Workshop 2.1: Data frames

Murray Logan
July 15, 2017

Table of contents

1. Data importation and exportation

1.1. Prior preparation

Download the `macnally.csv` file

- www.flutterbys.com.au/stats/ downloads/data/macnally.csv
- put it in a directory you wish to work from

Make sure you know where you have put it!

1.2. Prior preparation

Download the `macnally.csv` file

- www.flutterbys.com.au/stats/ downloads/data/macnally.csv
- put it in a directory you wish to work from

OR

```
> download.file('http://www.flutterbys.com.au/stats/downloads/data/macnally.csv',
+               '~/macnally.csv')
```

1.3. Working directory

- Querying the current working directory

```
> getwd()
```

```
[1] "/home/murray/Work/SUYR/downloads/slides"
```

1.4. Working directory

- Querying the current working directory

```
> getwd()
```

```
[1] "/home/murray/Work/SUYR/downloads/slides"
```

- Examples of navigating (moving current working directory)

```
> #Go to a subdirectory of the current directory
> setwd('data')
> #Go to the parent directory
> setwd('..')
> #Go to a sibling directory
> setwd('../data')
```

2. Working with files

2.1. Importing from text file

2.1.1. Comma separated file

1. Full path

```
> MACNALLY <- read.csv(
+  '/home/murray/Work/SUYR/downloads/data/macnally.csv',
+   header=T, row.names=1, strip.white=TRUE)
> MACNALLY
```

```
                       HABITAT  GST EYR
Reedy Lake               Mixed  3.4 0.0
Pearcedale          Gipps.Manna  3.4 9.2
Warneet             Gipps.Manna  8.4 3.8
Cranbourne          Gipps.Manna  3.0 5.0
Lysterfield              Mixed  5.6 5.6
Red Hill                 Mixed  8.1 4.1
Devilbend                Mixed  8.3 7.1
Olinda                   Mixed  4.6 5.3
Fern Tree Gum     Montane Forest  3.2 5.2
Sherwin       Foothills Woodland  4.6 1.2
Heathcote Ju      Montane Forest  3.7 2.5
Warburton         Montane Forest  3.8 6.5
Millgrove                Mixed  5.4 6.5
Ben Cairn                Mixed  3.1 9.3
Panton Gap        Montane Forest  3.8 3.8
OShannassy               Mixed  9.6 4.0
Ghin Ghin                Mixed  3.4 2.7
Minto                    Mixed  5.6 3.3
Hawke                    Mixed  1.7 2.6
St Andrews    Foothills Woodland  4.7 3.6
Nepean        Foothills Woodland 14.0 5.6
Cape Schanck             Mixed  6.0 4.9
Balnarring               Mixed  4.1 4.9
Bittern             Gipps.Manna  6.5 9.7
Bailieston         Box-Ironbark  6.5 2.5
Donna Buang              Mixed  1.5 0.0
Upper Yarra              Mixed  4.7 3.1
Gembrook                 Mixed  7.5 7.5
Arcadia           River Red Gum  3.1 0.0
Undera            River Red Gum  2.7 0.0
Coomboona         River Red Gum  4.4 0.0
Toolamba          River Red Gum  3.0 0.0
Rushworth          Box-Ironbark  2.1 1.1
Sayers             Box-Ironbark  2.6 0.0
Waranga                  Mixed  3.0 1.6
```

```
Costerfield         Box-Ironbark  7.1 2.2
Tallarook     Foothills Woodland  4.3 2.9
```

2.2. Importing from text file

2.2.1. Comma separated file

2. Relative path

```
> MACNALLY <- read.csv('../data/macnally.csv',
+    header=T, row.names=1, strip.white=TRUE)
> getwd() #to see the current working directory
```

```
[1] "/home/murray/Work/SUYR/downloads/slides"
```

```
> MACNALLY
```

```
                       HABITAT  GST EYR
Reedy Lake               Mixed  3.4 0.0
Pearcedale         Gipps.Manna  3.4 9.2
Warneet            Gipps.Manna  8.4 3.8
Cranbourne         Gipps.Manna  3.0 5.0
Lysterfield              Mixed  5.6 5.6
Red Hill                 Mixed  8.1 4.1
Devilbend                Mixed  8.3 7.1
Olinda                   Mixed  4.6 5.3
Fern Tree Gum     Montane Forest  3.2 5.2
Sherwin       Foothills Woodland  4.6 1.2
Heathcote Ju      Montane Forest  3.7 2.5
Warburton         Montane Forest  3.8 6.5
Millgrove                Mixed  5.4 6.5
Ben Cairn                Mixed  3.1 9.3
Panton Gap        Montane Forest  3.8 3.8
OShannassy               Mixed  9.6 4.0
Ghin Ghin                Mixed  3.4 2.7
Minto                    Mixed  5.6 3.3
Hawke                    Mixed  1.7 2.6
St Andrews    Foothills Woodland  4.7 3.6
Nepean        Foothills Woodland 14.0 5.6
Cape Schanck             Mixed  6.0 4.9
Balnarring               Mixed  4.1 4.9
Bittern            Gipps.Manna  6.5 9.7
Bailieston        Box-Ironbark  6.5 2.5
Donna Buang              Mixed  1.5 0.0
Upper Yarra              Mixed  4.7 3.1
Gembrook                 Mixed  7.5 7.5
Arcadia          River Red Gum  3.1 0.0
Undera           River Red Gum  2.7 0.0
Coomboona        River Red Gum  4.4 0.0
Toolamba         River Red Gum  3.0 0.0
Rushworth         Box-Ironbark  2.1 1.1
Sayers            Box-Ironbark  2.6 0.0
Waranga                  Mixed  3.0 1.6
Costerfield       Box-Ironbark  7.1 2.2
Tallarook     Foothills Woodland  4.3 2.9
```

