

דו"ר

The image shows the Hebrew word 'דו"ר' (Dr.) in a stylized, modern font. The letters are white with a slight shadow effect. The 'ד' (Dalet) is on the left, followed by 'ו' (Vav) and 'ר' (Resh). The 'ר' is particularly stylized, with a vertical stem and a curved base. There are four orange dots: one on the top of the 'ו', one on the top of the 'ר', one at the bottom of the 'ר', and one at the top of the vertical stem of the 'ר'.



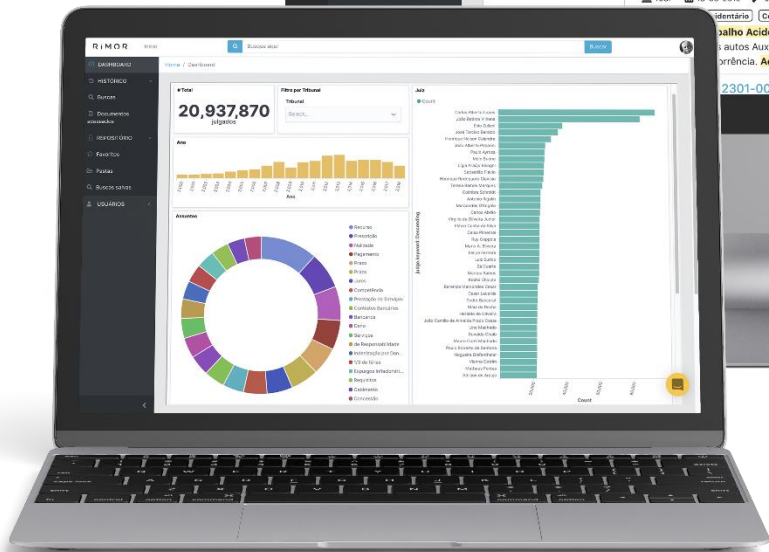
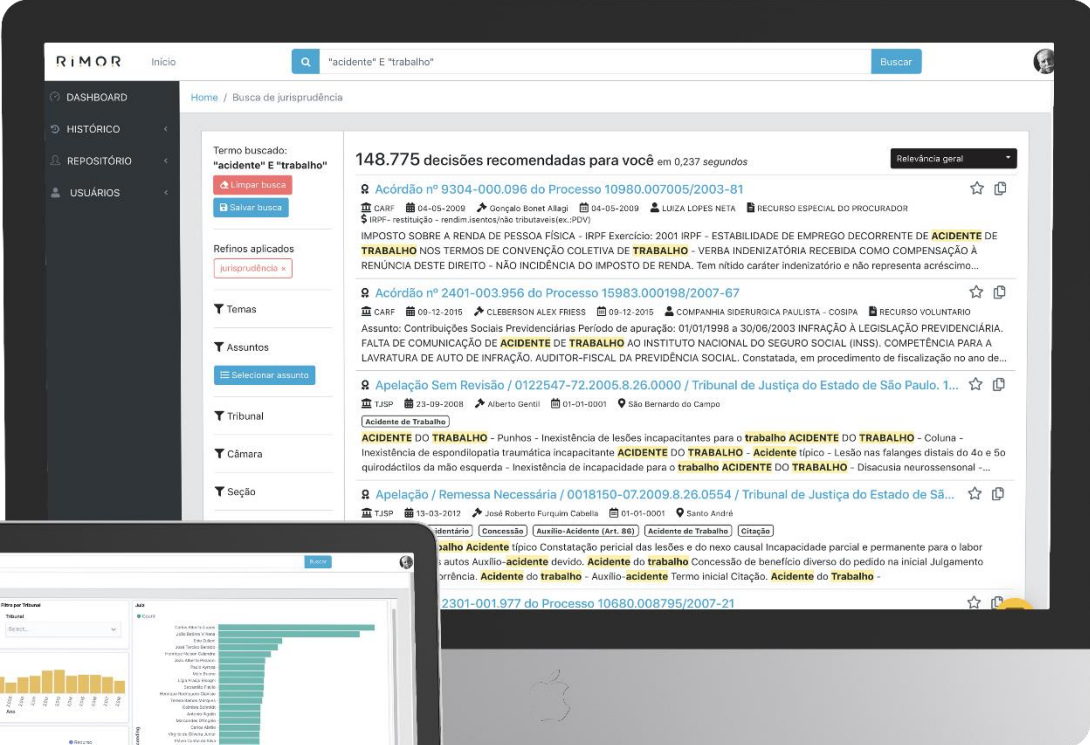
## Ciência de dados no universo jurídico

Alavancando a compreensão do complexo sistema judiciário brasileiro

# Quem sou eu?

- **Empreendedor. Fundador e CEO da JUIT** (legaltech focada em pesquisa jurídica e jurimetria sob demanda)
- **Bacharel em Direito / MBA em Big Data / Extensão em Ciência de Dados aplicado ao Direito**
- Anteriormente: atuação em escritório de Direito Digital, inaugurando uma área de análise de dados jurídicos, com posterior passagem pelo mercado financeiro atuando com FIDCs e ativos estressados.
- No campo acadêmico: atividades como palestrante e professor convidado (USP, Insper, FGV).



