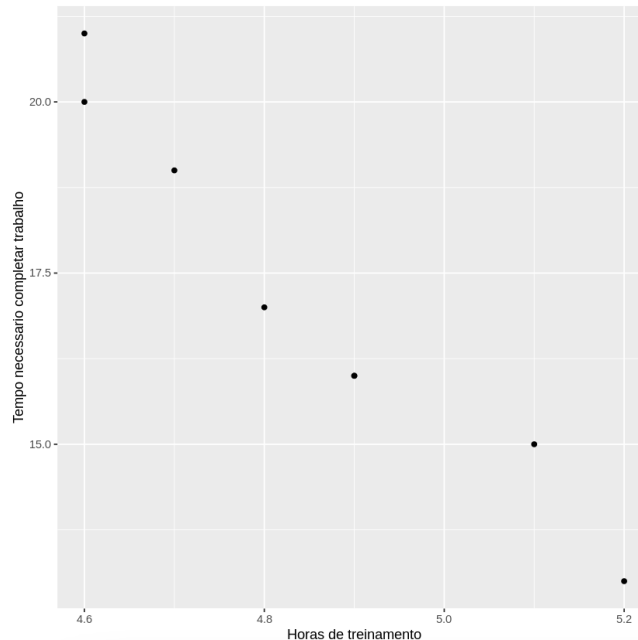


Trilha 4

Bruna Matos

Problema 1

a) Um gráfico de dispersão para os dados



b) Determine o modelo de regressão linear simples entre as variáveis x e y, sendo y a variável resposta.

```
Call:
lm(formula = y ~ x, data = dados)

Coefficients:
(Intercept)          x
       74.90        -11.91
```

c) Faça uma análise do modelo de regressão utilizando a função summary:

i) resíduos, significância estatística dos coeficientes, percentual de variância explicada pelo modelo

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.72059 -0.52941 -0.02941  0.27941  0.89706

