A R discussion with Born in Bradford and Bradford NHS

Ning Lu

2019-12-10

Contents

Chapter 1

Mission

This is a collection of R discussion in relation to Born in Bradford and Bradford NHS.

Chapter 2

Public Health API

Fingertips is a web platform that provides easy access to in-depth analysis of a wide range of health and health related data in thematic Profiles. All profiles can be accessed via:

http://fingertips.phe.org.uk/

The project was initially designed and built by the Eastern Region Public Health Observatory, and is now owned by PHE. The platform grows quickly both in functionality and content and is served by one common database called PHOLIO.

Fingertips is used to create the public health outcome framework data tool, and the healthier lives application. It also underpins Health profiles, the mental health intelligence network tools, tobacco control profiles, the NCMP Data tool, NHS Health Check, National General Practice Profiles and the children and young people's benchmarking tool among others.

R package fingertipsR is to interact with Public Health England's Fingertips data tool.

Stable version from CRAN

```
install.packages("fingertipsR")
```

The latest development version from github

```
# install.packages("devtools")
remotes::install_github("rOpenSci/fingertipsR",
                        build_vignettes = TRUE,
                        dependencies = "suggests")
```

This is a workflow example from the R package to demonstrate how to download data for the indicators on Healthy Life Expectancy at Birth from the Public Health Outcomes Framework profile.

The profiles() function presents all of the available profiles:

```
library(fingertipsR)
profs <- profiles()
profs <- profs[grepl("Public Health Outcomes Framework", profs$ProfileName),]
head(profs)
#> # A tibble: 6 x 4
#>   ProfileID ProfileName                  DomainID DomainName
#>       <int> <chr>                           <int> <chr>
#> 1        19 Public Health Outcomes Fr~    1.00e6 Overarching indicators
#> 2        19 Public Health Outcomes Fr~    1.00e6 Wider determinants of hea~
#> 3        19 Public Health Outcomes Fr~    1.00e6 Health improvement
#> 4        19 Public Health Outcomes Fr~    1.00e6 Health protection
#> 5        19 Public Health Outcomes Fr~    1.00e6 Healthcare and premature ~
#> 6        19 Public Health Outcomes Fr~    1.94e9 Supporting information
```

This table shows that the ProfileID for the Public Health Outcomes Framework is 19. This can be used as an input for the indicators() function:

```
inds <- indicators(ProfileID = 19)
print(inds[grepl("Healthy", inds$IndicatorName), c("IndicatorID", "IndicatorName")])
# A tibble: 1 x 2
  IndicatorID IndicatorName
        <int> <fct>
1       90362 A01a - Healthy life expectancy at birth
```

Healthy Life Expectancy at Birth has the IndicatorID equal to 90362.

