Introduction to Telling Stories with Data

Author: Nicholas G Reich

This material is part of the **statsTeachR** *project*

Communicating ideas with evidence

What is a narrative?

What is data?

Examples

[example headlines/images from blog posts/articles]

Examples

[example headlines/images from blog posts/articles]

How to tell a story using data

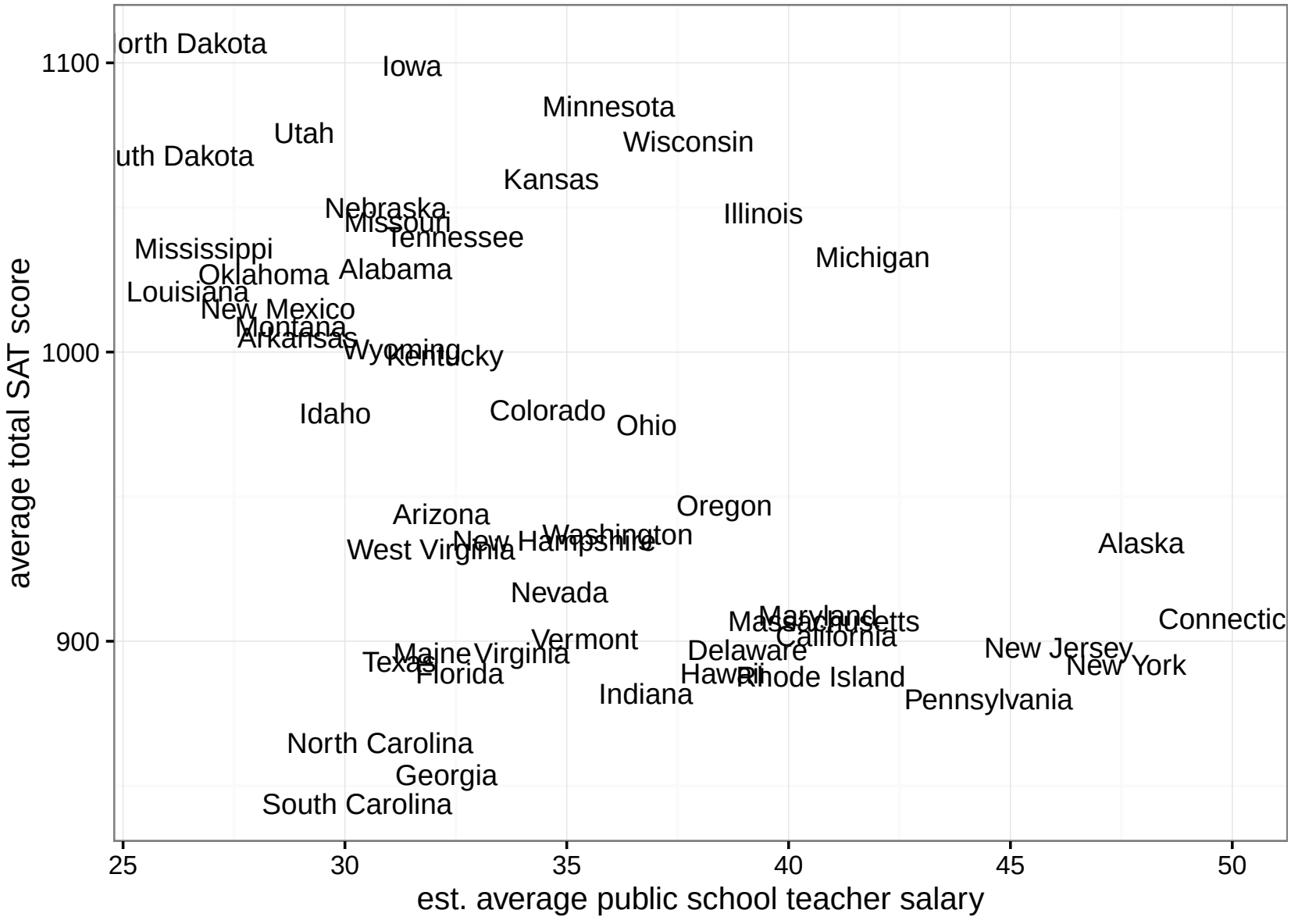
[add some images here] Telling stories with data requires

- detective work
- creativity, both scientific and artistic
- experimentation
- good data, and note good data does not nescessarily equal big data