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   74.897     5.514   13.58 9.88e-06 ***
x            -11.912     1.136  -10.49 4.42e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6624 on 6 degrees of freedom
Multiple R-squared:  0.9483,    Adjusted R-squared:  0.9396
F-statistic: 110 on 1 and 6 DF, p-value: 4.416e-05
A anova: 2 x 5
```

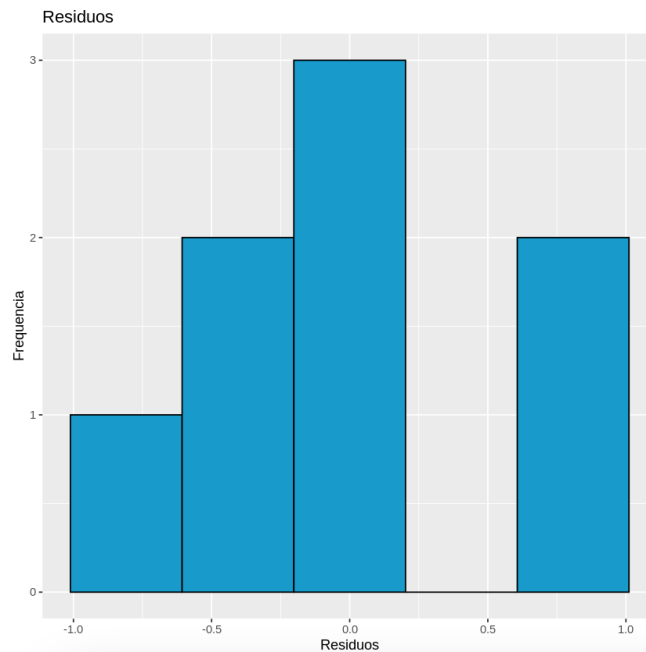
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
x	1	48.242647	48.2426471	109.9609	4.415513e-05
Residuals	6	2.632353	0.4387255	NA	NA

Analisando os dados podemos observar que:

I) o valor de Pr para as variáveis x e y são menores que 5%, o que significa que elas são relevantes para o sistema.

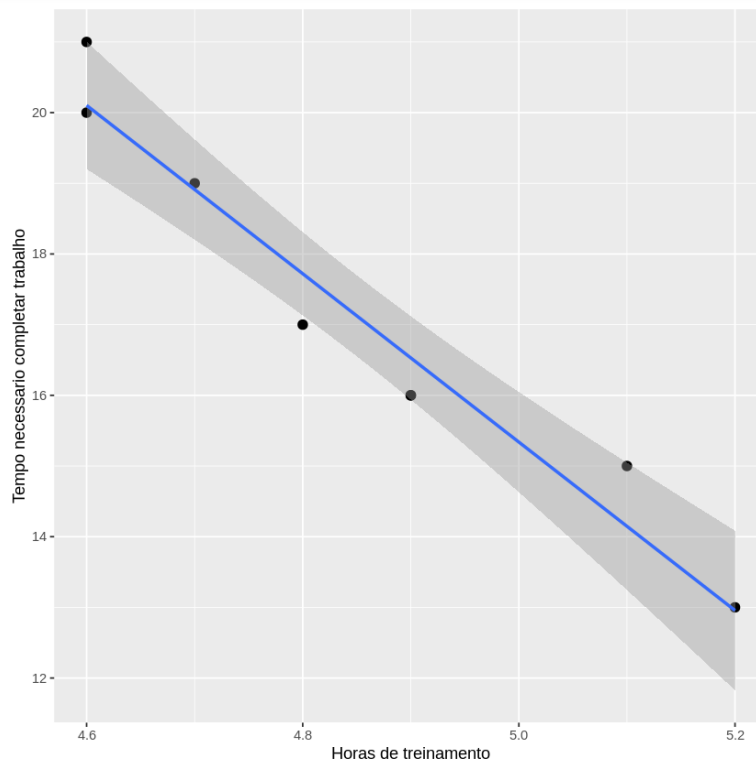
II) o valor de p-value é menor que 5%, o que indica também que a reta pode representar os dados

III) O valor de R-squared = 0.95, o que indica que 95% dos dados podem ser representados pela reta.



Ainda, analisando o histograma com os valores dos resíduos, podemos comprovar que eles seguem uma distribuição normal indicando que o modelo é coerente.

d) Trace, no gráfico anterior, a reta de regressão



Problema 2

a) Faça a importação dos dados, verifique a estrutura e faça um sumário estatístico.

A importação foi feita e utilizando os comando tail e head, pudemos perceber que para as variáveis ordered, elas estavam sendo lidas como factor. Com a remoção do parâmetro stringsAsFactor as variáveis estavam sendo lidas como Character.

Foi então que usando o comando transform modificamos as colunas para o tipo Ordered e pudemos fazer as nossas análises.

A data.frame: 6 × 6

	pais	ano	pop	continente	expVida	pibPercap
	<ord>	<int>	<dbl>	<ord>	<dbl>	<dbl>
1	Afghanistan	1952	8425333	Asia	28.801	779.4453
2	Afghanistan	1957	9240934	Asia	30.332	820.8530
3	Afghanistan	1962	10267083	Asia	31.997	853.1007
4	Afghanistan	1967	11537966	Asia	34.020	836.1971
5	Afghanistan	1972	13079460	Asia	36.088	739.9811
6	Afghanistan	1977	14880372	Asia	38.438	786.1134

A data.frame: 6 × 6

	pais	ano	pop	continente	expVida	pibPercap
	<ord>	<int>	<dbl>	<ord>	<dbl>	<dbl>
1699	Zimbabwe	1982	7636524	Africa	60.363	788.8550
1700	Zimbabwe	1987	9216418	Africa	62.351	706.1573
1701	Zimbabwe	1992	10704340	Africa	60.377	693.4208
1702	Zimbabwe	1997	11404948	Africa	46.809	792.4500
1703	Zimbabwe	2002	11926563	Africa	39.989	672.0386
1704	Zimbabwe	2007	12311143	Africa	43.487	469.7093

Na sequência, usamos o comando summary para ter uma overview das variáveis e entender um pouco dos tipos de dados e suas ordens de grandeza. Considerando os valores máximo e min e também as opções disponíveis, aparentemente não temos problemas nos dados.

	pais	ano	pop	continente
Afghanistan:	12	Min. :1952	Min. :6.001e+04	Africa :624
Albania :	12	1st Qu.:1966	1st Qu.:2.794e+06	Americas:300
Algeria :	12	Median :1980	Median :7.024e+06	Asia :396
Angola :	12	Mean :1980	Mean :2.960e+07	Europe :360
Argentina :	12	3rd Qu.:1993	3rd Qu.:1.959e+07	Oceania : 24
Australia :	12	Max. :2007	Max. :1.319e+09	
(Other) :	1632			
	expVida	pibPercap		
Min. :	23.60	Min. : 241.2		
1st Qu.:	48.20	1st Qu.: 1202.1		
Median :	60.71	Median : 3531.8		
Mean :	59.47	Mean : 7215.3		
3rd Qu.:	70.85	3rd Qu.: 9325.5		
Max. :	82.60	Max. :113523.1		

Na sequência, foi possível montar o sumário estatístico conforme apresentado na tabela abaixo:

Nome da Variável	Descrição da Variável	Tipo da Variável	Tipo de Mensuração	Valores possíveis da variável
pais	País analisado	Qualitativo	Nominal	São 142 levels "Afghanistan",...
ano	Ano do calculo	Quantitativo	Ordinal	De 1952 - 2007
pop	população	Quantitativo	Razão	De 6.001e+4 - 1.319e+9
continente	qual o continente que o país está	Qualitativo	Nominal	Africa, Americas, Asia, Europe, Oceania
expVida	a expectativa de vida	Quantitativo	Razão	23.6 - 82.60
pibPercap	PIB per capta	Quantitativo	Razão	241.2 - 113523.1

b) Verifique a estrutura dos dados (str)

Usando o comando str obtivemos o seguinte resultado:

```
'data.frame': 1704 obs. of 6 variables:
 $ pais      : Ord.factor w/ 142 levels "Afghanistan"<...: 1 1 1 1 1 1 1 1 1 1 ...
 $ ano       : int  1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...
 $ pop       : num  8425333 9240934 10267083 11537966 13079460 ...
 $ continente: Ord.factor w/ 5 levels "Africa"<"Americas"<...: 3 3 3 3 3 3 3 3 3 3 ...
 $ expVida   : num  28.8 30.3 32 34 36.1 ...
 $ pibPercap : num  779 821 853 836 740 ...
```

Ele é importante para nos certificarmos quais os tipos das variáveis e ainda, para as variáveis do tipo Factor, podemos saber quantas opções diferentes temos.

c) Classifique cada variável de acordo com seu tipo (qualitativa ordinal, nominal, quantitativa discreta, contínua, etc).

Essa classificação está disponível na tabela de sumário estatístico do item A nas colunas "Tipo de variável" e "Tipo de mensuração".

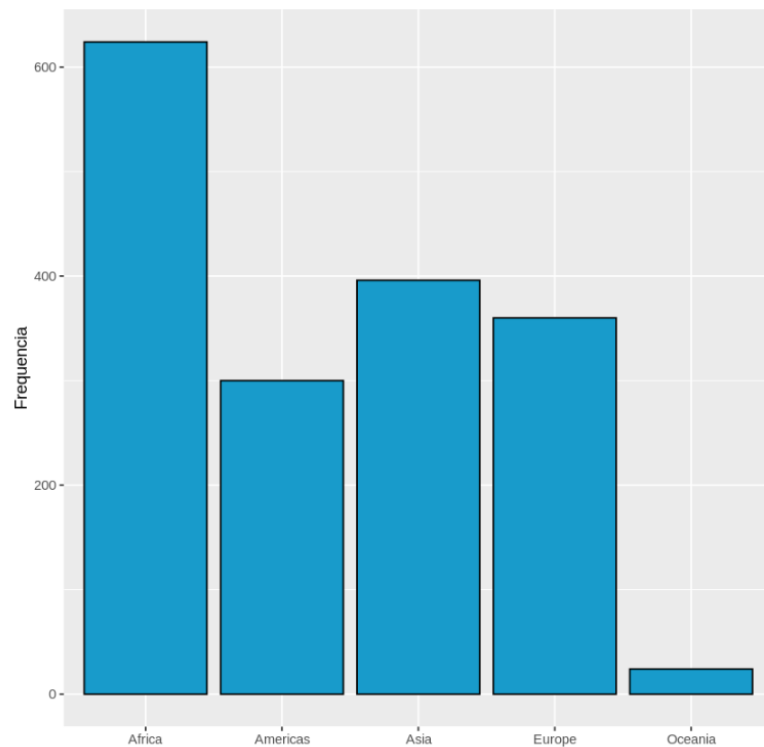
d) Faça um sumário estatístico dos dados

Já foi feito no item A

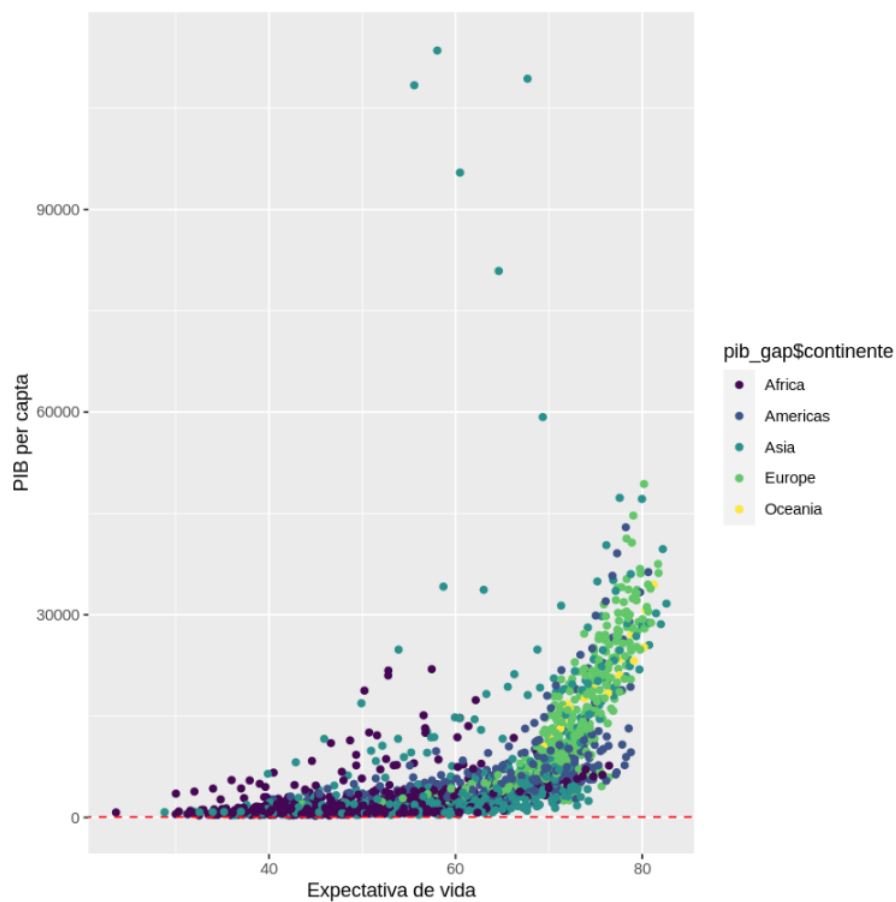
e) Faça uma tabela de frequência absoluta e uma tabela de frequência relativa para verificar o número de observações por continente.

Continentes			Continentes		
Frequencia_Absoluta			Frequencia_Relativa		
<ord>			<ord>		
<int>			<dbl>		
1	Africa	624	1	Africa	0.36619718
2	Americas	300	2	Americas	0.17605634
3	Asia	396	3	Asia	0.23239437
4	Europe	360	4	Europe	0.21126761
5	Oceania	24	5	Oceania	0.01408451

f) Faça um gráfico de barras da tabela de frequência absoluta dos continentes



g) Faça um gráfico apropriado para relacionar o PIB per capita à expectativa de vida.



Pudemos observar que a expectativa de vida e o PIB são maiores na Europa e Asia. Tbm percebemos uma curva exponencial.

h) Crie duas novas colunas nesta base de dados com o logaritmo de PIB per capita, e o logaritmo da expectativa de vida. Estas colunas devem ter os nomes: `lpibPercap` e `lexpVida`, respectivamente.

As colunas foram adicionadas conforme código abaixo

```
# log das colunas
lpibPercap = log(pib_gap$piibPercap)
lexpVida = log(pib_gap$expVida)

