

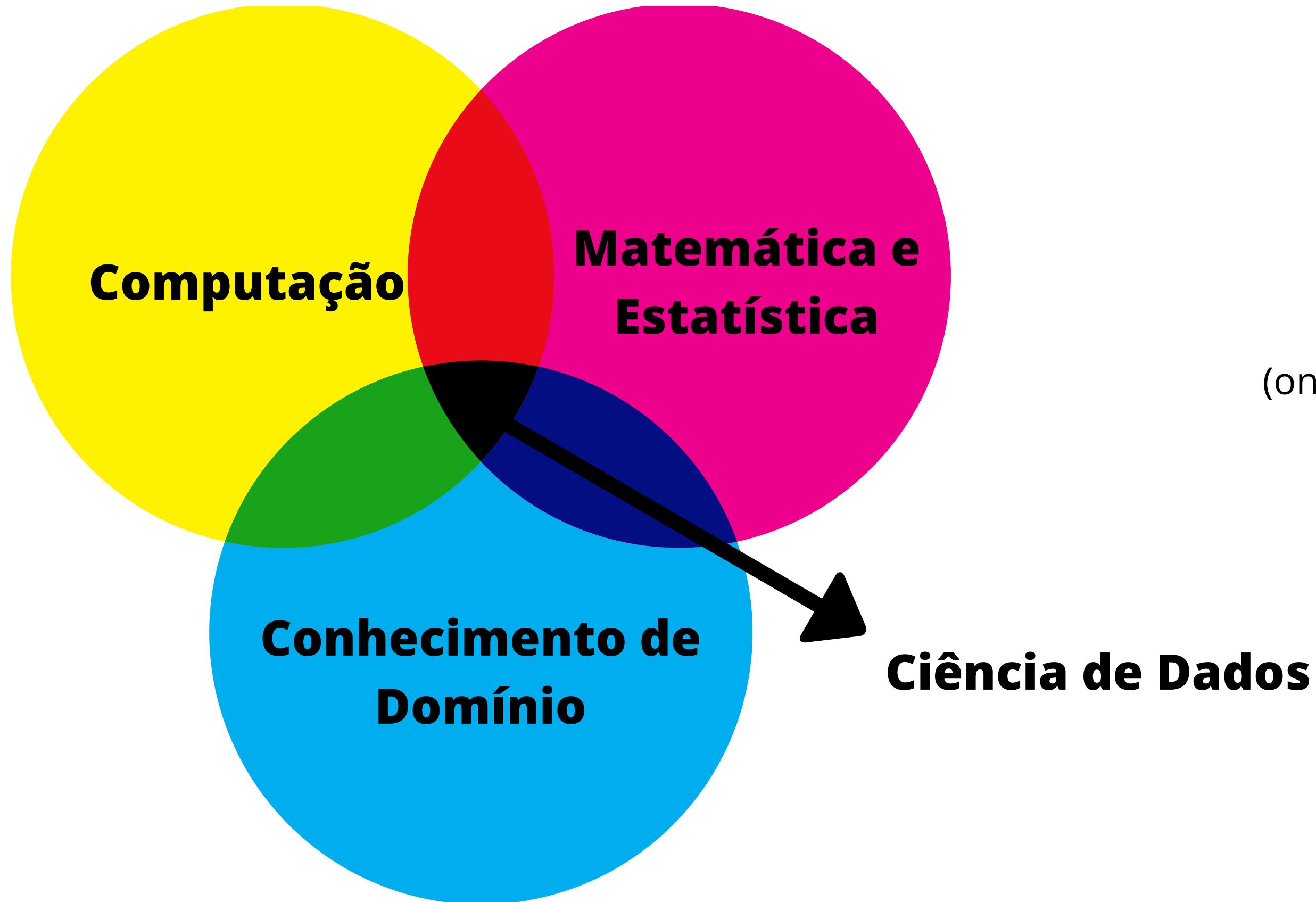
II Seminário de Pesquisa da FDRP  
Outubro/2020

## Jurimetria e Processamento de Linguagem Natural em pesquisas sobre acesso à justiça e desigualdades

Raysa Benatti (Instituto de Computação - Unicamp)  
Orientação: Esther Luna Colombini e Sandra Eliza Fontes de Avila  
Colaboração: Fabiana Cristina Severi e José de Jesus Filho

