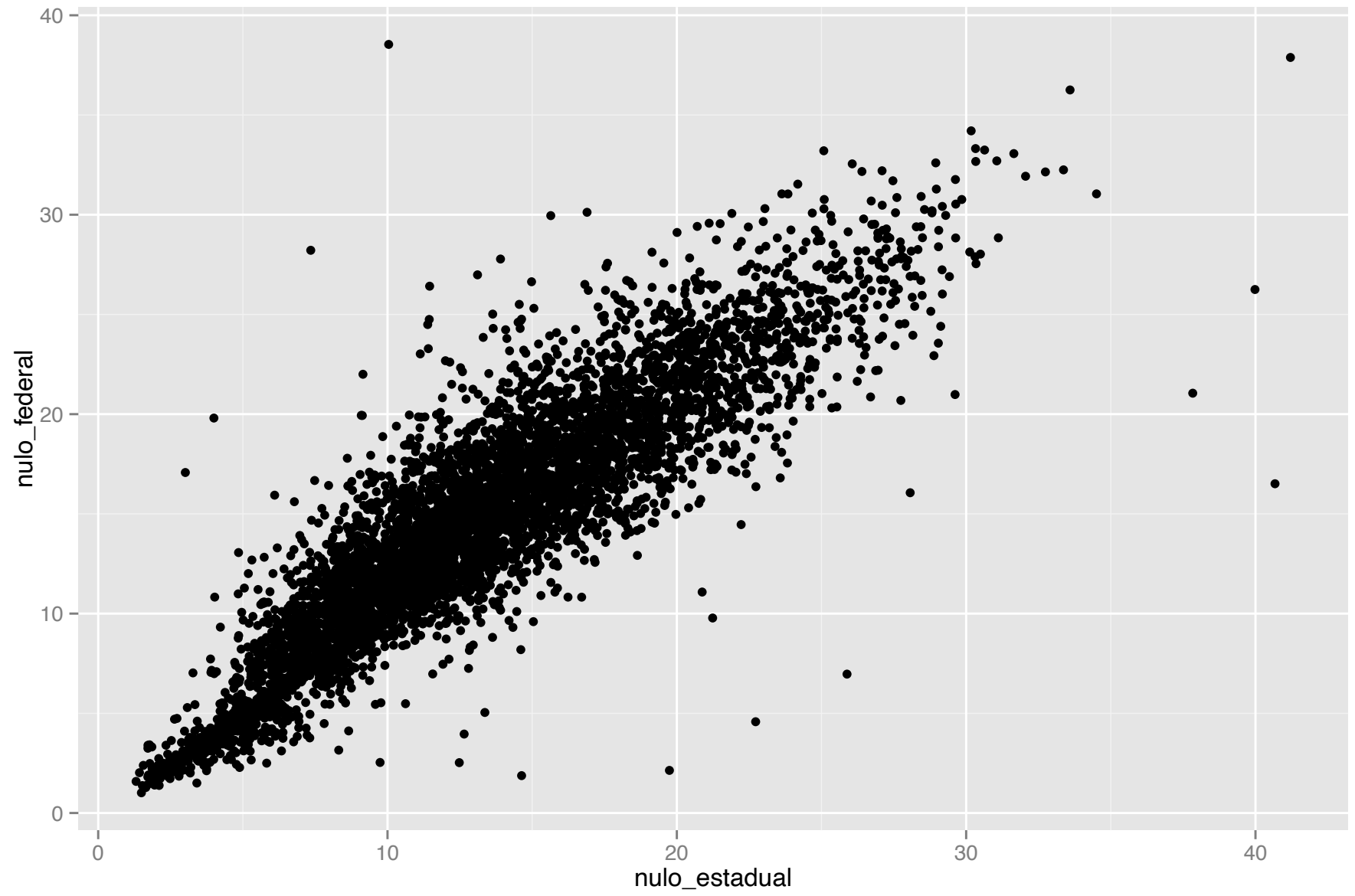


Diagrama de dispersão e correlação

Diagrama de dispersão (scatterplot)



