

title: "Case 4Intelligence

```
#install.packages("readxl")
#install.packages("tidyverse")
#install.packages("tidyr")
#install.packages("dplyr")
#install.packages("zoo")
#install.packages("seasonal")

library(readxl)

## Warning: package 'readxl' was built under R version 4.2.2

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.2.2

## — Attaching packages —————
tidyverse 1.3.2 —
## ✓ ggplot2 3.3.6      ✓ purrr   0.3.4
## ✓ tibble  3.1.8      ✓ dplyr  1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.4.1
## ✓ readr   2.1.3      ✓ forcats 0.5.2

## Warning: package 'tidyr' was built under R version 4.2.2
## Warning: package 'dplyr' was built under R version 4.2.2

## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()

library(tidyr)
library(dplyr)
library(zoo)

## Warning: package 'zoo' was built under R version 4.2.2

##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

library(seasonal)

## Warning: package 'seasonal' was built under R version 4.2.2
```

```
##
## Attaching package: 'seasonal'
##
## The following object is masked from 'package:tibble':
##
##     view
```

— A base de dados fornecida é um arquivo de formato xlsx, que contém três sheets, Diesel, Gasolina e Etanol, respectivamente. Antes de iniciar a análise de visualização é necessário realizar uma etapa de pré-processamento. Iremos importar cada um dos sheets do arquivo em formato de data frame, assim, separando-os por tipo de combustível e facilitando a análise. Após isso, é necessário transformar a estrutura dos dados no formato apropriado para séries temporais, isto é, será necessário “empilha-los”, também conhecido como tidy data.

```
df_diesel <-
read_xlsx("dados_desafiodatascientistintern_vendas_distribuidoras_anp.xlsx")
df_gasolina <-
read_xlsx("dados_desafiodatascientistintern_vendas_distribuidoras_anp.xlsx", sheet = "gasolina")
df_etanol <-
read_xlsx("dados_desafiodatascientistintern_vendas_distribuidoras_anp.xlsx", sheet = "etanol")

df_diesel_tidy <- gather(df_diesel, Ano, Diesel, -regiao, -meses)
df_gasolina_tidy <- gather(df_gasolina, Ano, Gasolina, -regiao, -meses)
df_etanol_tidy <- gather(df_etanol, Ano, Etanol, -regiao, -meses)

df_diesel_tidy$Data <- as.yearmon(paste(df_diesel_tidy$Ano,
df_diesel_tidy$meses), "%Y %m")
df_gasolina_tidy$Data <- as.yearmon(paste(df_gasolina_tidy$Ano,
df_gasolina_tidy$meses), "%Y %m")
df_etanol_tidy$Data <- as.yearmon(paste(df_etanol_tidy$Ano,
df_etanol_tidy$meses), "%Y %m")

df_diesel_tidy <- df_diesel_tidy[, -2:-3]
df_gasolina_tidy <- df_gasolina_tidy[, -2:-3]
df_etanol_tidy <- df_etanol_tidy[, -2:-3]

str(df_diesel_tidy)

## tibble [2,268 × 3] (S3: tbl_df/tbl/data.frame)
## $ regiao: chr [1:2268] "df" "df" "df" "df" ...
## $ Diesel: num [1:2268] 23311 25968 24591 24976 27225 ...
## $ Data : 'yearmon' num [1:2268] jan 2000 fev 2000 mar 2000 abr 2000
...

str(df_gasolina_tidy)
```

```
## tibble [2,268 × 3] (S3: tbl_df/tbl/data.frame)
## $ regioao : chr [1:2268] "df" "df" "df" "df" ...
## $ Gasolina: num [1:2268] 39258 49570 48423 48553 51085 ...
## $ Data    : 'yearmon' num [1:2268] jan 2000 fev 2000 mar 2000 abr 2000
2000 ...

str(df_etanol_tidy)

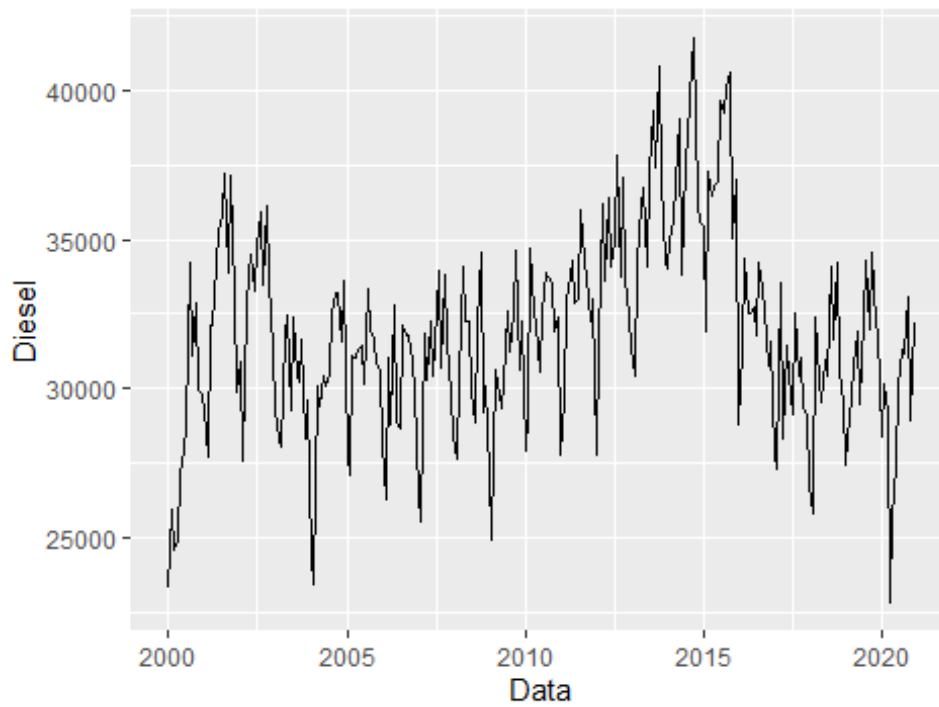
## tibble [2,268 × 3] (S3: tbl_df/tbl/data.frame)
## $ regioao: chr [1:2268] "df" "df" "df" "df" ...
## $ Etanol: num [1:2268] 11688 13408 13389 12681 13047 ...
## $ Data   : 'yearmon' num [1:2268] jan 2000 fev 2000 mar 2000 abr 2000
...

df_diesel_tidy$Data <- as.Date(df_diesel_tidy$Data)
df_gasolina_tidy$Data <- as.Date(df_gasolina_tidy$Data)
df_etanol_tidy$Data <- as.Date(df_etanol_tidy$Data)
```

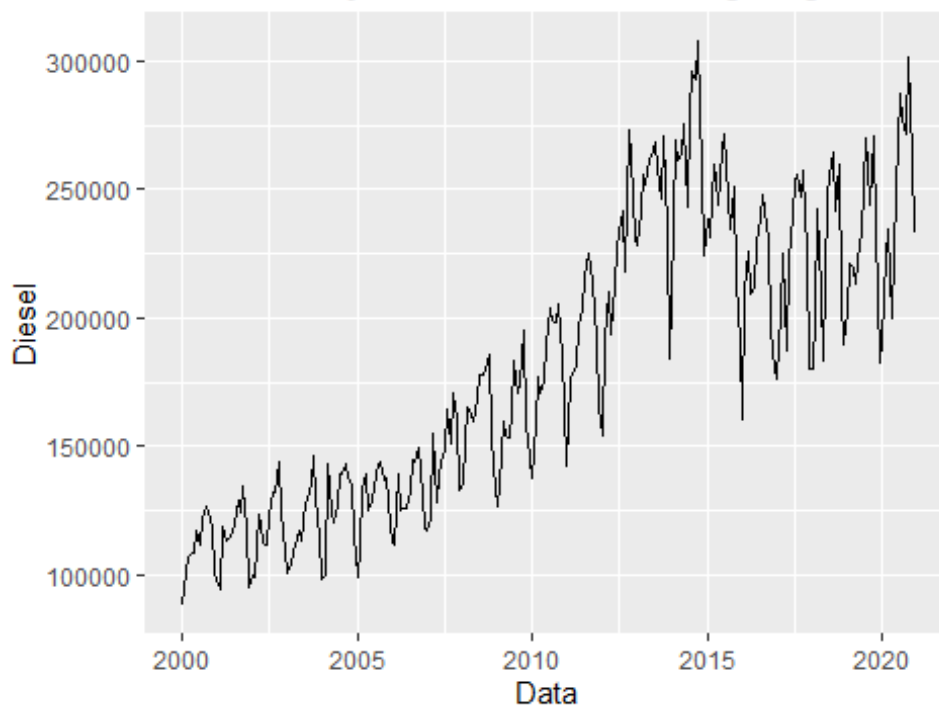
— Com a base de dados pronta, partiremos para a etapa de visualização dos dados. Com o objetivo de melhor entendimento do mercado de combustíveis, iremos criar gráficos para cada UF disponível no arquivo, além do nível nacional (br). Analisando as séries em nível nacional, para um entendimento geral do mercado de combustíveis, pode-se notar um comportamento semelhante entre as séries, com um forte componente sazonal no último trimestre do ano, sobretudo em dezembro/novembro, onde a venda de combustíveis costuma atingir seu máximo anual, e também um componente sazonal de baixa no primeiro trimestre do ano, sobretudo em janeiro/fevereiro, onde normalmente as vendas de combustível no Brasil atingem a mínima anual. É interessante destacar que para o álcool e gasolina, o pico anual costuma ocorrer em dezembro, provavelmente por conta da característica do mercado, onde esses combustíveis são majoritariamente utilizados em veículos de passeio, assim, dezembro costuma ser um mês de férias, onde muitas famílias viajam, o que pode explicar essa sazonalidade. Já com relação ao Diesel, nota-se que essa alta ocorre novembro/outubro, infere-se que essa alta acompanha o movimento do setor industrial e varejista que costuma se preparar para as vendas de black friday e vendas de final de ano. Essa sazonalidade é identificada também no comportamento dos estados, apesar disso, os estados possuem bastante diferenciação com relação à variância das observações, por exemplo, o diesel no estado do Pará possui uma amplitude visualmente identificada bem menor do que o estado de São Paulo, por exemplo.

```
#série original Diesel
for (i in unique(df_diesel_tidy$regiao)) {
  print(ggplot(df_diesel_tidy %>% filter(regiao == i), aes(Data, Diesel))
  +
    geom_line() + ggtitle(paste("Evolução Vendas Diesel - Região",
i, sep = " ")) + theme(plot.title = element_text(hjust = 0.5)))
}
```

