Survey design and statistical analysis for conservation science

Philipp Boersch-Supan
British Trust for Ornithology

@pboesu

with thanks to Alison Johnston

BTO
Looking out for birds

Fellowes

Survey design and statistical analysis for conservation science

SURVEY DESIGN AND SAMPLING

STATISTICAL ANALYSIS

COMMUNICATING RESULTS

About me

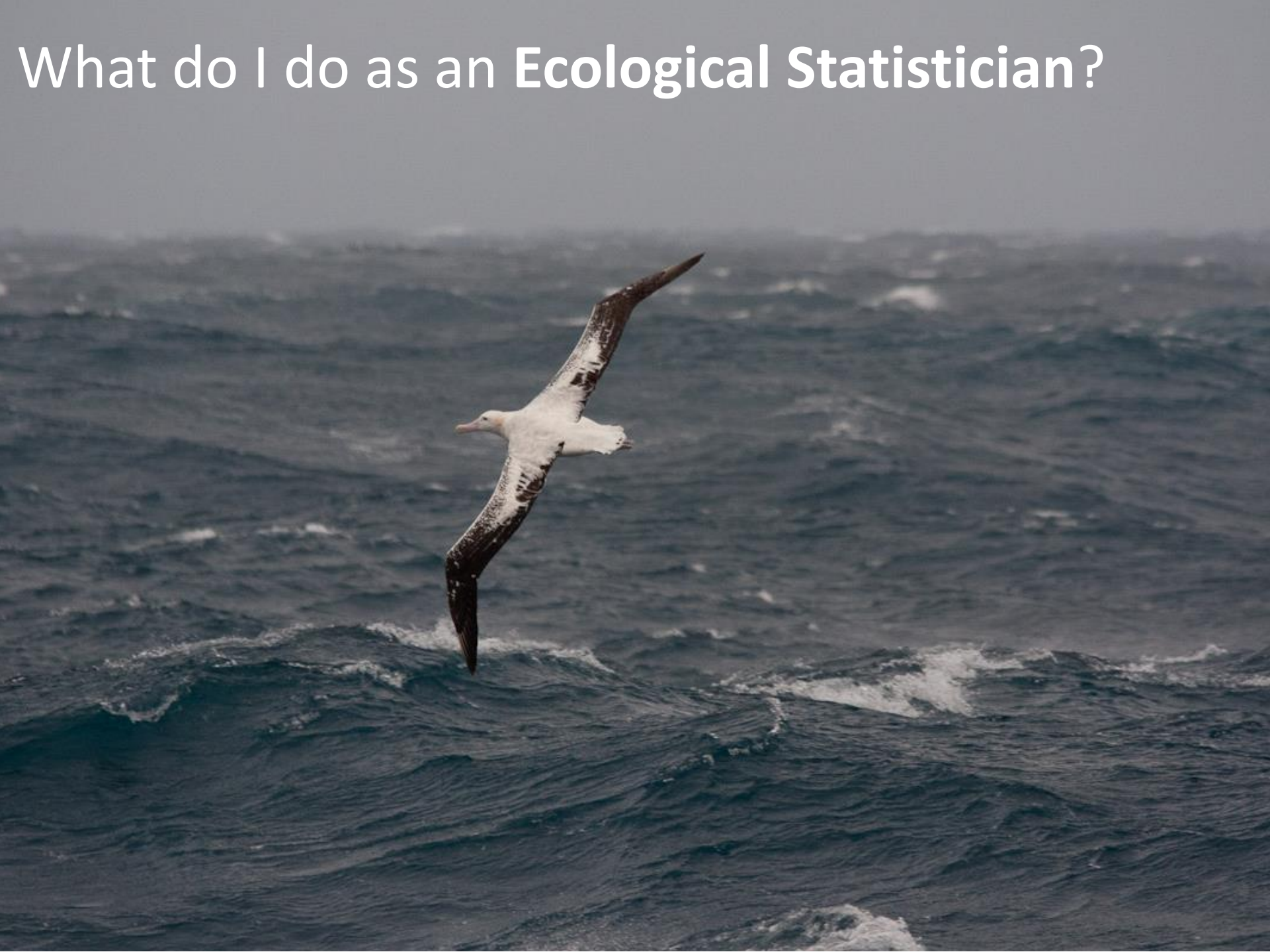
		Field	Lab	Maths	Stats	Coding
2005	Field Ornithologist	■				
2006-2008	UG in Chemistry		■	■		
2008-2009	MRes in Environmental Biology	■		■	■	■
2009-2013	PhD in Marine Ecology	■	■		■	■
2013-2014	Analyst at British Antarctic Survey				■	■
2014-2018	Postdoc Quantitative Ecology			■	■	■
2018-	Ecological Statistician				■	■

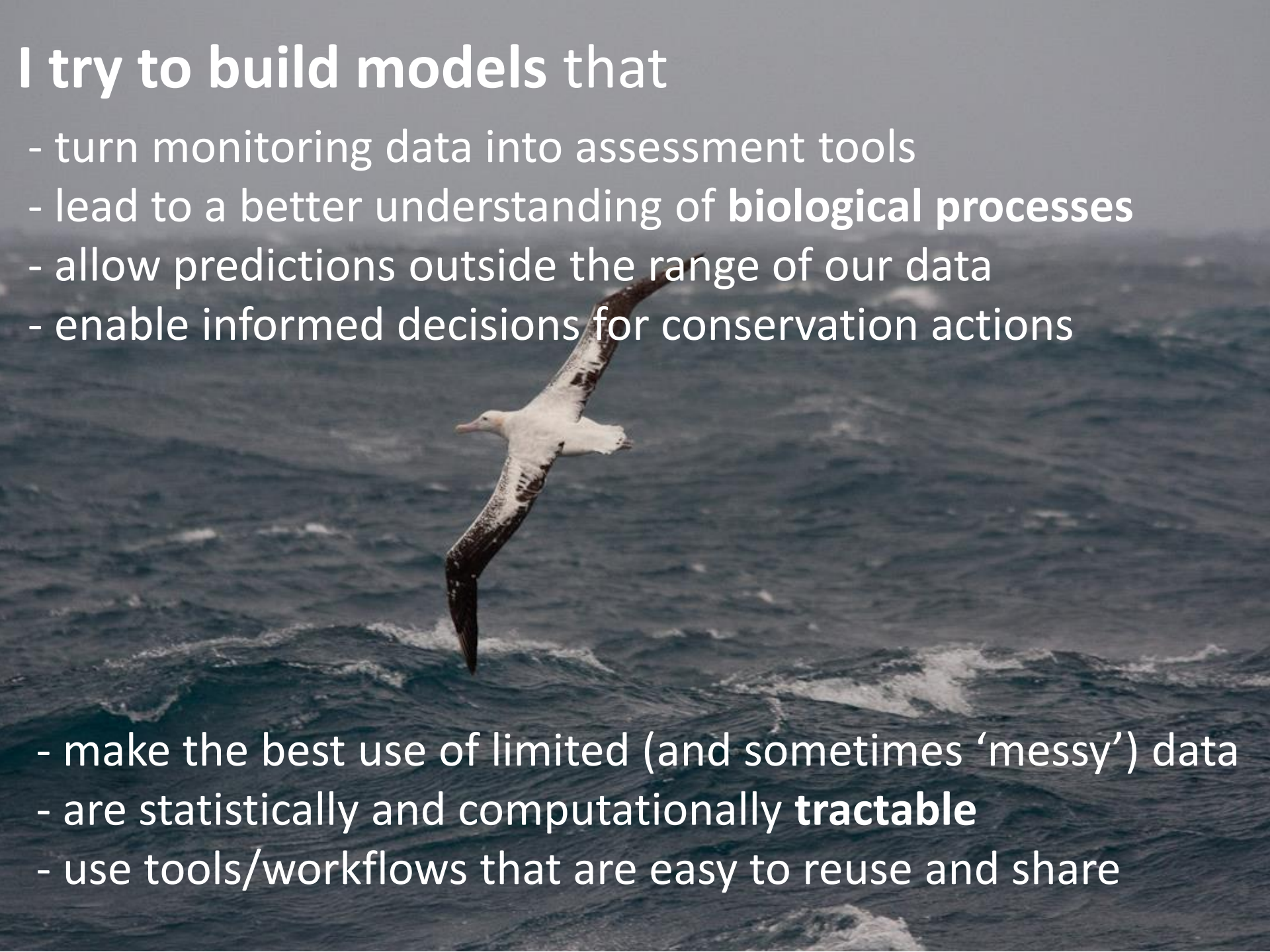
The British Trust for Ornithology

- We are an independent charitable research institute
- We monitor UK birds with the help of over 40,000 volunteers
- We employ c. 100 staff
  - c. 50 scientists
  - c. 50 communicators, fundraisers, volunteer coordinators, administrators, software/database/web developers
- We don't advocate but aim to provide impartial evidence to inform the public, opinion-formers and policy/decision-makers.

Image: DAVID TIPLING

What do I do as an **Ecological Statistician**?





I try to build models that

- turn monitoring data into assessment tools
- lead to a better understanding of **biological processes**
- allow predictions outside the range of our data
- enable informed decisions for conservation actions

- make the best use of limited (and sometimes ‘messy’) data
- are statistically and computationally **tractable**
- use tools/workflows that are easy to reuse and share

Ecological models

Tactical/Phenomenological

Strategic/Mechanistic

Describe patterns without elucidating mechanism

Prediction within range of data

Statistical models (regressions, etc.)

Focus on process/mechanisms

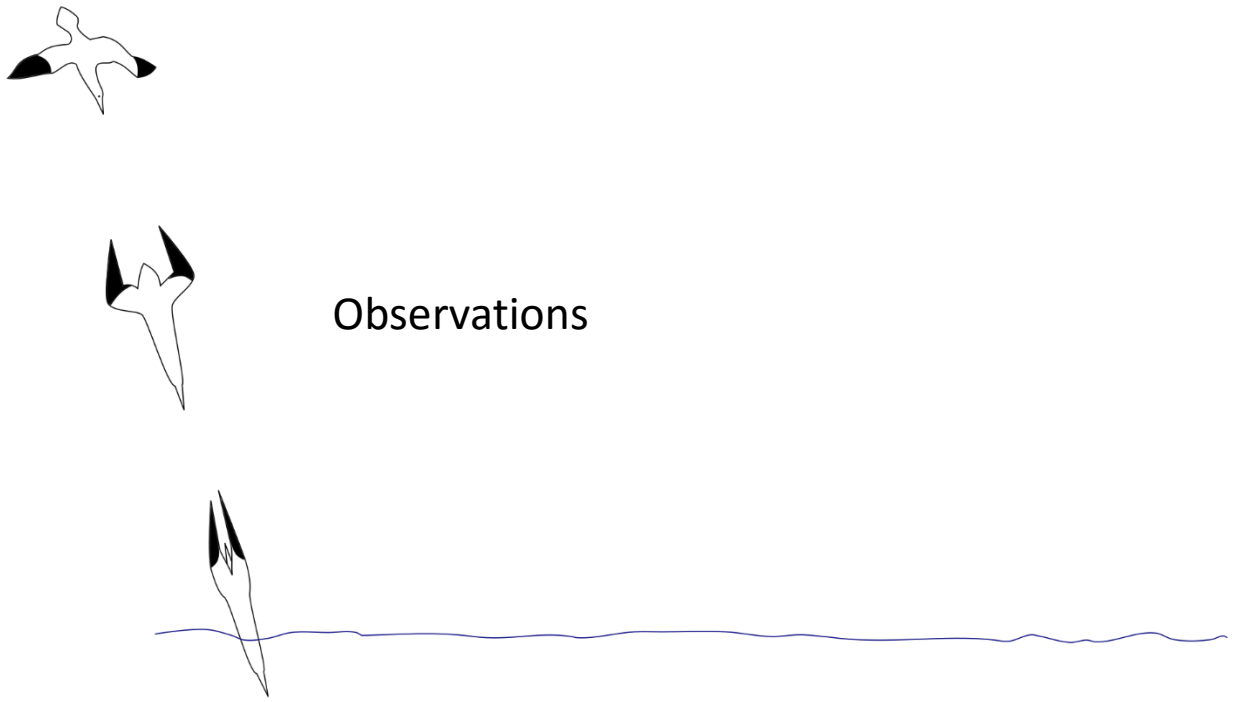
Explanation or understanding
Prediction outside range of data

Math models e.g. ODEs, PDEs
Individual/Agent based Models

Historically largely separate communities of researchers within ecology, but increasingly integration of data and theory-focussed approaches happens

Inspired by Leah Johnson

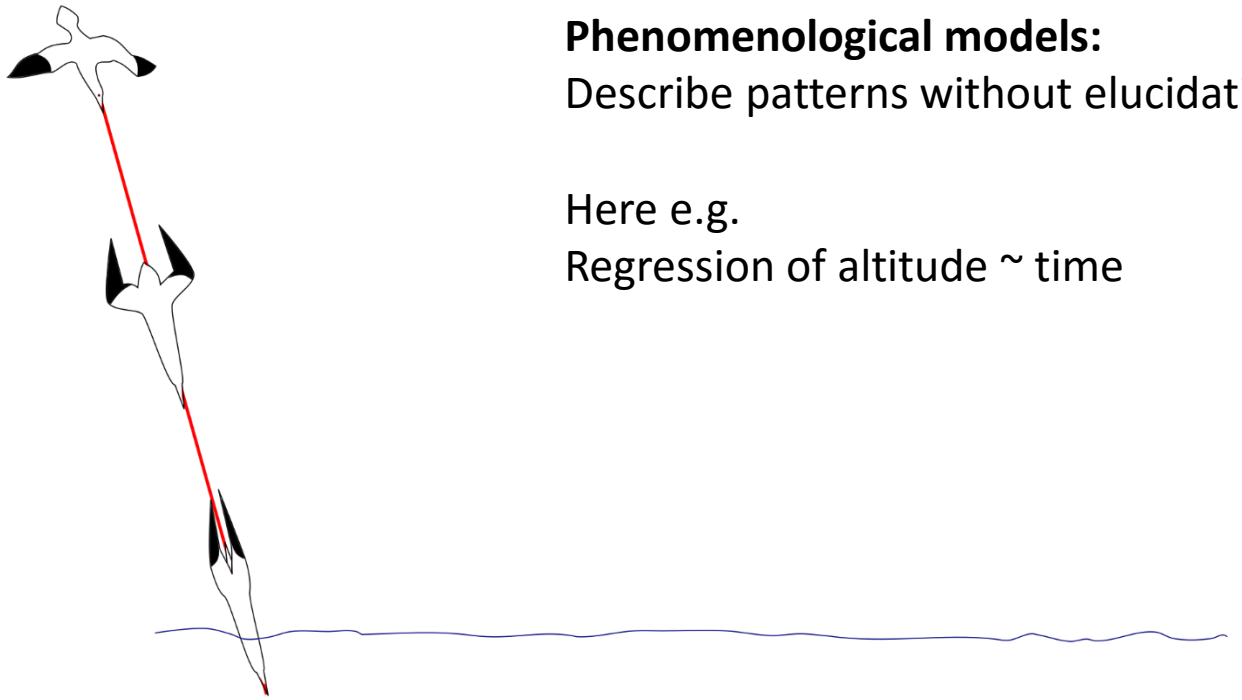
Models in ecology – an analogy



Observations

inspired by Maino et al. (2016) Curr Opp Insect Sci

Models in ecology – an analogy

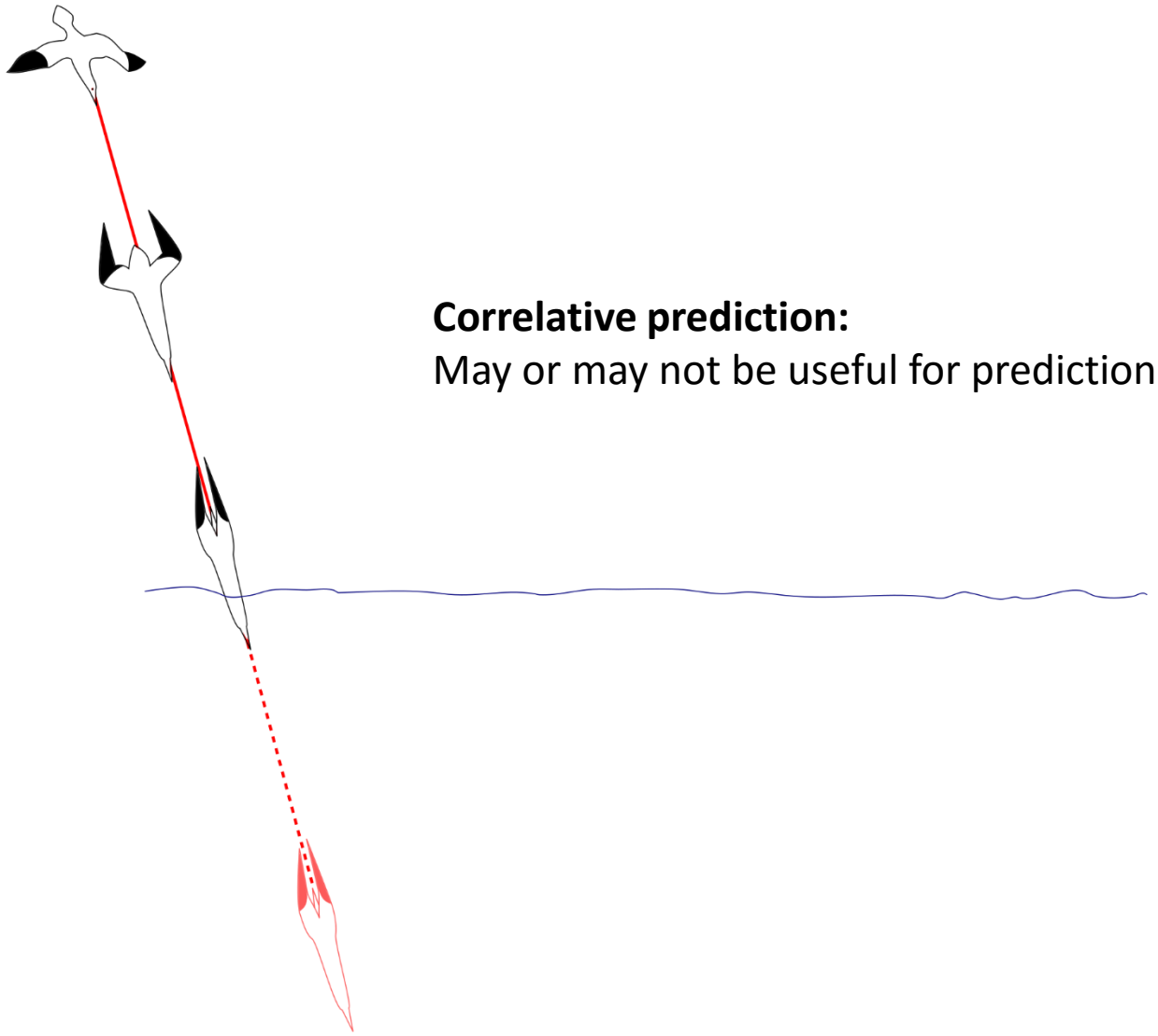


Phenomenological models:
Describe patterns without elucidating mechanism

Here e.g.
Regression of altitude ~ time

inspired by Maino et al. (2016) Curr Opp Insect Sci

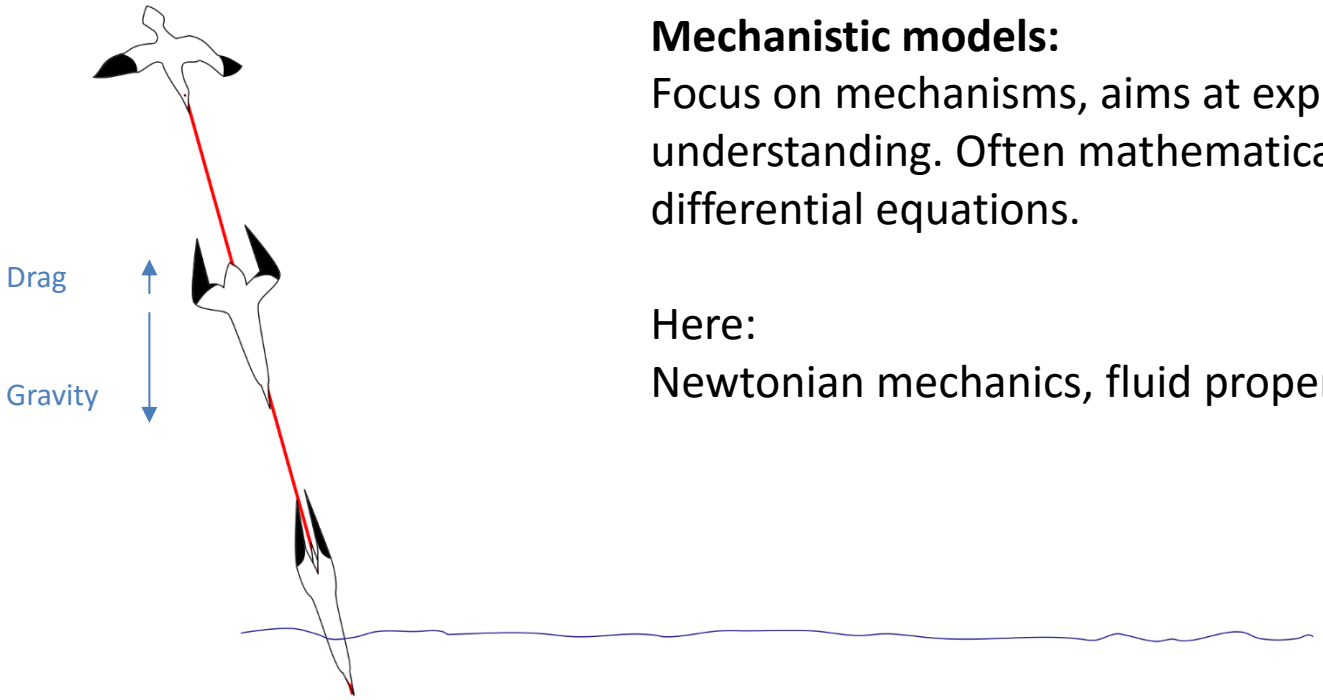
Models in ecology – an analogy



Correlative prediction:
May or may not be useful for prediction beyond data

inspired by Maino et al. (2016) Curr Opp Insect Sci

Models in ecology – an analogy



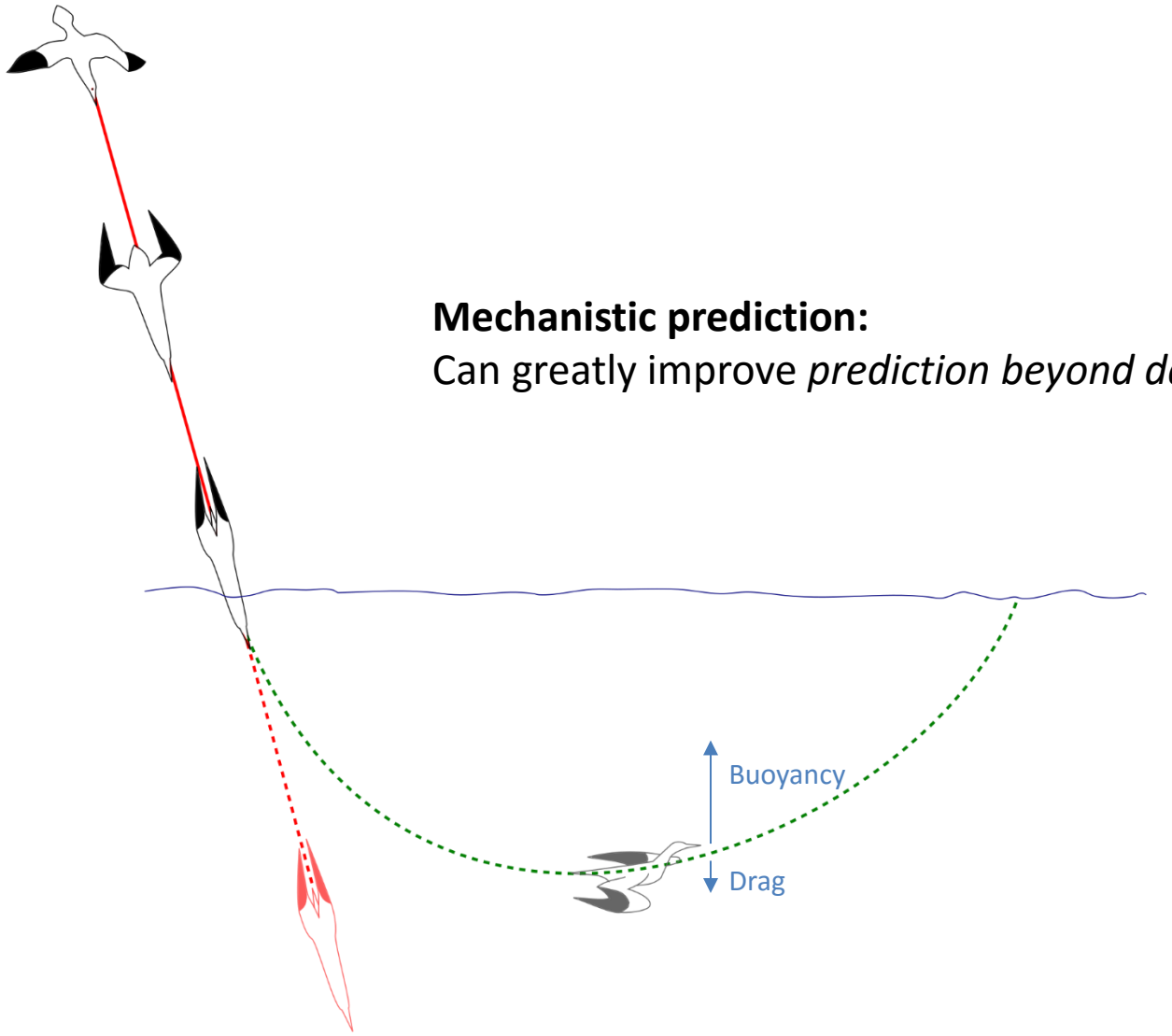
Mechanistic models:
Focus on mechanisms, aims at explanation or understanding. Often mathematical models like differential equations.

Here:
Newtonian mechanics, fluid properties

Other examples:
Lotka-Volterra predator-prey models
SIR model in epidemiology

inspired by Maino et al. (2016) Curr Opp Insect Sci

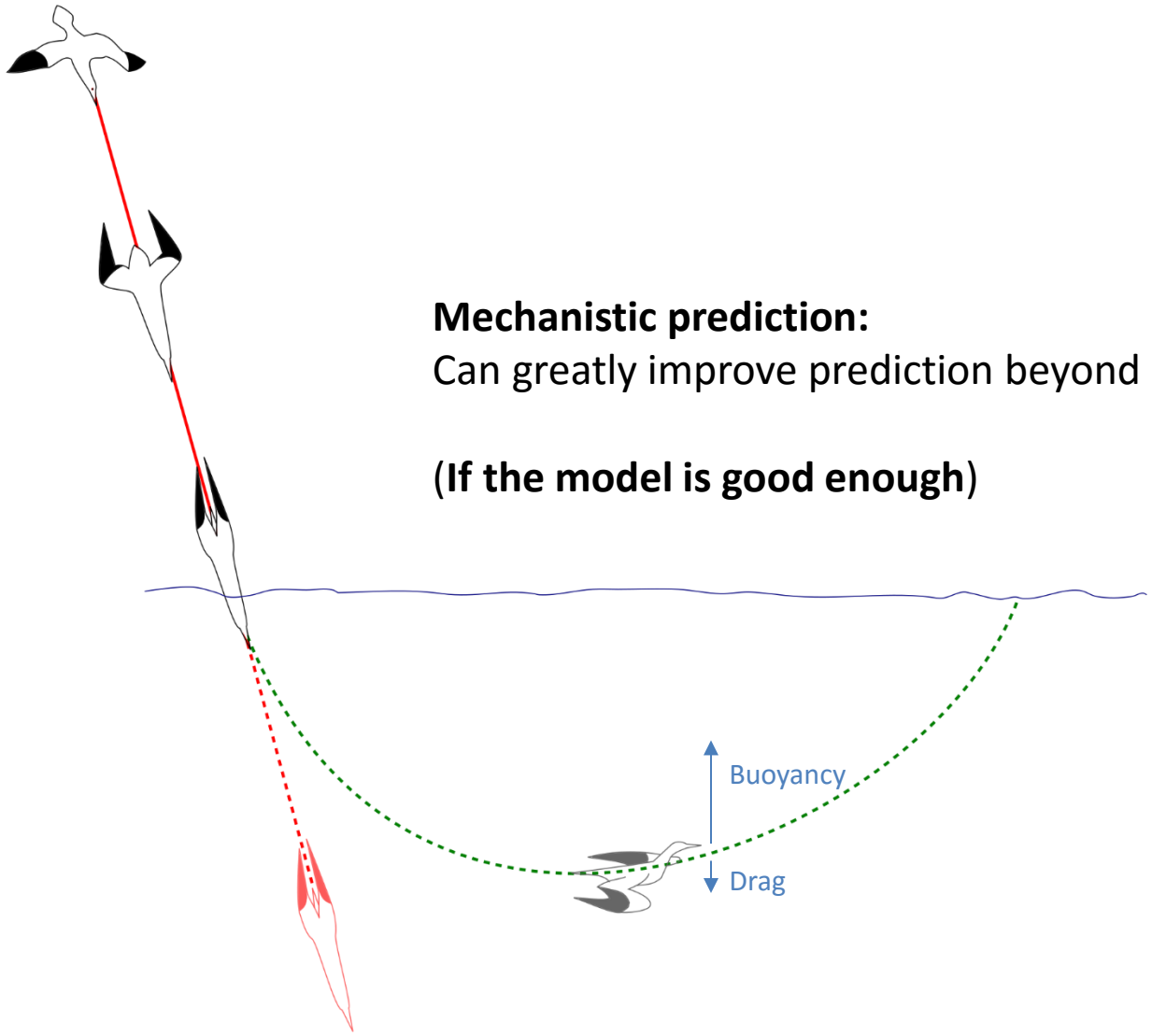
Models in ecology – an analogy



Mechanistic prediction:
Can greatly improve *prediction beyond data*

inspired by Maino et al. (2016) Curr Opp Insect Sci

Models in ecology – an analogy



Mechanistic prediction:
Can greatly improve prediction beyond data

(**If the model is good enough**)

inspired by Maino et al. (2016) Curr Opp Insect Sci

Tactical/Phenomenological

Strategic/Mechanistic

Describe patterns without elucidating mechanism

Prediction within range of data

Statistical models (regressions, etc.)

