

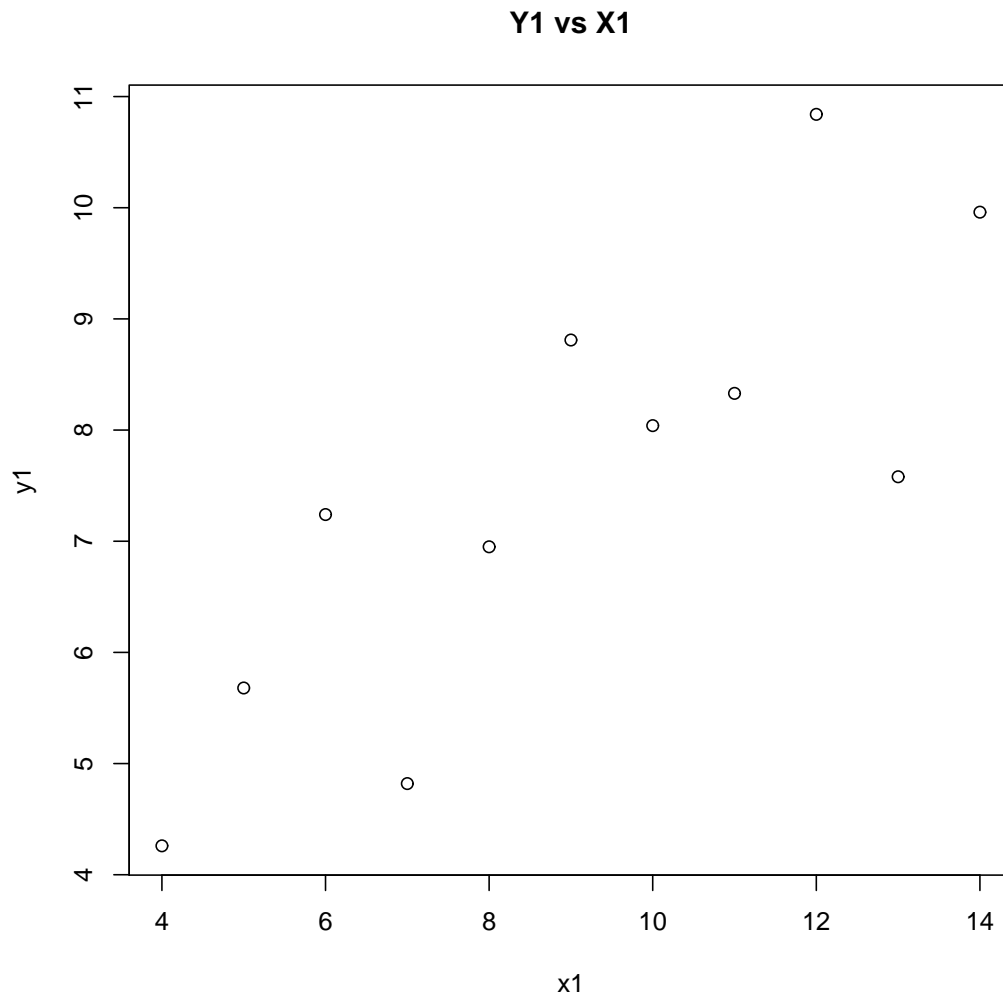
# Set 3: Paired data, Sections 2.5

Scatterplots:

- a graphical descriptive statistic
- for paired quantitative data  $(x_1, y_1), \dots, (x_n, y_n)$
- always label axes and provide a title
- focus is on the relationship between  $x$  and  $y$
- scatterplots aid in prediction
- interpolation versus extrapolation

Example:

x	10	8	13	9	11	14	6	4	12	7	5
y	8.04	6.95	7.58	8.81	8.33	9.96	7.24	4.26	10.84	4.82	5.68



Examples: data appropriate for a scatterplot?

(a) Consider 20 patients who take drug 1 and we record their blood pressure ( $x$ 's). There are 20 other patients who take drug 2 and we record their blood pressure ( $y$ 's).

(b) Consider the monthly immigration rates ( $x$ 's) into British Columbia and the monthly emigration rates from British Columbia ( $y$ 's).