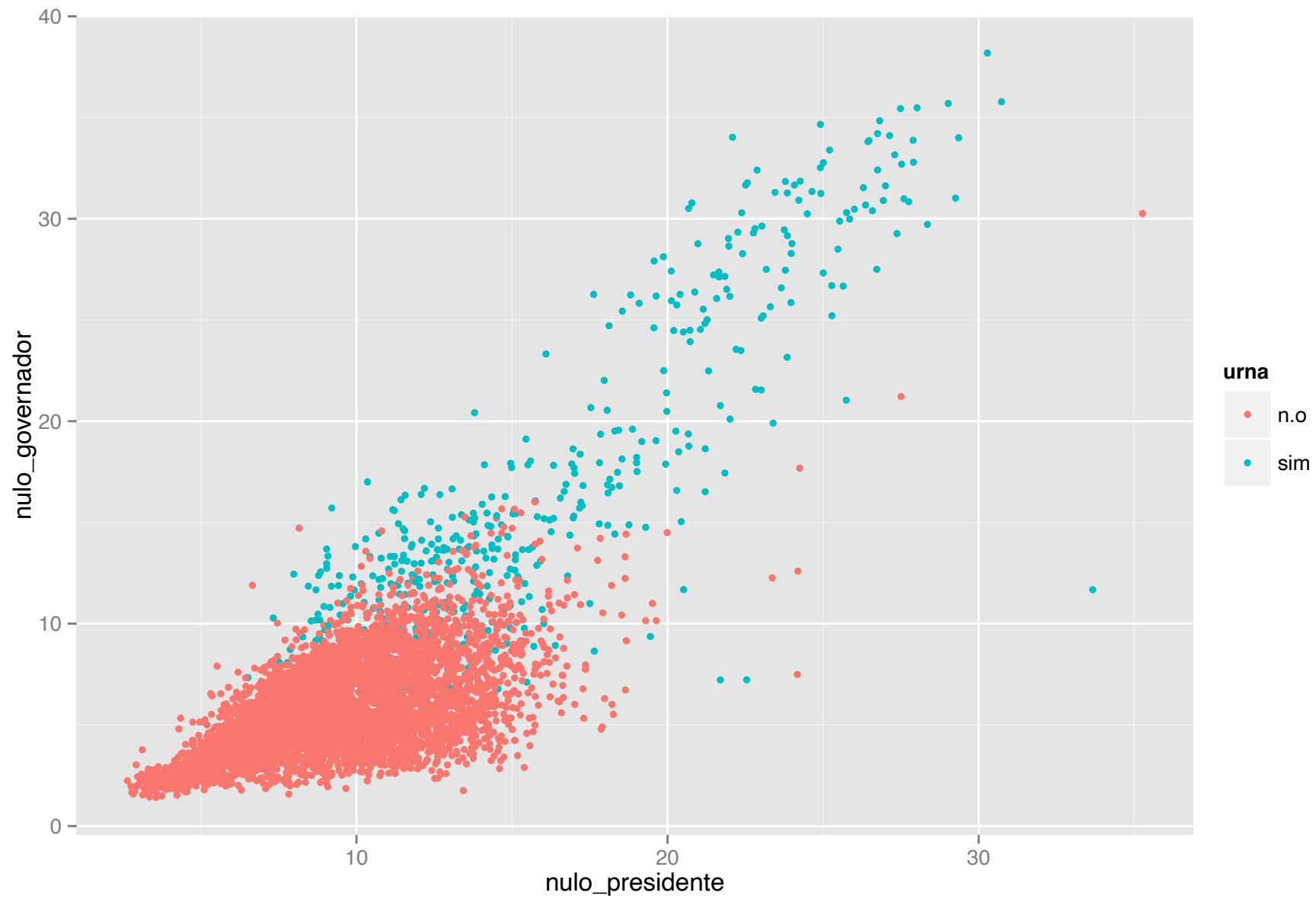


Diagrama de dispersão

- 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 -1 e $+1$.
- 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 sensível aos outliers. Um único valor outlier 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 outlier 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 selecionados

