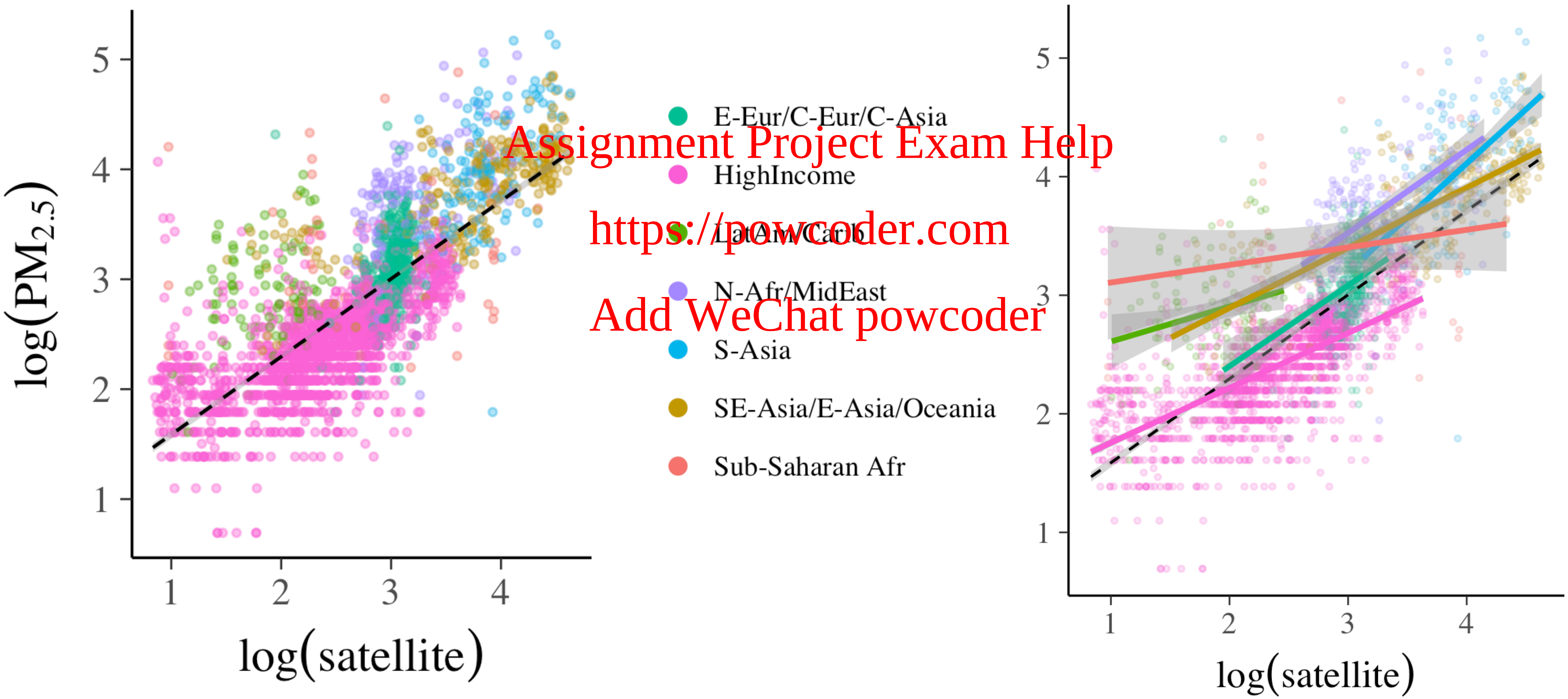
A QUICK LOOK INTO INLA

STA465: Theory and Methods for Complex Spatial Data

Instructor: Dr. Vianey Leos Barajas

BACK TO THE AIR POLLUTION

GLOBAL PM2.5 DATA



FITTING MULTILEVEL MODELS IN INLA

WHAT IS INLA?

- INLA stands for the Integrated Nested Laplace Approximation
- It is a clever way to compute posterior distributions for multilevel models and their spatial generalizations

- For the purposes of this course, it is magic.
- This magic takes the form of an R-package that can be downloaded from http://r-inla.org/

HOW TO CALL INLA

```r
library(INLA)

formula = log(pm25) ~ 1 + f(region_int, model="iid") +
  log(sat_2014) + f(region_sat, log(sat_2014), model="iid")

data = data.frame(pm25 = GM$pm25, sat_2014 = GM$sat_2014,
                  region_int = GM$super_region, region_sat = GM$super_region)

fit = inla(formula, data=data, family="gaussian")

summary(fit)
```


- INLA works just like the formula in lm but with some slight differences
- The f() function describe **random effects** (ie things that have more structure.
- The first two terms are μ_j . The second two are $\beta_j x_{ij}$

MORE INFORMATION ABOUT INLA

- Sections 4.6-4.9 on Blangiardo and Cameletti.

- Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny

WHAT COMES OUT?

```
##
## Call:
##    "inla(formula = formula, family = \"gaussian\", data = data)"
## Time used:
##     Pre = 2.51, Running = 1.87, Post = 0.104, Total = 4.48
## Fixed effects:
##                 mean    sd 0.025quant 0.5quant 0.975quant  mode kld
## (Intercept)    1.705 0.249      1.190    1.710      2.193 1.719   0
## log(sat_2014) 0.512 0.078      0.357    0.511      0.672 0.510   0
##
## Random effects:
##   Name      Model
##     region_int IID model
##    region_sat IID model
##
## Model hyperparameters:
##                                          mean     sd 0.025quant 0.5quant
## Precision for the Gaussian observations  8.57  0.223      8.136     8.57
## Precision for region_int                 3.84  2.490      0.969     3.23
## Precision for region_sat                39.89 26.666      9.816    33.23
##                                         0.975quant  mode
## Precision for the Gaussian observations       9.01  8.56
## Precision for region_int                     10.32  2.25
## Precision for region_sat                    109.59 22.87
##
## Expected number of effective parameters(stdev): 12.67(0.475)
## Number of equivalent replicates : 235.17
##
## Marginal log-Likelihood:  -1092.74
```

WHAT WERE THE PRIORS?

- Good question!

- You've got to dig into the documentation to find them.

$$\mu \sim N(0,1000)$$

$$\beta \sim N(0,100)$$

$$\tau_\mu^2 \sim Exp(100)$$

$$\tau_\beta^2 \sim Exp(100)$$

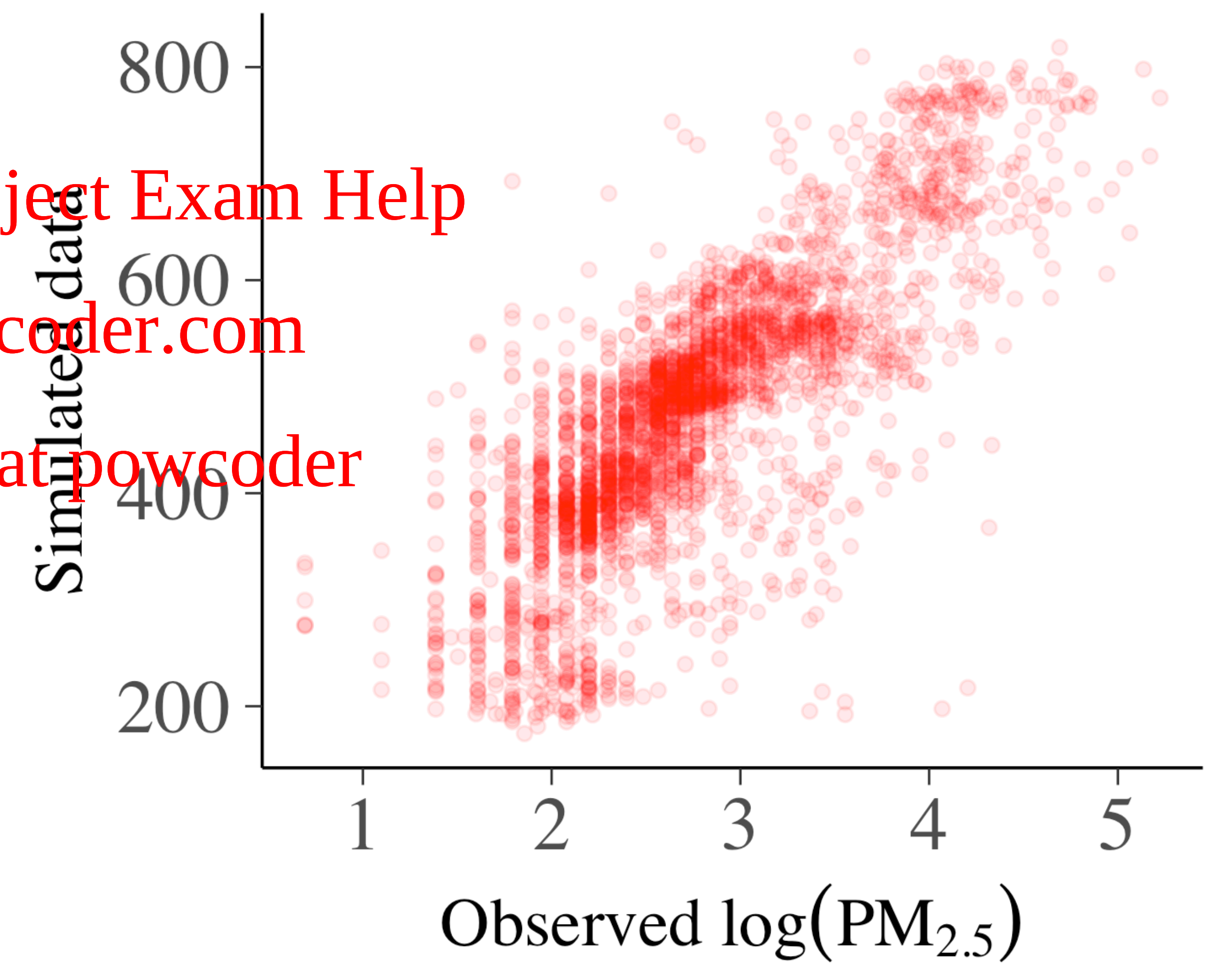
$$\sigma^2 \sim Exp(100)$$

- Are these sensible???

WE CAN SIMULATE FROM THE PRIORS

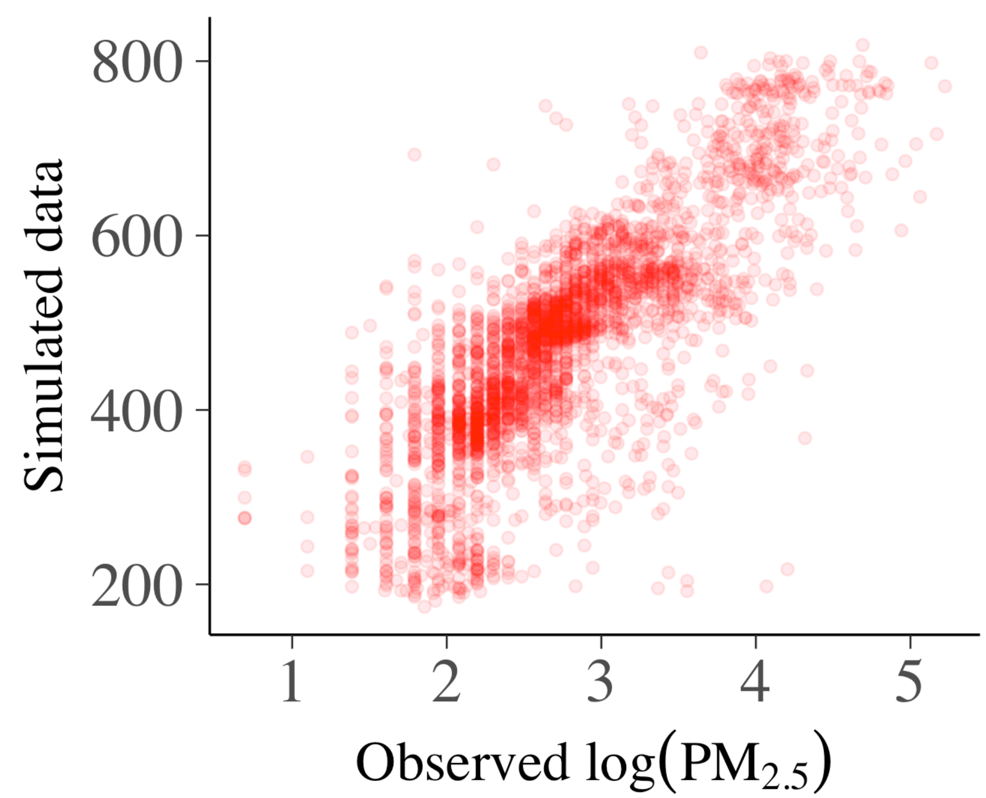
- Simulate

$$\mu \sim N(0,1000)$$
$$\beta \sim N(0,100)$$
$$\tau_\mu^2 \sim Exp(100)$$
$$\tau_\beta^2 \sim Exp(100)$$
$$\mu_j \sim N(\mu, \tau_\mu^2)$$
$$\beta_j \sim N(\beta, \tau_\beta^2)$$
$$\sigma^2 \sim Exp(100)$$
$$y_{ij} \sim N(\mu_j + \beta_j x_{ij}, \sigma^2)$$



WAIT?! WHAT?

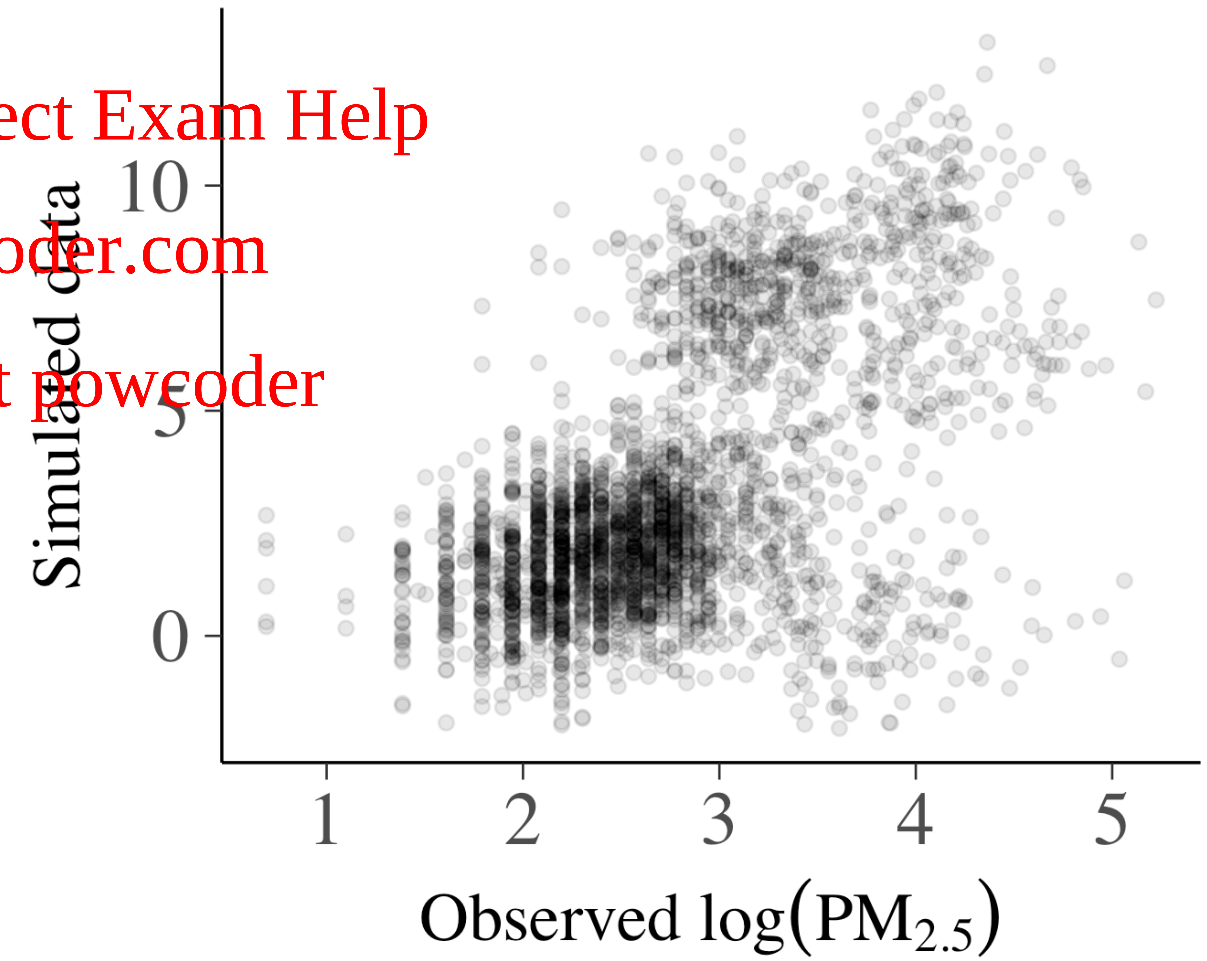
- The prior model is two orders of magnitude off the real data
- Two orders of magnitude on the log scale!
- Log density of neutron star only 60 $\mu g m^{-3}$
- What does this mean practically?
- The data will have to overcome the prior…



WE CAN DO BETTER

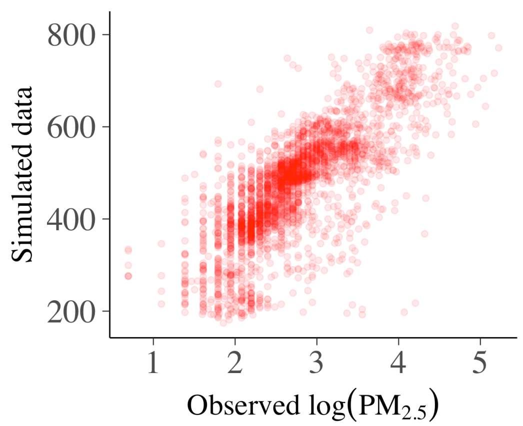
- With more sensible priors

$$\mu \sim N(0,1)$$
$$\beta \sim N(1,1)$$
$$\tau_\mu^2 \sim N_+(0,1)$$
$$\tau_\beta^2 \sim N_+(0,1)$$
$$\sigma^2 \sim N_+(0,1)$$

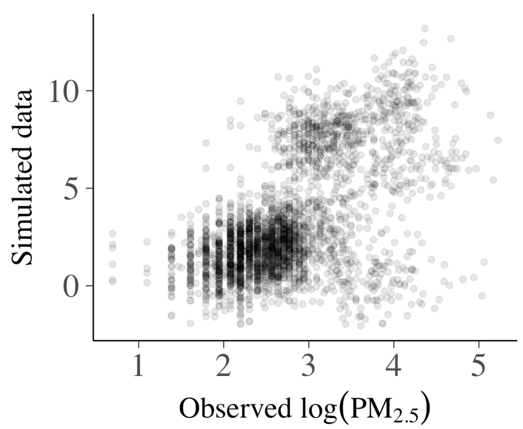


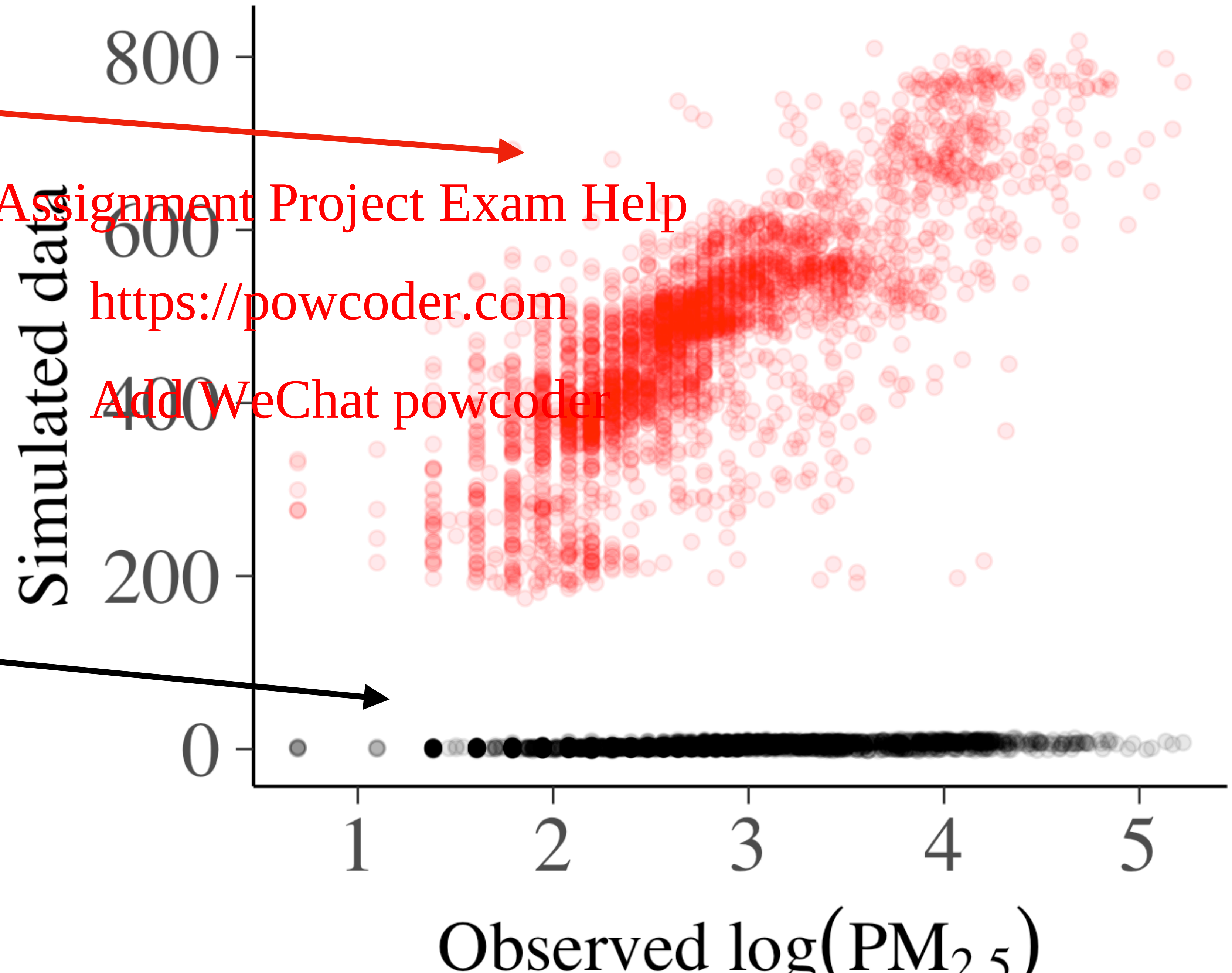
AND MAKE IT EASIER TO DEFEND YOUR MODELLING CHOICES

Non-informative



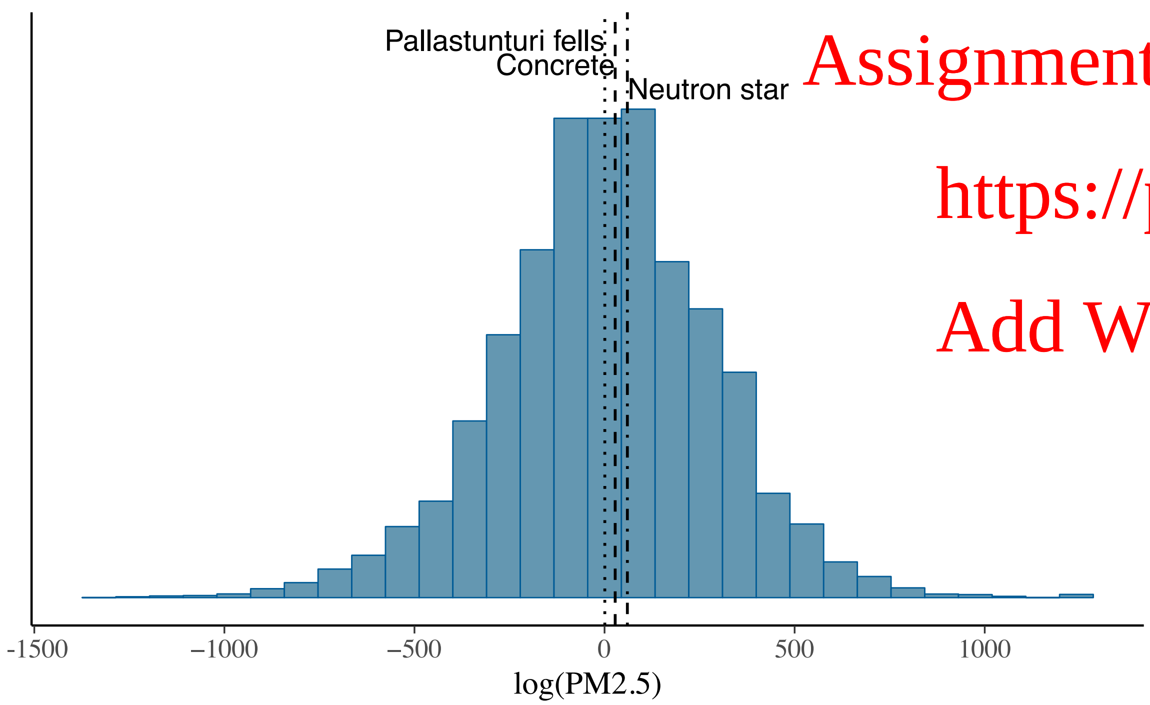
