

tidy_data.R

diego

Sun Jun 18 14:46:21 2017

```
# load packages

library("tidyr")
library("dplyr")
library("ggplot2")

parse_csv <- function(filename){

  base <- gsub(x = filename,
              pattern = ".csv",
              replacement = "")

  fname <- strsplit(x = basename(base), split = "-")[[1]]

  ano <- fname[1]
  grupo <- fname[2]

  tabela <- read.table(file = filename,
                      header = F,
                      sep = ",",
                      skip = 16,
                      nrow = 33,
                      dec = ".",
                      na.strings = "-")

  tabela$ano <- ano
  tabela$grupo <- grupo

  colnames(tabela) <- c(
    "Região",
    "Código_UF",
    "UF",
    "0 dias nos últimos sete dias n",
    "0 dias nos últimos sete dias %",
    "1 dia nos últimos sete dias n",
    "1 dia nos últimos sete dias %",
    "2 dias nos últimos sete dias n",
    "2 dias nos últimos sete dias %",
    "3 dias nos últimos sete dias n",
    "3 dias nos últimos sete dias %",
    "4 dias nos últimos sete dias n",
    "4 dias nos últimos sete dias %",
    "5 dias nos últimos sete dias n",
    "5 dias nos últimos sete dias %",
    "6 dias nos últimos sete dias n",
    "6 dias nos últimos sete dias %",
    "Todos os últimos sete dias n",
```

```

    "Todos os últimos sete dias %",
    "Total de indivíduos",
    "ano",
    "grupo")

  tabela
}

# Get list of files
csv_dir <- "tabelas_sisvan"
lista_arquivos <- list.files(csv_dir, full.names = T, pattern = "*.csv")

# Read data
dados_arquivos <- lapply(X = lista_arquivos, parse_csv)

# Merge
merged_data <- do.call(what = rbind, args = dados_arquivos)

# Clean up
batata <- "batata frita,batata de pacote,salgado frito"
bisc <- "biscoito doce, doces, balas, chocolates"
bs <- "biscoito salgado, salgadinho de pacote"
hb <- "hamburguer, embutidos"
lg <- "legumes, verduras cozidas"
merged_data$grupo[ merged_data$grupo == "batatafrita,batatadepacote,salgadofrito" ] <- batata
merged_data$grupo[ merged_data$grupo == "batatafrita,batatadepacote,salgadofritos" ] <- batata
merged_data$grupo[ merged_data$grupo == "batatafrita,batatadepacote,salgadofritosAcompConAlimentar (64)" ] <- batata
merged_data$grupo[ merged_data$grupo == "batatafrita,batatadepacote,salgadofritosAcompConAlimentar (66)" ] <- batata
merged_data$grupo[ merged_data$grupo == "batatafrita,batatadepacote,salgadofritosAcompConAlimentar (62)" ] <- batata
merged_data$grupo[ merged_data$grupo == "biscoitodoce,doces,balas,chocolates" ] <- bisc
merged_data$grupo[ merged_data$grupo == "biscoitodoce,doces,chocolates" ] <- bisc
merged_data$grupo[ merged_data$grupo == "biscoitosalgado,salgadinhodepacote" ] <- bs
merged_data$grupo[ merged_data$grupo == "biscoitosalgado,salgadinhodepacote" ] <- bs
merged_data$grupo[ merged_data$grupo == "biscoitosalgado,salgadinhodepacoteAcompConAlimentar (78)" ] <- bs
merged_data$grupo[ merged_data$grupo == "biscoitosalgado,salgadinhodepacotes" ] <- bs
merged_data$grupo[ merged_data$grupo == "biscoitosalgado,salgadinhosdepacotes" ] <- bs
merged_data$grupo[ merged_data$grupo == "biscoitosalgados,salgadinhosdepacoteAcompConAlimentar (79)" ] <- bs
merged_data$grupo[ merged_data$grupo == "frutasfrescas" ] <- "frutas frescas"
merged_data$grupo[ merged_data$grupo == "hamburgue,embutido" ] <- hb
merged_data$grupo[ merged_data$grupo == "hamburguer,embutidos" ] <- hb
merged_data$grupo[ merged_data$grupo == "hamburguer,embutidosAcompConAlimentar (76)" ] <- hb
merged_data$grupo[ merged_data$grupo == "legumes,verduracozidos" ] <- lg
merged_data$grupo[ merged_data$grupo == "legumes,verdurascozidas" ] <- lg
merged_data$grupo[ merged_data$grupo == "legumes,verdurascozidos" ] <- lg
merged_data$grupo[ merged_data$grupo == "saladacrua" ] <- "salada crua"

# Select CE
write.csv(merged_data, "tidy_data.csv")

ce <- merged_data %>%
  filter(UF == "CE") %>%
  select(ano, grupo, ends_with("n")) %>%
  gather(ano, grupo)