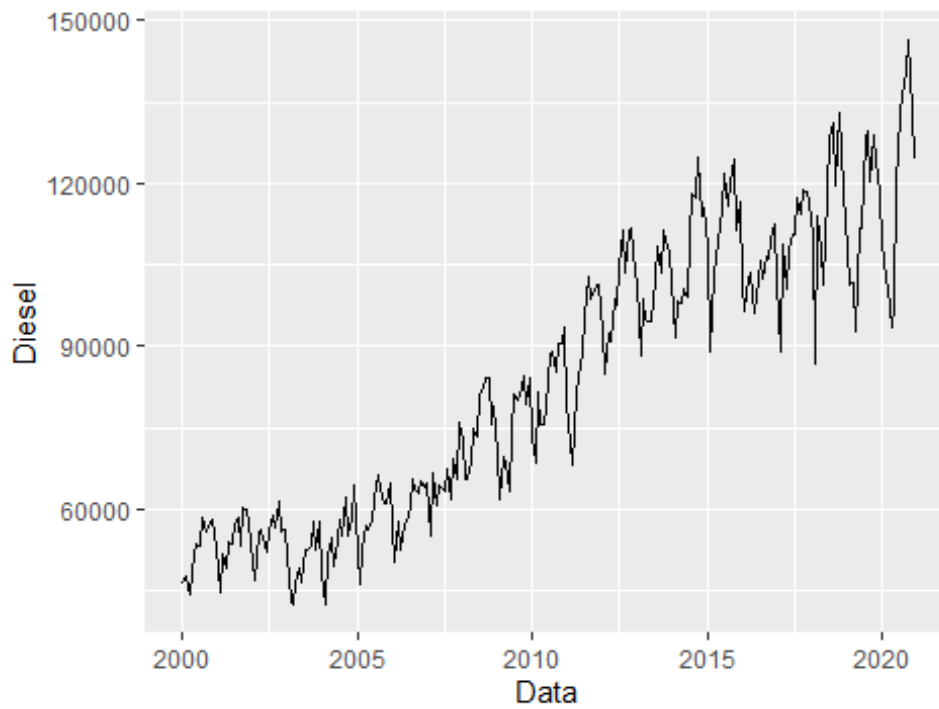
Evolução Vendas Diesel - Região df



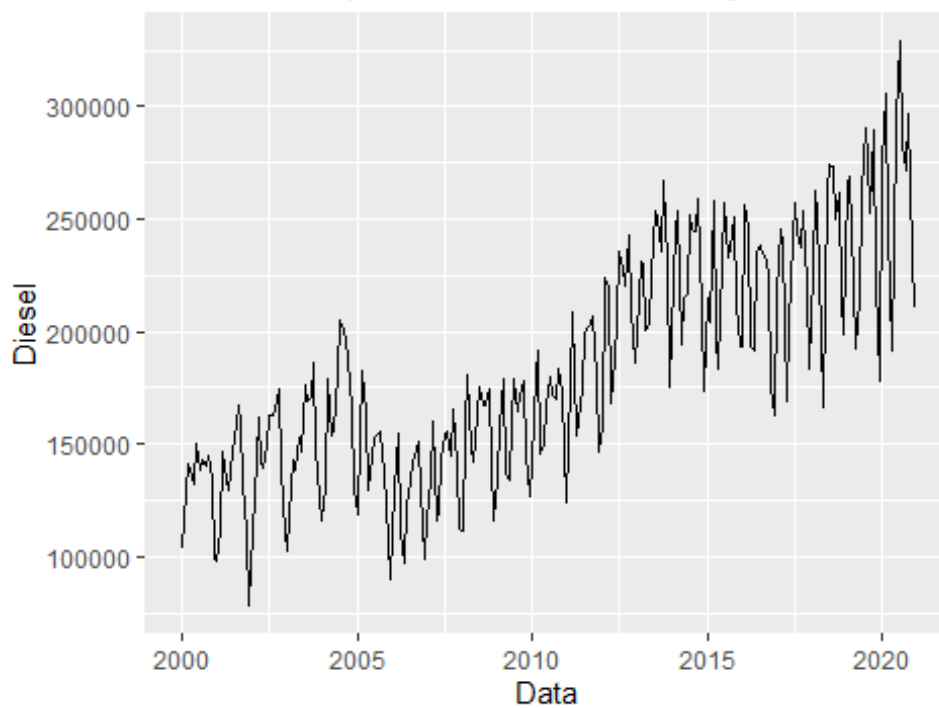
Evolução Vendas Diesel - Região go



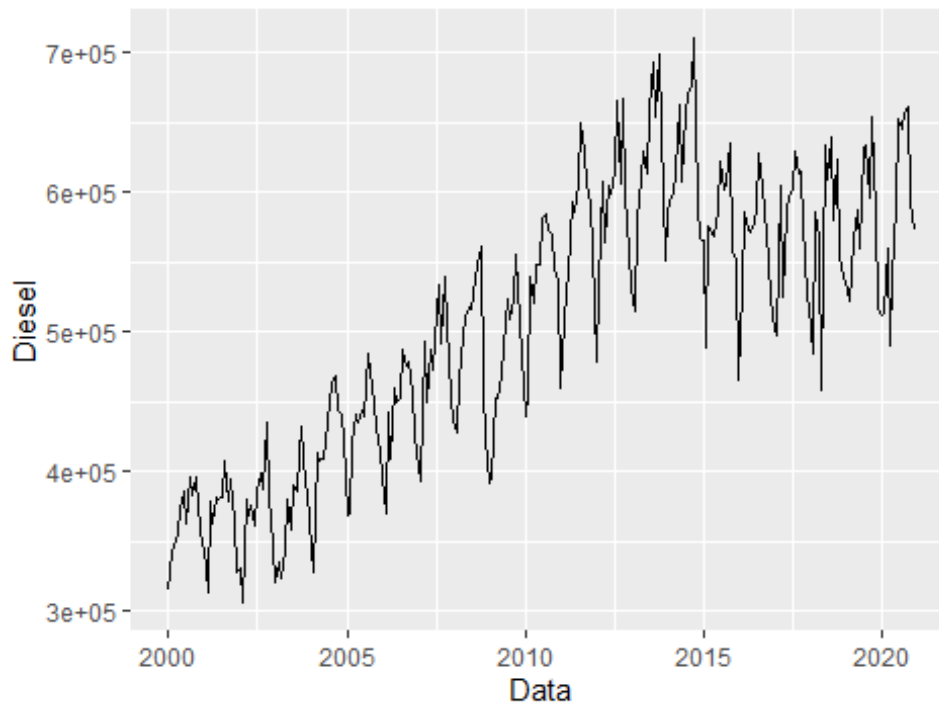
Evolução Vendas Diesel - Região ma



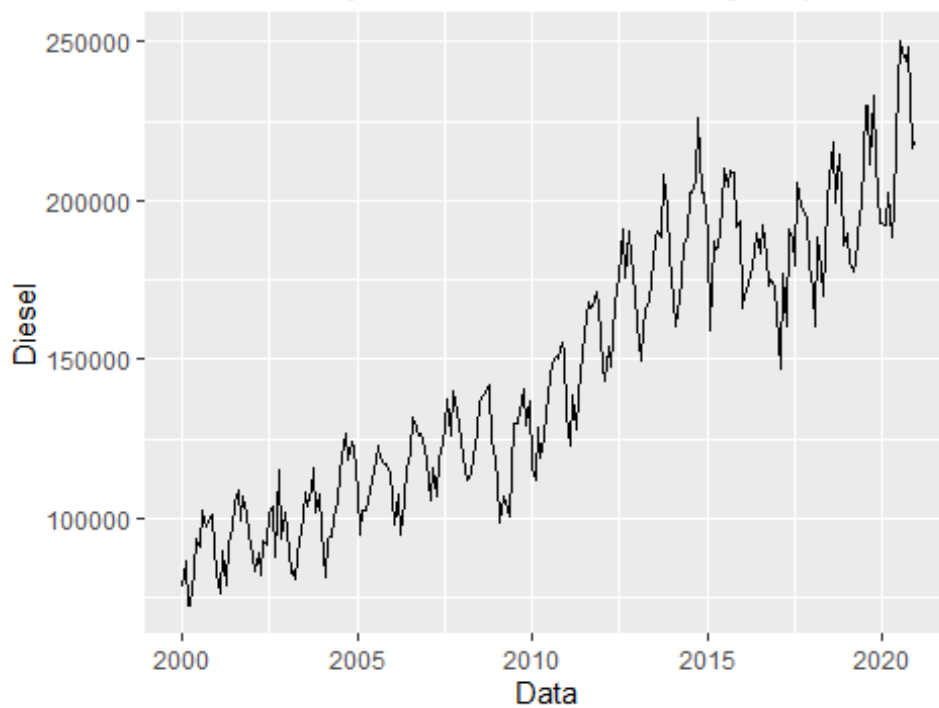
Evolução Vendas Diesel - Região mt



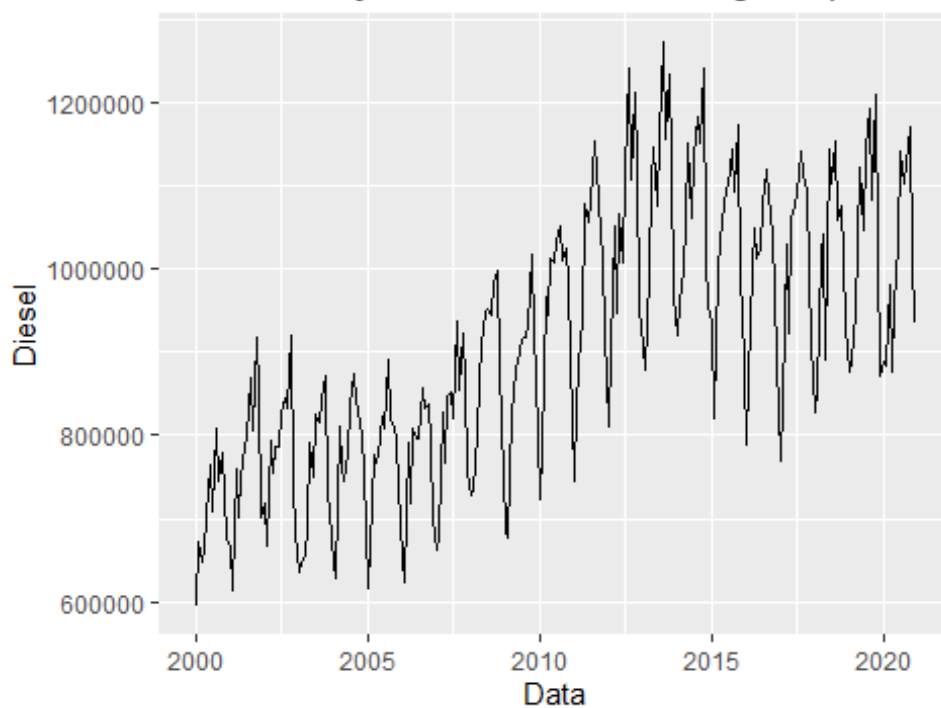
Evolução Vendas Diesel - Região mg



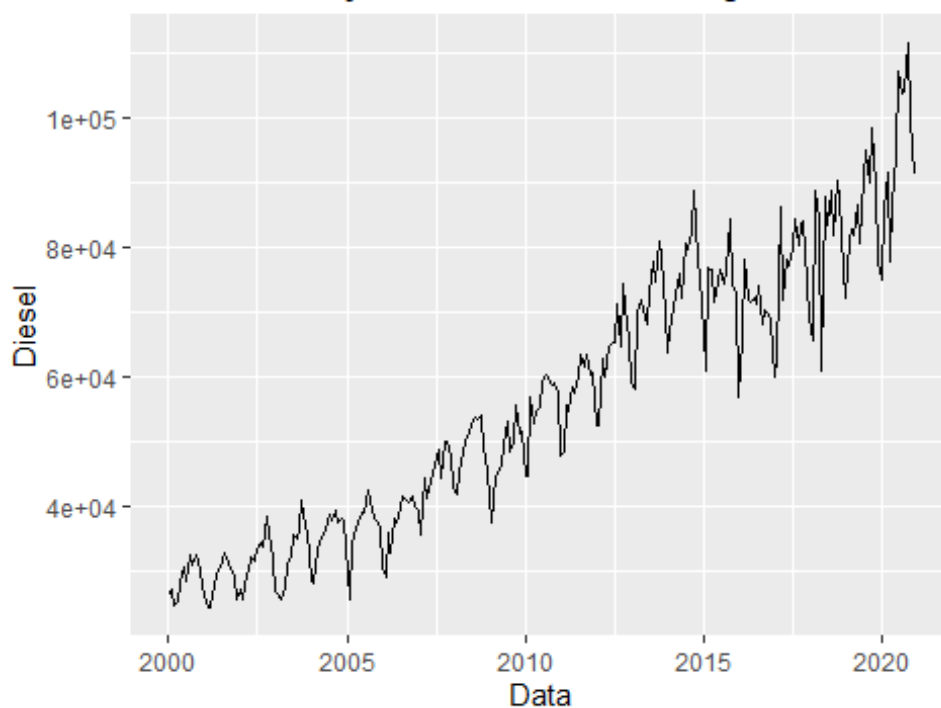
Evolução Vendas Diesel - Região pa

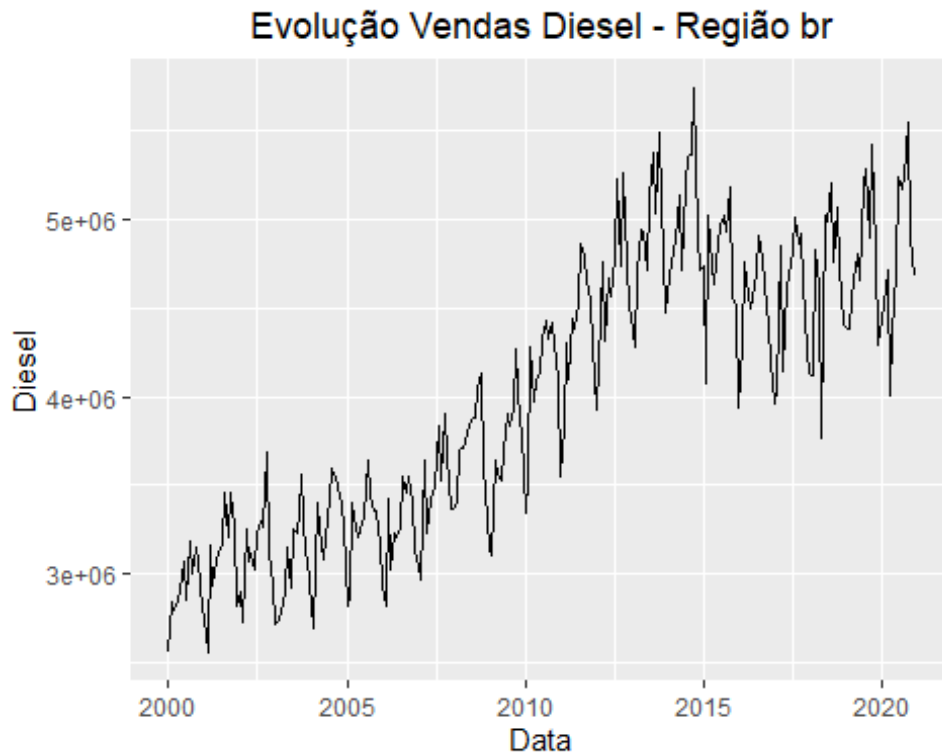


Evolução Vendas Diesel - Região sp



Evolução Vendas Diesel - Região to

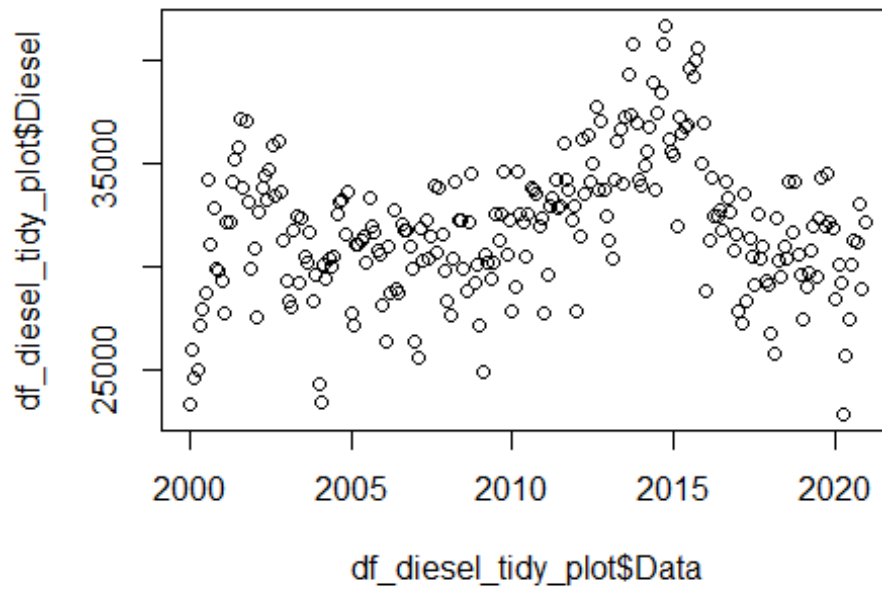




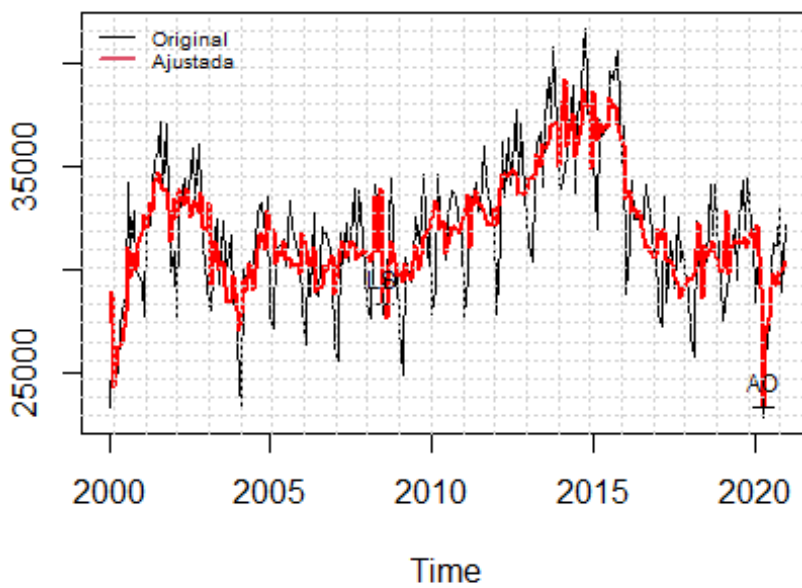
```
#Por região: série original vs ajuste sazonal (Diesel)
for (i in unique(df_diesel_tidy$regiao)) {
  df_diesel_tidy_plot <- df_diesel_tidy %>% filter(regiao == i)
  plot(df_diesel_tidy_plot$Data, df_diesel_tidy_plot$Diesel, main =
paste("Evolução Vendas Diesel - Região", i, sep = " "))

  regio_br <- ts(df_diesel_tidy_plot$Diesel, start = c(2000, 1) ,
frequency = 12)
  dessaz_br <- seas(regio_br, x11 = "")
  plot(dessaz_br, main = paste("Evolução Vendas Diesel com Ajuste Sazonal
- Região", i, sep = " "))
  grid(nx = 23)
  legend("topleft", legend = c("Original", "Ajustada"), col = c(1, 2),
lwd = c(1, 2),
      lty = 1, bty = "n", cex = 0.6)
}
```

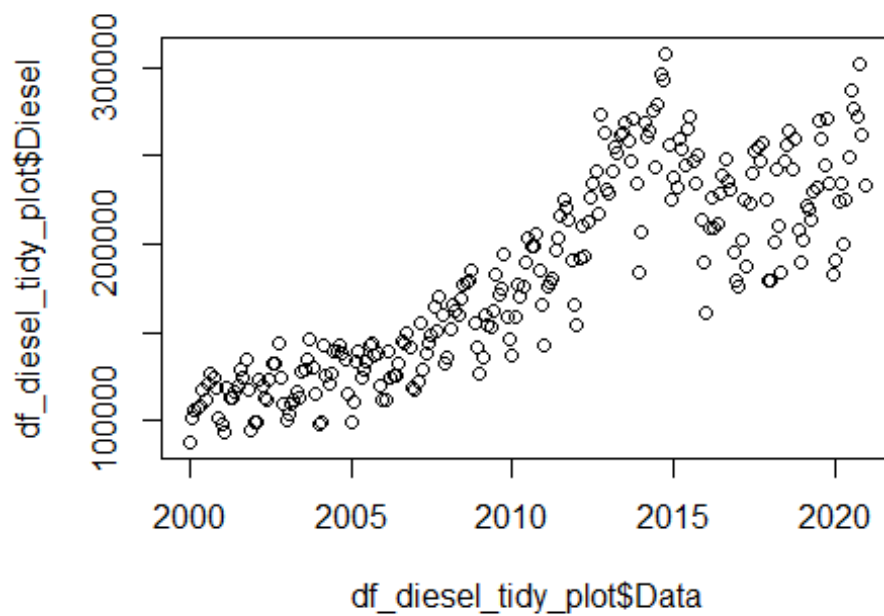

Evolução Vendas Diesel - Região df



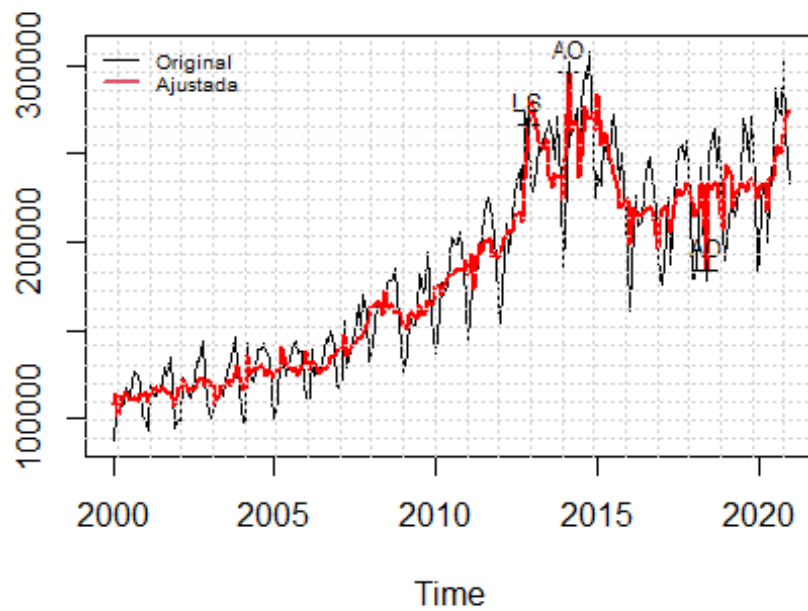
Evolução Vendas Diesel com Ajuste Sazonal - Regiã



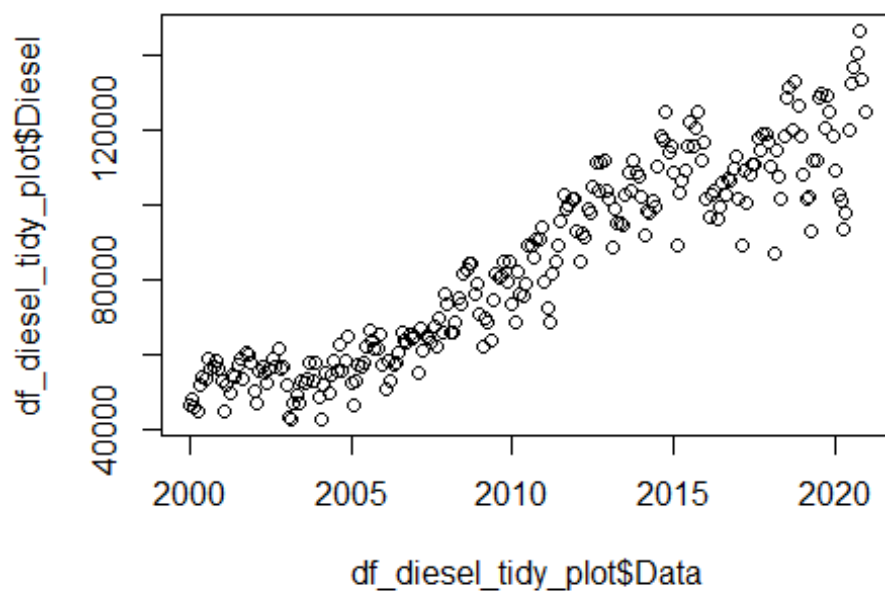
Evolução Vendas Diesel - Região go



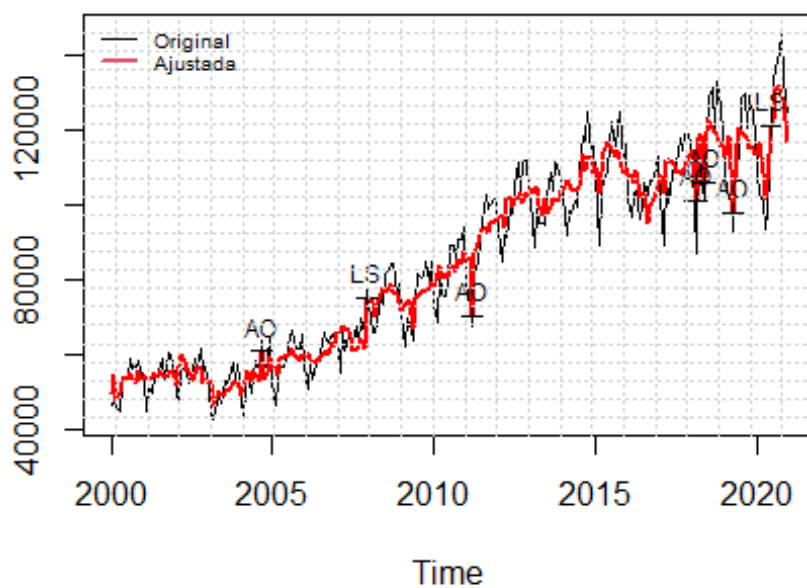
Evolução Vendas Diesel com Ajuste Sazonal - Região



