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

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())
```

```
Parsed with column specification:
cols(
  .default = col_character(),
  YEAR = col_integer(),
  Data_Value = col_double(),
  Data_Value_Std_Err = col_double(),
  Low_Confidence_Limit = col_double(),
  High_Confidence_Limit = col_double(),
  Sample_Size = col_integer(),
  DisplayOrder = col_integer()
)

See spec(...) for full column specifications.

# A tibble: 6 x 31
   YEAR LocationAbbr LocationDesc TopicType    TopicDesc  MeasureDesc
  <int> <chr>        <chr>        <chr>        <chr>      <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

Useful (+Free) Resources

- Homework will involve working through: http://tryr.codeschool.com/
- DataCamp http://www.datacamp.com
- UCLA Institute for Digital Research and Education: http://www.ats.ucla.edu/stat/r/
- R reference card: http://cran.r-project.org/doc/contrib/Short-refcard.pdf
- Undergrad Guide to R: https://sites.google.com/site/undergraduateguidetor/
- Quick R: http://statmethods.net/

Website

Website