



# World Happiness Report (2019)

Catarina, Isabella, Júlia, Lucas e Tiago

# Objetivo

Prever a **pontuação de felicidade** (Score) de diferentes países com base em variáveis socioeconômicas e ambientais, como **PIB per capita, suporte social e expectativa de vida** e assim, verificar **quais características contribuem para um país mais feliz.**

# Sobre o banco de dados

O banco de dados "**The World Happiness Report**" apresenta o **ranking dos países mais felizes em 2019** de acordo com vários critérios, como a sua economia, expectativa de vida e percepção de corrupção. Como objetivo, esse banco quer mostrar tanto **o que falta para os países serem mais felizes**, quanto **ajudar os governos a identificar onde necessita-se mais investimento**.

<https://www.kaggle.com/datasets/unsdsn/world-happiness?select=2019.csv>

## Variável resposta

Nome da variável	Explicação
Score	Pontuação geral de felicidade, que será prevista pelo modelo

# Variáveis explicativas

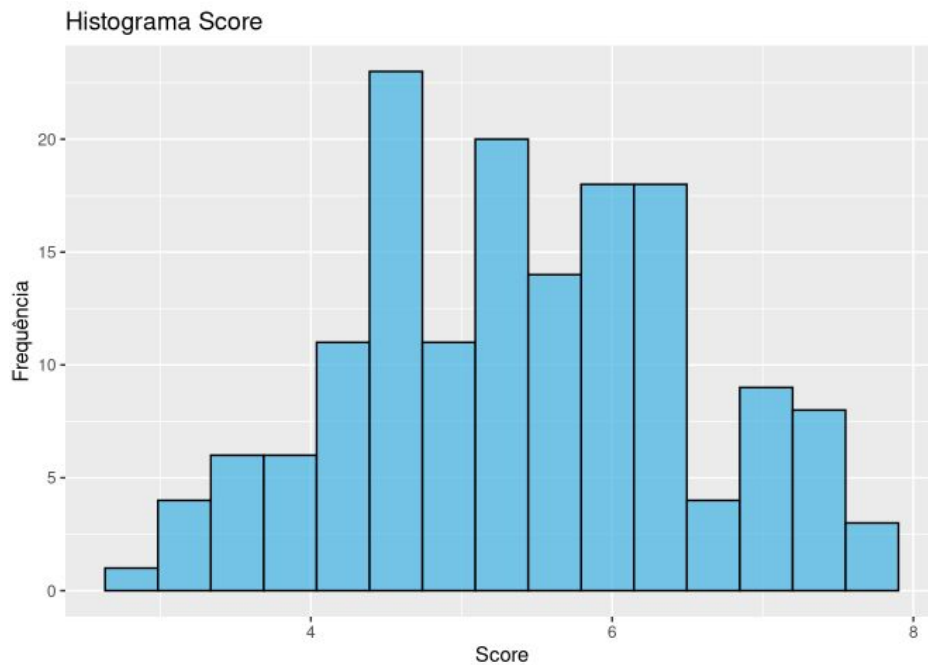
Nome da variável	Explicação
GDP per capita	Prosperidade econômica de um país
Social Support	Mede o suporte social disponível
Healthy expectancy	Reflete a saúde da população
Freedom to make life choices	Liberdade percebida nas decisões pessoais
Generosity	Disposição para ajudar os outros
Perceptions of corruptions	Percepção de corrupção no governo e nos negócios

# Análise descritiva



# SCORE

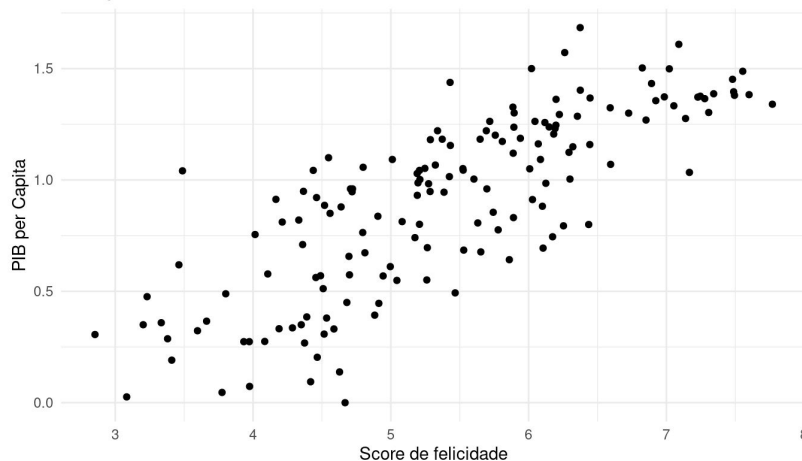
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	2.853	4.545	5.380	5.407	6.184	7.769



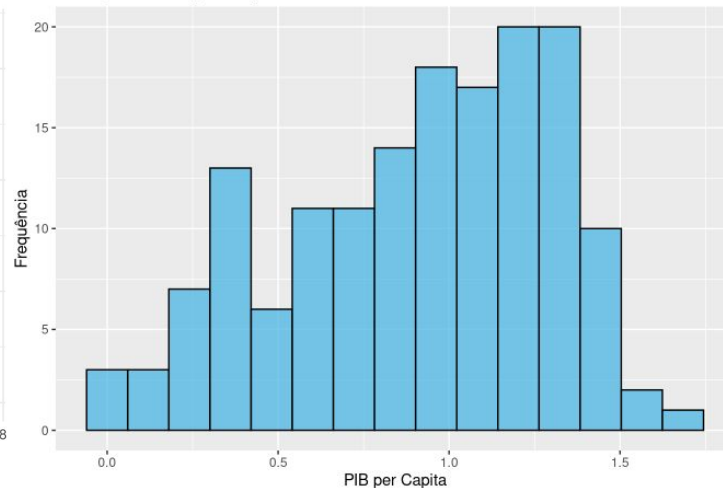
# PIB PER CAPITA

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0000	0.6028	0.9600	0.9051	1.2325	1.6840

Relação entre PIB per Capita e Felicidade



Histograma PIB per Capita



## O país com o maior PIB per capita é: Qatar - PIB per capita: 1.684

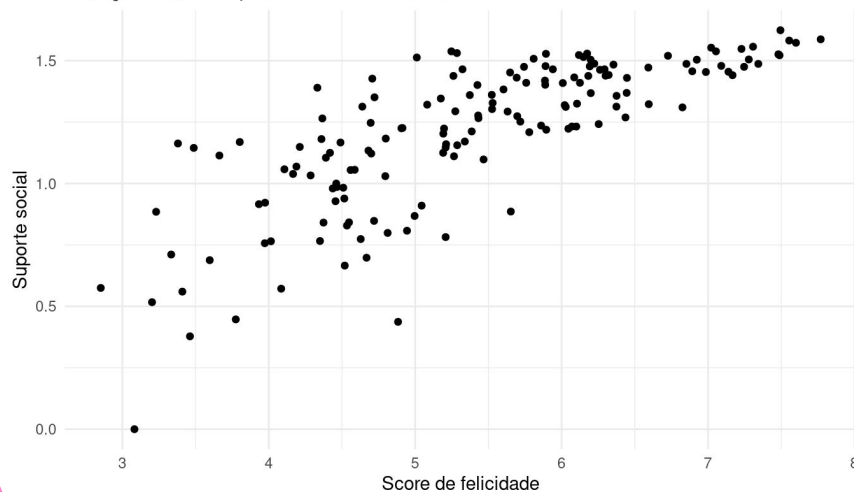
## O país com o menor PIB per capita é: Somalia - PIB per capita: 0



