

IASC-LARS WEBINAR ON COMPUTATIONAL STATISTICS AND DATA SCIENCE



IASC- LARS SCHOOL ON COMPUTATIONAL
STATISTICS AND DATA SCIENCE

DATA VISUALIZATION THEORY AND APPLICATIONS

NATALIA DA SILVA

UNIVERSIDAD DE LA REPUBLICA, URUGUAY

IFCEA
IESTA

NOVEMBER 22-24, 2021

INTERNATIONAL ORGANIZING COMMITTEE

ALBA MARTÍNEZ-RUIZ, MONDAY ADENOMON, PAULO CANAS RODRIGUES
HAN-MING (HANK) WU, SARA TASKINEN, CRISTIAN GATU
LUIS FIRINGUETTI-LIMONE, DAVID MUÑOZ-NEGRON

IASC-LARS Webinar on Computational Statistics and Data Science

DATA VISUALIZATION: THEORY AND APPLICATIONS

PROGRAM



IASC- LARS SCHOOL ON COMPUTATIONAL STATISTICS AND DATA SCIENCE

The Latin American Regional Section of the International Association for Statistical Computing (IASC-LARS), the IASC-LARS School on Computational Statistics and Data Science, the International Association for Statistical Computing (IASC), and the International Statistical Institute (ISI) are pleased to invite postgraduate and undergraduate students to attend the IASC-LARS Webinar Course “Data Visualization: Theory and Applications”. The course will be taught by Dr. Natalia da Silva from the Universidad de la República, Uruguay, November 22-24, 2021.

The IASC-LARS Courses aim (1) to spread the knowledge base and advances in Statistical Computing in Latin American and the world, (2) to provide an overview of the state-of-the-art of the ongoing research in computational statistics, (3) to provide an overall perspective of the application of computational statistics in data science problems, (4) to present applications where computational statistics have been crucial to solve problems in real-life applications, and (5) to increase the number of researchers and practitioners in computational statistics and data science.

IASC-LARS Webinar on Computational Statistics and Data Science

DATA VISUALIZATION: THEORY AND APPLICATIONS

PROGRAM

Data visualization plays an important role in every step of the statistical analysis workflow. Visualization is a key component for exploratory data analysis but also can help in model building and diagnosis. This workshop will cover the foundational principles of statistical graphics, computational tools for data visualization, and go over some challenges in advanced topics of visualization. This workshop is suitable for those interested in learning fundamentals in data visualization who are not very familiar with the grammar of graphics and its implementation in R and want to explore some advanced topics on data visualization. Students should know how to use R to make the most of this workshop.

As building blocks of data visualization, the workshop will cover the grammar of graphics, the relevance of graphics in the statistical workflow, and a glimpse of human perception and color theory. Then, the use of ggplot2 and other related packages to produce publication-ready plots, is presented as a hands-on tutorial. Finally, graphics tools for model diagnosis are explored, for example, display the model's main features in the context of the original data is key to understand how the model works. Other advanced topics, as visualization in high dimensions and graphical inference, are introduced.

INSTRUCTOR: Natalia da Silva (Instituto de Estadística, Universidad de la República).

AGENDA

(Local time: Montevideo – Uruguay, UCT -3 hours)

Monday – November 22

session 1: 09.00 – 10.30

session 2: 10.45 – 11.30

- Fundamentals of data visualization
 1. Importance of visualization.
 2. Basic data visualizations: (barchart, histogram, density plot, scatterplot, boxplot).
 3. Gestalt principles of visual perception: emergence, reification, multistability, invariance, laws of grouping (proximity, similarities, closure, continuity, connectedness).
 4. Aesthetic mappings and visual encodings of data.
 5. Use of color: how to select color based on the fundamental use cases (distinguish groups, represent data values and highlight).

6. Introduction to the grammar of graphics with ggplot2.

Tuesday – November 23

session 1: 09.00 – 10.30

session 2: 10.45 – 11.30

- Creating data displays in R
 1. How to do basic data visualizations with ggplot2.
 2. Multiple layers, facetting and tidying your data.
 3. Scales and color.
 4. Publication-ready plots: themes, axes and aspect ratios, annotations.
 5. Additional geoms, examples to visualize uncertainty.
 6. Introduction to interactive visualization with plotly.