país	Pib Per Capita (x)	Expectativa de vida (y)
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 selecionados

x	y	(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
20.0	77.32	-0.992	-0.546	0.542
22.7	78.63	0.770	1.102	0.849
20.8	77.17	-0.470	-0.735	0.345
18.6	76.39	-1.906	-1.716	3.271
21.5	78.51	-0.013	0.951	-0.012
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

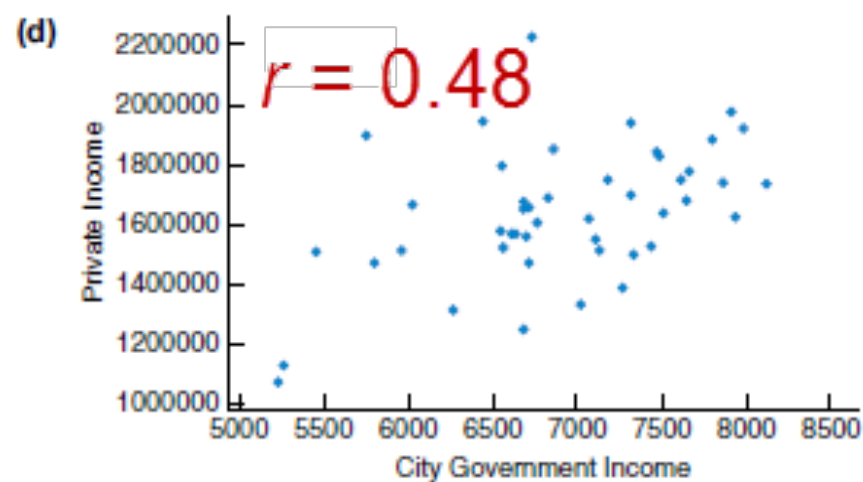
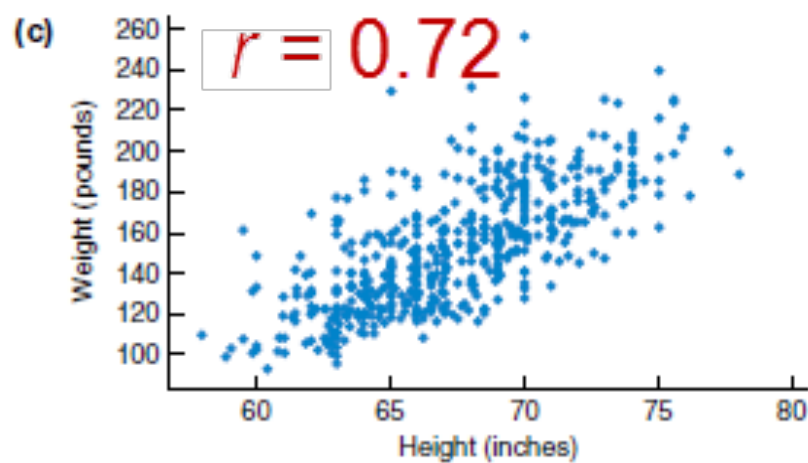
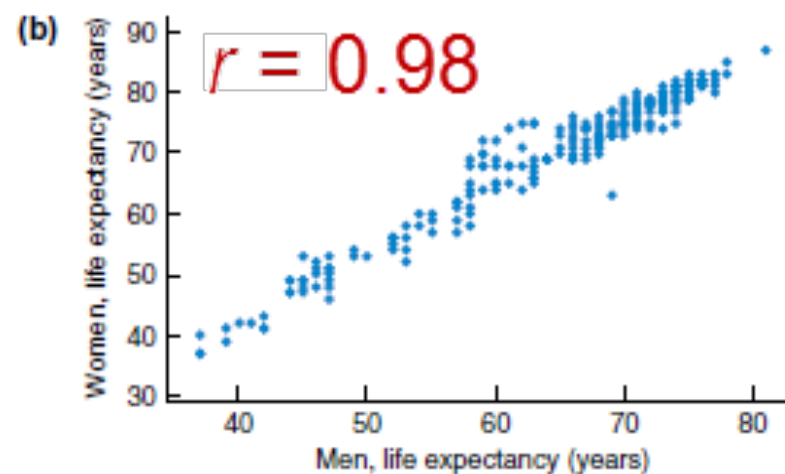
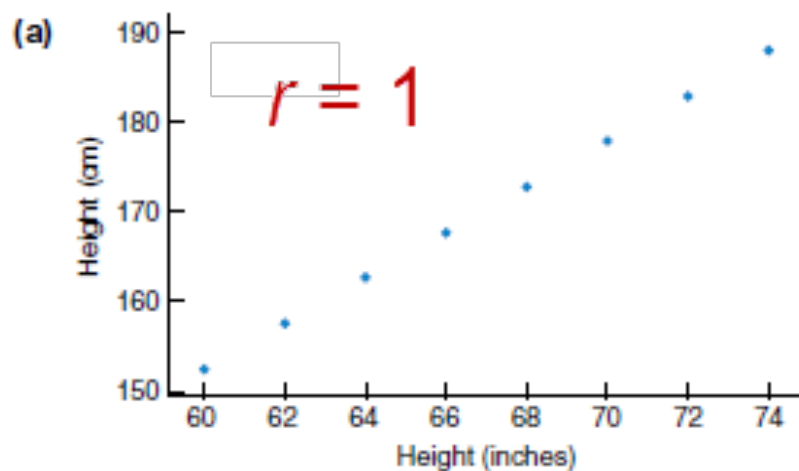
mx= 21.52 my= 77.75