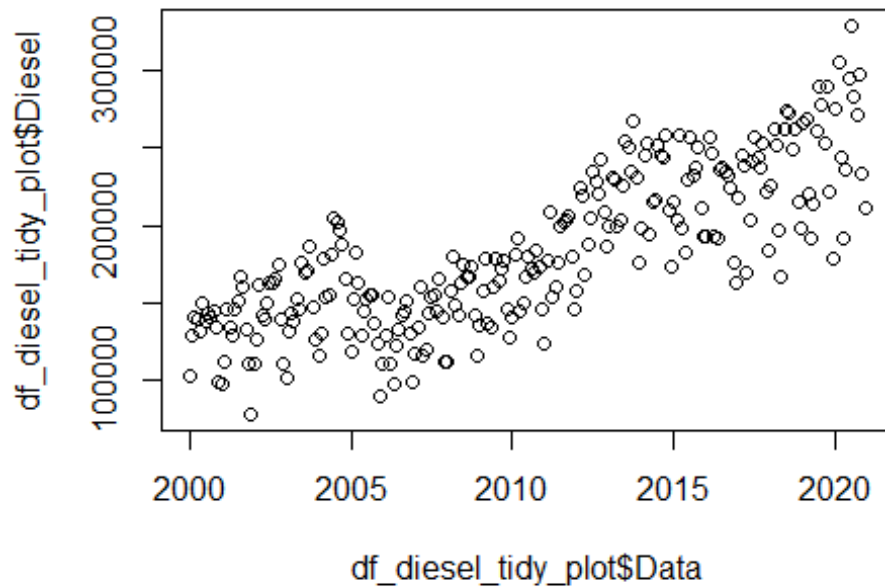
Evolução Vendas Diesel - Região ma



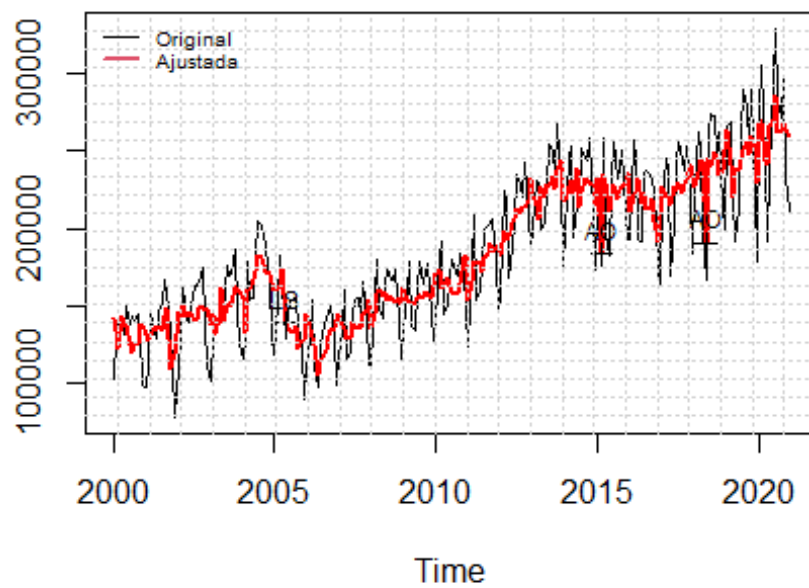
Evolução Vendas Diesel com Ajuste Sazonal - Região



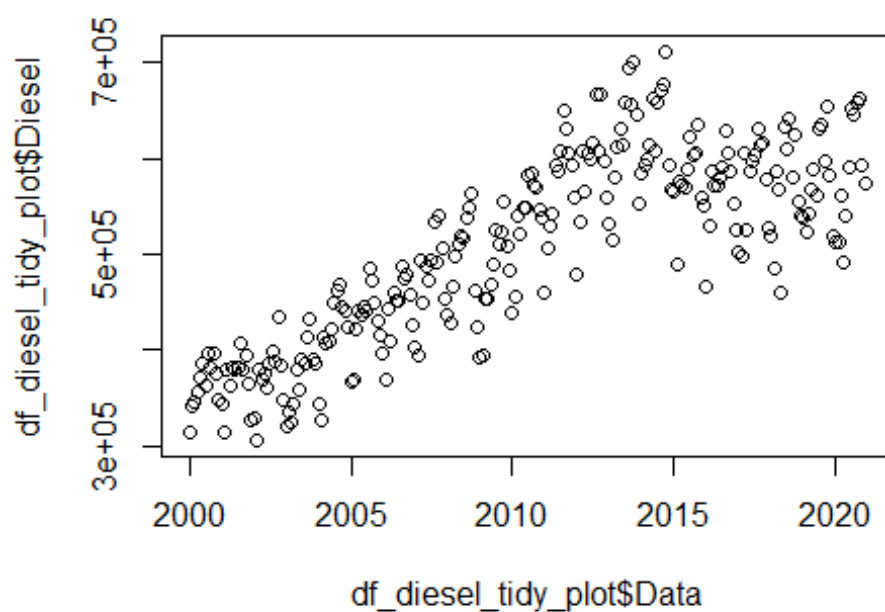
Evolução Vendas Diesel - Região mt



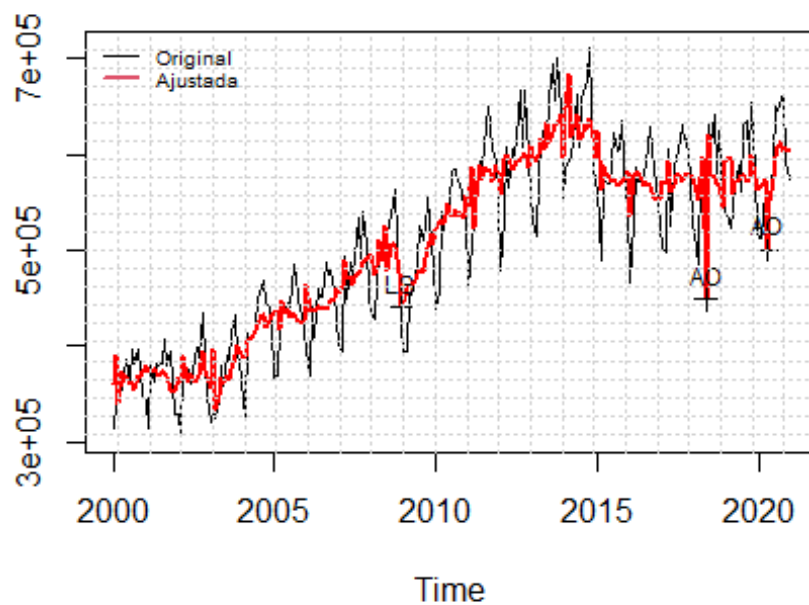
Evolução Vendas Diesel com Ajuste Sazonal - Região



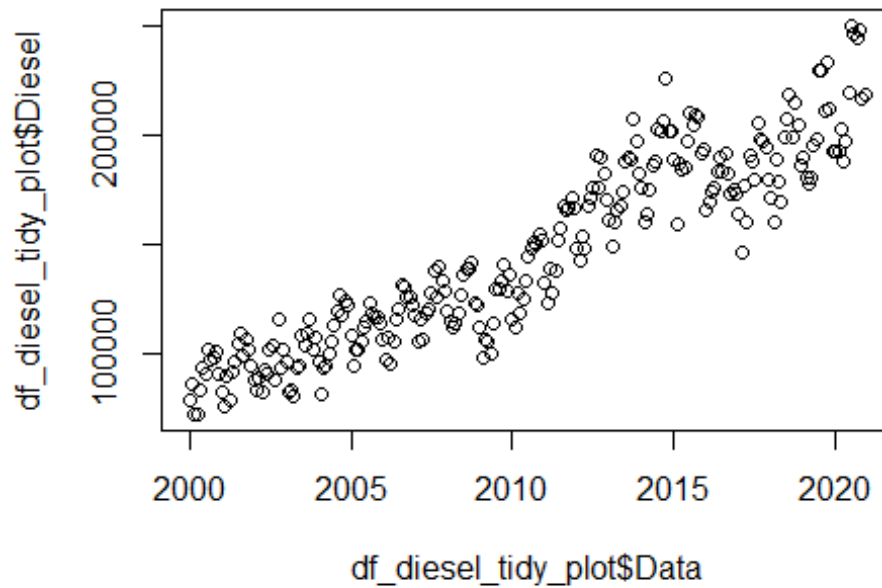
Evolução Vendas Diesel - Região mg



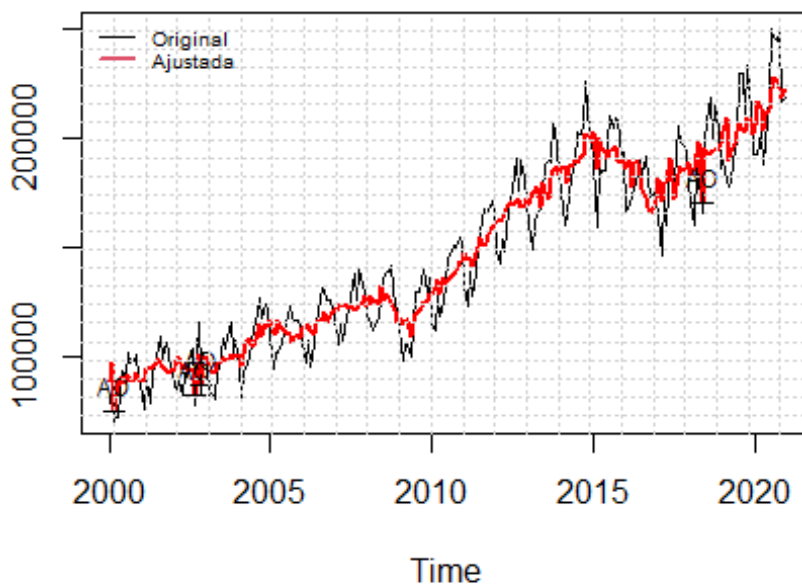
Evolução Vendas Diesel com Ajuste Sazonal - Região mg



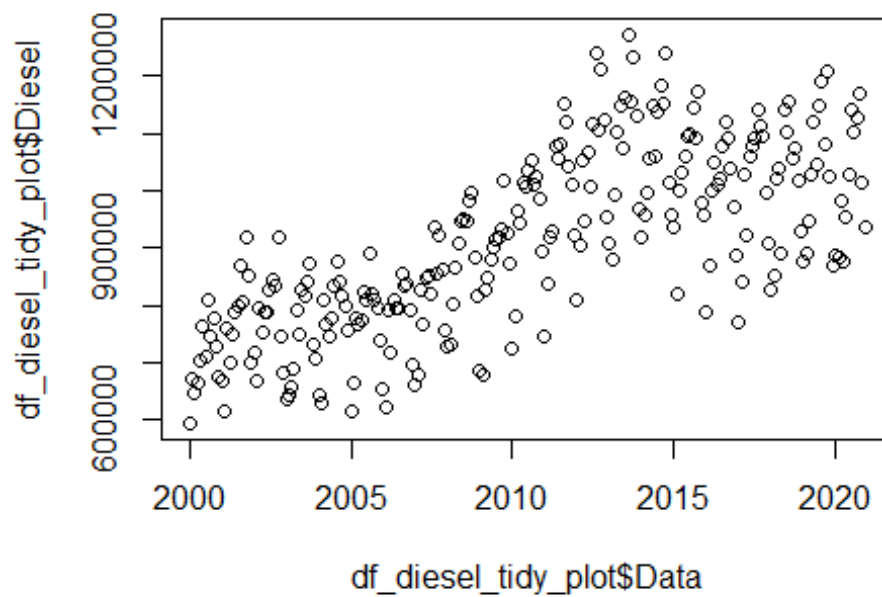
Evolução Vendas Diesel - Região pa



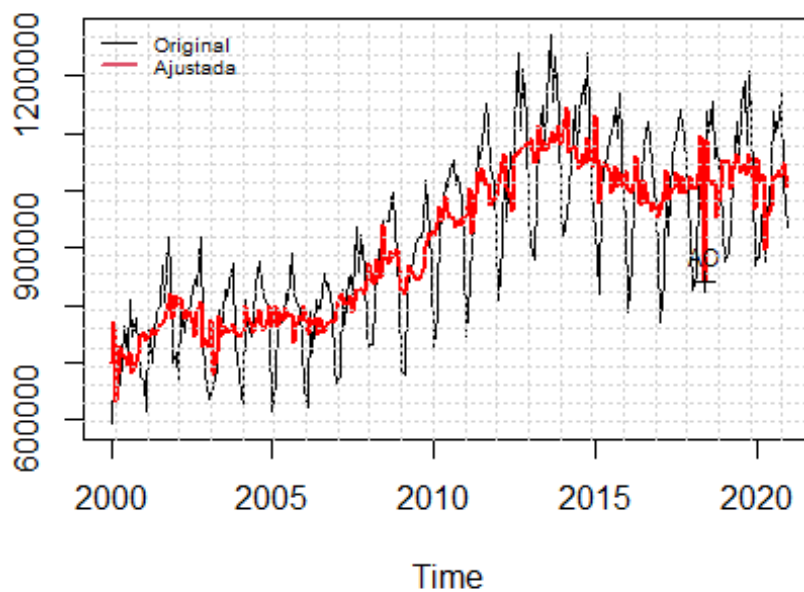
Evolução Vendas Diesel com Ajuste Sazonal - Região



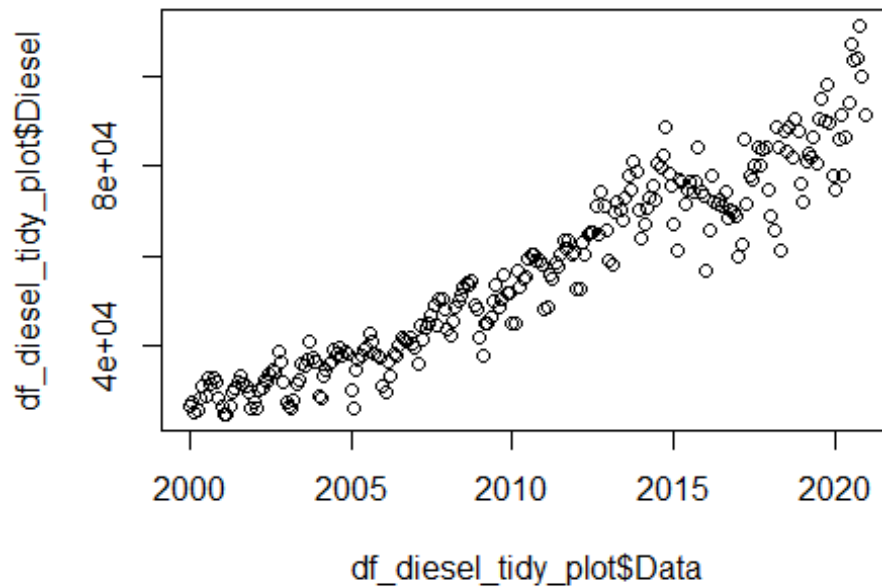
Evolução Vendas Diesel - Região sp



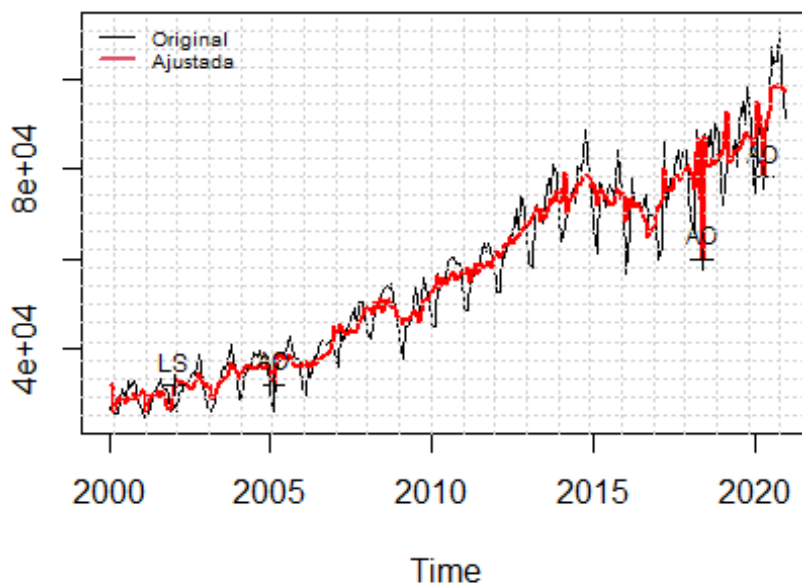
Evolução Vendas Diesel com Ajuste Sazonal - Região



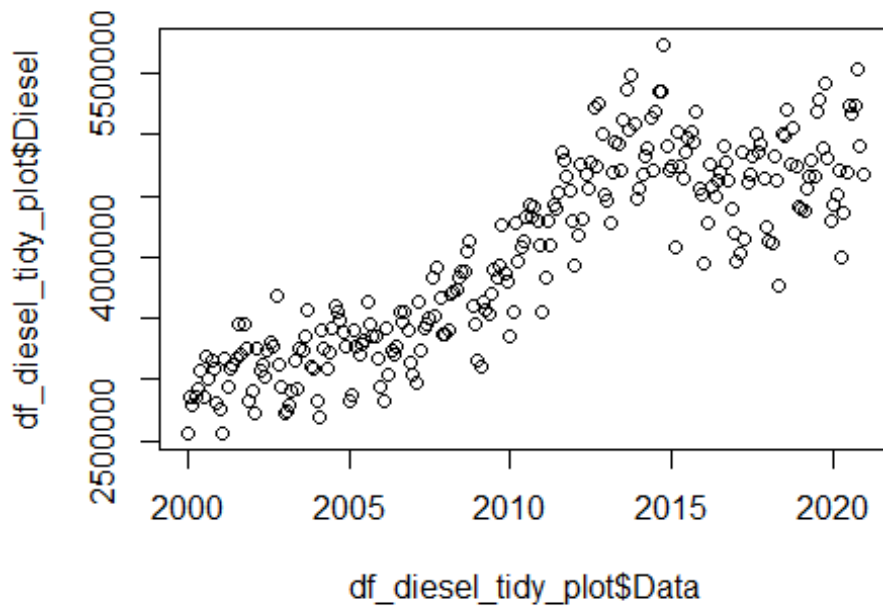
Evolução Vendas Diesel - Região to



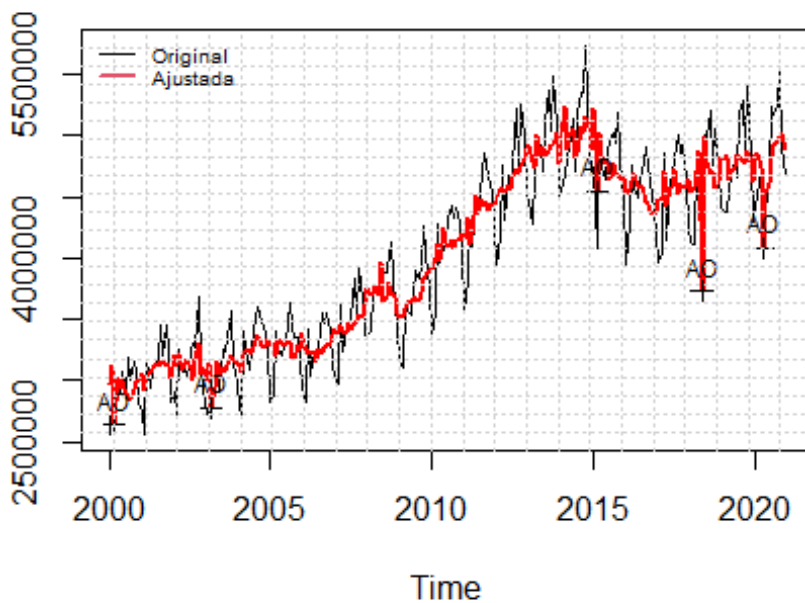
Evolução Vendas Diesel com Ajuste Sazonal - Regiã



