

How to better structure code using own R packages to analyse advertising campaigns

Núria Duran Adroher
Data scientist at Signifikant

Lucerne, November 25, 2020

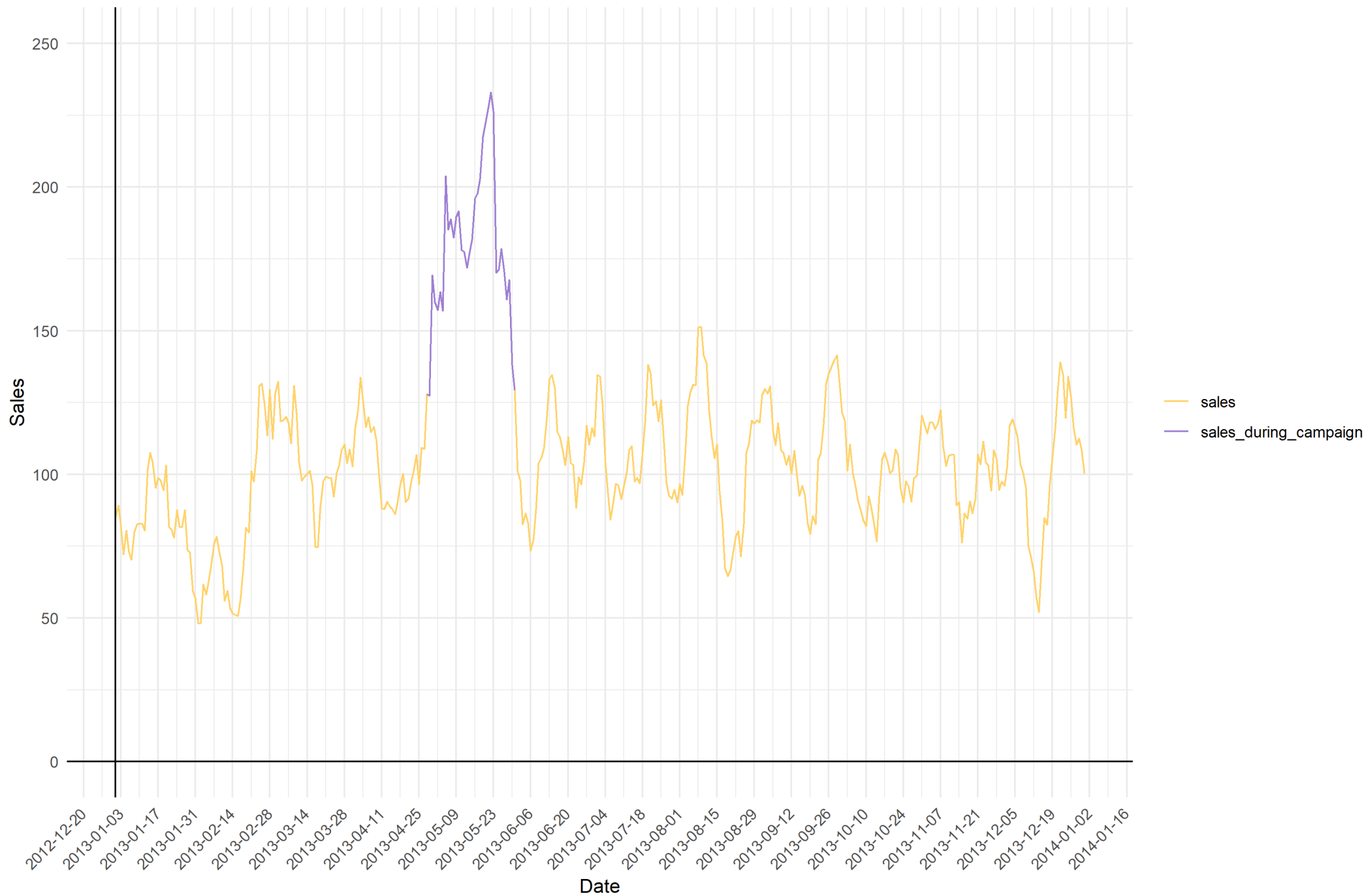
- Signifikant is a startup located at Technopark Luzern (Root D4)
- We analyse campaigns launched by customers from different fields (consumer goods, telecommunication, online shopping, banking...)
- Using data-driven algorithms, we answer questions such as:
 - How successful was the campaign?
 - How did the different media channels (online, tv, print, ...) perform in the campaign?
 - ...