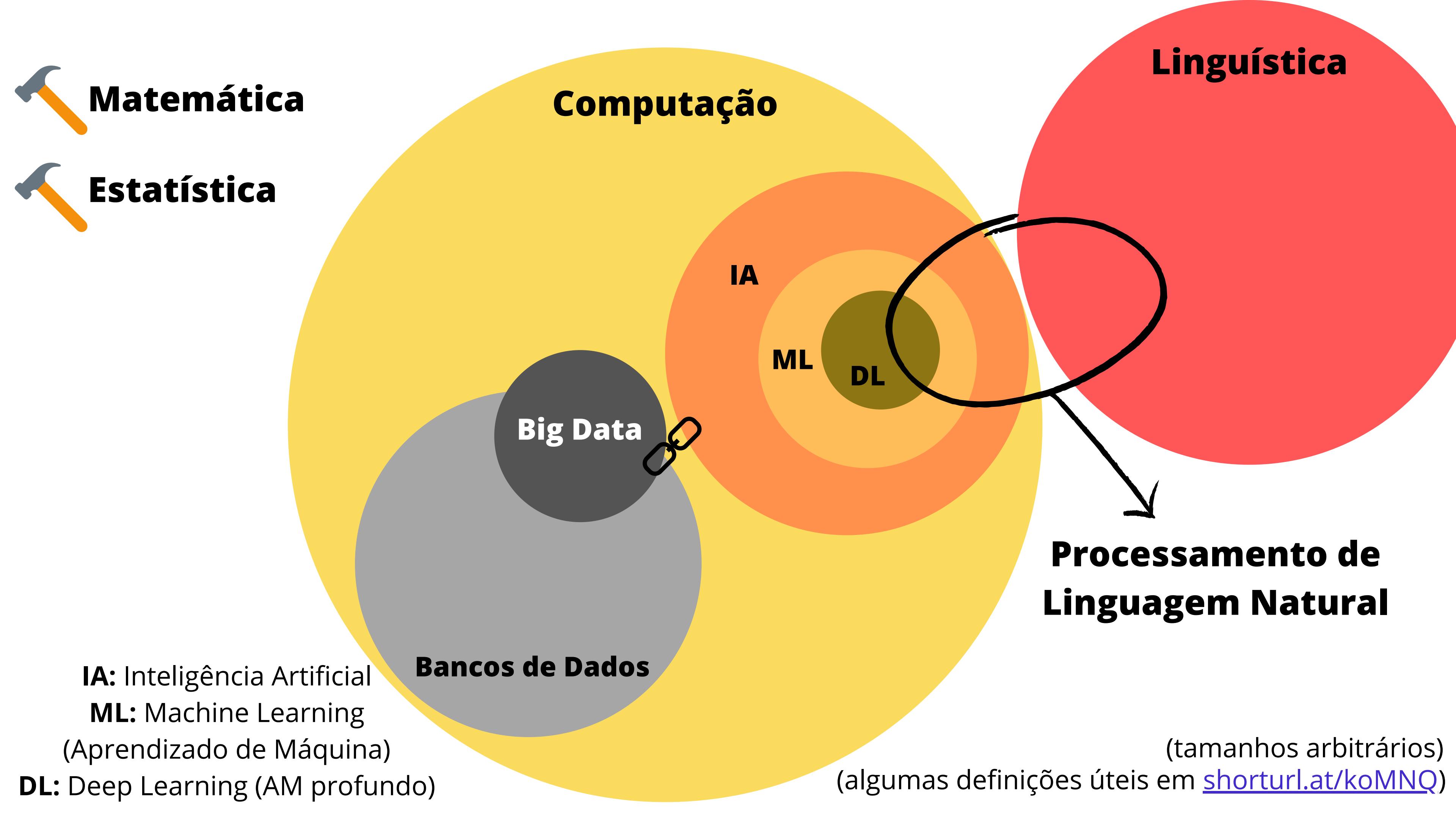
# **Agenda**

- Introdução - computação, inteligência artificial, ciência de dados e que tais
- Perspectiva histórica
- Processamento de linguagem natural
- Jurimetria
- Caso de uso e exercício

