R Code – Best practices

Ken Mwai

1 – Naming conventions

- R has no standardised naming conventions
- Always choose a naming convention to work with; for example
  - all lowercase: e.g. adjustcolor
  - underscore separated: e.g. numeric_version
  - lowerCamelCase: e.g. addTaskCallback
  - UpperCamelCase: e.g. SignatureMethod
- **Avoid SPACES** while naming files

“There are only two hard things in Computer Science: cache invalidation and naming things.” — Phil Karlton

- Strive for names that are concise and meaningful
- R file names should be meaningful and end in .R.

Object names

- Variable and function names should be lowercase.

```
# Good
day_one
day_1

# Bad
first_day_of_the_month
DayOne
dayone
djm1
```

- avoid using names of existing functions and variables.

2 – Files organisation

- File organisation makes code and data analysis project readable
- Data should be seperated from codes
- Documents should be seperated from codes
- Use project facility of RStudio each time you start working on a new project

3 - organise the code within each file

- Start each file with a comment saying who wrote it and when, what it contains, and how it fits into the larger program
- Load all required packages (at the top of the script)
- Source required data files if any
- Modularize your code.

```r
#-----------------------------------------------
## f_StaR Introduction
## Ken Mwai - April 2022
#-----------------------------------------------
#-----------------------------------------------
# 0 - Load librairies
#-----------------------------------------------
library(dplyr)
library(ggplot2)
#-----------------------------------------------
# 1 - Source Data
#-----------------------------------------------
df1 <- read_csv("data/my_data.csv")
#-----------------------------------------------
# 2 - Start my code
#-----------------------------------------------
mean(mtcars$mpg)
```

3 – Syntax

- Place spaces around all infix operators (`=, +, -, <-, etc.`).
- Use <–, not =, for object assignment in R.
- Use comments to mark off sections of code.
- Comment your code with care. Comments should explain the why, not the what
- Each line of a comment should begin with the comment symbol and a single space
- Keep your lines less than 80 characters.

```r
# This is a comment
# Good
# Object assignment in R
x <- 10

#Bad
x=10
```

```r
# Good
average <- mean(feet / 12 + inches, na.rm = TRUE)

# Bad
average<-mean(feet/12+inches,na.rm=TRUE)
```

Use <-, not =, for assignment.

```r
# Good
x <- 5
# Bad
x = 5
```

Take a first look at the data. Useful functions are `dim()`, `head()`,`str()` and `summary()`.

Solution

```
dim(mtcars)
```

```
## [1] 32 11
```

```
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
```

```
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp              
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0   Min.
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5   1st Q
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0   Media
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7   Mean
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0   3rd Q
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0   Max.
##       qsec             vs              am              gear          
##  Min.   :14.50   Min.   :0.0000   Min.   :0.0000   Min.   :3.000   Mi
##  1st Qu.:16.89   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:3.000   1st
##  Median :17.71   Median :0.0000   Median :0.0000   Median :4.000   Med
##  Mean   :17.85   Mean   :0.4375   Mean   :0.4062   Mean   :3.688   Mea
##  3rd Qu.:18.90   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:4.000   3rd
##  Max.   :22.90   Max.   :1.0000   Max.   :1.0000   Max.   :5.000   Max
```

Notes

- Do not include functions that change someone's computer i.e.
  - Installation of packages
  - setwd()
- Follow a style and be consistent.
  - https://google.github.io/styleguide/Rguide.html
  - https://style.tidyverse.org/files.html
  - https://bookdown.org/marius_mather/Rad/tips-for-effective-r-programming.html#writing-readable-code