microdadosBrasil

leitura fácil e rápida dos microdados brasileiros no R

Lucas F. Mation Nicolas S. Pinto

novembro de 2016

Agenda

- microdadosBrasil: overview

Agenda

- microdadosBrasil: overview
- Princípios

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso
  - Separação: código x metadados

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso
  - Separação: código x metadados
- microdadosBrasil: Estrutura

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso
  - Separação: código x metadados
- microdadosBrasil: Estrutura
  - CSV com Metadados

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso
  - Separação: código x metadados
- microdadosBrasil: Estrutura
  - CSV com Metadados
  - Função principal: `read_data`

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso
  - Separação: código x metadados
- microdadosBrasil: Estrutura
  - CSV com Metadados
  - Função principal: `read_data`
  - Wrapper functions

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso
  - Separação: código x metadados
- microdadosBrasil: Estrutura
  - CSV com Metadados
  - Função principal: `read_data`
  - Wrapper functions
  - Funções auxiliares

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso
  - Separação: código x metadados
- microdadosBrasil: Estrutura
  - CSV com Metadados
  - Função principal: `read_data`
  - Wrapper functions
  - Funções auxiliares
- Comparação com outros pacotes

Agenda

- microdadosBrasil: overview
- Princípios
  - Facilidade de uso
  - Separação: código x metadados
- microdadosBrasil: Estrutura
  - CSV com Metadados
  - Função principal: `read_data`
  - Wrapper functions
  - Funções auxiliares
- Comparação com outros pacotes
- Próximos passos

Dificuldades para importar microdados

Nome	Data de modificaç...	Tipo
micro_censo_edu_superior1995	19/07/2016 09:02	Pasta de arquivos
micro_censo_edu_superior1996	19/07/2016 09:02	Pasta de arquivos
micro_censo_edu_superior1997	19/07/2016 09:02	Pasta de arquivos
micro_censo_edu_superior1998	19/07/2016 09:02	Pasta de arquivos
micro_censo_edu_superior1999	19/07	
micro_censo_edu_superior2000	19/07	
micro_censo_edu_superior2001	19/07	
micro_censo_edu_superior2002	19/07	
micro_censo_edu_superior2003	19/07	
micro_censo_edu_superior2008	19/07	
micro_censo_edu_superior2010	19/07	
microdados_censo_educacao_superior_2...	19/07	
microdados_censo_educacao_superior_2...	19/07	
microdados_censo_superior_2009	19/07	
microdados_educacao_superior_2006	19/07	
microdados_educacao_superior_2007	19/07	
micro		
micro		
micro		

```
10  FILE HANDLE GRADUACAO_PRES NAME='D:\DADOS\GRADUACAO_PRESENCIAL.TXT'
11  /MODE=CHARACTER/LRECL=3400.
12  DATA LIST FIXED
13  FILE=GRADUACAO_PRES
14  /MASCARA                1-8 (A)
15  ANO                  9-16
16  CURSO                17-24
17  NOMEDOCURSO          25-224 (A)
18  CODMUNICCURSO            225-236 (A)
19  DTINIFUNCCURSO           237-244
20  NIVELCURSO           245-254 (A)
21  SUBNIVEL                 255-264 (A)
22  MOD_PRESENC          265-265 (A)
23  MOD_DISTANCIA            266-266 (A)
    EH_BACHARELADO       267-267 (A)
    EH_LICENCPLENA           268-268 (A)
    EH_LICCURTA          269-269 (A)
    EH_TECNO                 270-270 (A)
    EH_ESPECPROF             271-271 (A)
    AREACURSO            272-281 (A)
    NOMEAREACURSO        282-381 (A)
    CODAREAGERAL             382-391 (A)
    NOMEAREAGERAL        392-491 (A)
    CODAREAESPECIFICA        492-501 (A)
    NOMEAREAESPECIFICA       502-601 (A)
    CODAREADETALHADA         602-611 (A)
    NOMEAREADETALHADA        612-711 (A)
```

PNAD 2011 - Microdados

Dificuldades para importar microdados

- Encontrar versão oficial

Dificuldades para importar microdados

- Encontrar versão oficial
- Txt colunado(fixed width files)

Dificuldades para importar microdados

- Encontrar versão oficial
- Txt colunado(fixed width files)
  - Planilha com dicionário

Dificuldades para importar microdados

- Encontrar versão oficial
- Txt colunado(fixed width files)
  - Planilha com dicionário
  - Programa de leitura: SAS e SPSS

Dificuldades para importar microdados

- Encontrar versão oficial
- Txt colunado(fixed width files)
  - Planilha com dicionário
  - Programa de leitura: SAS e SPSS
- Falta de harmonização entre os anos

Dificuldades para importar microdados

- Encontrar versão oficial
- Txt colunado(fixed width files)
  - Planilha com dicionário
  - Programa de leitura: SAS e SPSS
- Falta de harmonização entre os anos
  - Nomes de arquivo

Dificuldades para importar microdados

- Encontrar versão oficial
- Txt colunado(fixed width files)
  - Planilha com dicionário
  - Programa de leitura: SAS e SPSS
- Falta de harmonização entre os anos
  - Nomes de arquivo
  - Nomes de variável

Dificuldades para importar microdados

- Encontrar versão oficial
- Txt colunado(fixed width files)
  - Planilha com dicionário
  - Programa de leitura: SAS e SPSS
- Falta de harmonização entre os anos
  - Nomes de arquivo
  - Nomes de variável
  - Ex: INSTITUICAO_SUP_97.txt, ies_superior_98.txt

