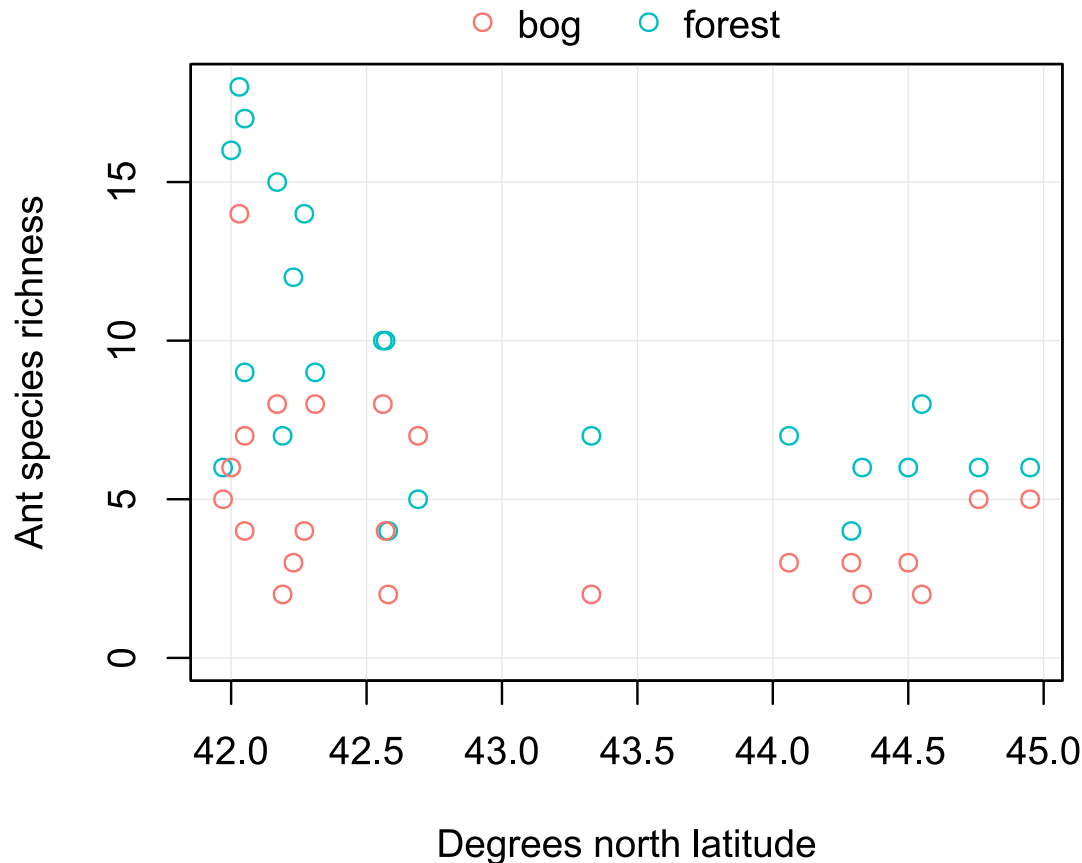# Today

- Richness-latitude design from scratch
- DGP when you have a dataset already
- Binomial DGP (birds dataset)

Design a study

- Richness declines with latitude
- Budget: adequate to sample 100 sites
- What did you do?
- What issues come up?

Data from an ant study

"We sampled 22 high-grade, undisturbed bogs and their surrounding forests in Vermont, Massachusetts, and Connecticut"

Make a statement about scope. Aspirational vs effective. Justify your statement.

