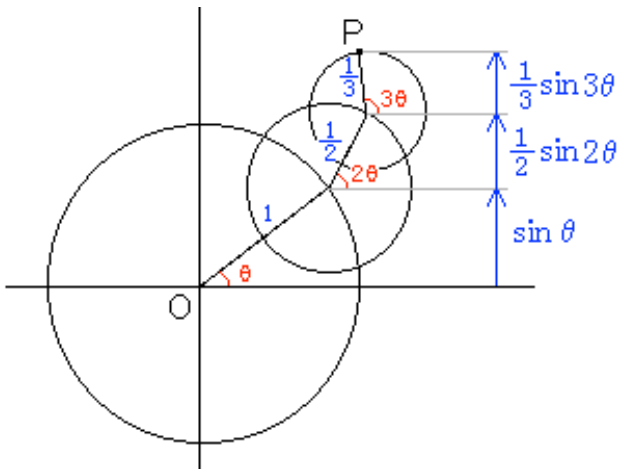
Statistical Rethinking

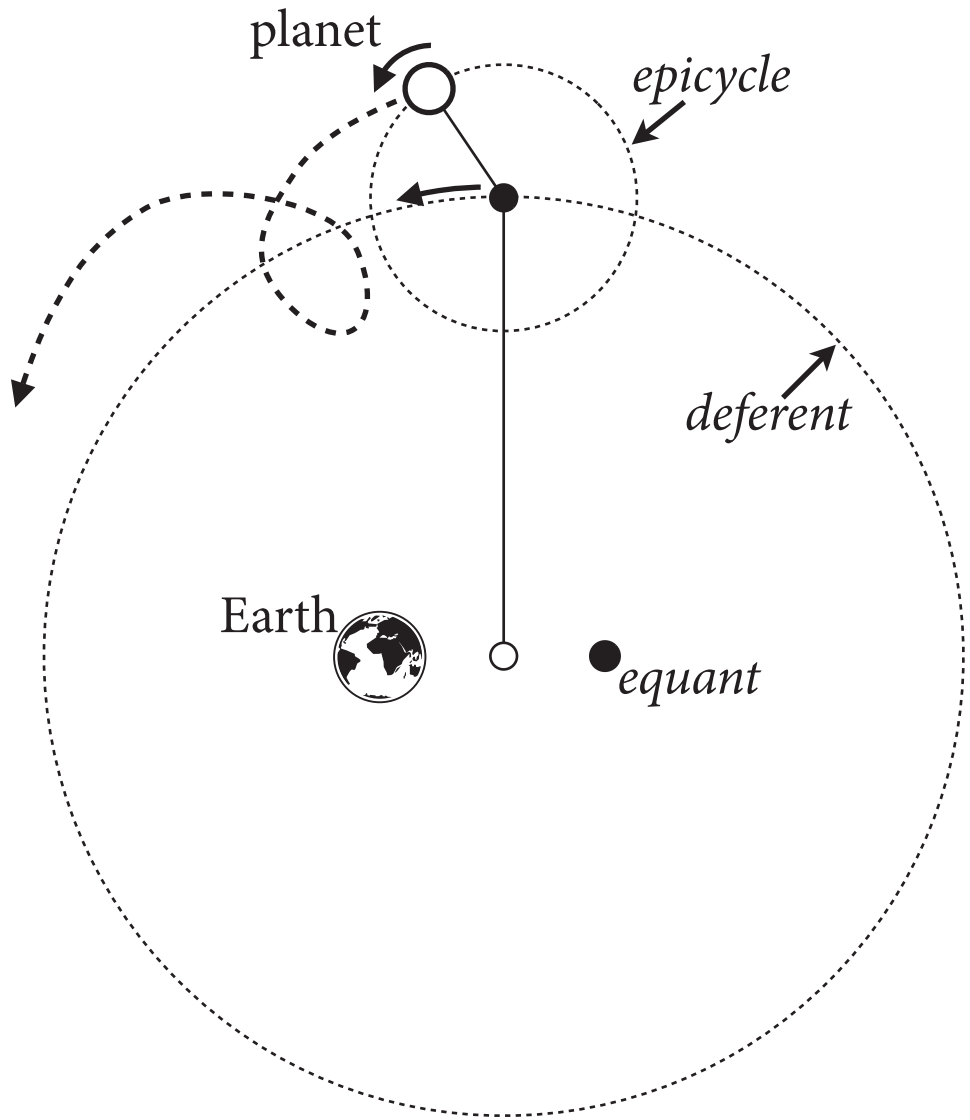
Week 2:
Linear Models

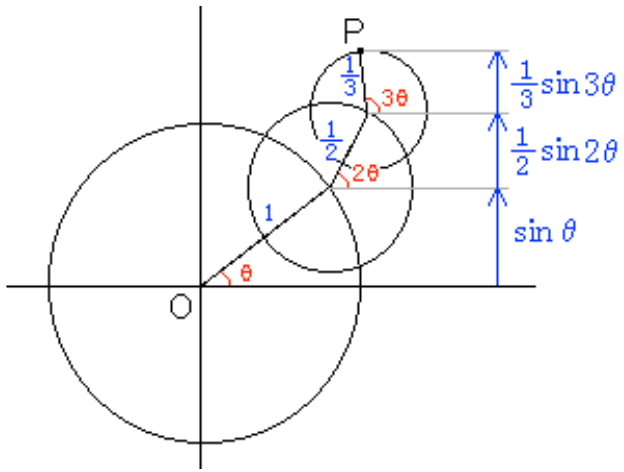
Richard McElreath

Triumph of Geocentrism

- Claudius Ptolemy (90–168)
  - Egyptian mathematician
  - Accurate model of planetary motion
  - Epicycles: orbits on orbits
  - Fourier series

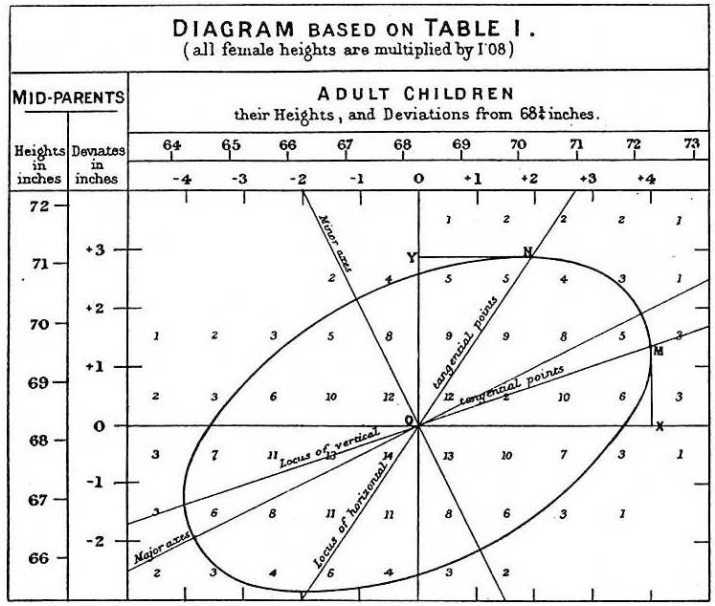






Geocentrism

- Descriptively accurate
- Mechanistically wrong
- General method of approximation
- Known to be wrong

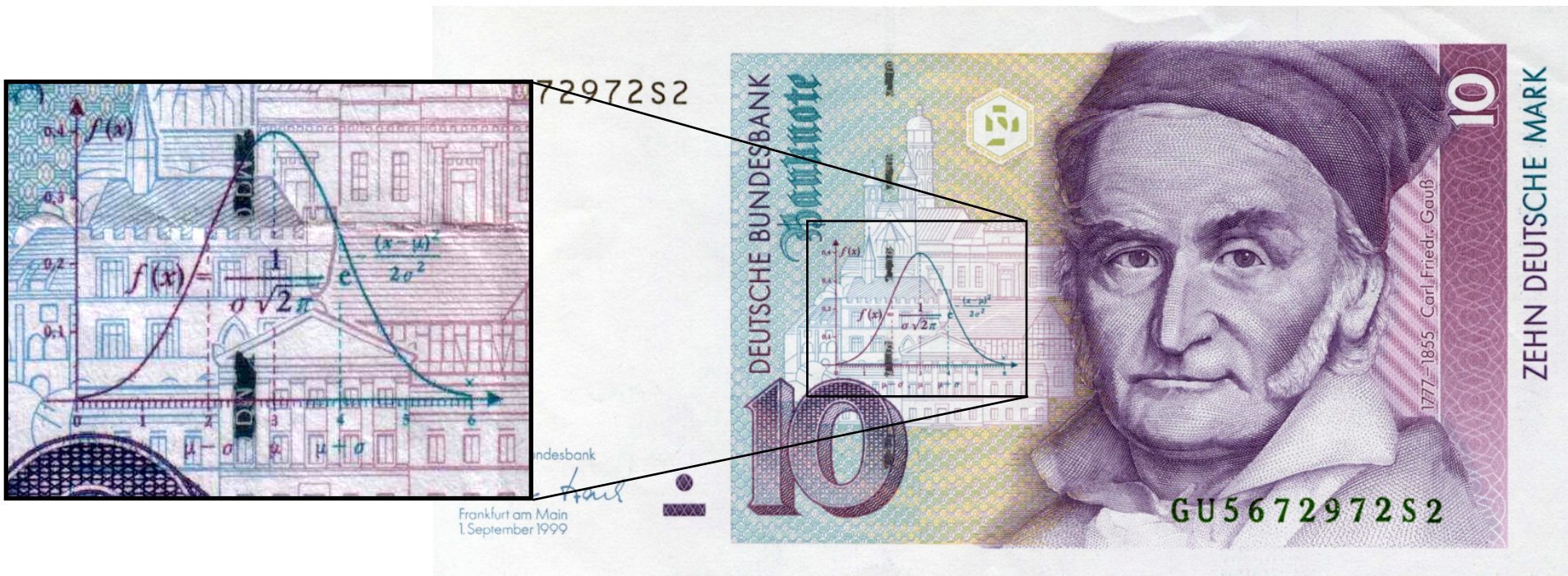


Regression

- Descriptively accurate
- Mechanistically wrong
- General method of approximation
- Taken too seriously

Linear regression

- Simple statistical golems
  - Model of mean and variance of normally (Gaussian) distributed measure
  - Mean as additive combination of predictors
  - Constant variance



THEORIA MOTVS CORPORVM COELESTIVM

IN

SECTIONIBVS CONICIS SOLEM AMBIENTIVM

AVCTORE

CAROLO FRIDERICO GAVSS

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

Venditur

PARISIIS ap. Treuttel & Würtz. LONDINI ap. R. H. Evans.
STOCKHOLMIAE ap. A. Wiborg. PETROPOLI ap. Klostermann.
MADRITI ap. Sancha. FLORENTIAE ap. Molini, Landi & C.
AMSTELODAMI in libraria: Kunst- und Industrie-Comptoir, dicta.