Wednesday – November 24

session 1: 09.00 – 10.30

session 2: 10.45 – 11.30

- Advanced topics
 1. Visualizing models: display the model in the data space, look at all members of a collection and explore the process of model fitting, not just the end result.
 2. Introduction to graphical inference.
 3. Visualization in high dimensions, focus on visualizing high-dimensional numerical data using linear projections.

References

- Buja, Andreas, Dianne Cook, Heike Hofmann, Michael Lawrence, EunKyung Lee, Deborah F Swayne, and Hadley Wickham. 2009. Statistical inference for exploratory data analysis and model diagnostics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1906): 4361-4383
- Cook, Dianne, Deborah F Swayne, and Andreas Buja. 2007. *Interactive and dynamic graphics for data analysis: with r and ggobi*. Springer.
- da Silva, Natalia, Dianne Cook, and Eun-Kyung Lee. 2017. Interactive graphics for visually diagnosing forest classifiers in r. *arXiv preprint arXiv:1704.02502*.
- Lee, Stuart, Dianne Cook, Natalia da Silva, Ursula Laa, Earo Wang, Nick Spyrison, and H Sherry Zhang. 2021. A review of the state-of-the-art on tours for dynamic visualization of high-dimensional data. *arXiv preprint arXiv:2104.08016*.
- Majumder, Mahbubul, Heike Hofmann, and Dianne Cook. 2013. Validation of visual statistical inference, applied to linear models. *Journal of the American Statistical Association* 108 (503): 942-956.
- Wickham, Hadley. July 2009. *ggplot2: Elegant graphics for data analysis*. useR. Springer.

- Wickham, Hadley, Dianne Cook, and Heike Hofmann. 2015. Visualizing statistical models: Removing the blindfold. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 8 (4): 203-225.
- Wickham, Hadley, Dianne Cook, Heike Hofmann, and Andreas Buja. 2010. Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 973-979.
- Wilke, Claus O. 2019. *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media.
- Wilkinson, Leland. 2012. The grammar of graphics. In *Handbook of computational statistics*, 375-414. Springer.

INSTRUCTOR



Natalia da Silva is an assistant professor at Instituto de Estadística, Universidad de la República, Montevideo, Uruguay (since 2017). She has a PhD in Statistics (2017) and a MSc in Statistics (2014) from Iowa State University. She has a BSc in Statistics (2007) and BSc in Economics (2008) from Universidad de la República. Some of her research interests are, supervised learning methods, prediction, exploratory data analysis, statistical graphics, reproducible research and meta-analysis.

IASC-LARS Webinar on Computational Statistics and Data Science

DATA VISUALIZATION: THEORY AND APPLICATIONS

REGISTRATION PROCEDURE

The IASC-LARS Webinar “Data Visualization: Theory and Applications” will be held virtually using the platform [GoToWebinar](#) from November 22 to 24, 2021. The official language is English. The deadline for registration is **November 10, 2021**. The course is free of charge.

Please complete the [IASC-LARS Webinar Registration Form](#) at <https://register.gotowebinar.com/register/1393724862580276494>

To become a IASC-LARS member, please complete the Membership Application Form at <https://www.isi-web.org/index.php/membership/individual-membership/iasc>.

All participants are expected to adhere to the ISI Community Principles and Conduct Policy (<https://www.isi-web.org/index.php/about-isi/policies/community-conduct>).

More information about GoToWebinar please visit <https://www.gotomeeting.com/webinar>. GoToWebinar application is also available for iOS, Android and Windows Phone: <https://support.goto.com/webinar/help/gotowebinar-for-mobile-devices-g2w050033>.

IASC-LARS Webinar on Computational Statistics and Data Science

DATA VISUALIZATION: THEORY AND APPLICATIONS GOTOWEBINAR PLATFORM

To attend this virtual course:

1. Please complete the IASC-LARS Webinar Registration Form at <https://register.gotowebinar.com/register/1393724862580276494> as soon as you decide to attend the course.
2. You will receive an email with an URL link to attend the webinar in the GTW platform.
3. Please, try to connect about 30 minutes before the start of the course.
4. We recommend joining via a high speed and wired connection and to use a USB headset for best sound quality.
5. Before to attend the course, please visit the webpage <https://www.gotomeeting.com/webinar/join-webinar> and see the video “GoToWebinar Attendee Quick Start”. For more information, you can also visit the YouTube channel <https://www.youtube.com/user/gotowebinar>.
6. If you have any questions or you need any help, please do not hesitate to contact us at iasc.lars.iasc@gmail.com.