

CURSO R - 2023



NAIARA ALCANTARA



AVISOS

Solicitamos que todos os que possuírem e puderem levar seus próprios notebooks, levem.

Os primeiros slides indicam como baixar o R Cran e R Studio. Podem instalar em seus computadores pessoais, de todo modo, quem não conseguir será auxiliado a fazer os downloads e instalações no primeiro dia do curso

Todos os materiais do curso:

1-Script

2-Bases de dados

3-PDF do curso

Devem ser baixados do Github e armazenados em uma pasta de fácil acesso

Nome da pasta: CURSO R UFPA

Link: <https://github.com/NaiaraSandi1995?tab=repositories>

Receberão certificação quem participar os 2 dias de curso e preencher as listas de presença disponibilizadas no curso.

TÓPICOS

1-INTRODUÇÃO AO R

2- ANÁLISE EXPLORATÓRIA E MANIPULAÇÃO
DOS DADOS

3- SALVAMENTO E ABERTURA

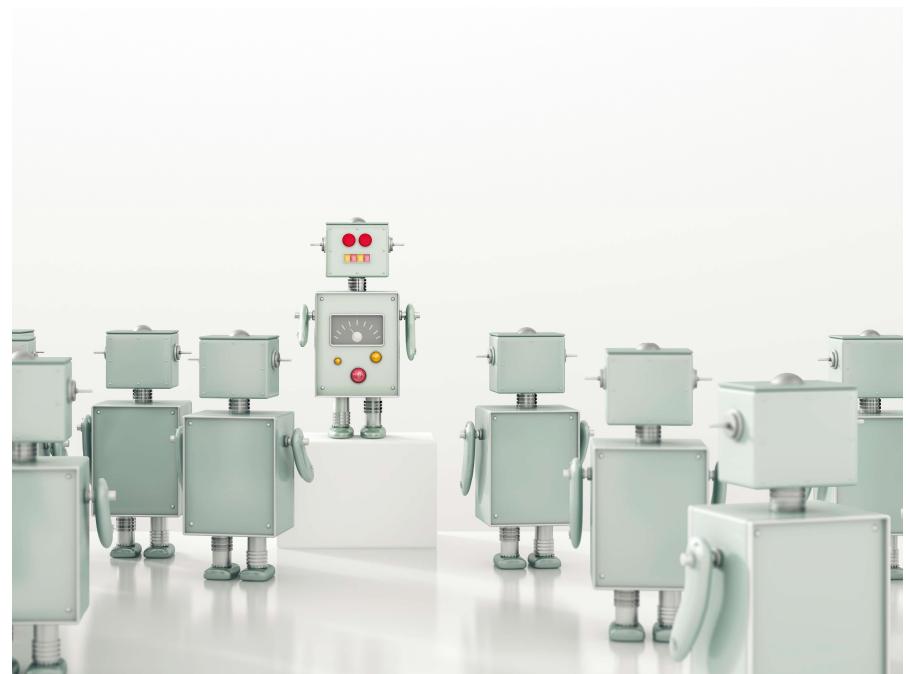
4- ANÁLISES DESCRIPTIVAS

5- APRESENTAÇÃO GRÁFICA

6- PROCESSAMENTO DE DADOS

7- ANÁLISES INFERENCIAIS

- Download do cran e rstudio
- Apresentação do R
- Funções básicas
- Criação de data frame





INTRODUÇÃO

O que é R?

Introdução a R

R é uma linguagem e ambiente para computação estatística e gráficos. É um projeto GNU que é semelhante à linguagem e ambiente S que foi desenvolvido na Bell Laboratories (anteriormente AT&T, agora Lucent Technologies) por John Chambers e colegas. R pode ser considerado como uma implementação diferente de S. Existem algumas diferenças importantes, mas muitos códigos escritos para S são executados inalterados em R.

R fornece uma ampla variedade de técnicas estatísticas (modelagem linear e não linear, testes estatísticos clássicos, análise de série temporal, classificação, agrupamento ...) e técnicas gráficas, e é altamente extensível. A linguagem S é frequentemente o veículo de escolha para pesquisa em metodologia estatística, e R fornece uma rota de código aberto para a participação nessa atividade.

Um dos pontos fortes de R é a facilidade com que gráficos de qualidade de publicação bem projetados podem ser produzidos, incluindo símbolos matemáticos e fórmulas quando necessário. Foi tomado muito cuidado com os padrões para as escolhas de design menores em gráficos, mas o usuário mantém o controle total.

R está disponível como Software Livre sob os termos da Free Software Foundation's GNU General Public License em forma de código fonte. Ele compila e roda em uma ampla variedade de plataformas UNIX e sistemas semelhantes (incluindo FreeBSD e Linux), Windows e Mac OS.

ENTENDA O QUE É O R - NO PRÓPRIO SITE DO R

O ambiente R

R é um conjunto integrado de recursos de software para manipulação de dados, cálculos e exibição gráfica. Inclui

- um manuseio eficaz de dados e instalação de armazenamento,
- um conjunto de operadores para cálculos em matrizes, em matrizes particulares,
- uma coleção grande, coerente e integrada de ferramentas intermediárias para análise de dados,
- recursos gráficos para análise e exibição de dados na tela ou em papel, e
- uma linguagem de programação bem desenvolvida, simples e eficaz que inclui condicionais, loops, funções recursivas definidas pelo usuário e recursos de entrada e saída.

O termo "ambiente" pretende caracterizá-lo como um sistema totalmente planejado e coerente, ao invés de um acréscimo incremental de ferramentas muito específicas e inflexíveis, como é frequentemente o caso com outros softwares de análise de dados.

R, como S, é projetado em torno de uma verdadeira linguagem de computador e permite aos usuários adicionar funcionalidades adicionais, definindo novas funções. Muito do sistema é escrito no dialeto R de S, o que torna mais fácil para os usuários seguirem as escolhas algorítmicas feitas. Para tarefas de computação intensiva, códigos C, C ++ e Fortran podem ser vinculados e chamados em tempo de execução. Os usuários avançados podem escrever código C para manipular objetos R diretamente.

Muitos usuários pensam em R como um sistema de estatísticas. Preferimos pensar nisso como um ambiente no qual as técnicas estatísticas são implementadas. R pode ser estendido (facilmente) por meio de pacotes. Existem cerca de oito pacotes fornecidos com a distribuição R e muitos mais estão disponíveis através da família CRAN de sites da Internet cobrindo uma ampla gama de estatísticas modernas.

RStudio

Assuma o controle do seu código R

RStudio é um ambiente de desenvolvimento integrado (IDE) para R. Inclui um console, editor de realce de sintaxe que oferece suporte à execução direta de código, bem como ferramentas para plotagem, histórico, depuração e gerenciamento de espaço de trabalho.

O RStudio está disponível em edições **comerciais** e de **código aberto** e é executado na área de trabalho (Windows, Mac e Linux) ou em um navegador conectado ao RStudio Server ou RStudio Workbench (Debian / Ubuntu / Red Hat / CentOS e SUSE Linux)

INSTALAÇÃO DO R

Vídeo demonstrando a instalação*

É possível que uma parte do vídeo esteja desatualizada, nesse caso, siga o passo a passo abaixo:

<https://www.youtube.com/watch?v=SWEizTHK57A>

(este vídeo foi produzido pelo aluno do mestrado em Ciência Política da UFPR, Renan Arnon de Souza, e disponibilizado para ser utilizado nessa oficina)

1º PASSO:

BAIXAR E INSTALAR O CRAN:

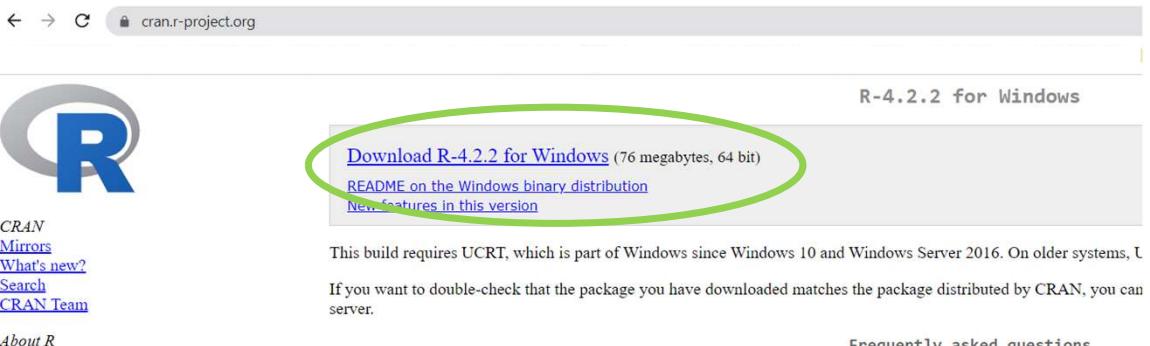
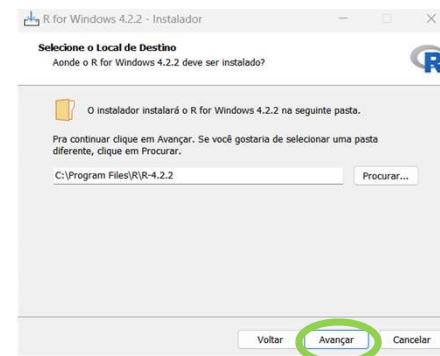
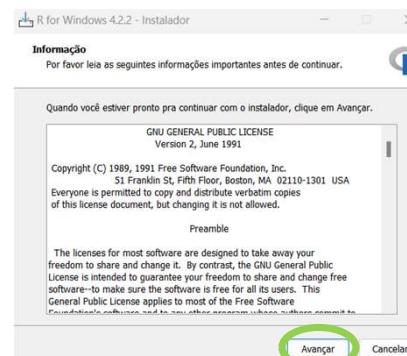
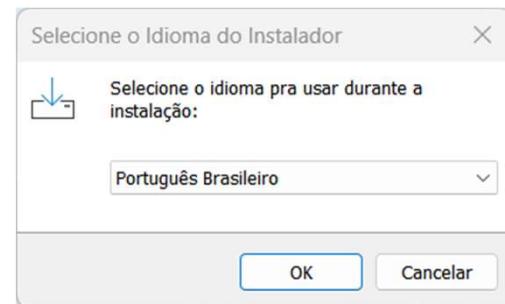
[HTTPS://CRAN.R-PROJECT.ORG/](https://cran.r-project.org/)

Obs.: Geralmente o próprio R já entende qual é o seu sistema operacional e indica o cran corretamente

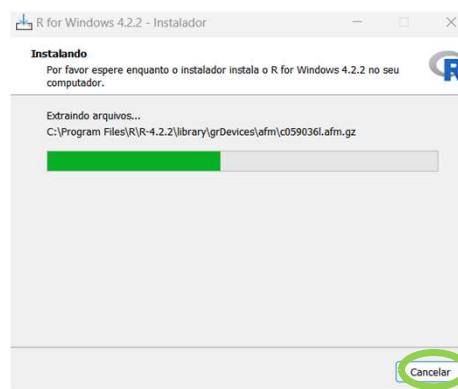
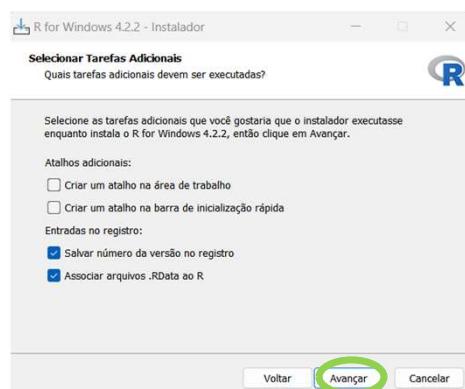
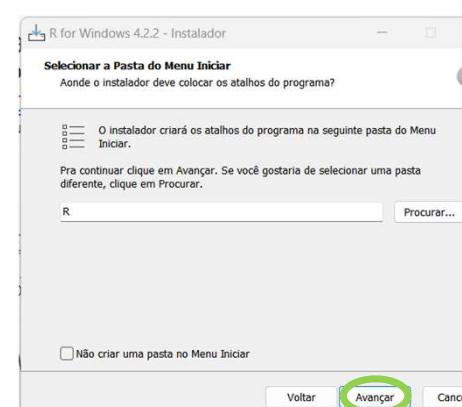
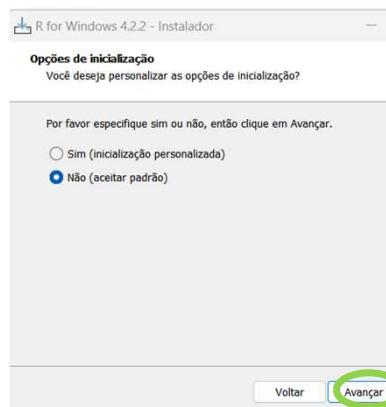
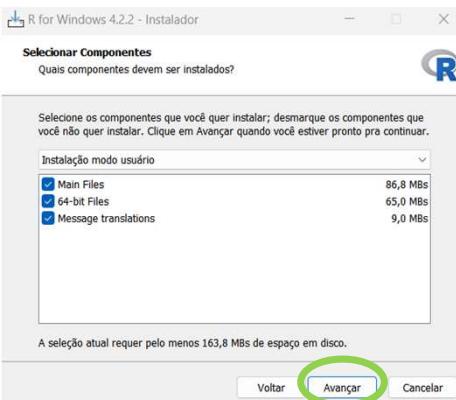
2º PASSO:

APÓS FINALIZAR O DOWNLOAD, CLIQUE PARA INSTALAR O PROGRAMA

Selecione o idioma português, o programa será em inglês, porém a instalação será em português.



A máquina tem que ter 64-bit



Após finalizar o carregamento conclua o processo

3º passo:

Baixar e instalar o R studio:

[HTTPS://POSIT.CO/DOWNLOAD/RSTUDIO-DESKTOP/](https://posit.co/download/rstudio-desktop/)

No site estará indicando que você deverá seguir um conjunto de passos, sendo o passo 1 baixar o cran, como você já fez isso, vá direto ao passo 2 e clique em download Rstudio.

