Correlation and Linear Regression Conceptual Review

Fels Bioinformatics Meetup 2018.11.02
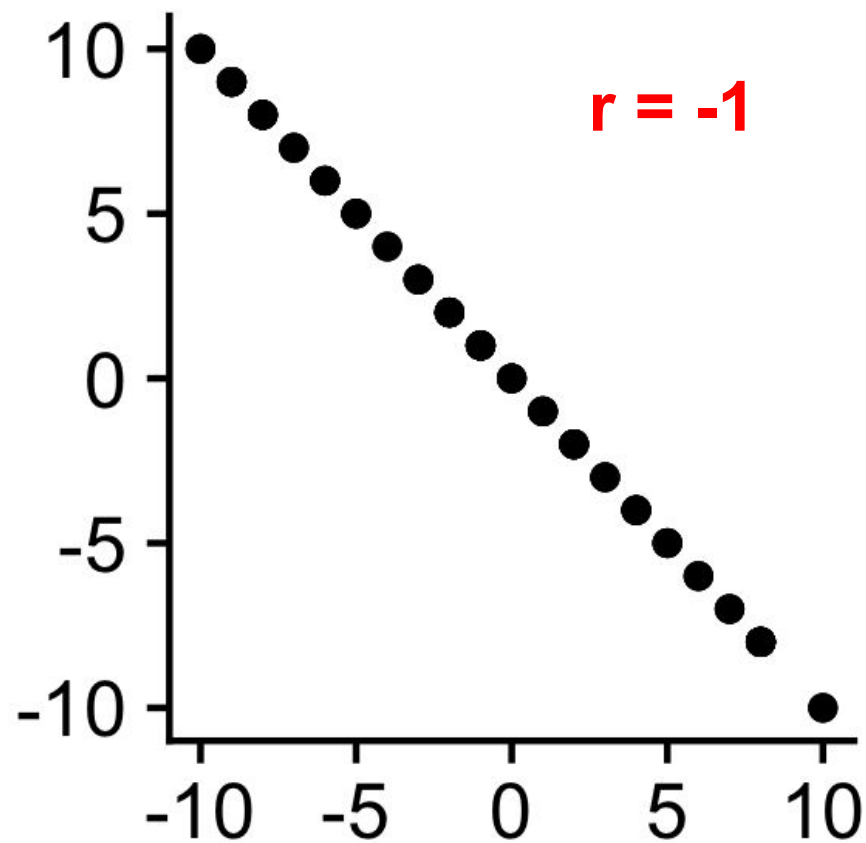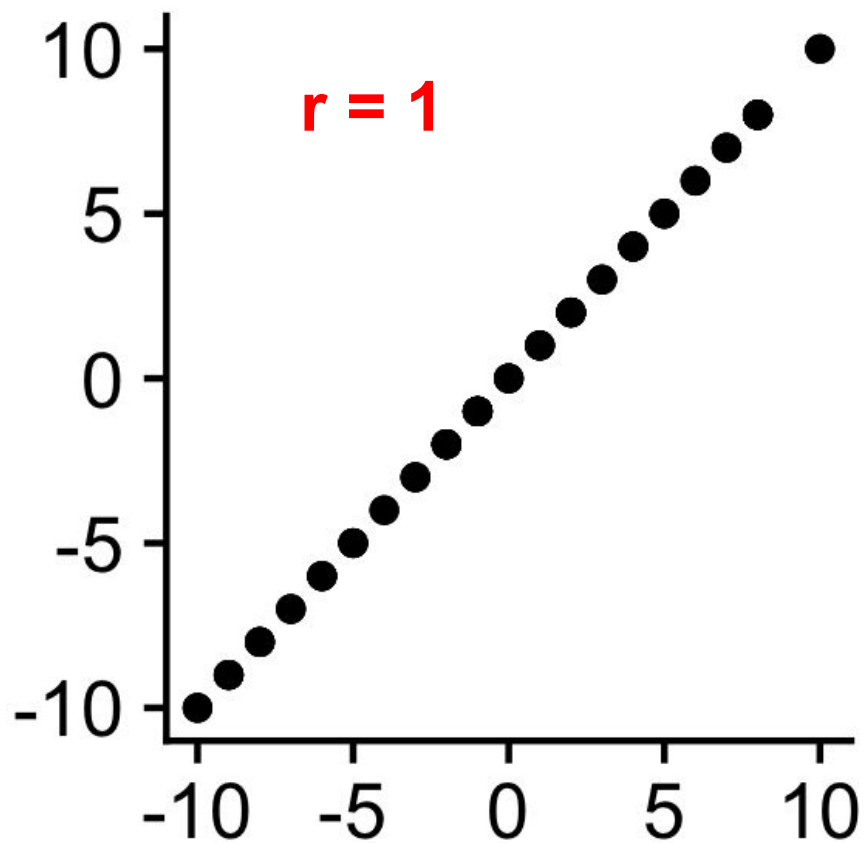