s_x=1.532 s_y=0.795

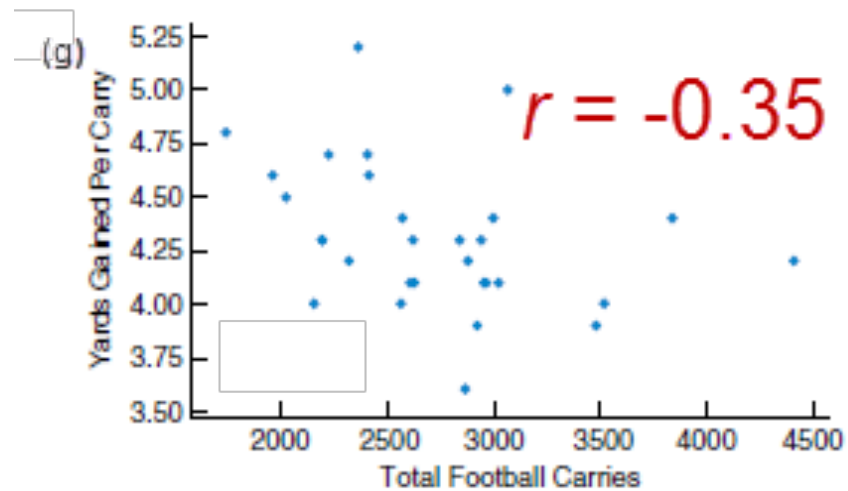
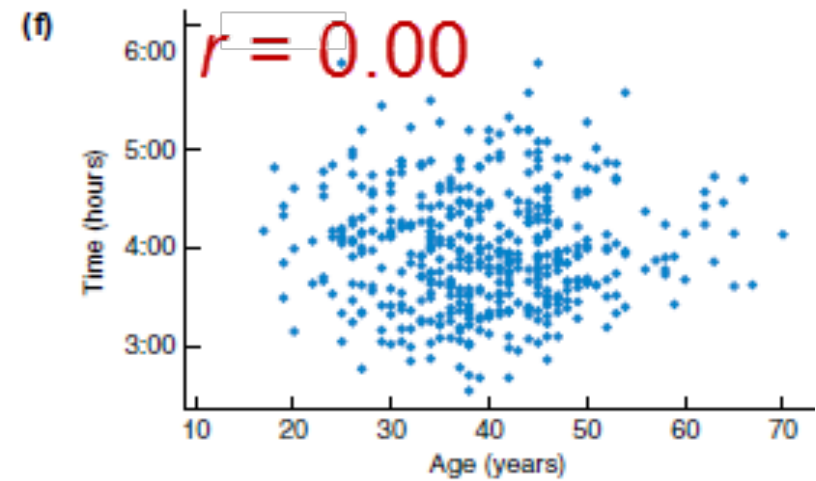
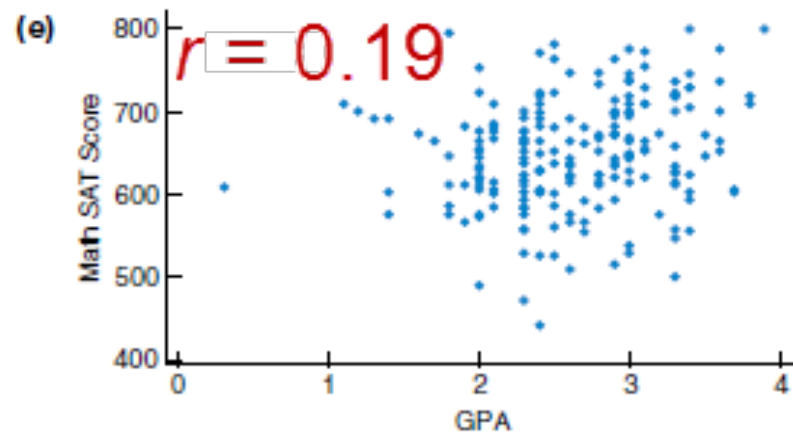
soma = 7.285

$$\begin{aligned}
 r &= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) \\
 &= \left(\frac{1}{10-1} \right) (7.285) \\
 &= 0.809
 \end{aligned}$$

