Problem Set 3

Alexander Brandt
SID: 24092167

September 30 2015

1 Debugging Reading

I chose to read option ii) as I am a huge fan of the Software Carpentry Foundation, as well as Titus Brown's work in general. My question was "After reading the Software Carpentry Foundation's paper on 'Best Practices for Scientific Computing', I found myself wondering about section 9, which says 'Document design and purpose, not mechanics.' When is code written in such a way that it is self-evident to an outside reader/developer? What are the hallmarks of code written in this style? I think the suggestion has good intentions, but often discerning the difficulty level associated block of code is challenging for developers. One person's 'easy' is often another person's 'hard,' and so in this way applying the principle of "the onerous should always be on the author to convince his or her peers of that" seems to be to be a recipe for disaster in the hands of some very well-meaning scientific programmers that I know that have forgotten how difficult it can be to work within a certain language, leading to less commentary and comprehension, not more. For example, would explaining a choice of data structure in a language (a dict vs. a list in Python, for example) be violating this principle?"

2 An Analysis of the Presidental Election Debates

My question is:

```
library(stringr)
library(XML)
library(curl)

split_block <- function (list_of_strings_solid) {
  current_name = ""
  current_block = ""
  my_list = list()
```

```
  list_of_strings <- unlist(strsplit(list_of_strings_solid,"\n"))
  # XML example class notes -- different HTML features
  for (i in 1:length(list_of_strings)) {
    if (((toupper(list_of_strings[[i]]) == list_of_strings[[i]]) &&
         !grepl("^\\([A-Z]+\\)$",list_of_strings[[i]]))) {
      # This is a "caps line" which doesn't contain useful information about
      # the text of the debate.  Mostly filler
      next
    }
    if (grepl("END",list_of_strings[[i]])) { break }
    name <- str_match(list_of_strings[[i]],regex("([A-Z]+):"))[,2]
    if ((!is.na(name)) && name != current_name) {
      if (current_name != "") {
        my_list[[current_name]] <- c(my_list[[current_name]],
        str_replace_all(current_block,paste(current_name,": ",sep=""),
        ""))
      }
      current_name <- name
      current_block <- ""
      # ... and maybe add a new block to the list?
    }
    if (length(current_name) != 0) {
      if (current_block != "") {
        current_block <- paste(current_block,
                               list_of_strings[[i]], sep=" ")
      }
      else {
        current_block <- list_of_strings[[i]]
      }
    }
  }
  my_list[[current_name]] <- c(my_list[[current_name]],
  str_replace_all(current_block,
                  paste(current_name,": ",sep=""),""))
  return(my_list)
}

create_debate_text <- function(file_url)
{
  xml_handle <- htmlParse(file_url)
  v <- xpathSApply(xml_handle,
                   "//div[@id = 'content-sm']",xmlValue)
  text_data <- lapply(v,str_replace_all,
                      "([A-Z]+:)","\n\n\\1")
  return(text_data[[1]])
```

```
}

debate_summary <- function(file_url)
{
text_data <- create_debate_text(file_url)

events = list()
debate_blocks <- split_block(text_data)
for (n in names(debate_blocks))
{
   events[[n]] <- table(
     str_extract_all(
       paste(debate_blocks[[n]],collapse=" ")
       ,"\\([A-Za-z]+\\)"))
   debate_blocks[[n]] <- lapply(debate_blocks[[n]],
                                str_replace_all,
                                "\\([A-Za-z]+\\)","")
}

words = list()
sentence = list()
word_counts = list()
patterns = c("I[^a-zA-Z]", "we[^a-zA-Z]", "America(n)?[^a-zA-Z]",
             "democra(cy|tic)[^a-zA-Z]", "republic[^a-zA-Z]",
             "Democrat(ic)?[^a-zA-Z]","Republican[^a-zA-Z]",
             "free(dom)?[^a-zA-Z]", "war[^a-zA-Z]",
             "God(?! bless)[^a-zA-Z]",
             "(Jesus|Christ|Christian)[^a-zA-Z]",
             "God bless[^a-zA-Z]")
for (n in names(debate_blocks))
{
  to_analyze <- paste(debate_blocks[[n]],collapse=" ")
  for (pattern in patterns)
  {
    word_counts[[n]][[pattern]] <- str_count(to_analyze,pattern)
  }
  names(word_counts[[n]]) <- c("I","we","America{n}",
                               "democra{cy,tic}","republic",
                               "Democrat{,ic}","Republican",
                               "free{,dom}","war",
                               "God (only)","God Bless",
                               "{Jesus, Christ, Christian}")
  # print(word_counts)
  words[[n]] <- str_extract_all(to_analyze,
'(([:alpha:]+(\'([:alpha:]+)?)?)|([:digit:]+(,([:digit:]+)?)?))')
