

13

CORRELAÇÃO

Profa. Raquel C. de Melo Minardi

- Análise de como variáveis quantitativas se relacionam e como uma afeta as outras
- Envolve a comparação de variáveis para verificar se a variação em um conjunto de valores interfere no outro sistematicamente, em qual grau e porque

1. Uma variável influencia na outra
2. As variáveis não se influenciam
3. As variáveis não se influenciam mutuamente mas são conectadas por uma terceira variável
4. As variáveis se correlacionam aparentemente mas devido à amostragem insuficiente ou enviesada

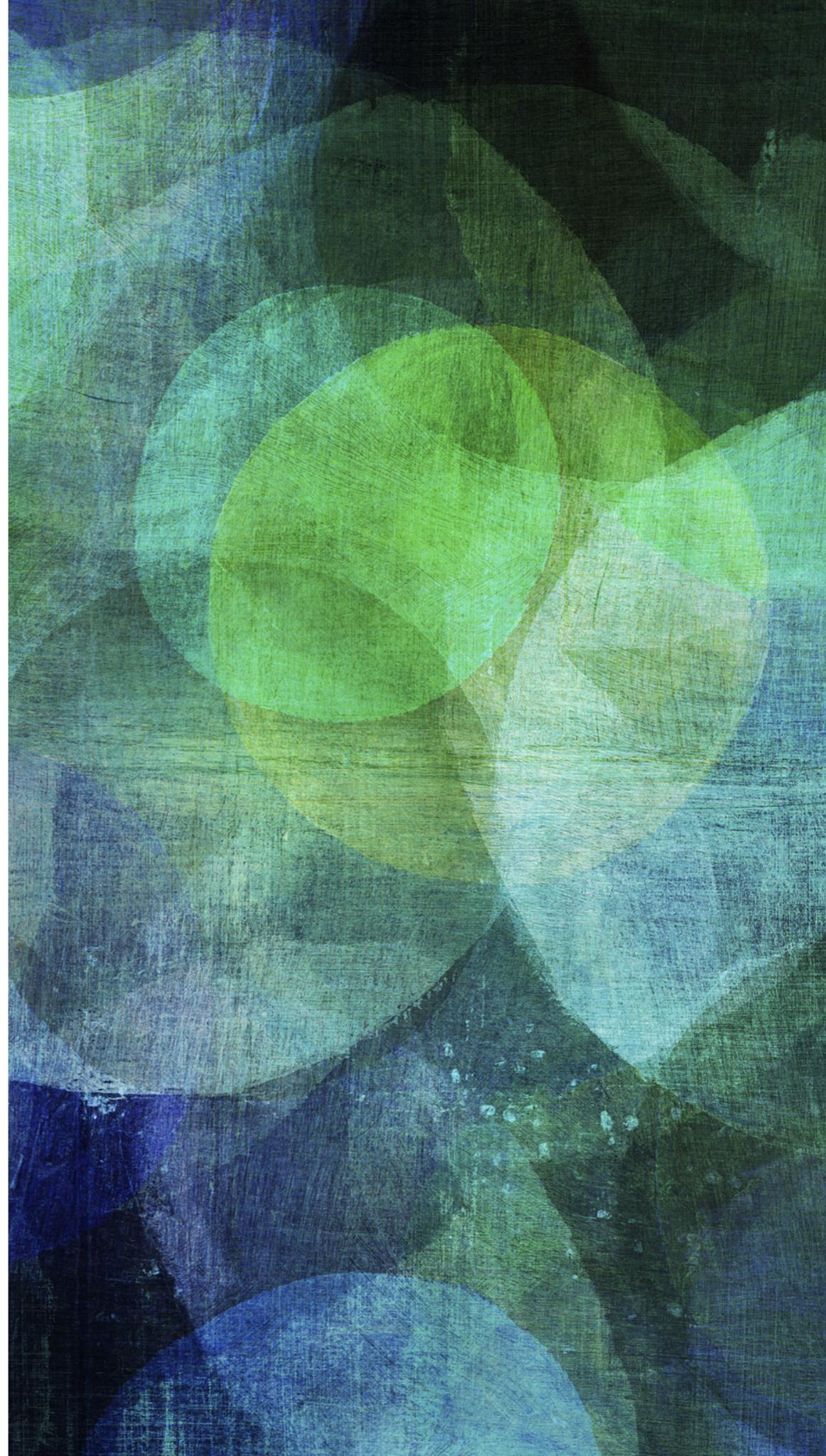
EXEMPLO DO CASO 3

- Pode-se mostrar uma forte correlação entre a taxa de mortalidade e a temperatura da cidade
- Pessoas mais velhas buscam viver em cidades mais quentes

EXEMPLO DO CASO 4

- Pode-se mostrar uma forte correlação entre salários de professores e o número de carros novos vendidos
- Ambas as variáveis são afetadas pelo estado da economia

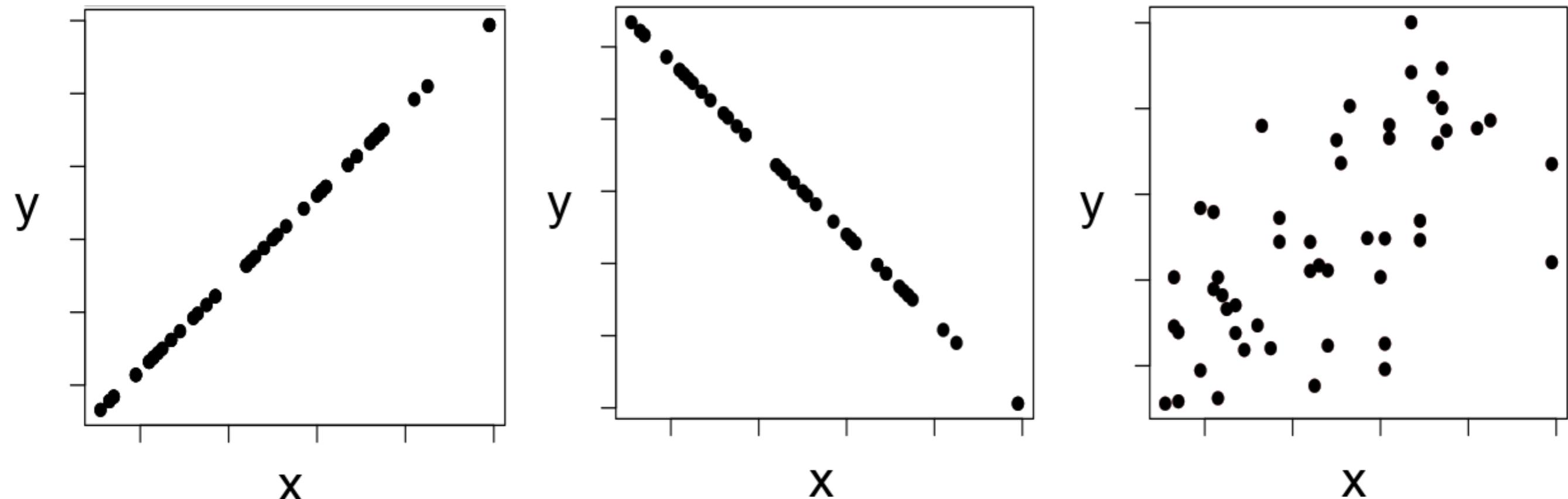
PADRÕES ANALÍTICOS



PADRÕES ANALÍTICOS

- Direção
- Grau
- Forma

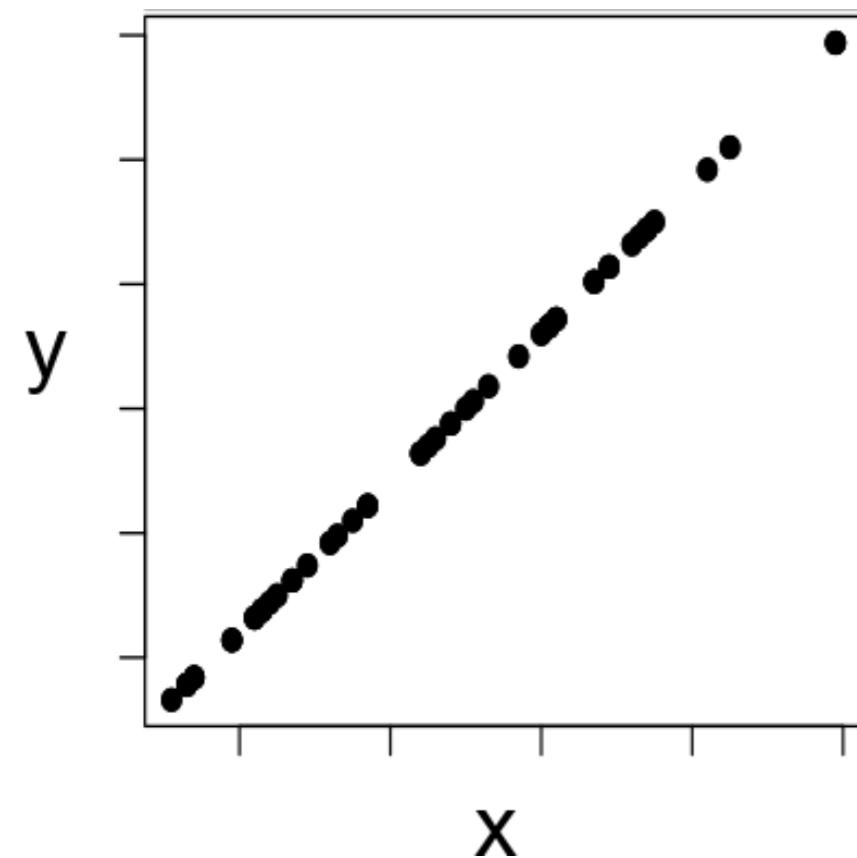
PADRÕES ANALÍTICOS



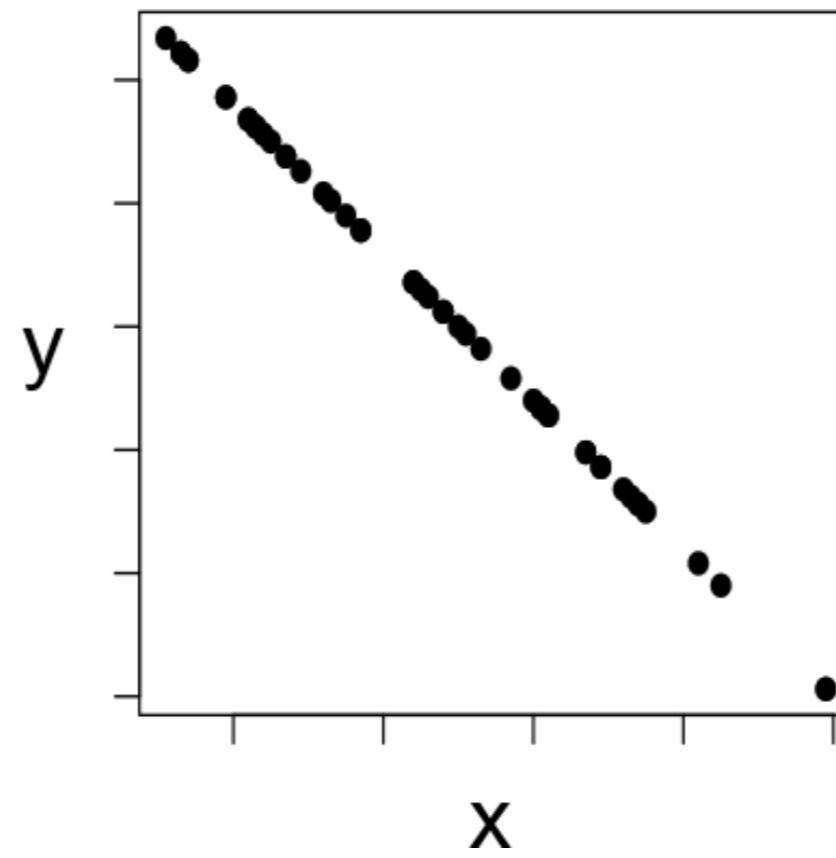
COEFICIENTE DE CORRELAÇÃO LINEAR

- O coeficiente de correlação linear r descreve a direção e o grau da correlação
 - $r = +1$: correlação forte e positiva
 - $r = -1$: correlação forte e negativa
 - $r = 0$: sem correlação

