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

```r
in file todayswork.r

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 )
```

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

### 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)
```

Example 2

```r
## my box-cox function
my.func <- function(a ,b) {
  if( b==0 ) {
    log( a )
  } else {
    (a^b- 1)/b
  }
}

## the Box-Cox transformation
myBoxCox <- function(x, lambda) {
  if (lambda == 0) {
    log(x)
  } else {
    (x^lambda - 1) / lambda
  }
}
```

Example 3

```r
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

<table>
<thead>
<tr>
<th>Alice</th>
<th>Sex</th>
<th>Age</th>
<th>IncomeGroup</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>7</td>
<td>1</td>
<td>67 79 71</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sophia</th>
<th>Sex</th>
<th>Age</th>
<th>IncomeGroup</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>8</td>
<td>2</td>
<td>89 90 93</td>
</tr>
</tbody>
</table>
### Method 2

<table>
<thead>
<tr>
<th>Class</th>
<th>Student</th>
<th>Time</th>
<th>Score</th>
<th>Sex</th>
<th>Age</th>
<th>IncomeGroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson</td>
<td>Alice</td>
<td>1</td>
<td>67</td>
<td>F</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Johnson</td>
<td>Alice</td>
<td>2</td>
<td>79</td>
<td>F</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Johnson</td>
<td>Alice</td>
<td>3</td>
<td>71</td>
<td>F</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Johnson</td>
<td>Bob</td>
<td>1</td>
<td>58</td>
<td>M</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Johnson</td>
<td>Bob</td>
<td>2</td>
<td>59</td>
<td>M</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Johnson</td>
<td>Bob</td>
<td>3</td>
<td>55</td>
<td>M</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Johnson</td>
<td>Carol</td>
<td>1</td>
<td>80</td>
<td>F</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Johnson</td>
<td>Carol</td>
<td>2</td>
<td>82</td>
<td>F</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Johnson</td>
<td>Carol</td>
<td>3</td>
<td>86</td>
<td>F</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Johnson</td>
<td>Daniel</td>
<td>1</td>
<td>55</td>
<td>M</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Johnson</td>
<td>Daniel</td>
<td>2</td>
<td>60</td>
<td>M</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Johnson</td>
<td>Daniel</td>
<td>3</td>
<td>42</td>
<td>M</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

### Method 3

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

__________________________________________________________________________

__________________________________________________________________________

__________________________________________________________________________

__________________________________________________________________________