"secret sauce of data science

Donoho, 2015
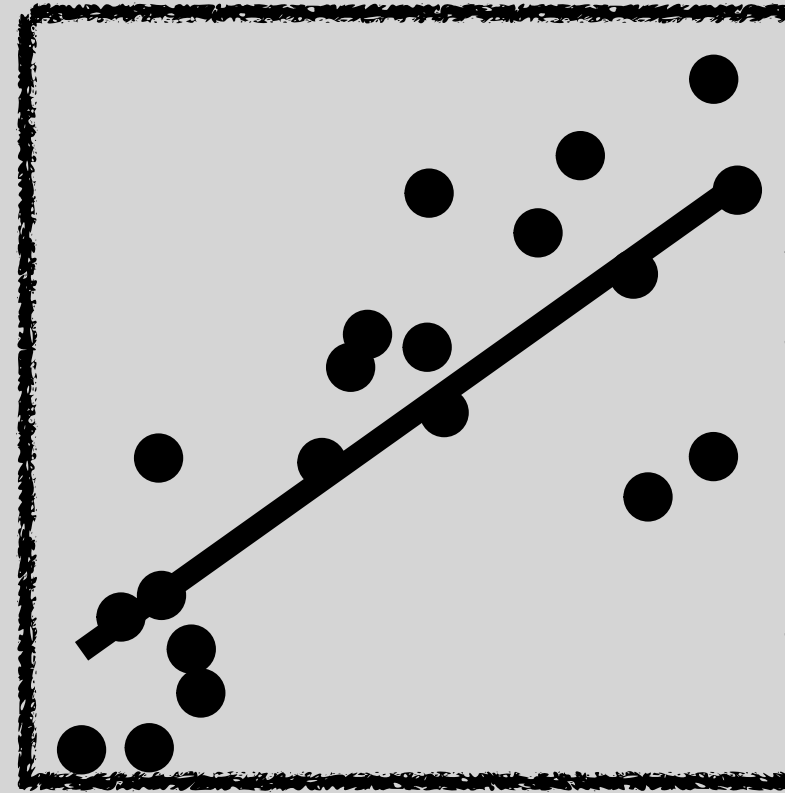
understanding fertility outcomes
by quantifying the (un)predictable

