

Projeto EAD5754TIC - Pesquisa TIC 2017

Fabiano Castello e Gustavo Galeale

Setembro de 2018

Intro

Conciliação e junção das bases de dados [base de microdados TIC](#). Mais informações veja o [Repositório no Github]
(<https://github.com/fabianocastello/NIC.br.base.CETIC>)

Carregando Libs no R

```
## Loading required package: ggplot2
## Loading required package: 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
## Loading required package: readr
## Warning: package 'readr' was built under R version 3.3.3
## ggplot2    dplyr    readr
##   TRUE      TRUE    TRUE
```

Carregando e tratando DOMICILIOS

```
library("readr")
dom15 <- read_csv2(
  "../BasesMicrodadosTIC/ticdom_2015_domicilios_base_de_microdados_v1.0.csv"
)
## Using ',' as decimal and '.' as grouping mark. Use read_delim() for
## more control.
## Parsed with column specification:
## cols(
##   .default = col_integer(),
```

```

## PESO = col_double()
## )

## See spec(...) for full column specifications.
names(dom15) <- tolower(names(dom15))

dom16 <- read_csv2(
  "../BasesMicrodadosTIC/ticdom_2016_domicilios_base_de_microdados_v1.0.csv"
)

## Using ',' as decimal and '.' as grouping mark. Use read_delim() for
## more control.

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   Peso = col_double()
## )

## See spec(...) for full column specifications.
names(dom16) <- tolower(names(dom16))

dom17 <- read_csv2(
  "../BasesMicrodadosTIC/ticdom_2017_domicilios_base_de_microdados_v1.1.csv"
)

## Using ',' as decimal and '.' as grouping mark. Use read_delim() for
## more control.

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   PESO = col_double()
## )

## See spec(...) for full column specifications.
names(dom17) <- tolower(names(dom17))

```

Retirando variáveis que não são comuns as três bases

```

dim(dom15)

## [1] 23465    55

dom15$a6 <- NULL
dom15$energia_eletrica <- NULL
dim(dom15)

```

```
## [1] 23465    53

dim(dom16)

## [1] 23721    60

dom16$a10 <- NULL
dom16$a5_outro_desc <- NULL
dom16$a5a <- NULL
dom16$a6 <- NULL
dom16$ajuda <- NULL
dom16$energia_eletrica <- NULL
dom16$tipo_pesq <- NULL
dim(dom16)

## [1] 23721    53

dim(dom17)

## [1] 23592    65

dom17$a1_agreg <- NULL
dom17$a1_exclusivos <- NULL
dom17$a1a4 <- NULL
dom17$a2_a_faixa <- NULL
dom17$a2_b_faixa <- NULL
dom17$a2_c_faixa <- NULL
dom17$a5_nenhum <- NULL
dom17$a5_outro_desc <- NULL
dom17$a5a <- NULL
dom17$a7_agreg <- NULL
dom17$a9_faixa <- NULL
dom17$renda_familiar_2 <- NULL
dim(dom17)

## [1] 23592    53
```

Ajustando nomes das variáveis que existem nos três anos com nomes diferentes mas com mesma codificação

```
names(dom15)[names(dom15)=="classe_cb2008"] <- "classe_2008"
names(dom16)[names(dom16)=="classe_cb2008"] <- "classe_2008"
names(dom15)[names(dom15)=="classe_cb2015"] <- "classe_2015"
names(dom16)[names(dom16)=="classe_cb2015"] <- "classe_2015"
```

