Poli 5D Social Science Data Analytics
Introduction to R

Shane Xinyang Xuan
ShaneXuan.com

February 15, 2017

Contact Information

Shane Xinyang Xuan

xxuan@ucsd.edu

The teaching staff is a team!

Professor Roberts	M	1600-1800 (SSB 299)
Jason Bigenho	Th	1000-1200 (Econ 116)
Shane Xuan	M	1100-1150 (SSB 332)
	Th	1200-1250 (SSB 332)

Supplemental Materials

UCLA STATA starter kit

`http://www.ats.ucla.edu/stat/stata/sk/`

Princeton data analysis

`http://dss.princeton.edu/training/`

Road map

Some quick notes before we start today's section:

- Make sure that you pass around the attendance sheet

Road map

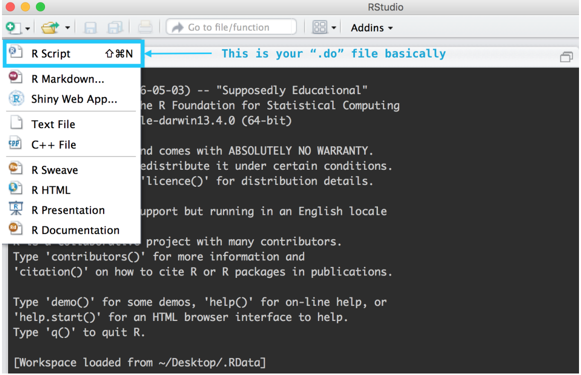
Some quick notes before we start today's section:

- Make sure that you pass around the attendance sheet
- Open the R console (RStudio recommended); I will be using my slides, but you should work in your console

Road map

Some quick notes before we start today's section:

- Make sure that you pass around the attendance sheet
- Open the R console (RStudio recommended); I will be using my slides, but you should work in your console



This is your ".do" file basically

```
6-05-03) -- "Supposedly Educational"
he R Foundation for Statistical Computing
le-darwin13.4.0 (64-bit)

nd comes with ABSOLUTELY NO WARRANTY.
edistribute it under certain conditions.
'licence()' for distribution details.

upport but running in an English locale

 project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/Desktop/.RData]
```

-

Set up

- Changing working directory (setwd):

Set up

- Changing working directory (setwd):
  setwd("/Users/Shane/Dropbox/Poli5D/dataforlecture")

Set up

- Changing working directory (setwd):
  setwd("/Users/Shane/Dropbox/Poli5D/dataforlecture")

– Your working directory depends on where you saved your folder

Set up

- Changing working directory (setwd): setwd("/Users/Shane/Dropbox/Poli5D/dataforlecture")

– Your working directory depends on where you saved your folder

- Read in data (read.csv):

Set up

- Changing working directory (setwd):
  setwd(“/Users/Shane/Dropbox/Poli5D/dataforlecture”)
- – Your working directory depends on where you saved your folder
- Read in data (read.csv):
  data ← read.csv (“moms_clean.csv”)

Set up

- Changing working directory (setwd):
  setwd("/Users/Shane/Dropbox/Poli5D/dataforlecture")
- Your working directory depends on where you saved your folder
- Read in data (read.csv):
  data ← read.csv ("moms_clean.csv")
- Note that ← is an assignment operator in R
- What you did with this command is to load the data, and throw them in an object called "data"

Basic functions

- names() gives you the names of an object

Basic functions

- names() gives you the names of an object
- head() or tail() returns the first or last parts of a vector

Basic functions

- names() gives you the names of an object
- head() or tail() returns the first or last parts of a vector
- summary() is similar to the summarize function in Stata

Basic functions

- names() gives you the names of an object
- head() or tail() returns the first or last parts of a vector
- summary() is similar to the summarize function in Stata
- nrow() and ncol() return the number of rows and columns of the data

Data frame

- In R, we use
  - Parentheses for function
  - Brackets for indices
  - Dollar signs for variables

Data frame

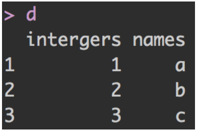
- In R, we use
  - Parentheses for function
  - Brackets for indices
  - Dollar signs for variables
- A data frame is similar to a matrix; you can use is.data.frame() to check if the data set is a data frame

Data frame (2)

- Create a data frame:
  d ← data.frame(integers=1:3, names=c(”a”,”b”,”c”))

Data frame (2)

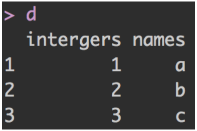
- Create a data frame:
  d ← data.frame(integers=1:3, names=c(”a”,”b”,”c”))



```
> d
  intergers names
1         1     a
2         2     b
3         3     c
```

Data frame (2)

- Create a data frame:
  d ← data.frame(integers=1:3, names=c(”a”,”b”,”c”))

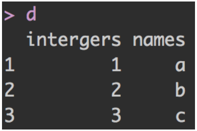
  

  ```
  > d
    intergers names
  1         1     a
  2         2     b
  3         3     c
  ```

- Row and column in a matrix: d[row, column]

Data frame (2)

- Create a data frame:
  d ← data.frame(integers=1:3, names=c(”a”,”b”,”c”))



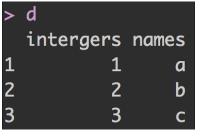
```
> d
  intergers names
1         1     a
2         2     b
3         3     c
```

- Row and column in a matrix: d[row, column]
  d[,1] is the first column

Data frame (2)

- Create a data frame:

  d ← data.frame(integers=1:3, names=c(”a”,”b”,”c”))

  ```
  > d
    intergers names
  1         1     a
  2         2     b
  3         3     c
  ```

- Row and column in a matrix: d[row, column]

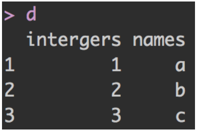
  d[,1] is the first column

  d[1,] is the first row

Data frame (2)

- Create a data frame:

  d ← data.frame(integers=1:3, names=c(”a”,”b”,”c”))

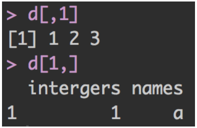


```
> d
  intergers names
1         1     a
2         2     b
3         3     c
```

- Row and column in a matrix: d[row, column]

  d[,1] is the first column

  d[1,] is the first row



```
> d[,1]
[1] 1 2 3
> d[1,]
  intergers names
1         1     a
```

Data frame (3)

- Specify a variable in the data frame: use $

Data frame (3)

- Specify a variable in the data frame: use $
- For example, if a data frame called poli5 contained a variable called pset, you can access the variable pset using poli5$pset

Other functions

- mean(); median()

Other functions

- mean(); median()
- Use seq() to create a sequence :

Other functions

- mean(); median()
- Use seq() to create a sequence :

  seq(from, to, by)

Other functions

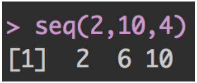
- mean(); median()
- Use seq() to create a sequence :

  seq(from, to, by)
- What is the output of seq(2,10,4)?

Other functions

- mean(); median()
- Use seq() to create a sequence :

  seq(from, to, by)
- What is the output of seq(2,10,4)?

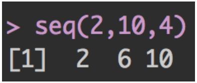


```
> seq(2,10,4)
[1]  2  6 10
```

Other functions

- mean(); median()
- Use seq() to create a sequence :

  seq(from, to, by)
- What is the output of seq(2,10,4)?

```
> seq(2,10,4)
[1]  2  6 10
```

- table()