

Bayesian estimation of missing traits and hidden disease dynamics

Maxwell B. Joseph

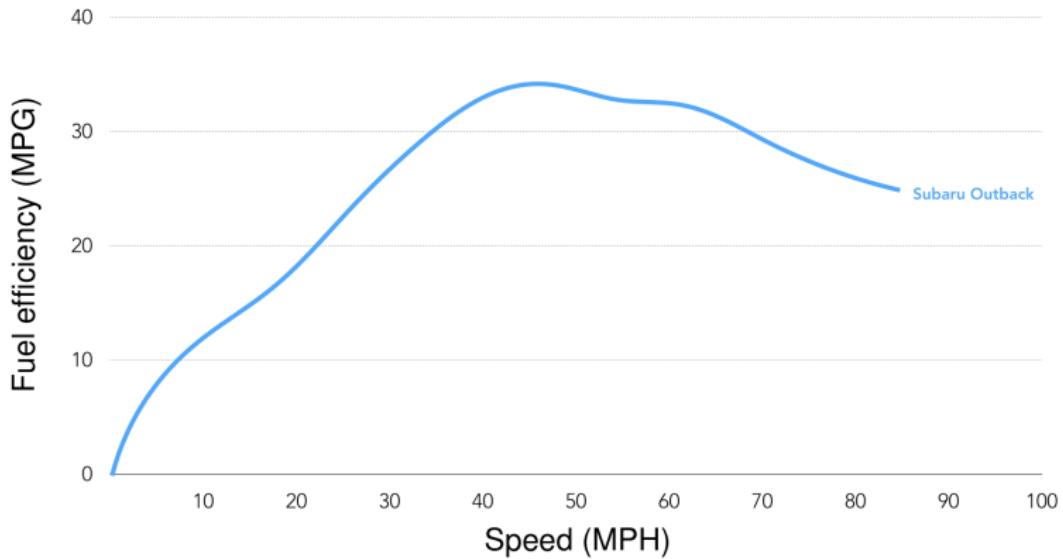
April 1, 2016

Claim: we use functions all the time

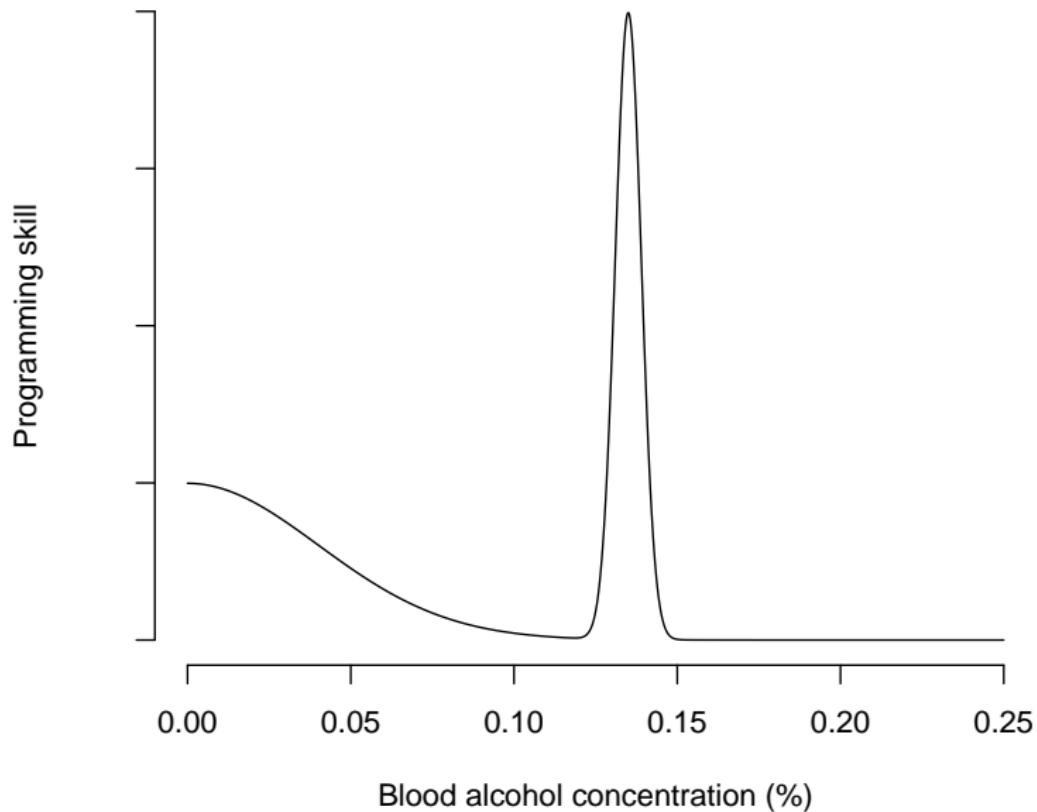
Claim: we use functions all the time

Reminder: a function $f(x)$ maps an input x to an output y

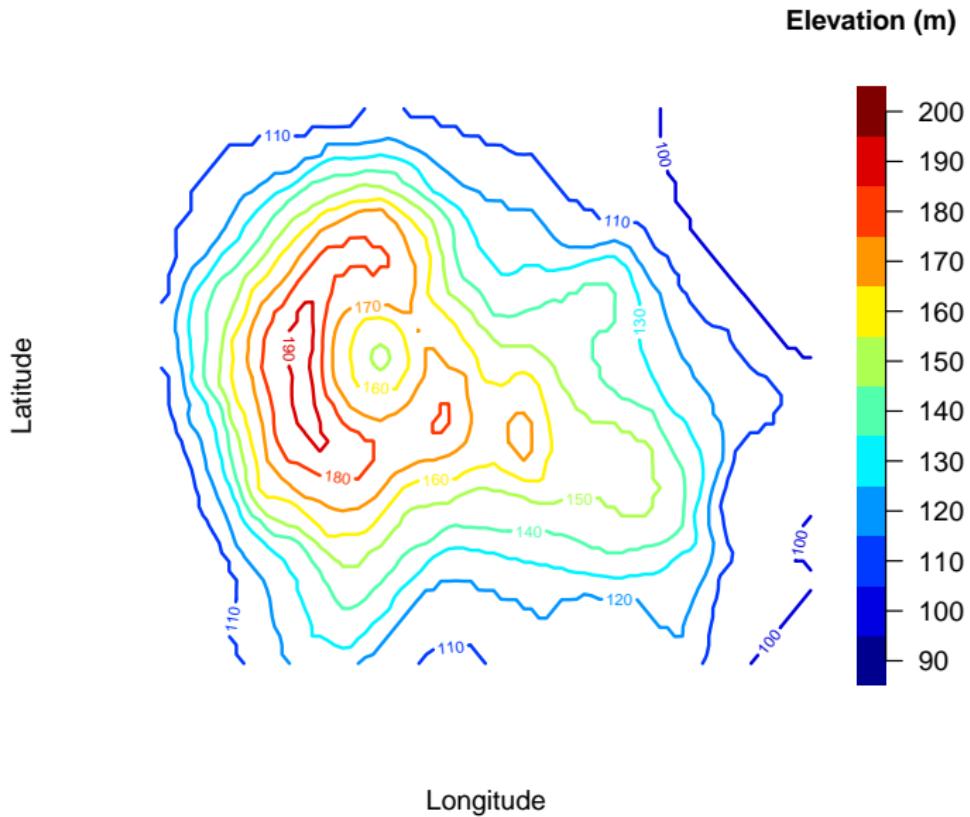
Example: how fast should I drive?



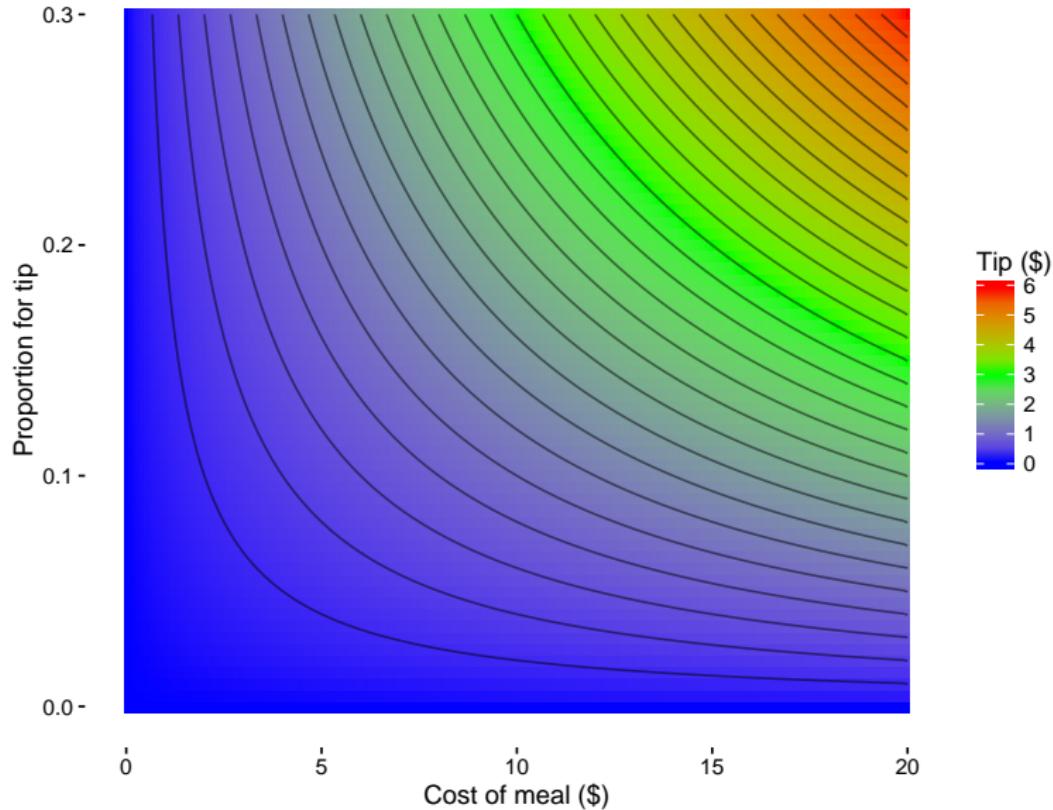
Example: computer programming



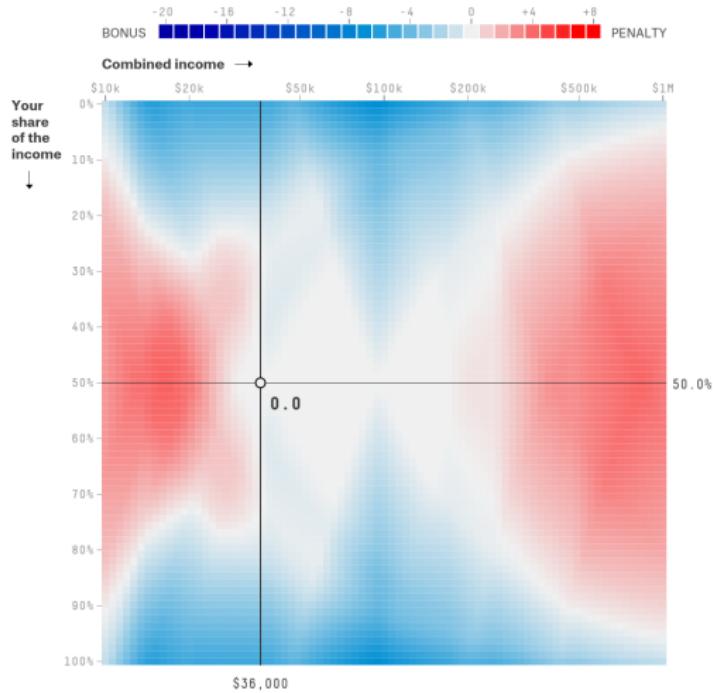
Example: climbing a mountain



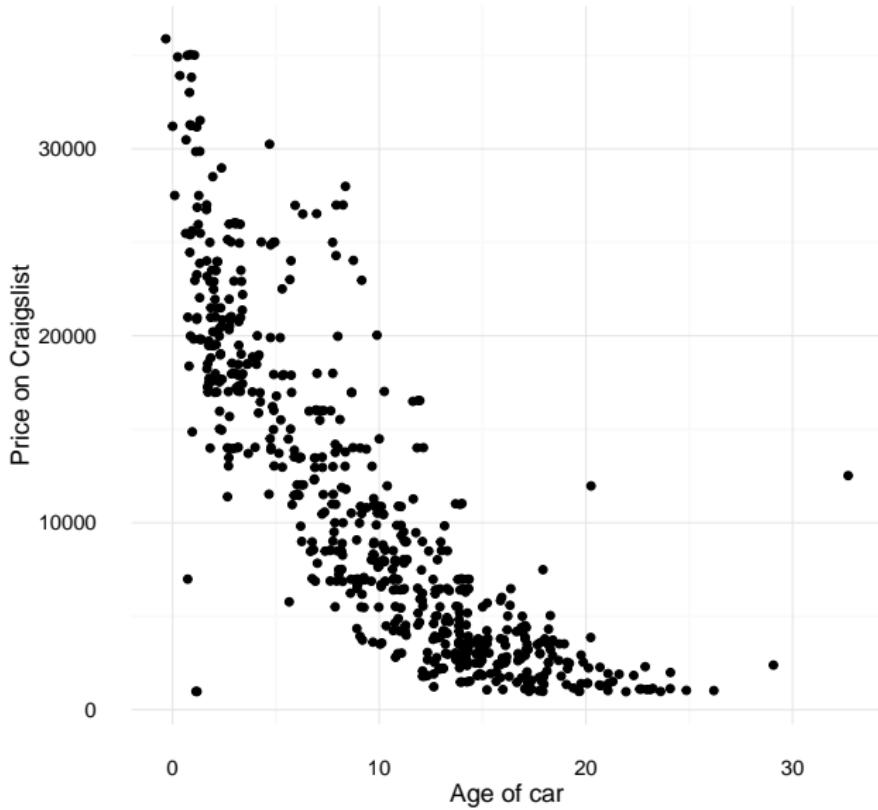
Example: leaving a tip



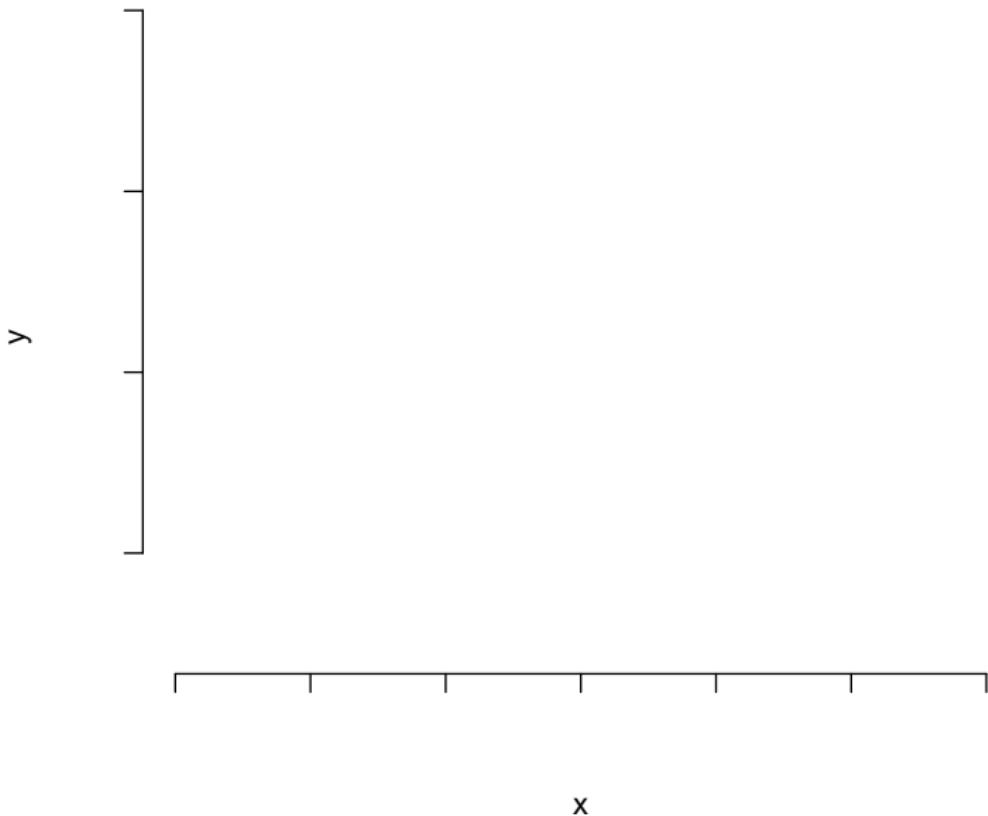
Example: deciding to marry a grad student

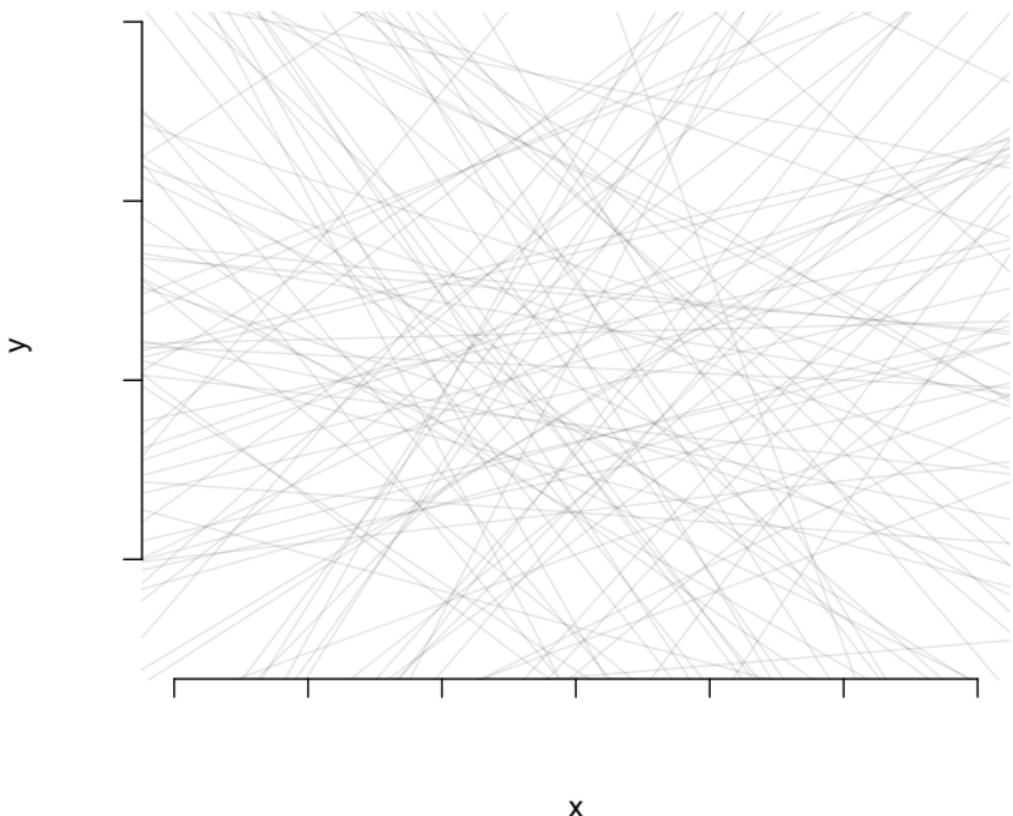


Example: buying/selling a Subaru in Colorado

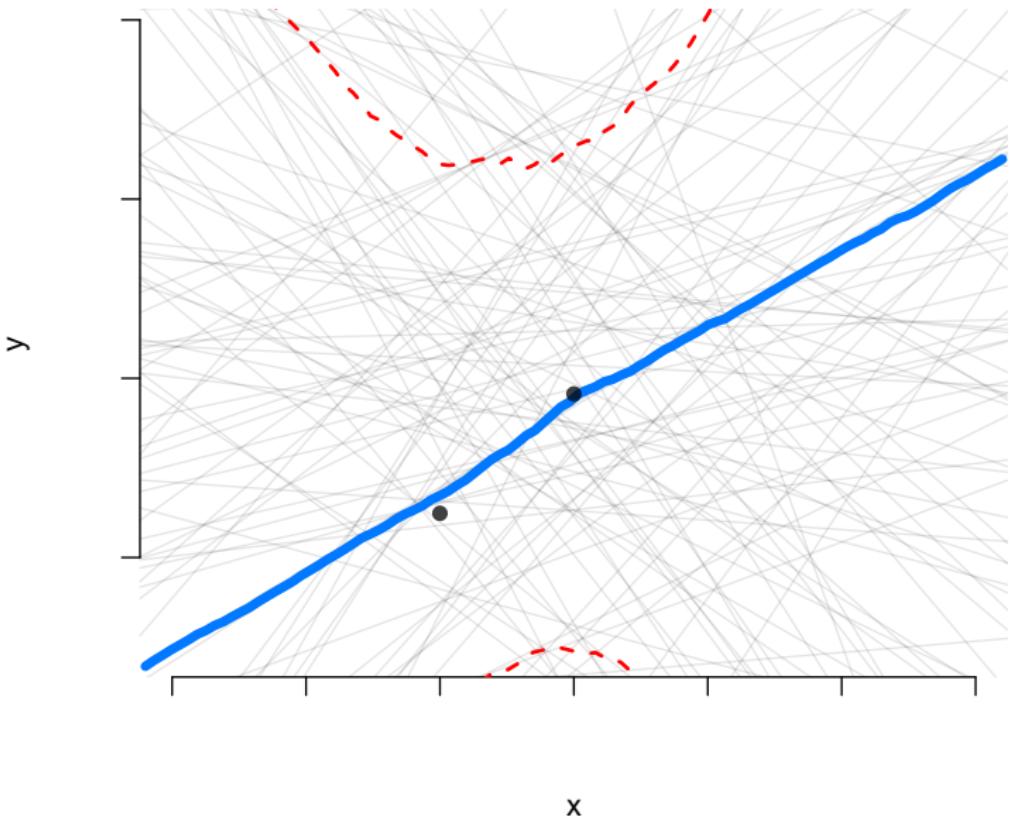


Use your imagination

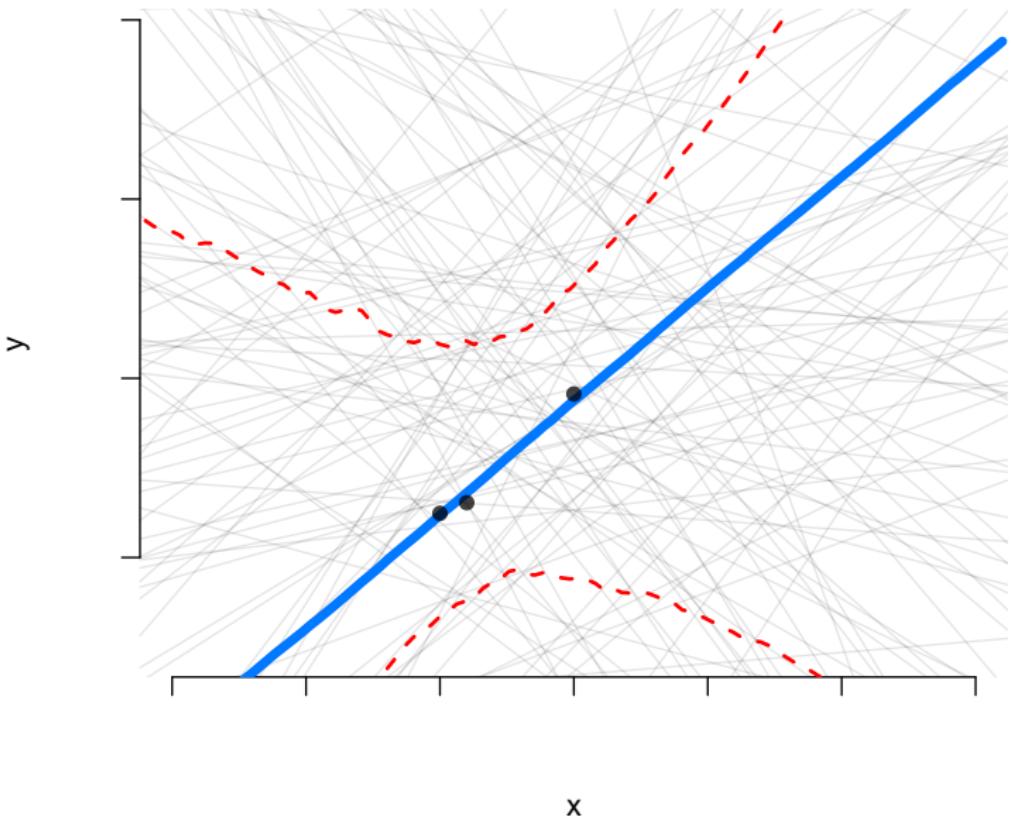




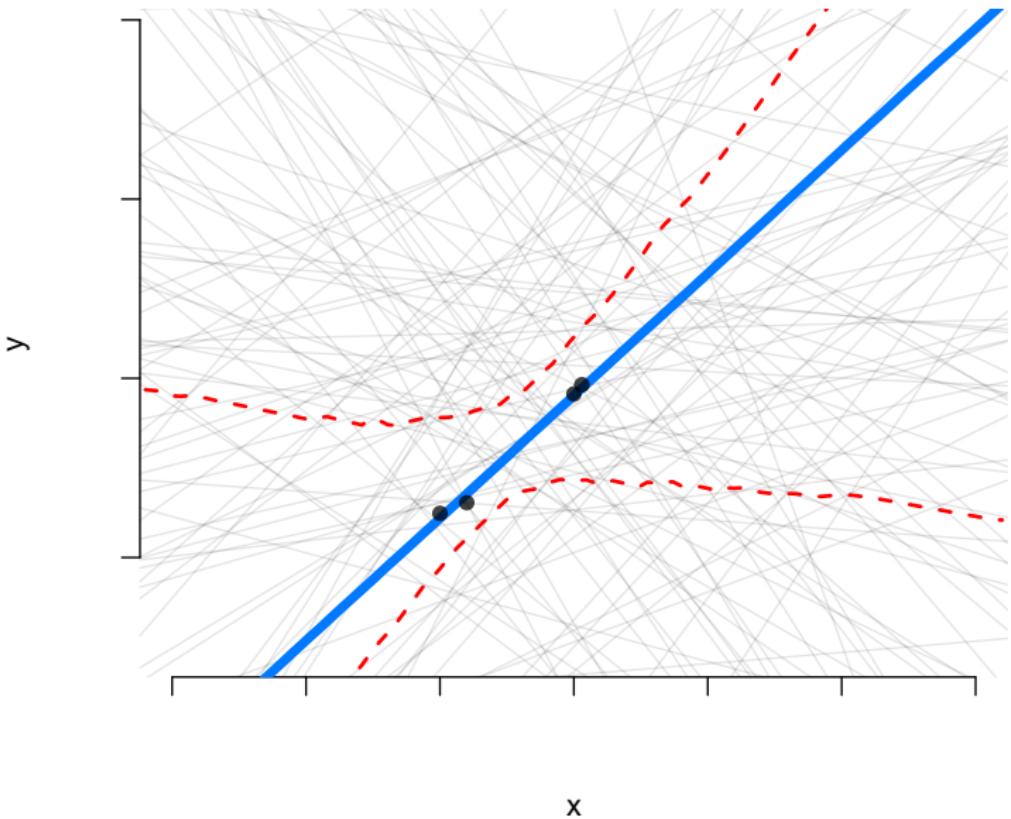
2 observations



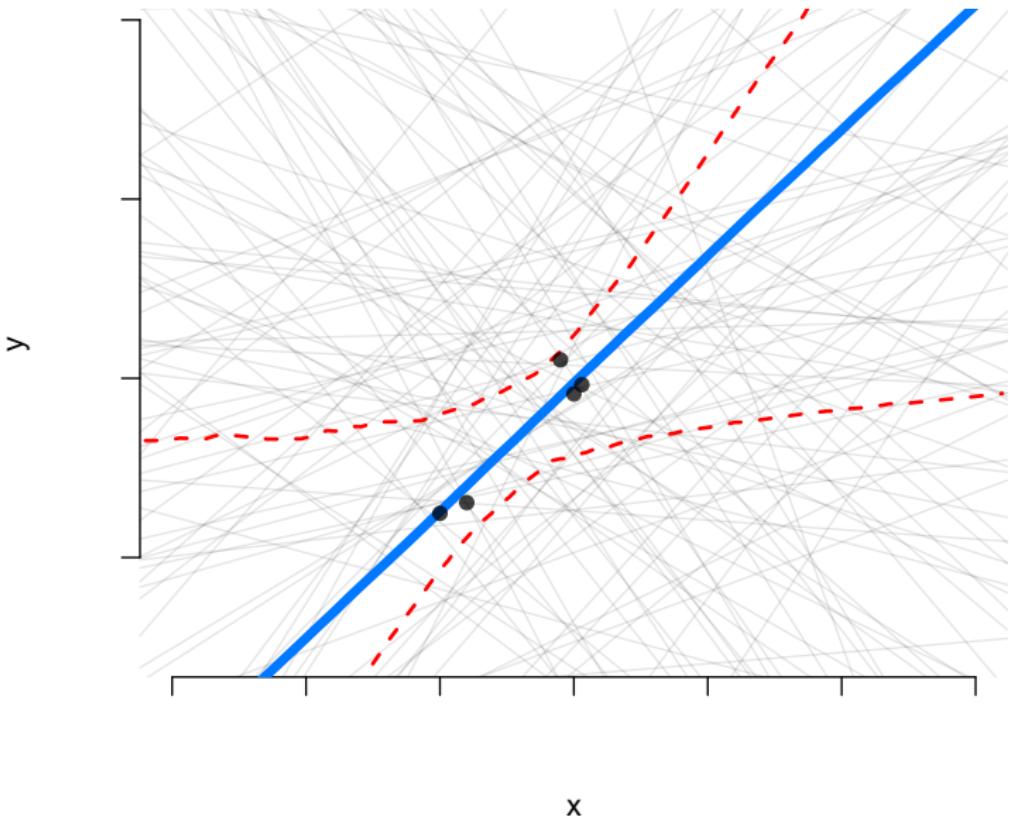
3 observations



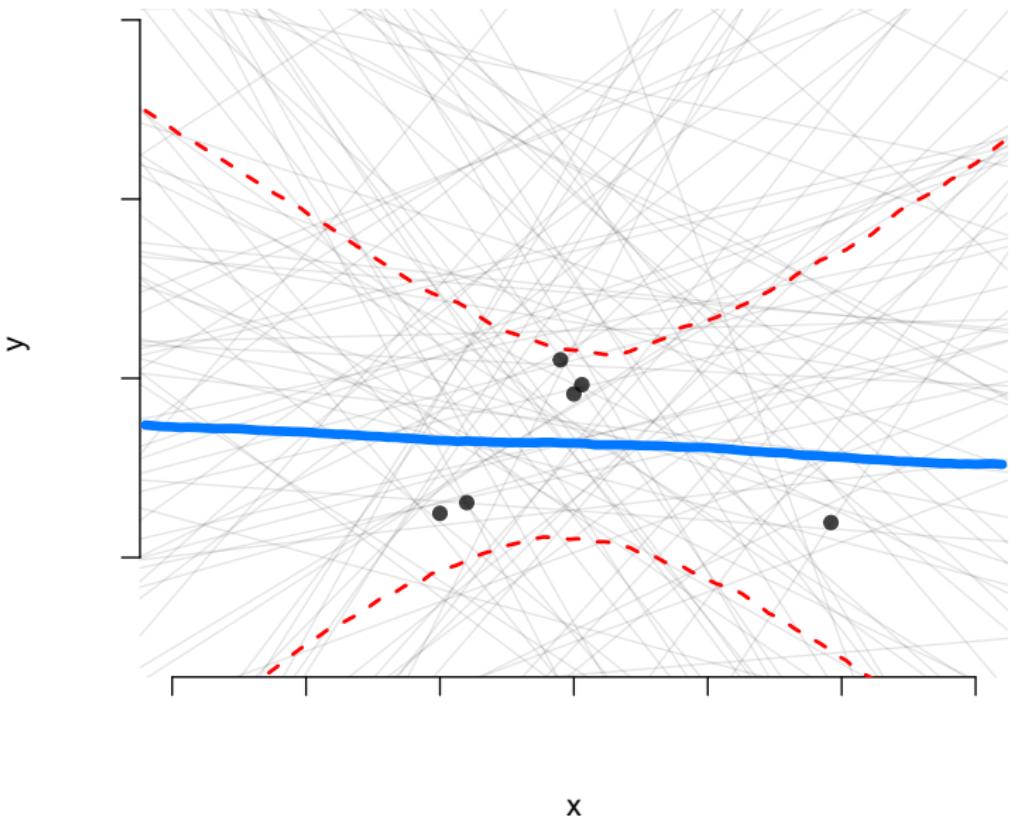
4 observations



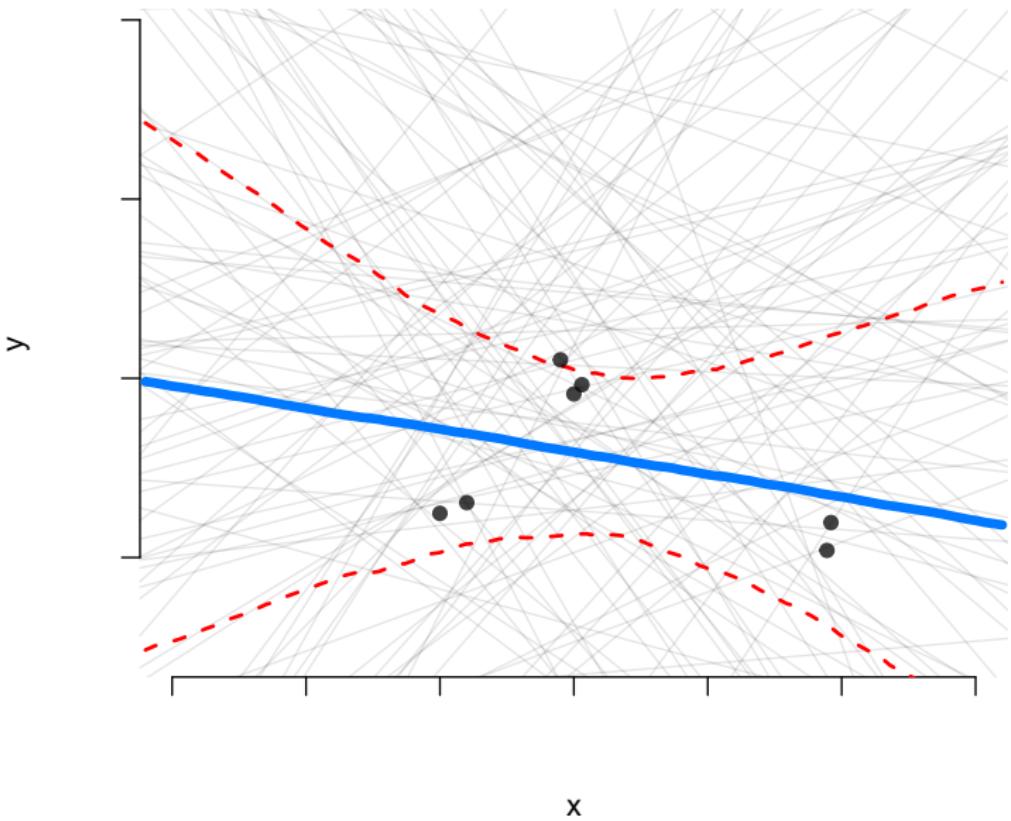
5 observations



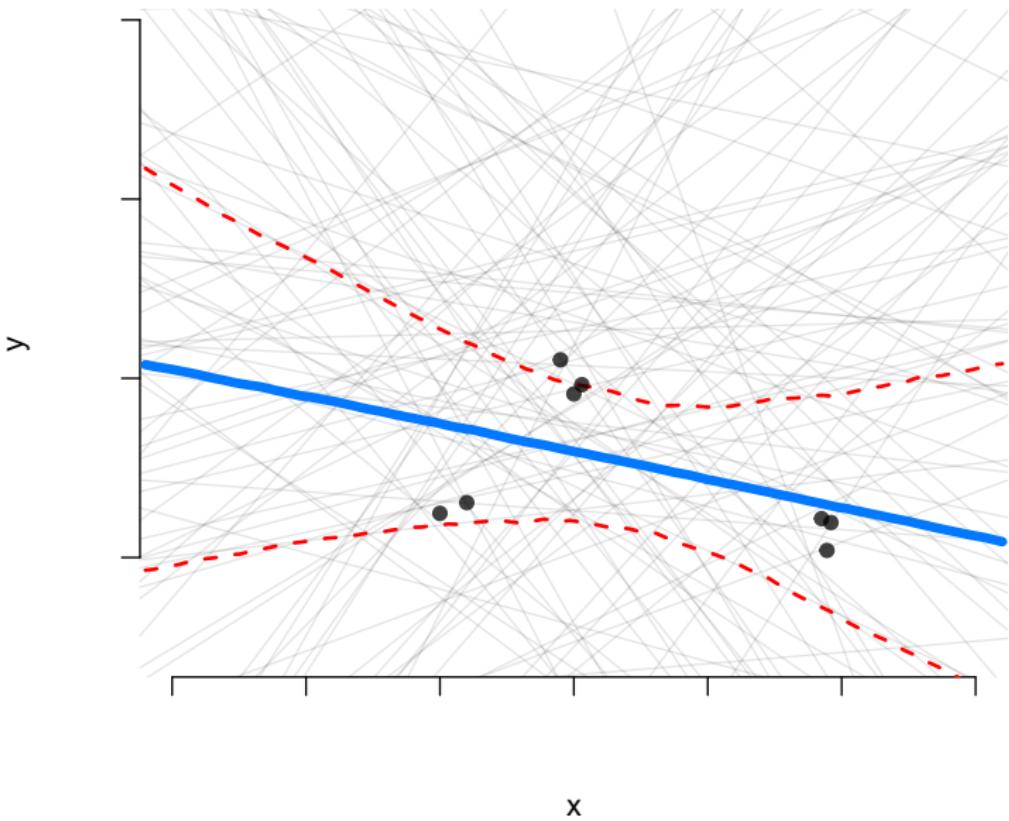
6 observations



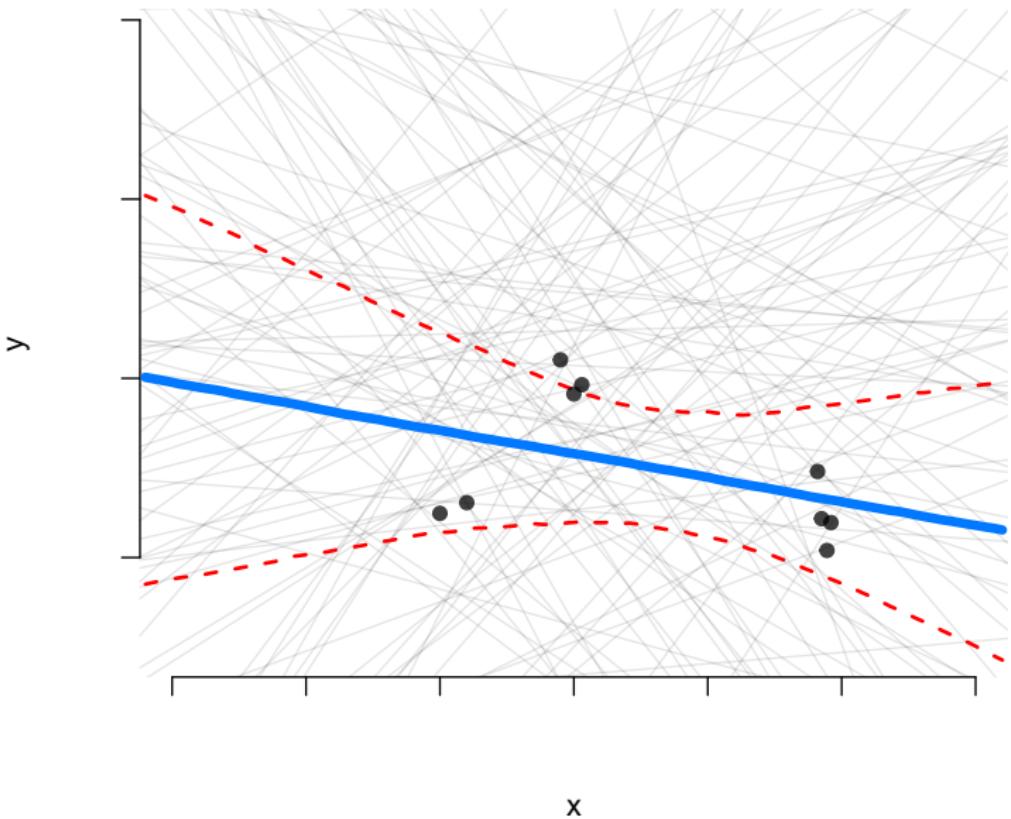
7 observations



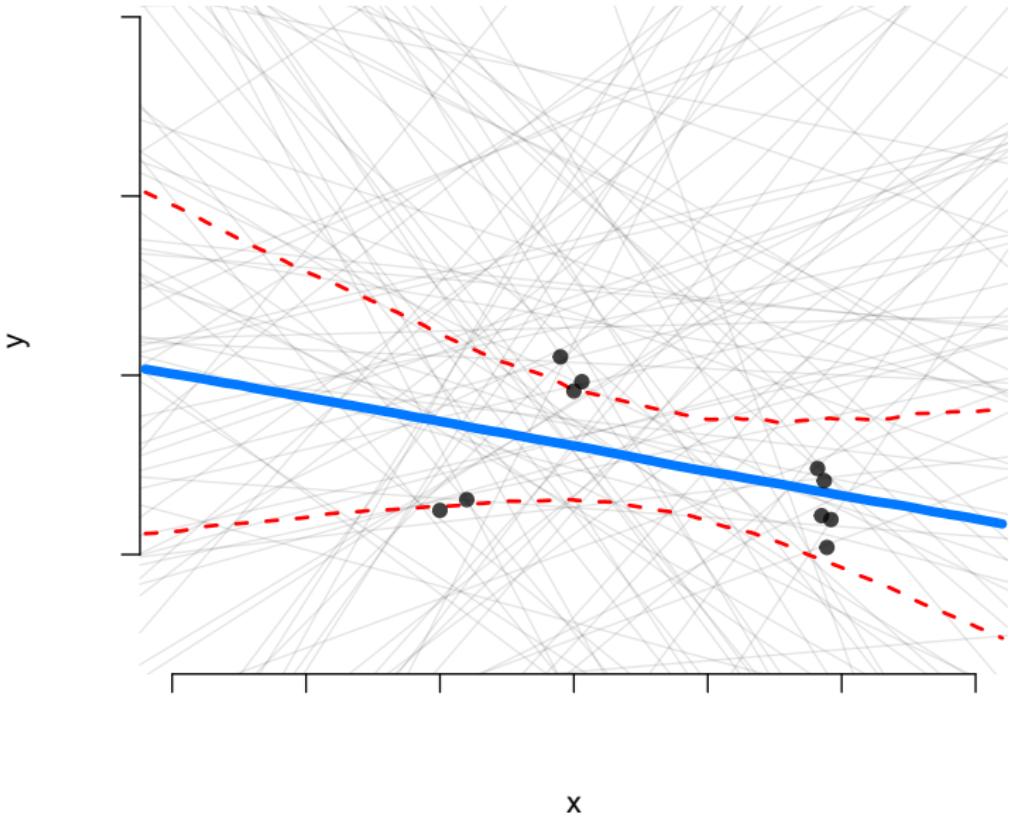
8 observations



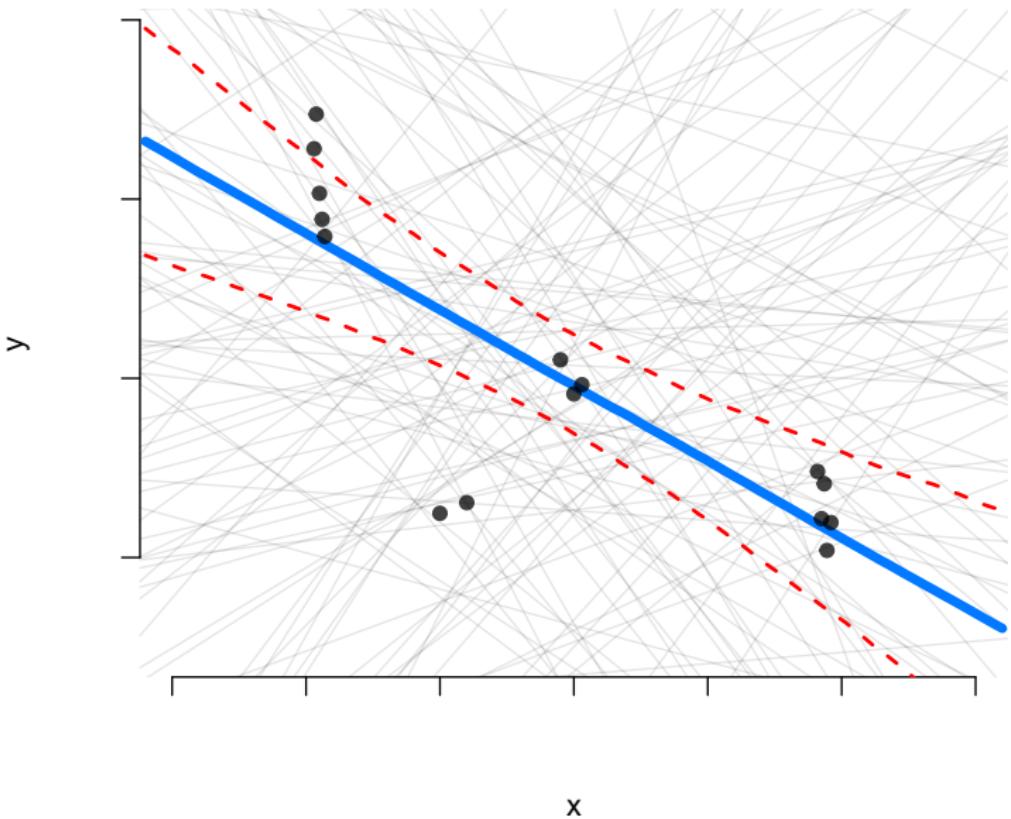
9 observations



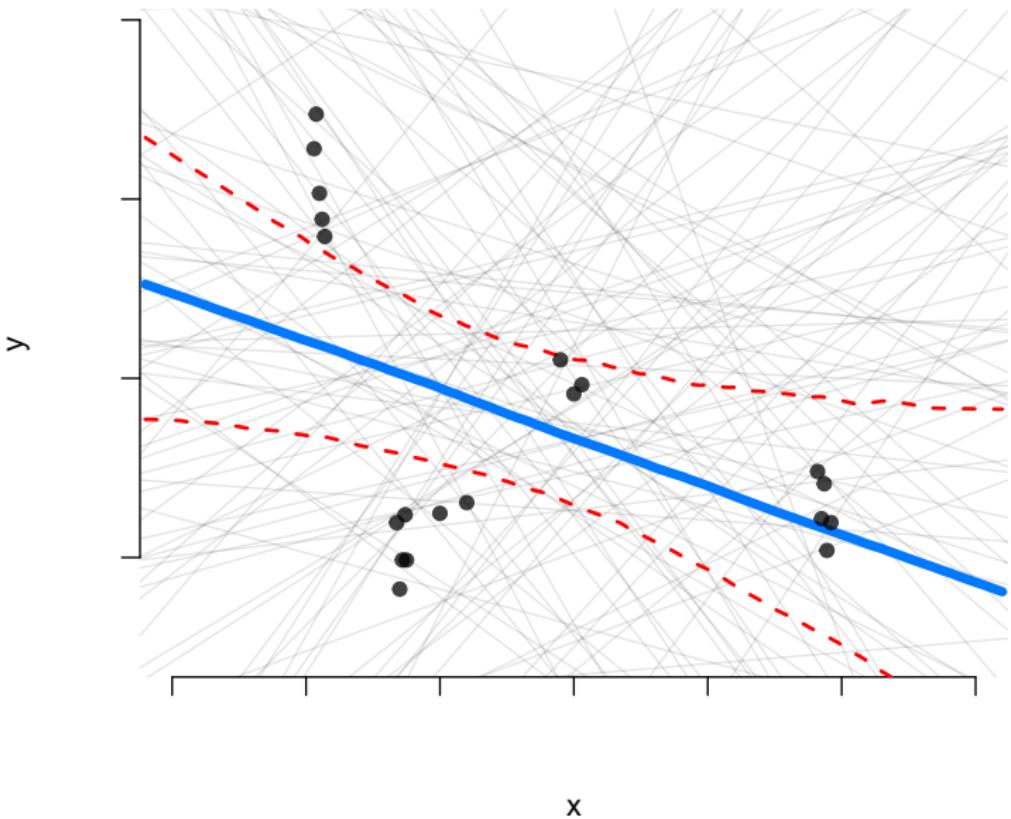
10 observations



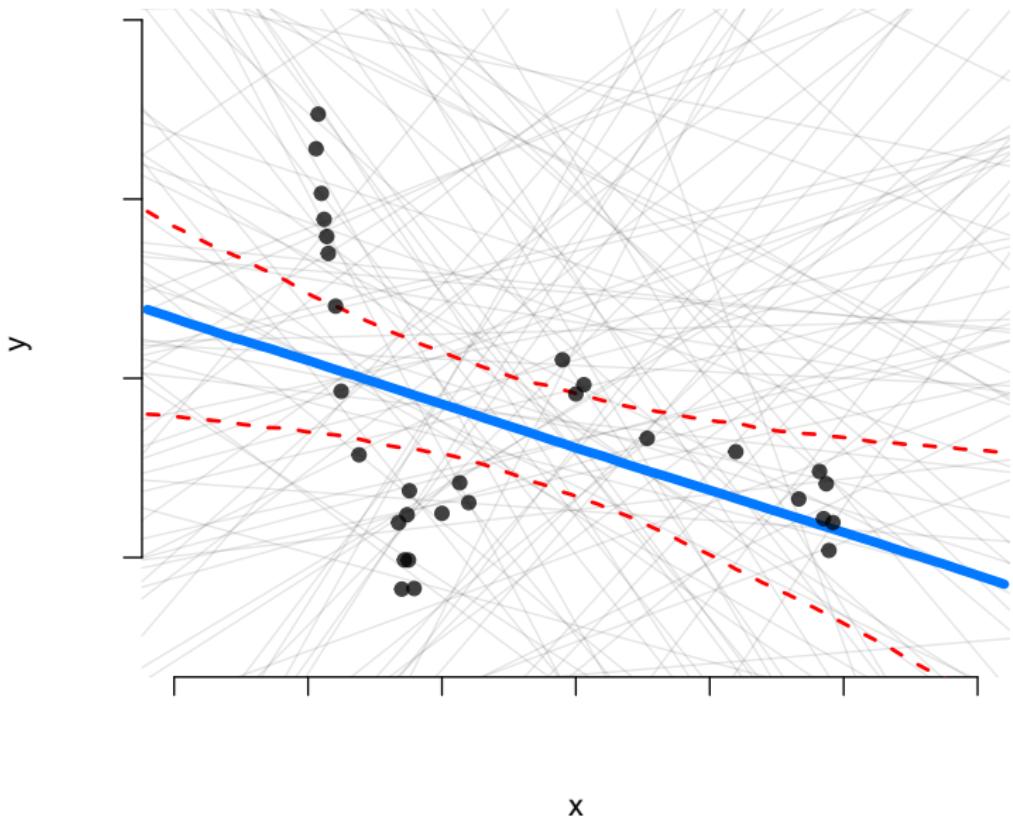
15 observations



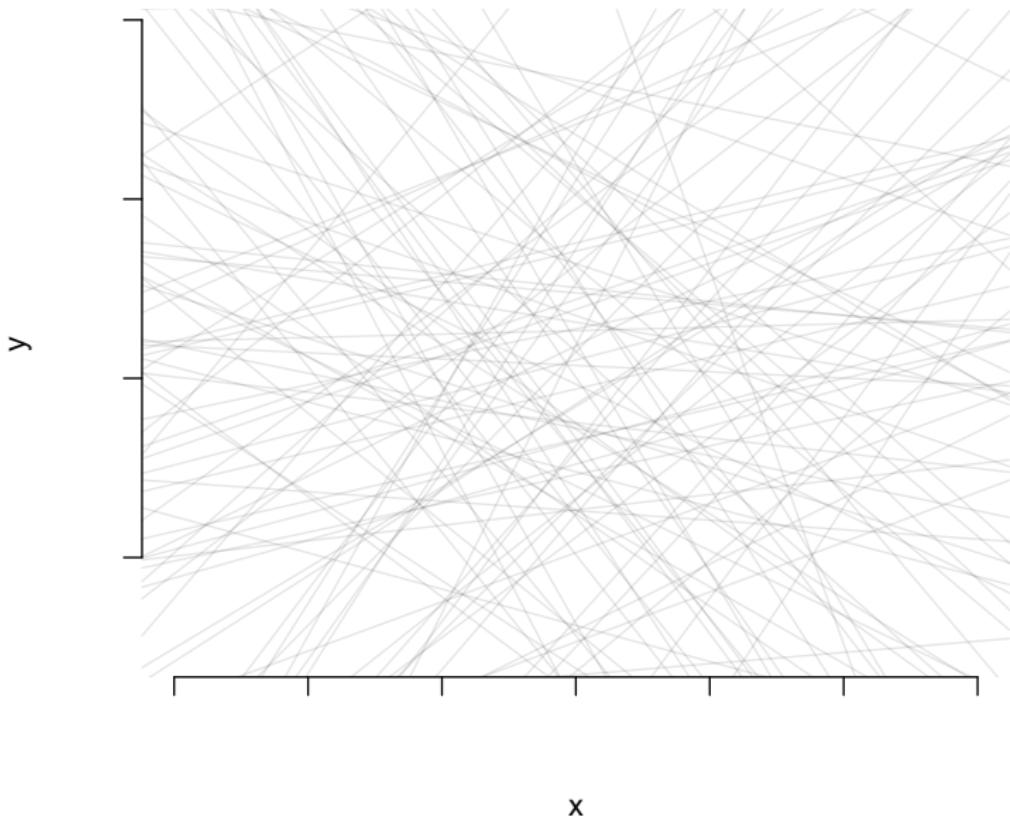
20 observations



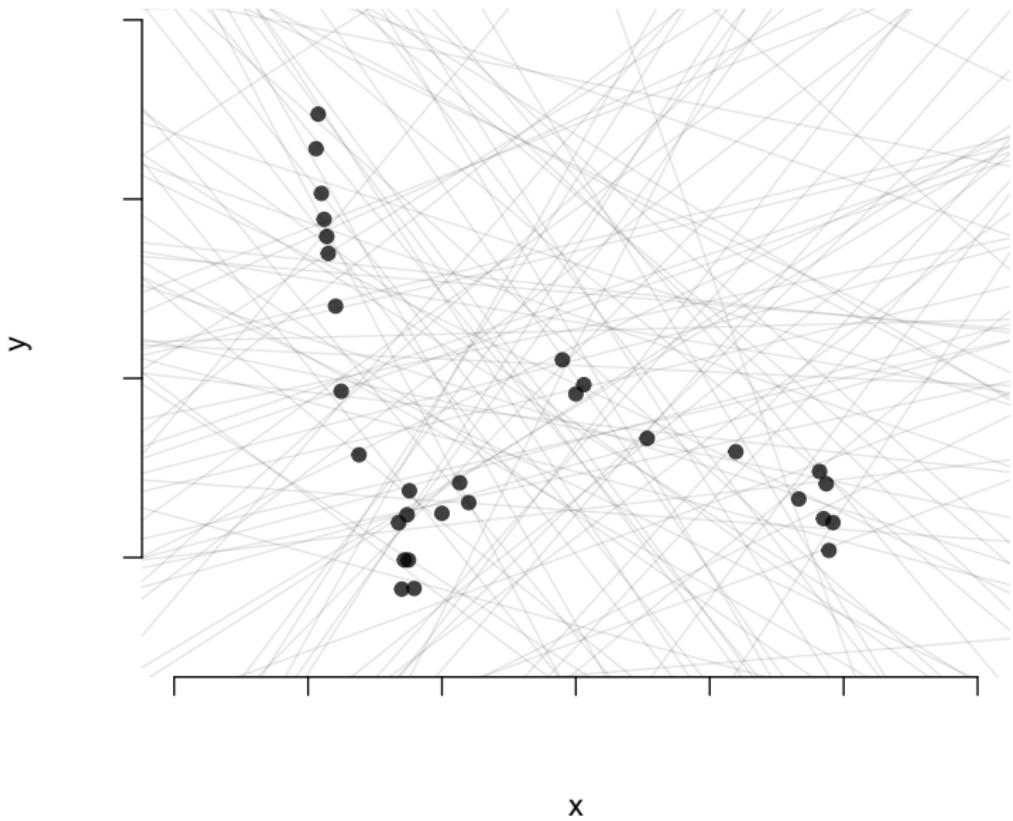
30 observations



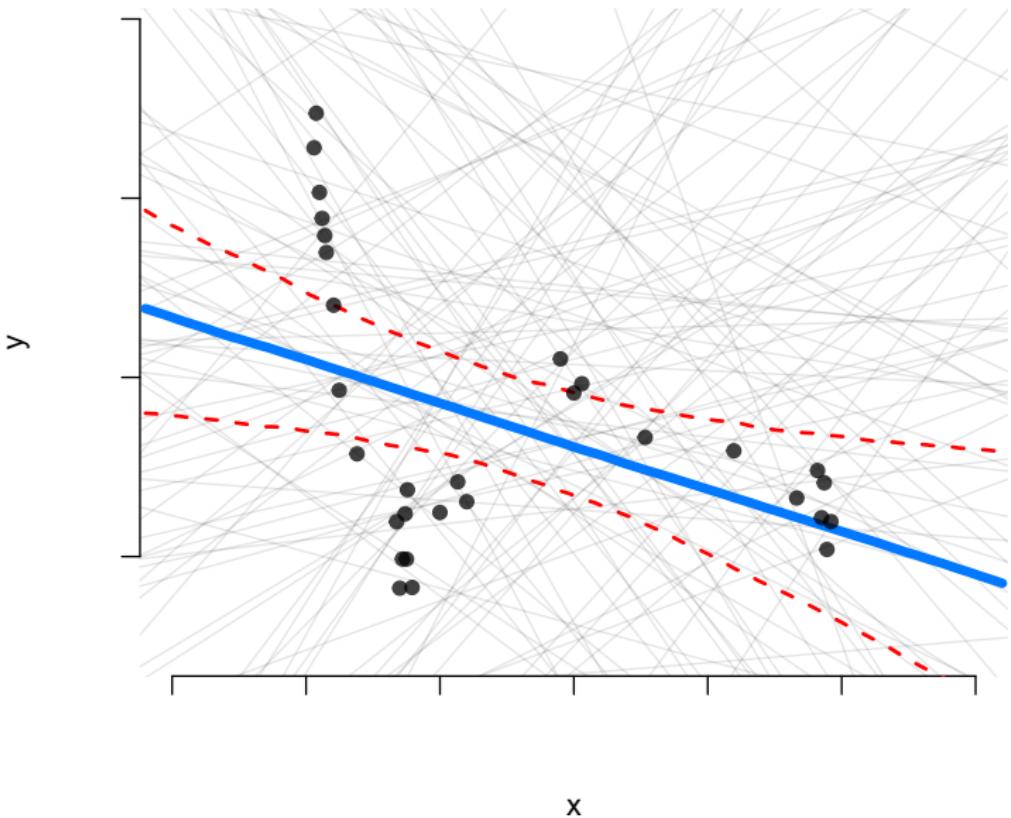
Prior beliefs



Priors with data



Priors with data → new understanding



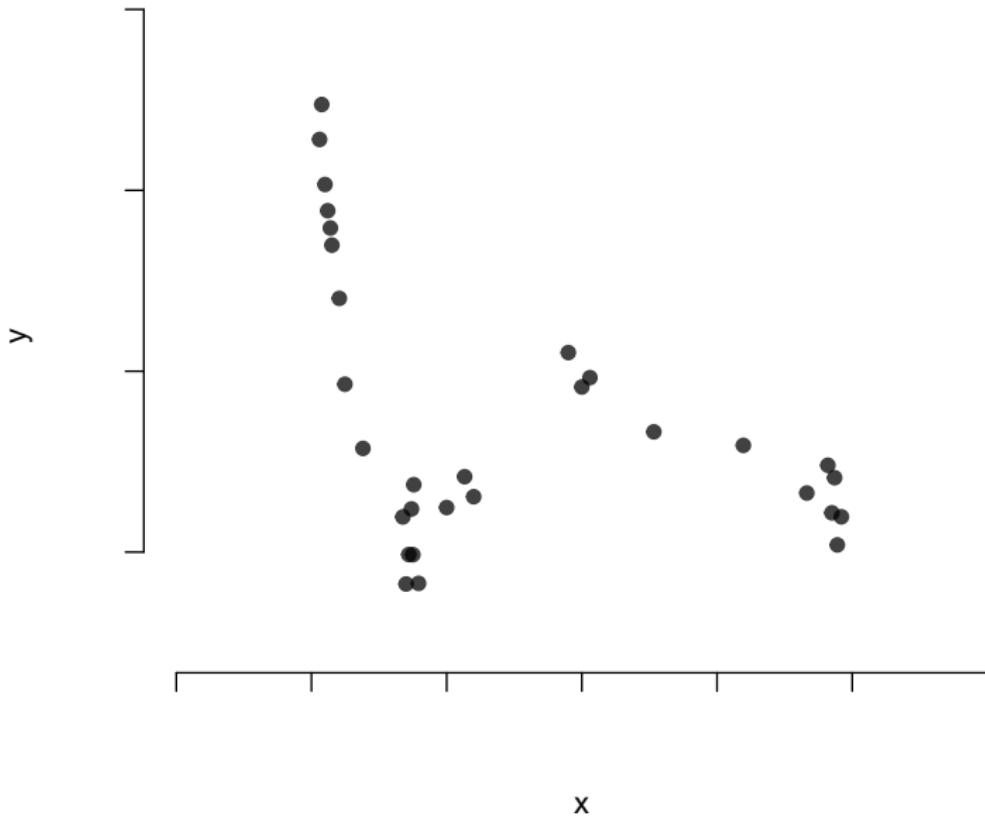
Bayes' theorem

$$\underbrace{p(\theta \mid y)}_{\text{Posterior}} \propto \underbrace{p(y \mid \theta)}_{\text{Likelihood}} \underbrace{p(\theta)}_{\text{Prior}}$$

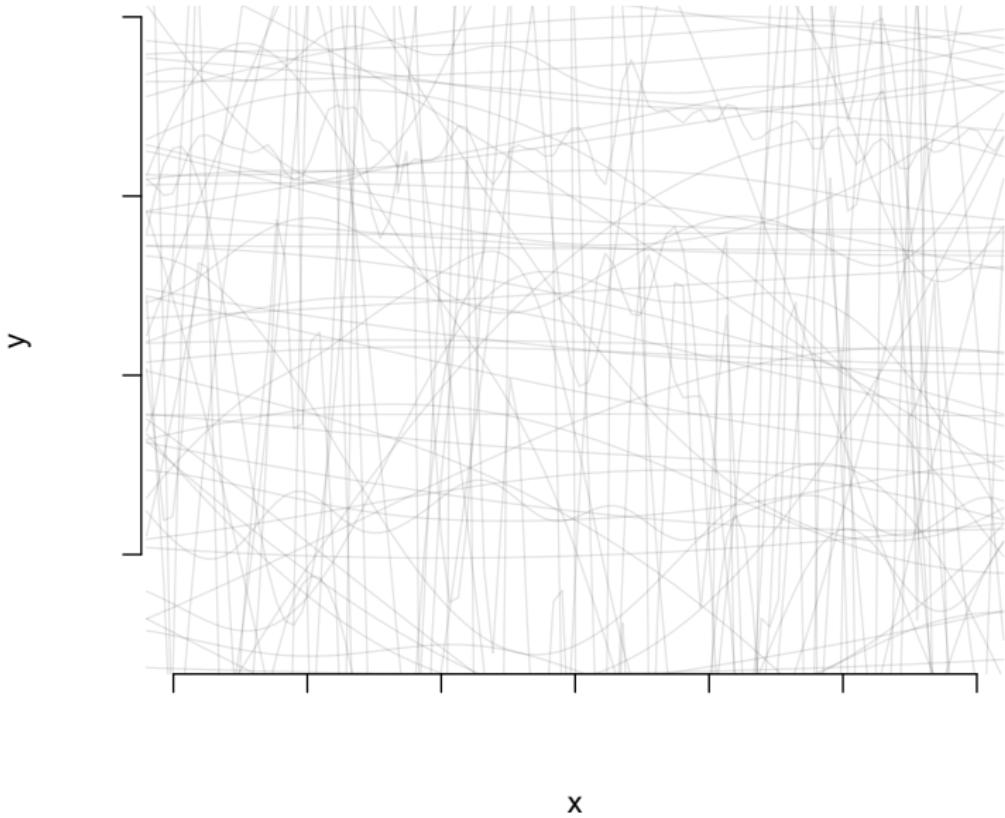
θ : unknown quantities

y : data

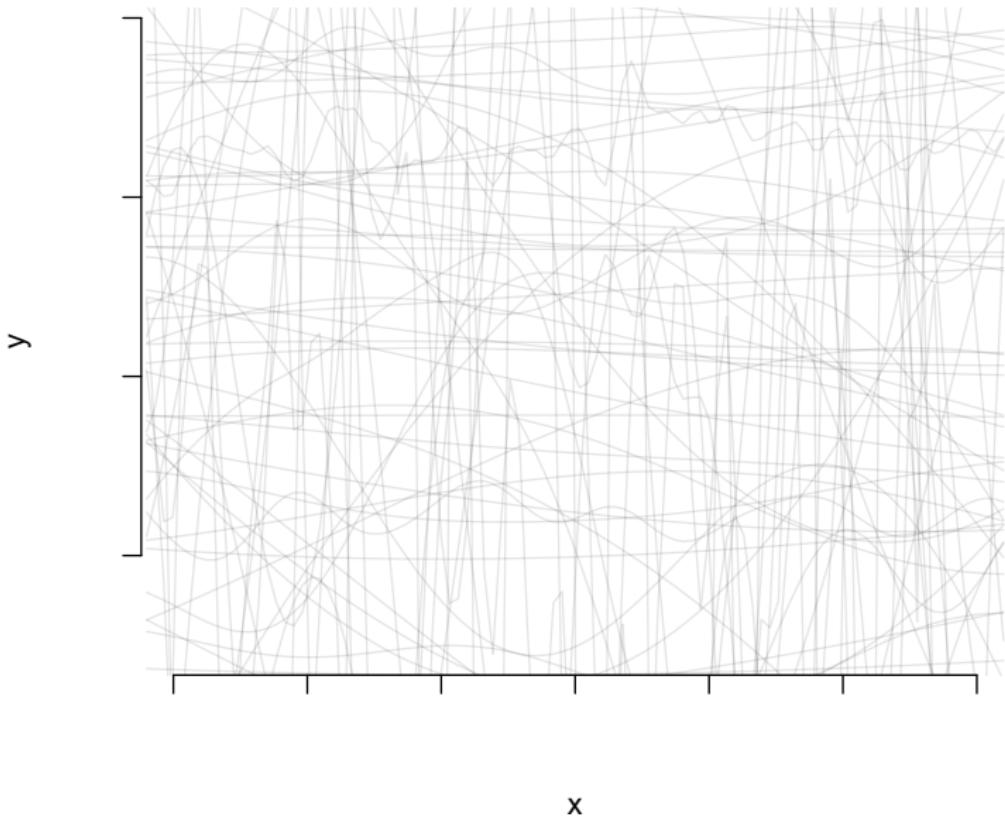
The unknown function problem



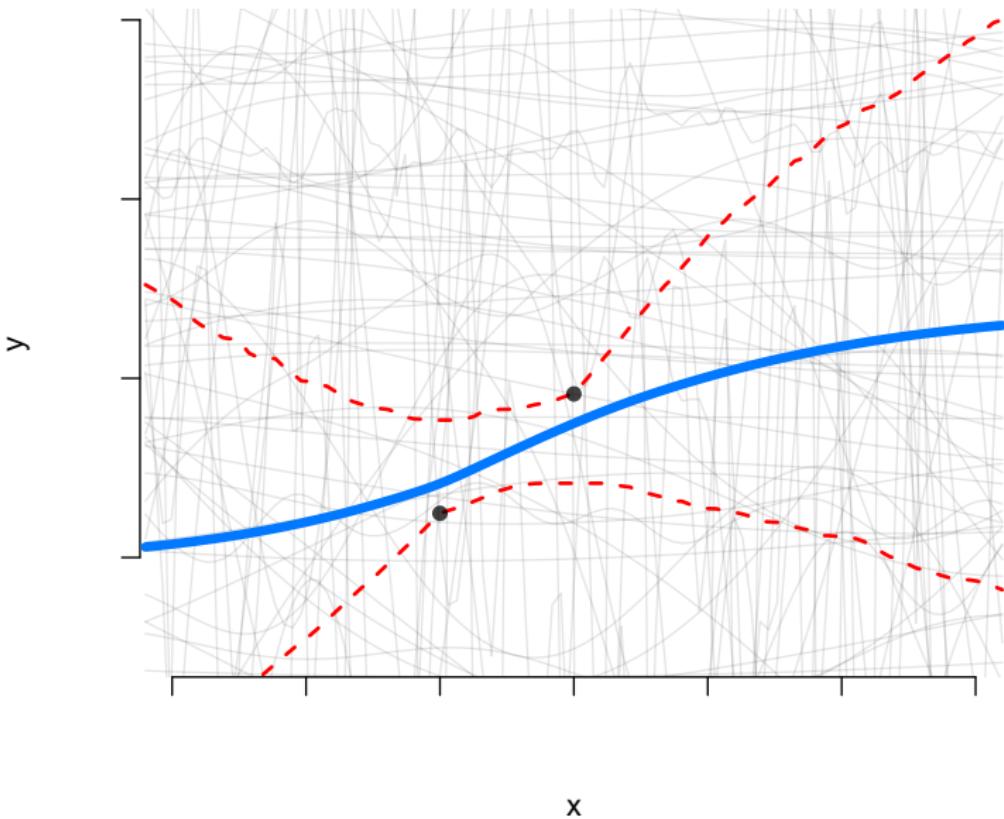
Prior for an unknown function



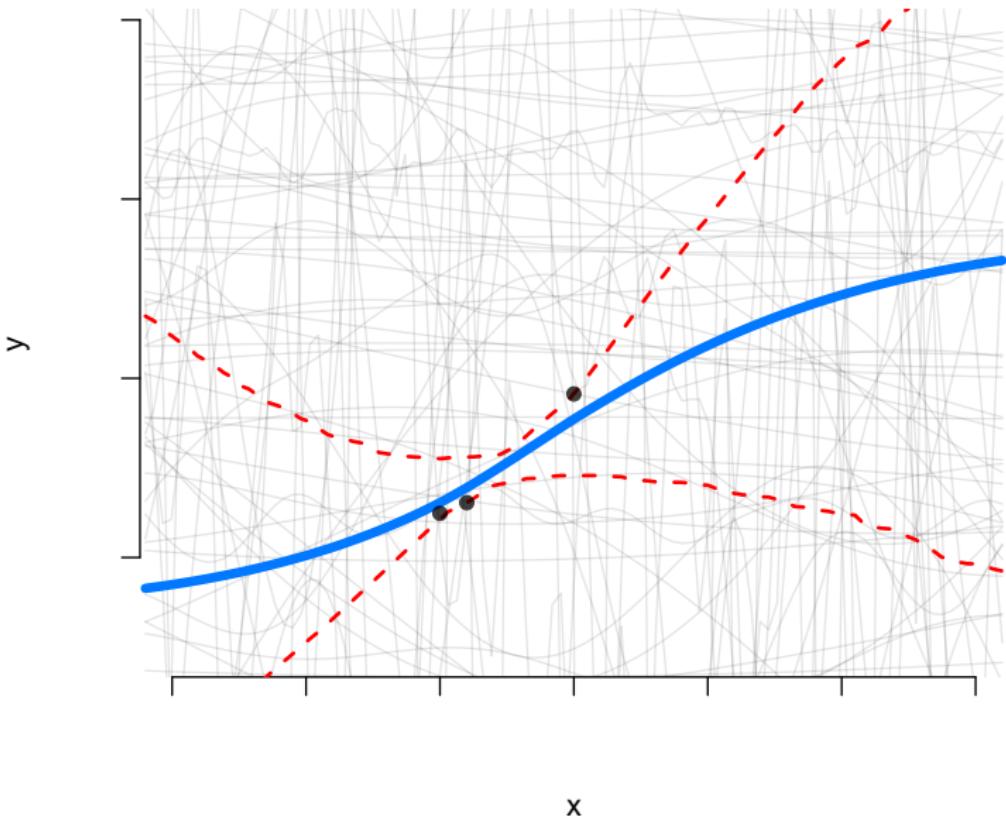
Gaussian processes: priors for unknown functions



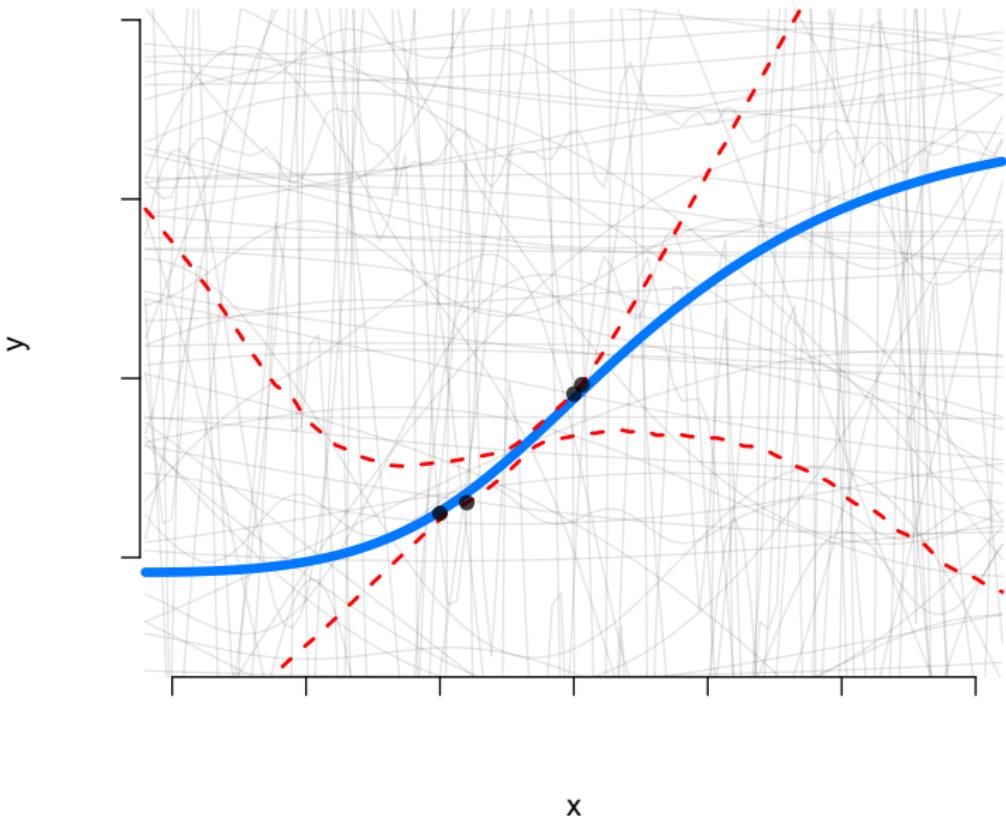
2 observations



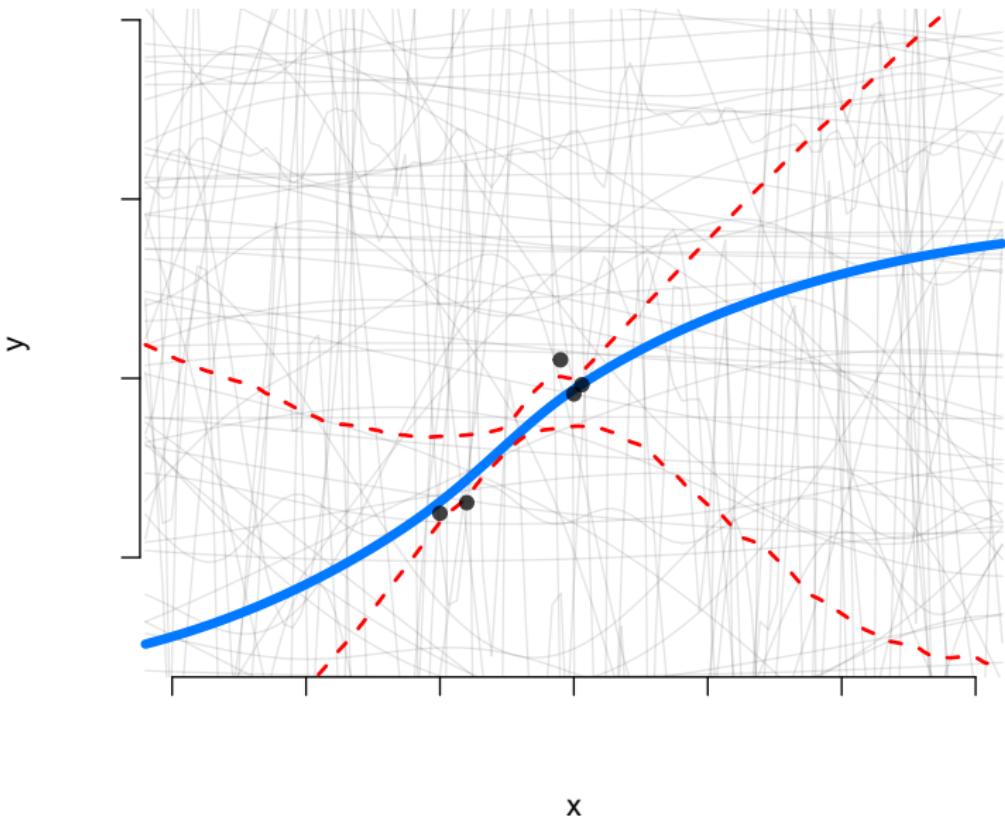
3 observations



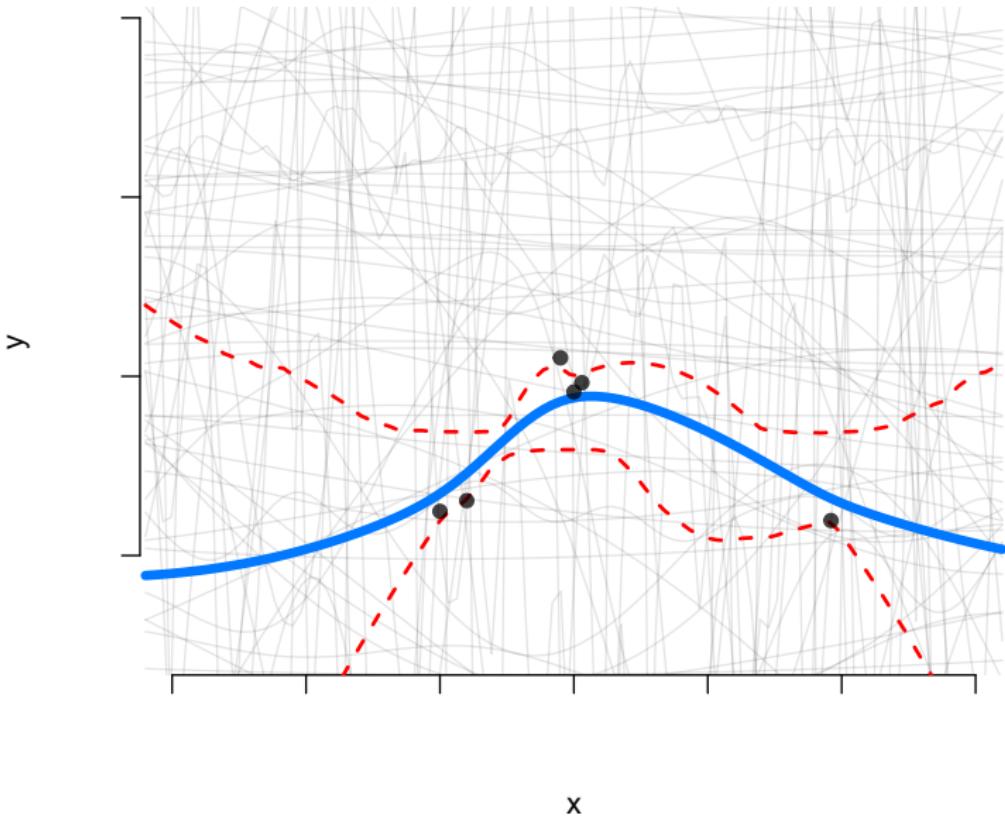
4 observations



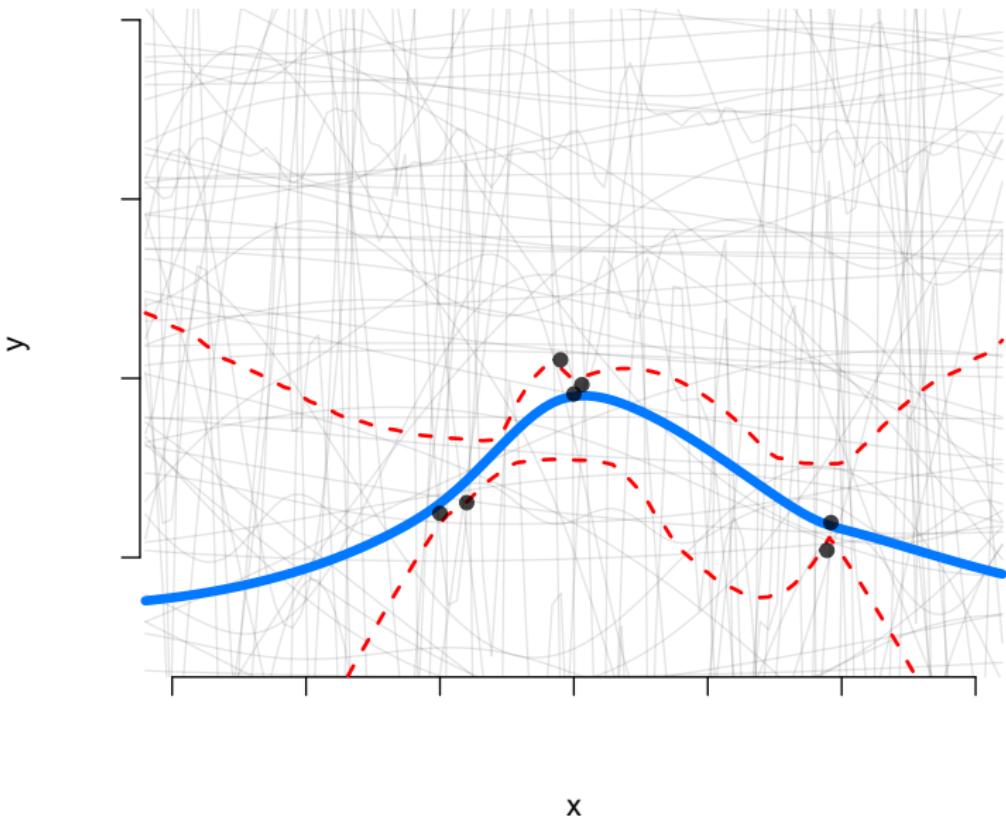
5 observations



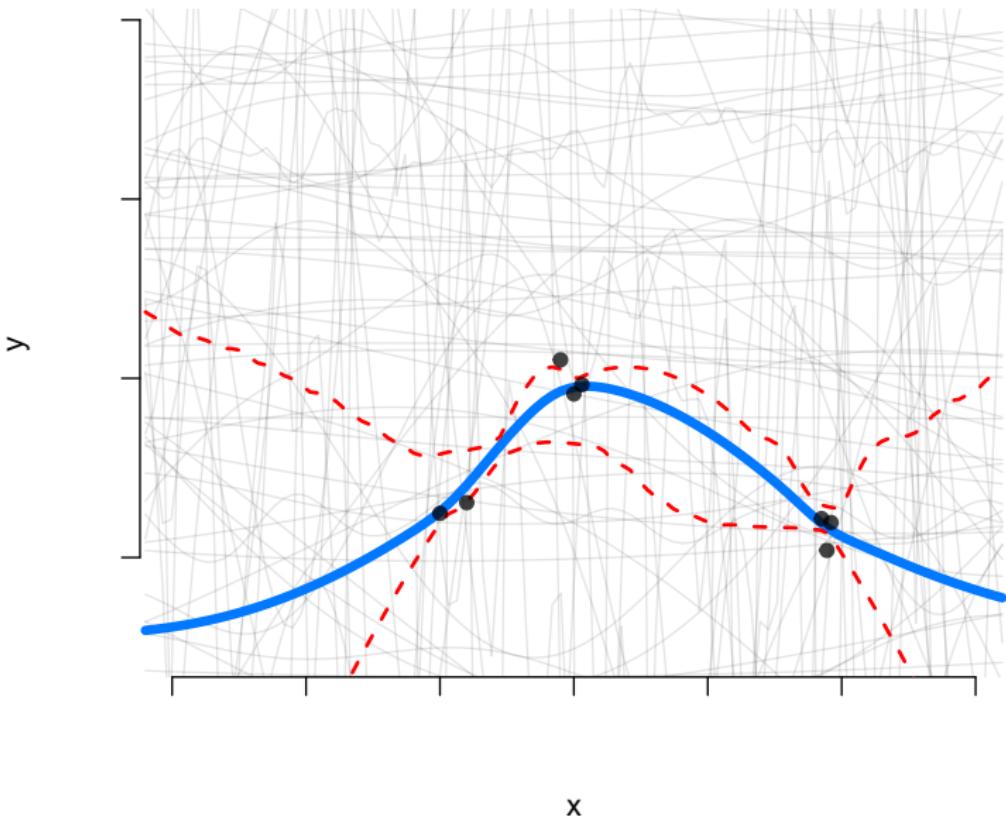
6 observations



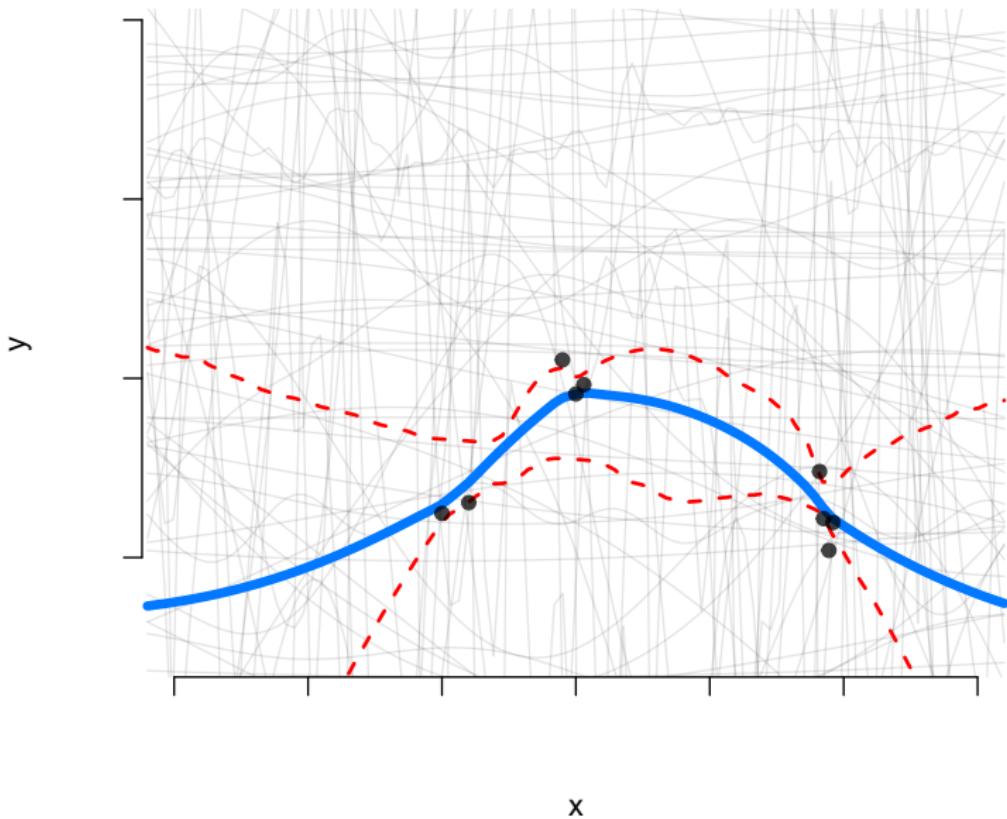
7 observations



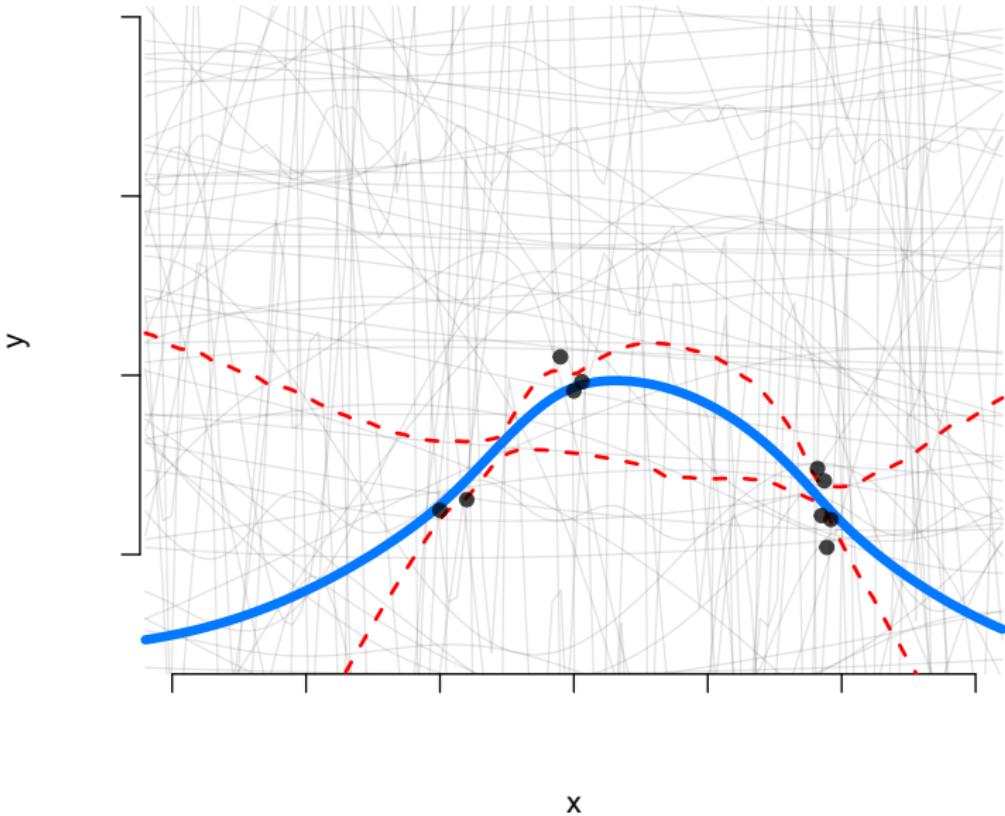
8 observations



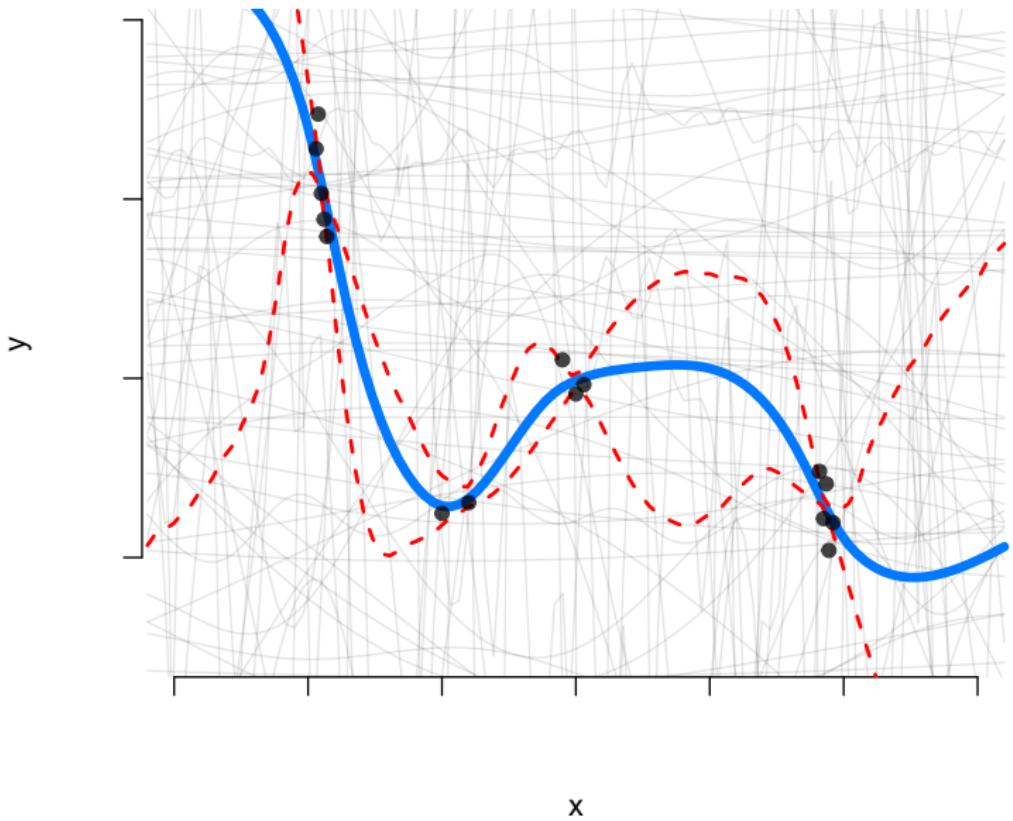
9 observations



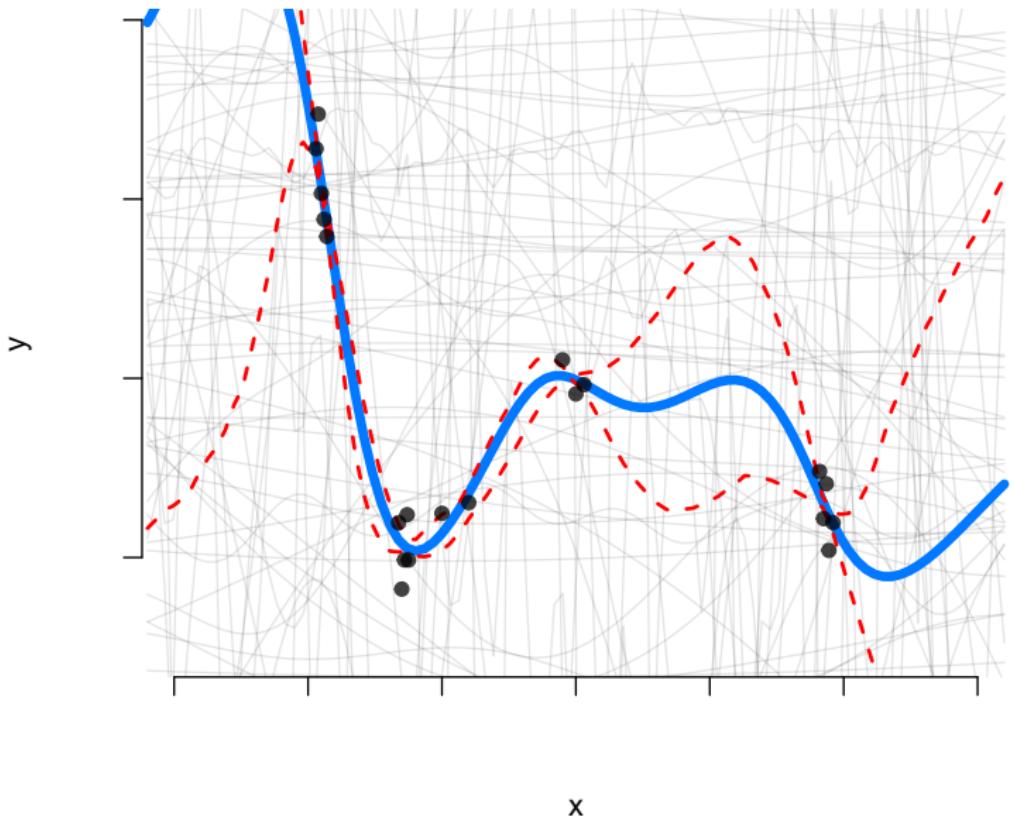
10 observations



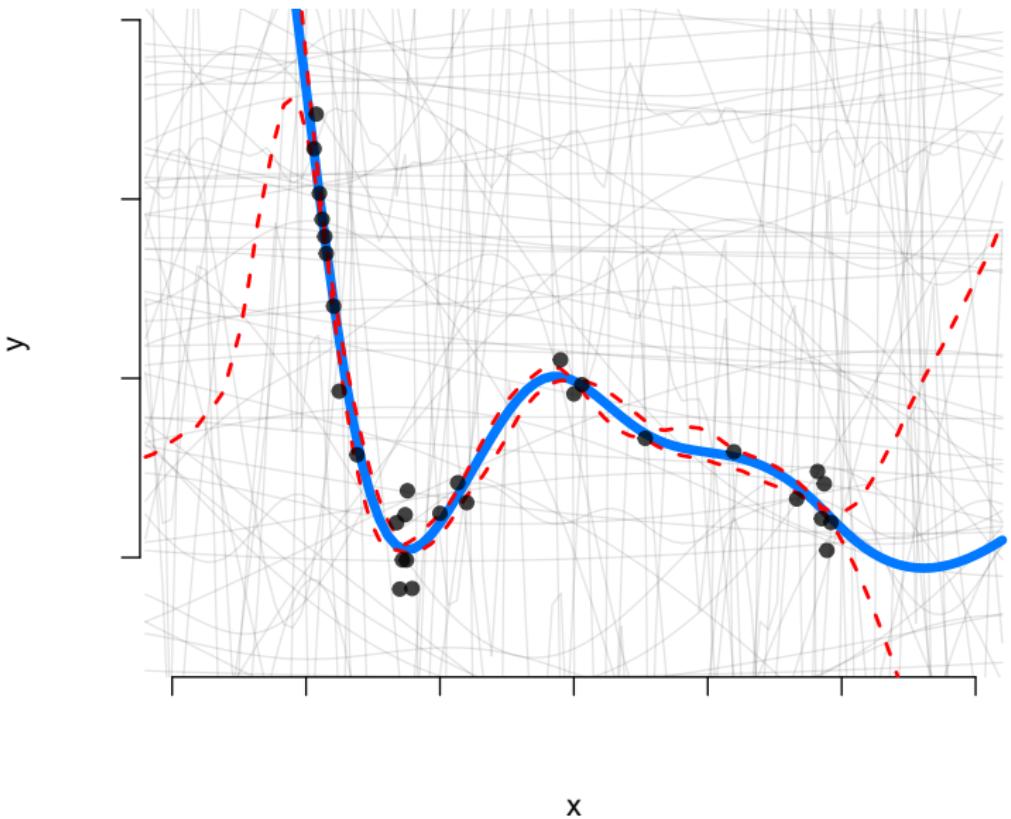
15 observations



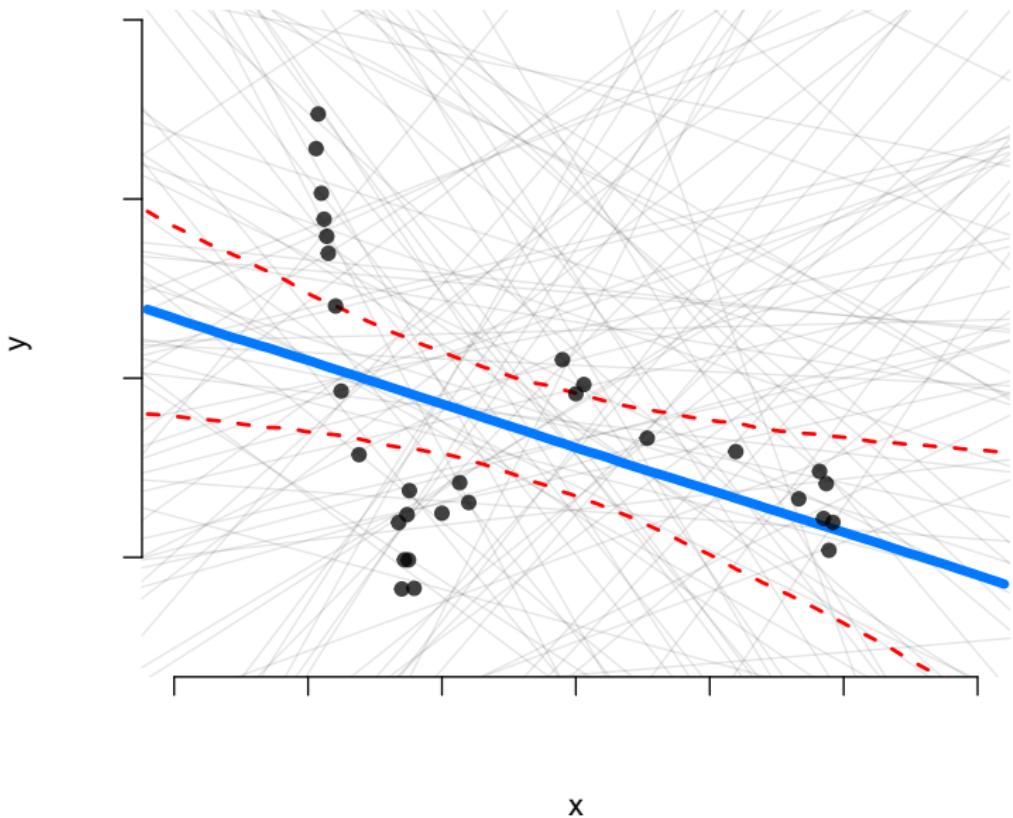
20 observations



30 observations

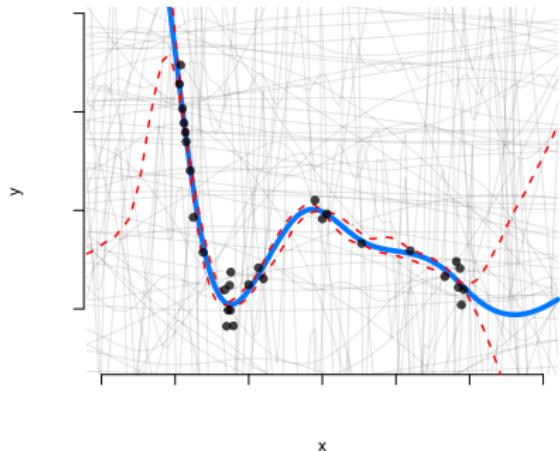
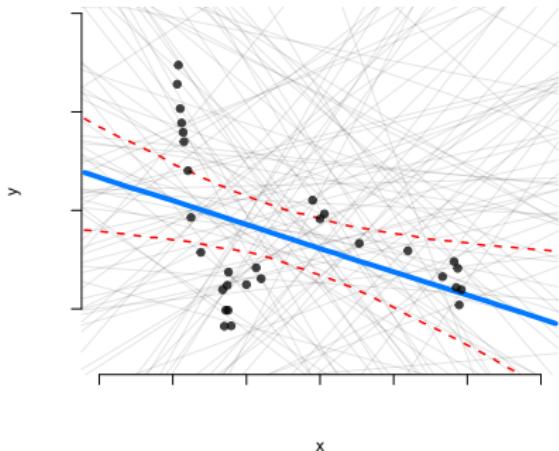


Linear regression: functional form known



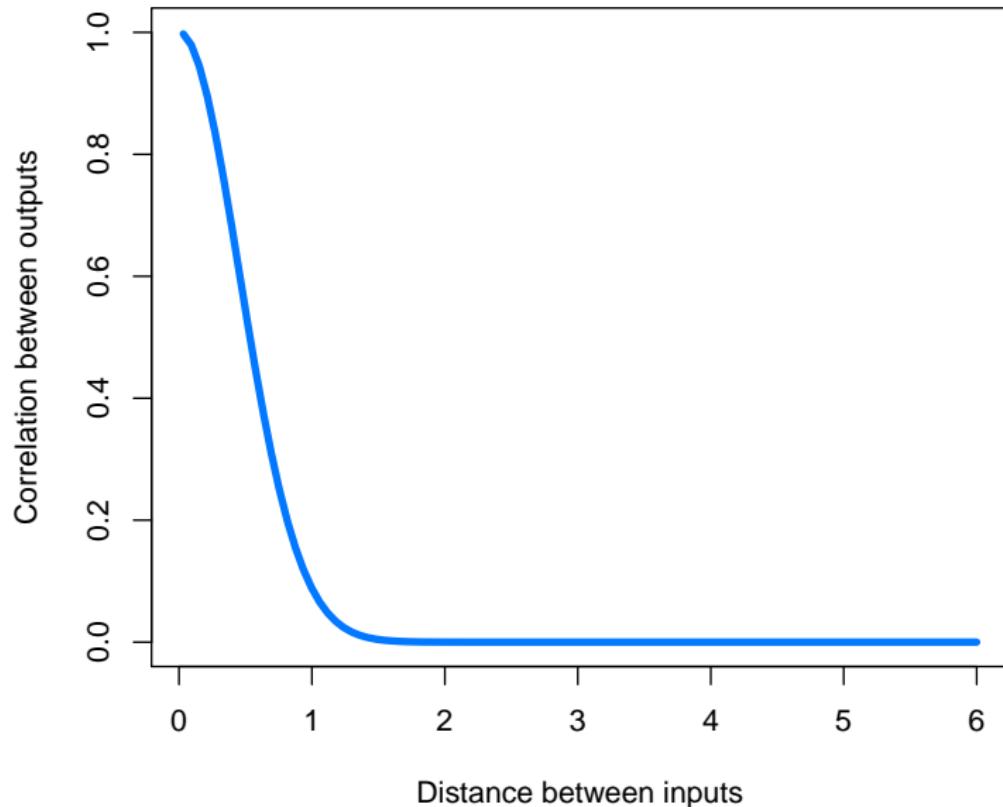
Linear regression: functional form known

Gaussian process regression: functional form unknown

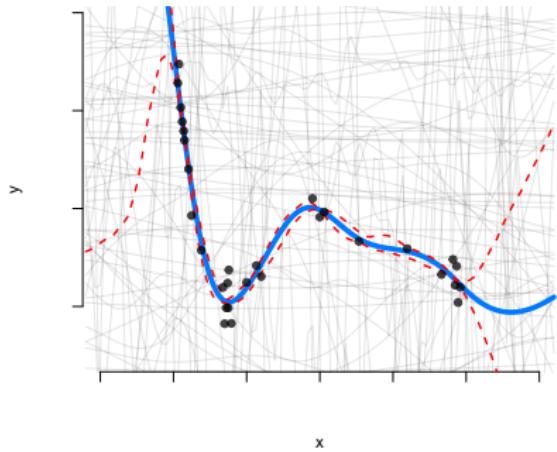
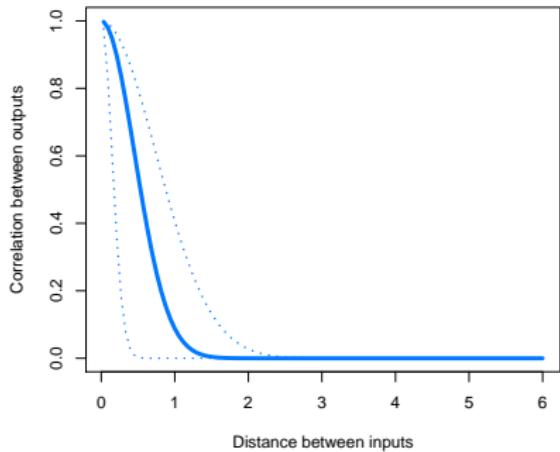


Correlation function

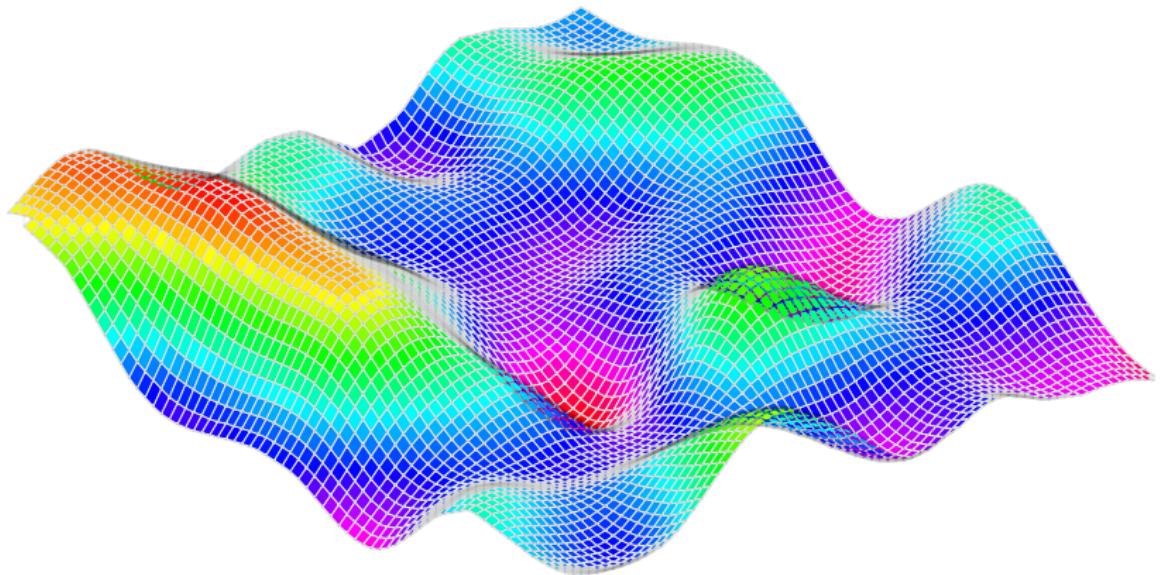
Nearby inputs produce similar outputs



Correlation function

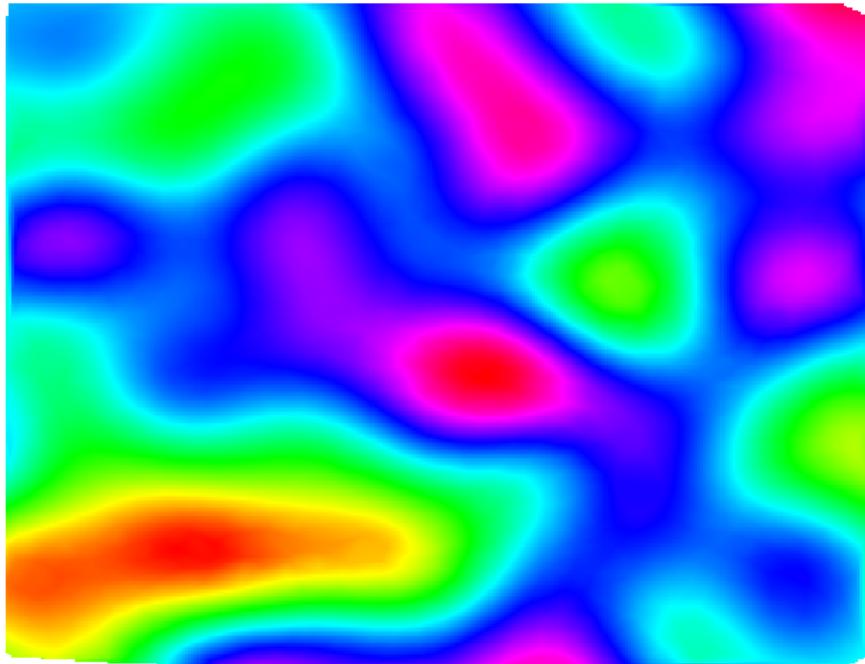


Gaussian processes in space



Gaussian processes in space

x_2

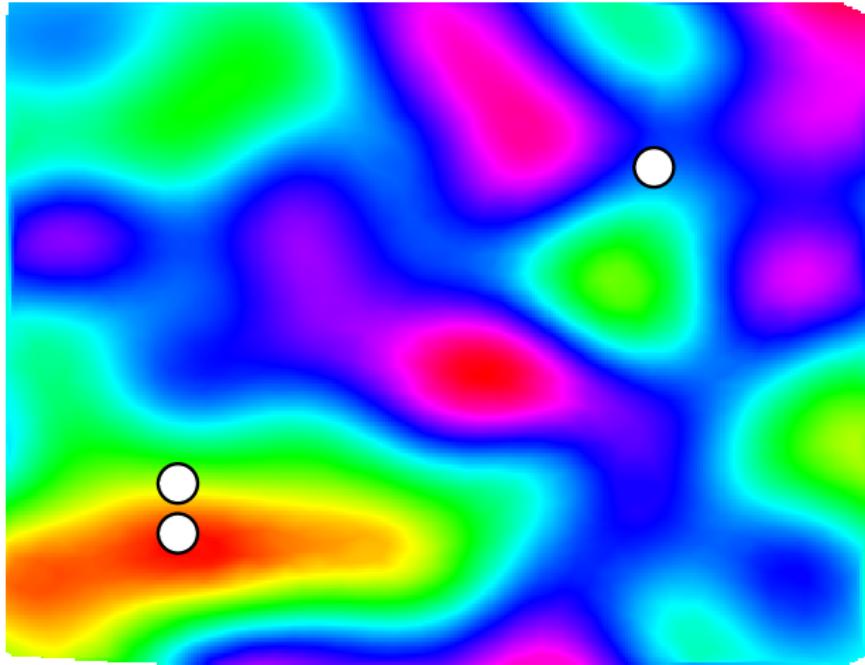


x_1

Nearby inputs → similar outputs

x_2

x_1



Two applications of Gaussian processes

1. Predicting missing species traits
2. Understanding spatiotemporal disease dynamics

Why care about missing species traits?

1. Traits are important for extinction risk



J. D'Orbigny

Tudor imp.

CHIROGALEUS MILII.

Why care about missing species traits?

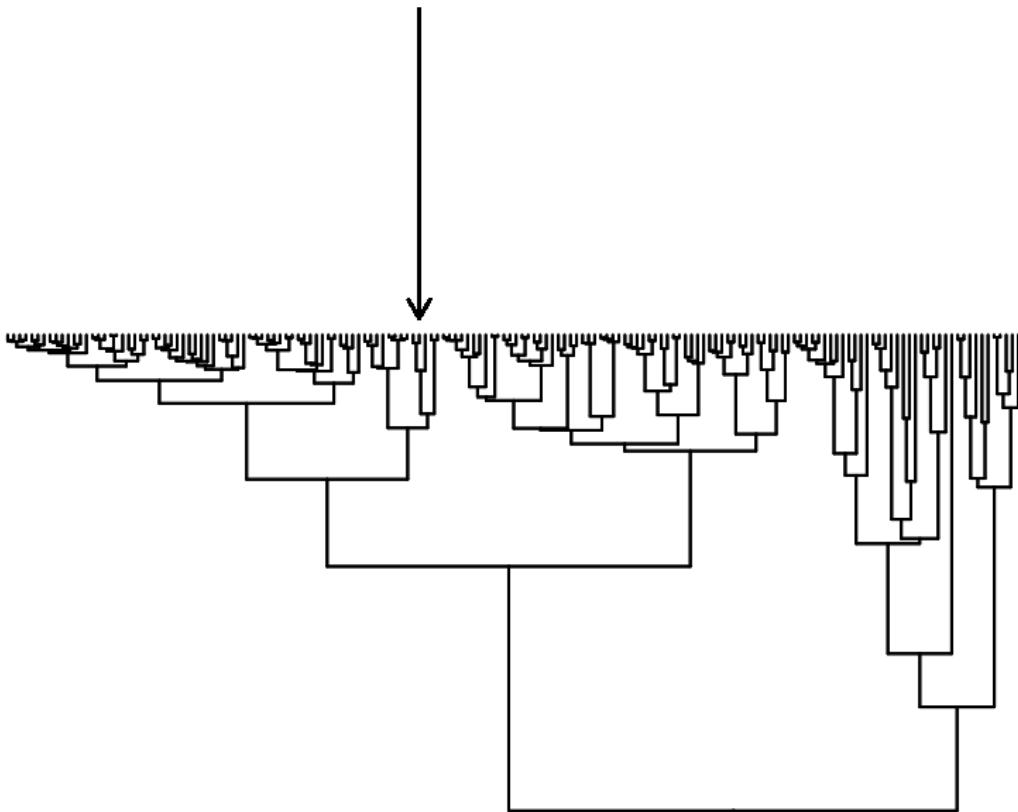
1. Traits are important for extinction risk
2. Data subsetting reduces power

Useful fact: relatives look alike

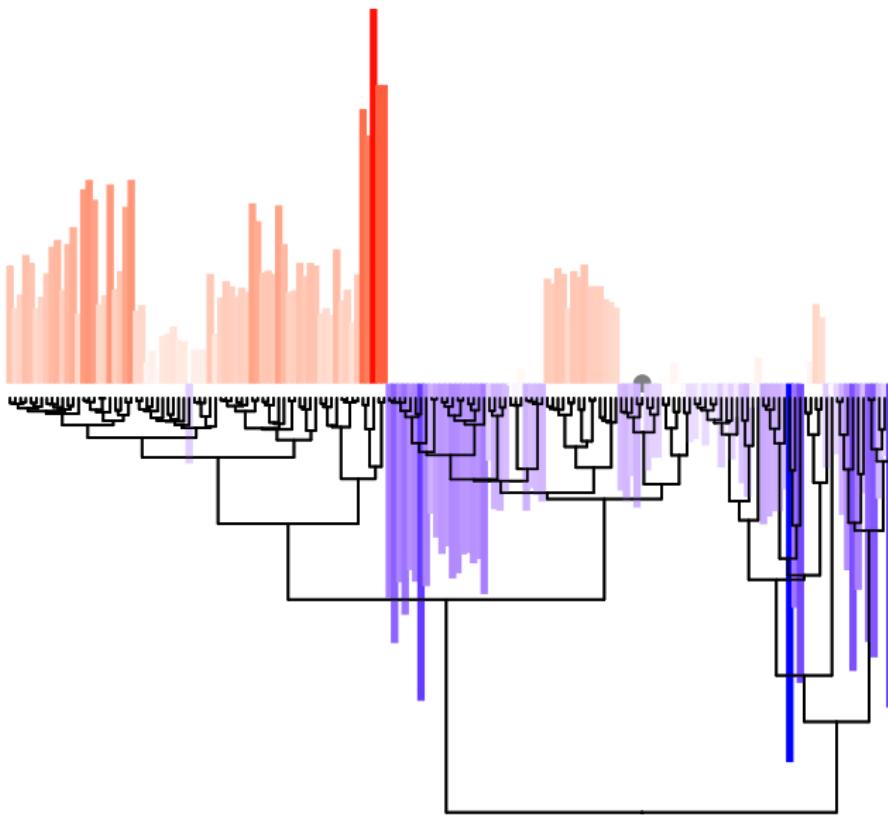




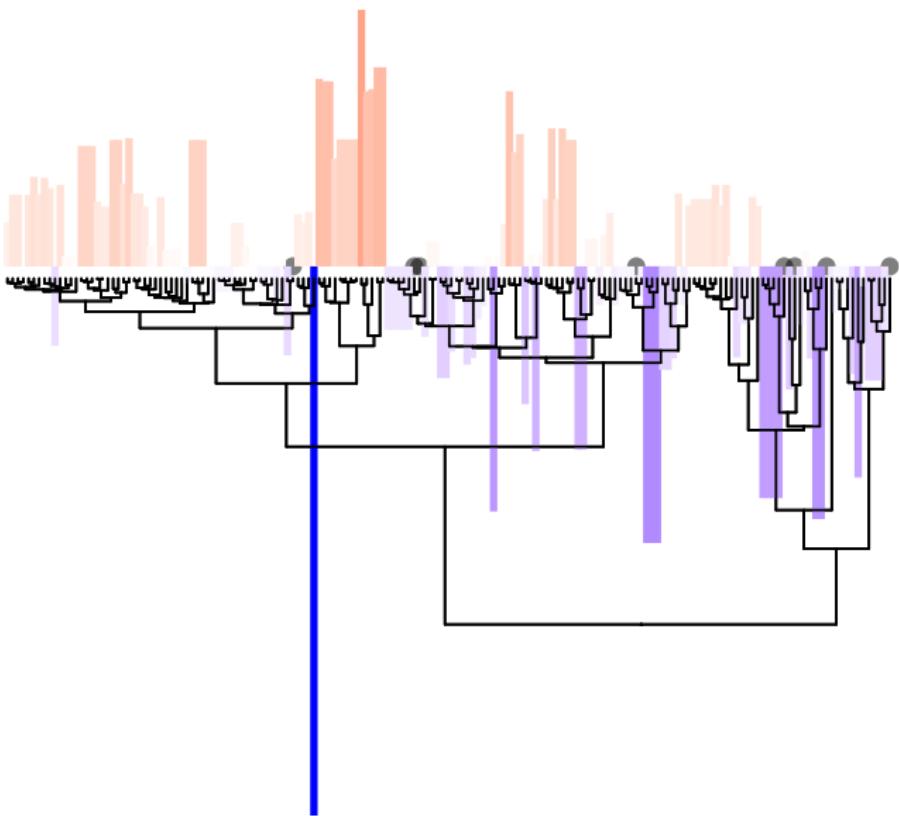
You are here



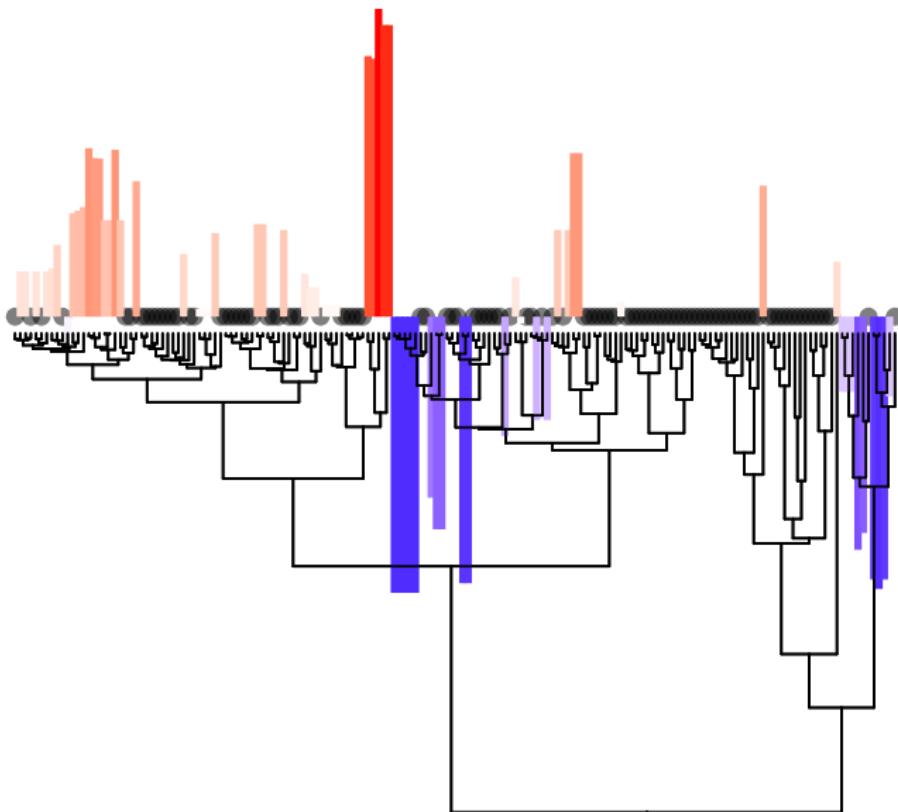
Primate body mass



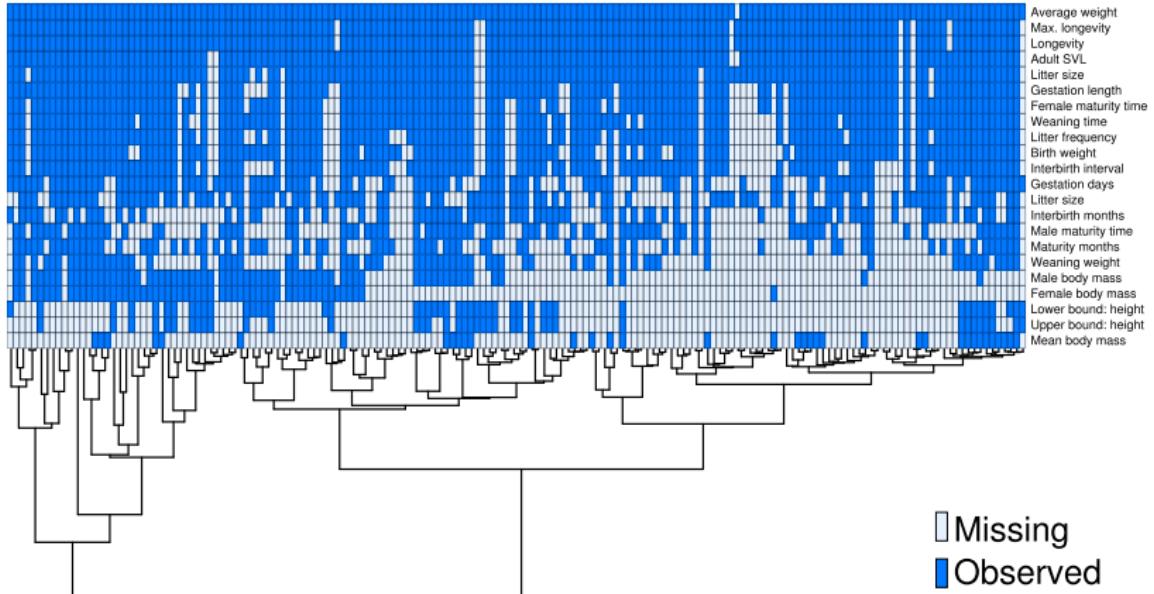
Primate longevity



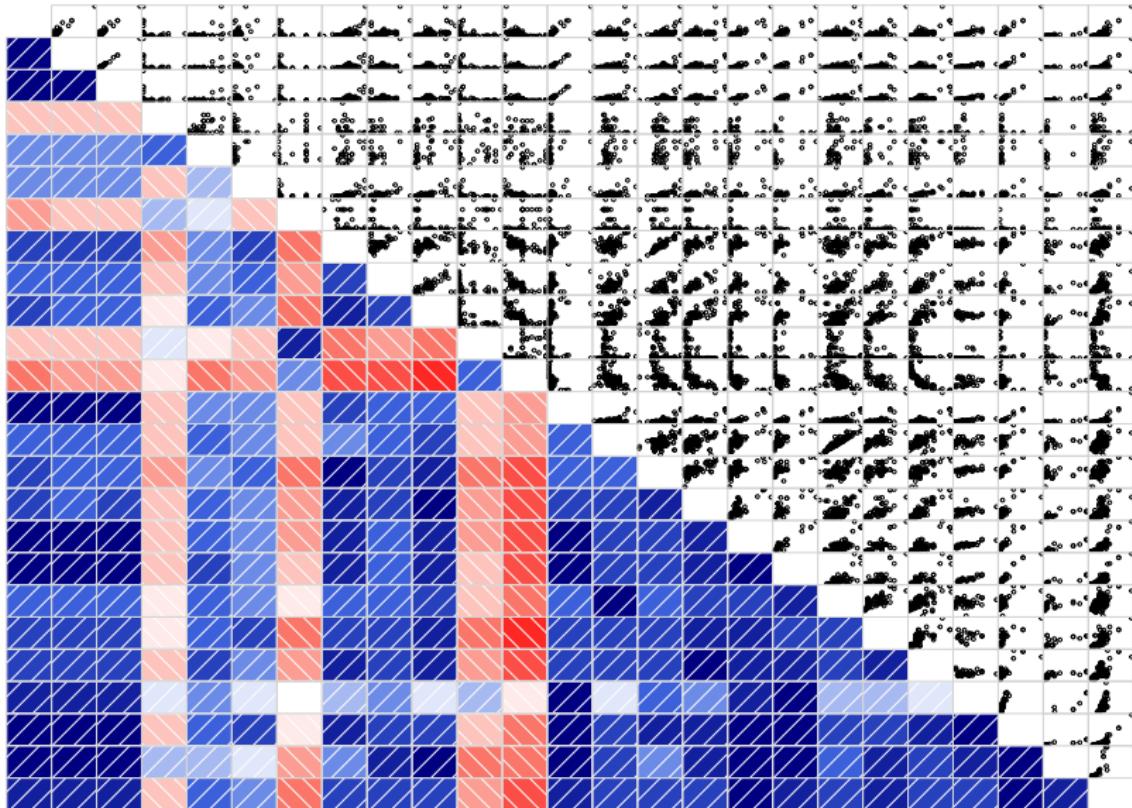
Primate weaning weight



Missing traits



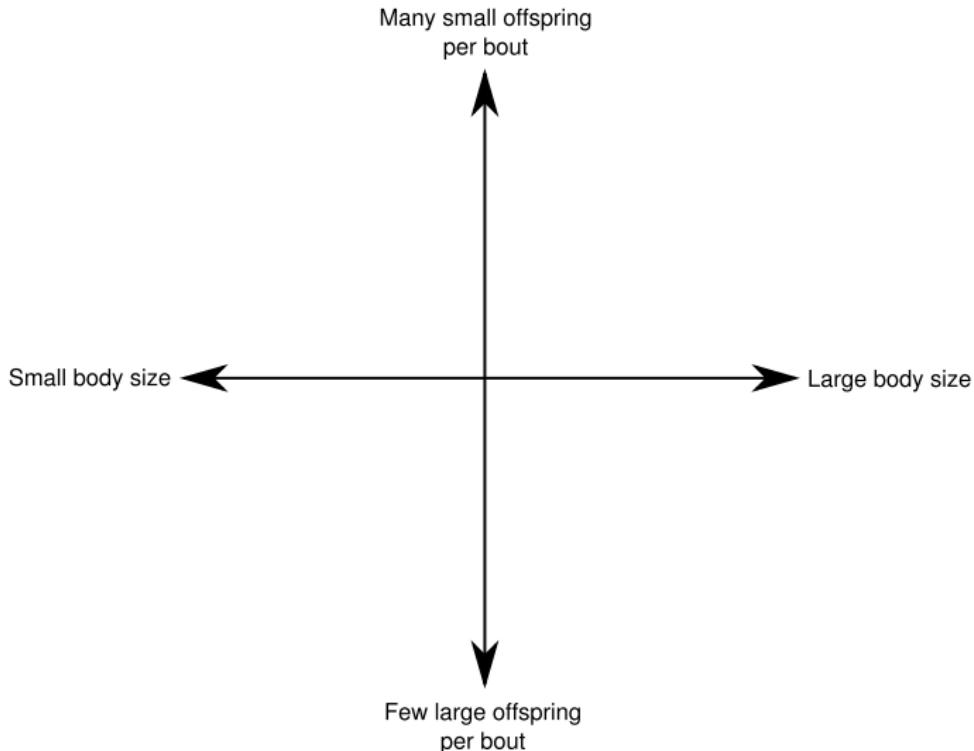
Observed traits



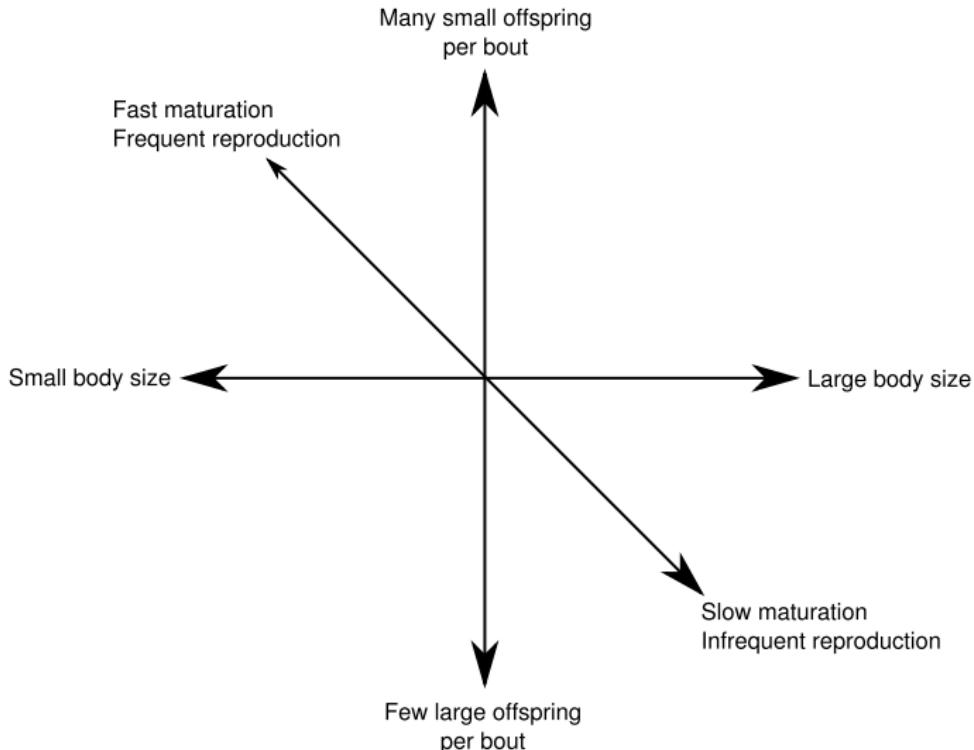
Dimension reduction and the fast-slow continuum



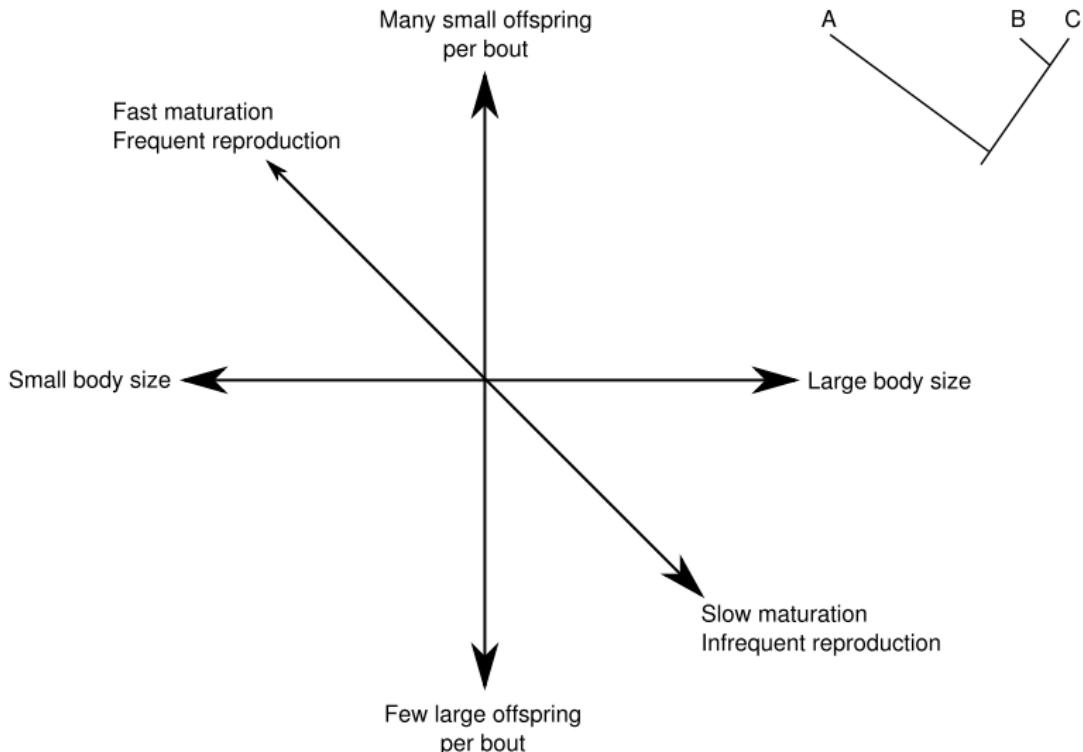
2 dimensional view



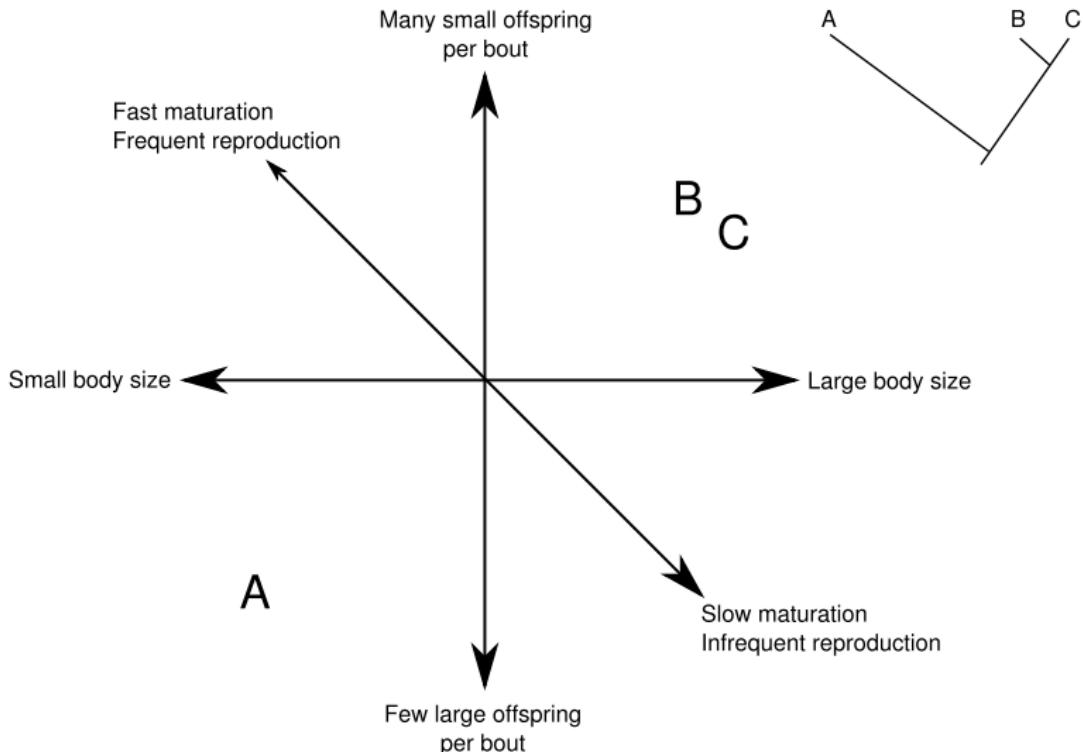
3 dimensional view



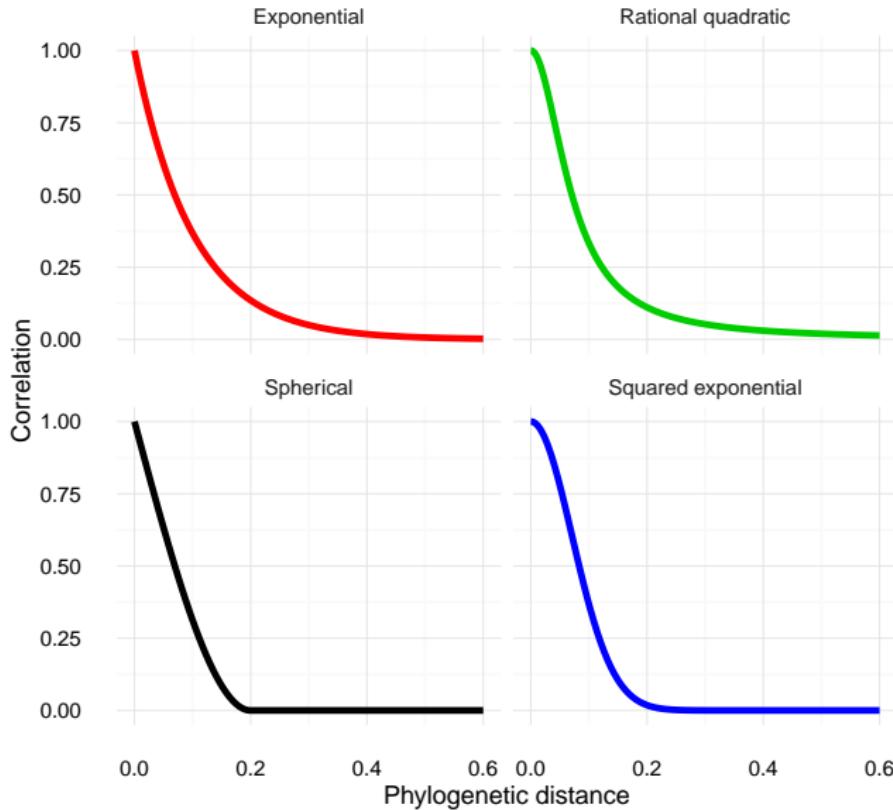
Placing species in latent factor space



Placing species in latent factor space



Phylogenetic distance and latent factor correlations



Questions

1. Can we predict traits?
2. Is phylogeny useful?
3. How many latent dimensions?

Primate data summary

- ▶ 22 traits
- ▶ 168 species
- ▶ 2351 of 3696 potential observations
- ▶ 36.4% missing

Model structure

$$Y = \Lambda F + \epsilon$$

Y : trait matrix

Λ : loading matrix

F : latent factors

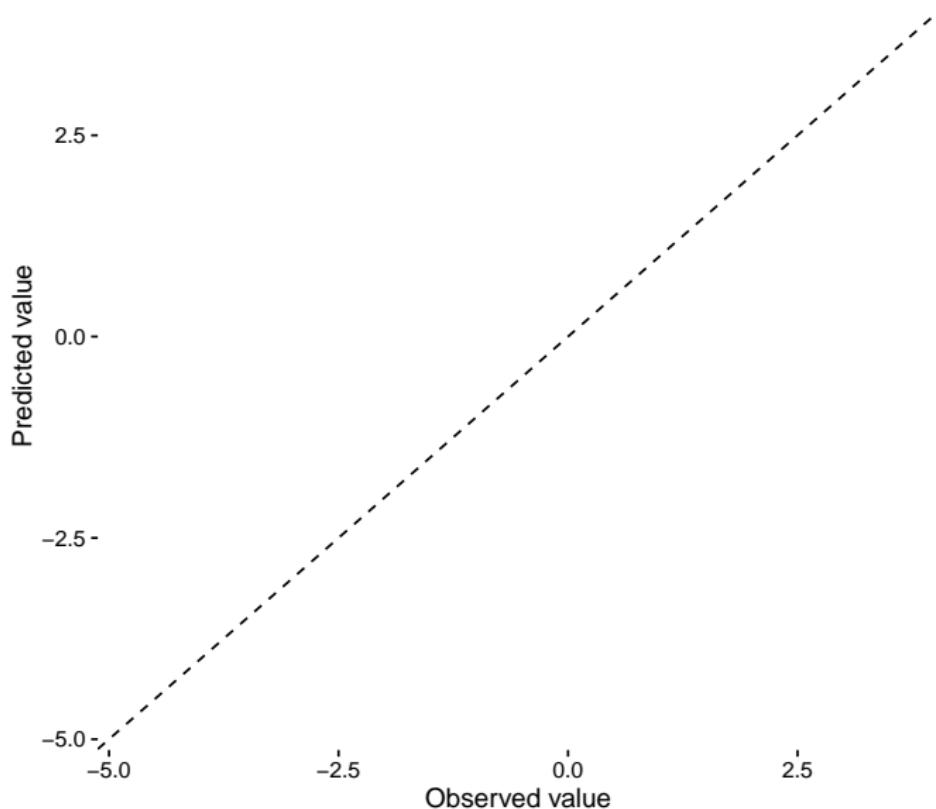
ϵ : unexplained variation

Model comparisons

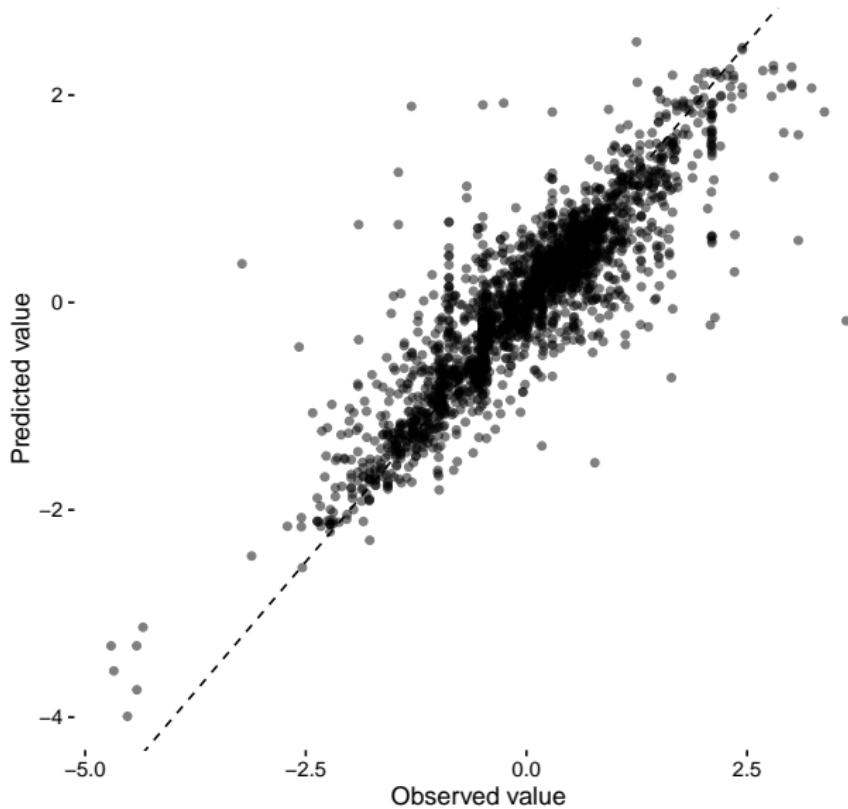
Model set varies:

- ▶ latent factor dimensionality
- ▶ correlation functions

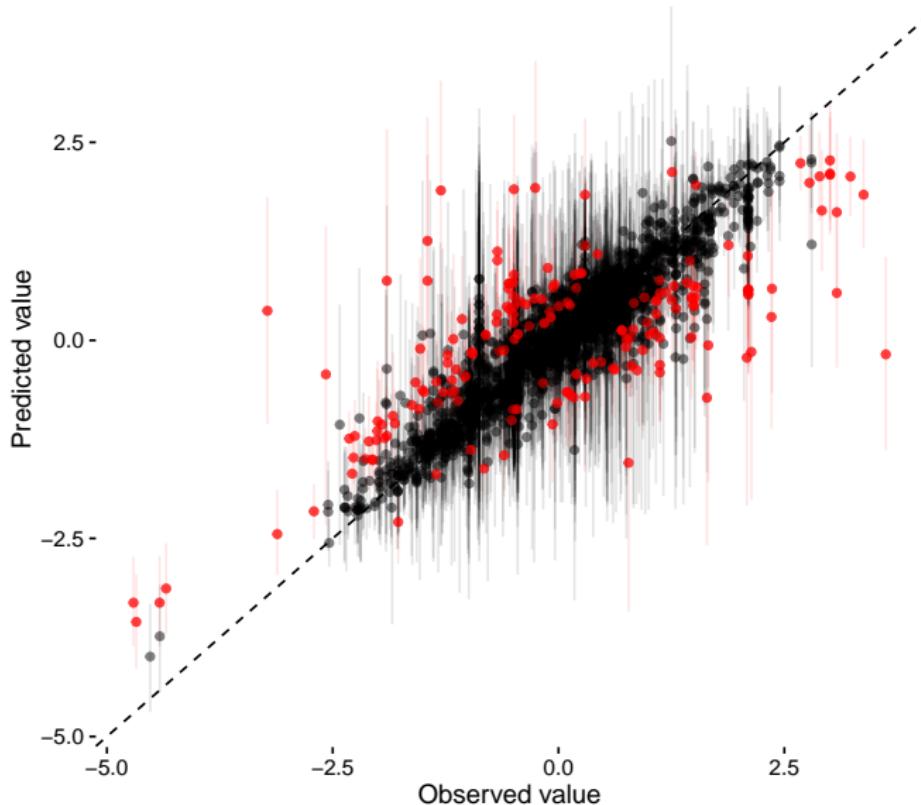
Results



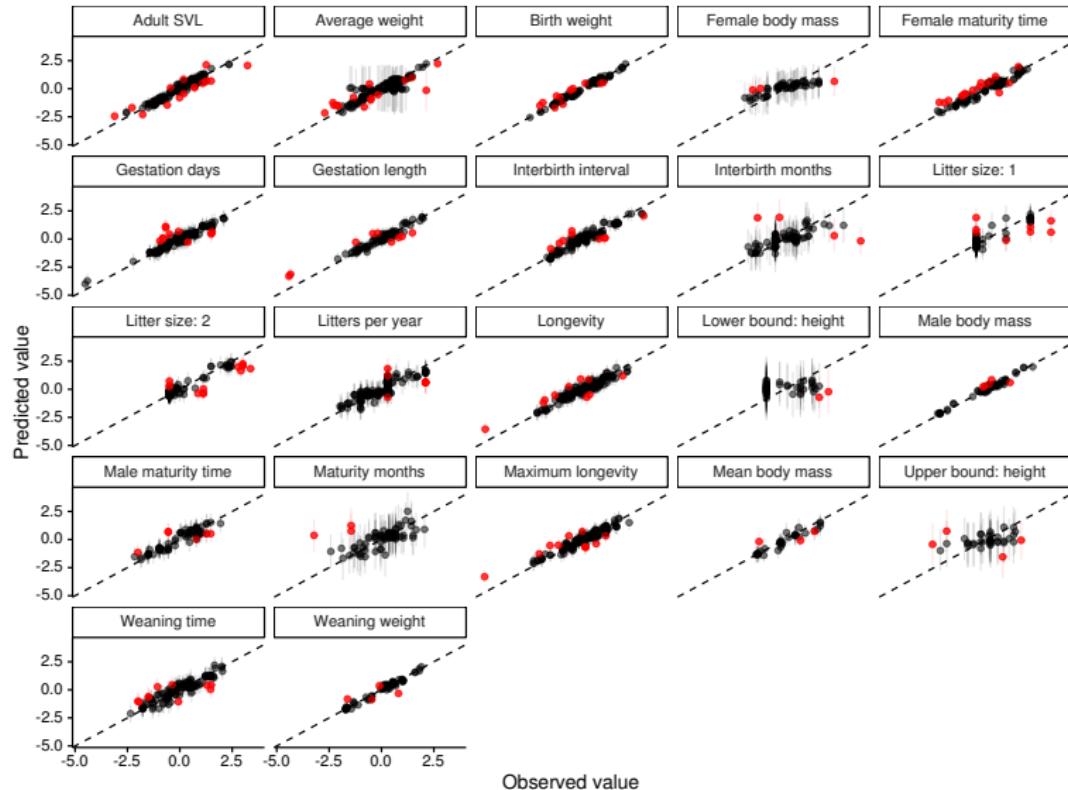
Results



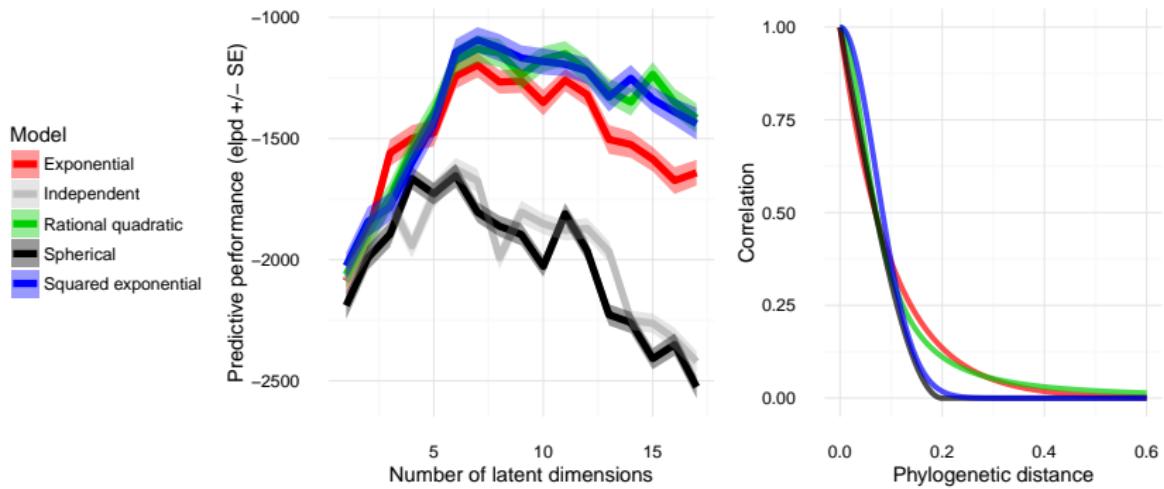
Results



Results by trait



How many latent factors?



Questions

1. Can we predict traits? Yes
2. Is phylogeny useful? Yes
3. How many latent dimensions? ≈ 7

Conservation in the face of uncertainty



J. G. Keulemans

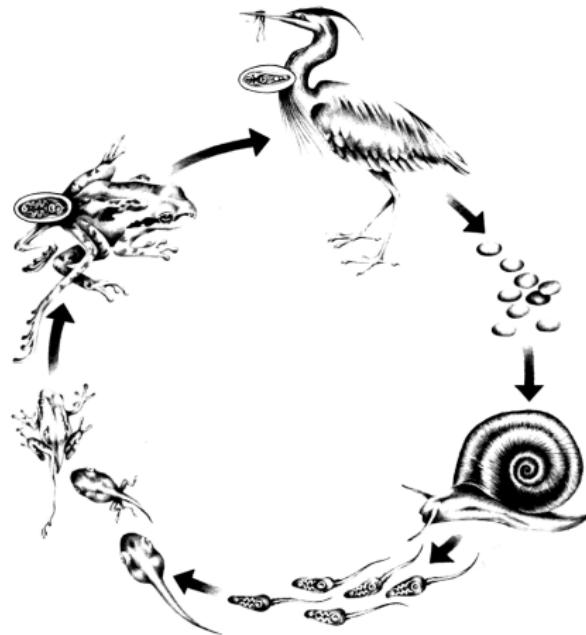
R. E. M. Bakkerart imp.

CHIROGALEUS MILII.

Outline

1. Gaussian processes for missing species traits
2. **Gaussian processes for spatiotemporal disease dynamics**

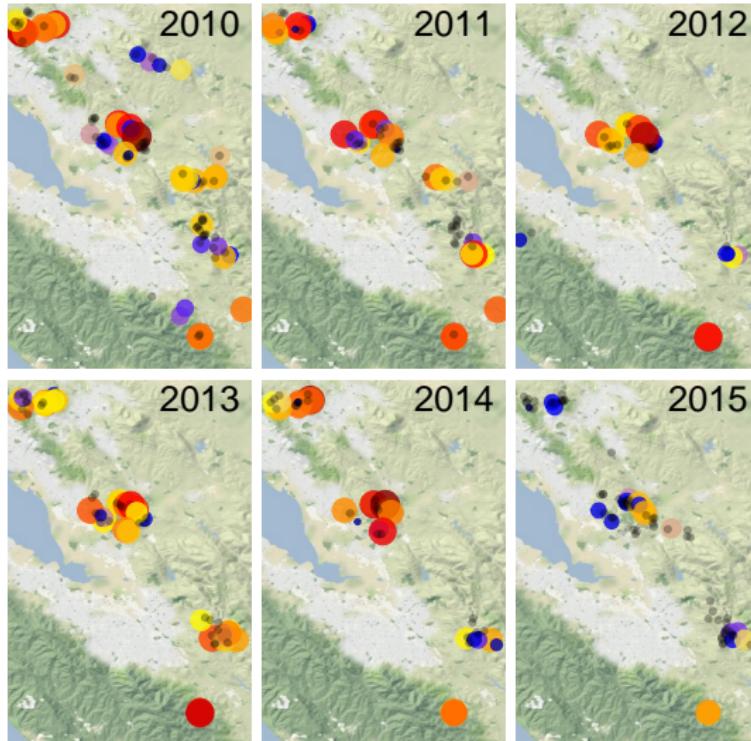
The parasite *Ribeiroia ondatrae*



The data

- ▶ 207 sampling locations from 2009 - 2015
- ▶ 6361 dip net sweeps
- ▶ 5139 frogs sampled
- ▶ 33644 snails sampled

Spatial hotspots across years



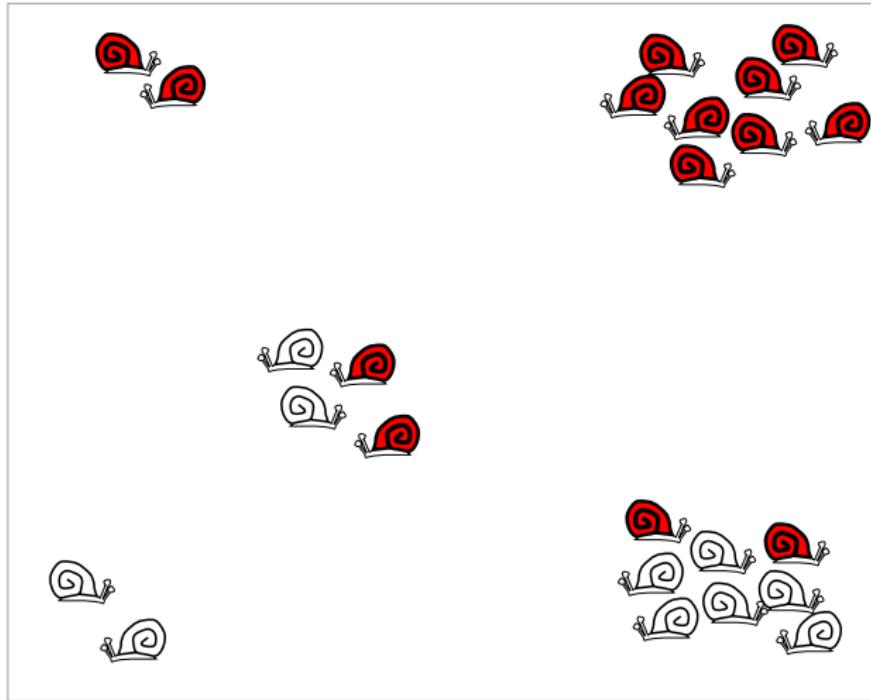
Mean infection per frog ● -2 ● 0 ● 2 ● 4

Questions

1. How can we link snail and frog infections?
2. What is the spatiotemporal nature of infection dynamics?

How can we link snail and frog infections?

Snail infection prevalence

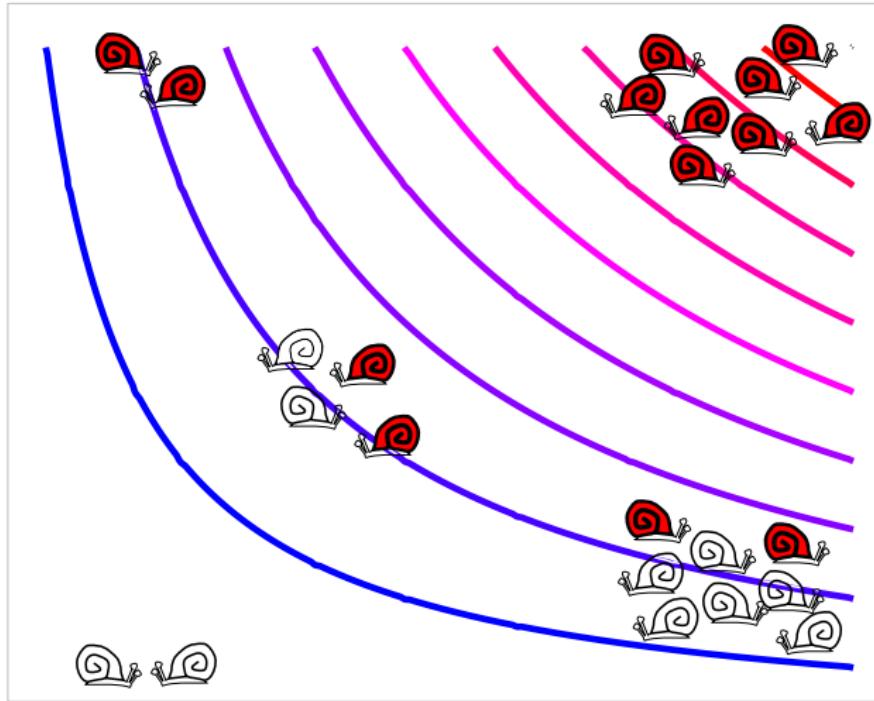


Snail density

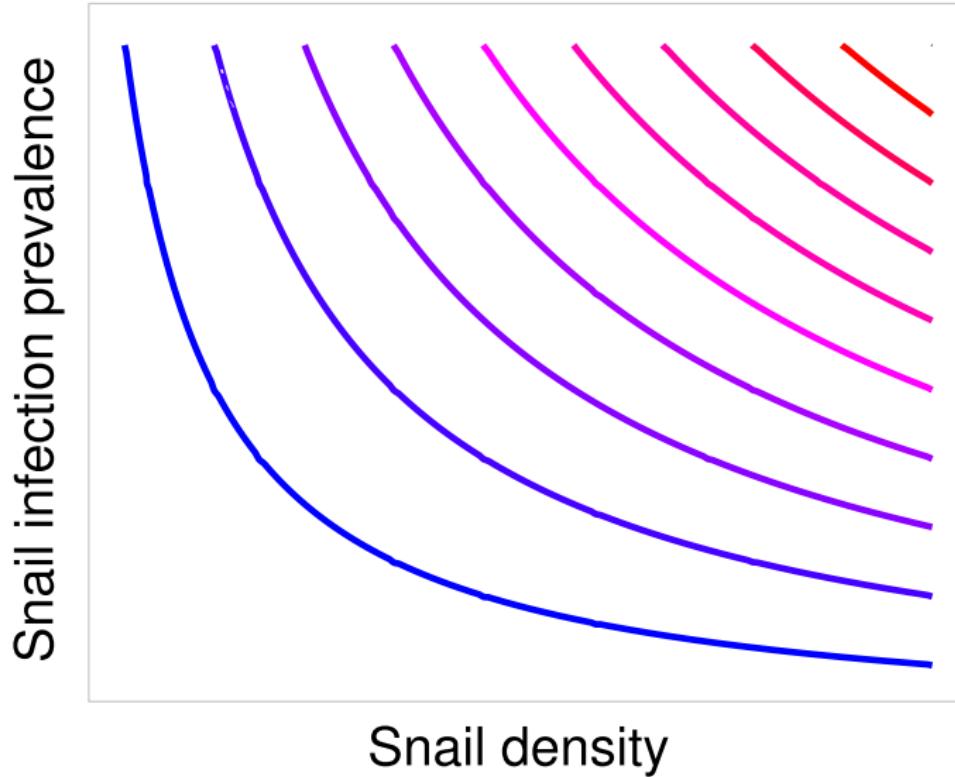
Infected snail density = density \times prevalence

Snail infection prevalence

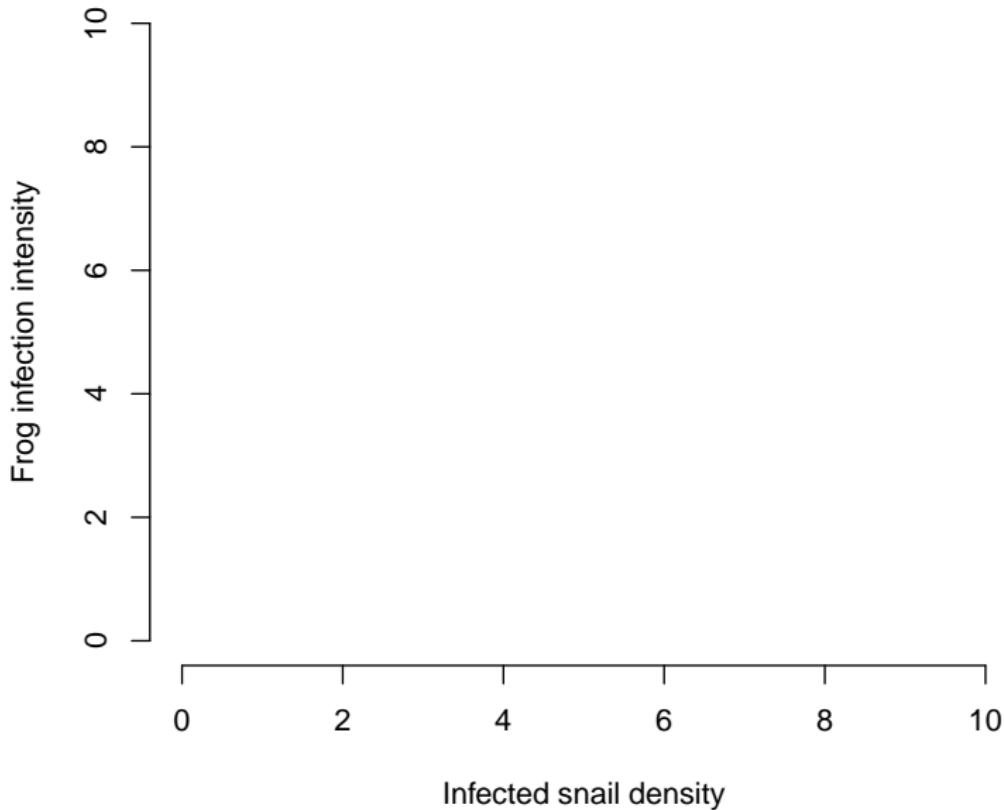
Snail density



Infected snail density = density \times prevalence



Goal: use ISD to explain frog infections

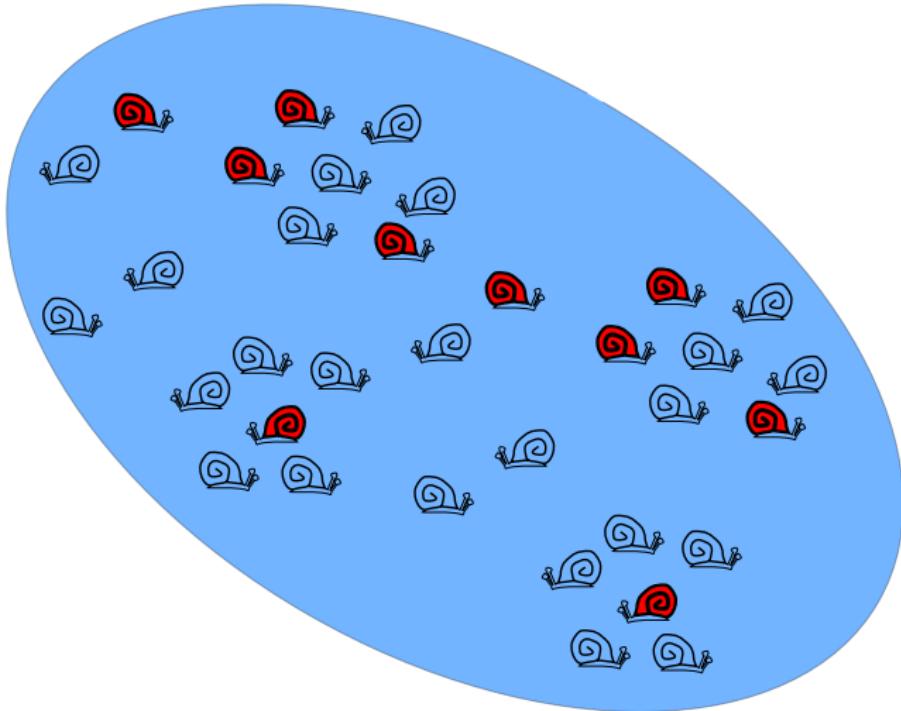


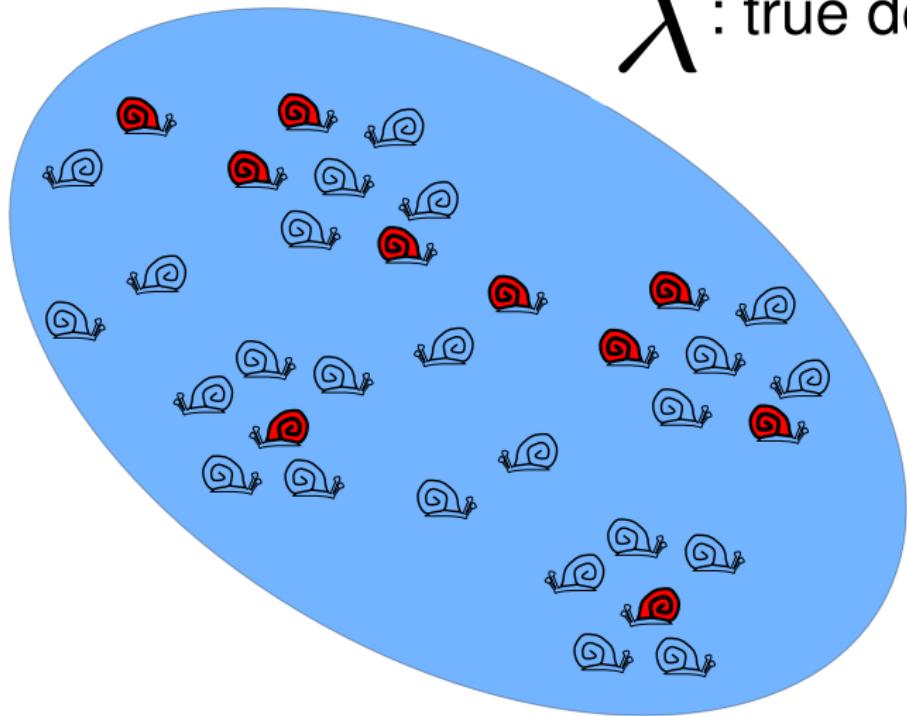
Infected snail density

Product of two unknowns:

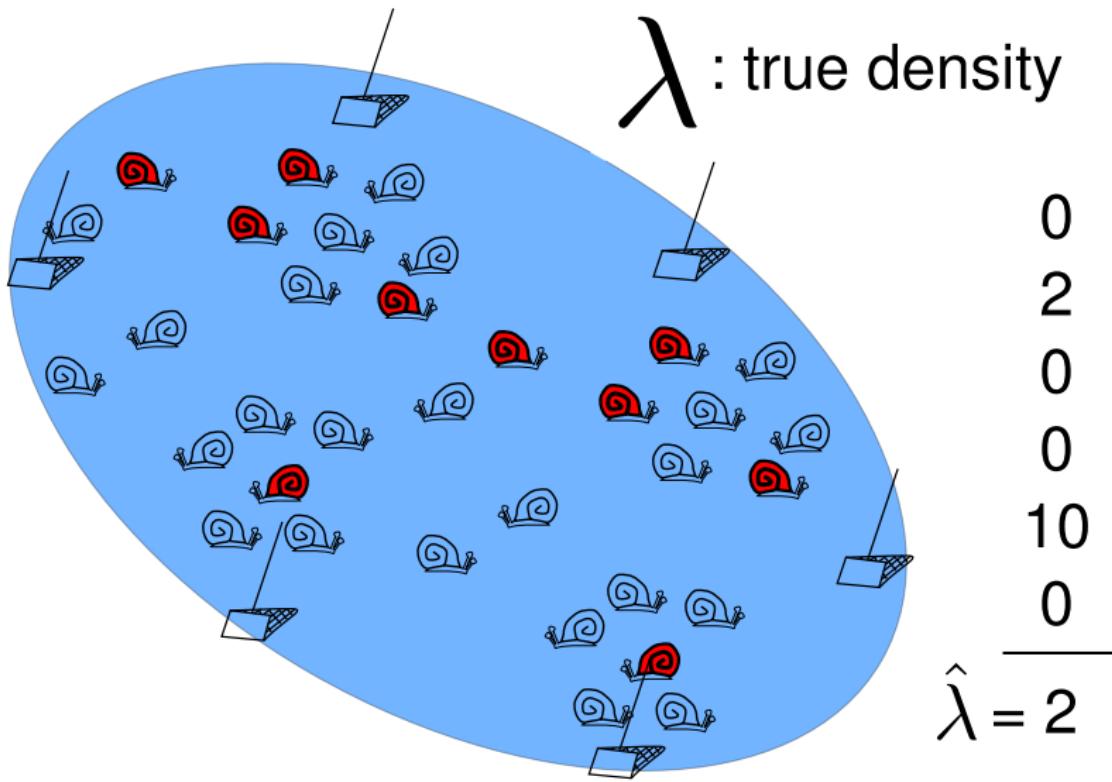
1. Snail density
2. Snail infection prevalence

Snail density

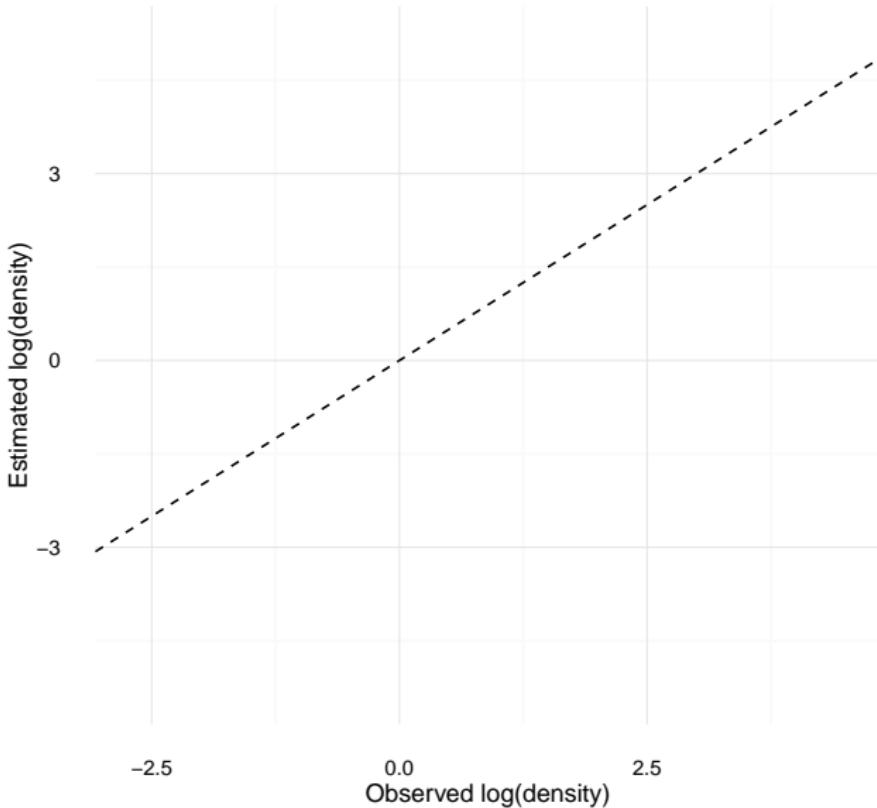


λ : true density

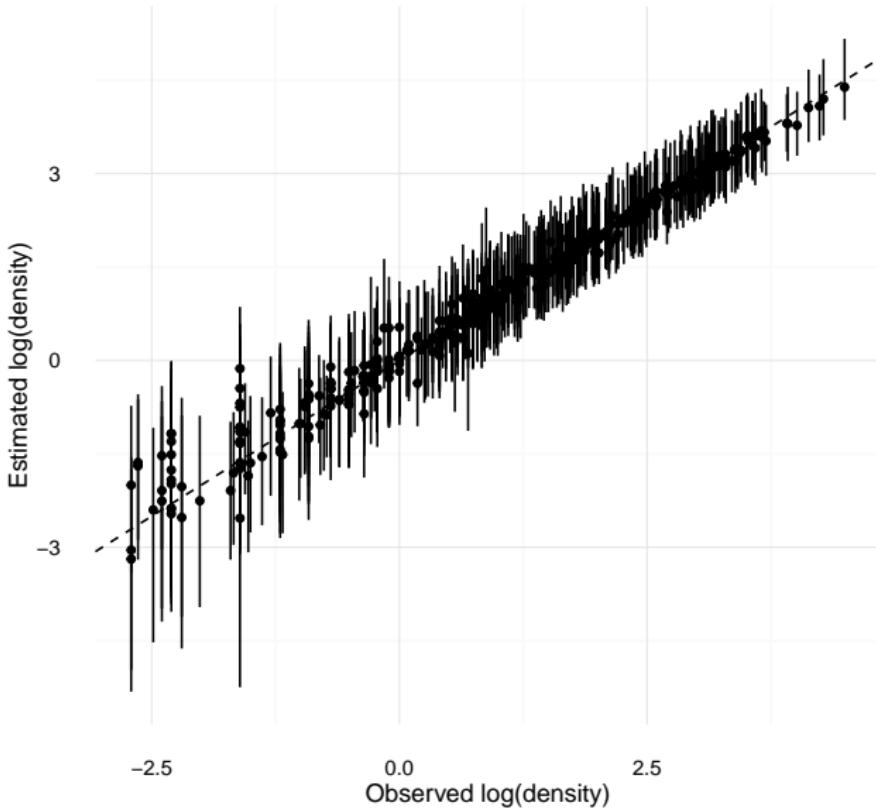
Data and a point estimate



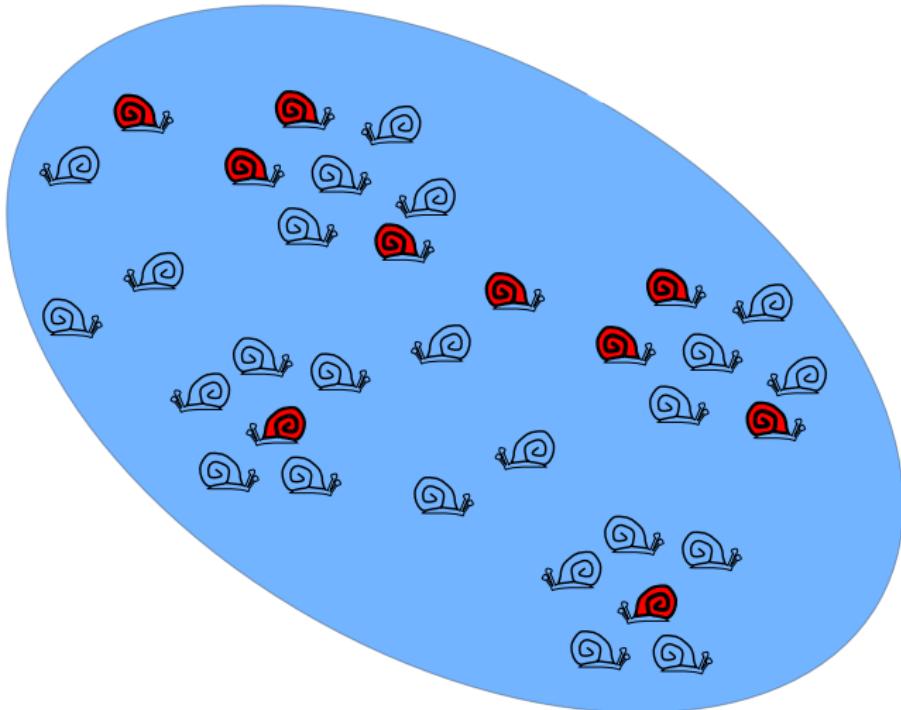
Snail density: observed vs. estimated



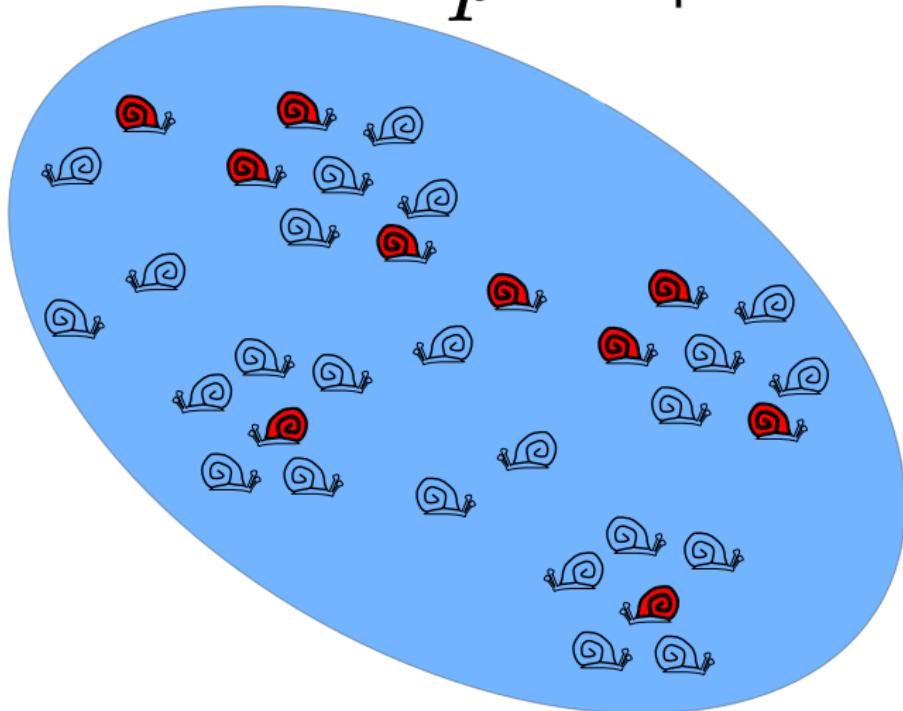
Snail density: observed vs. estimated



Snail infection prevalence



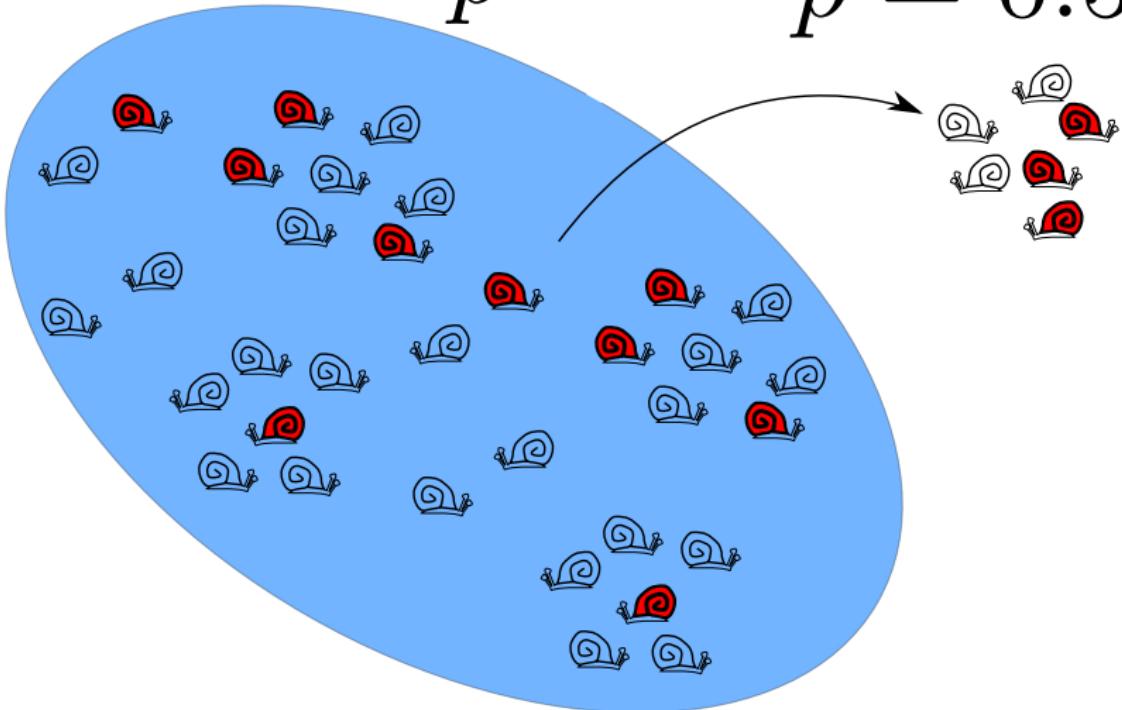
p : true prevalence



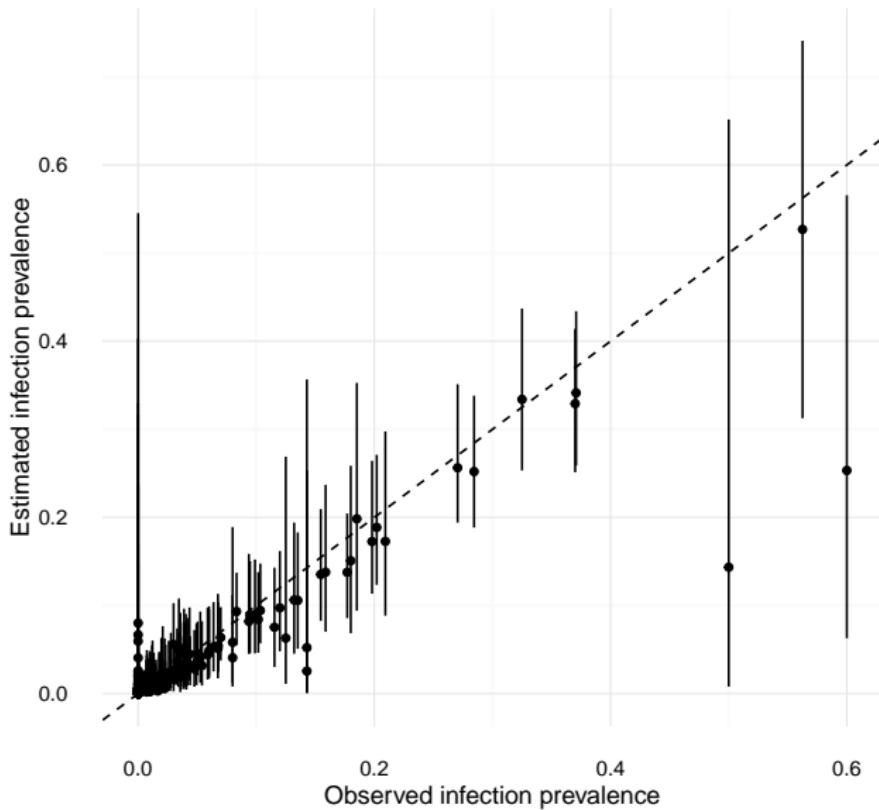
Data and a point estimate

p

$\hat{p} = 0.5$



Snail infection prevalence: observed vs. estimated



Two approaches

1. Naive point-estimate

$$ISD = \hat{p}\hat{\lambda}$$

Two approaches

1. Naive point-estimate

$$ISD = \hat{p}\hat{\lambda}$$

$$ISD = 0.5 \times 2$$

$$ISD = 1$$

Two approaches

1. Naive point-estimate

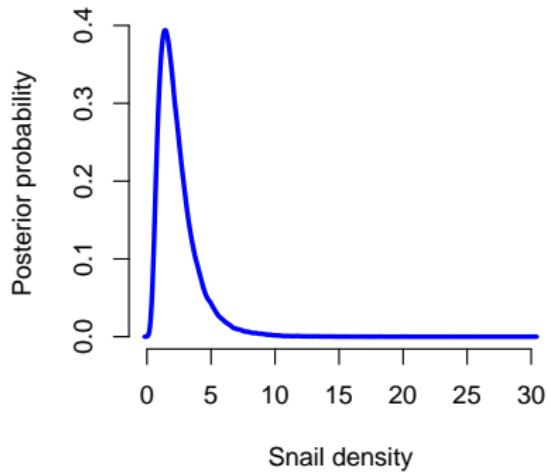
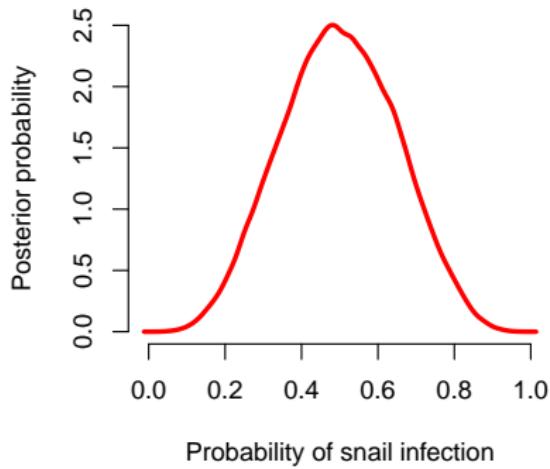
$$ISD = 1$$

2. Bayesian estimate

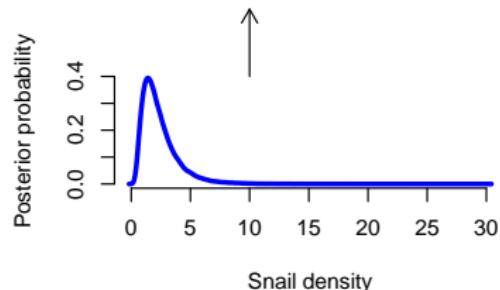
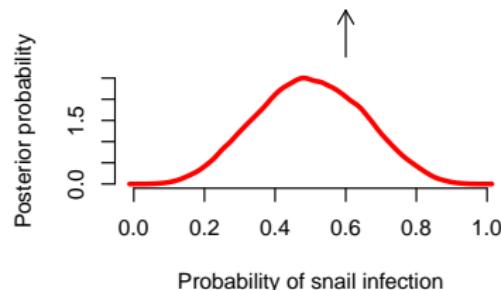
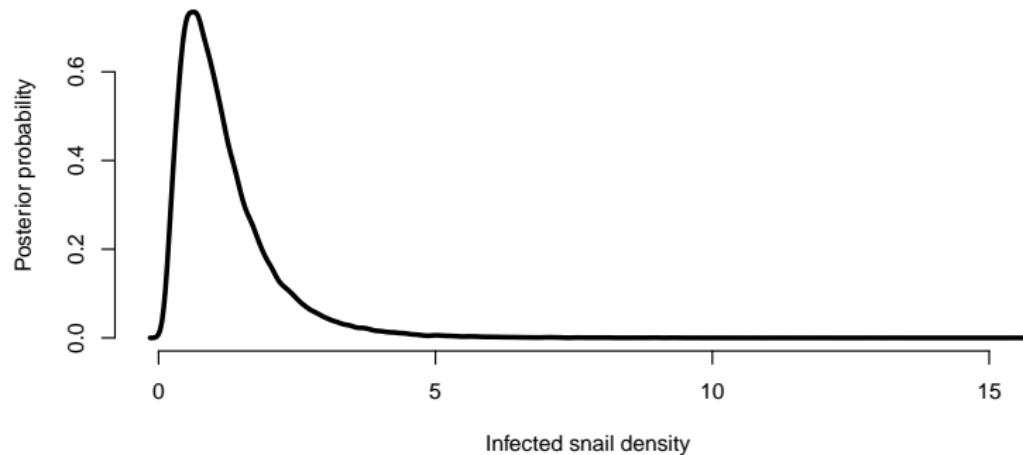
$$ISD_b = p_b \lambda_b$$

$$ISD_b = \textcolor{red}{p}_{\mathfrak{b}} \lambda_{\mathfrak{b}}$$

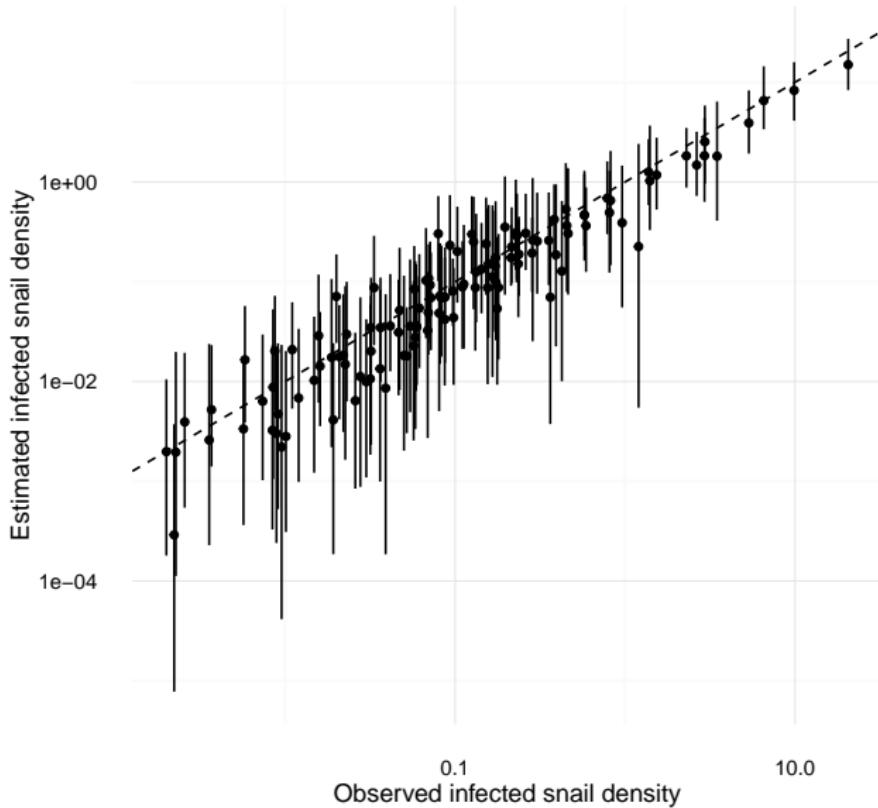
$$ISD_b = p_b \lambda_b$$



$$ISD_b = p_b \lambda_b$$



Infected snail density: observed vs. estimated



Why worry about error in an explanatory variable?

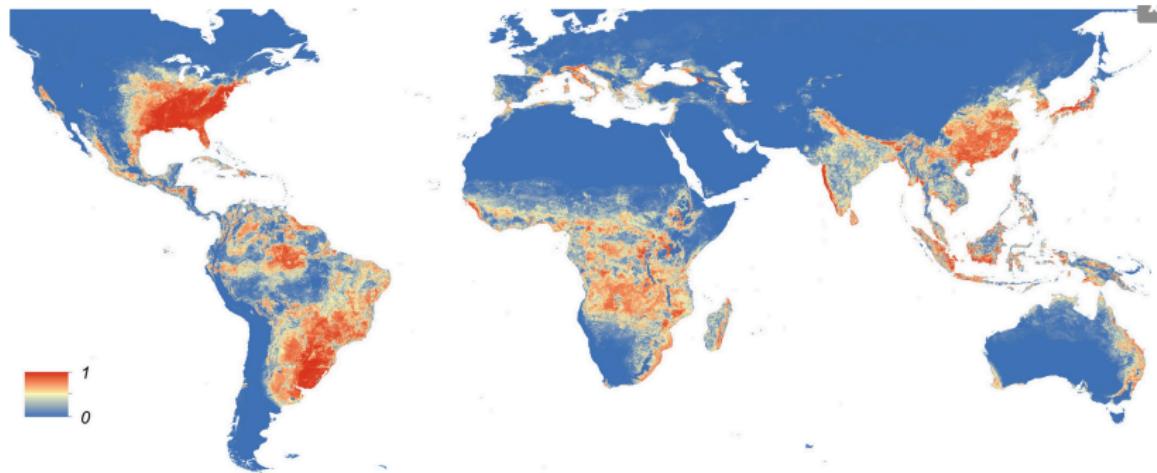
Why worry about error in an explanatory variable?

1. **Principle:** being honest about uncertainty

Why worry about error in an explanatory variable?

1. **Principle:** being honest about uncertainty
2. **Risk:** interval estimates are more useful

Example: what's the risk of Zika virus?



Kraemer et al. 2015

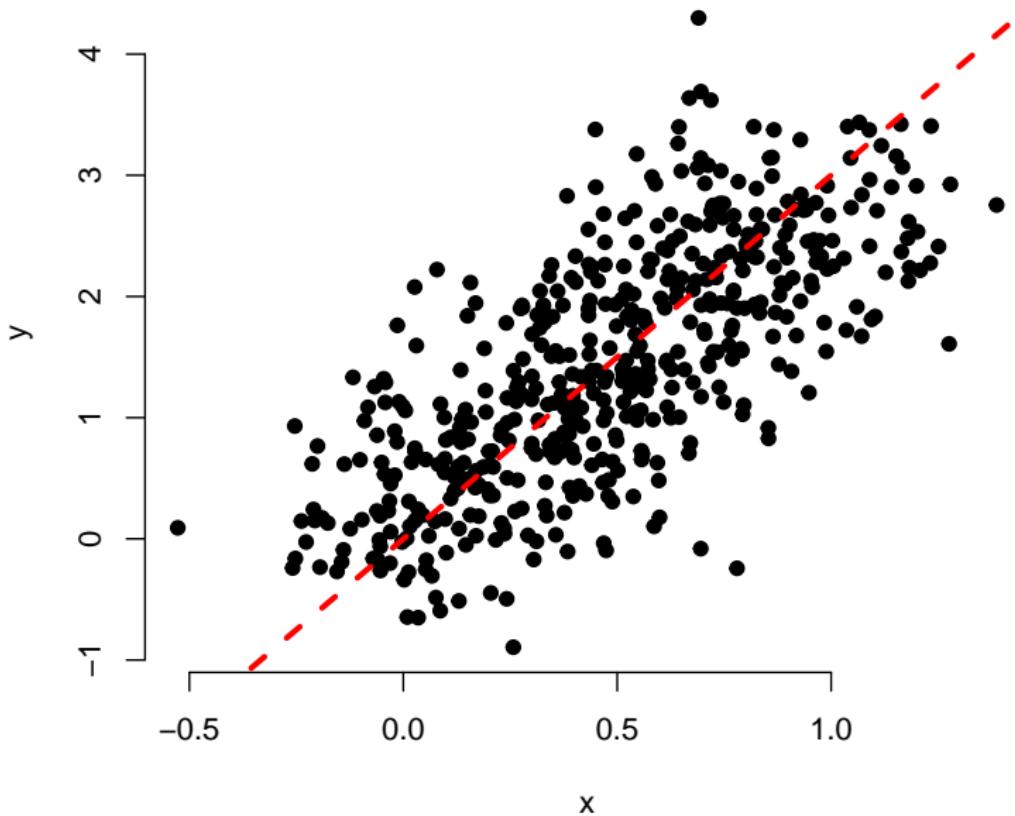
Why worry about error in an explanatory variable?

1. **Principle:** being honest about uncertainty
2. **Risk:** interval estimates are more useful
3. **Propriety:** error causes bias

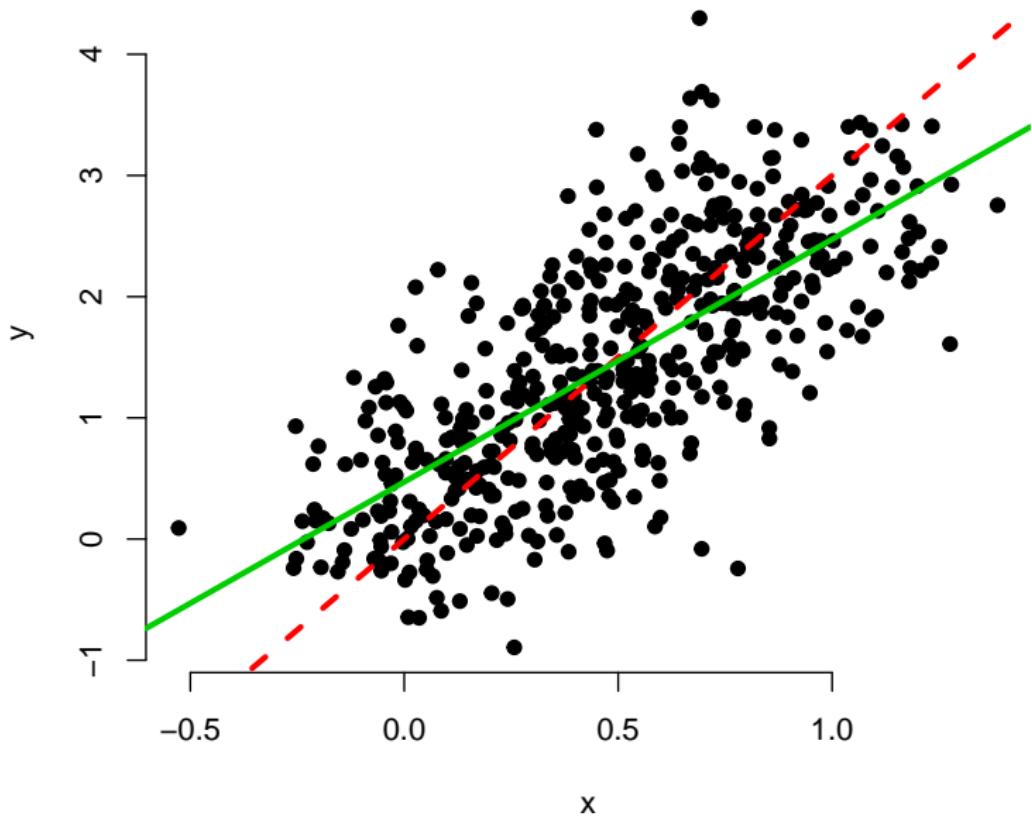
Karl Pearson



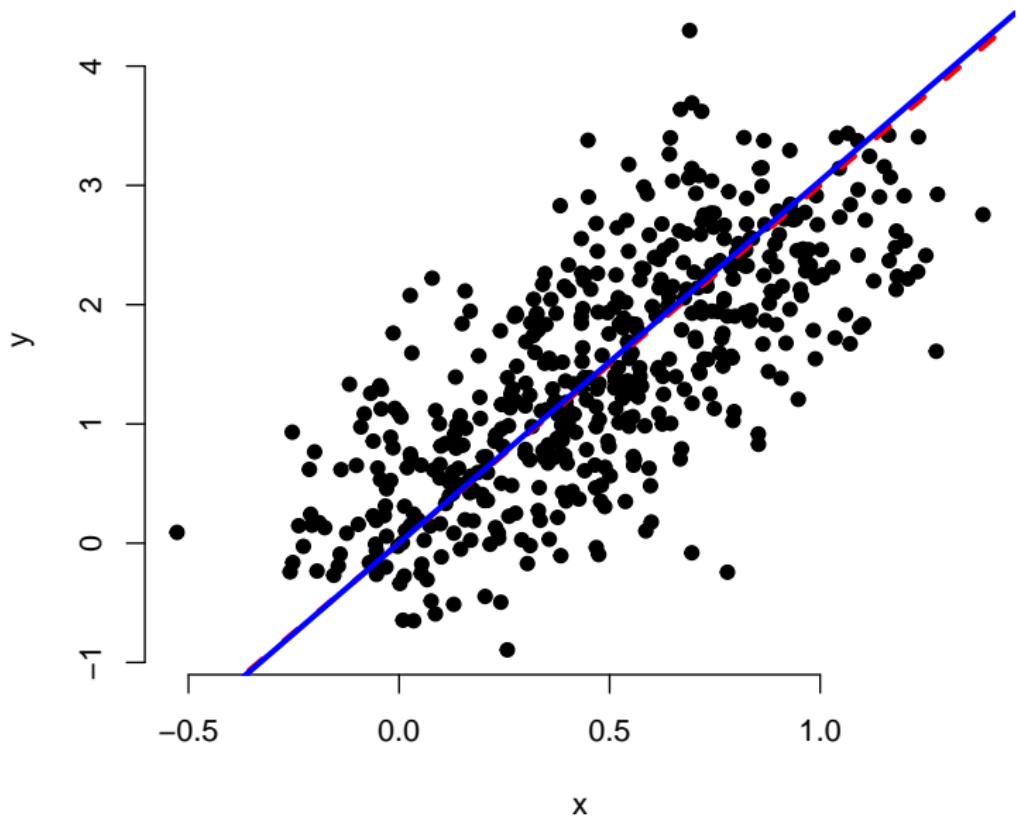
Errors in variables → biased slopes



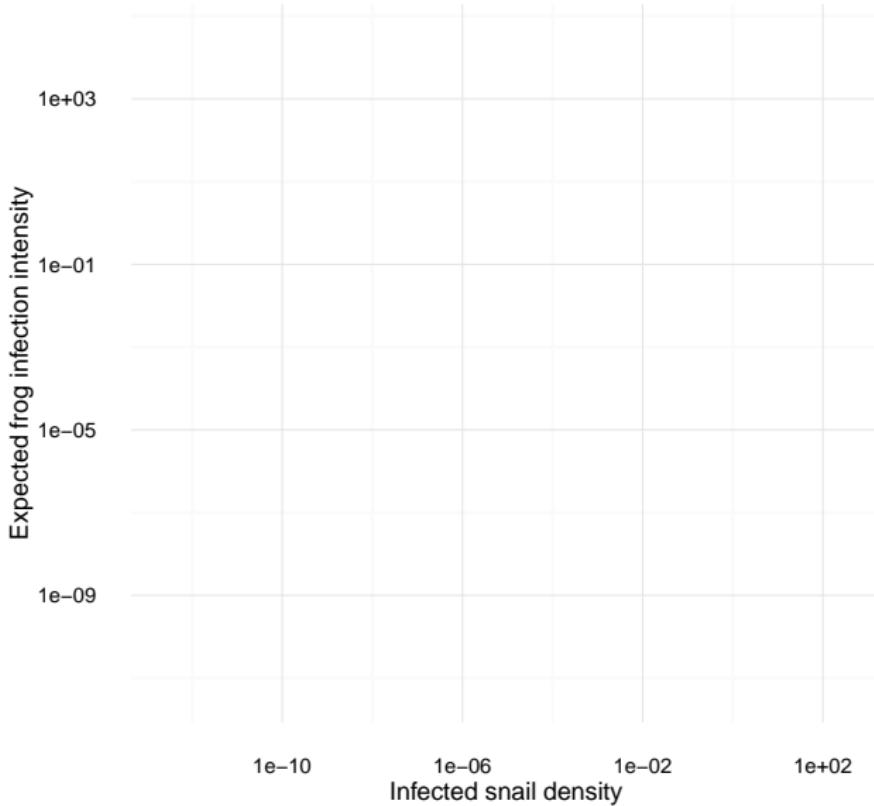
Errors in variables → biased slopes



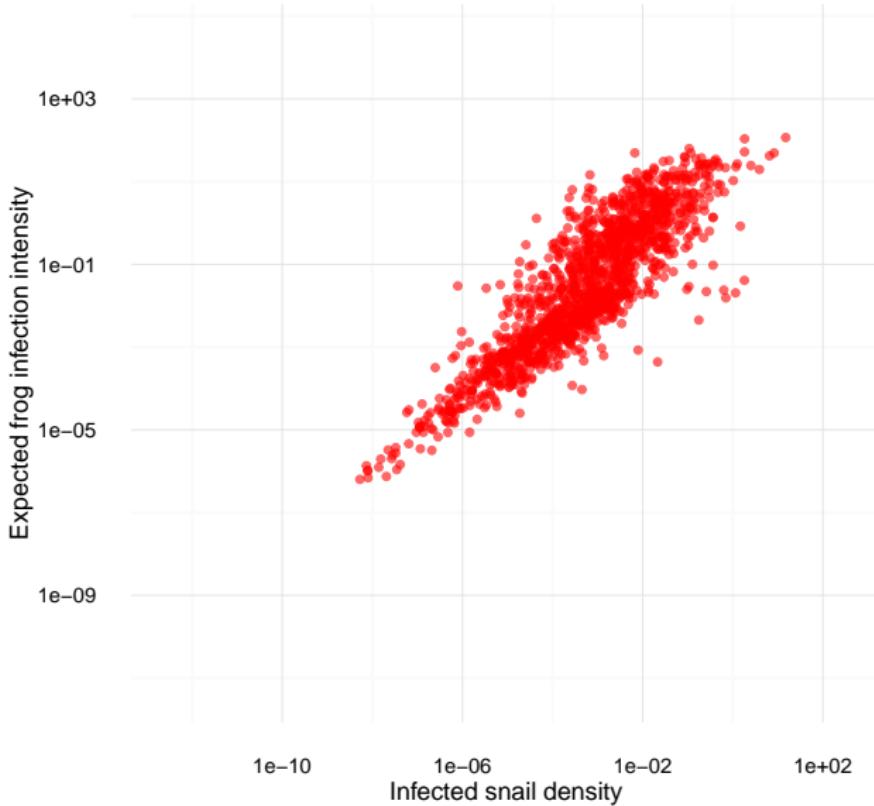
Accounting for measurement error → unbiased slopes



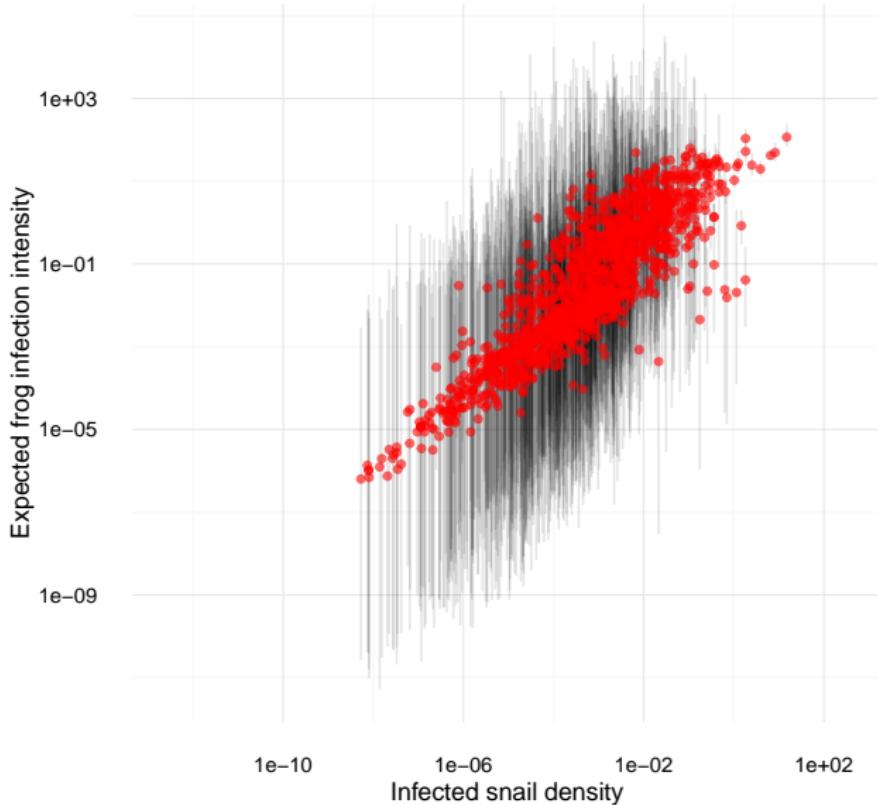
Infected snail density and frog infection



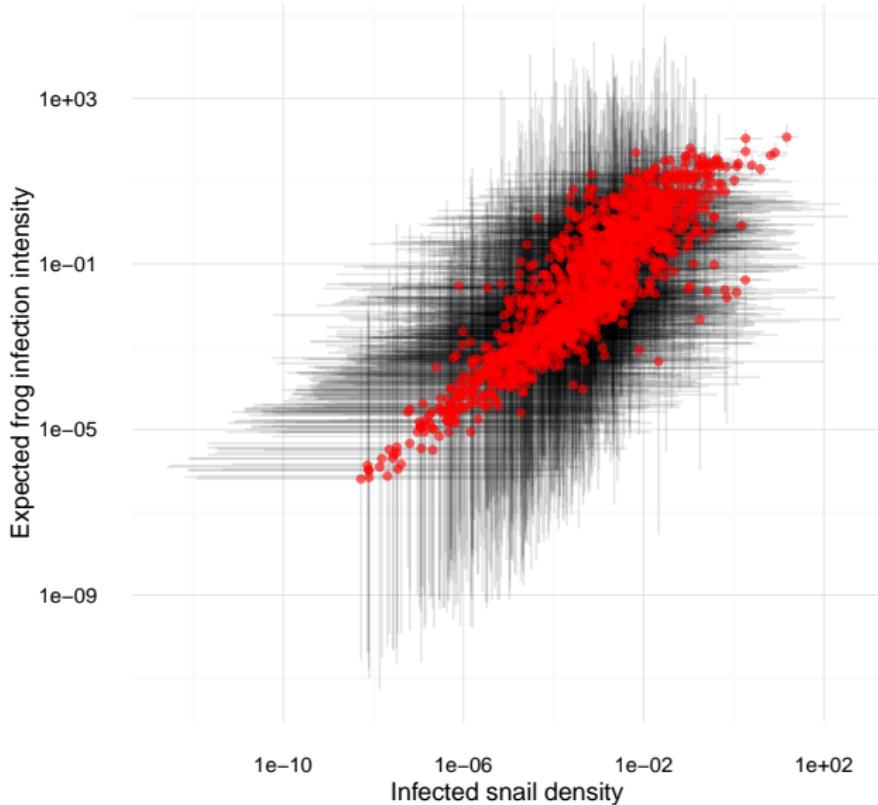
Infected snail density and frog infection



Uncertainty in y



Uncertainty in x and y



Questions

1. How can we link snail and frog infections?
 - ▶ Bayesian error propagation

Questions

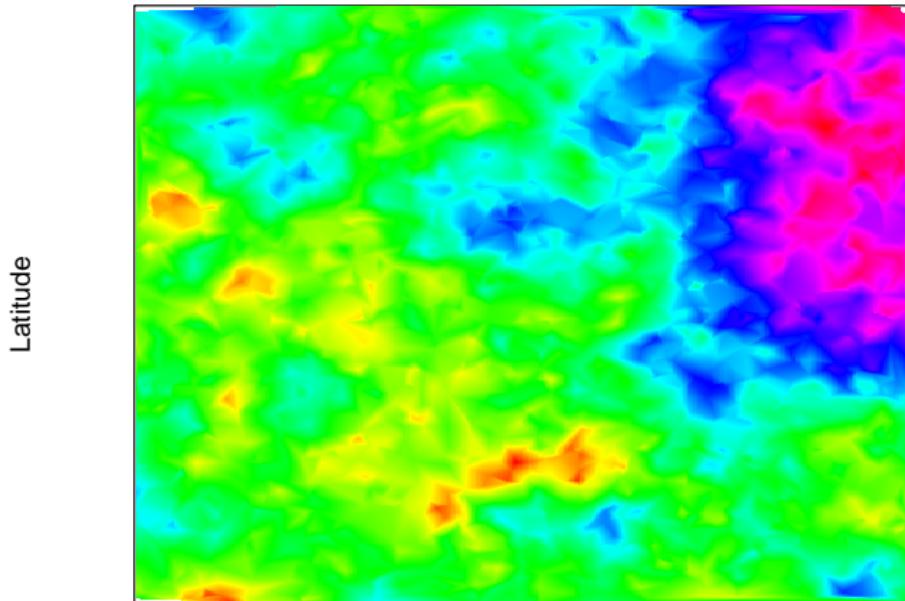
1. How can we link snail and frog infections?
2. **What is the spatiotemporal nature of infection dynamics?**

Three spatiotemporal processes

1. Snail density
2. Snail *Rib* infections
3. *Rib* infections in frogs

Spatiotemporal Gaussian process

Gaussian process in year 1

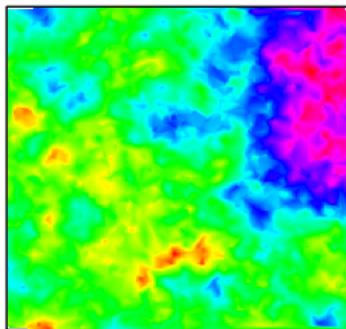


Longitude

Spatiotemporal Gaussian process

Gaussian process in year 1

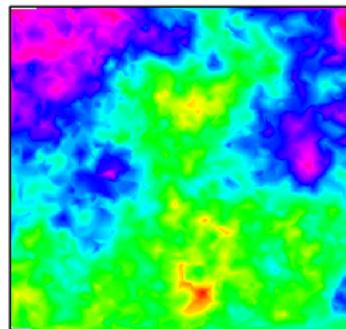
Latitude



Longitude

Δ Gaussian process to year 2

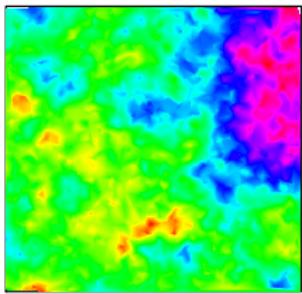
Latitude



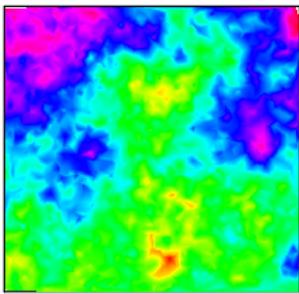
Longitude

Spatiotemporal Gaussian process

Gaussian process in year 1



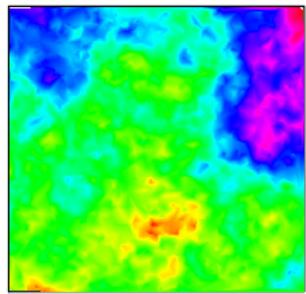
Δ Gaussian process to year 2



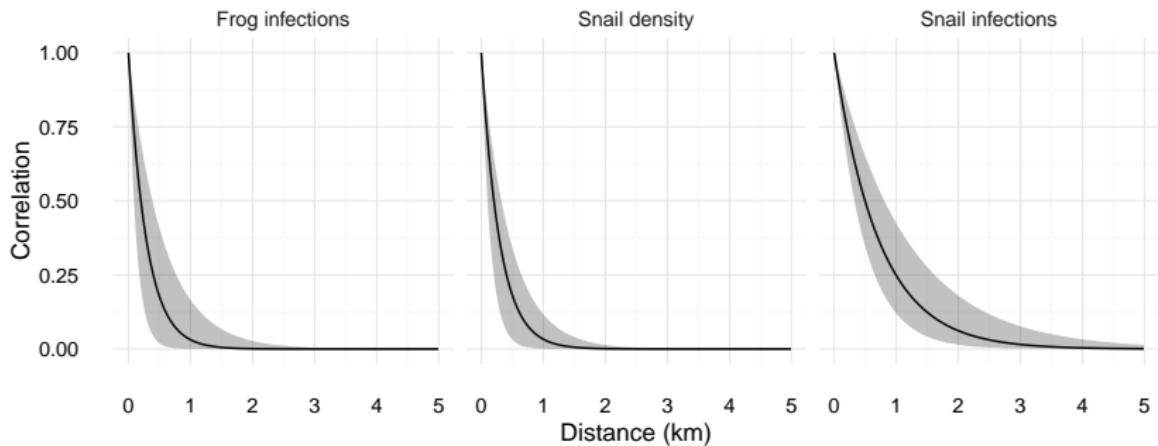
+

=

Gaussian process in year 2



Estimated spatial correlation functions

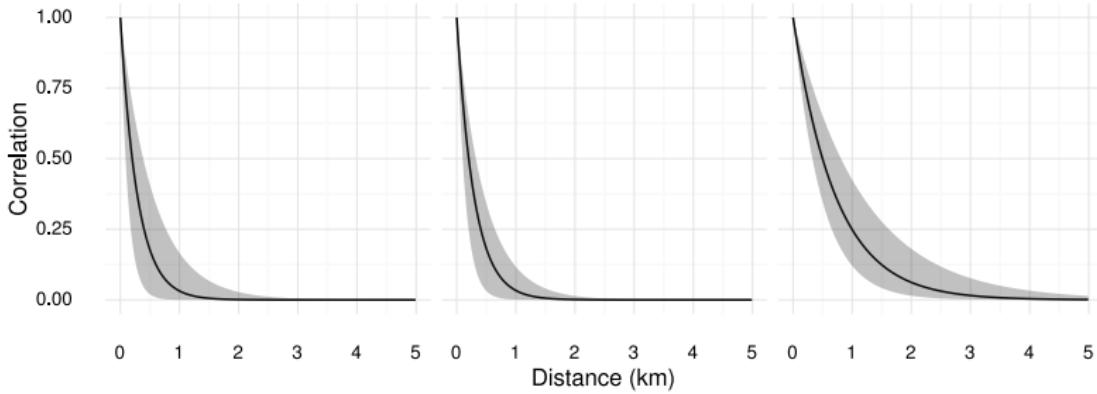




Frog infections

Snail density

Snail infections



Spatial correlation range on the landscape

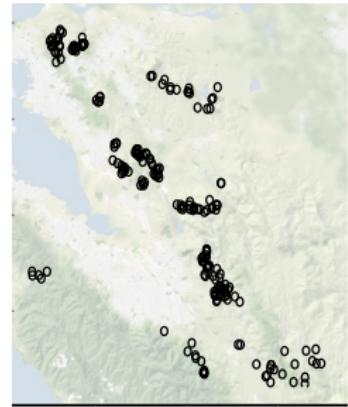
Frog infection



Snail density

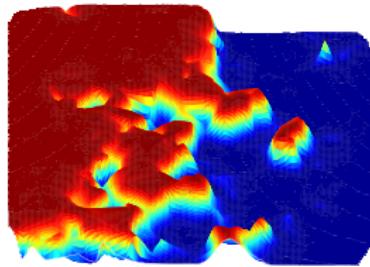


Snail infection

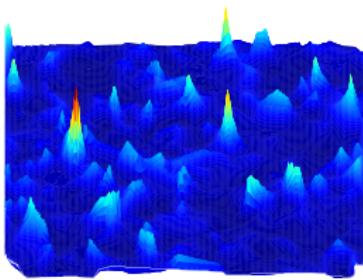


Disease risk: a product of processes

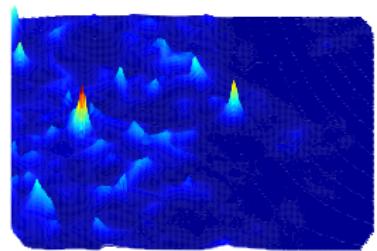
Snail infection probability



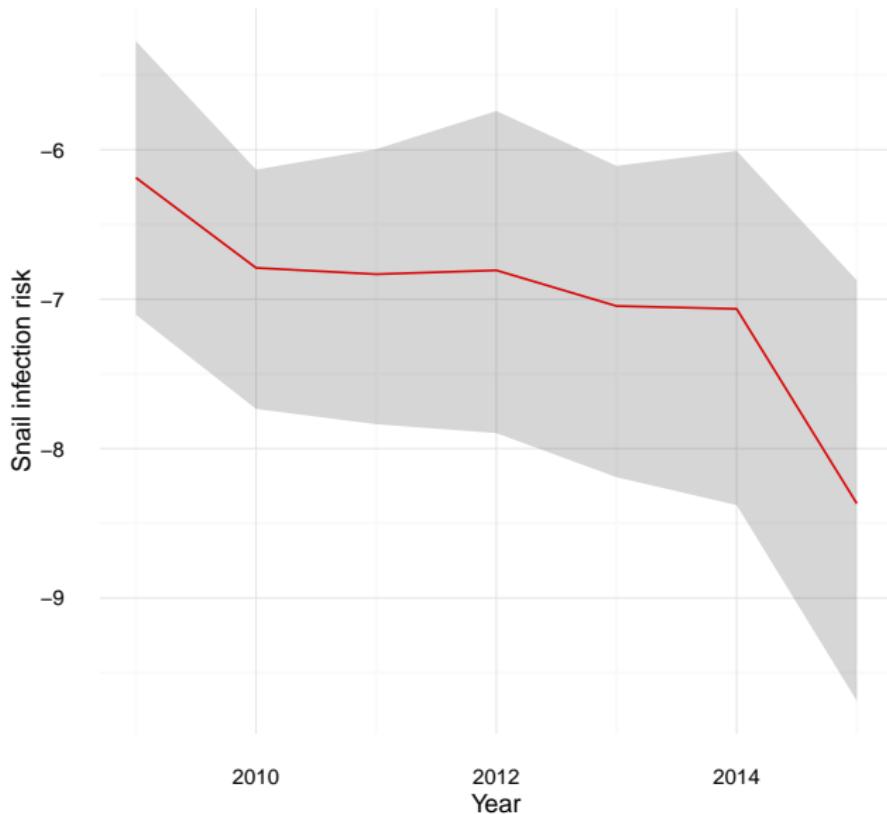
Snail density



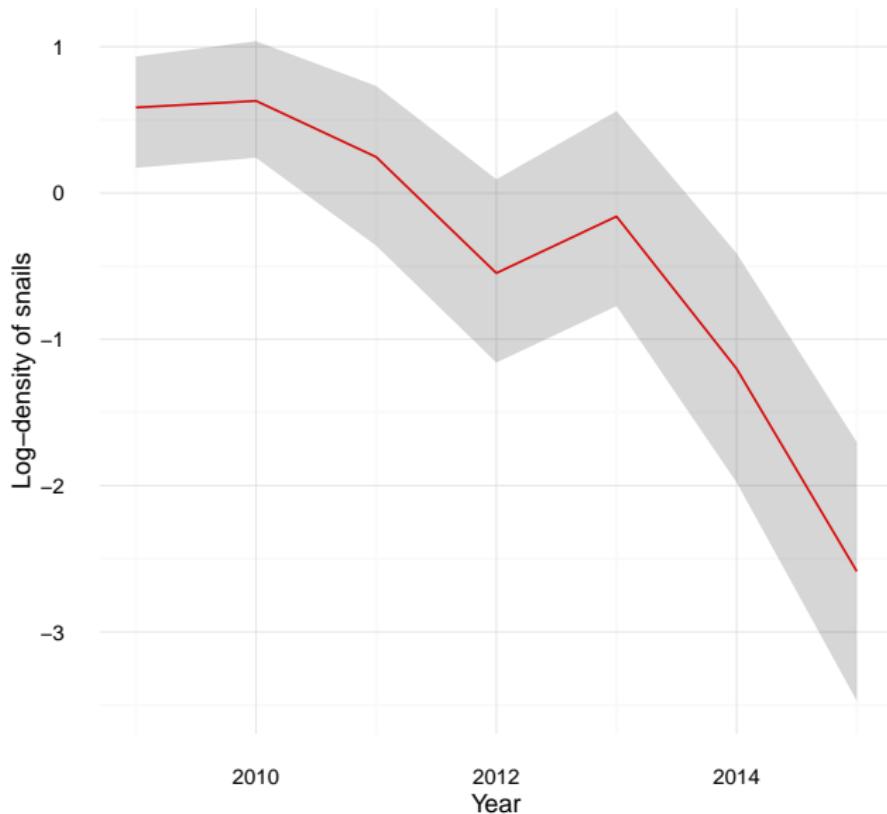
Infected snail density



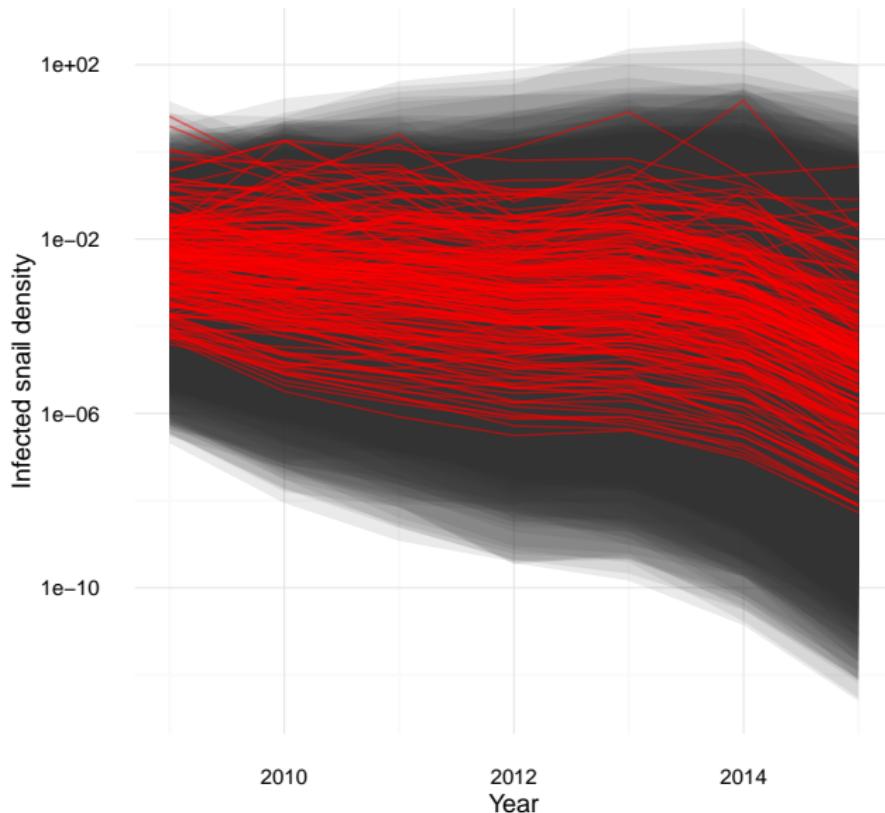
Landscape-level trend: snail infection



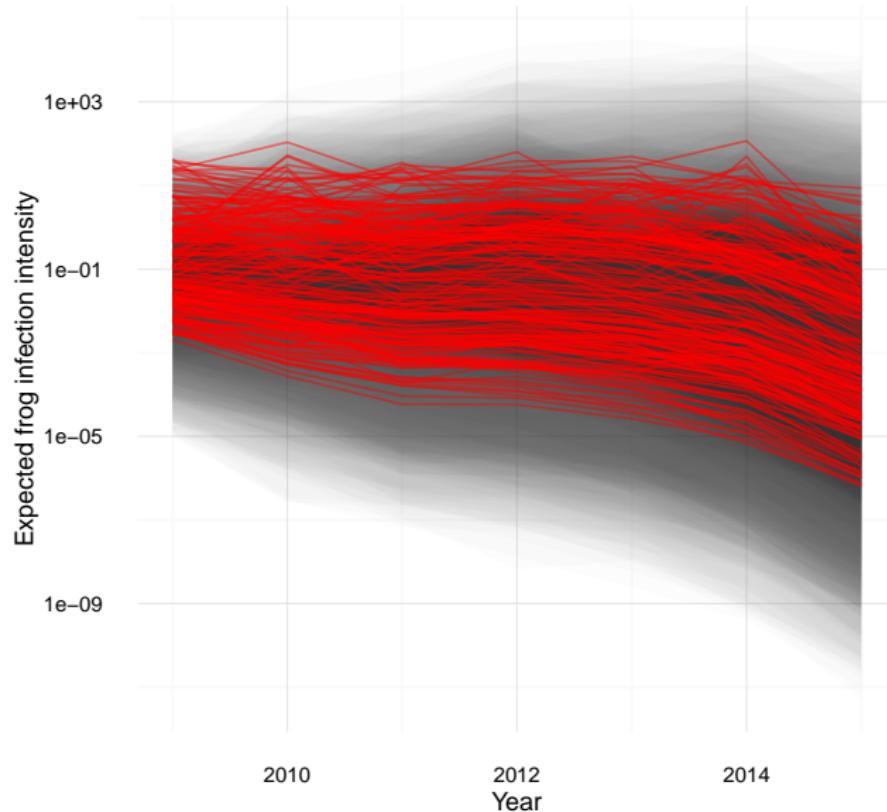
Landscape-level trend: snail density



Infected snail density through time



Frog infections through time

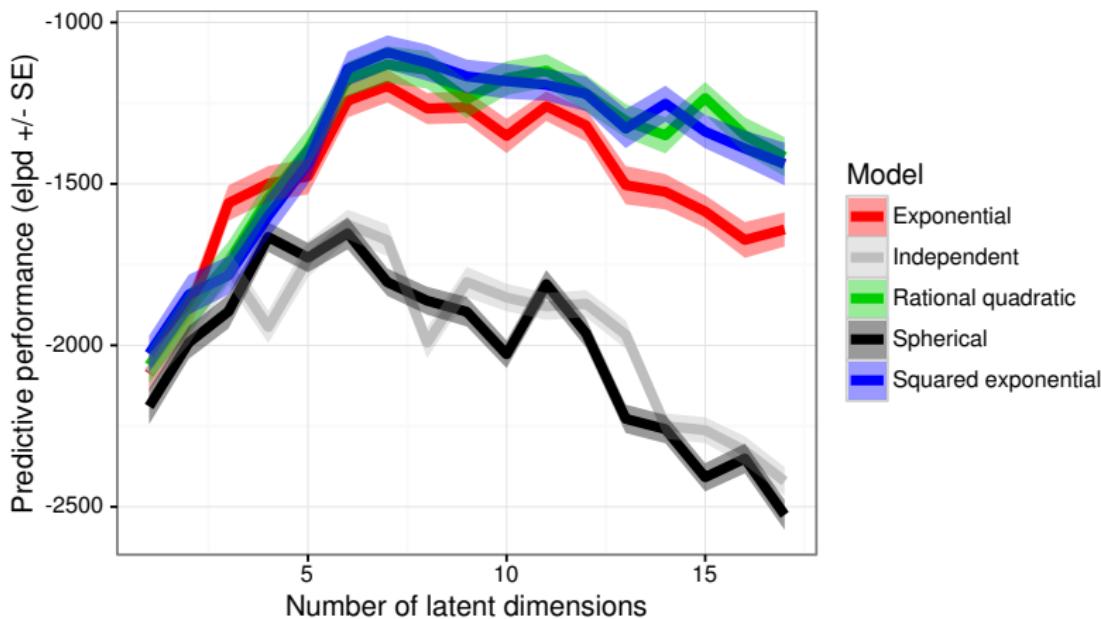


Questions

1. How can we link snail and frog infections?
 - ▶ Bayesian error propagation
2. What is the spatiotemporal nature of infection dynamics?
 - ▶ combination of long and short range processes
 - ▶ snail density decreasing through time → lower disease risk

Gaussian processes

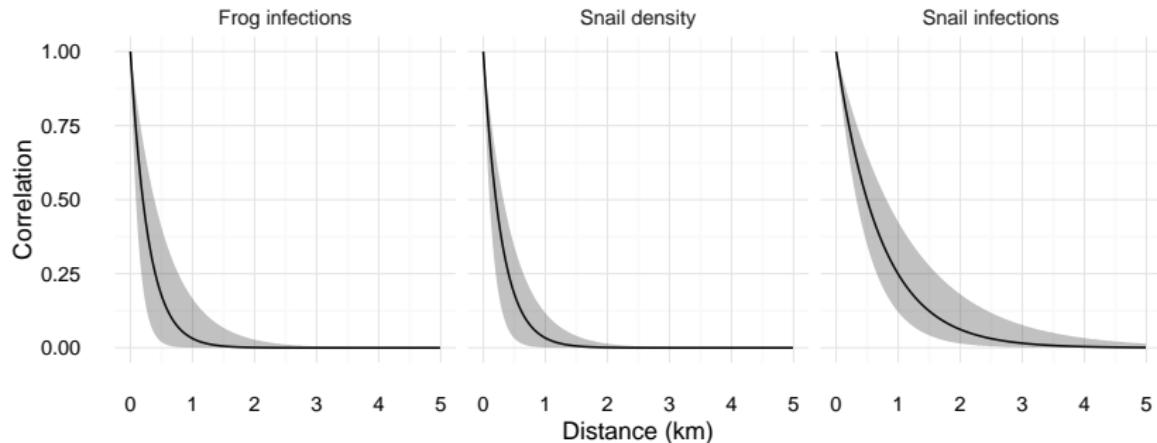
- ▶ priors over unknown functions
- ▶ useful in **tree space**



Gaussian processes

Gaussian processes:

- ▶ priors over unknown functions
- ▶ useful in tree space, and **spatial space**



Huge thanks

Other students and friends

Aidan Beers
Amanda Hund
Amy Churchill
Angela Boag
Caroline Tucker
Chris Steenbock
Christine Avena
Collin "Schwan Quixote" Schwantes
David Zonana
Evelyn Cheng
Gaurav Vaidya
Helen McCreery
Jeff McClenahan
Joey Hubbard
Jon Leff
Jordan Keueneman
Julian Resasco
Kevin Bracy Knight
Kika Tarsus
Lauren Shoemaker
Lisette Arellano
Matt "Marty Milkins" Wilkins
Megan Blanchard
Nathan Kleist
Nora Connor
Se Jin Song
Teal Potter
Tim Szewczyk
Toby Hammer
Topher Weiss-Lehman
Tim Farkas
Ty Tuff

Other faculty

Nichole Barger
John Basey
Dean Bowers
Kendi Davies
Michael Grant
Brett Melbourne
Rebecca Safran
Stacey Smith
Carol Wessman

Committee

Pieter Johnson
Valerie McKenzie
Andy Martin
Sam Flaxman
Vanja Dukic

Labmates

Joseph Mihaljevic
Sarah Orlofske
Sara Paull
Katherine Richgels
Travis McDevitt-Galles
Dan Preston
Miranda Redmond
Jazzmin Jenkins
Dain Calhoun
Will Stutz
Sarah Haas
Bethany Hoye
Jason Hoverman
Kim Medley
Yuri Springer

Climbing friends



Steph



www.aysanaimages.zenfolio.com

Family

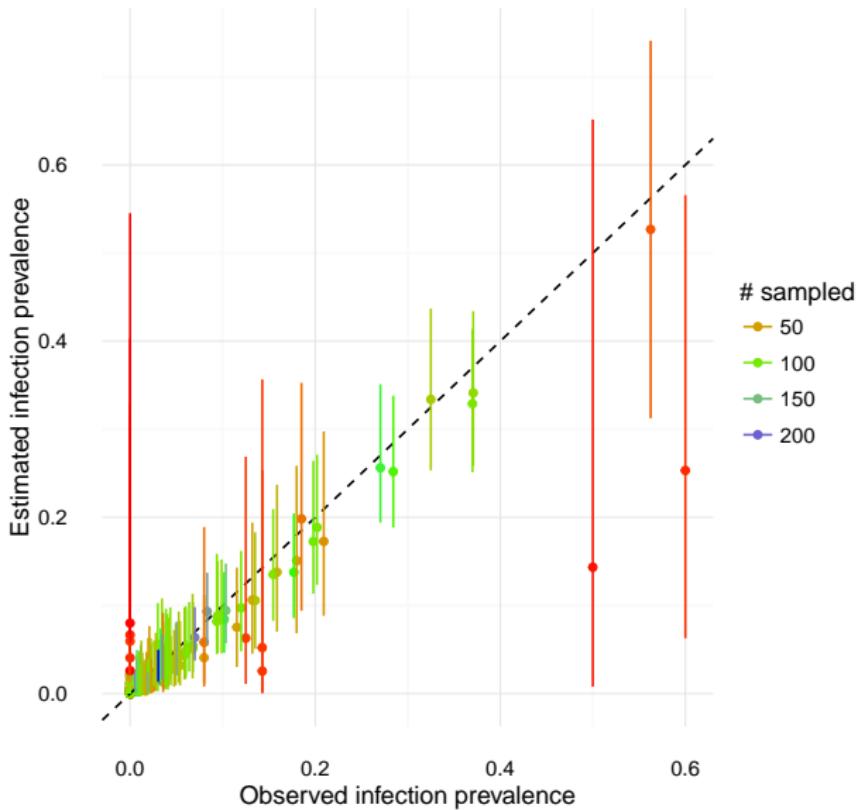


Non-human helpers



Thank you

Snail infection prevalence: observed vs. estimated



Spatiotemporal process ranges

- ▶ Snail density: 0.3 - 1.2 km
- ▶ Snail infections: 1.3 - 3.4 km
- ▶ Frog infections: 0.4 - 1.5 km