

Exploratory Data Analysis with R

Introduction to R



Matthew Renze

@matthewrenze | www.matthewrenze.com

Overview



Introduction to R

Transforming and Cleaning Data

Calculating Descriptive Statistics

Visualizing Data

Moving Beyond R and EDA

Prerequisites

Experience with a C-like language

Audience

Developers

Data Analysts

IT Professionals

A Flood of Data Is Coming

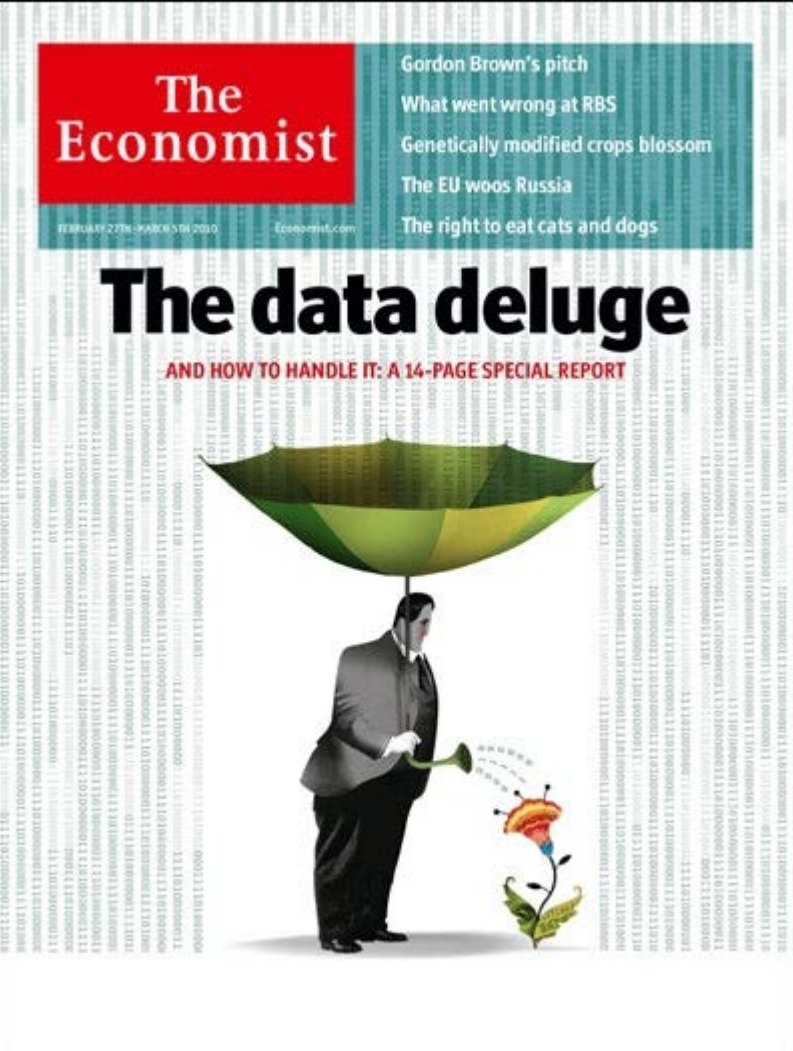


Sink

or



Swim



The New York Times

For Today's Graduate, Just One Word: Statistics

By STEVE LORIK
Published: August 5, 2009

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls “all the computer and math stuff” that was part of the job.

- [TWITTER](#)
- [LINKEDIN](#)
- [COMMENTS \(58\)](#)
- [SIGN IN TO E-MAIL](#)

AVERAGE SALARY FOR High Paying Skills and Experience		
SKILL	2013	YR/YR CHANGE
R	\$ 115,531	n/a
NoSQL	\$ 114,796	1.6%
MapReduce	\$ 114,396	n/a
PMBok	\$ 112,382	1.3%
Cassandra	\$ 112,382	n/a
Omnigraffle	\$ 111,039	0.3%
Pig	\$ 109,561	n/a
SOA (Service Oriented Architecture)	\$ 108,997	-0.5%
Hadoop	\$ 108,669	-5.6%
Mongo DB	\$ 107,825	-0.4%

Source: Dice 2014 Tech Salary Survey Results

How Does This Apply to Me?

