

PÓS **PUCPR** DIGITAL



Extração de Dados: Como
Obter o Melhor dos Dados

15/02/24

Agenda

- Combinados / Roteiro
- Proposta de divisão
- “Check-in”: Como chegamos?
- Recapitular/Dúvidas (até 10 min):
 - Por que programação? Por que R?
 - Intuições: pacotes, objetos, funções, fluxo da ciência de dados
- Extração de Dados - por que mesmo?
- “Receita de Bolo” de extração de dados
 - Como preparar os ingredientes?
 - Exemplo 1 - HC Coletivo
 - Exemplo 2 - Vocês decidem
- Dúvidas finais / Check-out
- Continua no próximo episódio - Datajud

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
- O tema pode ser bastante árido, mas estamos comprometid@s com o aprendizado de vocês
- Eu realmente acredito que vai valer a pena :)

Check-in

- Quem se sentir à vontade
 - No chat/por escrito
 - “Qual minha área?”
 - “Como estou chegando hoje?”
 - “Qual minha expectativa?”
-
- *(eu sei que parece titlelê, mas juro que ajuda)*

Recapitulação - R em uma Casca de Noz

1. Instalação do R / Uso do RStudio

- a. Não consegue de nenhuma forma? Sem problemas: <https://rstudio.cloud/>

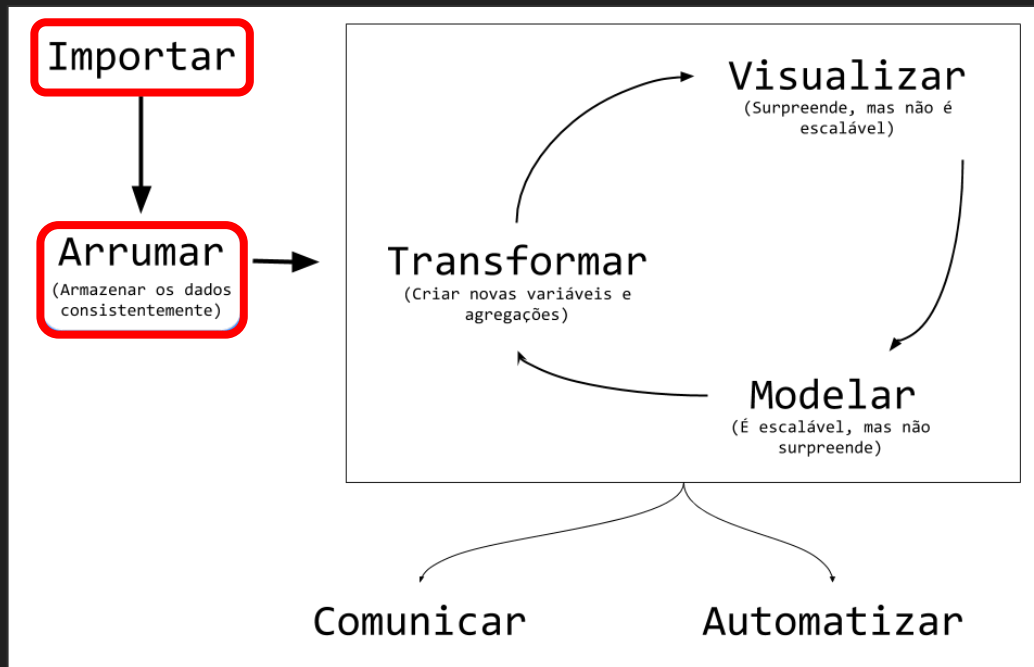
2. Instalação de Pacotes

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

Recapitulação II - Qual nosso fluxo de trabalho?



Mais dúvidas?

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




Mais dúvidas?



Lembrando: qual a nossa mecânica?





```
graph LR; A[Acessar a busca] --> B[Simular requisição]; B --> C[Copiar parâmetros]; C --> D[Repetir]; D --> E[Analisar!]
```

Acessar a busca

Simular
requisição

Copiar
parâmetros

Repetir

Analisar!



Orientações

- **Atenção:** A Consulta de Jurisprudência agora permite a seleção do tipo de decisão que se deseja consultar. Estão disponíveis os seguintes tipos de decisão: Acórdão, Decisão Monocrática e Homologação de Acordo. Para realizar sua pesquisa, lembre-se de selecionar o tipo de decisão que deve ser considerado na busca.
- Para detalhes sobre cada campo de busca, selecione o campo específico.

Consulta Completa

Pesquisa livre

:

E

OU

NÃO

" "

Como utilizar os filtros ☒ Pesquisar por sinônimos

Pesquisa por campos específicos

Ementa

:

Número do recurso

:

Número do registro

:

Relator(a)

:



Magistrado prolator

:



Classe

:



Assunto

:



Comarca

:



Órgão julgador

:



Data do julgamento

:

até

(dd/mm/aaaa)

Data de publicação

:

até

(dd/mm/aaaa)

Origem

:



2º grau



Colégios Recursais

Tipo de Publicação

:



Acórdãos



Homologações de Acordo



Decisões Monocráticas

Orientações

- **Atenção:** A Consulta de Jurisprudência agora permite a seleção do tipo de decisão que se deseja consultar. Estão disponíveis os seguintes tipos de decisão: Acórdão, Decisão Monocrática e Homologação de Acordo. Para realizar sua pesquisa, lembre-se de selecionar o tipo de decisão que deve ser considerado na busca.
- Para detalhes sobre cada campo de busca, selecione o campo específico.

Consulta Completa

Pesquisa livre

:

E

OU

NÃO

" "

Como utilizar os filtros ☒ Pesquisar por sinônimos

Pesquisa por campos específicos

Ementa

:

Número do recurso

:

Número do registro

:

Relator(a)

:



Magistrado prolator

:



Classe

:



Assunto

:



Comarca

:



Órgão julgador

:



Data do julgamento

:

até

(dd/mm/aaaa)

Data de publicação

:

até

(dd/mm/aaaa)

Origem

:



2º grau



Colégios Recursais

Tipo de Publicação

:



Acórdãos



Homologações de Acordo



Decisões Monocráticas

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

Mais dúvidas?

Vamos para o R!



https://bit.ly/pucpur_jurimetria2024



Exemplo II: ????????

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 = "."

Mais dúvidas?

Muito obrigado!

**TO BE
CONTINUED...** →

**Continuem mandando suas
dúvidas pelo sistema e até a
próxima live!**

Rodrigo Dornelles

rodornelles@gmail.com

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