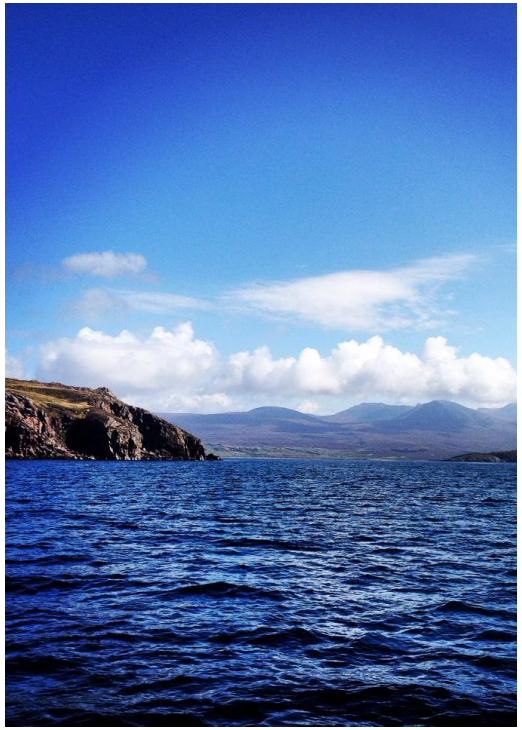


Demystifying statistics & programming in R

Dr Neil Burns – neil.burns@glasgow.ac.uk

How the session will work

Section headings in R will be highlighted like:

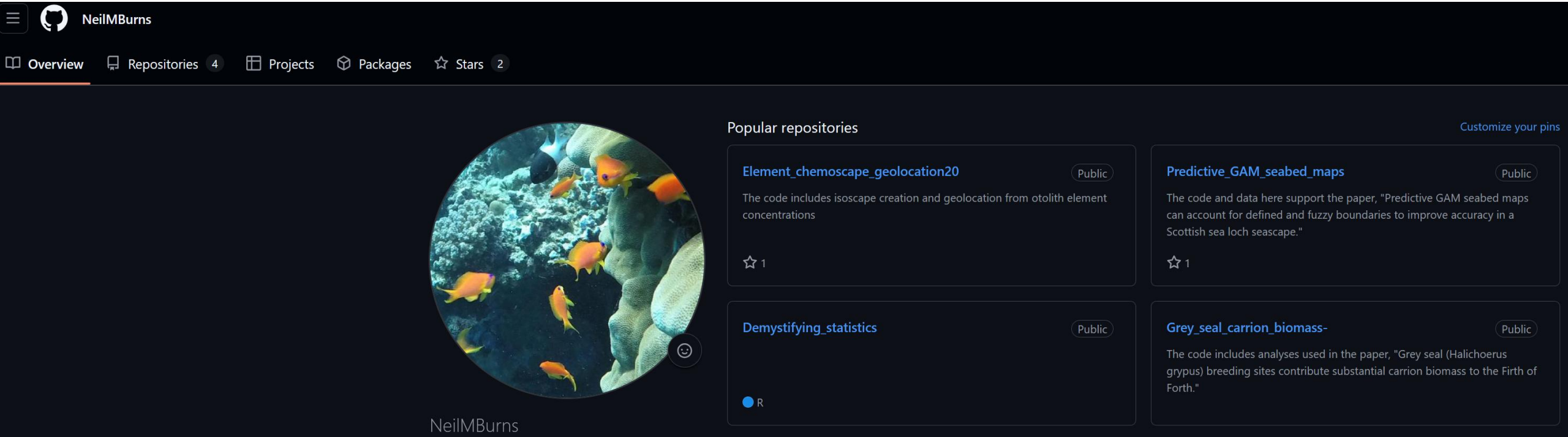
```
##########################################
#####        1.  some sort of title        #####
##########################################
```

Mix of presentation & R code

Ways to engage:

- Run the code alongside me. This may distract you from the presentation so only do this if you are comfortable. https://github.com/NeilMBurns

- Listen to the presentation and try the coding examples later at your own pace. This is what I recommend. That way, technical difficulties won't get in your way



How the session will work

Main course page

Fundamental scientific and data analysis resources

College of Social Sciences

School of Social and Environmental Sustainability

Fundamental scientific and data analysis resources

Distance based methods for analysing community composition using vegan

Demystifying statistics course materials

```
#############################################################
#         Demystifying statistics & programming in R        #
#                  Neil Burns © 2024 - 25                   #
#############################################################
```

Installing R, R-studio and code setup

```
#############################################################
#####            1. Setting up your script            #####
#############################################################
```

R install and code set up video

When you pop open the video to full screen or enlarge it the quality may take a couple of seconds to improve.

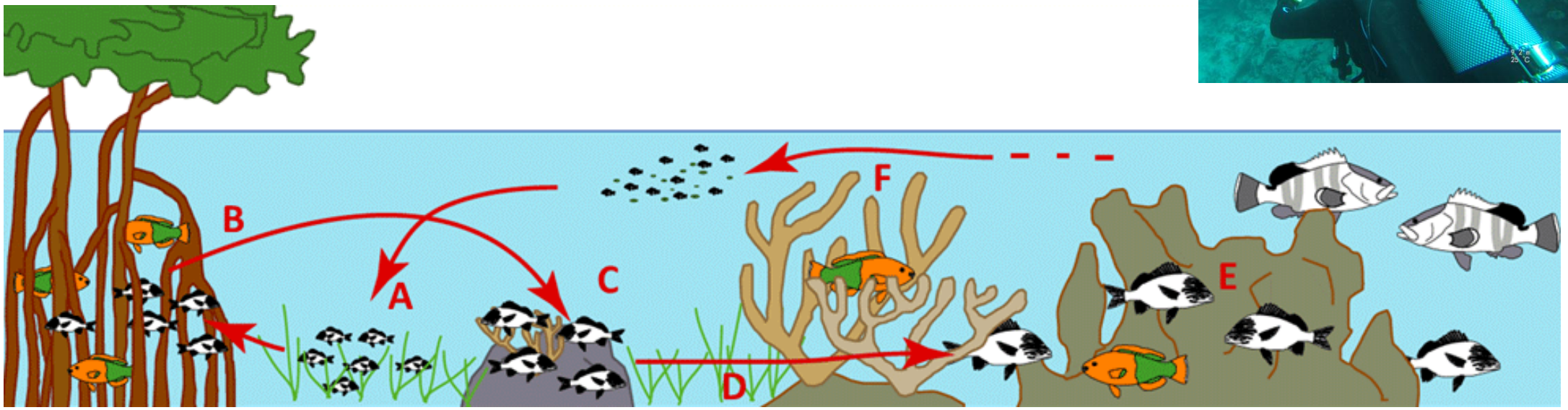
Contents

1. Introduction and rationale
2. Proposed example workflow and Introductory example
3. Don’t stare into the abyss – model interpretation
4. Frequentist, mixed effects models or and Bayesian approaches

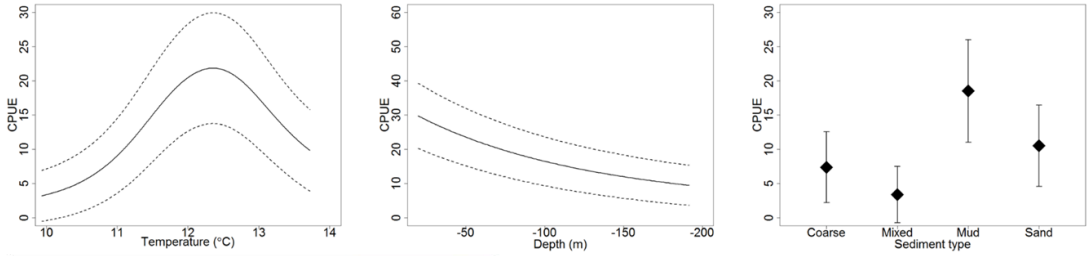
Research Interests

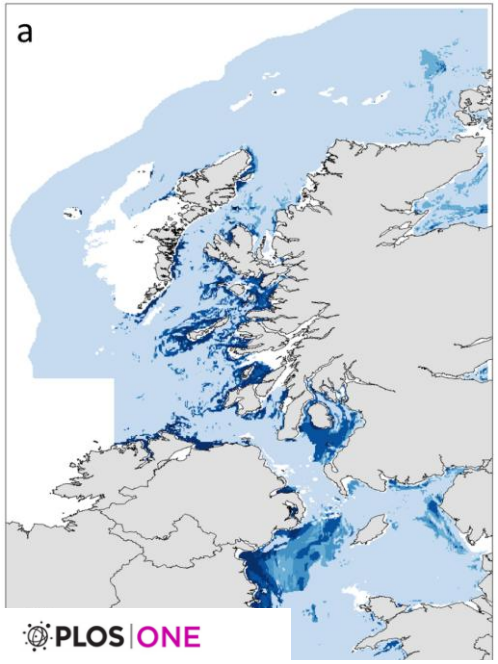
Population ecology and ecosystem health

- Develop methods to understand:
  - Demographic distributions of marine species
  - Spatial and temporal life stage connectivity
  - Relationship between connectivity and ecosystem health



Research background - ecology



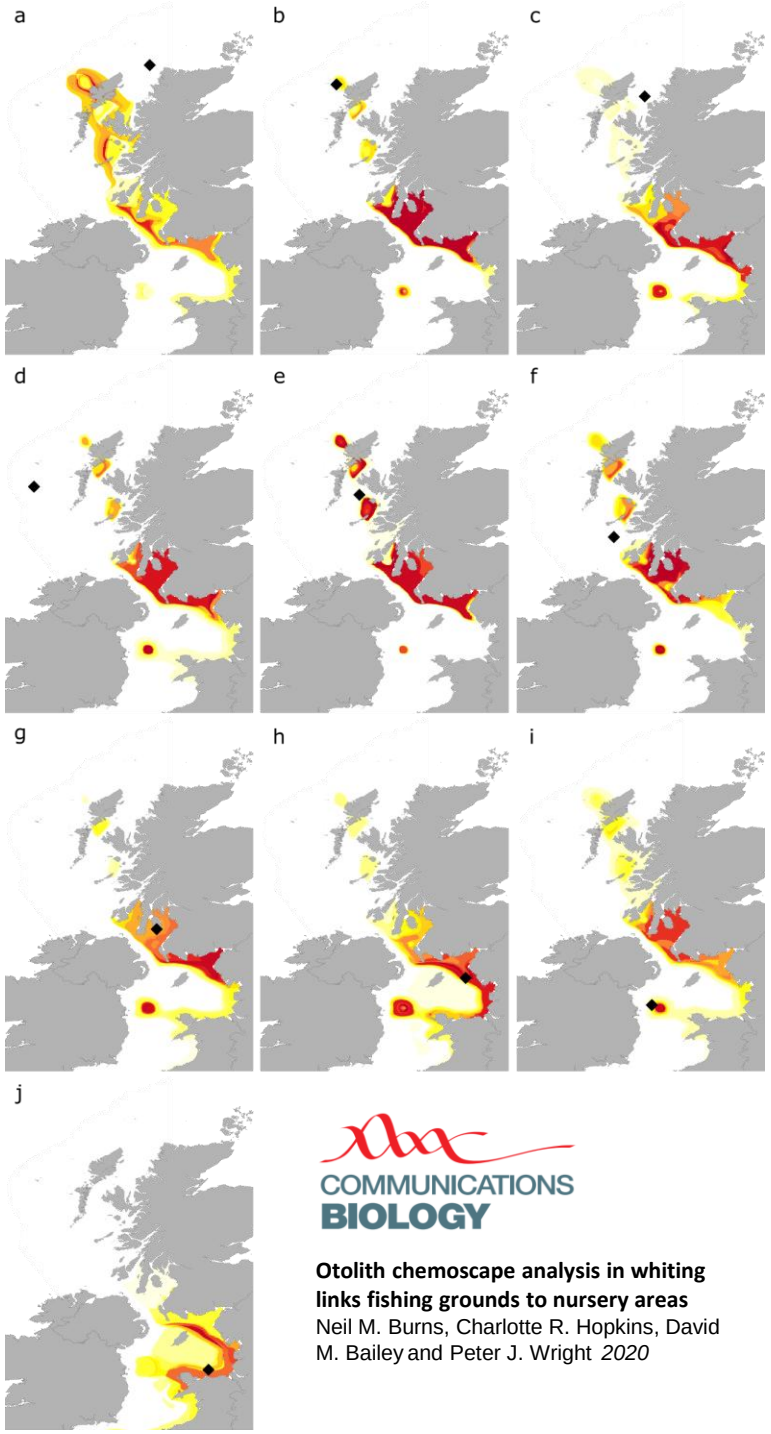


PLOS ONE

RESEARCH ARTICLE

A method to improve fishing selectivity through age targeted fishing using life stage distribution modelling

Neil M. Burns[1,2]*, David M. Bailey[1], Peter J. Wright[2]



COMMUNICATIONS BIOLOGY

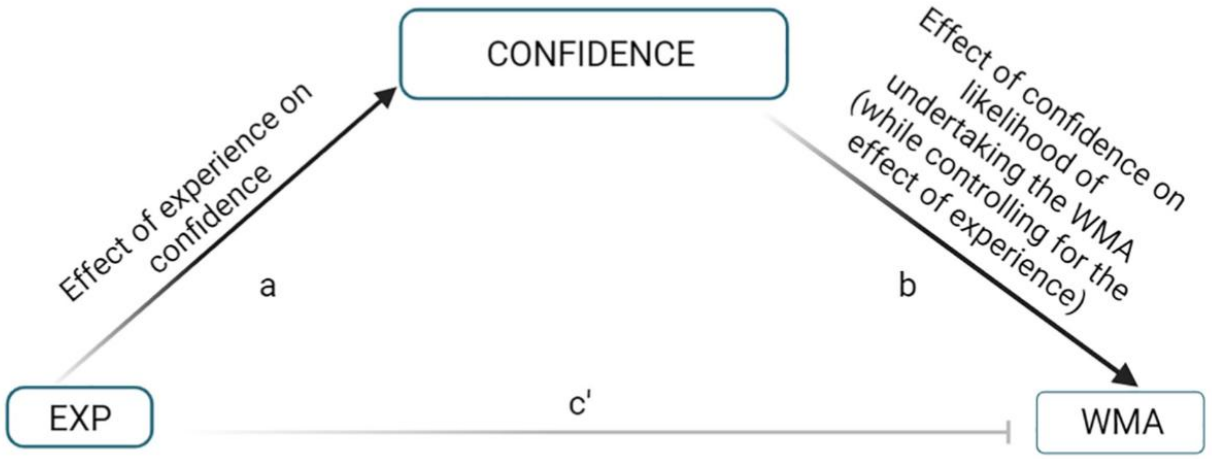
Otolith chemoscape analysis in whiting links fishing grounds to nursery areas
Neil M. Burns, Charlotte R. Hopkins, David M. Bailey and Peter J. Wright *2020*

Research background – other analysis applications

RESEARCH ARTICLE **PLOS ONE**

Confidence does not mediate a relationship between owner experience and likelihood of using weight management approaches for native ponies

Ashley B. Ward[1,2]*, Patricia A. Harris[3], Caroline McG. Argo[1], Christine A. Watson[1], Neil M. Burns[4], Madalina Neacsu[2], Wendy R. Russell[2], Dai Grove-White[5], Philippa K. Morrison[1]



Directional Acyclic Graph (DAG) demonstrating mediation model structures.

ELSEVIER

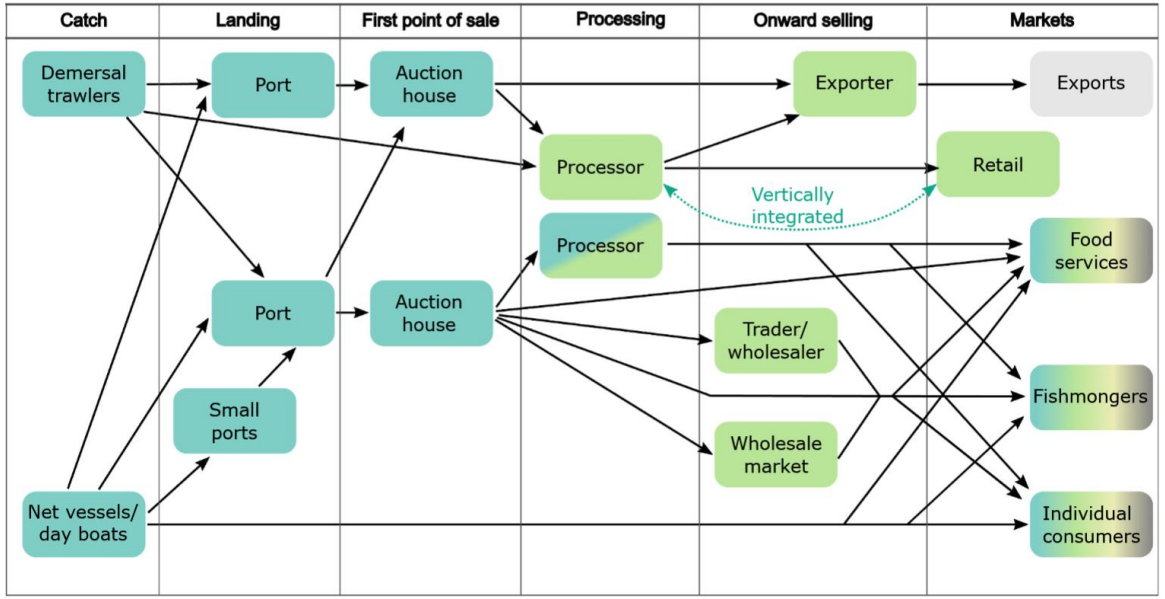
Contents lists available at ScienceDirect

Marine Policy

journal homepage: www.elsevier.com/locate/marpol

Improved traceability in seafood supply chains is achievable by minimising vulnerable nodes in processing and distribution networks

C.R. Hopkins[a,*], S.I. Roberts[b], A.J. Caveen[a], C. Graham[c], N.M. Burns[b]



Dover sole supply chain showing the first point of traceability loss during processing

Resources

https://cran.r-project.org/

CRAN
Mirrors
What's new?
Task Views
Search

About R

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them.

