

# Teste - 4intelligence

## PROVA DE CIÊNCIA DE DADOS

Candidata: Juliane do Carmo Duare Magalhães

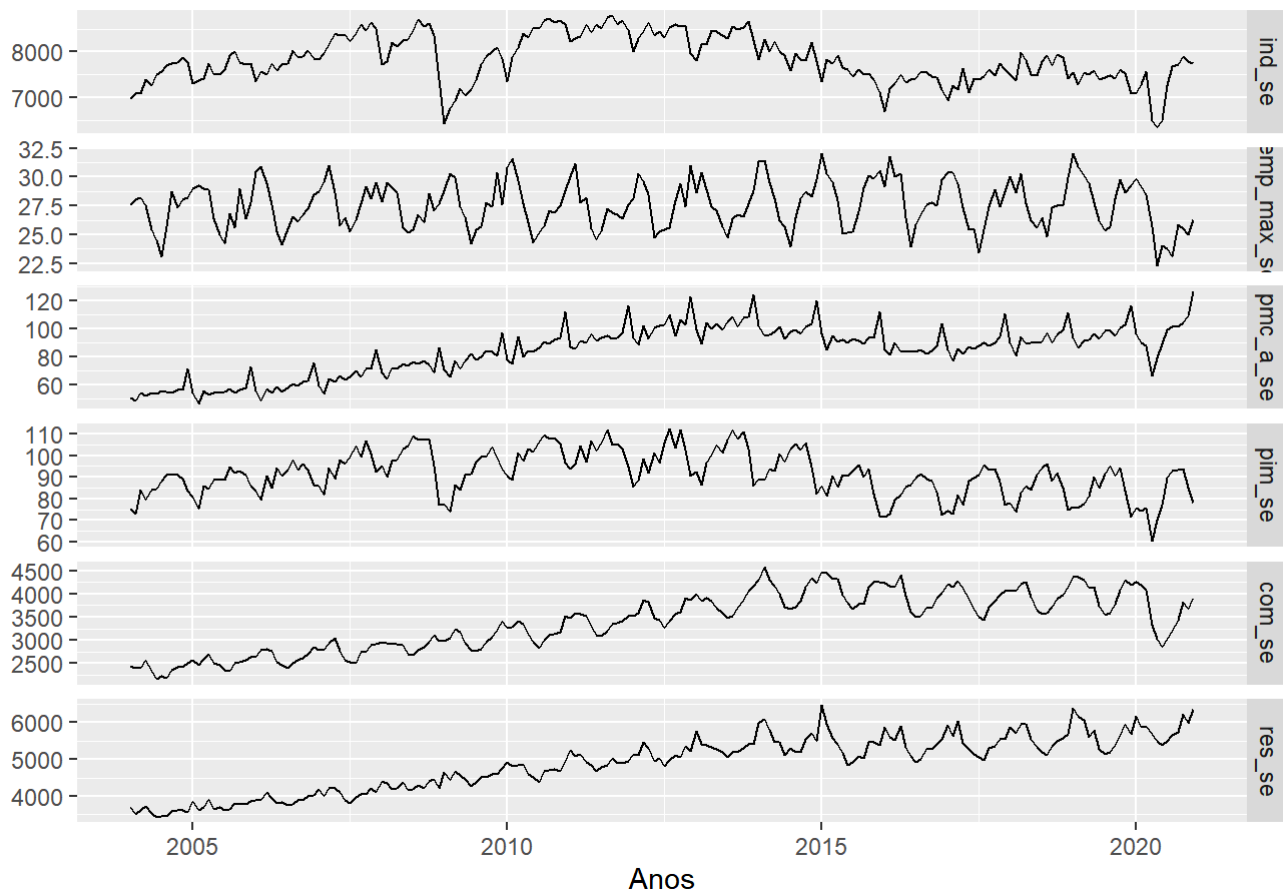
### 1.Dados

#### Análise Descritiva #

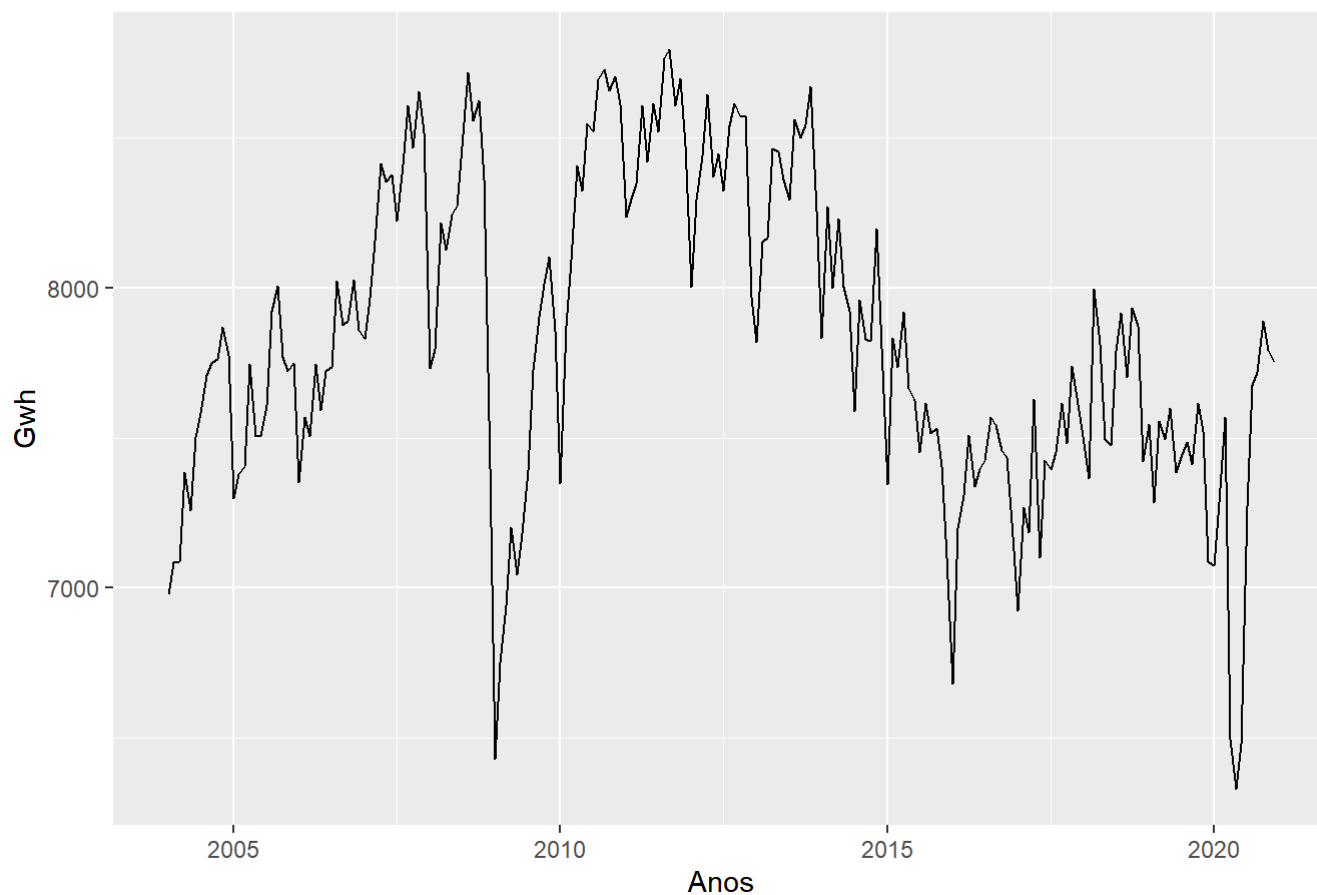
Foram selecionadas da base de dados requerida as variáveis: Temperatura máxima, Pesquisa mensal do comércio ampliada, Produção industrial, Consumo de energia do comércio e residencial, além da série de consumo da indústria, todas referentes à região Sudeste do país.

