OPEN HPI Hasso Plattner Institut

Fundamentals of Programming for Digital Health - Part 1: R

Berry Boessenkool & Team
Berry.Boessenkool@hpi.de

2022-12-07 09:48

Lucius Annaeus Seneca (4BC - 65AD):

to err is human

to  is better

Pirates need to know many languages

but their favorite language is 

- Did you use `rt_create()` as intended?
- Lab open during the week
- Exercises have no submission deadline
- submitting is optional (use it to see progress in openHPI)
- multiple submits are OK
- scripts must be executable (or later tasks are not solved)
- 2.1 Syntax task T5: function / command / statement / expression
- 2.3 Vectors T6: 3 ways to add 77 to an existing vector `x`

```
x[length(x) + 1] <- 77
x <- append(x, 77)
x <- c(x, 77)
```

- 2.4 Stats T3: If you used the `replace` function - how often?
- Which task in the exercises was the hardest?
- Questions

René Descartes (1596 - 1650):

I think, therefore I am

We think, therefore we 

- Exercises have no submission deadline
- submitting is optional (use it to see progress in openHPI)
- multiple submits are OK
- scripts must be executable (or later tasks are not solved)
- Skip difficult tasks first, get back to them later (exam!)
- Rstudio reproducibility settings
- Are there too many / too few exercises?

- Ex 2.6 Logic T3: Alternative to individually compare `3,"b","2",`

```
sort(c(3,"b","2","a","B",1))
## [1] "1" "2" "3" "a" "b" "B"
```

- Ex 2.7 Char T4: two ways to create "letter_A", "letter_B", ... "_Z"

```
paste0("letter_", LETTERS)
paste("letter", LETTERS, sep="_")
# handy in function with customizable sep
```

- Ex 2.7 char T11: `w` entries beginning with `b`

```
function(b,w) w[startsWith(w,b)]
function(b,w) grep(paste0("^",b), w, value=TRUE)
```

- Ex 2.8 Factors T6: codegolf bonus: shortest solution for

```
tapply(income, eyeColor, function(x) c(min(x), median(x),
                                       mean(x), max(x)))
tapply(income, eyeColor, function(x) summary(x)[-c(2,5)])
```

- Ex 2.7 Char T13: all dates with minus

```
date_with_minus <- function(x)
 {
 sel <- nchar(x)<10
 x[sel] <- paste(substr(x[sel],1,4), substr(x[sel],5,6),
                 substr(x[sel],7,8), sep="-")
 x
 }
```

```
date_with_minus <- function(x) {
 x <- gsub("-", "", x)
 paste(substr(x,1,4),substr(x,5,6),substr(x,7,8), sep="-")
}
```

```
function(x) sub("(\\d{4})-?(\\d{2})-?(\\d{2})",
                "\\1-\\2-\\3", x)
```

`\d` looks for any digit

`\d{4}` looks for 4 digits

`-?` looks for zero or one `-`

To get `\`, R needs `"\\"`

`(...)` groups `...` together

`\1` returns first captured group

Famous musicians have endorsed this course:

Madonna: Don't cry for me  gentina

Cutting Crew: I just died in your  ms tonight

Lady Gaga: Pok  face

Billy Joel: We didn't start the fi 

Lee Marvin: I was born under a wandering st 

Bonnie Tyler: Turn  ound

Nirvana: Here we  now

```
# given this functioning code:
pressure[["temperature"]]
##  [1]   0  20  40  60  80 100 120 140 160 180 200 220
## [13] 240 260 280 300 320 340 360
# what class do you expect?
class(pressure)
## [1] "data.frame"
```

The syntax works because data.frames (along with most advanced objects) are internally a list.

Per style guide (for humans!), we have agreed to use `[[` only for lists, not for data.frames.

```
# what are the two 'correct' ways to select temperature?
pressure[ ,"temperature"]
pressure$temperature
```

Ex 3.1 DataFrames T8: column selection with error message

```
iris$non_existing_column
## NULL
```

```
iris[,"non_existing_column"]
## Error in `[.data.frame`(iris, , "non_existing_column"):
undefined columns selected
```

Fix misspelled column names with the `TAB` key

When will `sapply` fail to simplify the result of applying a function to a list? -> if the output length is not always the same

```
sapply(Harman23.cor, length)
##    cov center  n.obs
##     64      8      1
```

can be abbreviated to

```
lengths(Harman23.cor)
##    cov center  n.obs
##     64      8      1
```

```
head(esoph) # age, alcohol, tobacco, cancer cases
##   agegp     alcgp    tobgp ncases ncontrols
## 1 25-34 0-39g/day 0-9g/day      0        40
## 2 25-34 0-39g/day    10-19      0        10
## 3 25-34 0-39g/day    20-29      0         6
## 4 25-34 0-39g/day      30+      0         5
## 5 25-34     40-79 0-9g/day      0        27
## 6 25-34     40-79    10-19      0         7
```

```
summary(esoph)
##    agegp          alcgp          tobgp          ncases          ncontrols
##  25-34:15   0-39g/day:23   0-9g/day:24   Min.   : 0.000   Min.   : 0.000
##  35-44:15   40-79    :23   10-19   :24   1st Qu.: 0.000   1st Qu.: 1.000
##  45-54:16   80-119   :21   20-29   :20   Median : 1.000   Median : 4.000
##  55-64:16   120+     :21   30+     :20   Mean   : 2.273   Mean   : 8.807
##  65-74:15                                3rd Qu.: 4.000   3rd Qu.:10.000
##  75+  :11                                Max.   :17.000   Max.   :60.000
```