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources

R-4.0.3 for Windows (32/64 bit)

Download R 4.0.3 for Windows (85 megabytes, 32/64 bit)

Installation and other instructions
New features in this version

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the md5sum of the .exe to the fingerprint on the master server. You will need a version of md5sum for windows: both graphical and command line versions are available.

Frequently asked questions

- Does R run under my version of Windows?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

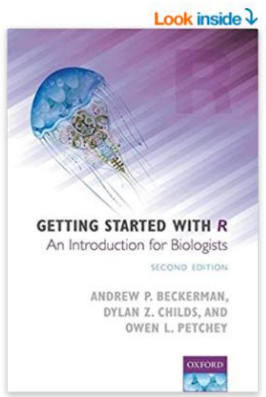
Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.

Resources

posit PRODUCTS SOLUTIONS LEARN & SUPPORT EXPLORE MORE DOWNLOAD RSTUDIO

DOWNLOAD

Download RStudio Desktop



Getting Started with R: An Introduction for Biologists Paperback – 26 Mar. 2017

by Andrew P. Beckerman (Author), Dylan Z. Childs (Contributor), Owen L. Petchey (Contributor)

4.4 48 ratings

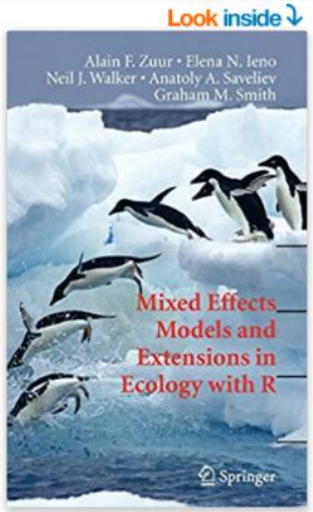
See all formats and editions

Kindle Edition	Hardcover	Paperback
£14.62	£50.99 prime	£22.99 prime
Read with Our Free App	5 New from £49.99	8 Used from £18.17 23 New from £18.79

FREE delivery: **Saturday, Dec 5**
Order within 10 hrs Details

Note: This item is eligible for **click and collect.** Details

R is rapidly becoming the standard software for statistical analyses, graphical presentation of data, and programming in the natural, physical, social, and engineering sciences. Getting Started with R is now the go-to introductory guide for biologists wanting to learn how to use R in their research. It teaches readers



Mixed Effects Models and Extensions in Ecology with R

Hardcover – 12 Mar. 2009

by Alain Zuur (Author), Elena N. Ieno (Author), Neil Walker (Author)

4.4 37 ratings

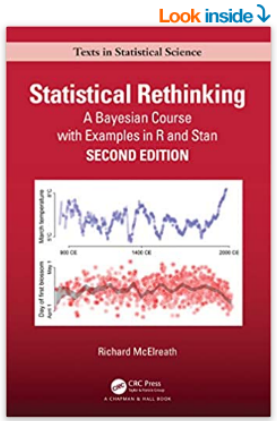
See all formats and editions

Kindle Edition	Hardcover	Paperback
£69.98	£73.66 prime	£109.99 prime
Read with Our Free App	3 Used from £90.51 11 New from £69.19	2 Used from £123.17 11 New from £91.14

FREE delivery: **Thursday, Dec 3** Details

Note: This item is eligible for **click and collect.** Details

This book discusses advanced statistical methods that can be used to analy
environmental collected data are measured repeatedly over time, or space a



Statistical Rethinking: A Bayesian Course with Examples in R and STAN (Chapman & Hall/CRC Texts in Statistical Science)

Hardcover – 16 Mar. 2020

by Richard McElreath (Author)

4.8 288 ratings

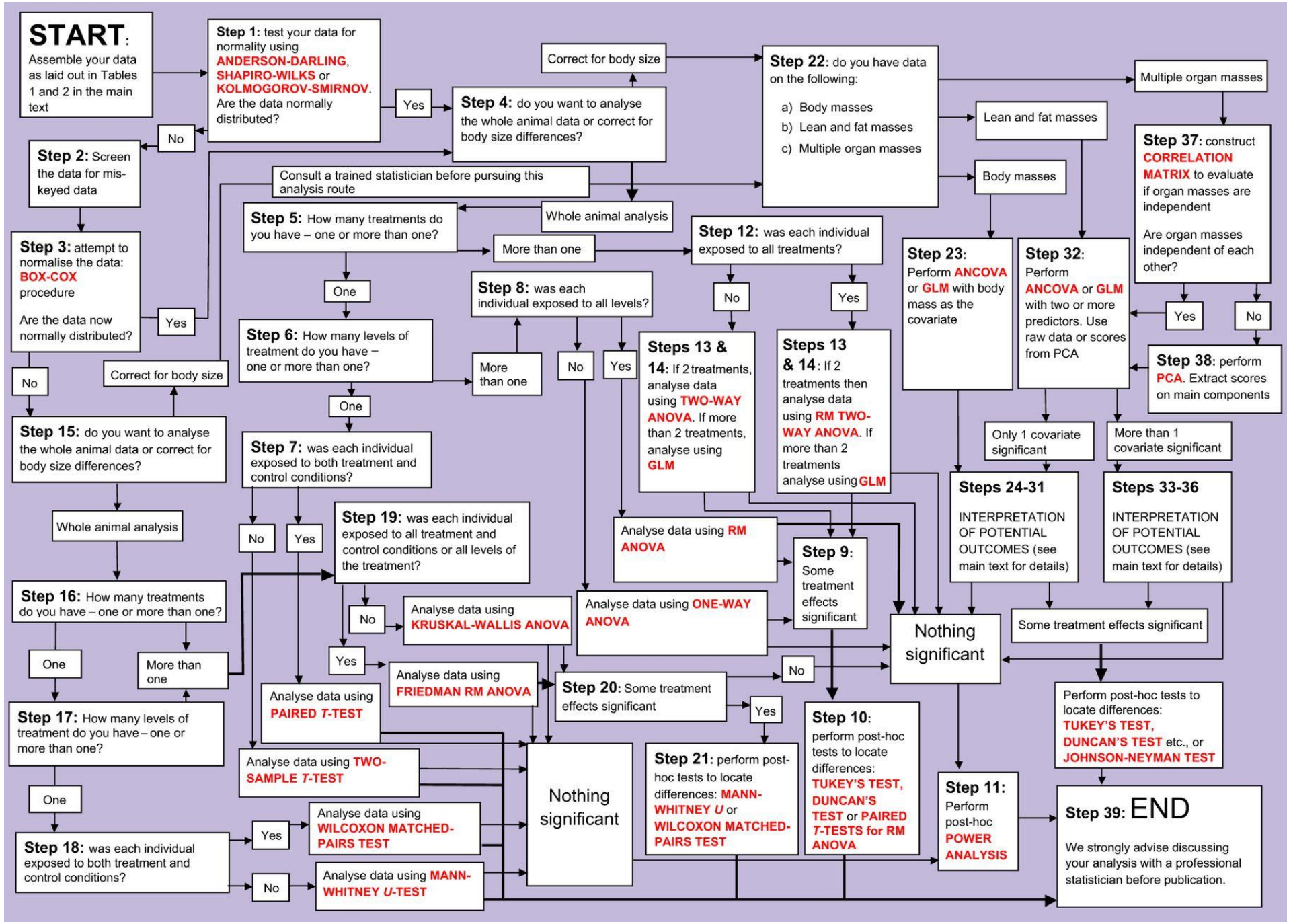
Part of: Chapman & Hall/CRC Texts in Statistical Science (129 books)

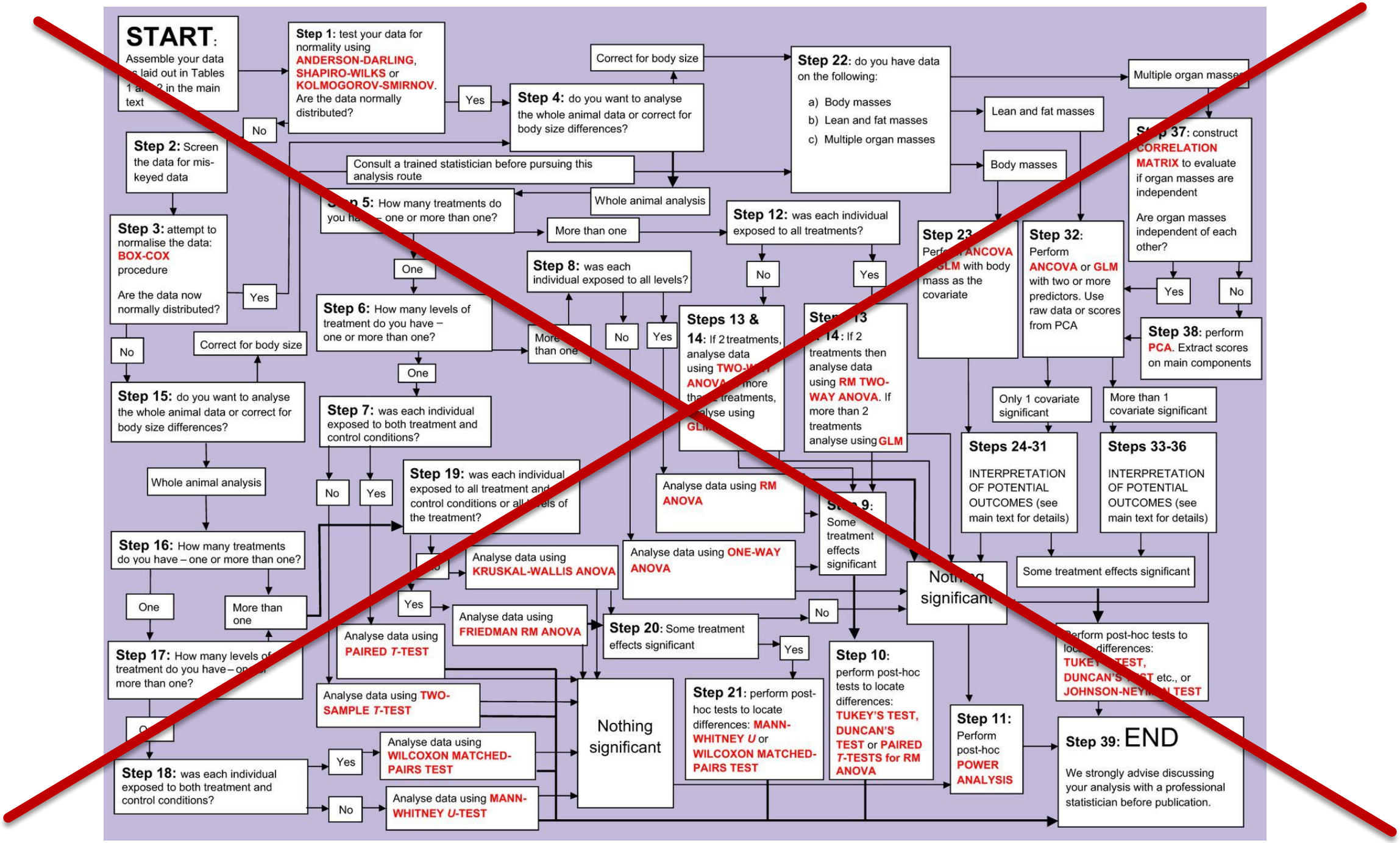
See all formats and editions

Kindle Edition	Hardcover
£7.38 incl. VAT	£66.69 incl. VAT prime
	5 Used from £58.99 23 New from £66.69

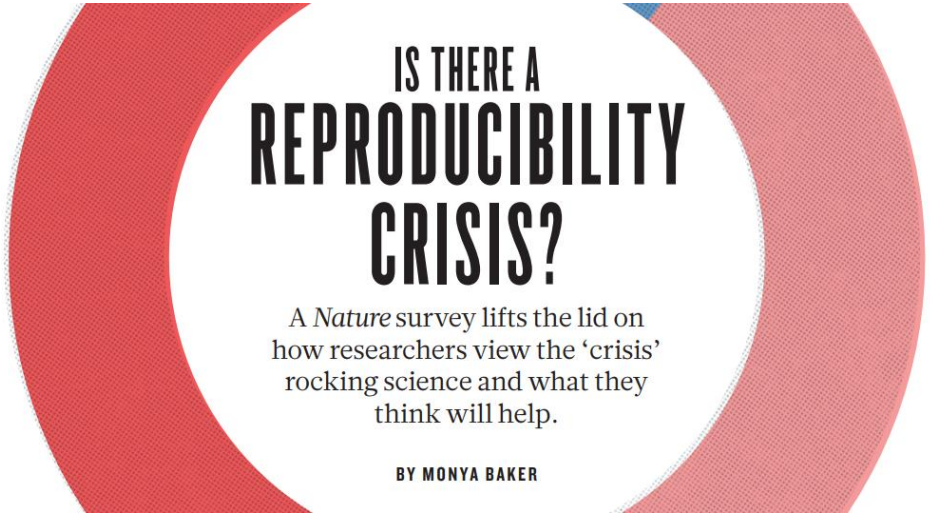
Note: This item is eligible for **FREE Click and Collect** without a minimum order subject to availability. Details

Statistical Rethinking: A Bayesian Course with Examples in R and Stan builds your knowledge of and confidence in making inferences from data. Reflecting the need for scripting in today's model-based statistics, the book pushes you to perform step-by-step calculations that are


START: Assemble your data as laid out in Tables 1 and 2 in the main text
Step 1: test your data for normality using ANDERSON-DARLING, SHAPIRO-WILKS or KOLMOGOROV-SMIRNOV. Are the data normally distributed?
Yes
No
Step 2: Screen the data for mis-keyed data
Step 3: attempt to normalise the data: BOX-COX procedure
Are the data now normally distributed?
Yes
No
Correct for body size
Consult a trained statistician before pursuing this analysis route
Step 4: do you want to analyse the whole animal data or correct for body size differences?
Correct for body size
Whole animal analysis
Step 5: How many treatments do you have – one or more than one?
One
More than one
Step 6: How many levels of treatment do you have – one or more than one?
One
More than one
Step 7: was each individual exposed to both treatment and control conditions?
No
Yes
Step 8: was each individual exposed to all levels?
No
Yes
Step 12: was each individual exposed to all treatments?
No
Yes
Steps 13 & 14: If 2 treatments, analyse data using TWO-WAY ANOVA. If more than 2 treatments, analyse using GLM
Steps 13 & 14: If 2 treatments then analyse data using RM TWO-WAY ANOVA. If more than 2 treatments analyse using GLM
Step 22: do you have data on the following:
a) Body masses
b) Lean and fat masses
c) Multiple organ masses
Multiple organ masses
Lean and fat masses
Body masses
Step 37: construct CORRELATION MATRIX to evaluate if organ masses are independent
Are organ masses independent of each other?
Yes
No
Step 38: perform PCA. Extract scores on main components
Step 23: Perform ANCOVA or GLM with body mass as the covariate
Step 32: Perform ANCOVA or GLM with two or more predictors. Use raw data or scores from PCA
Only 1 covariate significant
More than 1 covariate significant
Steps 24-31 INTERPRETATION OF POTENTIAL OUTCOMES (see main text for details)
Steps 33-36 INTERPRETATION OF POTENTIAL OUTCOMES (see main text for details)
Step 15: do you want to analyse the whole animal data or correct for body size differences?
Whole animal analysis
Step 16: How many treatments do you have – one or more than one?
One
More than one
Step 17: How many levels of treatment do you have – one or more than one?
One
Step 18: was each individual exposed to both treatment and control conditions?
Yes
No
Step 19: was each individual exposed to all treatment and control conditions or all levels of the treatment?
No
Yes
Analyse data using RM ANOVA
Step 9: Some treatment effects significant
Analyse data using ONE-WAY ANOVA
Analyse data using KRUSKAL-WALLIS ANOVA
Analyse data using FRIEDMAN RM ANOVA
Step 20: Some treatment effects significant
No
Yes
Nothing significant
Some treatment effects significant
Analyse data using PAIRED T-TEST
Analyse data using TWO-SAMPLE T-TEST
Analyse data using WILCOXON MATCHED-PAIRS TEST
Analyse data using MANN-WHITNEY U-TEST
Nothing significant
Step 21: perform post-hoc tests to locate differences: MANN-WHITNEY U or WILCOXON MATCHED-PAIRS TEST
Step 10: perform post-hoc tests to locate differences: TUKEY'S TEST, DUNCAN'S TEST or PAIRED T-TESTS for RM ANOVA
Perform post-hoc tests to locate differences: TUKEY'S TEST, DUNCAN'S TEST etc., or JOHNSON-NEYMAN TEST
Step 11: Perform post-hoc POWER ANALYSIS
Step 39: END
We strongly advise discussing your analysis with a professional statistician before publication.


