

Code and Data Styles

Each example below has two versions. How many differences can you find between them? Discuss which difference you prefer and why. If you discover anything you'd like to emulate (or not!), write it in your personal style guide. We'll come together as a whole group to discuss each example in turn.

Example 1

<i>in file</i> todayswork.r	<i>in file</i> analyze_height-2014-04-30.R
<pre>data <- read.csv("data.csv",sep="\t") attach(data) height_for_each_person<-Feet/12+inches V3 <- factor(V3) mymodel=lm(height_for_each_person~V3+V4) summary(mymodel)</pre>	<pre>### read and manipulate data ----- d <- read.csv("height.csv", sep="\t") d <- within(d, { Height <- Feet / 12 + Inches Sex <- factor(Sex, levels=1:2, labels=c("M", "F")) }) ### data analysis ----- model.height <- lm(Height ~ Sex + Age, data=d) summary(model.height)</pre>

Example 2

<pre>## my box-cox function my.func <- function(a ,b) { if(b==0) { log(a) } else (a^b- 1)/b }</pre>	<pre>## the Box-Cox transformation ## x: the data to be transformed ## lambda: the Box-Cox parameter ## output: the transformed data myBoxCox <- function(x, lambda) { if (lambda == 0) { log(x) } else { (x^lambda - 1) / lambda } }</pre>
--	--

Example 3

```
dat <- read.csv("mydata.csv")
dat_byday <- split(dat, dat$day)
mean.and.sd <- function(x) {c(m=mean(x), sd=sd(x))}
day1.summary <- mean.and.sd(dat_byday$'1'$response)
DayTwoSummary <- mean.and.sd(dat_byday$'2'$response)

dat <- read.csv("mydata.csv")
dat.byday <- split(dat, dat$day)
getMeanAndSd <- function(x) {c(m=mean(x), sd=sd(x))}
day1.summary <- getMeanAndSd(dat.byday$'1'$response)
day2.summary <- getMeanAndSd(dat.byday$'2'$response)
```

Data Management

Data management is the “creation, storage, analysis, dissemination, and preservation of your research data” (<https://www.lib.umn.edu/datamanagement>).

Why is this important? Why should it be part of your role?

Here are two examples of a spreadsheet for storing data. Each has certain elements you make like or not like. Discuss with your groups. What do you prefer, and why? Add to your personal style guide.

Example 1:

Class	Student	Time (Days)	Score	M/F	Age (Years)	IncomeGroup 1=lo, 2=med, 3=hi
Johnson	Alice	1	67	F	7	1
		2	79	F	7	1
		3	71	F	7	1
	Bob	1	58	M	7	2
		2	59	M	7	2
		3	55	M	7	2
Olson	Carol	1	80	F	8	2
		2	82	F	8	2
		3	86	F	8	2
	Daniel	1	55	M	7	3

Example 2:

Johnson class is Bold, Olson class is *Italic*
IncomeGroup: 1=lo, 2=med, 3=hi

Student	Score1	Score2	Score3	Change	Sex	Age	IncomeGroup
Alice	67	79	71	4	F	7 years	1
Bob	58	59	55	-3	M	7 years	2
Emma	70	76	81	11	F	9 years	1
Mason	80	83	70	-10	M	8 years	3
<i>Carol</i>	80	82	86	6	F	8 years	2
<i>Daniel</i>	55	60	42	-13	M	7 years	3
<i>Jacob</i>	50	60	65	15	M	9 years	1

Style Guide

“Good coding style is like using correct punctuation. You may think you can manage without it, but it sure makes things easier to read. As with styles of punctuation, there are many possible variations. [Well-known examples are by Hadley Wickham, Yihui Xie, and Google.] You don’t have to use [one of theirs]. However, you do need to have and to use a consistent style.” (from <http://adv-r.had.co.nz/Style.html>) Also see Karl Broman at UW-Madison for an example of a data style guide, <http://kbroman.org/dataorg/>.

Consider your own personal style guide. What would you include in each topic?

Coding:

naming (files/code variables/functions/data variables)

commenting

braces

indentation and spacing

Data:

columns (and column names)

rows (what should each row correspond to?)

cells (blanks/formatting/units, etc)

computations

Data Confidentiality

Discuss the balance between sharing your data and keeping it confidential. When are cases when one or the other might be necessary or preferred?

Name: _____

Which element of the code style guide do you think will be most helpful to you?
