

---



# useR 2019

**July 9-12, 2019 in Toulouse, France**

Amy Tzu-Yu Chen  
LA West R Users Meeting  
July 31, 2019

---

## What is useR?

- Main international R User Conference
- Usually in Europe, North America, or Oceania
- \$500ish to attend, add another \$100 if you attend tutorials

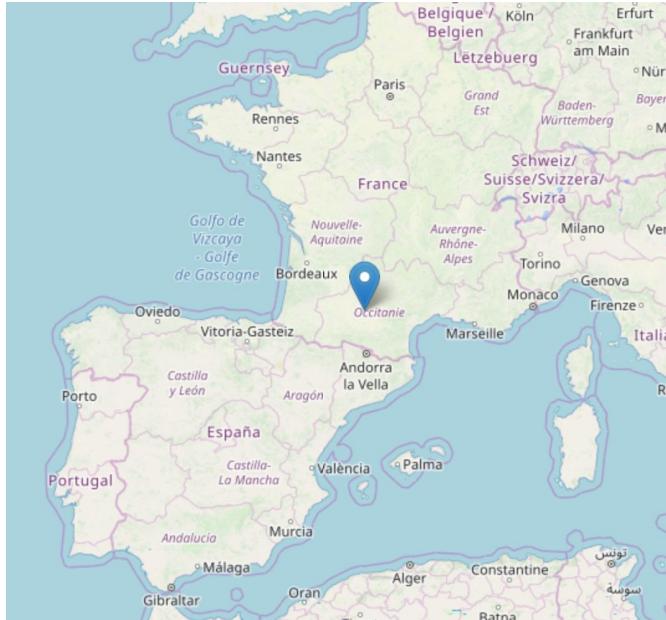
---

# Why I Like useR?

- Meet international influencRs, organizers, and package devs
- Get to know regional R users and how they use R
- useR! @ Toulouse:
  - More academic than American DS/DA conferences
  - Larger attendance
  - Keynotes

---

# Toulouse





# Tutorials

<http://www.user2019.fr/tutorials/>

Rasmus Bååth	⊕ AM1: Get up to speed with Bayesian data analysis in R
Jo-Fai Chow	⊕ AM2: Automatic and Explainable Machine Learning with H2O in R
Di Cook	⊕ AM3: Visualising High-Dimensional Data
Dirk Eddelbuettel	⊕ AM4: Extending R with C++
Colin Fay	⊕ AM5: Hacking RStudio: Advanced Use of your Favorite IDE
Anqi Fu and Balasubramanian Narasimhan	⊕ AM6: CVXR: An R Package for Disciplined Convex Optimization ( <i>joint work with S. Boyd</i> )
Colin Gillespie and Rhian Davies	⊕ AM7: Getting the most out of Git
Jim Hester, Hadley Wickham and Jenny Bryan	⊕ AM8: Package Development
Heather Turner	⊕ AM9: Generalized Nonlinear Models

# Tutorials

<http://www.user2019.fr/tutorials/>

Edzer Pebesma and Roger Bivand	⊕ PM1: Spatial and Spatiotemporal Data Analysis in R
Maria Prokofieva	⊕ PM2: Watch me: introduction to social media analytics
Emma Rand	⊕ PM3: Keeping an exotic pet in your home! Taming Python to live in RStudio because sometimes the best language is both!
Marco Scutari	⊕ PM4: bnlearn: Practical Bayesian Networks in R
Torsten Hothorn	⊕ PM5: Transformation Models
Mark Van Der Loo and Edwin De Jonge	⊕ PM6: Statistical Data Cleaning using R
Davis Vaughan	⊕ PM7: Design For Humans! A Toolkit For Creating Intuitive Modeling Packages <i>(joint work with M. Kuhn)</i>
Tobias Verbeke	⊕ PM8: Docker for Data Science: R, ShinyProxy and more
Achim Zeileis	⊕ PM9: R/exams: A One-For-All Exams Generator