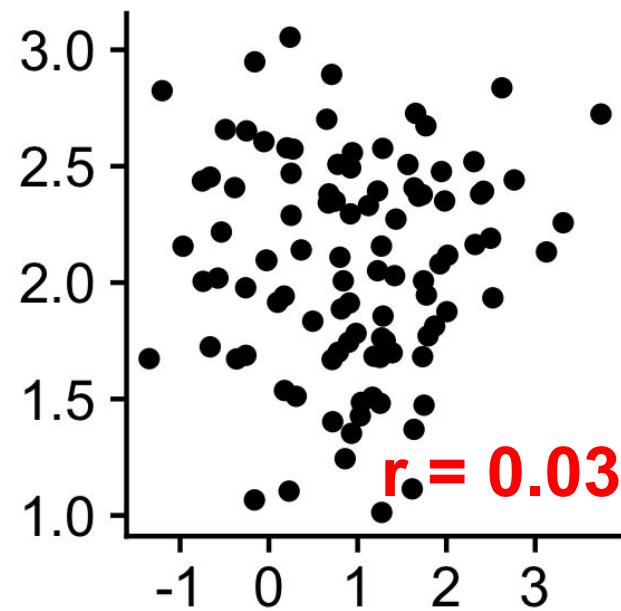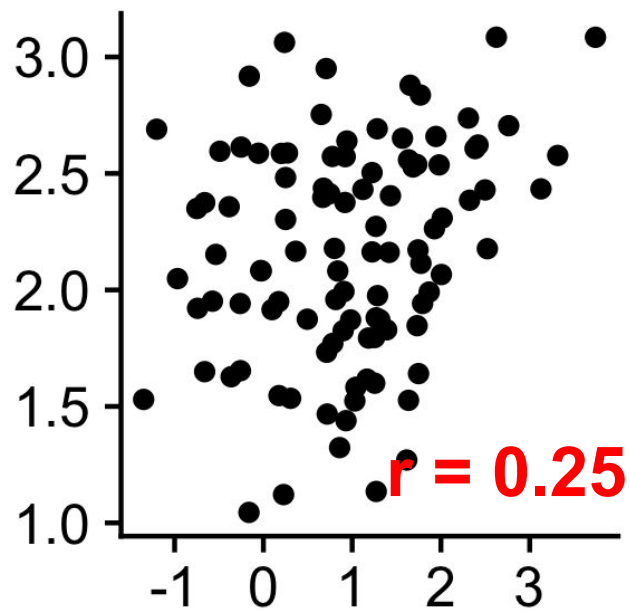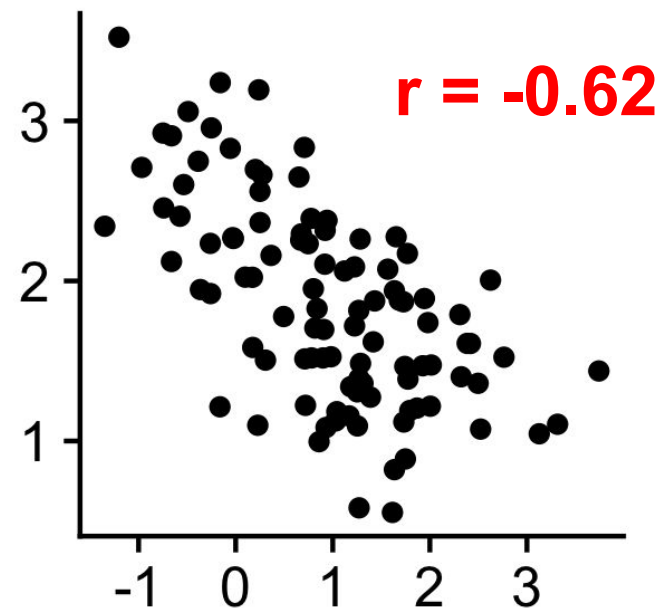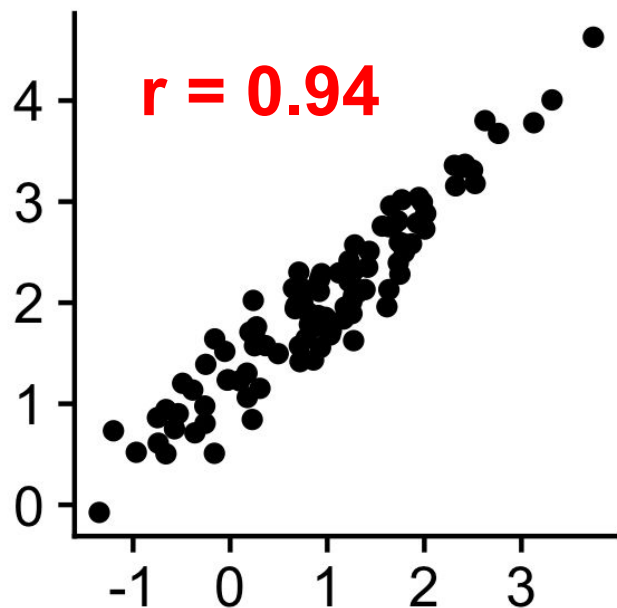
Correlation

