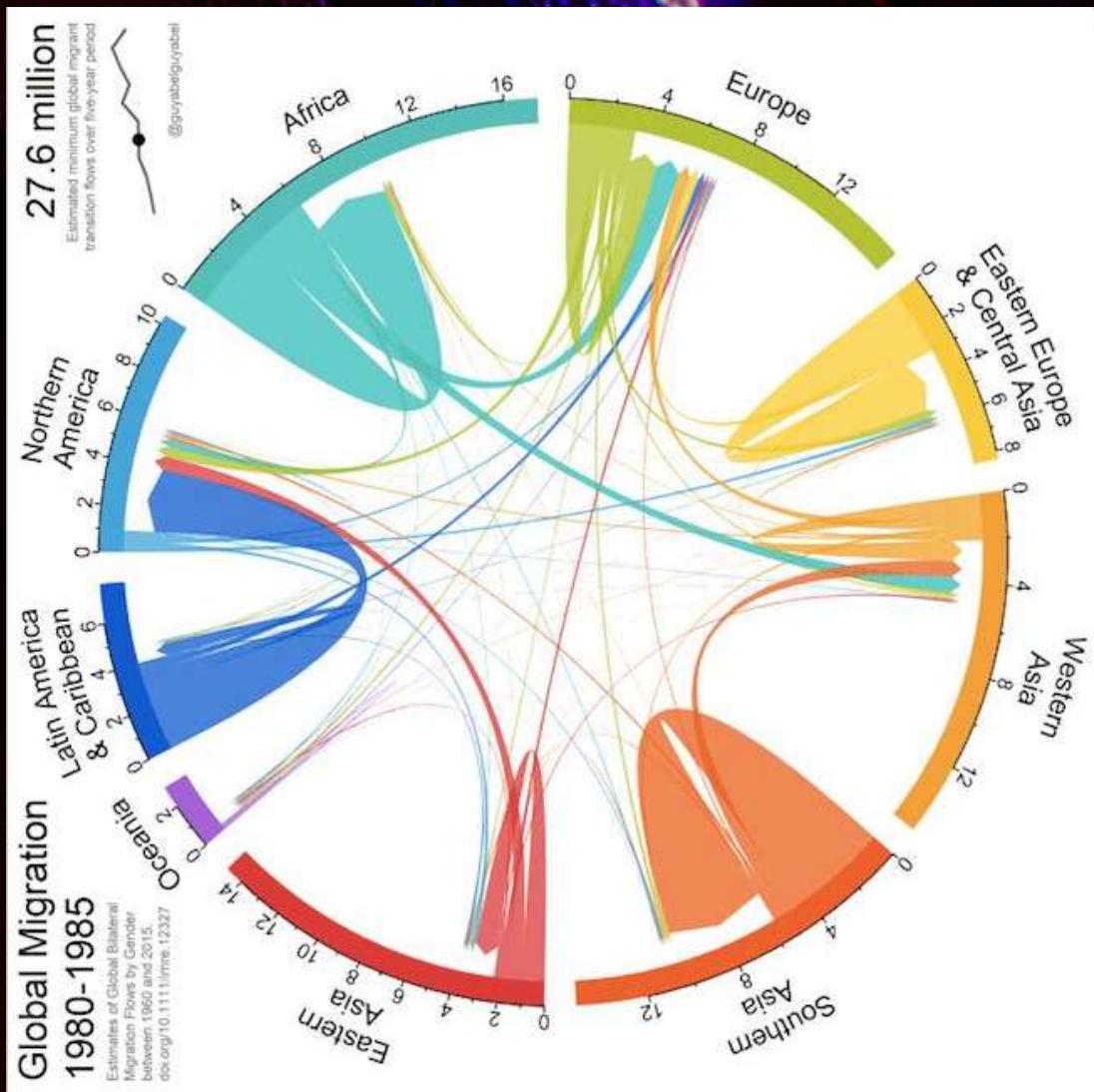
Migração global



**Fonte: David Smith,
R-Bloggers**

DIAGRAMA DE CORDA

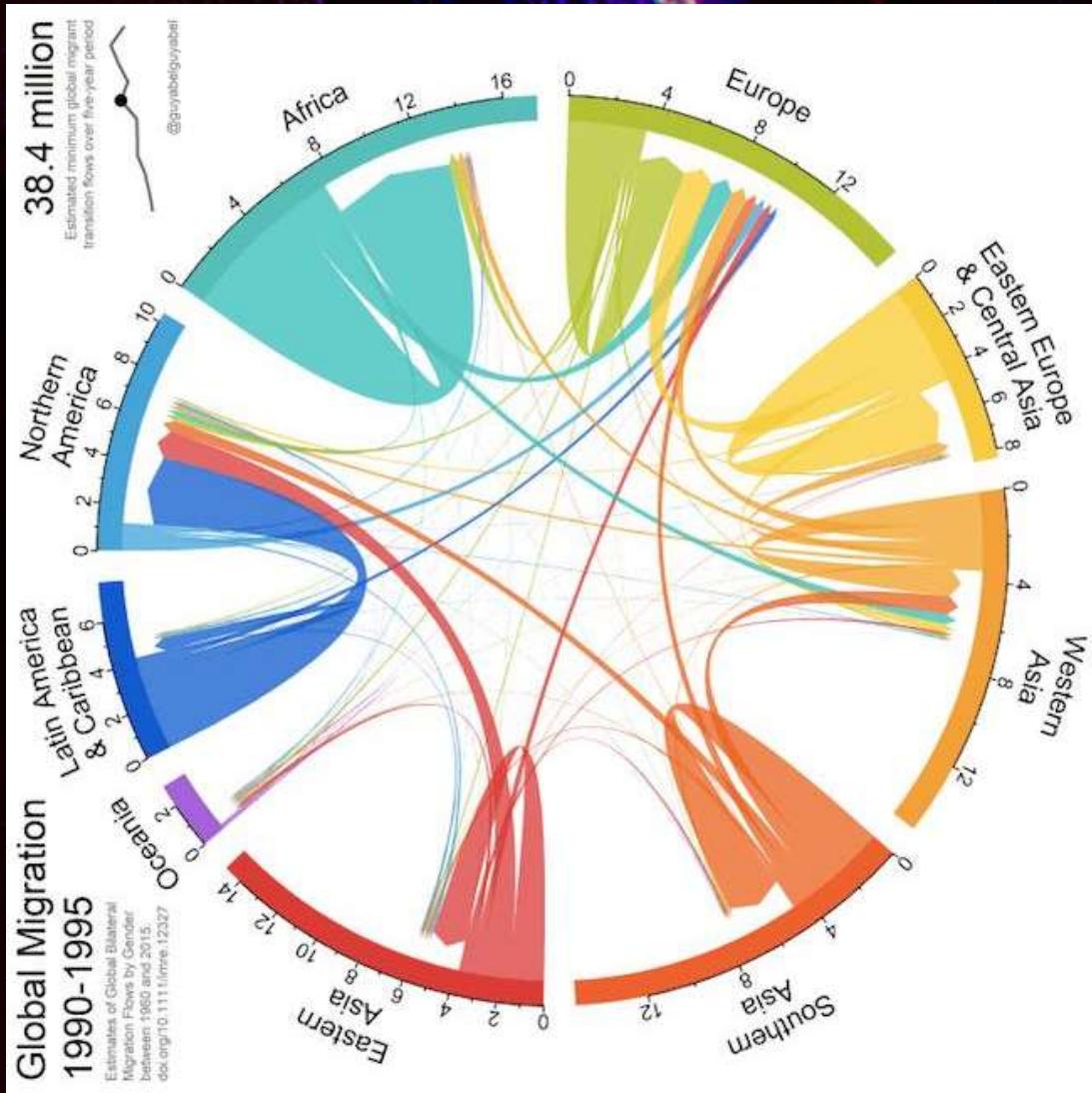
Migração global



**Fonte: David Smith,
R-Bloggers**

DIAGRAMA DE CORDA

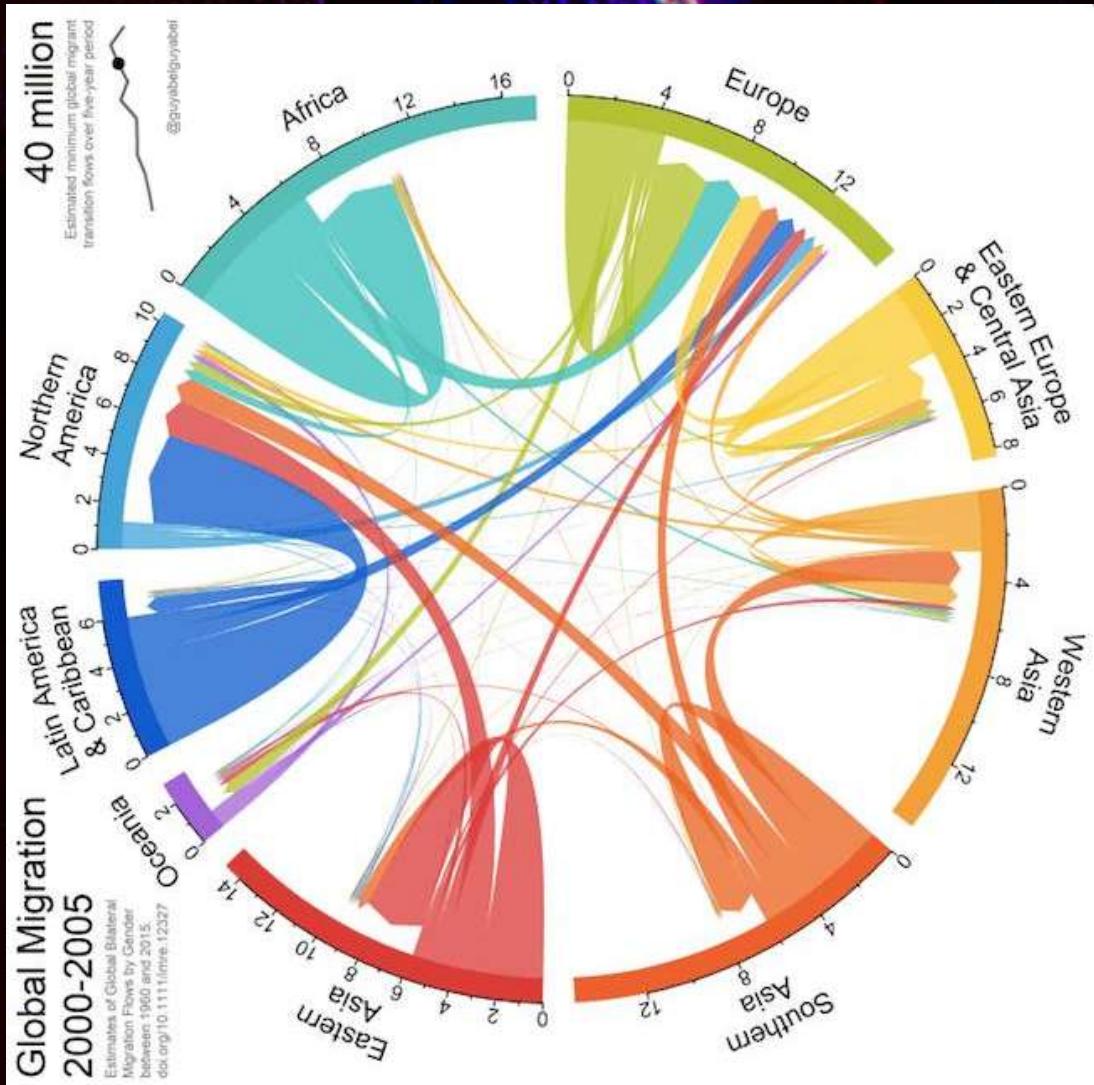
