

# ZIKA VIRUS DISEASE

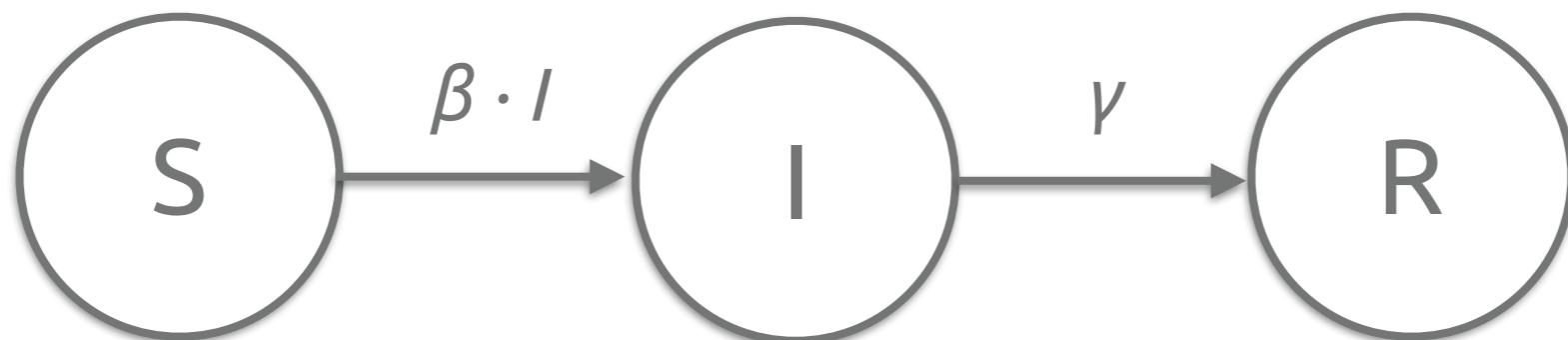
---

*Variation in the Basic Reproduction Number  
for the 2015/2016 Colombia Outbreak*