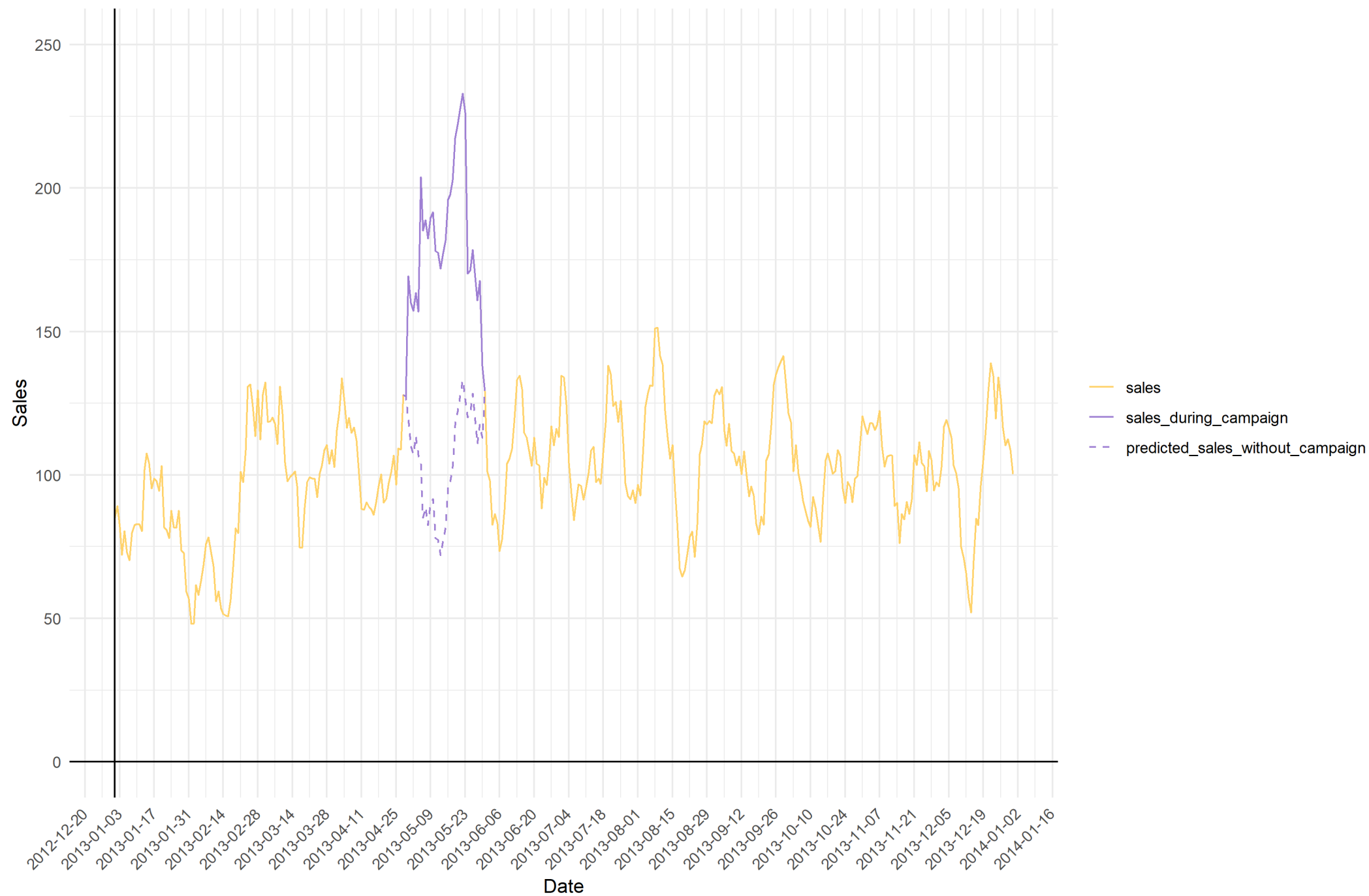
Example of a print campaign



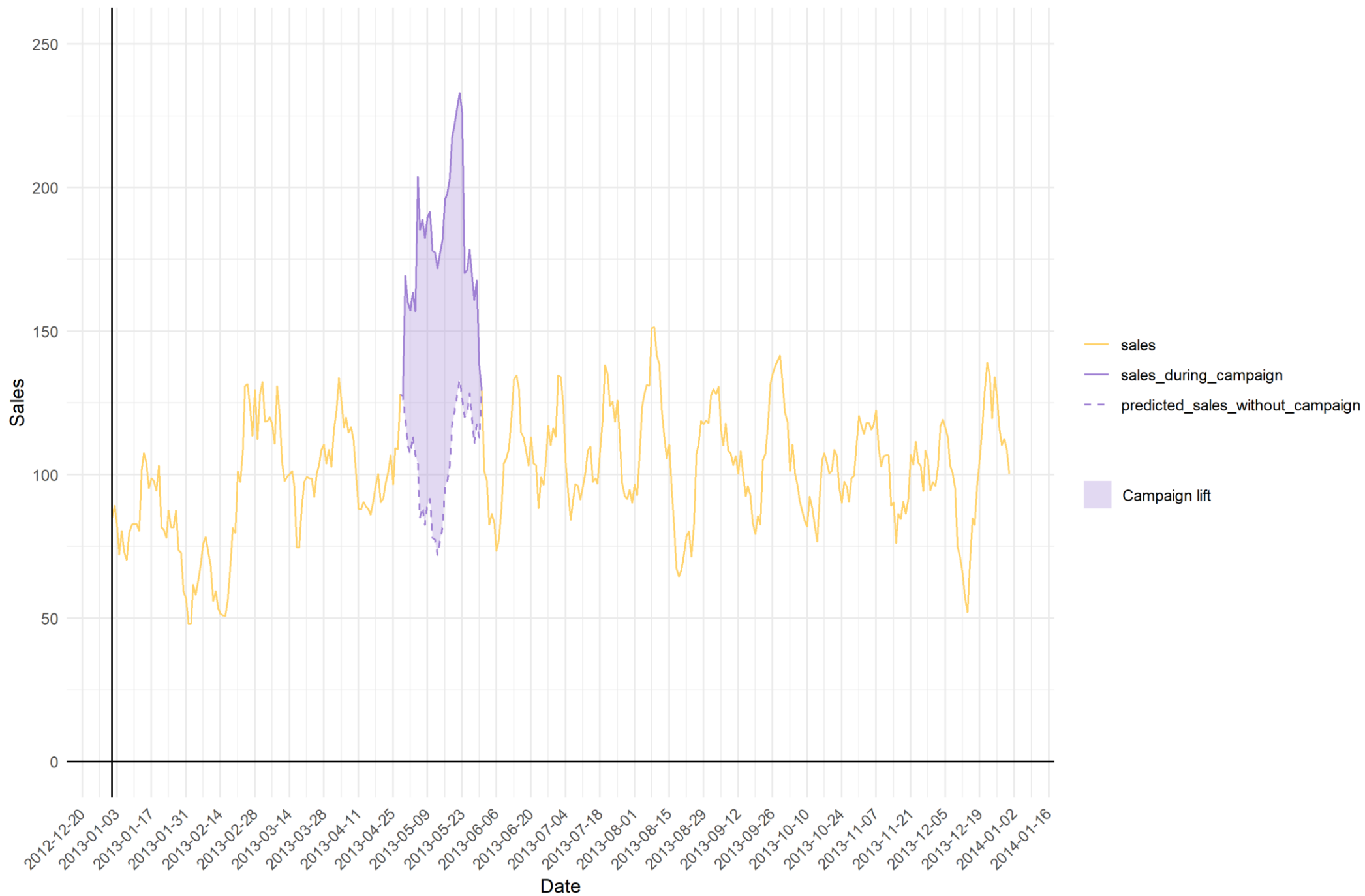
Daily sales of glasses (Dummy data)



Daily sales of glasses (Dummy data)

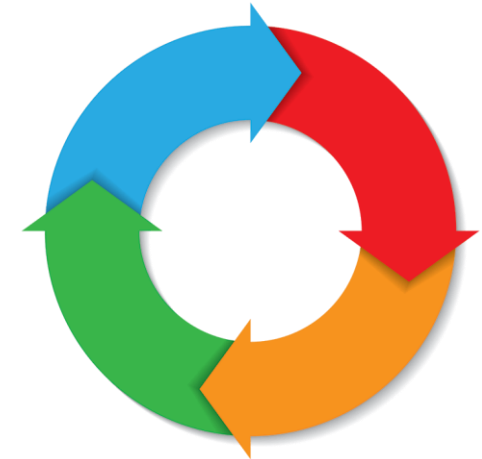


Daily sales of glasses (Dummy data)



Data analysis steps

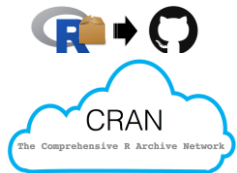
- Load and merge the different datasets
 - Sales
 - Media channels (online, TV, print, ...)
 - Weather
 - Competition
 - ...
- Predict sales in the hypothetical situation that the campaign had not taken place
- Calculate the performance of the media channels
- Transform the results to present them in a dashboard



Developing own R packages to structure data analysis

<https://r-pkgs.org/index.html>

WHY

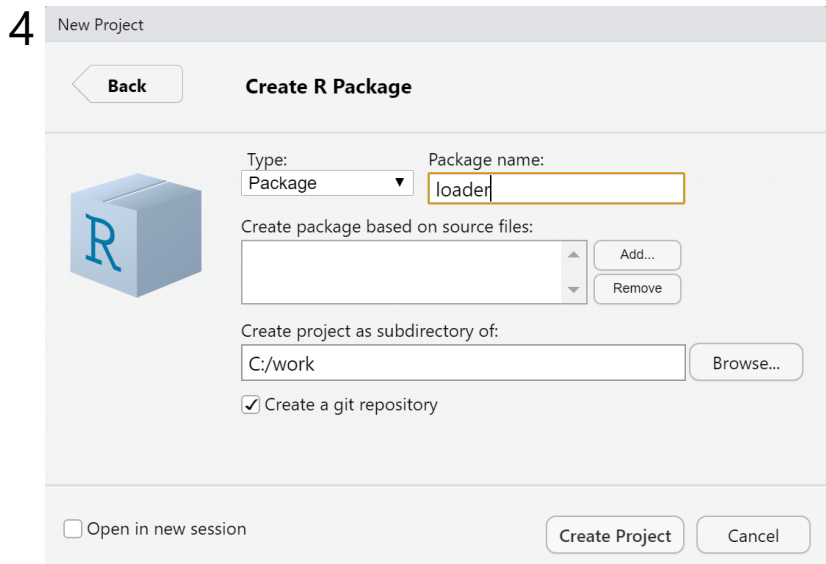
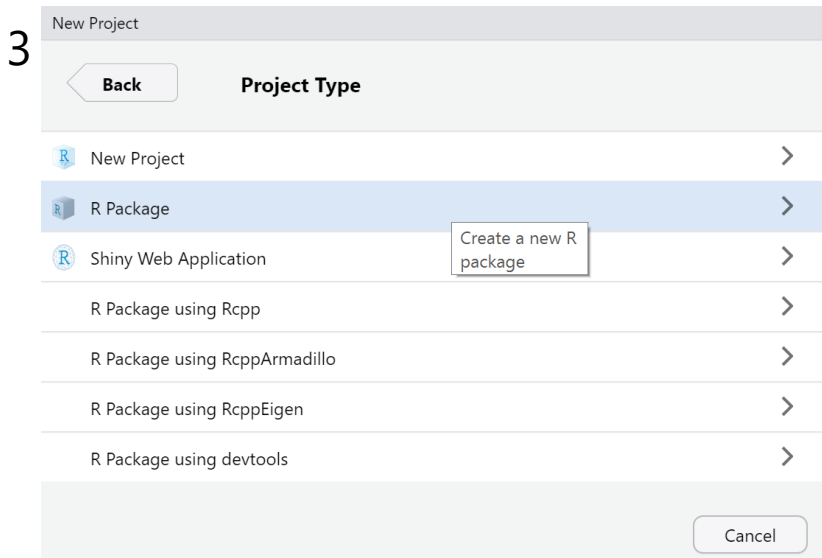
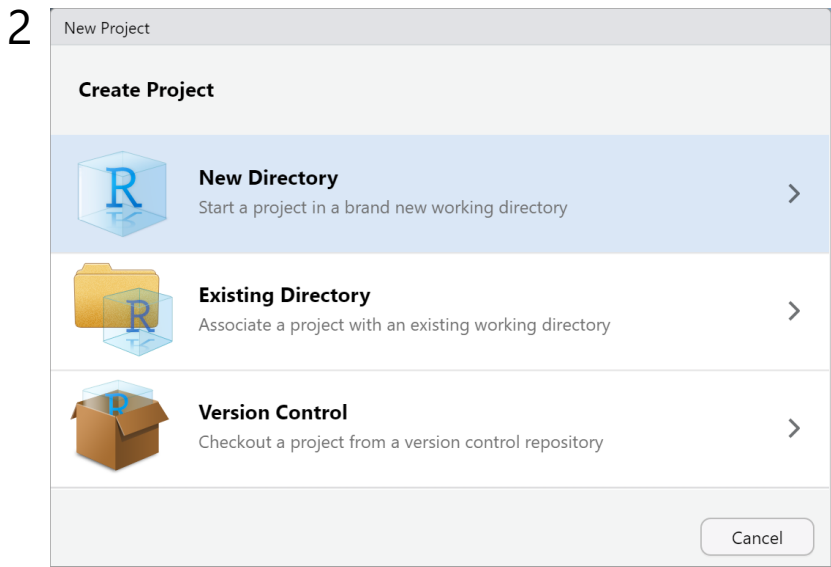
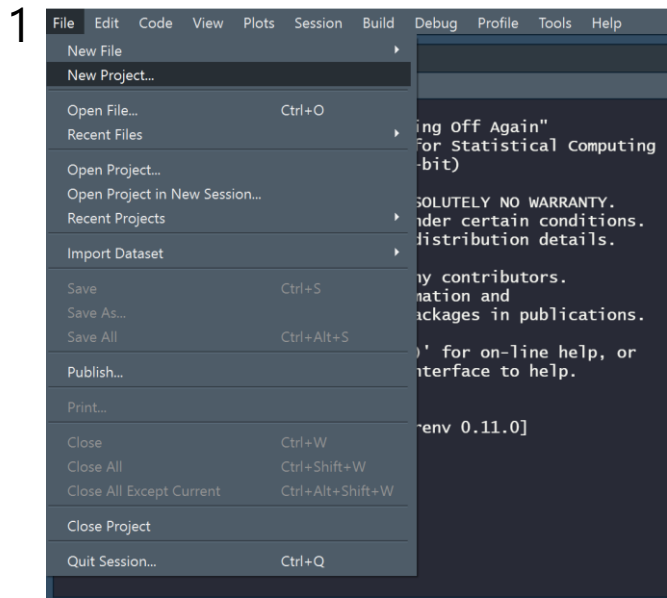


