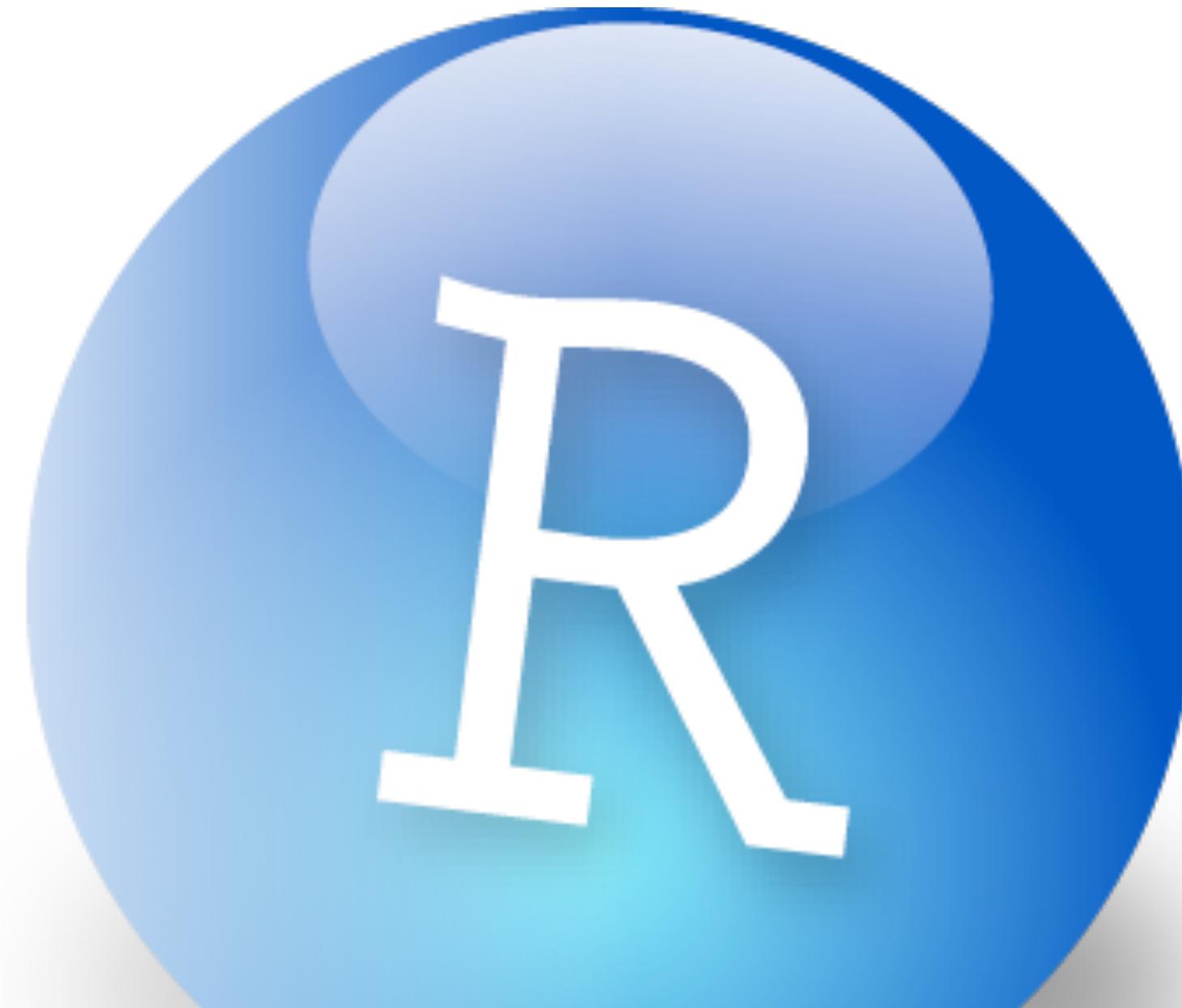


All Training materials are provided "as is" and without warranty and RStudio disclaims any and all express and implied warranties including without limitation the implied warranties of title, fitness for a particular purpose, merchantability and noninfringement.

The Training Materials are licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit<http://creativecommons.org/licenses/by-nc/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Introduction