2.3. Importing from text file

2.3.1. Tab separated file

Relative path

```
> MACNALLY <- read.table('../data/macnally.txt',
+    header=T, row.names=1, sep='\t', strip.white=TRUE)
> MACNALLY
```

```
                        HABITAT  GST EYR
Reedy Lake                Mixed  3.4 0.0
Pearcedale          Gipps.Manna  3.4 9.2
Warneet             Gipps.Manna  8.4 3.8
Cranbourne          Gipps.Manna  3.0 5.0
Lysterfield               Mixed  5.6 5.6
Red Hill                  Mixed  8.1 4.1
Devilbend                 Mixed  8.3 7.1
Olinda                    Mixed  4.6 5.3
Fern Tree Gum     Montane Forest  3.2 5.2
Sherwin       Foothills Woodland  4.6 1.2
Heathcote Ju      Montane Forest  3.7 2.5
Warburton         Montane Forest  3.8 6.5
Millgrove                 Mixed  5.4 6.5
Ben Cairn                 Mixed  3.1 9.3
Panton Gap        Montane Forest  3.8 3.8
OShannassy                Mixed  9.6 4.0
Ghin Ghin                 Mixed  3.4 2.7
Minto                     Mixed  5.6 3.3
Hawke                     Mixed  1.7 2.6
St Andrews    Foothills Woodland  4.7 3.6
Nepean        Foothills Woodland 14.0 5.6
Cape Schanck              Mixed  6.0 4.9
Balnarring                Mixed  4.1 4.9
Bittern             Gipps.Manna  6.5 9.7
Bailieston         Box-Ironbark  6.5 2.5
Donna Buang               Mixed  1.5 0.0
Upper Yarra               Mixed  4.7 3.1
Gembrook                  Mixed  7.5 7.5
Arcadia           River Red Gum  3.1 0.0
Undera            River Red Gum  2.7 0.0
Coomboona         River Red Gum  4.4 0.0
Toolamba          River Red Gum  3.0 0.0
Rushworth          Box-Ironbark  2.1 1.1
Sayers             Box-Ironbark  2.6 0.0
Waranga                   Mixed  3.0 1.6
Costerfield        Box-Ironbark  7.1 2.2
Tallarook     Foothills Woodland  4.3 2.9
```

2.4. Exporting to a text file

```
> write.table(MACNALLY, '../data/macnally.csv',
+       quote=FALSE, row.names=TRUE, sep=',')
```

2.5. R and Excel?

2.6. R and Excel?

2.6.1. Reading from Excel

```
> library(XLConnect)
> wb=loadWorkbook("../data/macnally.xlsx")
> macnally=readWorksheet(wb,sheet="Sheet1",header=TRUE)
> head(macnally)
```

```
     LOCATION     HABITAT GST EYR
1  Reedy Lake       Mixed 3.4 0.0
2  Pearcedale Gipps.Manna 3.4 9.2
3     Warneet Gipps.Manna 8.4 3.8
4  Cranbourne Gipps.Manna 3.0 5.0
5 Lysterfield       Mixed 5.6 5.6
6    Red Hill       Mixed 8.1 4.1
```

```
> ##OR
> library(gdata)
> macnally<- read.xls('../data/macnally.xlsx',sheet='Sheet1',header=TRUE)
> head(macnally)
```

```
     LOCATION     HABITAT GST EYR
1  Reedy Lake       Mixed 3.4 0.0
2  Pearcedale Gipps.Manna 3.4 9.2
3     Warneet Gipps.Manna 8.4 3.8
4  Cranbourne Gipps.Manna 3.0 5.0
5 Lysterfield       Mixed 5.6 5.6
6    Red Hill       Mixed 8.1 4.1
```

2.7. R and Excel?

2.7.1. Writing to Excel

```
> library(XLConnect)
> wb=loadWorkbook("../data/macnally1.xlsx", create=TRUE)
> createSheet(wb, name='MacNally')
> writeWorksheet(wb, macnally, sheet='MacNally')
> saveWorkbook(wb)
```