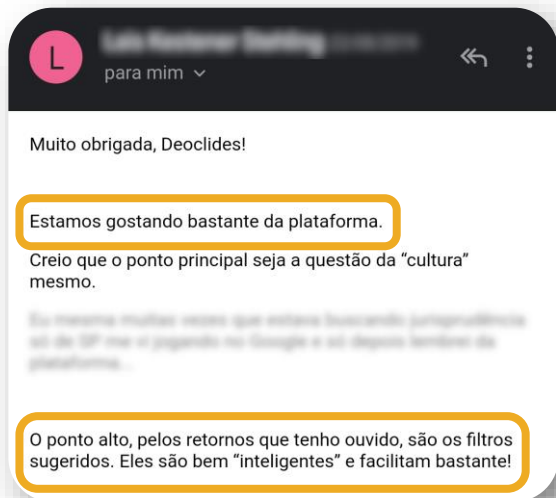


# Case

Situação de advogado tributarista procurando jurisprudência do CARF e TJSP

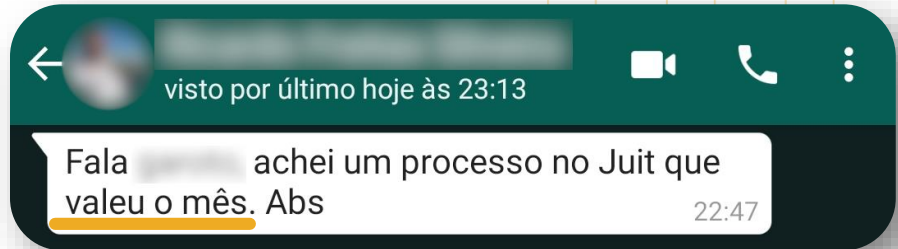
**SEM O JUIT RIMOR: 4 meses**

**COM O JUIT RIMOR: 15 minutos**



AB

To gostando pelo pouco que já mexi aqui! Achei umas juris que não encontrava há 4 meses pesquisando!



# Time



**Deoclides**  
Founder & CEO

- > Bacharel em Direito; MBA em Big Data e extensão em Ciência de Dados para o Direito
- > 7 anos de experiência com tecnologia para advogados e mercado financeiro (FIDCs e NPLs)



**Marcio**  
Co-founder & CTO

- > Bacharel Ciências da Computação; MBA em Big Data e Engenharia de Software
- > 17 anos de experiência (11 anos com engenharia de software na IOB Síntese – produtos com foco no mercado jurídico)



**João Marcelo**  
Fullstack Dev

- > Bacharel em Ciências da Computação
- > Experiência com frontend, backend e devops
- > Entusiasta de machine learning



**Terso**  
Backend Dev

- > Bacharel em Ciências da Computação
- > Ingestão de dados via web scraping e APIs
- > Manipulação de dados em bancos relacionais e não-relacionais



**Elie**  
P.O.

- > Doutorando e Mestre em Direito pela USP
- > Professor de Direito em diversas instituições de ensino
- > 10 anos de experiência na advocacia e ensino jurídico



# Premiações e reconhecimentos



Vencedora na categoria Dados Não-Estruturados da **Neo4j Graphie Award** (NYC, 2018)

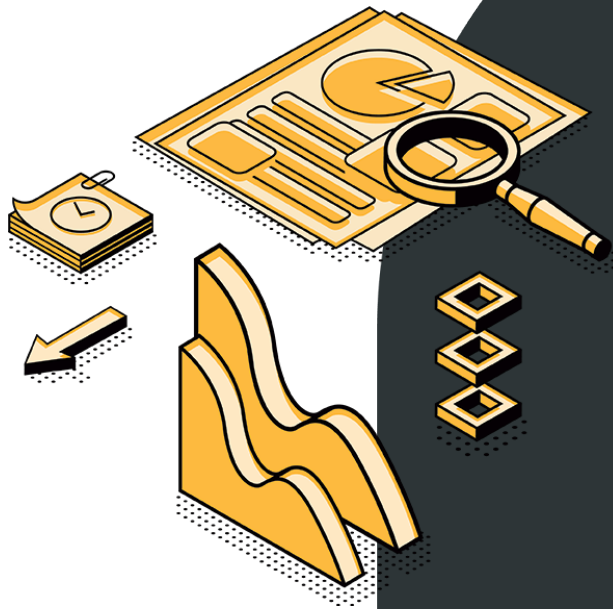


Vencedora como melhor legaltech do Brasil no **Legaltech Venture Days** (São Paulo, 2019)