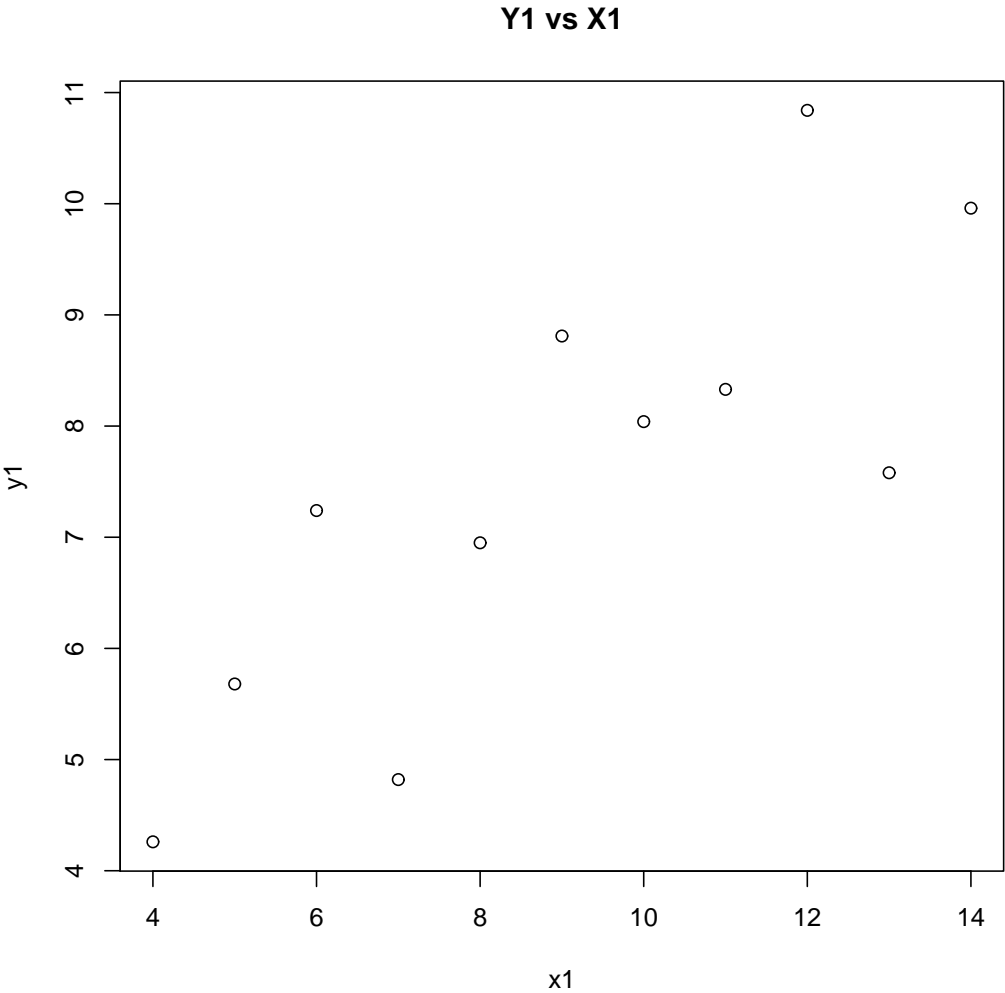
(c) We consider 10 different colours. In a neighbourhood, we count the number of houses of each colour.

Sample correlation coefficient  $r$ :

- a numerical descriptive statistic
- for paired quantitative data  $(x_1, y_1) \dots, (x_n, y_n)$
- $r$  describes linearity between  $x$  and  $y$
- $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$
- $r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)}}$  Computing Formula