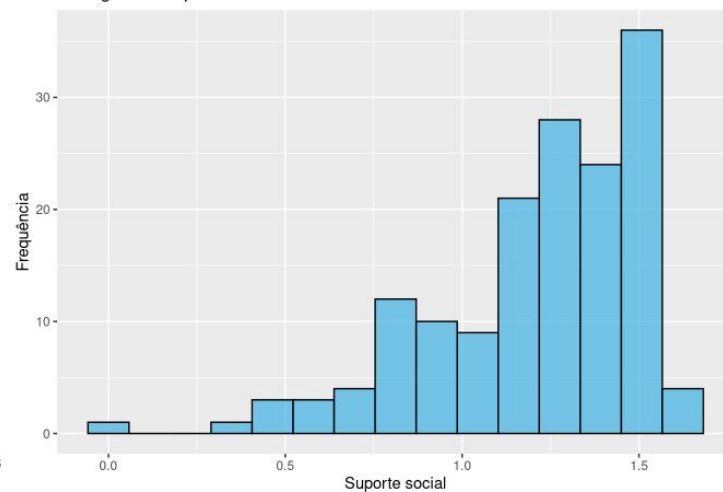
# SUORTE SOCIAL

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.000	1.056	1.272	1.209	1.452	1.624

Relação entre Suporte social e Felicidade



Histograma Suporte social



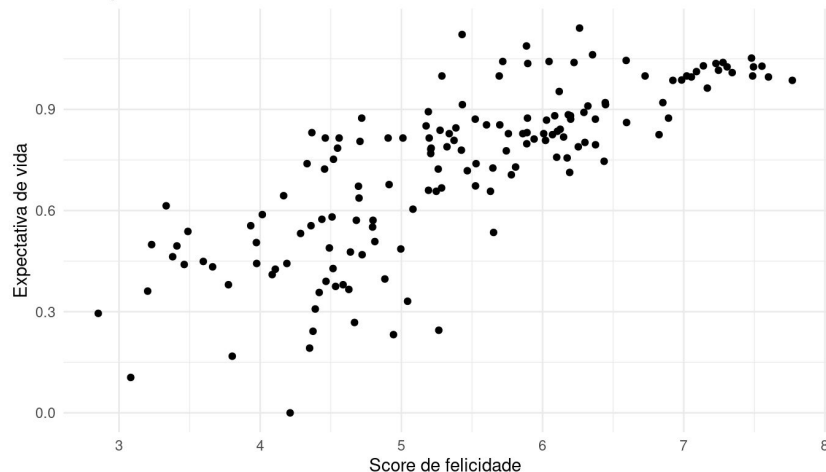
## O país com o maior suporte social é: Iceland - Suporte social 1.624

## O país com o menor suporte social é: Central African Republic - Suporte social: 0

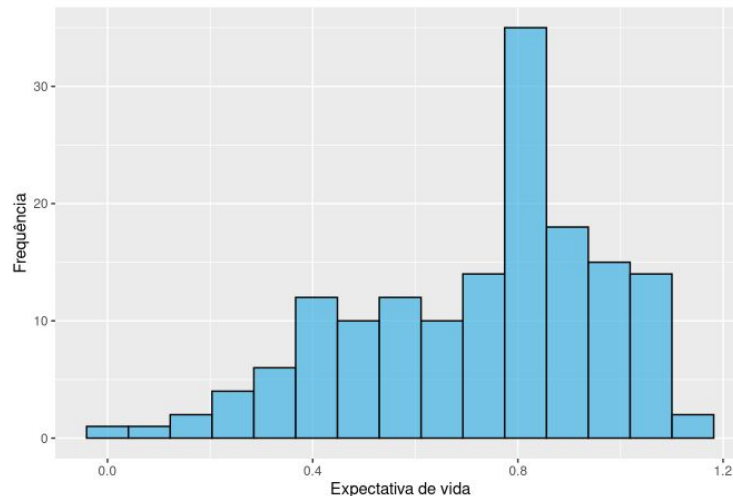
# EXPECTATIVA DE VIDA

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0000	0.5477	0.7890	0.7252	0.8818	1.1410

Relação entre Expectativa de vida e Felicidade



Histograma Expectativa de vida



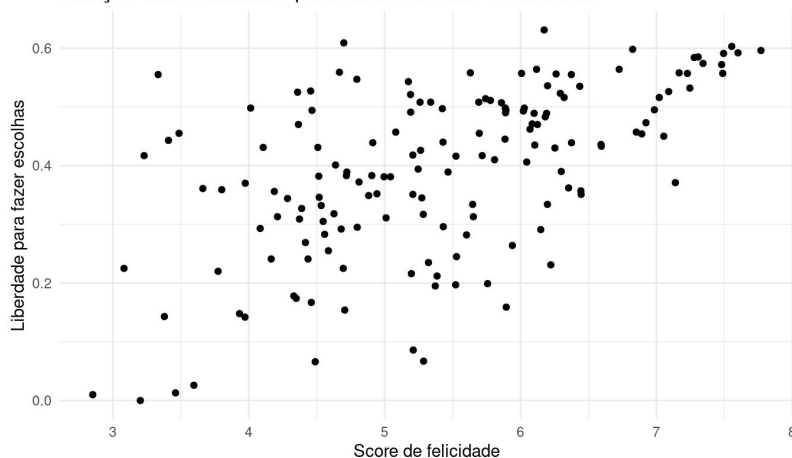
## O país com o maior expectativa de vida é: Singapore - Expectativa de vida: 1.141