Step 2: Install RStudio Desktop

[DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS](#)

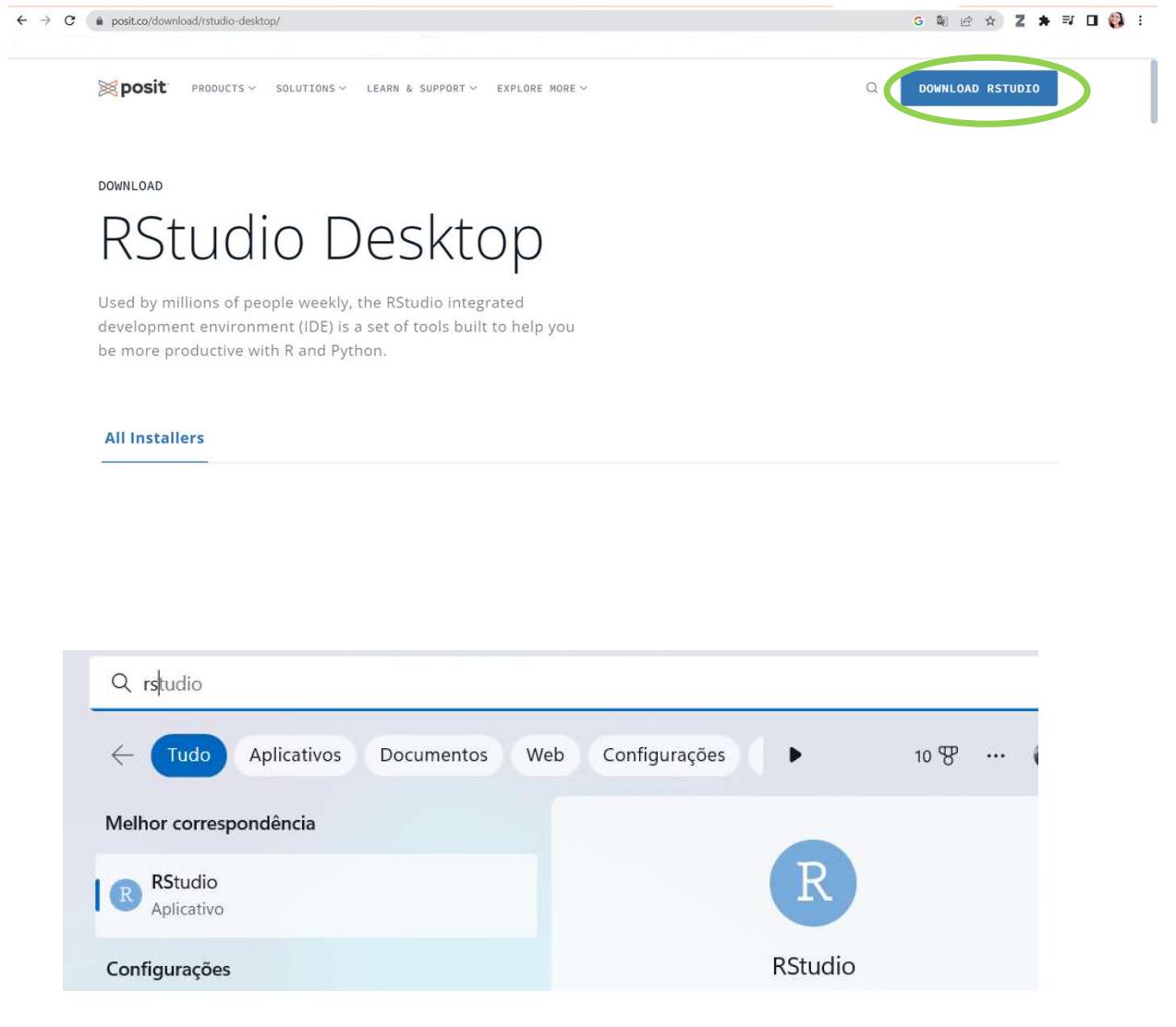
Size: 202.76MB | [SHA-256: FD8EA4B4](#) | Version: 2022.12.0+353 |

Released: 2022-12-15

4º passo:

Após finalizar o download, inicie a instalação sem alterar nenhum parâmetro, e clique em concluir

Agora o Rstudio já deve estar instalado, basta clicar em buscar e digitar Rstudio. Você pode pedir para fixa-lo em sua barra de tarefas.



ENTENDENDO A INTERFACE DO R



The screenshot shows the RStudio interface with several panes highlighted by yellow boxes:

- Script**: The top-left pane where R code is written.
- Environment/History/Connections**: The top-right pane containing tabs for Environment, History, and Connections.
- Console**: The bottom-left pane displaying the R startup message and help text.
- Files/Plots/Packages/Help/Viewer**: The bottom-right pane containing tabs for Files, Plots, Packages, Help, and Viewer.

The R startup message in the Console pane includes:

```
R version 4.0.3 (2020-10-10) -- "Bunny-wunnies Freak Out"
Copyright (C) 2020 The R Foundation for statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuir e modificá-lo sob certas condições.
Digitte 'license()' para ver as condições de distribuição.

R é um projeto da The R Foundation.
Digitte 'contributors()' ou 'citation()' para obter informações e citar autores.

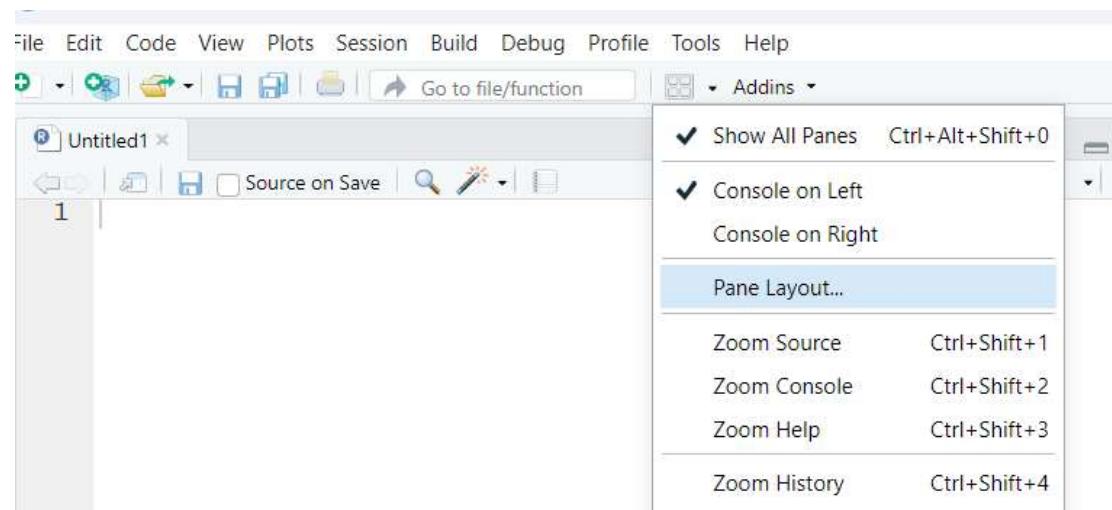
Digitte 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digitte 'q()' para sair do R.

[workspace loaded from ~/.RData]

> |
```

BOAS PRÁTICAS E MUDANÇA DE VISUAL

MUDANÇA DE VISUAL DO R:
1º WORKSPACE PANE > 2º PANE LAYOUT ... >

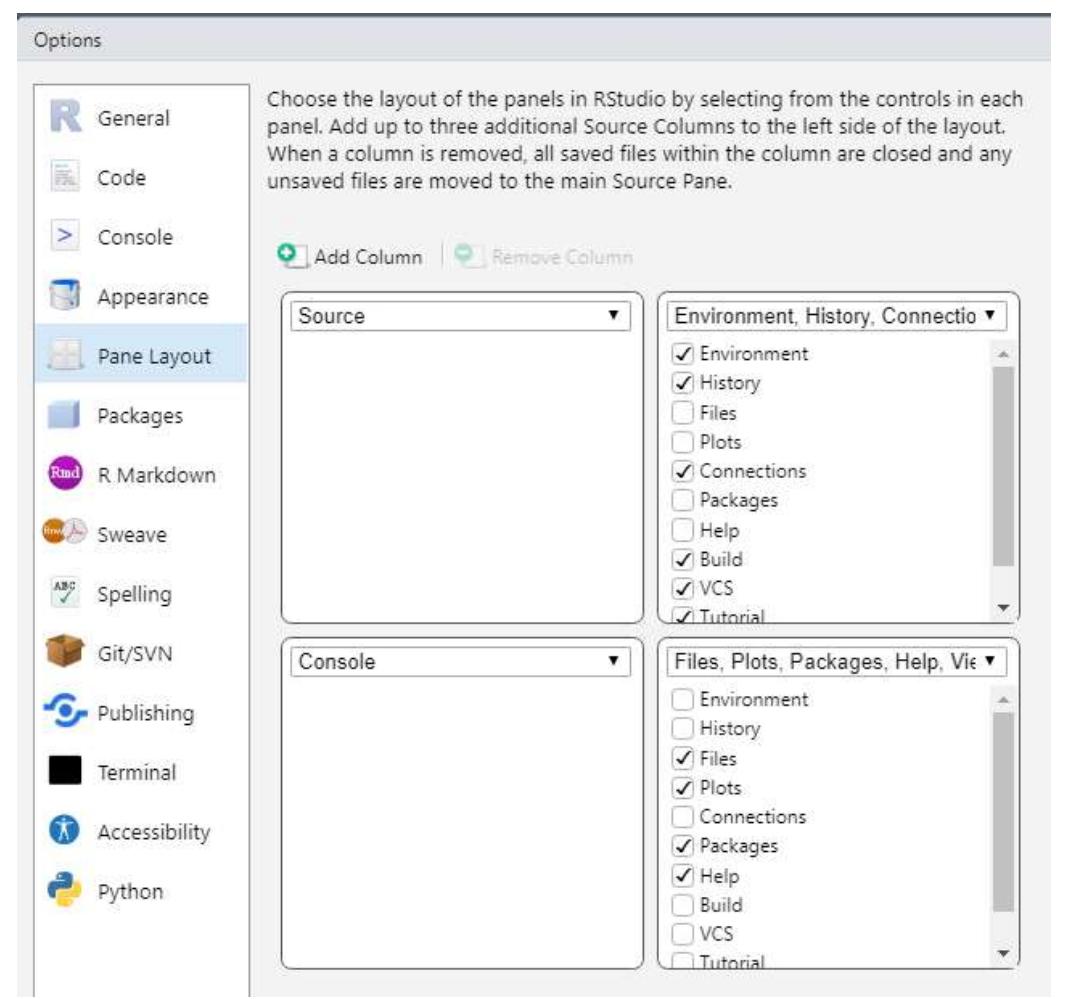
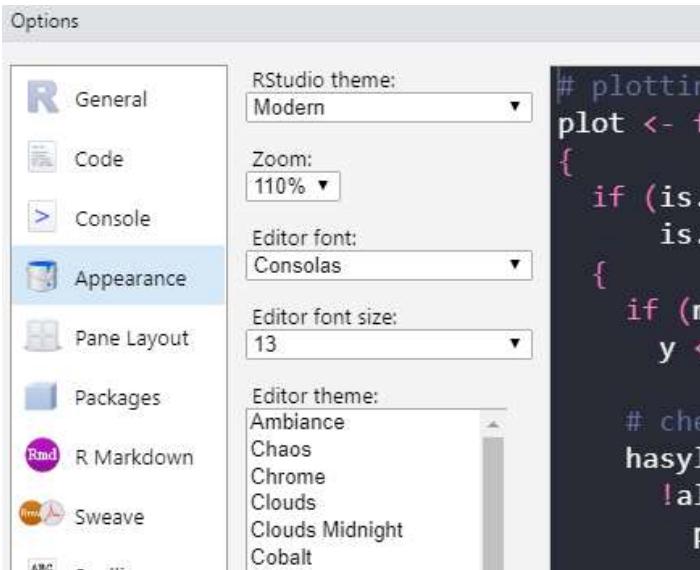


BOAS PRÁTICAS E MUDANÇA DE VISUAL

3º ABRIRÁ UMA CAIXA DE DIÁLOGO
CHAMADA “OPTION”

4º CLIQUE EM APPEARANCE

5º ESCOLHA A FONTE, O TAMANHO, O
ZOOM E ETC.

