# EMI TANAKA

## 💼 WORK EXPERIENCE

- **Senior Lecturer**

  Australian National University,
  Biological Data Science Institute

  2023/07 - current

  ongoing

- **Senior Lecturer / Lecturer**

  Monash University,
  Department of Econometrics and Business Statistics

  2020/01 - 2023/06

  ongoing

- **Lecturer**

  The University of Sydney,
  School of Mathematics and Statistics

  2017/01 - 2019/12

  contract

- **Research fellow**

  University of Wollongong,
  School of Mathematics and Applied Statistics

  2014/01 - 2017/01

  contract

## 🎓 EDUCATION

**PhD, Statistics**

Statistical Methods for Improving Motif Evaluation
Supervisor: Dr. Uri Keich
School of Mathematics and Statistics
The University of Sydney, Sydney, Australia, 2015

**Bachelor of Science (Advanced Mathematics), Honours I**

Major in Mathematics and Statistics
The University of Sydney, Sydney, Australia, 2010

## 📖 PUBLICATIONS

1. Tanaka, E (2022) Getting the most out of your experimental data with design. *Biometric Bulletin*, 39 (4), 9-13.
2. Tanaka, E & Amaliah, D (2022) Current state and prospects of r-packages for the design of experiments. 10.48550/arxiv.2206.07532

## 📇 Contact

- ✉ emi.tanaka@anu.edu.au

- 🐙 github.com/emitanaka

- 🐦 @statsgen

For more information, please contact me via email.

## 🔧 Skills

**Expert**: R, HTML/CSS, LaTeX
**Intermediate**: Git/GitHub, Python, Bash, JS
**Languages**: English (fluent) and Japanese (conversational)

## 🔖 Interests

**Research**: experimental design, mixed models, data visualisation, visual inference, bioinformatics, statistical genetics, selective breeding, software development, statistical workflow
**Non-Research**: drawing (but not good at it), reading (manga, manhwa and non-fiction books)

3. Amaliah, D, Cook, D, Tanaka, E, Hyde, K, & Tierney, N (2022) A journey from wild to textbook data to reproducibly refresh the wages data from the national longitudinal survey of youth database. *Journal of Statistics and Data Science Education*, 0 (ja), 1-27. 10.1080/26939169.2022.2094300

4. Tanaka, E, Leung, J, & Cook, D (2022) Commentary on "visualization in operations management research": Incorporating statistical thinking into visualization practices for decision making in operational management. *INFORMS Journal on Data Science* 10.1287/ijds.2021.0008

5. Cook, D, Reid, N, & Tanaka, E (2021) The foundation is available for thinking about data visualization inferentially. *Harvard Data Science Review* 10.1162/99608f92.8453435d *Citations: 5*.

6. Morota, G, Cheng, H, Cook, D, & Tanaka, E (2021) ASAS-NANP SYMPOSIUM: Prospects for interactive and dynamic graphics in the era of data-rich animal science. *Journal of animal science*, 99 (2) 10.1093/jas/skaa402 *Citations: 5*.

7. Tanaka, E (2020) Simple outlier detection for a multi-environmental field trial. *Biometrics*, 76 (4), 1374-1382. 10.1111/biom.13216 *Citations: 9*.

8. Tanaka, E & Hui, F (2019) Symbolic formulae for linear mixed models. *Statistics and data science*, 3-21. 10.1007/978-981-15-1960-4_1 *Citations: 3*.

9. Hui, F, Tanaka, E, & Warton, D (2018) Order selection and sparsity in latent variable models via the ordered factor LASSO. *Biometrics*, 74 (4), 1311-1319. 10.1111/biom.12888 *Citations: 24*.

10. Norman, A, Taylor, J, Tanaka, E, Telfer, P, Edwards, J, Martinant, J, & Kuchel, H (2017) Increased genomic prediction accuracy in wheat breeding using a large australian panel. *Theoretical and applied genetics*, 130 (12), 2543-2555. 10.1007/s00122-017-2975-4 *Citations: 32*.

11. Tanaka, E, Ral, J, Li, S, Gaire, R, Cavanagh, C, Cullis, B, & Whan, A (2017) Increased accuracy of starch granule type quantification using mixture distributions. *Plant methods*, 13, 107. 10.1186/s13007-017-0259-2 *Citations: 7*.

12. Tanaka, E (2014) Statistical methods for improving motif evaluation. *PhD Thesis*

13. Tanaka, E, Bailey, T, & Keich, U (2014) Improving MEME via a two-tiered significance analysis. *Bioinformatics*, 30 (14), 1965-1973. 10.1093/bioinformatics/btu163 *Citations: 25*.

14. Liachko, I, Tanaka, E, Cox, K, Chung, S, Yang, L, Seher, A, Hallas, L, Cha, E, Kang, G, Pace, H, Barrow, J, Inada, M, Tye, B, & Keich, U (2011) Novel features of ARS selection in budding yeast lachancea kluyveri. *BMC genomics*, 12, 633. 10.1186/1471-2164-12-633 *Citations: 27*.