Which row of `esoph` has the maximum ncases?

```
which.max(esoph$ncases)
## [1] 67
```

3.2 Matrices T8: mean values of the appropriate `iris` columns

```
colMeans(iris[,-5])
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##     5.843333     3.057333     3.758000     1.199333
```

Manually, create a data.frame with the two columns
- AAA, with the values 1,2,3,4,1,2,3,4
- BBB, with the values 8,7,..,1

Remember: vector recycling saves you the trouble of using the `rep` function.

```
data.frame(AAA=1:4, BBB=8:1)
##   AAA BBB
## 1   1   8
## 2   2   7
## 3   3   6
## 4   4   5
## 5   1   4
## 6   2   3
## 7   3   2
## 8   4   1
```

3.4 Arrays T9: sum of Titanic Deceased + Survivors (two approaches)

```
Titanic[,,,"No"] + Titanic[,,,"Yes"]
apply(Titanic, 1:3, sum)
```

- Feedback to shape next weeks: bit.ly/feedbackR

- Retake exercises by clicking *reset* in CodeOcean, copy to your script
- Questions / hardest task / task amount
- If a script cannot be run, outcomment the failing part (`CTRL` + `SHIFT` + `C`)

I have an idea:
I'll write an R-ticle.
It's going to be R-some!

I know, I should be R-rested for such bad puns.
After all, I'm not a famous R-tist or anything.
But I don't want to go to prison. There are R-sonists there!
I once saw a movie about how they set everything on fi-R.
It was so horrible it had to be rated R.

I probably better R-range a backup.
Something that makes me popul-R.
Then I might escape that R-ful fate...
I know! I'll teach an R course!

FP35tutorial: Delete the second till fourth element of the list 'nileEVS'.

```
nileEVS <- nileEVS[-2]
nileEVS <- nileEVS[-2]
nileEVS <- nileEVS[-2]
```

```
Nile <- nileEVS(-c(2,3,4))
```

```
nileEVS <- nileEVS[-c(2,3,4)]
```

```
nileEVS <- nileEVS[-c(2:4)]
```

```
nileEVS <- nileEVS[-(2:4)]
```

```
nileEVS$parameter <- NULL
nileEVS$distselector <- NULL
nileEVS$returnlev <- NULL
```

```
nileEVS[2:4] <- NULL
```

FP35tutorial: filtering to create a subset from dataframes like the built-in dataset 'iris'. The subset should only contain the entries for the versicolor Species.

```
only_vers <- function(df) df[df$Species == "versicolor", ]
```

ex 4.1 Read T9 (two different column separators)

```
days1 <-  readLines("ex_04_1_a09_days1.txt")
days1 <- read.table(text=gsub(":","-",days1), sep="-")
```

one-line:

```
days1 <- read.table(text=gsub("[:-]", " ",
                              readLines("days1.txt")))
```