---

# **Tutorial: Rcpp**

## **Dirk Eddelbuettel**

<http://www.user2019.fr/static/pres/tutoam4.pdf>

More than 1500 CRAN packages uses Rcpp (12% of CRAN packages)

Important to people who are interested in package dev or just understand the “backbone” of many pkgs

Check out other Dirk’s projects: <https://github.com/eddelbuettel>

---

# Tutorial: Statistical Data Cleaning

## Mark van der Loo and Edwin de Jonge

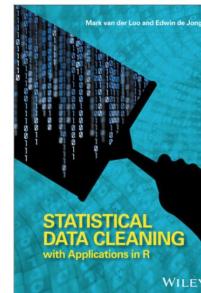
[https://github.com/data-cleaning/useR2019\\_tutorial](https://github.com/data-cleaning/useR2019_tutorial)

A suite of packages for data cleaning and manipulation

Also highlighted good practices for data cleaning

Amy's note: library(simputation) is quite cool

<https://github.com/data-cleaning>



READ AN EXCERPT ▾

Statistical Data Cleaning with Applications in R

Mark van der Loo, Edwin de Jonge

ISBN: 978-1-118-89715-7 | April 2018 | 320 Pages

E-BOOK

Starting at just \$70.99

PRINT

Starting at just \$88.50

O-BOOK

Hardcover

### DESCRIPTION

A comprehensive guide to automated statistical data cleaning



# Keynote

Julia Stewart Lowndes

<http://jules32.github.io>

Teach and organize for Openscapes, Rladies, Carpentries

Having open data science as a mindset; open data science as a shared workflow (tidyverse, github etc)

Team starts with openness: working openly, vertical & horizontal leadership, community

Feel safe and trust each other to do and share data science work

The importance of R community

---

## **Keynote:**

# **Shiny's Holy Grail: Interactivity with Reproducibility**

## **Joe Chang**

<https://github.com/rstudio/shinymeta> (very early stage, open for testing and contributions)

The problem: Shiny brings interactivity but reproducibility is often lost

Reproducible artifacts: code snippet, standalone .Rmd or .R, zip bundle

Makeup of a shiny: domain logic + reactive structure

Uses metaprogramming techniques to write code that serves dual purposes (execute interactively, and export static code)

Related blogpost: <https://towardsdatascience.com/shinymeta-a-revolution-for-reproducibility-bfda6b329f68>



# Keynote

How Bioconductor advances science while contributing to the R language  
and community

Martin Morgan

“Bioconductor provides tools for the analysis and comprehension of high- throughput genomic data”

1741 packages and growing

“Stricter” than CRAN

13 Core Members

<https://www.bioconductor.org>

---

# Keynote: A missing value tour in R

Julie Josse

[http://www.user2019.fr/static/pres/keynote\\_201907101600.pdf](http://www.user2019.fr/static/pres/keynote_201907101600.pdf)

A deep dive into how to handle missing values

“The idea of imputation is both seductive and dangerous”. It is seductive because it can lull the user into the pleasurable state of believing that the data are complete after all, and it is dangerous because it lumps together situations where the problem is sufficiently minor that it can be legitimately handled in this way and situations where standard estimators applied to the imputed data have substantial biases.” (Dempster & Rubin, 1983)