START:
Assemble your data as laid out in Tables 1 and 2 in the main text
Step 1: test your data for normality using ANDERSON-DARLING, SHAPIRO-WILKS or KOLMOGOROV-SMIRNOV. Are the data normally distributed?
Yes
No
Step 2: Screen the data for mis-keyed data
Step 3: attempt to normalise the data: BOX-COX procedure
Are the data now normally distributed?
Yes
No
Correct for body size
Step 4: do you want to analyse the whole animal data or correct for body size differences?
Correct for body size
Consult a trained statistician before pursuing this analysis route
Whole animal analysis
Step 5: How many treatments do you have – one or more than one?
One
More than one
Step 6: How many levels of treatment do you have – one or more than one?
One
More than one
Step 7: was each individual exposed to both treatment and control conditions?
No
Yes
Step 8: was each individual exposed to all levels?
No
Yes
Step 12: was each individual exposed to all treatments?
No
Yes
Step 15: do you want to analyse the whole animal data or correct for body size differences?
Whole animal analysis
Step 16: How many treatments do you have – one or more than one?
One
More than one
Step 17: How many levels of treatment do you have – one or more than one?
Step 18: was each individual exposed to both treatment and control conditions?
Yes
No
Step 19: was each individual exposed to all treatment and control conditions or all levels of the treatment?
Yes
Analyse data using KRUSKAL-WALLIS ANOVA
Analyse data using FRIEDMAN RM ANOVA
Analyse data using PAIRED T-TEST
Analyse data using TWO-SAMPLE T-TEST
Analyse data using WILCOXON MATCHED-PAIRS TEST
Analyse data using MANN-WHITNEY U-TEST
Step 20: Some treatment effects significant
Yes
No
Steps 13 & 14: If 2 treatments, analyse data using TWO-WAY ANOVA. If more than 2 treatments, analyse using GLM
Step 13 & 14: If 2 treatments then analyse data using RM TWO-WAY ANOVA. If more than 2 treatments analyse using GLM
Analyse data using RM ANOVA
Analyse data using ONE-WAY ANOVA
Step 9: Some treatment effects significant
Step 10: perform post-hoc tests to locate differences: TUKEY'S TEST, DUNCAN'S TEST or PAIRED T-TESTS for RM ANOVA
Step 21: perform post-hoc tests to locate differences: MANN-WHITNEY U or WILCOXON MATCHED-PAIRS TEST
Nothing significant
Step 22: do you have data on the following:
a) Body masses
b) Lean and fat masses
c) Multiple organ masses
Lean and fat masses
Body masses
Multiple organ masses
Step 23: Perform ANCOVA or GLM with body mass as the covariate
Step 32: Perform ANCOVA or GLM with two or more predictors. Use raw data or scores from PCA
Step 37: construct CORRELATION MATRIX to evaluate if organ masses are independent
Are organ masses independent of each other?
Yes
No
Step 38: perform PCA. Extract scores on main components
Only 1 covariate significant
More than 1 covariate significant
Steps 24-31
INTERPRETATION OF POTENTIAL OUTCOMES (see main text for details)
Steps 33-36
INTERPRETATION OF POTENTIAL OUTCOMES (see main text for details)
Nothing significant
Some treatment effects significant
Perform post-hoc tests to locate differences: TUKEY'S TEST, DUNCAN'S TEST etc., or JOHNSON-NEYMAN TEST
Step 11: Perform post-hoc POWER ANALYSIS
Step 39: END
We strongly advise discussing your analysis with a professional statistician before publication.

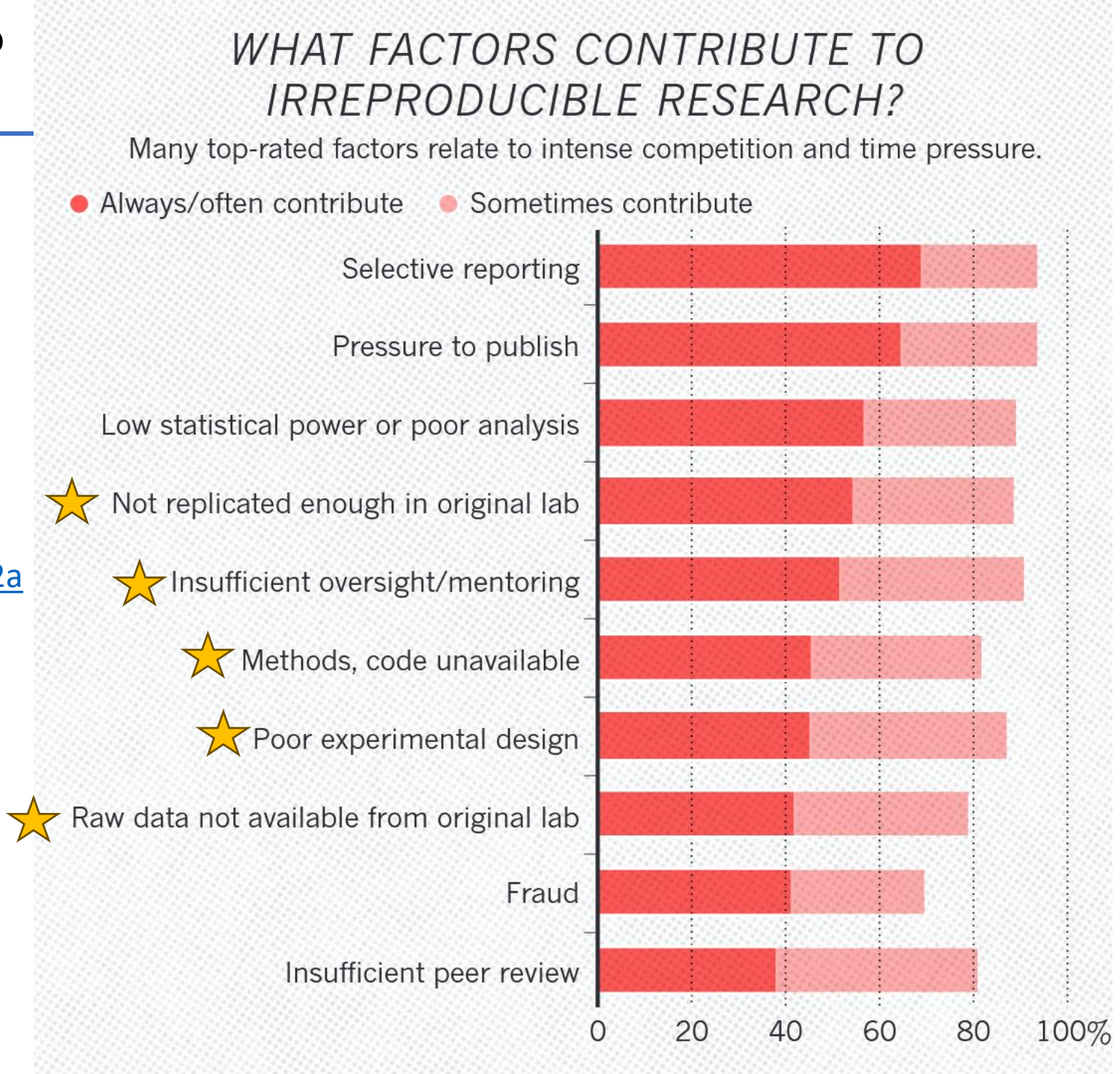
Why do we need to do better?



Nature 2016 https://www.nature.com/articles/533452a

https://stats.andrewheiss.com/hack-your-way/

Hack Your Way To Scientific Glory (RECREATION)



Why do we need to do better?

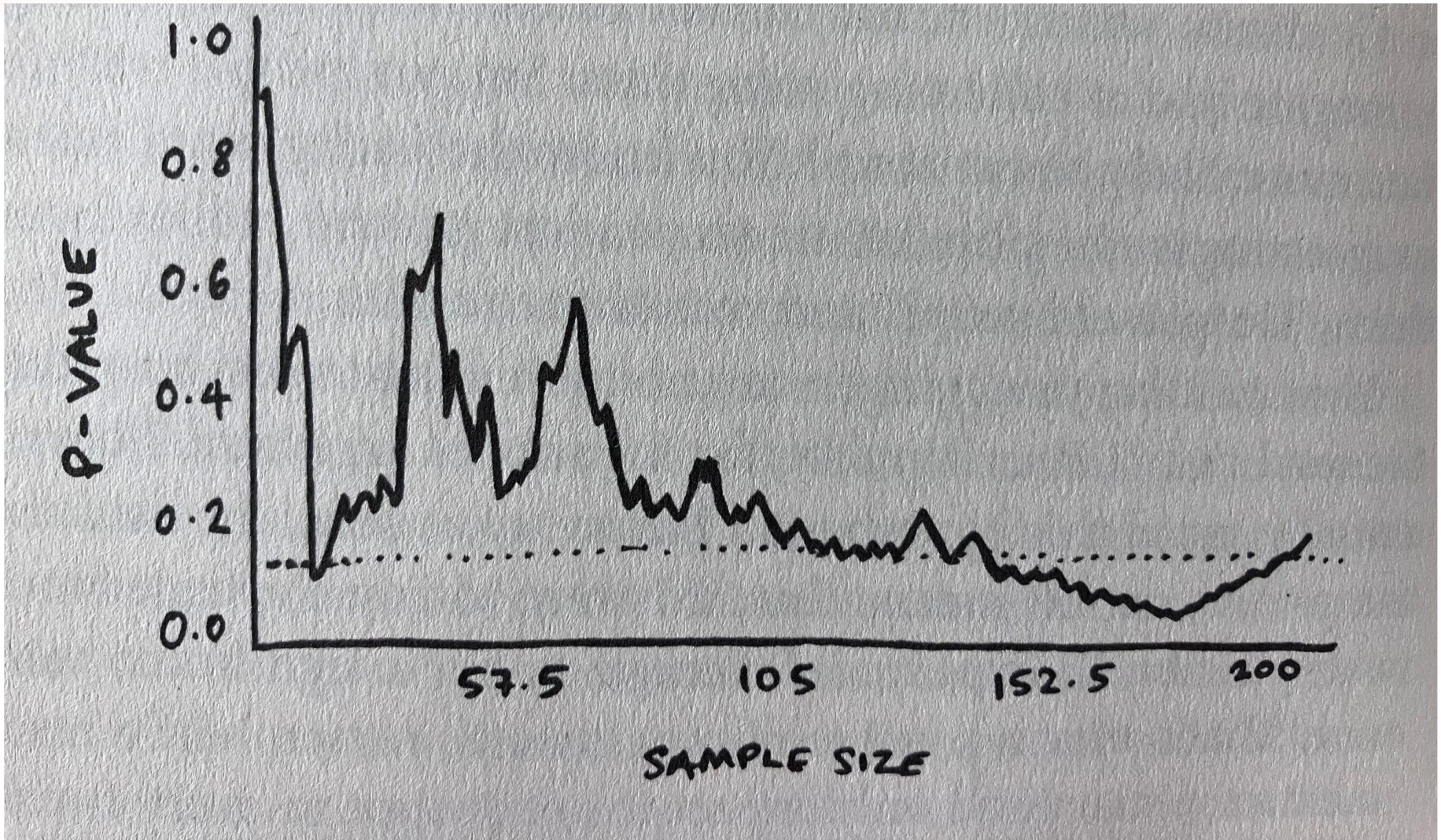
https://lakens.github.io/statistical_inferences/

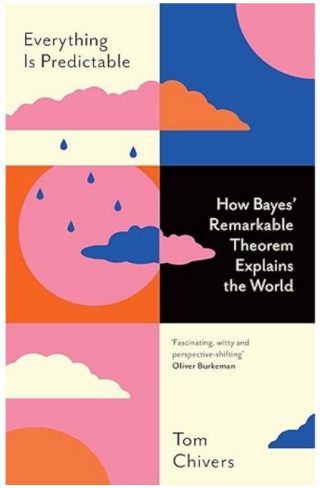
Improving Your Statistical Inferences

Improving Your Statistical Inferences

This open educational resource contains information to improve statistical inferences, design better experiments, and report scientific research more transparently.

AUTHOR
Daniël Lakens





Everything Is Predictable: How Bayes' Remarkable Theorem Explains the World Paperback – 25 April 2024
by Tom Chivers (Author)

What do we not teach you.

- Data organisation
- Data curation
- Procedures for test data*
- Managing contributions from multiple coders
- How to connect your hypothesis to data analysis logically

*

- You will make mistakes.
- Testing is not optional.
- Test your code!

What do we not teach you.

- How to connect your hypothesis to data analysis logically

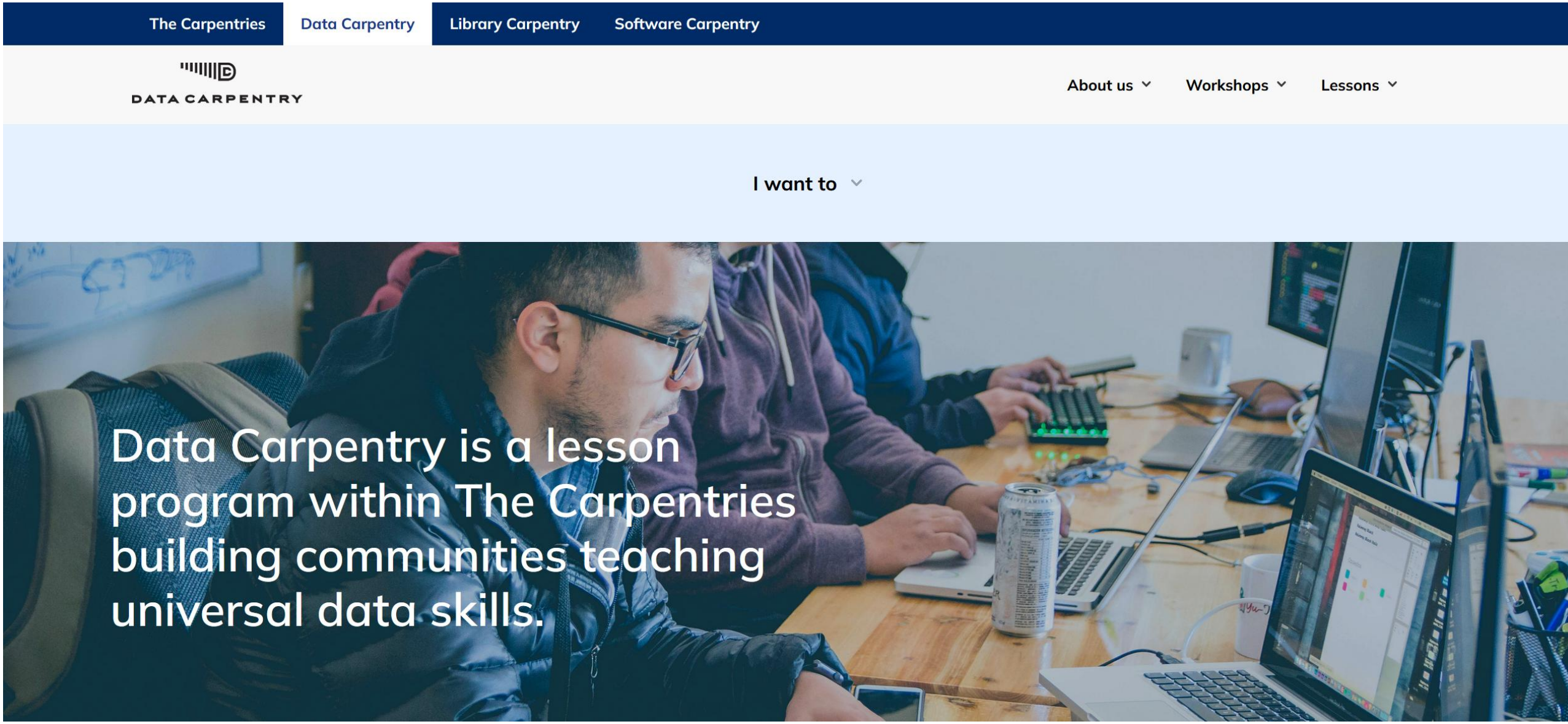
Ask yourself: Can my proposed analysis method, in principle, get the analysis correct first?

- Express your hypothesis
  - (diagrammatically or in code)
- Show that your planned analysis could work
- Test your analysis pipeline on synthetic data set.
  - Generative AI could help here
- Run the analysis on the gathered empirical data

Aim to make this all this process available for scrutiny

Modelling as a method helps to address a lot of the above.

What about the bits not covered here.



R script set up…oh yeah and making it look cool…although I might have forgotten to do this bit. Here's a link

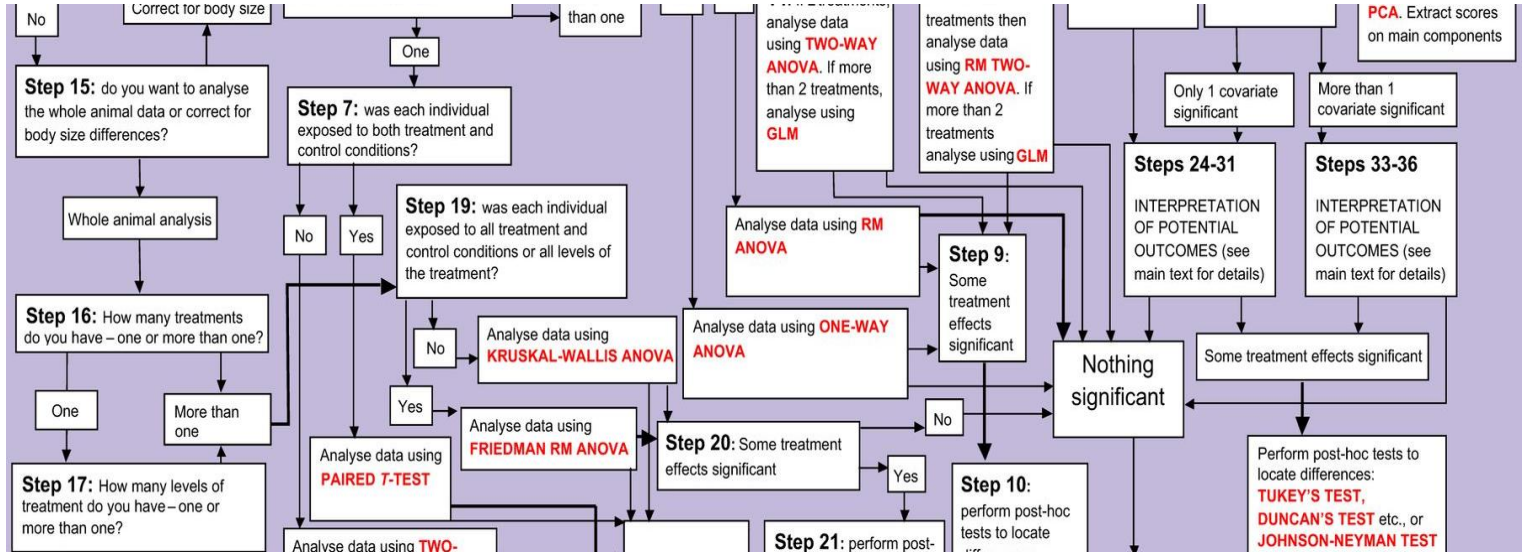
```
##########################################
#####        1. Setting up your script        #####
##########################################
```

Pause

They are just robots

Stained glass window Latin square, honouring [Ronald Fisher]

Why use someone else's robot?



