Developing tools for "real time" formative assessment of writing within large introductory statistics & data science lectures

Dr Anna Fergusson

Department of Statistics | Te Kura Tatauranga

University of Auckland | Waipapa Taumata Rau

a.fergusson@auckland.ac.nz

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

SCIENCE
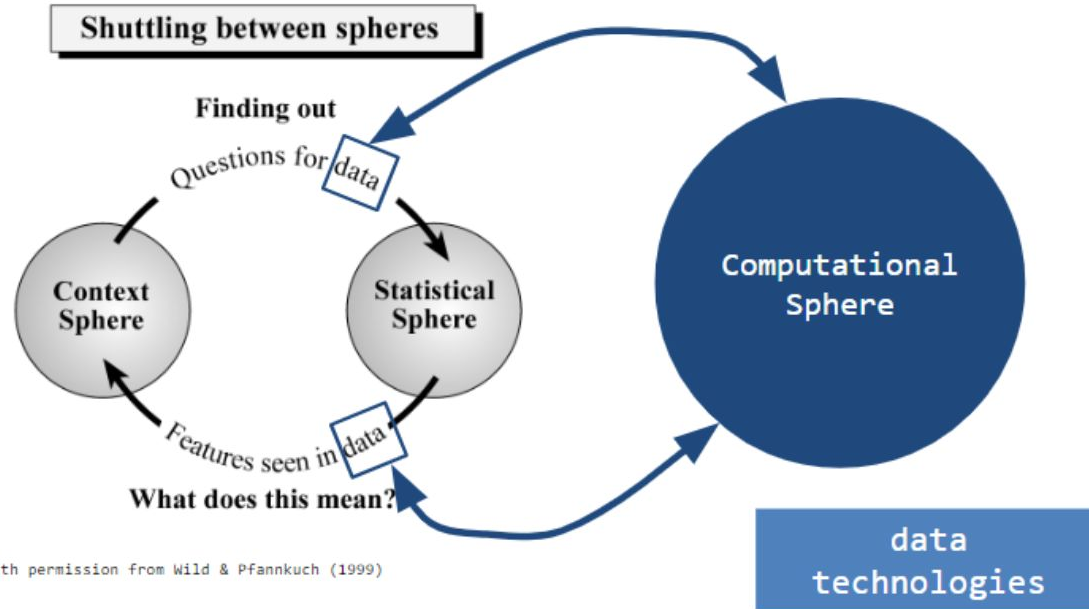DEPARTMENT OF STATISTICS

Statistics education research

We are known internationally for our statistics education research on statistical modelling, inferential reasoning, probabilistic thinking, dynamic visualisation tools and statistical literacy.

UoA Statistics Education Research

Data science + learning: Research at the interface

Tasks

Assessment

Shuttling between spheres

Finding out

Questions for data

Context Sphere

Statistical Sphere

Computational Sphere

Features seen in data

What does this mean?

Reprinted with permission from Wild & Pfannkuch (1999)