## O país com o menor expectativa de vida é: Swaziland - Expectativa de vida: 0

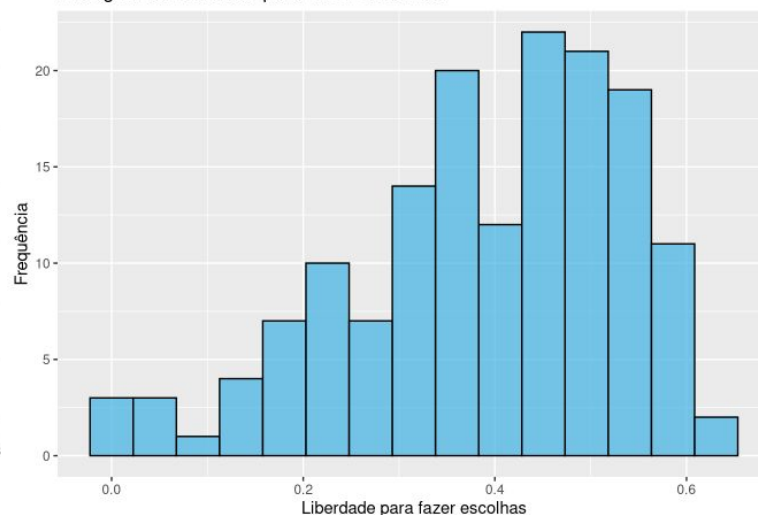
# LIBERDADE PARA FAZER ESCOLHAS

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0000	0.3080	0.4170	0.3926	0.5072	0.6310

Relação entre Liberdade para fazer escolhas e Felicidade



Histograma Liberdade para fazer escolhas



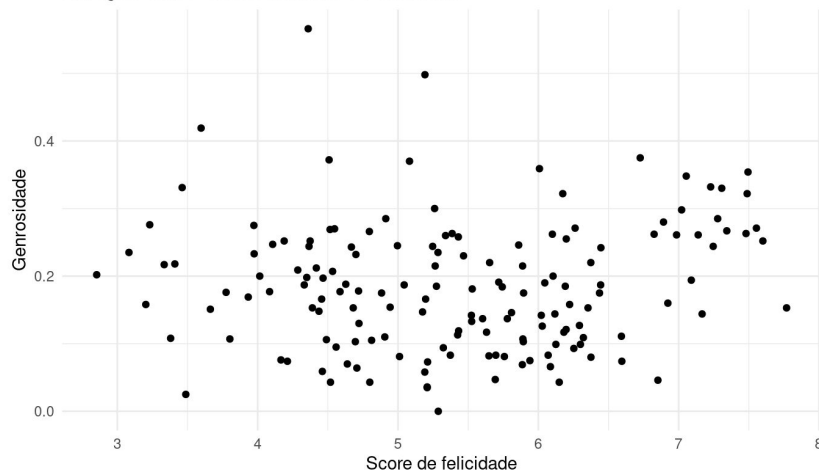
## O país com o maior liberdade para fazer escolhas é: Uzbekistan - Liberdade para fazer escolhas: 0.631

## O país com o menor liberdade para fazer escolhas é: Afghanistan - Liberdade para fazer escolhas: 0

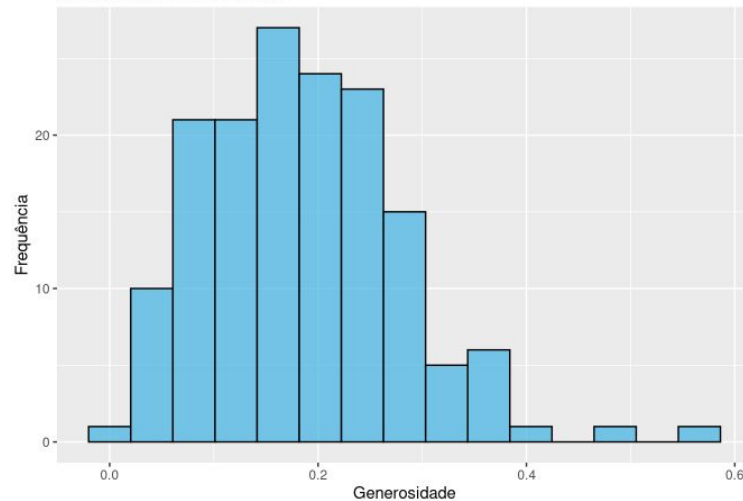
# GENEROSIDADE

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0000	0.1087	0.1775	0.1848	0.2482	0.5660

Relação entre Generosidade e Felicidade



Histograma Generosidade



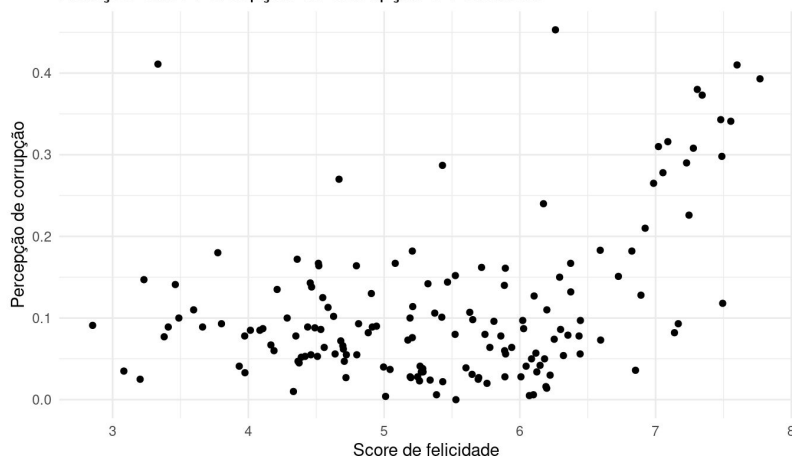
## O país com o maior generosidade é: Myanmar - Generosidade: 0.566

## O país com o menor generosidade é: Greece - Generosidade: 0

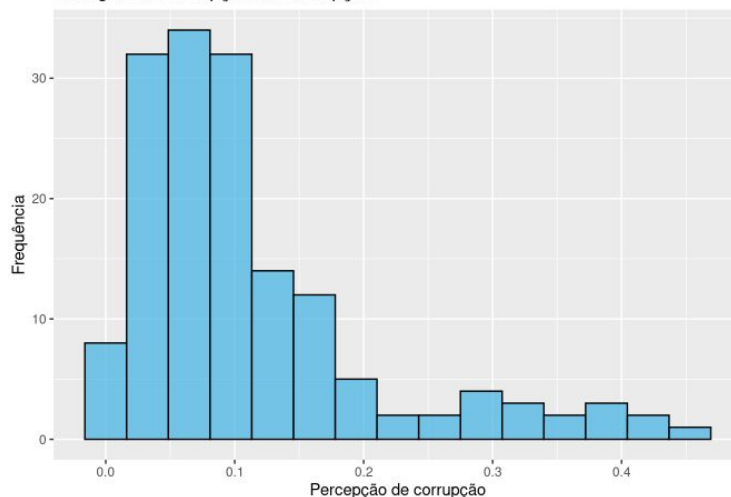
# PERCEPÇÃO DE CORRUPÇÃO

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0000	0.0470	0.0855	0.1106	0.1412	0.4530

