



Rensselaer

why not change the world?®



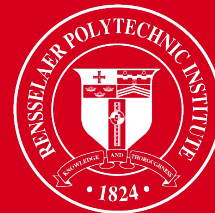
R as your Superpower!

John S. Erickson

RPIrates Monthly Meeting

04 Sep 2024

The Future of Computing Institute at Rensselaer
Rensselaer Polytechnic Institute



FOCI

Future of Computing Institute at Rensselaer

R as your SuperpowerR!

1. The global R Community as your **package** SuperpowerR!
2. RStudio as your **development environment** SuperpowerR!
3. The tidyverse as your **data manipulation** SuperpowerR!
4. ggplot2 as your **visualization** SuperpowerR!
5. R Notebooks as your **communications** SuperpowerR!
6. Shiny as your **app creation** SuperpowerR!
7. The FOCI Cluster as your **computing** SuperpowerR!
8. RPIrates as your **local support** SuperpowerR!

The global R Community as your SuperpowerR!



[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Get Involved: Contributing](#)

[Developer Pages](#)

[R Blog](#)

R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 4.4.1 (Race for Your Life)** has been released on 2024-06-14.
- We are deeply sorry to announce that our friend and colleague Friedrich (Fritz) Leisch has died. [Read our tribute to Fritz here](#).
- **R version 4.4.0 (Puppy Cup)** has been released on 2024-04-24.
- **R version 4.3.3 (Angel Food Cake)** (wrap-up of 4.3.x) was released on 2024-02-29.
- **Registration for useR! 2024** has opened with early bird deadline March 31 2024.
- You can support the R Foundation with a renewable subscription as a [supporting member](#).

News via Mastodon



[latin_conf](#)

Hola, hola! #LatinR llega a mastodon 🐘



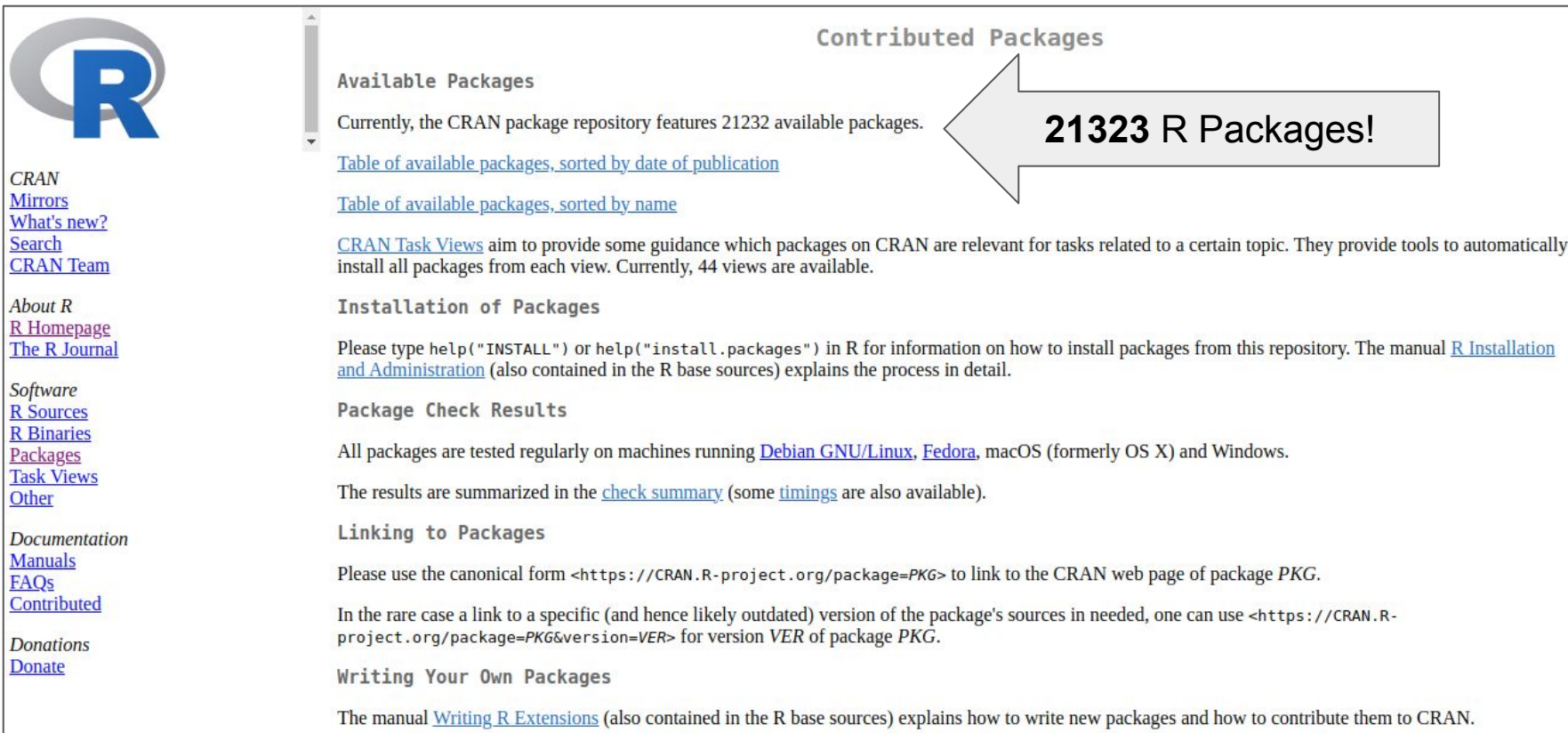
Rensselaer



FOCI

Future of Computing Institute of Rensselaer

The global R Community as your SuperpowerR!



