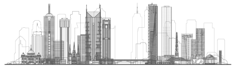
Elevating your research through software development

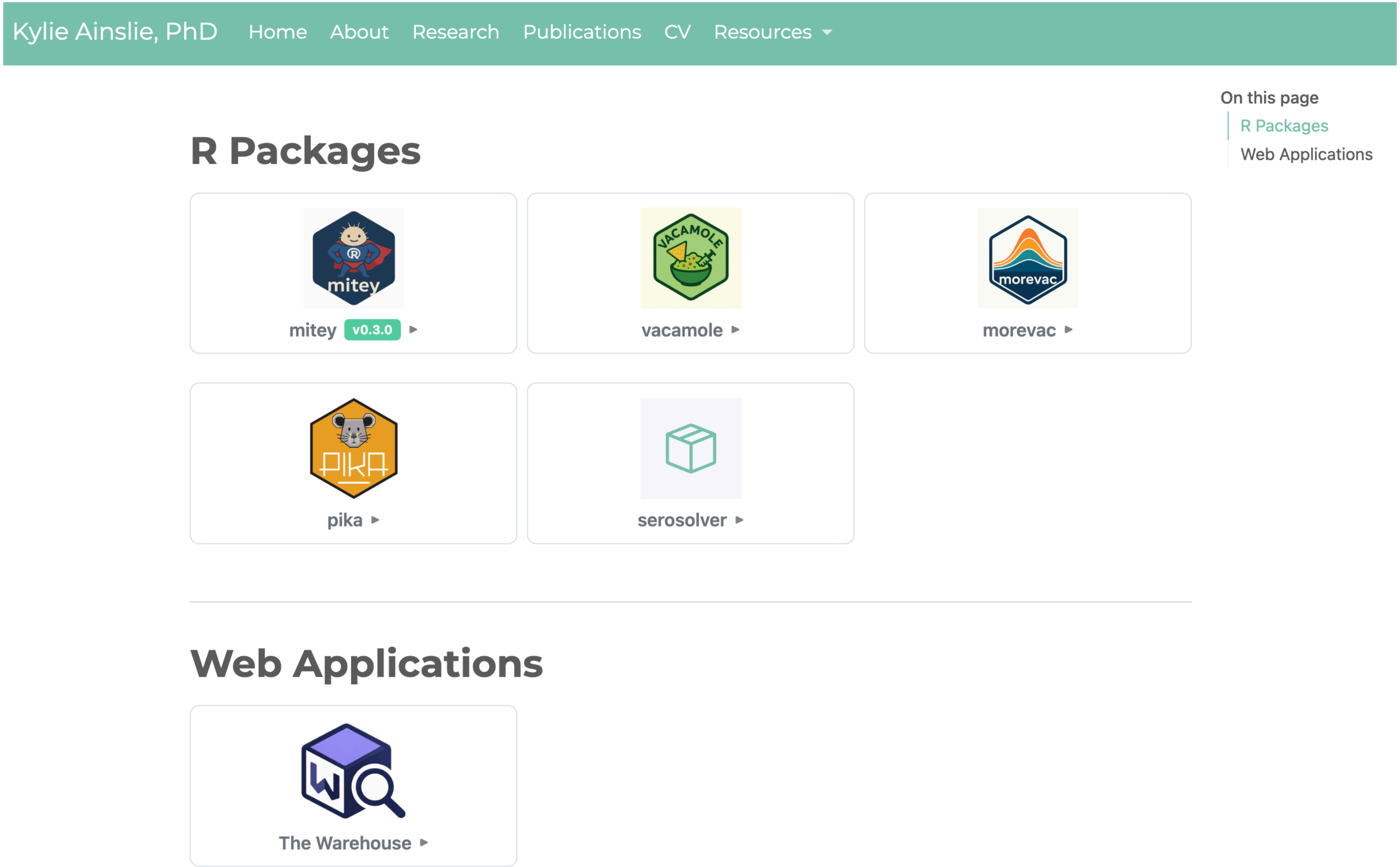
Kylie Ainslie



How many consider yourself software developers?

Me either.

And yet…

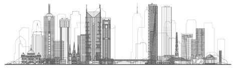


R Packages

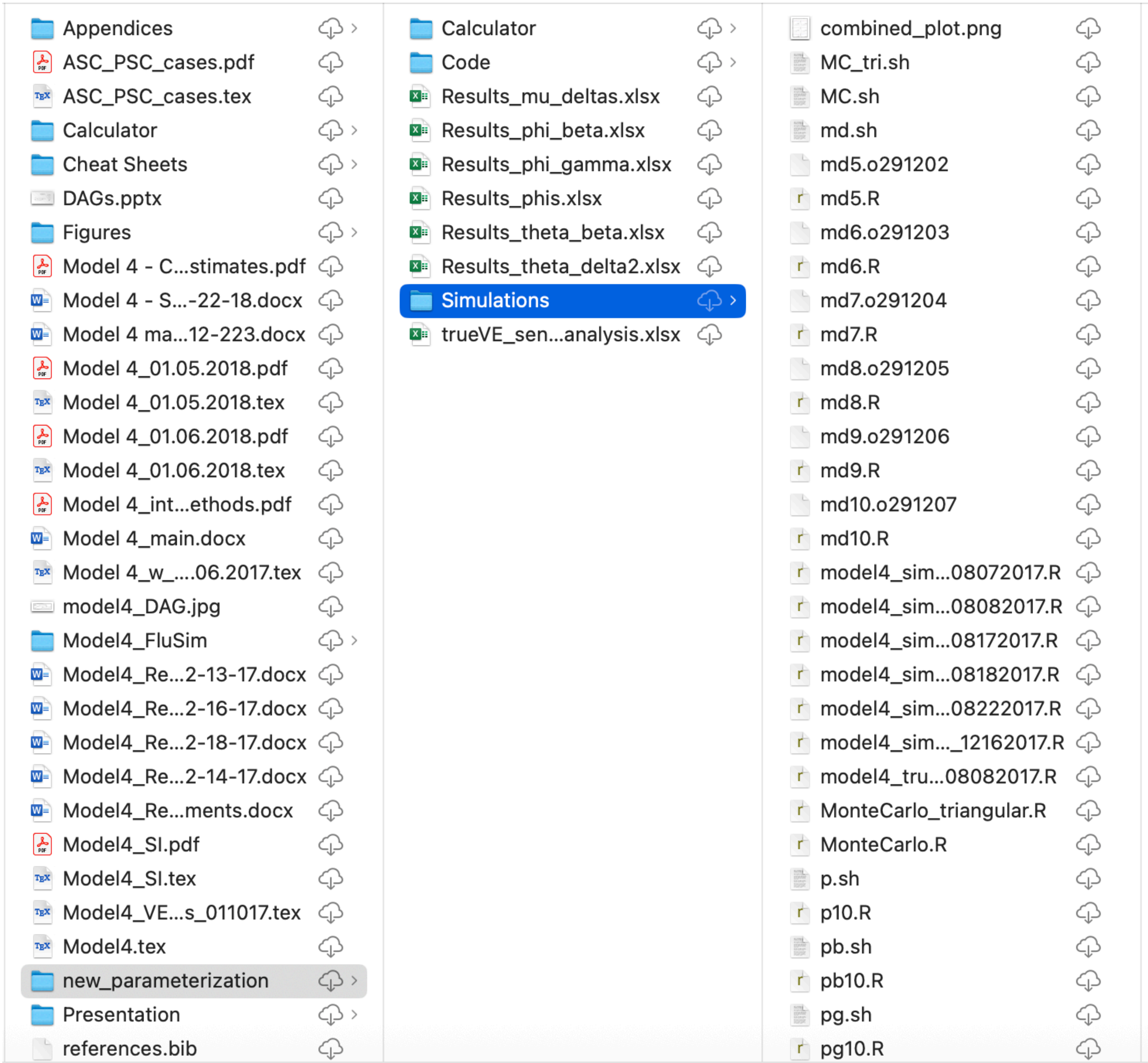
- mitey v0.3.0 ▸
- vacamole ▸
- morevac ▸
- pika ▸
- serosolver ▸

Web Applications

- The Warehouse ▸



Look familiar?