Linear models (lm)
Generalised Linear Models (glm)
Generalised Additive Models (gam)
Mixed effects versions of (lmm, glmm, gamm)

Statistical modelling using glms

Why R?

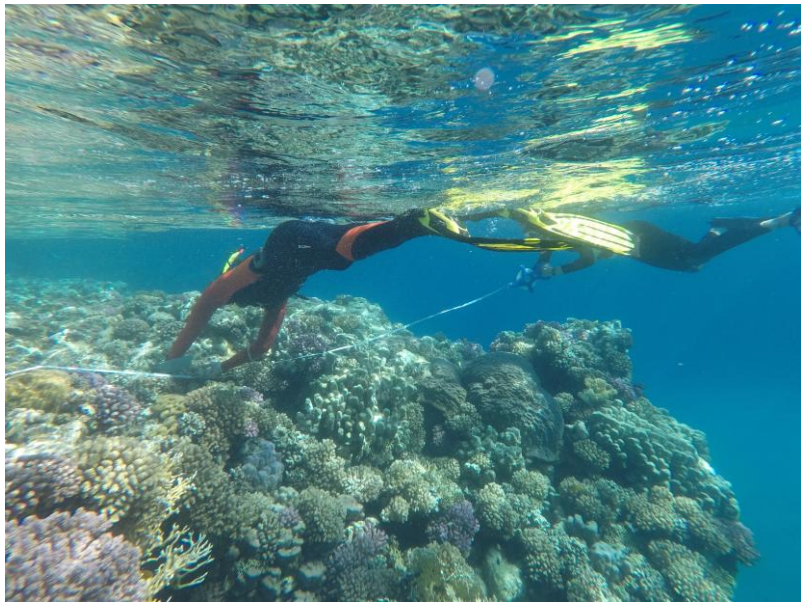
But…we are getting ahead of ourselves.

	A	B	C	D
1	Site	Coral_colour	Perc_cover	S_abund
2	1A	Blue	20.5	23
3	1A	Green	9.5	2
4	1A	Red	70	19
5	1B	Blue	50	18
6	1B	Green	25.5	22
7	1B	Red	25.5	6

- Data into R from a .csv file
- Each column is a variable (explanatory (x) and response (y))
- Keep the column names short and simple but meaningful (to you)
- No spaces in column names_use_as_separator
- Be consistent with capitals

The very 1st thing you need to do = UNDERSTAND YOUR DATA!!!!!

Our toy example



Coral percentage cover (x_1)

0 to 100%

Abundance of sharks (y_1) ~

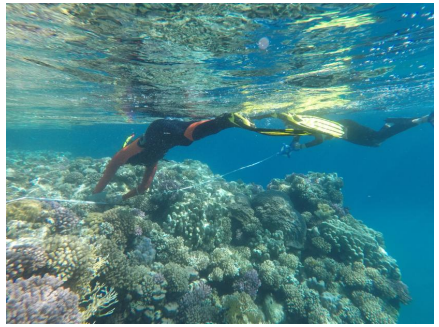
It's a whole number (<u>or maybe not</u>)
They are big and there are not loads of them so probably less than 60

Coral colour (x_2)

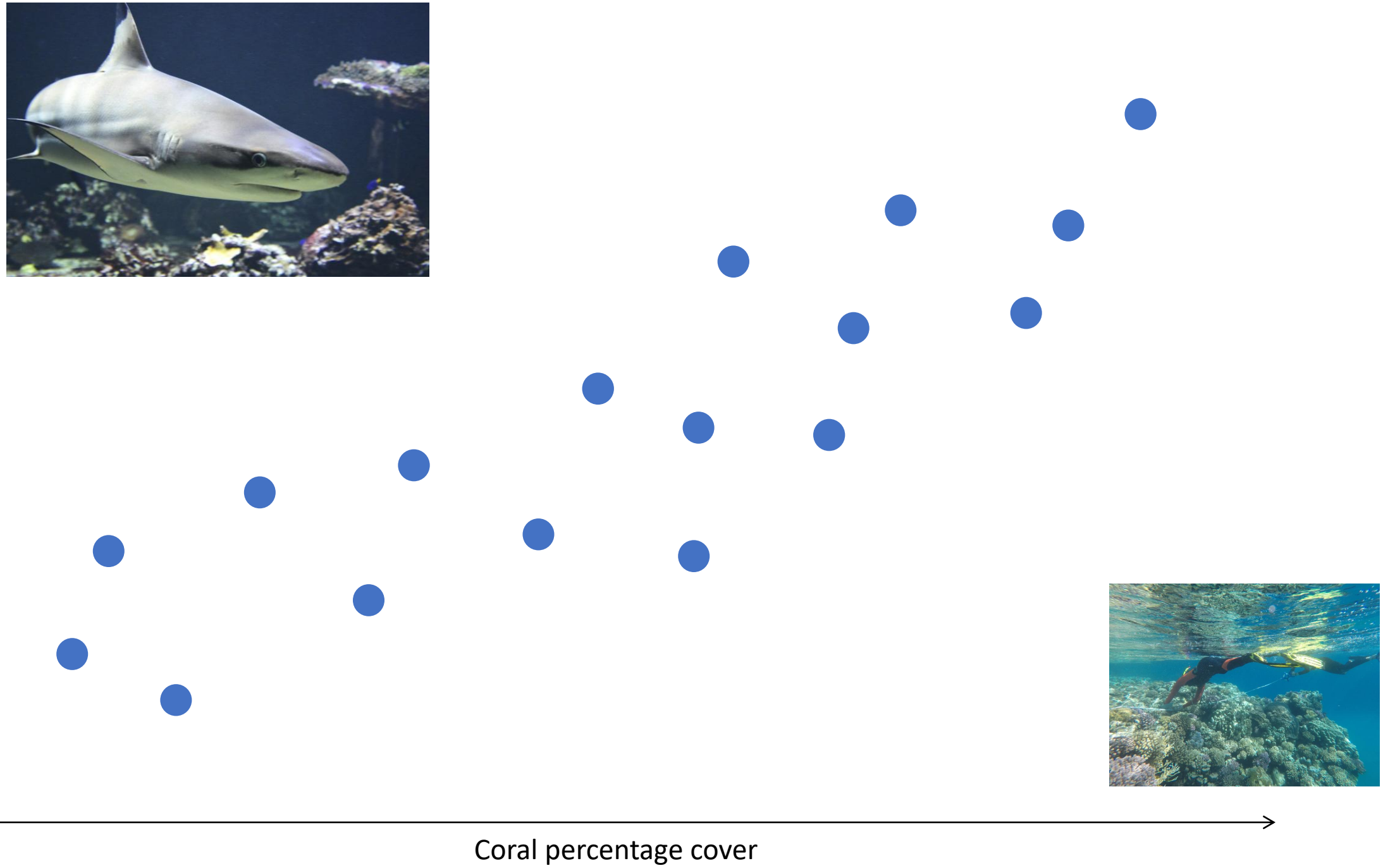
Blue, Brown & green

```
plot(count_shark~ perc_cov, data=dat,  xlim=c(0,100), ylim=c(0,30), pch=19, col='firebrick4',
     cex=2, xlab='Percentage cover', ylab='count_shark')
```

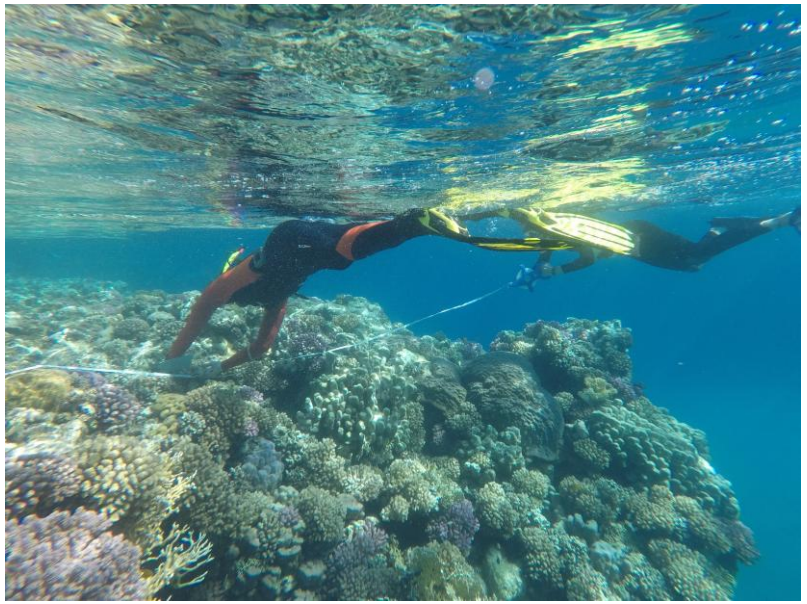
Shark abundance

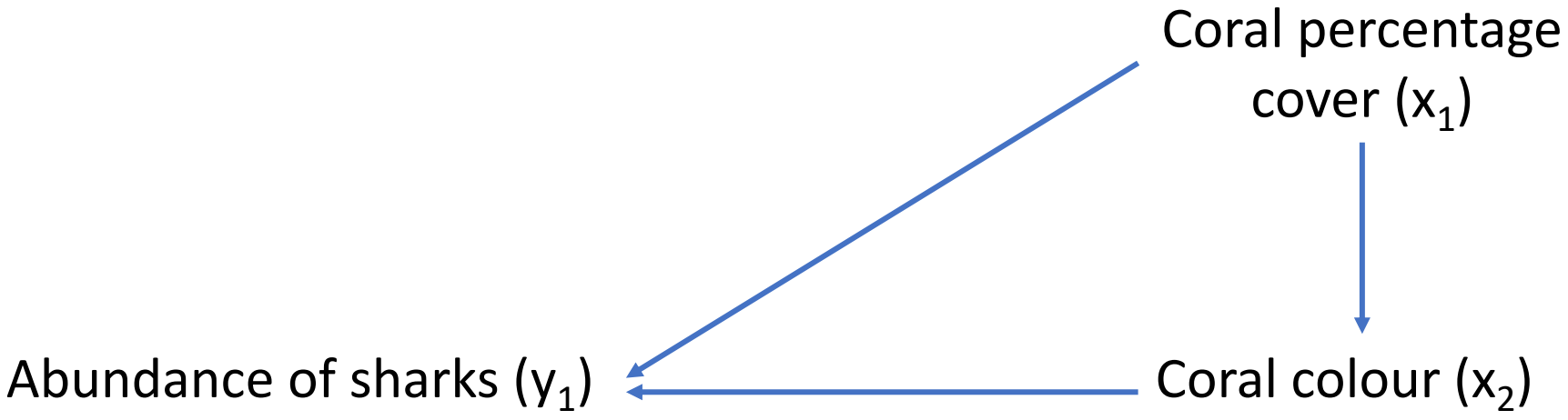


Coral percentage cover

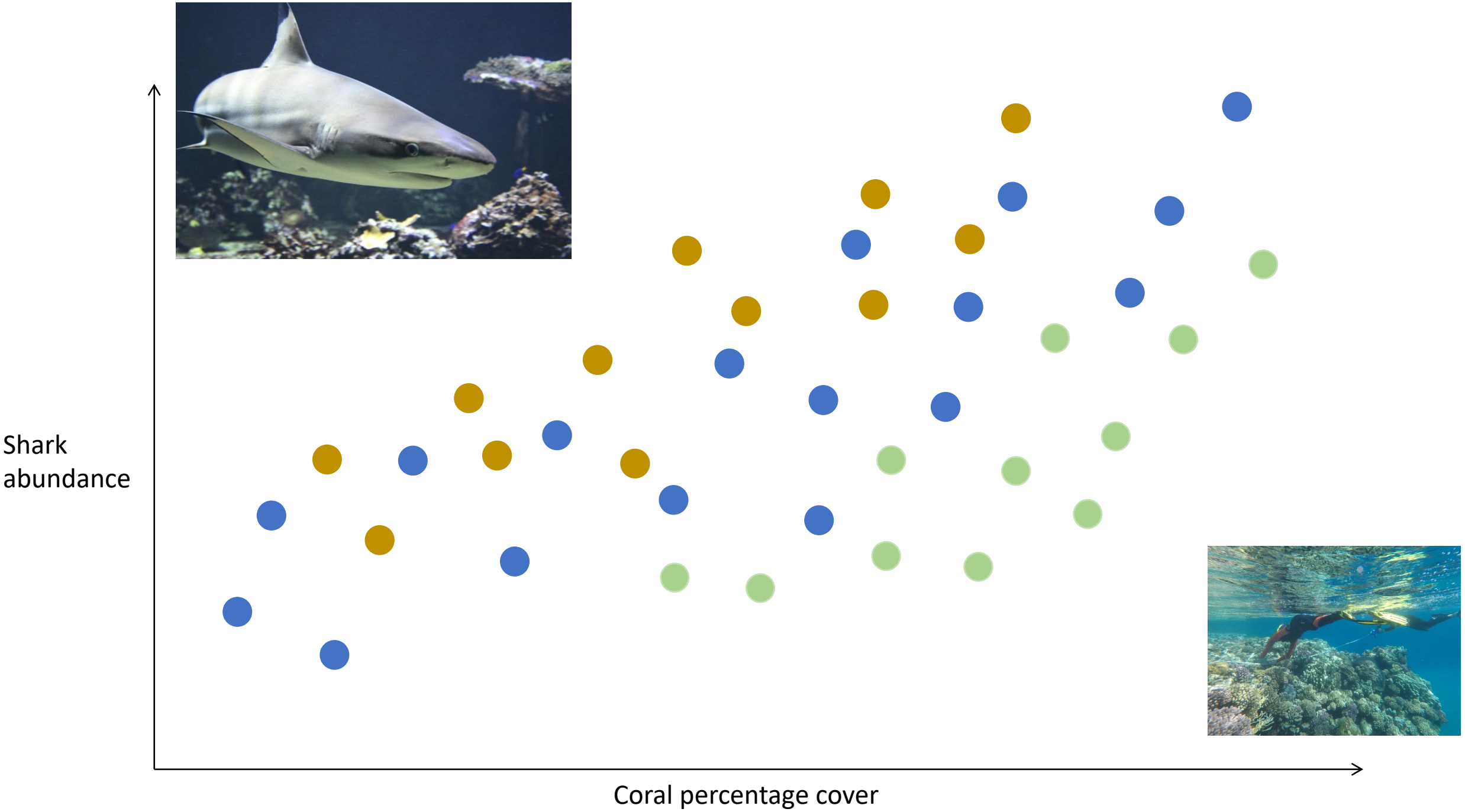

Shark abundance
Coral percentage cover

Express our hypothesis





Directional acyclic Graph (DAG)


Shark abundance
Coral percentage cover

R code section

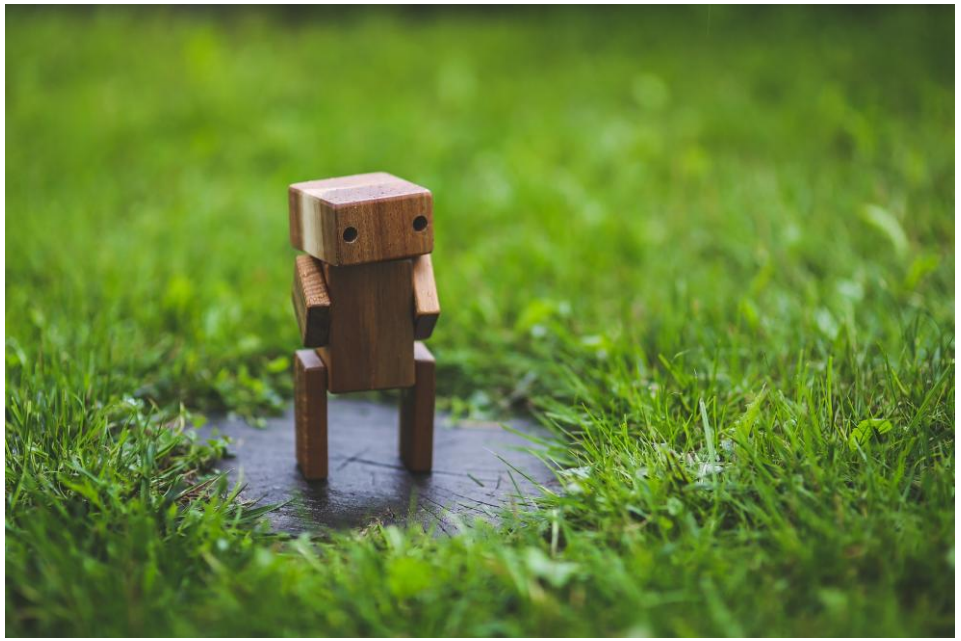
```
##########################################
#####          2. Understanding our data     #####
##########################################
```

Building useful robots

Build the model from the knowledge of your data then use information theory to select models (and make inferences).