Ajustando codificação da região

```
#Codificação da região em 2015 e 2016
#1 = ""Norte""
#2 = ""Nordeste""
#3 = ""Sudeste""
#4 = ""Sul""
#5 = ""Centro-Oeste"
```

```

#Codificação da região em 2017
#1 = ""Sudeste""
#2 = ""Nordeste""
#3 = ""Sul""
#4 = ""Norte""
#5 = ""Centro-Oeste""
#Codificação da região no consolidado (base 2017)
#1 = ""Sudeste""
#2 = ""Nordeste""
#3 = ""Sul""
#4 = ""Norte""
#5 = ""Centro-Oeste""

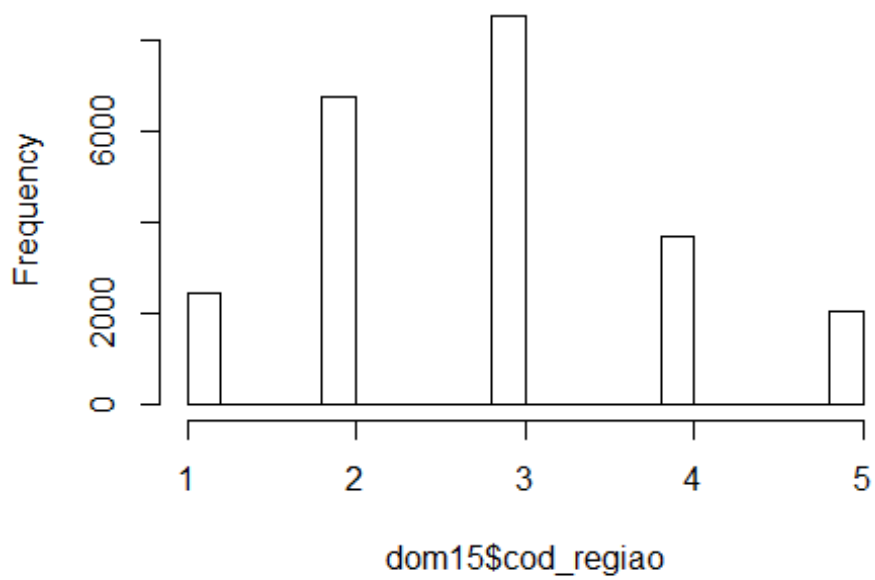
dom15["cod_regiao_3"] <- NA
dom15$cod_regiao_3[dom15$cod_regiao == 1] <- 4
dom15$cod_regiao_3[dom15$cod_regiao == 2] <- 2
dom15$cod_regiao_3[dom15$cod_regiao == 3] <- 1
dom15$cod_regiao_3[dom15$cod_regiao == 4] <- 3
dom15$cod_regiao_3[dom15$cod_regiao == 5] <- 5
dom16["cod_regiao_3"] <- NA
dom16$cod_regiao_3[dom16$cod_regiao == 1] <- 4
dom16$cod_regiao_3[dom16$cod_regiao == 2] <- 2
dom16$cod_regiao_3[dom16$cod_regiao == 3] <- 1
dom16$cod_regiao_3[dom16$cod_regiao == 4] <- 3
dom16$cod_regiao_3[dom16$cod_regiao == 5] <- 5
dom17["cod_regiao_3"] <- NA
dom17$cod_regiao_3[dom17$cod_regiao_2 == 1] <- 1
dom17$cod_regiao_3[dom17$cod_regiao_2 == 2] <- 2
dom17$cod_regiao_3[dom17$cod_regiao_2 == 3] <- 3
dom17$cod_regiao_3[dom17$cod_regiao_2 == 4] <- 4
dom17$cod_regiao_3[dom17$cod_regiao_2 == 5] <- 5

