Conference Report: *Why R?* 2018

by Michał Burdukiewicz, Marta Karas, Leon Eyrich Jessen, Marcin Kosiński, Bernd Bischl and Stefan Rödiger

Figure 1: *Why R?* 2018 conference banner used for social media promotion. The background displays banks of the Odra river in Wroclaw – the city of Poland where the conference was held.

Why R? 2018 conference

The primary purpose of the *Why R?* 2018 conference was to provide R programming language enthusiasts with an opportunity to meet and discuss experiences in R software development and analysis applications, for both academia and industry professionals. The event was held 2-5 August, 2018 in a city of Wroclaw, a strong academic and business center of Poland. The total of approximately 250 people from 6 countries attended the main conference event. Additionally, approximately 540 R users attended the pre-meetings in eleven cities across Europe (Figure 2).

Why R? 2018 conference is the continuation of the *Why R?*'s first edition that took place Sep 27-29, 2017 at the Warsaw University of Technology in Warsaw (Poland). Given the success of the first event, this year's conference extended its program concept and scope; importantly, *Why R?* 2018 conference was held as international.

Conference program

The format of the conference was aimed at exposing participants to R language recent developments as well as a wide range of application examples. It consisted of workshops, invited talks, field-specific series of talks, lighting-talks, special interest groups and a full-day programming hackathon.

The conference program had a strong focus on machine learning techniques and applications, with **mlr** (Bischl et al., 2016) R package – an interface to a large number of classification and regression methods – being emphasized in a number of presentations, as well as employed during workshops and the hackathon provided by the **mlr** team. The scope of conference program included statistical methodology, data visualization, R code performance, building products based on data analyses and R role in academia / industry.

The event offered extensive networking opportunities. The cocktail party was held at the conference venue on the 2nd conference day. In addition, convenient location in the close proximity of the old town market square facilitated many informal gatherings that were happening each conference day.

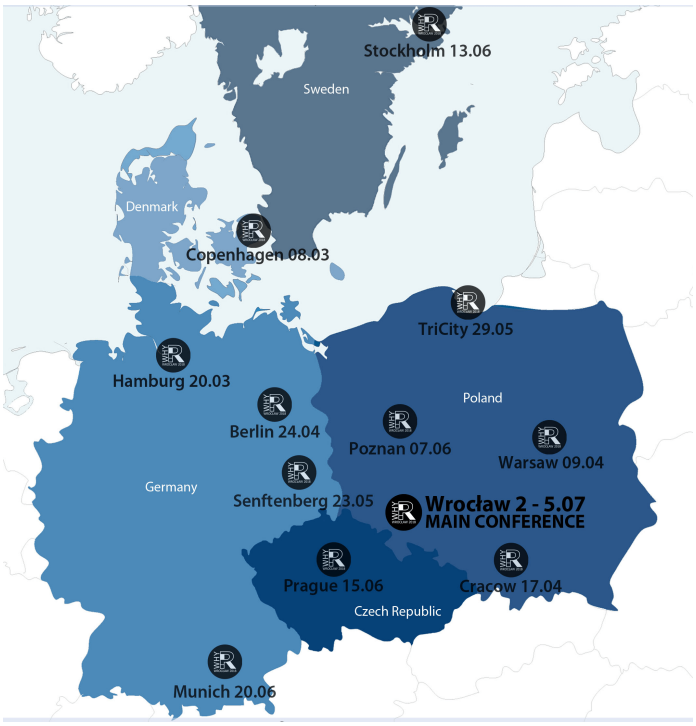

Figure 2: Locations and dates of the *Why R?* 2018 main conference event and 11 *Why R?*-branded pre-meetings.

Why R? Pre-meetings

The novel idea of pre-meetings has proved to be successful in popularizing *Why R?* conference in the international community of R users. Eleven pre-meetings took place in Czech Republic, Denmark, Germany, Poland and Sweden in the run-up to the *Why R?* main event. The pre-meetings either constituted a part of another conference, one day-long workshop and discussion event or a meeting of a local R user group.

As R provides a versatile framework for reproducible research in different scientific domains (Gentleman and Temple Lang, 2007; Gandrud, 2013; Leeper, 2014; Liu and Pounds, 2014; Rödiger et al., 2015), we considered the *Why R?* pre-meetings as a great opportunity to convey and popularize R as an analytics tool in groups of professionals from different fields. The pre-meeting held at International Biotechnology Innovation Days (*IBID*), an open-access conference held 23-25 May, 2018 at the Brandenburg University of Technology Cottbus - Senftenberg (Senftenberg, Germany)[1] is an example where the R came in close contact with scientist from other domains. *IBID* brought together specialists and experts in the fields of bioanalytics, biomedical and translational research, autoimmune diagnostics, digitalization and engineering; hence it posed an excellent platform to promote R and *Why R?* 2018 conference.