1809 Bayesian argument for normal error and least-squares estimation

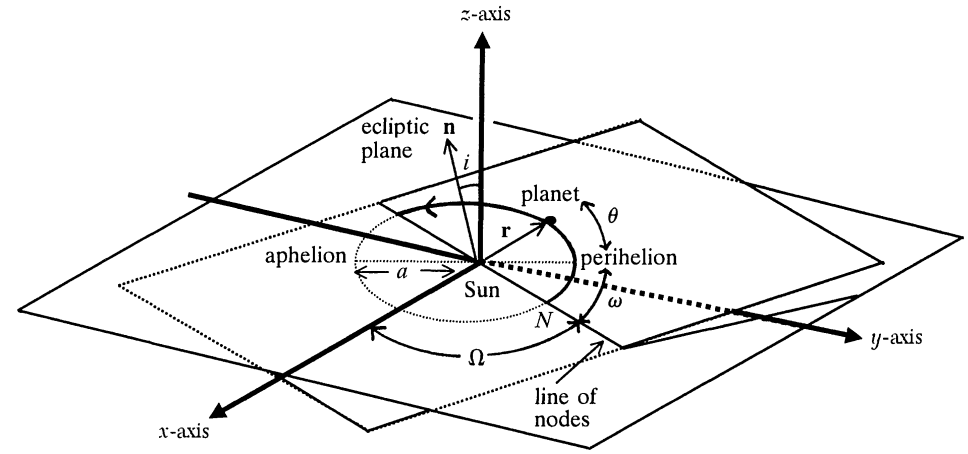
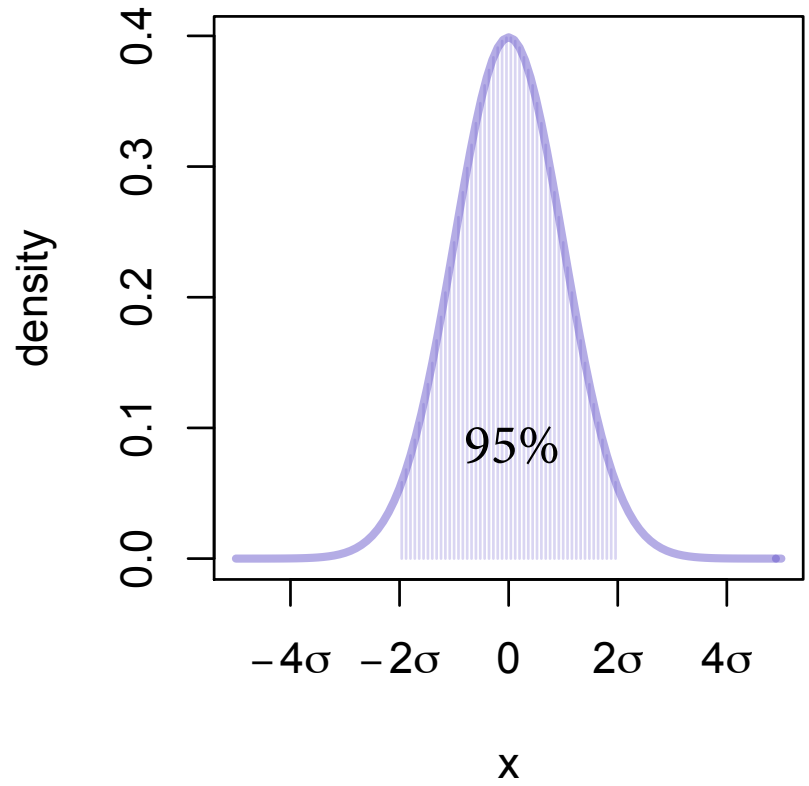


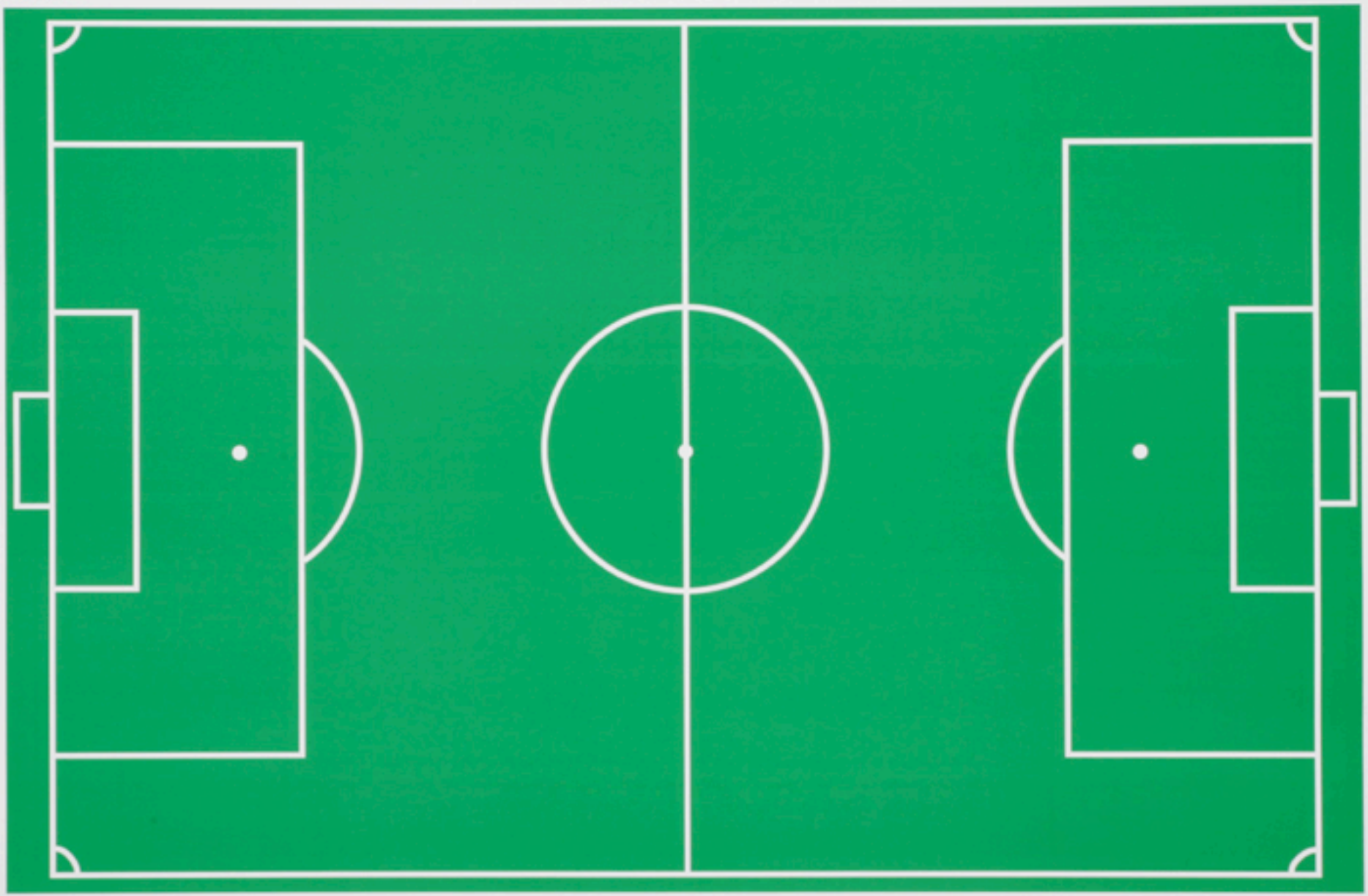
FIGURE 1
Parameters describing the planetary orbit

Why normal?

- Why are normal (Gaussian) distributions so common in statistics?
  1. Easy to calculate with
  2. Common in nature
  3. Very conservative assumption



Football

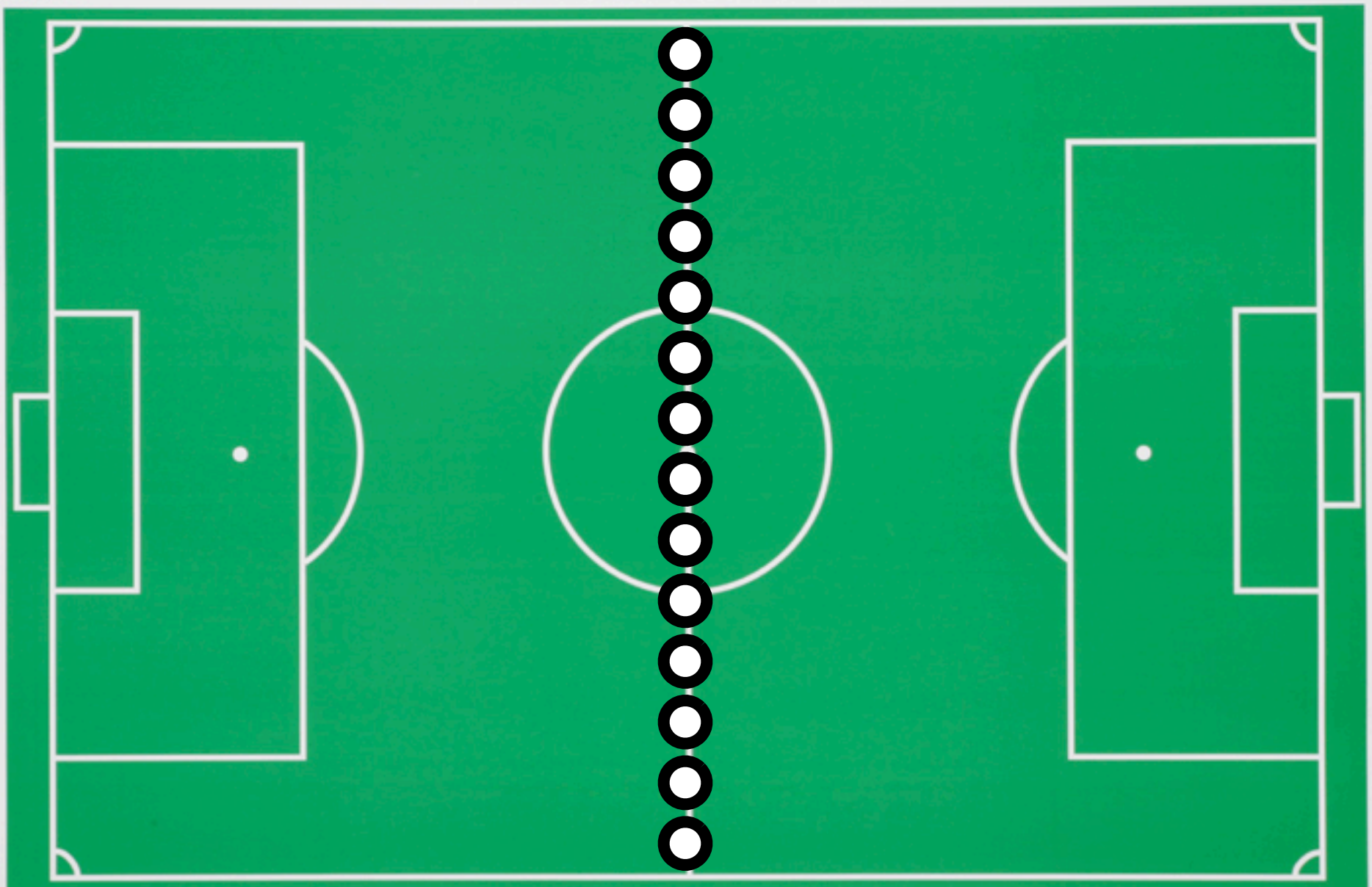


***Game Plan* - Strategy Overlay**

Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.

