

Projeto Análise de Séries Temporais

Mayara Yonemura

2022-08-08

Projeto de Conclusão da Disciplina Análise de Séries Temporais

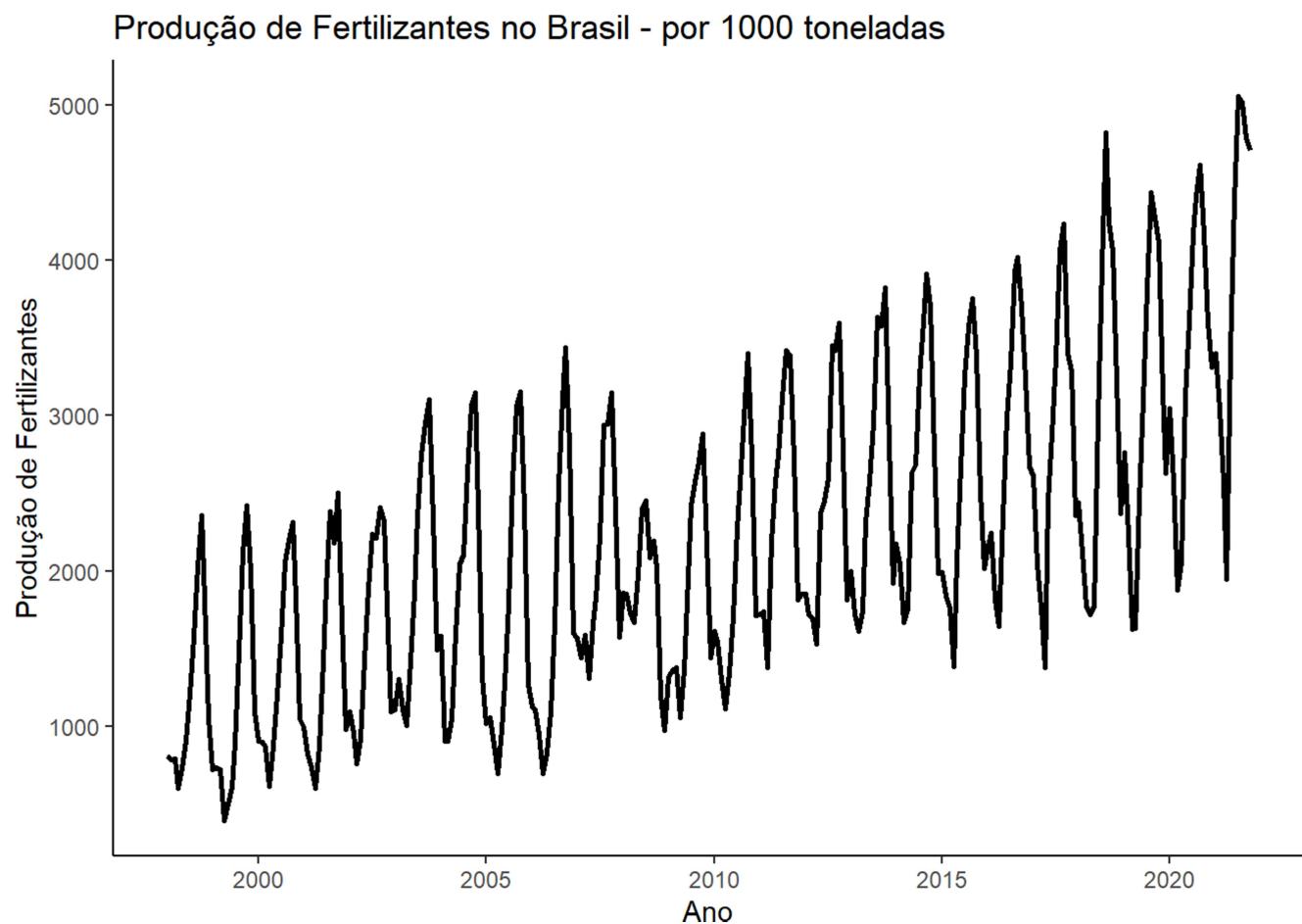
Aluna: Mayara Yonemura - A58337141

Briefing Os dados do arquivo “Fertilizantes”, disponibilizado pela FGV, para a disciplina de Análise de Séries Temporais, contempla a entrega de fertilizantes, produzidos no Brasil, ao mercado em mil toneladas no período mensal de janeiro de 1998 a setembro de 2021.

A fonte dos dados é o sítio da Associação Nacional para Difusão de Adubos –ANDA (<http://anda.org.br/estatisticas/> (<http://anda.org.br/estatisticas/>)).

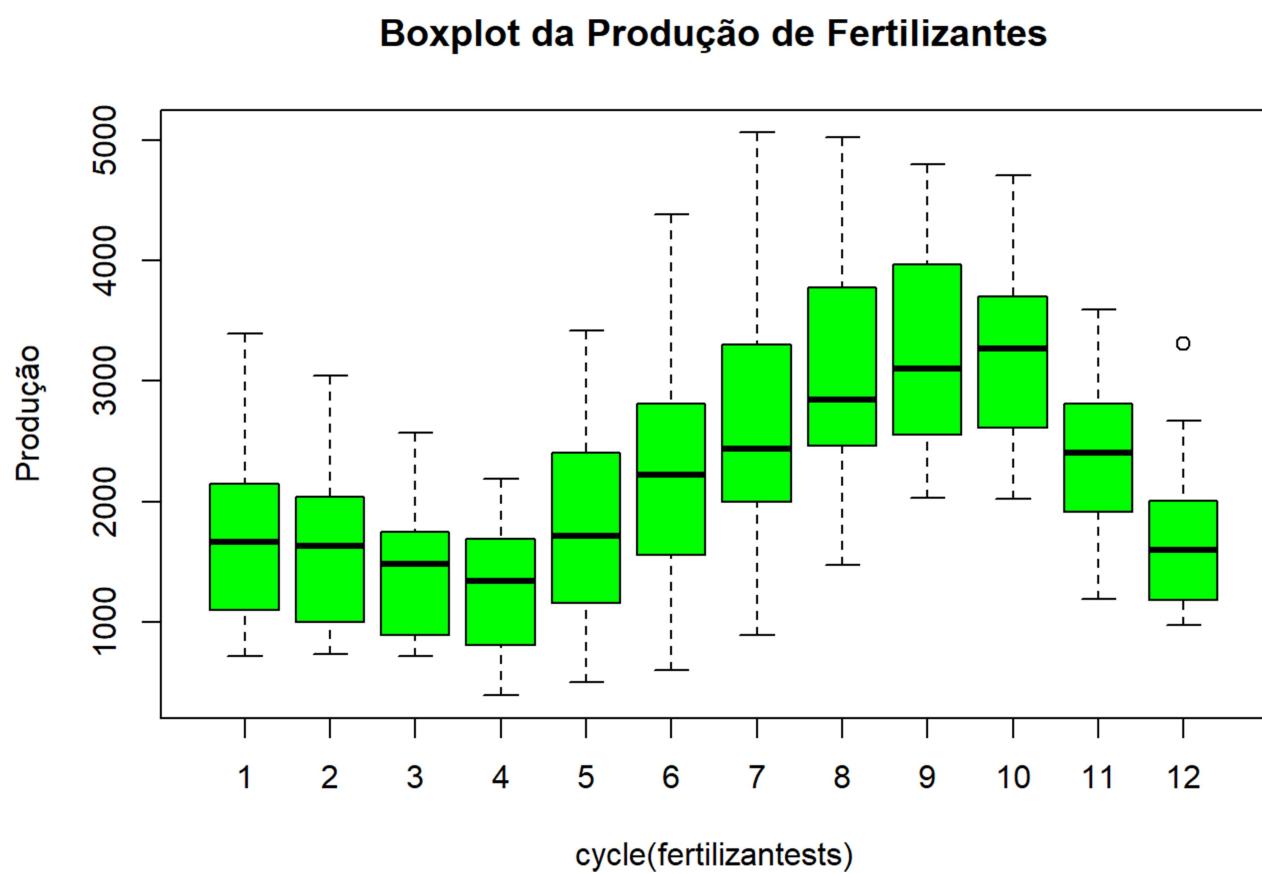
Com a visualização do Plot da Série de Fertilizantes, podemos ter alguns insights interessantes: - A produção está crescendo (tendência de alta) ano após ano, e não tivemos impacto de produção durante o período da pandemia covid-19 (após 2020 até o momento que temos os dados mapeados - Set/21); - Há um padrão de comportamento de produção, com picos e vales, repetidamente, durante todo o período, no qual observamos os ciclos da serie (padrões de comportamento com certa regularidade, no longo prazo); - Aparentemente, estamos visualizando uma serie do tipo Multiplicativa (ou heterocedástica - variância não é constante), pois os comprimentos das linhas entre o máximo e o mínimo (sazonalidade) fica maior ao passar dos anos. Exemplo:

No ano de 2000, a linha atravessa uma banda (1000 a 2000), mas em 2020, atravessa duas bandas (2000 a 4000);



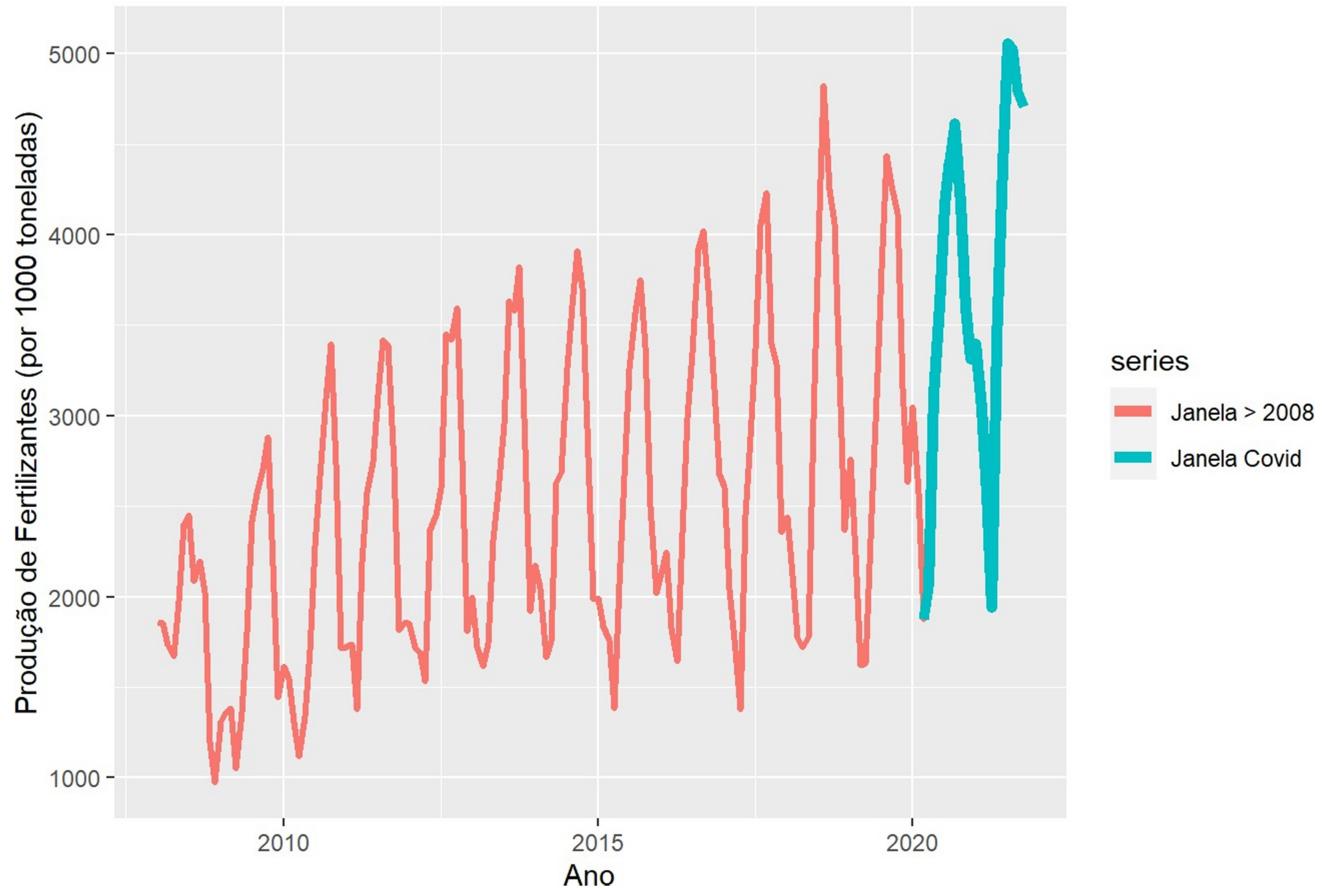
Visualização da Serie Temporal por Boxplot (sazonalidade): - Dentro do ciclo, podemos observar a sazonlidade entre os meses (picos e vales no período de 1 ano, sendo o 1º semestre do ano o mais fraco

(vales) e o 2º semestre mais forte (picos) - observados pelo Boxplot;



Visualização da Série Temporal a partir de Janeiro de 2008, com destaque para a janela temporal da pandemia (a partir de Março de 2020).