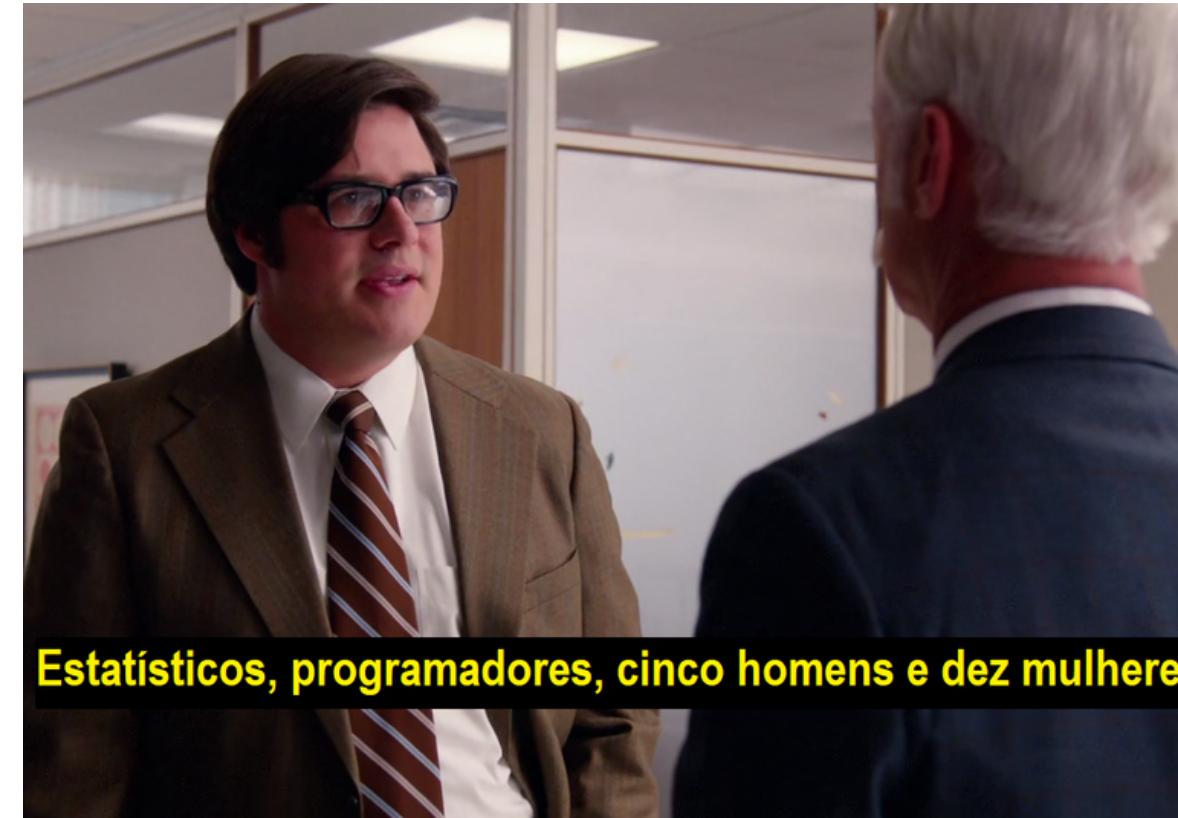
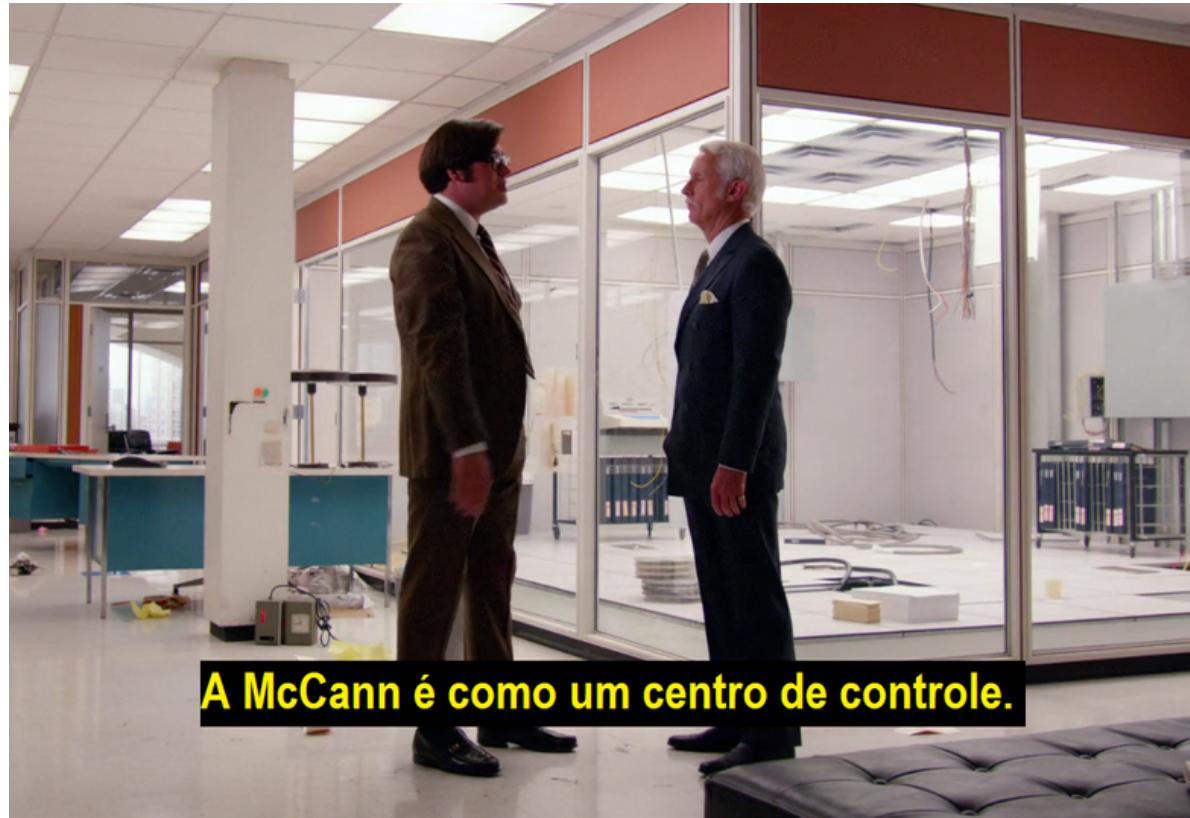


(onde estaria a Jurimetria?)