Gotelli & Ellison (2002) Ecology 83: 1604-1609

Exponential Poisson model

richness   "is distributed as"

$$y_i \sim \text{Poisson}(\mu_i)$$ Stochasticity

$$\mu_i = a\, e^{b\, x_i}$$ Bog   Deterministic skeleton

$$\mu_i = c\, e^{d\, x_i}$$ Forest

mean richness   latitude

Data story reads from bottom to top

Exponential Poisson model

Generalized linear model form

$$y_i \sim \text{Poisson}\left(\mu_i\right)$$

$$\ln\left(\mu_i\right) = \beta_0 + \beta_1 x_i \qquad \text{Bog}$$

$$\ln\left(\mu_i\right) = \beta_2 + \beta_3 x_i \qquad \text{Forest}$$

Log link function

Inverse link

$$\mu_i = e^{\beta_0} e^{\beta_1 x_i}$$

Poisson-lognormal model

Extra variation among plots

$$y_i \sim \text{Poisson}\left(\mu_i\right)$$

Deterministic

$$\ln\left(\mu_i\right) = \beta_0 + \beta_1 x_i + e_i \quad \text{Bog}$$

$$\ln\left(\mu_i\right) = \boxed{\beta_2 + \beta_3 x_i} + e_i \quad \text{Forest}$$

$$e_i \sim \text{Normal}\left(0, \sigma_e\right)$$

Stochastic

Generalized linear mixed model GLMM

Code

Ants: sketch of study design

Poisson-lognormal model

Extra variation among sites and plots

$$y_i \sim \text{Poisson}(\mu_i)$$

sites   plots

$$\ln(\mu_i) = \beta_0 + \beta_1 x_i + s_{j[i]} + e_i \qquad \text{Bog}$$

$$\ln(\mu_i) = \beta_2 + \beta_3 x_i + s_{j[i]} + e_i \qquad \text{Forest}$$

$$e_i \sim \text{Normal}(0, \sigma_e)$$

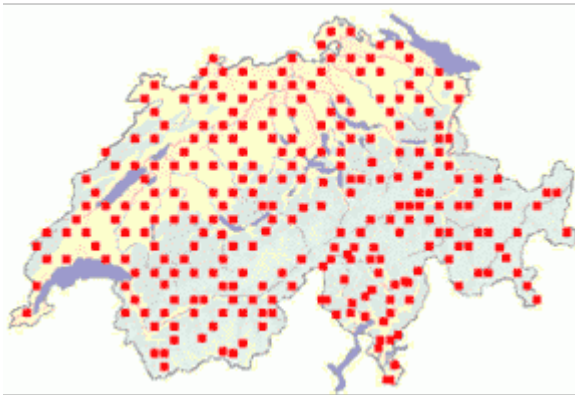$$s_j \sim \text{Normal}(0, \sigma_s)$$

GLMM

Spatial autocorrelation

- Plots within sites are correlated

- Their deviations are correlated. All the plots from site j share the same deviation s_j (plus a deviation, e_i, of their own).

