

AQUISIÇÃO DE DADOS COM R

Davi Alves

Contato: davi.alves@uniriotec.br

ANTES DE COMEÇAR ATAQUE DE OPORTUNIDADE

- [Conheça o Grupo de Apoio Estatístico](#)
- Instagram: [@gae_assessoria](#)
- [Evento GAE Assessoria](#)
- E-mail: gaeassessoria@unirio.br

O QUE SERÁ APRESENTADO?

- Porque utilizar o R para aquisição de dados?
- Como organizar os dados
- Como adquirir dados utilizando o R?

PORQUE USAR O R PARA AQUISIÇÃO DE DADOS?

- Pacotes hospedados no CRAN para aquisição de dados oficiais
 - Permitem aquisição e processamento dos dados
- Diversos Pacotes para aquisição de dados de diferentes fontes
- Automação de processos de aquisição
- Vasta biblioteca de pacotes para processamento de dados
- Possibilidade de trabalhar diretamente com SGBDs

COMO ORGANIZAR OS DADOS

- Os dados normalmente são armazenados em uma matriz (ou quadro, *planilha*) onde cada coluna se refere a uma variável e cada linha a um elemento
- É comum a existência de uma **coluna de identificação** do elemento
- A **primeira linha** é geralmente utilizada para identificar as colunas (variáveis)

COMO ORGANIZAR OS DADOS

EXEMPLOS

nº do quest.	3(a) didat.	3(b) conhec.	3(c) bibl.	3(d) labor.	3(e) disc.	3(f) curric.	3(g) satisf.	4(a) posit.	4(b) negat.	5 desemp.
1	2	4	2	1	2	2	2	1	2	1,95
2	2	3	2	1	2	3	3	9	1	1,72
3	3	2	1	1	3	2	3	3	3	2,39
4	2	2	3	1	4	4	3	3	5	2,57
5	3	3	4	3	3	4	2	3	1	2,51
6	2	2	2	1	3	1	3	9	2	2,04
7	4	3	1	1	4	2	5	1	9	1,99
8	2	3	2	2	2	3	3	.	1	2,69
9	3	3	2	3	4	4	4	5	2	2,57
10	3	4	2	1	3	4	4	1	1	2,10
11	3	3	2	2	3	3	3	2	2	3,61
12	4	4	2	3	4	3	4	1	2	2,37
13	2	3	3	4	4	3	4	3	1	1,62
14	2	2	3	2	3	3	3	1	2	1,87
15	2	3	3	2	4	3	3	.	.	2,47
16	3	3	1	2	3	4	3	2	1	2,61
17	2	4	3	4	4	2	3	3	1	2,73
18	4	4	1	1	4	4	5	9	2	2,50
19	3	4	2	1	4	3	3	1	4	3,12
20	2	2	1	1	3	3	3	9	1	3,19

Barbetta (2010)

microcod	micronome	uf	reg	tx_dcv_f	tx_dcv_m	tx_homi_f
11001	Porto Velho	RO	N	4.13	5.20	0.87
11002	Guajará-Mirim	RO	N	2.35	5.15	0.87
11003	Ariquemes	RO	N	3.90	4.30	0.69
11004	Ji-Paraná	RO	N	3.45	4.03	0.19
11005	Alvorada D'Oeste	RO	N	4.95	4.81	0.29
11006	Cacoal	RO	N	3.75	5.45	0.39
11007	Vilhena	RO	N	3.27	4.54	0.51
11008	Colorado Oeste	RO	N	3.24	4.94	0.13
12001	Cruzeiro do Sul	AC	N	3.25	6.16	0.53
12002	Tarauacá	AC	N	4.04	3.07	0.09
12003	Sena Madureira	AC	N	2.21	3.62	0.13
12004	Rio Branco	AC	N	4.26	5.12	0.57

Alves (2017)

COMO ORGANIZAR OS DADOS

EXEMPLOS

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa

Anderson (1935)

