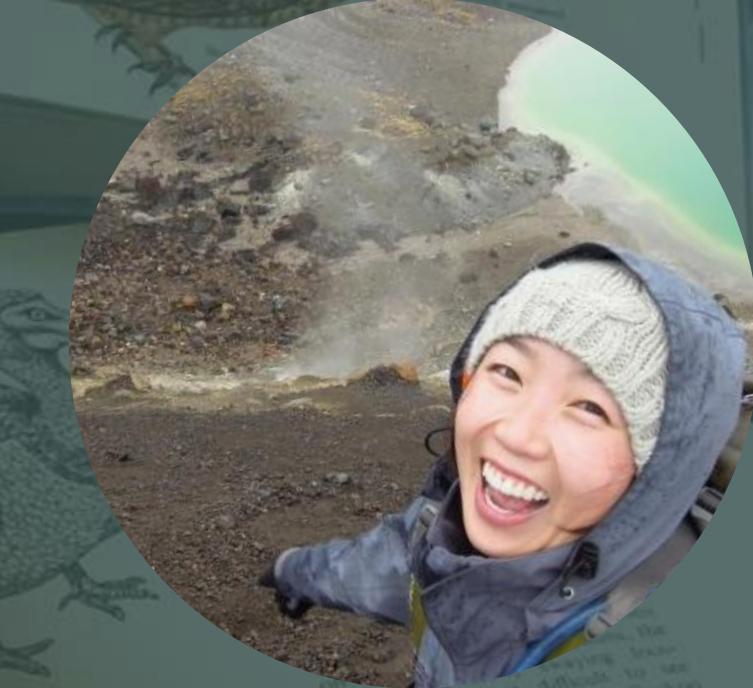


Dr. Fonti Kar (she/her)

University of New South Wales



Field Guide to Playful, Reproducible, and Inclusive Approaches to Teaching Data Science

Who am I?



Who am I?



Who am I?



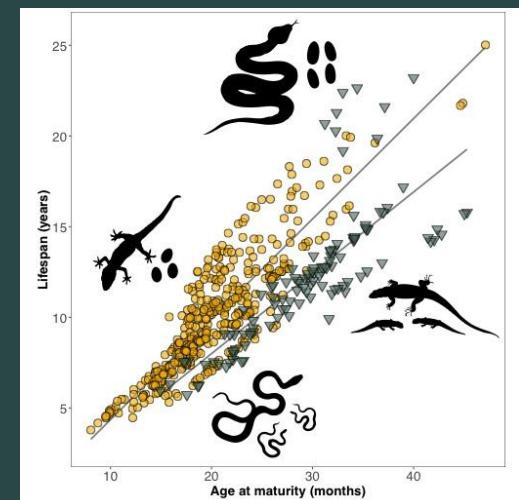
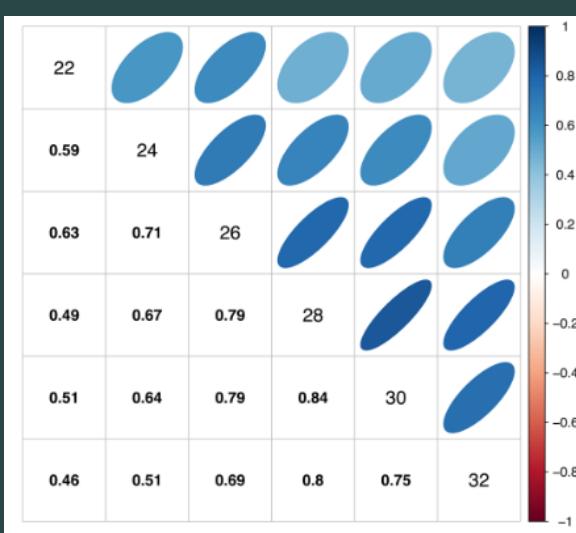
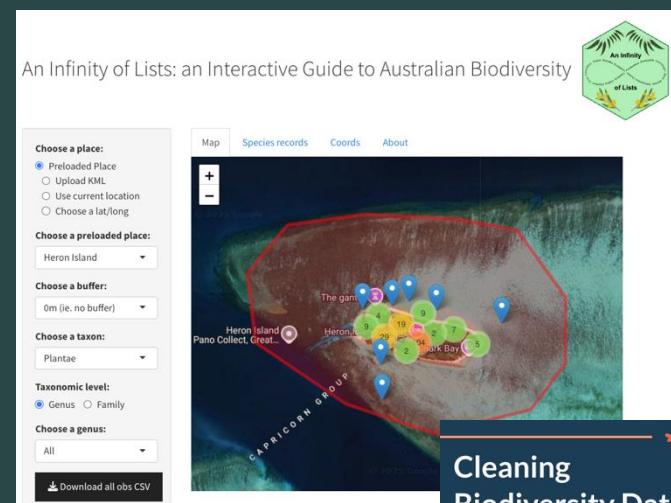
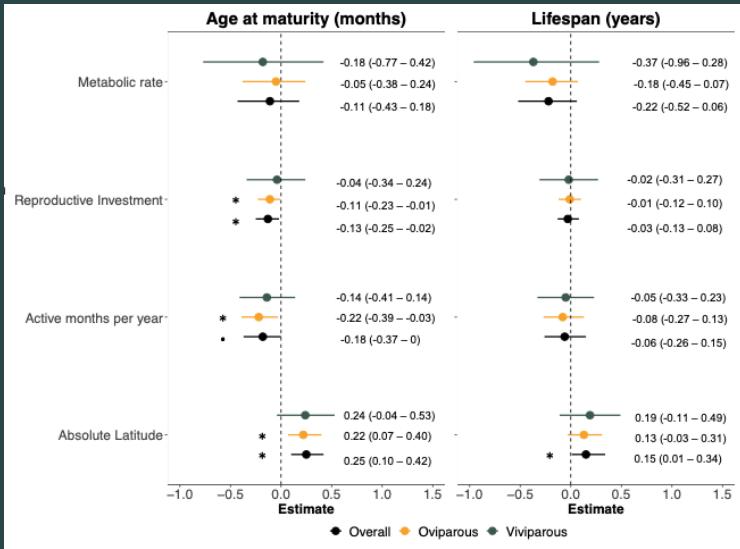
Who am I?
Why am I here?

> I love doing data science with R



Who am I?
Why am I here?

> I love solving data problems



Who am I?
Why am I here?

australs: : CHEAT SHEET

Explore the data
Use the following functions to explore and search the database to get a sense of trait/taxonomic coverage.

Download data
We recommend storing the Austraits database in your project folder e.g. in a `data/australs/` folder.
The path of this folder is important for loading the database into R without having to download the files again.

`get_versions(path)`

Retrieve all version information for the Austraits database

`date n_records doi version`

2024-05 22,764 95 4803

2023-11 10,538 10,519 2 2358

`n_taxa`

10386 53 2002

Search for terms with exact/part string matches in trait names/location properties/context properties

Functions return vector of string matches

`lookup_trait(database, trait_name = "all")`, `include_description = TRUE`

`lookup_methods(database, vars = c("methods"))`

`join_taxonomic_updates(database, vars = c("aligned_name"))`

`join_taxa(database, vars = c("family", "genus", "taxon_rank", "establishment_means"))`

`join_contributors(database, format = "single_column_pretty", vars = "all")`

`flatten_databases(database, ...)`

`bind_databases(database, ...)`

`binds multiple databases into one`

Structure
traits.build databases are **relational**. This means its data is organised in 13 tables as a list in R.

Names of the tables:

- traits
- locations
- contexts
- contributors
- excluded_data
- data
- contributors
- operations
- permissions
- metafiles
- build_info

Every record in the database is linked to across tables by a unique combination of identifiers.

You can access each table using \$

`australs$traits`

`australs$contexts`

Extract data

Depending on input, functions returns a subset

`extract_trait(database, trait_name)`

`extract_data(database, table, col, col_value)`

`extract_dataset(database, dataset_id)`

`extract_specific_dataset`

`australs %>% extract_dataset("Trait_2022")`

`extract_taxa(database, taxon_name/genus/family)`

`Extract specific taxa`

`australs %>% extract_taxa(name_property, "COT")`

`separate_trait_values(database$Trait)`

`Extract data that meet specified criteria`

`australs %>% extract_datatables(criteria_property, "COT")`

`Revert concatenated trait values back into multiple rows`

TRAIT VALUES

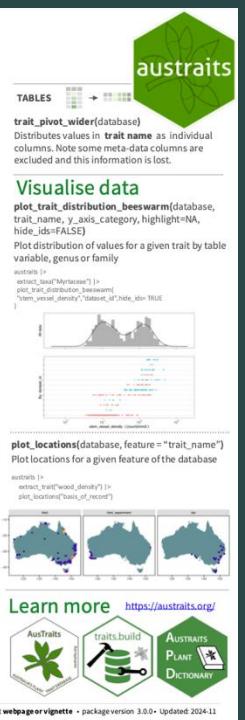
`bind_trait_values(database$Trait)`

`Concatenates trait values from measurements of the same observation into a string`

`separate_trait_values(database$Trait)`

`Revert concatenated trait values back into multiple rows`

Learn more <https://australs.org/>



> I love empowering others



Who am I?
Why am I here?