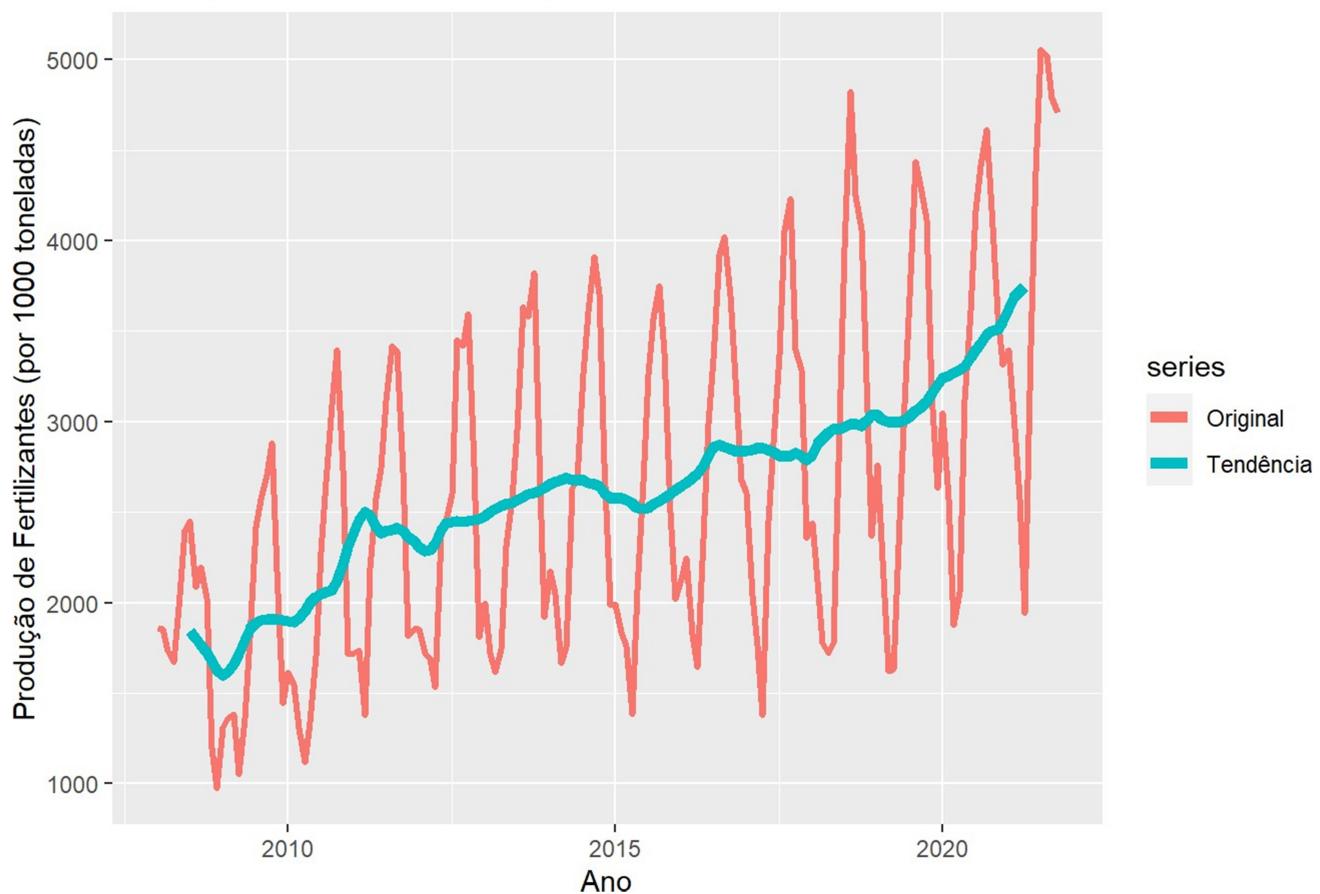
Produção de Fertilizantes - por 1000 toneladas



Para iniciarmos os estudos sobre estacionariedade, tendência, sazonalidade, realizamos a decomposição da serie, que já nos permite verificar a tendência de alta observada (em azul) no Plot

Original da Serie Temporal, que também se manteve na janela dos dados (a partir de 2008).

Produção de Fertilizantes - por 1000 toneladas



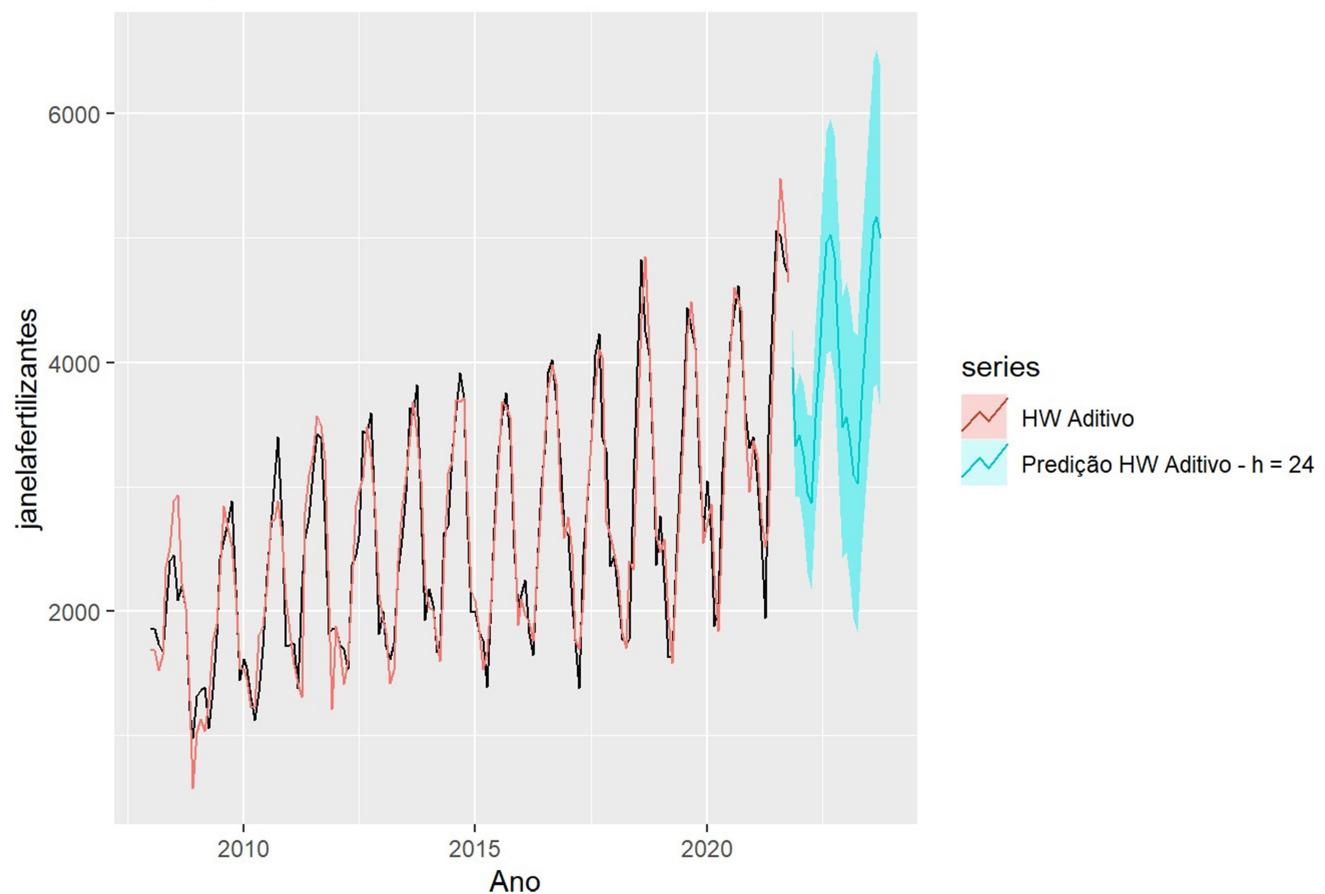
Podemos observar pelos números do RMSE que o modelo de Holt Winters Aditivo possui o menor erro, portanto, este é o melhor modelo para predição da serie temporal de fertilizantes (aditivo 280 x multiplicativo 300).

```
##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -2.160395 280.401 218.0567 -0.6553139 9.238996 0.7020704
##                      ACF1
## Training set 0.01382827
```

```
##               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
## Training set -6.523285 300.3069 231.4157 -1.625082 9.923476 0.7450819 0.2994625
```

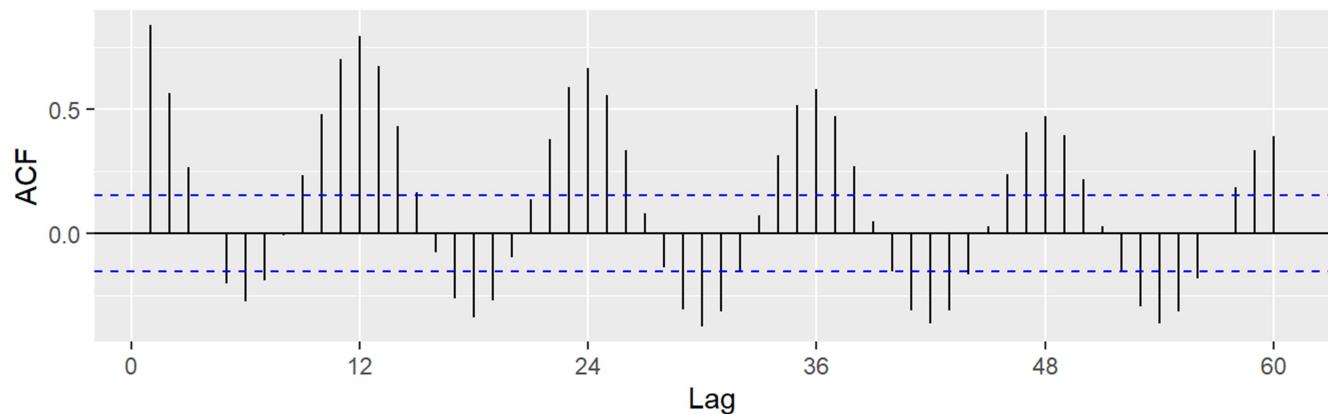
Plot do Modelo Aditivo de Holt-Winters (melhor modelo):

Produção de Fertilizantes RMSP - HW Aditivo

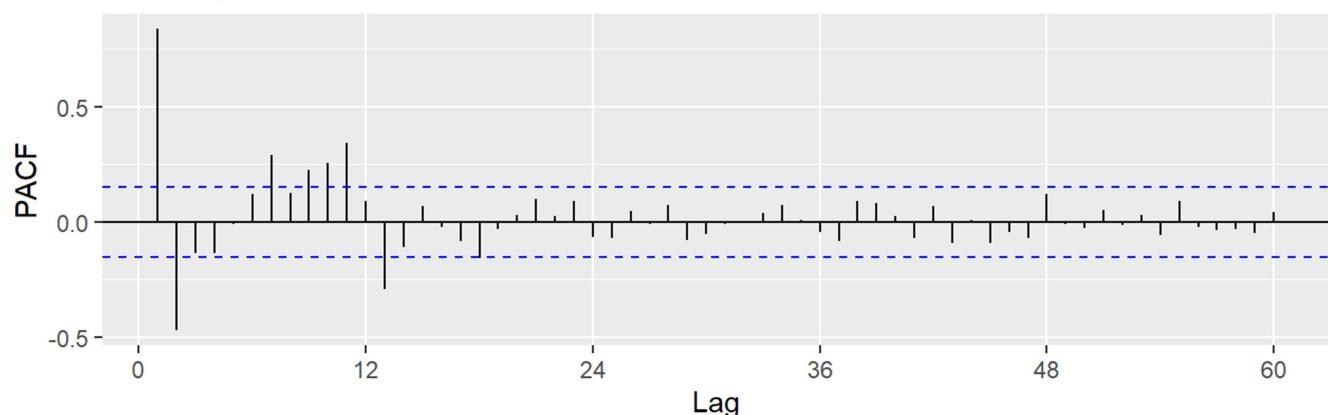


Análise da Estacionariedade da Série Temporal

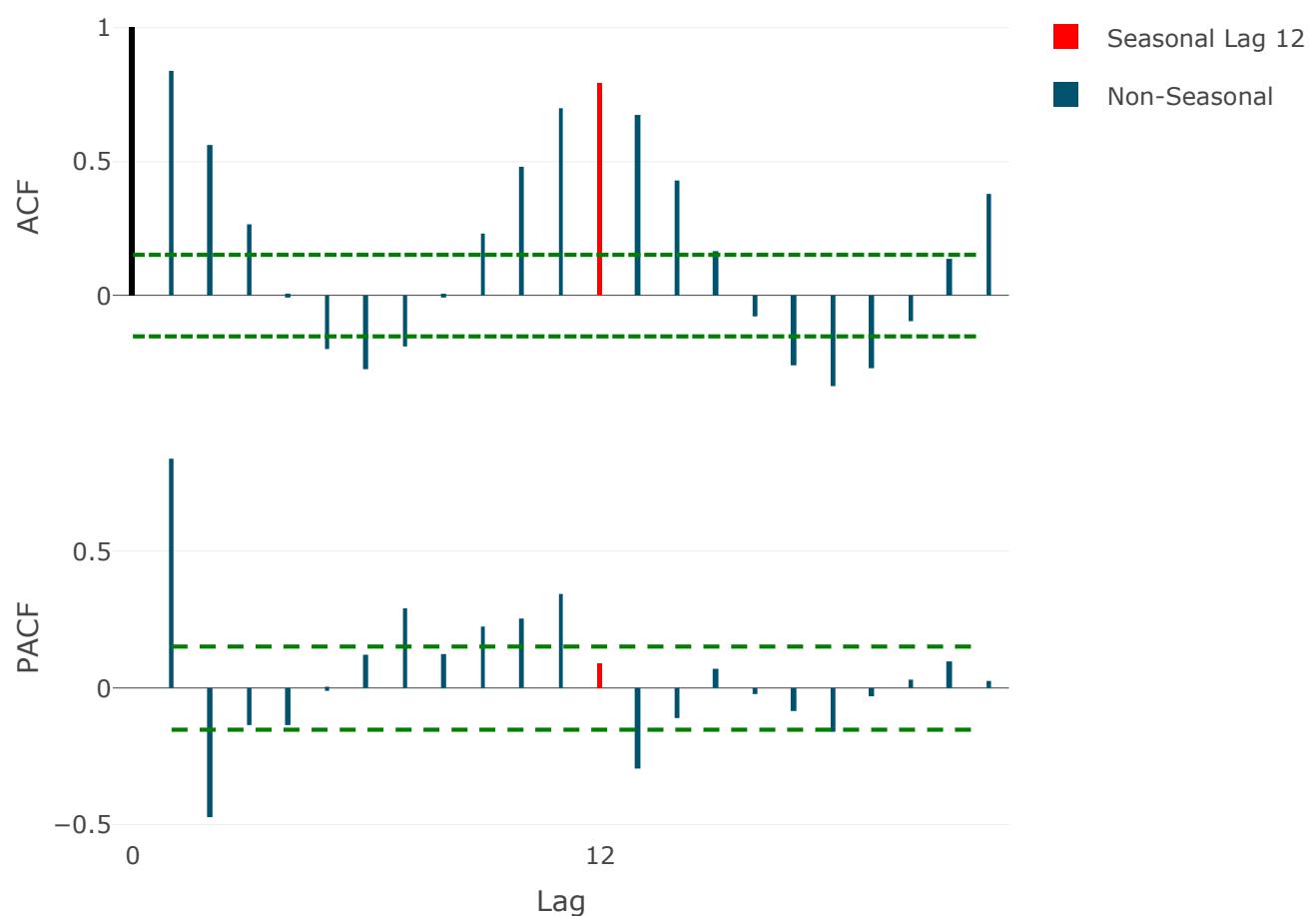
Serries: janelafertilizantes



Serries: janelafertilizantes



janelafertilizantes ACF and PACF Plots



Para entendermos os testes de raiz unitária, basta observarmos o p-valor, sendo válida a regra de Ouro: p-valor baixo rejeita H_0 , ou seja, ficamos com a hipótese alternativa.