*William Rittasse, S. M. Z. Rizvi, & Stephen Thomas*

# MODELING DISEASE OUTBREAKS

---

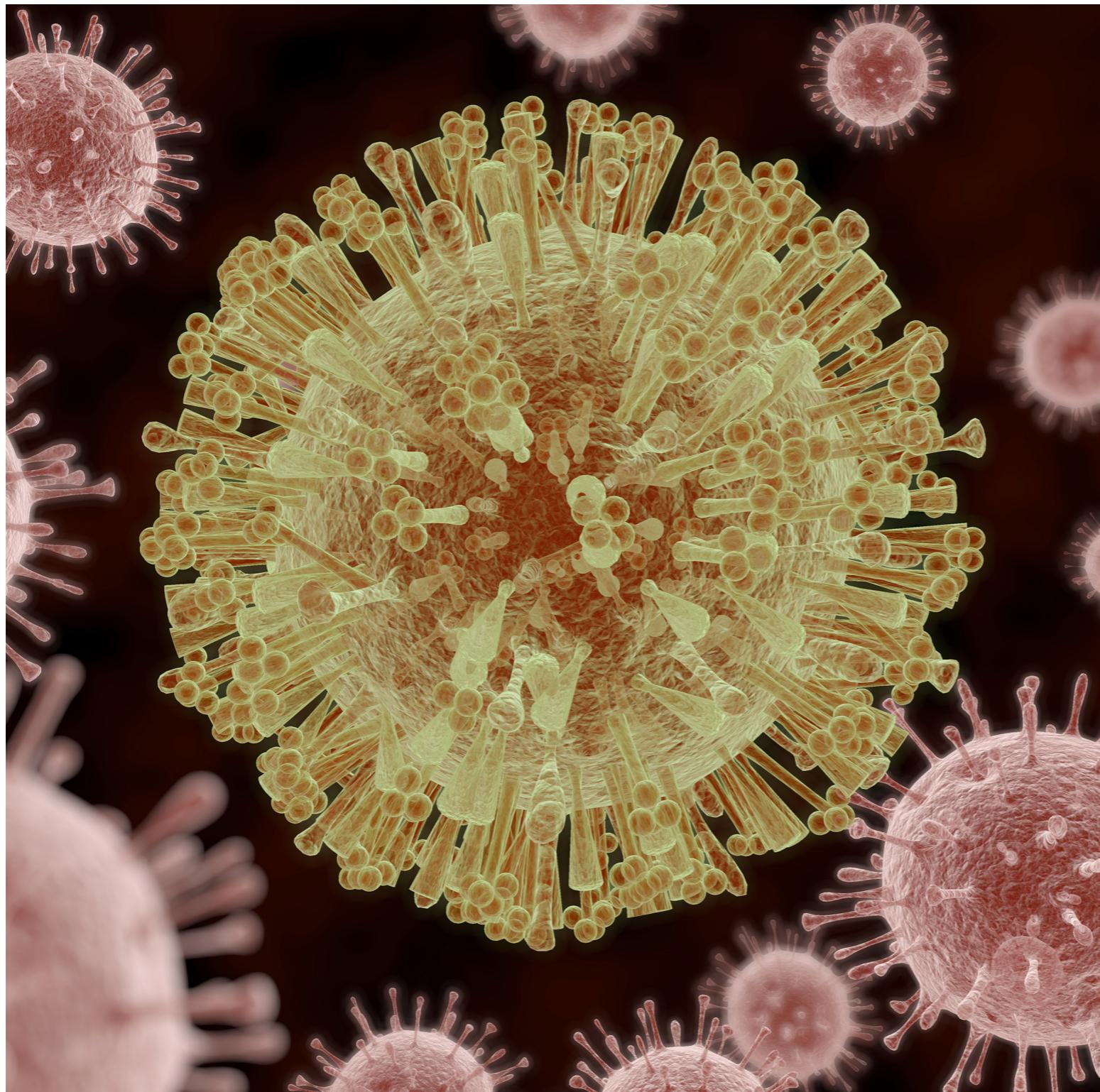


$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **115** (772): 700

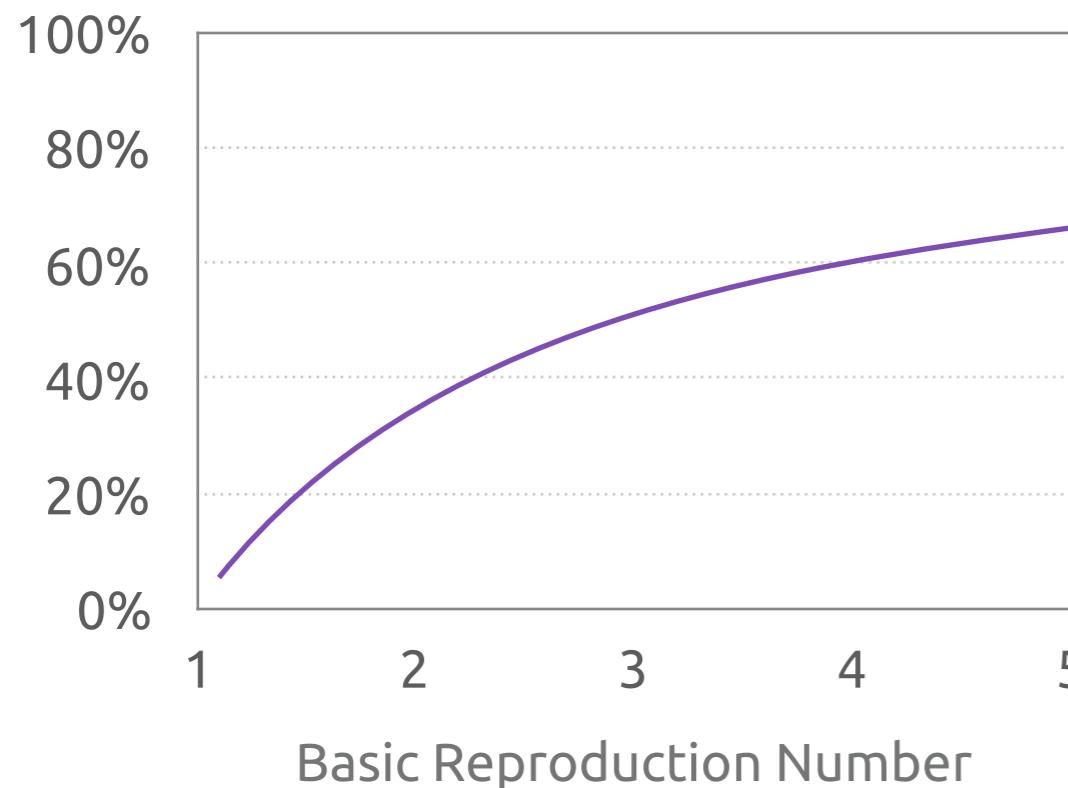


BASIC REPRODUCTION NUMBER: e.g.  $R_0 = \beta/\gamma$

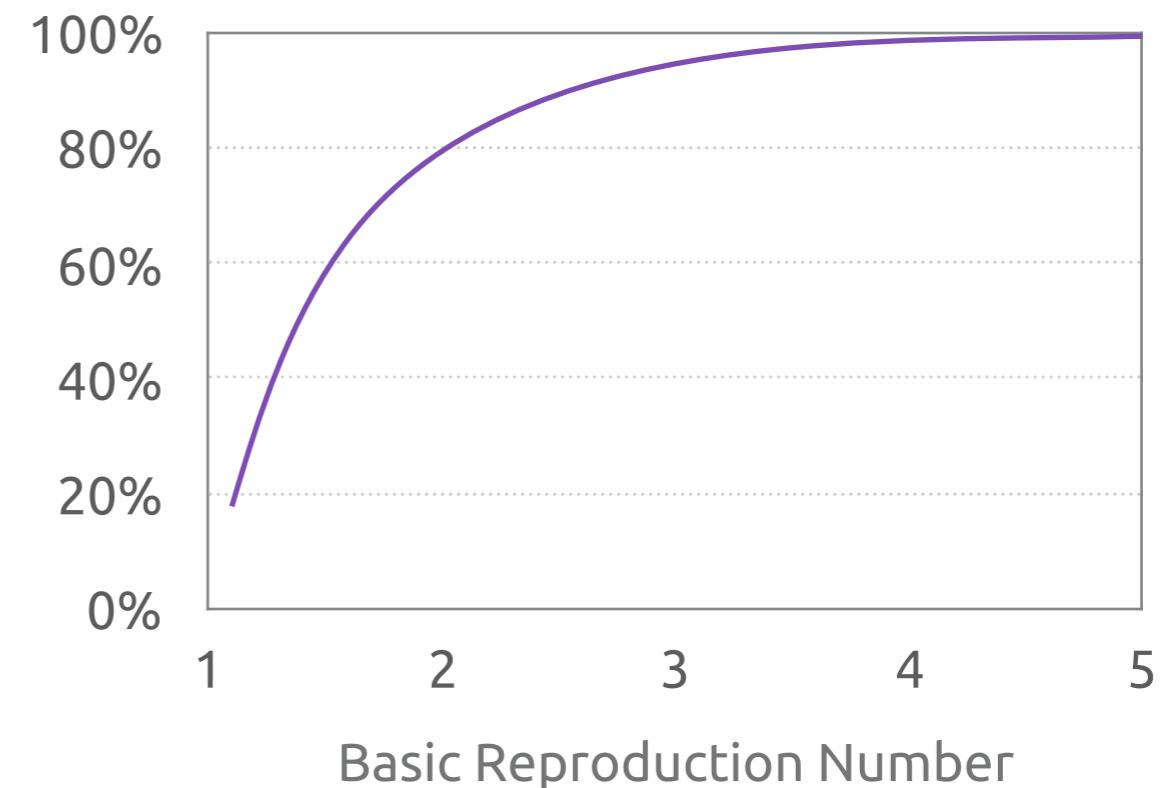
# $R_0$ DESCRIBES KEY CHARACTERISTICS OF AN EPIDEMIC

---

Peak Infection Rate



Final Size of Epidemic



Critical Vaccination Threshold:

$$P_c = \frac{1}{\varepsilon} \left( 1 - \frac{1}{R_0} \right)$$

# DATA: COLOMBIA NATIONAL INSTITUTE OF HEALTH

**BES**  
BOLETÍN EPIDEMIOLÓGICO SEMANAL

Contenido

Portada .....	1
Cumplimiento en la notificación .....	2
Inmunoprevenibles .....	5
Enfermedades transmitidas por alimentos .....	16
Enfermedades transmitidas por vectores .....	20
Infecciones de transmisión sexual .....	38
Micobacterias .....	42
Zoonosis .....	44
Factores de Riesgo Ambiental .....	50

Dirección de Vigilancia y Análisis del Riesgo en Salud Pública  
INSTITUTO NACIONAL DE SALUD  
TODOS POR UN NUEVO PAÍS  
PAZ EQUIDAD EDUCACIÓN

Semana epidemiológica número 01 de 2016 (03 ene. al 09 ene.)

**"FIN DEL BROTE MÁS RECENTE DE ÉBOLA EN LIBERIA, PERO NO PUEDEN DESCARTARSE NUEVOS BROTES".**

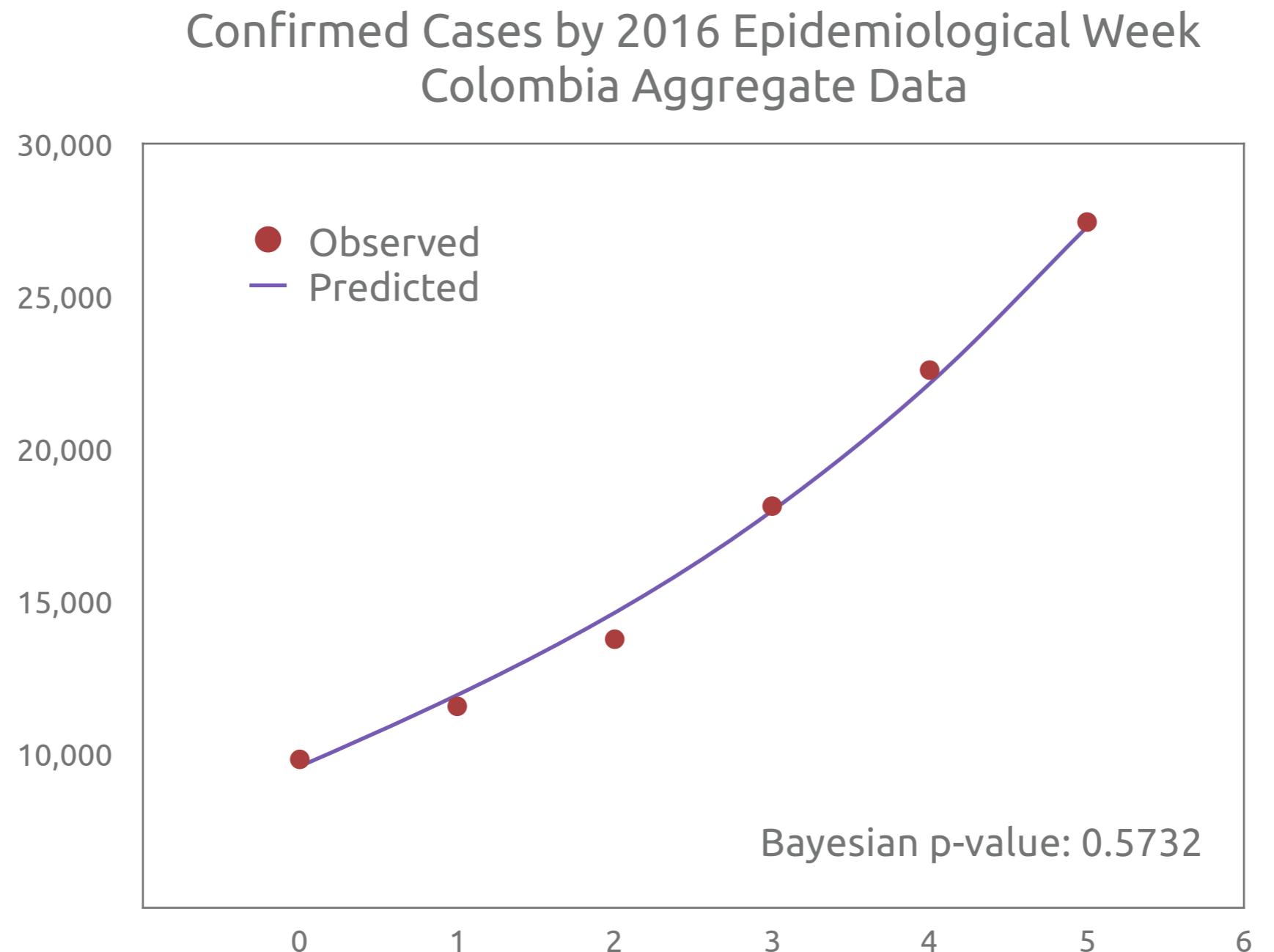
14 de enero de 2016 -- La OMS declara hoy el fin del brote más reciente de enfermedad por el virus del Ébola en Liberia. No obstante, la Organización advierte de que la tarea no se ha