The image shows a screenshot of the CRAN (Comprehensive R Archive Network) website. On the left side, there is a navigation menu with links for CRAN, Mirrors, What's new?, Search, CRAN Team, About R, R Homepage, The R Journal, Software, R Sources, R Binaries, Packages, Task Views, Other, Documentation, Manuals, FAQs, Contributed, Donations, and Donate. The main content area is titled 'Contributed Packages' and includes sections for 'Available Packages' (stating 21232 packages), 'Installation of Packages', 'Package Check Results', 'Linking to Packages', and 'Writing Your Own Packages'. A large grey arrow points from the right towards the 'Available Packages' section, with the text '21323 R Packages!' inside it.

Contributed Packages

Available Packages

Currently, the CRAN package repository features 21232 available packages.

[Table of available packages, sorted by date of publication](#)

[Table of available packages, sorted by name](#)

[CRAN Task Views](#) aim to provide some guidance which packages on CRAN are relevant for tasks related to a certain topic. They provide tools to automatically install all packages from each view. Currently, 44 views are available.

Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this repository. The manual [R Installation and Administration](#) (also contained in the R base sources) explains the process in detail.

Package Check Results

All packages are tested regularly on machines running [Debian GNU/Linux](#), [Fedora](#), macOS (formerly OS X) and Windows.

The results are summarized in the [check summary](#) (some [timings](#) are also available).

Linking to Packages

Please use the canonical form `<https://CRAN.R-project.org/package=PKG>` to link to the CRAN web page of package *PKG*.

In the rare case a link to a specific (and hence likely outdated) version of the package's sources is needed, one can use `<https://CRAN.R-project.org/package=PKG&version=VER>` for version *VER* of package *PKG*.

Writing Your Own Packages

The manual [Writing R Extensions](#) (also contained in the R base sources) explains how to write new packages and how to contribute them to CRAN.

RStudio as your SuperpowerR!

The screenshot displays the RStudio interface with the following components:

- Code Editor:** Contains R code for loading packages, sourcing files, and defining a Shiny application controller.
- Environment Pane:** Shows loaded data objects and their characteristics.
- Files Pane:** Lists files in the current project directory.
- Console:** Shows the execution of `replace_na` and `setwd` commands, along with warnings about masked objects.

```
1 library(shiny)
2 library(visNetwork)
3 library(tidyverse)
4 library(ggplot2)
5 #install.packages("useful")
6 library(useful)
7 library(assertthat)
8 library(lubridate)
9 library(RJSONIO)
10 library(parallel)
11 library(collapsE)
12
13 #source("app-only-auth-twitter.R")
14 source("src/floor.R")
15 source("src/wall.R")
16 source("src/network.R")
17 source("src/external-monitor.R")
18 source("src/utilities.R")
19 source("src/campfire_lib.R") # See edits by JSE [03 Sep 2024]
20
21 campfireApp(
22
23   controller = div(
24     h1("Controller"),
25
26   )
27 )
```

Environment Pane:

Object	Description
data	7405 obs. of 21 variables
edge_colname	List of 3
nodes	12 obs. of 14 variables
nodes_jse	12 obs. of 14 variables
parsed_json	List of 3
ServerValues.nodes	12 obs. of 15 variables

Files Pane:

Name	Size	Modified
..		
app.R	8 KB	Sep 3, 2024, 9:57 AM
ServerValues.nodes.Rds	16 KB	Sep 3, 2024, 9:54 AM
nodes_jse.Rds	15.9 KB	Sep 3, 2024, 8:42 AM
edges2_jse.Rds	597 B	Sep 3, 2024, 8:30 AM
edges_jse.Rds	1 KB	Sep 3, 2024, 8:30 AM
wall.css	480 B	Nov 11, 2021, 8:15 AM
todo	80 B	Jun 11, 2019, 6:33 PM
ttm_output.json	631 B	Jun 11, 2019, 6:33 PM
.gitignore	52 B	Jun 11, 2019, 6:33 PM
app-only-auth-twitter.R	1.5 KB	Jun 11, 2019, 6:33 PM
asd.json	631 B	Jun 11, 2019, 6:33 PM
TwitterNetworkRData.Rproj	205 B	Jun 11, 2019, 6:33 PM
data		
src		
test		

Console:

```
R 4.4.1 - ~/ShinyApps/TwitterTimeMachine/
replace_na

The following object is masked from 'package:plyr':
  join

The following object is masked from 'package:stats':
  D

> setwd("~/ShinyApps/TwitterTimeMachine")
```

The Tidyverse as your SuperpowerR!



R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

The Tidyverse as your SuperpowerR!



ggplot2 is a system for declaratively creating graphics, based on *The Grammar of Graphics*. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.



dplyr provides a *grammar of data manipulation*, providing a consistent set of verbs that solve the most common data manipulation challenges.



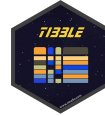
tidyr provides a set of functions that help you get to *tidy data*. Tidy data is data with a consistent form: in brief, every variable goes in a column, and every column is a variable.



readr provides a fast and friendly way to *read rectangular data* (like csv, tsv, and fwf). It is designed to flexibly parse many types of data found in the wild, while still cleanly failing when data unexpectedly changes.



purrr enhances R's functional programming (FP) toolkit by providing a complete and consistent set of tools for working with functions and vectors. Once you master the basic concepts, purrr allows you to replace many for loops with code that is easier to write and more expressive.



tibble is a modern re-imagining of the data frame, keeping what time has proven to be effective, and throwing out what it has not. Tibbles are data.frames that are lazy and surly: they do less and complain more forcing you to confront problems earlier, typically leading to cleaner, more expressive code.



stringr provides a cohesive set of functions designed to make working with strings as easy as possible. It is built on top of stringi, which uses the ICU C library to provide fast, correct implementations of common string manipulations.



forcats provides a suite of useful tools that solve common problems with factors. R uses factors to handle categorical variables, variables that have a fixed and known set of possible values.

