

Um guia para a montagem e identificação dos painéis da pnadc

Webinar - 01/12/2021

Rafael Guerreiro Osorio

Instituto de Pesquisa Econômica Aplicada – Ipea

Diretoria de Estudos e Políticas Sociais - Disoc

Centro Internacional de Políticas para o Crescimento Inclusivo – IPC-IG

www.ipea.gov.br www.ipcig.org

Aplicativo **pynad** (Texto para Discussão “no forno”):

- escrito em **python** - não requer programação
- *pip install pynad*
- depende de como o IBGE divulga as bases da pnadc (pastas no servidor, nomenclatura de arquivos, estrutura dos dicionários de dados)
- cria/atualiza arquivos **zip** contendo os arquivos originais e derivados
- baixa os arquivos originais e os atualiza conforme a necessidade
- organiza os metadados dos dicionários em arquivos **json**
- converte os microdados para **csv**
- monta os painéis de domicílios
- identifica os grupos domésticos e seus indivíduos
- monta os painéis de indivíduos
- BUG ao redefinir seleções de cópias parciais

Pynad - Pesquisa Nacional por Amostra de Domicílios Contínua

Arquivo Sincronizar Ferramentas Ajuda

Arquivos originais da Pnadc em 26/11/2021

	Status na cópia local
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> anual	
▼ <input checked="" type="checkbox"/> <input type="checkbox"/> trimestral	
▼ <input checked="" type="checkbox"/> <input type="checkbox"/> 2012	
<input checked="" type="checkbox"/> <input type="checkbox"/> PNADC_012012_20210610.zip	sincronizado
<input checked="" type="checkbox"/> <input type="checkbox"/> PNADC_022012_20210610.zip	sincronizado
<input checked="" type="checkbox"/> <input type="checkbox"/> PNADC_032012_20210610.zip	sincronizado
<input checked="" type="checkbox"/> <input type="checkbox"/> PNADC_042012_20210610.zip	s
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2013	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2014	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2015	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2016	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2017	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2018	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2019	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2020	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2021	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> atualizacoes_divulgacao	

pnadc_novembro_2011.zip
Última atualização: 26/11/2021

Pynad - Pesquisa Nacional por Amostra de Domicílios Contínua

Arquivo Sincronizar Ferramentas Ajuda

Confirmar

Mapear os arquivos da Pnadc em **ftp.ibge.gov.br** pode demorar alguns minutos

Evite verificações desnecessárias

2 dias desde a última verificação

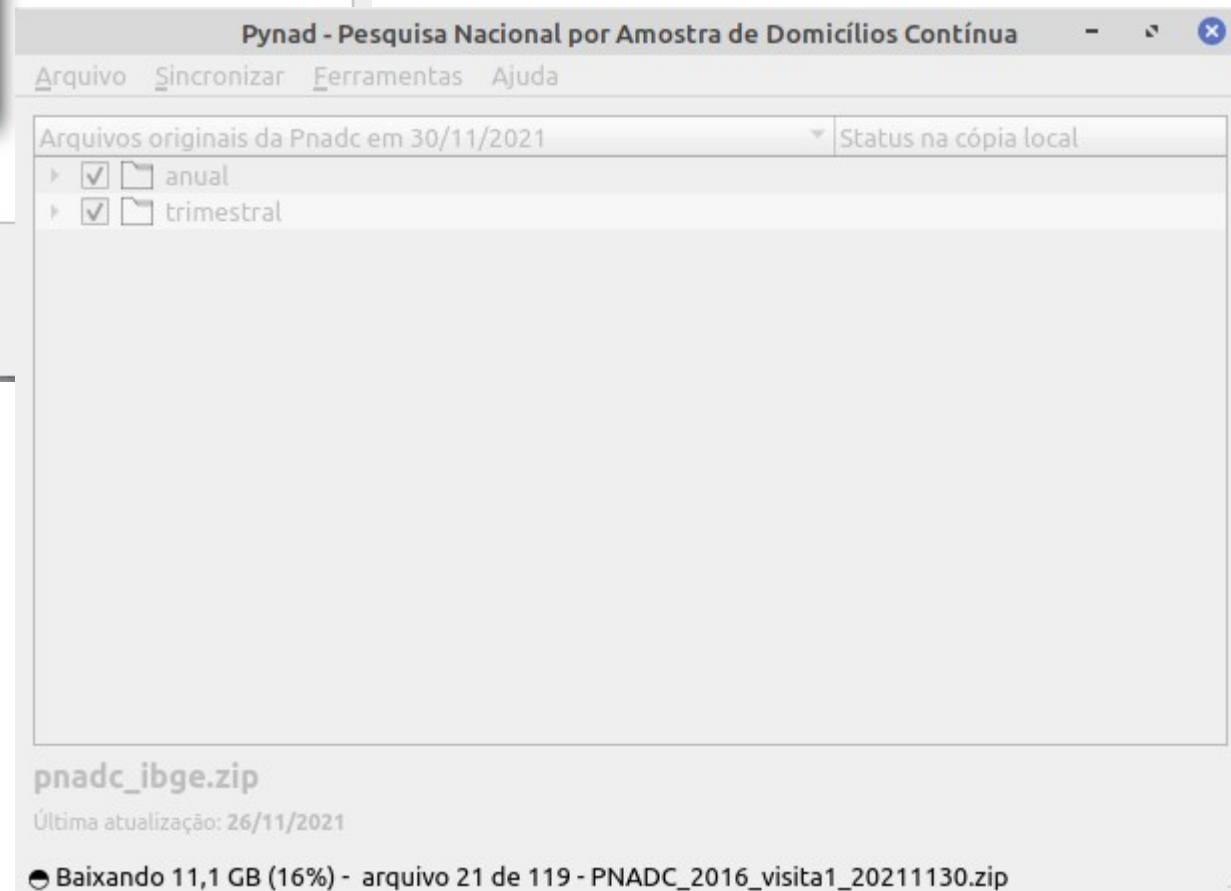
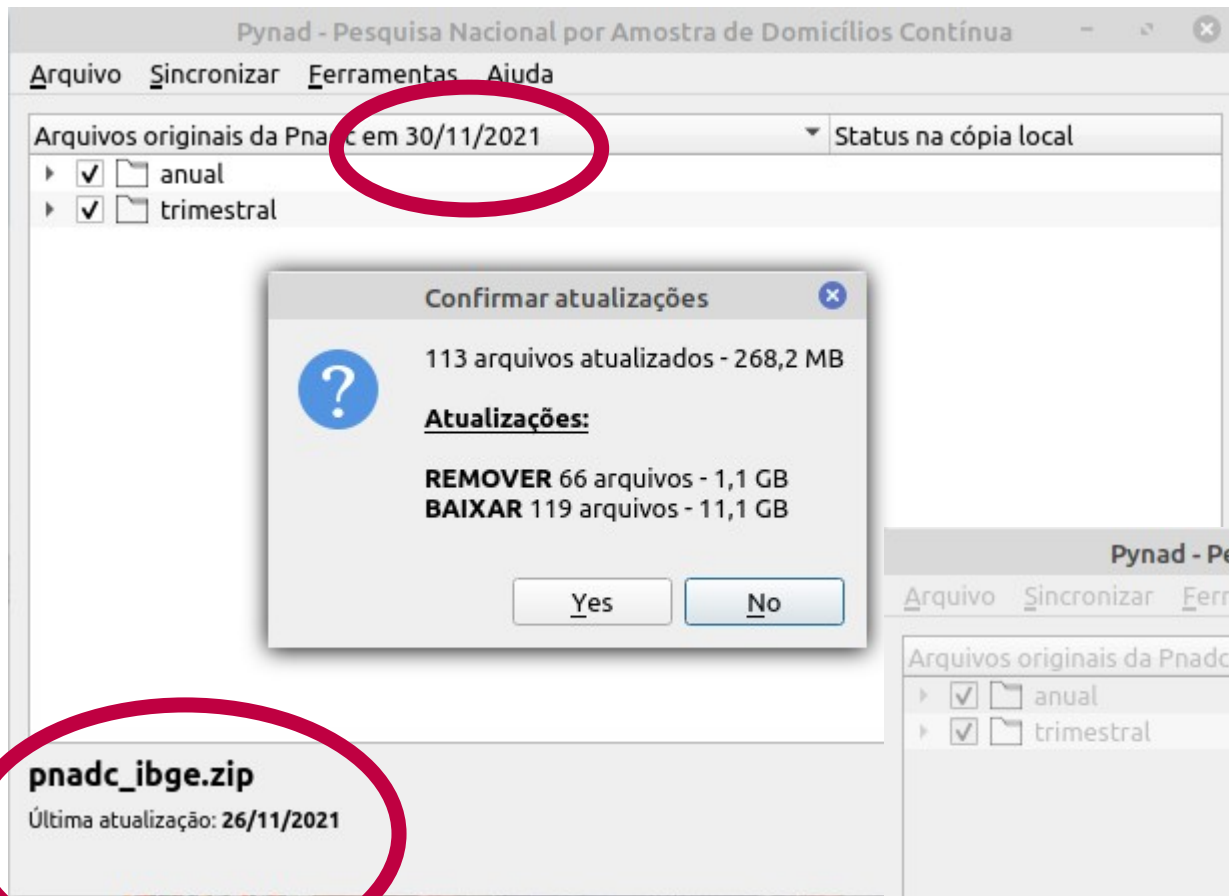
Verificar agora?