15. Tanaka, E, Bailey, T, Grant, C, Noble, W, & Keich, U (2011) Improved similarity scores for comparing motifs. *Bioinformatics*,

27 (12), 1603-1609. 10.1093/bioinformatics/btr257*Citations: 60*.

Citation counts are sourced from Google Scholar at 2023-05-27.

# </> SOFTWARE

My software contributions can be found at **https://github.com/emitanaka**.

**ⓡ edibble**

An R-package that implements the grammar of experimental design.
*Creator and maintainer. GitHub stars: 182*
🔗 Source: https://github.com/emitanaka/edibble
🔗 Documentation: https://edibble.emitanaka.org

**ⓡ flipbookr**

An R-package that parses code, creates partial code builds, delivers code movie
*Author. GitHub stars: 189*
🔗 Source: https://github.com/EvaMaeRey/flipbookr
🔗 Documentation: https://evamaerey.github.io/flipbookr/

**ⓡ anicon**

An R-package to insert animated icons for R markdown and Shiny apps.
*Creator. GitHub stars: 122*
🔗 Source: https://github.com/emitanaka/anicon

**ⓡ simulate**

An R-package for a parametric simulation framework to generate complex multivariate and multilevel data.
*Creator and maintainer. GitHub stars: 1*
🔗 Source: https://github.com/emitanaka/simulate

**ⓡ nestr**

An R-package to build hierarchical or nested sturctures.
*Creator and maintainer. GitHub stars: 13*
🔗 Source: https://github.com/emitanaka/nestr
🔗 Documentation: https://nestr.emitanaka.org

**ⓡ monash**

A utility R-package with consolidated tools and templates for staffs at Monash University.
*Creator and maintainer. GitHub stars: 15*
🔗 Source: https://github.com/numbats/monash

**ⓡ sugar**

An R-package for instructors to create a shiny app for particular unit to review grade and attendance of students.
*Supervisor.GitHub stars: 1*
🔗 Source: https://github.com/numbats/sugar
🔗 Documentation: https://numbats.github.io/sugar/

**ⓡ xaringan**

An R package for creating slideshows with remark.js through R Markdown.
*Contributor. My contribution is the adaptation of the ninja-theme shown in the Documentation. GitHub stars: 1443*
🔗 Source: https://github.com/yihui/xaringan
🔗 Documentation: https://github.com/emitanaka/ninja-theme

### ® anidb

An R-package to fetch data from the anime database at AniDB.
*Creator and maintainer. GitHub stars: 1*
🔗 Source: https://github.com/emitanaka/anidb

### ® ggmatplot

Plot Columns of Two Matrices Against Each Other Using 'ggplot2'
*Author. GitHub stars: 5*
🔗 Source: https://github.com/xuan-liang/ggmatplot
🔗 Documentation: https://xuan-liang.github.io/ggmatplot/

### ® gghdr

Plots of highest density regions (HDR) for ggplot2.
*Listed as author but my contribution is small. GitHub stars: 43*
🔗 Source: https://github.com/ropenscilabs/gghdr
🔗 Documentation: https://ropenscilabs.github.io/gghdr

### ® shinycustomloader

Add a custom loader for R shiny.
*Creator. GitHub stars: 113*
🔗 Source: https://github.com/emitanaka/shinycustomloader

### ® datalegreyar

Datalegreya, the typeface that melts text and data visualisation, for R markdown.
*Creator. GitHub stars: 42*
🔗 Source: https://github.com/emitanaka/datalegreyar

### </> Tomtom

A motif comparison tool part of the MEME Suite.
*The -incomplete-scores is an implementation as a result of Tanaka et al. (2011).*
🔗 Documentation: https://meme-suite.org/meme/doc/tomtom.html

# 💬 TALKS

Links to the slides (if available) are at **https://emitanaka.org/talks/**.

Below show the last 10 invited talks.

- ● Towards a unified language in experimental designs
  Australian Data Science Network
  2022-11-21

- ● The genesis of experimentation
  Ihaka Lecture Series

2022-07-28

- **Getting the Most Out of Your Experimental Data With Design**
  Monash Data Futures Insitute
  2022-07-07

- **Current state and prospects of R-packages for the design of experiments**
  Statistical Society of Australia Canberra Branch
  2022-06-29

- **An anthology of experimental designs**
  Toronto Data Workshop
  2022-03-25

- **The Art of Design: Software & Experimental Design**
  R-Ladies Melbourne
  2021-10-26

- **Current state and prospects of R-package for the design of experiments**
  University of Auckland Statistics Seminar
  2021-10-14

- **Advent of "Grammar": Bridging Statistics and Data Science for the Design of Experiments**
  Statistical Society of Australia
  2020-11-27

- **Rethinking the framework to specify the design of experiments**
  Department of Agriculture and Fisheries, Queensland
  2020-11-11