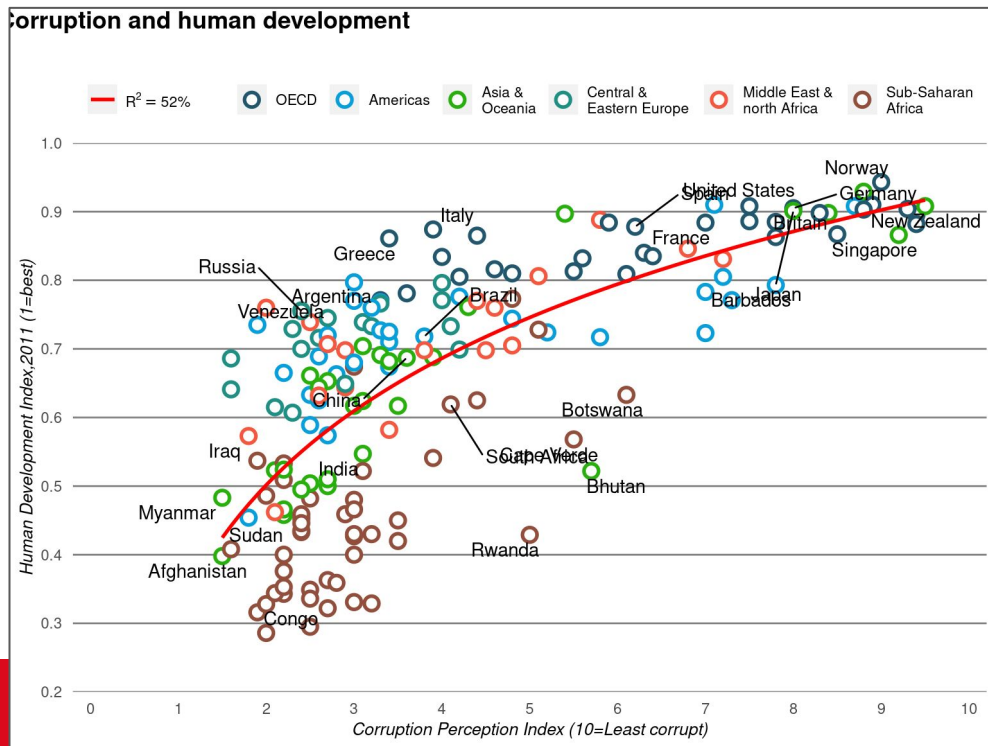


lubridate makes it easier to work with dates and times in R. *As of the latest tidyverse release, lubridate is part of core tidyverse!*

ggplot2 as your SuperpowerR!

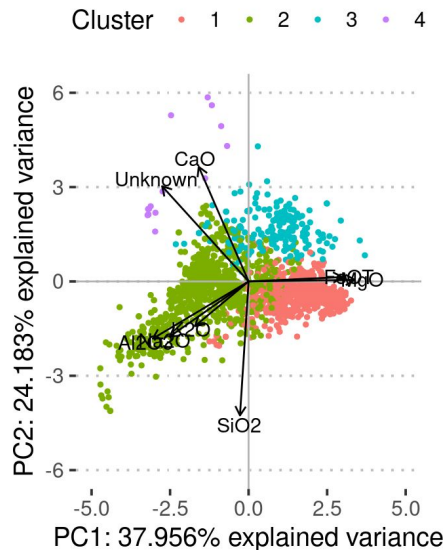
- **ggplot2** is a system for declaratively creating graphics, based on *The Grammar of Graphics*. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.
- It's hard to succinctly describe how ggplot2 works because it embodies a deep philosophy of visualisation. However, in most cases you...
 - ...start with `ggplot()`,
 - ...supply a **dataset** and aesthetic mapping (with `aes()`),
 - ...then add on layers (like `geom_point()` or `geom_histogram()`),
 - ...scales (like `scale_colour_brewer()`),
 - ...faceting specifications (like `facet_wrap()`),
 - ...and coordinate systems (like `coord_flip()`).

Right: A ggplot2-based visualization from "The Economist (ggplot2) Tutorial," John's R Bootcamp, 2018.

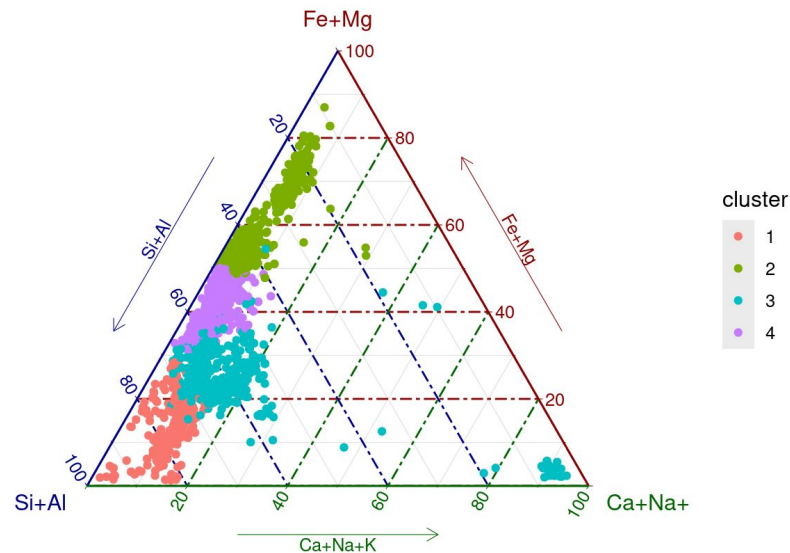


ggplot2 as your SuperpowerR!

