The Lab note book for recode Network Analysis of Survey

Ph.D. of Plant Pathology

Sith Jaisong

s.jaisong@irri.org

Beginning 5 February 2015

Contents

Friday, 6 February 2015

1 Raw data

The R script is to load the raw data from folder.This script start using gdata package [2]
The R script file:

```r
###########################################
#'title : 1-raw.R
#'date : January, 2015
#'purpose : Load raw data from
#' format
#'writed by : Sith Jaisong (s.jaisong@irri.org)
#'contact : International Rice Research Institute
#'input : the xts file
#'output : data.frame and RData
###########################################
# The is the script for loading the raw data from local computer
#------Load Library-----
 library(gdata)
#---- Set working directory
# set your working directory
#wd = '~/Documents/R.github/network.analysis.skep1'
#setwd(wd)

Filepath <- '~/Google Drive/1.SKEP1/SKEP1survey.xls'
#-----Load raw data (Survey data in SKEP 1)-----

data <- read.xls(Filepath,
                 sheet = 1,
                 header = TRUE,
                 stringsAsFactor = FALSE)

#------ Examine the raw data ------

head(data)
str(data) # check the class of each variable
summary(data)
#---save data to R object ----

save(data, file = "output/1-raw.skep1survey.RData")
```

2 correct class of data

The R script is to load the raw data from folder.This script start using plyr, dplyr, lubridate package [3, 4, 1]

```r
##################################################
#'title : 2-technically_correct
#'date : January, 2015
#'purpose :
#'writed by : Sith Jaisong (s.jaisong@irri.org)
#'contact : International Rice Research Institute
#'input : import excel file from the shared files and delete the
#'output : data frame and RData
##################################################
#
library(plyr)
library(dplyr)
library(lubridate)
#-----Load file from output folder-----

load(file = "output/1-raw.skep1survey.RData")


#----- clean define the missing value -----

data[data == "-"] <- NA # replace '-' with NA
data[data == ""] <- NA # replace 'missing data' with NA

#----- to lower variable names -----
names(data) <- tolower(names(data))

#----- This step is to seelect the numeric data set -----

data <- transform(data,
                  phase = as.factor(phase),
                  fno = as.character(fno),
                  identifier = as.character(identifier),
                  country = as.factor(country),
                  year = as.factor(year),
                  season = as.character(season),
                  lat = as.numeric(lat),
                  long = as.numeric(long),
                  village = as.character(village),
                  fa = as.numeric(fa),
                  fn = as.character(fn),
                  lfm = as.character(lfm),
                  pc = as.factor(pc),
                  fp = as.character(fp),
                  cem = as.factor(cem),
```

```
45              ast = as.factor(ast),
46              nplsqm = as.numeric(nplsqm),
47              ced = dmy(ced),
48              cedjul = as.numeric(cedjul),
49              hd = dmy(hd),
50              hdjul = as.numeric(hdjul),
51              ccd = as.character(ccd),
52              cvr = as.character(cvr),
53              vartype = as.character(vartype),
54              varcoded = as.character(varcoded),
55              fym = as.character(fym),
56              fym.coded = as.character(fym.coded),
57              n = as.numeric(n),
58              p = as.numeric(p) ,
59              k = as.numeric(k),
60              mf = as.numeric(mf),
61              wcp = as.character(wcp),
62              mu = as.character(mu) ,
63              iu = as.numeric(iu),
64              hu = as.numeric(hu),
65              fu = as.numeric(fu),
66              cs = as.factor(cs),
67              ldg = as.numeric(ldg),
68              yield = as.numeric(yield) ,
69              dscum = as.numeric(dscum),
70              wecum = as.numeric(wecum),
71              ntmax = as.numeric(ntmax),
72              npmax = as.numeric(npmax),
73              nltmax = as.numeric(nltmax),
74              nlhmax = as.numeric(nltmax),
75              waa = as.numeric(waa),
76              wba = as.numeric(wba) ,
77              dhx = as.numeric(dhx),
78              whx = as.numeric(whx),
79              ssx = as.numeric(ssx),
80              wma = as.numeric(wma),
81              lfa = as.numeric(lfa),
82              lma = as.numeric(lma),
83              rha = as.numeric(rha) ,
84              thrx = as.numeric(thrx),
85              pmx = as.numeric(pmx),
86              defa = as.numeric(defa) ,
87              bphx = as.numeric(bphx),
88              wbpx = as.numeric(wbpx),
89              awx = as.numeric(awx),
90              rbx =as.numeric(rbx),
91              rbbx = as.numeric(rbbx),
92              glhx = as.numeric(glhx),
93              stbx=as.numeric(stbx),
```

```
 94                      rbpx = as.numeric(rbpx),
 95                      hbx= as.numeric(hbx),
 96                      bbx = as.numeric(bbx),
 97                      blba = as.numeric(blba),
 98                      lba = as.numeric(lba),
 99                      bsa = as.numeric(bsa),
100                      blsa = as.numeric(blsa),
101                      nbsa = as.numeric(nbsa),
102                      rsa = as.numeric(rsa),
103                      lsa = as.numeric(lsa),
104                      shbx = as.numeric(shbx) ,
105                      shrx = as.numeric(shrx),
106                      srx= as.numeric(srx),
107                      fsmx = as.numeric(fsmx),
108                      nbx = as.numeric(nbx),
109                      dpx = as.numeric(dpx),
110                      rtdx = as.numeric(rtdx),
111                      rsdx = as.numeric(rsdx),
112                      gsdx =as.numeric(gsdx),
113                      rtx = as.numeric(rtx)
114 )
115
116 save(data, file="output/2-correct.class.skep1survey.RData")
```

Bibliography

[1] Garrett Grolemund and Hadley Wickham. Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25, 2011. 2

[2] Gregory R. Warnes, Ben Bolker, Gregor Gorjanc, Gabor Grothendieck, Ales Korosec, Thomas Lumley, Don MacQueen, Arni Magnusson, Jim Rogers, and others. *gdata: Various R programming tools for data manipulation*, 2014. R package version 2.13.3. 1

[3] Hadley Wickham. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29, 2011. 2

[4] Hadley Wickham and Romain Francois. *dplyr: A Grammar of Data Manipulation*, 2015. R package version 0.4.1. 2