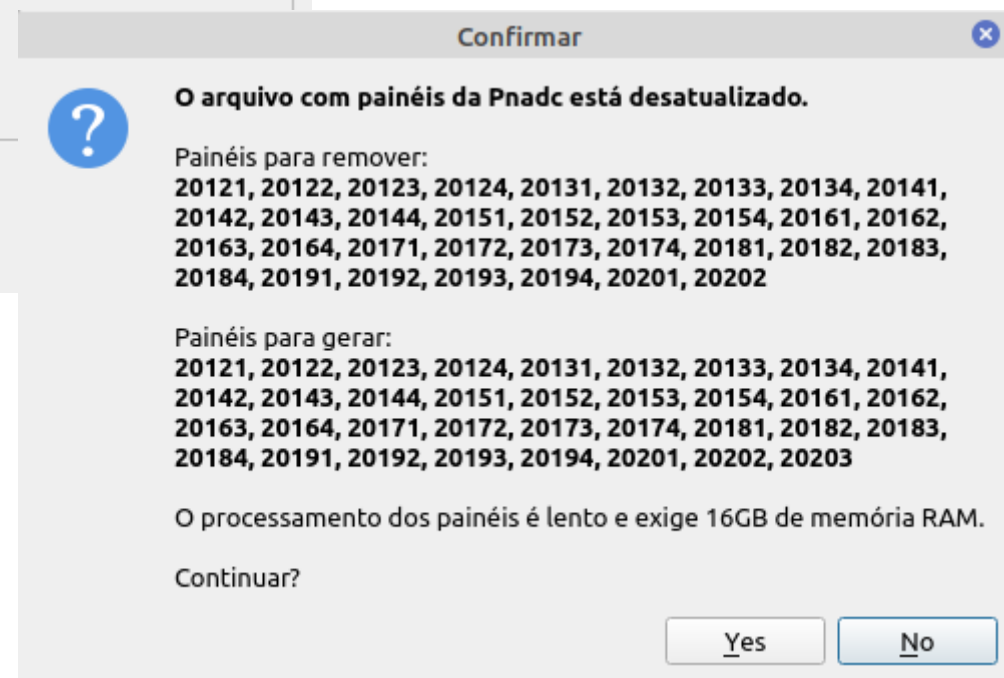
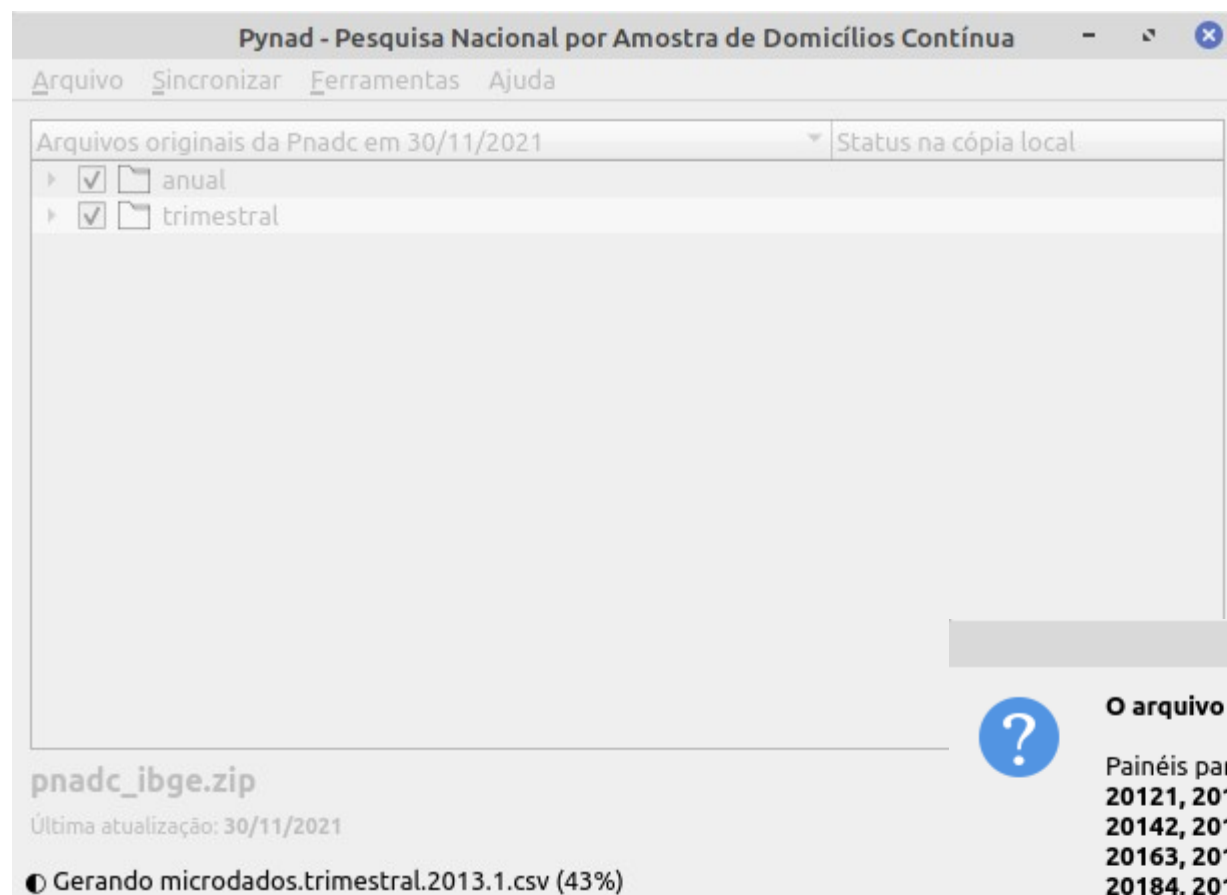
	Status na cópia local
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2016	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2017	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2018	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2019	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2020	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> 2021	
▶ <input checked="" type="checkbox"/> <input type="checkbox"/> atualizacoes_divulgacao_trimestral_2021...	sincronizado

pnadc_novembro_2011.zip
Última atualização: 26/11/2021

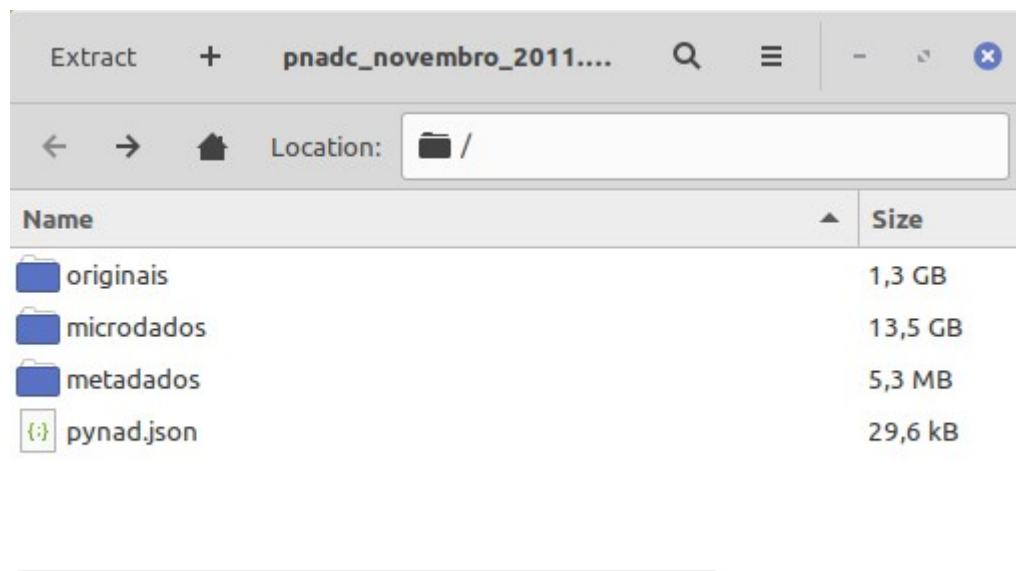
Verificando arquivos da Pnadc em ftp.ibge.gov.br: 36 arquivos em 7s

