

Segundo Caderno de Exercícios

Introdução ao Tratamento de dados em R

Meu Nome Aqui

20 de maio, 2019

Contents

| | |
|---|----------|
| Exercícios | 1 |
| Exercícios em sala (utilizando o dplyr, tidy, lubridate e ggplot) | 1 |
| Tarefa | 8 |
| Anotações | 8 |
| Anotações sobre o R | 8 |
| Anotações gerais | 8 |

Exercícios

Este é o segundo caderno de exercícios.

O arquivo Rmd, testado e funcionando será utilizado para a avaliação da aprendizagem no final do curso.

Este arquivo deve ser entregue, ao final do curso para sergiolmrivero@gmail.com

Baixe este programa para sua pasta **exercícios**

Lembre-se que você deverá modificar as referências às pastas onde armazena os dados para a sua pasta de dados. No caso deste arquivo, a pasta dados é referenciada aqui como “../Data”

Exercícios em sala (utilizando o dplyr, tidy, lubridate e ggplot)

Abaixo você tem um exemplo de uma tarefa completa de download e geração de gráficos e tabelas executados via R (e markdown).

Sua tarefa é comentar cada bloco de comandos explicando o que estes comandos fazem.

Abaixo há um conjunto de bibliotecas do R que estão sendo postas na memória. O que faz cada biblioteca?

```
library(jsonlite)
library(dplyr)
library(lubridate)
library(tidy)
library(ggplot2)
```

Resposta:

O que faz o comando `download.file`? O que é o endereço de internet dentro do comando? O que é a string “Saldo_OpCreditoPF_PA” no comando `download.file`? O que faz o comando `fromJSON`?

```
download.file("http://api.bcb.gov.br/dados/serie/bcdata.sgs.14015/dados?formato=json&",
              "../Dados/Saldo_OpCreditoPF_PA.json"
              )
Saldo_OpCreditoPF_PA <- fromJSON("../Dados/Saldo_OpCreditoPF_PA.json")
```

Resposta:

Abaixo são feitas várias operações modificando variáveis no dataframe `Saldo_OpCreditoPF_PA`. O que são estas operações e o que elas fazem?

```
Saldo_OpCreditoPF_PA$data <- as.Date(Saldo_OpCreditoPF_PA$data, "%d/%m/%Y")
Saldo_OpCreditoPF_PA$valor <- as.integer(Saldo_OpCreditoPF_PA$valor)
Saldo_OpCreditoPF_PA <- rename(Saldo_OpCreditoPF_PA, Saldo_OpCreditoPF = valor)
```

Resposta:

Abaixo há um conjunto completo de comandos. Dê uma visão geral do que eles fazem de de quais são as diferenças e semelhanças em relação aos comandos de download acima.

```
#####
#####
# Api do saldo de credito de pessoas juridicas para o Para
#####
#####

download.file("http://api.bcb.gov.br/dados/serie/bcdata.sgs.14042/dados?formato=json&", "../Dados/Sal
Saldo_OpCreditoPJ_PA <- fromJSON("../Dados/Saldo_OpCreditoPJ_PA.json")

#####
#####
# Transformando a primeira coluna em data, a segunda em inteiro e
# renomeando o nome da segunda coluna
#####
#####

Saldo_OpCreditoPJ_PA$data <- as.Date(Saldo_OpCreditoPJ_PA$data, "%d/%m/%Y")
Saldo_OpCreditoPJ_PA$valor <- as.integer(Saldo_OpCreditoPJ_PA$valor)
Saldo_OpCreditoPJ_PA <- rename(Saldo_OpCreditoPJ_PA, Saldo_OpCreditoPJ = valor)

#####
#####
# Api da porcentagem de inadimplencia de credito de pessoas fisicas
# para o Para
#####
#####

download.file("http://api.bcb.gov.br/dados/serie/bcdata.sgs.15874/dados?formato=json&", "../Dados/Inad
Inad_OpCreditoPF_PA <- fromJSON("../Dados/Inad_OpCreditoPF_PA.json")

#####
#####
```

```

# Transformando a primeira coluna em data, a segunda em numerico e
# renomeando o nome da segunda coluna
#####
#####

Inad_OpCreditoPF_PA$data <- as.Date(Inad_OpCreditoPF_PA$data, "%d/%m/%Y")
Inad_OpCreditoPF_PA$valor <- as.numeric(Inad_OpCreditoPF_PA$valor)
Inad_OpCreditoPF_PA <- rename(Inad_OpCreditoPF_PA, Inad_OpCreditoPF = valor)

#####
#####
# Api da porcentagem de inadimplencia de credito de pessoas juridicas
# para o Para
#####
#####

download.file("http://api.bcb.gov.br/dados/serie/bcdata.sgs.15906/dados?formato=json&","../Dados/Inad_OpCreditoPJ_PA.json")
Inad_OpCreditoPJ_PA <- fromJSON("../Dados/Inad_OpCreditoPJ_PA.json")

#####
#####
# Transformando a primeira coluna em data, a segunda em numerico e
# renomeando o nome da segunda coluna
#####
#####

Inad_OpCreditoPJ_PA$data <- as.Date(Inad_OpCreditoPJ_PA$data, "%d/%m/%Y")
Inad_OpCreditoPJ_PA$valor <- as.numeric(Inad_OpCreditoPJ_PA$valor)
Inad_OpCreditoPJ_PA <- rename(Inad_OpCreditoPJ_PA, Inad_OpCreditoPJ = valor)