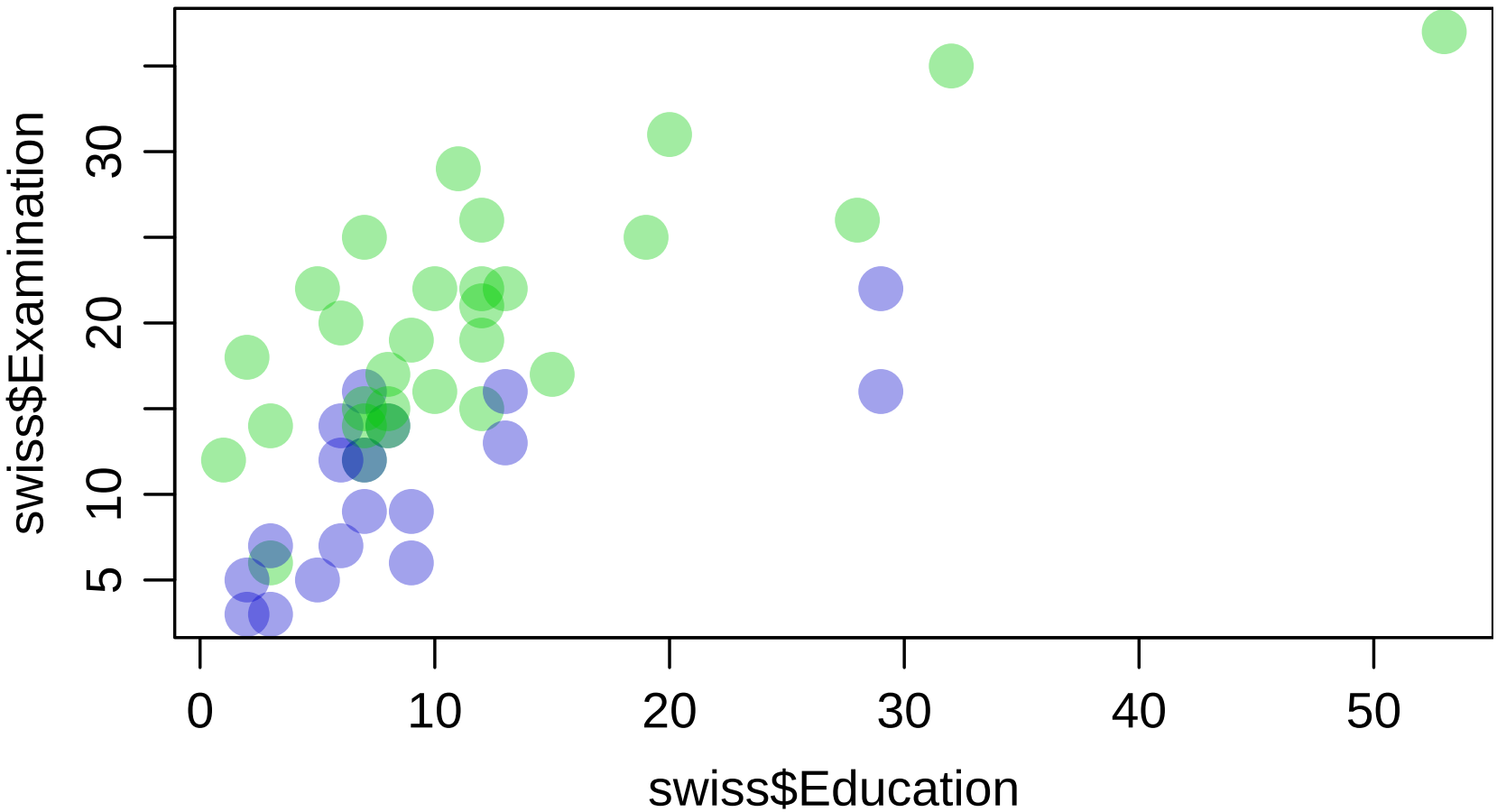
- feedback implemented?
- Run selection of code (even in comments)
- Questions

I went to an R-chery club last week.
Turns out they love memes!
We had a discussion about R-row score calculation. This came up:

Facebook R memes group

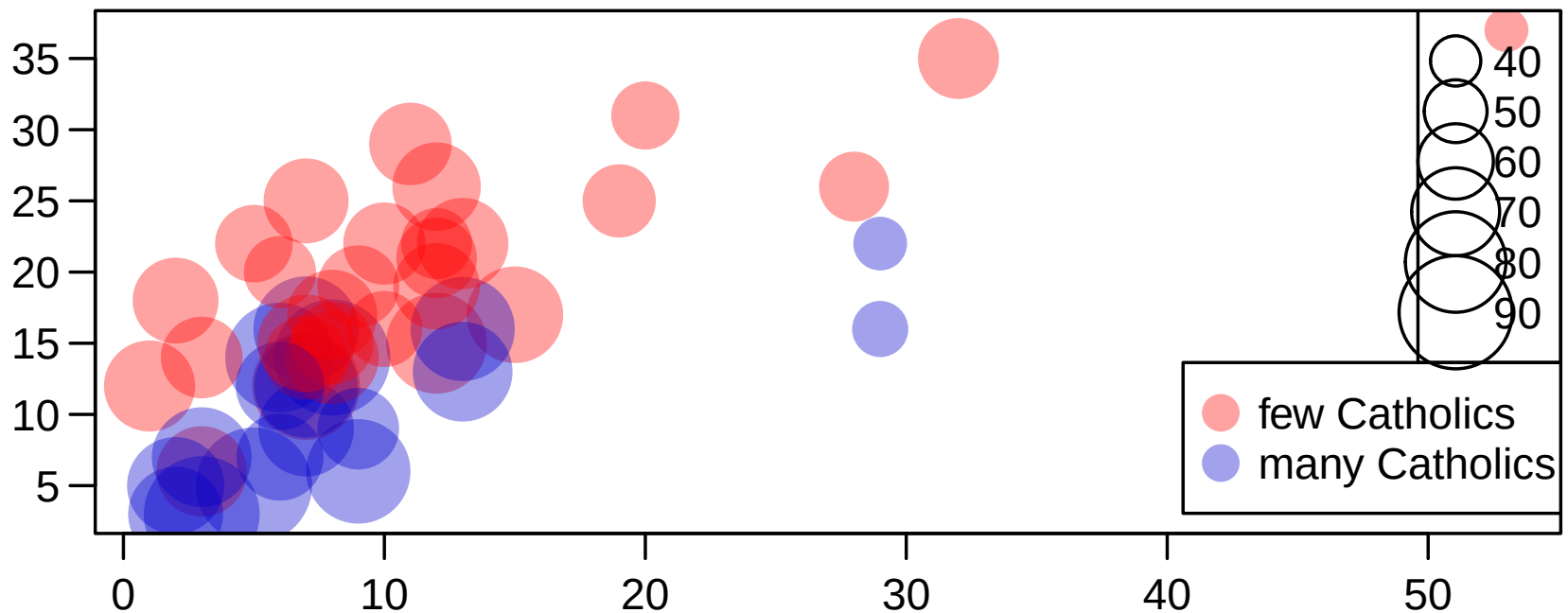
Custom colors depending on swiss Catholics group:

```
plot(x=swiss$Education, y=swiss$Examination, pch=16, cex=2,
     col=c("#00CD005D","#0000CD5D")[(swiss$Catholic>50)+1] )
```




```
plot(x=swiss$Education, y=swiss$Examination, pch=16, cex=2,
     col=ifelse(swiss$Catholic>50,"#0000CD5D","#00CD005D") )
```

```r
swissCols <- c("#FF00005D","#0000CD5D")
plot(x=swiss$Education, y=swiss$Examination, pch=16,
     cex=swiss$Fertility/15,
     col=ifelse(swiss$Catholic>50,swissCols[2],swissCols[1]) )
leg <- legend("topright", legend=seq(40,90,10), pch=1,
       pt.cex=seq(40,90,10)/15)
# for coordinates of next legend
legend(leg$rect$left-9, leg$rect$top - leg$rect$h,
       c("few Catholics", "many Catholics"), pch=16,
       pt.cex=2, col=swissCols)
```