Atualização de 30/11/21

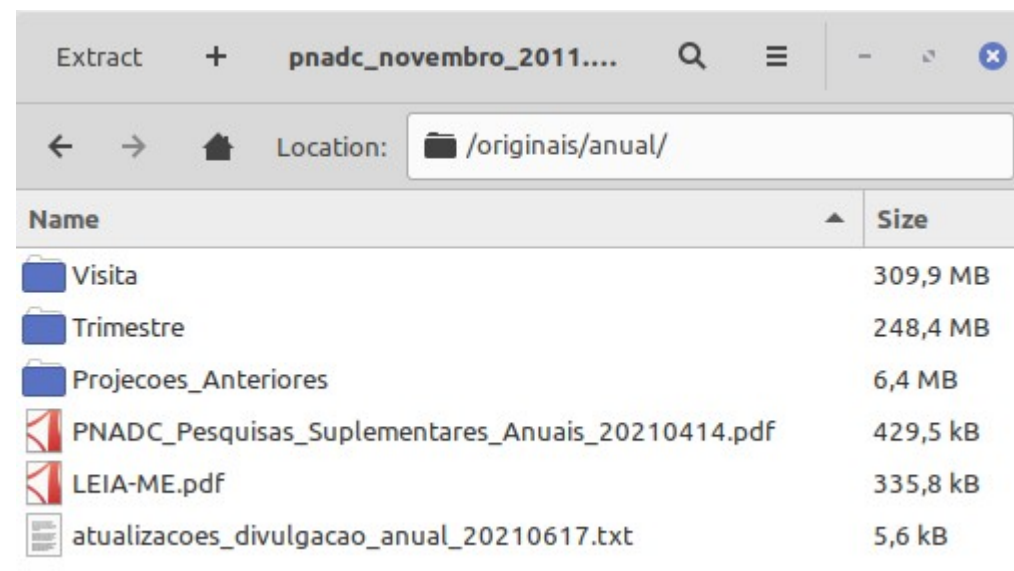




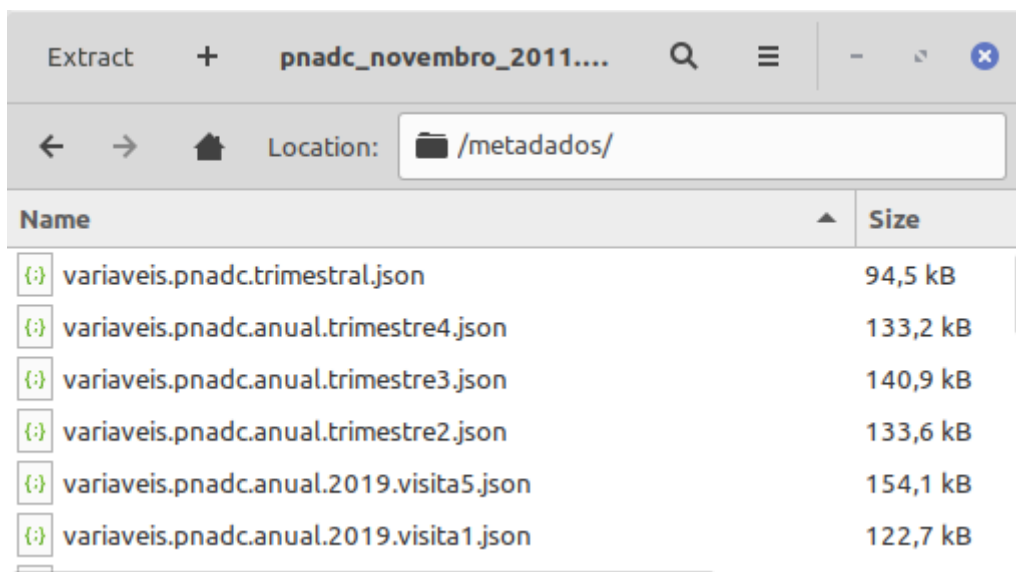
Arquivo com a cópia local – arquivos originais e derivados



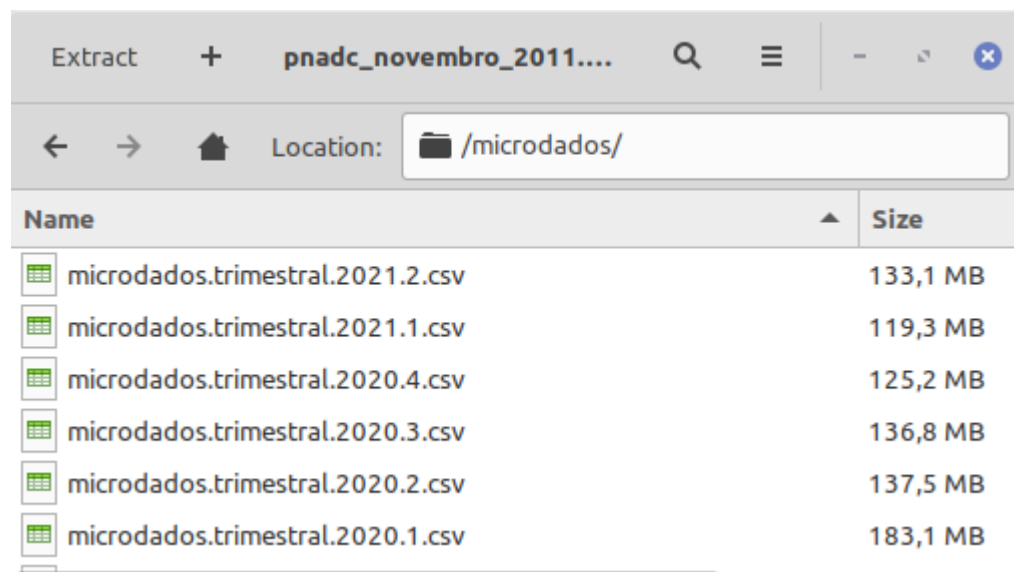
Name	Size
originais	1,3 GB
microdados	13,5 GB
metadados	5,3 MB
pynad.json	29,6 kB



Name	Size
Visita	309,9 MB
Trimestre	248,4 MB
Projecoes_Anteriores	6,4 MB
PNADC_Pesquisas_Suplementares_Anuais_20210414.pdf	429,5 kB
LEIA-ME.pdf	335,8 kB
atualizacoes_divulgacao_anual_20210617.txt	5,6 kB

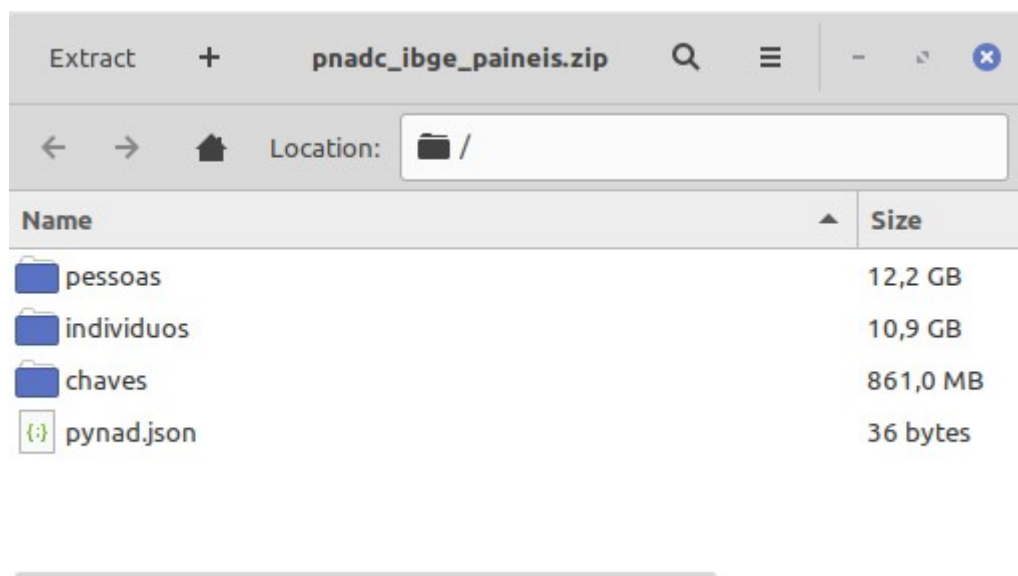


Name	Size
variaveis.pnadc.trimestral.json	94,5 kB
variaveis.pnadc.anual.trimestre4.json	133,2 kB
variaveis.pnadc.anual.trimestre3.json	140,9 kB
variaveis.pnadc.anual.trimestre2.json	133,6 kB
variaveis.pnadc.anual.2019.visita5.json	154,1 kB
variaveis.pnadc.anual.2019.visita1.json	122,7 kB



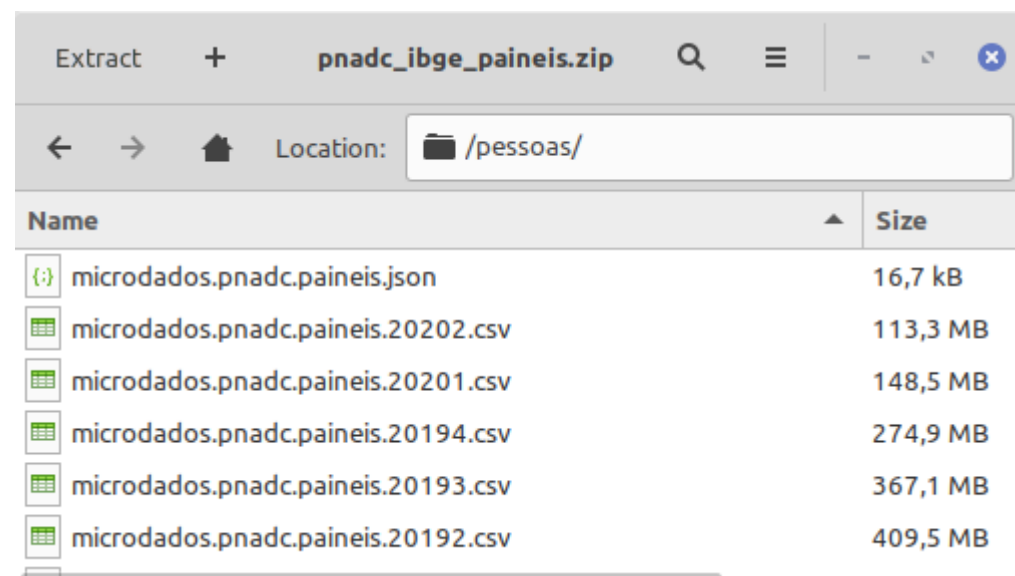
Name	Size
microdados.trimestral.2021.2.csv	133,1 MB
microdados.trimestral.2021.1.csv	119,3 MB
microdados.trimestral.2020.4.csv	125,2 MB
microdados.trimestral.2020.3.csv	136,8 MB
microdados.trimestral.2020.2.csv	137,5 MB
microdados.trimestral.2020.1.csv	183,1 MB

