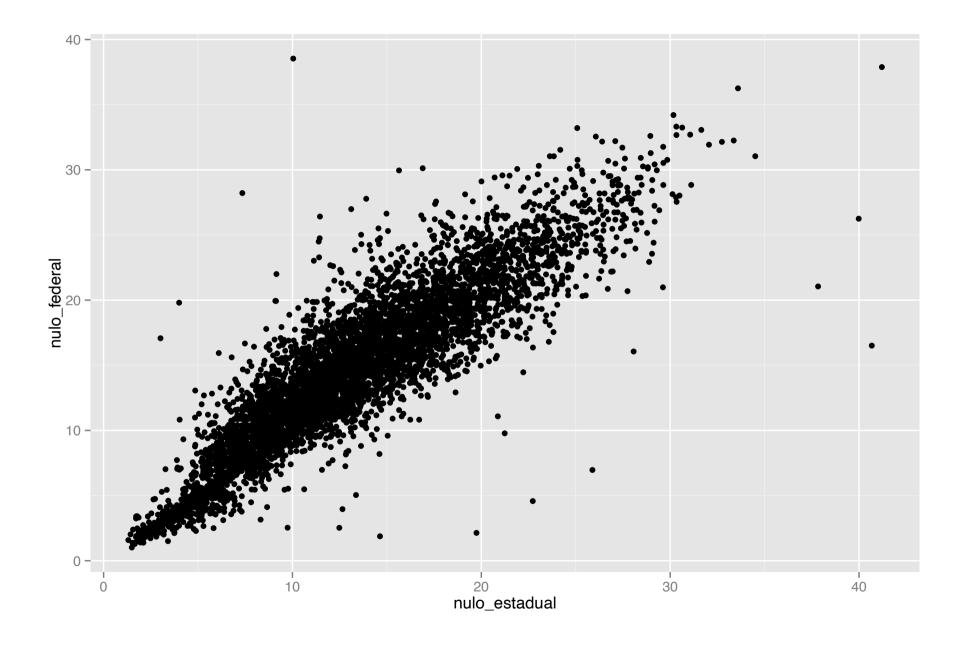
Diagrama de dispersão e correlação

Diagrama de dispersão (scatterplot)