data technologies

Research that connects data technologies and pedagogy
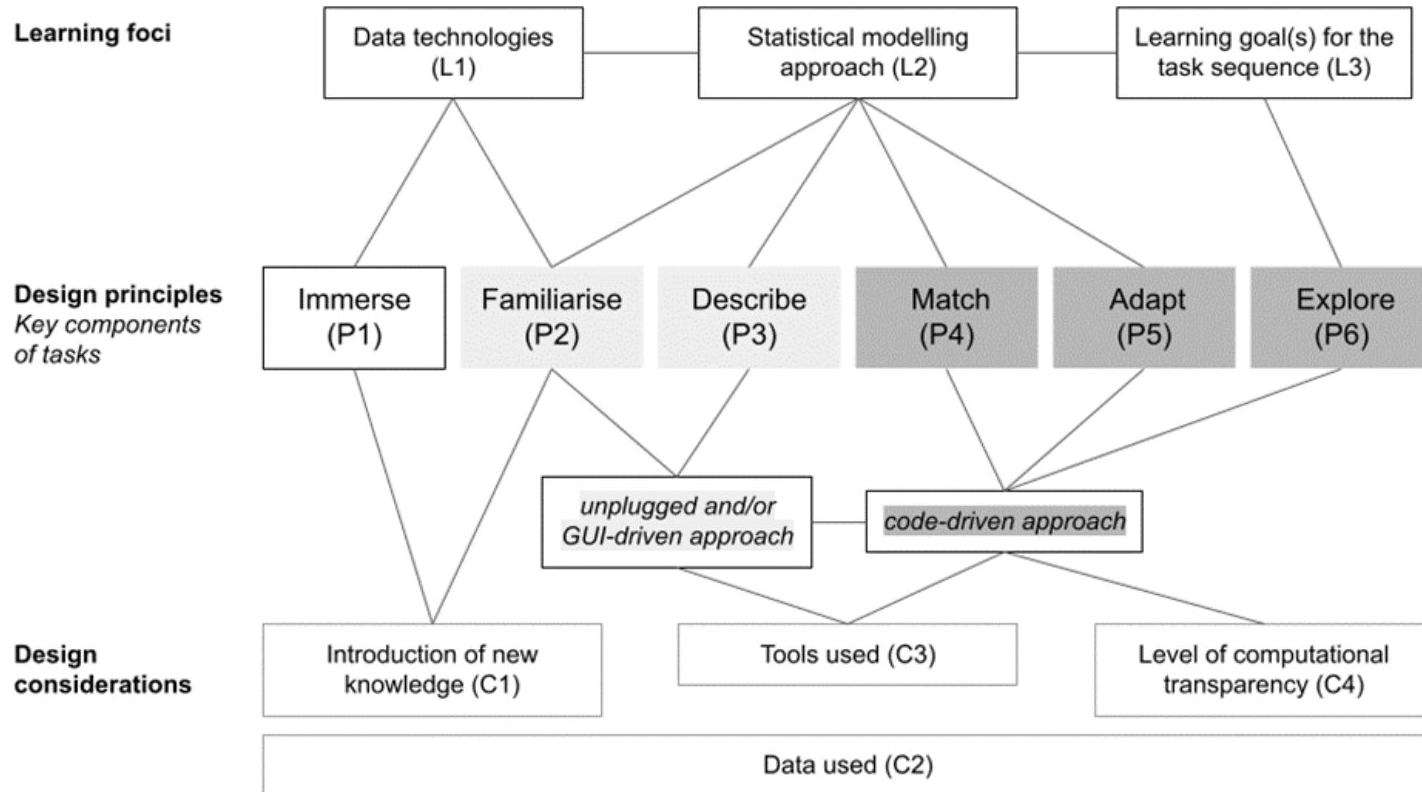
Quick example: Drawings of data!

My research as a tool + task designer

Introducing Code-driven Tools through Statistical Modelling framework

What is the challenge?

Generating feedback on statistical and computational writing

Why is this a challenge?

validity of feedback in terms of data science and statistical content

optimisation of feedback for sustainable learning outcomes

ethicality of data and pedagogical practices

engagement of a diverse range of learners within large-scale social environments

What courses are these tools used in?

STATS 101/108 Introduction to Statistics

⑦ Purposes + Decisions

- Recognise different purposes and motivations for making data-based decisions and the consequences of those decisions for affected communities.

👥 Data + People

- Describe ethical, responsible, and culturally-responsive data practices, acknowledging Māori Data Sovereignty.
- Use data generated from a range of sources, considering how decisions made affect its quality, diversity, and quantity.