Arquivo com os microdados de painéis



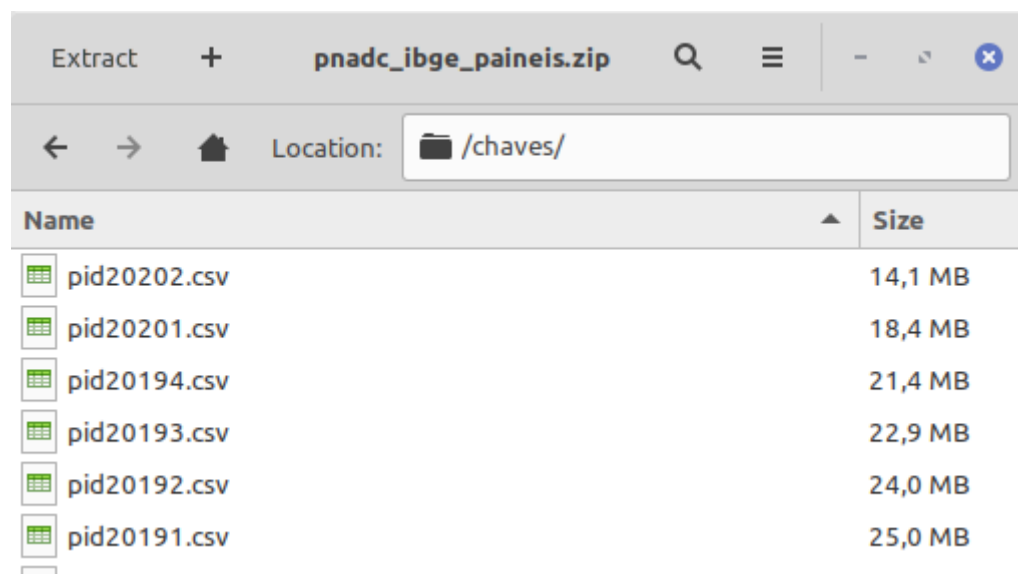
File explorer window showing the root directory of the extracted archive. The location is `/`. The table lists the following files and folders:

Name	Size
peessoas	12,2 GB
indivíduos	10,9 GB
chaves	861,0 MB
pynad.json	36 bytes



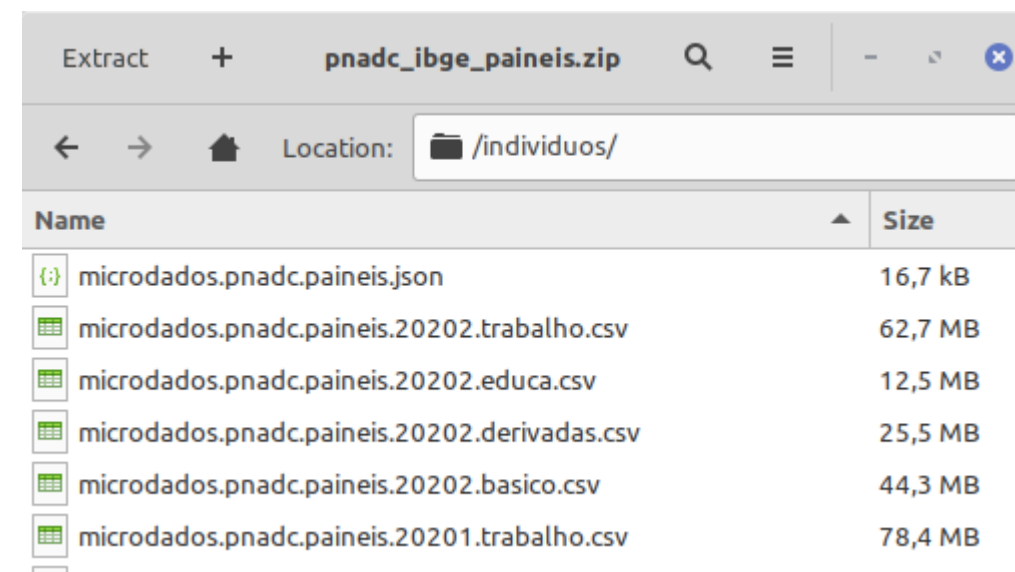
File explorer window showing the contents of the `/peessoas/` directory. The table lists the following files:

Name	Size
microdados.pnadc.paineis.json	16,7 kB
microdados.pnadc.paineis.20202.csv	113,3 MB
microdados.pnadc.paineis.20201.csv	148,5 MB
microdados.pnadc.paineis.20194.csv	274,9 MB
microdados.pnadc.paineis.20193.csv	367,1 MB
microdados.pnadc.paineis.20192.csv	409,5 MB



File explorer window showing the contents of the `/chaves/` directory. The table lists the following files:

Name	Size
pid20202.csv	14,1 MB
pid20201.csv	18,4 MB
pid20194.csv	21,4 MB
pid20193.csv	22,9 MB
pid20192.csv	24,0 MB
pid20191.csv	25,0 MB



File explorer window showing the contents of the `/indivíduos/` directory. The table lists the following files:

Name	Size
microdados.pnadc.paineis.json	16,7 kB
microdados.pnadc.paineis.20202.trabalho.csv	62,7 MB
microdados.pnadc.paineis.20202.educa.csv	12,5 MB
microdados.pnadc.paineis.20202.derivadas.csv	25,5 MB
microdados.pnadc.paineis.20202.basico.csv	44,3 MB
microdados.pnadc.paineis.20201.trabalho.csv	78,4 MB

Distribuições de microdados da PnadC

Trimestral:

- questionário básico das entrevistas do trimestre;
- domicílios de cinco painéis (identificados pelo número da visita);
- projeção de população no meio do trimestre

Anual de trimestre:

- questionário básico e suplementar de trimestre de todas as entrevistas do trimestre;
- entrevistas com domicílios de cinco painéis (identificados pelo número da visita);
- projeção de população no meio do trimestre

Anual de visita:

- questionário básico e suplementar de visita de todas as entrevistas do ano;
- entrevistas com domicílios de quatro painéis (identificados por ano e trimestre);
- projeção de população no meio do ano
- aproximadamente 80% da amostra de um trimestre

Os painéis da PnadC são compostos por domicílios:

- Um novo painel é iniciado a cada trimestre
- Os domicílios selecionados são visitados 5 vezes
- Um domicílio pode ter de 0 a 5 entrevistas realizadas
- As entrevistas são registradas por pessoa moradora na visita
- Sem entrevista não tem registro nos microdados
- Os domicílios são identificados: **upa v1008 v1014**

Uma base específica de microdados *cross-section* tem:

- apenas uma entrevista de cada domicílio
- entrevistas de domicílios de painéis diferentes

Para montar um painel de domicílios é preciso buscar o registro das pessoas entrevistadas nas cinco visitas em diferentes bases de microdados

ETAPA 1 Montar bases de microdados de painel com registros de pessoas

Exemplo - painel 2016-2:

- registros de pessoas em 10 arquivos de microdados!
- base trimestral está contida na base anual de trimestre
- Sete arquivos precisam ser lidos para obter os registros completos de pessoas nas cinco visitas do painel 2016-2:
 - visita 1: base anual do 2o trimestre (educação) de 2016; e anual de 1a visita de 2016;
 - visita 2: base trimestral do 3o trimestre de 2016;
 - visita 3: base anual do 4o trimestre (TICs) de 2016;
 - visita 4: base trimestral do 1o trimestre de 2017;
 - visita 5: base anual do 2o trimestre de 2017; e anual de 5a visita de 2017.

Pynad analisa a cópia local e separa os painéis, criando arquivos de microdados com os registros completos de todas as pessoas entrevistadas nas cinco visitas a um domicílio do painel (pasta pessoas)

Arquivo do painel 2016-2 com registros completos de pessoas

Data Editor (Browse)								
File Edit View Data Tools								
ano[1] 2016								
	ano	trimestre	upa	v1008	v1016	v2005	vd5008	vdi5008
1	2016	3	110000319	1	2	1	.	.
4	2016	4	110000319	1	3	1	.	2385
5	2017	1	110000319	1	4	1	.	.
6	2017	2	110000319	1	5	1	937	937
9	2016	2	110000319	2	1	1	278	278
11	2016	3	110000319	2	2	1	.	.
18	2016	4	110000319	2	3	1	.	429
20	2017	1	110000319	2	4	1	.	.
25	2017	2	110000319	2	5	1	650	650
31	2016	2	110000319	3	1	1	420	420
35	2016	3	110000319	3	2	1	.	.
37	2016	4	110000319	3	3	1	.	656
45	2017	1	110000319	3	4	1	.	.
48	2017	2	110000319	3	5	1	787	787

Ready

Vars: 8 of 539 Order: Modified

Obs: 187,577 of 568,064

Filter: On

Mode: Browse

ETAPA 2 Identificação de grupos domésticos e indivíduos

O grupo de moradores do domicílio pode variar por visita:

- 1 a 5 grupos domésticos distintos
- indivíduos que não estão em todas as entrevistas

Os grupos domésticos e seus indivíduos não estão identificados

- v2003 identifica a pessoa em uma visita com entrevista
- usado para montar o registro completo da pessoa na entrevista
- um único indivíduo pode receber números de ordem (v2003) de pessoa diferentes em cada entrevista
- a condição no domicílio (v2005) varia de acordo com a pessoa de referência escolhida pelos moradores na entrevista
- data de nascimento e sexo podem variar por entrevista
- outras características variam ainda mais (cor, alfabetização, trabalho...)

