

PÓS **PUCPR** DIGITAL



Extração de Dados: Como
Obter o Melhor dos Dados

22/02/24

Agenda

- Combinados / Roteiro
 - Tem muita coisa!!
- Dúvidas?
- Avisos Paroquiais
- Datajud
- Chat GPT
- Outros pacotes
- Final

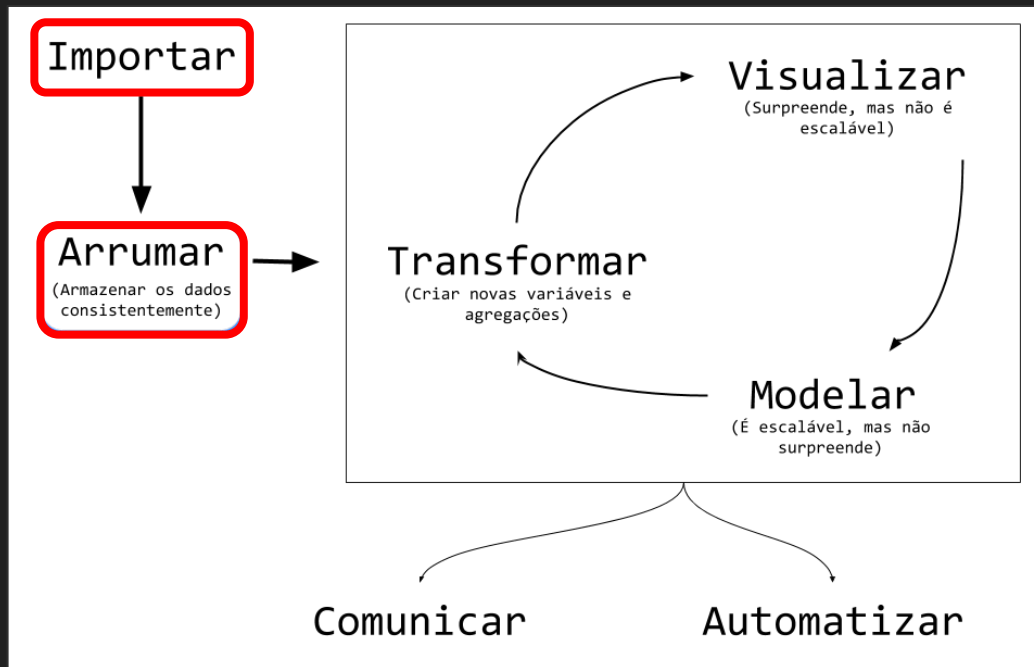
Renovando os Combinados

- Nenhuma dúvida é “idiota”
 - Mas pode ser que eu não consiga responder todas nesse exato momento
 - Há espaço para todo mundo na ciência de dados!
- Erros acontecem
 - Na vida real de cientistas de dados experientes e - por que não? - aqui também
- Mão na massa
 - Vou tentar falar o mínimo possível (apesar de ser advogado)
 - Tentativa: focarmos na prática

(Re)Vimos na Semana Passada

1. Instalação do R / Uso do RStudio
 - a. Plano B: <https://rstudio.cloud/>
2. Instalação de Pacotes
3. Objetos
 - a. Assumem “naturezas” variadas
 - i. Strings
 - ii. Números
 - iii. Funções
 - b. Podem ser gravados/sobrescritos/apagados
4. Vetores e dataframes são as principais formas dos objetos que vamos usar

(Re)Vimos na Terça



Receita de Bolo

- Passo 1: identificar a análise que queremos fazer
 - Que termos temos que buscar?
- Passo 2: análise de viabilidade
 - Testar a pesquisa: tem poucos? Tem muitos?
- Passo 3: raspagem dos dados
 - Equivale a entrar em um por um e “tirar um print”
 - Pode demandar autenticação / pode haver entraves técnicos
- Passo 4: leitura dos dados baixados
- Passo 5: limpeza e organização
- Passo 6: análises
- Passo 7: salvar resultados

Exemplo: HC Coletivo das Mães

Termo de busca	livre = ""
Classe processual	classe = ""
Assunto(s)	assunto = ""
Órgão Julgador	orgao_julgador = ""
Datas de início/fim do julgamento	inicio = "", fim = ""
Datas de início/fim da publicação ou registro	inicio_pb = "", fim_pb = ""
Outras opções	tipo = "A", n = NULL, diretorio = "."

É o que faremos no R a partir de agora: automatizar a “pesquisa” e extração dos dados, de maneira que possamos manuseá-los e extrair insights

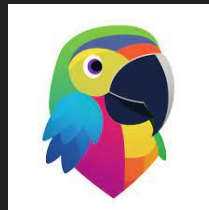
Update do pacote {tjsp}

- Lembram que falamos que tudo é muito dinâmico?
- É verdade :)
- Pacote {tjsp} estava com um problema na função de andamentos
 - Reportei aqui: <https://github.com/jjesusfilho/tjsp/issues/37>
- Prof. José prontamente atualizou
- Orientação: reinstalar
 - `remotes::update_packages()`
 - `remotes::install_github("jjesusfilho/tjsp")`

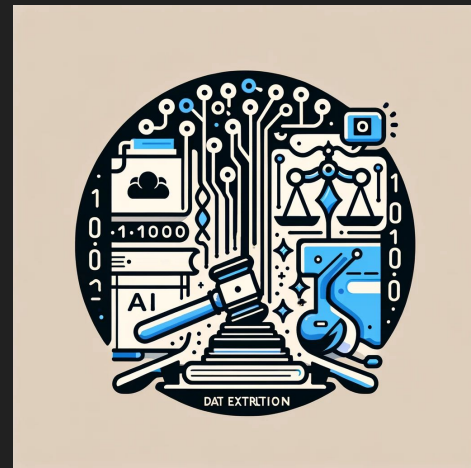
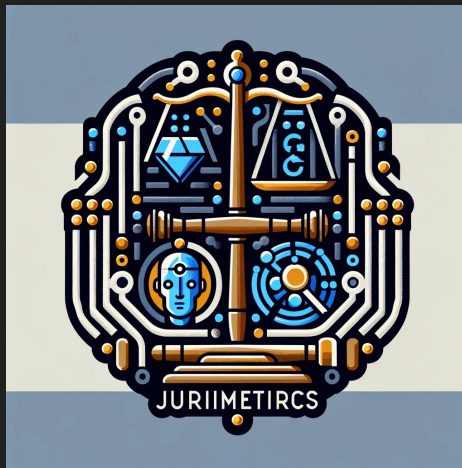
Datajud -

- Iniciativa do CNJ há alguns anos
- No final de 2023 foi aberta ao público
 - Com ressalvas importantes!
 - Não apresenta dados de partes
 - Não contém o teor de decisões
 - Etc
- Ainda assim, ferramenta muito poderosa e promissora
- Preparei um pacote (simples!) especialmente pra vocês
 - `remotes::install_github("rfdornelles/datajud")`
- Extrair dados de praticamente qualquer tribunal
- Atenção: LGPD e Termos de uso!!
- Link: <https://www.cnj.jus.br/sistemas/datajud/>

ChatGPT, Gemini, Maritalk e afins



- Existem muitas ferramentas de LLM disponíveis
- Usem (com bom senso): será essencial para nós
- Pode ajudar na geração de códigos: mas façam com responsabilidade, testando e validando
- Vamos pensar em alguns exemplos?



Vamos para o R!



https://bit.ly/pucpur_jurimetria2024



Muito obrigado!

Rodrigo Dornelles

rodornelles@gmail.com

<https://www.linkedin.com/in/rfdornelles/>

