The R Environment

Casimir Saternos

pluralsight
hardcore dev and IT training

Module Overview

What, Why, How?

R as a Calculator

Loading packages

Data Structures

Programming

What is R?

R is a language and environment for statistical computing and graphics.
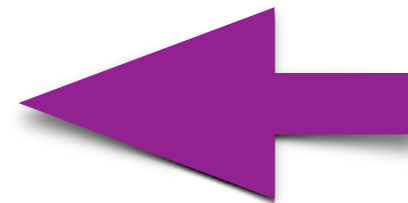
http://www.r-project.org/

Who uses R?

Spreadsheet Users

Programmers

Statisticians

Data Scientists

R Installation and Initial Setup

Select R Version

Download/Install R

Download/Install RStudio

Set up Text Editors

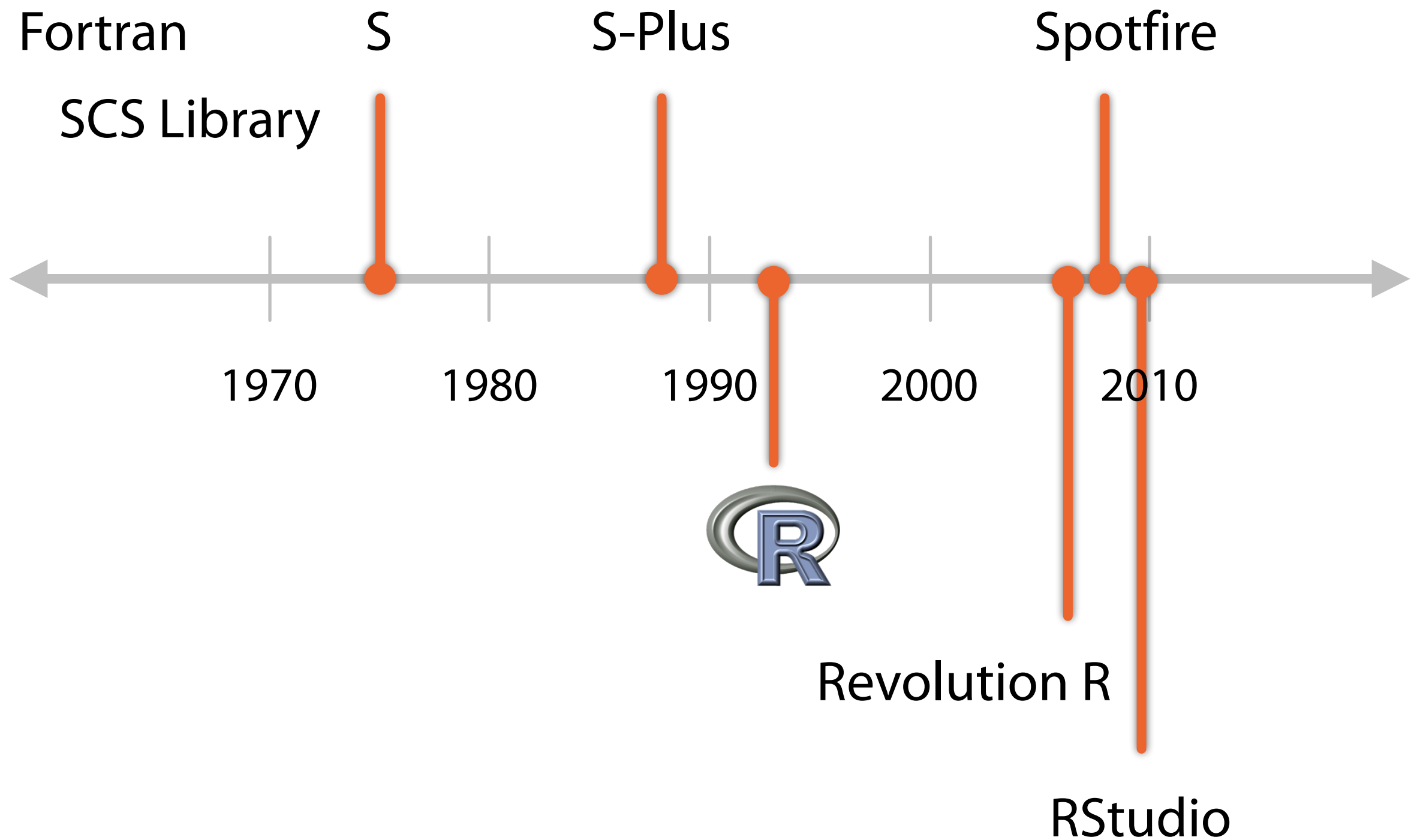Related Software

Select R Version

http://cran.r-project.org/

| Linux | (Mac) OS X | Windows |

A Brief History of R

Fortran

SCS Library

S

S-Plus

Spotfire

1970 1980 1990 2000 2010

Revolution R

RStudio

R and Text Editors

TextMate

Vim

Notepad++

Sublime Text

Emacs

BBEdit

RStudio

R and Related Software

http://strata.oreilly.com/2014/01/2013-data-science-salary-survey.html
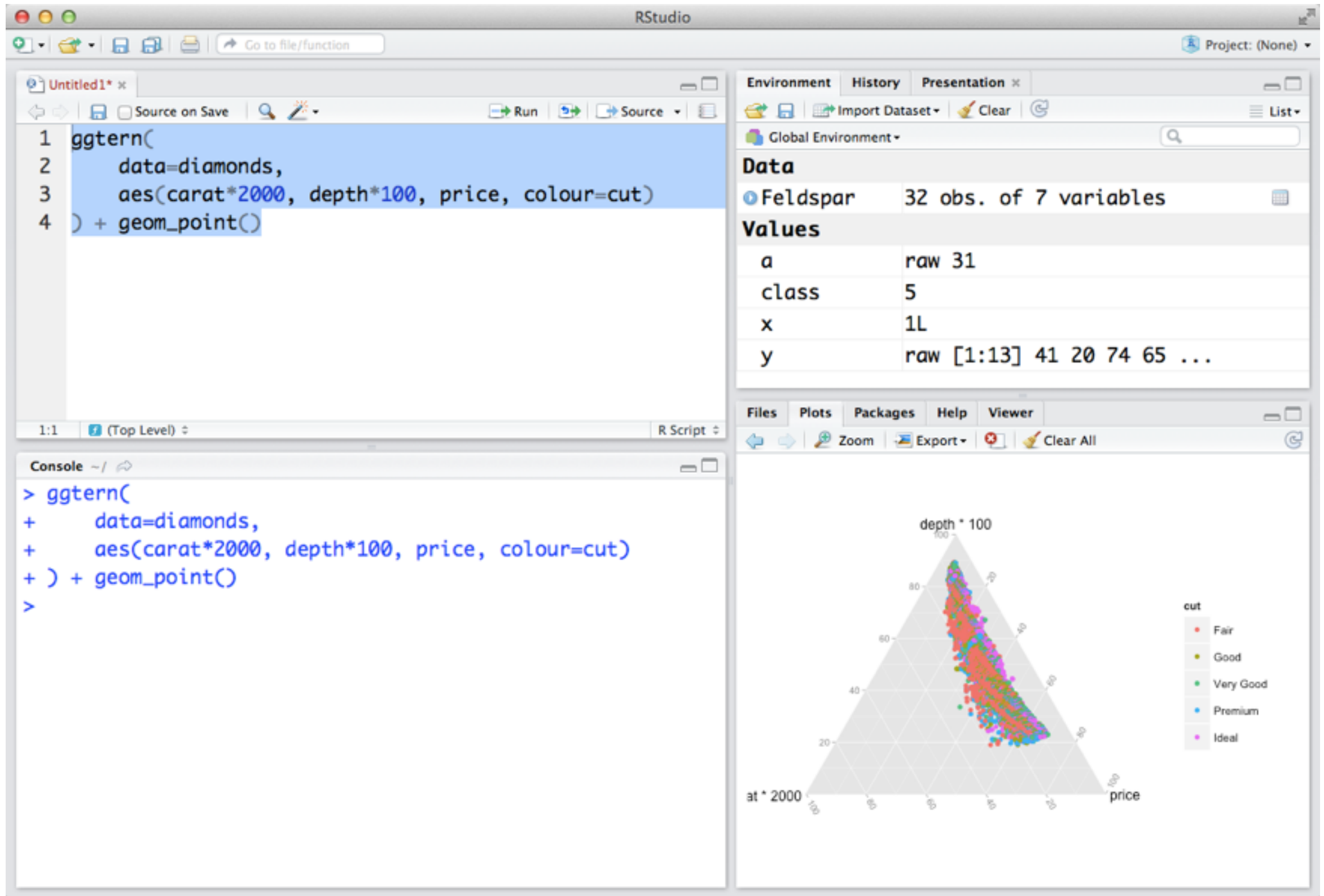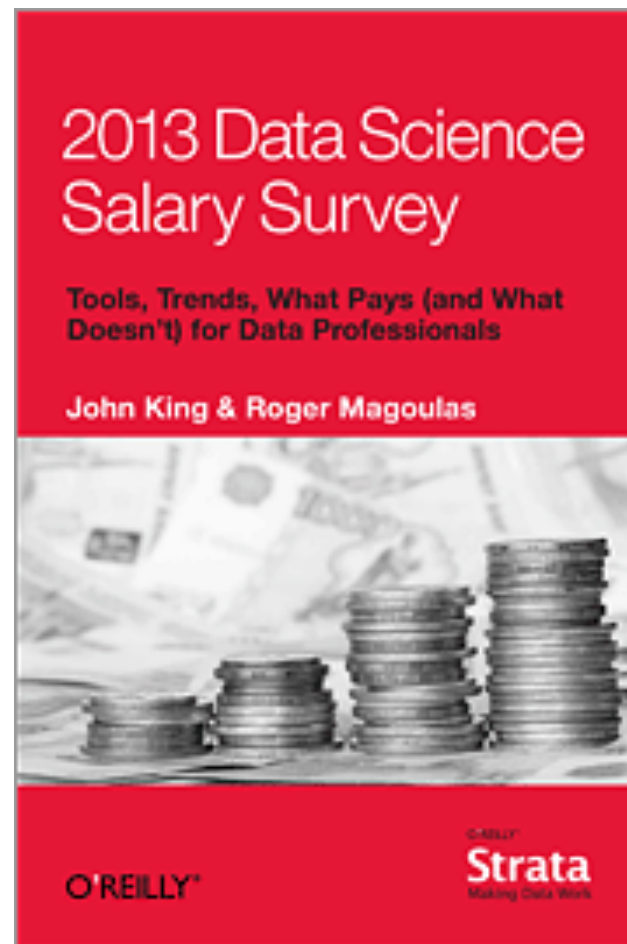
R and Related Software