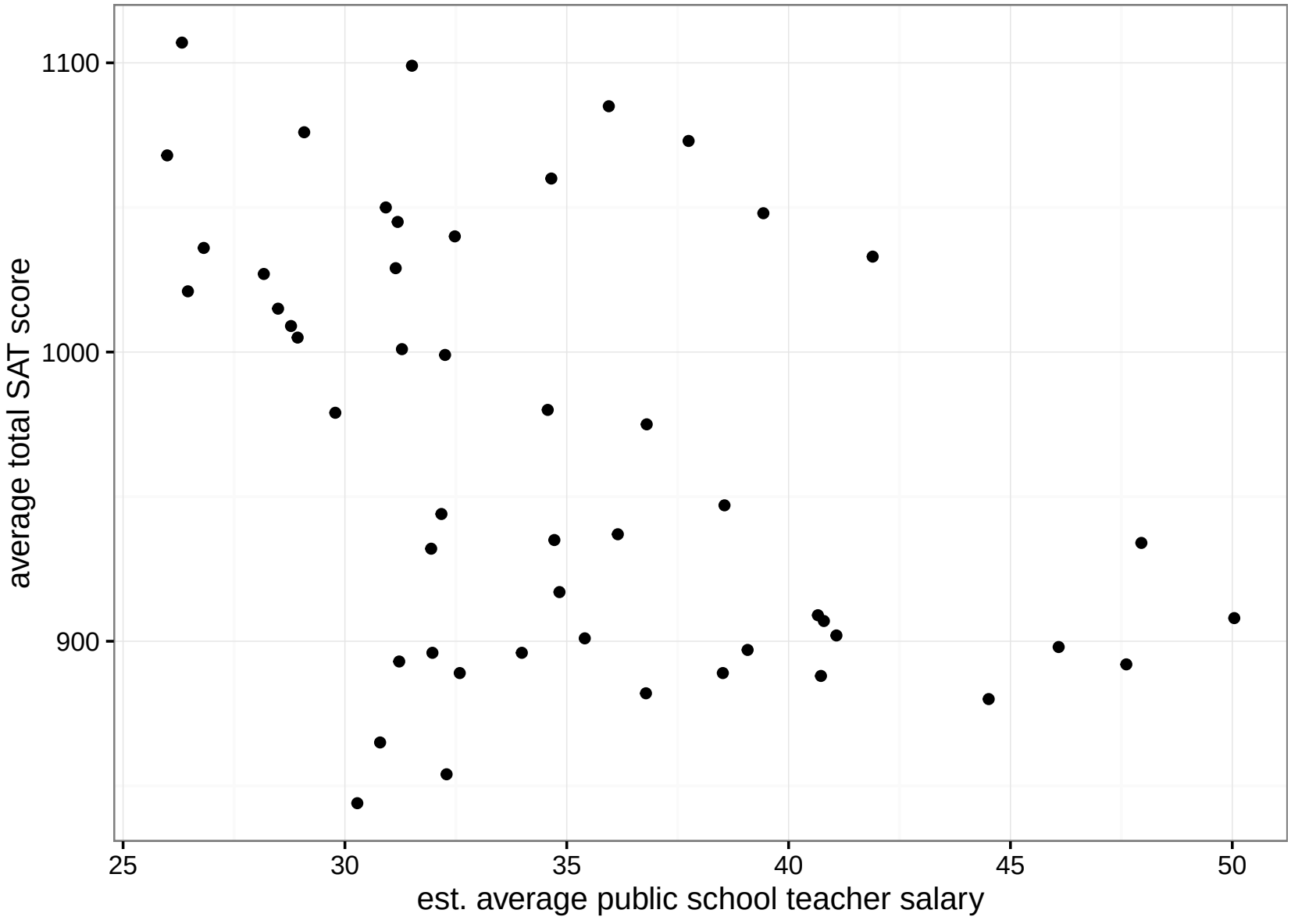
A common tool: regression

- The goal is to learn about the relationship between a covariate (predictor) of interest and an outcome of interest.
  - Some models focus on prediction.
  - Other models focus on description.
- Regression is an exercise in inferential statistics: we are drawing evidence and conclusions from data about "noisy" systems.