Workshops

Why R? 2018 conference had a wide portfolio of workshops:

- **Maps in R** by Piotr Sobczyk (OLX Group). Piotr showed how to create spatial data visualization efficiently in the R. He gave a plenty of tips to follow, pitfalls to avoid and a number of useful hacks. Starting from a basic plot function, he covered the usage of **ggplot2** as well as R packages that use interactive javascript libraries to prepare data reports.

[1] http://web.archive.org/web/20180701084524/https://ibid-2018.b2match.io/

- **iDash - Make your R slides awesome with xaringan** by Mikołaj Olszewski (iDash) and Mikołaj Bogucki (iDash). The workshop introduced the **xaringan** (Xie et al., 2018) package – an alternative approach to preparing a slide deck. **xaringan** allows customizing each slide entirely and previewing slides dynamically in RStudio; moreover, the export of the slide deck (natively in HTML) to a pixel-perfect PDF is fairly easy. As **xaringan** also uses RMarkdown, it allows for reproducible results.

- **Jumping Rivers - Shiny Basics** and **Advanced Shiny** by Roman Popat (Jumping Rivers). The instructor Roman Popat from Jumping Rivers conducted two workshops. In first (Shiny Basics), he gave an introduction to creating interactive visualizations of data using Shiny. Here, participants learned how to use **rmarkdown** and **htmlwidgets**, input and output bindings to interact with R data structures, input widgets and render functions to create complete page layouts using shiny and shiny dashboard. The advanced Shiny workshop explored how to add functionality to shiny apps using javascript packages and code. In particular, it was showed how one might deal with routines in a Shiny application that take a long time to run and how to provide a good experience for simultaneous users of an app. Finally, the instructor showed how to create a standalone web server API to the R code and how to integrate the use of it into a Shiny application using the **plumber** (Technology et al., 2018) package.

- **DALEX - Descriptive mAchine Learning EXplanations** by Mateusz Staniak(Uniwersytet Wrocławski). THe workshop covered tools for exploration, validation and explanation of complex machine learning models. The packages explored in this workshop include **mlr** (Bischl et al., 2016), **DALEX** (Biecek, 2018), **live** (Staniak and Biecek, 2018), **FactorMerger** (Sitko and Biecek, 2017), **archivist** (Biecek and Kosinski, 2017), **pdp** (Greenwell, 2017) and **ALEPlot** (Apley, 2018).

- **Constructing scales from survey questions** by Tomasz Żółtak (Educational Research Institute in Warsaw, Poland). Tomasz showed how to create scales based on sets of categorical variables using Categorical Exploratory/Confirmatory Factor Analysis (CEFA / CCFA) and IRT models. He used models with bi-factor rotation to deal with different forms of asking questions and corrected for differences in a style of answering questions asked using a Likert scale. In addition, it was showed how to correct self-assessment knowledge/skill indicators using fake items.

- **From RS data to knowledge – Remote Sensing in R** by Bartłomiej Kraszewski (Forest Research Institute in XXX). Remote sensing data from different sensors is a rich source of information for studying the natural environment, natural phenomena and monitoring some extreme phenomena, such as floods. Bartłomiej presented R language packages that can be used to work with remote sensing data. These included (a) for geographic information system analysis: **rgdal** (Bivand et al., 2018a), **rgeos** (Bivand et al., 2018b) and **sf** (Pebesma et al., 2018); (b) for raster data processing: **raster** (Hijmans et al., 2017); (c) for Airborne LaserScanning data processing: the **lidR** (Roussel et al., 2018) package.

- **Introduction to Deep Learning with Keras in R** by Michał Maj (Appsilon Data Science). The workshop covered many important aspects of Deep Learning with the Keras in R, including sequential model building, performing data ingestion and using pre-trained models and performing fine-tuning. The **keras** (Allaire et al., 2018) R package was explored.