A. Model selection
   - Picking the best of a bad bunch.

B. Model validation
   - So how rubbish is it?



Stepwise model selection aims to identify the most “cost” effective model. With “cost” being measured in degrees of freedom.

AIC balance between over and underfitting by estimating out-of-sample deviance without needing to do cross-validation (the gold standard)

A - Model selection

Using AIC (log-likelihood ratio tests are also useful for those who like a p-value)

<u>Selection recipe</u>

- Start with the most complex model and work "back" towards the most simple

- Use AIC to choose (3 rules)
  1. Simple models are best
  2. Small AIC is best
  3. If these rules contradict (ie the more complex model has smaller AIC) then AIC should be different by more than 2 AICs

"I used backwards stepwise model selection to …"

A - Model selection

```
# Eg with "normal error structure"
mod1<- glm(shark_per_h~ perc_cov + cor_col + perc_cov:cor_col,
           data=dat, family=gaussian)


mod2<- glm(shark_per_h~ perc_cov + cor_col, data=dat, family=gaussian)

AIC(mod1,mod2)
summary(mod2)

mod3<- glm(shark_per_h~ perc_cov, data=dat, family=gaussian)
AIC(mod2,mod3)

summary(mod3)


mod_null<- glm(shark_per_h~1, data=dat, family=gaussian)

AIC(mod3, mod_null)
```

"I used backwards stepwise model selection to …"

Table1. What to present in a paper

Model	**Variables**	**AIC**	**Log-Lik ratio …**
Mod1	Percent cover + colour + Percent cover: colour	824.83	…
Mod2	Percent cover + colour	821.90	…

Weirdly it’s not an exact science… (B)

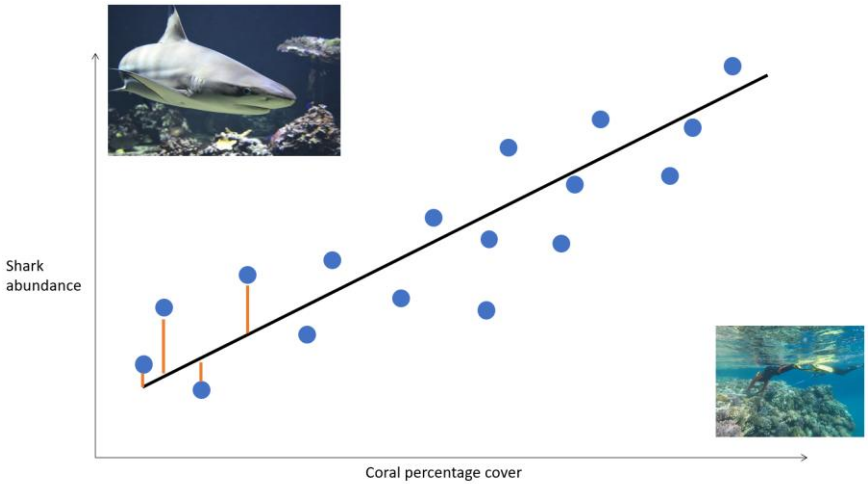
General aim of model validation is:

to have normally distributed residuals

and

no obvious patterns in the residuals

WTF is a residual?



Final interpretation

Simulation allows us to interpret our complex models

A word about coefficients and factors >>>>>>>>

$$Y = mX + c \quad (+\ er)$$

gradient intercept

$$Y = mX_1 + mX_2 + c \quad (+\ er)$$

Three possible outcomes from model selection

Our mod3

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.929914   0.417166   4.626 1.14e-05 ***
perc_cov    0.116220   0.007622  15.248  < 2e-16 ***
---
```

Our mod2

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.866607   0.513668  11.421  < 2e-16 ***
perc_cov      0.094528   0.008575  11.024  < 2e-16 ***
cor_colBrown -0.986790   0.143164  -6.893 5.75e-10 ***
cor_colGreen -0.580381   0.111171  -5.221 1.03e-06 ***
---
```

Our mod1

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           5.49784    0.60079   9.151 1.17e-14 ***
perc_cov              0.10071    0.01005  10.026  < 2e-16 ***
cor_colBrown         -1.61468    1.00986  -1.599  0.11320
cor_colGreen          3.69075    1.38125   2.672  0.00889 **
perc_cov:cor_colBrown 0.01524    0.02008   0.759  0.44973
perc_cov:cor_colGreen -0.08092    0.02581  -3.135  0.00229 **
---
```

Final interpretation

From the count data model

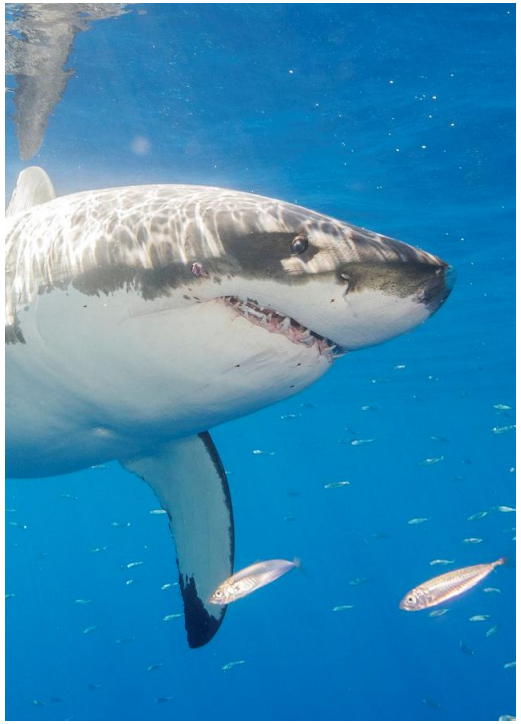
Our mod1

```
Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)           1.513482   0.098614  15.348  < 2e-16 ***
perc_cov              0.018233   0.001386  13.154  < 2e-16 ***
cor_colBrown         -0.110811   0.137758  -0.804 0.421172
cor_colGreen         -0.362180   0.135972  -2.664 0.007730 **
perc_cov:cor_colBrown  0.002505   0.001980   1.265 0.205711
perc_cov:cor_colGreen  0.006713   0.001975   3.399 0.000677 ***
```

R code section

```
##########################################
#####          3. Model selection          #####
##########################################
```

Pause



Don’t stare into the abyss: understanding your model.

Final interpretation – a word about staring into the abyss

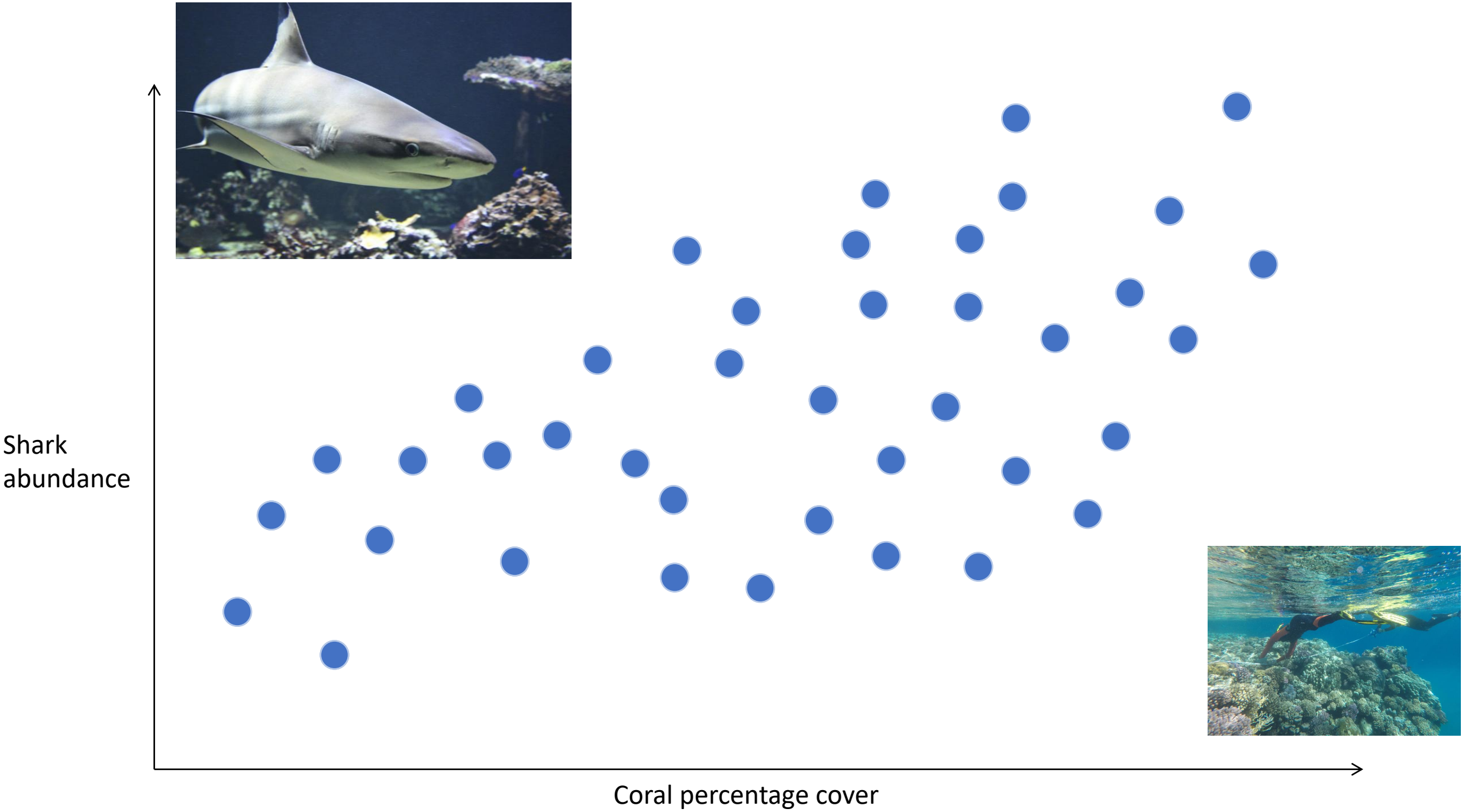
A very simple model

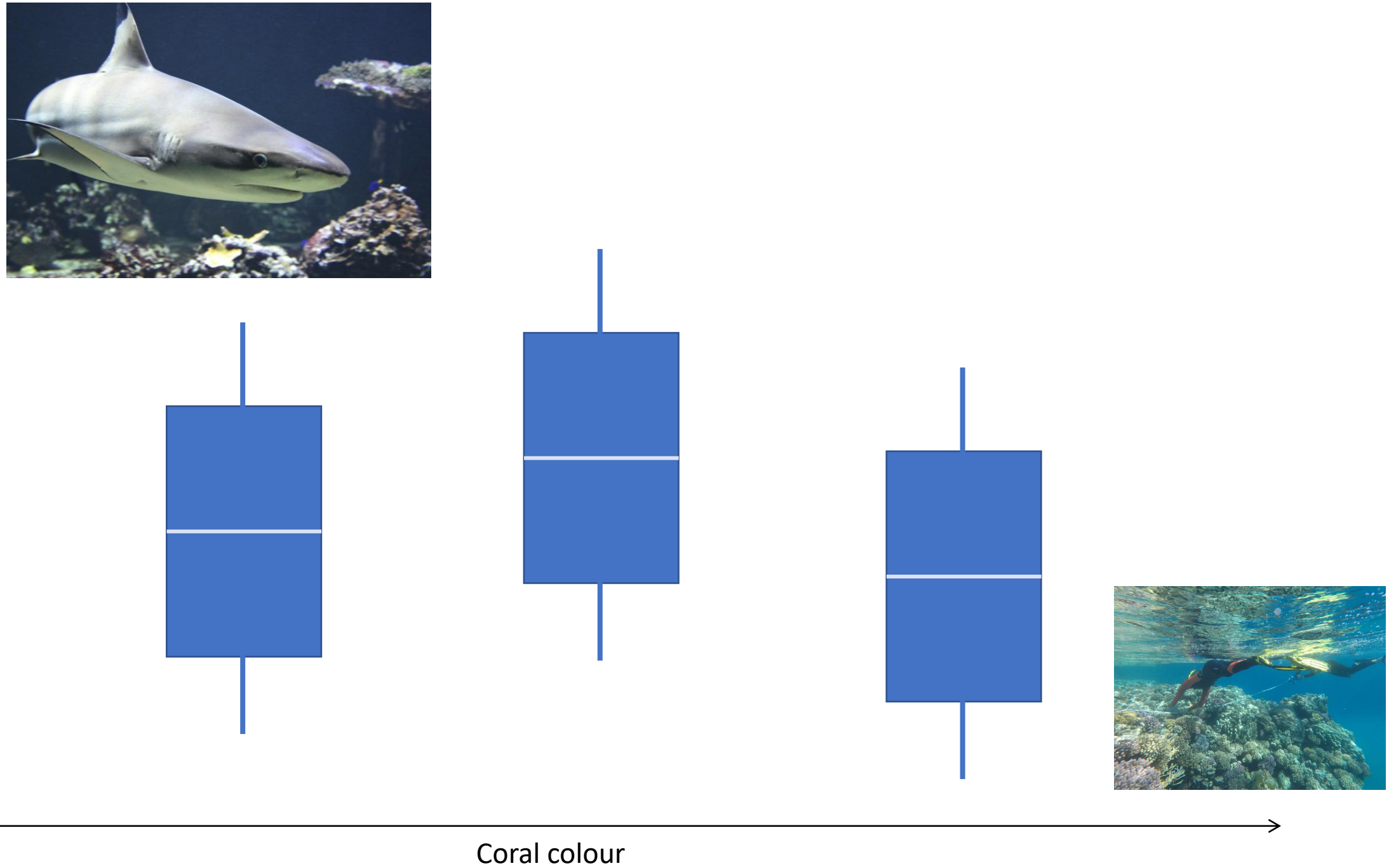
```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 1.913513   0.048185   39.71   <2e-16 ***
perc_cov    0.012233   0.000753   16.25   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

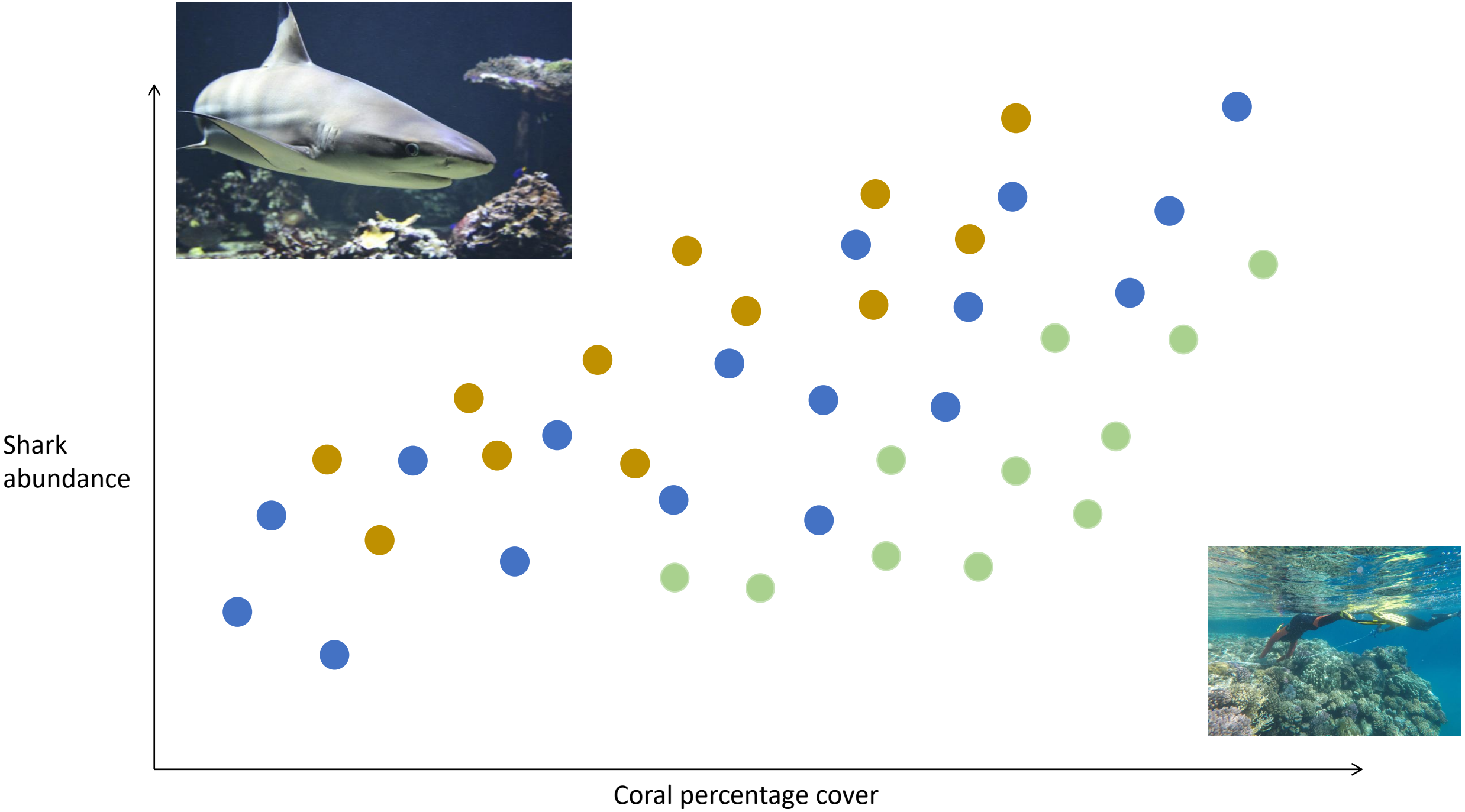
Even slightly more complex models become difficult to interpret

