

Análise exploratória: Exportações

Luiz Paulo T. Gonçalves

25/03/2023

Exportações brasileiras durante o século XIX

Os dados escolhidos para a presente análise exploratória são referentes aos dados de exportação de café (em sacas de 60kg) e de açúcar (em toneladas). Há tempo eu desenvolvi um pacote em linguagem R, {CliometricsBR}, o qual pode ser encontrado no meu github: <https://github.com/LuizPaulo23/CliometricsBR>. Com o {CliometricsBR} podemos importar tais dados direto do repositório do Instituto de Pesquisa Econômica Aplicada (IPEA) via API.

Importando os dados

```
library(CliometricsBR)
library(tidyverse)
library(GGally)
library(stargazer)

# Purando os metadados
#metadados = CliometricsBR::get_metadadosIPEA(codes = "all")
# Series que vão ser selecionadas: HIST_XACUCARQ e HIST_XCAFEQ

export_raw <- CliometricsBR::get_seriesIPEA(codes = c("HIST_XACUCARQ",
                                                    "HIST_XCAFEQ"))
```

Limando e organizando o dataset

Pois bem, com os dados importados via API usando o CliometricsBR. Pode-se, agora, organizar e limpar o dataset. Em primeiro, vamos filtrar os dados apenas de 1821 a 1900, pois queremos apenas os dados referentes ao século XIX. Em segundo, busca-se transformar os dados por logaritmo natural, assim, padronizando ambas variáveis na mesma escala.

```
db_export <- export_raw %>%
  dplyr::filter(date >= "1821-01-01" &
                date <= "1900-01-01") %>%
  relocate(date, .after = NULL) %>%
  mutate(log_value = log(value)) %>%
  tidyr::pivot_wider(id_cols = date,
                     names_from = code_series,
                     values_from = log_value) %>%
  janitor::clean_names() %>%
  dplyr::rename("Log do Açúcar" = "hist_xacucarq",
               "Log do Café" = "hist_xcafeq")
```

Visualizando: a relação de associação entre as exportações de café e açúcar do século XIX

