

**ANÁLISE ESTATÍSTICA DE CONSUMO DE ENERGIA ELÉTRICA EM MODELO DE SÉRIES
TEMPORAIS PARA O ESTADO DE SÃO PAULO**

Florianópolis, 2025

1. Grupo (7):

- **ALEC OLIVEIRA COELHO, 25200564**
- **CECÍLIA ZICA CAMARGO, 24208346**
- **JOÃO PAULO DECKER OLEINIK, 24202528**
- **NICOLE CRISTINA DA SILVA GOMES, 24105041**

2. Resumo

Este estudo tem como objetivo analisar e prever o consumo de energia elétrica no Estado de São Paulo para os próximos 12 meses, segregando a análise entre os setores Residencial e Industrial. Utilizando dados históricos a partir de 2015, investigou-se a influência de variáveis exógenas — Temperatura Média, Carga do Sistema (atividade econômica) e Bandeiras Tarifárias — sobre a demanda. A metodologia consistiu na seleção de variáveis via regressão linear com algoritmo *Stepwise Backward*, seguida pela competição de modelos de séries temporais (ARIMA, SARIMA e Holt-Winters). Os resultados demonstraram dinâmicas opostas: o setor Residencial mostrou-se inelástico a preço e fortemente sazonal (dependente do clima), sendo melhor modelado pelo SARIMA. O setor Industrial apresentou sensibilidade às tarifas e à atividade econômica, com melhor aderência ao modelo Holt-Winters. A abordagem setorial permitiu reduzir significativamente o erro de previsão (MAPE) em comparação a modelos genéricos.

3. Introdução e objetivos

O planejamento energético é um pilar fundamental para a estabilidade econômica e o bem-estar social. No Estado de São Paulo, o maior centro consumidor do país, a demanda por energia elétrica sofre influência de múltiplos fatores, desde oscilações climáticas extremas até variações na atividade industrial e políticas tarifárias. No entanto, tratar o consumo agregado do estado como uma massa única pode mascarar comportamentos distintos entre diferentes classes de consumidores.

O problema central deste trabalho reside na hipótese de que residências e indústrias reagem de formas divergentes aos mesmos estímulos (preço e temperatura), exigindo modelagens

específicas para garantir previsões assertivas. O objetivo principal deste estudo é desenvolver modelos estatísticos de predição para o consumo de energia elétrica (MWh) das classes **Residencial e Industrial** em São Paulo para um horizonte de 12 meses. Como objetivos específicos, busca-se:

1. Identificar estatisticamente quais variáveis (Clima, Economia ou Tarifa) impactam cada setor;
2. Comparar o desempenho de diferentes algoritmos de séries temporais (ARIMA vs. Suavização Exponencial);
3. Validar a precisão dos modelos através de métricas de erro e diagnóstico de resíduos.

4. Materiais e métodos

A análise foi conduzida utilizando dados mensais de consumo de energia elétrica (MWh), temperatura média, carga do sistema elétrico e bandeiras tarifárias vigentes, compreendendo o período de janeiro de 2015 a dezembro de 2023. Os dados de consumo de energia elétrica foram extraídos do repositório Base dos Dados, compilados a partir de registros oficiais de distribuição, e validados com as Séries Históricas do **Balanco Energético Nacional (BEN)**, disponibilizadas pela **Empresa de Pesquisa Energética (EPE)**.

A principal ferramenta para processamento e modelagem estatística foi a linguagem **R**, executada no ambiente **RStudio**. Como suporte ao desenvolvimento dos scripts, depuração de erros de sintaxe e otimização das rotinas de visualização de dados, foram realizadas consultas à ferramenta de inteligência artificial **ChatGPT - Wizard** (OpenAI).

O fluxo metodológico foi estruturado nas seguintes etapas:

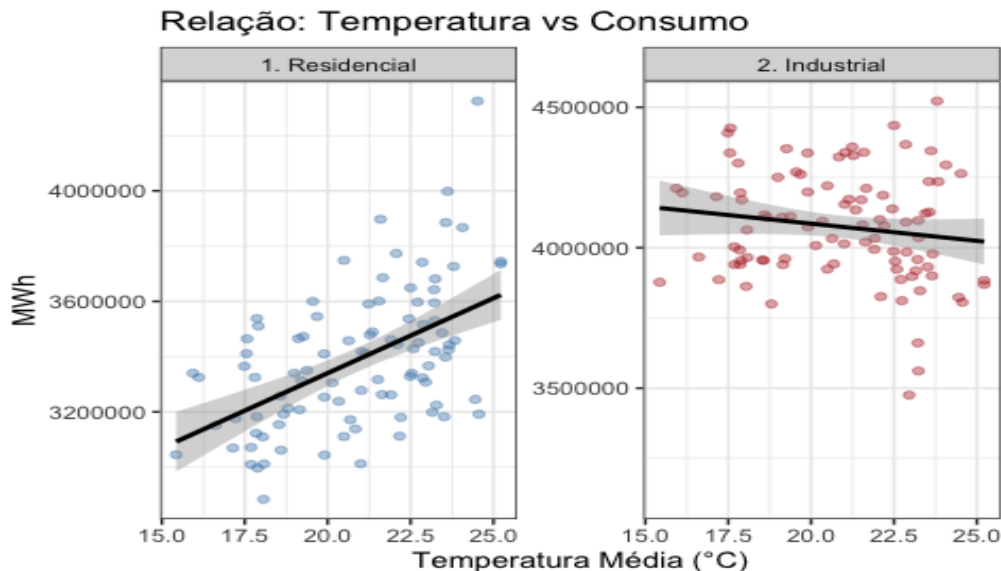
1. **Pré-processamento e Análise Estrutural:** Decomposição STL para isolar tendência e sazonalidade, e teste ADF para verificar a estacionariedade.
2. **Seleção de Variáveis:** Utilização de Regressão Linear com algoritmo *Stepwise Backward* para identificar quais fatores (Clima, Economia ou Tarifa) são estatisticamente relevantes para cada setor.
3. **Modelagem e Validação:** Comparação competitiva entre modelos ARIMA, SARIMA e Holt-Winters, com a escolha do melhor modelo baseado na minimização do erro (MAPE) em uma base de teste (últimos 12 meses).
4. **Diagnóstico:** Validação estatística dos resíduos via teste de Ljung-Box.

5. Resultados e discussões

A estratégia inicial deste estudo visava o desenvolvimento de um modelo unificado para a predição do consumo global de energia elétrica no estado de São Paulo. No entanto, as análises preliminares demonstraram que tratar a demanda agregada como uma massa única comprometia a acurácia das previsões, pois "diluía" os sinais específicos de cada classe de consumidor.

Optou-se, portanto, pela segregação analítica entre os setores Residencial e Industrial. Essa abordagem permitiu identificar que, embora ambos consumam a mesma *commodity* (energia elétrica), seus comportamentos são regidos por leis econômicas e físicas distintas, exigindo modelagens específicas para garantir maior assertividade.