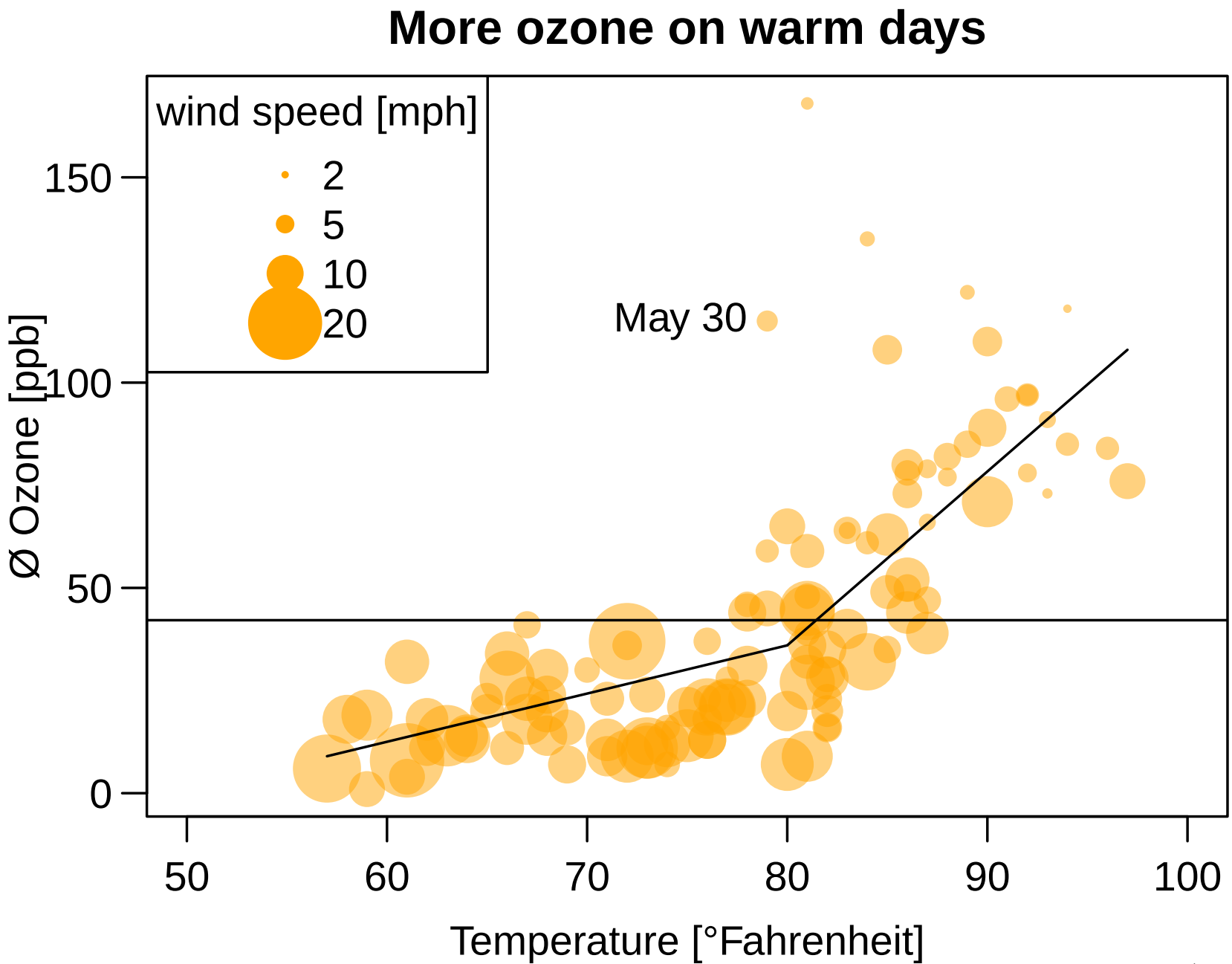


Tutorial exercise

```
segments(x0=c(57,80), y0=c( 9, 36),
         x1=c(80,97), y1=c(36,108))
```

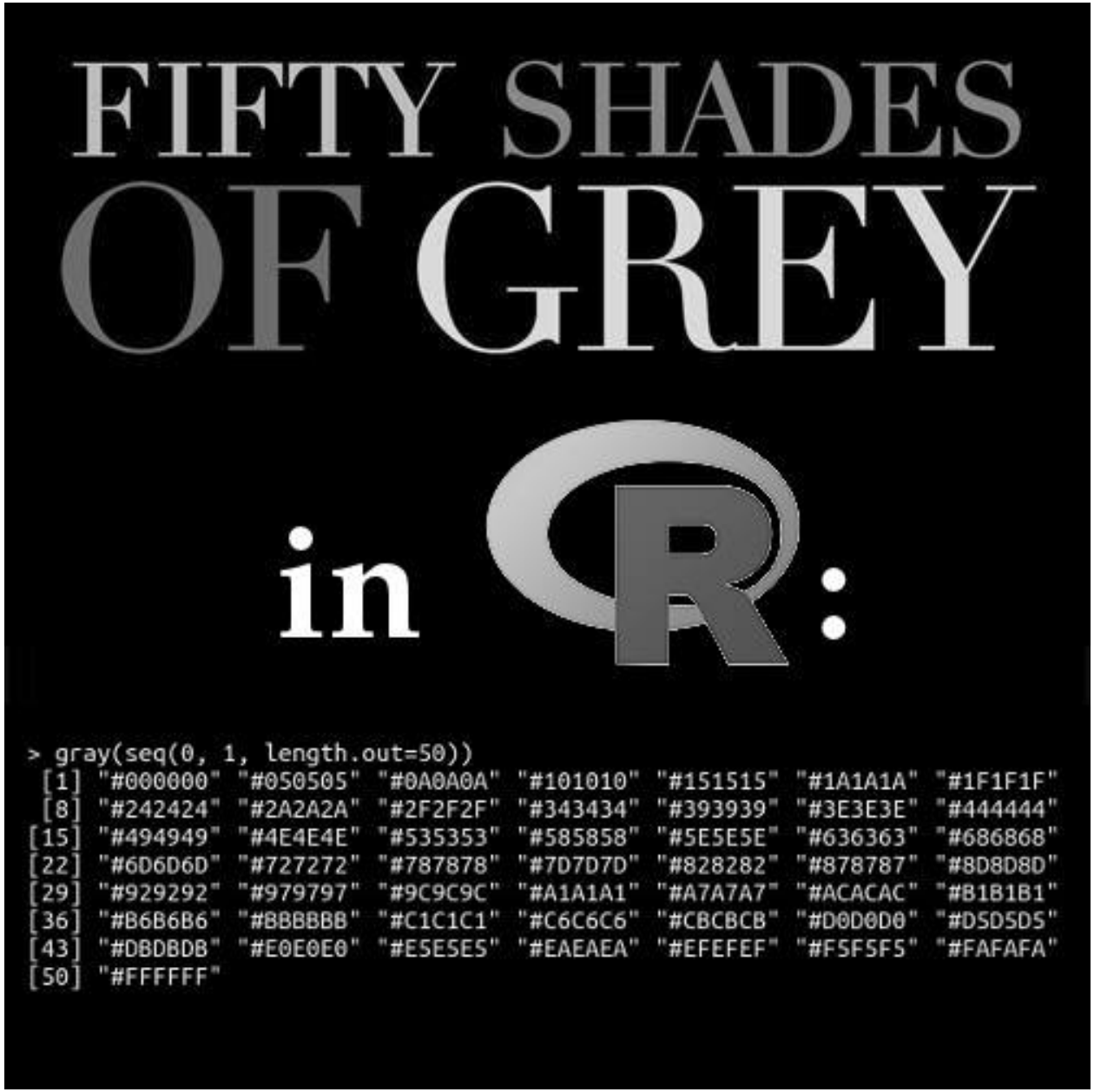
For this simple example, there is also simpler code:

```
lines(x=c(57, 80, 97), y=c(9, 36, 108))
```


More ozone on warm days
wind speed [mph]
2
5
10
20
May 30
Ø Ozone [ppb]
0
50
100
150
50
60
70
80
90
100
Temperature [°Fahrenheit]

Coding makes the blood rush through my  teries.

