Lecture 14: Linear Regression III

COSC 480 Data Science, Spring 2017

Michael Hay

Logistics

- Project

- Course schedule

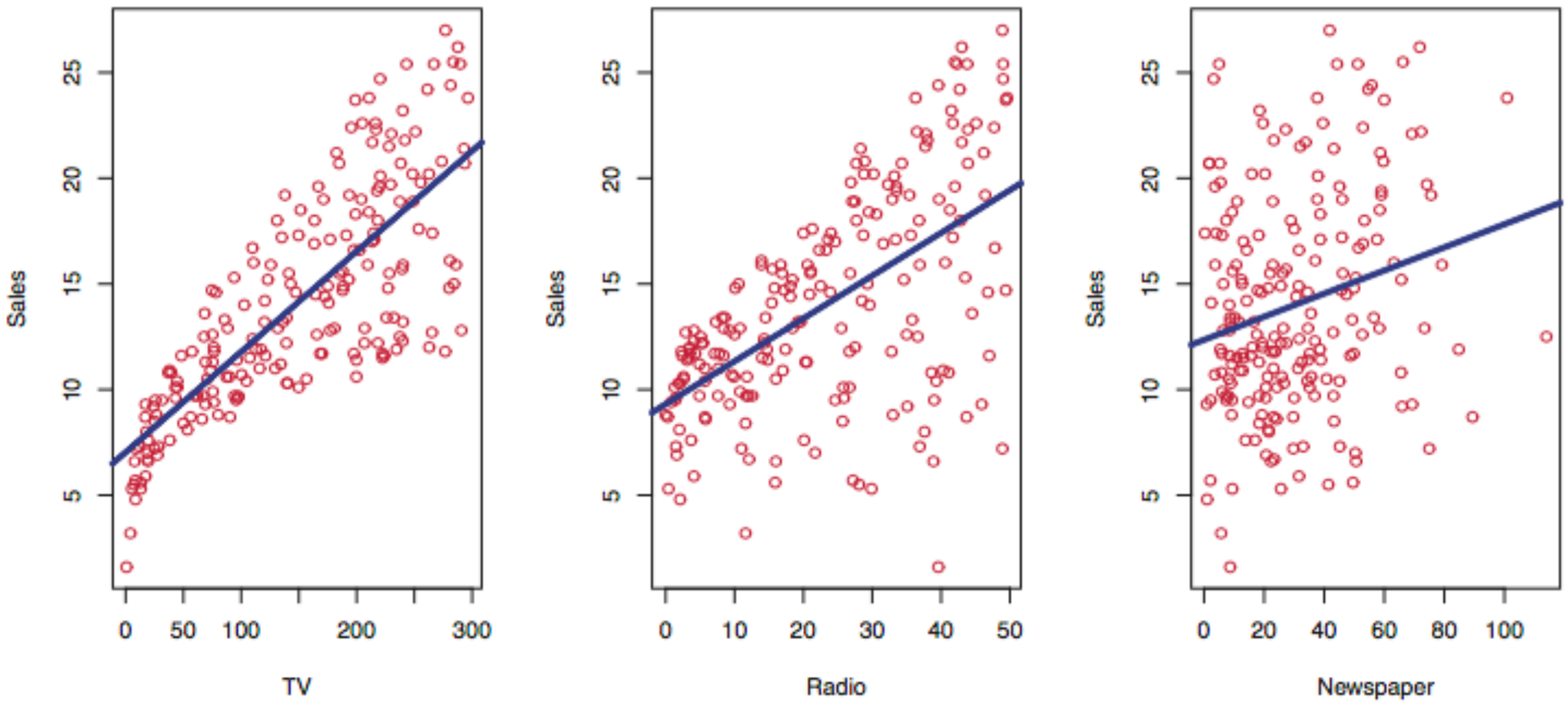

FIGURE 2.1. *The* `Advertising` *data set. The plot displays* `sales`*, in thousands of units, as a function of* `TV`*,* `radio`*, and* `newspaper` *budgets, in thousands of dollars, for* 200 *different markets. In each plot we show the simple least squares fit of* `sales` *to that variable, as described in Chapter 3. In other words, each blue line represents a simple model that can be used to predict* `sales` *using* `TV`*,* `radio`*, and* `newspaper`*, respectively.*