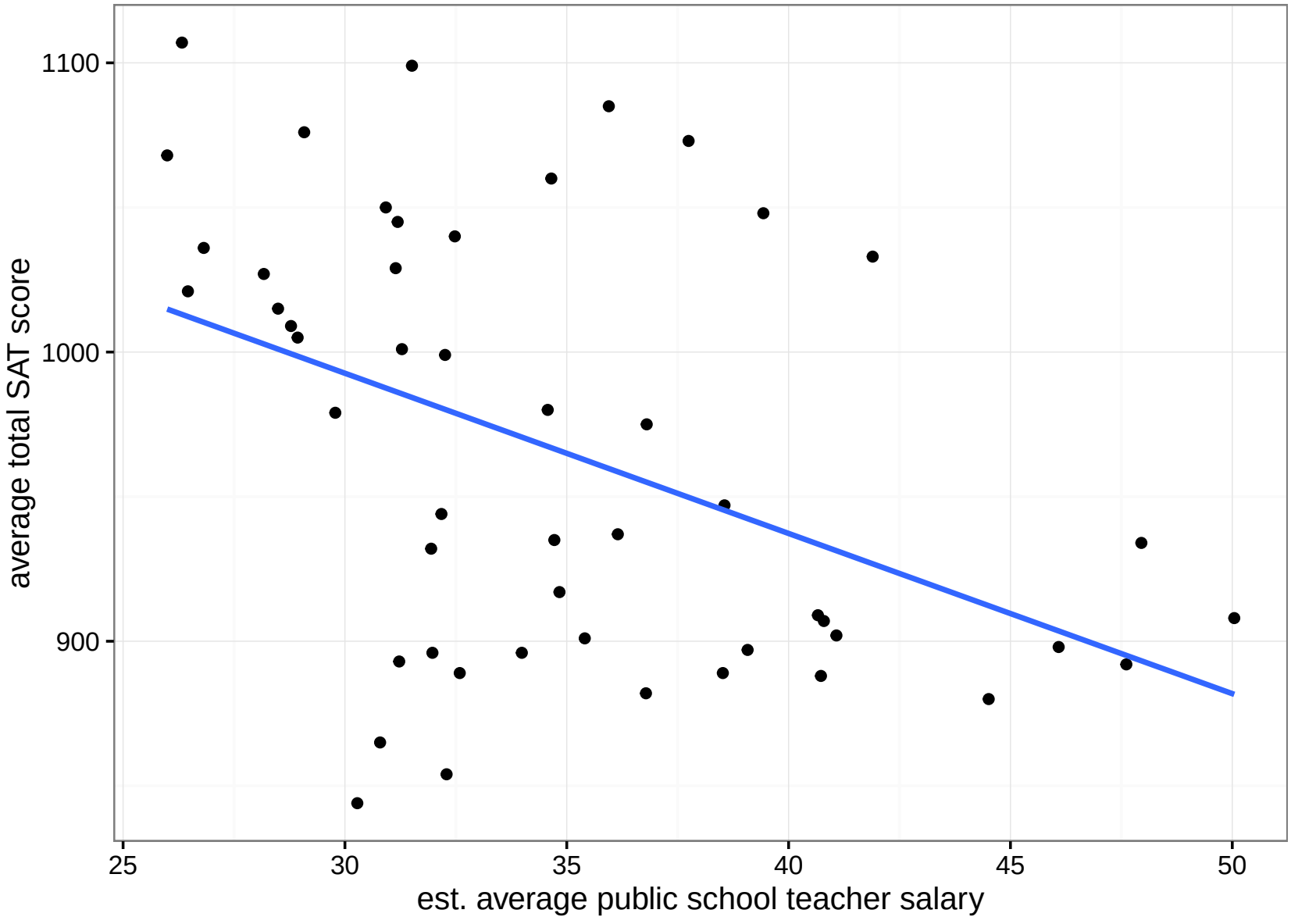
State-level SAT score data (1994-95)



State-level SAT score data (1994-95)



State-level SAT score data (1994-95)