2.8. Saving R objects

2.8.1. Saving an individual object

```
> save(MACNALLY, file='../data/macnally.RData')
```

2.8.2. Saving multiple objects

```
> #calculate the mean GST
> meanGST <- mean(MACNALLY$GST)
> #display the mean GST
> meanGST
> #save the MACNALLY data frame as well as the mean GST object
> save(MACNALLY, meanGST, file='macnallystats.RData')
```

2.9. Loading R objects

```
> load(file='../data/macnally.RData')
```

2.10. Scripting Advice #2

1. place `save()` and `load()` statements regularly
   - act as backup and **entry** points
2. cache slow code chunks

```
‘‘ ‘{r prepareData, cache=TRUE}
VAR3 <- 1:100
‘‘ ‘

‘‘ ‘{r processData, cache=TRUE, dependson=prepareData}
mean(VAR3)
‘‘ ‘
```

2.11. Including R objects in R scripts

1. Dump the object to console or file

```
> dump('MACNALLY','')
```

```
MACNALLY <-
structure(list(HABITAT = structure(c(4L, 3L, 3L, 3L, 4L, 4L,
4L, 4L, 5L, 2L, 5L, 5L, 4L, 4L, 5L, 4L, 4L, 4L, 4L, 2L, 2L, 4L,
4L, 3L, 1L, 4L, 4L, 4L, 6L, 6L, 6L, 6L, 1L, 1L, 4L, 1L, 2L), .Label = c("Box-Ironbark",
"Foothills Woodland", "Gipps.Manna", "Mixed", "Montane Forest",
"River Red Gum"), class = "factor"), GST = c(3.4, 3.4, 8.4, 3,
5.6, 8.1, 8.3, 4.6, 3.2, 4.6, 3.7, 3.8, 5.4, 3.1, 3.8, 9.6, 3.4,
5.6, 1.7, 4.7, 14, 6, 4.1, 6.5, 6.5, 1.5, 4.7, 7.5, 3.1, 2.7,
4.4, 3, 2.1, 2.6, 3, 7.1, 4.3), EYR = c(0, 9.2, 3.8, 5, 5.6,
4.1, 7.1, 5.3, 5.2, 1.2, 2.5, 6.5, 6.5, 9.3, 3.8, 4, 2.7, 3.3,
2.6, 3.6, 5.6, 4.9, 4.9, 9.7, 2.5, 0, 3.1, 7.5, 0, 0, 0, 0, 1.1,
0, 1.6, 2.2, 2.9)), .Names = c("HABITAT", "GST", "EYR"), class = "data.frame", row.names = c("Reedy Lake",
"Pearcedale", "Warneet", "Cranbourne", "Lysterfield", "Red Hill",
"Devilbend", "Olinda", "Fern Tree Gum", "Sherwin", "Heathcote Ju",
"Warburton", "Millgrove", "Ben Cairn", "Panton Gap", "OShannassy",
"Ghin Ghin", "Minto", "Hawke", "St Andrews", "Nepean", "Cape Schanck",
"Balnarring", "Bittern", "Bailieston", "Donna Buang", "Upper Yarra",
"Gembrook", "Arcadia", "Undera", "Coomboona", "Toolamba", "Rushworth",
"Sayers", "Waranga", "Costerfield", "Tallarook"))
```

2.12. Including R objects in R scripts

1. Dump the object to console or file

```
> dump('MACNALLY','')
```

2. Cut and paste into the top of a script

3. Data within data frames

3.1. Data within data frames

```
> DATA <- data.frame(LOCATION=gl(3,2,6, paste('Location',1:3)),
+                    TREATMENT = gl(2,3,6, LETTERS[1:2]),
+                    Y=rnorm(6,10,2)
+                    )
> DATA
```

```
    LOCATION TREATMENT         Y
1 Location 1         A  8.158481
2 Location 1         A  8.144742
3 Location 2         A  9.969023
4 Location 2         B  9.726616
5 Location 3         B  8.067003
6 Location 3         B 10.797749
```

Your turn

- create this data frame

3.2. Individual vectors

```
> str(DATA)
```

```
'data.frame':   6 obs. of  3 variables:
 $ LOCATION : Factor w/ 3 levels "Location 1","Location 2",..: 1 1 2 2 3 3
 $ TREATMENT: Factor w/ 2 levels "A","B": 1 1 1 2 2 2
 $ Y        : num  8.16 8.14 9.97 9.73 8.07 ...
```

3.3. Individual vectors

Remove individual vectors

```
> LOCATION
```

```
Error in eval(expr, envir, enclos): object 'LOCATION' not found
```

```
> DATA$LOCATION
```

```
[1] Location 1 Location 1 Location 2 Location 2 Location 3 Location 3
Levels: Location 1 Location 2 Location 3
```

3.4. Individual vectors

```
> with(DATA, LOCATION)
```

```
[1] Location 1 Location 1 Location 2 Location 2 Location 3 Location 3
Levels: Location 1 Location 2 Location 3
```

3.5. What next

All this is foundation is awesome...

If only we knew how to summarise and plot all of these data....

On to pres.2.4 and pres.5.2