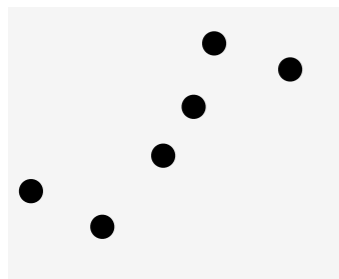


**Models  
explored  
interactively  
with the meifly package**

**Hadley Wickham**



**Models are  
data too!**



Observation

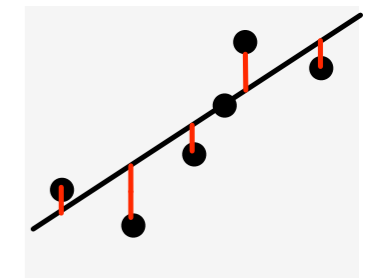
**Obs ID**

Original data  
Model-  
observation  
summaries

1

many

Model-Observation

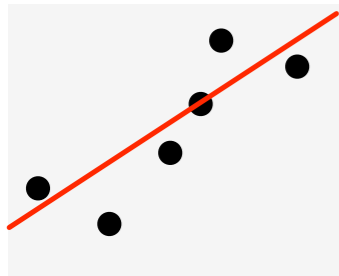


**Obs ID**

**Model ID**

Diagnostics  
Fit quality

many



Model

**Model ID**

Model fit  
statistics

1

many

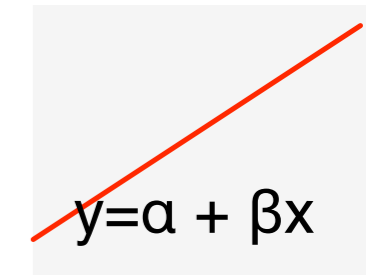
Model-Estimate

**Model ID**

**Estimate ID**

Raw  
Standardised  
Uncertainty

1



$f(\alpha, \beta)$

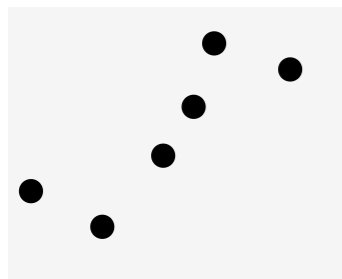
Estimate

**Estimate ID**

n  
model-estimate  
summaries

1

many



Observation

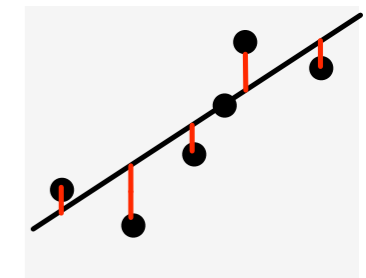