gert stulp — gertstulp.com — g.stulp@rug.nl
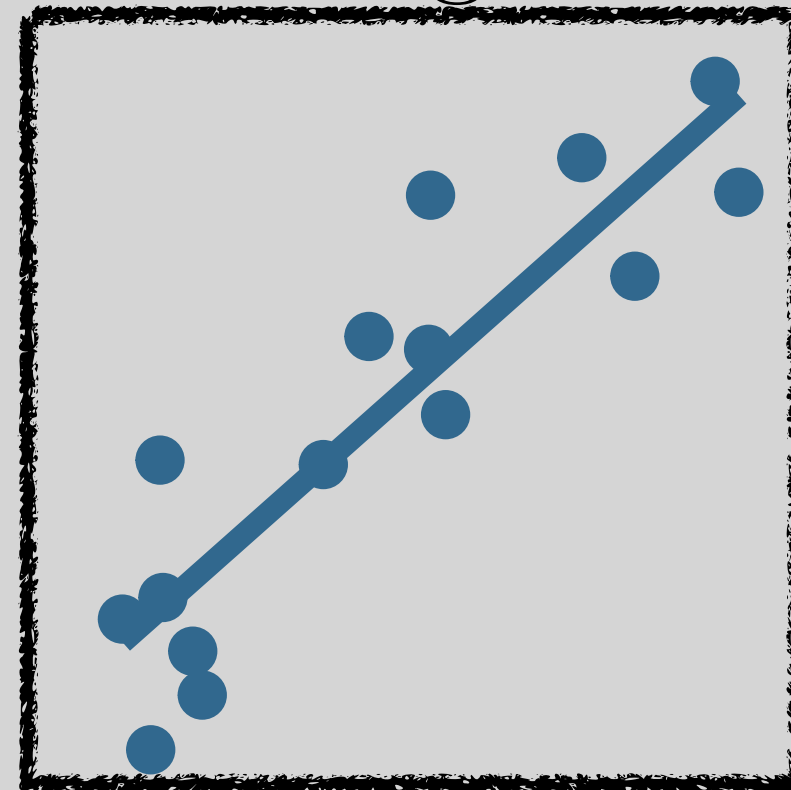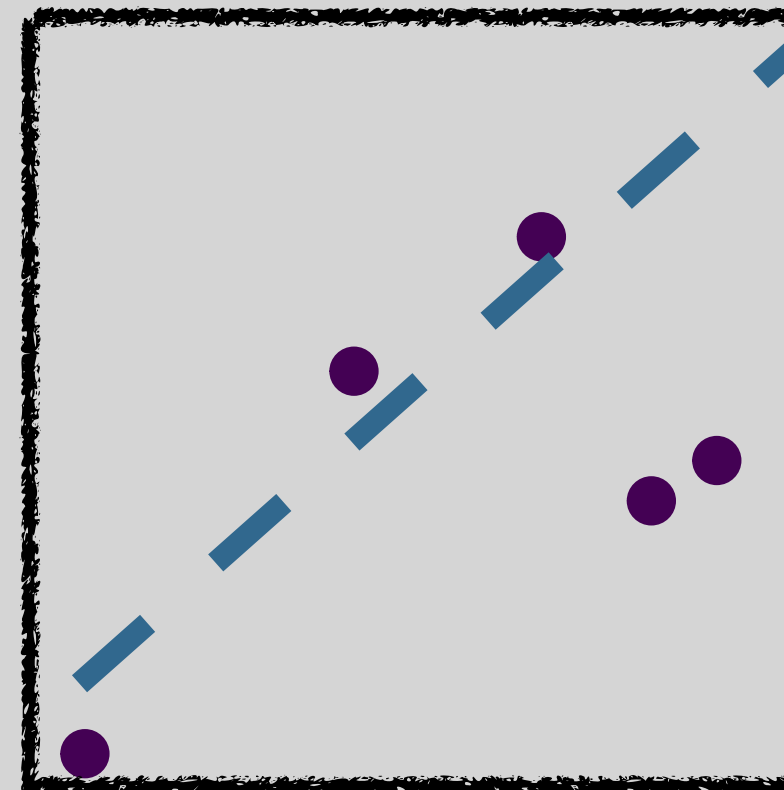
Out-of-Sample Prediction