Learn to create and run
programs in R



Garrett Grolemund

Master Instructor, RStudio

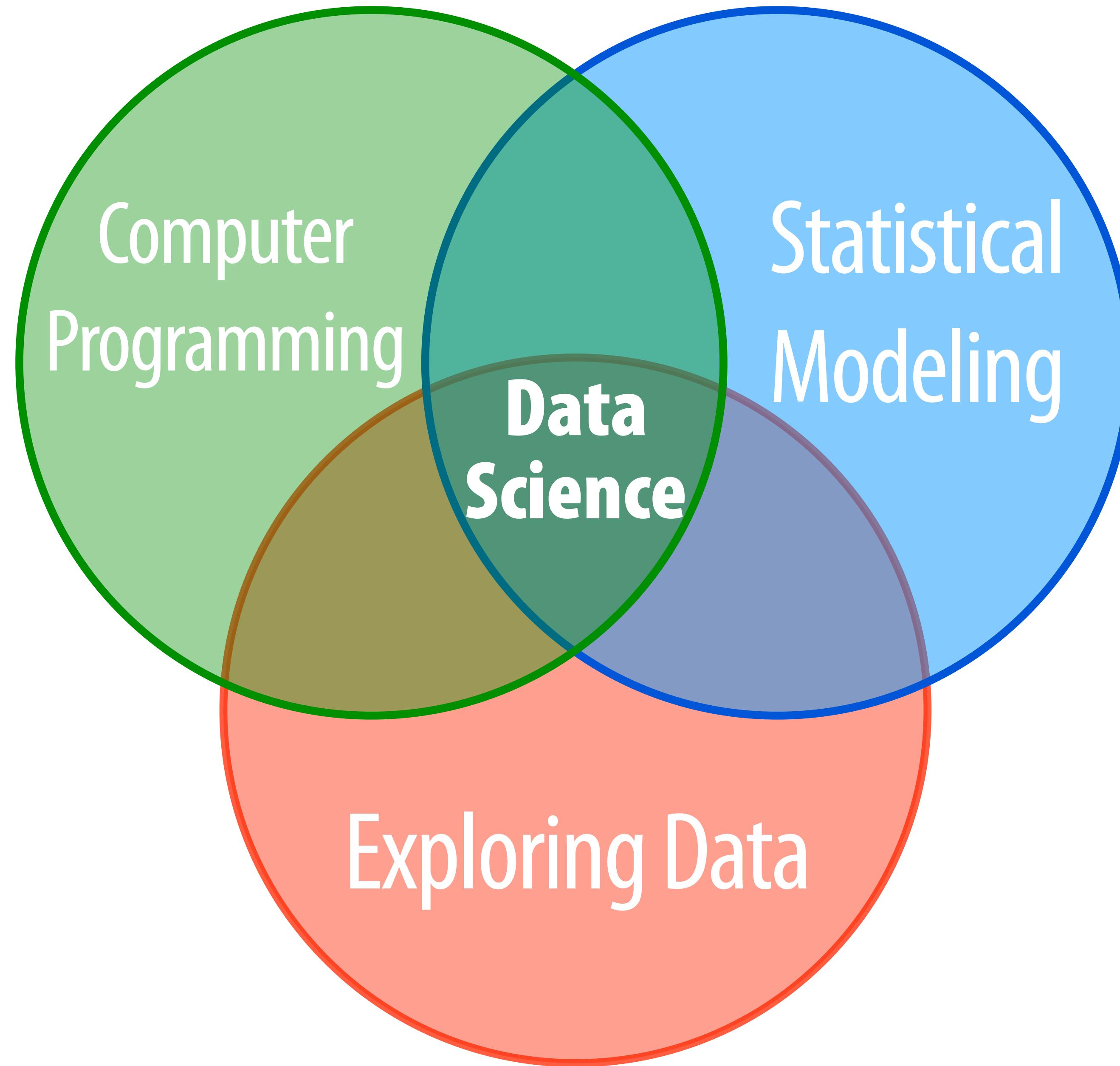
August 2014

HELLO

my name is

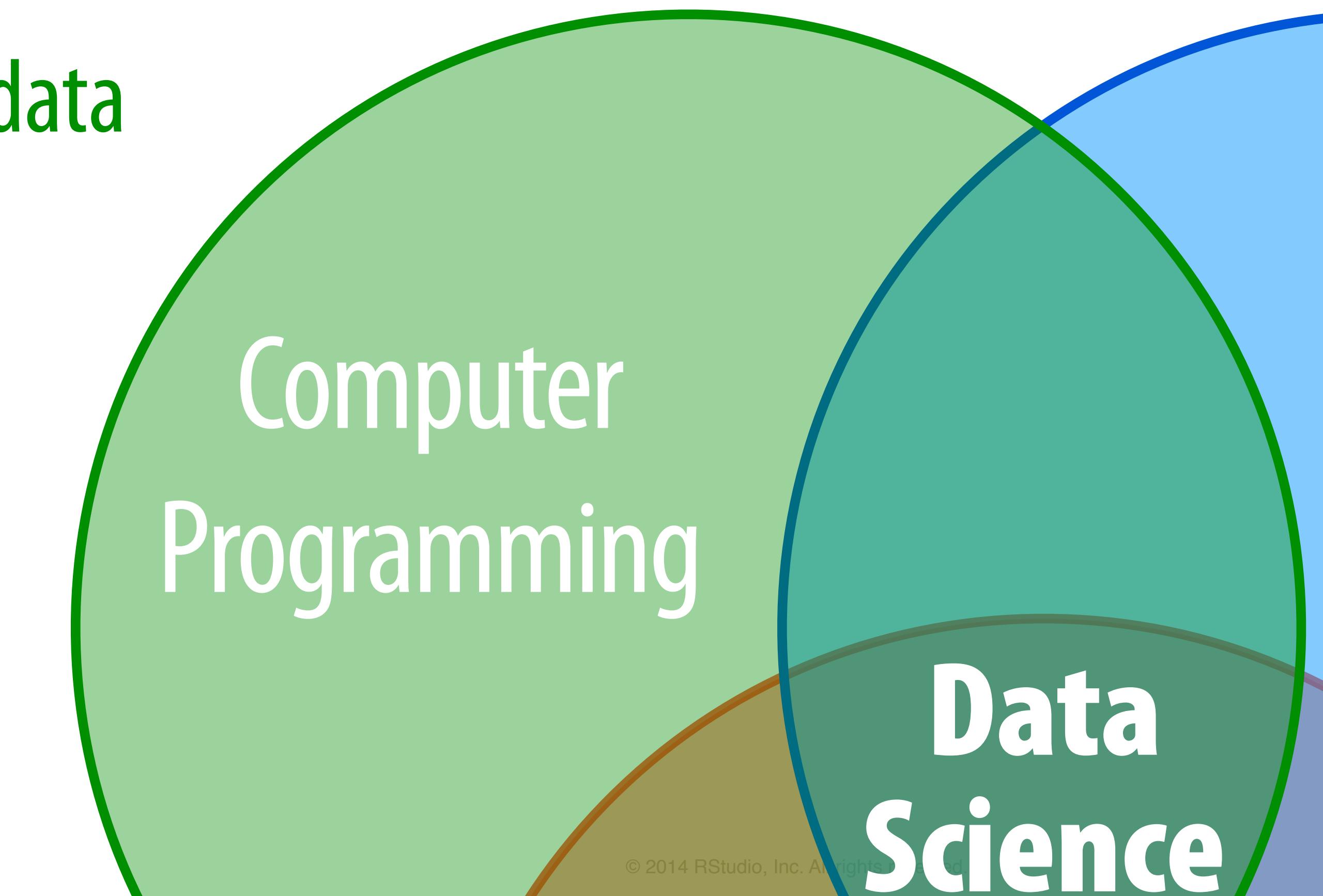
Garrett

garrett@rstudio.com



Modules 1 & 2

1. The R Language
2. Storing and retrieving data



Exploring Data

- 3. Visualizing data
- 4. Customizing plots

Modules 3 & 4

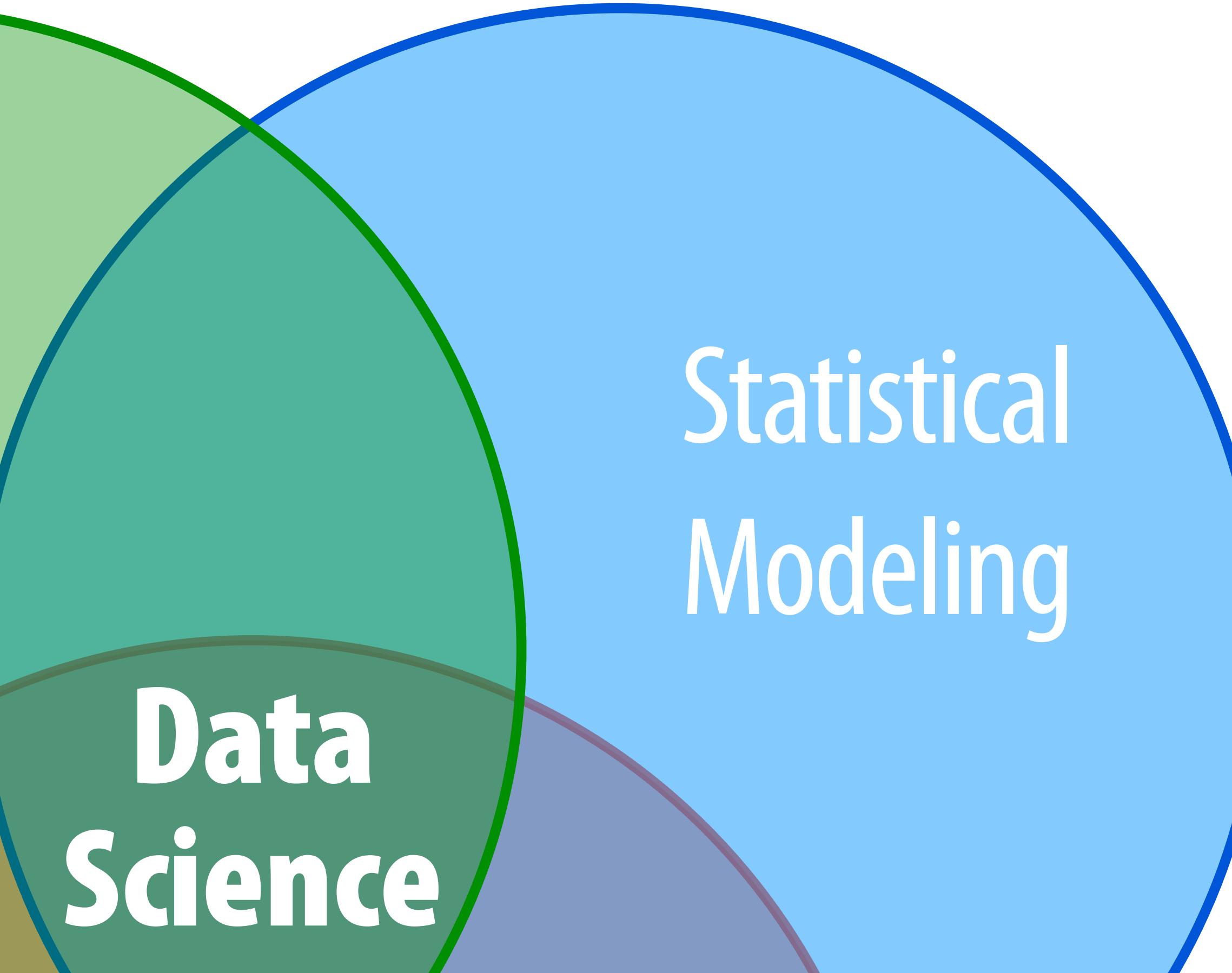
Exploring Data

- 5. Loading and cleaning data
- 6. Manipulating data

Modules 5 & 6

Modules 7 & 8

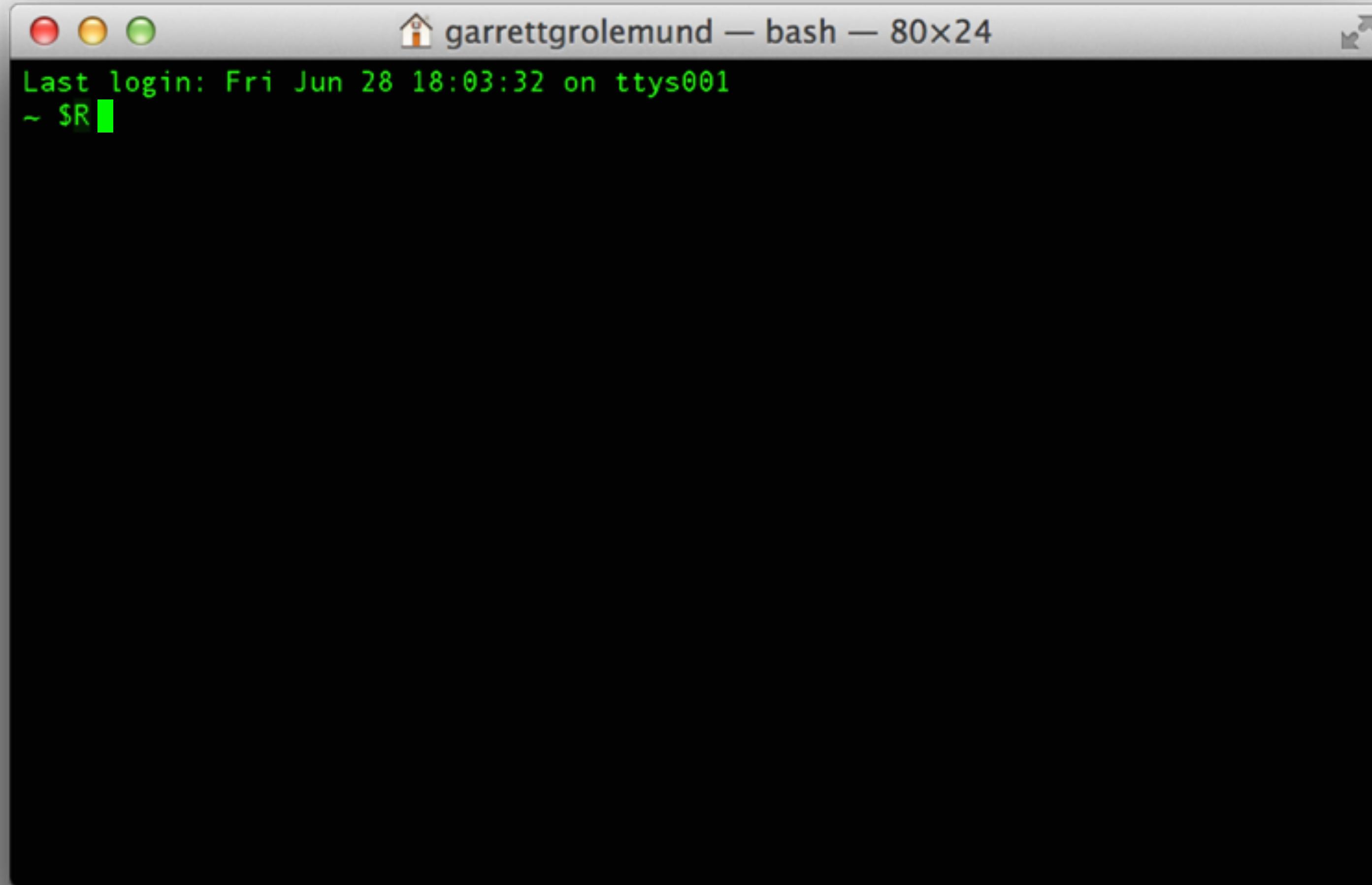
- 7. Linear regression
- 8. Non-linear models
 - a. Logistic regression
 - b. Modeling survey



