PLSC 503 – Spring 2017
Introduction

January 17, 2017

Housekeeping

- The github repo...
- Buy the books on-line...
- "Introduction to R and RStudio" on github...

Regression

"Regression," conceptually:

$$\Pr(Y|\mathbf{X}) = f(\mathbf{X})$$

Two important things:

- The distribution of Y is *conditional on all variables in* $\mathbf{X}$, and
- The conditional distribution of Y is conditional on the *joint distribution* of the elements of $\mathbf{X}$.

$\rightarrow$ Regression is hard...

Figure 1: Infant Mortality and Life Expectancy

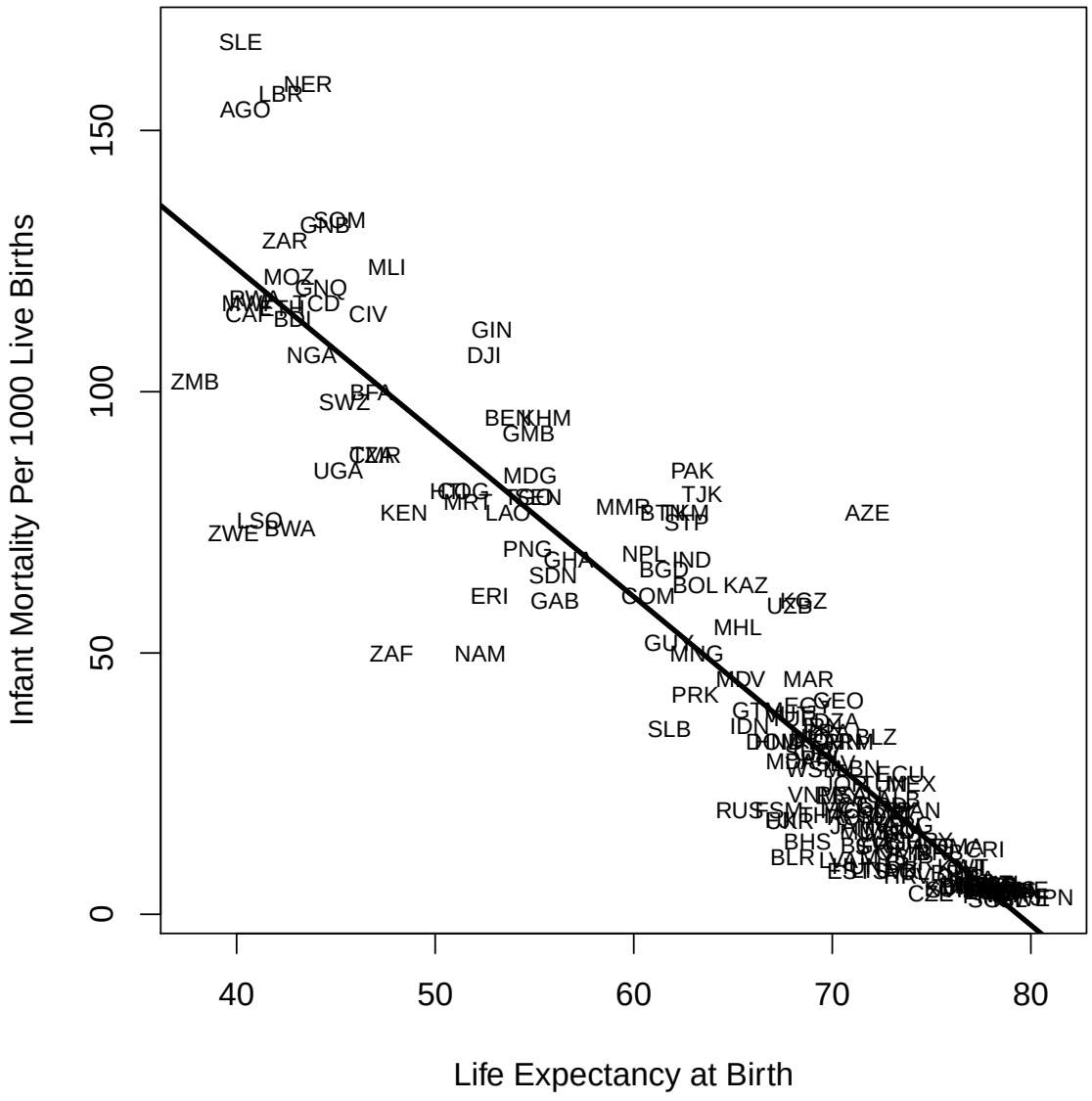


Figure 2: Infant Mortality and Life Expectancy: "Residuals"

