

Industry Case-studies

With

Data Mining & R

AbdulMajedRaja RS

Simply put,
What happens out there?
In the Real-World, With R

Credit - Thanks!

<https://github.com/ThinkR-open/companies-using-r>

<https://github.com/ColinFay>

The screenshot shows the GitHub repository page for `ThinkR-open/companies-using-r`. The page has a dark theme. At the top, there's a navigation bar with links for Pull requests, Issues, Marketplace, and Explore. Below the navigation, the repository name is displayed along with its fork status, star count (123), and fork count (8). A tab bar allows switching between Code, Issues (0), Pull requests (1), Projects (0), Wiki, and Insights. The main content area is titled "A Curated list of R uses in enterprise". It shows statistics: 66 commits, 1 branch, 0 releases, and 8 contributors. There are buttons for Branch: master, New pull request, Create new file, Upload files, Find File, and Clone or download. A commit history is shown with one entry from ColinFay updating README.md on Feb 8. The README file content is partially visible, featuring a section titled "Companies, Officials and NGO Using R" and a note about its purpose.

A Curated list of R uses in enterprise

66 commits 1 branch 0 releases 8 contributors

Branch: master ▾ New pull request Create new file Upload files Find File Clone or download ▾

ColinFay Update README.md Latest commit e9d2a07 on Feb 8

README.md Update README.md a month ago

Companies, Officials and NGO Using R

In case you need it, a curated list of testimonies / blogposts / packages from enterprises, officials and NGO using R in production and/or for research.

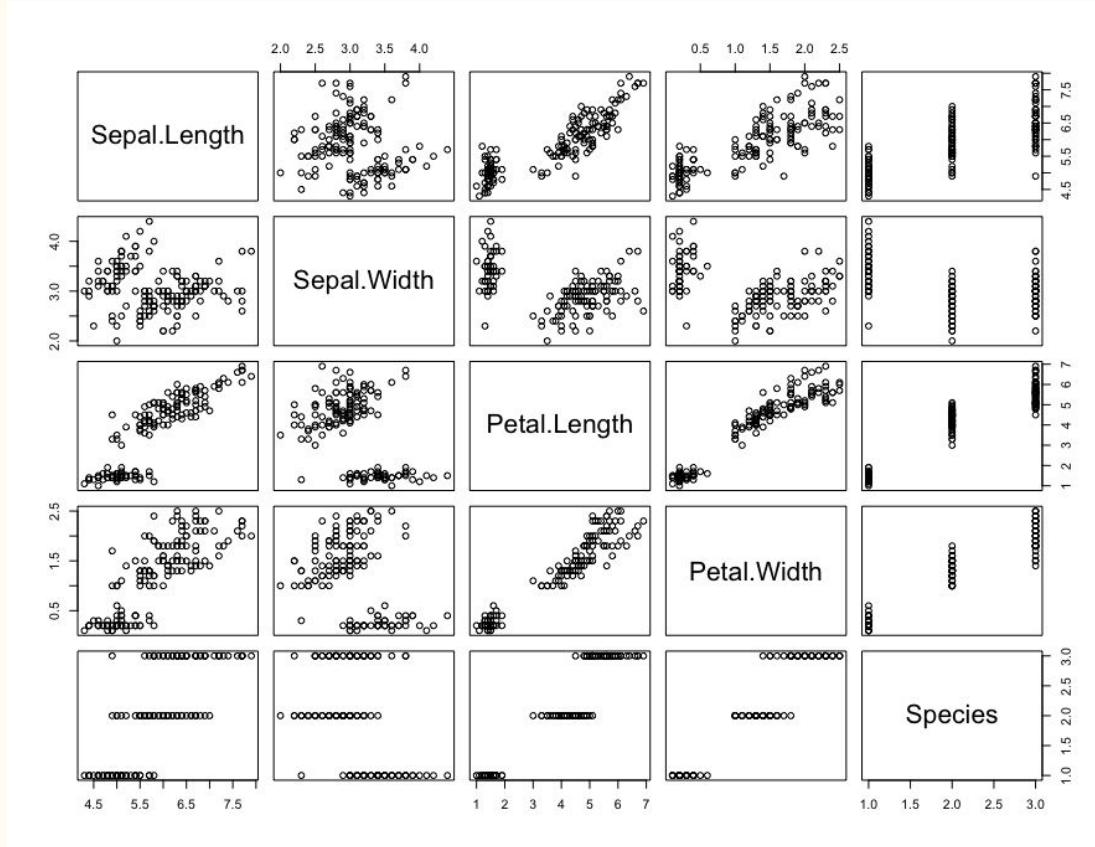
Feel free to contribute.