Bernoulli/Binomial DGP

- Swiss breeding bird survey (www.vogelwarte.ch)
- Skilled observers, 1 km^2 cells
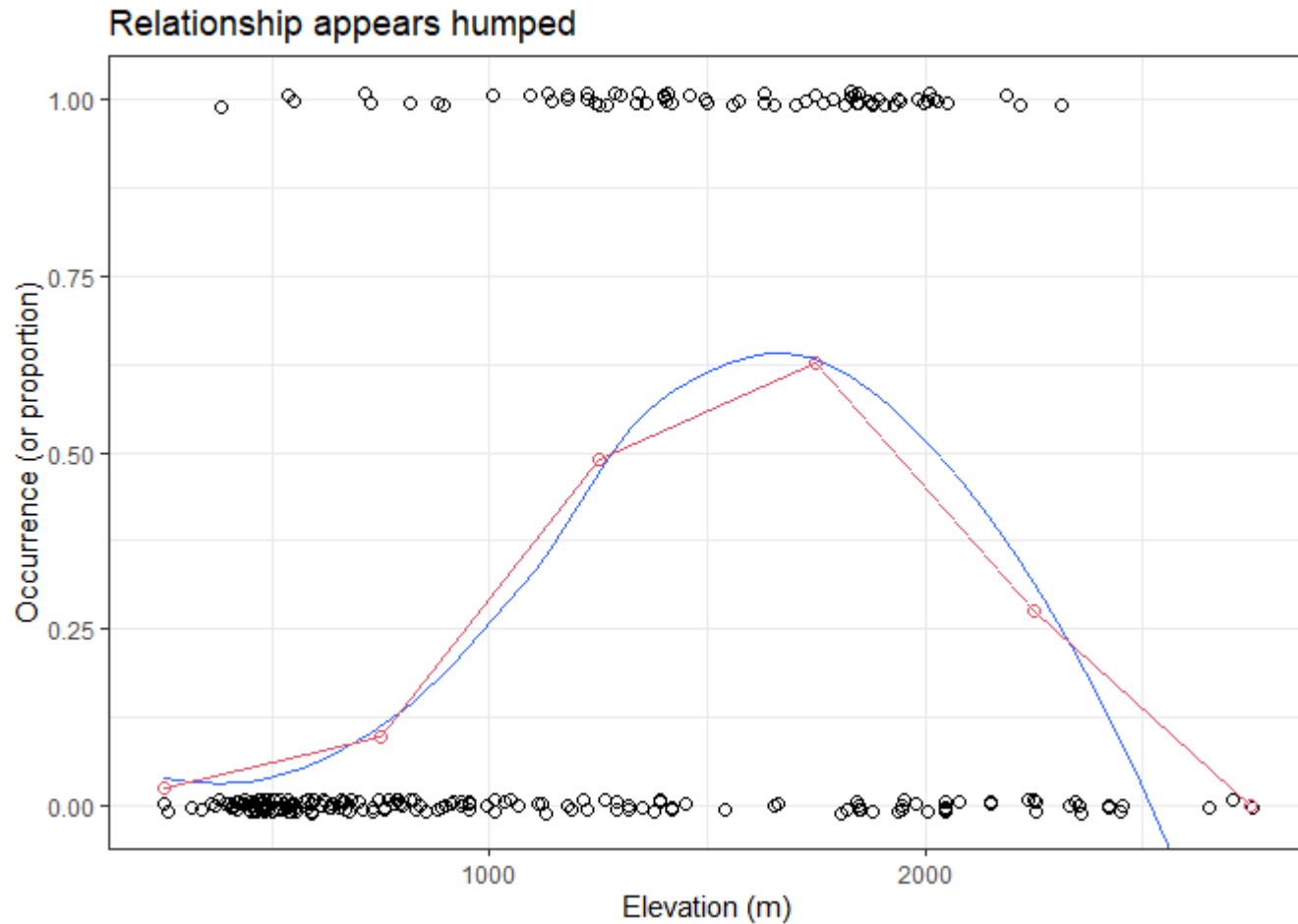

Switzerland; showing survey locations

A 1 km^2 survey cell

- Willow tit territory presence-absence in relation to altitude

see swissbbs_dgp.R

Royle JA, Dorazio RM (2008) Hierarchical Modeling and Inference in Ecology. Academic Press, Oxford. p 87.