```

Double check antes de deletar variáveis de região

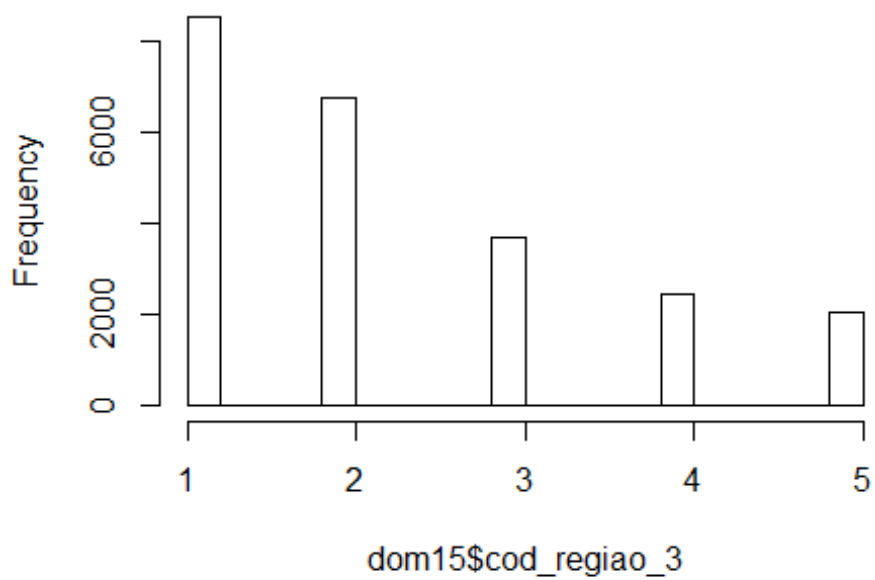
```
hist(dom15$cod_regiao)
```

Histogram of dom15\$cod_regiao

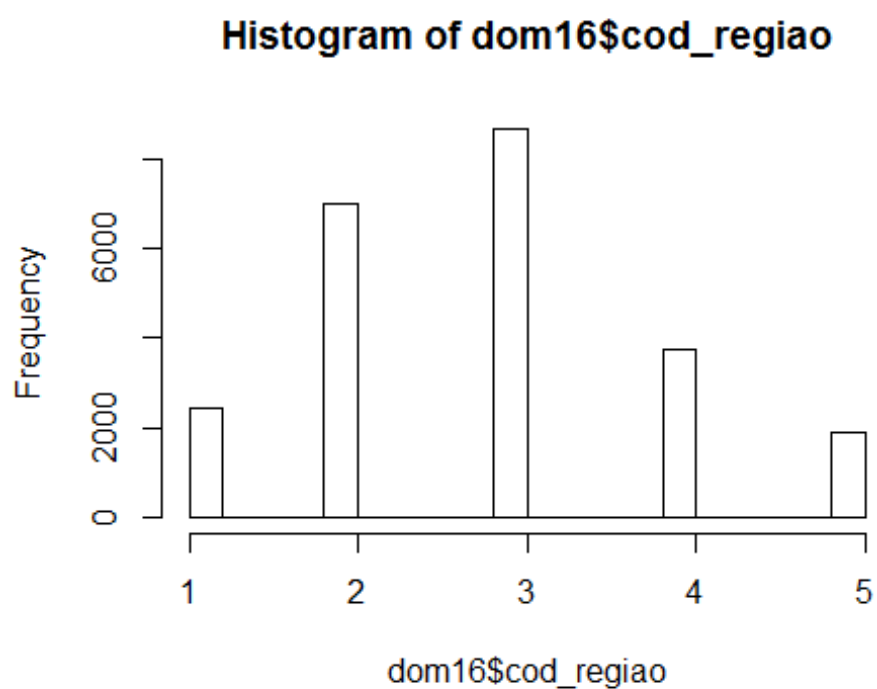


```
hist(dom15$cod_regiao_3)
```

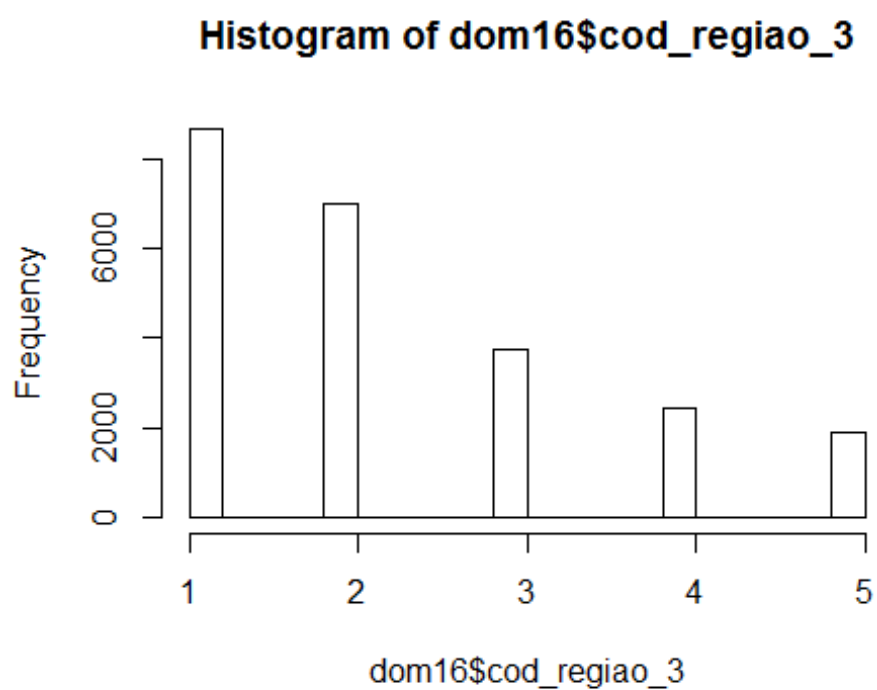
Histogram of dom15\$cod_regiao_3



```
hist(dom16$cod_regiao)
```