Focus on process/mechanisms

Explanation or understanding
Prediction outside range of data

Math models e.g. ODEs, PDEs
Individual/Agent based Models

BBS

WeBS

UK Bird Trends

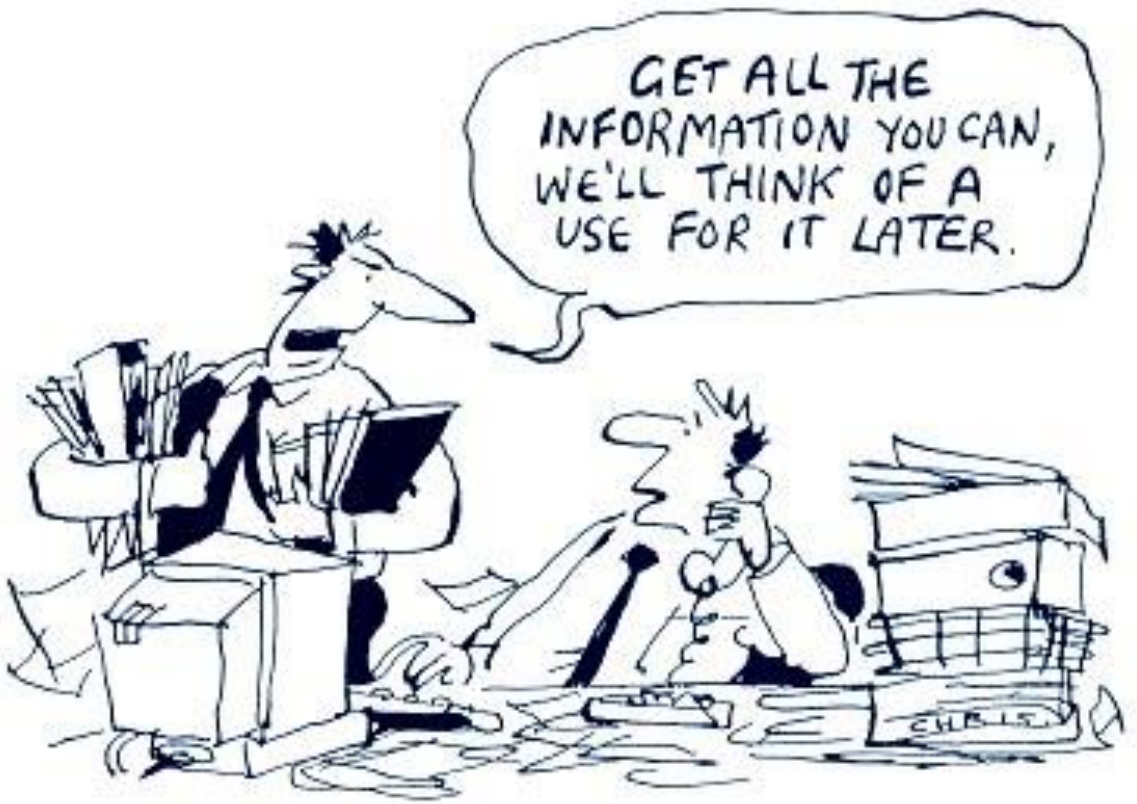
BTO Nest Record Scheme

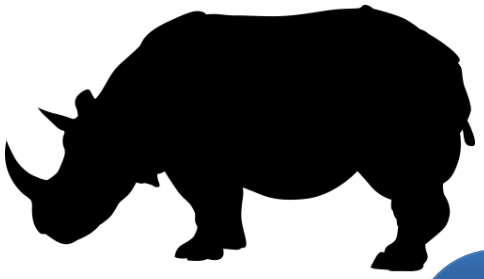
BBS

Ringing Schemes

Integrated Population Models

SURVEY DESIGN AND SAMPLING





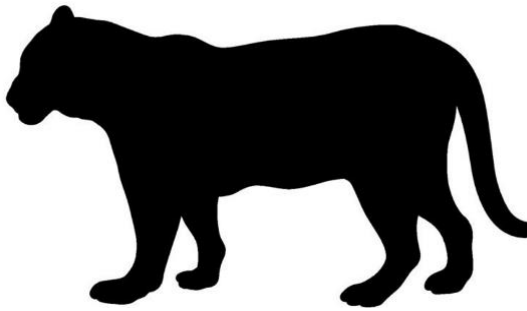
How many rhino horns are sold each year and where?

How does habitat affect blue tit nesting success?

How many tigers are in India?

How is sea temperature affecting coral reefs?

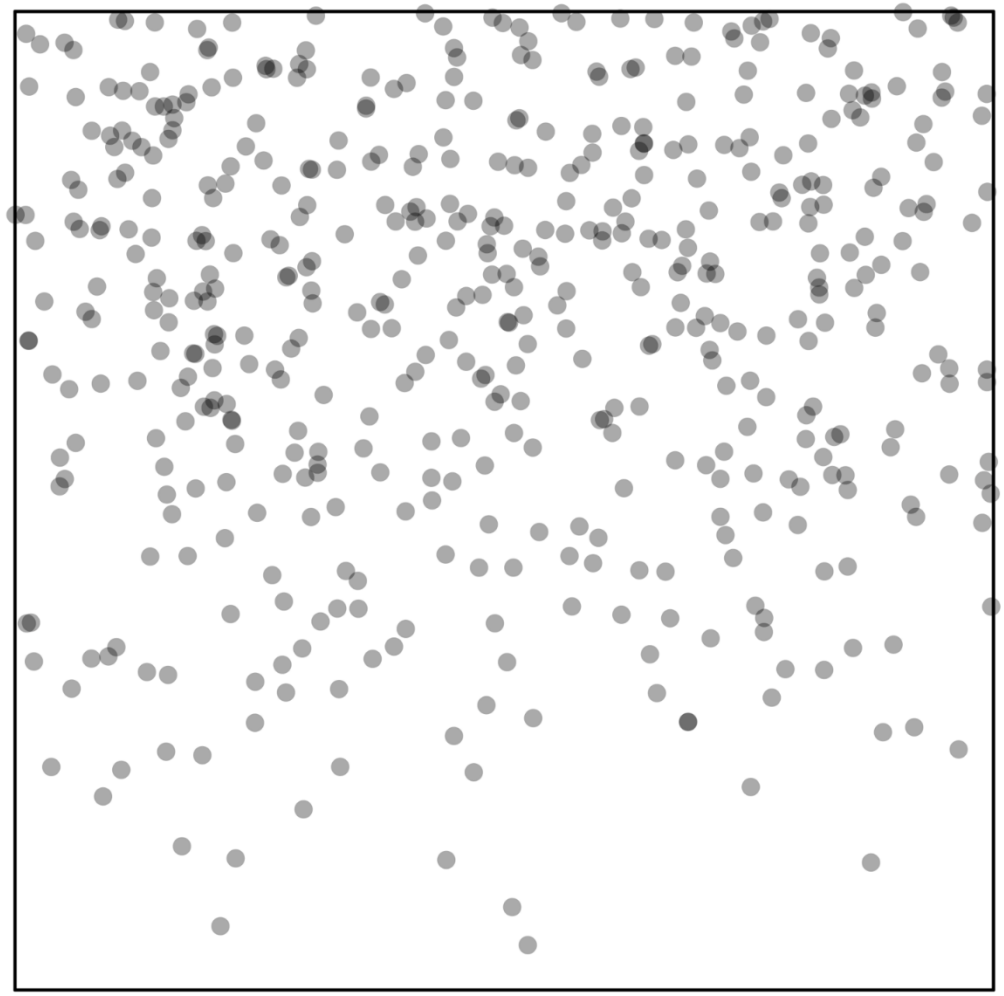
What's the maximum fragmentation that a tropical frog species can tolerate?



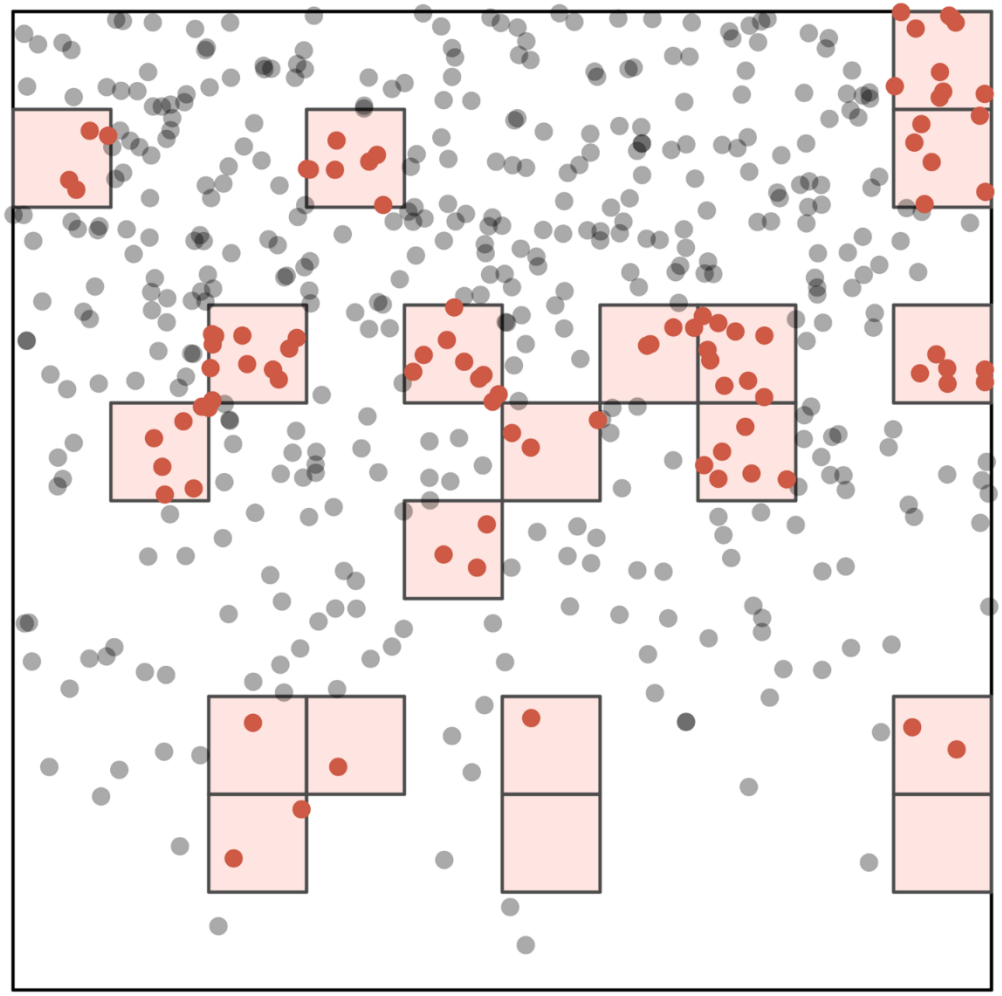
What data would you collect?

What compromises will you make?

Estimate number of trees

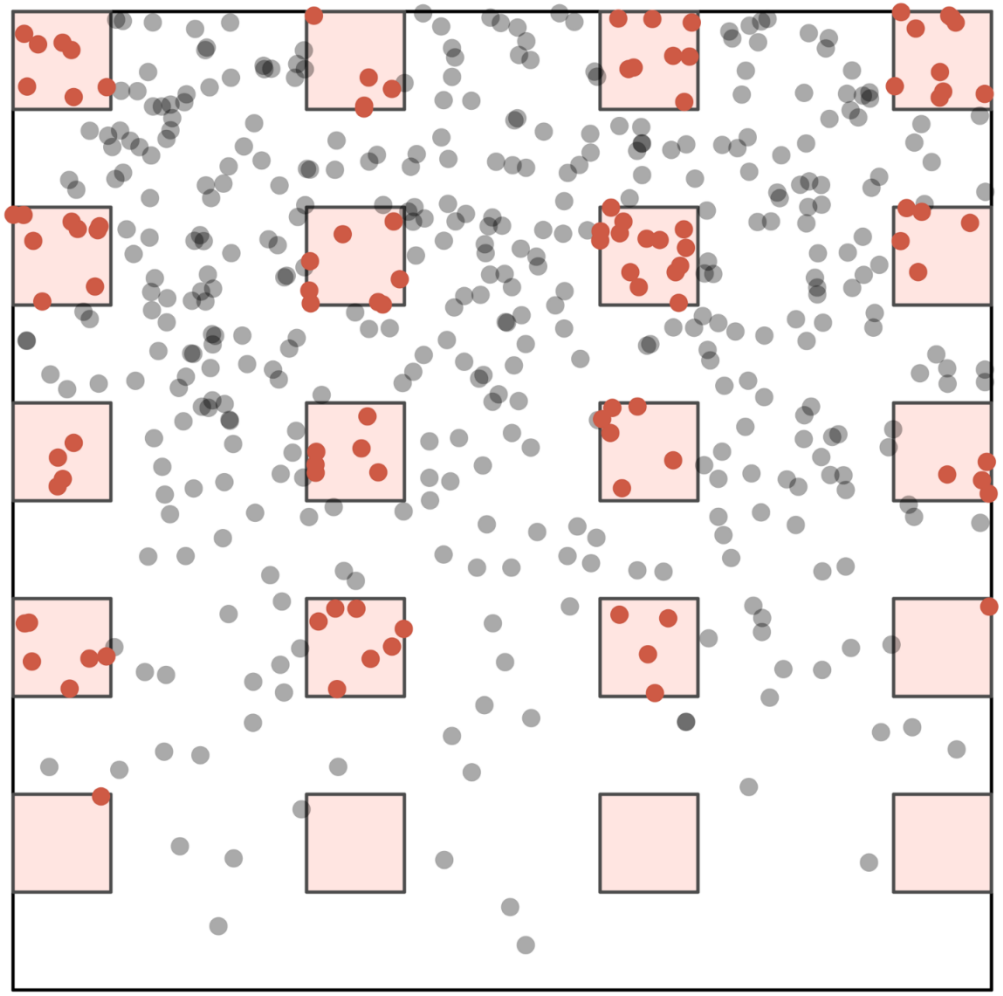


Estimate number of trees



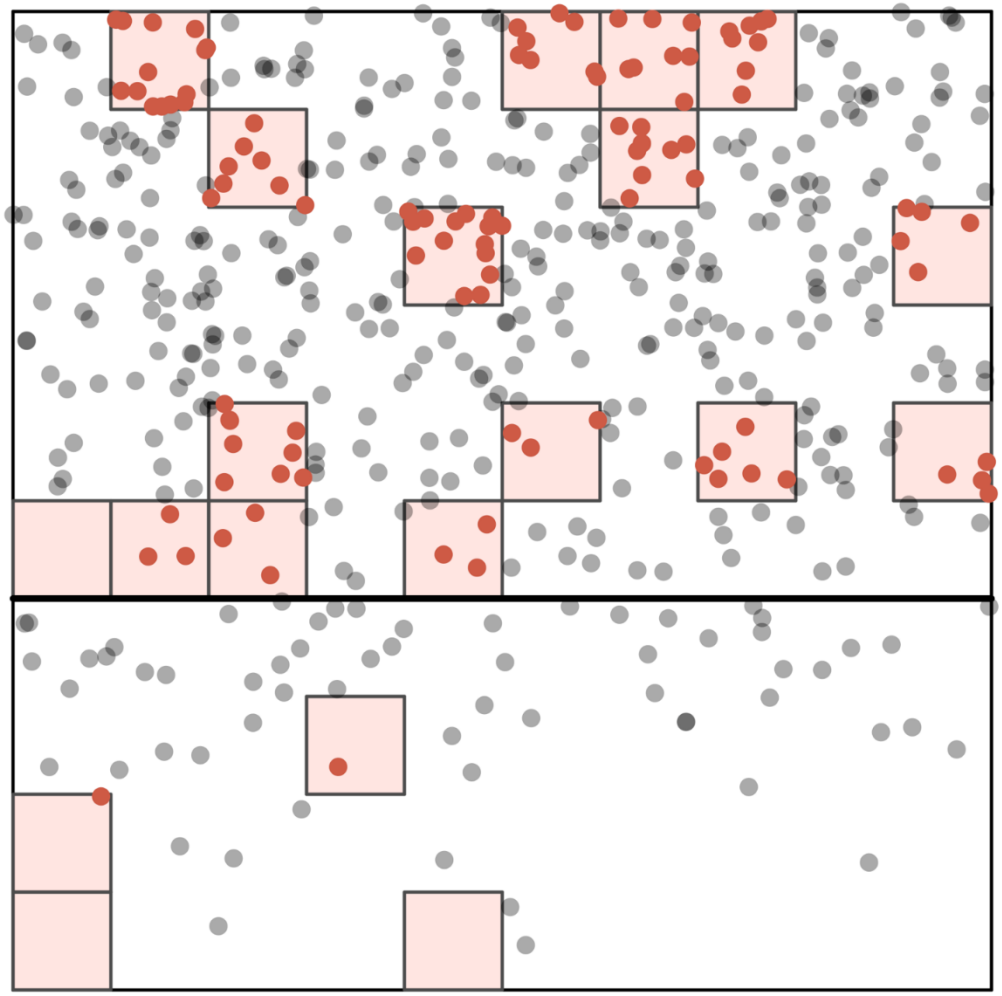
Random

Estimate number of trees



Random
Systematic

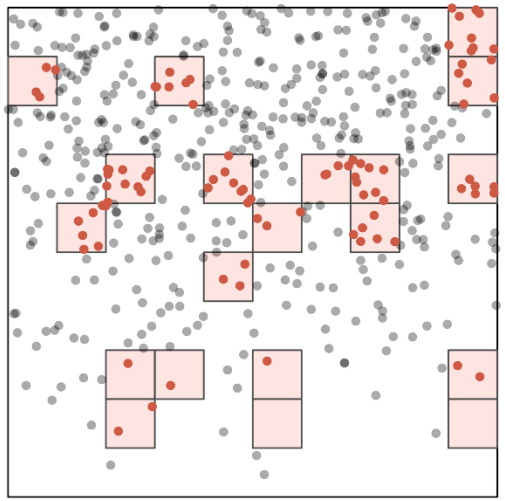
Estimate number of trees



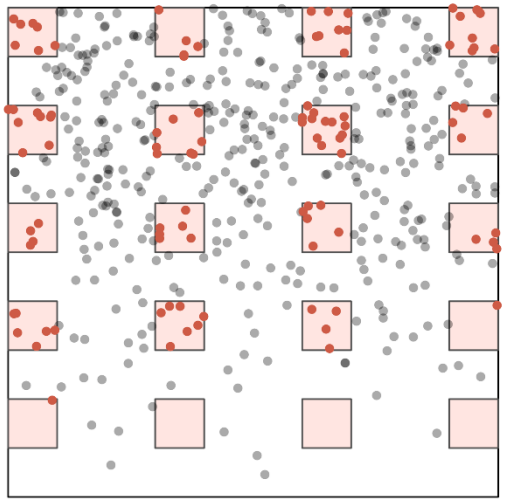
Random
Systematic
Stratified

Estimate number of trees

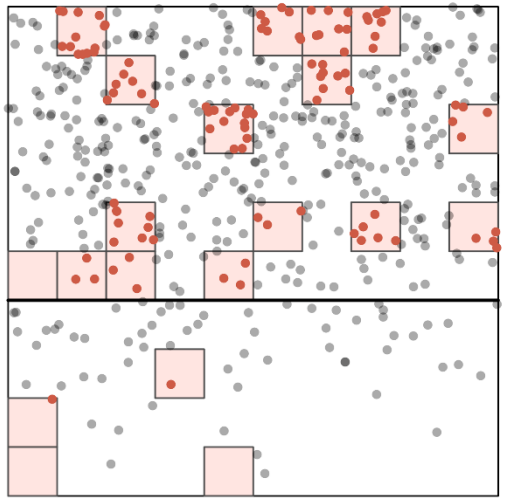
Target: REPRESENTATIVE SAMPLE



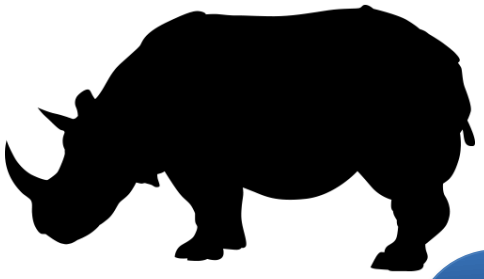
RANDOM



SYSTEMATIC



STRATIFIED



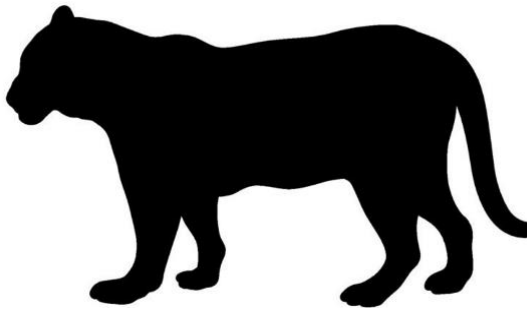
How many rhino horns are sold each year and where?

How does habitat affect blue tit nesting success?

How many tigers are in India?

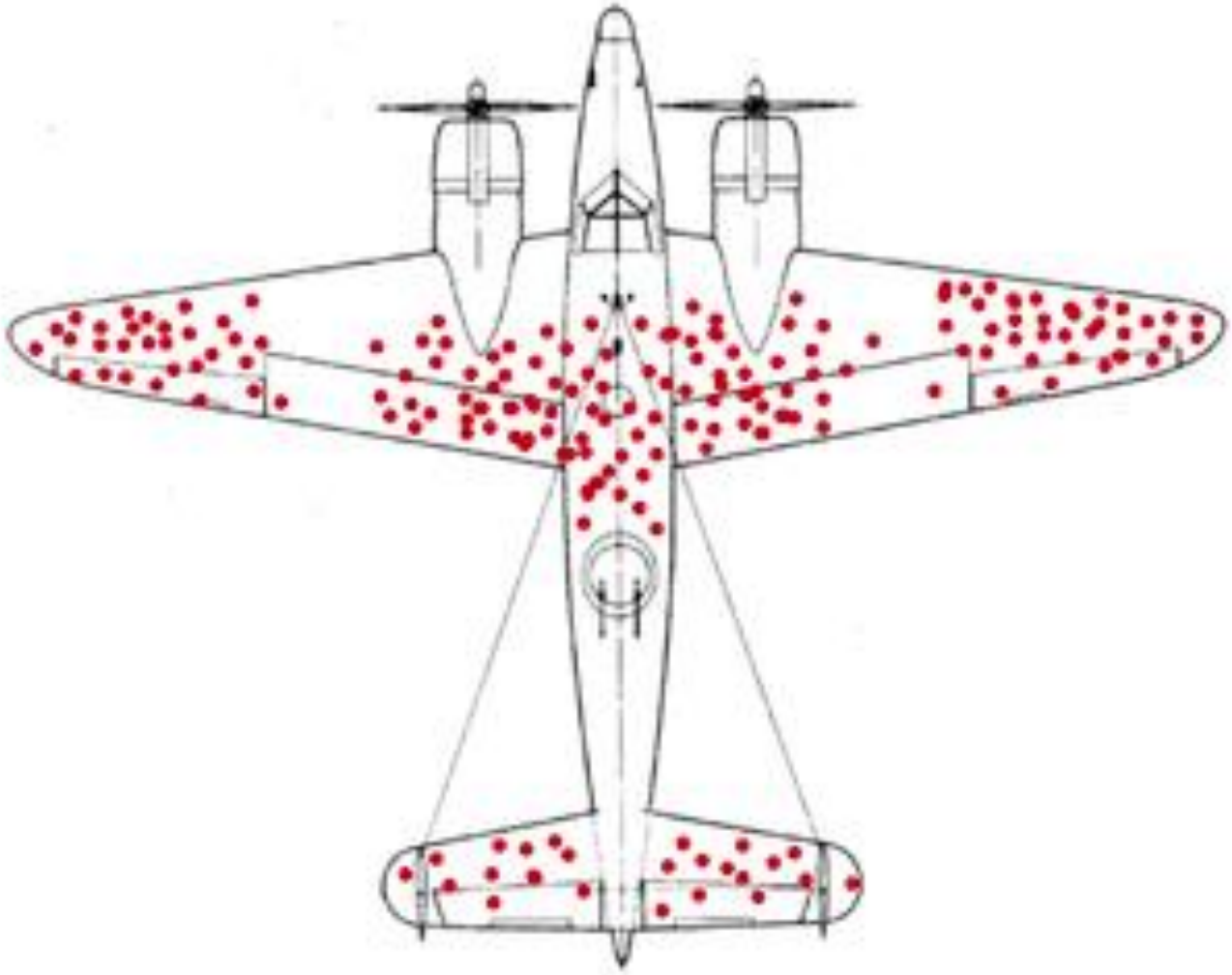
How is sea temperature affecting coral reefs?

What's the maximum fragmentation that a tropical frog species can tolerate?

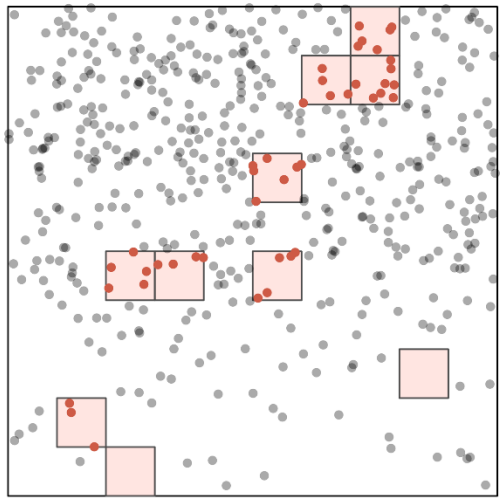


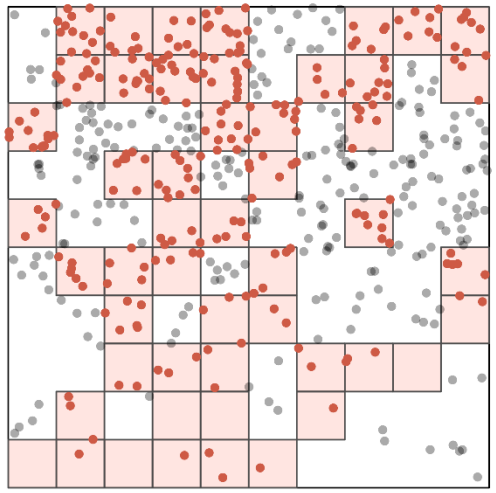
- Is your sample **representative**?
- Where do your results apply?

Can you actually observe what matters?



Replication





How many samples do I need?

How many samples?

The power of **POWER ANALYSIS**

1. State what you want to estimate

Number of trees in area within 20% of true number

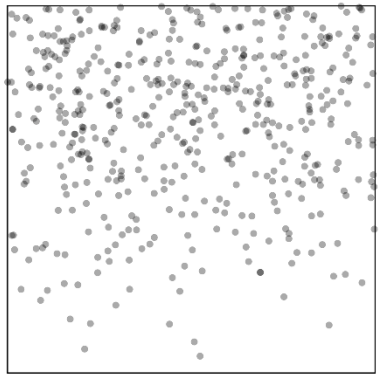
How many samples?

The power of **POWER ANALYSIS**

1. State what you want to estimate

2. Simulate a realistic version of what you want to estimate

Number of trees in area within 20% of true number

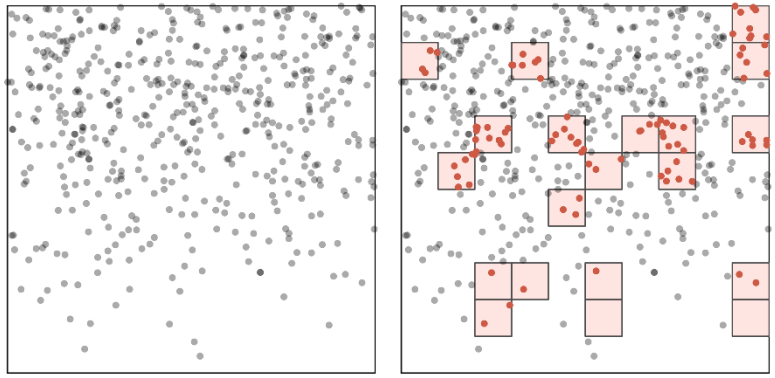


How many samples?

The power of **POWER ANALYSIS**

1. State what you want to estimate
2. Simulate a realistic version of what you want to estimate
3. Simulate sampling the population

Number of trees in area within 20% of true number

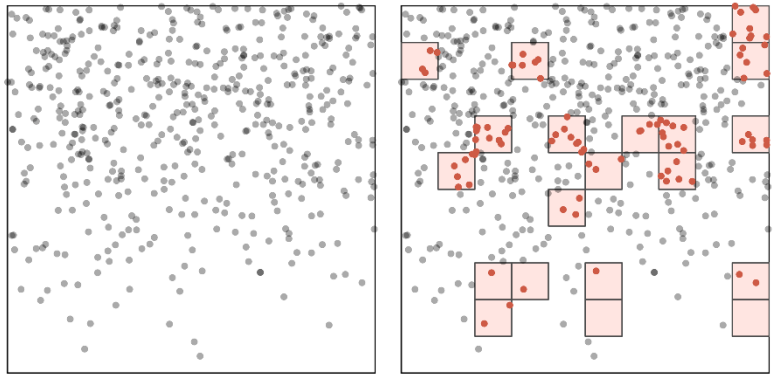


How many samples?

The power of **POWER ANALYSIS**

1. State what you want to estimate
2. Simulate a realistic version of what you want to estimate
3. Simulate sampling the population
4. Simulate analysing the sample

Number of trees in area within 20% of true number



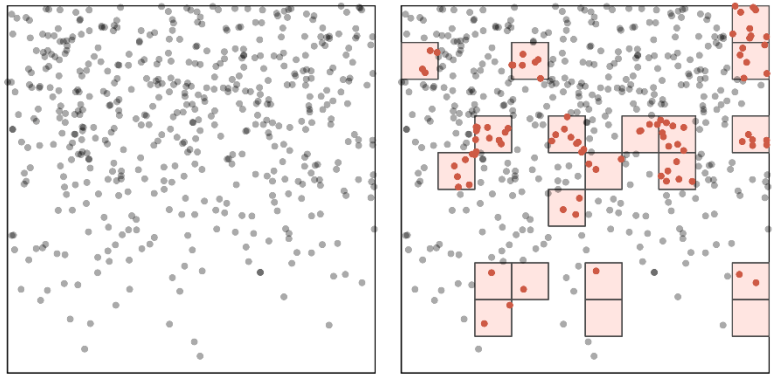
$$\frac{\text{Number observed}}{\text{Proportion of area sampled}}$$

How many samples?