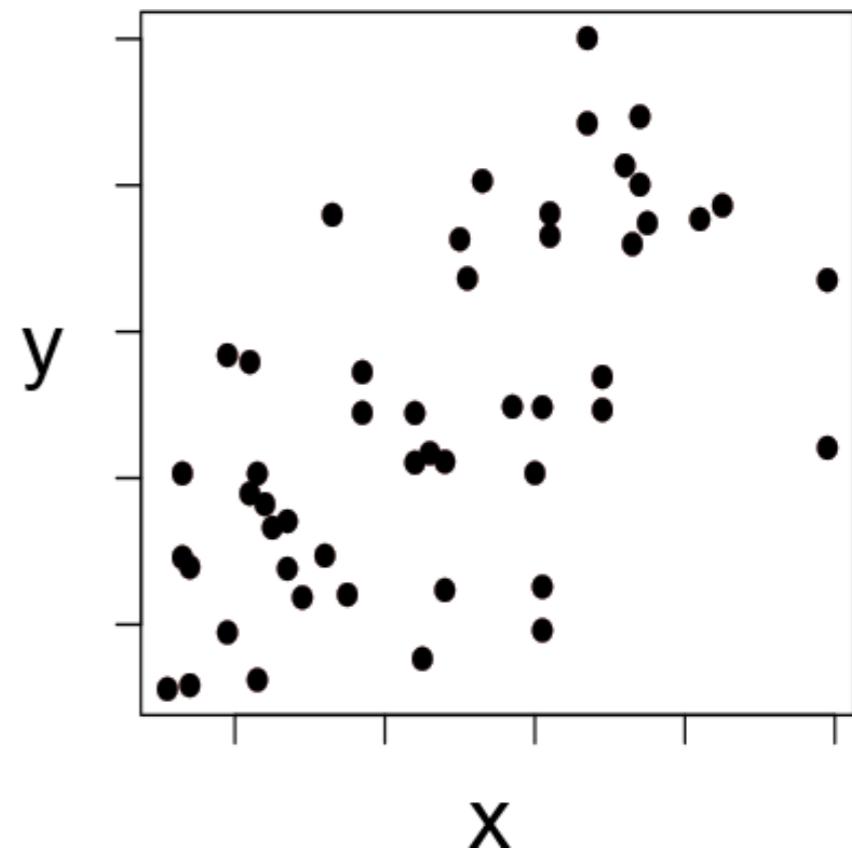
COEFICIENTE DE CORRELAÇÃO LINEAR



$$r = +1$$



$$r = -1$$



$$r = 0,62$$

Slide me



4

2

0

-2

-4

-2

0

2

4

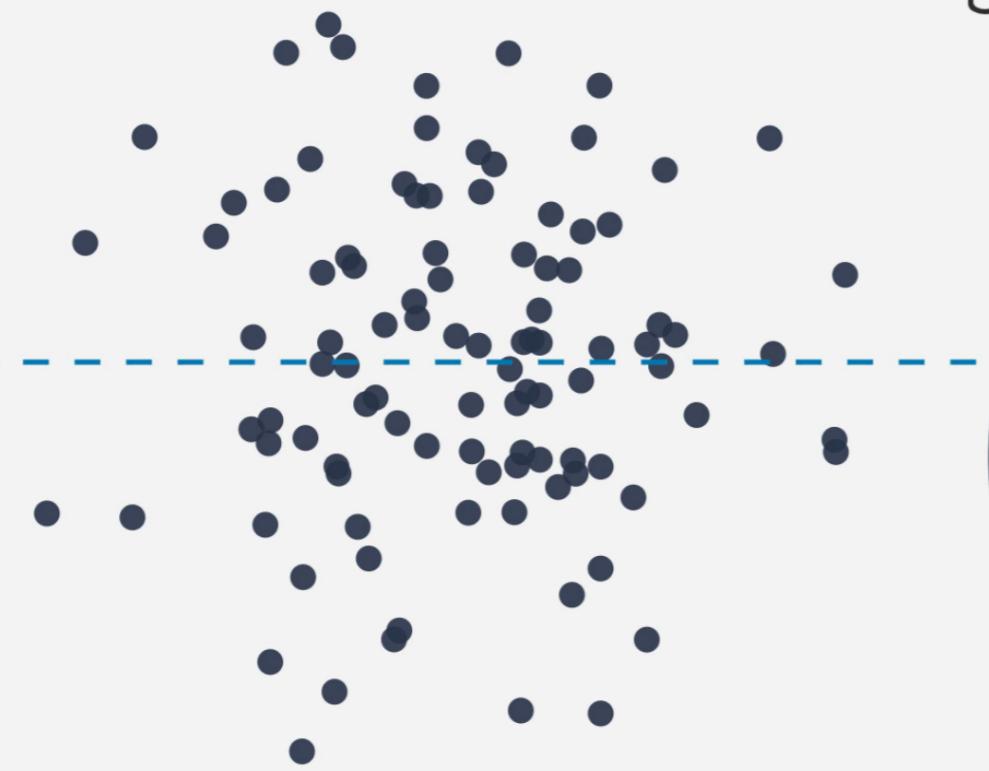
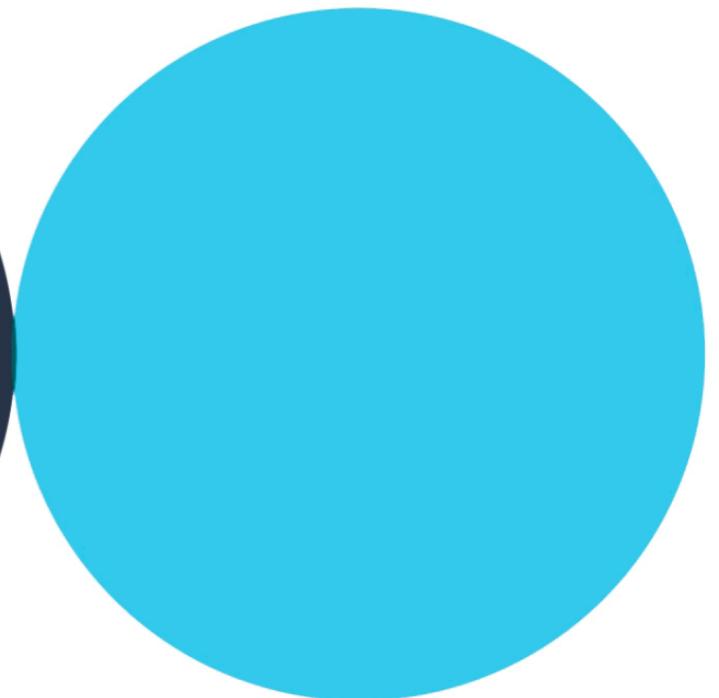
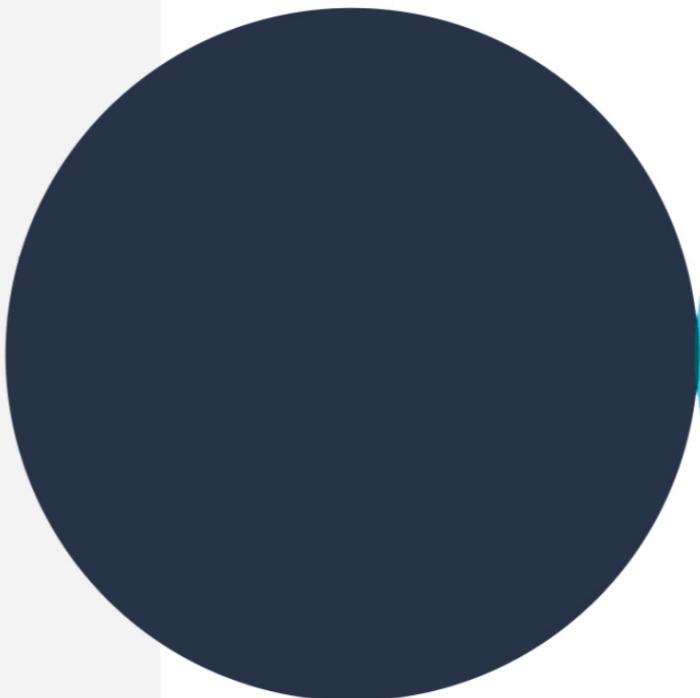
Correlation: 0

Sample size

100

New sample

Shared variance: 0%



COEFICIENTE DE DETERMINAÇÃO

- O coeficiente de determinação r^2 é sempre positivo não indicando a direção da correlação mas é útil pois retrata um percentual
- $r^2 = 0,986$ indica que 98,6% da mudança da variável dependente pode ser determinada pela variável independente

ANScombe's Quartet

Data set	1-3	1	2	3	4	4
Variable	x	y	y	y	x	y
Obs. no.						
1 :	10.0	8.04	9.14	7.46	8.0	6.58
2 :	8.0	6.95	8.14	6.77	8.0	5.76
3 :	13.0	7.58	8.74	12.74	8.0	7.71
4 :	9.0	8.81	8.77	7.11	8.0	8.84
5 :	11.0	8.33	9.26	7.81	8.0	8.47
6 :	14.0	9.96	8.10	8.84	8.0	7.04
7 :	6.0	7.24	6.13	6.08	8.0	5.25
8 :	4.0	4.26	3.10	5.39	19.0	12.50
9 :	12.0	10.84	9.13	8.15	8.0	5.56
10 :	7.0	4.82	7.26	6.42	8.0	7.91
11 :	5.0	5.68	4.74	5.73	8.0	6.89

TABLE. Four data sets, each comprising 11 (x , y) pairs.

Each of the four data sets yields the same standard output from a typical regression program, namely

Number of observations (n) = 11

Mean of the x 's (\bar{x}) = 9.0

Mean of the y 's (\bar{y}) = 7.5

Regression coefficient (b_1) of y on x = 0.5

Equation of regression line: $y = 3 + 0.5x$

Sum of squares of $x - \bar{x}$ = 110.0

Regression sum of squares = 27.50 (1 d.f.)

Residual sum of squares of y = 13.75 (9 d.f.)

Estimated standard error of b_1 = 0.118

Multiple R^2 = 0.667

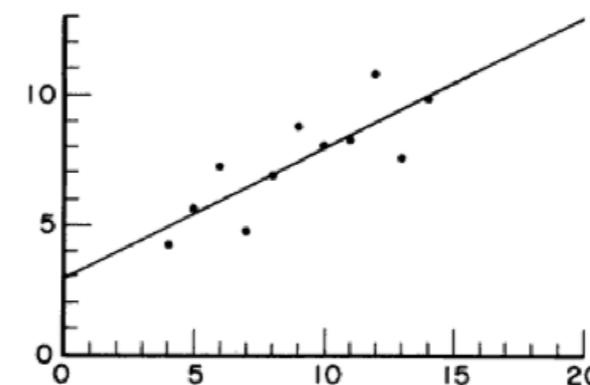


Figure 1

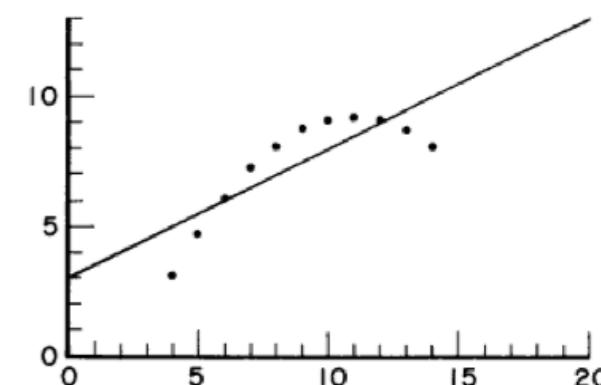


Figure 2

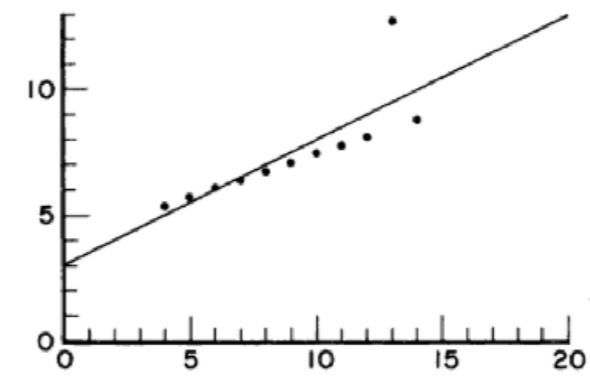


Figure 3

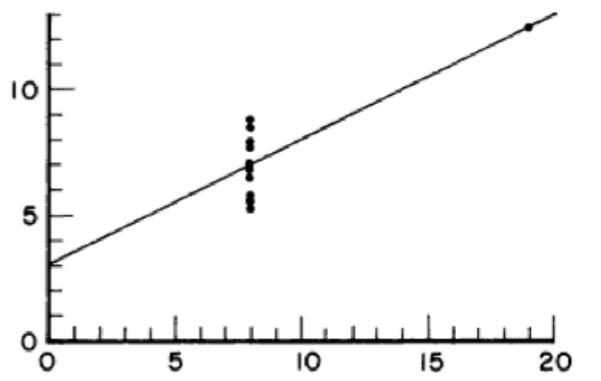
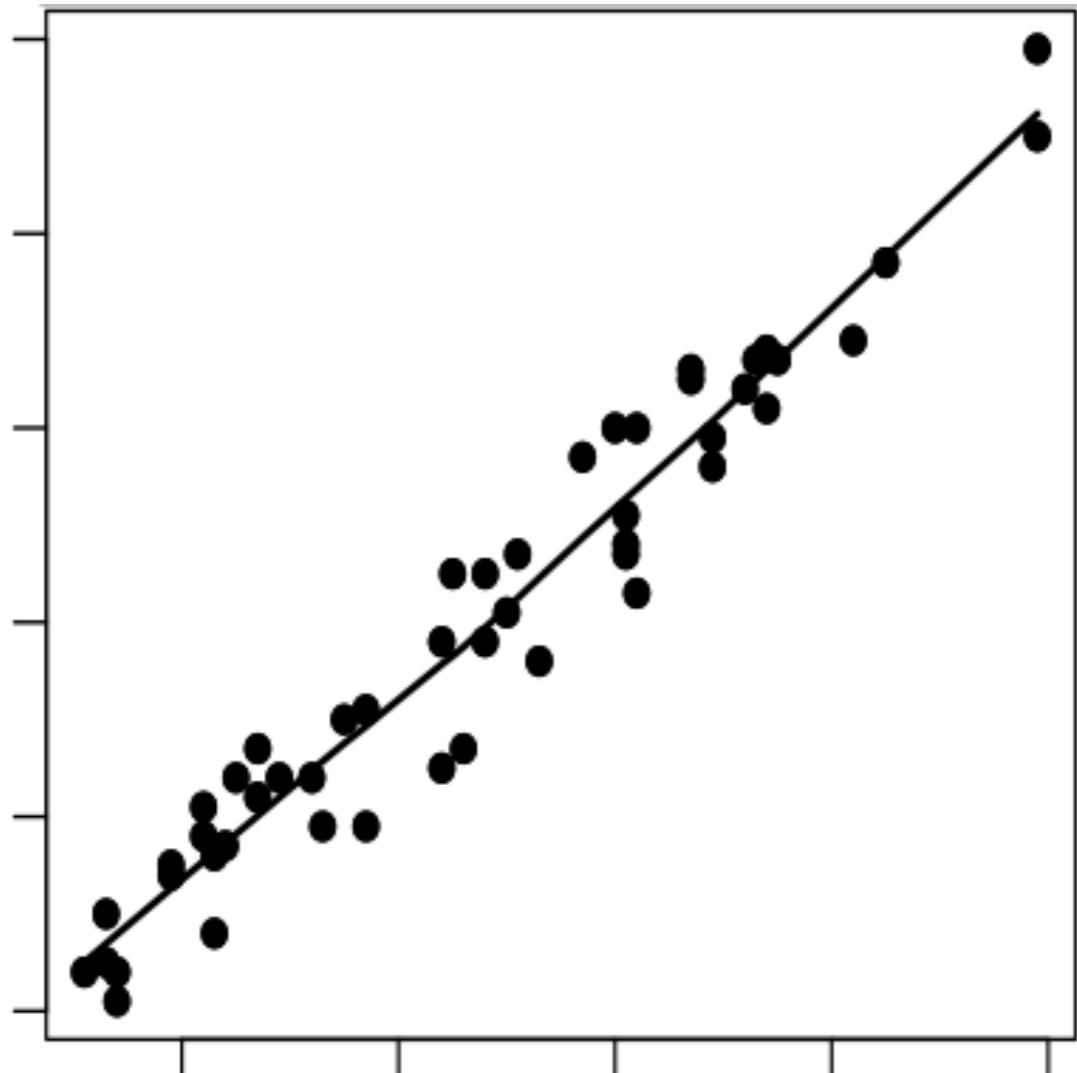


