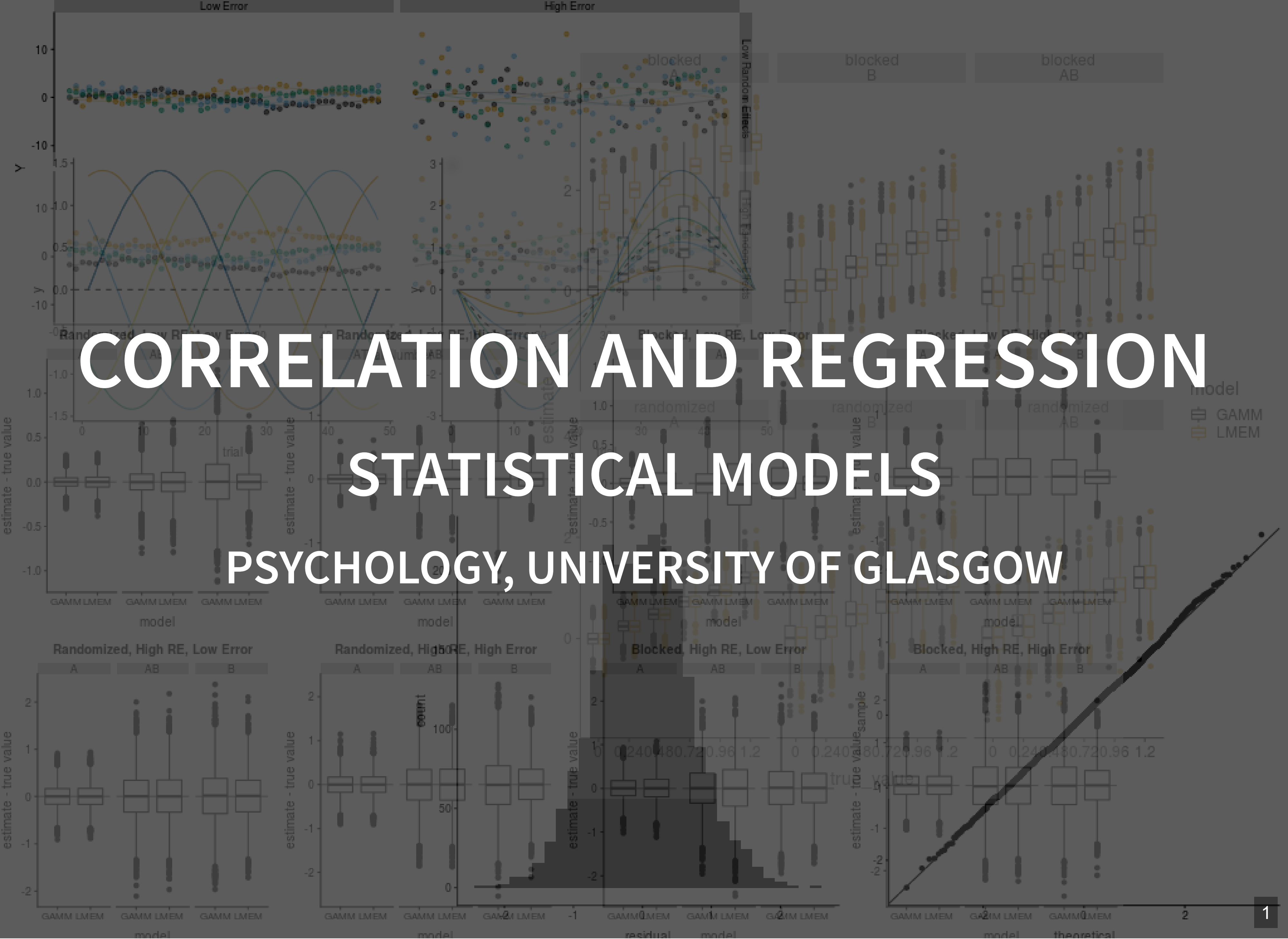


CORRELATION AND REGRESSION

STATISTICAL MODELS

PSYCHOLOGY, UNIVERSITY OF GLASGOW

WHAT YOU CAN EXPECT FROM ME

FEEDBACK IS A DIALOGUE

WHAT I AM SEEING

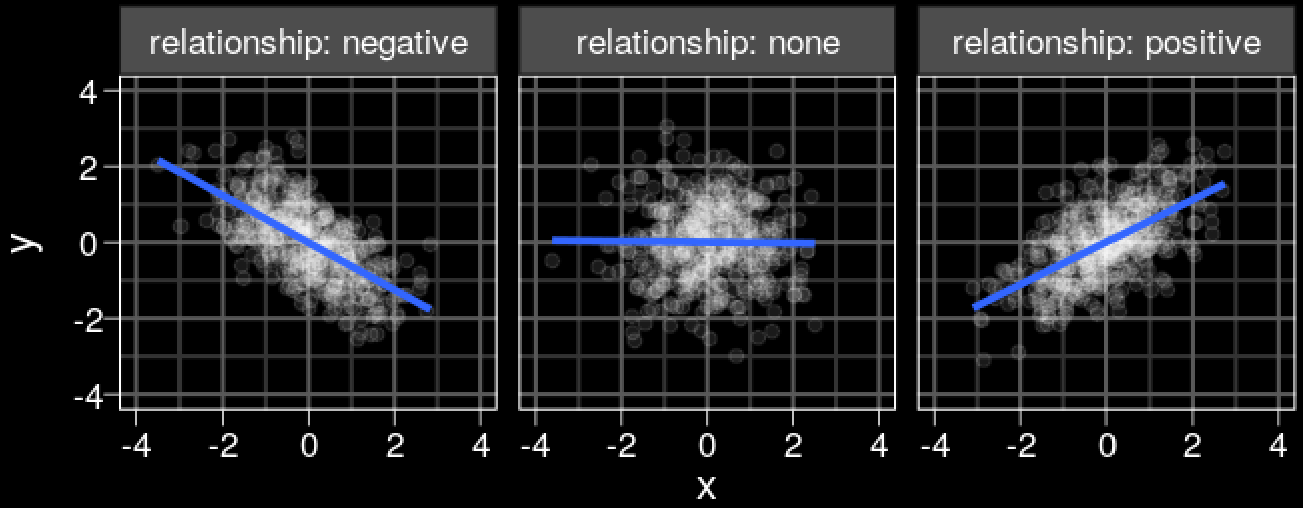
As of Oct 10:

- 20% of students turned in formative assignment 2
- far more students looked at solutions for assignment 1 (80%) than attempted it (50%)