Before we Begin

What we think about R?

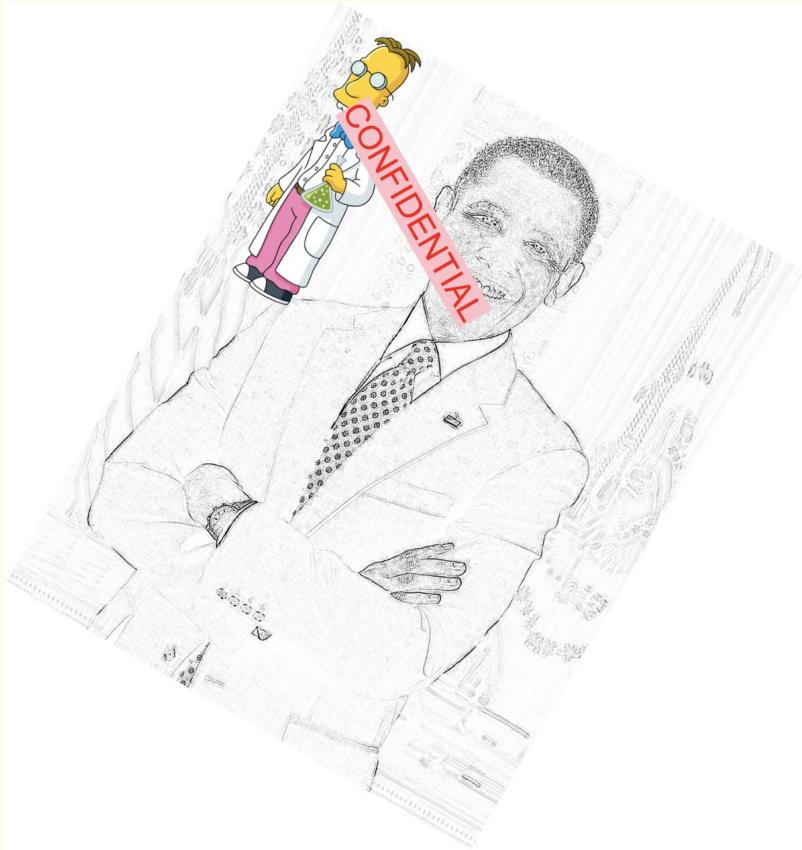
What comes to your mind when you someone mentions R?

This?

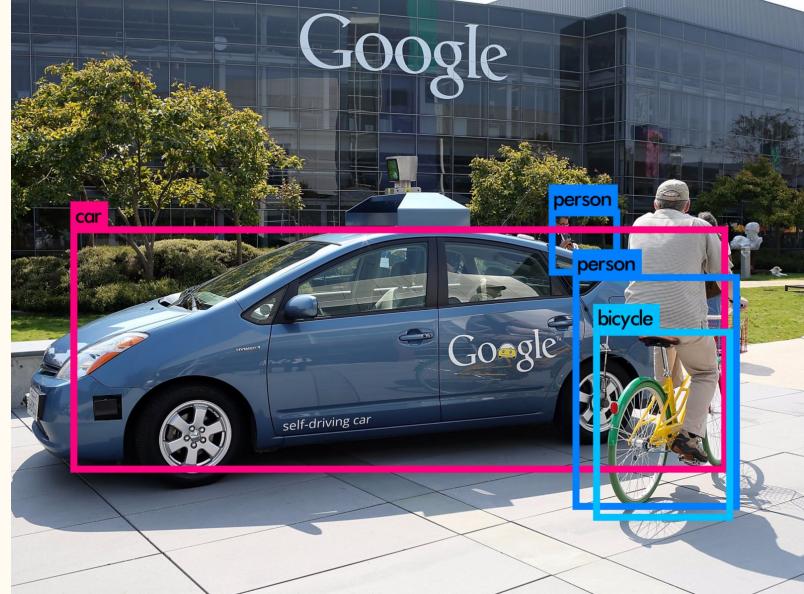


Is that all what R can do?