R &
RStudio

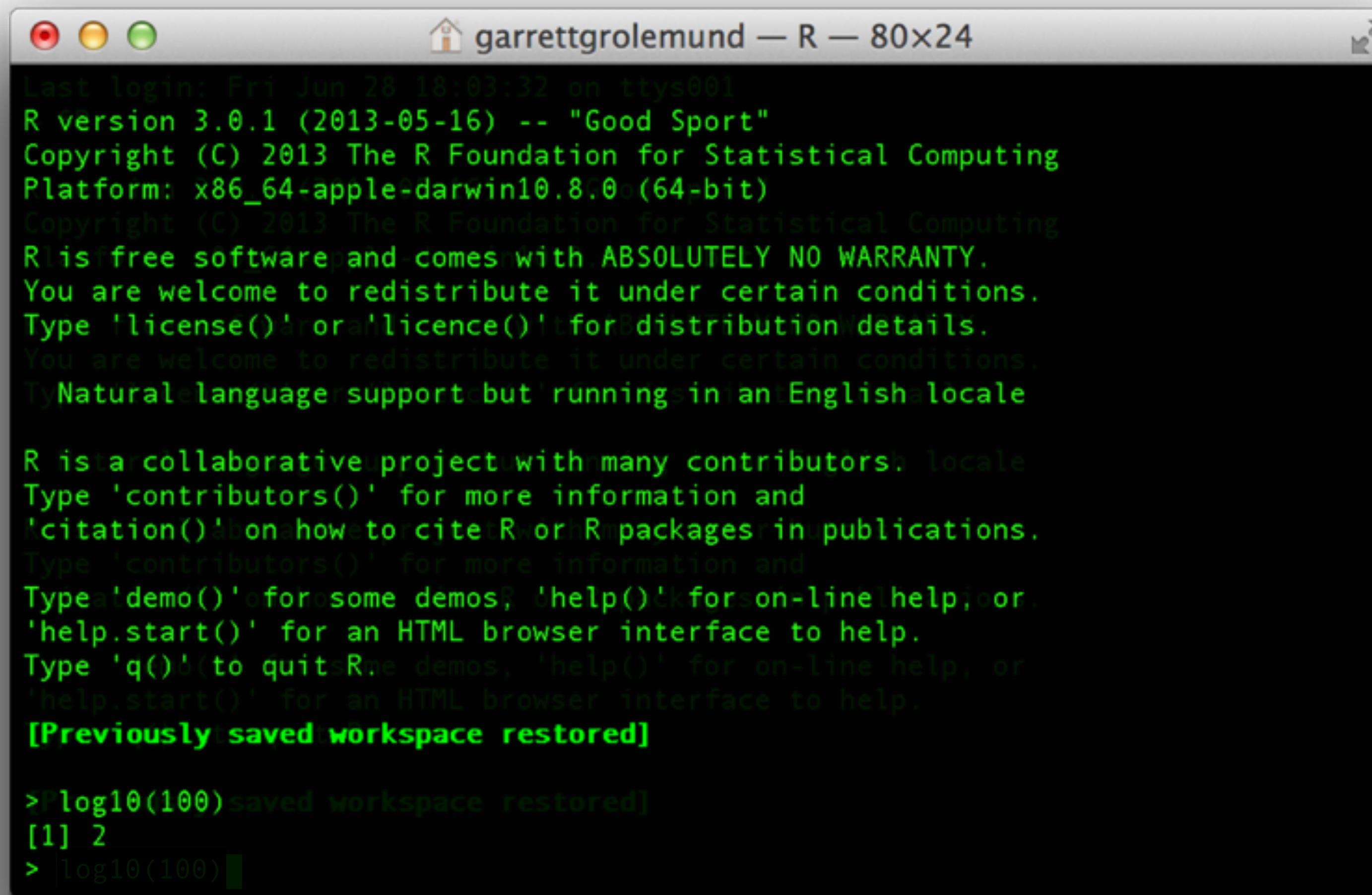


R: a computer programming language



```
Last login: Fri Jun 28 18:03:32 on ttys001
~ $R
```

R: a computer programming language



The screenshot shows the RStudio interface with the R console window open. The title bar reads "garrettgrolemund — R — 80x24". The console displays the standard R startup message, including the last login date, version information, and a detailed explanation of the GPL license. It also mentions that R is a collaborative project and provides links to contributors and citation information. The message concludes with the text "[Previously saved workspace restored]" in red. At the bottom of the console, there is a command prompt with the text "> log10(100) saved workspace restored" and the output "[1] 2".

```
Last login: Fri Jun 28 18:03:32 on ttys001
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin10.8.0 (64-bit)

Copyright (C) 2013 The R Foundation for Statistical Computing
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
You are welcome to redistribute it under certain conditions.
Type 'Natural language support but running in an English locale

R is a collaborative project with many contributors. See
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'contributors()' for more information and
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R or 'demos', 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
[Previously saved workspace restored]

> log10(100) saved workspace restored
[1] 2
> log10(100)
```

R: a computer programming language

1. Descends from S, Bell Laboratories
2. Evolved in university environment
3. Full scripting language
4. ...but can be used as a simple application

> mpg

	manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
1	audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
6	audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
7	audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact
8	audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
9	audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
10	audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p	compact
11	audi	a4 quattro	2.0	2008	4	auto(s6)	4	19	27	p	compact
12	audi	a4 quattro	2.8	1999	6	auto(l5)	4	15	25	p	compact
13	audi	a4 quattro	2.8	1999	6	manual(m5)	4	17	25	p	compact
14	audi	a4 quattro	3.1	2008	6	auto(s6)	4	17	25	p	compact
15	audi	a4 quattro	3.1	2008	6	auto(s6)	4	17	25	p	compact

```
> mean(mpg$hwy)
```

```
[1] 23.44017
```

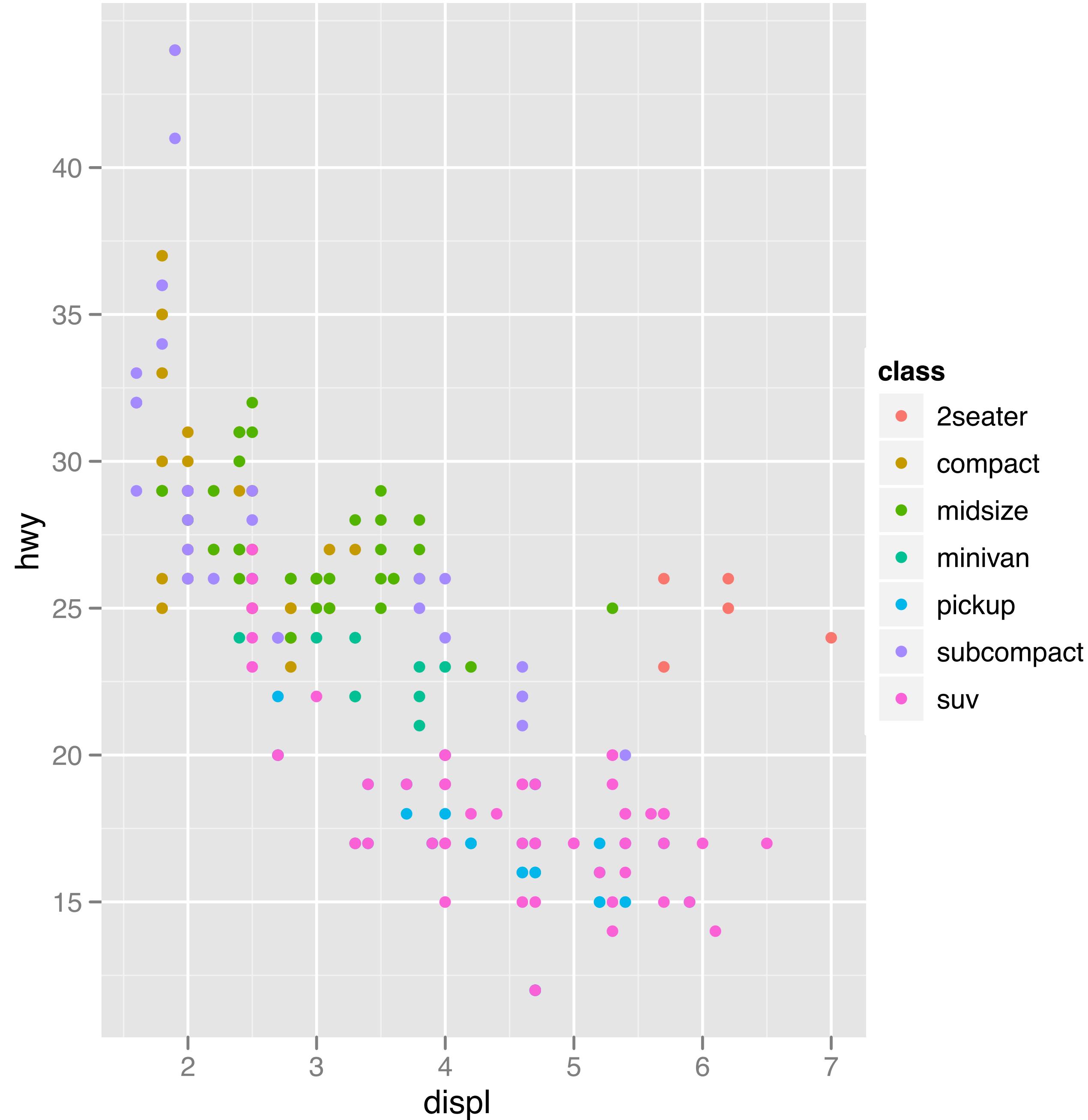
```
> summary(mpg$hwy)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
12.00	18.00	24.00	23.44	27.00	44.00