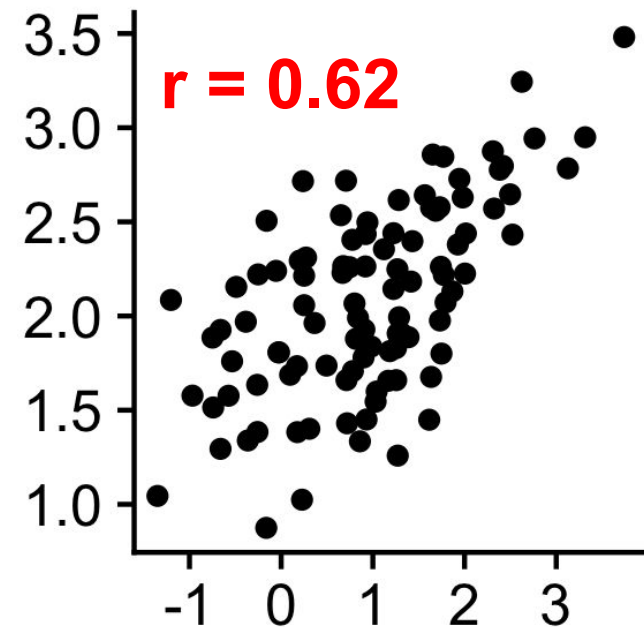Correlation gives you the direction and strength of the linear association between two numeric variables.
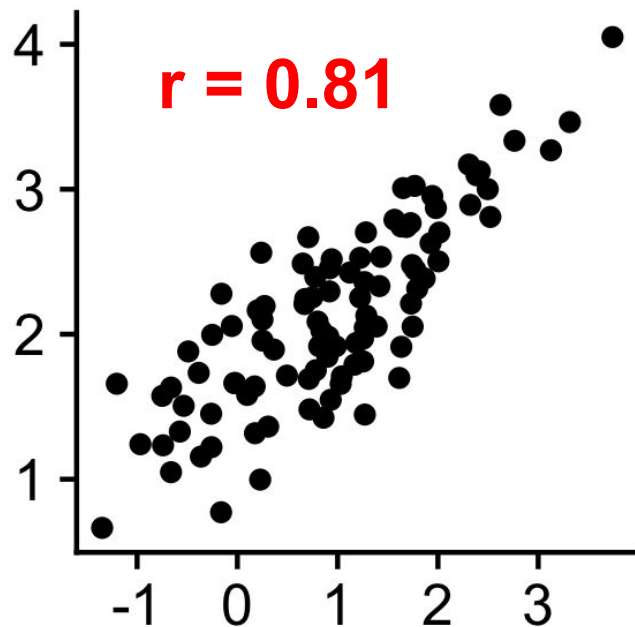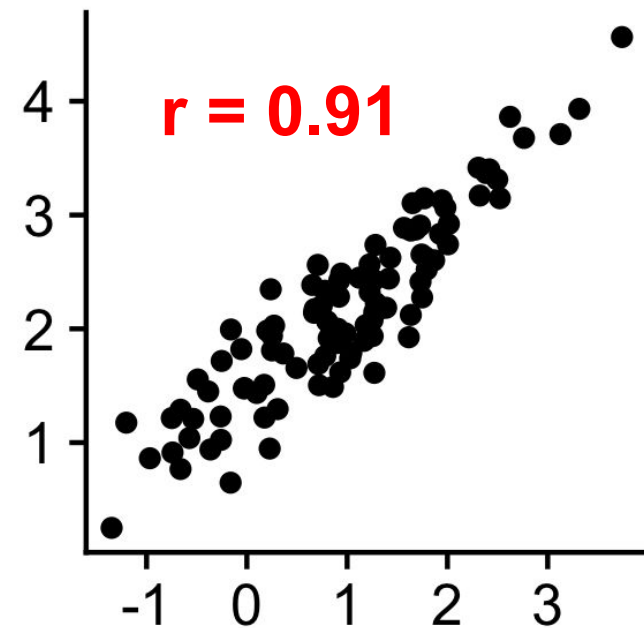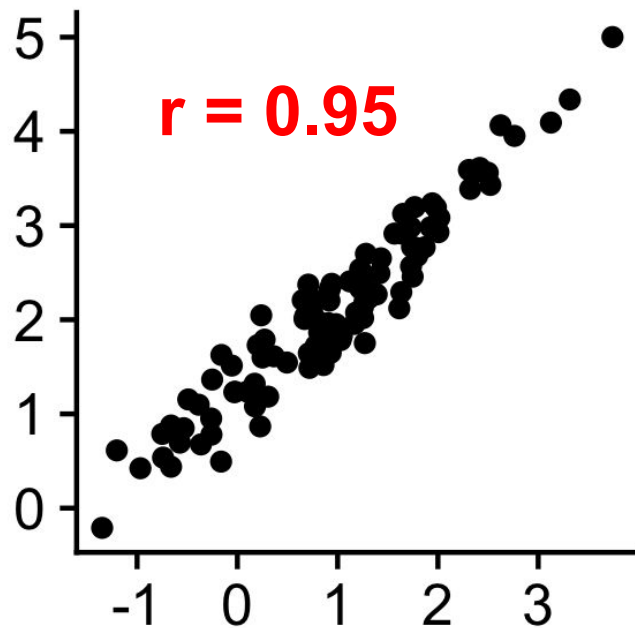
$$-1 \leq r \leq 1$$