```

```
  sentence[[n]] <- str_extract_all(to_analyze,
"([:alpha:])(([:alpha:]|[:space:]|[:digit:]|\'|,|-)*)(\\.|\\?|\\!)")
  print(paste("The average word length of ",
              n,"'s speach is:",sep = ""))
  print(mean(rapply(words[[n]],nchar)))
  print(paste("The number of characters (in the words) in ",
              n,"'s speach is:",sep = ""))
  print(sum(rapply(words[[n]],nchar)))
  print(paste("The number of words in ", n,"'s speach is:",
              sep = ""))
  print(length(unlist(words[[n]])))
  print(paste("The buzzwords in ", n,"'s speach is:",
              sep = ""))
  print(word_counts[[n]])
  print(paste("Event occurences in ", n,"'s speach is:",
              sep = ""))
  print(events[[n]])
}
return(debate_blocks)
}

menu_url="http://www.debates.org/index.php?page=debate-transcripts"
menu_xml_handle <- htmlParse(menu_url)
menu_nodes <- getNodeSet(menu_xml_handle,"//a[@href]")
all_debate_links <- xpathSApply(
  menu_xml_handle, "//a[@href]", xmlGetAttr, 'href')
years <- c("2012","2008","2004","2000","1996")

year_reg <- paste("(",
                  paste(paste(years,collapse="|"),
                        ").+(First)",
                        sep=""),
                  sep="")

my_debate_links <- all_debate_links[grepl(
  year_reg,
  sapply(menu_nodes,xmlValue))]
debate_blocks_list = list()
i <- 1
for (year in years)
{
  print(paste("The statistics for the first debate in",year,"..."))
  debate_blocks_list[[year]] <- debate_summary(my_debate_links[i])
  i <- i + 1
  cat("\n\n")
```

```
}
```

```
## [1] "The statistics for the first debate in 2012 ..."
## [1] "The average word length of LEHRER's speach is:"
## [1] 4.399606
## [1] "The number of characters (in the words) in LEHRER's speach is:"
## [1] 6705
## [1] "The number of words in LEHRER's speach is:"
## [1] 1524
## [1] "The buzzwords in LEHRER's speach is:"
##                          I                         we
##                         17                         20
##                America{n}            democra{cy,tic}
##                          1                          0
##                  republic              Democrat{,ic}
##                          0                          1
##                Republican                 free{,dom}
##                          1                          0
##                        war                 God (only)
##                          0                          0
##                 God Bless {Jesus, Christ, Christian}
##                          0                          0
## [1] "Event occurences in LEHRER's speach is:"
##
##  (APPLAUSE) (CROSSTALK) (inaudible)
##           1          10           4
## [1] "The average word length of OBAMA's speach is:"
## [1] 4.450814
## [1] "The number of characters (in the words) in OBAMA's speach is:"
## [1] 32531
## [1] "The number of words in OBAMA's speach is:"
## [1] 7309
## [1] "The buzzwords in OBAMA's speach is:"
##                          I                         we
##                        119                        172
##                America{n}            democra{cy,tic}
##                         18                          0
##                  republic              Democrat{,ic}
##                          0                          4
##                Republican                 free{,dom}
##                          5                          3
##                        war                 God (only)
##                          2                          0
##                 God Bless {Jesus, Christ, Christian}
##                          0                          0
```

