Evaluation of Interactive Systems

R Introduction
R: types

Basic types

Booleans (logical), Numbers (integer or double), Strings (character)

Vectors

```r
> a <- c(1,2,5,3,6,-2,4)  # numeric vector
> b <- c("one","two","three")  # character vector
> c <- c(TRUE,TRUE,TRUE,FALSE,TRUE,FALSE)  # logical vector
> a[c(2,4)]  # 2nd and 4th elements of vector (start index is 1)
[1] 2 6
```

Matrices: All columns in a matrix must have the same type (numeric, character, etc.) and the same length

```r
> cells <- c(1,26,24,68)
> rnames <- c("R1","R2")
> cnames <- c("C1","C2")
> mymatrix <- matrix(cells, nrow=2, ncol=2, byrow=TRUE,dimnames=list(rnames,cnames))
> mymatrix
   C1 C2
R1  1 26
R2 24 68
> mymatrix[1,]  # 1st row
   C1 C2
  1 26
> mymatrix[,2]  # 2nd column
   C1 C2
R1 R2
  26 68
```
R: types

Data frames: A data frame is more general than a matrix, in that different columns can have different types (numeric, character, factor, etc.)

Factors: categorical variables (nominal / ordinal). Storing data as factors ensures that the statistical models will treat such data correctly.

use `factor` function to create factors

```r
> d <- c(1,2,3,4)
> e <- c("red","white","red",NA)
> f <- c(TRUE,TRUE,TRUE,FALSE)
> mydata <- data.frame(d,e,f)
> names(mydata) <- c("ID", "Color", "Hit")
> mydata
    ID Color  Hit
  1  1   red  TRUE
  2  2 white  TRUE
  3  3   red  TRUE
  4  4     <NA> FALSE
> mydata$ID
[1] 1 2 3 4
```

```r
> AvailableTech <- c("none","keyboard","gesture","both")
> AvailableTech
[1] "none"     "keyboard" "gesture" "both"
> typeof(AvailableTech)
[1] "character"
> is.factor(AvailableTech)
[1] FALSE
> AvailableTech <- factor(AvailableTech)
> is.factor(AvailableTech)
[1] TRUE
> AvailableTech
[1] none     keyboard gesture both
Levels: both gesture keyboard none
> Difficulty <- ordered(c("easy","medium","hard"))
> Difficulty
[1] easy    medium hard
Levels: easy < hard < medium
```
R: types

Check vector type
is.numeric, is.character, is.factor, etc.

> d <- c(1,2,3,4)
> is.numeric(d)
[1] TRUE
> is.factor(d)
[1] FALSE

> e <- c("red", "white", "red", NA)
> is.numeric(e)
[1] FALSE
> is.factor(e)
[1] FALSE
> is.character(e)
[1] TRUE

> e <- factor(e)
> is.factor(e)
[1] TRUE
> is.character(e)
[1] TRUE

Change vector type
as.numeric, as.character, etc.

# from character vector to numeric vector
> num <- c("1","2","3","4")
> is.numeric(num)
[1] FALSE
> num <- as.numeric(num)
> is.numeric(num)
[1] TRUE

# from character factor to numeric vector
> fac <- factor(c("22", "44", "78"))
> is.numeric(fac)
[1] FALSE
> fac <- as.numeric(fac)
> is.numeric(fac)
[1] TRUE
# ...but the values are:
> fac
[1] 1 2 3
# Get a vector of the factor levels:
> fac <- factor(c("22", "44", "78"))
> fac <- as.numeric(levels(fac))
> fac
[1] 22 44 78
Data frame manipulations

```r
> a <- c(124, 235, 357, 489)
> b <- c(202, 306, 476, 604)
> c <- c(1, 2, 3, 4)
> d <- c("red", "white", "red", "blue")
> e <- c(TRUE, TRUE, TRUE, FALSE)
> mydata <- data.frame(a, b, c, d, e)
> names(mydata) <- c("starttime", "endtime", "ID", "Color", "Hit")
> mydata

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>124</td>
<td>202</td>
<td>1</td>
<td>red</td>
<td>TRUE</td>
</tr>
<tr>
<td>235</td>
<td>306</td>
<td>2</td>
<td>white</td>
<td>TRUE</td>
</tr>
<tr>
<td>357</td>
<td>476</td>
<td>3</td>
<td>red</td>
<td>TRUE</td>
</tr>
<tr>
<td>489</td>
<td>604</td>
<td>4</td>
<td>blue</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

> # filter out lines
> mydata <- mydata[! mydata$Color == "white", ]
```