Aprovação no programa de **aceleração do C6 Bank**, dentre mais de 360 startups





**Como chegamos  
nestes resultados?**

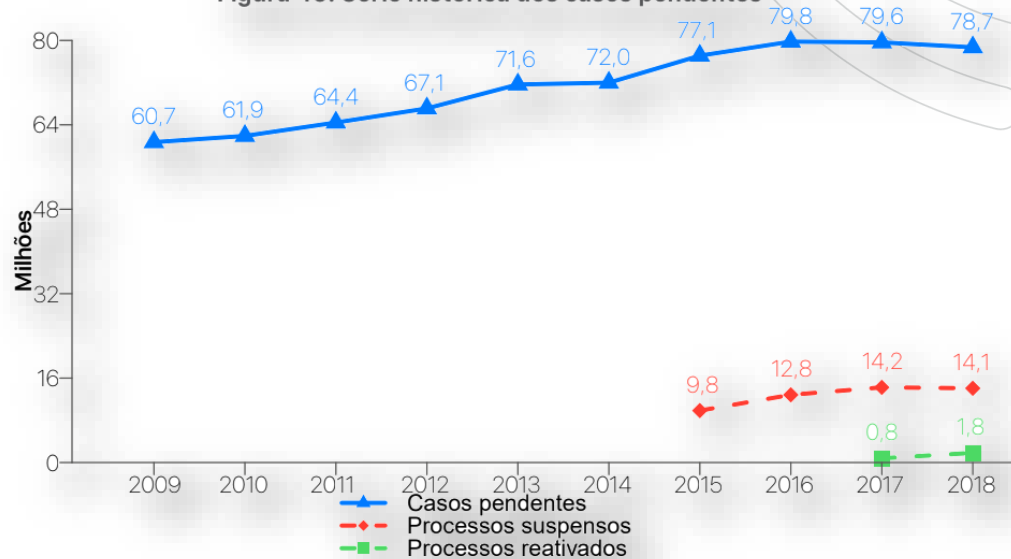


# **As dificuldades de um Tribunal afogado em demandas**

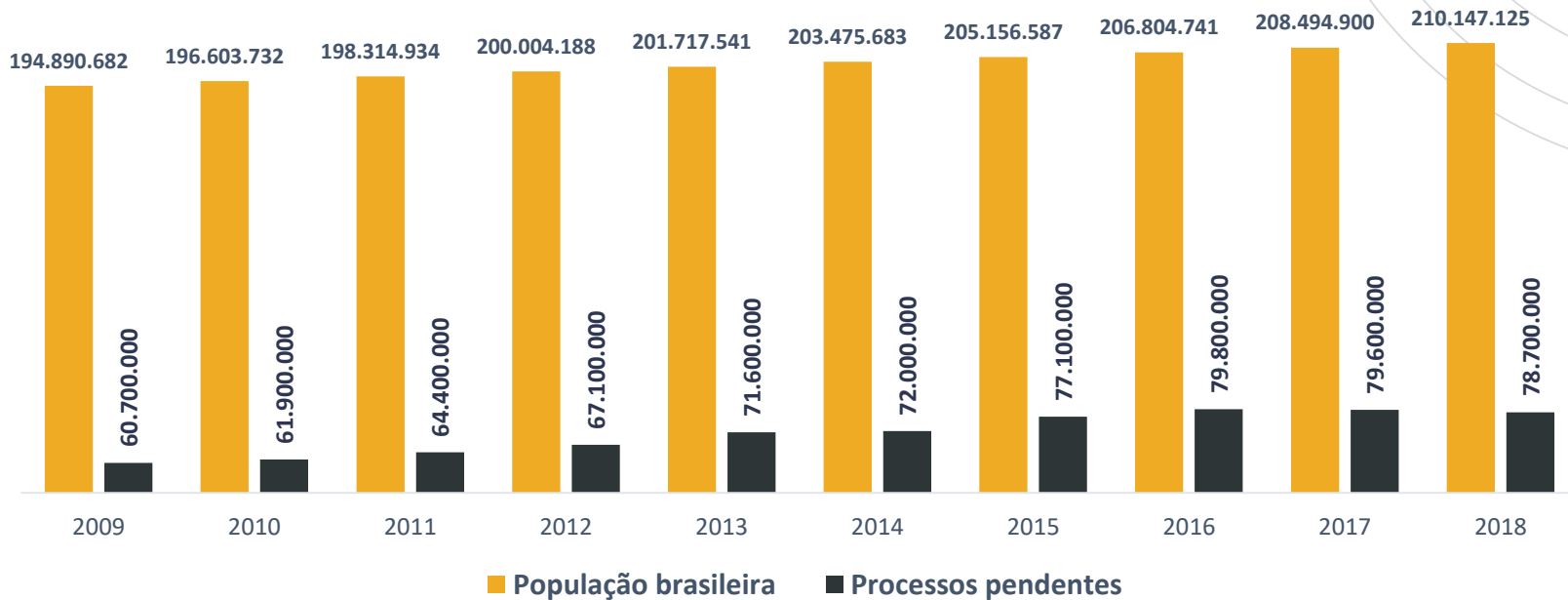
# Relatório Justiça em Números 2019

- 78,7 milhões de processos em trâmite
- 32,4 milhões de decisões terminativas proferidas

Figura 45: Série histórica dos casos pendentes



# População brasileira x nº de processos



+ 7,8%

+ 29,6%



**“Mesmo que não houvesse ingresso de novas demandas e fosse mantida a produtividade dos magistrados e dos servidores, seriam necessários aproximadamente 2 anos e 6 meses de trabalho para zerar o estoque”.**

**Conceito “Tempo de Giro do Acervo”**

# Quais são as principais dores do setor?

## Dor

Gestão administrativa e de processos (ERP)

Automação e gestão de documentos

Redes de profissionais

Resolução de conflitos online

Conteúdo jurídico, educação e consultoria

**Extração e monitoramento de dados públicos ← JUIT Rimor**

**Jurimetria e Legal Analytics ← JUIT Rimor**



# Jurimetria

Conceitos acadêmicos e dilemas  
práticos

# JURIMETRIA – Marcelo Guedes Nunes

## Definição

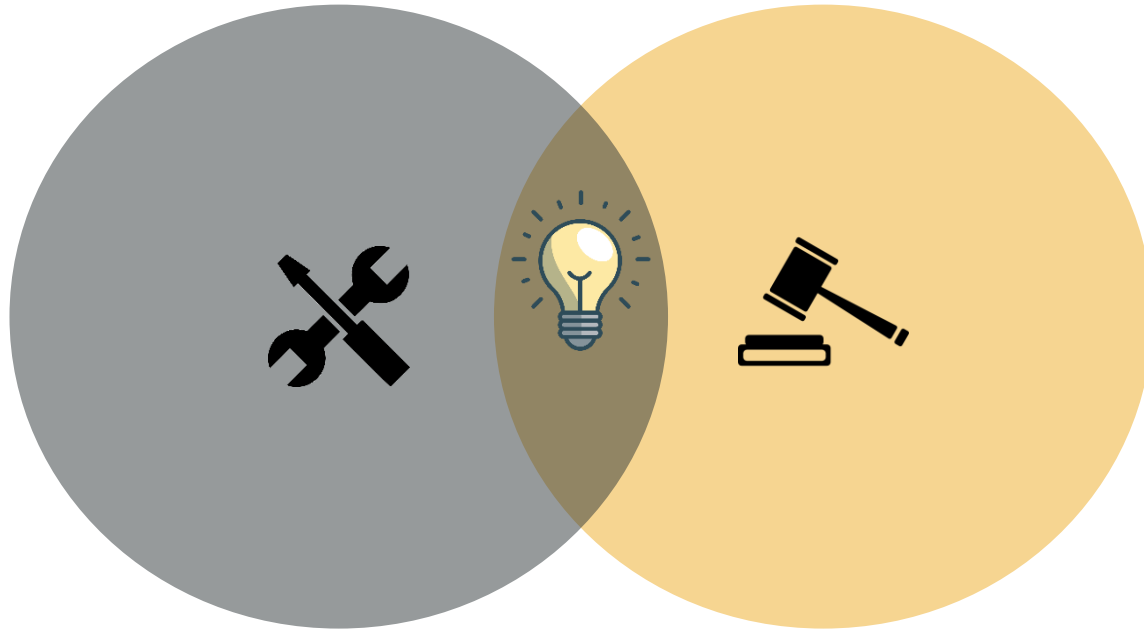
- “Disciplina do conhecimento que utiliza a metodologia estatística para investigar o funcionamento de uma ordem jurídica.”
- “Visa explicar como o Direito funciona, sendo desenvolvida por meio de inferências. Representando a conjunção de um método (estatística) com um objeto (norma jurídica)”.