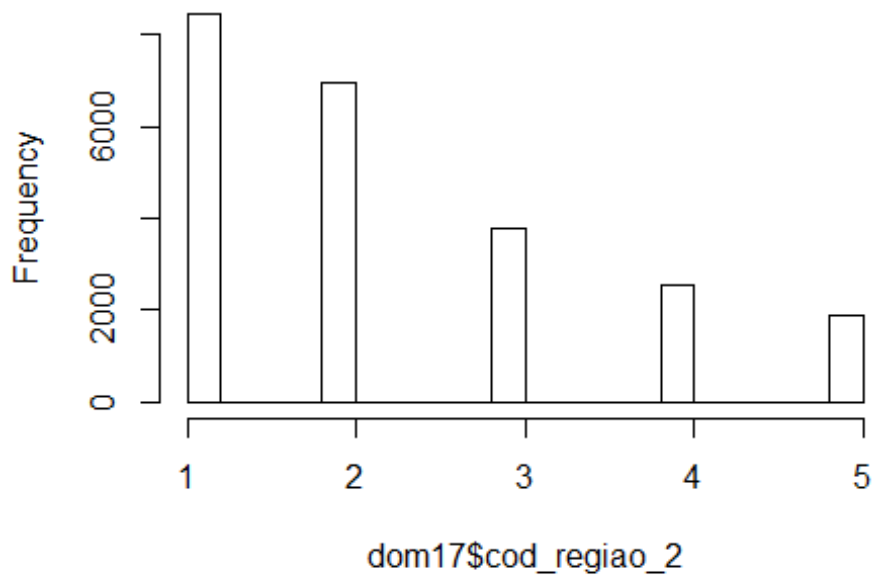


```
hist(dom16$cod_regiao_3)
```



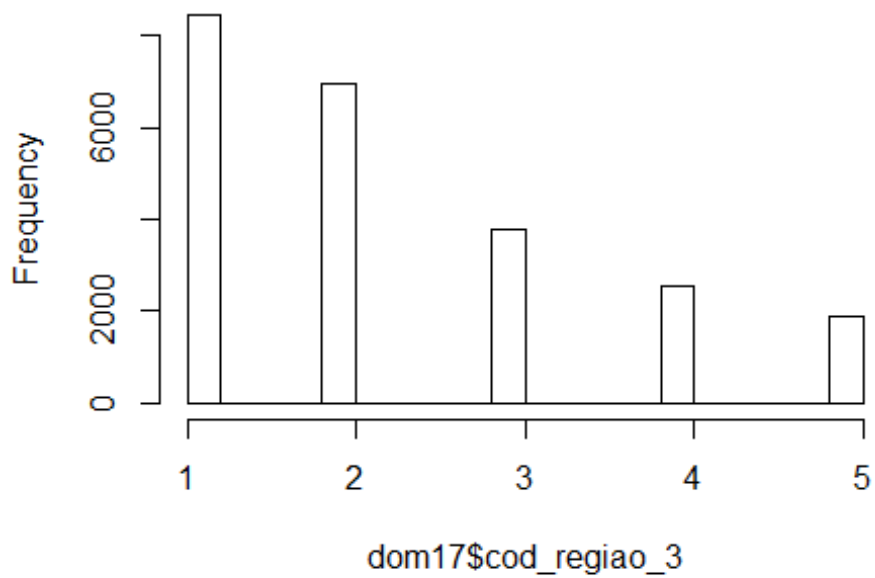
```
hist(dom17$cod_regiao_2)
```

Histogram of dom17\$cod_regiao_2



```
hist(dom17$cod_regiao_3)
```

Histogram of dom17\$cod_regiao_3



Eliminando variáveis de cod_região não mais utilizadas e comparando bases

```
dom15$cod_regiao <- NULL
dom16$cod_regiao <- NULL
dom17$cod_regiao_2 <- NULL

# Comparando as bases: se 15 = 16 e 16 = 17 então 15 = 17
sort(names(dom15)) == sort(names(dom16))

## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [15] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [29] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [43] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE

sort(names(dom16)) == sort(names(dom17))

## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [15] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [29] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [43] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Criando um campo para identificar o ano na pesquisa, agregando datasets e exportando