Dificuldades para importar microdados

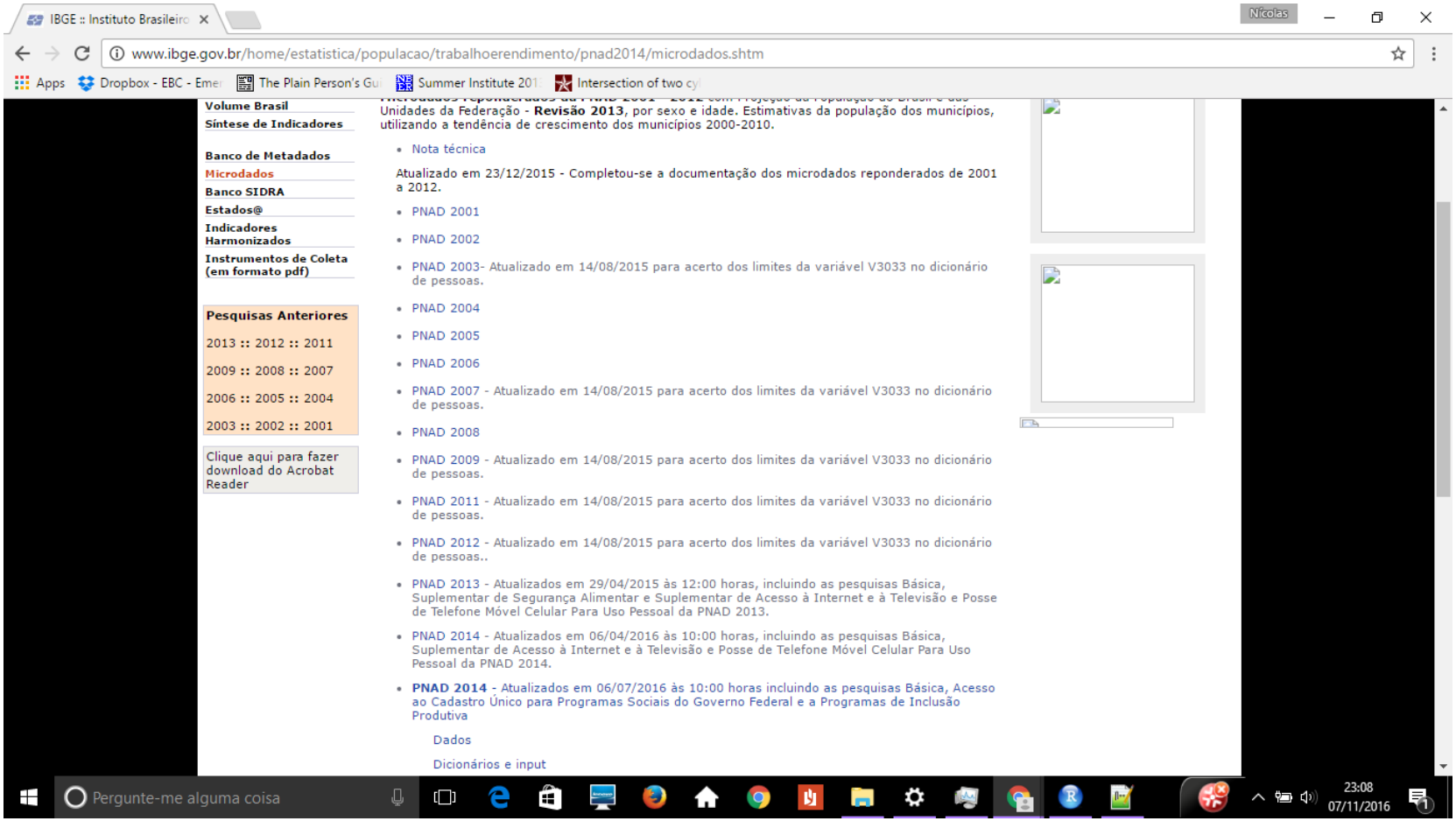
- Encontrar versão oficial
- Txt colunado(fixed width files)
  - Planilha com dicionário
  - Programa de leitura: SAS e SPSS
- Falta de harmonização entre os anos
  - Nomes de arquivo
  - Nomes de variável
  - Ex: INSTITUICAO_SUP_97.txt, ies_superior_98.txt
- Dados grandes subdivididos em muitos arquivos

Dificuldades para importar microdados

- Encontrar versão oficial
- Txt colunado(fixed width files)
  - Planilha com dicionário
  - Programa de leitura: SAS e SPSS
- Falta de harmonização entre os anos
  - Nomes de arquivo
  - Nomes de variável
  - Ex: INSTITUICAO_SUP_97.txt, ies_superior_98.txt
- Dados grandes subdivididos em muitos arquivos
  - Ex. RAIS: AC2014.txt,AL2014.txt, AM2014.txt,...