- Using existing R packages is very useful: it's nice to use functions from other people instead of having to code them by yourself
- Sharing code via R packages
 - R users can install it and learn how to use it
- Even if one does not want to share code, organising code in packages makes life easier
 - Reproducibility¹
 - Automation
 - It dramatically improves usability and visibility for your work

¹ Ben Marwick, Carl Boettiger & Lincoln Mullen (2018) Packaging Data Analytical Work Reproducibly Using R (and Friends), The American Statistician, 72:1, 80-88, DOI: 10.1080/00031305.2017.1375986

HOW

Developing own R packages to structure data analysis

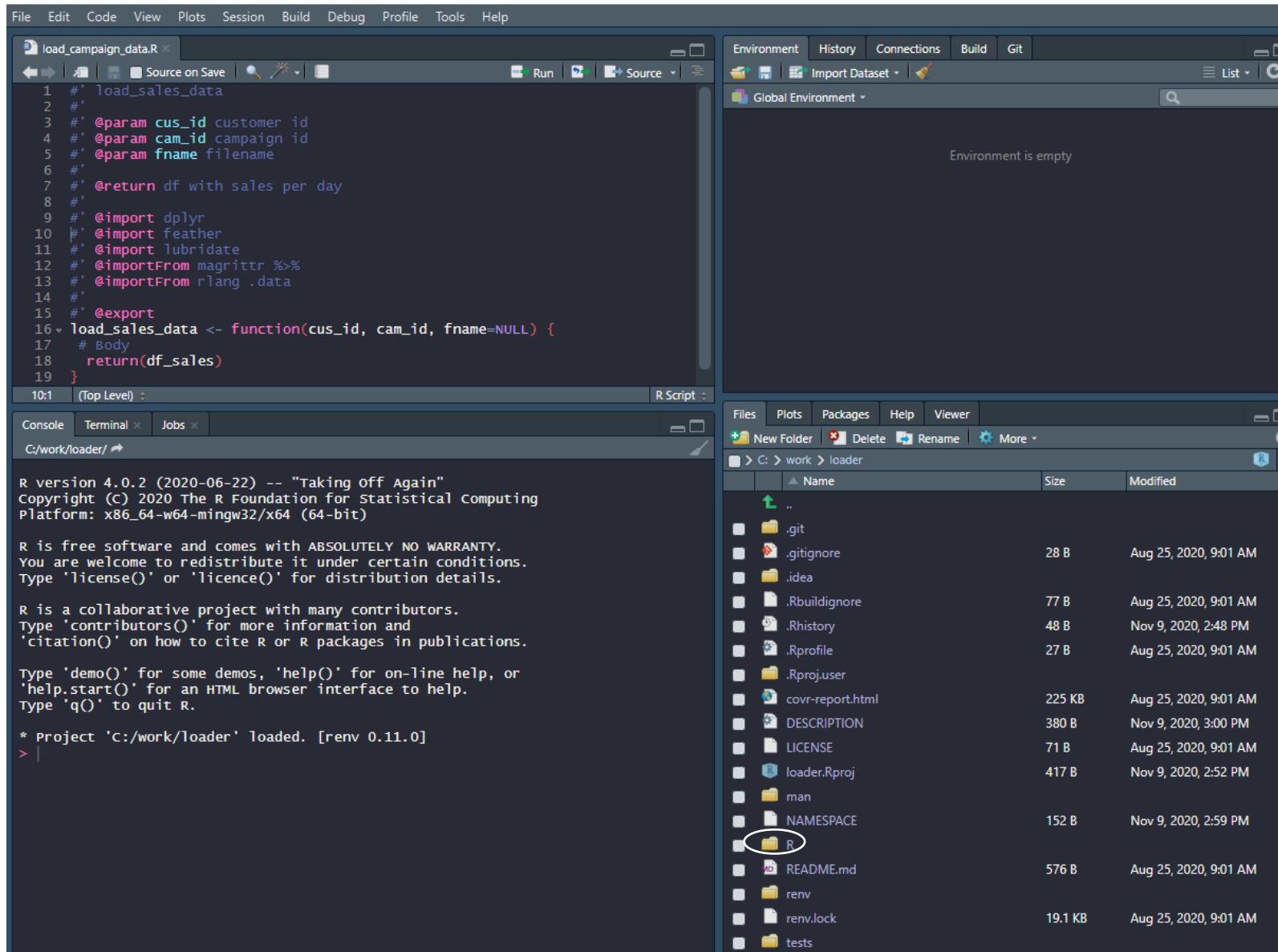


> This PC > Windows (C:) > work > loader >

Name	Date modified	Type	Size
.git	16/11/2020 19:00	File folder	
.idea	09/11/2020 14:49	File folder	
.Rproj.user	09/11/2020 14:49	File folder	
man	09/11/2020 15:16	File folder	
R	09/11/2020 14:56	File folder	
renv	09/11/2020 15:12	File folder	
tests	09/11/2020 19:22	File folder	
.gitignore	25/08/2020 09:01	Text Document	1 KB
.Rbuildignore	25/08/2020 09:01	RBUILDIGNORE File	1 KB
.Rhistory	16/11/2020 19:00	RHISTORY File	1 KB
.Rprofile	25/08/2020 09:01	RPROFILE File	1 KB
covr-report.html	25/08/2020 09:01	Microsoft Edge H...	226 KB
DESCRIPTION	09/11/2020 16:00	File	1 KB
LICENSE	25/08/2020 09:01	File	1 KB
loader.Rproj	16/11/2020 18:25	R Project	1 KB
NAMESPACE	09/11/2020 14:59	File	1 KB
README.md	25/08/2020 09:01	MD File	1 KB
renv.lock	09/11/2020 16:01	LOCK File	2 KB

