

# GSOC 2018: A Shiny User Interface to RobStatTM

Project Idea by Matias Salibian-Barrera and Doug Martin

# 1 Summary

RobStatTM is a new R robust statistics package available at <https://github.com/msalibian/RobStatTM>, and is maintained by Matias Salibian-Barrera. This package is a complement to the second edition of *Robust Statistics: Theory and Methods* by Maronna, R., Martin, R.D., Yohai, V. J., and Salibian-Barrera, M, to be published in 2018 (RSTM for short). The first edition of the book was by Maronna, Martin and Yohai and published in 2006. In the intervening years since the first edition a considerable amount of research by the co-authors and several colleagues has resulted in many new and improved robust methods.

Like the large majority of R packages, the robust methods in RobStatTM are available only to R command line users. Consequently the learning curve required for effective use the robust methods in RobStatTM is steep and many applied researchers in statistics and quantitative finance who could benefit from the package will not use it. The goal of this project is to develop a Shiny point and click user interface (UI) to RobStatTM, much in the spirit of the Graphical User Interface (GUI) type UI that was developed for the S-PLUS Robust Library many years ago. Details and the flavor of that UI are provided in Appendix A. We believe that the result of this project will be a very substantial increase in the number of users of RobStatTM, both in university courses in engineering, applied mathematics and statistics, and among practicing statisticians and finance quants, the latter market for which R already has deep inroads. The net result of this will be more frequent and effective use of modern robust statistical methods.

## 2 Background

The S-PLUS product, developed years ago by Statistical Science, Inc., subsequently renamed Insightful, Inc, and subsequently acquired by TIBCO, developed a point and click front end for the command-line based robust statistics library. This resulted in a substantially increased user base of statisticians who would not have used the command-line only library, and thereby increased the applications use of robust statistics in various market segments, including bio-statistics and bio-technology, and in quantitative finance.

In the past there have existed two main R packages for robust statistics, the `robust` package and the `robustbase` package. During its years of commercial viability S-PLUS focused on the `robust` package, with Kjell Konis as the main contribute to and maintainer of the complete `robust` package with the GUI partially described in Appendix A. But after the S-PLUS company Insightful,

Inc. was sold to TIBCO in 2008, S-PLUS was no longer actively marketed and interest in S-PLUS rapidly decreased. At about the same time interest in and use of R began to grow at an astonishingly rapid rate, and it became a dominant open source statistics development and solutions deployment language and environment. Corresponding to the rapidly decreasing interest in S-PLUS, the need for an R robust statistics package was fulfilled by the development of the `robustbase` package by an ETH Zurich centric team.

As it turned out during the creation of the 2nd edition of *Robust Statistics: Theory and Methods* (2018), henceforth referred to as RSTM, due to a significant number of new robust methods in the book, and changes that would have been needed to the `robustbase` package in order for it to serve the needs of the book, the authors decided to create the `RobStatTM` package, borrowing some components of `robustbase` where it made sense, to ideally support the 2nd edition for uses in both teaching and research.

### 3 Project Goals

An over-arching initial goal of the two mentors of this project, has been to develop a complete user manual complement to RSTM, along with the companion complete R script that creates all the examples in the book. The latter is in the spirit of “reproducible research”, which is of rapidly growing general interest.

Moreover, the dramatic growth in the capabilities of Shiny and its use for both teaching and industrial strength deployment of R code solutions, has made it clear that an expansion of the initial goal stated above to include development of a Shiny point and click UI front-end to the `RobStatTM` package is a hugely attractive overall goal. As the RStudio web site <https://www.rstudio.com/products/shiny-2/> states: “Shiny is an open source R package that provides an elegant and powerful web framework for building web applications using R. Shiny helps you turn your analyses into interactive web applications without requiring HTML, CSS, or JavaScript knowledge”

The over-arching goal of this project is to develop Shiny interfaces to the following robust methods described in RSTM: (1) Selected robust estimators of location and scale estimators as described in Chapter 2 of RSTM, (2) Robust linear regression for linear models with non-random independent variables as in Chapter 4 of RSTM; (3) Robust linear regression with random independent variables as in Chapter 5 of RSTM; (4) Robust covariance matrix estimation and robust principal components as in Chapter 6 of RSTM. The UI for the robust package which is described in Appendix A can provide a point of departure for designing an even better UI for the `RobStatTM` package.