- 50% of students have neither joined slack channel nor asked a question on Moodle
- 0% of the class have attended my consultation hour

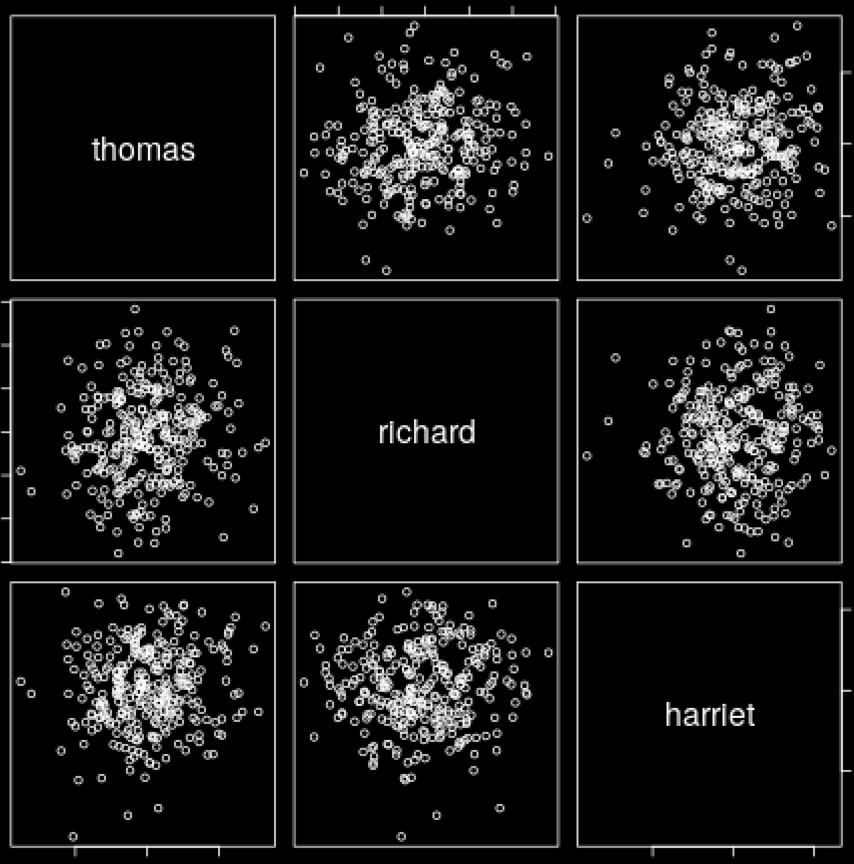
TODAY'S LECTURE

- correlations and correlation matrices
- simulating bivariate data
- relationship between correlation and regression