Developing own R packages to structure data analysis

HOW



The screenshot shows the RStudio interface. The script editor on the left contains the following code:

```
1 #' load_sales_data
2 #'
3 #' @param cus_id customer id
4 #' @param cam_id campaign id
5 #' @param fname filename
6 #'
7 #' @return df with sales per day
8 #'
9 #' @import dplyr
10 #' @import feather
11 #' @import lubridate
12 #' @importFrom magrittr %>%
13 #' @importFrom rlang .data
14 #'
15 #' @export
16 load_sales_data <- function(cus_id, cam_id, fname=NULL) {
17   # Body
18   return(df_sales)
19 }
```

The environment pane on the right shows "Global Environment" and "Environment is empty". The console at the bottom shows the R version and project information:

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

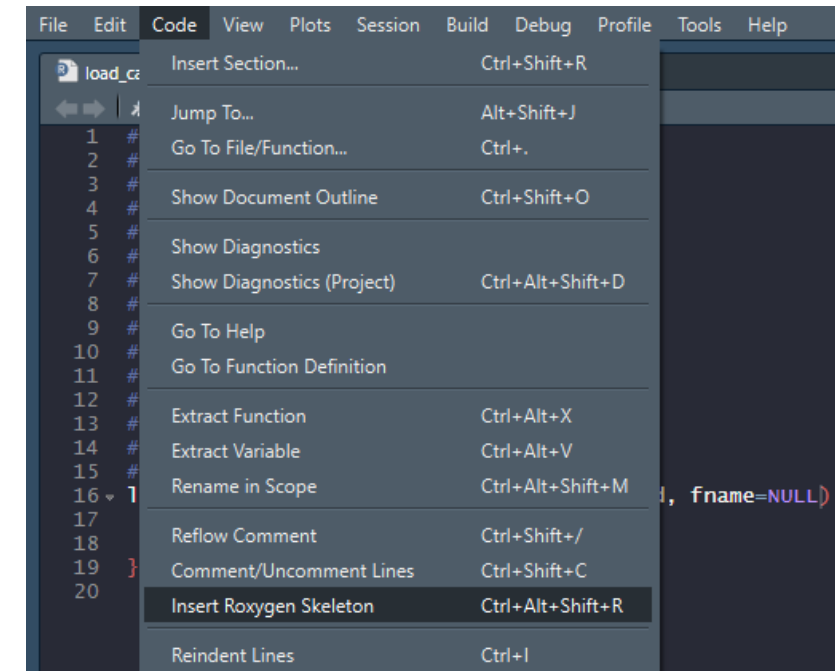
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

* Project 'c:/work/loader' loaded. [renv 0.11.0]
>
```

The file explorer on the right shows the project structure, with the "R" folder circled.

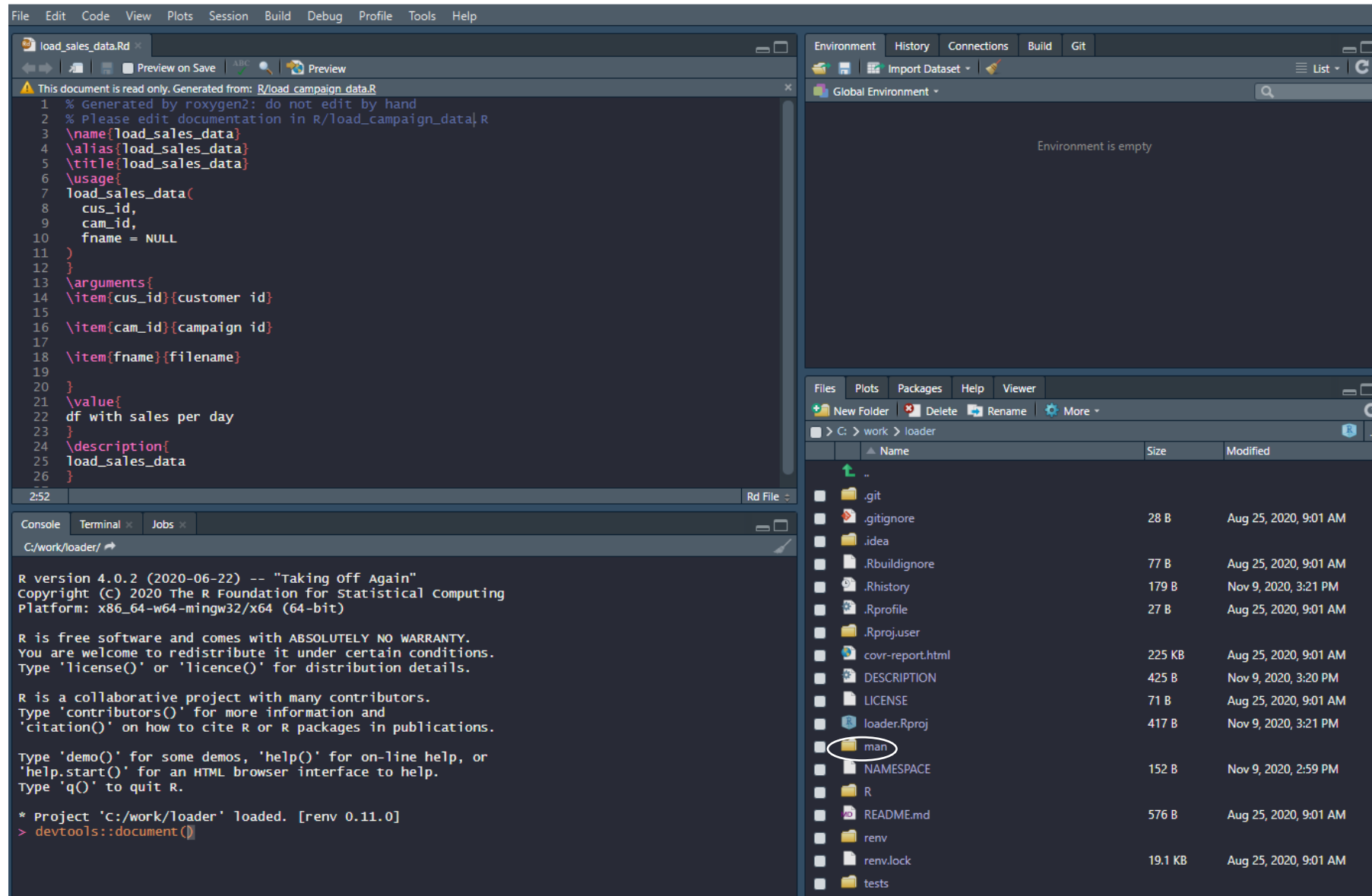
- Use **devtools**, **usethis**, and **roxygen2** packages in RStudio
- Follow a template: a package is a R project (e.g., '*loader.Rproj*') with predefined files and folders
- Functions must be stored under the *R* folder
- Functions must be documented using *Roxygen Skeleton*



The screenshot shows the RStudio interface with the 'Code' menu open. The 'Insert Roxygen Skeleton' option is highlighted, which is used to generate the initial documentation structure for an R package function.

Developing own R packages to structure data analysis

HOW



The screenshot shows the RStudio IDE interface. The main editor window displays the 'load_sales_data.Rd' file, which is a roxygen2-generated file. The console shows the R version (4.0.2) and the project path (C:/work/loader/). The file explorer on the right shows the directory structure, including a 'man' folder which is circled.