Migração
global



Fonte: David Smith,
R-Bloggers

DIAGRAMA DE CORDA

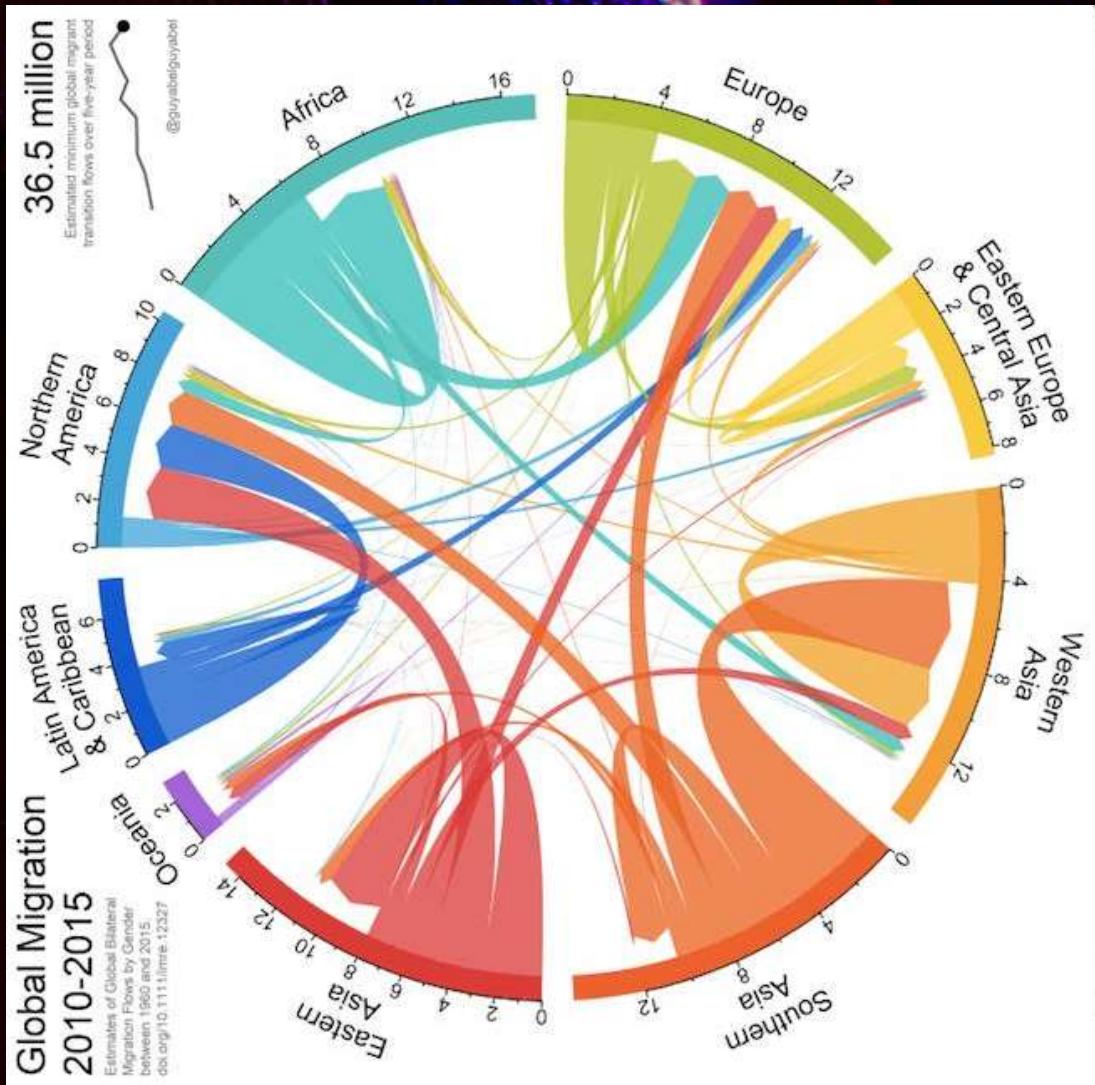
Migração global



**Fonte: David Smith,
R-Bloggers**

DIAGRAMA DE CORDA

Migração global



**Fonte: David Smith,
R-Bloggers**

REFERÊNCIAS

Visualização de dados. Capítulo: *Visualização de dados em big data.*
Alessandra M. Paz Milani et al. Editora SAGAH, 2020.

Newsmap.js 2022: notícias mais acessadas: <https://newsmap.js.herokuapp.com>

Our World in Data: acesso à Internet por país: <https://ourworldindata.org/internet>

World Bank: dados da população por país e ano: <https://data.worldbank.org/indicator/SP.POP.TOTL>

Health Dashboard: <https://interscity.org/health/>

Migração global animada com R: <https://www.r-bloggers.com/2018/06/global-migration-animated-with-r/>