Figura 1: Relação de Temperatura vc Consumo por Setor em São Paulo



A figura acima apresenta a relação entre a temperatura média mensal e consumo de energia elétrica nos setores residencial e industrial no estado de São Paulo.

Setor Residencial:

Apresenta uma relação positiva clara entre a temperatura média e consumo. A medida que a temperatura aumenta de 16 °C para 25 °C, observamos o crescimento do consumo mensal. O ajuste linear evidencia uma inclinação positiva acentuada, sugerindo que a elevação da temperatura está associada ao maior uso de equipamentos de climatização. A dispersão dos pontos é esperada para séries temporais longas, mas a tendência geral permanece consistente ao longo dos anos.

Concluimos que, em regiões urbanizadas e de clima quente, os domicílios tornam-se mais sensíveis às variações térmicas, elevando a demanda elétrica principalmente nos meses mais quentes. Como o setor residencial já apresentava tendência crescente (figura 1), a temperatura funciona como **determinante adicional**, reforçando a sazonalidade do consumo.

Setor Industrial:

No setor industrial observa-se uma tendência oposta. O ajuste linear indica inclinação levemente negativa, sugerindo que temperaturas mais altas estão associadas a uma redução marginal do consumo industrial. Embora a variabilidade seja maior — reflexo da sensibilidade da indústria a fatores como ciclos econômicos, produção, exportações e choques externos — ainda assim a tendência linear aponta que o consumo industrial não aumenta com o calor, diferentemente do setor residencial.

Existem explicações plausíveis para esse comportamento:

1. Processos industriais são menos sensíveis ao clima, pois operam com controle térmico interno.
2. A indústria tende a **reduzir o ritmo** em períodos de calor extremo, seja por questões operacionais (riscos térmicos para trabalhadores), seja por sazonalidade de determinados setores produtivos.
3. Parte significativa do declínio observado em temperaturas maiores pode refletir **sazonalidade econômica** e não apenas térmica — mas o padrão visual sugere que a temperatura não exerce o mesmo papel estimulante que no setor residencial.

O residencial tem o consumo motivado por conforto térmico e hábitos domiciliares, e o industrial tem o consumo determinado pelo ritmo produtivo, ou seja, tem dependência muito maior de variáveis econômicas e operacionais.

5.1. Determinantes do Consumo: O que move cada setor?

Inicialmente, investigou-se a estrutura das séries temporais por meio do teste de estacionariedade, etapa fundamental para definir a necessidade de diferenciação nos modelos preditivos.

Na sequência, aplicou-se o método de seleção de variáveis *Stepwise Backward* para filtrar quais fatores exógenos (Temperatura, Economia/Carga e Tarifas) possuem relevância estatística para explicar o consumo. Para o **Setor Residencial**, o modelo descartou as variáveis de preço (Bandeiras Tarifárias), mantendo apenas a Carga e a Temperatura. Já para o **Setor Industrial**, as Bandeiras Tarifárias (especialmente a de Escassez Hídrica) mostraram-se estatisticamente significativas ($p\text{-valor} < 0.05$).

Tabela 1: Resumo dos Coeficientes Seleccionados (Output do R):

```
## [1] ">>> Variáveis Seleccionadas para RESIDENCIAL:"
```

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	-447500.6270	2.261026e+05	-1.979193	5.078391e-02
##	carga_mwmed	106.4052	7.760296e+00	13.711492	5.857927e-24
##	temp_media_mes	-16977.9334	7.406685e+03	-2.292245	2.417196e-02

```
## [1] ">>> Variáveis Seleccionadas para INDUSTRIAL:"
```

##		Estimate	Std. Error	t value
##	(Intercept)	2513506.71250	320059.58434	7.8532462
##	carga_mwmed	71.36998	11.17402	6.3871348
##	temp_media_mes	-61141.45809	9836.23522	-6.2159410
##	nom_bandeira_acionadaEscassez Hídrica	161648.57190	71526.86500	2.2599700
##	nom_bandeira_acionadaVerde	-12858.87016	50063.56273	-0.2568509
##	nom_bandeira_acionadaVermelha P1	68762.81009	54878.96752	1.2529902
##	nom_bandeira_acionadaVermelha P2	46544.96603	63788.85583	0.7296724
##		Pr(> t)		
##	(Intercept)	9.230284e-12		

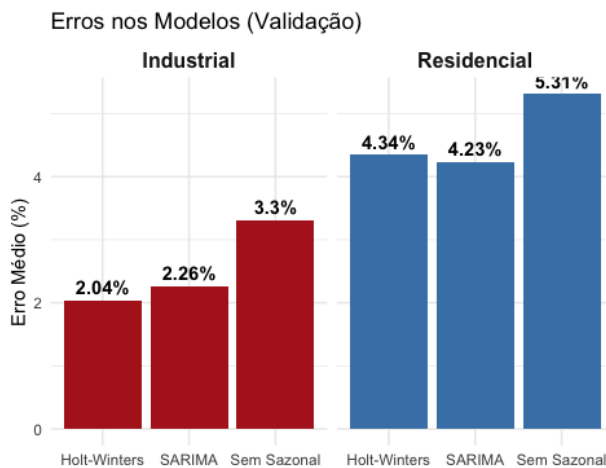
## carga_mwmed	7.759912e-09
## temp_media_mes	1.660691e-08
## nom_bandeira_acionadaEscassez H?drlica	2.629145e-02
## nom_bandeira_acionadaVerde	7.978935e-01
## nom_bandeira_acionadaVermelha P1	2.135285e-01
## nom_bandeira_acionadaVermelha P2	4.675283e-01

A divergência nos fatores determinantes confirma a hipótese de inelasticidade do consumidor residencial, que prioriza o conforto térmico sobre o custo. Em contrapartida, a indústria demonstra racionalidade econômica, ajustando sua demanda diante de tarifas punitivas (Bandeira de Escassez).

5.2. Seleção de Modelos: Qual foi mais preciso

Foram testados três algoritmos preditivos (ARIMA, SARIMA e Holt-Winters) em uma base de validação (últimos 12 meses conhecidos). A métrica de escolha foi o MAPE (Erro Percentual Absoluto Médio).

Figura 2: Comparativo de Erros (MAPE) por Modelo e Setor



Observa-se na **Figura 2** que:

Residencial: O modelo **SARIMA** venceu com um erro de **4,22%**, superando significativamente o modelo sem sazonalidade (5,31%).