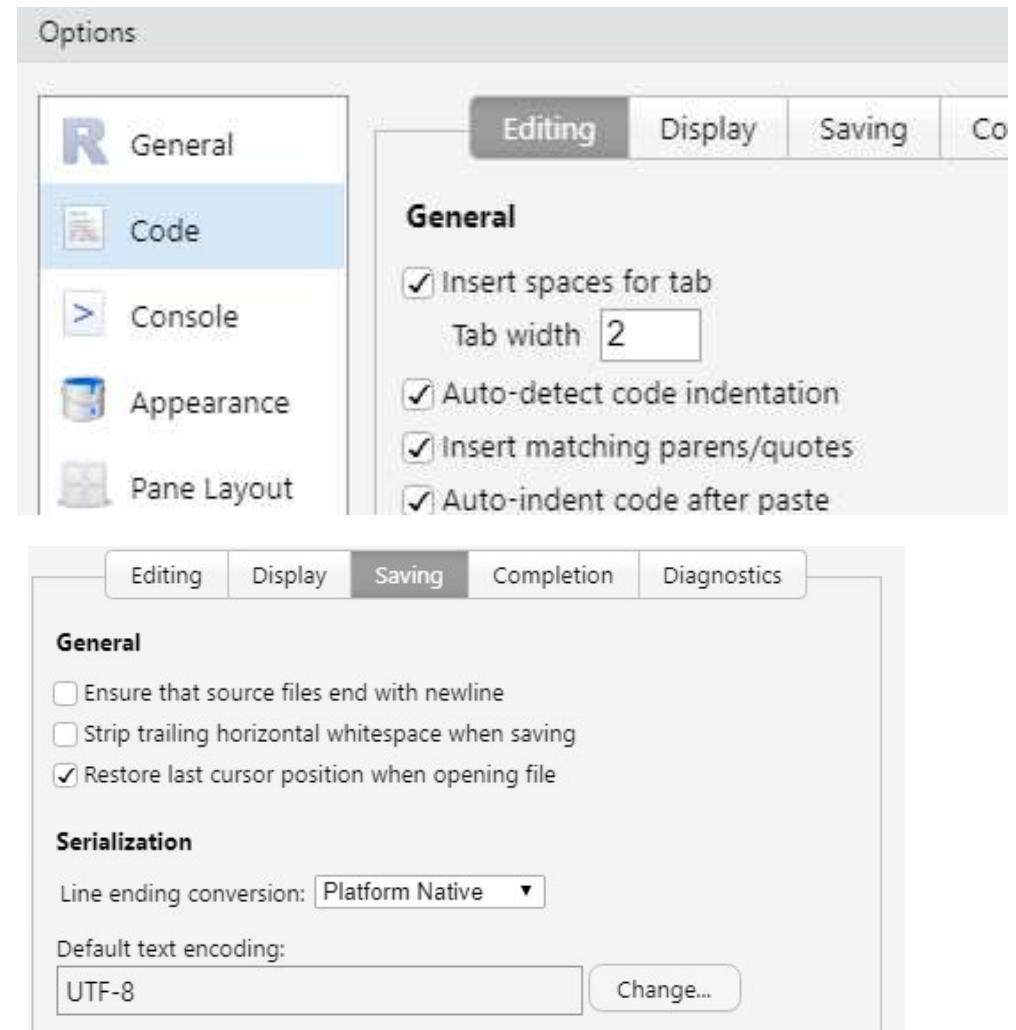


BOAS PRÁTICAS E MUDANÇA DE VISUAL

6º MUDANÇA DO CÓDIGO DE SALVAMENTO

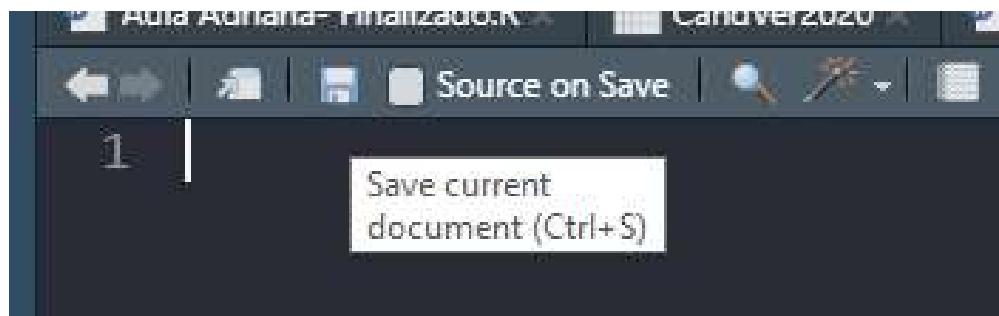
LINGUAGEM UNIVERSAL

- É ACONSELHÁVEL SALVAR OS SCRIPTS PARA QUE DEPOIS A ATIVIDADE DESENVOLVIDA POSSA SER RETOMADA, COMPARTILHADA E REPRODUZIDA.
- SALVE OS SCRIPTS EM UM FORMATO UNIVERSAL, POIS DESSA FORMA AO SER COMPARTILHADO O FORMATO NÃO SOFRERÁ MODIFICAÇÕES.
 - ENTÃO NA MESMA CAIXA DE DIÁLOGO CLIQUE EM CODE
- ESCOLHA O TIPO DE SALVAMENTO CLICANDO EM CHANGE > OPTE SEMPRE PELO TIPO UNIVERSAL UTF- 8
 - SALVE A LINGUAGEM DE ESCRITA EM PORTUGUÊS

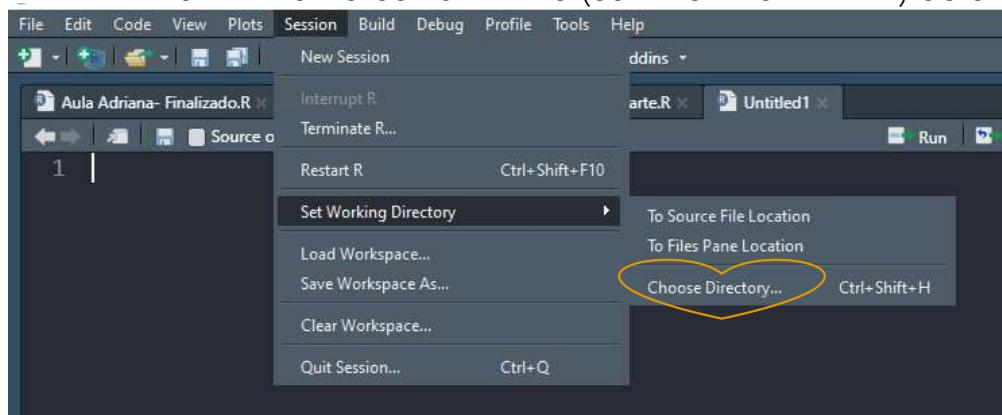


BOAS PRÁTICAS E MUDANÇA DE VISUAL

7º ORGANIZAÇÃO DO SUMÁRIO E SALVAR O SCRIPT



8º ENQUANTO ESCREVE O SCRIPT DE MANEIRA ORGANIZADA, SALVE EM SEU DIRETÓRIO DE TRABALHO,
PARA DEFINIR O DIRETÓRIO USE O ATALHO (CONTROL + SHIFT + H) OU SIGA OS PASSOS:



FUNÇÕES BÁSICAS DO R

Operadores	Descrição
<	Menor que
<=	Menor ou igual a
>	Maior que
>=	Maior ou igual a
==	Igual a
!=	Diferente de
!x	Não x
x y	x OU y
x & y	x E y

Objeto de atribuição
Atalho Alt - (menos)

<-

Pipi
Atalho Ctrl + shift + m

%>%

- Sempre que atribuímos valor a alguma coisa no R, essa coisa se chamará objeto.
- Tudo no R é um objeto.
- Nunca haverá dois objetos com o mesmo. Mas o mesmo valor poderá ter dois nomes diferentes.

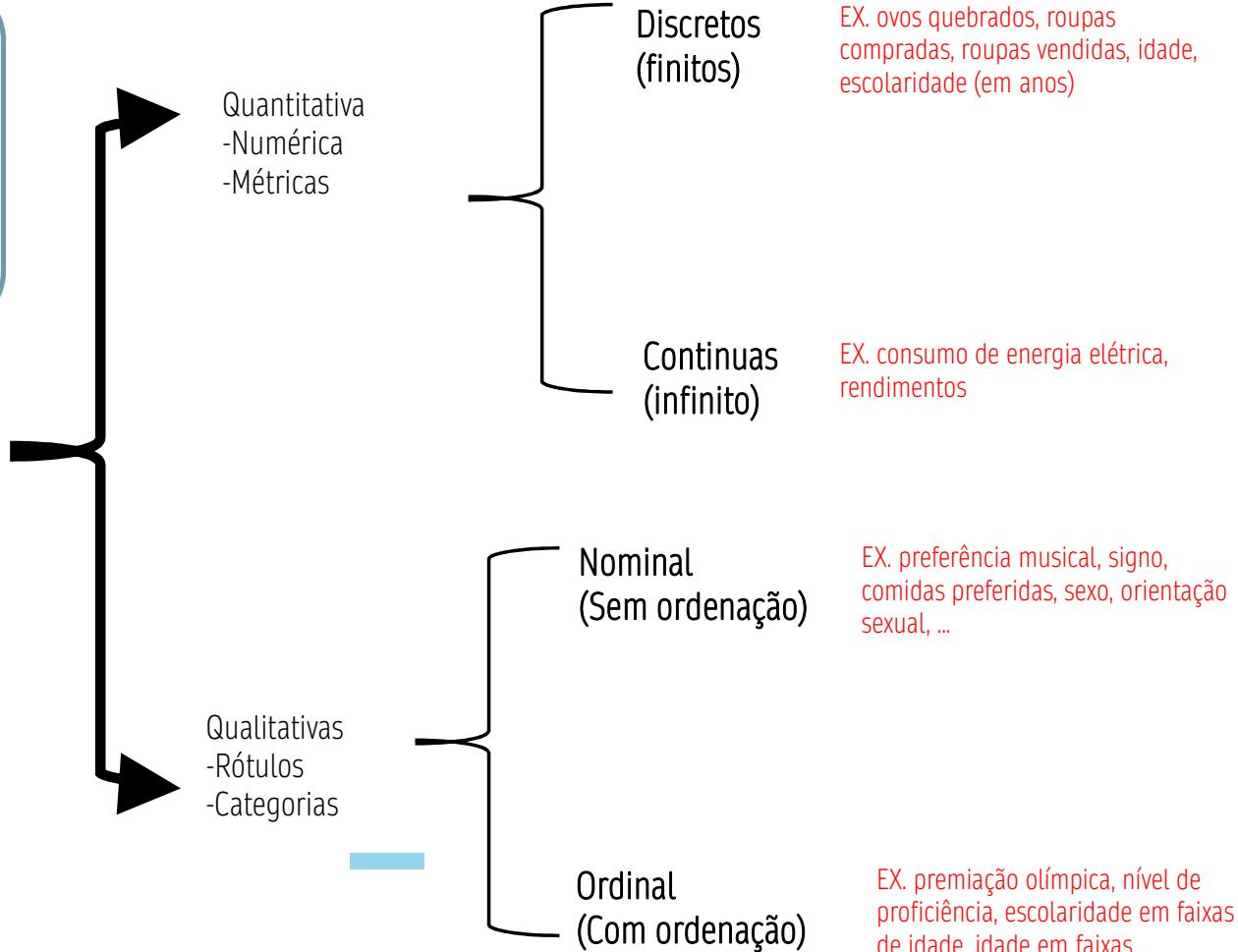
VARIÁVEIS



Antes da criação de um data frame é importante entender quais são os tipos de variáveis existentes, para:

- 1-Entender como inserir a variável na base
- 2-Escolher os testes mais adequados

Variável
1- Características
2-Atributos



CRIAÇÃO DE UM DATA FRAME

Primeiro criam-se os vetores > Depois juntamos os vetores através da função data.frame

```
#Criar data frame

Id ← c(1:4)
Nomes ← c("Ana", "Gilberto", "Rodrigo", "Marcela")
Peso ← c(75.6, 99, 62.8, 102)
Idades ← c(25, 18, 44, 23)
Escolaridade ← c("Graduação", "Mestrado", "Primário", "Graduação")
Exerc_Recomend ← c("Natação", "Pilates", "Musculação", "Corrida")
Comidas_preferidas ← c("Chocolate", "Sorvete", "Milho", "Pão")

Ficha_Pacientes ← data.frame(Id, Nomes, Peso,
                               Idades, Escolaridade,
                               Exerc_Recomend, Comidas_preferidas)
```



Lembre-se de salvar tanto o script, quanto a base de dados que você criou no diretório de trabalho correto.

```
#Função para ver o data frame
View(Ficha_Pacientes)

#Função para identificar o caminho em que o arquivo de script será salvo
getwd()
#[1] "C:/Users/nayar/OneDrive/8. AMBIENTE DE PROGRAMAÇÃO R/1. CURSO 2023"

#Função manual para selecionar o diretório de trabalho
setwd("C:/Users/nayar/OneDrive/8. AMBIENTE DE PROGRAMAÇÃO R/1. CURSO 2023")

#Função para salvar o data frame
save(Ficha_Pacientes, file = "Ficha_Pacientes.RData")
```

CRIAÇÃO DE UM DATA FRAME

Criamos um data frame, mas o que é um data frame?

Id	Nomes	Peso	Idades	Escolaridade	Exerc_Recomend	Comidas_preferidas	Escolaridade2
1	Ana	75.6	25	Graduação	Natação	Chocolate	Graduação
2	Gilberto	99.0	18	Mestrado	Pilates	Sorvete	Mestrado
3	Rodrigo	62.8	44	Primário	Musculação	Milho	Primário
4	Marcela	102.0	23	Graduação	Corrida	Pão	Graduação

É o mesmo que base de dados, trata-se de um conjunto de informações em que todas as colunas possuem o mesmo tamanho* e podem ser abertas/criadas/modificadas em qualquer ambiente de programação (python, stata, sql) ou software de análise de dados, como o Excel, Spss e afins.

Cada linha do data frame representa um caso/observação e cada coluna representa uma variável/questão. O cabeçalho das bases de dados devem ser compostos pelos nomes das variáveis.

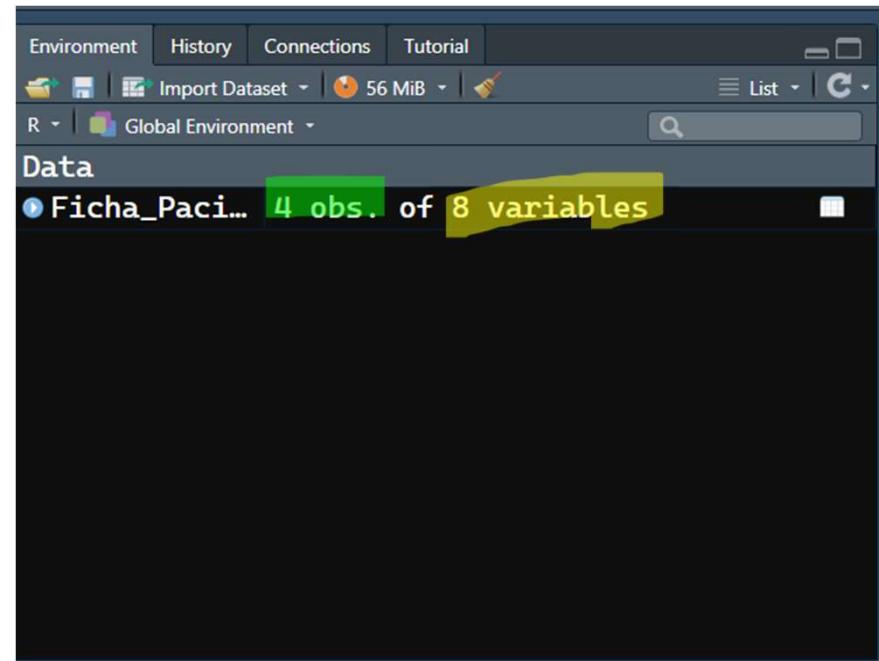
Variável é nome atribuído as questões estudadas, é um "objeto capaz de reter e representar um valor ou expressão". No ambiente de programação do R, tudo é objeto.