É importante visualizar as séries e assim identificar o comportamento sazonal, tendencioso e/ou cíclico. O comportamento de demanda por energia tende a apresentar sazonalidade devido às oscilações de temperatura e clima na região, como observamos nos gráficos abaixo:

Dados - Sudeste



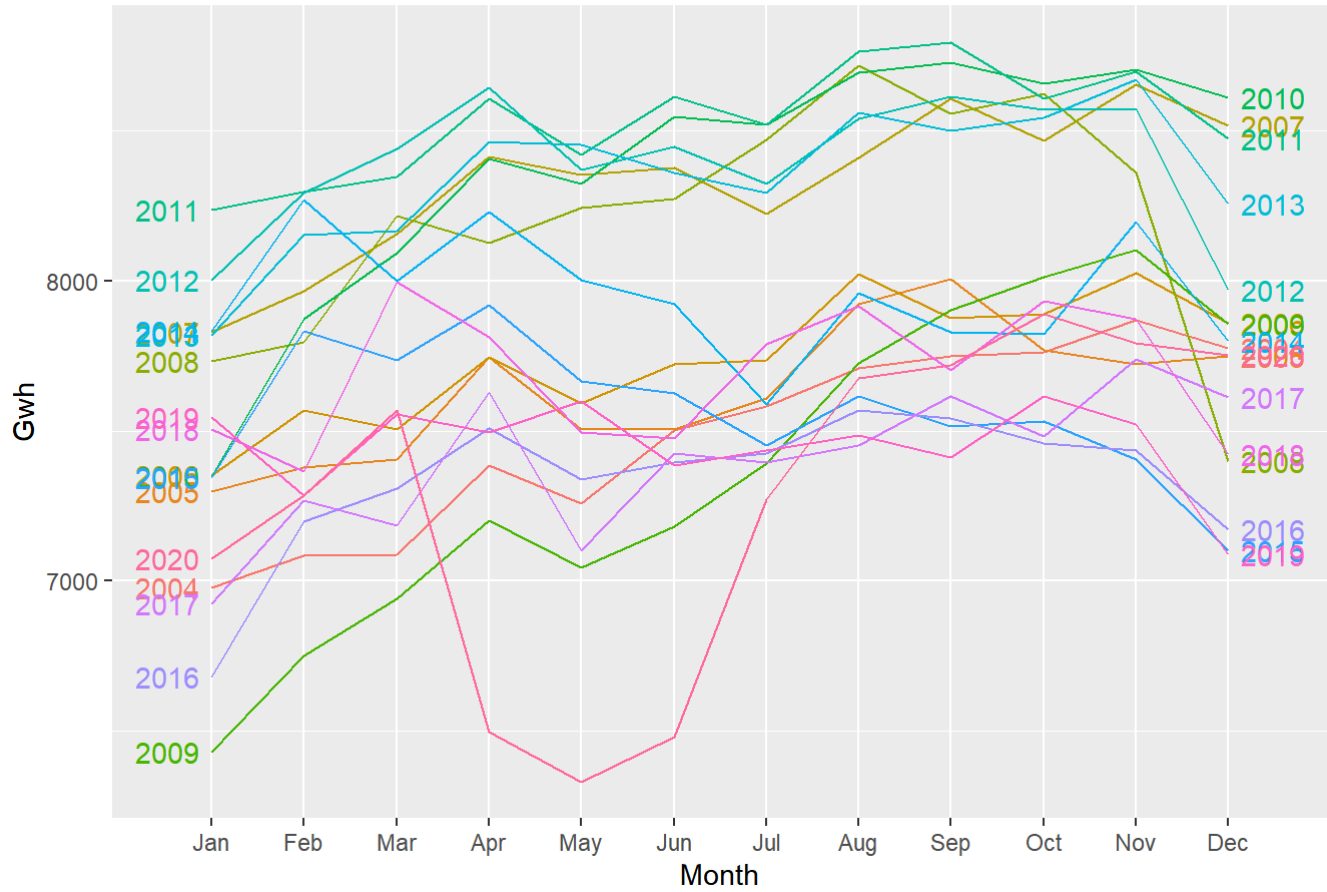
## Consumo de Energia na indústria - Sudeste - 2004-2020