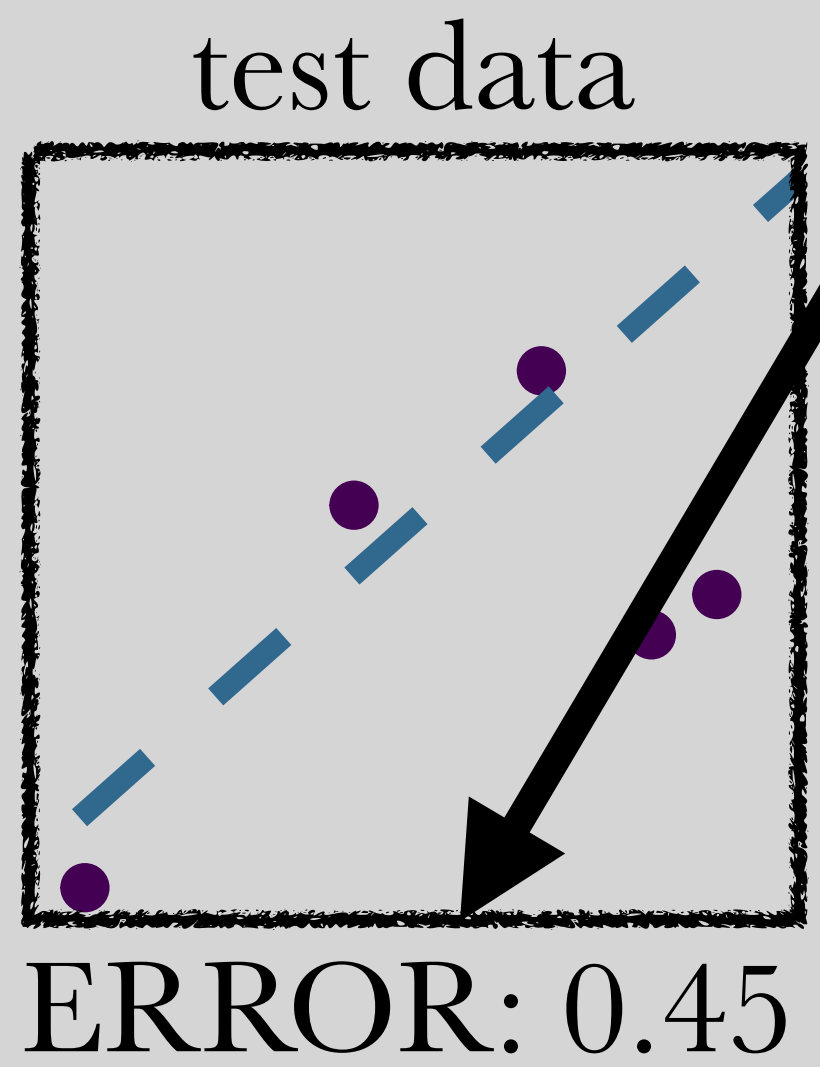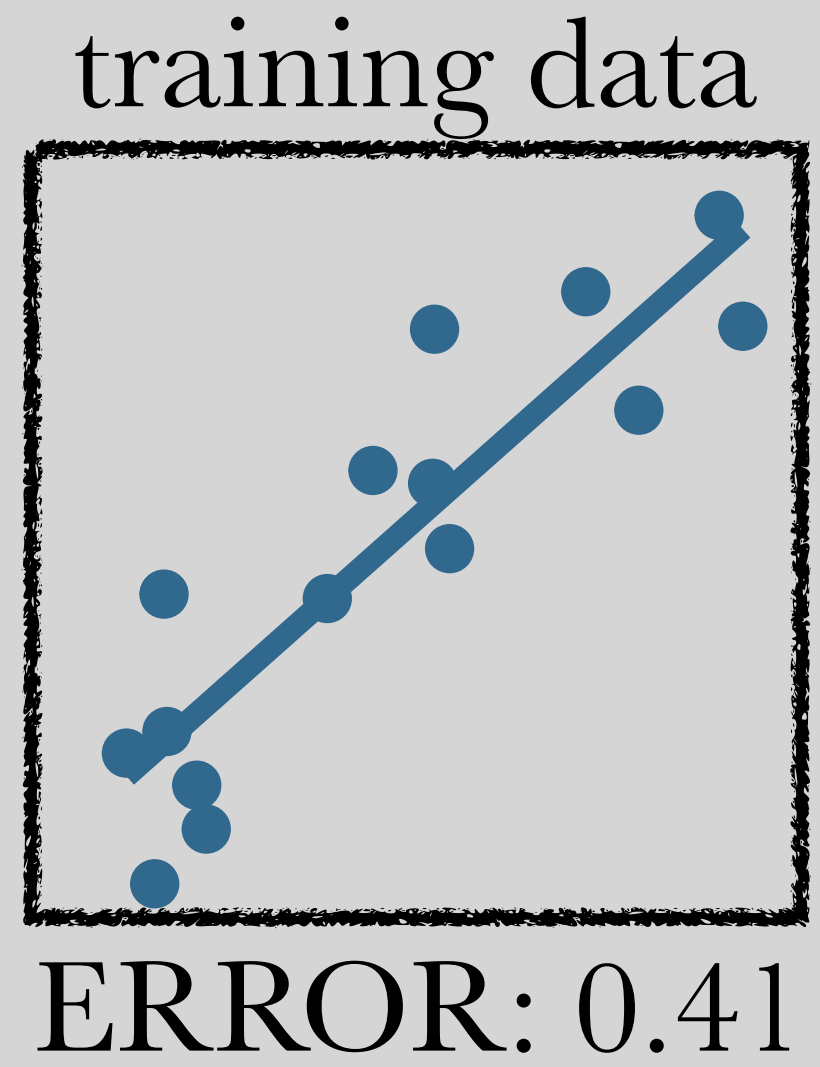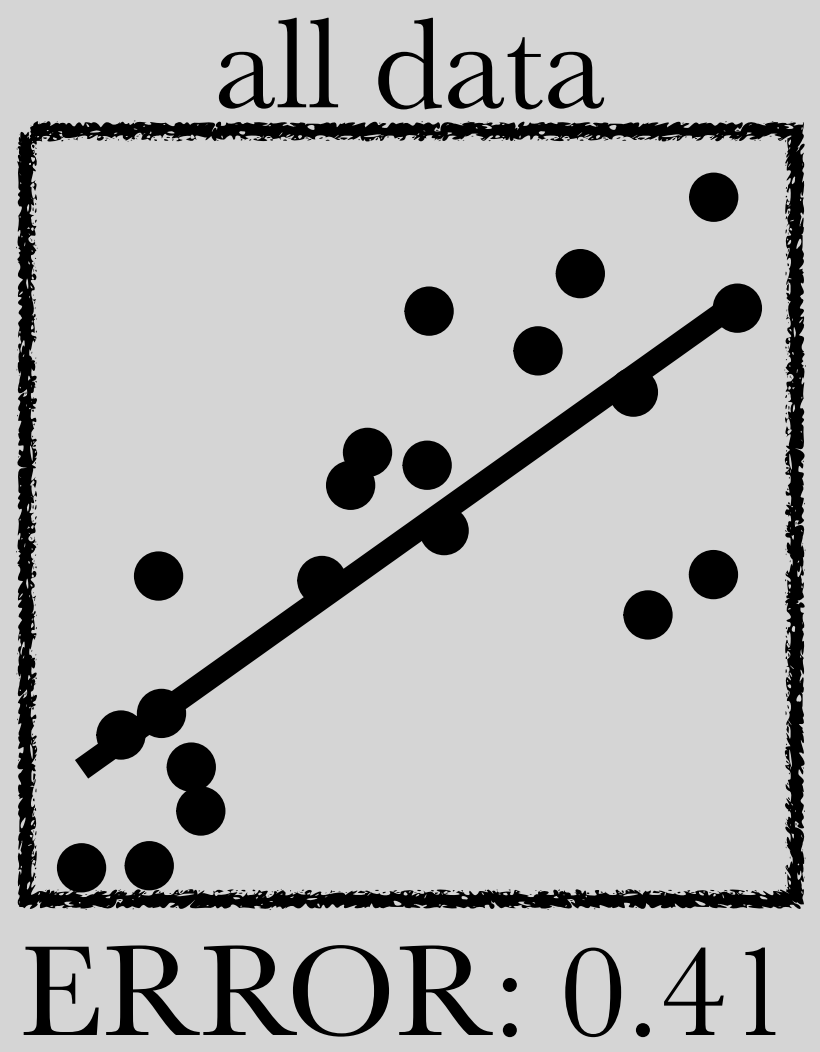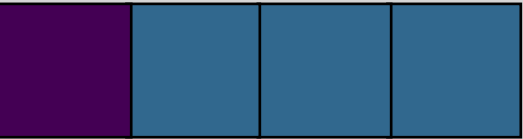