University of Auckland
BSc Biology + Statistics

UNSW + ANU
PhD



Macquarie University
MRes



UNSW
Postdoc



Atlas of Living Australia
Data Analyst

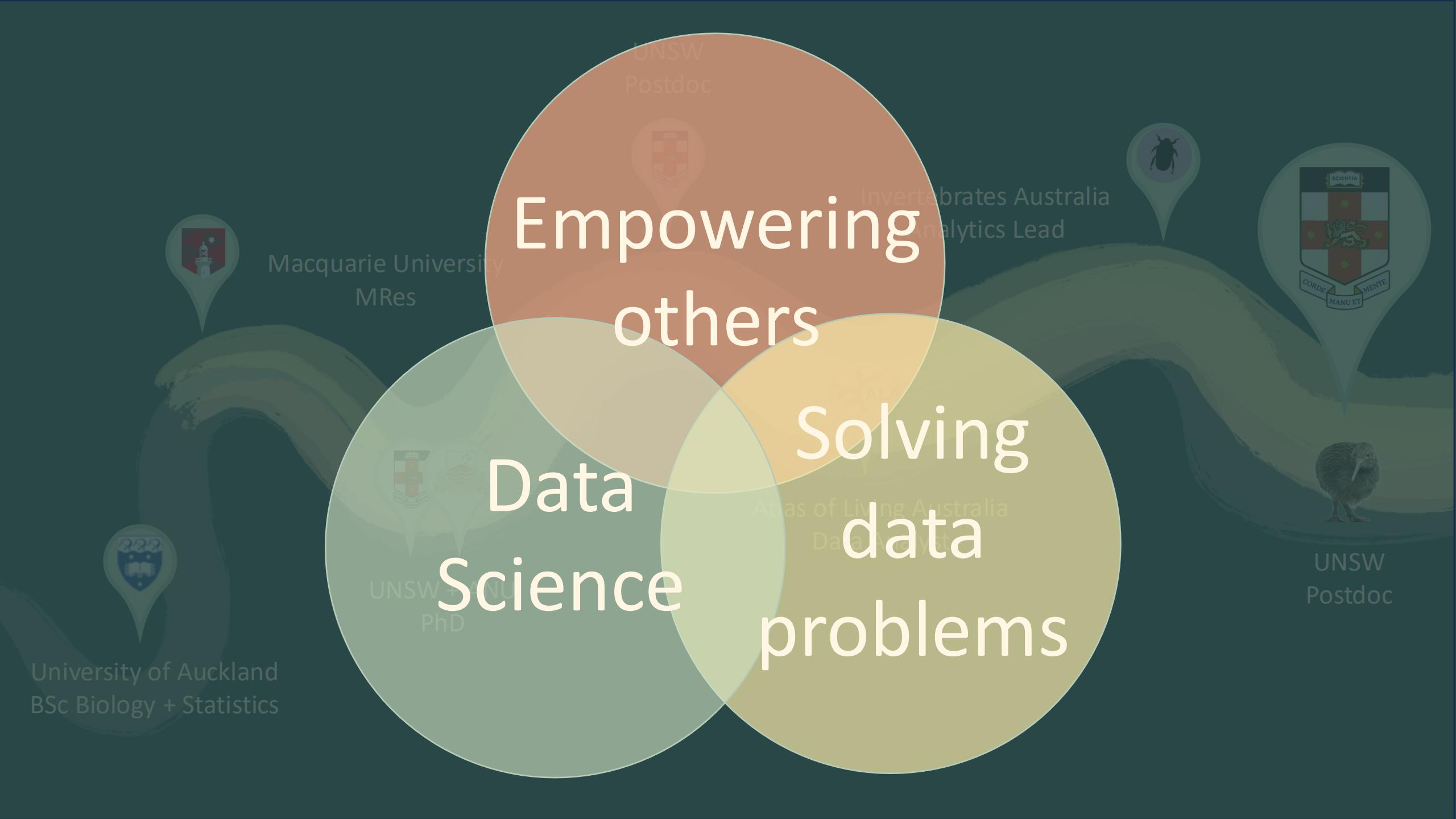


Invertebrates Australia
Analytics Lead



UNSW
Postdoc





Empowering others

Data
Science

Solving
data
problems

University of Auckland
BSc Biology + Statistics

Macquarie University
MRes

UNSW
Postdoc

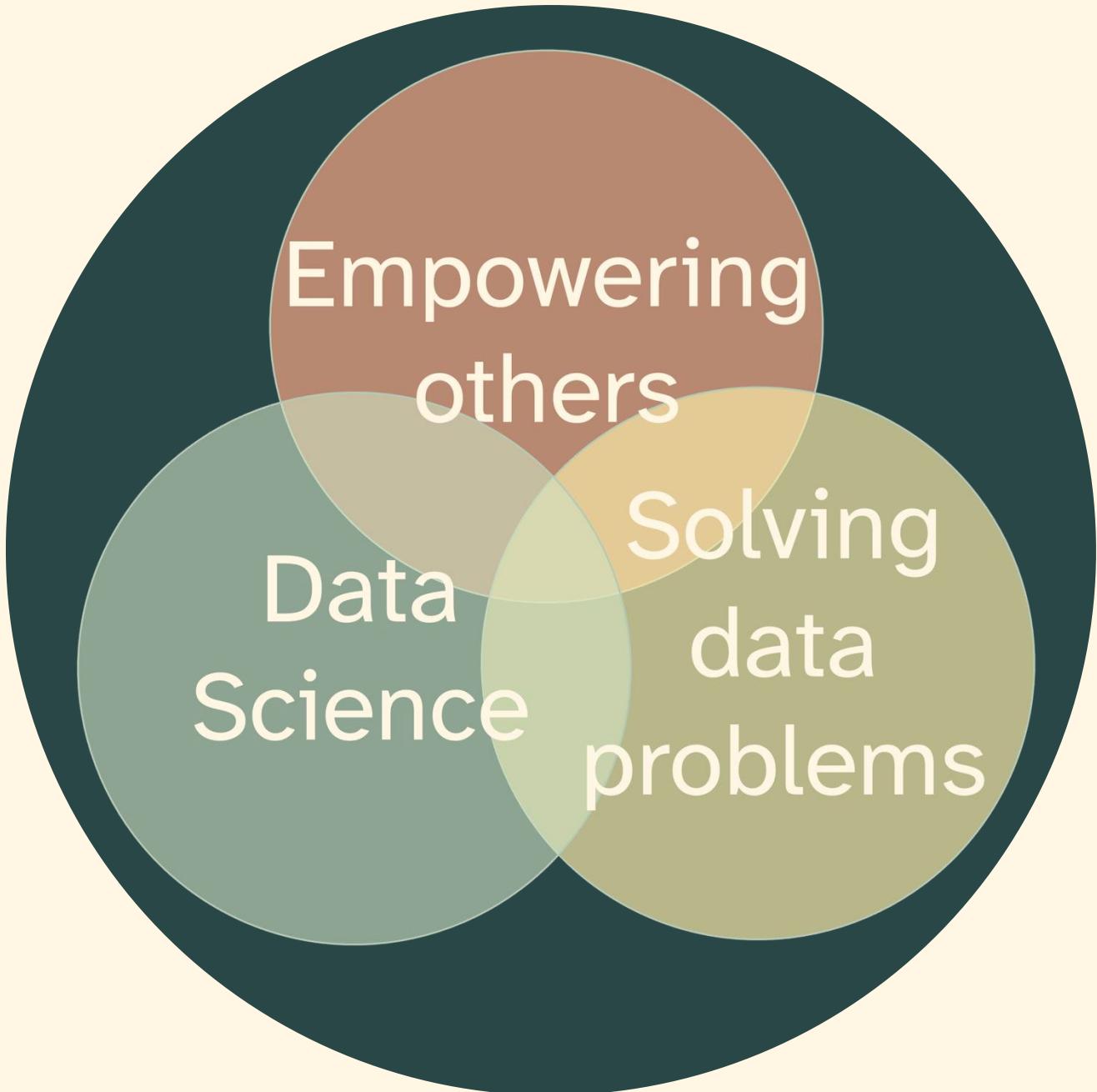
Invertebrates Australia
Analytics Lead

Atlas of Living Australia
Data Analyst

UNSW + ANU
PhD

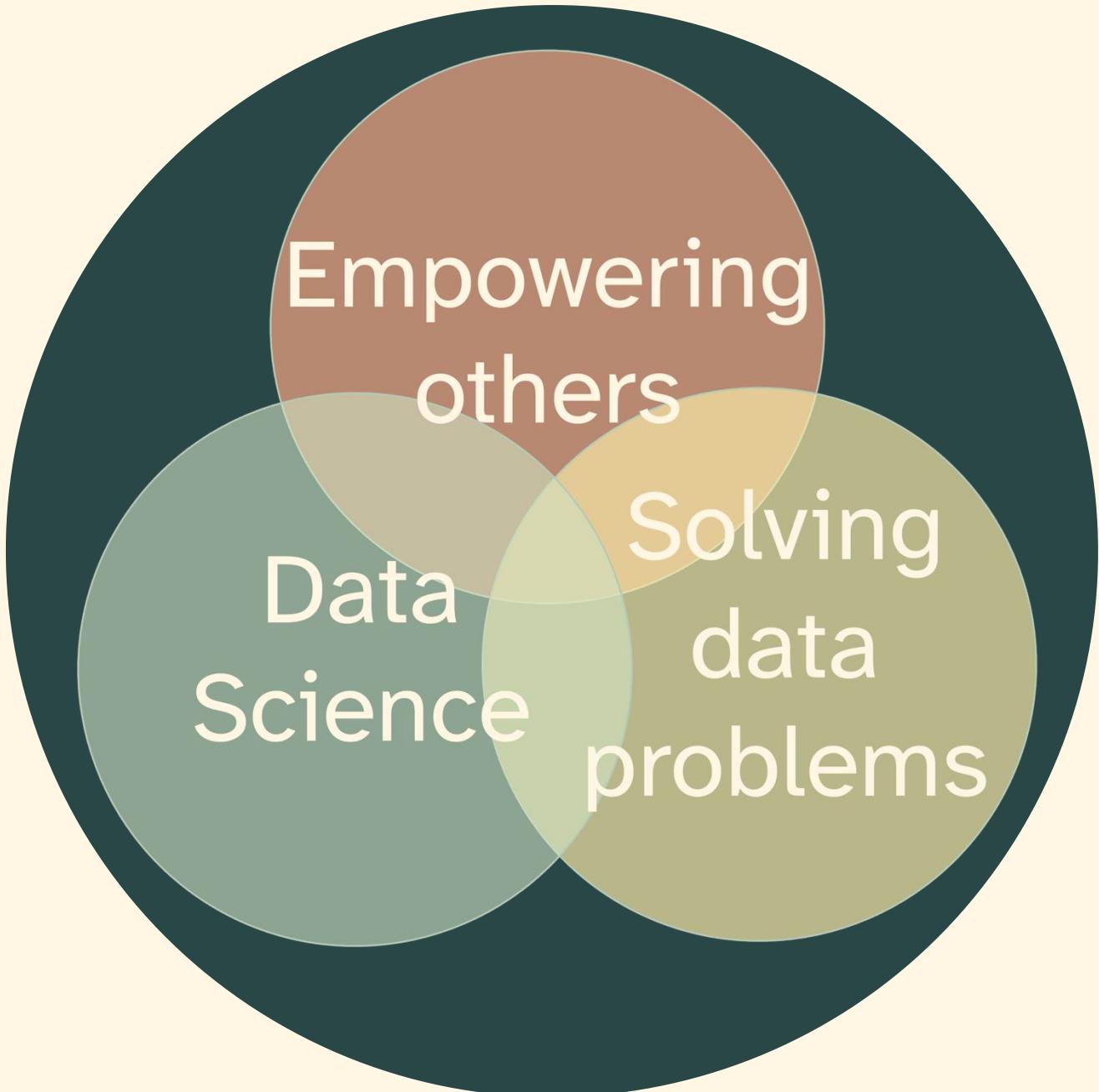
UNSW
Postdoc

Open
Reproducible
Transparent



Why?

- > Advance knowledge
- > Strengthens trust
- > Equal access
- > Safeguard future ‘us’





UNSW +
ANU
PhD
Evolutionary Biology





[Prof. Shinichi Nakagawa](#)



[A/Prof. Daniel Noble](#)

This
Idea
Must
Die

EDGE.ORG Presents Today's
Leading Thinkers

Jared Diamond, Richard Thaler,
Steven Pinker, Richard Dawkins,
Nassim Nicholas Taleb,
Ian McEwan, Rebecca Newberger
Goldstein, Freeman Dyson,
Daniel Goleman, Alan Guth,
and more

"A mind-blowing gathering
of innovative thinkers."
—Booklist

Scientific
Theories That
Are Blocking
Progress



Edited by
JOHN
BROCKMAN
Editor of
This Explains Everything

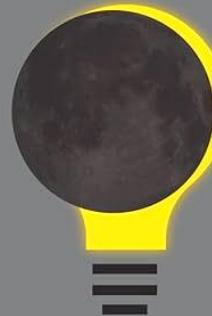
This
Idea
Must
Die

EDGE.ORG Presents Today's
Leading Thinkers

Jared Diamond, Richard Thaler,
Steven Pinker, Richard Dawkins,
Nassim Nicholas Taleb,
Ian McEwan, Rebecca Newberger
Goldstein, Freeman Dyson,
Daniel Goleman, Alan Guth,
and more

“A mind-blowing gathering
of innovative thinkers.”
—Booklist

Scientific
Theories That
Are Blocking
Progress



Edited by
JOHN
BROCKMAN
Editor of
This Explains Everything

THE AVERAGE

NICHOLAS A. CHRISTAKIS

Goldman Family Professor of Social and Natural Science, Yale University; coauthor (with James H. Fowler), Connected

Ever since the landmark invention of diverse statistical techniques 100 years ago that let us properly compare the difference between the averages of two groups, we have deluded ourselves into thinking that it's such differences that are the salient—and often the only—important one between groups. We've spent a century observing and interpreting such differences. We've become almost obsessed, and we should stop.

Yes, we can reliably say that men are taller than women, on average; that Norwegians are richer than Swedes; that first-born children are smarter than second-born children. And we can do experiments to detect tiny differences in means—between groups exposed and unexposed to a virus, or between groups with and without a particular allele of a gene. But this is too simple and too narrow a view of the natural world.

Our focus on averages should be retired. Or, if not retired, given an extended vacation. During this vacation, we should catch up on another sort of difference between groups which has gotten short shrift: We should focus on comparing the difference in variance—which captures the spread, or range, of measured values—between groups.

Part of the reason we've focused so much on the average is that the statistical tools for computing and comparing averages are so much easier and well developed. It's much harder to compare whether the variance of one group is different from the variance of

This Idea Must Die

EDGE.ORG Presents Today's
Leading Thinkers

Jared Diamond, Richard Thaler,
Steven Pinker, Richard Dawkins,
Nassim Nicholas Taleb,
Ian McEwan, Rebecca Newberger
Goldstein, Freeman Dyson,
Daniel Goleman, Alan Guth,
and more

"A mind-blowing gathering
of innovative thinkers."

—Booklist

Scientific
Theories That
Are Blocking
Progress



Edited by
**JOHN
BROCKMAN**

Editor of
This Explains Everything

THE AVERAGE

NICHOLAS A. CHRISTAKIS

Goldman Family Professor of Social and Natural Science, Yale University; coauthor (with James H. Fowler), Connected

Ever since the landmark invention of diverse statistical techniques 100 years ago that let us properly compare the difference between the averages of two groups, we have deluded ourselves into thinking that it's such differences that are the salient—and often the only—important one between groups. We've spent a century observing and interpreting such differences. We've become almost obsessed, and we should stop.

Yes, we can reliably say that men are taller than women, on average; that Norwegians are richer than Swedes; that first-born children are smarter than second-born children. And we can do experiments to detect tiny differences in means—between groups exposed and unexposed to a virus, or between groups with and without a particular allele of a gene. But this is too simple and too narrow a view of the natural world.

Our focus on averages should be retired. Or, if not retired, given an extended vacation. During this vacation, we should catch up on another sort of difference between groups which has gotten short shrift: We should focus on comparing the difference in variance—which captures the spread, or range, of measured values—between groups.

Part of the reason we've focused so much on the average is that the statistical tools for computing and comparing averages are so much easier and well developed. It's much harder to compare whether the variance of one group is different from the variance of

This Idea Must Die

EDGE.ORG Presents Today's
Leading Thinkers

Jared Diamond, Richard Thaler,
Steven Pinker, Richard Dawkins,
Nassim Nicholas Taleb,
Ian McEwan, Rebecca Newberger
Goldstein, Freeman Dyson,
Daniel Goleman, Alan Guth,
and more

THE AVERAGE

NICHOLAS A. CHRISTAKIS

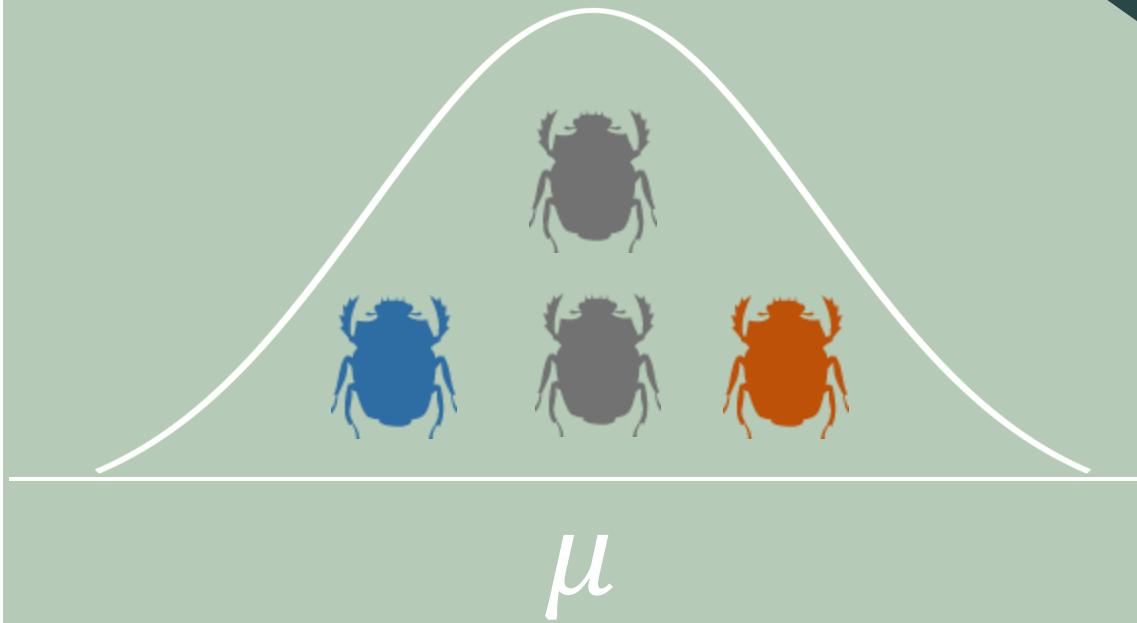
Goldman Family Professor of Social and Natural Science, Yale University; coauthor (with James H. Fowler), Connected

Ever since the landmark invention of diverse statistical techniques 100 years ago that let us properly compare the difference between the averages of two groups, we have deluded ourselves into thinking that it's such differences that are the salient—and often the only—important one between groups. We've spent a century observing and interpreting such differences. We've become almost obsessed, and we should stop.

Yes, we can reliably say that men are taller than women, on average; that Norwegians are richer than Swedes; that first-born children are smarter than second-born children. And we can do experiments to detect tiny differences in means—between groups exposed and unexposed to a virus, or between groups with and without a particular allele of a gene. But this is too simple and too narrow a view of the natural world.

Our focus on averages should be retired. Or, if not retired, given an extended vacation. During this vacation, we should catch up on another sort of difference between groups which has gotten short shrift: We should focus on comparing the difference in variance—which captures the spread, or range, of measured values—between groups.

Part of the reason we've focused so much on the average is that the statistical tools for computing and comparing averages are so much easier and well developed. It's much harder to compare whether the variance of one group is different from the variance of



> *Evolution requires variation*

This Idea Must Die

EDGE.ORG Presents Today's
Leading Thinkers

Jared Diamond, Richard Thaler,
Steven Pinker, Richard Dawkins,
Nassim Nicholas Taleb,
Ian McEwan, Rebecca Newberger
Goldstein, Freeman Dyson,
Daniel Goleman, Alan Guth,
and more

THE AVERAGE

NICHOLAS A. CHRISTAKIS

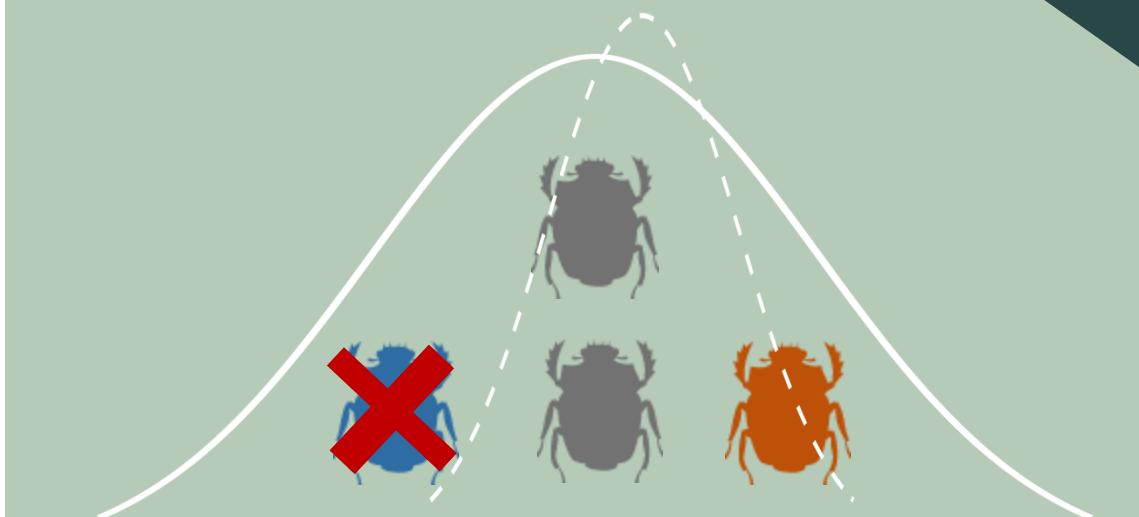
Goldman Family Professor of Social and Natural Science, Yale University; coauthor (with James H. Fowler), Connected

Ever since the landmark invention of diverse statistical techniques 100 years ago that let us properly compare the difference between the averages of two groups, we have deluded ourselves into thinking that it's such differences that are the salient—and often the only—important one between groups. We've spent a century observing and interpreting such differences. We've become almost obsessed, and we should stop.

Yes, we can reliably say that men are taller than women, on average; that Norwegians are richer than Swedes; that first-born children are smarter than second-born children. And we can do experiments to detect tiny differences in means—between groups exposed and unexposed to a virus, or between groups with and without a particular allele of a gene. But this is too simple and too narrow a view of the natural world.

Our focus on averages should be retired. Or, if not retired, given an extended vacation. During this vacation, we should catch up on another sort of difference between groups which has gotten short shrift: We should focus on comparing the difference in variance—which captures the spread, or range, of measured values—between groups.

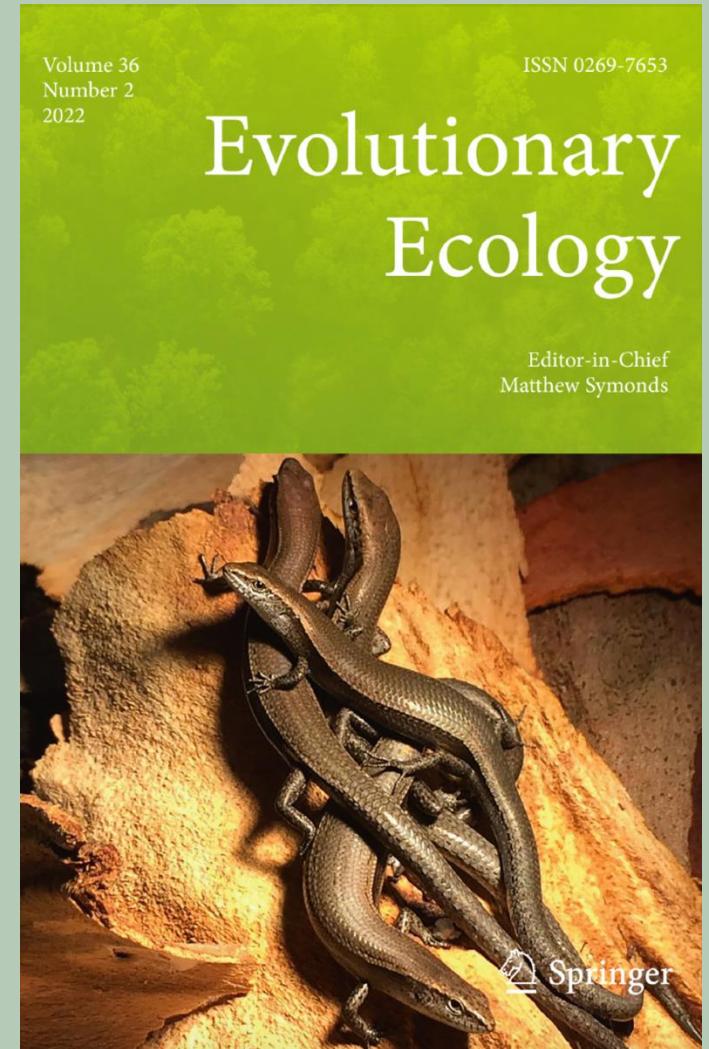
Part of the reason we've focused so much on the average is that the statistical tools for computing and comparing averages are so much easier and well developed. It's much harder to compare whether the variance of one group is different from the variance of



- > *Evolution requires variation*
- > *Natural selection erodes variation*

How will warming climate alter evolution?

