

Rio de Janeiro, 24 de Abril de 2016

Aluno: Carlos Mattoso

Matrícula: 1210553

Disciplina: INF1335 - Introdução a Data Science com R e Python

A1 - Dengue e Renda

Introdução

O trabalho abaixo apresenta duas análises da associação entre taxa de incidência de dengue e renda domiciliar por Unidade Federativa (UF). A primeira análise é feita com base em dados pré-processados da PNAD Contínua de renda domiciliar per capita publicados pelo datasus [2]. A segunda análise é feita com base nos microdados disponibilizados pelo IBGE [1], tendo sido a entregue na primeira entrega deste trabalho.

Setup

Primeiramente, é necessário que sejam baixados todos os *datasets* necessários para as análises e instaladas as bibliotecas que serão utilizadas. Rode o código abaixo para que o ambiente seja preparado por você. Caso você já tenha os anos baixados em disco, cada um em seu próprio sub-diretório com nome igual ao do ano, basta redefinir o valor de `base.dir` antes de executar o código abaixo.

```
In [1]: # -----  
# -----  
# SETUP Base  
# -----  
# -----  
printf <- function(...) invisible(print(sprintf(...))) # source: http:  
//stackoverflow.com/a/13023329  
  
DEBUG = T  
if (!DEBUG) {  
  noW <- function(...) invisible(suppressWarnings(...))  
} else {  
  noW <- function(...) invisible(...)  
}  
  
# Handy function to load RData files into object.  
load.rdata.into.obj <- function(input) {  
  env <- new.env()  
  nm <- load(input, env)[1]  
  env[[nm]]  
}
```

```

# Necessário devido ao encoding dos arquivos SAS
options(encoding = "windows-1252");

# Cria o diretório em que os datasets serão armazenados
base.dir = './data' # !! Mude aqui se já tiver tudo baixado
if (!dir.exists(base.dir)) {
  dir.create(base.dir)
}

# Baixa CSV de incidência de dengue por UF do github da disciplina
dengue.url = 'https://raw.githubusercontent.com/simonedjb/inf1335/master/dados/dengueUF.csv'
download.file(dengue.url, destfile=paste(base.dir, 'dengueUF.csv', sep='/' ), method='curl')

# já produz algumas variáveis úteis

# -----
# SETUP PARA ANÁLISE 1
# -----

# Baixa o arquivo do datasus com a renda média domiciliar anual per capita já pré-processada
url.datasus = 'https://gist.githubusercontent.com/calmmattoso/7d3ffa79bf7b5c9a530a1749180ecf42/raw/b3b6c95e9878d9belce7cf394fb5a85cce393929/datasus_2001_2008.csv'
datasus.file.path = paste(base.dir, 'pnad_renda_datasus_2001_2008.csv', sep='/')
if (!file.exists(datasus.file.path)) {
  temp <- tempfile()
  download.file(url.datasus, destfile=temp, method='curl')
  # Abre o csv e faz um rápido pré-processamento para eliminar colunas e linhas desnecessárias
  datasus.csv <- read.csv(temp, sep=';', skip=3, fileEncoding='cp1252')
  datasus.cols <- names(datasus.csv)
  datasus.csv <- datasus.csv[, !(datasus.cols %in% grep("Informa", datasus.cols, value=T)), drop=F] # remove colunas de infos estatísticas
  datasus.csv <- datasus.csv[1:28,] # remove linhas desnecessárias no final do arquivo
  write.csv(datasus.csv, file=datasus.file.path, row.names=F)

  unlink(temp)
} else {
  printf("Arquivo de renda do Datasus já existe!")
}

# -----
# SETUP PARA ANÁLISE 2
# -----

```

```
# Biblioteca para leitura de arquivos SAS
if (!require('SAScii')) {
  install.packages('SAScii', repos="http://cran.rstudio.com/")
}
library('SAScii')

years = 2001:2008
base.url = "ftp://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacio
nal_por_Amostra_de_Domicilios_anual/microdados/reponderacao_2001_2012/
";
for (year in years) {
  extra = ''
  if (year == 2003 || year == 2007) {
    extra = '_20150814' # isto pq os arquivos de 2003 e 2007 foram
    atualizados em 14/08/2015
  }
  file.name = paste("PNAD_reponderado_", year, extra, '.zip', sep='')
  url = paste(base.url, file.name, sep='')

  # Baixa e descompacta o arquivo do ano caso não já exista a pasta
  do ano
  if (!dir.exists(paste(base.dir, year, sep='/'))) {
    temp <- tempfile()
    download.file(url, destfile = temp, method='curl')
    unzip(temp, exdir = base.dir)
    unlink(temp) # deleta o .zip
  } else {
    printf('Dados já existentes pro ano %d.', year)
  }

  # Renomeia os arquivos da sub-pasta `Input` segundo um formato com
  um
  input.files = list.files(paste('data', year, 'Input', sep='/'),
    full.names=T)
  input.files.ren = gsub('_', ' ', tolower(input.files))
  file.rename(input.files, input.files.ren)
}
```

```
[1] "Arquivo de renda do Datasus já existe!"
```

```
Loading required package: SAScii
```

```
[1] "Dados já existentes pro ano 2001."
[1] "Dados já existentes pro ano 2002."
[1] "Dados já existentes pro ano 2003."
[1] "Dados já existentes pro ano 2004."
[1] "Dados já existentes pro ano 2005."
[1] "Dados já existentes pro ano 2006."
[1] "Dados já existentes pro ano 2007."
[1] "Dados já existentes pro ano 2008."
```

Item 1

Acho que seria interessante avaliar a associação para dado ano entre gastos em políticas públicas destinadas ao combate a dengue e a incidência da dengue em determinada UF. Se possível, creio que fazer esta análise a nível de município seria ainda mais útil, já que estaríamos atacando um problema menor. Além disso, seria válido avaliar investimentos gerais em saneamento básico e se isto estaria relacionado com quedas na incidência da dengue, já que é senso comum que a disseminação do mosquito se deve em grande parte ao descaso com o tratamento de resíduos, levando ao acúmulo de água.