```

1 % Generated by roxygen2: do not edit by hand
2 % Please edit documentation in R/load_campaign_data.R
3 \name{load_sales_data}
4 \alias{load_sales_data}
5 \title{load_sales_data}
6 \usage{
7   load_sales_data(
8     cus_id,
9     cam_id,
10    fname = NULL
11  )
12 }
13 \arguments{
14   \item{cus_id}{customer id}
15 }
16   \item{cam_id}{campaign id}
17 }
18   \item{fname}{filename}
19 }
20 }
21 \value{
22   df with sales per day
23 }
24 \description{
25   load_sales_data
26 }

```

R version 4.0.2 (2020-06-22) -- "Taking Off Again"
 Copyright (C) 2020 The R Foundation for Statistical Computing
 Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
 You are welcome to redistribute it under certain conditions.
 Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
 Type 'contributors()' for more information and
 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
 'help.start()' for an HTML browser interface to help.
 Type 'q()' to quit R.

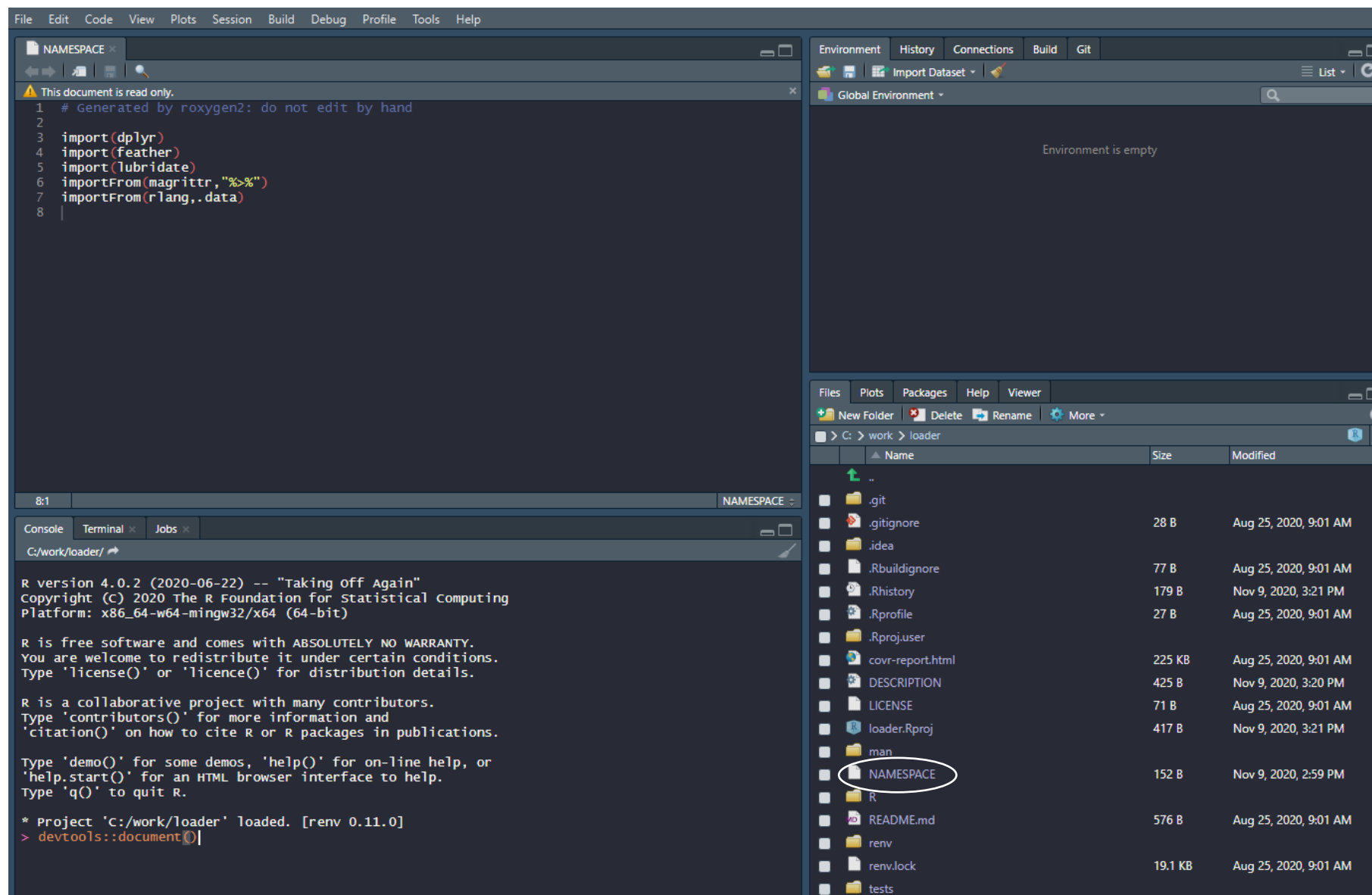
* Project 'C:/work/loader' loaded. [renv 0.11.0]
 > devtools::document()

Name	Size	Modified
..		
.git	28 B	Aug 25, 2020, 9:01 AM
.gitignore		
.idea		
.Rbuildignore	77 B	Aug 25, 2020, 9:01 AM
.Rhistory	179 B	Nov 9, 2020, 3:21 PM
.Rprofile	27 B	Aug 25, 2020, 9:01 AM
.Rproj.user		
covr-report.html	225 KB	Aug 25, 2020, 9:01 AM
DESCRIPTION	425 B	Nov 9, 2020, 3:20 PM
LICENSE	71 B	Aug 25, 2020, 9:01 AM
loader.Rproj	417 B	Nov 9, 2020, 3:21 PM
man		
NAMESPACE	152 B	Nov 9, 2020, 2:59 PM
R		
README.md	576 B	Aug 25, 2020, 9:01 AM
renv		
renv.lock	19.1 KB	Aug 25, 2020, 9:01 AM
tests		

- The *man* folder contains one *.Rd* file for each function
- The *.Rd* file is automatically generated with the command `devtools::document()`

Developing own R packages to structure data analysis

HOW

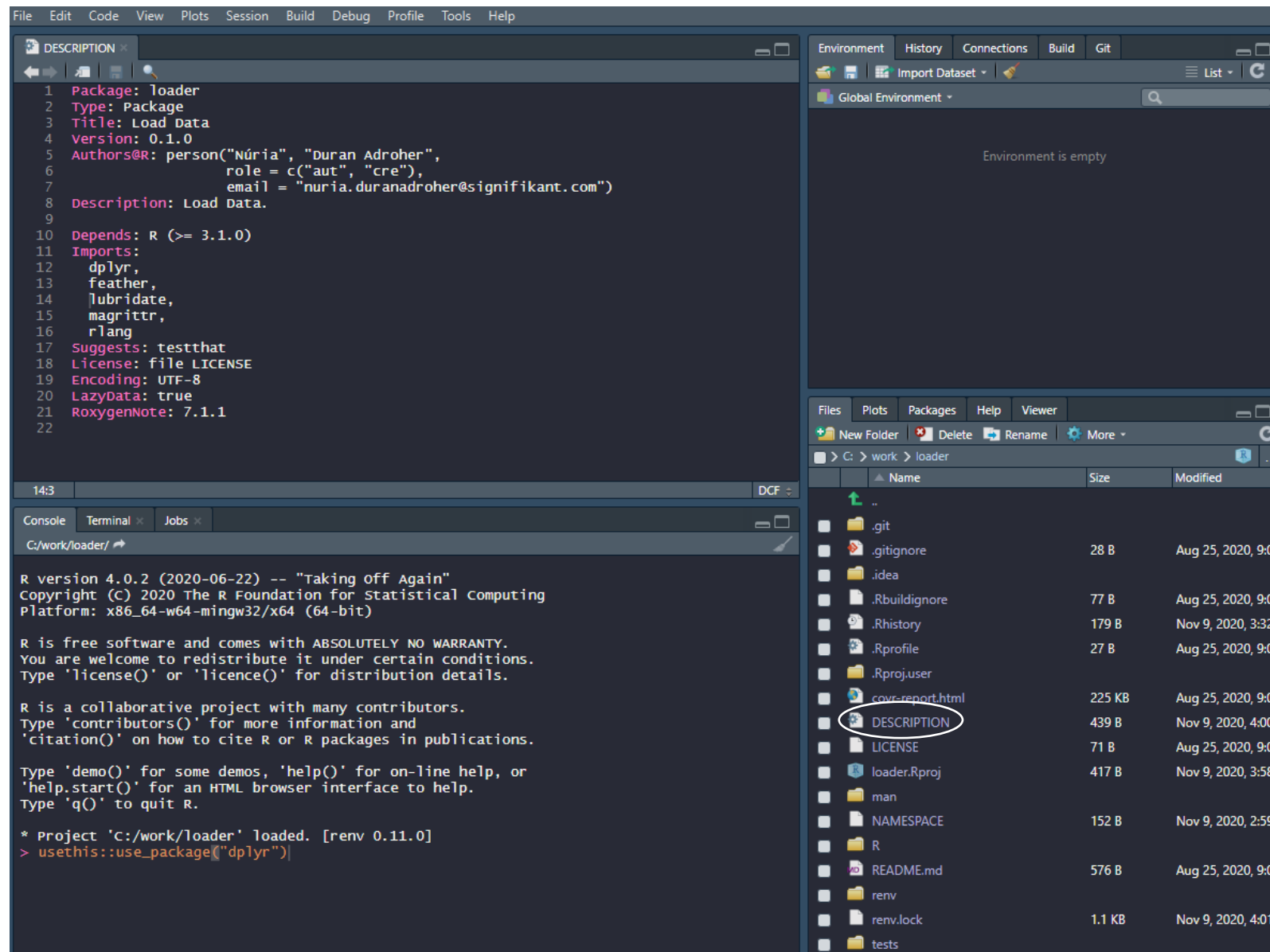


The screenshot displays the RStudio IDE interface during the development of an R package. The top-left pane shows the `NAMESPACE` file, which contains import statements for various packages: `dplyr`, `feather`, `lubridate`, `magrittr`, and `rlang`. The bottom-left pane shows the R console output, including the R version (4.0.2) and project information. The right pane shows the 'Environment' tab, which is currently empty. The bottom-right pane shows a file explorer view of the project directory, with the `NAMESPACE` file highlighted.

- The `NAMESPACE` file contains commands for importing or exporting functions, generated by roxygen2
- It is automatically updated using the command `devtools::document()`

Developing own R packages to structure data analysis

HOW



The screenshot shows the RStudio IDE interface. The left pane displays the `DESCRIPTION` file for a package named `loader`. The right pane shows the Environment pane (empty) and the File explorer (showing the package structure).

DESCRIPTION file content:

```

1 Package: loader
2 Type: Package
3 Title: Load Data
4 Version: 0.1.0
5 Authors@R: person("Núria", "Duran Adroher",
6                 role = c("aut", "cre"),
7                 email = "nuria.duranadroher@signifikant.com")
8 Description: Load Data.
9
10 Depends: R (>= 3.1.0)
11 Imports:
12   dplyr,
13   feather,
14   lubridate,
15   magrittr,
16   rlang
17 Suggests: testthat
18 License: file LICENSE
19 Encoding: UTF-8
20 LazyData: true
21 RoxygenNote: 7.1.1
22

```

Environment pane: Environment is empty.

File explorer: Shows the package structure. The `DESCRIPTION` file is highlighted.

Console/terminal output:

```

R version 4.0.2 (2020-06-22) -- "Taking off Again"
Copyright (c) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

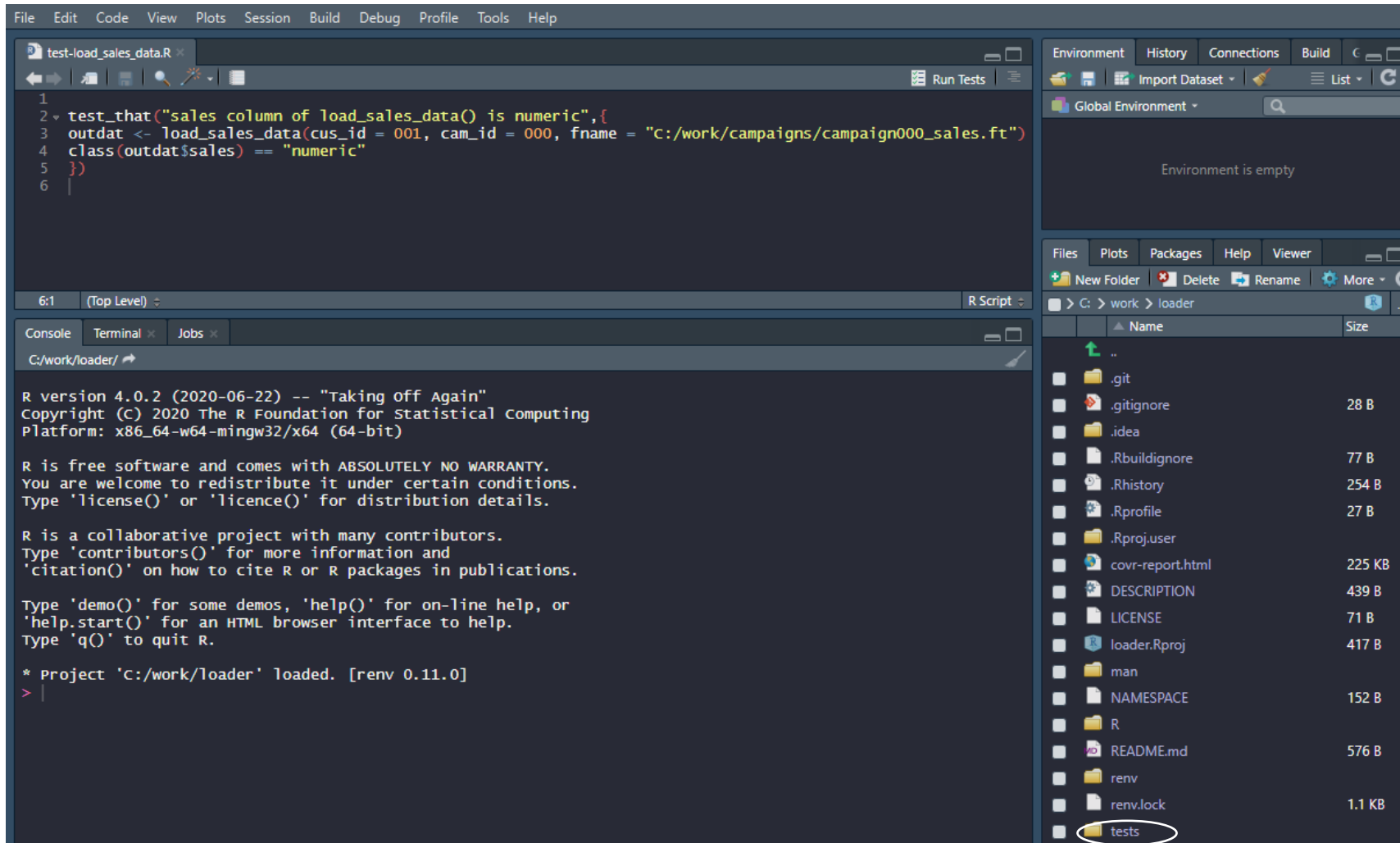
* Project 'c:/work/loader' loaded. [renv 0.11.0]
> usethis::use_package("dplyr")

```

- The *DESCRIPTION* file contains metadata
- The *Imports* section is automatically updated using the command `usethis::use_package("pkg_name")`

Developing own R packages to structure data analysis

HOW



The screenshot shows the RStudio IDE interface. The main editor window displays a test script for the `load_sales_data` function. The script includes a `test_that` block that checks if the `sales` column of `load_sales_data()` is numeric. The console window shows the R version (4.0.2) and the project path (`C:/work/loader`). The file explorer on the right shows the project structure, with the `tests` folder highlighted.

```
1 test_that("sales column of load_sales_data() is numeric",{
2   outdat <- load_sales_data(cus_id = 001, cam_id = 000, fname = "c:/work/campaigns/campaign000_sales.ft")
3   class(outdat$sales) == "numeric"
4 })
5
6
```

Console output:

```
R version 4.0.2 (2020-06-22) -- "Taking off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

