R-tistic

Introduction to R and RStudio

Lars Schoebitz

r-tistic@lse.de

March 16, 2017

R-tistic

Overview

1. Introduction to R and RStudio
    - basics of data visualisation
    - basics of reproducible research
2. Basics of data manipulation
3. Basics of using R for statistical analyses
4. Basics of Git, GitHub and collaborative programming
5. Advanced classes of the above

Why use R?

- R is able to perform every type data analyses.
- R is free and open source.
- R is a language and is interactive.
- R produces amazing graphics.
- R has a fast growing user network.
- ...and many more reasons.

Why use R?

Data Analyses

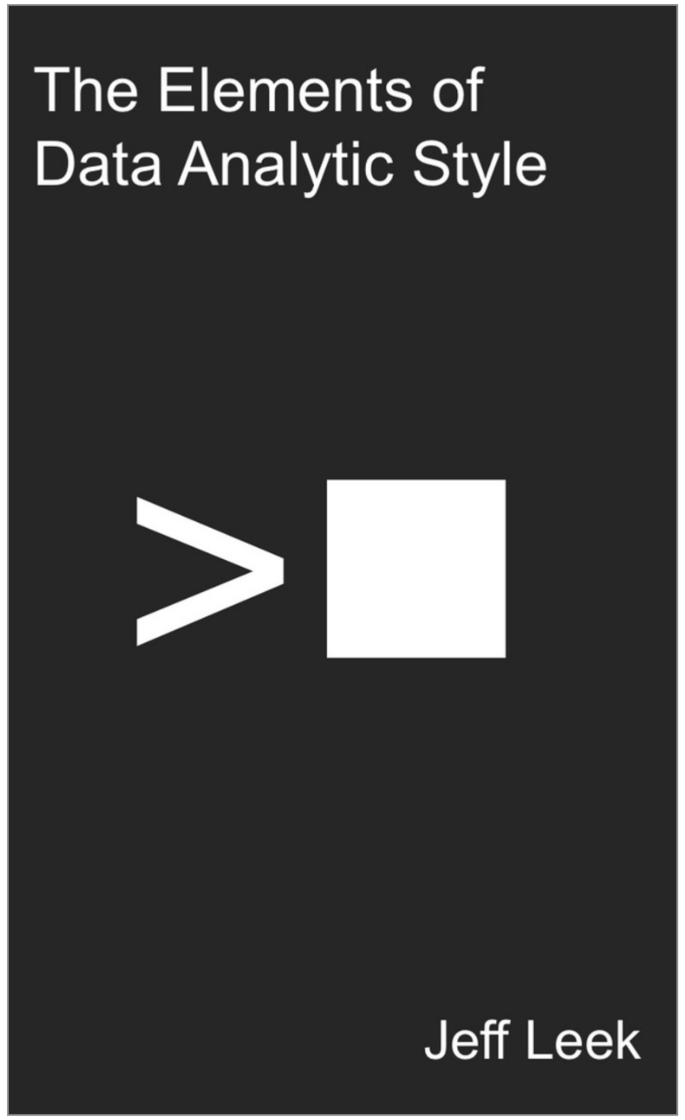
- why do we analyse data?
  - to summarise data in tables and graphs.
  - to explore relationships in data by using statistical analysis.
  - to visualise data for publications.
  - to get answers for our questions.

Why use R?

Data Analyses

- what types of data analyses do you know?
  - descriptive
  - exploratory
  - inferential
  - predictive
  - causal
  - mechanistic
- nothing that's not possible in R

- read...



Why use R?