Perfect Correlation

Different Correlations

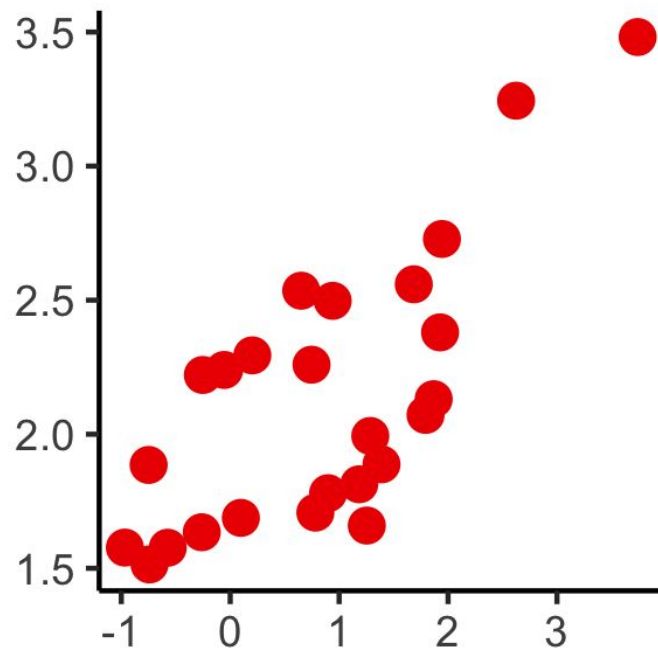r = 0.94

r = -0.62

r = 0.25

r = 0.03

Effect of
Variation on
Correlation

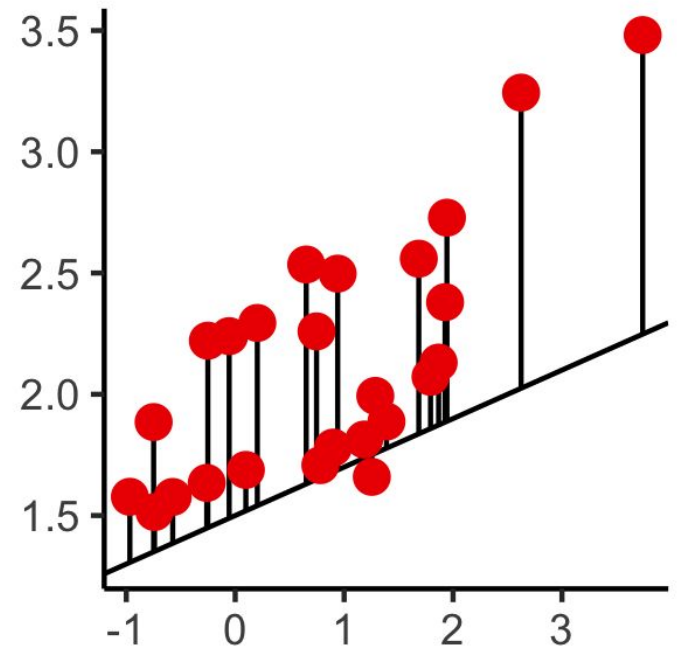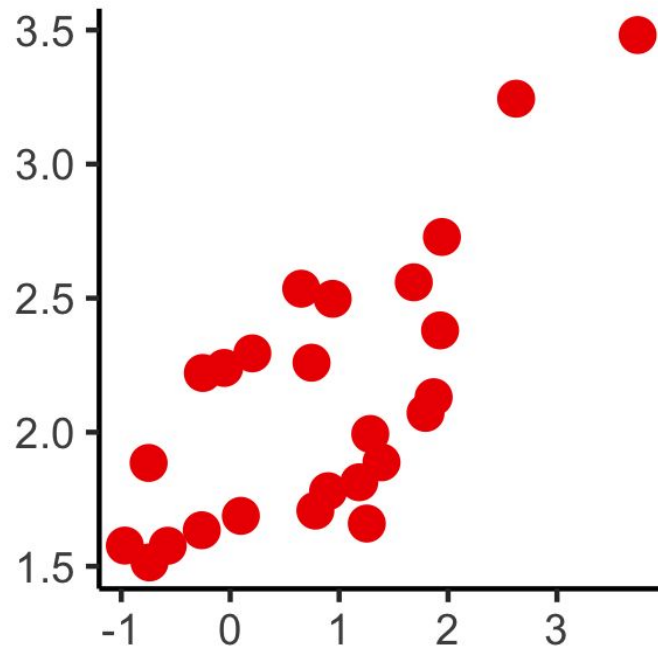r = 0.95

