

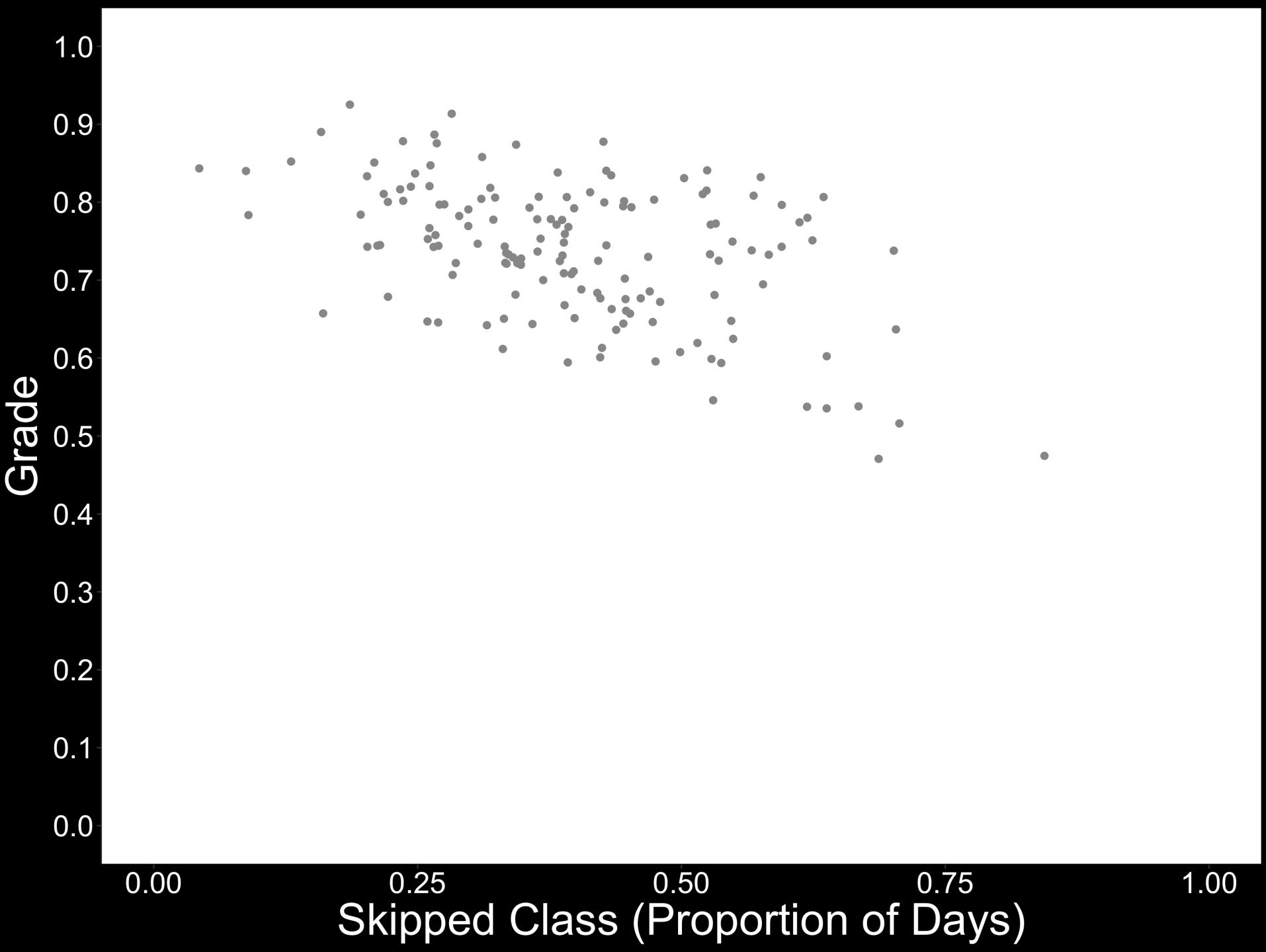
Multilevel Modeling

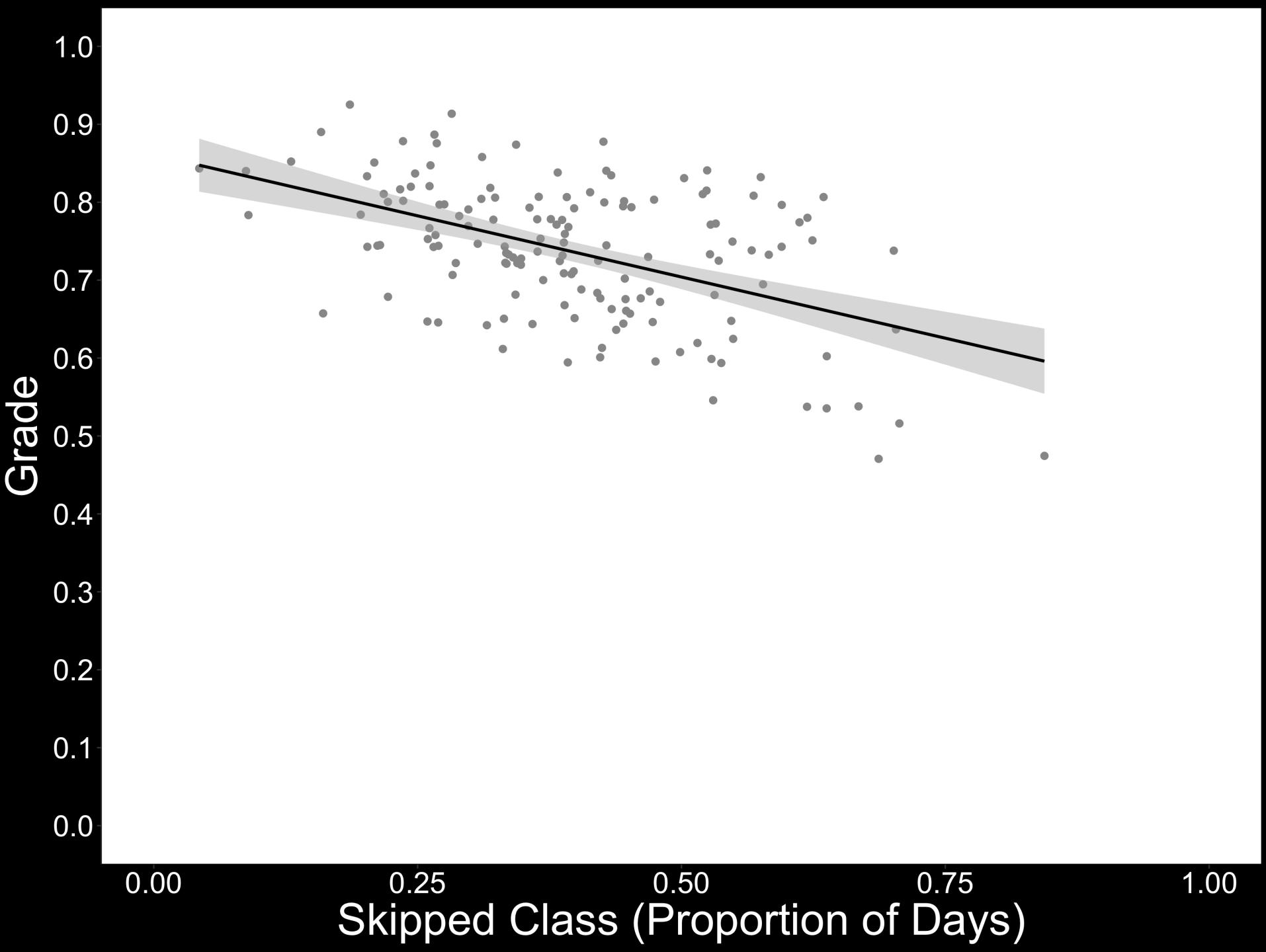
Day 1

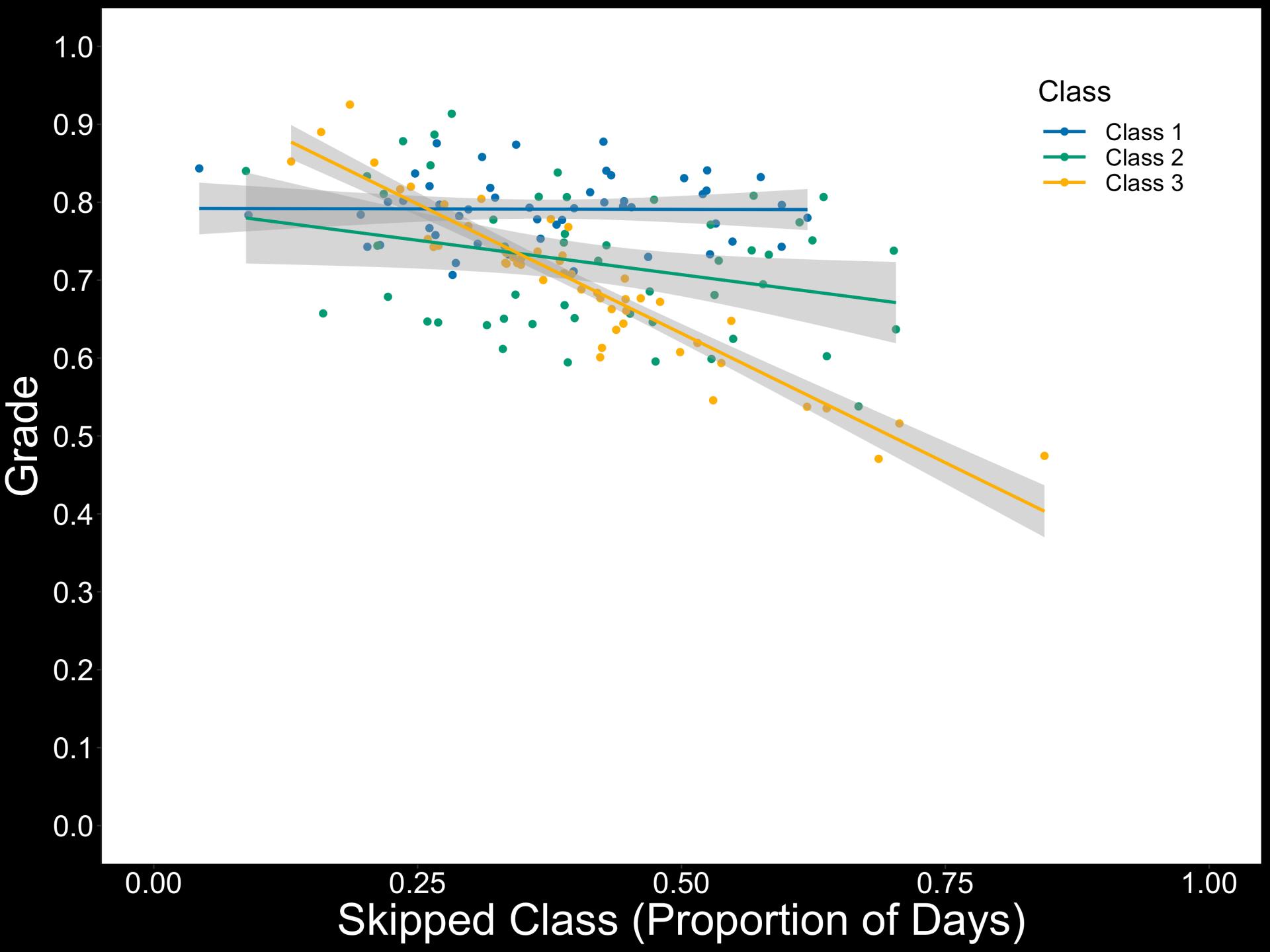
Violet Brown

Standard regression equation:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$







Regression assumptions

- **Normality assumption:** normally distributed residuals
- **Homogeneity of variance (i.e., homoscedasticity):** variance around the regression line (i.e., residual error) is constant across levels of predictors

Regression assumptions

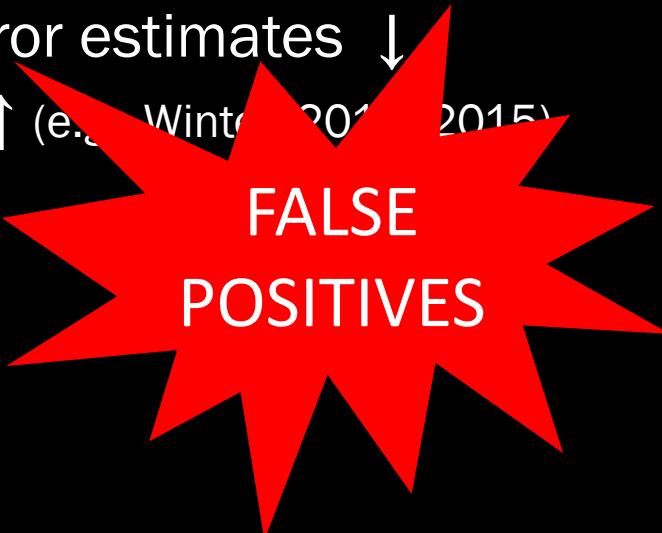