all data

ERROR: 0.41

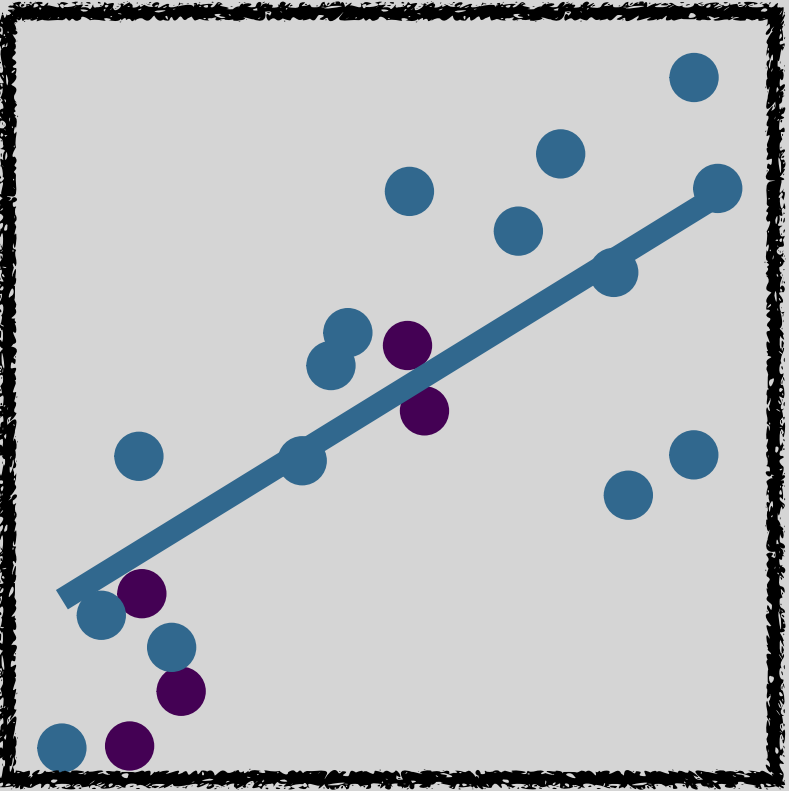training data

ERROR: 0.41

test data

ERROR: 0.45

Out-of-Sample Prediction

all data

ERROR: 0.41

out-of-sample
predictive ability

training data

ERROR: 0.41

test data

ERROR: 0.45

Cross-Validation

all data

ERROR: 0.41

fold 1

ERROR: 0.38

fold 2

ERROR: 0.62