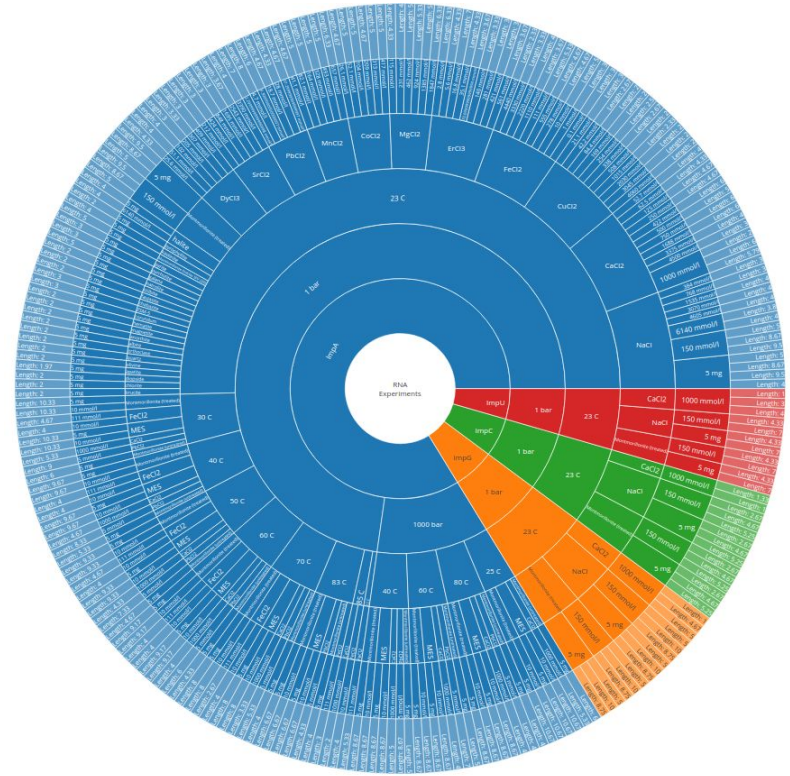
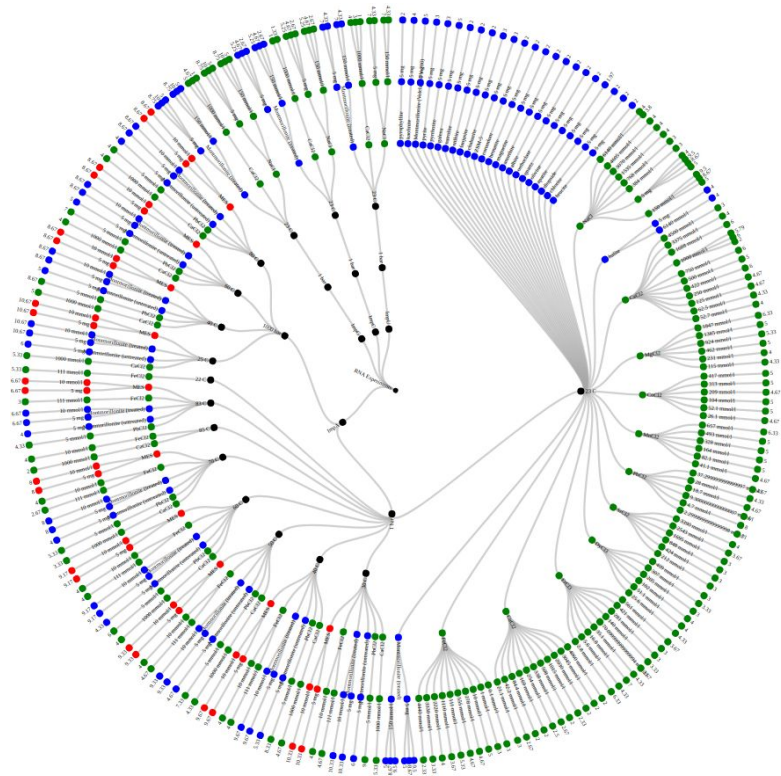
Mars LIBS PCA Biplot



Example of Mars 2020 LIBS ternary Plot

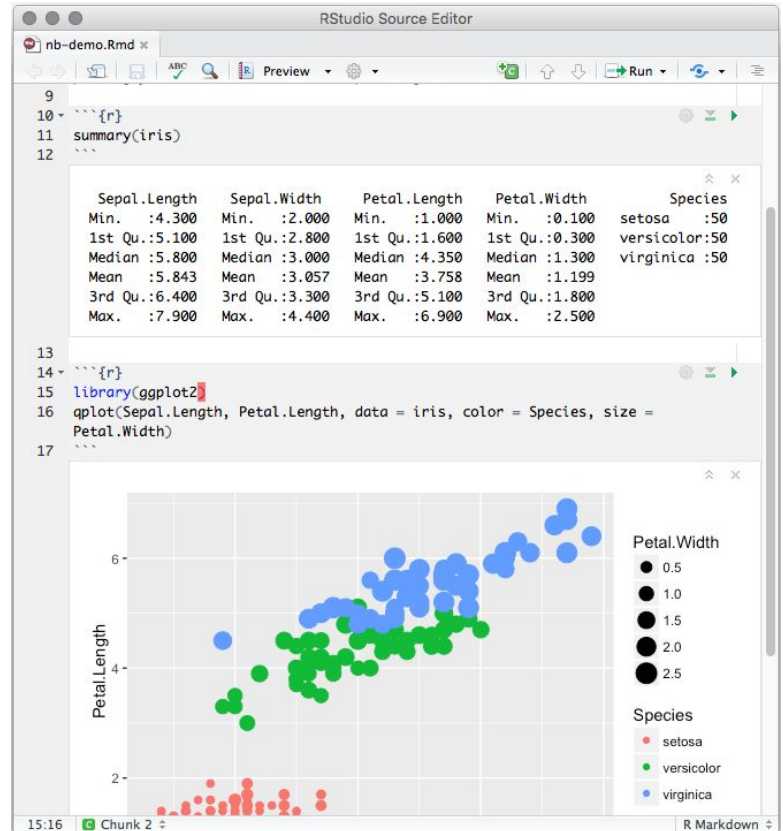


ggplot2 as your SuperpowerR!