É importante que seja um nome curto, um resumo e que represente a questão, como "Nome", "Peso", ...

A principal característica de uma variável é sua capacidade de variar, portanto não existe variável com valor único. Se não varia, não é variável, é constante.

CRIAÇÃO DE UM DATA FRAME

- Quando criamos um data frame suas informações aparecem no Environment
- Apesar da base criada aparecer, isso não significa que ela está salva, por isso é importante salvá-la a cada modificação, utilizando a função save, ensinada no slide 15.
- Sobre salvamentos e aberturas de base de dados trataremos mais detalhadamente na seção sobre esse assunto.





ANÁLISE EXPLORATÓRIA E ORGANIZAÇÃO DOS DADOS

TÓPICOS

1-INTRODUÇÃO AO R

2- ANÁLISE EXPLORATÓRIA E
MANIPULAÇÃO DOS DADOS

3- SALVAMENTO E ABERTURA

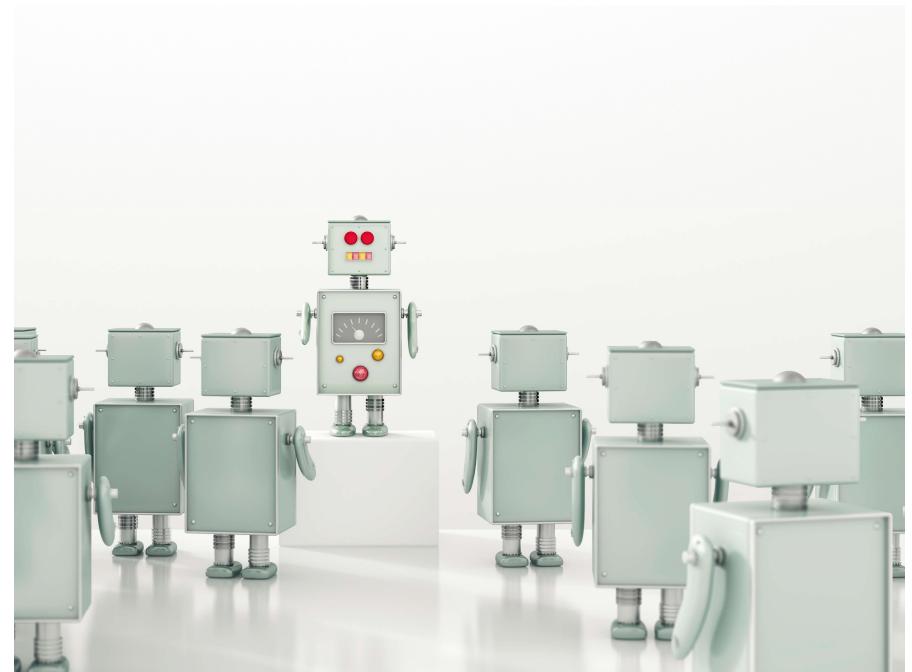
4- ANÁLISES DESCRIPTIVAS

5- APRESENTAÇÃO GRÁFICA

6- PROCESSAMENTO DE DADOS

7- ANÁLISES INFERENCIAIS

- Funções para observar a estrutura dos dados
- Seleção
- Exclusões
- Inserções e filtros
- Alterações
- Tipos de variáveis



ANÁLISE DO DATA FRAME

Estrutura dos dados

```
#Função para análise das classes de cada variável
str(Ficha_Pacientes)
# 'data.frame': 4 obs. of 7 variables:
# $ Id : int 1 2 3 4
# $ Nomes : chr "Ana" "Gilberto" "Rodrigo" "Marcela"
# $ Peso : num 75.6 99 62.8 102
# $ Idades : num 25 18 44 23
# $ Escolaridade : chr "Graduação" "Mestrado" "Primário" "Graduação"
# $ Exerc_Recomend : chr "Natação" "Pilates" "Musculação" "Corrida"
# $ Comidas_preferidas: chr "Chocolate" "Sorvete" "Milho" "Pão"
```

- número, valor real, *numeric, double*
- texto, string, *character*, caracteres
- lógico, *logical*, booleano, valor TRUE/FALSE

Int são os números inteiros

- O R identifica o tipo de variável e diz qual seu tipo.
- É possível que o tipo não esteja adequado, nesse caso é necessário alterar o tipo usando a função “as.” com o tipo que queremos que seja inserido na variável, e depois para verificar se o tipo está correto usaremos a função “is.”. Falaremos mais sobre isso, em uma etapa mais avançada do curso

```
#A base de dados deverá ter um nome de coluna e linhas
#Para saber esses nomes use as funções rownames e colnames
rownames(Ficha_Pacientes)
#[1] "1" "2" "3" "4"
colnames(Ficha_Pacientes) #Usando a função names também
#Conseguimos ver todas as colunas
# [1] "Id"          "Nomes"        "Peso"         "Idades"
# [5] "Escolaridade" "Exerc_Recomend" "Comidas_preferidas"
```

ANÁLISE DO DATA FRAME - SELEÇÕES

```
#Análise das médias
print(mean(Ficha_Pacientes$Idades))
#[1] 27.5
summary(Ficha_Pacientes$Idades)
# Min. 1st Qu. Median Mean 3rd Qu. Max.
# 18.00 21.75 24.00 27.50 29.75 44.00
summary(Ficha_Pacientes$Exerc_Recomend)
# Length Class Mode
# 4 character character
table(Ficha_Pacientes$Escolaridade)
# Graduação Mestrado Primário
# 2 1 1
```



O cífrão é um caractere importante, pois permite que selecionemos um objeto dentro de outro.

Permite chamar uma variável dentro de uma base específica

```
#Selação dos dados indicando 1º a linha e 2º a coluna
Ficha_Pacientes[2,5]
#[1] "Mestrado"
Ficha_Pacientes[2,5, drop=F]
# Escolaridade
# 2 Mestrado
Ficha_Pacientes[2,c("Peso", "Idades", "Nomes")]
# Peso Idades Nomes
# 2 99 18 Gilberto
```

MANIPULAÇÃO DO DATA FRAME - EXCLUSÕES

```
#Exclusão da primeira linha  
Ficha_Pacientes[-1, ]  
#Exclusão da primeira linha e última coluna  
Ficha_Pacientes[ -1 , -7]
```



Id	Nomes	Peso	Idades	Escolaridade	Exerc_Recomend	Comidas_preferidas
2	Gilberto	99.0	18	Mestrado	Pilates	Sorvete
3	Rodrigo	62.8	44	Primário	Musculação	Milho
4	Marcela	102.0	23	Graduação	Corrida	Pão



Lembre-se que os resultados somente estão impressos na base, mas não estão salvos

```
#Exclusões de colunas  
## deleta quant_filhos  
Ficha_Pacientes$Quant_filhos ← NULL  
  
## deleta colunas específicas,  
#O mesmo pode ser feito para linhas  
Ficha_Pacientes ← Ficha_Pacientes[, c(-4, -6)]
```

! Todas essas inclusões e exclusões serão muito importante na manipulação real de dados.

MANIPULAÇÃO DO DATA FRAME -INSERÇÕES E FILTROS

```
#Inserir novos dados em toda a base  
Ficha_Pacientes$Sexo ← c("F", "M", "M", "F")  
Ficha_Pacientes$Quant_filhos ← c(4:7)
```

```
#Uma outra forma de inserção de dados  
#é utilizando uma função muito importante no rbase  
# a função cbind
```

```
Ficha_Pacientes ← cbind(Ficha_Pacientes,  
                         Prim_emprego =  
                         c("sim", "nao", "nao", "sim"))
```

! O subset é uma forma de selecionar e criar bases menores, inserindo somente suas variáveis de interesse- Rbase

FILTRAR USANDO O SUBSET = SUBCONJUNTO

```
#Subset:  
#Quero uma base de dados com apenas 5 variáveis  
subset(Ficha_Pacientes, select = c("Id", "Nomes", "Peso",  
                                    "Idades", "Sexo"))  
  
Base_menor ← subset(Ficha_Pacientes, select =  
                      c("Id", "Nomes", "Peso",  
                        "Idades", "Sexo"))
```

```
#Mas eu queria apenas pessoas do sexo masculino  
Base_menor ← subset(Ficha_Pacientes,  
                     Sexo == "M",  
                     select = c("Id", "Nomes", "Peso", "Idades", "Sexo"))  
  
#Mas eu tbém queria pessoas com idades maiores ou  
#igual a 20 anos  
Base_menor ← subset(Ficha_Pacientes,  
                     Idades ≥ 20,  
                     select = c("Id", "Nomes", "Peso", "Idades", "Sexo"))
```

MANIPULAÇÃO DO DATA FRAME -FILTROS E ALTERAÇÕES

```
#Filtro usando o filter
# seleciona apenas colunas numéricas
Filter(is.numeric, Ficha_Pacientes)
# Id Peso Idades
# 1 1 75.6    25
# 2 2 99.0    18
# 3 3 62.8    44
# 4 4 102.0   23

# seleciona apenas colunas de texto
Filter(is.character, Ficha_Pacientes)
#      Nomes Escolaridade Exerc_Recomend Comidas_preferidas
# 1     Ana   Graduação       Natação      Chocolate
# 2 Gilberto Mestrado       Pilates       Sorvete
# 3 Rodrigo  Primário      Musculação      Milho
# 4 Marcela Graduação       Corrida        Pão
# Separando dados das colunas#####
library(stringr)

#Inserindo uma variável com 2 informações
Ficha_Pacientes$Nome_Mãe ←
c('John, Mae', 'Maude, Lebowski', 'Mia, Amy', 'Andy, James')

#Separando
str_split_fixed(Ficha_Pacientes$Nome_Mãe, ", ", 2)

#salvando na base |
Ficha_Pacientes$Nomes ←
  str_split_fixed(Ficha_Pacientes$Nome_Mãe, ", ", 2)
```

MANIPULAÇÃO DO DATA FRAME -ALTERAÇÕES

Alterando valores individuais

```
#Vamos alterar a idade da Marcela de 23 para 53  
Ficha_Pacientes$Idades[4] ← 53  
#Seguindo essa mesma lógica podemos alterar o nome das variáveis  
#Ou então criar uma cópia com um nome diferente|
```

Alterando grandes conjuntos de dados

```
library(memisc)  
  
#Quero transformar a idade, fazer com que ela deixe de  
#Ser numérica e se torne categorica (faixas de idade)  
Ficha_Pacientes$Idade_Faixas ← recode(Ficha_Pacientes$Idades,  
"Jovens" ← c(18:25), "Idoso" ← 44)  
#Assim podemos alterar qualquer variável, mas é importante  
#lembrar que tipo de variável estamos alterando|
```

! Antes de usar o recode
pense no tipo de variável
que está sendo
modificada

MANIPULAÇÃO DO DATA FRAME -TIPOS DE VARIÁVEIS NO AMBIENTE DE PROGRAMAÇÃO

- No slide 14 nós vimos que existem tipos diferentes de variáveis, essa informação é essencial em várias etapas da análise dos dados, desde a construção do data frame, até a utilização de testes mais robustos estatisticamente.
- Nesse momento veremos a importância dos tipos das variáveis para análise dos dados e recodificação
- No ambiente de programação do r, os tipos de variáveis recebem nomes em inglês, os principais são:

FACTOR são um tipo de dado especial usado principalmente para representar variáveis categóricas.
Quando você especifica um objeto como um fator, automaticamente o R irá atribuir uma ordem para esses dados

Nomes	Peso	Idades	Escolaridade
Gilberto	99.0	18	Mestrado
Marcela	102.0	23	Graduação
Ana	75.6	25	Graduação
Rodrigo	62.8	44	Primário

MANIPULAÇÃO DO DATA FRAME -TIPOS DE VARIÁVEIS NO AMBIENTE DE PROGRAMAÇÃO

```
#Podemos alterar qualquer variável, mas é importante  
#lembrar que tipo de variável estamos alterando  
#e entender como essa variável estará após as alterações|
```

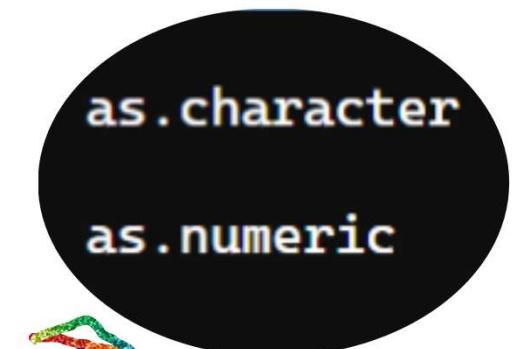
```
Ficha_Pacientes$Escolaridade ←  
  as.factor(Ficha_Pacientes$Escolaridade)
```

```
#Olhar a estrutura dos dados  
str(Ficha_Pacientes)
```

```
#Verificar os níveis  
Ficha_Pacientes$Escolaridade  
#Levels: Graduação Mestrado Primário
```

```
#Se eu não concordar com a ordem atribuída pelo R,  
#Se essa ordem não fizer sentido para as análises estatísticas  
#então eu posso alterar a ordem usando a função levels  
Ficha_Pacientes$Escolaridade2 ←  
  factor(Ficha_Pacientes$Escolaridade,  
levels = c("Primário", "Graduação", "Mestrado"))
```

```
#as.numeric/as.character####  
#Em geral o ambiente de programação entende quais são  
#os tipos das variáveis, porém pode ser que ele entenda errado  
#ou que em razão da abertura da base de dados seja necessário  
#fazer alguns ajustes, dizendo para o ambiente o que é numérico  
#e o que é caracter, nesses casos usamos as funções as. =como
```



as.character
as.numeric



SALVAMENTOS E ABERTURA DA BASE DE DADOS

TÓPICOS

- 1-INTRODUÇÃO AO R
- 2- ANÁLISE EXPLORATÓRIA E MANIPULAÇÃO DOS DADOS
- 3- SALVAMENTO E ABERTURA
- 4- ANÁLISES DESCRIPTIVAS
- 5- APRESENTAÇÃO GRÁFICA
- 6- PROCESSAMENTO DE DADOS
- 7- ANÁLISES INFERENCIAIS

- CSV
- XLSX
- STATA
- SPP
- Abertura clicável

EXTRA:

BASES DE DADOS COMUMENTE UTILIZADAS NA NOSSA ÁREA

- LAPOP- Barômetro das Americas
- CEPESP DATA
- Datafolha ibope
- CESOP - ESEB
- Latinobarômetro
- Freedom house
- Vdem
- Databank

SALVAMENTO E ABERTURA DA BASE DE DADOS

- Na seção inicial vimos uma forma de salvamento dos dados e abertura da base. Nessa seção veremos diversas outras formas.
- A forma vista para salvamento é mais comum dentro do ambiente do R, ".RData"
- A abertura da base ocorreu através da função "view", porém essa função somente irá funcionar se a base de dados já estiver carregada dentro do R, então ela serve somente para ver o que já está lá. Portanto, agora veremos formas de abrir bases de dados salvas anteriormente no computador.