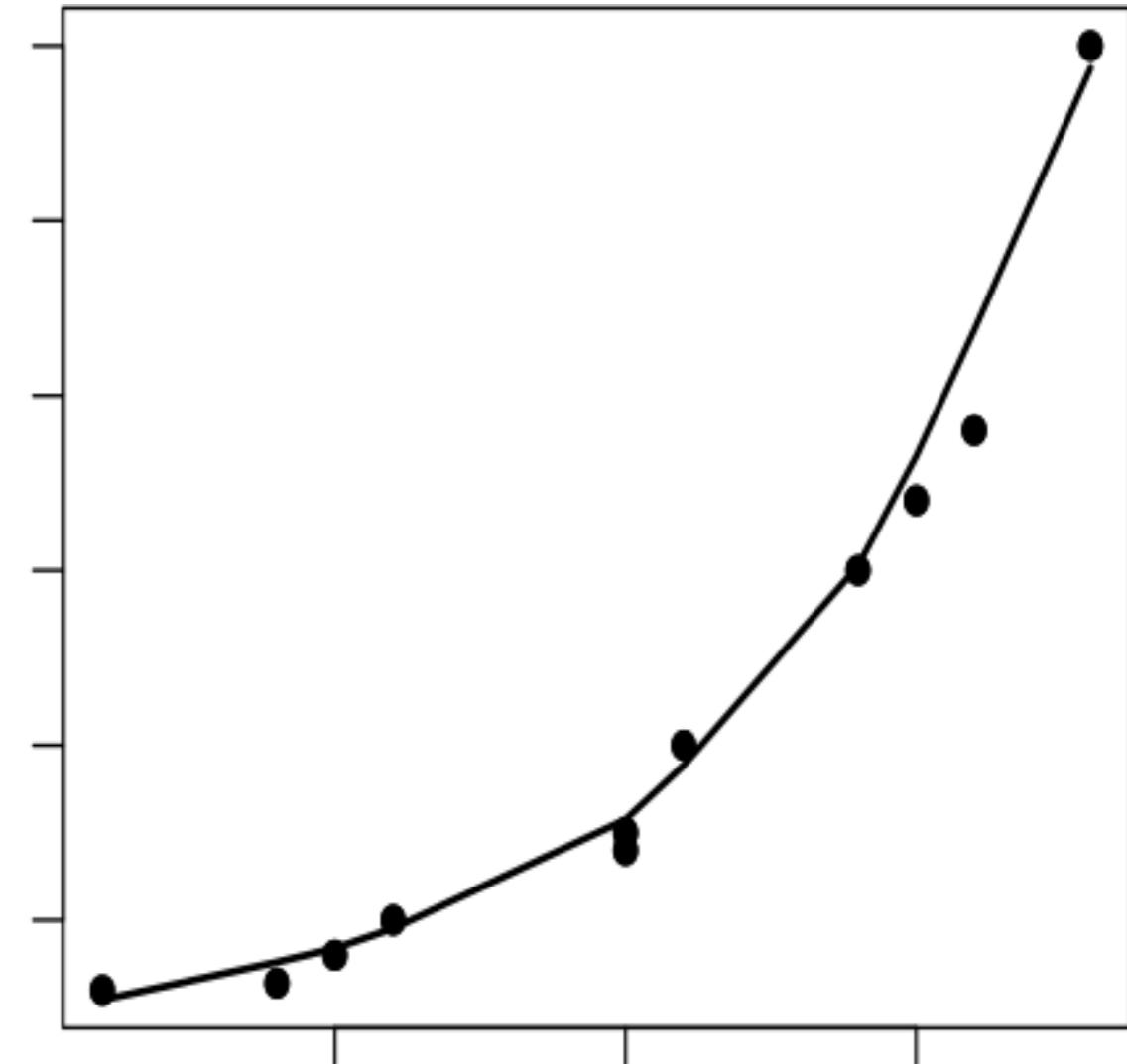
Figure 4

FORMA

Reta ou curva?



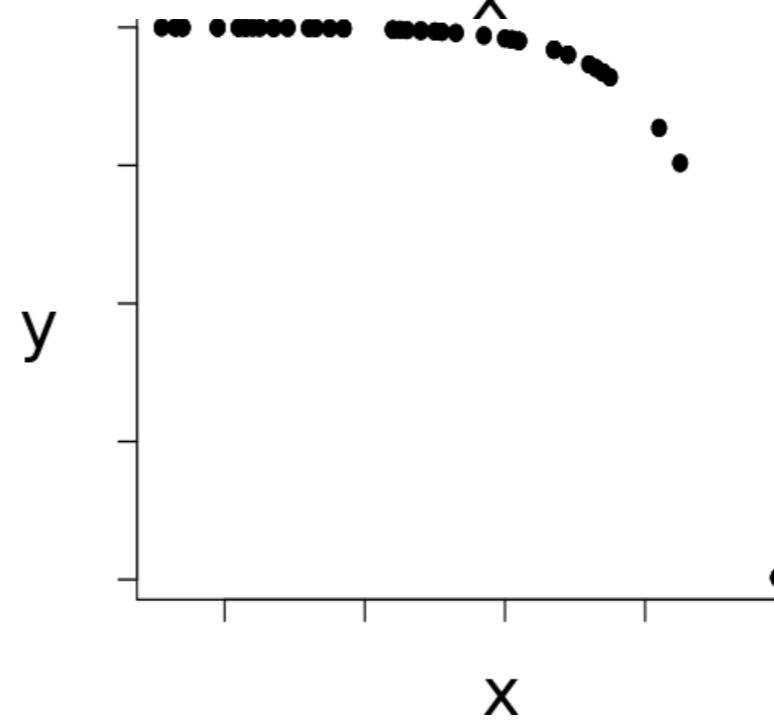
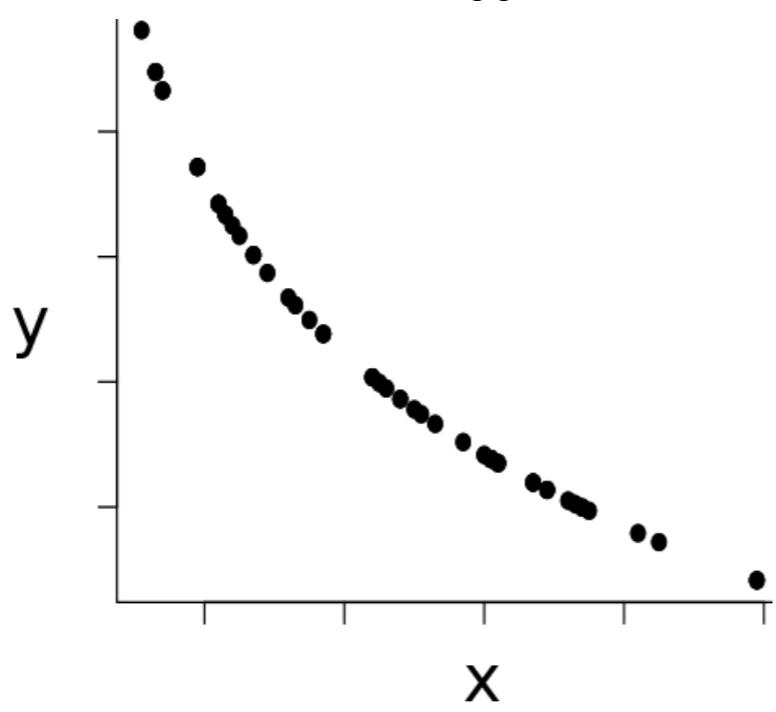
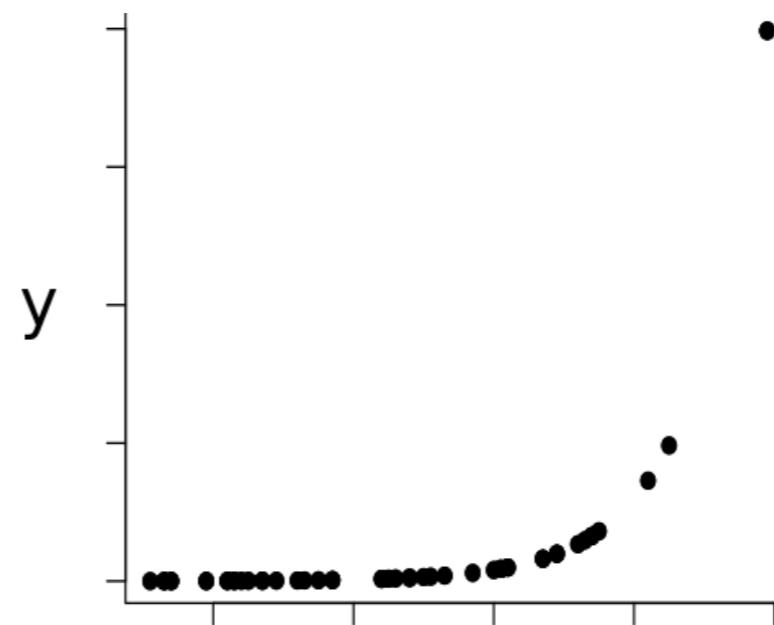
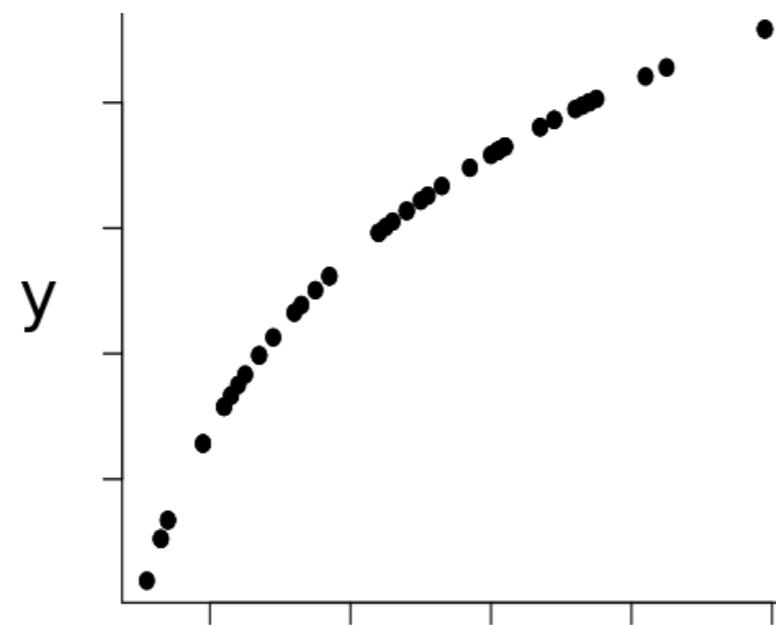
y



y

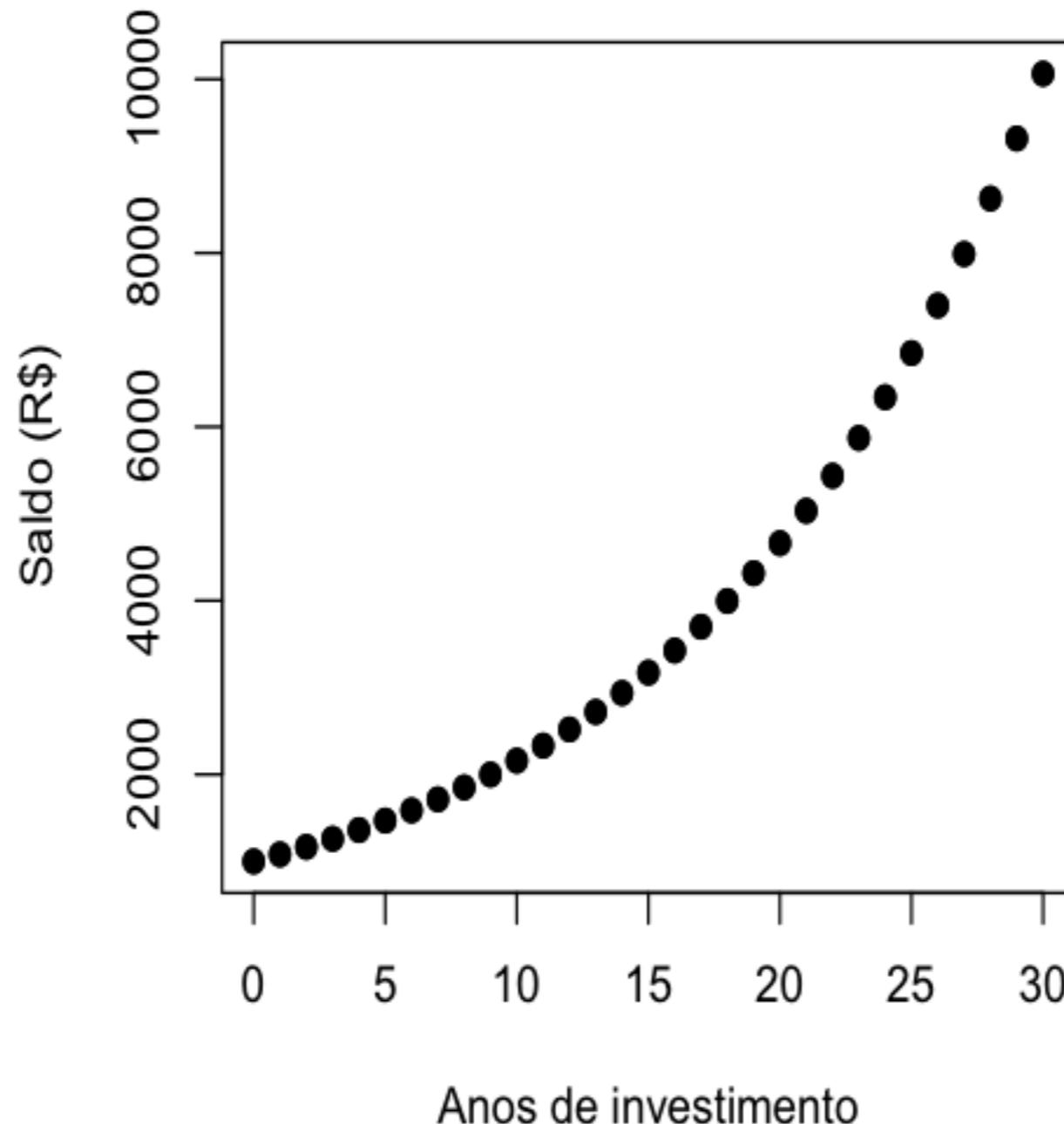
FORMA

Logarítmica ou exponencial?