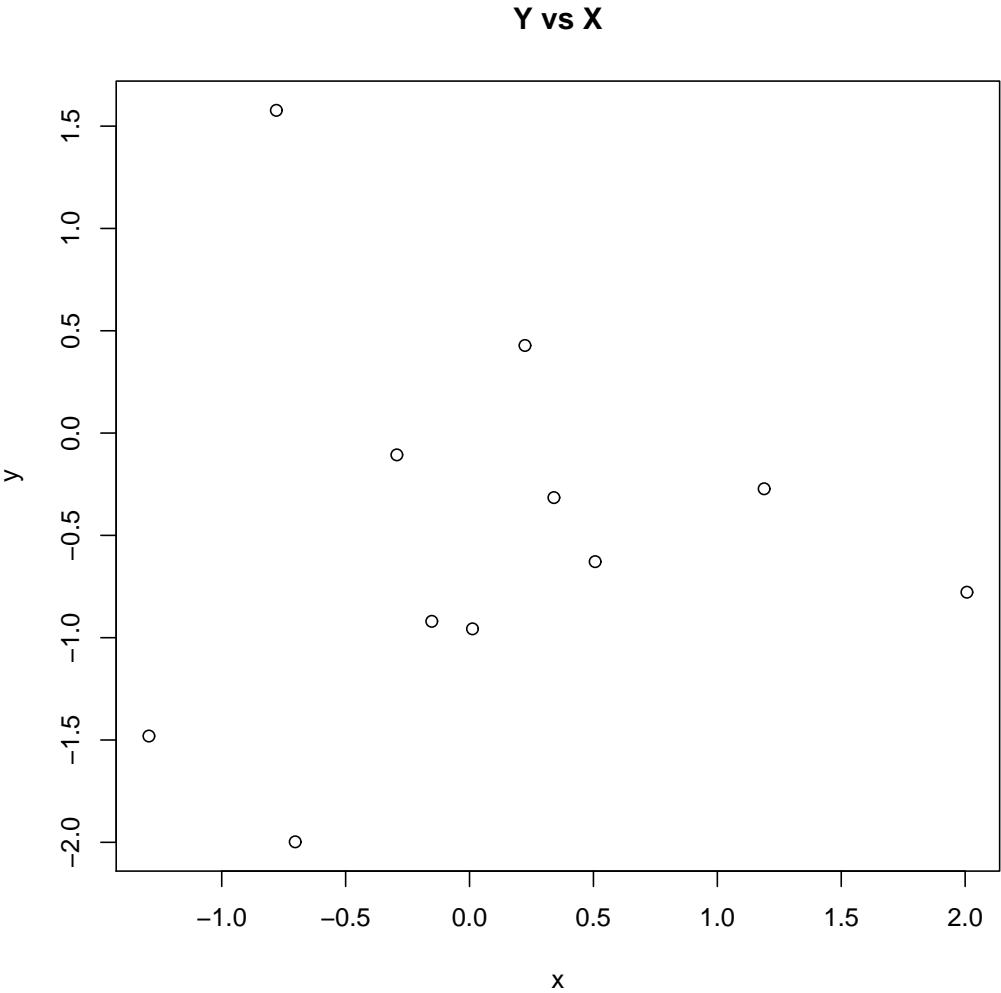
Example:  $r=.82$

x	10	8	13	9	11	14	6	4	12	7	5
y	8.04	6.95	7.58	8.81	8.33	9.96	7.24	4.26	10.84	4.82	5.68