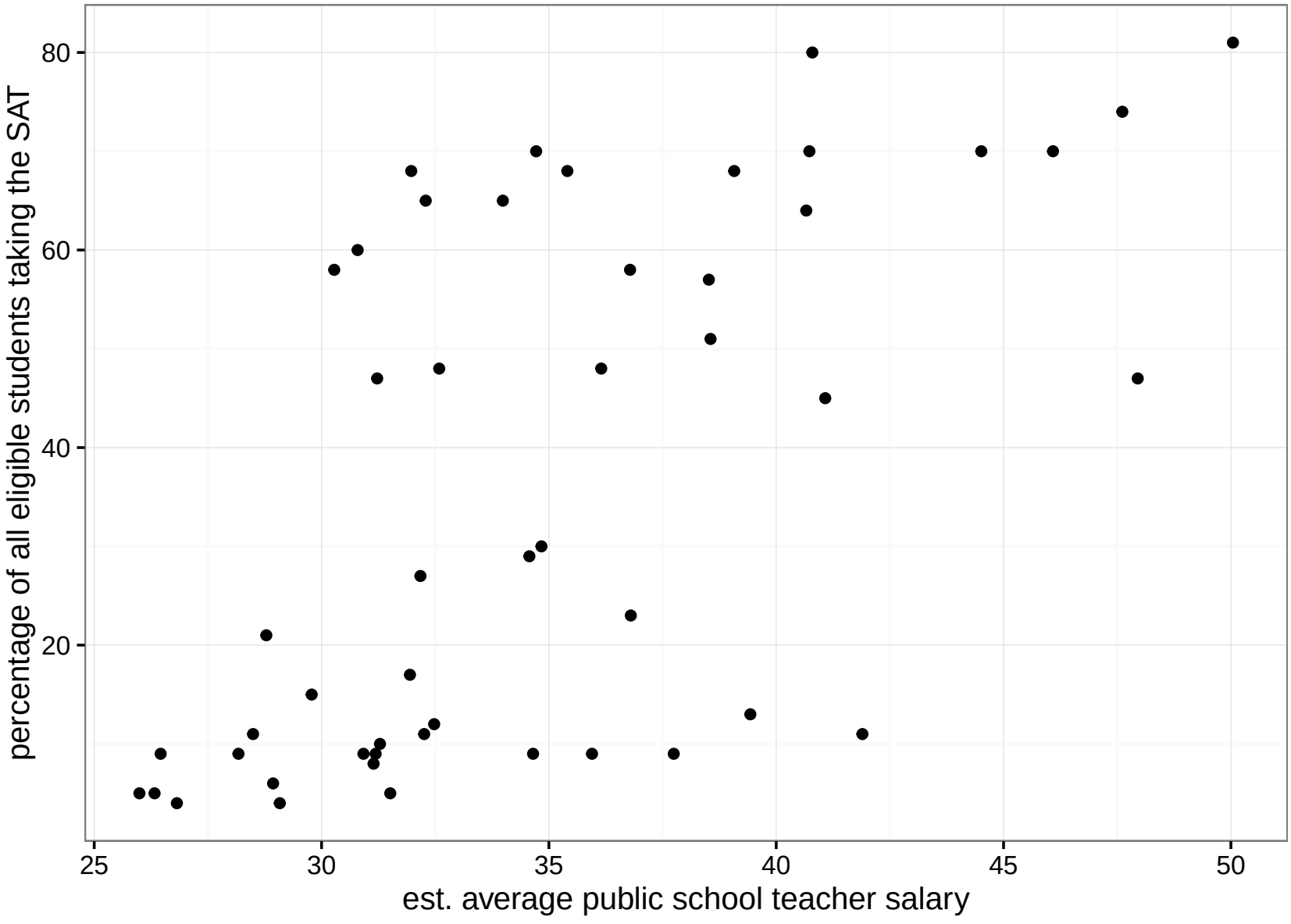
The SAT example

What is the outcome variable?

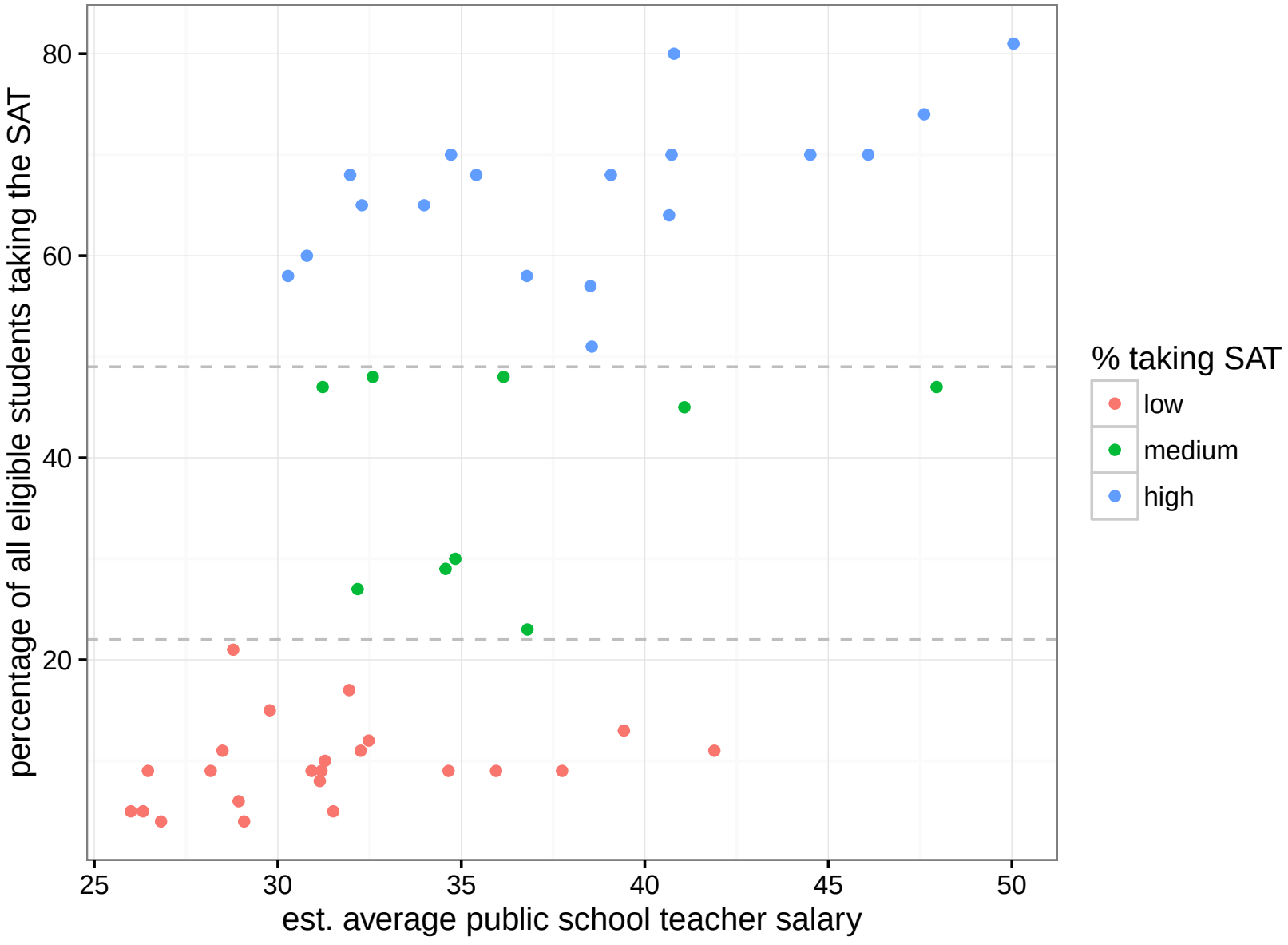
What is the covariate or predictor variable?

What other data might be part of this story?

