Useful R code

Matthew Malishev[1]*
[1] *Department of Biology, Emory University, 1510 Clifton Road NE, Atlanta, GA, USA, 30322*

Contents

Date: 2018-08-21
`R` version: 3.5.0
*Corresponding author: matthew.malishev@gmail.com
This document can be found at https://github.com/darwinanddavis/UsefulCode

Overview

This document outlines some useful `R` code for plotting, cool functions, and other random tidbits.

Install dependencies

```
packages <- c("rgdal","dplyr","zoo","RColorBrewer","viridis","plyr","digitize","jpeg","devtools","image
if (require(packages)) {
    install.packages(packages,dependencies = T)
    require(packages)
}
lapply(packages,library,character.only=T)
```

Classes

Convert character to factor to numeric without conversion error

```
read.table(f,header=T,sep=",",row.names=NULL,stringsAsFactors=FALSE, strip.white=TRUE)
f$V2<-as.numeric(f$V2)
```

See call options for class

```
methods(class="estUDm")
```

Set dynamic input for variable / assign variable to char vector

```
shadedens<-function(shadedens){ # set shade density to clumped (to match food) or sparse
  if (shadedens == "Random"){
    NLCommand("set Shade-density \"Random\" ")
    }else{
    NLCommand("set Shade-density \"Clumped\" ")
    }
  }
shadedens("Clumped") # set clumped resources
```

Dataframes

Optimal empty data frame

```
df <- data.frame(Date=as.Date(character()),
                 X=numeric(),
                 Y=integer(),
                 stringsAsFactors=FALSE)
```

Add df cols with `mutate`

```
df <- data.frame("a"=rnorm(10),"b"=(1:20))
df %>%
  mutate(
  "c"=rnorm(20),
  b = b *67
  )
```

ggplot functions

Remove annoying stock gridlines from plot window

```
plot + theme_bw() +
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
                            panel.grid.minor = element_blank(), axis.line = element_line(colour = "blac
# alternative (after loading ggridges library)
theme_ridges(grid=F,center_axis_labels = T)
```

Setting global graphics theme for ggplot

```
plot_it_gg <- function(bg,family){ # bg = colour to plot bg, family = font family
  theme_tufte(base_family = family) +
  theme(panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_rect(fill = bg,
                                        colour = bg),
        plot.background = element_rect(fill=bg)
  ) +
    theme(axis.line = element_line(color = "white")) +
    theme(axis.ticks = element_line(color = "white")) +
    theme(plot.title = element_text(colour = "white")) +
    theme(axis.title.x = element_text(colour = "white"),
          axis.title.y = element_text(colour = "white")) +
    theme(axis.text.x = element_text(color = "white"),
          axis.text.y = element_text(color = "white")) +
    theme(legend.key = element_rect(fill = bg)) + # fill bg of legend
    theme(legend.title = element_text(colour="white")) + # legend title
    theme(legend.text = element_text(colour="white")) # legend labels
}
```

Put plot in function to take dynamic data inputs
Ref: http://jcborras.net/carpet/visualizing-political-divergences-2012-local-elections-in-helsinki.html

```
hr.mass.plot <- function(d) {
  p <- ggplot(d, aes(HR, Mass, color = colfunc)) +
    geom_density_2d(data=d, aes(x = HR, y = Mass),
                    stat = "density2d",position="identity",
                    color=adjustcolor("orange",alpha=0.8), size=1.5, contour = T, lineend="square",line
  p <- p + geom_point(data=d, aes(x = HR, y = Mass),
                      color=colfunc,
                      fill=colfunc) +
    scale_color_manual(values = magma(8))
  p <- p + scale_y_continuous(limits=c(-200,200), name="Mass lost (g)")
  p <- p + scale_x_continuous(limits=c(0,0.35),name=expression("Home range area (km^2)"))
  p <- p + theme_classic()
  print(p)
}
hr.mass.plot(d)
```

Using ggplot when looping through for loop and saving to dir

```
pdf("mypdf.pdf",onefile = T)
for(i in 1:3){
par(bty="n", las = 1)
```

```
  grid.arrange(
  ggplot(data, aes(x = X, y = Y, fill=..x..)) + # geom_density_ridges()
    # scale = overlap
    geom_density_ridges_gradient(scale = 5, size=0.2,color="black", rel_min_height = 0.01,panel_scaling=
    geom_density_ridges(scale = 5, size=0.2,color="black", rel_min_height = 0.01,fill="white",alpha=0.2)
    # geom_density_ridges(scale = 5, size=0.2,color="white", rel_min_height = 0.01,fill=col,alpha=0.5) +
    scale_fill_viridis(name = "Diameter", alpha=0.1, option = "magma",direction=-1) + # "magma", "infer
    xlim(c(0,25)) +
    labs(title = paste0("Title_",i)) +
    xlab("X") +
    ylab("Y") +
    # plot_it_gg("white")
  )
} # end loop
dev.off()
```

NAs

Replace NAs with 0's

```
df[is.na(df)] <- 0
```

Replace X values less than given value (V) with 0

```
df$X[df$X<V] <- 0
```

Check for NAs

```
sapply(df, function(x) sum(is.na(x)))
```

Replace NaN and Inf values with NA

```
df$col1[which(!is.finite(df$col1))] <-  NA
```