```
## [1] "Event occurences in OBAMA's speach is:"
##
## (CROSSTALK)  (LAUGHTER)
##           4           3
## [1] "The average word length of ROMNEY's speach is:"
## [1] 4.322593
## [1] "The number of characters (in the words) in ROMNEY's speach is:"
## [1] 33807
## [1] "The number of words in ROMNEY's speach is:"
## [1] 7821
## [1] "The buzzwords in ROMNEY's speach is:"
##                          I                         we
##                        217                         94
##                America{n}            democra{cy,tic}
##                         34                          1
##                   republic              Democrat{,ic}
##                          0                          4
##                Republican                 free{,dom}
##                          5                          7
##                        war                 God (only)
##                          0                          0
##                 God Bless {Jesus, Christ, Christian}
##                          0                          0
## [1] "Event occurences in ROMNEY's speach is:"
##
## (CROSSTALK) (inaudible)  (LAUGHTER)
##          11           2           1
##
##
## [1] "The statistics for the first debate in 2008 ..."
## [1] "The average word length of LEHRER's speach is:"
## [1] 4.317448
## [1] "The number of characters (in the words) in LEHRER's speach is:"
## [1] 5617
## [1] "The number of words in LEHRER's speach is:"
## [1] 1301
## [1] "The buzzwords in LEHRER's speach is:"
##                          I                         we
##                         14                          8
##                America{n}            democra{cy,tic}
##                          0                          0
##                   republic              Democrat{,ic}
##                          0                          1
##                Republican                 free{,dom}
##                          1                          0
```

```
##                         war                 God (only)
##                           0                          0
##                  God Bless {Jesus, Christ, Christian}
##                           0                          0
## [1] "Event occurences in LEHRER's speach is:"
##
##  (APPLAUSE) (CROSSTALK)  (LAUGHTER)
##           1           4           1
## [1] "The average word length of OBAMA's speach is:"
## [1] 4.368359
## [1] "The number of characters (in the words) in OBAMA's speach is:"
## [1] 33383
## [1] "The number of words in OBAMA's speach is:"
## [1] 7642
## [1] "The buzzwords in OBAMA's speach is:"
##                           I                         we
##                         145                        220
##                 America{n}            democra{cy,tic}
##                          13                          1
##                    republic              Democrat{,ic}
##                           0                          0
##                 Republican                 free{,dom}
##                           2                          2
##                         war                 God (only)
##                          12                          0
##                  God Bless {Jesus, Christ, Christian}
##                           0                          0
## [1] "Event occurences in OBAMA's speach is:"
##
## (CROSSTALK)        (ph)
##           3           1
## [1] "The average word length of MCCAIN's speach is:"
## [1] 4.412685
## [1] "The number of characters (in the words) in MCCAIN's speach is:"
## [1] 31586
## [1] "The number of words in MCCAIN's speach is:"
## [1] 7158
## [1] "The buzzwords in MCCAIN's speach is:"
##                           I                         we
##                         213                        141
##                 America{n}            democra{cy,tic}
##                          18                          1
##                    republic              Democrat{,ic}
##                           0                          1
##                 Republican                 free{,dom}
```

```
##                           2                           3
##                         war                 God (only)
##                           5                           0
##                  God Bless {Jesus, Christ, Christian}
##                           0                           0
## [1] "Event occurences in MCCAIN's speach is:"
##
## (CROSSTALK)  (LAUGHTER)        (ph)       (sic)
##           1           1           1           1
##
##
## [1] "The statistics for the first debate in 2004 ..."
## [1] "The average word length of LEHRER's speach is:"
## [1] 4.715942
## [1] "The number of characters (in the words) in LEHRER's speach is:"
## [1] 6508
## [1] "The number of words in LEHRER's speach is:"
## [1] 1380
## [1] "The buzzwords in LEHRER's speach is:"
##                           I                          we
##                           9                           2
##                 America{n}            democra{cy,tic}
##                           2                           1
##                   republic              Democrat{,ic}
##                           0                           1
##                 Republican                 free{,dom}
##                           1                           0
##                         war                 God (only)
##                           3                           0
##                  God Bless {Jesus, Christ, Christian}
##                           0                           0
## [1] "Event occurences in LEHRER's speach is:"
##
##  (APPLAUSE) (CROSSTALK)
##           2           1
## [1] "The average word length of KERRY's speach is:"
## [1] 4.291059
## [1] "The number of characters (in the words) in KERRY's speach is:"
## [1] 30621
## [1] "The number of words in KERRY's speach is:"
## [1] 7136
## [1] "The buzzwords in KERRY's speach is:"
##                           I                          we
##                         197                         114
##                 America{n}            democra{cy,tic}
```

