

Getting Started with R

Installation and Libraries



Midcourse Projects

Upcoming projects:

- **Data Question 5: R, tidyverse, and machine learning basics**
 - Start Date - 11/26
 - Presentation Date - 12/17
- **Shiny Mini-Project using data from DQ5**
 - Start Date - 12/19
 - End Date - 12/21
- **Midcourse Project using R Shiny**
 - Start Date: 12/21
 - Presentation Date: 1/18

Midcourse Projects

Your task will be to create a Shiny web app to present your findings.

For the midcourse project, you must choose a dataset and a data question.

Proposal deadline: December 14

Midcourse Projects

Awesome Public Datasets:

<https://github.com/awesomedata/awesome-public-datasets>

Reddit Datasets:

<https://www.reddit.com/r/datasets/>

Datasetlist:

<https://www.datasetlist.com/>

Midcourse Projects

Dataquest Datasets List:

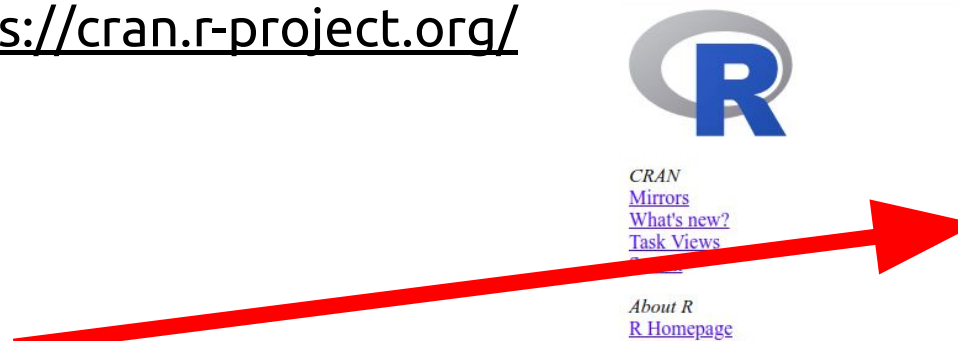
<https://www.dataquest.io/blog/free-datasets-for-projects/>

Data is Plural:

<https://tinyletter.com/data-is-plural/archive> OR

<https://docs.google.com/spreadsheets/d/1wZhPLMCHKJvwOkP4juclhjFgqIY8fQFMemwKL2c64vk/edit#gid=0>

Install R - <https://cran.r-project.org/>



The screenshot shows the CRAN website. A red arrow points from the left towards the 'Download and Install R' section. In this section, three links are highlighted with an orange box: 'Download R for Linux', 'Download R for (Mac) OS X', and 'Download R for Windows'. The text around these links indicates that these are precompiled binary distributions for Windows and Mac users, while Linux users should check with their package management system.

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

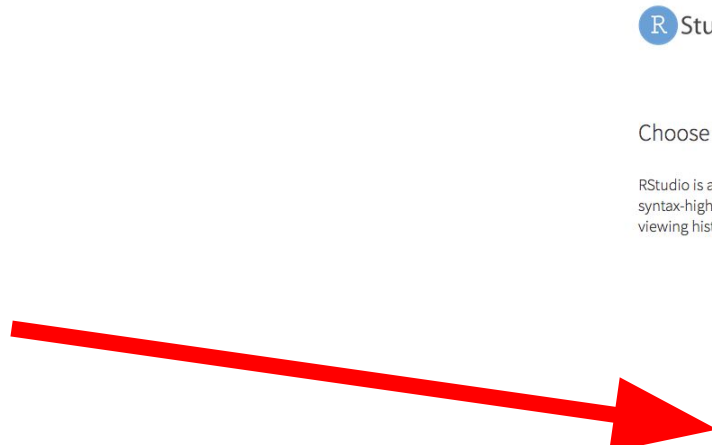
R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2019-07-05, Action of the Toes) [R-3.6.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and](#)

Install RStudio Desktop - <https://www.rstudio.com/products/rstudio/download/>



The screenshot shows the RStudio website's product page. A red arrow points from the left towards the 'DOWNLOAD' buttons for the Desktop versions. The page lists five options: RStudio Desktop Open Source License (FREE), RStudio Desktop Commercial License (\$995 per year), RStudio Server Open Source License (FREE), RStudio Server Pro Commercial License (\$9,995 per year), and RStudio Server Pro + RStudio Connect Commercial License (\$29,995 per year). Each option has a 'DOWNLOAD' or 'BUY' button and a 'Learn More' link. A table at the bottom compares features like 'Integrated Tools for R', 'Priority Support', and 'Access via Web Browser' across the different licenses.