Evolução Vendas Diesel - Região br



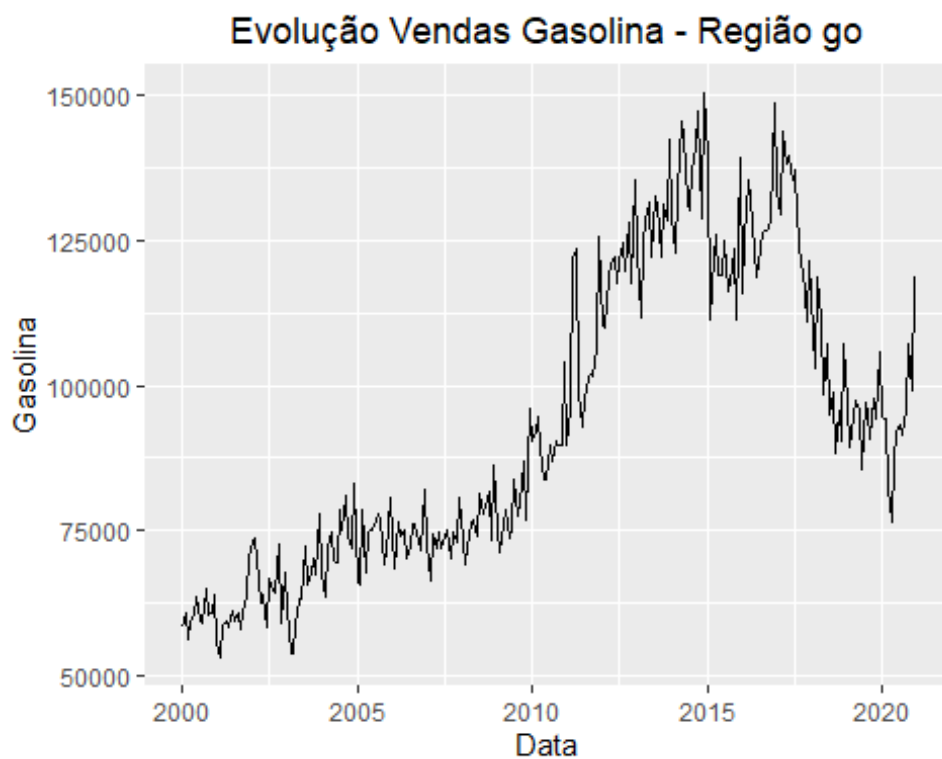
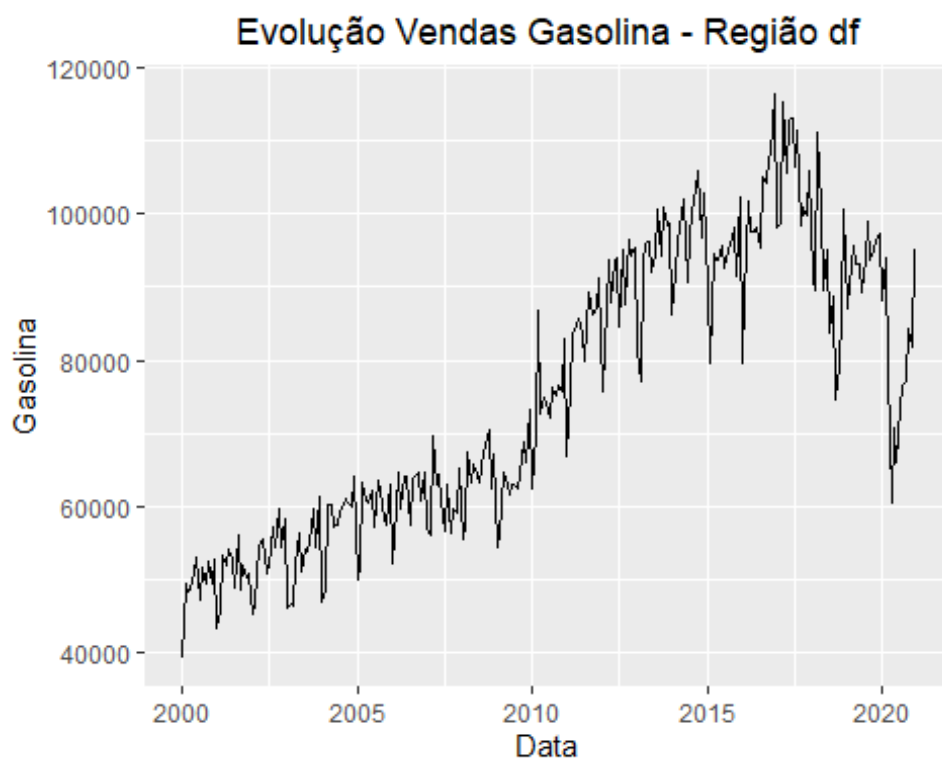
Evolução Vendas Diesel com Ajuste Sazonal - Regiã



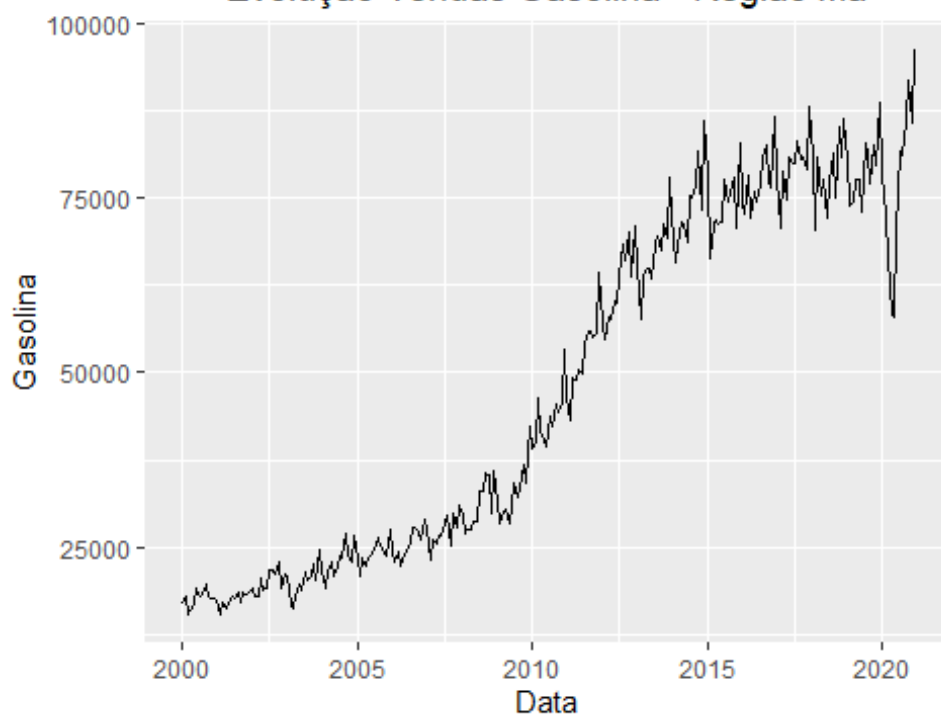
#série original Gasolina

```
for (i in unique(df_gasolina_tidy$regiao)) {
  print(ggplot(df_gasolina_tidy %>% filter(regiao == i), aes(Data,
    Gasolina)) +
```

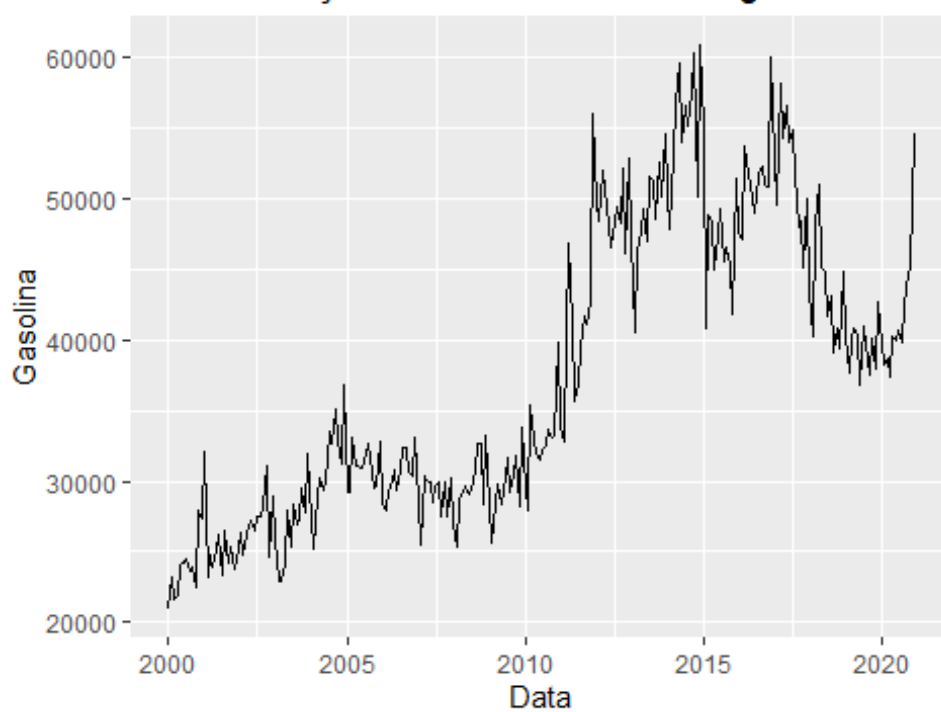
```
geom_line() + ggtitle(paste("Evolução Vendas Gasolina -  
Região", i, sep = " ")) + theme(plot.title = element_text(hjust = 0.5)))  
}
```