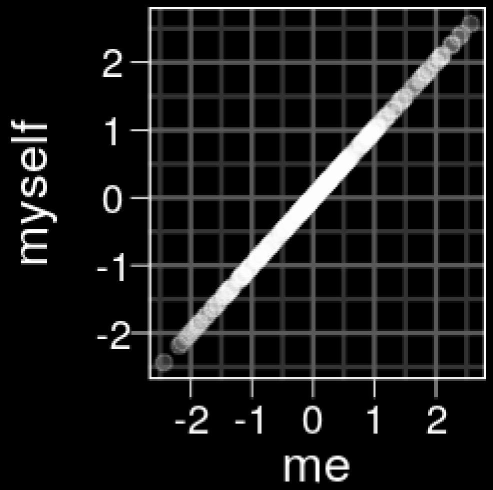
RELATIONSHIPS



MULTIPLE RELATIONSHIPS



THE PERFECT RELATIONSHIP



THE CORRELATION COEFFICIENT

Typicaly denoted as ρ (Greek symbol 'rho') or r

$$-1 \geq r \leq 1$$

- $r > 0$: positive relationship
- $r < 0$: negative relationship
- $r = 0$: no relationship

Estimated using Pearson or Spearman (rank) method.

In R: `cor()`, `cor.test()`

ASSUMPTIONS

- relationship between X and Y is **linear**
- deviations from line of best fit are **normally distributed**

MULTIPLE CORRELATIONS

For n variables, you have

$$\frac{n!}{2(n-2)!}$$

unique pairwise relationships, where $n!$ is the **factorial** of n.

In R: `choose(n, 2)`.

CORRELATION MATRICES

	IQ	verbal fluency	digit span
IQ	1.00	0.56	0.43
verbal fluency	0.56	1.00	-0.23
digit span	0.43	-0.23	1.00

In R: `corrr::correlate()`

CORRELATION MATRICES

	IQ	verbal fluency	digit span
IQ			
verbal fluency	0.56		
digit span	0.43	-0.23	

SIMULATING CORRELATIONAL DATA

To simulate bivariate (or multivariate) data in R, use `MASS::mvrnorm()`.