```
> table(mpg$class)
```

2seater	compact	midsize	minivan	pickup	suv
5	47	41	11	33	62

```
> qplot(displ, hwy, data = mpg, color = class)
```



```
> lm(hwy ~ displ, data = mpg)
```

Call:

```
lm(formula = hwy ~ displ, data = mpg)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.1039	-2.1646	-0.2242	2.0589	15.0105

Coefficients:

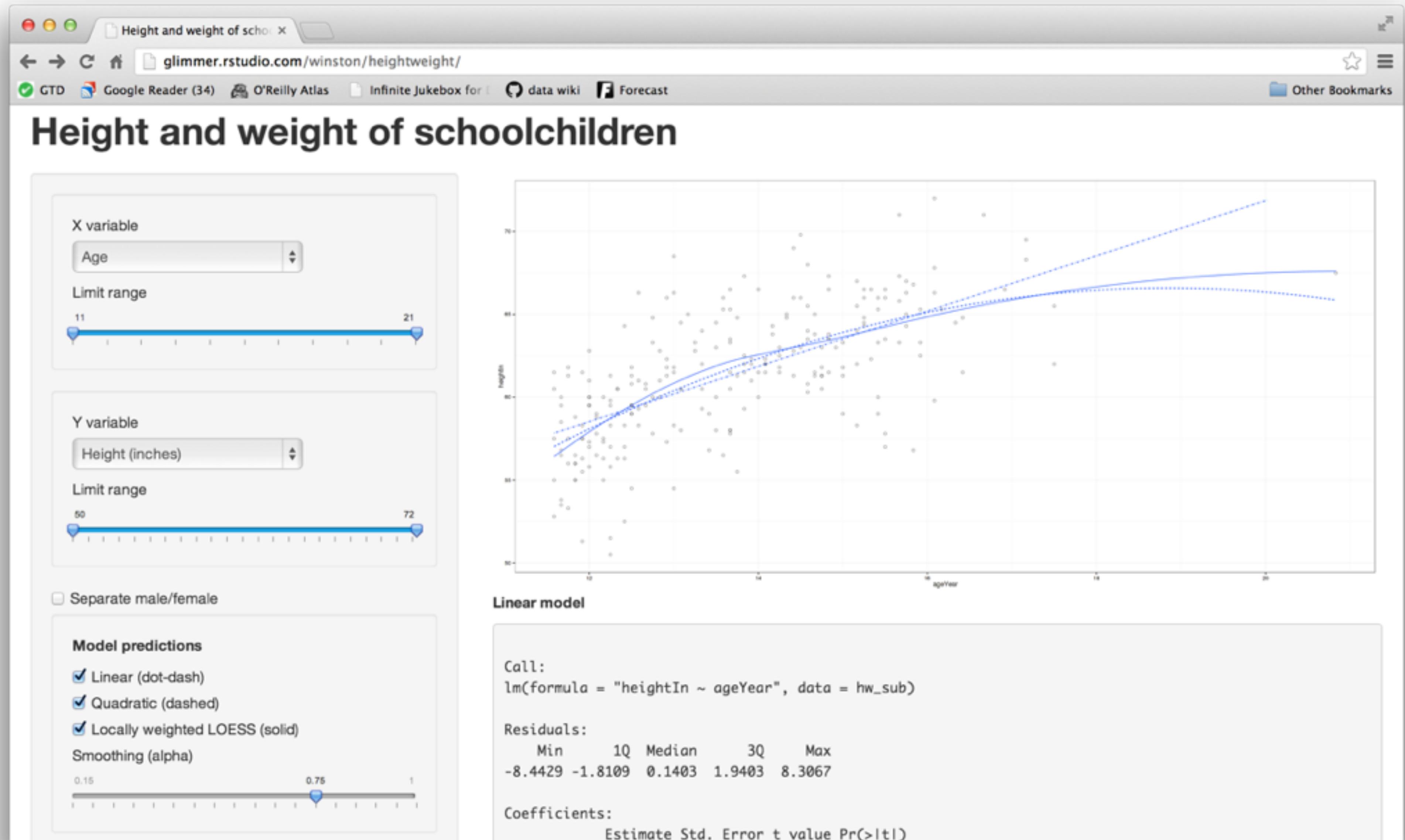
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	35.6977	0.7204	49.55	<2e-16 ***
displ	-3.5306	0.1945	-18.15	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

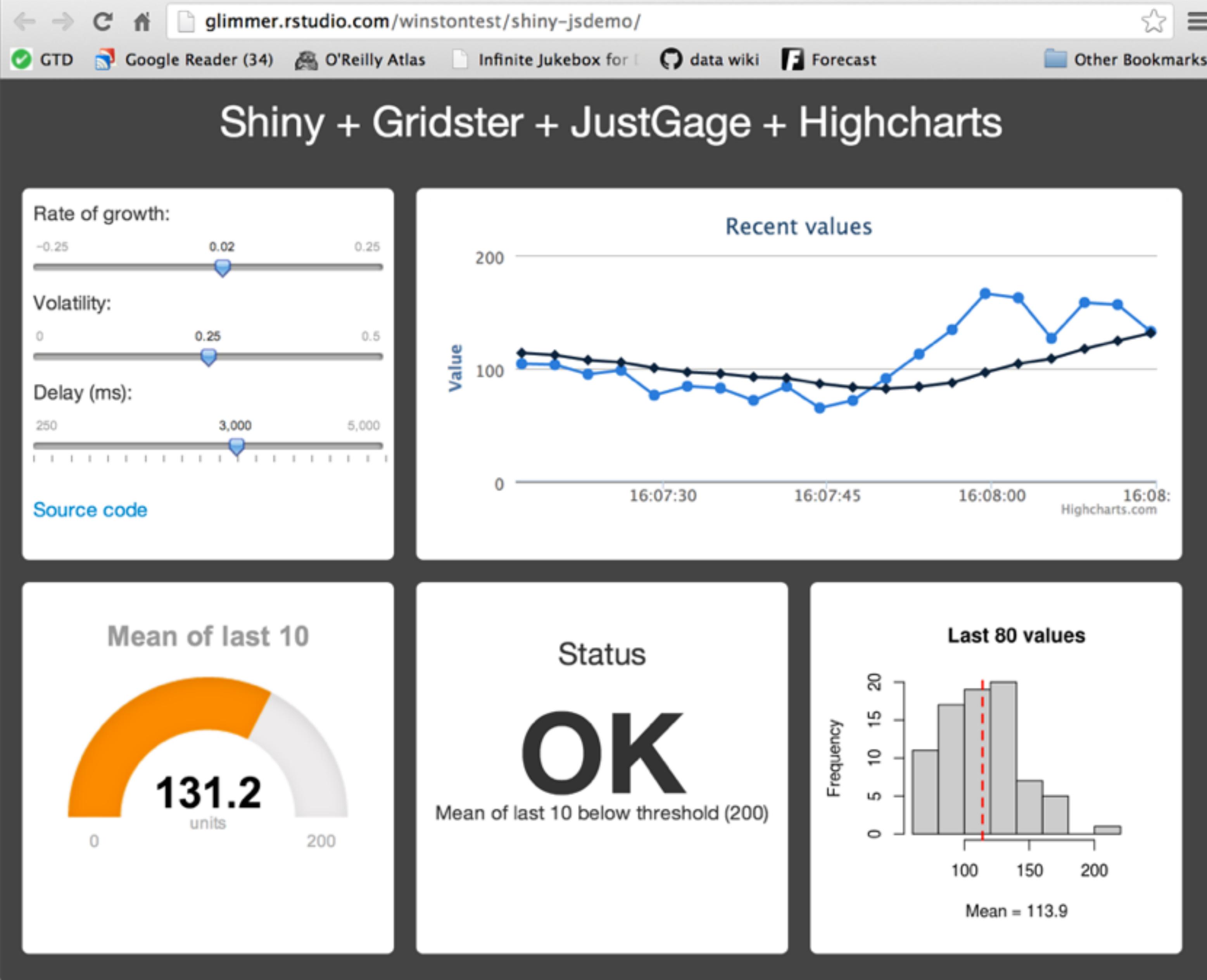
Residual standard error: 3.836 on 232 degrees of freedom