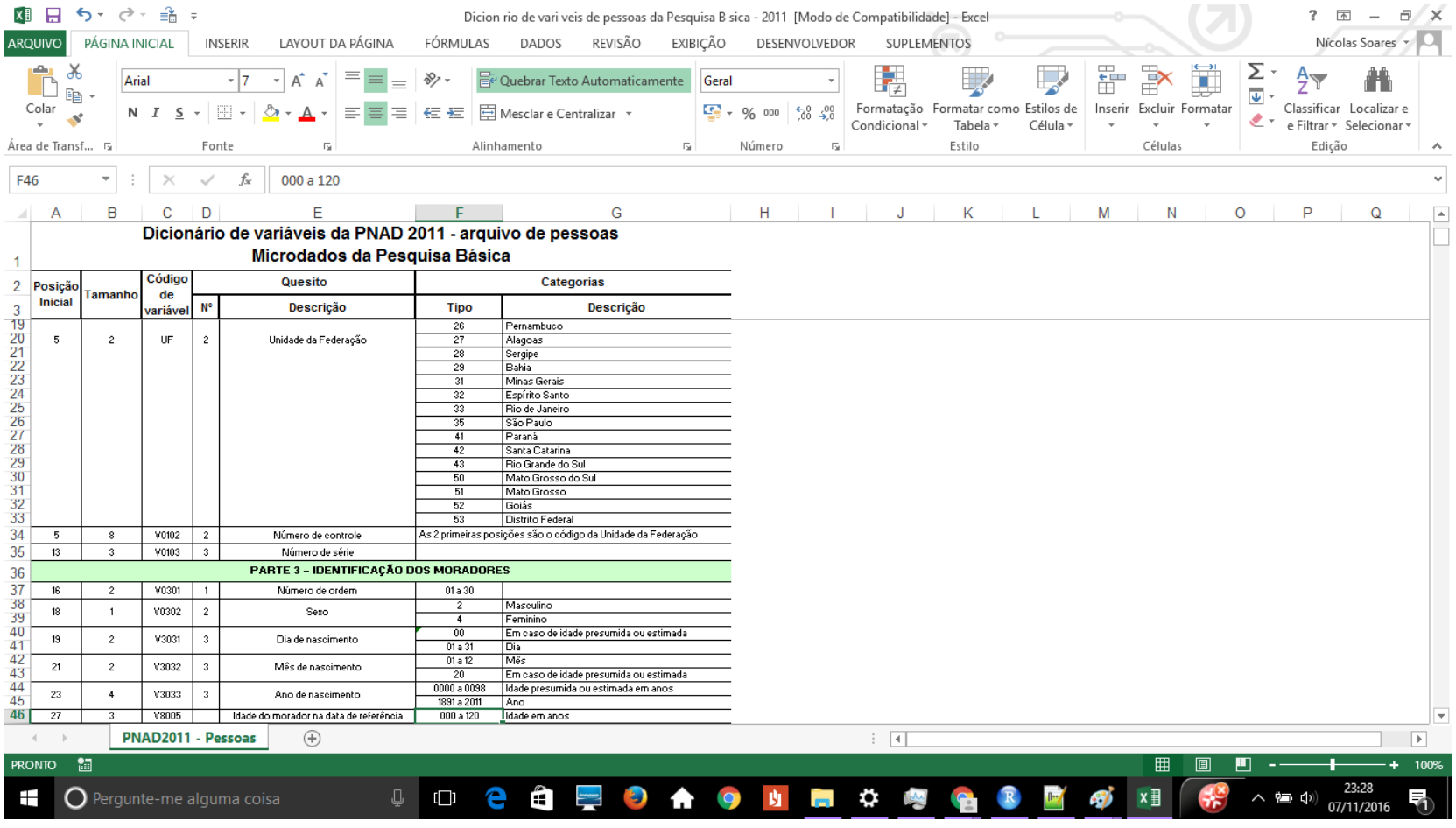
Método tradicional de importação de microdados

1. Baixar os dados do site oficial



Método tradicional de importação de microdados

2. Encontrar entre os arquivos as larguras de importação



Dicion rio de vari veis de pessoas da Pesquisa B sica - 2011 [Modo de Compatibilidade] - Excel

F46 | 000 a 120

Dicionário de variáveis da PNAD 2011 - arquivo de pessoas
Microdados da Pesquisa Básica

Posição Inicial	Tamanho	Código de variável	Quesito		Categorias	
			Nº	Descrição	Tipo	Descrição
5	2	UF	2	Unidade da Federação	26	Pernambuco
					27	Alagoas
					28	Sergipe
					29	Bahia
					31	Minas Gerais
					32	Espírito Santo
					33	Rio de Janeiro
					35	São Paulo
					41	Paraná
					42	Santa Catarina
					43	Rio Grande do Sul
					50	Mato Grosso do Sul
					51	Mato Grosso
					52	Goiás
					53	Distrito Federal
5	8	V0102	2	Número de controle	As 2 primeiras posições são o código da Unidade da Federação	
13	3	V0103	3	Número de série		
PARTE 3 – IDENTIFICAÇÃO DOS MORADORES						
16	2	V0301	1	Número de ordem	01 a 30	
18	1	V0302	2	Sexo	2	Masculino
					4	Feminino
19	2	V3031	3	Dia de nascimento	00	Em caso de idade presumida ou estimada
					01 a 31	Dia
21	2	V3032	3	Mês de nascimento	01 a 12	Mês
					20	Em caso de idade presumida ou estimada
23	4	V3033	3	Ano de nascimento	0000 a 0098	Idade presumida ou estimada em anos
					1891 a 2011	Ano
27	3	V8005		Idade do morador na data de referência	000 a 120	Idade em anos

PNAD2011 - Pessoas

PRONTO

Pergunte-me alguma coisa

23:28
07/11/2016

Método tradicional de importação de microdados

2. Encontrar entre os arquivos as larguras de importação

```
vars_width <-

  c(4,8,3,2,1,2,2,4,3,1,1,1,1,1,1,2,1,1,1,1,1,1,1,1,1,2,1,1
1,1,2,2,1,1,1,1,1,1,1,1,1,1,1,1,2,1,1,1,1,2,1,1,1,1,1,1,1,1
1,1,1,4,5,1,4,5,1,1,12,1,12,1,1,2,1,2,1,1,1,1,1,1,4,5,2,1,1
1,1,11,7,1,11,7,1,11,7,1,1,1,1,1,11,7,1,11,7,1,11,7,1,1,1,1
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,12,1,12,1,1
1,1,1,2,1,1,2,2,1,1,2,1,1,1,1,1,1,4,5,2,1,1,1,1,1,1,1,1,1,1,1
2,2,1,1,1,2,4,5,1,1,1,1,1,1,1,12,1,12,1,1,1,2,1,12,1,12,1,1
1,1,2,2,4,5,1,1,1,1,1,1,1,1,1,1,2,1,1,1,2,12,2,12,2,12,2,12
2,12,2,12,1,1,2,2,2,2,1,1,2,2,1,1,1,2,4,1,1,2,2,1,1,1,2,2,2
,1,1,2,2,1,1,1,1,2,2,2,12,12,12,12,12,2,2,1,1,5,5,1,1,1,2,1
,1,12,8)
```