**Obs ID**

Original data  
Model-  
observation  
summaries

1

many

Model-Observation

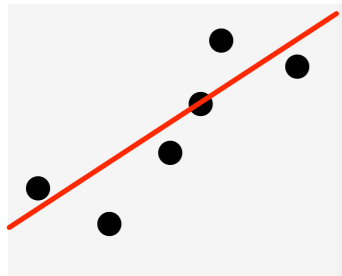


**Obs ID**

**Model ID**

Diagnostics  
Fit quality

many



Model

**Model ID**

Model fit  
statistics

1

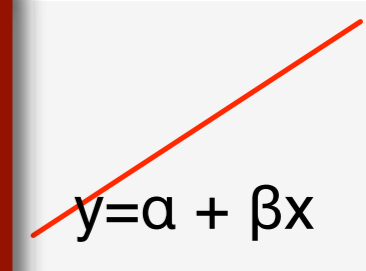
many

Model-Estimate

**Model ID**

**Estimate ID**

Raw  
Standardised  
Uncertainty



$f(\alpha, \beta)$

Estimate

**Estimate ID**

n  
model-estimate  
summaries

1

many

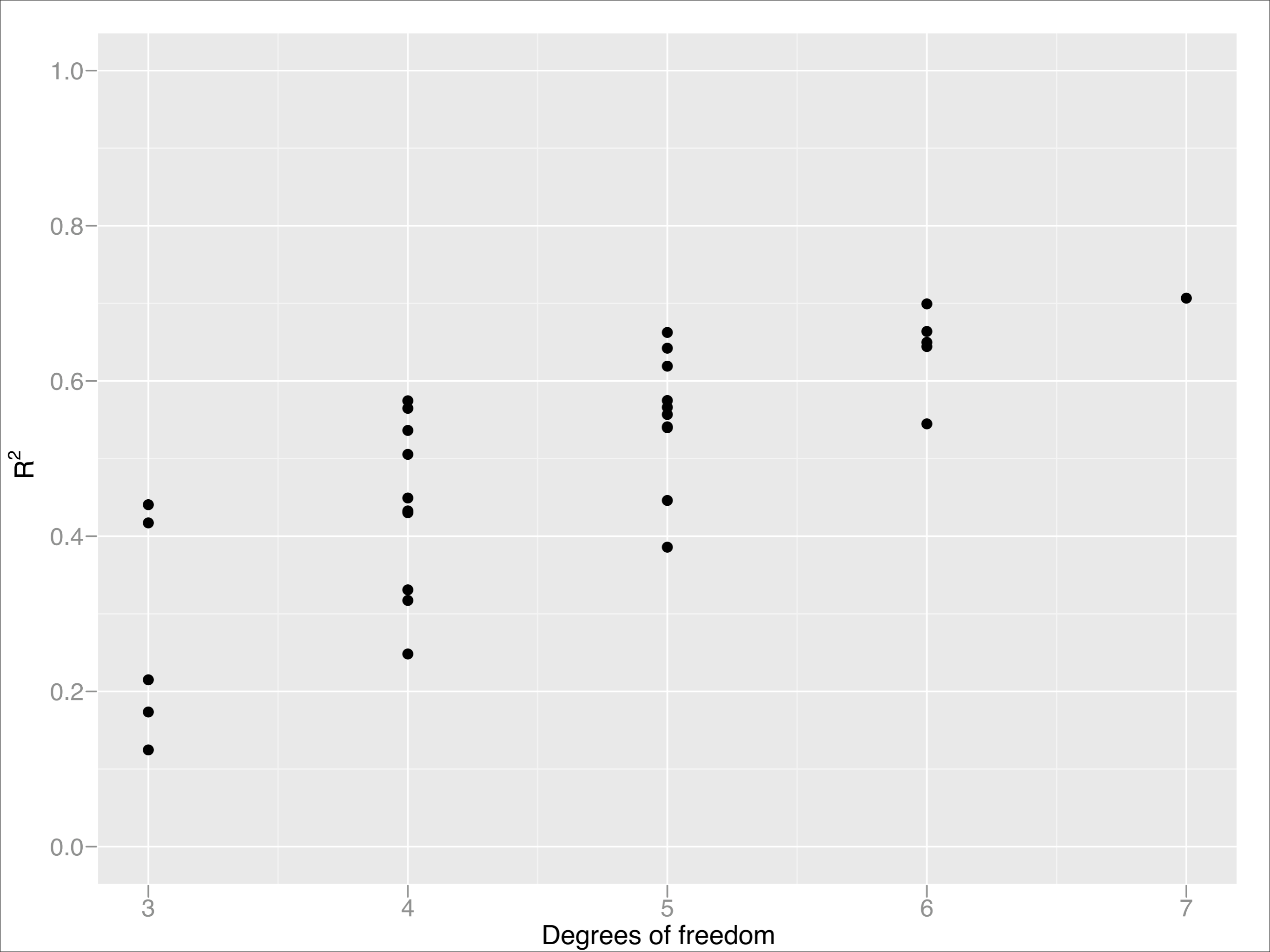
# Example

- Data from French-speaking Swiss provinces in the late 1800's
- Want to understand the relationship between fertility and:
  - proportion of agricultural workers
  - performance on army examination
  - higher education
  - proportion of Catholics
  - infant mortality