<http://www.ins.gov.co/Noticias/ZIKA/Forms/AllItems.aspx>

<https://github.com/cdcepi/zika>

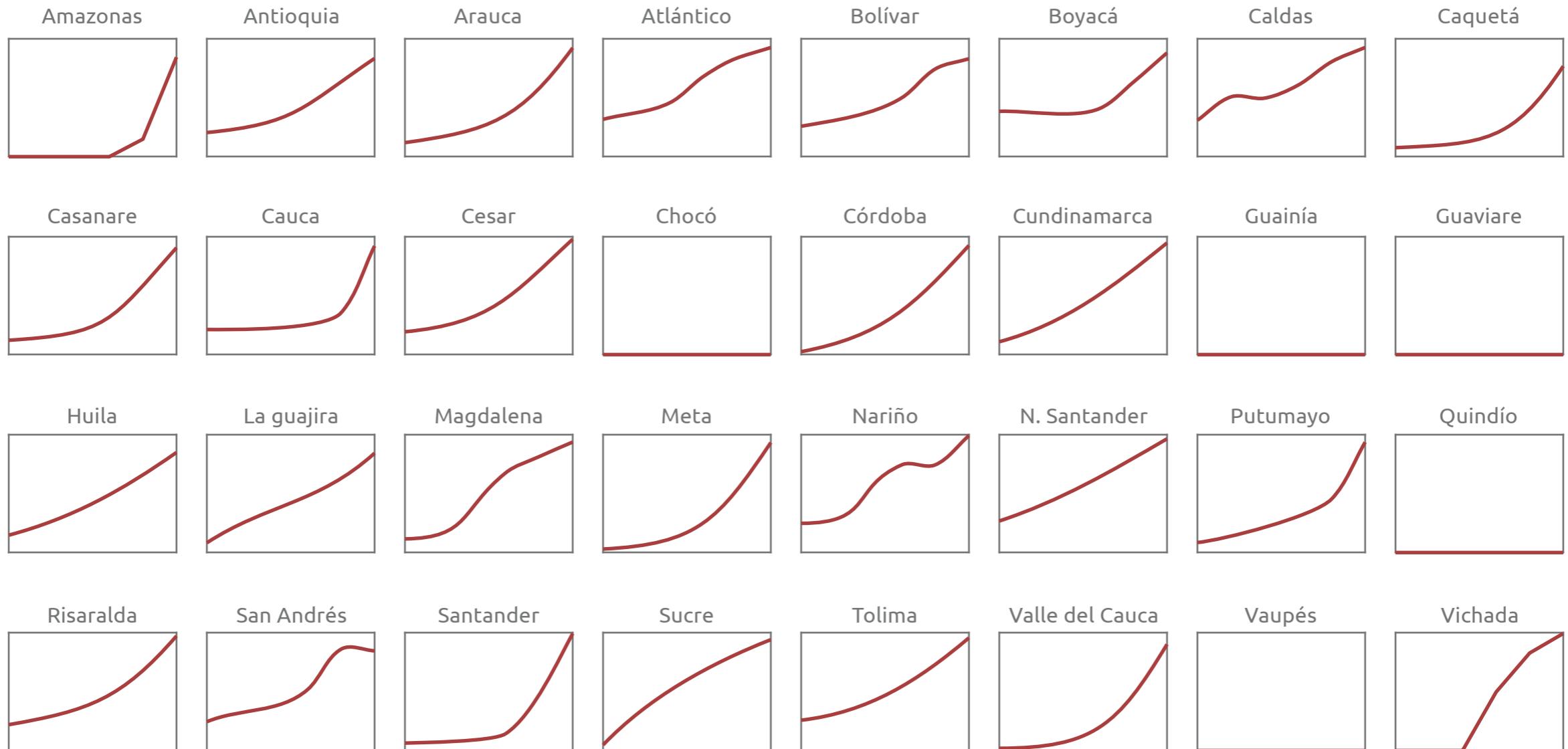
# VALIDATING THE DATA USING AGGREGATE CASES

---



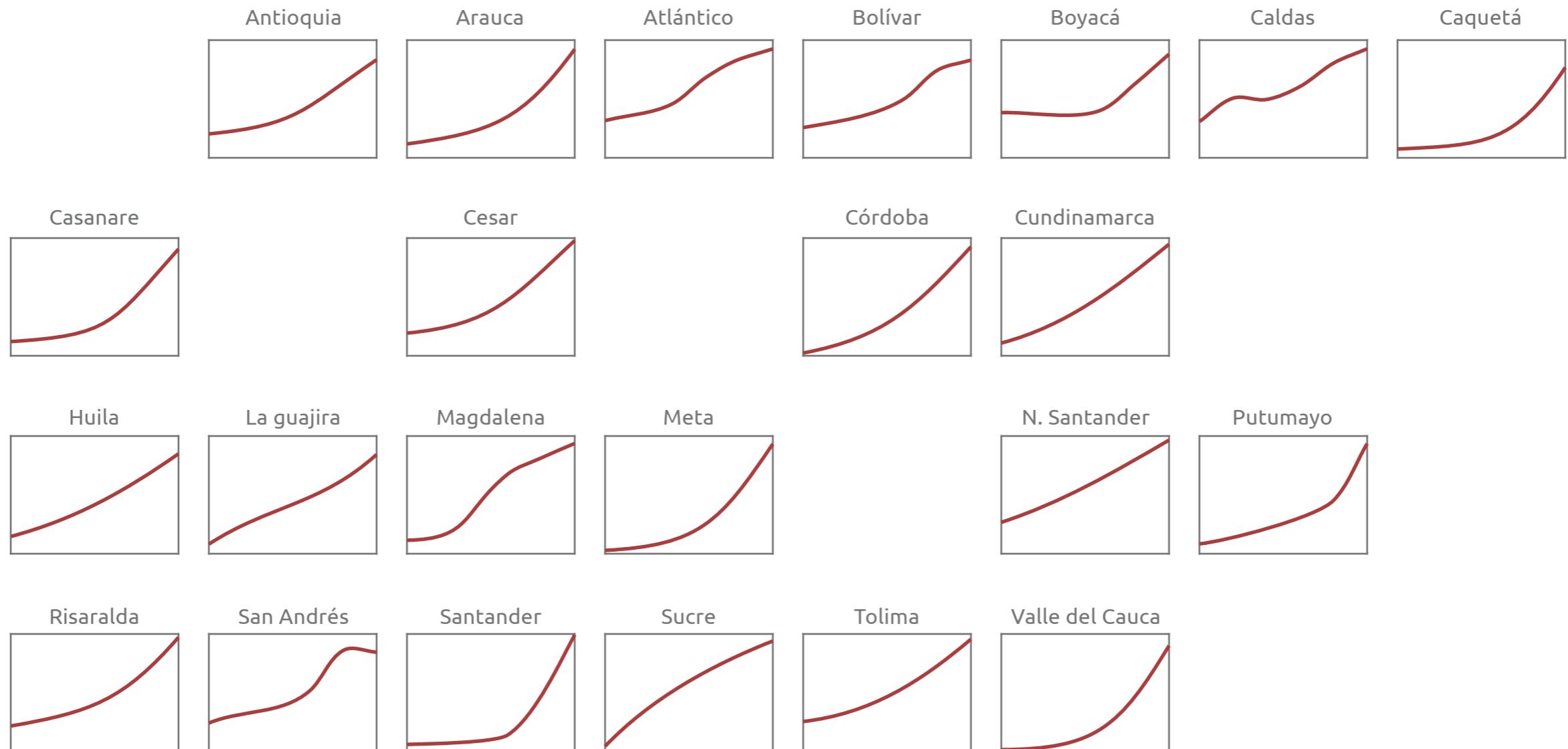
# CONFIRMED CASES BY DEPARTMENT

---



# CONFIRMED CASES BY DEPARTMENT (> 50 CASES)

---



# $R_0$ ESTIMATION MODEL

```
# This is a JAGS model for estimating the basic reproduction
# number ( $R_0$ ) for Colombia departments. It relies on a
# hierarchical linear regression with two levels of
# hierarchy. The input data consists of a set of
# observations where each observation includes an
# independent variable ( $x$ ) representing the week number, a
# dependent variable ( $y$ ) representing the cumulative
# confirmed cases, and an integer-valued department
# identification.
#
# The model assumes that the dependent variable  $y$  is
# exponentially related to the independent variable, and
# observations within a department are more closely related
# than observations between departments. In particular, it
# assumes that the slope parameters for individual groups
# are normally distributed around an overall slope.
# Intercept parameters, however, are assumed independent for
# each group and are derived from the data.
#
# The model reports goodness of fit using Bayesian p-value.
#
# Observation data is supplied in three vectors:
#   dept[] - department id
#   x[]    - week number
#   y[]    - cumulative confirmed cases
#
# In addition, a vector of observed intercepts for
# each department is supplied in the `intercept[]` vector.

data {
  # Transform Inputs
  #
  # Logarithmic transformation to allow linear regression
  ln_y <- log(y)
}

model {
  # Prior Probabilities
  #
  # Express all priors as reference (non-informative)
  # priors. The parameters for the model are:
  #
  #   beta_mu      - the mean for the aggregate slope
  #                   (all depts)
  #   beta_sigma   - the standard deviation for the
  #                   aggregate slope (all depts)
  #   sigma        - the standard deviation for the
  #                   residuals (assumed to be the same
  #                   for all depts)

  beta_mu ~ dnorm(0, 0.0001)
  beta_sigma ~ dunif(0.001, 1000)
  sigma ~ dunif(0.001, 1000)

  # The Likelihood
  #
  # Model the dependent variable `y` as Normal with mean
  # `mu` and standard deviation `sigma`. Model the mean
  # `mu` as a linear function of the independent variable
  # `x`. The parameters of that linear relationship vary
  # based on the department, and are:
  #
  #   alpha[dept[i]] - intercept for department
  #   beta[dept[i]]  - slope for the department
  #
  # The standard deviation `sigma` is assumed to be the
  # same for all observations, regardless of dept.
  #
  # Apply the model to all observations.
  for (i in 1:length(y)) {
    ln_y[i] ~ dnorm(mu[i], 1/(sigma^2))
    mu[i] <- alpha[dept[i]] + beta[dept[i]]*x[i]
  }