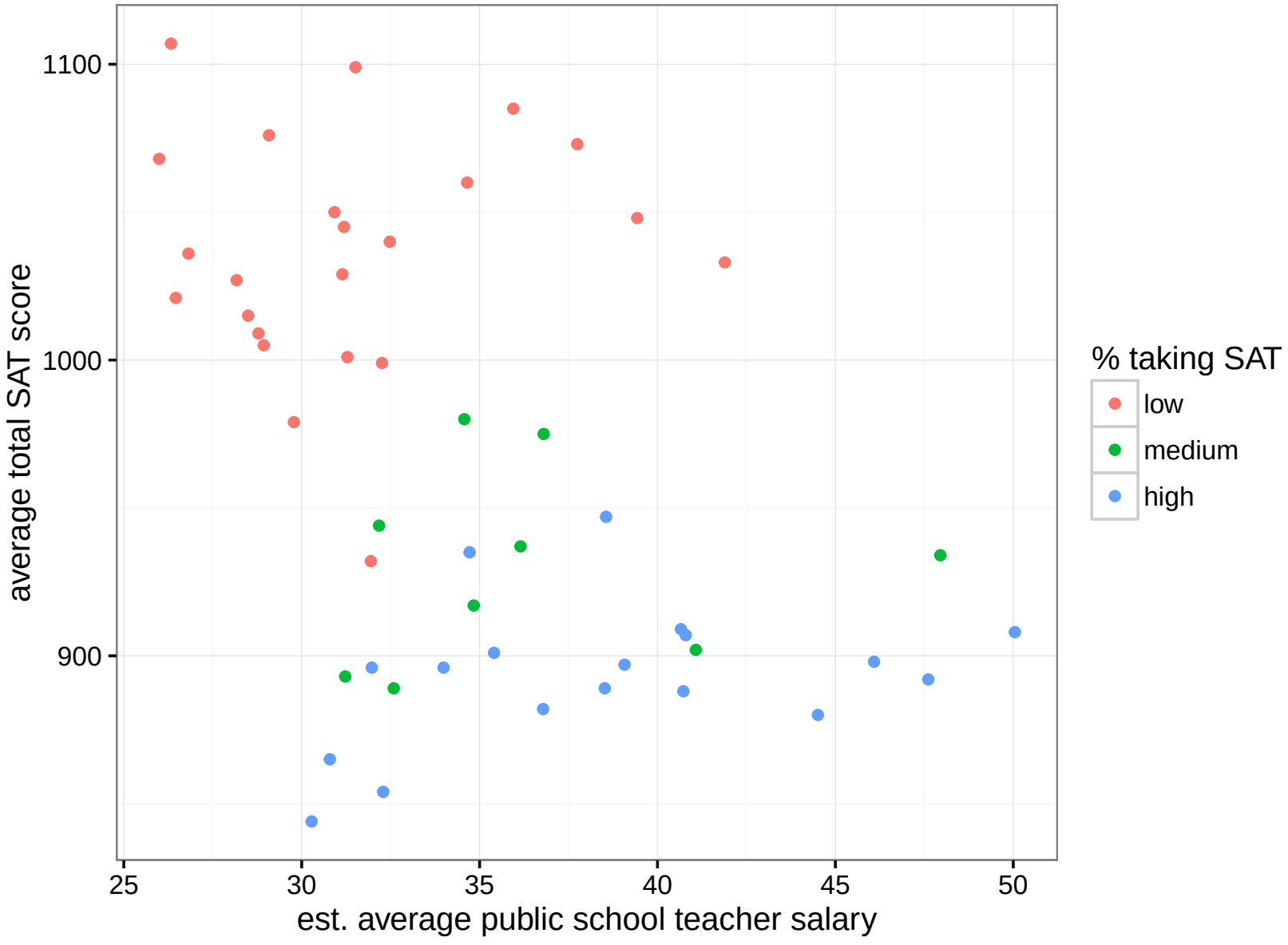
State-level SAT score data (1994-95)