By a significant margin, more respondents used SQL than any other tool (71% of respondents, compared to 43% for the next highest ranked tool, R).

— O'Reilly 2013 Data Science Salary Survey

R and Related Software

…open source tools (R, Python, Hadoop frameworks, and several scalable machine learning tools)…

— O'Reilly 2013 Data Science Salary Survey

R and Related Software

It seems very likely that knowing how to use tools such as R, Python, Hadoop frameworks, D3, and scalable machine learning tools qualifies an analyst for more highly paid positions.

— O'Reilly 2013 Data Science Salary Survey

R and Related Software

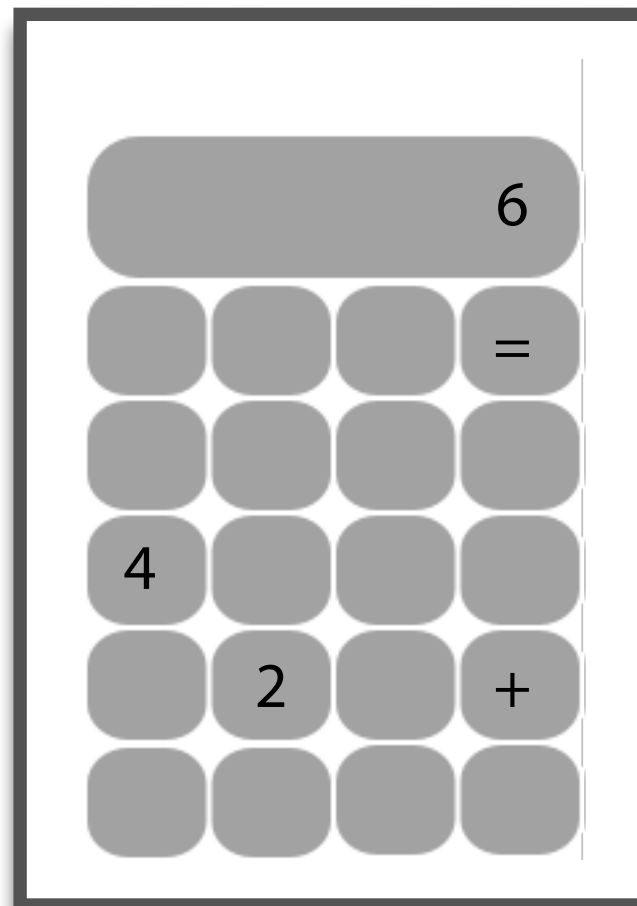D3   Hadoop

Ruby   Java

Python   SQL

R as a Calculator

R More Than a Calculator

Variables

```
myVar = 5

myVar <- 5
```
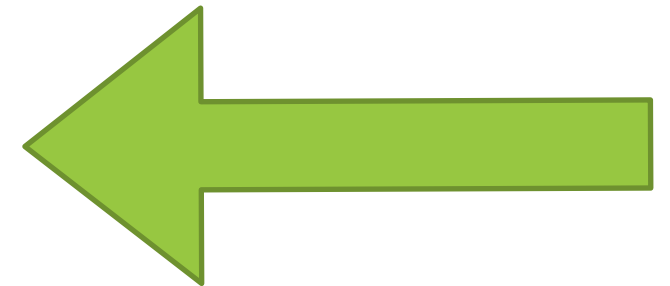
Vectorized Operations

Vector

`c()`

96   6   7

[1]   [2]   [3]

R as a Calculator

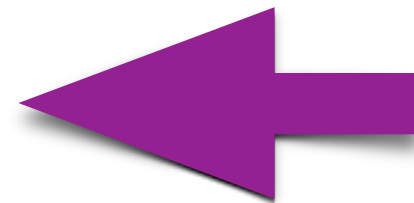**R**ead
Eval
Print
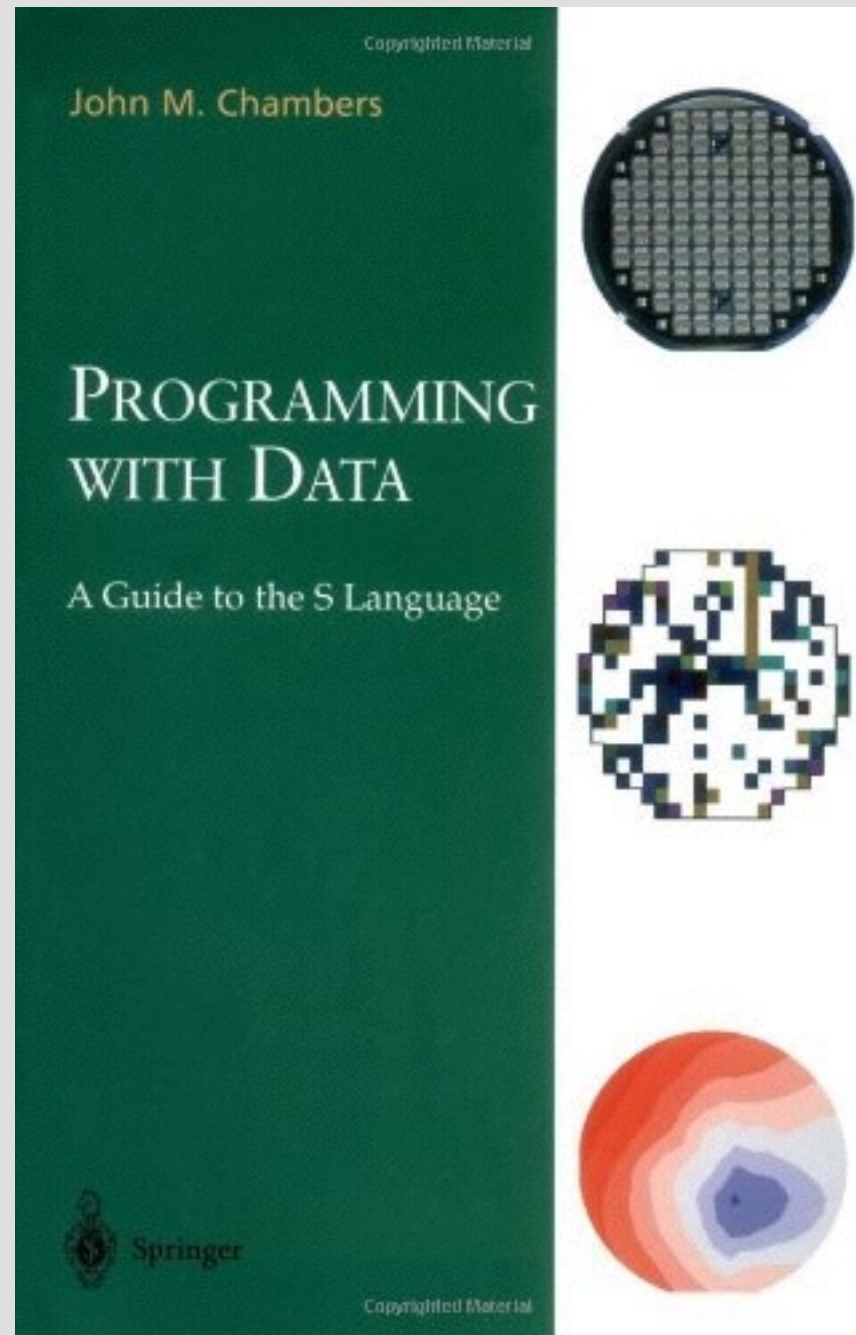**L**oop