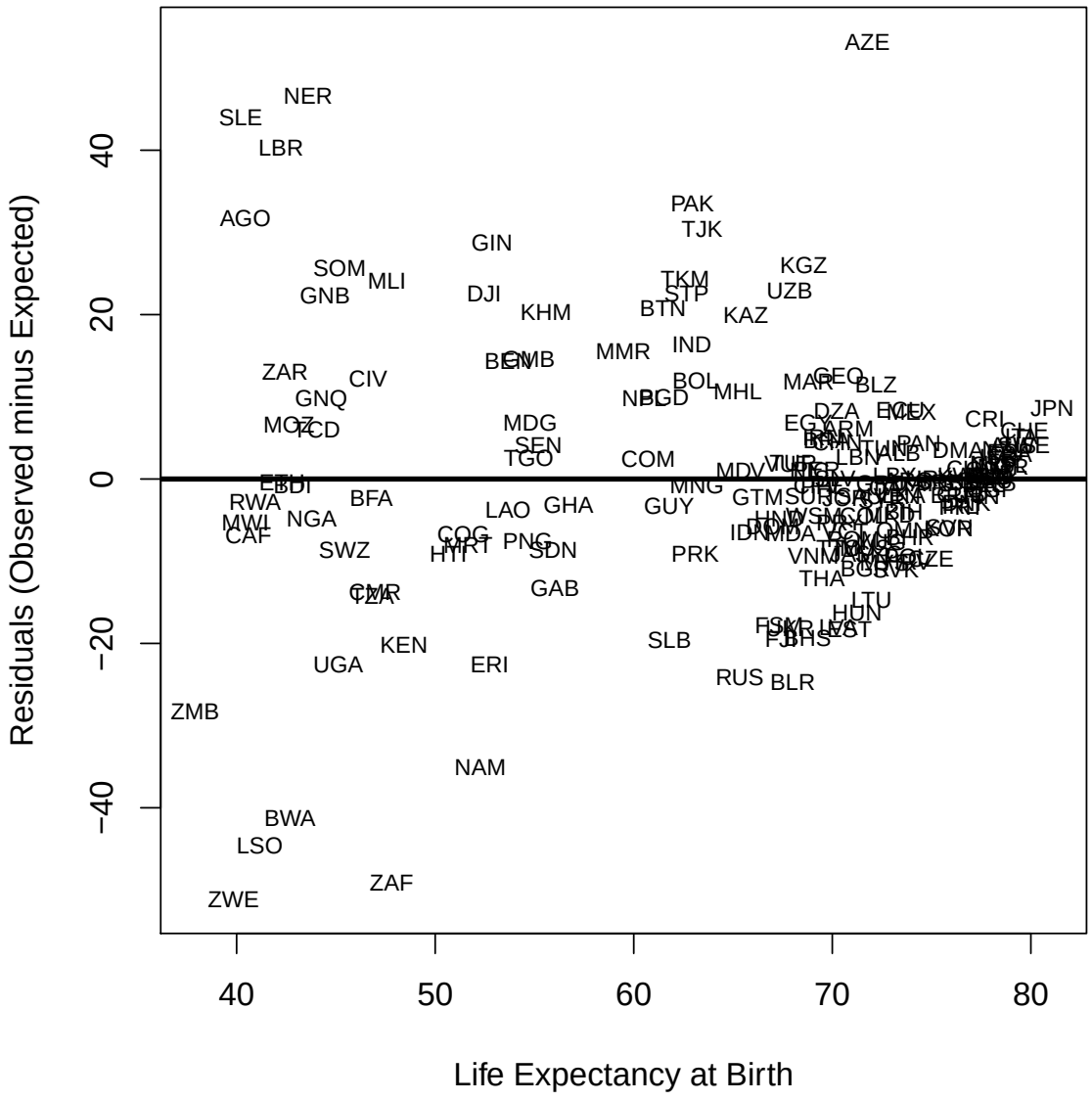


Figure 3: Infant Mortality and Fertility

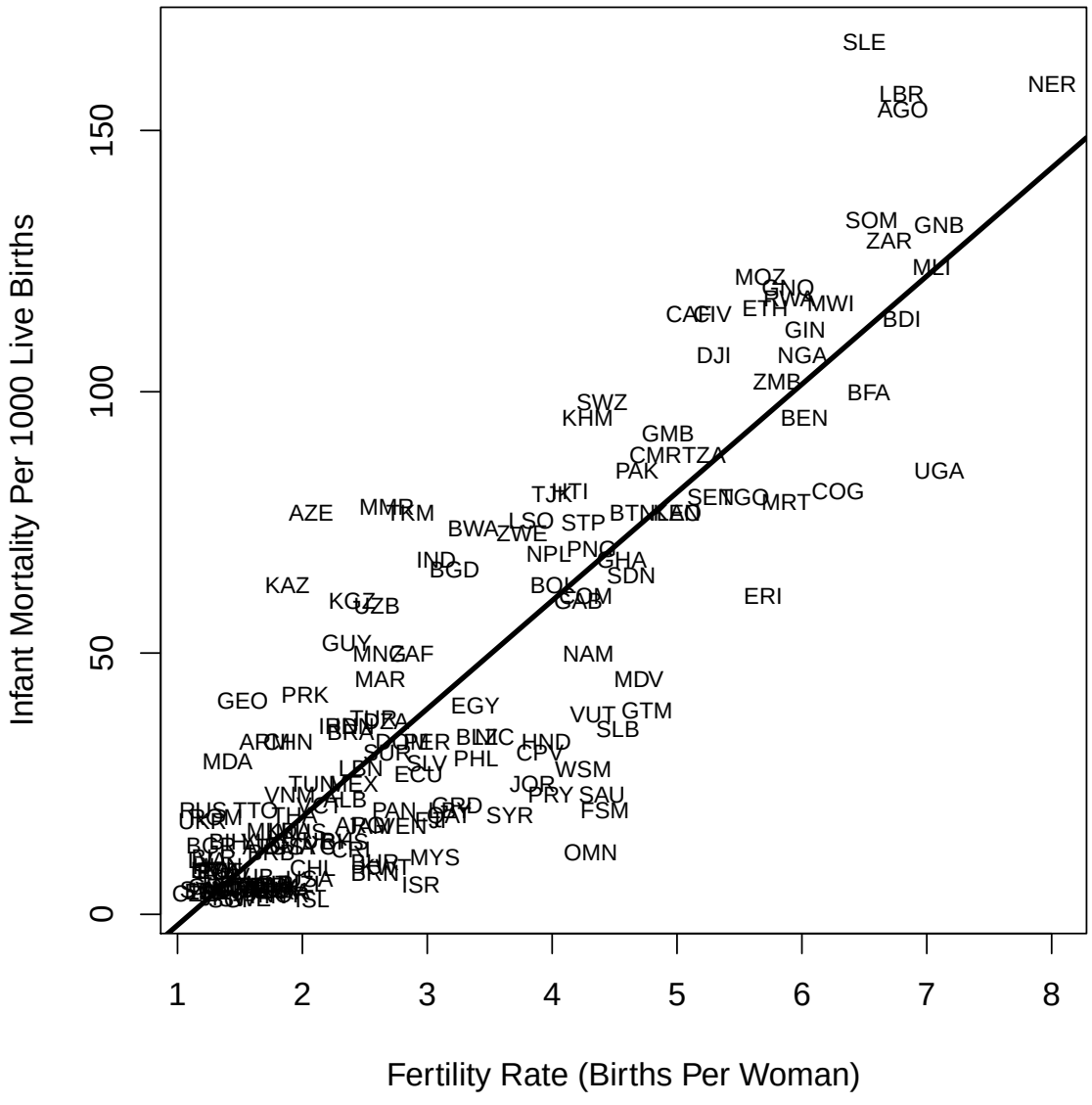


Figure 4: Infant Mortality and Wealth