1º CSV

```
#CSV salvamento#####
write.csv2(Ficha_Pacientes, file = "Ficha2.csv",
           quote = F, #Dividir por ponto e vírgulas
           row.names = F,#remover a primeira coluna com id do sistema
           fileEncoding = "latin1")#função pra definir a linguagem
#CSV abertura#####
#Precisaremos do pacote readr
library(readr)
#abertura
Ficha2 ← read_delim("Ficha2.csv", delim = ";",
                      locale = locale(encoding = "Latin1"))
```



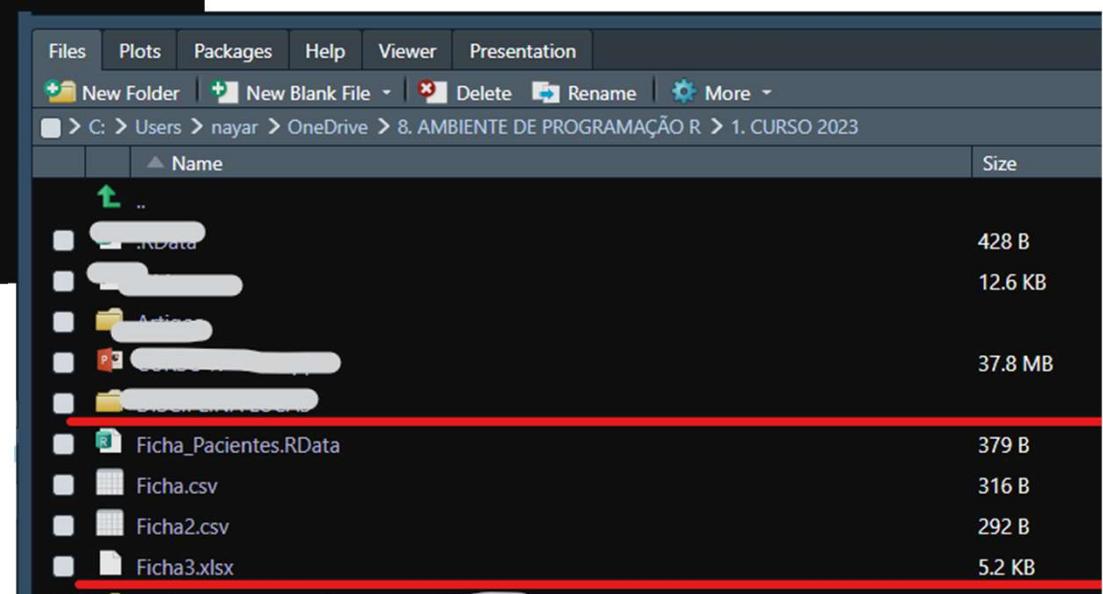
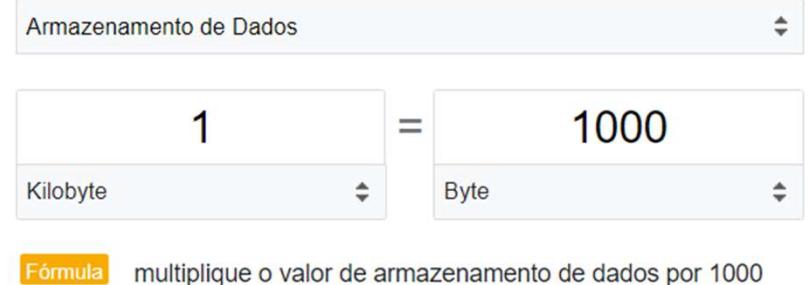
Lembre-se de sempre definir o seu diretório de trabalho antes de tentar abrir ou salvar qualquer base.

SALVAMENTO E ABERTURA DA BASE DE DADOS

2º XLSX

```
#Precisaremos do pacote  
library(writexl)  
  
#salvando  
writexl::write_xlsx(Ficha_Pacientes, path = "Ficha3.xlsx")  
  
#Xlsx Abertura####  
#Pacote:  
library(readxl)  
  
Ficha3 ← read_excel("Ficha3.xlsx")  
  
View(Ficha3)
```

! Atenção para o tamanho dos documentos



SALVAMENTO E ABERTURA DA BASE DE DADOS

3º Abertura em stata e spp

Brazil	2014	Elizabeth Zechmeister
Data Files and Notes		
Brazil 2014 Tech Info	Brazil 2014 Stata	Brazil 2014 SPSS
Brazil 2014 Questionnaire - Portuguese		
<p>To download the files, click on the links shown or right-click on the file name and save it. You will not be able to open STATA or SPSS files unless your computer has the corresponding software available. The Questionnaires and Technical Information documents are saved as .pdf files which can be opened with pdf reader software available for free online such as Adobe Acrobat Reader get.adobe.com/reader/ and Foxit Reader www.foxitsoftware.com. Stata files contain labels in both English and Spanish. To change the language, use the commands: <code>label language en</code> OR <code>label language es</code>.</p>		
<p>Change Log:329664568Brazil AmericasBarometer 2014 Change Log v3.0_W.pdf</p>		

Site do lapop:

<http://datasets.americasbarometer.org/database/index.php?freeUser=true>

Atualmente os dados do LAPOP são disponibilizados somente em Stata, porém até 2014 eram disponibilizados em:

- Stata
- Spss
- e Tech Info

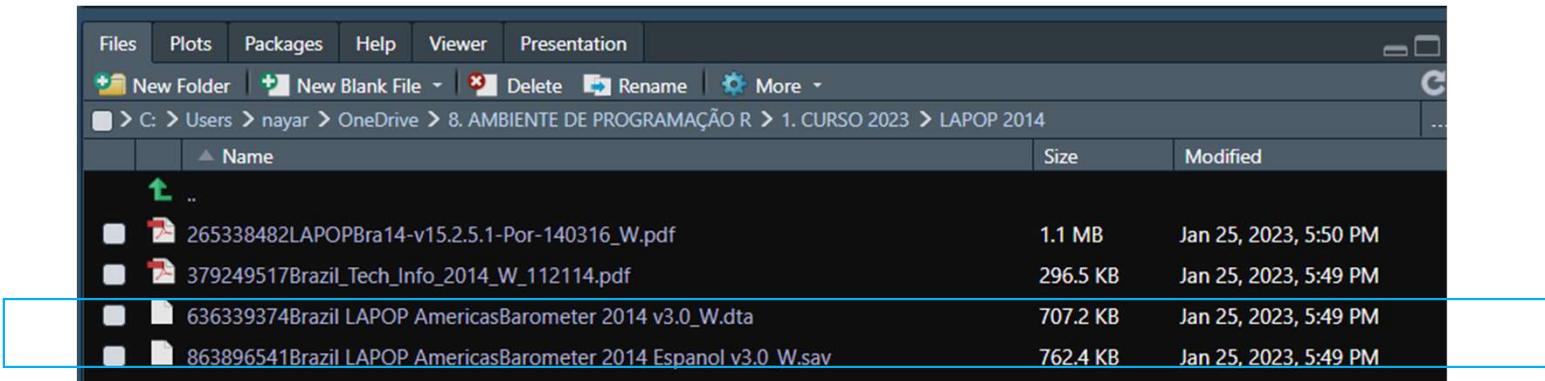
Brincadeirinha! Não vá confundir as informações técnicas com formato da base rsrs



SALVAMENTO E ABERTURA DA BASE DE DADOS

Baixe em seu diretório de trabalho todo aquele matéria: base de dados, questionário e informações técnicas.

Para quem comumente utiliza SPSS, sabe que o formato de salvamento é sav, já do STATA é dta



The screenshot shows a file explorer window with the following details:

- Toolbar:** Files, Plots, Packages, Help, Viewer, Presentation.
- Top Bar:** New Folder, New Blank File, Delete, Rename, More.
- Path:** C: > Users > nayar > OneDrive > 8. AMBIENTE DE PROGRAMAÇÃO R > 1. CURSO 2023 > LAPOP 2014
- File List:** A table showing four files:

Name	Size	Modified
265338482LAPOPBra14-v15.2.5.1-Por-140316_W.pdf	1.1 MB	Jan 25, 2023, 5:50 PM
379249517Brazil_Tech_Info_2014_W_112114.pdf	296.5 KB	Jan 25, 2023, 5:49 PM
636339374Brazil LAPOP AmericasBarometer 2014 v3.0_W.dta	707.2 KB	Jan 25, 2023, 5:49 PM
863896541Brazil LAPOP AmericasBarometer 2014 Espanol v3.0_W.sav	762.4 KB	Jan 25, 2023, 5:49 PM

SALVAMENTO E ABERTURA DA BASE DE DADOS

💡 Vamos abrir e salvar bases em **Sav**

```
#Abertura de dados reais#####
#Abertura em SAV#####

#Pacote:
library(haven)

#Como o nome da base é bastante grande, vamos chamar por um
#nome mais resumido
Brasil2014 ← read_sav("LAPOP 2014/863896541Brazil_LAPOP_AmericasBarometer 2014_Espanol_v3.0_W.sav")
View(X863896541Brazil_LAPOP_AmericasBarometer_2014_Espanol_v3_0_W)

#Após realizar as alterações na base
#se quiser salvar no mesmo formato, use:
write_sav(data = Brasil2014, path = "Brasil2014.sav")
```

💡 Em dta o processo é semelhante veja:

```
#Abertura em DTA#####
#Use o mesmo pacote haven
Brasil2014dta ←
  read_dta("LAPOP 2014/636339374Brazil_LAPOP_AmericasBarometer 2014_v3.0_W.dta")

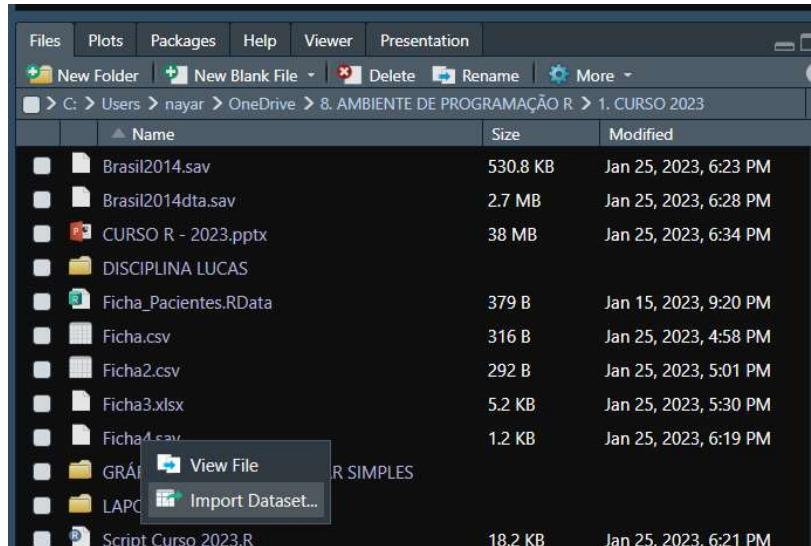
#Após realizar as alterações na base
#se quiser salvar no mesmo formato, use:
write_dta(data = Brasil2014dta, path = "Brasil2014dta.sav")
```

Em seu diretório de trabalho as pastas devem aparecer dessa forma:

Brasil2014.sav	✓
Brasil2014dta.sav	✓

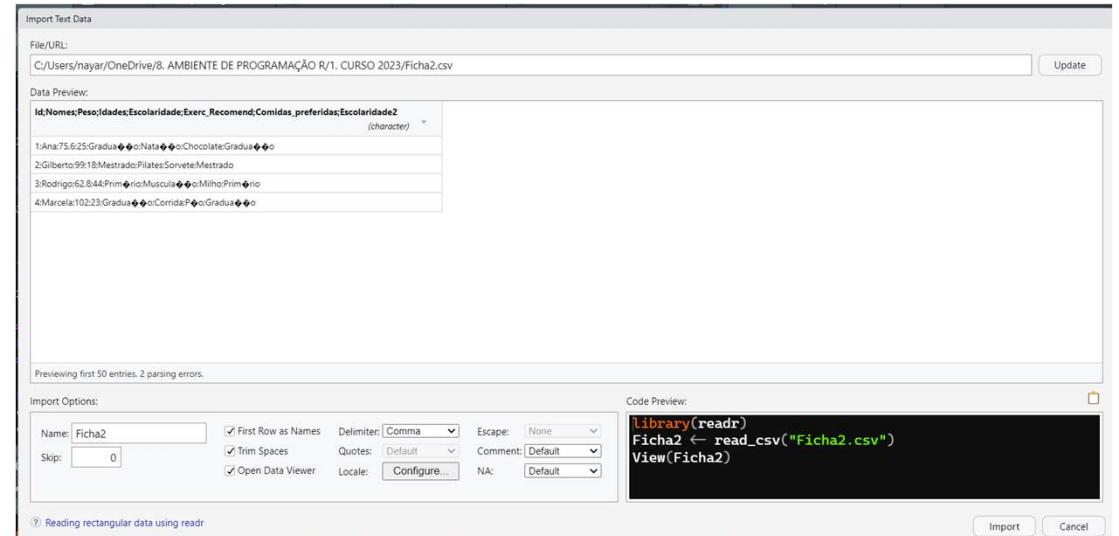
SALVAMENTO E ABERTURA DA BASE DE DADOS

Até aqui vimos algumas formas de abertura de base de dados, essas são as mais utilizadas, contudo é importante que você esteja ciente o **ambiente de programação do R consegue ler bases nos mais diferentes formatos**, por isso agora vamos aprender uma forma manual de abrir. Isso poderá lhe auxiliar com outros formatos.