Método tradicional de importação de microdados

3. Encontrar os nomes de pastas e arquivos

```
data_path<- paste0("C:/Users/Nícolas/Documents/",
                   "PNAD_reponderado_2011_20150814/2011/Da
```

Método tradicional de importação de microdados

4. Encontrar os nomes das variáveis

```
vars<- c("V0101","V0102","V0103","V0301","V0302","V3031","V
"V0402","V0403","V0404","V0405","V0406","V0407","V0408","V0
"V4112","V4011","v0412","V0501","V0502","V5030","V0504","V0
"V5063","V5064","V5065","V0507","V5080","V5090","V0510","V0
"V5123","V5124","V5125","V5126","V0601","V0602","V6002","V6
"V0604","V0605","V0606","V6007","V6070","V0608","V0609","V0
"V06112" "V0612","V0701","V0702","V0703","V0704","V0705","V
"V7090","V7100","V0711","V7121","V7122","V7124","V7125","V7
"V0714","V0715","V0716","V9001","V9002","V9003","V9004","V9
"V9008","V9009","V9010","V9011","V9012","V9013","V9014","V9
"V9156","V9157","V9159","V9161","V9162","V9164","V9016","V9
 "V9201","V9202","V9204","V9206","V9207","V9209","V9211","V
 "V9022","V9023","V9024","V9025","V9026","V9027","V9028","V
 "V9032","V9033","V9034","V9035","V9036","V9037","V9038","V
 "V9042","V9043","V9044","V9045","V9046","V9047","V9048","V
 "V9052","V9531","V9532","V9534","V9535","V9537","V90531",
```

Método tradicional de importação de microdados

5. Importar no R

```
d<- read.fwf(file = paste0(data_path, "PES2011.txt"),
                                widths = vars_width,
                                 col.names = vars)
```

Dificuldades para importar microdados

- E se quisermos importar dados de um período diferente?

Dificuldades para importar microdados

- E se quisermos importar dados de um período diferente?
- Seria necessário começar tudo de novo

microdadosBrasil: overview

- A importação de microdados deveria ser fácil e simples:

microdadosBrasil: overview

- A importação de microdados deveria ser fácil e simples:
  - função de download

microdadosBrasil: overview

- A importação de microdados deveria ser fácil e simples:
  - função de download
  - função de leitura

funções de leitura

Fonte	Função	Período	Nível
IBGE	read_PNAD	2001 a 2014	domicilios, pessoas
IBGE	read_CENSO	2000 a 2010	domicilios, pessoas
IBGE	read_PME	2002.01 a 2015.12	vinculos
IBGE	read_POF	2008	several, see details
INEP	read_CensoEscolar	1995 a 2015	escolas, . . . , see details
MTE	read_CAGED	2009.01 a 2016.05	vinculos
MTE	read_RAIS	1998 a 2014	estabelecimentos,vincu

funções de leitura

```
read_*(
  ft = filetype (subbase),
  i = periodo,
  root_path = pasta raiz (opcional),
  file = "C:/.../base.txt" # Se essa opção for preenchida,

)
```

Instalação

```
# github.com/lucasmation/microdadosBrasil

# README_PT.md : documentação em português

# Instalação:

devtools::install_github("lucasmation/microdadosBrasil")
```

PNAD

```
download_sourceData(dataset = "PNAD",
                    i = 2011, unzip = T)

d<- read_PNAD(ft = "pessoas", i = 2011)
d<- read_PNAD(ft = "domicilios", i = 2011)
```

PNAD

```
download_sourceData(dataset = "PNADcontinua",
                    i = "2012.01", unzip = T)

d<- read_PNADcontinua(ft = "pessoas", i = "2012.1")
```

CENSO DEMOGRÁFICO

```
download_sourceData("CENSO", 2000, unzip = T)
d <- read_CENSO('domicilios',2000)
d <- read_CENSO('pessoas',2000)
```

PME

```r
#It will download files for all months and the selected ye
download_sourceData("PME", i = "2012.01")
#'Period' argument should with quotes and formated as "YYY
d <- read_PME("vinculos", "2012.01")
```

Censo Escolar

```
download_sourceData(dataset = "CensoEscolar",
                    i = 2007, unzip = T)

d<- read_CensoEscolar(ft = "matricula", i = 2007)
d<- read_CensoEscolar(ft = "docente", i = 2007)
d<- read_CensoEscolar(ft = "turma", i = 2007)
d<- read_CensoEscolar(ft = "escola", i = 2007)
```

RAIS

```
download_sourceData("RAIS", i = "2000")

# Todas as UFs
d<- read_RAIS('vinculos', i = 2000)

# Apenas algumas UFs
d<- read_RAIS('vinculos', i = 2000, UF = c("DF","GO"))
```

Separação: código x metadados

Hadley Wickham, mai/2016, Managing many models with R

Vanilla cupcakes

The hummingbird bakery cookbook

1 cup flour
a scant ¾ cup sugar
1 ½ t baking powder
3 T unsalted butter
½ cup whole milk
1 egg
¼ t pure vanilla extract

Preheat oven to 350°F.