Exploratory Data Analysis

Relationship appears humped

Smoothing and binning are useful to visualize binary data

Logit Binomial model

Generalized linear model form

$$y_i \sim \text{Binomial}(p_i, n=1)$$

occurrence

number of trials = 1

Logit function

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_i$$
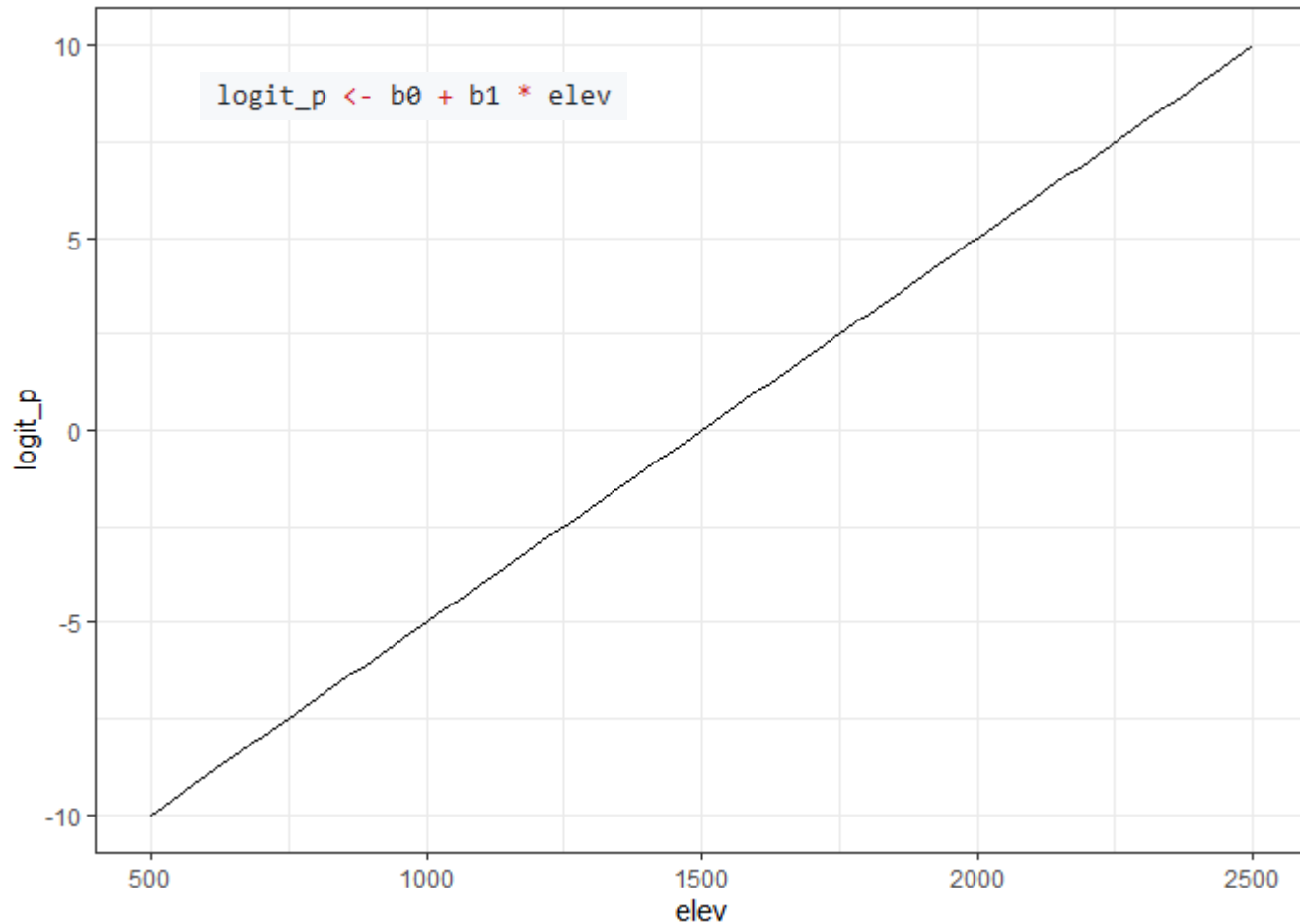
Logit link function