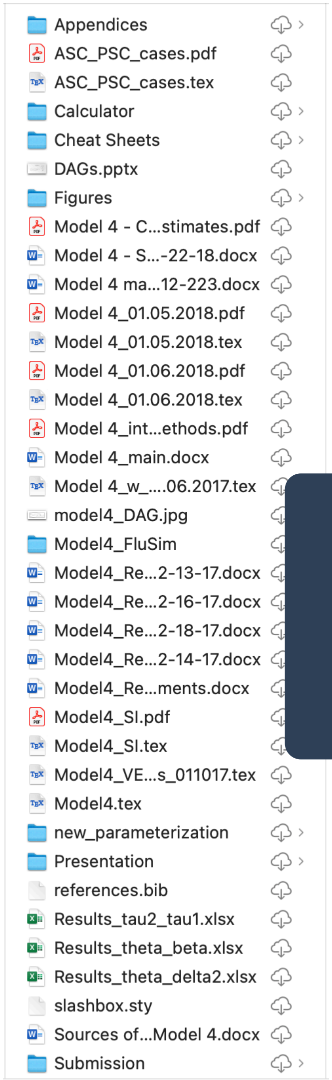


Appendices
ASC_PSC_cases.pdf
ASC_PSC_cases.tex
Calculator
Cheat Sheets
DAGs.pptx
Figures
Model 4 - C...stimates.pdf
Model 4 - S...-22-18.docx
Model 4 ma...12-223.docx
Model 4_01.05.2018.pdf
Model 4_01.05.2018.tex
Model 4_01.06.2018.pdf
Model 4_01.06.2018.tex
Model 4_int...ethods.pdf
Model 4_main.docx
Model 4_w_....06.2017.tex
model4_DAG.jpg
Model4_FluSim
Model4_Re...2-13-17.docx
Model4_Re...2-16-17.docx
Model4_Re...2-18-17.docx
Model4_Re...2-14-17.docx
Model4_Re...ments.docx
Model4_SI.pdf
Model4_SI.tex
Model4_VE...s_011017.tex
Model4.tex
new_parameterization
Presentation
references.bib

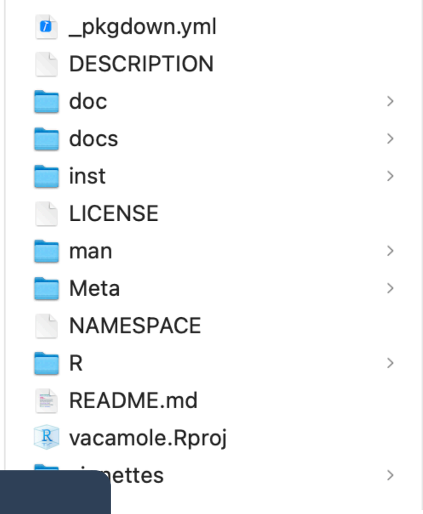
Calculator
Code
Results_mu_deltas.xlsx
Results_phi_beta.xlsx
Results_phi_gamma.xlsx
Results_phis.xlsx
Results_theta_beta.xlsx
Results_theta_delta2.xlsx
Simulations
trueVE_sen...analysis.xlsx

combined_plot.png
MC_tri.sh
MC.sh
md.sh
md5.o291202
md5.R
md6.o291203
md6.R
md7.o291204
md7.R
md8.o291205
md8.R
md9.o291206
md9.R
md10.o291207
md10.R
model4_sim...08072017.R
model4_sim...08082017.R
model4_sim...08172017.R
model4_sim...08182017.R
model4_sim...08222017.R
model4_sim..._12162017.R
model4_tru...08082017.R
MonteCarlo_triangular.R
MonteCarlo.R
p.sh
p10.R
pb.sh
pb10.R
pg.sh
pg10.R

What if there was a better way?



Appendices
ASC_PSC_cases.pdf
ASC_PSC_cases.tex
Calculator
Cheat Sheets
DAGs.pptx
Figures
Model 4 - C...stimates.pdf
Model 4 - S...-22-18.docx
Model 4 ma...12-223.docx
Model 4_01.05.2018.pdf
Model 4_01.05.2018.tex
Model 4_01.06.2018.pdf
Model 4_01.06.2018.tex
Model 4_int...ethods.pdf
Model 4_main.docx
Model 4_w_....06.2017.tex
model4_DAG.jpg
Model4_FluSim
Model4_Re...2-13-17.docx
Model4_Re...2-16-17.docx
Model4_Re...2-18-17.docx
Model4_Re...2-14-17.docx
Model4_Re...ments.docx
Model4_SI.pdf
Model4_SI.tex
Model4_VE...s_011017.tex
Model4.tex
new_parameterization
Presentation
references.bib
Results_tau2_tau1.xlsx
Results_theta_beta.xlsx
Results_theta_delta2.xlsx
slashbox.sty
Sources of...Model 4.docx
Submission



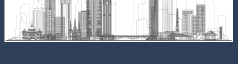
_pkgdown.yml
DESCRIPTION
doc
docs
inst
LICENSE
man
Meta
NAMESPACE
R
README.md
vacamole.Rproj
ettes

How?

Structure your project like a software package

1. Organise
2. Share
3. Innovate

Organise



Organise

❌ Typical Project

```
myproject/
├── analysis.R
├── analysis_FINAL.R
├── analysis_FINAL_v2.R
├── code_guide.pdf
├── data_raw.csv
├── final_report.pdf
├── final_report_FINAL.pdf
├── functions.R
├── methodology.docx
└── more_functions.R
```

✅ Package Structure

```
📦 mypackage/
├── R/             # Code
├── inst/extdata/  # Data + Scripts
├── man/           # Help files
└── vignettes/     # Analysis
```

Organise

❌ Typical Project

```
myproject/
├── analysis.R
├── analysis_FINAL.R
├── analysis_FINAL_v2.R
├── code_guide.pdf
├── data_raw.csv
├── final_report.pdf
├── final_report_FINAL.pdf
├── functions.R
├── methodology.docx
└── more_functions.R
```

✅ Package Structure

```
📦 mypackage/
├── R/                # Code
│   ├── functions.R
│   └── more_functions.R
├── inst/extdata/ # Data + Scripts
├── man/          # Help files
└── vignettes/    # Analysis
```

Organise

❌ Typical Project

```
myproject/
├── analysis.R
├── analysis_FINAL.R
├── analysis_FINAL_v2.R
├── code_guide.pdf
├── data_raw.csv
├── final_report.pdf
├── final_report_FINAL.pdf
├── functions.R
├── methodology.docx
└── more_functions.R
```

✅ Package Structure

```
📦 mypackage/
├── R/             # Code
├── inst/extdata/  # Data + Scripts
    ├── analysis.R
    ├── analysis_FINAL.R
    ├── analysis_FINAL_v2.R
    └── data_raw.csv
├── man/           # Help files
└── vignettes/     # Analysis
```

Organise

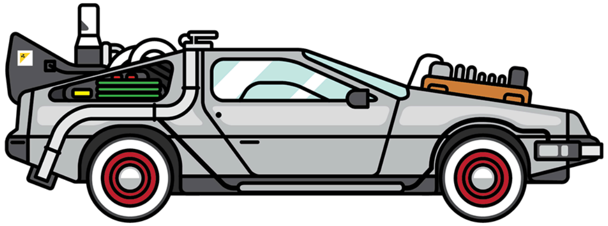
❌ Typical Project

```
myproject/
├── analysis.R
├── analysis_FINAL.R
├── analysis_FINAL_v2.R
├── code_guide.pdf
├── data_raw.csv
├── final_report.pdf
├── final_report_FINAL.pdf
├── functions.R
├── methodology.docx
└── more_functions.R
```