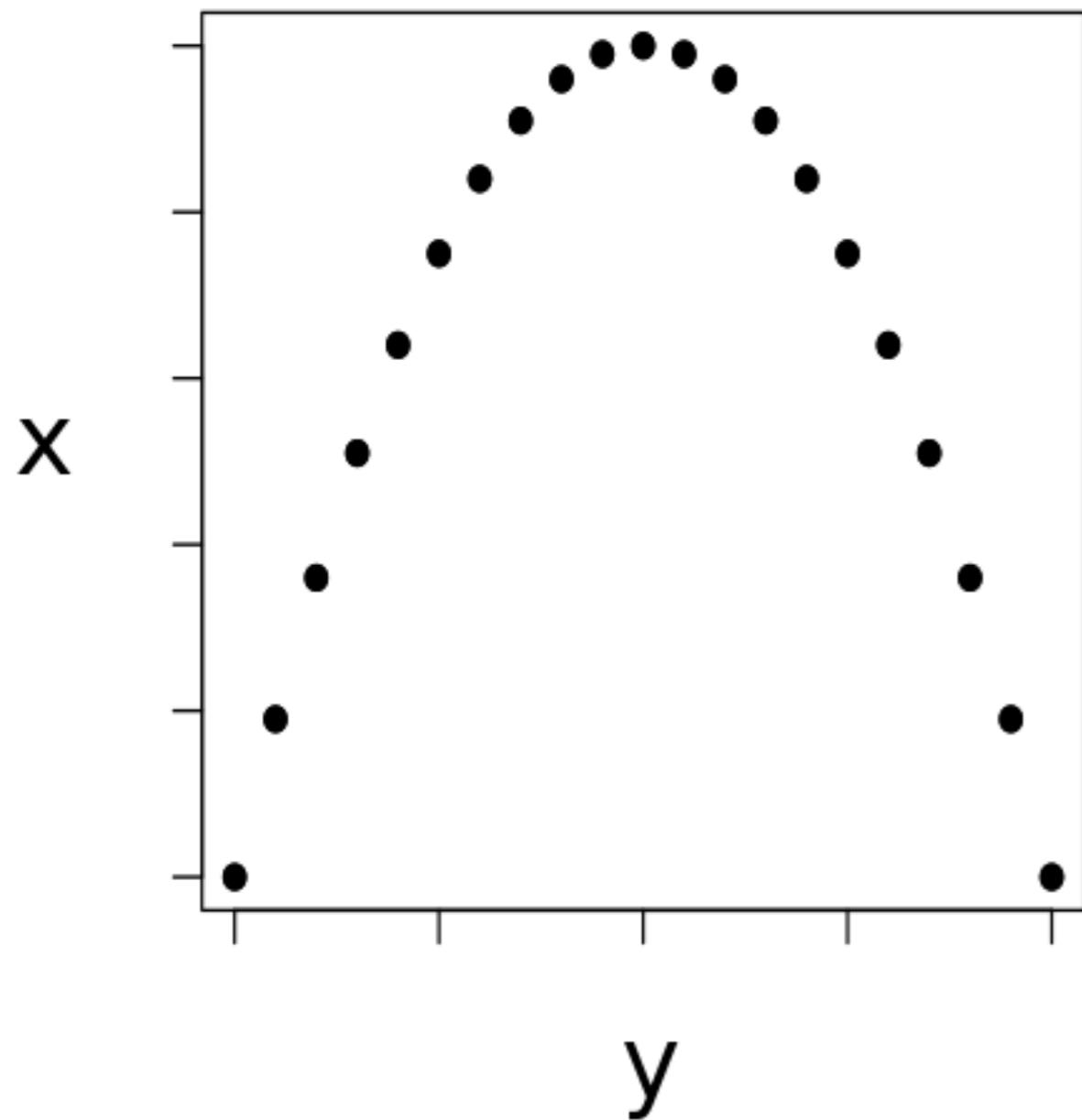
FORMA

R\$1.000,00 a uma taxa de 8% ao ano

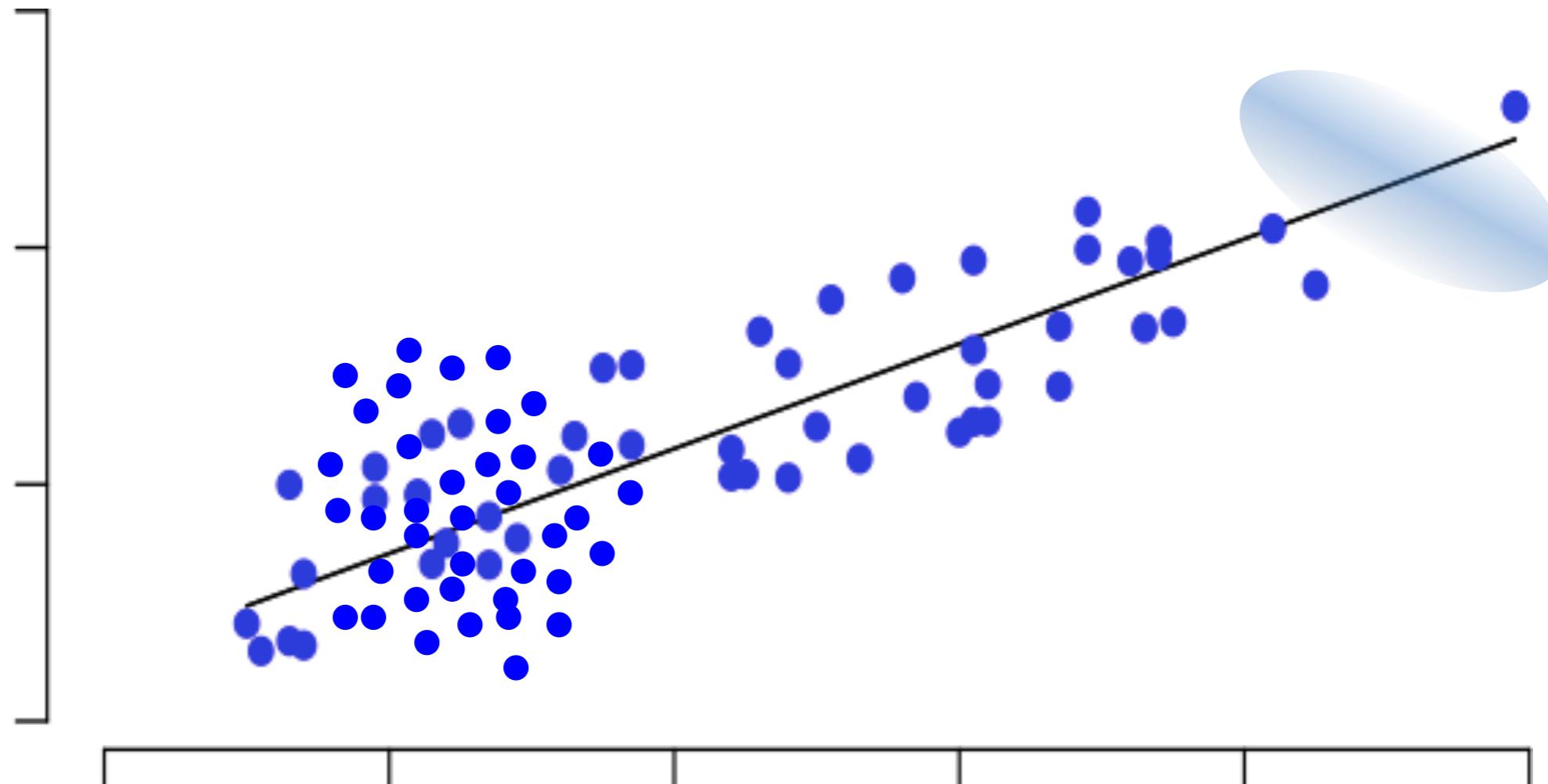


FORMA

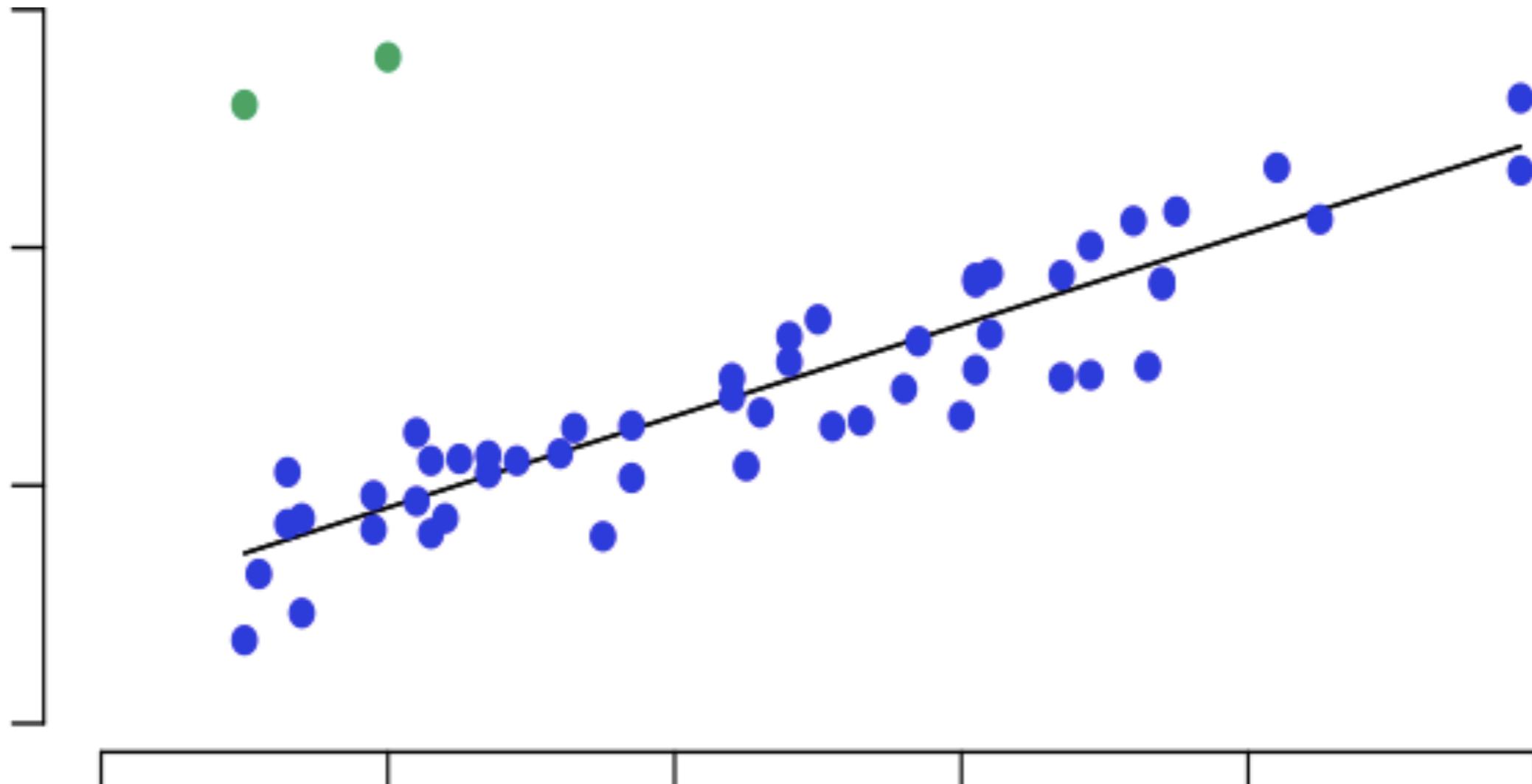
Certas correlações podem ser não-monotônicas