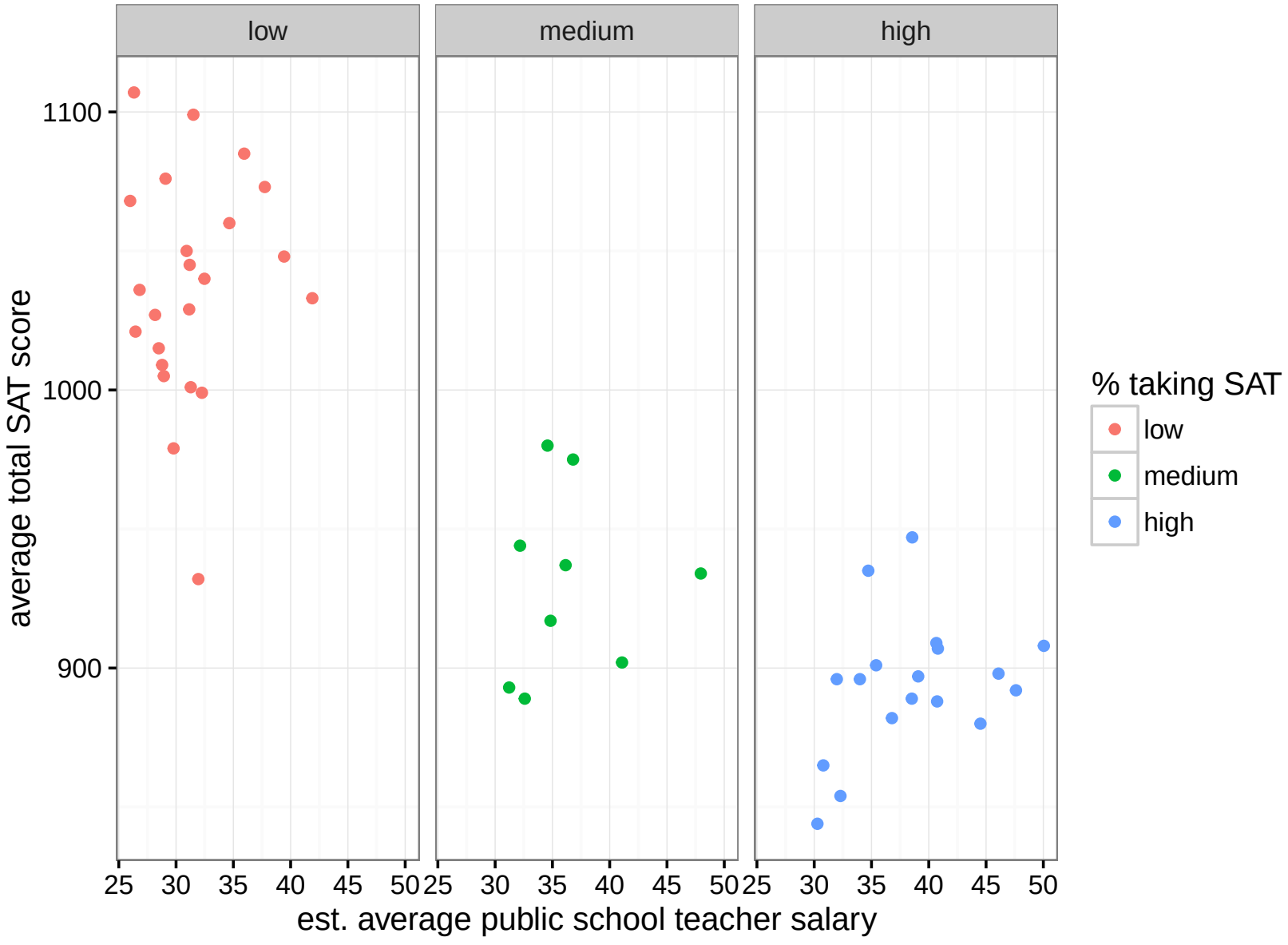
State-level SAT score data (1994-95)



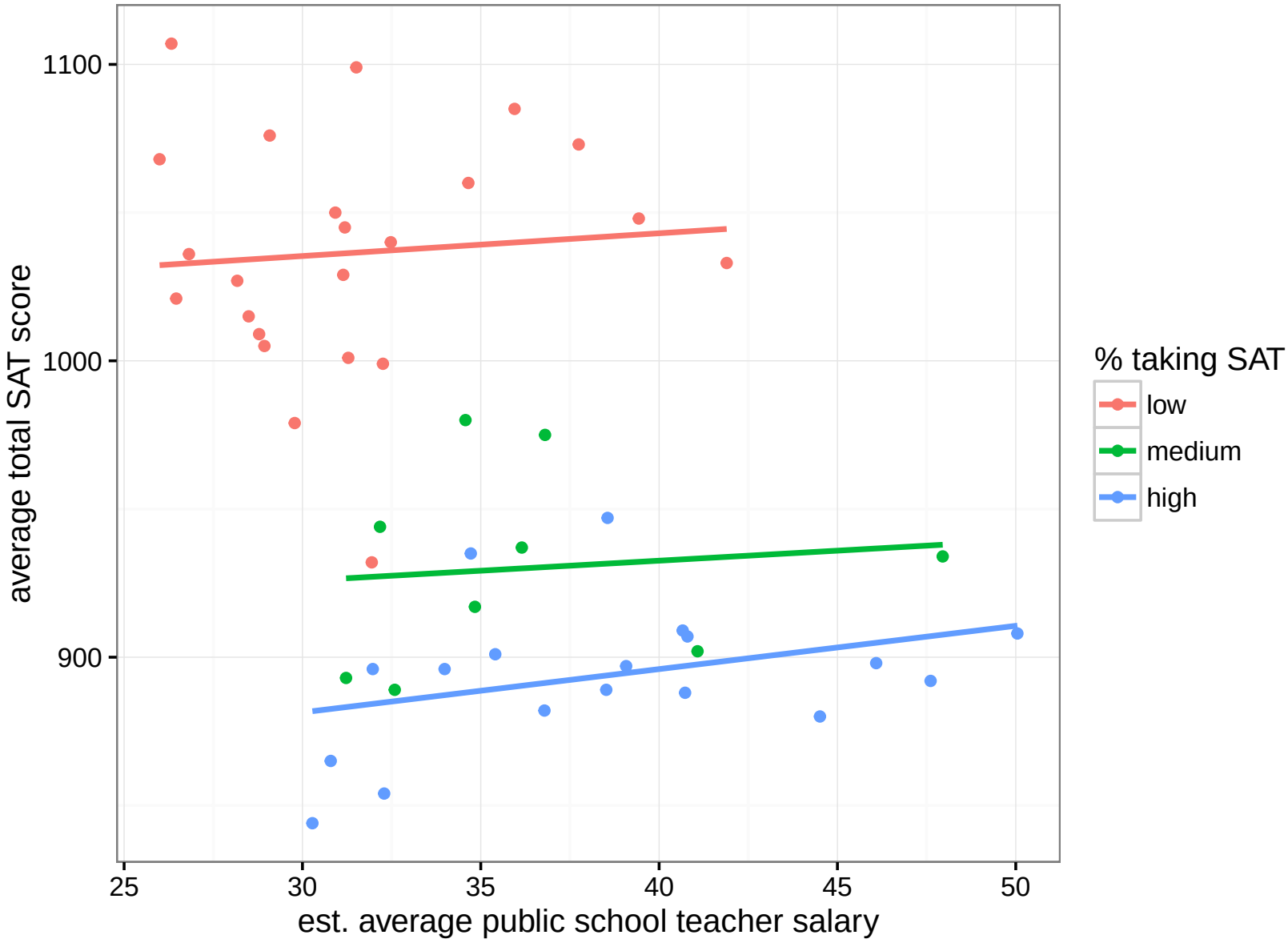
State-level SAT score data (1994-95)