# The Mayor Welcomed Us!



Romain François 🇫🇷  
@romain\_francois

跟随



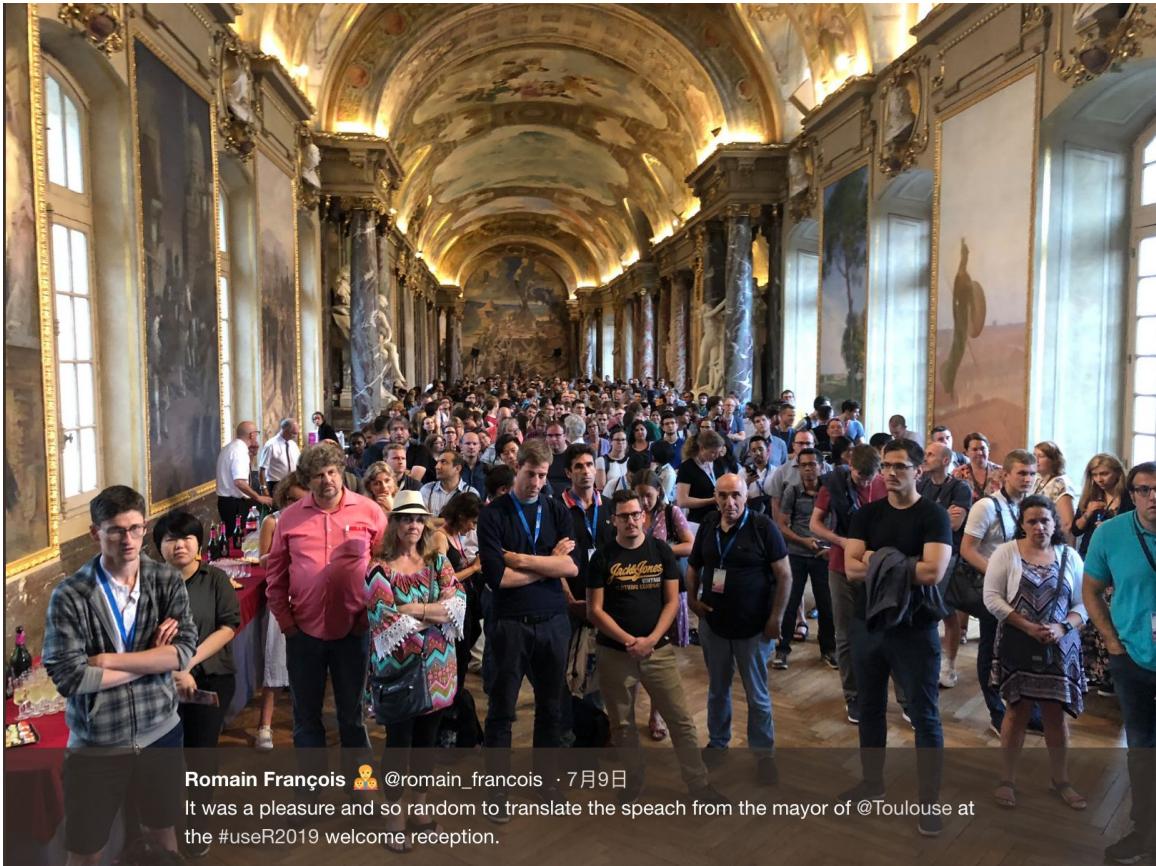
It was a pleasure and so random to translate the speech from the mayor of @Toulouse at the #useR2019 welcome reception.



上午10:12 - 2019年7月9日



[https://twitter.com/romain\\_francois/status/1148641180332318721?s=21](https://twitter.com/romain_francois/status/1148641180332318721?s=21)



Romain François 🇫🇷 @romain\_francois · 7月9日

It was a pleasure and so random to translate the speech from the mayor of @Toulouse at the #useR2019 welcome reception.

[https://twitter.com/romain\\_francois/status/1148641180332318721?s=21](https://twitter.com/romain_francois/status/1148641180332318721?s=21)

# Space Center

---



# Space Center

---



# Space Center

---



# Space Center

---





# Recommended Talks

If you are a tidyverse useR, you will probably be interested in these two talks

Data handling	Hadley Wickham	Enhancements to data tidying
Data handling	Romain Francois	n() cool #dplyr things