LACUNAS E CONCENTRAÇÕES



EXCEÇÕES



REPRESENTAÇÕES VISUAIS OU TÉCNICAS DE VISUALIZAÇÃO

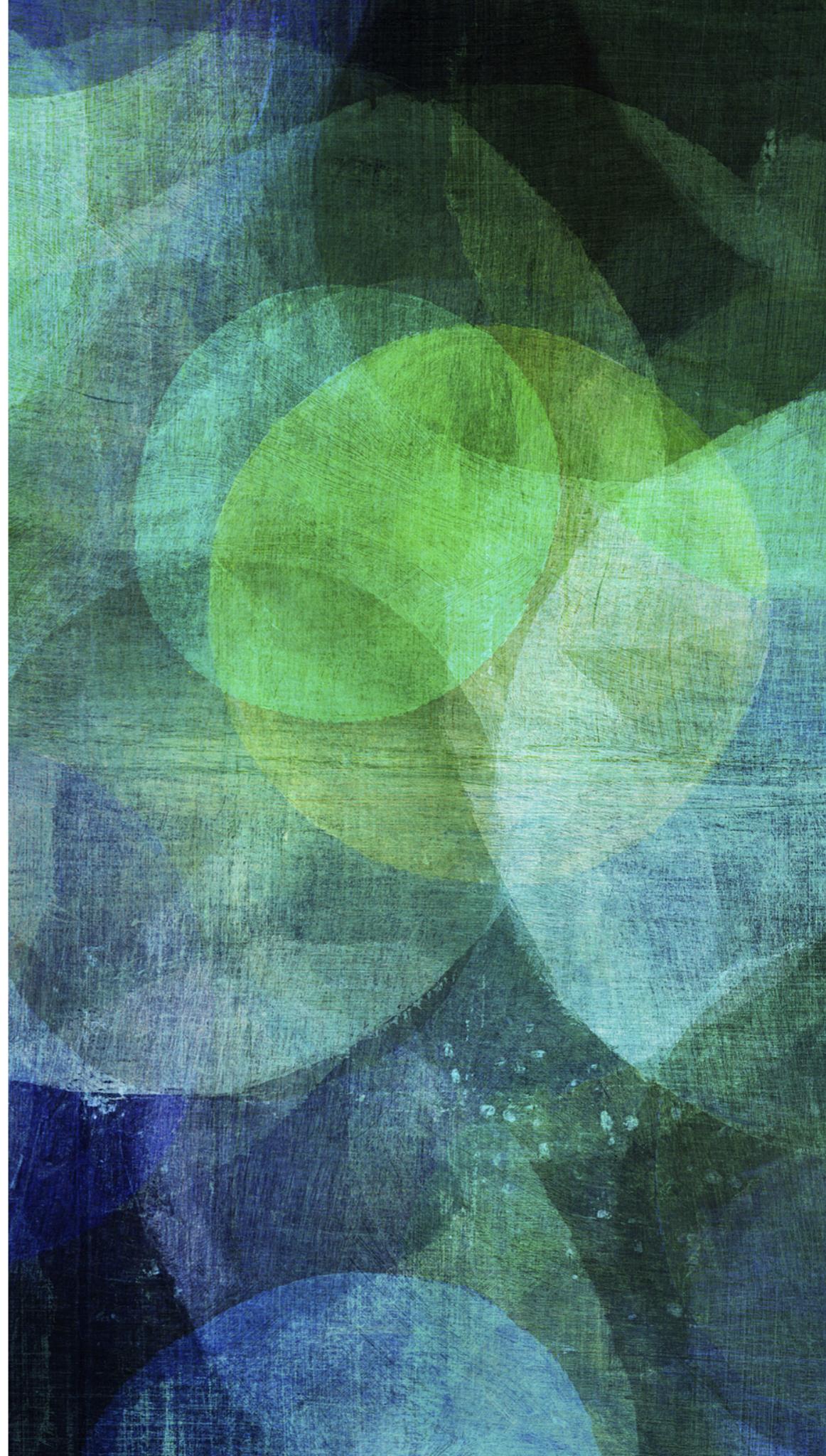
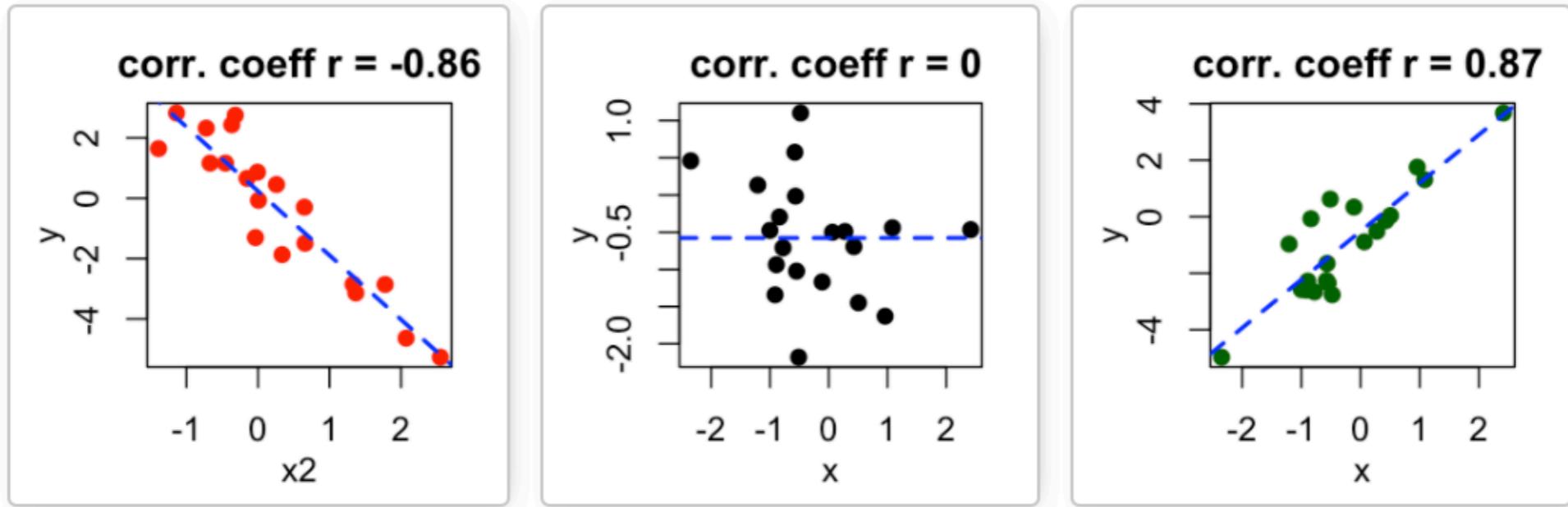
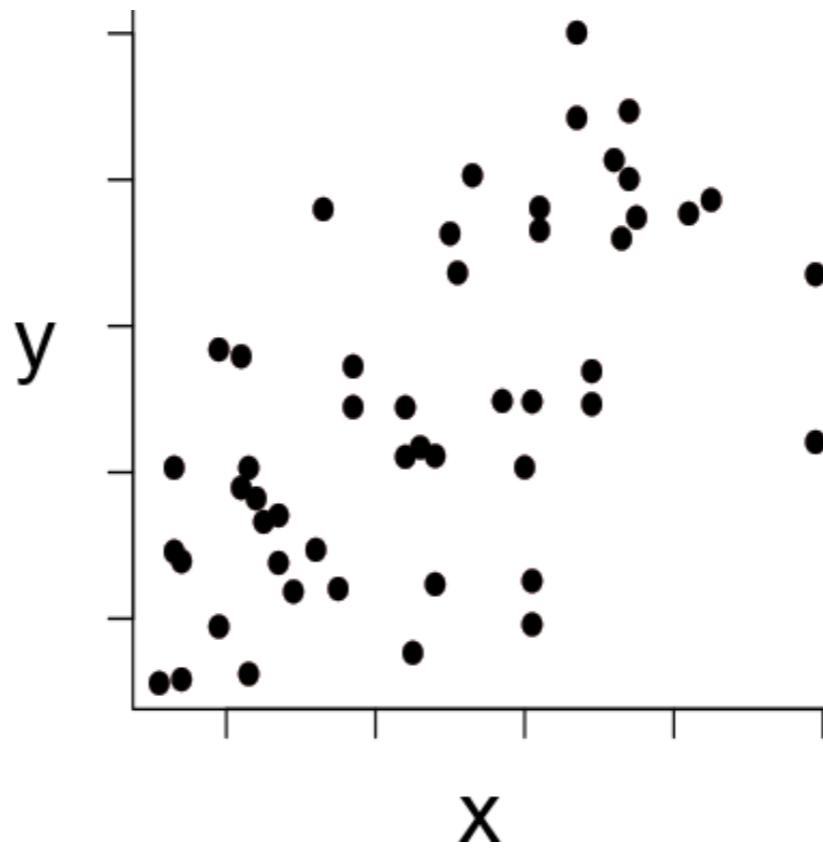
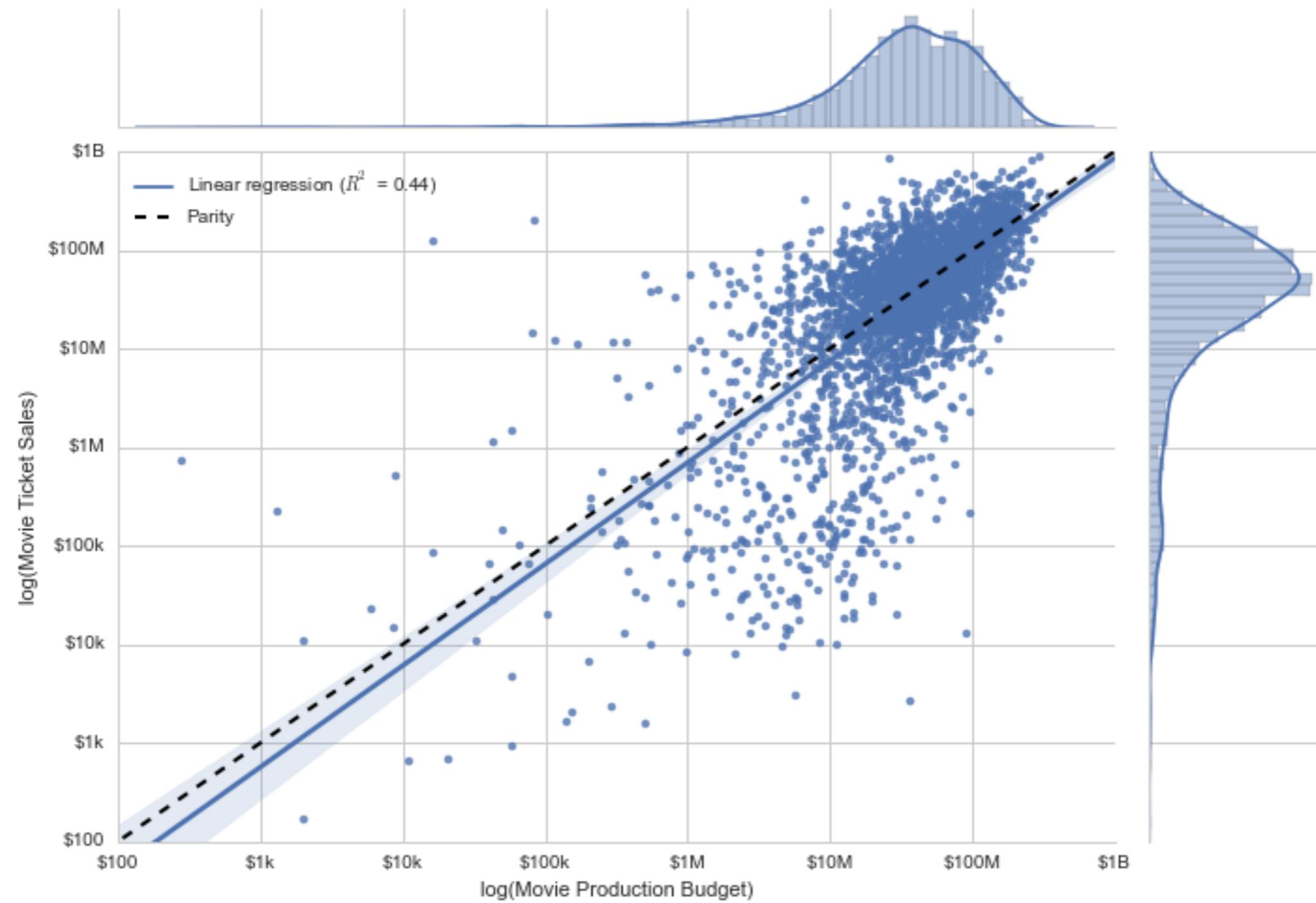
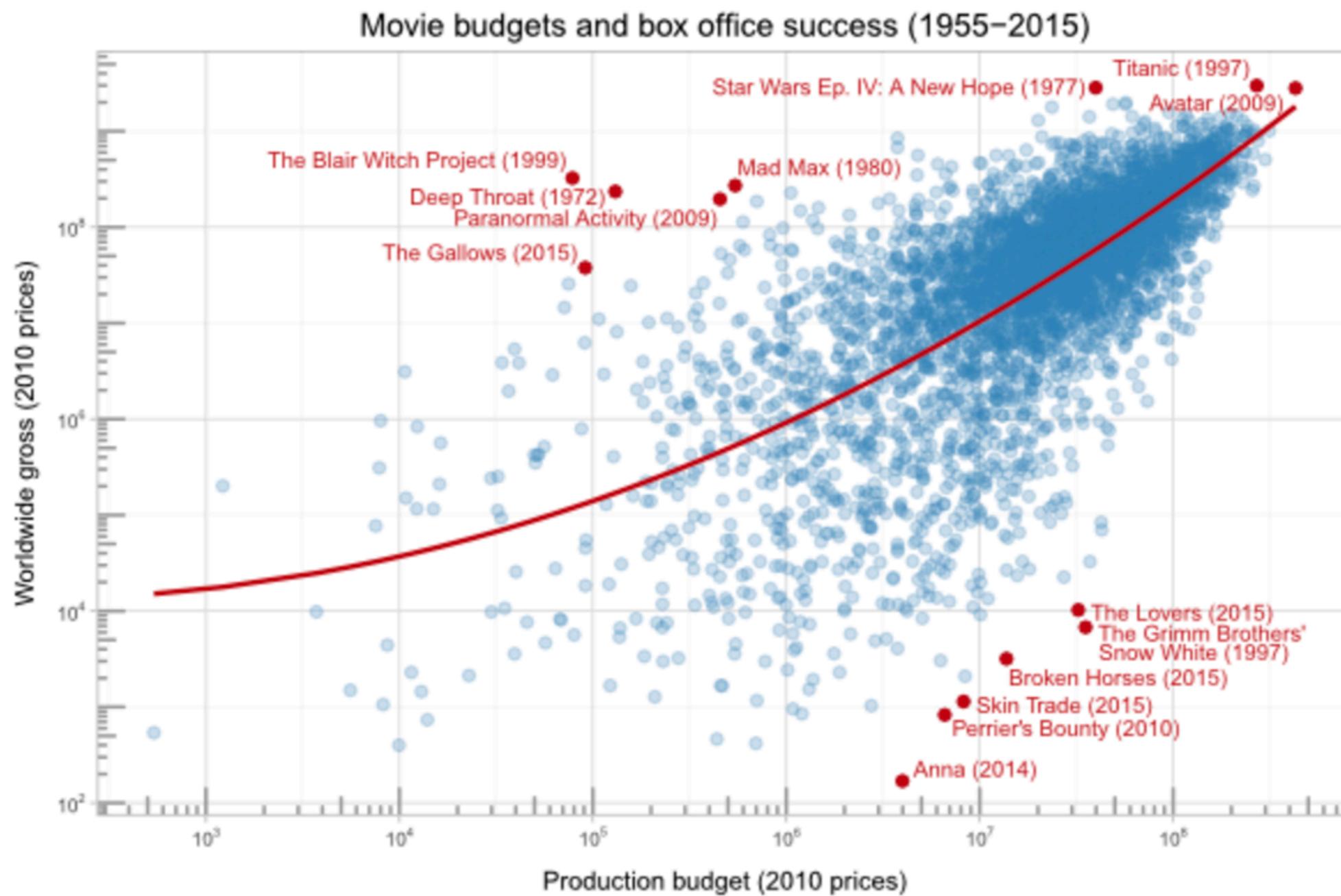


GRÁFICO DE DISPERSÃO

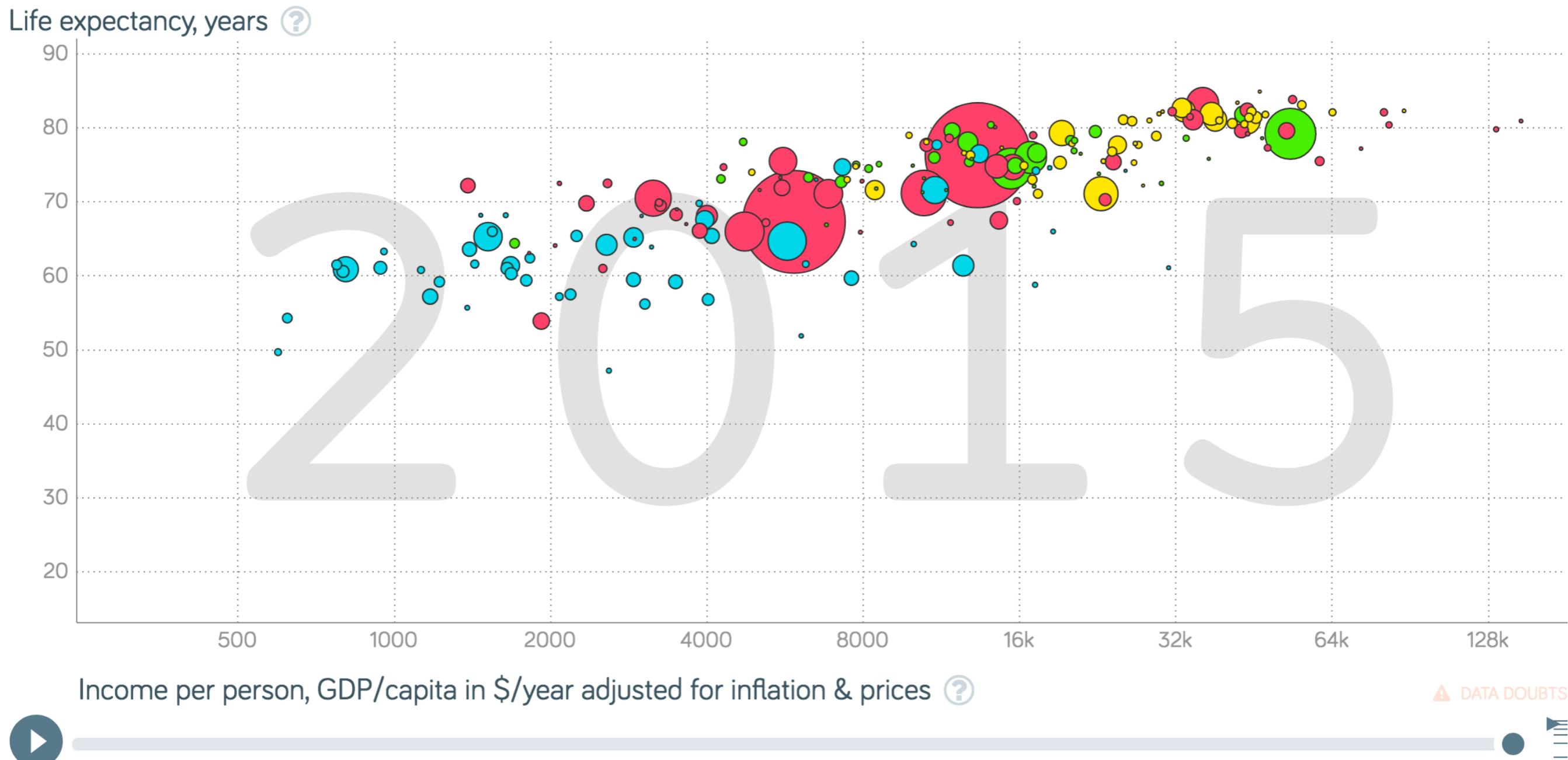
Úteis para analisar a correlação entre duas variáveis







