

Octombrie 2018 volum 5
numărul 9 (52)



octombrie 2018
Colectiv redacțional:
Nicoleta CARAGEA
Ana Maria CIUHU
Andreea IORDAN
Bogdan PĂTĂRLĂGEANU
Bogdan DRĂGHIA
comunicare@insse.ro

<https://www.aanmelder.nl/uoros2018>
<https://www.uorosconf.org>
Twitter: @urosconf

ISSN: 2537 - 3129

The Use of R in Official Statistics - uRos2018



Conferința Internațională **The Use of R in Official Statistics - uRos2018** a avut loc în perioada 10-14 septembrie 2018 și este a 6-a conferință inițiată de către INS - România, fiind organizată anul acesta la Haga, împreună cu CBS (Centraal Bureau voor de Statistiek), Olanda. Partenerii conferinței au fost Oficiul de Statistică din Austria, Universitatea București și Universitatea Ecologică din București. Seria de evenimente uRos promovează utilizarea și dezvoltarea software-ului de analiză R în statistica oficială, dar și a altor medii care integrează pachete și funcții ale R. De-a lungul celor șase ediții ale evenimentului participarea internațională a crescut continuu, de la 6 țări în anul 2014, la 32 de țări anul acesta.

The International Conference **The Use of R in Official Statistics - uRos2018** (September 10-14, 2018) was the sixth of a series of events, initiated by Romanian NIS. This edition was organized in Netherlands, the Hague, together with CBS (Centraal Bureau Voor De Statistiek). The partners of the conference were *Statistics Austria, University of Bucharest and Ecological University of Bucharest*. uRos series is promoting the use and development in official statistics of R and other environments that integrate the R packages and their functionalities. Throughout the six editions of the event, the international participation has increased exponentially, from 6 countries in 2014 to 32 countries in 2018.

UTILIZAREA SOFT-ULUI „R” ÎN STATISTICA OFICIALĂ

Conferința a reunit 100 de statisticieni și specialiști din oficiile de statistică și din mediul academic din 32 de țări (Africa de Sud, Austria, Belgia, Canada, Coreea, Danemarca, Elveția, Estonia, Finlanda, Franța, Germania, Guatemala, Islanda, India, Irlanda, Italia, Japonia, Letonia, Lituania, Luxemburg, Marea Britanie, Noua Zeelandă, Olanda, Polonia, Portugalia, România, Slovenia, Spania, Suedia, SUA, Tanzania, Ungaria) de pe 6 continente. De asemenea, au fost prezenți reprezentanți ai Eurostat și ai altor organizații internaționale care au ca obiect de activitate statisticile oficiale (OECD, Banca Mondială, Biroul de Statistică a Muncii - SUA).

Lucrările conferinței s-au desfășurat pe secțiuni paralele și au cuprins 60 de prezentări și 5 tutoriale tematice. Agenda a cuprins și o secțiune de proiecte de inovare pe teme din diverse domenii ale statisticii oficiale, denumită generic *unConf*, precum și cursuri tematice.

În deschiderea conferinței, Wim van Nunspeet (Directorul Departamentului



de IT și metodologie) de la CBS a subliniat importanța evenimentului pentru dezvoltarea statisticilor oficiale și a colaborării internaționale la cel mai înalt nivel științific. De asemenea, conferința a avut ca invitați speciali/*keynote speakers* doi dintre cei mai cunoscuți experți din comunitatea R. Primul invitat special care a deschis conferința, Alina Matei, autoarea pachetului *sampling*, profesor

la Universitatea Neuchâtel din Elveția a prezentat "*Sample and coordination*", expunând o metodă de coordonare a eșantioanelor pentru cercetări statistice. Al doilea *keynote speaker* a fost Jerome Ooms (unul dintre autorii implementării R pentru Windows). Prezentarea lui a cuprins informații privitoare la infrastructura sistemului care stă la baza R și la librăriile externe utilizate.



Proiectele elaborate de cele 4 echipe internaționale la maratonul *unConf* au fost următoarele:

1) **R wrapper for PDOK location server:** <https://github.com/uRos2018/nlgeocoder>

Echipa: Willy Tadema (Provincie Groningen, Olanda), Egge-Jan Pollé (Tensing GIS Consultancy, Olanda), Edwin de Jonge (CBS, Olanda), Juris Breidaks (Central Statistical Bureau, Letonia);

2) **Categorical: a better factor:** <https://github.com/uRos2018/categorical>

Echipa: Yukako Toko (National Statistics Center, Japonia), Arne Schöman (IT.NRW Germania), Jan van der Laan (CBS, Olanda);

3) **Graphical frontend for the R package „validate”:** <https://github.com/uRos2018/validate>