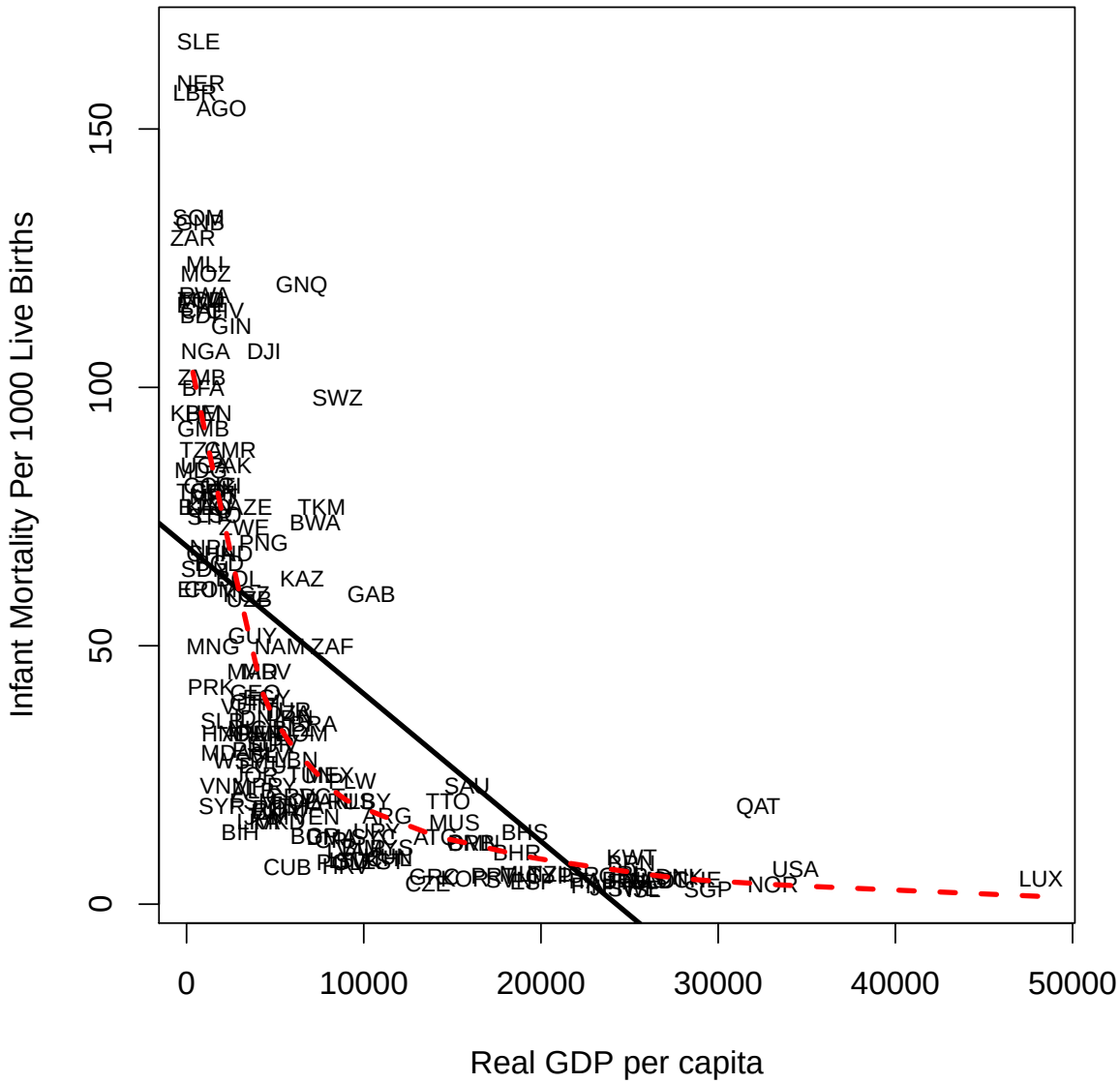


Figure 5: (Logged) Infant Mortality and (Logged) Wealth

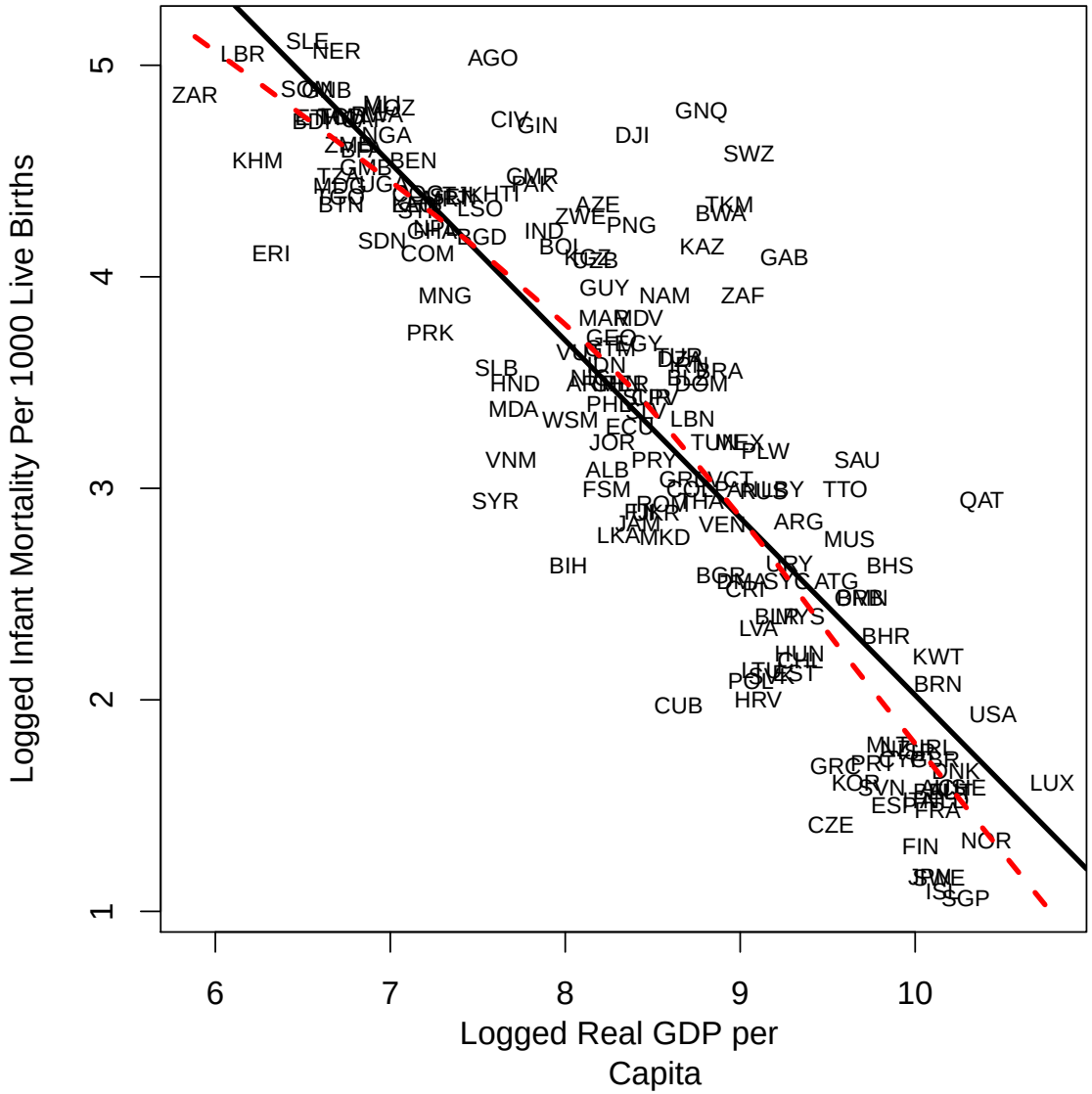