O gráfico a seguir mostra uma queda na temperatura máxima nos meses do junho e julho.

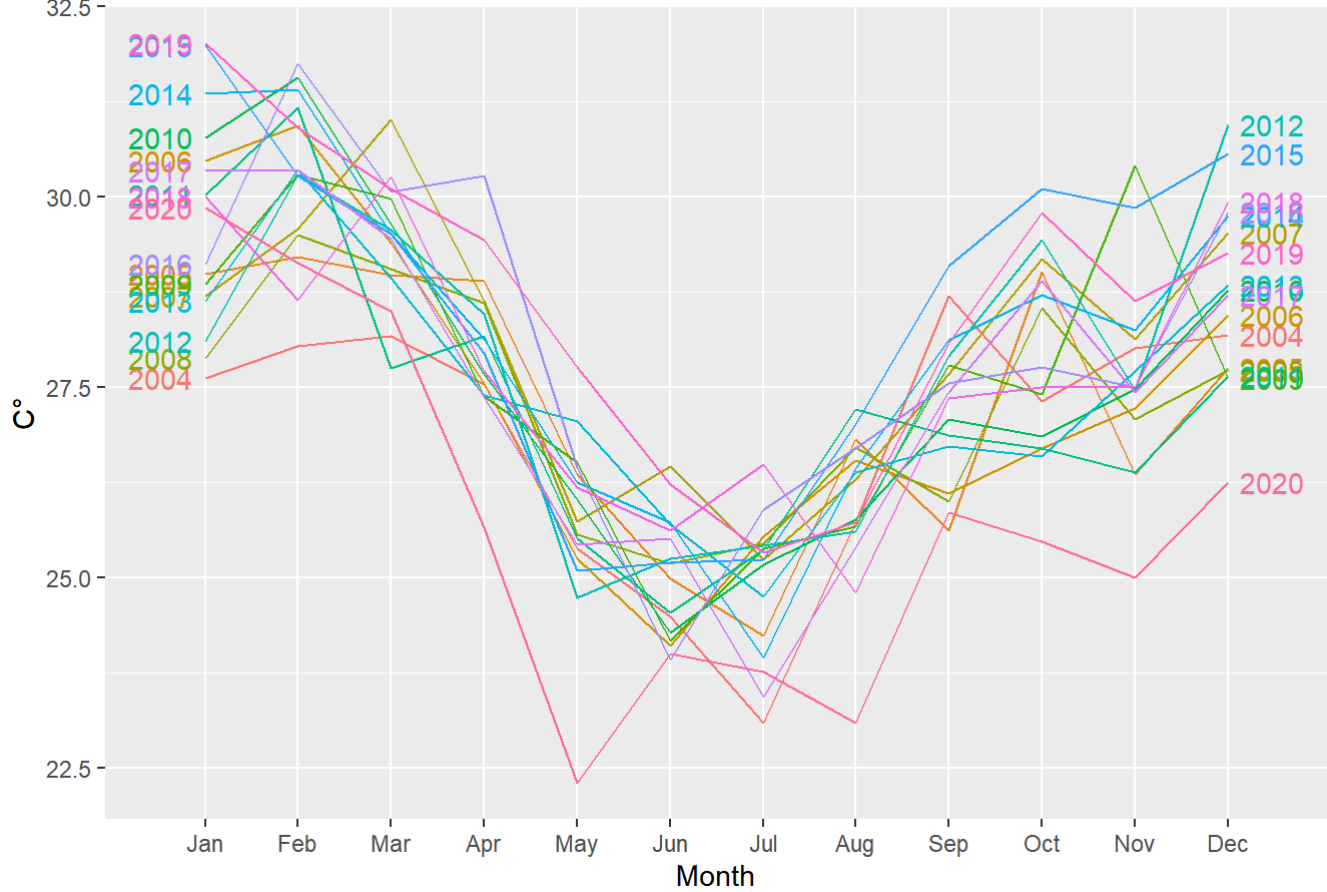
```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.
```