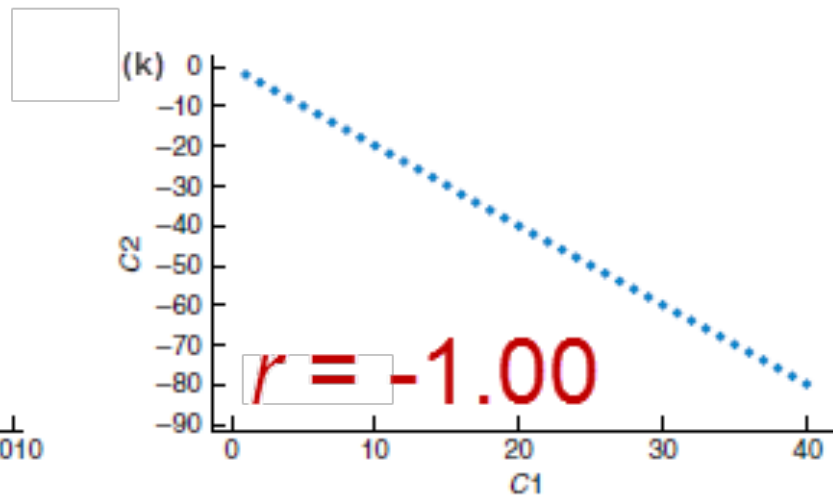
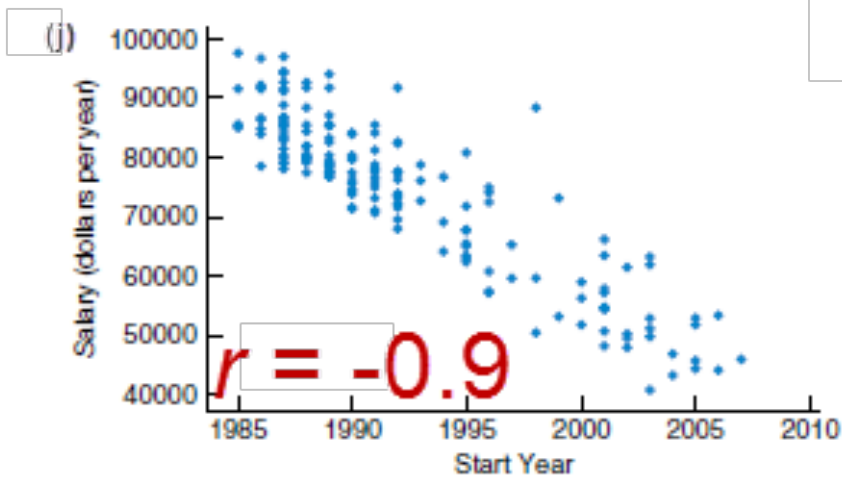
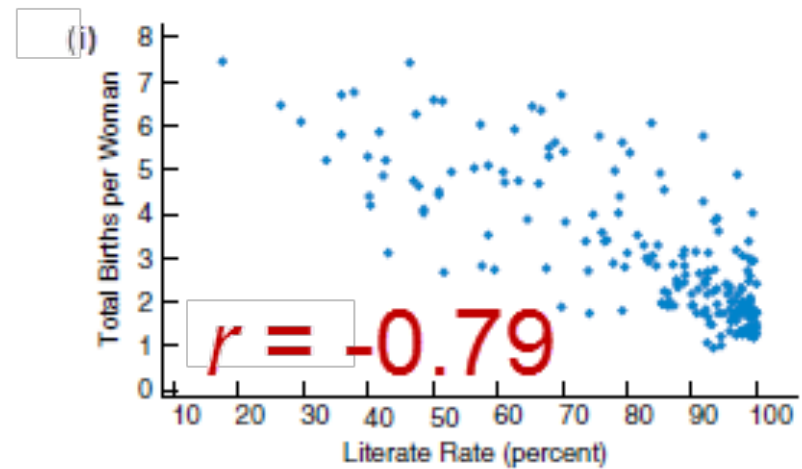
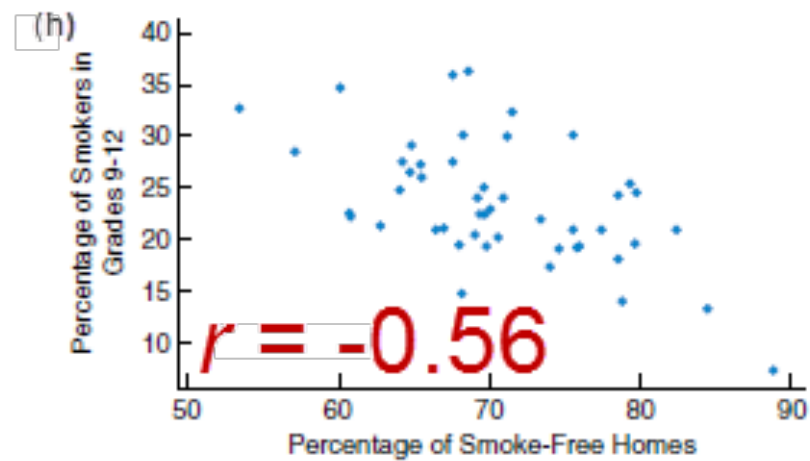
Correlação positiva



Correlação fraca ou ausente



Correlação negativa



$r = 0.90$