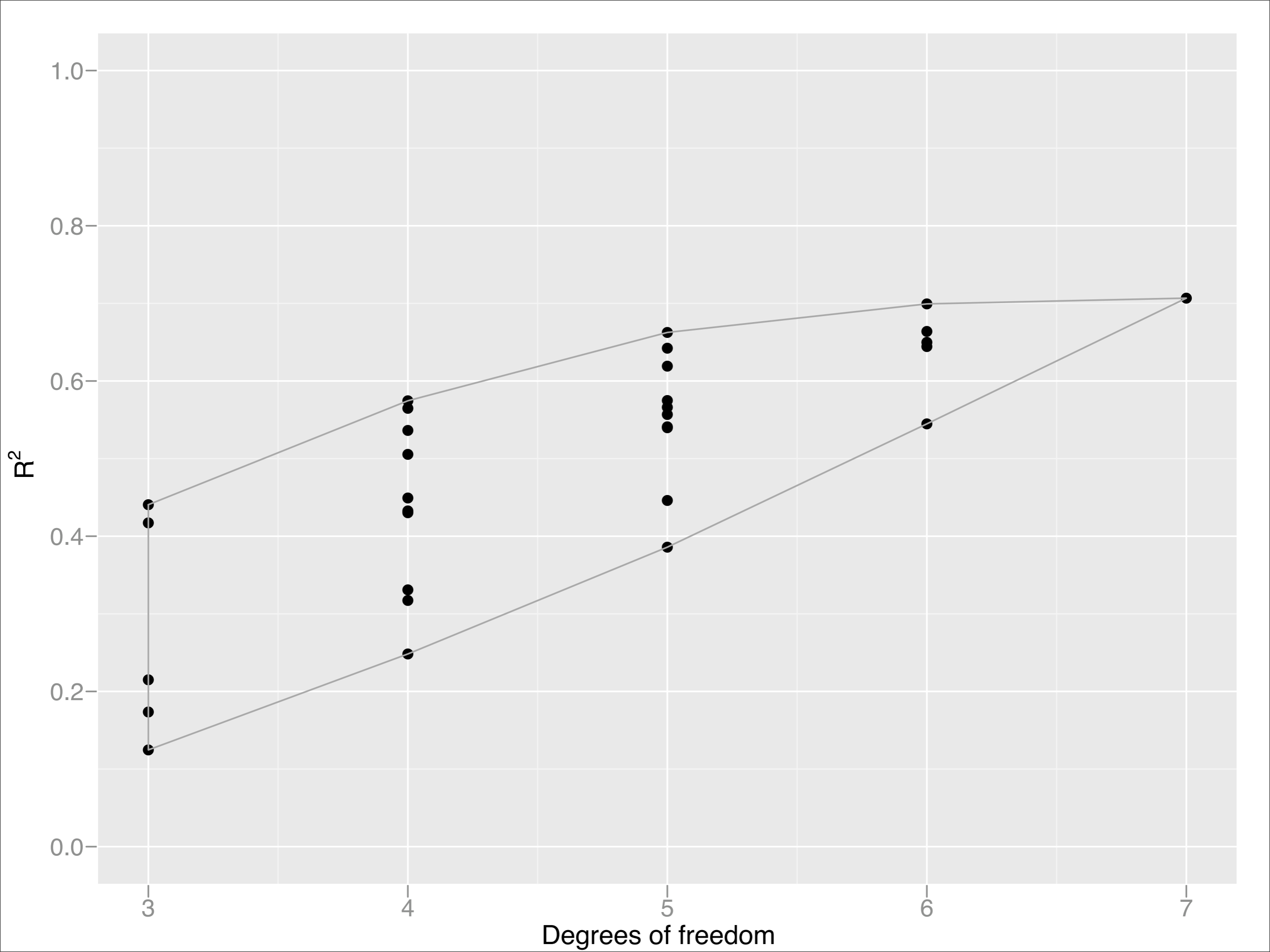
**Focus on  
understanding,  
not prediction**