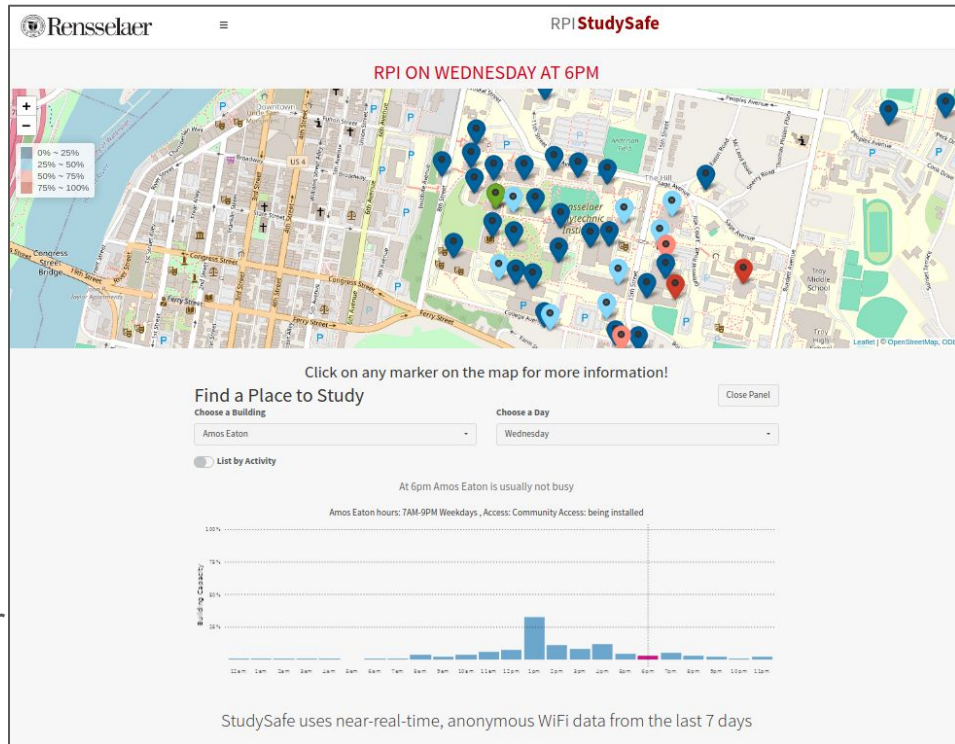
R Notebooks as your SuperpowerR!

- An R Notebook is an **R Markdown document** with chunks that can be executed independently and interactively, with output visible immediately beneath the input.
- R Notebooks are an implementation of **Literate Programming** that allows for direct interaction with R while producing a **reproducible document** with publication-quality output.
- Any R Markdown document can be used as a notebook, and all R Notebooks can be rendered to other R Markdown document types. *A notebook can therefore be thought of as a special execution mode for R Markdown documents.*
- The immediacy of notebook mode makes it a good choice while authoring the R Markdown document and iterating on code. When you are ready to publish the document, you can share the notebook directly, or render it to a publication format with the Knit button.



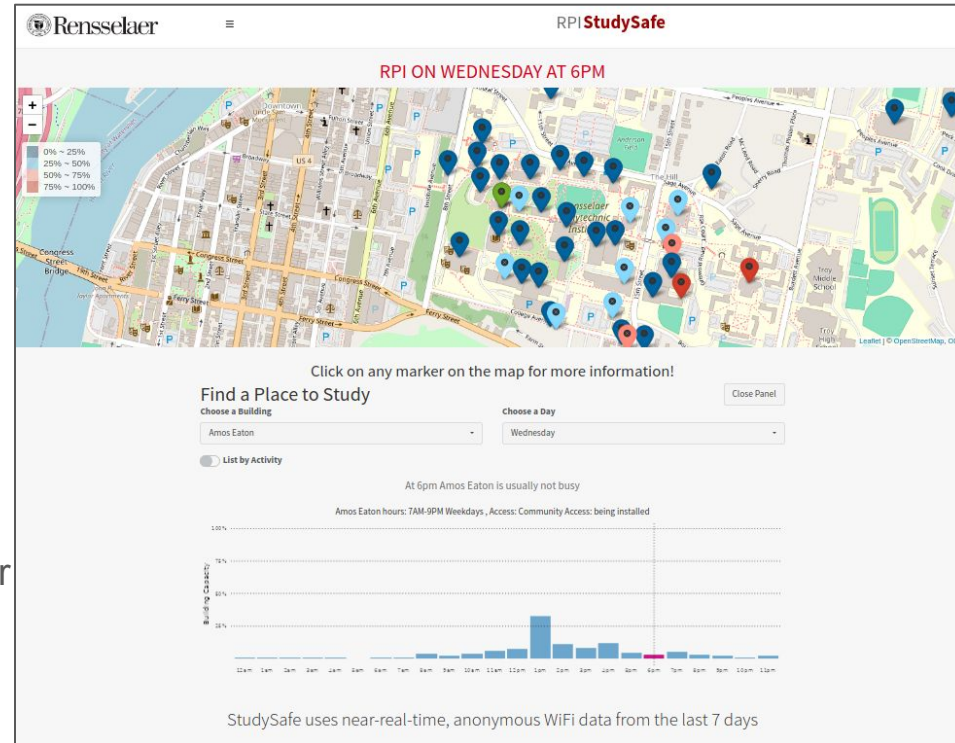
Shiny as your SuperpowerR!

- Shiny is an R package that makes it easy to build interactive web app apps straight from R
- Shiny apps are contained in a single script called app.R.
- app.R has three components:
 - a user interface object
 - a server function
 - a call to the shinyApp function
- The user interface (ui) object controls the layout and appearance of your app; the server function contains the instructions that your computer needs to build your app. Finally the shinyApp function creates Shiny app objects from an explicit UI/server pair.



Shiny as your Superpower!