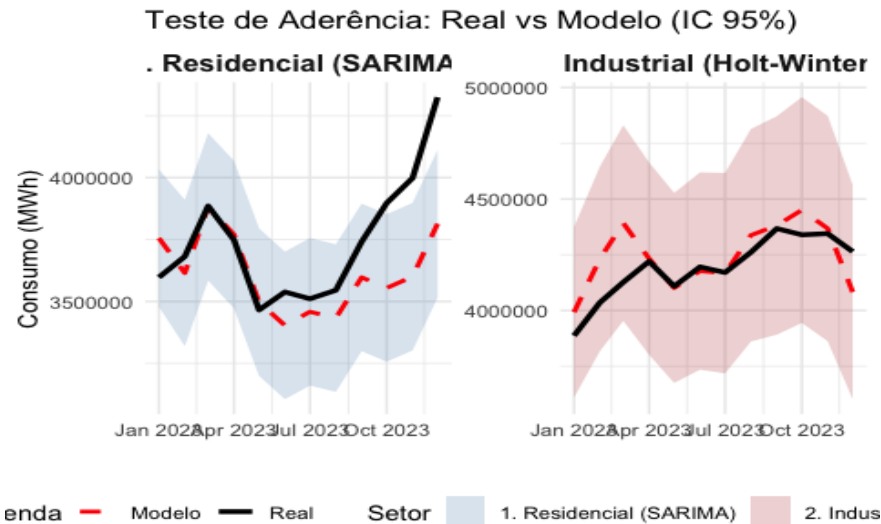
Industrial: O modelo **Holt-Winters** obteve a melhor performance, com um erro de apenas **2,03%**, mas ainda muito próximo do Sarima.

Nota-se que o erro do setor industrial é praticamente a metade do erro residencial. Isso ocorre porque a demanda industrial é menos volátil a picos de calor aleatórios, seguindo uma tendência econômica mais estável.

5.3. Validação: Teste de Aderência

Para garantir que os modelos não apenas minimizam o erro médio, mas também capturam a dinâmica da série, plotou-se a previsão contra os dados reais na base de teste, incluindo o intervalo de confiança de 95%.

Figura 3: Aderência dos Modelos (Real vs. Previsto) com Intervalo de Confiança:



A Figura 3 demonstra a robustez dos modelos selecionados:

No painel **Industrial (Vermelho)**, a linha da realidade (preta) mantém-se consistentemente dentro do intervalo de confiança, indicando alta previsibilidade. Já no painel **Residencial (Azul)**, o modelo capturou corretamente a tendência de alta e a sazonalidade, embora eventos extremos de temperatura no final do período tenham testado os limites superiores do intervalo de confiança.

5.4. Previsão de Cenários Futuros

Com os modelos validados (SARIMA para Residencial e Holt-Winters para Industrial), projetou-se o consumo para o horizonte de 12 meses.

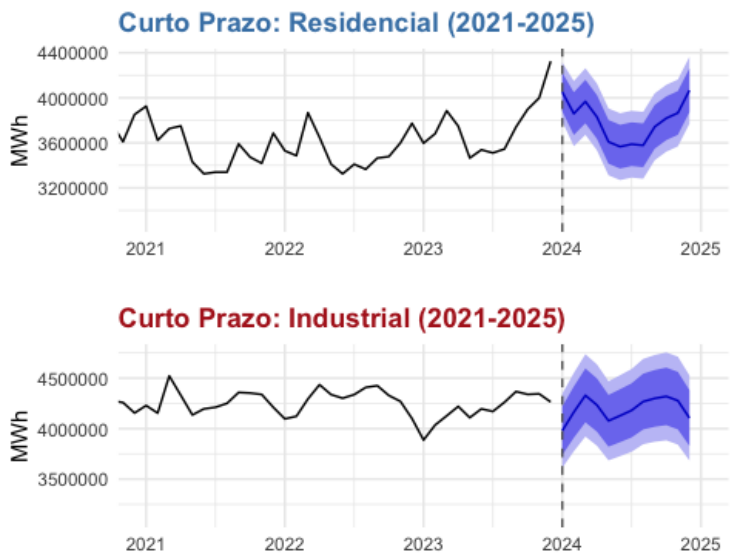


Figura 4: Projeção de 1 ano

A projeção de curto prazo indica a manutenção dos padrões sazonais:

Residencial: Previsão de pico acentuado nos meses de verão (Janeiro/Fevereiro), exigindo atenção do sistema elétrico para garantir o suprimento em horários de ponta. **Industrial:** Previsão de recuperação cíclica, com menor amplitude de variação mensal, refletindo a estabilidade da carga econômica projetada.

Para uma visão estrutural, a Figura 4 apresenta a tendência de longo prazo, contextualizando a previsão dentro da série histórica iniciada em 2015.

Figura 5: Panorama Histórico e Tendência (2015-2025):



A análise visual de longo prazo confirma que o setor residencial apresenta uma tendência de crescimento vegetativo constante, impulsionada pela eletrificação das famílias, enquanto o setor industrial exhibe cicatrizes de crises econômicas passadas, mas com viés de recuperação no horizonte projetado.

6. Considerações finais

O presente estudo cumpriu seu objetivo principal de desenvolver modelos estatísticos de predição de demanda de energia elétrica para o Estado de São Paulo, confirmando a tese inicial de que tratar o consumo agregado mascara dinâmicas fundamentais. A segregação entre as classes Residencial e Industrial provou-se mandatória para a obtenção de previsões assertivas.

A aplicação da metodologia estatística permitiu chegar a algumas conclusões centrais:

Dicotomia de Comportamento: Ficou estatisticamente comprovado que os setores operam sob lógicas opostas. O setor Residencial é regido pelo **conforto térmico** e apresenta demanda inelástica ao preço (não reduz consumo significativamente com tarifas altas). Em contrapartida, o setor Industrial é regido pela **racionalidade econômica**, demonstrando sensibilidade estatística às Bandeiras Tarifárias e ajustando sua produção (e consumo) conforme o custo da energia.

Desempenho dos Modelos: A competição de modelos validou que não existe um "algoritmo universal". A forte sazonalidade climática do setor residencial exigiu a robustez do modelo **SARIMA** (MAPE 4,22%), enquanto a estabilidade cíclica da indústria foi melhor capturada pela suavização exponencial do **Holt-Winters** (MAPE 2,03%).

Estabilidade vs. Volatilidade: O erro de previsão do setor industrial foi substancialmente menor que o do residencial. Conclui-se que é mais fácil prever o comportamento planejado de fábricas do que a reação de milhões de famílias.

Limitações e Trabalhos Futuros: Uma limitação observada foi a dificuldade do modelo residencial em antecipar picos de calor extremos que fogem do padrão histórico (como os observados sob efeito do fenômeno *El Niño* no final de 2023).