Sazonalidade - Consumo



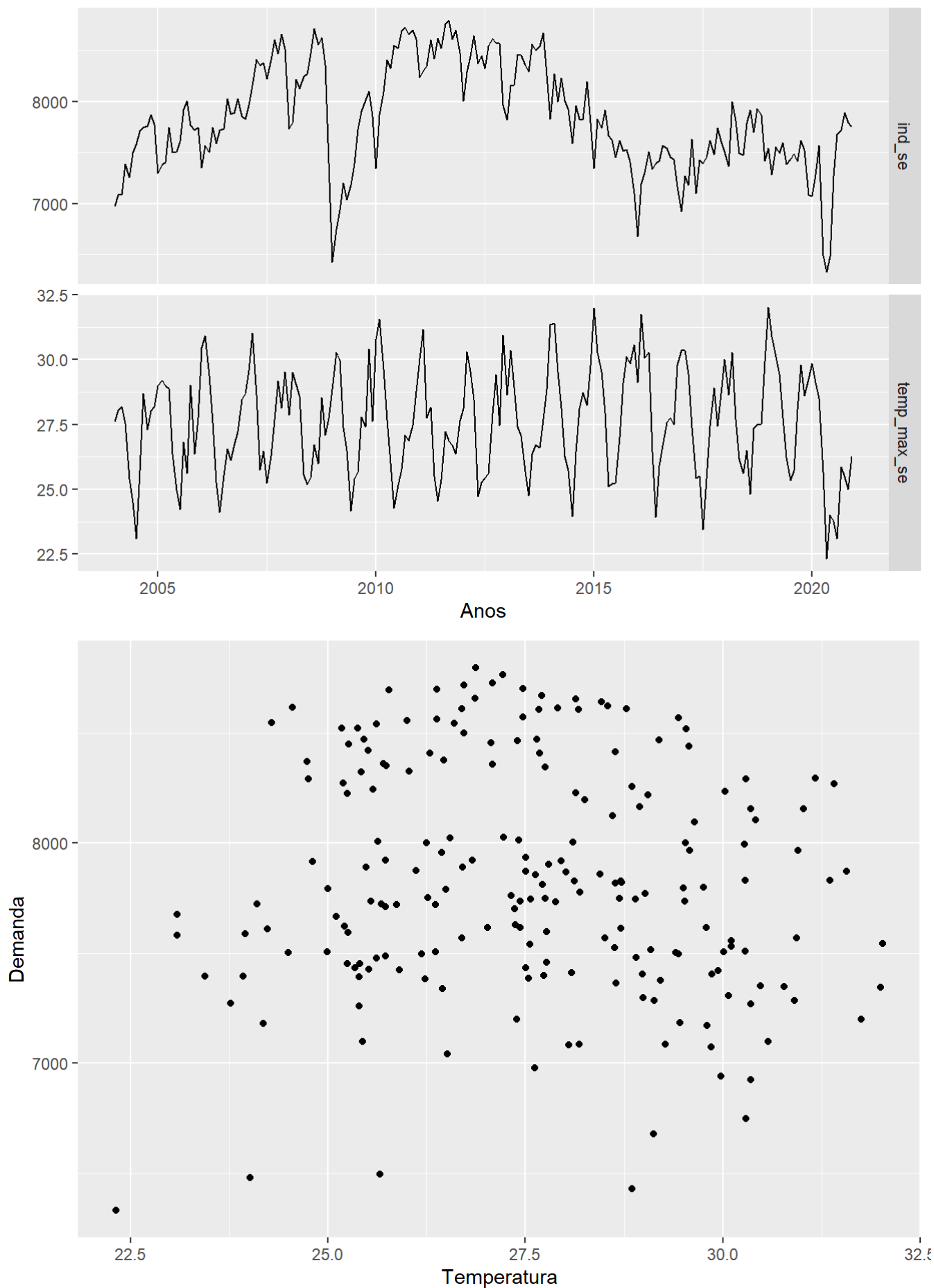
```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.
```