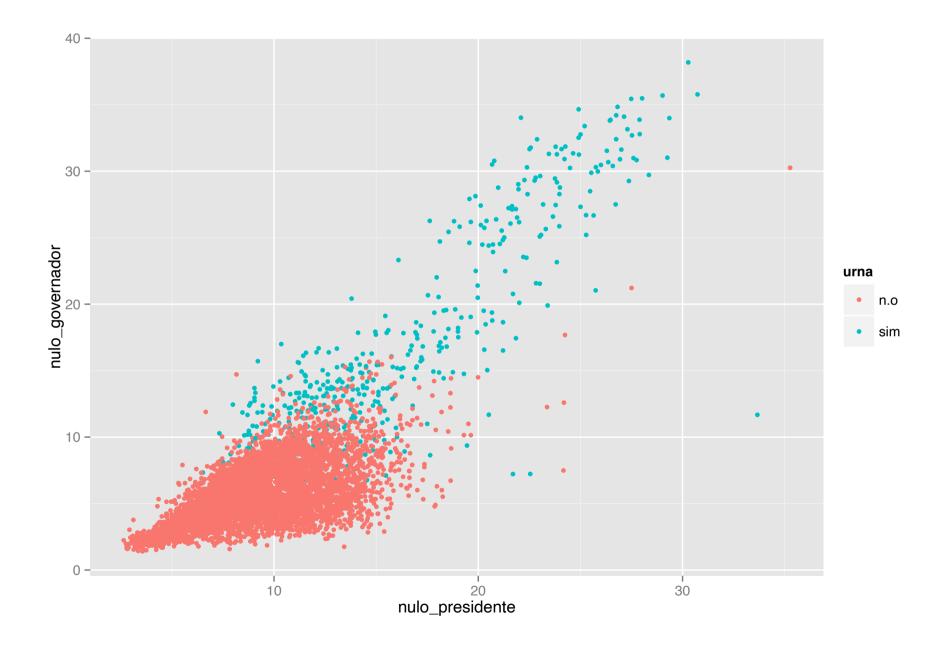


Diagrama de dispersão

- Excelente forma de observar associações entre duas variáveis quantitativas.
- Direção:
- ✓ Um padrão que segue do canto superior esquerdo para o canto inferior à direita tem uma direção negativa.
- ✓ Um padrão seguindo caminho inverso tem uma direção positiva.

Correlação

Propriedades da correlação

- O sinal de um coeficiente de correlação revela a direção da associação.
- Correlação é sempre entre I e + I.
- Correlação pode ser exatamente igual a -1 ou +1, mas esses valores são incomuns para dados reais, pois eles significam que todos os pontos de dados caem exatamente sobre uma linha reta.
- Uma correlação próxima de zero corresponde a uma fraca associação linear.

Propriedades da correlação

- A correlação mensura a força de uma relação linear entre duas variáveis.
- As variáveis podem ter uma forte associação, mas uma reduzida correlação se a associação não é linear.
- A correlação é sensivel aos outliers. Um único valor outilier pode fazer uma reduzida correlação tornar se forte, ou vice-versa.

Propriedades da correlação

- Outliers podem distorcer a correlação acentuadamente.
- Um outlier pode fazer uma pequena correlação parecer grande ou esconder uma forte correlação.
- Ele pode até mesmo transformar um coeficiente de uma associação positiva em negativa (e vice- versa).
- Quando encontrar um oulier é uma boa ideia apresentar as correlações com e sem aquele ponto.

Correlação de Pearson = r

 A correlação mensura a força de uma relação linear entre duas variáveis.

$$r = \frac{\sum \left[\frac{(x - \bar{x})}{s_x} \frac{(y - \bar{y})}{s_y}\right]}{n - 1}$$

Calculando o coeficiente de correlação (r)

Exemplo: PIB per Capita e expectativa de vida em países europeus selcionados

pais	Pib Per Capita (<i>x</i>)	Expectativa de vida (<i>y</i>)			
Austria	21.4	77.48			
Bélgica	23.2	77.53			
Finlandia	20.0	77.32			
França	22.7	78.63			
Alemanha	20.8	77.17			
Irlanda	18.6	76.39			
Italia	21.5	78.51			
Holanda	22.0	78.15			
Suíça	23.8	78.99			
Reino Unido	21.2	77.37			

Calculando o coeficiente de correlação (r)

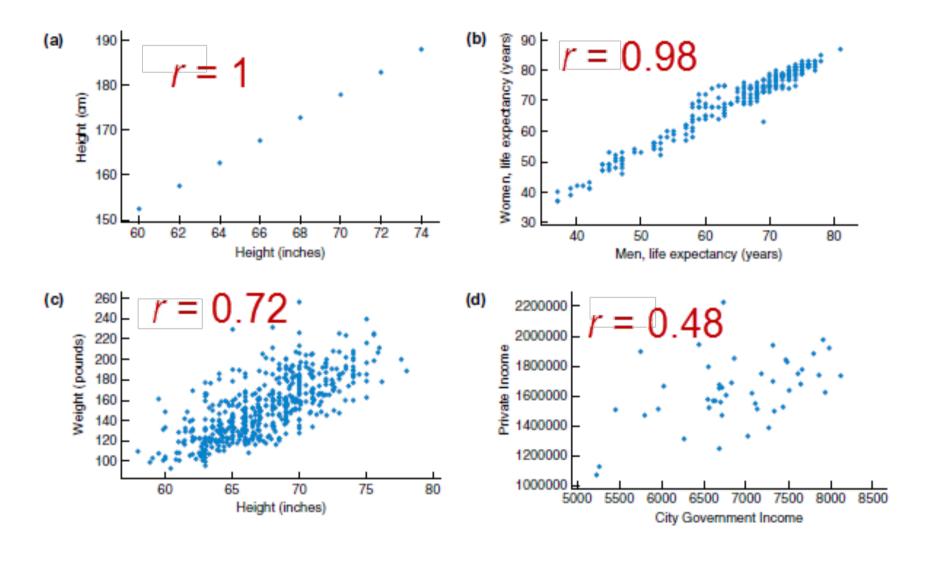
Exemplo: PIB per Capita e expectativa de vida em países europeus selcionados