Credit: http://www-bcf.usc.edu/~gareth/ISL/

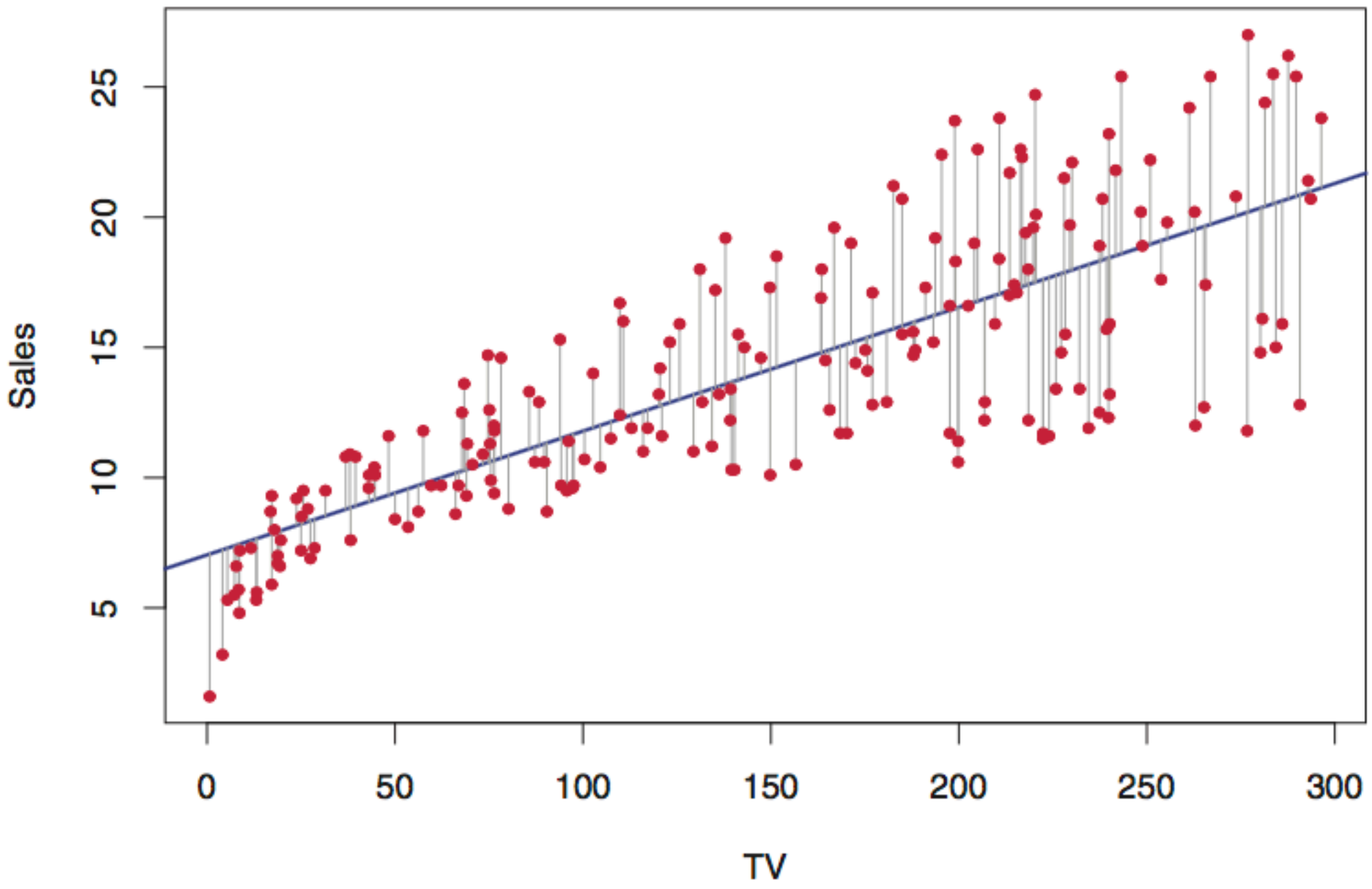

FIGURE 3.1. *For the* `Advertising` *data, the least squares fit for the regression of* `sales` *onto* `TV` *is shown. The fit is found by minimizing the sum of squared errors. Each grey line segment represents an error, and the fit makes a compromise by averaging their squares. In this case a linear fit captures the essence of the relationship, although it is somewhat deficient in the left of the plot.*

Credit: http://www-bcf.usc.edu/~gareth/ISL/

Goals for today

- Goodness of fit?
- What do coefficients mean?
- Which predictor variables are important?
  - Hypothesis testing
  - Bootstrap sampling
- Relaxing the additive assumption (if time; if not, see the reading Ch. 3 of James)