⌄ Models + Generalisations

- Develop models using data, representations and critical reasoning, considering the applicability and generalisability of models and model-based claims.
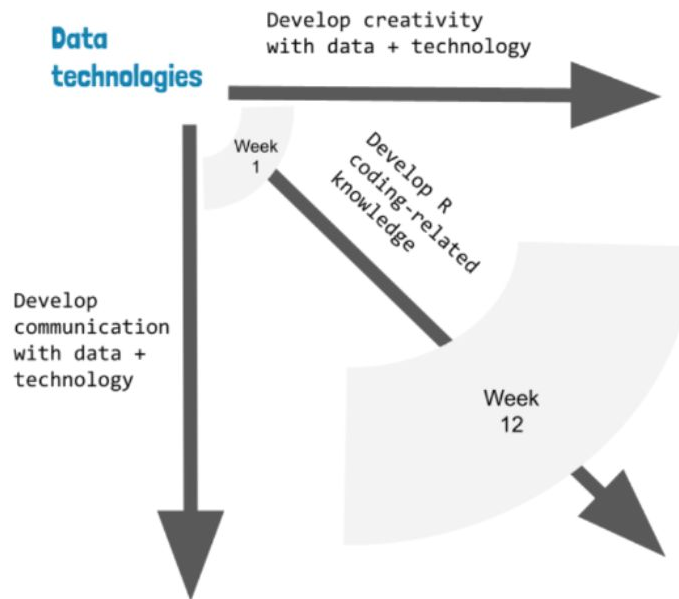
⌨ Technology + Automation

- Select and apply appropriate technology to analyse data, considering automated and reproducible approaches.

🧑‍🏫 Communication + Uncertainty

- Produce written summaries that communicate the uncertainty associated with data and interpret and critique communications produced by others.

STATS 220 Data Technologies

Where are these tools used?

Why do we care about formative assessment of writing within large-scale social learning environments?

UoA Signature Pedagogical Practices

Relational learning

Developing communities and cohorts of learning is scaffolded through appropriate learning environments, collaborative practices and relevant learning experiences.

Assessment for learning

This recasts assessment as learning, shifting the focus from end-point testing. Here we judge students' abilities to meet learning outcomes in situations where they will be applied.

Technology-enhanced learning

TEL supports a wide variety of teaching modes from blended and flexible learning, multimodal teaching to local and distance students, and online or face-to-face learning.

https://teachwell.auckland.ac.nz/signature-pedagogies/

But first ….. data!
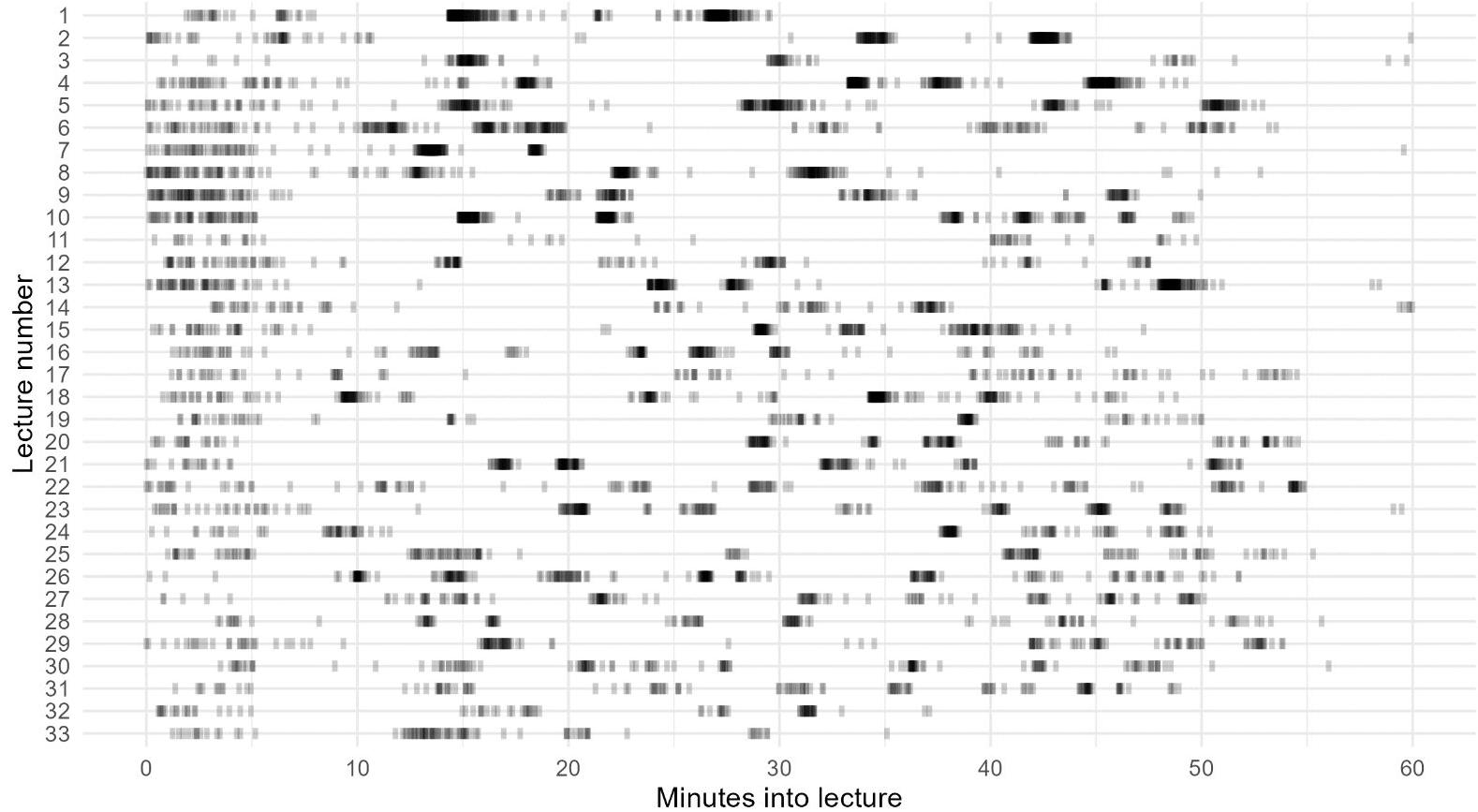
Quick Write Tool (QWT) demo