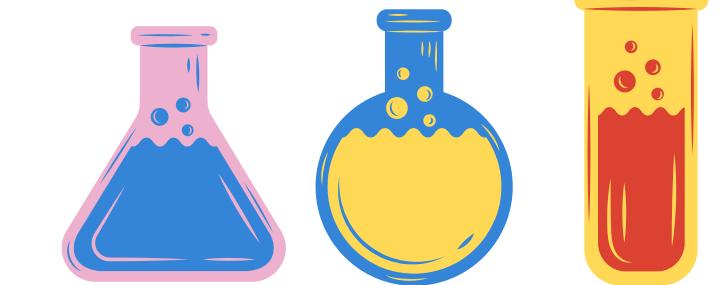
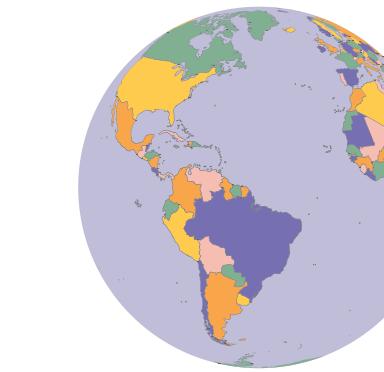
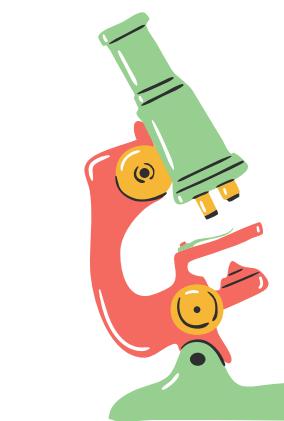
(adaptado de Conway, 2010)