www.pnpsports.co.uk

Football

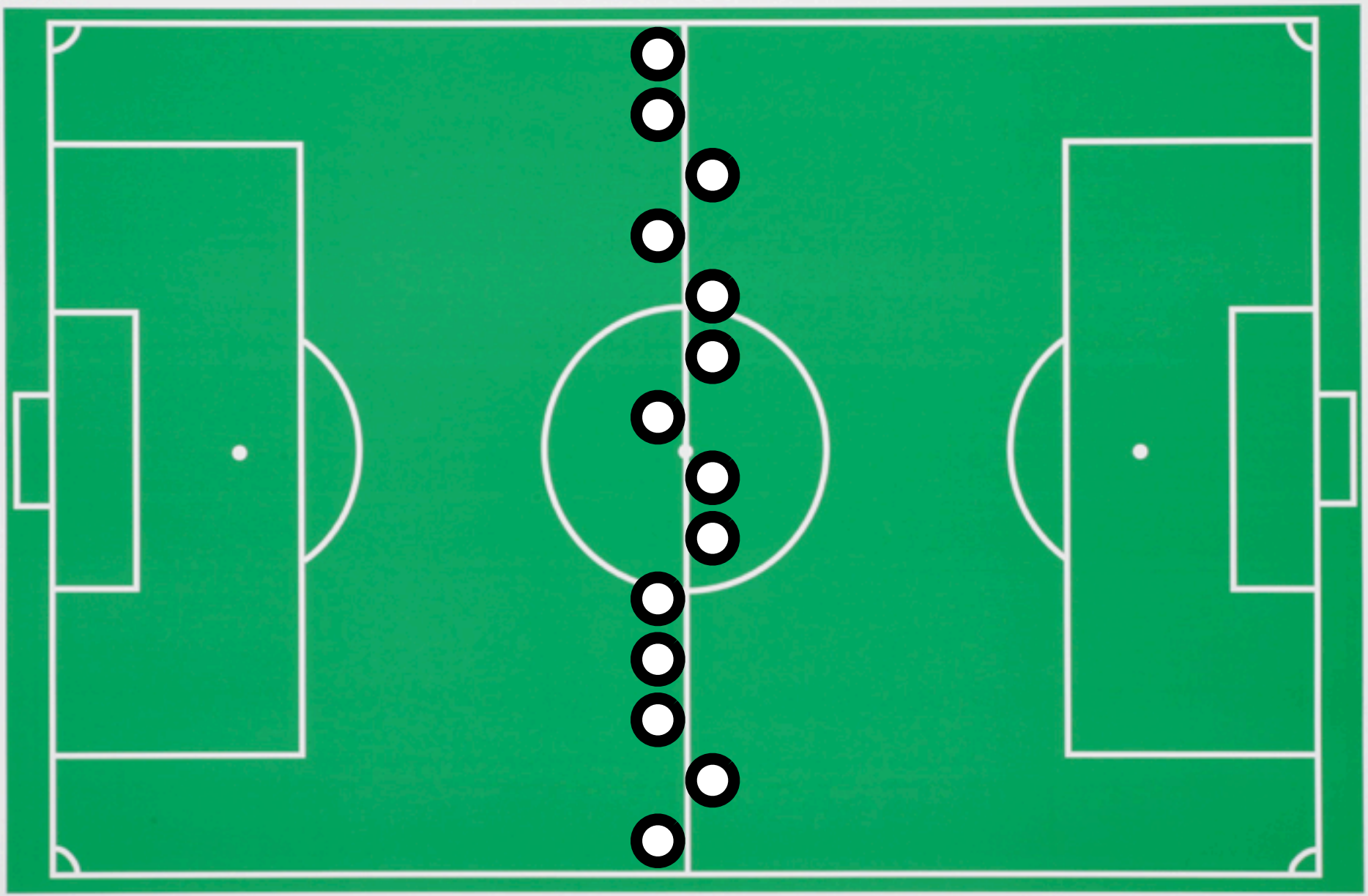


***Game Plan* - Strategy Overlay**

Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.

www.pnpsports.co.uk

Football

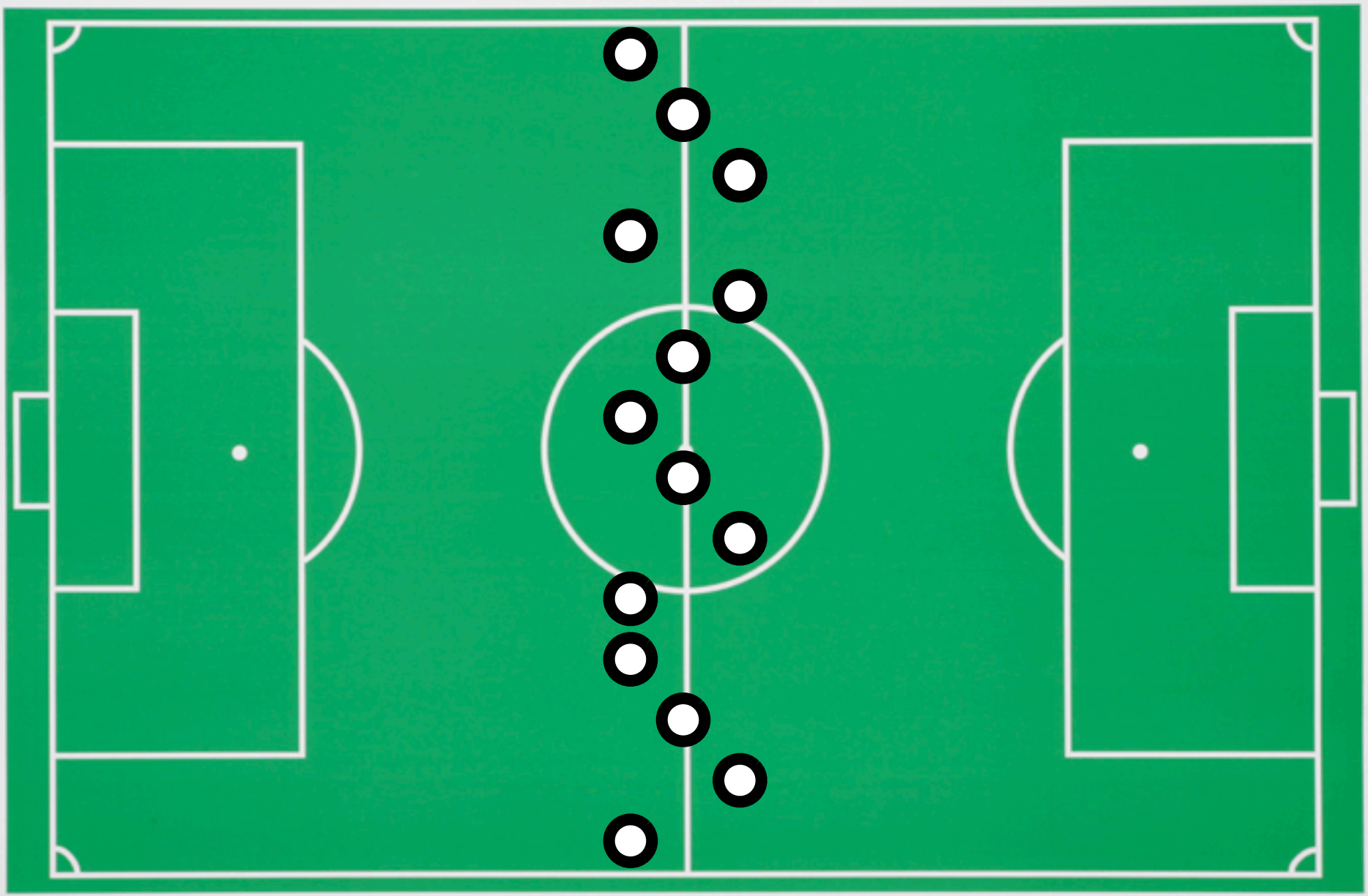


Game Plan - Strategy Overlay

Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.

www.pnpsports.co.uk

Football

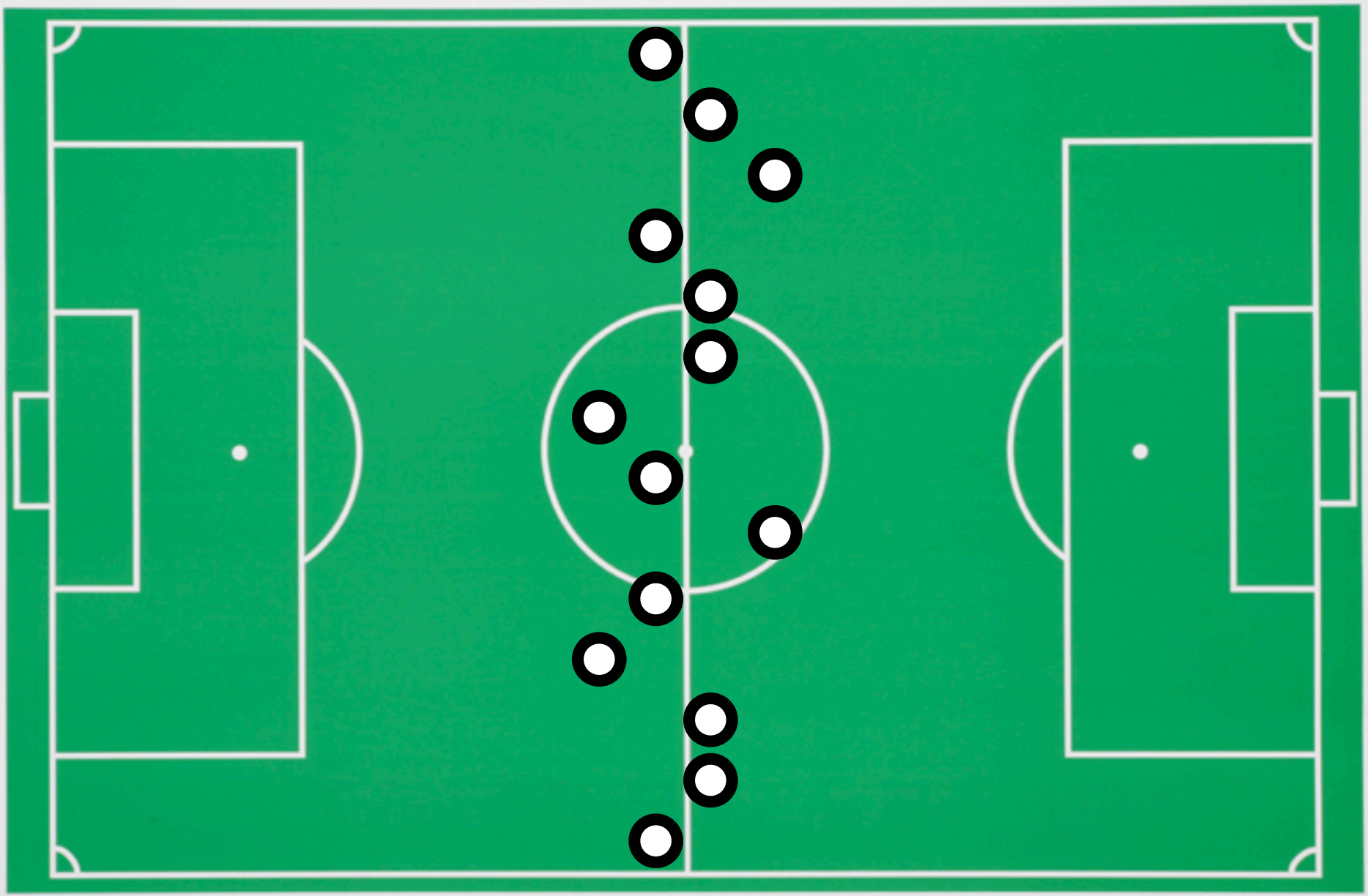


Game Plan - Strategy Overlay

Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.

www.pnpsports.co.uk

Football



Game Plan - Strategy Overlay

Use in conjunction with a magnetic board such as the A2 Folding Wedge, use only dry-wipe markers, clean with a soft cloth as harsh or abrasive solvents will damage the surface.