r = 0.91

r = 0.81

r = 0.62

Linear Regression

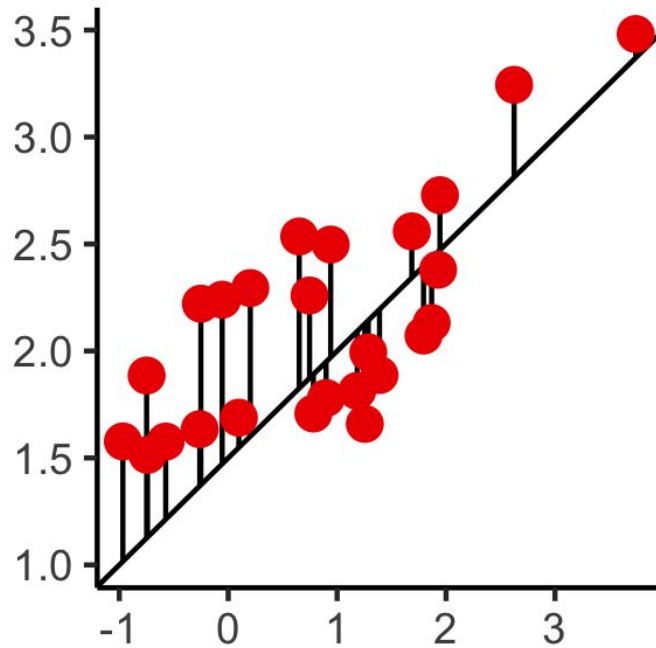Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data (by minimizing residuals).

Linear
Regression
(aka Line of
Best Fit)

Linear
Regression
(aka Line of
Best Fit)