elevation

Inverse link

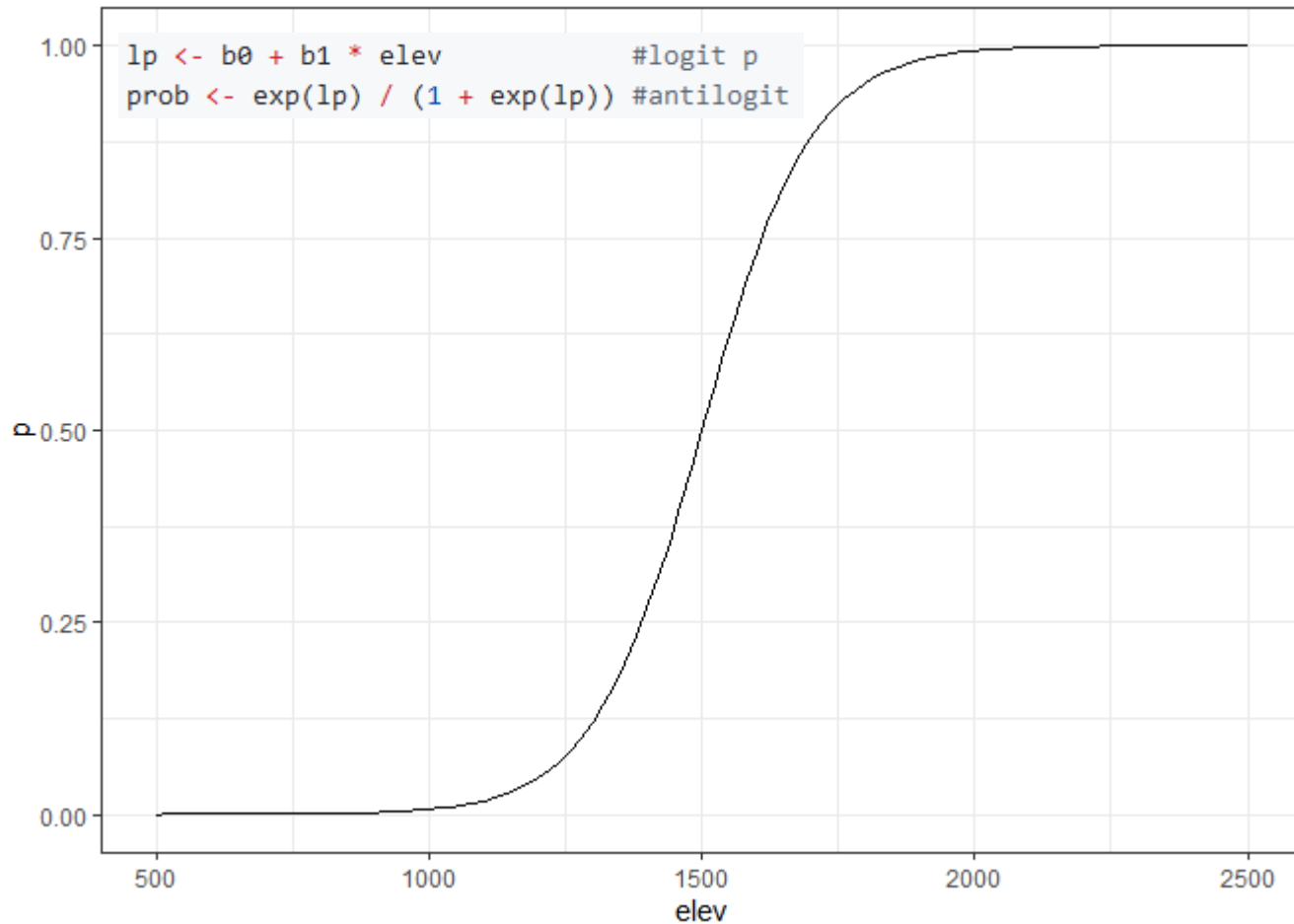$$p = \frac{\exp(\text{logit}(p))}{1 + \exp(\text{logit}(p))}$$