Multiple R-squared: 0.5868, Adjusted R-squared: 0.585

F-statistic: 329.5 on 1 and 232 DF, p-value: < 2.2e-16

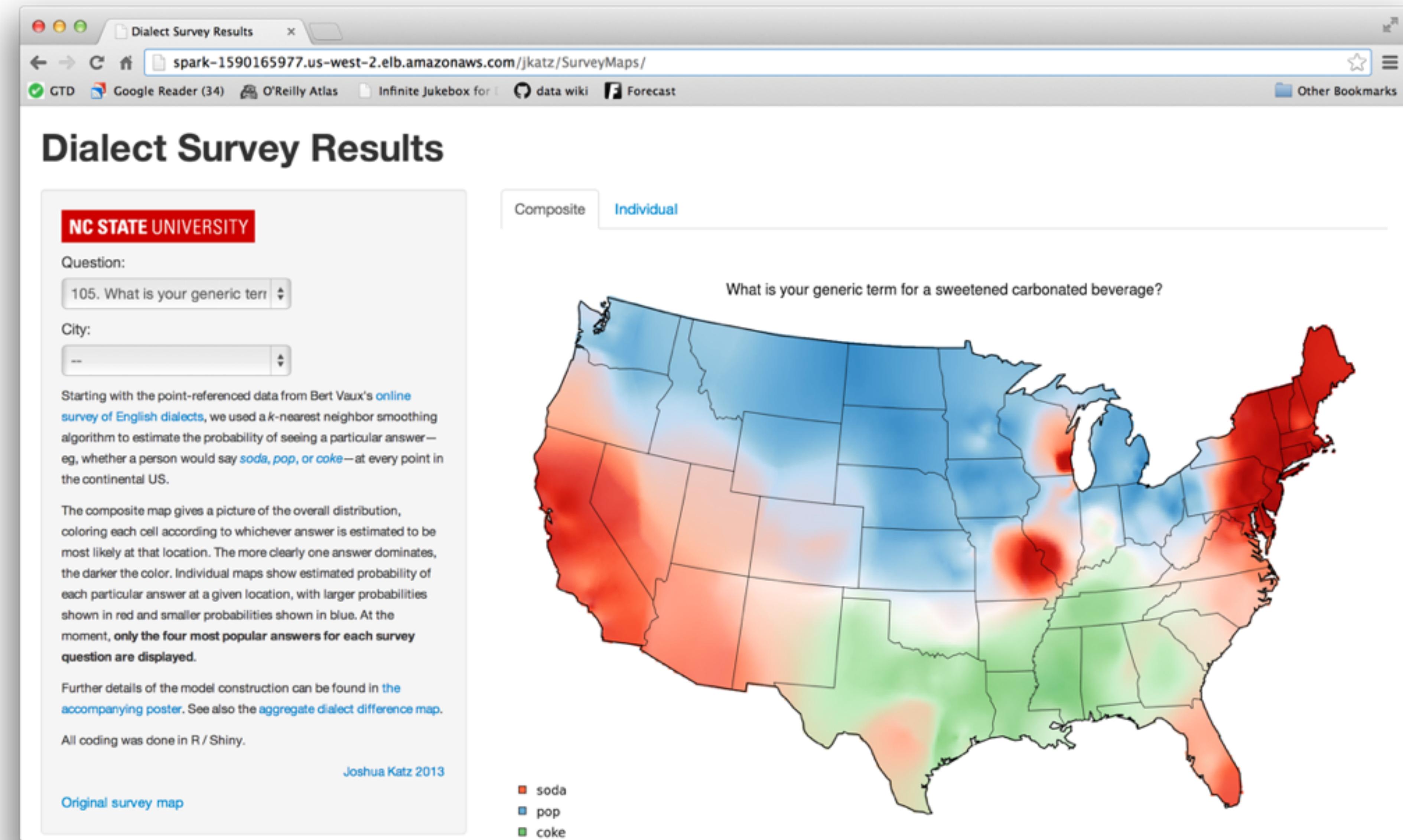


<http://glimmer.rstudio.com/winston/heightweight/>

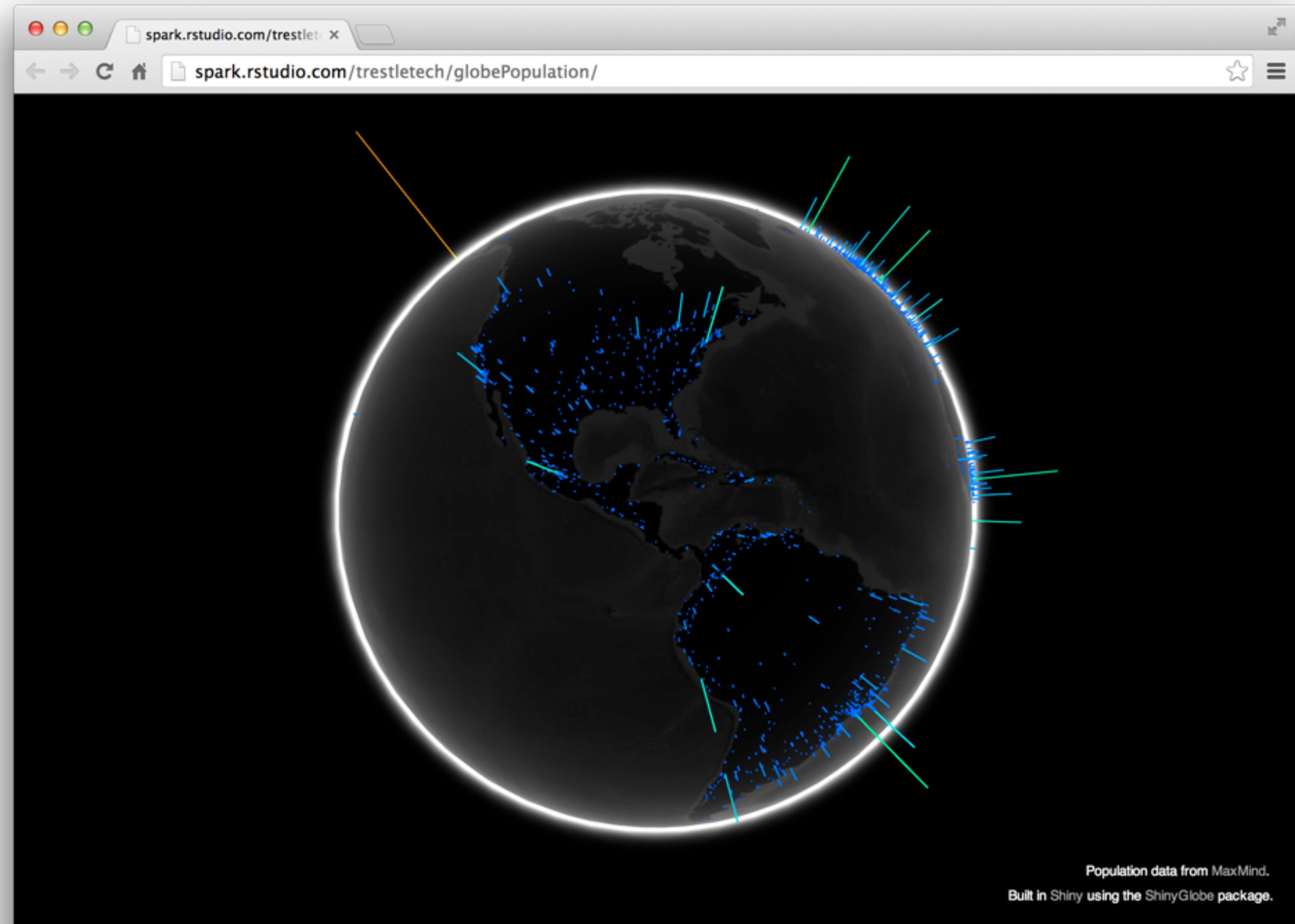


<http://glimmer.rstudio.com/winstontest/shiny-jsdemo/>

© 2014 RStudio, Inc. All rights reserved.



