

# Desafio de Ciência de Dados

Análise Macro

Cairê Britto Barletta

21/08/2021

```
#carregando as bibliotecas necessarias
```

```
library(tidyverse)
library(lubridate)
library(RcppRoll)
library(gridExtra)
library(ggpubr)
library(sidr)
library(ipeadatar)
library(BETS)
library(readxl)
library(seasonal)
library(gtrendsR)
```

```
#funcao que sera utilizada, retirada da documentacao do pacote 'tstools'
```

```
acum_i <- function(data, n){

  data_ma_n <- RcppRoll::roll_meanr(data, n)
  data_lag_n <- dplyr::lag(data_ma_n, n)
  data_acum_n = (((data_ma_n/data_lag_n) - 1) * 100)

  return(data_acum_n)
}
```

1. Utilizando o pacote `sidr`, você deve pegar os números-índices da produção industrial geral, com e sem ajuste sazonal.

```
#puxando a PIM com ajuste
```

```
PIM_com_ajuste <-
  get_sidra(api = "/t/3653/n1/all/v/3134/p/all/c544/129314/d/v3134%201") %>%
  mutate(date = parse_date(`Mês (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
  rename(ind_geral_ca = Valor) %>%
  as.data.frame()
```

```
#puxando a PIM sem ajuste
```

```
PIM_sem_ajuste <-
  get_sidra(api = "/t/3653/n1/all/v/3135/p/all/c544/129314/d/v3135%201") %>%
  mutate(date = parse_date(`Mês (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
```

```
rename(ind_geral_sa = Valor) %>%
as.data.frame()
```

```
#unindo as duas series em um unico df
dados <- left_join(PIM_com_ajuste, PIM_sem_ajuste, by = "date")
```

(i) Calcule a variação marginal, a variação interanual e a variação acumulada em 12 meses da produção industrial.

```
#calculando a variacao marginal
var_marg <- dados %>%
  mutate(var_marg_ca = (ind_geral_ca/lag(ind_geral_ca, 1) - 1) * 100) %>%
  select(c(date, var_marg_ca)) %>%
  na.omit()

#calculando a variacao interanual
var_interanual <- dados %>%
  mutate(var_interanual_sa = (ind_geral_sa/lag(ind_geral_sa, 12) - 1) * 100) %>%
  select(c(date, var_interanual_sa)) %>%
  na.omit()

#variacao acumulada em 12m
var_12m <- dados %>%
  mutate(var_12m_sa = acum_i(ind_geral_sa, 12)) %>%
  select(c(date, var_12m_sa)) %>%
  na.omit()
```

(ii) Gere gráficos dos números índices da produção industrial e de suas métricas de crescimento.

```
#criando graficos da industria geral com e sem ajuste
g1 <- dados %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = ind_geral_ca, colour = "Com ajuste sazonal"), size = 0.5) +
  geom_line(aes(y = ind_geral_sa, colour = "Sem ajuste sazonal"), size = 0.5) +
  scale_x_date(breaks = "1 year",
               labels = scales::date_format("%Y")) +
  scale_color_manual(values = c("Com ajuste sazonal" = "darkblue",
                                "Sem ajuste sazonal" = "darkgray")) +
  labs(title = "Gráfico 1",
        subtitle = "Números-Índice, Base: média de 2012 = 100.",
        x = element_blank(),
        y = element_blank()) +
  theme_light() + theme(legend.title = element_blank(),
                        legend.position = "bottom",
                        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

#criando o grafico da variacao marginal
g2 <- var_marg %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = var_marg_ca), color = "darkblue", size = 0.5) +
  geom_line(aes(y = mean(var_marg_ca)), linetype = 2, color = "darkgray", size = 0.5) +
  scale_x_date(breaks = "1 year",
               labels = scales::date_format("%Y")) +
```

```

scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    decimal.mark = ',')) +

labs(title = "Gráfico 2",
      subtitle = "Variação marginal (%)",
      x = element_blank(),
      y = element_blank()) +
theme_light() + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

#criando o grafico da variacao interanual
g3 <- var_interanual %>%
ggplot(aes(x = date)) +
geom_line(aes(y = var_interanual_sa), color = "darkblue", size = 0.5) +
geom_line(aes(y = mean(var_interanual_sa)), linetype = 2, color = "darkgray",
          size = 0.5) +
scale_x_date(breaks = "1 year",
             labels = scales::date_format("%Y")) +
scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    decimal.mark = ',')) +

labs(title = "Gráfico 3",
      subtitle = "Variação interanual (%)",
      x = element_blank(),
      y = element_blank()) +
theme_light() + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

#criando o grafico da variacao acumulada em 12m
g4 <- var_12m %>%
ggplot(aes(x = date)) +
geom_line(aes(y = var_12m_sa), color = "darkblue", size = 0.5) +
geom_line(aes(y = mean(var_12m_sa)), linetype = 2, color = "darkgray", size = 0.5) +
scale_x_date(breaks = "1 year",
             labels = scales::date_format("%Y")) +
scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    decimal.mark = ',')) +

labs(title = "Gráfico 4",
      subtitle = "Variação acumulada em 12m (%)",
      x = element_blank(),
      y = element_blank()) +
theme_light() + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

#arranjando os graficos em quatro linhas
graficos <- arrangeGrob(g1, g2, g3, g4, nrow = 4)

#plotando os quatro primeiros graficos, com titulo e fonte unicos
annotate_figure(graficos,
               top = text_grob("Produção Industrial Geral e Métricas",
                              face = "bold", size = 14),
               bottom = text_grob("Fonte: IBGE-SIDRA. Elaboração do Autor.",
                                  face = "italic", size = 10, x = 0.87))

```

## Produção Industrial Geral e Métricas

Gráfico 1

Números-Índice, Base: média de 2012 = 100.

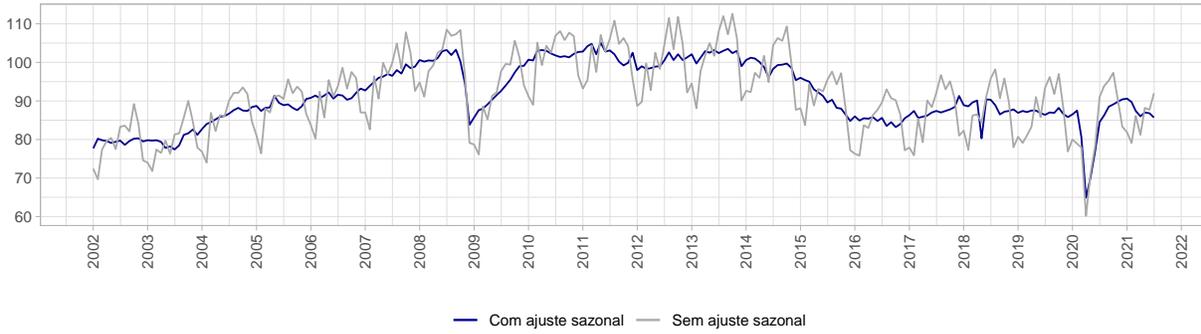


Gráfico 2

Varição marginal (%)