Add in memes and my  ms lift themselves up!



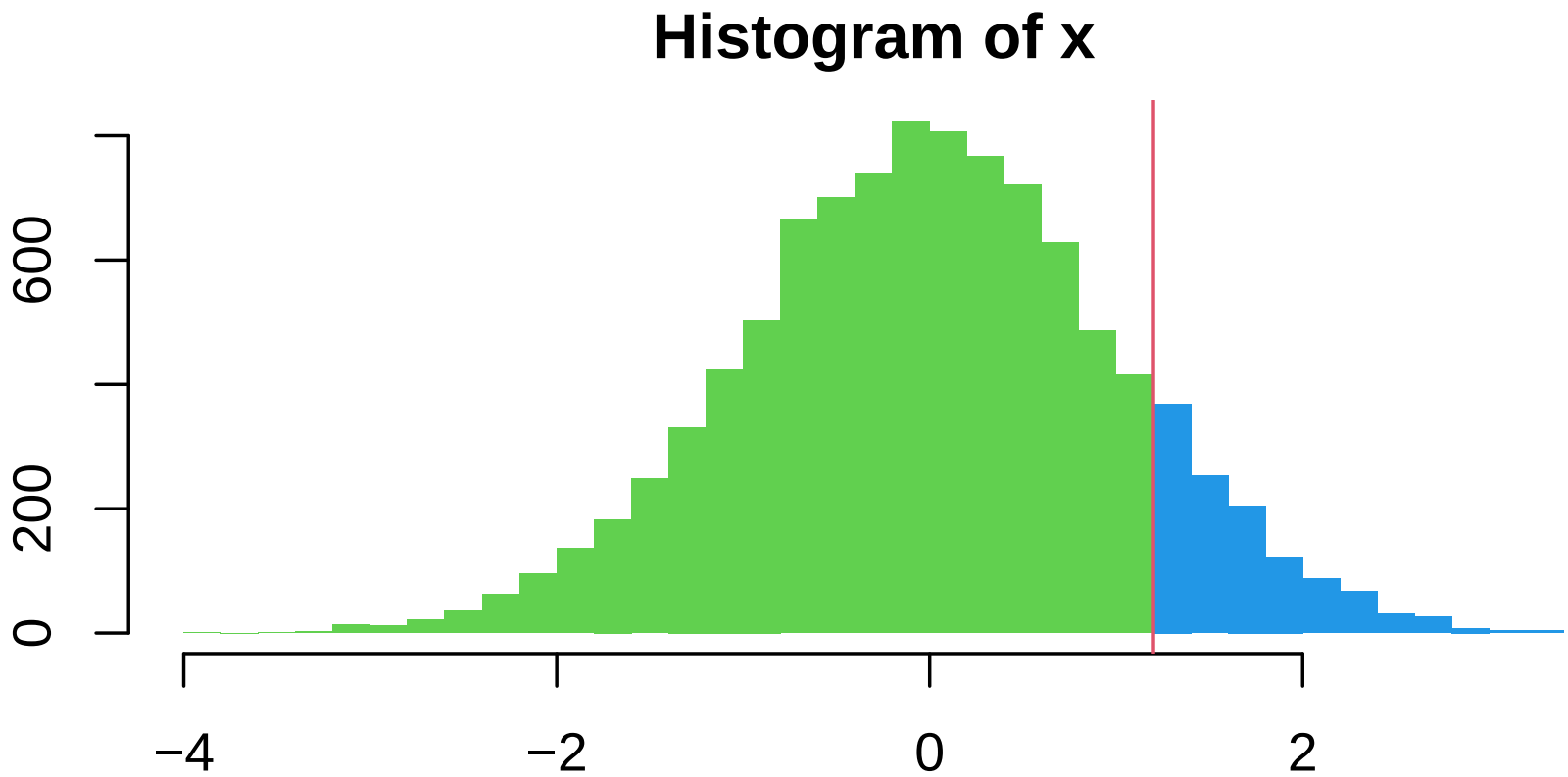
reddit.com/r/rstatsmemes

reddit.com/r/rstatsmemes

facebook.com/Rmemes0

```
fn <- system.file("success.jpg", package="meme")
meme::meme(fn, "Luckily", "we can create memes in R!", size=5)
```

```r
hist2cols <- function(x, split, cols, border=NA, ...)
  {
  h <- hist(x, ..., plot=FALSE)
  big <- h$breaks>=split
  big <- head(big, -1) # exclude last value (right end of bin)
  hist(x, col=ifelse(big, cols[2], cols[1]), border=border, ...)
  }
hist2cols(rnorm(1e4), 1.2, cols=3:4, breaks=30) ;  abline(v=1.2,col=2)
```



- Exam with more tasks + relative grade?
- Questions

This is .