www.pnpsports.co.uk

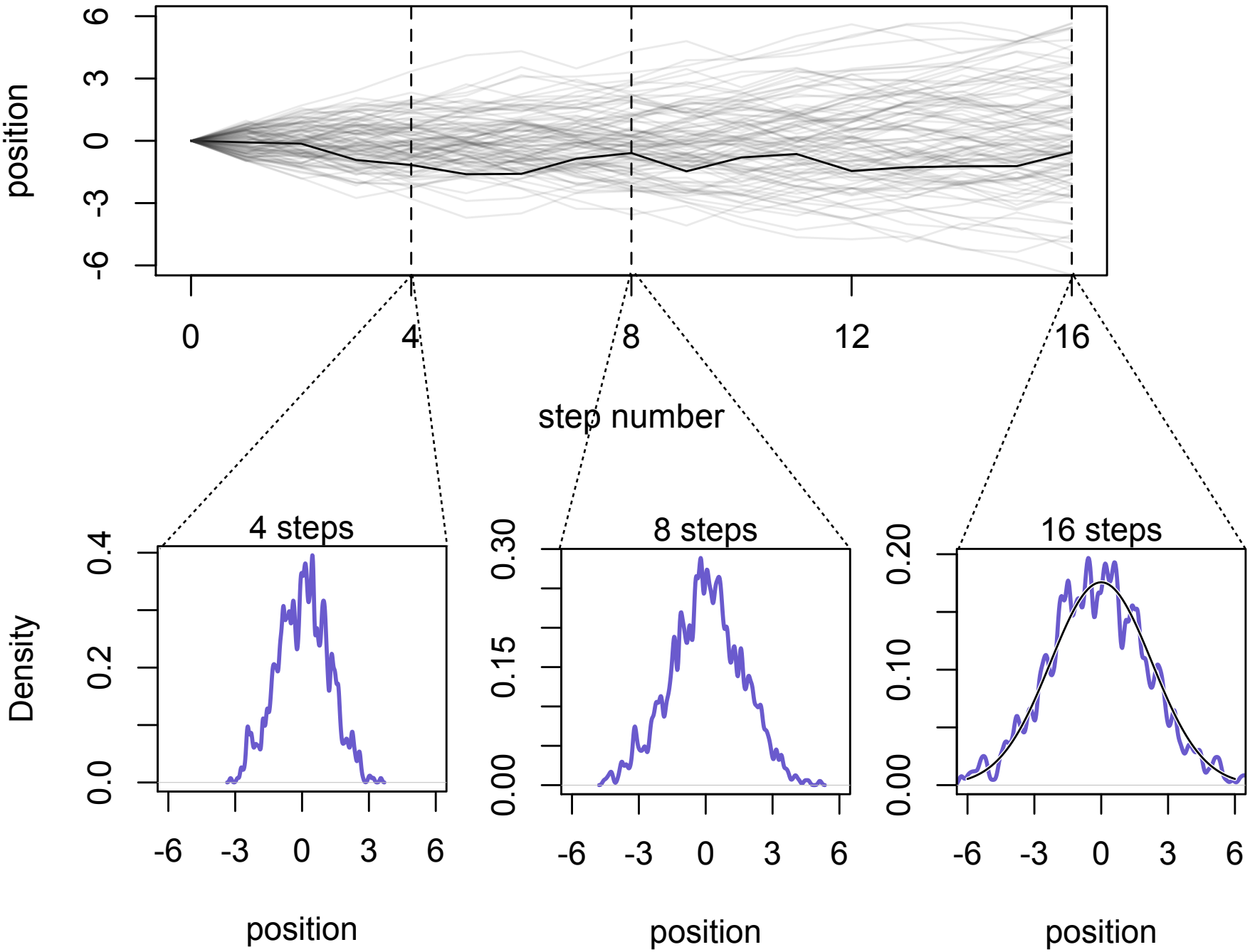
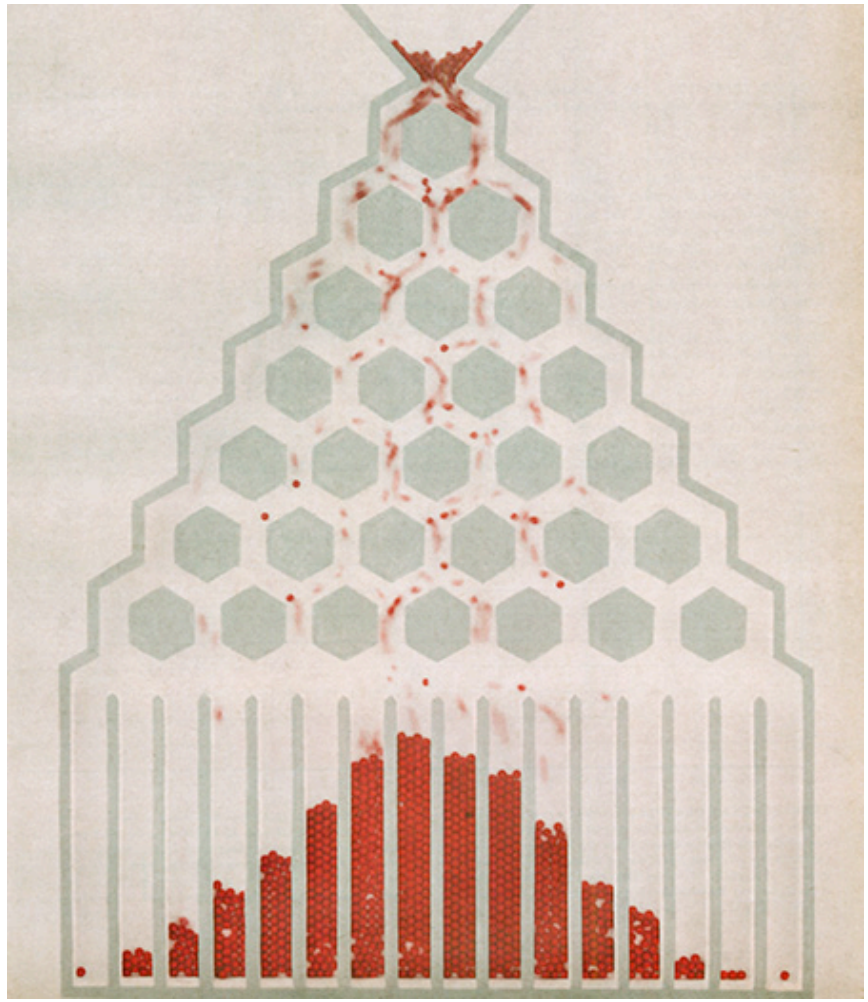

position
6
3
0
-3
-6
0
4
8
12
16
step number
4 steps
8 steps
16 steps
Density
0.4
0.2
0.0
0.30
0.15
0.00
0.20
0.10
0.00
-6
-3
0
3
6
position

Figure 4.2

Why normal?

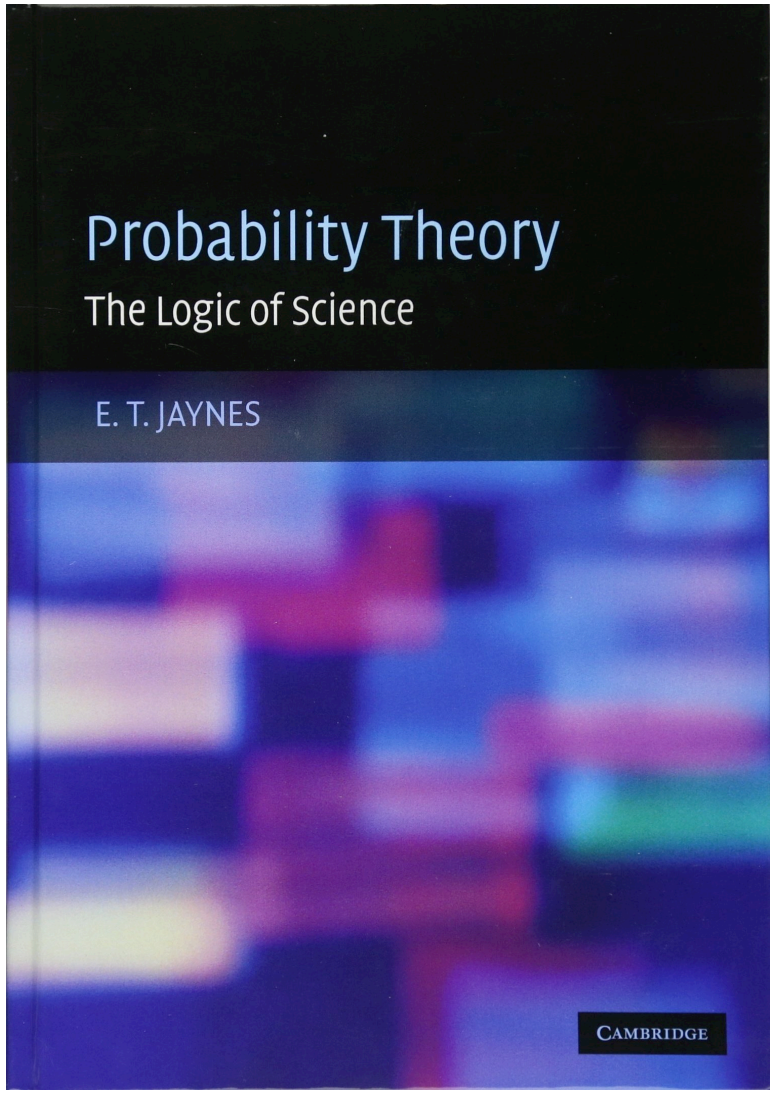
- Processes that produce normal distributions
  - Addition
  - Products of small deviations
  - Logarithms of products



Francis Galton's 1894 "bean machine" for simulating normal distributions

Why normal?

- Ontological perspective
  - Processes which add fluctuations result in dampening
  - Damped fluctuations end up Gaussian
  - No information left, except mean and variance
  - Can't infer process from distribution!
- Epistemological perspective
  - Know only *mean* and *variance*
  - Then least surprising and most conservative (*maximum entropy*) distribution is Gaussian
  - Nature likes maximum entropy distributions



Linear models

- Models of normally distributed data common
  - "General Linear Model": *t*-test, single regression, multiple regression, ANOVA, ANCOVA, MANOVA, MANCOVA, yadda yadda yadda
  - All the same thing
- Learn strategy, not procedure

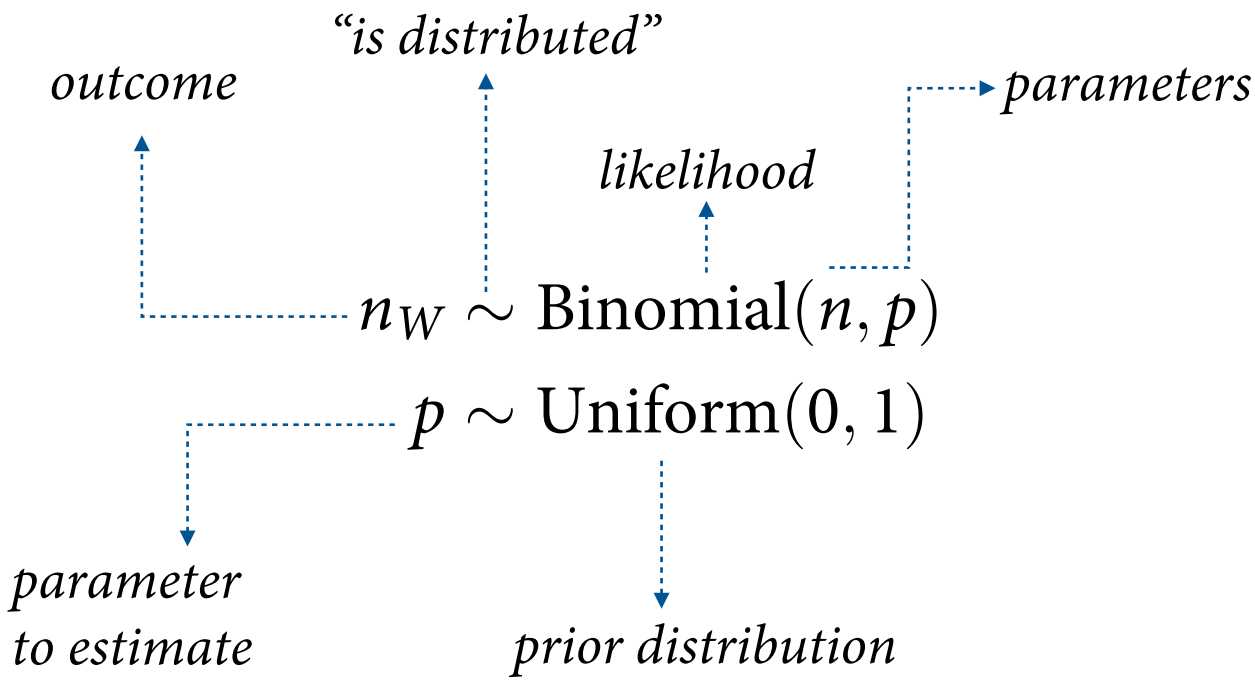
Language for modeling

- Questions to answer
  1. What are the outcomes?
  2. How are the outcomes constrained (what is *likelihood*)?
  3. What are the predictors, if any?
  4. How do predictors relate to *likelihood*?
  5. What are the *priors*?