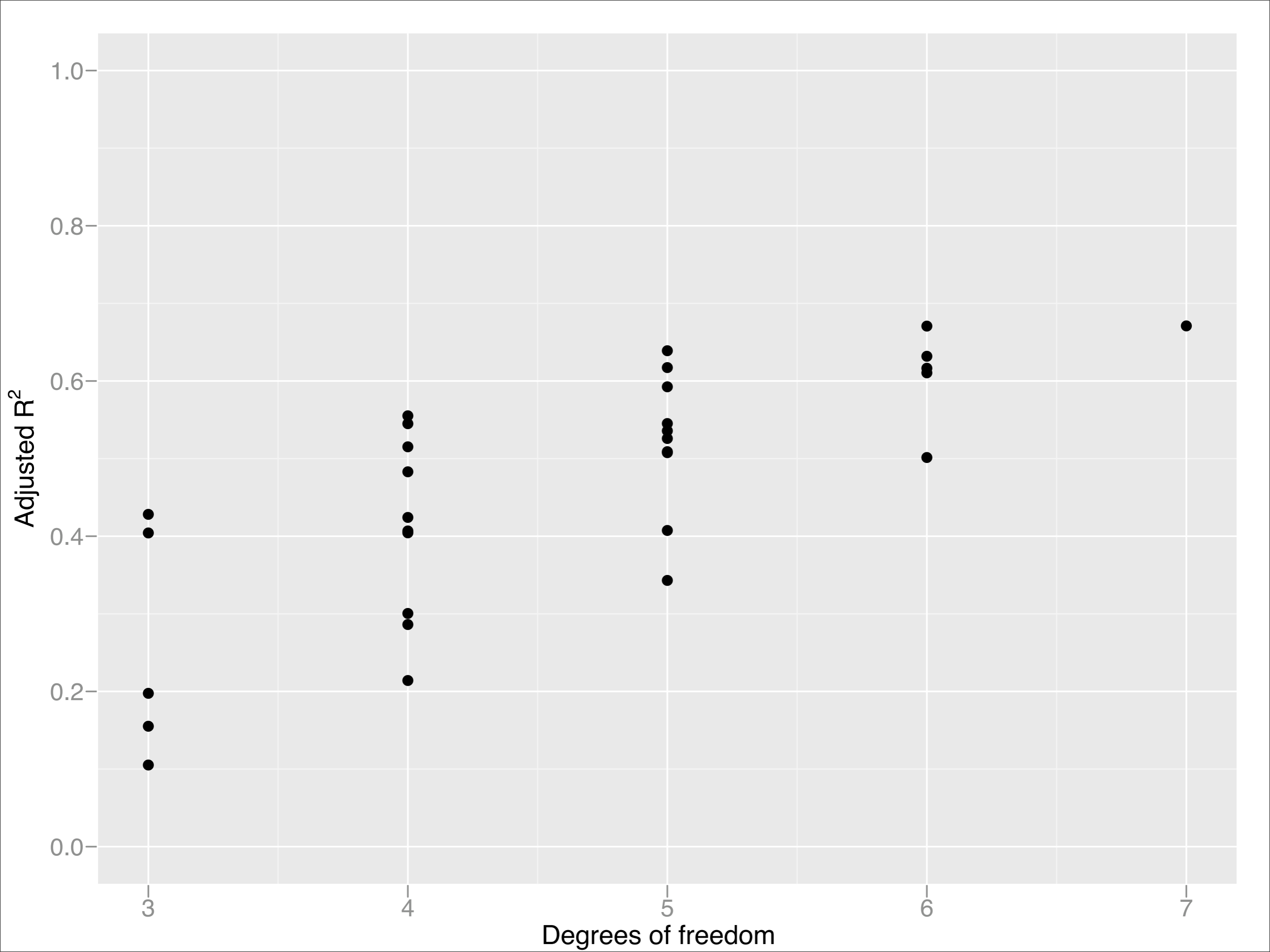
# Model level

- Fit all  $2^5 - 1 = 31$  possible linear models
- Summarise with:
  - degrees of freedom
  - $R^2$ , adjusted  $R^2$
  - Log-likelihood, BIC, AIC