## Finalidade

- Aumentar nível de segurança jurídica, diminuindo incertezas.



# CONJUNÇÃO MÉTODO + OBJETO





# Perspectivas da Jurimetria

## OBJETIVA

- Refere-se ao objeto ‘norma jurídica’; não a norma isoladamente considerada, mas a norma jurídica articulada.
- Considera-se, de um lado, o efeito do comportamento dos reguladores, e de outro, a causa no comportamento de seus destinatários.
- É o “o que?”

## METODOLÓGICA

- Usa a estatística para restabelecer um elemento de causalidade e investigar fatores (sociais, econômicos, geográficos, éticos, dentre outros) que influenciam o comportamento dos agentes jurídicos
- É o “como?”



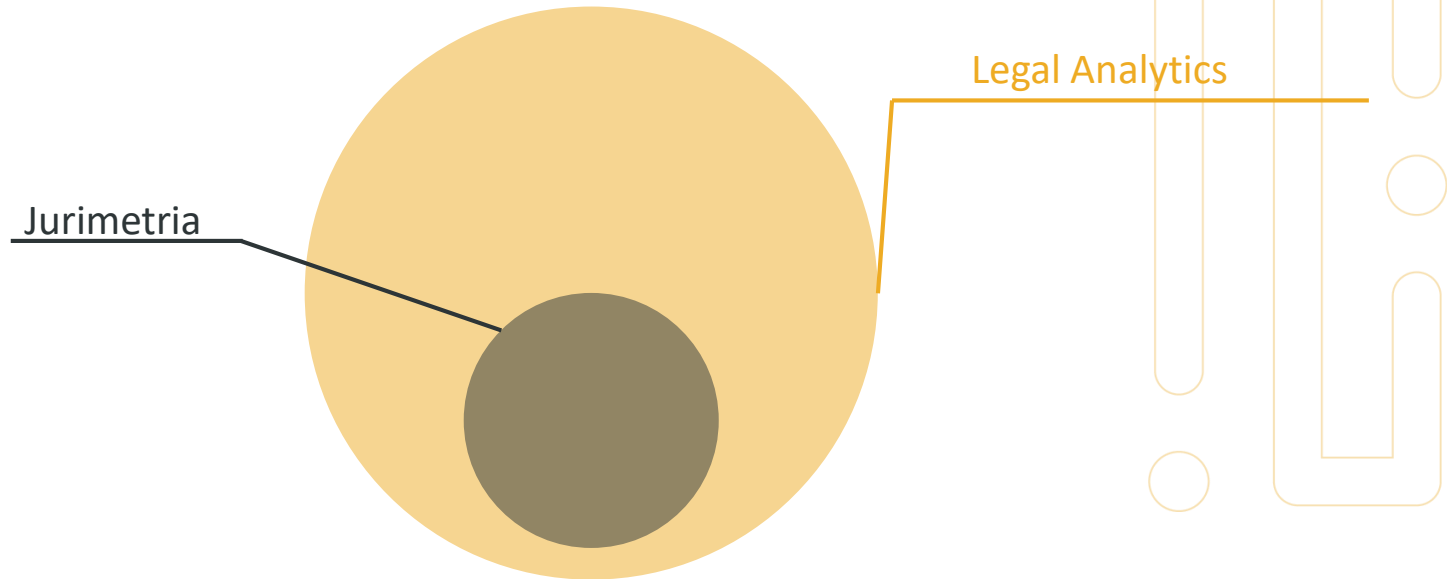
# 7 passos da metodologia estatística

1. Definição do problema
2. Planejamento
3. Coleta de dados
4. Limpeza de dados
5. Apuração
6. Apresentação
7. Análise dos dados



# Jurimetria vs Legal Analytics

**Legal Analytics:** Análise exploratória quantitativa e qualitativa de dados oriundos de processos e Tribunais, do qual Jurimetria é uma parte.





# Aplicações práticas

- ✓ Criação de estratégias jurídicas;
- ✓ Busca por resolução alternativa de conflitos (celeridade e economia processual) – ODRs;
- ✓ Limpeza de carteira de processos – ERPs;
- ✓ Estudo jurisprudencial;
- ✓ Compreensão das ‘endemias sociais’ para desafogamento dos Tribunais. Exemplo: recursos repetitivos (CPC, art. 1036);
- ✓ Outros.



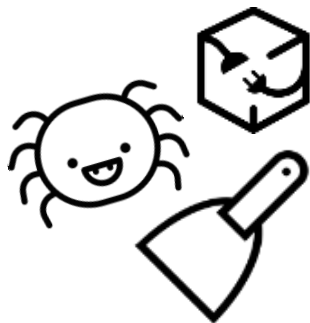
**O que utilizamos de  
tecnologia?**



*“In God we trust.  
All others must bring DATA.”*

Dr. William Edwards Deming, *estadístico*

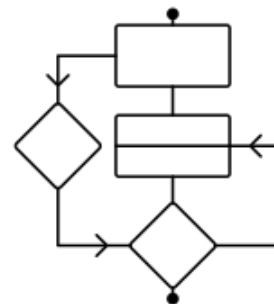
# Backend



API  
Crawlers  
Scrapers  
Parsers



Banco de dados  
SQL  
NoSQL



Processamento  
dos dados  
Limpeza  
Contagem



# Backend – crawlers e scrapers

## Crawler

Robôs que varrem websites indexando novas páginas.

## Scraper

Robôs que salvam as informação das páginas como grandes objetos (HTML, usualmente).

## Parser

Robôs que dividem os dados em partes menores e os armazenam em outra estrutura de dados.

## Googlebot

Googlebot é o robô de rastreamento da Web do Google (também chamado de "indexador"). O rastreamento é o processo pelo qual o Googlebot descobre páginas novas e atualizadas para serem incluídas no índice do Google.

O processo de rastreamento do Googlebot começa com uma lista de URLs de páginas da Web, gerada a partir de processos anteriores de rastreamento e aumentada com dados dos [sitemaps](#) fornecidos por webmasters. Conforme o Googlebot visita cada um desses sites, ele detecta os links (SRC e HREF) de cada página e os inclui na lista de páginas a serem rastreadas. Novos sites, alterações em sites existentes e links inativos serão detectados e usados para atualizar o índice do Google.