✅ Package Structure

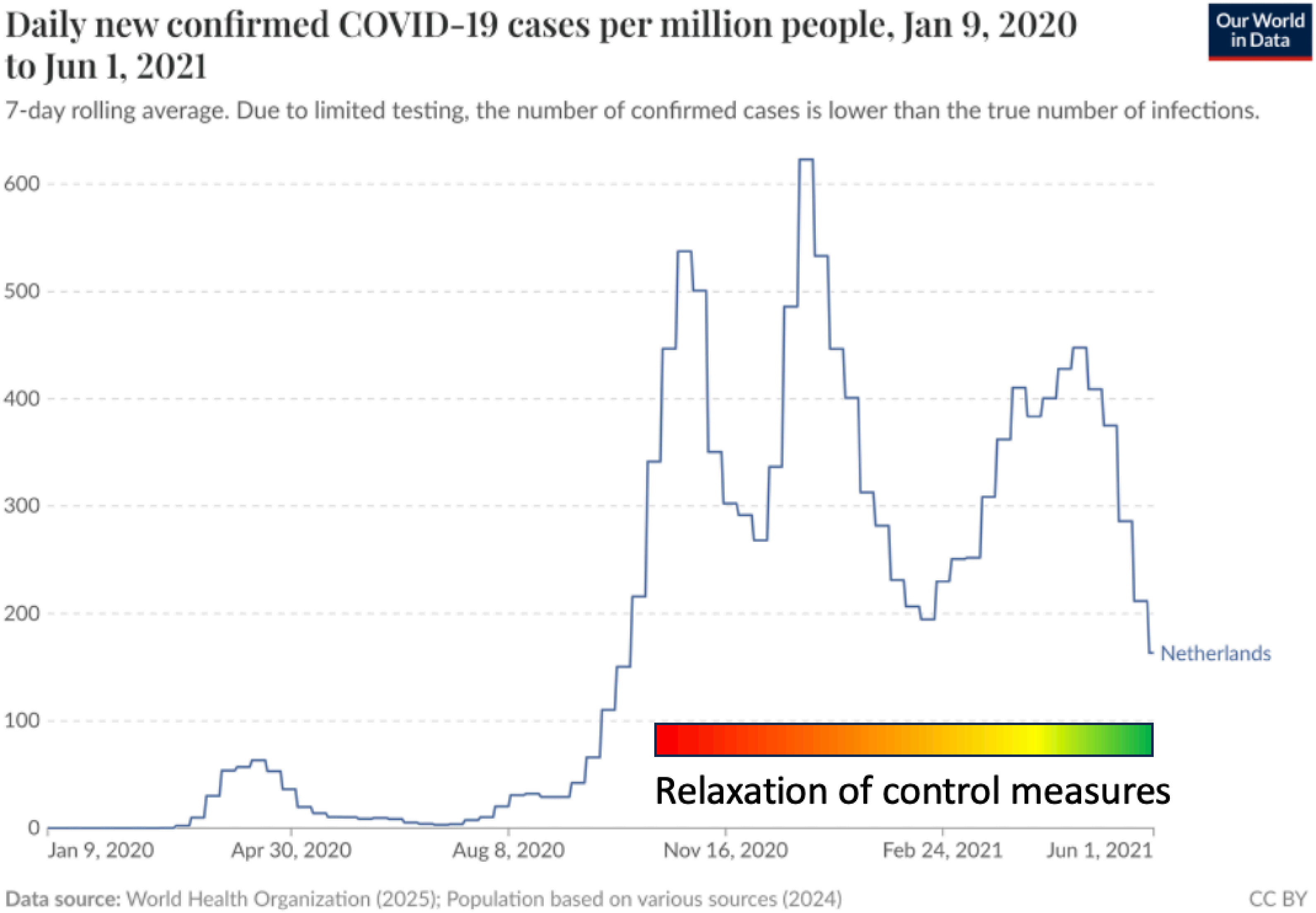
```
📦 mypackage/
├── R/             # Code
├── inst/extdata/  # Data + Scripts
├── man/           # Help files
└── vignettes/     # Analysis
    ├── code_guide.Rmd
    ├── methodology.Rmd
    └── report.Rmd
```

Let’s travel back in time



June 2021 in the Netherlands

June 2021 in the Netherlands



Daily new confirmed COVID-19 cases per million people, Jan 9, 2020 to Jun 1, 2021

7-day rolling average. Due to limited testing, the number of confirmed cases is lower than the true number of infections.

Relaxation of control measures

Data source: World Health Organization (2025); Population based on various sources (2024)

CC BY

June 2021 in the Netherlands

Should the Netherlands extend COVID-19 vaccination to adolescents and children?

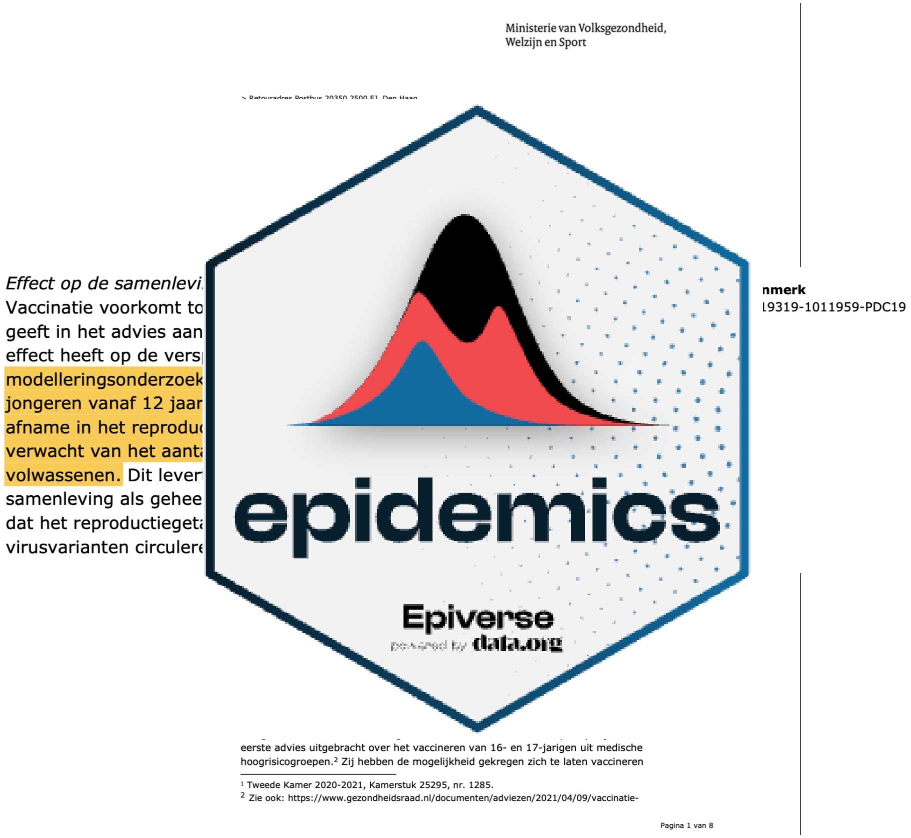
Structure

✅ **Package Structure**

```
📦 mypackage/
├── R/              # Code
├── inst/extdata/ # Data + Scripts
├── man/            # Help files
└── vignettes/      # Analysis
```

Structure → Efficiency

- 🔄 Models rapidly updated as new data came in
- ⏰ Analysis delivered on time
- 💉 Directly informed national health policy



Ministerie van Volksgezondheid,
Welzijn en Sport

> Retouradres Postbus 20350 2500 EJ Den Haag

nmerk
9319-1011959-PDC19

Effect op de samenlevi
Vaccinatie voorkomt to
geeft in het advies aan
effect heeft op de vers
modelleringsonderzoek
jongeren vanaf 12 jaar
afname in het reproduc
verwacht van het aanta
volwassenen. Dit lever
samenleving als gehee
dat het reproductiegeta
virusvarianten circulere

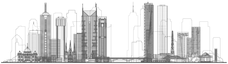
eerste advies uitgebracht over het vaccineren van 16- en 17-jarigen uit medische
hoogrisicogroepen.[2] Zij hebben de mogelijkheid gekregen zich te laten vaccineren