The power of **POWER ANALYSIS**

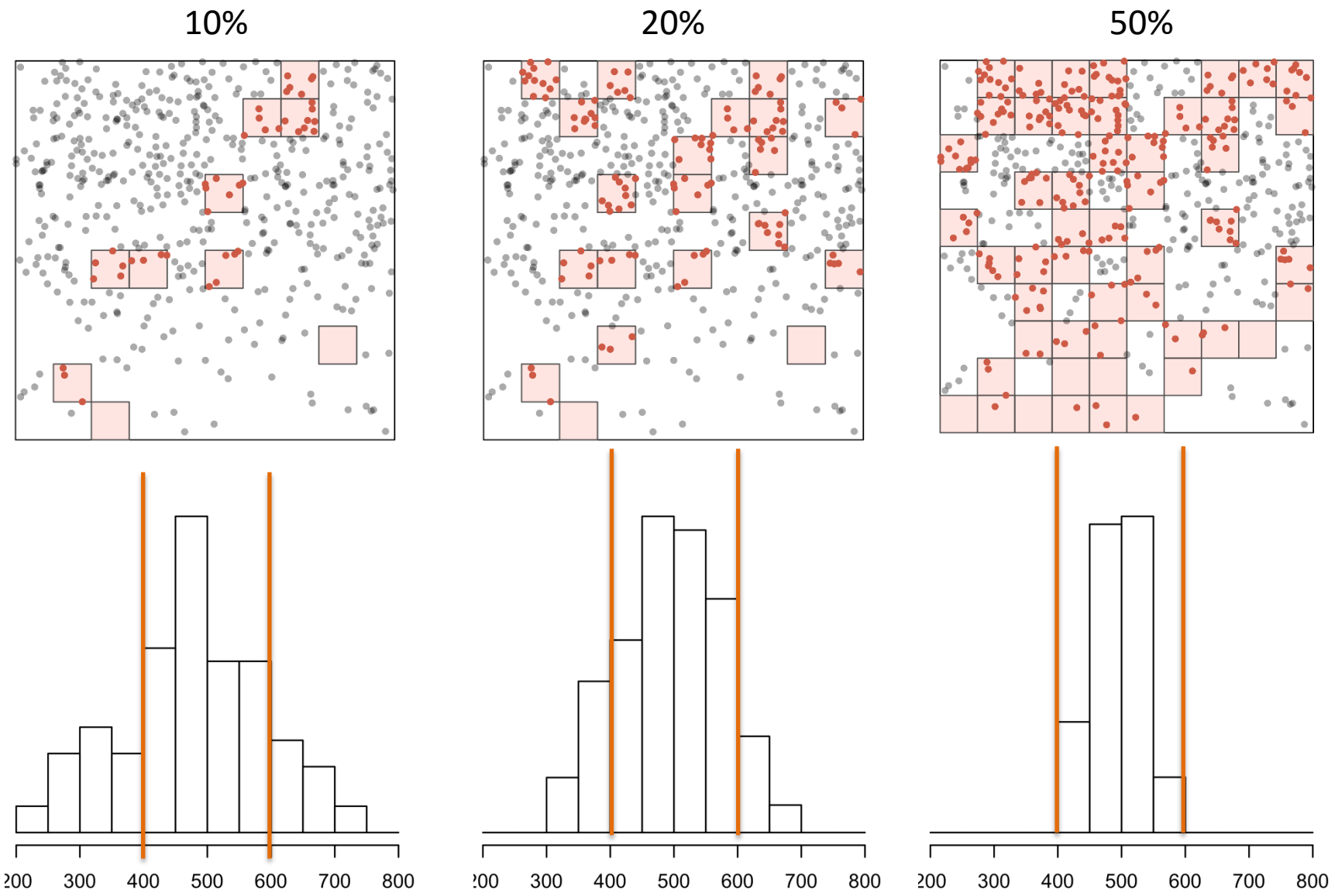
1. State what you want to estimate
2. Simulate a realistic version of what you want to estimate
3. Simulate sampling the population
4. Simulate analysing the sample
5. Calculate *how often* you meet your target

Number of trees in area within 20% of true number



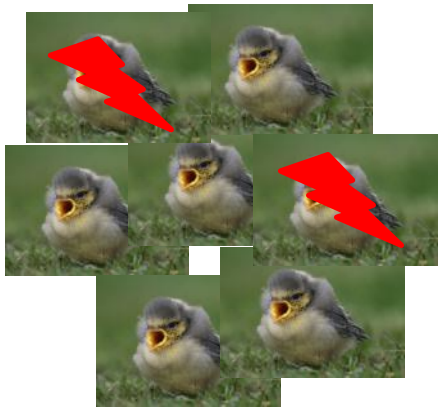
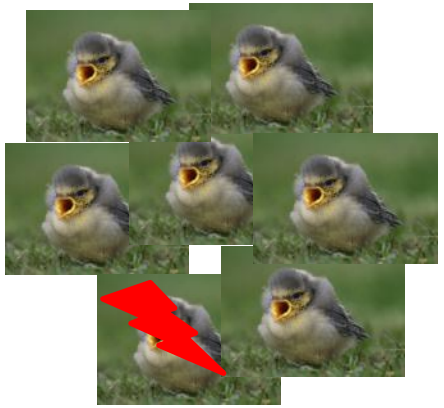
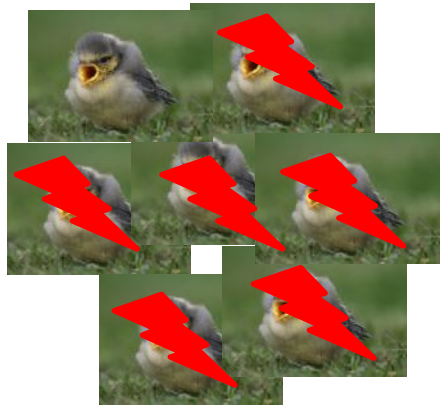
$$\frac{\text{Number observed}}{\text{Proportion of area sampled}}$$

Estimating number of trees



What is the unit of independence?



	Extra food	No extra food
Survive to fledging	11	1
Dead	3	6

21 chicks in 3 nest boxes. What's the sample size?

Percentage data

10 x 10 grid in quadrat

1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

What's the sample size?

Survey design and statistical analysis for conservation science

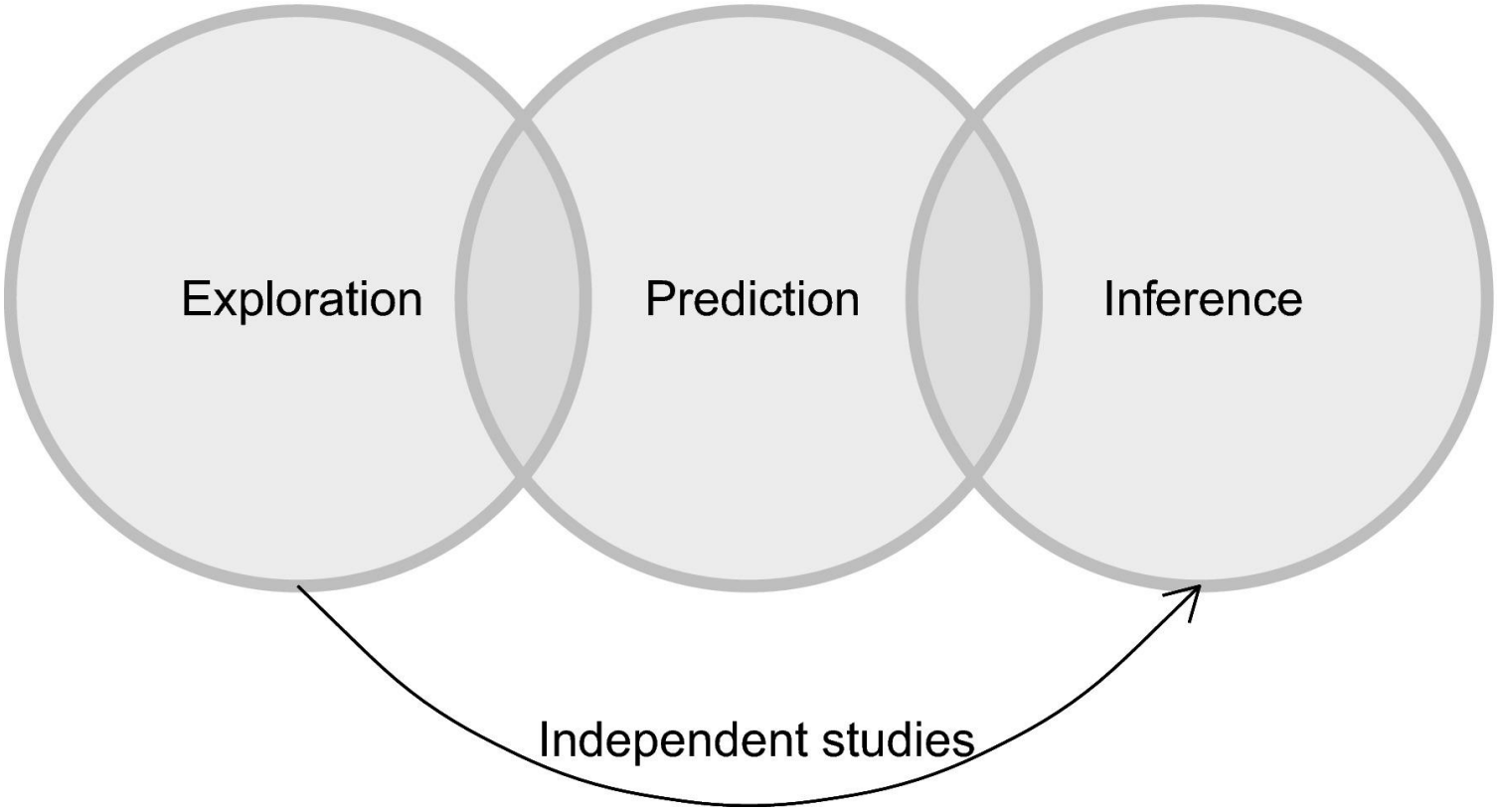
SURVEY DESIGN AND SAMPLING

1. REPRESENTATIVE samples
2. STRATIFICATION for sampling design
3. POWER ANALYSIS to assess sampling design
4. PSEUDO-REPLICATION in sampling

STATISTICAL ANALYSIS

COMMUNICATING RESULTS

STATISTICAL ANALYSIS



Tredennick et al. 2021, Ecology, https://doi.org/10.1002/ecy.3336

What is your ecological goal?

DESCRIBE a pattern

EXPLAIN a pattern

PREDICT an outcome

How many rhino horns are sold each year and where from?

How does habitat affect blue tit nesting success?

How many tigers are in India?

How is sea temperature affecting coral reefs?

What's the maximum fragmentation that a tropical frog species can tolerate?

What is your ecological goal?

DESCRIBE a pattern

EXPLAIN a pattern

PREDICT an outcome

How many rhino horns are sold each year and where from?

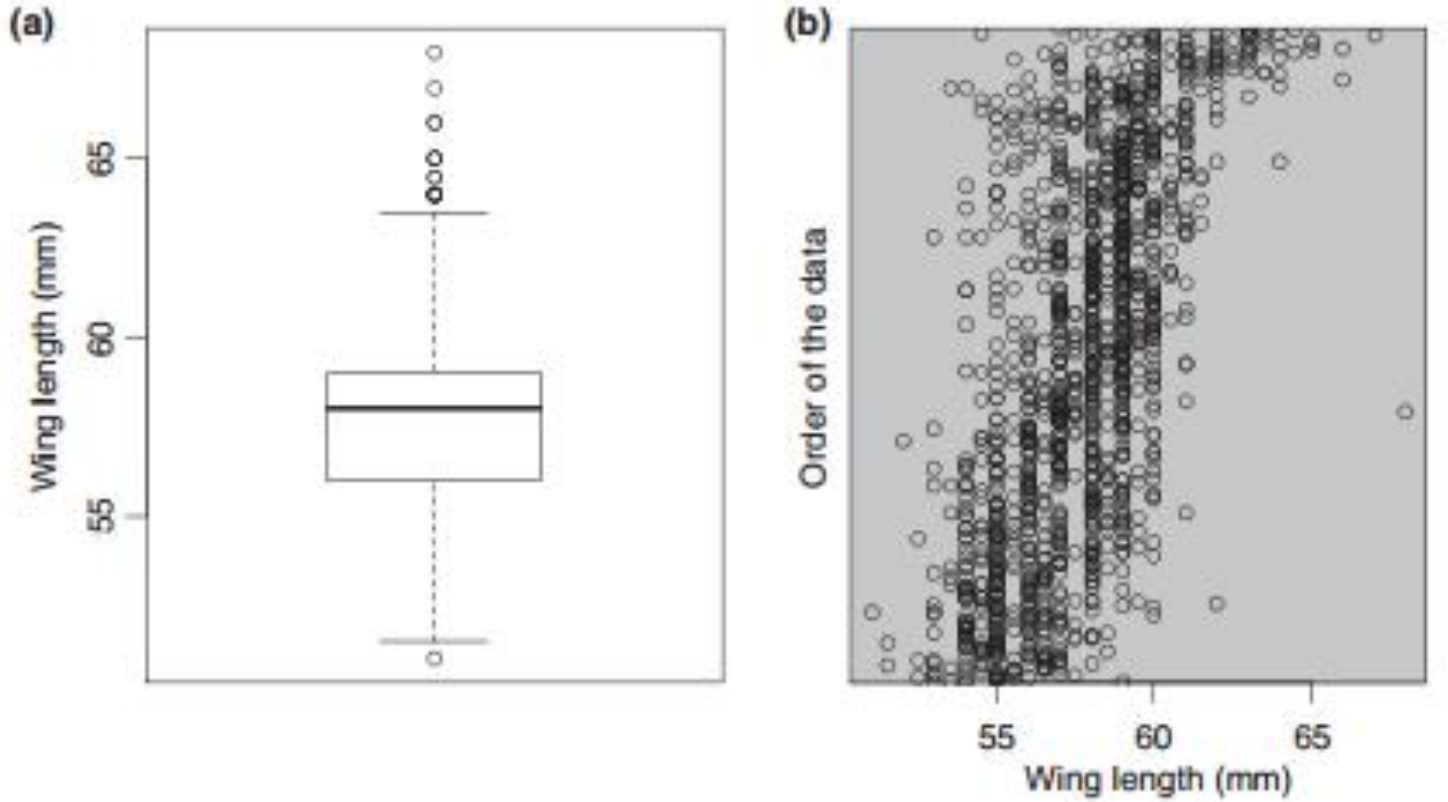
How does habitat affect blue tit nesting success?

How many tigers are in India?

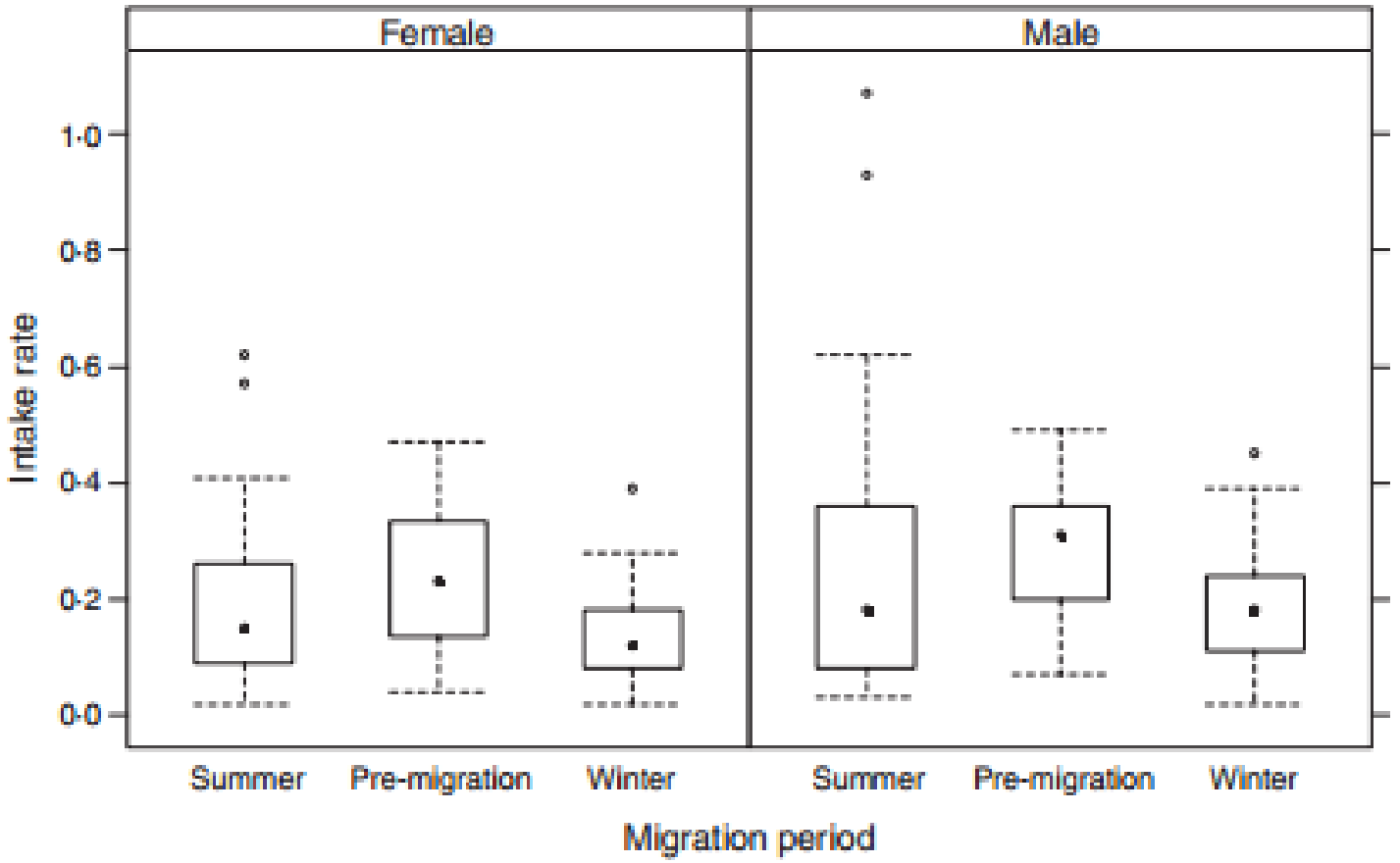
How is sea temperature affecting coral reefs?

What's the maximum fragmentation that a tropical frog species can tolerate?

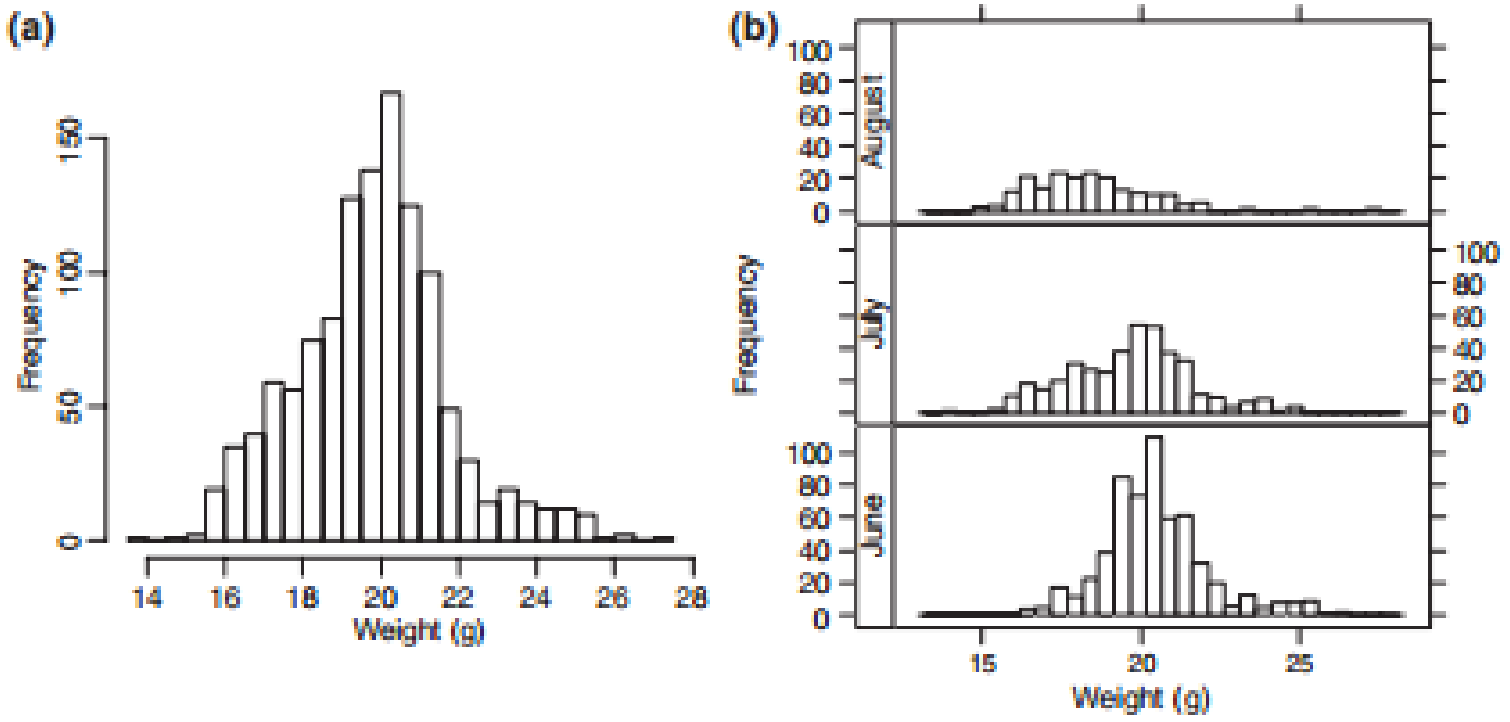
Look at your data!

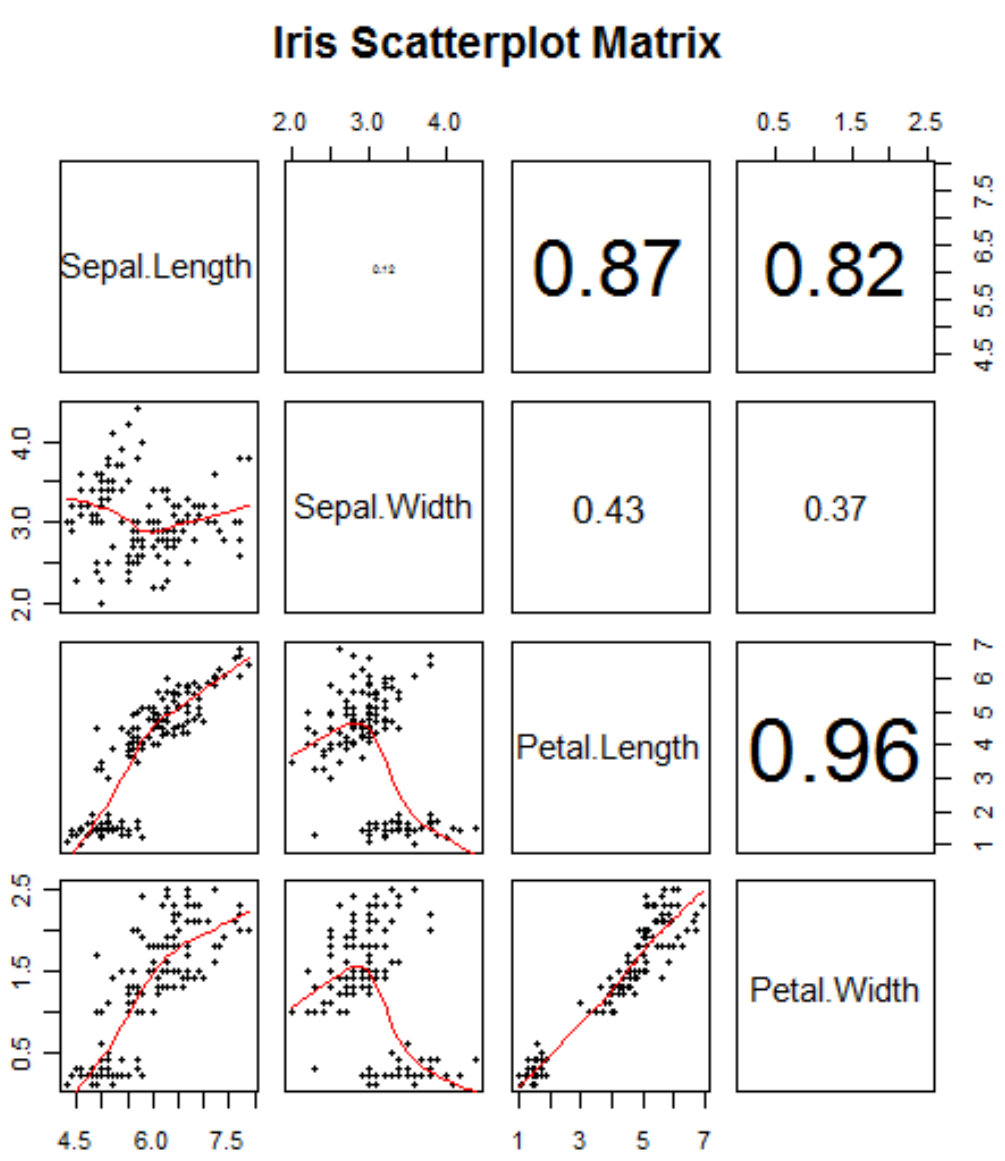


Look at your data!



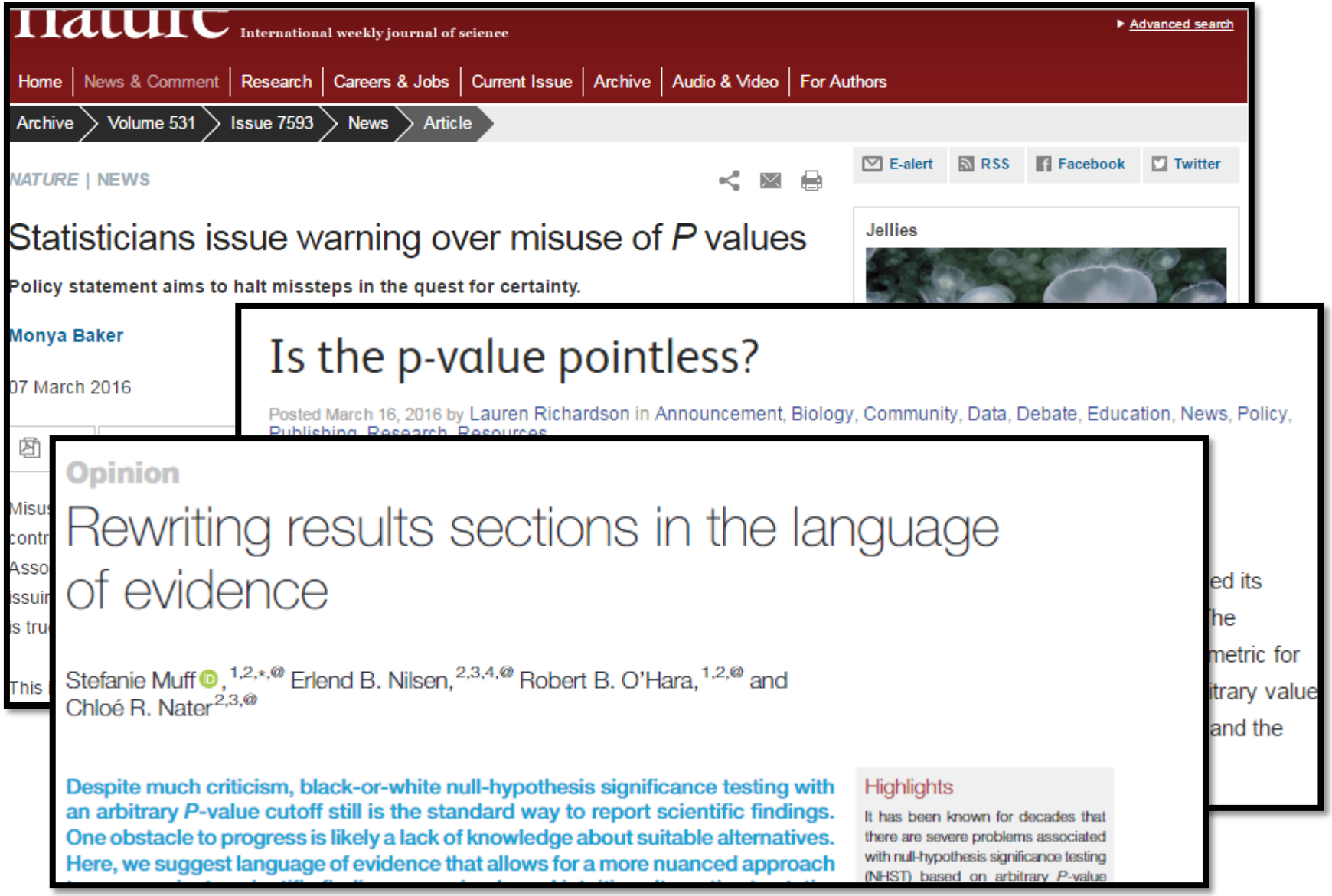
Look at your data!




Iris Scatterplot Matrix
Sepal.Length
Sepal.Width
Petal.Length
Petal.Width
0.87
0.82
0.43
0.37
0.96
2.0
3.0
4.0
0.5
1.5
2.5
4.5
5.5
6.5
7.5
1
2
3
4
5
6
7
4.5
6.0
7.5

Inference - Which are significant?

	est	SE	Z	P-value	significance
Diurnal temp range	0.009693	0.0077	1.26	0.2069	
Precipitation	0.002372	0.0010	-2.49	0.086	
Temperature	0.546524	0.0650	8.41	<0.0001	***
Vapour pressure	-1.089962	0.1595	-3.834	0.0453	*
Frost days	-0.010056	0.0031	-6.786	0.0012	***
Aspect	1.54897	0.9873	-5.439	0.0556	



Muff et al. 2022, https://doi.org/10.1016/j.tree.2021.10.009

What is a p-value?

p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true

This is not the probability that the null hypothesis is true:

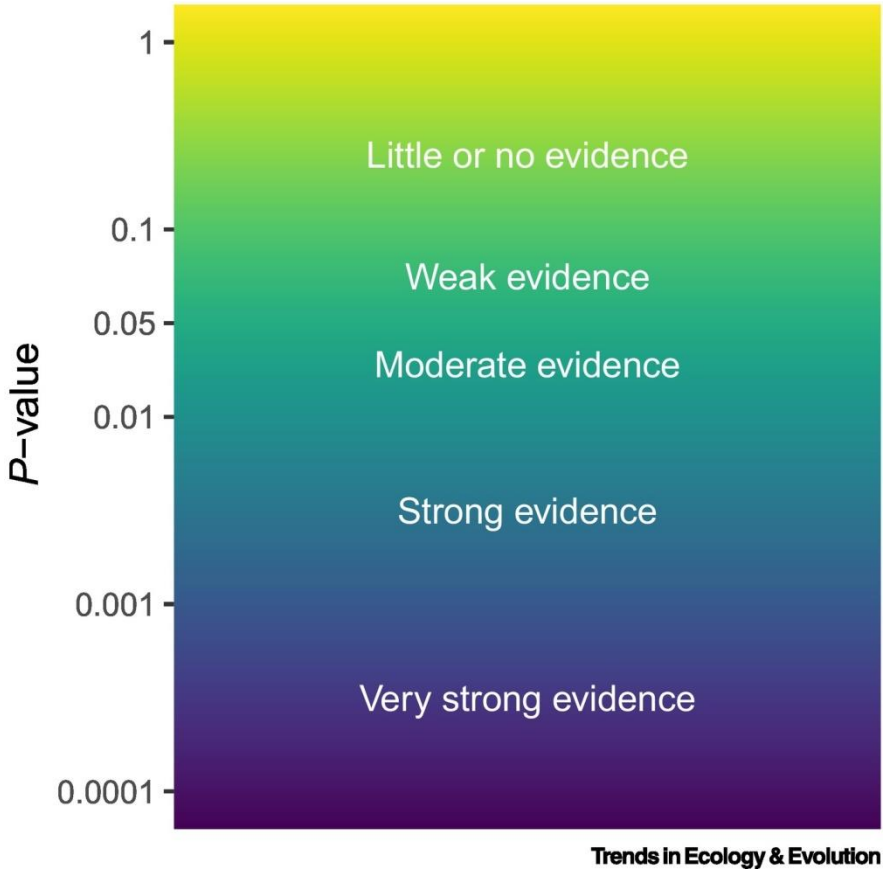
The probability of observing a result given that some hypothesis is true
is not equivalent to
the probability that a hypothesis is true given that some result has been observed.