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0

Henderson & Velleman (1981)

AQUISIÇÃO DE DADOS *HTML*

EXEMPLO “SIMPLES”

- Aquisição de dados de uma página com uma única tabela
- Motivação: Aquisição de dados sobre Acreditação Hospitalar ONA



AQUISIÇÃO DE DADOS *HTML*

EXEMPLO “SIMPLES”

MAPA DE ACREDITAÇÕES

TIPO

ESTADO

NOME

CERTIFICAÇÃO

CIDADE

TIPO DE GESTÃO

Buscar

AC

AM

RR

PA

AP

MA

PI

CE

RN

PB

PE

AL

SE

TO

MT

BA

GO

DF

MG

ES

MS

RO

Tipo

Total de certificações

Processamento de roupas para saúde

Serviço de Engenharia Clínica para a Saúde

Serviço de Higienização para a Saúde

Serviço de Manipulação e Atenção Farmacêutica para a Saúde

Serviço de Processamento de Produtos para a Saúde

Serviço de Processamento de Roupas para a Saúde

Serviço odontológico

MAPA DE ACREDITAÇÕES				
TIPO				
ESTADO				
NOME				
CERTIFICAÇÃO				
CIDADE				
TIPO DE GESTÃO				
Buscar				
1409 resultados encontrados.				
Nome	Cidade/UF	Certificação	Tipo	Gestão
CLINILAVES	Araquari/SC	Selo de Qualificação	Processamento de roupas para saúde	PRIVADO
Lavebrás MT Gestão de Texteis Ltda	Cuiabá/MT	Selo de Qualificação	Processamento de roupas para saúde	PRIVADO

AQUISIÇÃO DE DADOS *HTML*

EXEMPLO “SIMPLES”

- 1º Passo: instalar pacotes
 - `install.packages(“rvest”)`
 - `install.packages(“dplyr”)`
 - `install.packages(“stringr”)`
- 2º Passo: carregar pacotes
 - `library(“rvest”)`
 - `library(“dplyr”)`
 - `library(“stringr”)`