7. Referências bibliográficas

- [1] Rstudio, Posit. [Online]. Disponível: <https://posit.co/download/rstudio-desktop/>
- [2] R, Posit. [Online]. Disponível: <https://www.r-project.org/>
- [3] BASE DOS DADOS [Online]. *Consumo de Energia Elétrica no Brasil*. Disponível em: <https://basedosdados.org/dataset/782c5607-9f69-4e12-b0d5-aa0f1a7a94e2?table=28d16282-d100-4ea8-9dde-36c05c8f1ca2>.
- [4] EPE - EMPRESA DE PESQUISA ENERGÉTICA. [Online]. *Balanco Energético Nacional (BEN): Séries Históricas Completas*. Disponível em: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/BEN-Series-Historicas-Completas>.

8. Sintaxe R

```
library(tidyverse)

## Warning: package 'ggplot2' was built under R version 4.5.2

## — Attaching core tidyverse packages —————
tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats   1.0.1      ✓ stringr    1.5.1
## ✓ ggplot2    4.0.1      ✓ tibble     3.3.0
## ✓ lubridate  1.9.4      ✓ tidyr      1.3.1
## ✓ purrr      1.1.0
## — Conflicts —————
tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

library(forecast)

## Registered S3 method overwritten by 'quantmod':
##   method              from
## as.zoo.data.frame zoo

library(lubridate)
library(ggplot2)
library(gridExtra)

##
## Attaching package: 'gridExtra'
##
```

```

## The following object is masked from 'package:dplyr':
##
##      combine

library(tseries)

dados <- readRDS("results/base_sp_analitica.rds") %>%
  filter(ano_mes >= "2015-01-01")

# Criando Séries Temporais
ts_res <- ts(dados$consumo_residencial, start = c(2015, 1), frequency = 12)
ts_ind <- ts(dados$consumo_industrial, start = c(2015, 1), frequency = 12)

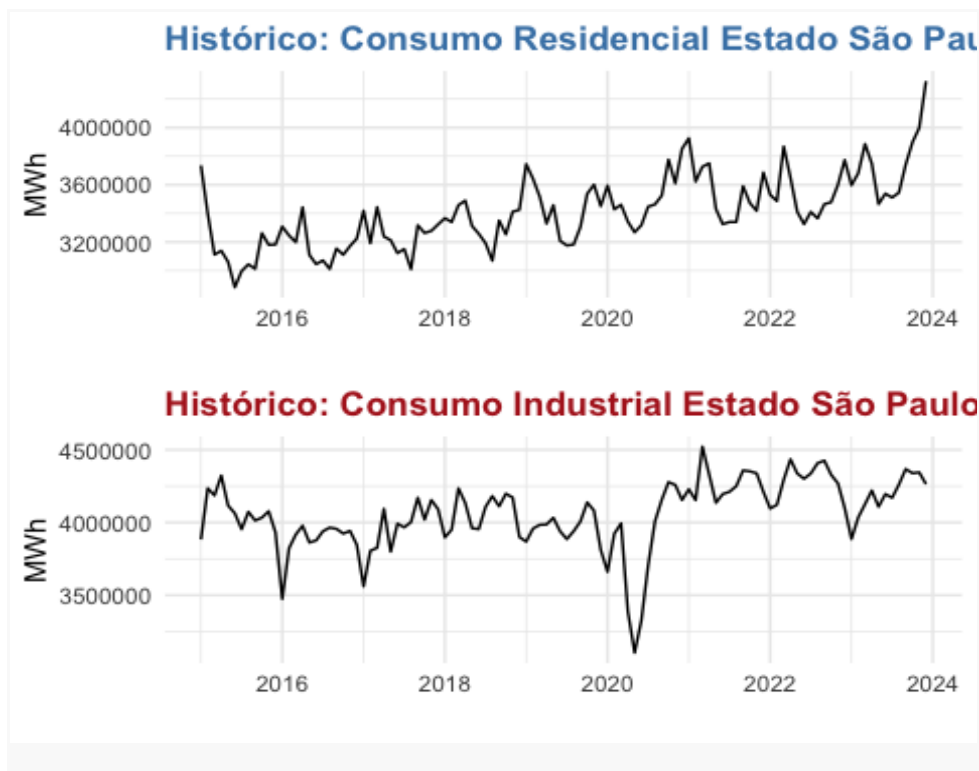
# Divisão Treino vs Teste
tam <- length(ts_res)
treino_res <- subset(ts_res, end = tam - 12); teste_res <- subset(ts_res, start
= tam - 11)
treino_ind <- subset(ts_ind, end = tam - 12); teste_ind <- subset(ts_ind, start
= tam - 11)

g_hist_res <- autoplot(ts_res) +
  labs(title = "Histórico: Consumo Residencial Estado São Paulo", y="MWh",
x="") +
  theme_minimal() + theme(plot.title = element_text(color="steelblue",
face="bold"))

g_hist_ind <- autoplot(ts_ind) +
  labs(title = "Histórico: Consumo Industrial Estado São Paulo", y="MWh", x="")
+
  theme_minimal() + theme(plot.title = element_text(color="firebrick",
face="bold"))

grid.arrange(g_hist_res, g_hist_ind, ncol = 1)

```



```
# Objetivo: "ver" correlações e estrutura antes de calcular.
# RELAÇÃO COM A TEMPERATURA
df_visual <- dados %>%
  select(ano_mes, temp_media_mes, consumo_residencial, consumo_industrial,
nom_bandeira_acionada) %>%
  pivot_longer(cols = c(consumo_residencial, consumo_industrial),
names_to = "Setor", values_to = "Consumo") %>%
  mutate(Setor = ifelse(Setor == "consumo_residencial", "1. Residencial", "2.
Industrial"))

g_temp <- ggplot(df_visual, aes(x = temp_media_mes, y = Consumo, color =
Setor)) +
  geom_point(alpha = 0.4) +
  geom_smooth(method = "lm", color = "black") +
  facet_wrap(~Setor, scales = "free_y") +
  scale_color_manual(values = c("steelblue", "firebrick")) +
  labs(title = "Relação: Temperatura vs Consumo",
x = "Temperatura Média (°C)", y = "MWh") +
  theme_bw() + theme(legend.position = "none")

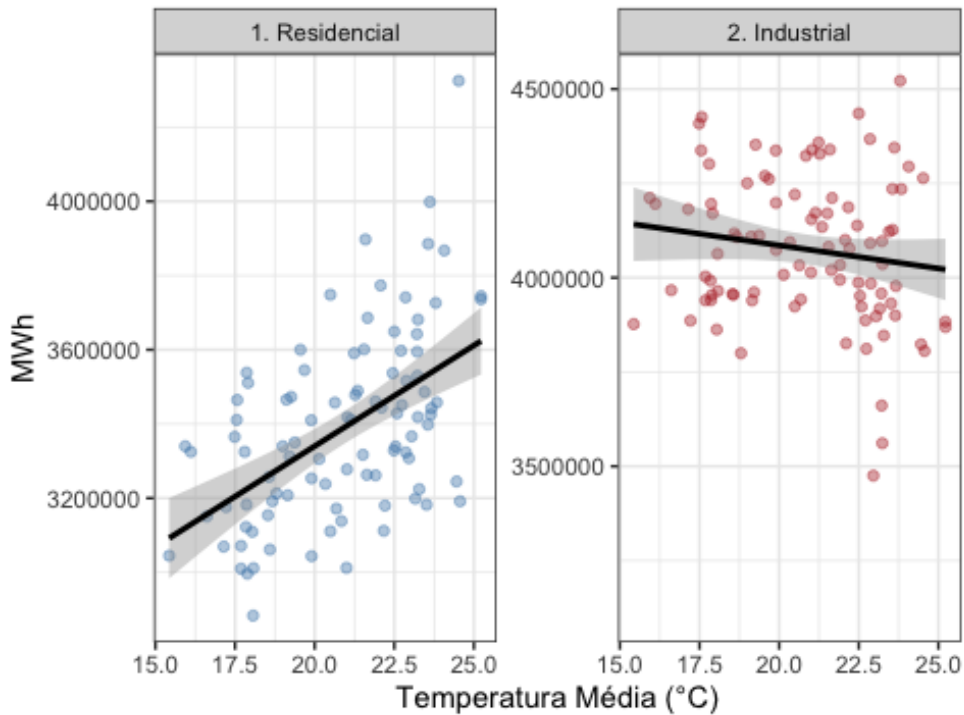
print(g_temp)

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

## Warning: Removed 26 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 26 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```

Relação: Temperatura vs Consumo



