# STATS 3001 / STATS 4104 / STATS 7054
# Statistical Modelling III
# Workshop 8 - linear mixed effects I

John Maclean

You've learned how to fit some variance structures in the framework of `gls`. We'll now use a similar approach to model *nested data*.

## Case Study

Your data is `RIKZ.txt` in the Zuur Dataset. It consists of measurements taken at various beaches along the Dutch coastline:

- `Richness`, measures of the abundance of biological species in each measurement
- `Exposure`, a measure of how exposed the beach surface is to waves and wind
- `NAP`, the height of the sample compared to mean tidal level
- `Beach`, a factor labelling the source beach.

The question of interest is the relationship between `Richness` and the predictors `Exposure` and `NAP`.

## Linear Mixed Effects Models

Anticipate discussion here: how do we model the fact that some of our data comes from the same beach, and some from different beaches? I don't really care about any one beach - I want to know a relationship between my predictors and response variables that holds across all beaches.

Can I use a standard linear model? What are the downsides?

The LME model is

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i \ .$$

Parts of this are standard:

- $Y_i$ is response variable
- $X_i$ is an $n \times p$ design matrix and $\beta$ is the vector of coefficients
- $\epsilon_i$ is iid Gaussian noise

New parts:

- $Z_i$ is a second design matrix of size $n \times q$
- $b_i$ is a second random noise variable, $b_i \sim \mathcal{N}(0, D)$, to be estimated.

## Workshop Questions

1. Does the data have any missing values? What values does exposure take?
2. Load the `nlme` package.

3. You want to use the `lme` function (linear mixed effects). Use this function like `lm` but add one new argument: `random`. This argument sets $Z$, the design matrix for the random effects. You have already used something similar in `gls`.
4. Try fitting a *random intercept model* where the random effect is `random = ~1 | beach` (beach a factor). Plot the model predictions and hence interpret the summary.
5. Try fitting a *random intercept and slope model* with `random = ~ 1 + NAP | beach`. Again - how do you view model predictions?
6. Which part of the model answers our question of interest? What is the role of the other part?