Introduction to R

Introduction to R for Public Health Researchers

Welcome to class!

1. Introductions
2. Class overview
3. Getting R up and running

About Us

John Muschelli

Assistant Scientist, Department of Biostatistics

PhD in Biostatistics, ScM in Biostatistics

Email: jmusche1@jhu.edu

About Us

Kayode Sosina

PhD Student Biostatistics

Email: ksosina1@jhu.edu

About Us

Andrew Jaffe

Lead Investigator, Lieber Institute for Brain Development

Assistant Professor, Department of Mental Health, JHSPH

PhD in Epidemiology, MHS in Bioinformatics

Email: ajaffe@jhu.edu

What is R?

- R is a language and environment for statistical computing and graphics
- R is the open source implementation of the S language, which was developed by Bell laboratories
- R is both open source and open development

(source: http://www.r-project.org/)

Why R?

- Powerful and flexible
- Free (open source)
- Extensive add-on software (packages)
- Designed for statistical computing
- High level language

Why not R?

- Fairly steep learning curve
  - “Programming” oriented
  - Minimal interface
- Little centralized support, relies on online community and package developers
- Annoying to update
- Slower, and more memory intensive, than the more traditional programming languages (C, Java, Perl, Python)

Introductions

What do you hope to get out of the class?

Why else to use R?

Course Website

http://johnmuschelli.com/intro_to_r

Materials will be uploaded the night before class

Course Package

We have an `R` package called `jhur` that will make sure all the packages are installed.

```
install.packages("remotes")
remotes::install_github("muschellij2/jhur")
```

```
library(jhur)
head(read_yts()) # read youth tobacco data
```

```
# A tibble: 6 x 31
   YEAR LocationAbbr LocationDesc TopicType TopicDesc MeasureDesc
  <int> <chr>        <chr>        <chr>     <chr>     <chr>
1  2015 AZ           Arizona      Tobacco … Cessatio… Percent of…
2  2015 AZ           Arizona      Tobacco … Cessatio… Percent of…
3  2015 AZ           Arizona      Tobacco … Cessatio… Percent of…
4  2015 AZ           Arizona      Tobacco … Cessatio… Quit Attem…
5  2015 AZ           Arizona      Tobacco … Cessatio… Quit Attem…
6  2015 AZ           Arizona      Tobacco … Cessatio… Quit Attem…
# ... with 25 more variables: DataSource <chr>, Response <chr>,
#   Data_Value_Unit <chr>, Data_Value_Type <chr>, Data_Value <dbl>,
#   Data_Value_Footnote_Symbol <chr>, Data_Value_Footnote <chr>,
#   Data_Value_Std_Err <dbl>, Low_Confidence_Limit <dbl>,
#   High_Confidence_Limit <dbl>, Sample_Size <int>, Gender <chr>,
#   Race <chr>, Age <chr>, Education <chr>, GeoLocation <chr>,
#   TopicTypeId <chr>, TopicId <chr>, MeasureId <chr>,
#   StratificationID1 <chr>, StratificationID2 <chr>,
```

Learning Objectives

- Reading data into R
- Recoding and manipulating data
- Writing R functions and using add-on packages
- Making exploratory plots
- Understanding basic programming syntax
- Performing basic statistical tests

Installing R

- Install the latest version from: http://cran.r-project.org/
- Install RStudio

Collection of R packages

We have an R package called jhur that will make sure all the packages are installed.

You can just copy and paste the below code into your console - we'll explain what it all means in the next day or two

```
install.packages("remotes")
remotes::install_github("muschellij2/jhur")
```

Note it may take ~5-10 minutes to run.

Useful (+Free) Resources

- R for Data Science: http://r4ds.had.co.nz/
- Various “Cheat Sheets”: https://www.rstudio.com/resources/cheatsheets/
- DataCamp http://www.datacamp.com
- R reference card: http://cran.r-project.org/doc/contrib/Short-refcard.pdf
- UCLA Institute for Digital Research and Education: http://www.ats.ucla.edu/stat/r/
- Quick R: http://statmethods.net/

Website

Website