Remember too this was a model of count data so: exp(1.51) = 4.54

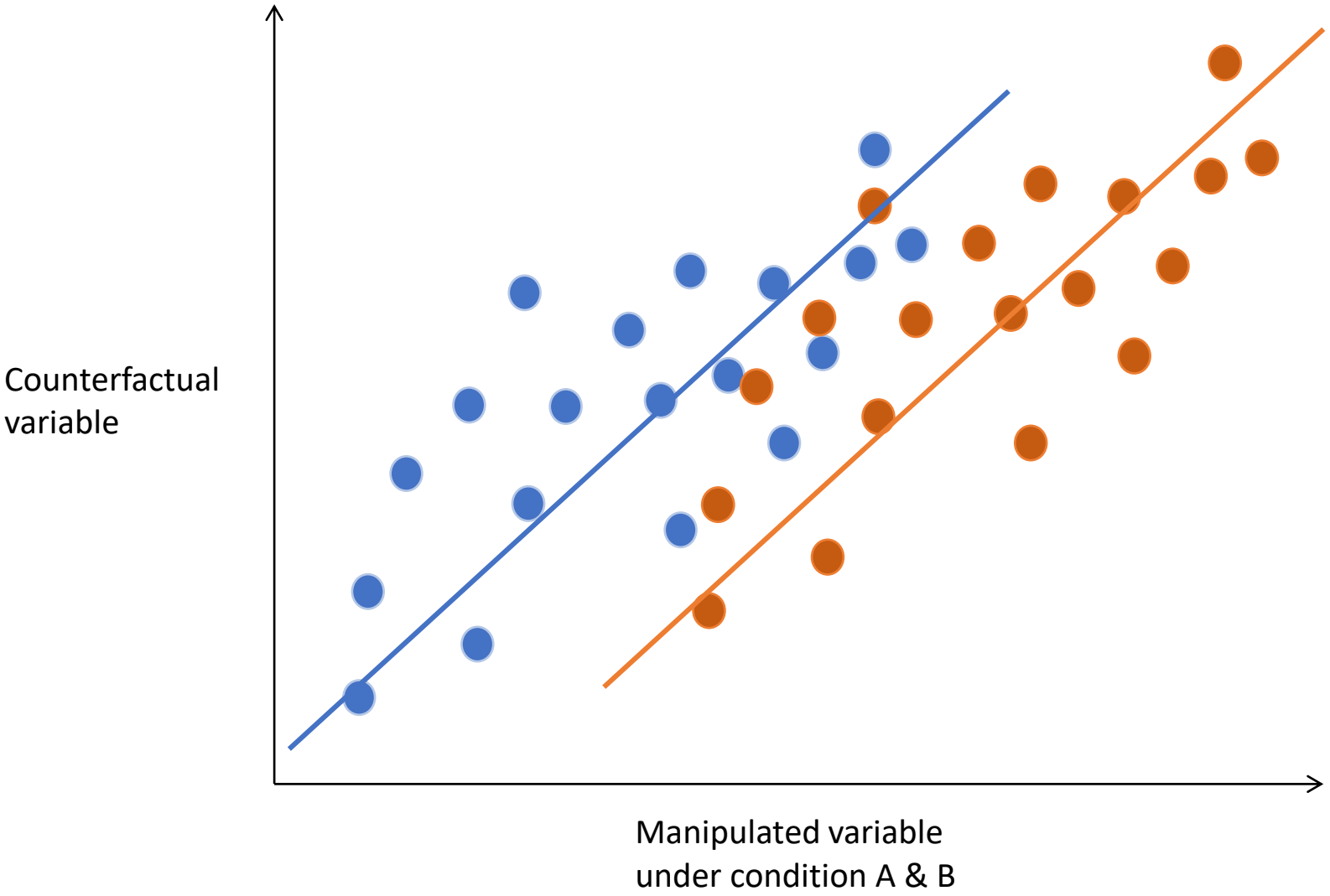
```
Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          1.513482   0.098614  15.348  < 2e-16 ***
perc_cov             0.018233   0.001386  13.154  < 2e-16 ***
cor_colBrown        -0.110811   0.137758  -0.804 0.421172
cor_colGreen        -0.362180   0.135972  -2.664 0.007730 **
perc_cov:cor_colBrown  0.002505   0.001980   1.265 0.205711
perc_cov:cor_colGreen  0.006713   0.001975   3.399 0.000677 ***
```


Shark abundance
Coral percentage cover


Shark abundance
Coral colour


Shark abundance
Coral percentage cover

Counterfactuals – One continuous variable under a set of conditions (factor/treatment)



Counterfactuals – in practice

We provide R a set of values for the X variables and ask R to predict y (shark abundance).

```
Neils_new_dat <- data.frame(
  expand.grid(perc_cov= c(0:100), cor_col= levels(dat$cor_col)
              ))
```

```
preds<- predict(mod1.p, newdata = Neils_new_dat, type='response' )
pred.dat<- cbind(Neils_new_dat, preds)
```

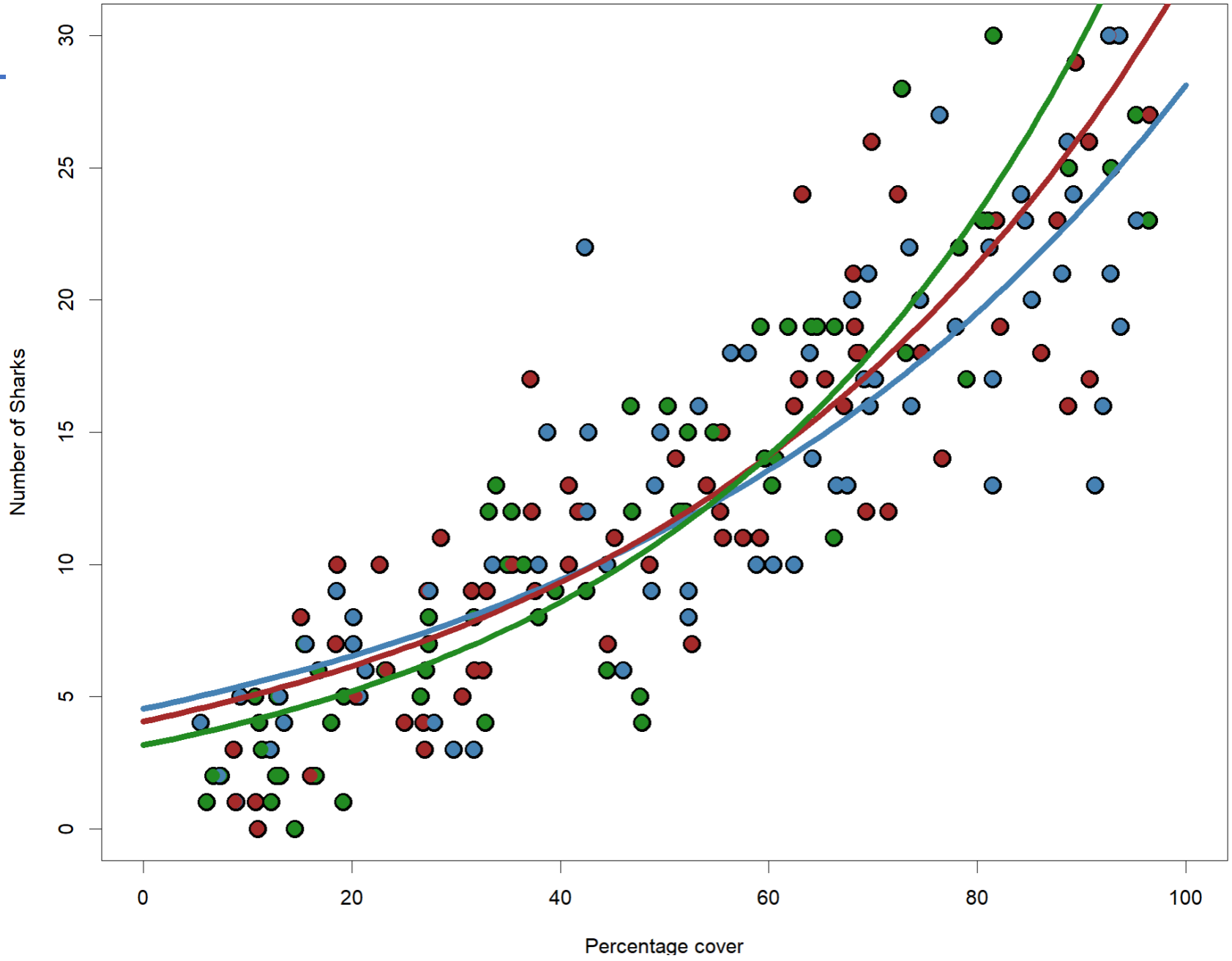
```
  perc_cov cor_col    preds
1        0    Blue 4.542519
2        1    Blue 4.626101
3        2    Blue 4.711222
4        3    Blue 4.797908
5        4    Blue 4.886190
6        5    Blue 4.976096
```

We can then carve up the data (pred.dat) to display each coral colour on its own line

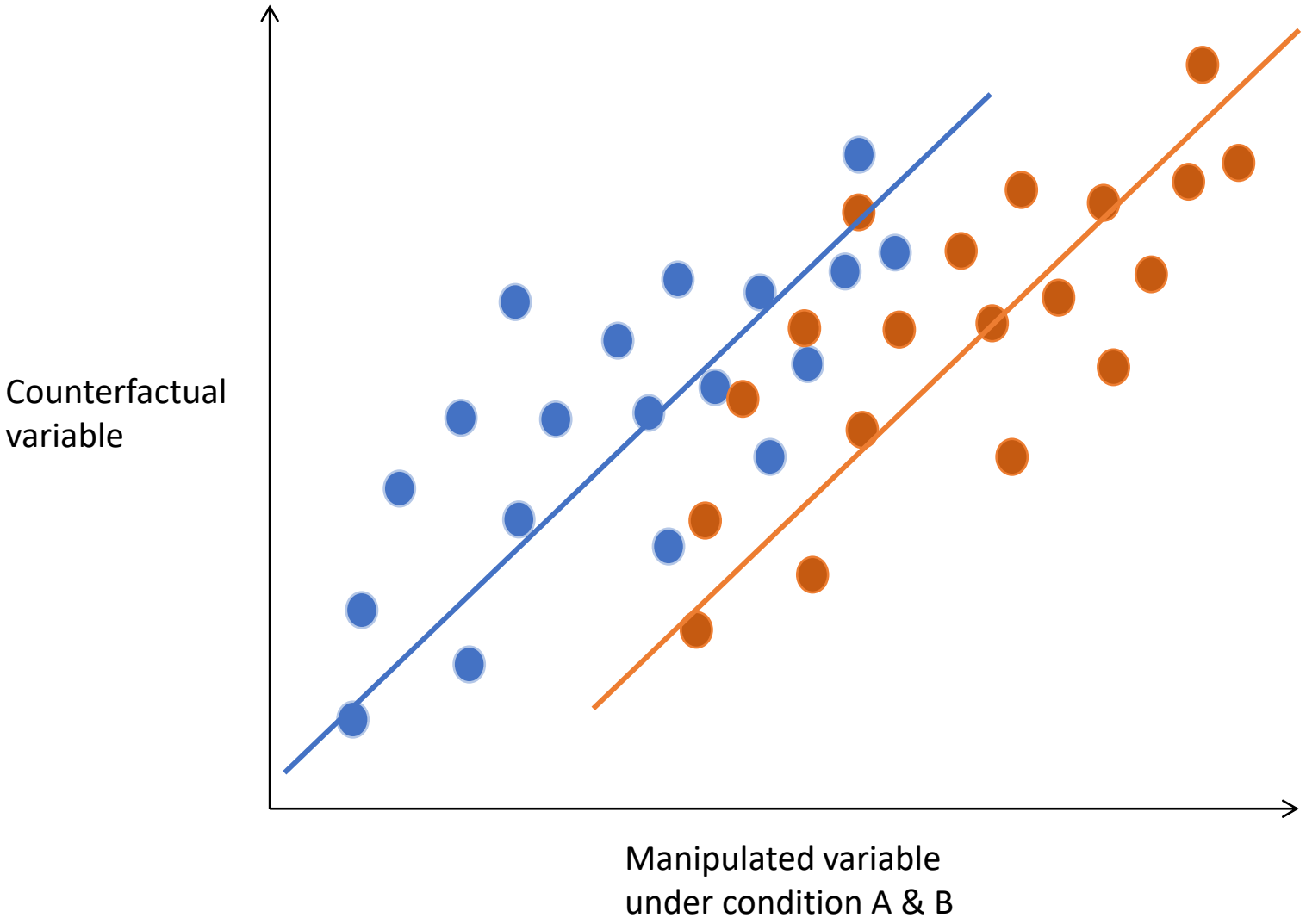
```
line_blue <- subset(pred.dat, pred.dat$cor_col=='Blue')
line_brown <- subset(pred.dat, pred.dat$cor_col=='Brown')
line_green <- subset(pred.dat, pred.dat$cor_col=='Green')
```

Counterfactuals

What is the effect of coral cover on the number of sharks given the dominant colour of coral present?



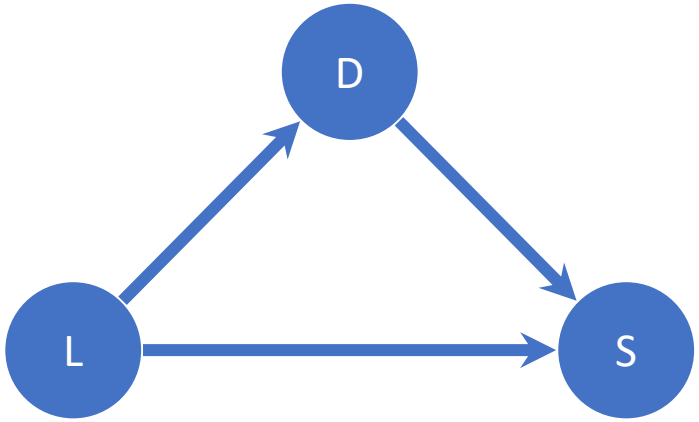
Counterfactuals – One continuous variable under a set of conditions (factor/treatment)



What about several continuous variables?

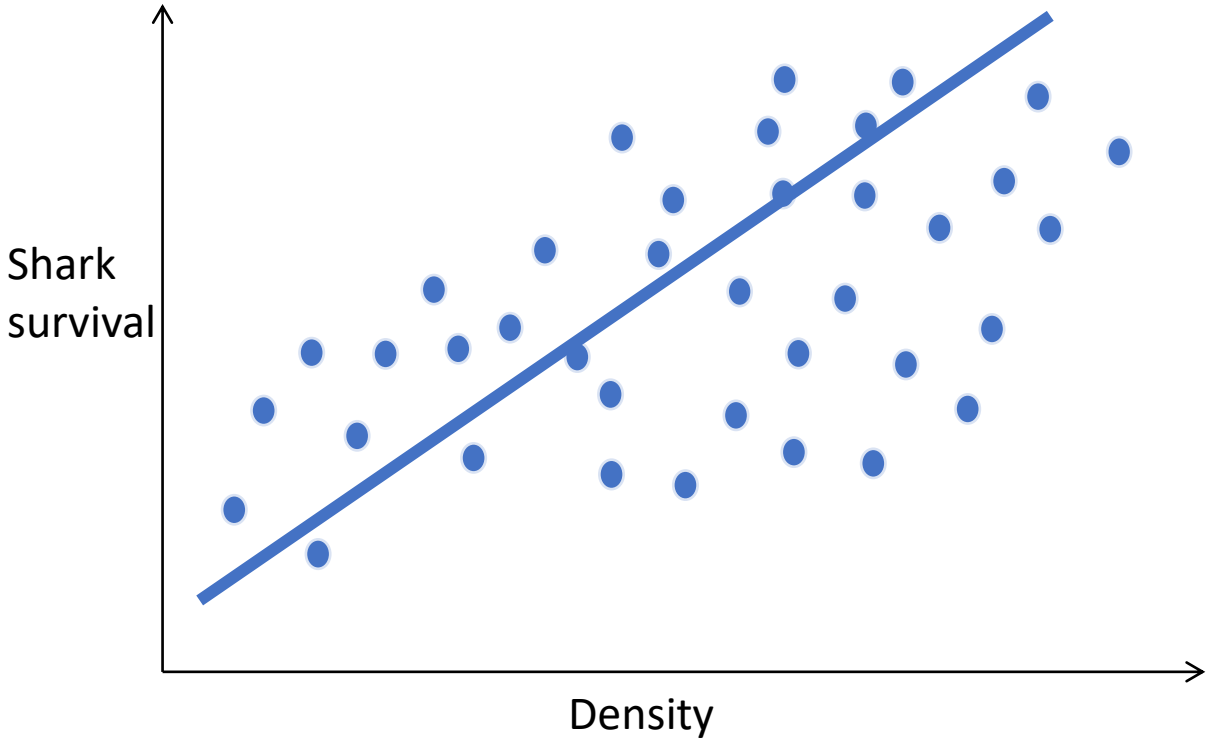
Example 2 – Shark…survival

L= shark length
D = shark density
S = survival

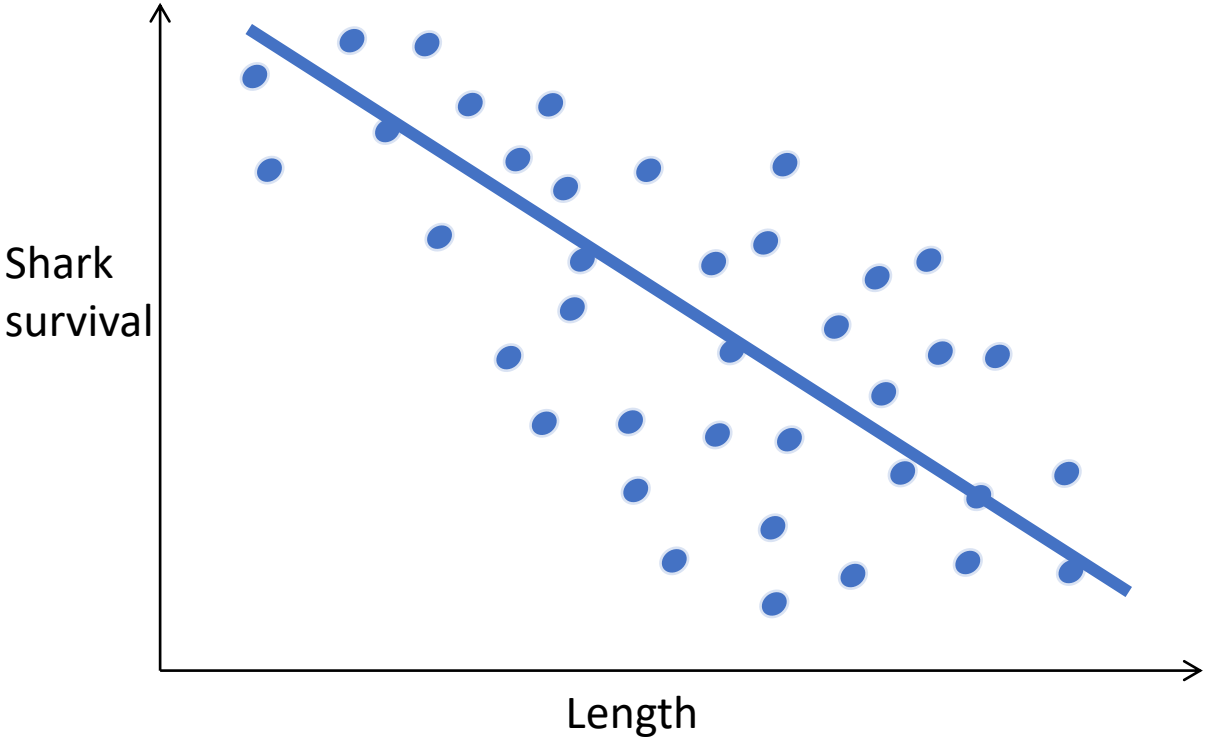


Method to deal with multiple continuous variable counterfactual

Counterfactual holding length at 3 m



Counterfactual holding density at 5 per km^2




```
##########################################
######### 4. Example 2. Shark survival   #########
##########################################
```