  # Model the slope for each dept as Normal with a mean
  # and standard deviation from hyper-priors. Apply the
  # model to all depts.
  for (j in 1:max(dept[])) {
    beta[j] ~ dnorm(beta_mu, 1/(beta_sigma^2))
  }

  # The intercept values for each dept are included in
  # the observations. Use the observed value as the
  # mean for a Normal distribution with a standard
  # deviation equal to the overall standard deviation.
  for (j in 1:max(dept[])) {
    alpha[j] ~ dnorm(intercept[j], 1/(sigma^2))
  }

  # Posterior Predictive Check
  #
  # Calculate the p-value using a sum-of-squares
  # test for fitness. To do that, first calculate
  # the sum of squares of the residuals. (The
  # residuals are the differences between the observed
  # y-values and the values that the current iteration's
  # parameters would predict.)
  for (i in 1:length(y)) {
    sq.res[i] <- pow((ln_y[i] - mu[i]), 2)
  }

  # Next, generate a new y-value given the current
  # iteration's parameters and compare that new
  # value with what the parameters would predict.

  for (i in 1:length(y)) {
    ln_y.new[i] ~ dnorm(mu.new[i], 1/(sigma^2))
    mu.new[i] <- alpha[dept[i]] + beta[dept[i]]*x[i]
    sq.new[i] <- pow((ln_y.new[i] - mu[i]), 2)
  }

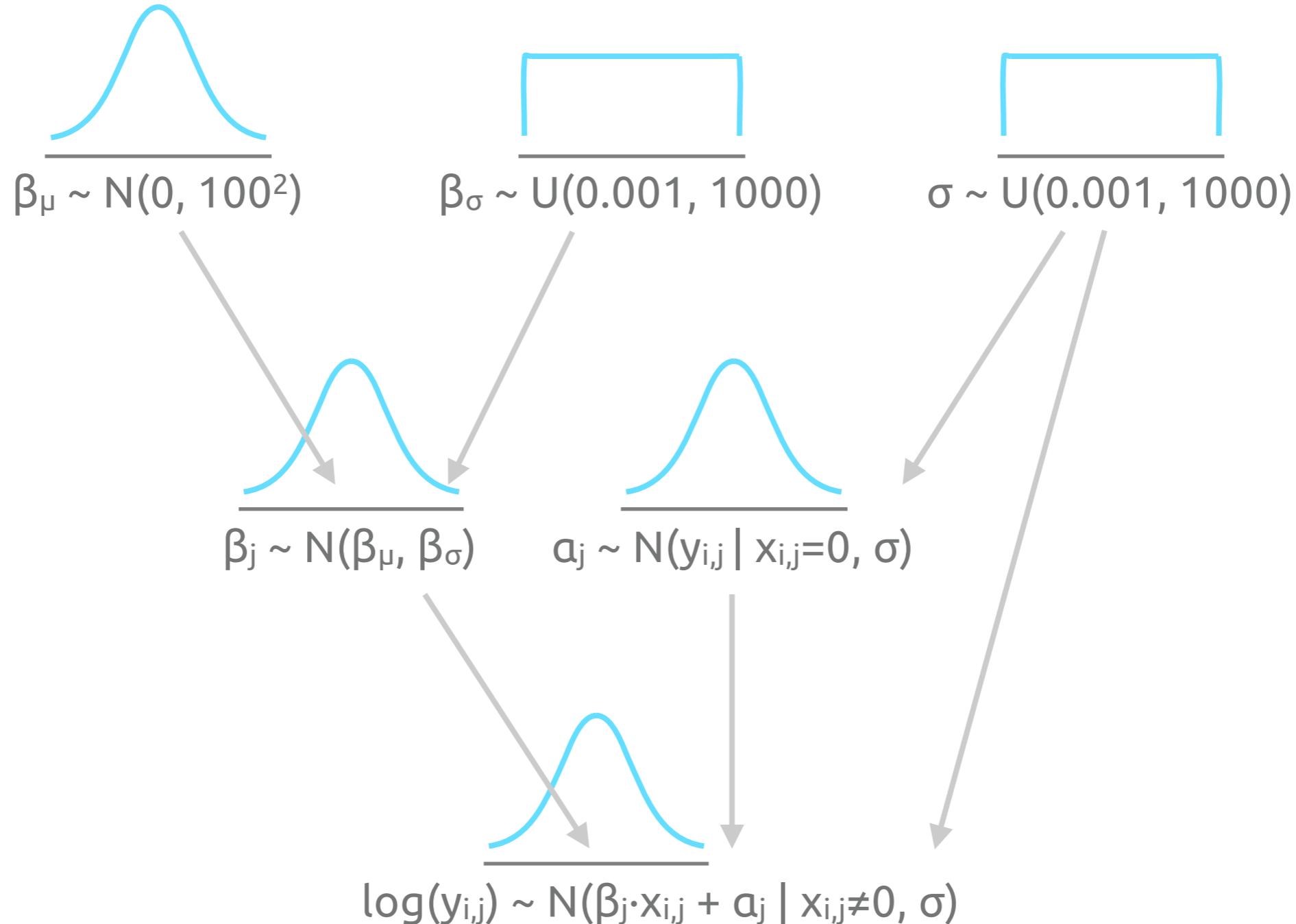
  # For the goodness of fit test, compare the sums
  # of the two squared values to see which is greater.
  # If the model is a good fit, neither sum is more
  # likely to be greater than the other. The
  # resulting p-value should, therefore, be close to
  # 0.5.
  test <- step(sum(sq.new[]) - sum(sq.res[]))
  pvalue <- mean(test)

  # Additional Processing
  #
  # Although not part of the regression itself, the
  # model can be used to estimate  $R_0$  from the
  # regression parameters. The following code relies
  # on the approach documented in
  #
  #   Heffernan, J M, Smith, R J, & Wahl, L M (2005).
  #   "Perspectives on the basic reproductive ratio."
  #   Journal of the Royal Society Interface, 2(4),
  #   281-293. http://doi.org/10.1098/rsif.2005.0042
  #
  # For the serial interval time, use data reported in
  #
  #   Majumder M S, Cohn E, Fish D & Brownstein J S.
  #   Estimating a feasible serial interval range for
  #   Zika fever [Submitted]. Bull World Health Organ,
  #   E-pub: 9 Feb 2016.
  #   http://dx.doi.org/10.2471/BLT.16.171009
  #
  # Majumder et al report only a range for their
  # estimates without a distribution, so the following
  # code assumes a uniform distribution within the
  # estimated range. (Note that Majumder et al's range
  # is specified in days. It must be converted to weeks
  # to correspond to the observations.)
  #
  # The initial growth rate ( $r$ ) in the calculation is
  # simply the estimated slope  $\beta$ .
  for (j in 1:max(dept[])) {
    si[j] ~ dunif(10, 23)
    R0[j] <- 1 + beta[j]*si[j]/7
  }

  # Also model an overall  $R_0$ 
  si_mean ~ dunif(10, 23)
  R0_mean <- 1 + beta_mu*si_mean/7
}
```

