

# Ciência de Dados e suas Aplicações em Cenários de Pandemia



**Trio Nordeste**

Andressa Gabrielly M. Marçal  
Décio Gonçalves de Aguiar Neto  
Diego Alysson Braga Moreira

# Hipótese e Questão de Pesquisa

**Hipótese:** O clima pode ter influência na quantidade de novos casos

**Questão de Pesquisa:** Variações de temperatura e umidade apresentam algum tipo de correlação com a quantidade de novos casos nas cidades analisadas?

# Trabalhos Relacionados

— — —

	Variável observada	Conclusão	Região analisada
Tobias e Molina	Temperatura	Aumento no número de casos	Mundial
Sobral	Temperatura	Redução no número de casos	Mundial
Peng Shi	Temperatura	Redução no número de casos	China
Prata	Temperatura	Redução no número de casos	Capitais Brasileiras

# Materiais e Métodos

# Bases de dados

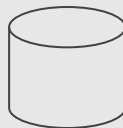
— — —

- Covid-19 (Brasil.io)
  - Contêm dados relacionados ao Covid no Brasil (Novos casos, Óbitos, População)
  - Os dados são de todas as cidades
- Instituto Nacional de Meteorologia (INMET)
  - Contêm dados meteorológicos (**temperatura** e **umidade**)
  - Possui dados de algumas estações, **não contemplando** todas as cidades do Brasil

INMET



Brasil.io

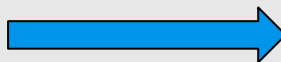
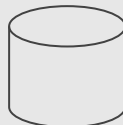


Seleção dos  
campos de  
interesse das  
bases de dados

INMET -  
Cidade



Covid -  
Cidade



Junção das bases  
com base nas  
datas

Covid -  
Clima



## Modelo Utilizado: Regressão OLS(Método dos Mínimos Quadrados - MMQ)




— — —

- Procura encontrar o melhor ajuste para um conjunto de dados;
- Tenta minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados;



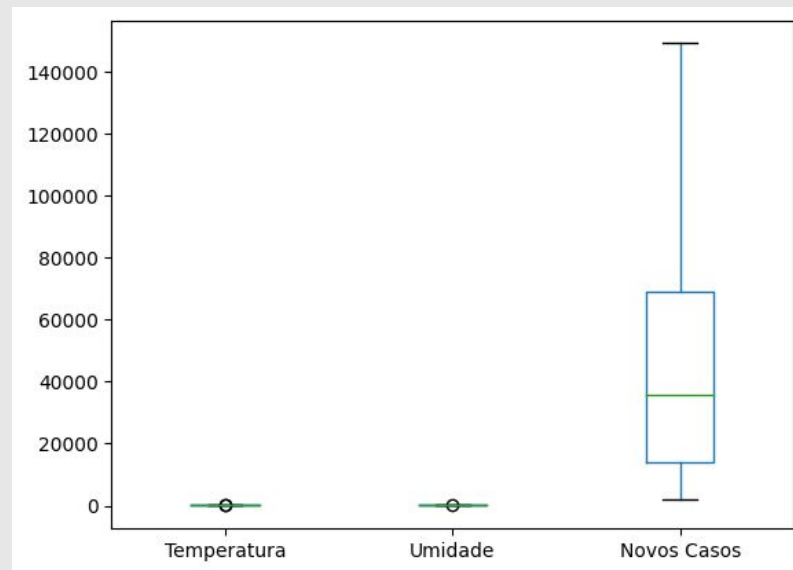
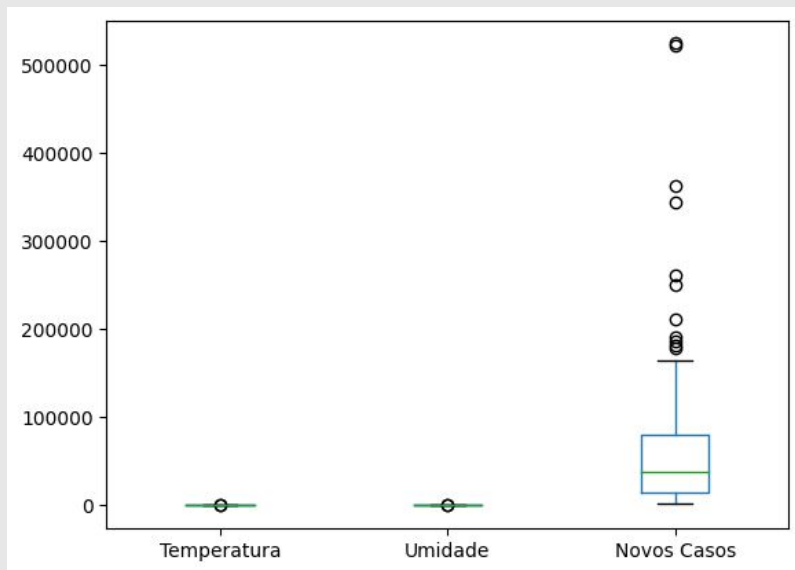
# Resultados