logistic function

Logit Binomial model

Logit scale is linear –INF to +INF


```
logit_p <- b0 + b1 * elev
```

Logit Binomial model

Probability (antilogit) scale is sigmoid 0 to 1, monotonic


```
lp <- b0 + b1 * elev          #logit p
prob <- exp(lp) / (1 + exp(lp)) #antilogit
```
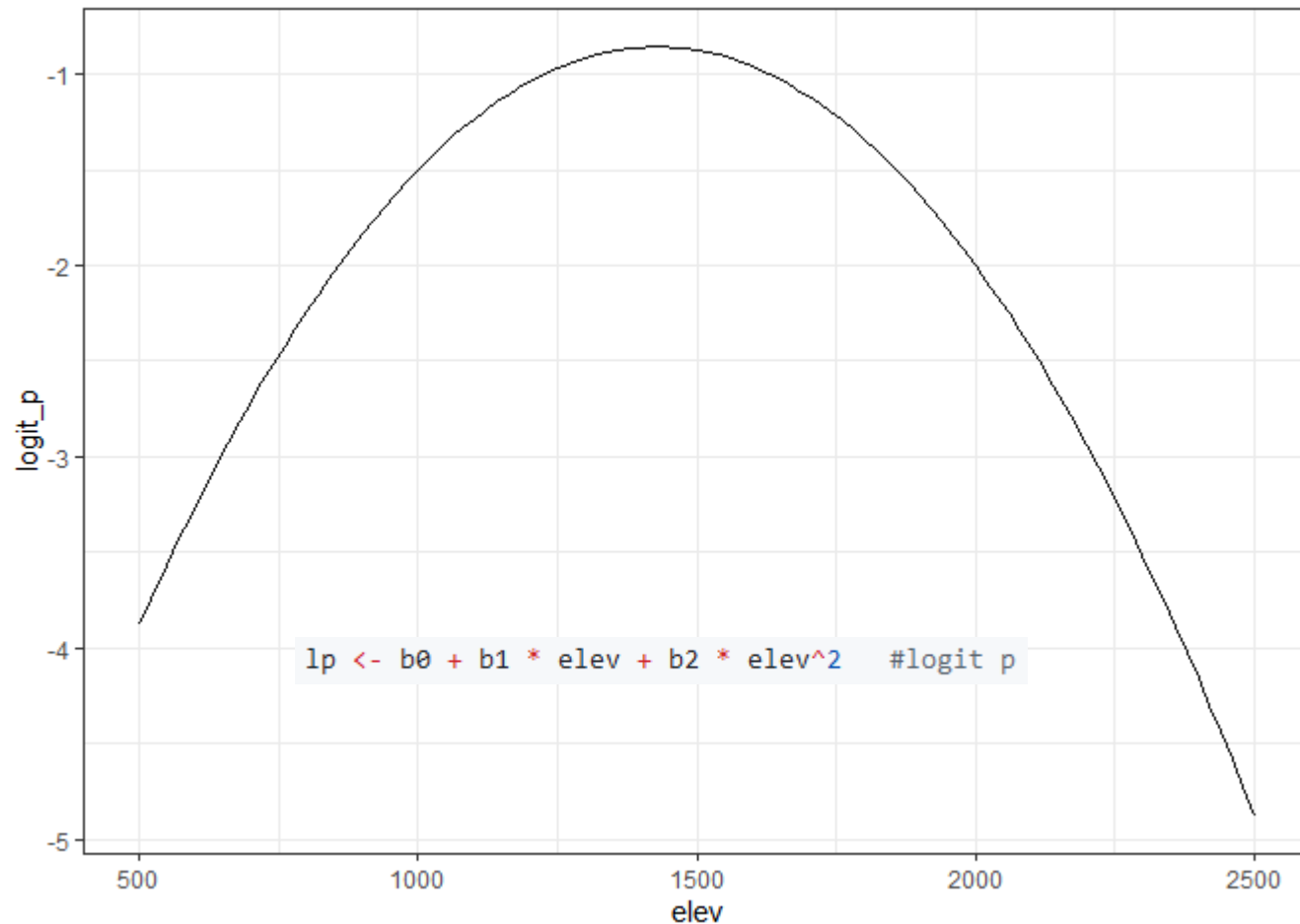
The humped model

