

Lab 3

Lin Cheng

7/25/2019

```
library(Hmisc)
library(dplyr)
library(tidyr)
library(readr)
```

1 Examine and Modify an R Function

```
mysd1<-function (x, na.rm=FALSE)
sqrt(sum((x-mean(x))^2/(length(x)-1)))
```

```
mysd1(rivers)
```

```
## [1] 493.8708
```

```
sd(rivers)
```

```
## [1] 493.8708
```

2 Modify an R function to accommodate something different

```
mysd2<-function (x, na.rm=FALSE)
if (is.data.frame(x)) {
  apply(x,2,sd) #2 means to go by columns
} else {
  sqrt(sum((x-mean(x))^2/(length(x)-1)))
}
```

```
mysd2(rivers)
```

```
## [1] 493.8708
```

```
mysd2(USArrests)
```

```
##      Murder      Assault  UrbanPop      Rape  
## 4.355510 83.337661 14.474763 9.366385
```

3 The Pythagorean Theorem

```
pythag<-function(a,b)  
  sqrt(a^2+b^2)
```

```
pythag(4,5)
```

```
## [1] 6.403124
```

```
pythag(4:10,2:8)
```

```
## [1] 4.472136 5.830952 7.211103 8.602325 10.000000 11.401754 12.806248
```

```
library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##  
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':  
##  
##      format.pval, units
```

```
pythag<-function(a,b)
if (!all.is.numeric(c(a,b))){
  stop("I need numeric values to make this work")
} else if (all(c(a,b)<0)) {
  stop("values need to be positive")
} else {
  c<-sqrt(a^2+b^2)
  list("hypoteneuse"=c,"sidea"=a,"sideb"=b)
}
```

```
pythag(4,5)
```

```
## $hypoteneuse
## [1] 6.403124
##
## $sidea
## [1] 4
##
## $sideb
## [1] 5
```

```
pythag("A","B")
```

```
## Error in pythag("A", "B"): I need numeric values to make this work
```

```
pythag(-4,-5)
```

```
## Error in pythag(-4, -5): values need to be positive
```

```
pythag(4:10,2:8)
```

```
## $hypoteneuse
## [1] 4.472136 5.830952 7.211103 8.602325 10.000000 11.401754 12.806248
##
## $sidea
## [1] 4 5 6 7 8 9 10
##
## $sideb
## [1] 2 3 4 5 6 7 8
```

4 Loops vs. lapply

```
library(plyr)
```

```
##  
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:Hmisc':  
##  
##      is.discrete, summarize
```

```
attach(baseball)  
apply(baseball, 2, class)
```

```
##           id           year           stint           team           lg           g  
## "character" "character" "character" "character" "character" "character"  
##           ab           r           h           X2b           X3b           hr  
## "character" "character" "character" "character" "character" "character"  
##           rbi           sb           cs           bb           so           ibb  
## "character" "character" "character" "character" "character" "character"  
##           hbp           sh           sf           gidp  
## "character" "character" "character" "character"
```

```
baseball_1<-baseball[,c(1,2,6:12)]
```

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':  
##  
##      arrange, count, desc, failwith, id, mutate, rename, summarise,  
##      summarize
```

```
## The following objects are masked from 'package:Hmisc':  
##  
##      src, summarize
```

```
## The following objects are masked from 'package:stats':  
##  
##      filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##      intersect, setdiff, setequal, union
```

```
baseball2<- baseball_1 %>%
  group_by(year) %>%
  do(mutate(. ,g=g/max(g),ab=ab/max(ab),r=r/max(r),h=h/max(h),X2b=X2b/max(X2b),X3b=X3b/max(X3b),hr=hr/max(hr)))
head(baseball2,n=7)
```

```
## # A tibble: 7 x 9
## # Groups:   year [1]
##   id      year      g      ab      r      h      X2b      X3b      hr
##   <chr>    <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 ansonca01  1871 0.758 0.741 0.644 0.672 1      0.429 0
## 2 forceda01  1871 0.970 1      1      0.776 0.818 0.571 0
## 3 mathebo01  1871 0.576 0.549 0.333 0.414 0.273 0.143 0
## 4 startjo01  1871 1      0.994 0.778 1      0.455 0.143 0.333
## 5 suttoez01  1871 0.879 0.790 0.778 0.776 0.273 1      1
## 6 whitede01  1871 0.879 0.901 0.889 0.810 0.545 0.714 0.333
## 7 yorkto01  1871 0.879 0.895 0.8    0.638 0.455 1      0.667
```

```
str(baseball2,give.attr=FALSE)
```

```
## Classes 'grouped_df', 'tbl_df', 'tbl' and 'data.frame':  21699 obs. of  9 variables:
## $ id : chr  "ansonca01" "forceda01" "mathebo01" "startjo01" ...
## $ year: int  1871 1871 1871 1871 1871 1871 1871 1871 1872 1872 1872 ...
## $ g : num  0.758 0.97 0.576 1 0.879 ...
## $ ab : num  0.741 1 0.549 0.994 0.79 ...
## $ r : num  0.644 1 0.333 0.778 0.778 ...
## $ h : num  0.672 0.776 0.414 1 0.776 ...
## $ X2b : num  1 0.818 0.273 0.455 0.273 ...
## $ X3b : num  0.429 0.571 0.143 0.143 1 ...
## $ hr : num  0 0 0 0.333 1 ...
```

5 Interacting with files outside of R

A

```
someURL<-"http://www.stat.ucla.edu/~vlew/datasets/spssSTUFF.zip"
download.file(someURL,"spssSTUFF.zip")
unzip("spssSTUFF.zip", exdir = "spssSTUFFdata")
list.files("spssSTUFFdata")
```

```
## [1] "__MACOSX"      "yrbs2013.sav" "yrbss.sav"
```

B

```
library(readr)
```

```
read_csv(list.files(path='/Users/Caroline.C',pattern="death+"))
```

```
## # A tibble: 1 x 1  
##   death00.csv  
##   <chr>  
## 1 death01.csv
```