```
dom15$ano <- "2015"
dom16$ano <- "2016"
dom17$ano <- "2017"

dom <- rbind(dom15, dom16, dom17)

write.csv2(dom,
            "../BasesMicrodadosTIC/ticdom_base_domicilios.csv",
            append = FALSE)

## Warning in write.csv2(dom, "../BasesMicrodadosTIC/
## ticdom_base_domicilios.csv", : attempt to set 'append' ignored
```

Wrap-up e finalização de DOMICÍLIOS

```
rm(dom17)
rm(dom16)
rm(dom15)
```


Carregando e tratando INDIVÍDUOS

```
library("readr")

ind15 <- read_csv2(

  "../BasesMicrodadosTIC/ticdom_2015_individuos_base_de_microdados_v1.0.csv"
)

## Using ',' as decimal and '.' as grouping mark. Use read_delim() for
## more control.

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   Peso = col_double()
## )

## See spec(...) for full column specifications.
names(ind15) <- tolower(names(ind15))

ind16 <- read_csv2(

  "../BasesMicrodadosTIC/ticdom_2016_individuos_base_de_microdados_v1.0.csv"
)

## Using ',' as decimal and '.' as grouping mark. Use read_delim() for
## more control.

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   Peso = col_double()
## )

## See spec(...) for full column specifications.
names(ind16) <- tolower(names(ind16))

ind17 <- read_csv2(

  "../BasesMicrodadosTIC/ticdom_2017_individuos_base_de_microdados_v1.1.csv"
)

## Using ',' as decimal and '.' as grouping mark. Use read_delim() for
## more control.

## Parsed with column specification:
## cols(
##   .default = col_integer(),
```

```

## PESO = col_double()
## )

## See spec(...) for full column specifications.

names(ind17) <- tolower(names(ind17))

dim(ind15)

## [1] 20397 163

ind15$c4 <- NULL
ind15$c6_h <- NULL
ind15$c9_e <- NULL
ind15$h1 <- NULL
ind15$h3_a <- NULL
ind15$h3_b <- NULL
ind15$h3_c <- NULL
ind15$h3_d <- NULL
ind15$h3_e <- NULL
ind15$h3_f <- NULL
ind15$h3_g <- NULL
ind15$h3_h <- NULL
ind15$h3_i <- NULL
ind15$h3_j <- NULL
ind15$h3_outro <- NULL
ind15$h3_outro_desc <- NULL
ind15$h4 <- NULL
ind15$j2_o <- NULL
ind15$j3b_a <- NULL
ind15$j3b_b <- NULL
ind15$j3b_c <- NULL
ind15$j3b_d <- NULL
ind15$j3b_e <- NULL
ind15$j3b_f <- NULL
ind15$j3b_g <- NULL
ind15$j3b_h <- NULL
ind15$j3b_outro <- NULL
ind15$j3b_outro_desc <- NULL
ind15$j4 <- NULL
ind15$renda_pessoal <- NULL
dim(ind15)

## [1] 20397 133

dim(ind16)