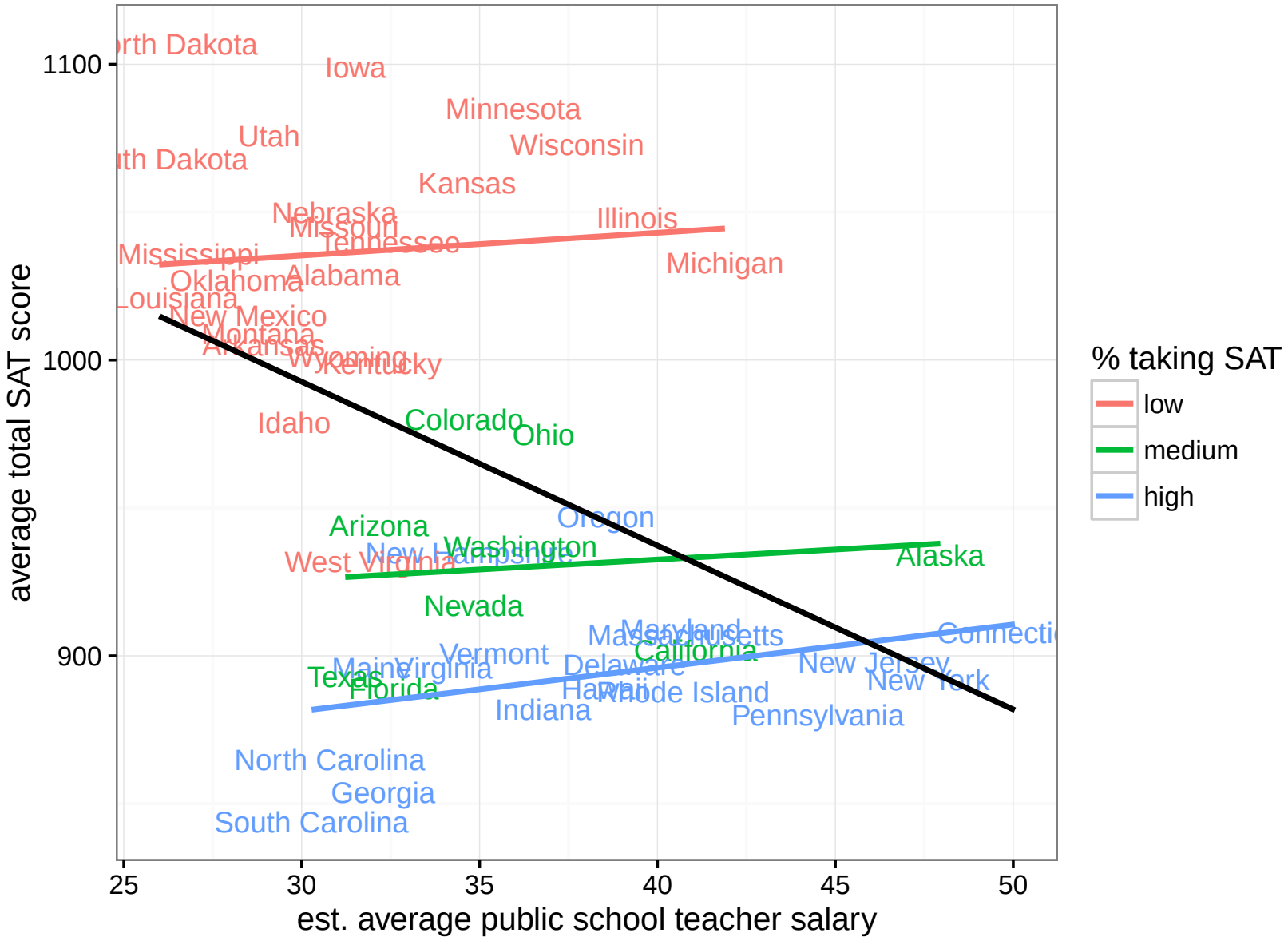
State-level SAT score data (1994-95)