fold 3

ERROR: 0.45

fold 4

ERROR: 0.38

Cross-Validation

all data
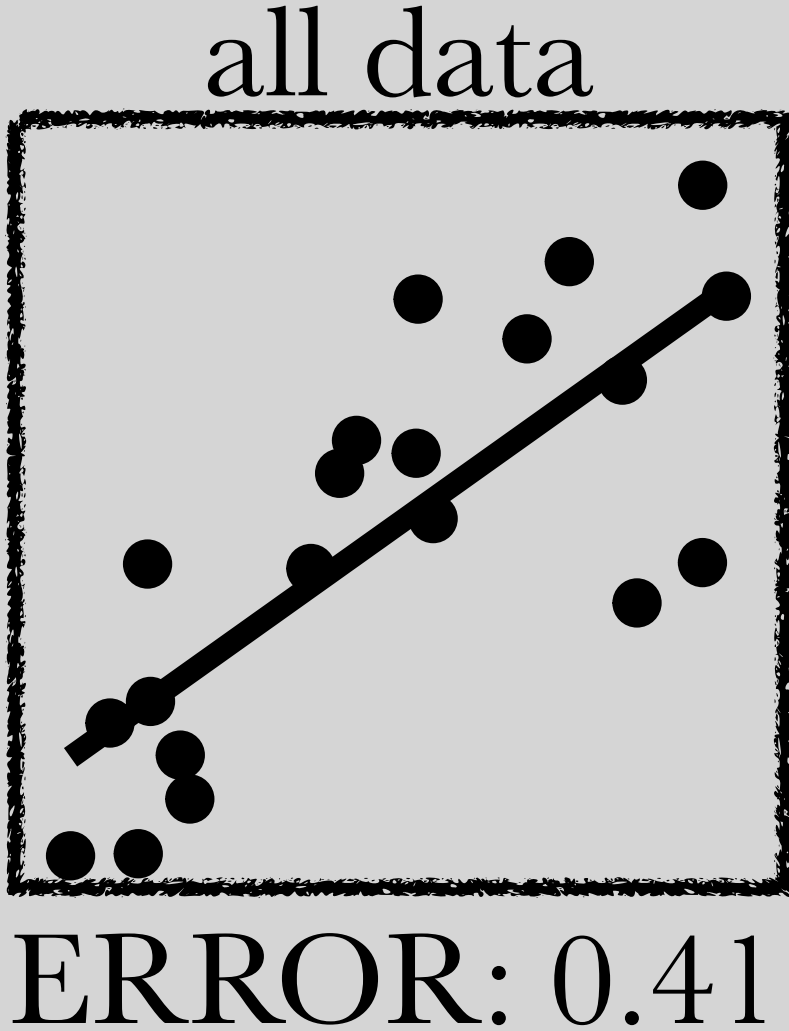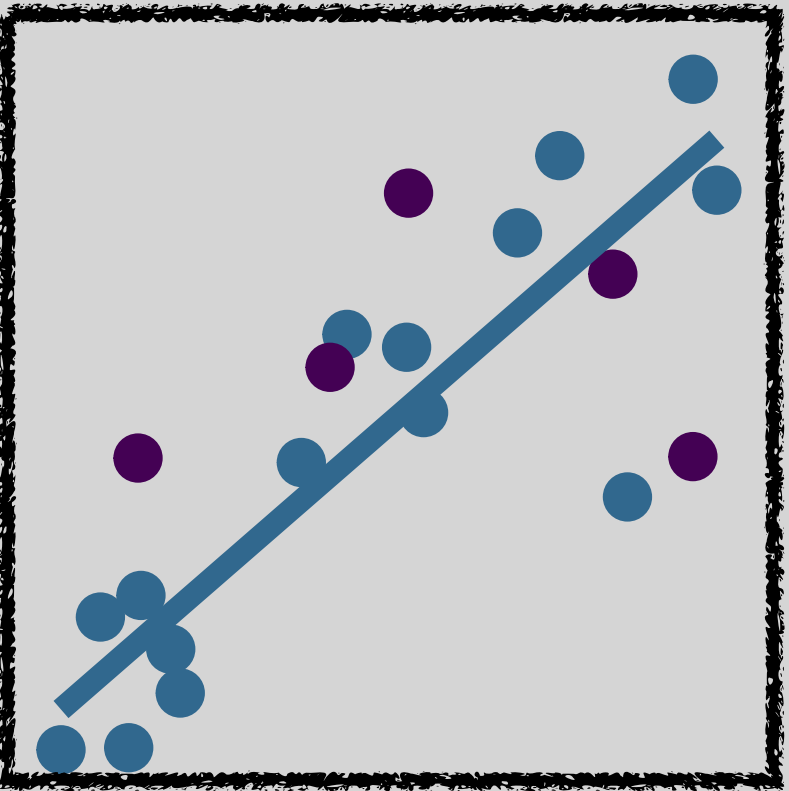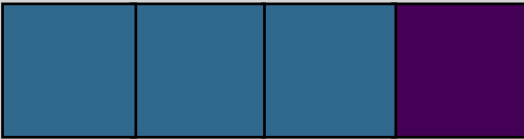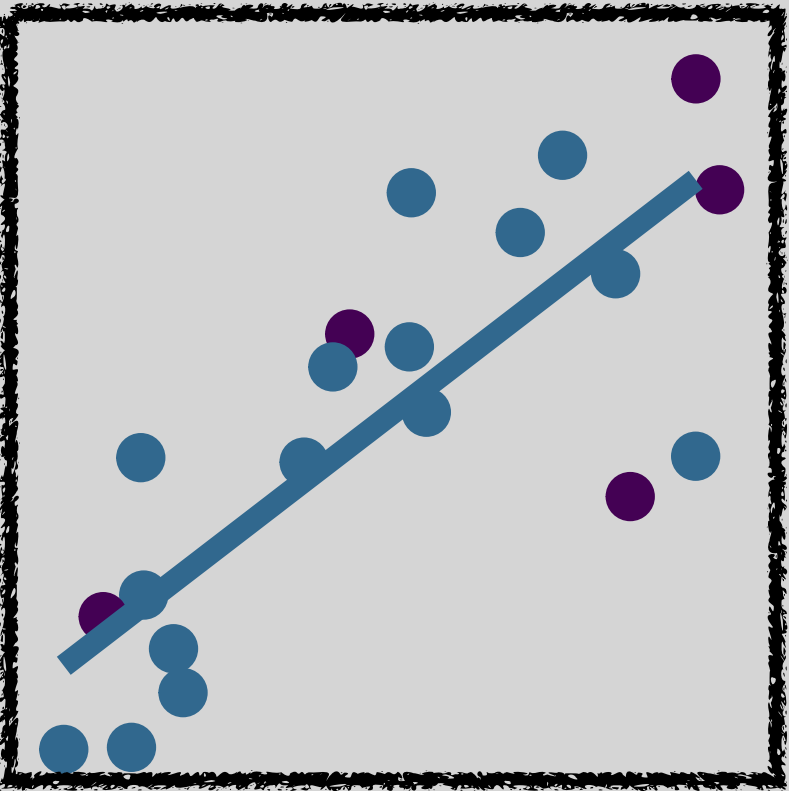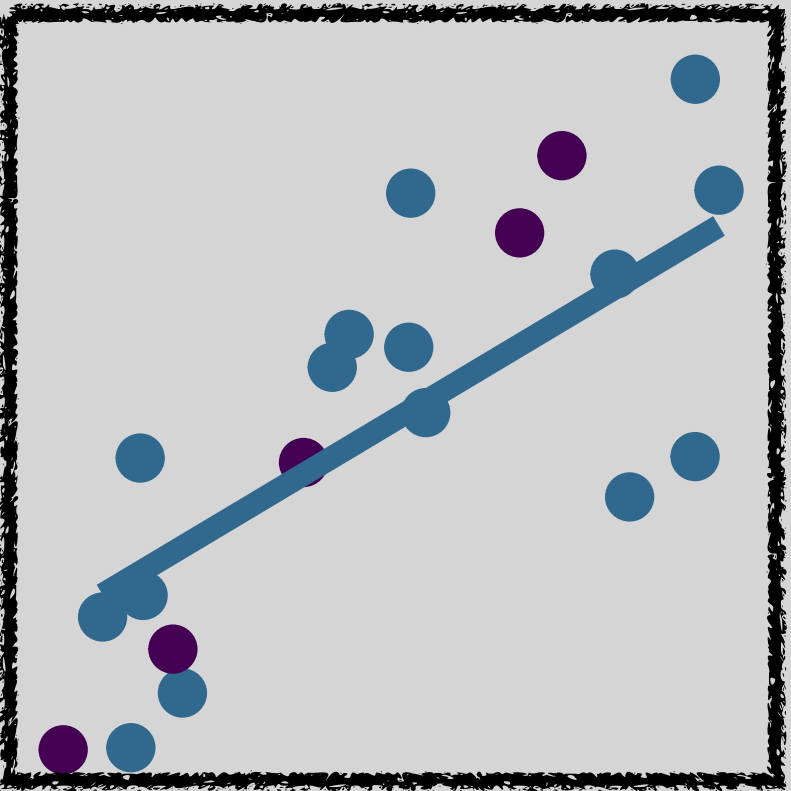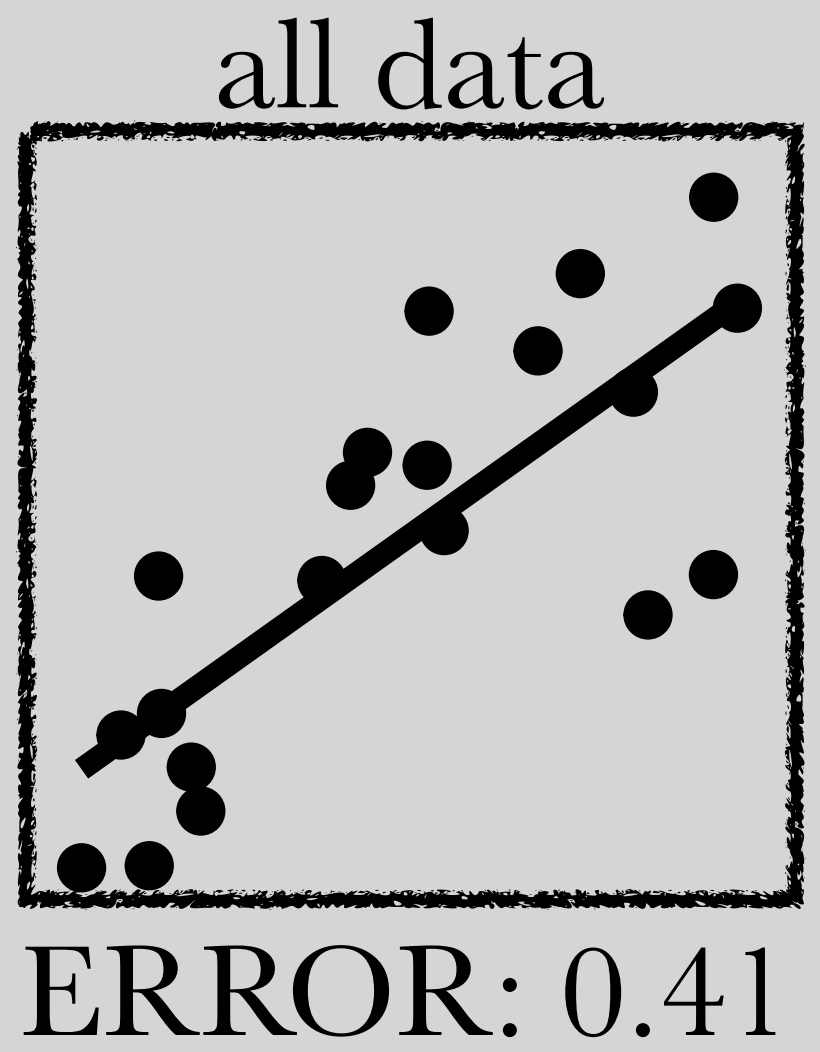
ERROR: 0.41