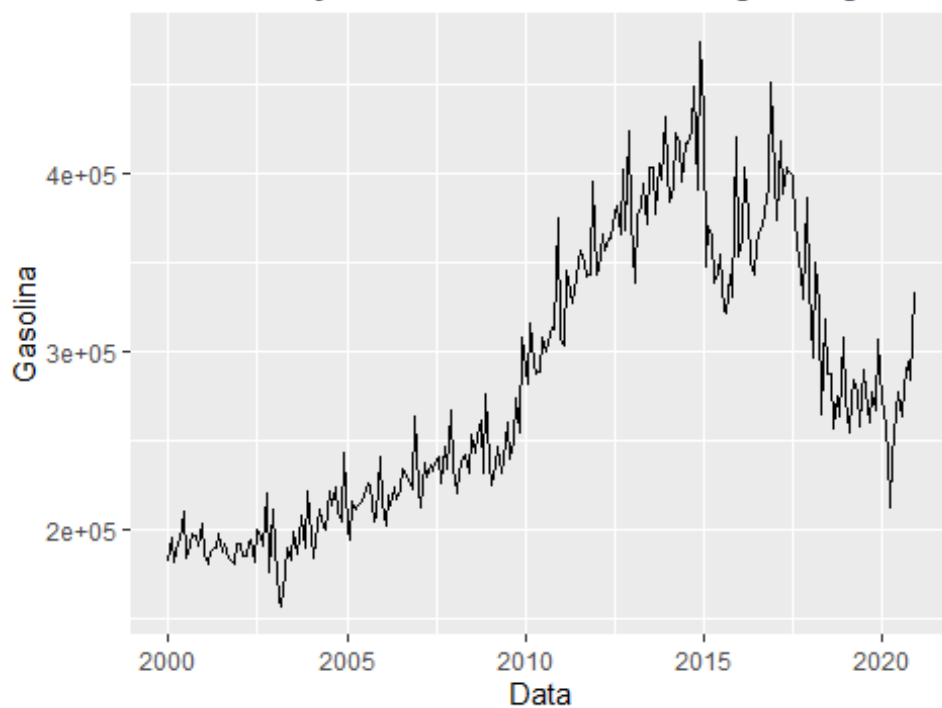
Evolução Vendas Gasolina - Região ma



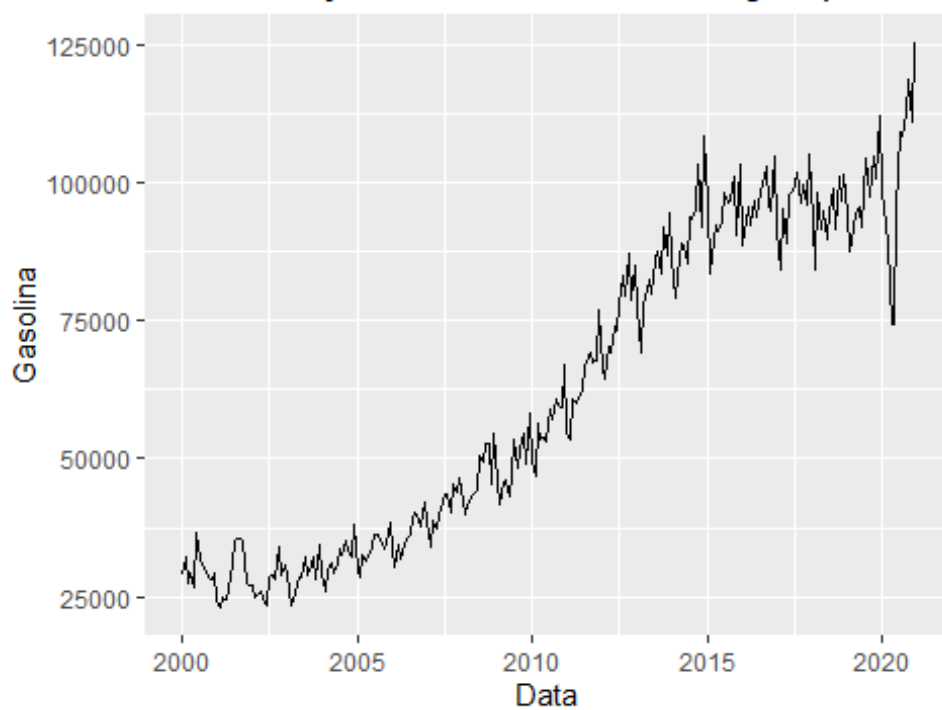
Evolução Vendas Gasolina - Região mt



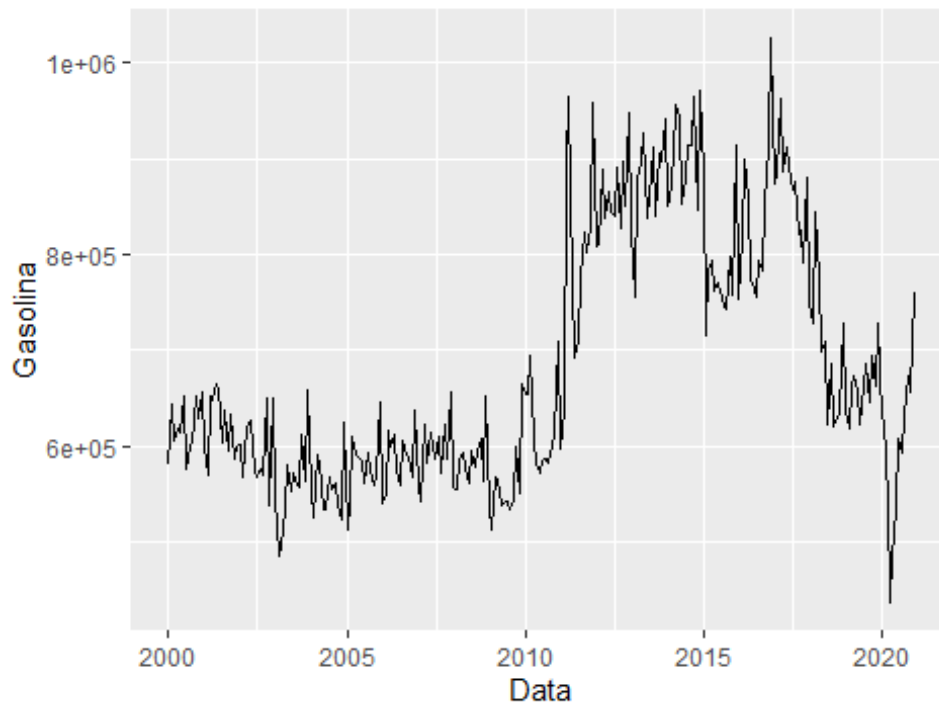
Evolução Vendas Gasolina - Região mg



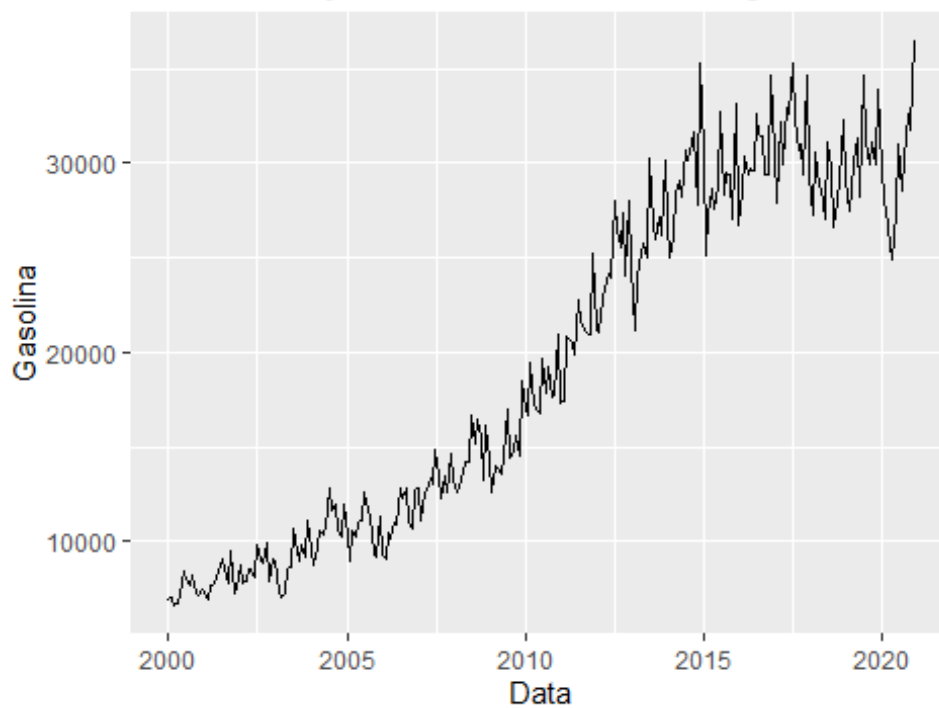
Evolução Vendas Gasolina - Região pa

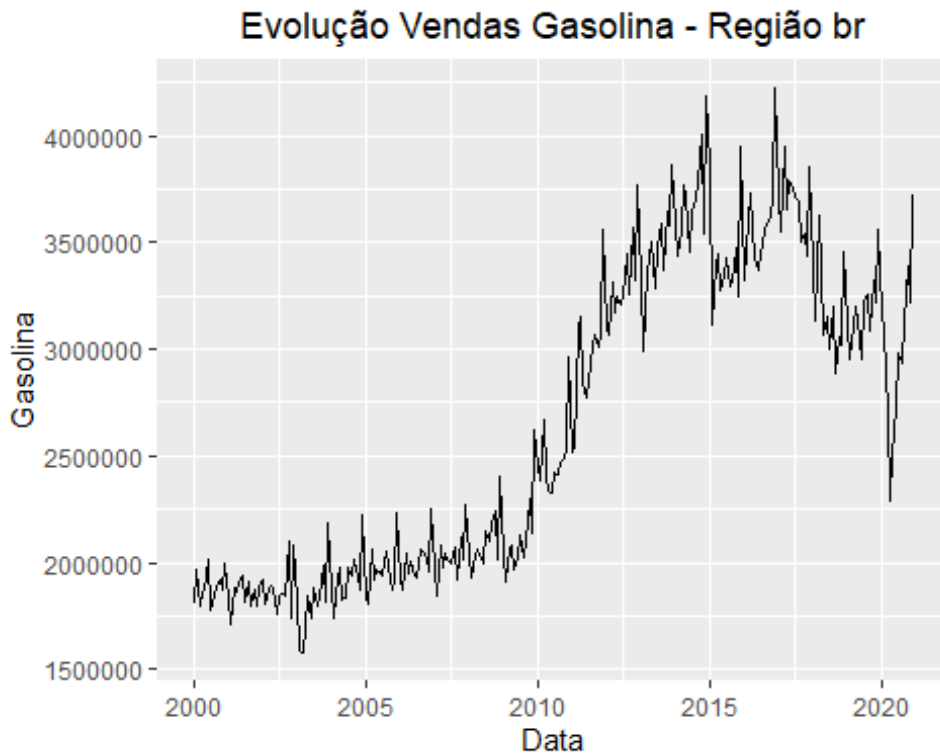


Evolução Vendas Gasolina - Região sp



Evolução Vendas Gasolina - Região to

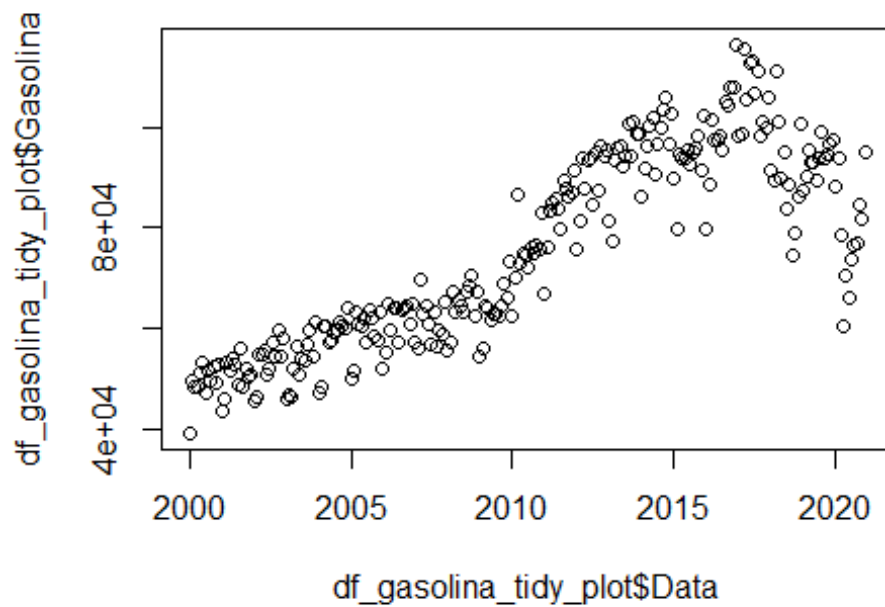




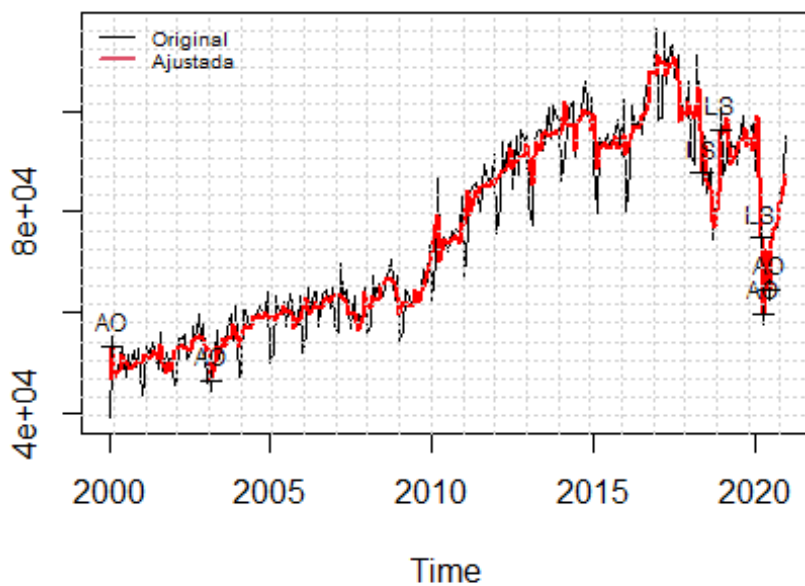
```
#Por região: série original vs ajuste sazonal (Gasolina)
for (i in unique(df_gasolina_tidy$regiao)) {
  df_gasolina_tidy_plot <- df_gasolina_tidy %>% filter(regiao == i)
  plot(df_gasolina_tidy_plot$Data, df_gasolina_tidy_plot$Gasolina, main =
paste("Evolução Vendas Gasolina - Região", i, sep = " "))

  regiao_br <- ts(df_gasolina_tidy_plot$Gasolina, start = c(2000, 1) ,
frequency = 12)
  dessaz_br <- seas(regiao_br, x11 = "")
  plot(dessaz_br, main = paste("Evolução Vendas Gasolina com Ajuste
Sazonal - Região", i, sep = " "))
  grid(nx = 23)
  legend("topleft", legend = c("Original", "Ajustada"), col = c(1, 2),
lwd = c(1, 2),
      lty = 1, bty = "n", cex = 0.6)
}
```

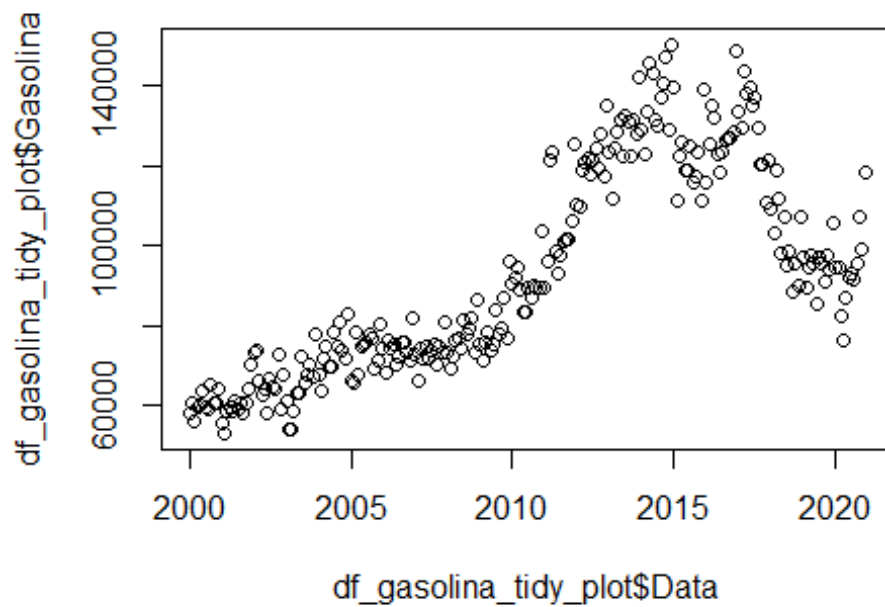
Evolução Vendas Gasolina - Região df



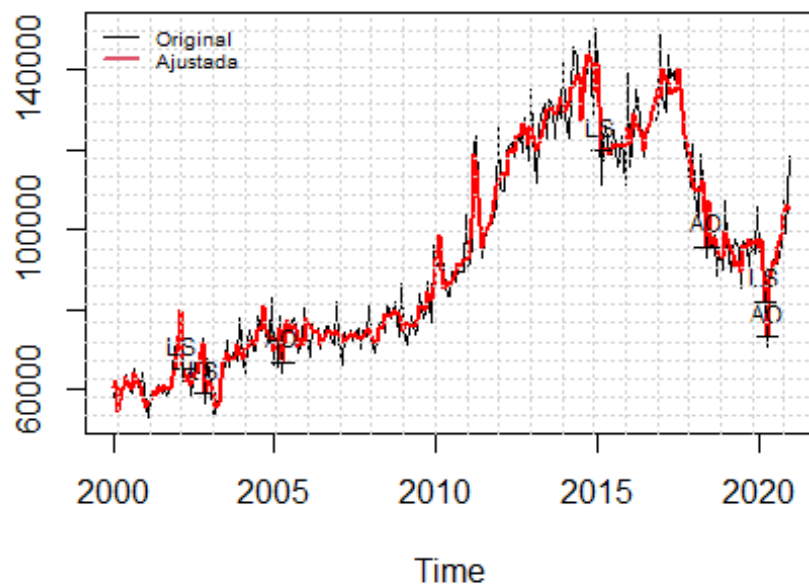
Evolução Vendas Gasolina com Ajuste Sazonal - Regi



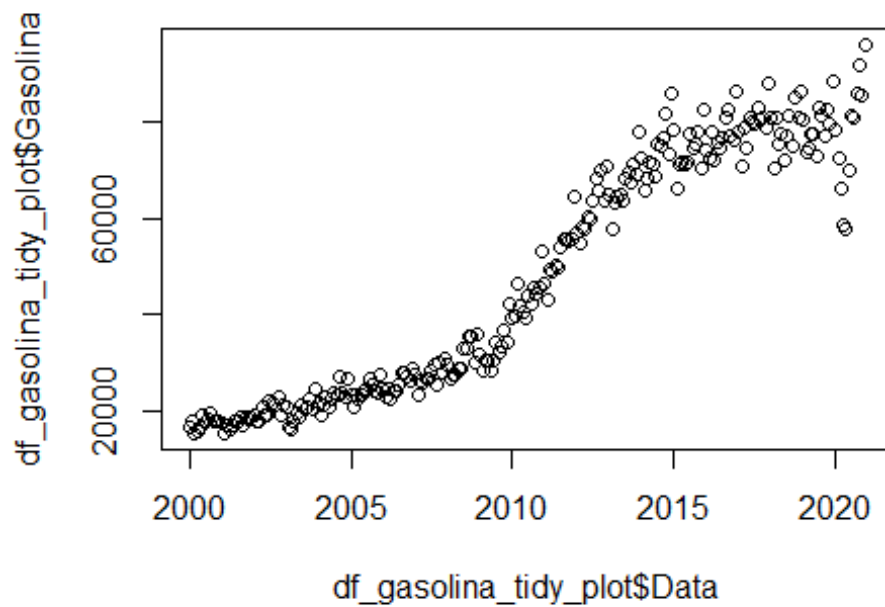
Evolução Vendas Gasolina - Região go



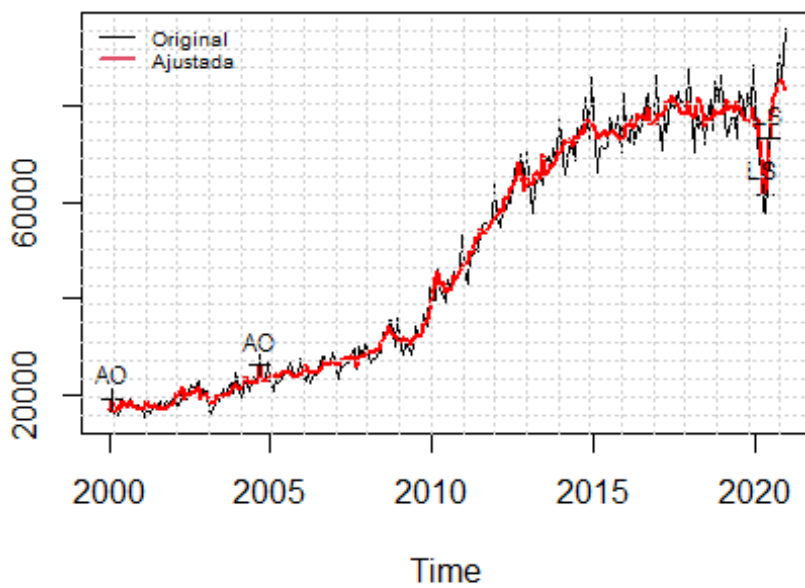
volução Vendas Gasolina com Ajuste Sazonal - Regi



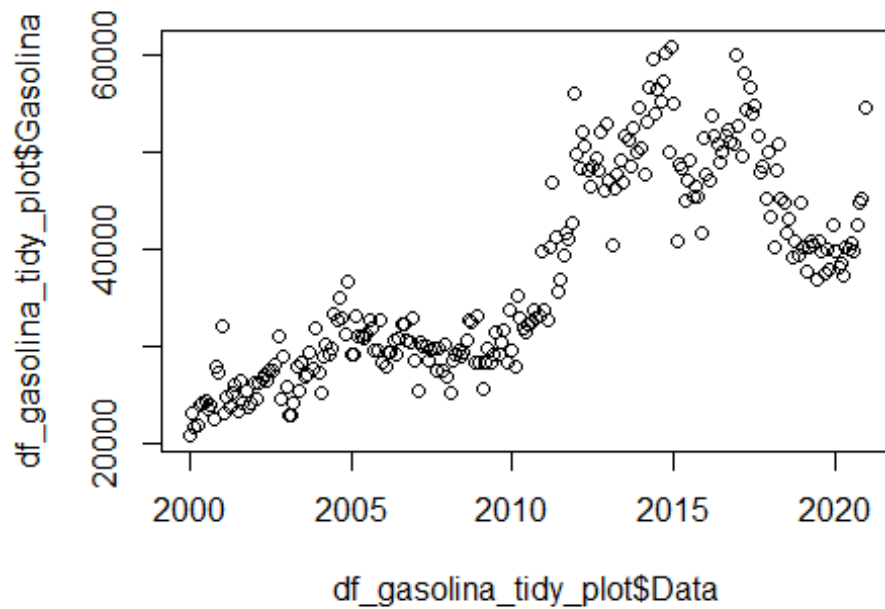
Evolução Vendas Gasolina - Região ma



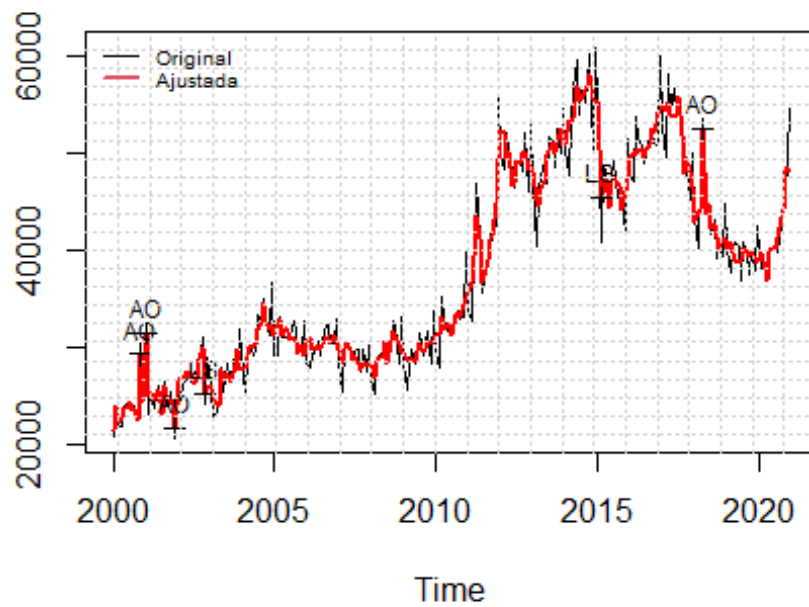
volução Vendas Gasolina com Ajuste Sazonal - Regi