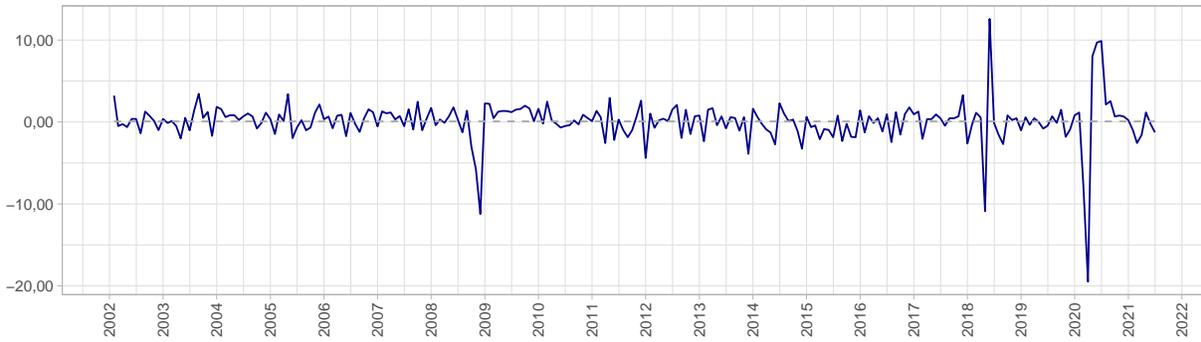


Gráfico 3

Varição interanual (%)

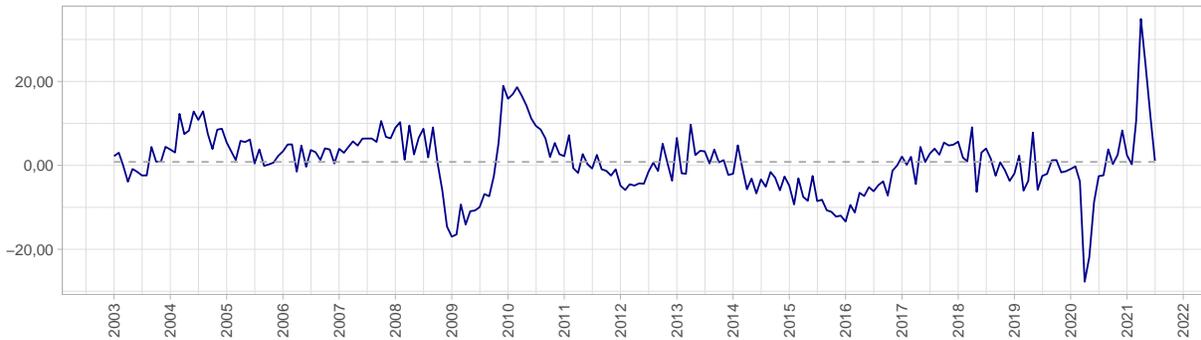
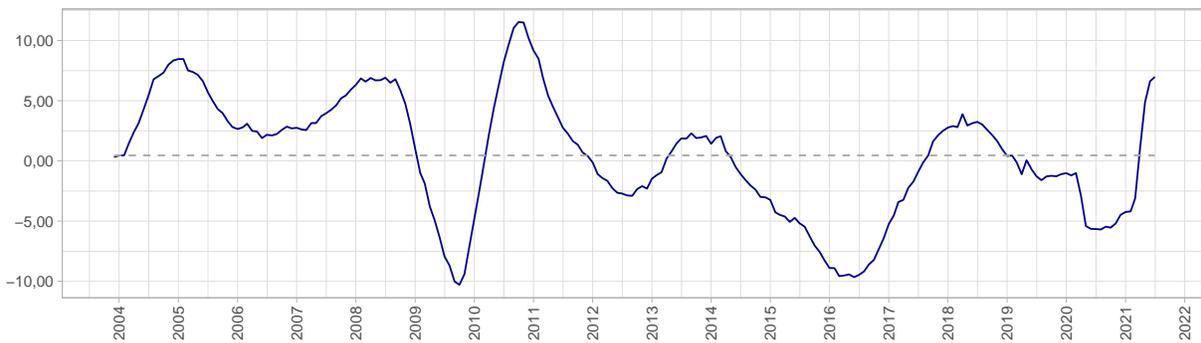


Gráfico 4

Varição acumulada em 12m (%)



Fonte: IBGE-SIDRA. Elaboração do Autor.

## Qual a diferença entre elas?

Pode-se observar algumas diferenças quando nos deparamos com as séries de Produção Industrial Geral e suas diversas métricas de apresentação do desempenho da indústria. Primeiramente, conforme o *Gráfico 1*, que representa a evolução do número-índice, podemos observar pela linha em cinza a clara sazonalidade anual (com crescimento no primeiro semestre e arrefecimento no segundo).

A variação marginal, conforme o *Gráfico 2*, representa a variação mensal do número-índice. Com esta métrica, perdemos a primeira observação; e não temos o efeito de ciclicidade e sazonalidade que vemos nas demais séries.

Já a variação interanual, representada pelo *Gráfico 3*, é a divisão do número-índice em um mês, pelo número-índice no mesmo mês, porém do ano anterior. Nesta série, perdemos as doze primeiras observações e podemos destacar que a sazonalidade observada no *Gráfico 1* é eliminada.

Por fim, temos a variação acumulada em doze meses (*Gráfico 4*), que é a média móvel do crescimento em 12m do mês em  $t$ , em relação a média móvel do crescimento em 12m do mês em  $t-1$ . Nesta série, perdemos as vinte e três primeiras observações; e fica clara a ciclicidade (suavizada, pela métrica se tratar de uma média móvel).

**2. Utilizando o pacote `sidrar`, você deve pegar os números-índices adequados da produção industrial geral, do volume do Varejo Ampliado e do volume dos Serviços. Você deve (i) criar as variações acumuladas em 12 meses das três pesquisas e plotar no mesmo gráfico de linhas com datas a partir de janeiro de 2014; (ii) criar um mesmo gráfico de barras com datas a partir de dezembro de 2018 com as variações na margem das três pesquisas.**

```
#puxando dados do volume do varejo ampliado sem ajuste a partir de 2012
vol_varejo_amp_sa <-
  get_sidra(api = "/t/3417/n1/all/v/1186/p/all/c11046/40311/d/v1186%201") %>%
  mutate(date = parse_date(`Mês (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
  filter(date >= "2012-02-01") %>%
  rename(varejo_ampliado_sa = Valor) %>%
  as.data.frame()

#puxando dados do volume dos servicos sem ajuste a partir de 2012
vol_servicos_sa <-
  get_sidra(api = "/t/6442/n1/all/v/8677/p/all/c11046/40311/d/v8677%201") %>%
  mutate(date = parse_date(`Mês (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
  filter(date >= "2012-02-01") %>%
  rename(servicos_sa = Valor) %>%
  as.data.frame()

#unindo varejo, servicos e industria em um unico df
dados_acum_12m <- left_join(vol_varejo_amp_sa, PIM_sem_ajuste, by = "date") %>%
  filter(date <= "2021-06-01") %>%
  cbind(vol_servicos_sa$servicos_sa) %>%
  rename(servicos_sa = `vol_servicos_sa$servicos_sa`) %>%
  mutate(varejo_amp = acum_i(varejo_ampliado_sa, 12),
         servicos = acum_i(servicos_sa, 12),
         PIM = acum_i(ind_geral_sa, 12)) %>%
  select(c(date, varejo_amp, servicos, PIM)) %>%
  na.omit()
```

```

#puxando dados do volume do varejo ampliado com ajuste a partir de 2012
vol_varejo_amp_ca <-
  get_sidra(api = "/t/3417/n1/all/v/1186/p/all/c11046/40312/d/v1186%201") %>%
  mutate(date = parse_date(`Mês (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
  filter(date >= "2012-02-01") %>%
  rename(varejo_ampliado_ca = Valor) %>%
  as.data.frame()

#puxando dados do volume dos servicos com ajuste a partir de 2012
vol_servicos_ca <-
  get_sidra(api = "/t/6442/n1/all/v/8677/p/all/c11046/40312/d/v8677%201") %>%
  mutate(date = parse_date(`Mês (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
  filter(date >= "2012-02-01") %>%
  rename(servicos_ca = Valor) %>%
  as.data.frame()

#juntando tudo em um dataframe
dados_var_marg <- left_join(vol_varejo_amp_ca, PIM_com_ajuste, by = "date") %>%
  filter(date <= "2021-06-01") %>%
  cbind(vol_servicos_ca$servicos_ca) %>%
  rename(servicos_ca = `vol_servicos_ca$servicos_ca`) %>%
  mutate(varejo_amp = (varejo_ampliado_ca/lag(varejo_ampliado_ca, 1) - 1) * 100,
         servicos = (servicos_ca/lag(servicos_ca, 1) - 1) * 100,
         PIM = (ind_geral_ca/lag(ind_geral_ca, 1) - 1) * 100) %>%
  select(c(date, varejo_amp, servicos, PIM)) %>%
  filter(date >= "2018-12-01") %>%
  na.omit()

#criando o grafico de linha em conjunto dos acumulados em 12m a partir de 2014
g5 <- dados_acum_12m %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = varejo_amp, colour = "Varejo Ampliado"), size = 0.5) +
  geom_line(aes(y = servicos, colour = "Serviços"), size = 0.5) +
  geom_line(aes(y = PIM, colour = "Produção Industrial"), size = 0.5) +
  scale_x_date(breaks = "6 months",
              labels = scales::date_format("%m/%Y")) +
  scale_color_manual(values = c("Varejo Ampliado" = "darkgrey",
                                "Serviços" = "darkblue",
                                "Produção Industrial" = "lightblue")) +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    decimal.mark = ',,')) +
  labs(title = "Variação acumulada em 12m (%)",
       x = element_blank(),
       y = element_blank()) +
  theme_light() + theme(legend.title = element_blank(),
                        legend.position = "bottom",
                        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

#gerando um grafico de barras com variacao marginal do varejo ampliado a partir de 2018
g6 <- dados_var_marg %>%
  ggplot(aes(x = date)) +