> *Experimental design*



[Kar et al. 2021 Oikos](#)

[Kar et al. 2022 Evolutionary Ecology](#)

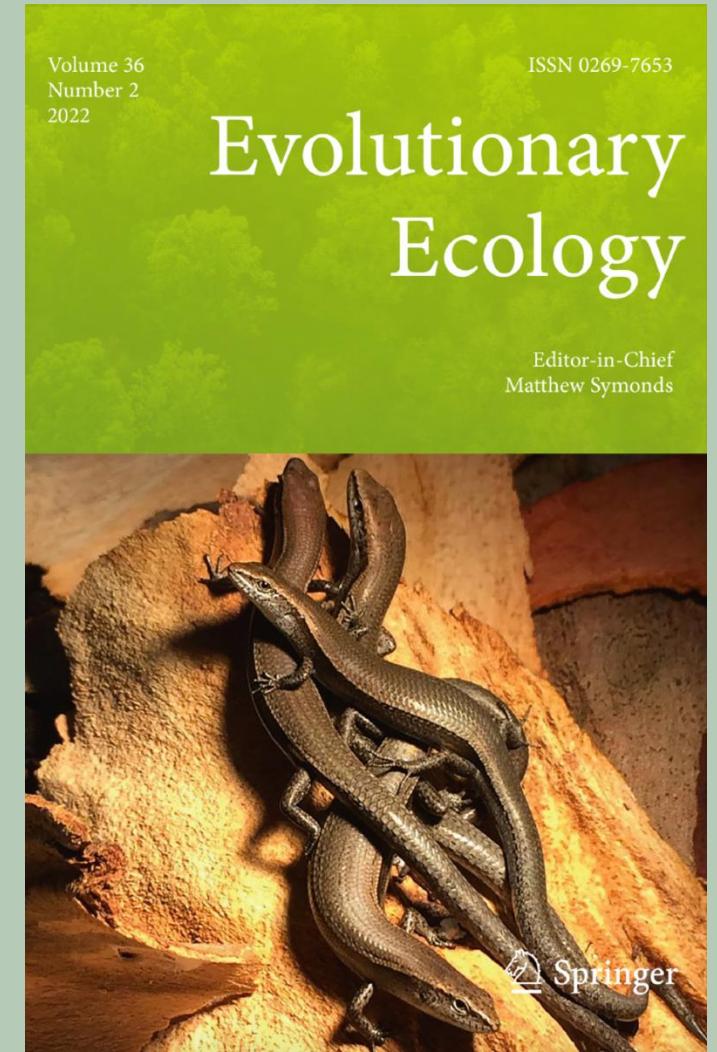
How will warming climate alter evolution?

> *Experimental design*



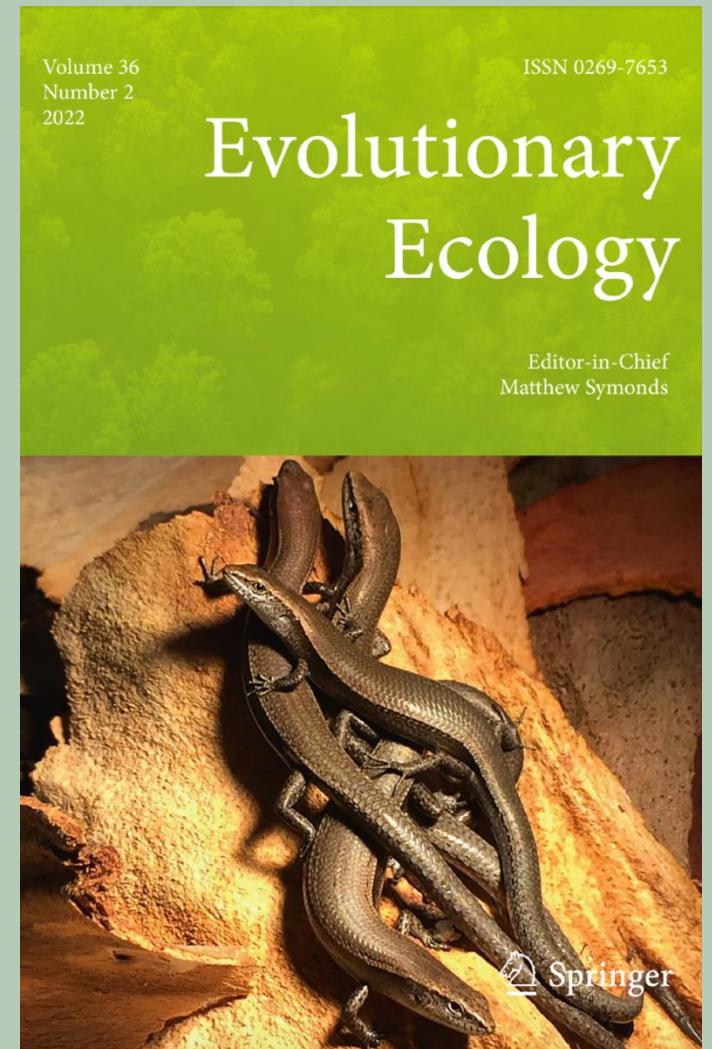
[Kar et al. 2021 Oikos](#)

[Kar et al. 2022 Evolutionary Ecology](#)



How will warming climate alter evolution?

> *Experimental design*

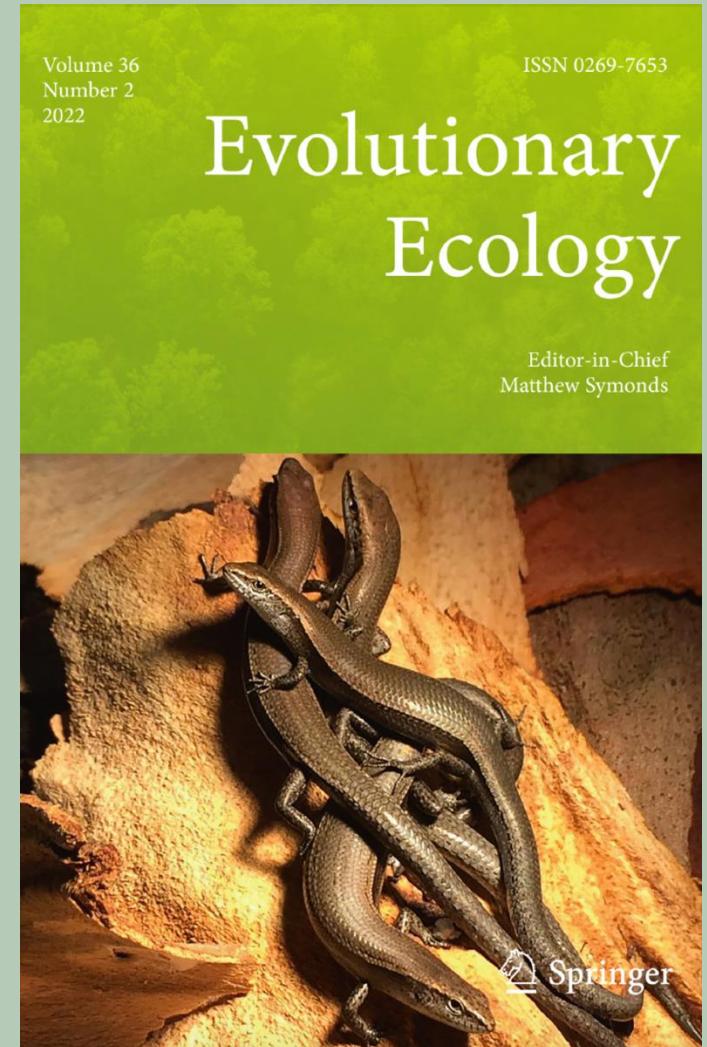
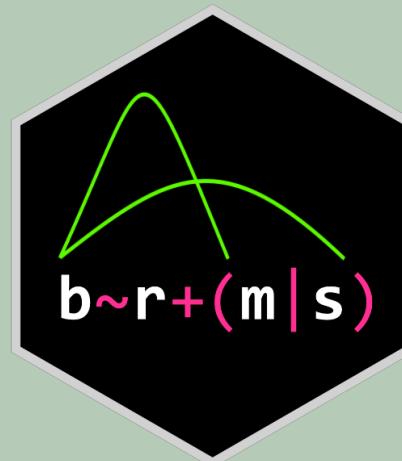


[Kar et al. 2021 Oikos](#)

[Kar et al. 2022 Evolutionary Ecology](#)

How will warming climate alter evolution?

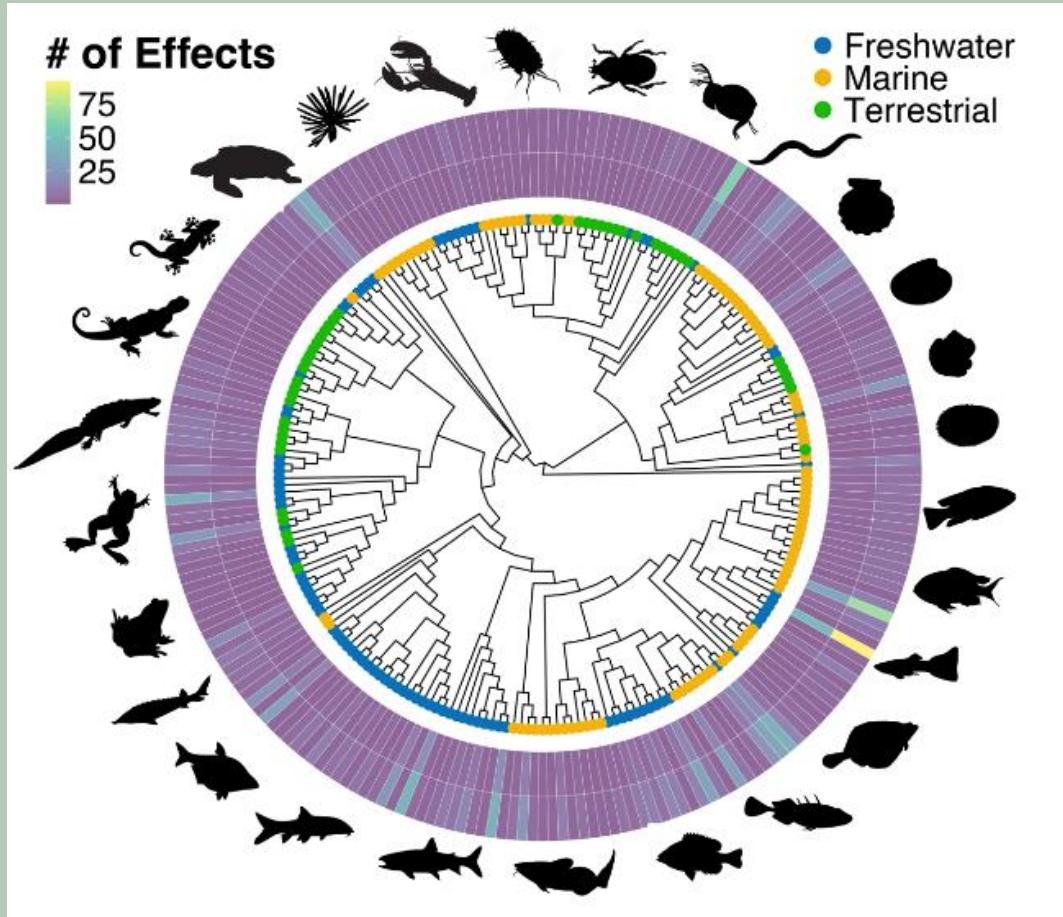
- > *Experimental design*
- > *Repeated measures*
- > *Bayesian hierarchical models*
- > *Model-based missing data imputation*



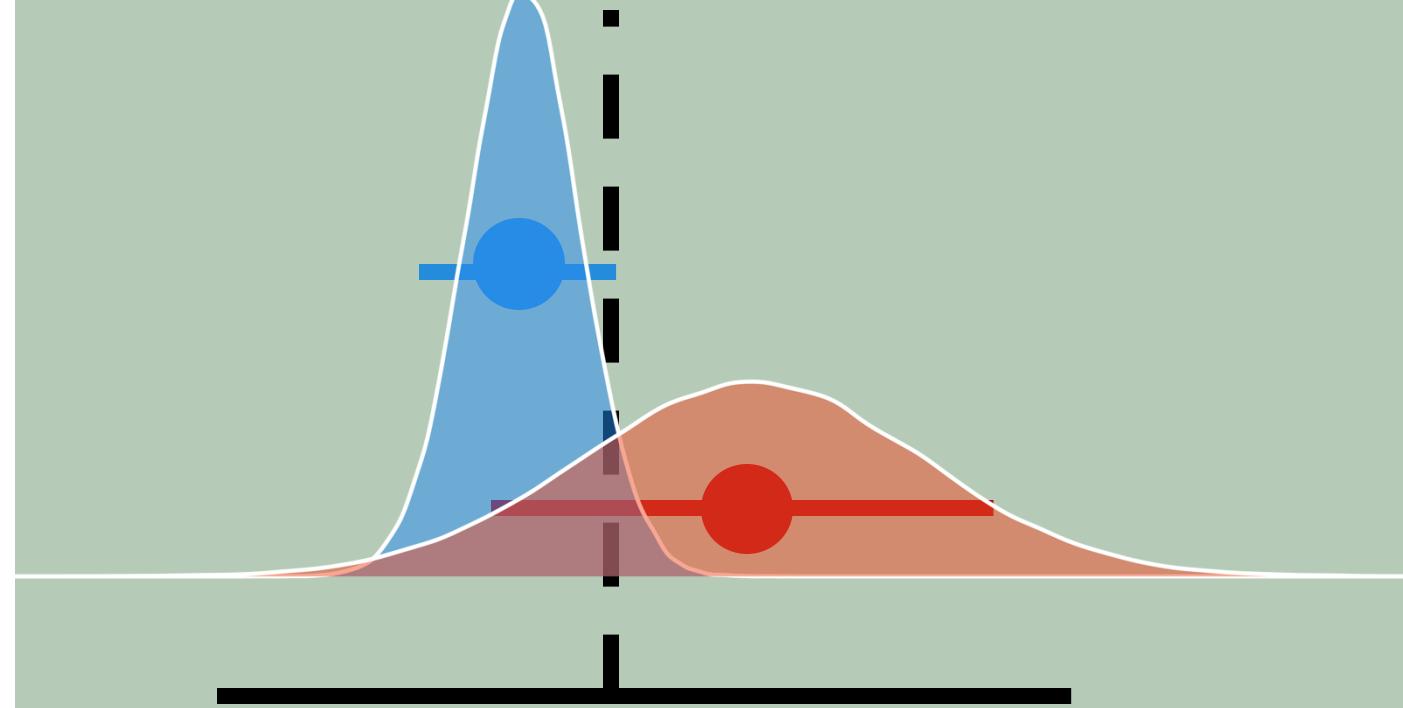
[Kar et al. 2021 Oikos](#)

[Kar et al. 2022 Evolutionary Ecology](#)

How will warming climate alter evolution?



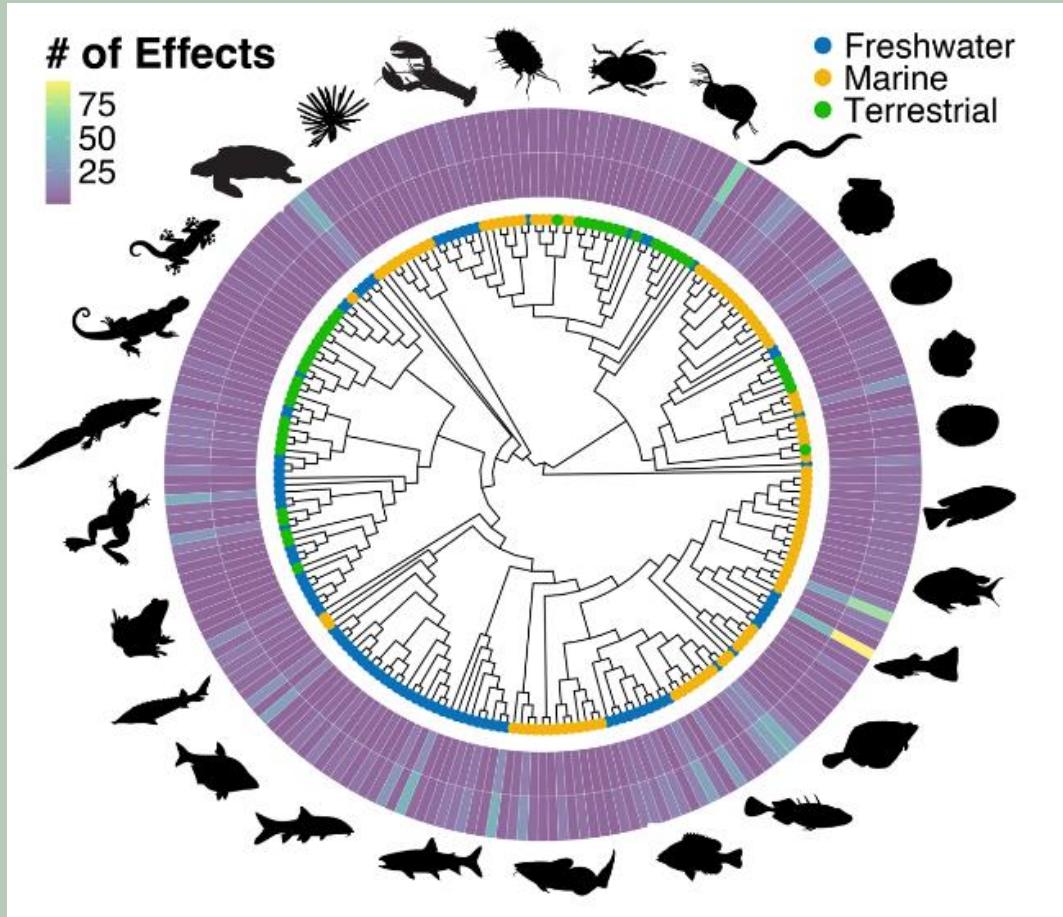
> *Meta-analysis of Variances*



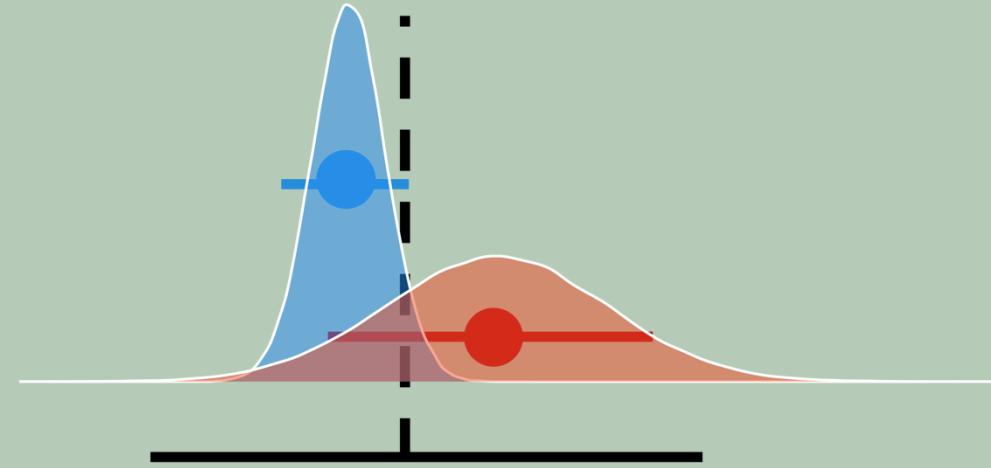
[Nakagawa et al. 2015](#)

[Noble et al. in press](#)

How will warming climate alter evolution?



