Recap of rstudio::conf(2018)

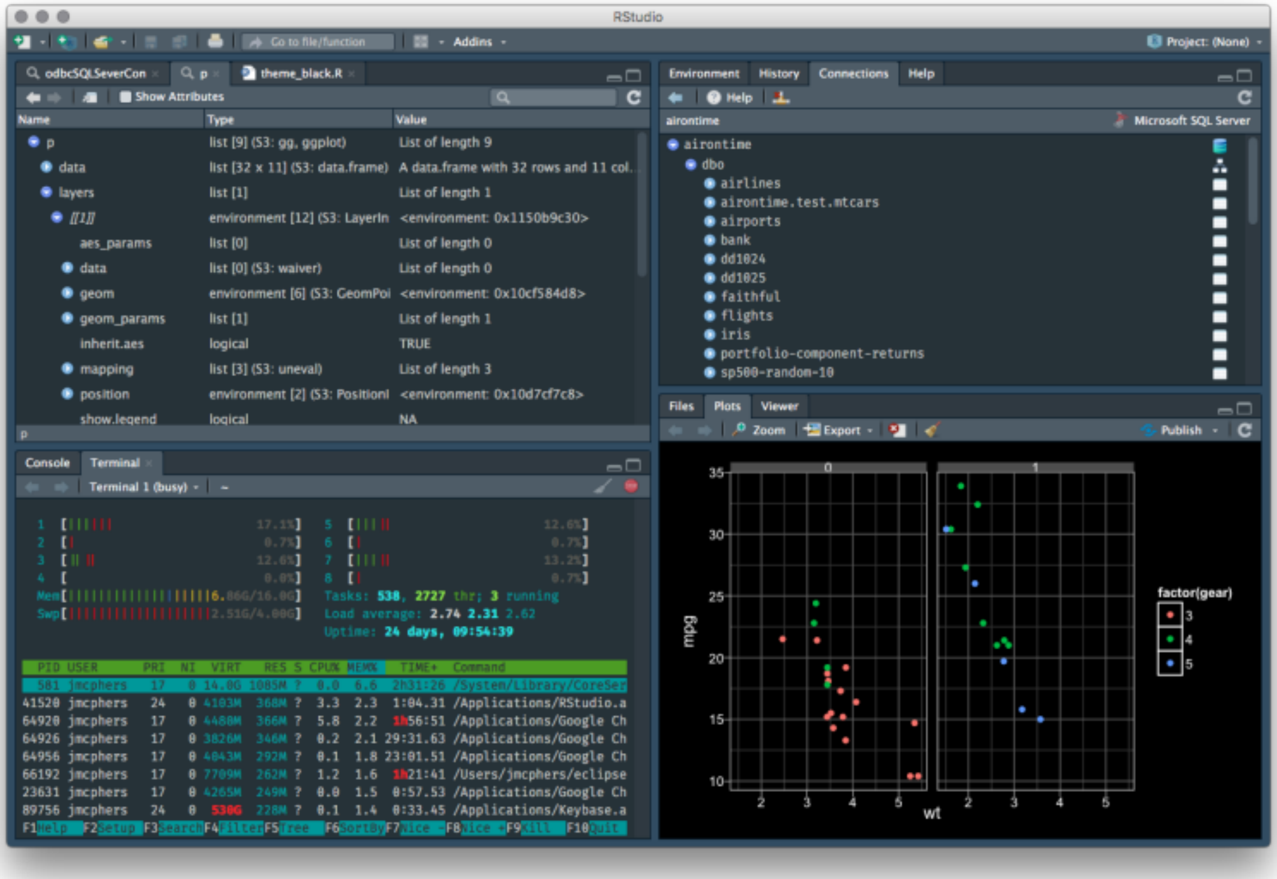
Colby Ford

From January 31st through February 3rd, Andy Lathrop and I attended rstudio::conf 2018. The conference offered two days of training events and two days of actual conference activities and was a lot of fun. Hosted at the beautiful Manchester Grand Hyatt right on the San Diego harbor, the convention included lots of different talks covering an array of subjects for all skill levels. Couldn't attend? Don't worry, I'll recap the best parts for you.

Next year's conference will be in Austin, Texas, January 15-18, 2019.