```
# DECOMPOSIÇÃO E TESTE DE ESTACIONARIEDADE
```

```
# DECOMPOSIÇÃO STL (Visualizar Tendência vs Sazonalidade)
```

```
# Residencial
```

```
decomp_res <- decompose(ts_res, type = "multiplicative")
```

```
g_dec_res <- autoplot(decomp_res) +
```

```
  labs(title = "Decomposição: Residencial") + theme_minimal()
```

```
# Industrial
```

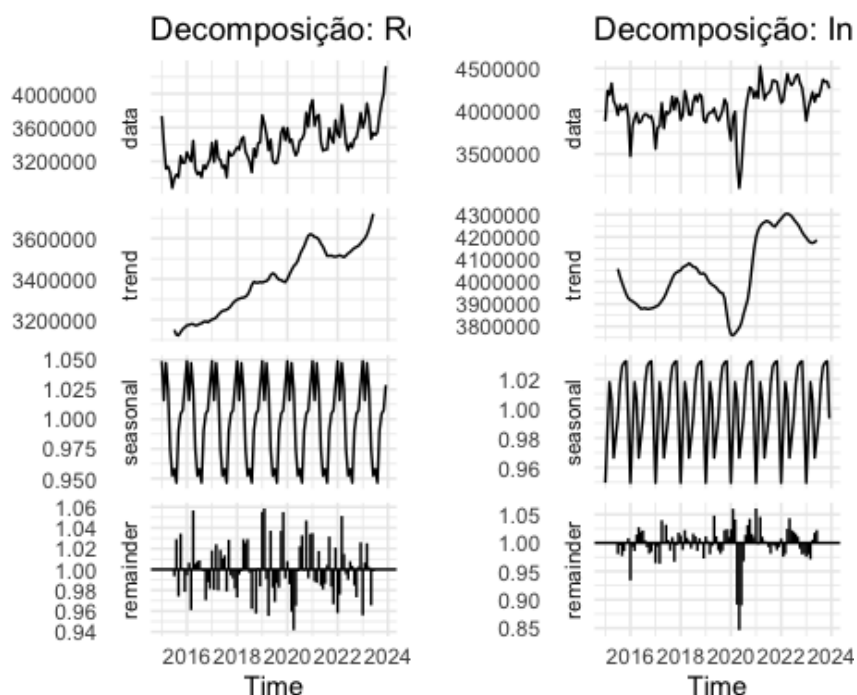
```
decomp_ind <- decompose(ts_ind, type = "multiplicative")
```

```
g_dec_ind <- autoplot(decomp_ind) +
```

```
  labs(title = "Decomposição: Industrial") + theme_minimal()
```

```
# Plotando Lado a Lado
```

```
grid.arrange(g_dec_res, g_dec_ind, ncol = 2)
```



2. TESTE ADF (Augmented Dickey-Fuller)

Objetivo: Provar estatisticamente que a série não é estacionária

```
print(">>> TESTE DE ESTACIONARIEDADE (ADF) <<<")
```

```
## [1] ">>> TESTE DE ESTACIONARIEDADE (ADF) <<<"
```

```
print(paste("Residencial p-valor:", round(adf.test(ts_res)$p.value, 4)))
```

```
## Warning in adf.test(ts_res): p-value smaller than printed p-value
```

```
## [1] "Residencial p-valor: 0.01"
```

```
print(paste("Industrial p-valor: ", round(adf.test(ts_ind)$p.value, 4)))
```

```
## [1] "Industrial p-valor: 0.0559"
```

MÉTODO DE SELEÇÃO DE VARIÁVEIS (STEPWISE)

```
# --- SETOR RESIDENCIAL ---
```

```
df_res <- dados %>% filter(!is.na(consumo_residencial), !is.na(temp_media_mes))
```

```
full_res <- lm(consumo_residencial ~ carga_mwmed + temp_media_mes +
```

```
nom_bandeira_acionada, data=df_res)
```

```
step_res <- step(full_res, direction="backward", trace=0)
```

```
print(">>> Variáveis Seleccionadas para RESIDENCIAL:")
```

```
## [1] ">>> Variáveis Seleccionadas para RESIDENCIAL:"
```

```
print(summary(step_res)$coefficients)
```

```
##              Estimate  Std. Error  t value    Pr(>|t|)
## (Intercept) -447500.6270 2.261026e+05 -1.979193 5.078391e-02
## carga_mwmed    106.4052 7.760296e+00 13.711492 5.857927e-24
## temp_media_mes -16977.9334 7.406685e+03 -2.292245 2.417196e-02
```

```

# --- SETOR INDUSTRIAL ---
df_ind <- dados %>% filter(!is.na(consumo_industrial), !is.na(temp_media_mes))
full_ind <- lm(consumo_industrial ~ carga_mwmed + temp_media_mes +
nom_bandeira_acionada, data=df_ind)
step_ind <- step(full_ind, direction="backward", trace=0)

print(">>> Variáveis Seleccionadas para INDUSTRIAL:")

## [1] ">>> Variáveis Seleccionadas para INDUSTRIAL:"

print(summary(step_ind)$coefficients)