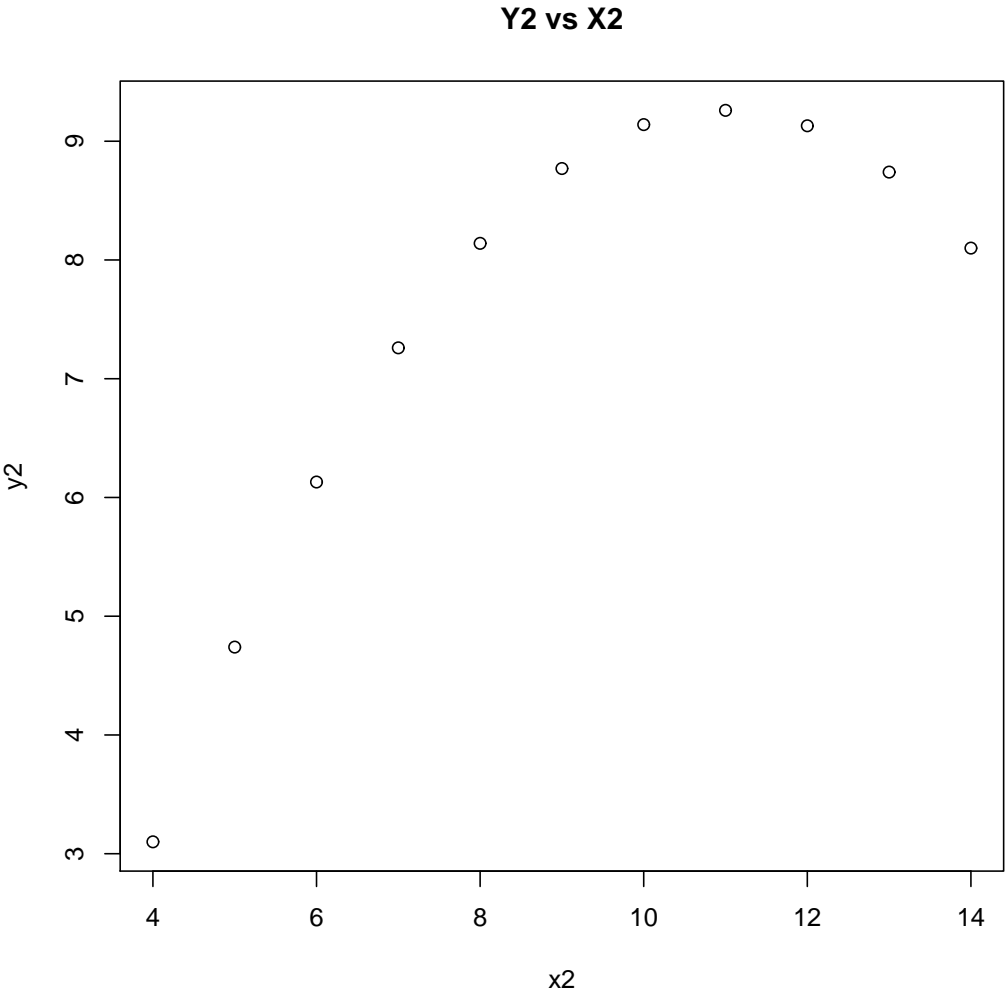
Example:  $r=0.06$

x	-1.29	-.78	.01	-.15	-.70	1.19	.34	.51	-.29	.22	2.01
y	-1.48	1.58	-.96	-.92	-2.00	-.27	-.32	-.63	-.11	.43	-.78



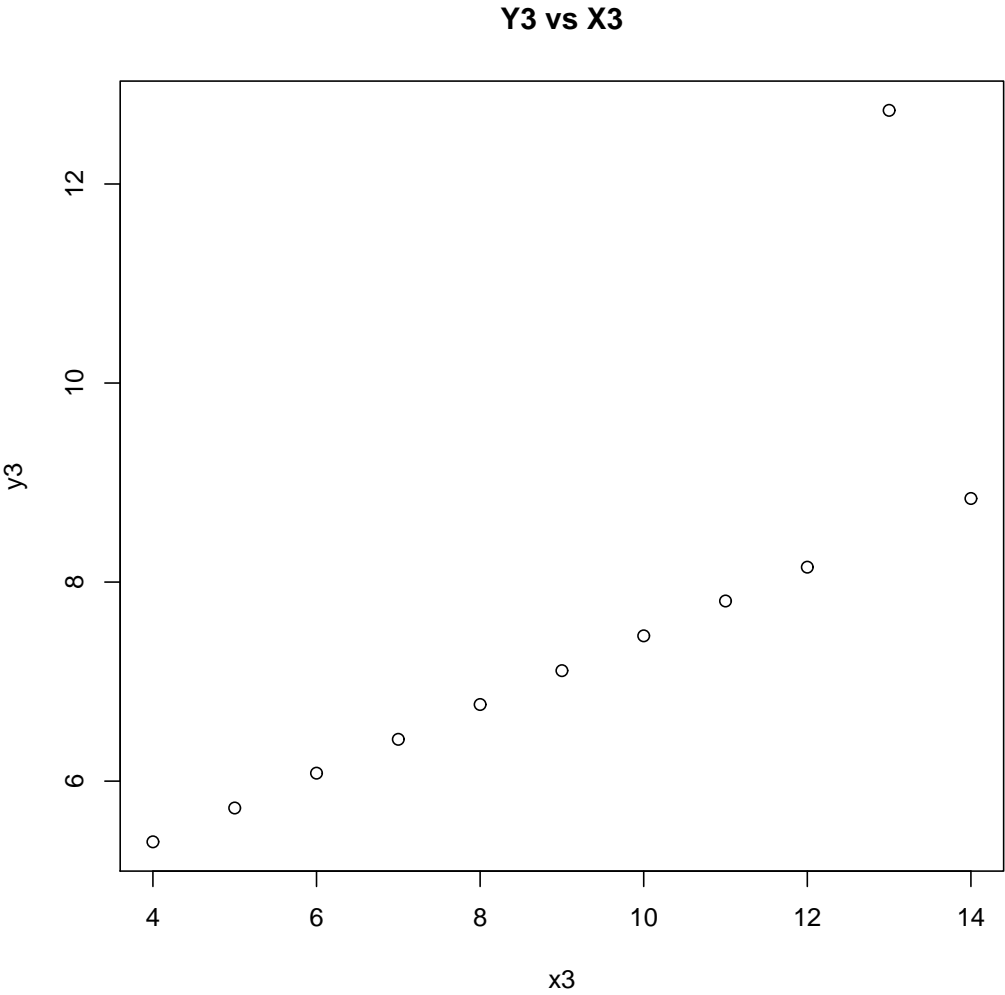
Example:  $r=.82$

x	10	8	13	9	11	14	6	4	12	7	5
y	9.14	8.14	8.74	8.77	9.26	8.1	6.13	3.1	9.13	7.26	4.74