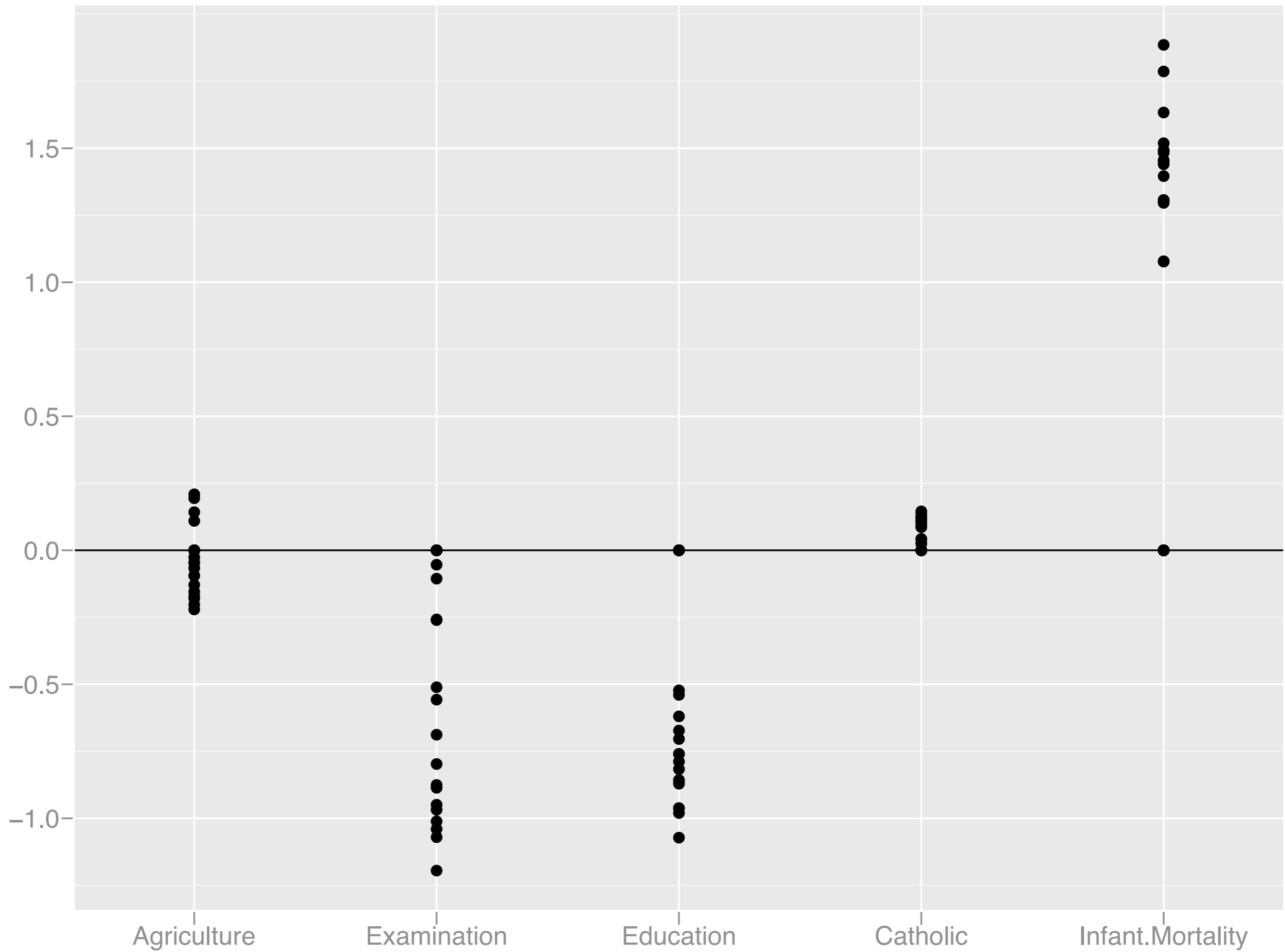




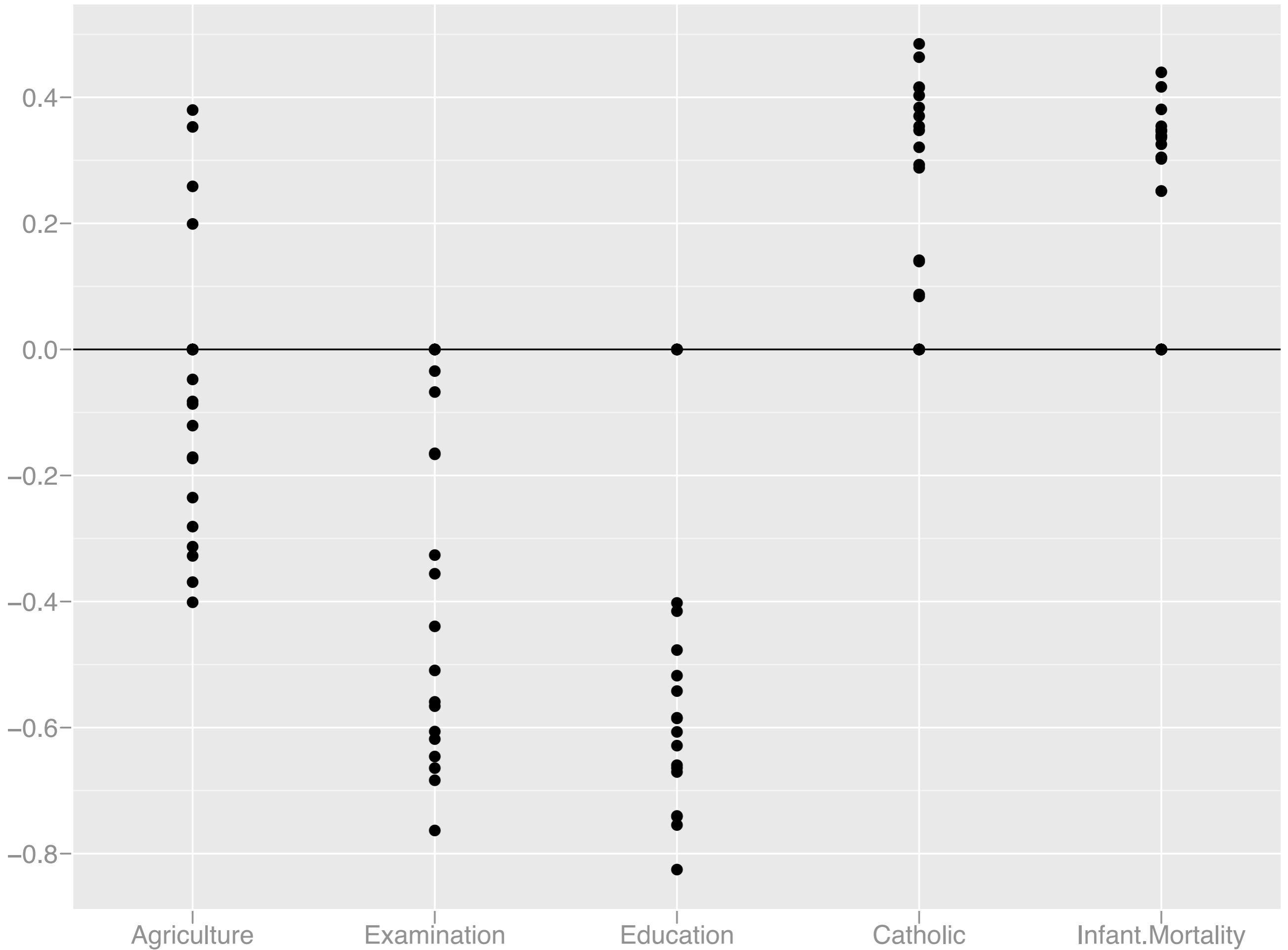
# Model-estimate level

- Raw and standardised estimates
- Standard error
- t-value, absolute t-value
- Explore variance-covariance matrix of predictors