---

# You don't need Spark for this - larger-than-RAM data manipulation with `disk.frame`

Zhuo Jia Dai

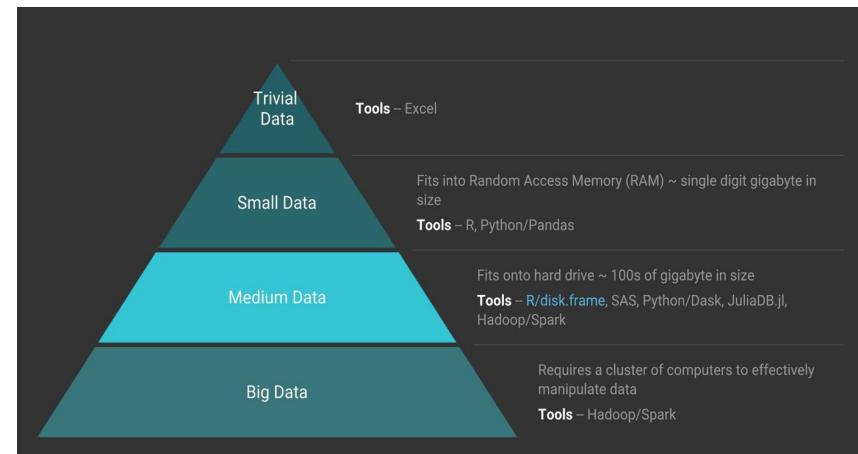
<https://github.com/xiaodaigh/disk.frame>

Interesting and has a lot of potential if you want to test and contribute

“`disk.frame`” format: divide data into chunks and each chunk is a `fst` file

Built on `data.table`, `future`, `fsp`, and `dplyr` for syntax

Current status: `Group_by` and `merge` works; working on `sort`





# Movement & Transport

Movement & transport	Rocio Joo	Navigating through the R packages for movement
Movement & transport	Mohammad Mehdi Moradi	Classes, methods and data analysis for trajectories
Movement & transport	Christine Thomas-Agnan	Modelling spatial flows with R
Movement & transport	Robin Lovelace	R for Transport Planning

#rspatial  
#geocompr

---

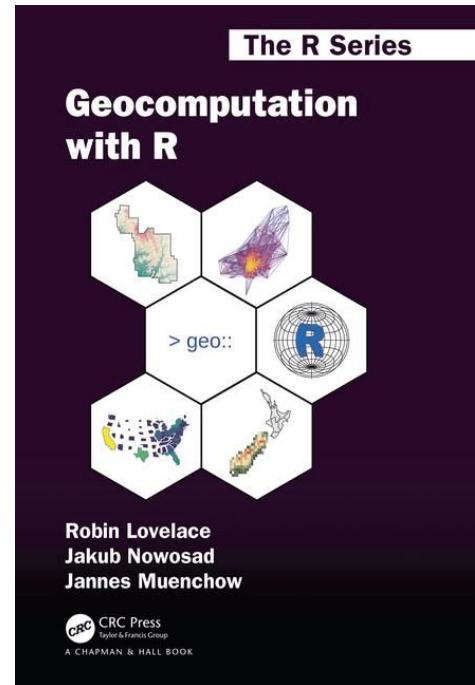
# R for Transport Planning

## Robin Lovelace

<https://www.robinlovelace.net/presentations/user2019-r-for-transport-planning.html#1>

Make transport planning “data-driven, modular, and open”

- library(stplanr)
  - ROpenSci project <https://github.com/ropensci/stplanr>
  - <https://journal.r-project.org/archive/2018/RJ-2018-053/index.html>



<https://geocompr.robinlovelace.net>

---

# Lightning: {polite} - web etiquette for R users

Dmytro Perepolkin

<http://www.user2019.fr/static/pres/l257430.pdf>

<https://github.com/dmi3kno/polite>

Package for managing your web scraping manners!

- **bow()**
  - Establishes “polite” session, checks *robots.txt*
  - For “quick bow”, when changing path (e.g. inside loop), use `nod()`
- **scrape()**
  - Wrapper over `httr::GET`, rate-limited and `memoise'd`
  - For polite batch-downloading use `rip()`



---



# **Lightning: Die Nutella oder Das Nutella? Grammatical Gender Prediction of German Nouns**

## **Chung-hong Chan**

<http://www.user2019.fr/static/pres/lit242481.pdf>

Interesting linguistic/NLP project

LSTM model on 329,876 German nouns and their respective gender

High accuracy but does not work on some common *short* words (i.e. the model can predict nouns like *der Terrorismus / Terrorism* but failed on *die Messer / Knife*)