Há 3 testes a serem aplicados:

- Augmented Dickey_Fuller (p-valor baixo rejeita H_0)
- Kwiatkowski–Phillips–Schmidt–Shin (KPSS) (p valor baixo aceita H_0)
- Phillip-Perron (p-valor baixo rejeita H_0)

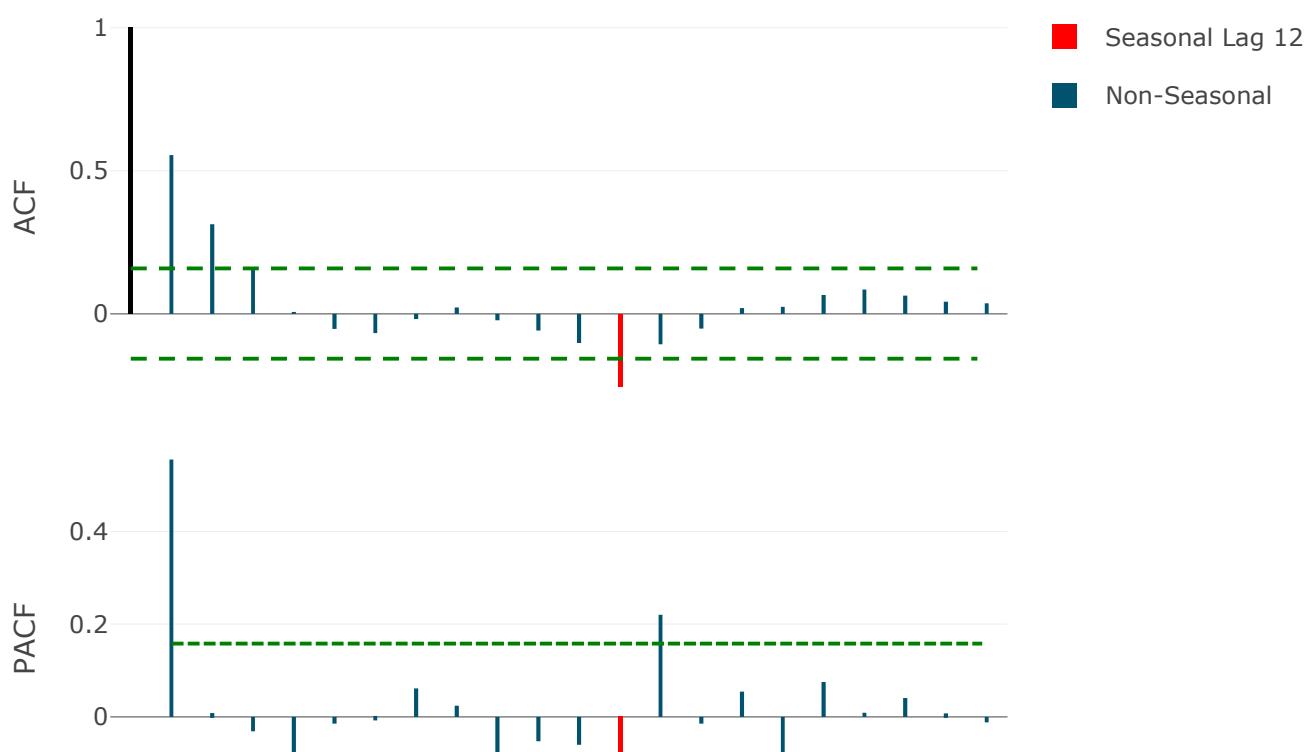
Considerando o grau de confiança de 95% (p valor abaixo de 0,05), de acordo com os testes feitos, a série é estacionária.

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: diff(janelafertilizantes, lag = 12, differences = 1)  
## Dickey-Fuller = -5.0565, Lag order = 5, p-value = 0.01  
## alternative hypothesis: stationary
```

```
##  
## KPSS Test for Level Stationarity  
##  
## data: diff(janelafertilizantes, lag = 12, differences = 1)  
## KPSS Level = 0.18707, Truncation lag parameter = 4, p-value = 0.1
```

```
##  
## Phillips-Perron Unit Root Test  
##  
## data: diff(janelafertilizantes, lag = 12, differences = 1)  
## Dickey-Fuller Z(alpha) = -67.386, Truncation lag parameter = 4, p-value  
## = 0.01  
## alternative hypothesis: stationary
```

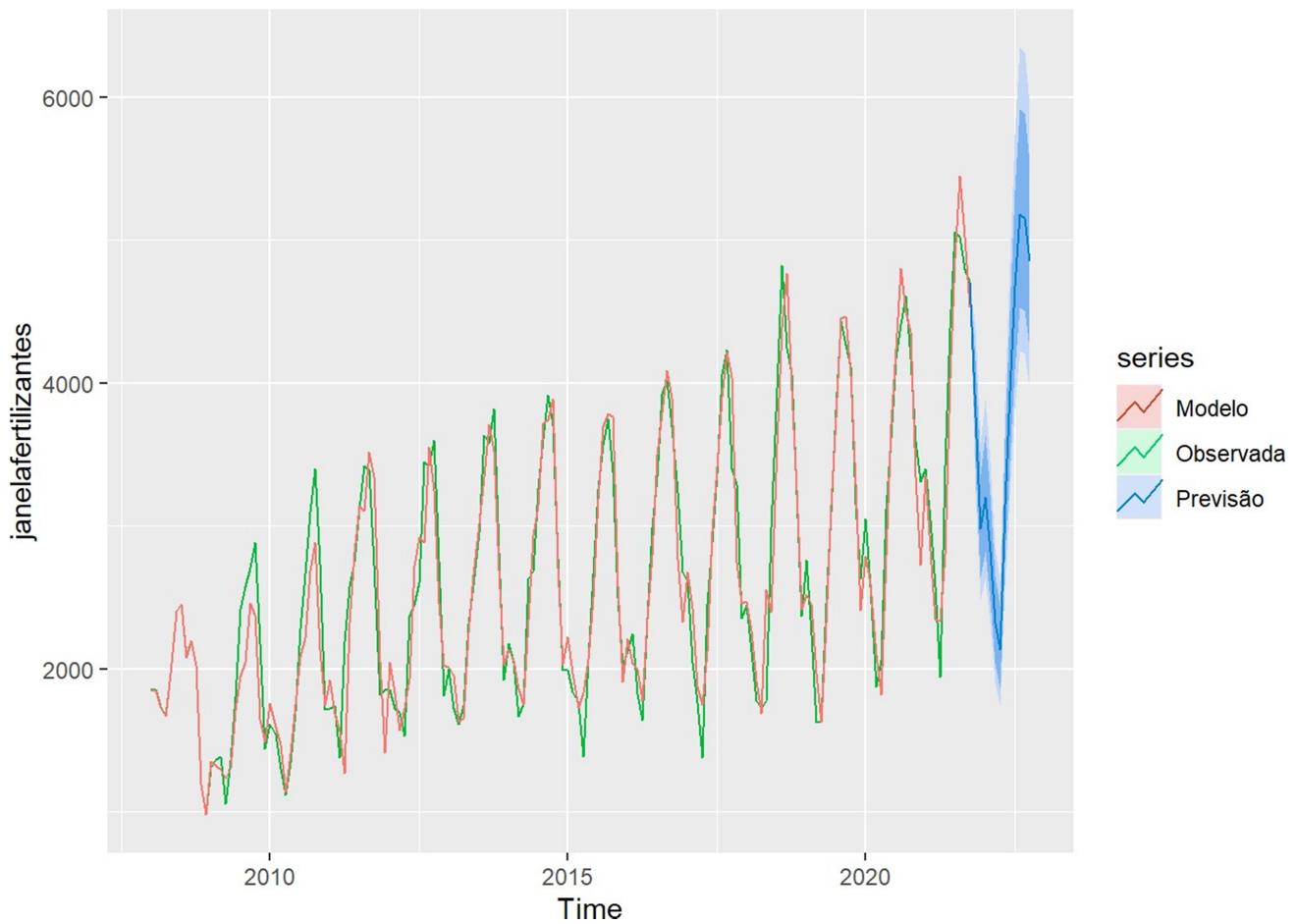
diff(janelafertilizantes, lag = 12, differences = 1) ACF and PACF Plots





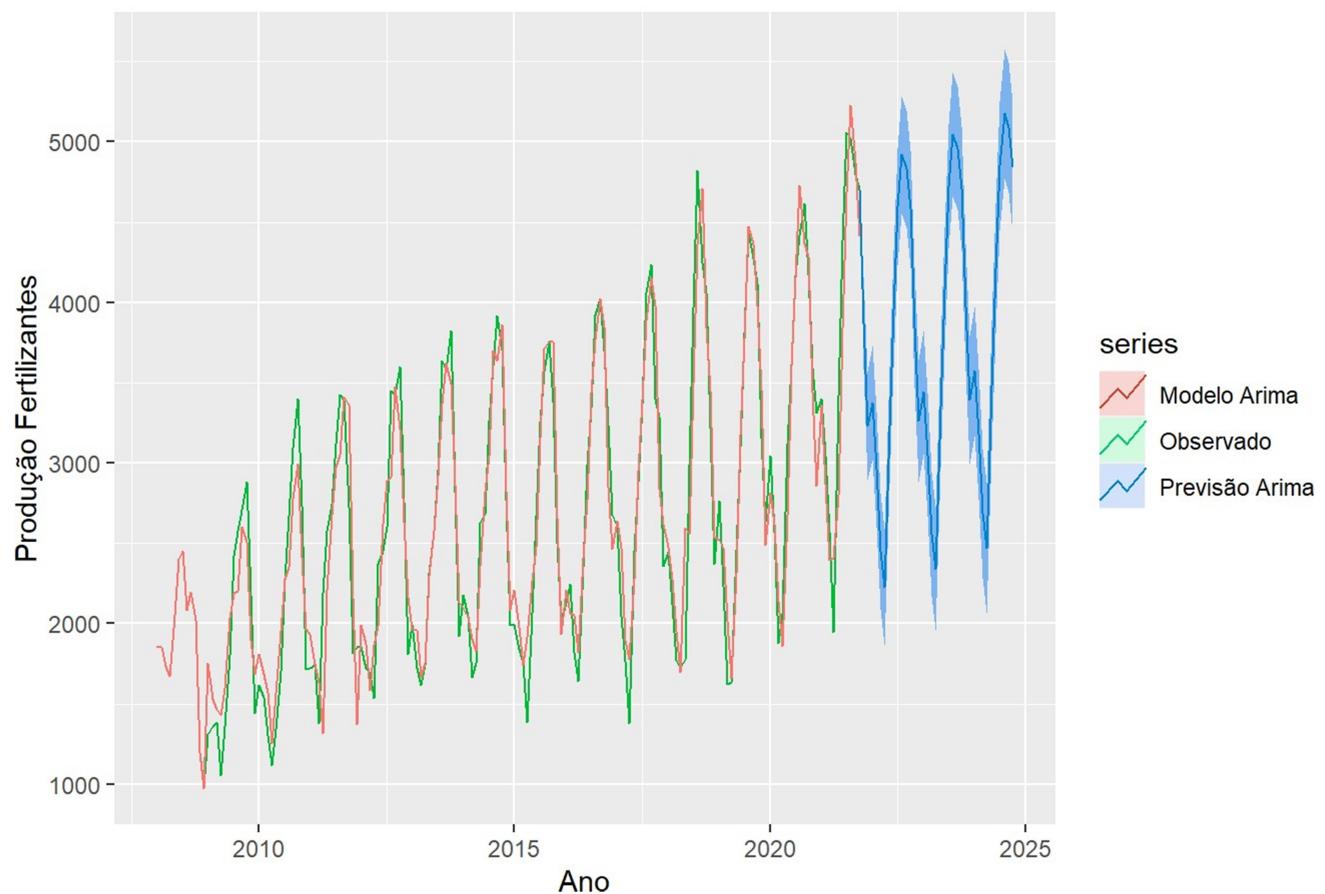
Previsibilidade da Serie Temporal Observando os gráficos de ACF e PACF, com lag = 12, conseguimos descrever um modelo de SARIMA para previsibilidade da serie:

```
## Series: janelafertilizantes
## ARIMA(1,1,3)(1,1,1)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##             ar1      ma1      ma2      ma3      sar1      sma1
##             0.6801  -1.0310 -0.0105  0.0415  0.0151 -0.8012
## s.e.     0.1446   0.1639   0.1396  0.1326  0.1146   0.0962
##
## sigma^2 = 0.0139: log likelihood = 104.17
## AIC=-194.34  AICc=-193.57  BIC=-173.13
```



E vamos testar a mesma previsibilidade com o modelo Auto-Arima:

Produção Fertilizantes + Predição Arima



Comparando os dois modelos (Auto Arima e SARIMA) com o método AKAIKE – O melhor modelo, com menor erro, é o modelo de SARIMA.

```
## [1] 2180.374
```

```
## [1] -194.3417
```

Ao encontrar o melhor modelo de previsibilidade, podemos de fato, verificar o quanto o modelo foi assertivo nos meses subsequentes com os valores reais já divulgados:

	Point Forecast	Lo 60	Hi 60	Lo 80	Hi 80
## Nov 2021	3765.782	3407.884	4161.268	3234.553	4384.258
## Dec 2021	2980.137	2643.931	3359.095	2483.568	3575.990
## Jan 2022	3201.192	2819.903	3634.037	2639.029	3883.107
## Feb 2022	2826.143	2481.128	3219.134	2317.886	3445.847
## Mar 2022	2323.466	2036.381	2651.023	1900.725	2840.229
## Apr 2022	2135.530	1870.027	2438.728	1744.653	2613.980
## May 2022	3156.137	2762.435	3605.949	2576.592	3866.037
## Jun 2022	3946.121	3452.951	4509.729	3220.203	4835.680
## Jul 2022	4679.913	4094.389	5349.171	3818.089	5736.269
## Aug 2022	5179.864	4531.344	5921.200	4225.339	6350.020
## Sep 2022	5154.739	4509.070	5892.864	4204.427	6319.847
## Oct 2022	4853.986	4245.788	5549.308	3958.834	5951.546

Dados Observados (reais) e dados previstos, respectivamente modelo Arima e HW Aditivo:

- Outubro-21 4.706
- Novembro-21 4.201 / 3765 (89%) / 3958 (94%)
- Dezembro-21 3.314 / 2980 (89%) / 3330 (100%)
- Janeiro-22 3.223 / 3201 (99%) / 3414 (100%)
- Fevereiro-22 2.509 / 2826 (112%) / 3256 (129%)
- Março-22 2.952 / 2323 (78%) / 2944 (100%)
- Abril-22 2.713 / 2135 (78%) / 2870 (100%)

Conclusão sobre a Serie Temporal de Fertilizantes

Os níveis de assertividade do modelo Arima estão dentro das bandas esperadas de intervalo de confiança, e seria para uma projeção mais conversadora, que particularmente, eu adotaria - como estamos tratando de commodities seria melhor prever menos e ter um retorno maior, do que projetar e realizar exatamente o projetado, pois o não atingimento impactaria muito mais o mercado. Assim, os resultados do modelo Arima fazem total sentido para que possamos fazer uma leitura, e consequentemente uma expectativa bastante realista (e conservadora) da produção de fertilizantes. Os meses finais do ano refere-se aos ciclos de alta (picos) e os primeiros meses do ano referem-se aos ciclos de baixa (vales), assim sendo, podemos observar que estamos crescendo além do esperado pelo modelo (a serie história está crescendo mais), pois todas as previsões estão abaixo - as bandas entre High 60 e High 80 poderiam virar as referências de meta de produção, ainda trabalhando dentro do modelo com números factíveis de atingimento (projeções de produção, receita, valores, etc). Podemos também concluir que, por se tratar de commodities, os fertilizantes não sofreram o impacto da pandemia como demais setores sofreram, e que a projeção/tendência de alta se manteve entre os ciclos e as previsibilidades tem se concretizados ainda melhores.

Códigos Utilizados e respectivos comentários

Carregando as bibliotecas