# Backend – APIs

## APIs

Application Programming Interface - Conector de dados estruturados que permite integração com outras aplicações.

APIs facilitam a construção de aplicações por prover blocos de construção que são montados por desenvolvedores.



```
1  "document_type": "jurisprudência",
2  "locality": "",
3  "authority": "Tribunal Superior do Trabalho. 2ª Turma",
4  "title": "AIRR - 1128-96.2010.5.15.0055 / Tribunal Superior do Trabalho. 2ª Turma",
5  "titleclean": "AIRR - 1128-96.2010.5.15.0055 / Tribunal Superior do Trabalho. 2ª Turma",
6  "date": "21-09-2016",
7  "date_publish": "30-09-2016",
8  "date_register": "01-01-0001",
9  "date_session": "01-01-0001",
10 "description": "AGRAVO DE INSTRUMENTO. RECURSO DE REVISTA INTERPOSTO NA VIGÊNCIA DA LEI Nº13.015/
11 2014.\\n\\nEXECUÇÃO. AUXÍLIO ALIMENTAÇÃO. ABONO ESPECIAL DE ANIVERSÁRIO. LIMITAÇÃO. COISA JULGADA. Não se
vislumbra afronta ao inciso XXXVI do artigo 5º da Constituição Federal, pois a Corte Regional deixou
claro que \\nquando da apresentação dos cálculos pela reclamante, já havia sido proferida decisão na
ação direta de inconstitucionalidade proposta perante o TJSP pela Procuradoria Geral de Justiça do
Estado em face das Leis Municipais 1.624-93 e 2.924-10, tendo sido julgada procedente em 16-11-2011\\n",
desse modo, concluiu o Regional que \\n a coisa julgada sofreu relativização, atraindo a aplicação do
disposto no art. 884 § 5º, da CLT\\n. Agravo de instrumento não provido.",
12 "urn": "https://jurisprudencia-backend.tst.jus.br/rest/documentos/58a0a46f907282b07fd9e9838a365a59",
13 "urls": [
14   "https://jurisprudencia-backend.tst.jus.br/rest/documentos/58a0a46f907282b07fd9e9838a365a59",
15   "http://aplicacao5.tst.jus.br/consultaDocumento/acordao.do?anoProcInt=2014&numProcInt=290186&
dtPublicacaoStr=30/09/2016%2007:00:00&nia=6761220"
16 ],
17 "source": "TST",
18 "judge": "Maria Helena Mallmann",
19 "defendant": "",
20 "appeal_type": "Acórdão",
21 "authority_main": "Tribunal Superior do Trabalho",
22 "authority_section": "",
23 "authority_chamber": "",
24 "authority_class": "2ª Turma",
25 "other_authority": "",
26 "process_id": "1128-96.2010.5.15.0055",
27 "decision_id": "",
28 "state_code": [],
29 "state_description": [],
30 "district": "",
31 "year": 2016,
32 "month": 9,
33 "day": 21,
34 "court_code": "TST",
35 "decision_text": "",
36 "amendment": "AGRAVO DE INSTRUMENTO. RECURSO DE REVISTA INTERPOSTO NA VIGÊNCIA DA LEI Nº13.015/
2014.\\n\\nEXECUÇÃO. AUXÍLIO ALIMENTAÇÃO. ABONO ESPECIAL DE ANIVERSÁRIO. LIMITAÇÃO. COISA JULGADA. Não se
vislumbra afronta ao inciso XXXVI do artigo 5º da Constituição Federal, pois a Corte Regional deixou
claro que \\nquando da apresentação dos cálculos pela reclamante, já havia sido proferida decisão na
ação direta de inconstitucionalidade proposta perante o TJSP pela Procuradoria Geral de Justiça do
Estado em face das Leis Municipais 1.624-93 e 2.924-10, tendo sido julgada procedente em 16-11-2011\\n",
desse modo, concluiu o Regional que \\n a coisa julgada sofreu relativização, atraindo a aplicação do
disposto no art. 884 § 5º, da CLT\\n. Agravo de instrumento não provido.",
37 "amendment_list": [
38   "AGRAVO DE INSTRUMENTO",
39   "RECURSO DE REVISTA INTERPOSTO NA VIGÊNCIA DA LEI Nº13.015/2014.\\n\\nEXECUÇÃO",
40   "AUXÍLIO ALIMENTAÇÃO",
41   "ABONO ESPECIAL DE ANIVERSÁRIO",
42   "LIMITAÇÃO",
43   "COISA JULGADA",
44   "Não se vislumbra afronta ao inciso XXXVI do artigo 5º da Constituição Federal, pois a Corte
Regional deixou claro que \\nquando da apresentação dos cálculos pela reclamante, já havia sido
proferida decisão na ação direta de inconstitucionalidade proposta perante o TJSP pela Procuradoria
Geral de Justiça do Estado em face das Leis Municipais 1.624-93 e 2.924-10, tendo sido julgada
procedente em 16-11-2011\\n", desse modo, concluiu o Regional que \\n a coisa julgada sofreu
relativização, atraindo a aplicação do disposto no art.",
45   "884 § 5º, da CLT\\n.",
46   "Agravo de instrumento não provido."
47 ]
48 }
```