- **Normality assumption:** normally distributed residuals
- **Homogeneity of variance (i.e., homoscedasticity):** variance around the regression line (i.e., residual error) is constant across levels of predictors
- **Independence assumption:** errors are not correlated

The independence assumption

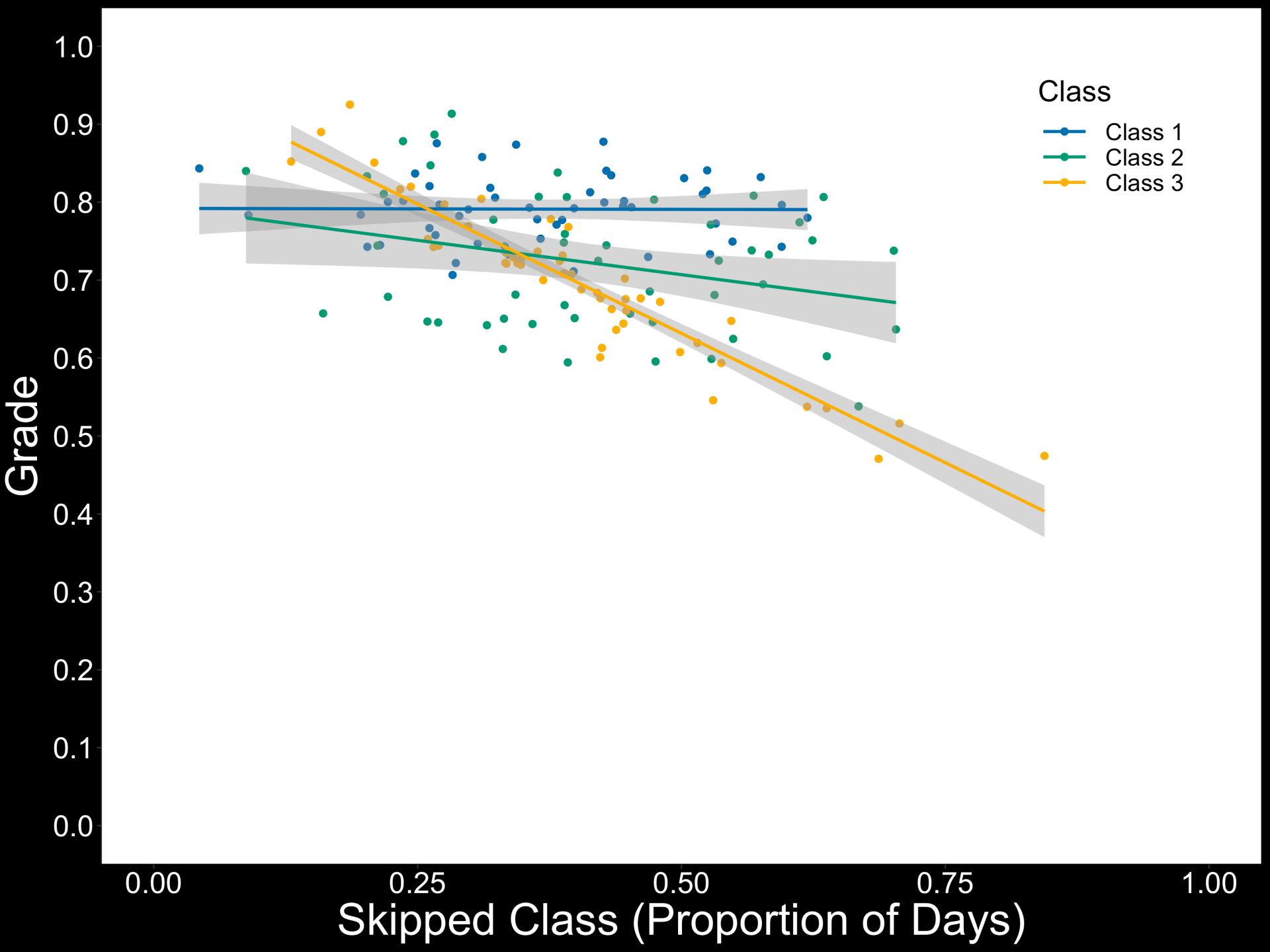
- Violating it is bad
 - Amounts to artificially inflating sample size

The independence assumption

- Violating it is bad
 - Amounts to artificially inflating sample size
 - Standard error estimates ↓
 - Type I error ↑ (e.g. Winter 2014–2015)