# Estudo dos dados para o Brasil por mês

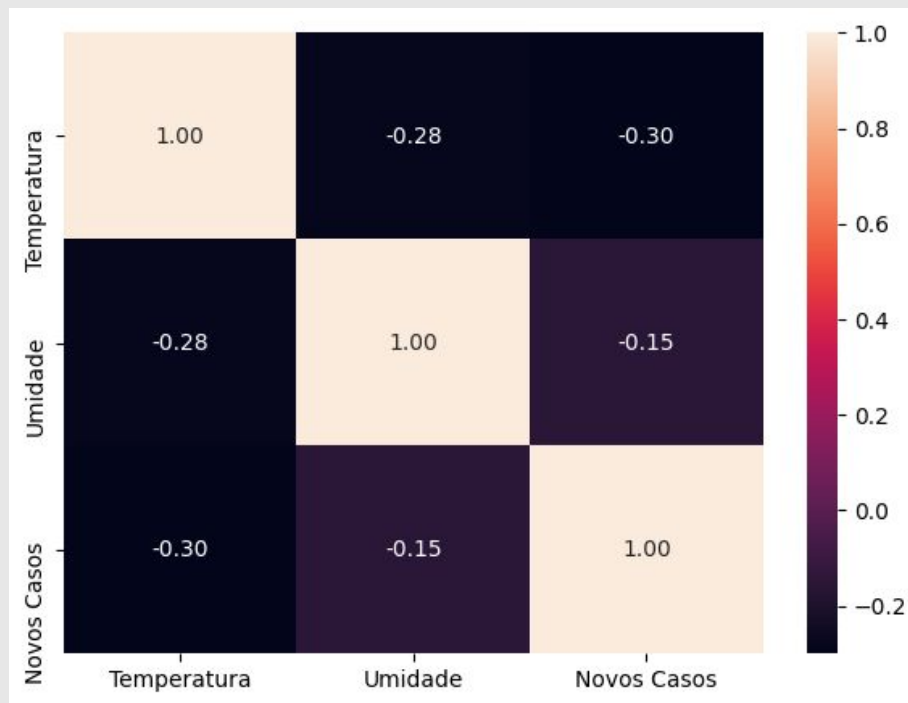
	TEMPERATURA DO AR - ... flo... 12.435 - 30.074	UMIDADE RELATIVA DO ... flo... 36.486 - 86.012	new_confirmed int64 1816 - 524076
			
<b>26</b>	21.992840430752445	71.69502415512773	4674
<b>29</b>	27.033429600425833	74.56607793478507	2054
<b>30</b>	26.489006050990323	83.11805798953543	10158
<b>31</b>	26.04722063419965	86.01150847777305	2140
<b>32</b>	24.92150137916182	79.25649092443824	5300
Expand rows 5 - 202			
<b>239</b>	20.60941272153713	69.06826004461095	148964
<b>240</b>	19.281860596653832	74.76223861954652	210808
<b>241</b>	25.924959938692453	71.57900857492768	11644
<b>242</b>	21.963868142685417	63.83724739573044	251052
<b>243</b>	27.042033448063147	71.98496309400113	12694