Item 2

Análise Nova - *Dataset* da PNADc pré-processado

Esta análise utiliza um *dataset* pré-processado da PNADc que apresenta os dados de renda anual domiciliar per capita por UF para o período de 2001 a 2008, extraído na plataforma TABNET do DataSUS [3]. Na seção de *Setup* deste *notebook* tal *dataset* já foi baixado e parcialmente pré-processado para estar formatado de maneira a facilitar a análise. Abaixo, exibe-se seu conteúdo:

```
In [2]: datasus.csv <- read.csv(datasus.file.path, dec=",")
names(datasus.csv) <- c('UF', 2001:2008)
datasus.csv <- datasus.csv[1:27,]

datasus.csv
```

Out[2]:

	UF	2001	2002	2003	2004	2005	2006	2007	2008
1	Rondônia	513.04	551.3	482.18	501.13	531.55	578.56	539.47	573.7
2	Acre	618.51	603.7	504.81	405.38	422.79	510.85	529.08	520.71
3	Amazonas	439.19	417.79	397.31	389.29	421.36	440.01	444.1	475.56
4	Roraima	473.38	409.74	448.92	328.37	381.98	524.53	464.19	562.05
5	Pará	413.99	431.97	361.52	380.09	375.02	404.09	441.66	457.6
6	Amapá	540.96	445.1	466.59	395.07	491.78	473.23	500.64	480.35
7	Tocantins	412.81	369.52	385.99	422.79	423.24	462.42	504.69	575.05
8	Maranhão	273.38	275.78	268.36	295.77	263.43	356.22	350.87	360.99
9	Piauá	303.32	324.55	294.96	316.87	333.13	394.2	437.89	446.98
10	Ceará	338.77	332.68	304.49	326.23	351.24	373.8	388.86	436.99
11	Rio Grande do Norte	374.73	379.15	340.13	374.29	445.42	458.65	483.32	519.82
12	Paraíba	326.34	361.95	330.73	359.22	398.61	434.96	458.29	491.66
13	Pernambuco	369.52	373.27	321.23	369.58	379.82	417.08	408.5	448.85
14	Alagoas	294.76	292.76	288.43	275.07	292.67	384.65	412.2	394.37
15	Sergipe	348.85	371.23	382.86	416.51	409.66	454.62	464.86	483.84

UF	2001	2002	2003	2004	2005	2006	2007	2008
16 Bahia	332.13	335.58	324.29	333.95	360.33	406.89	417.38	460.51
17 Minas Gerais	538.95	550.5	518.41	546.93	581.89	650.04	660.76	713.18
18 Esp�rito Santo	562.54	607.52	547.49	594.71	630.2	665.47	694.11	724.62
19 Rio de Janeiro	829.7	818.01	790.4	809.49	826.48	924.9	902.27	954.39
20 S�o Paulo	877.93	864.71	798.67	783.05	864.82	928.65	941.87	960.21
21 Paran�	657.18	664.23	649.05	715.37	724.71	768.22	861.4	853.57
22 Santa Catarina	739.04	709.13	741.5	734.29	811.36	899.01	918.49	949.65
23 Rio Grande do Sul	744.57	735.13	724.09	749.24	757.25	819.18	830.49	895.82
24 Mato Grosso do Sul	591.72	620.42	566.04	569.45	608	689.8	793.78	757.81
25 Mato Grosso	585.52	604.48	526.64	594.92	580.56	633.8	621.5	788.59
26 Goi�s	553.32	566.84	526.31	584.94	617.76	637.47	706.94	734.01
27 Distrito Federal	1118.71	1195.75	1095.35	1117.8	1215.74	1350.02	1491.24	1528.67

Como pode-se ver h  pequenos problemas de formata  o nos nomes das UFs. Para corrigir isso, extraem-se seus nomes do arquivo `dengue.csv`:

```
In [3]: years.under.analysis = 2001:2008
dengue = read.csv(paste(base.dir, '/dengueUF.csv', sep=''), dec=".",
na.strings = '-')
dengue.cols = sapply('X', paste, years.under.analysis, sep='',
USE.NAMES = FALSE)
dengue.subset = dengue[1:27, dengue.cols]

dengue.new.cols = vector(mode="character", length =length(dengue.cols)
)
for (i in 1:length(names(dengue.subset))) {
  dengue.new.cols[i] = substr(names(dengue.subset)[i], 2, 5)
}
names(dengue.subset) = dengue.new.cols
row.names(dengue.subset) = dengue$UF[1:27]
datasus.csv$UF = dengue$UF[1:27]
datasus.csv
```

```
Out[3]:
```

UF	2001	2002	2003	2004	2005	2006	2007	2008
----	------	------	------	------	------	------	------	------

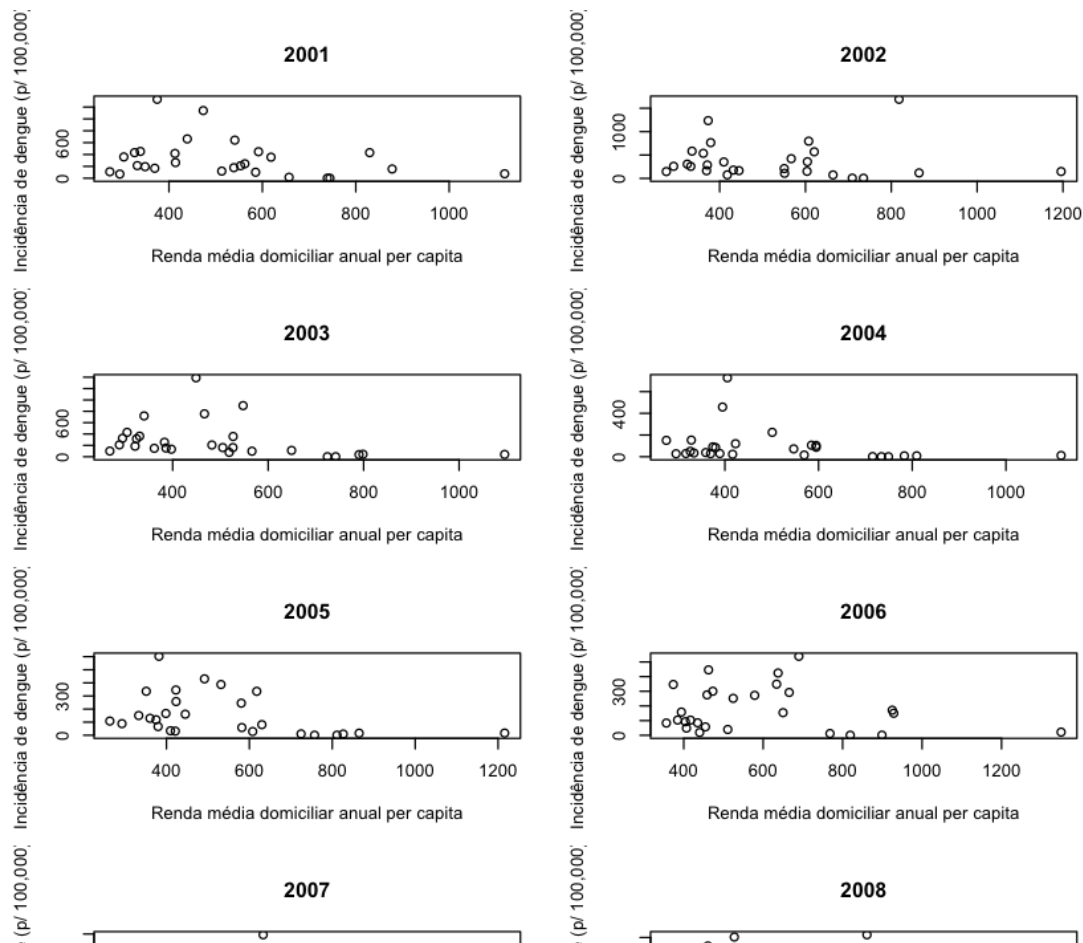
	UF	2001	2002	2003	2004	2005	2006	2007	2008
1	Rondonia	513.04	551.3	482.18	501.13	531.55	578.56	539.47	573.7
2	Acre	618.51	603.7	504.81	405.38	422.79	510.85	529.08	520.71
3	Amazonas	439.19	417.79	397.31	389.29	421.36	440.01	444.1	475.56
4	Roraima	473.38	409.74	448.92	328.37	381.98	524.53	464.19	562.05
5	Para	413.99	431.97	361.52	380.09	375.02	404.09	441.66	457.6
6	Amapa	540.96	445.1	466.59	395.07	491.78	473.23	500.64	480.35
7	Tocantins	412.81	369.52	385.99	422.79	423.24	462.42	504.69	575.05
8	Maranhao	273.38	275.78	268.36	295.77	263.43	356.22	350.87	360.99
9	Piaui	303.32	324.55	294.96	316.87	333.13	394.2	437.89	446.98
10	Ceara	338.77	332.68	304.49	326.23	351.24	373.8	388.86	436.99
11	Rio Grande do Norte	374.73	379.15	340.13	374.29	445.42	458.65	483.32	519.82
12	Paraiba	326.34	361.95	330.73	359.22	398.61	434.96	458.29	491.66
13	Pernambuco	369.52	373.27	321.23	369.58	379.82	417.08	408.5	448.85
14	Alagoas	294.76	292.76	288.43	275.07	292.67	384.65	412.2	394.37
15	Sergipe	348.85	371.23	382.86	416.51	409.66	454.62	464.86	483.84
16	Bahia	332.13	335.58	324.29	333.95	360.33	406.89	417.38	460.51
17	Minas Gerais	538.95	550.5	518.41	546.93	581.89	650.04	660.76	713.18
18	Espirito Santo	562.54	607.52	547.49	594.71	630.2	665.47	694.11	724.62
19	Rio de Janeiro	829.7	818.01	790.4	809.49	826.48	924.9	902.27	954.39
20	Sao Paulo	877.93	864.71	798.67	783.05	864.82	928.65	941.87	960.21
21	Parana	657.18	664.23	649.05	715.37	724.71	768.22	861.4	853.57
22	Santa Catarina	739.04	709.13	741.5	734.29	811.36	899.01	918.49	949.65
23	Rio Grande do Sul	744.57	735.13	724.09	749.24	757.25	819.18	830.49	895.82
24	Mato Grosso do Sul	591.72	620.42	566.04	569.45	608	689.8	793.78	757.81
25	Mato Grosso	585.52	604.48	526.64	594.92	580.56	633.8	621.5	788.59
26	Goiias	553.32	566.84	526.31	584.94	617.76	637.47	706.94	734.01