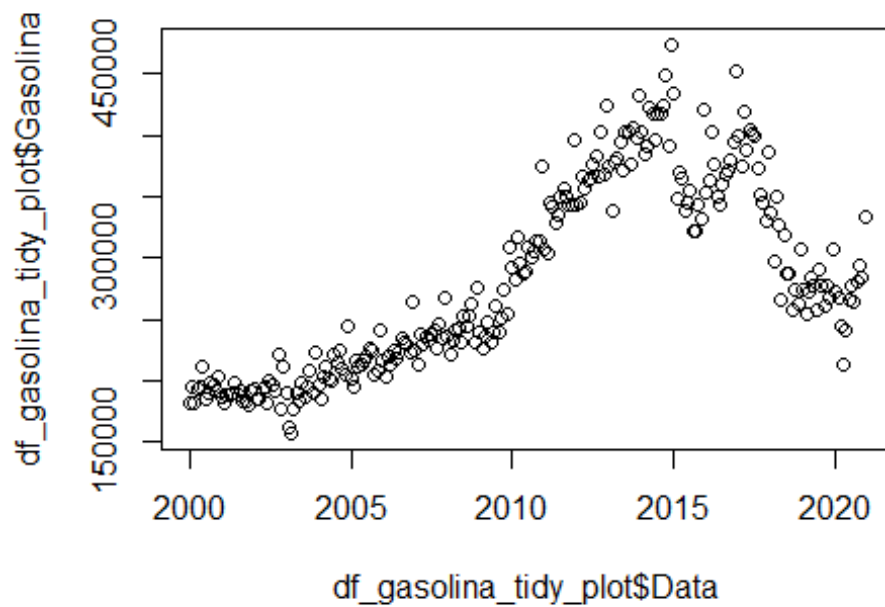
Evolução Vendas Gasolina - Região mt



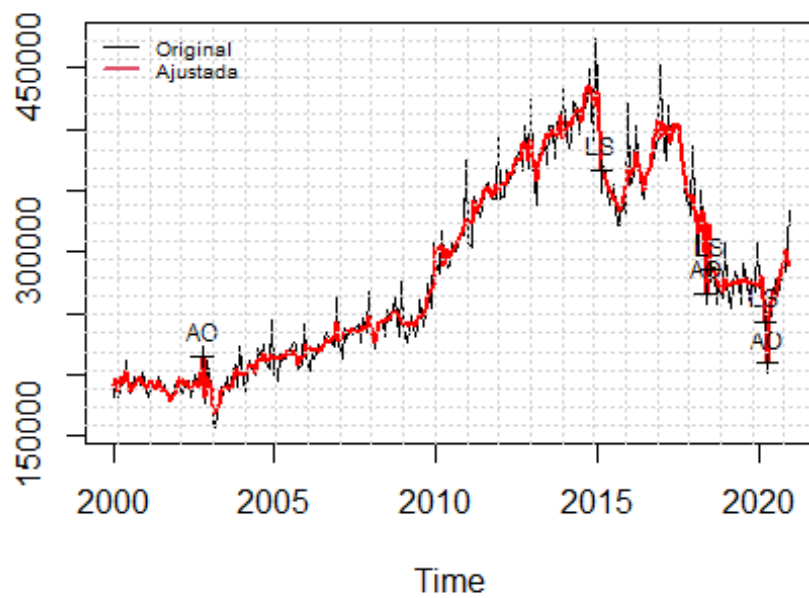
volução Vendas Gasolina com Ajuste Sazonal - Regi



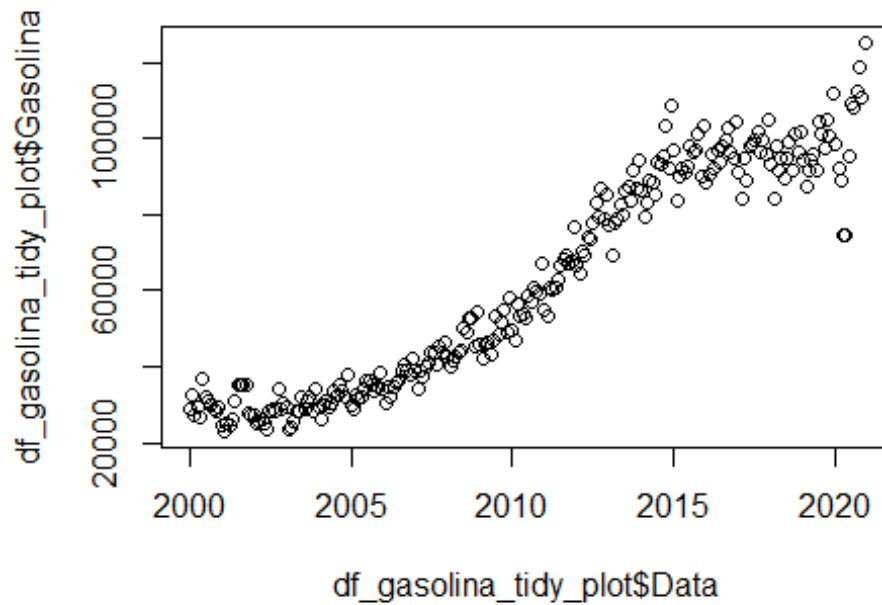
Evolução Vendas Gasolina - Região mg



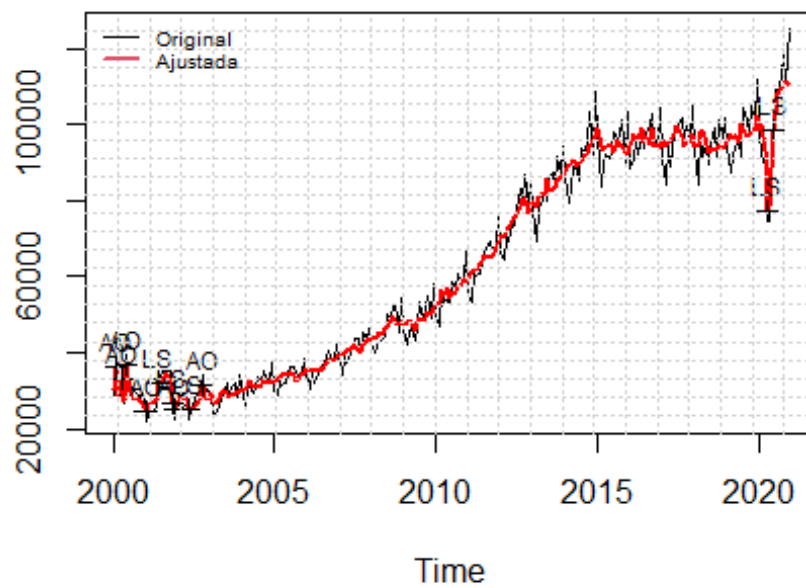
volução Vendas Gasolina com Ajuste Sazonal - Regi



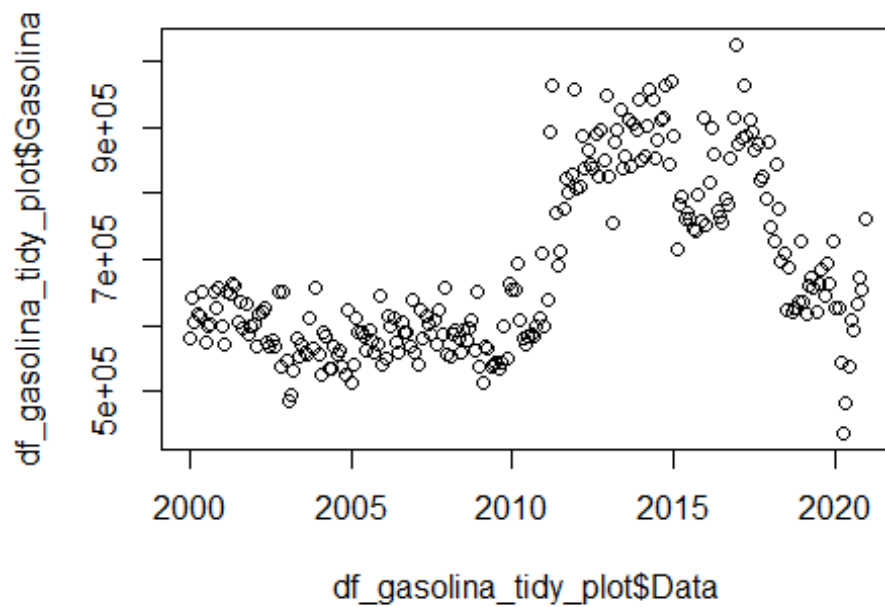
Evolução Vendas Gasolina - Região pa



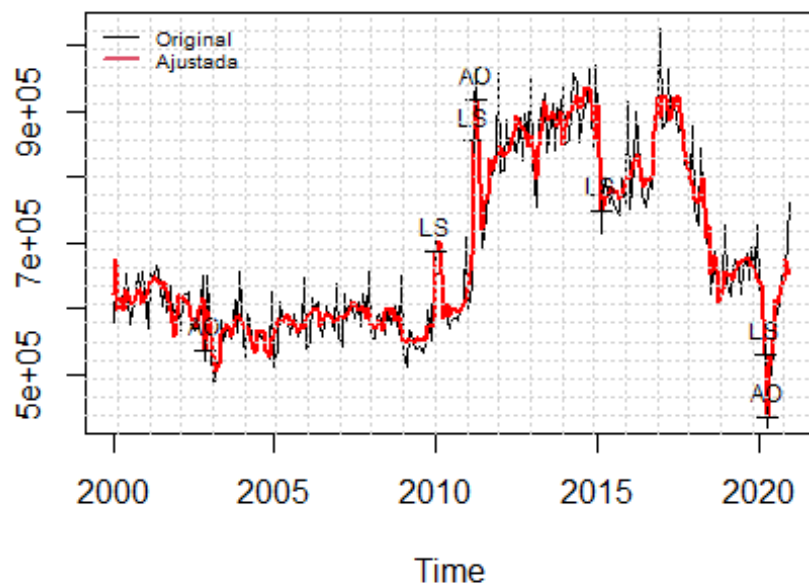
volução Vendas Gasolina com Ajuste Sazonal - Regi



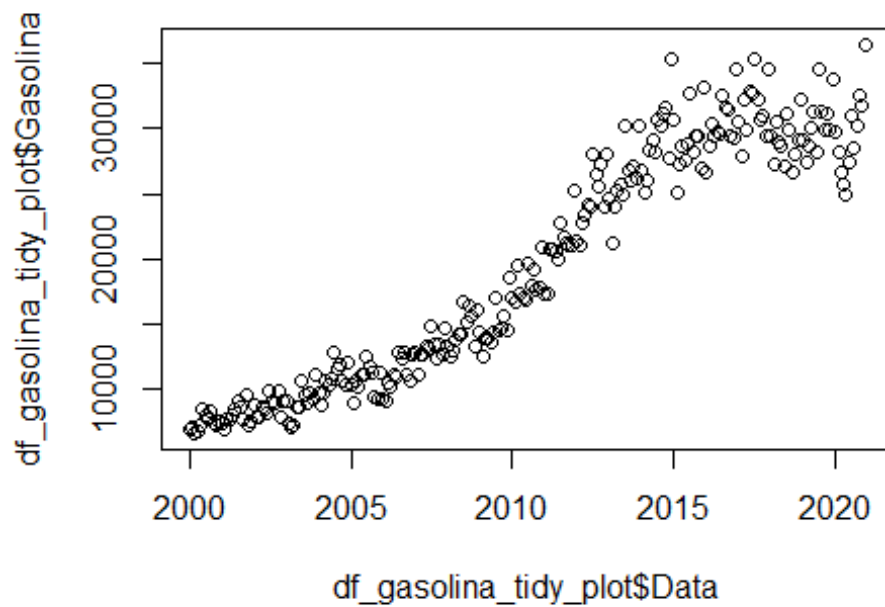
Evolução Vendas Gasolina - Região sp



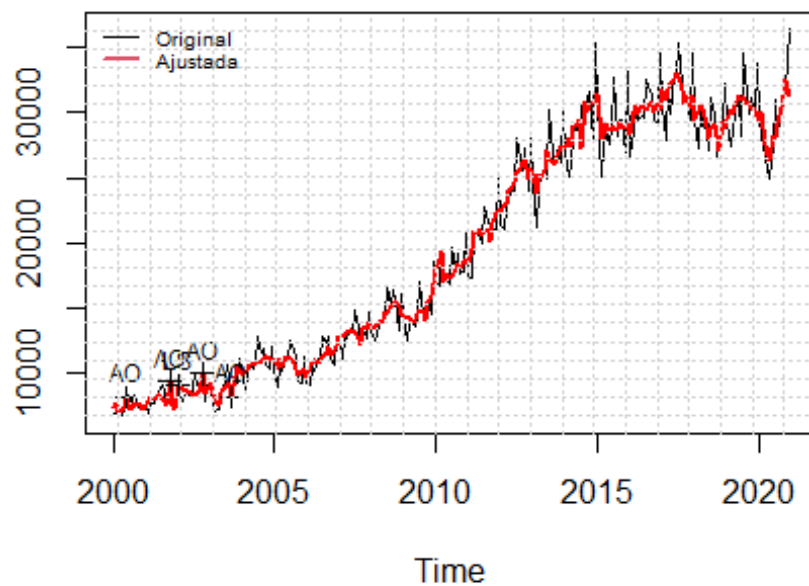
volução Vendas Gasolina com Ajuste Sazonal - Regi



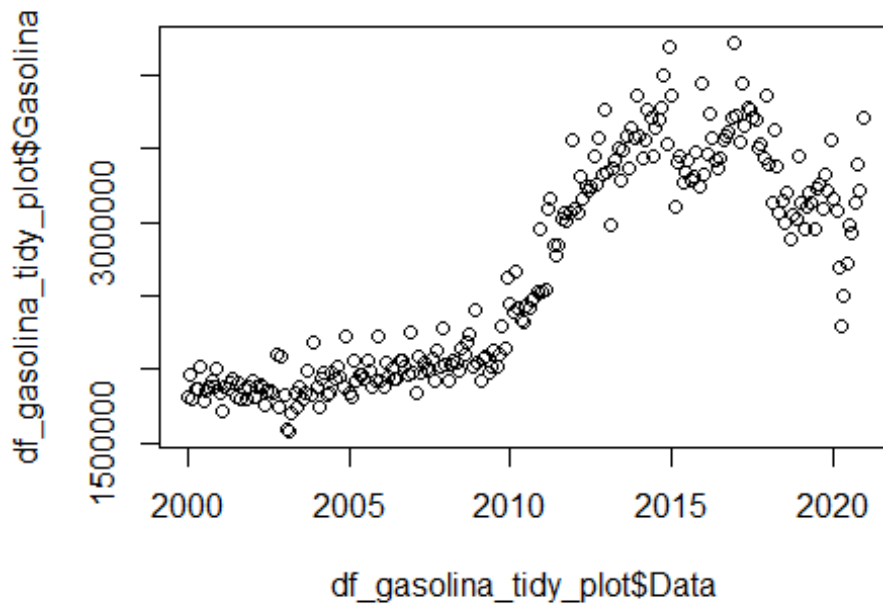
Evolução Vendas Gasolina - Região to



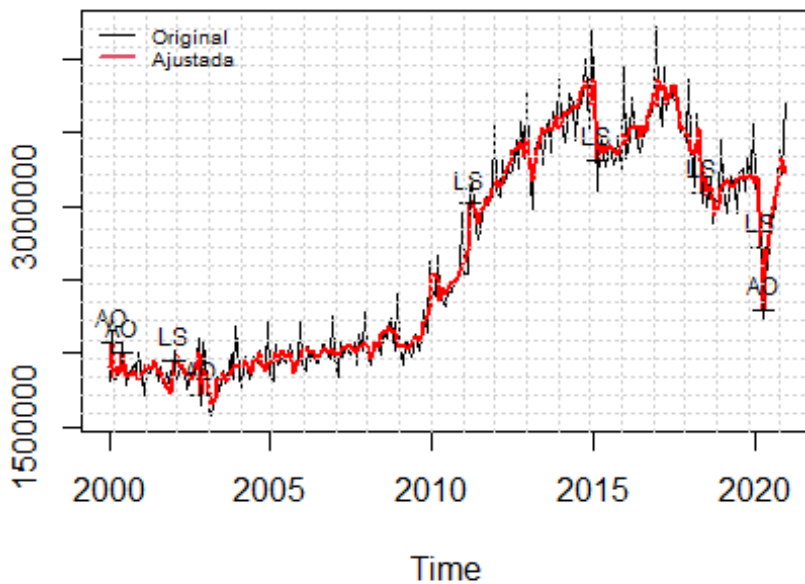
Evolução Vendas Gasolina com Ajuste Sazonal - Regi



Evolução Vendas Gasolina - Região br



volução Vendas Gasolina com Ajuste Sazonal - Regi



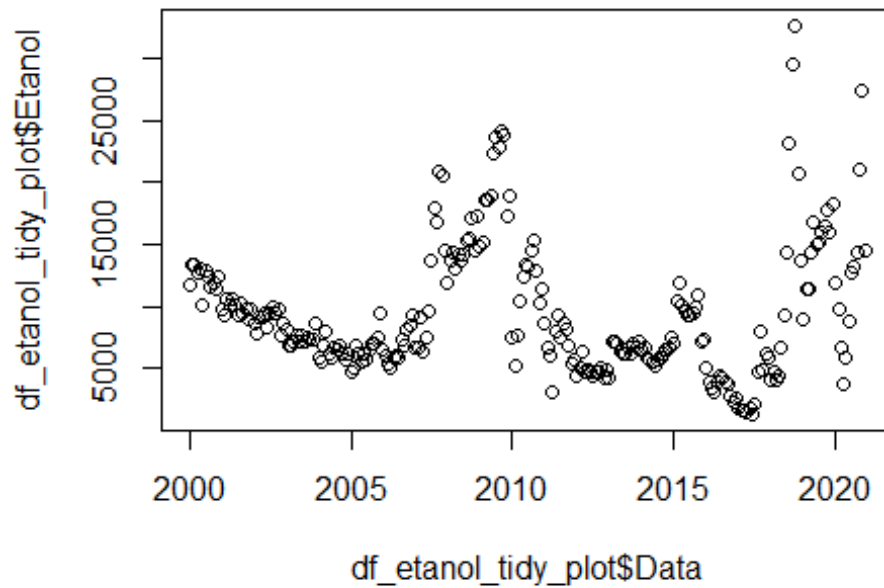
#Por região: série original vs ajuste sazonal (Gasolina)

```
for (i in unique(df_etanol_tidy$regiao)) {
  df_etanol_tidy_plot <- df_etanol_tidy %>% filter(regiao == i)
  plot(df_etanol_tidy_plot$Data, df_etanol_tidy_plot$Etanol, main =
```

