STAT540: Computing in Statistics
Getting Started

Dr. Yen-Yi Ho (hoyen@stat.sc.edu)

Objectives

- ▶ Why R?
- ▶ The R Environment (Console and Scripts)
- ▶ RStudio
- ▶ Installing R packages

Origins of R

- ▶ S language - developed by John Chambers of Bell Labs[1].
- ▶ Commercial implementation: S-plus
- ▶ Free software implementation: R (named in honor of founders Robert Gentleman and Ross Ihaka)
- ▶ Object-oriented language
- ▶ Each object has a "type" (vector, factor, matrix, function, data frame, etc..)
- ▶ Important to be aware of what type each object is.

---

[1]Becker, Richard A. "A brief history of S." Computational statistics (1994): 81-110.

The New York Times

Business Computing

Data Analysts Captivated by R's Power

Left, Stuart Isett for The New York Times; right, Kieran Scott for The New York Times

R first appeared in 1996, when the statistics professors Robert Gentleman, left, and Ross Ihaka released the code as a free software package.

By ASHLEE VANCE
Published: January 6, 2009

To some people R is just the 18th letter of the alphabet. To others, it's the rating on racy movies, a measure of an attic's insulation or what pirates in movies say.

Related

R is also the name of a popular programming language used by a

FACEBOOK

TWITTER

GOOGLE+

SAVE

EMAIL

R computing & graphics package

- ▶ R is free and open source[2].
- ▶ It runs on all major platforms: Windows, Mac OS, UNIX/Linux.
- ▶ Popular with many researchers due to contributed packages: R functions to do specialized, advanced, & often complex statistical analysis.
- ▶ R can also do many important, routine calculations, analysis, and provide common graphical displays used in this course.
- ▶ Installed in several of the computing labs across campus, e.g. Sloan 108 & 109, Gambrell 003.
- ▶ You can download it and install it from CRAN: http://cran.r-project.org/

---

[2]https://opensource.org/history

R: Pros and Cons

Pros
+ Free
+ Available for all major platforms
+ Powerful graphics
+ Comprehensive
+ Easy interface with other languages (such as C, Fortran)
+ Well-designed programming language (object-oriented)
+ Unlimited extensibility
+ Widely used by statisticians
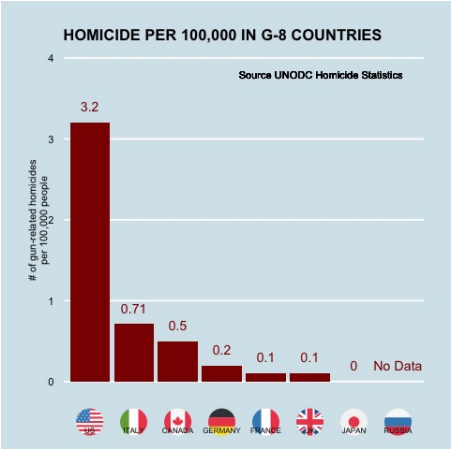+ Increasingly used for genomic analyses (Bioconductor)

Cons
- No dedicated support
- Complex Syntax

- Not point-and-click
- No warranty

- Relatively slow

R Console

As a quick example, try using the console to calculate a 15% tip on a meal that cost $19.71.
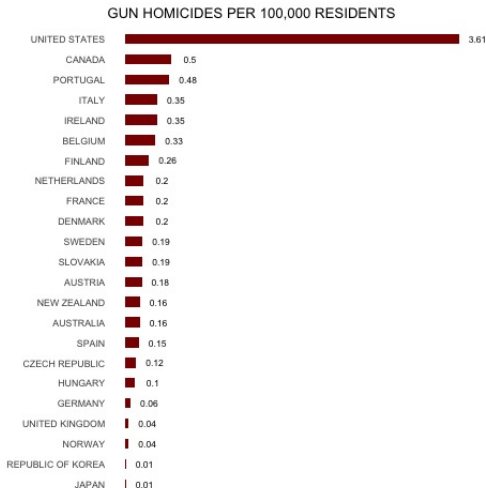
Motivation

US Gun Homicide Rate[3]

HOMICIDE PER 100,000 IN G-8 COUNTRIES

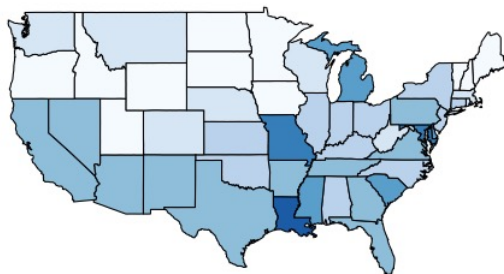Source UNODC Homicide Statistics

---

[3]http://abcnews.go.com/blogs/headlines/2012/12/us-gun-ownership-homicide-rate-higher-than-other-developed-countries/

Motivation

US Gun Homicide Rate from everytown [4]

GUN HOMICIDES PER 100,000 RESIDENTS

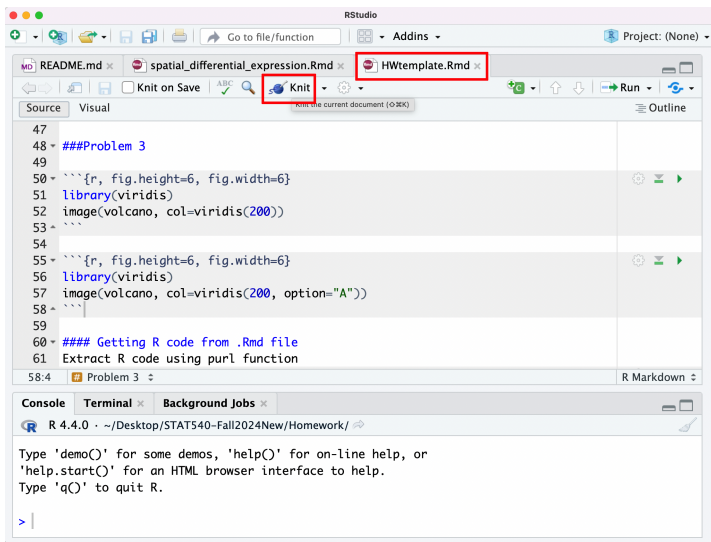|	
UNITED STATES	3.61
CANADA	0.5
PORTUGAL	0.48
ITALY	0.35
IRELAND	0.35
BELGIUM	0.33
FINLAND	0.26
NETHERLANDS	0.2
FRANCE	0.2
DENMARK	0.2
SWEDEN	0.19
SLOVAKIA	0.19
AUSTRIA	0.18
NEW ZEALAND	0.16
AUSTRALIA	0.16
SPAIN	0.15
CZECH REPUBLIC	0.12
HUNGARY	0.1
GERMANY	0.06
UNITED KINGDOM	0.04
NORWAY	0.04
REPUBLIC OF KOREA	0.01
JAPAN	0.01

---

[4] everytown.org2

Motivation

- ▶ California, for example, has a larger population than Canada,
- ▶ 20 US states have populations larger than that of Norway.
- ▶ Furthermore, although not included in the charts above, the murder rates in Lithuania, Ukraine, and Russia are higher than 4 per 100,000.

R Studio: Do this now!!!

Global options Installing R packages

Objects