

Lab02 - Aula 2

Vanderlin Júnior - Curso Verão 2016 - R

25 de fevereiro de 2016

Aula 02 - Laboratório - Curso R

Ler o arquivo arq.txt que se encontra na pasta “/curso-r-2016/arq.txt”

```
arq <- 'arq.txt'
dados <- read.table(file = arq, header=TRUE, fill = TRUE)
dados <- read.table(file = arq, sep = ";")
str(dados)

## 'data.frame': 101 obs. of 3 variables:
## $ V1: Factor w/ 11 levels "-0,0907448251260999",...: 11 3 10 6 2 1 4 5 9 8 ...
## $ V2: Factor w/ 101 levels "0,00994513742625713",...: 101 27 84 32 76 37 3 2 79 40 ...
## $ V3: Factor w/ 4 levels "amarelo","azul",...: 3 2 4 1 2 1 1 1 1 1 ...

dados <- read.table(file = arq, sep = ";", dec = ",")
str(dados)

## 'data.frame': 101 obs. of 3 variables:
## $ V1: Factor w/ 11 levels "-0,0907448251260999",...: 11 3 10 6 2 1 4 5 9 8 ...
## $ V2: Factor w/ 101 levels "0,00994513742625713",...: 101 27 84 32 76 37 3 2 79 40 ...
## $ V3: Factor w/ 4 levels "amarelo","azul",...: 3 2 4 1 2 1 1 1 1 1 ...

dados <- read.table(file = arq, sep = ";", dec = ",", header = T)
str(dados)

## 'data.frame': 100 obs. of 3 variables:
## $ Aleatorio : num -0.7158 1.9053 0.8586 -0.7131 -0.0907 ...
## $ aleatorio2: num 0.297 0.876 0.324 0.824 0.379 ...
## $ cor       : Factor w/ 3 levels "amarelo","azul",...: 2 3 1 2 1 1 1 1 3 ...

dados <- read.table(file = arq, sep = ";", dec = ",", header = T, stringsAsFactors = F)
str(dados)

## 'data.frame': 100 obs. of 3 variables:
## $ Aleatorio : num -0.7158 1.9053 0.8586 -0.7131 -0.0907 ...
## $ aleatorio2: num 0.297 0.876 0.324 0.824 0.379 ...
## $ cor       : chr "azul" "vermelho" "amarelo" "azul" ...

summary(dados)
```

```

##      Aleatorio      aleatorio2      cor
##  Min.   :-1.5865   Min.   :0.009945  Length:100
##  1st Qu.:-0.7131   1st Qu.:0.292952  Class  :character
##  Median : 0.5795   Median :0.577933  Mode   :character
##  Mean   : 0.3675   Mean   :0.540947
##  3rd Qu.: 1.1027   3rd Qu.:0.823797
##  Max.   : 1.9053   Max.   :0.990911

```

Incluindo library's

Abrindo bibliotecas

```

library(readr)
library(dplyr)

```

```

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

```

```
library(httr)
```

Parte 2: descritiva básica

```

## Response [https://dl.dropboxusercontent.com/content_link/QKP9iJzDhjoV3StBzSQZwnB7VHd6Ipv0vxm442KuCPE
##   Date: 2016-02-25 20:46
##   Status: 200
##   Content-Type: application/octet-stream
##   Size: 419 kB
## <ON DISK>  C:\Users\Unifesp\AppData\Local\Temp\RtmpExZElz\file1edc73601f07

## [1] TRUE

##    municipio          ufn          ano          idhm
##  Length:16695    Length:16695    Min.   :1991   Min.   :0.1200
##  Class  :character  Class  :character  1st Qu.:1991   1st Qu.:0.4090
##  Mode   :character  Mode   :character  Median :2000   Median :0.5390
##                                         Mean   :2000   Mean   :0.5213
##                                         3rd Qu.:2010   3rd Qu.:0.6385
##                                         Max.   :2010   Max.   :0.8620
##    espvida        pesotot        gini        rdpc
##  Min.   :50.97   Min.   : 555   Min.   :0.2700   Min.   : 33.24
##  1st Qu.:64.83   1st Qu.: 5035  1st Qu.:0.4700   1st Qu.: 182.10
##  Median :69.19   Median : 10326  Median :0.5200   Median : 299.38
##  Mean   :68.41   Mean   : 30391  Mean   :0.5223   Mean   : 355.66
##  3rd Qu.:72.65   3rd Qu.: 21042  3rd Qu.:0.5700   3rd Qu.: 487.24