Put the flour, sugar, baking powder, salt, and butter in a freestanding electric mixer with a paddle attachment and beat on slow speed until you get a sandy consistency and everything is combined.

Whisk the milk, egg, and vanilla together in a pitcher, then slowly pour about half into the flour mixture, beat to combine, and turn the mixer up to high speed to get rid of any lumps.

Turn the mixer down to a slower speed and slowly pour in the remaining milk mixture. Continue mixing for a couple of more minutes until the batter is smooth but do not overmix.

Spoon the batter into paper cases until 2/3 full and bake in the preheated oven for 20-25 minutes, or until the cake bounces back when touched.

Separação: código x metadados

Vanilla cupcakes

The hummingbird bakery cookbook

1 cup flour
a scant ¾ cup sugar
1 ½ t baking powder
3 T unsalted butter
½ cup whole milk
1 egg
¼ t pure vanilla extract

Preheat oven to 350°F.

Put the flour, sugar, baking powder, salt, and butter in a freestanding electric mixer with a paddle attachment and beat on slow speed until you get a sandy consistency and everything is combined.

Whisk the milk, egg, and vanilla together in a pitcher, then slowly pour about half into the flour mixture, beat to combine, and turn the mixer up to high speed to get rid of any lumps.

Turn the mixer down to a slower speed and slowly pour in the remaining milk mixture. Continue mixing for a couple of more minutes until the batter is smooth but do not overmix.

Spoon the batter into paper cases until 2/3 full and bake in the preheated oven for 20-25 minutes, or until the cake bounces back when touched.

Separação: código x metadados

Chocolate cupcakes

The hummingbird bakery cookbook

¾ cup + 2T flour
2 ½ T cocoa powder
a scant ¾ cup sugar
1 ½ t baking powder
3 T unsalted butter
½ cup whole milk
1 egg
¼ t pure vanilla extract

Preheat oven to 350°F.

Put the flour, cocoa, sugar, baking powder, salt, and butter in a freestanding electric mixer with a paddle attachment and beat on slow speed until you get a sandy consistency and everything is combined.

Whisk the milk, egg, and vanilla together in a pitcher, then slowly pour about half into the flour mixture, beat to combine, and turn the mixer up to high speed to get rid of any lumps.

Turn the mixer down to a slower speed and slowly pour in the remaining milk mixture. Continue mixing for a couple of more minutes until the batter is smooth but do not overmix.

Spoon the batter into paper cases until 2/3 full and bake in the preheated oven for 20-25 minutes, or until the cake bounces back when touched.

Separação: código x metadados

Chocolate cupcakes

The hummingbird bakery cookbook

¾ cup + 2T flour
2 ½ T cocoa powder
a scant ¾ cup sugar
1 ½ t baking powder
3 T unsalted butter
½ cup whole milk
1 egg
¼ t pure vanilla extract

Preheat oven to 350°F.

Put the flour, cocoa, sugar, baking powder, salt, and butter in a freestanding electric mixer with a paddle attachment and beat on slow speed until you get a sandy consistency and everything is combined.

Whisk the milk, egg, and vanilla together in a pitcher, then slowly pour about half into the flour mixture, beat to combine, and turn the mixer up to high speed to get rid of any lumps.

Turn the mixer down to a slower speed and slowly pour in the remaining milk mixture. Continue mixing for a couple of more minutes until the batter is smooth but do not overmix.

Spoon the batter into paper cases until 2/3 full and bake in the preheated oven for 20-25 minutes, or until the cake bounces back when touched.

Separação: código x metadados

Vanilla cupcakes

The hummingbird bakery cookbook

120g flour
140g sugar
1.5 t baking powder
40g unsalted butter
120ml milk
1 egg
0.25 t pure vanilla extract

Preheat oven to 170°C.

Put the flour, sugar, baking powder, salt, and butter in a freestanding electric mixer with a paddle attachment and beat on slow speed until you get a sandy consistency and everything is combined.

Whisk the milk, egg, and vanilla together in a pitcher, then slowly pour about half into the flour mixture, beat to combine, and turn the mixer up to high speed to get rid of any lumps.

Turn the mixer down to a slower speed and slowly pour in the remaining milk mixture. Continue mixing for a couple of more minutes until the batter is smooth but do not overmix.

Spoon the batter into paper cases until 2/3 full and bake in the preheated oven for 20-25 minutes, or until the cake bounces back when touched.

1. Convert units

Separação: código x metadados

Vanilla cupcakes

The hummingbird bakery cookbook

120g flour
140g sugar
1.5 t baking powder
40g butter
120ml milk
1 egg
0.25 t vanilla

Beat flour, sugar, baking powder, salt, and butter until sandy.

Whisk milk, egg, and vanilla. Mix half into flour mixture until smooth (use high speed). Beat in remaining half. Mix until smooth.

Bake 20-25 min at 170°C.

2. Rely on domain knowledge

Separaçãoo: código x metadados

Vanilla cupcakes

The hummingbird bakery cookbook

120g flour
140g sugar
1.5 t baking powder
40g butter
120ml milk
1 egg
0.25 t vanilla

Beat dry ingredients + butter until sandy.

Whisk together wet ingredients. Mix half into dry until smooth (use high speed). Beat in remaining half. Mix until smooth.

Bake 20-25 min at 170°C.