Linear
Regression
(aka Line of
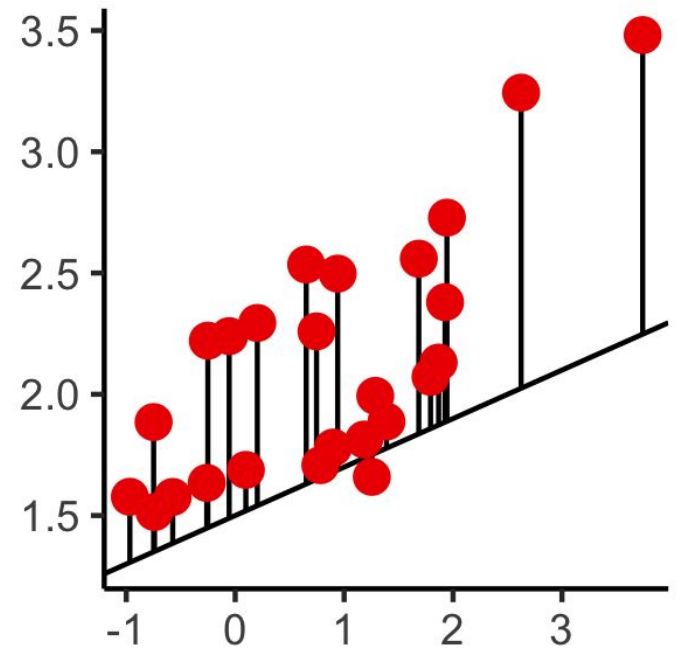Best Fit)
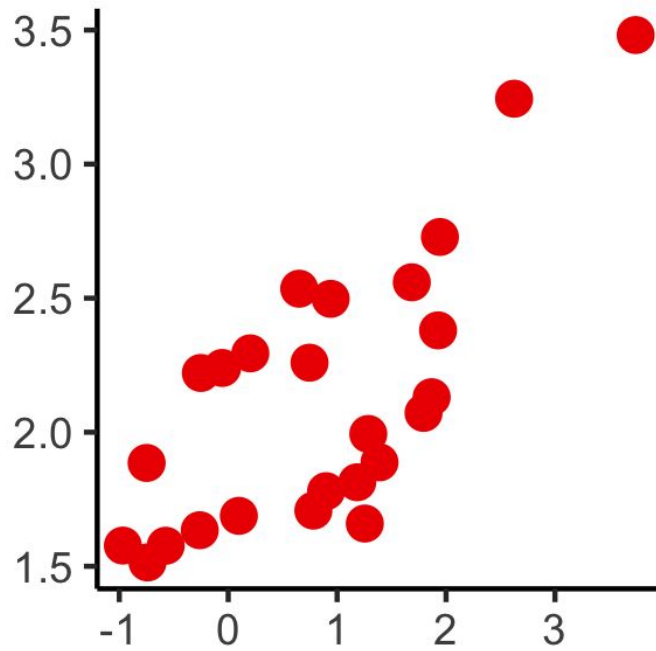
Linear
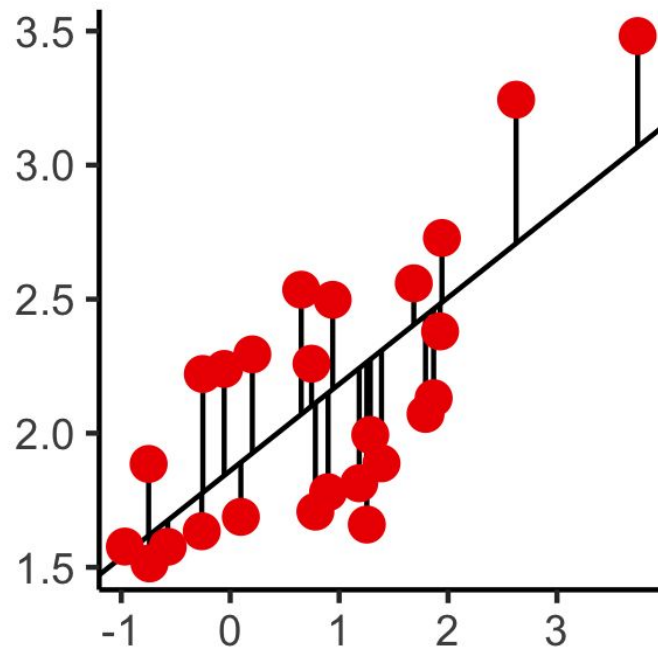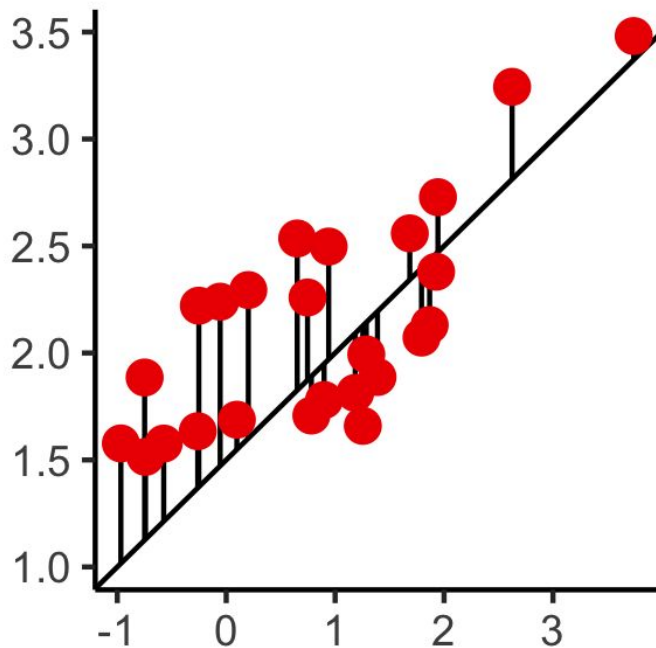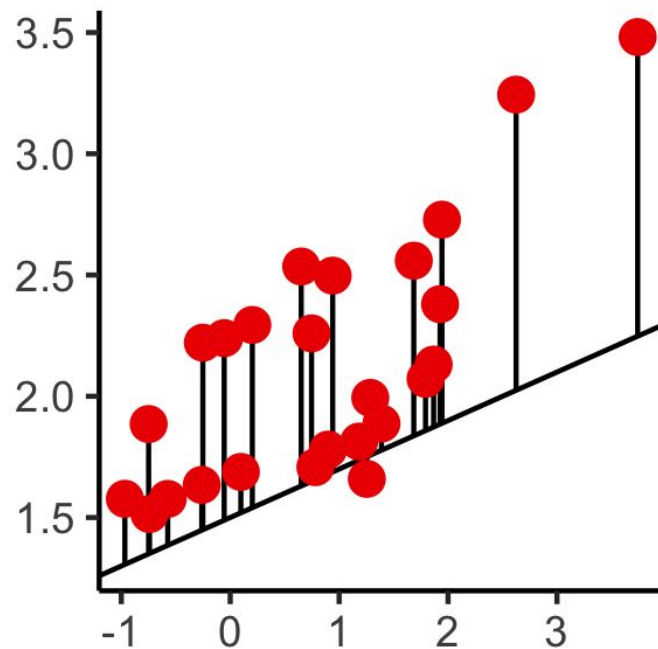