Relação entre Percepção de corrupção e Felicidade



Histograma Percepção de corrupção



## O país com o maior percepção de corrupção é: Singapore - Percepção de corrupção: 0.453

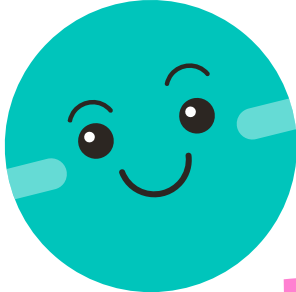
## O país com o menor percepção de corrupção é: Moldova - Percepção de corrupção: 0

# MODELO

$$\text{Score} = \beta_0 + \beta_1(\text{GDP per capita}) + \beta_2(\text{Social support}) + \beta_3(\text{Healthy life expectancy}) + \beta_4(\text{Freedom to make life choices}) + \beta_5(\text{Generosity}) + \beta_6(\text{Perceptions of corruption}) + \varepsilon$$



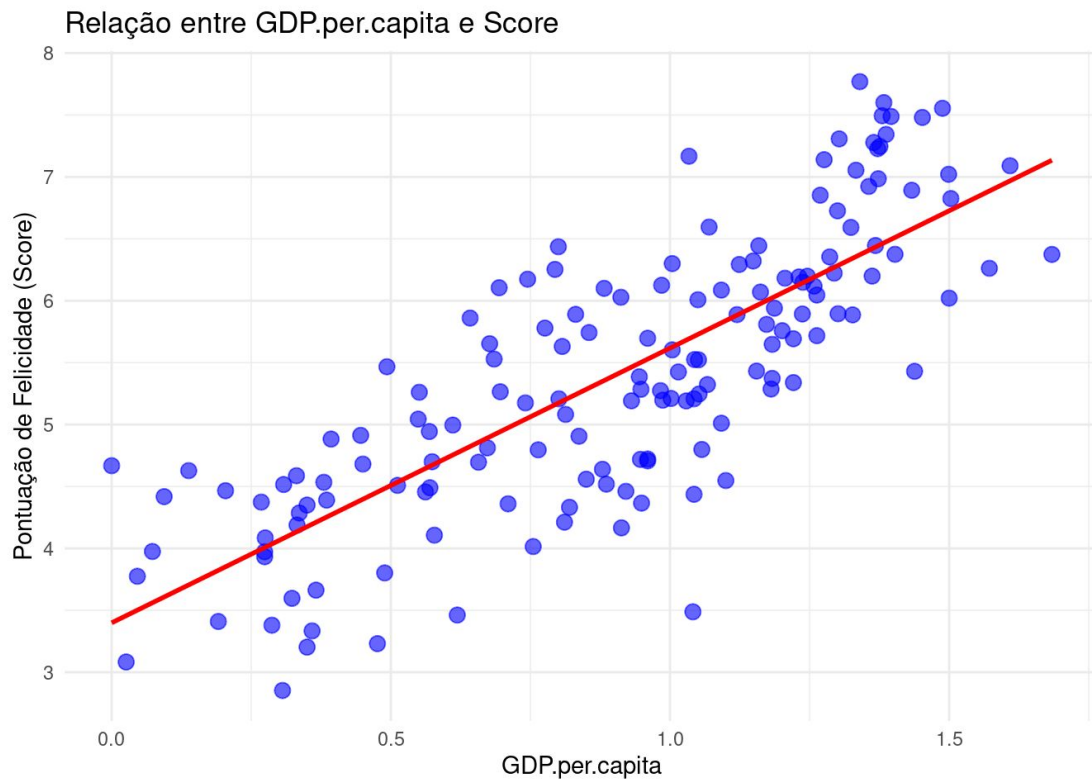
$$\begin{aligned} \text{Score} = & 1.7052 + 0.7754 * \text{GPD.per.capita} + 1.1242 * \text{Social.support} + \\ & 1.0781 * \text{Healthy.life.expectancy} + 1.4584 * \\ & \text{Freedom.to.make.life.choices} + 0.4898 * \text{Generosity} + 0.9723 * \\ & \text{Perceptions.of.corruption} + \varepsilon \end{aligned}$$



# RETAS DE REGRESSÃO INDIVIDUAIS

Relação individual das variáveis explicativas(GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity e Perceptions of corruption) com a variável resposta(Score)