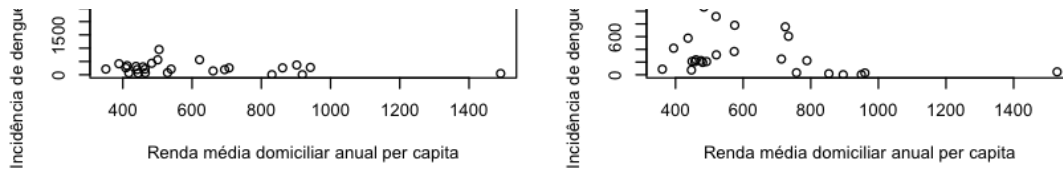
27	Distrito Federal	1118.71	1195.75	1095.35	1117.8	1215.74	1350.02	1491.24	1528.67
-----------	------------------	---------	---------	---------	--------	---------	---------	---------	---------

Pode-se agora fazer a análise. Primeiramente, são exibidos os gráficos, por ano, que relacionam a renda domiciliar anual per capita e a taxa de incidência de dengue para cada UF.

```
In [4]: layout(matrix(c(1,2,3,4,5,6,7,8), 4, 2, byrow = TRUE))
xlab.str = "Renda média domiciliar anual per capita"

for (i in 1:length(years.under.analysis)) {
  year = as.character(years.under.analysis[i])
  noW(plot(datasus.csv[,year], dengue.subset[,year], ylab="Incidência de dengue (p/ 100,000)",
    xlab=xlab.str, main = year))
}
```





Não parece haver, visualmente, uma associação entre as variáveis, visto que independente do aumento da renda não se observa um padrão de aumento ou queda na taxa de incidência de dengue. Na verdade, pode-se ver que muitos estados com maior renda apresentam taxas bastante elevadas de dengue, o que vai contra uma hipótese bastante razoável de que estados "mais ricos" teriam menos casos de dengue. A fim de reassegurar esta conclusão, apresentam-se abaixo as correlações destas variáveis, por ano:

```
In [5]: a <- datasus.csv[,2:ncol(datasus.csv)]
        b <- dengue.subset[,1:ncol(dengue.subset)]
        inter.cor = diag(cor(a, b))
        printf('Média da correlação: %f', mean(inter.cor))
        inter.cor
```

```
[1] "Média da correlação: -0.192377"
```

```
Out[5]:      2001 -0.26847840905719
          2002 -0.0156216954296057
          2003 -0.305521976287597
          2004 -0.266735770070378
          2005 -0.394319134596965
          2006 -0.149569088616857
          2007  0.036105898117288
          2008 -0.174874386138641
```

Para um valor tão baixo, de aproximadamente -0.20 , pode-se concluir que as variáveis de renda e incidência da dengue de fato não apresentam uma associação, reforçando a conclusão extraída a partir dos gráficos.