> *Meta-analysis of Variances*



> *Bayesian hierarchical models*

$$n_{\text{studies}} = 256$$

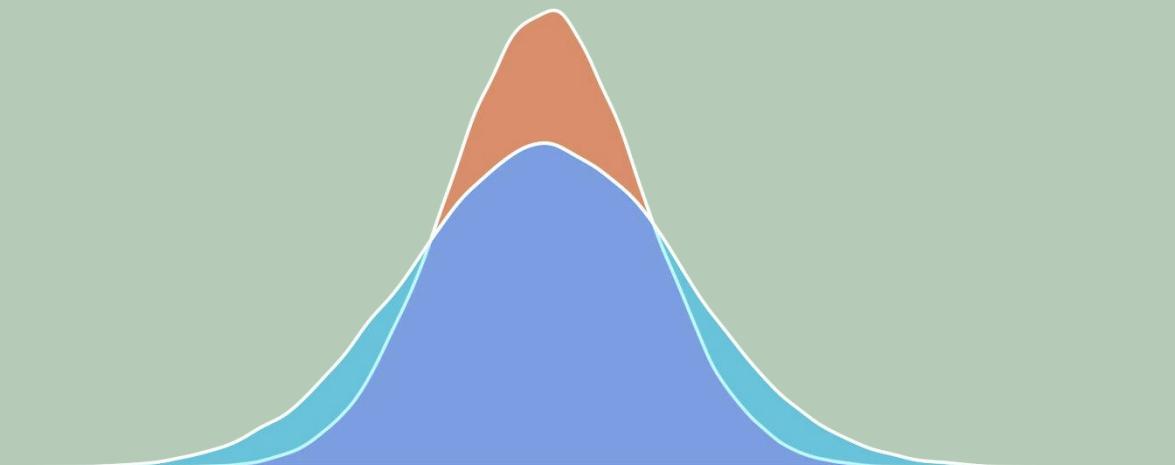
$$n_{\text{animals}} = 226$$

$$n_{\text{effect sizes}} = 1115$$

[Nakagawa et al. 2015](#)

[Noble et al. in press](#)

In my experiments,
variance **reduced**
in **warming** climate

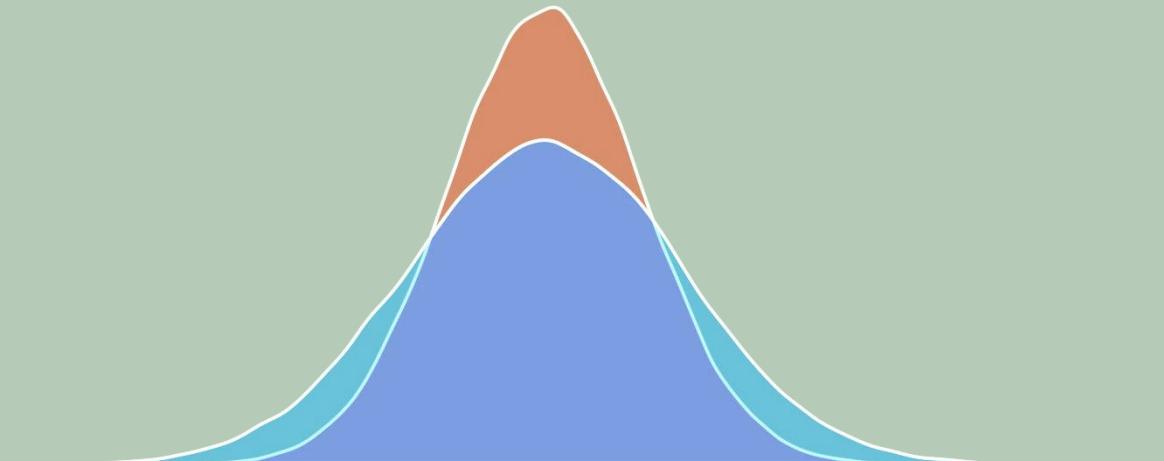


[Kar et al. 2021 Oikos](#)

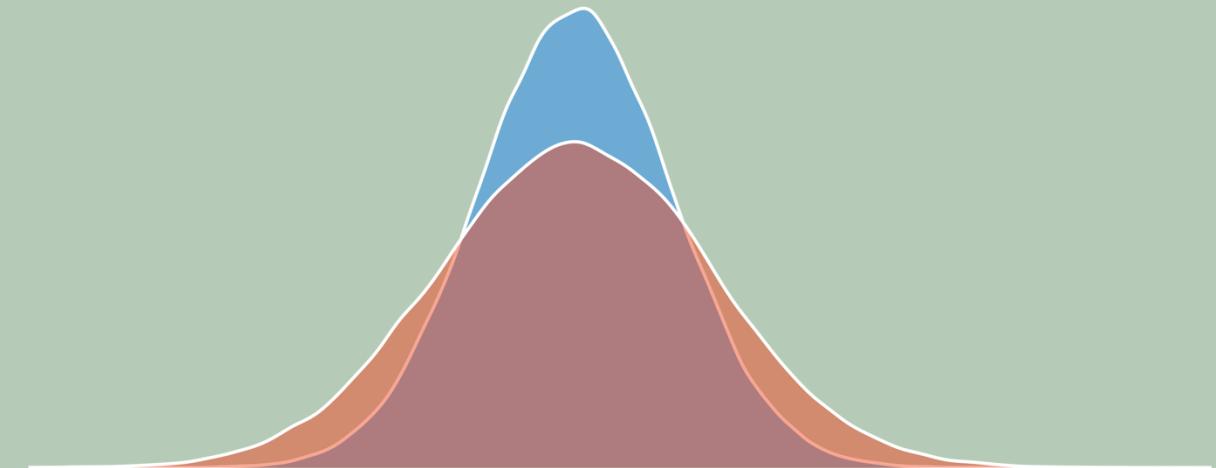
[Kar et al. 2022 Evolutionary Ecology](#)

[Noble et al. in press](#)

In my experiments,
variance **reduced**
in **warming** climate



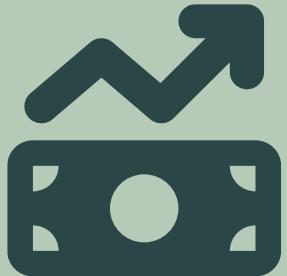
Meta-analysis showed
that variance **increased** in
warming climate*



*Terrestrial < Marine + Freshwater

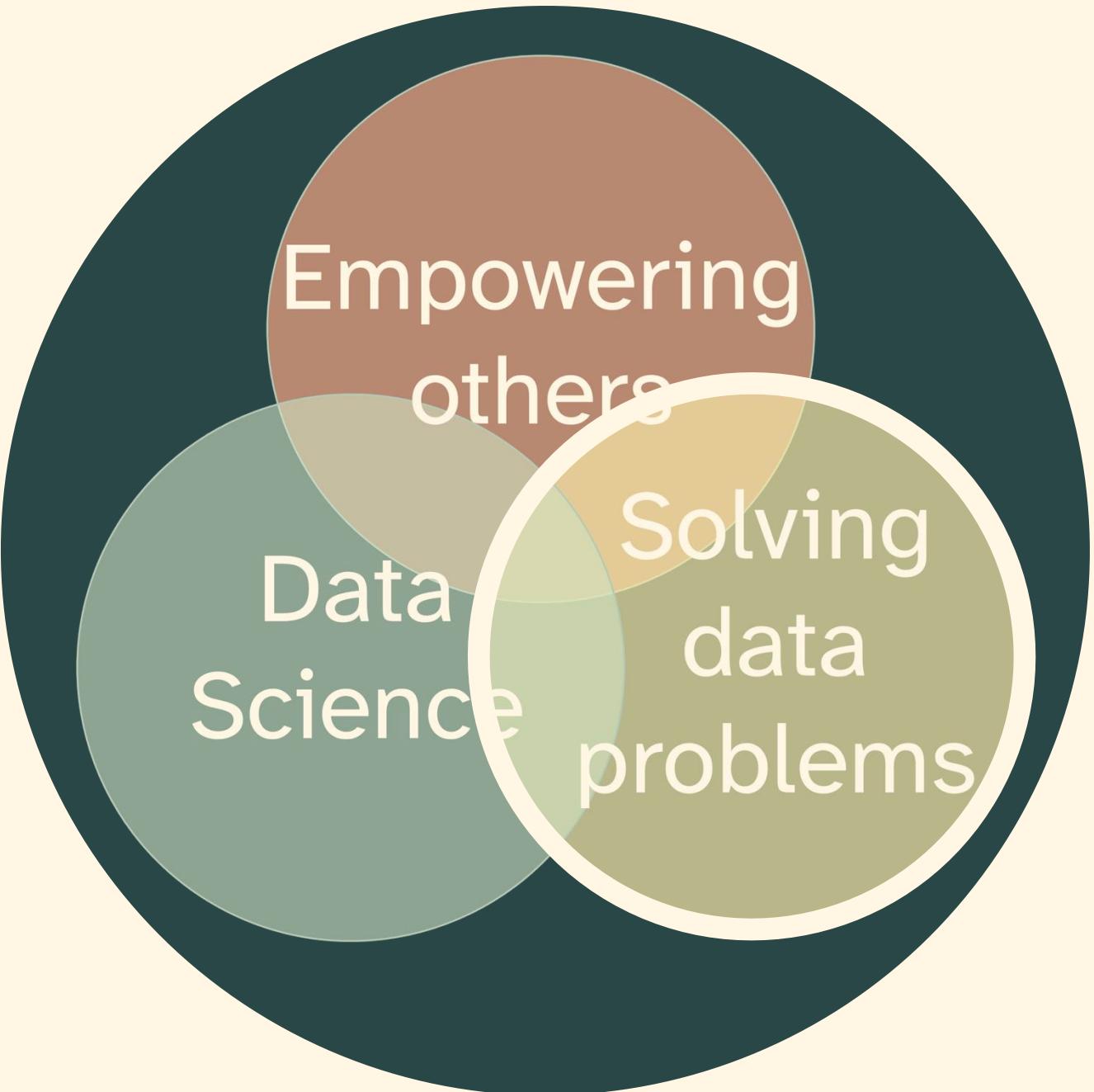


- *Variance-centric thinking*
- *Bayesian hierarchical models*
- *Missing data imputation*
- *Research synthesis*





UNSW
Postdoc
Research Software
Engineer



Data
Science



GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data



GBIF

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES



Search

What is GBIF?

About GBIF Australia

Atlas of Living Australia

ala.org.au

Open access to Australia's biodiversity data

150,259,838 occurrence records

10,164 datasets

Search species, datasets, and more...

Search

Problem with Big Biodiversity data

- > *It's out there...but :*
- > **Access** is tricky



Problem with Big Biodiversity data

- > *It's out there...but :*
 - > **Access** is tricky
 - > *Data is messy*



Streamline data cleaning



Dr. Elizabeth
Wenk



Dr. Ruby
Stephens



A/Prof. Daniel
Falster



Prof. Will
Cornwell

Streamline data cleaning

> Plant names keep **changing**



Dr. Elizabeth
Wenk



Dr. Ruby
Stephens



A/Prof. Daniel
Falster



Prof. Will
Cornwell

Streamline data cleaning

- > Plant names keep **changing**
- > Cross-match to “the truth” in a **reproducible + transparent way**



Dr. Elizabeth
Wenk



Dr. Ruby
Stephens



A/Prof. Daniel
Falster



Prof. Will
Cornwell

Streamline data cleaning

- > Plant names keep **changing**
- > Cross-match to “the truth” in a **reproducible + transparent way**



Dr. Elizabeth
Wenk



Dr. Ruby
Stephens



A/Prof. Daniel
Falster



Prof. Will
Cornwell



APCAlign :: CHEAT SHEET

Wenk Elizabeth H., Cornwell William K., Fuchs Anne, Kar Fonti, Monroe Anna M., Sauvaget Hervé, Stephens Ruby E., Falster Daniel S. (2024) APCAlign: an R package workflow and app for aligning and updating flora names to the Australian Plant Census. *Australian Journal of Botany* 72, BT24014.

APCAlign uses the Australian Plant Census and the Australian Plant Name Index to align and update plant taxon names to current accepted standards. It also provides useful information about diversity and establishment means of Australian plants across states.

Workflow

Aligning and updating your plant taxon names can be achieved in one simple step:

```
create_taxonomic_update_lookup(
  taxa,
  ...
  )
```

Calling this **wrapper** function will:

1. Retrieve taxonomic reference lists (APC and APNI)


```
load_taxonomic_resources(...)
```
2. Align your taxon names to names in the APC and APNI using our matching algorithms


```
align_taxa(...)
```

- Function will first attempt to align to names that exist in the APC
- If an alignment cannot be found, APNI will be used
- Includes fuzzy/partial string matches

3. Update aligned names to currently accepted names


```
update_taxonomy(...)
```

NOTE: We recommend calling each sub-function individually if you require finer control of each step's options

Data sources

Australian Plant Census (APC)

The APC is the national database of accepted names for Australian vascular plants. The APC provides information on:

- synonyms, and misapplications of scientific names
- establishment means (native/introduced)
- plant distribution across states and territories.

Australian Plant Name Index (APNI)

The APNI is a list for all Australian plants in scientific literature. It is used for standardising synonyms. APNI does not provide recommendations of taxonomy or nomenclature.

Retrieve data

Version-controlled taxonomic data are stored in our GitHub Repository at <https://github.com/traitecoevo/APCAlign/releases>

Get all versions of **stable** taxonomic datasets

versions

```
get_versions()
```

Download either the **latest** taxonomic data or a **version-controlled** taxonomic dataset

Function returns list object of APC and APNI databases

resources + **load_taxonomic_resources**(
resources = **resources**,
use_stable_or_current_data = "stable",
version = "2024-09-25",
quiet = FALSE
)

Update names

Update name matched to accepted name in APC. Function returns a tibble of original names, aligned names, accepted names and suggested names

```
update_taxonomy(  
aligned_data = alignments,  
taxonomic_splits = "most_likely_species",  
quiet = TRUE,  
output = NULL,  
resources = resources  
)
```

Taxonomic splits

- **most_likely_species**
 - Returns the species name in use before the split; alternative names are returned in a separate column
- **"return_all"**
 - Adds additional rows to the output, one for each possible taxon concept
- **"collapse_to_higher_faxon"**
 - Returns the genus with possible names in square brackets.

Establishment Means

native_anywhere_in_australia(
species,
resources = **resources**
)

Identify establishment status of taxon name

state_diversity_counts(
state,
resources = **resources**
)

Summarise counts of establishment means by state

Learn more

Don't want to code? No worries! Try our **Shiny** interface: <https://unsw.shinyapps.io/APCAlign-app/>

CC BY-AD Daniel Falster · daniel.falster@unsw.edu.au · <https://traitecoevo.github.io/APCAlign/> · Learn more at [webpage](#) or [vignette](#) · package version 1.1.2 · Updated: 2024-11

Streamline data cleaning

- > Encourages **reproducibility + transparency**



```
resources <- load_taxonomic_resources(  
  stable_or_current_data = "stable",  
  version = "2025-01-28"  
)
```

Streamline data cleaning

- > Encourages **reproducibility + transparency**



```
resources <- load_taxonomic_resources(  
  stable_or_current_data = "stable",  
  version = "2025-01-28"  
)
```

Streamline data cleaning

- > Encourages **reproducibility + transparency**

Commersonia rosea → *Androcalva rosea*



Streamline data cleaning



- > Encourages **reproducibility + transparency**

Commersonia rosea → *Androcalva rosea*

Strip down
name

Coarse match
catch

Fuzzy match
first word

Fuzzy match
second word

... →

Suggested
name

```
aligned <- align_taxa(  
  "Commersonia rosea",  
  full = TRUE,  
  resources = resources  
)
```

A tool to **discover** biodiversity

- > *Location-based discovery for citizen scientists*

“You don’t know what you’re looking for...”



