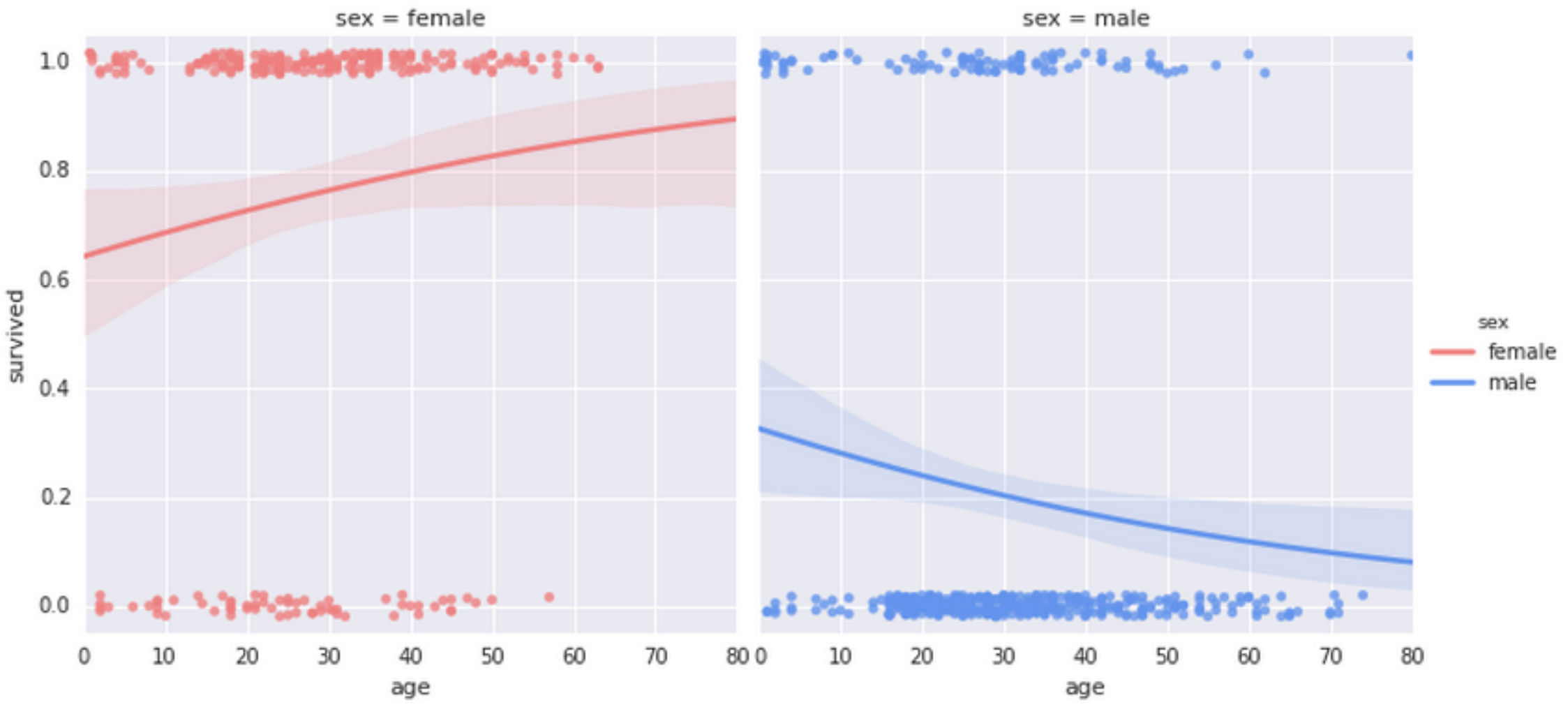
# Logistic regression example: Probability of surviving the sinking of the Titanic



# When to use logistic regression

Logistic regression is used when the dependent variable is discrete (often binary). The explanatory variables may be either continuous or discrete.

Examples:

- whether a gene is turned off (=0) or on (=1) as a function of levels of various proteins
- whether an individual is healthy (=0) or diseased (=1) as a function of various risk factors.
- whether an individual animal died (=0) or survived (=1) some selective event as a function of one or more moprhological traits.

# Mathematical formulation

Model the binary responses as:

$$P(Y = 1|X_1, \ldots, X_p) = f(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

So we're modeling the probability of the states as a function of a linear combination of the predictor variables.