- Shiny is an R package that makes it easy to build interactive web applications (apps) straight from R.
- Shiny apps are contained in a single script called app.R.
- app.R has three components:
 - a user interface object
 - a server function
 - a call to the shinyApp function
- The user interface (ui) object controls the layout and appearance of your app; the server function contains the instructions that your computer needs to build your app. Finally the shinyApp function creates Shiny app objects from an explicit UI/server pair.



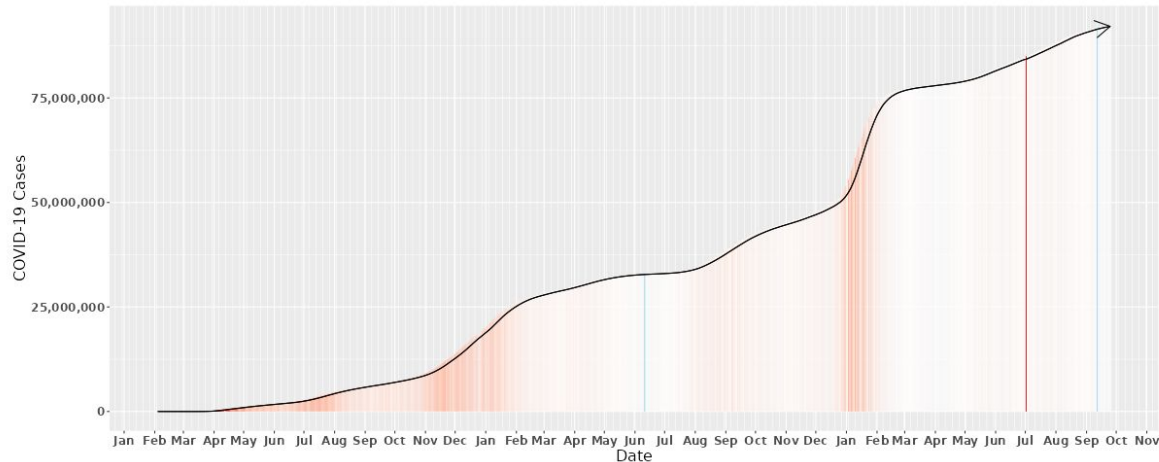
Shiny as your SuperpowerR!

COVIDMINDER: Where you live matters

United States Overview

United States COVID-19 Case Curve

How have United States overall COVID-19 Cases changed over time?



Shiny as your SuperpowerR!

COVIDMINDER: Where you live matters

State Selector

New York

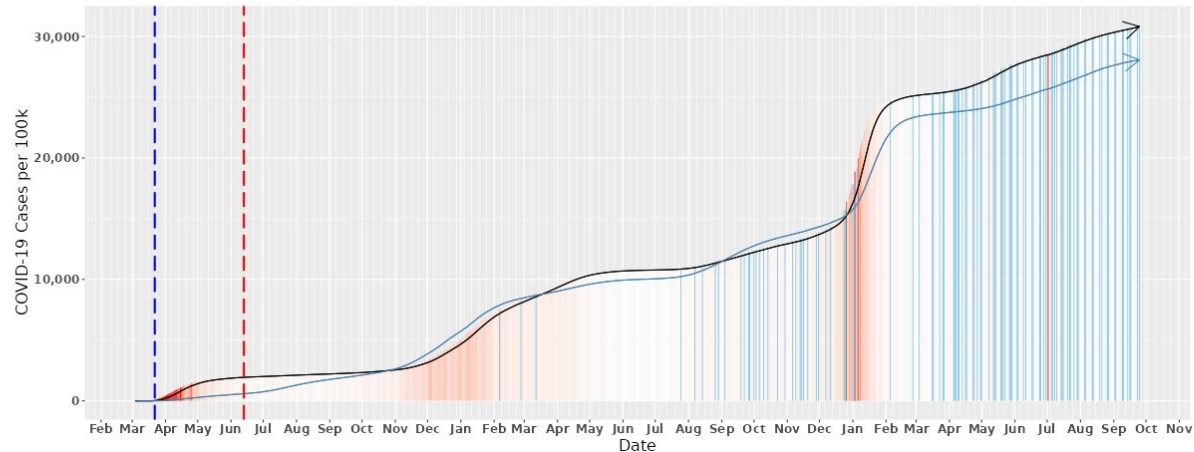
New York Overview

State rank: 40 *

* States are ranked best to worst by their percentage change in COVID-19 cases over the past 7 days.

New York COVID-19 Case Curve

How have New York COVID-19 Cases Over Time per 100k changed compared to US?



Shiny as your SuperpowerR!

MORTALITYMINDER

State: Ohio

Cause of Death: Deaths of Despair

Nationwide View: What are the trends in midlife mortality rates for Deaths of Despair across the United States and in Ohio?

MortalityMinder analyzes trends of premature death in the United States which are caused by:

- All Causes
- Cancer
- Deaths of Despair
- Cardiovascular Disease

MortalityMinder is an interactive presentation that examines county-level factors associated with midlife mortality trends.

Choose **State** and **Cause of Death** on the menu bar at the top of the page (and **Risk Factor** on Factor View page) to see how mortality rates in the selected state and the United States have changed from 2000 to 2017.

Click **BACK** and **NEXT** or the left and right arrow keys to move between the **Nationwide**, **State** and **Factor** views.