Image Processing?



Computer Vision?



Fun NLP Game?

https://apps.garrickadenbuie.com/trump-tweet-time/

Trump Tweet Time

When did Trump tweet...?

Russia Dossier reporter now doubts dopey Christopher Steele's claims! "When you get into the details of the Steele Dossier, the specific allegations, we have not seen the evidence to support them. There's good grounds to think that some of the more sensational allegations.....



Your Score: 1

During...

Lunch Executive Time Travel

Based on White House schedules released by Axios.
App by @grrrck using rtweet, Shiny, and nessy.

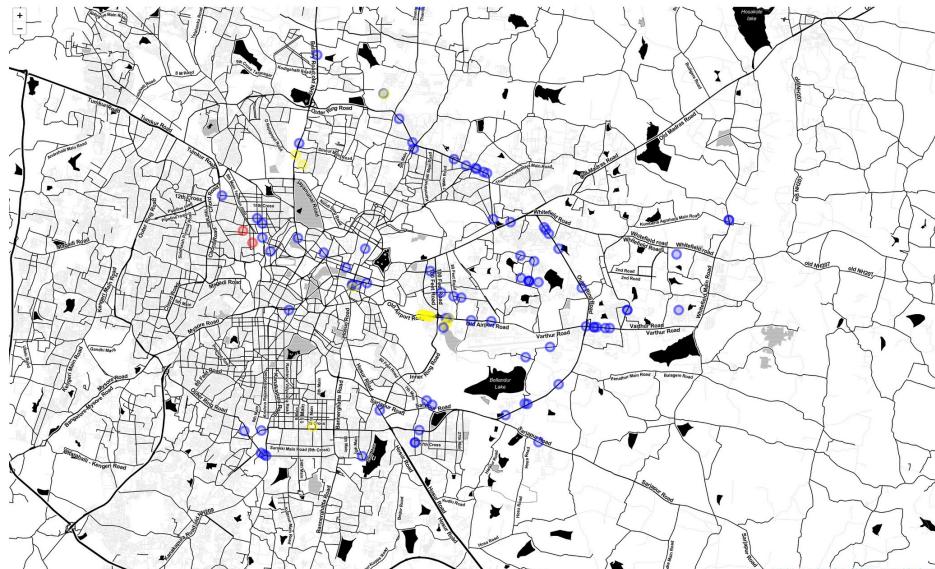
 View source on GitHub

Geospatial Analysis + Interactive Viz?

Building a WiFi spots Map of networks
around you with WiGLE and R



AMR
Feb 20 · 4 min read



Map generated using the below code

Book / Webpage / Blog / Web App?

https://amrrs.github.io/introduction_to_kaggle_tamil/index.html

1 முகவரை 2 ஏற்றாளர் பற்றி 3 கந்தகம் 4 கேக்ஸேவெக் 5 பயன்கள் நூல்கள்

கேகள் - ஒரு குட்டி அறிமுகம் - தமிழில்

அப்துல் மதீத் ராஜா

1 முகவரை

The screenshot shows the Kaggle homepage in Tamil. At the top, there's a navigation bar with links for '1 முகவரை', '2 ஏற்றாளர் பற்றி', '3 கந்தகம்', '4 கேக்ஸேவெக்', '5 பயன்கள்', and 'நூல்கள்'. Below the navigation is a heading 'கேகள் - ஒரு குட்டி அறிமுகம் - தமிழில்' and a subtitle 'அப்துல் மதீத் ராஜா'. A large '1 முகவரை' button is prominently displayed. The main content area features the Kaggle logo, a search bar, and links for 'Competitions', 'Datasets', 'Kernels', 'Discussion', 'Jobs', and 'Sign In'. The central part of the page has a dark background with white text: 'The Home of Data Science & Machine Learning', 'Kaggle helps you learn, work, and play', 'Create an account', 'or', 'Host a competition'. To the right is an illustration of an astronaut. Below this are sections for 'Competitions', 'Datasets', and 'Kernels'.