Fill in missing data values in sequence with NA

```
# /Users/malishev/Documents/Manuscripts/Chapter4/Sims/Chapter4_figs.R
library(zoo)
data <- data.frame(index = c(1:4, 6:10),
  data = c(1.5,4.3,5.6,6.7,7.1,12.5,14.5,16.8,3.4))
#you can create a series
z <- zoo(data$data, data$index)
#end extend it to the grid 1:10
z <- merge(zoo(,1:10), z)

#worked example
# fill in missing Tb values
minTb.d <- zoo(minTb$Tick,minTb$Days)
minTb.d <- merge(zoo(NULL,1:days), minTb.d) # make the minTb series match the temp series (117 days)
minTb.d <- as.numeric(minTb.d) # = time individuals reached VTMIN in ticks
minTb <- minTb.d - temp$Tick # get diff between starting time and time to reach VTMIN
minTb <- minTb/2 # convert ticks to minutes
minTb <- minTb/60 #convert to hours
minTb <- data.frame("Days"=1:days,"Time"=minTb)

# then fill in missing values
```

```
approx(minTb$Time,method = "linear")
```

Plotting

Plot one plot window above and two below

```
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
```

Bookend axis ticks for plot E.g. at 0 and 100 when data is 1:99

```
axis(1,at=c(0,length(loco$X)),labels=c("",""))# bookending axis tick marks
```

Optimal legend formatting for base

```
legend("right",legend=c("Small","Intermediate","Large"),col=c(colfunc[colvec[1:3]]),
       bty="n",pch=20,pt.cex=1.5,cex=0.7,y.intersp = 0.5, xjust = 0.5,
       title="Size class",title.adj = 0.3,text.font=2,
       trace=T,inset=0.1)
```

Plot inset plot in current plot (https://stackoverflow.com/questions/17041246/how-to-add-an-inset-subplot-to-topright-of-an-r-plot)

```
# calculate position of inset
plotdim <- par("plt")# get plot window dims as fraction of current plot dims
xleft    = plotdim[2] - (plotdim[2] - plotdim[1]) * 0.5
xright   = plotdim[2]  #
ybottom  = plotdim[4] - (plotdim[4] - plotdim[3]) * 0.5  #
ytop     = plotdim[4]  #

# set position for plot inset
par(fig = c(xleft, xright, ybottom, ytop),mar=c(0,0,0,0),new=TRUE)

boxplot(Eggs~Size,data=meso2,
                col=adjustcolor(colfunc[colvec[1:3]],alpha=0.5),
                notch = T,xlab="Week",ylab="Diameter (mm)",
                xaxs = "i", yaxs = "i"
                )
```

Interactive plots with rCharts (javascript and d3 viz)
http://ramnathv.github.io/rCharts/

```
require(devtools)
install_github('rCharts', 'ramnathv')
```

Cluster plot
https://rpubs.com/dgrtwo/technology-clusters

```
library(readr)
library(dplyr)
library(igraph)
library(ggraph)
library(ggforce)

# This shared file contains the number of question that have each pair of tags
# This counts only questions that are not deleted and have a positive score
tag_pair_data <- read_csv("http://varianceexplained.org/files/tag_pairs.csv.gz")
```

```
relationships <- tag_pair_data %>%
  mutate(Fraction = Cooccur / Tag1Total) %>%
  filter(Fraction >= .35) %>%
  distinct(Tag1)

v <- tag_pair_data %>%
  select(Tag1, Tag1Total) %>%
  distinct(Tag1) %>%
  filter(Tag1 %in% relationships$Tag1 |
         Tag1 %in% relationships$Tag2) %>%
  arrange(desc(Tag1Total))

a <- grid::arrow(length = grid::unit(.08, "inches"), ends = "first", type = "closed")

set.seed(2016)

relationships %>%
  graph_from_data_frame(vertices = v) %>%
  ggraph(layout = "fr") +
  geom_edge_link(aes(alpha = Fraction), arrow = a) +
  geom_node_point(aes(size = Tag1Total), color = "lightblue") +
  geom_node_text(aes(size = Tag1Total, label = name), check_overlap = TRUE) +
  scale_size_continuous(range = c(2, 9)) +
  ggforce::theme_no_axes() +
  theme(legend.position = "none")
```