State-level SAT score data (1994-95)



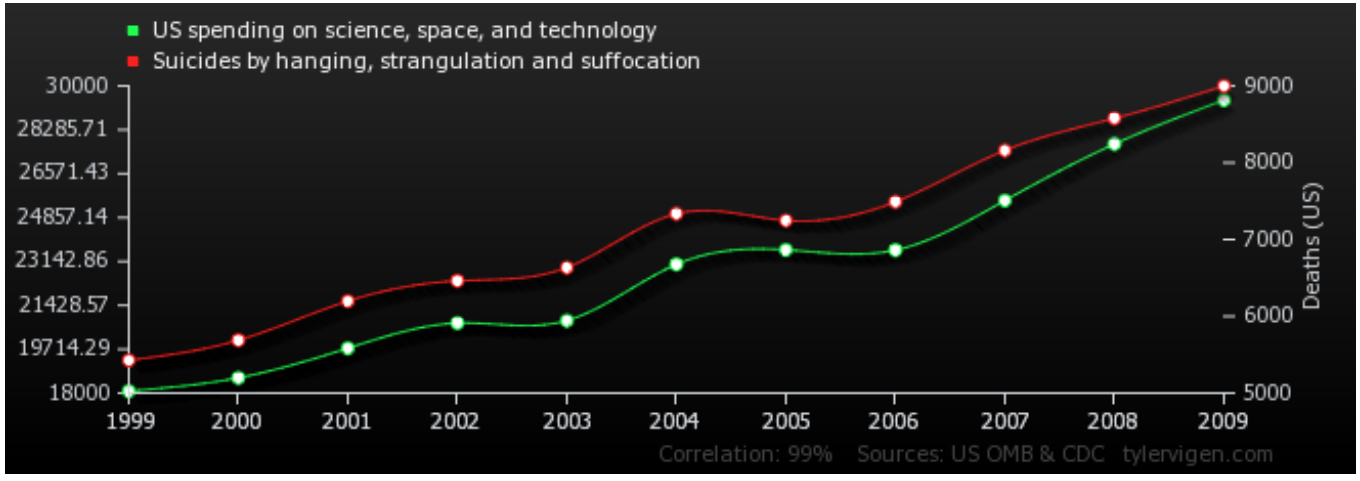
State-level SAT score data (1994-95)



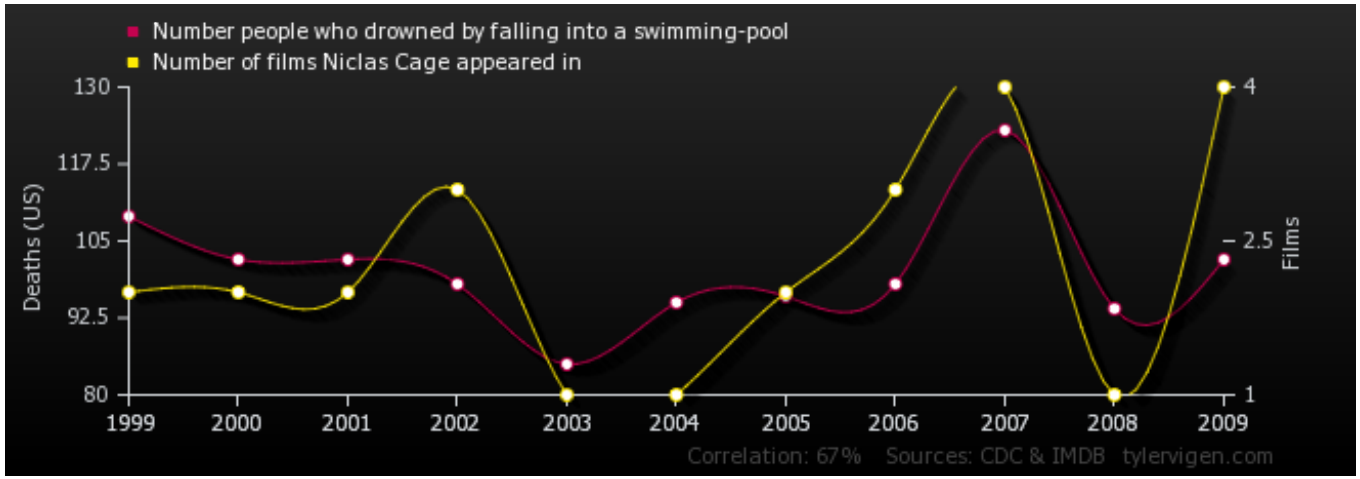
State-level SAT score data (1994-95)

What can we conclude from all of this? (BTW, this is an example of "Simpson's Paradox".)

Beware of correlation!



Beware of correlation!



1

[1] Hat tip to www.tylervigen.com

Regression modeling

The process of using data to describe the relationship between outcomes and predictors is called modeling.

- Models are models, not reality.
- "All models are wrong, but some are useful."
- Introduce structure to our model that balances realism with "goodness of fit".

Things to come

- Tools to help tell stories with data.
  - Software
  - Statistical methods
- Practice developing and conceiving models/stories.