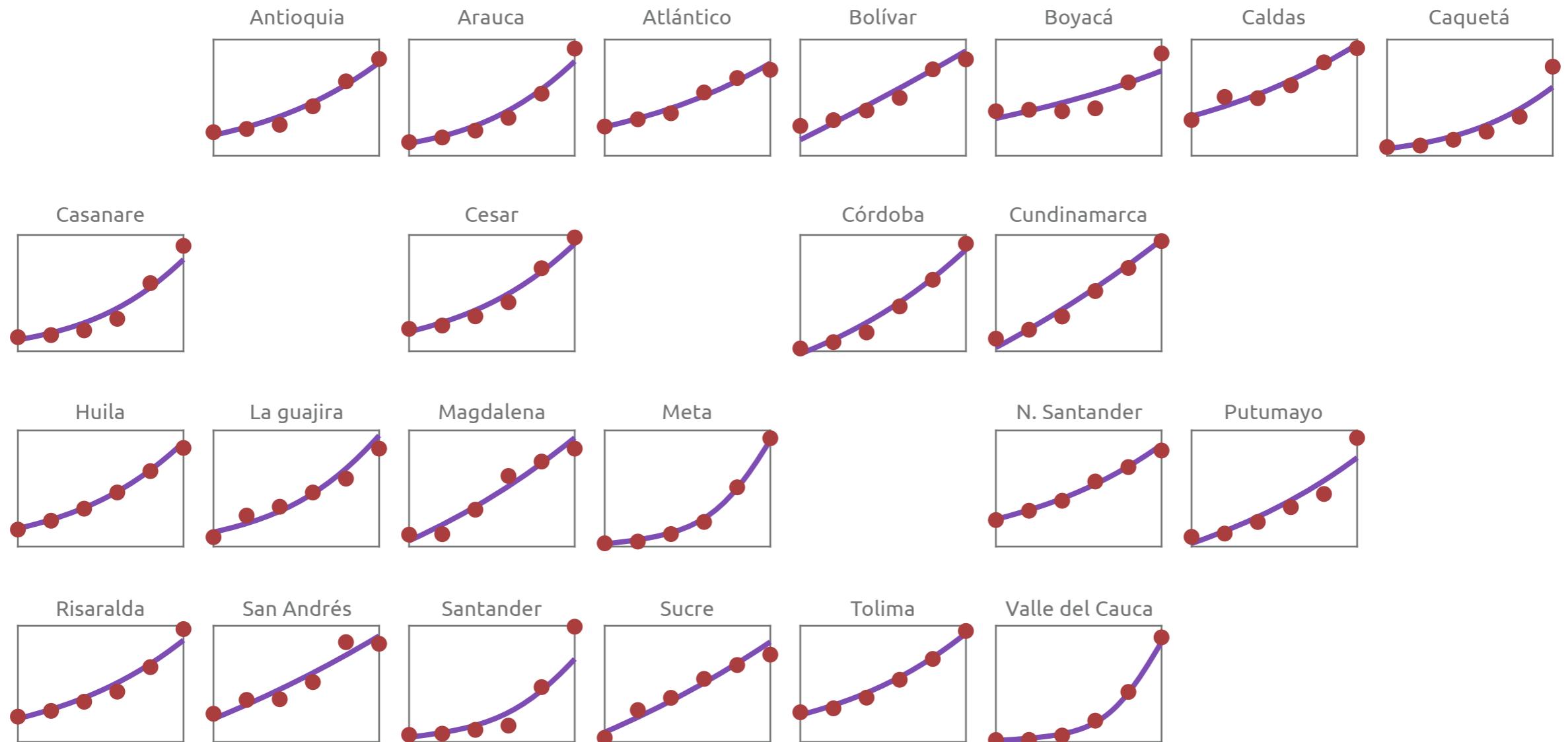
# HIERARCHICAL REGRESSION MODEL (LOG TRANSFORM)