```
library(knitr)
library(rmarkdown)
library(htmltools)
library(gridExtra)

library(readxl)
library(sf)
library(spatstat)
library(tidyverse)

library(tseries)
library(stats)
library(ggplot2)
library(ggthemes)
library(ggmap)
library(fpp2)
library(BETS)
library(urca)
library(TSA)
library(forecast)
library(MVar)
library(TSstudio)
library(bayesforecast)
library(NlinTS)
```

```
head(fertilizantes)
```

```
## # A tibble: 6 × 2
##   data           fertilizantes
##   <dttm>             <dbl>
## 1 1998-01-01 00:00:00     808
## 2 1998-02-01 00:00:00     784
## 3 1998-03-01 00:00:00     793
## 4 1998-04-01 00:00:00     600
## 5 1998-05-01 00:00:00     726
## 6 1998-06-01 00:00:00     911
```

Podemos observar pelo head que a frequência temporal é mensal ($f = 12$).

– Dimensões do Dataset - linhas e colunas, respectivamente:

```
## [1] 286    2
```

Contagem de dados nulos por coluna:

```
colSums(is.na(fertilizantes))
```

```
##          data fertilizantes
##          0             0
```

Datatypes do Dataset:

```
str(fertilizantes)
```

```
## # tibble [286 x 2] (S3: tbl_df/tbl/data.frame)
## $ data      : POSIXct[1:286], format: "1998-01-01" "1998-02-01" ...
## $ fertilizantes: num [1:286] 808 784 793 600 726 ...
```

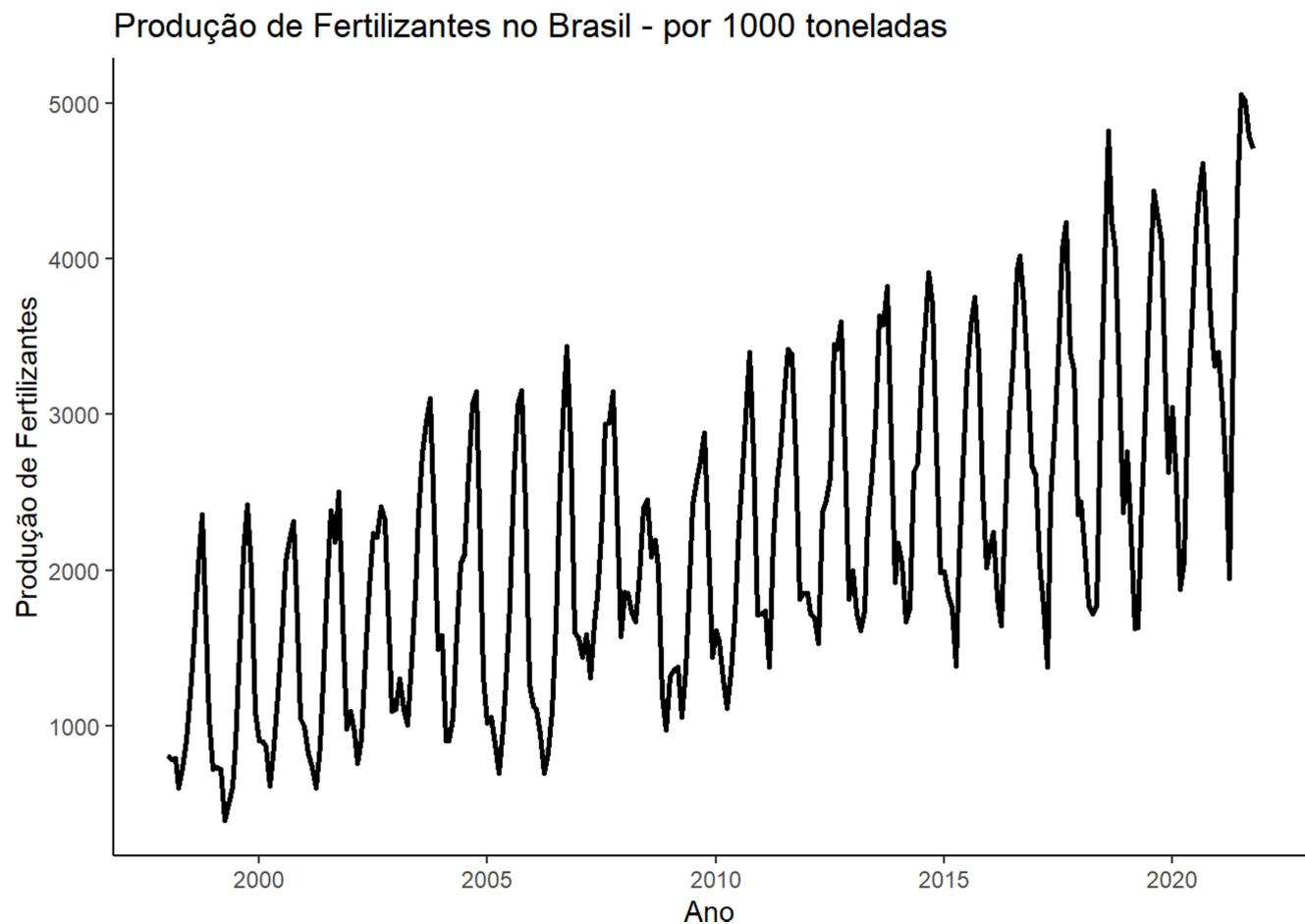
Como a classe de 'fertilizantes' é numérica, para Time Series precisa transformar em formato ts:

```
## [1] "ts"
```

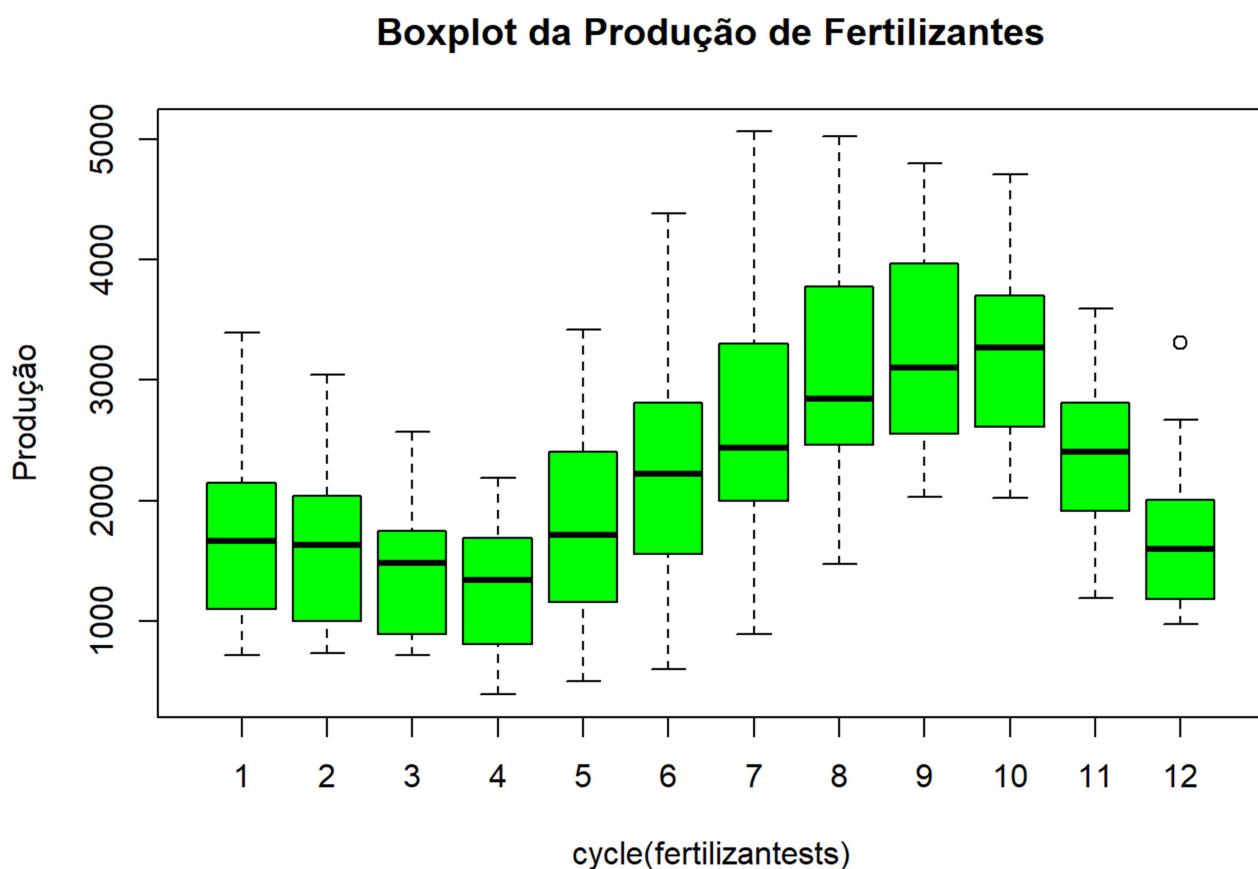
Visualização do formato da ts após transformação (númerico > timeseries)

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1998  808 784 793 600 726 911 1207 1575 2030 2361 1813 1048
## 1999  721 737 720 396 502 608 892 1478 2119 2422 1998 1095
## 2000  902 900 866 616 871 1204 1581 2056 2203 2311 1828 1051
## 2001 1001 823 732 603 875 1414 1898 2381 2180 2506 1678 978
## 2002 1094 953 760 920 1428 1889 2238 2214 2411 2319 1789 1096
## 2003 1110 1304 1108 1007 1423 1859 2350 2748 2954 3105 2336 1490
## 2004 1584 902 902 1052 1598 2044 2104 2636 3063 3150 2403 1325
## 2005 1017 1059 886 699 983 1278 1792 2588 3058 3154 2409 1271
## 2006 1124 1110 950 697 829 1092 1666 2554 3112 3438 2813 1597
## 2007 1571 1442 1587 1307 1654 1829 2219 2937 2949 3150 2387 1574
## 2008 1857 1852 1733 1671 1973 2404 2452 2084 2197 2023 1196 977
## 2009 1313 1362 1382 1054 1359 1804 2419 2574 2705 2883 2174 1442
## 2010 1614 1542 1306 1116 1342 1705 2279 2667 3095 3400 2732 1715
## 2011 1720 1740 1378 2192 2578 2745 3117 3422 3391 2725 1816 1852
## 2012 1852 1722 1692 1532 2370 2451 2601 3450 3422 3597 2754 1812
## 2013 1996 1719 1614 1742 2314 2579 2954 3635 3579 3823 2817 1923
## 2014 2176 2045 1667 1755 2629 2683 3262 3606 3914 3706 2772 1988
## 2015 1994 1839 1761 1383 2067 2667 3257 3569 3754 3383 2503 2020
## 2016 2129 2245 1824 1643 2354 2986 3346 3924 4022 3698 3235 2676
## 2017 2609 2044 1765 1379 2450 2883 3369 4058 4234 3398 3288 2358
## 2018 2443 2119 1774 1722 1775 2985 3964 4824 4247 4061 3220 2369
## 2019 2762 2235 1624 1634 2458 3099 3811 4437 4270 4116 3155 2632
## 2020 3047 2572 1876 2060 3120 3611 4169 4415 4614 4173 3595 3312
## 2021 3398 3042 2572 1945 3418 4384 5059 5021 4794 4706
```

Plot da Serie Temporal



Visualização da Serie Temporal por Boxplot (sazonalidade):

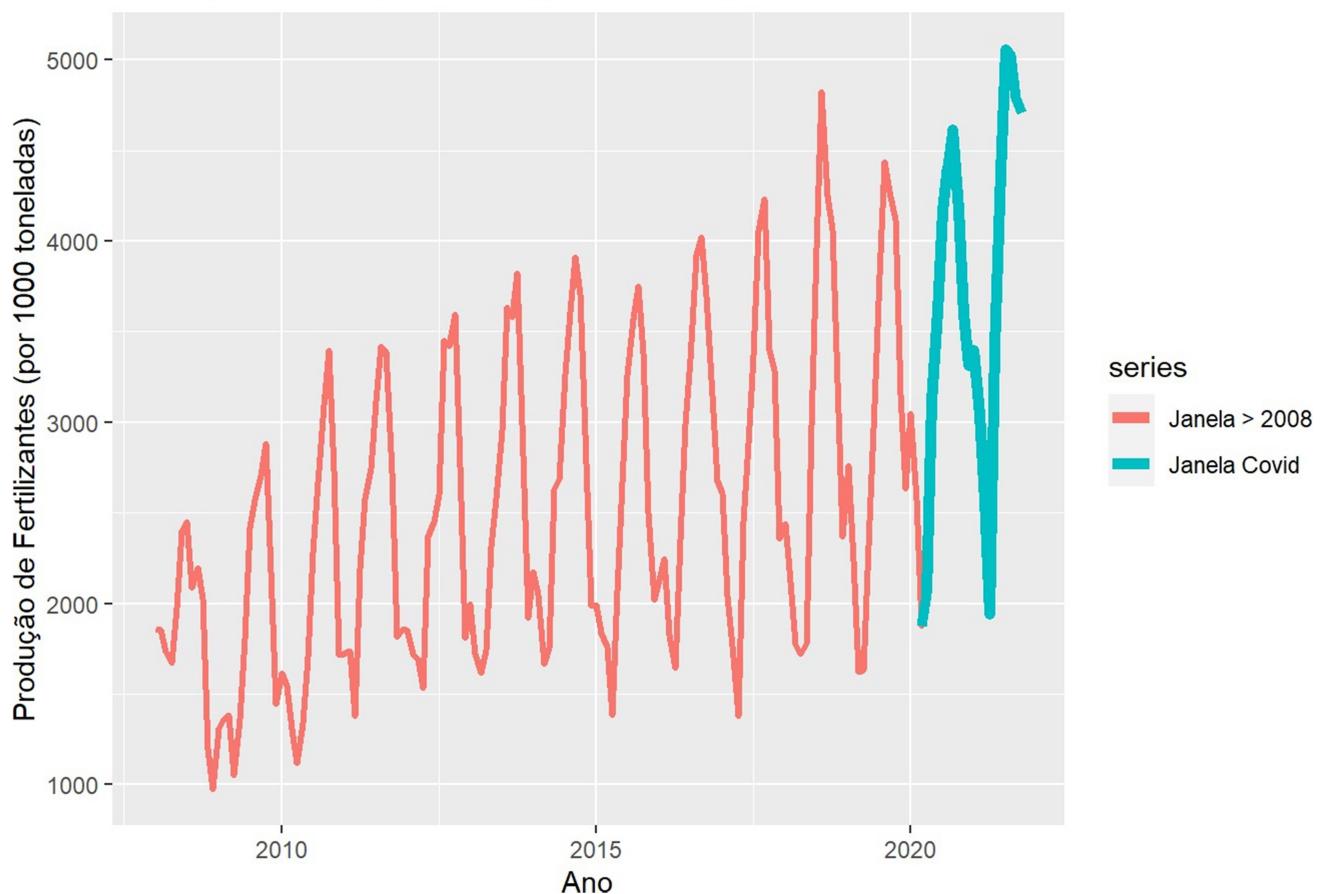


Com a visualização do Plot da Série de Fertilizantes, podemos ter alguns insights interessantes: - A produção

está crescendo (tendência de alta) ano após ano, e não tivemos impacto de produção durante o período da pandemia covid-19 (após 2020 até o momento que temos os dados mapeados - Set/21); - Há um padrão de comportamento de produção, com picos e vales, repetidamente, durante todo o período, no qual observamos os ciclos da série (padrões de comportamento com certa regularidade, no longo prazo); - Dentro do ciclo, podemos observar a sazonalidade entre os meses (picos e vales no período de 1 ano, sendo o 1º semestre do ano o mais fraco (vales) e o 2º semestre mais forte (picos) - observados pelo Boxplot; - Aparentemente, estamos visualizando uma série do tipo Multiplicativa (ou heterocedástica - variância não é constante), pois os comprimentos das linhas entre o máximo e o mínimo (sazonalidade) fica maior ao passar dos anos. Exemplo: No ano de 2000, a linha atravessa uma banda (1000 a 2000), mas em 2020, atravessa duas bandas (2000 a 4000);

Visualização da Série Temporal a partir de Janeiro de 2008, com destaque para a janela temporal da pandemia (a partir de Março de 2020).