```
##                          43                           2
##                    republic              Democrat{,ic}
##                           0                           0
##                  Republican                  free{,dom}
##                           1                           3
##                         war                  God (only)
##                          35                           0
##                   God Bless {Jesus, Christ, Christian}
##                           0                           1
## [1] "Event occurences in KERRY's speach is:"
##
## (LAUGHTER)
##          2
## [1] "The average word length of BUSH's speach is:"
## [1] 4.319169
## [1] "The number of characters (in the words) in BUSH's speach is:"
## [1] 27444
## [1] "The number of words in BUSH's speach is:"
## [1] 6354
## [1] "The buzzwords in BUSH's speach is:"
##                           I                          we
##                         179                         122
##                 America{n}            democra{cy,tic}
##                          24                           4
##                    republic              Democrat{,ic}
##                           0                           0
##                  Republican                  free{,dom}
##                           0                          36
##                         war                  God (only)
##                          24                           1
##                   God Bless {Jesus, Christ, Christian}
##                           0                           0
## [1] "Event occurences in BUSH's speach is:"
##
## (LAUGHTER)
##          1
##
##
## [1] "The statistics for the first debate in 2000 ..."
## [1] "The average word length of MODERATOR's speach is:"
## [1] 4.558824
## [1] "The number of characters (in the words) in MODERATOR's speach is:"
## [1] 7750
## [1] "The number of words in MODERATOR's speach is:"
## [1] 1700
```

```
## [1] "The buzzwords in MODERATOR's speach is:"
##                           I                          we
##                          14                          11
##                 America{n}            democra{cy,tic}
##                           0                           1
##                   republic              Democrat{,ic}
##                           0                           1
##                 Republican                 free{,dom}
##                           1                           0
##                         war                 God (only)
##                           0                           0
##                  God Bless {Jesus, Christ, Christian}
##                           0                           0
## [1] "Event occurences in MODERATOR's speach is:"
##
## (Applause) (APPLAUSE)
##          1          1
## [1] "The average word length of GORE's speach is:"
## [1] 4.339315
## [1] "The number of characters (in the words) in GORE's speach is:"
## [1] 31434
## [1] "The number of words in GORE's speach is:"
## [1] 7244
## [1] "The buzzwords in GORE's speach is:"
##                           I                          we
##                         230                          72
##                 America{n}            democra{cy,tic}
##                          13                           1
##                   republic              Democrat{,ic}
##                           0                           1
##                 Republican                 free{,dom}
##                           1                           1
##                         war                 God (only)
##                           3                           0
##                  God Bless {Jesus, Christ, Christian}
##                           0                           0
## [1] "Event occurences in GORE's speach is:"
## < table of extent 0 >
## [1] "The average word length of BUSH's speach is:"
## [1] 4.3035
## [1] "The number of characters (in the words) in BUSH's speach is:"
## [1] 32216
## [1] "The number of words in BUSH's speach is:"
## [1] 7486
## [1] "The buzzwords in BUSH's speach is:"
```

```
##                           I                          we
##                         213                          83
##                 America{n}            democra{cy,tic}
##                          19                           1
##                   republic              Democrat{,ic}
##                           0                           2
##                 Republican                 free{,dom}
##                           1                           3
##                         war                 God (only)
##                           4                           0
##                 God Bless {Jesus, Christ, Christian}
##                           0                           0
## [1] "Event occurences in BUSH's speach is:"
## < table of extent 0 >
##
##
## [1] "The statistics for the first debate in 1996 ..."
## [1] "The average word length of LEHRER's speach is:"
## [1] 4.691332
## [1] "The number of characters (in the words) in LEHRER's speach is:"
## [1] 4438
## [1] "The number of words in LEHRER's speach is:"
## [1] 946
## [1] "The buzzwords in LEHRER's speach is:"
##                           I                          we
##                           6                           6
##                 America{n}            democra{cy,tic}
##                           0                           0
##                   republic              Democrat{,ic}
##                           0                           1
##                 Republican                 free{,dom}
##                           2                           0
##                         war                 God (only)
##                           0                           0
##                 God Bless {Jesus, Christ, Christian}
##                           0                           0
## [1] "Event occurences in LEHRER's speach is:"
## < table of extent 0 >
## [1] "The average word length of CLINTON's speach is:"
## [1] 4.365762
## [1] "The number of characters (in the words) in CLINTON's speach is:"
## [1] 33612
## [1] "The number of words in CLINTON's speach is:"
## [1] 7699
## [1] "The buzzwords in CLINTON's speach is:"
```

```
##                           I                          we
##                         243                         113
##                 America{n}             democra{cy,tic}
##                          34                           4
##                    republic               Democrat{,ic}
##                           0                           1
##                  Republican                  free{,dom}
##                           7                           8
##                         war                  God (only)
##                           2                           0
##                  God Bless {Jesus, Christ, Christian}
##                           0                           0
## [1] "Event occurences in CLINTON's speach is:"
## < table of extent 0 >
## [1] "The average word length of DOLE's speach is:"
## [1] 4.308865
## [1] "The number of characters (in the words) in DOLE's speach is:"
## [1] 35044
## [1] "The number of words in DOLE's speach is:"
## [1] 8133
## [1] "The buzzwords in DOLE's speach is:"
##                           I                          we
##                         276                         109
##                 America{n}             democra{cy,tic}
##                          42                           0
##                    republic               Democrat{,ic}
##                           0                           7
##                  Republican                  free{,dom}
##                          11                           1
##                         war                  God (only)
##                           2                           0
##                  God Bless {Jesus, Christ, Christian}
##                           0                           1
## [1] "Event occurences in DOLE's speach is:"
##
##    (ph) (staff)
##       3       1
```

3 Practice with S4 – Illustrating a Random Walk

```
rw <- setClass(
  "rw",
  # The basic slots associated with our class
```

```
  slots = c(
    start = "numeric",
    steps   = "numeric",
    trajectory_recording = "logical",
    .trajectory = "matrix"),

  # Now we declare our default values
  prototype=list(
    start = c(0,0),
    #steps = 10,
    trajectory_recording = TRUE
  ),
  # Look for things that might be amiss
  validity=function(object)