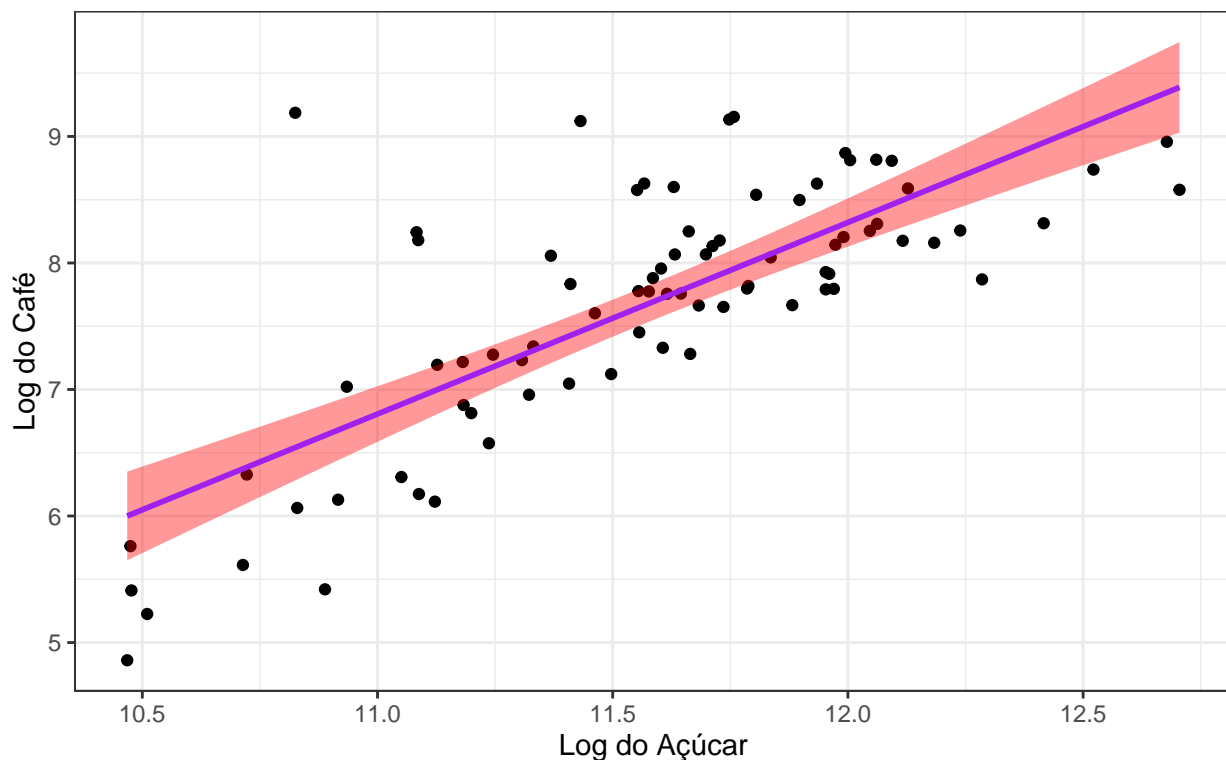
Com o dataset organizado vamos a análise exploratória com alguns plots e testes estatísticos

```
# db_export %>% glimpse()

ggplot2::ggplot(db_export)+
  aes(x = `Log do Açúcar`,
      y = `Log do Café`)+
  geom_point(pch = 19, colour = "black")+
  geom_smooth(method = "lm",
             fill = "red",
             colour = "purple")+
  labs(title = "Log das exportações de café vs Log das exportações de açúcar",
       caption = "Elaboração de Luiz Paulo T. Gonçalves (2023)")+
  theme_bw()+
  theme(plot.caption = element_text(hjust = 0))
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Log das exportações de café vs Log das exportações de açúcar



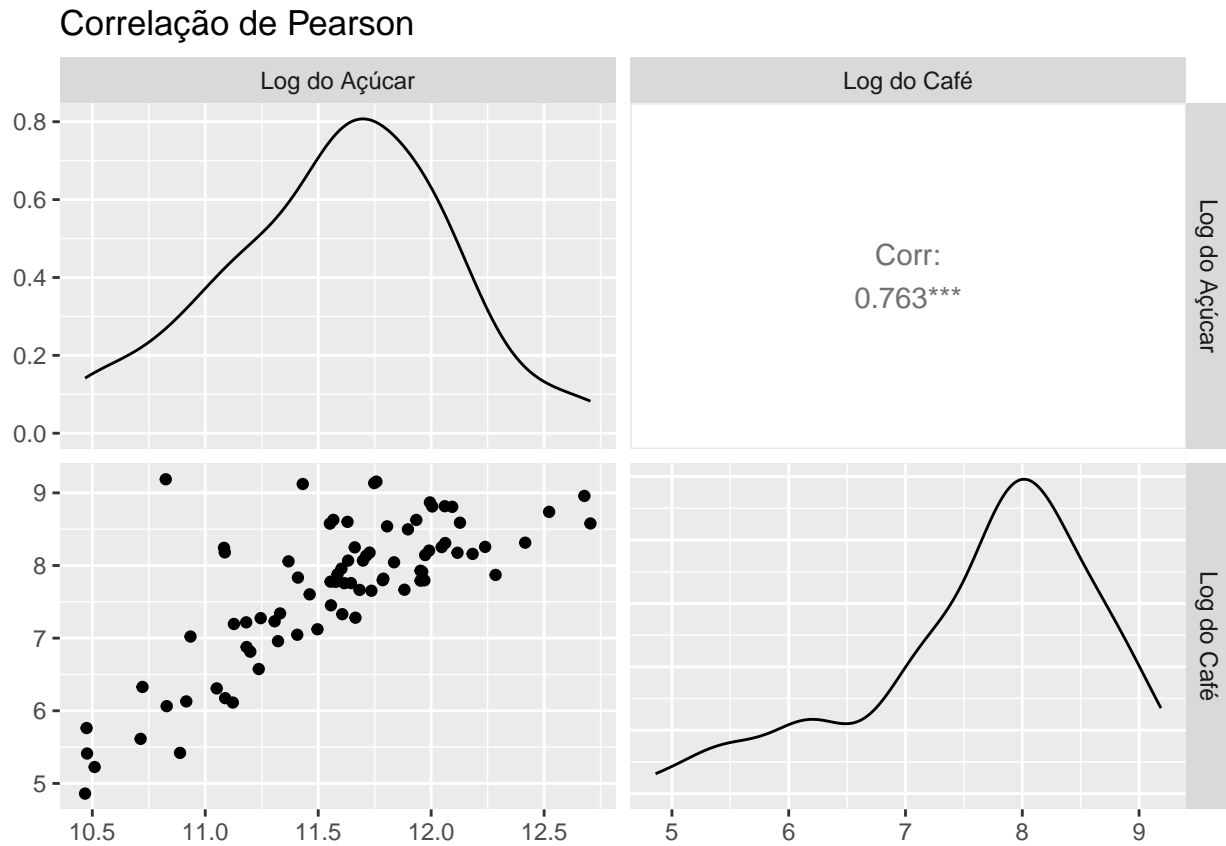
Elaboração de Luiz Paulo T. Gonçalves (2023)

Fica evidente que encontramos uma relação positiva entre as exportações de café e açúcar no século XIX, isto é, correlação positivas entre as variáveis. Pois bem, vamos aplicar um teste de Pearson para observar o grau de associação linear entre as variáveis. A correlação de Pearson pode ser dado matematicamente como segue:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Assim, temos:

```
GGally::ggpairs(db_export[, 2:3],
  title = "Correlação de Pearson")
```



Pois bem, retornou uma correlação positiva na casa de 0.76 Poderíamos ir além e modelar um modelo de regressão linear, assim, como ambas variáveis foram transformadas por log natural ficaria uma modelo log-log de regressão linear simple:

```
model_lm = lm(formula = `Log do Café` ~ `Log do Açúcar`,
  data = db_export)
```

```
stargazer::stargazer(model_lm,
  type = "text",
  title = "Modelo Log-Log")
```

```
##
## Modelo Log-Log
## =====
##               Dependent variable:
##               -----
##               `Log do Café`
## -----
## `Log do Açúcar`      1.514***
##                      (0.145)
##
## Constant            -9.851***
##                      (1.683)
##
```

```
## -----
## Observations          80
## R2                    0.582
## Adjusted R2           0.577
## Residual Std. Error   0.651 (df = 78)
## F Statistic           108.648*** (df = 1; 78)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

Observe que retornou significativo e com R^2 na casa de 0,58. O coeficiente indica que a variação de 1% das exportações de açúcar aumentavam, em média, 1,5% as exportações de café. Uma relação positiva e elástica.