FALSE
POSITIVES

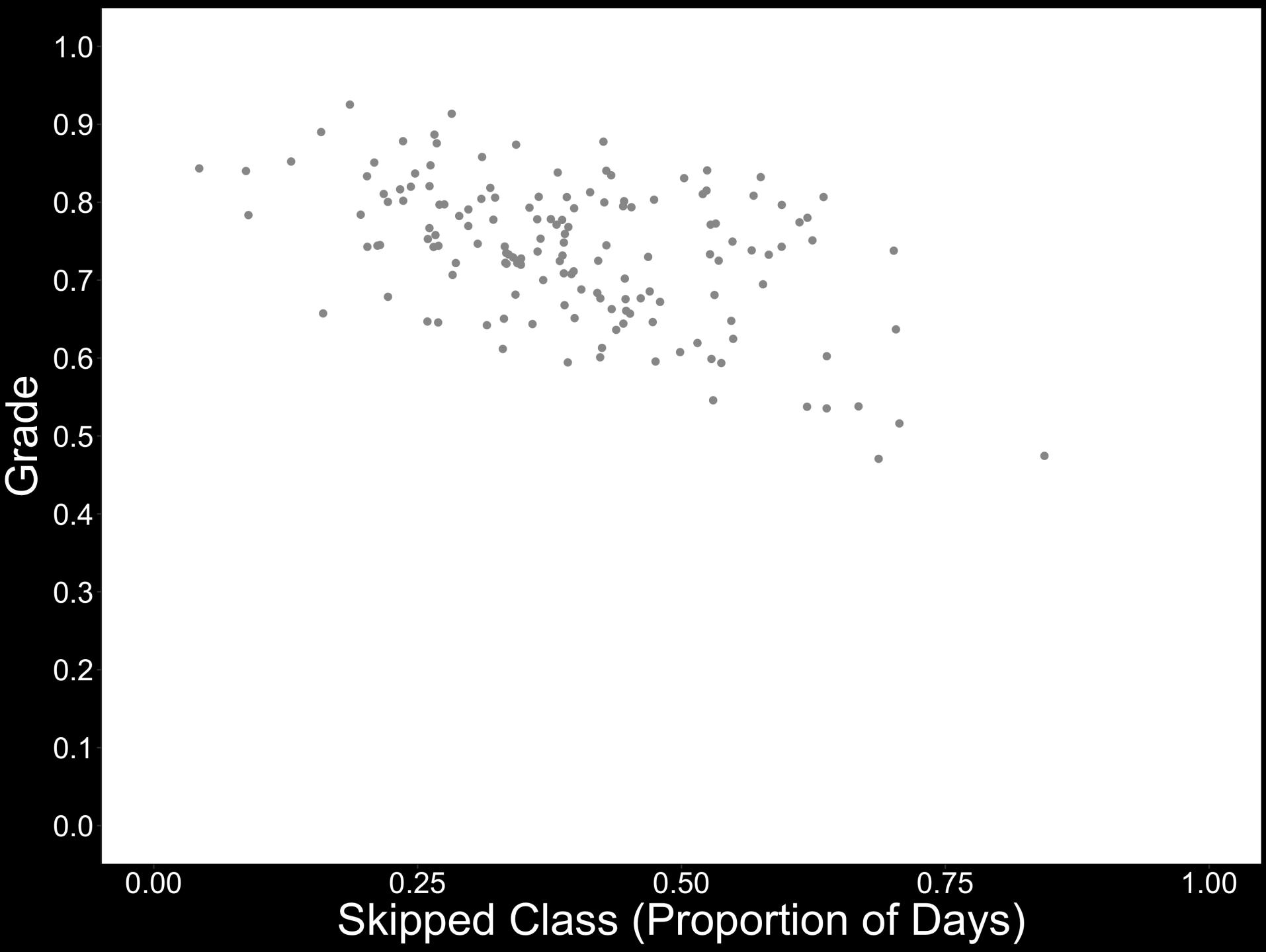


So you have dependencies in your data...now what?

- Design choices (e.g., one student per class)
 - Reduces statistical power
 - Isn't feasible for some experiments (e.g., response time studies)

So you have dependencies in your data...now what?

- Design choices (e.g., one student per class)
 - Reduces statistical power
 - Isn't feasible for some experiments (e.g., response time studies)
- Aggregation (i.e., means are entered into the analysis)
 - The model will underestimate variability





So you have dependencies in your data...now what?

- Design choices (e.g., one student per class)
 - Reduces statistical power
 - Isn't feasible for some experiments (e.g., response time studies)
- Aggregation (i.e., means are entered into the analysis)
 - The model will underestimate variability
 - You lose information about variability within the grouping factor
 - The model has less information to work with!

Home Price

0.8

0.4

0.0

0.00

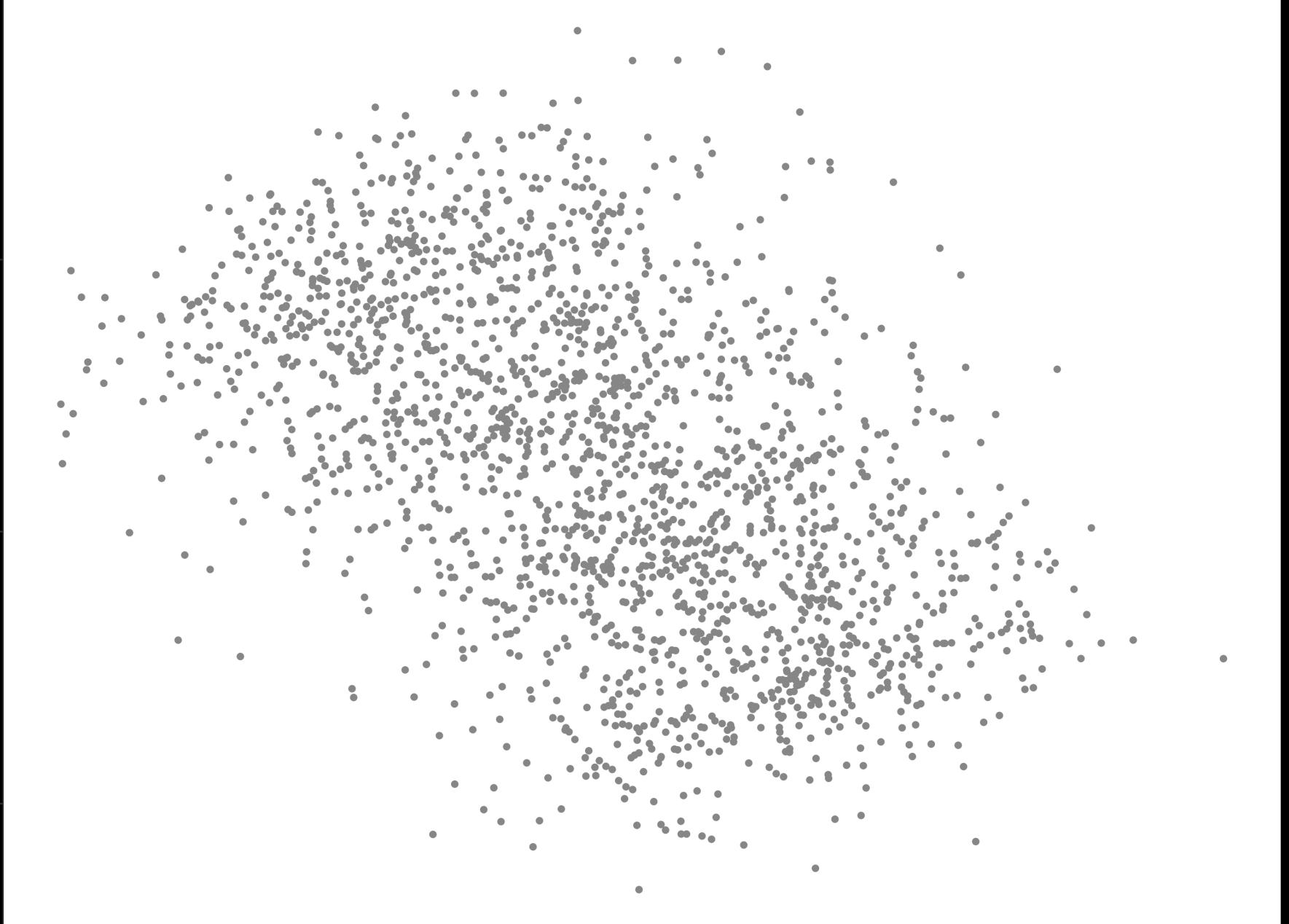
0.25

0.50

0.75

1.00

Number of Bedrooms



Coefficients:

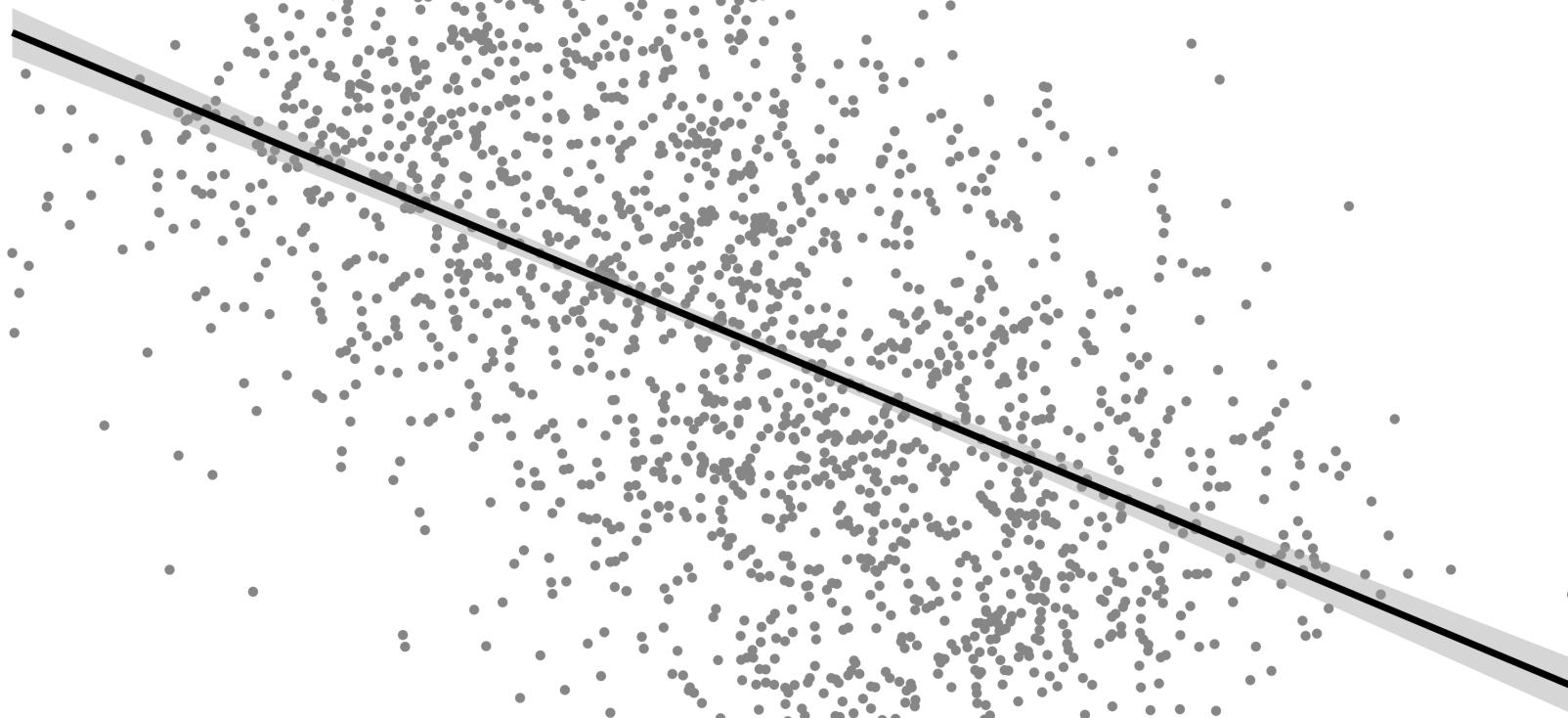
| | Estimate | Std. Error | t value | Pr(> t) | |
|-------------|----------|------------|---------|------------|--|
| (Intercept) | 0.85236 | 0.01449 | 58.84 | <2e-16 *** | |
| n_bedrooms | -0.75994 | 0.02923 | -26.00 | <2e-16 *** | |