## **Backend – Banco de dados**

# Backend – Banco de dados (tipos)

## Relacionais (SQL) Structured Query Language

Valor  
R\$ 200.000,00

Vara  
24ª VARA CÍVEL

Comarca  
FORTALEZA - FÓRUM CLÓVIS BEVILÁQUA

Data Distribuição  
12/03/2008

Data Início  
11/03/2008

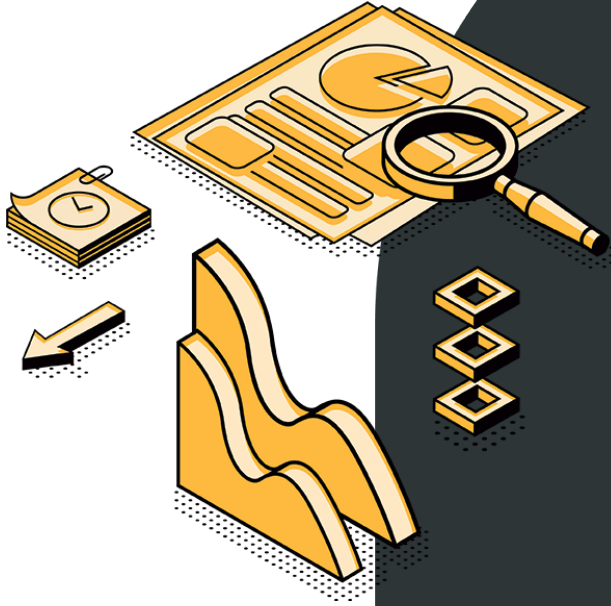
Data Última Atualização  
19/07/2016

Assuntos  
INDENIZACAO POR DANO MORAL

Classes  
INDENIZAÇÃO POR DANO MORAL

## Não-Relacionais (NoSQL) Not Only SQL





## Processamento de dados

Métodos e modelos estatísticos para transformar dados e extrair insights

# NLP – Processamento de linguagem natural

Foco em compreender as entidades que existem em um conjunto de dados

Essencial para criar modelos de machine learning.

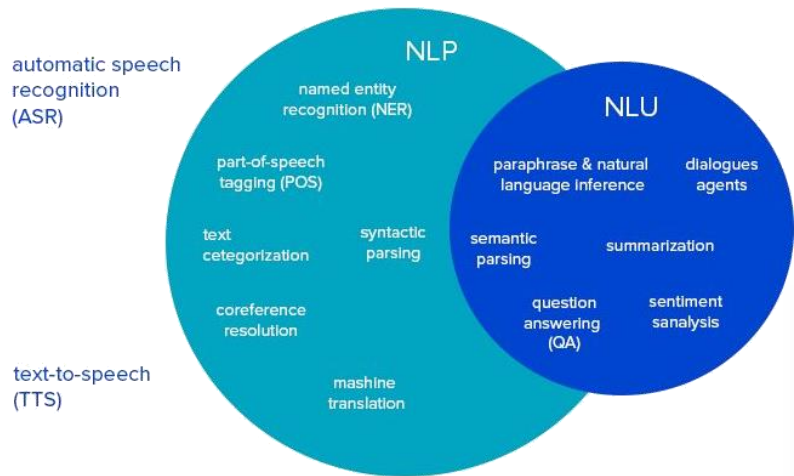
Link: [Doccano](#)

Person p Loc l Org o Event e Date d Other z

Barack Hussein Obama II (born August 4, 1961) is an American attorney and politician who served as the 44th President of the United States from January 20, 2009, to January 20, 2017. A member of the Democratic Party, he was the first African American to serve as president. He was previously a United States Senator from Illinois and a member of the Illinois State Senate.



# NLP = NLU + NLG



<https://www.blog.google/products/search/search-language-understanding-bert/>

Google

The Keyword

Latest Stories

Product Updates

Company News

SEARCH

## Understanding searches better than ever before

Pandu Nayak

Google Fellow and  
Vice President, Search

Published Oct 26, 2019

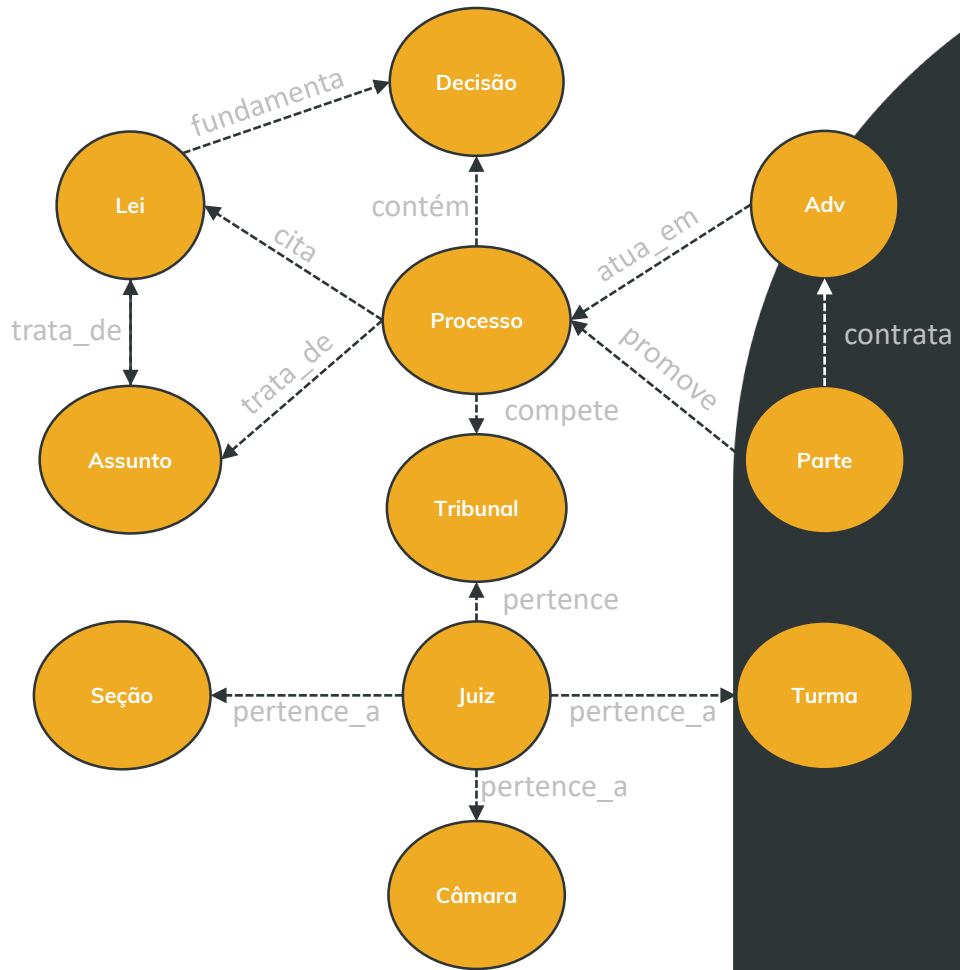
If there's one thing I've learned over the 15 years working on Google Search, it's that people's curiosity is endless. We see billions of searches every day, and 15 percent of those queries are ones we haven't seen before--so we've built ways to return results for queries we can't anticipate.