Finally, the data can be extracted using the fingertips_data() function using that IndicatorID and filter() function in dplyr.

```
df <- fingertips_data(IndicatorID = 90362)
head(df %>% filter(AreaName == "Yorkshire and the Humber region"))
  IndicatorID                    IndicatorName ParentCode ParentName  AreaCode
1       90362 Healthy life expectancy at birth  E92000001    England E12000003
2       90362 Healthy life expectancy at birth  E92000001    England E12000003
3       90362 Healthy life expectancy at birth  E92000001    England E12000003
4       90362 Healthy life expectancy at birth  E92000001    England E12000003
5       90362 Healthy life expectancy at birth  E92000001    England E12000003
6       90362 Healthy life expectancy at birth  E92000001    England E12000003
                         AreaName AreaType    Sex      Age CategoryType Category Timepe
1 Yorkshire and the Humber region   Region   Male All ages         <NA>     <NA>  2009
```

```
2 Yorkshire and the Humber region   Region Female All ages          <NA>      <NA>  2009 - 11
3 Yorkshire and the Humber region   Region   Male All ages          <NA>      <NA>  2010 - 12
4 Yorkshire and the Humber region   Region Female All ages          <NA>      <NA>  2010 - 12
5 Yorkshire and the Humber region   Region   Male All ages          <NA>      <NA>  2011 - 13
6 Yorkshire and the Humber region   Region Female All ages          <NA>      <NA>  2011 - 13
     Value LowerCI95.0limit UpperCI95.0limit LowerCI99.8limit UpperCI99.8limit Count
1 60.84033         60.38649         61.29417               NA               NA    NA
2 61.97605         61.51676         62.43533               NA               NA    NA
3 60.90318         60.44972         61.35665               NA               NA    NA
4 61.78648         61.30470         62.26827               NA               NA    NA
5 60.95582         60.47015         61.44150               NA               NA    NA
6 61.60701         61.10430         62.10971               NA               NA    NA
  Denominator Valuenote RecentTrend ComparedtoEnglandvalueorpercentiles
1          NA      <NA>        <NA>                                Worse
2          NA      <NA>        <NA>                                Worse
3          NA      <NA>        <NA>                                Worse
4          NA      <NA>        <NA>                                Worse
5          NA      <NA>        <NA>                                Worse
6          NA      <NA>        <NA>                                Worse
  ComparedtoRegionvalueorpercentiles TimeperiodSortable Newdata Comparedtogoal
1                       Not compared           20090000    <NA>          <NA>
2                       Not compared           20090000    <NA>          <NA>
3                       Not compared           20100000    <NA>          <NA>
4                       Not compared           20100000    <NA>          <NA>
5                       Not compared           20110000    <NA>          <NA>
6                       Not compared           20110000    <NA>          <NA>
```

Chapter 3

Mari Kondo

Marie Kondo is the author of The Life-Changing Magic of Tidying Up: The Japanese Art of Decluttering and Organizing.

> The best way to choose what to keep and what to throw away is to take each item in one's hand and ask: "Does this spark joy?" If it does, keep it. If not, dispose of it. This is not only the simplest but also the most accurate yardstick by which to judge.

3.1 Tidy your files

There're a number of ways to organise data science work. The key is to set up Rproject and GitHub. Here we list two main approaches to achieve this end.

The first way is the **pull** way where we get both Rproject and git integrated from outside - GitHub. You use the `github` function from `usethis` package and put down ("OWNER/REPO_NAME") and opt for https when you get asked on git protocol.

```
> usethis::create_from_github("dataning/learn_usethis")

Which git protocol to use? (enter 0 to exit)

1: ssh   <-- presumes that you have set up ssh keys
2: https <-- choose this if you don't have ssh keys (or don't know if you do)

Selection: 2
 Tip: To suppress this menu in future, put
  `options(usethis.protocol = "https")`
```

```
  in your script or in a user- or project-level startup file, '.Rprofile'.
  Call `usethis::edit_r_profile()` to open it for editing.
 Cloning repo from 'https://github.com/dataning/learn_usethis.git' into '/Users/sushic
 Setting active project to '/Users/sushicat/Desktop/learn_usethis'
 Writing 'learn_usethis.Rproj'
 Adding '.Rproj.user' to '.gitignore'
 Opening '/Users/sushicat/Desktop/learn_usethis/' in new RStudio session
 Setting active project to '<no active project>'
```

The second way is to imagine you're working in a random folder and you wish to set up the Rproject

```
> library(usethis)
> library(here)

here() starts at /Users/sushicat/Dropbox/R_Me
```

```
> here::here()
[1] "/Users/sushicat/Dropbox/R_Me"
```

```
> path <- file.path(here(), "learn_usethis")
create_project(path)

 Creating '/Users/sushicat/Dropbox/R_Me/learn_usethis/'
 Setting active project to '/Users/sushicat/Dropbox/R_Me/learn_usethis'
 Creating 'R/'
 Writing 'learn_usethis.Rproj'
 Adding '.Rproj.user' to '.gitignore'
 Opening '/Users/sushicat/Dropbox/R_Me/learn_usethis/' in new RStudio session
 Setting active project to '<no active project>'
```

Chapter 4

Care Quality API

We describe our methods in this chapter.