```

```

colnames(ce) <- c("ano", "grupo", "frequência", "n")

## Save plots

png(file = "gráficos/gráficos_de_barras.png", width = 1000, height = 500)

ggplot(ce, aes(x = ano, y = n, fill = frequência)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ grupo) +
  ylab("Frequência") +
  ggtitle("Frequência Simples de Consumo de Grupos Alimentares por Ano, Ceará")

dev.off()

## pdf
## 2

png(file = "gráficos/gráficos_de_barras_per.png", width = 1000, height = 500)

ggplot(ce, aes(x = ano, y = n, fill = frequência)) +
  geom_bar(stat = "identity", position="fill") +
  facet_wrap(~ grupo)+
  ylab("Porcentagem")+
  ggtitle("Frequência Percentual de Consumo de Grupos Alimentares por Ano, Ceará")

dev.off()

## pdf
## 2

# Select Brazil
br <- merged_data %>%
  filter(Região == "TOTAL BRASIL") %>%
  select(ano, grupo, ends_with("n")) %>%
  gather(ano, grupo)

colnames(br) <- c("ano", "grupo", "frequência", "n")

# Save Plots
png(file = "gráficos/gráficos_de_barras_brasil.png", width = 1000, height = 500)

ggplot(br, aes(x = ano, y = n, fill = frequência)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ grupo) +
  ylab("Frequência") +
  ggtitle("Frequência Simples de Consumo de Grupos Alimentares por Ano, Brasil")

dev.off()

## pdf
## 2

```

```
png(file = "gráficos/gráficos_de_barras_brasil_per.png", width = 1000, height = 500)

ggplot(br, aes(x = ano, y = n, fill = frequência)) +
  geom_bar(stat = "identity", position="fill") +
  facet_wrap(~ grupo)+
  ylab("Porcentagem")+
  ggtitle("Frequência Percentual de Consumo de Grupos Alimentares por Ano, Brasil")

dev.off()
```

```
## pdf
## 2
```

```
# Select Nordeste
nr <- merged_data %>%
  filter(Região == "TOTAL REGIÃO NORDESTE") %>%
  select(ano, grupo, ends_with("n")) %>%
  gather(ano, grupo)

colnames(nr) <- c("ano", "grupo", "frequência", "n")

# Save Plots
png(file = "gráficos/gráficos_de_barras_ne.png", width = 1000, height = 500)

ggplot(nr, aes(x = ano, y = n, fill = frequência)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ grupo) +
  ylab("Frequência") +
  ggtitle("Frequência Simples de Consumo de Grupos Alimentares por Ano, Nordeste do Brasil")

dev.off()
```

```
## pdf
## 2
```

```
png(file = "gráficos/gráficos_de_barras_ne_per.png", width = 1000, height = 500)

ggplot(nr, aes(x = ano, y = n, fill = frequência)) +
  geom_bar(stat = "identity", position="fill") +
  facet_wrap(~ grupo)+
  ylab("Porcentagem")+
  ggtitle("Frequência Percentual de Consumo de Grupos Alimentares por Ano, Nordeste do Brasil")

dev.off()
```

```
## pdf
## 2
```

```
sessionInfo()
```

```
## R version 3.4.0 (2017-04-21)
```

```

## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 16.04.2 LTS
##
## Matrix products: default
## BLAS: /usr/lib/libblas/libblas.so.3.6.0
## LAPACK: /usr/lib/lapack/liblapack.so.3.6.0
##
## locale:
##  [1] LC_CTYPE=pt_BR.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=pt_BR.UTF-8      LC_COLLATE=pt_BR.UTF-8
##  [5] LC_MONETARY=pt_BR.UTF-8  LC_MESSAGES=pt_BR.UTF-8
##  [7] LC_PAPER=pt_BR.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=pt_BR.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] rmarkdown_1.5  ggplot2_2.2.1  dplyr_0.5.0    tidyr_0.6.1
## [5] nvimcom_0.9-28
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.10    rprojroot_1.2    digest_0.6.12    assertthat_0.2.0
##  [5] grid_3.4.0      R6_2.2.0         plyr_1.8.4       backports_1.0.5
##  [9] gtable_0.2.0    DBI_0.6-1        magrittr_1.5     evaluate_0.10
## [13] scales_0.4.1    stringi_1.1.5    lazyeval_0.2.0   labeling_0.3
## [17] tools_3.4.0     stringr_1.2.0    munsell_0.4.3    compiler_3.4.0
## [21] colorspace_1.3-2 htmltools_0.3.6  knitr_1.15.1     tibble_1.3.0

```