---

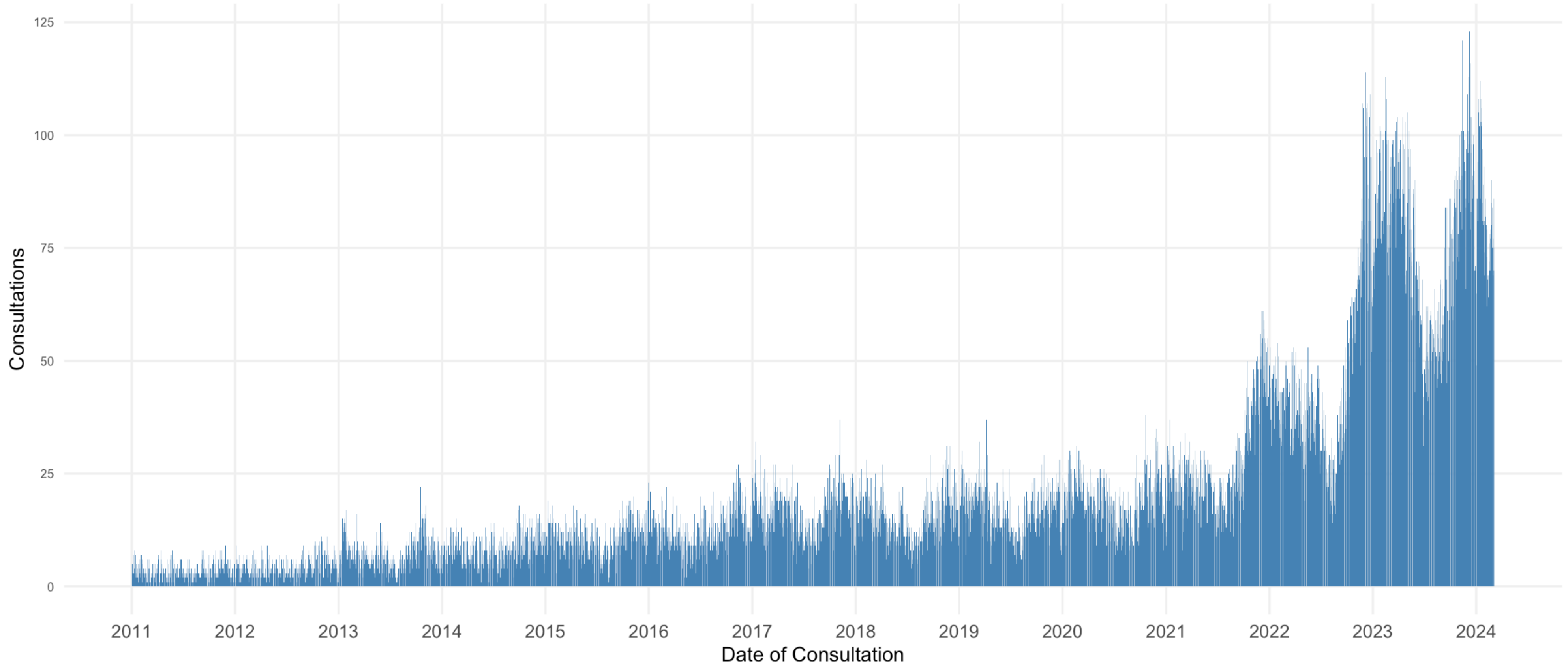
[1] Tweede Kamer 2020-2021, Kamerstuk 25295, nr. 1285.
[2] Zie ook: https://www.gezondheidsraad.nl/documenten/adviezen/2021/04/09/vaccinatie-

Pagina 1 van 8

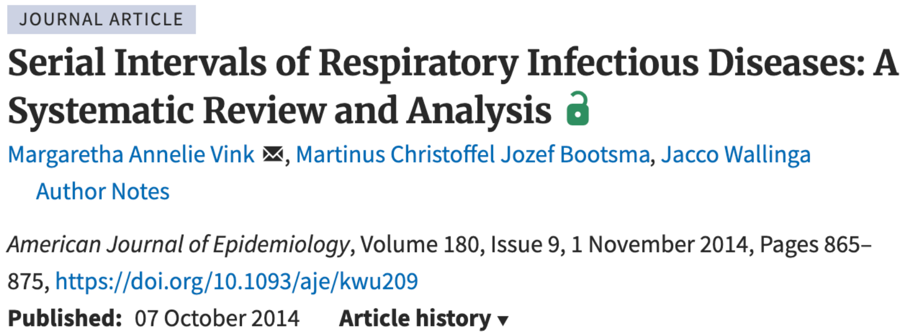
Share



Scabies in the Netherlands



Vink et al.



JOURNAL ARTICLE

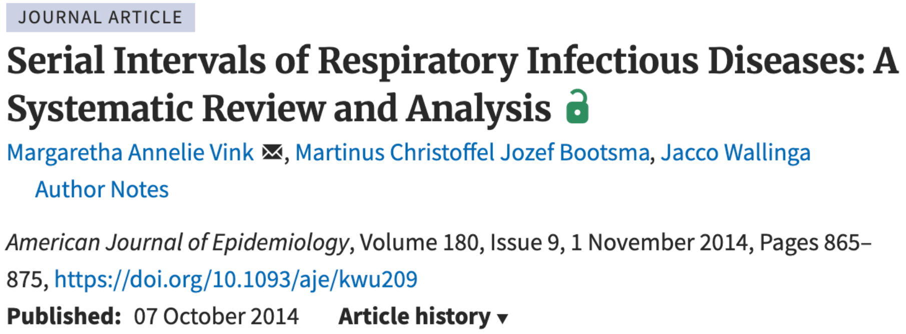
Serial Intervals of Respiratory Infectious Diseases: A Systematic Review and Analysis

Margaretha Annelie Vink ✉, Martinus Christoffel Jozef Bootsma, Jacco Wallinga
Author Notes

American Journal of Epidemiology, Volume 180, Issue 9, 1 November 2014, Pages 865–875, https://doi.org/10.1093/aje/kwu209

Published: 07 October 2014 **Article history** ▾

Vink et al.



JOURNAL ARTICLE

Serial Intervals of Respiratory Infectious Diseases: A Systematic Review and Analysis

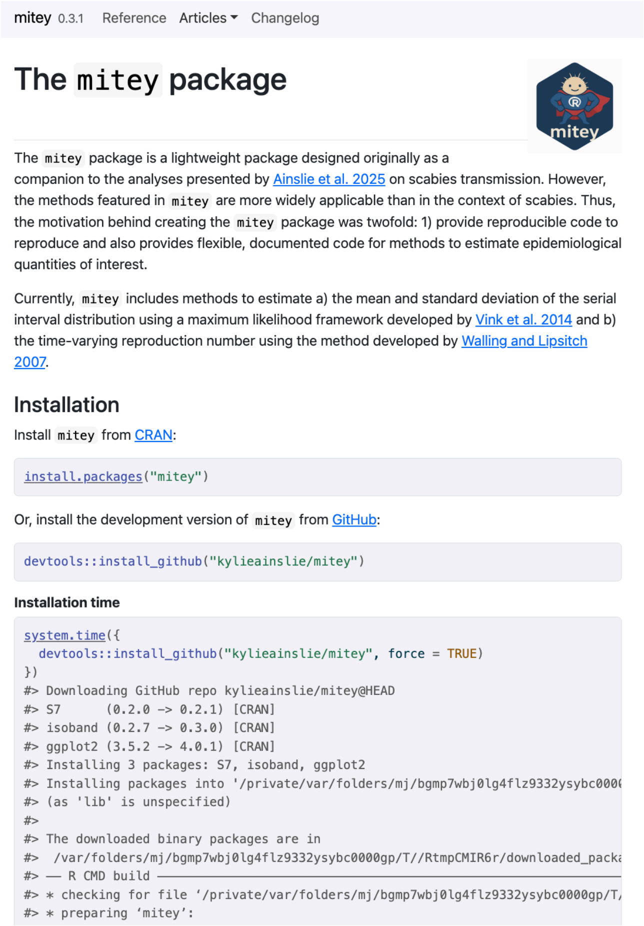
Margaretha Annelie Vink, Martinus Christoffel Jozef Bootsma, Jacco Wallinga

Author Notes

American Journal of Epidemiology, Volume 180, Issue 9, 1 November 2014, Pages 865–875, https://doi.org/10.1093/aje/kwu209

Published: 07 October 2014 **Article history** ▾

Document



mitey 0.3.1 Reference Articles Changelog

The `mitey` package

The `mitey` package is a lightweight package designed originally as a companion to the analyses presented by Ainslie et al. 2025 on scabies transmission. However, the methods featured in `mitey` are more widely applicable than in the context of scabies. Thus, the motivation behind creating the `mitey` package was twofold: 1) provide reproducible code to reproduce and also provides flexible, documented code for methods to estimate epidemiological quantities of interest.