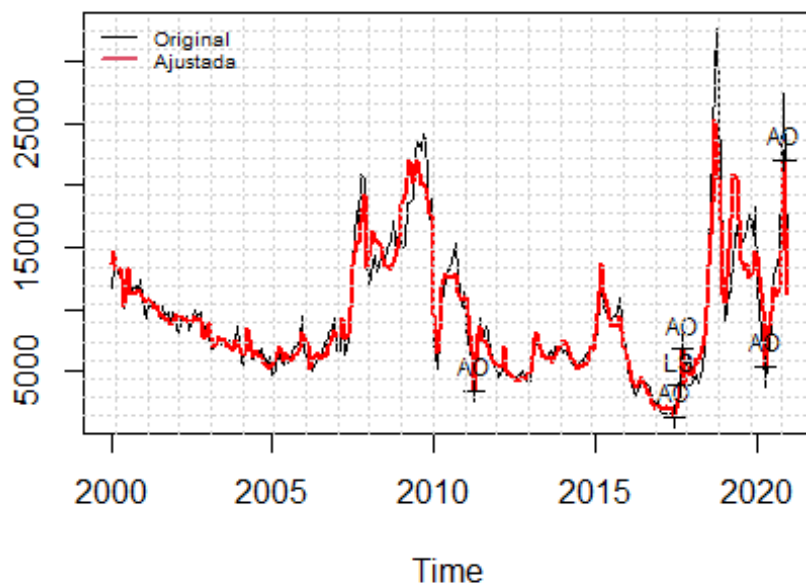
```
paste("Evolução Vendas Etanol - Região", i, sep = " "))

  regioao_br <- ts(df_etanol_tidy_plot$Etanol, start = c(2000, 1) ,
frequency = 12)
  dessaz_br <- seas(regiao_br, x11 = "")
  plot(dessaz_br, main = paste("Evolução Vendas Etanol com Ajuste Sazonal
- Região", i, sep = " "))
  grid(nx = 23)
  legend("topleft", legend = c("Original", "Ajustada"), col = c(1, 2),
lwd = c(1, 2),
      lty = 1, bty = "n", cex = 0.6)
}
```

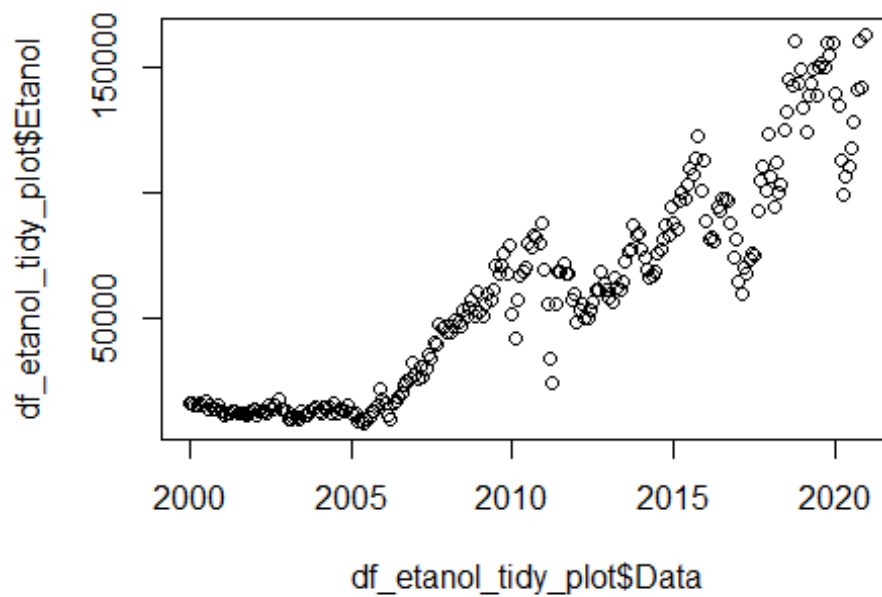

Evolução Vendas Etanol - Região df



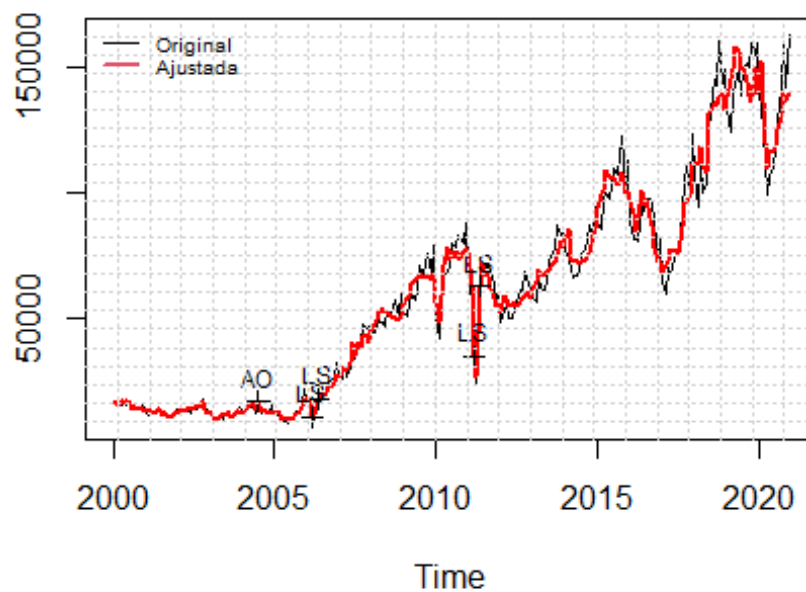
Evolução Vendas Etanol com Ajuste Sazonal - Regiã



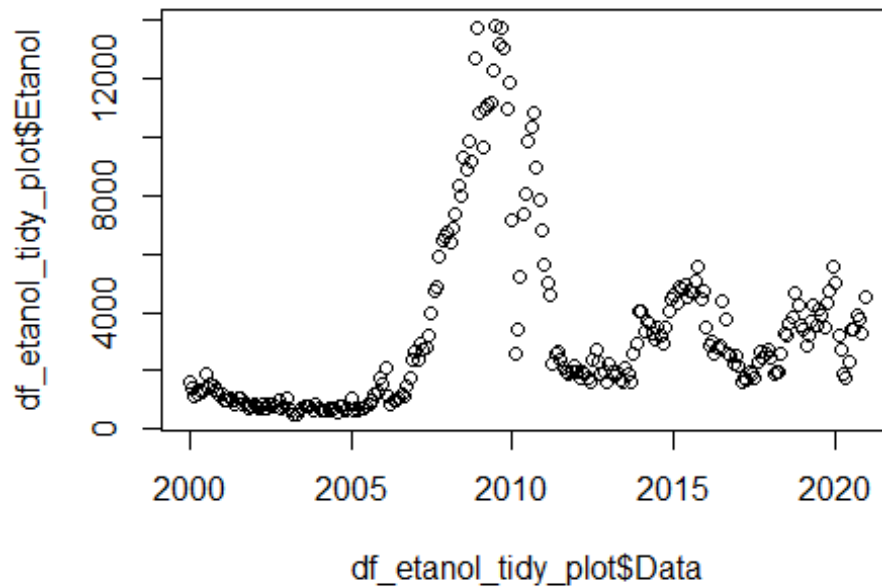
Evolução Vendas Etanol - Região go



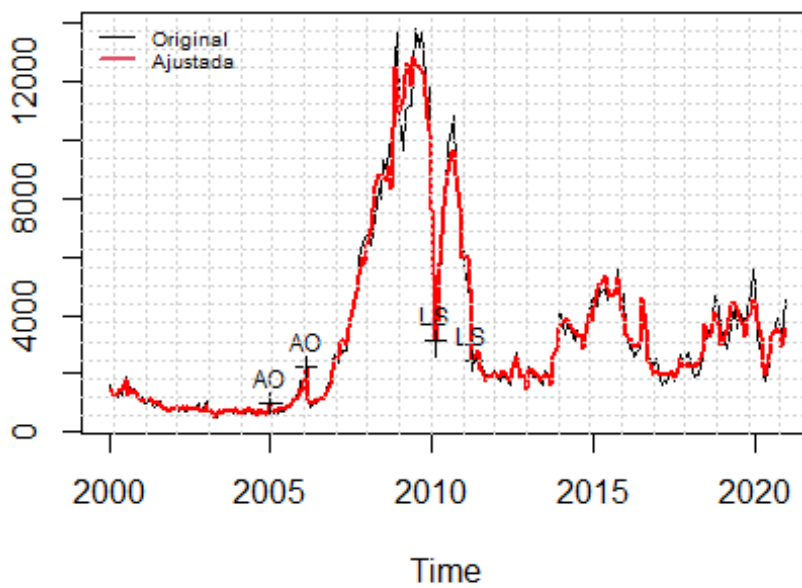
Evolução Vendas Etanol com Ajuste Sazonal - Região



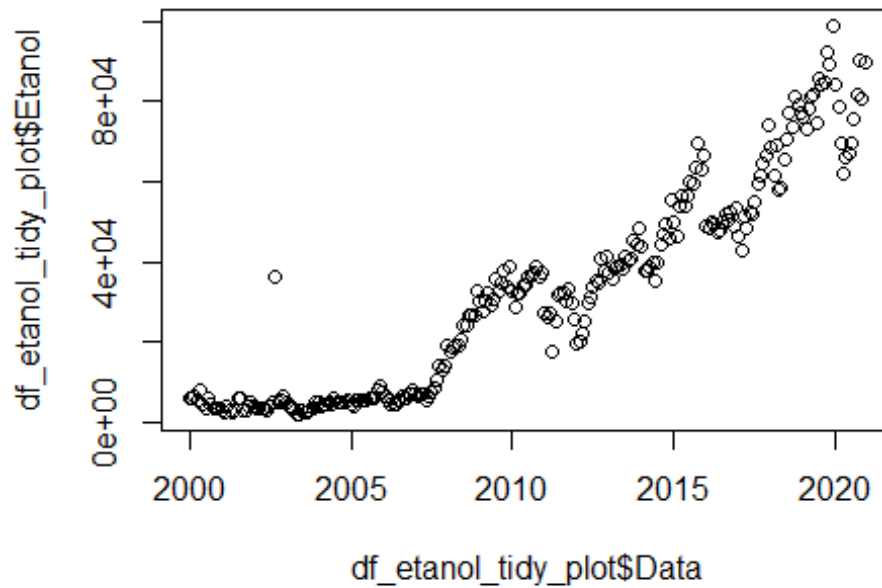
Evolução Vendas Etanol - Região ma



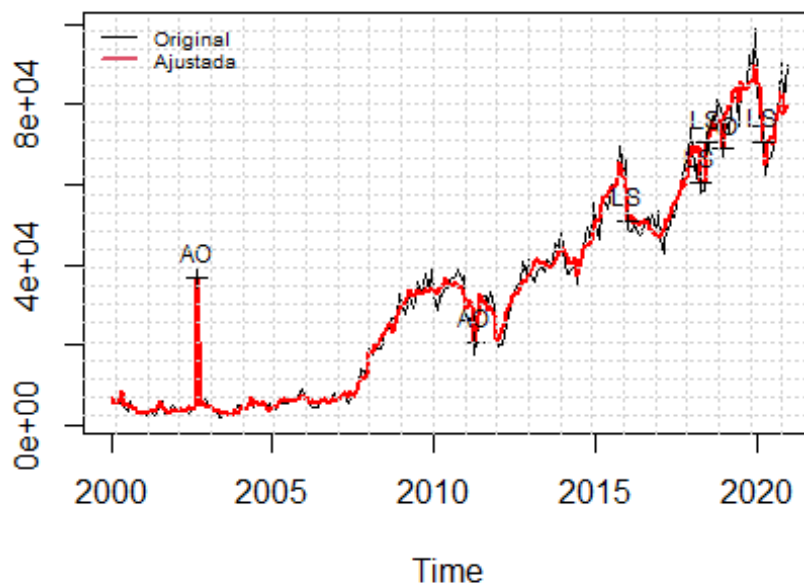
Evolução Vendas Etanol com Ajuste Sazonal - Região



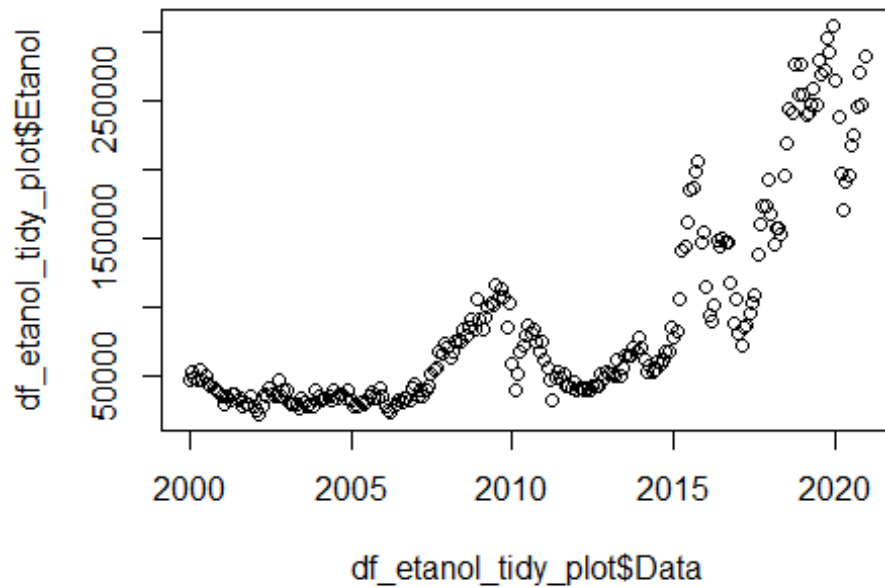
Evolução Vendas Etanol - Região mt



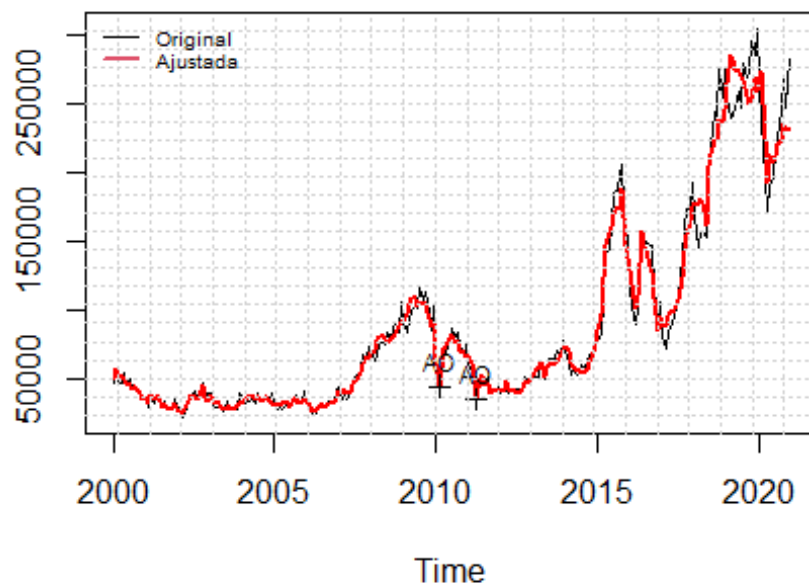
Evolução Vendas Etanol com Ajuste Sazonal - Região



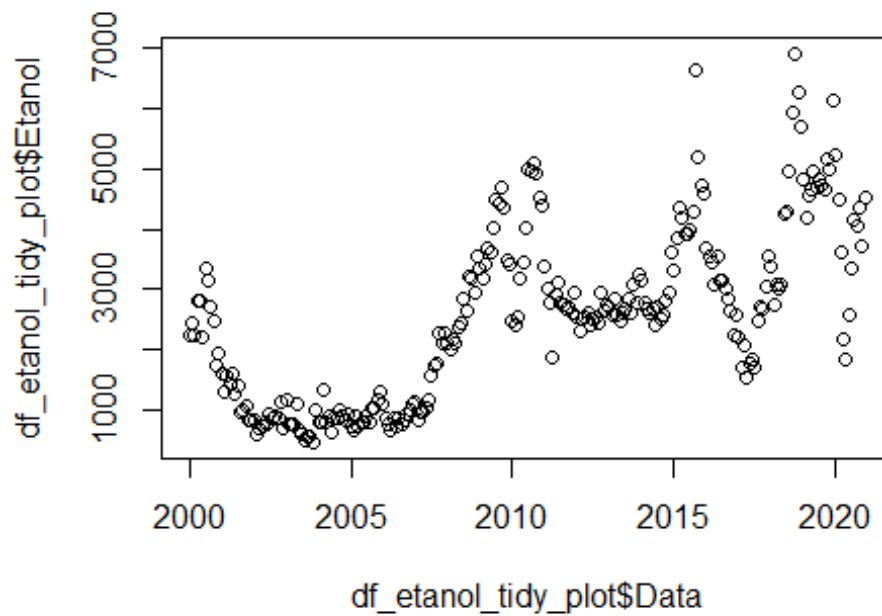
Evolução Vendas Etanol - Região mg



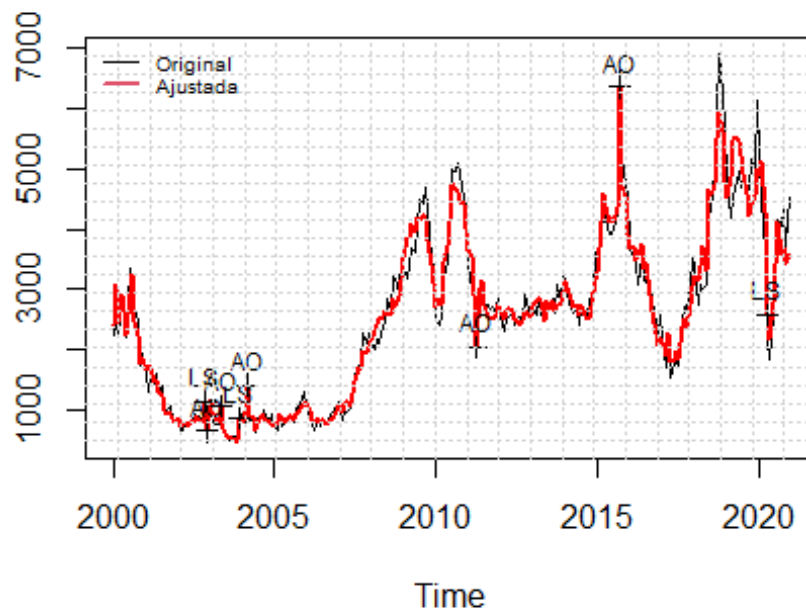
Evolução Vendas Etanol com Ajuste Sazonal - Região mg