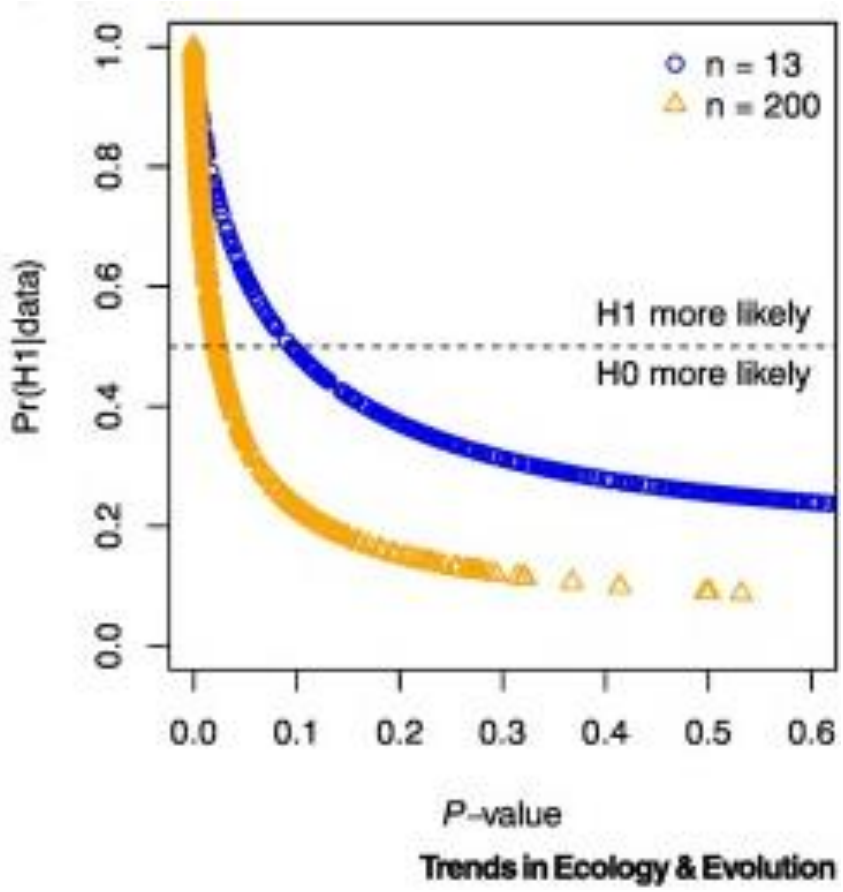
What is a p-value?

By itself, a *p*-value does not provide a good measure of evidence regarding a model or hypothesis

Whether and where to discretizing the p-value into significant/not-significant (e.g. at 0.05) is completely arbitrary!

Statistical significance is continuous, but also context dependent

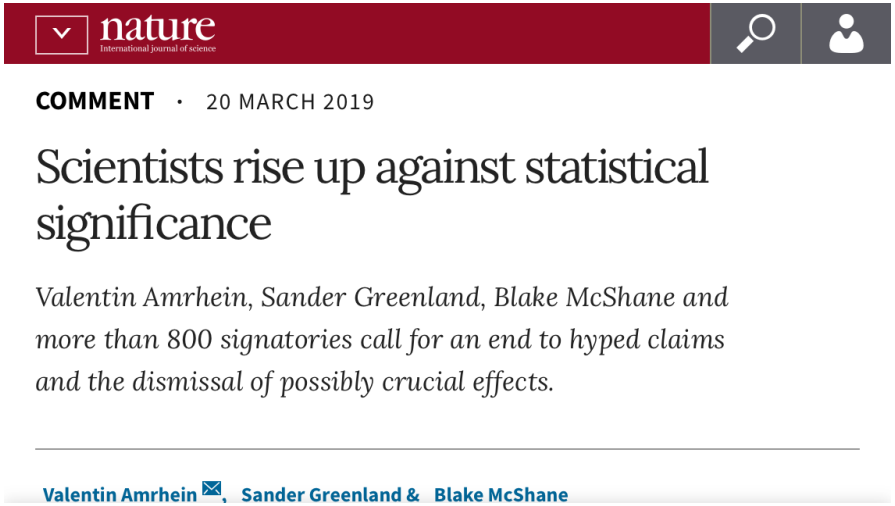




Muff et al. 2022, https://doi.org/10.1016/j.tree.2021.10.009
Hartig et al. 2022, https://doi.org/10.1016/j.tree.2022.02.011

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

3. *Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.*



nature
International journal of science

COMMENT · 20 MARCH 2019

Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Valentin Amrhein ✉, Sander Greenland & Blake McShane

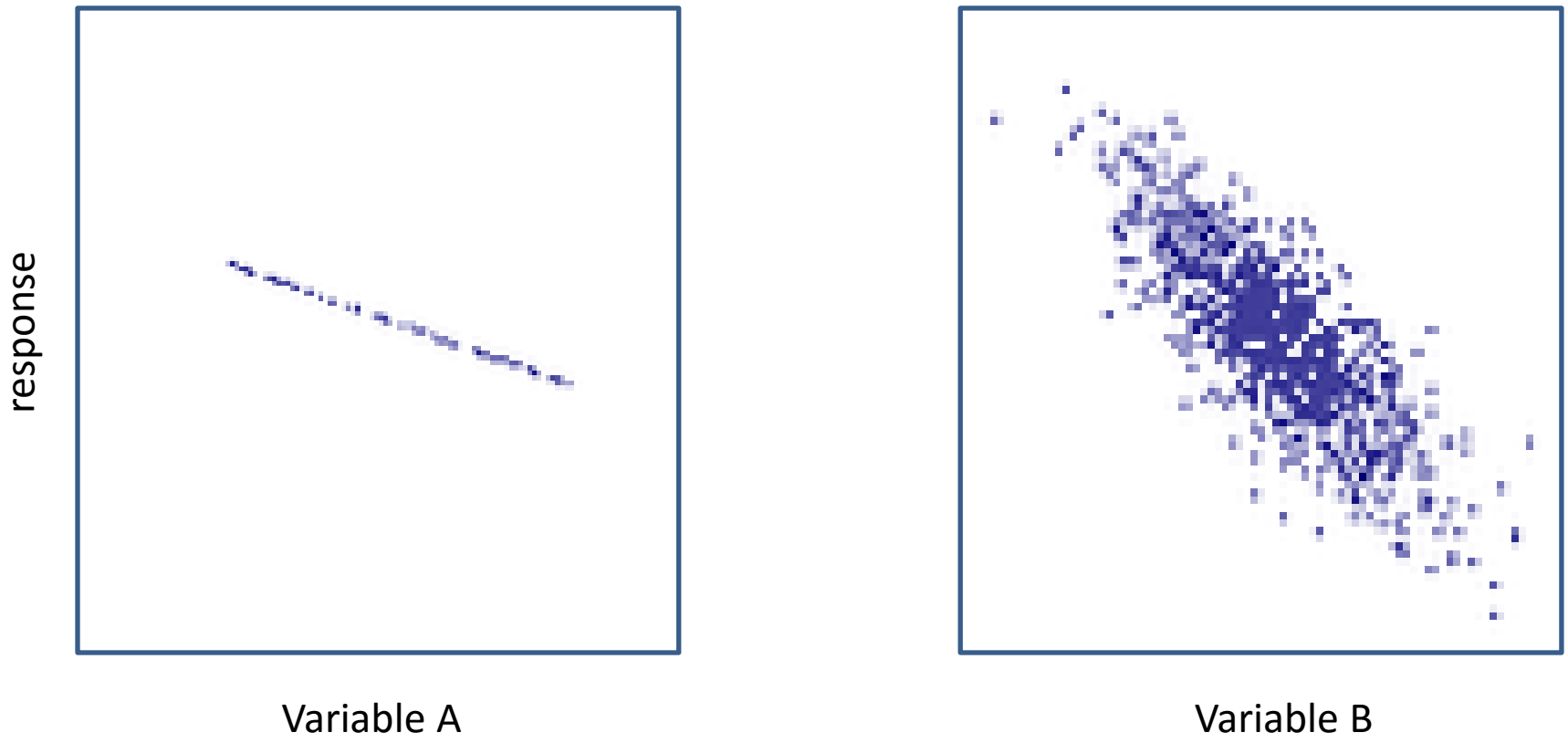
All biologists should be ultimately interested in **biological importance**,

which may be assessed using the **magnitude of an effect**,

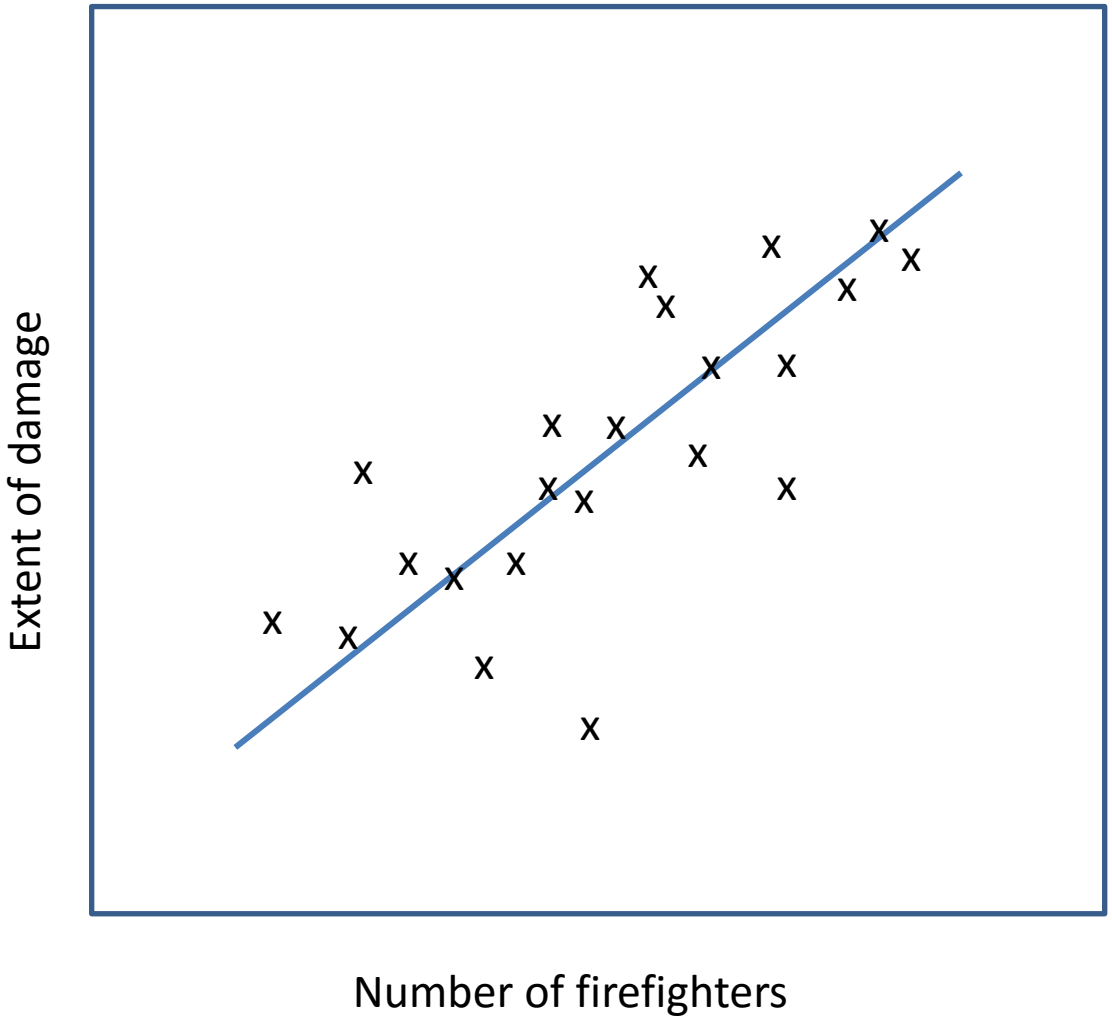
but not its **statistical significance**.

It is almost always more useful for interpretation to look at / report **effect estimates and their uncertainty range** (e.g. confidence interval), than the p-value alone.

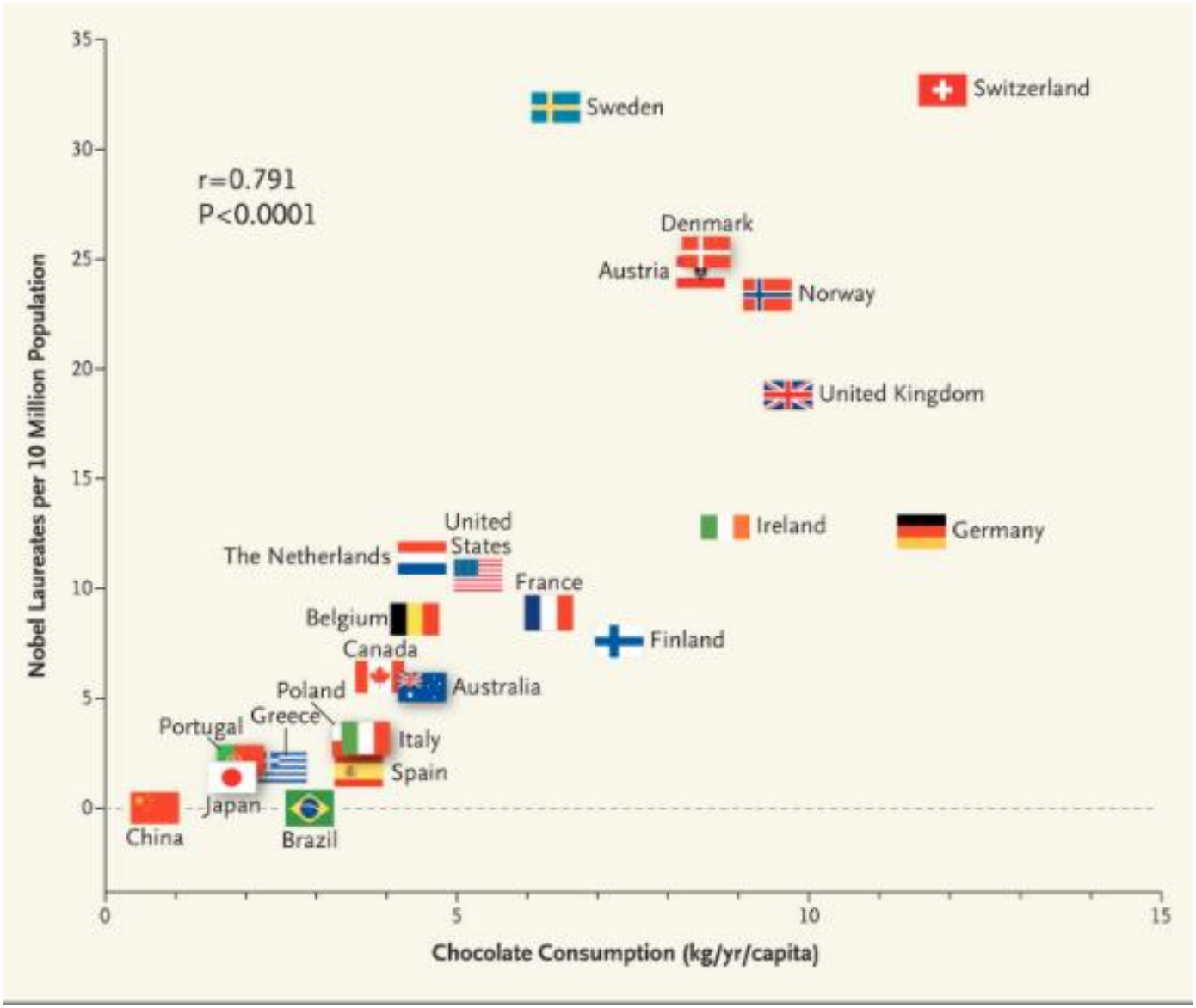
Nakagawa & Cuthill (2007)
Effect size, confidence interval and statistical significance: a practical guide for biologists
Biological Reviews 82(4):591 – 605. https://doi.org/10.1111/j.1469-185X.2007.00027.x


response
Variable A
Variable B

correlation *is not equal to* **causation**



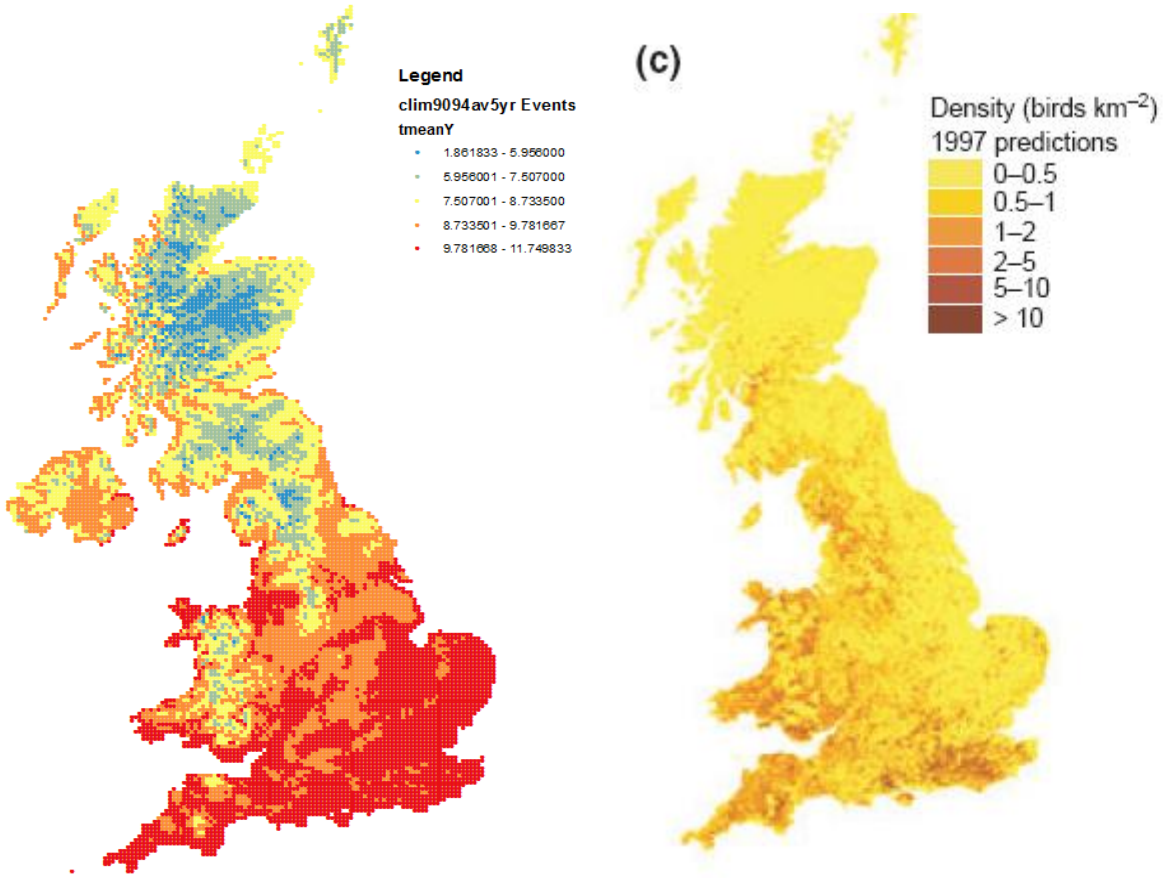
Conclusion: firefighters cause damage



"It remains to be determined whether the consumption of chocolate is the underlying mechanism for the observed association with improved cognitive function."

Messerli (2012) *New England Journal of Medicine*

Climate envelope models



Renwick et al. (2011) *Diversity & Distributions*

Thinking about uncertainty

- How big is the uncertainty?
- How can you best communicate the uncertainty?
- What are the consequences of concluding the wrong thing?
- How much uncertainty are you comfortable with?
- How much uncertainty are policy makers comfortable with?

Survey design and statistical analysis for conservation science

SURVEY DESIGN AND SAMPLING

1. REPRESENTATIVE samples
2. STRATIFICATION for sampling design
3. POWER ANALYSIS to assess sampling design
4. PSEUDO-REPLICATION in sampling

STATISTICAL ANALYSIS

5. DESCRIPTION or EXPLANATION
6. SIGNIFICANCE is continuous
7. Separate SIGNIFICANCE and EFFECT SIZE
8. Separate CORRELATION AND CAUSATION
9. Consider UNCERTAINTY

COMMUNICATING RESULTS

COMMUNICATING ANALYSES AND RESULTS

- Methods:
  - Always fully describe the analysis you are conducting
  - Write out the exact model you are fitting
  - Provide details on the software you are using
  - If you need to be terse in the main manuscript, put a full description in an appendix/supplement

COMMUNICATING ANALYSES AND RESULTS

Describe the model structures fully

Name & cite the software

Analysis

All statistical analyses were carried out in *R v. 3.1* (R Development Core Team, 2013). We used generalized linear models with Poisson-distributed errors to estimate the species–area relationship by predator status for all islands (Supplementary Material 4). The models of species richness and island area assume there is a linear increase in the number of species as island area increases but in reality this is limited to the number of seabird species in the region. We assume that the maximum species richness is constrained by the number of species included in the analysis (n = 16). Model assumptions of a linear mean variance relationship and of spatially uncorrelated errors were checked using the dispersion test of the *AER* package in *R* (Kleiber & Zeileis, 2008), and spatial correlation tests of the *sp*, *ape* and *gstat* packages (Supplementary Material 4; Paradis et al., 2004; Pebesma, 2004; Bivand et al., 2013).

We considered three models for the species–area relationship by predator status, all being Poisson regressions of a semilog model (i.e. species count, S, as response, $\log_{10}(A)$ as predictor). The three models are all Poisson generalized linear models: $S \sim \text{Poisson}(\mu)$ with log link, such that $\log \mu(S) = \beta \boldsymbol{X}_i$, where $\boldsymbol{X}$ and β are the design matrix and the parameter vector, respectively.

Effect of log(area) only:

$$\log\mu(S) = \beta_0 + \beta_1 \log_{10} A \quad (1)$$

Additive effects of log(area) and predator status (i.e. separate intercept per predator status but same slope):

$$\log\mu(S) = \beta_0 + \beta_1 \log_{10} A + \beta_2 Predator \quad (2)$$

Interaction of log(area) and predator status (i.e. separate slope and intercept per predator status):

$$\log\mu(S) = \beta_1 Predator + \beta_2 \log_{10} A \times Predator \quad (3)$$

Describe the model and its assumptions (using words)

Equations or pseudocode can be less ambiguous than words

Borrelle et al. 2016, *Oryx*, http://doi.org/10.1017/S0030605316000880

COMMUNICATING ANALYSES AND RESULTS

- Results:
  - Always provide full model summary tables (if need be in an appendix)
  - BUT key to a good paper/report is translating statistical summaries into human language and images

COMMUNICATING RESULTS

– Always provide full model summary tables (if need be in an appendix)

TABLE 2 Generalized linear model parameter estimates for the species–area relationship by predator status. Parameter values are given on the link scale.

Parameter	Parameter estimate ± SE	P
Predator_statusCleared	1.462 ± 0.160	< 0.001
Predator_statusInvaded	−0.480 ± 0.285	0.092
Predator_statusUninvaded	0.890 ± 0.156	< 0.001
log(ha):Predator_statusCleared	0.038 ± 0.036	0.295
log(ha):Predator_statusInvaded	0.249 ± 0.096	0.009
log(ha):Predator_statusUninvaded	0.259 ± 0.062	< 0.001

Borrelle et al. 2016, *Oryx*, http://doi.org/10.1017/S0030605316000880

COMMUNICATING RESULTS

Restate the purpose / ecological question of the analysis

TABLE 2 Generalized linear model parameter estimates for the species–area relationship by predator status. Parameter values are given on the link scale.

Parameter	Parameter estimate ± SE	P
Predator_statusCleared	1.462 ± 0.160	< 0.001
Predator_statusInvaded	−0.480 ± 0.285	0.092
Predator_statusUninvaded	0.890 ± 0.156	< 0.001
log(ha):Predator_statusCleared	0.038 ± 0.036	0.295
log(ha):Predator_statusInvaded	0.249 ± 0.096	0.009
log(ha):Predator_statusUninvaded	0.259 ± 0.062	< 0.001

P-values / significance measures

Use human-readable descriptions of your variables

Parameter estimates and their uncertainties

Borrelle et al. 2016, *Oryx*, http://doi.org/10.1017/S0030605316000880

COMMUNICATING RESULTS

– BUT good communication means translating statistical summaries into human language!

Restate the model

The final model for the species–area relationship included separate slopes and intercepts by predator status. No significant overdispersion was detected (overdispersion test: dispersion parameter = 0.99; P = 0.53). Moran's I (I = 0.1, P = 0.011) indicated weak but statistically significant spatial autocorrelation in the residuals, and visual inspection of spatial residuals indicated that this was because the model overpredicted species richness for the islands of the inner Hauraki Gulf (Supplementary Fig. S1). Generalized linear model parameter estimates are in Table 2, and model predictions are illustrated in Fig. 2.

Translate parameter estimates back to words about ecology!

For uninvaded islands the intercept (i.e. expected species count on a 1 ha island) is 2.43 (95% CI 1.77–3.28; $\hat{\beta} = 0.89$, $P < 0.001$), and for a 10-fold increase in area the species count is expected to increase 1.81-fold (95% CI 1.36–2.39; $\hat{\beta} = 0.26$, $P < 0.001$). For cleared islands the intercept is 4.31 (95% CI 3.12–5.84; $\hat{\beta} = 1.46$, $P < 0.001$), and no significant relationship was found between species richness and area ($\hat{\beta} = 0.04$, P = 0.295). For invaded islands the intercept is 0.62 (95% CI 0.34–1.05; $\hat{\beta} = -0.48$, P = 0.092), and for a 10-fold increase in area the species count is expected to increase 1.77-fold (95% CI 1.14–2.71; $\beta = 0.25$, P = 0.009). Regardless of size, invaded islands did not have more than four species (Fig. 2).

Describe how you checked the model

Estimate and CI on interpretable scale are more useful than raw coeffcients/p-values

Borrelle et al. 2016, *Oryx*, http://doi.org/10.1017/S0030605316000880

COMMUNICATING RESULTS

– or images!

You can use model predictions to illustrate the key take-aways from your model

Label Axes and Symbols clearly
Use simple colours

Use the caption to summarise the findings. This could be improved here, e.g.:
"Species richness increases with increasing area, and is higher on uninvaded and cleared islands"

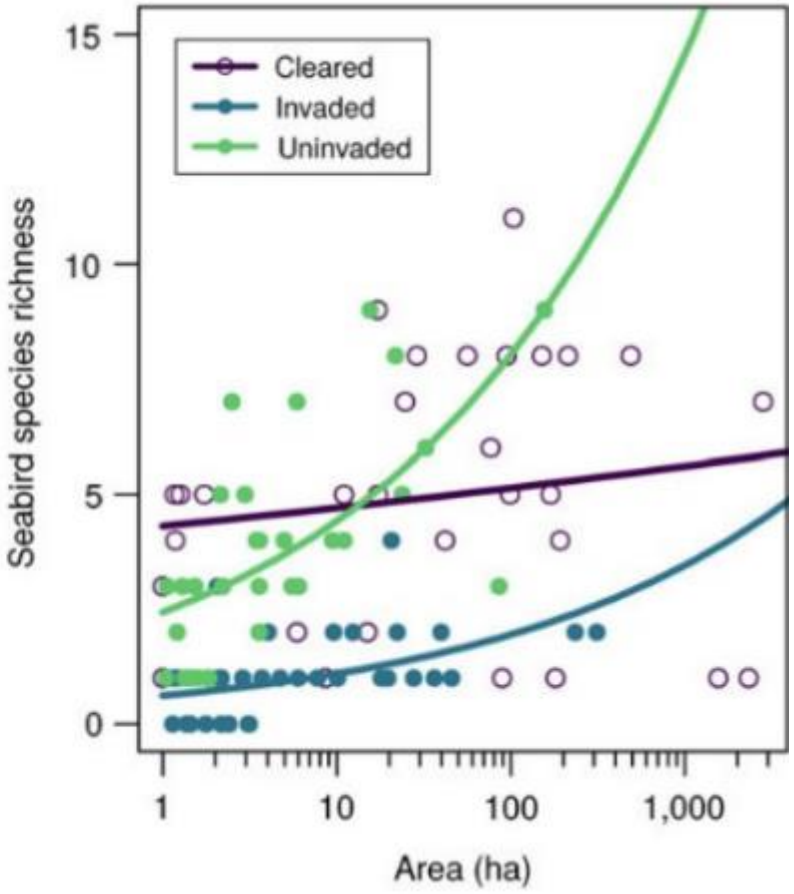


FIG. 2 Observed species richness and generalized linear model predictions as a function of island area for uninvaded (intercept = 2.43, slope = 0.26), cleared (intercept = 4.31, slope = 0.04), and invaded islands (intercept = 0.62, slope = 0.25) in the Hauraki Gulf (Fig. 1). Model parameter estimates are in Table 2.

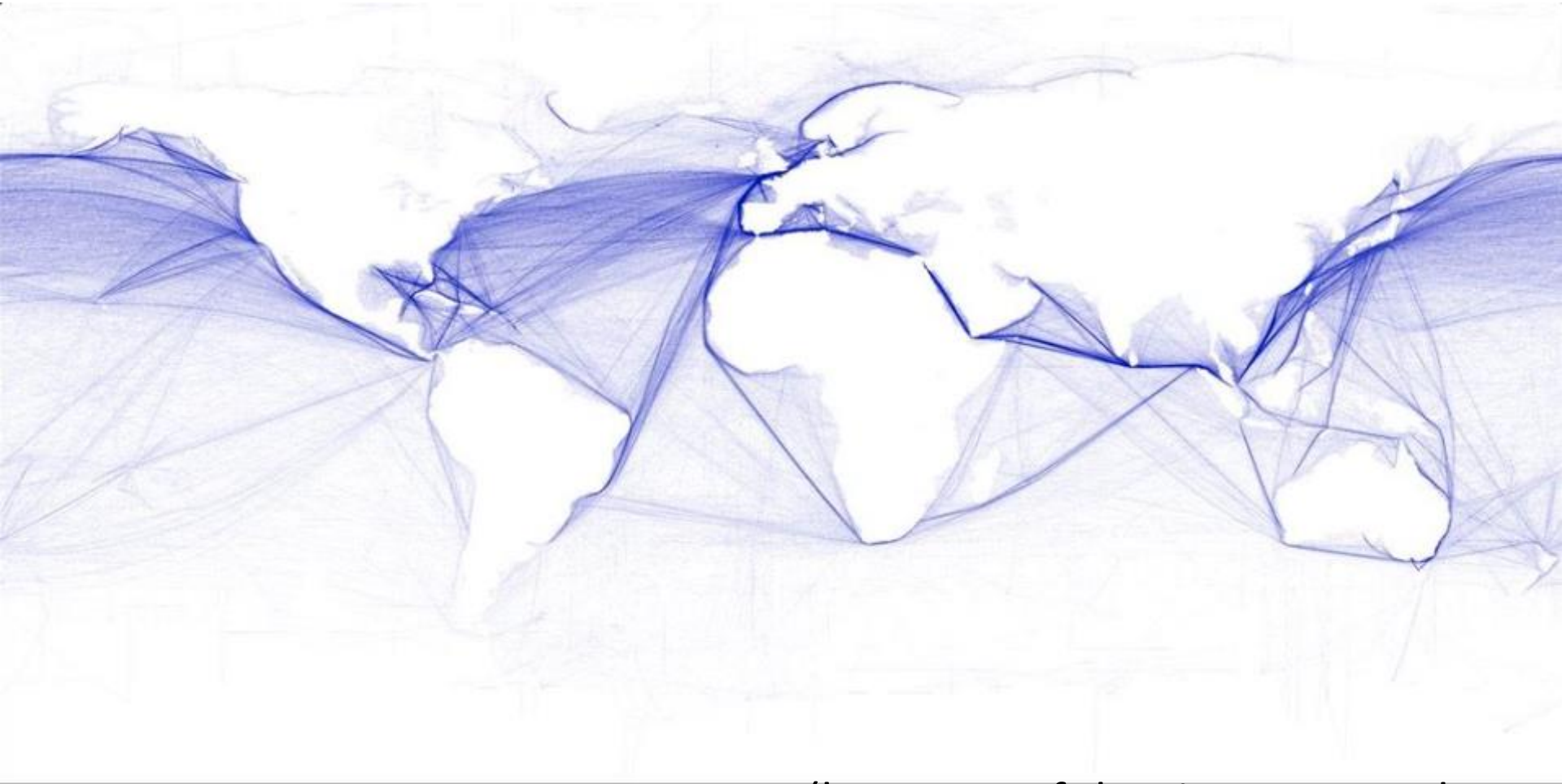
Borrelle et al. 2016, *Oryx*, http://doi.org/10.1017/S0030605316000880

COMMUNICATING RESULTS: GRAPHS

What is the **key** message you want the graph to convey?
What is the **simplest** way of communicating that message?