---



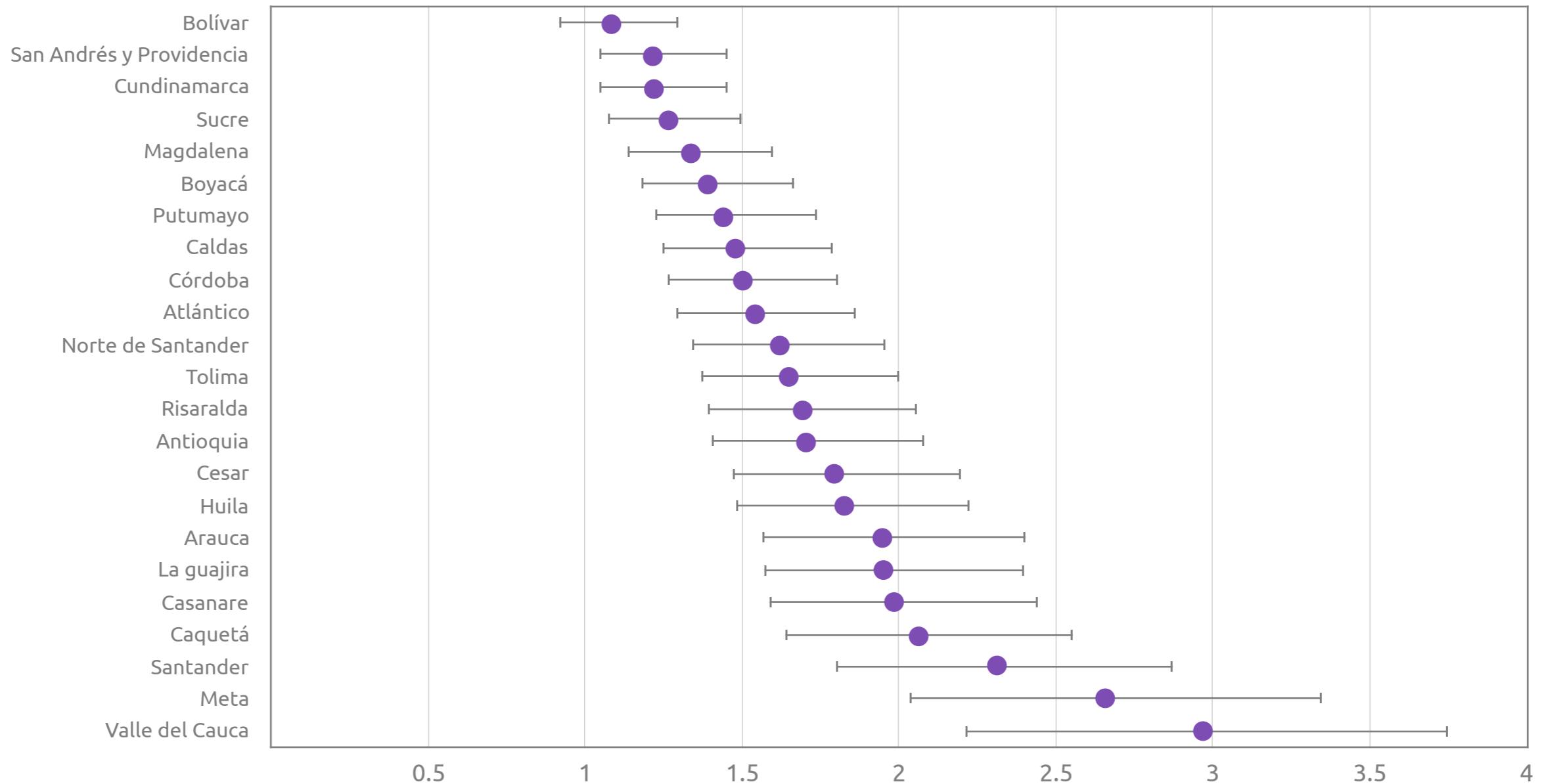
# REGRESSION RESULTS

---



# 95% CREDIBLE INTERVALS FOR $R_0$ BY DEPARTMENT

---

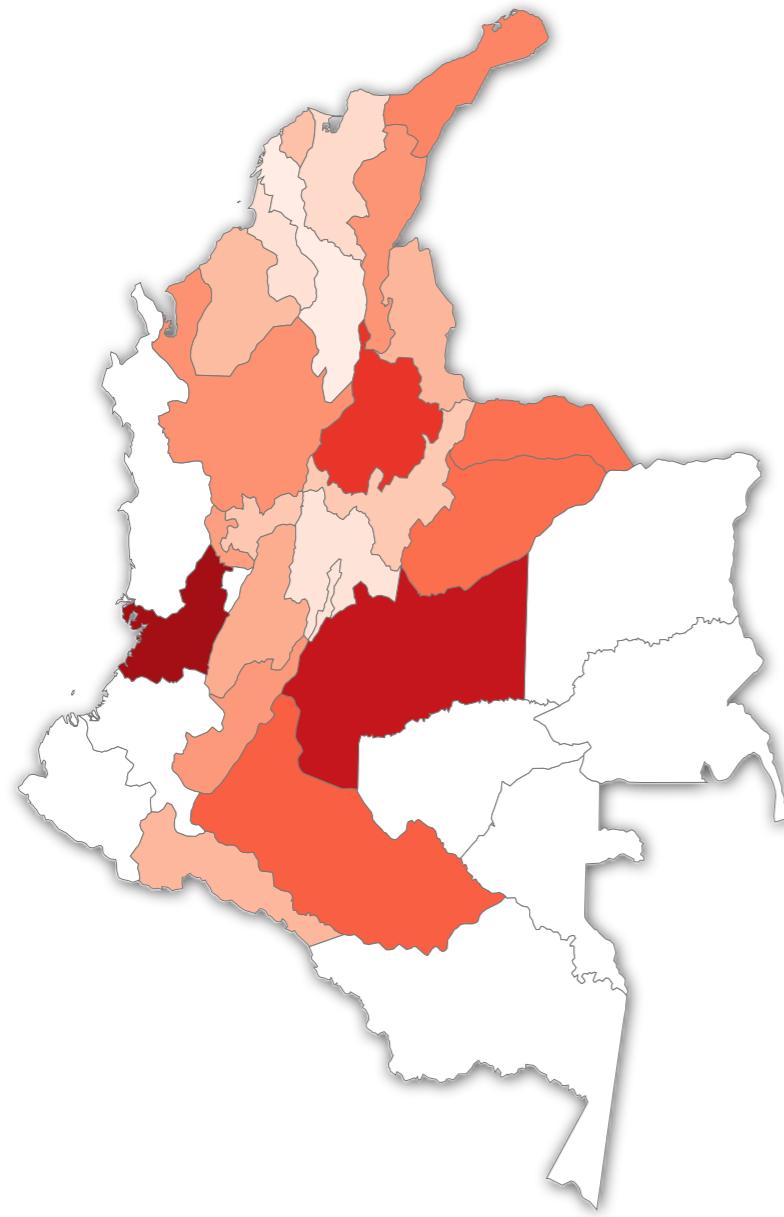


Compare with Nishiura, H., Kinoshita, R., Mizumoto, K., Yasuda, Y., & Nah, K. (2016). "Transmission potential of Zika virus infection in the South Pacific." *International Journal of Infectious Diseases*, 45, 95–97. <http://doi.org/10.1016/j.ijid.2016.02.017>

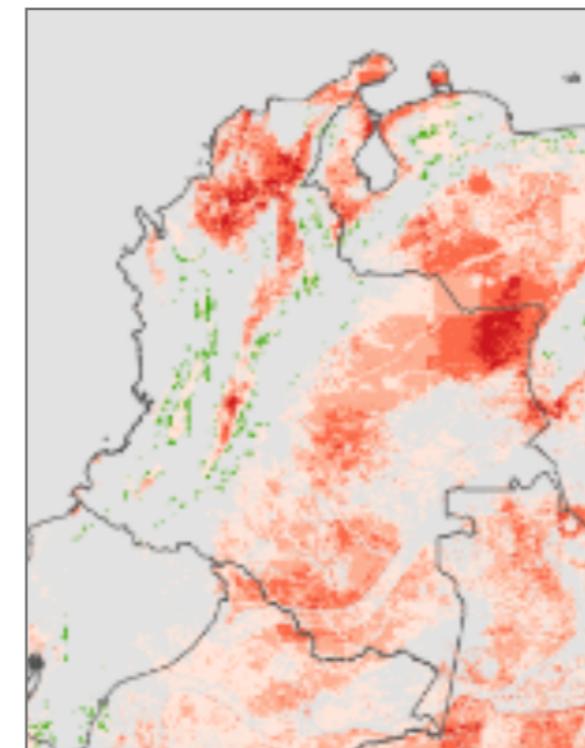
Yap Island: 2.7 - 5.0, French Polynesia: 1.8 - 2.0

# GEO-SPATIAL VARIATION IN $R_0$

---



Compare with projected  $R_0$  from Perkins, et al, "Model-based projections of Zika virus infections in childbearing women in the Americas" based on discrete model parameters, e.g. daily temperature, mosquito mortality, transmission probabilities, mosquito biting rate, etc.



# EVALUATE DOMINANT CLIMATE CLASSIFICATION

---

```
# This is a JAGS model for performing a one-way ANOVA
# analysis of R0 based on climate classification. The
# input data consists of the estimated R0 values in
# y[] and an index into the climate classification
# in x[].

model {

  # For consistency, the following indices are used
  # throughout the model:
  #
  # i: the individual observations
  # j: the group number

  # Prior Probabilities
  # ----

  # All stochastic parameters use reference
  # (non-informative) priors.

  a0 ~ dnorm(0, 0.0001)
  sigma ~ dunif(0.001, 1000)

  # Note that no prior is needed for a[1] as
  # the sum-to-zero constraint will force a value
  # for a[1] based on the remaining a[] elements.

  for (j in 2:max(x[])) {
    a[j] ~ dnorm(0, 0.0001)
  }

  # Constraints
  # -----
  #
  # Implement the sum-to-zero constraint
  a[1] <- - sum(a[2:max(x[])])

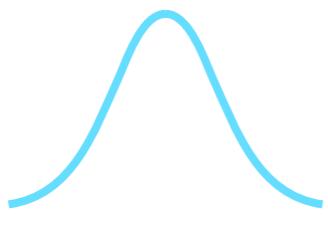
  # The Likelihood
  # -----
  for (i in 1:length(y)) {

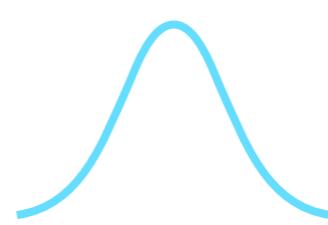
    # Assume the measured value is normal with mean mu
    # and standard deviation sigma. Each observation
    # has its own mean, and we assume a common value
    # for the standard deviation.
    #
    # The mean for each measurement is calculated as the
    # sum of a base value (a0) and the effect of the
    # group. This latter quantity is a[j], where j is
    # the group index. The value for j is read from the
    # data as x[i], where i is the observation index.
    # Note that this assumes that groups are coded as
    # 1, 2, ...

    y[i] ~ dnorm(a0 + a[x[i]], 1/(sigma^2))
  }
}
```

# ANOVA MODEL

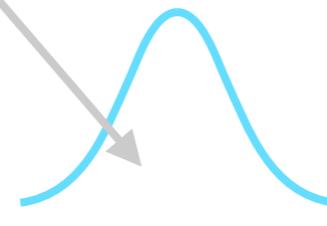
---


$$a_0 \sim N(0, 100^2)$$


$$a_{ij} \sim N(0, 100^2)$$

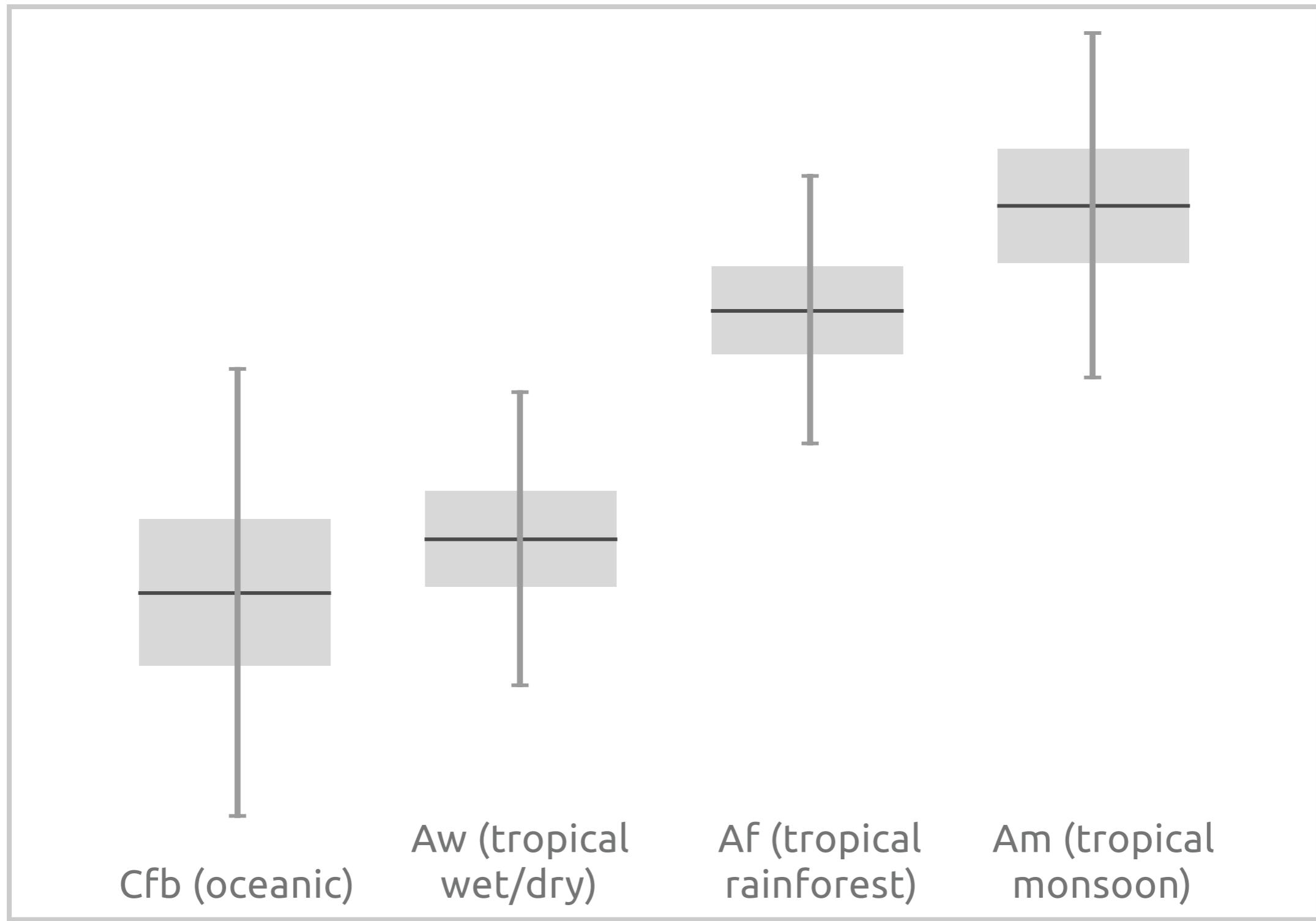

$$\sigma \sim U(0.001, 1000)$$

$$\mu_i \leftarrow a_0 + \sum_j a_{ij} \cdot x_{ij}(i)$$


$$y_i \sim N(\mu_i, \sigma)$$

# EFFECTS OF DOMINANT CLIMATE CLASSIFICATION ON $R_0$

---



# MULTIPLE LINEAR REGRESSION

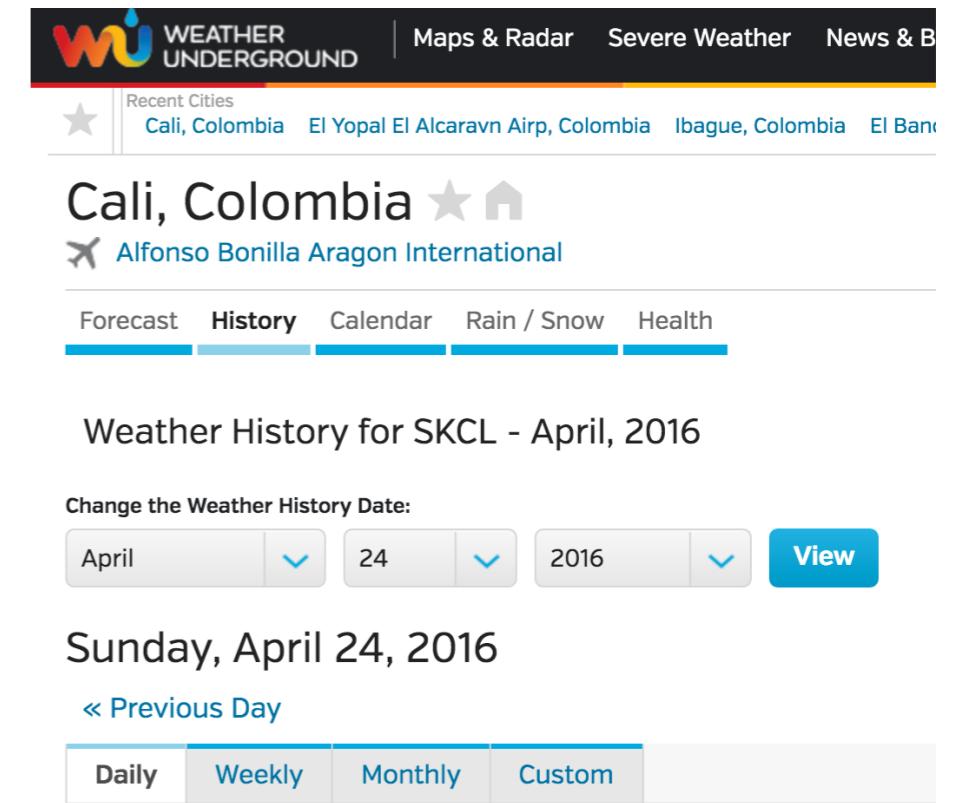
---

- What is likely to impact the R0 in the different departments?
  - Human density
    - Use census data.
    - Latest is from 2005.
  - Mosquito density
    - Not available
    - Inherent effects of disease
      - Not available
- So what do we use to replicate these effects using commonly available data?