```

Resposta:

Abaixo está sendo executado um full_join. O que este comando está fazendo nos 3 dataframes abaixo? O que ele está gerando?

```
PA <- full_join(Saldo_OpCreditoPF_PA, Saldo_OpCreditoPJ_PA)
```

```
## Joining, by = "data"
```

```
PA <- full_join(PA, Inad_OpCreditoPF_PA)
```

```
## Joining, by = "data"
```

```
PA <- full_join(PA, Inad_OpCreditoPJ_PA)
```

```
## Joining, by = "data"
```

O que faz o comando abaixo? Qual a sua utilidade para o trabalho executado? (Dica, há mais de um estado)

```
PA$UF <- "PA"
```

** Resposta: **

Abaixo há um conjunto completo de comandos, repetindo os comandos acima. Dê uma visão geral do que eles fazem de de quais são as diferenças em relação aos comandos de download acima.(Dica: o estado é Espírito Santo)

```
#####
# Espírito Santo
#####

download.file("http://api.bcb.gov.br/dados/serie/bcdata.sgs.14009/dados?formato=json&", "../Dados/Sal
Saldo_OpCreditoPF_ES <- fromJSON("../Dados/Saldo_OpCreditoPF_ES.json")

Saldo_OpCreditoPF_ES$data <- as.Date(Saldo_OpCreditoPF_ES$data, "%d/%m/%Y")
Saldo_OpCreditoPF_ES$valor <- as.integer(Saldo_OpCreditoPF_ES$valor)
Saldo_OpCreditoPF_ES <- rename(Saldo_OpCreditoPF_ES, Saldo_OpCreditoPF = valor)

download.file("http://api.bcb.gov.br/dados/serie/bcdata.sgs.14036/dados?formato=json&", "../Dados/Sal
Saldo_OpCreditoPJ_ES <- fromJSON("../Dados/Saldo_OpCreditoPJ_ES.json")

Saldo_OpCreditoPJ_ES$data <- as.Date(Saldo_OpCreditoPJ_ES$data, "%d/%m/%Y")
Saldo_OpCreditoPJ_ES$valor <- as.integer(Saldo_OpCreditoPJ_ES$valor)
Saldo_OpCreditoPJ_ES <- rename(Saldo_OpCreditoPJ_ES, Saldo_OpCreditoPJ = valor)

download.file("http://api.bcb.gov.br/dados/serie/bcdata.sgs.15868/dados?formato=json&", "../Dados/Inad
Inad_OpCreditoPF_ES <- fromJSON("../Dados/Inad_OpCreditoPF_ES.json")

Inad_OpCreditoPF_ES$data <- as.Date(Inad_OpCreditoPF_ES$data, "%d/%m/%Y")
Inad_OpCreditoPF_ES$valor <- as.numeric(Inad_OpCreditoPF_ES$valor)
Inad_OpCreditoPF_ES <- rename(Inad_OpCreditoPF_ES, Inad_OpCreditoPF = valor)

download.file("http://api.bcb.gov.br/dados/serie/bcdata.sgs.15900/dados?formato=json&", "../Dados/Inad
Inad_OpCreditoPJ_ES <- fromJSON("../Dados/Inad_OpCreditoPJ_ES")

Inad_OpCreditoPJ_ES$data <- as.Date(Inad_OpCreditoPJ_ES$data, "%d/%m/%Y")
Inad_OpCreditoPJ_ES$valor <- as.numeric(Inad_OpCreditoPJ_ES$valor)
Inad_OpCreditoPJ_ES <- rename(Inad_OpCreditoPJ_ES, Inad_OpCreditoPJ = valor)

ES <- full_join(Saldo_OpCreditoPF_ES, Saldo_OpCreditoPJ_ES)

## Joining, by = "data"
ES <- full_join(ES, Inad_OpCreditoPF_ES)

## Joining, by = "data"
ES <- full_join(ES, Inad_OpCreditoPJ_ES)

## Joining, by = "data"
ES$UF <- "ES"
```