Invited talks

The invited talks topics included domain knowledge from statistics, computer science, natural sciences and economics. The speakers list presents as follows:

- Tomasz Niedzielski (University of Wroclaw): *Forecasting streamflow using the HydroProg system developed in R*,

- Daria Szmurło (McKinsey & Company): *The age of automation – What does it mean for data scientists?*,
- Agnieszka Suchwałko (Wroclaw University of Technology): *Project evolution – from university to commerce*,
- Bernd Bischl (Ludwig-Maximilians-University of Munich): *Machine learning in R*,
- Artur Suchwałko (QuantUp): *A business view on predictive modeling: goals, assumptions, implementation*,
- Maciej Eder (Institute of Polish Language): *New advances in text mining: exploring word embeddings*,
- Thomas Petzoldt (Dresden University of Technology): *Simulation of dynamic models in R*,
- Leon Eyrich Jessen (Technical University of Denmark): *Deep Learning with R using TensorFlow*.

Special Interest Groups

Three Special Interest Groups were organized to facilitate topic-specific discussion between conference participants.

- **Diversity in Data Science**, moderated by R-Ladies Warsaw, aimed to discuss boosting the diversity of R community and inspire members of affinity groups to pursue careers in data science.
- **The Career planning in data science**, moderated by Artur Suchwałko (QuantUp) and Marcin Kosiński (Why R? Foundation), gave participants a chance to learn from experienced R enthusiasts about their career paths.
- **Teaching of data science**, moderated by Leon Eyrich Jessen (Technical University of Denmark) and Stefan Rödiger (Brandenburg Technical University Cottbus-Senftenberg), gathered data science experts from academia an industry to share their experiences and discuss challenges and solutions in teaching different concepts of data science.

Conference organizers

The quality of the scientific program of the conference was the achievement of Marcin Kosiński, Alicja Gosiewska, Aleksandra Grudziąż, Malte Grosser, Andrej-Nikolai Spiess, Przemysław Gagat, Joanna Szyda, Paweł Mackiewicz, Bartosz Sękiewicz, Przemysław Biecek, Piotr Sobczyk, Marta Karaś, Marcin Krzystanek, Marcin Łukaszewicz, Agnieszka Borsuk - De Moor, Jarosław Chilimoniuk, Michał Maj and Michał Kurtys. The organization was in the hands of Michał Burdukiewicz (chair).

The organizers want to acknowledge R user groups from Berlin, Copenhagen, Cracow, Hamburg, Munich, Poznan, Prague, Stockholm, TriCity, Wroclaw and Warsaw.

Acknowledgements

We would like to say thank you to all the sponsors, the University of Wrocław, Wrocław Center of Biotechnology Consortium, the local organizers of the pre-meetings, the **mlr** team and student helpers.

Additional information

***Why R?* 2018 website** http://whyr.pl/2018
Corporate sponsors: McKinsey & Company, Wrocław Center for Biotechnology, KRUK S.A., iDash s.c., R Consortium, WLOG Solutions, Jumping Rivers Ltd., RStudio, Inc., Analyx®GmbH and Pearson IOKI.

Bibliography

J. J. Allaire, F. Chollet, RStudio, Google, Y. Tang, D. Falbel, W. V. D. Bijl, and M. Studer. keras: R Interface to 'Keras', Apr. 2018. URL https://CRAN.R-project.org/package=keras. [p3]

D. Apley. ALEPlot: Accumulated Local Effects (ALE) Plots and Partial Dependence (PD) Plots, May 2018. URL https://CRAN.R-project.org/package=ALEPlot. [p3]

P. Biecek. DALEX: Descriptive mAchine Learning EXplanations, June 2018. URL https://CRAN.R-project.org/package=DALEX. [p3]

P. Biecek and M. Kosinski. archivist: An R package for managing, recording and restoring data analysis results. *Journal of Statistical Software*, 82(11):1–28, 2017. doi: 10.18637/jss.v082.i11. [p3]

B. Bischl, M. Lang, L. Kotthoff, J. Schiffner, J. Richter, E. Studerus, G. Casalicchio, and Z. M. Jones. mlr: Machine learning in r. *Journal of Machine Learning Research*, 17(170):1–5, 2016. URL http://jmlr.org/papers/v17/15-066.html. [p1, 3]

R. Bivand, T. Keitt, B. Rowlingson, E. Pebesma, M. Sumner, R. Hijmans, E. Rouault, F. Warmerdam, J. Ooms, and C. Rundel. rgdal: Bindings for the 'Geospatial' Data Abstraction Library, June 2018a. URL https://CRAN.R-project.org/package=rgdal. [p3]

R. Bivand, C. Rundel, E. Pebesma, R. Stuetz, K. O. Hufthammer, P. Giraudoux, M. Davis, and S. Santilli. rgeos: Interface to Geometry Engine - Open Source ('GEOS'), June 2018b. URL https://CRAN.R-project.org/package=rgeos. [p3]