208 rows x 3 columns

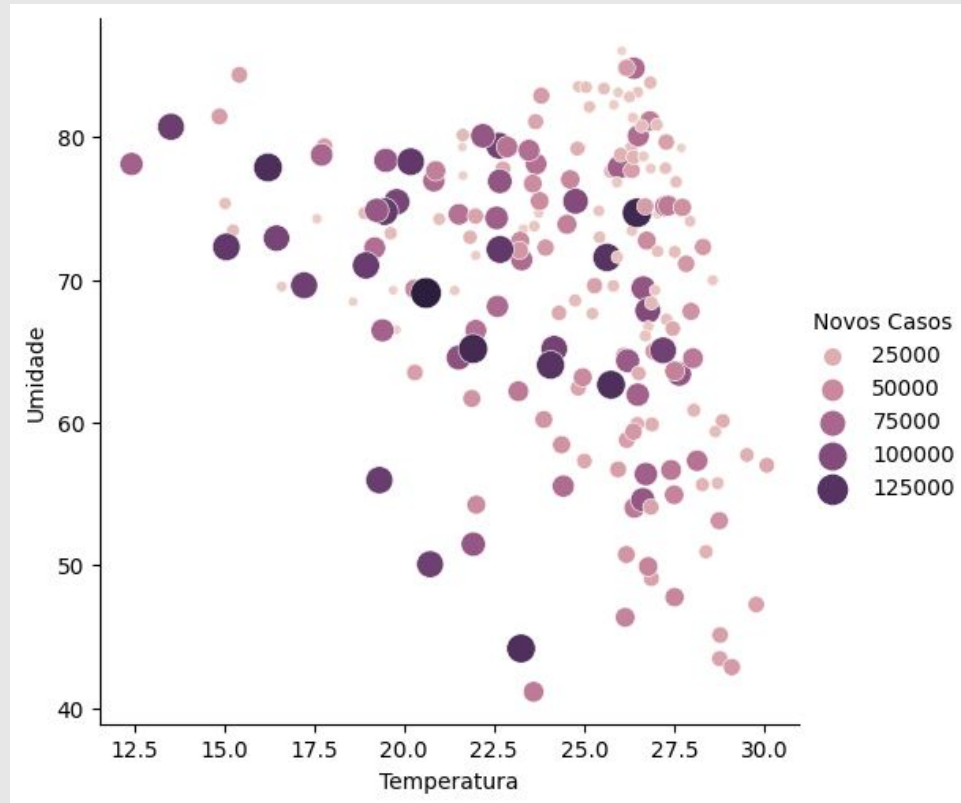
## Outliers - Brasil



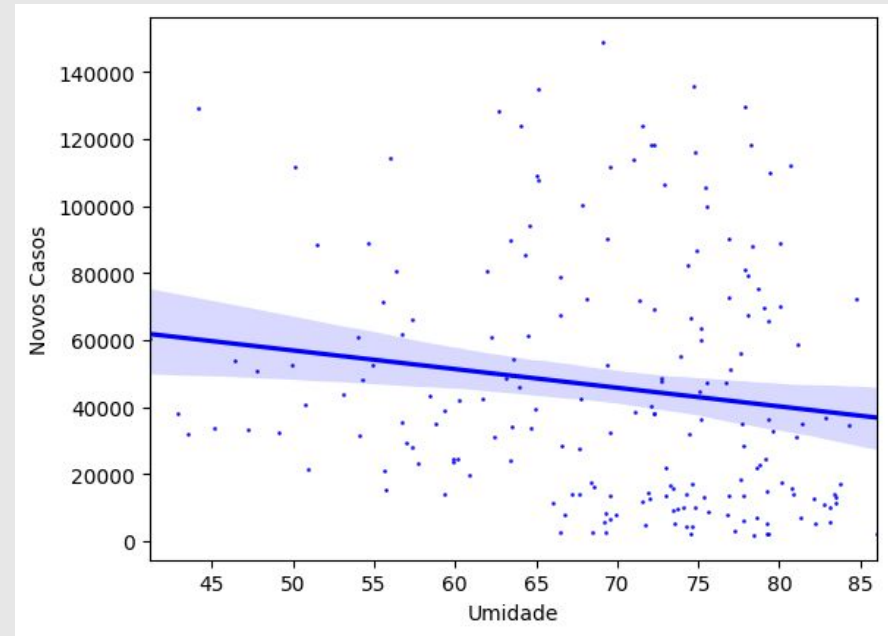
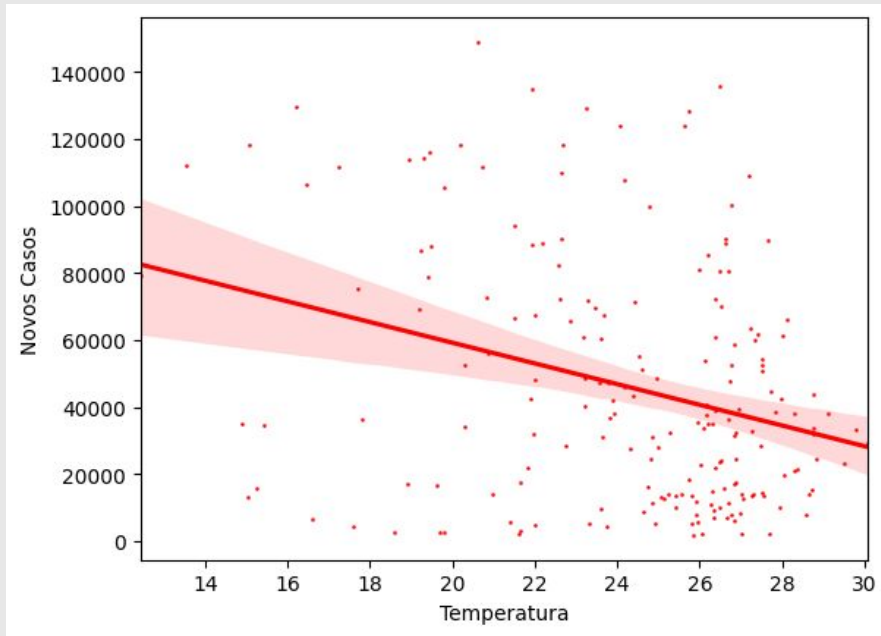
## Spearman - Brasil