Interactive
Shell

R as a Calculator

Constants

Operators

Functions

Variables

Programming with Data

S is a programming language and environment for all kinds of computing involving data.

— John M. Chambers

S specializes in *computing with data*: any application with interesting requirements for <u>organizing</u> <u>analyzing</u> or <u>presenting</u> data is a candidate.

— John M. Chambers

Organizing, Analyzing & Presenting Data

| Organize | Analyze | Present |

Data in R

Data is Collective

Data Structures

Data Types

Data is Collective

Data … occurs in two constructions: as a plural noun…. and as an abstract mass noun (like information) .

— http://www.merriam-webster.com/dictionary/data

R Data Structures

Dimensions	Types	
	1	**n**
1	vector `c()`	list `list()`
2	matrix `matrix()`	data frame `data.frame()`
n	array `array()`	

Source: http://adv-r.had.co.nz/Data-structures.html

R Data Types

	`typeof()`	`mode()`	`storage.mode()`
TRUE	logical	logical	logical
1L	integer	numeric	integer
1	double	numeric	double
'1'	character	character	character
2i	complex	complex	complex
charToRaw('1')	raw	raw	raw

R Data Types

Significance of R's Type System

Weakly Typed

Type Coercion

Flexible

Surprising

Programming

Paradigms

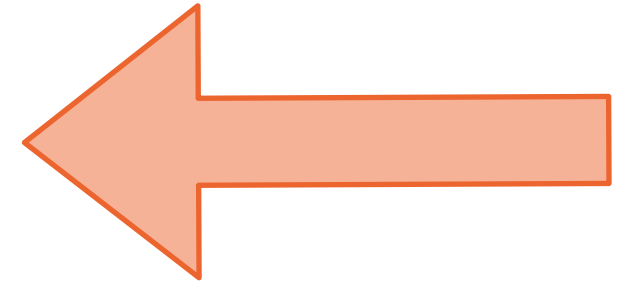|	
Structured	Sequence, Selection, Repetition
Functional	1st Class, Lazy
Object Oriented	S3 (informal), S4 (formal), others
Dynamic	Flexible Type System, eval

Module Overview

What, Why, How?

R as a Calculator

Loading packages

Data Structures

Programming

Course Overview

✔ Environment	Importing Data	Creating Visualizations
Packages	Exporting/ Presenting Data	R RStudio