R introduction: data import and output

Qiang Shen

Dec. 5, 2017

R version

- version 3.3.3 (2017.3.7)

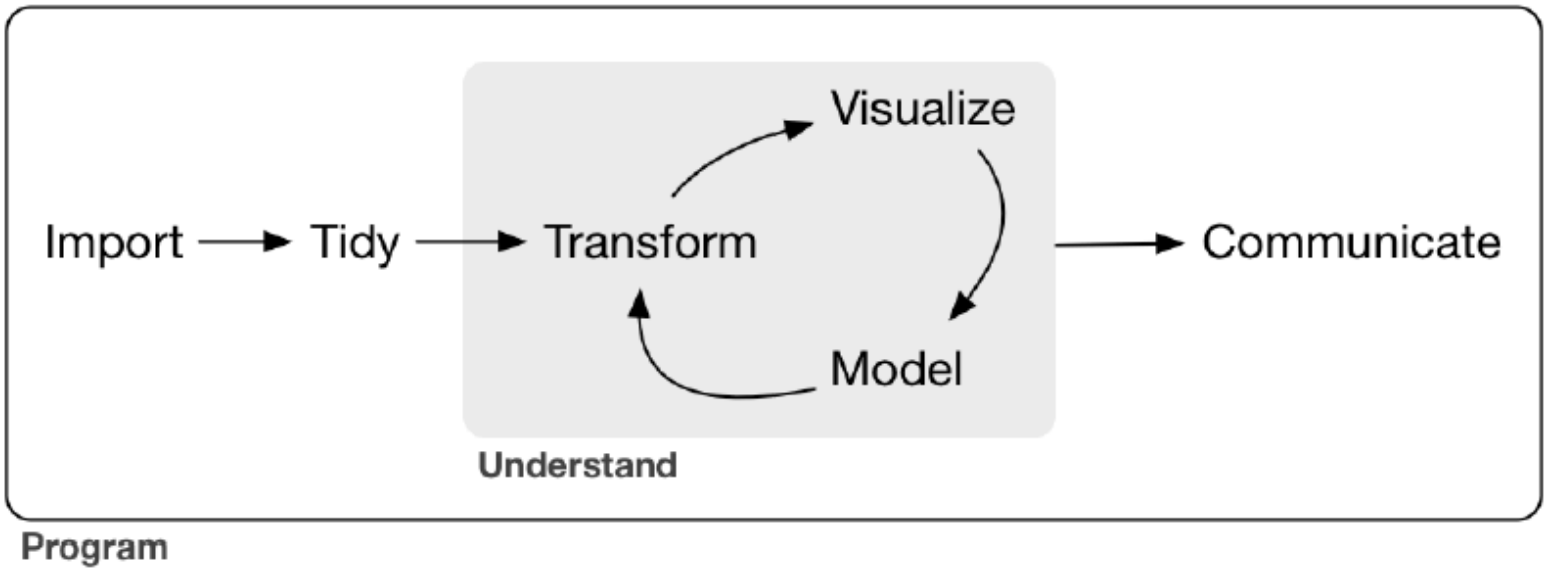
R version

- version 3.3.3 (2017.3.7)
- 32 bit / 64 bit (3.0)