R Studio Products Resources Pricing About Us Blogs

Choose Your Version of RStudio

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. [Learn More about RStudio features.](#)

	RStudio Desktop Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License	RStudio Server Pro + RStudio Connect Commercial License
Price	FREE	\$995 per year	FREE	\$9,995 per year	\$29,995 per year
Action	DOWNLOAD	BUY	DOWNLOAD	DOWNLOAD	TALK
Learn More	Learn More	Learn More	Learn More	Learn More	Learn More
Integrated Tools for R	●	●	●	●	●
Priority Support		●		●	●
Access via Web Browser			●	●	●

Try RStudio Server Pro for free!

**DO NOT INSTALL
RSTUDIO FROM
ANACONDA!**

Libraries in R

Windows users will also need to install Rtools so that they can install new libraries.

<https://cran.r-project.org/bin/windows/Rtools/>

RStudio

Source Pane
Files you are editing appear here.

Console
Code is executed here.

RStudio

Project: (None)

Environment History Connections Presentation

Global Environment

Data

not_states	List of 4
: chr "AA"	
: chr "GU"	
: chr "ZZ"	
: chr "AE"	

Values

x	Factor w/ 2 levels "yes","no": 1 2 1
---	--------------------------------------

Files Plots Packages Help Viewer

Zoom Export Publish

Oxycontin Prescriptions

Gender

- F
- M

```
1 library(tidyverse)
2
3 opioids <- as.data.frame(read_csv('./data/opioids.csv'))
4 overdoses <- as.data.frame(read_csv('./data/overdoses.csv'))
5
6
7 prescribers <- as.data.frame(read_csv('./data/prescriber-info.csv'))
8 not_states <- list('AA', 'GU', 'ZZ', 'AE')
9
10 # remove non-state entries
11 prescribers <- prescribers[!(prescribers$State %in% not_states), ]
12
13 oxycontin <- prescribers %>%
14   group_by(State, Gender) %>%
15   summarize(sum(OXYCONTIN))
16
17
18
19
20
```

19:1 (Top Level) R Script

Console Terminal

~/kaggle/opiates/

```
> View(overdoses)
>
```


Environment/History
Variables you have assigned appear here.


**Files/Plots/
Packages/Help**


Tools > Global Options


Options


R General


 Code


 Appearance


 Pane Layout


 Packages


 R Markdown

 Sweave

 Spelling

 Git/SVN

 Publishing

 Terminal

Default working directory (when not in a project):

~ Browse...

☐ Re-use idle sessions for project links

☐ Restore most recently opened project at startup

☐ Restore previously open source documents at startup

☐ Restore .RData into workspace at startup

Save workspace to .RData on exit: Never

☐ Always save history (even when not saving .RData)

