Anthony Damico <ajdamico@gmail.com>

PNAD
12 messages

Gabriel Cepaluni <gabi.cepal@gmail.com>   Tue, May 6, 2014 at 5:24 PM
To: djalma.pessoa@ibge.gov.br, ajdamico@gmail.com

Dear Damico,

I read some of your posts on PNAD (and also on Brazilian Census). Congratulations! They were very good.

I want to compare the development of German and Italian municipalities in Rio Grande do Sul with PNAD (I claim they are a natural experiment of history). I was able to marge individual level data from PNAD from 2012 to 2001. I stopped at 1999 because I couldn't find some variables I wanted, but I'm still working on it.

I have two questions. The first one is if you can suggest something to speed up my code. The second (a more substantive question) is if you believe PNAD is comparable from its outset until 2012. Above you will find my code for 2012 and 2011 PNAD.

2012 PNAD

```
load('/Users/gabrielcepaluni/Dropbox/Nat_Exp_RGS/PNAD/2012/Leitura_em_R/dicPNAD2012.RData')
dicpes2012 <- dicpes2012[-2, ]
load('/Users/gabrielcepaluni/Dropbox/Nat_Exp_RGS/PNAD/2012/Dados/PES2012.RData')
names(pnad12) <- dicpes2012$cod
pnad12 <- pnad12[, c('V0101', 'V0102', 'V0103', 'V0301', 'V0302', 'V3031', 'V3032', 'V3033', 'V0601', 'V0602', 'V6003', 'V6006', 'V6007', 'V6010', 'V0611',
'V4729')]
```

2011 PNAD
```
load('/Users/gabrielcepaluni/Dropbox/Nat_Exp_RGS/PNAD/2011/pnad2011.RData')
pnad11$UF <- NULL
pnad11 <- read.spss('/Users/gabrielcepaluni/Dropbox/Nat_Exp_RGS/PNAD/2011/pes_2011.sav',
            to.data.frame = TRUE, trim.factor.names = TRUE)
pnad11 <- pnad11[, c('V0101', 'V0102', 'V0103', 'V0301', 'V0302', 'V3031', 'V3032', 'V3033', 'V0601', 'V0602', 'V6003', 'V6006', 'V6007', 'V6010', 'V0611',
'V4729')]
load('/Users/gabrielcepaluni/Dropbox/Nat_Exp_RGS/PNAD/2011/Leitura em R/dicPNAD2011.RData')
dicdom2011 <- dicdom2011[-2, ]
pnad11dom <- read.fwf("/Users/gabrielcepaluni/Dropbox/Nat_Exp_RGS/PNAD/2011/Dados/DOM2011.TXT", widths = dicdom2011$tamanho,
            comment.char="", strip.white=TRUE)
names(pnad11dom) <- dicdom2011$cod
```

pnad11dom <- pnad11dom[, c('V0101', 'V0102', 'V0103', 'V4618', 'V4617')]
pnad11 <- merge(pnad11, pnad11dom)

pnad0512 <- rbind(pnad11, pnad12)

I want to have more data far back in time as possible. On the other hand, my computer is getting really slow, so that if PNAD samples is not so good in previous period, I might consider working with more recent samples.

Best,
--
Gabriel Cepaluni
Professor de Relações Internacionais
Assistant Professor of International Relations
Unesp-Franca

Anthony Damico <ajdamico@gmail.com>   Tue, May 6, 2014 at 7:43 PM
To: Gabriel Cepaluni <gabi.cepal@gmail.com>
Cc: djalma.pessoa@ibge.gov.br

hi gabriel, great to hear from you! i do not know anything about your substantive question but your code is slow because you are using `read.fwf` and `read.spss` instead of using the download automation script that we wrote for you! ;) our code puts everything into a database, but you can pull out tables if you have enough ram with `pnad12 <- dbReadTable( db , 'pnad2012' )` or specific variables with `pnad12.sub <- dbGetQuery( db , 'select v0101 , v0102 , v0103 , v4618 , v4617 from pnad2012' )` it is still smart to leave it running overnight, but at least with the download automation scripts, you know that they will work perfectly and you only have to do things once. if you run all of your analysis in SQL you should never run out of ram -- instead of `merge` use `INNER JOIN` and instead of `rbind` use `UNION ALL` -- but it is nicer to stick with r syntax, so just keep your "windows task manager" window open so you can monitor how much ram you are using while you work.. and search my archive to use rm() and gc() functions effectively. keep in touch

[Quoted text hidden]

Gabriel Cepaluni <gabi.cepal@gmail.com>   Tue, May 6, 2014 at 8:33 PM
To: Anthony Damico <ajdamico@gmail.com>
Cc: djalma.pessoa@ibge.gov.br

Thank you so much for answer. I was actually considering running your code, but I was in doubt I could update it to get all the individual level data from 1976 in this website. I'm a political scientist which means my computer skills are rather limited :).