VAMOS AO



AQUISIÇÃO DE DADOS *HTML*

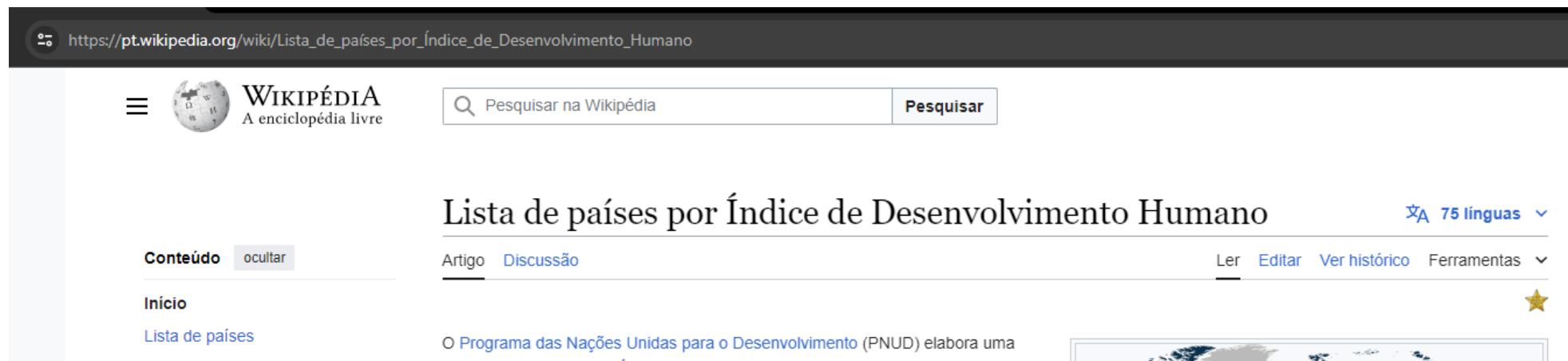
PÁGINA COM UMA ÚNICA TABELA ESTRUTURADA

Nome	Cidade/UF	Certificação	Tipo	Gestão	Validade
CLINILAVES	Araquari/SC	Selo de Qualificação	Processamento de roupas para saúde	PRIVADO	07/08/2025
Lavebrás MT Gestão de Textéis Ltda	Cuiabá/MT	Selo de Qualificação	Processamento de roupas para saúde	PRIVADO	09/05/2026
MEDLAV LAVANDERIA HOSPITALAR	Rafard/SP	Selo de Qualificação	Processamento de roupas para saúde	PRIVADO	09/10/2025
OPY Health - Metropolitano	Contagem/MG	Selo de Qualificação	Processamento de roupas para saúde	PRIVADO	27/01/2025
EQUIPACARE	Volta Redonda/RJ	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	04/03/2026
MVS ENGENHARIA CLINICA	Fortaleza/CE	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	15/02/2025
OPY HEALTH - DELPHINA AZIZ	Manaus/AM	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	11/12/2025
OPY Health - Metropolitano	Belo Horizonte/MG	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	26/01/2025
ORBIS ENGENHARIA CLÍNICA	Goiânia/GO	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	15/09/2024
ORBIS ENGENHARIA CLÍNICA	Aparecida de Goiânia/GO	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	15/09/2024
ORBIS ENGENHARIA HOSPITALAR	Goiânia/GO	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	15/09/2024
SEOH SERVIÇO DE ENGENHARIA OTIMIZADA HOSPITALAR	Goiânia/GO	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	30/06/2025
SulMed Engenharia Hospitalar	Pelotas/RS	Selo de Qualificação	Serviço de Engenharia Clínica para a Saúde	PRIVADO	19/12/2025
AST FACILITIES	Porto Alegre/RS	Selo de Qualificação	Serviço de Higienização para a Saúde	PRIVADO	18/03/2026
JB SERVICOS DE LIMPEZA	Campo Grande/MS	Selo de Qualificação	Serviço de Higienização para a Saúde	PRIVADO	14/08/2025
MANSERV FACILITIES	São Caetano do Sul/SP	Selo de Qualificação	Serviço de Higienização para a Saúde	PRIVADO	12/12/2024
OPY HEALTH - DELPHINA AZIZ	Manaus/AM	Selo de Qualificação	Serviço de Higienização para a Saúde	PRIVADO	11/12/2025
OPY Health - Metropolitano	Belo Horizonte/MG	Selo de Qualificação	Serviço de Higienização para a Saúde	PRIVADO	27/01/2025
QUALITECH TERCEIRIZAÇÃO LTDA	São Paulo/SP	Selo de Qualificação	Serviço de Higienização para a Saúde	PRIVADO	14/12/2024

AQUISIÇÃO DE DADOS *HTML*

EXEMPLO “COMPLEXO”

- Aquisição de dados de uma página com várias tabelas, que podem ou não estar estruturadas
- Motivação: Aquisição de dados da Wikipedia



AQUISIÇÃO DE DADOS *HTML*

EXEMPLO “COMPLEXO”

Conteúdo

ocultar

Início

Lista de países

IDH por regiões e grupos

✓ Países/territórios que não constaram no último relatório

Estados membros da ONU (não calculados pelo PNUD)

Estados que não são membros da ONU (últimos dados do PNUD)

Estados que não são membros da ONU (não calculado pelo PNUD)

Ver também

Notas

Referências

Ligações externas

Lista de países por Índice de Desenvolvimento Humano

75 línguas

Artigo Discussão

Ler Editar Ver histórico Ferramentas

★

O Programa das Nações Unidas para o Desenvolvimento (PNUD) elabora uma **lista de 193 países por Índice de Desenvolvimento Humano (IDH)** no **Relatório de Desenvolvimento Humano**. O índice considera **saúde, educação e renda** em um determinado país para fornecer uma medida de **desenvolvimento humano** que seja comparável entre países e ao longo do tempo^{[1][2]}