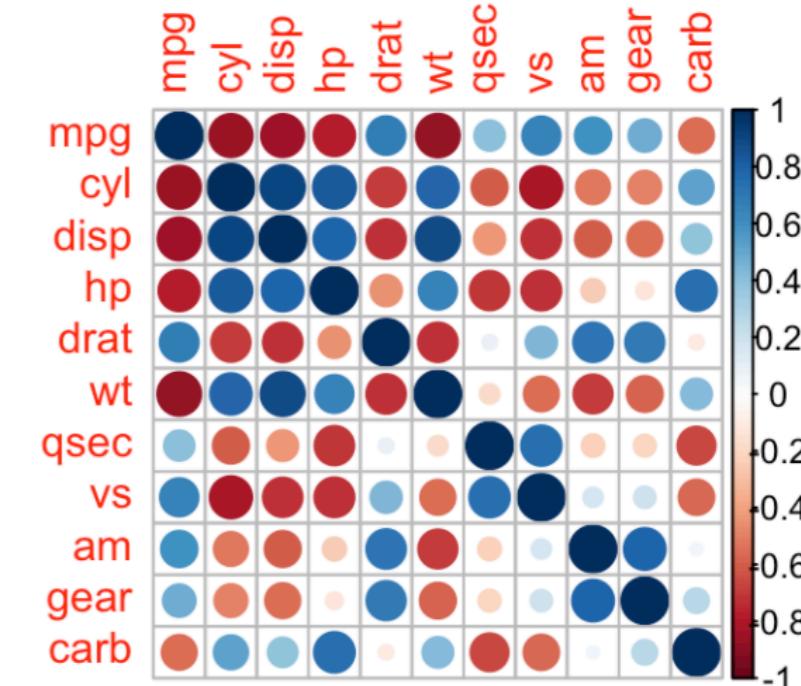
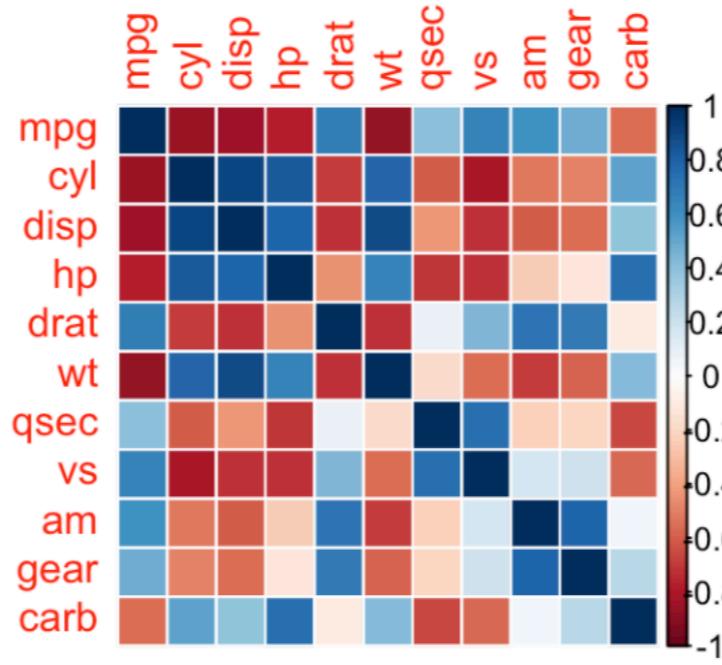
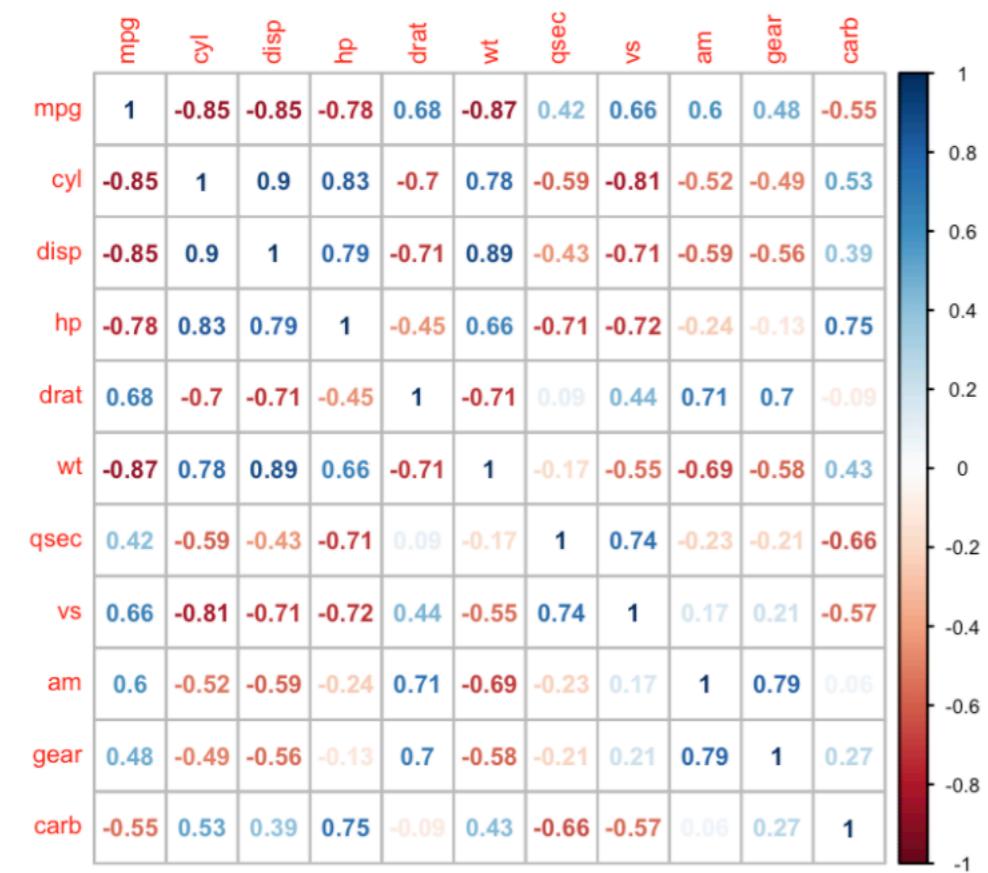
MOTION CHART



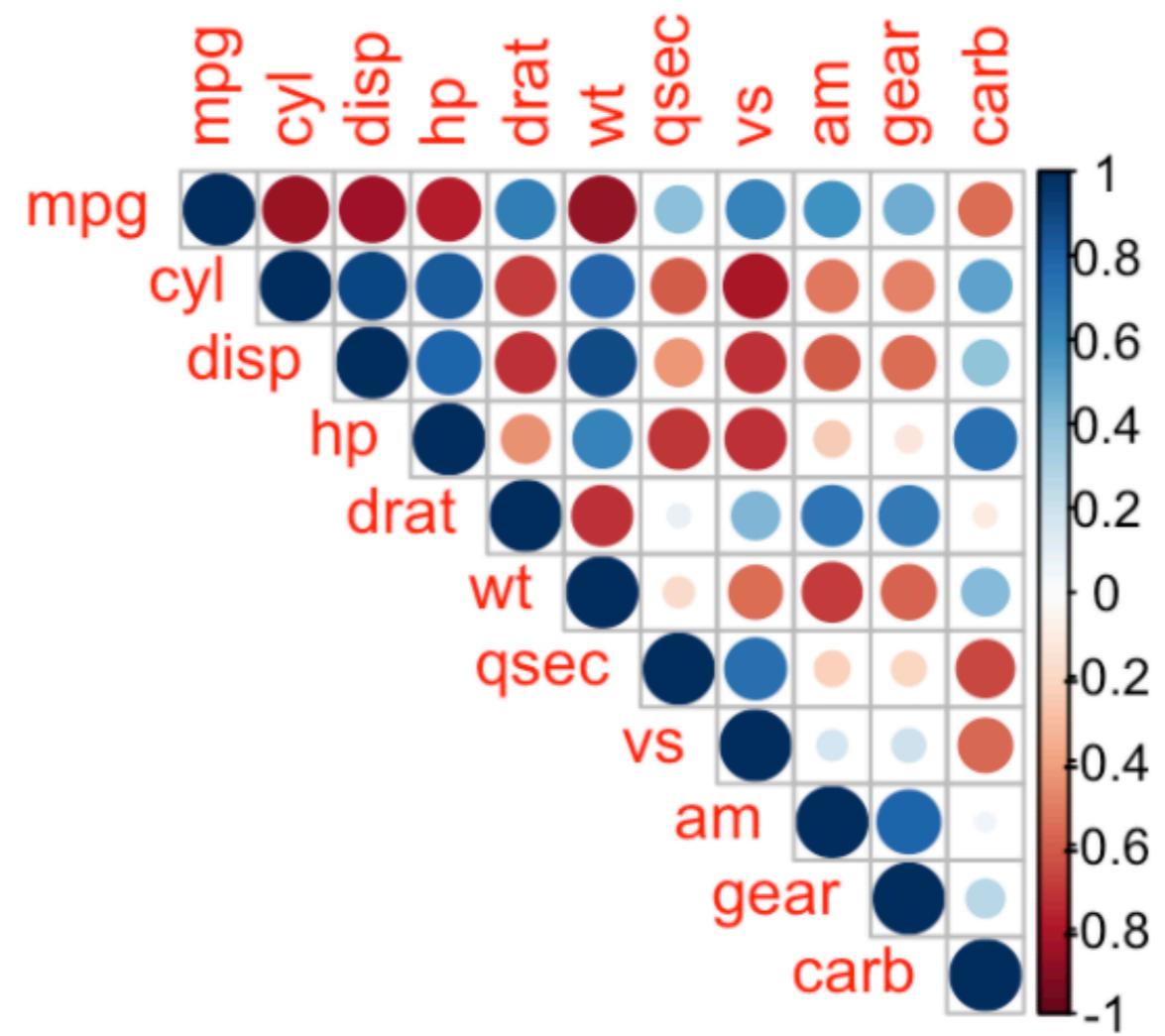
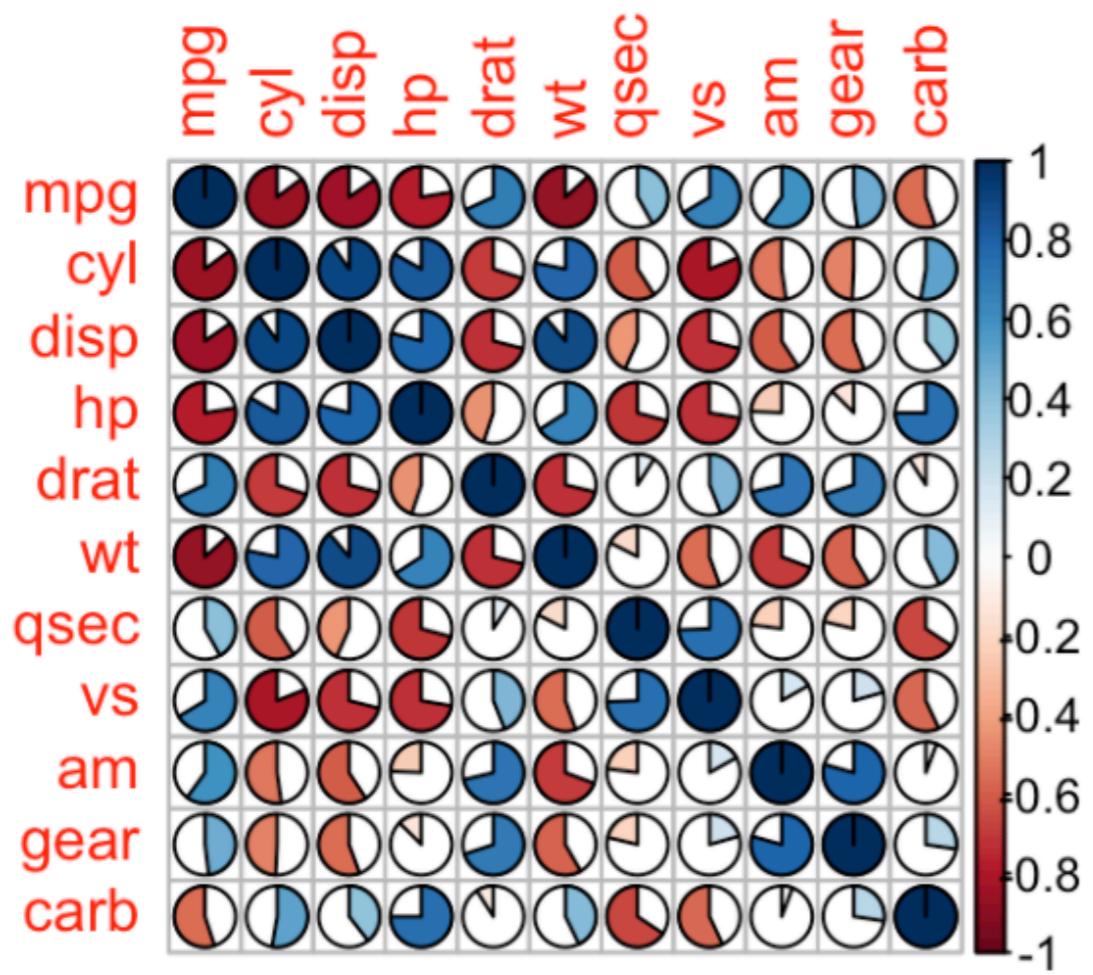
MATRIZES DE CORRELAÇÃO