Currently, `mitey` includes methods to estimate a) the mean and standard deviation of the serial interval distribution using a maximum likelihood framework developed by Vink et al. 2014 and b) the time-varying reproduction number using the method developed by Walling and Lipsitch 2007.

Installation

Install `mitey` from CRAN:

```
install.packages("mitey")
```

Or, install the development version of `mitey` from GitHub:

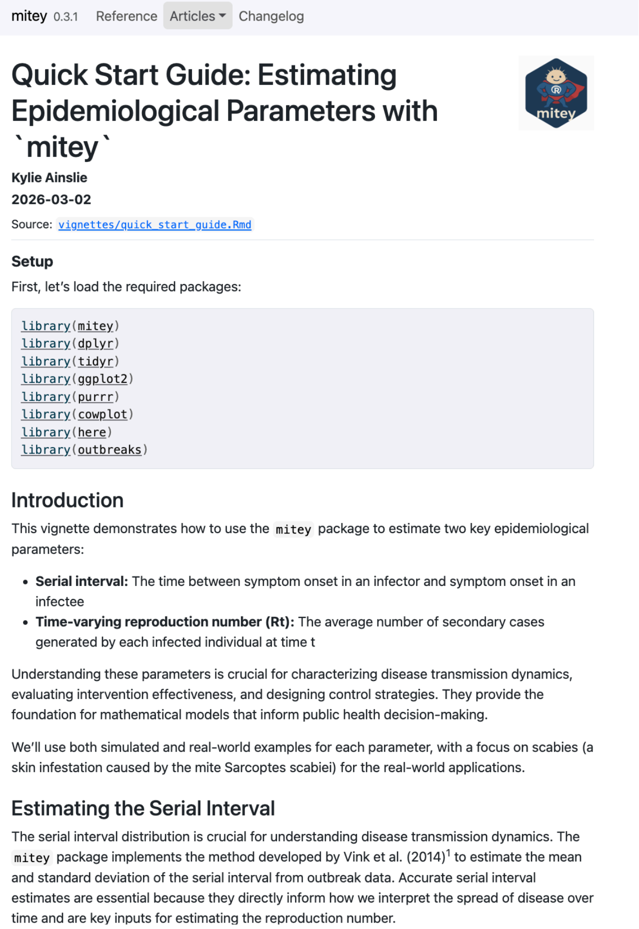
```
devtools::install_github("kylieainslie/mitey")
```

Installation time

```
system.time({
  devtools::install_github("kylieainslie/mitey", force = TRUE)
})
#> Downloading GitHub repo kylieainslie/mitey@HEAD
#> S7      (0.2.0 -> 0.2.1) [CRAN]
#> isoband (0.2.7 -> 0.3.0) [CRAN]
#> ggplot2 (3.5.2 -> 4.0.1) [CRAN]
#> Installing 3 packages: S7, isoband, ggplot2
#> Installing packages into '/private/var/folders/mj/bgmp7wbj0lg4flz9332ysybc000
#> (as 'lib' is unspecified)
#>
#> The downloaded binary packages are in
#>  /var/folders/mj/bgmp7wbj0lg4flz9332ysybc0000gp/T//RtmpCMIR6r/downloaded_packa
#> ── R CMD build ──────────────────────────────────────────────
#> * checking for file ‘/private/var/folders/mj/bgmp7wbj0lg4flz9332ysybc0000gp/T
#> * preparing ‘mitey’:
```

Document

- 🚀 Quick start guide



Quick Start Guide: Estimating Epidemiological Parameters with `mitey`

Kylie Ainslie

2026-03-02

Source: `vignettes/quick_start_guide.Rmd`

Setup

First, let's load the required packages:

```
library(mitey)
library(dplyr)
library(tidyr)
library(ggplot2)
library(purrr)
library(cowplot)
library(here)
library(outbreaks)
```

Introduction

This vignette demonstrates how to use the `mitey` package to estimate two key epidemiological parameters:

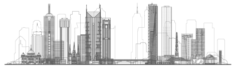
- **Serial interval:** The time between symptom onset in an infector and symptom onset in an infectee
- **Time-varying reproduction number (Rt):** The average number of secondary cases generated by each infected individual at time t

Understanding these parameters is crucial for characterizing disease transmission dynamics, evaluating intervention effectiveness, and designing control strategies. They provide the foundation for mathematical models that inform public health decision-making.

We'll use both simulated and real-world examples for each parameter, with a focus on scabies (a skin infestation caused by the mite Sarcoptes scabiei) for the real-world applications.

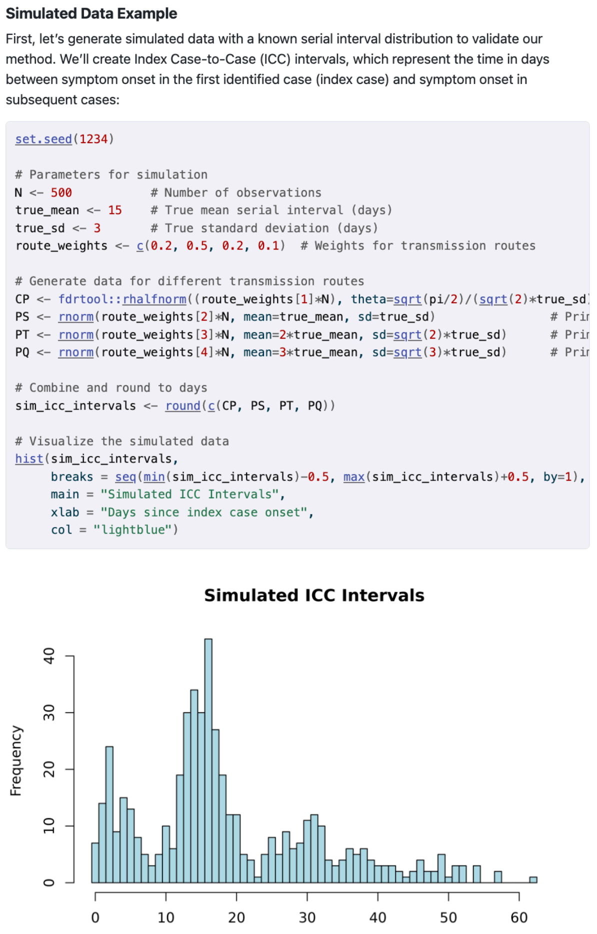
Estimating the Serial Interval

The serial interval distribution is crucial for understanding disease transmission dynamics. The `mitey` package implements the method developed by Vink et al. (2014)[1] to estimate the mean and standard deviation of the serial interval from outbreak data. Accurate serial interval estimates are essential because they directly inform how we interpret the spread of disease over time and are key inputs for estimating the reproduction number.