## Gráfico de Relação - Brasil



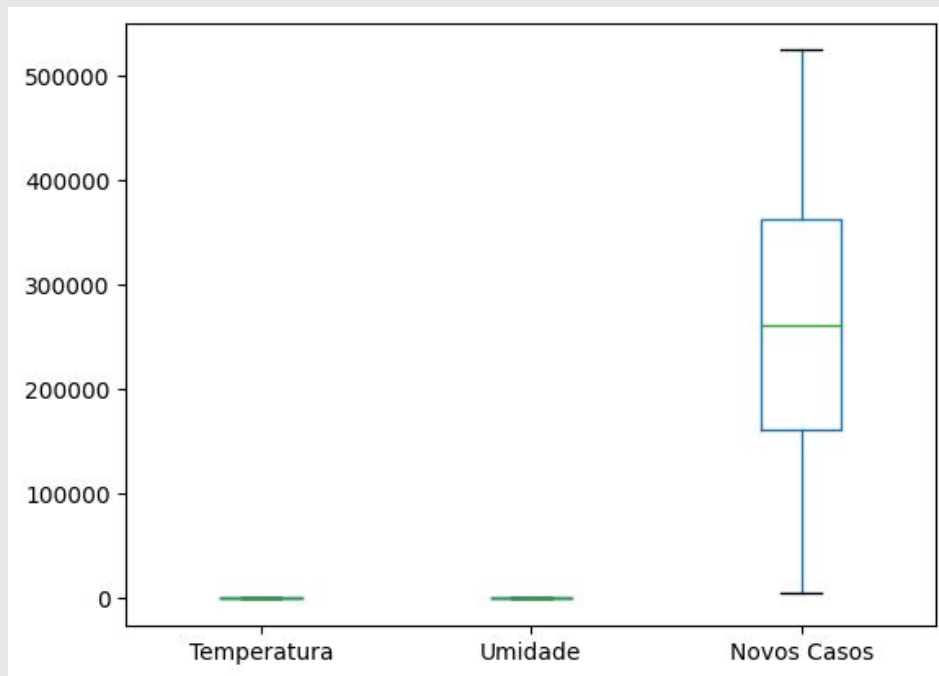
## Plot da Regressão - Brasil



# Dados para o Estado de São Paulo por mês

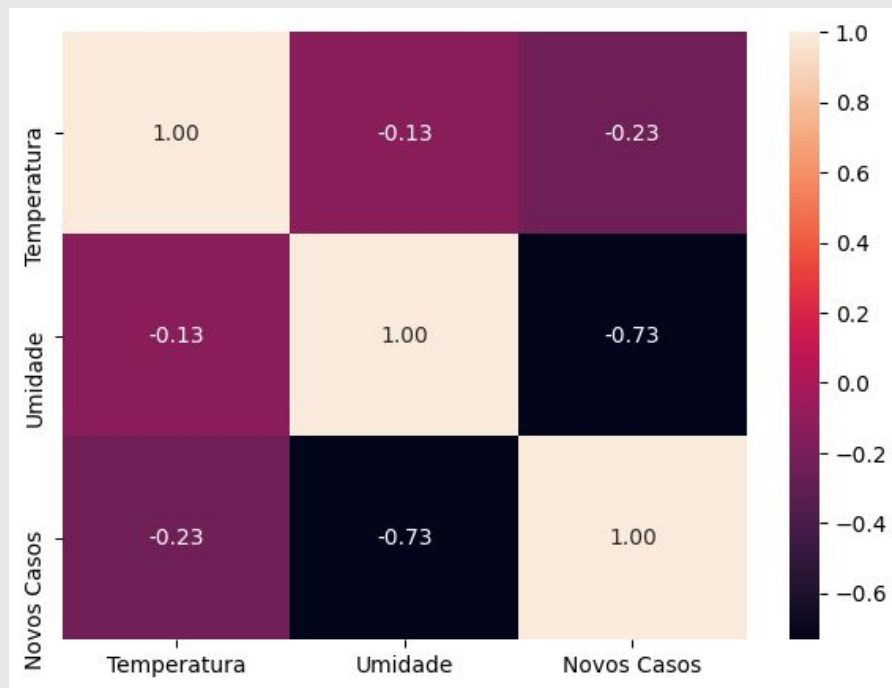
TEMPERATURA DO AR - ... flo...	UMIDADE RELATIVA DO ... flo...	new_confirmed int64
17.219 - 22.411	54.447 - 71.695	4674 - 524076
		
21.992840430752445	71.69502415512773	4674
20.294385870518344	69.35066747797886	52718
17.21874106154236	67.21720797375787	162000
18.47931505225771	71.04116726539237	343364
17.932724417758738	63.44949648472444	521848
17.853578071197855	61.546176873456794	524076
22.286553966343366	54.447286550999635	362572
22.4111279985193	62.50065449773793	260998
21.963868142685417	63.83724739573044	251052