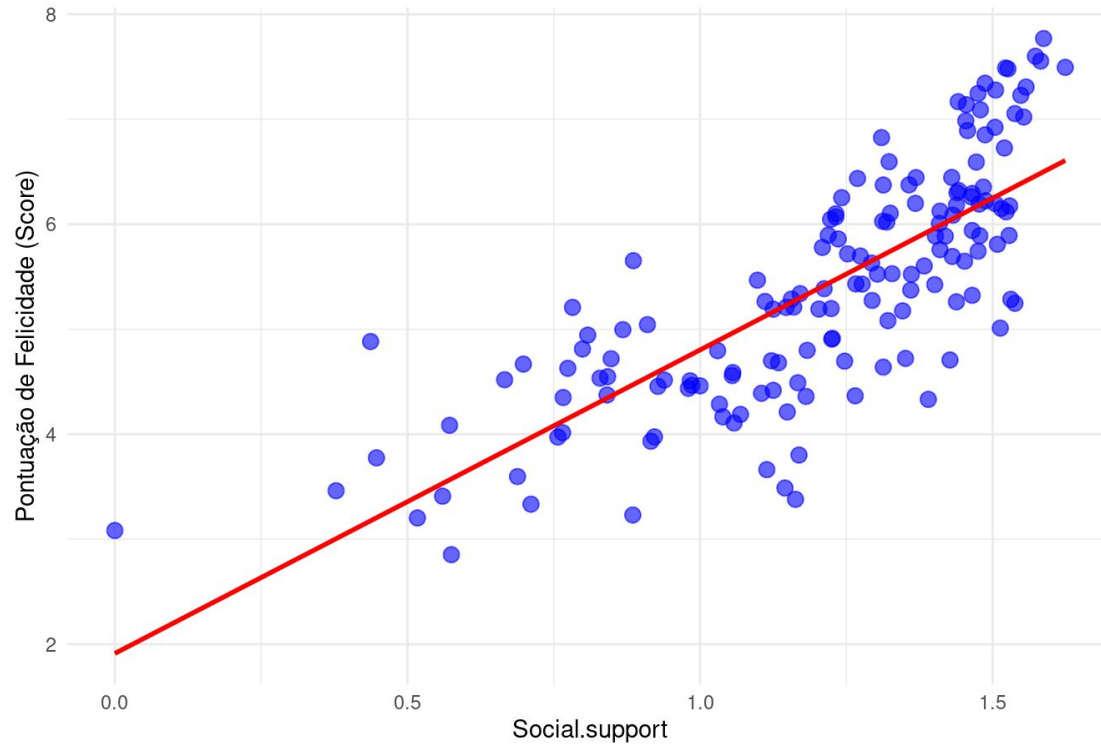
# GPD PER CAPITA E SCORE





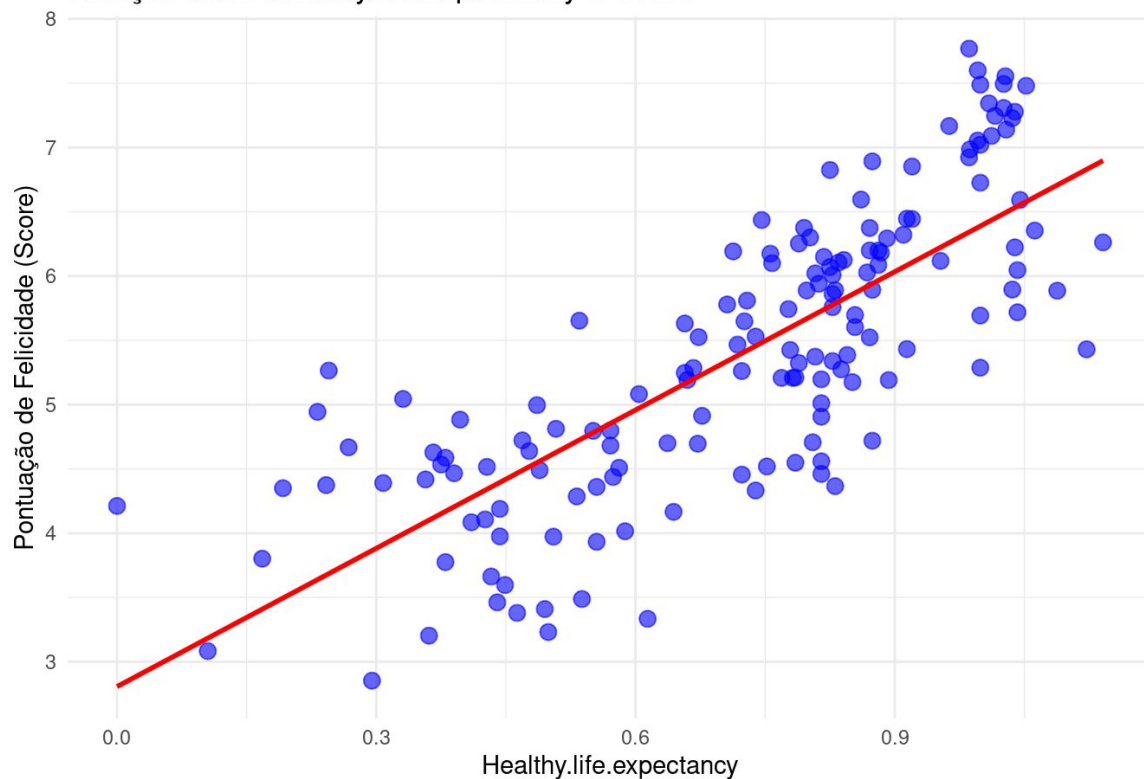
# SOCIAL SUPPORT SCORE

Relação entre Social.support e Score



# HEALTHY LIFE EXPECTANCY E SCORE

Relação entre Healthy.life.expectancy e Score

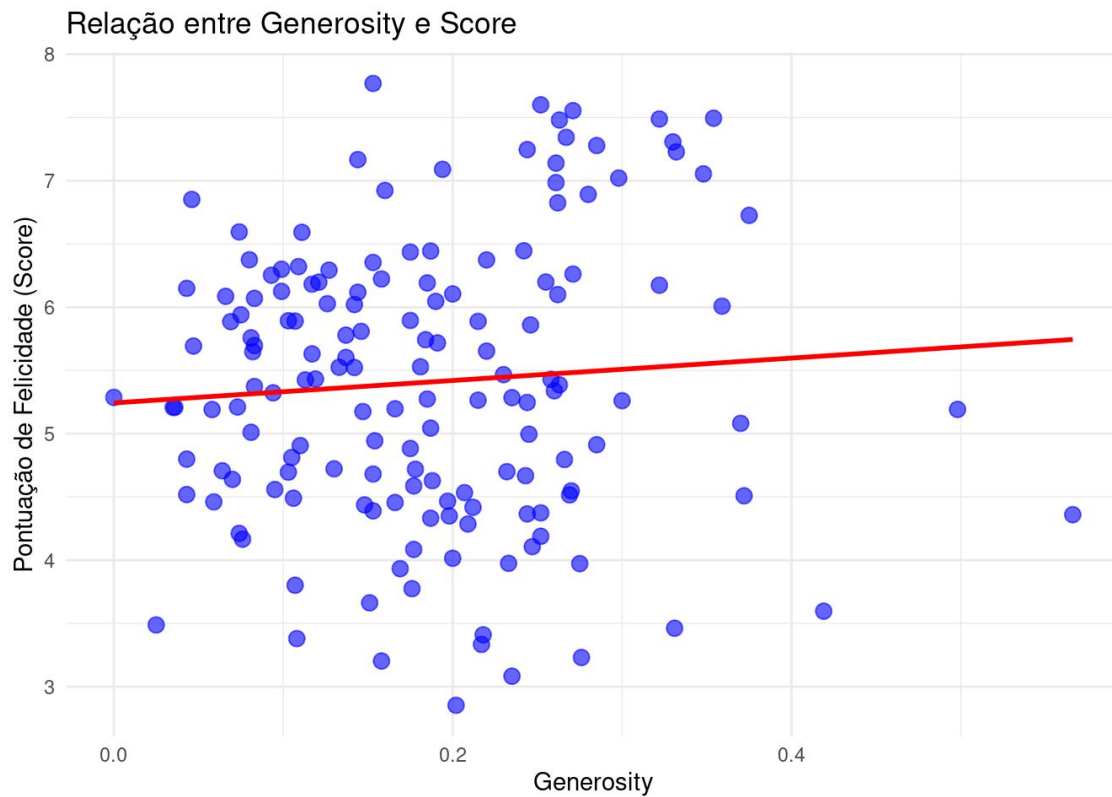


# FREEDOM TO MAKE LIFE CHOICES E SCORE

Relação entre Freedom.to.make.life.choices e Score

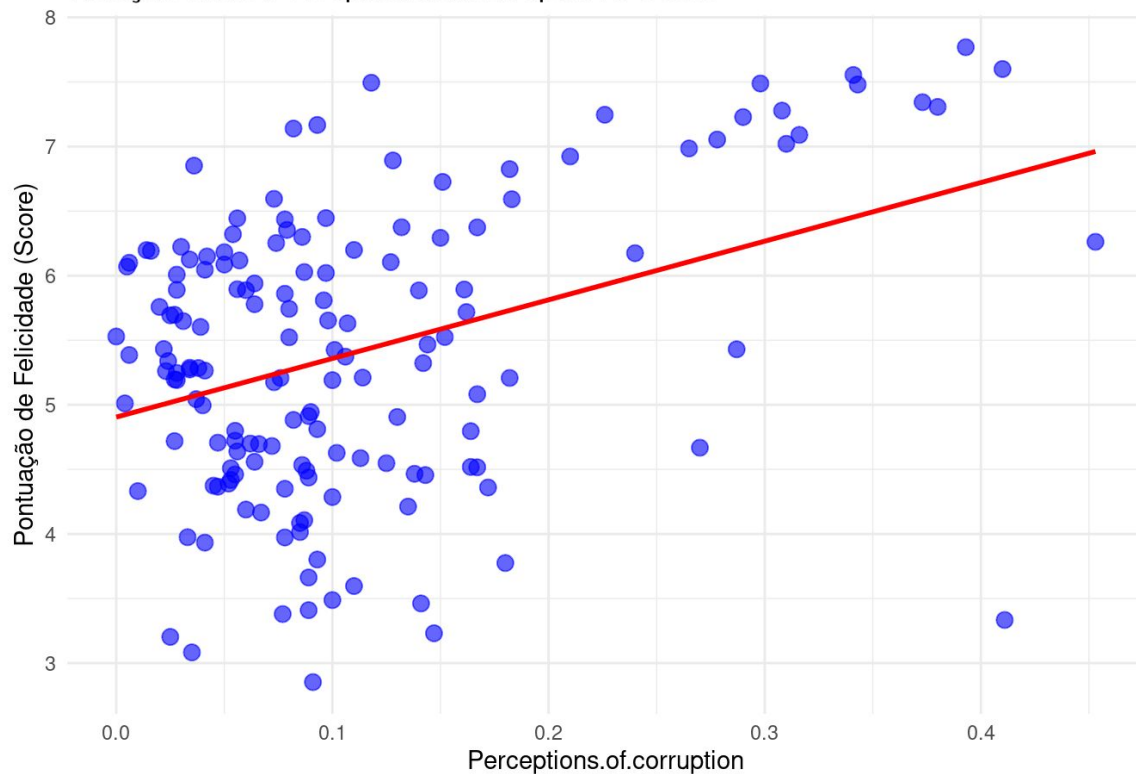


# GENEROSITY E SCORE



# PERCEPTIONS OF CORRUPTION SCORE

Relação entre Perceptions.of.corruption e Score





# INTERPRETAÇÃO DOS COEFICIENTES

## Intercepto ( $B_0 = 1.7952$ ):

- Não tem uma interpretação prática, pois valores zero não fazem sentido neste contexto.

## GDP per capita ( $B_1 = 0.7754$ ):

- Cada aumento de 1 unidade no GDP.per.capita está associado a um aumento médio de 0.7754 unidades em Score, mantendo as outras variáveis constantes.

## Social support ( $B_2 = 1.1242$ ):

- Cada aumento de 1 unidade no Social.support está associado a um aumento médio de 1.1242 unidades em Score, mantendo as outras variáveis constantes.

### Healthy life expectancy ( $B_3 = 1.0782$ ):

- Cada aumento de 1 unidade na expectativa de vida está associado a um aumento médio de 1.0781 unidades em Score, mantendo as outras variáveis constantes.

### Freedom to make life choices ( $B_4 = 1.4548$ ):

- Cada aumento de 1 unidade na liberdade para tomar decisões está associado a um aumento médio de 1.4548 unidades em Score, mantendo as outras variáveis constantes.

### Generosity ( $B_5 = 0.4898$ ):

- Cada aumento de 1 unidade na generosidade está associado a um aumento médio de 0.4898 unidades em Score, mantendo as outras variáveis constantes.

### Perceptions of corruption ( $B_6 = 0.9723$ ):

- Cada aumento de 1 unidade na percepção de corrupção está associado a um aumento médio de 0.9723 unidades em Score, mantendo as outras variáveis constantes.

