

Desafio

Emanuela Mendes Cardoso

2022-12-21

Diretório

```
# meu computador (mudar de acordo)
setwd("C:/Introdução R/RGV410046/faostat")
```

Carregar pacotes

```
library(rio)
library(tidyverse)
library(metan)
library(DataExplorer)
```

Importação dos dados

```
df_1 <-
  import("area.xlsx")
df_2 <-
  import("producao.xlsx")
```

Ajustes de tabela

```
#Tabela produção

#Separando a informação de ano e produção em duas colunas individuais
df_3_prod <- df_2 |>
  separate(ANO_PRODUCAO
, into = c("ANO", "PRODUCAO"), sep = "_")

#Ajustar todas as palavras em maiúsculas, espaços, tabulação e caracteres de pontuação
df_4_prod <- tidy_strings(df_3_prod)

#Passando campo de caracter para numérico
df_4_prod$PRODUCAO <- as.numeric(as.character(df_4_prod$PRODUCAO))
```

```
## Warning: NAs introduzidos por coerção
```

```
#Tabela área
```

```
#Girar tabela de área
```

```
long <-  
  pivot_longer(df_1,  
               cols = 2:61,  
               names_to = "ANO",  
               values_to = "AREA")  
long
```

```
## # A tibble: 6,240 x 3  
##   Pais      ANO      AREA  
##   <chr>   <chr>  <dbl>  
## 1 Paraguay 1961   70000  
## 2 Paraguay 1962   71200  
## 3 Paraguay 1963   71000  
## 4 Paraguay 1964  103200  
## 5 Paraguay 1965  108000  
## 6 Paraguay 1966  102000  
## 7 Paraguay 1967   97330  
## 8 Paraguay 1968  100700  
## 9 Paraguay 1969  103270  
## 10 Paraguay 1970  103300  
## # ... with 6,230 more rows
```

```
#Ajustar todas as palavras em maiúsculas, espaços, tabulação e caracteres de pontuação  
df_4_area <- tidy_strings(long)
```

```
#Ajuste nome PAIS
```

```
df_4_area <- df_4_area  
names(df_4_area)[c(1)] <- c("PAIS")  
names(df_4_area)
```

```
## [1] "PAIS" "ANO" "AREA"
```

```
#Junção tabela de área e produção
```

```
df_5_uniao <-  
  left_join(df_4_area, df_4_prod, by = c("PAIS", "ANO"))
```

```
#Mostrar somente a primeira linha da tabela
```

```
head(df_5_uniao, n=1)
```

```
## # A tibble: 1 x 4  
##   PAIS      ANO      AREA PRODUCAO  
##   <chr>   <chr>  <dbl>   <dbl>  
## 1 PARAGUAY 1961   70000   994000
```