O que fazem este comando `full_join` e o `save`?

```
PAES <- full_join(PA, ES)
```

```
## Joining, by = c("data", "Saldo_OpCreditoPF", "Saldo_OpCreditoPJ", "Inad_OpCreditoPF", "Inad_OpCreditoPJ")
```

```
save(PAES, file="../Dados/CreditoPAES.Rda")
```

Explique o que faz este comando gather (Dica: olhe o dataframe DF_PAES_1 que foi gerado)

```
DF_PAES_1 <- gather(PAES, "Saldo_OpCreditoPF", "Saldo_OpCreditoPJ",  
  "Inad_OpCreditoPF", "Inad_OpCreditoPJ", key = Variaveis,  
  value = Valor)
```

Explique o que faz este comando DF_PAES_2 (Dica: olhe o dataframe DF_PAES_2 que foi gerado)

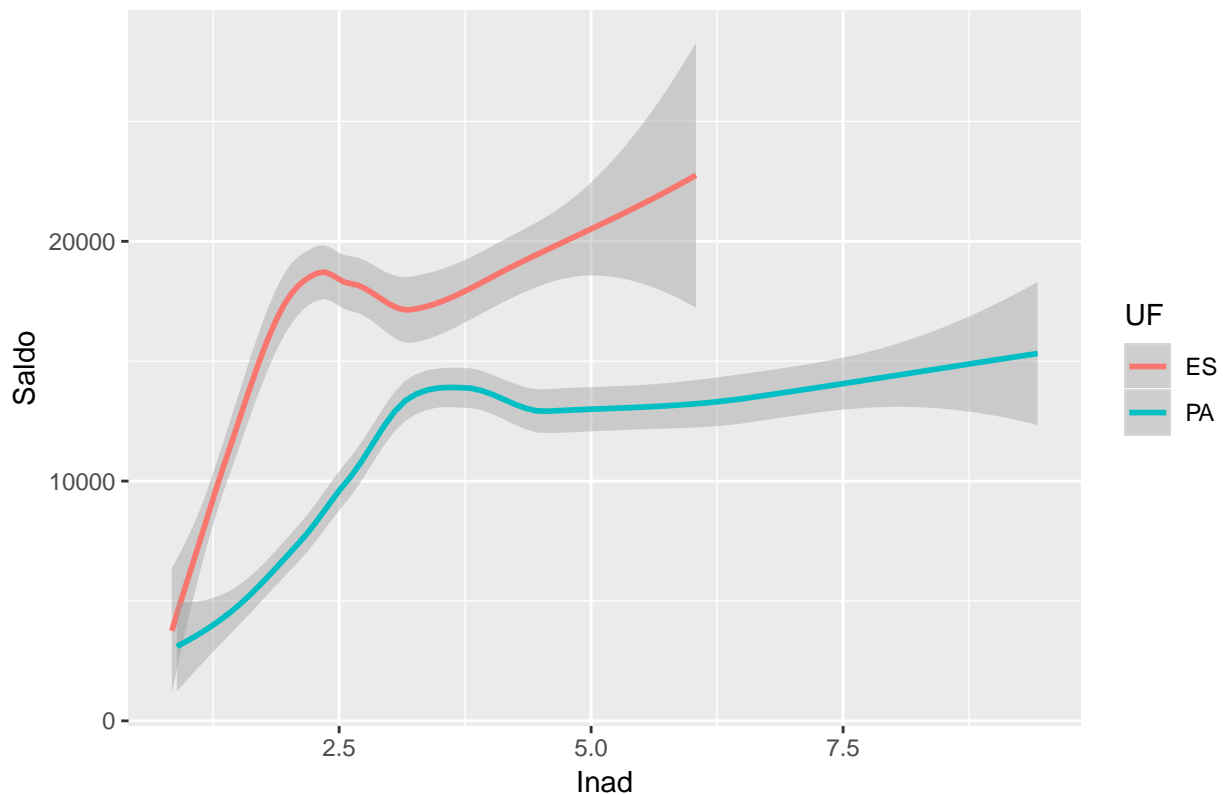
```
DF_PAES_2 <- DF_PAES_1 %>% spread(Variaveis, Valor)
```