Echipa: Matthias Gomolka (Deutsche Bundesbank, Germania), Mariko Kawano (National Statistics Center, Japonia), Mark van der Loo (CBS, Olanda), Marianna Mitratza (University of Amsterdam/CBS, Olanda), Kazumi Wada (National Statistics Center, Japonia);

4) **R package for voronoi treemaps with added interactivity by shiny:** <https://github.com/uRos2018/voronoiTreemap>

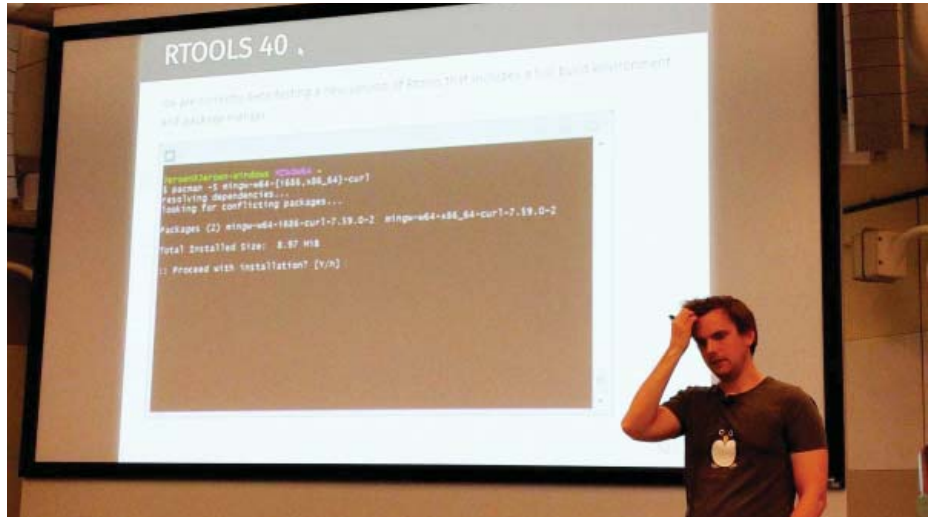
Echipa: Manolo Malaver Vojvodic (Statistics Canada), Alexander Kowarik (Statistics Austria), Bernhard Meindl (Statistics Austria) Reprezentanții INS au făcut parte din comitetul de evaluare al participanților la sesiunea *unConf*.

THE USE OF R IN OFFICIAL STATISTICS - UROS2018

The conference brought together 100 statisticians and professionals from official statistics and academia from 32 countries (South Africa, Austria, Belgium, Canada, Korea, Denmark, Switzerland, Estonia, Finland, France, Germany, Guatemala, Iceland, India, Ireland, Italy, Japan, Latvia, Lithuania, Luxembourg, United Kingdom, New Zealand, Netherlands, Poland, Portugal, Romania, Slovenia, Spain, Sweden, USA, Tanzania, Hungary) located on 6 continents. Moreover, we had as guests representatives from Eurostat and other international organizations (OECD, World Bank, Bureau for Labour Statistics USA)

The conference included parallel sessions with 60 presentations and 5 tutorials. The agenda also included a section of innovation projects on topics from various fields of official statistics, commonly called *unConf*.

In the opening, Wim van Nunspeet (Director of Corporate Services, IT and Methodology Department) from CBS underlined the importance of the



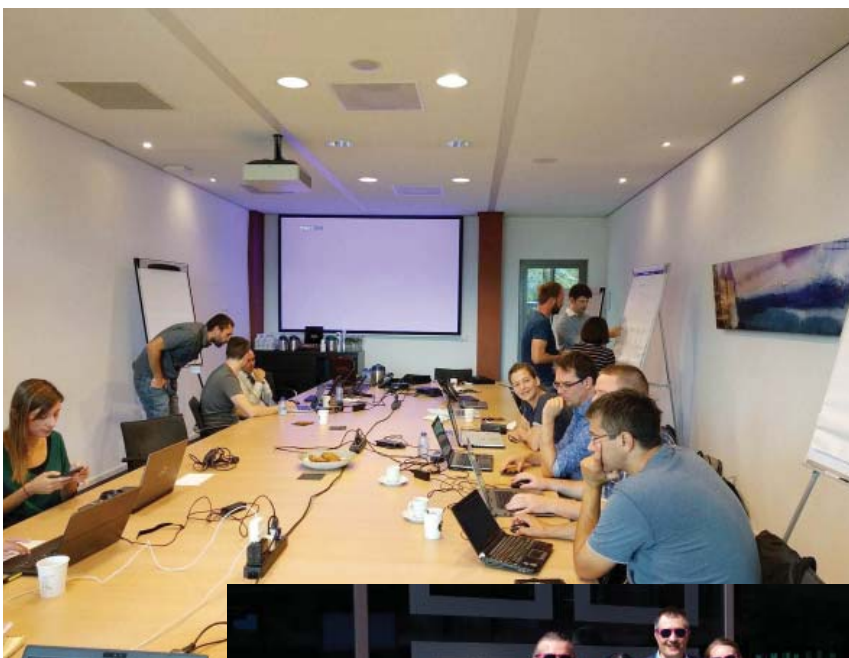
event for the development of official statistics and high-level international collaboration.