## Outliers - São Paulo/Mês

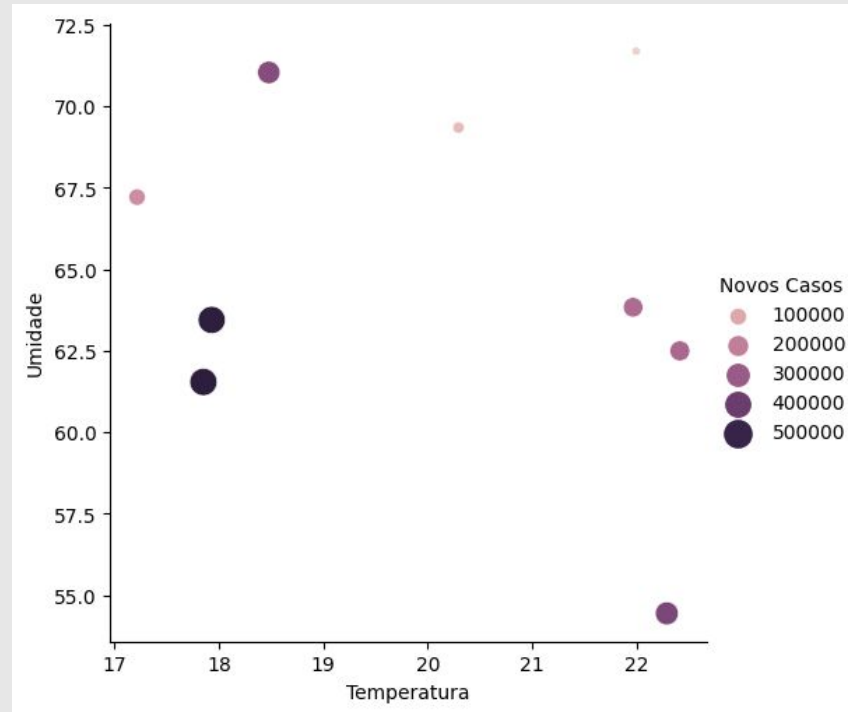




## Spearman - São Paulo/Mês



## Gráfico de Relação - São Paulo/Mês



# OLS Regression Results

```

=====
Dep. Variable:      Novos Casos   R-squared (uncentered):      0.675
Model:              OLS           Adj. R-squared (uncentered):  0.582
Method:             Least Squares F-statistic:                    7.263
Date:               Tue, 15 Dec 2020 Prob (F-statistic):           0.0196
Time:               01:53:50      Log-Likelihood:              -121.97
No. Observations:   9             AIC:                          247.9
Df Residuals:       7             BIC:                          248.3
Df Model:           2
Covariance Type:    nonrobust
=====

```

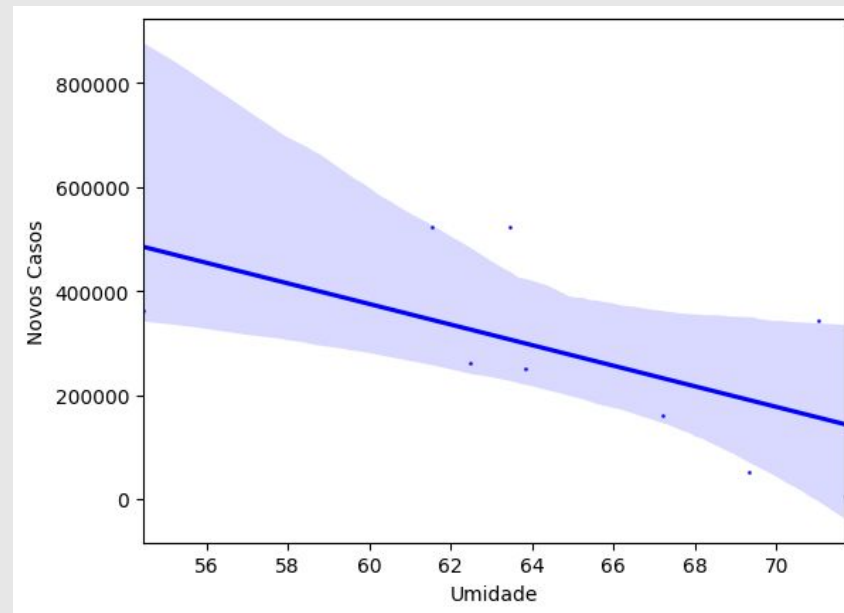
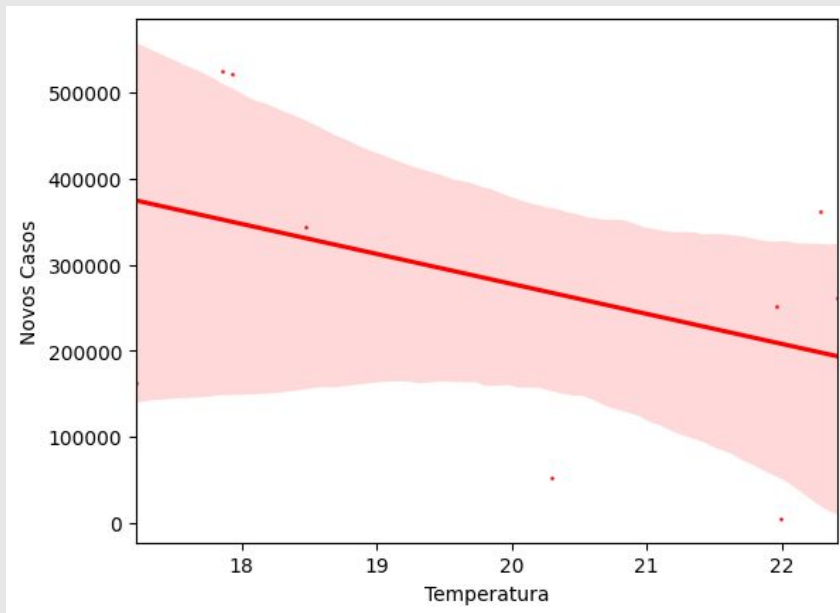
	coef	std err	t	P> t	[0.025	0.975]
Temperatura	7127.8670	2.45e+04	0.291	0.779	-5.07e+04	6.5e+04
Umidade	1914.3597	7561.940	0.253	0.807	-1.6e+04	1.98e+04

```

=====
Omnibus:            0.239   Durbin-Watson:           0.433
Prob(Omnibus):      0.887   Jarque-Bera (JB):       0.388
Skew:               -0.016   Prob(JB):               0.823
Kurtosis:           1.983   Cond. No.                24.8
=====

