

Code and Data Styles

With the people sitting next to you, see how many differences can you find in each of these examples. 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

```

in file todayswork.r                                     in file analyze_height-2014-04-30.R

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

```

Example 2

```

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

myBoxCox <- function(x, lambda) {
  if (lambda == 0) {
    log(x)
  } else {
    (x^lambda - 1) / lambda
  }
}

```

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)

```


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>)

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

file naming

variable naming (both in code and in data frames)

function naming

braces

indentation

commenting

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?

What are pros and cons of each of the following ways of storing a data set?

Method 1

Johnson class is Bold, Olson class is Italic

Alice	Sex	F		
	Age	7		
	IncomGroup	1		
	Scores	67	79	71
<i>Sophia</i>	Sex	F		
	Age	8		
	IncomeGroup	2		
	Scores	89	90	93

Method 2

Class	Student	Time	Score	Sex	Age	IncomeGroup
Johnson	Alice	1	67	F	7	1
Johnson	Alice	2	79	F	7	1
Johnson	Alice	3	71	F	7	1
Johnson	Bob	1	58	M	7	2
Johnson	Bob	2	59	M	7	2
Johnson	Bob	3	55	M	7	2
Johnson	Carol	1	80	F	8	2
Johnson	Carol	2	82	F	8	2
Johnson	Carol	3	86	F	8	2
Johnson	Daniel	1	55	M	7	3
Johnson	Daniel	2	60	M	7	3
Johnson	Daniel	3	42	M	7	3

Method 3

Class	Student	Score1	Score2	Score3	Sex	Age	IncomeGroup
Johnson	Alice	67	79	71	F	7	1
Johnson	Bob	58	59	55	M	7	2
Johnson	Carol	80	82	86	F	8	2
Johnson	Daniel	55	60	42	M	7	3
Olson	Sophia	89	90	93	F	8	2
Olson	Emma	70	76	81	F	9	1
Olson	Jacob	50	60	65	M	9	1
Olson	Mason	80	83	70	M	8	3

How do you look for errors in a data set? Are there specific things you should look for? Consider both categorical and continuous variables.

What do you do when you find an error? How about an outlier that you're not sure what to do with?

What kinds of metadata do you wish you had on this data set? How might you store it?

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?