Teacher view

Show writings from last 500 minutes

| user interface | Accessibility, stability, compatibility | Design tools with the user in mind. | Easy to use and get around it quickly |

| Easy to use and give correct results | Easy to use. Intuitive. | hello world | Easy to setup and get going with a simple example. |

Quick, write!

Submit writing

Data goes where?

	A	G
1	timestamp	comment
	10/05/2024 13:13:56	Consider the scope of the study. What could be improved? If the randomly
18564	10/05/2024 13:13:58	the scope of the study is 40-84 year old men, and so results found can on
18565	10/05/2024 13:14:03	Observational, lots of confounding variables not controlled
18566	10/05/2024 13:14:12	Experiment, Control, blinding, placebo, no blocking Only men were analysed
18567	10/05/2024 13:14:18	Mostly experiment. Everything could be improved. No control!
18568	10/05/2024 13:14:22	control (Placebo) group, no blocking though
18569	10/05/2024 13:14:29	control group
18570	10/05/2024 13:14:40	It negatively affects people (if they dont have aspirin theyre more likely to
18571	10/05/2024 13:14:42	it wasnt ethical??
18572	10/05/2024 13:14:45	No blocking
18573	10/05/2024 13:14:50	there is blinding, there is a control group, there is no blocking
18574	10/05/2024 13:14:50	Blinding
18575	10/05/2024 13:15:01	the scope of the study (and therefore inference) is limited to men and can
18576	10/05/2024 13:15:08	what side effects came from taking the drugs so often, did any of the parti
18577	10/05/2024 13:15:16	women are not included in the experiment
18578	10/05/2024 13:15:24	Only men, typical medicine
18579	10/05/2024 13:15:33	There is blinding There is blocking

How does it get there?

For now ….

```
var t = HtmlService.createTemplateFromFile(page);
t.code = code;
return t.evaluate().setXFrameOptionsMode(HtmlService.XFrameOptionsMode.ALLOWALL);
```

Google sheets

```
function addRow(newRow, ssid)
{
  var ss = SpreadsheetApp.openById(ssid);
  var sheet = ss.getSheetByName("Sheet1");
  var current_time = new Date();
  newRow.unshift(current_time);
  sheet.appendRow(newRow);
}
```

HTML5

Why did I incorporate certain features for the teacher interface?

Separate

Why did I incorporate certain features for the teacher interface?

Anonymous

Show writings from last 3 minutes

there is no evidence for the alternative hypothesis. there is no evidence that the underlying proportion is not the same for each bird type

there is some evidence p=0.04 that.... I wasn't listening

There is no evidence to support the alternative hypothesis.