இந்த புத்தகம் தமிழ் மொழியில் பயிலும் மாணவர்கள் நன்றாக கேகள் பற்றிய ஒரு சிறிய அறிமுகம்

Next

That useless thing

R vs Python

The image shows a Google search interface. At the top is the Google logo. Below it is a search bar containing the text "r vs |". A dropdown menu lists several search suggestions: "r vs python", "r vs sas", "r vs matlab", "r vs python for data science", "r vs brown", "r vs python quora", "r vs prince", "r vs williams", "r vs tolson", and "r vs java". At the bottom of the search bar are two buttons: "Google Search" and "I'm Feeling Lucky". Below the search bar is a link to "Report inappropriate predictions".

r vs |

- r vs **python**
- r vs **sas**
- r vs **matlab**
- r vs **python for data science**
- r vs **brown**
- r vs **python quora**
- r vs **prince**
- r vs **williams**
- r vs **tolson**
- r vs **java**

Google Search I'm Feeling Lucky

Report inappropriate predictions

This useful thing

reticulate ----- R + Python

Combining the power of Python with R

R + Py

This is a simple (could be called, *naïve* as well) attempt to show how we can combine the power of Python with R and create a new superpower.



Like this one, If you have watched *The Incredibles* before!

Superstar - Reticulate

The superstar who's making this possible is the R package `reticulate` by RStudio.

Let us start with the code!!

The R Code

```
#loading required R libraries

library(tidyverse)
library(ggthemes)
library(knitr)
tweets <- read_csv("justdoit_tweets_2018_09_07_2.csv")
text <- tweets$tweet_full_text
set.seed(123)
text_10 <- text[sample(1:nrow(tweets),100)]
```

The Python Code

```
import spacy
import pandas as pd
nlp = spacy.load('en_core_web_sm')
```

A screenshot of the RStudio interface. On the left, there is an R script named 'py_plus_r.Rmd' containing R code for reading a CSV file and selecting 100 rows. On the right, there is a Python script named 'readme.md' containing Python code for loading the spaCy English model, creating a DataFrame, and performing POS tagging. The RStudio interface includes a top bar with tabs for 'Console', 'Terminal', 'R Markdown', and 'Jobs'. The 'Console' tab shows R code and its output. The 'R Markdown' tab shows the Python code. The 'Files' tab shows the project structure. The 'Plots' tab shows a histogram of POS tags. The 'Packages' tab shows available packages. The 'Help' and 'Viewer' tabs are also visible.

Let's begin!

Companies & Case studies

- BBC
 - Uber
 - Airbnb
 - Microsoft
 - Stack Overflow
-

BBC

BBC - Visual & Data Journalism with R Graphics

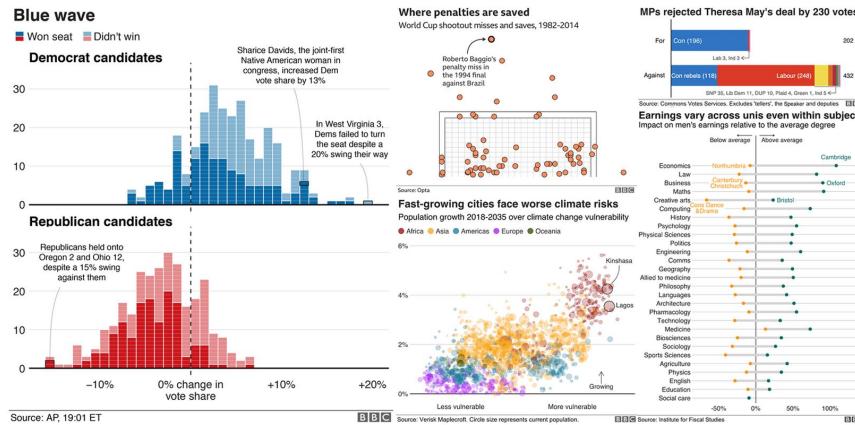
BBC Visual and Data Journalism cookbook for R graphics

Last updated: 2019-01-24

How to create BBC style graphics

At the BBC data team, we have developed an R package and an R cookbook to make the process of creating publication-ready graphics in our in-house style using R's ggplot2 library a more reproducible process, as well as making it easier for people new to R to create graphics.