Weakly informative



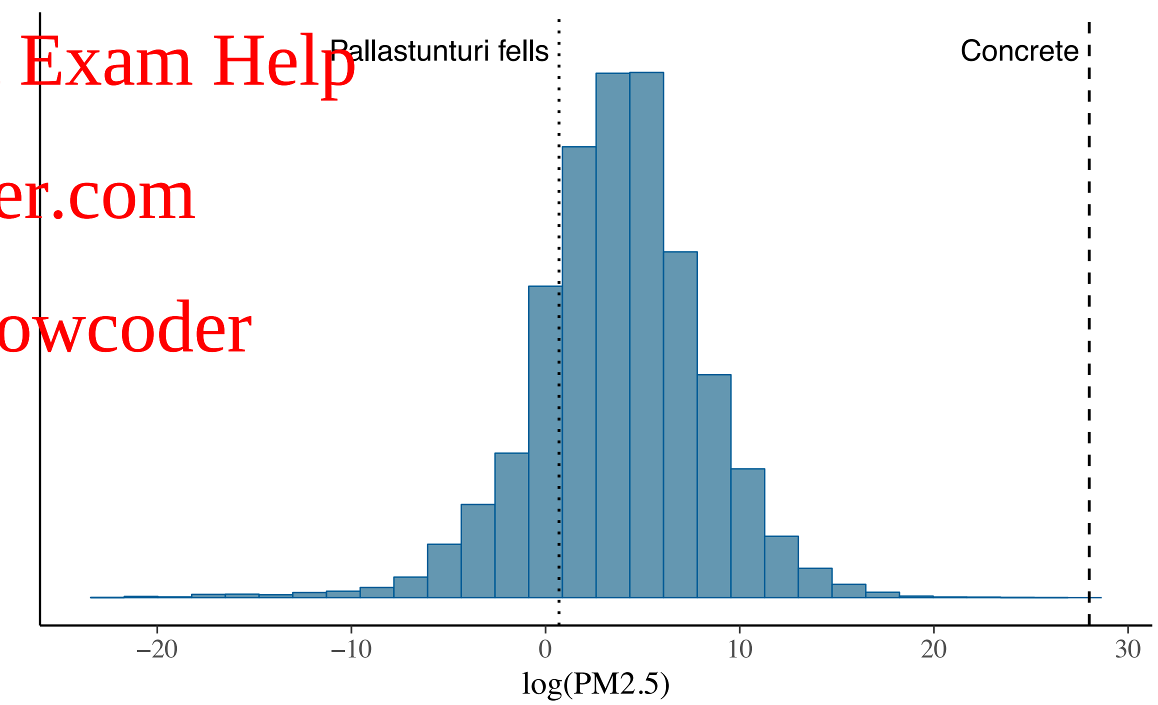


A DIFFERENT VISUALIZATION

Prior predictive distribution with vague prior



Prior predictive distribution with weakly informative prior



SPATIAL STRUCTURES AND INLA

NEXT WEEK…

- Next week we will get into spatial models. We will use INLA for the remainder of the term.

- So far, we have focused on two general topics:
    - Simulation from Bayesian models
    - Maps

- For the rest of the term, we'll put those two together!
    - We'll simulate from models.
    - We'll plot our model results with maps.

GENERAL SPATIAL DATA STRUCTURES

- General spatial data structure: $Z(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d$
- **Areal Data**

- **Geostatistical Data**

- **Point patterns**