There is no evidence that the alternative hypothesis is true (Pval = 0.2776) thus the null model is possible

There is no evidence not compatible with null so it could be same as null

There is no evidence (p value 0.2)that the underlying proportion of diffrent types of bird are diffrent.

Why did I incorporate certain features for the teacher interface?

Reusable

Why did I incorporate certain features for the teacher interface?

how writings from last |3 | minutes

there is no evidence for the
alternative
evidence t
proportio

There is no evidence to suppor

There is no evidence not compatible
with null so it could be same as null

there is some evidence p=0.04 that.... I
wasn't listening

There is no evidence that the
alternativ
0.2776) th

There is no evidence (p value 0.2)that
the underlying proportion of diffrent
types of bird are diffrent.

There is some evidence (p=0.28) for
the Null hypothesis that the
underlying proportion is the same for
each bird type.

Why did I incorporate certain features for the teacher interface?

Editable

Show writings from last 6 minutes

there is no evidence (p that the underlying proportion is not the same for each bird type as p=0.28?

there is no evidence for the alternative hypothesis that the underlying proportion of each bird type are not the same

Why did I incorporate certain features for the teacher interface?

Safety

how writings from last 3 minutes

The uhnderlying proportion is not the sane for each bird type

🟢 this is too hard

tells us that there is no evidence of the alternative model. 🍍

the underlying peoportion is compatible with the null model

There is no evidence that the alternative hypothesis is true (Pval = 0.2776) thus the null model (being that all birds were sent out in equal proportions) is possible

There is no evidence (p value 0.2)that the underlying proportion of diffrent types of bird are diffrent.

Why did I incorporate certain features for the teacher interface?