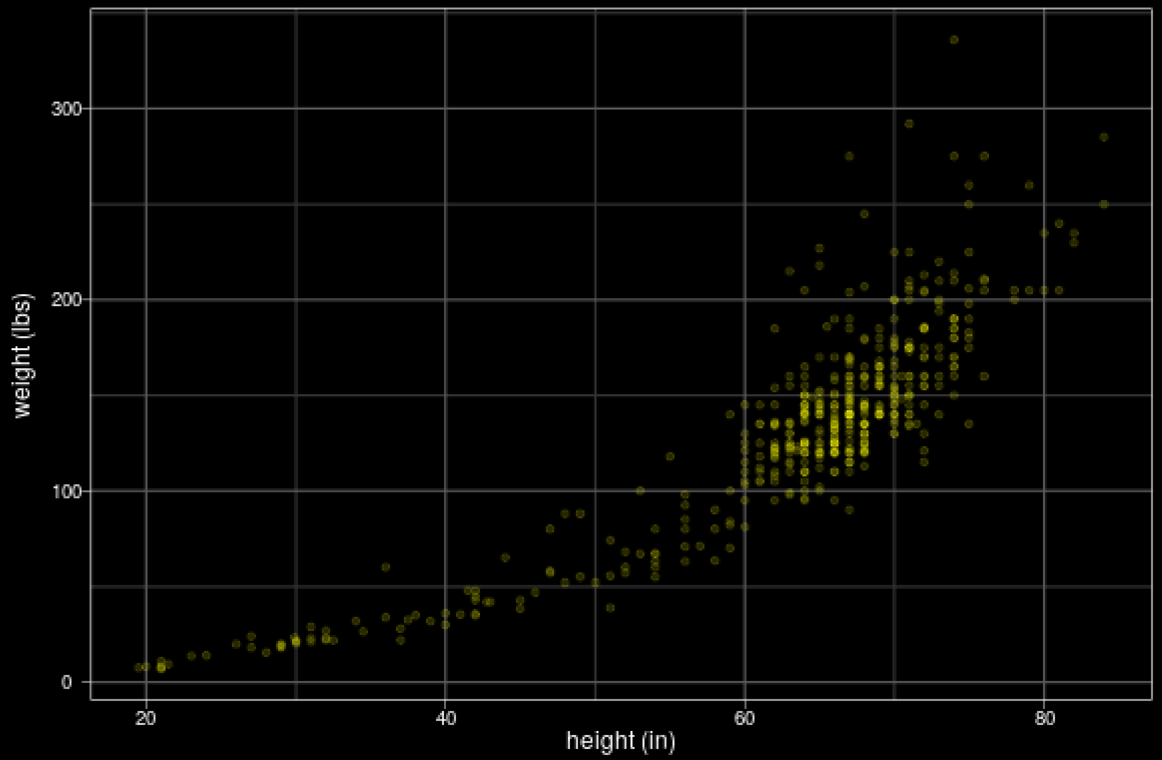
```
mvrnorm(n, mu, Sigma, ...)
```

You need the following information:

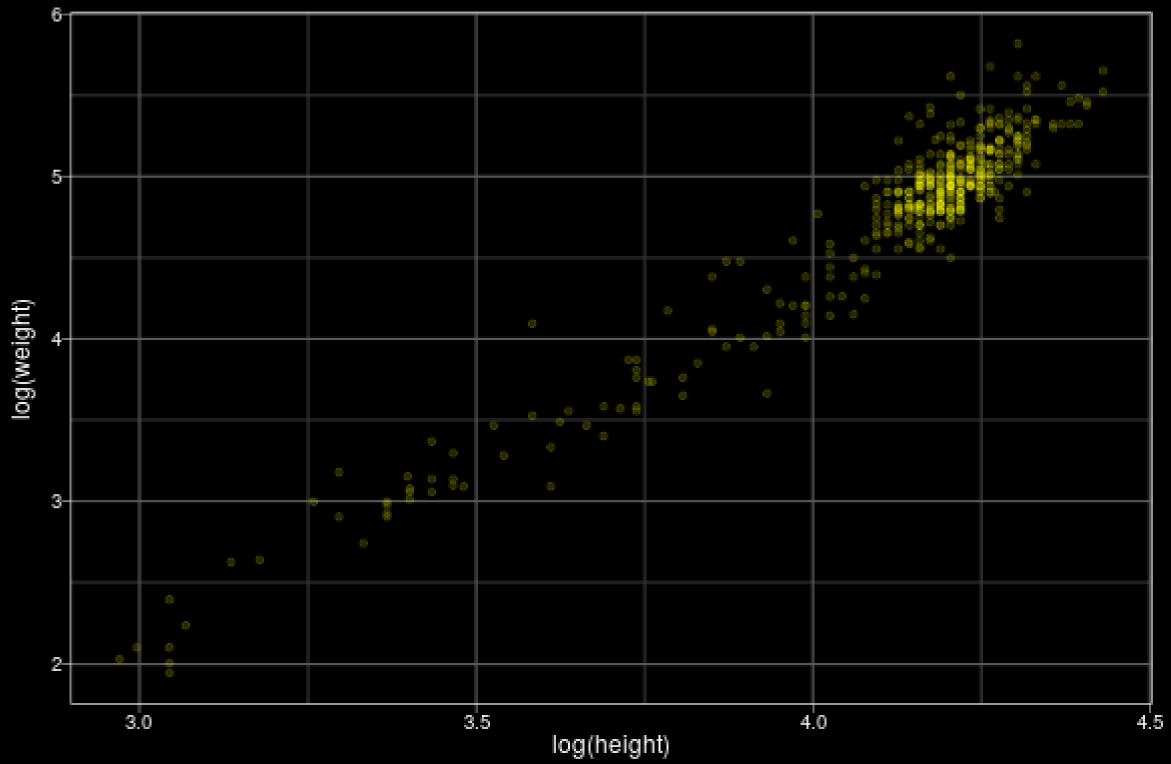
- means of X and Y, $\bar{X}$ and $\bar{Y}$
- standard deviations of X and Y, σ_X and σ_Y.
- correlation coefficient ρ_{XY}.