X	У	(a) xi-x/s	(b) yi –y/s	(a) X (b)	
21.4	77.48	-0.078	-0.345	0.027	_
23.2	77.53	1.097	-0.282	-0.309	$1 \sum_{i=1}^{n} \left(x_{i} - \overline{x} \right) \left(y_{i} \right)$
20.0	77.32	-0.992	-0.546	0.542	$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i}{s_x} \right)$
22.7	78.63	0.770	1.102	0.849	,
20.8	77.17	-0.470	-0.735	0.345	$=\left(\frac{1}{10-1}\right)(7.285)$
18.6	76.39	-1.906	-1.716	3.271	(10-1)
21.5	78.51	-0.013	0.951	-0.012	= 0.809
22.0	78.15	0.313	0.498	0.156	
23.8	78.99	1.489	1.555	2.315	
21.2	77.37	-0.209	-0.483	0.101	
mx= 21.52	my= 77.75			- 7.005	

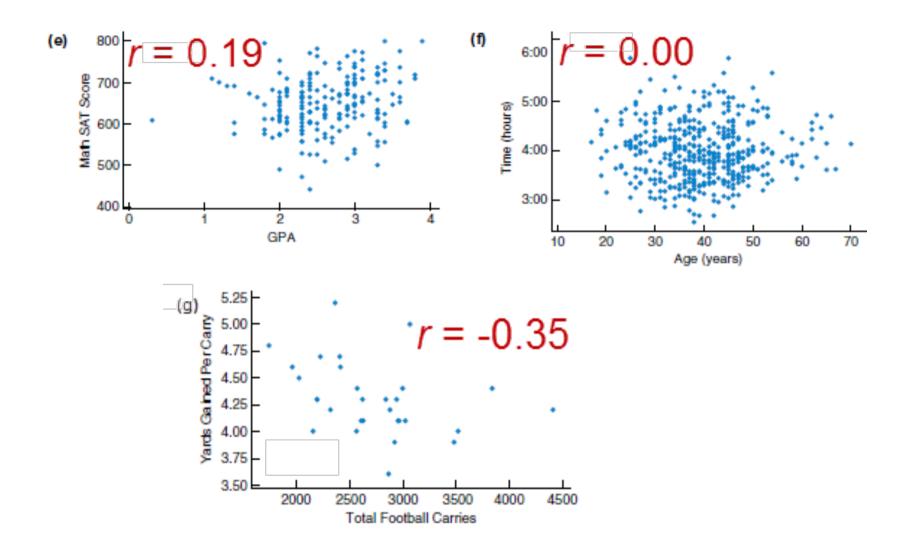
 $s_x = 1.532$ $s_v = 0.795$

soma = 7.285

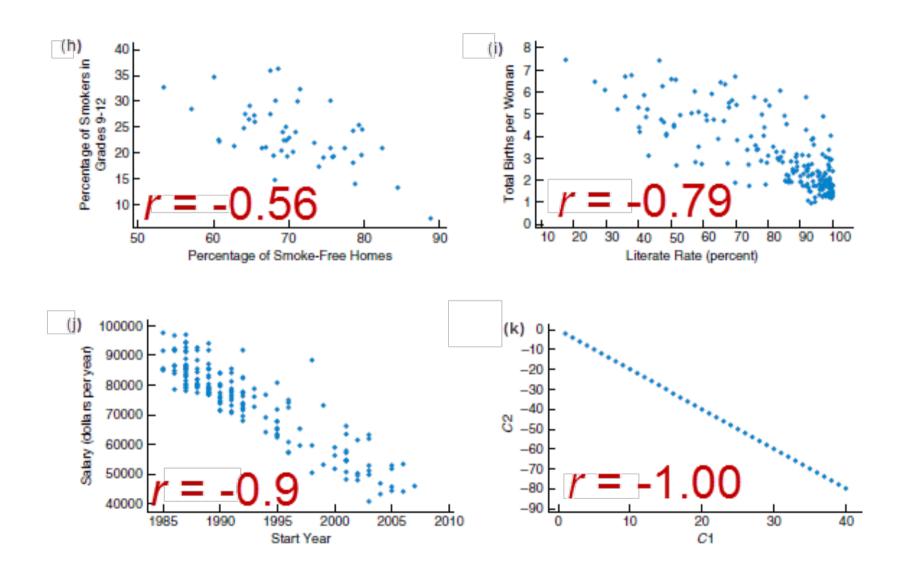
Correlação positiva



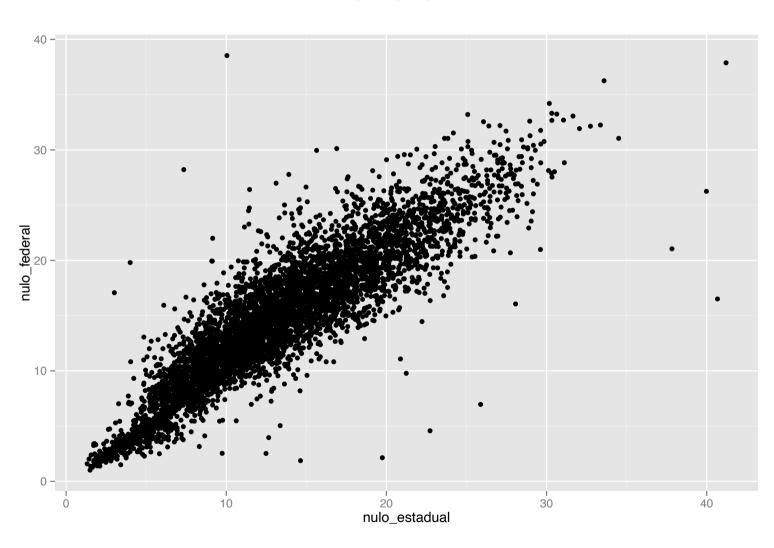
Correlação fraca ou ausente