##              Estimate   Std. Error   t value
## (Intercept)    2513506.71250 320059.58434  7.8532462
## carga_mwmed      71.36998    11.17402   6.3871348
## temp_media_mes  -61141.45809   9836.23522  -6.2159410
## nom_bandeira_acionadaEscassez H?drlica 161648.57190  71526.86500  2.2599700
## nom_bandeira_acionadaVerde    -12858.87016   50063.56273  -0.2568509
## nom_bandeira_acionadaVermelha P1    68762.81009   54878.96752  1.2529902
## nom_bandeira_acionadaVermelha P2    46544.96603   63788.85583  0.7296724
##              Pr(>|t|)
## (Intercept)    9.230284e-12
## carga_mwmed    7.759912e-09
## temp_media_mes 1.660691e-08
## nom_bandeira_acionadaEscassez H?drlica 2.629145e-02
## nom_bandeira_acionadaVerde    7.978935e-01
## nom_bandeira_acionadaVermelha P1    2.135285e-01
## nom_bandeira_acionadaVermelha P2    4.675283e-01

# MÉTODO DE SELEÇÃO DE MODELO (MAPE)

get_mape <- function(model, test_data) {
  mean(abs((test_data - forecast(model, h=12)$mean)/test_data)) * 100
}

# TREINANDO CANDIDATOS RESIDENCIAL
m_res_sarima <- auto.arima(treino_res, seasonal=TRUE) # Sazonal
m_res_arima <- auto.arima(treino_res, seasonal=FALSE) # Sem Sazonal
m_res_hw <- hw(treino_res, seasonal="multiplicative") # Holt-Winters

# TREINANDO CANDIDATOS INDUSTRIAL
m_ind_sarima <- auto.arima(treino_ind, seasonal=TRUE)
m_ind_arima <- auto.arima(treino_ind, seasonal=FALSE)
m_ind_hw <- hw(treino_ind, seasonal="multiplicative")

# Tabela de Resultados
tabela_erros <- data.frame(
  Modelo = c("SARIMA (Sazonal)", "ARIMA (Sem Sazonal)", "Holt-Winters"),
  Erro_Residencial = c(get_mape(m_res_sarima, teste_res),
    get_mape(m_res_arima, teste_res),
    get_mape(m_res_hw, teste_res)),
  Erro_Industrial = c(get_mape(m_ind_sarima, teste_ind),
    get_mape(m_ind_arima, teste_ind),
    get_mape(m_ind_hw, teste_ind))
)

```

```

print(">>> Tabela de Erros (MAPE):")
## [1] ">>> Tabela de Erros (MAPE):"
print(tabela_erros)

##           Modelo Erro_Residencial Erro_Industrial
## 1    SARIMA (Sazonal)         4.225781         2.263560
## 2  ARIMA (Sem Sazonal)         5.310543         3.304000
## 3      Holt-Winters           4.344330         2.036283

# VISUALIZAÇÃO DOS ERROS (BARRAS)

df_mape <- data.frame(
  Setor = c(rep("Residencial", 3), rep("Industrial", 3)),
  Modelo = c("SARIMA", "Sem Sazonal", "Holt-Winters",
             "SARIMA", "Sem Sazonal", "Holt-Winters"),
  MAPE = c(tabela_erros$Erro_Residencial, tabela_erros$Erro_Industrial)
)

g_erro <- ggplot(df_mape, aes(x = reorder(Modelo, MAPE), y = MAPE, fill =
Setor)) +
  geom_col(position = "dodge") +
  geom_text(aes(label = paste0(round(MAPE, 2), "%")),
            position = position_dodge(width = 0.9), vjust = -0.5, fontface =
"bold") +
  facet_wrap(~Setor, scales = "free_x") +
  scale_fill_manual(values = c("firebrick", "steelblue")) +
  labs(title = "Erros nos Modelos (Validação)",
       y = "Erro Médio (%)", x = "") +
  theme_minimal() +
  theme(legend.position = "none",
        strip.text = element_text(size = 12, face="bold"))

print(g_erro)

```

Erros nos Modelos (Validação)

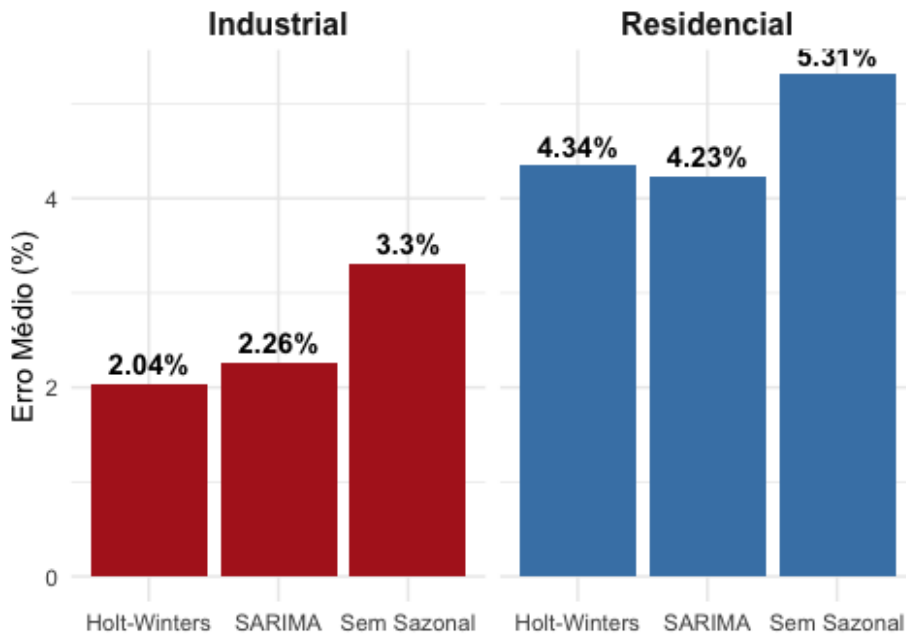


GRÁFICO DE VALIDAÇÃO (TESTE vs REAL + IC)

```
mod_camp_res <- auto.arima(treino_res, seasonal = TRUE)
prev_teste_res <- forecast(mod_camp_res, h = 12, level = 95)
prev_teste_ind <- hw(treino_ind, seasonal = "multiplicative", h = 12, level = 95)

datas_teste <- seq(as.Date("2015-01-01"), by = "month", length.out =
length(ts_res)) %>%
  tail(12)

montar_df <- function(prev_obj, real_data, nome_setor) {
  data.frame(
    Data = datas_teste,
    Setor = nome_setor,
    Real = as.numeric(real_data),
    Previsto = as.numeric(prev_obj$mean),
    Lower = as.numeric(prev_obj$lower),
    Upper = as.numeric(prev_obj$upper)
  )
}

df_validacao <- bind_rows(
  montar_df(prev_teste_res, teste_res, "1. Residencial (SARIMA)"),
  montar_df(prev_teste_ind, teste_ind, "2. Industrial (Holt-Winters)")
)

g_validacao <- ggplot(df_validacao, aes(x = Data)) +
  geom_ribbon(aes(ymin = Lower, ymax = Upper, fill = Setor), alpha = 0.2) +
  geom_line(aes(y = Previsto, color = "Modelo"), linetype = "dashed", size = 1)
+
  geom_line(aes(y = Real, color = "Real"), size = 1.2) +
  facet_wrap(~Setor, scales = "free_y") +
```