Is the end graph intuitive to understand?
What might distract readers from the key message?
Add the data where possible

Sometimes the raw data is all you need:
World shipping routes



Amazing Maps @Amazing_Maps · Mar 11
World shipping routes pic.twitter.com/Ywn4UfwWKV

(but most of the time we need to work a bit harder…)

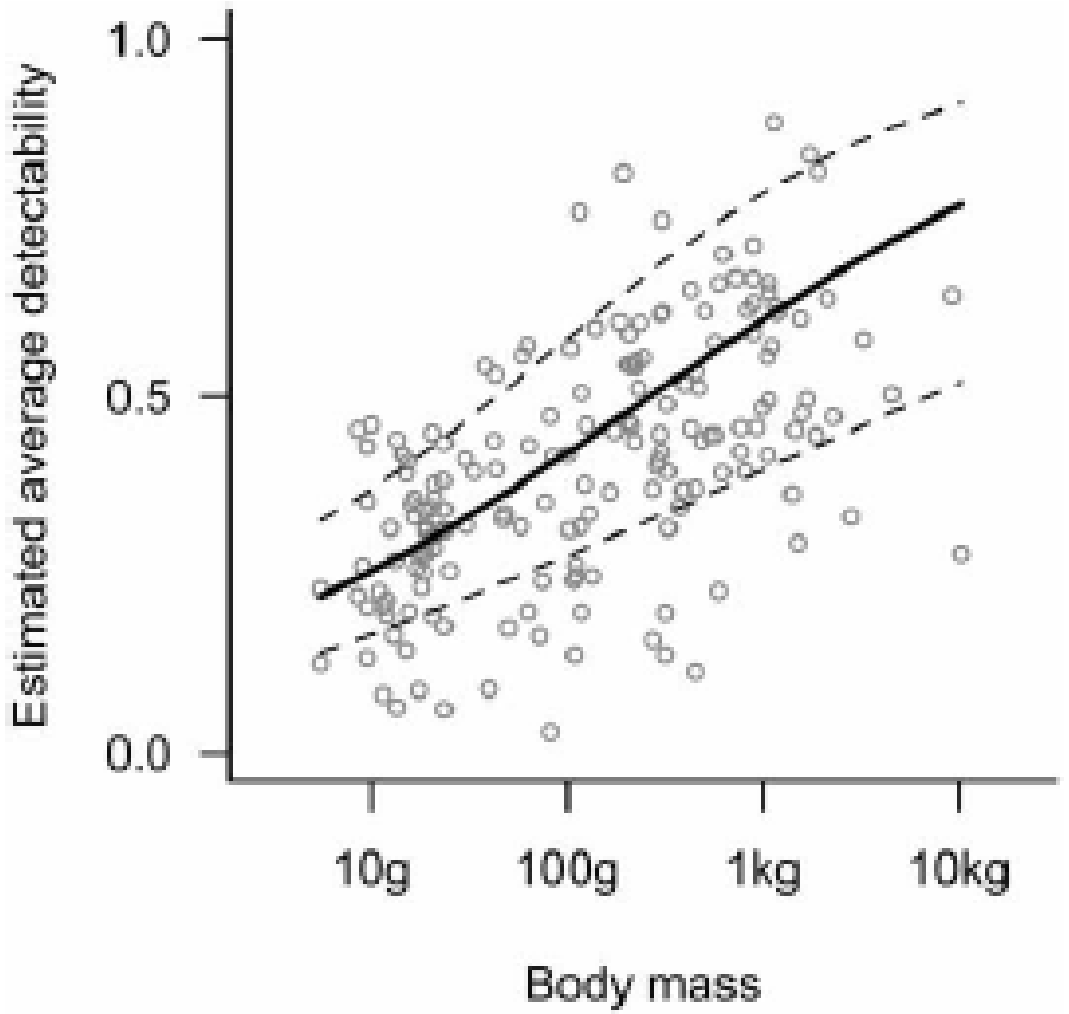
Use clear labels

Include units

Show the data

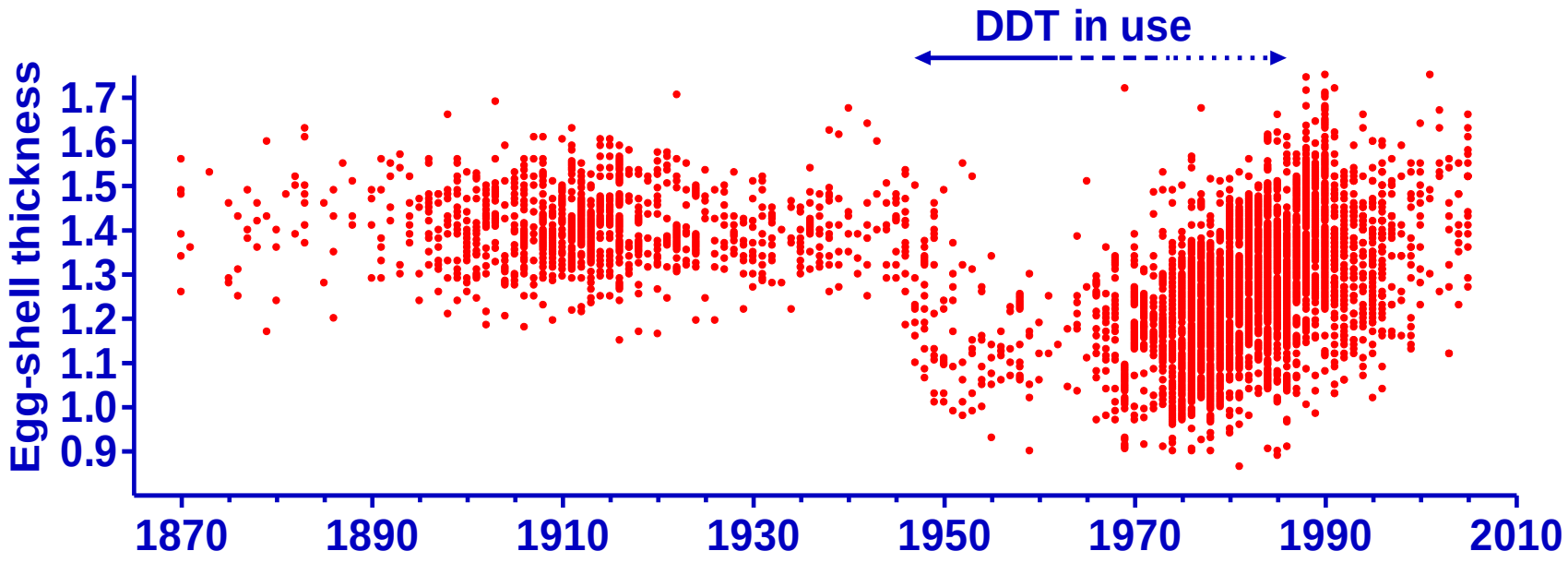
Show the model

Show uncertainty

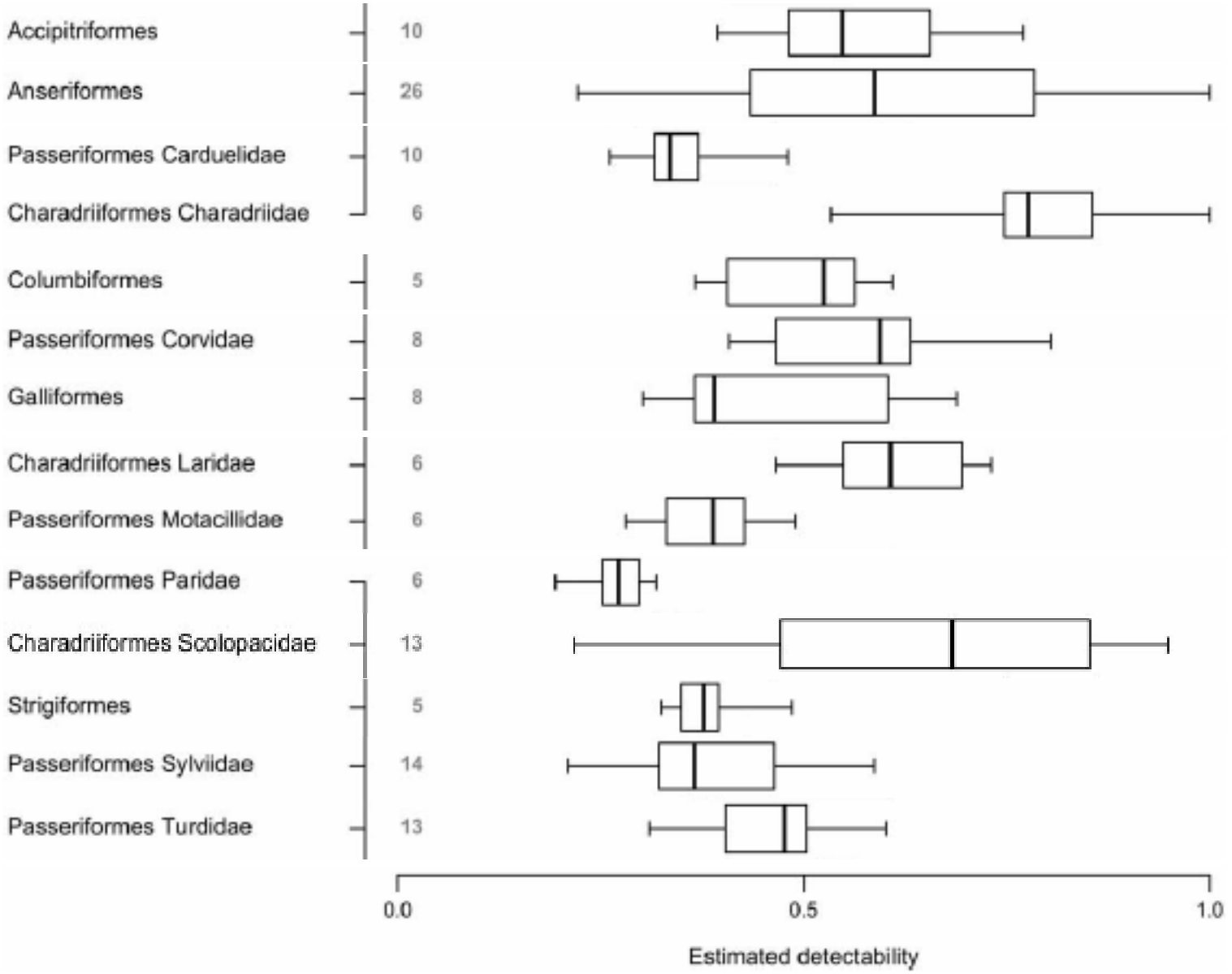


Using annotations to add context:

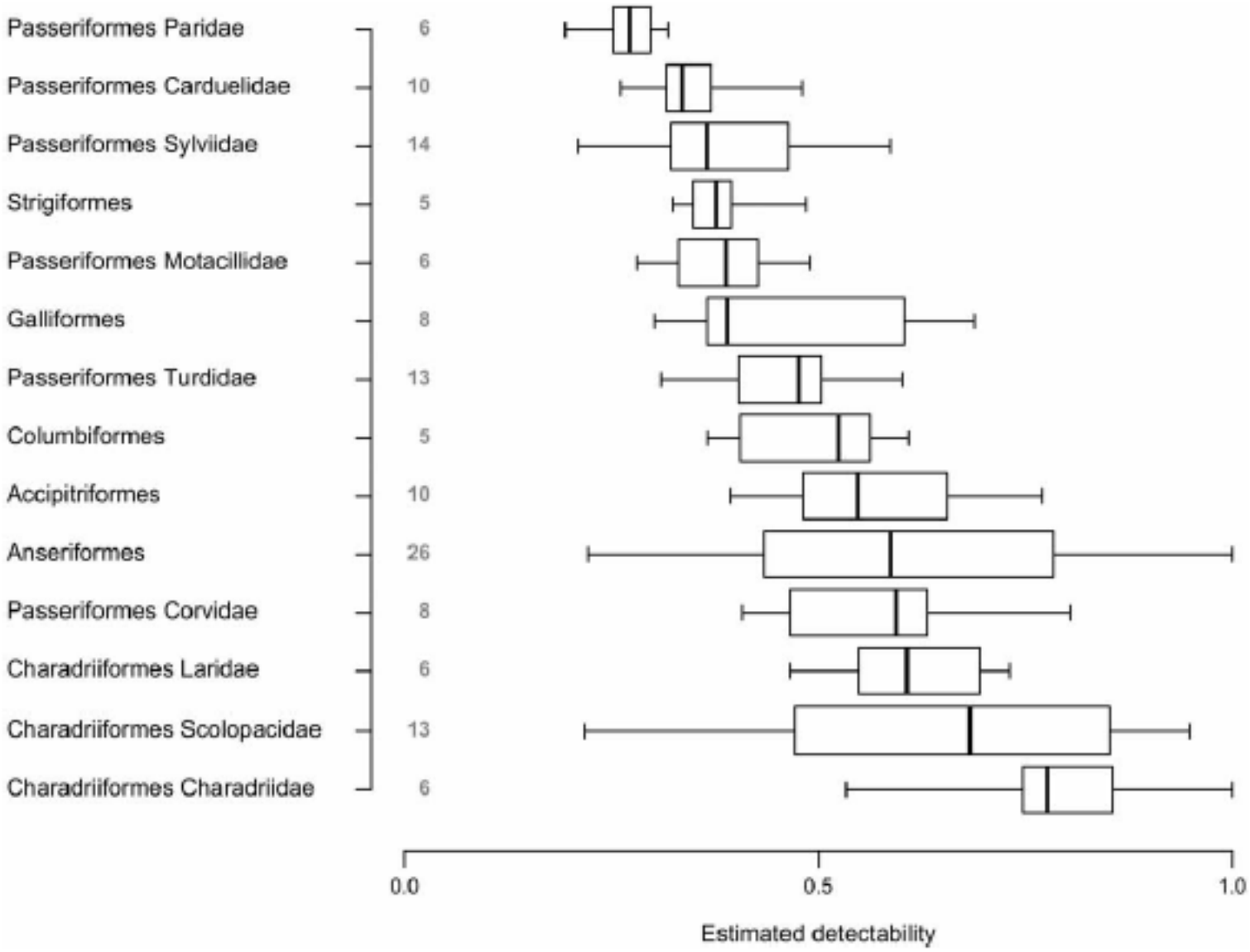
Sparrowhawk eggs and pesticide contamination



Thinking about composition:
Ordering of categories can be important



Thinking about composition:
Ordering of categories can be important



Colours

Select **intuitive** colours to communicate your message

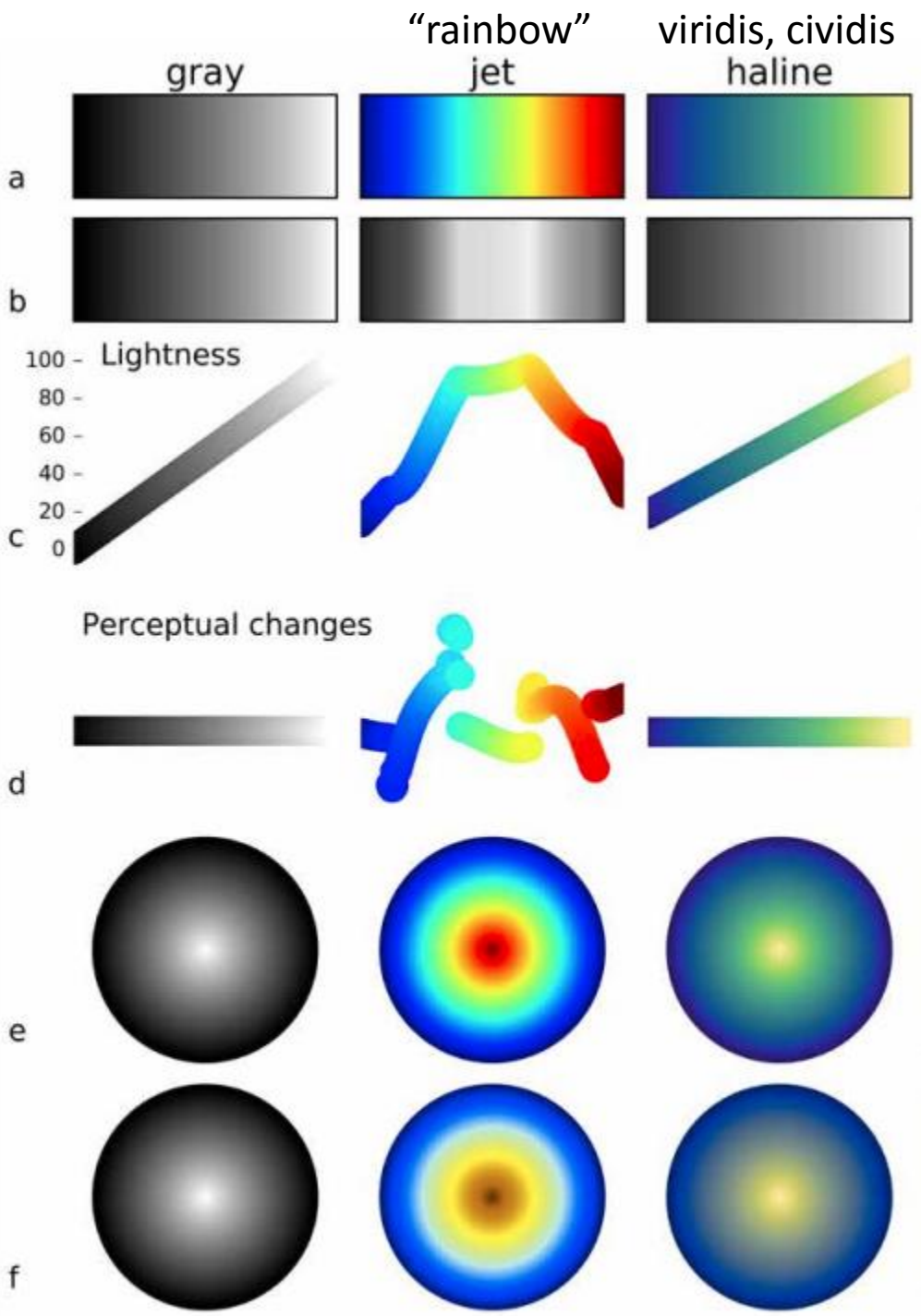
Use the **simplest** colour scheme that communicates the story

Think about the data:
Continuous or Categorical?
Ordered? Diverging?

Think about the medium:
print vs screen, photocopying?

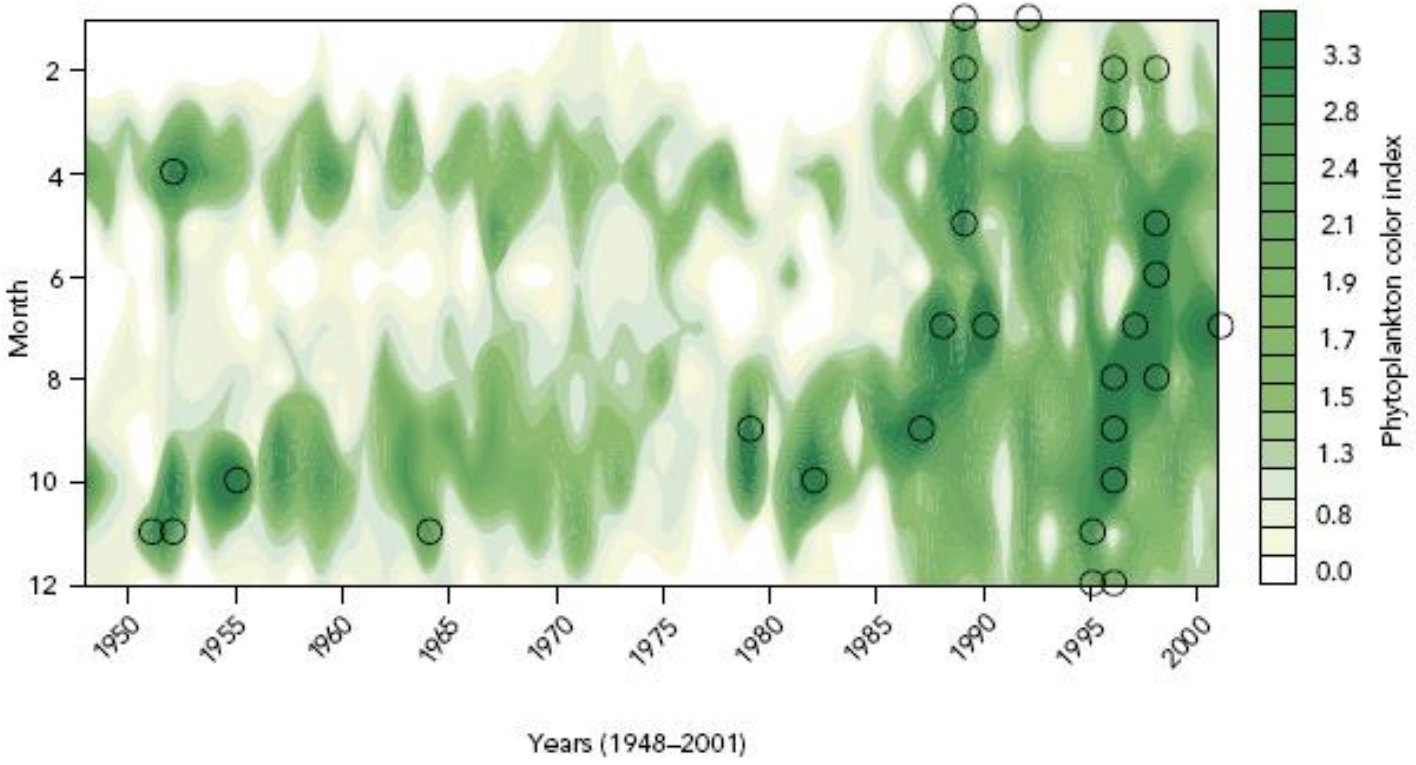
Think about the audience:
Colour blindness?

Tool for good palettes:
https://colorbrewer2.org/



Using colour as the third dimension: Changing timing of Phytoplankton bloom

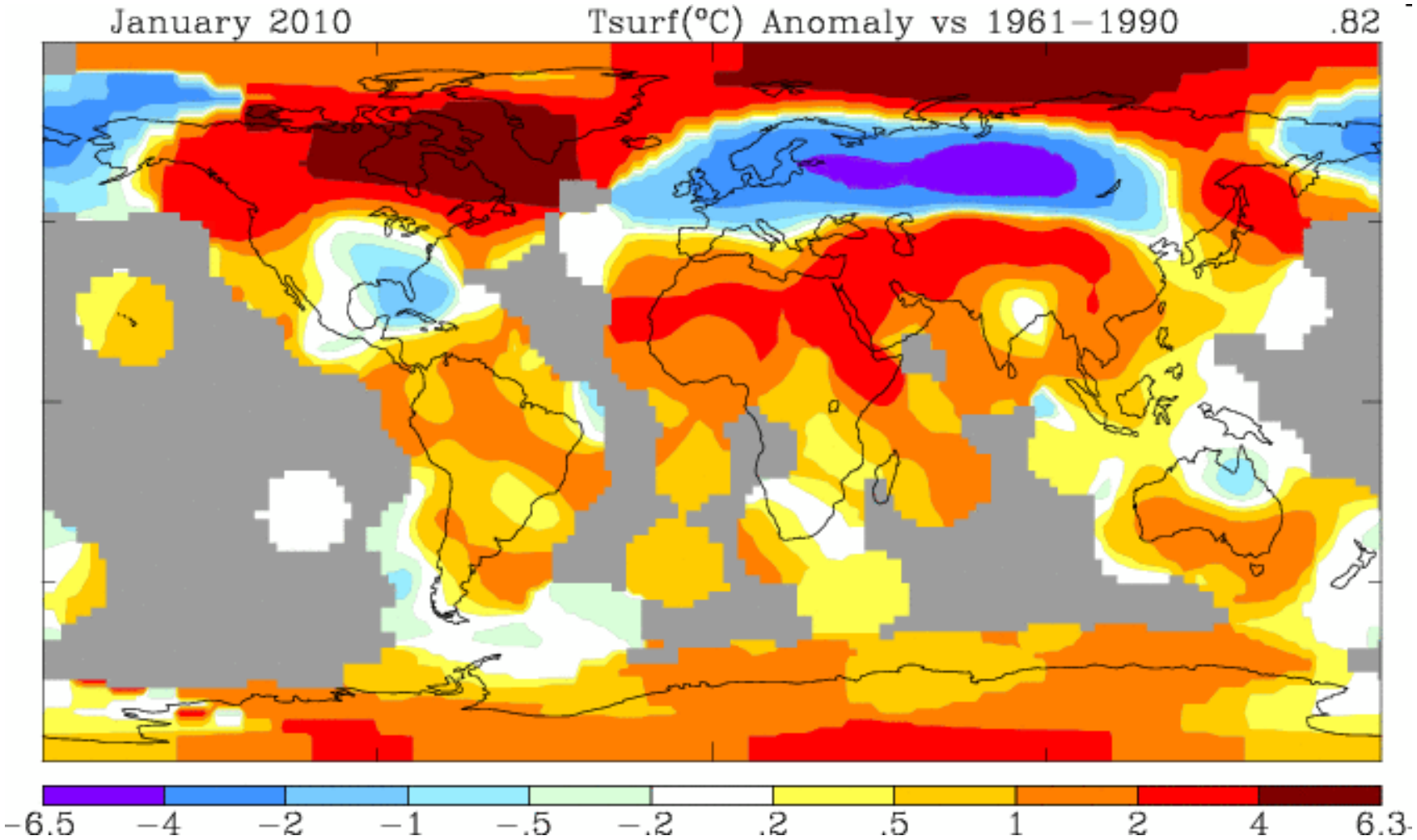
Single colour gradient -> darker = more



Source: M. Edwards, (SAHFOS), 2003.

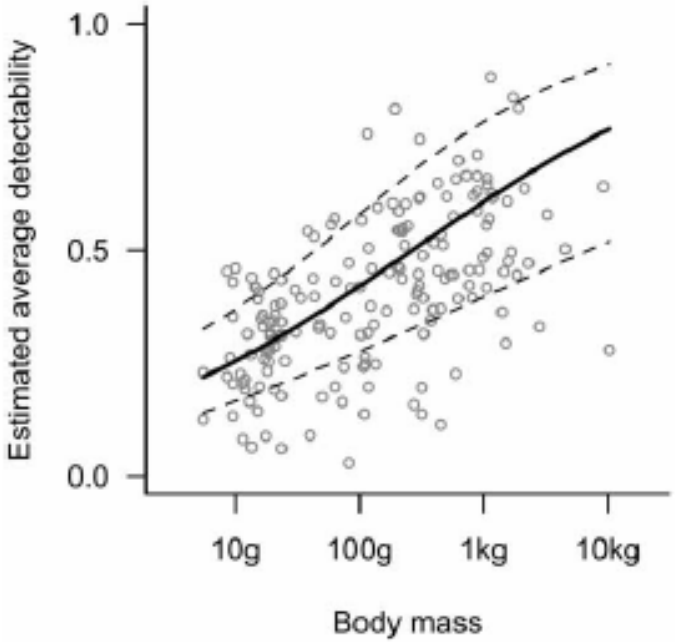
EEA Report 2004/2

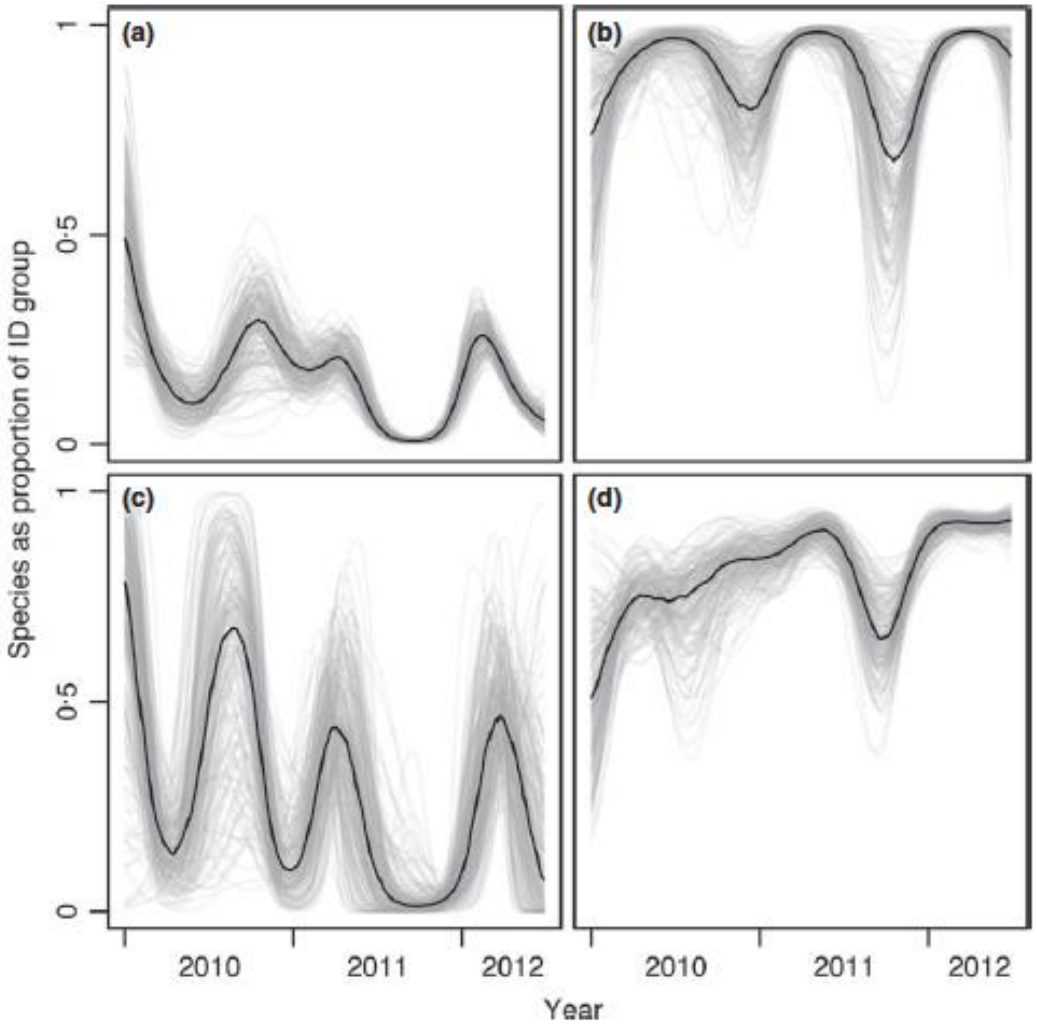
Using colour as the third dimension:
Diverging colour schemes highlight deviations from a baseline