```

```

geom_bar(aes(y = varejo_amp), stat = "identity", color = "darkblue",
         fill = "darkgrey") +
scale_x_date(breaks = "2 months",
             labels = scales::date_format("%m/%Y")) +
scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                decimal.mark = ',')) +

labs(title = "Volume do Varejo Ampliado",
     x = element_blank(),
     y = "Variação marginal (%)") +
theme_light() + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

#gerando um grafico de barras com variacao marginal dos servicos a partir de 2018
g7 <- dados_var_marg %>%
ggplot(aes(x = date)) +
geom_bar(aes(y = servicos), stat = "identity", color = "darkblue", fill = "lightblue") +
scale_x_date(breaks = "2 months",
             labels = scales::date_format("%m/%Y")) +
scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                decimal.mark = ',')) +

labs(title = "Volume dos Serviços",
     x = element_blank(),
     y = "Variação marginal (%)") +
theme_light() + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

#gerando um grafico de barras com variacao marginal da PIM a partir de 2018
g8 <- dados_var_marg %>%
ggplot(aes(x = date)) +
geom_bar(aes(y = PIM), stat = "identity", color = "darkblue", fill = "lightgreen") +
scale_x_date(breaks = "2 months",
             labels = scales::date_format("%m/%Y")) +
scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                decimal.mark = ',')) +

labs(title = "Produção Industrial Geral",
     x = element_blank(),
     y = "Variação marginal (%)") +
theme_light() + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

```

```

#arranjando os graficos em quatro linhas
graficos2 <- arrangeGrob(g5, g6, g7, g8, nrow = 4)

```

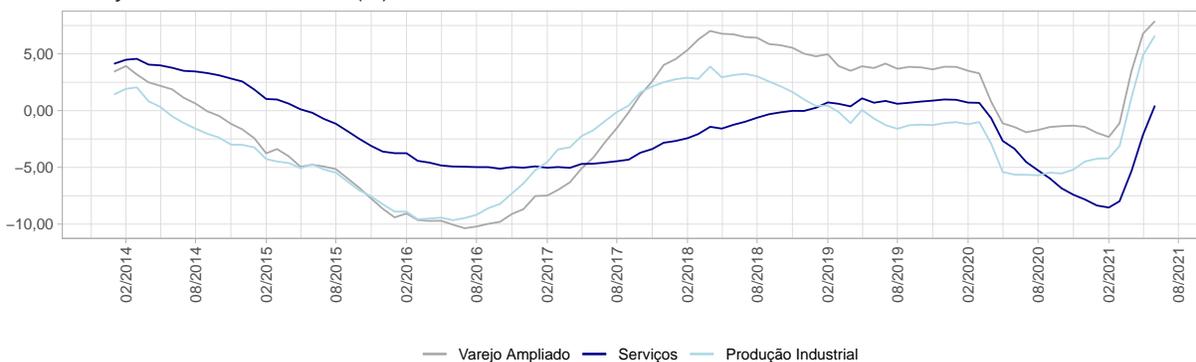
```

#plotando os quatro segundos graficos, com titulo e fonte unicos
annotate_figure(graficos2,
               top = text_grob("Varejo, Serviços e Indústria",
                              face = "bold", size = 14),
               bottom = text_grob("Fonte: IBGE-SIDRA. Elaboração do Autor.",
                                  face = "italic", size = 10, x = 0.87))

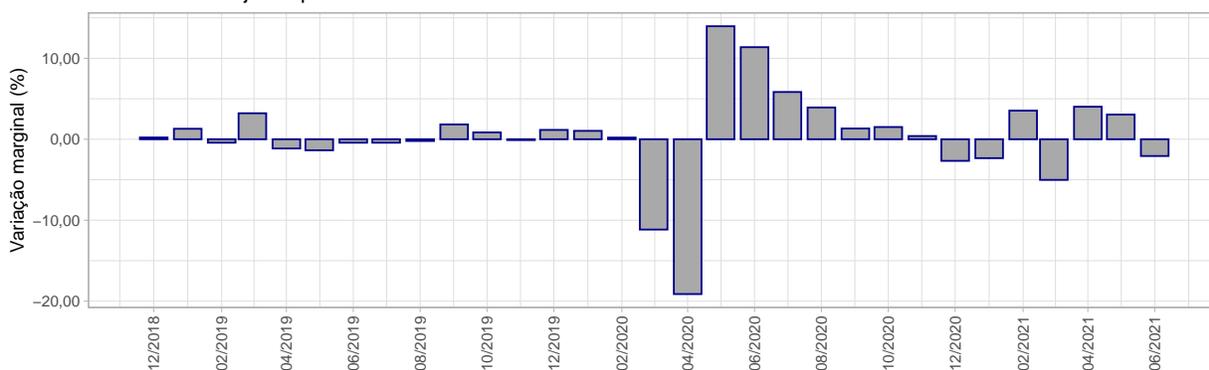
```