2014-05-06 20:43 GMT-03:00 Anthony Damico <ajdamico@gmail.com>:

[Quoted text hidden]
[Quoted text hidden]

djalma.pessoa@ibge.gov.br <djalma.pessoa@ibge.gov.br>   Wed, May 7, 2014 at 10:39 AM
To: Gabriel Cepaluni <gabi.cepal@gmail.com>
Cc: ajdamico@gmail.com

Dear Gabriel

As far as know the PNADs from years before 2001 are comparable with the subsequent PNADs with respect to the concepts adopted but they might not include the same variables.
 It is worth mentioning that PNAD was not designed to afford precise estimates for the municipality level. If you compute the cvs you will get very high values and most of the municipalities don't have households in the sample.

Best regards,

Djalma

De:	Gabriel Cepaluni <gabi.cepal@gmail.com>
Para:	<djalma.pessoa@ibge.gov.br>, <ajdamico@gmail.com>
Data:	07/05/2014 08:07
Assunto:	PNAD

[Quoted text hidden]

Gabriel Cepaluni <gabi.cepal@gmail.com>   Wed, May 7, 2014 at 11:17 AM
To: djalma.pessoa@ibge.gov.br
Cc: ajdamico@gmail.com

Djalma,

Thank you so much for your answer. I don't know the scope of your work, but this site has more observations than at the IBGE website. They compiled all the microdata possible from PNAD and Census - http://www.fflch.usp.br/centrodametropole/1148 and http://www.fflch.usp.br/centrodametropole/v3/bases. php?retorno=1147&language=pt_br.

Actually merging all these files have been a pain, so that I will probably have to stop and learn one thing or two (or more :).

Best,
Gabriel

2014-05-07 11:39 GMT-03:00 <djalma.pessoa@ibge.gov.br>:

[Quoted text hidden]

[Quoted text hidden]

Anthony Damico <ajdamico@gmail.com>   Sat, Oct 11, 2014 at 9:48 AM
To: "djalma.pessoa@ibge.gov.br" <djalma.pessoa@ibge.gov.br>

hi!

i am writing a blog post on how to create beautiful survey-weighted maps. i have attached an example map. it looks great, doesn't it?? :)   notice that the attached map does not have borders. this is very important, because it uses a technique that allows **survey weights** instead of computing values at each specific geography.

in my blog post, i want to include an example that uses brazilian microdata!

in order to make the map correctly, i actually need two different data sets:

(1) the pnad, to actually compute the values that i am interested in
(2) the censo demografico, to spread those values across small area

i have attached both codebooks..

==================

i have three questions for you..

(1) what is the smallest geography available in the pnad that is also in the censo? (your discussion with gabriel is below..)

(2) what is the smallest geography available in the 2010 censo?

(3) where can i find a shapefile from ibge with the smallest 2010 censo geography?

[[i think that #3 is one of these, but i am not sure which??]]

thanks!!

[Quoted text hidden]

3 attachments

A922E[1].png
120K

Layout_microdados_Amostra.xls
158K

Dicionário de variáveis de domicílios da Pesquisa Básica - 2013.xls
51K

djalma.pessoa@ibge.gov.br <djalma.pessoa@ibge.gov.br>   Wed, Oct 15, 2014 at 11:27 AM
To: Anthony Damico <ajdamico@gmail.com>

Hi!

Awesome!

1) Unfortunatelly the smallest geography available for PNAD also in thecenso is UF (state). There is no municipality code in the PNAD.

2) the smallest geography available in the 2010 censo is the "area de ponderacao" (weighting area). This is a census EA conglomerate having at most one municipality One municipality can have more than one weighting area..

3) You're right. It is one of them, depending on what you want to do.

Probably you will have to use estimates from the census sample. From PNAD estimates you will not get great maps.