There is no if. Only how.

$\sim$ *Simon 'Yoda' Blomberg, R-help (April 2005)*

Tshirt

 -kenntnis ist der erste Schritt auf dem Weg zur Besserung

Awareness ( -knowledge) is the first step towards improvement

Expanding on 6.1 conditional return

```
stat <- function(x, fun) { # conditional returns need no 'else'
  if(fun=="mean"  ) return(  mean(x))
  if(fun=="median") return(median(x))
  if(fun=="max"   ) return(   max(x))
  message(paste("method ",fun," is not implemented.",sep="'"))
  }
```

How could this still be improved further?

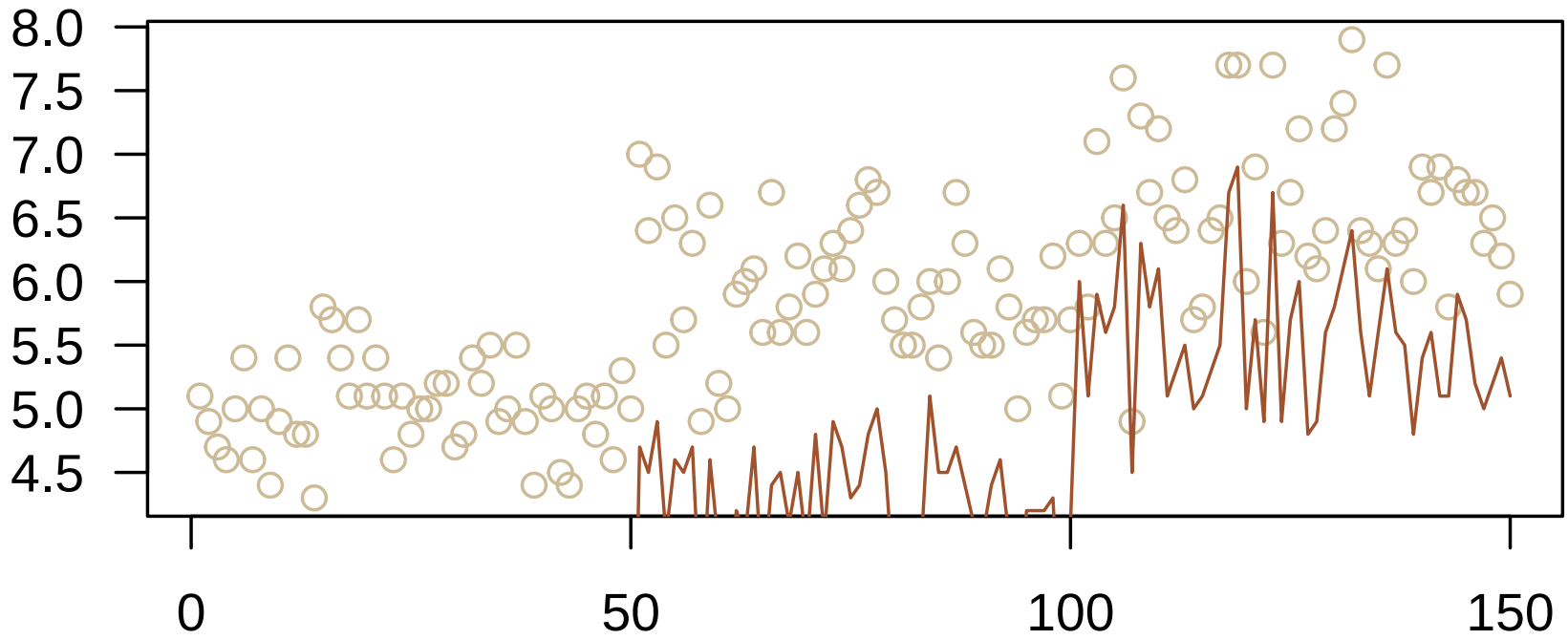
```
stat <- function(x, fun)  switch(fun,
   mean   =  mean(x),
   median=median(x),
   max    =   max(x),
   message("method '",fun,"' is not implemented.")
   )
stat(1:10, "max") # 10
stat(1:10, "sum") # method 'sum' is not implemented.
```

Shorter idea for this particular example?

```
stat <- function(x, fun) fun(x) # no custom warning message
stat(1:10, sum)
```

How to set up an empty plot with enough space for later lines?

```
plot(iris$Sepal.Length, col="wheat3")
lines(iris$Petal.Length, col="sienna")
```




```
plot(iris$Sepal.Length, ylim=range(iris[,1:4]), type="n")
```