Análise Antiga - Microdados "crus" da PNADc

Primeiramente, execute o código abaixo a fim de produzir os resultados utilizados para gerar e exibir os gráficos.

```
In [6]: # Basic setup
        year.periods = list(c(2001:2003), c(2004:2008))
        states = c("Rondonia", "Acre", "Amazonas", "Roraima", "Para",
                   "Amazonas", "Mato Grosso", "Mato Grosso do Sul", "Piauí", "Ceará")
```



```

        Amapa , Tocantins , Roraima , Rondonia , Ceara ,
        "Rio Grande do Norte", "Paraiba", "Pernambuco", "Alagoa
s",
        "Sergipe", "Bahia", "Minas Gerais", "Espirito Santo",
        "Rio de Janeiro", "Sao Paulo", "Parana", "Santa Catarin
a",
        "Rio Grande do Sul", "Mato Grosso do Sul", "Mato Grosso
",
        "Goias", "Distrito Federal");
for (years in year.periods) {
  # Ignora os cálculos de renda caso já exista um arquivo com tais r
esultados
  period.begin = years[1]
  period.end = years[length(years)]
  results.data.file = paste(base.dir, '/results_', period.begin, '_'
, period.end, '.RData', sep='');

  if (!file.exists(results.data.file)) {
    results = data.frame(row.names = states) # data frame com rend
a média e taxa de incidência de dengue por UF

    for (year in years) {
      file.header <- paste(base.dir, year, '', sep='/')
      file.path <- paste(file.header, 'Dados/DOM', year, '.txt',
sep='')

      print(paste("Processando os dados para o ano ", year, "...",
sep=''))

      if(year >= 2007) {
        # Para as PNADc de 2007 e 2008 já são disponibilizados dat
a frames com os dados
        dic <- load.rdata.into.obj(paste(file.header, '/Leitura em
R/dicPNAD',
                                     year, '.RData', sep=''))
        pnad.data <- read.fwf(file.path, widths = dic$tamanho)
        # rename the columns to the standard spec
        names(pnad.data) <- dic$cod
      }
      else {
        # Para os demais anos há um arquivo do tipo SAS na sub-pas
ta `Input` de cada
        # ano que contém a descrição das variáveis do dataset. Bas
ta então ler o dataset
        # diretamente usando a biblioteca de SAS carregada no Setu
p.
        pnad.data <- noW(read.SAScii(file.path,
                                     paste(file.header, 'Input/input d
om', year, '.txt', sep=''),
                                     intervals.to.print = 10000,
bufferize = 10000))
      }

      # Agregação dos dados ao df `results`
      # Calcula a renda per capita segundo UF, ignorando _missing
data_

```

```

    # Antes de 2004 a renda per capita por domicílio segundo UF não
    # é reportada diretamente
    # no dataset. Logo, a renda mensal domiciliar é usada.
    income.code = 'V4621'
    if (year < 2004) {
      income.code = 'V4614'
    }

    # Algumas entidades apresentam uma renda per capita ou mensa
    # l muito elevada,
    # o que acaba ocasionando uma média bastante desviada para
    # a direita (right skew).
    # Sendo assim, valores extremamente elevados são ignorados
    # por distorcerem os resultados.
    valid.data = pnad.data[income.code] < 9000000000000
    avg.income = tapply(pnad.data[income.code][valid.data],
                        pnad.data$UF[valid.data], mean, na.rm=TR
UE)

    names(avg.income) <- states

    results <- cbind(results, avg.income)

    # Remove as variáveis usadas para o ano atual
    remove(pnad.data);
    gc();
  }

  # Renomeia as colunas segundo o formato `<year>.income`.
  names(results) = lapply(years, paste, '.income', sep='')
  save(results, file=results.data.file)
} else {
  results = load.rdata.into.obj(results.data.file)
}

years.under.analysis = years
names(results) = years.under.analysis

# now, read the dengue incidence rate dataset
dengue = read.csv(paste(base.dir, '/dengueUF.csv', sep=''), dec=",
", na.strings = '-')
dengue.cols = sapply('X', paste, years.under.analysis, sep='',
USE.NAMES = FALSE)
dengue.subset = dengue[1:27, dengue.cols]

dengue.new.cols = vector(mode="character", length=length(dengue.c
ols))
for (i in 1:length(names(dengue.subset))) {
  dengue.new.cols[i] = substr(names(dengue.subset)[i], 2, 5)
}

names(dengue.subset) = dengue.new.cols
row.names(dengue.subset) = dengue$UF[1:27]

# Calculate the cross-UF correlation. This tells us whether higher
income states

```

```

# have a relatively lower dengue incidence rate when compared to o
thers
cross.state.cor = diag(cor(results, dengue.subset))
names(cross.state.cor) = years.under.analysis

# Calculate the intra-UF correlation. This tells us whether higher
income in a
# state is associated with a lower dengue incidence rate **within*
* that state.
# In this case, the state is evaluated on its own.
A <- as.matrix(results)
B <- as.matrix(dengue.subset)
intra.state.cor = sapply(seq.int(dim(A)[1]), function(i) cor(A[i,]
, B[i,]))
names(intra.state.cor) = dengue$UF[1:27]
}

```

A análise foi feita com base nos microdados do PNAD disponíveis na página do IBGE [1] e nos dados de incidência da dengue disponíveis no GitHub do curso. Devido a ausência de dados para os anos de 1990 a 2000 no site do IBGE, a análise que segue se limitou aos anos de 2001 a 2008. Além disso, em razão da ausência da variável de renda anual per capita até o ano de 2003, os anos de 2001 a 2003 foram analisados a parte, sendo utilizada a variável de renda mensal domiciliar para efeitos de noção de renda. É importante destacar que nestes anos certas regiões desfavorecidas não foram abrangidas pelo PNAD. [2]

Um outro problema a ser apontado é o elevado valor encontrado para as rendas, em ambos os casos. Certamente, há o peso de outliers que levam a distribuição de renda média a ser *right-skewed*, mas creio haver um problema no pré-processamento dos dados. Em todo caso, os valores calculados estão com as proporções relativas esperadas, similares a fontes confiáveis que publicam tabelas com os dados de renda já processados. [3] Além disto, faltou ser realizado o cálculo de valor presente da renda com base na inflação para ajustá-la a um ano base. Apesar destes problemas com o pré-processamento dos dados, optei por fazer a análise toda em R por acreditar que este fosse o espírito do exercício. Em resumo, não foi encontrada, para o período de 2001 a 2008, correlação entre incidência de dengue e renda nas UFs. Abaixo apresenta-se uma análise detalhada.

Para o período de 2004 a 2008 foi calculada uma correlação média de aproximadamente -0,22 entre a renda per capita na UF e incidência de dengue nesta relativa as demais. Para um valor tão baixo, pode-se dizer que as variáveis de renda e incidência da dengue não apresentam correlação. Abaixo pode-se observar os valores para cada anos:

```

In [7]: printf("Média da correlação inter-estadual: %f", mean(cross.state.cor)
)
cross.state.cor

```

```
[1] "Média da correlação inter-estadual: -0.217749"
```

```

Out[7]:
2004 -0.283522307887389
2005 -0.396186297453112
2006 -0.227995158468646
2007 0.0055382245471229