# Análise e processamento de dados não foram criados ontem...



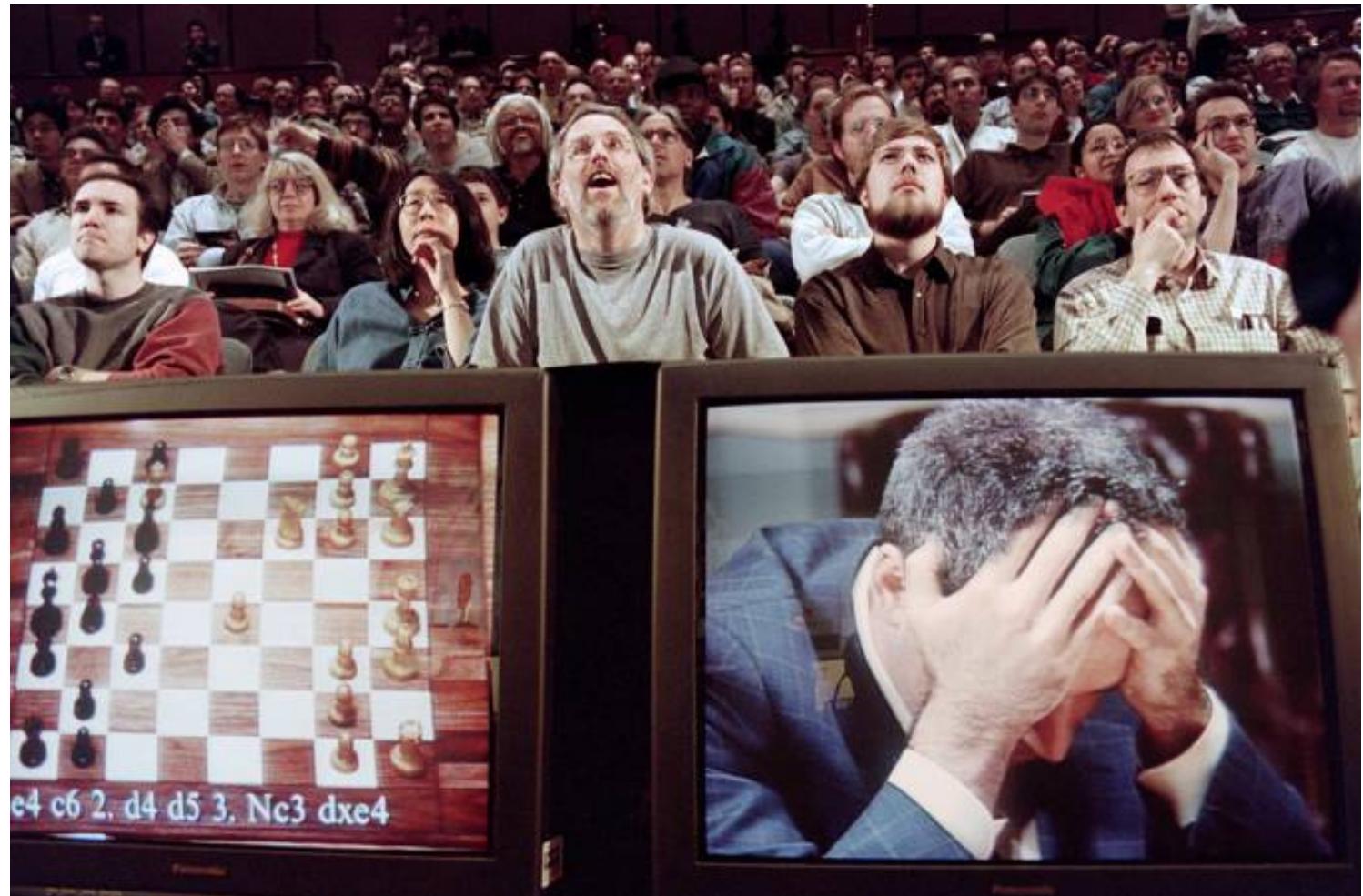
*Mad Men, 2015*



**Ciência experimental, políticas públicas, negócios**

# ...Inteligência Artificial também não

```
Welcome to  
EEEEEELL    IIIIZZZZZZ  AAAAAA  
EE      LL    II       ZZ  AA  AA  
EEEEEELL    II       ZZZ  AAAAAAAA  
EE      LL    II       ZZ  AA  AA  
EEEEEELL    IIIIZZZZZZ  AA  AA  
  
Eliza is a mock Rogerian psychotherapist.  
The original program was described by Joseph Weizenbaum in 1966.  
This implementation by Norbert Landsteiner 2005.  
  
ELIZA: Is something troubling you ?  
YOU: Men are all alike.  
ELIZA: What is the connection, do you suppose ?  
YOU: They're always bugging us about something or other.  
ELIZA: Can you think of a specific example ?  
YOU: Well, my boyfriend made me come here.  
ELIZA: Is it important to you that your boyfriend made you come here ?  
YOU: He says I'm depressed much of the time.  
ELIZA: I am sorry to hear that you are depressed.  
YOU: It's true. I am unhappy.  
ELIZA: Can you explain what made you unhappy ?  
YOU: █
```