3. Use variables

Separação: código x metadados

Cupcakes

	Vanilla	Chocolate
Beat dry ingredients + butter until sandy.	120g flour	100g flour
		20g cocoa
Whisk together wet ingredients. Mix half into dry until smooth (use high speed). Beat in remaining half. Mix until smooth.	140g sugar	140g sugar
	1.5t baking powder	1.5t baking powder
	40g butter	40g butter
	120ml milk	120ml milk
	1 egg	1 egg
Bake 20-25 min at 170°C.	0.25 t vanilla	0.25 t vanilla

4. Extract out common code

Separação: código x metadados

Cupcakes

	Flour	Baking powder	Sugar	Butter	Egg	Extra
Vanilla	120	1.5	140	40	1	0.25t vanilla
Chocolate	100	1.5	140	40	1	20g cocoa • 0.25t vanilla
Lemon	120	1.5	140	40	1	2T lemon zest
Red velvet	150	0	150	60	1	10g cocoa • 20ml red colouring • 1.5t vinegar • 0.5 t baking soda

4. Convert to data

CSV com Metadados

- Exemplo da estrutura dos metadados - Censo Escolar

CSV com Metadados

- Exemplo da estrutura dos metadados - Censo Escolar
  - **period**: 2000

CSV com Metadados

- Exemplo da estrutura dos metadados - Censo Escolar
  - **period**: 2000
  - **format**: fwf

CSV com Metadados

- Exemplo da estrutura dos metadados - Censo Escolar
  - **period**: 2000
  - **format**: fwf
  - **download_path**: `http://download.inep.gov.br/...`/micro_censo_escolar2000.zip

CSV com Metadados

- Exemplo da estrutura dos metadados - Censo Escolar
  - **period**: 2000
  - **format**: fwf
  - **download_path**: `http://download.inep.gov.br/...`/micro_censo_escolar2000.zip
  - **data_folder** : DADOS

CSV com Metadados

- Exemplo da estrutura dos metadados - Censo Escolar
  - **period**: 2000
  - **format**: fwf
  - **download_path**: `http://download.inep.gov.br/...`/micro_censo_escolar2000.zip
  - **data_folder** : DADOS
  - **input_folder** : INPUTS_SAS_SPSS

CSV com Metadados

- Exemplo da estrutura dos metadados - Censo Escolar
  - **period**: 2000
  - **format**: fwf
  - **download_path**: `http://download.inep.gov.br/...`/micro_censo_escolar2000.zip
  - **data_folder** : DADOS
  - **input_folder** : INPUTS_SAS_SPSS
  - **missing_symbols** : “{ñ c&{ñ”

CSV com Metadados

- Exemplo da estrutura dos metadados - Censo Escolar
  - **period**: 2000
  - **format**: fwf
  - **download_path**: `http://download.inep.gov.br/...`/micro_censo_escolar2000.zip
  - **data_folder** : DADOS
  - **input_folder** : INPUTS_SAS_SPSS
  - **missing_symbols** : “{ñ c&{ñ”
  - **ft_escola** : INPUT_SAS_MEDPROF.sas&DADOS_MEDPROF.TXT

CSV com Metadados

period	format	download_path	download_mode	path	inputs_folder	data_fold	missing_s	ft_escola	ft_escola_em_e_emprof
1995	fwf	http://download.inep.g	source	micro_censo_escolar1995	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	INPUT_SAS_CURSO.SAS&DADOS_CURSO.txt
1996	fwf	http://download.inep.g	source	micro_censo_escolar1996	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSO96.txt	NA
1997	fwf	http://download.inep.g	source	micro_censo_escolar1997	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	INPUT_SAS_EM12.SAS&DADOS_EM12.TXT
1998	fwf	http://download.inep.g	source	micro_censo_escolar1998	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	INPUT_SAS_EM12.SAS&DADOS_EM12.TXT
1999	fwf	http://download.inep.g	source	micro_censo_escolar1999	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	NA
2000	fwf	http://download.inep.g	source	micro_censo_escolar2000	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	INPUT_SAS_MEDPROF.sas&DADOS_MEDPROF.TX
2001	fwf	http://download.inep.g	source	micro_censo_escolar2001	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	INPUT_SAS_MEDPROF.sas&DADOS_MEDPROF.TX
2002	fwf	http://download.inep.g	source	micro_censo_escolar2002	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	INPUT_SAS_MEDPROF.sas&DADOS_MEDPROF.TX
2003	fwf	http://download.inep.g	source	micro_censo_escolar2003/M	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	INPUT_SAS_MEDPROF.sas&DADOS_MEDPROF.TX
2004	fwf	http://download.inep.g	source	microdados_censo_escolar_	INPUTS_SAS_SPSS	DADOS	NA	INPUT_SAS_CENSOESC.sas&DADOS_CENSOESC.txt	NA

Dicionários:

- Armazenamento em listas padronizadas

```
str(PNAD_dics[1:3] ,max.level = 2)
```

```
## List of 3
##  $ 2001:List of 2
##   ..$ dic_domicilios_2001:'data.frame':  59 obs. of  9 v
##   ..$ dic_pessoas_2001   :'data.frame':  363 obs. of  9
##  $ 2002:List of 2
##   ..$ dic_domicilios_2002:'data.frame':  63 obs. of  9 v
##   ..$ dic_pessoas_2002   :'data.frame':  323 obs. of  9
##  $ 2003:List of 2
##   ..$ dic_domicilios_2003:'data.frame':  63 obs. of  9 v
##   ..$ dic_pessoas_2003   :'data.frame':  421 obs. of  9
```

Dicionários:

- Armazenamento em listas padronizadas
  - dataset_dics.rda -> ano -> dic_filetype_ano

```
str(PNAD_dics[1:3] ,max.level = 2)
```

```
## List of 3
##  $ 2001:List of 2
##   ..$ dic_domicilios_2001:'data.frame':  59 obs. of  9 v
##   ..$ dic_pessoas_2001   :'data.frame':  363 obs. of  9
##  $ 2002:List of 2
##   ..$ dic_domicilios_2002:'data.frame':  63 obs. of  9 v
##   ..$ dic_pessoas_2002   :'data.frame':  323 obs. of  9
##  $ 2003:List of 2
##   ..$ dic_domicilios_2003:'data.frame':  63 obs. of  9 v
##   ..$ dic_pessoas_2003   :'data.frame':  421 obs. of  9
```