Calculo da produtividade

```
#Calculo da produtividade
df_4_prod <- df_5_uniao %>% mutate(prod = PRODUCAO/AREA)

#Ajustar tabela transformar valores 0 para NA
df_4_prod[df_4_prod == 0] <- NA
df_4_prod <- df_4_prod |> remove_rows_na()
```

```
## Warning: Row(s) 712, 713, 714, 715, 716, 717, 718, 719, 720, 997, 998, 999,
## 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012,
## 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1825, 1826, 1827, 1828, 1829,
## 1830, 1831, 1832, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856,
## 1857, 1858, 1859, 1860, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220,
## 2987, 2988, 2989, 2990, 2991, 2992, 2993, 2994, 2995, 2996, 2997, 2998, 2999,
## 3000, 3241, 3242, 3243, 3244, 3245, 3246, 3247, 3248, 3249, 3250, 3251, 3252,
## 3253, 3254, 3255, 3256, 3257, 3258, 3259, 3260, 3261, 3262, 3263, 3264, 3265,
## 3266, 3267, 3268, 3269, 3270, 3271, 3272, 3273, 3274, 3275, 3276, 3277, 3278,
## 3279, 3280, 3281, 3282, 3283, 3284, 3285, 3286, 3287, 3288, 3289, 3290, 3291,
## 3292, 3293, 3294, 3295, 3296, 3297, 3298, 3299, 3300, 3661, 3662, 3663, 3664,
## 3665, 3666, 3667, 3668, 3669, 3670, 3671, 3672, 3673, 3674, 3675, 3676, 3677,
## 3678, 3679, 3680, 3681, 3682, 3683, 4021, 4022, 4023, 4024, 4025, 4026, 4027,
## 4028, 4029, 4030, 4031, 4032, 4033, 4034, 4035, 4036, 4037, 4038, 4039, 4040,
## 4041, 4042, 4043, 4044, 4045, 4046, 4047, 4187, 4188, 4189, 4190, 4191, 4192,
## 4193, 4194, 4195, 4196, 4197, 4198, 4199, 4200, 4487, 4488, 4489, 4490, 4491,
## 4492, 4493, 4494, 4495, 4496, 4497, 4498, 4499, 4500, 4501, 4502, 4503, 4504,
## 4505, 4506, 4507, 4508, 4509, 4510, 4511, 4512, 4513, 4514, 4515, 4516, 4517,
## 4518, 4519, 4520, 4521, 4522, 4523, 4524, 4525, 4526, 4527, 4528, 4529, 4530,
## 4531, 4532, 4533, 4534, 4535, 4536, 4537, 4538, 4539, 4540, 4541, 4542, 4543,
## 4544, 4545, 4546, 4547, 4548, 4549, 4550, 4551, 4861, 4862, 4863, 4864, 4865,
## 4866, 4867, 4868, 4869, 4870, 4871, 4872, 4873, 4874, 4875, 4876, 4877, 4878,
## 4879, 4880, 4881, 4882, 4883, 4884, 4885, 4886, 4887, 4888, 4889, 4890, 4891,
## 4892, 4893, 4894, 4981, 4982, 4983, 4984, 4985, 4986, 4987, 4988, 5461, 5462,
## 5463, 5464, 5465, 5466, 5467, 5468, 5469, 5470, 5471, 5472, 5473, 5474, 5475,
## 5476, 5477, 5478, 5479, 5480, 5481, 5482, 5483, 5484, 5485, 5486, 5487, 5488,
## 5701, 5702, 5703, 5704, 5705, 5706, 5707, 5708, 5709, 5710, 5711, 5712, 5713,
## 5714, 5715, 5716, 5717, 5718, 5719, 5720, 5721, 5722, 5723, 5724, 5725, 5726,
## 5727, 5728, 5729, 5730, 5731, 5732, 5733, 5734, 5735, 5736, 5737, 5738, 5739,
## 5740, 5741, 5742, 5743, 5744, 5745, 5746, 5747, 5748, 5749, 5750, 5751, 5752,
## 5753, 5754, 5755, 5756, 5757, 5758, 5759, 5760, 5813, 5814, 5815, 5816, 5817,
## 5818, 5819, 5820 with NA values deleted.
```

```
#Retornar somente a primeira linha da tabela
head(df_4_prod,n=1)
```

```
## # A tibble: 1 x 5
##   PAIS      ANO   AREA PRODUCAO  prod
##   <chr>    <chr> <dbl>    <dbl> <dbl>
## 1 PARAGUAY 1961  70000   994000  14.2
```

Atividade(1)

Calcular a produtividade média, máximo, mínimo e erro padrão da média (Mg ha⁻¹) de cada país considerando todos os anos, ordenando os países pelo valor decrescente da produtividade média.

```
#função erro padrão da média
erro <- function(prod_media){
  sd(prod_media,na.rm=TRUE)/sqrt(length(na.omit(prod_media)))
}
prod_media <- df_4_prod |>
  group_by(PAIS) |>
  summarise(across(prod,.fns=list(min=min,media=mean,max=max,erro=erro, n=length))) |>
  arrange(desc(prod_media))
head(prod_media,n=1)
```

```
## # A tibble: 1 x 6
##   PAIS      prod_min prod_media prod_max prod_erro prod_n
##   <chr>      <dbl>      <dbl>    <dbl>    <dbl>  <int>
## 1 BARBADOS    16.4        22.6     31.9     0.591     60
```

Atividade (2)

Calcular a diferença (Mg ha⁻¹) da produtividade média dos três países com maior produtividade em relação à produtividade média dos países restantes (excluindo os três primeiros colocados)

```
#top3
top3 <-
  prod_media |>
  slice(1:3) |>
  pull(PAIS)
top3

## [1] "BARBADOS"      "INDIA"          "COOK_ISLANDS"

media_top3 <- df_4_prod %>%
  mutate(top3 = ifelse(PAIS %in% top3,"yes","no")) |>
  group_by(top3) |>
  summarise(mean=mean(prod, na.rm=TRUE)) |>
  pivot_wider(names_from = top3,
              values_from = mean) |>
  mutate(dif = yes - no)
media_top3
```

```
## # A tibble: 1 x 3
##       no  yes  dif
##   <dbl> <dbl> <dbl>
## 1  9.20  21.4  12.2
```

```
#exportar os dados
export(list(
  diferenca = media_top3,
  prod_media = prod_media
),
"atividades_1_2.xlsx")
```

Atividade (3)

Criar um gráfico de tendência mostrando a produtividade anual dos três países com maior produtividade.