Cross-Validation
$\overline{\text{ERROR}}$: 0.46

out-of-sample
predictive ability

fold 1

ERROR: 0.38

fold 2

ERROR: 0.62
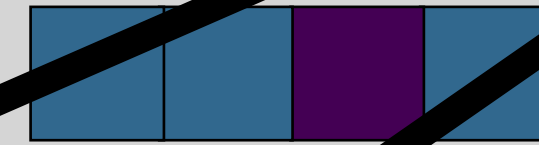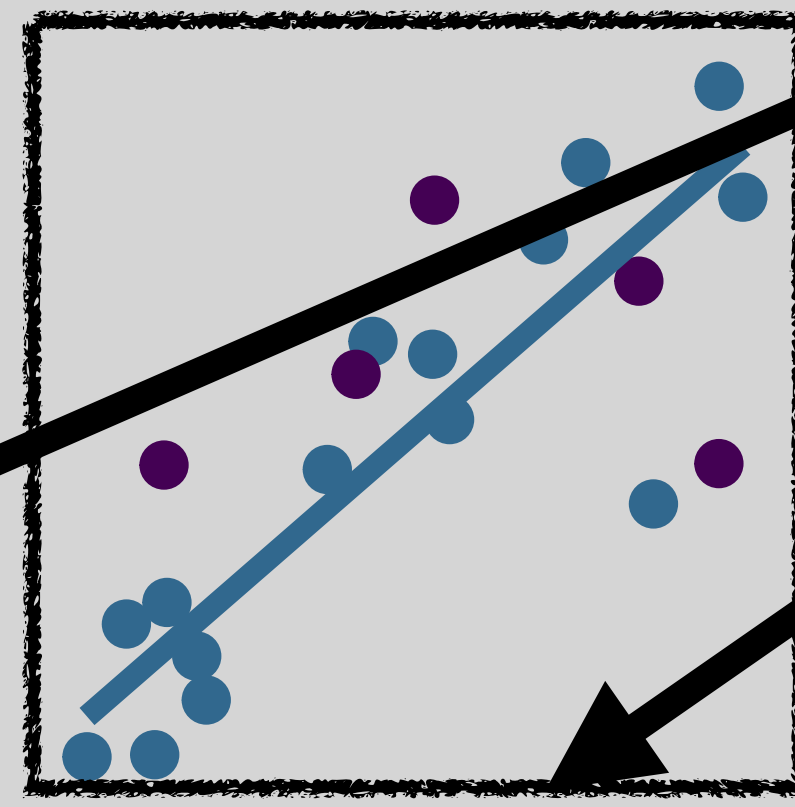
fold 3
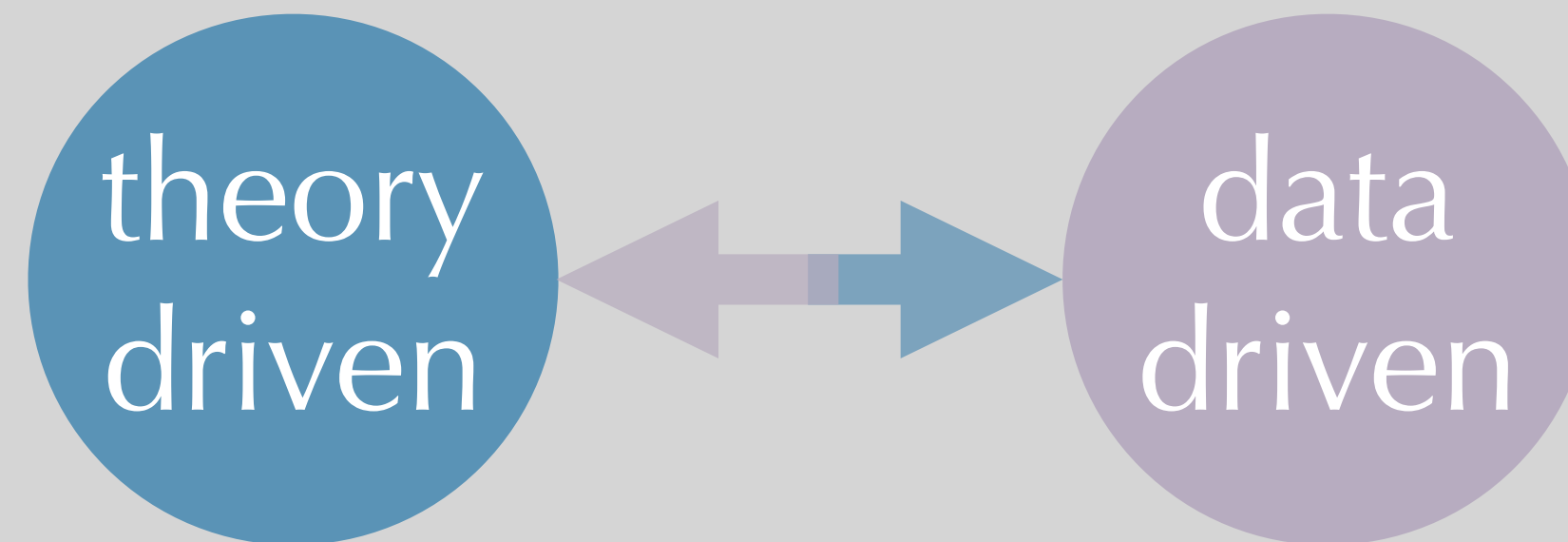