Document

- 🚀 Quick start guide
- 📊 Worked examples



Simulated Data Example

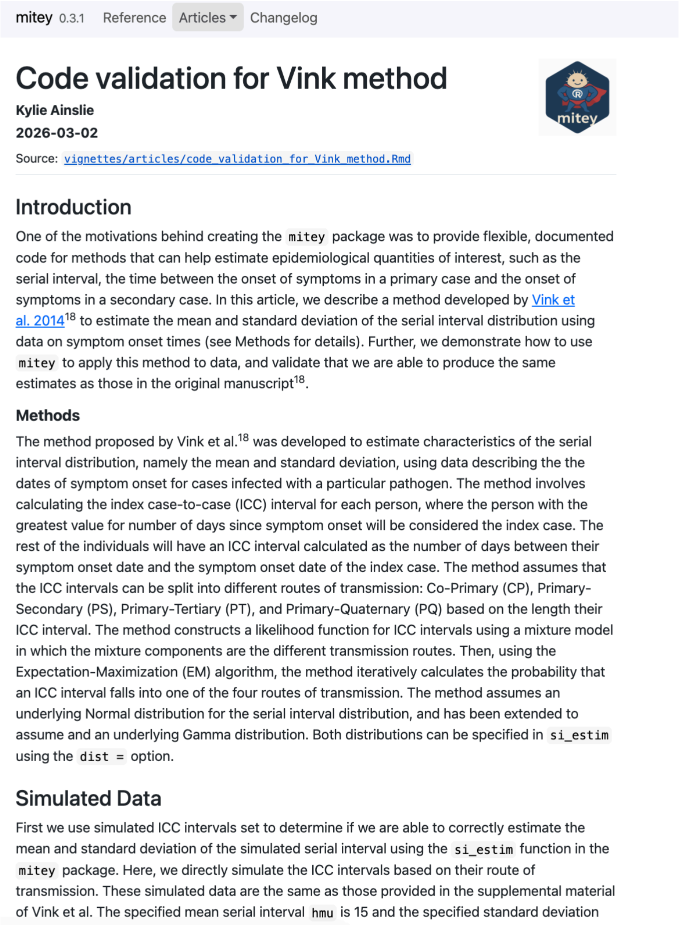
First, let's generate simulated data with a known serial interval distribution to validate our method. We'll create Index Case-to-Case (ICC) intervals, which represent the time in days between symptom onset in the first identified case (index case) and symptom onset in subsequent cases:

```r
set.seed(1234)

# Parameters for simulation
N <- 500              # Number of observations
true_mean <- 15       # True mean serial interval (days)
true_sd <- 3          # True standard deviation (days)
route_weights <- c(0.2, 0.5, 0.2, 0.1)  # Weights for transmission routes

# Generate data for different transmission routes
CP <- fdrtool::rhalfnorm((route_weights[1]*N), theta=sqrt(pi/2)/(sqrt(2)*true_sd
PS <- rnorm(route_weights[2]*N, mean=true_mean, sd=true_sd)                # Pri
PT <- rnorm(route_weights[3]*N, mean=2*true_mean, sd=sqrt(2)*true_sd)      # Pri
PQ <- rnorm(route_weights[4]*N, mean=3*true_mean, sd=sqrt(3)*true_sd)      # Pri

# Combine and round to days
sim_icc_intervals <- round(c(CP, PS, PT, PQ))

# Visualize the simulated data
hist(sim_icc_intervals,
     breaks = seq(min(sim_icc_intervals)-0.5, max(sim_icc_intervals)+0.5, by=1),
     main = "Simulated ICC Intervals",
     xlab = "Days since index case onset",
     col = "lightblue")
```

Simulated ICC Intervals

Document

- 🚀 Quick start guide
- 📊 Worked examples
- ✅ Validation



Code validation for Vink method

Kylie Ainslie

2026-03-02

Source: `vignettes/articles/code_validation_for_Vink_method.Rmd`

Introduction

One of the motivations behind creating the `mitey` package was to provide flexible, documented code for methods that can help estimate epidemiological quantities of interest, such as the serial interval, the time between the onset of symptoms in a primary case and the onset of symptoms in a secondary case. In this article, we describe a method developed by Vink et al. 2014[18] to estimate the mean and standard deviation of the serial interval distribution using data on symptom onset times (see Methods for details). Further, we demonstrate how to use `mitey` to apply this method to data, and validate that we are able to produce the same estimates as those in the original manuscript[18].

Methods

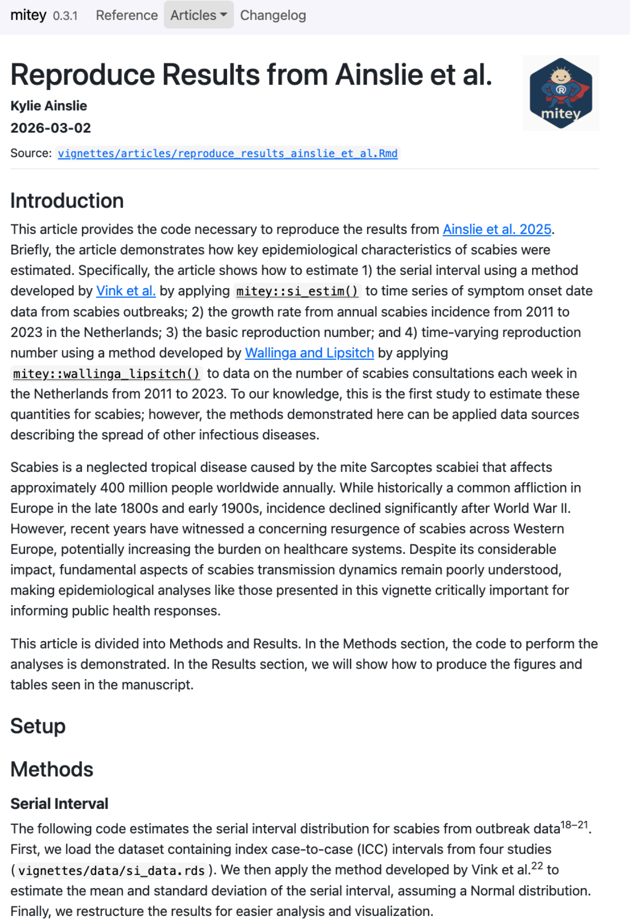
The method proposed by Vink et al.[18] was developed to estimate characteristics of the serial interval distribution, namely the mean and standard deviation, using data describing the the dates of symptom onset for cases infected with a particular pathogen. The method involves calculating the index case-to-case (ICC) interval for each person, where the person with the greatest value for number of days since symptom onset will be considered the index case. The rest of the individuals will have an ICC interval calculated as the number of days between their symptom onset date and the symptom onset date of the index case. The method assumes that the ICC intervals can be split into different routes of transmission: Co-Primary (CP), Primary-Secondary (PS), Primary-Tertiary (PT), and Primary-Quaternary (PQ) based on the length their ICC interval. The method constructs a likelihood function for ICC intervals using a mixture model in which the mixture components are the different transmission routes. Then, using the Expectation-Maximization (EM) algorithm, the method iteratively calculates the probability that an ICC interval falls into one of the four routes of transmission. The method assumes an underlying Normal distribution for the serial interval distribution, and has been extended to assume and an underlying Gamma distribution. Both distributions can be specified in `si_estim` using the `dist =` option.

Simulated Data

First we use simulated ICC intervals set to determine if we are able to correctly estimate the mean and standard deviation of the simulated serial interval using the `si_estim` function in the `mitey` package. Here, we directly simulate the ICC intervals based on their route of transmission. These simulated data are the same as those provided in the supplemental material of Vink et al. The specified mean serial interval `hmu` is 15 and the specified standard deviation

Document

- 🚀 Quick start guide
- 📊 Worked examples
- ✅ Validation
- 📄 Reproduce paper results



Reproduce Results from Ainslie et al.

Kylie Ainslie

2026-03-02

Source: `vignettes/articles/reproduce_results_ainslie_et_al.Rmd`

Introduction

This article provides the code necessary to reproduce the results from Ainslie et al. 2025. Briefly, the article demonstrates how key epidemiological characteristics of scabies were estimated. Specifically, the article shows how to estimate 1) the serial interval using a method developed by Vink et al. by applying `mitey::si_estim()` to time series of symptom onset date data from scabies outbreaks; 2) the growth rate from annual scabies incidence from 2011 to 2023 in the Netherlands; 3) the basic reproduction number; and 4) time-varying reproduction number using a method developed by Wallinga and Lipsitch by applying `mitey::wallinga_lipsitch()` to data on the number of scabies consultations each week in the Netherlands from 2011 to 2023. To our knowledge, this is the first study to estimate these quantities for scabies; however, the methods demonstrated here can be applied data sources describing the spread of other infectious diseases.