About RStudio

This section is mainly for the non-R readers in our midst. RStudio, located in Boston, is a tech company founded by J.J. Allaire, an acclaimed software engineer and entrepreneur. RStudio Desktop is the company's flagship product and is the most popular R integrated development environment (IDE) software, used by data scientists and statisticians across the planet. It allows for users to write code, see objects in the coding environment, view plots and tables, and more.



Name	Type	Value
p	list [9] (S3: gg, ggplot)	List of length 9
data	list [32 x 11] (S3: data.frame)	A data.frame with 32 rows and 11 col...
layers	list [1]	List of length 1
[[1]]	environment [12] (S3: LayerIn	<environment: 0x1150b9c30>
aes_params	list [0]	List of length 0
data	list [0] (S3: waiver)	List of length 0
geom	environment [6] (S3: GeomPoi	<environment: 0x10cf584d8>
geom_params	list [1]	List of length 1
inherit.aes	logical	TRUE
mapping	list [3] (S3: uneval)	List of length 3
position	environment [2] (S3: PositionI	<environment: 0x10d7cf7c8>
show.legend	logical	NA

p

airontime

```
airontime
  dbo
    airlines
    airontime.test.mtcars
    airports
    bank
    dd1024
    dd1025
    faithful
    flights
    iris
    portfolio-component-returns
    sp500-random-10
```

```
  1  [||||||                    17.1%]   5  [|||||                     12.6%]
  2  [|                          0.7%]   6  [|                          0.7%]
  3  [|| ||                     12.6%]   7  [|||||                     13.2%]
  4  [                           0.0%]   8  [|                          0.7%]
  Mem[|||||||||||||||||||6.86G/16.0G]   Tasks: 538, 2727 thr; 3 running
  Swp[|||||||||||||||||||2.51G/4.00G]   Load average: 2.74 2.31 2.62
                                         Uptime: 24 days, 09:54:39

  PID USER      PRI  NI  VIRT   RES S CPU% MEM%   TIME+  Command
  581 jmcphers   17   0 14.0G 1085M ?  0.0  6.6  2h31:26 /System/Library/CoreSer
41520 jmcphers   24   0 4103M  368M ?  3.3  2.3  1:04.31 /Applications/RStudio.a
64920 jmcphers   17   0 4488M  366M ?  5.8  2.2  1h56:51 /Applications/Google Ch
64926 jmcphers   17   0 3826M  346M ?  0.2  2.1 29:31.63 /Applications/Google Ch
64956 jmcphers   17   0 4043M  292M ?  0.1  1.8 23:01.51 /Applications/Google Ch
66192 jmcphers   17   0 7709M  262M ?  1.2  1.6  1h21:41 /Users/jmcphers/eclipse
23631 jmcphers   17   0 4265M  249M ?  0.0  1.5  0:57.53 /Applications/Google Ch
89756 jmcphers   24   0  530G  228M ?  0.1  1.4  0:33.45 /Applications/Keybase.a
F1Help  F2Setup F3SearchF4FilterF5Tree  F6SortByF7Nice -F8Nice +F9Kill  F10Quit
```

At the conference, one talk covered the enhancements to the software in version 1.1. Highlights included:

- **Connections tab:** Allows the user to connect to, explore, and view data in a variety of databases.
- **Terminal tab:** Provides shell integration within the IDE so that users can now execute terminal commands without leaving the program.
- **Object explorer enhancements:** Easily navigate deeply nested R data structures such as complex lists and S3 objects.
- **UI enhancements:** Two new UI themes are available (a modern, flat theme and a dark theme), as well as Retina-quality icons throughout.

For more information on the 1.1 enhancements, see the RStudio Blog.

In addition to RStudio Desktop, the company also has an array of other products, such as:

- RStudio Server - A browser-based IDE useful for centralizing access and computation across an organization.
- Shiny and Shiny Server Pro - Software that provides an elegant and powerful web framework for building web applications using R.
- RStudio Connect - A new publishing platform to share Shiny applications, R Markdown reports, plumber APIs, dashboards, plots, and more in one convenient place.

Many of RStudio's products are open source (AGPL v3), but offer commercial licenses with additional features. Check out the website for more information.