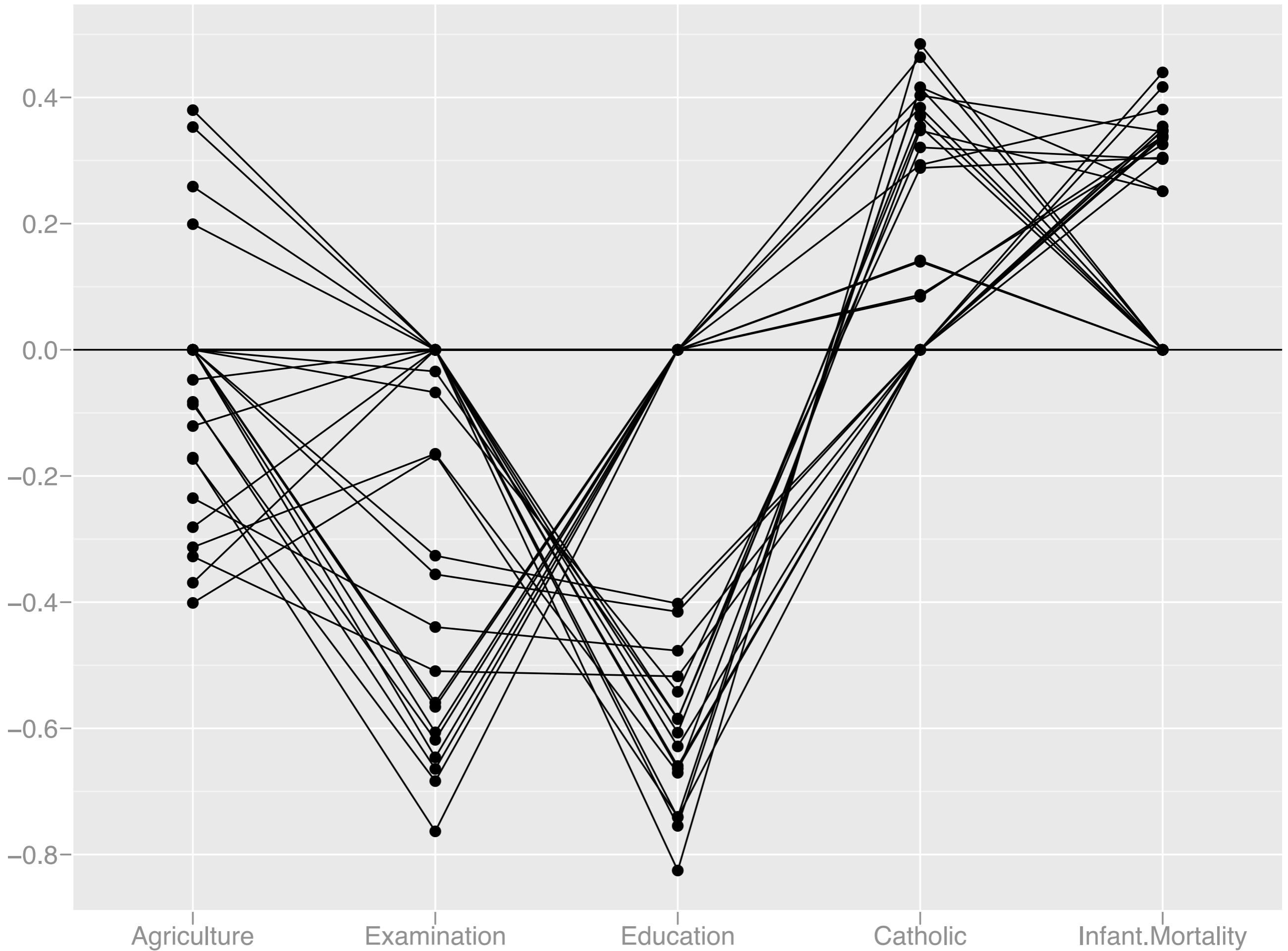
Raw coefficient



Standardised coefficient



Standardised coefficient



**Model +  
model-estimate**

# Conclusion

- Ensemble of models can be used to aid understanding, not just improving prediction
- Number of models, extend to non-parametric/Bayesian models, but would have to sample model space in more sophisticated way



<http://had.co.nz/meifly>