The cookbook below should hopefully help anyone who wants to make graphics like these:



We'll get to how you can put together the various elements of these graphics, but let's get the admin out of the way first...

- Bbplot - bbc_style()
- R cookbook - Reference guide for ggplot2

Uber

Uber - Why R is a great tool for UBER

Business needs

Why R is a great tool for UBER



Reproducibility

Uber operates in various markets at different life stages. Reproducibility helps teams share knowledge, reduce duplication of effort and achieve greater consistency. Rmarkdown allows for reproducibility within the organisation.



Speed

Lean teams with competing priorities need to be able to quickly analyse data and set up reports. The Tidyverse ecosystem of packages allows for rapid, tidy data analysis.



Integration

We need a tool that can fetch data from various systems and send standardised output to different end points with little overhead. Internal packages allows us to meet this need.

- ubeR - Internal R package
- shinydashboard.uber / uberplot
- The tidyverse
- googlesheets
- Xgboost
- CausalImpact

Process automation in
R

Airbnb

Airbnb - Using R packages and education to scale Data Science at Airbnb

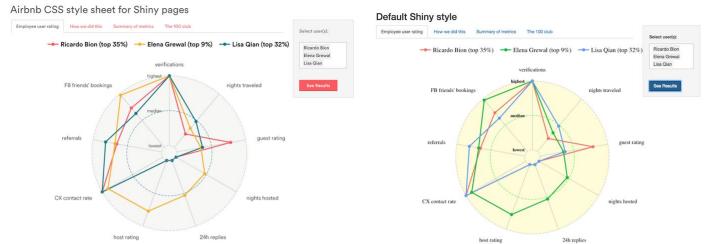


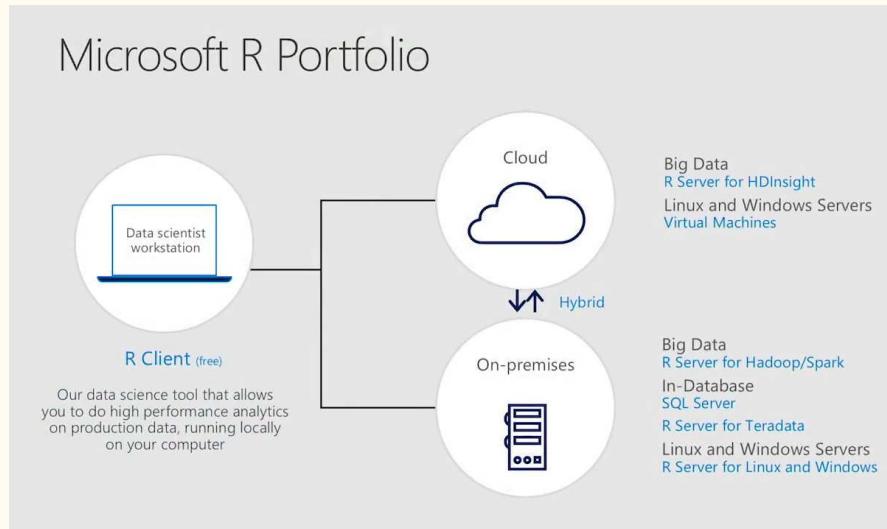
Figure 3. Airbnb branded CSS style sheet for Shiny exploring employee Airbnb usage. These style sheets give our internal tools an aesthetic that is consistent with the Airbnb.com website, making them more familiar and engaging to users.

```
1 ---  
2 title: "Your title"  
3 output:  
4 html_document:  
5 fig_height: 6  
6 fig_width: 12  
7 keep_md: yes  
8 ---  
9  
10 **Author**: <a href=  
11 **Date Created**: 2  
12 **Date Updated**: 2  
13 **Tags**: #knowledge  
14 **TLDR**: This is s  
15  
16  
17 ````{r global_option  
18 library(knitr)  
19 library(Rbnb)}  
20 opts_chunk$set(fig.  
21 ````  
22  
23 _NOTE: In the TLDR, optimize for **clarity** and **comprehensiveness**. The goal is to  
convey the post with the least amount of friction. Make the reader get a correct  
understanding of the post's takeaway, and the points supporting that takeaway without  
having to strain through paragraphs and tons of prose. Bullet points are great here, but  
are up to you. Try to avoid academic paper style abstracts..
```