Clique com o botão direito na base e peça para “Import Dataset...”

Alguns formatos como o csv você poderá configurar a base antes de abrir. Para isso essa janela de comunicação abrirá:



SALVAMENTO E ABERTURA DA BASE DE DADOS

Import Text Data

File/URL: C:/Users/nayar/OneDrive/8. AMBIENTE DE PROGRAMAÇÃO R/1. CURSO 2023/Ficha2.csv Update

Data Preview:

	Id;Nomes;Peso;Idades;Escolaridade;Exerc_Recomend;Comidas_preferidas;Escolaridade2 (character)
1	Ana;75,6;25;Gradua o;Nata o;Chocolate;Gradua o
2	Gilberto;99,18;Mestrado;Pilates;Sorvete;Mestrado
3	Rodrigo;62,8;44;Prim rio;Muscula o;Milho;Prim rio
4	Marcela;102,23;Gradua o;Corrida;P o;Gradua o

1º defina como será a divisão dos dados aqui: Escolha ponto e vírgula

Delimiter: Semicolon

Previewing first 50 entries. 2 parsing errors.

Import Options:

Name: Ficha2	<input checked="" type="checkbox"/> First Row as Names	Delimiter: Comma	Escape: None
Skip: 0	<input checked="" type="checkbox"/> Trim Spaces	Quotes: Default	Comment: Default
	<input checked="" type="checkbox"/> Open Data Viewer	Locale: Configure...	NA: Default

Code Preview:

```
library(readr)
Ficha2 ← read_csv("Ficha2.csv")
View(Ficha2)
```

② Reading rectangular data using readr Import Cancel

SALVAMENTO E ABERTURA DA BASE DE DADOS

Import Text Data

File/URL:
C:/Users/nayar/OneDrive/8. AMBIENTE DE PROGRAMAÇÃO R/1. CURSO 2023/Ficha2.csv

Data Preview:

Id (double)	Nomes (character)	Peso (double)	Idades (double)	Escolaridade (character)	Exerc_Recomend (character)	Cor (character)
1	Ana	756	25	Graduaço	Nataço	Cho
2	Gilberto	99	18	Mestrado	Pilates	Sor
3	Rodrigo	628	44	Prim&rio	Musculaço	Mil
4	Marcela	102	23	Graduaço	Corrida	Pç

Previewing first 50 entries.

Import Options:

Name: Ficha2 First Row as Names Trim Spaces Open Data Viewer

Skip: 0 Delimiter: Semicolon Quotes: Default Comment: Default NA: Default

Configure Locale

Date Name: en Encoding: UTF-8
Date Format: %AD Time Format: UTC
Decimal Mark: . Grouping Mark: ,
Time Zone: UTC

Locales in reader

Config Other...

Code Preview:

```
library(readr)
Ficha2 <- read_delim("Ficha2.csv", delim = ";",
                      escape_double = FALSE, trim_ws = TRUE)
View(Ficha2)
```

Import Cancel

2º Configure a base para a linguagem que está sendo utilizada, no caso Latin1, isso fará com que a base não fique com caracteres estranhos no lugares dos acentos.
Depois de escrever Latin1 dê ok.

Encoding Identifier

Please enter an encoding identifier. For a list of valid encodings run iconvlist().

Latin1 OK Cancel

SALVAMENTO E ABERTURA DA BASE DE DADOS

Import Text Data

File/URL: C:/Users/nayar/OneDrive/8. AMBIENTE DE PROGRAMAÇÃO R/1. CURSO 2023/Ficha2.csv Update

Data Preview:

Id (double)	Nomes (character)	Peso (double)	Idades (double)	Escolaridade (character)	Exerc_Recomend (character)	Comidas_preferidas (character)	Escolaridade2 (character)
1	Ana	756	25	Graduação	Natação	Chocolate	Graduação
2	Gilberto	99	18	Mestrado	Pilates	Sorvete	Mestrado
3	Rodrigo	628	44	Primário	Musculação	Milho	Primário
4	Marcela	102	23	Graduação	Corrida	Pão	Graduação

Previewing first 50 entries.

Import Options:

Name: Ficha2	<input checked="" type="checkbox"/> First Row as Names	Delimiter: Semicolon	Escape: None
Skip: 0	<input checked="" type="checkbox"/> Trim Spaces	Quotes: Default	Comment: Default
	<input checked="" type="checkbox"/> Open Data Viewer	Locale: Configure...	NA: Default

Code Preview:

```
library(readr)
Ficha2 ← read_delim("Ficha2.csv", delim = ";",
                      escape_double = FALSE, locale = locale(encoding =
"Latin1"),
                      trim_ws = TRUE)
```

! Quando a base de dados estiver correta, basta clicar em Import

Lembre-se a maioria das bases pode ser aberta dessa forma.

Import Cancel

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

Quais são as outras agendas de pesquisas da nossa área que demandam bases de dados?

1-Questões eleitorais? Sucesso eleitoral?

Possibilidades para aquisição dos dados:

1- <https://cepesdata.io/>

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?



PÁGINA INICIAL SOBRE CONSULTAS API MAPAS ELEITORAIS

CEPESP DATA

O Centro de Política e Economia do Setor Público (CEPESP) da Fundação Getulio Vargas (FGV) disponibiliza uma série de bases de dados eleitorais pesquisáveis, o CepespData. O CepespData tem por objetivo facilitar o acesso integrado às bases de dados disponibilizadas pelo TSE em seu repositório de dados. Foi desenvolvido com apoio da FAPESP ao projeto As Instituições Políticas Subnacionais: Um Estudo Comparativo dos Estados Brasileiros (processo: 2013/15658-1) e do Tribunal Superior Eleitoral – TSE.

 Material de apoio:	Como começar	Leia-me	Dicionário de variáveis	Tabelas auxiliares
Consultar Eleições	Consultar prestação de contas	Consultar informações partidárias	Aplicativos	
 Resultados eleitorais VEJA MAIS →	 Bens de candidatos VEJA MAIS →	 Filiados VEJA MAIS →	 Indicadores eleitorais VEJA MAIS →	
 Resumo da eleição VEJA MAIS →	 Doações eleitorais (EM BREVE) VEJA MAIS →	 Secretários VEJA MAIS →	 Carreiras VEJA MAIS →	
 Bancos desagregados VEJA MAIS →	 Despesas eleitorais (EM BREVE) VEJA MAIS →	 Diretórios (EM BREVE) VEJA MAIS →	 Mapas eleitorais VEJA MAIS →	

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?



EAESP
EESP

CENTRO DE POLÍTICA
E ECONOMIA DO SETOR
PÚBLICO



IR PARA O CEPESP

PÁGINA INICIAL SOBRE CONSULTAS API MAPAS ELEITORAIS

Resultado de eleições por cargo

Consulta às bases de dados relativas às eleições especificando um cargo eletivo. As consultas podem ser também agregadas por partido e nível regional.

Cargo:

Agregação Regional:

Agregação Política:

Ano:

2014
2010
2006

Adicionar **VOTOS BRANCOS**

Adicionar **VOTOS NULOS**

Somente Eleitos

▶ Consultar

Selecionar colunas

PÁGINA INICIAL SOBRE CONSULTAS API MAPAS ELEITORAIS

CONSULTAS API MAPAS ELEITORAIS

Mapas por Município
Mapas por Local de Votação

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

2- Dados direto do sistema do TSE

estatística eleitoral

X | ☰ | 🎤 | 📸 | 🔎

Todas Imagens Notícias Videos Shopping Mais Ferramentas

Aproximadamente 3.790.000 resultados (0,41 segundos)

tse.jus.br
https://www.tse.jus.br › estatisticas › estatisticas ▾

Estatísticas - Tribunal Superior Eleitoral

Feitas por meio da replicação diária dos dados dos tribunais regionais **eleitorais** para uma base consolidada no TSE, organizada em quatro seções: ...

<https://www.tse.jus.br/eleicoes/estatisticas/estatisticas>

<https://www.tse.jus.br/eleicoes/estatisticas/estatisticas-de-eleitorado/evolucao-do-eleitorado>

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

The screenshot shows the homepage of the Tribunal Superior Eleitoral (TSE) website. The top navigation bar includes links for Serviços eleitorais, Eleições, Partidos, Comunicação, Jurisprudência, Legislação, Serviços judiciais, and Institucional, along with a search icon. Below the navigation is a breadcrumb trail: Home / Eleições / Estatísticas. On the left, a sidebar menu under 'Eleições' lists Biometria, Calendário eleitoral, Canal do Mesário, Contas eleitorais, Eleições 2022, Eleições, plebiscitos e referendos, and 'Estatísticas', which is highlighted with a yellow background. The main content area features a large heading 'Estatísticas'. Below it, a text block states: 'São feitas com base na replicação diária dos dados dos tribunais regionais eleitorais para uma base consolidada no Tribunal Superior Eleitoral, em horários predeterminados.' A section titled 'Conheça as estatísticas relacionadas às eleições:' contains a bullet point: '• Estatísticas eleitorais'. Another section titled 'Outras estatísticas' lists: '• Estatísticas do eleitorado', '• Relatório das eleições', and '• Portal de Dados Abertos'. At the bottom, there is a contact email: 'Dúvidas e sugestões devem ser encaminhadas para estatistica@tse.ius.br'.

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

 Estatísticas eleitorais

 Portal do TSE  Portal de dados abertos

[Página inicial](#)

[Estatísticas de eleição](#) >

[Estatísticas mensais](#) >

[Conjuntos de dados](#) >

[Arquivos gerados](#)

[Mapa do site](#)

Mapa do site

Estatísticas de eleição

Candidaturas

- [Estatísticas de candidaturas](#)
 - > Candidaturas por vaga
 - > Cargo
 - > Evolução de candidaturas
 - > Federação
 - > Motivo da cassação
 - > Partido
 - > Quantitativo e situação
 - > Reeleição
 - > Situação da candidatura
- [Perfil da candidatura](#)
 - > Cor/raça
 - > Estado civil
 - > Faixa etária
 - > Gênero
 - > Grau de instrução

Comparecimento/abstenção

- [Comparecimento/abstenção](#)
 - > Comparecimento/abstenção
 - > Estado civil
 - > Evolução do quantitativo
 - > Faixa etária
 - > Gênero
 - > Grau de instrução
 - > Pirâmide etária
 - > Tipo de deficiência

Eleitorado da eleição

- [Estatísticas do eleitorado](#)
 - > Detalhes municípios
 - > Eleitorado
 - > Evolução do eleitorado
 - > Perfil do eleitorado
 - > Estado civil
 - > Faixa etária
 - > Gênero
 - > Grau de instrução
 - > Nome social
 - > Pirâmide etária
- [Eleitorado com deficiência](#)
 - > Eleitorado com deficiência
- [Capilaridade](#)
 - > Capilaridade
 - > Seções

Filiação partidária da eleição

- [Estatísticas de filiação](#)
 - > Estado civil
 - > Evolução da filiação
 - > Faixa etária
 - > Filiação do eleitorado
 - > Gênero
 - > Grau de instrução
 - > Município
 - > Partido
 - > Pirâmide etária

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

2- Dados direto do sistema do TSE

Se quisermos somente os dados dos eleitores?



Para que ?

Para um cálculo amostral, *

Para a análise apenas do eleitorado

The screenshot shows the TSE website's sidebar menu and the 'Estatísticas' section content.

Sidebar Menu:

- Eleições, plebiscitos e referendos
- Estatísticas** (highlighted)
- História
- Resultados das eleições
- Urna eletrônica

Tags:

- #Eleitorado
- #Estatística

Gestor responsável:

Secretaria de Modernização, Gestão Estratégica e Socioambiental +

Estatísticas Section Content:

Depois de auditadas, as informações servem de base para a divulgação do perfil do eleitorado brasileiro nas eleições daquele mesmo ano.

Dados referentes ao perfil do eleitor brasileiro para as Eleições podem ser acessados na página de **Estatísticas do eleitorado da eleição**.

As informações – em especial, sobre faixa etária – levam em conta a situação do eleitor apto a votar no dia da votação no 1º turno.

Informamos, ainda, que a classificação das idades do eleitorado em faixas-etárias mensal é realizada considerando a data de extração dos dados.

As estatísticas disponíveis, a partir dos links abaixo, por sua vez, refletem os dados do eleitor no mês em que sua última atualização foi informada.

Clique nas opções disponíveis para ter acesso às estatísticas desejadas.