<http://spark-1590165977.us-west-2.elb.amazonaws.com/jkatz/SurveyMaps/>



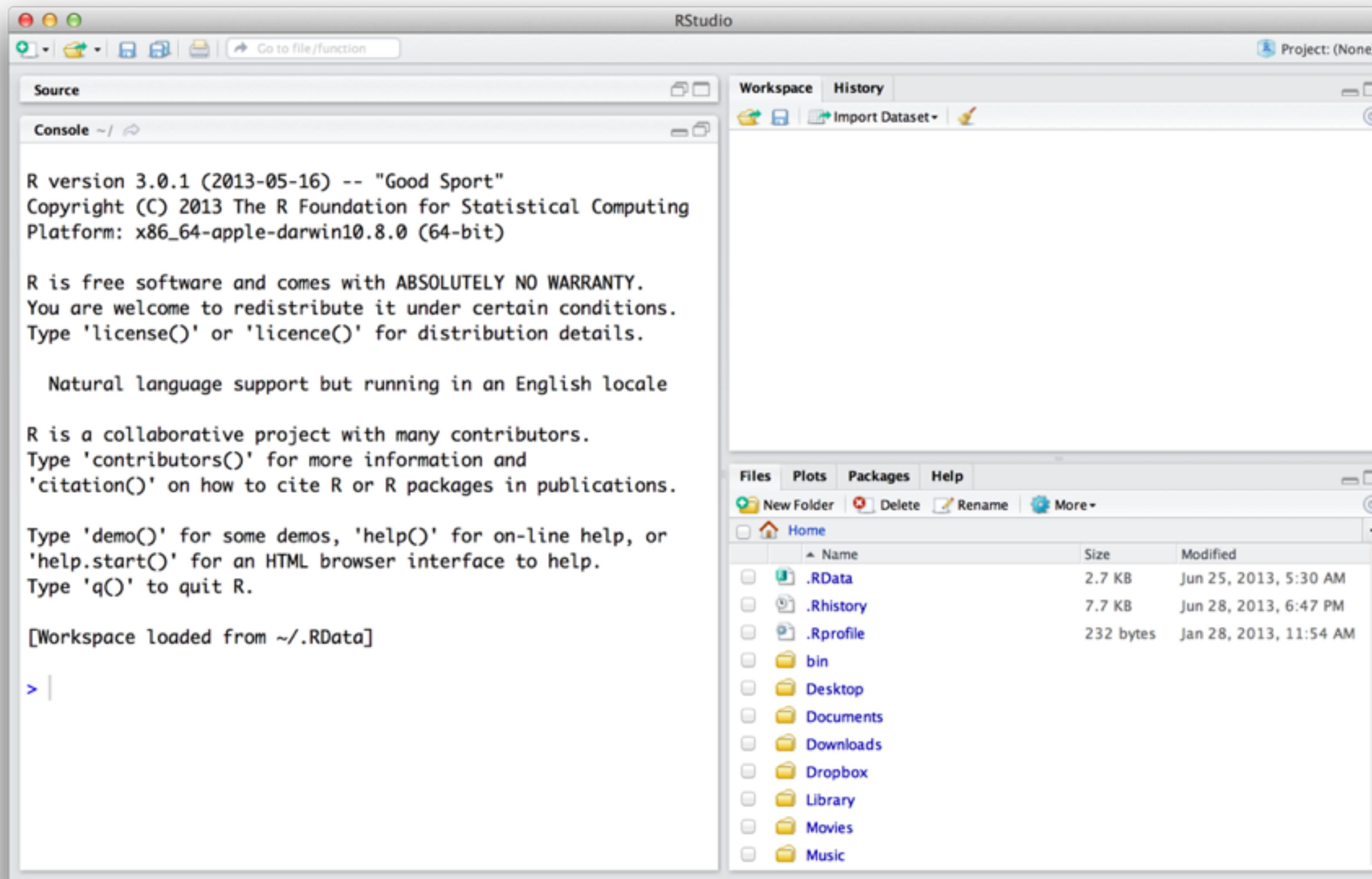
<http://spark.rstudio.com/trestletech/globePopulation/>



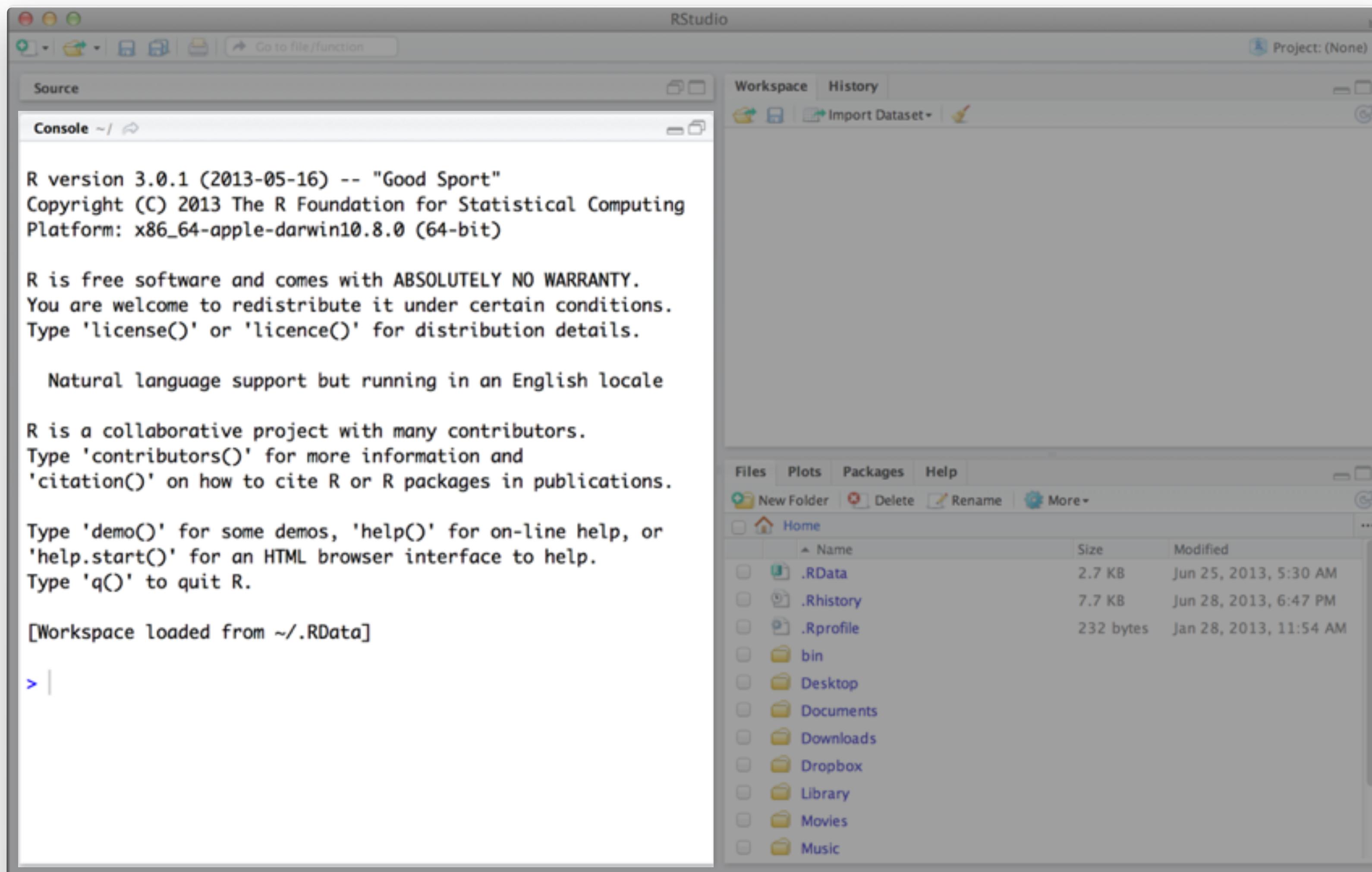
RStudio: a software program

1. like Microsoft Word, Excel, etc.
2. built to help you write R code, run R code, and analyze data with R
3. text editor, version control, latex integration, keyboard shortcuts, debugging tools, and much more