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

Language for modeling

- Revisit globe tossing model:



Language for modeling

- Revisit globe tossing model:

$$n_W \sim \text{Binomial}(n, p)$$
$$p \sim \text{Uniform}(0, 1)$$

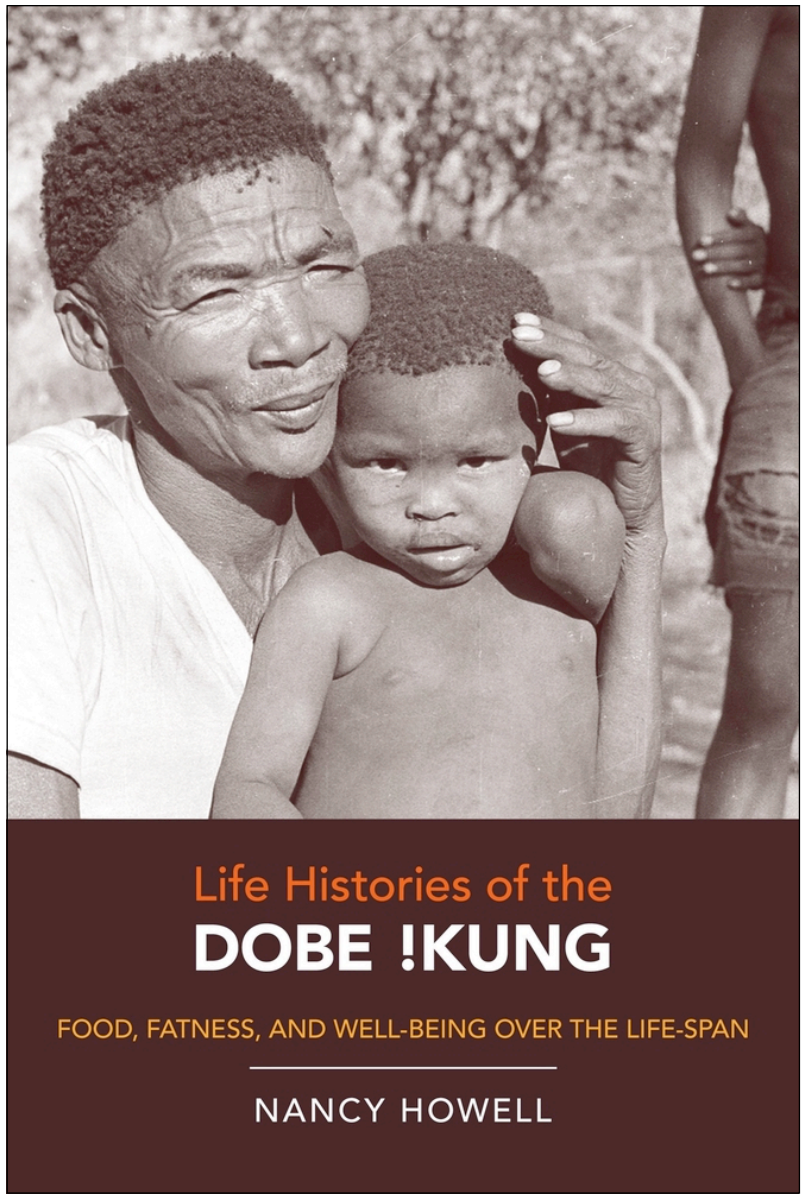
The count n_W is distributed binomially with sample size n and probability p. The prior for p is assumed to be uniform between zero and one.

Some data: Kalahari foragers

R code 4.6

```
library(rethinking)
data(Howell1)
d <- Howell1
```

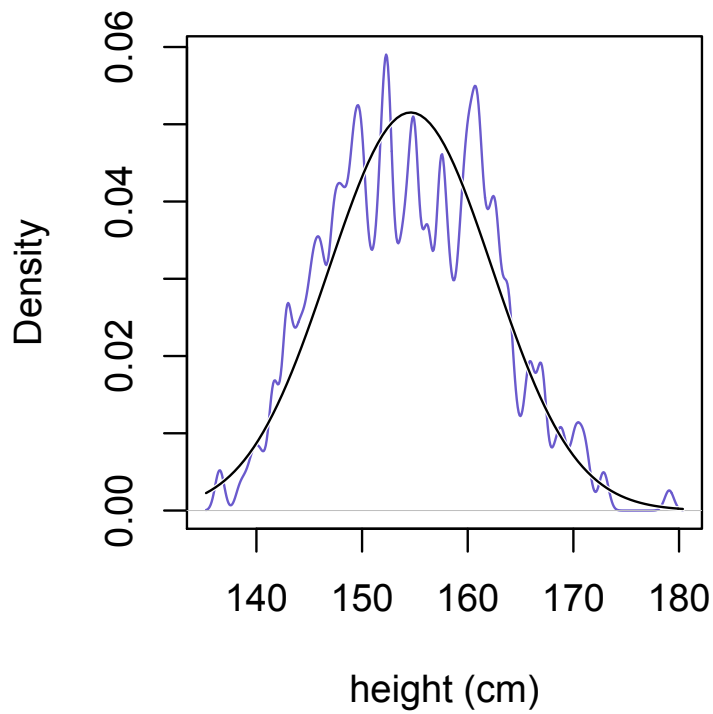
```
     height  weight   age male
1   151.765 47.82561  63    1
2   139.700 36.48581  63    0
3   136.525 31.86484  65    0
4   156.845 53.04191  41    1
5   145.415 41.27687  51    0
6   163.830 62.99259  35    1
                ...
544 158.750 52.53162  68    1
```



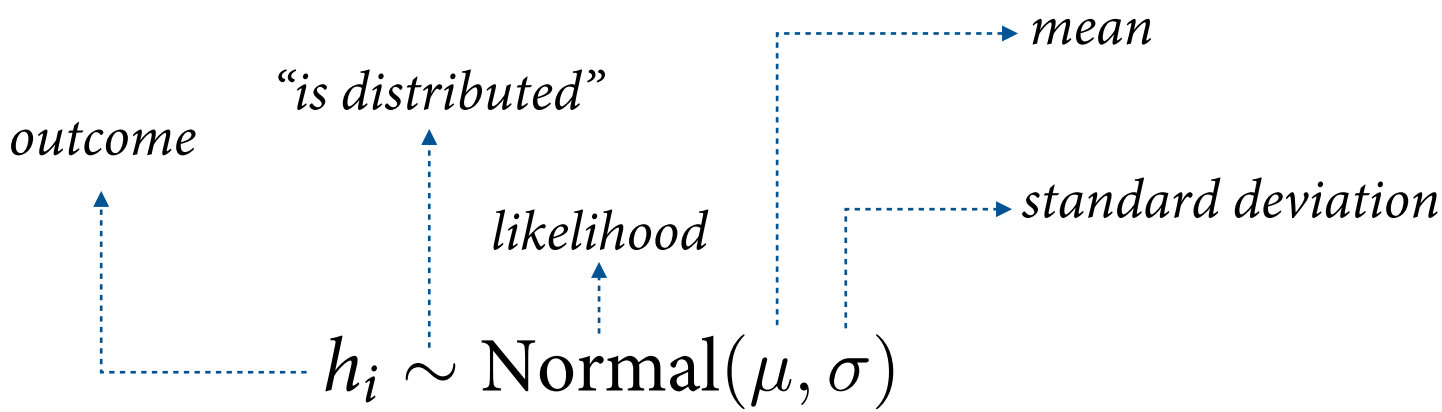
Gaussian model

- A first model:

$$h_i \sim \text{Normal}(\mu, \sigma)$$



Gaussian model

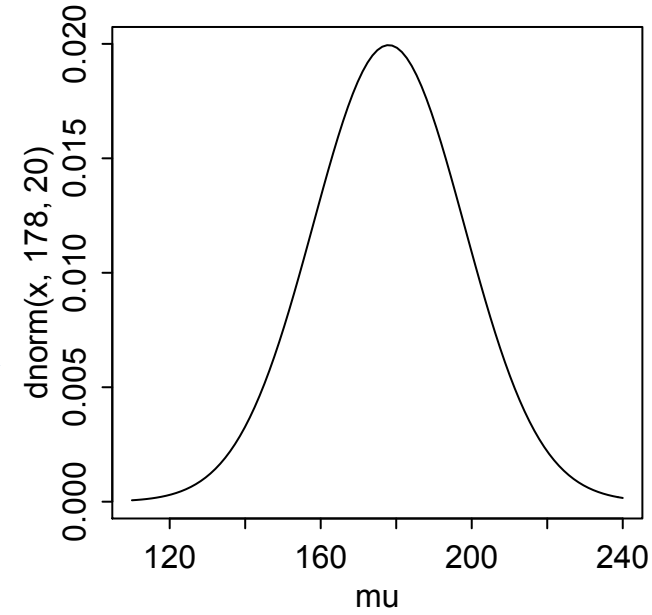


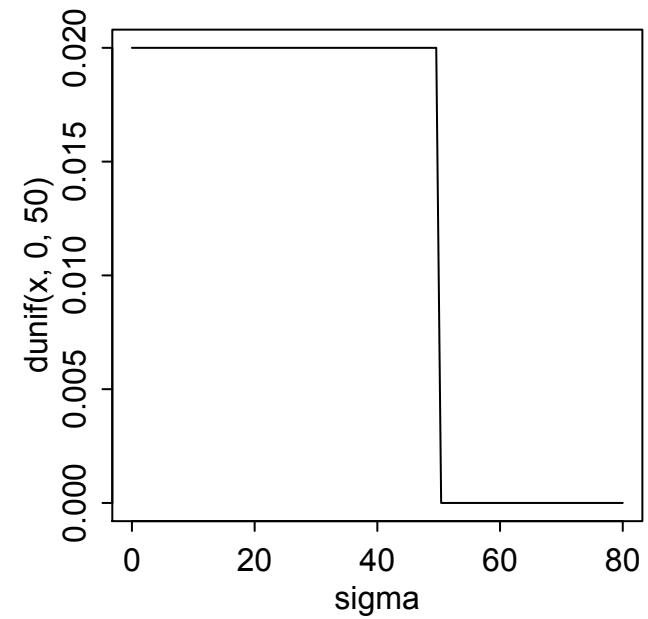
Height h_i of an individual i is distributed normally, with mean *mu* and standard deviation *sigma*.