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

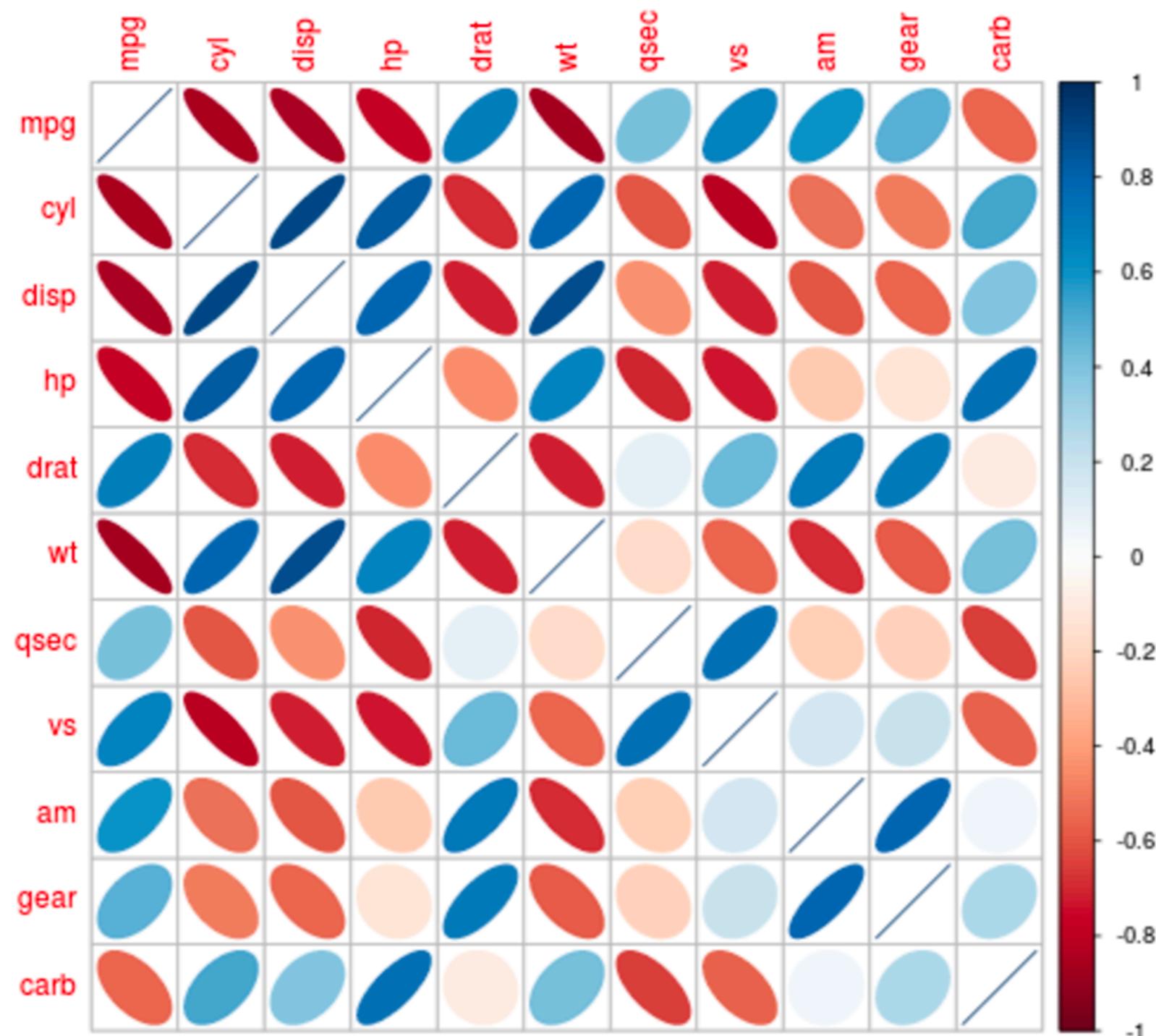
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1.00	-0.85	-0.85	-0.78	0.68	-0.87	0.42	0.66	0.60	0.48	-0.55
cyl	-0.85	1.00	0.90	0.83	-0.70	0.78	-0.59	-0.81	-0.52	-0.52	-0.49
disp	-0.85	0.90	1.00	0.79	-0.71	0.89	-0.43	-0.71	-0.59	-0.59	-0.56
hp	-0.78	0.83	0.79	1.00	-0.45	0.66	-0.71	-0.72	-0.24	-0.13	0.75
drat	0.68	-0.7	-0.71	-0.45	1	-0.71	0.09	0.44	0.71	0.7	-0.09
wt	-0.87	0.78	0.89	0.66	-0.71	1	-0.17	-0.55	-0.69	-0.58	0.43



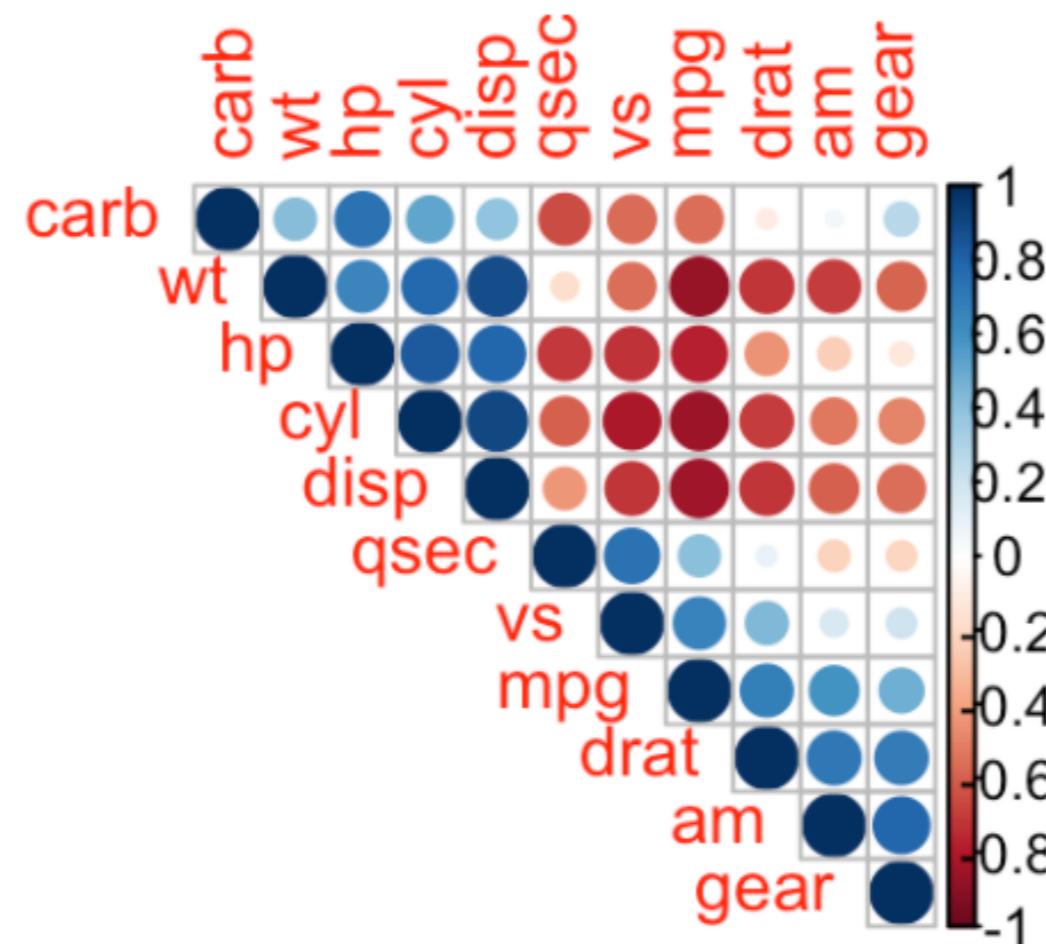
MATRIZES DE CORRELAÇÃO - VARIAÇÕES



MATRIZES DE CORRELAÇÃO - VARIAÇÕES

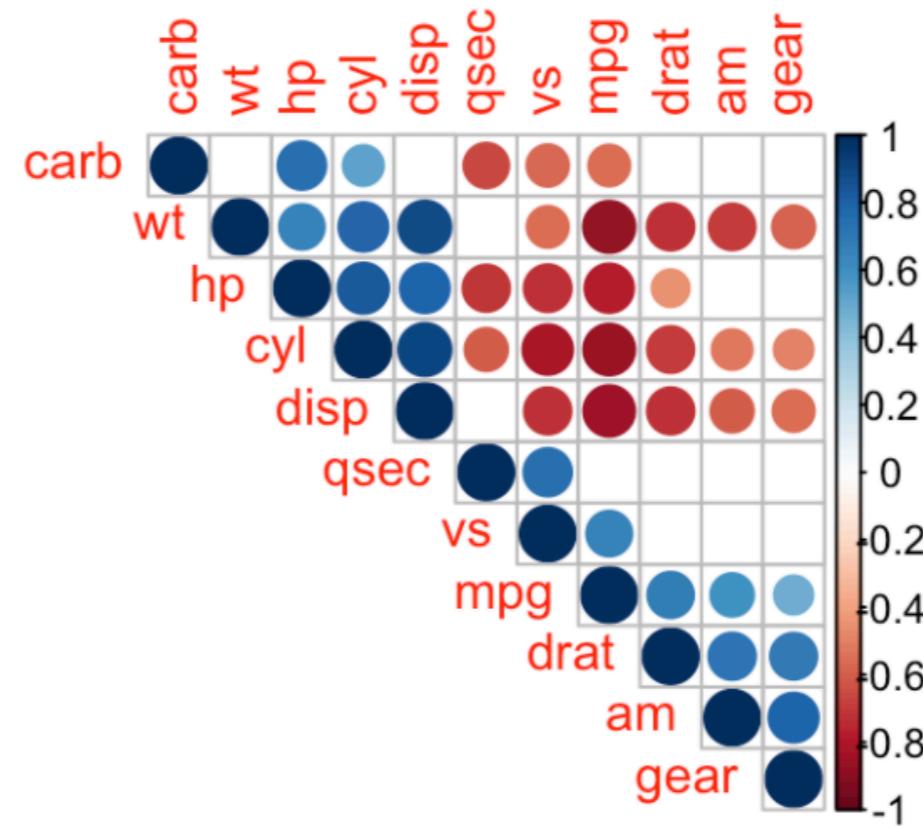
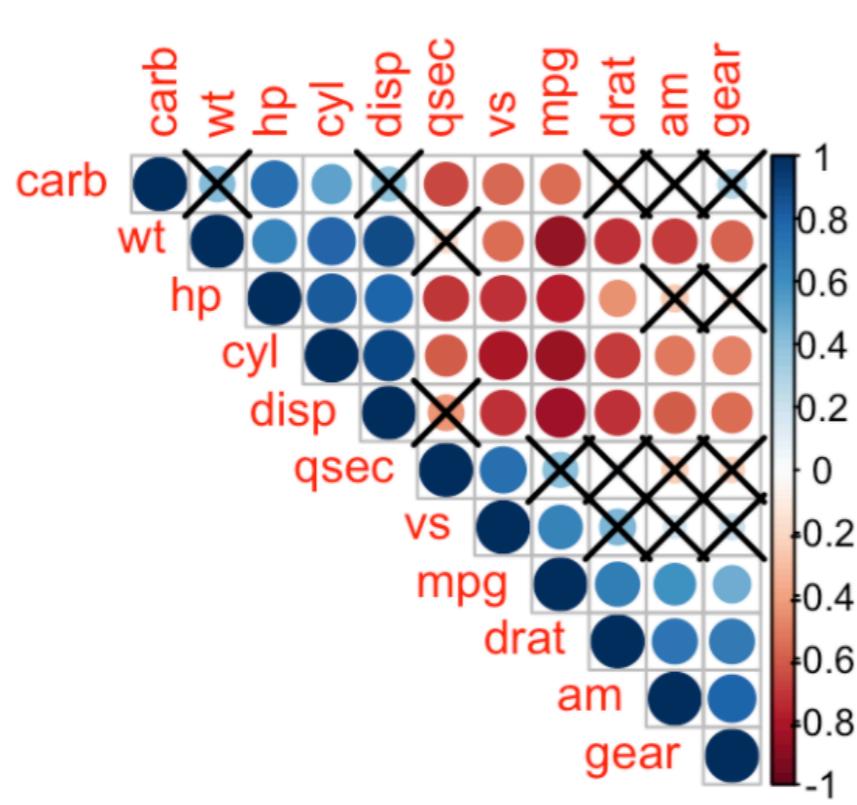


MATRIZES DE CORRELAÇÃO - AGRUPAMENTO DE COLUNAS

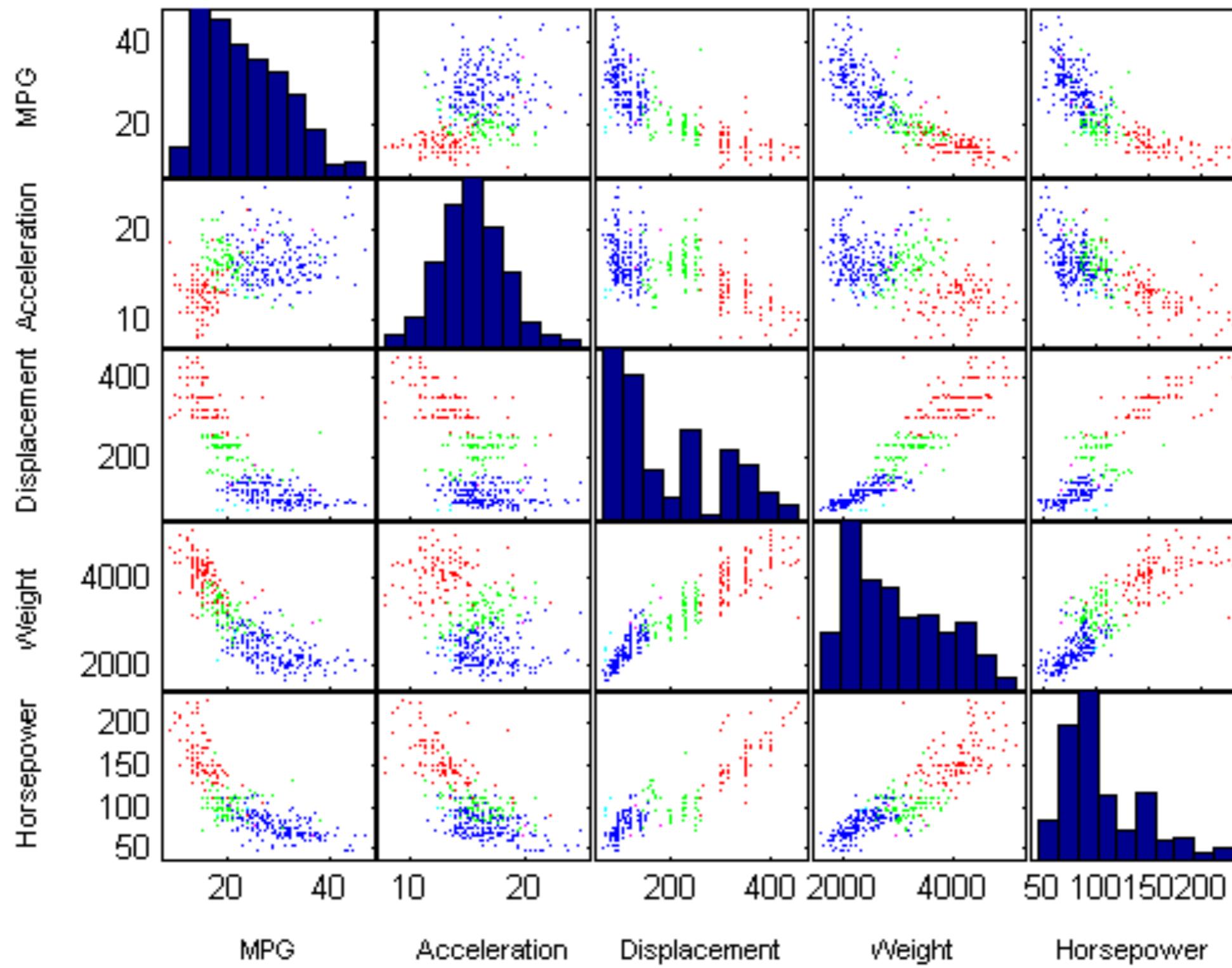


MATRIZES DE CORRELAÇÃO - NÍVEL DE SIGNIFICÂNCIA

	mpg	cyl	disp	hp	drat
mpg	0.000e+00	6.113e-10	9.380e-10	1.788e-07	1.776e-05
cyl	6.113e-10	0.000e+00	1.803e-12	3.478e-09	8.245e-06
disp	9.380e-10	1.803e-12	0.000e+00	7.143e-08	5.282e-06
hp	1.788e-07	3.478e-09	7.143e-08	0.000e+00	9.989e-03
drat	1.776e-05	8.245e-06	5.282e-06	9.989e-03	0.000e+00
wt	1.294e-10	1.218e-07	1.222e-11	4.146e-05	4.784e-06



MATRIZES DE GRÁFICOS DE PONTOS



FilmStrips

Movies, money and metadata, visualised.