```

2008 -0.186577084863445

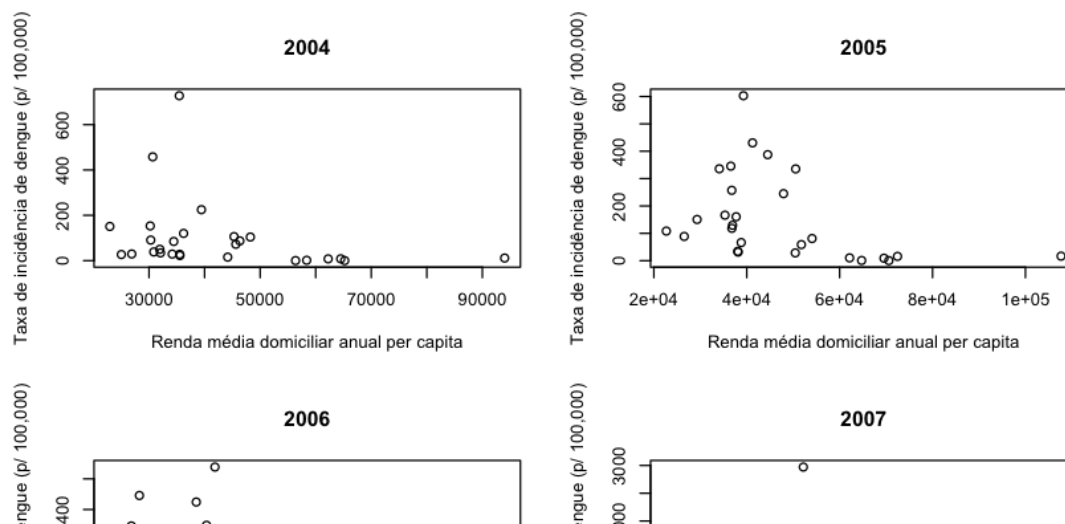
Além disso, para ficar clara a visualização deste resultado, pode-se observar os scatterplots por ano, que demonstram a ausência de associação entre as variáveis durante todos os anos sob estudo.

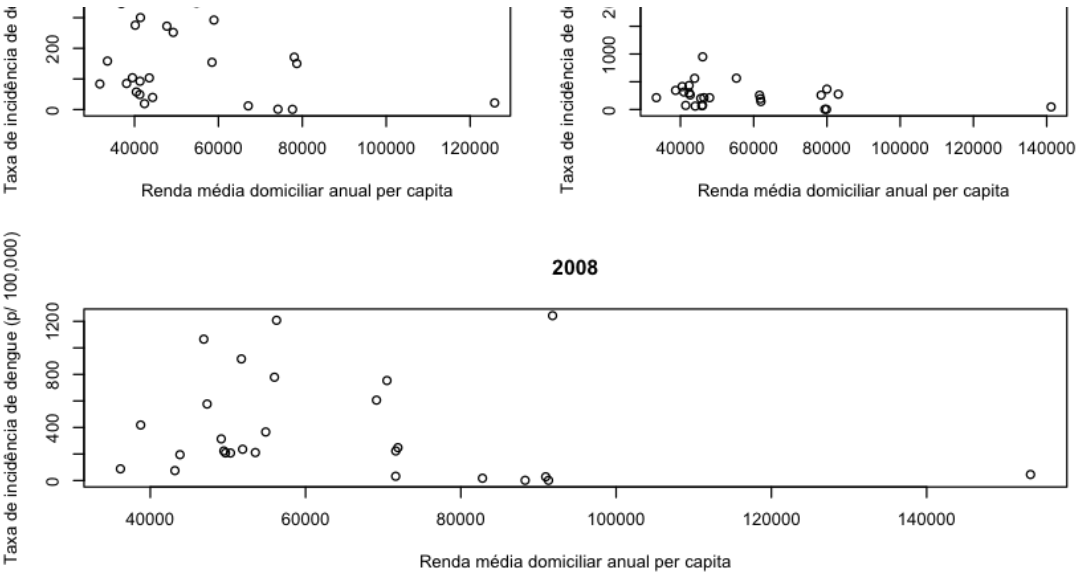
```
In [8]: cross.title = paste('Correlação interestadual de renda e dengue de',
                             years.under.analysis[1], 'a', tail(years.under.analysis, n=1))
intra.title = paste('Correlação intraestadual de renda e dengue de',
                    years.under.analysis[1], 'a', tail(years.under.analysis, n=1))

layout(matrix(c(1,2,3,4,5,5), 3, 2, byrow = TRUE))
xlab.str = "Renda média domiciliar anual per capita"

for (i in 1:length(years.under.analysis)) {
  year = as.character(years.under.analysis[i])
  noW(plot(results[,year], dengue.subset[,year], ylab="Taxa de incidência de dengue (p/ 100,000)",
           xlab=xlab.str, main = year))
}

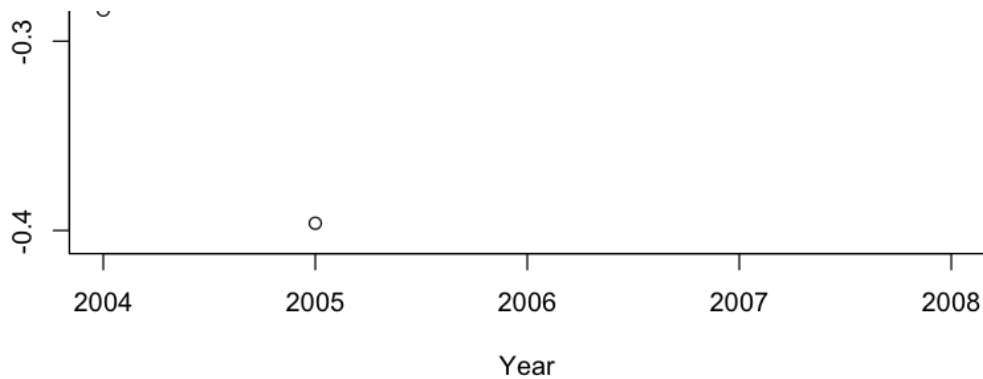
par(mfrow=c(1,1), par(xpd=T))
noW(plot(cross.state.cor, xlab="Year", ylab="Correlação", xaxt = "n",
main=cross.title))
axis(side=1, at=1:length(cross.state.cor), labels=names(cross.state.cor))
```