Dicionários:

int_pos	var_name	x
1	MASCARA	8.
9	ANO	5.
14	CODMUNIC	$CHAR12.

length	decimal_places	fin_pos	col_type	CHAR
8	0	8	i	FALSE
5	0	13	i	FALSE
12	0	25	c	TRUE

Função principal: `read_data`

```
read_data(dataset, ft, i,
          metadata = NULL,
          dic_list = NULL,
          var_translator = NULL,
          root_path = NULL,
          file = NULL
          )
```

- Função de importação genérica

Função principal: `read_data`

- Parâmetros:

Função principal: `read_data`

- Parâmetros:
  - *dataset*: Referência básica para a função de importação, procurará o dicionário e os metadados com base nesses nomes.

Função principal: `read_data`

- Parâmetros:
  - *dataset*: Referência básica para a função de importação, procurará o dicionário e os metadados com base nesses nomes.
  - Este parâmetro deve ser suficiente para identificar todos os componentes auxiliares necessários para a leitura

Função principal: `read_data`

- Parâmetros:
  - *dataset*: Referência básica para a função de importação, procurará o dicionário e os metadados com base nesses nomes.
  - Este parâmetro deve ser suficiente para identificar todos os componentes auxiliares necessários para a leitura
    - Arquivo com metadatados: inst/extadata/**dataset**_files_harmonization.csv

Função principal: `read_data`

- Parâmetros:
  - *dataset*: Referência básica para a função de importação, procurará o dicionário e os metadados com base nesses nomes.
  - Este parâmetro deve ser suficiente para identificar todos os componentes auxiliares necessários para a leitura
    - Arquivo com metadatados: inst/extadata/**dataset_**files_harmonization.csv
    - Arquivos com dicionarios: data/**dataset_**dics.rda

Função principal: `read_data`

- Parâmetros:

Função principal: `read_data`

- Parâmetros:
  - *ft*: Tipo de arquivo que será utilizado(Ex: “matrículas” dentro do Censo Escolar)

Função principal: `read_data`

- Parâmetros:
  - *ft*: Tipo de arquivo que será utilizado(Ex: "matrículas" dentro do Censo Escolar)
  - *i*: Período do arquivo que será utilizado(Ex: 2012)

Função principal: `read_data`

- Parâmetros:
  - *ft*: Tipo de arquivo que será utilizado(Ex: "matrículas" dentro do Censo Escolar)
  - *i*: Período do arquivo que será utilizado(Ex: 2012)
  - *root_path*: local dos arquivos, se não for fornecido, procurará no Working Directory

Função principal: `read_data`

- Parâmetros:
  - *ft*: Tipo de arquivo que será utilizado(Ex: "matrículas" dentro do Censo Escolar)
  - *i*: Período do arquivo que será utilizado(Ex: 2012)
  - *root_path*: local dos arquivos, se não for fornecido, procurará no Working Directory
  - *metadata*: `data.frame` contendo as informações necessárias para a leitura da base de dados, se = `NULL` é utilizada a tabela encontrada pelo nome padrão do arquivo `inst/extadata/dataset_files_harmonization.csv`

Função principal: `read_data`

- Parâmetros:
  - *ft*: Tipo de arquivo que será utilizado(Ex: "matrículas" dentro do Censo Escolar)
  - *i*: Período do arquivo que será utilizado(Ex: 2012)
  - *root_path*: local dos arquivos, se não for fornecido, procurará no Working Directory
  - *metadata*: `data.frame` contendo as informações necessárias para a leitura da base de dados, se = `NULL` é utilizada a tabela encontrada pelo nome padrão do arquivo `inst/extadata/dataset_files_harmonization.csv`
  - *file*: se preenchido, ignora os metadatados e o `root_path` lê direto o arquivo indicado,

Wrapper functions

- Traduzir a função `read_data` em argumentos mais amigáveis ao usuário.

```
read_PNAD<- function(ft,i,root_path=NULL,file = NULL){

  data<-read_data(dataset = "PNAD",
                  ft, i,
                  root_path =  root_path,
                  file = file)

}
```

Wrapper functions

- Traduzir a função `read_data` em argumentos mais amigáveis ao usuário.
- Lidar com exceções

```
read_PNAD<- function(ft,i,root_path=NULL,file = NULL){

  data<-read_data(dataset = "PNAD",
                  ft, i,
                  root_path =  root_path,
                  file = file)

}
```

Wrapper functions

- Traduzir a função `read_data` em argumentos mais amigáveis ao usuário.
- Lidar com exceções
- O identificador *dataset* é suficiente para encontrar metadados para uma dada pesquisa.

```
read_PNAD<- function(ft,i,root_path=NULL,file = NULL){

  data<-read_data(dataset = "PNAD",
                  ft, i,
                  root_path =  root_path,
                  file = file)

}
```

Funções auxiliares

```
download_sourceData(dataset = "PNAD",
                    i   = 2012,
                    unzip = T)
```