O IDH foi publicado pela primeira vez em 1990 com o objetivo de ser uma medida mais abrangente do desenvolvimento humano do que medidas puramente econômicas, como o **produto interno bruto (PIB)**. O índice incorpora três dimensões do desenvolvimento humano: uma vida longa e saudável, conhecimento e padrões de vida dignos. Vários indicadores são usados para quantificar o desempenho dos países em cada dimensão. Os indicadores utilizados no relatório de 2022 foram **expectativa de vida ao nascer**, anos esperados de **escolaridade** das crianças; média de anos de escolaridade dos adultos; e **renda nacional bruta** per capita. Os indicadores são usados para criar um índice de saúde, um índice de educação e um índice de renda, cada um com um valor entre 0 e 1. A **média geométrica** dos três índices – ou seja, a raiz cúbica do produto dos índices – é o Índice de Desenvolvimento Humano. Um valor acima de 0,800 é classificado como "muito alto", entre 0,700 e 0,799 como "alto", 0,550 a 0,699 como "médio" e abaixo de 0,550 como "baixo".^{[1][3][4]}

Os dados usados para calcular o IDH são provenientes principalmente de agências das **Nações Unidas** e instituições internacionais, como a **Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO)**, o **Departamento de Assuntos Econômicos e Sociais das Nações Unidas**, o **Banco Mundial**, o **Fundo Monetário Internacional** e a **Organização para a Cooperação e Desenvolvimento Econômico (OCDE)**. Raramente, quando falta um dos indicadores, são usados modelos de regressão entre países. Devido a atualizações de dados e metodologias aprimoradas, os valores do IDH não são comparáveis entre os relatórios de desenvolvimento humano; em vez disso, cada relatório recalcula o IDH de alguns anos anteriores.^{[4][5]}

O IDH é o indicador de desenvolvimento humano mais utilizado e mudou a forma



Mapa-múndi representando as quatro categorias do **Índice de Desenvolvimento Humano**, baseado no relatório publicado em 2024, com dados referentes a 2022.

0,800 – 1,000 (muito alto)	0,350 – 0,554 (baixo)
0,700 – 0,799 (alto)	Sem dados
0,555 – 0,699 (médio)	



Mapa-múndi representando faixas do **Índice de Desenvolvimento Humano** em incrementos de 0,05, baseado no relatório publicado em 2024, com dados referentes a 2022.

≥ 0,950	0,650 – 0,699	0,400 – 0,449
0,900 – 0,950	0,600 – 0,649	≤ 0,399
0,850 – 0,899	0,550 – 0,599	Sem dados
0,800 – 0,849	0,500 – 0,549	

AQUISIÇÃO DE DADOS *HTML*

EXEMPLO “COMPLEXO”

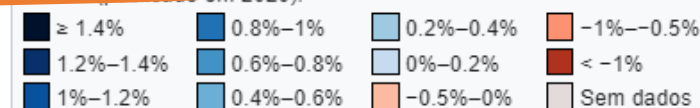
- A mudança de posição em relação a 2020:

- ▲ = aumento;
- = = estável;
- ▼ = perda;

- Os valores similares do IDH na lista atual não conduzem a relações classificatórias, já que a classificação do IDH é realmente determinada usando valores do IDH ao sexto ponto decimal.