Example:  $r=.82$

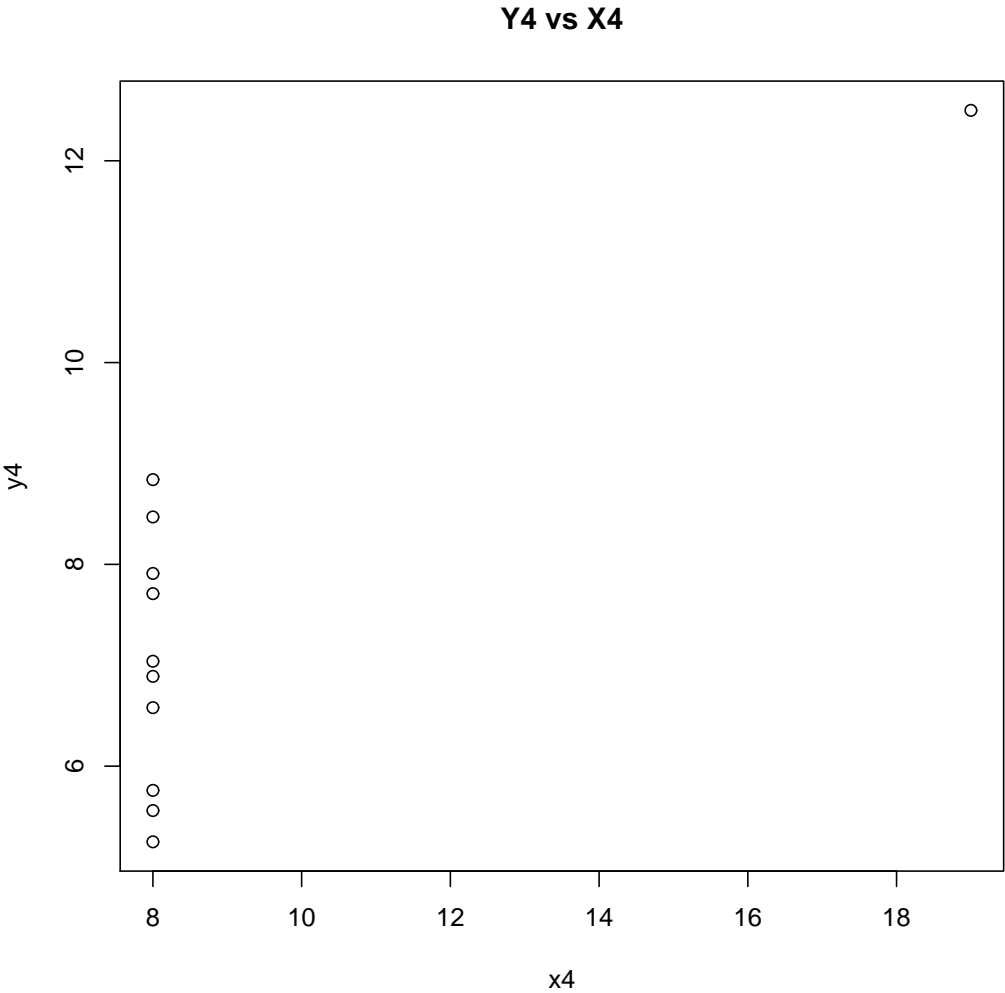
x	10	8	13	9	11	14	6	4	12	7	5
y	7.46	6.77	12.74	7.11	7.81	8.84	6.08	5.39	8.15	6.42	5.73





Example:  $r=.82$

x	8	8	8	8	8	8	8	19	8	8	8
y	6.58	5.76	7.71	8.84	8.47	7.04	5.25	12.5	5.56	7.91	6.89



Association versus cause-effect:

- correlation does not imply causation
- the role of lurking variables in causation
- observational studies
- randomized experiments

Example for discussion: