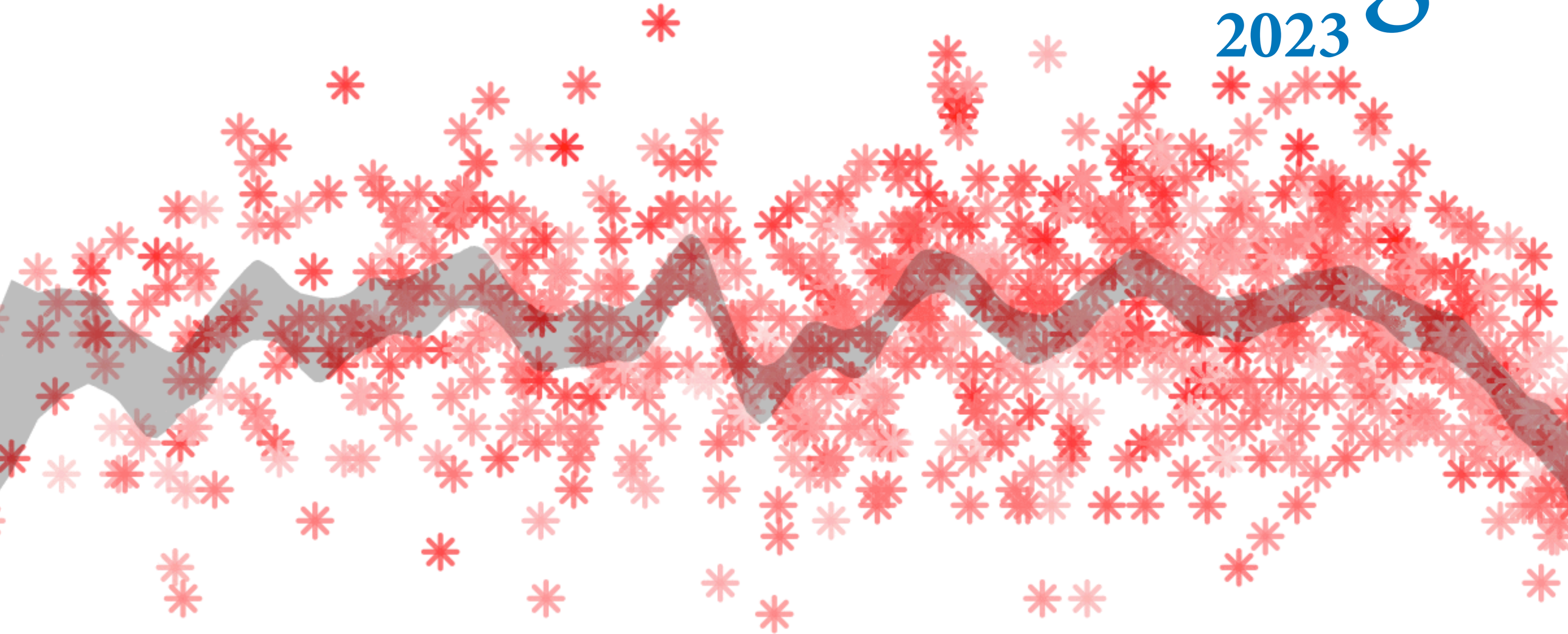
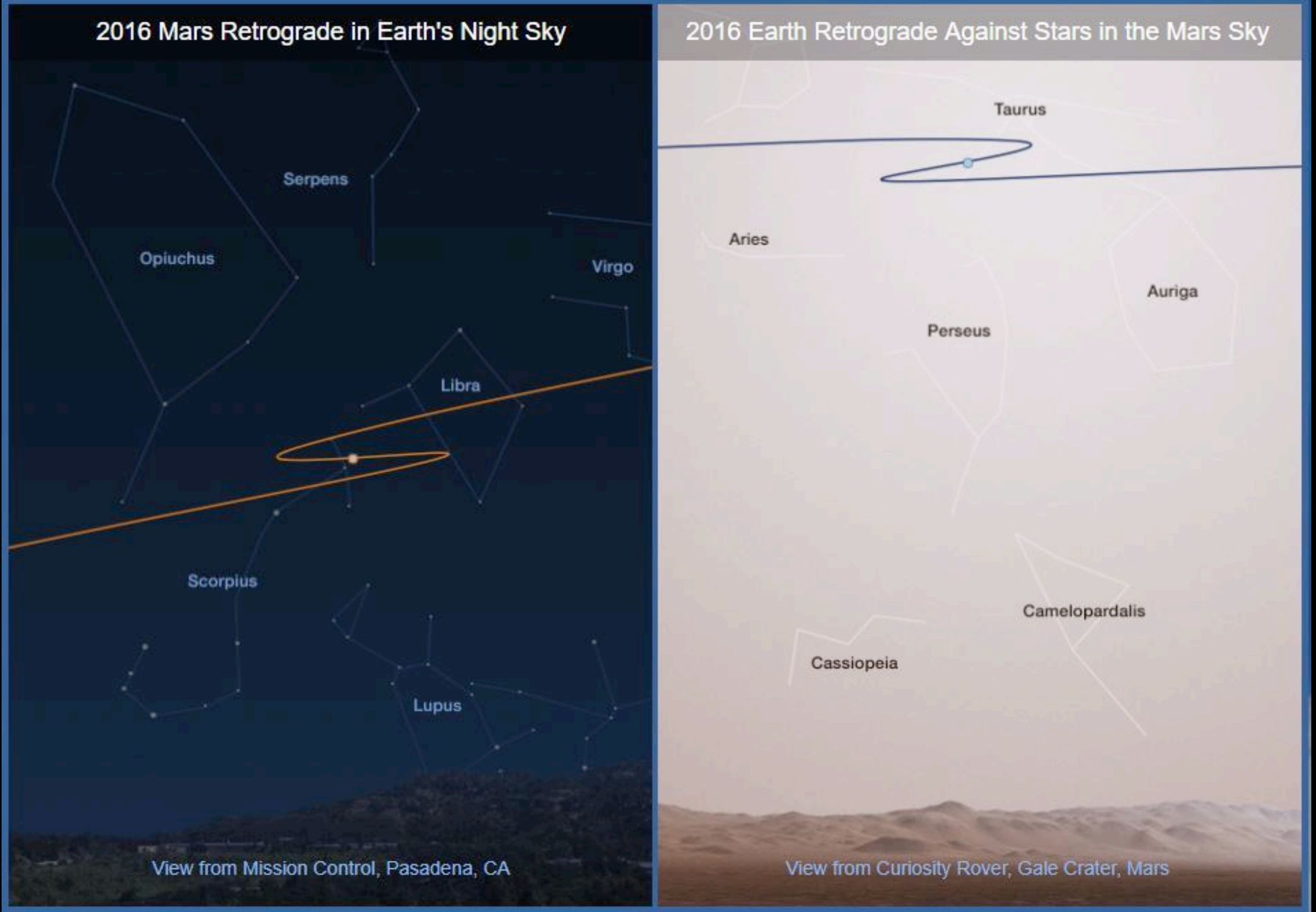
Statistical Rethinking

2023



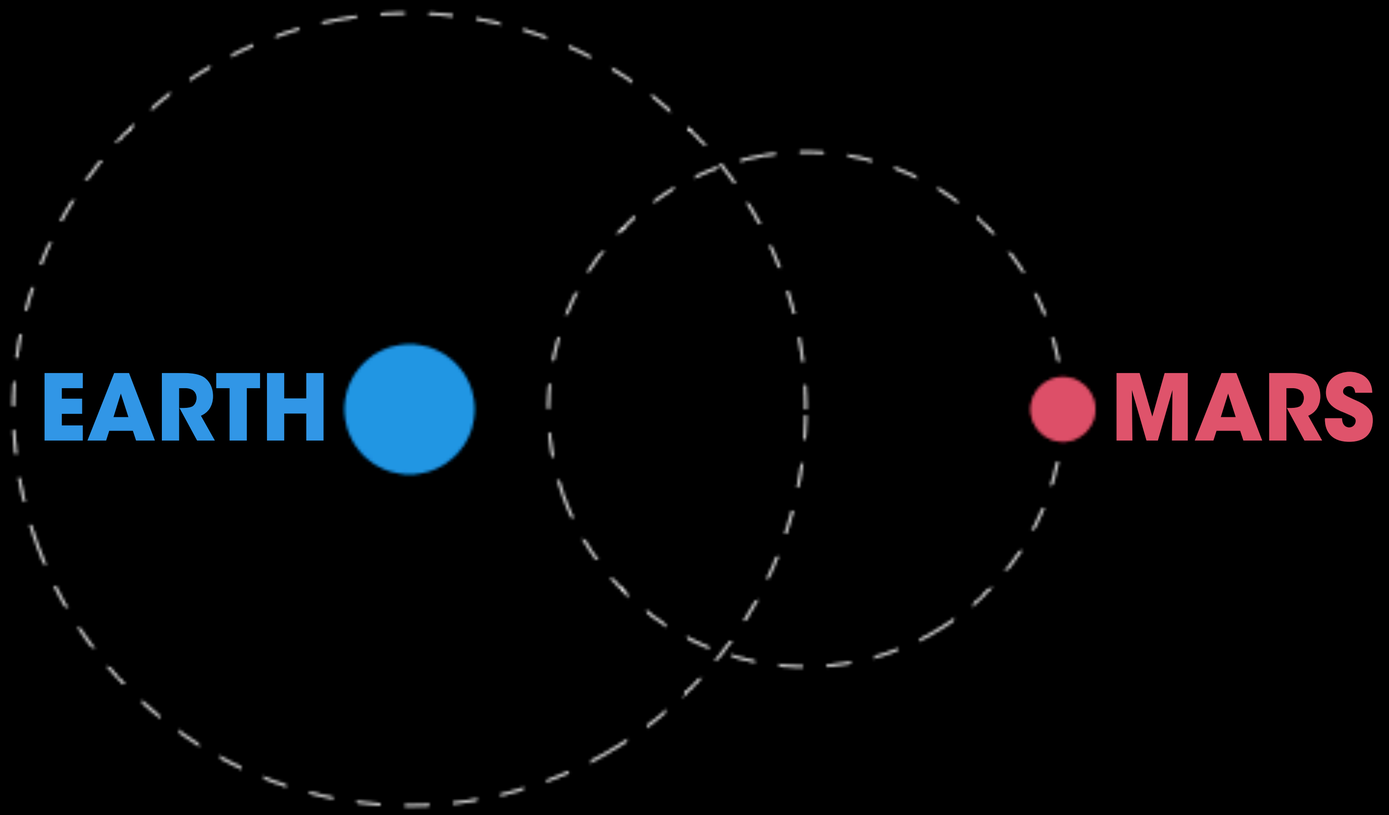
3. Geocentric Models

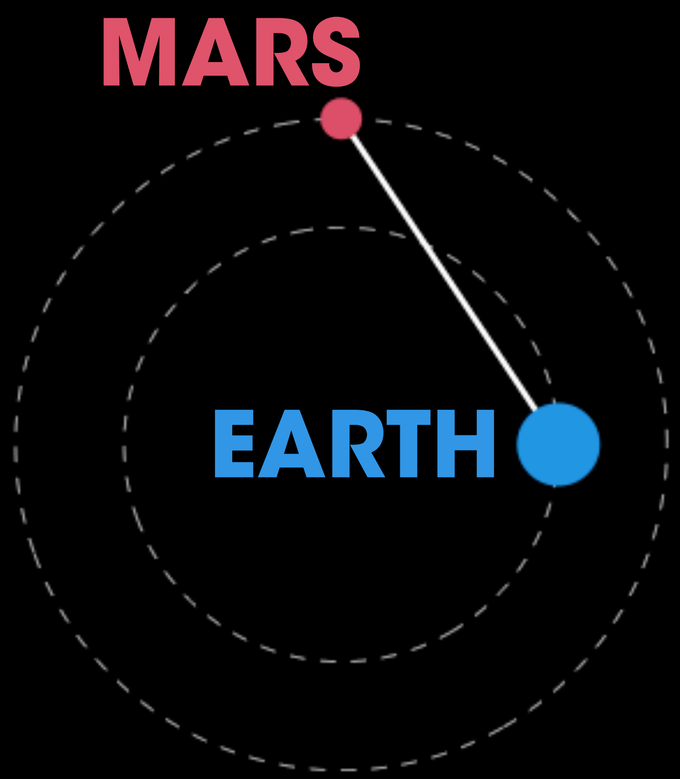
Mars: June 2020 until Feb 2021 — Tunç Tezel — https://vimeo.com/user48630149

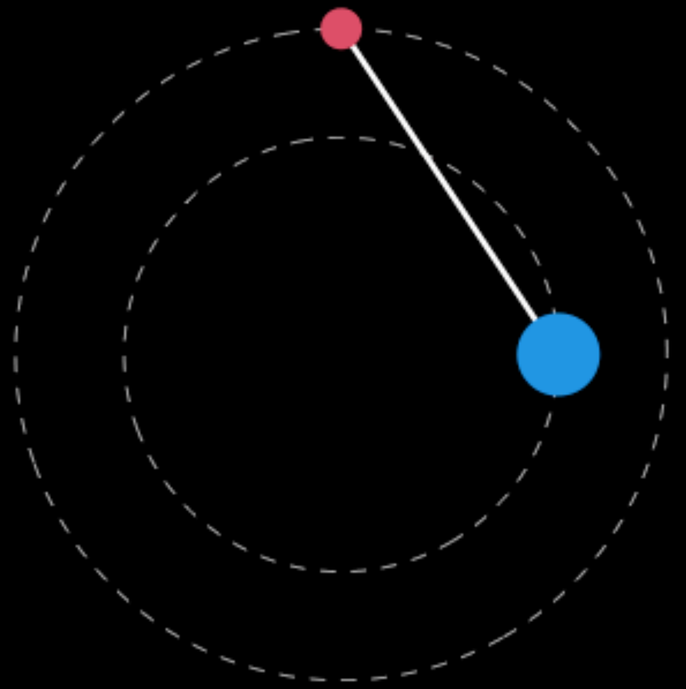

2016 Mars Retrograde in Earth's Night Sky
Serpens
Opiuchus
Virgo
Libra
Scorpius
Lupus
View from Mission Control, Pasadena, CA
2016 Earth Retrograde Against Stars in the Mars Sky
Taurus
Aries
Auriga
Perseus
Camelopardalis
Cassiopeia
View from Curiosity Rover, Gale Crater, Mars

Geocentric Model

Prediction
Without
Explanation

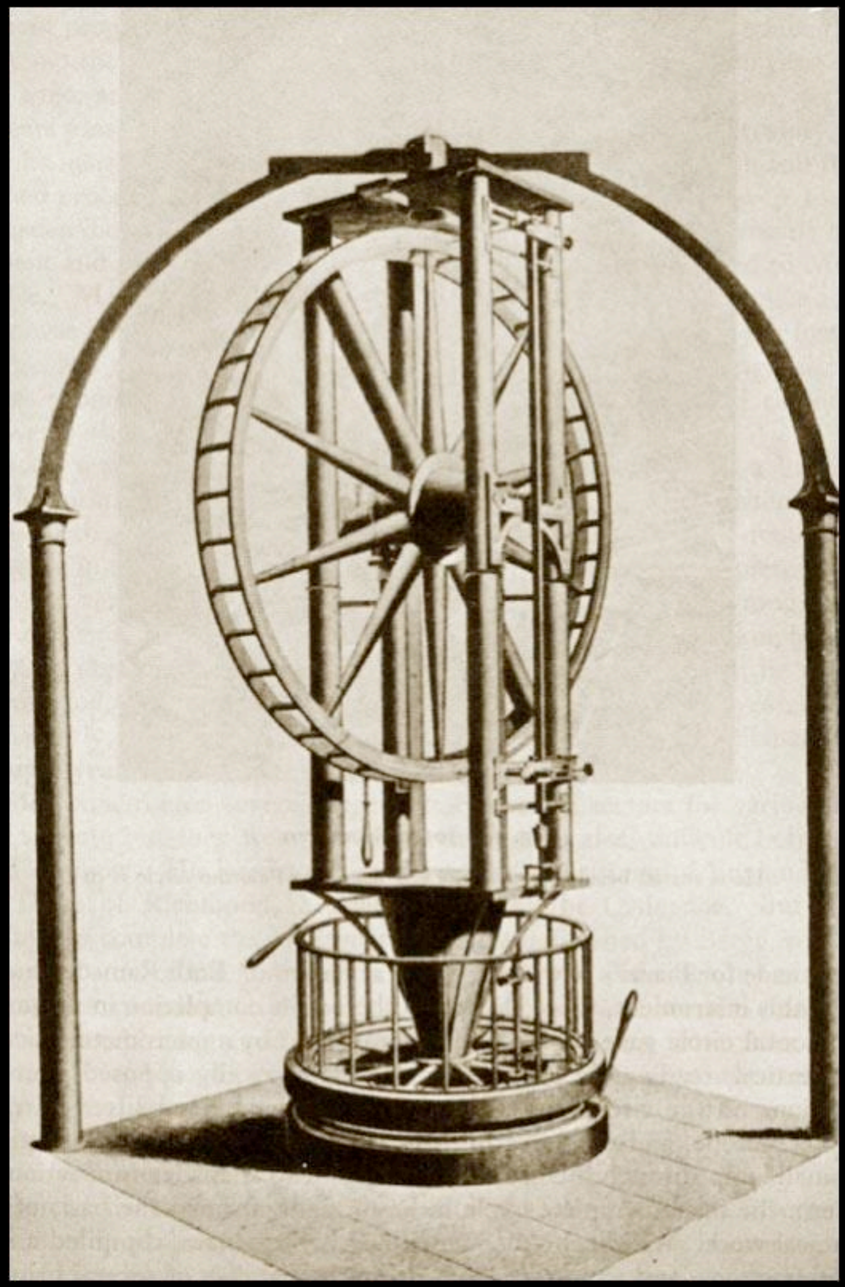



MARS
EARTH

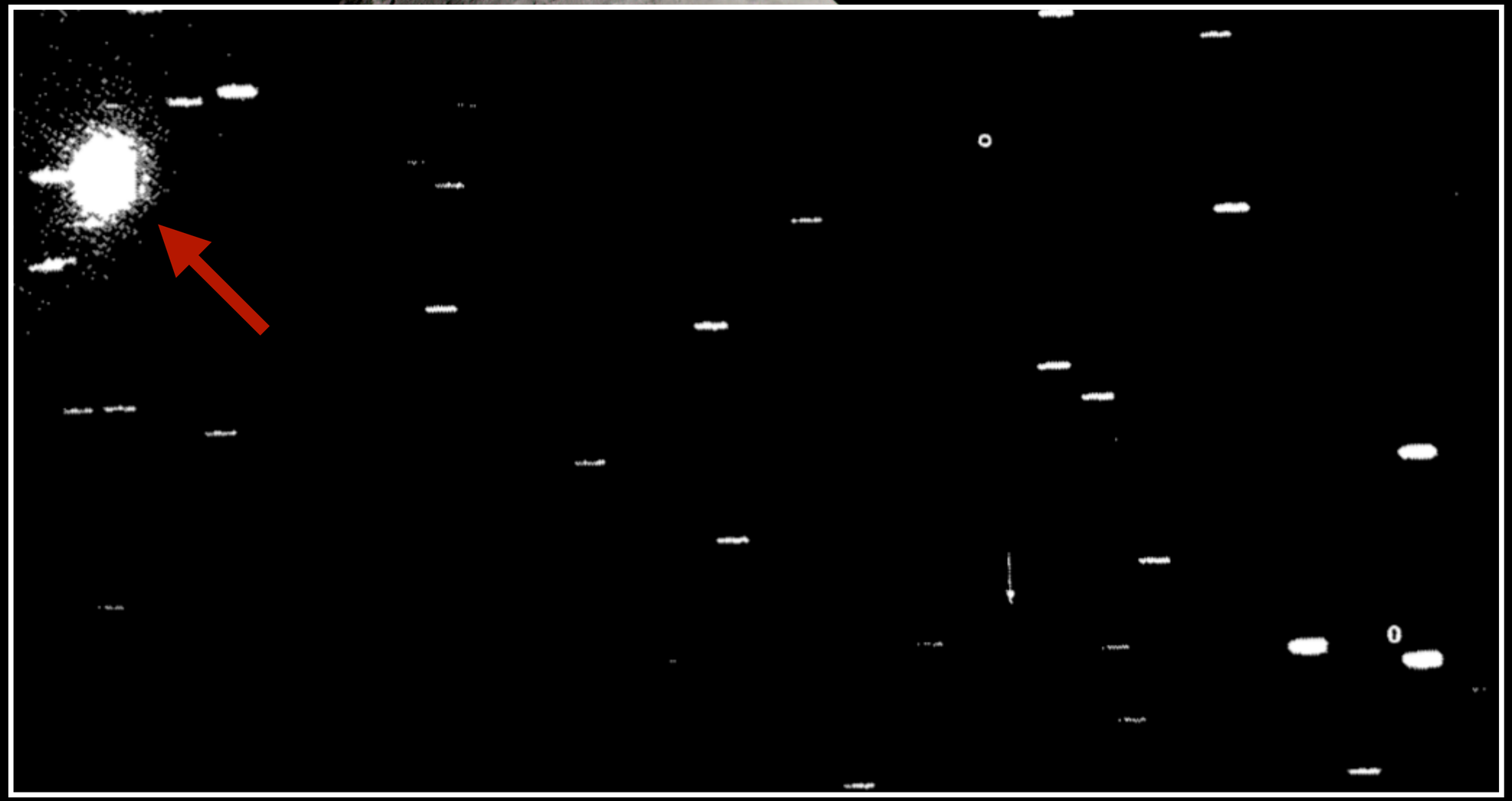


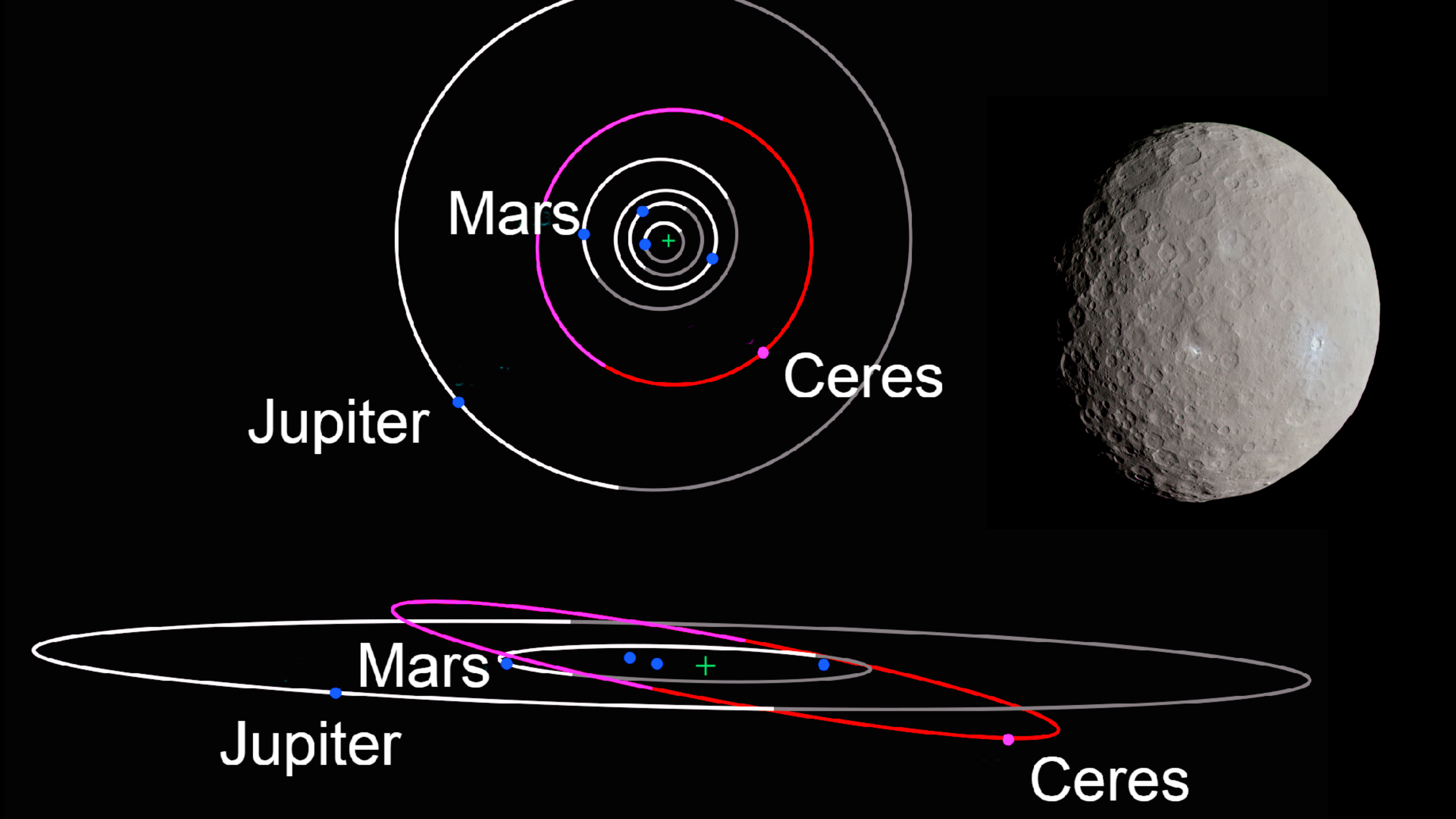
1.1.1801

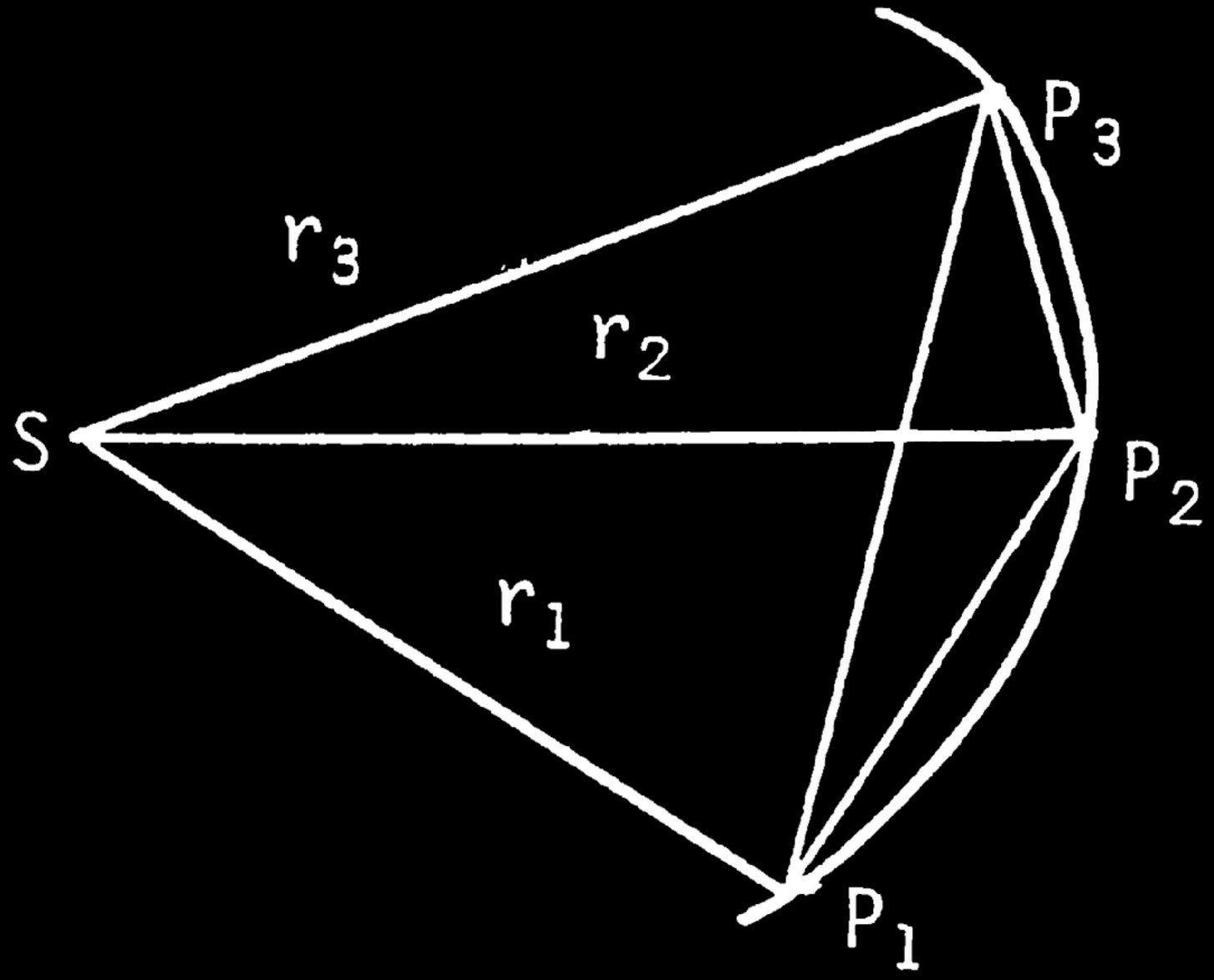
Giuseppe Piazzi 1746–1826



Palermo Circle




Mars
Ceres
Jupiter
Mars
Jupiter
Ceres



Carl Friedrich Gauss 1777–1855 (portrait in 1803)

THEORIA
MOTVS CORPORVM
COELESTIVM

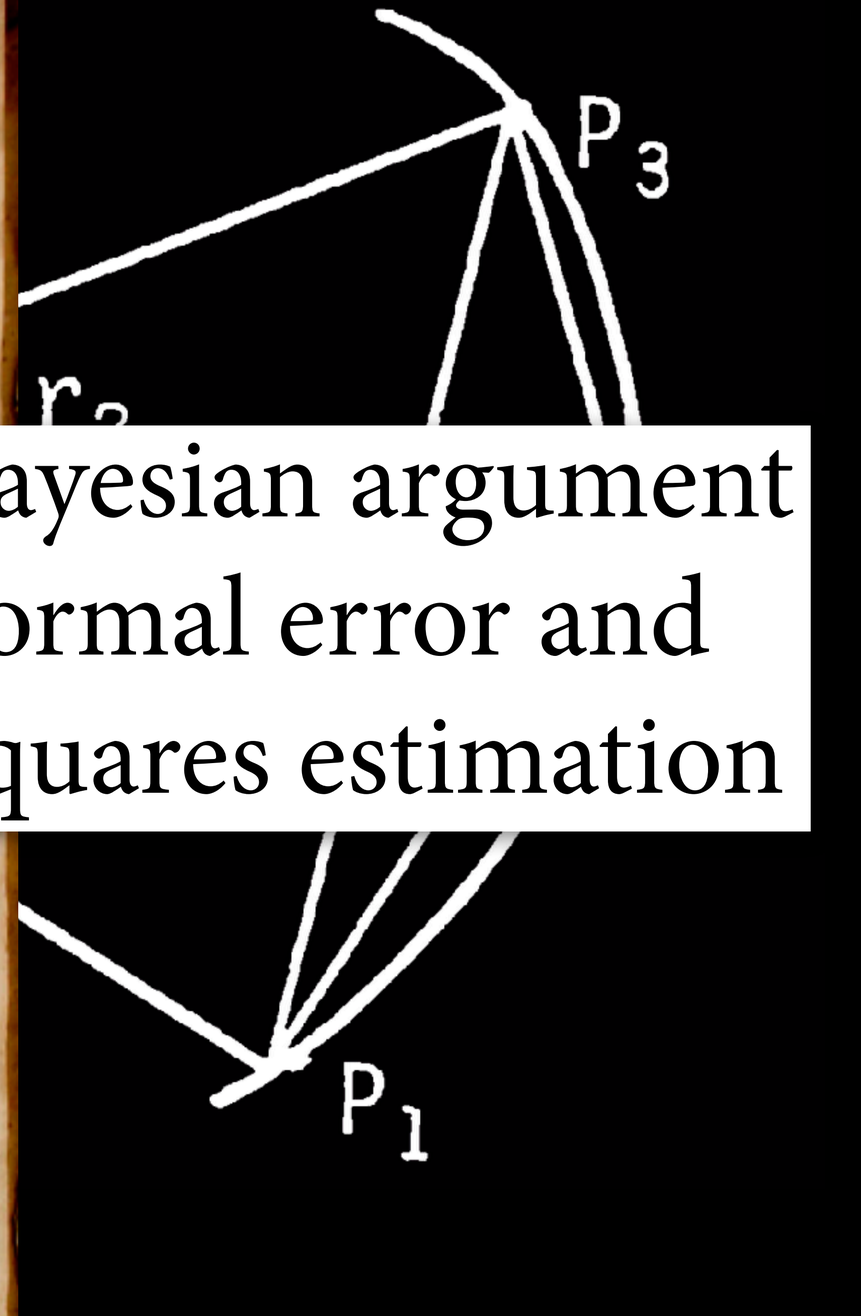
IN

SECTIONIBVS CONICIS SOLEM AMBIENTIVM

AVCTORE

CAROLO FRIDERICO GAVSS

HAMBVRGI SVMTIBVS FRID. PERTHES ET I. H. BESSER
1809.



1809 Bayesian argument for normal error and least-squares estimation

TH

MOTVS

COE

SECTIONIBVS CO

CAROLO F

HAMBVRGI SVMTIBVS FRID. PE

1809.

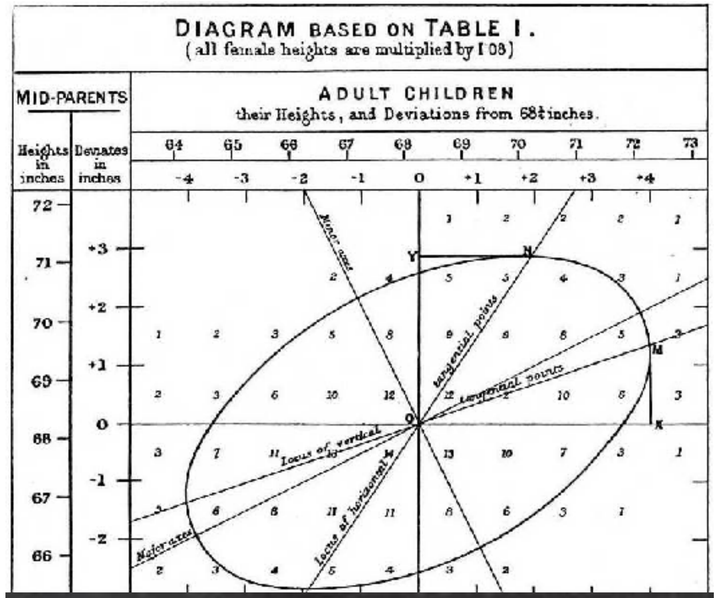
Linear Regression

Geocentric: Describes associations, makes predictions, but mechanistically wrong

Gaussian: Abstracts from generative error model, replaces with normal distribution, mechanistically silent

Useful when handled with care

Many special cases: ANOVA, ANCOVA, t-test, others

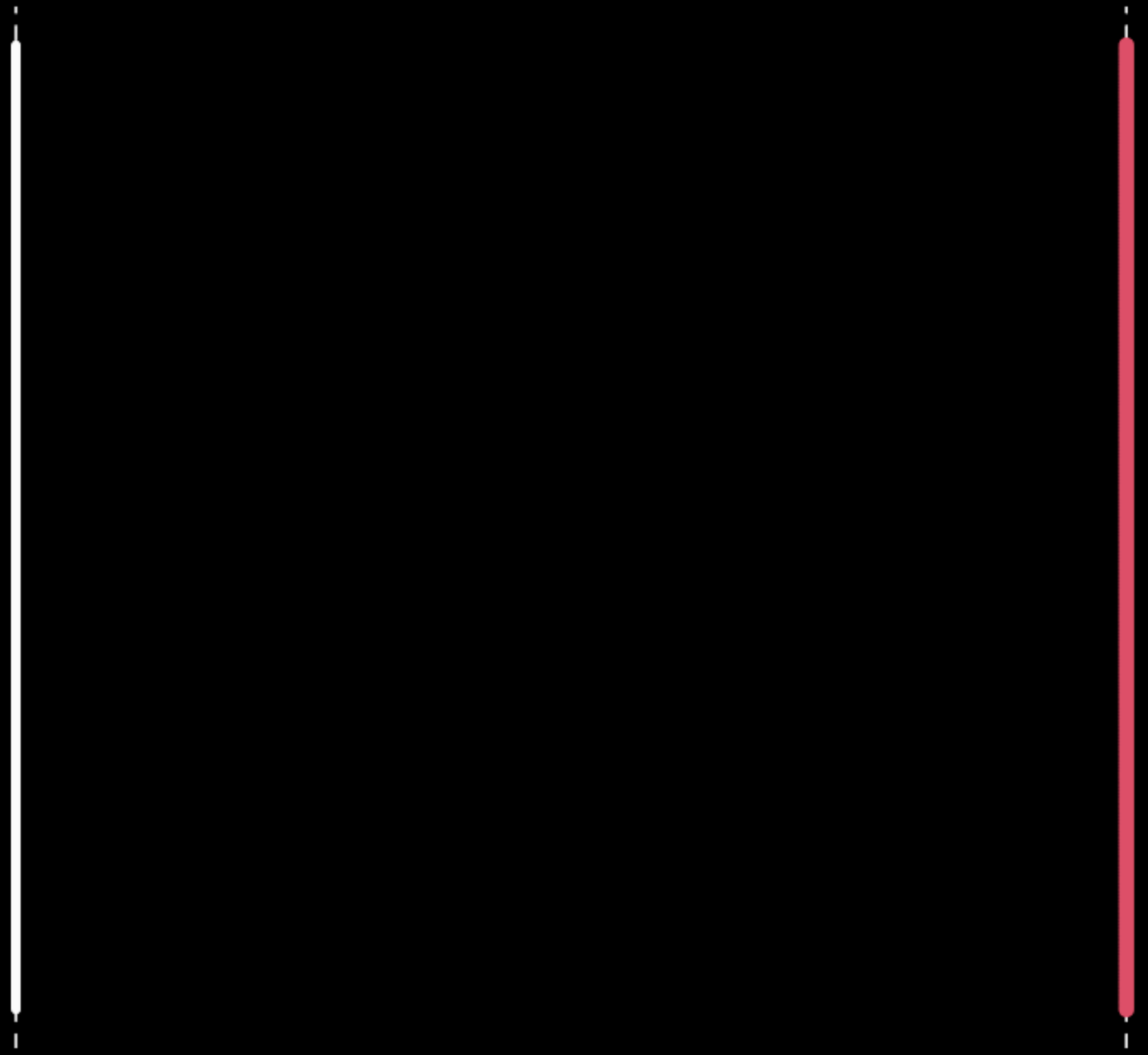


From *Breath of Bones: A Tale of the Golem*

Positions

Positions

Distribution



Why Normal?

Two arguments

(1) Generative: Summed fluctuations tend towards normal distribution

(2) Inferential: For estimating mean and variance, normal distribution is least informative distribution (maxent)

Variable does not have to be normally distributed for normal model to be useful. It's a machine for estimating mean/variance.

Making Geocentric Models

Skill development goals:

(1) Language for representing models

(2) Calculate posterior distributions with multiple unknowns

(3) Constructing & understanding linear models


FLOW

Owl-drawing workflow

(1) State a clear question

(2) Sketch your causal assumptions

(3) Use the sketch to define a generative model

(4) Use generative model to build estimator

(5) Profit

Linear Regression

Drawing the Owl

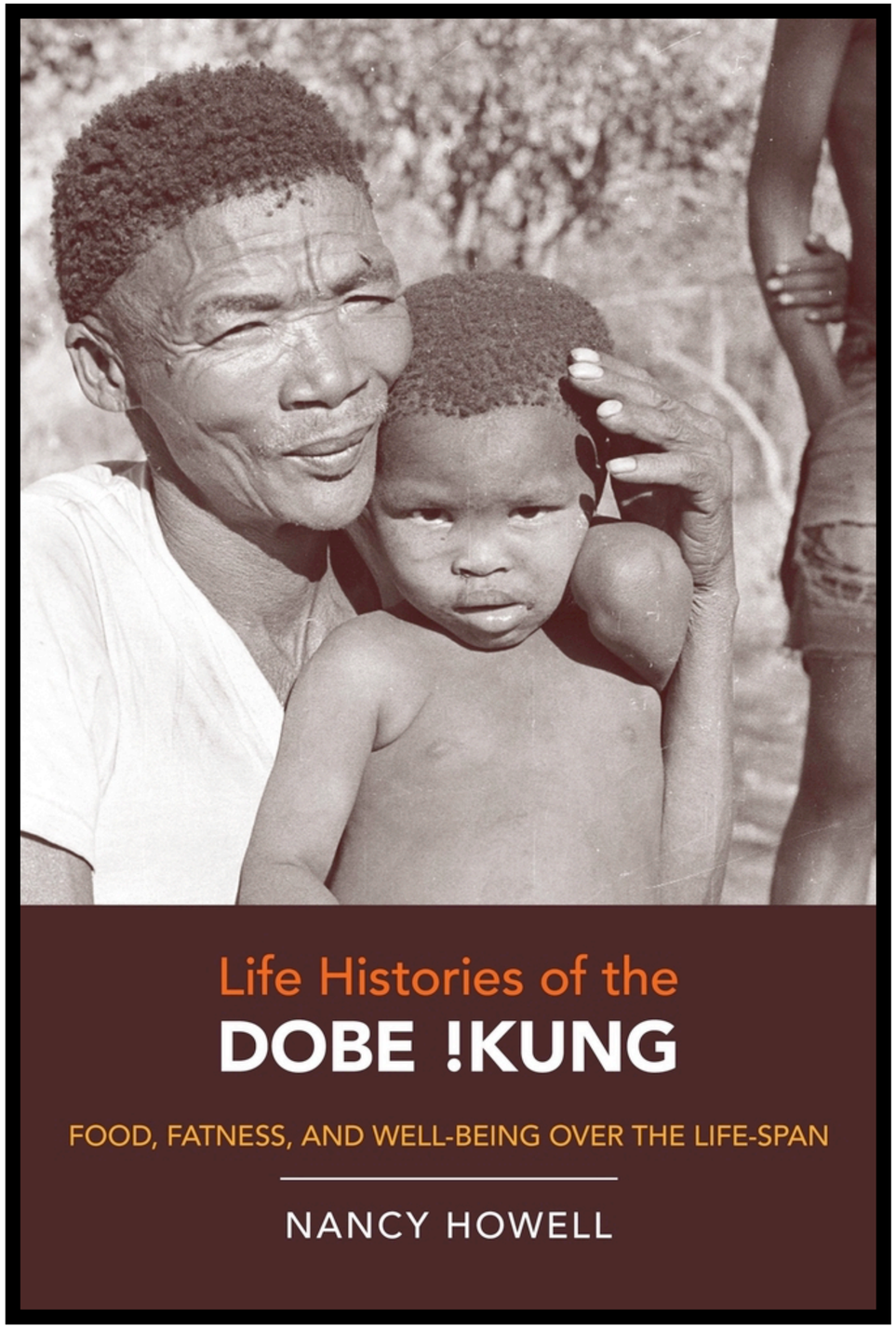
(1) Question/goal/estimand

(2) Scientific model

(3) Statistical model(s)

(4) Validate model

(5) Analyze data



Linear Regression

```
library(rethinking)
data(Howell1)
```

Drawing the Owl

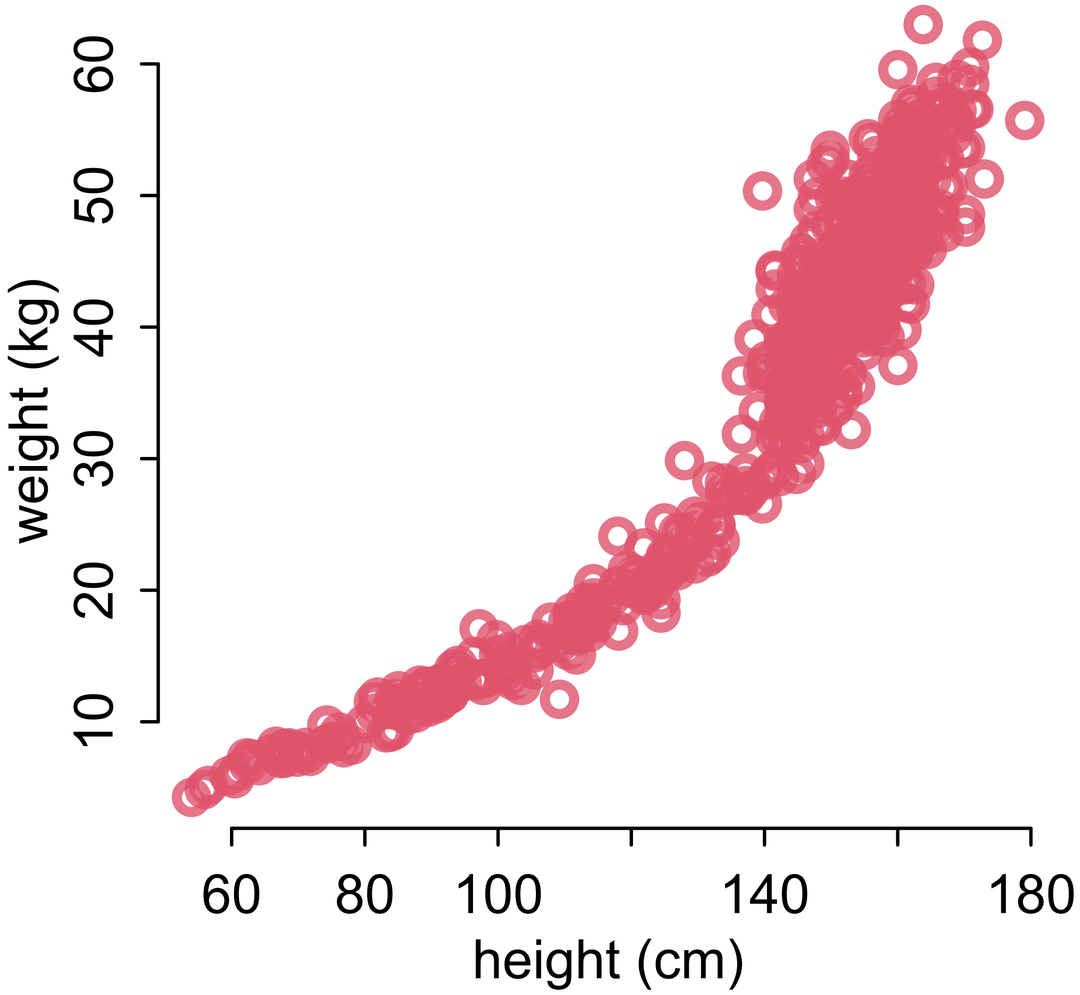
(1) Question/goal/estimand

(2) Scientific model

(3) Statistical model(s)

(4) Validate model

(5) Analyze data



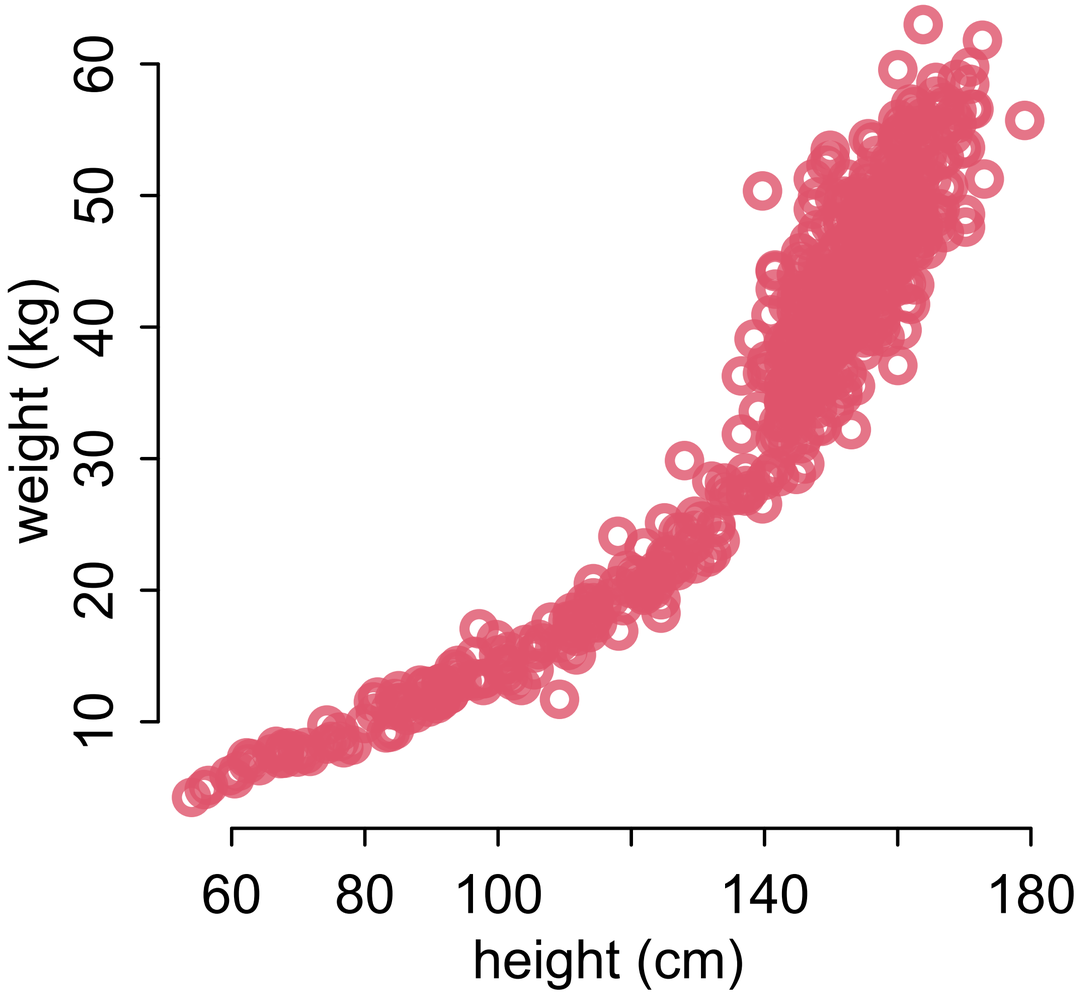
Linear Regression

```
library(rethinking)
data(Howell1)
```

Drawing the Owl

(1) Question/goal/estimand

Describe association between **weight** and **height**



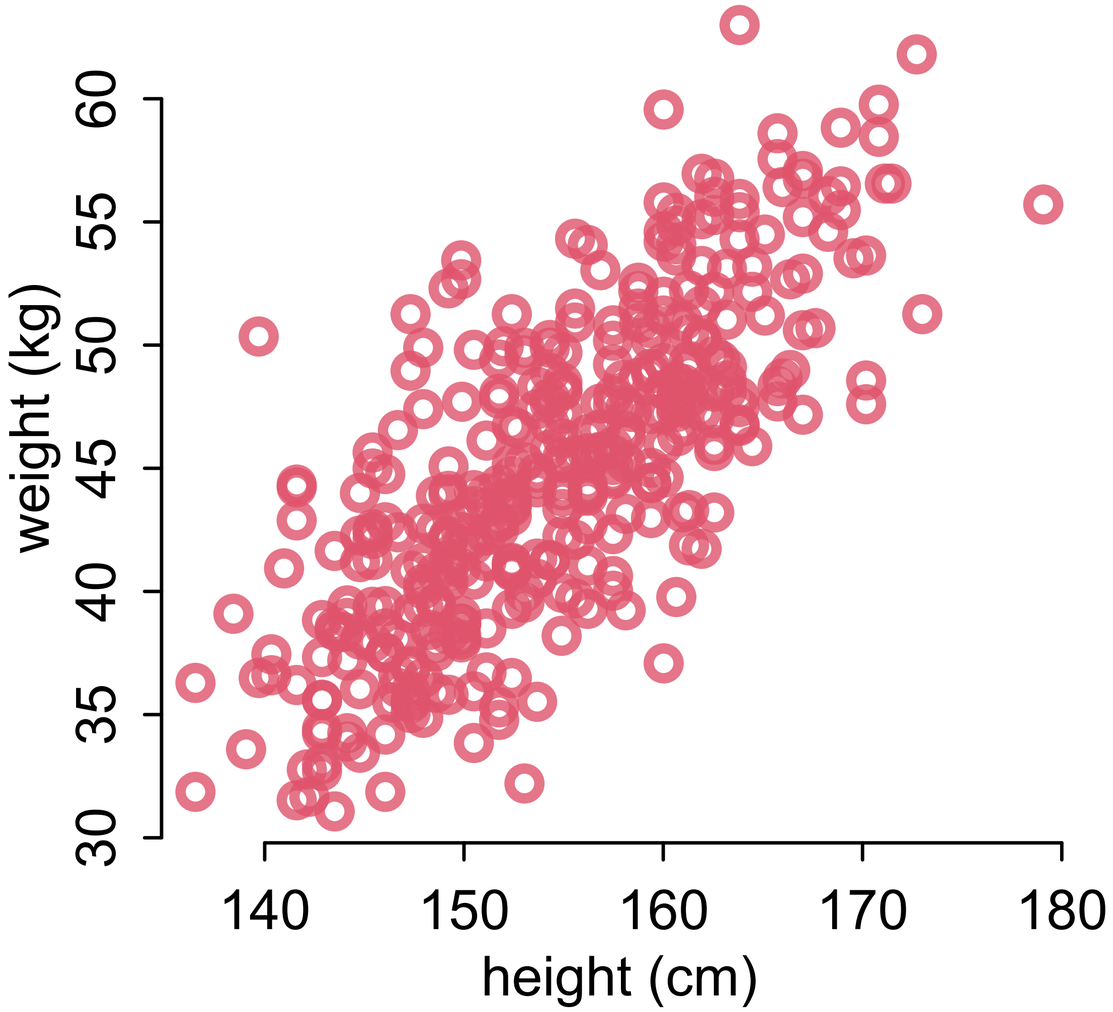
Linear Regression

```
data(Howell1)
d <- Howell1[Howell1$age>=18,]
```

Drawing the Owl

(1) Question/goal/estimand

Describe association between ADULT **weight** and **height**



Linear Regression

```
data(Howell1)
d <- Howell1[Howell1$age>=18,]
```

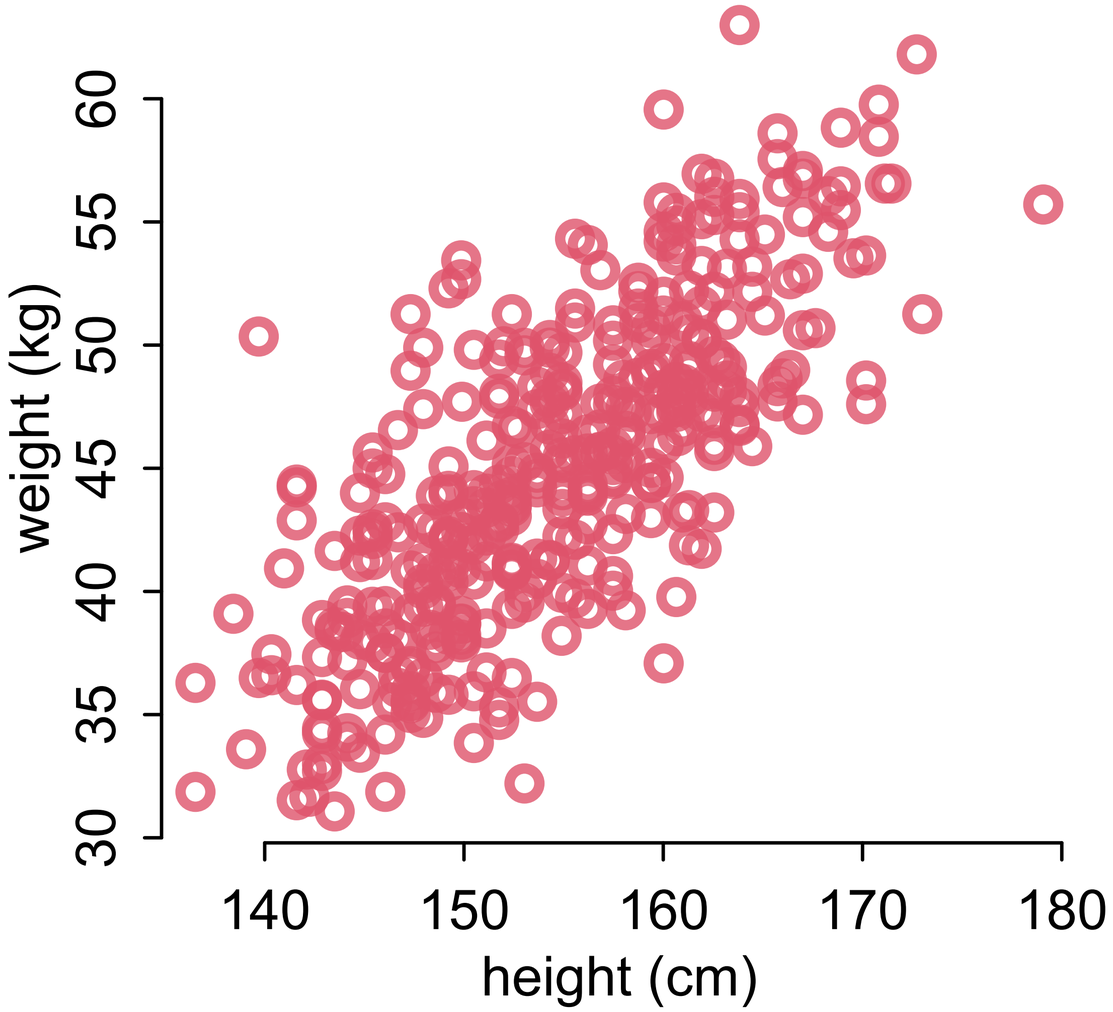
(2) Scientific model

How does **height** influence **weight**?

$$H \longrightarrow W$$

$$W = f(H)$$

"Weight is some function of height"

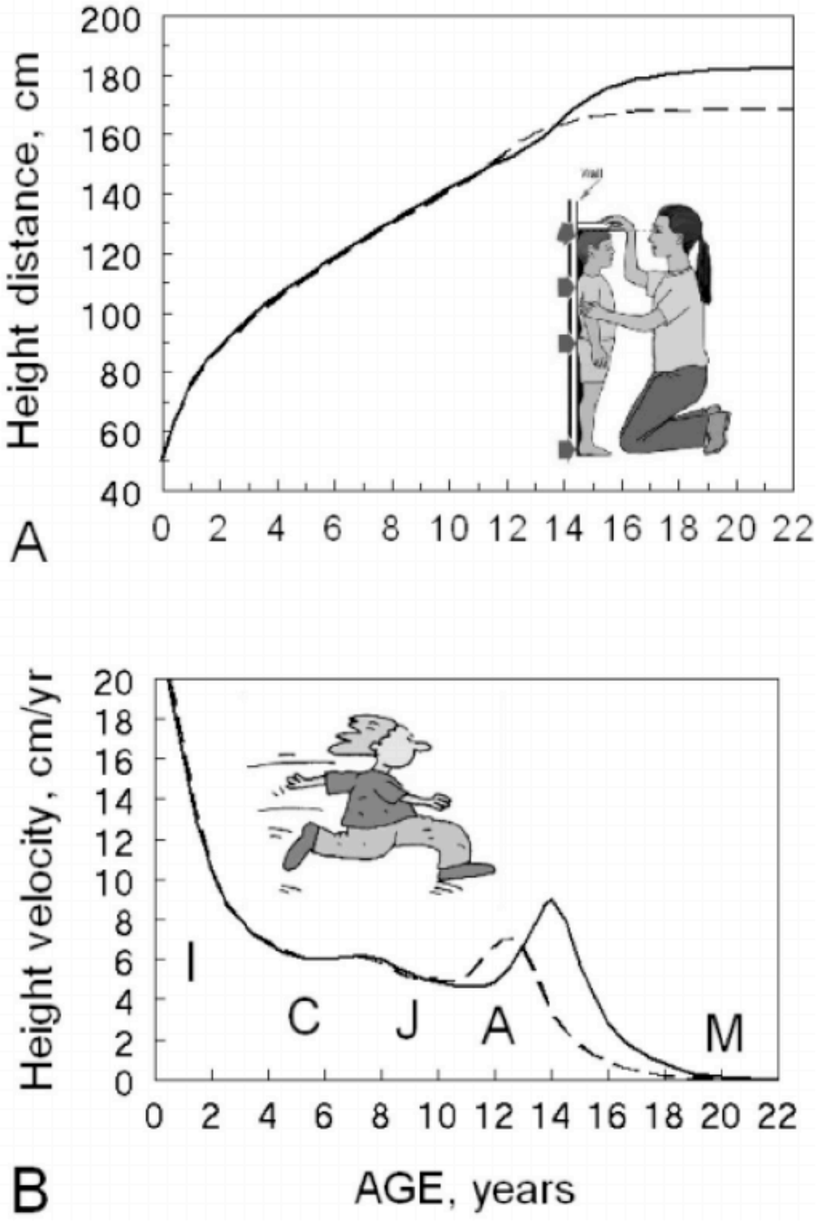


Generative models

Options

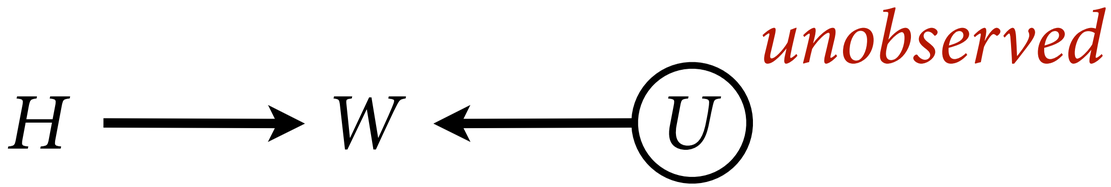
(1) **Dynamic**: Incremental growth of organism; both mass and height (length) derive from growth pattern; Gaussian variation result of summed fluctuations

(2) **Static**: Changes in height result in changes in weight, but no mechanism; Gaussian variation result of growth history



(2) Scientific model

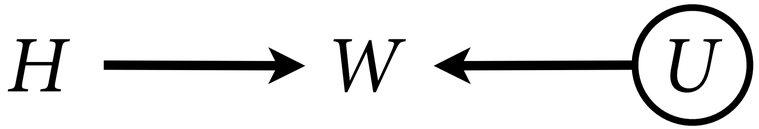
How does **height** influence **weight**?



$$W = f(H, U)$$

"Weight is some function of height and unobserved stuff"

Generative model: H → W



For adults, weight is a proportion of height plus the influence of unobserved causes:

$$W = \beta H + U$$

Generative model: H → W

$$W = \beta H + U$$

Generative code:

R code 3.2

```
# function to simulate weights of individuals from height
sim_weight <- function(H,b,sd) {
    U <- rnorm( length(H) , 0 , sd )
    W <- b*H + U
    return(W)
}
```

R code 3.2

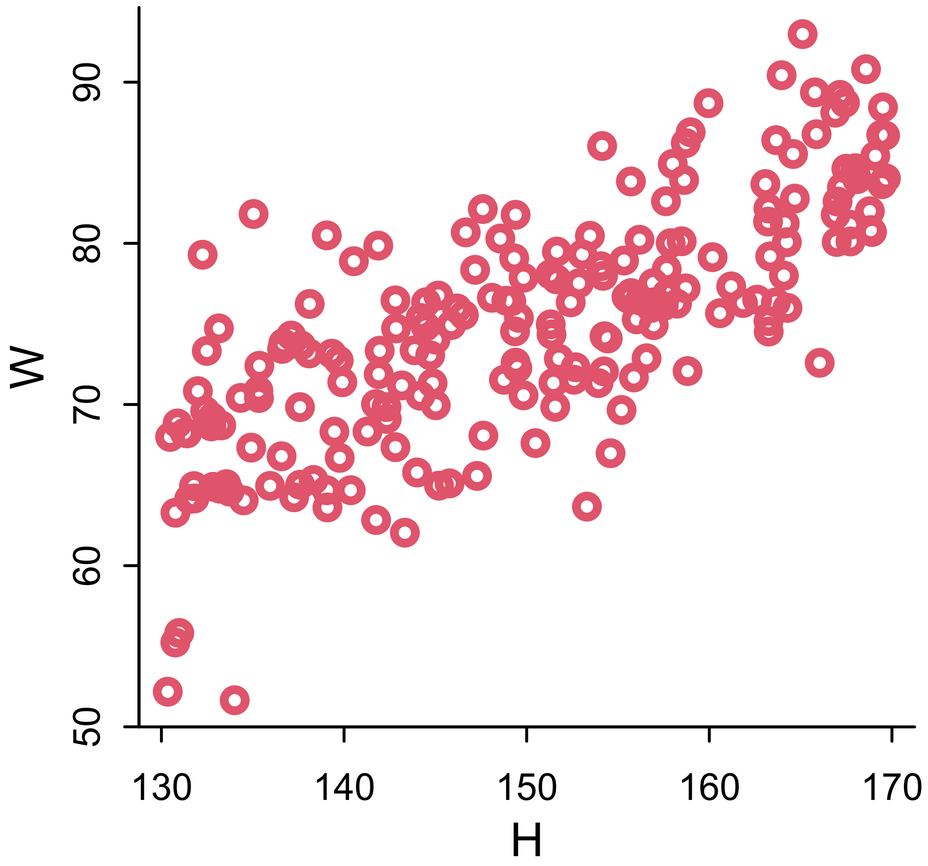
```
# function to simulate weights of individuals from height
sim_weight <- function(H,b,sd) {
    U <- rnorm( length(H) , 0 , sd )
    W <- b*H + U
    return(W)
}
```

R code 3.3

```
H <- runif( 200 , min=130 , max=170 )
W <- sim_weight( H , b=0.5 , sd=5 )
plot( W ~ H , col=2 , lwd=3 )
```

R code 3.3

```
H <- runif( 200 , min=130 , max=170 )
W <- sim_weight( H , b=0.5 , sd=5 )
plot( W ~ H , col=2 , lwd=3 )
```



Describing models

Conventional statistical model notation:

(1) List the variables

(2) Define each variable as a deterministic or distributional function of the other variables

Describing models

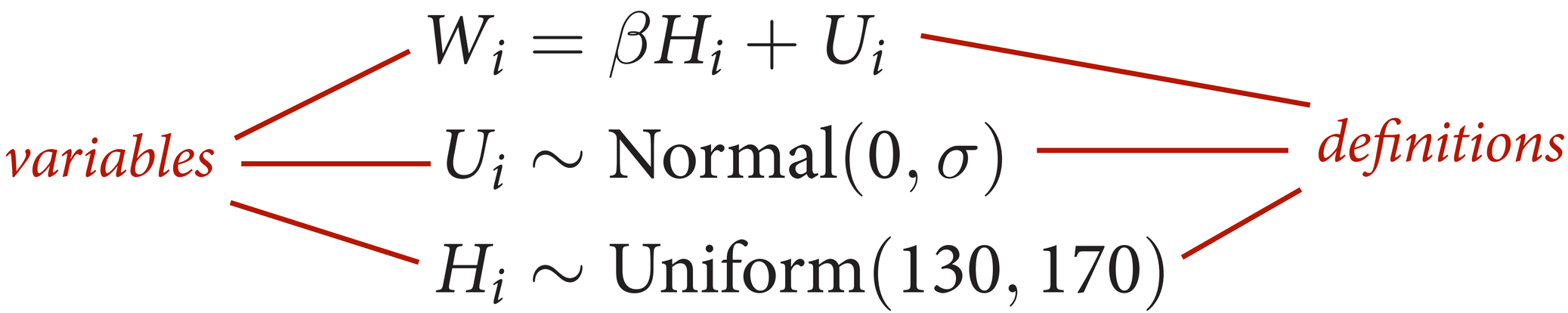
R code 3.2

```
# function to simulate weights of individuals from height
sim_weight <- function(H,b,sd) {
    U <- rnorm( length(H) , 0 , sd )
    W <- b*H + U
    return(W)
}
```

$$W_i = \beta H_i + U_i$$

$$U_i \sim \text{Normal}(0, \sigma)$$

$$H_i \sim \text{Uniform}(130, 170)$$


variables
$W_i = \beta H_i + U_i$
$U_i \sim \text{Normal}(0, \sigma)$
$H_i \sim \text{Uniform}(130, 170)$
definitions

$$W_i = \beta H_i + U_i$$

$$U_i \sim \text{Normal}(0, \sigma)$$

$$H_i \sim \text{Uniform}(130, 170)$$

individuals

deterministic

$$W_i = \beta H_i + U_i$$

$$U_i \sim \text{Normal}(0, \sigma)$$

$$H_i \sim \text{Uniform}(130, 170)$$

distributed as

Equation for expected weight

$$W_i = \beta H_i + U_i$$

$$U_i \sim \text{Normal}(0, \sigma)$$

$$H_i \sim \text{Uniform}(130, 170)$$

$$W_i = \beta H_i + U_i$$

$$U_i \sim \text{Normal}(0, \sigma)$$

$$H_i \sim \text{Uniform}(130, 170)$$

Gaussian error with standard deviation sigma

$$W_i = \beta H_i + U_i$$

$$U_i \sim \text{Normal}(0, \sigma)$$

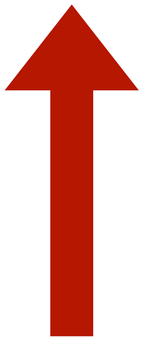
$$H_i \sim \text{Uniform}(130, 170)$$

Height uniformly distributed from 130cm to 170cm

$$W_i = \beta H_i + U_i$$

$$U_i \sim \text{Normal}(0, \sigma)$$

$$H_i \sim \text{Uniform}(130, 170)$$



R code 3.2

```
# function to simulate weights of individuals from height
sim_weight <- function(H,b,sd) {
    U <- rnorm( length(H) , 0 , sd )
    W <- b*H + U
    return(W)
}
```

PAUSE

Linear Regression

```
data(Howell1)
d <- Howell1[Howell1$age>=18,]
```

Drawing the Owl

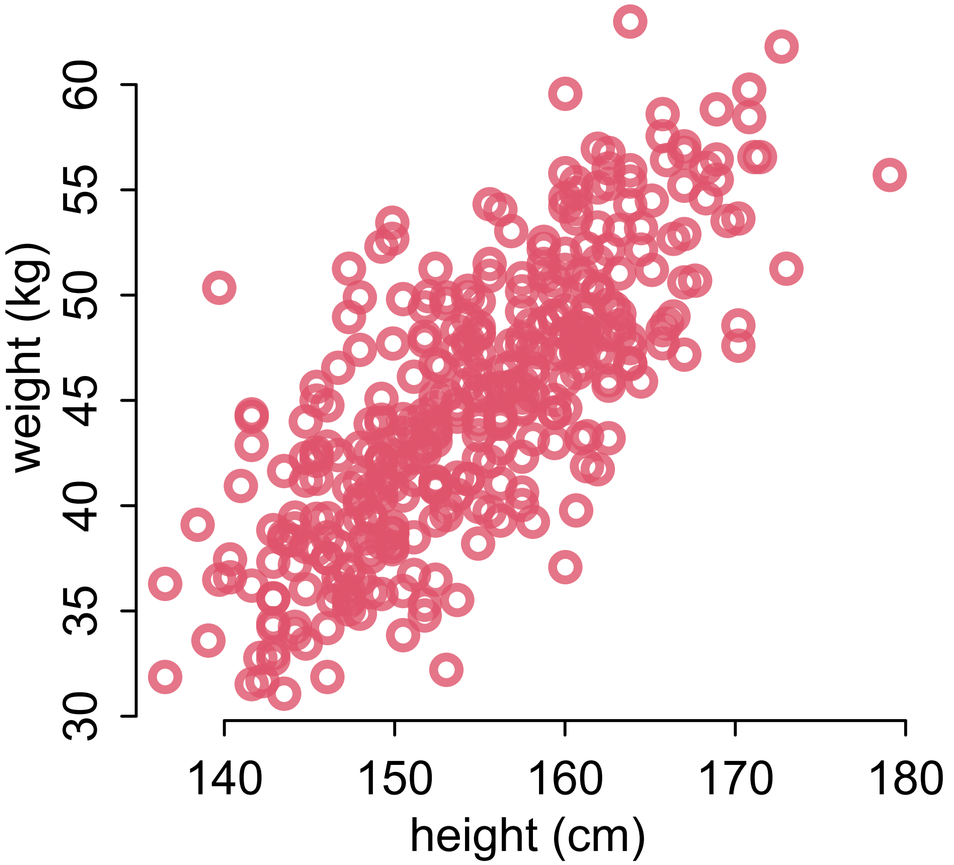
(1) Question/goal/estimand

(2) Scientific model

(3) Statistical model(s)

(4) Validate model

(5) Analyze data



Estimator

We want to estimate how the average weight changes with height.

$$\mathrm{E}(W_i|H_i) = \alpha + \beta H_i$$

Average weight conditional on height

intercept

slope

Posterior distribution

$$\Pr(\alpha, \beta, \sigma | H_i, W_i) = \frac{\Pr(W_i | H_i, \alpha, \beta, \sigma) \Pr(\alpha, \beta, \sigma)}{Z}$$

Posterior distribution

$$\Pr(\alpha, \beta, \sigma | H_i, W_i) = \frac{\Pr(W_i | H_i, \alpha, \beta, \sigma) \Pr(\alpha, \beta, \sigma)}{Z}$$

posterior probability of specific line

Posterior distribution

$$\Pr(\alpha, \beta, \sigma | H_i, W_i) = \frac{\Pr(W_i | H_i, \alpha, \beta, \sigma) \Pr(\alpha, \beta, \sigma)}{Z}$$

garden of forking data

posterior probability of specific line

Posterior distribution

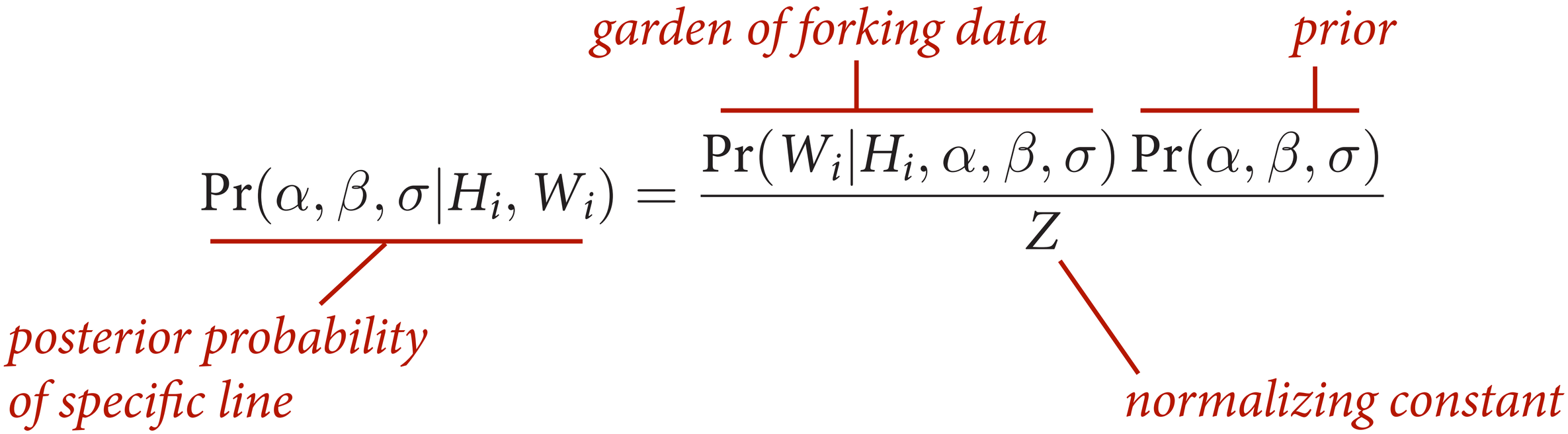
$$\Pr(\alpha, \beta, \sigma | H_i, W_i) = \frac{\Pr(W_i | H_i, \alpha, \beta, \sigma) \Pr(\alpha, \beta, \sigma)}{Z}$$

garden of forking data

prior

posterior probability of specific line

Posterior distribution



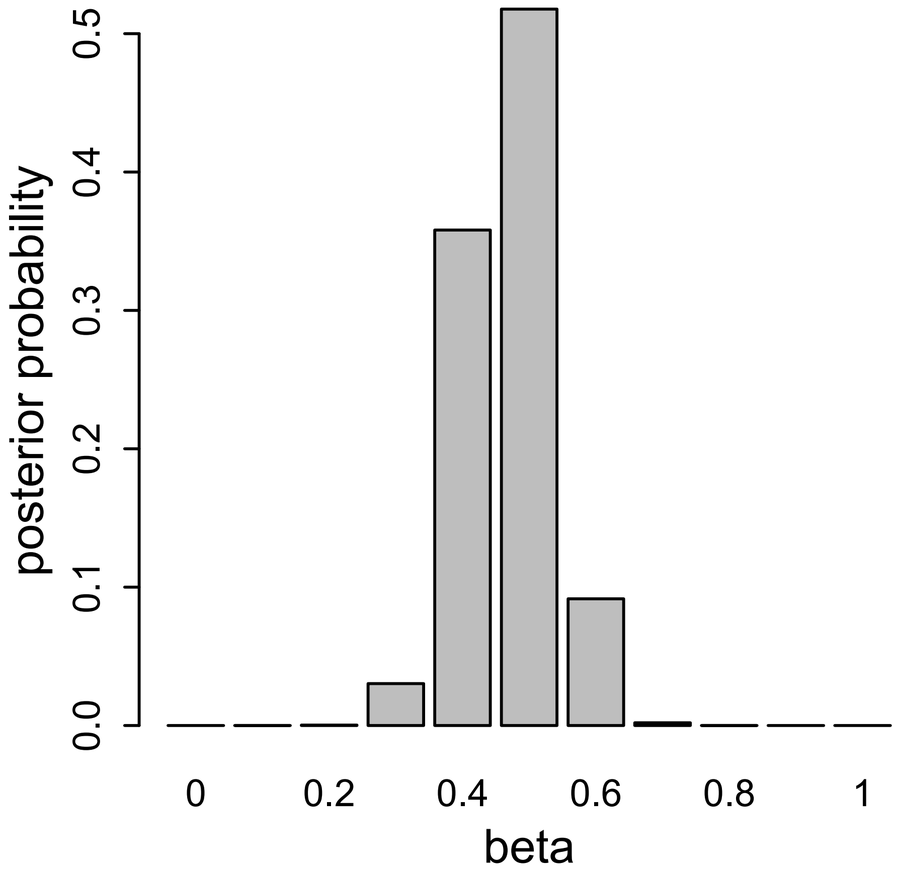
$$\Pr(\alpha, \beta, \sigma | H_i, W_i) = \frac{\Pr(W_i | H_i, \alpha, \beta, \sigma) \Pr(\alpha, \beta, \sigma)}{Z}$$

$$W_i \sim \text{Normal}(\mu_i, \sigma)$$

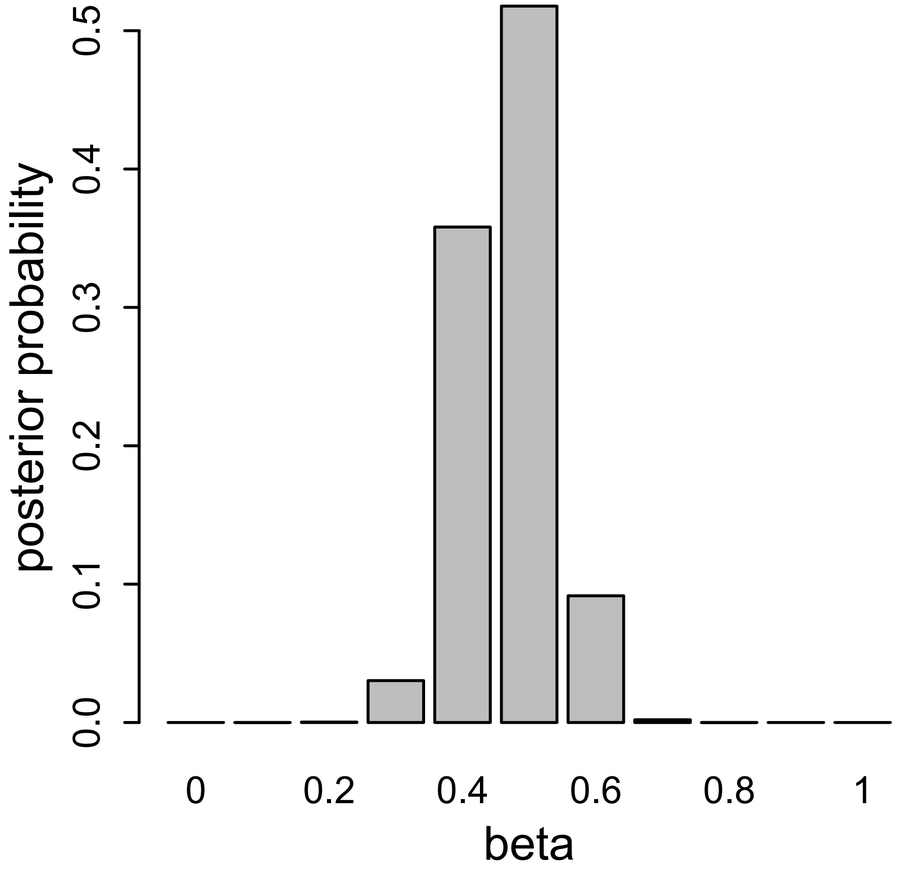
$$\mu_i = \alpha + \beta H_i$$

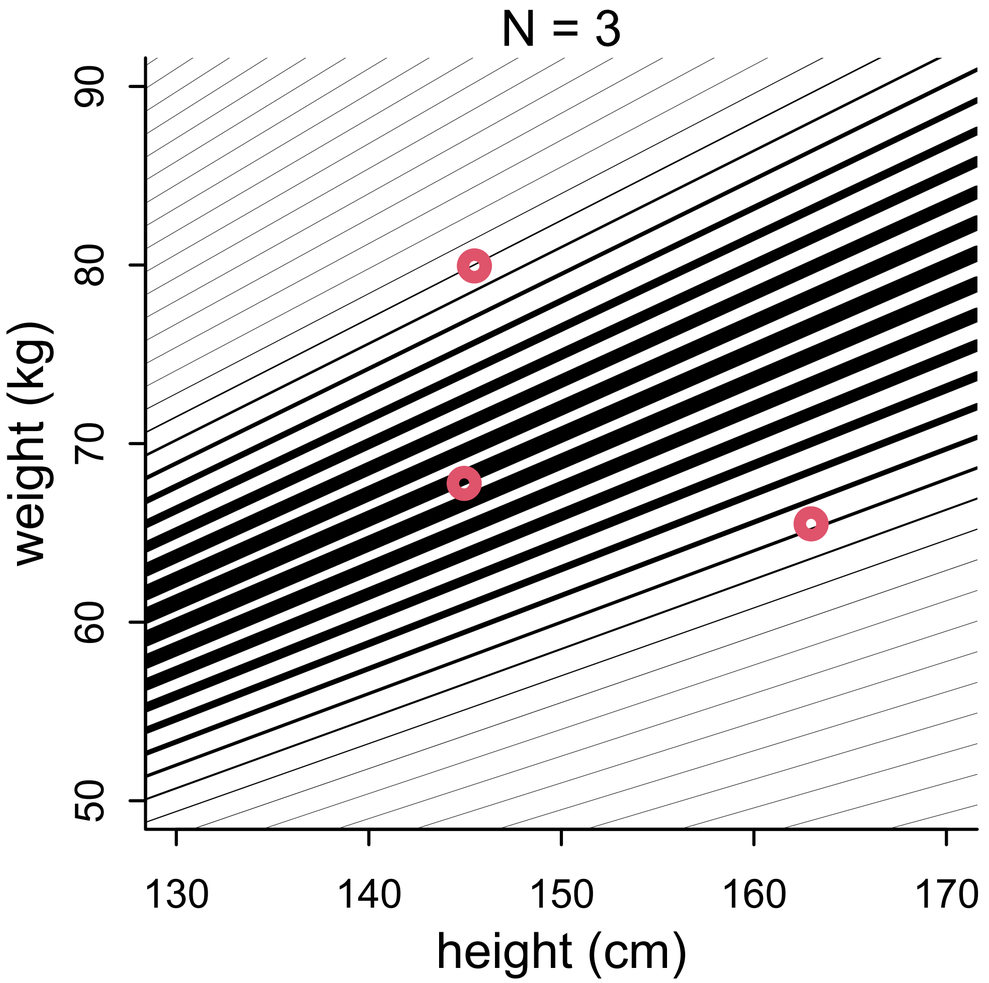
W is distributed normally with mean that is a linear function of H

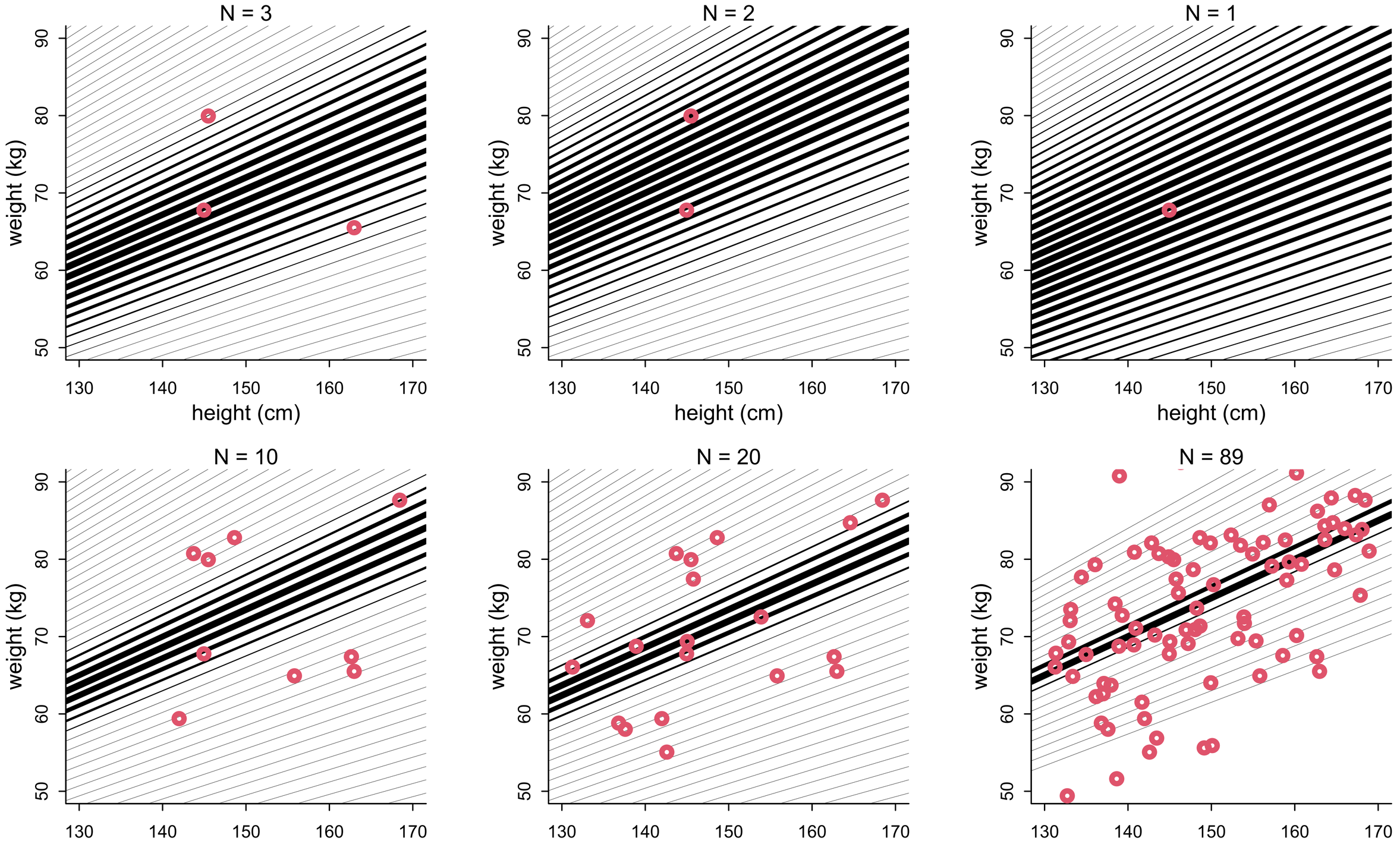
Grid approximate posterior



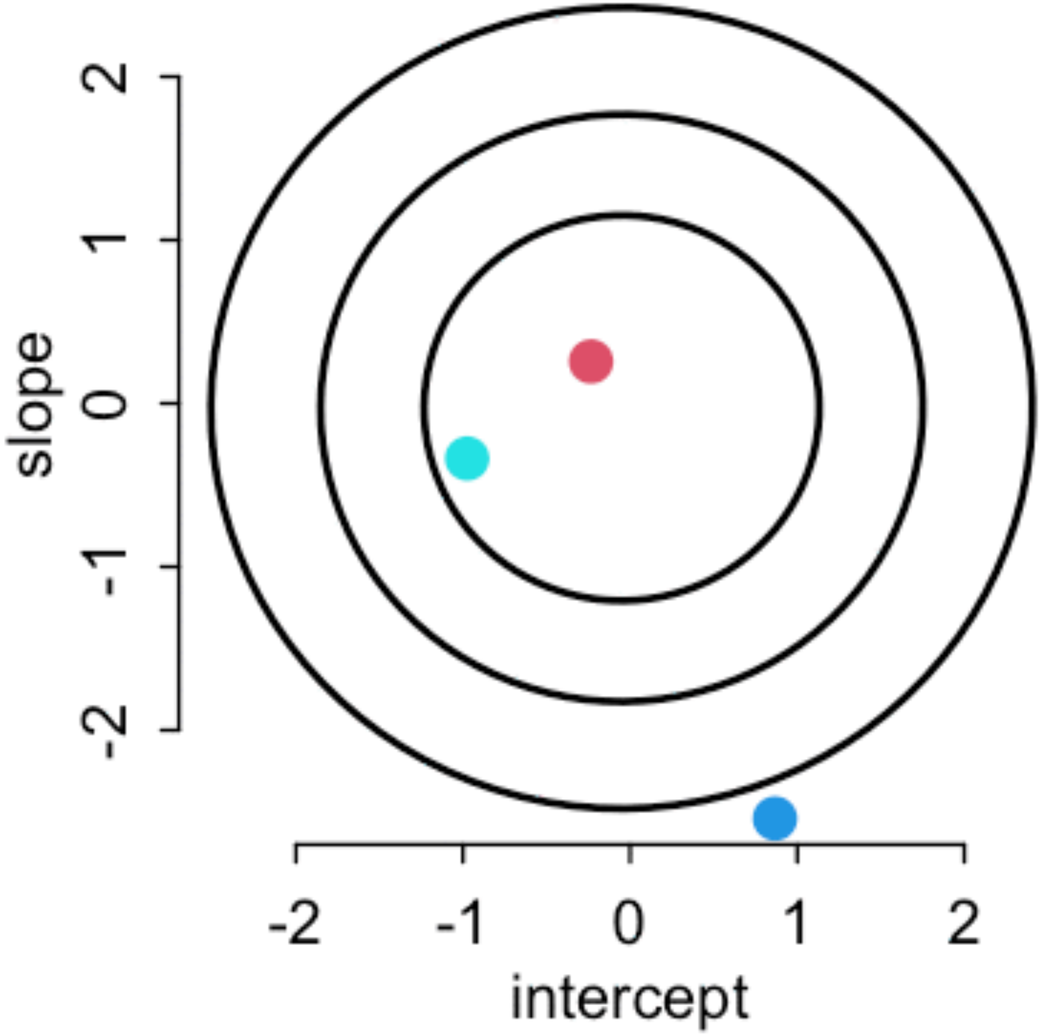
Grid approximate posterior

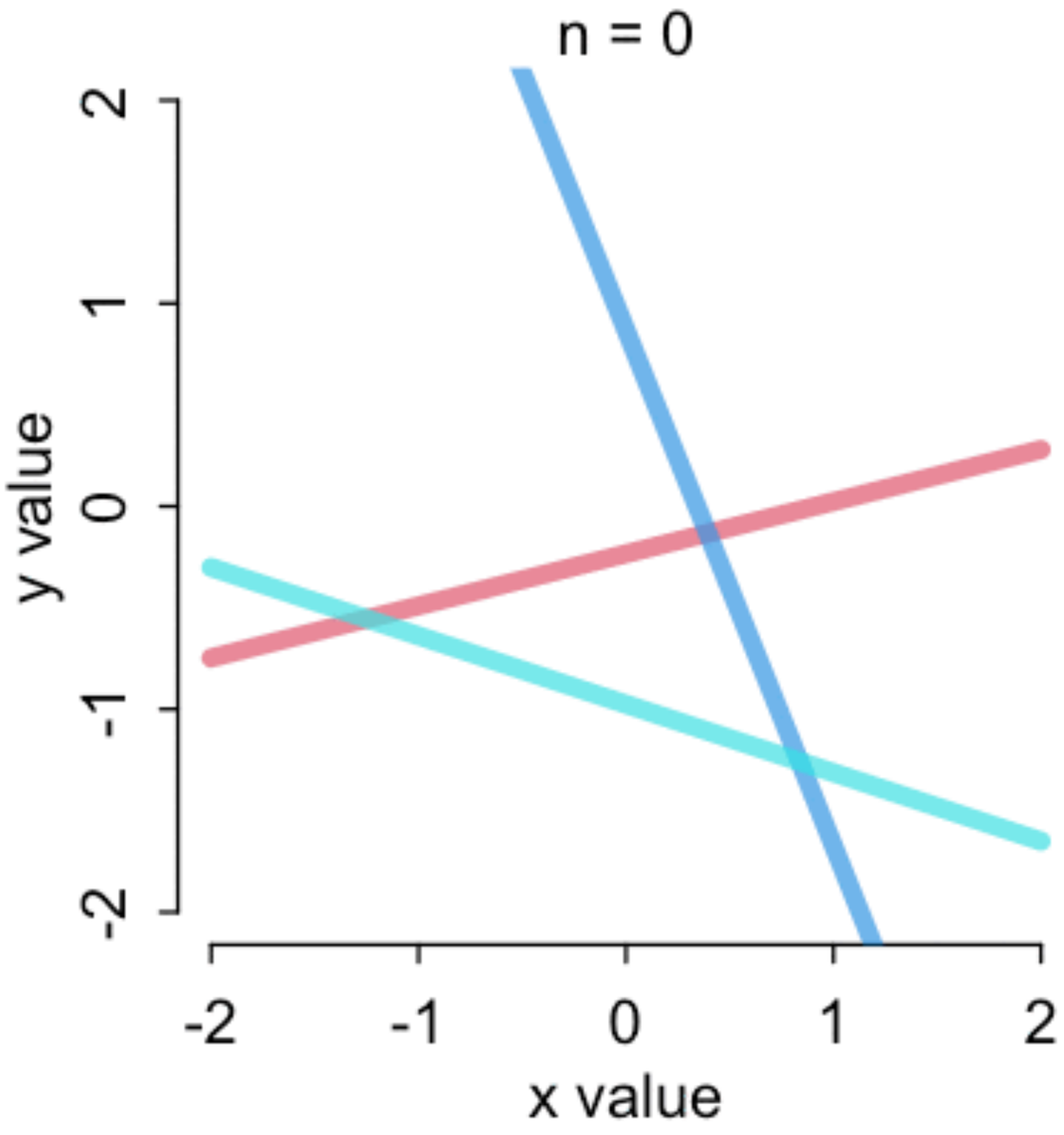




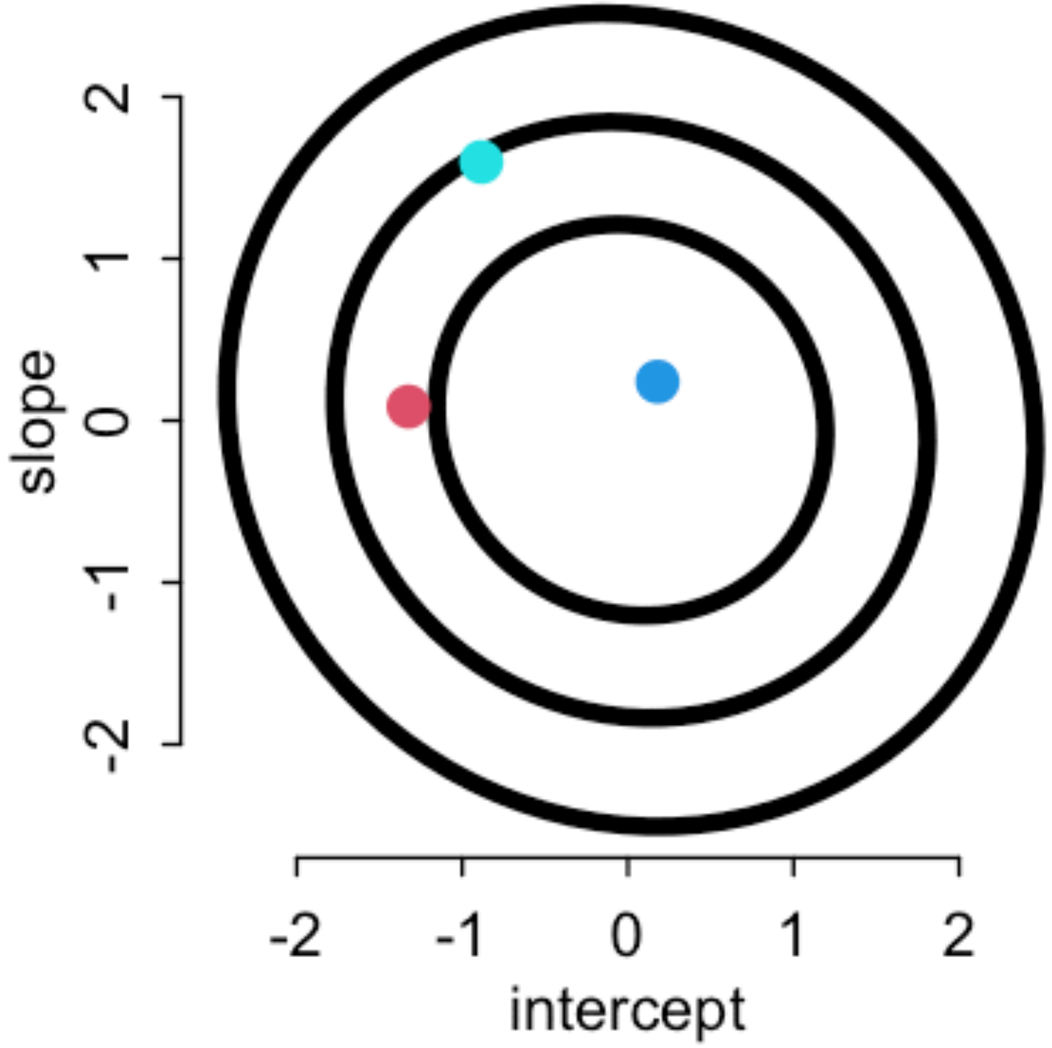

N = 3
N = 2
N = 1
N = 10
N = 20
N = 89
weight (kg)
height (cm)
50
60
70
80
90
130
140
150
160
170

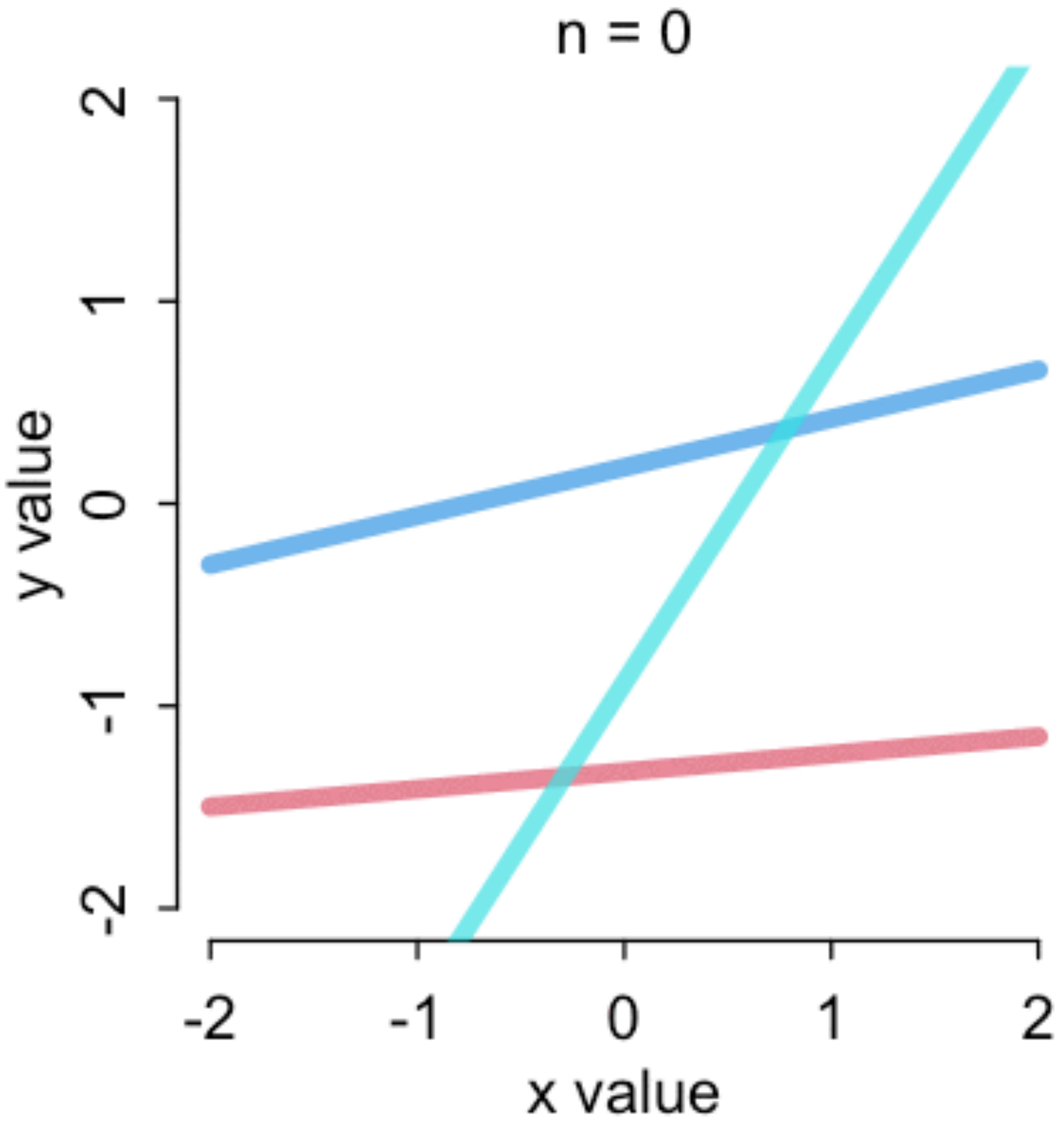
Updating the posterior





Updating the posterior





Enough grid approximation

We'll use quadratic approximation for the rest of the first half of the course.

$$
\begin{aligned}
W_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta H_i \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Uniform}(0, 1) \\
\sigma &\sim \text{Uniform}(0, 10)
\end{aligned}
$$

```
m3.1 <- quap(
    alist(
        W ~ dnorm(mu,sigma),
        mu <- a + b*H,
        a ~ dnorm(0,10),
        b ~ dunif(0,1),
        sigma ~ dunif(0,10)
    ) , data=list(W=W,H=H) )
```

Enough grid approximation

We'll use quadratic approximation for the rest of the first half of the course.

$$W_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta H_i$$
$$\alpha \sim \text{Normal}(0, 10)$$
$$\beta \sim \text{Uniform}(0, 1)$$
$$\sigma \sim \text{Uniform}(0, 10)$$

```
m3.1 <- quap(
    alist(
        W ~ dnorm(mu,sigma),
        mu <- a + b*H,
        a ~ dnorm(0,10),
        b ~ dunif(0,1),
        sigma ~ dunif(0,10)
    ) , data=list(W=W,H=H) )
```

Prior predictive distribution

Priors should express scientific knowledge, but **softly**

When $H = 0$, $W = 0$

Weight increases (on avg) with height

Weight (kg) is less than height (cm)

sigma must be positive

$$W_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta H_i$$

$$\alpha \sim \text{Normal}(0, 10)$$

$$\beta \sim \text{Uniform}(0, 1)$$

$$\sigma \sim \text{Uniform}(0, 10)$$

Prior predictive distribution

Understand the implications of priors through simulation

What do the observable variables look like with these priors?

$$W_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta H_i$$

$$\alpha \sim \text{Normal}(0, 10)$$

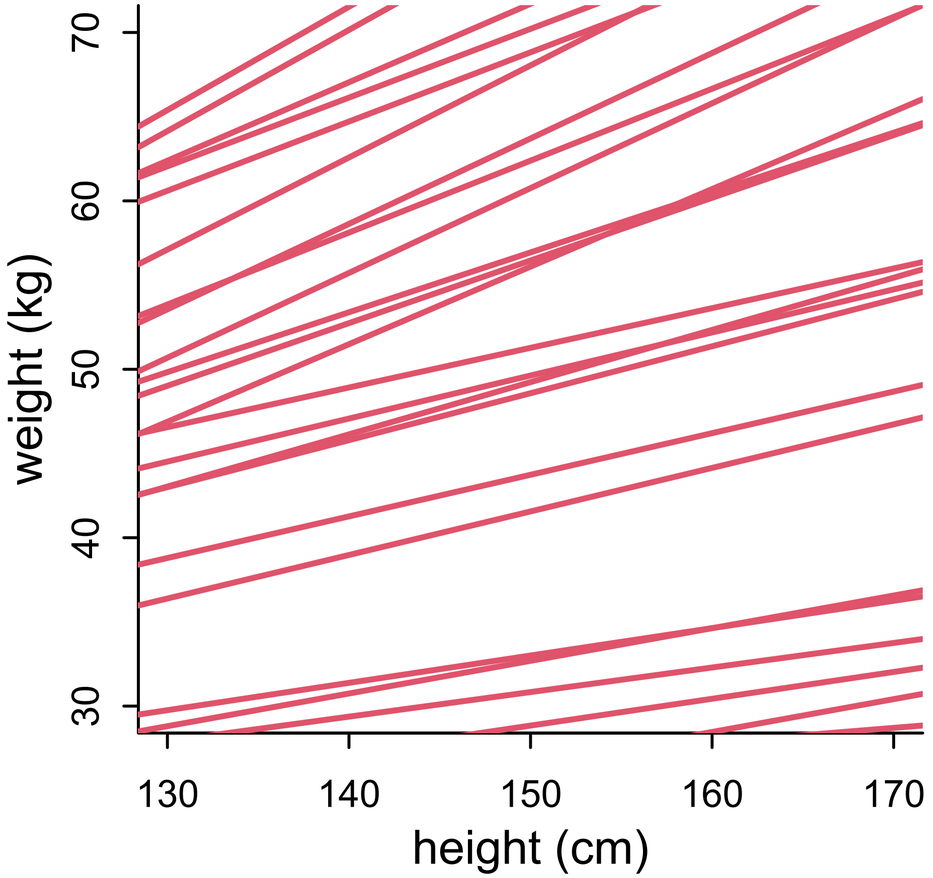
$$\beta \sim \text{Uniform}(0, 1)$$

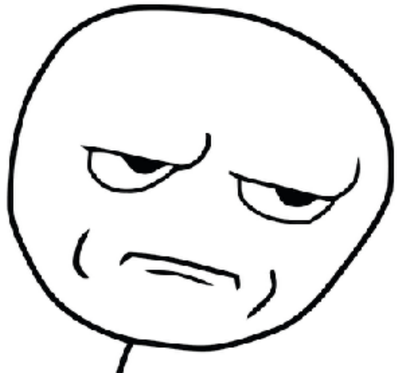
$$\sigma \sim \text{Uniform}(0, 10)$$

```
n <- 1e3
a <- rnorm(n,0,10)
b <- runif(n,0,1)
plot( NULL , xlim=c(130,170) , ylim=c(50,90) ,
    xlab="height (cm)" , ylab="weight (kg)" )
for ( j in 1:50 ) abline( a=a[j] , b=b[j] , lwd=2 , col=2 )
```

$$
\begin{aligned}
W_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta H_i \\
\alpha &\sim \text{Normal}(0, 10) \\
\beta &\sim \text{Uniform}(0, 1) \\
\sigma &\sim \text{Uniform}(0, 10)
\end{aligned}
$$

```
n <- 1e3
a <- rnorm(n,0,10)
b <- runif(n,0,1)
plot( NULL , xlim=c(130,170) , ylim=c(50,90) ,
    xlab="height (cm)" , ylab="weight (kg)" )
for ( j in 1:50 ) abline( a=a[j] , b=b[j] , lwd=2 , col=2 )
```





Sermon on priors

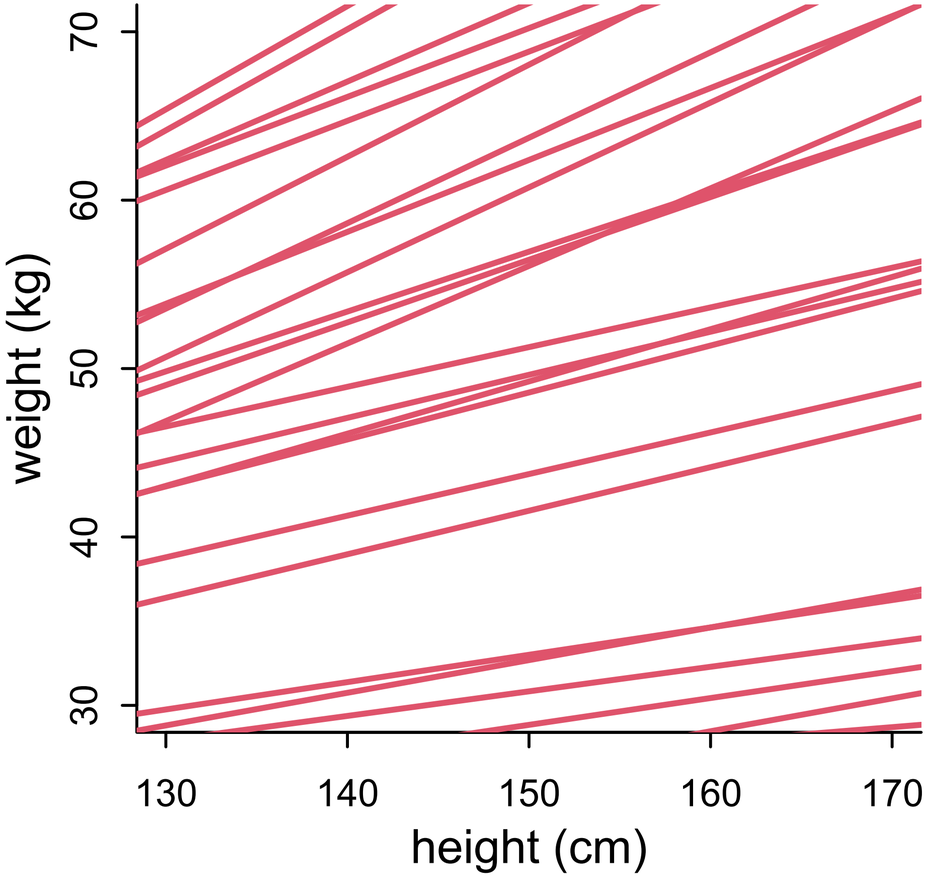
There are no correct priors, only scientifically justifiable priors

Justify with information outside the data — **like rest of model**

Priors not so important in simple models

Very important/useful in complex models

Need to practice now: simulate, understand



Linear Regression

```
data(Howell1)
d <- Howell1[Howell1$age>=18,]
```

Drawing the Owl

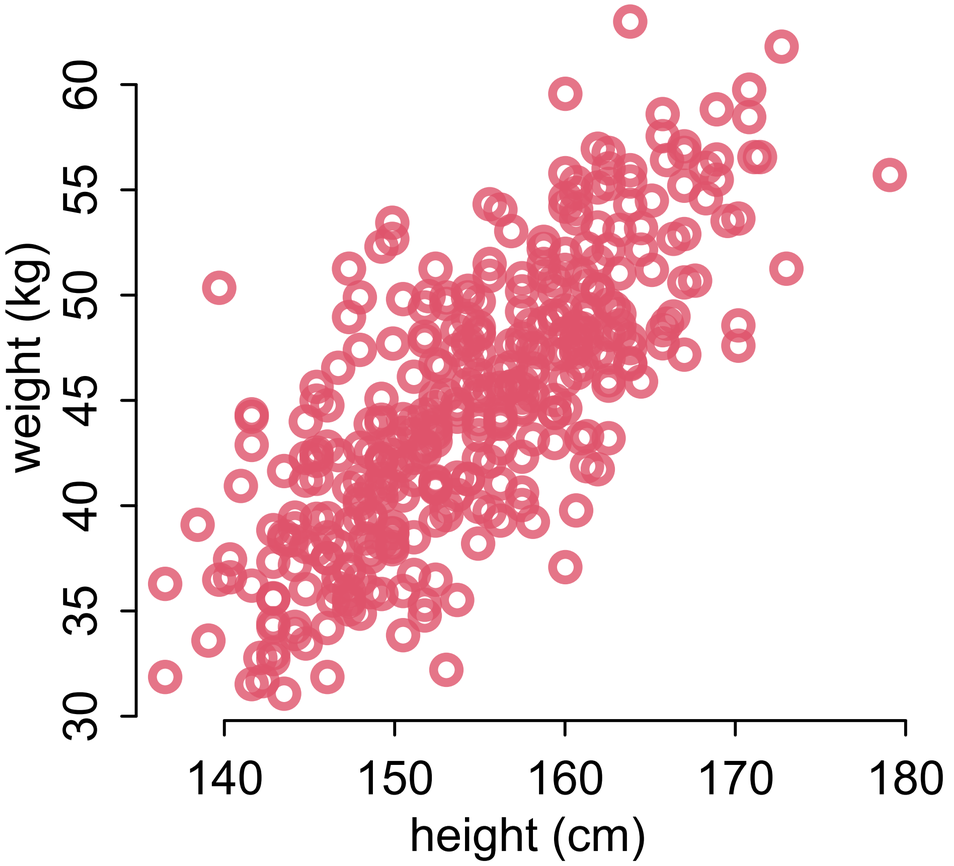
(1) Question/goal/estimand

(2) Scientific model

(3) Statistical model(s)

(4) Validate model

(5) Analyze data



Simulation-Based Validation

Bare minimum: Test statistical model with simulated observations from scientific model

Golem might be broken

Even working golems might not deliver what you hoped

Strong test: Simulation-Based Calibration

Fahrvergnügen

```
# simulate a sample of 10 people
set.seed(93)
H <- runif(10,130,170)
W <- sim_weight(H,b=0.5,sd=5)

# run the model
library(rethinking)
m3.1 <- quap(
    alist(
        W ~ dnorm(mu,sigma),
        mu <- a + b*H,
        a ~ dnorm(0,10),
        b ~ dunif(0,1),
        sigma ~ dunif(0,10)
    ) , data=list(W=W,H=H) )

# summary
precis( m3.1 )
```

```
       mean   sd  5.5% 94.5%
a      5.19 9.43 -9.88 20.26
b      0.49 0.07  0.38  0.59
sigma 5.64 1.29  3.57  7.71
```

Vary slope and make sure posterior mean tracks it

Use a large sample to see that it converges to data generating value

Same for other unknowns (parameters)

Linear Regression

```
data(Howell1)
d <- Howell1[Howell1$age>=18,]
```

Drawing the Owl

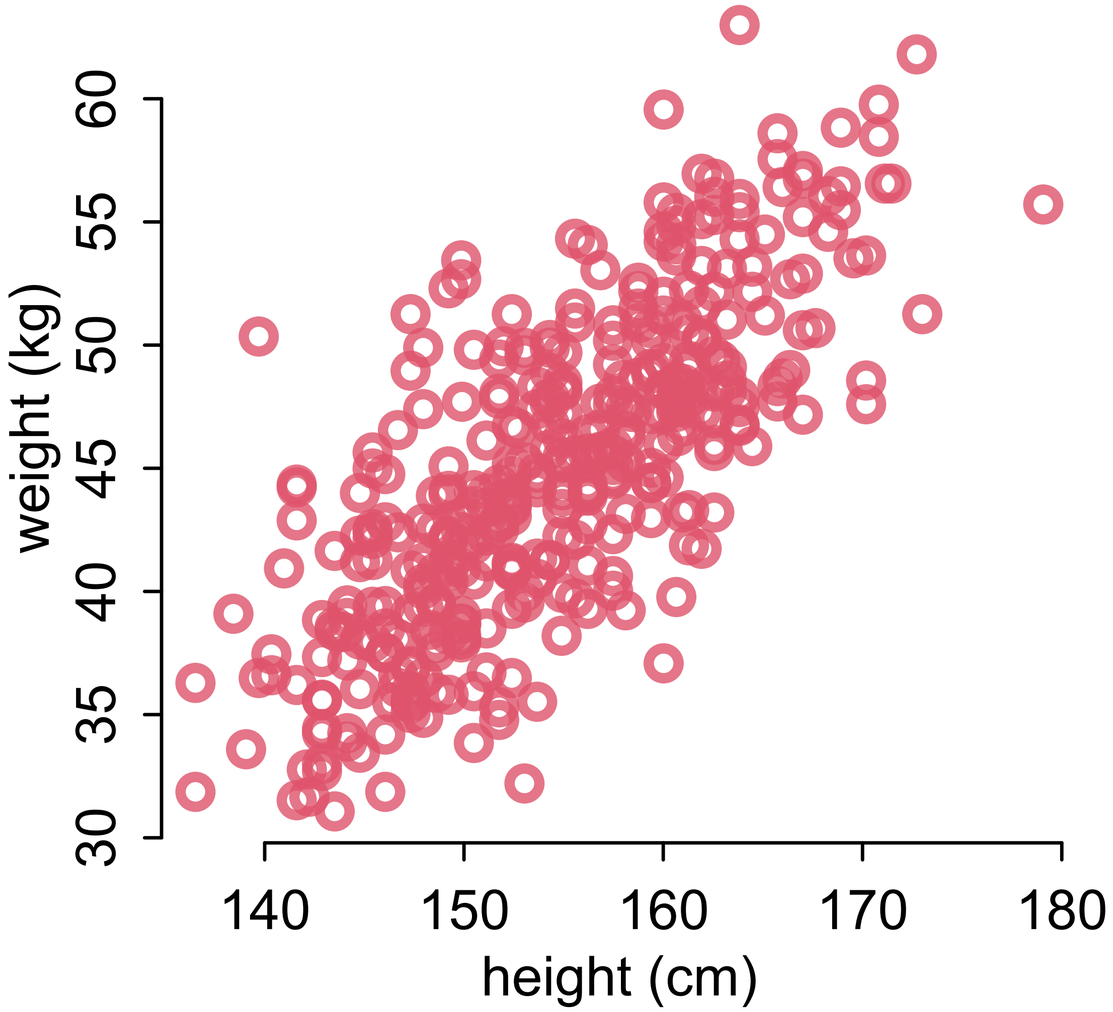
(1) Question/goal/estimand

(2) Scientific model

(3) Statistical model(s)

(4) Validate model

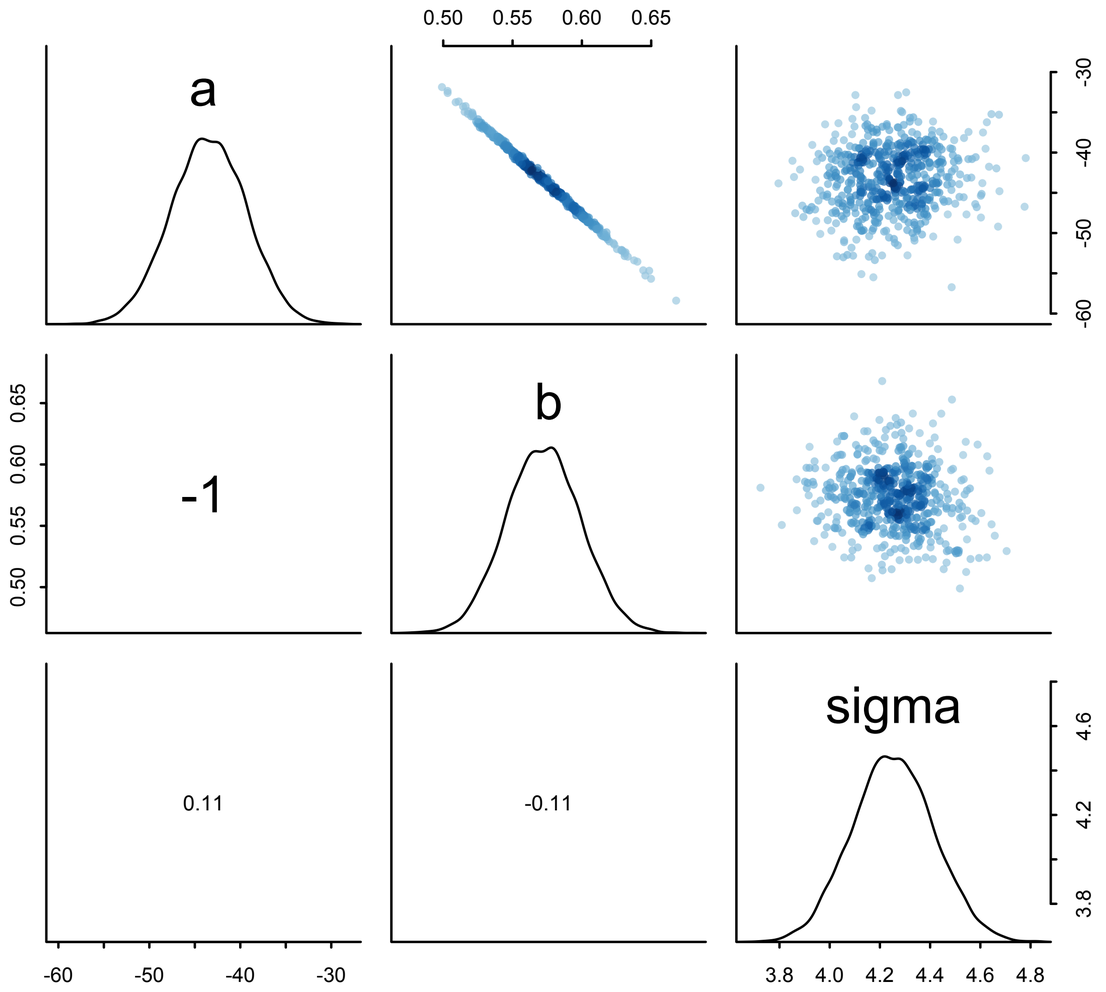
(5) Analyze data



Analyze the data

```
dat <- list(W=d2$weight,H=d2$height)
m3.2 <- quap(
    alist(
        W ~ dnorm(mu,sigma),
        mu <- a + b*H,
        a ~ dnorm(0,10),
        b ~ dunif(0,1),
        sigma ~ dunif(0,10)
    ) , data=dat )
precis( m3.2 )
```

```
          mean   sd   5.5%  94.5%
a      -43.38 4.17 -50.04 -36.71
b        0.57 0.03   0.53   0.61
sigma    4.25 0.16   3.99   4.51
```



Obey The Law

First Law of Statistical Interpretation:

The **parameters are not independent** of one another and cannot always be independently interpreted

Instead:

Push out **posterior predictions** and describe/interpret those

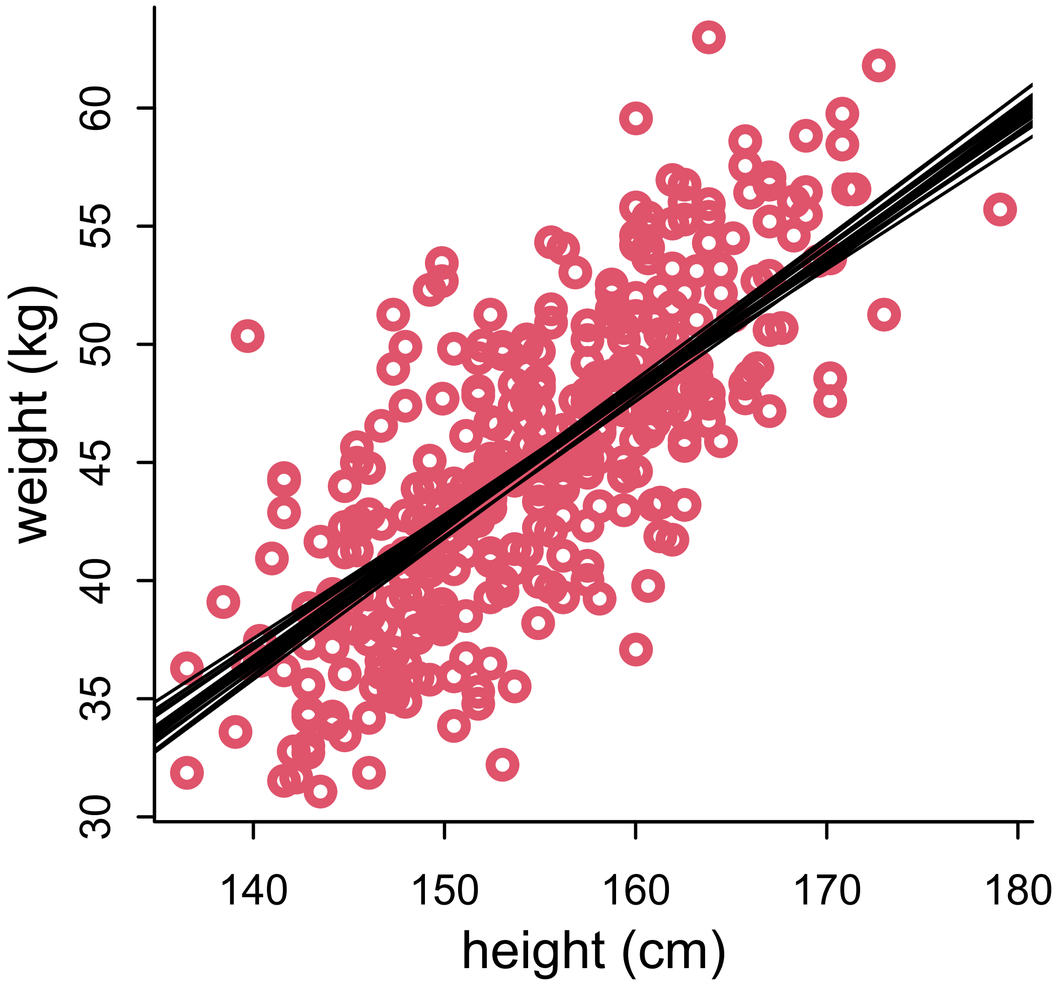
```
> precis(m_adults)
       mean   sd  5.5% 94.5%
a     45.00 0.23 44.64 45.36
b      0.63 0.03  0.58  0.68
sigma  4.23 0.16  3.97  4.48
>
```

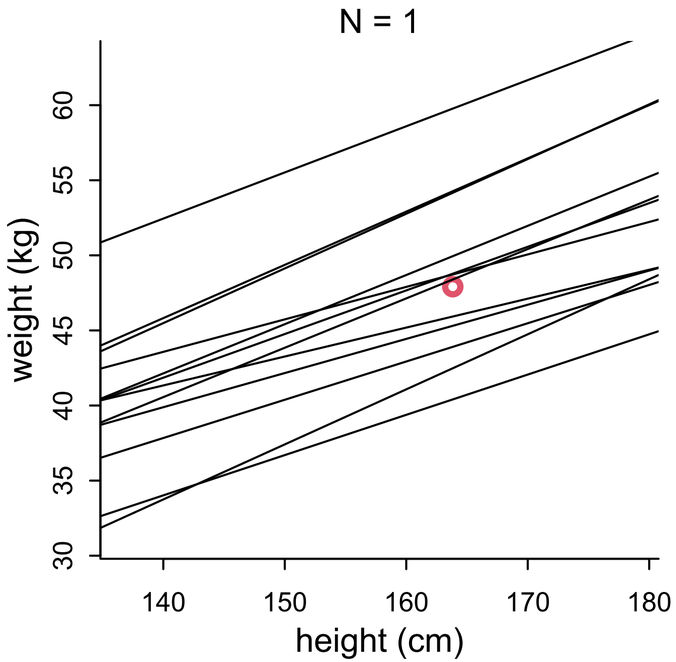
```
> post <- extract.samples(m_adults)
> head(post)
         a         b    sigma
1 45.14733 0.7045790 4.380254
2 44.97759 0.6461353 4.372925
3 44.94856 0.6537192 4.111149
4 44.85016 0.6597310 4.379347
5 44.75898 0.6532690 4.200026
6 44.91711 0.6090434 4.105432
>
```

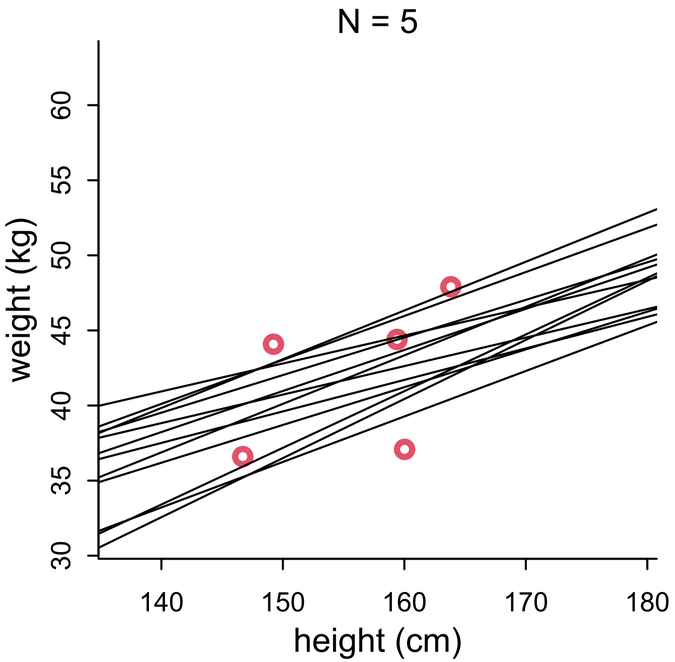
Posterior predictive distribution

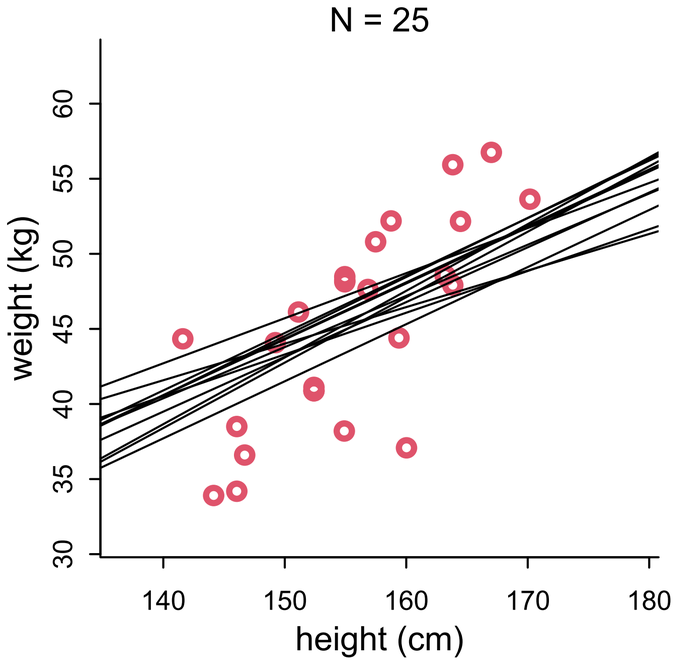
The posterior is full of lines

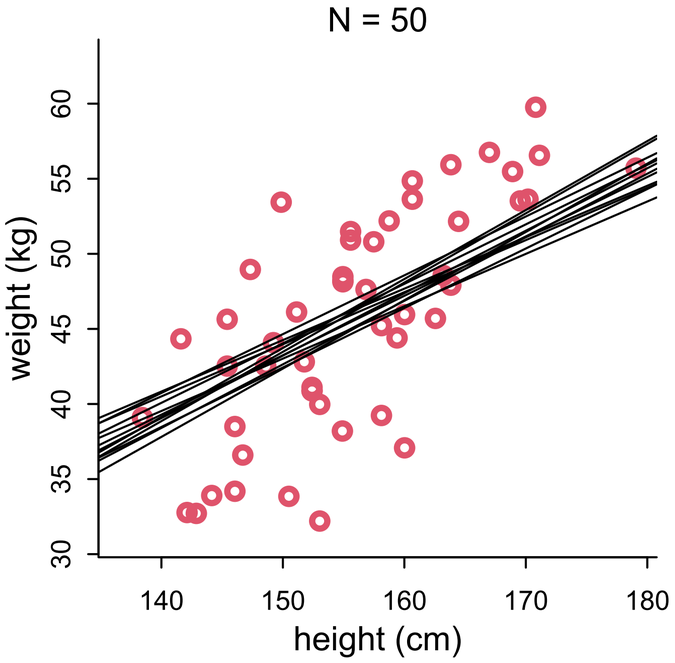
```
post <- extract.samples(m3.2)
plot( d2$height , d2$weight , col=2 , lwd=3 ,
    xlab="height (cm)" , ylab="weight (kg)" )
for ( j in 1:20 )
    abline( a=post$a[j] , b=post$b[j] , lwd=1 )
```

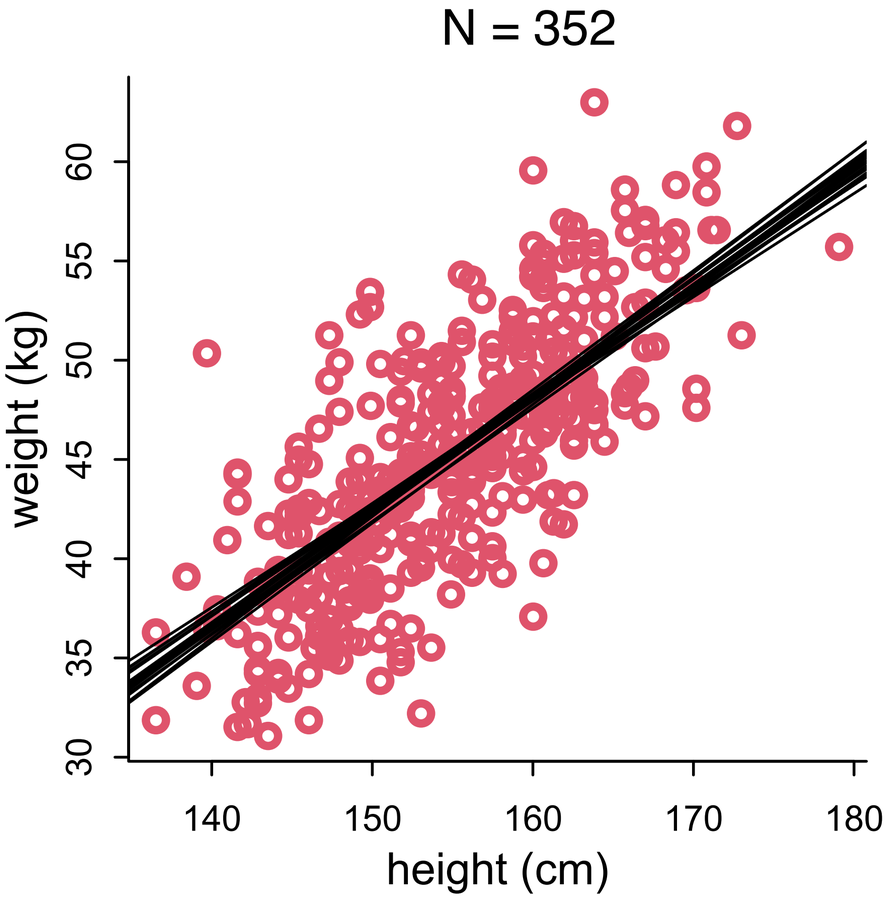



N = 1
weight (kg)
height (cm)
30
35
40
45
50
55
60
140
150
160
170
180


N = 5
weight (kg)
height (cm)
30
35
40
45
50
55
60
140
150
160
170
180


N = 25
weight (kg)
height (cm)
30
35
40
45
50
55
60
140
150
160
170
180


N = 50
weight (kg)
height (cm)
30
35
40
45
50
55
60
140
150
160
170
180


N = 352
weight (kg)
height (cm)
30
35
40
45
50
55
60
140
150
160
170
180

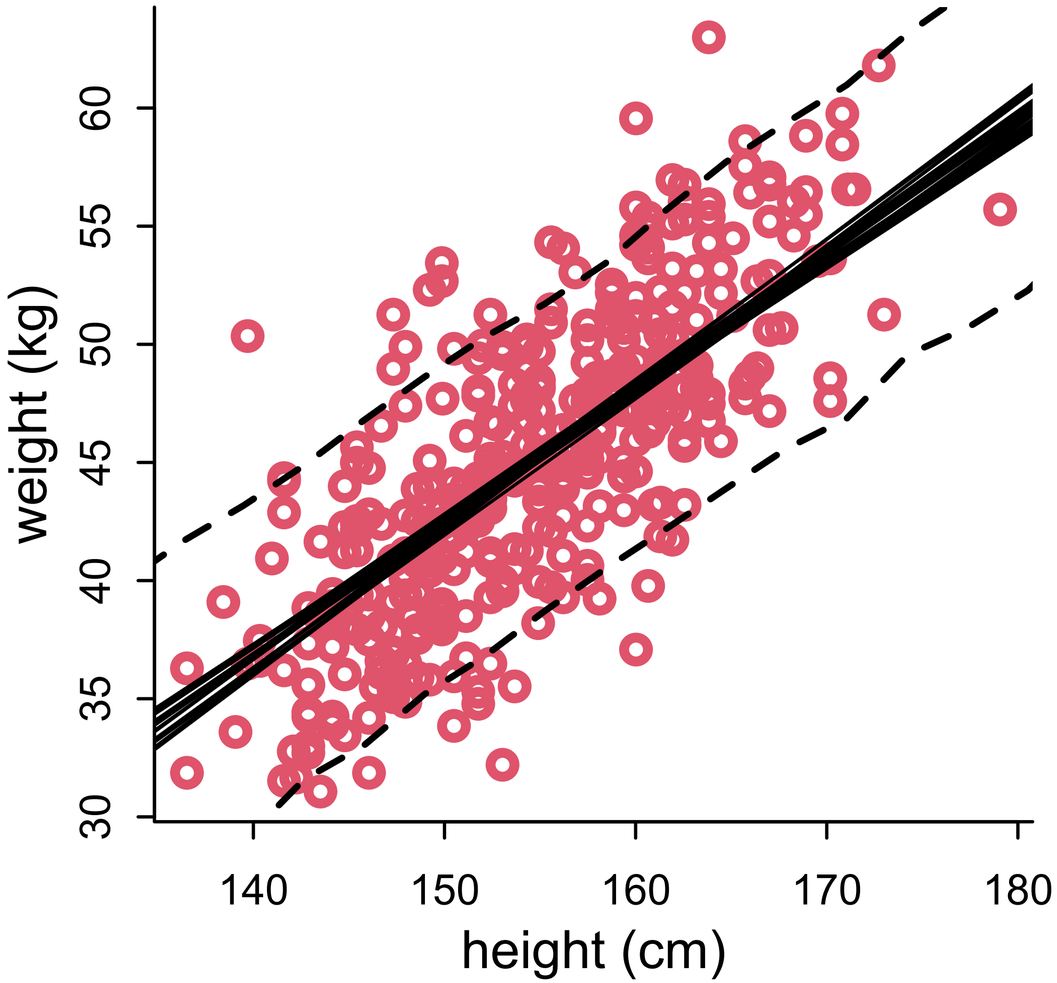
Posterior predictive distribution

The posterior is full of lines

```
post <- extract.samples(m3.2)
plot( d2$height , d2$weight , col=2 , lwd=3 ,
    xlab="height (cm)" , ylab="weight (kg)" )
for ( j in 1:20 )
    abline( a=post$a[j] , b=post$b[j] , lwd=1 )
```

The posterior is full of people

```
height_seq <- seq(130,190,len=20)
W_postpred <- sim( m3.2 ,
data=list(H=height_seq) )
W_PI <- apply( W_postpred , 2 , PI )
lines( height_seq , W_PI[1,] , lty=2 , lwd=2 )
lines( height_seq , W_PI[2,] , lty=2 , lwd=2 )
```

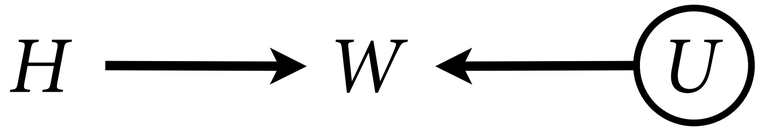



120
110
100
90
80
70
60
50
40
30
20
BLOOD HEAT
SUMR HEAT
TEMPR RATE
FREEZ ING

Flexible Linear Thermometers

Generative model

How does **height** influence **weight**?



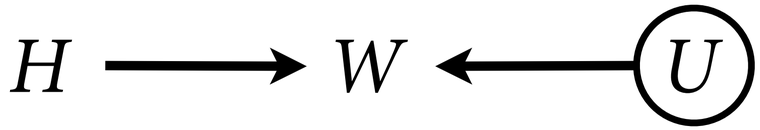
$$W = f(H, U)$$

"Weight is some function of height & unmeasured stuff"

Flexible Linear Thermometers

Generative model

How does **height** influence **weight**?



$$W = f(H, U)$$

"Weight is some function of height & unmeasured stuff"

Statistical model

How does average **weight** change with **height**?

$$W_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta H_i$$
$$\alpha \sim \text{Normal}(0, 10)$$
$$\beta \sim \text{Uniform}(0, 1)$$
$$\sigma \sim \text{Uniform}(0, 10)$$

Course Schedule

Week 1	Bayesian inference	Chapters 1, 2, 3
Week 2	Linear models & Causal Inference	Chapter 4
Week 3	Causes, Confounds & Colliders	Chapters 5 & 6
Week 4	Overfitting / Interactions	Chapters 7 & 8
Week 5	MCMC & Generalized Linear Models	Chapters 9, 10, 11
Week 6	Integers & Other Monsters	Chapters 11 & 12
Week 7	Multilevel models I	Chapter 13
Week 8	Multilevel models II	Chapter 14
Week 9	Measurement & Missingness	Chapter 15
Week 10	Generalized Linear Madness	Chapter 16

`https://github.com/rmcelreath/stat_rethinking_2023`