## Varejo, Serviços e Indústria

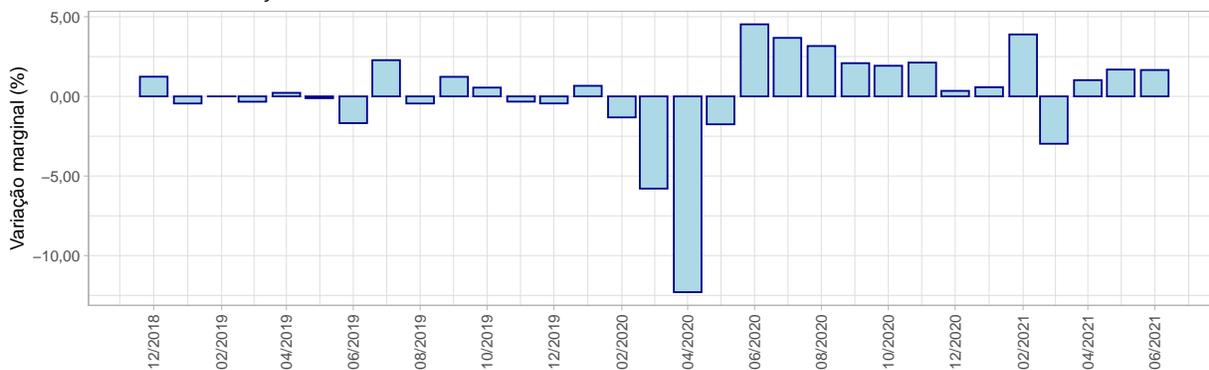
Varição acumulada em 12m (%)



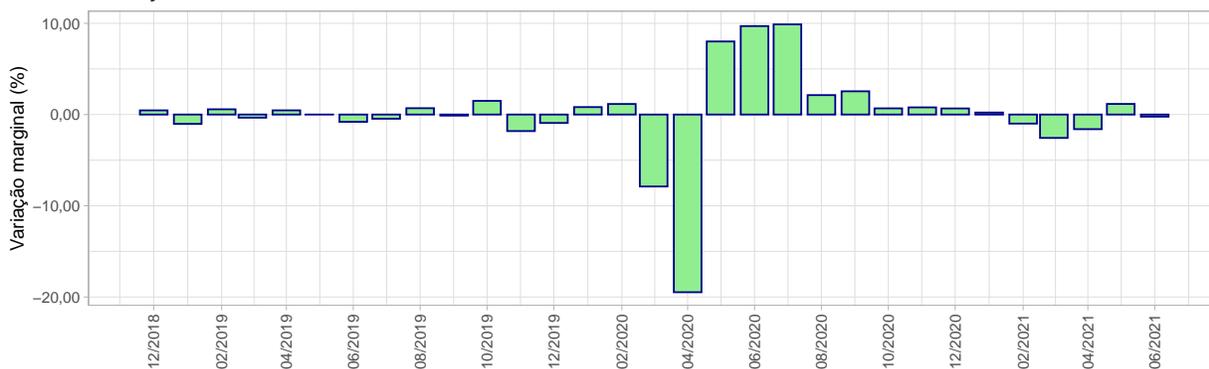
Volume do Varejo Ampliado



Volume dos Serviços



Produção Industrial Geral



Fonte: IBGE-SIDRA. Elaboração do Autor.

3. Construa um gráfico de correlação entre a produção industrial geral e a produção de veículos divulgada pela ANFAVEA, utilizando para isso as medidas que julgar mais adequada.

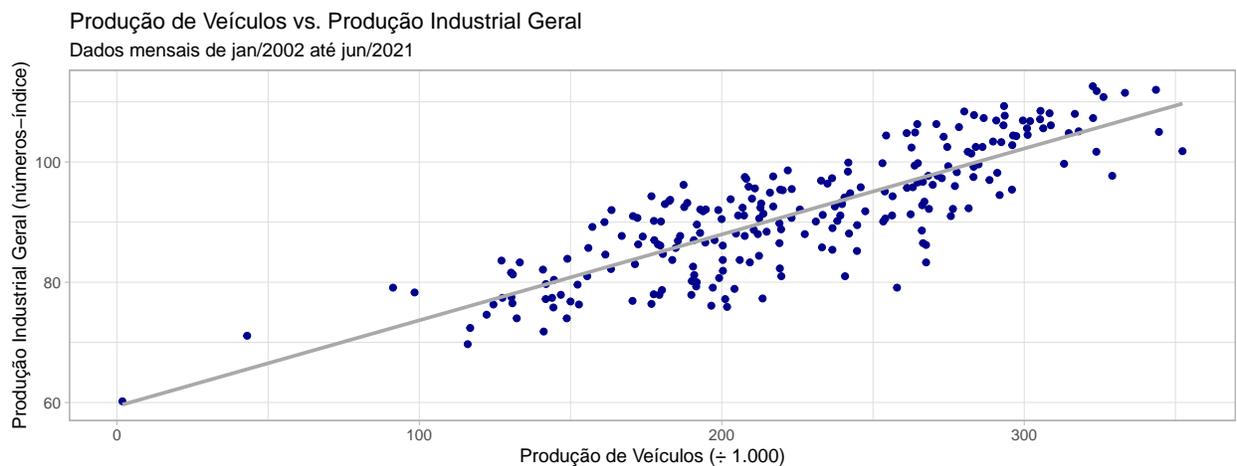
```
#puxando o arquivo .xslm do site da ANFAVEA e baixando em nosso diretorio
url <- "http://www.anfavea.com.br/docs/SeriesTemporais_Autoveiculos.xslm"
download.file(url, destfile = "prod_veiculos.xslm", mode = "wb")

#lendo o arquivo baixado
prod_veiculos <- read_excel("prod_veiculos.xslm",
                           skip = 4,
                           col_types = c("date", rep("numeric", 25)))
names(prod_veiculos)[1] <- "date"

#selecionando somente a data, producao total de autoveiculos e filtrando a data
prod_veiculos <- prod_veiculos %>%
  select(c(date, Produção...5)) %>%
  rename(producao_veiculos = Produção...5)

#juntando em um unico dataframe
dados_PIM_prod <- left_join(PIM_sem_ajuste, prod_veiculos, by = "date")

#criando o grafico de correlacao entre ind_geral_sa ~ producao_veiculos
dados_PIM_prod %>%
  ggplot(aes(x = producao_veiculos/1000, y = ind_geral_sa)) +
  geom_point(color = "darkblue") +
  geom_smooth(method = "lm", se = F, colour = "darkgrey") + #b1 = 0.5919; b2 = 1.4350
  labs(title = "Produção de Veículos vs. Produção Industrial Geral",
       subtitle = "Dados mensais de jan/2002 até jun/2021",
       x = "Produção de Veículos (÷ 1.000)",
       y = "Produção Industrial Geral (números-índice)",
       caption = "Fonte: ANFAVEA e IBGE-SIDRA. Elaboração do Autor.") + theme_light()
```