A tool to discover biodiversity

- > **Location-based discovery for citizen scientists**
- > **Global scope**
- > **Efficient storage + quick loading**
- > **Portable**



Thomas
Mesaglio



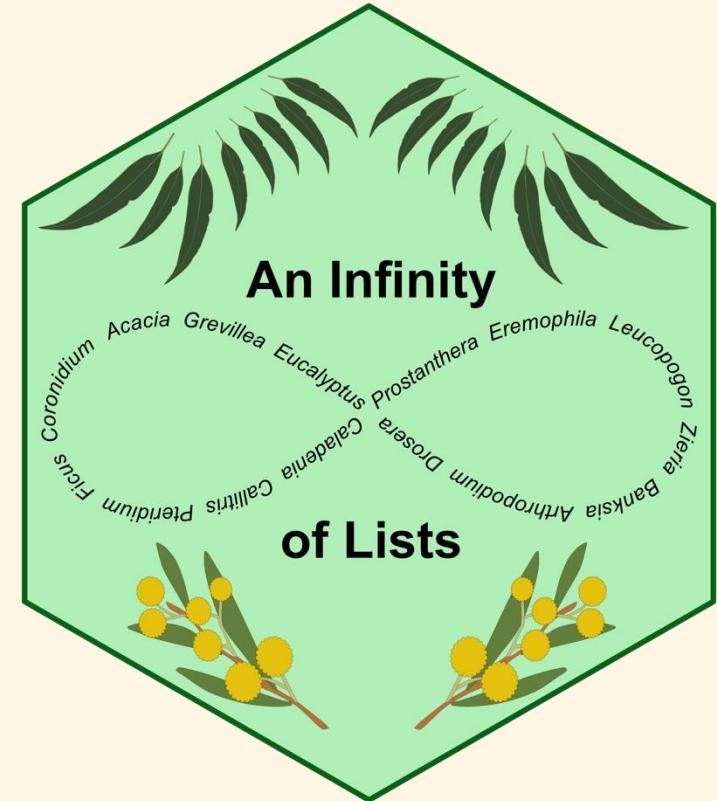
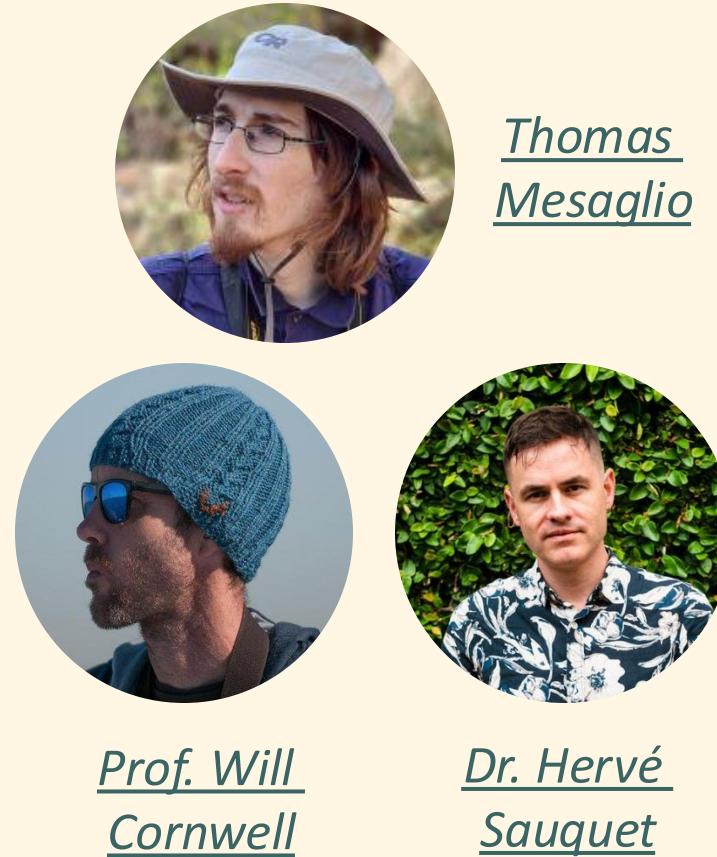
Prof. Will
Cornwell



Dr. Hervé
Sauquet

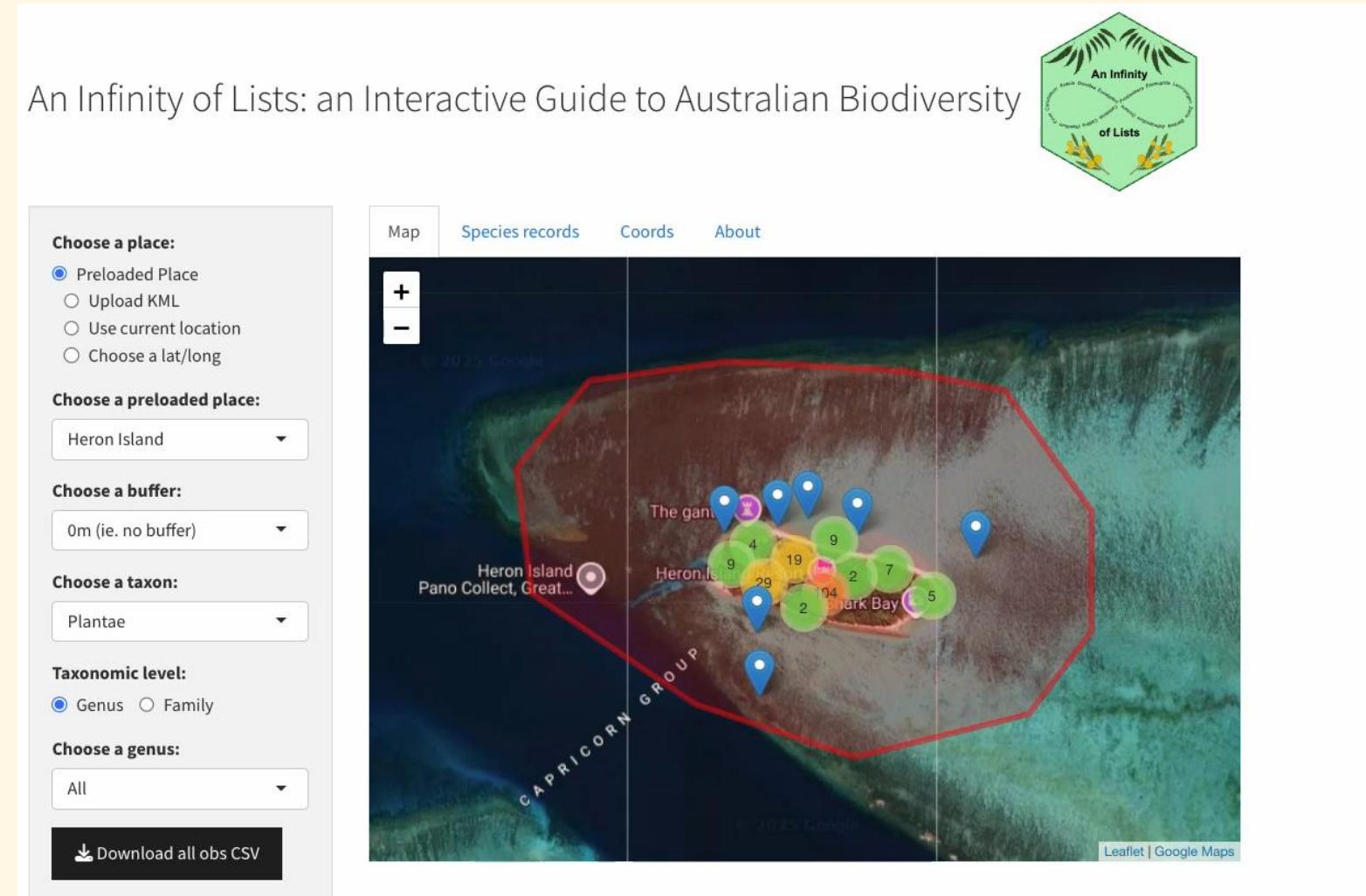
A tool to discover biodiversity

- > **Location-based discovery for citizen scientists**
- > **Global scope**
- > **Efficient storage + quick loading with .parquets**
- > **Portable**

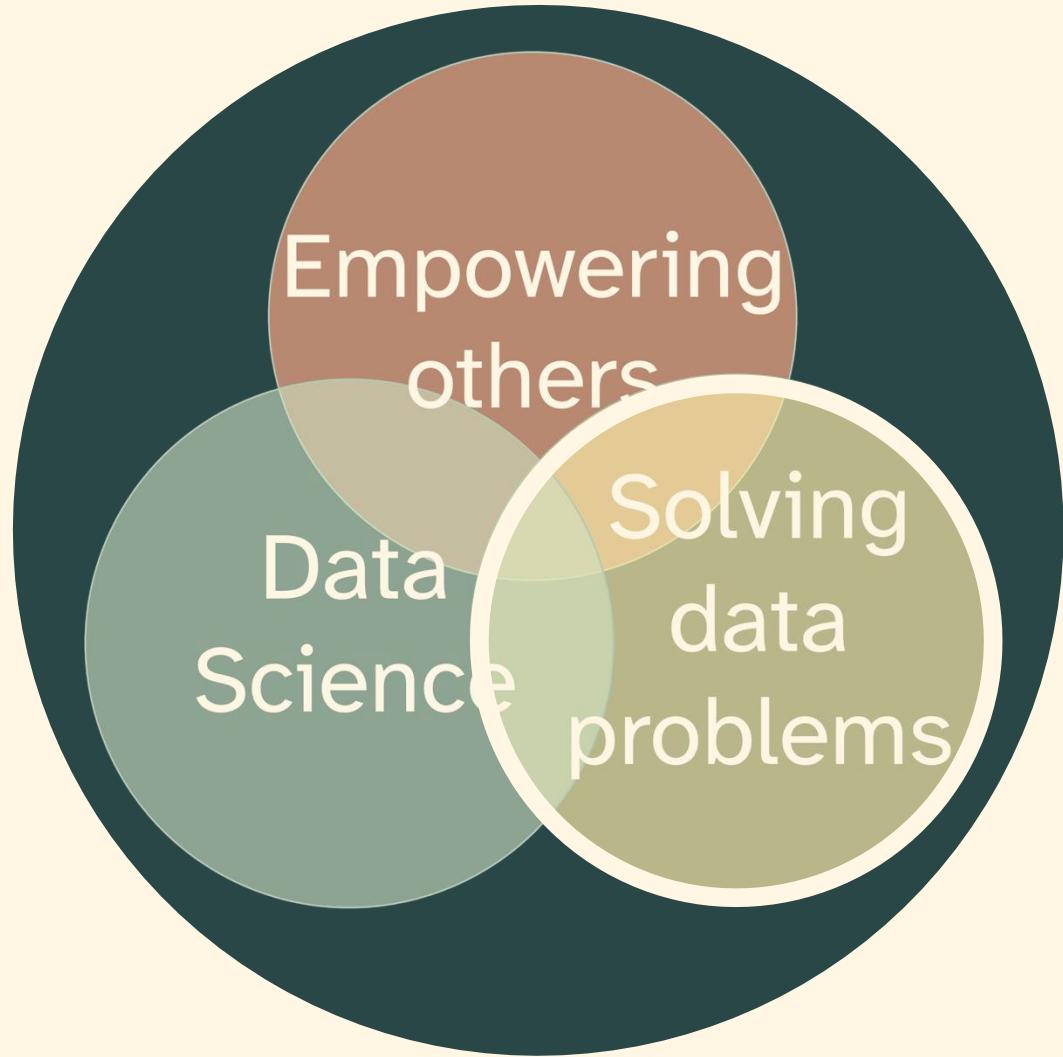


A tool to discover biodiversity

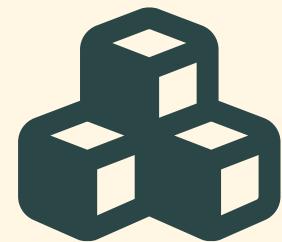
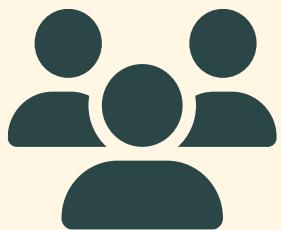
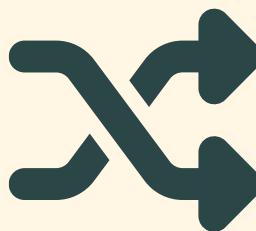
- > *Location-based discovery for citizen scientists*
- > *Global scope*
- > *Efficient storage + quick loading with .parquets*
- > *Portable*

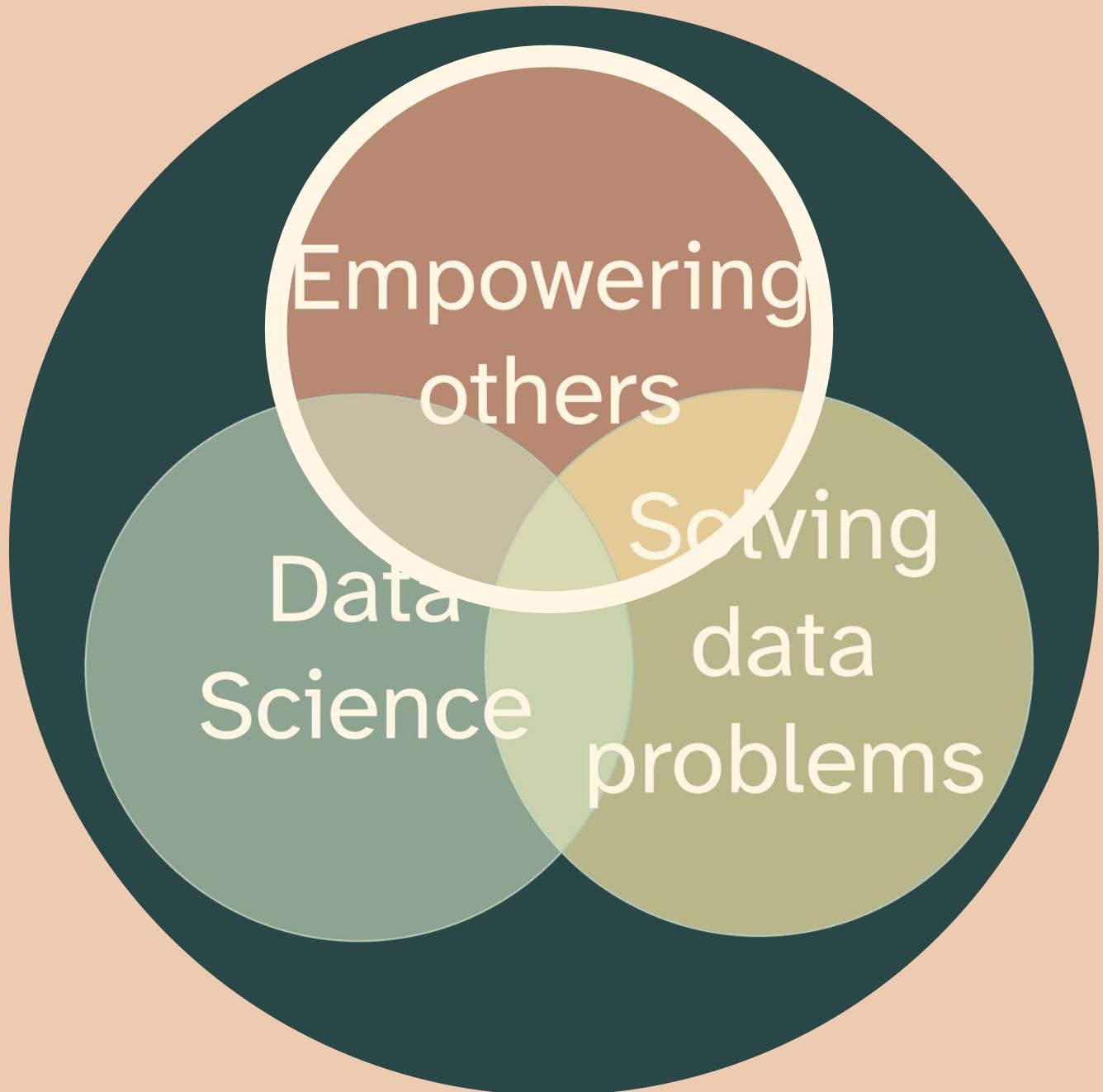


Solve
problems



- > *Streamline data flows*
- > *Build useful, open-source software / dashboards*





Empower

- Build R packages
- Teach others to build R packages



- Build R packages
- Teach others to build R packages

problem:
“...I've never built an R
package before...”



learning new R skills is hard.



1.
I got this.

learning new R skills is hard.



learning new R skills is hard.