C. Gandrud. *Reproducible Research with R and RStudio*. Chapman and Hall/CRC, July 2013. ISBN 978-1-4665-7284-3. [p2]

R. Gentleman and D. Temple Lang. Statistical Analyses and Reproducible Research. *Journal of Computational and Graphical Statistics*, 16(1):1–23, Mar. 2007. ISSN 1061-8600, 1537-2715. doi: 10.1198/106186007X178663. URL http://www.tandfonline.com/doi/abs/10.1198/106186007X178663. [p2]

B. M. Greenwell. pdp: An r package for constructing partial dependence plots. *The R Journal*, 9(1):421–436, 2017. URL https://journal.r-project.org/archive/2017/RJ-2017-016/index.html. [p3]

R. J. Hijmans, J. van Etten, J. Cheng, M. Mattiuzzi, M. Sumner, J. A. Greenberg, O. P. Lamigueiro, A. Bevan, E. B. Racine, A. Shortridge, and A. Ghosh. raster: Geographic Data Analysis and Modeling, Nov. 2017. URL https://CRAN.R-project.org/package=raster. [p3]

T. J. Leeper. Archiving Reproducible Research with R and Dataverse. *The R Journal*, 6(1):151–158, June 2014. URL http://journal.r-project.org/archive/2014-1/leeper.pdf. [p2]

Z. Liu and S. Pounds. An R package that automatically collects and archives details for reproducible computing. *BMC Bioinformatics*, 15(1):138, May 2014. ISSN 1471-2105. doi: 10.1186/1471-2105-15-138. URL http://www.biomedcentral.com/1471-2105/15/138/abstract. [p2]

E. Pebesma, R. Bivand, E. Racine, M. Sumner, I. Cook, T. Keitt, R. Lovelace, H. Wickham, J. Ooms, and K. Müller. sf: Simple Features for R, May 2018. URL https://CRAN.R-project.org/package=sf. [p3]

S. Rödiger, M. Burdukiewicz, K. A. Blagodatskikh, and P. Schierack. R as an Environment for the Reproducible Analysis of DNA Amplification Experiments. *The R Journal*, 7(2): 127–150, 2015. URL http://journal.r-project.org/archive/2015-1/RJ-2015-1.pdf. [p2]

J.-R. Roussel, D. A. R. the documentation), F. D. B. F. a. bugs improved catalog features), and A. S. M. I. lassnags). lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications, June 2018. URL https://CRAN.R-project.org/package=lidR. [p3]

A. Sitko and P. Biecek. *The Merging Path Plot: adaptive fusing of k-groups with likelihood-based model selection*, 2017. URL https://arxiv.org/abs/1709.04412. [p3]

M. Staniak and P. Biecek. Explanations of model predictions with live and breakDown packages. *ArXiv e-prints*, Apr 2018. URL https://arxiv.org/abs/1804.01955. [p3]

T. Technology, LLC, J. Allen, F. van Dunné, S. Vandewoude, and S. Software (swagger-ui). plumber: An API Generator for R, June 2018. URL https://CRAN.R-project.org/package=plumber. [p3]

Y. Xie, C. T. Ekstrøm, D. Lang, G. Aden-Buie, O. P. B. C. in rmarkdown/templates/xaringan/resources/default.css), P. Schratz, and S. Lopp. xaringan: Presentation Ninja, Feb. 2018. URL https://CRAN.R-project.org/package=xaringan. [p3]

Michał Burdukiewicz
Warsaw University of Technology, Why R? Foundation
Pl. Politechniki 1, 00-661 Warsaw
Poland
michalburdukiewicz@gmail.com

Marta Karas
Johns Hopkins Bloomberg School of Public Health
615 N Wolfe St Rm E3527
Baltimore, MD
USA
marta.karass@gmail.com

Leon Eyrich Jessen
Technical University of Denmark
Anker Engelunds Vej 1, 2800 Kgs. Lyngby, Denmark
Denmark
author1@work

Marcin Kosiński
Gradient Metrics, Why R? Foundation
Warsaw
Poland
m.p.kosinski@gmail.com

Bernd Bischl
Ludwig Maximilian University of Munich
Ludwigstraße 33, München

Germany
bernd_bischl@gmx.net

Stefan Rödiger
Brandenburg University of Technology Cottbus–Senftenberg
Universitätsplatz 1, Senftenberg
Germany
stefan.roediger@b-tu.de