De: Anthony Damico <ajdamico@gmail.com>
Para: "djalma.pessoa@ibge.gov.br" <djalma.pessoa@ibge.gov.br>,
Data: 13/10/2014 08:17
Assunto: Fwd: PNAD

[Quoted text hidden]

[anexo "A922E[1].png" removido por Djalma Galvao Pessoa/DPE/IBGE][anexo "Layout_microdados_Amostra.xls" removido por Djalma Galvao Pessoa/DPE/IBGE][anexo "Dicionário de variáveis de domicílios da Pesquisa Básica - 2013.xls" removido por Djalma Galvao Pessoa/DPE/IBGE]

Anthony Damico <ajdamico@gmail.com>   Wed, Oct 15, 2014 at 11:42 AM
To: "djalma.pessoa@ibge.gov.br" <djalma.pessoa@ibge.gov.br>

for #1 -- but state isn't the smallest geography in the pnad, right? is it possible to combine geographies in the censo and match some sub-state estimate? or do they not crosswalk?

for #3 -- i want the shapefile of the area de ponderacaos from the 2010 censo.. do you know which one it is?

here are some more pre-smoothed maps. the final product will be smoothed like the last one i sent :)

[Quoted text hidden]

---

2 attachments

alternative pre-smoothed map.png
48K

pre-smoothed map.png
39K

---

djalma.pessoa@ibge.gov.br <djalma.pessoa@ibge.gov.br>   Wed, Oct 15, 2014 at 2:36 PM
To: Anthony Damico <ajdamico@gmail.com>

#1 - yes, there is smaller geography like urban and rural also metropolitan region. But the shape file is not available for these geographies.
#2   ftp://geoftp.ibge.gov.br/malhas_digitais/censo_2010/areas_de_ponderacao

De:	Anthony Damico <ajdamico@gmail.com>
Para:	"djalma.pessoa@ibge.gov.br" <djalma.pessoa@ibge.gov.br>,
Data:	15/10/2014 14:29
Assunto:	Re: Fwd: PNAD

[Quoted text hidden]

[anexo "alternative pre-smoothed map.png" removido por Djalma Galvao Pessoa/DPE/IBGE][anexo "pre-smoothed map.png" removido por Djalma Galvao Pessoa/DPE/IBGE]

Anthony Damico <ajdamico@gmail.com>   Wed, Oct 15, 2014 at 2:41 PM
To: "djalma.pessoa@ibge.gov.br" <djalma.pessoa@ibge.gov.br>

wait wait wait so then can i merge urban/rural and/or metropolitan region onto the censo 2010 any way? if both the censo and the pnad have state +
urban/rural, then that is very useful. i only need the shapefile for the units in the censo 2010, i only need the pnad so that i can crosswalk it :) is it possible to
crosswalk any sub-state geographies from pnad to censo?

thanks for pointing me to link #2 ;) you saved me a lot of time hehe

[Quoted text hidden]

---

djalma.pessoa@ibge.gov.br <djalma.pessoa@ibge.gov.br>   Thu, Oct 16, 2014 at 9:28 AM
To: Anthony Damico <ajdamico@gmail.com>

I fowarded your question to a coleague.

One map of great interest in Brazil is the so called **Poverty Map**. This map shows the proportion of poor people by municipality. Its not
possible to construct it just by using data from any survey from IBGE.
PNAD and POF don't provide estimates with the necessary precision at the municipality level.

I take part in a group that uses models to combine surveys and census results to get good estimates but this is a bit complex to serve as
example.

Another approach is to use direct estimates from the census survey data for the 2010 year. The computations are very simple and are is in
one script included in
brazilian census blog at asdfree, used to produce a barplot.

This might produce a nice map showing the big contrasts in Brazil.

-----Anthony Damico <ajdamico@gmail.com> escreveu: -----
Para: "djalma.pessoa@ibge.gov.br" <djalma.pessoa@ibge.gov.br>

De: Anthony Damico <ajdamico@gmail.com>
Data: 15/10/2014 04:35 PM
Assunto: Re: Fwd: PNAD

wait wait wait so then can i merge urban/rural and/or metropolitan region onto the censo 2010 any way? if both the censo and the pnad have state + urban/rural, then that is very useful. i only need the shapefile for the units in the censo 2010, i only need the pnad so that i can crosswalk it :) is it possible to crosswalk any sub-state geographies from pnad to censo?