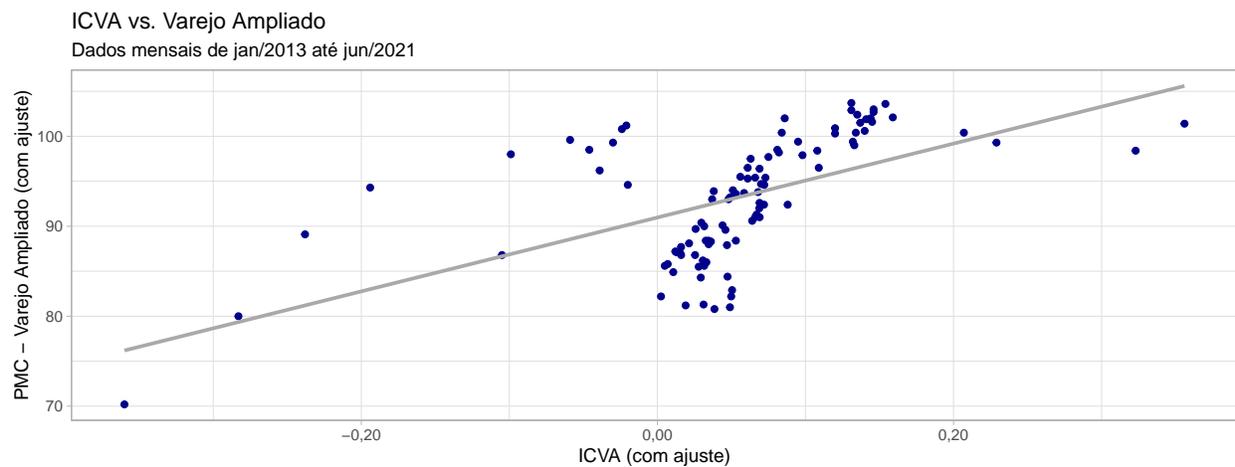
4. Construa um gráfico de correlação entre o ICVA e a Pesquisa Mensal do Comércio, utilizando para isso as medidas que julgar mais adequada.

```
#puxando o arquivo .xlsx do site da Cielo e baixando em nosso diretorio
url <- "https://apicatalog.mziq.com/filemanager/v2/d/4d1ebe73-b068-4443-992a-3d72d573238c/3e864198-0b72"
download.file(url, destfile = "icva.xlsx", mode = "wb")

#lendo o arquivo baixado
icva <- read_excel("icva.xlsx")
colnames(icva) <- c("date", "icva_sa", "icva_ca", "icva_real_sa", "icva_real_ca")

#juntando em um unico dataframe com volume do varejo ampliado (ambos com ajuste)
dados_icva_vamp <- left_join(icva, vol_varejo_amp_ca, by = "date") %>%
  select(c(date, icva_ca, varejo_ampliado_ca))
```

```
dados_icva_vamp %>%
  ggplot(aes(x = icva_ca, y = varejo_ampliado_ca)) +
  geom_point(color = "darkblue") +
  geom_smooth(method = "lm", se = F, colour = "darkgrey") +
  scale_x_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    decimal.mark = ',')) +
  labs(title = "ICVA vs. Varejo Ampliado",
       subtitle = "Dados mensais de jan/2013 até jun/2021",
       x = "ICVA (com ajuste)",
       y = "PMC - Varejo Ampliado (com ajuste)",
       caption = "Fonte: Cielo e IBGE-SIDRA. Elaboração do Autor.") + theme_light()
```



5. Coloque no mesmo gráfico a variação acumulada em 12 meses do IBC-Br e a variação acumulada em quatro trimestres do PIB.

```
#puxando dados do ibc-br do BCB
ibc_br <- BETSget(24363, data.frame = T) %>%
  rename(ibc_br = value)

#calculando a variacao acumulada em 12m do ibc-br
var_12m_ibc_br <- ibc_br %>%
  mutate(var_12m_ibc_br = acum_i(ibc_br, 12)) %>%
  select(c(date, var_12m_ibc_br)) %>%
  na.omit()

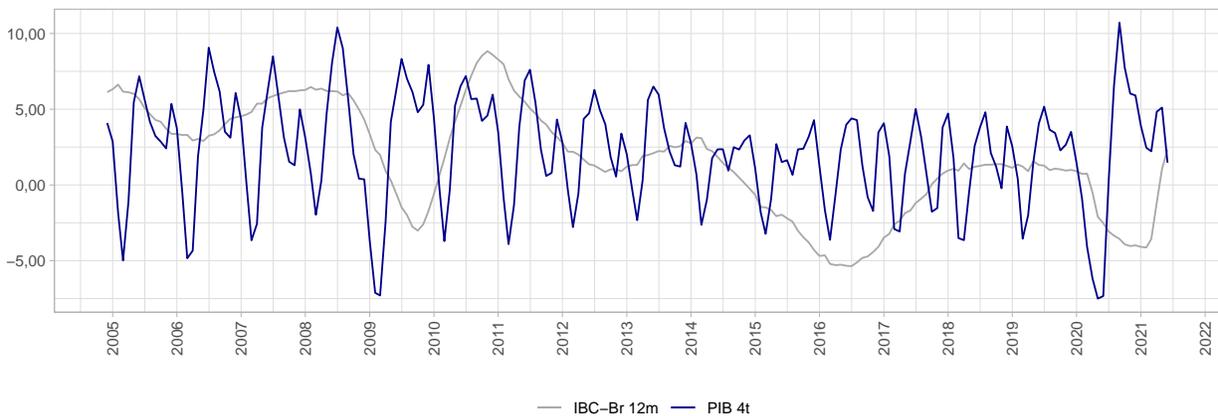
#puxando dados do pib do IPEA
pib <- ipeadata("BM12_PIB12", language = "br", quiet = T) %>%
  select(c(date, value)) %>%
  rename(pib = value)

#calculando a variacao acumulada em 4t do pib
var_4t_pib <- pib %>%
  mutate(var_4t_pib = acum_i(pib, 3)) %>%
  select(c(date, var_4t_pib)) %>%
  na.omit()

#unindo as variacoes em um mesmo dataframe
dados_pib_ibc_br <- left_join(var_12m_ibc_br, var_4t_pib, by = "date")

#criando o grafico do acumulado IBC-Br 12m e PIB 4t
dados_pib_ibc_br %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = var_12m_ibc_br, colour = "IBC-Br 12m"), size = 0.5) +
  geom_line(aes(y = var_4t_pib, colour = "PIB 4t"), size = 0.5) +
  scale_x_date(breaks = "1 year",
              labels = scales::date_format("%Y")) +
  scale_color_manual(values = c("IBC-Br 12m" = "darkgrey",
                              "PIB 4t" = "darkblue")) +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    decimal.mark = ',')) +
  labs(title = "Variação acumulada IBC-Br 12m vs. PIB 4t",
       x = element_blank(),
       y = element_blank(),
       caption = "Fonte: IPEADATA, BCB, IBGE. Elaboração do Autor.") +
  theme_light() + theme(legend.title = element_blank(),
                      legend.position = "bottom",
                      axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```

Varição acumulada IBC-Br 12m vs. PIB 4t



Fonte: IPEADATA, BCB, IBGE. Elaboração do Autor.

6. Baixe os dados da PEA e da População Desocupada a partir da base da PNAD Contínua contida no SIDRA/IBGE com o pacote `sidrar`. Você deve criar a taxa de desemprego com esses dados. Em seguida, utilizando o pacote `seasonal`, você deve fazer um ajuste sazonal na taxa de desemprego. Por fim, crie um gráfico de linha que contenha a taxa de desemprego e a taxa de desemprego com ajuste sazonal.