Figure 6: Infant Mortality and Democracy

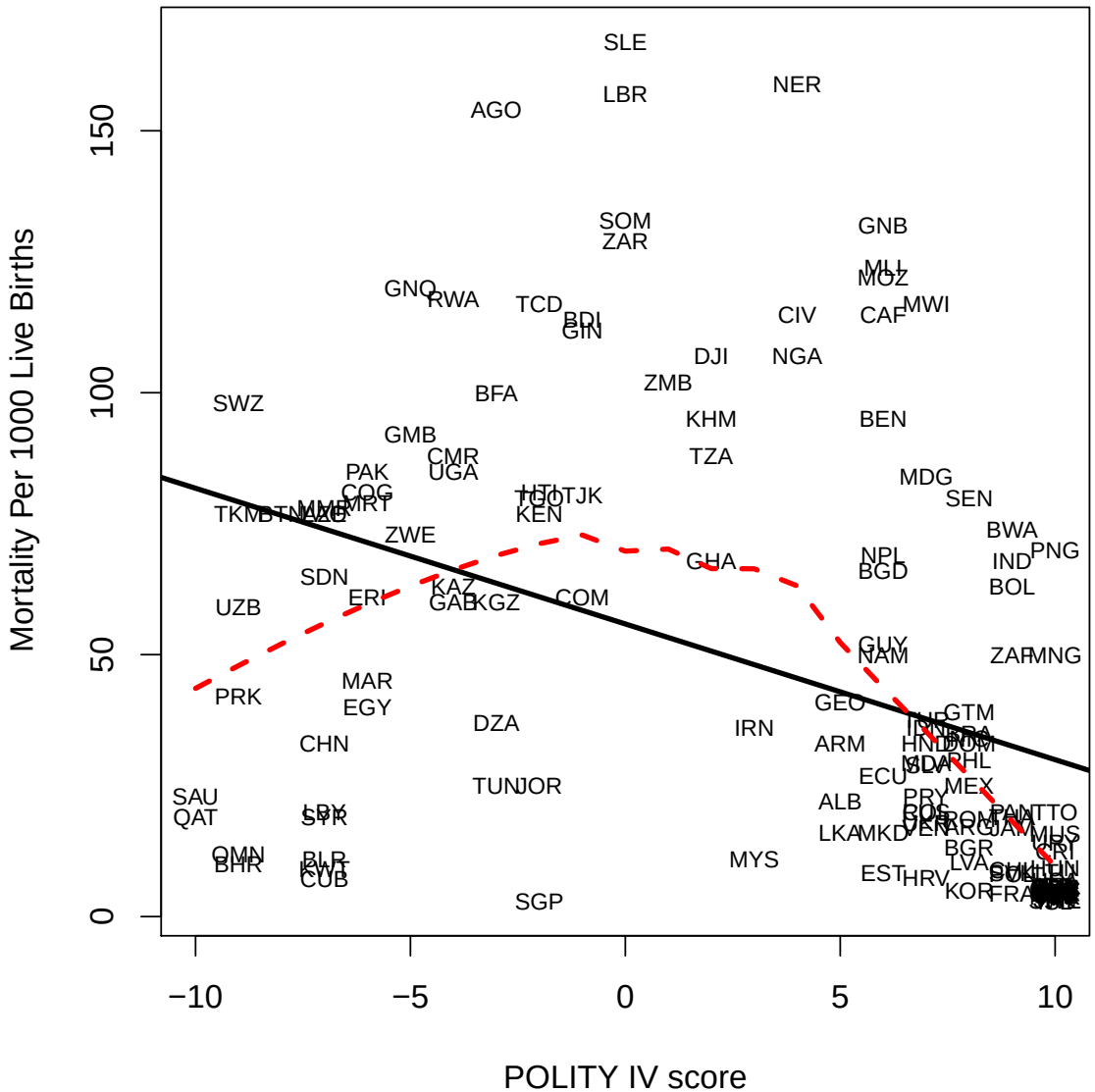


Figure 7: Infant Mortality, (Dichotomized) Wealth, and Democracy