Flexible

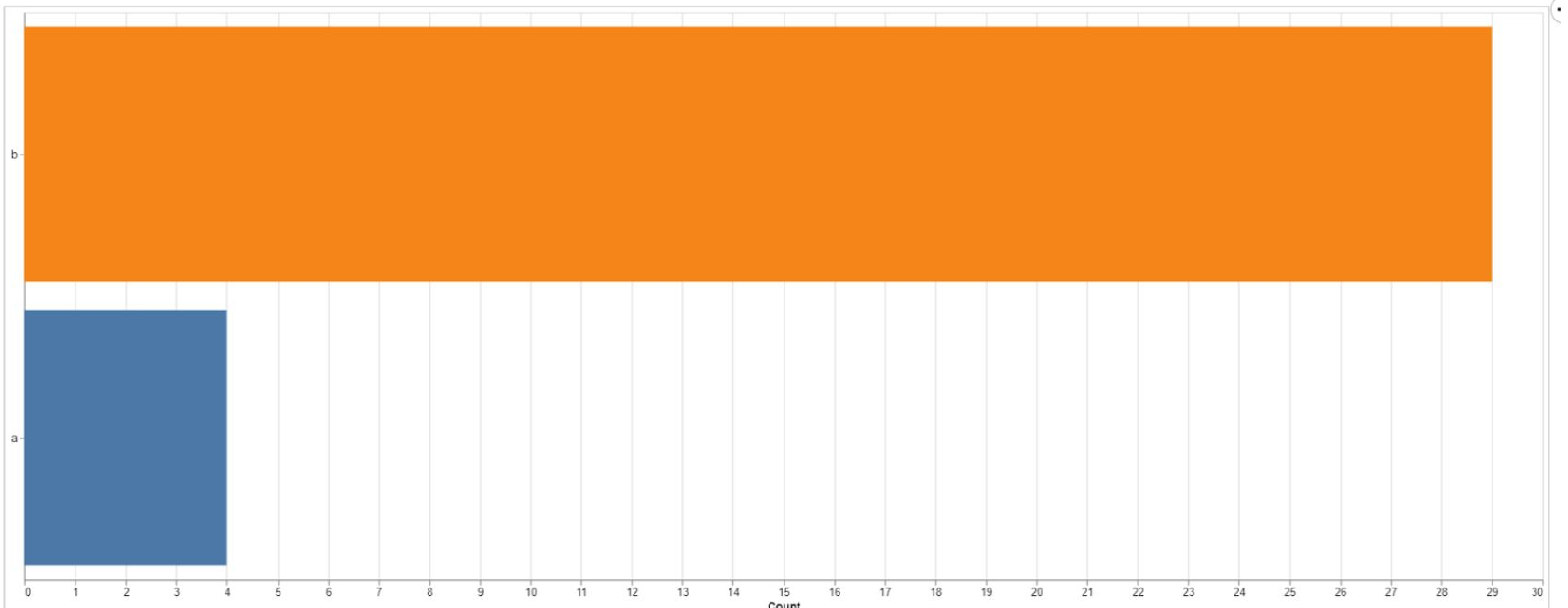Poll students   Visualise results

B	B	B

A	b	A

Why did I incorporate certain features for the teacher interface?

Flexible

Poll students   Visualise results

Using students to give feedback

We trialled giving students pairs of writing examples and asked them to select which one they thought was "better"

there is no evidence (p= 0.2776) that the true proportions in each category are not equal

There is no evidence (pvalue = 0.2776) that the underlying proportion is not the same for each bird type.

We're currently exploring whether using ranking (via Bradley-Terry models) might be helpful for supporting learning.

Ranked number 1

There is no evidence (pvalue = 0.2776) that the underlying proportion is not the same for each bird type.

Ranked number 2

there is no evidence (p= 0.2776) that the true proportions in each category are not equal
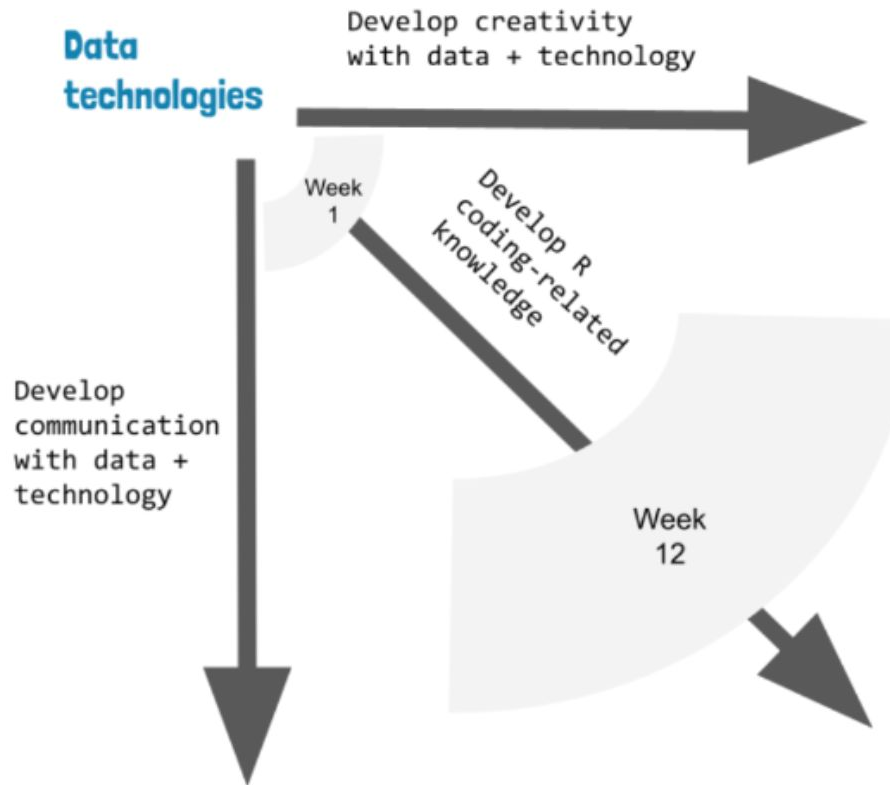
Ranked number 3

There is no evidence not compatible with null so it could be same as null

Ranked number 4

There is some evidence (p=0.28) for the Null hypothesis that the underlying proportion is the same for each bird type.

What about feedback on writing code?

Markdown

Interactive and/or live demonstration!