The 2018 edition of uRos marked the first occasion outside of the place where it was conceived: Romania.

The first keynote speaker was Alina Matei, author of R package *sampling* and professor at University from

Neuchâtel, Switzerland. In her presentation „Sample and coordination” she exposed a method for coordinating samples for statistical surveys.

The second keynote speaker was Jerome Ooms (the official maintainer for R for Windows). His presentation was about the infrastructure behind R and external libraries.



The output projects of *unConf* are the following:

1) *R wrapper for PDOK location server*: <https://github.com/uRos2018/nlgeocoder>

Team: Willy Tadema (Provincie Groningen, Netherlands), Egge-Jan Pollé (Tensing GIS Consultancy, Netherlands), Edwin de Jonge (CBS, Netherlands), Juris Breidaks (Central Statistical Bureau, Latvia);

2) *Categorical: a better factor*: <https://github.com/uRos2018/categorical>

Team: Yukako Toko (National Statistics Center, Japan), Arne Schöman (IT.NRW Germany), Jan van der Laan (CBS, Netherlands);

3) *Graphical frontend for the R package ,validate’*: <https://github.com/uRos2018/validaty>

Team: Matthias Gomolka (Deutsche Bundesbank, Germany), Mariko Kawano (National Statistics Center, Japan), Mark van der Loo (CBS, Netherlands), Marianna Mitratza (University of Amsterdam/CBS, Netherlands), Kazumi Wada (National Statistics Center, Japan);

4) *R package for voronoi treemaps with added interactivity by shiny*: <https://github.com/uRos2018/voronoiTreemap>

Team: Manolo Malaver Vojvodica (Statistics Canada), Alexander Kowarik (Statistics Austria), Bernhard Meindl (Statistics Austria)

Cursuri și tutoriale



Una din zilele conferinței a fost dedicată cursurilor tematice/tutoriale care au avut ca lectori de prestigiu autori de pachete R și de metodologii de referință din statisticile oficiale din întreaga lume. Temele tutorialurilor au fost următoarele:

1. *Spatial Analysis in R with open Geodata*

Lectori: Egge-Jan Pollé și Willy Tadema (Tensing GIS Consultants, Provincie Groningen, Olanda)

2. *So your data is Tidy, but is it Clean?*

Lectori: Edwin de Jonge și Mark van der Loo (CBS, Olanda)

3. *Fast and efficient data manipulation with data.table*

Lectori: Jaap Walhout (CBS, Olanda)

4. *Plotting Spatial data in R*

Lector: Martijn Tennekes (CBS, Olanda)

5. *Use of R package SamplingStrata for the Optimal Stratification of Sampling Frames for Multipurpose Sampling Surveys*

Lector: Giulio Barcaroli (ISTAT, Italia)



Tematici

Prezentările din cadrul conferinței s-au desfășurat în cadrul a 11 secțiuni paralele: Sampling and estimation, R in organization, Data cleaning, R in production: data analysis, Methods for official statistics, Shiny applications, Time series, Report and GUI programming, R in production: automation, Big data și Dissemination and visualization.

Acestea s-au axat pe următoarele tematici:

- Aplicarea software-ului R în producția statistică.
- Aspecte organizatorice și tehnice ale introducerii R în statisticile oficiale naționale.
- Exemple de aplicare a R în diseminarea statisticilor (vizualizare, aplicații, raportare).
- Analiza datelor de dimensiuni mari – Big Data.
- Implementarea metodologiei statistice în domeniile eșantionării, editării, modelării și controlului estimării sau dezvoltării.
- Abordări privind învățarea software-ului R de către utilizatorii de statistici oficiale.

Rețeaua de socializare Twitter a fost de-a dreptul „inundată” cu postări ale participanților care își doresc să continue colaborarea cu echipa R din România.

Twitter has been full of posts from participants who want to continue working with Romanian R team.