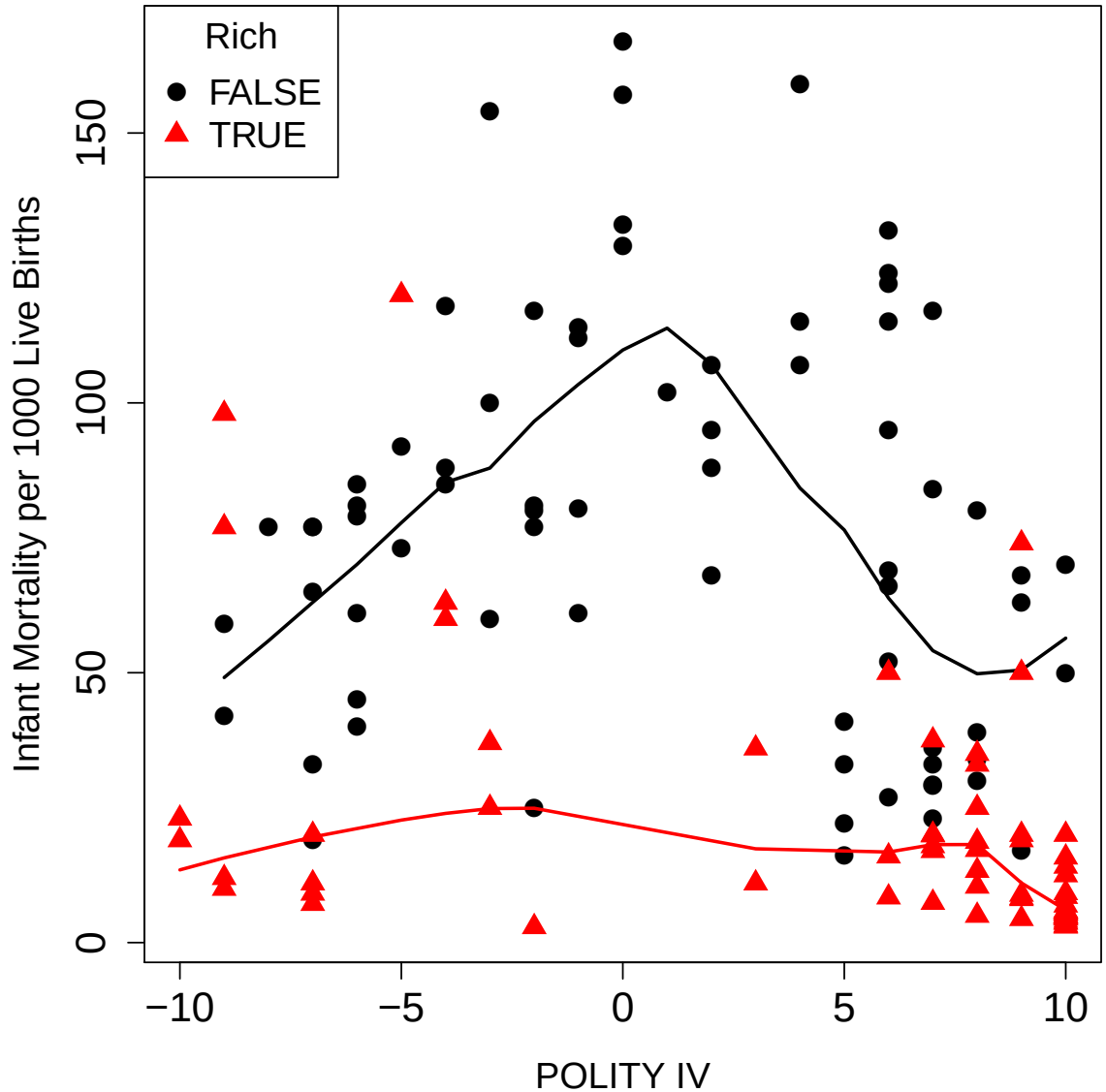


Figure 8: Measurement: National Health Indicators

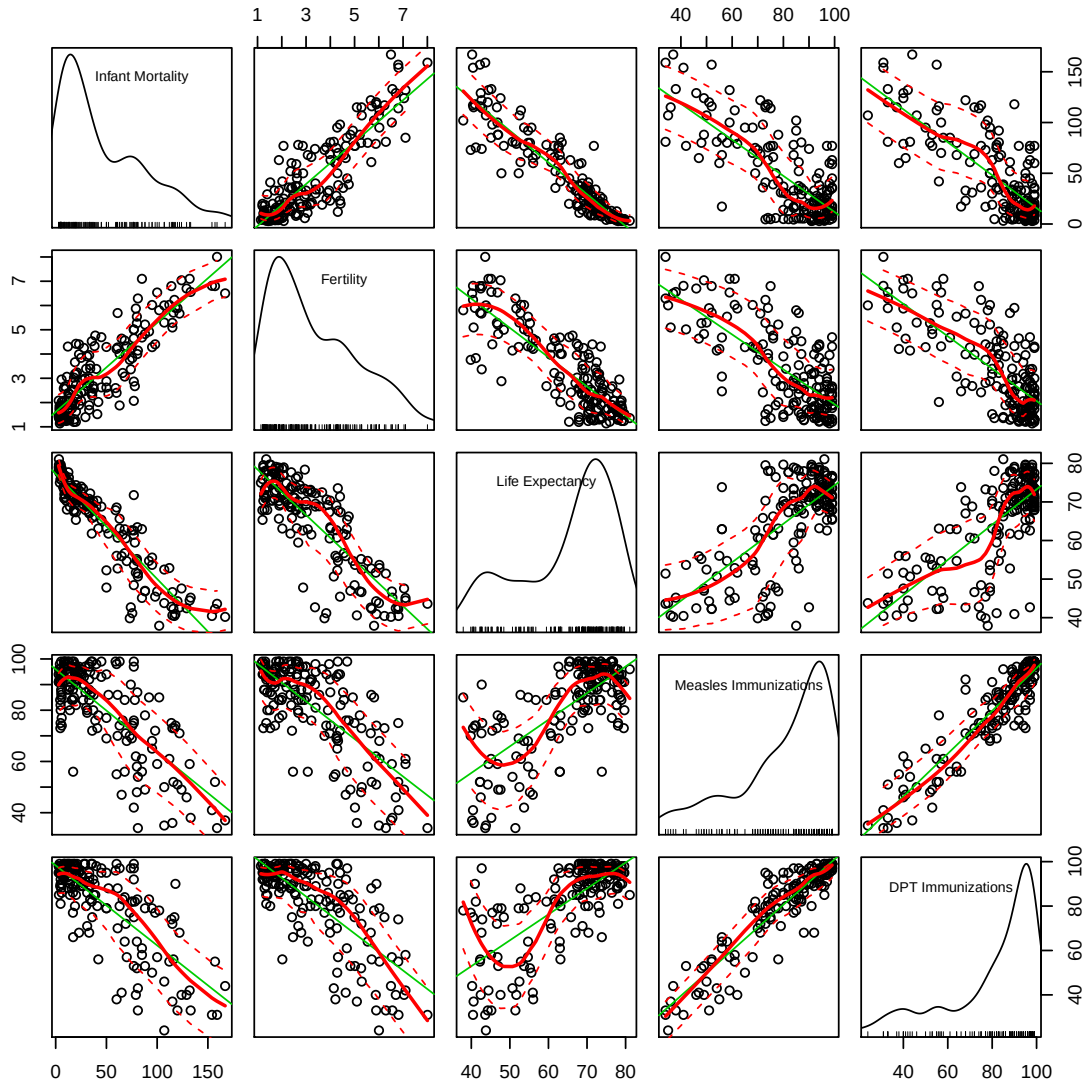
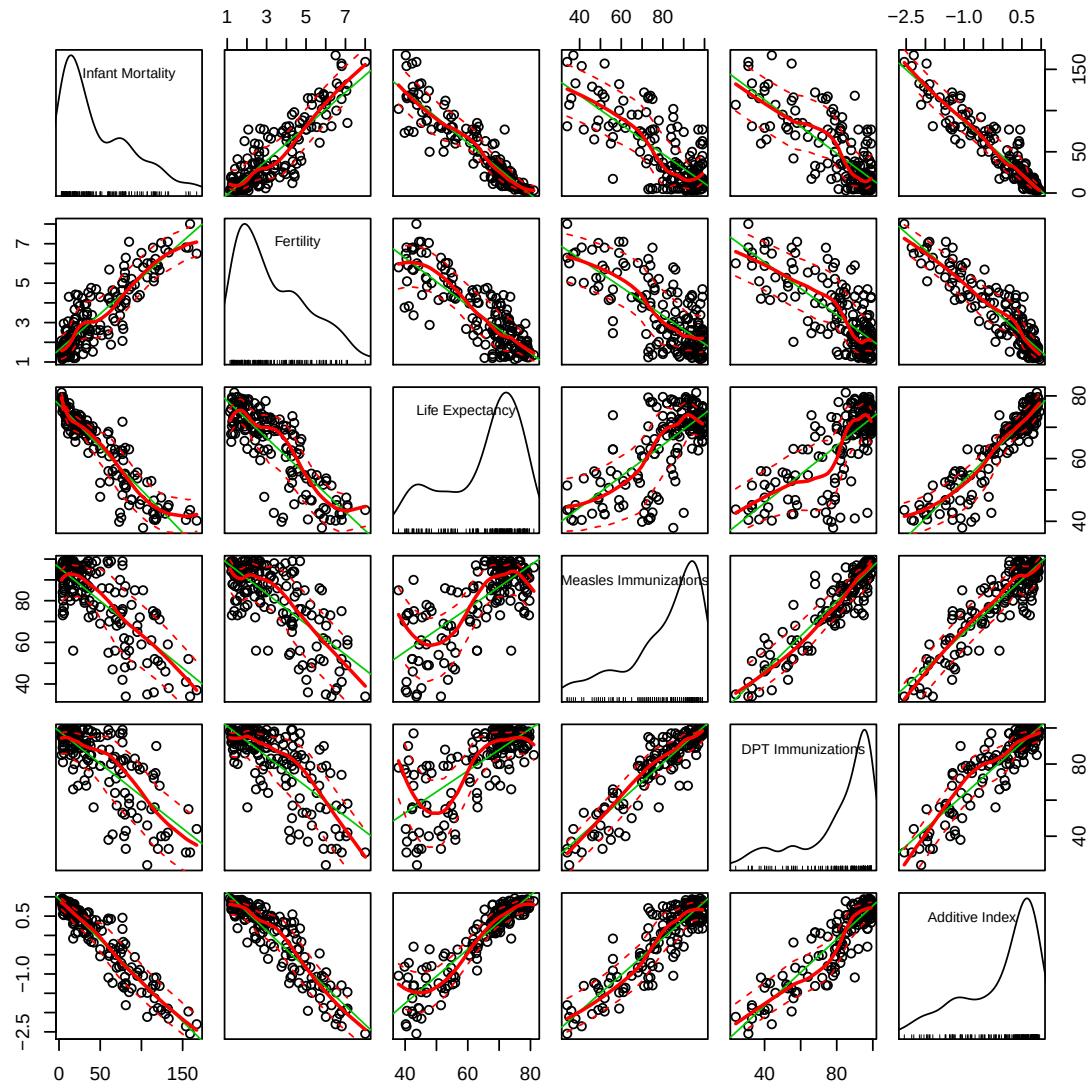


Figure 9: Measurement: National Health Indicators, Plus Additive Index



Wherefore Regression?

	Description	**Explanation**	**Prediction**
Task	Summarize data	Correlation/causation	Forecast OOS / future data
Emphasis	Data	Theory / Hypotheses	Outcomes
Focus	Univariate	Multivariate	Multivariate
Typical Application	Summarize / "reduce" data	Discuss marginal associations between predictors and an outcome of interest	Optimize out-of-sample predictive power / minimize prediction error