Chapter 5

Road Accidents

5.1 Google data

5.2 DVLA data

Chapter 6

Alfred Pennyworth

Alfred Pennyworth is Bruce Wayne (Batman)'s loyal and tireless butler, housekeeper, legal guardian, best friend, aide-de-camp.

> What was the point of all those push-ups if you can't even lift up a bloody log?

6.1 Log your work

First, choose *New Project* and *New Directiory*.

Second, choose *Book Project using bookdown* and pick a name as well as your preferred directory. RStudio will automatically set up the Rproj as well as the folder skeleton.

Third, tie the existing project with Git through the `usethis` package. It will re-organise the existing project folder and prepare the Git integration.

```
> usethis::use_git()

  Setting active project to '/Users/sushicat/Dropbox/R_Me/R_DE'
  Initialising Git repo
  Adding '.Rhistory', '.RData', '.Rproj.user' to '.gitignore'
There are 15 uncommitted files:
* '_bookdown.yml'
* '_output.yml'
* '.gitignore'
* '01-intro.Rmd'
* '02-literature.Rmd'
```

```
* '03-method.Rmd'
* '04-application.Rmd'
* '05-summary.Rmd'
* '06-references.Rmd'
* 'book.bib'
* 'index.Rmd'
* 'preamble.tex'
* 'R_DE.Rproj'
* 'README.md'
* 'style.css'
Is it ok to commit them?

1: Not now
2: For sure
3: No way

Selection: 2
 Adding files
 Commit with message 'Initial commit'
 A restart of RStudio is required to activate the Git pane
Restart now?

1: Not now
2: Yup
3: Absolutely not

Selection: 2
```

Fourth, create a GitHub repo through the `usethis` package and if the project name is available on the owner's repos. When facing git protocol, choose `https`.

```
> usethis::use_github()

 Setting active project to '/Users/sushicat/Dropbox/R_Me/R_DE'
 Checking that current branch is 'master'
Which git protocol to use? (enter 0 to exit)

1: ssh   <-- presumes that you have set up ssh keys
2: https <-- choose this if you don't have ssh keys (or don't know if you do)

Selection: 2
 Tip: To suppress this menu in future, put
  `options(usethis.protocol = "https")`
  in your script or in a user- or project-level startup file, '.Rprofile'.
  Call `usethis::edit_r_profile()` to open it for editing.
```

```
  Check title and description
   Name:        Bradford
   Description:
Are title and description ok?

1: Yeah
2: Not now
3: Absolutely not

Selection: 1
  Creating GitHub repository
  Setting remote 'origin' to 'https://github.com/dataning/R_DE.git'
  Pushing 'master' branch to GitHub and setting remote tracking branch
  Opening URL 'https://github.com/dataning/R_DE'
```

Fifth, create and save a random R.script in the current project. The commit and push the change of the project to your GitHub repo. You can go to your GitHub repo and check if the R script has been added. This should tell you whether your Rproj and GitHub Repo are fully synced/integrated.

Sixth, go to Netlify and deploy your GitHub repo on Netlify. This will give you the ability to perform continuous deployment as well as deployment to custom domain.

Type in your Rpoj's GitHub repo name.

You need to put down `_book` in *Publish directory*.

Chapter 7

Elastigirl

Elastigirl, also known as Mrs. Incredible, is a fictional character who appears in The Incredibles.

> When designing the Incredible family, Brad Bird wanted each of their superpowers to be related to their personality. He felt that as a mother, Helen was required by society to be pulled in many different directions, which led to her being given an elastic ability.

The same we can say to all sort of data science projects. We are always required by different stakeholders to be pull in many different directions. For us, we have to nail down where we are and how to initiate a new project first.

