Beat The Code

FCA Collin, Ph.D.

Thursday, March 18, 2021

Contents

2021-03-11, FCA Collin

`Tidyverse`, Not a Golden Hammer

The Law of the instrument describes a cognitive bias:

> "I call it the law of the instrument, and it may be formulated as follows: Give a small boy a hammer, and he will find that everything he encounters needs pounding." (Abraham Kaplan, 1964)

It was identified as an *AntiPattern*, a programming practice to be avoided (William Brown et al, 1998). One of the pit fall is expressed as:

> "the tendency of jobs to be adapted to tools, rather than adapting tools to jobs" (Silvan Tomkins, 1963).

The tidyverse package helps end-user in R-coding delimited statistic tasks. It is a very good idea to use it if your purpose is to walk through an analysis from a point A (the dataset) to the point B (the result) for procedures of limited complexity. Indeed, with a limited number of human-readable functions you can get the expected result while helping the future reader to follow the procedure. However, as soon as you want to resolve statistical problems in a more systemic way, by creating functions that will help you to get your result in a more concise (because accurate) code and tested for and documented and robust, tidyverse is not the most suitable choice.

The package vignette itself enclose a clear disclaimer about the package rational:

> *"the biggest difference is in priorities: base R is highly focussed on stability, whereas the tidyverse will make breaking changes in the search for better interfaces."* *Welcome to the Tidyverse* vignette, 2019

The trade-off between stability and interface evolution is also acknowledged:

> Do you expect the tidyverse to be the part of core R packages some day?
>
> **Hadley Wickham:** *"It's extremely unlikely because the core packages are extremely conservative so that base R code is stable, and backward compatible. I prefer to have a more utopian approach where I can be quite aggressive about making backward incompatible changes while trying to figure out a better API."* quora

There is no doubt that tidyverse is a set of high quality tools, but it is designed to serve some purpose: easy and highly readable code at the cost of stability which is a strategy which can't serve all developments. Besides, the over reliance of craftsperson on a known tool, brings to see the challenge not as it is but how it fit to the tool. As a matter of fact, an over reliance on tidyverse risk to introduce a cognitive bias, increasing the risk of of deviation from initial target as fitting your purpose to the problem instead of making the method suitable to answer specific question. The over-reliance can be evidence by a large tidyverse block which have obviously lost the main sells argument of the package: readability.

In order to increase the range of possible ways to address a question, so as to minimise the risk of programming cognitive bias, it is good to demonstrate alternatives to the tidyverse approach put in some context, and present the R base alternative. There will be a trade-off switching from one to the other about readability, performance and code stability. But, maybe, this will also help thinking about different approaches to address statistical problems.

Filter and Select

```
library(tidyverse)
```

Tidyverse Non-Standard Evaluation:

```
iris %>%
  filter(Species == "setosa") %>%
  select(Sepal.Width, Sepal.Length) %>%
  head
```

```
##   Sepal.Width Sepal.Length
```

```
## 1           3.5          5.1
## 2           3.0          4.9
## 3           3.2          4.7
## 4           3.1          4.6
## 5           3.6          5.0
## 6           3.9          5.4
```

Base Non-Standard Evaluation:

```
sel <- subset(
  iris,
  subset = Species == "setosa",
  select = c(Sepal.Width, Sepal.Length)
)
head(sel)
```

```
##   Sepal.Width Sepal.Length
## 1           3.5          5.1
## 2           3.0          4.9
## 3           3.2          4.7
## 4           3.1          4.6
## 5           3.6          5.0
## 6           3.9          5.4
```

Base data.frame accessors:

```
sel <- iris[
  iris$Species == "setosa",
  c("Sepal.Width", "Sepal.Length")
  ]
head(sel)
```

```
##   Sepal.Width Sepal.Length
## 1           3.5          5.1
## 2           3.0          4.9
## 3           3.2          4.7
## 4           3.1          4.6
## 5           3.6          5.0
## 6           3.9          5.4
```

Mutate

```
df[df$age > 90, ] <- NA
```

```
sessionInfo()
```

```
## R version 4.0.4 (2021-02-15)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 10 (buster)
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/libopenblasp-r0.3.5.so
##
## locale:
##  [1] LC_CTYPE=en_GB.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_GB.UTF-8        LC_COLLATE=en_GB.UTF-8
##  [5] LC_MONETARY=en_GB.UTF-8    LC_MESSAGES=en_GB.UTF-8
##  [7] LC_PAPER=en_GB.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets
methods   base
##
## other attached packages:
## [1] forcats_0.5.1   stringr_1.4.0   dplyr_1.0.5
purrr_0.3.4
## [5] readr_1.4.0     tidyr_1.1.3     tibble_3.1.0
ggplot2_3.3.3
## [9] tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.1.0  xfun_0.22         haven_2.3.1
colorspace_2.0-0
##  [5] vctrs_0.3.6       generics_0.1.0    htmltools_0.5.1.1 yaml_2.2.1
##  [9] utf8_1.2.1        rlang_0.4.10      pillar_1.5.1
withr_2.4.1
## [13] glue_1.4.2        DBI_1.1.1         dbplyr_2.0.0
modelr_0.1.8
## [17] readxl_1.3.1      lifecycle_1.0.0   munsell_0.5.0
gtable_0.3.0
```

```
## [21] cellranger_1.1.0   rvest_1.0.0        evaluate_0.14
knitr_1.31
## [25] ps_1.6.0           fansi_0.4.2        broom_0.7.5
Rcpp_1.0.6
## [29] scales_1.1.1       backports_1.2.1    jsonlite_1.7.2
fs_1.5.0
## [33] hms_1.0.0          digest_0.6.27      stringi_1.5.3
grid_4.0.4
## [37] cli_2.3.1          tools_4.0.4        magrittr_2.0.1
crayon_1.4.1
## [41] pkgconfig_2.0.3    ellipsis_0.3.1     xml2_1.3.2
reprex_1.0.0
## [45] lubridate_1.7.10   assertthat_0.2.1   rmarkdown_2.6
httr_1.4.2
## [49] rstudioapi_0.13    R6_2.5.0           compiler_4.0.4
```