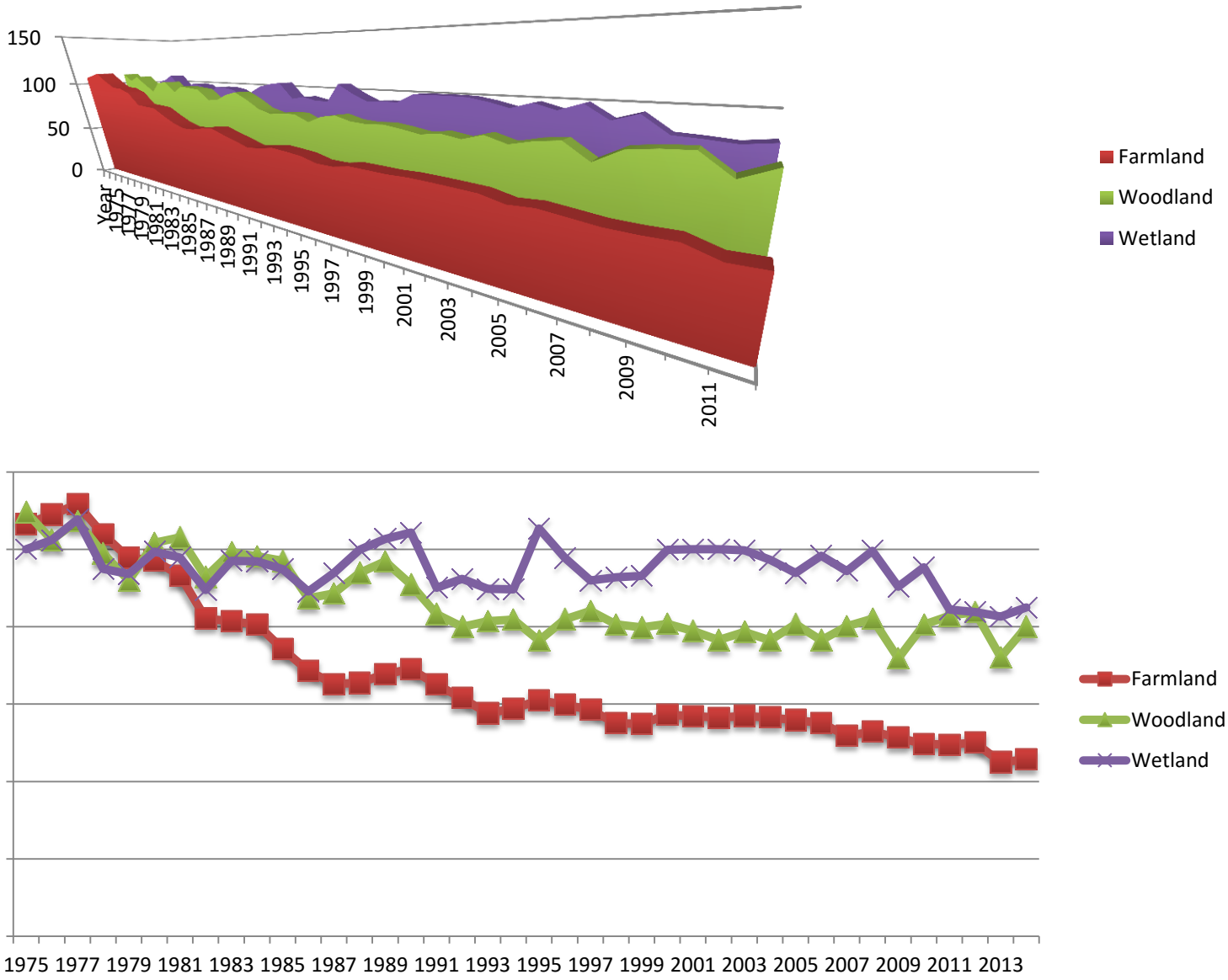
Two single colour gradients -> darker = more extreme
Blue <-> red to encode direction

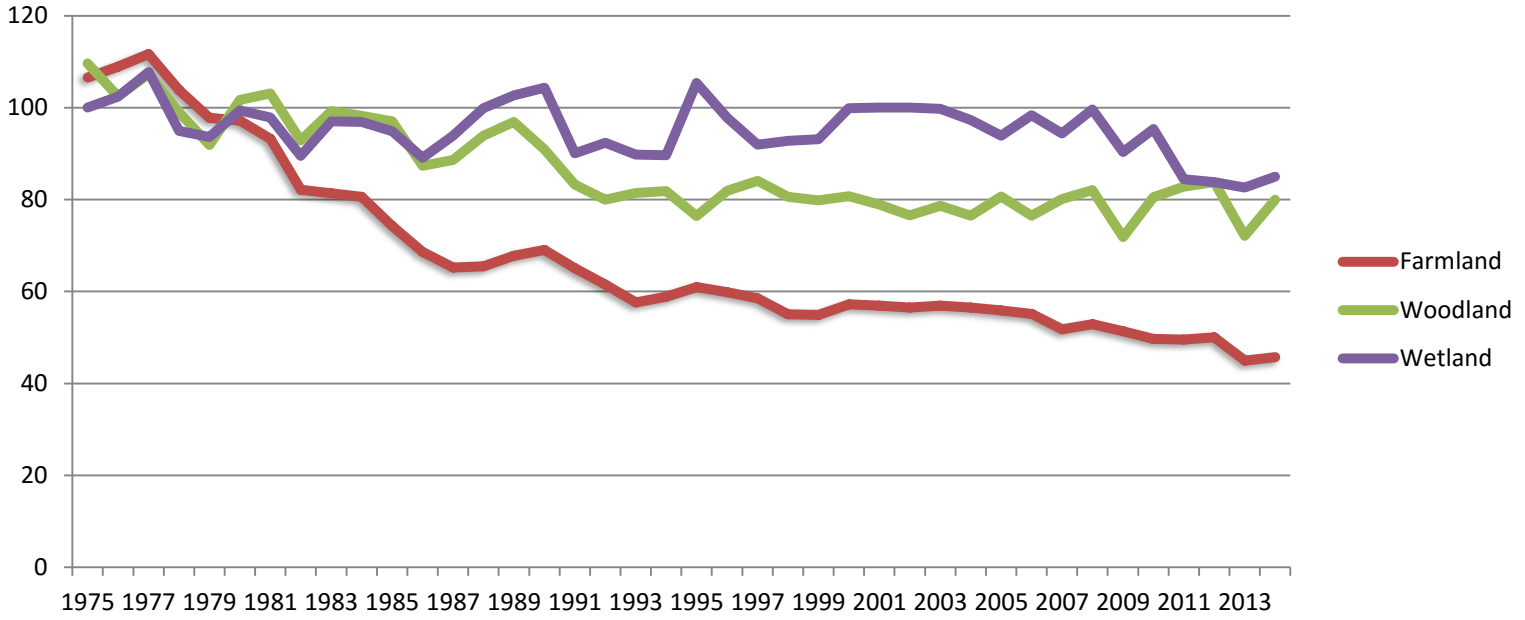
Visualising uncertainty:
Confidence bands and "spaghetti plots"

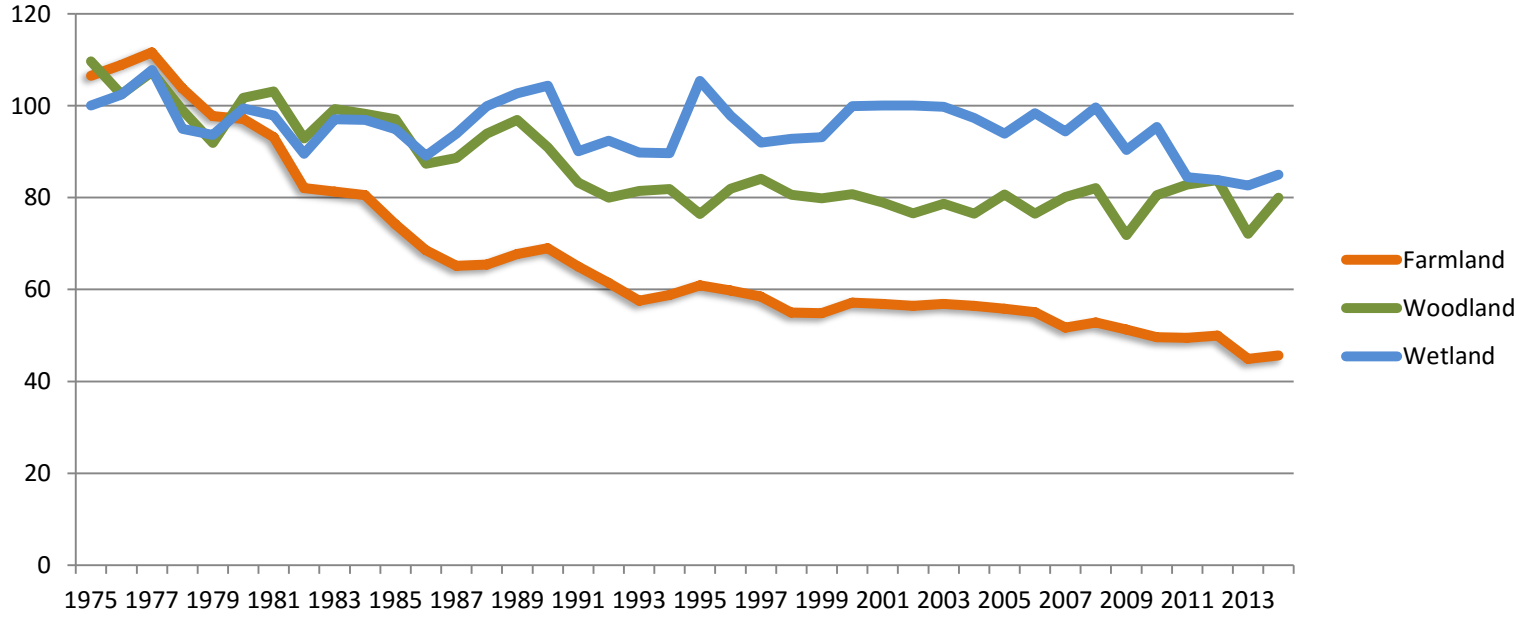


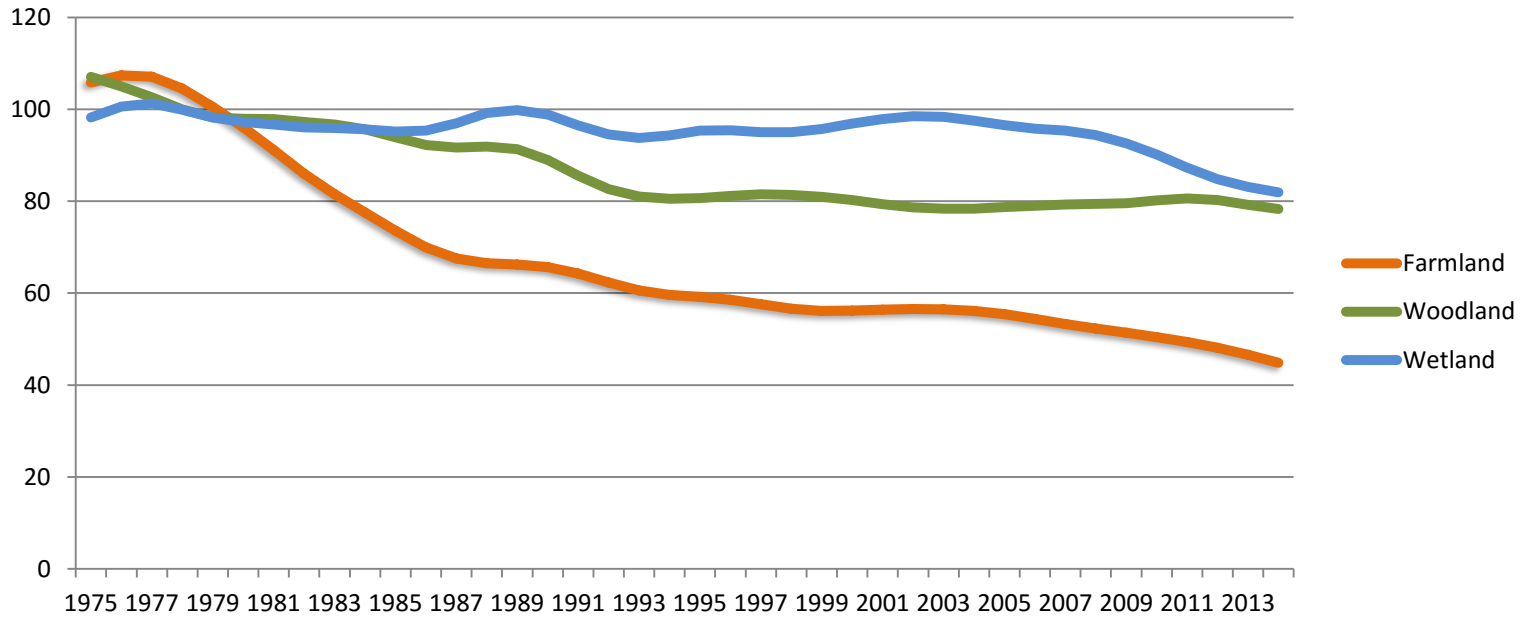


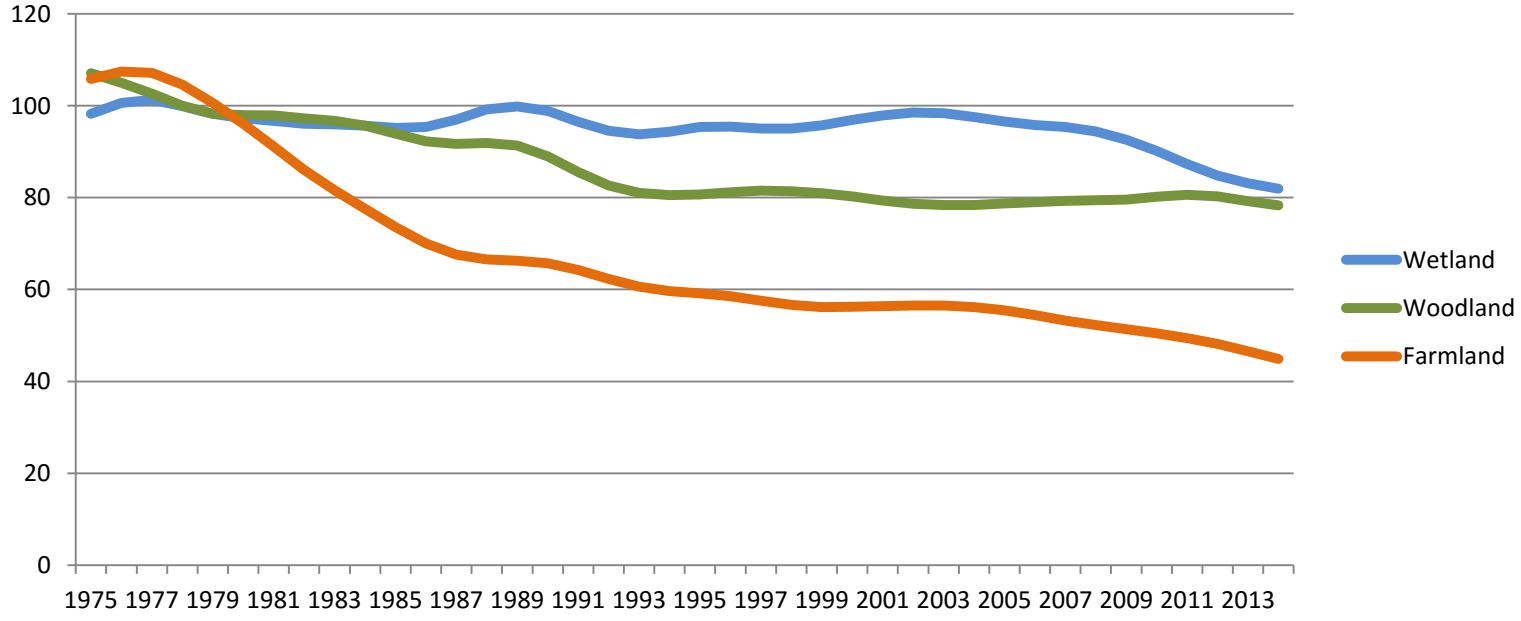
Simplifying graphs to the key message

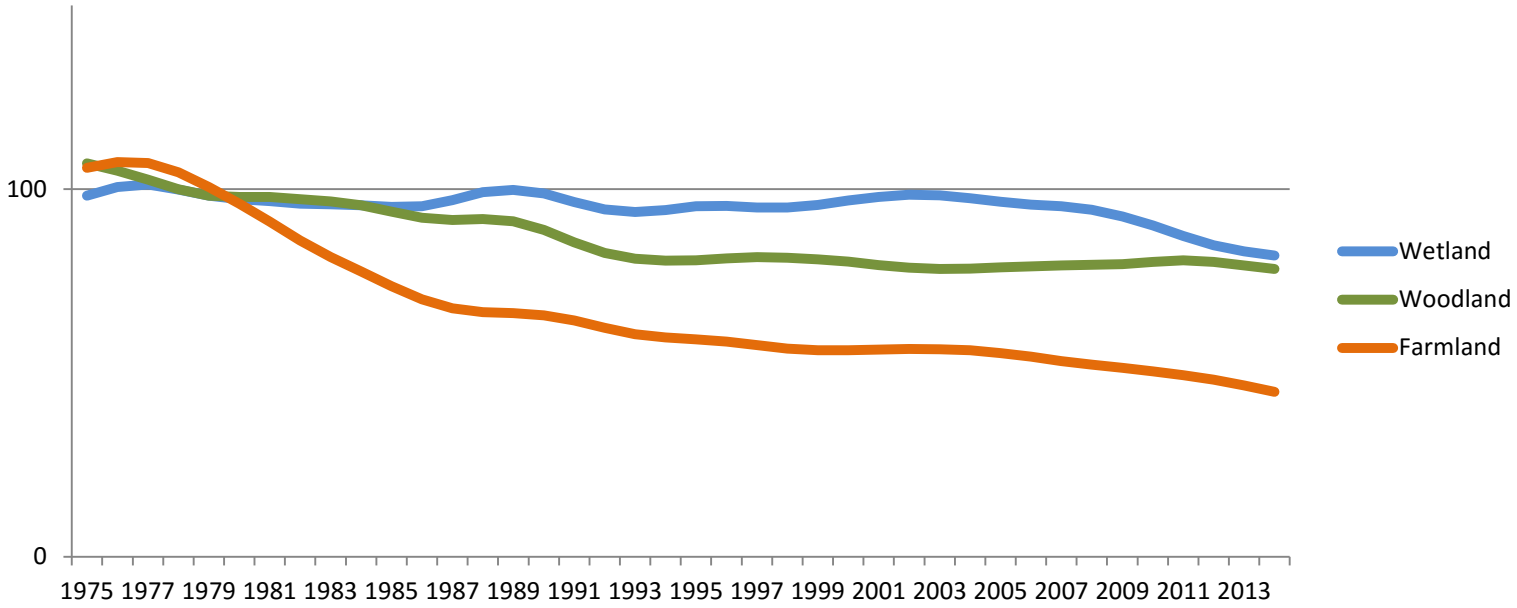


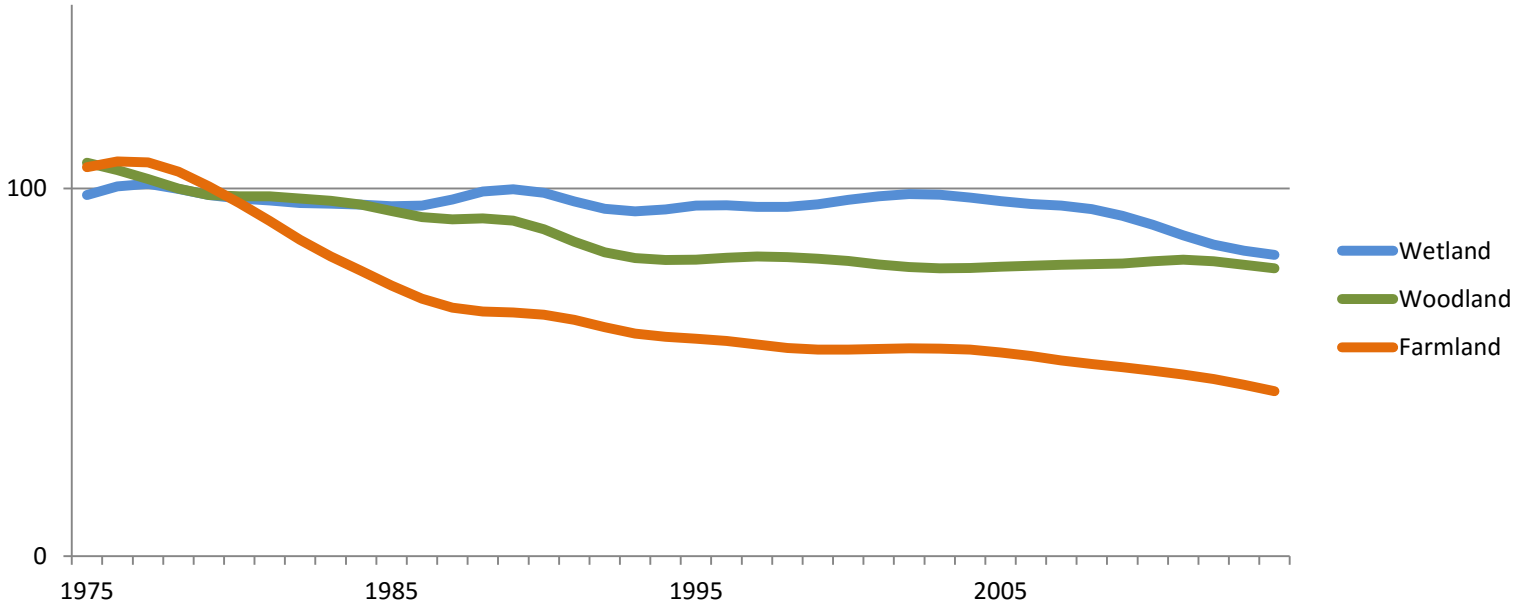

120
100
80
60
40
20
0
1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013
Farmland
Woodland
Wetland

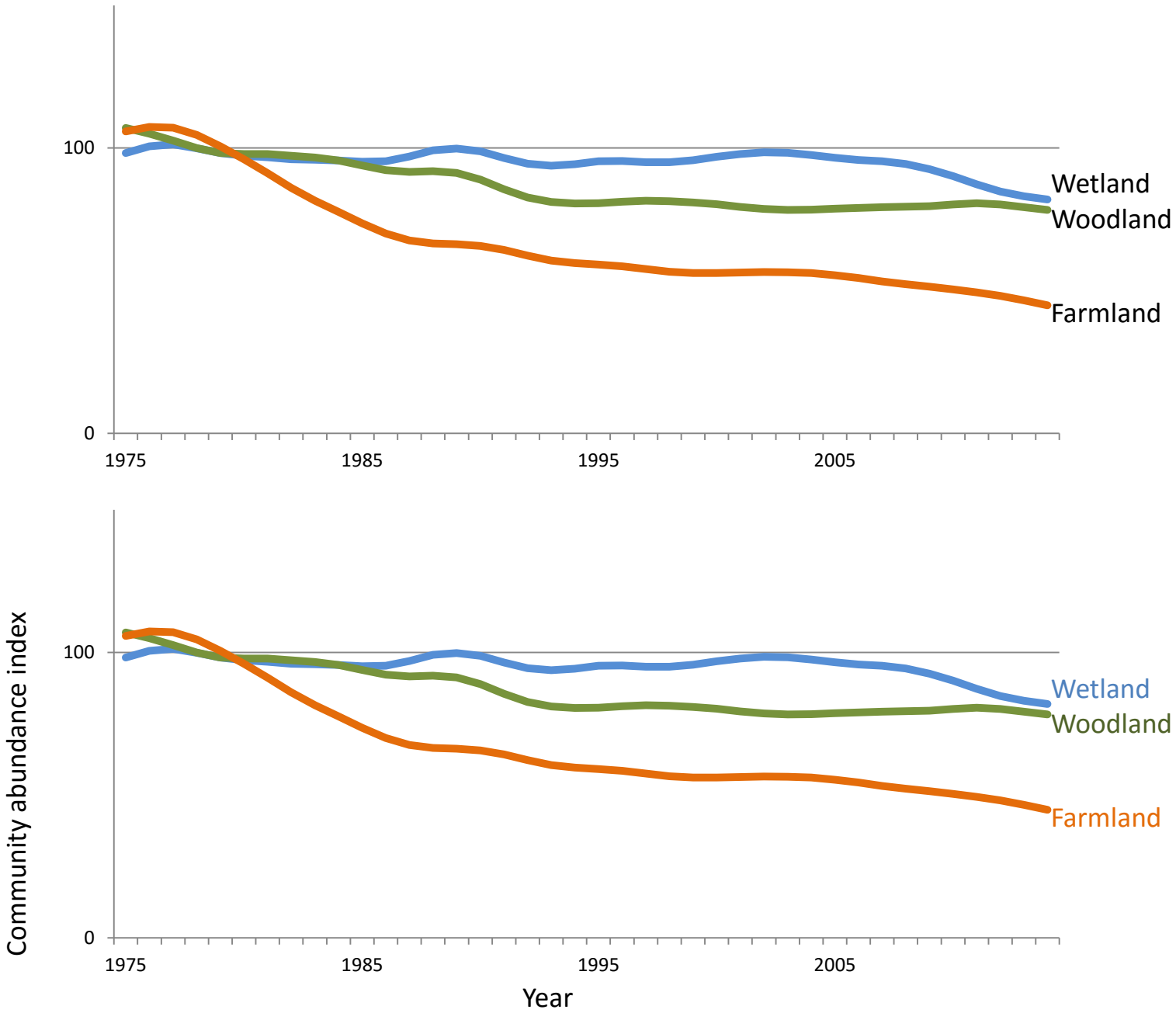

120
100
80
60
40
20
0
1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013
Farmland
Woodland
Wetland

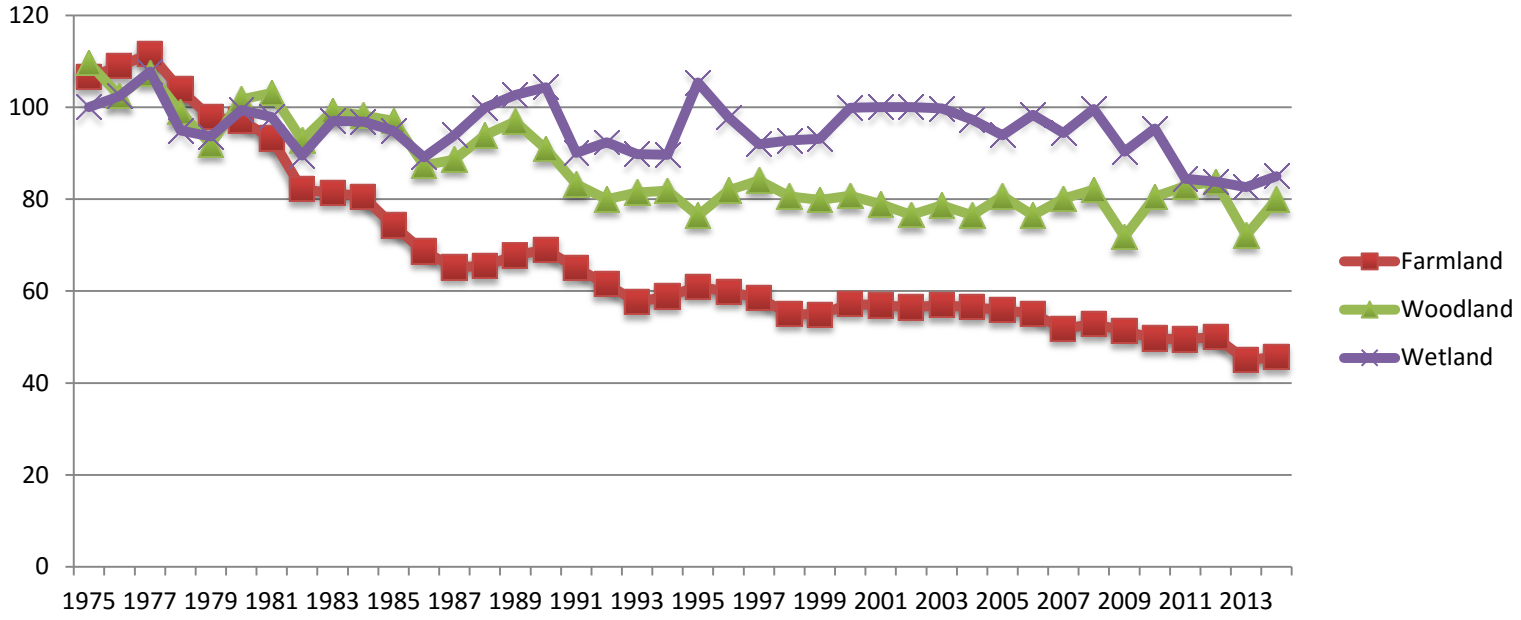

120
100
80
60
40
20
0
1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013
Farmland
Woodland
Wetland

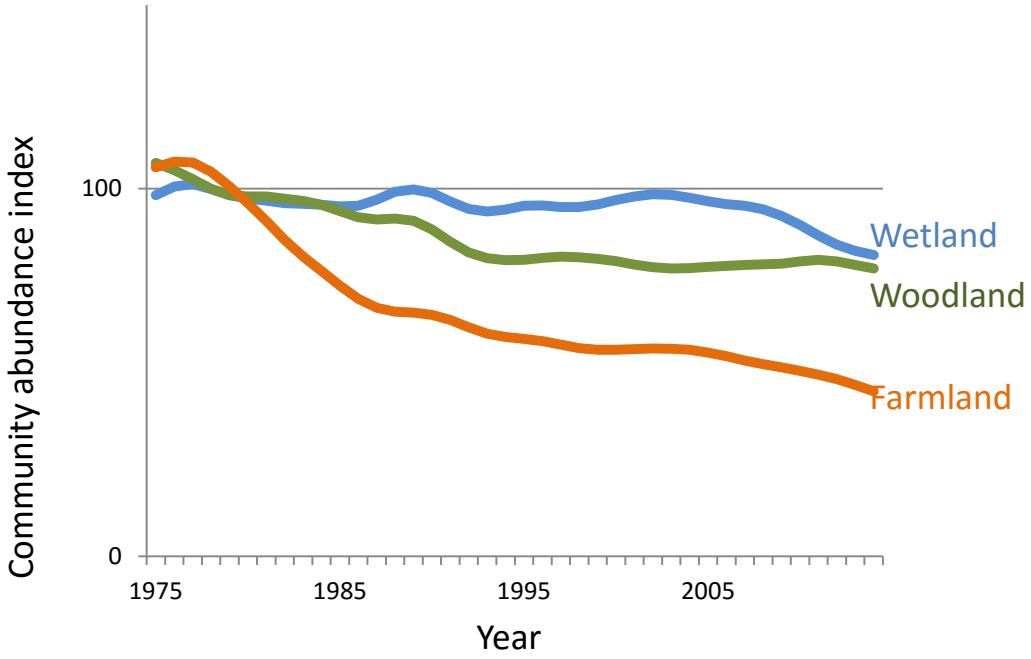

120
100
80
60
40
20
0
1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013
Wetland
Woodland
Farmland


100
0
1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013
Wetland
Woodland
Farmland