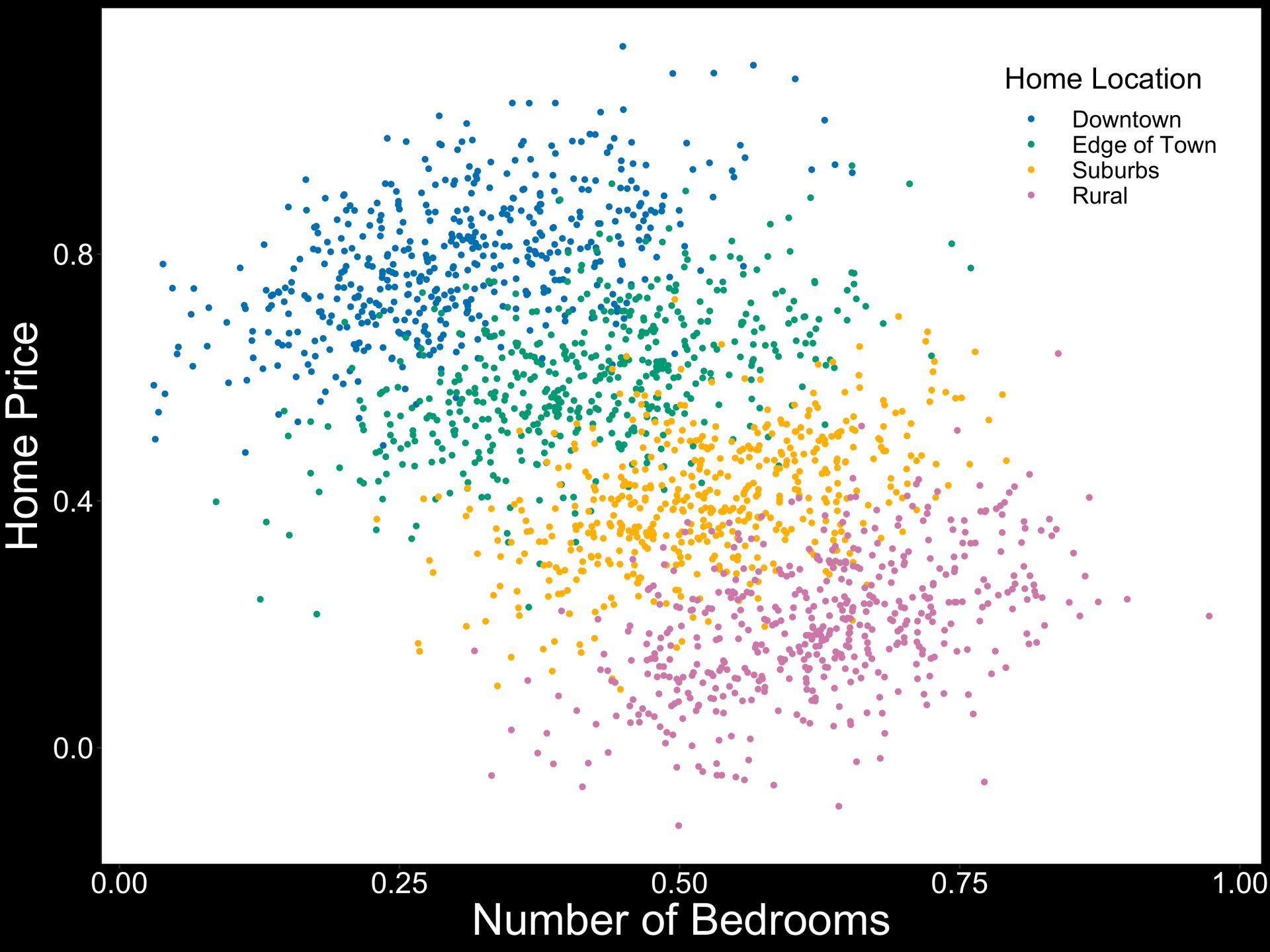
Home Price

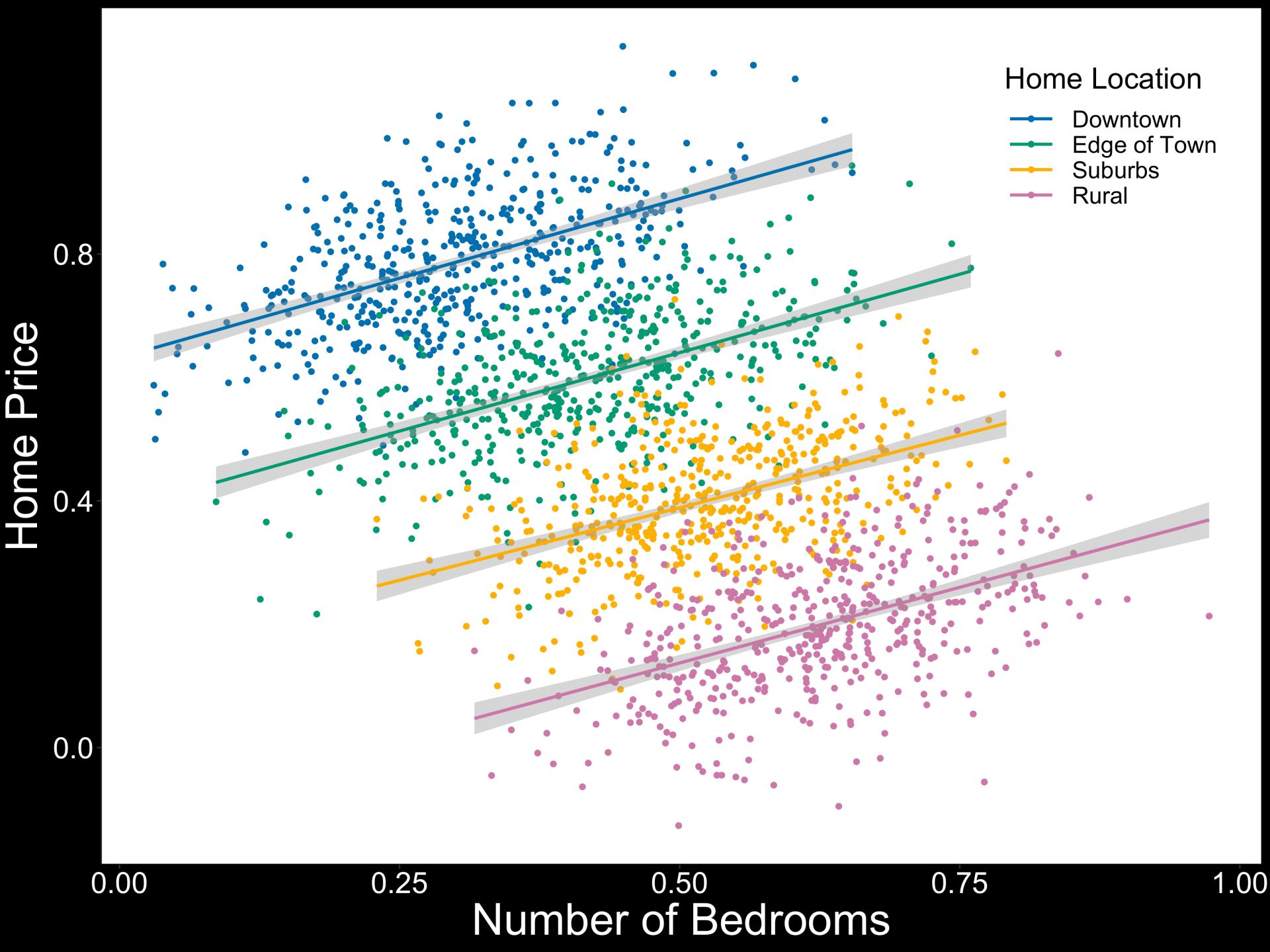
0.8
0.4
0.0

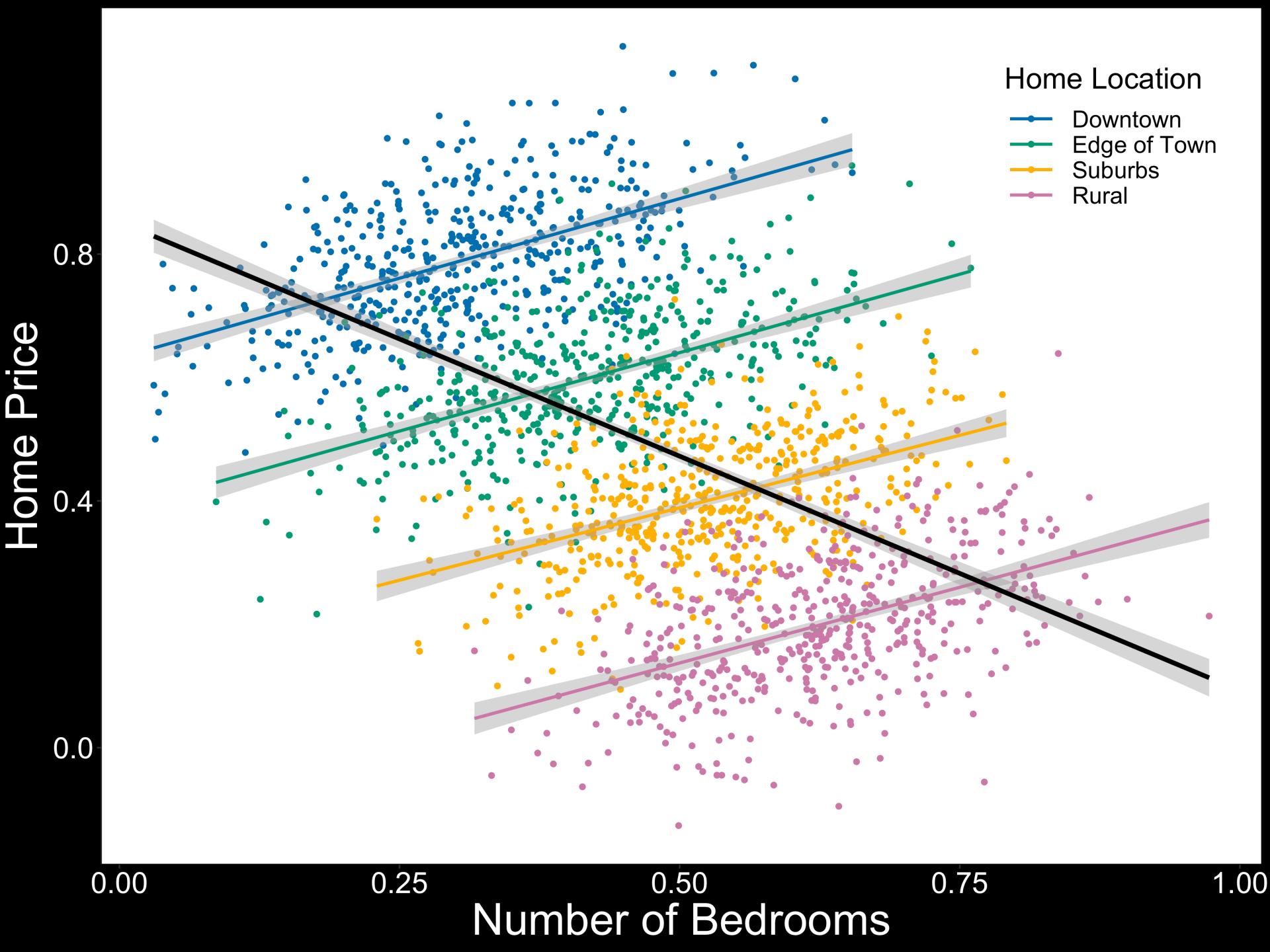
0.00 0.25 0.50 0.75 1.00

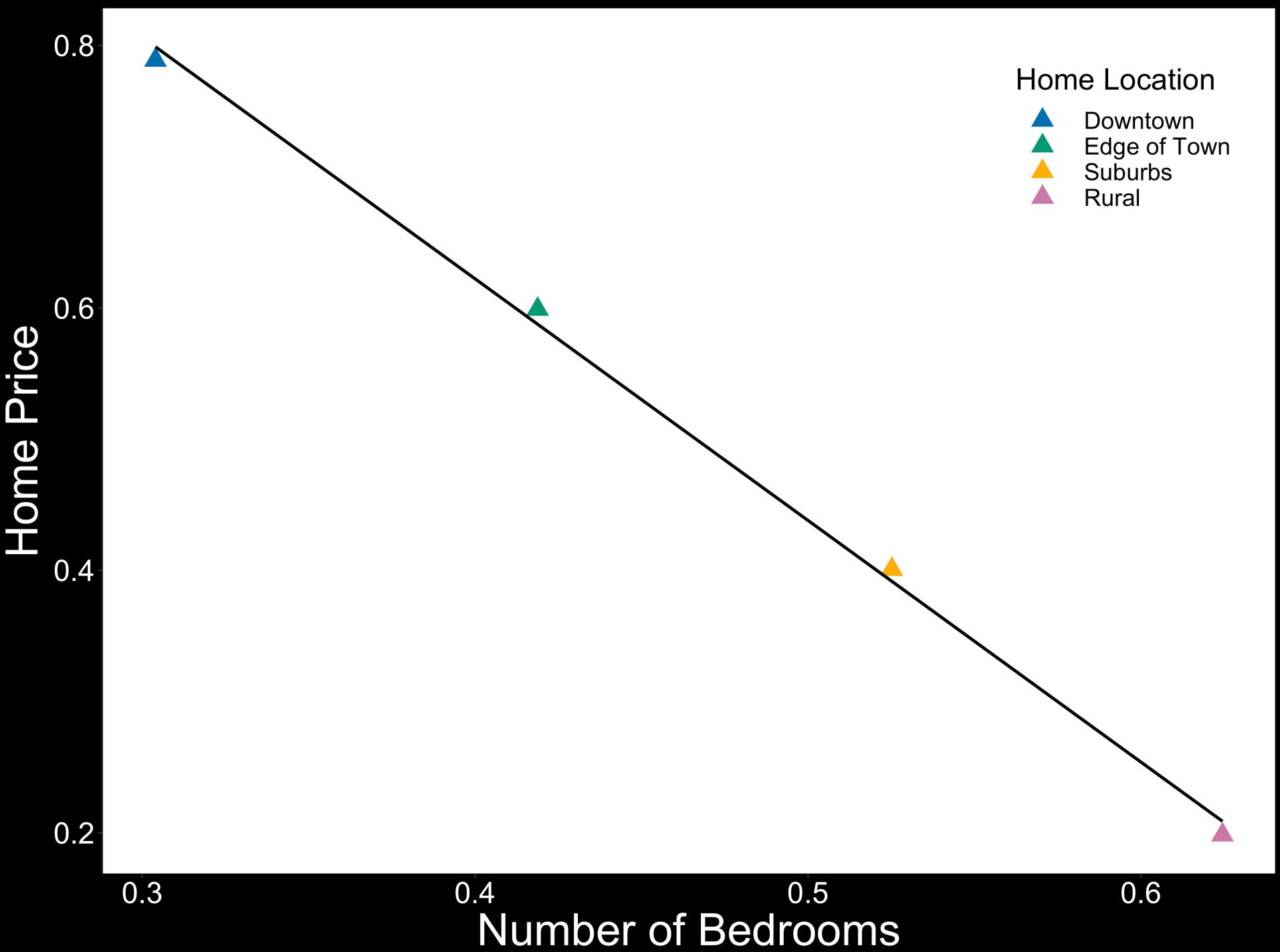
Number of Bedrooms

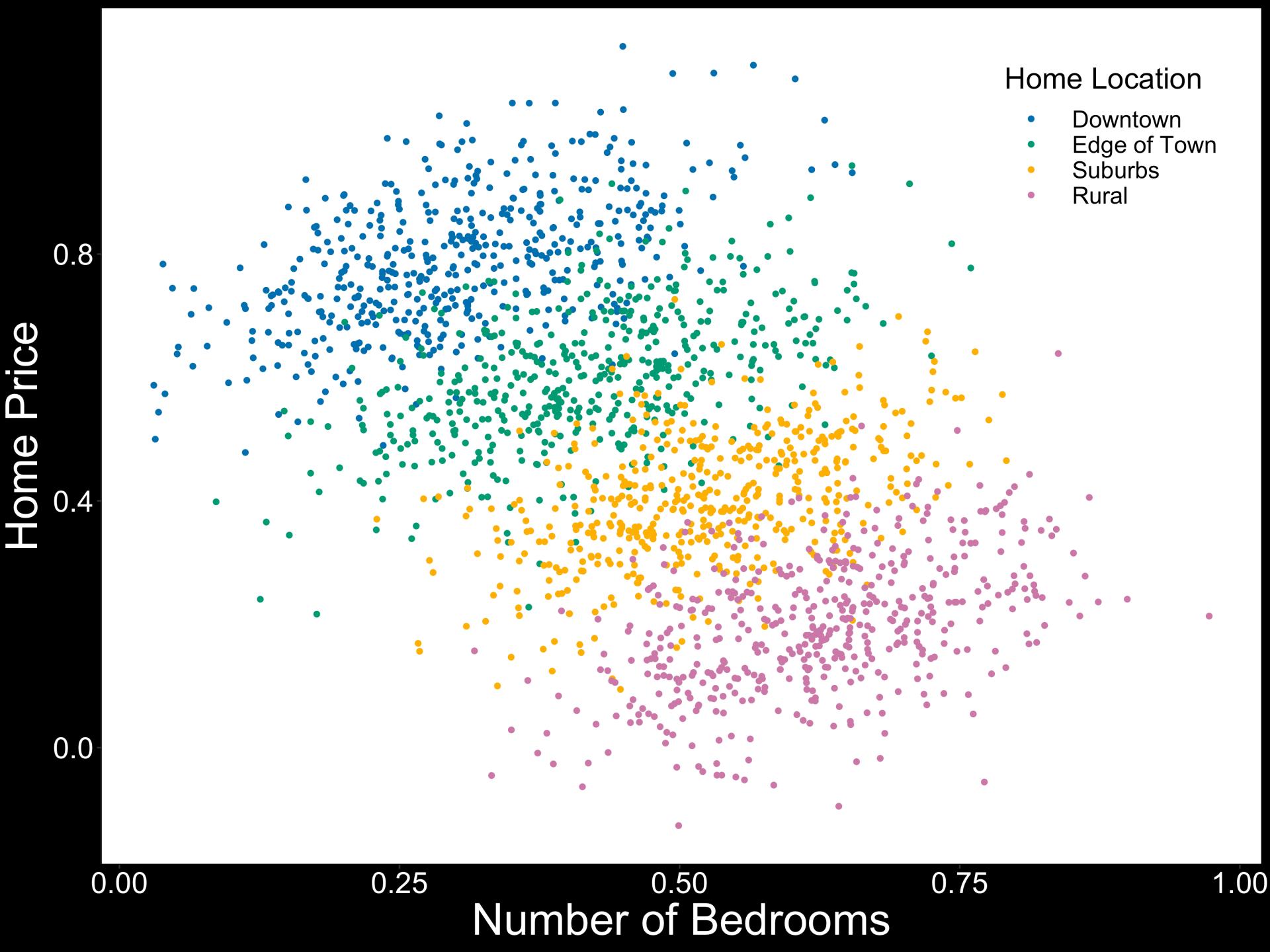


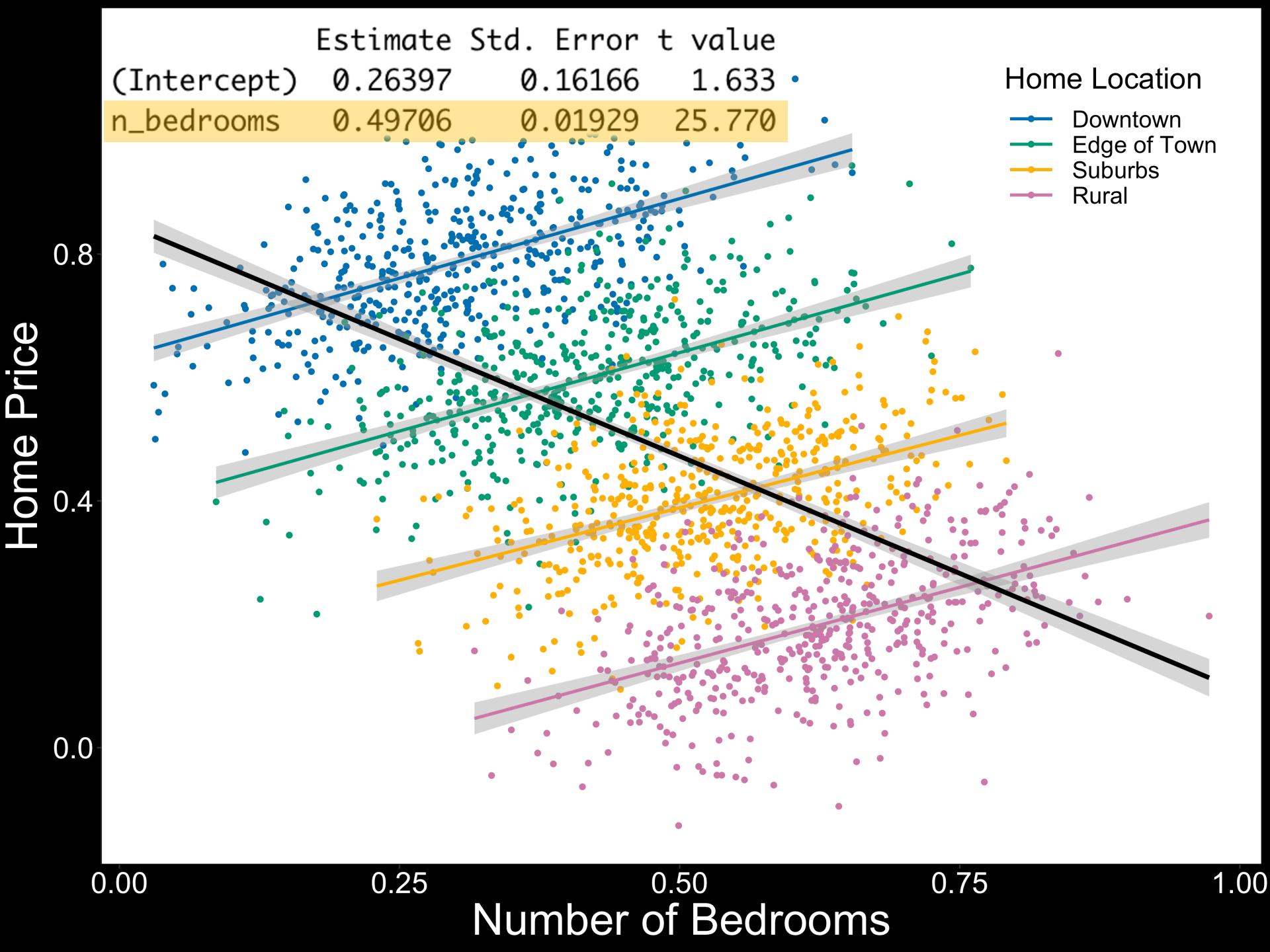












Repeated measures ANOVAs

- Do not tell you about the direction or magnitude of the effect
- Aggregate across either participants or items

Repeated measures ANOVAs

- Do not tell you about the direction or magnitude of the effect
- Aggregate across either participants or items
 - By-participant and by-item analyses can produce different results

Repeated measures ANOVAs

- Do not tell you about the direction or magnitude of the effect
- Aggregate across either participants or items
 - By-participant and by-item analyses can produce different results
- Assume continuous dependent variable and categorical independent variables

Repeated measures ANOVAs

- Do not tell you about the direction or magnitude of the effect
- Aggregate across either participants or items
 - By-participant and by-item analyses can produce different results
- Assume continuous dependent variable and categorical independent variables
- LMER is more flexible; framework extends to `glm`

Mixed-effects models take the stage

- Examples of nesting and dependencies:
 - Students within classes (within schools!)
 - Observations within people