Estatísticas do eleitorado

Eleitores faltosos

Evolução do eleitorado (highlighted with a yellow oval)

Justificativa eleitoral

Votação manual e eletrônica

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

Estatísticas do eleitorado – Evolução do eleitorado

□ Parâmetros de pesquisa

Inicio: Fevereiro - 2023 Fim: Fevereiro - 2023
Abrangência: Região Região: Todas
Biometria: Todos **PESQUISAR**

□ Evolução do eleitorado

Abrangência	Inicio	Fim	%
CENTRO-OESTE	11.571.909	11.571.909	0,000
EXTERIOR	707.481	707.481	0,000
NORDESTE	42.373.706	42.373.706	0,000
NORTE	12.583.641	12.583.641	0,000
SUDESTE	66.705.538	66.705.538	0,000
SUL	22.550.759	22.550.759	0,000
TOTAL	156.493.034	156.493.034	0,000

[Exportar dados](#)

1 - 7

<https://dadosabertos.tse.jus.br/dataset/?groups=eleitorado> >
<https://dadosabertos.tse.jus.br/dataset/eleitorado-atual>

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

Eleitorado Atual

Organização


Tribunal Superior Eleitoral

TSE/AGEL
Assessoria de Gestão Eleitoral - AGEL

Leia mais

 **Social**

 **Twitter**

 **Facebook**

 **Licença**

Creative Commons Atribuição
[OPEN DATA](#)

 Conjunto de dados

 Grupos

 Fluxo de Atividades

Eleitorado Atual

Eleitorado Atual - Eleitorado Atual por Local de Votação - Perfil do Eleitorado por Seção Eleitoral

ATENÇÃO!

UTILIZE SOFTWARE ADEQUADO PARA A COMPLETA VISUALIZAÇÃO DOS DADOS!

Arquivos de dados com um grande número de linhas (particularmente aqueles com extensão .csv e .txt) podem não ser visualizados em sua totalidade a depender do software utilizado. Por exemplo, o Microsoft Excel tem limitação de 1.048.576 linhas. Para evitar o carregamento incompleto do arquivo, utilize softwares com suporte para abrir grande volume de dados ou aplique técnicas específicas para seu processamento.

Confira algumas alternativas a seguir:

- Use ferramentas de análise estatística, de Business Intelligence (BI), de banco de dados ou de análise de dados;
- Abra o arquivo no Excel usando a opção Obter Dados: utilize o Power Query para carregar o conjunto de dados completo e analisá-lo com tabelas dinâmicas;
- Use programas de editores de texto com suporte para grandes volumes de dados;
- Use qualquer outra ferramenta que suporte arquivos de grande volume de dados.

Dados e recursos

 **Eleitorado - Atual**
Todas as UFs 

 **Eleitorado por local de votação - Atual**
Todas as UFs 

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

2- Dados direto do sistema do TSE

Dados dos candidatos



Para que ?

Quero entender o sucesso eleitoral. Acredito que o que faz um candidato ter sucesso eleitoral é o sexo (homens têm mais chances que mulheres) acredito que é receita do candidato, mais dinheiro = mais votos (?).

The screenshot shows the TSE Open Data Portal homepage. The main navigation bar includes links for Conjuntos de dados, Organizações, Grupos, Sobre, Contato, and Perguntas frequentes. A search bar is also present. Below the navigation, a breadcrumb trail shows the user is at the 'Conjuntos de dados' section. On the left, there's a sidebar with filters for Organizações (TSE/AGEL), Grupos (Candidatos - 34), and Etiquetas (Ano 1933, Ano 1934, Ano 1945, Ano 1947). The main content area displays a search bar and a list of 34 datasets found for the 'Candidatos' group. The results are ordered by Relevância. A blue callout box with a large arrow points from the right towards the 'Candidatos' group filter, containing the text: 'Não tem o total de votos e nem as receitas do candidato'.

Portal de Dados Abertos do TSE

Conjuntos de dados Organizações Grupos Sobre Contato Perguntas frequentes Pesquisar

Home / Conjuntos de dados

Organizações
TSE/AGEL 34

Grupos
Candidatos 34

Etiquetas
Ano 1933 1
Ano 1934 1
Ano 1945 1
Ano 1947 1

Buscar conjuntos de dados...

34 conjuntos de dados encontrados

Ordenar por: Relevância

Grupos: Candidatos

Candidatos - 2022

Candidatos - Bens de Candidatos - Coligações - Vagas - Motivo da cassação - Redes sociais de candidatos - Fotos de candidatos - Proposta de governo - Certidões criminais - Notas...

Não tem o total de votos e nem as receitas do candidato

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

FILTROS OPCIONAIS

Campo
Cargo

Valor
Senador

- Eleitorado mensal
- Filiação partidária mensal
- Pesquisas eleitorais
- Resultados

Votação candidato

Dimensões
Ano de eleição
Cargo
Código município
Município
Nome candidato
Número candidato
Partido
Região
Situação totalização
Turno
UF
Zona

Métricas

Votos nominais
Votos válidos

Mas ainda não tem as
receitas

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

Conjunto de dados

Eleitorado da eleição
Eleitorado mensal
Eleitorado falso
Filiação partidária da eleição
Filiação partidária mensal
Mesárias e mesários
Justificativas
Pesquisas eleitorais
Prestação de contas
Processos eleitorais
Resultados
Arquivos gerados
Mapa do site

Conjuntos de dados \ **prestacao_contas**

CONJUNTOS DE DADOS

Prestação de contas eleitorais - Fundo partidário

Dimensões

- Ano de eleição
- Cor/raça
- Esfera partidária
- Gênero
- Município
- Partido
- UF

Métricas

- Quantidade de candidatos
- Recursos declarados
- Valor FP aplicado na campanha

Prestação de contas eleitorais - Receitas de candidatos

Dimensões

- Ano de eleição
- Cargo**
- Esfera
- Esfera partidária
- Espécie receita
- Fonte de receita
- Município
- Nome candidato
- Nome doador
- Origem da receita
- Partido
- Região
- Situação de candidatura
- Situação de totalização
- UF

Métricas

- Valor de receita**

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

cand_2022_BRASIL.csv

leiam.pdf

receita_sen.csv

votos_sen.csv

Agora eu tenho todos os dados, porém em bases separadas como podemos fazer a junção de tudo

Arquivo de Valore

PDF X

Arquivo de Valore

Arquivo de Valore

Para junção da base de dados usamos uma função chamada merge, que significa mesclar.

Esse é uma possibilidade, porém existem vários meios de junção de base de dados a depender do que você precisa fazer.

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

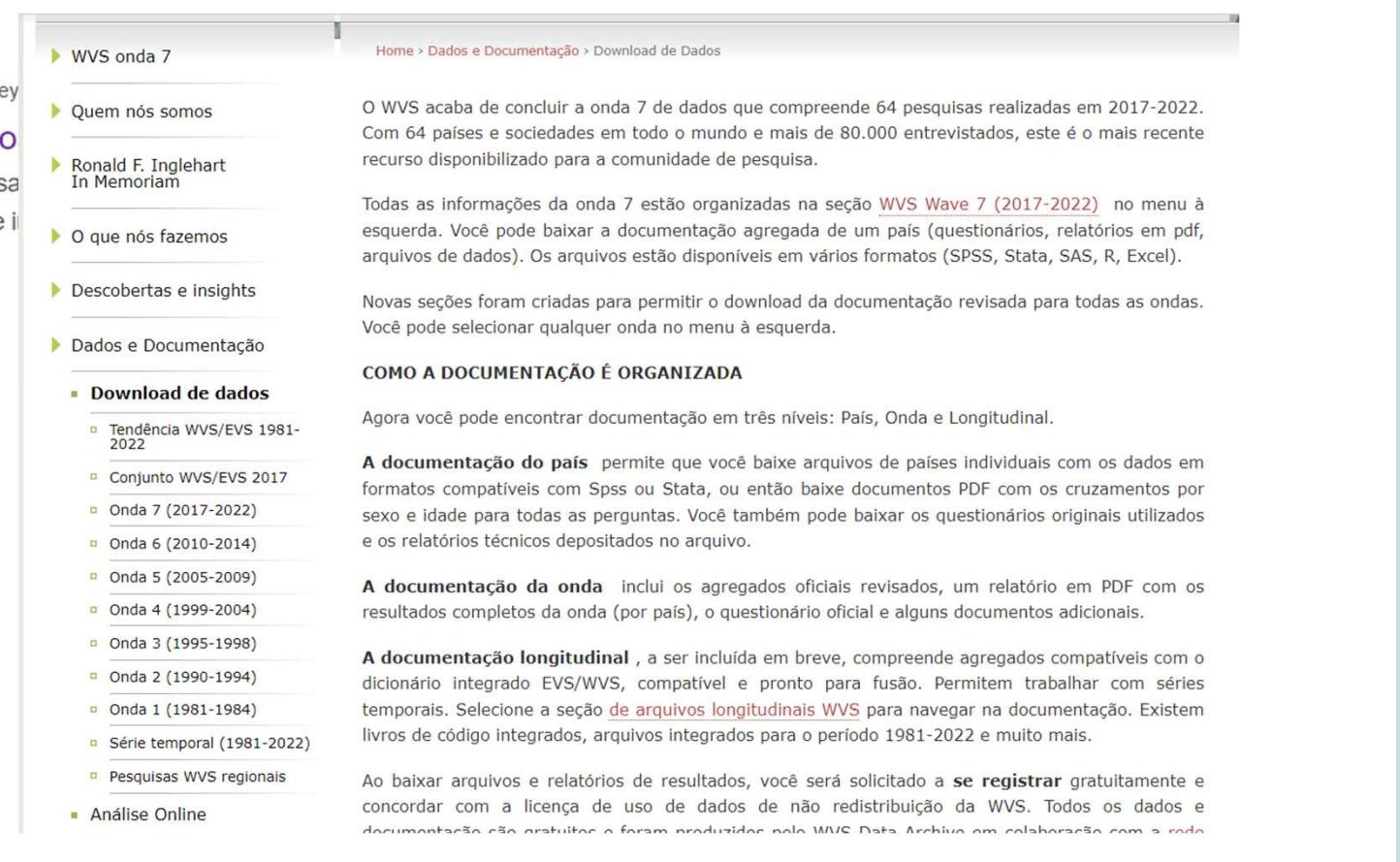
3- Outros dados de opinião pública <https://www.worldvaluessurvey.org/WVSContents.jsp>
WVS

 basedosdados.org

<https://basedosdados.org> › dataset › world-values-survey

World Values Survey (WVS) - Base do

O World Values Survey (**WVS**), um projeto de pesquisa pessoas, como elas mudam ao longo do tempo e que i



WVS onda 7

Home > Dados e Documentação > Download de Dados

O WVS acaba de concluir a onda 7 de dados que compreende 64 pesquisas realizadas em 2017-2022. Com 64 países e sociedades em todo o mundo e mais de 80.000 entrevistados, este é o mais recente recurso disponibilizado para a comunidade de pesquisa.

Todas as informações da onda 7 estão organizadas na seção [WVS Wave 7 \(2017-2022\)](#) no menu à esquerda. Você pode baixar a documentação agregada de um país (questionários, relatórios em pdf, arquivos de dados). Os arquivos estão disponíveis em vários formatos (SPSS, Stata, SAS, R, Excel).

Novas seções foram criadas para permitir o download da documentação revisada para todas as ondas. Você pode selecionar qualquer onda no menu à esquerda.

COMO A DOCUMENTAÇÃO É ORGANIZADA

Agora você pode encontrar documentação em três níveis: País, Onda e Longitudinal.

A documentação do país permite que você baixe arquivos de países individuais com os dados em formatos compatíveis com Spss ou Stata, ou então baixe documentos PDF com os cruzamentos por sexo e idade para todas as perguntas. Você também pode baixar os questionários originais utilizados e os relatórios técnicos depositados no arquivo.

A documentação da onda inclui os agregados oficiais revisados, um relatório em PDF com os resultados completos da onda (por país), o questionário oficial e alguns documentos adicionais.

A documentação longitudinal, a ser incluída em breve, compreende agregados compatíveis com o dicionário integrado EVS/WVS, compatível e pronto para fusão. Permitem trabalhar com séries temporais. Selecione a seção [de arquivos longitudinais WVS](#) para navegar na documentação. Existem livros de código integrados, arquivos integrados para o período 1981-2022 e muito mais.

Ao baixar arquivos e relatórios de resultados, você será solicitado a **se registrar** gratuitamente e concordar com a licença de uso de dados de não redistribuição da WVS. Todos os dados e documentação são gratuitos e foram produzidos pelo WVS Data Archive em colaboração com a [Fundaçao](#).

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

3- Outros dados de opinião pública –Latino Barómetro

<https://www.latinobarometro.org/latContents.jsp>

Início Corporación Datos Documentos Análisis Online Novedades English

 Latinobarómetro^{MR}
OPINIÓN PÚBLICA LATINOAMERICANA

Inicio /

Todas las descargas generan un fichero ZIP con los cuestionarios en español e inglés, y los datos en el formato seleccionado, en idioma español el inglés.

Año	SPSS	Stata	SAS	R rdata	R rds
Latinobarómetro 2020 NEW!	ZIP				
Latinobarómetro 2018	ZIP				
Latinobarómetro 2017	ZIP				
Latinobarómetro 2016	ZIP				
Latinobarómetro 2015	ZIP				
Latinobarómetro 2013	ZIP				
Latinobarómetro 2011	ZIP	ZIP	ZIP	ZIP	
Latinobarómetro 2010	ZIP	ZIP	ZIP	ZIP	
Latinobarómetro 2009	ZIP	ZIP	ZIP	ZIP	
Latinobarómetro 2008	ZIP	ZIP	ZIP	ZIP	

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

3- Outros dados de opinião pública –CESOP

https://www.cesop.unicamp.br/por/banco_de_dados

É necessário fazer um cadastro para conseguir baixar as bases de dados

The screenshot shows the CESOP website with a red header. The logo 'cesop centro de estudos de opinião pública' is on the left. On the right, there are links for 'cadastro' (registration), 'login', and 'english'. Below the header, there are navigation links: 'início', 'revista opinião pública', 'banco de dados' (which is highlighted in white), 'eseb', and 'pesquisa'. A dark red footer bar contains links for 'catálogo', 'apresentação', and 'equipe'. A small blue icon of a person with a speech bubble is in the bottom right corner.