Dentro de um domicílio, os atributos para identificar os registros de pessoa que correspondem a um indivíduo são:

número de ordem, condição no domicílio, data de nascimento, sexo

Para um indivíduo estes atributos podem variar:

- **por acréscimo ou exclusão de moradores**, que pode alterar a condição no domicílio e a ordem atribuída pelo aplicativo da coleta
- **por mudança da pessoa responsável**, que pode alterar a condição no domicílio e a ordem
- **por retificação da declaração ou do registro dos atributos de identificação**, quando em uma ou mais entrevistas: há um erro de declaração da parte do entrevistado, ou de registro da resposta pelo entrevistador; ou a data de nascimento de algum morador é ignorada; ou há uma mudança real de sexo

Variação da ordem e condição no domicílio segundo a pessoa responsável

Morador	Data de nascimento	Número de ordem (v2003) e condição no domicílio (v2005)				
		Entrevista 1	Entrevista 2	Entrevista 3	Entrevista 4	Entrevista 5
<i>Maria Ferreira</i>	15/02/1936	6 Sogra	6 Mãe	1 Responsável	7 Convivente	7 Avó
<i>Mário Cruz</i>	20/11/1952	1 Responsável	2 Cônjuge	3 Genro	2 Pai	3 Padrasto
<i>Teresa Ferreira</i>	04/05/1956	2 Cônjuge	1 Responsável	2 Filha resp.	3 Madrasta	4 Mãe
<i>João Cruz</i>	11/09/1977	4 Filho resp.	5 Enteado	7 Convivente	1 Responsável	6 Irmão
<i>Mariana Ferreira</i>	19/07/1982	5 Enteada	4 Filho resp.	4 Neta	5 Irmã	1 Responsável
<i>Júlia Ferreira Cruz</i>	26/06/1992	3 Filha ambos	3 Filha ambos	5 Neta	4 Irmã	5 Irmã
<i>Bruno Ferreira</i>	15/08/2012	7 Agregado	7 Neto	6 Bisneto	6 Agregado	2 Filho resp.

O número de ordem é atribuído de acordo com a condição no domicílio e a data de nascimento (ou idade)

Neste exemplo, a identificação poderia ser feita por idade

Mas se o sexo e a data de nascimento mudassem em algumas entrevistas, a identificação não seria perfeita (e.g. data ignorada 99/99/9999)

ETAPA 2A Tratamento das datas de nascimento ignoradas

Se a data de nascimento de uma pessoa é ignorada, a idade é estimada

Frequentemente:

- a data de nascimento não é ignorada em todas as entrevistas
- só há uma pessoa com data ignorada
- quando há mais de uma, as idades estimadas são diferentes

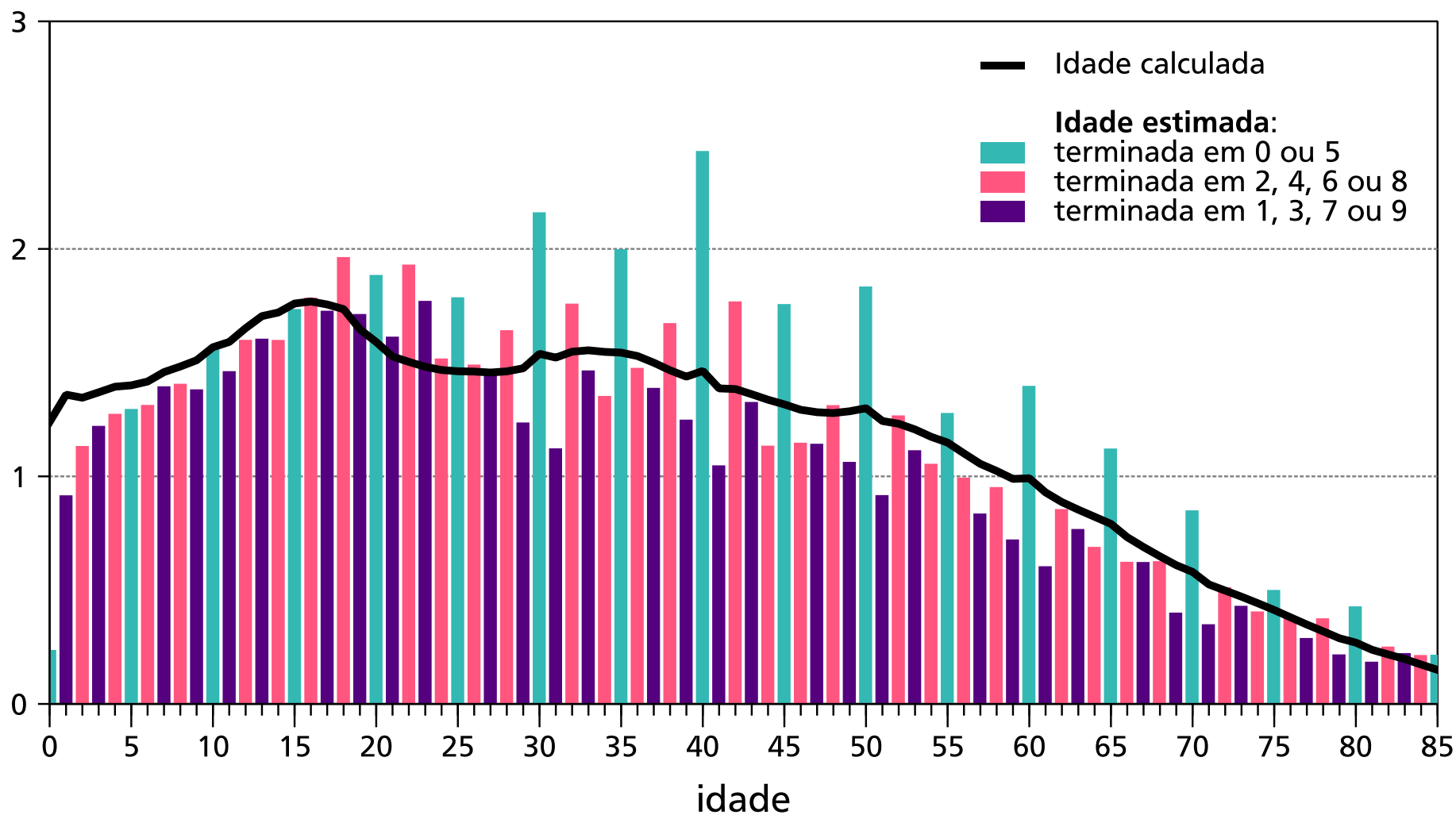
Logo, é possível achar a data de nascimento declarada em outra entrevista

Mas a idade estimada sofre a “atração por dígito”, o arredondamento das idades para múltiplos de 5 ou pares

Se em uma visita de 2012 um morador tinha data de nascimento ignorada e idade estimada em 40 anos, o ano de nascimento da idade estimada é 1972

Se em outras visitas a data de nascimento foi declarada, o ano de nascimento pode ser diferente, p.ex. 1971 ou 1973

Idade calculada e estimada



Por domicílio, considerando todas as pessoas entrevistadas

Para cada pessoa com data de nascimento ignorada são listadas as pessoas com data de nascimento conhecida que:

- não estão na mesma entrevista
- são do mesmo sexo
- ano de nascimento estimado e declarado com até 3 anos de diferença
- condição no domicílio igual, ou “equivalente” na troca de responsável (pessoas responsáveis e cônjuges; filhos e enteados; pai/mãe, sogro)
- ordena pela diferença entre anos de nascimento

A pessoa com data ignorada recebe a data de nascimento da pessoa mais próxima

Se não houver uma pessoa compatível com data declarada, a pessoa permanece com a data ignorada, mas o ano de nascimento passa a ser o implicado pela idade estimada: 99/99/AAAA

ETAPA 2B Identificação dos grupos domésticos

O domicílio é ocupado por mais de um grupo doméstico quando há grupos de indivíduos com conjuntos exclusivos de entrevistas

A identificação de um grupo doméstico não necessita da identificação de todos os indivíduos

Se um indivíduo “tem” pessoas em todas as visitas, há um grupo doméstico

Maior parte dos grupos domésticos possui ao menos um indivíduo com atributos estáveis – sexo e data de nascimento – nas entrevistas

Facilita a identificação do grupo doméstico, a despeito de eventuais mudanças e erros de declaração ou registro nos atributos dos outros moradores

Registros com erros poderiam levar a fusões ou separações impróprias de grupos domésticos

A separação dos grupos domésticos parte das pessoas do domicílio agrupadas por entrevistas:

- seleciona-se a entrevista com maior número de pessoas com data de nascimento declarada
- cria-se um grupo doméstico com as pessoas dessa entrevista
- atribui ao grupo doméstico todas as pessoas das outras entrevistas em que havia ao menos uma pessoa com o sexo e data de nascimento idênticos ao de uma pessoa já atribuída
- havendo entrevistas não atribuídas, existe outro grupo doméstico, e o processo é repetido desconsiderando as pessoas que já foram atribuídas