Correlação interestadual de renda e dengue de 2004 a 2008





Contudo, procurei determinar se obteria um resultado diferente caso analisasse cada UF isoladamente. Afinal, na análise anterior avalia-se o conjunto das UFs por ano, obtendo-se assim um dado que relativiza uma UF as demais; por outro lado, entender o comportamento isolado de cada UF é interessante. O curioso é que, neste caso, observa-se uma correlação positiva, em muitos casos forte, entre renda anual per capita e incidência de dengue. Embora não se possa estabelecer uma relação causal, há uma clara associação entre as variáveis para a maioria das UFs, quando vistas isoladamente:

```
In [9]: printf("Média da correlação intra-estadual: %f", mean(intra.state.cor)
)
intra.state.cor

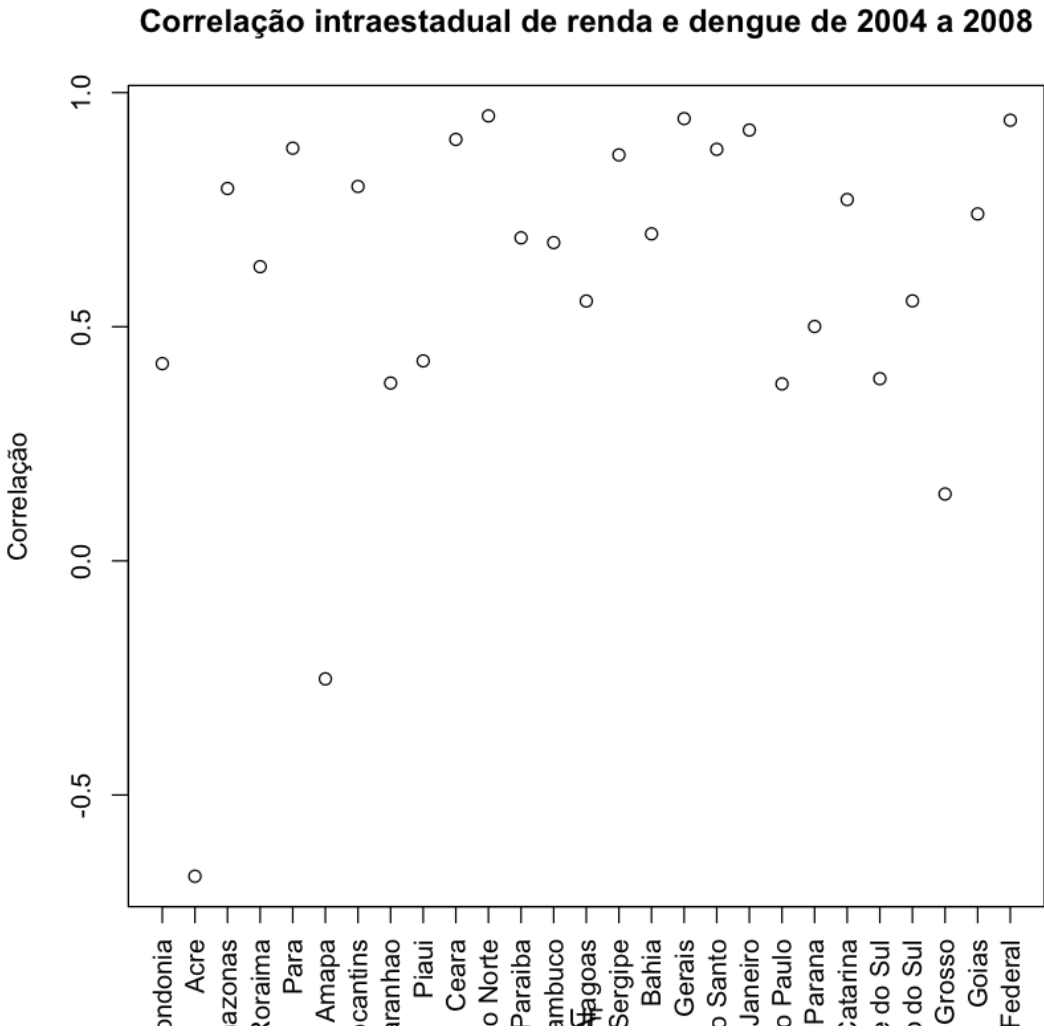
noW(plot(intra.state.cor, xlab="UF", ylab="Correlação", xaxt = "n",
main=intra.title))
axis(side=1, at=1:length(intra.state.cor), labels=names(intra.state.co
r), las=2)

[1] "Média da correlação intra-estadual: 0.589229"
```

```
Out[9]:
```

Rondonia	0.421243646967899
Acre	-0.673461070724564
Amazonas	0.795207995194116
Roraima	0.62827876014297
Para	0.881278649023513
Amapa	-0.251991676635566
Tocantins	0.799584708885441
Maranhao	0.379710144856956
Piaui	0.427084360888044
Ceara	0.90002819779847
Rio Grande do Norte	0.950216905102666
Paraiba	0.689856211393912
Pernambuco	0.679686210621467
Alagoas	0.554865995348426

Sergipe	0.866918282281269
Bahia	0.6983337919176
Minas Gerais	0.944427926816881
Espirito Santo	0.878869263289449
Rio de Janeiro	0.920013388168327
Sao Paulo	0.377953543041061
Parana	0.50055070574889
Santa Catarina	0.771545927918804
Rio Grande do Sul	0.389107238206488
Mato Grosso do Sul	0.5552415377209
Mato Grosso	0.142776720850559
Goias	0.740865268857545
Distrito Federal	0.940984840712227



R² An² F T_c M_e le d erni , nas jiri de Sa ta C inde ssc ato rito

Isto não é necessariamente contra-intuitivo: uma possível hipótese é a de que um UF mais rico pode empregar mais recursos no combate a doença, mas o crescimento desenfreado urbano e maior produção de resíduos podem levar a um aumento da doença no UF apesar disso. Outra hipótese seria de que, conforme as UF ficaram mais ricas, elas foram capazes de melhor medir casos de incidência de dengue. Novamente, é preciso enfatizar que embora levante tais hipóteses, este estudo não as afirma nem as explora, limitando-se a expor apenas a associação observada entre dengue e renda no caso intra-estadual.

Além disso, vale destacar os casos do Acre e Amapá: como pode ser observado nos dados acima, ambos apresentam uma correlação negativa entre a renda e incidência de dengue, sendo esta bastante forte no caso do Acre. Esta descoberta poderia motivar um estudo que procure entender a fundo o motivo de tal discrepância.