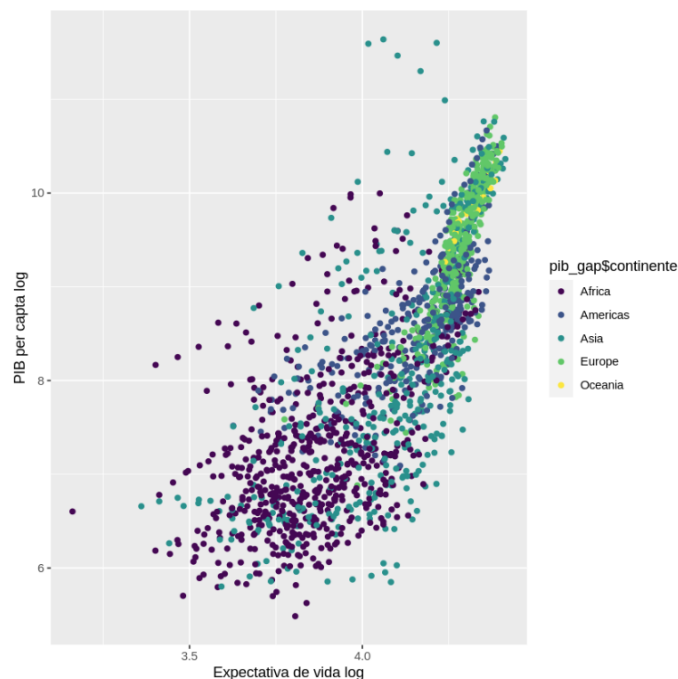
pib_gap <- mutate(pib_gap, lpibPercap=lpibPercap)
pib_gap <- mutate(pib_gap, lexpVida=lexpVida)

head(pib_gap)
```

A data.frame: 6 × 8

	<code>pais</code>	<code>ano</code>	<code>pop</code>	<code>continente</code>	<code>expVida</code>	<code>piibPercap</code>	<code>lpibPercap</code>	<code>lexpVida</code>
	<code><ord></code>	<code><int></code>	<code><dbl></code>	<code><ord></code>	<code><dbl></code>	<code><dbl></code>	<code><dbl></code>	<code><dbl></code>
1	Afghanistan	1952	8425333	Asia	28.801	779.4453	6.658583	3.360410
2	Afghanistan	1957	9240934	Asia	30.332	820.8530	6.710344	3.412203
3	Afghanistan	1962	10267083	Asia	31.997	853.1007	6.748878	3.465642
4	Afghanistan	1967	11537966	Asia	34.020	836.1971	6.728864	3.526949
5	Afghanistan	1972	13079460	Asia	36.088	739.9811	6.606625	3.585960
6	Afghanistan	1977	14880372	Asia	38.438	786.1134	6.667101	3.649047

i) Faça um gráfico apropriado para relacionar estas duas novas variáveis.



Com o gráfico de dispersão podemos observar que a curva que antes tinha uma aparência exponencial agora se aproxima de uma reta.

j) Ajuste um modelo linear aos dados, utilizando as duas novas variáveis criadas, sendo lexpVida a variável resposta

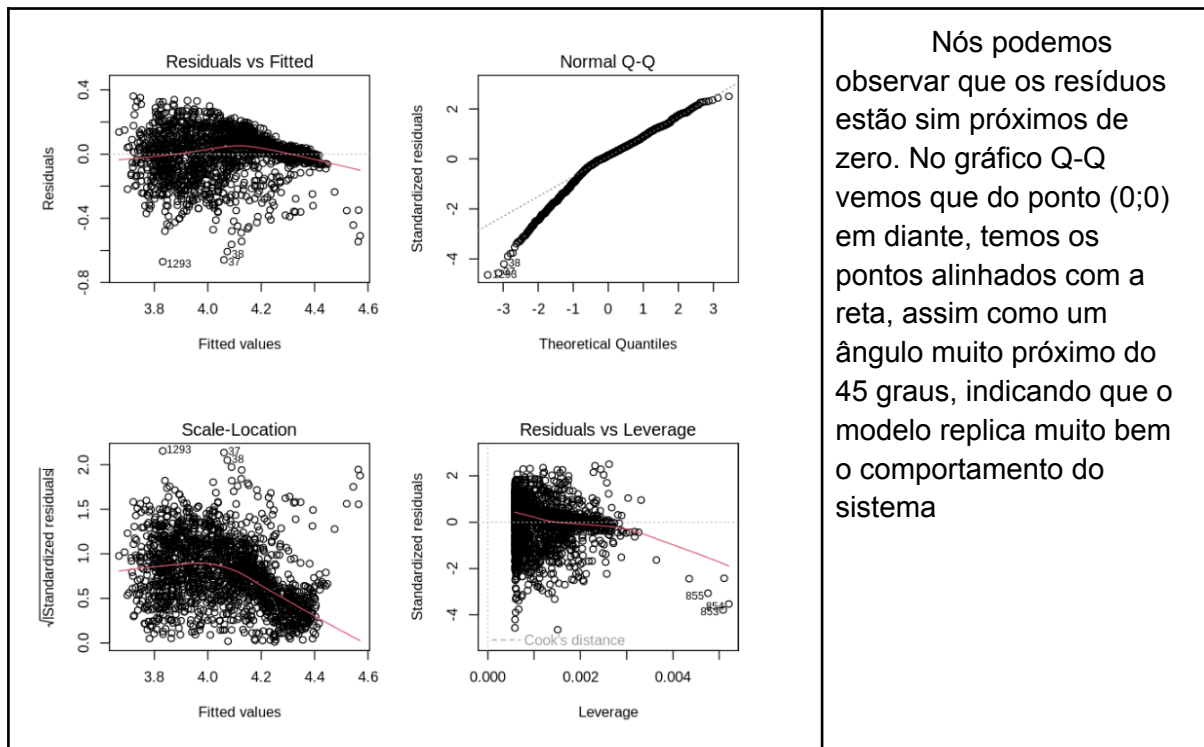
O modelo linear obtido foi:

```
Call:
lm(formula = lexpVida ~ lpibPercap, data = pib_gap)

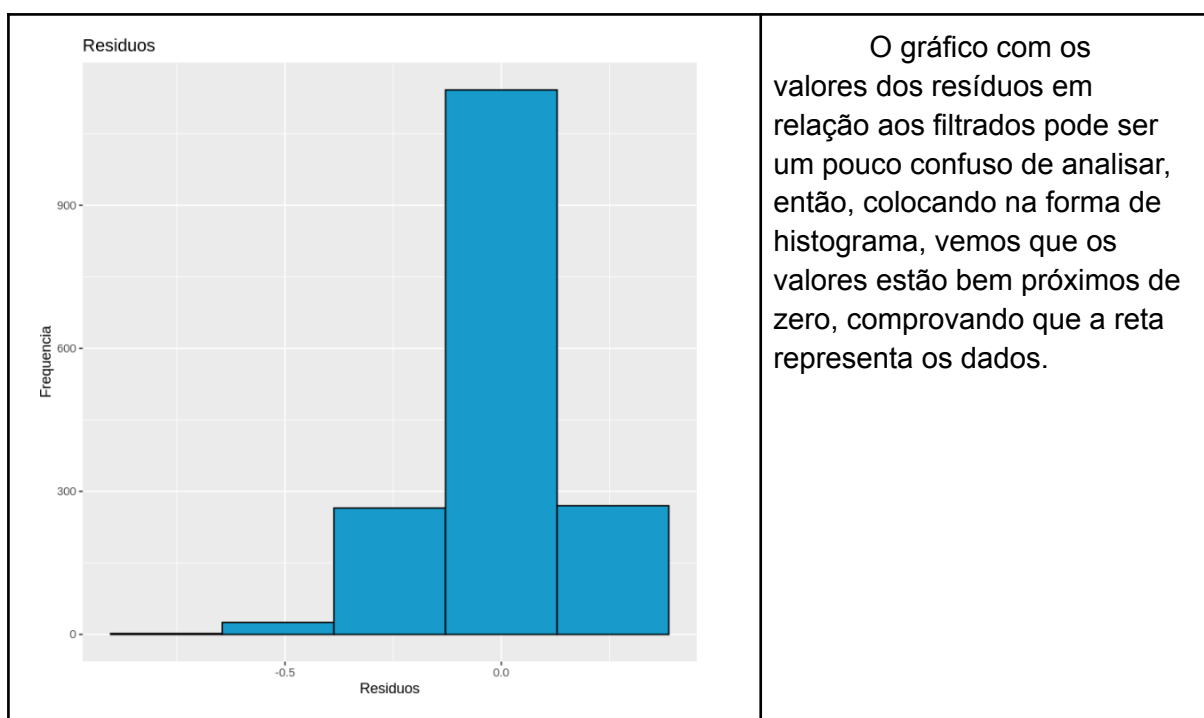
Coefficients:
(Intercept)      lpibPercap 
    2.8642         0.1465
```

k) Faça todas as análises da regressão, julgando:

i) Os gráficos diagnósticos



Nós podemos observar que os resíduos estão sim próximos de zero. No gráfico Q-Q vemos que do ponto (0;0) em diante, temos os pontos alinhados com a reta, assim como um ângulo muito próximo do 45 graus, indicando que o modelo replica muito bem o comportamento do sistema



O gráfico com os valores dos resíduos em relação aos filtrados pode ser um pouco confuso de analisar, então, colocando na forma de histograma, vemos que os valores estão bem próximos de zero, comprovando que a reta representa os dados.

ii) Os parâmetros obtidos (avaliar o summary do modelo)

```

Residuals:
    Min       1Q   Median       3Q      Max
-0.67059 -0.06453  0.01978  0.09086  0.36156

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.864177   0.023283  123.02  <2e-16 ***
lpibPercap   0.146549   0.002821   51.95  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1445 on 1702 degrees of freedom
Multiple R-squared:  0.6132,    Adjusted R-squared:  0.613
F-statistic: 2698 on 1 and 1702 DF,  p-value: < 2.2e-16
      Anova: 2 x 5

      Df  Sum Sq  Mean Sq  F value  Pr(>F)
<int>  <dbl>    <dbl>    <dbl>    <dbl>
lpibPercap      1  56.34280  56.34280459  2698.233      0
Residuals    1702  35.54009   0.02088137      NA      NA

```

Considerando os parâmetros obtidos, podemos observar que Pr de cada variável está bem abaixo de 5%, o que indica que ambas são relevantes para o modelo. O p-value do modelo também está bem abaixo de 5%, indicando que a reta representa sim os dados do sistema. o que difere seria o fator R-squared que foi de apenas 61.32%. Isso se deve ao fato de termos resíduos tanto para cima quanto para baixo.

iii) O poder de explicação do modelo.

Conforme mencionado diversas vezes anteriormente, o modelo possui uma alta taxa de explicação do modelo.

Problema 3

a) Criar um dicionário de dados, para entender o significado o tipo de cada variável; veja no link fornecido se há documentação disponível.

Nome da Variável	Descrição da Variável	Tipo da Variável	Tipo de Mensuração	Valores possíveis da variável
make	Contrutora	Qualitativo	Nominal	21 levels "alfa-romero"
fuel.type	Tipo de combustivel	Qualitativo	Nominal	diesel,gas
aspiration	Se possui turbo	Qualitativo	Nominal	std,turbo
num.doors	numero de portas	Quantitativo	Nominal	2,4
body.style	tipo da carcaça	Qualitativo	Nominal	5 levels "convertible"....
drive.wheels	tipo de tração	Qualitativo	Nominal	4wd,fwd,rwd
engine.location	localização do motor	Qualitativo	Nominal	front,rear
wheel.base	é a distância horizontal entre os centros das rodas dianteiras e traseiras	Quantitativo	Razão	93 - 1209
lenght	tamanho	Quantitativo	Razão	150 - 2081
width	altura	Quantitativo	Razão	64 - 717
height	peso	Quantitativo	Razão	51 - 598
curb.weight	peso do veiculo considerando tanque cheio	Quantitativo	Razão	1488 - 4066
engine.type	tip de motor	Qualitativo	Nominal	dohc,l,ohc,ohcf,ohcv
num.cylinders	numero de cilindros	Quantitativo	Ordinal	3 - 12
engine.size	tamanho do motor	Quantitativo	Ordinal	61 - 326
fuel.system	sistema de injeção	Qualitativo	Nominal	1bbl,2bbl,idi,mfi,mpfi,spdi,spfi
bore	diametro interno do cilindro	Quantitativo	Razão	35 - 394
stroke	posicao pistao no momento da explosao	Quantitativo	Razão	28 - 417
compression.ratio	media de compressao	Quantitativo	Razão	7 - 941
horsepower	potencia do motor	Quantitativo	Ordinal	48 - 262
peak.rpm	valor de pico da rotacao	Quantitativo	Ordinal	4150 - 6600
city.mpg	pontuação que um carro terá em média em condições de cidade, com paradas e partidas em velocidades mais baixas	Quantitativo	Ordinal	13 - 49
highway.mpg	pontuação que um carro terá em média em condições de rodovias, com paradas e partidas em velocidades mais baixas	Quantitativo	Ordinal	16 - 54
price	preço	Quantitativo	Ordinal	5118 - 45400

b) Carregar a base para o R, certificando-se de que os dados estão corretos, de acordo com o dicionário de dados.

Ao carregar os dados foi possível observar alguns problemas em relação aos tipos. as devidas transformações foram feitas e o resultado pode ser observado nas imagens abaixo:

A data.frame: 6 x 24

	make	fuel.type	aspiration	num.doors	body.style	drive.wheels	engine.location	wheel.base	length	width	engine.size	fuel.system	bore	stroke	compression.ratio	horsepower	peak.rpm	city.mpg	highway.mpg
	<ord>	<ord>	<ord>	<int>	<ord>	<ord>	<ord>	<dbl>	<dbl>	<dbl>	<int>	<ord>	<dbl>	<dbl>	<dbl>	<int>	<int>	<int>	<int>
1	alfa-romero	gas	std	2	convertible	rwd	front	886	1688	641	130	mpfi	347	268	9	111	5000	21	2
2	alfa-romero	gas	std	2	convertible	rwd	front	886	1688	641	130	mpfi	347	268	9	111	5000	21	2
3	alfa-romero	gas	std	2	hatchback	rwd	front	945	1712	655	152	mpfi	268	347	9	154	5000	19	2
4	audi	gas	std	4	sedan	fwd	front	998	1766	662	109	mpfi	319	34	10	102	5500	24	3
5	audi	gas	std	4	sedan	4wd	front	994	1766	664	136	mpfi	319	34	8	115	5500	18	2
6	audi	gas	std	2	sedan	fwd	front	998	1773	663	136	mpfi	319	34	85	110	5500	19	2

A data.frame: 6 x 24

	make	fuel.type	aspiration	num.doors	body.style	drive.wheels	engine.location	wheel.base	length	width	engine.size	fuel.system	bore	stroke	compression.ratio	horsepower	peak.rpm	city.mpg	highway.mpg
	<ord>	<ord>	<ord>	<int>	<ord>	<ord>	<ord>	<dbl>	<dbl>	<dbl>	<int>	<ord>	<dbl>	<dbl>	<dbl>	<int>	<int>	<int>	<int>
188	volvo	gas	turbo	4	wagon	rwd	front	1043	1888	672	130	mpfi	362	315	75	162	5100	17	22
189	volvo	gas	std	4	sedan	rwd	front	1091	1888	689	141	mpfi	378	315	95	114	5400	23	28
190	volvo	gas	turbo	4	sedan	rwd	front	1091	1888	688	141	mpfi	378	315	87	160	5300	19	25
191	volvo	gas	std	4	sedan	rwd	front	1091	1888	689	173	mpfi	358	287	88	134	5500	18	23
192	volvo	diesel	turbo	4	sedan	rwd	front	1091	1888	689	145	idi	301	34	23	106	4800	26	27
193	volvo	gas	turbo	4	sedan	rwd	front	1091	1888	689	141	mpfi	378	315	95	114	5400	19	25

c) Explore a base de dados:

- Sumários estatísticos dos dados
- Tabelas quando apropriado
- Gráficos exploratórios apropriados

make	fuel.type	aspiration	num.doors	body.style
toyota	:32	diesel: 19	std :158	Min. :2.000
nissan	:18	gas :174	turbo: 35	1st Qu.:2.000
honda	:13			Median :4.000
mitsubishi:	13			Mean :3.161
mazda	:12			3rd Qu.:4.000
subaru	:12			Max. :4.000
(Other)	:93			

drive.wheels	engine.location	wheel.base	length	width
4wd: 8	front:190	Min. : 93.0	Min. : 150	Min. : 64.0
fwd:114	rear : 3	1st Qu.: 937.0	1st Qu.:1653	1st Qu.:638.0
rwd: 71		Median : 965.0	Median :1724	Median :654.0
		Mean : 921.4	Mean :1623	Mean :613.4
		3rd Qu.:1012.0	3rd Qu.:1831	3rd Qu.:665.0
		Max. :1209.0	Max. :2081	Max. :717.0

height	curb.weight	engine.type	num.cylinders	engine.size
Min. : 51.0	Min. :1488	dohc: 12	Min. : 3.00	Min. : 61.0
1st Qu.:508.0	1st Qu.:2145	l : 12	1st Qu.: 4.00	1st Qu.: 98.0
Median :541.0	Median :2414	ohc :141	Median : 4.00	Median :120.0
Mean :497.1	Mean :2562	ohcf: 15	Mean : 4.42	Mean :128.1
3rd Qu.:556.0	3rd Qu.:2952	ohcv: 13	3rd Qu.: 4.00	3rd Qu.:146.0
Max. :598.0	Max. :4066		Max. :12.00	Max. :326.0

fuel.system	bore	stroke	compression.ratio	horsepower
1bbl:11	Min. : 35.0	Min. : 28.0	Min. : 7.0	Min. : 48.0
2bbl:64	1st Qu.:305.0	1st Qu.:268.0	1st Qu.: 9.0	1st Qu.: 70.0
idi :19	Median :327.0	Median :319.0	Median : 85.0	Median : 95.0
mfi : 1	Mean :315.9	Mean :273.6	Mean : 73.7	Mean :103.5
mpfi:88	3rd Qu.:354.0	3rd Qu.:339.0	3rd Qu.: 94.0	3rd Qu.:116.0
spdi: 9	Max. :394.0	Max. :417.0	Max. :941.0	Max. :262.0
spfi: 1				

peak.rpm	city.mpg	highway.mpg	price
Min. :4150	Min. :13.00	Min. :16.00	Min. : 5118
1st Qu.:4800	1st Qu.:19.00	1st Qu.:25.00	1st Qu.: 7738
Median :5100	Median :25.00	Median :30.00	Median :10245
Mean :5100	Mean :25.33	Mean :30.79	Mean :13285
3rd Qu.:5500	3rd Qu.:30.00	3rd Qu.:34.00	3rd Qu.:16515
Max. :6600	Max. :49.00	Max. :54.00	Max. :45400

```
'data.frame': 193 obs. of 24 variables:
 $ make      : Ord.factor w/ 21 levels "alfa-romero"<...: 1 1 1 2 2 2 2 2 3 ...
 $ fuel.type  : Ord.factor w/ 2 levels "diesel"<"gas": 2 2 2 2 2 2 2 2 2 ...
 $ aspiration : Ord.factor w/ 2 levels "std"<"turbo": 1 1 1 1 1 1 1 1 2 1 ...
 $ num.doors  : int 2 2 2 4 4 2 4 4 4 2 ...
 $ body.style : Ord.factor w/ 5 levels "convertible"<...: 1 1 3 4 4 4 4 5 4 4 ...
 $ drive.wheels : Ord.factor w/ 3 levels "4wd"<"fwd"<"rwd": 3 3 3 2 1 2 2 2 2 3 ...
 $ engine.location : Ord.factor w/ 2 levels "front"<"rear": 1 1 1 1 1 1 1 1 1 1 ...
 $ wheel.base : num 886 886 945 998 994 ...
 $ length     : num 1688 1688 1712 1766 1766 ...
 $ width      : num 641 641 655 662 664 663 714 714 714 648 ...
 $ height     : num 488 488 524 543 543 531 557 557 559 543 ...
 $ curb.weight : int 2548 2548 2823 2337 2824 2507 2844 2954 3086 2395 ...
 $ engine.type : Ord.factor w/ 5 levels "dohc"<"l"<"ohc"<...: 1 1 5 3 3 3 3 3 3 3 ...
 $ num.cylinders : int 4 4 6 4 5 5 5 5 5 4 ...
 $ engine.size : int 130 130 152 109 136 136 136 131 108 ...
 $ fuel.system  : Ord.factor w/ 7 levels "lbb1"<"2bb1"<...: 5 5 5 5 5 5 5 5 5 ...
 $ bore        : num 347 347 268 319 319 319 319 319 313 35 ...
```

d) Considerando então apenas uma variável preditora (explicativa) no modelo (horsepower), tente ajustar um modelo para explicar o preço (price) dos carros:

i) Qual sua intuição sobre o relacionamento da “potência” de um carro com o seu preço?

```
Call:
lm(formula = price ~ horsepower, data = autos_imp)

Coefficients:
(Intercept)    horsepower
   -4630.7         173.1

Call:
lm(formula = price ~ horsepower, data = autos_imp)

Residuals:
    Min       1Q   Median       3Q      Max
-10296.1  -2243.5   -450.1   1794.7  18174.9