## 4 Milestones

### Startup

Study the key methodology parts of RSTM Chapters 1-3, and the related R functions that are available in RobStatTM.

Complete the design and implementation of the Shiny interface for selected robust location and scale estimators in Chapter 2 of RSTM.

### Phase 1

Study the key methodology parts of RSTM Chapter 4: Linear Regression 1, and the related R functions that are available in RobStatTM.

Complete the design and implementation of the Shiny interface for the kinds of linear models described in Chapter 4.

### Phase 2

Study the key methodology parts of RSTM Chapter 5: Linear Regression 2, and the related R functions that are available in RobStatTM.

Complete the design and implementation of the Shiny interface for the kinds of linear models described in Chapter 5.

### Phase 3

Study the key robust covariance matrix estimation and PCA methodology parts of RSTM Chapter 6: Multivariate Analysis, , and the related R functions that are available in RobStatTM.

Complete the design and implementation of the Shiny interface for the robust covariance matrix estimators and robust PCA described in Chapter 6.

## 5 Expected Impact

The result of this project will make a wide range of very modern robust statistics methods available to a very broad range of applied statisticians, quantitative finance professionals and engineers, who would not otherwise use robust statistics that are available in purely command-line access software.

## 6 Skills Required

- Experience in use of R and writing R code
- Demonstrable experience in use of github, or ability to rapidly develop such knowledge by May 24, 2018.
- Knowledge of R package development, and in particular solid knowledge of this as presented in the book *R Packages* (2015) by Hadley Wickham, or ability to rapidly develop such knowledge rapidly by May 24, 2018.
- Ability to rapidly acquire Shiny applications development.
- MS degree level statistical theory and applications knowledge
- Ability to comfortably read and understand key robust statistical methods (but not the mathematical details and proofs) described in Chapters 1 through 6 of *Robust Statistics: Theory and Methods* (2018), 2nd edition by Maronna, Martin, Yohai and Salibian-Barrera (RSTM for short).

## 7 Project Proposal Requirements

The successful applicant will:

1. Define the Phase 1 Shiny interface, and design, implement and test it. Write Part 1 of a vignette.
2. Define the Phase 2 Shiny interface, and design, implement and test it. Write Part 2 of the vignette.
3. Define the Phase 3 Shiny interface, and design, implement and test it. Write Part 3 of the vignette.

NOTE: Testing above implies not only developer testing, but getting other interested individuals to test the UI and provide suggestions for improvements.

## 8 Mentors

- Matias Salibian Barrera (Lead Mentor). Co-Author of RSTM 2018 book, and Maintainer of RobStat™ package. Professor of Statistics, The University of British Columbia, matias@stat.ubc.ca.
- R. Douglas Martin, Professor Emeritus, Departments of Applied Mathematics and Statistics, University of Washington. Founder and former Director of the UW Computational Finance and Risk Management MS Degree program. doug@amath.washington.edu.

## 9 Collaborators

- RSTM co-authors Ricardo Maronna and Victor J. Yohai.
- Kjell Konis, PhD, Developer and Maintainer of the the robust R package, University of Washington Staff Member , and former faculty member of MS CFRM degree program in Applied Mathematics, University of Washington.
- Ruben Zamar, Extensive track record of ongoing research on robust statistical methods, Professor of Statistics, The University of British Columbia, ruben@stat.ubc.ca.
- Others TBD, particularly for usability testing of the Shiny UI