- A revisão do índice foi divulgada em **13 de março de 2024**, com dados referentes ao ano de **2022**.^[1]

- Os números colocados entre parênteses representam a diferença entre as posições relativas aos dados de 2020.^[17]



Posição		País	Índice de Desenvolvimento Humano		
2022	Variação em relação a 2021		IDH 2022	IDH 2021	Variação em relação a 2021
IDH Muito Alto					
1	=	 Suíça	0,967	0,962	▲ 0,002
2	=	 Noruega	0,966	0,961	▲ 0,002
3	=	 Islândia	0,959	0,959	▲ 0,002
4	=	 Hong Kong	0,956	0,952	▼ 0,003
5	▲ 1	 Dinamarca	0,952	0,948	▲ 0,005
5	▲ 2	 Suécia	0,952	0,947	▲ 0,003
7	▲ 2	 Alemanha	0,950	0,942	▲ 0,002
7	▲ 1	 Irlanda	0,950	0,945	▲ 0,004
9	▲ 3	 Singapura	0,949	0,939	▲ 0,007

AQUISIÇÃO DE DADOS *HTML* EXEMPLO “COMPLEXO”

VAMOS AO



AQUISIÇÃO DE MICRODADOS DO SUS

- Utilizaremos o pacote [microdatasus](#)
- 1º Passo: instalar pacotes
 - `install.packages("devtools")`
 - `devtools::install_github("rfsaldanha/microdatasus")`
- 2º Passo: carregar pacotes
 - `library("microdatasus")`

VAMOS AO



AQUISIÇÃO DE MICRODADOS DO SUS

- A função `fetch_datasus()`:
 - Função para aquisição de microdados
 - Argumentos principais:
 - `year_start`: ano de início dos dados a serem adquiridos
 - `year_end`: ano de fim dos dados a serem adquiridos
 - `month_start`: mês de início dos dados a serem adquiridos
 - `month_end`: mês de fim dos dados a serem adquiridos
 - `information_system`: sistema de informação fonte dos dados
 - `uf`: siglas das unidades da federação sobre as quais os dados serão coletados (entre aspas e maiúsculo, podem ser concatenados mais de um estado)
 - `vars`: indica as variáveis que se deseja coletar
- Importante:
 - Lembrar de sempre alocar os dados adquiridos em um objeto:
 - `bd <- fetch_datasus(...)`

AQUISIÇÃO DE MICRODADOS DO SUS

SISTEMA	SUBDIVISÃO	SIGLA	COBERTURA	MENSAL
SIH	AIH reduzida	SIH-RD	1992 – atual	Sim
	AIH rejeitada	SIH-RJ	1992 – atual	Sim
	AIH serviços profissionais	SIH-SP	1992 – atual	Sim
	AIH rejeitadas com código de erro	SIH-ER	1992 – atual	Sim
SIM	Declarações de óbitos	SIM-DO	1979 – atual	Não
	Declarações de óbitos fetais	SIM-DOFET	1979 – atual	Não
	Declarações de óbitos por causas externas	SIM-DOEXT	1979 – atual	Não
	Declarações de óbitos infantis	SIM-DOINF	1979 – atual	Não
	Declarações de óbitos maternos	SIM-DOMAT	1996 – atual	Não
SINASC	Declarações de nascidos vivos	SINASC	1994 – atual	Não
CNES	Leitos	CNES-LT	Outubro/2005 – atual	Sim
	Estabelecimentos	CNES-ST	Agosto/2005 – atual	Sim
	Dados complementares	CNES-EQ	Agosto/2005 – atual	Sim
	Equipamentos	CNES-EQ	Agosto/2005 – atual	Sim
	Serviço especializado	CNES-SR	Agosto/2005 – atual	Sim
	Habilitação	CNES-HB	Março/2007 – atual	Sim
	Profissional	CNES-PF	Agosto/2005 – atual	Sim
	Equipes	CNES-EP	Abril/2007 – atual	Sim
	Regra contratual	CNES-RC	Março/2007 – atual	Sim
	Incentivos	CNES-IN	Novembro/ 2007 – atual	Sim
	Estabelecimentos de ensino	CNES-EE	Março/2007 – atual	Sim
	Estabelecimento filantrópico	CNES-EF	Março/2007 – atual	Sim
	Gestão e metas	CNES-GM	junho/2007 – atual	Sim
SIA	APAC de acompanhamento à cirurgia bariátrica	SIA-AB	janeiro/2008 – Março/2013	Sim
	APAC de acompanhamento pós-cirurgia bariátrica	SIA-ABO	Abril/2013 – atual	Sim
	APAC de confecção de fistula arteriovenosa	SIA-ACF	junho/2014 – atual	Sim
	APAC de laudos diversos	SIA-AD	janeiro/2008 – atual	Sim
	APAC de medicamentos	SIA-AM	janeiro/2008 – atual	Sim
	APAC de nefrologia	SIA-AN	janeiro/2008 – Outubro/2014	Sim
	APAC de quimioterapia	SIA-AQ	janeiro/2008 – atual	Sim
	APAC de radioterapia	SIA-AR	janeiro/2008 – atual	Sim
	APAC de tratamento dialítico	SIA-ATD	junho/2014 – atual	Sim
	Produção ambulatorial	SIA-PA	junho/1994 – atual	Sim
	Psicossocial	SIA-PS	janeiro/2013 – atual	Sim
	Atenção domiciliar	SIA-SAD	Novembro/2012 – atual	Sim