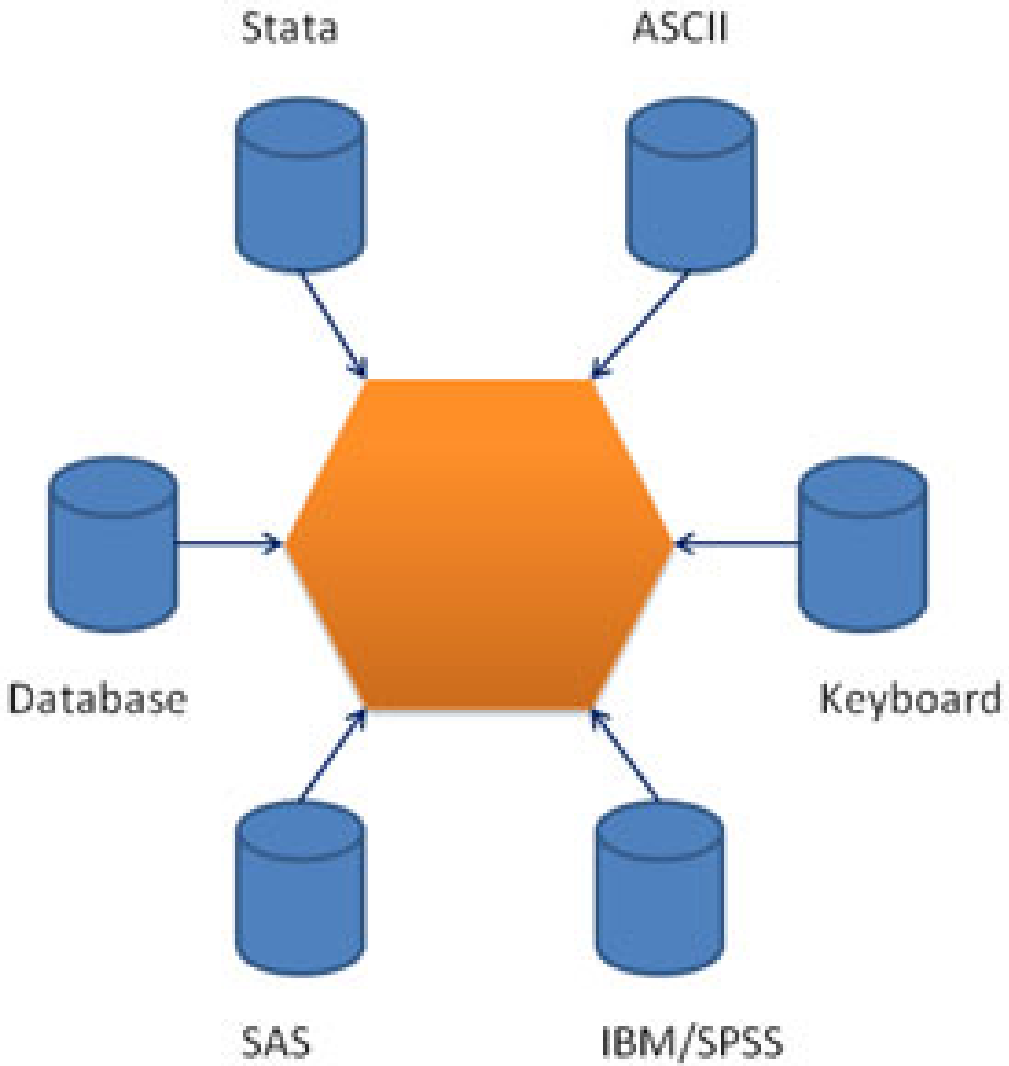
R version

- version 3.3.3 (2017.3.7)
- 32 bit / 64 bit (3.0)
- Microsoft R Open 3.3.2 (revolutionanalytics)