```
#puxando dados da pea da pnad continua
pea <- get_sidra(api = "/t/4092/n1/all/v/1641/p/all/c629/32386") %>%
  mutate(date = parse_date(`Trimestre (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
  rename(pea = Valor) %>%
  as.data.frame()

#puxando dados da populacao ocupada da pnad continua
pop_ocup <- get_sidra(api = "/t/4092/n1/all/v/1641/p/all/c629/32387") %>%
  mutate(date = parse_date(`Trimestre (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
  rename(pop_ocup = Valor) %>%
  as.data.frame()

#puxando dados da populacao desocupada da pnad continua
pop_desocup <- get_sidra(api = "/t/4092/n1/all/v/1641/p/all/c629/32446") %>%
  mutate(date = parse_date(`Trimestre (Código)`, format = "%Y%m")) %>%
  select(date, Valor) %>%
  rename(pop_desocup = Valor) %>%
  as.data.frame()

#juntando em um unico dataframe e calculando a taxa de desemprego
dados_pnad <- left_join(pea, pop_ocup, by = "date") %>%
  cbind(pop_desocup$pop_desocup) %>%
  rename(pop_desocup = `pop_desocup$pop_desocup`) %>%
  mutate(tx_desemp = (pea - pop_ocup) / pea * 100)

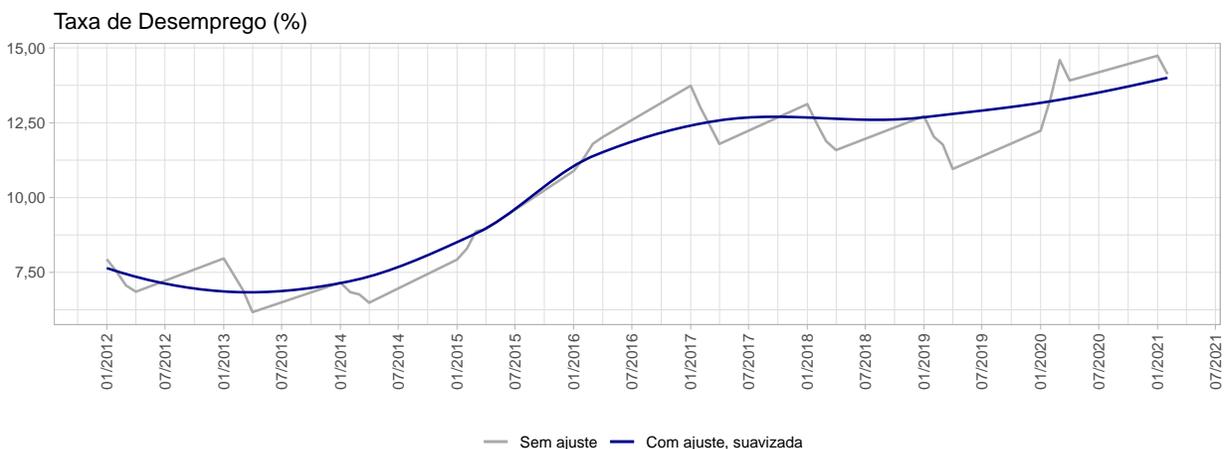
#dessazonalizando a serie de desemprego
tx_desemp_ts <- ts(dados_pnad$tx_desemp, freq = 12, start = c(2012, 1))
tx_desemp_ca_seas <- seas(tx_desemp_ts, x11 = "")$series$d11 %>%
  as.data.frame()
```

```

#dataframe com a serie com e sem ajuste
taxas <- cbind(dados_pnad, tx_desemp_ca_seas) %>%
  select(c(date, tx_desemp, x)) %>%
  rename(tx_desemp_ca = x)

#criando o grafico das taxas com e sem ajuste
taxas %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = tx_desemp, colour = "Sem ajuste"), size = 0.7) +
  geom_smooth(aes(y = tx_desemp_ca, colour = "Com ajuste, suavizada"), size = 0.7, se = F) +
  scale_x_date(breaks = "6 months",
              labels = scales::date_format("%m/%Y")) +
  scale_color_manual(values = c("Sem ajuste" = "darkgrey",
                                "Com ajuste, suavizada" = "darkblue")) +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    decimal.mark = ',')) +
  labs(title = "Taxa de Desemprego (%)",
       x = element_blank(),
       y = element_blank(),
       caption = "Fonte: IBGE - PNAD Contínua. Elaboração do Autor.") +
  theme_light() + theme(legend.title = element_blank(),
                       legend.position = "bottom",
                       axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))

```



Fonte: IBGE – PNAD Contínua. Elaboração do Autor.

7. Utilizando os dados do Novo CAGED, crie um gráfico de barras que mostre o avanço das demissões e admissões ao longo de 2020.

```

#puxando dados das admissoes do Novo CAGED
admissoes_novocaged <- ipeadata("CAGED12_ADMISN12", language = "br", quiet = T) %>%
  select(c(date, value)) %>%
  rename(admissoes = value)

#puxando dados das demissoes do Novo CAGED
demissoes_novocaged <- ipeadata("CAGED12_DESLIGN12", language = "br", quiet = T) %>%

```

```
select(c(date, value)) %>%
  rename(demissoes = value)
```

*#juntando em um unico dataframe*

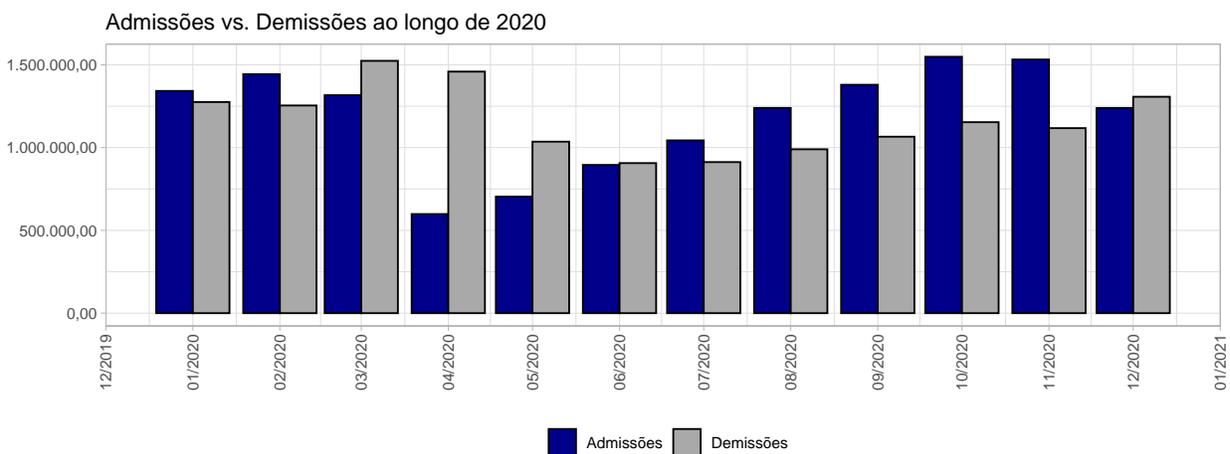
```
dados_novocaged <- inner_join(admissoes_novocaged, demissoes_novocaged, by = "date")
```

*#transformando nossos dados para o formato long*

```
dados_novocaged_long <- gather(dados_novocaged,
                                key = variavel,
                                value = value, admissoes, demissoes)
```

*#criando um grafico de barras do novo caged*

```
dados_novocaged_long %>%
  filter(date >= "2020-01-01" & date <= "2020-12-01") %>%
  ggplot(aes(x = date, y = value, fill = variavel)) +
  geom_bar(stat = "identity", position = "dodge", color = "black") +
  scale_x_date(breaks = "1 month",
               labels = scales::date_format("%m/%Y")) +
  scale_fill_manual(labels = c("Admissões", "Demissões"),
                    values = c("admissoes" = "darkblue",
                                "demissoes" = "darkgray")) +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    big.mark = ".",
                                                    decimal.mark = ',')) +
  labs(title = "Admissões vs. Demissões ao longo de 2020",
        x = element_blank(),
        y = element_blank(),
        caption = "Fonte: Novo CAGED. Elaboração do Autor.") +
  theme_light() + theme(legend.title = element_blank(),
                        legend.position = "bottom",
                        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



8. Utilizando os dados do CAGED antigo, crie um gráfico com o saldo de admissões e demissões desazonalizado, com o pacote seasonal.