Manolo Malaver-Vojvodic - Canada



Icarus @mmala0 · Sep 14

Special thanks to @statistiekpbs for being outstanding hosts, to the @uRos2018 organizers @markvdloo @Alexkvienna and to all the amazing people that I met at the conference! Looking forward to #uros2019 in Bucharest @ro_statistics @alexcipro @AnaCiuhu @NicolStat

Oscar de León - Guatemala



Oscar de León @odeleongt · Sep 15

Hey! @uRos2018 was an amazing conference. #awesome, some would say.

So good that, if you want to find me late in May next year, go to Bucharest :)

Conferința a constituit o oportunitate de a dezvolta noi idei și cooperări în domeniul statisticii oficiale și mediul academic. Acest eveniment a adus în prim plan rolul Institutului Național de Statistică în statistica oficială. Prezentările uRos2018 vor fi disponibile pe site-ul oficial al conferinței: www.urosconf.org.

Lucrările conferinței se vor publica în două reviste: *Revista Română de Statistică* și *Austrian Journal of Statistics*. Calitatea lucrărilor prezentate, bogata activitate desfășurată de autorii

cercetărilor, dar și rezultatele obținute prin cooperare internațională pot constitui tot atâtea puncte forte, de referință, pentru a concluziona că România este pe harta utilizatorilor mediului de analiză statistică R. Conferința a fost promovată printr-un logo aplicat pe o pereche de ochelari de culoare roz, ca un simbol al acestui eveniment (awesomesofficialstatistics.org). Promovarea se va continua pe rețelele de socializare cu utilizarea acestor ochelari și menționarea @urosconf.



Conferința **uros2019** va avea loc la București, iar în următorii ani este posibil ca evenimentul să se desfășoare, ca și în acest an, în alte sedii ale oficiilor de statistică din alte țări din Europa și nu numai.

Tutorials



The conference has included one full day of tutorials lectured by prestigious professionals from official statistics, authors of R packages and authors of important methodologies highly used in worldwide official statistics:

1. Spatial Analysis in R with open Geodata

Lecturers: Egge-Jan Pollé and Willy Tadema (Tensing GIS Consultants, Provincie Groningen, Netherlands)

2. So your data is Tidy, but is it Clean?

Lecturers: Edwin de Jonge and Mark van der Loo (CBS, Netherlands)

3. Fast and efficient data manipulation with data.table

Lecturer: Jaap Walhout (CBS, Netherlands)

4. Plotting Spatial data in R

Lecturer: Martjn Tennekes (CBS, Netherlands)

5. Use of R package SamplingStrata for the Optimal Stratification of Sampling Frames for Multipurpose Sampling Surveys

Lecturer: Giulio Barcaroli (ISTAT, Italy)

The conference was an opportunity to develop new ideas and cooperation in the field of official statistics and academia. The event enabled once again the significant role of National Institute of Statistics in the official statistics.

The presentations of the Conference uRos2018 will be available soon on the following website: www.urosconf.org.

A maximum of ten selected research papers will be published in a special issue of the *Romanian Statistical*

Review and additionally a maximum of five papers will be published in a special issue of *Austrian Journal of Statistics*.

Quality of the works presented, high research activity carried out by the authors, and the results may be as many strong points of reference, to conclude that Romania is on the map of userS.

The uRos2018 has been promoted by a logo applied on pink sunglasses, as a symbol of the event (awesomeofficialstatistics.org). The



Themes

The conference included 11 parallel sessions: Sampling and estimation, R in organization, Data cleaning, R in production: data analysis, Methods for official statistics, Shiny applications, Time series, Report and GUI programming, R in production: automation, Big data and Dissemination and visualization.

They focused on the following themes:

- Examples of applying R in statistical production.
- Examples of applying R in dissemination of statistics (visualisation, apps, reporting).
- Analyses of big data and/or application of machine learning for official statistics.
- R packages implementing statistical methodology
- R packages connecting R to other standard tools/technical standards
- Organisational and technical aspects of introducing R to the statistical office.
- Examples of accessing or using official statistics publications with R in other fields



advertising will continue on social media by using the sunglasses in pictures and the handle [@urosconf](https://twitter.com/urosconf). Conference uros2019 will be held in Bucharest in May, and in the next years probably the event will move to other statistical offices across the world.



Echipa României a fost formată din Nicoleta Caragea și Ciprian Alexandru, Direcția de Calcul al Indicatorilor privind Populația și Migrația Externă, Ana-Maria Ciuhu, Direcția de Statistică Socială, cărora li s-a alăturat Adrian Dușa, Universitatea București și Alina Matei, Universitatea Neuchâtel.

Romanian representatives at uRos 2018 were: Nicoleta Caragea and Ciprian Alexandru, Direction of Population and International Migration Statistics, Ana-Maria Ciuhu, Department of Social Statistics, Adrian Dușa, University of Bucharest and Alina Matei, University of Neuchâtel.

