Thom Benjamin Volker

Utrecht, The Netherlands

thomvolker.github.io t.b.volker@uu.nl GitHub LinkedIn

EDUCATION

MSc Methodology and Statistics **2019 - 2022**
Utrecht University, Netherlands *Cum laude (GPA 8.9/10)*

- Valedictorian (Faculty of Social and Behavioral Sciences, Utrecht University).
- Thesis - Combining support for hypotheses over heterogeneous studies with Bayesian Evidence Synthesis: A simulation study (supervised by Irene Klugkist).

MSc Sociology and Social Research **2019 - 2022**
Utrecht University, Netherlands *Cum laude (GPA 8.7/10)*

- Thesis - The future is made today: Concerns for reputation foster trust and cooperation (supervised by Vincent Buskens & Werner Raub).

BA Liberal Arts & Sciences **2016 - 2019**
Utrecht University, Netherlands *(GPA 7.7/10)*

- Major in Pedagogical Sciences; minor in Sociology and Social Research

EXPERIENCE

PhD-Candidate Synthetic data **2022 - Current**
Utrecht University

- I investigate how to optimize the creation of synthetic data in terms of the risk-utility trade-off.
- Supervisors: Erik-Jan van Kesteren, Peter-Paul de Wolf, Stef van Buuren

Internship MICE Group **2020 - 2022**
Utrecht University

- With Gerko Vink I established a workflow on how to generate synthetic data with the `R`-package `mice` to preserve privacy and confidentiality.

Research Assistant **2019 - 2022**
Utrecht University

- For Peter Lugtig, I worked on research and teaching related topics, (e.g., creating content for the research master level course 'Survey Data Analysis', by developing a Shiny App, creating data visualizations and creating the webpage https://www.peterlugtig.com.
- For Rebecca Kuiper, I created data visualizations and assisted PhD-candidates by revising text.
- For Laura Hofstee and Rens van de Schoot, I contributed to ASReview, an automated text screening service to assist in systematic reviews.

TEACHING

- **Post graduate-level courses**
  - Multiple Imputation in Practice [2019 - Current]: *Supervising practicals.*
  - Statistical Programming with R [2021]: *Supervising practicals.*
  - Advanced Survey Design [2021]: *Supervising practicals.*
- **Master level courses**
  - Battling the curse of dimensionality [2022 - Current]: *Developing course materials, supervising practical sessions and grading exams.*
  - Network Analysis [2020 - 2021]: *Developing course materials, supervising practicals and grading exams.*
  - Data wrangling [2021 - 2023]: *Supervising practicals and grading exams.*

  - Methodological and Statistical Aspects of Social Science Research [2019 - 2020]: *Supervising practicals.*
- **Bachelor level courses**
  - Theory Construction and Statistical Modeling [2019- 2020]: *Supervising practicals*
  - Various undergraduate courses on standard statistical methods [2018-2021].

SUPERVISION

- Heleen Brüggen (Master's thesis, co-supervised with Hanne Oberman & Gerko Vink, 2023-2024).
- Mirthe Hendriks (Master's thesis, co-supervised with Gerko Vink, 2021).
- Stijn van den Broek (Master's thesis, co-supervised with Gerko Vink, 2021).
- Romain Rey (Bachelor's thesis, co-supervised with Irene Klugkist, 2020).

MISCELLANEA

IOPS PhD Representative [2022 - Current]: Represent PhD candidates within the Interuniversity Graduate School of Psychometrics and Sociometrics.
Social committee [2022 - Current]: Organising frequent social activities for the Methods & Statistics department.
Debuut [2019 - 2020]: Participate in buddy programme at Utrecht University to match potential first generation university students ($\sim$ 11 year olds) with actual students to get a flavour of what studying at a university entails.

SOFTWARE PROFICIENCY

- `R`
- `MPlus`
- `SPSS`
- `HLM`
- `JAGS`
- Learning `python`
- Learning `C++`

RECEIVED FUNDING

SynthEval **2023-2024**
Project lead
- An R package for evaluating and improving the quality of synthetic data sets. For €15.000.

PUBLICATIONS

- **Volker, T. B.** & Klugkist, I. (*Submitted*). Combining support for hypotheses over heterogeneous studies with Bayesian Evidence Synthesis: A simulation study. arXiv.
- Klugkist, I. & **Volker, T. B.** (2023). Bayesian Evidence Synthesis for Informative Hypotheses: An introduction. *Psychological Methods, Advance online publication* Paper.
- Research Data Management Support (2023). Data Privacy Handbook. Online book
- Van Leeuwen, A., Bagheri, A., **Volker, T. B.** & Van Brakel, Charlotte (2023). Verkenning van de inzet van topic modelling bij het analyseren van schrijfopdrachten in de context van de flipped classroom. [Exploring the use of topic modelling for the analysis of written assignments in a flipped classroom context.] *Tijdschrift voor Hoger Onderwijs, 41* (2), 84-98. Paper.
- **Volker, T. B.** & Vink, G. (2021). Anony*mice*d Shareable Data: Using `mice` to Create and Analyze Multiply Imputed Synthetic Datasets. *psych, 3*, 703-716. Paper.

- **Volker, T. B.**, Buskens, V. & Raub, W. (In preparation). The future is made today: Concerns for reputation foster trust and cooperation. Paper.

WORKSHOPS & PRESENTATIONS

- Measuring utility of synthetic data [in Dutch], Knowledge Network Synthetic Data, 19 December 2023, Utrecht, Netherlands. Presentation
- Assessing the utility of synthetic data: A density ratio perspective [with Peter-Paul de Wolf & Erik-Jan van Kesteren], UNECE Expert Meeting on Statistical Data Confidentiality, 26-28 September 2023, Wiesbaden, Germany. Presentation
- Fake it 'till you make it: Generating synthetic data with high utility in `R` [with Erik-Jan van Kesteren], Tübingen Open Science Winter School, 13 February 2023, Tübingen, Germany. Course materials
- Creating synthetic data with `mice` in `R` [with Gerko Vink], UMCU, 4 November 2022, Utrecht, the Netherlands. Course materials

SOFTWARE VIGNETTES

- **Volker, T. B.** & Vink, G. (2022). `futuremice`: The future starts today. `https://www.gerkovink.com/futuremice/Vignette_futuremice.html`

SOFTWARE DEVELOPMENT

- `densityratio` [aut] - Developed `R`-package for directly estimating the ratio beteen two sample densities.
- `mice` [ctb] - Developed functionality to pool inferences from synthetic data and implemented `futuremice` to speed up imputation by using user-friendly parallel processing.
- `ggmice` [ctb] - Adjusted functionality to allow for visualization of synthetic data.