## Testando um chatbot com NLP

Aplicando NLU e NLG para  
interagir com a Justiça



## MODELO DO DADO

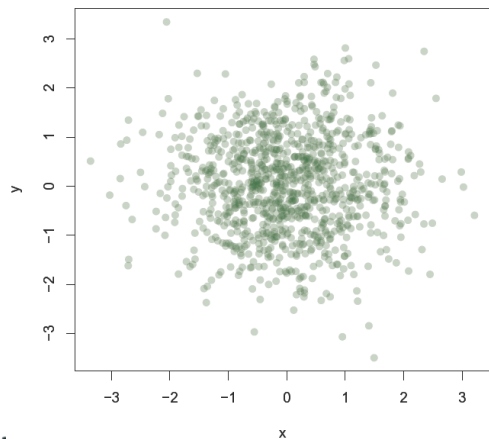
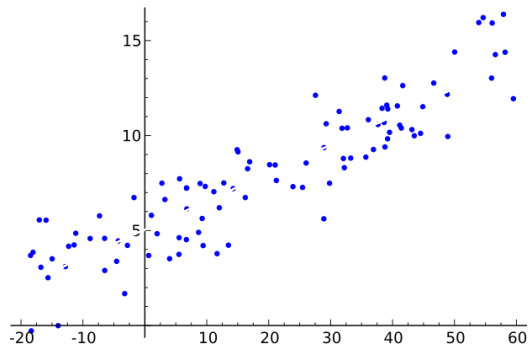
- ✓ Conectado
- ✓ Semântico
- ✓ Relações explícitas



# Possíveis entidades de um bd jurídico

1. Tribunal
2. Órgão julgante
3. Juiz(a) / Desembargador(a)
4. Número CNJ
5. Partes
  - a) Autor
  - b) Réu
6. Advogados
  - a) Do Autor
  - b) Do Réu
7. Datas
  - a) Início
  - b) Andamentos
  - c) Audiências
  - d) Encerramento
8. Classe
9. Assunto
10. Valores
  1. Pleiteados
  2. Concedidos
11. Segredo de justiça (S/N)
12. Resultado da sentença
  1. Procedência
13. Resultado do acórdão
  1. Conhecimento
  2. Provimento



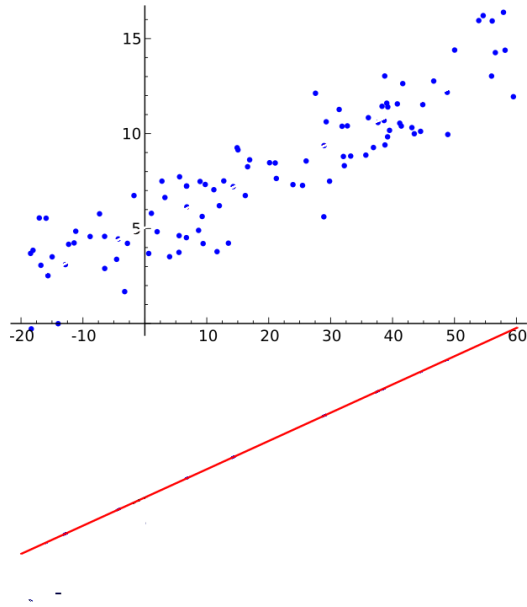


## Dispersão

Mostra o quão dispersa ou concisa é a distribuição dos dados.

**Exemplo:** valores de condenação de processos do mesmo assunto pelo mesmo juiz

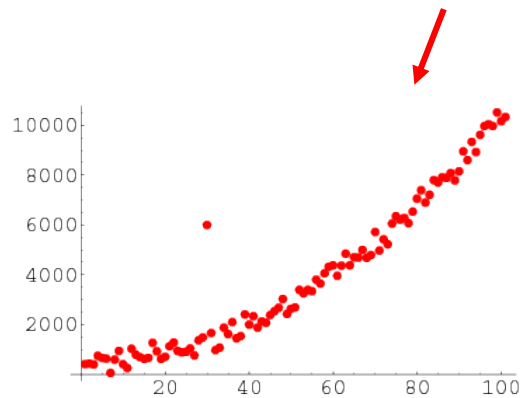




## Regressão linear

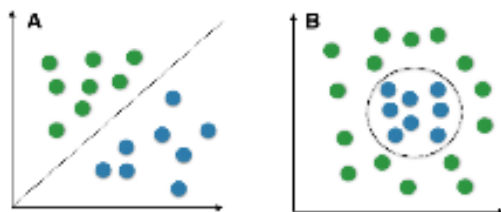
Trata de valores que não se consegue estimar inicialmente (valor esperado de uma variável 'y', dados os valores de outras variáveis 'x').

**Exemplo:** estimar valor de condenação (y), com base assunto (x).



## Outliers

Fenômeno que apresenta grande afastamento dos demais dados analisados, ou que é inconsistente. A existência de outliers implica, tipicamente, em prejuízos a interpretação dos resultados dos testes estatísticos aplicados às amostras.



## Naive Bayes

Modelo de classificação probabilístico com base na frequência de palavras de determinado conjunto de dados.