Gaussian model

- Add priors:

$$h_i \sim \text{Normal}(\mu, \sigma)$$
$$\mu \sim \text{Normal}(178, 20)$$
$$\sigma \sim \text{Uniform}(0, 50)$$



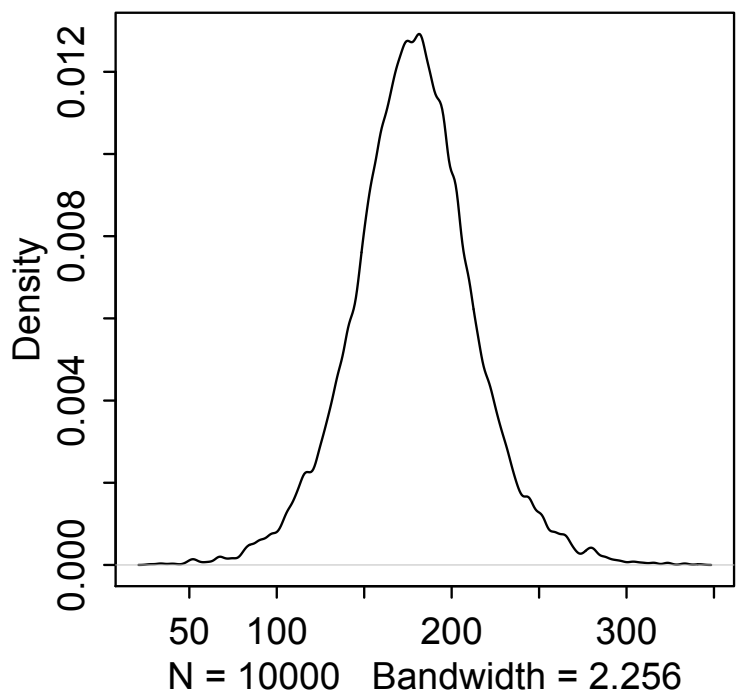


Gaussian model

- What do these priors imply about height, before we see data? Simulate! => *prior predictive distribution*

```
sample_mu <- rnorm( 1e4 , 178 , 20 )
sample_sigma <- runif( 1e4 , 0 , 50 )
prior_h <- rnorm( 1e4 , sample_mu , sample_sigma )
dens( prior_h )
```

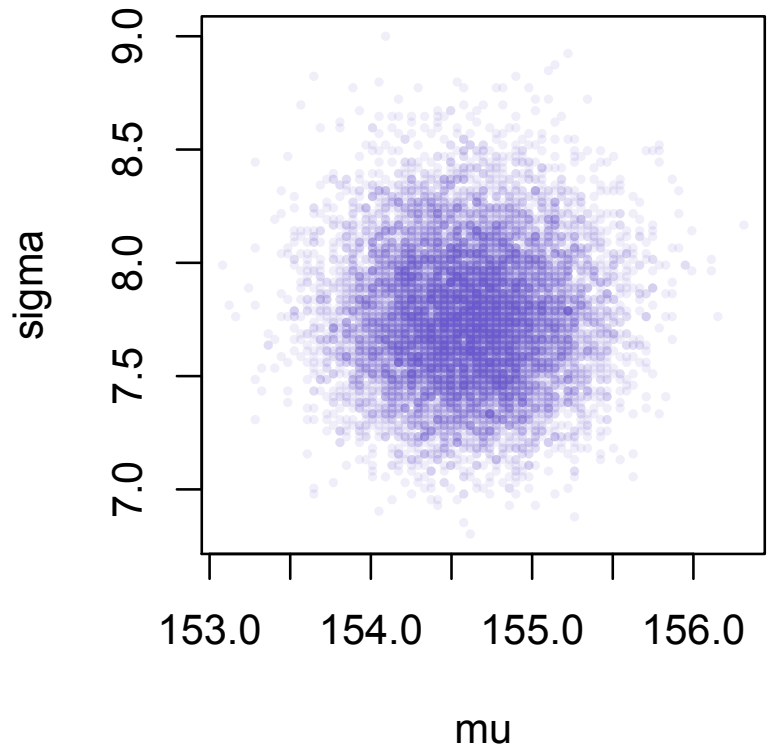
R code 4.13



100 cm = 3.3 feet
200 cm = 6.5 feet

Estimating *mu* and *sigma*

- Aim for the posterior distribution, which is now 2-dimensional
- Grid approximation: Compute posterior for many combinations of *mu* and *sigma*



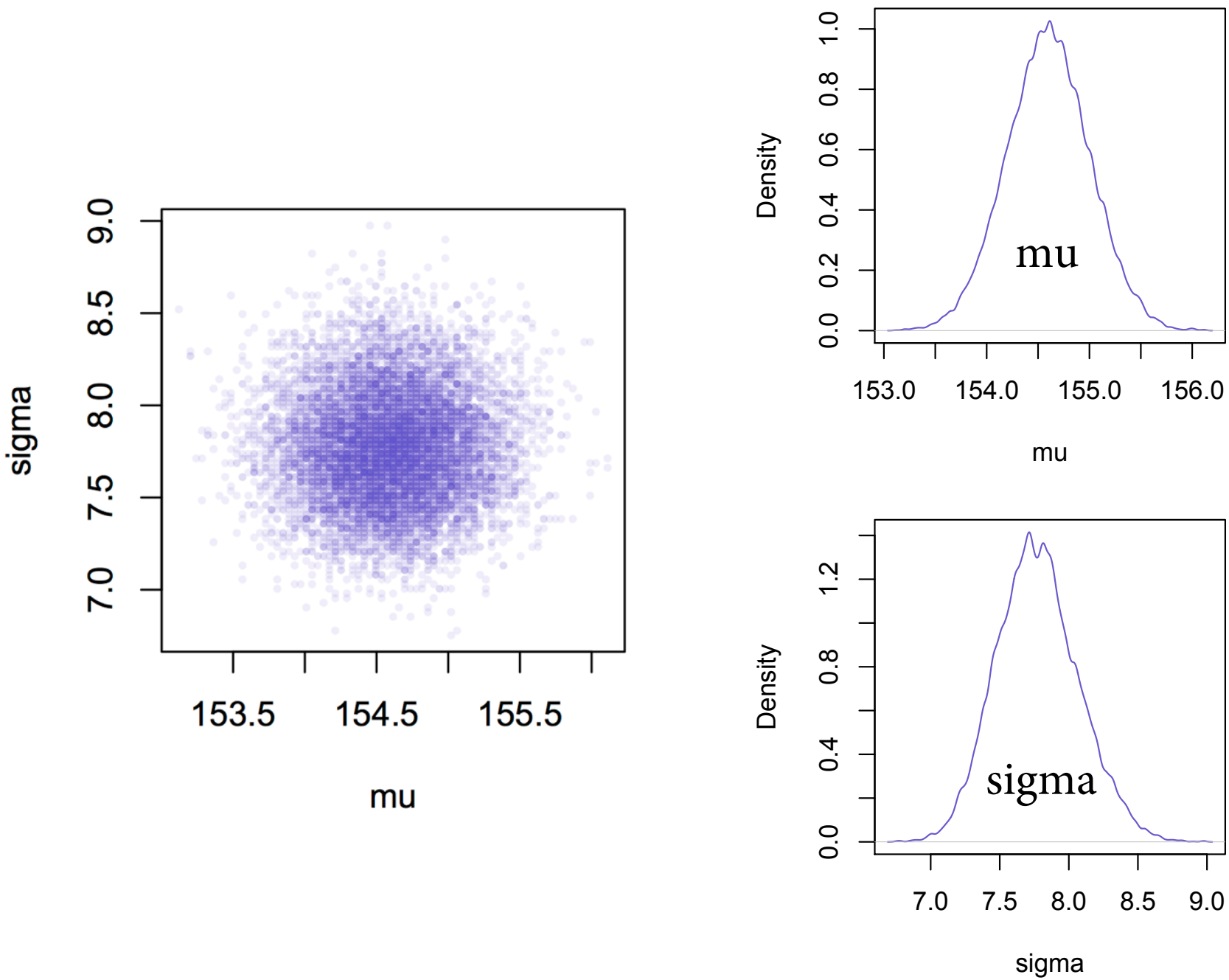

sigma
9.0
8.5
8.0
7.5
7.0
153.5
154.5
155.5
mu
Density
1.0
0.8
0.6
0.4
0.2
0.0
mu
153.0
154.0
155.0
156.0
mu
Density
1.2
0.8
0.4
0.0
sigma
7.0
7.5
8.0
8.5
9.0
sigma

Figure 4.3

Quadratic approximation

- Approximate posterior as Gaussian
- Can estimate with two numbers:
  - Peak of posterior, *maximum a posteriori* (MAP)
  - Standard deviation of posterior
- Lots of algorithms
- With flat priors, same as conventional *maximum likelihood estimation*

Using `map`

Maximum a posteriori

```
flist <- alist(
    height ~ dnorm( mu , sigma ) ,
    mu ~ dnorm( 178 , 20 ) ,
    sigma ~ dunif( 0 , 50 )
)
```

$$h_i \sim \text{Normal}(\mu, \sigma)$$
$$\mu \sim \text{Normal}(178, 20)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

Using `map`

```
flist <- alist(
    height ~ dnorm( mu , sigma ) ,
    mu ~ dnorm( 178 , 20 ) ,
    sigma ~ dunif( 0 , 50 )
)
```