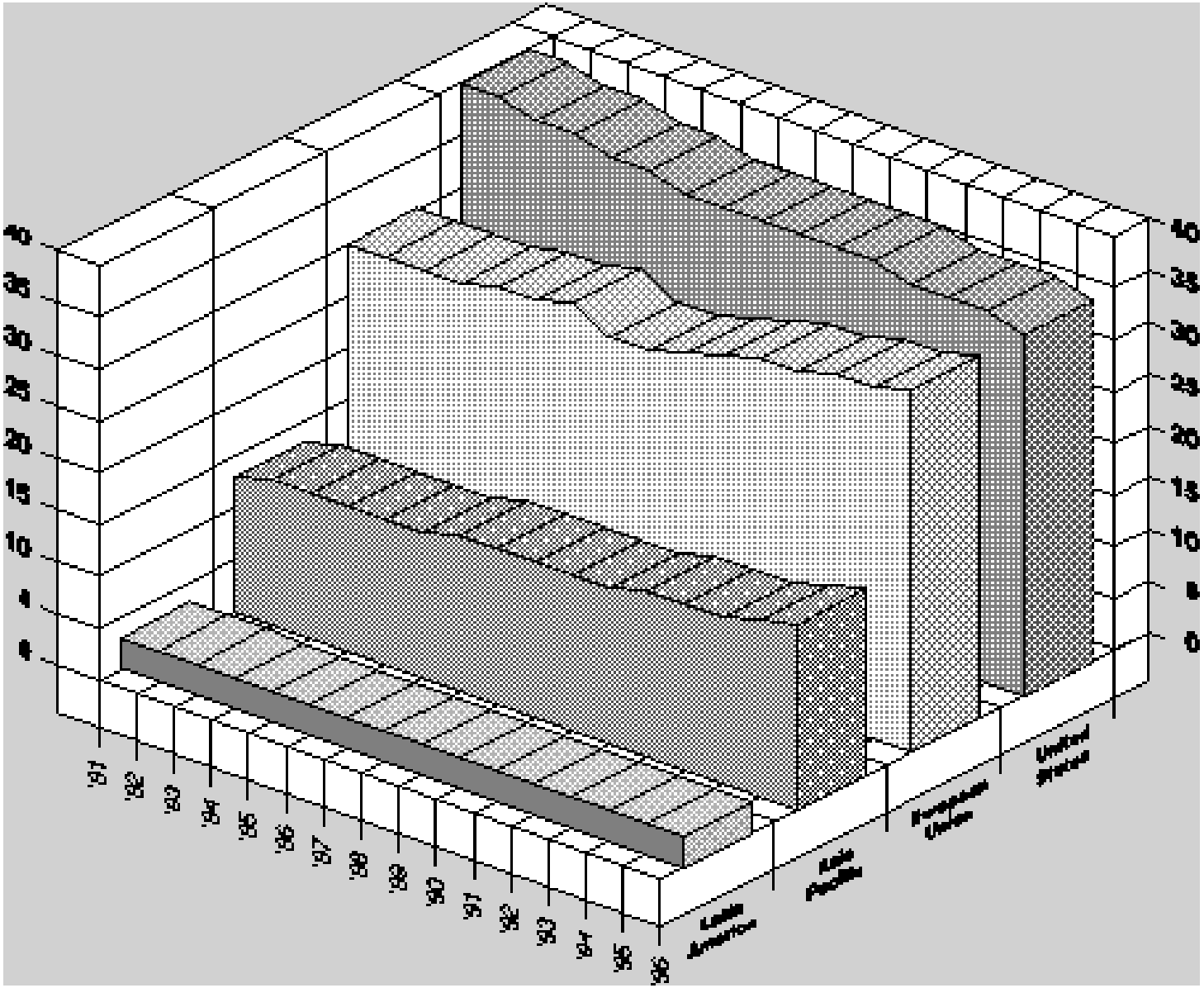

100
0
1975
1985
1995
2005
Wetland
Woodland
Farmland


100
0
1975
1985
1995
2005
Wetland
Woodland
Farmland
Community abundance index
Year

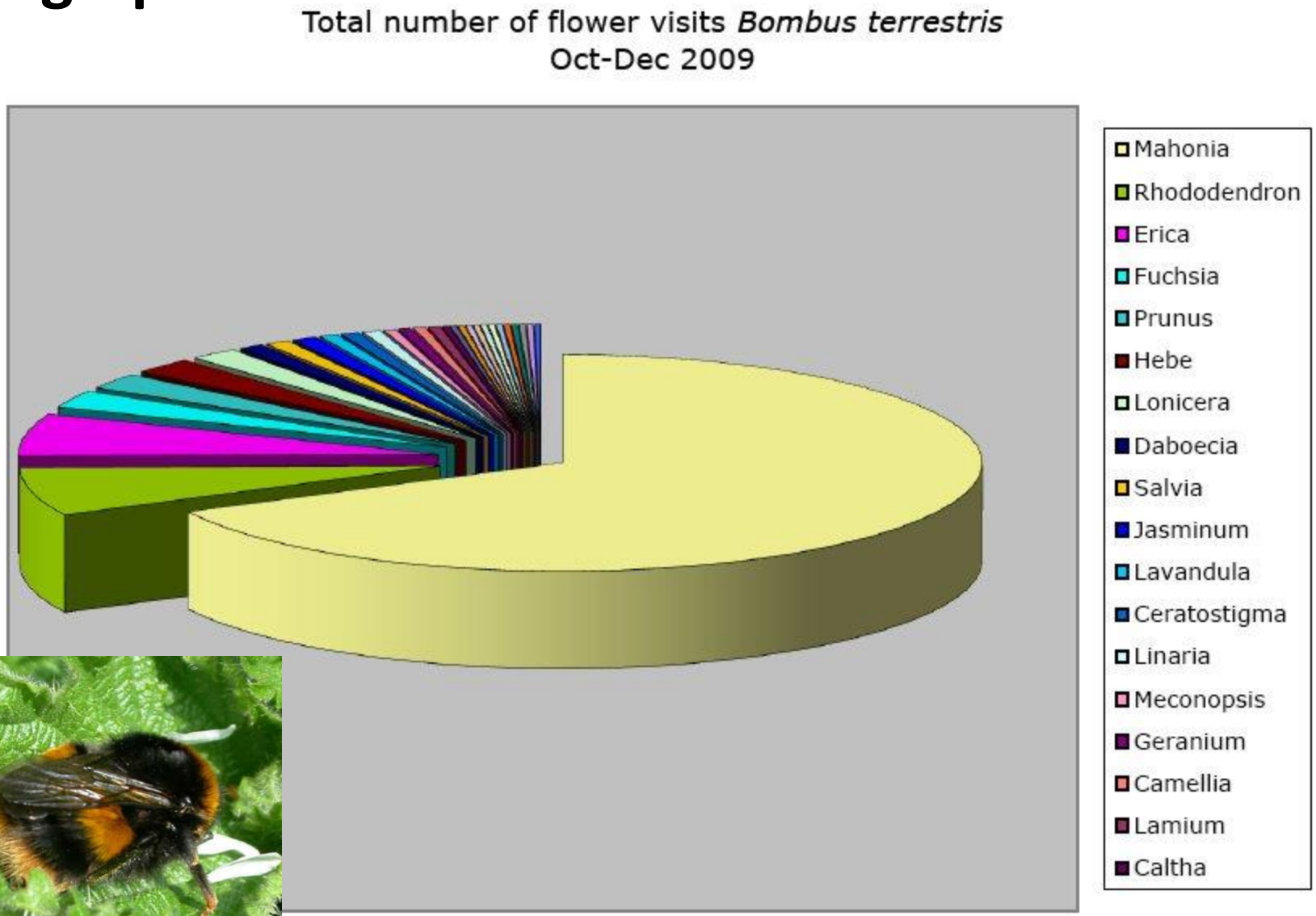

120
100
80
60
40
20
0
1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013
Farmland
Woodland
Wetland


Community abundance index
100
0
1975
1985
1995
2005
Year
Wetland
Woodland
Farmland

Bad graphs


40
35
30
25
20
15
10
5
0
'81
'82
'83
'84
'85
'86
'87
'88
'89
'90
'91
'92
'93
'94
'95
'96
Latin America
Asia Pacific
European Union
United States

Bad graphs


Total number of flower visits *Bombus terrestris*
Oct-Dec 2009
Mahonia
Rhododendron
Erica
Fuchsia
Prunus
Hebe
Lonicera
Daboecia
Salvia
Jasminum
Lavandula
Ceratostigma
Linaria
Meconopsis
Geranium
Camellia
Lamium
Caltha

Bad graphs

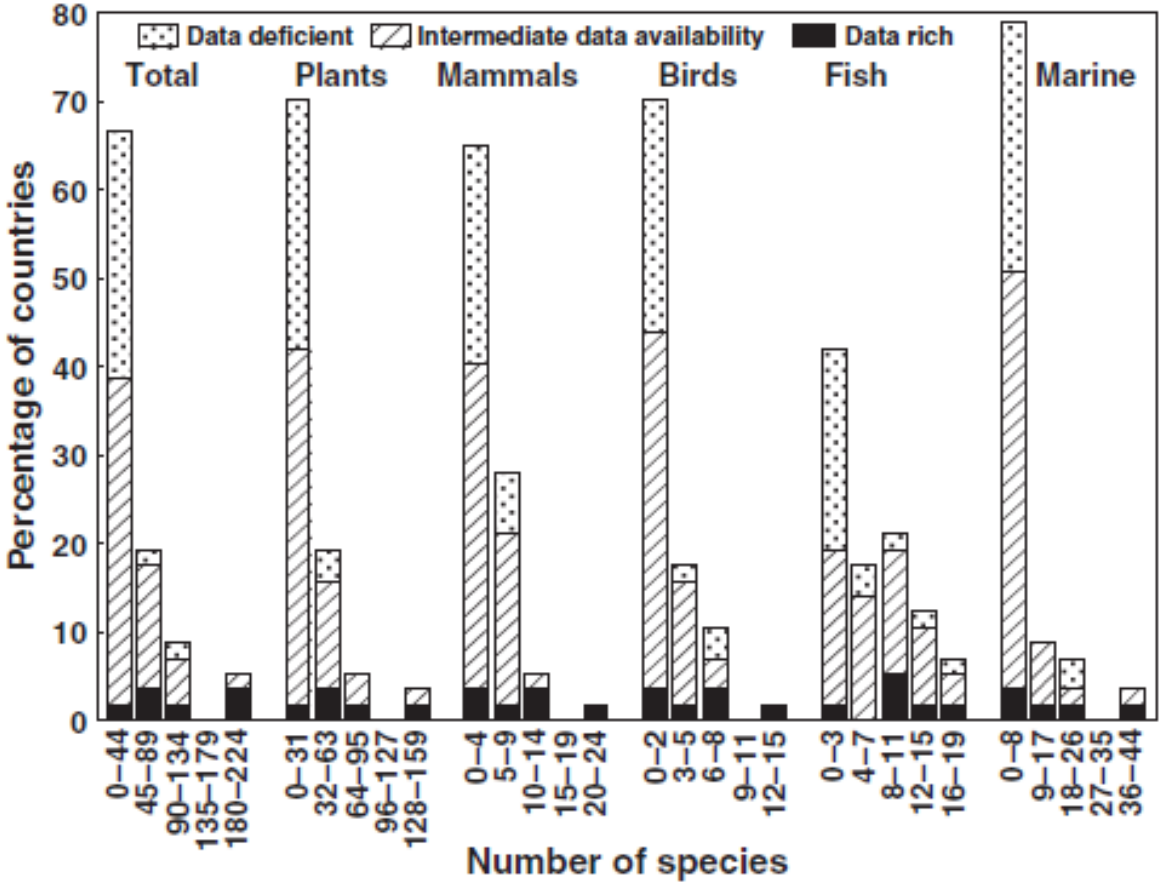
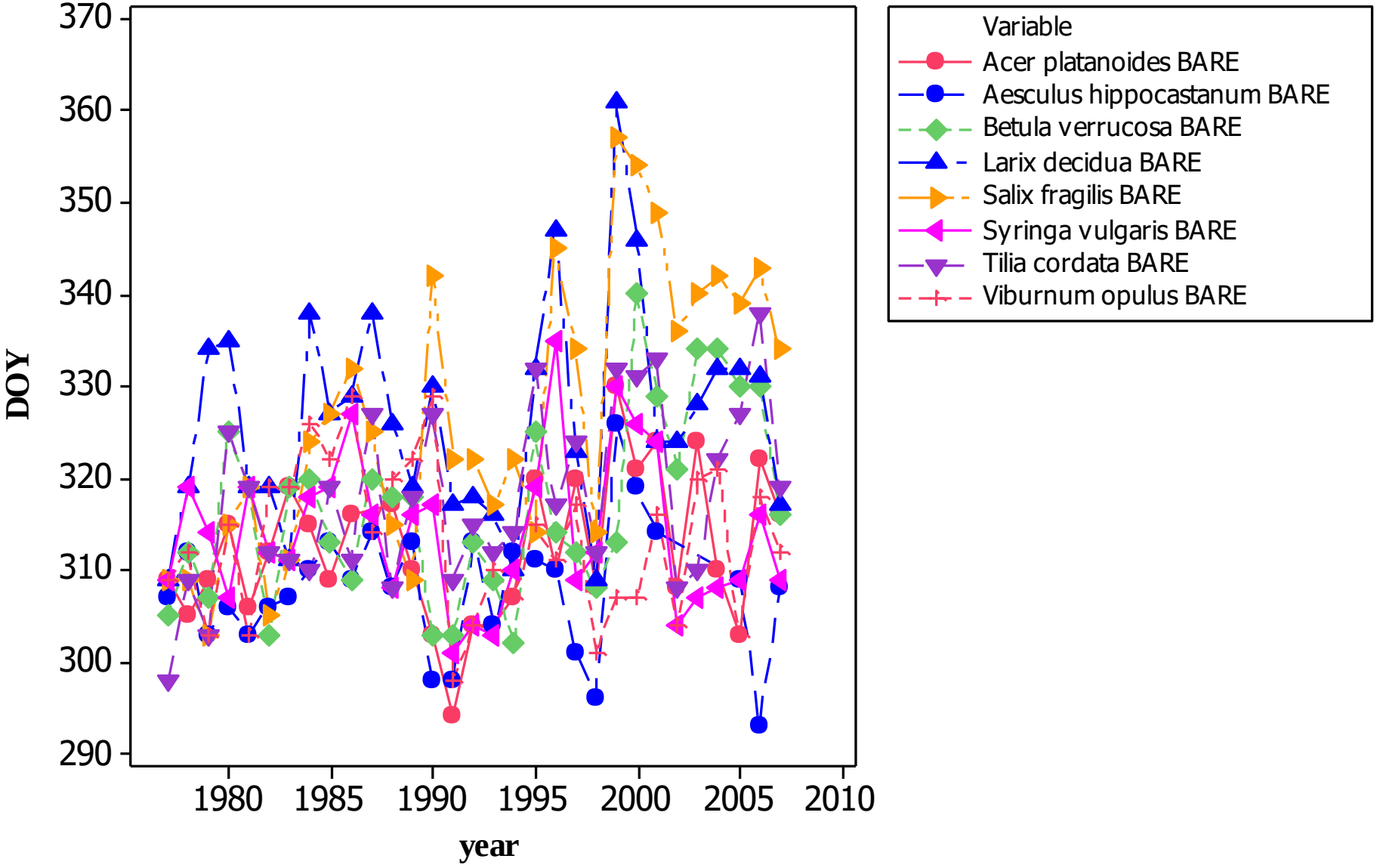


Figure 2 Frequency distributions of the number of Invasive alien species (IAS) across countries (n = 57) with different degrees of data adequacy (n = 18 data deficient (DD), n = 33 intermediate (IDA) and n = 6 data rich (DR) countries). (Vascular plants, mammals, birds, freshwater fish and marine organisms; amphibians not shown individually here because of low frequencies.)

Bad graphs


370
360
350
340
330
320
310
300
290
DOY
1980
1985
1990
1995
2000
2005
2010
year
Variable
Acer platanoides BARE
Aesculus hippocastanum BARE
Betula verrucosa BARE
Larix decidua BARE
Salix fragilis BARE
Syringa vulgaris BARE
Tilia cordata BARE
Viburnum opulus BARE

Survey design and statistical analysis for conservation science

SURVEY DESIGN AND SAMPLING

1. REPRESENTATIVE samples
2. STRATIFICATION for sampling design
3. POWER ANALYSIS to assess sampling design
4. PSEUDO-REPLICATION in sampling

STATISTICAL ANALYSIS

5. DESCRIPTION or EXPLANATION
6. SIGNIFICANCE is continuous
7. Separate SIGNIFICANCE and EFFECT SIZE
8. Separate CORRELATION AND CAUSATION
9. Consider UNCERTAINTY

COMMUNICATING RESULTS

10. Translate statistics to HUMAN LANGUAGE
11. Draw GOOD GRAPHS

Where to get support - Statistics

Statistics books are a bit like shoes – different people fit different types…

- **Regression and Other Stories**
  - Free PDF: https://avehtari.github.io/ROS-Examples/
  - Comprehensive introduction to modern applied statistics
- **Statistical Rethinking**
  - https://xcelab.net/rm/statistical-rethinking/
  - one of the most readable (intermediate - advanced) statistics books
  - free lecture series on youtube

- Wildlife population assessment (bird ringing, camera traps, etc) often require techniques that differ from regression models
  - Books by Kery, Schaub, Royle provide a good starting point
  - E.g. “Bayesian Population Analysis using WinBUGS”
  - https://www.mbr-pwrc.usgs.gov/pubanalysis/roylebook/

Where to get support - Coding

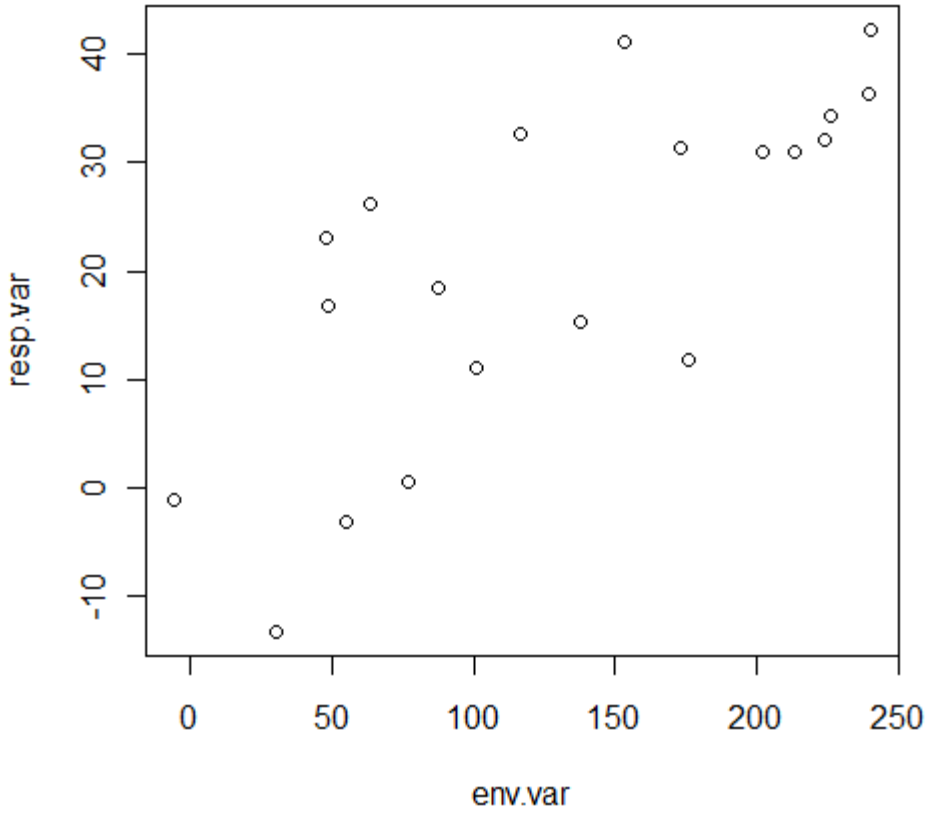
- The Carpentries www.carpentries.org
  - Non-profit that runs low cost data analysis / coding workshops worldwide (R, python, UNIX, …)
  - All materials freely available
  - most are suitable for self study, e.g. https://datacarpentry.org/ecology-workshop/
- Coding Club https://ourcodingclub.github.io/
  - Free online data science course for ecologists (R, some python)
  - Includes statistics tutorials
- R-Ladies https://rladies.org/
  - Global diversity initiative within in the R community

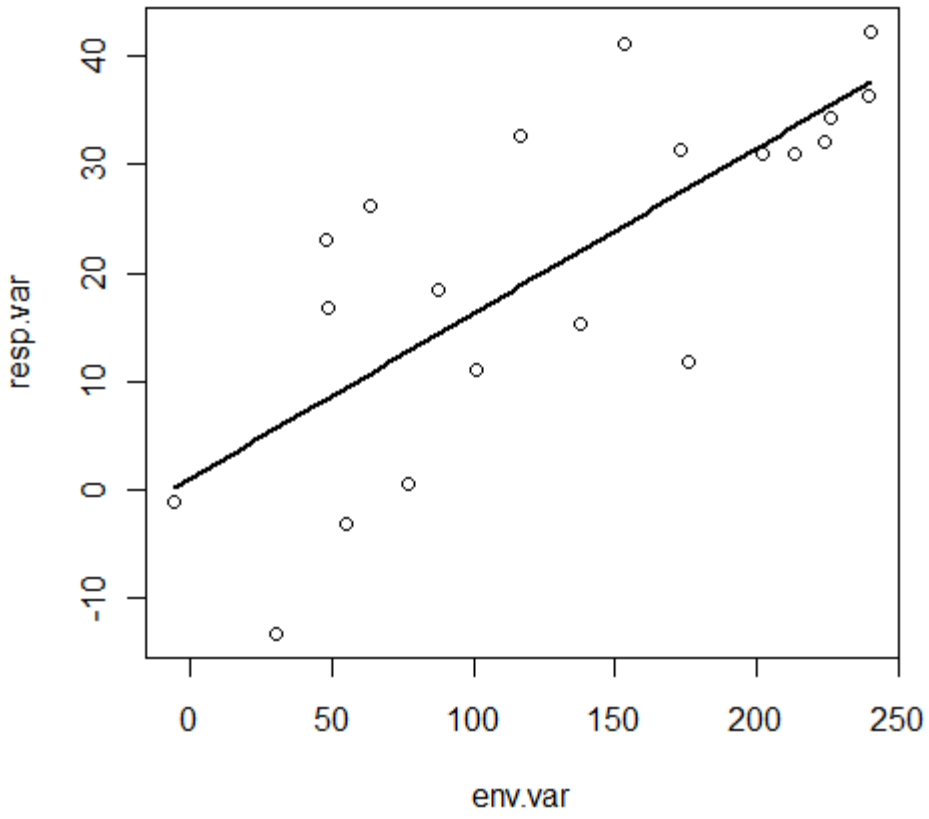
Where to get support - Collaboration

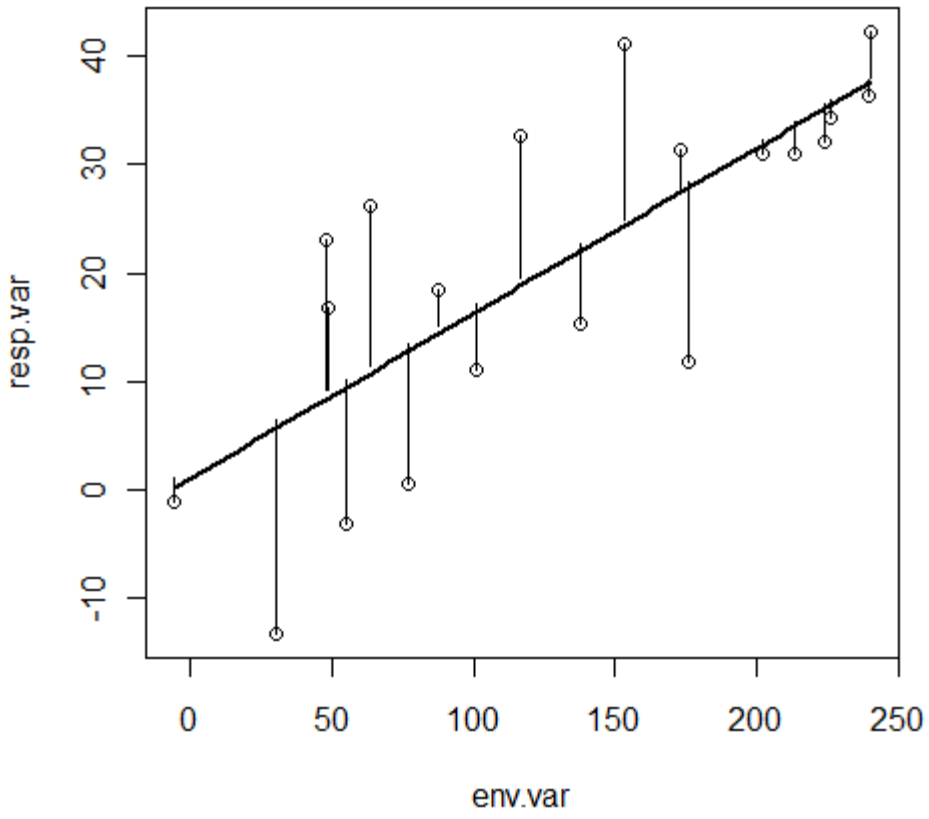
- You don’t have to learn or do everything yourself! (Even in your Masters/PhD dissertation)
- Conservation Science is highly interdisciplinary and collaborative, so think about building collaborations with Statisticians
  - the earlier the better!
- Statistics form a core intellectual contribution to a study, so give credit where it’s due!
  - inclusion in grant proposals, budgets, authorship
  - beer, chocolate, etc

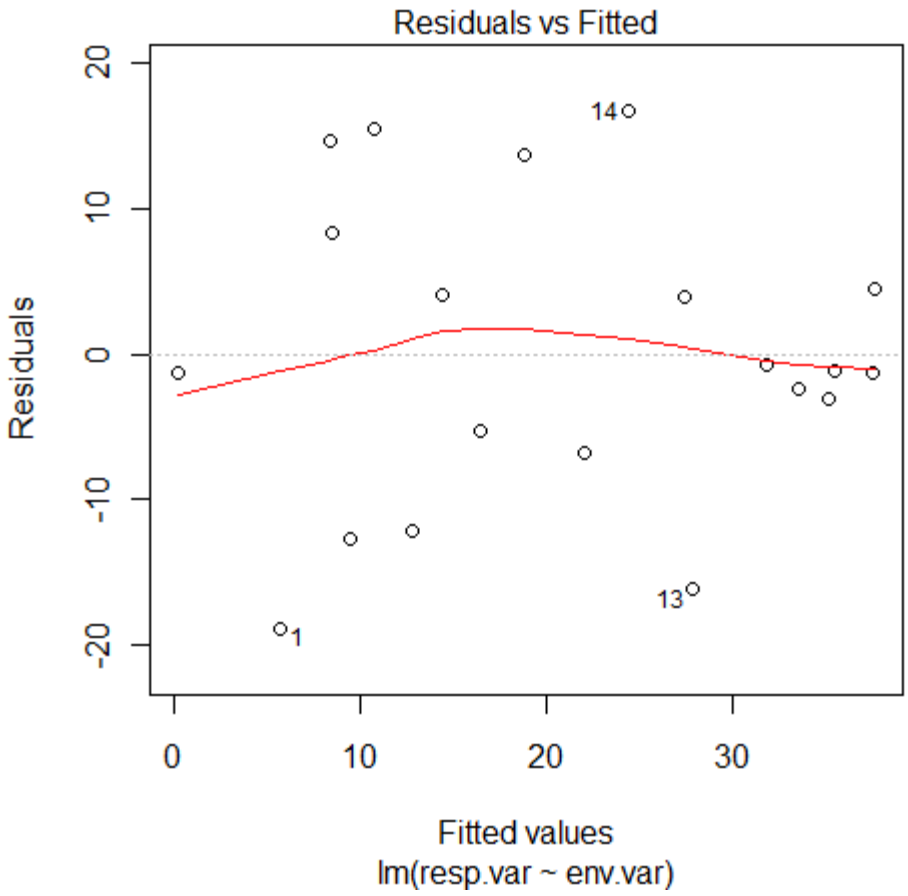
5. Do NOT ignore the model fit

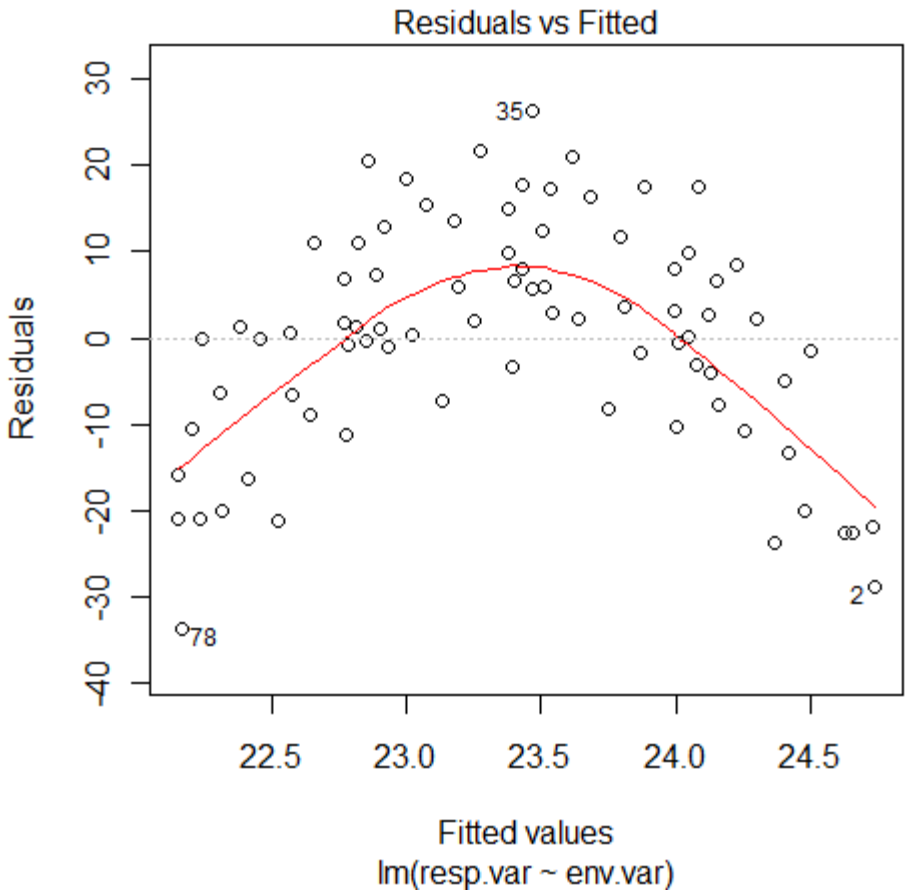
Residuals

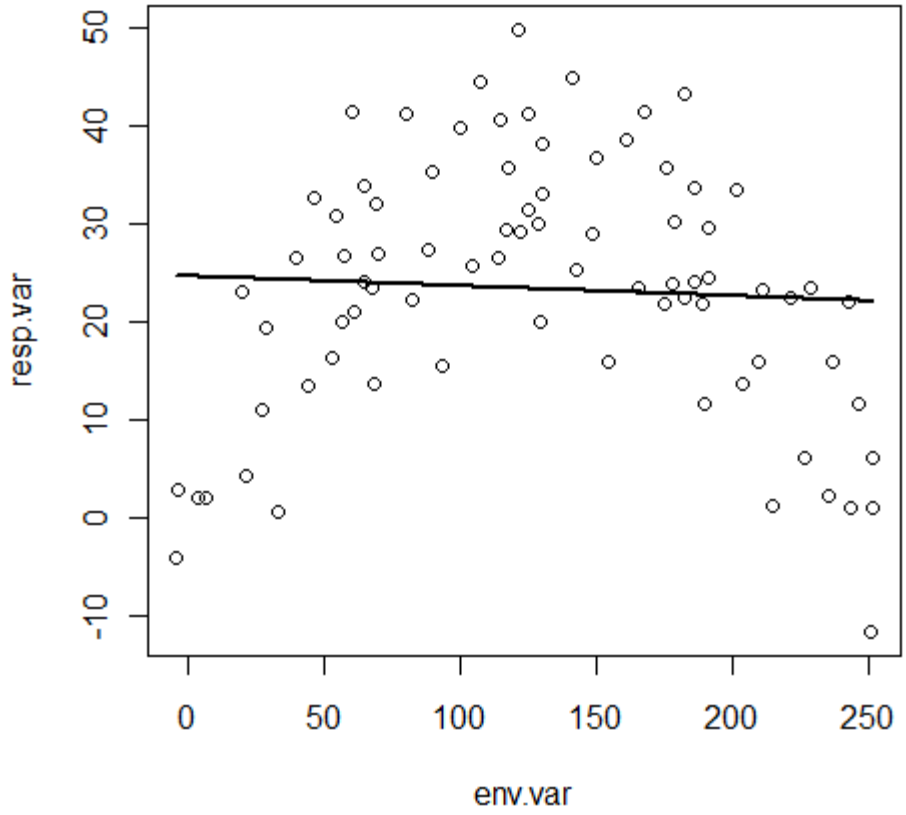



resp.var
-10
0
10
20
30
40
0
50
100
150
200
250
env.var


resp.var
-10
0
10
20
30
40
0
50
100
150
200
250
env.var

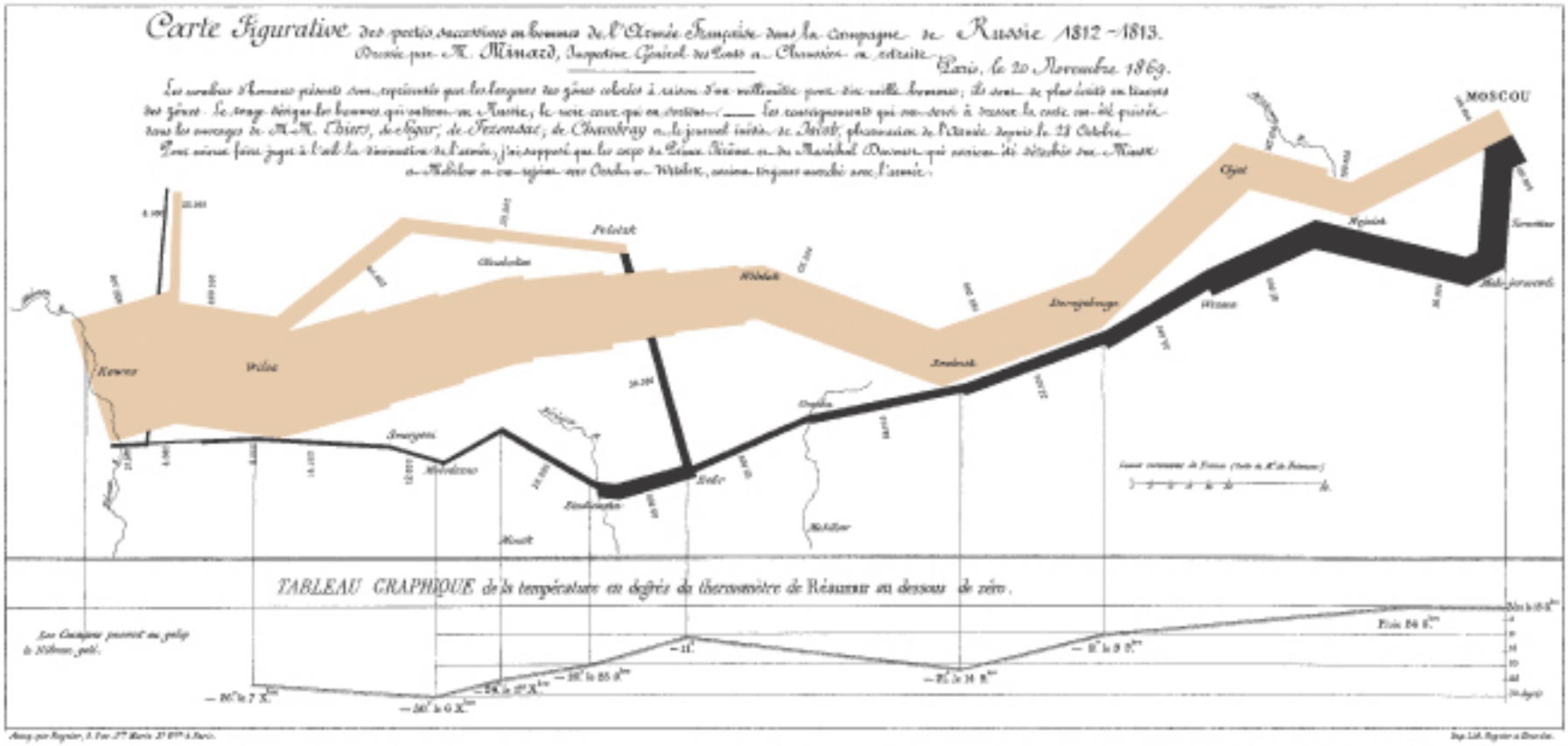

Residuals vs Fitted
Residuals
20
10
0
-10
-20
14
13
1
0
10
20
30
Fitted values
lm(resp.var ~ env.var)