Rensselaer

Deaths of Despair Midlife Mortality Rates Over Time

"Deaths of Despair" are deaths due to suicide, overdose, substance abuse and poisonings

2000-02

2003-05

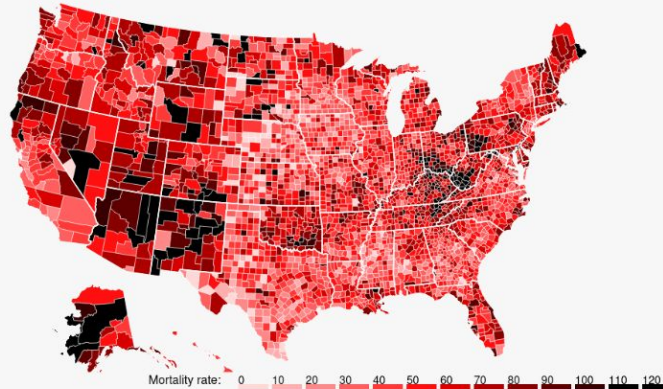
2006-08

2009-11

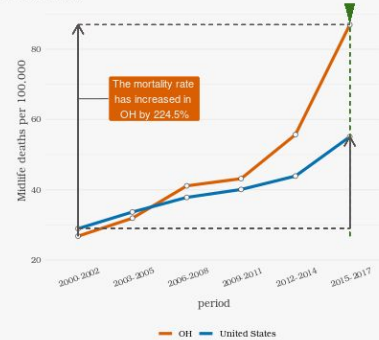
2012-14

2015-17

Nationwide Deaths of Despair Rates for 2015-2017



Midlife Mortality Rates for Deaths of Despair in Ohio vs. United States

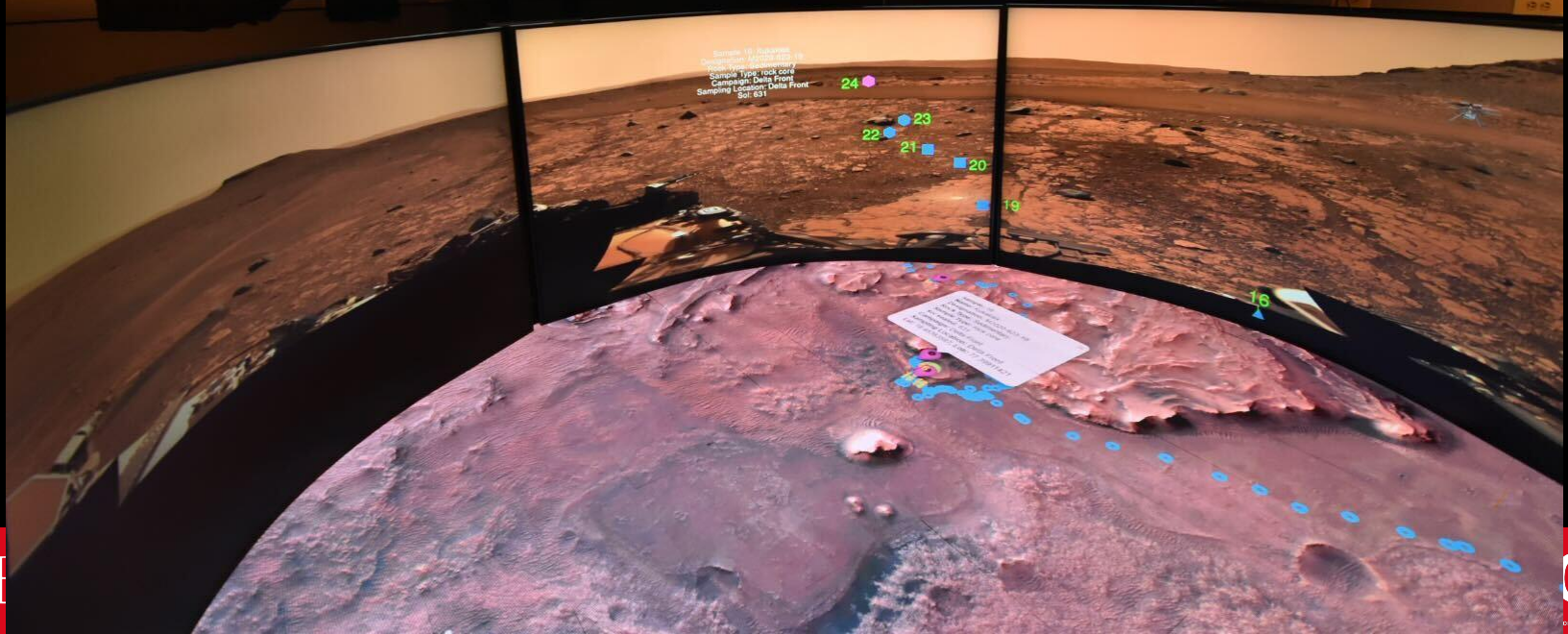
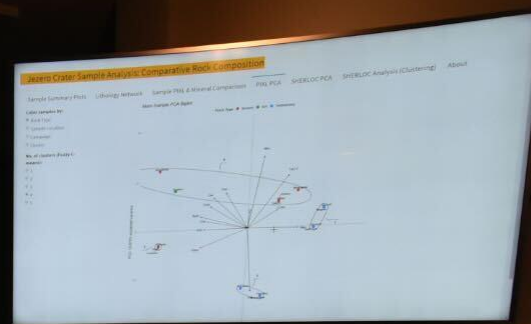


Midlife Mortality Rate: Deaths per 100,000 for adults ages 25-64 due to Deaths of Despair for three year periods for counties (left) and state and nation (right).

Data Source: CDC WONDER

Analysis: The Rensselaer Institute for Data Exploration and Applications ([The Rensselaer IDEA](#))

Shiny as your Superpower!



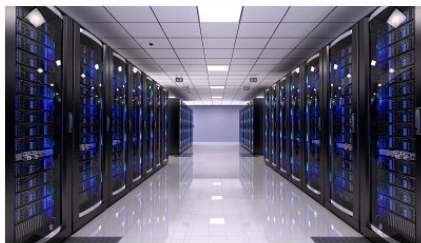
The FOCI Cluster as your SuperpowerR!

Accessing the FOCI Cluster (April 2024)

View Edit Delete Revisions

NOTE: If you're looking to [access RPI's Quantum One system via the FOCI Cluster](#), go [here](#) instead...