☐ Remove duplicate entries in history

☐ Show .Last.value in environment listing

☒ Use debug error handler only when my code contains errors

☐ Automatically expand tracebacks in error inspector

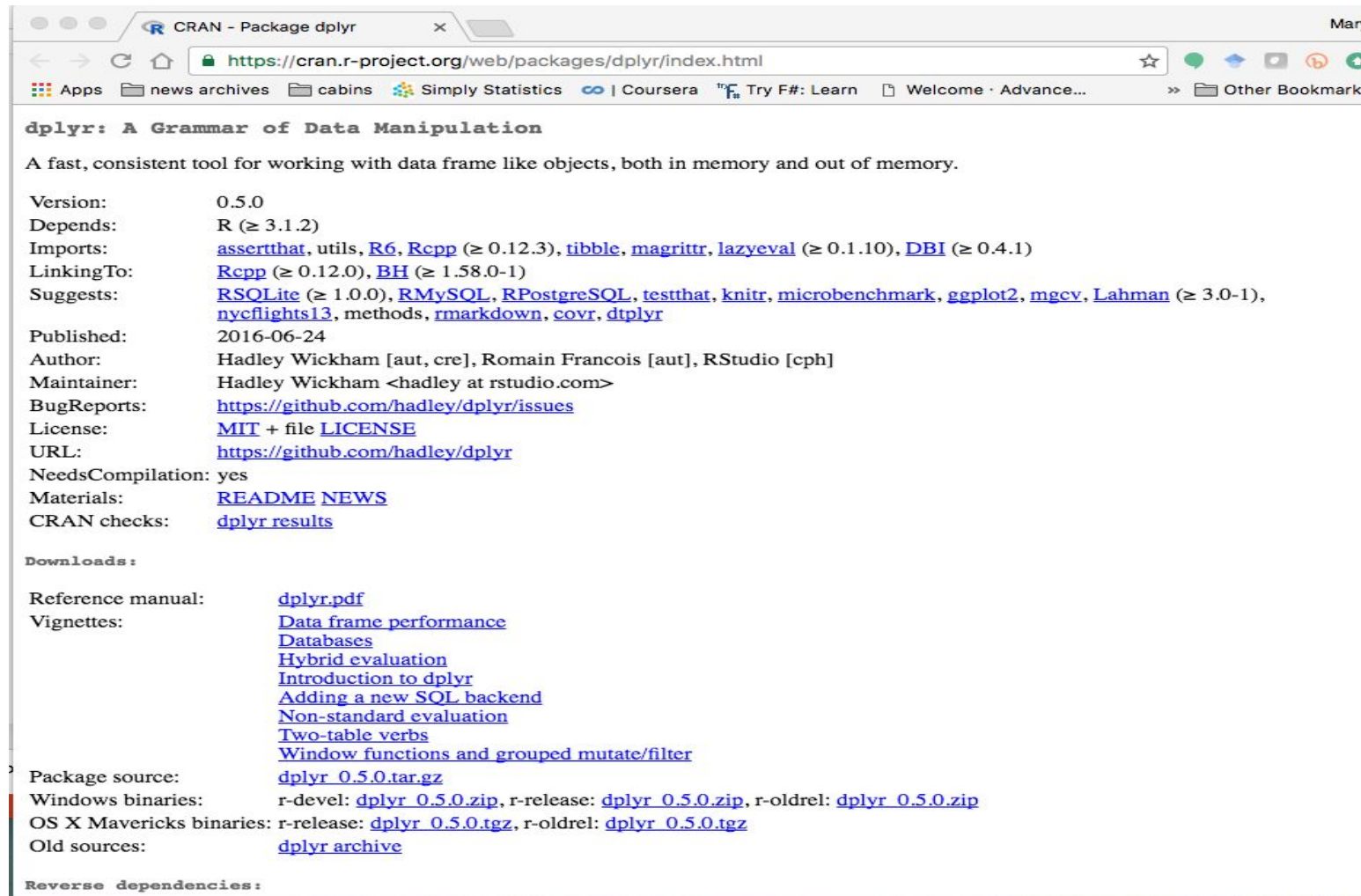
☐ Wrap around when navigating to previous/next tab

☒ Automatically notify me of updates to RStudio

OK Cancel Apply

Using CRAN – the Comprehensive R Archive Network (<https://cran.r-project.org/>)