Explique o que fazem os comandos ggplot abaixo

```
gf4 <- ggplot(DF_PAES_2, aes(y = Saldo_OpCreditoPJ, x = Inad_OpCreditoPJ,  
  color = UF)) +  
  #geom_point() +  
  geom_smooth() +  
  ggtitle("Relação entre Saldo e inad de crédito para PJ") +  
  labs(x="Inad", y="Saldo", color = "UF")  
plot(gf4)
```

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

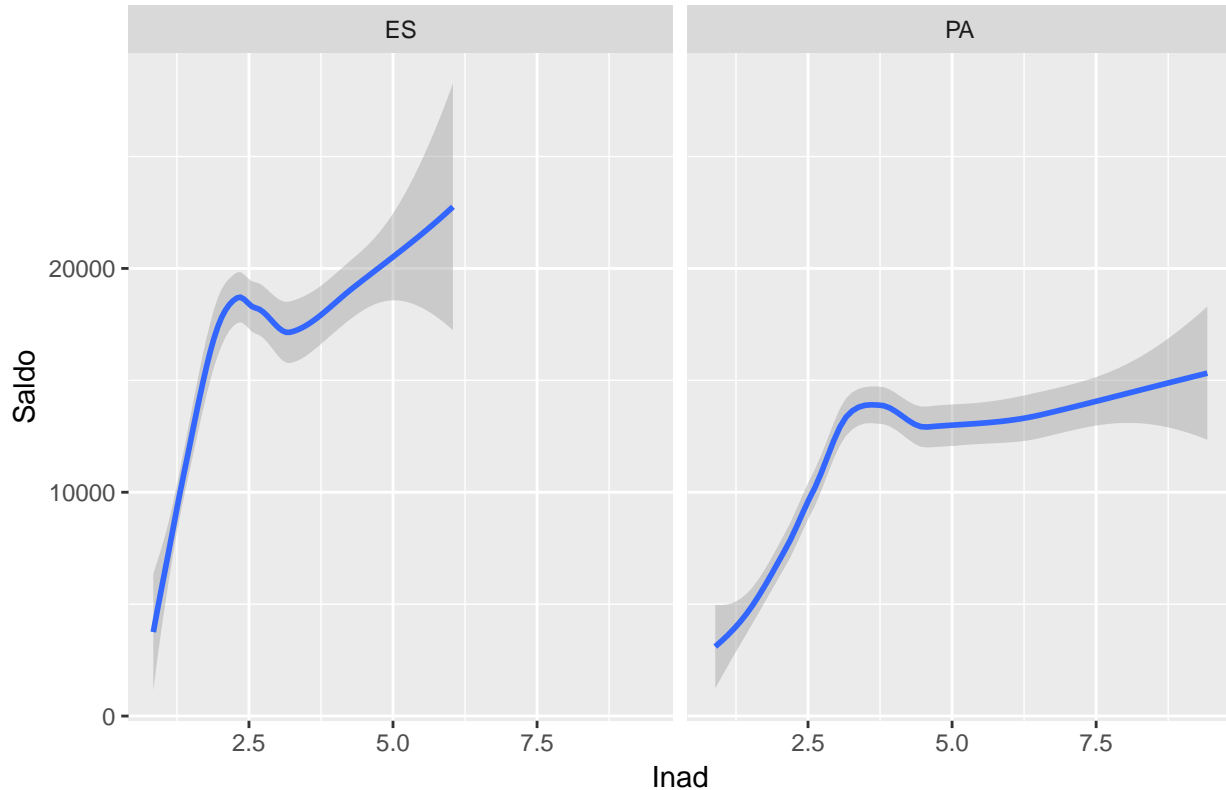
Relação entre Saldo e inad de crédito para PJ



```
gf5 <- ggplot(DF_PAES_2, aes(y = Saldo_OpCreditoPJ, x = Inad_OpCreditoPJ)) +
  #geom_point() +
  geom_smooth() +
  facet_wrap(~UF) +
  ggtitle("Relação entre Saldo e inad de crédito para PJ") +
  labs(x="Inad", y="Saldo", color = "UF")
plot(gf5)
```