Rapidamente, façamos uma análise dos anos de 2001-2003. Para estes anos o PNAD não publicou uma variável de renda anual per capita, sendo necessário o uso da variável de renda mensal domiciliar, não ajustada ao número de integrantes por domicílio. No caso geral, encontrou-se uma correlação negativa bem fraca entre as variáveis, podendo-se dizer que elas não são correlacionadas.

```
In [10]: # Bootstrapping das variáveis para 2001->2003 (não o método estatístico)
years.under.analysis = c(2001:2003)
period.begin = years.under.analysis[1]
period.end = years.under.analysis[length(years.under.analysis)]
results.data.file = paste(base.dir, '/results_', period.begin, '_',
period.end, '.RData', sep='');
results = load.rdata.into.obj(results.data.file)
names(results) = years.under.analysis

dengue = read.csv(paste(base.dir, '/dengueUF.csv', sep=''), dec="," ,
na.strings = '-')
dengue.cols = sapply('X', paste, years.under.analysis, sep=' ',
USE.NAMES = FALSE)
dengue.subset = dengue[1:27, dengue.cols]

dengue.new.cols = vector(mode="character", length =length(dengue.cols)
\
```



```

',
for (i in 1:length(names(dengue.subset))) {
  dengue.new.cols[i] = substr(names(dengue.subset)[i], 2, 5)
}
names(dengue.subset) = dengue.new.cols
row.names(dengue.subset) = dengue$UF[1:27]

cross.state.cor = diag(cor(results, dengue.subset))
names(cross.state.cor) = years.under.analysis

A <- as.matrix(results)
B <- as.matrix(dengue.subset)

intra.state.cor = sapply(seq.int(dim(A)[1]), function(i) cor(A[i,], B[i,]))
names(intra.state.cor) = dengue$UF[1:27]

# Código relevante
printf("Média da correlação inter-estadual: %f", mean(cross.state.cor)
)
cross.state.cor

cross.title = paste('Correlação interestadual de renda e dengue de',
                    years.under.analysis[1], 'a', tail(years.under.analysis, n=1))
intra.title = paste('Correlação intraestadual de renda e dengue de',
                    years.under.analysis[1], 'a', tail(years.under.analysis, n=1))

layout(matrix(c(1,2,3,3), 2, 2, byrow = TRUE))
xlab.str = "Renda média domiciliar mensal"

for (i in 1:length(years.under.analysis)) {
  year = as.character(years.under.analysis[i])
  noW(plot(results[,year], dengue.subset[,year], ylab="Taxa de incidência de dengue (p/ 100,000)",
          xlab=xlab.str, main = year))
}

par(mfrow=c(1,1), par(xpd=T))
noW(plot(cross.state.cor, xlab="Year", ylab="Correlação", xaxt = "n",
main=cross.title))
axis(side=1, at=1:length(cross.state.cor), labels=names(cross.state.cor))

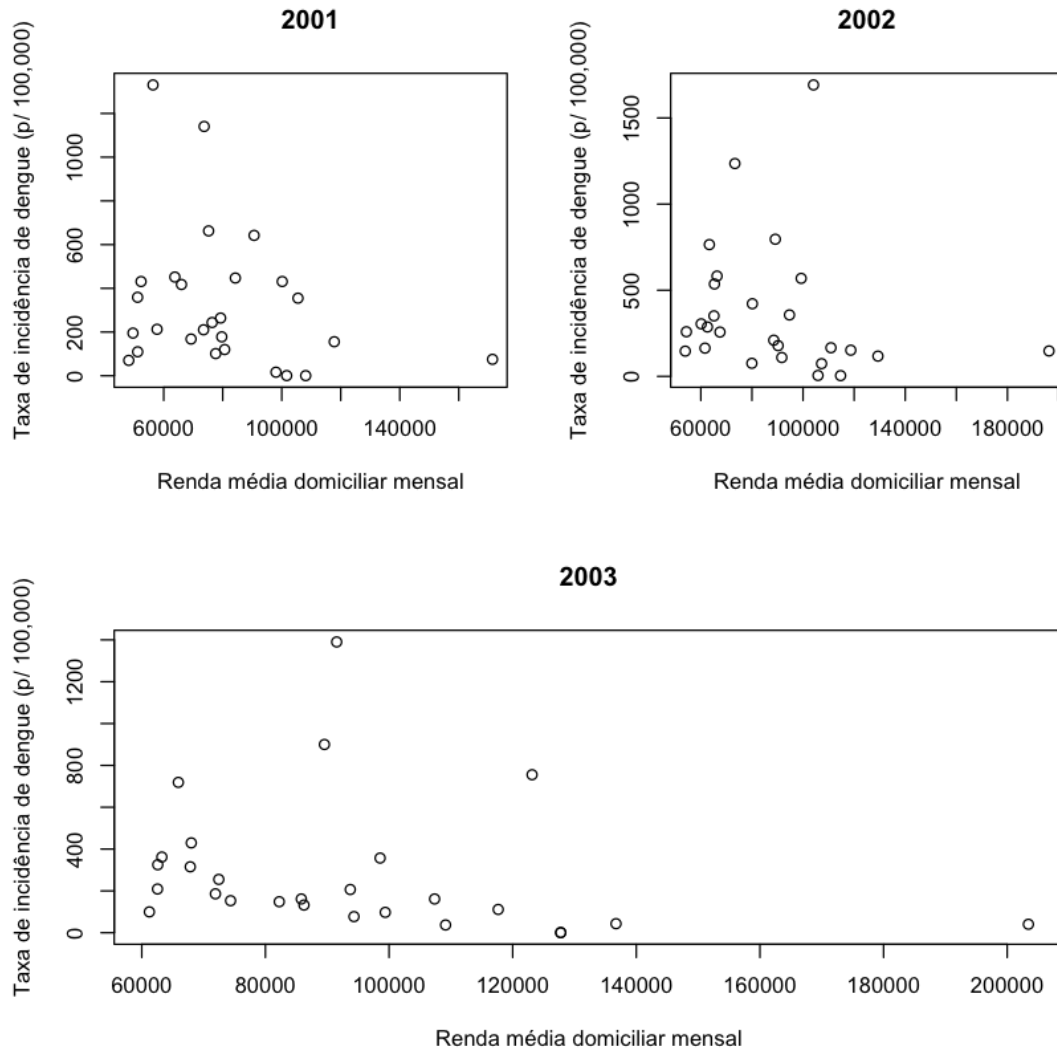
```

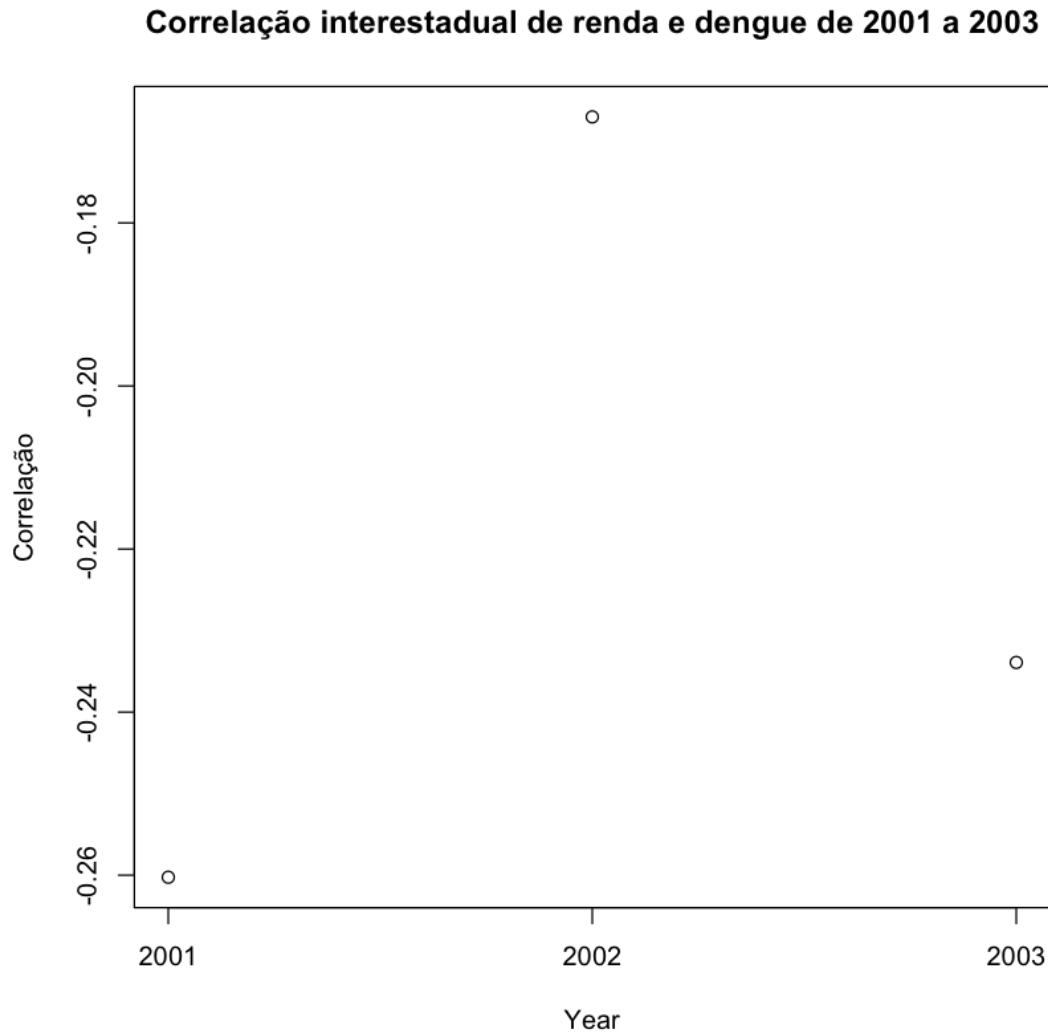
```
[1] "Média da correlação inter-estadual: -0.220389"
```

```

Out[10]:
2001 -0.260253989434337
2002 -0.167002415247218
2003 -0.233912085538855