```

scale_color_manual(values = c("Modelo" = "red", "Real" = "black")) +
scale_fill_manual(values = c("steelblue", "firebrick")) +
labs(title = "Teste de Aderência: Real vs Modelo (IC 95%)",
     y = "Consumo (MWh)", x = "", color = "Legenda", fill = "Setor") +
theme_minimal() +
theme(legend.position = "bottom",
     strip.text = element_text(face="bold", size=12))

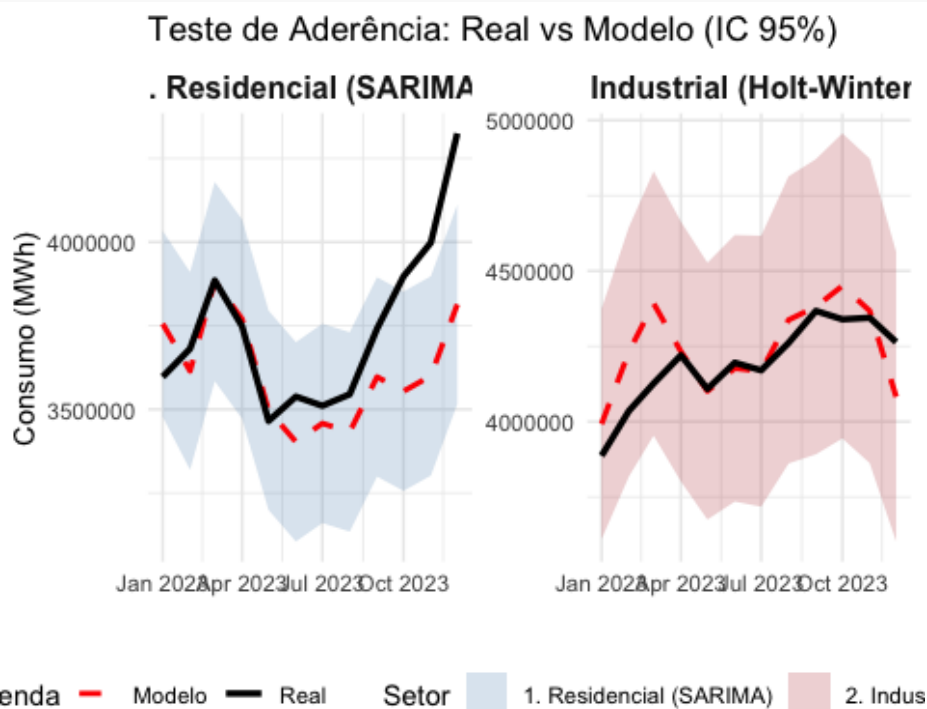
```

```

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

```
print(g_validacao)
```



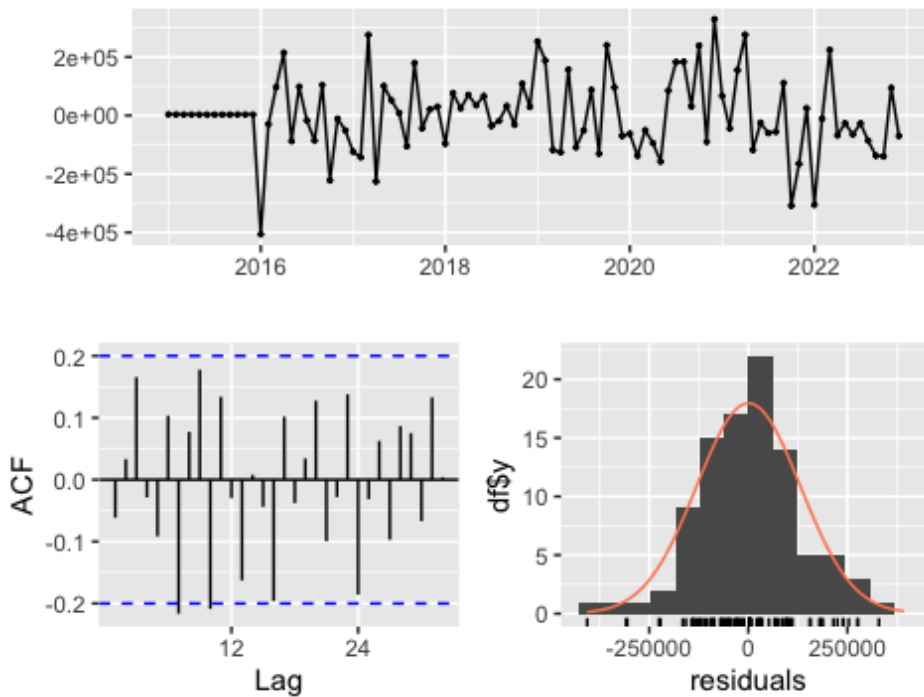
```
# DIAGNÓSTICO DE RESÍDUO
```

```

# Residencial
checkresiduals(mod_camp_res)

```

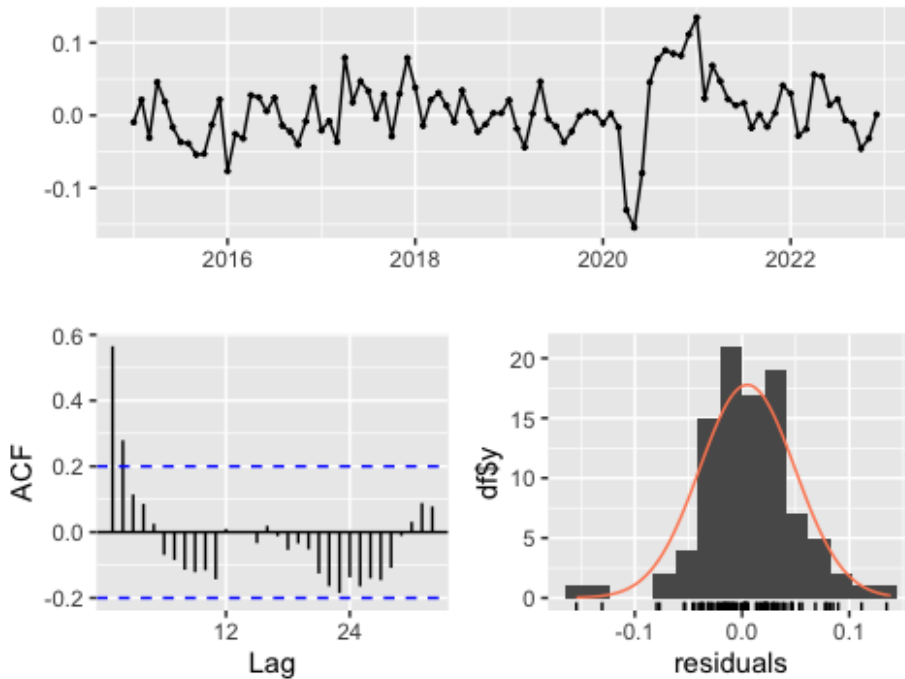
Residuals from ARIMA(1,0,0)(1,1,0)[12] with drift