DeepDream: generate random German nouns

# Lightning: Machine Learning with R: do it with a framework

## Eric Lecoutre

- Why use a framework? it's a full process and we need consistency
- Functionalities of a ML framework:
  - Preprocessing (missing values, scaling, outliers)
  - Feature selection and extraction (feature engineering)
- Packages
  - caret+caretEnsemble
    - more stackoverflow
  - mlr
    - use future for parallelization
    - recommendation: use when you want to do more ML
    - mlr3 (a rewrite of mlr)
- Notes
  - Frameworks don't always work with xgboost.
  - Difference between ML framework and tidy modelling pkg like library(broom)

	data preparation	modeling	ensemble/stacking
caret	*	*	
caretEnsemble			*
mlr		*	*
mlrCPO	*		
mlr3		*	*
recipes	*		
SuperLearner		*	*
slr3		*	*

---

# **Poster: The Evolution of R Programming Evolution**

**Chung-hong Chan & Chia-Yi Yen**

<https://github.com/chainsawriot/rstyle>

Analyzed CRAN packages over time

Interesting to see how people's code style changed as we gradually agreed on what's the good coding style

# Performance Talks

Performance	Gábor Csárdi	pak: a fresh approach to package installation	<a href="https://github.com/r-lib/pak">https://github.com/r-lib/pak</a>
Performance	Arun Srinivasan	Summary of developments in R's data.table package	<a href="http://www.user2019.fr/static/pres/t258038.pdf">http://www.user2019.fr/static/pres/t258038.pdf</a> <a href="https://github.com/Rdatatable/data.table">https://github.com/Rdatatable/data.table</a>
Performance	Jim Hester	Real-time file import with the vroom package	<a href="http://www.user2019.fr/static/pres/t258038.pdf">http://www.user2019.fr/static/pres/t258038.pdf</a> (very cool slides) <a href="https://github.com/r-lib/vroom">https://github.com/r-lib/vroom</a>
Performance	Henrik Bengtsson	A Future for R: Simplified Parallel and Distributed Processing	<a href="https://www.jottr.org/presentations/useR2019/BengtssonH_201..pdf">https://www.jottr.org/presentations/useR2019/BengtssonH_201..pdf</a> <a href="https://github.com/HenrikBengtsson/future">https://github.com/HenrikBengtsson/future</a>

# Future: now with Progress Bar

Henrik Bengtsson

New development since SatRday LA in April

[https://www.jottr.org/presentations/useR2019/BengtssonH\\_201...pdf](https://www.jottr.org/presentations/useR2019/BengtssonH_201...pdf)

<https://github.com/HenrikBengtsson/future>

User chooses how progress is presented

```
# without progress updates  
x <- 1:10  
y <- slow_sum(x)
```

```
handlers("beepr")  
with_progress(y <- slow_sum(x))  
♪ ♪ ♪ ♪ ... ♪
```

```
handlers("txtprogressbar")  
with_progress(y <- slow_sum(x))  
|=====| 40%
```

```
handlers("progress", "beepr")  
with_progress(y <- slow_sum(x))  
[=====-----] 40% Add 4  
♪ ♪ ♪ ♪ ... ♪
```

```
handlers("progress")  
with_progress(y <- slow_sum(x))  
[=====-----] 40% Add 4
```

# Community Talks

R4DS



AfricaR

R-Ladies

SatRdays

- [R4DS.online](#)
- Jesse Mostipak ([@kierisi](#)) founded as *R for Data Science* study group
- Twitter: [@R4DSCommunity](#)
- Slack: [bit.ly/R4DSSlack](https://bit.ly/R4DSSlack)



**R-Ladies** Founded in 2012

Members

**46,784**

Groups

**163**

Countries

**46**

Ask Amy or visit  
<https://satrdays.org/>

---

# Thank you!

Check <http://www.user2019.fr> for full schedule and talk slides. They shall upload the videos soon!

See you at useR! 2020 in St. Louis, July 7 ~10, 2020 ?



<https://twitter.com/user2020stl>