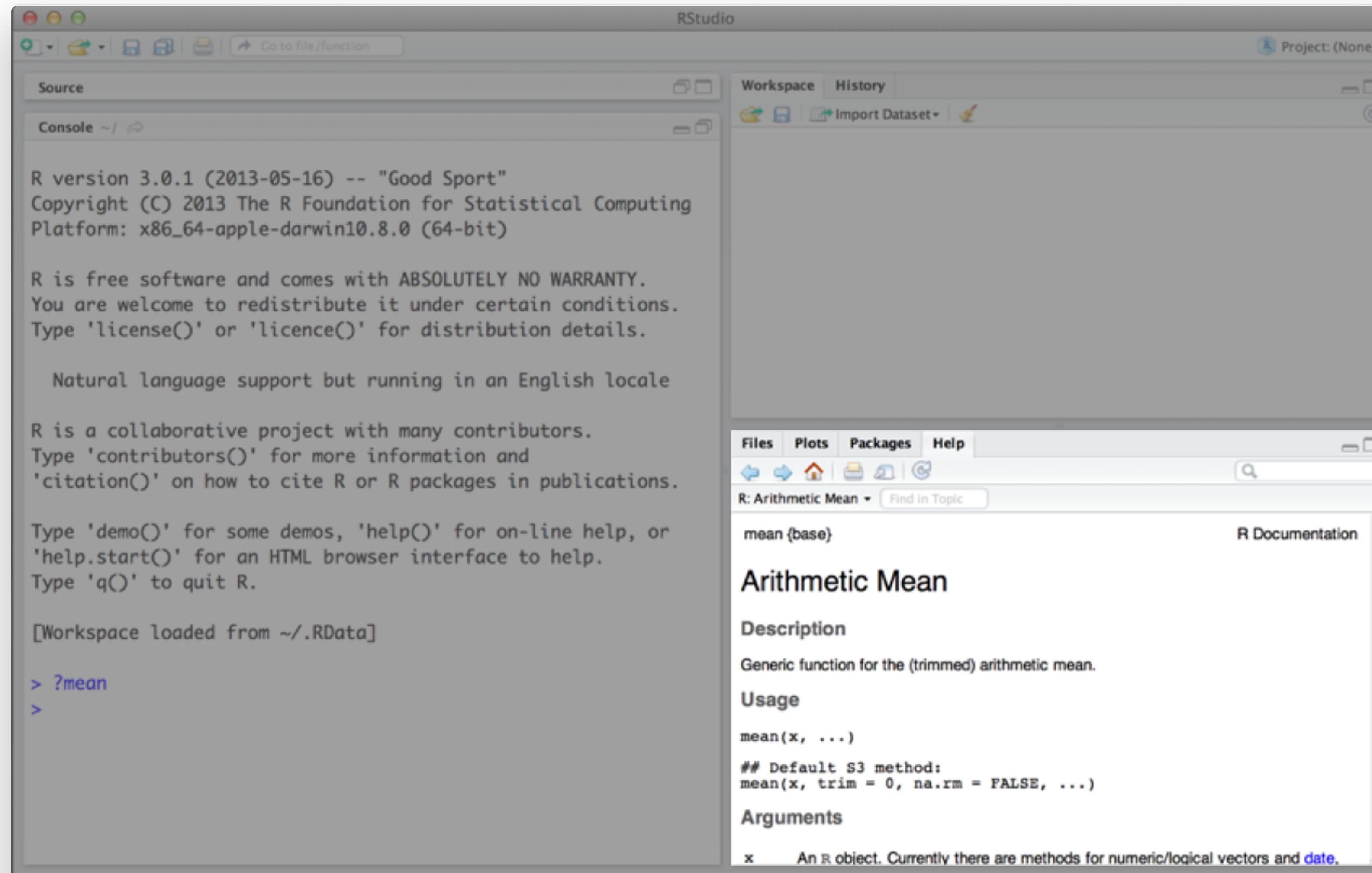
RStudio: a software program



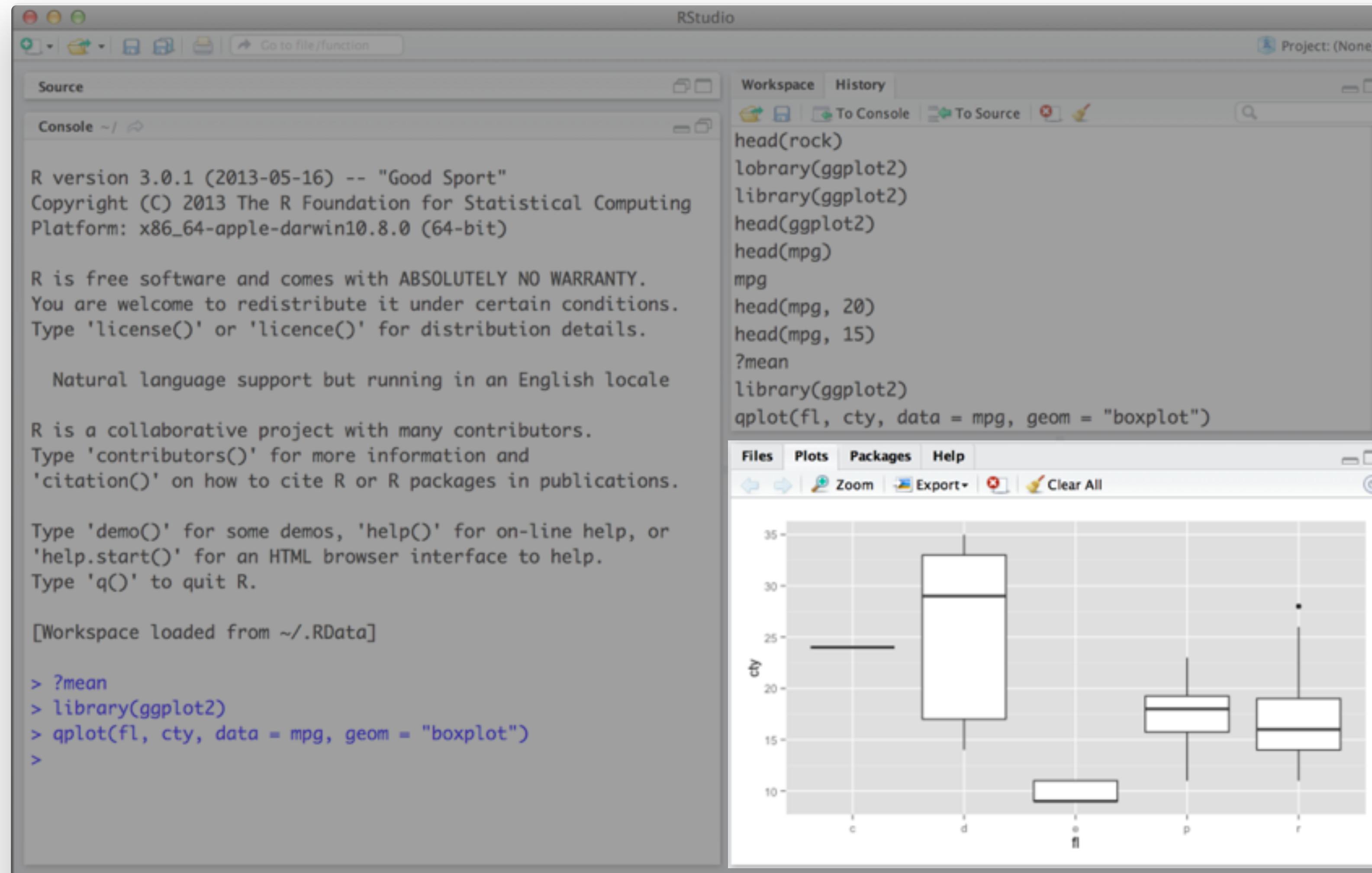
RStudio: a software program



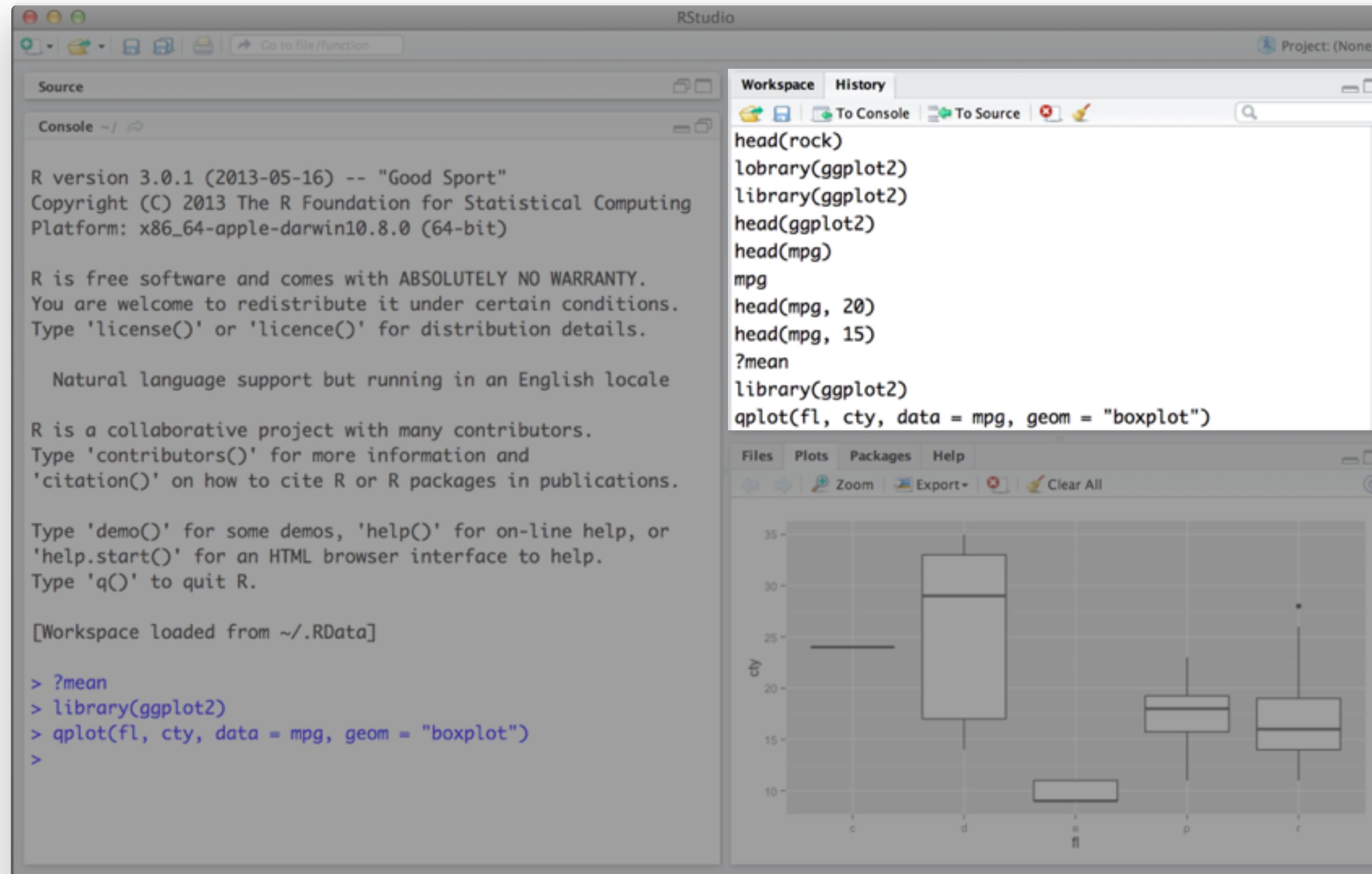
RStudio: a software program



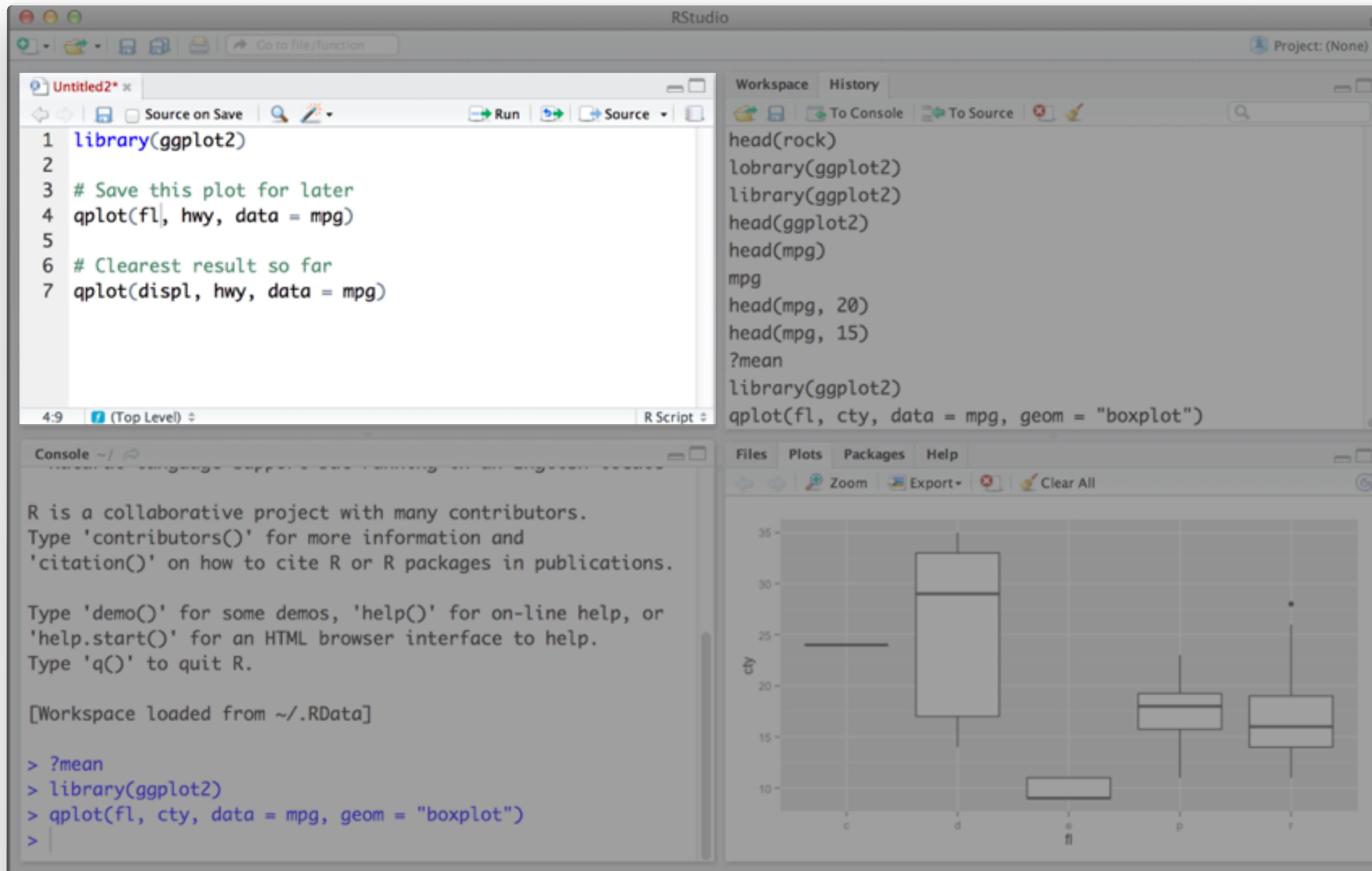
RStudio: a software program



RStudio: a software program



RStudio: a software program



RStudio: a software program

The screenshot displays the RStudio interface with the following components:

- Top Bar:** Shows the RStudio logo and the text "Project: (None)".
- Left Panel (Workspace):** Contains two tabs: "Untitled2*" and "mpg". The "mpg" tab shows a data frame with 234 observations of 11 variables, including manufacturer, model, displ, year, cyl, trans, drv, cty, and hwy.
- Right Panel (History):** Displays a history of R commands entered in the console, such as library(ggplot2), head(mpg), and View(mpg).
- Bottom Left Panel (Console):** Shows the R startup message and the command history from the session.
- Bottom Right Panel (Plots):** Displays a boxplot comparing city fuel economy (cty) across different vehicle types (c, d, p, f).

Tips

Ask yourself questions.

Work with frustration.