AIH: Autorização de Internação Hospitalar; APAC: Autorização de Procedimento de Alta Complexidade; CNES: Cadastro Nacional dos Estabelecimentos de Saúde; SIA: Sistema de Informações Ambulatoriais; SIH: Sistema de Internações Hospitalares; SIM: Sistema de Informações sobre Mortalidade; SINASC: Sistema de Informações sobre Nascidos Vivos.

Saldanha *et al.* (2017)

AQUISIÇÃO DE MICRODADOS DO SUS

- Pré-processamento dos dados:
 - Funções `process_sim`, `process_sih`, `process_sinasc`, `process_sia`, `process_cnes`, `process_sinan`
 - Prepara as variáveis do banco de acordo com o Sistema de Informação fonte de acordo com o dicionário de variáveis e arquivos auxiliares de tabulação
- Importante:
 - Na versão atual todas as variáveis são identificadas como caractere, logo, é necessário converter as variáveis numéricas
 - No caso da mortalidade, é necessária atenção especial ao processamento da idade

AQUISIÇÃO DE MICRODADOS DO SUS

- Alguns cuidados:
 - O tempo para aquisição de dados pode variar de acordo com a velocidade e estabilidade de sua conexão com a internet
 - Por se tratar de dados de eventos de saúde de todo o país, os dados podem grande volume
 - São milhões de interações por ano por exemplo!
 - O pré-processamento pode ser muito demorado quando o volume de dados é muito grande

AQUISIÇÃO DE DADOS DA PESQUISA NACIONAL DE SAÚDE

- A [PNS](#) é um inquérito populacional desenvolvido pelo Ministério da Saúde em parceria com o IBGE:
 - Existem dados dos levantamentos de 2013 e 2019
 - O inquérito levanta diversas informações acerca da saúde da população brasileira
 - Informações gerais são levantadas sobre o domicílio e todos os moradores
 - Um indivíduo é selecionado para responder a um questionário completo, com mais informações acerca de sua saúde

AQUISIÇÃO DE DADOS DA PESQUISA NACIONAL DE SAÚDE

- Utilizaremos o pacote [PNSIBGE](#)
- 1º Passo: instalar o pacote
 - `install.packages("PNSIBGE")`
- 2º Passo: carregar pacotes
 - `library("PNSIBGE")`

VAMOS AO



AQUISIÇÃO DE DADOS DA PESQUISA NACIONAL DE SAÚDE

- A função `get_pns()`:
 - Função para aquisição dos dados
 - Argumentos principais:
 - `year`: ano do levantamento cujo os dados serão adquiridos (2013 ou 2019)
 - `selected`: Se `TRUE` o questionário específico dos moradores selecionados serão usados, caso contrário questionário de residência e moradores
 - `anthropometry`: Se `TRUE` o questionário específico para o módulo de antropometria dos residentes selecionados será adquirido
 - `vars`: vetor com as variáveis de interesse
 - `labels`: Se `TRUE`, as categorias das variáveis serão apresentadas conforme o dicionário de variável
 - `design`: Se `TRUE`, o objeto baixado conterá as informações do desenho amostral para análise (*survey.design*)