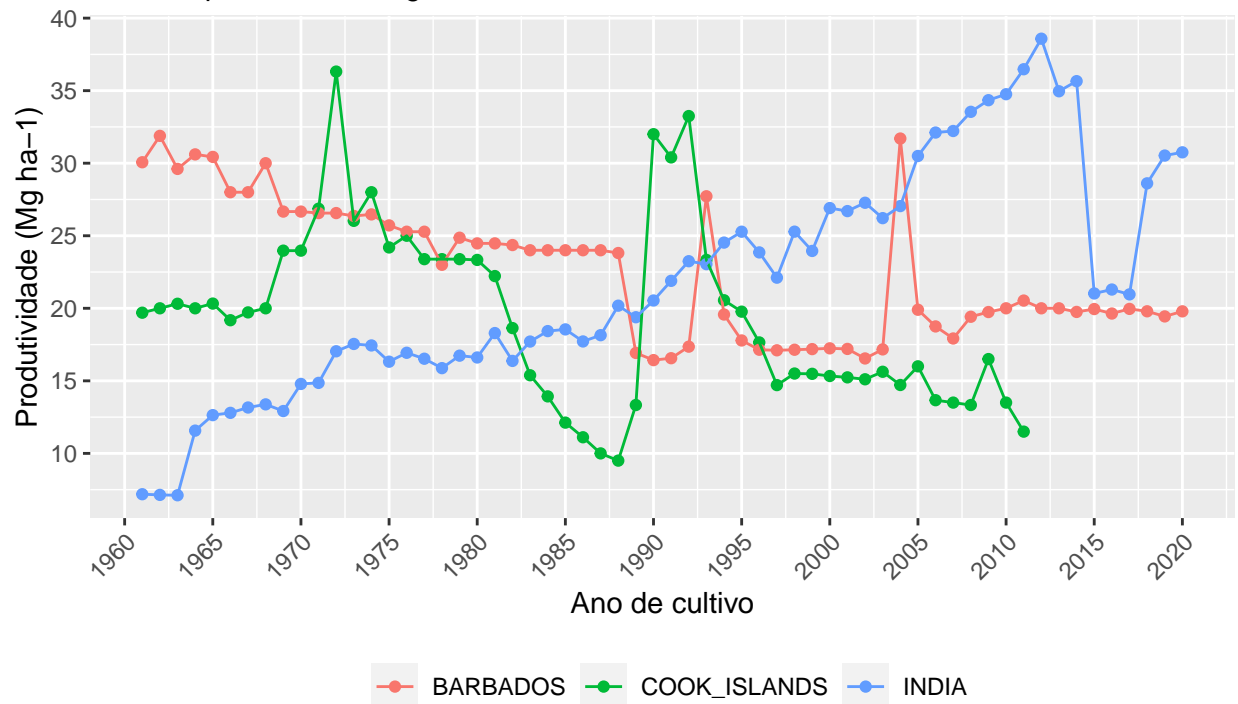
```
#Seleção gráfico
df_grafico <- df_4_prod |>
filter(PAIS %in% top3) |>
  mutate(ANO = as.numeric(ANO))

#Produção gráfico a partir da seleção

ggplot(df_grafico, aes(ANO, prod, color = PAIS, group = PAIS)) +
  geom_point() +
  geom_line() +
  scale_x_continuous(breaks = seq(1960,2020, by=5),
                     (date_labels = "Ano de cultivo")) +
  scale_y_continuous(breaks = seq(10,40, by=5))+
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) +
  labs(title = "Produtividade de mandioca de 1960 a 2020 dos TOP 3 países com maior produtividade",
        subtitle = "Fonte:https://www.fao.org/faostat/en/#data/QCL",
        caption = "Elaboração: Emanuela Mendes Cardoso",
        x = "Ano de cultivo",
        y = "Produtividade (Mg ha-1)",
        color = NULL)
```

Produtividade de mandioca de 1960 a 2020 dos TOP 3 países com maior p

Fonte: <https://www.fao.org/faostat/en/#data/QCL>



Elaboração: Emanuela Mendes Cardoso

Discussão

Os dados são derivados da Food and Agriculture Organization, e mostram a produção e área plantada de mandioca, em 102 países, de 1961 a 2020. A partir disso foi possível identificarmos a produtividade mínima, máxima e média de cada país. Pelo cálculo da produtividade média observamos que os três países com maior produtividade, são: Barbados, Cook Islands e Índia. Sendo que a produtividade média, desses três é de 21.35668(Mg ha⁻¹), enquanto dos outros 99 países restantes é de 9.19996(Mg ha⁻¹), gerando uma diferença entre eles de 12.15672(Mg ha⁻¹). Em relação ao gráfico das três principais, podemos visualizar oscilações de produtividade ao longo dos anos. Barbados em 1961 estava em destaque, mas passou por momentos de queda, recuperação e atualmente mantém-se mais estável. Já Cook Islands, estava em 1961 em posição mediana, teve períodos de alta, mas em geral segue em queda. A Índia, começou bem abaixo, mas teve grande crescimento, tendo uma queda em 2015. mas segue em recuperação.