Sample solution to Ex 6.2 functions T2-5 `circle` function

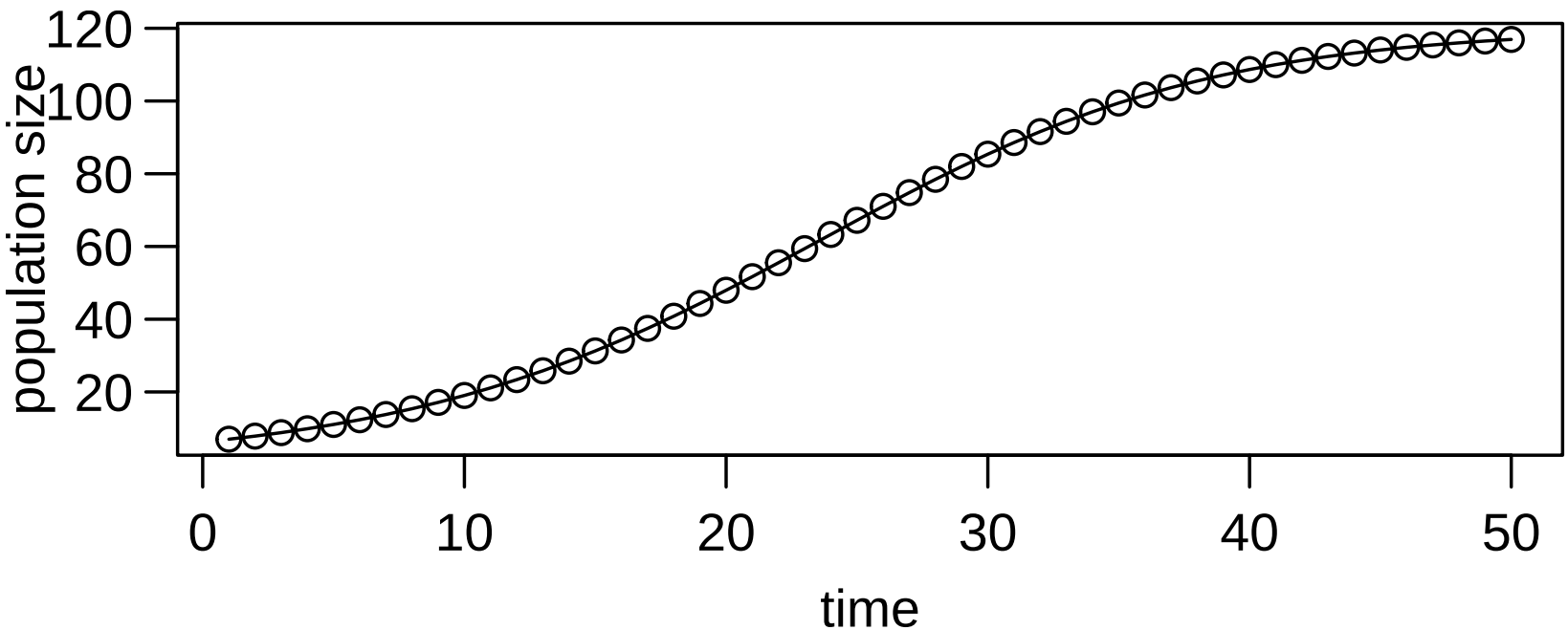
```
circle <- function(x,y,r, ...)
  {
  checkInput <- function(i)
    {
    v <- get(i)
    if(!is.numeric(v)) stop(i," must be numeric, not ", class(v))
    if(length(v)>1){warning("Only the first element of ",i," is used")
                    v <- v[1]}
    v
    }
  x <- checkInput("x")
  y <- checkInput("y")
  r <- checkInput("r")
  p <- seq(0, 2*pi, len=50)
  cx <- x+r*cos(p) ; cy <- y+r*sin(p)
  polygon(cx, cy, ...)
  invisible(data.frame(x=cx, y=cy))
  }
```

Alternative solution to Ex 6.2 functions T2-5 `circle` function

```
checkInput <- function(v)
  {
  n <- deparse(substitute(v)) # do not evaluate v, get the input code
  if(!is.numeric(v)) stop(n," must be numeric, not ", class(v))
  if(length(v)>1){warning("Only the first element of ",n," is used")
                  v <- v[1]}
  v
  }
circle <- function(x,y,r, ...)
  {
  x <- checkInput(x)
  y <- checkInput(y)
  r <- checkInput(r)
  p <- seq(0, 2*pi, len=50)
  cx <- x+r*cos(p) ; cy <- y+r*sin(p)
  polygon(cx, cy, ...)
  invisible(data.frame(x=cx, y=cy))
  }
```

The last task on the midterm will require a `for` loop to solve.
Since it is a real-world application, it needs a few words as intro.
To save you time in the midterm, you can already read that here.

In logistic growth models, a population e.g of bacteria grows towards the capacity `K` of the ecosystem (e.g petri dish).
At each time step, the change in population size is determined with `r*N*(K-N)/K` with `N` being the current population size.
Hence e.g. `N[9] = N[8] + r*N[8]*(K-N[8])/K`
Read more at SolvingDifferentialEquations.pdf



- R draw christmas tree (file to be sourced)
- midterm prep extra tutorial session?
- midterm questions: post **DM** in chat -> breakout session
- graphical task very free, see Exam Info
- Questions