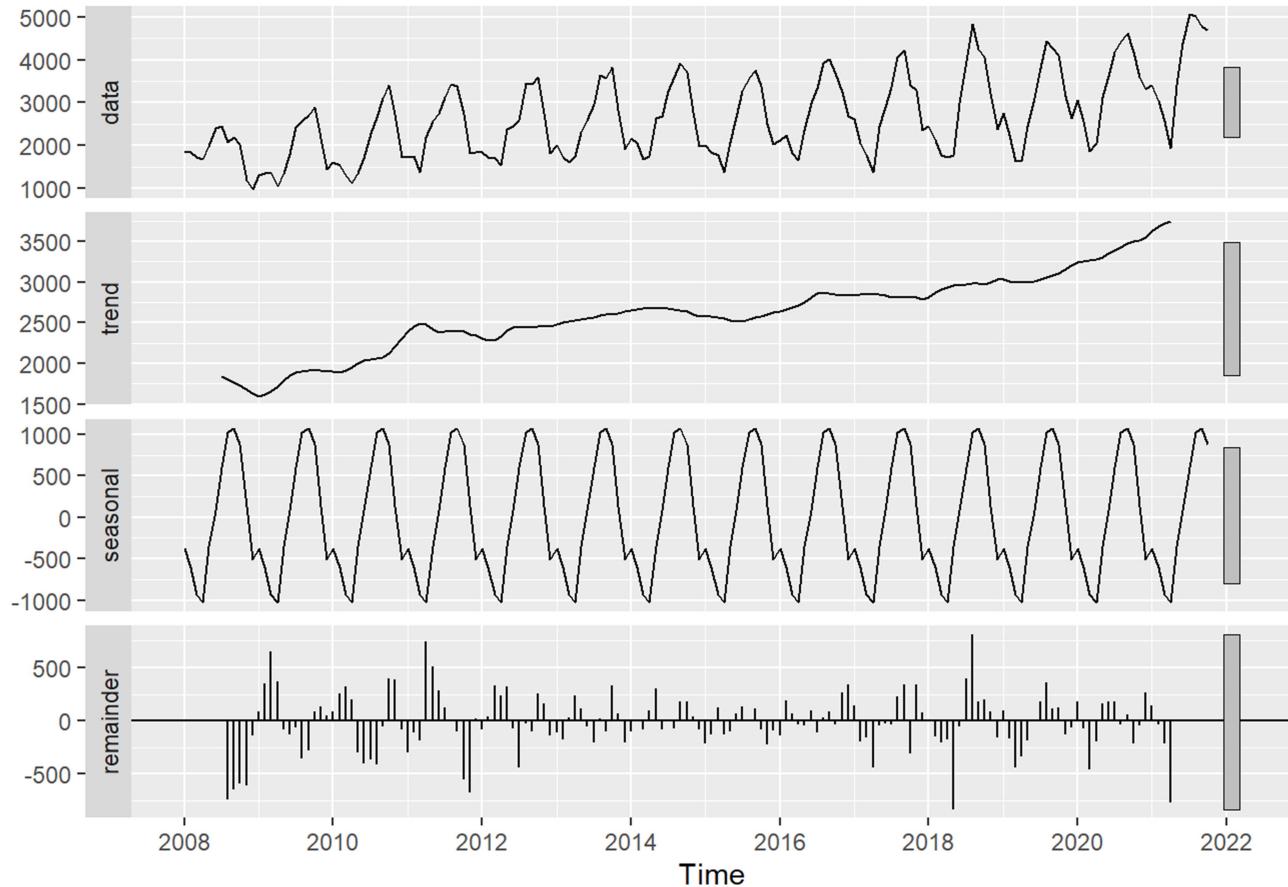
Produção de Fertilizantes - por 1000 toneladas



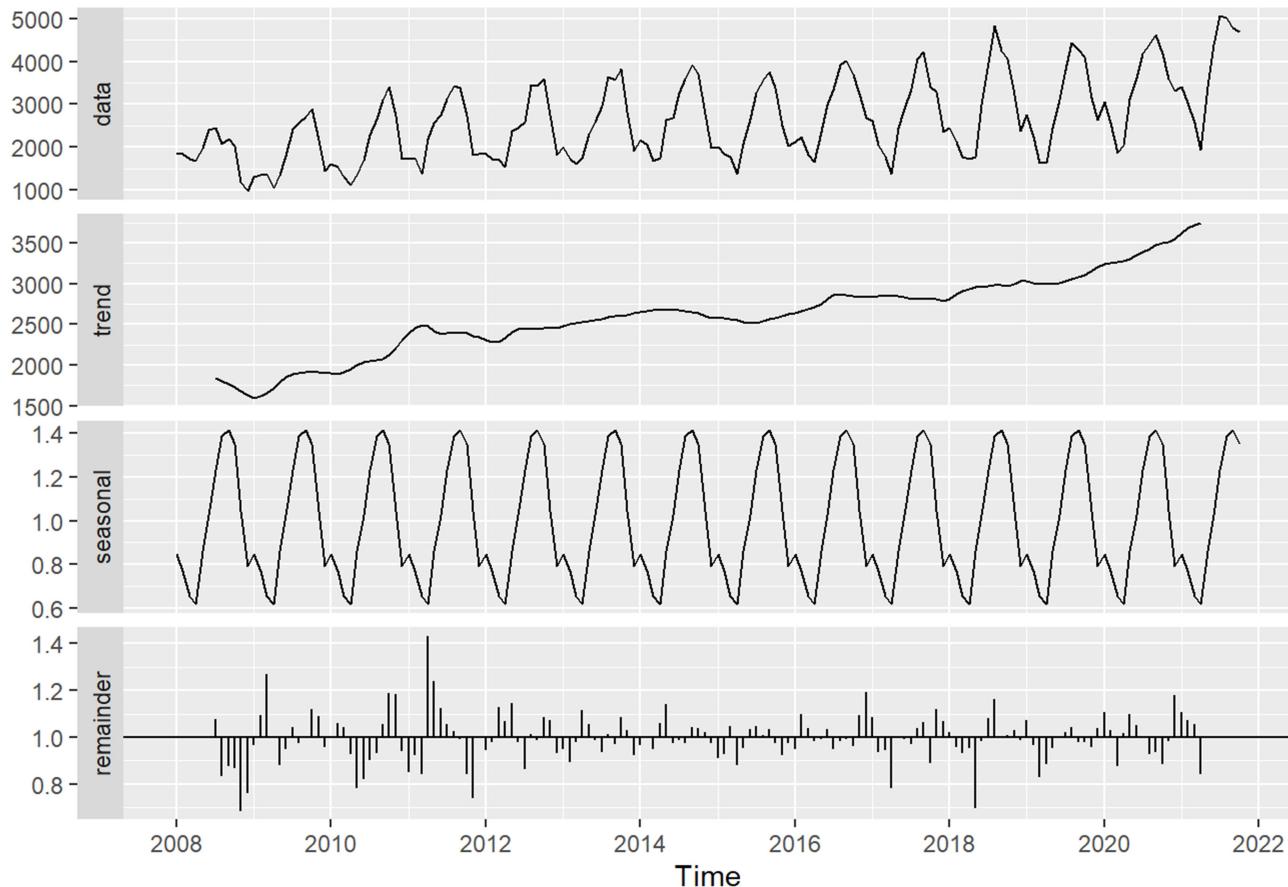
Decomposição da série

- Para encontrarmos o melhor modelo, o fato decisivo será o erro encontrado em ambos os modelos, sendo o melhor modelo aquele que possui o menor erro.

Decomposition of additive time series

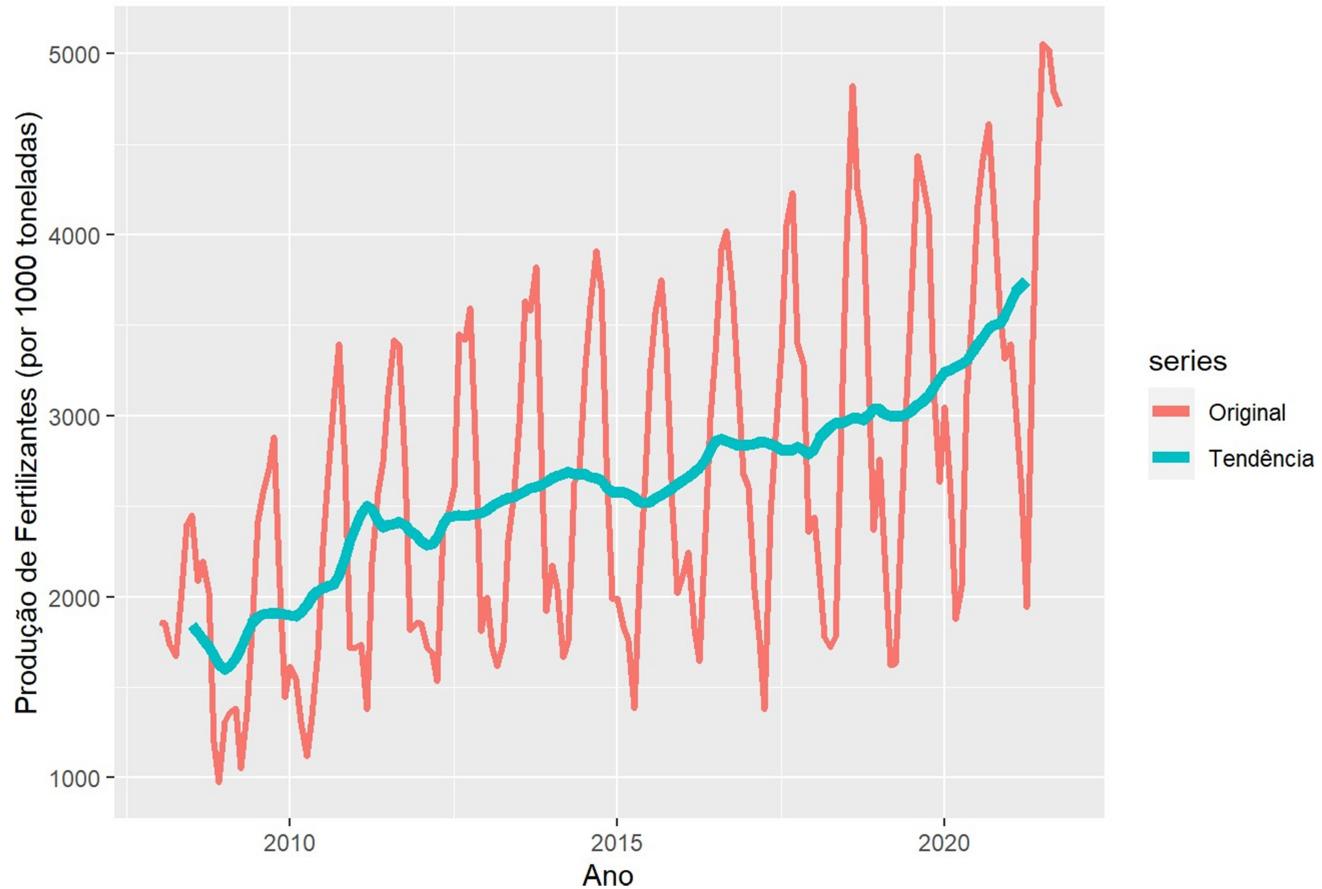


Decomposition of multiplicative time series



Visualização da série da janela dos dados (a partir de 2008) com a tendência de alta destacada em azul:

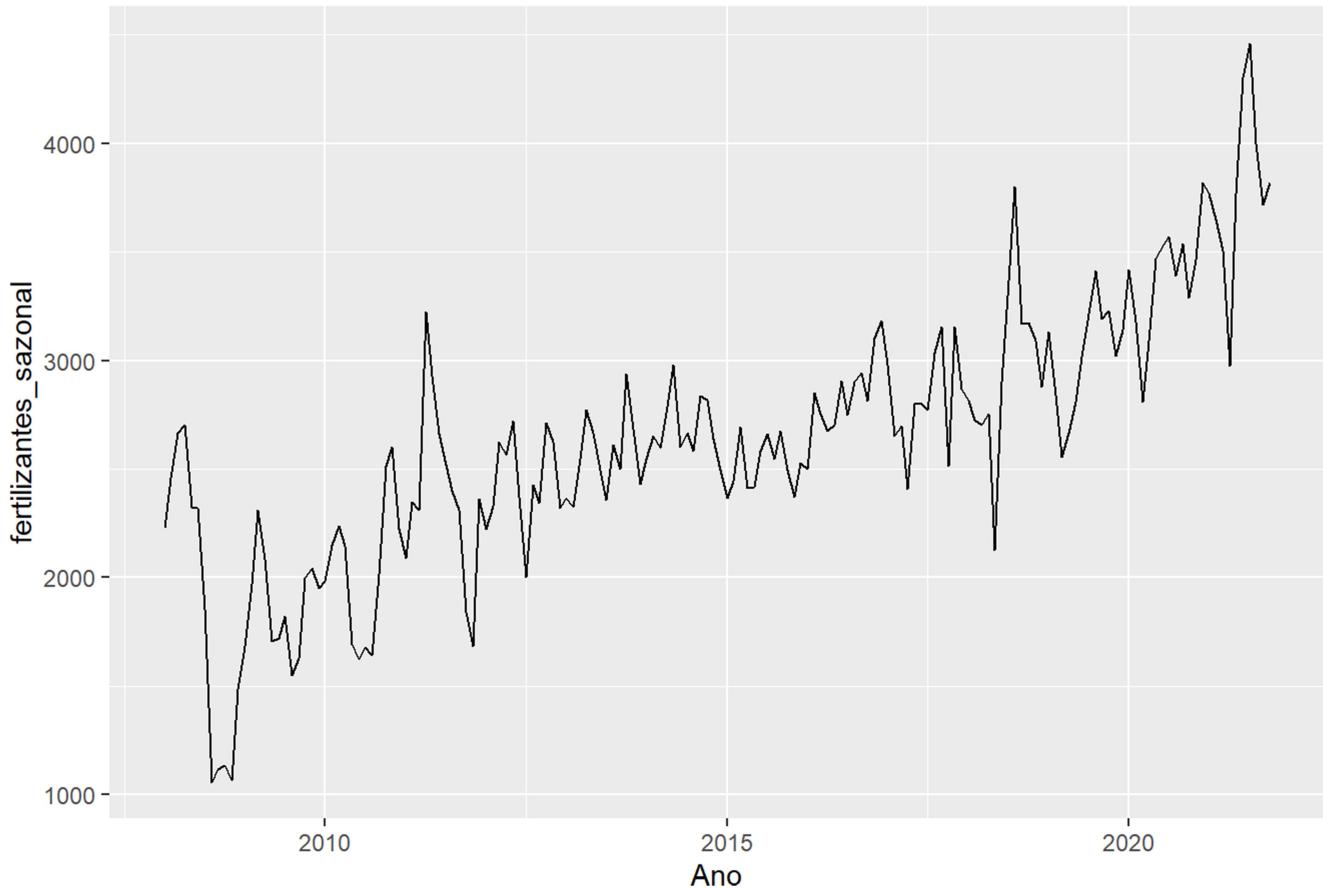
Produção de Fertilizantes - por 1000 toneladas



Análise exploratória da Sazonalidade da Série

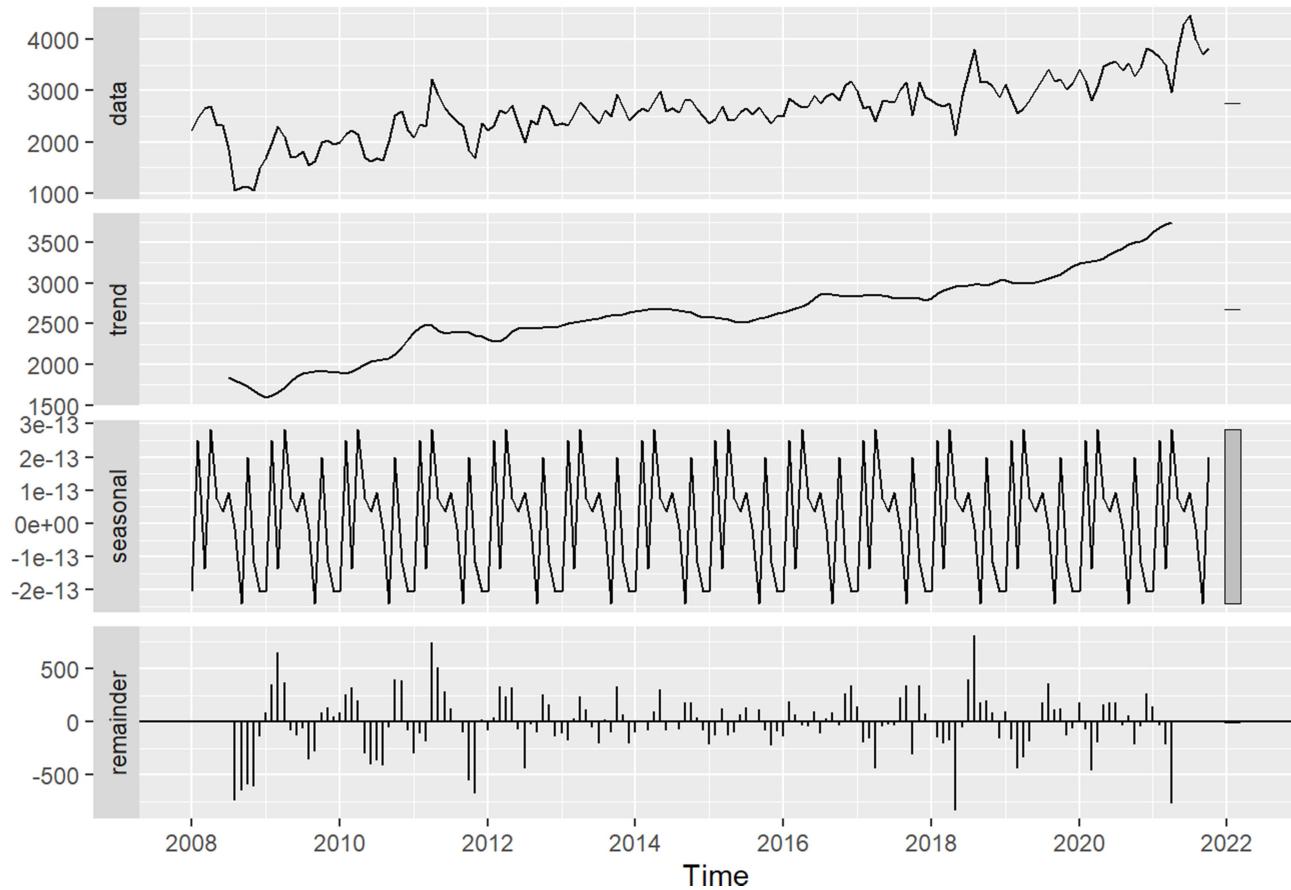
Plot da Sazonalidade da série:

Plot da Sazonalidade da Serie



Dessazonalidade da série (removendo a sazonalidade da serie original):

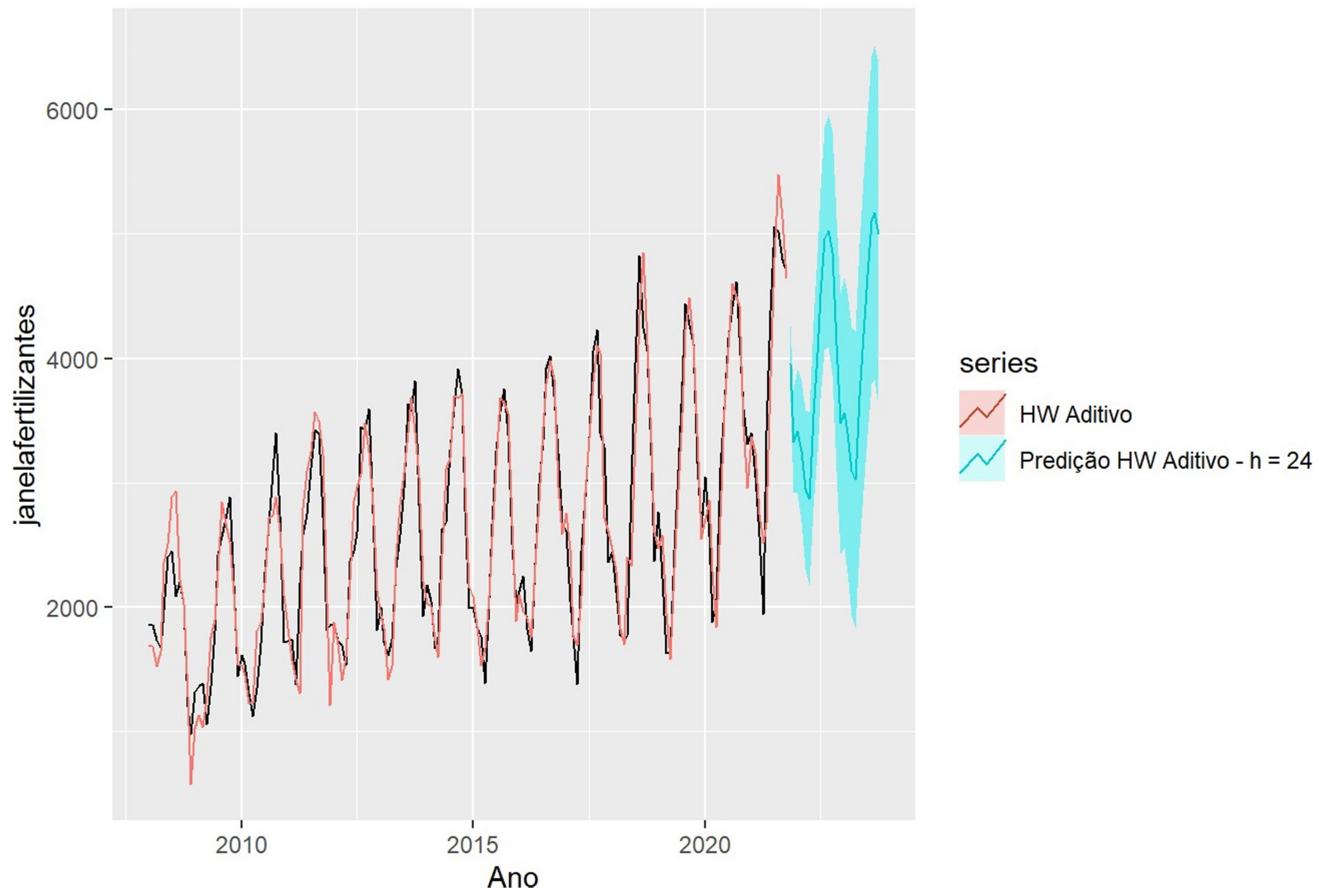
Decomposition of additive time series



Ao retirar a sazonalidade da série, reduzimos significativamente o erro do modelo (nos plots acima, o gráfico que demonstra sazonalidade está em escala de 0.75 a 1.5, e removendo a sazonalidade, o mesmo gráfico é plotado em escala menores - 0 com 13 casas decimais);

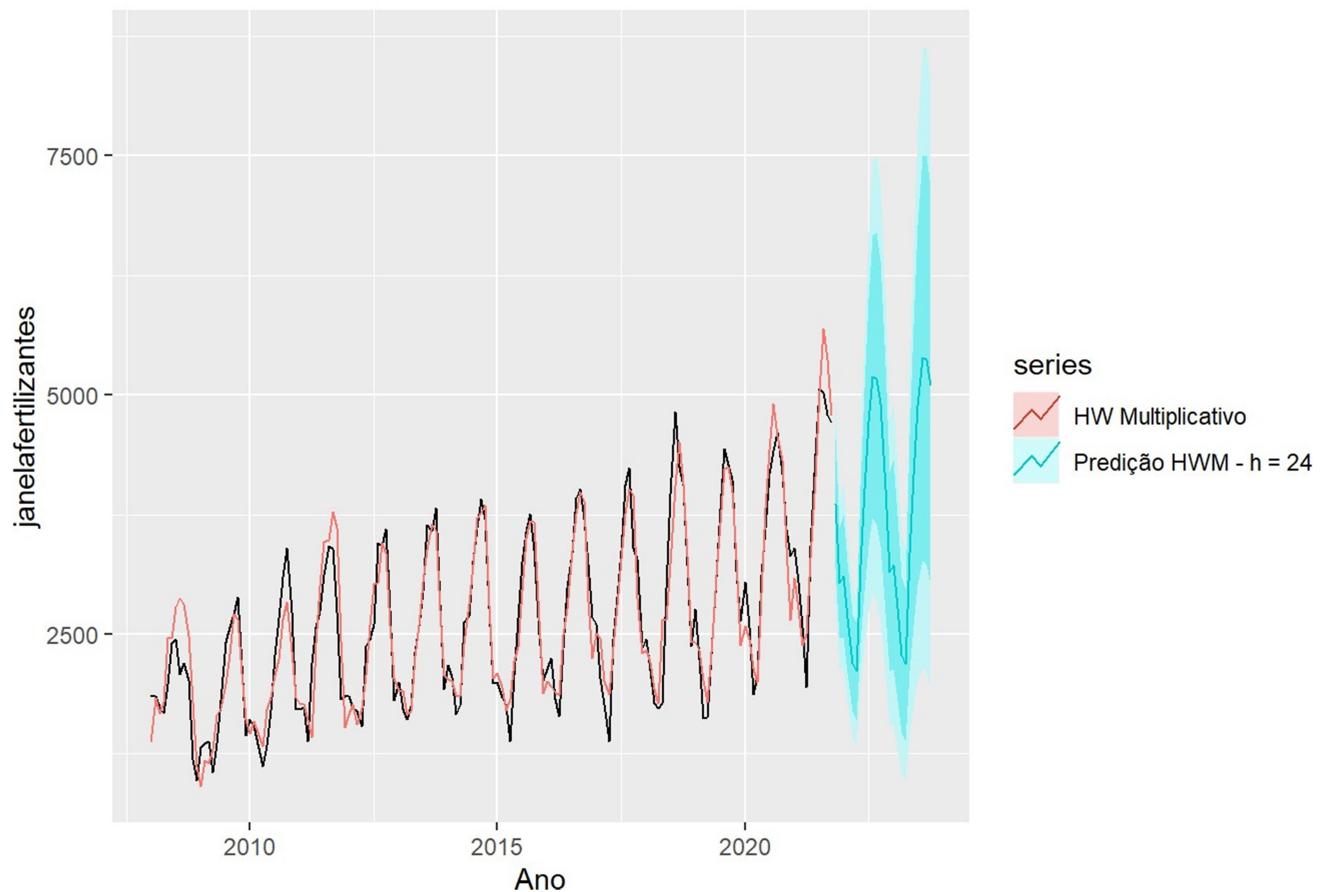
Métodos de Holt-Winters (HW) - Aditivo

Produção de Fertilizantes RMSP - HW Aditivo



- Multiplicativo

Produção de Fertilizantes RMSP - HWm



- Fator de Decisão para escolha do melhor modelo : menor erro (RMSE)