Mixed-effects models take the stage

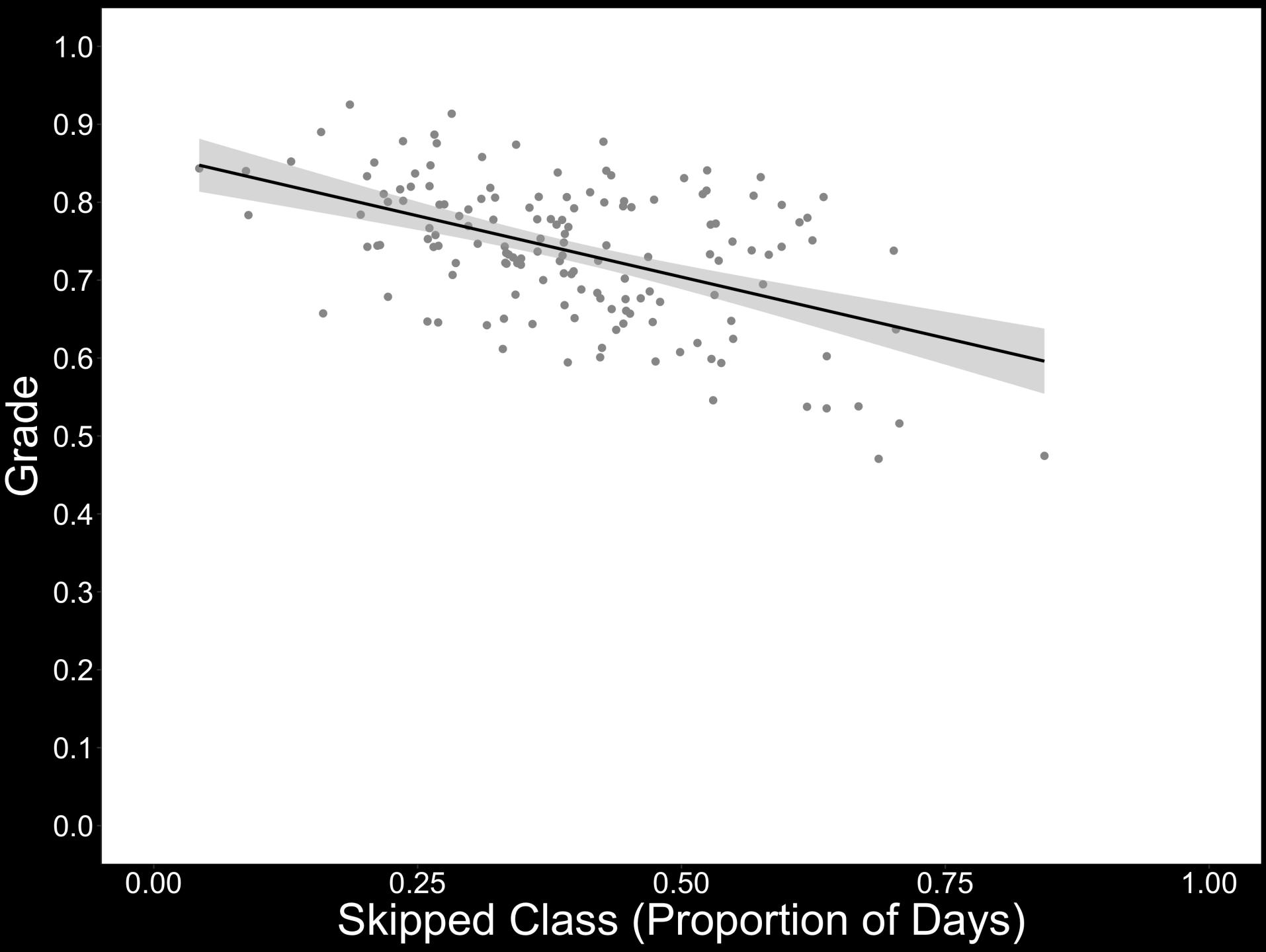
- Examples of nesting and dependencies:
 - Students within classes (within schools!)
 - Observations within people
 - Observations within items
 - People within counties (within states!)
 - Etc. etc.
- Mixed models can simultaneously model multiple dependencies

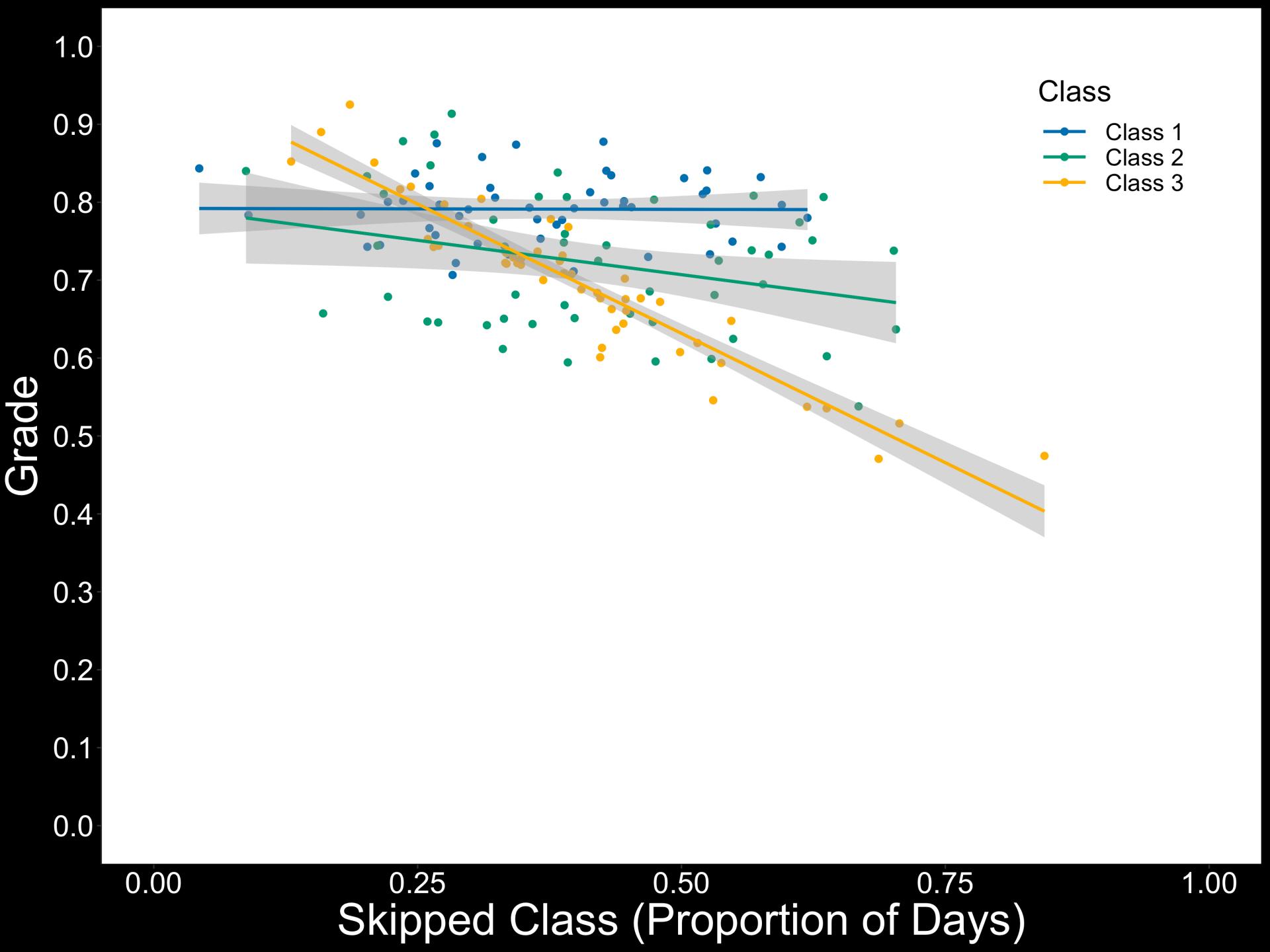
Fixed and random effects

- **Fixed effects:** population-level (i.e., average) trends that should persist across experiments
- **Random effects:** cluster-level trends

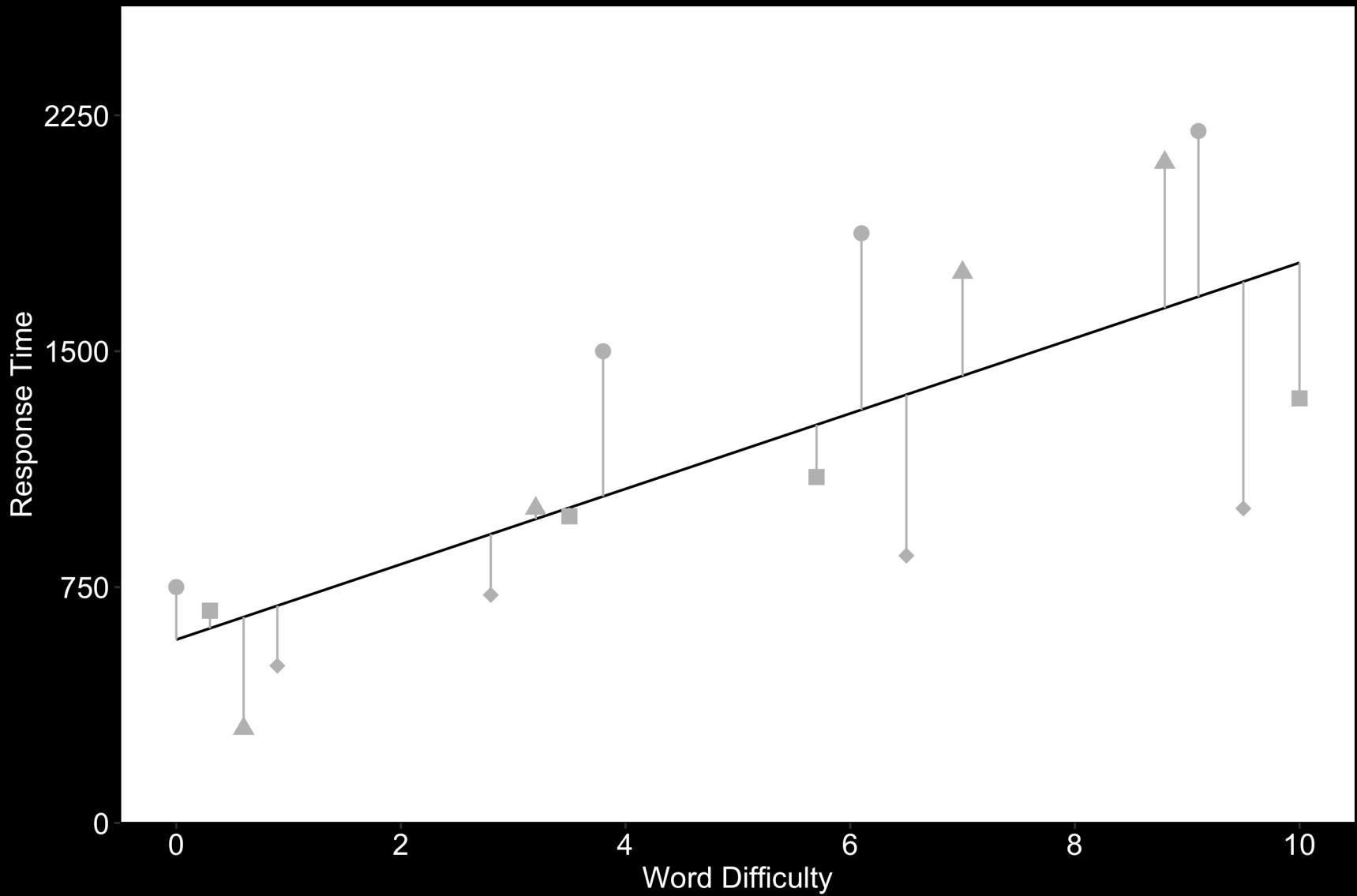
Fixed and random effects

- **Fixed effects:** population-level (i.e., average) trends that should persist across experiments
- **Random effects:** cluster-level trends
 - Clusters of *dependent* data points coming from the same higher-level group (e.g., participant, item, classroom, school)
 - Particular participants or items may behave differently from the average trend