Logit function

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x + \beta_2 x^2$$

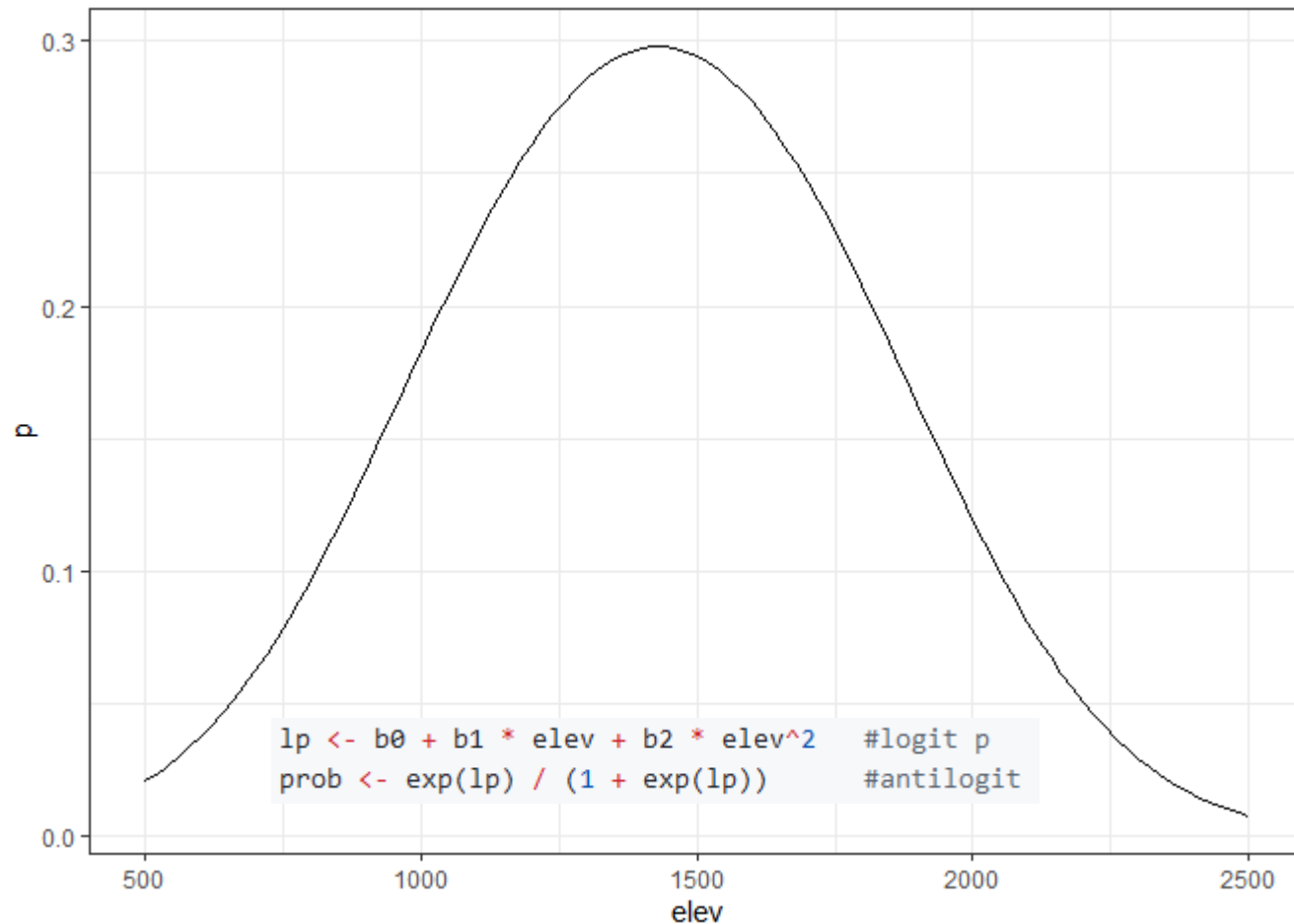Quadratic function allows for hump

Elevation

The humped model

Logit scale


```
lp <- b0 + b1 * elev + b2 * elev^2    #logit p
```

The humped model

Probability (antilogit) scale

```
lp <- b0 + b1 * elev + b2 * elev^2      #logit p
prob <- exp(lp) / (1 + exp(lp))         #antilogit
```

Code the DGP