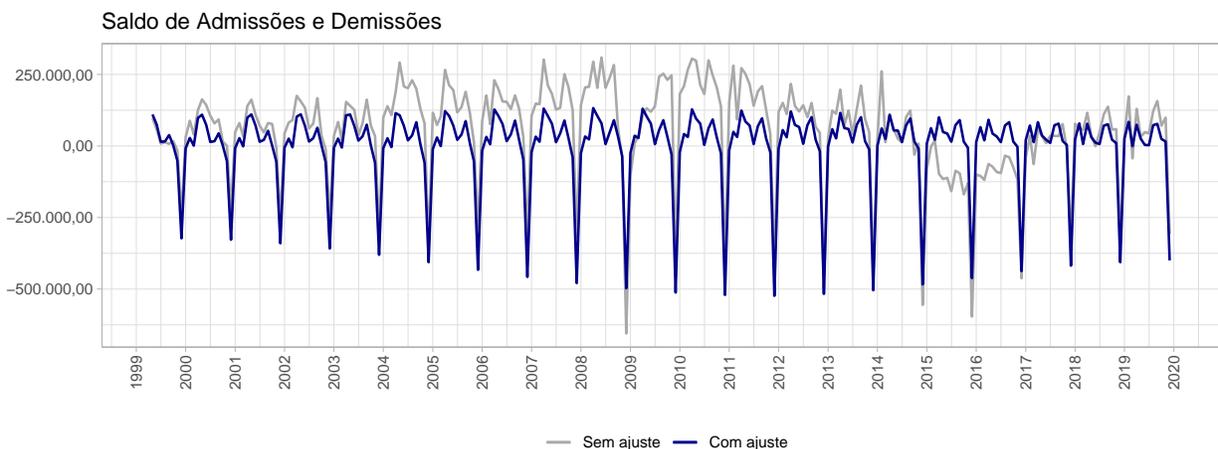
```
#puxando dados dos saldos das admissões vs. demissões do CAGED Antigo
saldo_cagedantigo <- ipeadata("CAGED12_SALDO12", language = "br", quiet = T) %>%
  select(c(date, value)) %>%
  rename(saldo = value)

#dessazonalizando a série de saldo do CAGED Antigo
saldo_ts <- ts(saldo_cagedantigo$saldo, freq = 12, start = c(1999, 5))
saldo_seas <- seas(saldo_ts, x11 = "")$series$d10 %>%
  as.data.frame()

#juntando em um unico dataframe, com e sem ajuste
saldos <- cbind(saldo_cagedantigo, saldo_seas) %>%
  rename(saldo_sa = saldo, saldo_ca = x)
```

```
#criando um grafico dos saldos do caged antigo
saldos %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = saldo_sa, colour = "Sem ajuste"), size = 0.7) +
  geom_line(aes(y = saldo_ca, colour = "Com ajuste"), size = 0.7) +
  scale_x_date(breaks = "1 year",
              labels = scales::date_format("%Y")) +
  scale_color_manual(values = c("Sem ajuste" = "darkgrey",
                                "Com ajuste" = "darkblue")) +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    big.mark = ".",
                                                    decimal.mark = ',')) +

  labs(title = "Saldo de Admissões e Demissões",
       x = element_blank(),
       y = element_blank(),
       caption = "Fonte: CAGED Antigo. Elaboração do Autor.") +
  theme_light() + theme(legend.title = element_blank(),
                       legend.position = "bottom",
                       axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



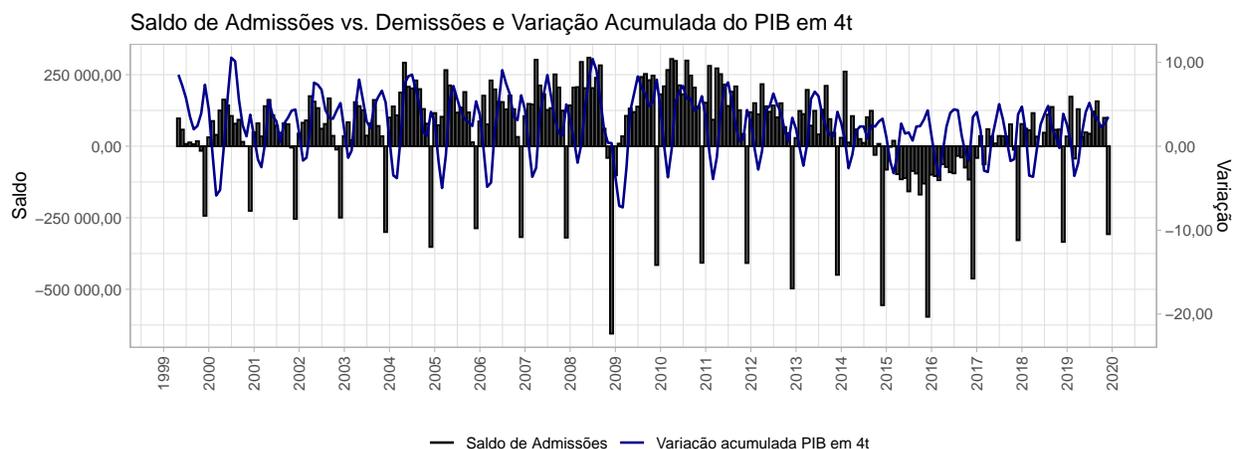
Fonte: CAGED Antigo. Elaboração do Autor.

9. Plote no mesmo gráfico os dados do saldo do CAGED antigo com a variação acumulada em quatro trimestres do PIB.

```
#juntando a serie de variacao acumulada do pib em 4t e saldo do caged antigo
dados_pib_caged <- left_join(var_4t_pib, saldo_cagedantigo, by = "date") %>%
  na.omit()

#criando o fator de proporcao para escala
fator <- max(dados_pib_caged$var_4t_pib) / max(dados_pib_caged$saldo)

#criando um grafico com os saldos de admissoes e acumulado PIB 4t
dados_pib_caged %>%
  ggplot(aes(x = date)) +
  geom_bar(aes(y = saldo, colour = "Saldo de Admissões"),
           stat = "identity", position = "dodge", color = "black") +
  geom_line(aes(y = var_4t_pib/fator, colour = "Variação acumulada PIB em 4t"),
           size = 0.7) +
  scale_x_date(breaks = "1 year",
              labels = scales::date_format("%Y")) +
  scale_colour_manual(values = c("Saldo de Admissões" = "black",
                                "Variação acumulada PIB em 4t" = "darkblue")) +
  scale_y_continuous("Saldo",
                    sec.axis = sec_axis(
                      ~. * fator, name = "Variação",
                      labels = scales::number_format(accuracy = 0.01,
                                                       decimal.mark = ',')),
                    labels = scales::number_format(accuracy = 0.01,
                                                       decimal.mark = ',')) +
  labs(title = "Saldo de Admissões vs. Demissões e Variação Acumulada do PIB em 4t",
       x = element_blank(),
       y = element_blank(),
       caption = "Fonte: Novo CAGED e BCB. Elaboração do Autor.") +
  theme_light() + theme(legend.title = element_blank(),
                       legend.position = "bottom",
                       axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



Fonte: Novo CAGED e BCB. Elaboração do Autor.