thanks for pointing me to link #2 ;) you saved me a lot of time hehe

On Wed, Oct 15, 2014 at 2:36 PM, <djalma.pessoa@ibge.gov.br > wrote:
#1 - yes, there is smaller geography likeurban and rural also metropolitan region. But the shape file is not available for these geographies.
#2 ftp://geoftp.ibge.gov.br/malhas_digitais/censo _2010/areas_de_ponderacao

De: Anthony Damico < ajdamico@gmail.com>
Para: "djalma.pessoa@ibge.gov.br "<djalma.pessoa@ibge.gov.br >,
Data: 15/10/2014 14:29
Assunto: Re: Fwd: PNAD
_____

for #1 -- but state isn't the smallest geography in thepnad, right? is it possible to combine geographies in the censo and match some sub-state estimate? or do they not crosswalk?

for #3 -- i want the shapefile of the area de ponderacaos from the 2010 censo.. do you know which one it is?

here are some more pre-smoothed maps. the final product will be smoothed like the last one i sent :)

On Wed, Oct 15, 2014 at 11:27 AM, < djalma.pessoa@ibge.gov.br >wrote:
Hi!

Awesome!

1) Unfortunatelly the smallest geography available for PNAD also in thecenso is UF (state). There is no municipality code in the PNAD.

2) the smallest geography available in the 2010 censo is the "area de ponderacao" (weighting area). This is a census EA conglomeratehaving at most one municipality One municipality can have more than one weighting area..

3) You're right. It is one of them, depending on what you want to do.

Probably you will have to use estimates from the census sample. From PNAD estimates you will not get great maps.

De:	AnthonyDamico < ajdamico@gmail.com >
Para:	" djalma.pessoa@ibge.gov.br "< djalma.pessoa@ibge.gov.br >,
Data:	13/10/201408:17
Assunto:	Fwd:PNAD

hi!

i am writing a blog post on how to create beautiful survey-weighted maps. i have attached an example map. it looks great, doesn't it?? :)   notice that the attached map does not have borders. this is very important, because it uses a technique that allows **survey weights** instead of computing values at each specific geography.

in my blog post, i want to include an example that uses brazilian microdata!

in order to make the map correctly, i actually need two different data sets:

(1) the pnad, to actually compute the values that i am interested in
(2) the censo demografico, to spread those values across small area

i have attached both codebooks..

===================

i have three questions for you..

(1) what is the smallest geography available in the pnad that is also in the censo? (your discussion with gabriel is below..)

(2) what is the smallest geography available in the 2010 censo?

(3) where can i find a shapefile from ibge with the smallest 2010 censo geography?

[[i think that #3 is oneof these , but i am not sure which??]]

thanks!!

---------- Forwarded message ----------
From: **Gabriel Cepaluni** < gabi.cepal@gmail.com>
Date: Wed, May 7, 2014 at 11:17 AM
Subject: Re: PNAD

To: djalma.pessoa@ibge.gov.br
Cc: ajdamico@gmail.com

Djalma,

Thank you so much for your answer. I don't know the scope of your work, but this site has more observations than at the IBGE website. They compilled all the microdata possible from PNAD and Census - http://www.fflch.usp.br/centrodametropole/1148 and http://www.fflch.usp.br/centrodametropole/v3/bases.php?retorno=1147&language=pt_br .

Actually merging all these files have been a pain, so that I will probably have to stop and learn one thing or two (or more :).

Best,
Gabriel

2014-05-07 11:39 GMT-03:00 < djalma.pessoa@ibge.gov.br >:

Dear Gabriel

As far as know the PNADs from years before 2001 are comparable with the subsequent PNADs with respect to the concepts adopted but they might not include the same variables.
 It is worth mentioning that PNAD was not designed to afford precise estimates for the municipality level. If you compute the cvs you will get very high values and most of the municipalities don't have households in the sample.

Best regards,

Djalma

De: GabrielCepaluni < gabi.cepal@gmail.com >

Para: < djalma.pessoa@ibge.gov.br >,< ajdamico@gmail.com >

Data: 07/05/201408:07

Assunto: PNAD

[Quoted text hidden]

[anexo "A922E[1].png" removidopor Djalma Galvao Pessoa/DPE/IBGE][anexo "Layout_microdados_Amostra.xls" removido por Djalma Galvao Pessoa/DPE/IBGE][anexo "Dicionário de variáveis de domicílios

da Pesquisa Básica - 2013.xls" removido por Djalma Galvao Pessoa/DPE/IBGE]

[anexo "alternative pre-smoothedmap.png" removido por Djalma Galvao Pessoa/DPE/IBGE][anexo "pre-smoothed map.png" removido por Djalma Galvao Pessoa/DPE/IBGE]

---

djalma.pessoa@ibge.gov.br <djalma.pessoa@ibge.gov.br>   Thu, Oct 16, 2014 at 2:58 PM
To: Anthony Damico <ajdamico@gmail.com>

Hi, that's the answer I got from Vivi. She's a reliable source. I hope you still recall your portuguese. If you have any doubts, please ask me.