  {
    # REMEMBER TO ADD INTEGER CHECKS.
    if(object@steps<0) {
      return("Please enter
             a positive number of steps.")
    }
    if(as.integer(object@steps)!=object@steps) {
      return("Please enter
             an integer valued number of steps.")
    }
    if(length(object@start)!=2) {
      return("This program is
             only written for 2D (for now!).")
    }
    return(TRUE)
  }
)


# Found "OOP in R" (http://practicalcomputing.org/node/80) to be very useful.
setGeneric("start<-", function(self, value) standardGeneric("start<-"))

## [1] "start<-"

setReplaceMethod("start",
  "rw",
  function(self,value) {
  self@start <- value
  self
  }
)

## [1] "start<-"
```

```
setMethod(
  f="[",
  signature="rw",
  definition=function(x,i,drop){
    mypath=slot(x,".trajectory");
    xs=sum(mypath[1:i,1]);
    ys=sum(mypath[1:i,2]);
    return(c(x@start[1]+xs, x@start[2]+ys))
  }
)
```

```
## [1] "["
```

```
setMethod(
  f="plot",
  signature="rw",
  definition=function(x){
    mypath=slot(x,".trajectory");
    xs=cumsum(mypath[,1]);
    ys=cumsum(mypath[,2]);
    plot(x@start[2]+ys,x@start[1]+xs, type='o');
  }
)
```

```
## Creating a generic function for 'plot' from package 'graphics' in
the global environment
```

```
## [1] "plot"
```

```
setMethod(
  f="print",
  signature="rw",
  definition=function(x){
    print("Starting position:")
    print(slot(x,"start"))
    print("After this many steps...:")
    print(slot(x,"steps"))