```

```

##   Max.    :78.64    Max.    :11253503    Max.    :0.9200    Max.    :2043.74
##   mort1      razdep      t_analf15m
##   Min.    : 8.49    Min.    :19.69    Min.    : 0.91
##   1st Qu.: 18.44    1st Qu.: 51.00    1st Qu.:11.32
##   Median  : 27.00    Median : 59.07    Median :19.20
##   Mean    : 33.27    Mean    : 62.98    Mean    :23.00
##   3rd Qu.: 41.62    3rd Qu.: 71.83    3rd Qu.:31.96
##   Max.    :120.12    Max.    :137.21    Max.    :88.39

## Source: local data frame [16,695 x 10]
##
##           municipio      ufn      ano     idhm espvida pesotot     gini     rdpc
##           (chr)      (chr)    (dbl)    (dbl)    (dbl)    (dbl)    (dbl)    (dbl)
## 1  ALTA FLORESTA D'OESTE Rondônia 1991 0.329  62.01  23546  0.63 198.46
## 2          ARIQUEMES Rondônia 1991 0.432  66.02  56812  0.57 319.47
## 3          CABIXI Rondônia 1991 0.309  63.16  7676  0.70 116.38
## 4          CACOAL Rondônia 1991 0.407  65.03  69725  0.66 320.24
## 5          CEREJEIRAS Rondônia 1991 0.386  62.73  19371  0.60 240.10
## 6 COLORADO DO OESTE Rondônia 1991 0.376  64.46  25664  0.62 224.82
## 7          CORUMBIARA Rondônia 1991 0.203  59.32  11897  0.59 81.38
## 8          COSTA MARQUES Rondônia 1991 0.425  62.76  7806  0.65 250.08
## 9          ESPIGÃO D'OESTE Rondônia 1991 0.388  64.18  23156  0.63 263.03
## 10         GUAJARÁ-MIRIM Rondônia 1991 0.468  64.71  32583  0.60 391.37
## ..
## ...
## Variables not shown: mort1 (dbl), razdep (dbl)

```

Quais são os seis municípios com os maiores IDH municipais em 2010.

```

pnud2 <- pnud[order(pnud$idhm,decreasing=TRUE),]

pnud2 %>%
  ##pnud2[order(pnud2$idhm,decreasing=TRUE),]  %>%
  filter(ano==2010)  %>%
  select(municipio,idhm)

```

```

## Source: local data frame [5,565 x 2]
##
##           municipio     idhm
##           (chr)      (dbl)
## 1  SÃO CAETANO DO SUL 0.862
## 2  ÁGUAS DE SÃO PEDRO 0.854
## 3          FLORIANÓPOLIS 0.847
## 4          VITÓRIA 0.845
## 5  BALNEÁRIO CAMBORIÚ 0.845
## 6          SANTOS 0.840
## 7          NITERÓI 0.837
## 8          JOAÇABA 0.827
## 9          BRASÍLIA 0.824
## 10         CURITIBA 0.823
## ..
## ...

```

Qual é a unidade federativa com menor expectativa de vida média,

```

pnud3 <- pnud[order(pnud$espvida, decreasing=FALSE),]
pnud3 %>%
  group_by(ano, ufn, espvida) %>%
  select(ano, ufn, espvida)

```

```

## Source: local data frame [16,695 x 3]
## Groups: ano, ufn, espvida [8561]
##
##     ano      ufn espvida
##   (dbl)    (chr)   (dbl)
## 1  1991 Alagoas   50.97
## 2  1991 Piauí    51.18
## 3  1991 Paraíba  51.26
## 4  1991 Bahia    51.32
## 5  1991 Bahia    51.32
## 6  1991 Bahia    51.32
## 7  1991 Piauí    51.42
## 8  1991 Pernambuco 51.47
## 9  1991 Bahia    51.48
## 10 1991 Pernambuco 51.61
## .. ...