All Things Shiny

Many of the sessions at the event centered around Shiny, a web application development framework for R. Without any knowledge of HTML or CSS, R programmers can create interactive web pages for analyzing data or showing results. At the conference, there were many sessions around designing amazing Shiny dashboards, using API's with the plumber and Shiny to ingest and return data, and even returning results in real-time of machine learning model training.

I have used Shiny in the past and I must admit that it makes some really nice web apps. So, I'm excited to see the versatility in the platform grow!

See the Shiny Gallery for great examples.

Keeping it Tidy

Another big theme at the conference was the tidyverse. Originally philosophized by Hadley Wickham, RStudio's Chief Scientist, the tidyverse is a set of packages that share an underlying design philosophy, grammar, and data structures. Hadley's one-sentence definition of tidy is "Every column is variable, every variable is a column, and every row is an observation." This attempts to standardize how R functions and packages are written so that each package can work together harmoniously. It was great to see other developers (outside of RStudio) publishing and sharing packages that follow these design practices. I attended multiple sessions where developers were presenting their work and demonstrating the interoperability with the tidyverse library.

tidyverse

Visit the tidyverse website to learn more.

Prior to the actual conference beginning, a couple hundred of us attended a training session called "Extending the tidyverse". This session covered some really useful packages for creating a package of your own, testing your packages, and documenting with ease. This really served as a nice walkthrough of the end-to-end process of making a tidy-compliant package in R.

Some big takeaways:

- The *usethis* package and R Project files make life easy when you want to make a package.
  - This creates all the files necessary for publishing a package such as *NAMESPACE*, *DESCRIPTION*, *LICENSE*, etc. and an .Rproj file for your code.
  - `usethis::create_package("~/Documents/mynewpackage")`
- Test your code with *testthat* to make sure your functions give you what you expect.
  - Automatically create test files that correspond to each of your functions.
  - `usethis::use_test("myfunction")`
  - `devtools::test()`
- The *purrr* package is nice for handling redundant function needs and you can ensure stuff won't fail.
  - `purrr::map(input, function)`
  - `purrr::safely(function)`

- We all know documentation is important. With *roxygen2*, documentation is as easy as using a special comment.

    -
      ```
      #' My header goes here.
      #'
      #' Some text will go here...
      #'
      #' @param var1 Explanation of var1
      ```

For further reading, check out the R Packages book by Hadley Wickham.

Deep Learning with R

The keynote on day 2 of the conference was given by J.J. Allaire. He and other developers have created the Keras, Estimator, and Core APIs for R, which open the door for deep learning using TensorFlow. For those of you who are unfamiliar, TensorFlow is an open-source machine learning/computational framework (initially developed at Google) and Keras is a high-level interface into many deep-learning frameworks (including TensorFlow, the Microsoft Cognitive Toolkit, and Theano). However, Keras was originally written as a Python library. Using the R Interface to TensorFlow from RStudio, R developers can train complex neural networks and other machine learning models with ease.

Yes, you read that correctly -- R & deep learning in the same sentence? No, R is still a single-threaded, memory/CPU-based language, which typically isn't conducive to deep learning. However, R is a great interfacing language that can connect to other environments like a charm. So, this new package simply operates as the liaison between R and TensorFlow. If you don't have multiple GPU's in your desktop workstation, fear not! The keynote also demonstrated the ease of use by writing the R training script locally and then deploying it out to a cloud platform to use pay-as-you-go GPU's for faster training of complex models. Personally, I find this exciting as it opens up R to be more commonly used in large-scale machine learning projects than ever before.

To learn more about the Keras package for R, see the RStudio documentation.

It's Really All About Stickers

The data science equivalent of collecting Pokémon cards is collecting hex stickers. These are the stickerized versions of the logos for many of our favorite R packages. These stickers often find their home on the laptops of developers as a pseudo-status symbol in the R community. At the conference, it was almost a competition as to who could find and collect the most hex stickers. Even the RStudio employees were making riddles to hint to their sticker hiding places.

I'd say I ended up with a pretty decent collection:

Resources

Special thanks to Petr Simecek for curating the rstudio::conf 2018 resources. You can view his GitHub repository for all the links here.

Want to learn more about how BlueGranite can help solve your organization's advanced analytics challenges? Contact us today!