----- Encaminhado por Djalma Galvao Pessoa/DPE/IBGE em 16/10/2014 03:54 PM -----
Para: Djalma Galvao Pessoa/DPE/IBGE@IBGE, pessoad@gmail.com
De: Viviane Cirillo Carvalho Quintaes/DPE/IBGE
Data: 16/10/2014 03:04 PM
Assunto: Re: Enc: Re: Fwd: PNAD

Oi Djalma, boa tarde.

Para se ter todos os arquivos de microdados das pesquisas (PNAD ou Censo) e os shapefiles, é necessário seguir os seguintes passos:

1º) Na PNAD

1.1) Construir um código que agrega a variável UF e a V4105 para ter a codificação Urbana e Rural.

1.2) Construir um código que agrega a variável UF e a V4107 para ter a codificação Região metropolitana

2º) No Censo de 2010

2.1) Construir um código que agrega a variável de UF (V0001) e a   V1006 para ter a codificação Urbana e Rural

2.2) A variável V1004 possui o c ódigo completo da região metropolitana. Porém este código completo não é igual ao da PNAD que ele acabou de construir. Assim, é necessário criar uma codificação igual ao da PNAD a partir da variável V1004. Vale ressaltar, que a PNAD não contém todas as regiões metropolitanas, somente aquelas onde consta o município da capital. Segue abaixo, a composição dos códigos de município das capitais.

# Vetor com os códigos das capitais
capitais<-c(1100205,1200401,1302603,1400100,1501402,1600303,1721000,2111300,2211001,2304400,

2408102,2507507,2611606,2704302,2800308,2927408,3106200,3205309,3304557,3550308,
4106902,4205407,4314902,5002704,5103403,5208707,5300108)

# Vetor com os nomes das capitais
capitaisnomes<-c("Porto Velho","Rio Branco","Manaus","Boa Vista","Belém",
"Macapá","Palmas","São Luís","Teresina","Fortaleza","Natal","João Pessoa",
"Recife","Maceió","Aracaju","Salvador","Belo Horizonte","Vitória","Rio de Janeiro",
"São Paulo","Curitiba","Florianópolis","Porto Alegre","Campo Grande","Cuiabá",
"Goiânia","Brasília")

3º) Construção das Malhas Digitais

3.1) Para se ter a malha digital da região metropolitana deve-se juntar o shapefile de município com o arquivo de dados do Censo por município. A partir daí pode-se utilizar a variável construída do Censo igual ao da PNAD e criar o shapefile da região metropolitana. E também, a partir da variável construída de urbano e rural do Censo fazer o seu shapefile.

3.2) Depois de feito isto ele poder á juntar todos os arquivo da PNAD, Censo com seus respectivos shapefiles.

Viviane Quintaes
_____ _____
Estatística
Instituto Brasileiro de Geografia e Estatística -IBGE
Diretoria de Pesquisas (DPE) - Coordenação de Métodos e Qualidade (COMEQ)
Tel: (0xx21) 21420458
viviane.quintaes@ibge.gov.br
Av. República do Chile, 500 - 10°. andar
20031-170 - Centro - Rio de Janeiro.

Inactive hide details for Djalma Galvao Pessoa---16/10/2014 10:05:57---Vivi, Você saberia responder isso? Djalma Galvao Pessoa---
16/10/2014 10:05:57---Vivi, Você saberia responder isso?

De: Djalma Galvao Pessoa/DPE/IBGE
Para: Viviane Cirillo Carvalho Quintaes/DPE/IBGE@IBGE
Data: 16/10/2014 10:05
Assunto: Enc: Re: Fwd: PNAD

Vivi,
Você saberia responder isso?
Abs
Djalma

----- Encaminhado por Djalma Galvao Pessoa/DPE/IBGE em 16/10/2014 10:04 AM -----

[Quoted text hidden]