Framework


Visualize
Import
Tidy
Transform
Communicate
Model
Understand
Program

Data import and export



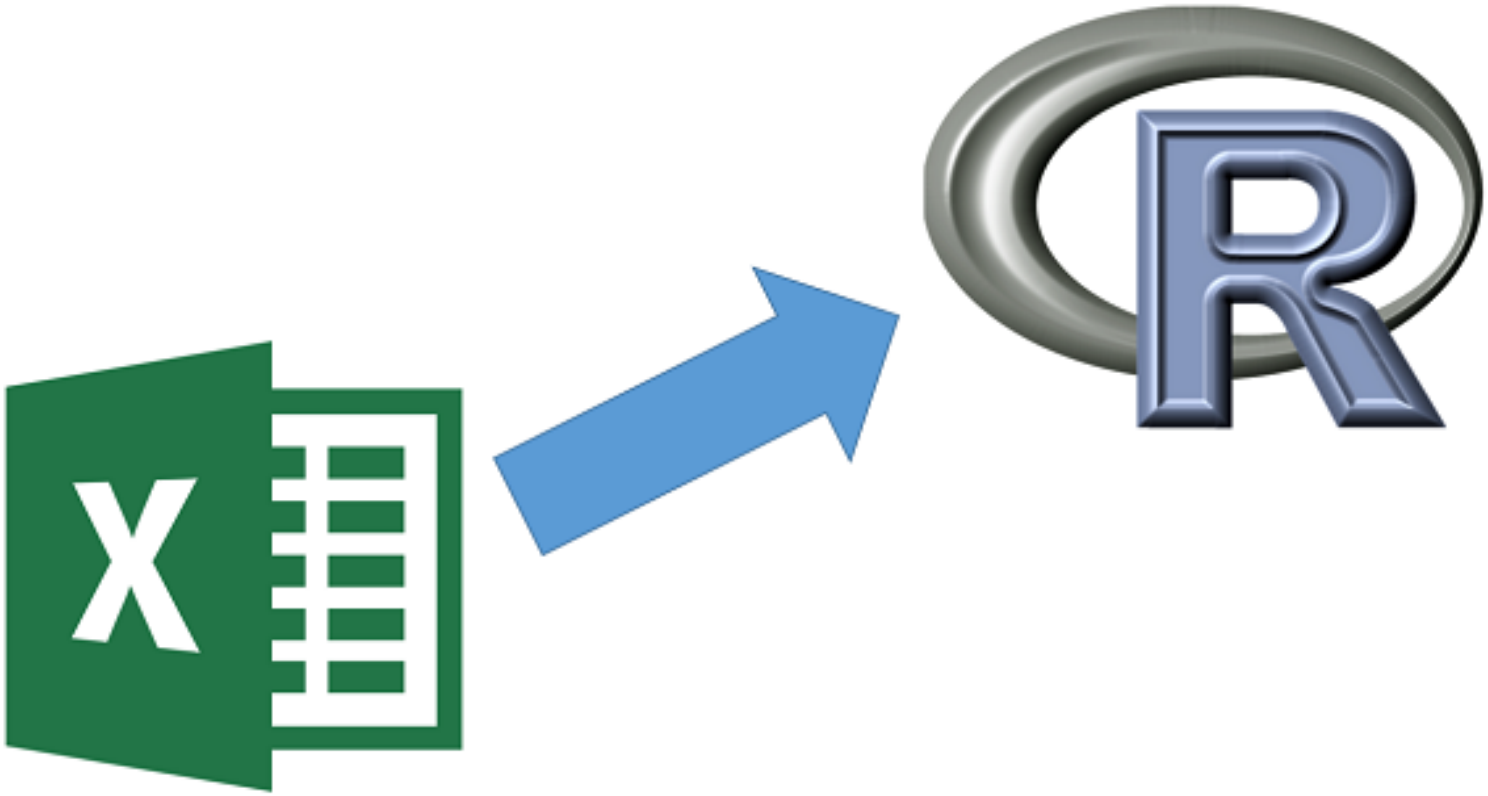
Import and export

- text file
- excel file
- Stata,SPSS,SAS
- online table
- databases

Import text file

```
str(read.table)
# write.table(iris,file='data/iris.txt',sep=','
#,quote=F,row.names=F)
#stringsAsFactors = F
iris_txt<-read.table("data/iris.txt",sep=",",header=T,strin
#stringsAsFactors = F
system.time(dat2 <- read.table("data/NYNEWYOR.txt",
    col.names=c("DAY","MONTH","YEAR","TEMP")))
iris_csv<-read.csv('data/iris.csv')
```

Import excel file



Import excel file

```
library(xlsx)
indicator_xls1<-read.xlsx("data/HIV.xlsx", 1)
system.time(indicator_xls<-read.xlsx("data/HIV.xlsx", 1))
 ##outdated.
library(XLConnect)
wb <- loadWorkbook("data/HIV.xlsx")
indicator_xls2 <- readWorksheet(wb, sheet=1)
detach(package:devtools)
devtools::install_github("hadley/readxl")
# library(readxl)
# system.time(indicator_xls<-read_excel("data/HIV.xlsx", 1
```

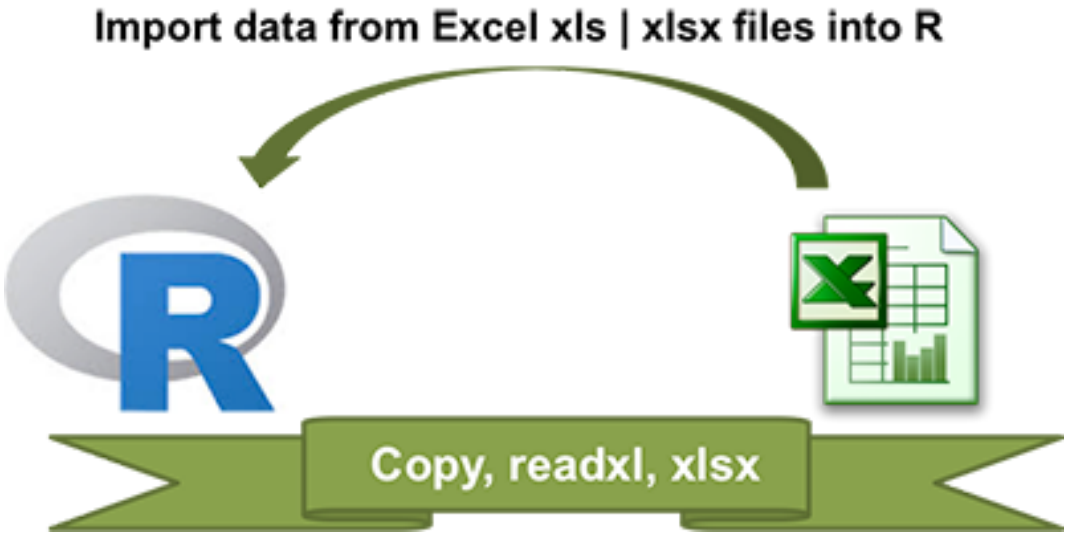
import data from SPSS and Stata

```
library(foreign)
Restaurant<-read.dta('data/Restaurant.dta')
head(Restaurant)
Restaurant_spss<-read.spss('data/restaurant.sav',
                           to.data.frame=TRUE)
head(Restaurant_spss)[1:5,]
#Restaurant<-read.ssd('restaurant.ssd',
                           #to.data.frame=TRUE)
```