Catálogo de pesquisas

The screenshot shows the search interface for the CESOP catalog. It has four main search fields: 'Palavras-chave' (with a button 'Escolher palavras-chave'), 'Local da pesquisa' (set to 'Brasil'), 'Data da pesquisa / início' (set to 'Janeiro 2010'), and 'Data da pesquisa / fim' (set to 'Janeiro 2023'). Below these are three dropdown fields: 'Instituto de pesquisa', 'Número da pesquisa', and 'Conteúdo do questionário'. A note below the 'Conteúdo do questionário' field says 'Busca textual nos questionários.' At the bottom is a large green search button with a white magnifying glass icon.

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

The screenshot shows a user profile page from the cesop website. At the top, there is a navigation bar with the cesop logo and links for 'início', 'revista opinião pública', 'banco de dados', 'eseb', and 'pesquisa'. Below the navigation bar is a banner featuring a photograph of people in what appears to be a library or study area. The main content area has a light gray background. A large, bold title 'Conta de Naiara Sandi de Almeida Alcantara' is centered above a horizontal line. To the left of the line, the text 'ACESSO AO BANCO DE DADOS' is displayed. Below this, a section titled 'Limite de acessos as pesquisas' provides information about monthly research access limits. It shows a 'LIMITE MENSAL' of '10 pesquisas' and 'UTILIZADAS' of '0 pesquisa'. A link 'Veja todas as pesquisas salvas' is also present. To the right of the line, a yellow sidebar contains the user's profile information: 'Seu cadastro', 'E-MAIL' (nayara_sandy@hotmail.com), 'NOME COMPLETO' (Naiara Sandi de Almeida Alcantara), 'INSTITUIÇÃO' (Universidade Federal do Paraná), and a link 'Alterar senha e dados pessoais'.

cesop centro de estudos de opinião pública

início revista opinião pública banco de dados eseb pesquisa

Conta de Naiara Sandi de Almeida Alcantara

ACESSO AO BANCO DE DADOS

Limite de acessos as pesquisas

LIMITE MENSAL	UTILIZADAS
10 pesquisas	0 pesquisa

[Veja todas as pesquisas salvas](#)

Seu cadastro

E-MAIL
nayara_sandy@hotmail.com

NOME COMPLETO
Naiara Sandi de Almeida Alcantara

INSTITUIÇÃO
Universidade Federal do Paraná

[Alterar senha e dados pessoais](#)

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

cesop centro de estudos de opinião pública

início revista opinião pública banco de dados eseb pesquisa

início publicações ondas

Sobre o ESEB

O Estudo Eleitoral Brasileiro (ESEB), *survey* nacional pós-eleitoral de cunho acadêmico, é realizado pelo CESOP desde 2002, sob a coordenação da Profa. Dra. Rachel Meneguello. O ESEB é um projeto vinculado ao *Comparative Study of Electoral Systems Project (CSES)*, coordenado pela Universidade de Michigan (www.cses.org) e que conta com a participação de dezenas de instituições de vários países. O projeto que embasa o CSES tem como premissa geral a ideia de que os contextos sócio-políticos e, em específico, os arranjos institucionais que regem as dinâmicas eleitorais, afetam a natureza e a qualidade da escolha democrática. O projeto busca identificar como variáveis contextuais, especialmente as instituições, moldam crenças e comportamentos dos cidadãos e, através de eleições, definem a capacidade ou qualidade do regime democrático. Busca ainda compreender a natureza dos alinhamentos e clivagens sociais e políticos, e como os cidadãos, vivendo sob distintos arranjos políticos, avaliam os processos políticos e as instituições democráticas.



BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

3- Outros dados de opinião pública – DATAFOLHA e outros institutos públicos e privados

https://www.cesop.unicamp.br/por/bando_de_dados

Informações divulgadas pelo data
folha, explicando como suas pesquisas
são realizadas

https://datafolha.folha.uol.com.br/divididas/pesquisas_eleitorais.shtml

2. Qual o número mínimo de entrevistas para uma pesquisa eleitoral?

As amostras nacionais do **Datafolha** têm entre 2.000 e 2.500 entrevistas, mas não há tamanho mínimo ou ideal para uma amostra eleitoral. O mais importante é a sua representatividade, ou seja, como são selecionados os entrevistados.

Palavras-chave	Local da pesquisa
Escolher palavras-chave	<input type="text"/> <input type="button" value=""/>
Instituto de pesquisa	Número da pesquisa
DATAFOLHA	
CESOP-CEMARX-UFG CESOP-FGV CESOP-IBOPE CESOP-IPSOS OPINION CESOP-NUPPS/USP CESOP-VOX POPULI CRITERIUM DATAFOLHA DATAUFF/UFG FPA-FUNDAÇÃO PERSEU ABRAMO IBOPE IDESP-INSTITUTO DE ESTUDOS ECONÔMICOS, SOCIAIS E POLÍTICOS DE SÃO PAULO INCT - Instituto da Democracia IPEC-INTELIGÊNCIA EM PESQUISA E CONSULTORIA VANDERBILT UNIVERSITY/UFG VOX POPULI	
<input type="button" value="Número da pesquisa"/>	
Número da pesquisa	Título

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

3- Outros dados de opinião pública –
FREEDOM HOUSE

<https://freedomhouse.org/countries/freedom-world/scores>

Be the first to know what's happening.

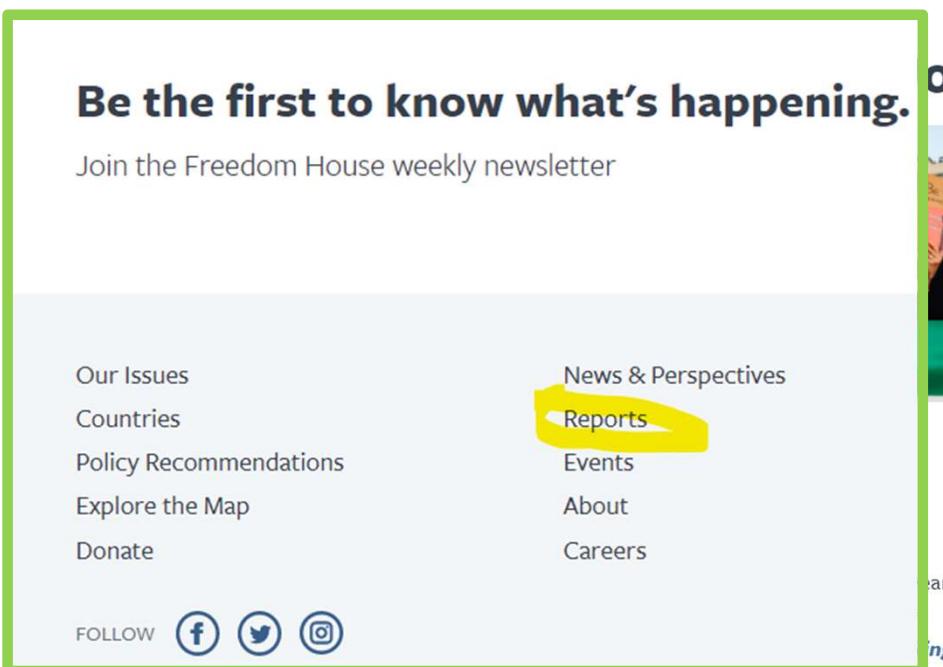
Join the Freedom House weekly newsletter

Our Issues Countries Policy Recommendations Explore the Map Donate

FOLLOW   

News & Perspectives Reports Events About Careers

50 Years in the Struggle for Democracy to learn more about the latest global trends.



About the Report

Published since 1973, *Freedom in the World* is Freedom House's flagship annual report, assessing the condition of political rights and civil liberties around the world.

- Research Methodology

Explore Country Data & Narratives

Visit our Countries & Territories page to explore the numerical ratings and supporting country narrative texts for the 195 countries and 15 territories assessed in the *Freedom in the World 2023* report.

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

3- Outros dados de opinião pública – FREEDOM HOUSE

A Freedom House classifica o acesso das pessoas aos direitos políticos e às liberdades civis em 210 países e territórios por meio de seu relatório anual *Freedom in the World*. As liberdades individuais – desde o direito de voto à liberdade de expressão e igualdade perante a lei – podem ser afetadas por atores estatais ou não estatais. Clique no nome de um país abaixo para acessar o relatório narrativo completo do país.

Visualizando:

[Pontuações globais de liberdade](#) [Pontuações de liberdade na Internet](#) [Pontuações de Democracia](#)

Busca rápida:

Viewing:

[Global Freedom Scores](#) [Internet Freedom Scores](#) [Democracy Scores](#)

Comparative and Historical Data Files:

- All Data, FIW 2013-2023 (Excel Download)
- Country and Territory Ratings and Statuses, 1973-2023 (Excel Download)
- Aggregate Category and Subcategory Scores, 2003-2023 (Excel Download)
- List of Electoral Democracies, FIW 2023 (Excel Download)

Country	Total Score and Status	Political Rights	Civil Liberties
Brazil	72 Free	30	42

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

3- Outros dados de opinião pública –
V=DEM

<https://www.v-dem.net/>



Lar [Sobre](#) ▾ [Nosso trabalho](#) ▾ [Conjuntos de dados](#) ▾ [gráficos](#) ▾ [Publicações](#) ▾ [Notícias](#)

[Projeto V-Dem](#) [Instituto V-Dem](#) [Centros Regionais](#) [Histórico V-Dem](#) [Financiadores](#) [Estágios](#) [Contato](#) [Perguntas frequentes](#)

O Projeto V-Dem

Variedades de Democracia (V-Dem) é uma abordagem única para conceituar e medir a democracia. Fornecemos um conjunto de dados multidimensional e desagregado que reflete a complexidade do conceito de democracia como um sistema de governo que vai além da simples presença de eleições. O projeto V-Dem distingue entre cinco princípios de democracia de alto nível: eleitoral, liberal, participativo, deliberativo e igualitário, e coleta dados para medir esses princípios.

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

The screenshot shows a portion of the V-Dem website. At the top, there is a navigation bar with links: Home, About, Our Work, Datasets, Graphing, Publications, and News. The 'Datasets' link is currently active, as indicated by a dropdown menu. This menu lists several datasets: The V-Dem Dataset, ERT Dataset, V-Party Dataset, Dataset Archive, and Reference Documents. Below this menu, the main content area features a large image of a hand reaching out over a green surface, with the text 'DEMOCRACY REPORT 2023' and 'in the Face of Autocratization'. To the right of the image, the text 'Report 2023' is displayed in large, bold letters, followed by 'DEMOCRACY in the Face of Autocratization'. A red button at the bottom of the image area contains the text 'Click here to Download the Democracy Report'.

<https://www.v-dem.net/data/the-v-dem-dataset/>

Para fazer o download é necessário responder um pequeno cadastro.

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

3- Outros dados de opinião pública –
Data bank

<https://databank.worldbank.org/>

The screenshot shows the homepage of DataBank, a service provided by The World Bank. The top navigation bar includes links for Home, About, Data, Search, Learn, News, Projects & Operations, Publications, Countries, Topics, and English. A feedback link 'Ajude-nos a melhorar esta seção do site. Podemos obter o seu feedback? Clique aqui' is also present. The main content area features a section titled 'Explorar. Criar. Compartilhe: Dados de Desenvolvimento' with a brief description of the tool's capabilities. On the left, a sidebar lists options like 'Início do banco de dados', 'bancos de dados', 'Criar relatório', 'Relatórios salvos', and 'Conjuntos de dados salvos'. On the right, a sidebar titled 'DataBank' lists 'DataBank Home', 'Databases' (which is highlighted with a yellow bar), 'Create Report', 'Saved Reports', 'Saved Datasets', and 'Metadata Glossary'.

BASES DE DADOS MUITO UTILIZADAS NA CIÊNCIAS SOCIAIS/CIÊNCIA POLÍTICA?

The screenshot shows the DataBank | World Development Indicators interface. The top navigation bar includes links for Help us improve this section of the site, Click here, Sign In, and various language options (English, Español, Français, عربى, 中文). The main content area is divided into two main sections: Preview and Visualização.

Preview Section: This section displays a message: "Please select variables from each of the following dimensions to view a report. You can select from left panel or by clicking the links above." It lists three dimensions: Country, Series, and Time. The Time dimension is currently selected, indicated by an underline. A sidebar on the right provides download options for the selected data, including Metadados, excel, CSV, and TXT com guias, with a link to Baixar formato CSV.

Visualização Section: This section shows a table of data for Brazil across five years (2017-2021) for six different indicators. The indicators include access to clean cooking fuels, rural access, urban access, electricity access, bank ownership, and adjusted savings. The table highlights data for bank ownership (85,62) and electricity access (99,80).

Left Panel (Variables): This panel lists various dimensions and their counts: Database (Available 85, Selected 1), Country (Available 266, Selected 0), Series (Available 1442, Selected 6), and Time (Available 5, Selected 1). It also includes a search bar for keywords and a list of selected variables at the bottom.

MATERIAIS COMPLEMENTARES

ENTOS > EVENTOS 2023 > CURSO DADOS 1SEM > MATERIAIS TEÓRICOS COMPLEMENTARES

Nome	Data de modificação	Tipo	Tamanho
Textos 2 Barbetta.pdf	21/03/2023 10:34	PDF X	7.839
Textos 3 Babbie.pdf	20/01/2017 23:12	PDF X	80.003
Textos 5a R_CientistasSociais.pdf	04/08/2017 15:18	PDF X	2.230
Textos 6 Triola_Estatística.pdf	23/06/2020 10:42	PDF X	17.904
Textos Complementar_Survey.pdf	13/05/2020 11:24	PDF X	10.991

TEXTOS PUBLICADOS PELO GRUPO COMUNICAÇÃO POLÍTICA E OPINIÃO PÚBLICA OU PELO
PROF. EMERSON CERVI

<https://cpop.ufpr.br/publicacoes-cpop/>

Métodos Quantitativos para Iniciantes

V.1

Publicações CPOP 2017

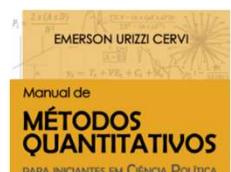


E-BOOK

Métodos Quantitativos para Iniciantes

V.2

Publicações CPOP 2019



E-BOOK



OBRIGADO

NAIARA ALCANTARA
NAYARA_SANDY@HOTMAIL.COM