7.1 Tiki-taka your workflow

First of all, we find where we stand.

```
> here::here()

[1] "/Users/sushicat/Dropbox/R_Me/Hero_book"
```

Second, we find out what we are being surrounded.

```
> fs::dir_ls()

01-think.Rmd           02-pm.Rmd              03-load-data.Rmd       04-tidy-data.Rmd
05-bayesian.Rmd        06-Elastigirl_1.Rmd 06_Elastigirl_1.R   20-references.Rmd
CreditCard             Creditcard_hack.R      Data                   Hero_book.Rproj
```

```
Hero_book.log          README.md              _book                  _bookdown.yml
_bookdown_files        _output.yml            book.bib               index.Rmd
packages.bib           preamble.tex           style.css
```

Third, we pick somewhere (in this case - the data folder) to explore further.

```
> fs::dir_ls("Data")
> fs::dir_ls("Data/Subway_delays")

Data/Subway_delays/Subway&SRT_Logs_April_2018.xlsx
Data/Subway_delays/Subway&SRT_Logs_February_2018.xlsx
Data/Subway_delays/Subway&SRT_Logs_March_2018.xlsx
Data/Subway_delays/Subway&SRT_Logs_May_2018.xlsx
Data/Subway_delays/Subway_&_SRT_Logs_(August_2018).xlsx
Data/Subway_delays/Subway_&_SRT_Logs_(September_2018).xlsx
Data/Subway_delays/Subway_&_SRT_Logs_December_2018.xlsx
Data/Subway_delays/Subway_&_SRT_Logs_November_2018.xlsx
Data/Subway_delays/Subway_SRT_Logs(January 2018).xlsx
Data/Subway_delays/Subway_SRT_Logs(July_2018).xlsx
Data/Subway_delays/Subway_SRT_Logs(June2018).xlsx
Data/Subway_delays/Subway_SRT_Logs(October 2018).xlsx
```

Alternatively, we can use the tree strcuture to show the folder.

```
> fs::dir_tree("Data/Subway_delays")

Data/Subway_delays
   Subway&SRT_Logs_April_2018.xlsx
   Subway&SRT_Logs_February_2018.xlsx
   Subway&SRT_Logs_March_2018.xlsx
   Subway&SRT_Logs_May_2018.xlsx
   Subway_&_SRT_Logs_(August_2018).xlsx
   Subway_&_SRT_Logs_(September_2018).xlsx
   Subway_&_SRT_Logs_December_2018.xlsx
   Subway_&_SRT_Logs_November_2018.xlsx
   Subway_SRT_Logs(January 2018).xlsx
   Subway_SRT_Logs(July_2018).xlsx
   Subway_SRT_Logs(June2018).xlsx
   Subway_SRT_Logs(October 2018).xlsx
```

Fourth, we make a shortcut if this is where we'd like to use or come back later.

```
> fs::dir_tree(here::here("Data", "Subway_delays"))

/Users/goal/Dropbox/R_Me/Hero_book/Data/Subway_delays
   Subway&SRT_Logs_April_2018.xlsx
   Subway&SRT_Logs_February_2018.xlsx
   Subway&SRT_Logs_March_2018.xlsx
   Subway&SRT_Logs_May_2018.xlsx
   Subway_&_SRT_Logs_(August_2018).xlsx
   Subway_&_SRT_Logs_(September_2018).xlsx
   Subway_&_SRT_Logs_December_2018.xlsx
   Subway_&_SRT_Logs_November_2018.xlsx
   Subway_SRT_Logs(January 2018).xlsx
   Subway_SRT_Logs(July_2018).xlsx
   Subway_SRT_Logs(June2018).xlsx
   Subway_SRT_Logs(October 2018).xlsx
```

Let's chain everything together. We present the folder with the dataset - it's like placing the meat and veggie into an oven tray. We then put the tray to an oven called `purrr` and it would import all the spreadsheet files in this particular folder - it's like an oven. Finally, we use the cleaning wipe from `janitor` and clean up the the column names - the ambiguity bit.

```
delays_clean <- fs::dir_ls(here::here("Data", "Subway_delays")) %>%
  purrr::map_dfr(readxl::read_excel) %>%
  janitor::clean_names()
```

Chapter 8

Final Words

We have finished a nice book.