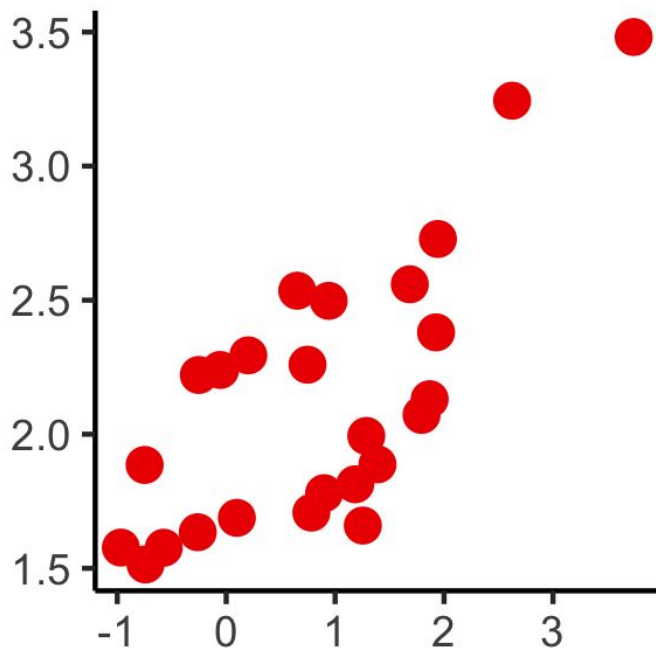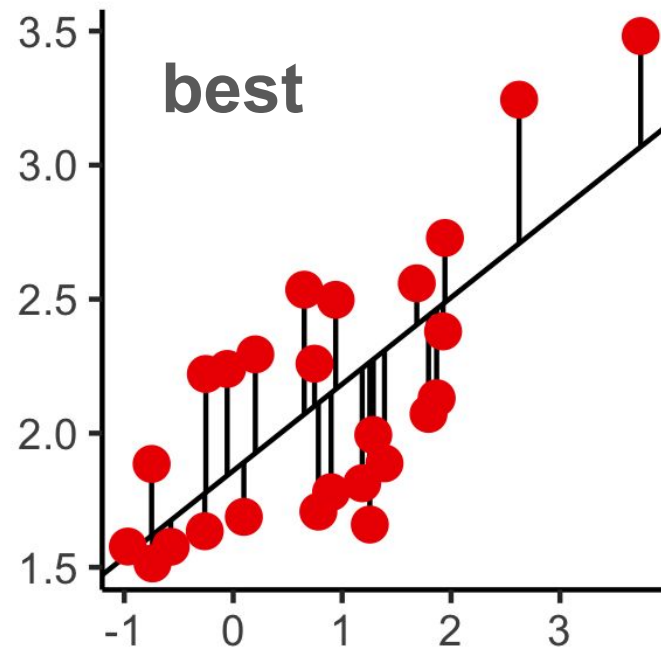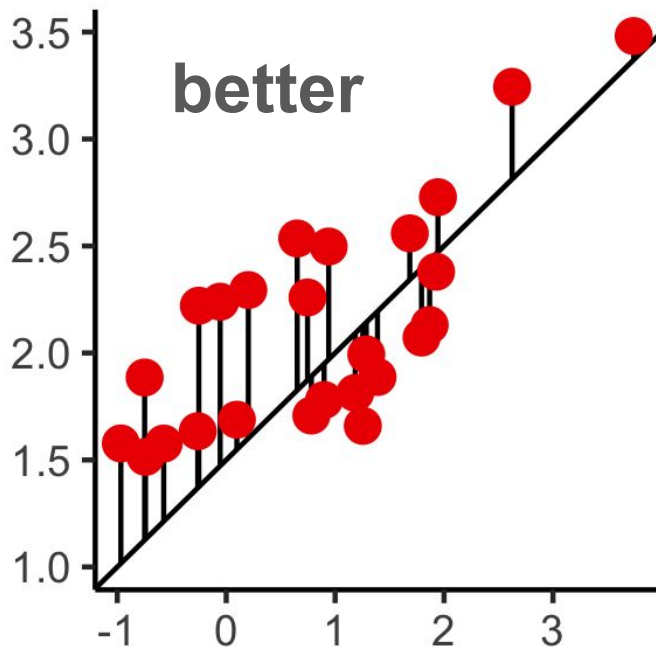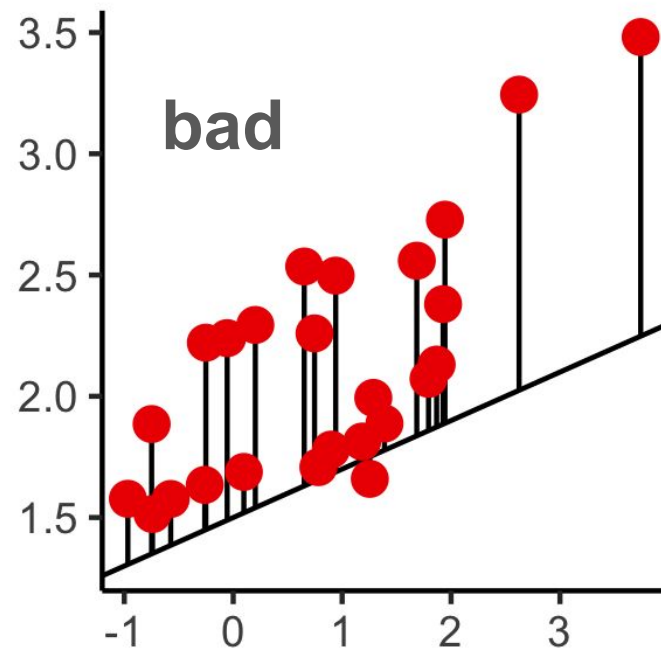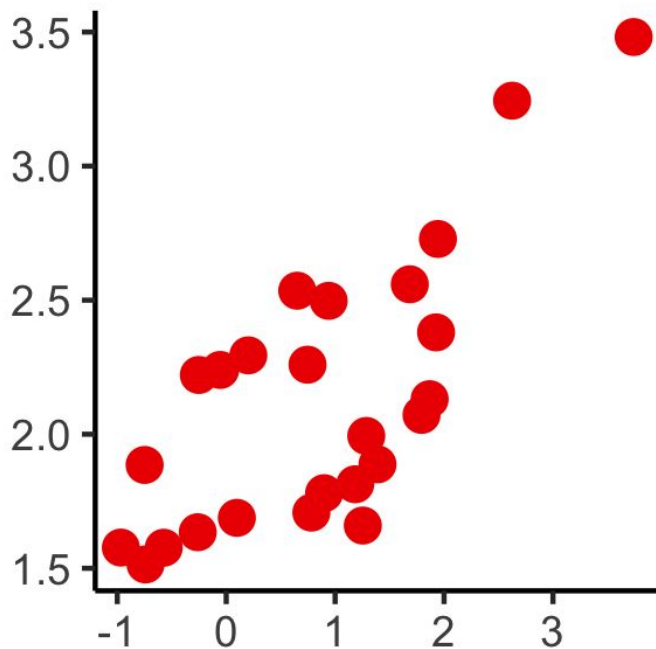Regression
(aka Line of
Best Fit)

Linear
Regression
(aka Line of
Best Fit)

Why care about correlation or linear regression?

Is this gene associated with my phenotype?

CORRELATION

I can predict my phenotype using these gene(s)?

REGRESSION

Is there a significant difference between these 3 (or more) groups?

ANOVA