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4630.70     990.58  -4.675 5.55e-06 ***
horsepower    173.13       8.99   19.259 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4728 on 191 degrees of freedom
Multiple R-squared:  0.6601,    Adjusted R-squared:  0.6583
F-statistic: 370.9 on 1 and 191 DF,  p-value: < 2.2e-16
A anova: 2 x 5
```

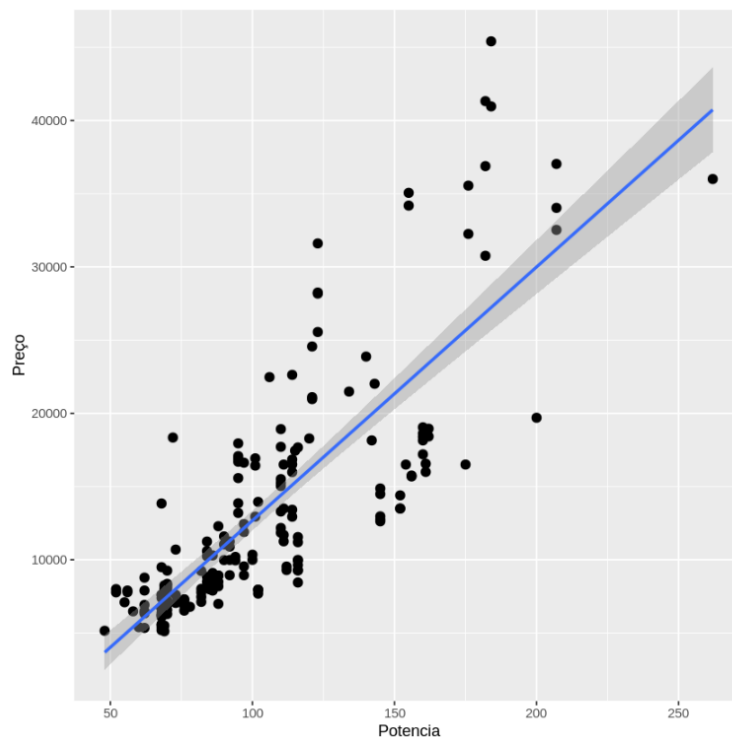
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
horsepower	1	8292710641	8292710641	370.8973	1.248407e-46
Residuals	191	4270475650	22358511	NA	NA

Analisando os valores obtidos podemos dizer que tendo Pr de cada variável menor que 5% e com o p-value menor que 5% também, o modelo pode ser considerado representativo. R-squared está em 66%, o que não é um valor muito elevado, porém, considerando as demais variáveis, ainda assim seria uma função válida para representar os dados do sistema.

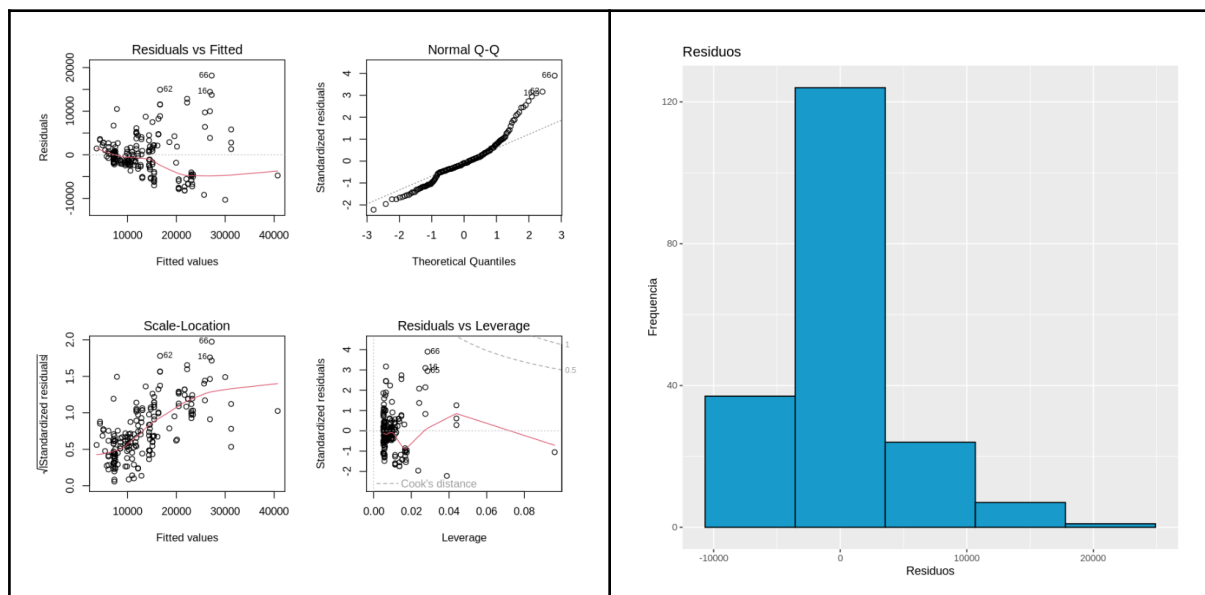
e) Faça todas as análises da regressão (modelo), julgando:

i) Os gráficos diagnósticos

ii) Os parâmetros obtidos (avaliar o summary do modelo)



Considerando a representação gráfica, vemos que para baixas potências o modelo é representativo, porém para potências mais altas passamos a ter uma maior distribuição dos pontos, o que está prejudicando a acurácia da equação.



Analisando os gráficos acima podemos afirmar que a função não representa o sistema de uma forma aceitável, uma vez que os resíduos não estão próximos de zero e na curva Q-Q temos uma discrepância muito grande na parte final da curva.

f) Interprete os resultados do ajuste:

i) Em que posição a reta corta o eixo Y? Isso faz sentido?

-4630.7

Não, esse valor não faz nenhum sentido, uma vez que não existem preços negativos.

ii) Como corrigir um modelo que apresenta este comportamento?

Podemos remover alguns outliers ou então retrabalhar para utilizar outra variável que possa se adequar mais.

Existe também a opção do uso da função logarítmica para se trabalhar com valores paralelos ao do sistema real.

Outra opção seria adicionar mais variáveis ao sistema para que elas possam auxiliar na análise.

g) Analise: Será que apenas a potência de um carro é suficiente para termos uma boa previsão do preço deste carro?

i) O que indica isso no seu ajuste?

Como a função obtida não parece representar bem o sistema, a utilização de mais variáveis para representar os dados poderia ser uma opção válida.