# MULTIPLE LINEAR REGRESSION – WEATHER AND CLIMATE

---

- Weather and climate data
  - Easily available ([wunderground.com](http://wunderground.com))
  - Many impact on mosquito population
- Data available during time of study
  - Temperature (minimum, maximum, mean)
  - Sea level pressure
  - Rainfall



$$R_{0,n} = \beta_0 + \beta_1 \cdot (\text{pop density}) + \beta_2 \cdot (\text{Temp}_{\min}) + \beta_3 \cdot (\text{Temp}_{\text{mean}}) + \beta_4 \cdot (\text{Temp}_{\max}) + \beta_5 \cdot (\text{Rain}_{\text{tot}}) + \beta_6 \cdot (\text{SLP})$$

# MULTIPLE LINEAR REGRESSION – POPULATION DENSITY

---

- Population Density has large, non-linear spread
  - Key factor is magnitude
  - Log transformation
- DIC Values
  - Linear: 57.9
  - Log: 46.3

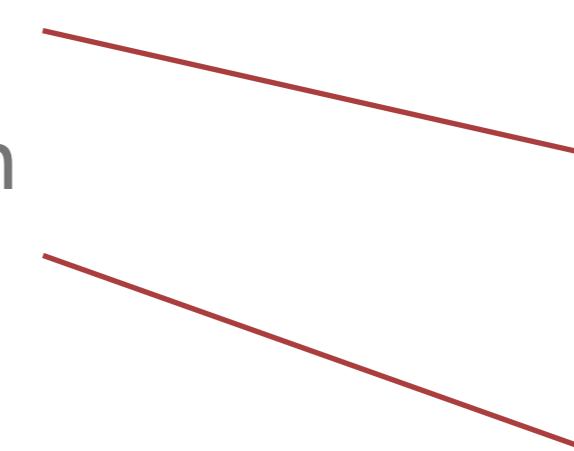
Department	Population Density
Antioquia	89.16
Arauca	8.76
Atlántico	623.41
Bolívar	71.62
Boyacá	52.23
Caldas	115.22
Caquetá	4.55
Casanare	6.33
Cesar	38.42
Córdoba	58.86
Cundinamarca	92.05
Huila	50.62
La guajira	29.89
Magdalena	49.03
Meta	9.22
Norte de Santander	56.70
Putumayo	12.03
Risaralda	208.61
San Andrés y Providencia	1145.63
Santander	62.75
Sucre	71.72
Tolima	56.67
Valle del Cauca	183.39

$$R_{0,n} = \beta_0 + \beta_1 \cdot \log(\text{pop density}) + \beta_2 \cdot (\text{Temp}_{\min}) + \beta_3 \cdot (\text{Temp}_{\text{mean}}) + \beta_4 \cdot (\text{Temp}_{\max}) + \beta_5 \cdot (\text{Rain}_{\text{tot}}) + \beta_6 \cdot (\text{SLP})$$

# MULTIPLE LINEAR REGRESSION - MODEL SELECTION

- Use DIC to eliminate variables

1. Remove minimum temperature
2. Remove maximum temperature



DIC	Dbar	Dhat	pD	DIC
<i>Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes</i>				
R0.1	69.108	65.305	3.803	72.911
R0.2	62.875	58.907	3.968	66.843
R0.3	43.323	39.512	3.812	47.135
R0.4	58.867	54.864	4.002	62.869
total	234.173	218.588	15.585	249.758

DIC	Dbar	Dhat	pD	DIC
<i>Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes</i>				
R0.1	62.538	57.743	4.795	67.333
R0.2	39.559	34.758	4.801	44.361
R0.3	46.807	41.968	4.838	51.645
R0.4	44.155	39.350	4.806	48.961
R0.5	59.831	54.833	4.998	64.828
total	252.890	228.651	24.239	277.129

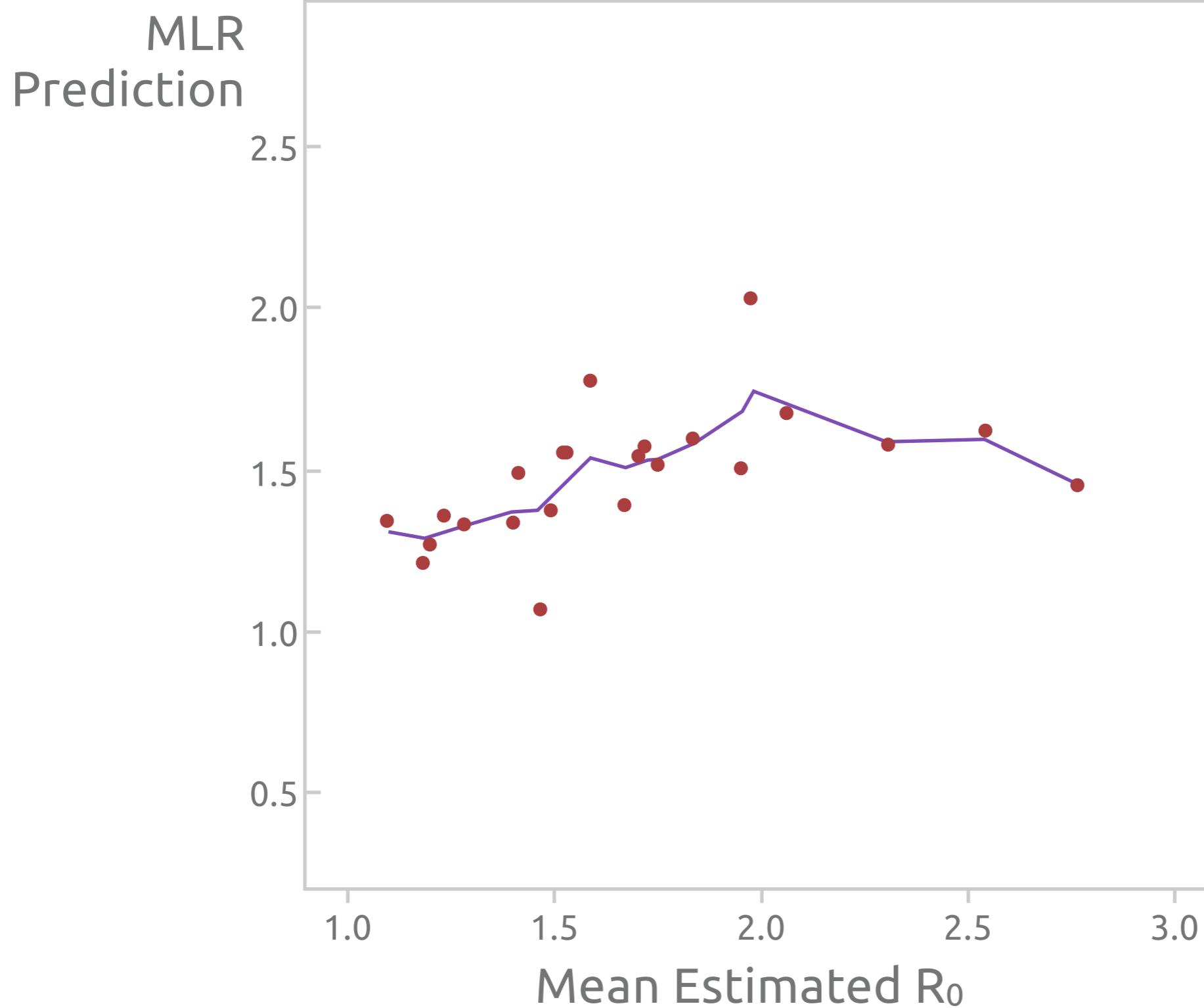
  

DIC	Dbar	Dhat	pD	DIC
<i>Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes</i>				
R0.1	52.839	47.054	5.785	58.625
R0.2	40.000	34.192	5.808	45.808
R0.3	40.535	34.743	5.792	46.326
R0.4	42.635	36.805	5.830	48.465
R0.5	45.170	39.365	5.805	50.975
R0.6	59.170	53.171	5.999	65.169
total	280.348	245.329	35.019	315.368

$$R_{0,n} = \beta_0 + \beta_1 \cdot \log(\text{pop density}) + \beta_2 \cdot (\text{Temp}_{\text{mean}}) + \beta_5 \cdot (\text{Rain}_{\text{tot}}) + \beta_6 \cdot (\text{SLP})$$

# MULTIPLE LINEAR REGRESSION - MODEL FIT

---



# MULTIPLE LINEAR REGRESSION - RESULTS

---

- 95% Credible Sets
  - $\beta_1$  Population density: -0.1572 to -0.0752
  - $\beta_2$  Mean temperature: 0.0234 to 0.0584
  - $\beta_3$  Total rainfall: 0.0288 to 0.3421
  - $\beta_4$  Mean sea level pressure: 0.9610 to 2.7030
- Higher total rainfall and temperature → robust mosquito population
- Higher population density → urban areas less hospitable to mosquitos
- Higher seal level pressure → more tranquil weather?