Empower



Empower

a whale of inspiration



Empower

a whale of inspiration



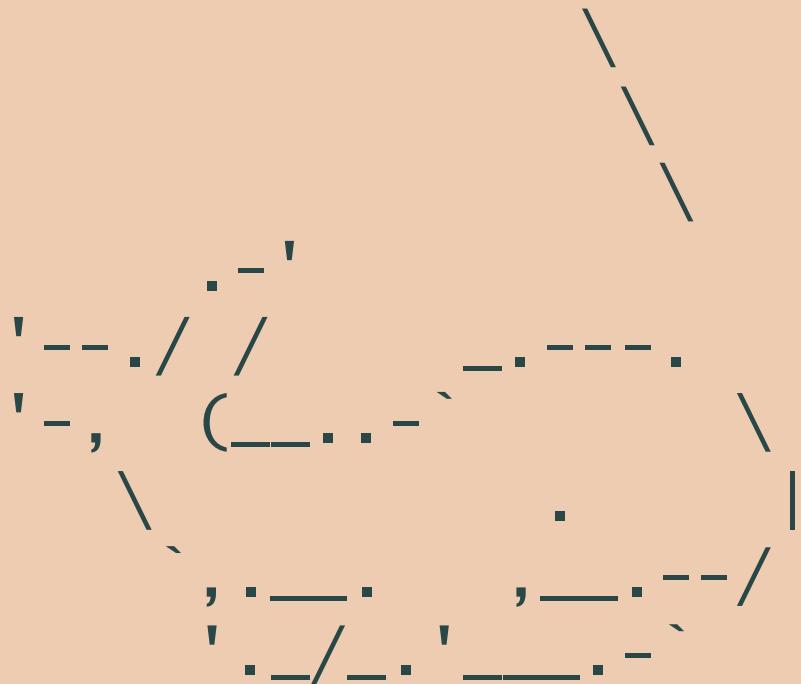
Empower

a whale of inspiration



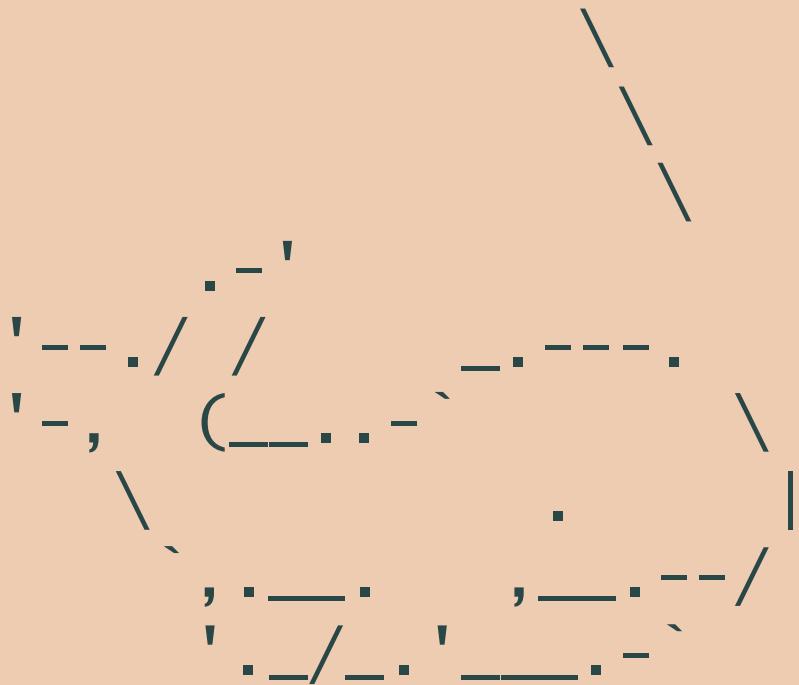
```
> say()
```

```
-----  
I'm having a whale of a time!  
-----
```



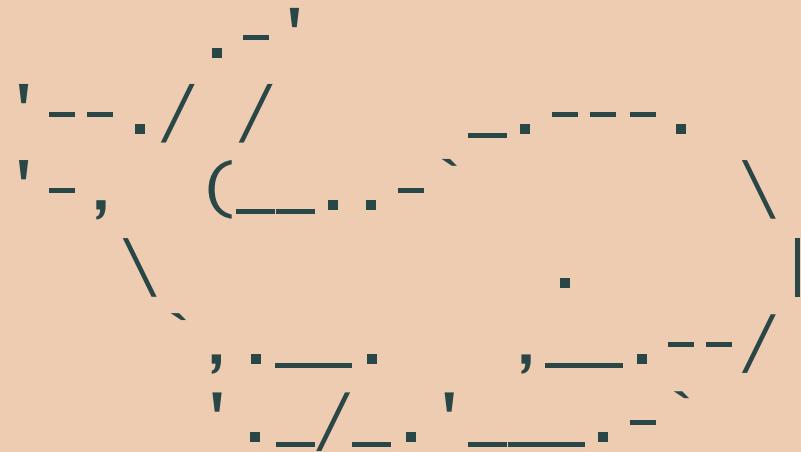
> say()

Gee...this is over-whalming!



```
> say("Thank you so much for hosting me EBS!!!")
```

Thank you so much for hosting me EBS!!!



why is silly so effective for learning?



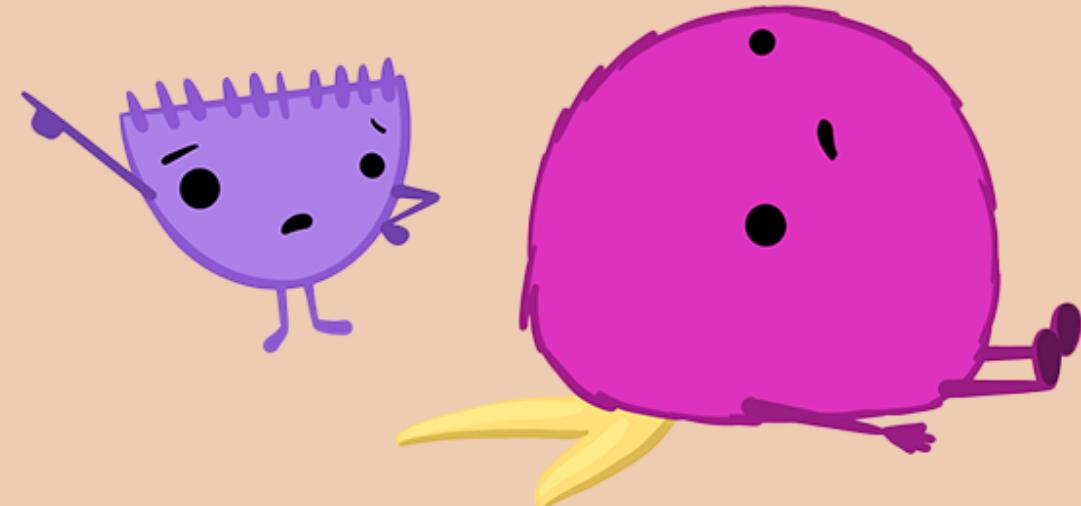
Empower

silliness is engaging

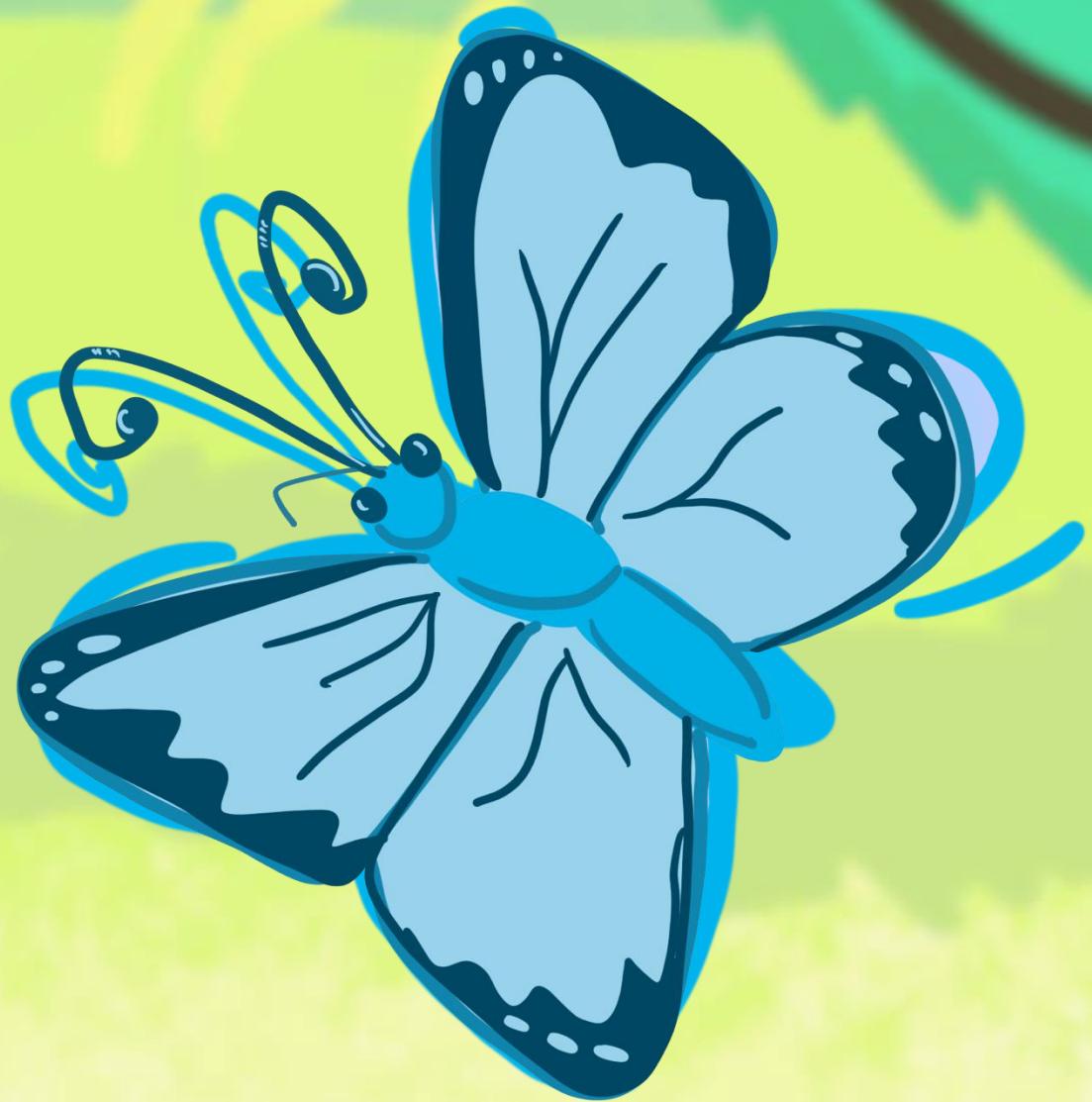




silly frees up brain space



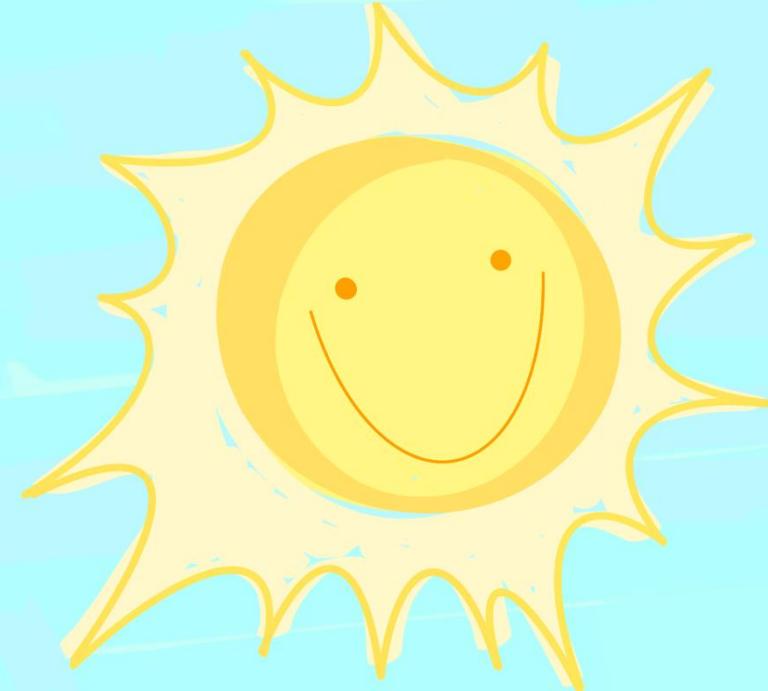
Empower



silly brings focus

Empower

silly is safe

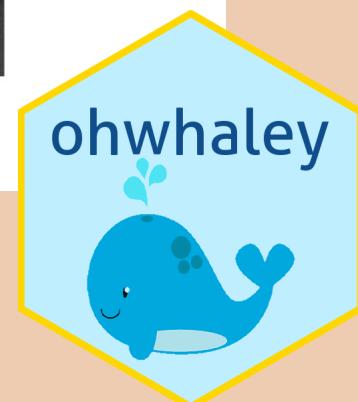


Share knowledge

D.I.Y R package



Fonti Kar
Postdoc UNSW



it's a whale
of a time!

Share knowledge

D.I.Y R package

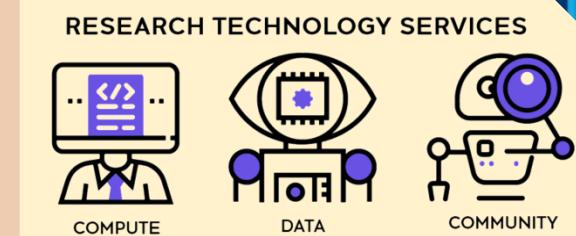
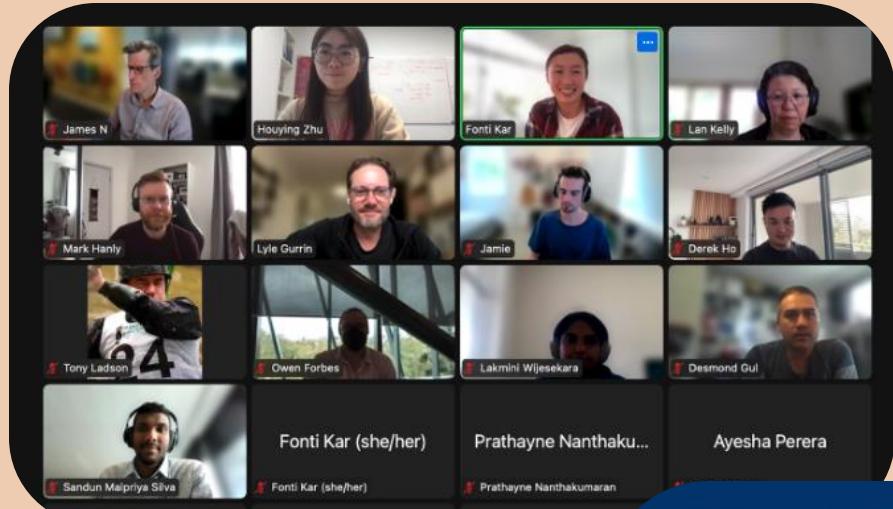
I've learn soooo much. I'm very grateful to you Fonti, for providing the documentation, your patience in keeping us together and your willingness to work through the breaks. Thank you!

"Cause if you like it then
you could've put it in an R package"

Fonti Kar
Postdoc UNSW



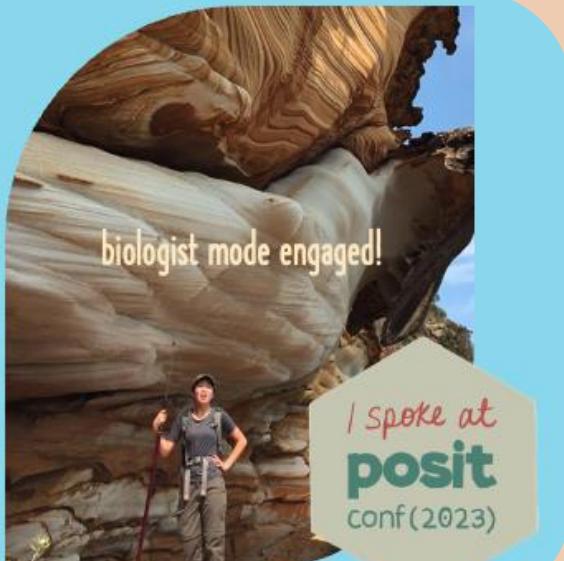
it's a whale
of a time!