Sazonalidade - Temperatura



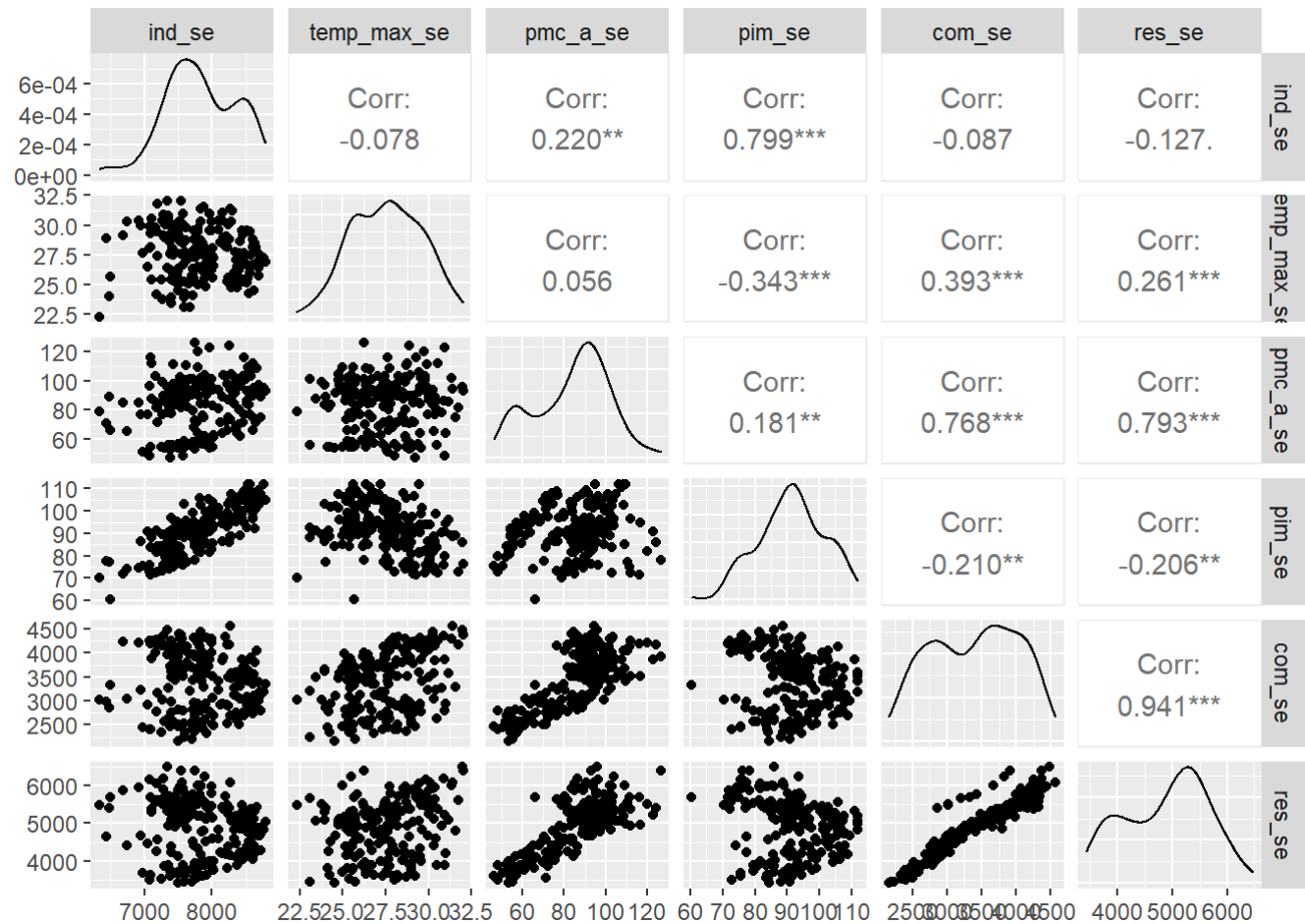
A relação entre as séries é fator de destaque. Observemos a relação entre demanda de energia pela indústria e as temperaturas máximas observadas:

### Uso de energia e Temperaturas máximas no Sudeste



Não há uma relação crescente óbvia como poderíamos supor. Agora analisando a relação entre todas as séries:

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
```



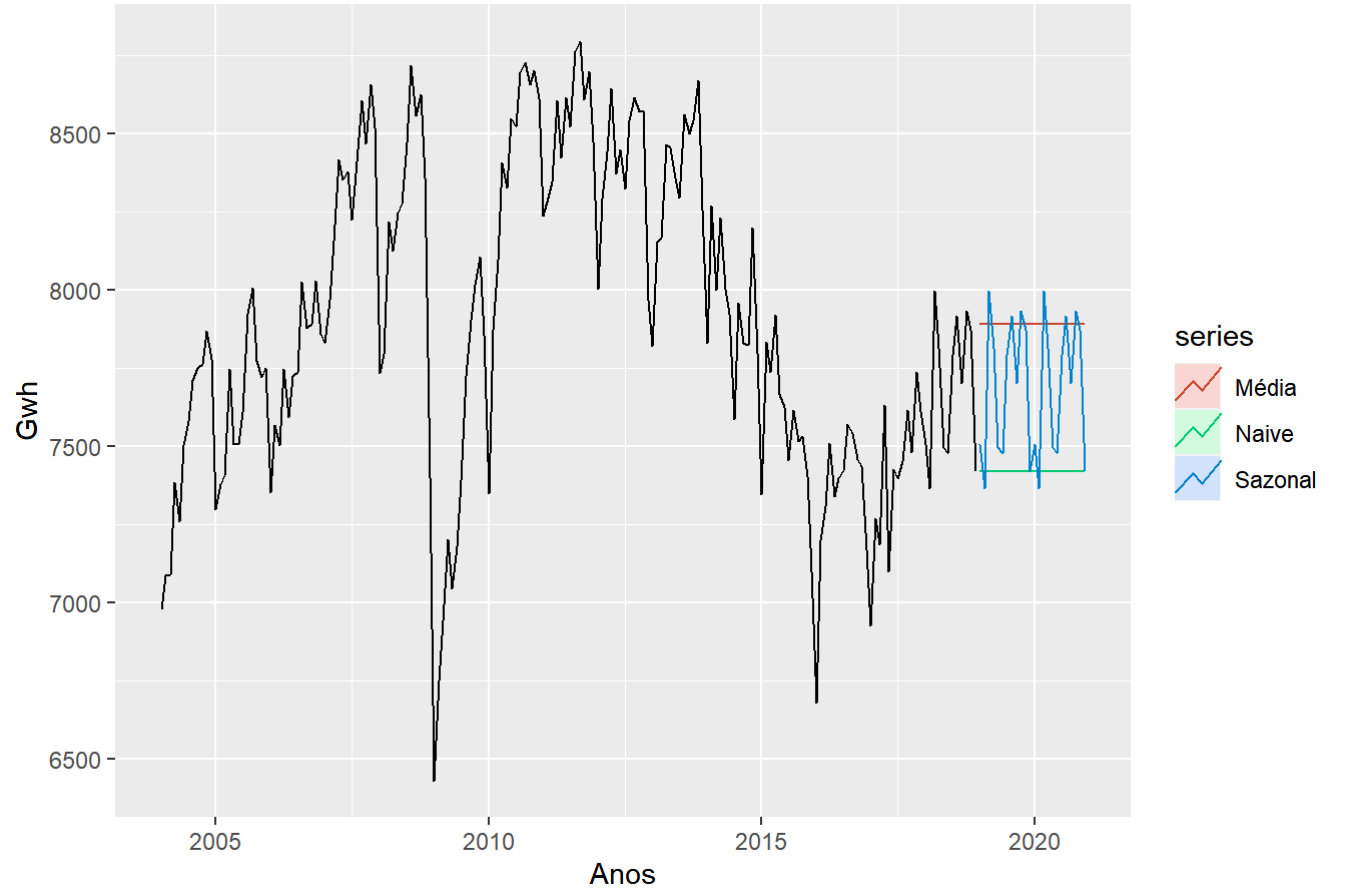
O aumento na produção industrial é o principal condutor de aumentos na demanda por energia na indústria. Outras informações básicas sobre as séries podem ser obtidas através de um simples comando: summary()