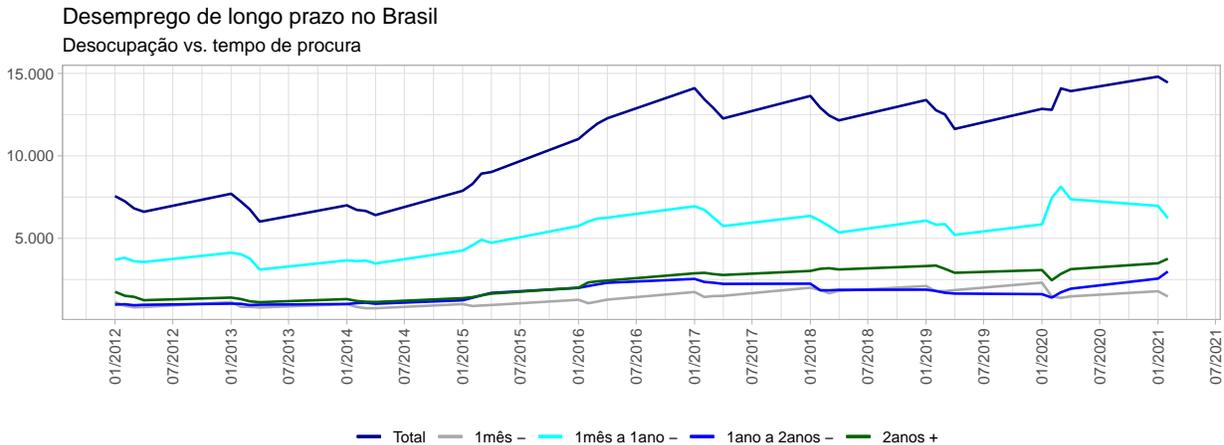
10. Você deve criar um gráfico de correlação entre os pedidos de seguro desemprego e a taxa de desemprego medida pela PNAD Contínua.

Não foi possível o acesso ao site do Ministério do Trabalho (<http://pdet.mte.gov.br/seguro-desemprego>) para a obtenção da série, sendo impossibilitado de resolver a questão.

11. A PNAD Contínua, no seu corte trimestral, traz informações relevantes sobre o desemprego de longo prazo. A tabela 1616 do SIDRA/IBGE, por exemplo, relaciona a desocupação por tempo de procura de trabalho. Pedese que você crie um gráfico de linha que relacione as diferentes medidas de tempo de procura de trabalho.

```
#puxando os dados da relacao de desocupados por tempo de procura e tratando os dados
tempo_procura <- get_sidra(api = "/t/1616/n1/all/v/4092/p/all/c1965/all") %>%
  mutate(date = parse_date(`Trimestre (Código)`, format = "%Y%m")) %>%
  select(date, "Tempo de procura de trabalho", Valor) %>%
  spread("Tempo de procura de trabalho", Valor) %>%
  rename(tp_2anos_mais = `2 anos ou mais`,
         tp_1ano_2anos = `De 1 ano a menos de 2 anos`,
         tp_1mes_1ano = `De 1 mês a menos de 1 ano`,
         tp_1mes_menos = `Menos de 1 mês`,
         tp_total = `Total`)
```

```
#criando um grafico de desempregados vs. tempo de procura
tempo_procura %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = tp_total, colour = "Total"), size = 0.7) +
  geom_line(aes(y = tp_1mes_menos, colour = "1mês -"), size = 0.7) +
  geom_line(aes(y = tp_1mes_1ano, colour = "1mês a 1ano -"), size = 0.7) +
  geom_line(aes(y = tp_1ano_2anos, colour = "1ano a 2anos -"), size = 0.7) +
  geom_line(aes(y = tp_2anos_mais, colour = "2anos +"), size = 0.7) +
  scale_x_date(breaks = "6 months",
              labels = scales::date_format("%m/%Y")) +
  scale_color_manual(values = c("Total" = "darkblue",
                                "1mês -" = "darkgray",
                                "1mês a 1ano -" = "cyan",
                                "1ano a 2anos -" = "blue",
                                "2anos +" = "darkgreen")) +
  scale_y_continuous(labels = scales::number_format(big.mark = ".")) +
  labs(title = "Desemprego de longo prazo no Brasil",
       subtitle = "Desocupação vs. tempo de procura",
       x = element_blank(),
       y = element_blank(),
       caption = "Fonte: IBGE-SIDRA. Elaboração do Autor.") +
  theme_light() + theme(legend.title = element_blank(),
                       legend.position = "bottom",
                       axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



Fonte: IBGE-SIDRA. Elaboração do Autor.

12. (Questão Desafio) O pacote `gtrendsR` pode ser utilizado para pegar os famosos dados do Google Trends. Com o código abaixo, por exemplo, nós pegamos os dados das palavras chaves `seguro desemprego` e `emprego` para o Brasil.

```
#pesquisa da palavra "seguro desemprego" no google
gtrends_seg_desemp <- gtrends(keyword = "seguro desemprego",
                              geo = "BR",
                              time = "all",
                              onlyInterest = T)$interest_over_time %>%

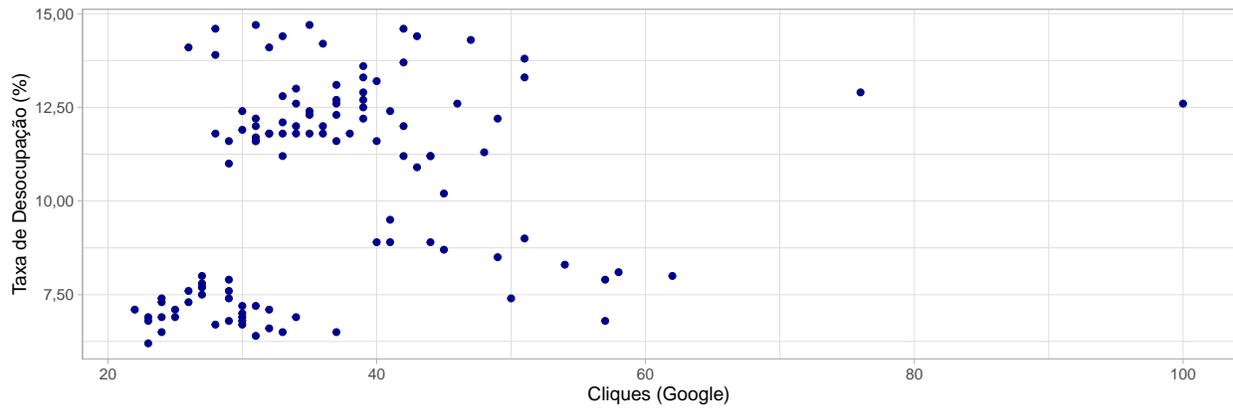
  select(c(date, hits))

#puxando a serie da tx desocupacao pnad continua
tx_desocup <- get_sidra(api = "/t/6381/n1/all/v/4099/p/all/d/v4099%201") %>%
  mutate(date = parse_date(`Trimestre Móvel (Código)`, format = "%Y%m")) %>%
  select(c(date, Valor))

#juntando em um unico dataframe
dados_sd_td <- inner_join(gtrends_seg_desemp, tx_desocup, by = "date") %>%
  rename(cliques_gtrends = hits, tx_desocup = Valor)

#criando um grafico da taxa de desemprego vs. procura do termo 'seguro desemprego'
dados_sd_td %>%
  ggplot(aes(x = cliques_gtrends, y = tx_desocup)) +
  geom_point(color = "darkblue") +
  scale_y_continuous(labels = scales::number_format(accuracy = 0.01,
                                                    decimal.mark = ",")) +
  labs(title = "Procura pelo termo 'seguro desemprego' vs. Taxa de Desocupação (%)",
       subtitle = "Dados mensais de mar/2012 até mai/2021",
       x = "Cliques (Google)",
       y = "Taxa de Desocupação (%)",
       caption = "Fonte: Google e IBGE-SIDRA. Elaboração do Autor.") + theme_light()
```

Procura pelo termo 'seguro desemprego' vs. Taxa de Desocupação (%)  
Dados mensais de mar/2012 até mai/2021



Fonte: Google e IBGE-SIDRA. Elaboração do Autor.