Practice!

And practice consciously: make a prediction, then test it, then reflect.

warm ups

Your turn

Which of these are numbers? Which are words? How can you tell?

1

"1"

"one"

Your turn

How many dimensions does a vector have?

- 1
- 2
- 3
- 4
- 5
- 6

Your turn

How many dimensions does a vector have?



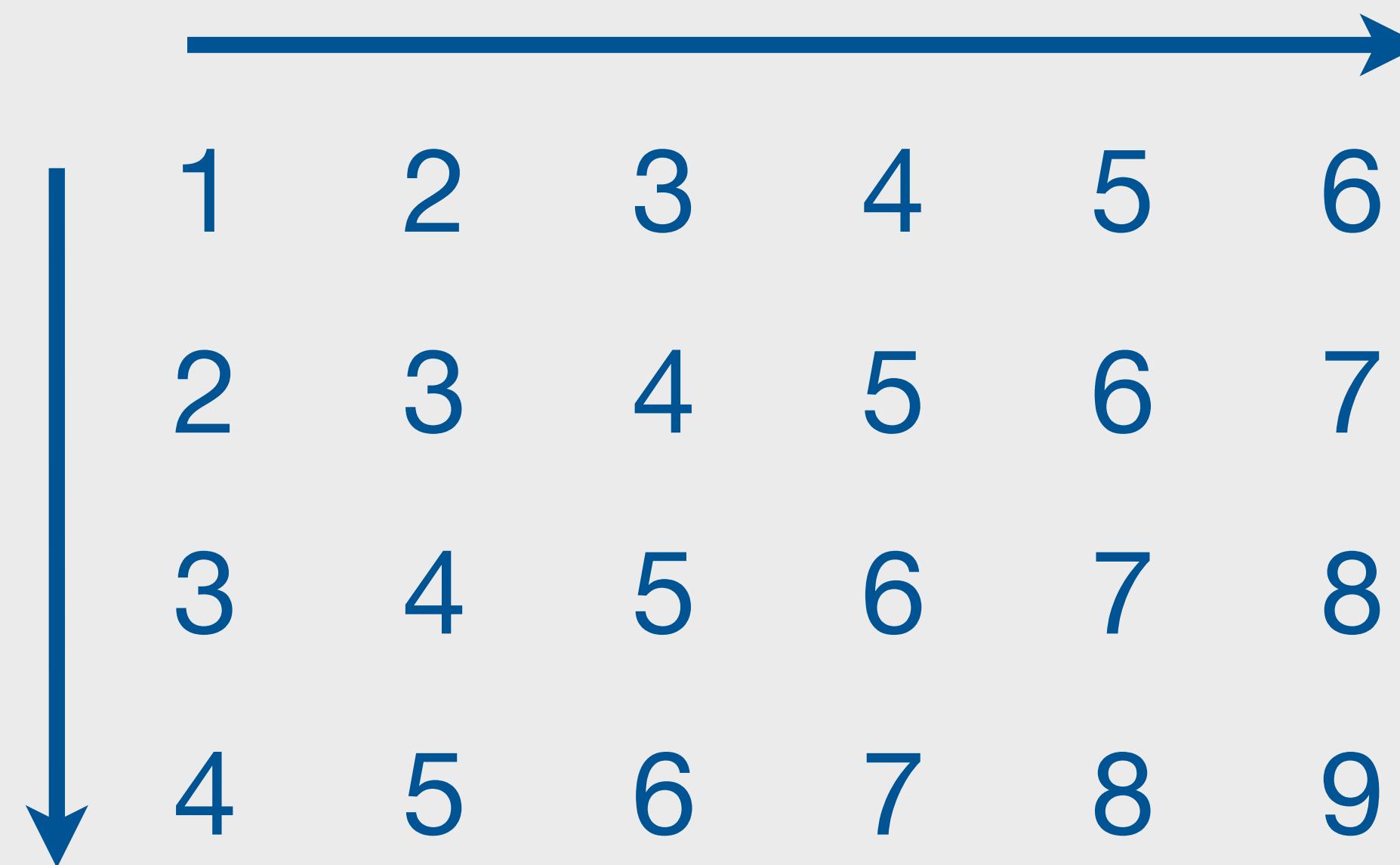
Your turn

How many dimensions does a matrix have?

1	2	3	4	5	6
2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9

Your turn

How many dimensions does a matrix have?



1	2	3	4	5	6
2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9

Your turn

The matrix below is named **M**. What is the value of **M₃₄**?

0	1	2	3	4	5
6	7	8	9	10	11
12	13	14	15	16	17
18	19	20	21	22	23

Your turn

The matrix below is named **M**. What is the value of **M₃₄**?

	0	1	2	3	4	5
6	7	8	9	10	11	
12	13	14	15	16	17	
18	19	20	21	22	23	



Your turn

The matrix below is named **M**. What is the value of **M₃₄**?

	0	1	2	3	4	5
	6	7	8	9	10	11
↓	12	13	14	15	16	17
	18	19	20	21	22	23

Your turn

The matrix below is named **M**. What is the value of **M₃₄**?

0	1	2	3	4	5
6	7	8	9	10	11
12	13	14	15	16	17
18	19	20	21	22	23