Counterfactuals – with several continuous variables

We provide R a set of values for the X variables and ask R to predict y (Survivorship (fit)).

```
Neils_new_dat3<- data.frame(expand.grid(length=(5:10), density = c(1:6)))
```

```
preds<- data.frame(predict(modA, newdata = Neils_new_dat3, type='response', se.fit = TRUE ))
```

```
## attach all the data - makes it easier to do graphing later
pred.dat<- cbind(preds, up, lo, Neils_new_dat3)
```

```
         fit      se.fit residual.scale         up         lo length density
1 0.25602707 0.013662821     0.01370157 0.28280620 0.22924794      5       1
2 0.21757956 0.009437034     0.01370157 0.23607615 0.19908297      6       1
3 0.17913205 0.006020575     0.01370157 0.19093237 0.16733172      7       1
4 0.14068454 0.005314302     0.01370157 0.15110057 0.13026850      8       1
5 0.10223703 0.008064804     0.01370157 0.11804404 0.08643001      9       1
6 0.06378952 0.012109711     0.01370157 0.08752455 0.04005448     10       1
```

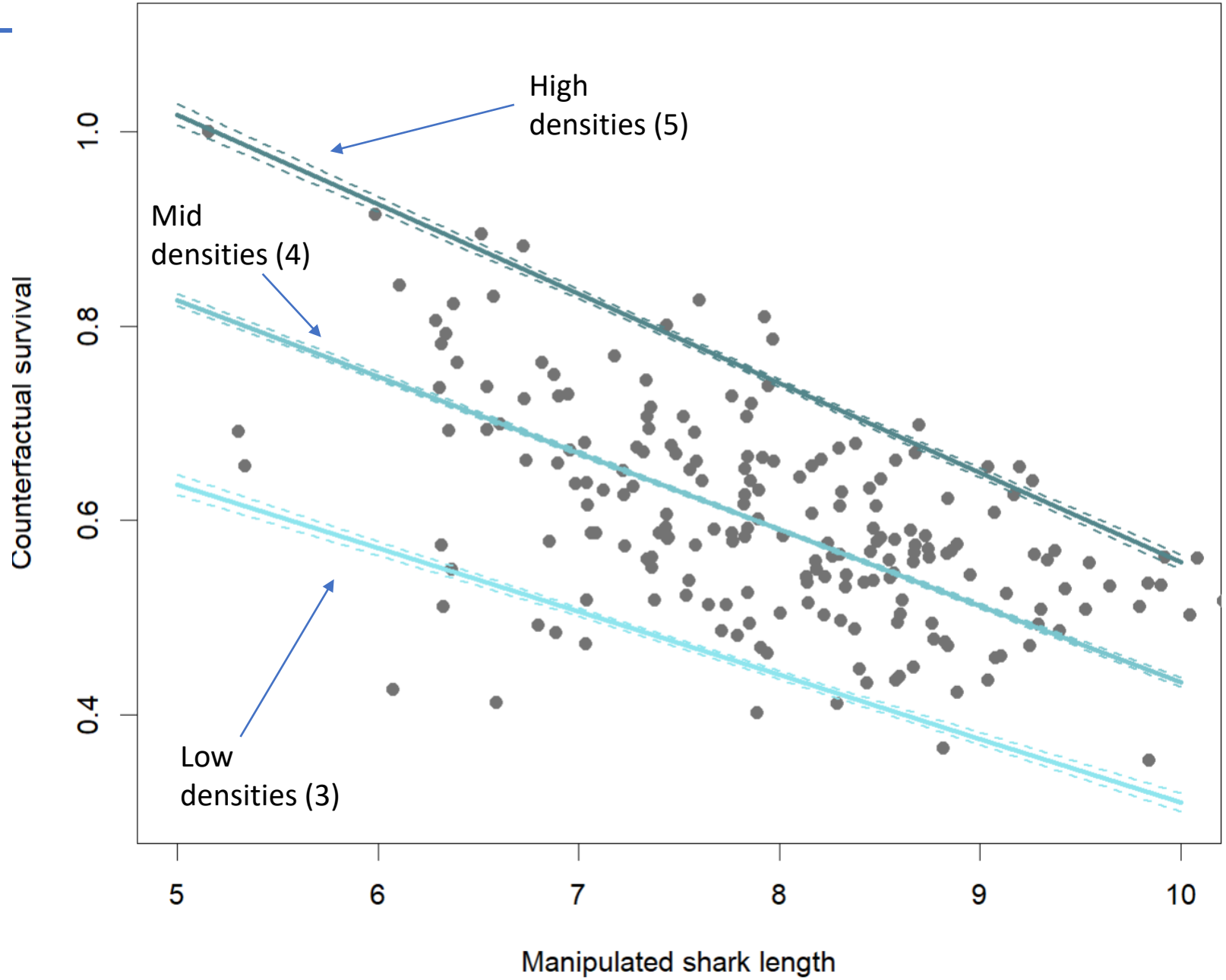
We can then carve up the data (pred.dat) to display values of y (fit) for our chosen values of X. E.G. across all lengths at 3 values of density

```
at.mid.dens<- subset(pred.dat, pred.dat$density== 4)
at.lo.dens<- subset(pred.dat, pred.dat$density== 3)
at.hi.dens<- subset(pred.dat, pred.dat$density== 5)
```

Counterfactuals: continuous variables

What happens to our survivability when we encounter larger sharks?

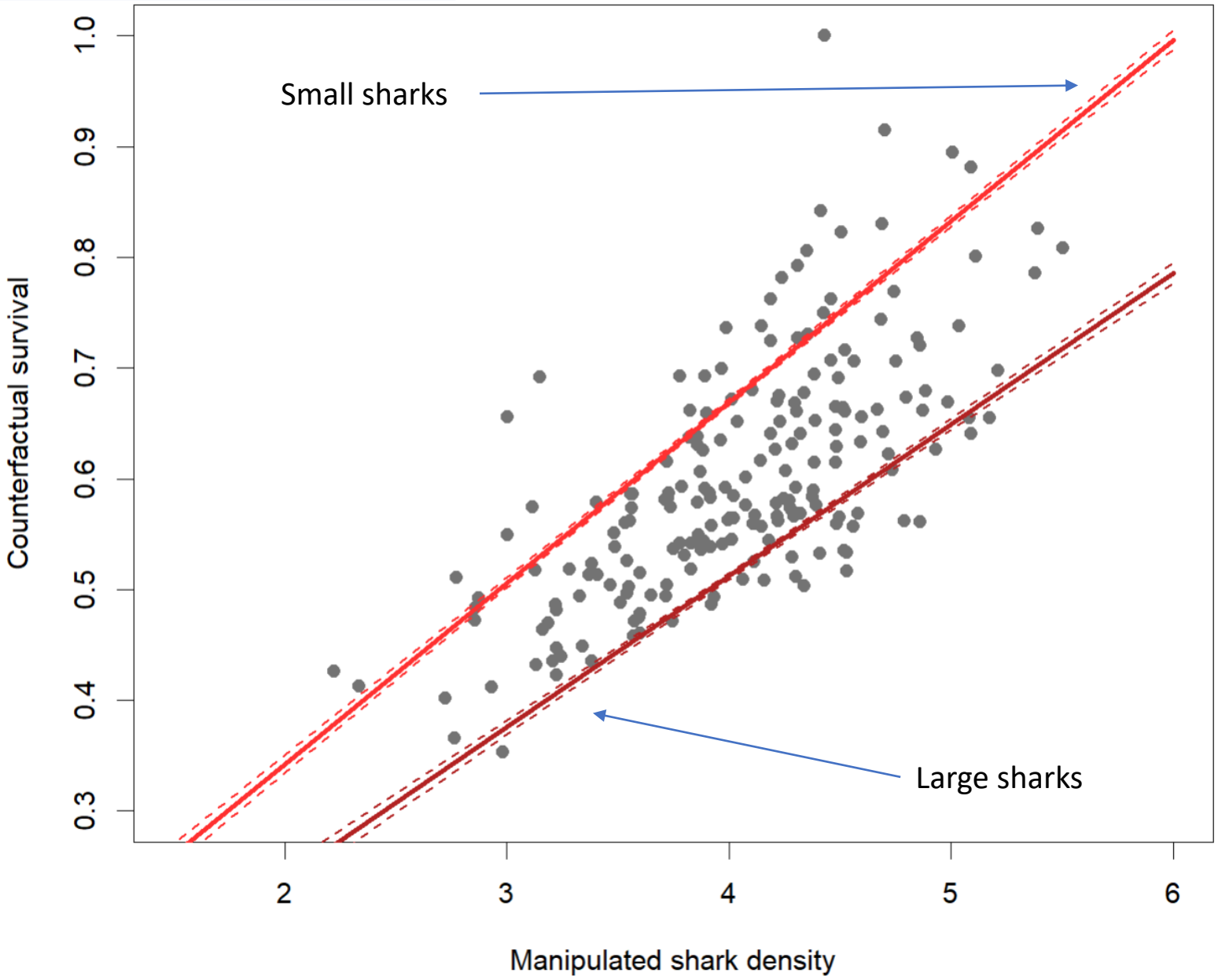
What happens when we encounter sharks of different lengths at different densities?



Counterfactuals: continuous variables

What happens to our survivability when we encounter larger sharks?

What happens when we encounter sharks of different length at different densities in the?



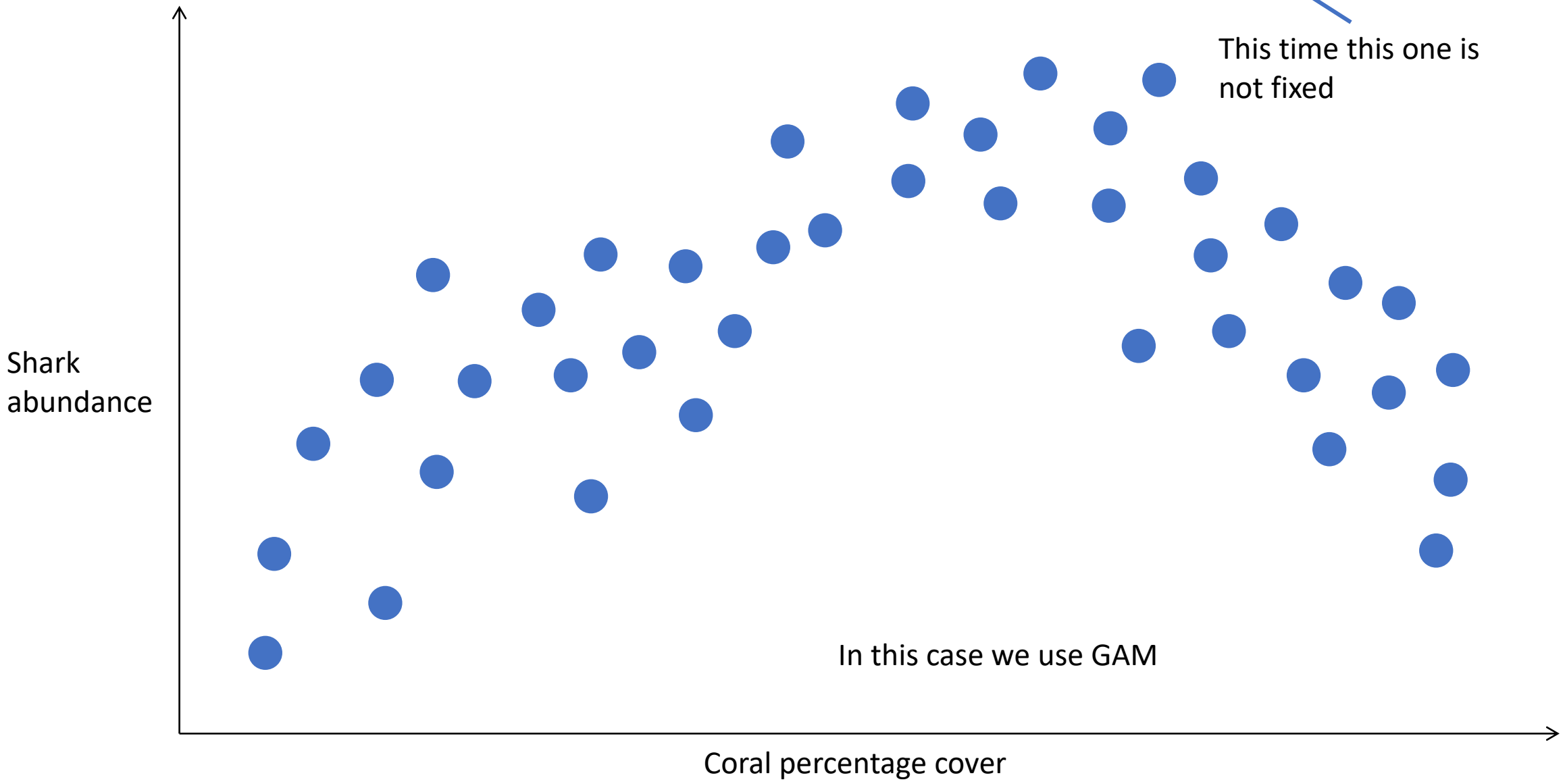
R code section

```
##########################################
#####          3. Model interpretation          #####
##########################################
```

Ahh wait, but what if…?