```





Para o caso intra-estado, nota-se uma significativa diferença frente ao período de 2004 a 2008, analisado anteriormente. Há uma altíssima variabilidade entre os resultados por estado, de modo que a média é praticamente nula.

```
In [11]: printf("Média da correlação intra-estadual: %f", mean(intra.state.cor)
)
intra.state.cor

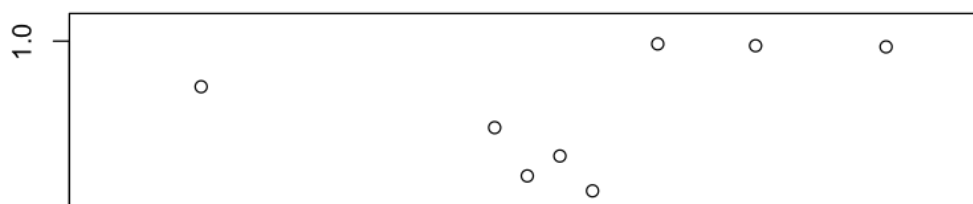
noW(plot(intra.state.cor, xlab="UF", ylab="Correlação", xaxt = "n",
main=intra.title))
axis(side=1, at=1:length(intra.state.cor), labels=names(intra.state.co
r), las=2)
```

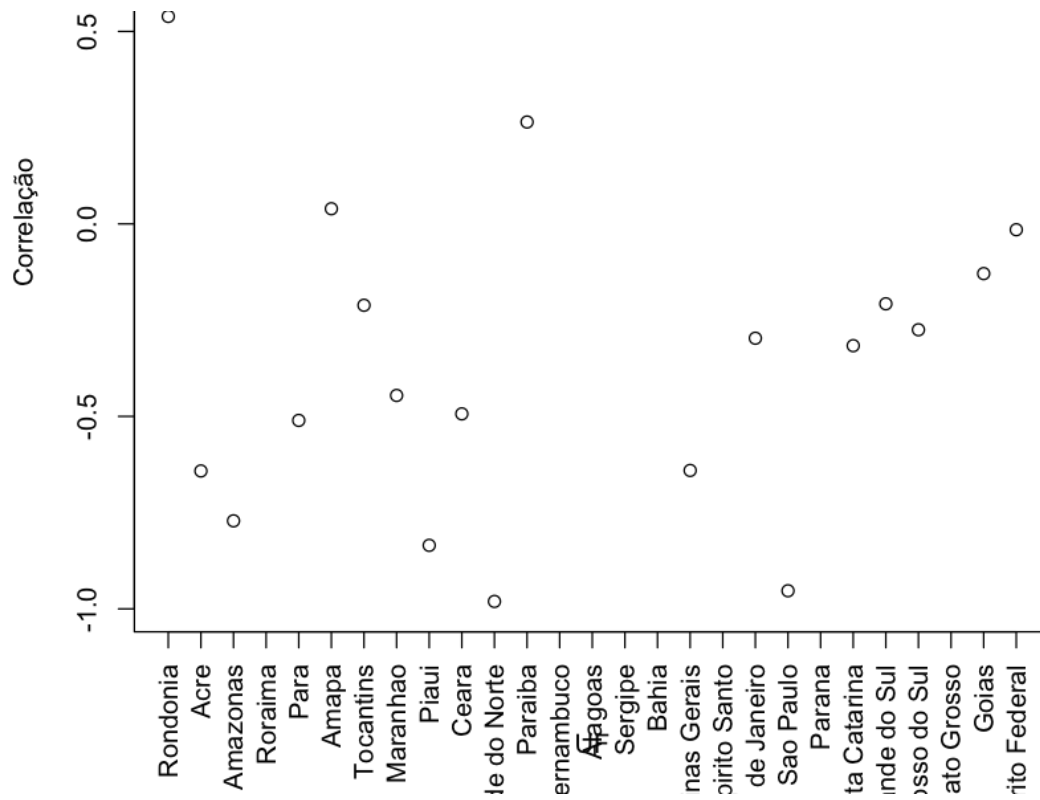
```
[1] "Média da correlação intra-estadual: -0.010882"
```

```
Out[11]:
          Rondonia  0.538976825765701
          Acre      -0.641774789838603
          ...       ...
```

Amazonas	-0.771362175622402
Roraima	0.881641527884426
Para	-0.510435881236817
Amapa	0.0394363051044103
Tocantins	-0.211309800499015
Maranhao	-0.445619190654787
Piaui	-0.834995627685089
Ceara	-0.493438413408773
Rio Grande do Norte	-0.980693887127669
Paraiba	0.264683505267009
Pernambuco	0.775466494010279
Alagoas	0.650129850218746
Sergipe	0.701831416846076
Bahia	0.611266864997467
Minas Gerais	-0.640450873024803
Espirito Santo	0.992699370520271
Rio de Janeiro	-0.29696216854196
Sao Paulo	-0.952991827105752
Parana	0.988057751190195
Santa Catarina	-0.316324677936931
Rio Grande do Sul	-0.207538669735278
Mato Grosso do Sul	-0.274967888467686
Mato Grosso	0.98508605073035
Goiias	-0.129114048594767
Distrito Federal	-0.0150981384985202

Correlação intraestadual de renda e dengue de 2001 a 2003





Referências

- [1] <http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2014/microdados.shtm>
(<http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2014/microdados.shtm>)
- [2] http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2013/default_sintese.shtm
(http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2013/default_sintese.shtm)
- [3] <http://tabnet.datasus.gov.br/cgi/tabnet.exe?idb2012/b08c.def>
(<http://tabnet.datasus.gov.br/cgi/tabnet.exe?idb2012/b08c.def>)