Domicílios com mais de um grupo doméstico não servem para o painel

A maior parte dos domicílios tem apenas um grupo doméstico em todas as entrevistas: 84% nos painéis de 2012, chegando a 95% no 2019-4

ETAPA 2C Identificação dos indivíduos

As etapas anteriores consideravam todas as pessoas entrevistadas nas visitas a um domicílio – esta considera todas as pessoas de um grupo doméstico (na maior parte dos domicílios, é a mesma coisa)

Em um domicílio com cinco entrevistas e apenas um grupo doméstico, um indivíduo que aparece em todas as entrevistas possui cinco registros de pessoas

Na identificação dos registros de pessoas que correspondem às aparições de um indivíduo são usados classificadores distintos

Os classificadores são tentados até que se obtenha uma identificação: do mais simples ao mais complexo

Os indivíduos são classificados segundo o método de identificação

A classificação fornece um parâmetro da qualidade da identificação

GRUPOS DOMÉSTICOS SEM MUDANÇA

classe 1: sexo, data de nascimento, condição no domicílio e ordem constantes

classe 2: sexo e data de nascimento constantes (controle para gêmeos); troca da pessoa responsável ou mudança de condição no domicílio

GRUPOS DOMÉSTICOS COM MUDANÇA

classe 3: sexo e data de nascimento constantes, **presentes em todas as entrevistas** (indivíduo sem mudança em grupo com mudança)

classe 4: sexo e data de nascimento constantes, mas **não estão presentes em todas as entrevistas do grupo doméstico** – indivíduos que nasceram, morreram, saíram do grupo ou se juntaram ao mesmo – casamentos e separações, jovens que saem de casa...

GRUPOS DOMÉSTICOS COM MUDANÇA

Suponha um grupo doméstico com vários indivíduos que não aparecem em todas as entrevistas de um grupo doméstico:

A {1, 2, 4, 5}; B {3}; C {1, 5}; D {2, 3}; E {4}

Os indivíduos que não aparecem na mesma entrevista seriam classe 4

Mas podem ser “fragmentos” de um indivíduo, separados por erros ou mudanças de declaração ou registro: A e B podem ser um só indivíduo; ou C, D e E

classe 5: sexo e data de nascimento constantes em grupo doméstico onde há outros indivíduos com conjuntos mutuamente exclusivos de entrevistas; há sempre 2 ou mais indivíduos classe 5

Todos os pares de indivíduos classe 5 são comparados: se há mudanças de sexo e/ou mudança **parcial** de data de nascimento, os indivíduos são agregados e um novo código é atribuído

GRUPOS DOMÉSTICOS COM MUDANÇA

classe 6: agregação de indivíduos com variação na data de nascimento ou sexo e condição no domicílio constante

classe 7: agregação de indivíduos com condição no domicílio variável

Exemplo: mãe da pessoa responsável no painel 2015-1 tem duas idades estimadas: 88 {1, 2, 3} e 90 {4, 5}, gerando 3 datas de nascimento

99/99/1927 99/99/1925 99/99/1926

Inicialmente gera 3 indivíduos classe 5; mas como a única diferença entre eles é o ano de nascimento, são agregados em um indivíduo classe 6

Se tivesse havido, depois da primeira entrevista, troca de pessoa responsável e reclassificação da mãe como sogra, a junção geraria um indivíduo classe 7

Se a diferença de anos de nascimento fosse maior que 3 anos, os fragmentos não são agregados unidos, ficando 3 indivíduos classe 5.

Painel 2016-2, UPA 350583328, domicílio 11

Indivíduo	Sexo	Data de nascimento	Ordem	Condição no domicílio	Entrevistas
1 – classe 3	masculino	61 anos - 99/99/1955	1	pessoa responsável	1
		01/09/1955			2, 3, 4, 5
2 – classe 3	feminino	30 anos - 99/99/1986	2	filha	1
		11/08/1986			2, 3, 4, 5
3 – classe 6	masculino	6 anos - 99/99/2010	3	neto	1
		27/12/2010			2, 3
		27/12/2009			4, 5
4 – classe 3	masculino	31 anos - 99/99/1985	4	outro parente	1
		28/03/1985			2, 3, 4, 5

Painel 2012-1, UPA 430070316, domicílio 8

Indivíduo	Sexo	Data de nascimento	Ordem	Condição no domicílio	Entrevistas
1 – classe 3	masculino	15/02/1975	1	pessoa responsável	1, 2, 3, 4, 5
2 – classe 3	feminino	20/11/1974	2	cônjuge	1, 2, 3, 4, 5
3 – classe 5	feminino	11/09/1995	3	filha de ambos	1, 3, 4, 5
4 – classe 6	masculino	04/05/2007	4	filho de ambos	1, 3, 4
		05/05/2007			5
		04/05/2007	3		2
5 – classe 5	feminino	19/07/2007	4	filha de ambos	2

O indivíduo 5 é improvável: filha de ambos, 4 anos, só em uma entrevista?

Seriam os indivíduos 3 e 5 fragmentos de um só indivíduo?

Note-se que a presença do ind. 5 na 2ª entrevista alterou a ordem: na mesma categoria de condição, a ordem é do mais velho ao mais novo

Painel 2012-1, UPA 320043538, domicílio 3

Indivíduo	Sexo	Data de nascimento	Ordem	Condição no domicílio	Entrevistas
1 – classe 3	masculino	12/03/35	1	pessoa responsável	1, 2, 3, 5
2 – classe 6	feminino	11/08/37	2	cônjuge	1, 2, 3
		11/08/36			5
3 – classe 7	masculino	09/05/51	3	outro parente	1, 2, 3
	feminino			filha	5

Indivíduo 2: mudança de data de nascimento quando muda o ano (5ª ent.)

Indivíduo 3: mudança de sexo e de condição no domicílio (seria classe 6 sem a mudança da condição)

Foi um erro ou uma troca real de sexo?

A identificação não envolve a decisão sobre qual atributo é “correto”

Painel 2012-1, UPA 110000548, domicílio 5, grupo doméstico 1

Indivíduo	Sexo	Data de nascimento	Ordem	Condição no domicílio	Entrevistas
1 – classe 1	masculino	04/02/1974	1	pessoa responsável	1
2 – classe 1	feminino	18/03/1992	2	cônjuge	1
3 – classe 1	masculino	10/07/2011	3	filho de ambos	1
4 – classe 1	masculino	26/03/1998	4	filho do responsável	1
5 – classe 1	masculino	19/02/2001	5	filho do responsável	1

Nada especial aqui: é um grupo doméstico com apenas uma entrevista

O grupo interessante é o 2º deste domicílio...

Painel 2012-1, UPA 110000548, domicílio 5, grupo doméstico 2

Indivíduo	Sexo	Data de nascimento	Ordem	Condição no domicílio	Entrevistas
1 – classe 3	masculino	16/03/1942	1	pessoa responsável	2
		70 anos - 99/99/1942	2	cônjuge	3
		72 anos - 99/99/1941			5
2 – classe 5	feminino	24/06/1943	2	cônjuge	2
			1	pessoa responsável	3
3 – classe 5	feminino	03/09/1964	1	pessoa responsável	5

Erro produzido pelo tratamento de data de nascimento?

O casal da entrevista 5 é realmente o mesmo da 2 e 3?

Casos assim são interessantes, mas são relativamente poucos:

- Até 2015, 87-88% dos indivíduos são classe 1 a 4
- A partir de 2016, 94% dos indivíduos