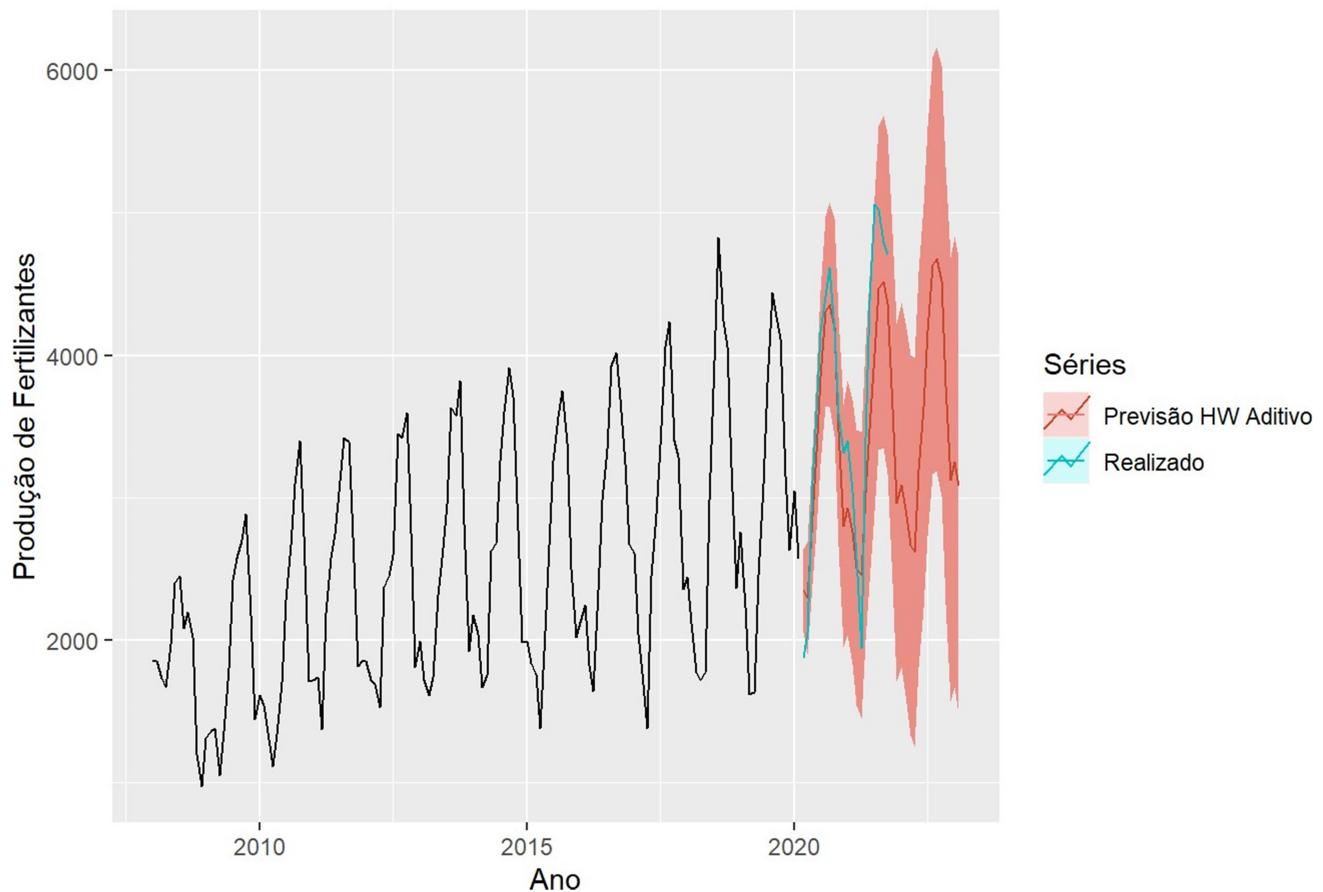
```
##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -2.160395 280.401 218.0567 -0.6553139 9.238996 0.7020704
##                         ACF1
## Training set 0.01382827
```

```
##               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
## Training set -6.523285 300.3069 231.4157 -1.625082 9.923476 0.7450819 0.2994625
```

Podemos observar pelos números do RMSE que o modelo de Holt Winters Aditivo possui o menor erro, portanto, este é o melhor modelo para predição da série temporal de fertilizantes (aditivo 280 x multiplicativo 300).

Treinando Modelo de Predição da Serie Temporal pelo modelo Aditivo (melhor modelo segundo erros):

Produção de Fertilizantes RMSP

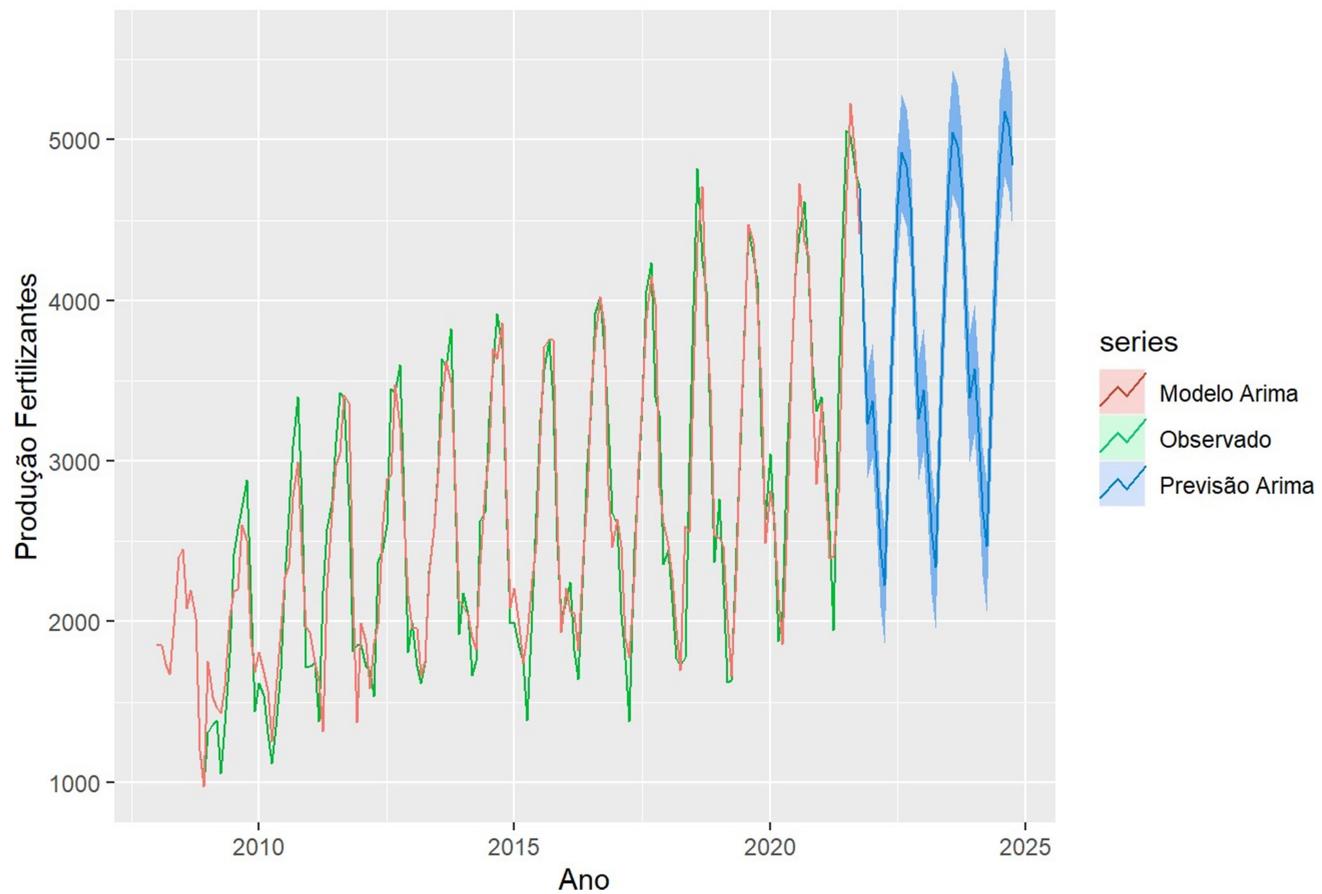


O período selecionado foi intencional para entender o comportamento da série (reflexo da pandemia), devido ao insight inicial de que não houve efeito na produção, e podemos observar que, de fato, o realizado ultrapassou o modelo preditivo, para ambos os lados, superou o máximo e o mínimo (por pouco desvio das linhas, observado no plot) durante a pandemia; dessa forma, concluímos que a pandemia, para a produção de fertilizantes, não teve impacto negativo pois se manteve dentro das bandas preditoras da sazonalidade.

Modelo SARIMA:

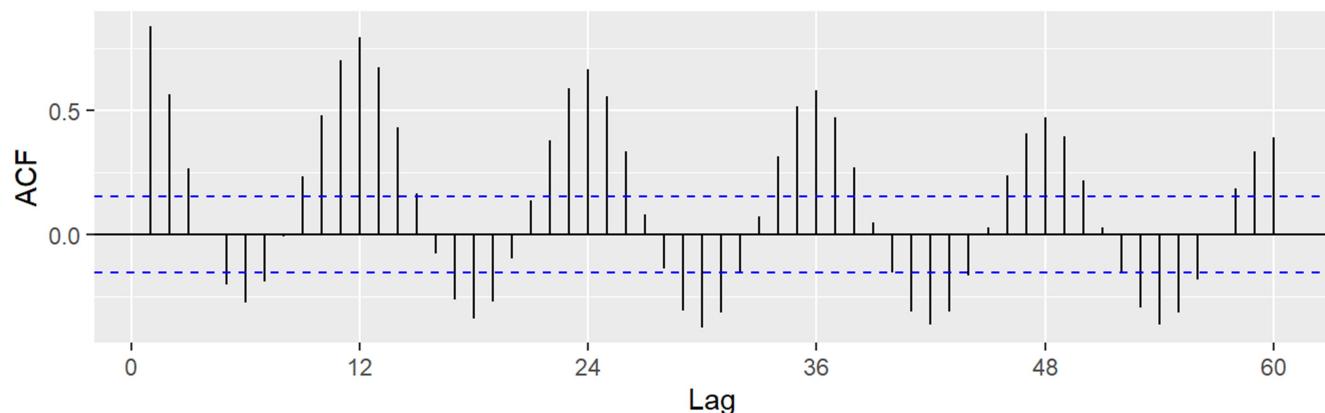
```
## Series: janelafertilizantes
## ARIMA(1,0,0)(0,1,1)[12] with drift
##
## Coefficients:
##          ar1     sma1     drift
##          0.6206 -0.6659 10.7654
## s.e.  0.0644  0.0777  1.9237
##
## sigma^2 = 76065: log likelihood = -1086.19
## AIC=2180.37   AICc=2180.64   BIC=2192.52
```