# INTERPRETAÇÃO DO COEFICIENTE DE DETERMINAÇÃO $R^2$

## Coeficiente de Determinação Ajustado ( $R^2$ ajustado=0.7703)

- O  $R^2$  ajustado indica que 77,03% da variação total na pontuação de felicidade (Score) é explicada pelas variáveis independentes no modelo de regressão.
- O  $R^2$  ajustado leva em conta o número de variáveis independentes no modelo e penaliza a inclusão de variáveis que não contribuem significativamente para explicar a variável resposta.
- Aqui, o  $R^2$  ajustado próximo de 1 significa que o modelo tem um bom ajuste, indicando que o modelo é eficiente e as variáveis escolhidas têm relevância, mas algumas podem ter contribuição menor.



# SIGNIFICÂNCIA DOS ESTIMADORES (ANOVA)

## Tabela de Análise de Variância

Resposta: Pontuação	Graus de liberdade	Soma dos Quadrados	Média dos Quadrados	Valor F	Pr(>F)
GDP.per.capita	1	121.040	121.040	425.2349	< 2.2e-16
Social.support	1	14.108	14.108	49.5625	6.558e-08
Healthy.life.expectancy	1	4.344	4.344	15.2601	0.0001417
Freedom.to.make.life.choices	1	8.567	8.567	30.0969	0,0001722
Generosity	1	0.666	0.666	2.3401	0.1281989
Perceptions.of.corruption	1	0.915	0.915	3.2137	0.0750526
Residuals	149	42.412	0.285		

Conclusões Baseadas no p-valor de 0.05:

### ALTAMENTE SIGNIFICANTE

**GDP per capita ( $p < 2.2e-16$ )**

**Social support ( $p = 6.558e-11$ )**

**Healthy life expectancy ( $p = 1.417e-7$ )**

### SIGNIFICANTE

**Freedom to make life choices ( $p = 1.722e-4$ )**

### NÃO SIGNIFICANTE

**Generosity ( $p = 0.128$ )**

**Perceptions of corruption ( $p=0.07$ )**

**Perceptions of corruption:** Embora a percepção de corrupção possa ter algum impacto, neste modelo, ela não é estatisticamente relevante para explicar a pontuação de felicidade, não tem um impacto claro no Score.

# MULTICOLINEARIDADE

## Matriz de Variância e Covariância

	(Intercept)	GDP.per.capita	Social.support	Healthy.life.expectancy	Freedom.to.make.life.choices	Generosity	Perceptions.of.corruption
(Intercept)	0.044552	0.013234	-0.028946	-0.014560	-0.007298	-0.037256	0.005391
GDP.per.capita	0.013234	0.047622	-0.020714	-0.045006	0.002025	0.015134	-0.020354
Social.support	-0.028946	-0.020714	0.056122	-0.018059	-0.026122	0.003116	0.023793
Healthy.life.expectancy	-0.014560	-0.045006	-0.018059	0.111916	-0.005195	-0.002971	-0.013116
Freedom.to.make.life.choices	-0.007298	0.002025	-0.026122	-0.005195	0.140878	-0.040051	-0.064111
Generosity	-0.037256	0.015134	0.003116	-0.002971	-0.040051	0.247751	-0.073484
Perceptions.of.corruption	0.005391	-0.020354	0.023793	-0.013116	-0.064111	-0.073484	0.294155

## Valores de VIF para Diagnóstico de Multicolinearidade

GDP.per.capita	4.115838
Social.support	2.735651
Healthy.life.expectancy	3.572728
Freedom.to.make.life.choices	1.575090
Generosity	1.224101
Perceptions.of.corruption	1.431594

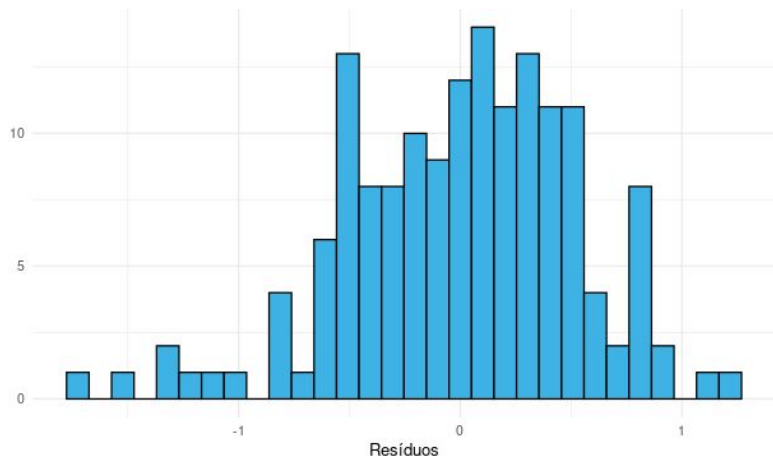
### Conclusão sobre a multicolinearidade:

- Com os resultados apresentados, evidencia-se que não há multicolinearidade entre as variáveis. Nenhuma das variáveis independentes apresentou  $VIF > 10$ , o que indica que a multicolinearidade no modelo é moderada e aceitável.

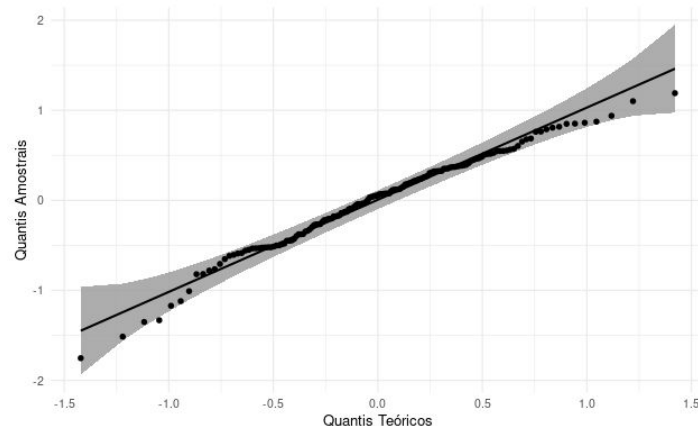
# ANÁLISE DE RESÍDUOS

## Normalidade

### Histograma



### QQ plot



## Testes de normalidade

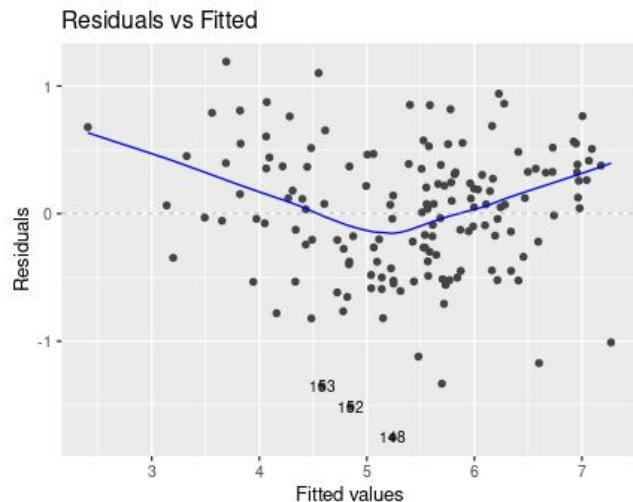
Teste	Statistic	P.value
Shapiro-Wilk	0.9818012	0.0374733
Kolmogorov-Smirnov	0.1850644	0.0000457

**Conclusão:** Através do histograma podemos notar um comportamento não normal dos resíduos. Pelo qq plot, é possível avaliar a não normalidade dos dados, principalmente nas caudas. E os testes de Shapiro-Wilk e Kolmogorov-Smirnov rejeitaram a hipótese nula de que os dados sejam normalmente distribuídos. Logo, os resíduos desse modelo não seguem distribuição normal.

# ANÁLISE DE RESÍDUOS

## Homocedasticidade

### Gráfico dos Resíduo x Valores ajustados

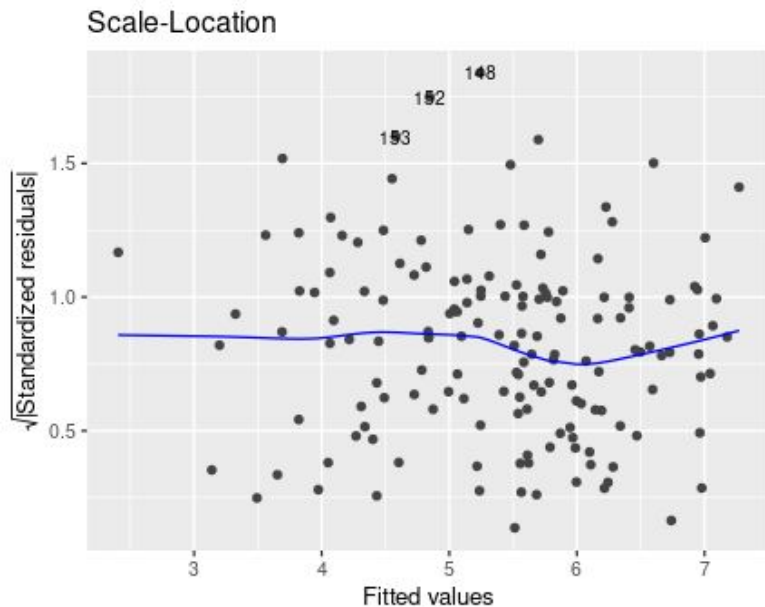


### Teste de homocedasticidade

Teste	Statistic	P.value
Breusch-Pagan	21.99838	0.0012117

**Conclusão:** Pelo gráfico dos resíduo x valores ajustados, os resíduos parecem se comportar de forma aleatória. Porém, o teste de Breusch-Pagan rejeitou a hipótese nula de homoscedasticidade. Logo os resíduos não apresentam variância constante.

# OUTLIERS



Quantidade de outliers identificados

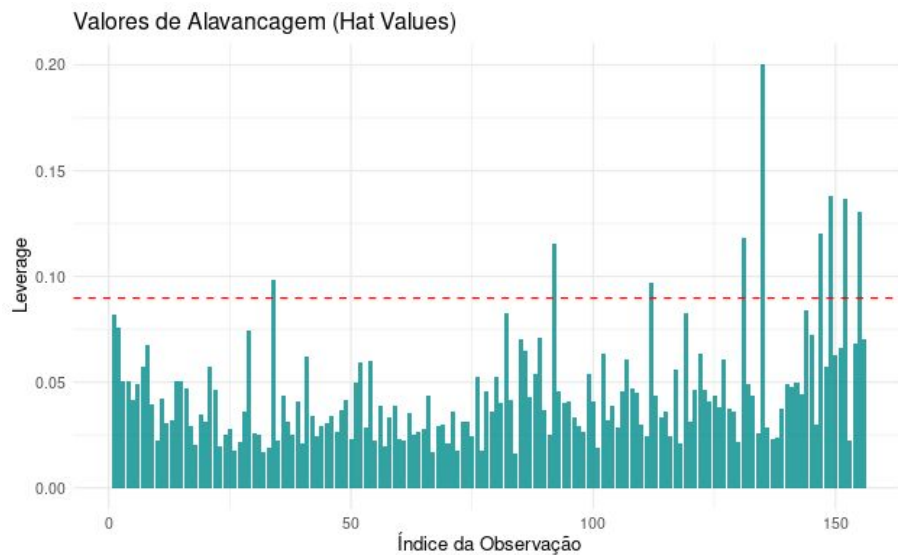
2

Pontos outliers

148 e 152

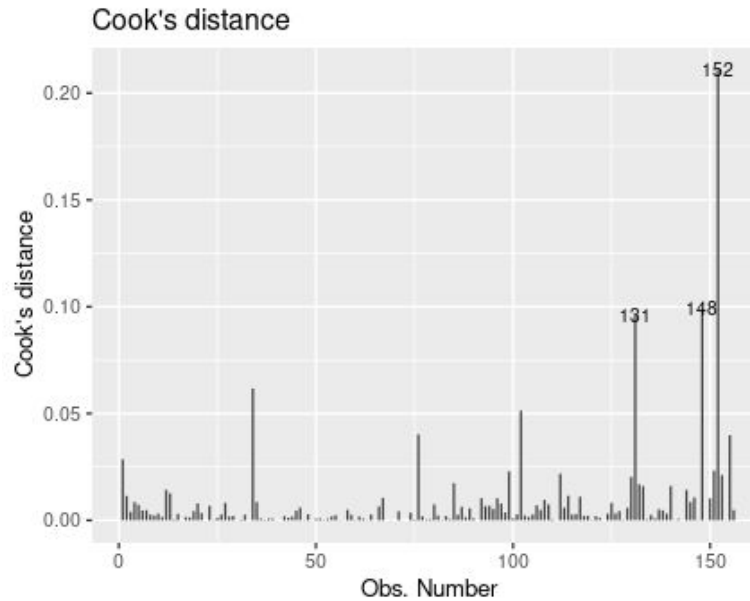
**Conclusão:** É possível perceber que há dois outliers, as observações 148 e 152. Essas observações desviam significativamente do padrão esperado do modelo, e nesse caso, são os países Botswana e Rwanda, ambos países da África. Além disso, com base no gráfico, também é possível observar que esses países estão mais afastados da reta de regressão esperada.

# ALAVANCAS



Percebemos que há 9 observações que excedem o limite como indicador de alavancagem, as quais são: **34, 92, 112, 131, 135, 147, 149, 152 e 155.**

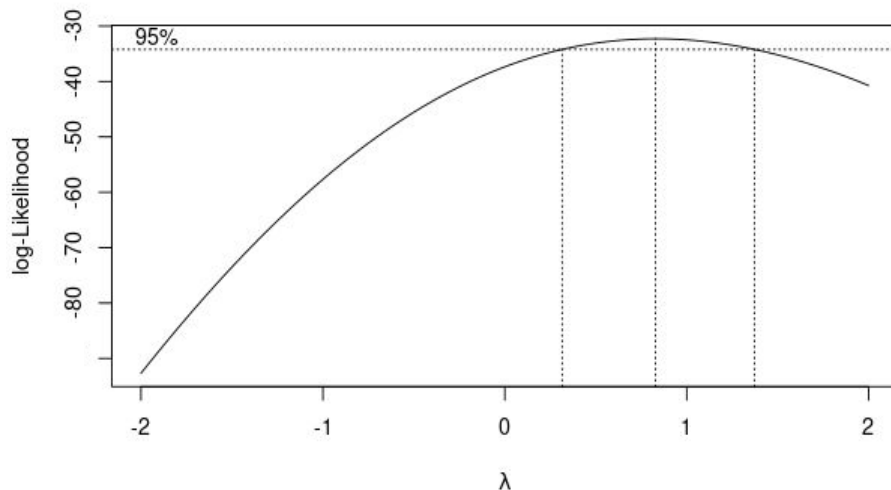
# PONTOS INFLUENTES



Com base nos cálculos e no gráfico, há 8 pontos influentes, os quais são as observações:  
**1, 34, 76, 102, 131, 148, 152, 155.**

# TRANSFORMAÇÃO

## Gráfico Box-Cox



Com base na análise de Box-Cox, o parâmetro  $\lambda$  que maximiza o log foi estimado como  $\lambda \approx 1$ , com um intervalo de confiança de 95% que também inclui este valor. Isso significa que a transformação ideal para os dados seria praticamente equivalente a manter a variável dependente ( $Y$ ) em sua escala original.

Dessa forma, não é necessário aplicar nenhuma transformação significativa aos dados, pois o método indica que o modelo ajustado com os valores originais já é adequado. Assim, podemos prosseguir com a análise sem realizar alterações na escala da variável dependente.



# CONCLUSÃO

Com base nos resultados obtidos, conclui-se que o modelo de regressão linear apresentou um bom ajuste, explicando aproximadamente 77,92% da variação na pontuação de felicidade (Score) dos países analisados.

Entre as variáveis explicativas, PIB per capita, suporte social, expectativa de vida e liberdade para tomar decisões destacaram-se como fatores significativos e positivamente associados à felicidade, reforçando a importância de investimentos econômicos e sociais para a melhoria do bem-estar das populações.

Por outro lado, generosidade e percepção de corrupção não apresentaram significância estatística no modelo, sugerindo que esses fatores, embora relevantes no contexto geral, não desempenham papel determinante na explicação do Score nesse conjunto de dados.

Esses resultados oferecem informações valiosas para governos e formuladores de políticas públicas sobre os principais determinantes da felicidade em diferentes países.

**OBRIGADO!**