Using markdown to write about birthday songs!

Level one header, list the top three songs, include the URL/link to the chart for that week

Why did I incorporate certain features for the teacher interface?

Markdown

stimulus · md · html · rendered · bonus

```
# LP
1. Beautiful
2. All I Have
[Link to song chart]
(https://nztop40.co.nz/chart/singles
```

```
# top 3 songs on the week I was
born

1. They Can't Take That Away – Ben
Lummis
```

```
# Top 3 songs when the goat was
born
* F**K IT
* SUPERSTAR
* TOXIC
```

```
# my header
* ignition remix
* crazy in love
[Link to song chart]
(https://nztop40.co.nz/chart/singles
```

```
# Birthday songs!
1. Lose Yourself – Eminem
2. All the Things She Said – t.A.T.u
3. NU-FLOW – Big Brovaz
[Link to song chart]
```

```
# Weewoo weewoo
* Just lose it – Eminem
* Stop the music – P-Money ft.
Scribe
* These words – Natasha Bedingfield
```

```
# My Birthday Songs

*Drop It Like It's hot*
*Swing*
*These Words*
```

```
# The Songs
* They can't take that away
* Burn
* Fools love
[Link to song chart: ]
```

Why did I incorporate certain features for the teacher interface?

Markdown

stimulus   md   **html**   rendered   bonus

Show writings from last `3`   minutes

<h1>LP</h1>

Beautiful
All I Have
<a

<h1>top 3 songs on the week I was born</h1>

They Can't Take That Away – Ben Lummis

<h1>Top 3 songs when the goat was born</h1>

F**K IT
SUPERSTAR

<h1>my header</h1>

ignition remix
crazy in love
<a

<h1>Birthday songs!</h1>

Lose Yourself – Eminem
All the Things She Said – t.A.T.u

<h1>Weewoo weewoo</h1>

Just lose it – Eminem
Stop the music – P-Money ft. Scribe

<h1>My Birthday Songs</h1>
<p>Drop It Like It's hot
Swing
These Words

<h1>The Songs</h1>

They can't take that away
Burn

Why did I incorporate certain features for the teacher interface?

Markdown

stimulus   md   html   rendered   bonus

Show writings from last 3 minutes

LP

1. Beautiful
2. All I Have Link to song chart

top 3 songs on the week I was born

1. They Can't Take That Away – Ben Lummis
2. Yeah – Usher feat. Lil John & Ludacris
3. My Band – D12

song chart link

Top 3 songs when the goat was born

- F**K IT
- SUPERSTAR
- TOXIC Link to song chart

my header

- ignition remix
- crazy in love Link to song chart

Birthday songs!

1. Lose Yourself – Eminem
2. All the Things She Said – t.A.T.u
3. NU–FLOW – Big Brovaz Link to song

Weewoo weewoo

- Just lose it – Eminem

My Birthday Songs

Drop It Like It's hot Swing These Words Link to song chart

The Songs

- They can't take that away

Why did I incorporate certain features for the teacher interface?

Markdown

stimulus   md   html   rendered   **bonus**

What are the challenges to generating feedback on statistical and computational writing?

validity of feedback in terms of data science and statistical content

optimisation of feedback for sustainable learning outcomes

ethicality of data and pedagogical practices

engagement of a diverse range of learners within large-scale social environments

Where to next with research?

Re-design QWT as an ethical open-source research tool
- Share code and produce documentation
- Introduce new features for data collection, including tracking changes to submissions made by users, and storing the data generated by users privately and in de-identifiable format

Leverage large-scale social learning environment for peer assessment
- Incorporate pair-wise comparative judgements
- Rank writing using Bradley-Terry modelling approaches

Use data collected, with appropriate ethics approval:
- Explore learner engagement and submissions over time
- Investigate AI integrations for supporting feedback systems

Dr Anna Fergusson

Department of Statistics | Te Kura Tatauranga

University of Auckland | Waipapa Taumata Rau

a.fergusson@auckland.ac.nz

- [UoA profile](#)
- [Google Scholar](#)
- [Teaching blog](#)
- [Bluesky](#)