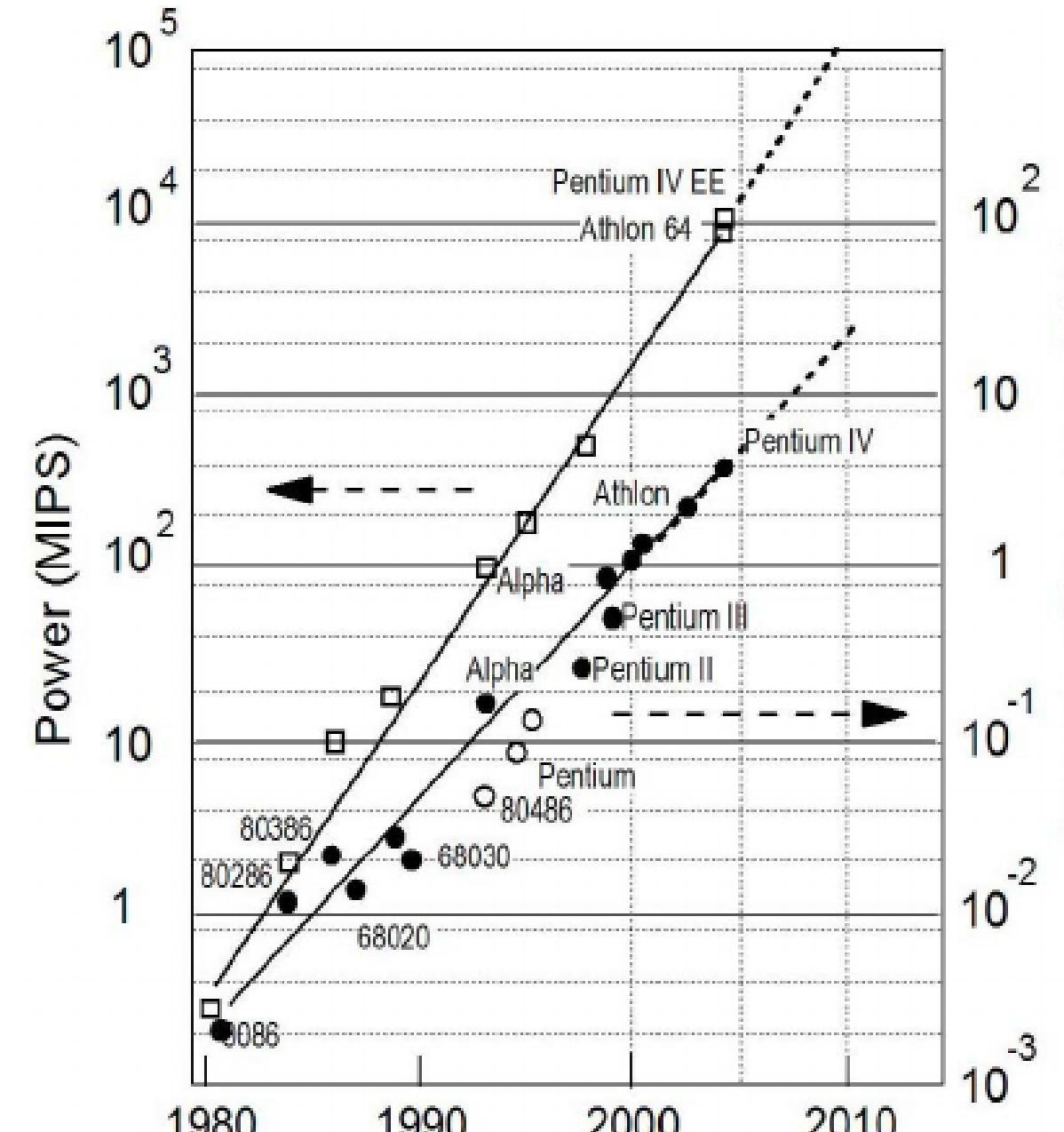
Weizenbaum (MIT), 1964-1966

Deep Blue (IBM) vs. Kasparov (11/05/1997, Stan Honda/AFP)