Residuals vs Fitted
Residuals
30
20
10
0
-10
-20
-30
-40
35
78
2
22.5
23.0
23.5
24.0
24.5
Fitted values
lm(resp.var ~ env.var)


resp.var
env.var
-10
0
10
20
30
40
50
0
50
100
150
200
250

Bad model fit:

a) The wrong distribution
b) Key missing variables
c) Poor model set up



Napoleon's March to Moscow The War of 1812

Charles Joseph Minard

This classic of Charles Joseph Minard (1781–1870), the French engineer, shows the terrible fate of Napoleon's army in Russia. Described by E. J. Marey as seeming to defy the pen of the historian by its brutal eloquence, this combination of data map and time-series, drawn in 1869, portrays the devastating losses suffered in Napoleon's Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Niemen River, the thick band shows the size of the army (422,000 men) as it invaded Russia in June 1812. The width of the band indicates the size of the army at each place on the map. In September, the army reached Moscow, which was by then sacked and deserted, with 100,000 men. The path of Napoleon's retreat from Moscow is depicted by the darker, lower band, which is linked to a temperature scale and dates at the bottom of the chart. It was a bitterly cold winter, and many froze on the march out of Russia. As the graphic shows, the crossing of the Berezina River was a disaster, and the army finally struggled back into Poland with only 10,000 men remaining. Also shown are the movements of auxiliary troops, as they sought to protect the rear and the flank of the advancing army. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. It may well be the best statistical graphic ever drawn.

Edward R. Tufte, The Visual Display of Quantitative Information Graphics Press Box 430 Cheshire, Connecticut 06410

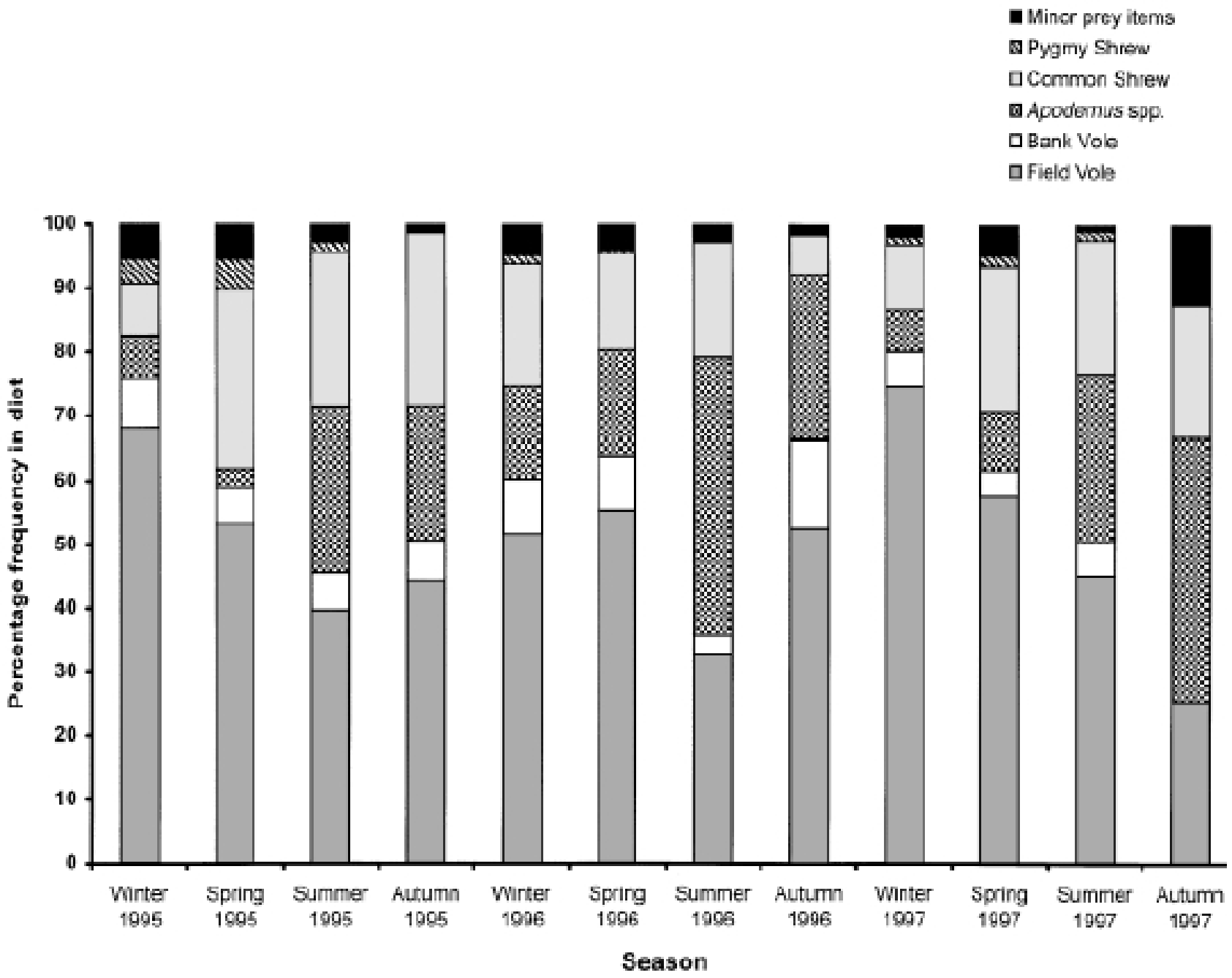


Fig. 2. Seasonal variation in the diet of Barn Owls in East Sussex 1995–97.

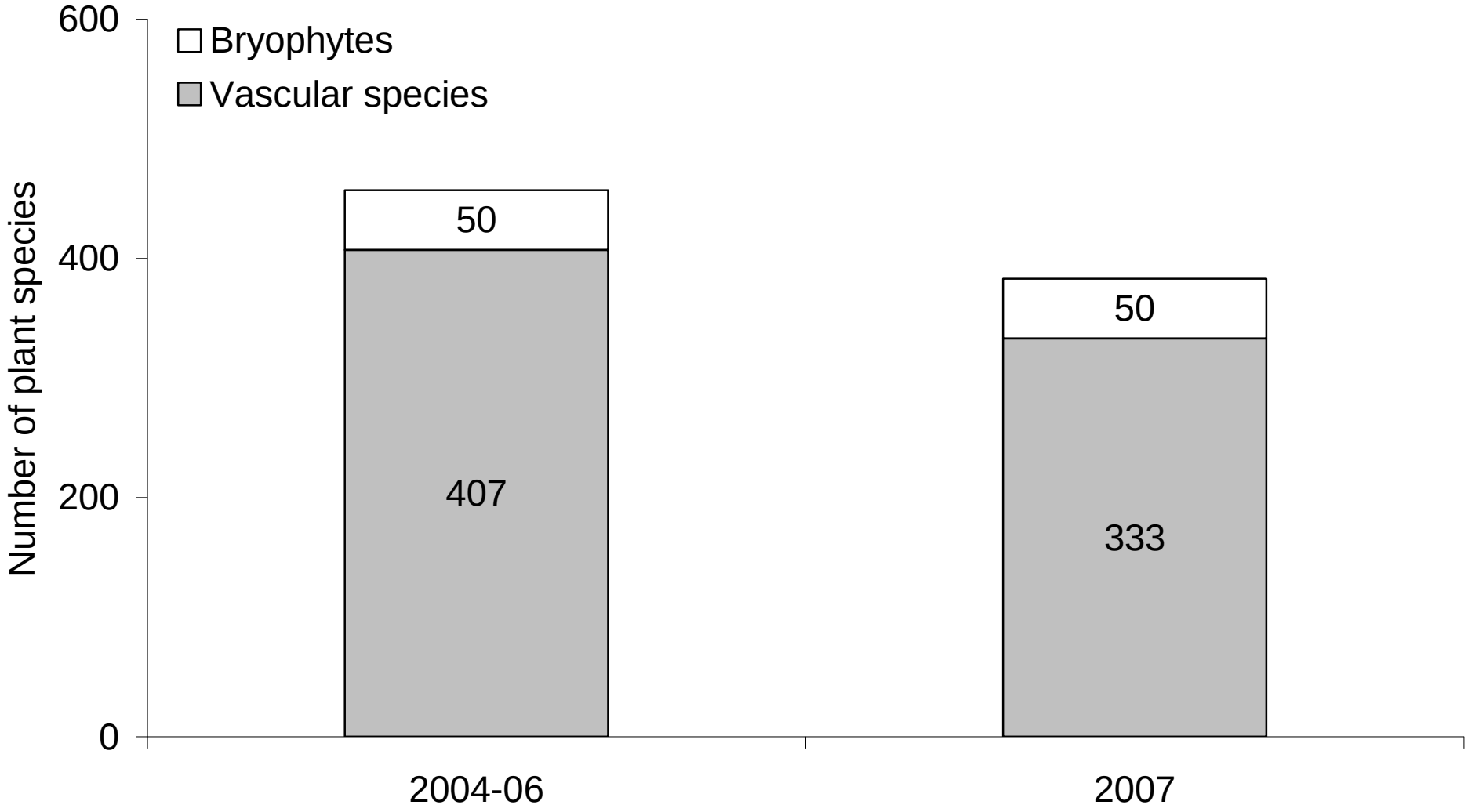
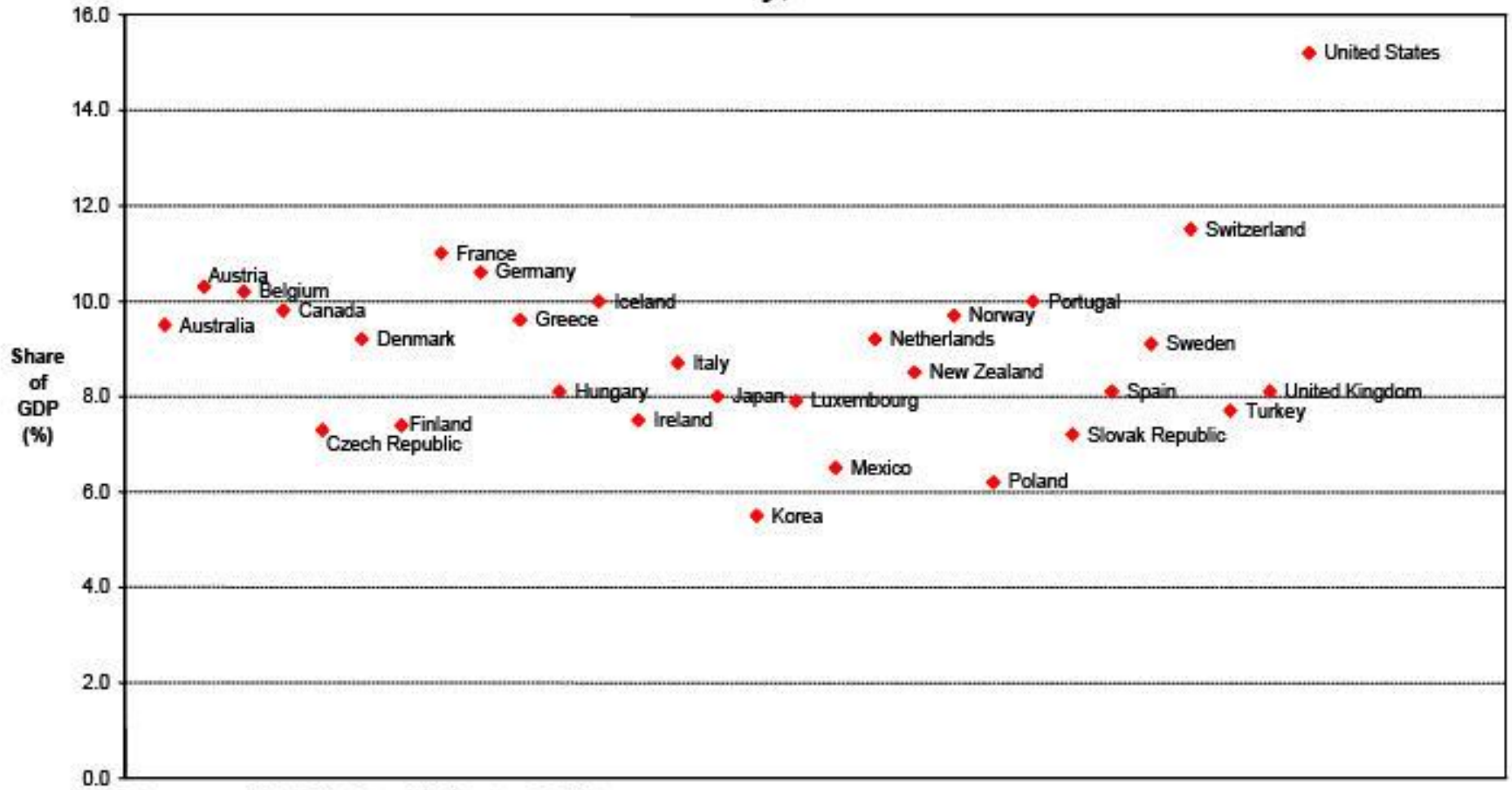


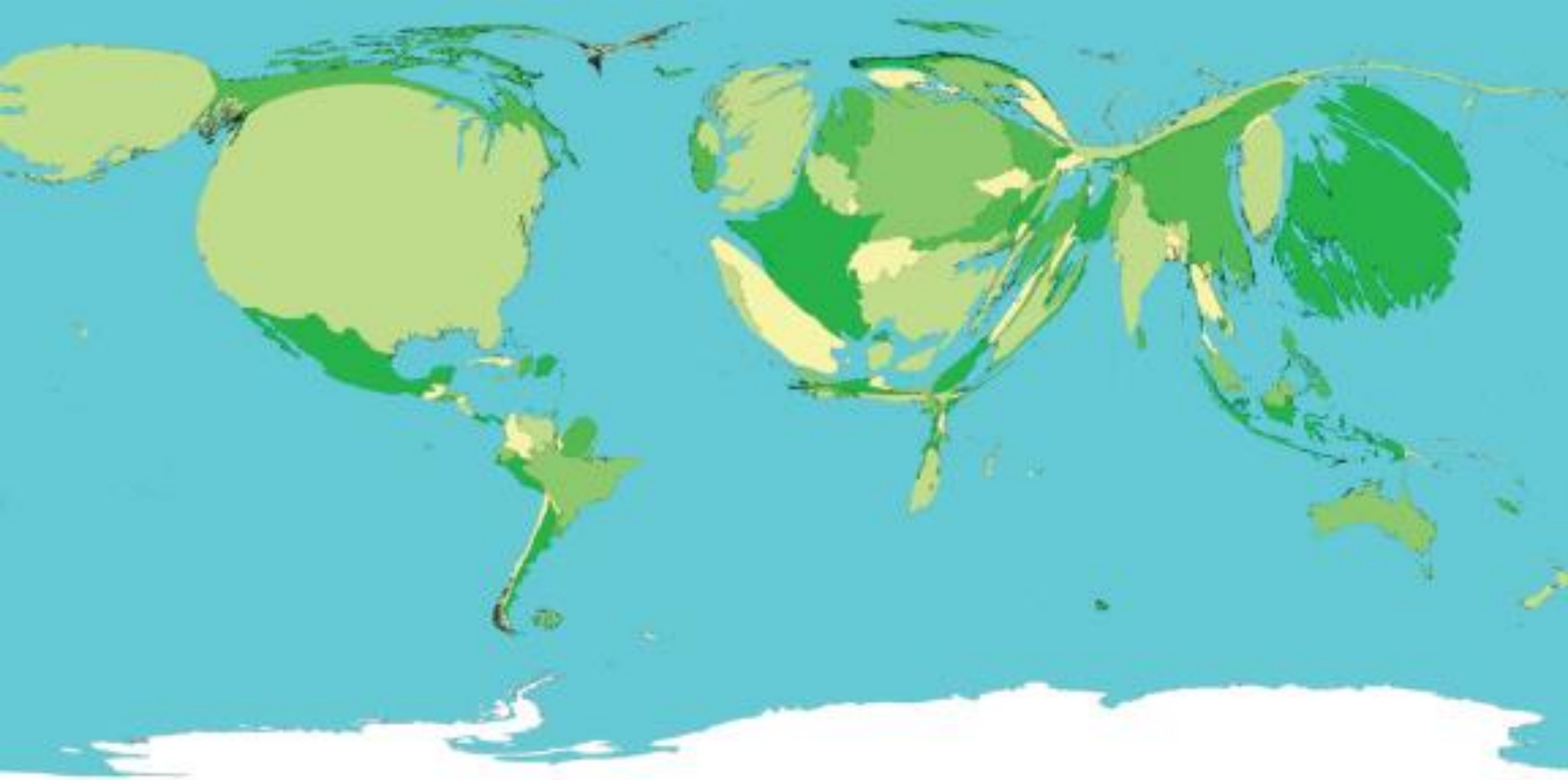
Figure 11: Comparison between plant species lists compiled by 24 groups of surveyors between March and July 2004 to 2006 and the plant species list compiled by one surveyor in May 2007 from 129 woods in Northern Ireland.

Chart 2 - Total Expenditures on Health as a Percentage Share of GDP, by OECD Country, 2004

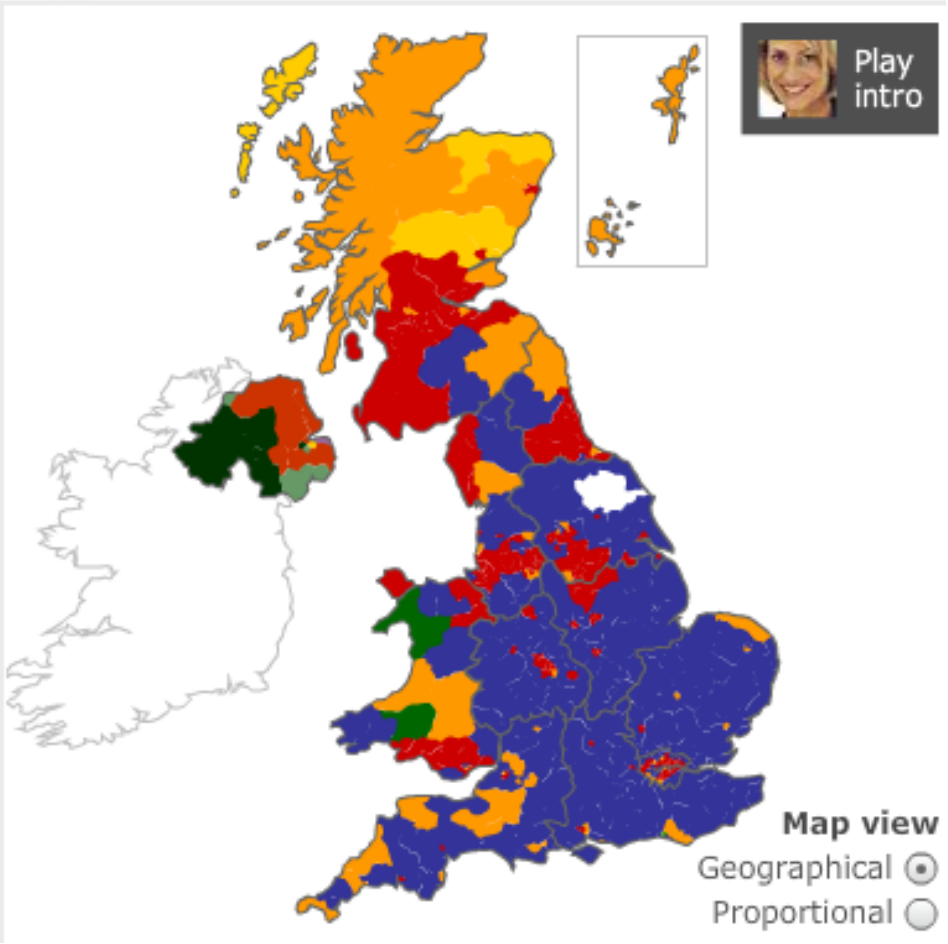


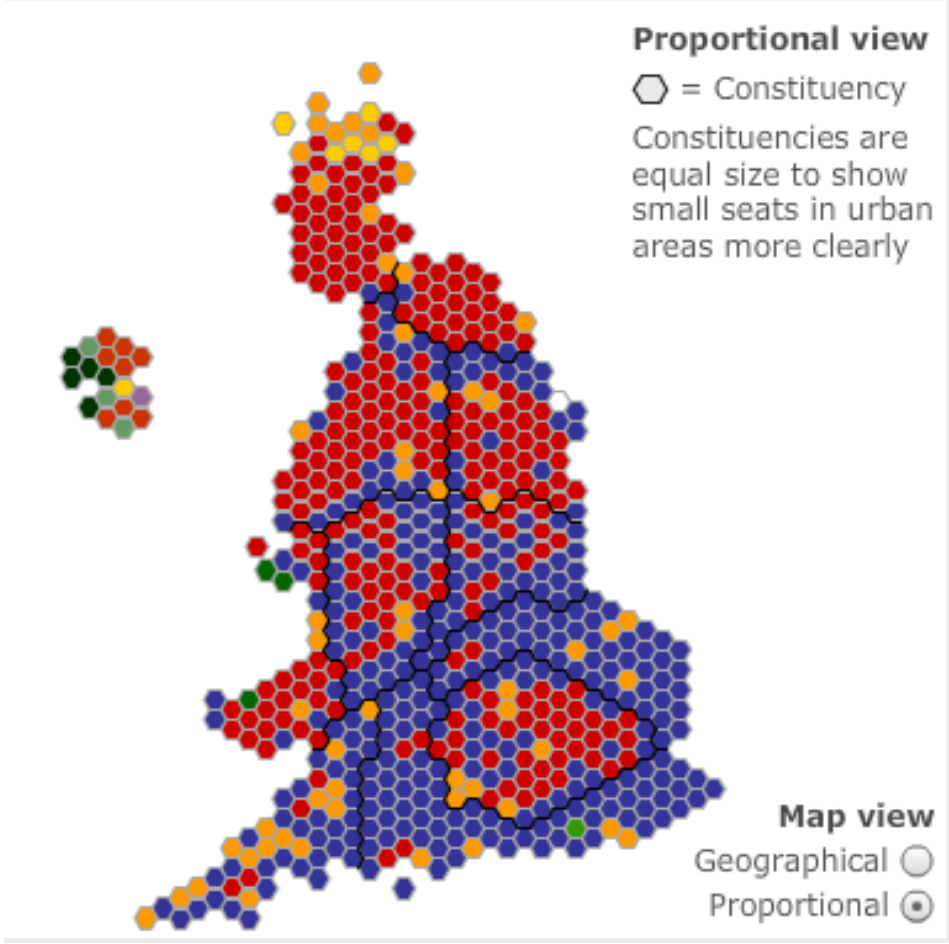
Source: OECD Health Data 2007.

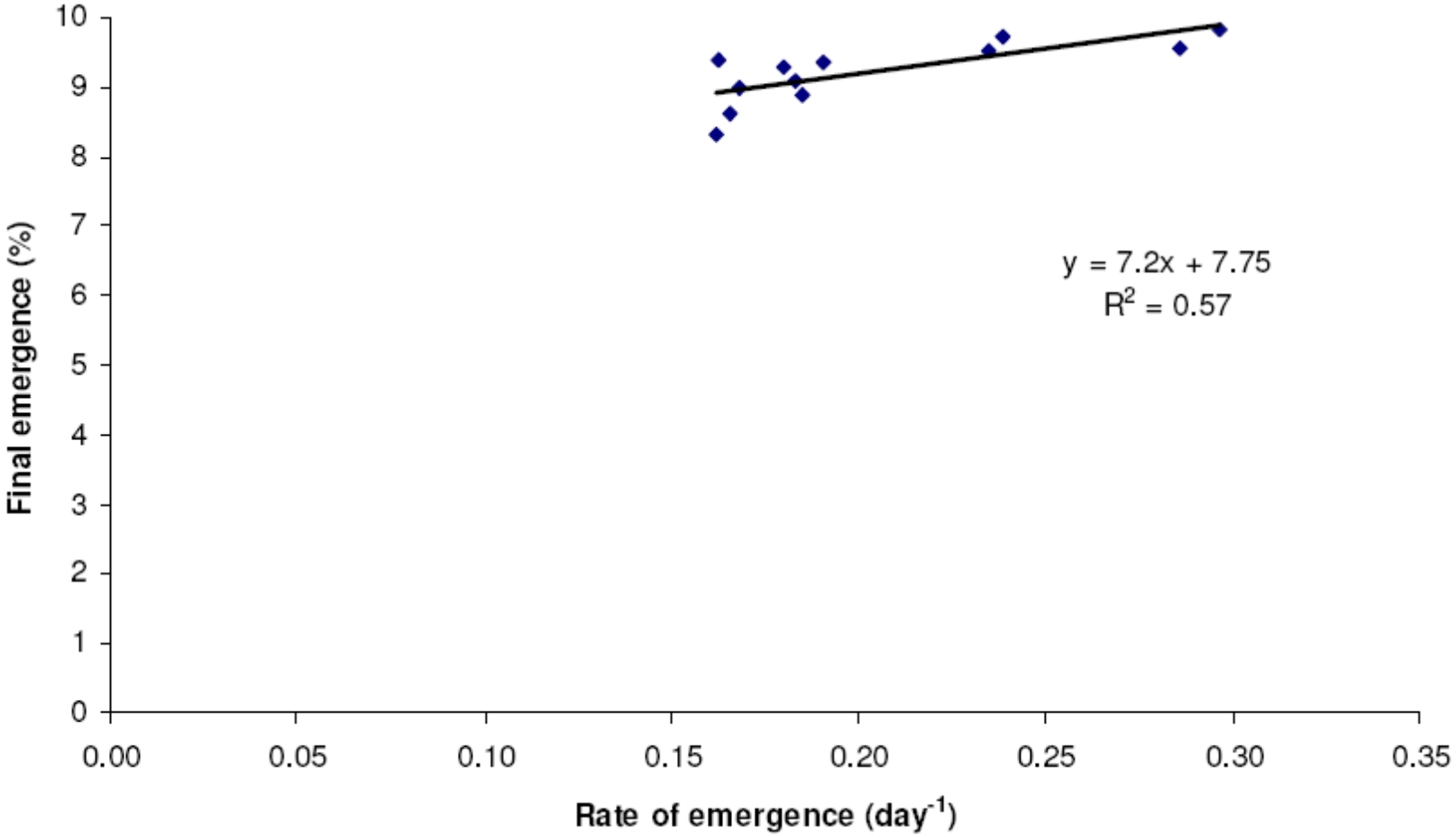
Note: For the United States the 2004 data reported here do not match the 2004 data point for the United States in Chart 1 since the OECD uses a slightly different definition of "total expenditures on health" than that used in the National Health Expenditure Accounts.



Map of the world according to the nations GDP
Copyright: Mark Newman, Department of Physics and Center for the Study of Complex Systems, University of Michigan.
URL: http://www.personal.umich.edu/~mejn/cartograms/gdp1024x512.png


Play intro
Map view
Geographical
Proportional


Proportional view
= Constituency
Constituencies are equal size to show small seats in urban areas more clearly
Map view
Geographical
Proportional


Final emergence (%)
10
9
8
7
6
5
4
3
2
1
0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Rate of emergence (day-1)
y = 7.2x + 7.75
R2 = 0.57