Statistical
Society of
Australia

The whale splash effect...

field guide
to writing
your first
R package



@zerogruen 7 months ago

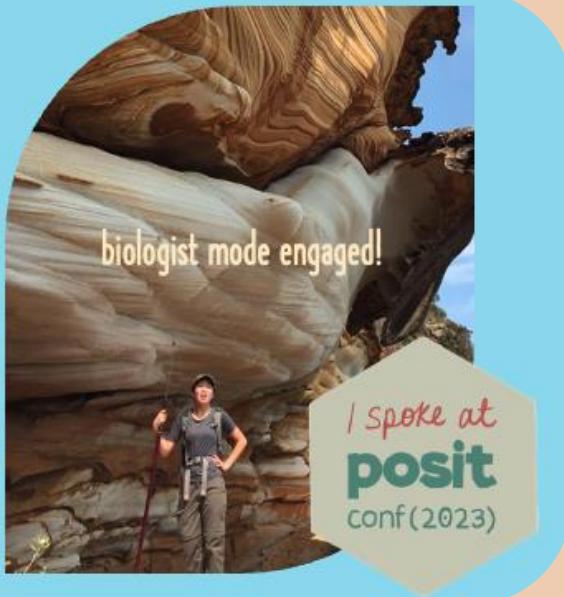
Hey Fonti, thanks for the nice talk. I found your picture with the stream of functions, the description campsite and so on very helpful, so I copied it to my notes. Thanks



Empower

The whale splash effect...

field guide
to writing
your first
R package



@zerogruen 7 months ago

Hey Fonti, thanks for the nice talk. I found your picture with the stream of functions, the description campsite and so on very helpful, so I copied it to my notes. Thanks



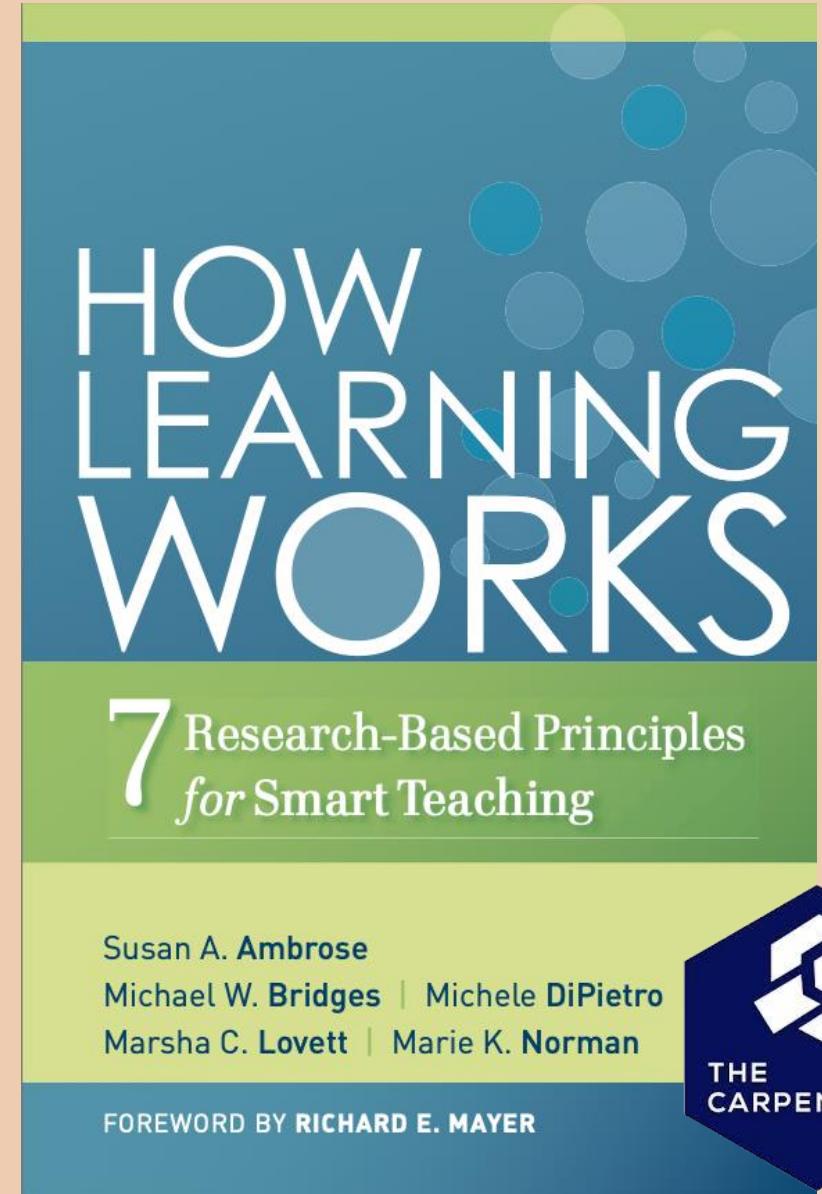
Lee Durbin



Seth Colbert-Pollack

Empower

Teaching



Teaching

- > *In person + virtual*
- > *Biostatistics under-graduate courses*
- > *Contributed to workshops series*

 Environmental Computing

Better figures with 

Fonti Kar

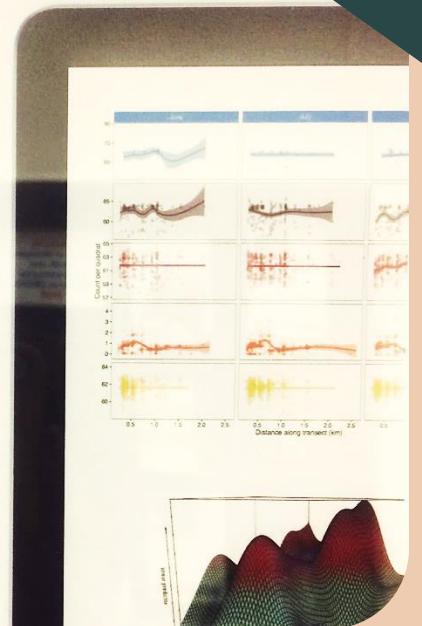


UNSW ENVIRONMENTAL COMPUTING SEMINAR SERIES

Project organization | Writing your own functions | tidyR and dplyr | Power and sample size | Mixed models | parallelization and clusters | UNSW Katana | Publication quality plotting | Test driven development | Making an R package | Raster and vector handling | Big Data | Inference for time and space | modeling standardized data | Loops and functions | Models for biomass data | Version Control | Help sessions

NO REGISTRATION NECESSARY

FRIDAY 2-3PM
20th JULY - 1st SEPT 2017
BioSciences Bldg RM640



 Skills transfer workshop

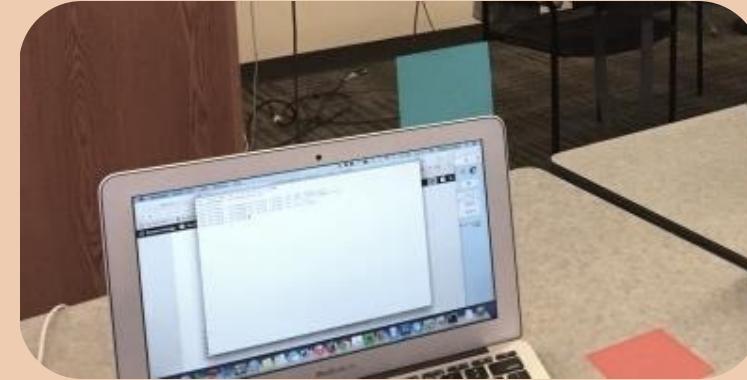
Katana 101 : How to be an analysis ninja



Fonti Kar   @fonti_kar

Teaching

- > *In person + virtual*
- > *Biostatistics under-graduate courses*
- > *Contributed to workshops series*



What is the purpose of the DESCRIPTION file in an R package

Provide meta-data and dependencies of the software

100%

Serves as an abstract to the software

0%

Where functions for package are declared

0%



Environmental Computing

Better figures with



Fonti Kar



@

fonti_kar



@fonti_kar



@fonti_kar



E&ERC Skills transfer workshop

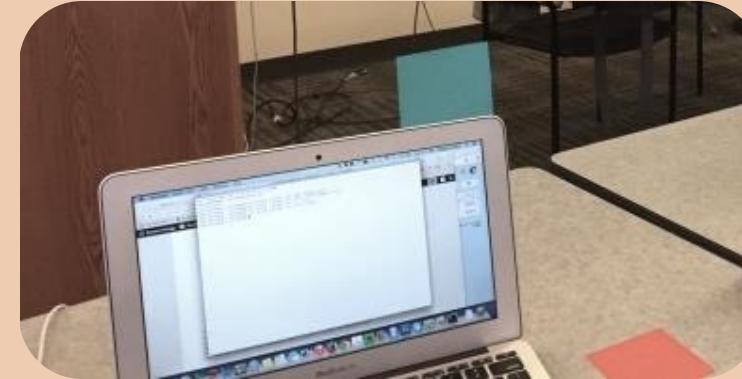
Katana 101 : How to be an analysis ninja



Fonti Kar @fonti_kar

Teaching

- > *In person + virtual*
- > *Biostatistics under-graduate courses*
- > *Contributed to workshops series*



What is the purpose of the DESCRIPTION file in an R package

Provide meta-data and dependencies of the software

100%

Serves as an abstract to the software

0%

Where functions for package are declared

0%



Environmental Computing

Better figures with

Fonti Kar

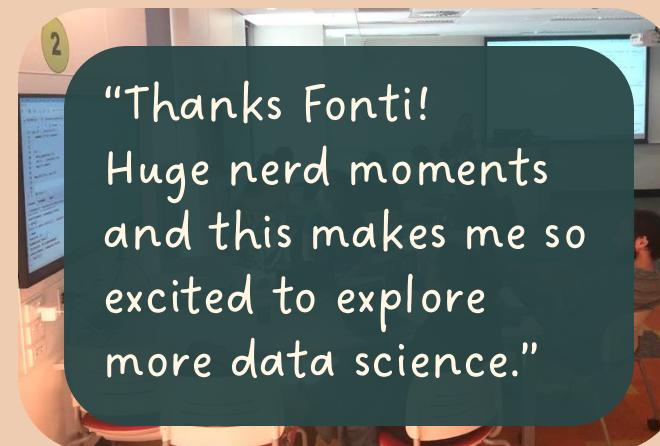


@

fonti_kar



@fonti_kar



"Thanks Fonti!
Huge nerd moments
and this makes me so
excited to explore
more data science."



E&ERC Skills transfer workshop

Katana 101 : How to be an analysis ninja



Fonti Kar @fonti_kar

Empower

Build capacity

- > *Upskilling post-graduates*



A/Prof. Daniel
Falster



Prof. Will Cornwell



Empower

Build capacity

- > *Upskilling post-graduates*
- > *Beginner*
 - > *Project management*
 - > *Data wrangle + visualise*



A/Prof. Daniel
Falster



Prof. Will Cornwell



Build capacity

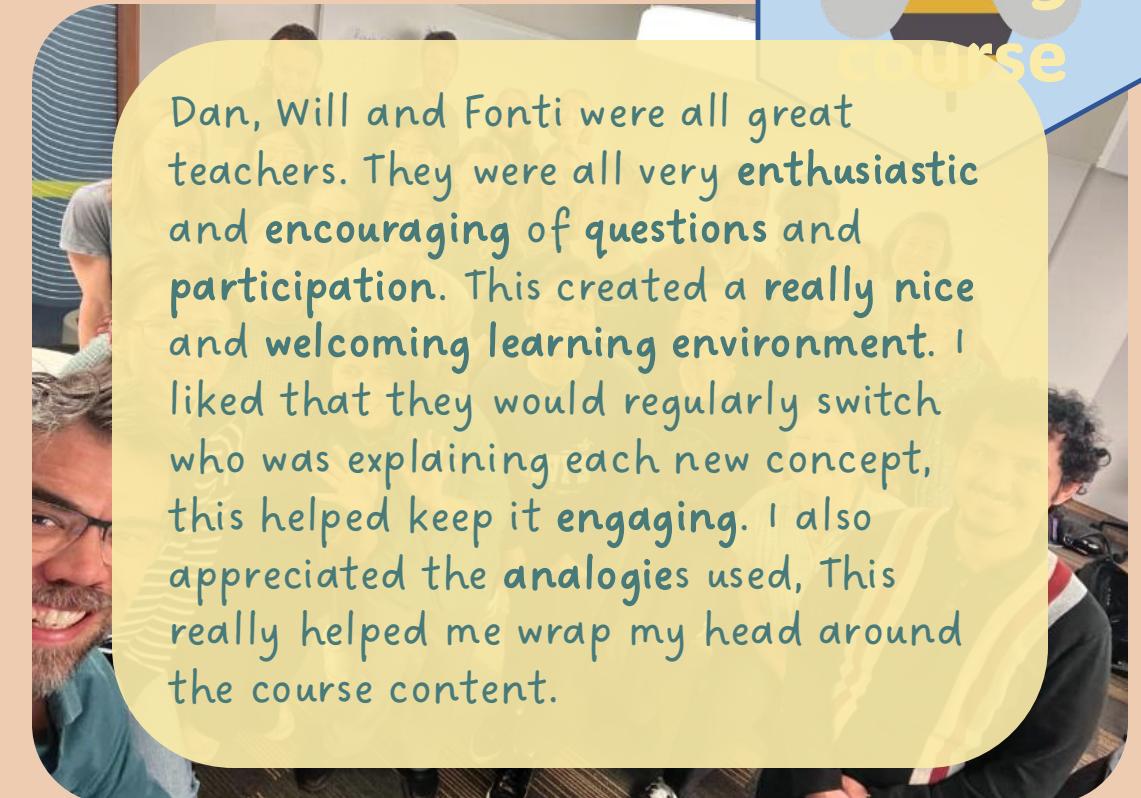
- > *Upskilling post-graduates*
 - > *Beginner*
 - > *Project management*
 - > *Data wrangle + visualise*
 - > *Intermediate*
 - > *Reproducible practices*
 - > *Split-apply-combine*
 - > *Functional programming*



A/Prof. Daniel Falster



Prof. Will Cornwell



Empower

Interactive curricula + assessment



Ethan Ou

Interactive curricula + assessment

Working with `dplyr`

Learn more from the [dplyr guide](#), its really comprehensive!

Extracting rows

The `filter()` function allows you to take subsets of a data frame. You have to specify the **condition** for which rows to keep.

Take this example, the code below gives me data for only `penguins` from the `Dream` Island.

R Code

```

1 penguins %>
2   filter(island == "Dream")
3

# A tibble: 124 x 8
# ... with 8 variables:
#   species     island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
#   <fct>      <fct>        <dbl>        <dbl>          <int>        <int>
# 1 Adelie     Dream         39.5       16.7            178        3250
# 2 Adelie     Dream         37.2       18.1            178        3900
# 3 Adelie     Dream         39.5       17.8            188        3300
# 4 Adelie     Dream         40.9       18.9            184        3900
# 5 Adelie     Dream         36.4        17             195        3325
# 6 Adelie     Dream         39.2       21.1            196        4150
# 7 Adelie     Dream         38.8        20             190        3950
# 8 Adelie     Dream         42.2       18.5            180        3550
# 9 Adelie     Dream         37.6       19.3            181        3300
# 10 Adelie    Dream         39.8       19.1            184        4650
# # ... with 114 more rows
# # ... with 2 more variables: sex <fct>, year <int>

```

Let's walk through the code:

1. The `penguins` dataset is piped to the `filter()` function.
2. Within the `filter()` function is where we specify our condition. I've used the **equal-to operator** represented by



Ethan Ou



- > Focus on **core concepts**
- > Accessible + Flexible

Interactive curricula + assessment

R Refresher

Learning to code in R is like learning to speak a new language. Here we will go through some conventions and common vocabulary used in R. With practice, these will become familiar and intuitive - let's get started!



R then and now, Artwork by @allison_horst



Ethan Ou



Mitchell
O'Hara-Wild



Dr. Emi Tanaka



+



- > Focus on **core concepts**
- > Accessible + Flexible
- > Formative assessment

<https://github.com/numbats/moodlequiz>

Interactive curricula + assessment

Working with `dplyr`

Learn more from the [dplyr guide](#), its really comprehensive!

Extracting rows

The `filter()` function allows you to take subsets of a data frame. You have to specify the **condition** for which rows to keep.

Take this example, the code below gives me data for only `penguins` from the `Dream` Island.

R Code
[Start Over](#)
[Run Code](#)

```

1 penguins |>
2   filter(island == "Dream")
3

# A tibble: 124 x 8
  species island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
  <fct>   <fct>     <dbl>        <dbl>          <int>        <int>
1 Adelie  Dream       39.5        16.7           178        3250
2 Adelie  Dream       37.2        18.1           178        3900
3 Adelie  Dream       39.5        17.8           188        3300
4 Adelie  Dream       40.9        18.9           184        3900
5 Adelie  Dream       36.4         17            195        3325
6 Adelie  Dream       39.2        21.1           196        4150
7 Adelie  Dream       38.8         20            190        3950
8 Adelie  Dream       42.2        18.5           180        3550
9 Adelie  Dream       37.6        19.3           181        3300
10 Adelie Dream       39.8        19.1           184        4650
# i 114 more rows
# i 2 more variables: sex <fct>, year <int>

```

Let's walk through the code:

1. The `penguins` dataset is piped to the `filter()` function.
2. Within the `filter()` function is where we specify our condition. I've used the **equal-to operator** represented by `==` here to tell R, I only want rows from the variable `island` where it's value matches the word "Dream"

`## Working with `dplyr``

<h3>Working with `dplyr`</h3>

> Learn more from [\[the dplyr guide\]](#)(<https://dplyr.tidyverse.org/articles/dplyr.html>), its really comprehensive!

<h4>Extracting rows</h4>

The `filter()` function allows you to take subsets of a data frame. You have to specify the ****condition**** for which rows to keep.

Take this example, the code below gives me data for only `penguins` from the `Dream` Island.

```

<div id="r-editor-114"><pre>
penguins |>
  filter(island == "Dream")
</pre>
</div>

```

Let's walk through the code:

1. The `penguins` dataset is piped to the `filter()` function.
2. Within the `filter()` function is where we specify our condition. I've used the ****equal-to operator**** represented by `==` here to tell R, I only want rows from the variable `island` where it's value matches the word "Dream"

> *Reproducible*

Empower



Prof.
Alistair Poore



A/Prof. Daniel
Falster



Prof.
Will Cornwell

Open resources for
data science education

Empower

WELCOME TO ENVIRONMENTAL COMPUTING



Search... x

About This Site
Getting Started with R
Data Manipulation
Coding Skills
Graphics
Statistics
Datasets

Clear History

[GitHub](#) [Star 16](#)

Built with Hugo via blogdown, using the
Learn theme

Getting started with R, data manipulation, graphics, and statistics for the environmental sciences

Students and researchers in the environmental sciences require a wide range of quantitative skills in analytical and data processing software, including R, geographic information systems (GIS) and the processing of remotely sensed data. There is increasingly a need to ensure transparency of data processing supported by statistical analyses to justify conclusions of scientific research and monitoring for management and policy. This site is a brief introduction to techniques for data organisation, graphics and statistical analyses.



<https://environmentalcomputing.net>



Prof.
Alistair Poore



A/Prof. Daniel Falster



Prof.
Will Cornwell

- > *Diverse contributors*
- > *Resource for under- / post- graduates*
- > *Scaffolded learning*

Open resources for
data science education



Search... x

- About This Site
- Getting Started with R
- Data Manipulation
- Coding Skills
- Graphics
- Statistics
- Datasets
- Clear History

[GitHub](#) [Star 16](#)

Built with Hugo via blogdown, using the Learn theme

WELCOME TO ENVIRONMENTAL COMPUTING

Getting started with R, data manipulation, graphics, and statistics for the environmental sciences

Students and researchers in the environmental sciences require a wide range of quantitative skills in analytical and data processing software, including R, geographic information systems (GIS) and the processing of remotely sensed data. There is increasingly a need to ensure transparency of data processing supported by statistical analyses to justify conclusions of scientific research and monitoring for management and policy. This site is a brief introduction to techniques for data organisation, graphics and statistical analyses.