- ☑ Perform log file analysis
- ☑ Analyze performance
- ☑ Analyze code metrics
- ☑ Detect anomalies
- ☑ Transform or clean data
- ☑ Help decision makers

What Is R?

Open source

Language and environment

Numerical and graphical

Cross platform



What Is R?

Active development

Large user community

Modular and extensible

6700+ extensions



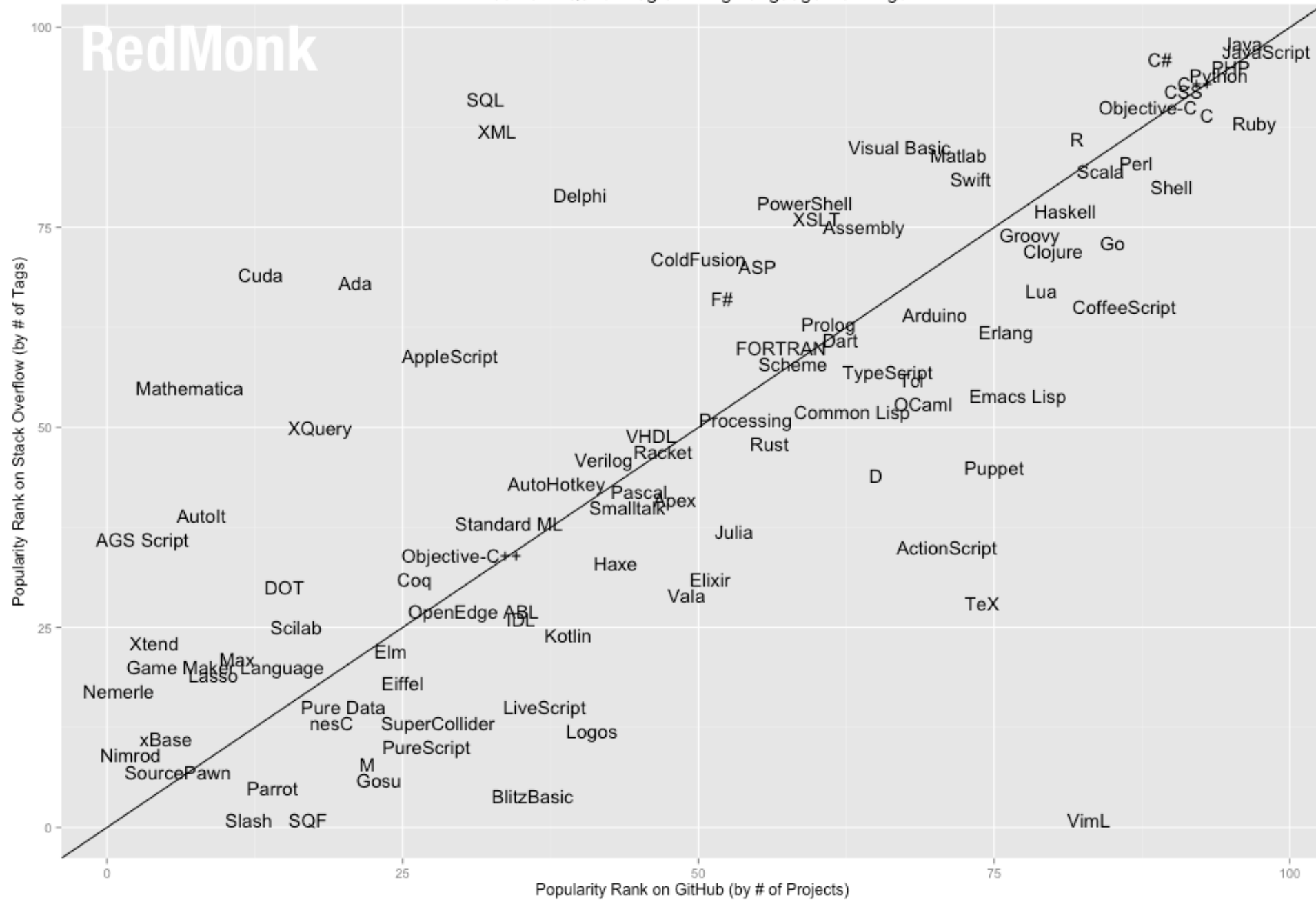


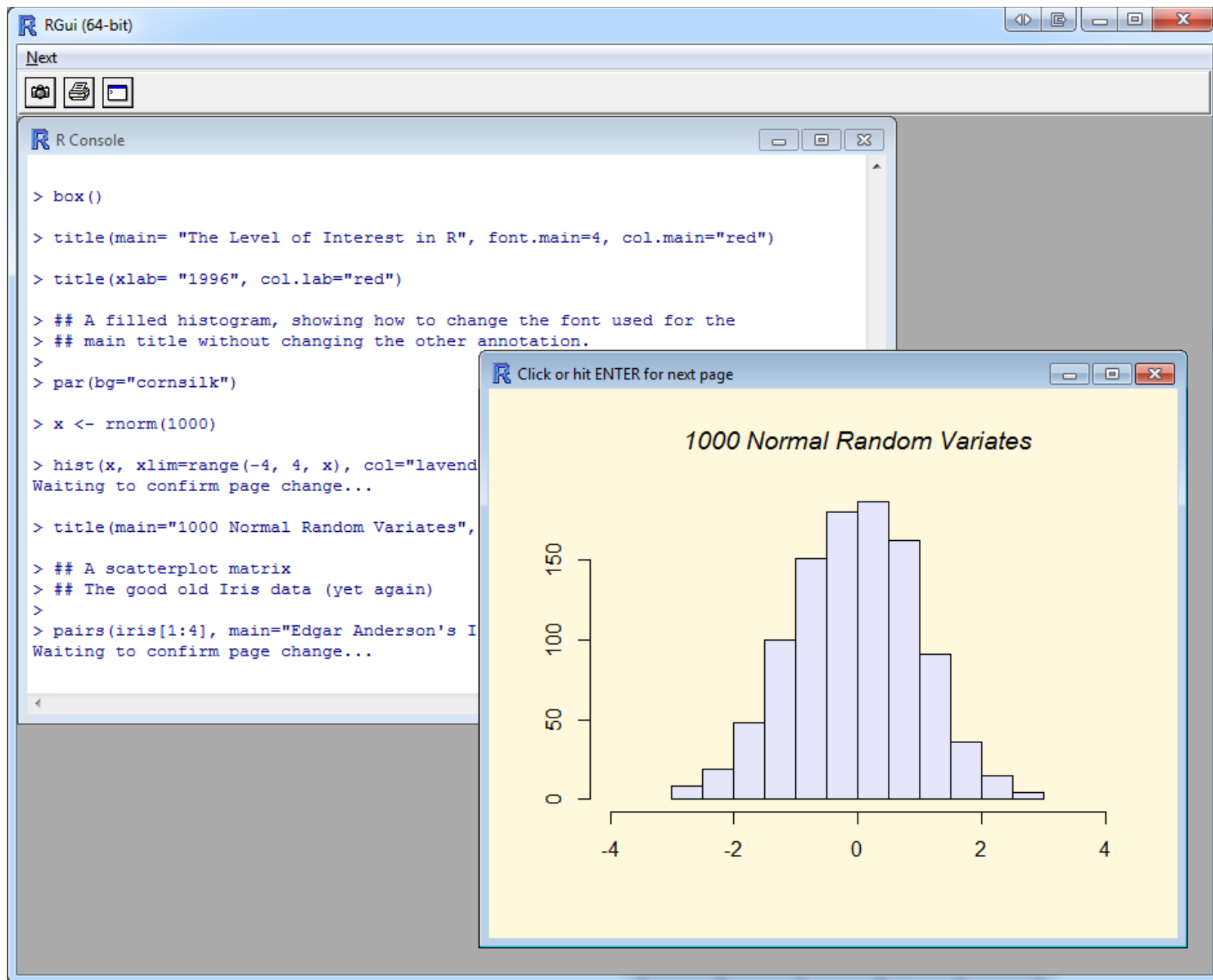
FREE

FREE

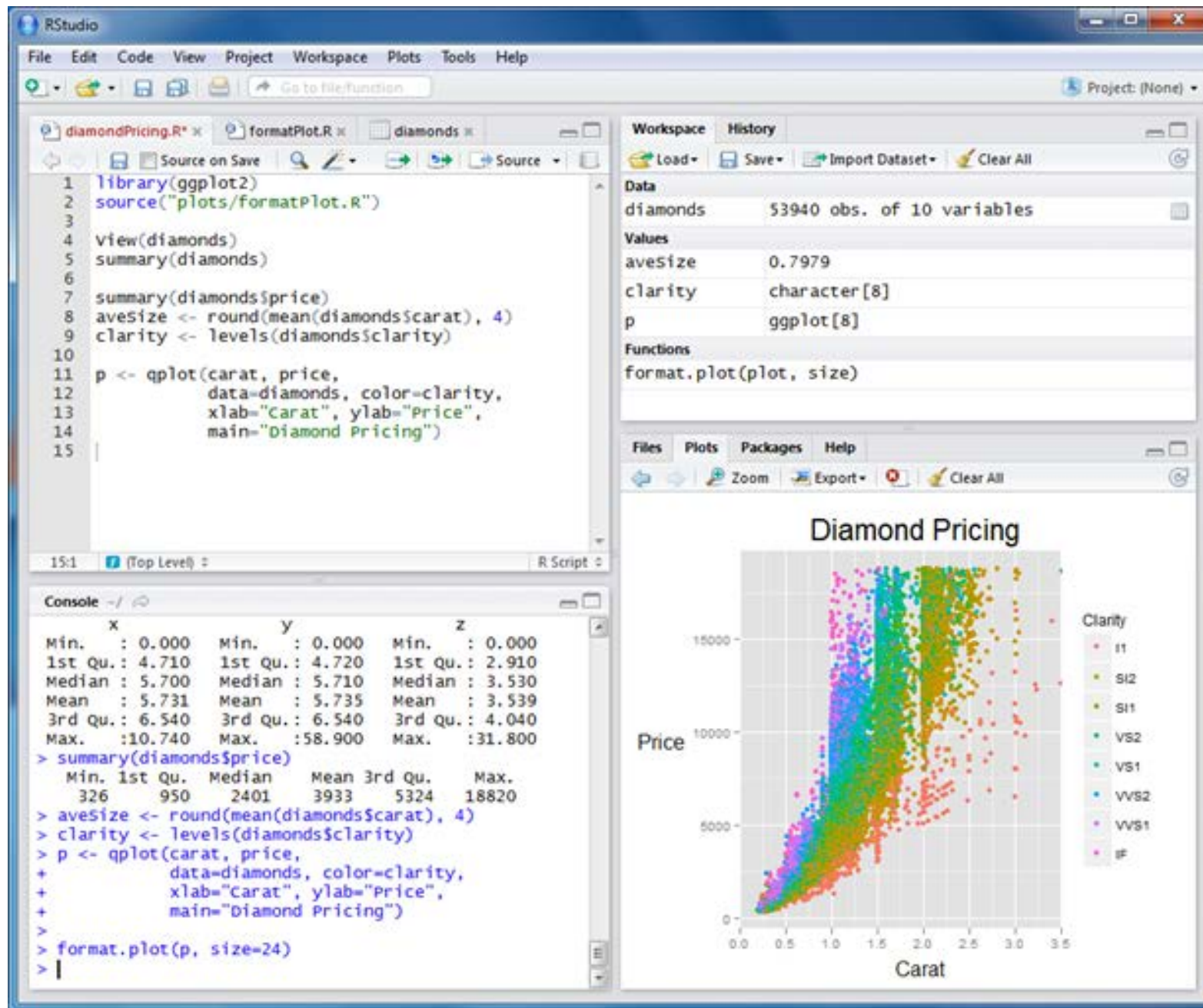


RedMonk Q315 Programming Language Rankings





Hello World Demo



Language Basics Demo

Scripts vs. Console

Hello World in R Studio

Assignment Operators

Implicit Printing

Creating Variables

Displaying Variables

Functions

Vectors

Sequences

Matrices

Arrays

Lists

Factors

Data Frame

Indexing

Subsetting

Subsetting (cont.)

Vectorized Operations

Named Arguments

Installing Packages

Getting Help

Other Language Features

Control structures

Mathematical operators

String manipulation

Summary

Introduction to R
RStudio
Language basics