AQUISIÇÃO DE DADOS DA PESQUISA NACIONAL DE SAÚDE

- Importante
 - É necessário indicar um objeto no qual os dados adquiridos devem ser alocados:
 - `pns <- get_pns(...)`
- Todas informações do plano amostral podem ser incorporadas no momento da aquisição dos dados

AQUISIÇÃO DE DADOS DA PNAD CONTÍNUA

- A PNAD CONTÍNUA foi implantada a partir de janeiro de 2012 em caráter definitivo e visa acompanhar as flutuações trimestrais e a evolução das informações necessárias para o estudo do desenvolvimento socioeconômico do País.
 - Amostra planejada para produzir resultados para o Brasil e diversos recortes regionais

AQUISIÇÃO DE DADOS DA PNAD CONTÍNUA

- Utilizaremos o pacote [PNADcIBGE](#)
- 1º Passo: instalar o pacote
 - `install.packages("PNADcIBGE")`
- 2º Passo: carregar pacotes
 - `library("PNADcIBGE")`

VAMOS AO



AQUISIÇÃO DE DADOS DA PNAD CONTÍNUA

- A função `get_pnadc()`:
 - Função para aquisição dos dados
 - Argumentos principais:
 - `year`: ano cujo os dados serão adquiridos (a partir de 2012)
 - `quarter`: trimestre do ano cujo o dado será adquirido (1 a 4)
 - `interview`: Número da entrevista do dado que será adquirido (1 a 5). Se não for informado, `quarter` ou `topic` deverá ser informado
 - `topic`: trimestre do tópico do dado que será adquirido (1 a 4). Se não for informado, `quarter` ou `interview` deverá ser informado
 - `selected`: Se `TRUE`, o questionário específico para morador selecionado será adquirido, caso contrário os dados adquiridos serão aqueles referentes ao questionário básico de moradia e residentes
 - `vars`: variáveis a serem adquiridas
 - `labels`: Se `TRUE`, as categorias das variáveis serão apresentadas conforme o dicionário de variável
 - `design`: Se `TRUE`, o objeto baixado conterá as informações do desenho amostral para análise (*survey.design*)

AQUISIÇÃO DE DADOS DA PNAD COVID19

- A PNAD COVID19 foi uma pesquisa desenvolvida durante o ano o ano de 2020 com objetivo de avaliar sintomas grupais autorreferidos e os impactos da pandemia no mercado de trabalho
 - Cerca de 48 mil domicílios foram entrevistados semanalmente por telefone totalizando uma amostra fixa de cerca de 193 mil domicílios
 - O questionário abordava questões relacionadas à saúde, trabalho e renda entrevistados

AQUISIÇÃO DE DADOS DA PNAD COVID19

- Utilizaremos o pacote [COVIDIBGE](#)
- 1º Passo: instalar o pacote
 - `install.packages("COVIDIBGE")`
- 2º Passo: carregar pacotes
 - `library("COVIDIBGE")`

VAMOS AO



AQUISIÇÃO DE DADOS DA PNAD COVID19

- A função `get_covid()`:
 - Função para aquisição dos dados
 - Argumentos principais:
 - `year`: ano cujo os dados serão adquiridos (2020)
 - `month`: mês cujo os dados se referem (5 a 11)
 - `vars`: vetor com as variáveis de interesse
 - `labels`: Se `TRUE`, as categorias das variáveis serão apresentadas conforme o dicionário de variável
 - `design`: Se `TRUE`, o objeto baixado conterá as informações do desenho amostral para análise (*survey.design*)