Choose which data to view with the panel on the right.

Use the mouse-wheel to zoom in and out; click-and-drag to scroll.

Move your cursor over over a strip to see more information.

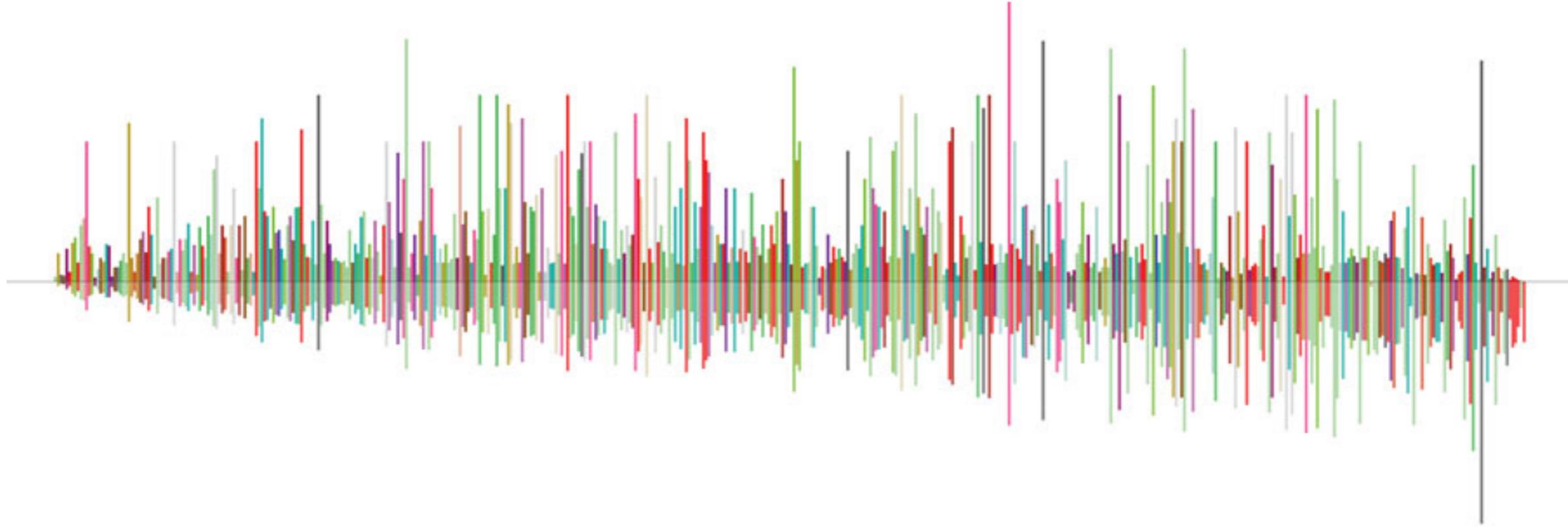
Data by Information is Beautiful Visualisation by Tom Evans

What data would you like to see?

Above the line: ▾

Below the line: ▾

Sort by: ▾



Story Type key

THE TABLE LENS: MERGING GRAPHICAL AND SYMBOLIC REPRESENTATIONS IN AN INTERACTIVE FOCUS+CONTEST VISUALIZATION FOR TABULAR RECOMMENDATION

*Ramana Rao e Stuart Card
Proceedings of the ACM SIGCHI
1994*

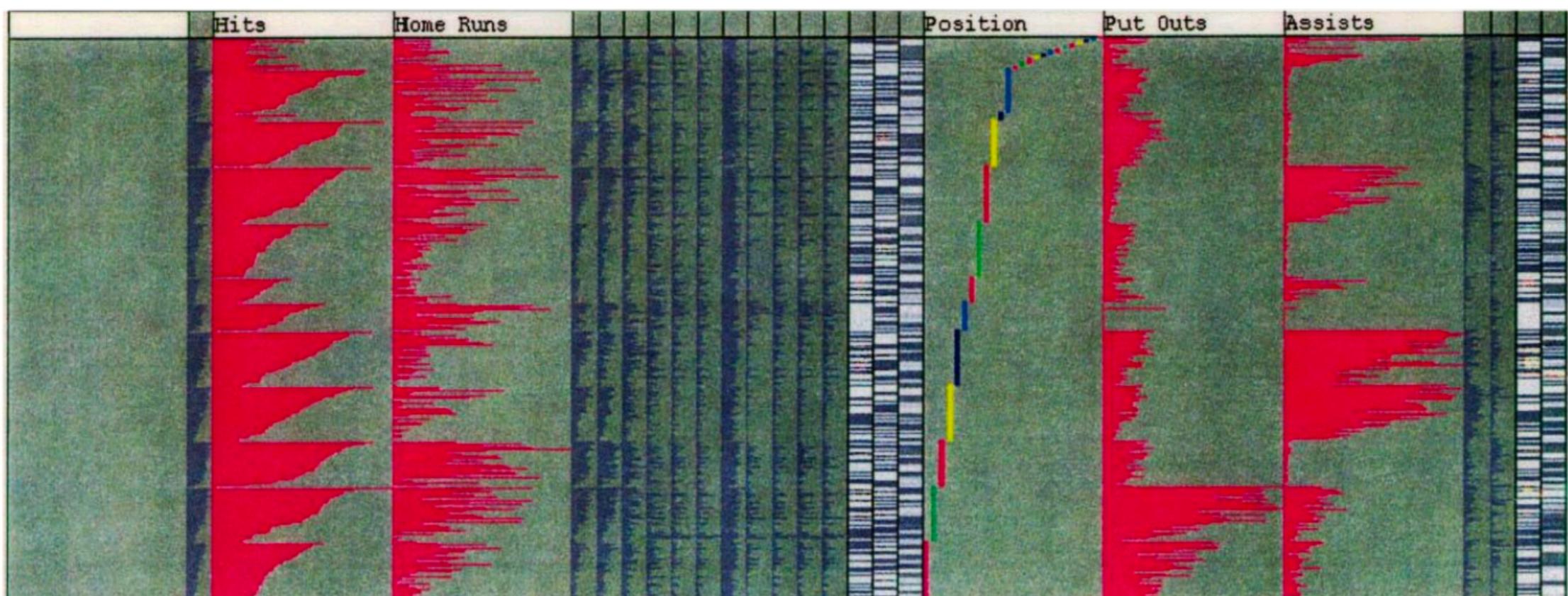
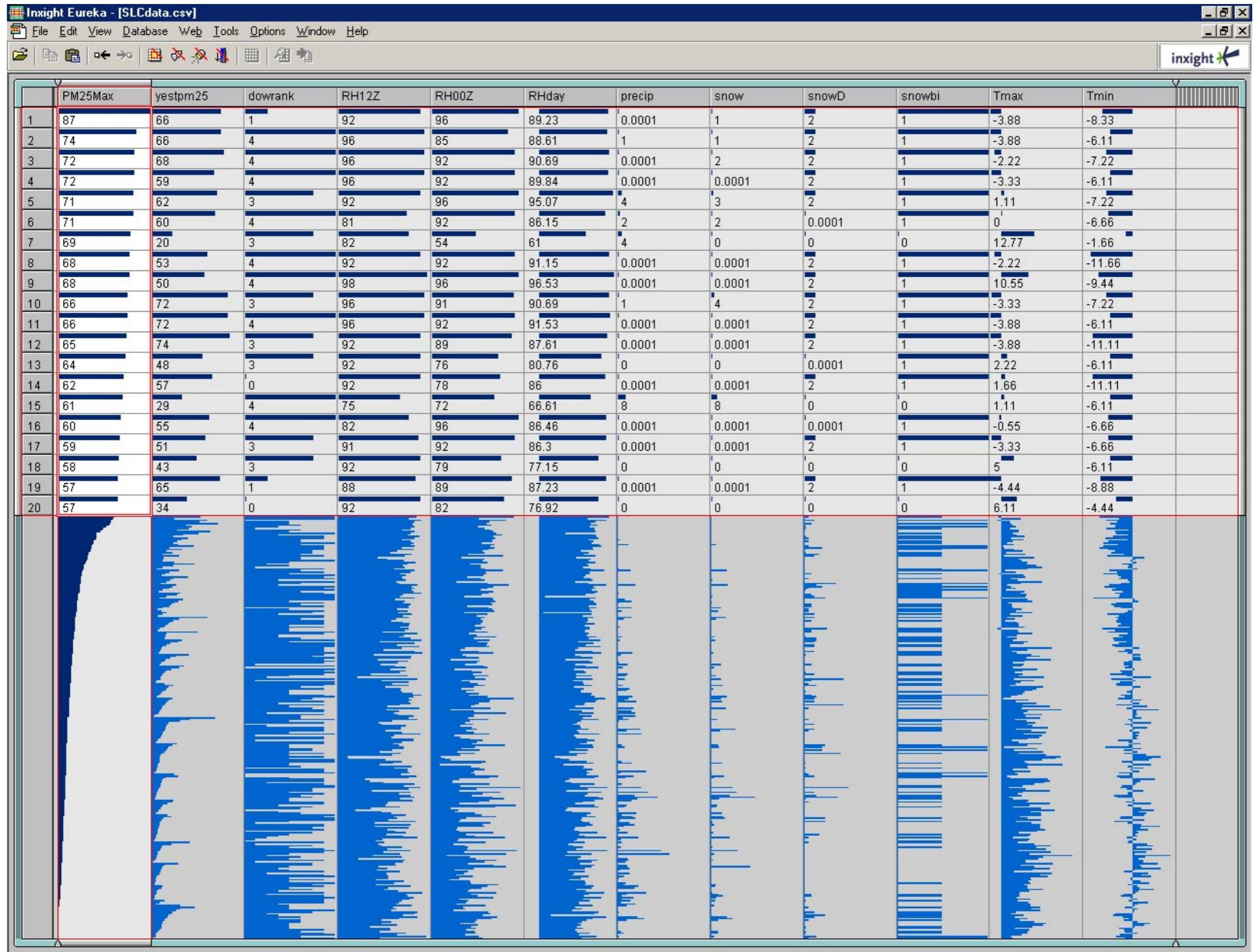
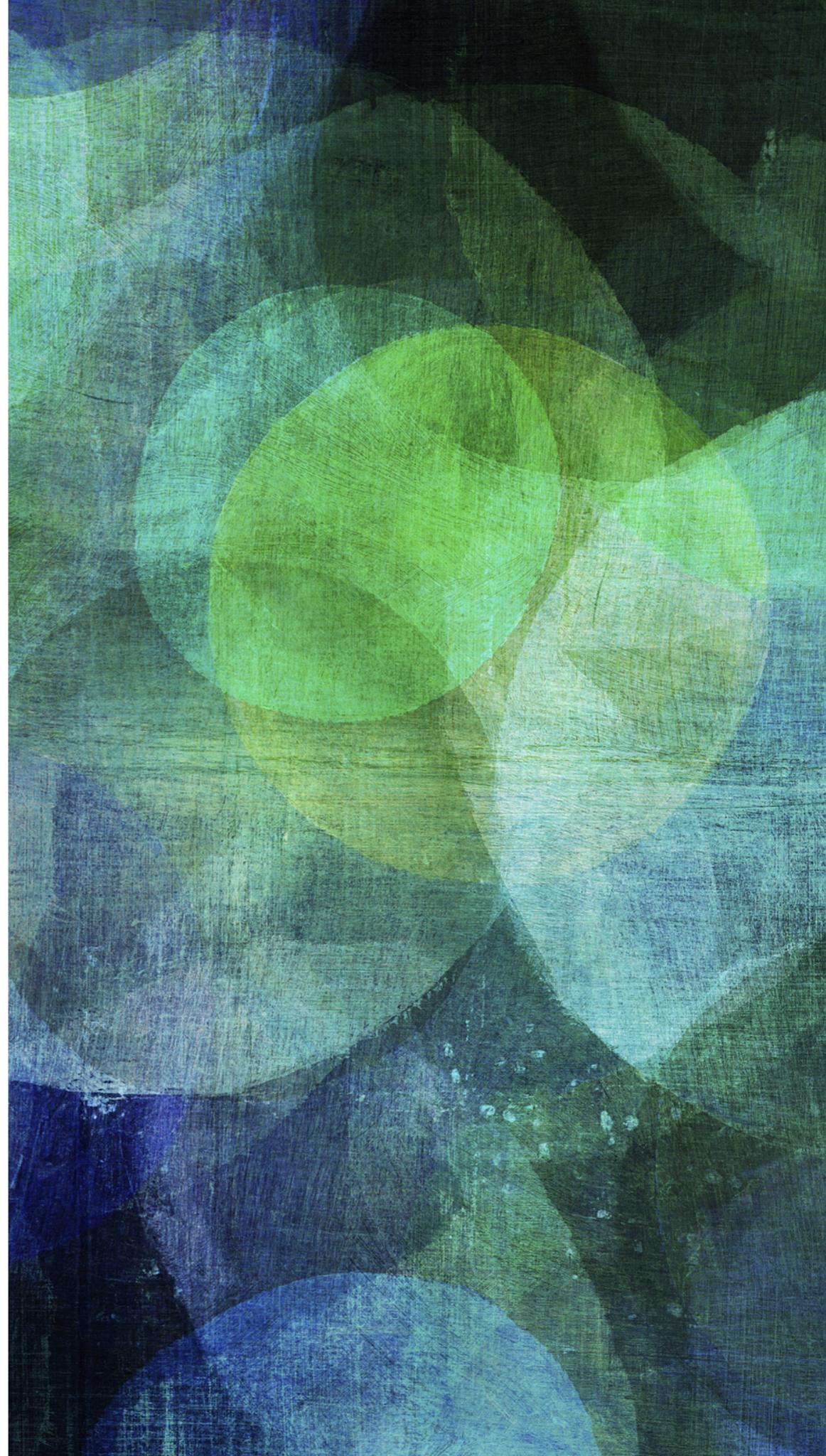


TABLE LENSES



BOAS PRÁTICAS



OTIMIZE A RAZÃO DE ASPECTO

- Em gráficos de pontos para análise de correlações, use sempre a razão do aspecto 1