```

## Plot da Regressão - São Paulo/Mês

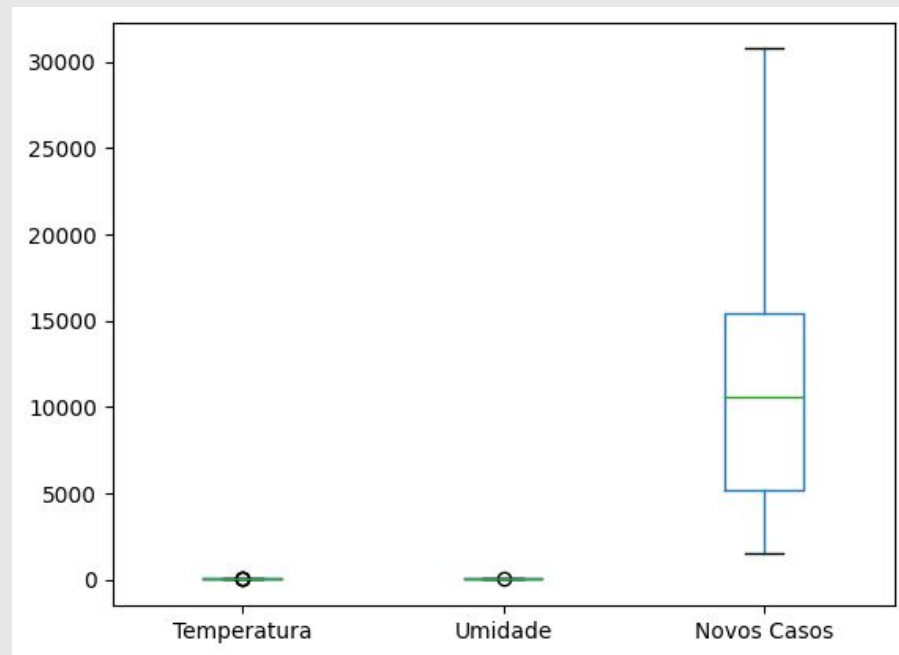
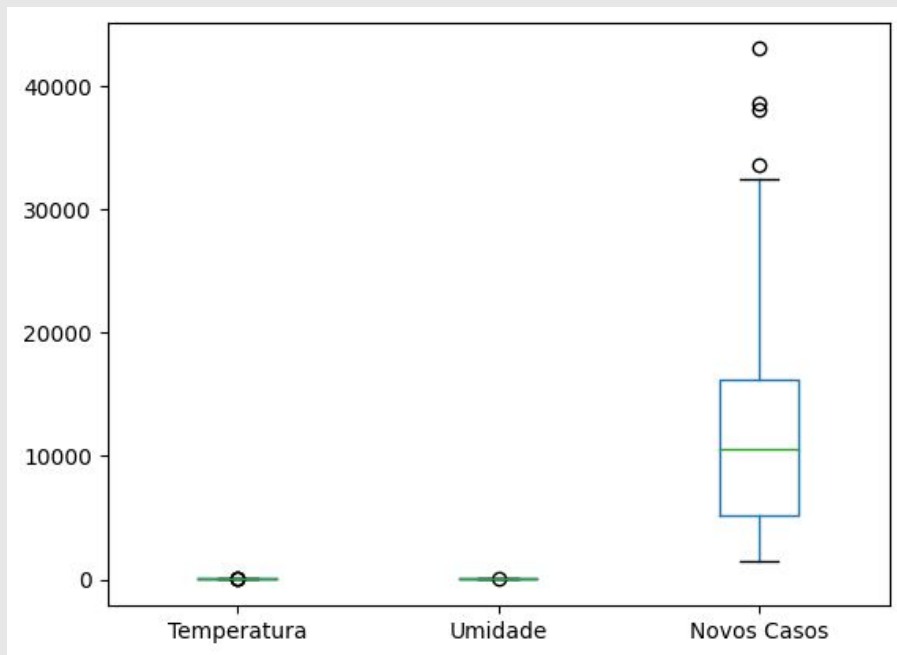