Prof.
Alistair Poore

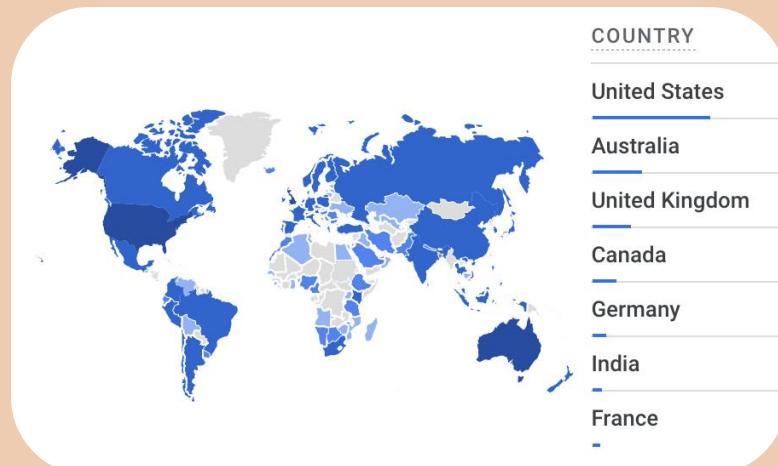


A/Prof. Daniel
Falster



Prof.
Will Cornwell

- > *Diverse contributors*
- > *Resource for under- / post- graduates*
- > *Scaffolded learning*
- > *34K visitors in past year*





Search... x

- About This Site
- Getting Started with R
- Data Manipulation
- Coding Skills
- Graphics
- Statistics
- Datasets
- Clear History

[GitHub](#) Star 16
Built with Hugo via blogdown, using the Learn theme

WELCOME TO ENVIRONMENTAL COMPUTING

Getting started with R, data manipulation, graphics, and statistics for the environmental sciences

Students and researchers in the environmental sciences require a wide range of quantitative skills in analytical and data processing software, including R, geographic information systems (GIS) and the processing of remotely sensed data. There is increasingly a need to ensure transparency of data processing supported by statistical analyses to justify conclusions of scientific research and monitoring for management and policy. This site is a brief introduction to techniques for data organisation, graphics and statistical analyses.



<https://environmentalcomputing.net>

Document Bienvenido

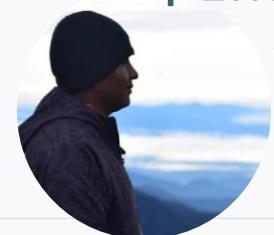
Bienvenido

Bienvenidos al sitio web de la versión en español de [Environmental Computing](#). Aquí se presenta información detallada sobre cómo emplear el lenguaje de programación R para la gestión y análisis de datos ambientales. Aprenderás sobre manipulación, análisis y visualización de datos, desarrollando habilidades clave para asegurar la reproducibilidad en los procesos de gestión de datos. Este sitio es un recurso esencial para aquellos en el campo de las ciencias biológicas y ambientales que buscan utilizar R en su trabajo.

Sobre la Traducción

La versión en español de [Environmental Computing](#) es un proyecto desarrollado con el fin de proporcionar a los profesionales de las ciencias biológicas una herramienta de consulta accesible para el uso de R. La traducción fue realizada por [Paul Efren Santos Andrade](#). Extendemos nuestro agradecimiento al profesor [Will Cornwell - Alistair Poore](#) por su generosidad al compartir la información y otorgarnos la autorización para difundir la traducción del contenido original de [Environmental Computing](#).

¡ Environmental Computing es traducido al español !



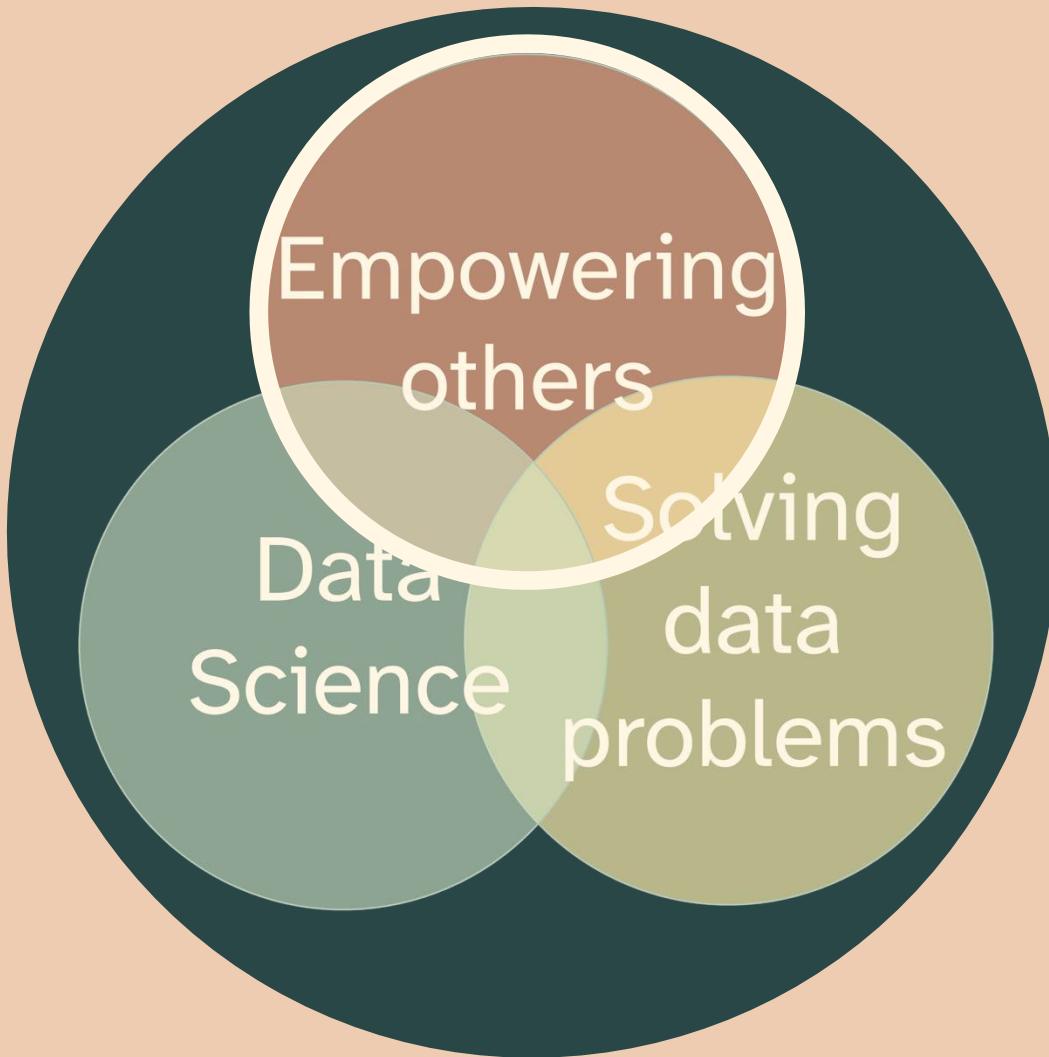
[Dr. Paul Santos](#)

[Sobre el sitio web →](#)

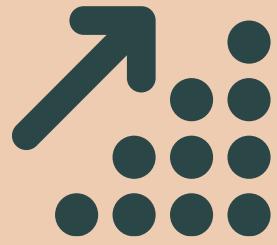
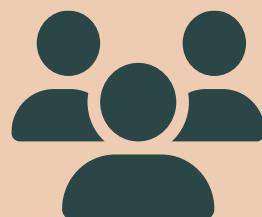
Environmental Computing - Spanish |
Made with [Quarto](#).

<https://environmentalcomputingsp.netlify.app/>

Empower



- > *Gentle on-ramps to data science concepts*
- > *Interactive, explorative learning*
- > *Scalable curricula + assessment workflows*



University of Auckland
BSc Biology + Statistics

UNSW + ANU
PhD



Macquarie University
MRes



UNSW
Postdoc



Atlas of Living Australia
Data Analyst

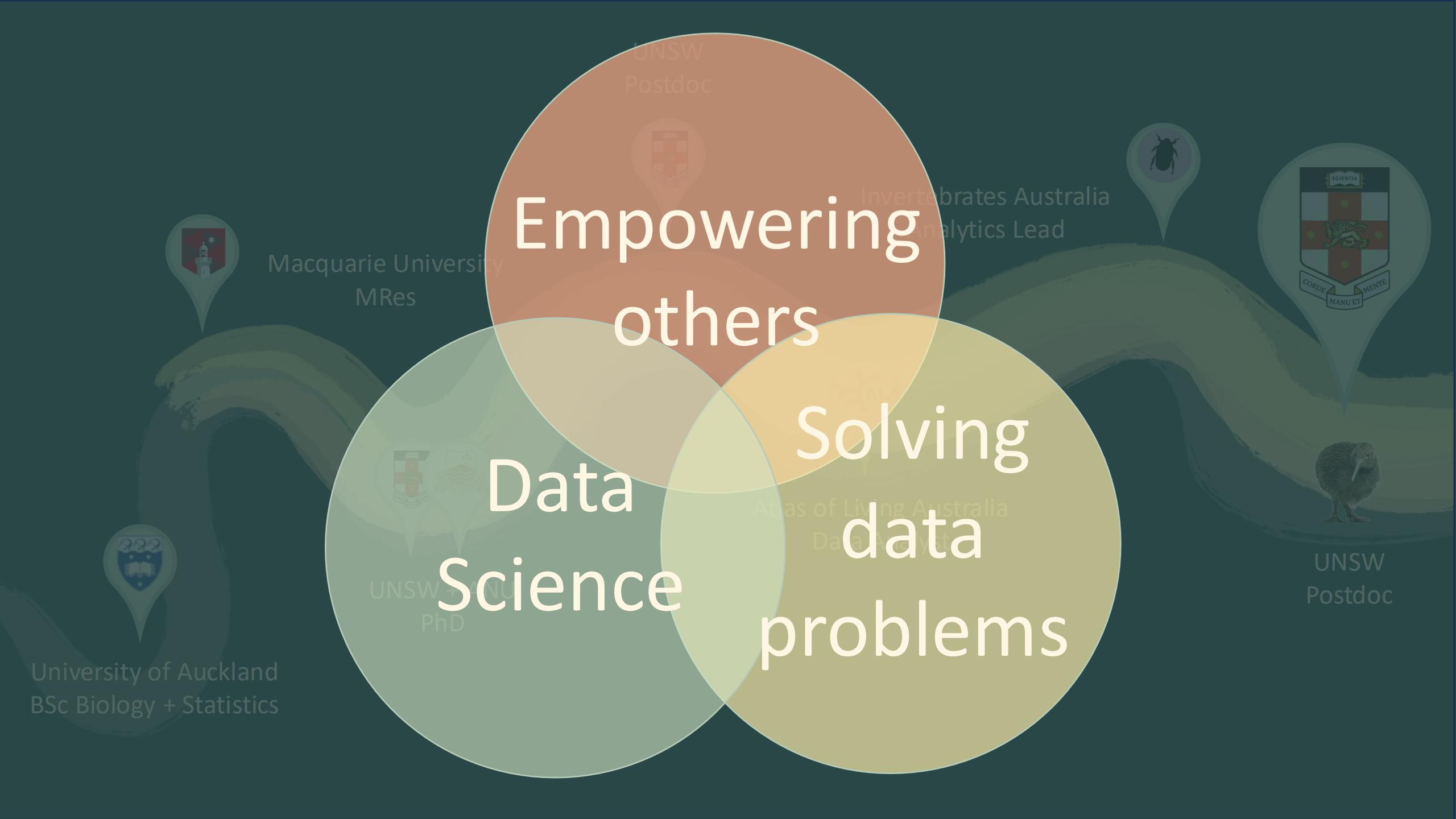


Invertebrates Australia
Analytics Lead



UNSW
Postdoc





Empowering others

Data
Science

Solving
data
problems

University of Auckland
BSc Biology + Statistics

Macquarie University
MRes

UNSW
Postdoc

Invertebrates Australia
Analytics Lead

Atlas of Living Australia
Data Analyst

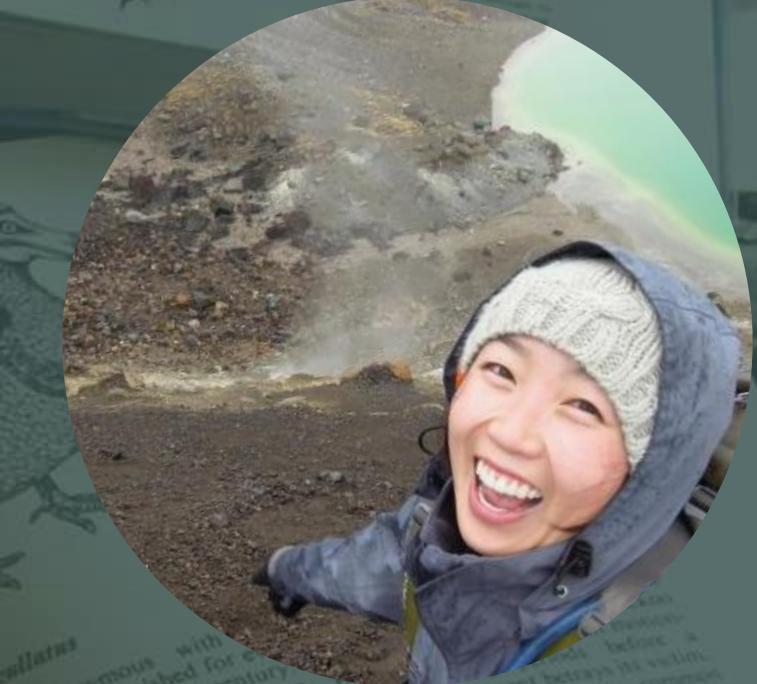
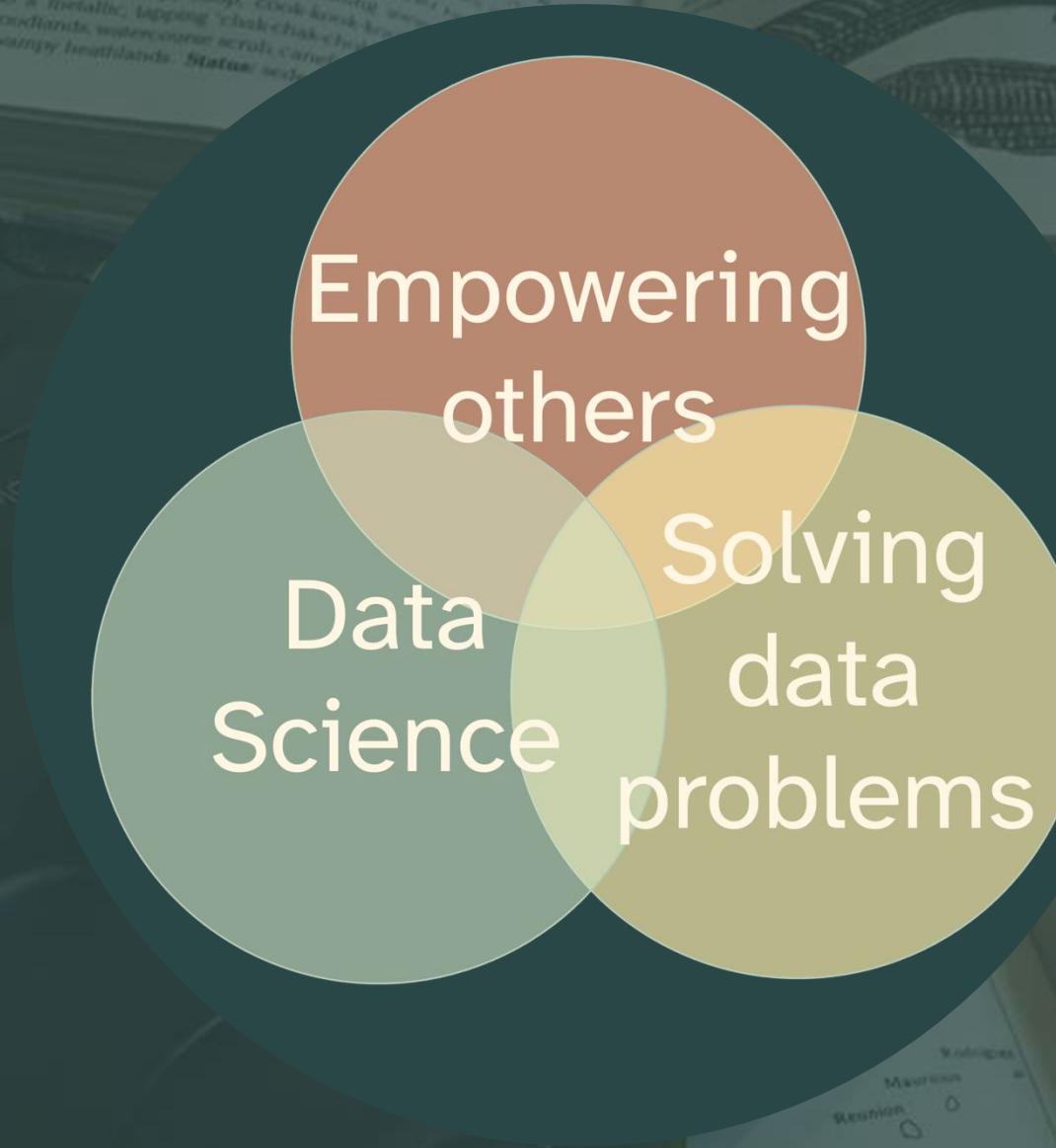
UNSW + ANU
PhD

UNSW
Postdoc

Original illustrations by Allison Horst (@allison_horst)

Dr. Fonti Kar

<https://fontikar.github.io/>



Thank you + let's connect!