1. R Manuals
2. Search Packages (date/package-name/title-description)
3. Package Manuals and Vignettes



The screenshot shows a web browser window with the address bar displaying <https://cran.r-project.org/web/packages/dplyr/index.html>. The page title is "dplyr: A Grammar of Data Manipulation". The main content area lists package details: Version (0.5.0), Depends (R ≥ 3.1.2), Imports (assertthat, utils, R6, Rcpp ≥ 0.12.3, tibble, magrittr, lazyeval ≥ 0.1.10, DBI ≥ 0.4.1), LinkingTo (Rcpp ≥ 0.12.0, BH ≥ 1.58.0-1), Suggests (RSQLite ≥ 1.0.0, RMySQL, RPostgreSQL, testthat, knitr, microbenchmark, ggplot2, mgcv, Lahman ≥ 3.0-1, nycflights13, methods, rmarkdown, covr, dtplyr), Published (2016-06-24), Author (Hadley Wickham [aut, cre], Romain Francois [aut], RStudio [cph]), Maintainer (Hadley Wickham <hadley at rstudio.com>), BugReports (https://github.com/hadley/dplyr/issues), License (MIT + file LICENSE), URL (https://github.com/hadley/dplyr), NeedsCompilation (yes), Materials (README NEWS), and CRAN checks (dplyr results). Below this, the "Downloads:" section lists the Reference manual (dplyr.pdf) and Vignettes (Data frame performance, Databases, Hybrid evaluation, Introduction to dplyr, Adding a new SQL backend, Non-standard evaluation, Two-table verbs, Window functions and grouped mutate/filter). The bottom section lists Package source (dplyr_0.5.0.tar.gz), Windows binaries (r-devel: dplyr_0.5.0.zip, r-release: dplyr_0.5.0.zip, r-oldrel: dplyr_0.5.0.zip), OS X Mavericks binaries (r-release: dplyr_0.5.0.tgz, r-oldrel: dplyr_0.5.0.tgz), Old sources (dplyr archive), and Reverse dependencies.

CRAN - Package dplyr

<https://cran.r-project.org/web/packages/dplyr/index.html>

Apps news archives cabins Simply Statistics | Coursera Try F#: Learn Welcome · Advance... Other Bookmarks

dplyr: A Grammar of Data Manipulation

A fast, consistent tool for working with data frame like objects, both in memory and out of memory.

Version: 0.5.0

Depends: R (≥ 3.1.2)

Imports: [assertthat](#), [utils](#), [R6](#), [Rcpp](#) (≥ 0.12.3), [tibble](#), [magrittr](#), [lazyeval](#) (≥ 0.1.10), [DBI](#) (≥ 0.4.1)

LinkingTo: [Rcpp](#) (≥ 0.12.0), [BH](#) (≥ 1.58.0-1)

Suggests: [RSQLite](#) (≥ 1.0.0), [RMySQL](#), [RPostgreSQL](#), [testthat](#), [knitr](#), [microbenchmark](#), [ggplot2](#), [mgcv](#), [Lahman](#) (≥ 3.0-1), [nycflights13](#), [methods](#), [rmarkdown](#), [covr](#), [dtplyr](#)

Published: 2016-06-24

Author: Hadley Wickham [aut, cre], Romain Francois [aut], RStudio [cph]

Maintainer: Hadley Wickham <hadley at rstudio.com>

BugReports: <https://github.com/hadley/dplyr/issues>

License: [MIT](#) + file [LICENSE](#)

URL: <https://github.com/hadley/dplyr>

NeedsCompilation: yes

Materials: [README NEWS](#)

CRAN checks: [dplyr results](#)

Downloads:

Reference manual: [dplyr.pdf](#)

Vignettes: [Data frame performance](#), [Databases](#), [Hybrid evaluation](#), [Introduction to dplyr](#), [Adding a new SQL backend](#), [Non-standard evaluation](#), [Two-table verbs](#), [Window functions and grouped mutate/filter](#)

Package source: [dplyr_0.5.0.tar.gz](#)

Windows binaries: r-devel: [dplyr_0.5.0.zip](#), r-release: [dplyr_0.5.0.zip](#), r-oldrel: [dplyr_0.5.0.zip](#)

OS X Mavericks binaries: r-release: [dplyr_0.5.0.tgz](#), r-oldrel: [dplyr_0.5.0.tgz](#)

Old sources: [dplyr archive](#)

Reverse dependencies:

Libraries in R

To import a library, simply use `library(<libraryname>)`

```
> library(tidyverse)
```

If you need to install a library, you can use `install.packages('<libraryname>')`

```
> install.packages('tidyverse')
```

To update a package, you can again use

```
> install.packages('tidyverse')
```

To see a list of packages that can be updated, use

```
> old.packages()
```

Tasks for Tonight

Install R and RStudio (and Rtools if on Windows)

Install the following libraries:

- glmnet
- ranger
- coefplot
- rpart
- rpart.plot
- censusapi
- tidyverse
- caret

Send a **direct message to me** to nominate at least one of your classmates for recognition for an outstanding job working on Data Question 4. Tell me your nominee and a sentence or two explaining your choice. You can nominate more than one, but you must nominate **at least one!**