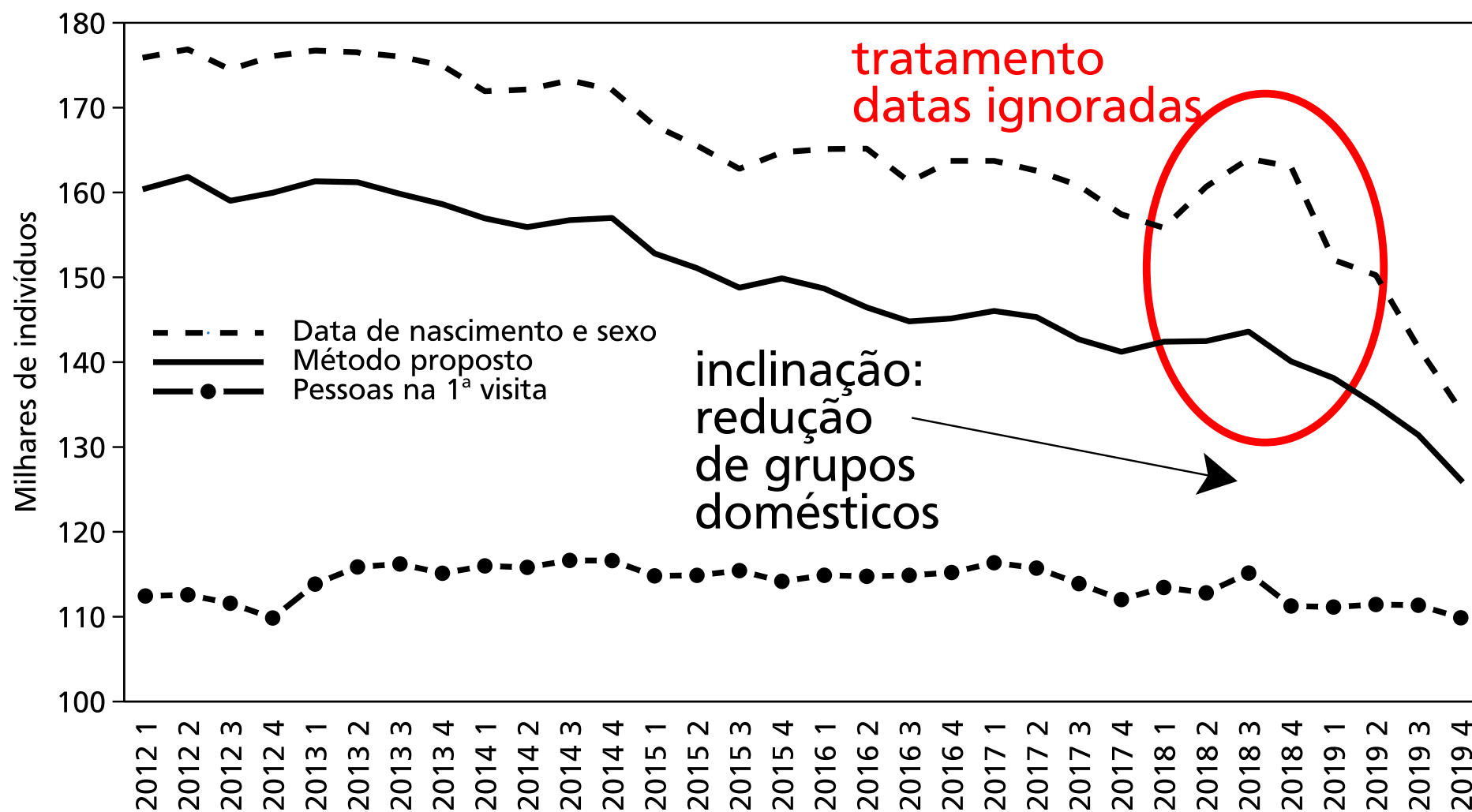
Muito difícil a identificação estar errada nestes casos

O tratamento de datas de nascimento ignorada é o “pulo do gato”: por conta do problema da atração por dígito

“Benchmarking”

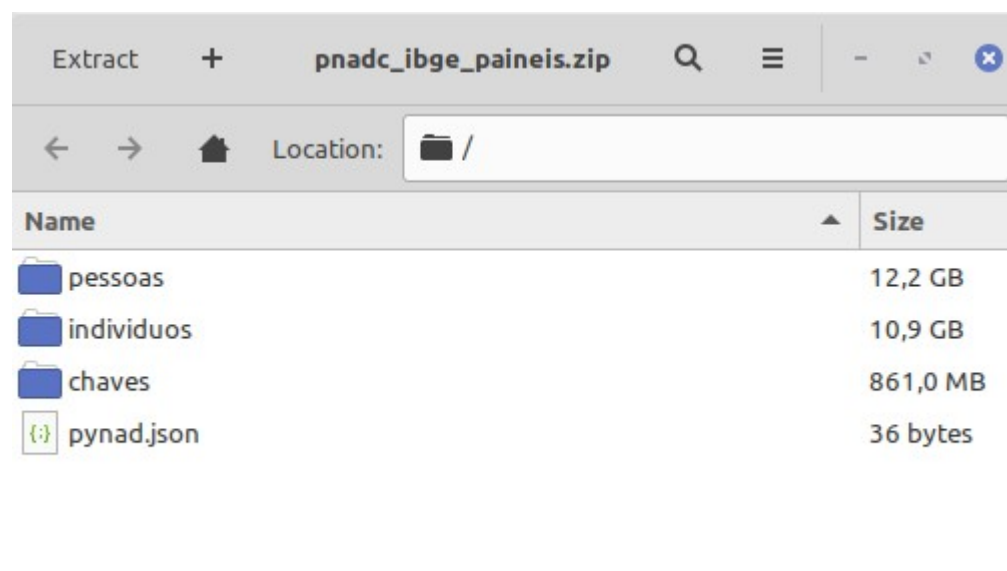
Se os grupos domésticos não mudassem de composição e se houvesse apenas um grupo por domicílio:

o número de indivíduos seria igual ao número de pessoas na 1ª entrevista



Os arquivos com as variáveis de identificação – prefixo “pid” – são armazenados na pasta “chaves” do arquivo de painéis, e as variáveis para fazer o *merge* com os painéis de domicílios na pasta “pessoas” são

upa v1008 v1016 v2003



Os arquivos já identificados e reformatados (cada linha é um indivíduo), separados por temas (básico, educação, trabalho, derivadas...), estão na pasta “indivíduos”

Arquivo do painel 2016-2 com registros de indivíduos (básico)

Data Editor (Browse)											
File Edit View Data Tools											
5R x 9C 9											
	v1008	pidgrp	pidind	pidcla	v2007_1	v2007_2	v2007_3	v2007_4	v2007_5		
27	9	1	1	3	1	1	1	.	.		
28	9	1	2	3	2	2	2	.	.		
29	9	1	3	4	2		
30	9	1	4	4	1		
31	9	1	5	4	1		
32	10	1	1	3	1	1	1	1	1		
33	10	1	2	3	2	2	2	2	2		
34	10	1	3	3	1	1	1	1	1		
35	10	1	4	4	.	.	.	1	1		
36	11	1	1	1	1	1	1	1	1		
37	11	1	2	1	2	2	2	2	2		
38	11	1	3	1	2	2	2	2	2		
39	12	1	1	1	1	1	1	1	1		
40	12	1	2	1	2	2	2	2	2		
41	12	1	3	1	1	1	1	1	1		
42	14	1	1	1	1	1	1	1	1		

Ready Vars: 10 of 149 Order: Dataset Obs: 146,454 Filter: Off Mode: Browse

Arquivo do painel 2016-2 com registros de indivíduos (básico)

Data Editor (Browse)										
File Edit View Data Tools										
5R x 9C 9										
	v1008	pidgrpent	pidindent	pidcla	v2007_1	v2007_2	v2007_3	v2007_4	v2007_5	
27	9	7	7	3	1	1	1	.	.	
28	9	7	7	3	2	2	2	.	.	
29	9	7	1	4	2	
30	9	7	1	4	1	
31	9	7	1	4	1	
32	10	31	31	3	1	1	1	1	1	
33	10	31	31	3	2	2	2	2	2	
34	10	31	31	3	1	1	1	1	1	
35	10	31	24	4	.	.	.	1	1	
36	11	31	31	1	1	1	1	1	1	
37	11	31	31	1	2	2	2	2	2	
38	11	31	31	1	2	2	2	2	2	
39	12	31	31	1	1	1	1	1	1	
40	12	31	31	1	2	2	2	2	2	
41	12	31	31	1	1	1	1	1	1	
42	14	31	31	1	1	1	1	1	1	

Ready Vars: 10 of 149 Order: Dataset Obs: 146,454 Filter: Off Mode: Browse

Variáveis do painel

pid	código do painel no formato AAAAT e.g. 20121, 20122
pidgrp	número sequencial do grupo doméstico no domicílio [1, 5]
pidgrpent	código das visitas com entrevistas do grupo doméstico [1, 31]
pidind	número sequencial do indivíduo no grupo doméstico
pidindent	código das visitas com entrevistas do indivíduo [1, 31]
pidcla	código da classe de identificação do indivíduo [1, 7]
piddnd	dia de nascimento imputado para data ignorada
piddnm	mês de nascimento imputado para data ignorada
piddna	ano de nascimento imputado para data ignorada

Chave única de indivíduos:

pid upa v1008 pidgrp pidind

Um painel contém apenas 20% da amostra trimestral (25% anual de visita)

Declarando o desenho com *survey* (Stata) e população na 1ª visita:

```
svyset upa [w=v1027_1], strata(estrato) poststrata(posest) postweight(v1029_1) singleunit(scaled)
```

Expand para a população do trimestre na primeira visita (*variavel_1*)

Para obter estimativas de ponto iguais às obtidas com *survey*, *pynad* calcula novos pesos para as bases de indivíduos (equivale a v1028):

$$pidpeso_{-v_{ij}} = \frac{w_{ij} N_i}{\sum_{j=1} w_{ij}}$$

Há um peso por projeção de população dos 5 trimestres de um painel

A escolha “natural” é usar o peso da 1ª visita, mas isto cria problemas para usar indivíduos que não aparecem nela:

- uma criança que nasce pode levar o peso do domicílio

- mas e um novo membro adulto, que estava na população da 1ª visita?