$$Y = mx + c \quad (+ er)$$

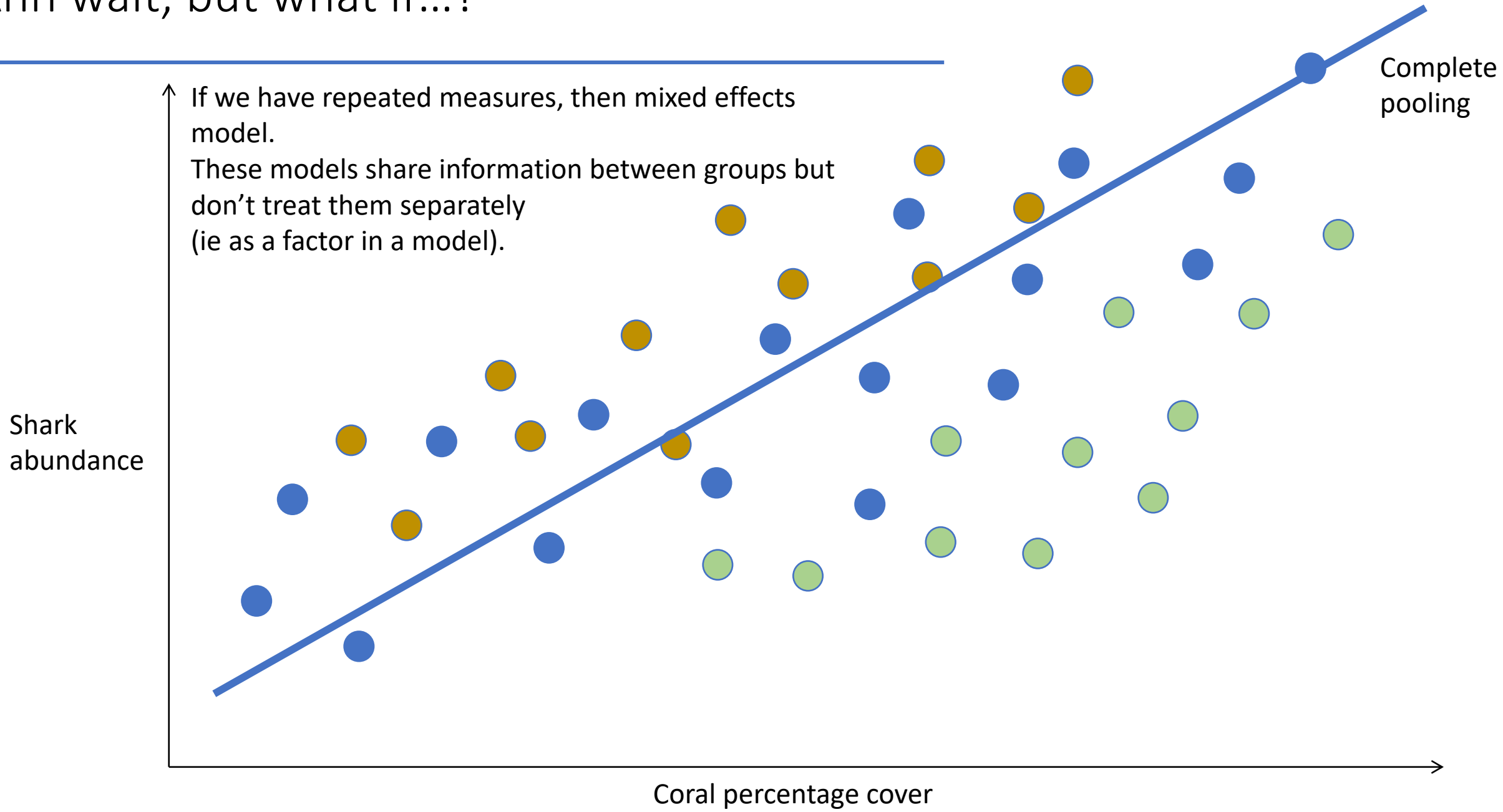
This time this one is not fixed



In this case we use GAM

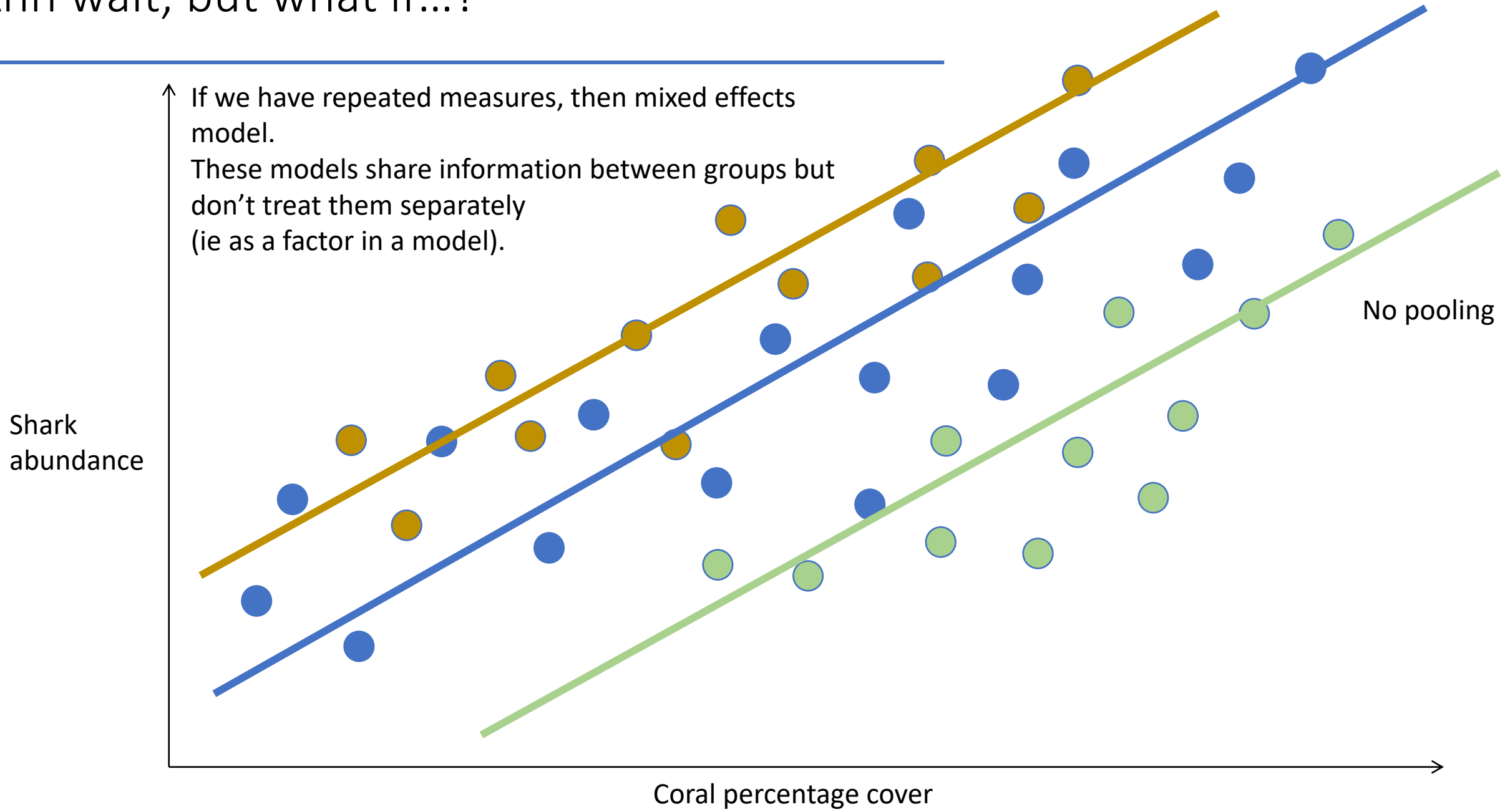
Ahh wait, but what if…?

If we have repeated measures, then mixed effects model.
These models share information between groups but don't treat them separately
(ie as a factor in a model).



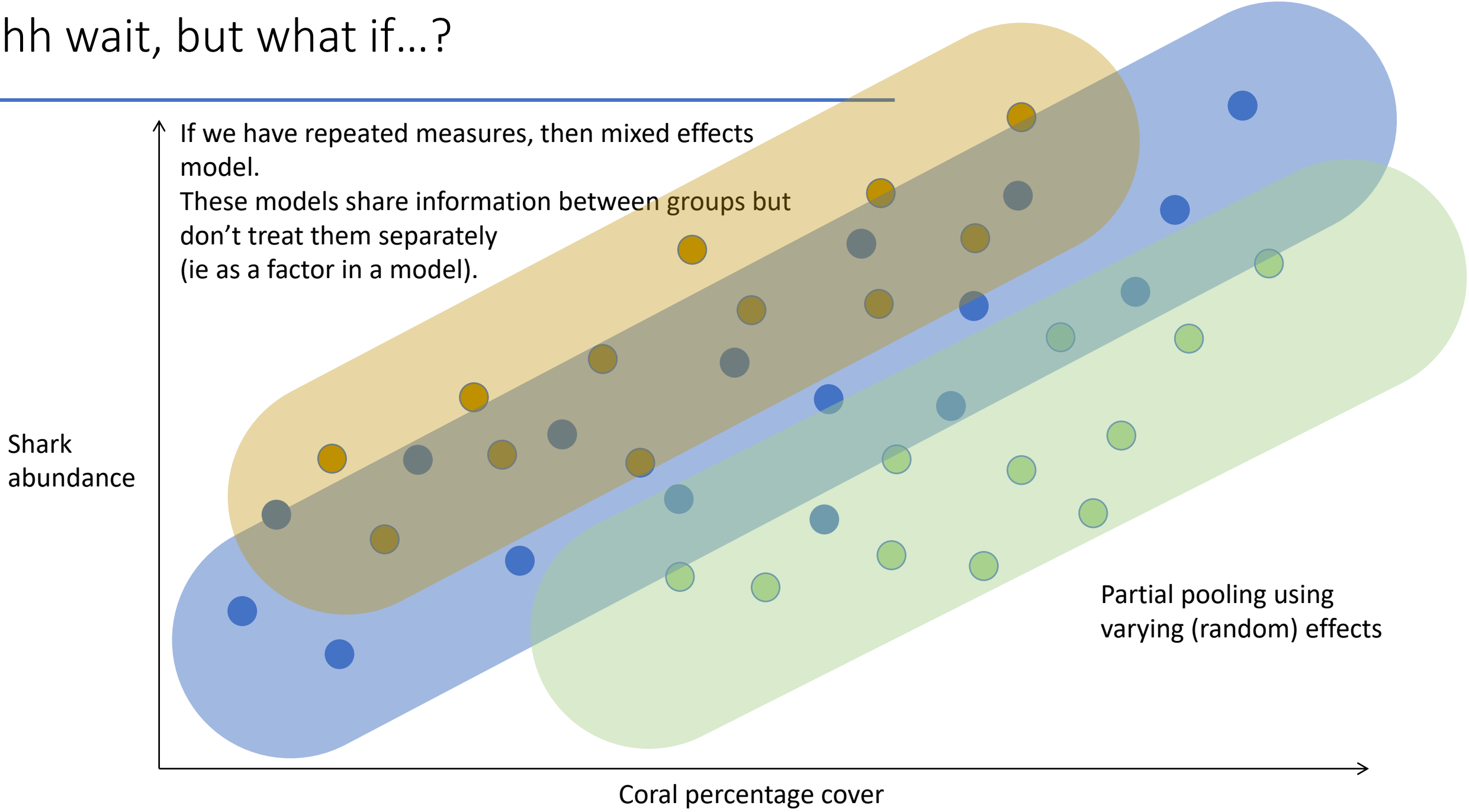
Ahh wait, but what if…?

If we have repeated measures, then mixed effects model.
These models share information between groups but don't treat them separately
(ie as a factor in a model).



Ahh wait, but what if…?

If we have repeated measures, then mixed effects model.
These models share information between groups but don't treat them separately
(ie as a factor in a model).

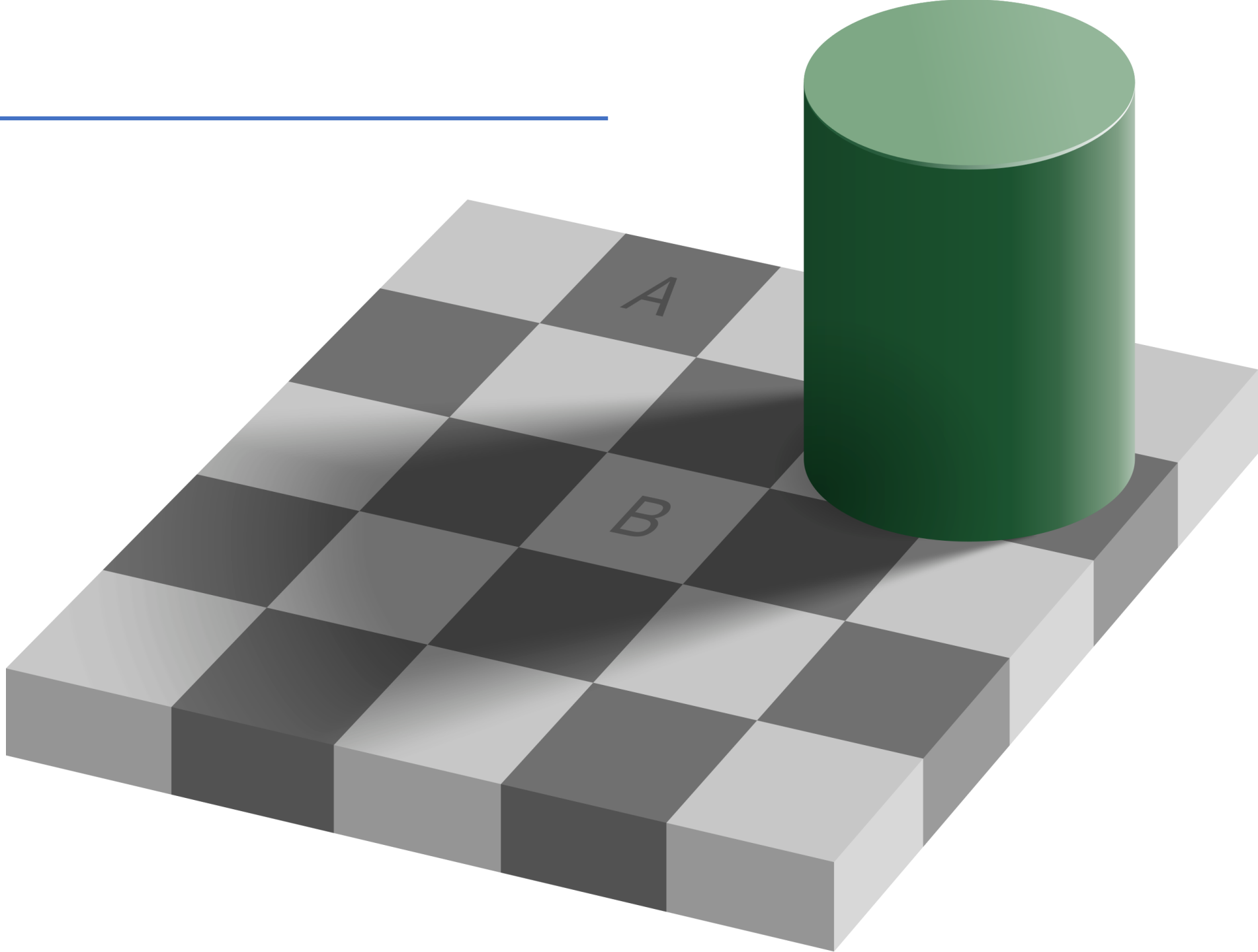


Bayesian methods

It’s the way you think any way

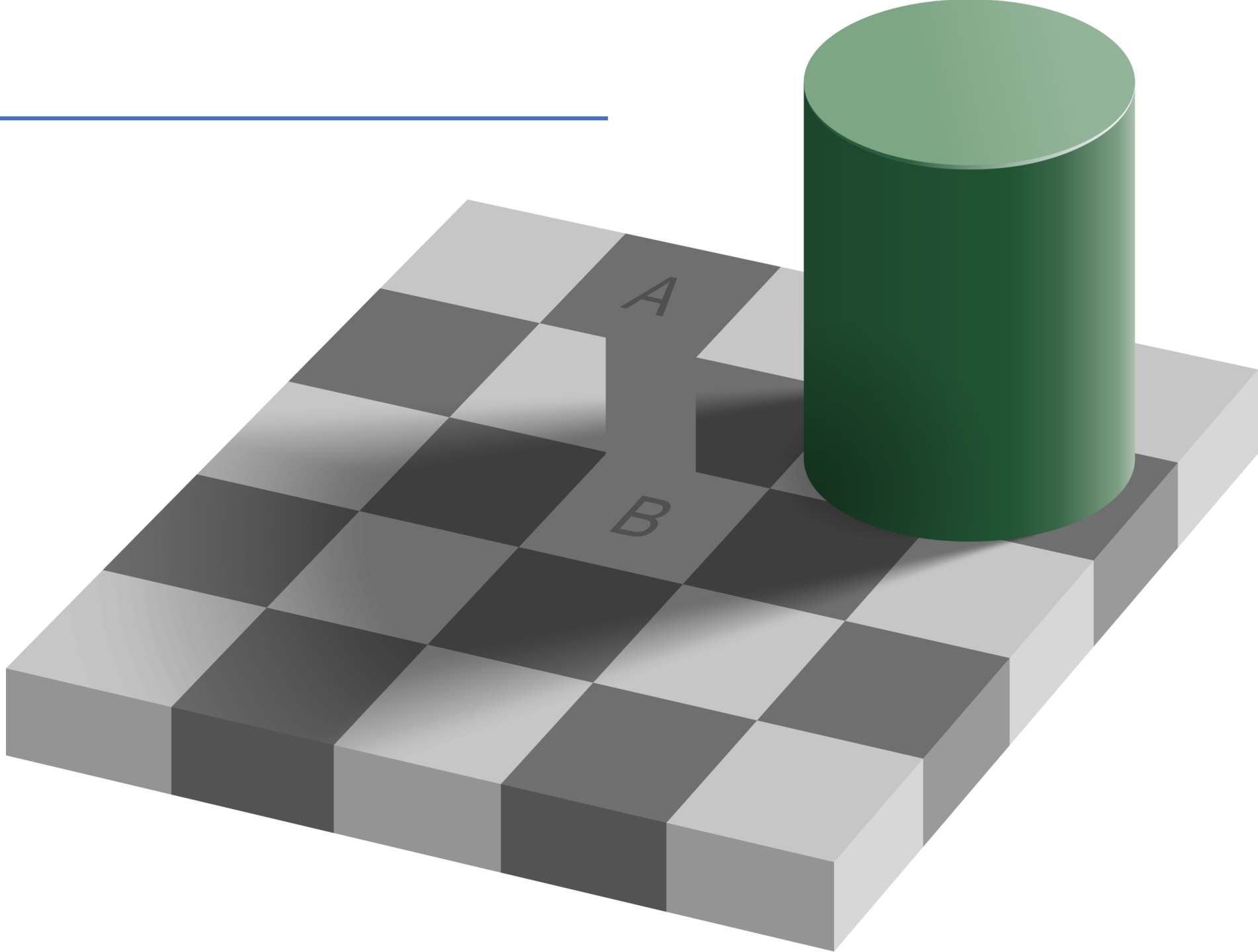
Bayesian methods

It’s the way you think any way



Bayesian methods

It’s the way you think any way



Summary

- Graph and understand your data
- Avoid spending too long down rabbit holes
- Use the tools we have discussed to build the robot
  - and the “recipe” to perform model selection
- We can use the *PGR Stats forum on Moodle* to discuss and resolve problems

PGR development

- You direct focus

- Suggestions
  - Continuation of this session
    - GLMM
    - GAM
    - Bayesian
  - Multivariate modelling
  - Spatial data analysis