- **Rbnb - Airbnb internal R Package**
 - a **consistent API** to move data between different places in the data infrastructure
 - **branded visualization** themes, scales, and geoms for ggplot2
 - **R Markdown templates** for different types of reports
 - custom functions to **optimize** different parts of our **workflow**.
- **Shiny**
 - Using Shiny to make decisions at Scale

Microsoft

Microsoft - contributing to R by tightly integrating R with its products

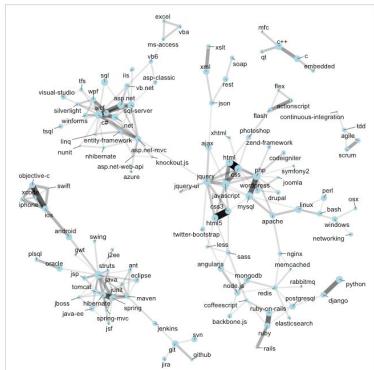
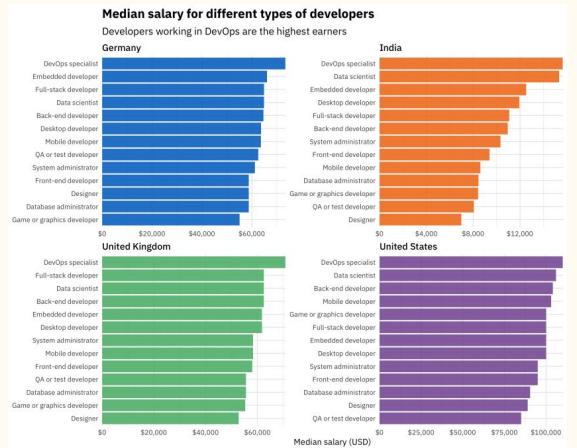


- **Microsoft R Open** - Microsoft's distribution of open-source R.
- **MRAN**, which is the download repository for Microsoft R Open, and also hosts daily archive snapshots of the entire CRAN system (from 2015 to the present).
- R available from within Microsoft products
- Open-sourcing *foreach* Package

Personalised Marketing at Scale with Propensity Models using R and Spark

Stack Overflow

Stack Overflow



- **Survey Analytics - Stack Overflow Developer Survey**
- **A/B Testing - R-based Tools and Functions to assist in A/B Testing**
- **Research, Blogging, Communicating - Rmarkdown**

Much More!!!

Sports Betting and R: How R is changing the sports betting world

Jul 12, 2017 at 9:07AM by Marco Blume
 ★★★★★ 4 ratings 1 comment

PINNACLE

Sports betting & R:

Marco Blume
 marco.blume@pinnacle.com

UseR_2017

T-mobile uses R for Customer Service AI

T-Mobile, the global telecommunication company, is using R in production to automatically classify text messages to customer service and route them to an agent that can help. The AI@T-mobile team used the keras library in R to build a natural language processing engine with Tensorflow, and deployed it to production as a docker container. The MRAN Time Machine ensures the container gets fixed R package versions for reproducibility.

While you may think first of Python for building production AI models, the keras library in R has already proven to be a first-class interface for building AI models. The T-mobile team was able to build this system in just four months and with a small budget because they were able to explore in R and immediately deploy in R:

"Despite being an incredibly popular language for exploratory analysis, data scientists are repeatedly told that R is not sufficient for machine learning – especially if those ML models are destined for production. Though much exploratory analysis and modeling is done in R, these models must be rebuilt in Python to meet DevOps and enterprise requirements. Our team doesn't see the value in this double work. We explore in R and immediately deploy in R. Our APIs are neural network models using R and TensorFlow in docker containers that are small and maintainable enough to make our DevOps team happy!"

AI AI @ T-Mobile

If you'd like to try this out yourself, the T-mobile team has published an open source version of their message-classification container on [Github](#), and you can read the details in the blog post by team leads at the link below.

T-Mobile blog: [Enterprise Web Services with Neural Networks Using R and TensorFlow](#)

R in production: theRe be dRagons!