- **What every young statistician should know: Developing your identity, networking and marketing**
  Statistical Society of Australia Western Australian Branch
  2020-09-30

## 🧰 WORKSHOPS

| Workshop | Date | Location | Host |
|---|---|---|---|
| Effective Data Visualisation with R | 6 Dec 2022 | Online | WOMBAT |
| Data Visualisation with R | 28 Nov 2022 | Inverloch | Australasian Applied Statistics Conference |

| Workshop | Date | Location | Host |
|---|---|---|---|
| Data Visualisation with R | 21-22 Feb 2021 | Online | methods @ manchester |
| Advanced Data Visualisation with R | 8-9 Dec 2021 | Online | Statistical Society of Australia, Canberra Branch |
| Data Visualisation with R | 6 Dec 2021 | Online | Statistical Society of Australia, NSW Branch |
| Data Visualisation with R | 15-16 Apr 2021 | Online | Statistical Society of Australia, Canberra Branch |
| Data Wrangling with R | 1-2 Dec 2020 | Online | Statistical Society of Australia, NSW Branch |
| Data Visualisation with R | 11-12 Nov 2020 | Online | Statistical Society of Australia, Victoria Branch |
| Data Visualisation with R | 28-29 Jul 2020 | Online | Statistical Society of Australia, Victoria Branch |
| Tidyverse & R Markdown | 1 Dec 2019 | Adelaide | International Biometrics Society, Australasia Region |
| Building R Packages & R Markdown | 19 Nov 2019 | Melbourne | Statistical Society of Australia, Victoria Branch |
| Communicating with Data via R Markdown | 4 Oct 2019 | Sydney | COMBINE (subcommittee of ABACBS) |
| Gaining skills in biostatistical consultancy | 4 Jul 2019 | Melbourne | Statistical Society of Australia, Biostatistics Section |
| Statistical Methods for Omics Assisted Breeding | 12-15 Nov 2018 | Tokyo | University of Tokyo |

# 🏫 TEACHING

- ETC5523: Communicating with Data
- ETC5512: Wild-Caught Data
- ETC5521: Explonatory Data Analysis
- STAT3012: Applied Linear Models
- STAT5002: Introduction to Statistics
- STAT906: Experimental Design

# ♜ PROFESSIONAL SERVICE

- **Associate Editor**

  R Journal, https://journal.r-project.org/

  2022/02-ongoing

- **President**

  Statistical Society of Australia, Victoria Branch

  2021/03-ongoing

- **Vice President**

  Statistical Society of Australia, Victoria Branch

  2023/03-2024/03

- **Assistant Secretary**

  Statistical Society of Australia, NSW Branch

  2017/03-2018/03

- **Secretary**

  Statistical Society of Australia, NSW Branch

  2018/03-2020/03

- **Social Media Coordinator**

  International Biometrics Society, Australaisan Region

  2018/01-ongoing

- **Member**

  useR! 2021, Program Committee

  2021/03-2021/03

- **Member**

  Australasian Applied Statistics Conference 2020/2022, Organising Committee

  2020/01-ongoing

- **Member**

  Australasian Applied Statistics Conference 2016, Program and Review Committee

  2016/07-2016/12

- **Chair**

  Australasian Applied Statistics Conference 2016, Local Organisation Committee

  2016/01-2016/12

# 👥 PROFESSIONAL MEMBERSHIPS

- Statistical Society of Australia
- International Biometrics Society
- American Statistical Association

# 🏅 AWARDS & DISTINCTIONS

- **SSA President's Award for Leadership in Statistics, awarded to the 2020-2022 Branch Councils**

  Statistical Society of Australia

  2022

- **ARC Linkage Project**

  *Chief Investigator*
  Three-dimensional Bayesian Modelling of Geological and Geophysical data
  Australian Research Council

  2022

  $554,188

- **Distinguished Presenter's Award**

  Statistical Society of Australia

  2021

- **ARC Centre of Excellence**

  *Associate Investigator*
  ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)
  Australian Research Council

  2020

- **ARC Industrial Transformation Training Centre**

  *Chief Investigator*
  Data Analytics for Resources and Environments (DARE)
  Australian Research Council

  2019

  $3,973,202

- **R Consortium Grant**

  *Lead Investigator*
  Symbolic Formulae for Linear Mixed Models
  R Consortium

  2019

  $6,000

- **Sydney Institute of Agriculture Research Project Grant**

*Chief Investigator*
Land surface models of carbon and water do not work in agricultural landscapes where it actually matters they work
Sydney Institute of Agriculture

2018

$90,000

- Faculty of Science Research Equipment and Infrastructure Scheme

  *Chief Investigator*
  Sydney bioinformatics and biometrics big data research server
  The University of Sydney

  2018

  $36,040

- J.B. Douglas Award

  Nomination and Prize
  Statistical Society of Australia, NSW Branch

  2011

  $500

- Tim Brown Prize II

  The University of Sydney

  2008

  $275

- Barker Scholarship IV

  The University of Sydney

  2008

  $750