

School Performance¹

August 2016

¹Copyright 2016 R. A. Reitmeyer. Released under Creative Commons CC-BY 4.0 license.

Exercise

- ▶ Your objective for the rest of the night is to try to model school district performance on the CAASPP for some bay area communities
 - ▶ CA Assessment of Student Performance and Progress
- ▶ QUESTION: What is “percentage standard met and above”, in terms of other columns?
 - ▶ Excluding other percentage..standard columns, of course!
- ▶ What's the best model for this, in the smallest-BIC sense?
- ▶ Data was assembled from CA Department of Education, US Census and was turned into geoJSON maps
- ▶ See <http://www.opensmc.org/LD3> for basic visualization
- ▶ Please grab data, and these slides, from [github://rareitmeyer/linear_regression_class](https://github.com/rareitmeyer/linear_regression_class)

Load Into R

- ▶ Won't use mapping for this, so use jsonlite package
 - ▶ Install if you don't have it via `install.packages('jsonlite')`
- ▶ Read in the data and drop the geo bits

```
library(jsonlite)
g5m_geo <- jsonlite::fromJSON(
  'grade_5_mathematics.geojson')
g5m <- g5m_geo$features$properties
dim(g5m)
```

```
## [1] 74 56
```

BIC

- ▶ As you try models, check the BIC

```
BIC(lm(percentage_standard_met_and_above ~  
      total_caaspp_enrollment, data=g5m))
```

```
## [1] 671.2097
```

```
BIC(lm(percentage_standard_met_and_above ~  
      CurrentExpenseADA, data=g5m))
```

```
## [1] 663.1436
```

- ▶ Advice: Don't hop straight into model-making. Start with graphs!
 - ▶ `library(ggplot2)` and `qplot(x, y, data=g5m)` are your friends!
 - ▶ Use `summary(g5m)` to see all columns and some info
 - ▶ Install the `car` package and look at `residualPlot` or `residualPlots!`