TIM SWEETSER AND KYLE SCHMAUS

June 15, 2017 - San Francisco, CA

[Tweet this post!](#) [Post on LinkedIn](#)



Here at Stitch Fix, we have an ambivalent relationship with R. We love R for its bleeding-edge statistical models, e.g. `mbest` and `bsts`, and for the robust data science toolchain known as the `tidyverse`. When we're working interactively, we love using R to explore new models and visualize our data with `ggplot2`. But when we're running production jobs, such as fitting models inside Docker images on cloud computers, R

More Resources

Videos, Blogs

R Studio

Products Resources Pricing About Us Blogs

rstudio::conf 2019

RStudio Connect Past, present, and future - Jeff Allen

RStudio Connect is a publishing platform that helps to operationalize the data science work you're doing. We'll review the current state and unveil some exciting new features.

WATCH VIDEO >

Shiny in production: Principles, practices, and tools - Joe Cheng

Shiny is a web framework for R. As such, Shiny has always faced

WATCH VIDEO >

Welcome and RStudio Vision - Tareef Kawaf

WATCH VIDEO >

Why TensorFlow eager execution matters - Sigrid Keydana

In current deep learning with Keras and TensorFlow, when

WATCH VIDEO >

Visualizing uncertainty with hypothetical outcomes plots - Claus Wilke

Uncertainty is a key component of

WATCH VIDEO >

R consortium

Supporting the R Community, the R Foundation and organizations developing, maintaining and distributing R software.

R Consortium 2,721 subscribers

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT SUBSCRIBE 2.7K

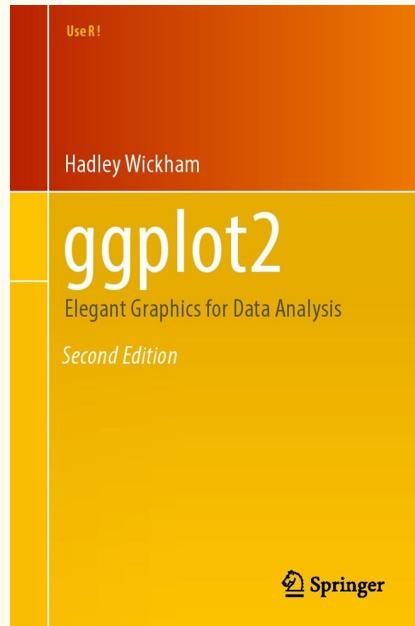
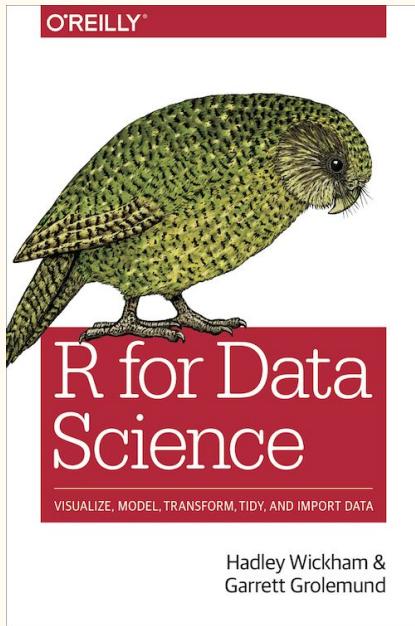
Uploads PLAY ALL SORT BY

Video Title	Views	Published
Code smells and feels	6.3K	7 months ago
Teaching R to New Users: From tidy to Tidyverse	4.3K	7 months ago
The Grammar of Animation	3.8K	7 months ago
Shiny meets Electron: Turn your Shiny app into a	3.3K	7 months ago
Recipes for Data Processing - Part 1	3.2K	7 months ago
Tidy forecasting in R	2.4K	7 months ago
What is R?	2.3K	1 year ago
Production-ready R: Getting started with R and docker	1.9K	7 months ago
Wrangling data in the Tidyverse - Part 1	1.6K	8 months ago
Moving from Prototype to Production - Part 1	1.5K	7 months ago
Tree-Based Machine Learning for Insurance	1.4K	7 months ago
tsibble: Tidy data structures to support exploration and	1.3K	7 months ago



R Studio Community

Books (FREE)



Thank you!