Correlação negativa



r = 0.90



	á	esidente nulo se	anador	governador nulo e	stadual	estadual	aderal branco	Diesidente	senador	overnador
	nulo J	nulo 3	branco	nulo	brance	nulo 3	brance	brance	rulo 9	1
nulo_presidente	1	0.64	0.16	0.02	0.01	-0.05	0.06	0.05	0.76	- 0.8
nulo_senador	0.64	1	0.36	0.21	0.28	0.11	0.16	0.17	0.68	- 0.6
branco_governador	0.16	0.36	1	0.34	0.52	0.33	0.72	0.75	-0.22	- 0.4
nulo_estadual	0.02	0.21	0.34	1	0.62	0.9	0.29	0.3	-0.01	- 0.2
branco_estadual	0.01	0.28	0.52	0.62	1	0.52	0.48	0.6	-0.04	- 0
nulo_federal	-0.05	0.11	0.33	0.9	0.52	1	0.29	0.33	-0.09	0.2
branco_presidente	0.06	0.16	0.72	0.29	0.48	0.29	1	0.66	-0.21	0.4
branco_senador	0.05	0.17	0.75	0.3	0.6	0.33	0.66	1	-0.22	0.6
nulo_governador	0.76	0.68	-0.22	-0.01	-0.04	-0.09	-0.21	-0.22	1	-0.8