    print("We arrive at:")
    print(x[slot(x,"steps")])
    if (slot(x,"trajectory_recording"))
    { for (i in 1:slot(x,"steps")) {
        print(x[i]) }}
  }
)
```

```
## Creating a generic function for 'print' from package 'base' in the
global environment
```

```
## [1] "print"
```

```
setGeneric("simulate",
           function(.Object){standardGeneric("simulate")})
```

```
## Creating a new generic function for 'simulate' in the global environment
```

```
## [1] "simulate"
```

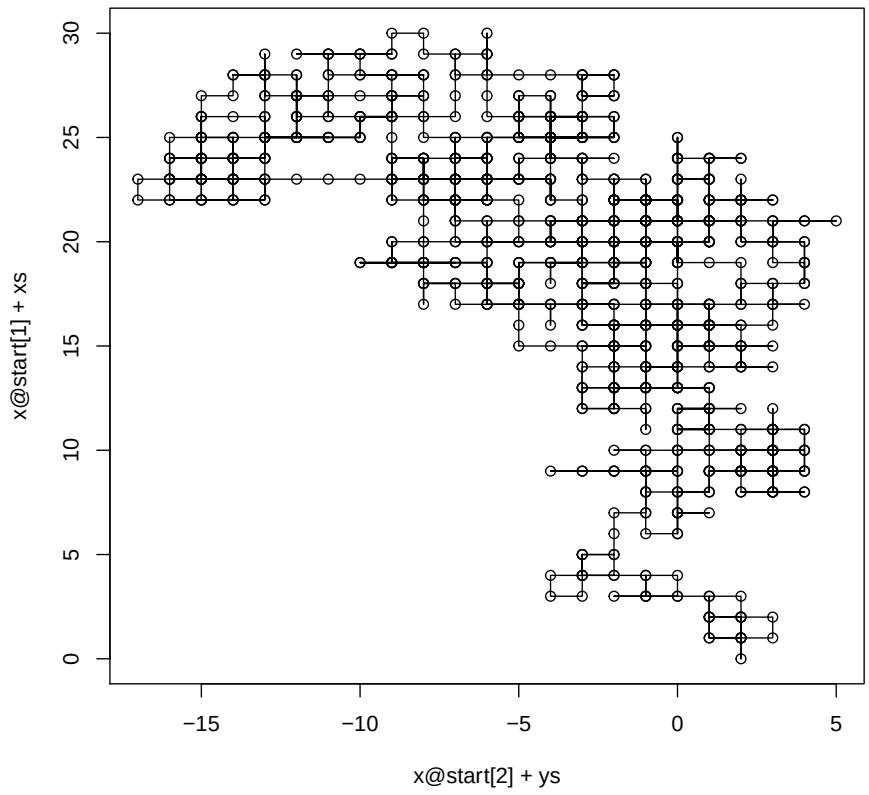
```
setMethod(
  f="simulate",
  signature="rw",
  definition=function(.Object){
    slot(.Object,".trajectory") <-
      matrix(c(0, 1, -1, 0, 1, 0, 0, -1),
             nrow=4,
             ncol=2)[sample(4,size=slot(.Object,"steps"),
             replace=TRUE),];
    return(.Object)
  }
)
```

```
## [1] "simulate"
```

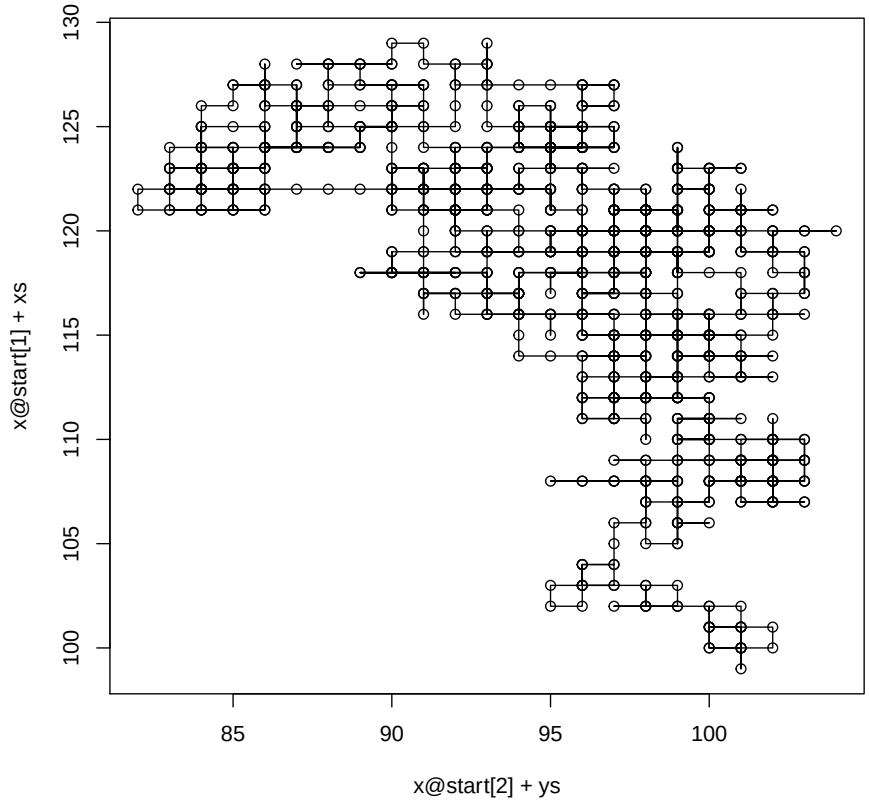
```
# Testing the replacement class
my_walk <- new("rw", start=c(1,1),steps=1000,trajectory_recording=FALSE)
# To circumvent this, I would need to use the "assign" function, which
# the S4 manual cautions against!
# (https://cran.r-project.org/doc/contrib/Genolini-S4tutorialV0-5en.pdf)
my_walk <- simulate(my_walk)
my_walk[50]
```

```
## [1] 10  0
```

```
plot(my_walk)
```




```
start(my_walk)<-c(100,100)
plot(my_walk)
```




```
print(my_walk)
```

```
## [1] "Starting position:"
## [1] 100 100
## [1] "After this many steps...:"
## [1] 1000
## [1] "We arrive at:"
## [1] 124  88
```