Produção Fertilizantes + Predição Arima

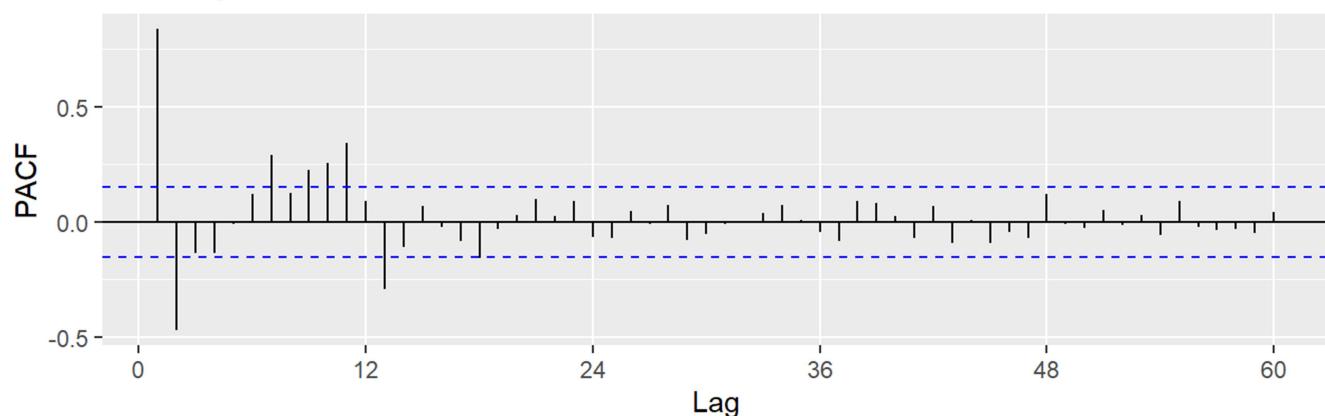


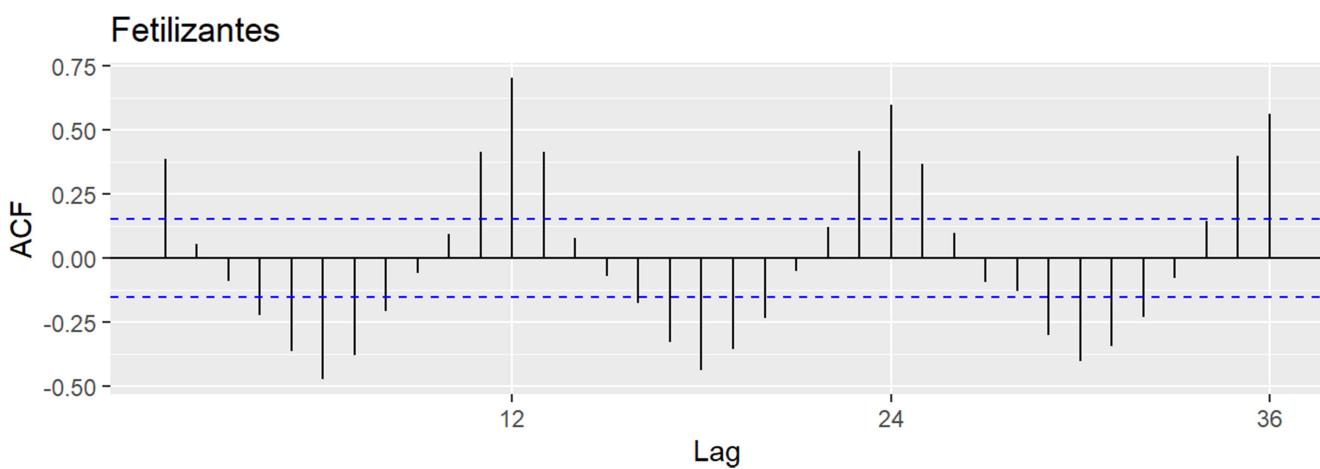
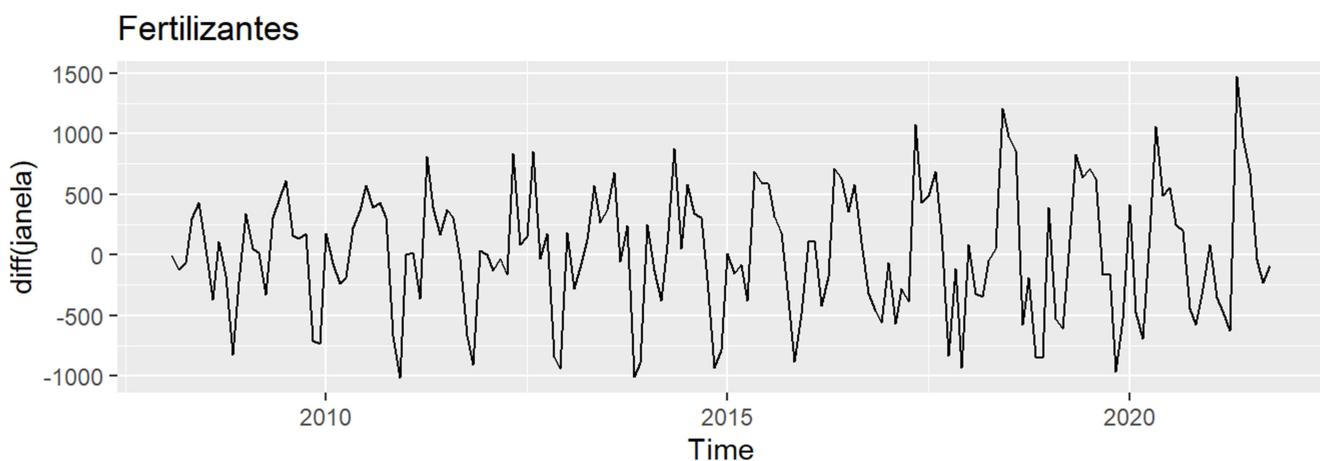
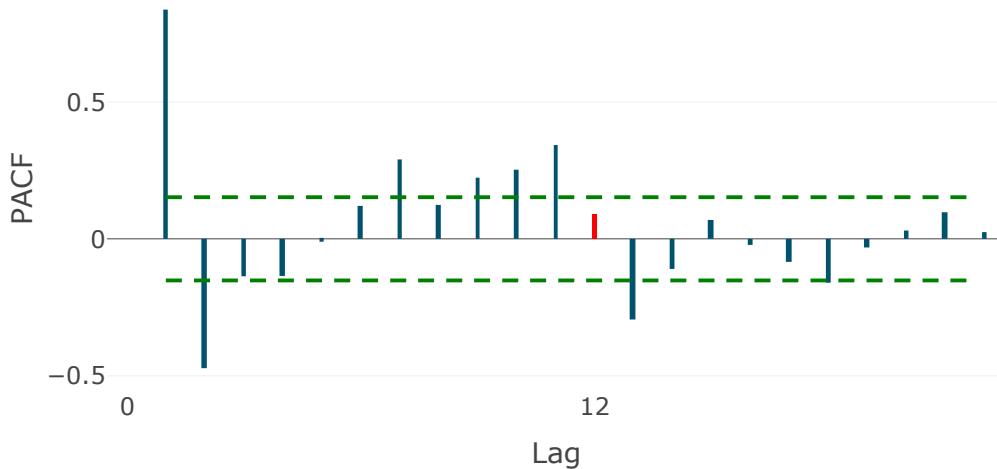
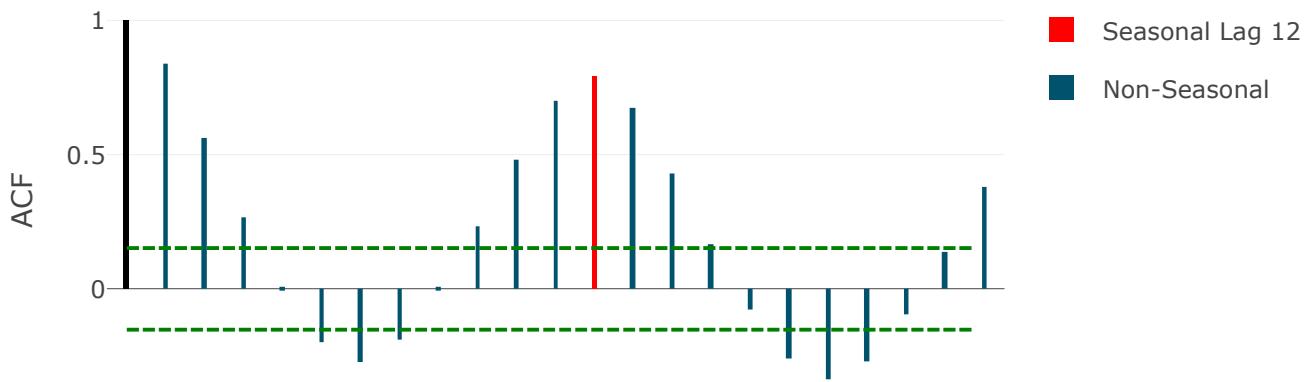
Auto Correlação Serial

Series: janelafertilizantes



Series: janelafertilizantes





Testes de Estacionariedade Regra de Ouro: p-valor baixo rejeita H0

```

## 
## Augmented Dickey-Fuller Test
## 
## data: diff(janelafertilizantes, lag = 12, differences = 1)
## Dickey-Fuller = -5.0565, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary

```

```

## 
## KPSS Test for Level Stationarity
## 
## data: diff(janelafertilizantes, lag = 12, differences = 1)
## KPSS Level = 0.18707, Truncation lag parameter = 4, p-value = 0.1

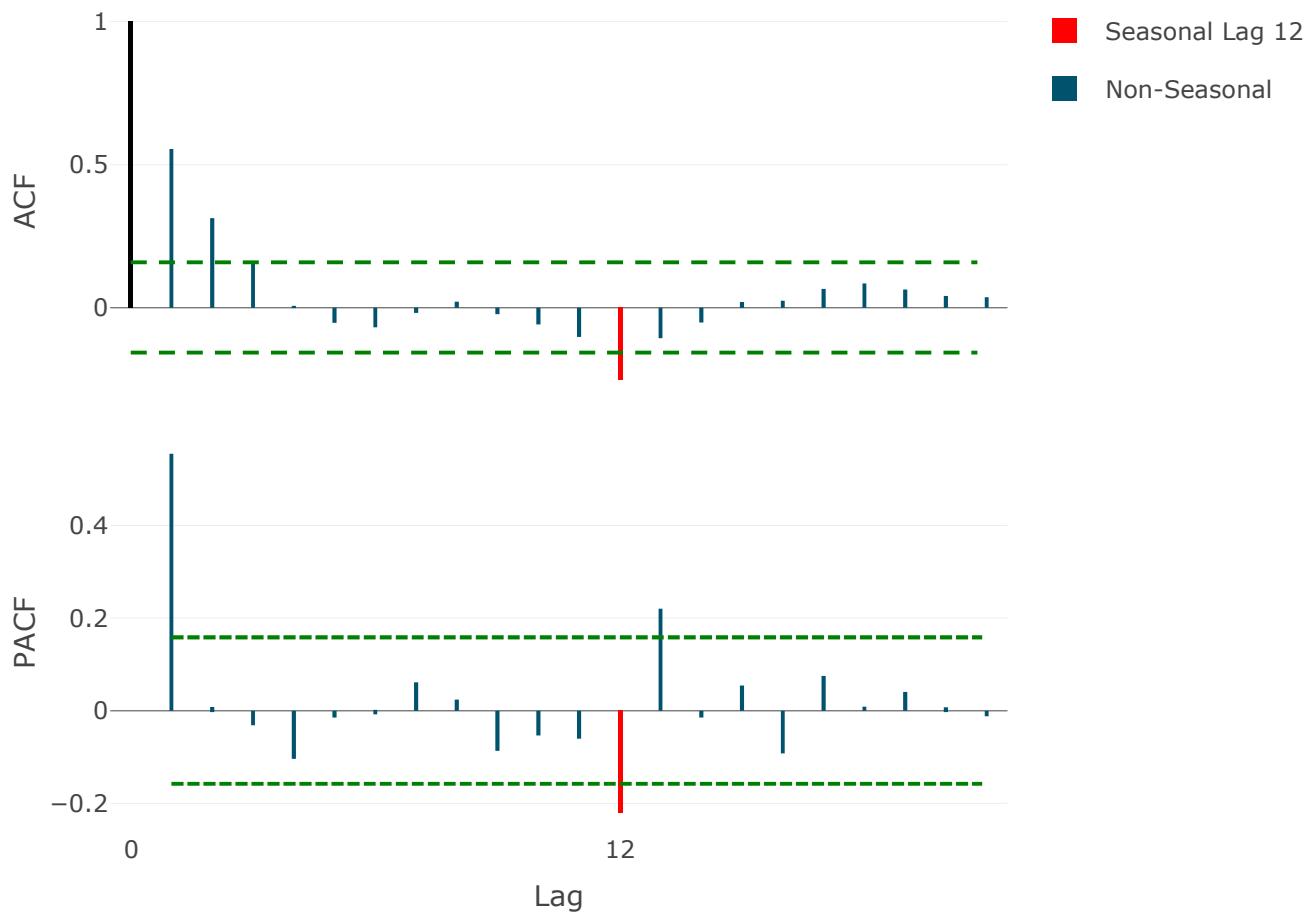
```

```

## 
## Phillips-Perron Unit Root Test
## 
## data: diff(janelafertilizantes, lag = 12, differences = 1)
## Dickey-Fuller Z(alpha) = -67.386, Truncation lag parameter = 4, p-value
## = 0.01
## alternative hypothesis: stationary

```

diff(janelafertilizantes, lag = 12, differences = 1) ACF and PACF Plots



Considerando o grau de confiança de 95%, rejeitamos H_0 devido ao valor de $p\text{-value}$ baixo e aceitamos a hipótese alternativa, ou seja, a série é estacionária.

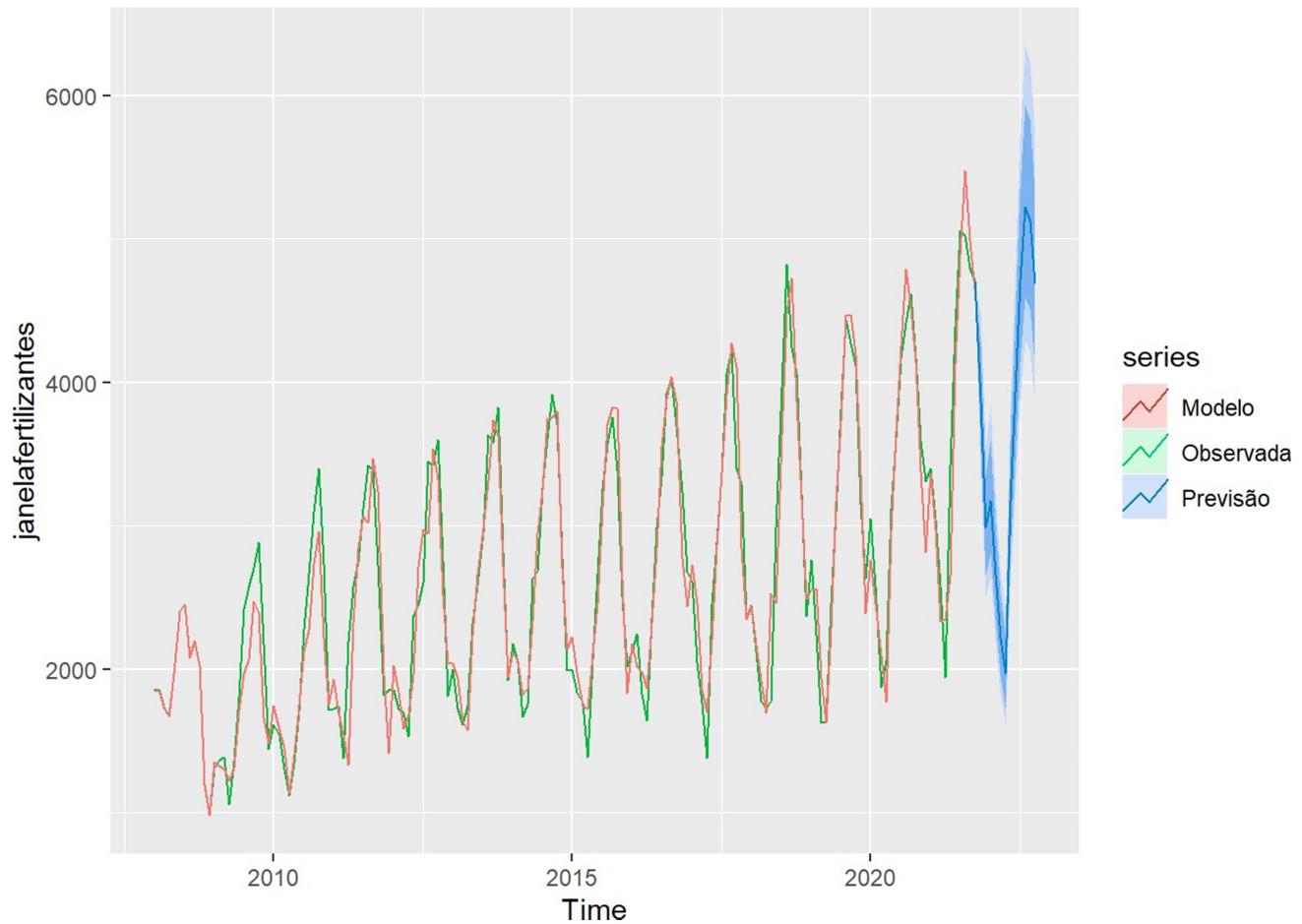
A parte PACF conversa com o modelo Auto Regressivo (AR), e a parte ACF conversa com a parte linear.

Parte Não Sazonal - O PACF fala com a parte auto regressiva AR(p) = 1 - O ACF fala com a parte MA(q) = 2 - A defasagem (d) = 1 — (p,d,q) = (1,1,2)

Parte Sazonal: - P = 1 (ver PCAF linha vertical vermelha) - Q = 3 (ver ACF linha vertical Vermelha) - D = 1 — (P,D,Q) = (1,1,3)

Dessa forma, o modelo SARIMA fica (1,1,2) (1,1,3) [12]

```
## Series: janelafertilizantes
## ARIMA(1,1,2)(1,1,3)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##             ar1      ma1      ma2      sar1      sma1      sma2      sma3
##             0.6439   -1.006   0.0060   -0.7543   0.0882   -0.7790   0.1327
## s.e.     0.1048    0.140   0.1351    0.1008   0.1943    0.1581   0.1196
## 
## sigma^2 = 0.01254: log likelihood = 107.44
## AIC=-198.88  AICc=-197.88  BIC=-174.64
```



Comparando os dois modelos (Auto Arima e SARIMA) com o método AKAIKE

```
## [1] 2180.374
```

```
## [1] -198.8825
```

Comparação do Modelo Sarima com os dados fornecidos pelo Professor (dados reais observados após o período do dataset)

```
##           Point Forecast    Lo 60     Hi 60     Lo 80     Hi 80
## Nov 2021      3858.464 3501.765 4251.498 3328.635 4472.627
## Dec 2021      2985.411 2659.637 3351.088 2503.753 3559.727
## Jan 2022      3175.753 2809.211 3590.121 2634.773 3827.809
## Feb 2022      2659.130 2345.033 3015.297 2195.904 3220.072
## Mar 2022      2220.220 1955.277 2521.063 1829.619 2694.210
## Apr 2022      1967.169 1731.299 2235.174 1619.486 2389.495
## May 2022      3169.402 2788.480 3602.360 2607.951 3851.725
## Jun 2022      3927.804 3455.142 4465.126 3231.164 4774.640
## Jul 2022      4709.848 4142.673 5354.675 3873.928 5726.144
## Aug 2022      5220.919 4591.914 5936.086 4293.886 6348.094
## Sep 2022      5128.377 4510.315 5831.135 4217.482 6236.008
## Oct 2022      4688.148 4122.975 5330.793 3855.210 5701.046
```