Scabies is a neglected tropical disease caused by the mite Sarcoptes scabiei that affects approximately 400 million people worldwide annually. While historically a common affliction in Europe in the late 1800s and early 1900s, incidence declined significantly after World War II. However, recent years have witnessed a concerning resurgence of scabies across Western Europe, potentially increasing the burden on healthcare systems. Despite its considerable impact, fundamental aspects of scabies transmission dynamics remain poorly understood, making epidemiological analyses like those presented in this vignette critically important for informing public health responses.

This article is divided into Methods and Results. In the Methods section, the code to perform the analyses is demonstrated. In the Results section, we will show how to produce the figures and tables seen in the manuscript.

Setup

Methods

Serial Interval

The following code estimates the serial interval distribution for scabies from outbreak data[18–21]. First, we load the dataset containing index case-to-case (ICC) intervals from four studies (`vignettes/data/si_data.rds`). We then apply the method developed by Vink et al.[22] to estimate the mean and standard deviation of the serial interval, assuming a Normal distribution. Finally, we restructure the results for easier analysis and visualization.

One piece of research. Two outputs

nature communications

Article

https://doi.org/10.1038/s41467-025-65544-y

Estimation of the epidemiological characteristics of scabies

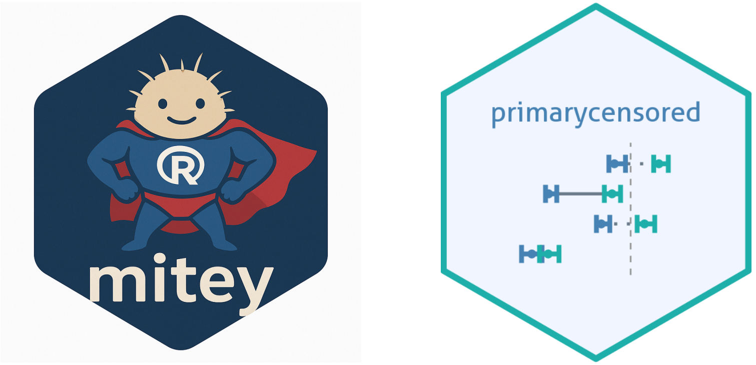
Received: 27 March 2025

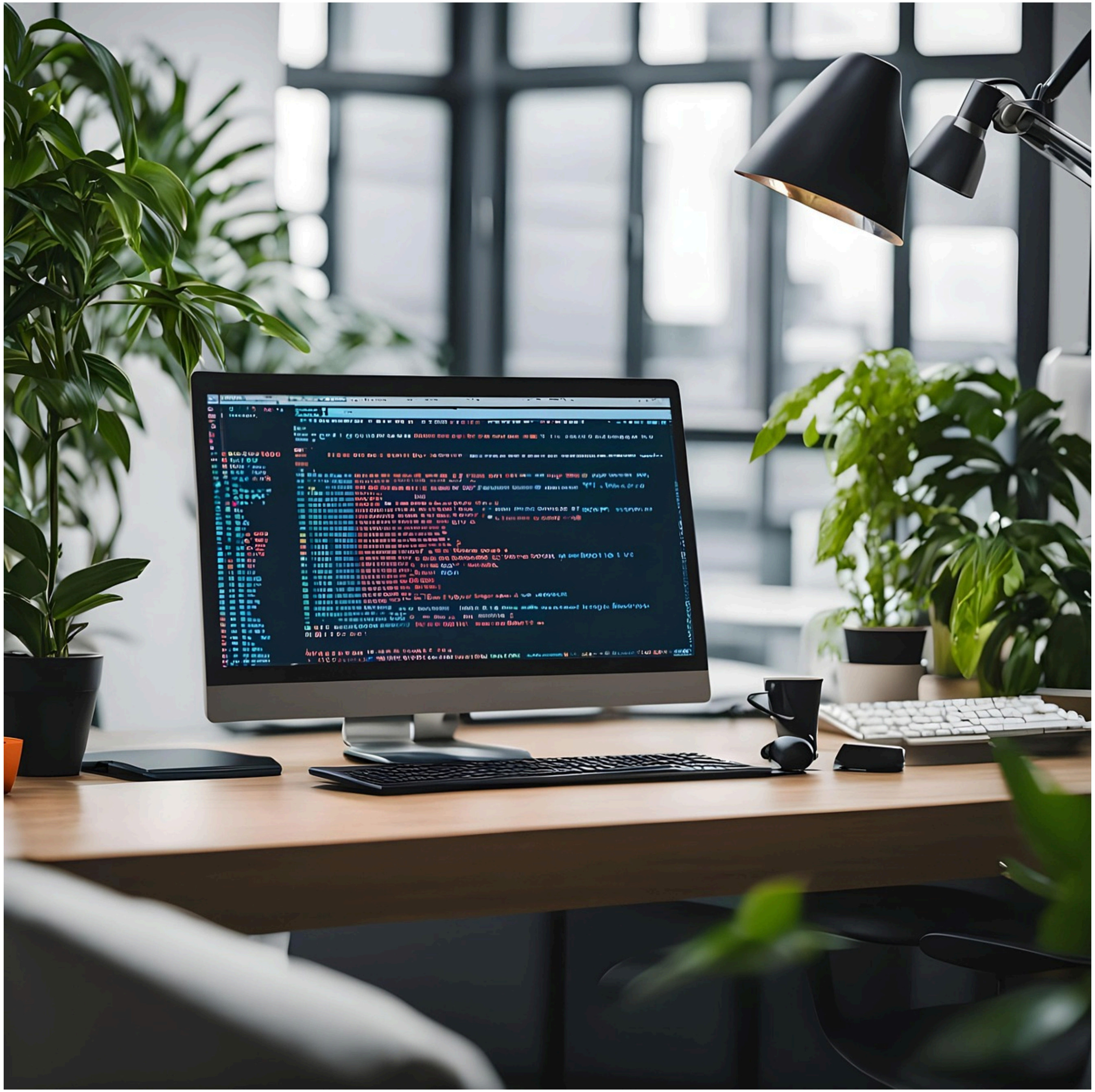
Accepted: 14 October 2025

Kylie E. C. Ainslie [1,2], Mariëtte Hooiveld [3] & Jacco Wallinga [1,4]