- The [Rensselaer FOCI Computing Cluster](#) (formerly the IDEA Cluster) is a [high performance computing environment](#) with five Intel Xeon and two AMD compute servers in various configurations ranging from 24-40 cores (48-80), 256GB-1TB RAM, and up to four GPUs per machine (Tesla K40m or Titan RTX). The IDEA Cluster includes two dedicated storage servers totaling more than 40TB of usable space.
 - RCS credentials and VPN are required
 - Email John Erickson for access privileges: erickj4@rpi.edu
- Student Access:**
 - Students of FOCI/IDEA-sponsored courses (e.g. MATP-4400, MATP-4900) are automatically added for the term of the course. Class privileges end with the end of the term, but home directories are preserved
 - The Cluster provides access to: RStudio, Jupyter, Python, MATLAB, GPUs (on Nodes 2, 5, 6 and 7); lots of storage and computes
 - Access via the RPI physical network or VPN required
 - Priority is given to FOCI researchers and students in FOCI/IDEA-sponsored courses (e.g. MATP-4910, MATP-4400)



- Web links to **RStudio** and **Jupyter** GUIs:
 - RStudio Studio:
 - idea-node-01: <http://lp01.idea.rpi.edu/rstudio-ose/>
 - idea-node-03: <http://lp01.idea.rpi.edu/rstudio-ose-3/>
 - idea-node-06: <https://lp01.idea.rpi.edu/rstudio-ose-6/>
 - idea-node-07: <https://lp01.idea.rpi.edu/rstudio-ose-7/>
 - Jupyter:
 - idea-node-02: <https://lp01.idea.rpi.edu/jupyter> (NEW LINK!)
 - idea-node-05: <https://lp01.idea.rpi.edu/hub/hub/login>
 - idea-node-06: <https://lp01.idea.rpi.edu/hub06/>
 - idea-node-07: <https://lp01.idea.rpi.edu/hub07/>
- GUI access to MATLAB possible via port forwarding; command line recommended! (Contact John Erickson)
- Linux terminal accessible from within RStudio "Terminal" or via ssh (below)
- Shared Data** on the Cluster:
 - All **idea_users** have access to shared storage via **/data** ("data" in your home directories)
 - Permissions are best managed via the Linux terminal (see above)
- Shell access** to individual nodes: You must access "landing pad" first, then a specific compute node:
 - `ssh your_rcs@lp01.idea.rpi.edu`, then `ssh idea-node-XX`
 - For example:
 - `ssh erickj4@lp01.idea.rpi.edu`
 - Then, `ssh idea-node-02` (to access "Node 02," a GPU node)

Help for R users new to Linux & github

- R Tutorial: [Intro to RStudio Server on the IDEA Cluster](#)
- [Navigating the Linux File System via the RStudio Terminal Tab](#)
- [Git usage "cheat sheet"](#)

RPIrates as your SuperpowerR!

RPIrates: The RPI R Users Group

View Edit Delete Revisions

Welcome to **RPIrates: The RPI R Users**

Group! We are a growing community --- over 300 email subscribers! --- of students, faculty and staff at Rensselaer who gather virtually and in person to support each others' use of the **R open source analytics platform**. In semi-

weekly meetings we explore topics of interest, share our "feats of strength," and of course "air grievances!"

During the Fall and Spring semester we have monthly in-person RPIrates meetings, usually the first Weds evening of the month in AE217; certain meetings may also be shared via Webex or. We also have an RPIrates email list through which we share information and ask questions: rpi-r-users@cs.lists.rpi.edu

Fall 2024 Schedule: During the Fall 2024 term we will be holding **monthly** RPIrates meetings at 6p in AE217, with pizza, salad and other goodies!

RPIrates and Data INCITE Tutorials: Watch for our growing list of R tutorials!

- [Intro to RStudio Server on the IDEA Cluster](#) (Jan 2021)
- [Rock your Dataviz World with ggplot2!](#) (Feb 2021)
- [Welcome to the tidyverse!](#) (Feb 2021)
- [Fun with OpenAI, GPTStudio and R!](#) (Feb 2023)
- [Using the RPI github](#) (Jun 2023)
- [What They Forgot to Teach You About R](#) (Feb 2024)

Ask us about RPI R Bootcamps! Occasionally the Rensselaer IDEA will host an **R Bootcamp**. These are two- or three-day, day-long, intensive, interactive deep-dives in the R language and platform, structured to accommodate the needs of the RPI research community.



Amazing R Feats-of-Strength by RPI R Users

- See the [Data INCITE projects page](#) for a growing list and details

Helpful R Resources

- [Accessing the IDEA Compute Cluster](#) (for IDEA and Data INCITE users)
- [R for Data Science](#) (Hadley Wickham, Garrett Grolemond)
 - **The** guide to the [tidyverse](#)
 - The recommended text for IDEA's INCITE courses!
- R Cheatsheets (from R Studio): <https://www.rstudio.com/resources/cheatsheets/>
 - [RStudio Download Cheatsheet](#) (PDF)
 - [Data Visualization via ggplot2 Cheat Sheet](#) (PDF)
 - [R Markdown Reference Guide](#) (PDF)
 - [R Markdown Tutorial](#) (Online)
- R Cheatsheets (from Data INCITE):
 - [Matrix Algebra in R](#) (PDF)
- R Studio Online Learning Resources: <https://www.rstudio.com/online-learning/>
- DataCamp Online Courses: <https://www.datacamp.com/courses/free-introduction-to-r>
- RPI R Bootcamp (github): <https://github.com/TheRensselaerIDEA/r-bootcamp-2018-rpi>
- [Using the RPI github](#): A concise guide to cloning and contributing to repositories on <http://github.rpi.edu>

Sign up for the RPI R Users email list (currently over 350 subscribers)!

- Go to rpi-r-users@cs.lists.rpi.edu