Data visualization
4 Exploratory Data Analysis

esempio per capire che la visualizzazone fa
vedere cose che le statistiche non mostrano.

UNIVERSITÀ
DEGLI STUDI
FIRENZE

Beware of descriptive statistics: e.g., Anscombe's quartet
A set of four artificially designed datasets with (almost) identical statistics

1		2		3		4	
X1	Y1	X2	Y2	X3	Y3	X4	Y4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

[Source: Wikipedia]

UNIVERSITÀ
DEGLI STUDI
FIRENZE

In tutti i dataset le statische
sono sempre le stesse.

Beware of descriptive statistics: e.g., Anscombe's quartet
A set of four artificially designed datasets with (almost) identical statistics

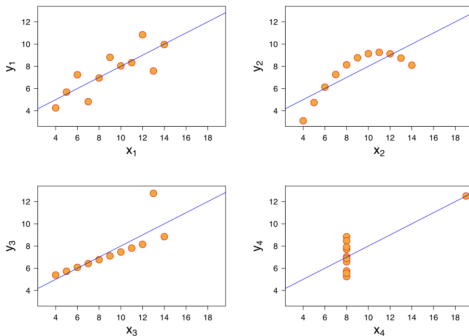| Property	Value	Accuracy
Mean of x	9	Exact
Variance of x	11	Exact
Mean of y	7.50	Up to 2 decimal places
Variance of y	4.125	Plus/Minus 0.003
Correlation between x and y	0.816	Up to 3 decimal places
Linear regression line	y = 5x + 3	Up to 2/3 decimal places

[Source: Wikipedia]

UNIVERSITÀ
DEGLI STUDI
FIRENZE

I plot però sono totalmente
diverso per i quattro dataset

Beware of descriptive statistics: e.g., Anscombe's quartet
A set of four artificially designed datasets with (almost) identical statistics

[Source: Wikipedia]