Funções auxiliares

```
parses_SAS_import_dic(file = "DICIONARIO.SAS")
```

- Método para traduzir dicionários SAS

Funções auxiliares

```
parses_SAS_import_dic(file = "DICIONARIO.SAS")
```

- Método para traduzir dicionários SAS
- Facilidade de replicação

Funções auxiliares

```
get_import_dictionary("CensoEscolar", 2000, 'escola')
```

```
##      int_pos var_name         x label length decimal_place
## 1          1  MASCARA        8.              8
## 2          9      ANO        5.              5
## 3         14 CODMUNIC $CHAR12.             12
## 4         26       UF $CHAR50.             50
## 5         76    SIGLA  $CHAR2.              2
## 6         78    MUNIC $CHAR50.             50
## 7        128      DEP $CHAR10.             10
## 8        138      LOC $CHAR10.             10
## 9        148  CODFUNC $CHAR11.             11
## 10       159 NIVELCRE  $CHAR1.              1
## 11       160 NIVELPRE  $CHAR1.              1
## 12       161 NIVELALF  $CHAR1.              1
## 13       162 NIV_F1A4  $CHAR1.              1
## 14       163 NIV_F5A8  $CHAR1.              1
## 15       164 NIVELMED  $CHAR1.              1
```

Utilização: adaptando novas bases para o microdadosBrasil

1. Criação de um arquivo "NomeDaBase_files_metadata_harmonzation.csv"

Utilização: adaptando novas bases para o microdadosBrasil

1. Criação de um arquivo "NomeDaBase_files_metadata_harmonzation.csv"
2. Criação de uma lista "NomeDaBase_dics.rda"

Utilização: adaptando novas bases para o microdadosBrasil

1. Criação de um arquivo "NomeDaBase_files_metadata_harmonzation.csv"
2. Criação de uma lista "NomeDaBase_dics.rda"
3. Criação de uma função "read_NomeDaBase()"

Ambiente de produção do pacote

- Produzido utilizando controle de versão

Ambiente de produção do pacote

- Produzido utilizando controle de versão
  - GIT

Ambiente de produção do pacote

- Produzido utilizando controle de versão
  - GIT
- Hospedado no GitHub

Ambiente de produção do pacote

- Produzido utilizando controle de versão
  - GIT
- Hospedado no GitHub
  - `https://github.com/lucasmation/microdadosBrasil`

Ambiente de produção do pacote

- Produzido utilizando controle de versão
  - GIT
- Hospedado no GitHub
  - `https://github.com/lucasmation/microdadosBrasil`
  - *software livre* : qualquer um pode colaborar

Ambiente de produção do pacote

- Produzido utilizando controle de versão
  - GIT
- Hospedado no GitHub
  - `https://github.com/lucasmation/microdadosBrasil`
  - *software livre* : qualquer um pode colaborar
- Utilização interna de pacotes de leitura modernos

Ambiente de produção do pacote

- Produzido utilizando controle de versão
  - GIT
- Hospedado no GitHub
  - `https://github.com/lucasmation/microdadosBrasil`
  - *software livre* : qualquer um pode colaborar
- Utilização interna de pacotes de leitura modernos
  - `data.table` para leitura de arquivo com delimitador

Ambiente de produção do pacote

- Produzido utilizando controle de versão
  - GIT
- Hospedado no GitHub
  - `https://github.com/lucasmation/microdadosBrasil`
  - *software livre* : qualquer um pode colaborar
- Utilização interna de pacotes de leitura modernos
  - `data.table` para leitura de arquivo com delimitador
  - `readr` para leitura de *fixed width files*

Comparação com outros pacotes

	dicIBGE.	datazoom	adsfree	microd
Software	R	Stata	R	R
Download				X
Dicionários de Importação	X	X	X	X
Função de importação		X	X	X
Simplicidade e ergonomicidade		X		X
Bases IBGE		X	X	X
Outras bases INEP - MTE				X
Dados $>$ RAM			X	
Desenho Amostral Complexo			X	
Interface gráfica		X		
Harmonização de variáveis		X		

Comparação com outros pacotes

.	Função	DataZoom	microdadosBrasil
Linhas	programa	4828	692
Linhas	metadado	NA	191
Caracteres	programa	257601	17305
Caracteres	metadado	NA	43404

Próximos passos

- Harmonização de nomes de variáveis

Próximos passos

- Harmonização de nomes de variáveis
- DADOS $>$ memória RAM: MonetDBLite

Próximos passos

- Harmonização de nomes de variáveis
- DADOS $>$ memória RAM: MonetDBLite
- Desenho amostral complexo(PNAD, PNAD contínua)

Próximos passos

- Harmonização de nomes de variáveis
- DADOS $>$ memória RAM: MonetDBLite
- Desenho amostral complexo(PNAD, PNAD contínua)
- Inclusão de mais bases

Próximos passos

- Harmonização de nomes de variáveis
- DADOS $>$ memória RAM: MonetDBLite
- Desenho amostral complexo(PNAD, PNAD contínua)
- Inclusão de mais bases
  - Provas do INEP

Próximos passos

- Harmonização de nomes de variáveis
- DADOS $>$ memória RAM: MonetDBLite
- Desenho amostral complexo(PNAD, PNAD contínua)
- Inclusão de mais bases
  - Provas do INEP
  - IBGE: PNS, ...

Próximos passos

- Harmonização de nomes de variáveis
- DADOS $>$ memória RAM: MonetDBLite
- Desenho amostral complexo(PNAD, PNAD contínua)
- Inclusão de mais bases
  - Provas do INEP
  - IBGE: PNS, ...
  - Versões mais antigas: PNAD desde 1976, Censo desde 1970.

Próximos passos

- Harmonização de nomes de variáveis
- DADOS $>$ memória RAM: MonetDBLite
- Desenho amostral complexo(PNAD, PNAD contínua)
- Inclusão de mais bases
  - Provas do INEP
  - IBGE: PNS, ...
  - Versões mais antigas: PNAD desde 1976, Censo desde 1970.
  - Sugestões...