## [1] 20722 178

ind16$c1_cob_a <- NULL
ind16$c1_cob_b <- NULL
ind16$c1_cob_c <- NULL

```

```

ind16$c1_cob_d <- NULL
ind16$c2a <- NULL
ind16$c3_cob <- NULL
ind16$c4 <- NULL
ind16$c4_cob <- NULL
ind16$c5_cob_a <- NULL
ind16$c5_cob_b <- NULL
ind16$c5_cob_c <- NULL
ind16$c5_cob_d <- NULL
ind16$c5_cob_e <- NULL
ind16$c5_cob_f <- NULL
ind16$c5_cob_g <- NULL
ind16$c5_cob_g_desc <- NULL
ind16$c6_cob_a <- NULL
ind16$c6_cob_b <- NULL
ind16$c6_cob_c <- NULL
ind16$c6_cob_d <- NULL
ind16$c6_cob_e <- NULL
ind16$c6_cob_f <- NULL
ind16$c6_cob_g <- NULL
ind16$c6_cob_h <- NULL
ind16$c6_cob_outro <- NULL
ind16$c6_cob_outro_desc <- NULL
ind16$c6_h <- NULL
ind16$c9_e <- NULL
ind16$h1 <- NULL
ind16$h3_a <- NULL
ind16$h3_b <- NULL
ind16$h3_c <- NULL
ind16$h3_d <- NULL
ind16$h3_e <- NULL
ind16$h3_f <- NULL
ind16$h3_g <- NULL
ind16$h3_h <- NULL
ind16$h3_i <- NULL
ind16$h3_j <- NULL
ind16$h3_outro <- NULL
ind16$h3_outro_desc <- NULL
ind16$h4 <- NULL
ind16$j4 <- NULL
ind16$renda_pessoal <- NULL
ind16$tipo_pesq <- NULL
dim(ind16)

## [1] 20722 133

dim(ind17)

## [1] 20490 180

```

```

ind17$c1_cob_a <- NULL
ind17$c1_cob_agreg <- NULL
ind17$c1_cob_b <- NULL
ind17$c1_cob_c <- NULL
ind17$c1_cob_d <- NULL
ind17$c12_e1 <- NULL
ind17$c12_f1 <- NULL
ind17$c2a <- NULL
ind17$c3_cob <- NULL
ind17$c3j3 <- NULL
ind17$c4 <- NULL
ind17$c4_cob <- NULL
ind17$c5_agreg <- NULL
ind17$c5_cob_a <- NULL
ind17$c5_cob_agreg <- NULL
ind17$c5_cob_b <- NULL
ind17$c5_cob_c <- NULL
ind17$c5_cob_d <- NULL
ind17$c5_cob_e <- NULL
ind17$c5_cob_f <- NULL
ind17$c5_cob_g <- NULL
ind17$c5_cob_g_desc <- NULL
ind17$c5_cob_nenhum <- NULL
ind17$c5_dispositivos <- NULL
ind17$c6_cob_a <- NULL
ind17$c6_cob_b <- NULL
ind17$c6_cob_c <- NULL
ind17$c6_cob_d <- NULL
ind17$c6_cob_e <- NULL
ind17$c6_cob_f <- NULL
ind17$c6_cob_g <- NULL
ind17$c6_cob_outro <- NULL
ind17$c6_cob_outro_desc <- NULL
ind17$c8_nao <- NULL
ind17$c9_f <- NULL
ind17$cod_regiao_2 <- NULL
ind17$g1_agreg <- NULL
ind17$g3_nenhum <- NULL
ind17$g3_outro <- NULL
ind17$g3_outro_desc <- NULL
ind17$i1_nenhum <- NULL
ind17$j2_nenhum <- NULL
ind17$j3_agreg <- NULL
ind17$j5_faixas <- NULL
ind17$renda_familiar <- NULL
ind17$renda_familiar_2 <- NULL
ind17$renda_pessoal <- NULL
dim(ind17)

```

```
## [1] 20490 133
```

```
#Ajustando nomes das variáveis que existem nos três anos com nomes diferentes  
# nada a ajustar!
```

```
# Comparando as bases: se 15 = 16 e 16 = 17 então 15 = 17  
sort(names(ind15)) == sort(names(ind16))
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [15] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [29] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [43] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [57] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [71] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [85] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [99] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [113] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [127] TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
sort(names(ind16)) == sort(names(ind17))
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [15] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [29] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [43] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [57] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [71] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [85] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [99] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [113] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE  
TRUE  
## [127] TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Criando um campo para identificar o ano na pesquisa, agregando datasets e exportando

```
ind15$ano <- "2015"
ind16$ano <- "2016"
ind17$ano <- "2017"

ind <- rbind(ind15, ind16, ind17)

write.csv2(ind,
            "../BasesMicrodadosTIC/ticdom_base_individuos.csv",
            append = FALSE)

## Warning in write.csv2(ind, "../BasesMicrodadosTIC/
## ticdom_base_individuos.csv", : attempt to set 'append' ignored
```

Wrap-up e finalização de INDIVÍDUOS

```
rm(ind17)
rm(ind16)
rm(ind15)
```

FIM DO MARKDOWN!