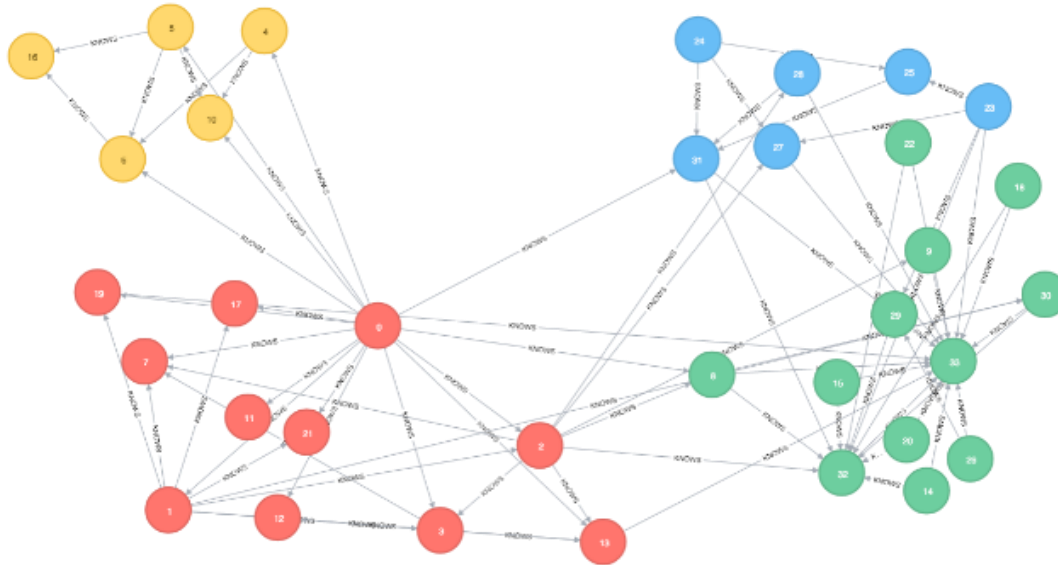
# k-means

Método de agrupamento de dados de acordo com suas propriedades – quanto mais semelhantes, mais próximos ficarão uns dos outros. Dados não-clusterizados.



# k-means

Método de agrupamento de dados de acordo com suas propriedades – quanto mais semelhantes, mais próximos ficarão uns dos outros. **Dados clusterizados.**



# Algumas ideias

- **Word2vec**: reconhecimento de entidades
- **Doc2vec**: semelhança entre trechos de decisões
- **Node2vec**: análise de semelhança entre julgados



# Mais alguns exemplos







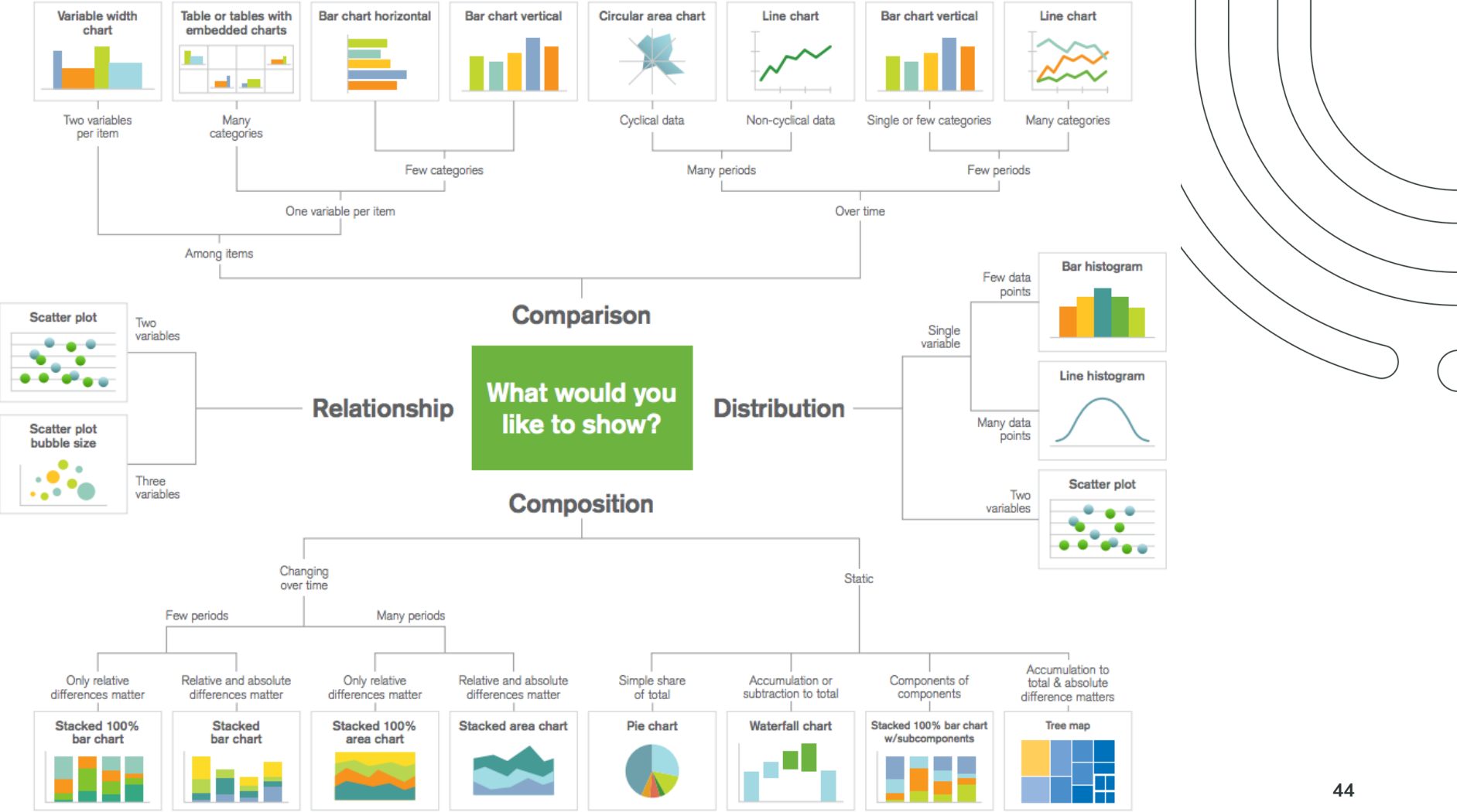
## FRONTEND

- ✓ Visualização dos dados coletados
- ✓ Apresentação de relatórios
- ✓ Acompanhamento de métricas

# FRONTEND

- Definir objetivo de cada visualização:
  - Composição
  - Distribuição
  - Comparação
  - Relação
- Escolher visualizações apropriadas de acordo com:
  - Quantidade de itens analisados
  - Categorias
  - Variáveis
- Apresentar
- Colher feedback
- Melhorar





**DÚVIDAS?**

Estou aqui para responde-las



## DEOCLIDES NETO

Co-Founder & CEO

e-mail [deocliides@juit.io](mailto:deocliides@juit.io)

<https://juit.io/>

