Contents

Preface

Chapter 1

Introduction

This is the introduction for this project. Before that, I would like to introduce some terms involved in this project.

- R language : R is a language and environment for statistical computing and graphics which can be extended easily via packages and provide an Open Source route to participation in statistical methodology. It is available as Free Software under the term of the Free Software Foundation's GNU General Public License(gnu, 1991).

- R Studio : It is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

- R Packages : They are collections of functions and data sets developed by the community. They increase the power of R by improving existing base R functionalities, or by adding new ones.

- CRAN : It is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R.

- CRAN mirror : It is a website containing differently located servers, which aims to facilitate people from different regions and countries to access CRAN more smoothly and quickly. And each server is called a mirror.

- API : It is the abbreviation of Application Programming Interface, which is a software intermediary that allows two applications to talk to each other.

1.1   rtrends

And then we come to the related literatures, which are as followed:

- `rtrends` This is a package used for simple analysis of download logs from the CRAN RStudio Mirror, which contains one dataset and five functions. It's mainly functions are:

(1) abstract all CRAN download logs for a specific date
(2) abstract CRAN download logs for a specific package
(3) abstract CRAN download logs for a several package in a range of dates
(4) merge CRAN download logs with county names
(5) provide sample formatted CRAN logs

- `packageRank`. It is also a package that relies on package `cranlog` and extends its functionality.It can publish the total download logs from CRAN with a more user-friendly interface and providing generic R `plot()` methods to make visualization easy. Also, many package creators are very concerned about the popularity of their packages, and this package `packageRank` can show the percentage of packages with less downloads. Furthermore, when we count the download logs, the exact thing we want is the number of times the package has been completely downloaded. In fact, we often encounter some wrongly sized download records, that is to say, users only download a part of the normal sized package. One is called "medium", the other is "small", which exists either as a part of pair with a full download or one of triplet with a "medium" and a full downloads. In order to deal with this, the creator filters the "small" entries that are less 1000 bytes and use triplet-specific filter to do with the "medium" entries.

- `cranlogs` - to download the summary download data. It is a R package used to publish the total download logs from CRAN mirror in last day, last week or even a specific time range, for one package or several packages. And it can used to figure out the top downloaded packages in a time.

- `adjustedcranlogs`. We already know that the main function of `cranlogs` is to show the total number of download logs, but in practice, so the statistics of that will face the following problems: including the number of shared download logs, the same user downloads the same package again for some reason, and when a package is updated, the number of download logs will surge. Due to that, `adjustedcranlogs` is a wrapper around the `cranlogs` package that removes these three types of download logs. Therefore, using the data processed by this package to plot, we will find that the fluctuation of the download logs will be smaller.

- `cran.stats` This package can also be used to get the total number of download logs first, but it is different from the previous package in that it gets seasonal data, such as annual, monthly and weekly. Besides, packages can have dependencies with each other, which means when a package is downloaded, its dependent package is also downloaded. But this will cause a problem, that is, when users download a package with its dependent package also downloaded, the download logs may not reflect the actual

popularity of that dependent package. Therefore, another function of this package can solve this problem well by identifying and subtracting the downloads due to it's dependent packages. Also, there is one function used to plot direct downloads and downloads due to dependencies separately.

- `dlstats`

- `pkgsearch` This is a package used to query packages in multiple dimensions. First, users can search by the usage of the package, and `pkgsearch` will return the results with popular packages before less frequently used ones. Or there is a `CRAN package search` addin in menu, with which users can search packages by clicking easily. And it can publish the version history of certain package or display the recent trending packages in detail. Besides, it can provide the latest releases or archivals. Moreover, the search server uses the stems of the words in the indexed metadata, and the search phrase. In this way, the deceleration result will not be affected by the grammar of spelling and tense, which increases the fault tolerance rate and become more user-friendly. At this point, there is also a functional extension, that is, the search engine prefers matching whole phrases over single words.This means that it will consider more about the meaning of the entire search term than match a word, which makes the results more accurate. For example, the search phrase "permutation test" will rank coin higher than testthat, even though testthat is a much better result for the single word "test". And it uses weighted scoring to show the ranking of the results, which includes reverse dependencies, matching degree of title and description and the number of downloads. Finally, it also provides ASCII folding for easy retrieval. In general, this is a comprehensive, practical and interesting search package.

- `Visualize.CRAN.Downloads` This package provides a variety of methods and different types of presentation to visualize the downloads of packages, including static plots, interactive plots and comparison plot. These visualizations mainly focus on the single or comparative study of package downloads in a certain period of time. It includes the change of daily download volume with time, the change of cumulative total download volume, the fluctuation of average download volume and download volume within the standard deviation. In addition, it also provides many function options for splitting or combining figures to make visualization as comprehensive and flexible as possible.

1.2 Helpful tips

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 1. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environ-

ments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```
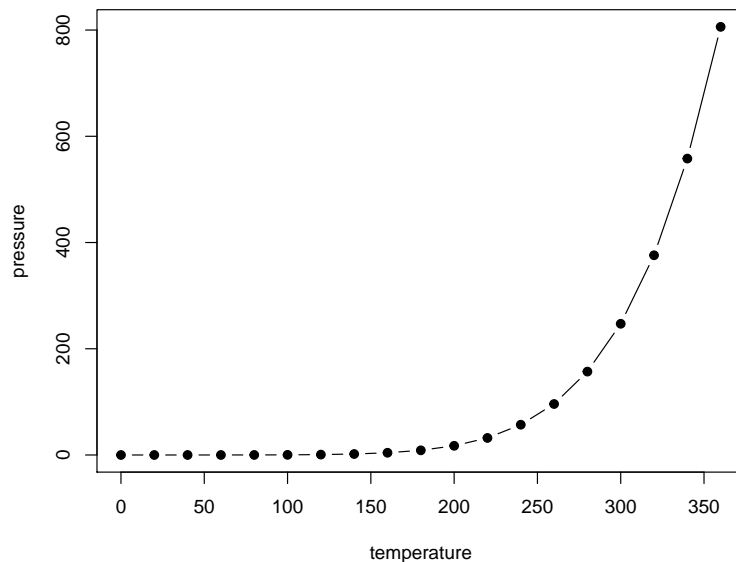

Figure 1.1: Here is a nice figure!

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure
1.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g.,
see Table 1.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package
(Xie, 2020) in this sample book, which was built on top of R Markdown and
knitr (Xie, 2015).

Table 1.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 2

Analysis

Chapter 3

Summary

Bibliography

(1991). The gnu operating system and the free software movement.

Xie, Y. (2015). *Dynamic Documents with R and knitr.* Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown.* R package version 0.21.