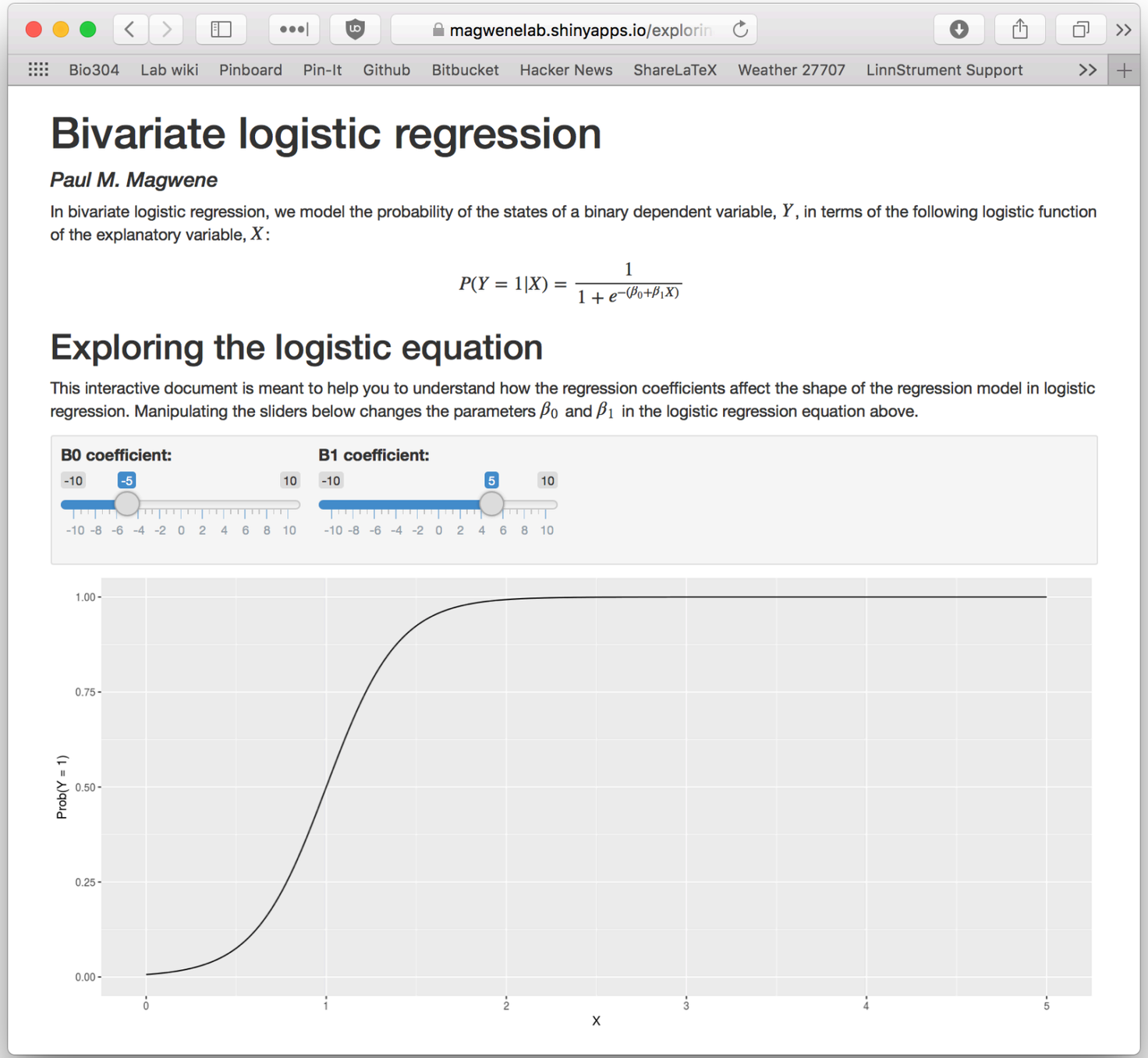
For logistic regression, we use the  logistic function for f:

$$f(z) = \frac{1}{1 + e^{-z}}$$

Substituting our linear combination of the predictor variables into the logistic function, for the bivariate case we get:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

# How does modifying the coefficients change the logistic curve?




**See the link to the web app in the lecture notes!**