# Qual é, então, a novidade?

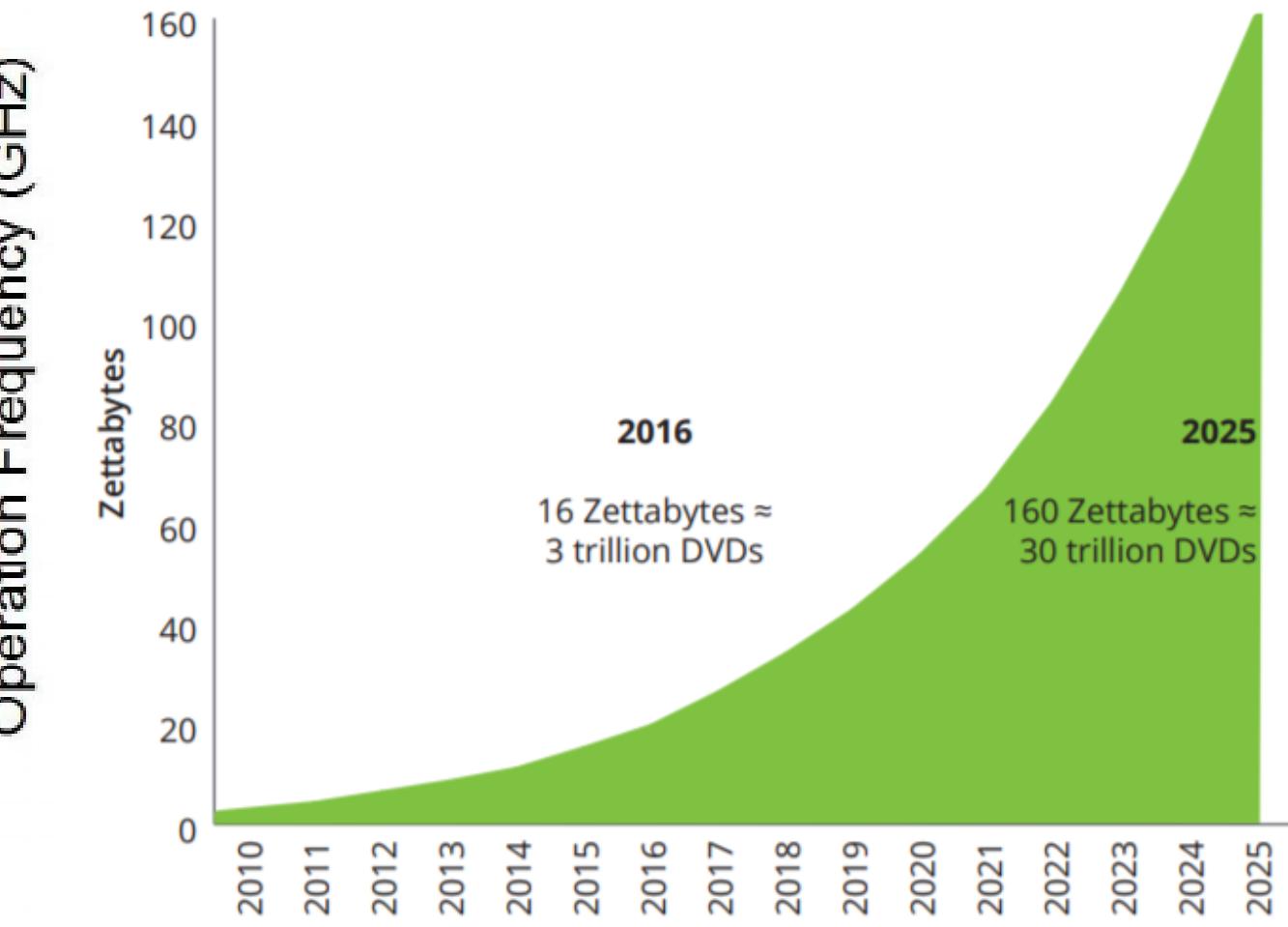


2315 *disk cartridge* (IBM, 1964, 1MB) vs. matriz de discos para armazenamento em nuvem (2009, ~TB). Computer History Museum ([computerhistory.org](http://computerhistory.org))



Evolução de poder computacional  
(Crouzet et al., 2005)

**Mais armazenamento**



Volume global estimado de dados  
(Deloitte / International Data Corporation, 2017)

**Mais processamento**

**Mais dados**

# Processamento de linguagem natural

- Interação entre computadores e linguagem humana
- Processamento e análise de grandes quantidades de dados em língua natural
- Geração e compreensão automática de línguas humanas naturais
- Processos computacionais para manipulação da linguagem humana (Linguística Computacional)

(Wikipedia)

- Que tipo de coisas as pessoas dizem?
- O que essas coisas dizem/demandam do mundo?



Na prática, bons sistemas de PLN devem ser capazes de tomar decisões de desambiguação em relação a sentido de palavras, categoria de palavras, estrutura sintática, escopo semântico

(Manning e Schütze, 1999)



## Exemplos

- Análise de sentimento em textos jornalísticos sobre economia, crime, meio ambiente e relações exteriores (Young e Soroka, 2011)
- Classificação ideológica de discursos proferidos no Senado estadunidense (Diermeier et al., 2011)
- Modelo de tópicos para detecção de prioridades expostas em discursos políticos (Grimmer, 2010)
- Classificação de resenhas de filmes baseada em análise de sentimento (Pang et al., 2002)
- Sumarização automática do conteúdo de documentos legais (Kanapala et al., 2018)
- Análise de conteúdo de peças para triagem processual (STJ, STF, tribunais estaduais)
- Análise de sentimento para identificação de resultado de decisão judicial (Aires, 2019; Morais, 2019 - UnB)

Em geral, técnicas comuns em ciências sociais incluem: sumarização automática; análise de discurso; identificação de tópicos; análise de subjetividade (sentimento, opinião, polaridade)

# E a Jurimetria?

- "Investigação científica de problemas legais" - proposição de questões sobre: comportamento de testemunhas, juízes e legisladores; procedimento e registros legais; "desajustes" e "aberrações de comportamento"; injúrias; técnicas "macrolegais" de investigação (Loevinger, 1949)
- "Estatística aplicada do Direito"; estudo do Direito "através das marcas que ele deixa na sociedade"; Judiciário como gerador de dados (Associação Brasileira de Jurimetria)
- Analogias: econometria, biometria, psicometria, antropometria

## Métodos quantitativos aplicados ao Direito

**Estatística  
Matemática  
Computação**

...

Essa análise começa com uma boa  
**pergunta de pesquisa...**

**Conhecimento de  
domínio**



...que pode vir **daqui**

(No tema **acesso à justiça e desigualdade de gênero**, por exemplo, em quais perguntas você pensaria?)