# Dados para o Estado de São Paulo por dia

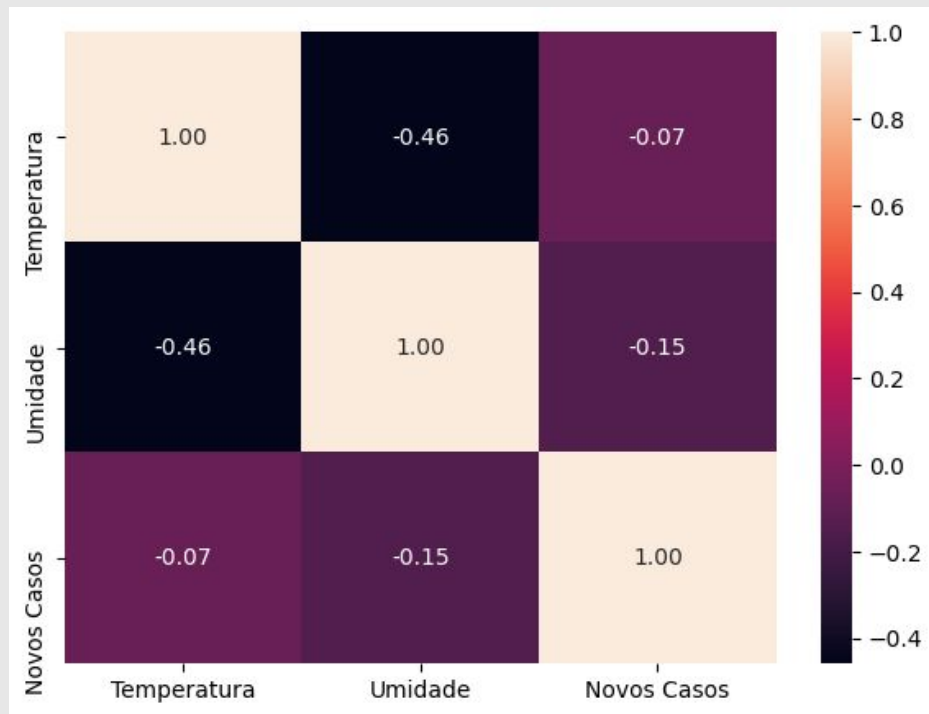
TEMPERATURA DO AR - ... flo...	UMIDADE RELATIVA DO ... flo...	new_confirmed int64
10.498 - 28.413	38.499 - 88	1544 - 43030
		
21.11239130434782	79.61495535714286	1632
20.879193899782138	76.95011086474501	2052
18.94328678839957	69.97379912663756	1544
18.809600862998906	75.00549450549451	3344
18.93419913419912	71.20953436807095	2546
id rows 5 - 206		
24.529618768328447	58.677611940298505	9046
23.99865871833084	60.53039513677812	8640
24.030803571428585	63.07142857142857	9014
24.93068702290075	58.42089093701997	4758
24.701555209953373	61.403174603174605	2360