* Project 'C:/work/loader' loaded. [renv 0.11.0]
>
```

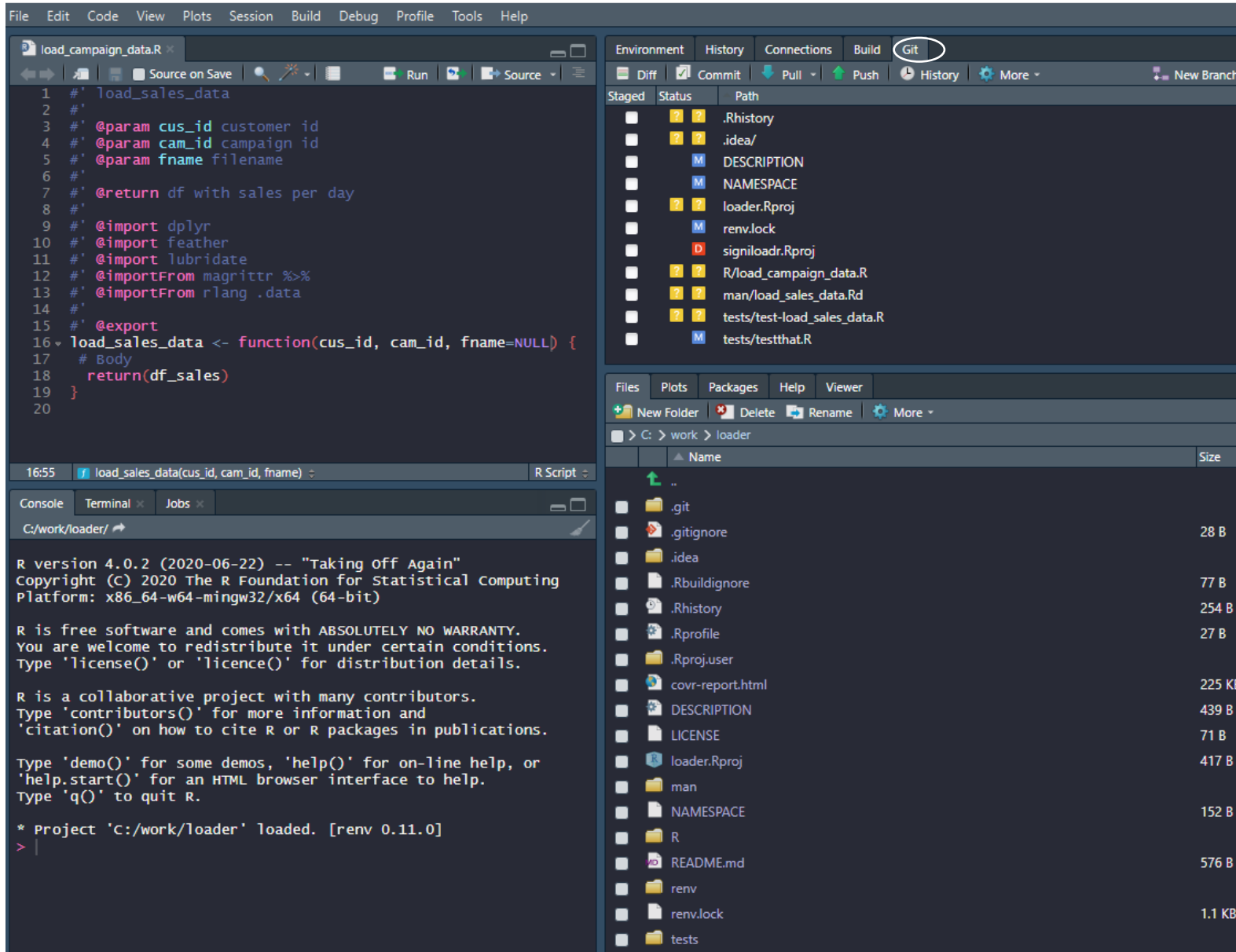
File explorer contents:

Name	Size
..	
.git	
.gitignore	28 B
.idea	
.Rbuildignore	77 B
.Rhistory	254 B
.Rprofile	27 B
.Rproj.user	
covr-report.html	225 KB
DESCRIPTION	439 B
LICENSE	71 B
loader.Rproj	417 B
man	
NAMESPACE	152 B
R	
README.md	576 B
renv	
renv.lock	1.1 KB
tests	

- Write unit tests under the *tests* folder

Developing own R packages to structure data analysis

HOW



The screenshot shows the RStudio IDE interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top toolbar has buttons for Source on Save, Run, and Source. The main editor window displays R code for a function named `load_sales_data`. The console window shows the R version (4.0.2) and the project path (C:/work/loader). The file explorer on the right shows the project structure, including files like `.Rhistory`, `.idea/`, `DESCRIPTION`, `NAMESPACE`, `loader.Rproj`, `renv.lock`, `signiloadr.Rproj`, `R/load_campaign_data.R`, `man/load_sales_data.Rd`, `tests/test-load_sales_data.R`, and `tests/testthat.R`.

```

1 #' load_sales_data
2 #'
3 #' @param cus_id customer id
4 #' @param cam_id campaign id
5 #' @param fname filename
6 #'
7 #' @return df with sales per day
8 #'
9 #' @import dplyr
10 #' @import feather
11 #' @import lubridate
12 #' @importFrom magrittr %>%
13 #' @importFrom rlang .data
14 #'
15 #' @export
16 load_sales_data <- function(cus_id, cam_id, fname=NULL) {
17   # Body
18   return(df_sales)
19 }
20

```

Console output:

```

R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (c) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

* Project 'C:/work/loader' loaded. [renv 0.11.0]
>

```

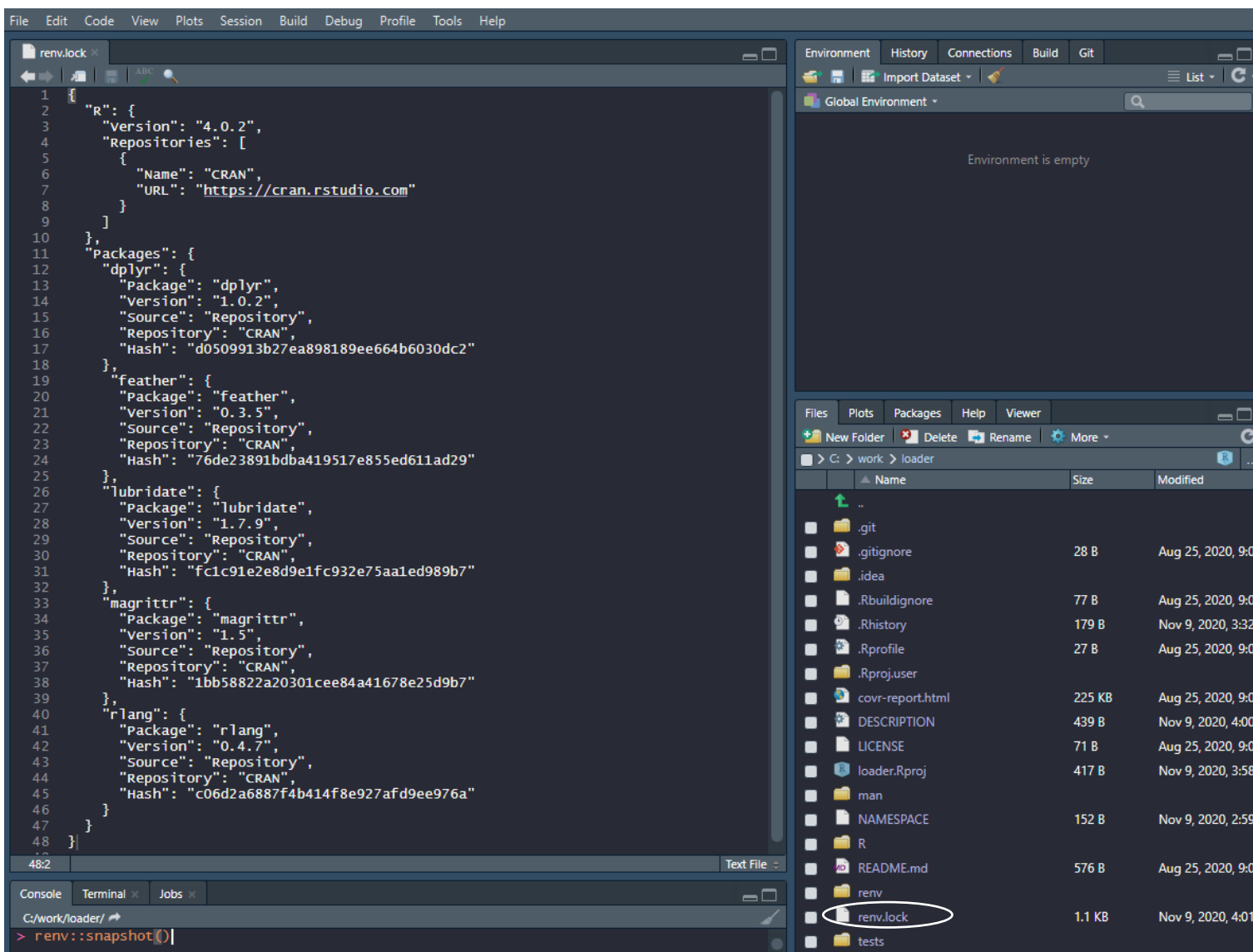
File explorer (C:/work/loader):

Name	Size
..	
.git	
.gitignore	28 B
.idea	
.Rbuildignore	77 B
.Rhistory	254 B
.Rprofile	27 B
.Rproj.user	
covr-report.html	225 KB
DESCRIPTION	439 B
LICENSE	71 B
loader.Rproj	417 B
man	
NAMESPACE	152 B
R	
README.md	576 B
renv	
renv.lock	1.1 KB
tests	

- Commit and push

Developing own R packages to structure data analysis

HOW (optional)



The screenshot shows the RStudio interface. The main editor displays the content of the `renv.lock` file, which is a JSON document specifying the R environment. The file lists the R version (4.0.2) and the CRAN repository. It also lists several installed packages with their versions, sources, and hashes:

```

{
  "R": {
    "Version": "4.0.2",
    "Repositories": [
      {
        "Name": "CRAN",
        "URL": "https://cran.rstudio.com"
      }
    ]
  },
  "Packages": {
    "dplyr": {
      "Package": "dplyr",
      "Version": "1.0.2",
      "Source": "Repository",
      "Repository": "CRAN",
      "Hash": "d0509913b27ea898189ee664b6030dc2"
    },
    "feather": {
      "Package": "feather",
      "Version": "0.3.5",
      "Source": "Repository",
      "Repository": "CRAN",
      "Hash": "76de23891bdba419517e855ed611ad29"
    },
    "lubridate": {
      "Package": "lubridate",
      "Version": "1.7.9",
      "Source": "Repository",
      "Repository": "CRAN",
      "Hash": "fc1c91e2e8d9e1fc932e75aa1ed989b7"
    },
    "magrittr": {
      "Package": "magrittr",
      "Version": "1.5",
      "Source": "Repository",
      "Repository": "CRAN",
      "Hash": "1bb58822a20301cee84a41678e25d9b7"
    },
    "rlang": {
      "Package": "rlang",
      "Version": "0.4.7",
      "Source": "Repository",
      "Repository": "CRAN",
      "Hash": "c06d2a6887f4b414f8e927afd9ee976a"
    }
  }
}

```

The file explorer on the right shows the project structure. The `renv` folder is highlighted, and the `renv.lock` file is circled. The console at the bottom shows the command `renv::snapshot()` being executed.

- Save and load the state of a project library using the package `renv`
 - The `renv.lock` file captures the state of the library at some point in time

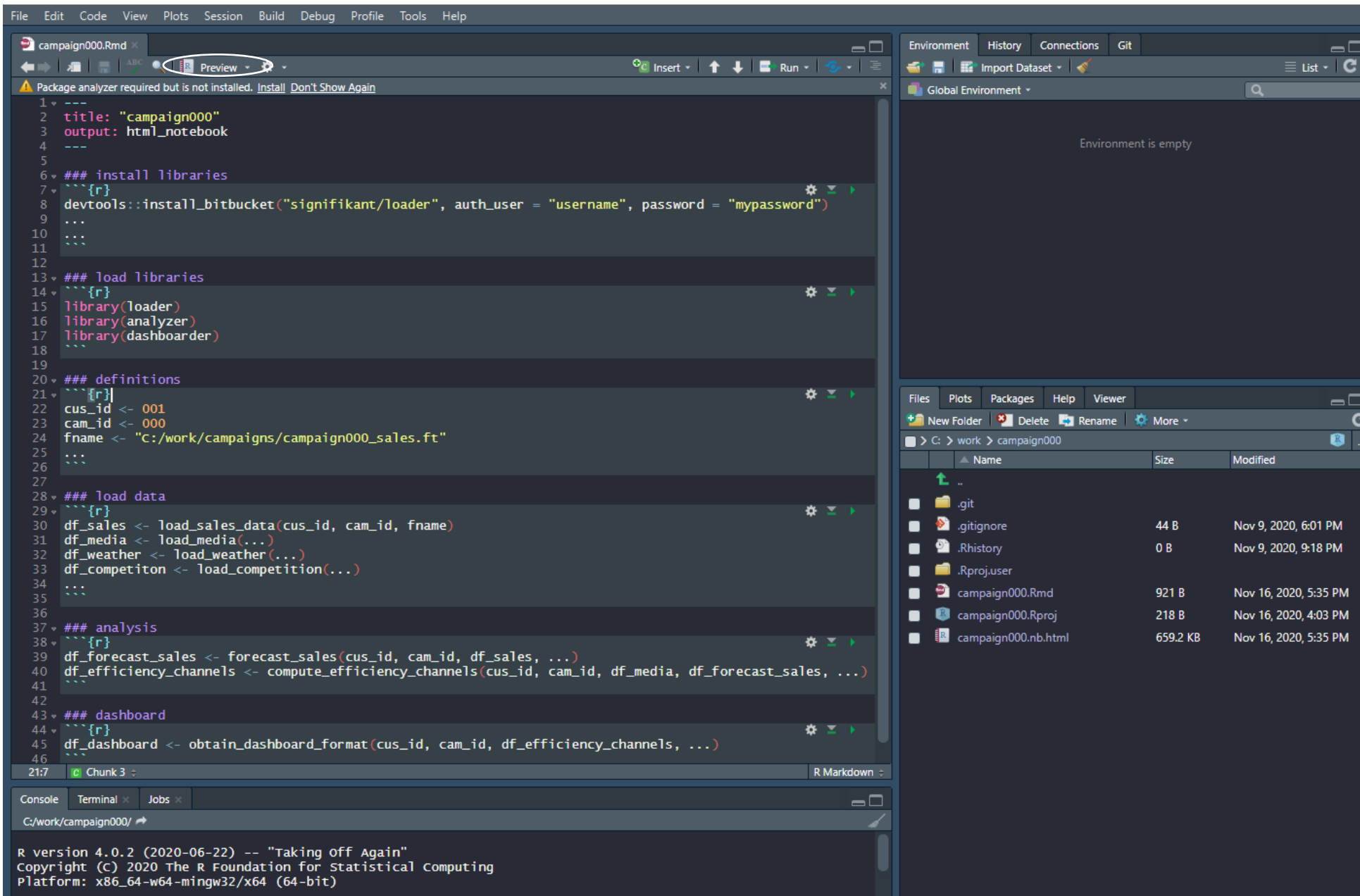
Using own R packages for the analysis



- Upload them in **bitbucket** (similar repository to **github**)
- In a R script (.R) or R Markdown file (.Rmd), install them via the command `devtools::install_bitbucket(...)`

Using the packages for the analysis

- Create a .Rmd file containing the steps of the analysis process



The screenshot displays the RStudio IDE interface with the following components:

- Source Editor:** Contains an R Markdown file named 'campaign000.Rmd'. The code is organized into sections: title, output, install libraries, load libraries, definitions, load data, analysis, and dashboard. A warning message at the top states: 'Package analyzer required but is not installed. Install Don't Show Again'.
- Environment Pane:** Shows 'Global Environment' with the message 'Environment is empty'.
- Files Pane:** Displays the project structure for 'campaign000'. The files listed are: .git, .gitignore (44 B, Nov 9, 2020, 6:01 PM), .Rhistory (0 B, Nov 9, 2020, 9:18 PM), .Rproj.user, campaign000.Rmd (921 B, Nov 16, 2020, 5:35 PM), campaign000.Rproj (218 B, Nov 16, 2020, 4:03 PM), and campaign000.nb.html (659.2 KB, Nov 16, 2020, 5:35 PM).
- Console:** Shows the R version (4.0.2 (2020-06-22)) and platform (x86_64-w64-mingw32/x64 (64-bit)).

campaign000

Code

install libraries

Hide

```
devtools::install_bitbucket("signifikant/loader", auth_user = "username", password = "mypassword")  
...  
...
```

load libraries

Hide

```
library(loader)  
library(analyzer)  
library(dashboarder)
```

definitions

Hide

```
cus_id <- 001  
cam_id <- 000  
fname <- "C:/work/campaigns/campaign000_sales.ft"  
...
```

load data

Hide

```
df_sales <- load_sales_data(cus_id, cam_id, fname)  
df_media <- load_media(...)  
df_weather <- load_weather(...)  
df_competition <- load_competition(...)  
...
```

analysis

Hide

```
df_forecast_sales <- forecast_sales(cus_id, cam_id, df_sales, ...)  
df_efficiency_channels <- compute_efficiency_channels(cus_id, cam_id, df_media, df_forecast_sales, ...)
```

dashboard

Hide

```
df_dashboard <- obtain_dashboard_format(cus_id, cam_id, df_efficiency_channels, ...)
```

Summary

- R packages can be used for any kind of project; we've seen the example of defining functions which analyse ad campaign data
- To develop a R package one needs to create a project and adapt to the specified template of files and folders
- Reproducibility enhanced by using **renv**
- Automation obtained by running a main R script / Markdown file which mainly calls functions defined in the own developed packages
- Usability is improved by documented functions using **roxygen2**
- Visibility is improved by clean code in R script / Markdown file

Thanks for your attention!

nuria.duranadroher@signifikant.com