LET’S MAKE SYNTHETIC HUMANS

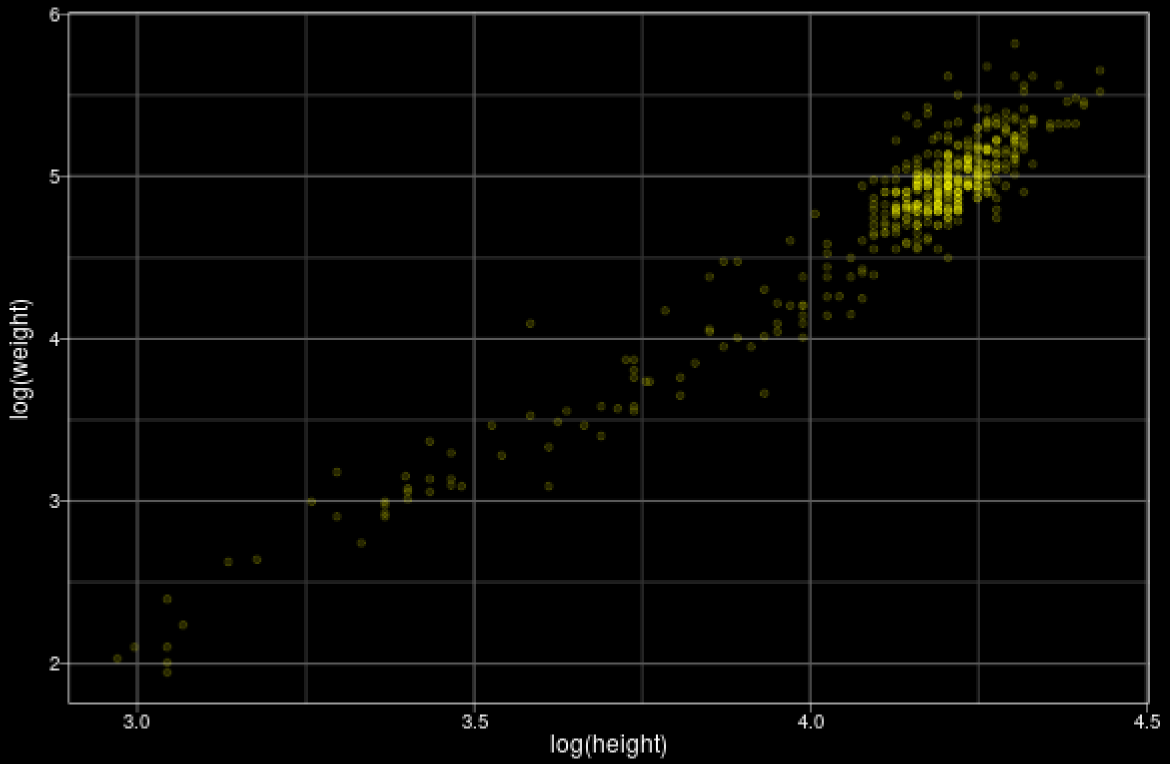
height and weight measurements for 435 people, taken from here



LOG-TRANSFORMED DATA



SUMMARY STATISTICS



$\bar{X}$	4.11
$\bar{Y}$	4.74
σ_X	.26
σ_Y	.65
ρ_{XY}	.96

COVARIANCE MATRIX

$$\boldsymbol{\Sigma}$$

A square matrix that characterizes the variances and their interrelationships (covariances).

$$\begin{pmatrix} {\sigma_x}^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{yx}\sigma_y\sigma_x & {\sigma_y}^2 \end{pmatrix}$$

Must be **symmetric** and **positive definite**

CALCULATIONS

$$\begin{pmatrix} {\sigma_x}^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{yx}\sigma_y\sigma_x & {\sigma_y}^2 \end{pmatrix}$$