212 rows x 3 columns

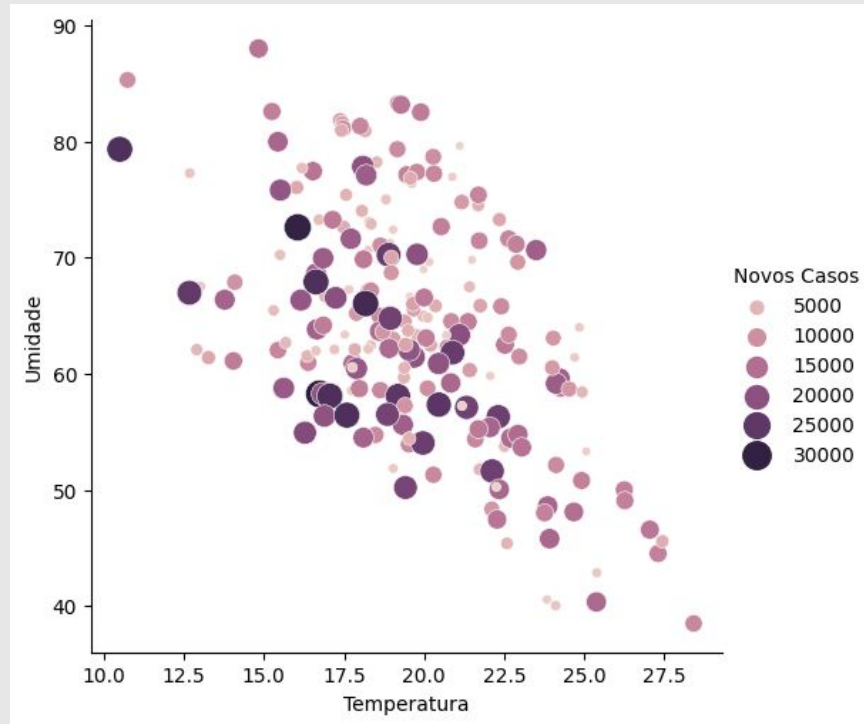
## Outliers - São Paulo/Dia



## Spearman - São Paulo/Dia



## Gráfico de Relação - São Paulo/Dia

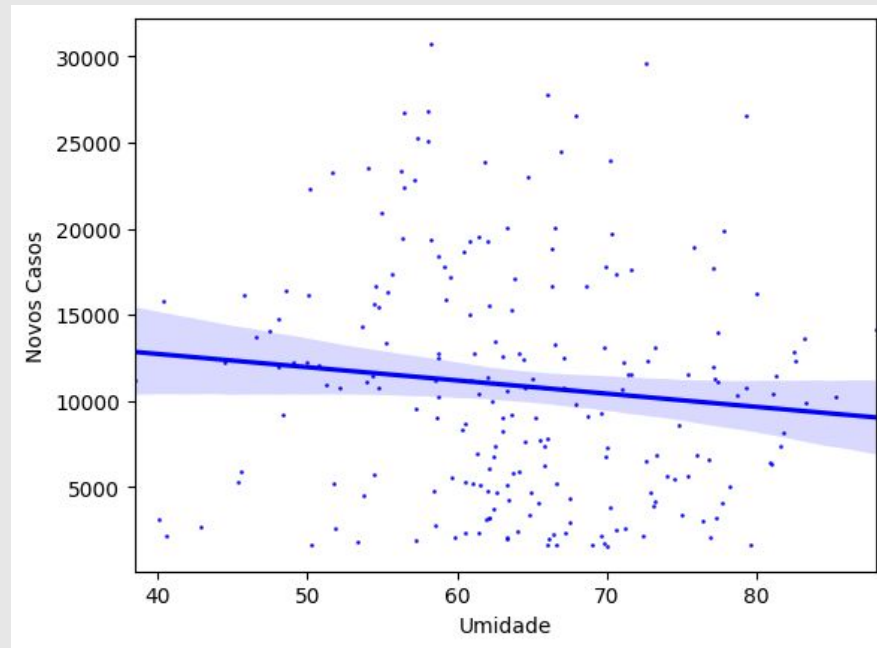
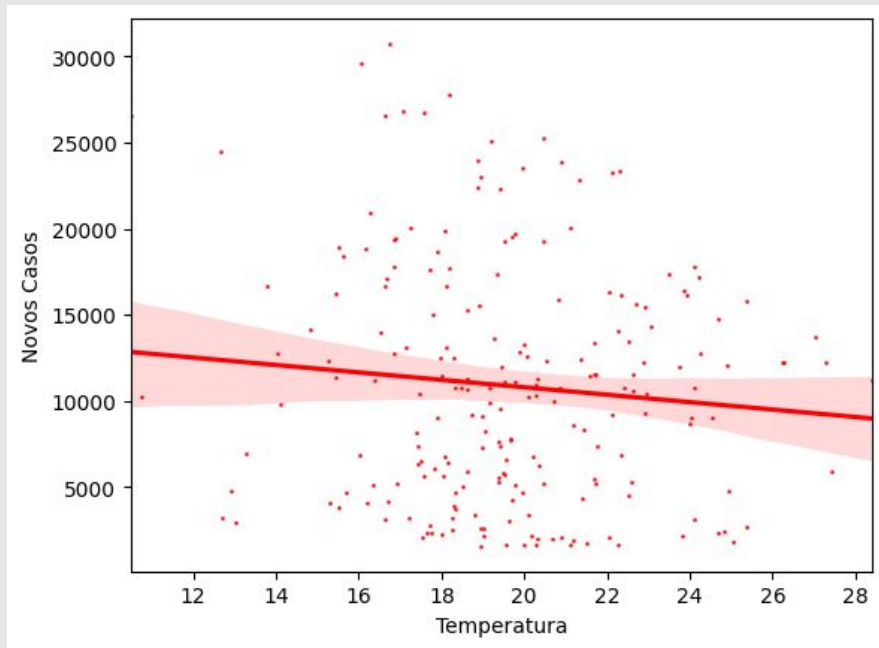




## Regressão OLS - Resultados

OLS Regression Results						
=====						
Dep. Variable:	Novos Casos	R-squared (uncentered):	0.697			
Model:	OLS	Adj. R-squared (uncentered):	0.694			
Method:	Least Squares	F-statistic:	241.2			
Date:	Tue, 15 Dec 2020	Prob (F-statistic):	3.96e-55			
Time:	13:17:28	Log-Likelihood:	-2180.0			
No. Observations:	212	AIC:	4364.			
Df Residuals:	210	BIC:	4371.			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
Temperatura	273.0935	92.046	2.967	0.003	91.641	454.546
Umidade	83.0586	28.186	2.947	0.004	27.495	138.622
=====						
Omnibus:	15.754	Durbin-Watson:	1.041			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	17.881			
Skew:	0.708	Prob(JB):	0.000131			
Kurtosis:	2.862	Cond. No.	13.3			
=====						

## Plot da Regressão - São Paulo/Dia



# Limitações e Trabalhos Futuros

— — —

- Uso de um modelo mais robusto que considere o dia em que iniciou a pandemia em cada lugar
- Adaptação do modelo para considerar mais variáveis que possam ser importantes para o contexto (Ex: Qualidade do Ar)

# Modelagem Futura

— — —

- *Generalized Additive Model* (GAM)
- O GAM pode ser utilizado para **explorar efeitos climáticos** lineares e não lineares e **resultados de saúde**

$$\log(y_{it}) = \beta_0 + \beta_1 x^3 - \beta_2 x^2 + \beta_2 x + s(mt_i) + s(dd_i) + s(ep_i)$$

$$\log(y_t) = \beta_0 + s(mt_t) + s(hu_t) + s(ep)$$