tidyverse: menu

- readr: flat file (csv)
- readxl: excel
- haven: SPSS,SAS,Stata
- *openxlsx*

readxl


Import data from Excel xls | xlsx files into R
Copy, readxl, xlsx

readxl+openxlsx

```
l <- list("iris" = iris, "mtcars" = mtcars, chickwts = chi
library(openxlsx)
# detach(package:xlsx)
write.xlsx(l, file = "data/datasets.xlsx")
library(readxl)
sample<-read_excel('data/datasets.xlsx','mtcars')
excel_sheets('data/datasets.xlsx')
```

```
[1] "iris"     "mtcars"   "chickwts" "quakes"
```

Online table

http://mirrors.ustc.edu.cn/CRAN/web/packages/

```
## example 1:packages
library(XML)
theURL=paste("http://mirrors.ustc.edu.cn/CRAN/web/packages/
 sep="")
Rpackages = readHTMLTable(theURL, header=T,
          which=1,stringsAsFactors=F)
dim(Rpackages)
head(Rpackages)[1:4,1:3]
write.csv(Rpackages,'Rpackages.csv',sep='\t',1)
```

Online table 2

```
library(XML)
url <- "http://www.uefa.com/uefachampionsleague/season=2016
tbs <- readHTMLTable(doc=url, which=1)
head(tbs)
```

```
      Goals scored  4 <U+00A0>  4
1   Possession (%) 51 <U+00A0> 49
2   Total attempts 16 <U+00A0> 14
3        on target  9 <U+00A0>  9
4       off target  4 <U+00A0>  4
5          blocked  3 <U+00A0>  1
6 against woodwork  0           1
```

Databases: MySQL

Databases: MySQL

```
# library(DBI);library(RMySQL)
#  conn <- dbConnect(MySQL(),user='root',password='123456'
#                    ,host='localhost',dbname = "rmysql")
#  users = dbGetQuery(conn, "SELECT * FROM f_demo")
#  dbDisconnect(conn)
# head(users)
```

Export data

- text,csv,xlsx

```
write.table(Restaurant,'output/Restaurant.txt',
            sep="\t",quote=F,row.name=F)
write.csv(Restaurant,'output/Restaurant.csv',
            row.names=F)
write.xlsx(Restaurant,'output/Restaurant.xlsx',
            row.name=F)
# writeWorksheetToFile('indicator.xlsx',
# data=df, sheet="FirstSheet")
```

matlab and R: write data frame

```
library(rmatio)
##write data frame
data<-data.frame(c(1,2,NA),c(4,5,6))
names(data)<-c('a','b')
data
```

```
   a b
1  1 4
2  2 5
3 NA 6
```

```
write.mat(data, 'data/dataframe.mat')
```

matlab: write list

```
data2<-list(a=c(1:10),b=c(4,5,8))
data2
```

```
$a
 [1]  1  2  3  4  5  6  7  8  9 10

$b
[1] 4 5 8
```

```
write.mat(data2, 'data/list.mat')
```

matlab: write nested list.

```
data(iris)
head(iris)[1,1:3]
```

```
  Sepal.Length Sepal.Width Petal.Length
1          5.1         3.5          1.4
```

```
names(iris)<-sub("\\.", "_", names(iris))
out<-split(iris[,c(1:4)],f=iris$Species)
write.mat(out,'data/iris_nested.mat')
```

read mat file.

```
read.mat('data/list.mat')
```

```
$a
 [1]  1  2  3  4  5  6  7  8  9 10

$b
[1] 4 5 8
```