```

Ponderada pela população dos municípios em 2000.

```

pnud %>%
  filter(ano==2000) %>%
  group_by(ufn) %>%
  summarise(n=n(),
            idhm_medio=mean(idhm),
            espvida_medio=mean(espvida),
            pesotot_medio=mean(pesotot),
            gini_medio=mean(gini),
            rdpc_medio=mean(rdpc),
            mort1_medio=mean(mort1),
            razdep_medio=mean(razdep),
            t_analf15m_medio=mean(t_analf15m)) %>%
  arrange(desc(idhm_medio)) %>% print(n=20)

```

```

## Source: local data frame [27 x 10]
##
##           ufn     n idhm_medio espvida_medio pesotot_medio
##   (chr) (int)       (dbl)        (dbl)        (dbl)
## 1 Distrito Federal 1  0.7250000  73.86000 2051146.000
## 2 São Paulo 645  0.6444760  71.95019  57414.581
## 3 Santa Catarina 293  0.6131160  72.52068 18281.102
## 4 Rio Grande do Sul 496  0.6051915  72.39167 20539.923
## 5 Rio de Janeiro 92  0.6010870  69.24304 156427.022
## 6 Paraná 399  0.5930150  70.04764 23968.571
## 7 Espírito Santo 78  0.5781795  69.89538 39708.141
## 8 Goiás 246  0.5571585  70.45508 20338.333
## 9 Mato Grosso do Sul 78  0.5511667  69.72269 26641.051
## 10 Minas Gerais 853  0.5478042  70.16120 20974.794

```

```

## 11         Mato Grosso    141  0.5391915    69.37624  17761.418
## 12             Amapá       16  0.4936250    67.21563 29814.500
## 13           Roraima      15  0.4934000    67.07600 21626.467
## 14        Rondônia      52  0.4748269    66.51596 26534.385
## 15 Rio Grande do Norte 167  0.4632515    65.78461 16617.090
## 16        Tocantins     139  0.4569424    65.39763  8324.446
## 17            Ceará      184  0.4490435    66.20391 40384.060
## 18       Pernambuco     185  0.4458486    64.95659 42803.270
## 19          Sergipe      75  0.4410267    63.93747 23793.013
## 20            Pará      143  0.4322238    67.48657 43302.839
## ...
## Variables not shown: gini_medio (dbl), rdpc_medio (dbl), mort1_medio
##   (dbl), razdep_medio (dbl), t_analf15m_medio (dbl)

```

Quais são os municípios outliers com relação ao índice de Gini em 1991 e em 2010 (Dica: utilize como critério para determinar outliers valores maiores que a média mais duas vezes o desvio padrão).

```

summary(pnud$gini)

##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
## 0.2700 0.4700 0.5200 0.5223 0.5700 0.9200

##sqrt(pnud$gini)*2

pnud %>%
  select(ano, ufn, municipio, gini) %>%
  filter(ufn=='São Paulo', ano==1991, (gini < summary(pnud$gini)[2] | gini > summary(pnud$gini)[5]))


## Source: local data frame [258 x 4]
##
##      ano      ufn      municipio  gini
##      (dbl)    (chr)    (chr)    (dbl)
## 1 1991 São Paulo      ADOLFO  0.41
## 2 1991 São Paulo ÁGUAS DE LINDÓIA 0.59
## 3 1991 São Paulo      ALTAIR  0.42
## 4 1991 São Paulo      ALUMÍNIO 0.39
## 5 1991 São Paulo ÁLVARO DE CARVALHO 0.45
## 6 1991 São Paulo      ALVINLÂNDIA 0.45
## 7 1991 São Paulo AMÉRICO BRASILIENSE 0.34
## 8 1991 São Paulo      ANHUMAS  0.45
## 9 1991 São Paulo      ARAMINA  0.46
## 10 1991 São Paulo     ARCO-ÍRIS 0.77
## ...

pnud %>%
  select(ano, ufn, municipio, gini) %>%
  filter(ano==1991, (gini < summary(pnud$gini)[2] | gini > summary(pnud$gini)[5]))


## Source: local data frame [2,357 x 4]
##
##      ano      ufn      municipio  gini
##      (dbl)    (chr)    (chr)    (dbl)