```
##      ind_se      temp_max_se      pmc_a_se      pim_se
## Min.   :6331   Min.   :22.31   Min.   : 46.53   Min.   : 60.32
## 1st Qu.:7472   1st Qu.:25.84   1st Qu.: 71.54   1st Qu.: 84.97
## Median :7784   Median :27.59   Median : 87.68   Median : 91.26
## Mean   :7830   Mean   :27.57   Mean   : 84.05   Mean   : 91.43
## 3rd Qu.:8278   3rd Qu.:29.10   3rd Qu.: 96.13   3rd Qu.: 98.13
## Max.   :8796   Max.   :32.02   Max.   :126.26   Max.   :112.05
##      com_se      res_se
## Min.   :2159   Min.   :3433
## 1st Qu.:2844   1st Qu.:4223
## Median :3483   Median :5035
## Mean   :3394   Mean   :4889
## 3rd Qu.:3906   3rd Qu.:5456
## Max.   :4572   Max.   :6477
```

Forecasting

Existem métodos simples de previsão baseados nas informações passadas chamados métodos ingênuos ou naive methods. Como a série apresenta sazonalidade, podemos usar um método que leve esse componente em conta:

Previsões para consumo de energia da indústria



##		ME	RMSE	MAE	MPE	MAPE	MASE
##	Training set	-2.880853e-13	503.1220	421.4815	-0.4137474	5.381909	1.165809
##	Test set	-5.265481e+02	663.5598	526.5481	-7.4988882	7.498888	1.456421
##		ACF1 Theil's U					
##	Training set	0.8574341	NA				
##	Test set	0.6432656	2.080246				

##		ME	RMSE	MAE	MPE	MAPE	MASE
##	Training set	2.47495	258.2037	198.0928	-0.02240576	2.552542	0.5479205
##	Test set	-55.32443	407.5776	286.0708	-1.07999786	4.077776	0.7912660
##		ACF1 Theil's U					
##	Training set	-0.07262856	NA				
##	Test set	0.64326564	1.286634				

##		ME	RMSE	MAE	MPE	MAPE	MASE
##	Training set	15.11996	495.3689	361.5356	-0.001967266	4.624914	1.000000
##	Test set	-325.01832	497.7508	365.9540	-4.688935882	5.220406	1.012221
##		ACF1 Theil's U					
##	Training set	0.8936234	NA				
##	Test set	0.6076240	1.559025				

O método básico que se mostrou mais adequado entre os acima foi o método “naive”. Outros métodos são: Holt, Holt-Winters, entre outros.