Define global plotting graphics function

```
plot_it <- function(manuscript,bg,cp,alpha,family){ # plotting function (plot for MS or not, set bg col
  graphics.off()
  if(manuscript==0){
    if(bg=="black"){
      colvec<-magma(200,1)
      par(bg = colvec[1],col.axis="white",col.lab="white",col.main="white",fg="white",bty="n",las=1,mar=
      border=adjustcolor("purple",alpha=0.5)
    }else{
      colvec<-bpy.colors(200)
      par(bg = colvec[1],col.axis="white",col.lab="white",col.main="white",fg="white",bty="n",las=1,mar=
      border=adjustcolor("blue",alpha=0.5)
    }
  }else{
#    graphics.off()
    par(bty="n",las=1,family=family)
  }
  # color palettes
  # ifelse(manuscript==1,colvec<-adjustcolor(brewer.pal(9,cp)[9], alpha = alpha),colvec <- adjustcolor(
   # colfunc <<- colorRampPalette(brewer.pal(9,cp),alpha=alpha)
  colfunc <<- adjustcolor(brewer.pal(9,cp),alpha=alpha) # USES <<- OPERATOR
}

# Setting ggplot theme graphics
plot_it_gg <- function(bg){ # bg = colour to plot bg, family = font family
  if(bg=="white"){
    bg <- "white"
```

```
    fg <- "black"
  theme_tufte(base_family = "HersheySans") +
    theme(panel.border = element_blank(),panel.grid.major = element_blank(),panel.grid.minor = element_
    theme(axis.line = element_line(color = fg)) +theme(axis.ticks = element_line(color = fg)) + theme(pl
}
  }# end gg

# Set global plotting parameters
print("1/0, set colour, set colour palette 'display.brewer.all()',set alpha for col,set font")
plot_it(0,"blue","YlOrRd",1,"HersheySans") # set col function params
plot_it_gg("white") # same as above
```

Reading in files/data

Read in file manually

```
get.file.vol <- read.table(file.choose())#read file manually
v.file <- get.file.vol[1:100,1]#get the volume
```

Loop through files from dir and append to list

```
# reading in spdf (hrpath) files from drive
setwd("/Users/camel/Desktop/Matt2016/Manuscripts/MalishevBullKearney/Resubmission/2016/barcoo sims/barc
file.list<-list.files()
hrs75<-as.list(rep(1,100)) # empty list
for (f in 1:100){
  load(file.list[f])
  hrs75[f]<-hrpath
}

# working version
#converting spdf into mcp(spdf,100,unout="m2)
ghr<-list()
for (i in hrs75[1:10]) {
  m<-mcp(i,100,unout='m2')
  ghr<-c(ghr,m)
};ghr
```

Regular expressions (regex)

Get just numbers or characters

```
vec <- "16-Feb-2018 20:08:04 PM"
vecN <- gsub("[^[:digit:]]", "", vec); vec; print(paste0("Just numbers: ",vecN))
```

```
[1] "16-Feb-2018 20:08:04 PM"
```

```
[1] "Just numbers: 162018200804"
```

```
vecC <- gsub("[[:digit:]]", "", vec); vec; print(paste0("Just characters: ", vecC))
```

```
[1] "16-Feb-2018 20:08:04 PM"
```

```
[1] "Just characters: -Feb- :: PM"
```

```r
# with tidyr. requires data frame
require(tidyr)
df <- data.frame(N1=c("APPLE348744", "BANANA77845", "OATS2647892", "EGG98586456"))
print("tidyr doesn't work with strings separated by spaces")
```

```
[1] "tidyr doesn't work with strings separated by spaces"
```

```r
df %>%
  separate(N1, into = c("text", "num"), sep = "(?<=[A-Za-z])(?=[0-9])")
```

```
# A tibble: 4 x 2
  text   num
  <chr>  <chr>
1 APPLE  348744
2 BANANA 77845
3 OATS   2647892
4 EGG    98586456
```

Insert or replace a character in a string at a specific location

```r
require(stringi)
vec <- "ABCEF"
stri_sub(vec, 4, 2) <- "d"
print(paste0("Original: ABCEF")); print(paste0("New: ",vec))
```

```
[1] "Original: ABCEF"
```

```
[1] "New: ABCdEF"
```

Subsetting

Select specific rows E.g. select rows of sfeed_move not in foodh

```r
library(sqldf)
a1NotIna2_h  <- sqldf('SELECT * FROM sfeed_move EXCEPT SELECT * FROM foodh')
a1NotIna2_l  <- sqldf('SELECT * FROM sfeed_move EXCEPT SELECT * FROM foodl')
# select rows from sfeed_move that also appear in foodh
a1Ina2_h  <- sqldf('SELECT * FROM sfeed_move INTERSECT SELECT * FROM foodh')
a1Ina2_l  <- sqldf('SELECT * FROM sfeed_move INTERSECT SELECT * FROM foodl')
```

Count occurrences of values in data frame

```r
table(unlist(df$X))
```

Remove a specific column from a data frame

```r
within(df, rm("Col1"))
```

R Markdown

Hide unwanted code output, such as inherent examples for functions

```r
# ```{r, cache = TRUE, tidy = TRUE, lazy = TRUE, results='markup'}
```

Web scraping

Scraping web tables
http://web.mit.edu/~r/current/arch/i386_linux26/lib/R/library/XML/html/readHTMLTable.html%5Bhttp://web.mit.edu/~r/current/arch/i386_linux26/lib/R/library/XML/html/readHTMLTable.html%5D

```
library(XML)
readHTMLTable()
```

Scraping Twitter timelines
See complete example at http://varianceexplained.org/r/trump-tweets/

```
# https://cran.r-project.org/web/packages/twitteR/
library(dplyr)
library(purrr)
library(twitteR)
```