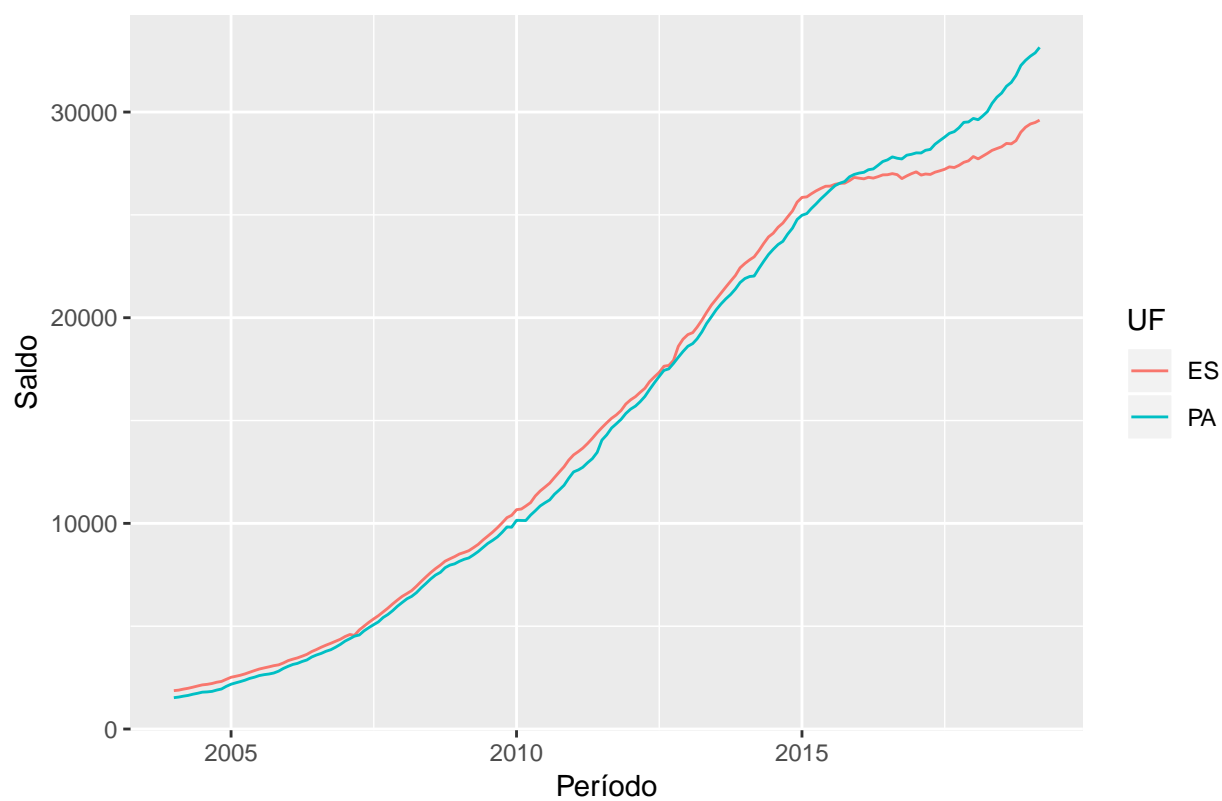
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Relação entre Saldo e inad de crédito para PJ