R code 4.25

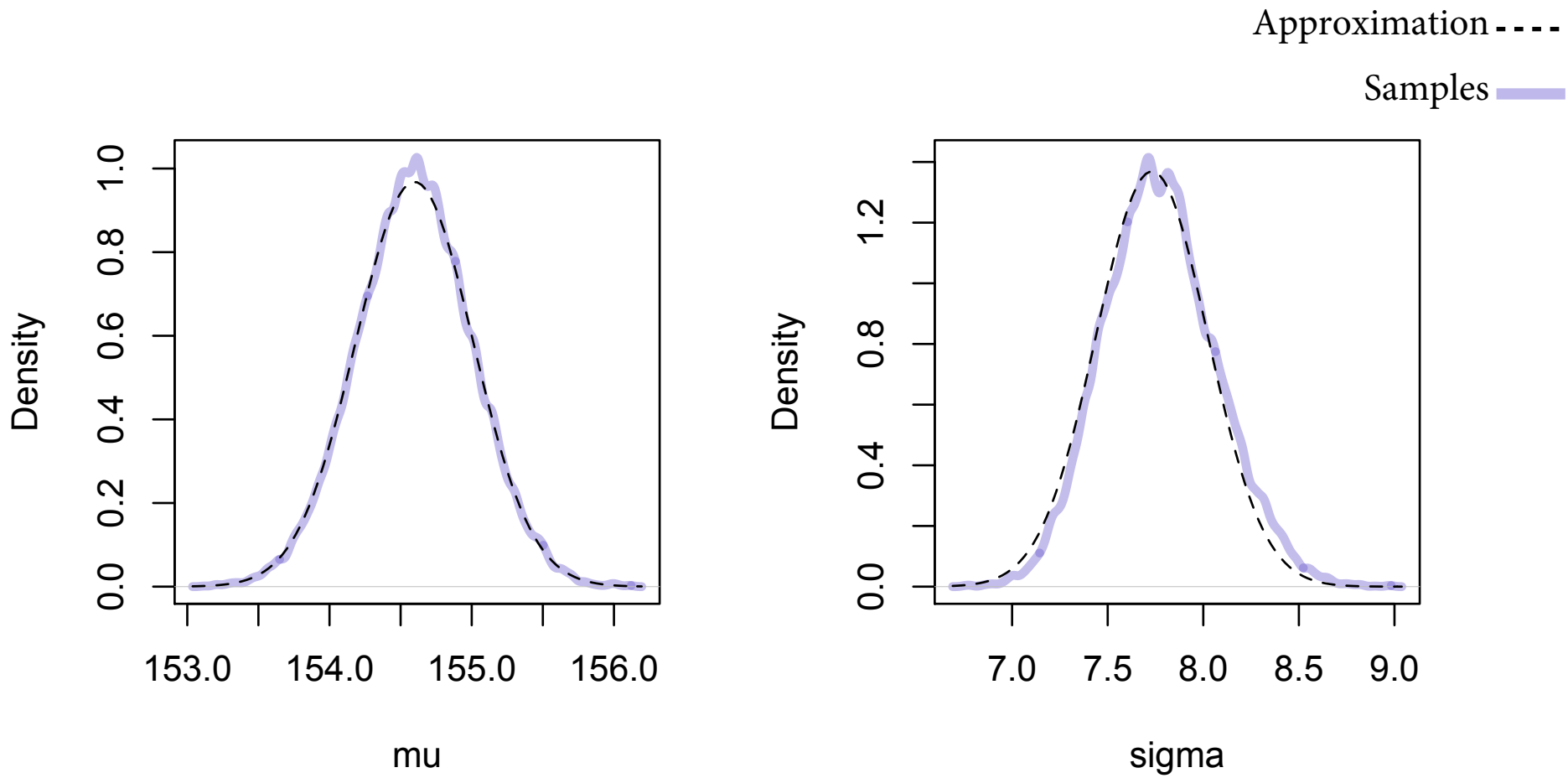
```
m4.1 <- map( flist , data=d2 )
```

R code 4.26

```
precis( m4.1 )
```

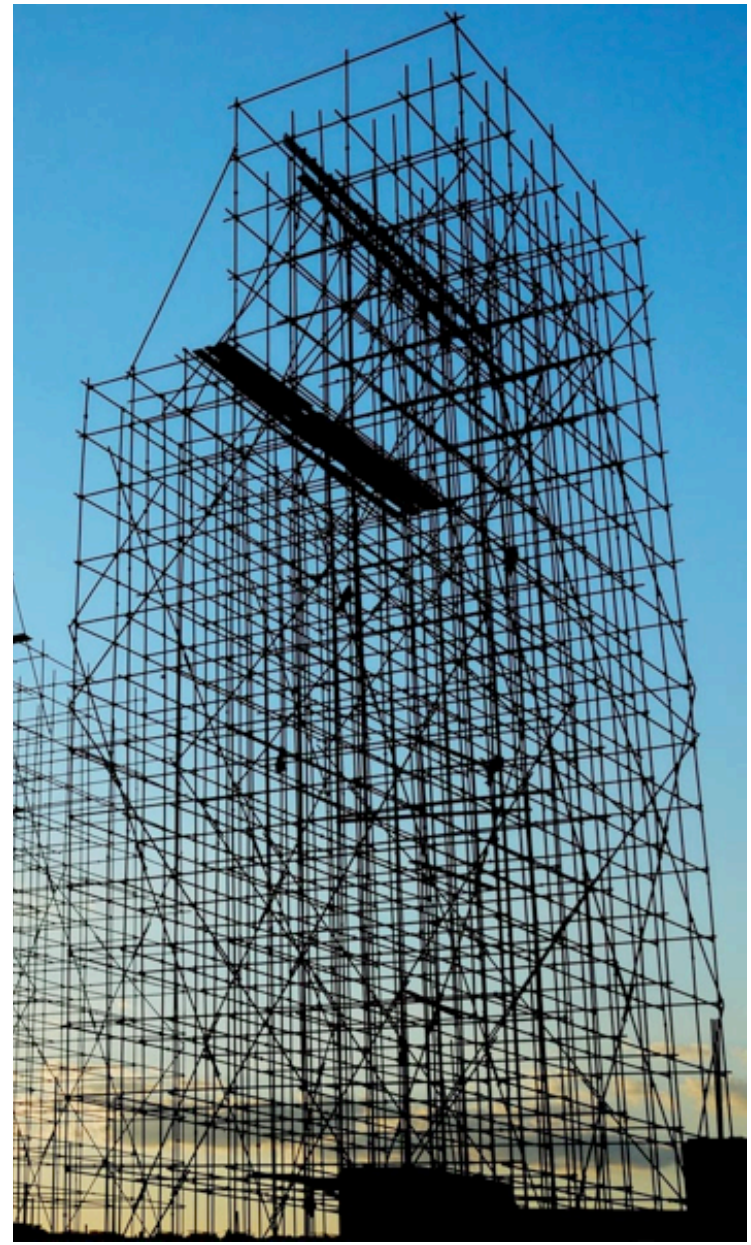
R code 4.27

```
          Mean StdDev   5.5%  94.5%
mu     154.61   0.41 153.95 155.27
sigma    7.73   0.29   7.27   8.20
```



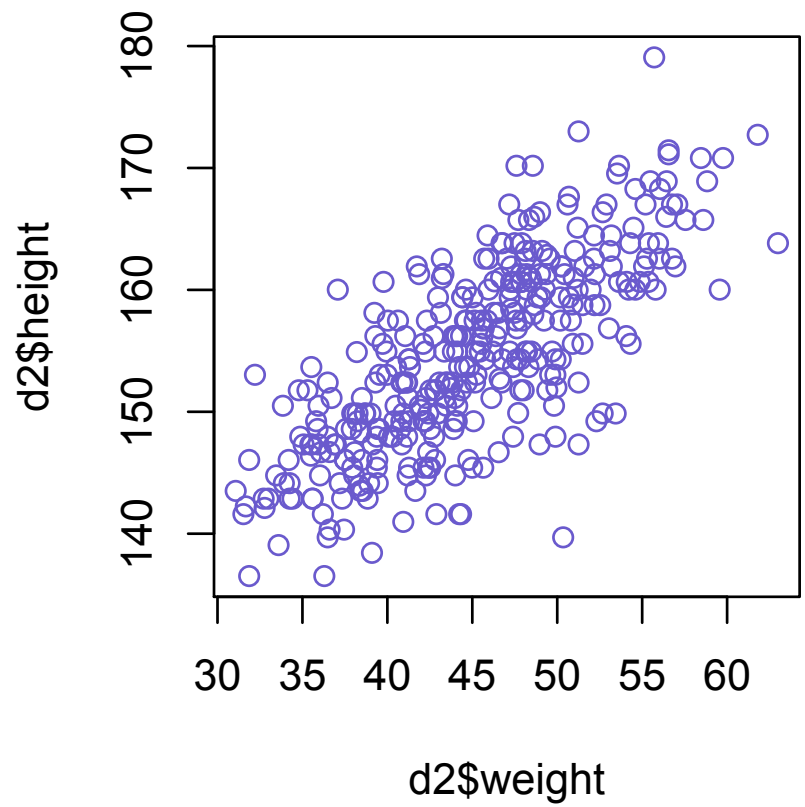
Scaffolds

- `map` is a scaffold
  - Forces full specification of model, so you learn it
  - Works with a very wide class of models
  - Not really a good way to approximate posterior



Adding a predictor variable

- How does weight describe height?

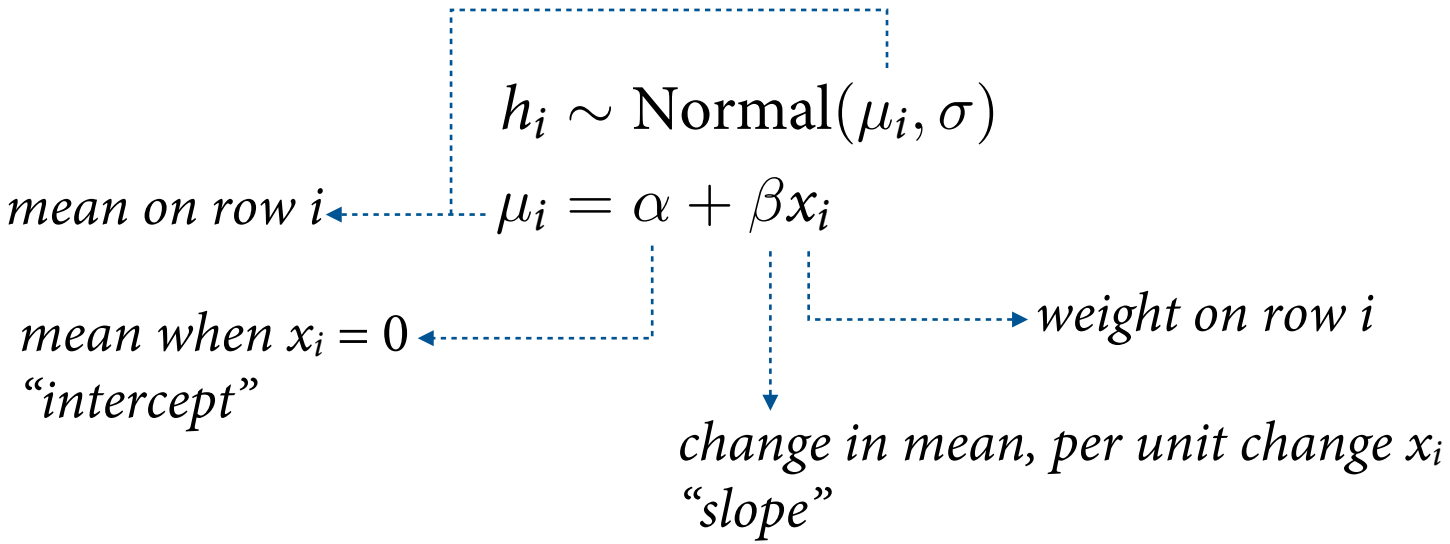


Adding a predictor variable

- Use a linear model of the mean, *mu*:

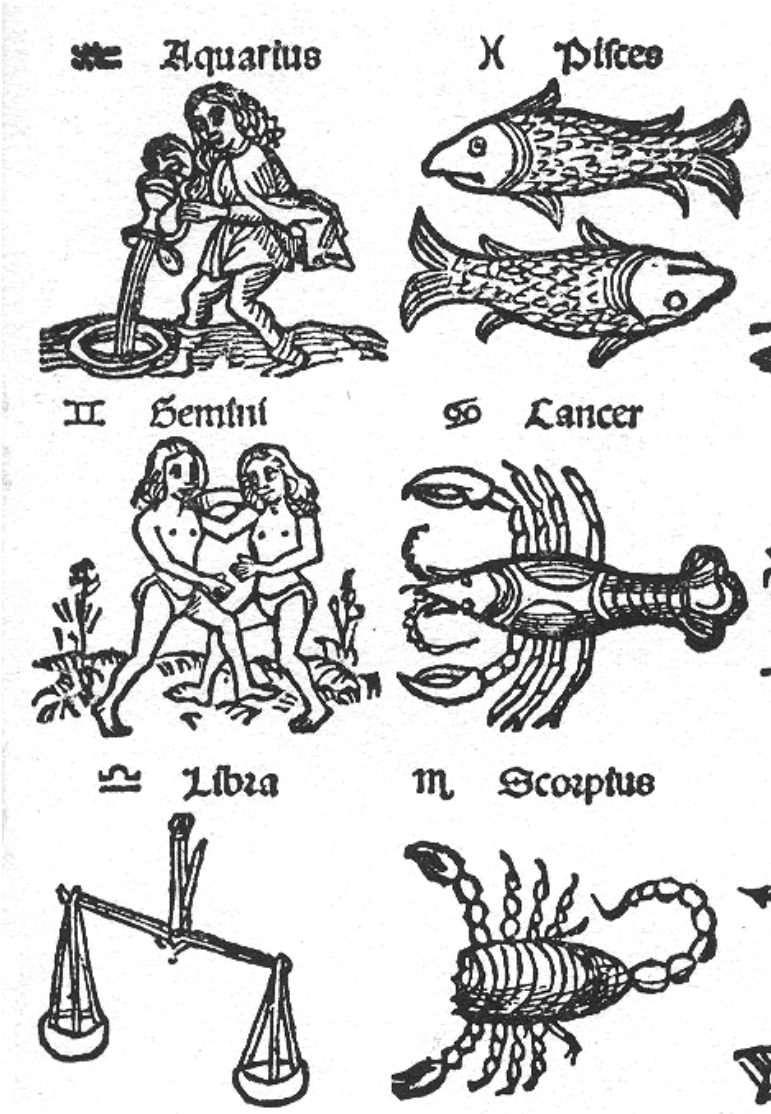
$$h_i \sim \text{Normal}(\mu_i, \sigma) \quad \text{[likelihood]}$$
$$\mu_i = \alpha + \beta x_i \quad \text{[linear model]}$$
$$\alpha \sim \text{Normal}(178, 100) \quad [\alpha \text{ prior}]$$
$$\beta \sim \text{Normal}(0, 10) \quad [\beta \text{ prior}]$$
$$\sigma \sim \text{Uniform}(0, 50) \quad [\sigma \text{ prior}]$$