ERROR: 0.45

fold 4

ERROR: 0.38
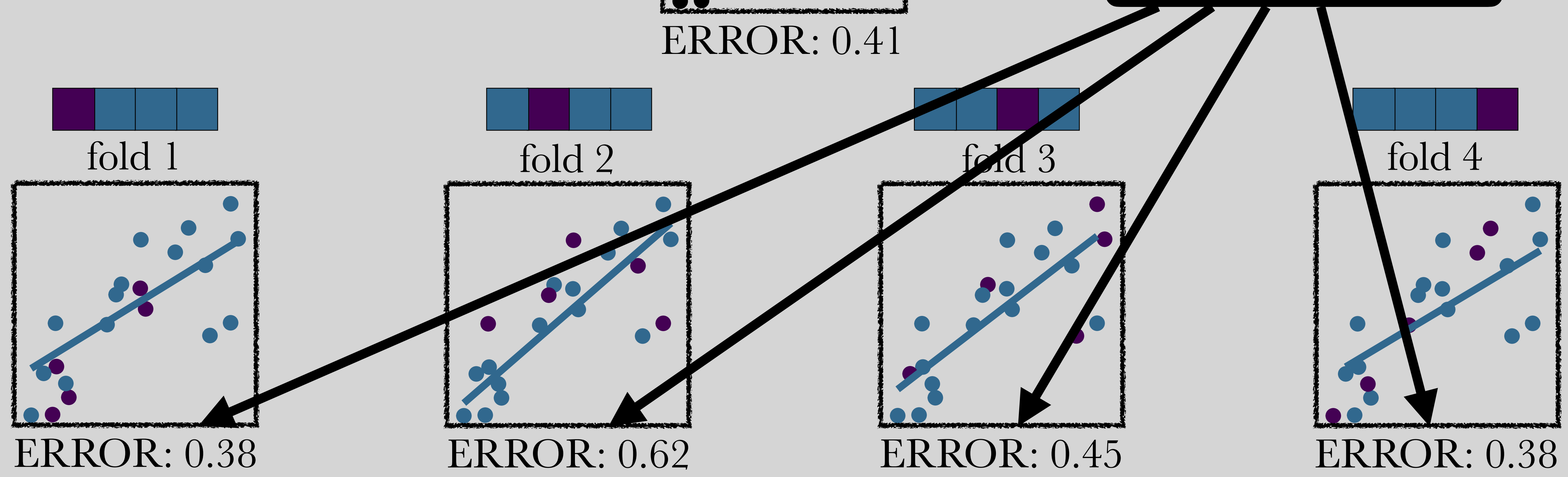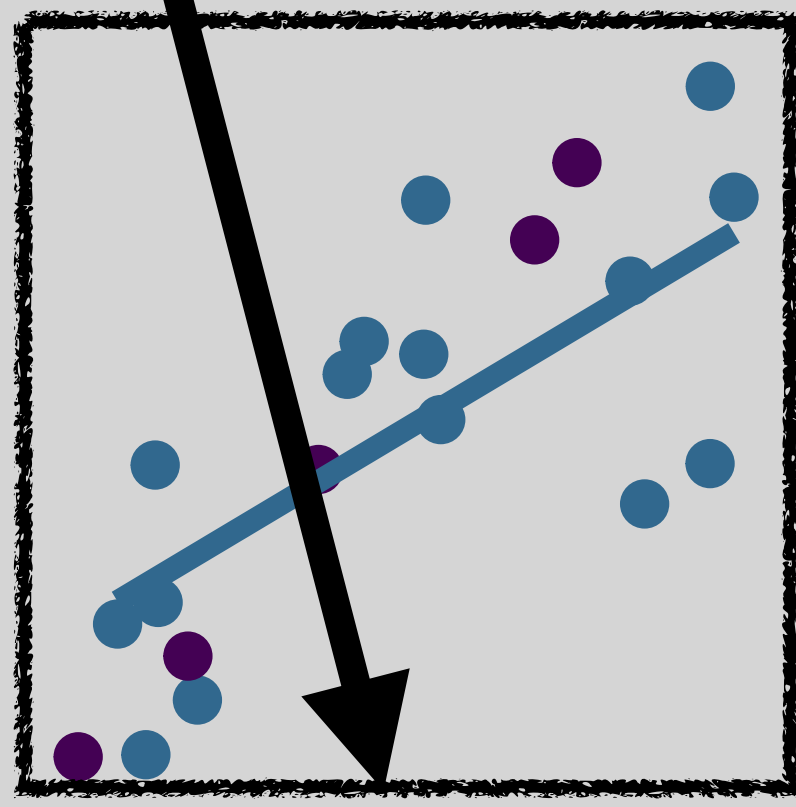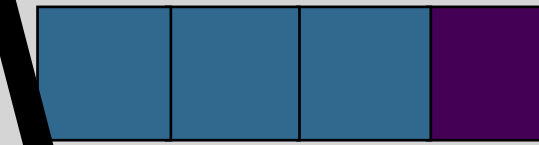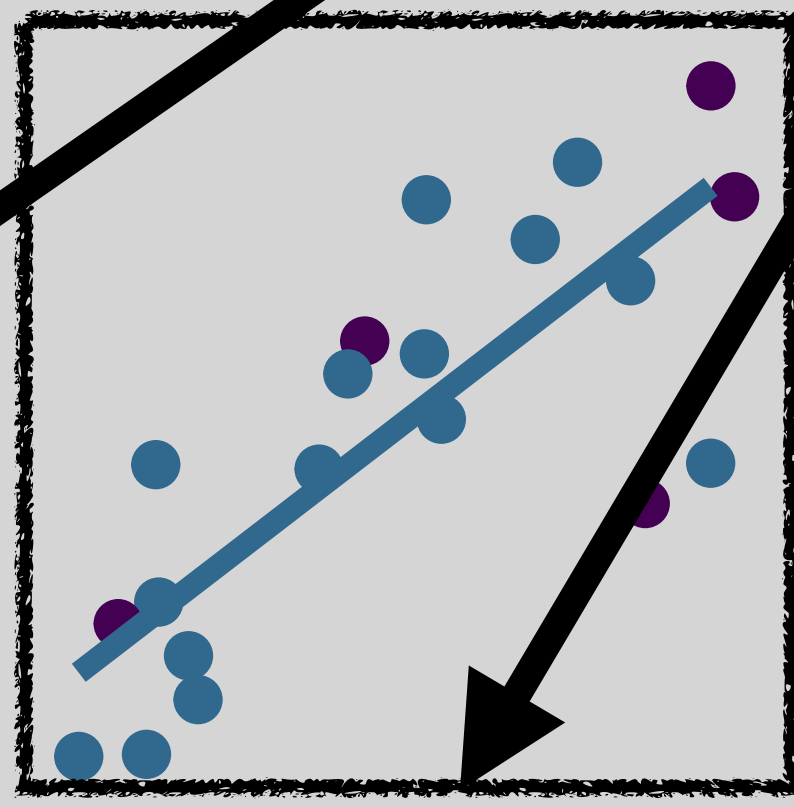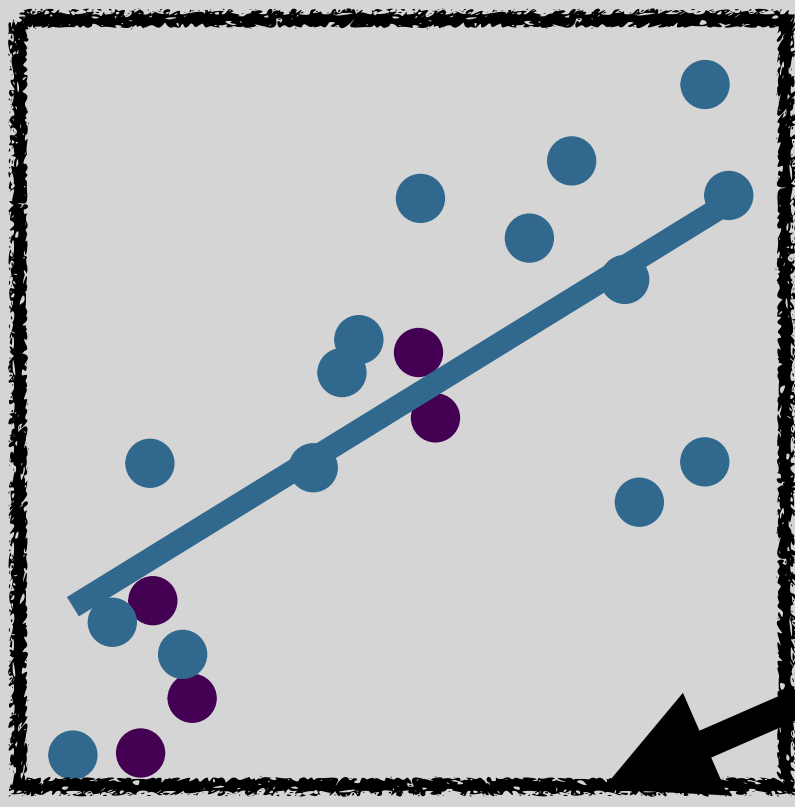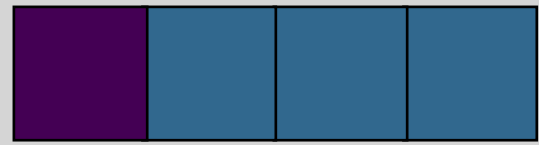
a shift towards **prediction** leads to a more reliable and useful social science

out-of-sample predictive ability:

clear measure of effect size

facilitates dialogue theory- and data-driven models

measure of distance theory and practice