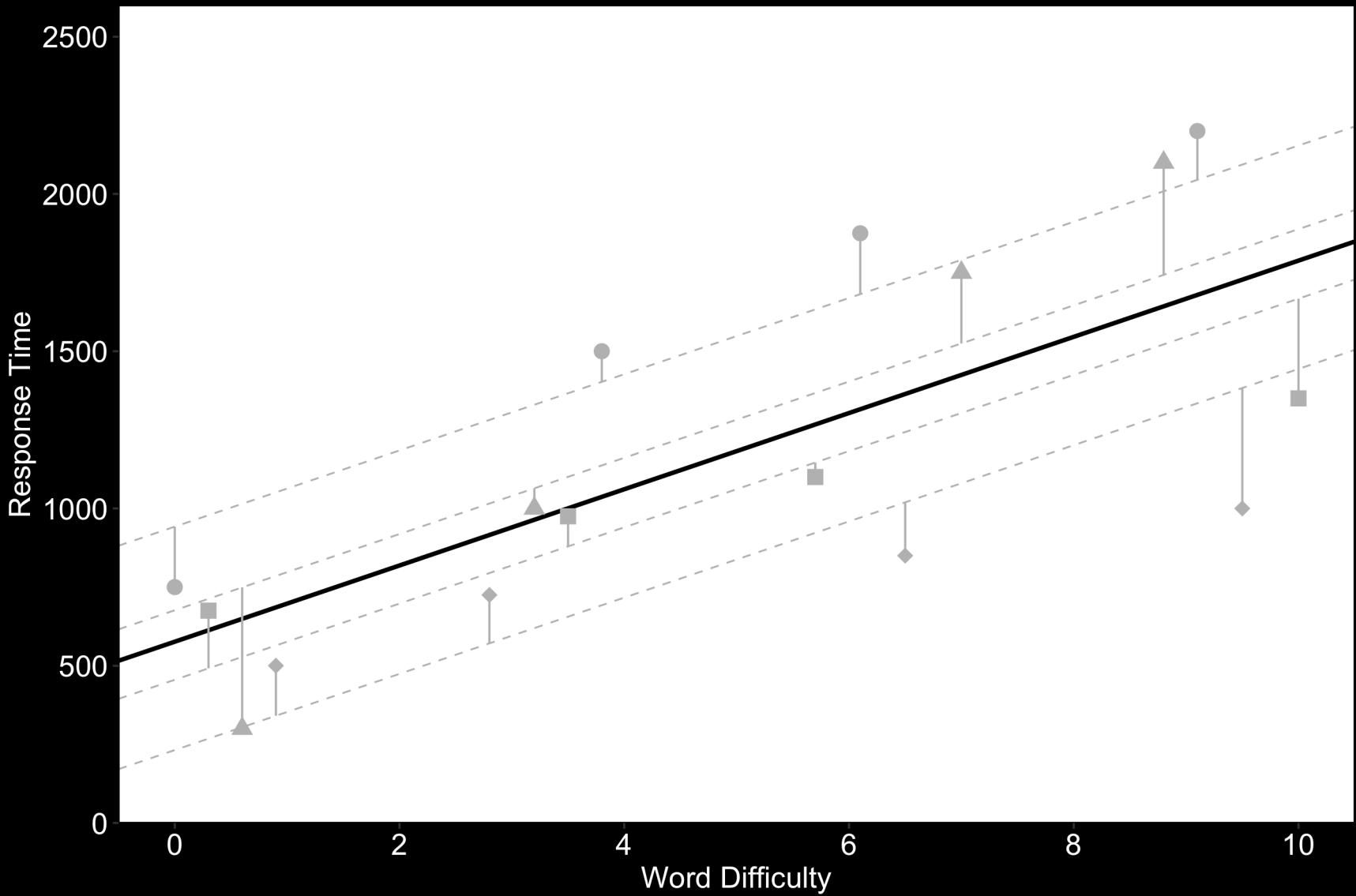




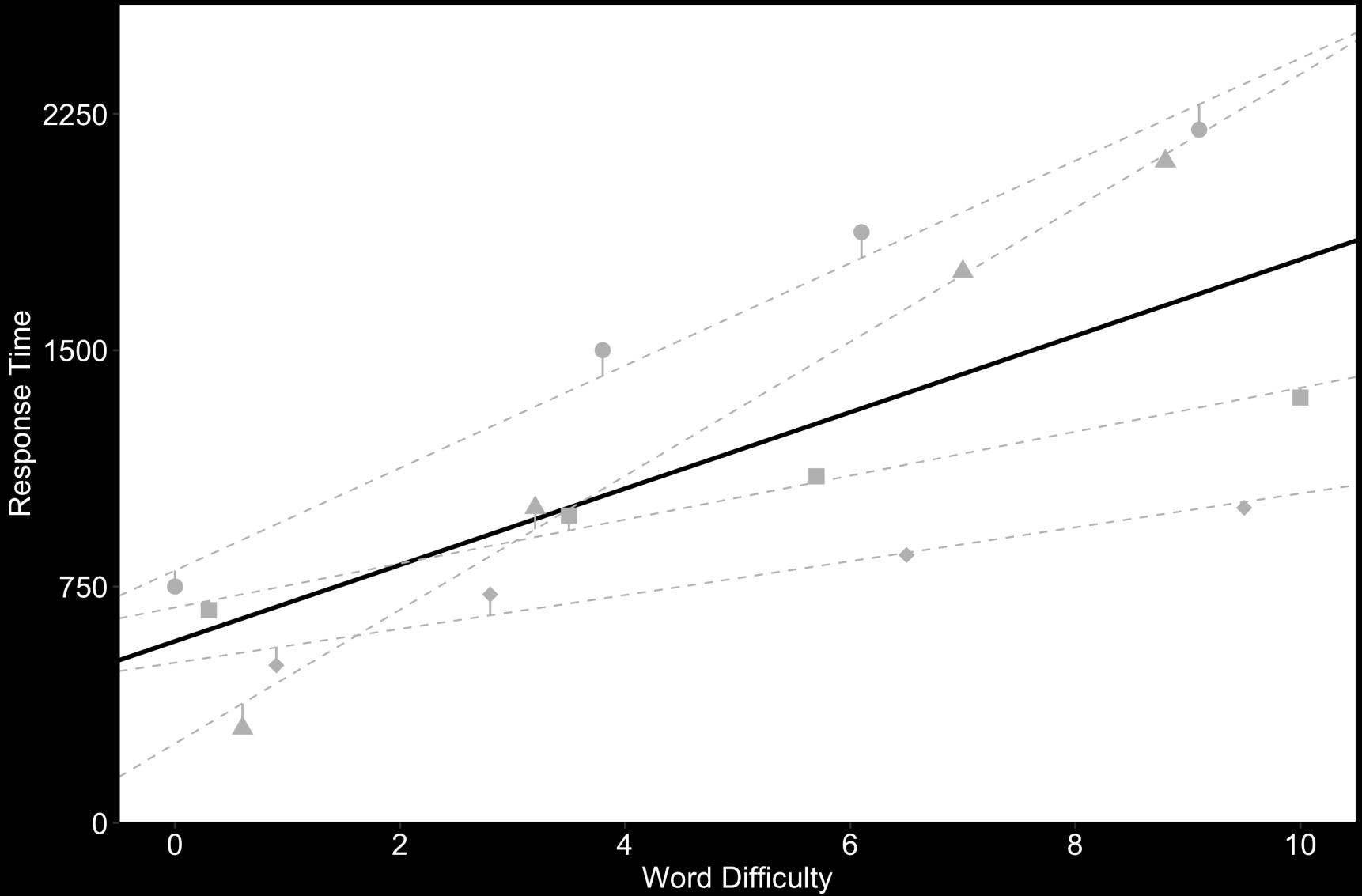
Fixed effects only regression



Adding random intercepts



Adding random slopes



Summary

- The independence assumption is important
- Aggregation masks variability
- Repeated-measures ANOVAs have issues
- Mixed-effects models are the answer

Summary

- The independence assumption is important
- Aggregation masks variability
- Repeated-measures ANOVAs have issues
- Mixed-effects models are the answer
- Fixed and random effects

β u γ ! ϵ