Atrito (medido como perda de domicílios da amostra)

O painel é de domicílios, não de grupos domésticos ou indivíduos

Assumindo que começamos sempre com a 1ª visita:

- o painel deve ter as 5 entrevistas
- o painel deve ter a 1ª e a 5ª (dados de outras rendas)

1ª fonte de atrito é a não realização das entrevistas subsequentes:

- **atrito médio 18% (5 entrevistas) e 10% (1ª e 5ª)**

O atrito era maior até 2015, diminuiu, e aumentou em 2019: 37% e 25%

2ª fonte de atrito é a troca de grupo doméstico:

- **atrito médio 26% e 21%**

No último painel de 2019, 43% e 31%

3ª fonte de atrito é a perda de indivíduos (depende do estudo)

Painel é um subgrupo de indivíduos:

- em domicílios com apenas um grupo doméstico
- presentes nas entrevistas selecionadas

Viés do atrito: maior quando os eventos que levam a mudanças de (ou no) grupo doméstico, ou que impedem a realização de todas as entrevistas, estão correlacionados com os indicadores relevantes

Exemplo: frequência a curso superior (de graduação, pós ou especialização) para os jovens de 20 a 24 anos:

A saída de casa e a trajetória escolar estão relacionadas e são influenciadas por fatores comuns. Jovens de famílias pobres têm propensão a deixar a escola, sair de casa, começar a trabalhar, e formar suas próprias famílias mais cedo que jovens de famílias de maior renda

Há razões para desconfiar que o atrito nesta faixa etária produza um indicador com viés

Jovens de 20 a 24 anos estudando em curso superior:

22,9% ($\pm 0,7$) no segundo trimestre de 2018, todas as visitas

23,2% ($\pm 1,5$) no segundo trimestre de 2018, primeira visita

26,1% ($\pm 1,8$) na primeira visita no segundo trimestre de 2018 (painel)

20,6% ($\pm 0,6$) no segundo trimestre de 2019, todas as visitas (21 a 25 anos)

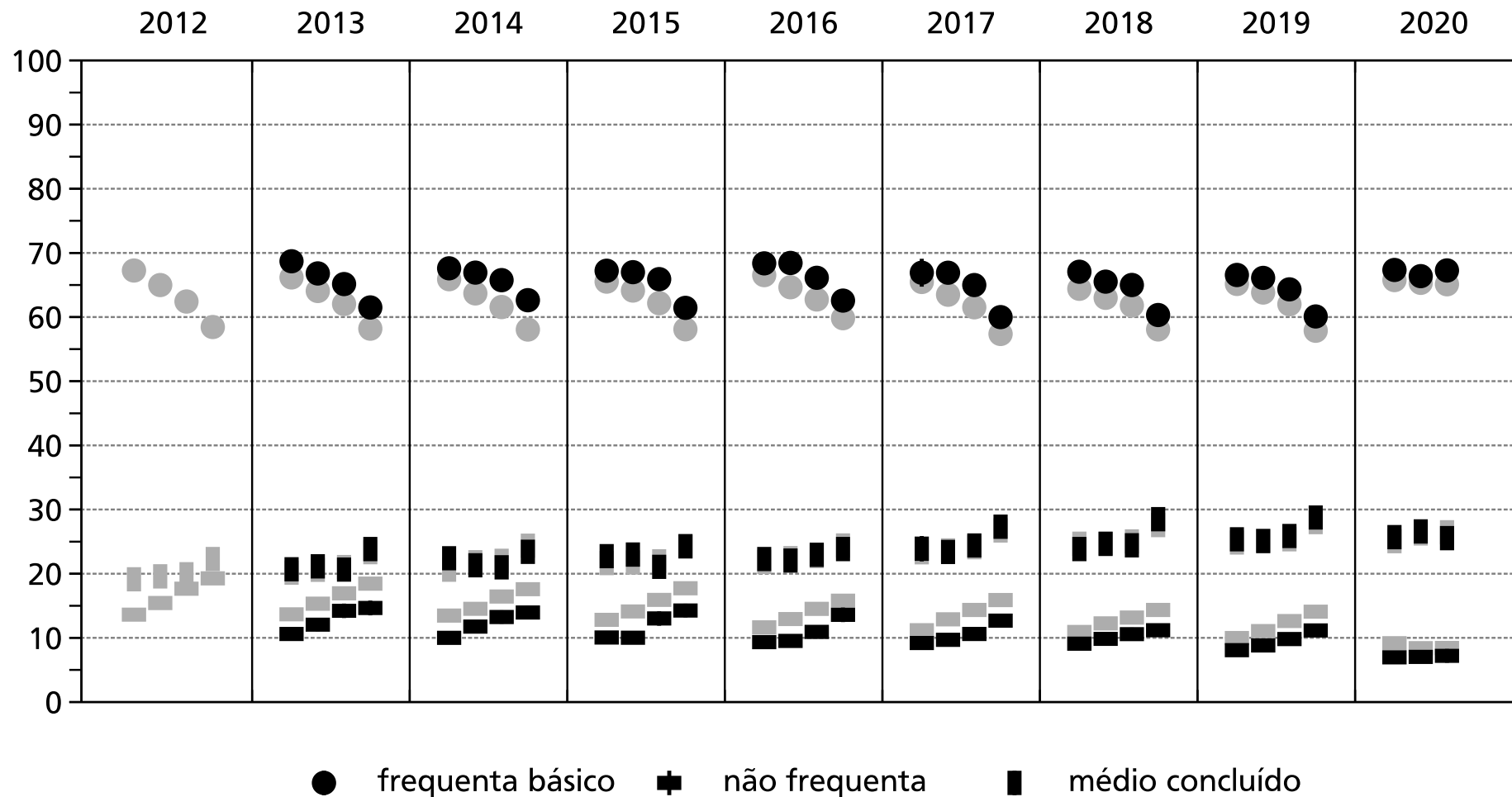
20,3% ($\pm 1,4$) no segundo trimestre de 2019, quinta visita (21 a 25 anos)

22,9% ($\pm 1,7$) na quinta visita no segundo trimestre de 2019 (painel)

Jovens que não saem de casa ou estão em grupos domésticos estáveis tem indicadores melhores...

Os dados em cross-section fornecem uma base para avaliar o viés do atrito do painel

Frequência a escola na 5ª visita do painel vs tri. (14-18 para 15-19 anos)



Consistência dos dados

Todo dado pode ter sido registrado com erro, por vários motivos

Em uma pesquisa *cross-section* com uma amostra enorme, como a PnadC trimestral, erros aleatórios na captação ou no registro dos dados ocorrem, mas provavelmente se anulam, e não são detectados

Com os dados em painel, alguns erros que passariam despercebidos surgem como inconsistências




Há indivíduos no painel com atributos que implicam evoluções, transições e mudanças improváveis, ou mesmo impossíveis

O problema é que o objetivo de ter um painel é justamente estudar a evolução dos atributos

Em alguns casos, como educação, é possível discernir variações inconsistentes, mas em outros não

Próx. Slide: gráfico de transição 14-18 para 15-19 anos

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20
1º tri.	<div>1 11 9</div> <div>2 67 0</div> <div>5 5 0</div>	<div>0 12 10</div> <div>2 65 1</div> <div>5 4 0</div>	<div>0 12 11</div> <div>2 65 1</div> <div>5 5 0</div>	<div>0 11 11</div> <div>1 67 0</div> <div>5 4 0</div>	<div>0 13 11</div> <div>1 65 0</div> <div>5 4 0</div>	<div>0 12 11</div> <div>1 66 0</div> <div>5 4 0</div>	<div>0 13 12</div> <div>1 65 0</div> <div>4 3 0</div>	<div>0 13 12</div> <div>1 66 0</div> <div>3 4 0</div>
2º tri.	<div>1 11 9</div> <div>2 64 0</div> <div>6 6 0</div>	<div>0 12 9</div> <div>2 64 0</div> <div>6 5 0</div>	<div>0 11 12</div> <div>2 65 0</div> <div>5 4 0</div>	<div>1 11 10</div> <div>2 66 0</div> <div>5 4 0</div>	<div>0 11 12</div> <div>1 65 0</div> <div>5 4 0</div>	<div>0 13 11</div> <div>1 64 0</div> <div>6 4 0</div>	<div>0 13 12</div> <div>1 65 0</div> <div>5 4 0</div>	<div>1 13 13</div> <div>1 65 0</div> <div>4 3 0</div>
3º tri.	<div>1 11 9</div> <div>2 62 0</div> <div>8 6 1</div>	<div>1 11 10</div> <div>2 63 1</div> <div>8 5 0</div>	<div>1 11 10</div> <div>2 64 0</div> <div>8 4 0</div>	<div>1 11 11</div> <div>2 64 1</div> <div>6 4 0</div>	<div>0 13 11</div> <div>2 63 0</div> <div>6 4 0</div>	<div>0 12 12</div> <div>1 63 0</div> <div>6 4 0</div>	<div>1 13 13</div> <div>1 63 0</div> <div>5 4 0</div>	<div>0 14 11</div> <div>1 65 0</div> <div>5 2 0</div>
4º tri.	<div>1 12 12</div> <div>2 59 0</div> <div>9 5 0</div>	<div>1 11 12</div> <div>2 60 1</div> <div>8 5 1</div>	<div>0 11 13</div> <div>2 59 0</div> <div>8 6 0</div>	<div>1 11 13</div> <div>2 60 1</div> <div>9 5 0</div>	<div>0 13 14</div> <div>1 58 0</div> <div>7 5 0</div>	<div>1 12 16</div> <div>2 58 0</div> <div>7 4 0</div>	<div>1 13 15</div> <div>1 58 0</div> <div>7 4 0</div>	<div></div> <div></div> <div></div>

 transição positiva
 transição negativa
 transição impossível

5ª visita	CM	1	11	9
	FQ	2	67	0
	NF	5	5	0
		NF	FQ	CM
		1ª visita		

NF = não frequente

FQ = frequente

CM = concluiu médio

Taxa de evasão escolar – Censo Educacional e Pnadc