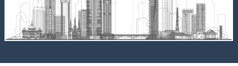
Share → Collaboration

Share → Collaboration

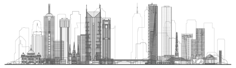




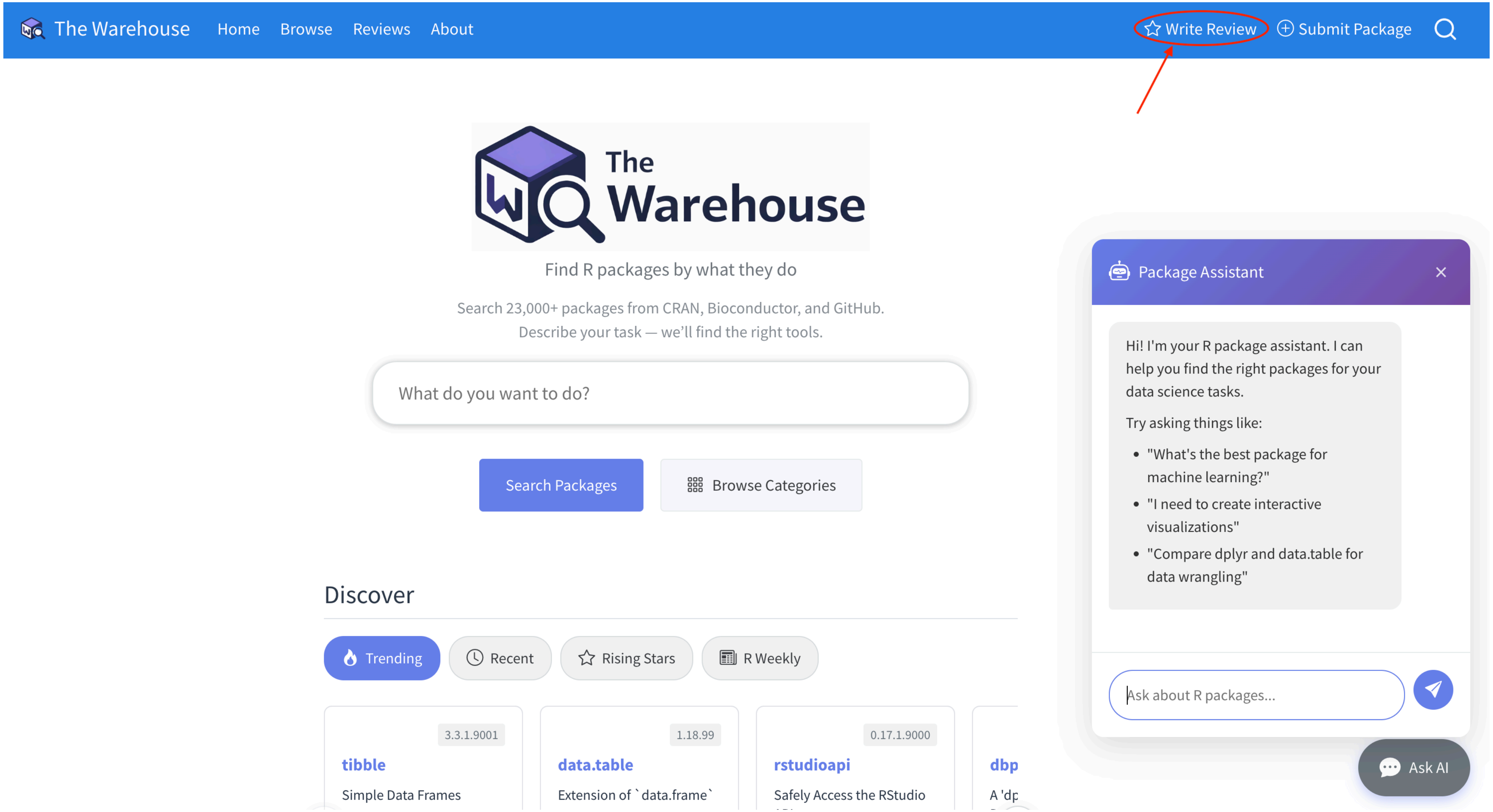
Innovate



?



The Warehouse



The Warehouse

Find R packages by what they do

Search 23,000+ packages from CRAN, Bioconductor, and GitHub.
Describe your task — we'll find the right tools.

What do you want to do?

Discover

3.3.1.9001
tibble
Simple Data Frames

1.18.99
data.table
Extension of `data.frame`

0.17.1.9000
rstudioapi
Safely Access the RStudio

dbp
A 'dp

Package Assistant

Hi! I'm your R package assistant. I can help you find the right packages for your data science tasks.

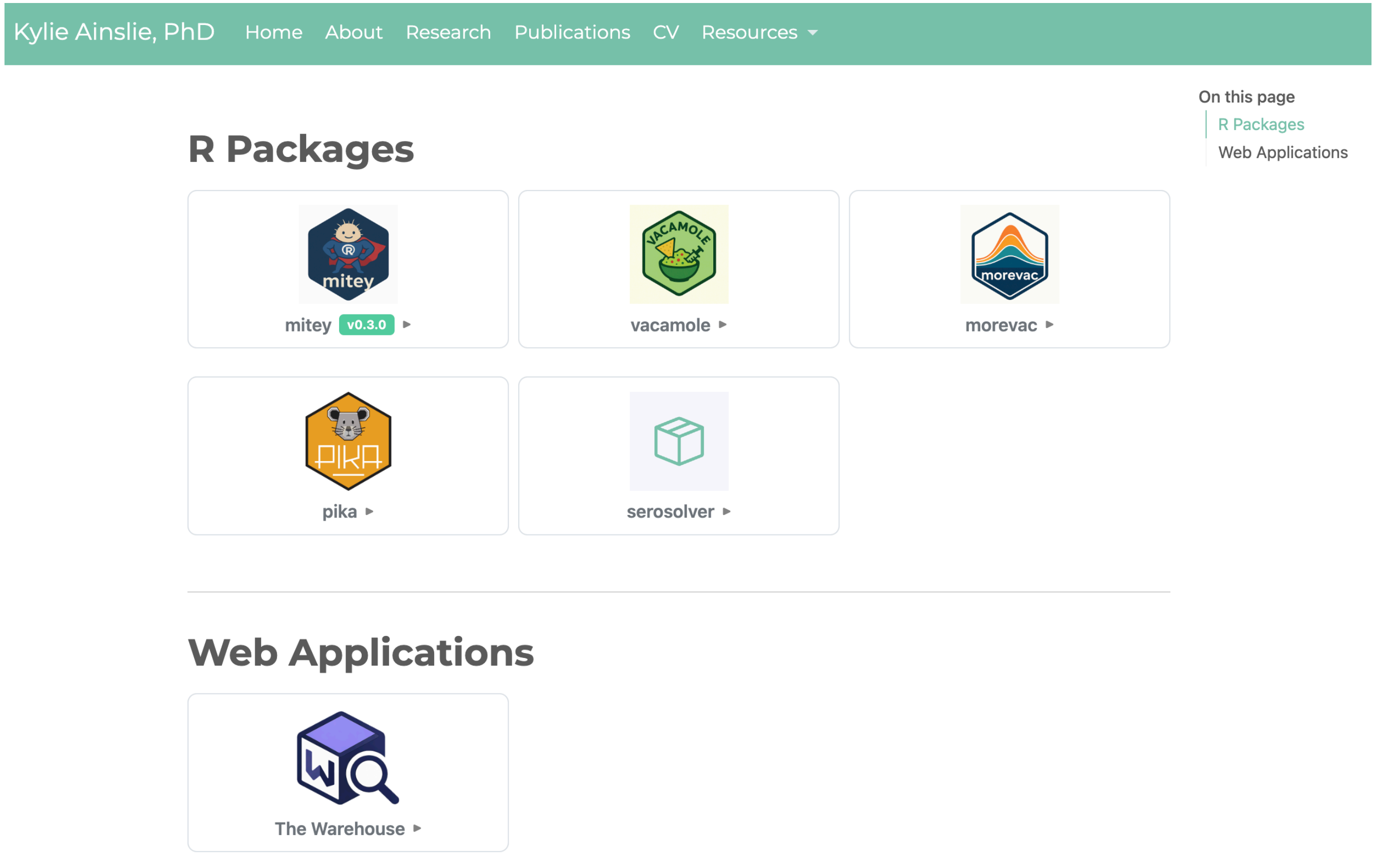
Try asking things like:

- "What's the best package for machine learning?"
- "I need to create interactive visualizations"
- "Compare dplyr and data.table for data wrangling"

Ask about R packages...

The Warehouse

- 💡 Identified a problem shared by the whole community
- 🛠️ Had the confidence to build the solution
- 🤖 AI-powered semantic search
- 🧑‍🤝‍🧑 Built for researchers, by a researcher



Kylie Ainslie, PhD Home About Research Publications CV Resources ▾

On this page
R Packages
Web Applications

R Packages

mitey v0.3.0 ▸

vacamole ▸

morevac ▸

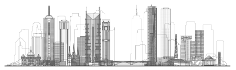
pika ▸

serosolver ▸

Web Applications

The Warehouse ▸

It started with a folder.

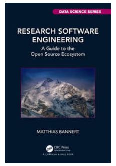


Your turn!

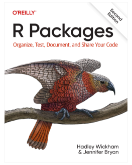
- Try a package structure for your next project
- Tons of tools and learning materials available
- **You'll never go back to messy folders**

The Turing Way

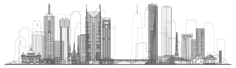
Software Carpentry



Research Software Engineering
Bannert



R Packages
(R users)



Get in touch!

- linkedin.com/in/kylieainslie
- @kylieainslie.bsky.social
- github.com/kylieainslie
- kylieainslie.github.io
- rwarehouse.netlify.app