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0)(1,1,0)[12] with drift
## Q* = 30.506, df = 17, p-value = 0.02292
##
## Model df: 2.   Total lags used: 19

# Industrial
mod_camp_ind <- hw(treino_ind, seasonal = "multiplicative")
checkresiduals(mod_camp_ind)
```

Residuals from Holt-Winters' multiplicative method



```
##
##  Ljung-Box test
##
## data:  Residuals from Holt-Winters' multiplicative method
## Q* = 50.267, df = 19, p-value = 0.0001197
##
## Model df: 0.   Total lags used: 19

#PREVISÃO DO FUTURO

# 1. Recalcular os modelos com a base COMPLETA (Histórico todo)
# Residencial (SARIMA)
modelo_final_res <- auto.arima(ts_res, seasonal = TRUE)
final_res <- forecast(modelo_final_res, h = 12)

# Industrial (Holt-Winters)
final_ind <- hw(ts_ind, seasonal = "multiplicative", h = 12)

# 2. Imprimir os números
print("--- Previsão Residencial (MWh) ---"); print(final_res)

## [1] "--- Previsão Residencial (MWh) ---"

##          Point Forecast   Lo 80   Hi 80   Lo 95   Hi 95
## Jan 2024      4051403 3879999 4222806 3789264 4313541
## Feb 2024      3859246 3669987 4048505 3569799 4148693
## Mar 2024      3967204 3774250 4160159 3672106 4262303
## Apr 2024      3829894 3636138 4023650 3533570 4126218
## May 2024      3608676 3414744 3802607 3312083 3905268
## Jun 2024      3566233 3372263 3760203 3269582 3862884
## Jul 2024      3588066 3394088 3782044 3291402 3884730
## Aug 2024      3578141 3384161 3772121 3281475 3874808
## Sep 2024      3742910 3548929 3936890 3446242 4039577
## Oct 2024      3818613 3624632 4012593 3521946 4115280
```

```

## Nov 2024      3865501 3671522 4059480 3568835 4162166
## Dec 2024      4066516 3872542 4260490 3769858 4363173

print("--- Previsão Industrial (MWh) ---"); print(final_ind)

## [1] "--- Previsão Industrial (MWh) ---"

##      Point Forecast   Lo 80   Hi 80   Lo 95   Hi 95
## Jan 2024      3978800 3738919 4218681 3611934 4345666
## Feb 2024      4163491 3909783 4417199 3775479 4551504
## Mar 2024      4329312 4062721 4595903 3921596 4737027
## Apr 2024      4233376 3969996 4496757 3830570 4636182
## May 2024      4077370 3821116 4333624 3685463 4469276
## Jun 2024      4127953 3865927 4389979 3727219 4528687
## Jul 2024      4180941 3912943 4448940 3771073 4590809
## Aug 2024      4267999 3991775 4544223 3845551 4690448
## Sep 2024      4299703 4018779 4580627 3870067 4729339
## Oct 2024      4319480 4034620 4604339 3883825 4755134
## Nov 2024      4277309 3992629 4561988 3841929 4712689
## Dec 2024      4105570 3829839 4381301 3683875 4527264

inicio_prev <- min(time(final_res$mean))

# GRÁFICO A: ZOOM NOS ULTIMOS DADOS - Com Intervalo de Confiança
p_zoom_res <- autoplot(final_res) +
  geom_vline(xintercept = inicio_prev, linetype="dashed", color="grey40") +
  labs(title = "Curto Prazo: Residencial (2021-2025)", y="MWh", x="") +
  coord_cartesian(xlim = c(2021, 2025)) +
  theme_minimal() + theme(plot.title = element_text(color="steelblue",
face="bold"))

p_zoom_ind <- autoplot(final_ind) +
  geom_vline(xintercept = inicio_prev, linetype="dashed", color="grey40") +
  labs(title = "Curto Prazo: Industrial (2021-2025)", y="MWh", x="") +
  coord_cartesian(xlim = c(2021, 2025)) +
  theme_minimal() + theme(plot.title = element_text(color="firebrick",
face="bold"))

grid.arrange(p_zoom_res, p_zoom_ind, ncol = 1)

```



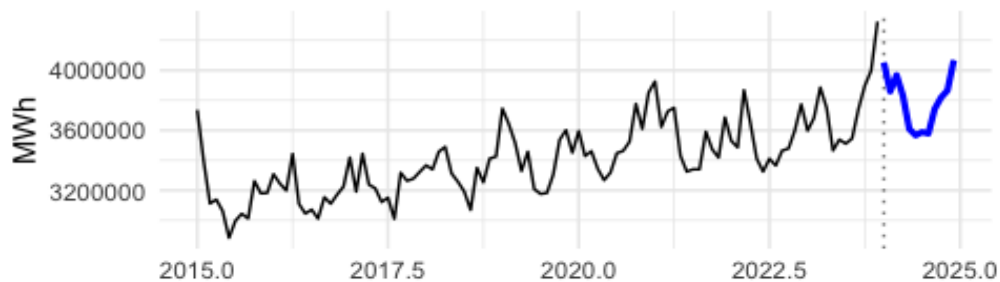
GRÁFICO B: PANORAMA COMPLETO

```
p_full_res <- autoplot(final_res, PI = FALSE) +
  geom_vline(xintercept = inicio_prev, linetype="dotted", color="grey50") +
  autolayer(final_res$mean, series="Previsão", size=1.1, color="blue") +
  labs(title = "Histórico Completo + Previsao: Residencial SP (2015-2025)",
    y="MWh", x="") +
  theme_minimal() + theme(plot.title = element_text(color="steelblue",
    face="bold"), legend.position = "none")

p_full_ind <- autoplot(final_ind, PI = FALSE) +
  geom_vline(xintercept = inicio_prev, linetype="dotted", color="grey50") +
  autolayer(final_ind$mean, series="Previsão", size=1.1, color="red") +
  labs(title = "Histórico Completo + Previsao: Industrial SP (2015-2025)",
    subtitle = "Ciclos Econômicos", y="MWh", x="") +
  theme_minimal() + theme(plot.title = element_text(color="firebrick",
    face="bold"), legend.position = "none")

grid.arrange(p_full_res, p_full_ind, ncol = 1)
```

Histórico Completo + Previsao: Residencial SP (



Histórico Completo + Previsao: Industrial SP (20

Ciclos Econômicos