Filter out values

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>124</td>
<td>202</td>
<td>1</td>
<td>red</td>
<td>TRUE</td>
</tr>
<tr>
<td>357</td>
<td>476</td>
<td>3</td>
<td>red</td>
<td>TRUE</td>
</tr>
<tr>
<td>489</td>
<td>604</td>
<td>4</td>
<td>blue</td>
<td>FALSE</td>
</tr>
</tbody>
</table>
### Data frame manipulations

> # add a column
> mydata$completiontime <- mydata$endtime - mydata$starttime

```r
> mydata
  starttime endtime ID Color   Hit completiontime
1        124     202  1   red  TRUE             78
3        357     476  3   red  TRUE            119
4        489     604  4  blue FALSE            115
```

> # add a column based on a condition
> mydata$difficult <- ifelse(mydata$ID > 2, "yes", "no")

```r
> mydata
  starttime endtime ID Color   Hit completiontime difficult
1        124     202  1   red  TRUE             78      no
3        357     476  3   red  TRUE            119      yes
4        489     604  4  blue FALSE            115      yes
```
R Studio: Set up a project

Create a new project

Import a dataset (e.g., allLogs.csv)

This creates a data frame in the project environment named after the file (e.g., allLogs)

The dataset is displayed in a tabular format in the dataset viewer
Create / Open an R script (e.g., helper_functions.R)
Select functions to add in the environment
Run them
Helper functions

Helper functions used in this class come from
http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2)/#helper-functions

The `summarySE` function treats factor as if they were between-subject factors.

The `summarySEwithin` function is more general and treats both between-subject and within-subject factors.

For between-subject designs, use the `summarySE` function.

For within-subject designs, use the `summarySEwithin` function.

```r
> allLogs_summary <- summarySE(allLogs, measurevar="visualSearchTime", groupvars=c("VV","OC"))
> allLogs_summary
   VV   OC   N visualSearchTime   sd    se   ci
1 Color Large  42  1217.33 167.8740 25.9035 52.3132
2 Color Medium 42  1193.74 133.3084 20.5699 41.5418
3 Color Small  42  1247.36 354.4045 54.6858 110.4402
4 ColorSize Large 42  2737.21 743.1175 114.6655 231.5717
5 ColorSize Medium 42  2481.95 881.9314 136.0849 274.8292
6 ColorSize Small  42  2393.00 1058.59 163.3435 329.8789
7 Size Large  42  1313.12 205.7042 31.7408 64.1019
8 Size Medium 43  1315.91 288.3521 43.9733 88.7417
9 Size Small  42  1270.76 220.8337 34.0754 68.8166

> allLogs_summary <- summarySEwithin(allLogs, measurevar="visualSearchTime", withinvars=c("VV","OC"), idvar="ParticipantID")
> allLogs_summary
   VV   OC   N visualSearchTime visualSearchTime_norm   sd    se   ci
1 Color Large  42  1217.33       1216.86  181.0708  27.9398  56.4256
2 Color Medium 42  1193.74       1193.26  166.1324  25.6348  51.7705
3 Color Small  42  1247.36       1246.88  417.4093  64.4077  130.0739
4 ColorSize Large 42  2737.21       2736.74  787.4788 121.5106 245.3957
...
summary provides descriptive statistics of the dataset
counts for character variables (e.g., Technique)
quartiles, min, max, mean for numeric variables (e.g., Time)

R: default summary function

summary(cellings)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ParticipantID</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>TrialID</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Block</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Trial</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Practice</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VV</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>visualSearchTime</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ErrorCount</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Length</td>
<td>379</td>
<td>22000</td>
<td>19630</td>
<td>14970</td>
<td>24240</td>
<td>82000</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>visualSearchTime</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ErrorCount</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Length</td>
<td>379</td>
<td>2.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class</td>
<td>Mode</td>
<td>character</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
<td>Max.</td>
</tr>
<tr>
<td></td>
<td>character</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>