

Thursday

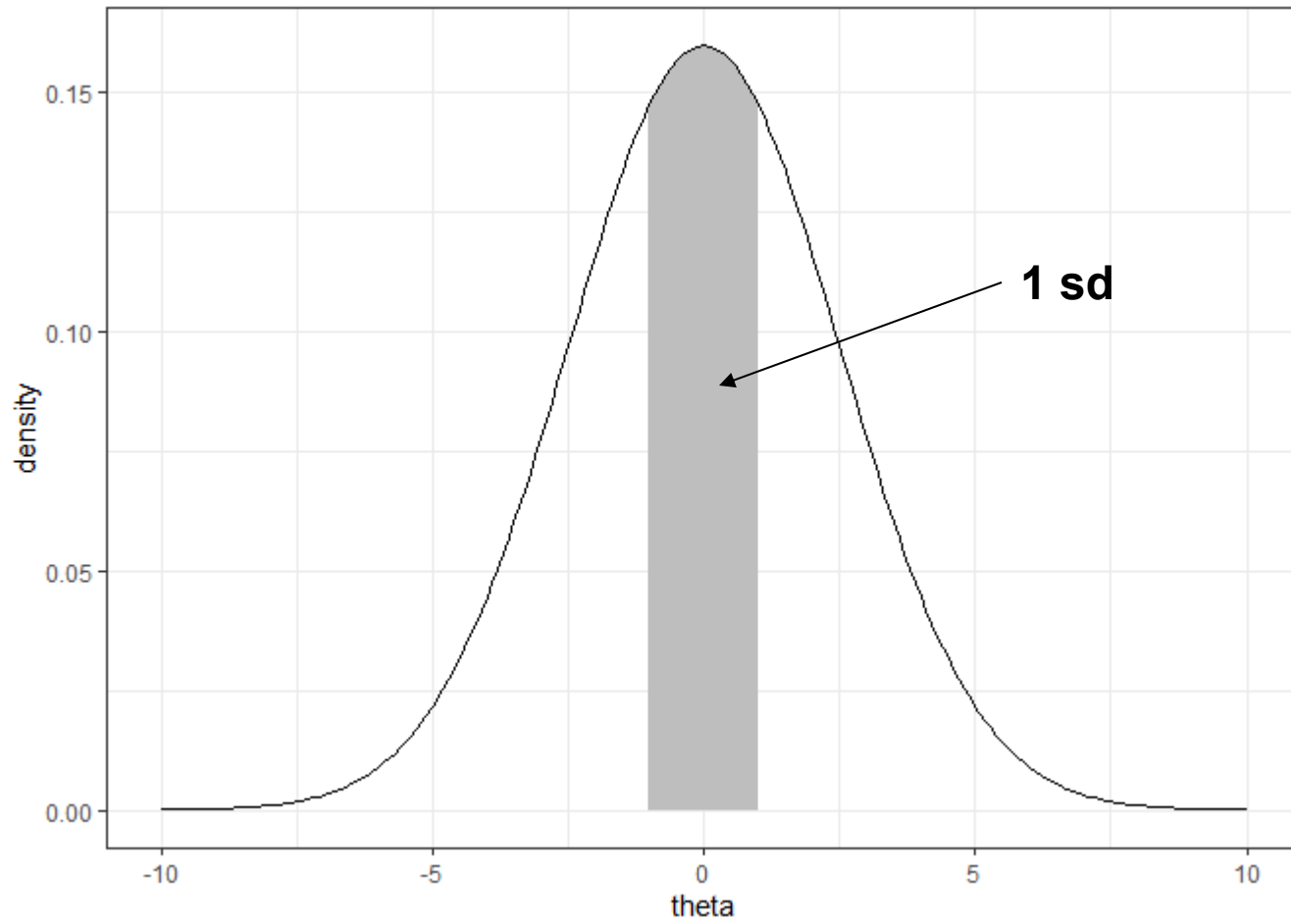
- Individual projects
 - see my initial comment in your repo
 - add more details if necessary (see HW)
 - individual meetings (5 mins)

Today

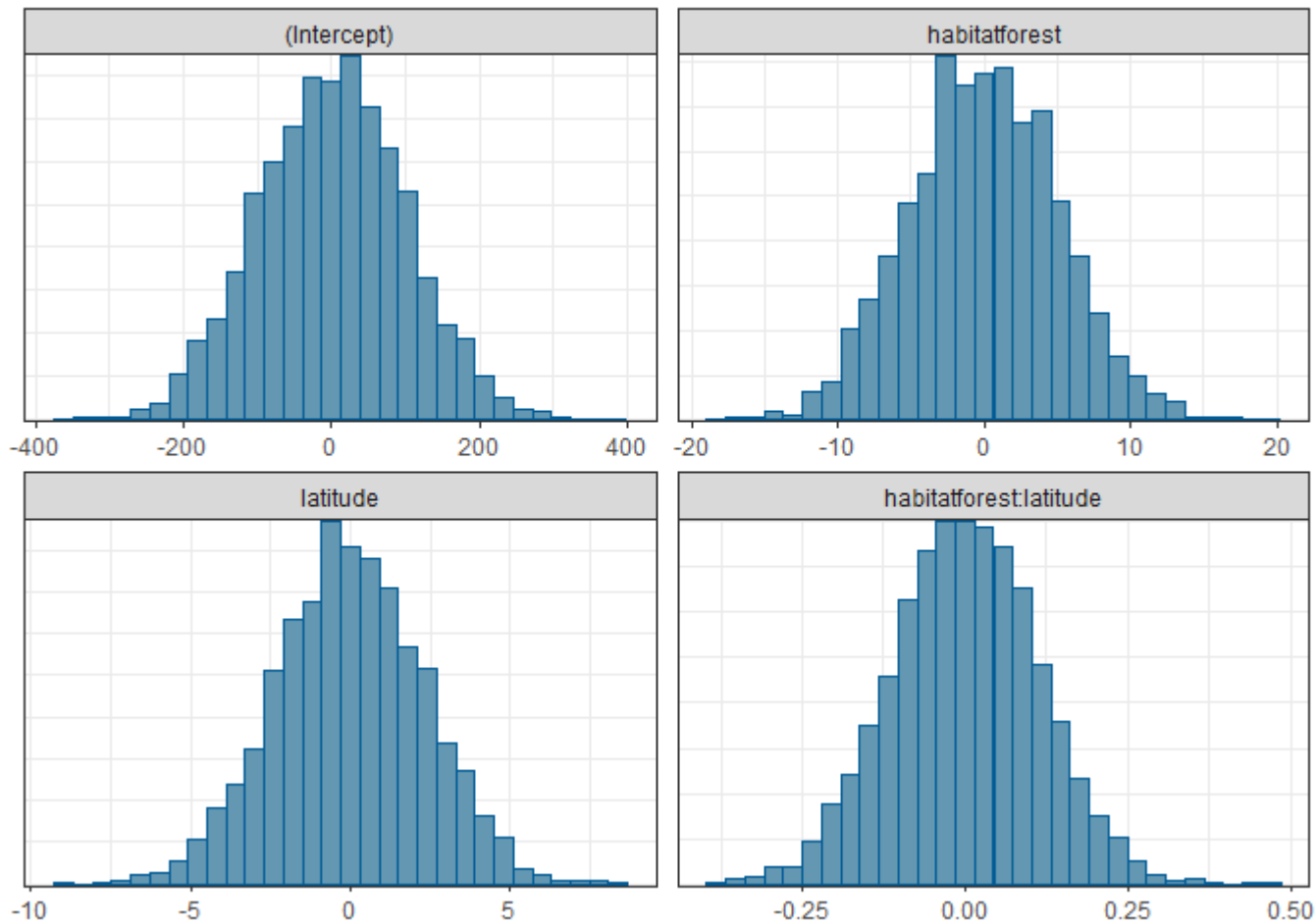
- Priors
- Bayesian model check
- Multilevel models (radon case study)
 - scales and groups
 - study design
 - models and code

Priors

Normal(0, 2.5) weakly informative



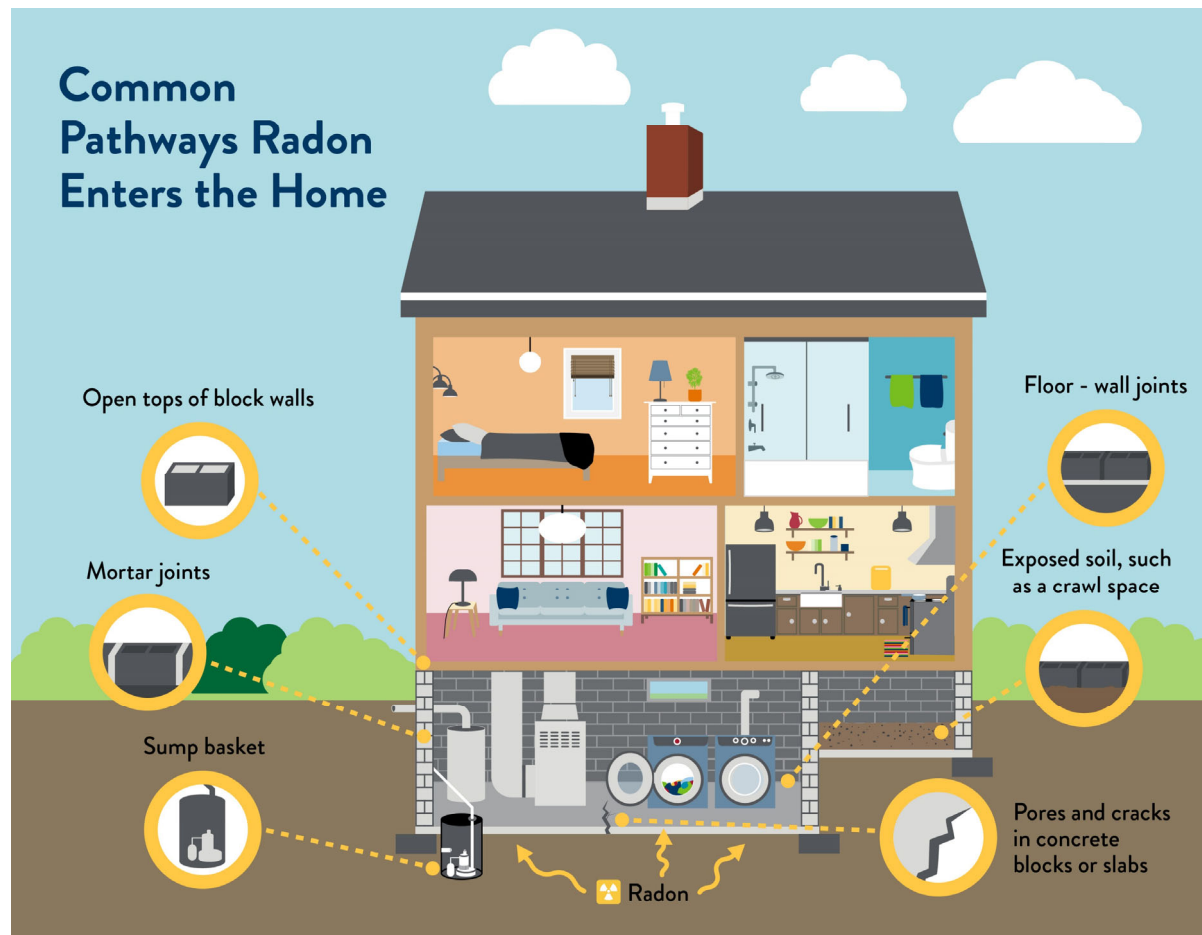
Prior predictive distribution



Bayesian model check

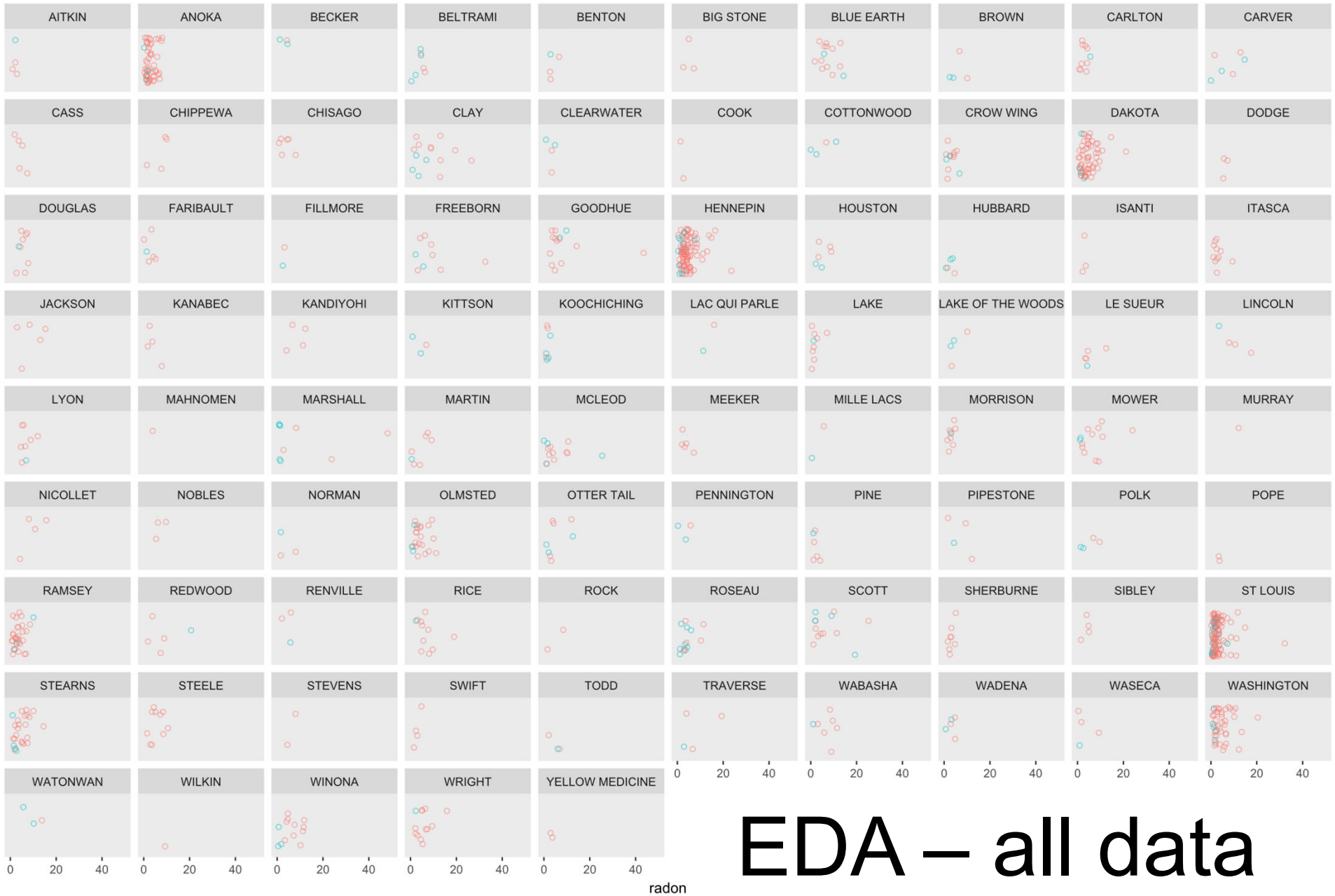
See script

Radon case study



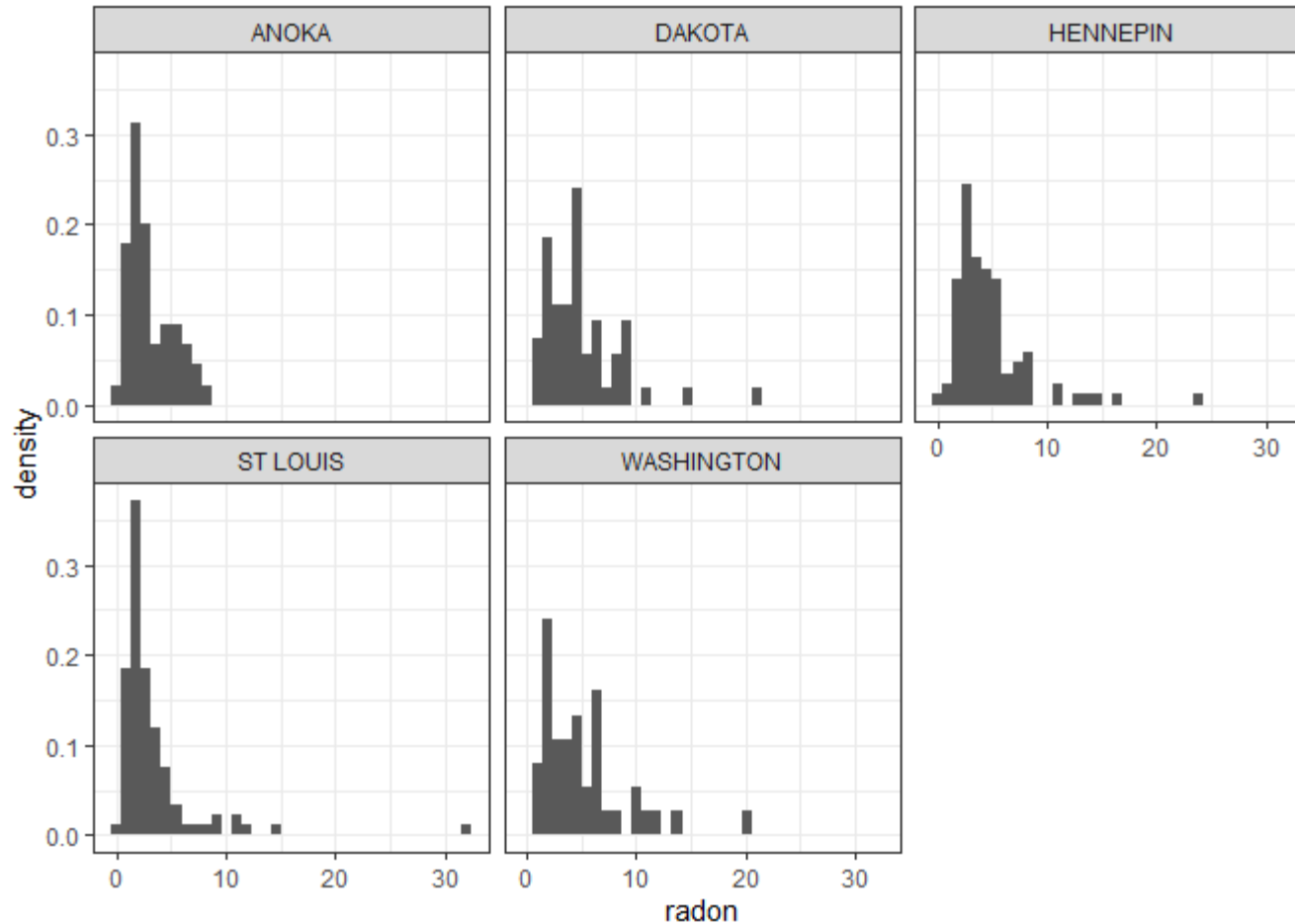
**Expect:
basement more susceptible**

floor ○ basement ○ first

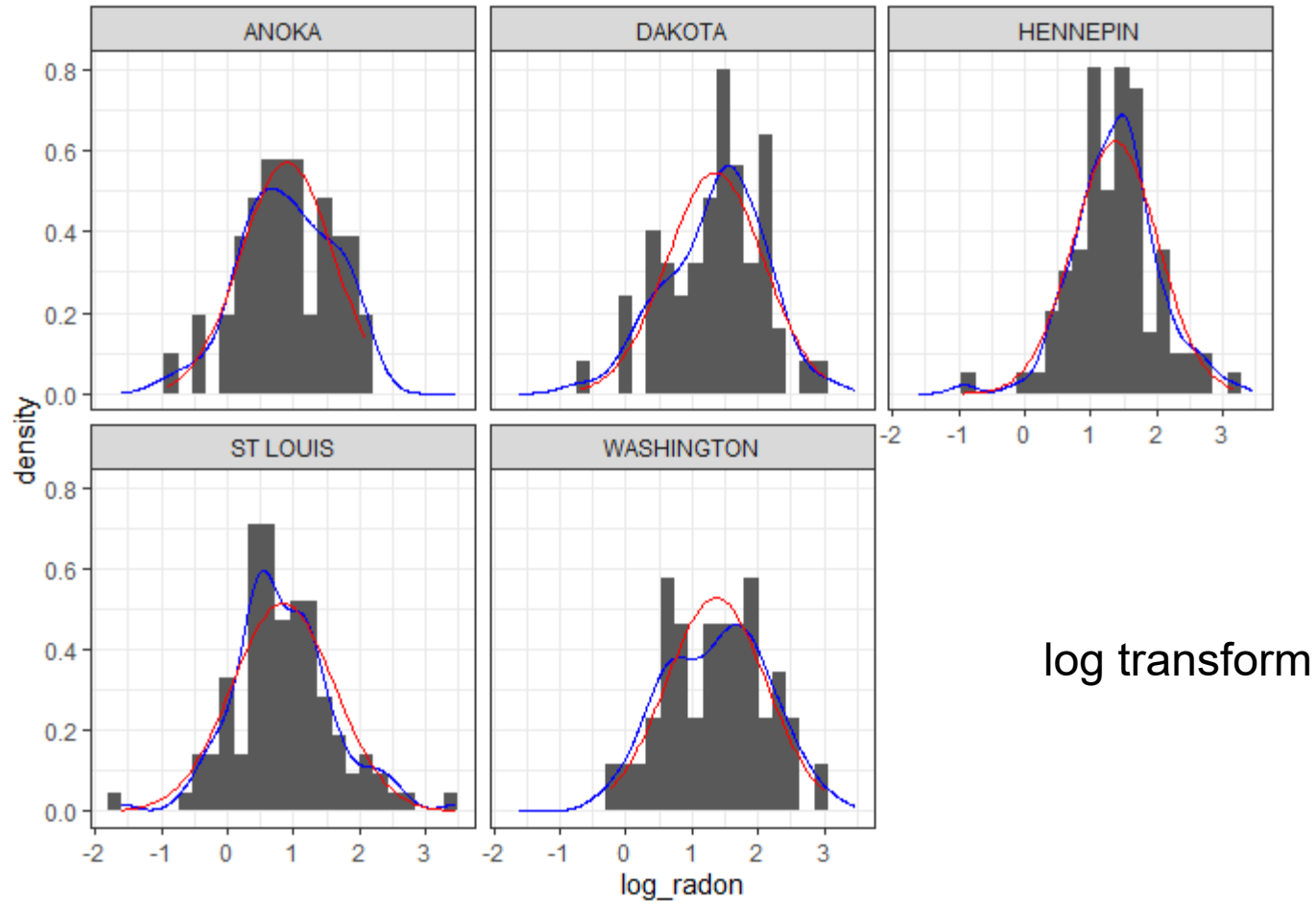


EDA – all data

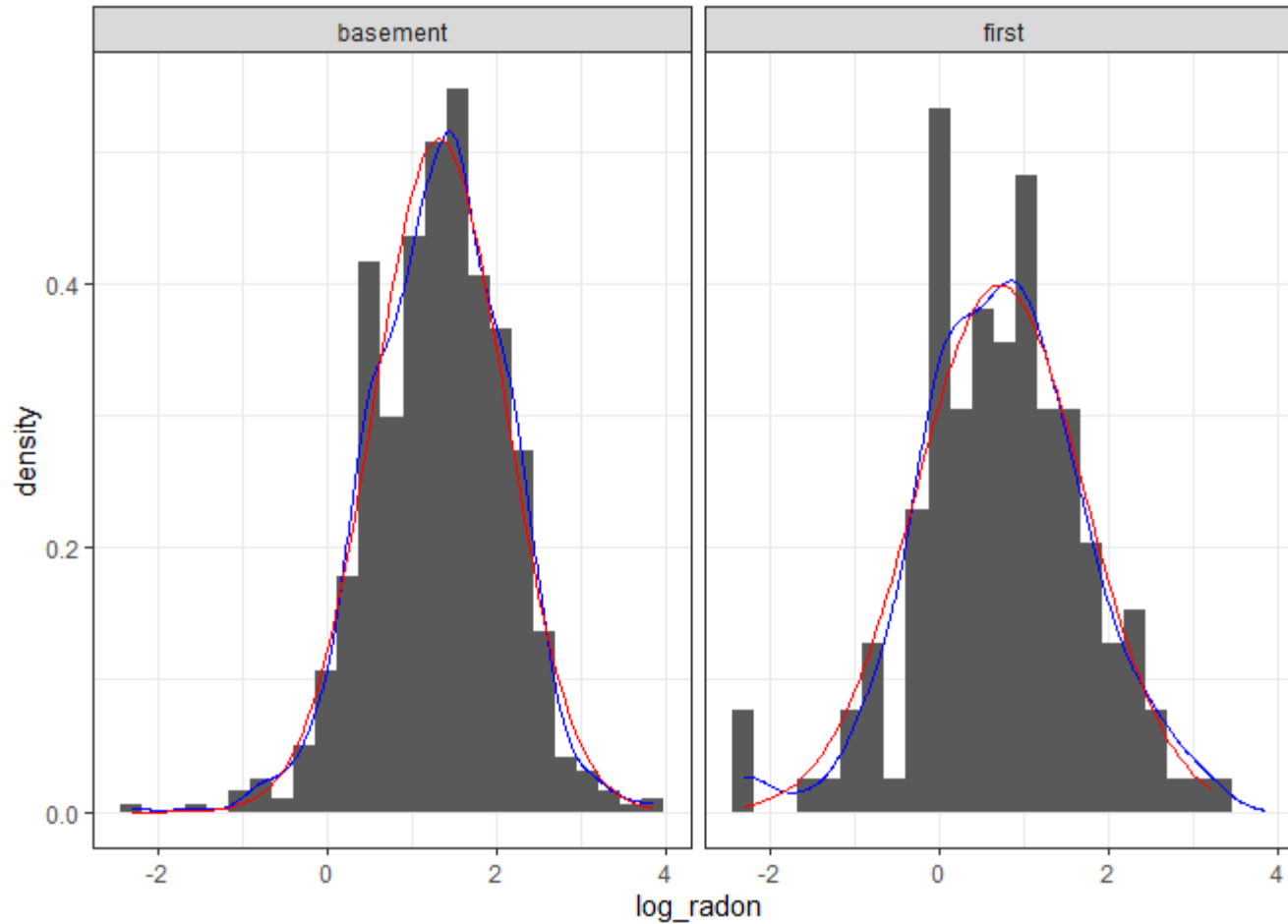
EDA – informal modeling



EDA – informal modeling



EDA – informal modeling



Art of multilevel models

- Sketch data design
- Sketch the data generating process
- Math equations
- Linear model syntax

Scales in radon data

- You identified these scales:
- State, county, house, floor, sensor
- Hierarchical structure: nested
- Data unit: measurement per sensor
- Potential groups: state, county, house, floor, sensor ... but 1 sample per house so house is smallest (unit scale)

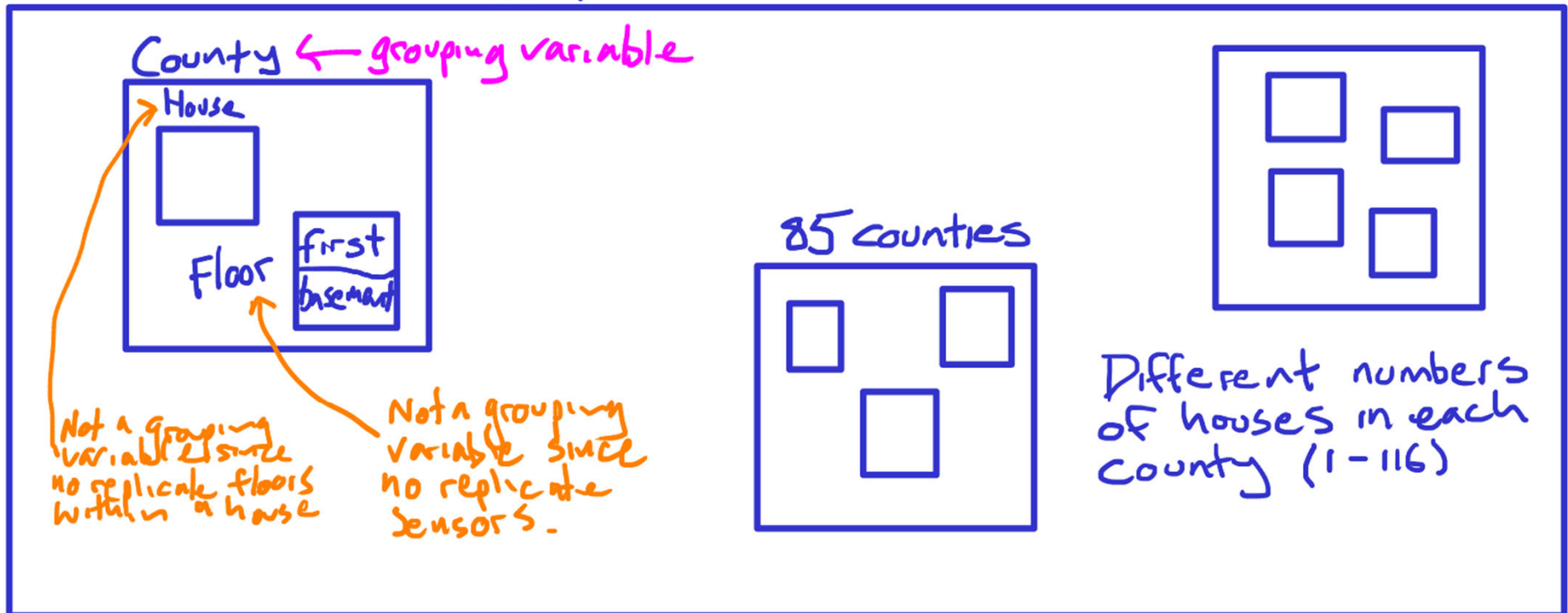
Scope of inference

- Houses in Minnesota ... potentially restricted by whatever criteria were used to include houses in the study
- If random sample of houses then scope would be houses generally in Minnesota
- It clearly wasn't a random sample, so warning: results might be biased
- Limited to basement and 1st floor

Sketch

State = extent

scope of inference



Fixed vs random

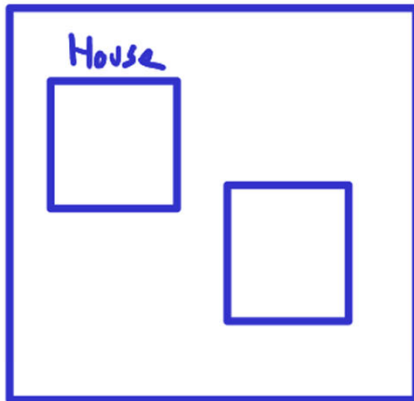
- Floor: fixed effect
 - deterministic model for parameters
 - shared characteristic
- County: random effect
 - stochastic model for parameters
 - model by group (parameters vary by group)
 - model accounts for correlation among houses within counties

We could alternatively model county as a fixed effect: model parameters would be deterministic

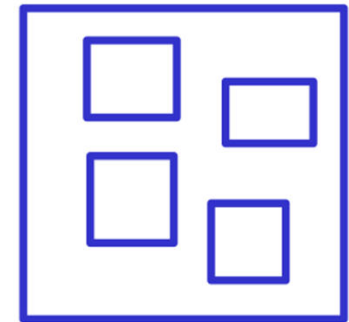
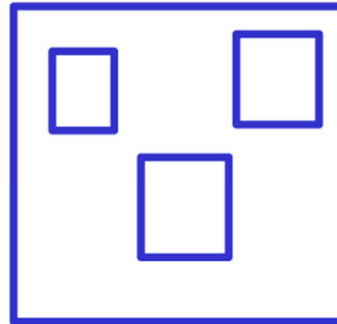
Model 1: random effects

State

County ← grouping variable

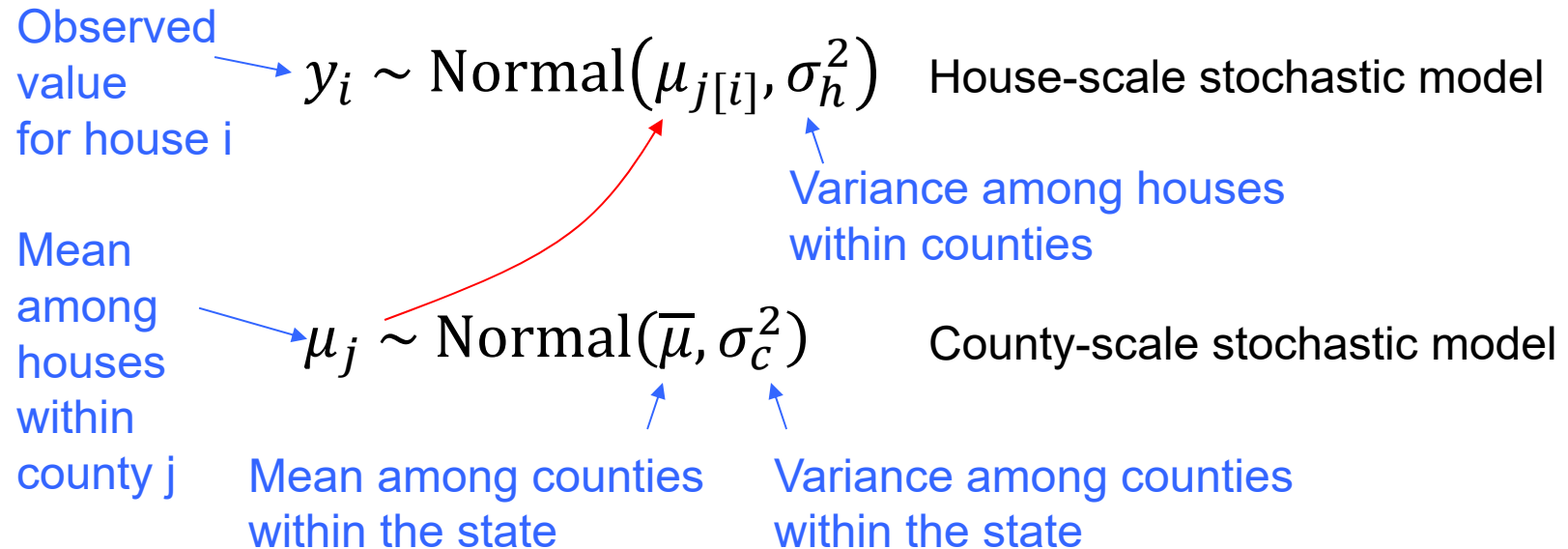


85 counties



Writing model 1

Basic multilevel model, no predictors



$j[i]$ is the county (j) of house i

Simulate model 1

- algorithm

```
set mu_bar
set sigma_c
set sigma_h
set n_c #number of counties
set n_h #number of houses

for j in 1:n_c
  mu_j = normal(mu_bar, sigma_c)

for i in 1:n_h
  y_i = normal(mu_j[i], sigma_h)
```

Data story:

First, county determines
background radiation level

Then, house determines
inside radiation level

Alt parameterization

Additive decomposition

Observed value for house i → $y_i = \bar{\mu} + c_{j[i]} + h_i$ ← Deviation of house i from county j mean (aka residual deviation, e_i)

Mean among counties (aka β_0) → $\bar{\mu}$

Deviation of county j mean from mean among counties → $c_{j[i]}$

h_i

$c_j \sim \text{Normal}(0, \sigma_c^2)$ ← Variance among counties within state

$h_i \sim \text{Normal}(0, \sigma_h^2)$ ← Variance among houses within counties

County-scale stochastic model

House-scale stochastic model

Sim alt parameterization

- algorithm

```
set mu_bar
set sigma_c
set sigma_h
set n_c #number of counties
set n_h #number of houses
```

```
for j in 1:n_c
  c_j = normal(0, sigma_c)
```

```
for i in 1:n_h
  h_i = normal(0, sigma_h)
  y_i = mu_bar + c_j[i] + h_i
```

Linear model syntax

```
log_radon ~ 1 + (1|county)
```

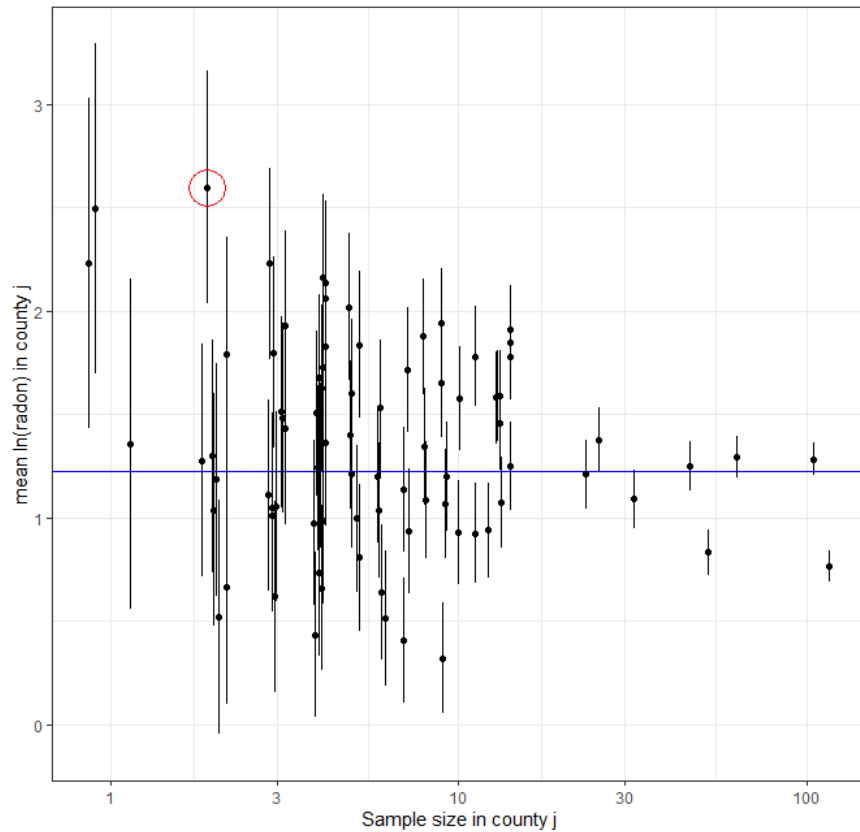
Code for the radon example

radon_multilevel_EDA.md

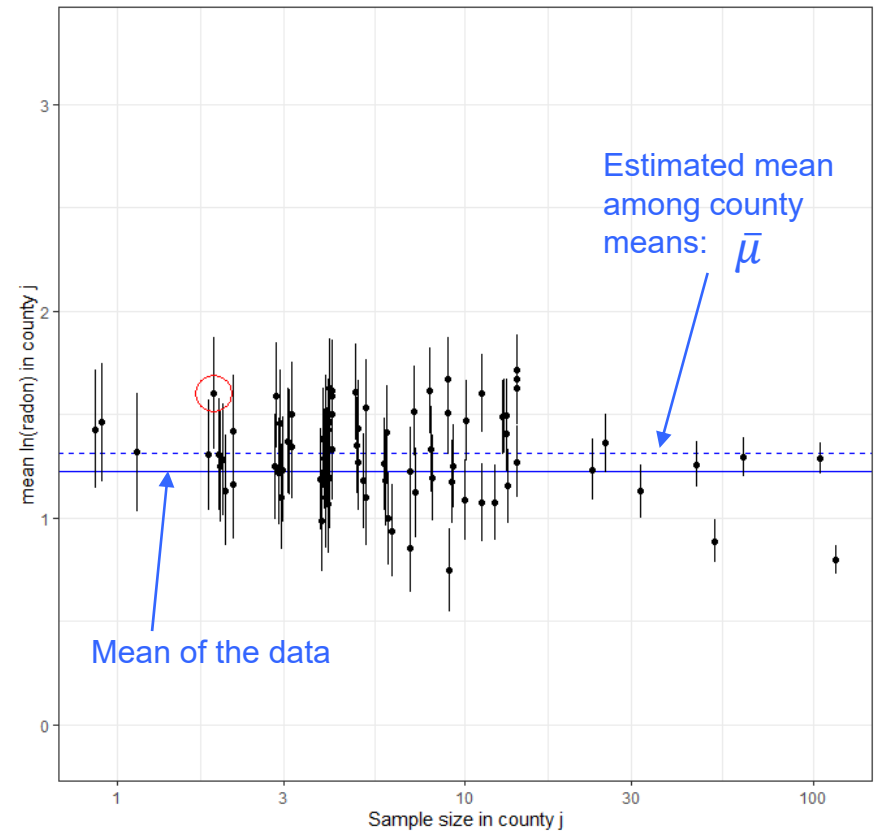
radon_multilevel_M1.md

Multilevel model - radon

No pooling: estimates from linear fixed-effects model



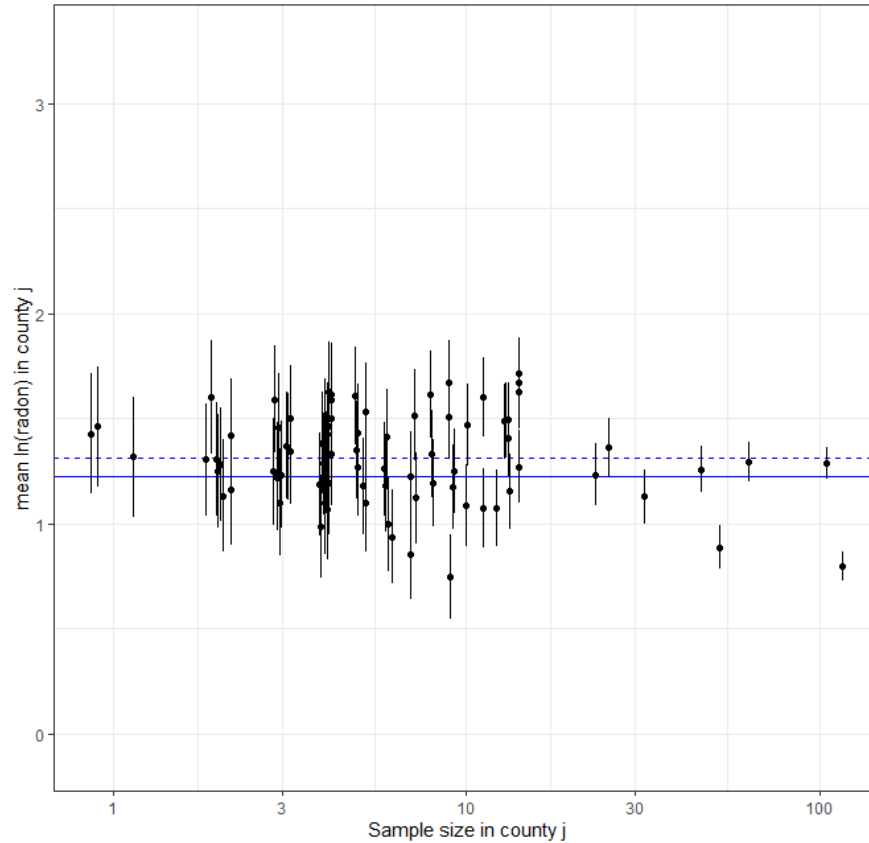
Partial pooling: multilevel model, max likelihood



Shrinkage

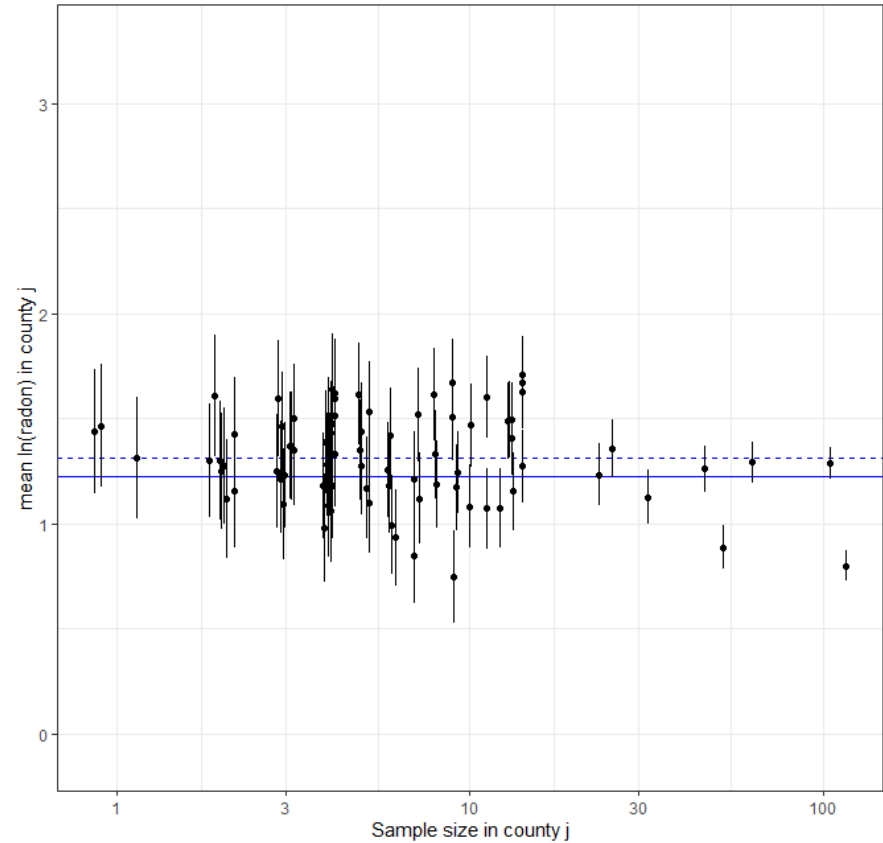
Multilevel model - radon

Partial pooling: multilevel model, max likelihood



Maximum likelihood

Partial pooling: multilevel model, Bayesian



Bayesian