AQUISIÇÃO DE DADOS DO IPEADATA

- O [IPEADATA](#) é uma base de dados gratuita mantida pelo Ipea. Oferece acesso a dados macroeconômicos, financeiros e regionais do Brasil, disponibilizando um catálogo informações acerca dos dados
- Existem dados de diversas naturezas no IPEADATA acerca de diferentes tópicos

AQUISIÇÃO DE DADOS DO IPEADATA

- Utilizaremos o pacote [ipeadatar](#)
- 1º Passo: instalar o pacote
 - `install.packages("ipeadatar")`
- 2º Passo: carregar pacotes
 - `library("ipeadatar")`

VAMOS AO



AQUISIÇÃO DE DADOS DO IPEADATA

- As principais funções:
 - `available_countries`: Lista os países disponíveis na base de dados
 - `available_series`: Lista as séries de dados disponíveis
 - `available_subjects`: Lista os temas abordados pelas séries de dados
 - `available_territories`: Lista as divisões territoriais disponíveis (estados, regiões, etc.)
 - `metadata`: Retorna informações detalhadas (metadados) sobre a série solicitada
 - `search_series`: Busca séries de dados por palavra-chave
 - `ipeadata`: Retorna os dados da série solicitada

AQUISIÇÃO DE DADOS DO IPEADATA

- Função `search_series()`
 - Permite identificar séries históricas de interesse:
 - Argumentos principais:
 - `terms`: permite buscar por termos específicos em campos específicos
 - `fields`: identifica o campo específico no qual o termo será buscado
- Função `metadata()`
 - Permite visualizar os metadados da série
 - A partir destas informações é possível identificar o código da série que deverá ser utilizado para aquisição dos dados de interesse

REFERÊNCIAS

- BARBETTA, Pedro A. **Estatística aplicada às ciências sociais**. 8. ed. Florianópolis: Editora da UFSC, 2010.
- ANDERSON, Edgar. The irises of the Gaspé Peninsula. **Bulletin of the American Iris Society**, v. 59, p. 2-5, 1935.
- HENDERSON, H. V.; VELLEMAN, P. F. Building multiple regression models interactively. **Biometrics**, v. 37, p. 391-411, 1981.
- ASSUNÇÃO, Guilherme; HIDALGO, Lorena. **COVIDIBGE: Downloading, Reading and Analyzing PNAD COVID19 Microdata**. Versão 0.2.2. 2024. Disponível em: <https://CRAN.R-project.org/package=COVIDIBGE>.
- ASSUNÇÃO, Guilherme; HIDALGO, Lorena. **PNSIBGE: Downloading, Reading and Analyzing PNS Microdata**. Versão 0.2.1. 2024. Disponível em: <https://CRAN.R-project.org/package=PNSIBGE>.
- BRAGA, Danielle; ASSUNÇÃO, Guilherme. **PNADcIBGE: Downloading, Reading and Analyzing PNADC Microdata**. Versão 0.7.5. 2024. Disponível em: <https://CRAN.R-project.org/package=PNADcIBGE>.
- GOMES, Luiz Paulo. **ipeadatar: API Wrapper for 'Ipeadata'**. Versão 0.1.6. 2022. Disponível em: <https://CRAN.R-project.org/package=ipeadatar>.
- SALDANHA, Ricardo F. **microdatasus: Download and preprocess DataSUS files**. 2023. Disponível em: <https://github.com/rfsaldanha/microdatasus>, <https://rfsaldanha.github.io/microdatasus/>.
- WICKHAM, Hadley. **rvest: Easily Harvest (Scrape) Web Pages**. Versão 1.0.4. 2024. Disponível em: <https://CRAN.R-project.org/package=rvest>.
- WICKHAM, Hadley et al. **dplyr: A Grammar of Data Manipulation**. Versão 1.1.3. 2023. Disponível em: <https://CRAN.R-project.org/package=dplyr>.
- WICKHAM, Hadley. **stringr: Simple, Consistent Wrappers for Common String Operations**. Versão 1.5.0. 2022. Disponível em: <https://CRAN.R-project.org/package=stringr>.