```

```

## 1 1991 Rondônia ALTA FLORESTA D'OESTE 0.63
## 2 1991 Rondônia CABIXI 0.70
## 3 1991 Rondônia CACOAL 0.66
## 4 1991 Rondônia CEREJEIRAS 0.60
## 5 1991 Rondônia COLORADO DO OESTE 0.62
## 6 1991 Rondônia CORUMBIARA 0.59
## 7 1991 Rondônia COSTA MARQUES 0.65
## 8 1991 Rondônia ESPIGÃO D'OESTE 0.63
## 9 1991 Rondônia GUAJARÁ-MIRIM 0.60
## 10 1991 Rondônia MACHADINHO D'OESTE 0.60
## ... ... ...
pnud %>%
  select(ano, ufn, municipio, gini) %>%
  filter(ano==2010, (gini < summary(pnud$gini)[2] | gini > summary(pnud$gini)[5]))


## Source: local data frame [2,405 x 4]
##
##    ano      ufn          municipio   gini
##    (dbl)    (chr)        (chr)     (dbl)
## 1 2010 Rondônia ALTA FLORESTA D'OESTE 0.58
## 2 2010 Rondônia RIO CRESPO 0.58
## 3 2010 Rondônia SÃO MIGUEL DO GUAPORÉ 0.62
## 4 2010 Rondônia CAMPO NOVO DE RONDÔNIA 0.67
## 5 2010 Rondônia MIRANTE DA SERRA 0.60
## 6 2010 Rondônia SERINGUEIRAS 0.63
## 7 2010 Rondônia TEIXEIRÓPOLIS 0.43
## 8 2010 Rondônia VALE DO PARAÍSO 0.59
## 9 2010 Acre ASSIS BRASIL 0.61
## 10 2010 Acre BRASILÉIA 0.58
## ... ...
pnud %>%
  select(ano, ufn, municipio, gini) %>%
  filter(ano==1991, (gini > (summary(pnud$gini)[4]+(2*sd(pnud$gini)))))


## Source: local data frame [167 x 4]
##
##    ano      ufn          municipio   gini
##    (dbl)    (chr)        (chr)     (dbl)
## 1 1991 Rondônia CABIXI 0.70
## 2 1991 Rondônia CANDEIAS DO JAMARI 0.71
## 3 1991 Rondônia GOVERNADOR JORGE TEIXEIRA 0.70
## 4 1991 Rondônia MINISTRO ANDREAZZA 0.78
## 5 1991 Rondônia SÃO FRANCISCO DO GUAPORÉ 0.71
## 6 1991 Rondônia SERINGUEIRAS 0.69
## 7 1991 Rondônia VALE DO ANARI 0.75
## 8 1991 Amazonas AMATURÁ 0.70
## 9 1991 Amazonas ATALAIA DO NORTE 0.76
## 10 1991 Amazonas AUTAZES 0.70
## ... ...

```

```

pesquisa <- pnuod %>%
  select(ano, ufn, municipio, gini) %>%
  filter(ano==2010, (gini > (summary(pnuod$gini)[4]+(2*sd(pnuod$gini)))))

head(pesquisa, 61)

```

```

## Source: local data frame [61 x 4]
##
##   ano      ufn          municipio  gini
##   (dbl)    (chr)        (chr)    (dbl)
## 1 2010 Rondônia CAMPO NOVO DE RONDÔNIA 0.67
## 2 2010     Acre           FEIJÓ  0.68
## 3 2010     Acre          JORDÃO  0.71
## 4 2010     Acre SANTA ROSA DO PURUS  0.78
## 5 2010 Amazonas        BARCELOS 0.74
## 6 2010 Amazonas       GUAJARÁ  0.74
## 7 2010 Amazonas      ITAMARATI  0.80
## 8 2010 Amazonas     ITAPIRANGA 0.67
## 9 2010 Amazonas       JUTAÍ  0.69
## 10 2010 Amazonas      MARAÃ  0.69
## ...   ...

```