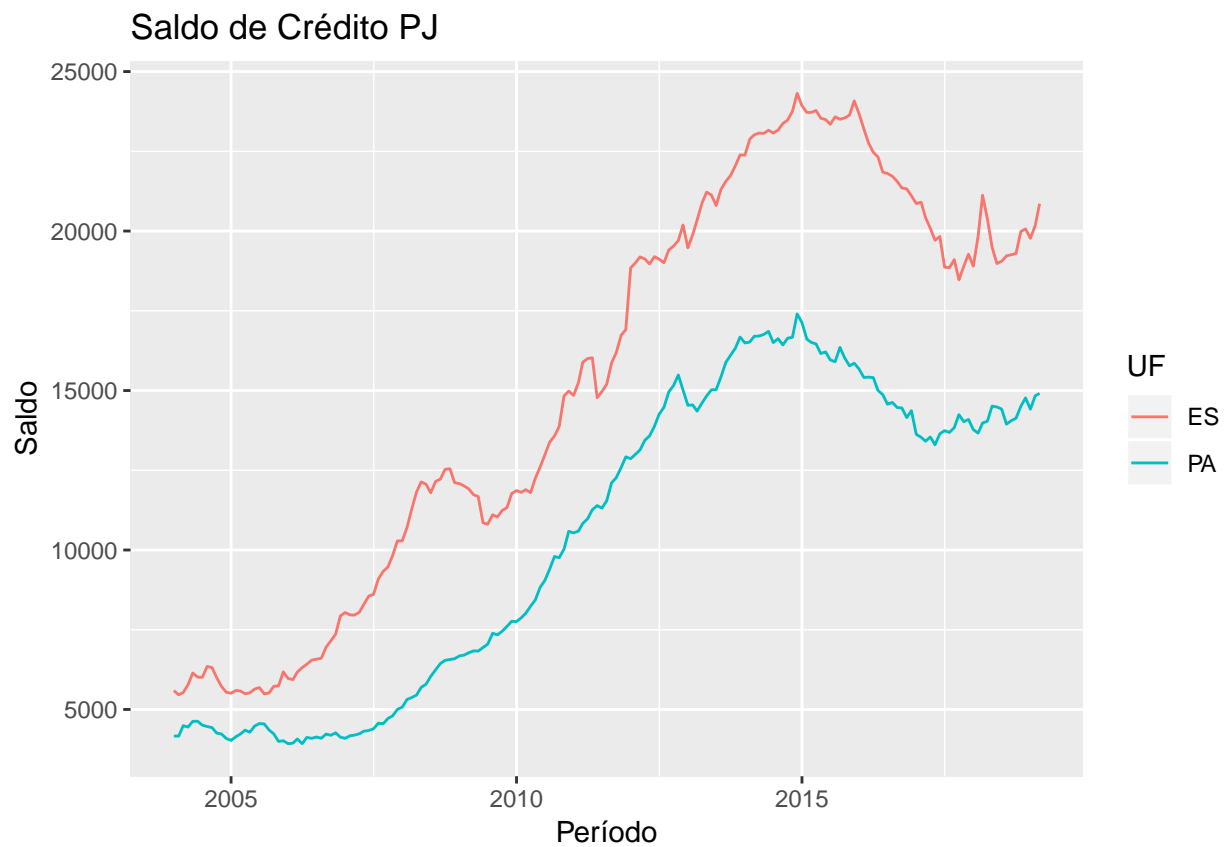


```
gf6 <- ggplot(DF_PAES_2, aes(y = Saldo_OpCreditoPF, x = data, color = UF)) +
  geom_line() +
  #geom_smooth() +
  #facet_wrap(~UF) +
  ggtitle("Saldo de Crédito PF") +
  labs(x="Período", y="Saldo", color = "UF")
plot(gf6)
```

Saldo de Crédito PF



```
gf7 <- ggplot(DF_PAES_2, aes(y = Saldo_OpCreditoPJ, x = data, color = UF)) +  
  geom_line() +  
  #geom_smooth() +  
  #facet_wrap(~UF) +  
  ggtitle("Saldo de Crédito PJ") +  
  labs(x="Período", y="Saldo", color = "UF")  
plot(gf7)
```



Tarefa

Avaliando um problema de análise de dados em sua área proponha a utilização do R e mark-down para resolvê-lo,

Anotações

Anotações sobre o R

Anotações gerais