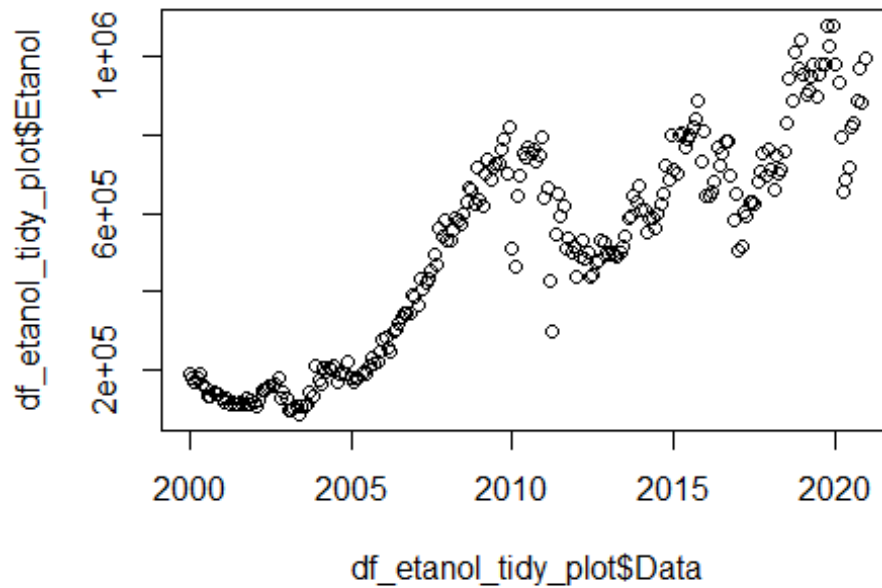
Evolução Vendas Etanol - Região pa



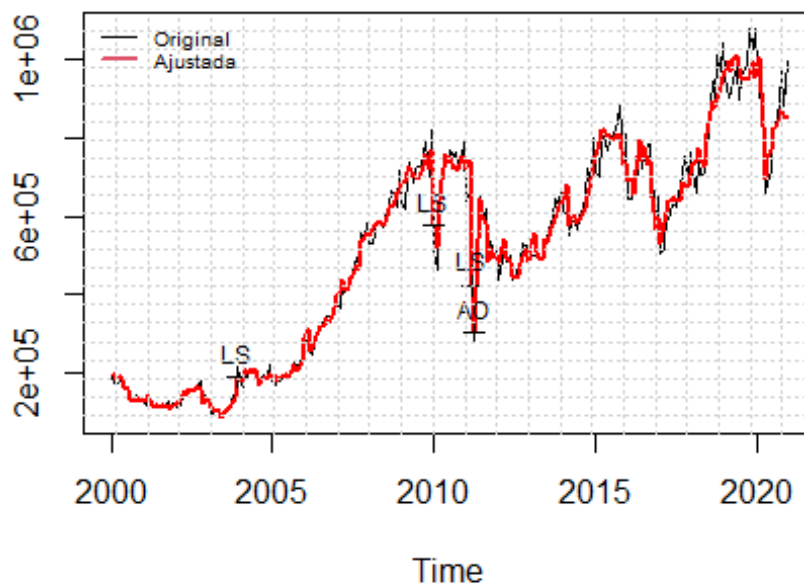
Evolução Vendas Etanol com Ajuste Sazonal - Região



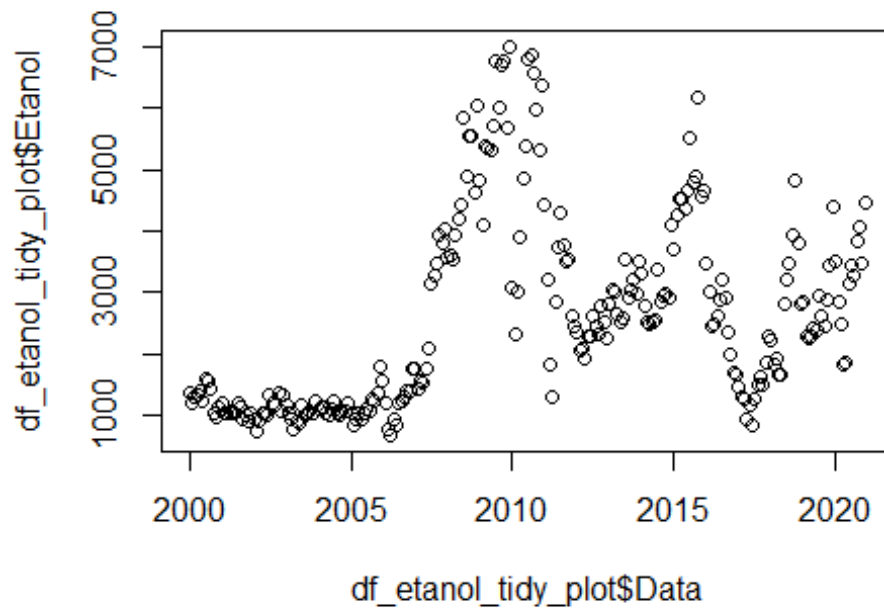
Evolução Vendas Etanol - Região sp



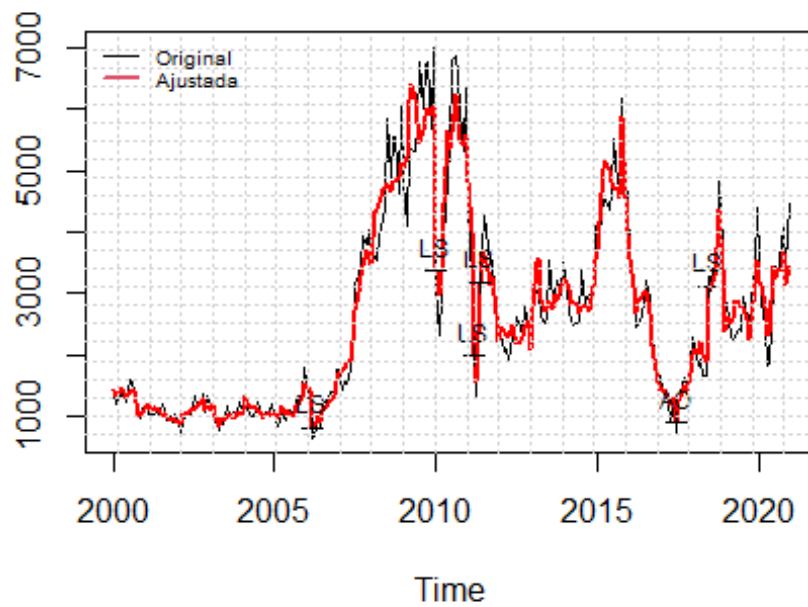
Evolução Vendas Etanol com Ajuste Sazonal - Região



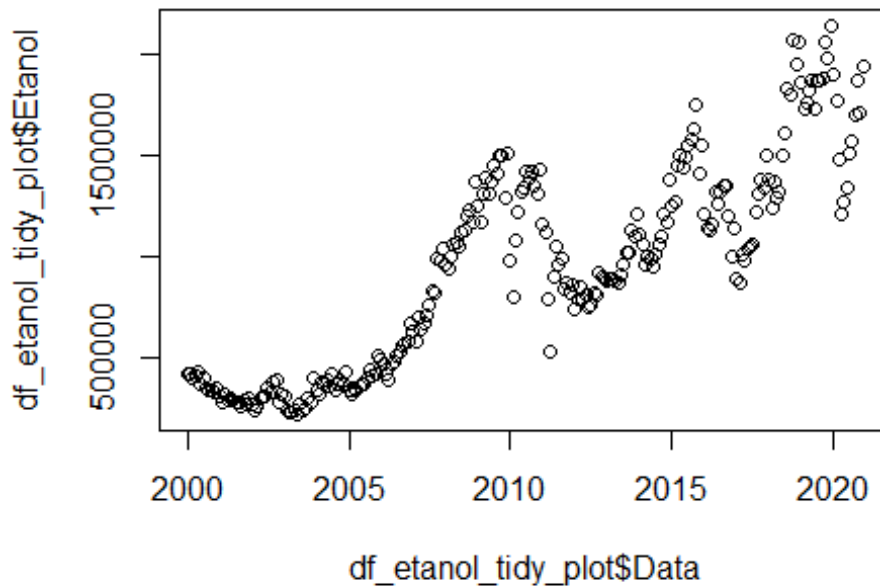
Evolução Vendas Etanol - Região to



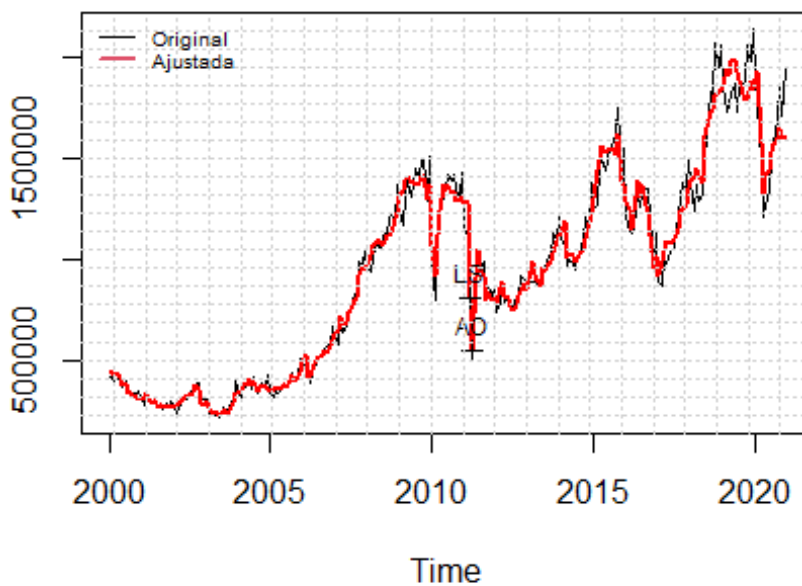
Evolução Vendas Etanol com Ajuste Sazonal - Regiã



Evolução Vendas Etanol - Região br



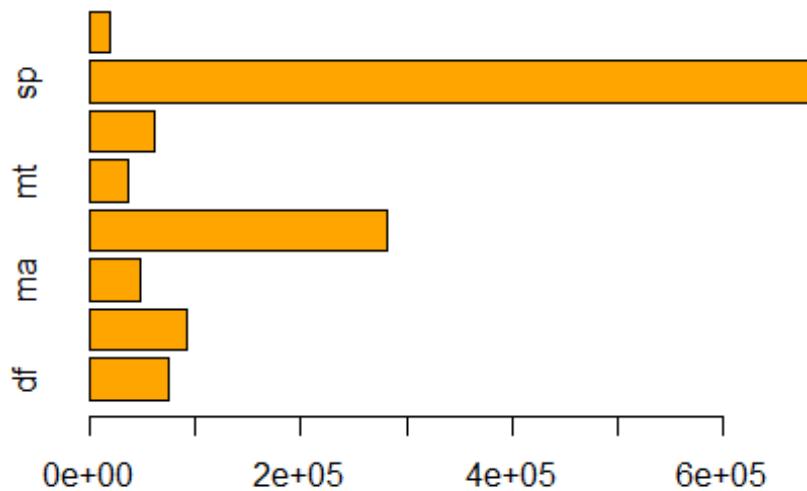
Evolução Vendas Etanol com Ajuste Sazonal - Região



— Abaixo, encontra-se a distribuição dos tipos de combustível por estado. Analisando os gráficos de barra, pode-se concluir que as vendas de combustível são maiores no estado de São Paulo, seguido por Minas Gerais.

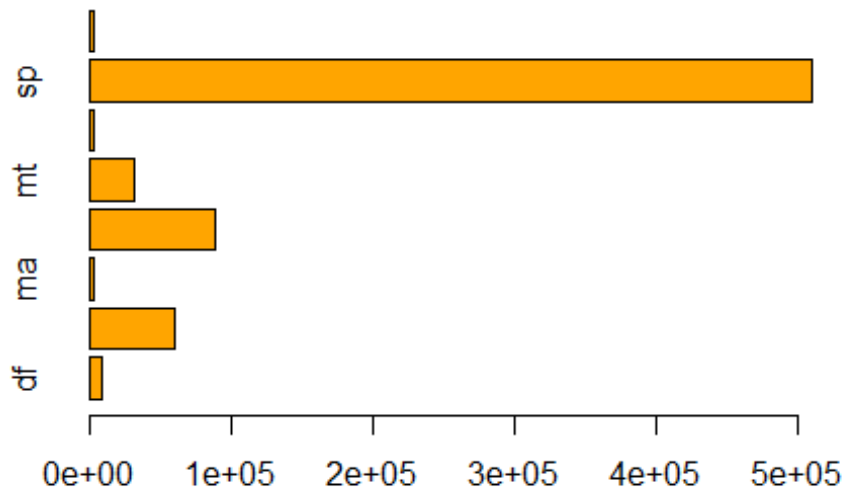
```
#Consumo de Gasolina por estado média dos últimos 20 anos
df_gasolina_tidy_semBR <- df_gasolina_tidy %>% filter(!regiao == "br")
barplot(tapply(df_gasolina_tidy_semBR$Gasolina,
df_gasolina_tidy_semBR$regiao, mean),
        col = "orange", horiz = T, main = "Média das vendas de Gasolina
por estado 2000 - 2020")
```

Média das vendas de Gasolina por estado 2000 - 2020



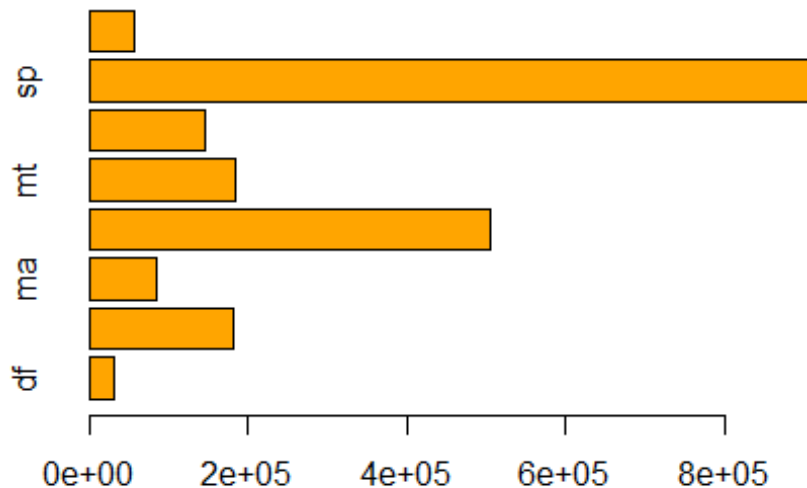
```
#Consumo de Etanol por estado, média dos últimos 20 anos
df_etanol_tidy_semBR <- df_etanol_tidy %>% filter(!regiao == "br")
barplot(tapply(df_etanol_tidy_semBR$Etanol, df_etanol_tidy_semBR$regiao,
mean),
        col = "orange", horiz = T, main = "Média das vendas de Etanol por
estado 2000 - 2020")
```

Média das vendas de Etanol por estado 2000 - 2020



```
#Consumo de Diesel por estado média dos últimos 20 anos  
df_diesel_tidy_semBR <- df_diesel_tidy %>% filter(!regiao == "br")  
barplot(tapply(df_diesel_tidy_semBR$Diesel, df_diesel_tidy_semBR$regiao,  
mean),  
        col = "orange", horiz = T, main = "Média das vendas de Diesel por  
estado 2000 - 2020")
```

Média das vendas de Diesel por estado 2000 - 2020



— O tamanho total médio do mercado nos últimos 20 anos é de 7.584.254 metros cúbicos de combustíveis. Além disso, analisando o a evolução do mercado nos últimos 5 anos pode-se notar uma tendência positiva, mesmo com o impacto da pandemia de COVID-19 em 2020.

```
#Analisando o tamanho total do mercado e a evolução deste nos últimos 5 anos#
df_consolidado <- full_join(x = df_diesel_tidy, y = df_gasolina_tidy)

## Joining, by = c("regiao", "Data")

df_consolidado <- full_join(x = df_consolidado, y = df_etanol_tidy)

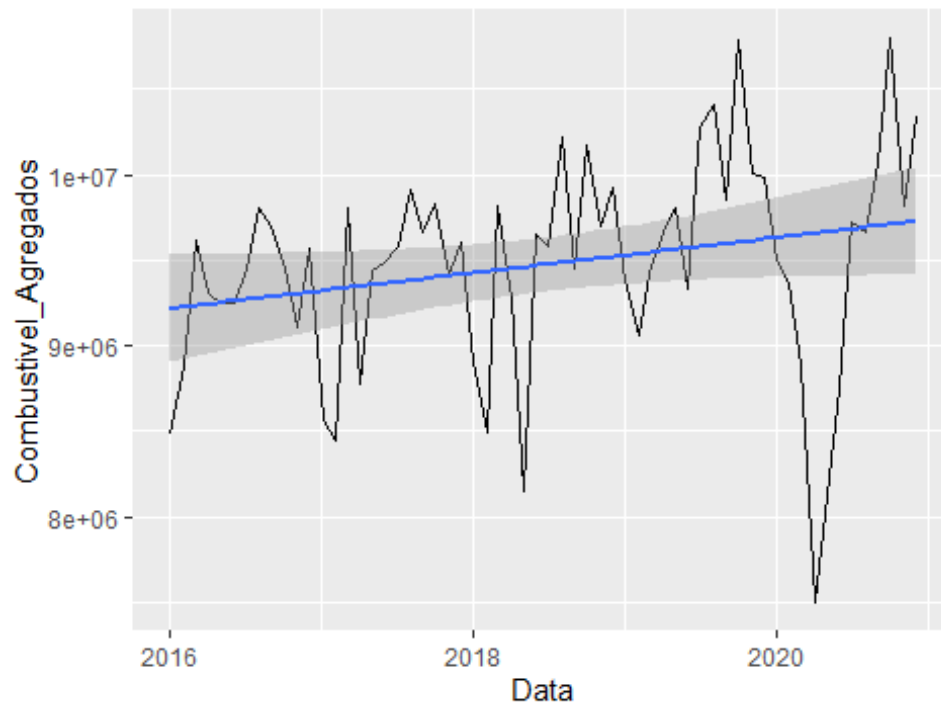
## Joining, by = c("regiao", "Data")

df_consolidado_tidy_BR <- df_consolidado %>% filter(regiao == "br")
df_tamanho_do_mercado <- df_consolidado_tidy_BR %>% group_by(Data) %>%
summarise(Combustivel_Agregados=sum(Diesel, Gasolina, Etanol))

ggplot(df_tamanho_do_mercado %>% filter(Data >= "2016-01-01"), aes(Data,
Combustivel_Agregados)) +
  geom_line() + geom_smooth(method = lm) + ggtitle("Evolução Vendas de
Combustíveis Total") + theme(plot.title = element_text(hjust = 0.5))

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

Evolução Vendas de Combustíveis Total



```
mean(df_tamanho_do_mercado$Combustivel_Agregados)
```

```
## [1] 7584254
```