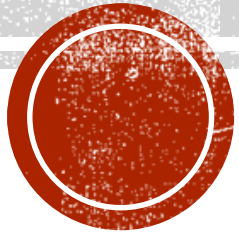
DATA ANALYSIS

Data Import & Export



READING AND WRITING DATA IN R

- There are a few principal functions reading data into R.
- read.table, read.csv, for reading tabular data
- readLines, for reading lines of a text file
- source, for reading in R code files (inverse of dump)
- dget, for reading in R code files (inverse of dput)
- load, for reading in saved workspaces
- unserialize, for reading single R objects in binary form

READING AND WRITING DATA IN R

- There are analogous functions for writing data to files

write.table

Write.csv

writeLines

dump

dput

save

serialize

READING DATA INTO R

- R can read data from a variety of file formats—for example, files created as text, or in Excel, SPSS or Stata.

- We will mainly be reading files in text format .txt or .csv (comma-separated, usually created in Excel).

READING DATA INTO R

- To read an entire data frame directly, the external file will normally have a special form

- The first line of the file should have a name for each variable in the data frame.

- Each additional line of the file has as its first item a row label and the values for each variable

READING DATA INTO R

- Here we use the example dataset called airquality.csv and airquality.txt

Input file form with names and row labels:

	Ozone	Solar.R	Wind	Temp	Month	Day
1	41	190	7.4	67	5	1
2	36	118	8.0	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5

READING DATA INTO R

- By default numeric items (except row labels) are read as numeric variables. This can be changed if necessary.

- The function read.table() can then be used to read the data frame directly

Ex:

```
> airqual <- read.table("airquality.txt")
```

READING DATA INTO R

- Similarly, to read .csv files the read.csv() function can be used to read in the data frame directly

Ex

```
> airqual <- read.csv("airquality.csv")
```

READING DATA INTO R

Ex2:

airqual <- read.table(" C:/Desktop/airquality.txt")

airqual <- read.csv(" C:/Desktop/airquality.csv")

Note: Occasionally you'll need to do a double slash in your path //. This seems to depend on the machine.

airqual <- read.table(" C://Desktop//airquality.txt")

airqual <- read.csv(" C://Desktop//airquality.csv")

READING DATA INTO R

Ex2:

airqual <- read.table(" C:/Desktop/airquality.txt")

airqual <- read.csv(" C:/Desktop/airquality.csv")

Note: Occasionally you'll need to do a double slash in your path //. This seems to depend on the machine.

airqual <- read.table(" C://Desktop//airquality.txt")

airqual <- read.csv(" C://Desktop//airquality.csv")

ASSIGNING COLUMN AND ROW NAMES

- Occasionally, you will need to read in data that does not already have column name information.

- For example, the dataset BOD.txt looks like this:

```
1    8.3
2   10.3
3   19.0
4   16.0
5   15.6
7   19.8
```

ASSIGNING COLUMN AND ROW NAMES

- Read the data from the text file BOD.txt

```
> bod <- read.table("BOD.txt", header=F)
> bod
  V1   V2
1  1  8.3
2  2 10.3
3  3 19.0
4  4 16.0
5  5 15.6
6  7 19.8
```

ASSIGNING COLUMN AND ROW NAMES

- The command to assign column names

```
> colnames(bod) <- c("Time","demand")
>
> bod
  Time demand
1    1    8.3
2    2   10.3
3    3   19.0
4    4   16.0
5    5   15.6
6    7   19.8
```

ASSIGNING COLUMN AND ROW NAMES

- The command to assign row names

```
> rownames(bod) <- c("R1","R2","R3","R4","R5","R6")
> bod
   Time demand
R1    1    8.3
R2    2   10.3
R3    3   19.0
R4    4   16.0
R5    5   15.6
R6    7   19.8
```

ASSIGNING COLUMN AND ROW NAMES

- Creating a new text file mod_bod.txt to save the updates.

```
> write.table(bod,"mod_bod.txt")
> read.table("mod_bod.txt")
   Time demand
R1    1    8.3
R2    2   10.3
R3    3   19.0
R4    4   16.0
R5    5   15.6
R6    7   19.8
```

WRITING DATA TO A FILE

- After working with a dataset, we might like to save it for future use.

- Before we do this, let's first set up a working directory so we know where we can find all our data sets and files later.

SETTING UP A DIRECTORY

- In the R window, click on "File" and then on "Change dir".

- You should then see a box pop up titled "Choose directory".

Ex:

- To choose the directory "Desktop" by clicking on "Browse", then select "Desktop" and click "OK".

SETTING UP A DIRECTORY

- Alternatively, you can use the setwd() function to assign as working directory.

```
> setwd("C:/Desktop")
```

- To find out what your current working directory is, type

```
> getwd()
```

WRITING DATA TO A FILE

- If you wat to save the data (with quotes on strings)

```
> write.table(cars, file="carsl.txt")
> testl <- read.table("carsl.txt")
> testl
   speed dist
1      4    2
2      4   10
3      7    4
4      7   22
5      8   16
6      9   10
7     10   18
```

WRITING DATA TO A FILE

- If you wat to save the data (without quotes)

```
>write.table(cars, file="cars1.txt", quote=F)
```

- You can see the variation in cars1.txt file by opening it externally