σ_X	.26
σ_Y	.65
ρ_{XY}	.96

SIMULATING WITH MASS::mvrnorm()

```r
my_cov <- .96 * .26 * .65
my_Sigma <- matrix(c(.26^2, my_cov,
                     my_cov, .65^2),
                   ncol = 2)
my_Sigma
```

```
        [,1]    [,2]
[1,] 0.06760 0.16224
[2,] 0.16224 0.42250
```

```r
set.seed(62) # so we get the same numbers

## DON'T put library(MASS) in your script!
newpeeps <-
  MASS::mvrnorm(6, mu = c(height = 4.11,
                          weight = 4.74),
                Sigma = my_Sigma)

newpeeps
```

```
       height   weight
[1,] 4.254209 5.282913
[2,] 4.257828 4.895222
[3,] 3.722376 3.759767
[4,] 4.191287 4.764229
[5,] 4.739967 6.185191
[6,] 4.058105 4.806485
```

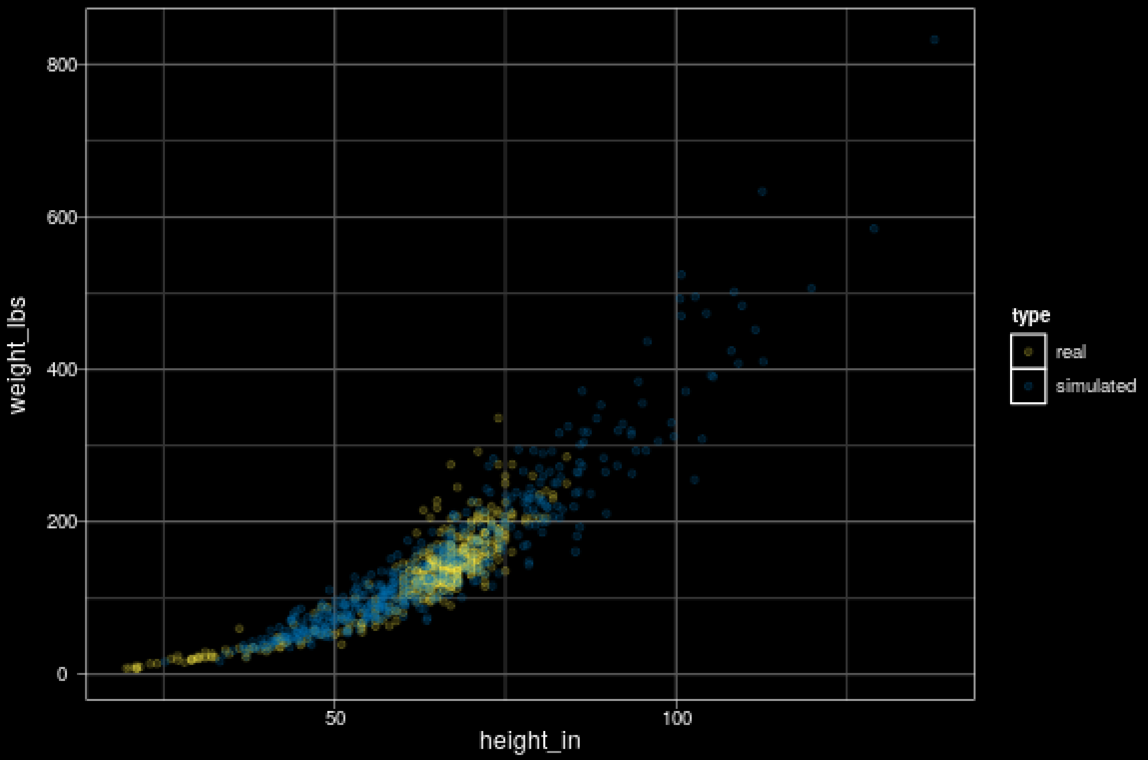
TRANSFORM BACK TO RAW UNITS

The `exp()` function is the inverse of `log()`.

```
exp(newpeeps)
```

```
        height    weight
[1,]  70.40108 196.94276
[2,]  70.65632 133.64963
[3,]  41.36254  42.93844
[4,]  66.10779 117.24065
[5,] 114.43045 485.50576
[6,]  57.86453 122.30092
```

OUR SYNTHETIC HUMANS



THE `bivariate` APP

http://shiny.psy.gla.ac.uk/Dale/bivariate

CORRELATION AND THE GLM

$$Y_i = \beta_0 + \beta_1 + e_i$$

$$e_i \sim N\left(0, \sigma^2\right)$$

$$\beta_1 = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$