# Caso de uso: acesso à justiça e desigualdade de gênero

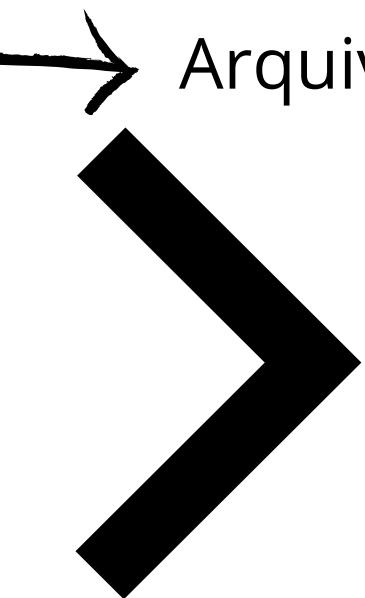
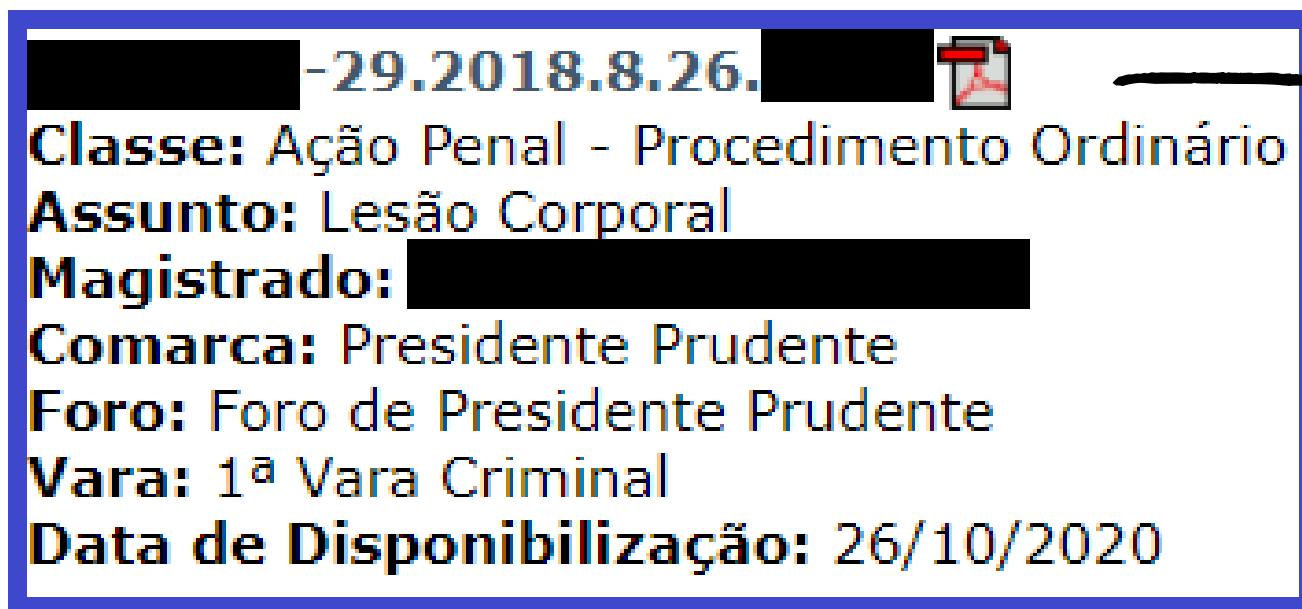
O que o **domínio** pode nos informar a respeito?

- Conceito de gênero;
- Tratamento legal sobre gênero;
- Análises qualitativas de situações de violação de direitos;
- Argumentos comuns em decisões;
- Experiência da vítima de violência de gênero submetida aos procedimentos institucionais;
- Qualidade das sentenças;
- Relatos;
- ...

Eventualmente, podemos esbarrar em problemas que **métodos quantitativos** podem ajudar a solucionar...

- Para uma determinada classe de violações, qual é o tempo médio de duração de um processo?;
- Existem padrões no conteúdo dos textos analisados quanto a q ( $Q = \{\text{resultado processual para pedidos semelhantes, argumentos, vieses, sentimentos, opiniões, ...\}}\}$ )?
- Em quais z ( $Z = \{\text{comarcas, varas, anos, ...}\}$ ) o achado w ( $W = \{\text{processos sobre BLA, penas mais severas, penas menos severas, ...}\}$ ) é mais comum?
- Há correlação entre x e y? ( $X, Y = \{\text{tempo de duração do processo, resultado processual, qualidade da sentença, gênero do julgador, Z, W, Q, ...}\}$ )
- ...

# Caso de uso: acesso à justiça e desigualdades



Arquivo com conteúdo textual da decisão

Metadados (dados sobre os dados)

Um dos resultados exibidos em busca por palavra-chave "violência doméstica" no site do TJSP

Sabendo que essas informações estão disponíveis e considerando possíveis perguntas de pesquisa (incluindo as sugeridas no slide anterior):

- Quais procedimentos seriam necessários, da coleta de dados à interpretação dos resultados, para buscar soluções?
- Para cada um desses procedimentos, qual(is) área(s) é(são) mais diretamente implicada(s)? (Pensar nisso pode ajudar a compreender quais recursos - humanos ou não - são necessários em cada etapa)
- O que se deve considerar antes de coletar os dados? ("Good data does not guarantee good decisions, but bad data almost always results in bad decisions" - Mike Tanier)
- O que podemos fazer com a solução em mãos?

**Obrigada!**

[raysa.benatti@gmail.com](mailto:raysa.benatti@gmail.com)  
[github.com/ra-ysa](https://github.com/ra-ysa)