Adding a predictor variable



Linear regression priors

- Horoscopic advice
  - **Intercept**, "alpha": no idea where it might end up, so broad Gaussian prior
  - **Slopes**, "beta": Gaussian, center on zero, scale so extreme estimates ruled out, "regularization" (Chapter 6)
  - **Scale**, "sigma": uniform with reasonable upper bound usually fine; later we'll use Cauchy or exponential for regularization
  - *Check prior predictive for sanity*



$$h_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta x_i$$
$$\alpha \sim \text{Normal}(178, 100)$$
$$\beta \sim \text{Normal}(0, 10)$$
$$\sigma \sim \text{Uniform}(0, 50)$$

```
height ~ dnorm(mu,sigma)
    mu <- a + b*weight
     a ~ dnorm(178,100)
     b ~ dnorm(0,10)
 sigma ~ dunif(0,50)
```

```
# fit model
m4.3 <- map(
    alist(
        height ~ dnorm( mu , sigma ) ,
        mu <- a + b*weight ,
        a ~ dnorm( 178 , 100 ) ,
        b ~ dnorm( 0 , 10 ) ,
        sigma ~ dunif( 0 , 50 )
    ) ,
    data=d2 )
```

These priors are terrible, but harmless here, because so much data

R code 4.40

```
precis( m4.3 )
```

```
          Mean StdDev   5.5%  94.5%
a       113.90   1.91 110.85 116.94
b         0.90   0.04   0.84   0.97
sigma     5.07   0.19   4.77   5.38
```

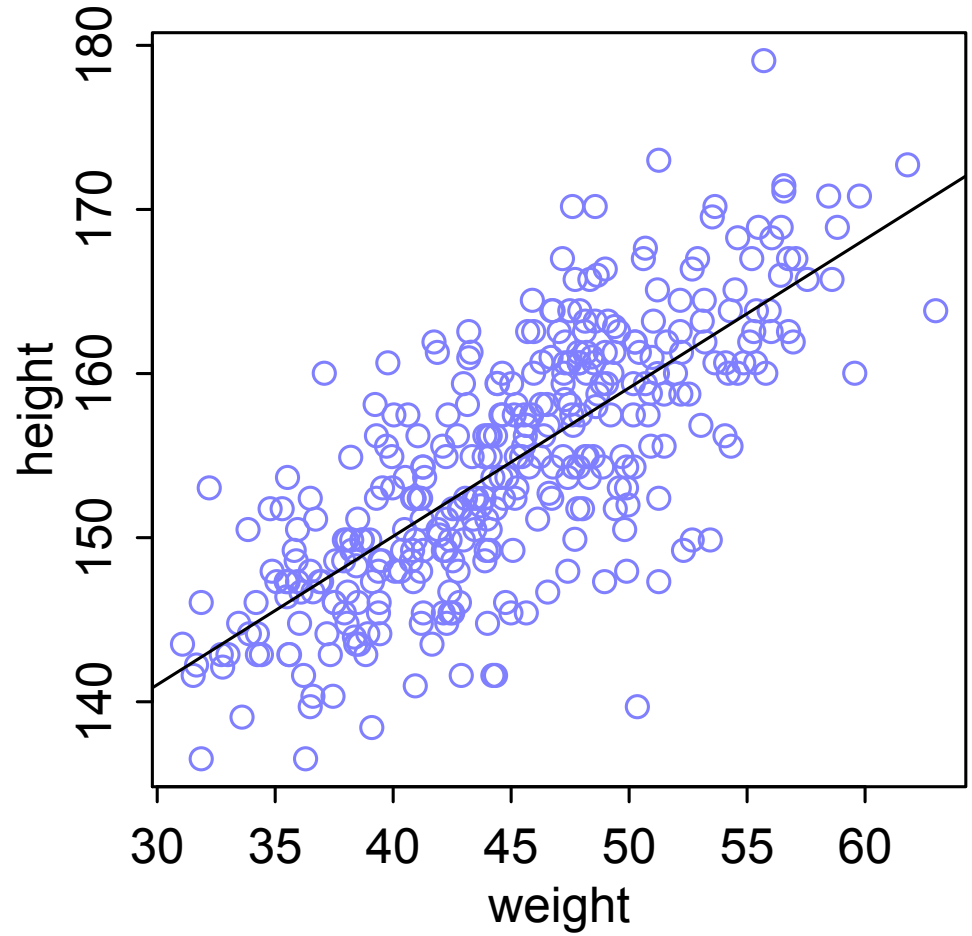


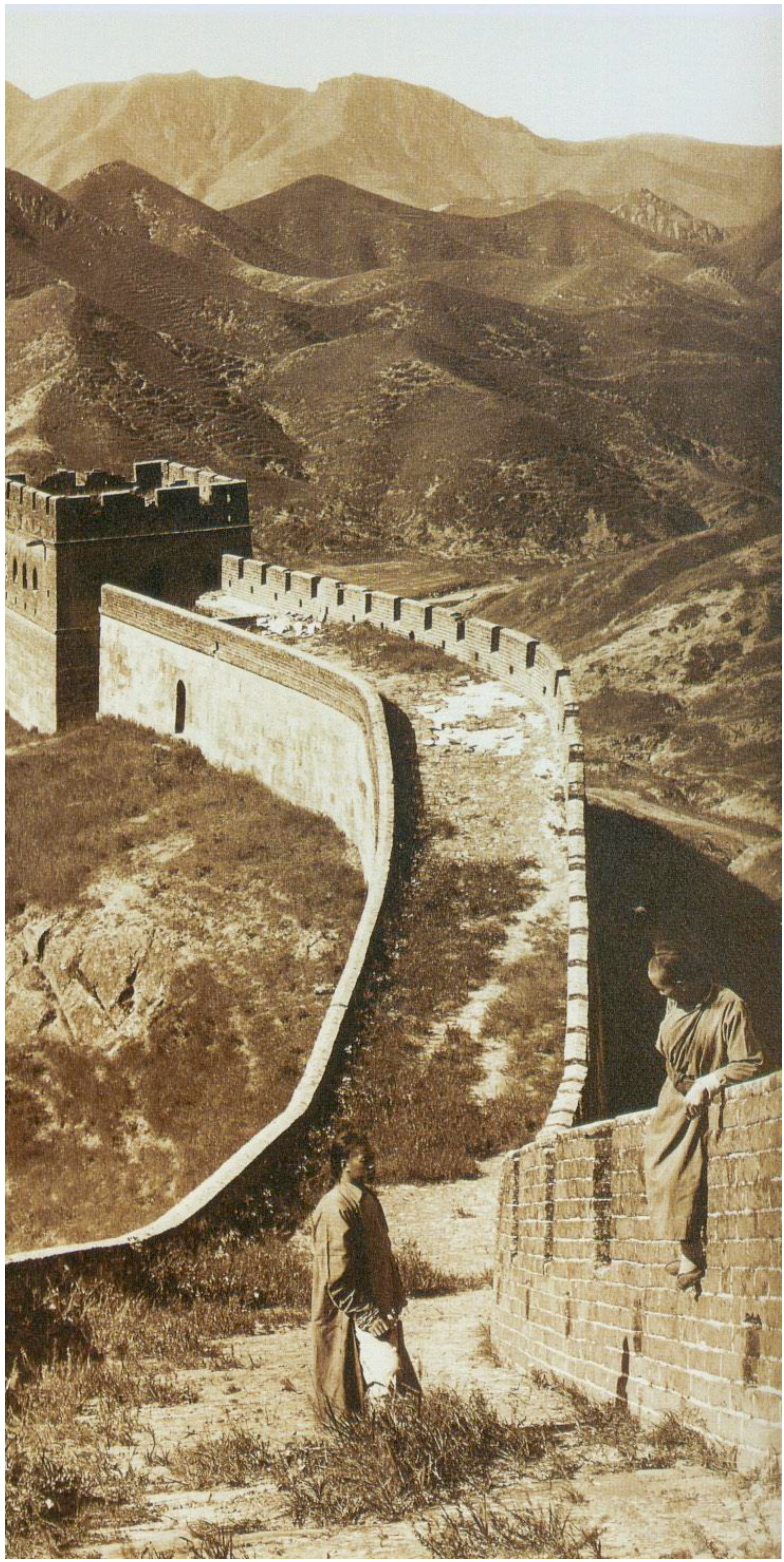
FIGURE 4.4. Height in centimeters (vertical) plotted against weight in kilograms (horizontal), with the *maximum a posteriori* line for the mean height at each weight plotted in black.

Sampling from the posterior

- Want to get uncertainty onto that graph
- Again, sample from posterior
  1. Use MAP and standard deviation to approximate posterior
  2. Sample from *multivariate normal* distribution of parameters
  3. Use samples to generate predictions that "integrate over" the uncertainty

Historical obstacles

- Prior education impedes learning
- “Sampling” in frequentist stats is a device to construct uncertainty around an estimate
- “Sampling” in Bayesian stats is a way to perform integral calculus (or to simulate observations)



Sampling from the posterior

```
post <- extract.samples( m4.3 )
```

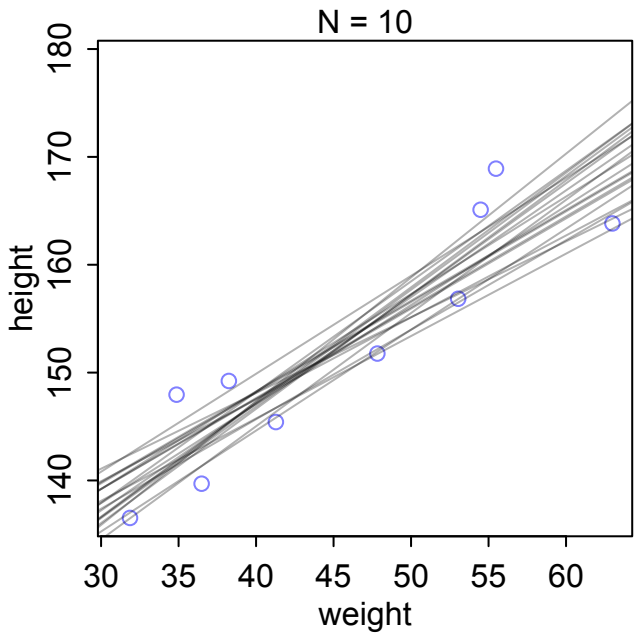
R code 4.46

```
post[1:5,]
```

R code 4.47

```
         a         b    sigma
1 114.7880 0.8822921 5.121102
2 112.7115 0.9230855 4.907987
3 114.4557 0.9018482 5.276036
4 114.7696 0.8831561 5.021958
5 112.6333 0.9383632 4.898554
```

Posterior is full of lines




```
post[1:5,]
```

```
          a    sigma         b
1  115.1964 4.992267 0.8776393
2  111.0389 5.169515 0.9758554
3  115.4833 5.133463 0.8726757
4  109.6488 5.005837 0.9812692
5  112.4637 4.678314 0.9384814
```

Figure 4.5