Data Visualisation

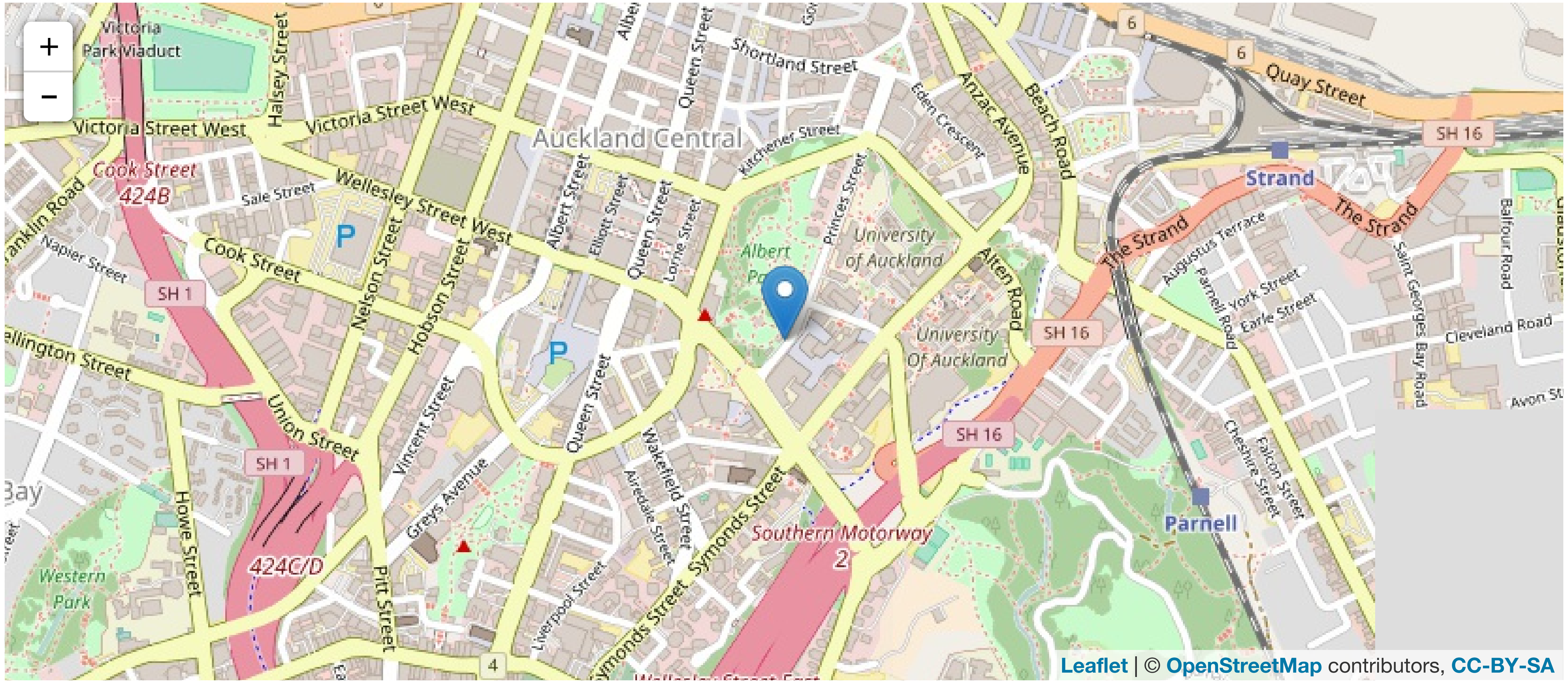
Talk:
Stefanie Posavec on Data Visualization at Awwwards Conference London

Great resource:
Data visualisation catalogue

Another great resource:
R Graph Gallery

Why use R?

Maps with leaflet package.



Leaflet | © OpenStreetMap contributors, CC-BY-SA

Click here to learn more!

Why use R?

Interactive plots with plotly package

Learn more about plotly

Why use R?

Interactive plots with plotly package

Why use R?

More than 10'000 other packages

Nice curated list of R packages here:
Awesome R

Why use R?

Because R Markdown!

- fully reproducible documents in all formats
    - HTML
    - PDF
    - MS Word
    - HTML5 slides
    - Tufte-style handouts
    - books
    - dashboards
    - shiny applications
    - scientific articles
    - websites
    - ...

Click here to learn more

R Markdown

Install R Markdown

- type `install.packages("rmarkdown")` into the Console and hit ↵

R Markdown

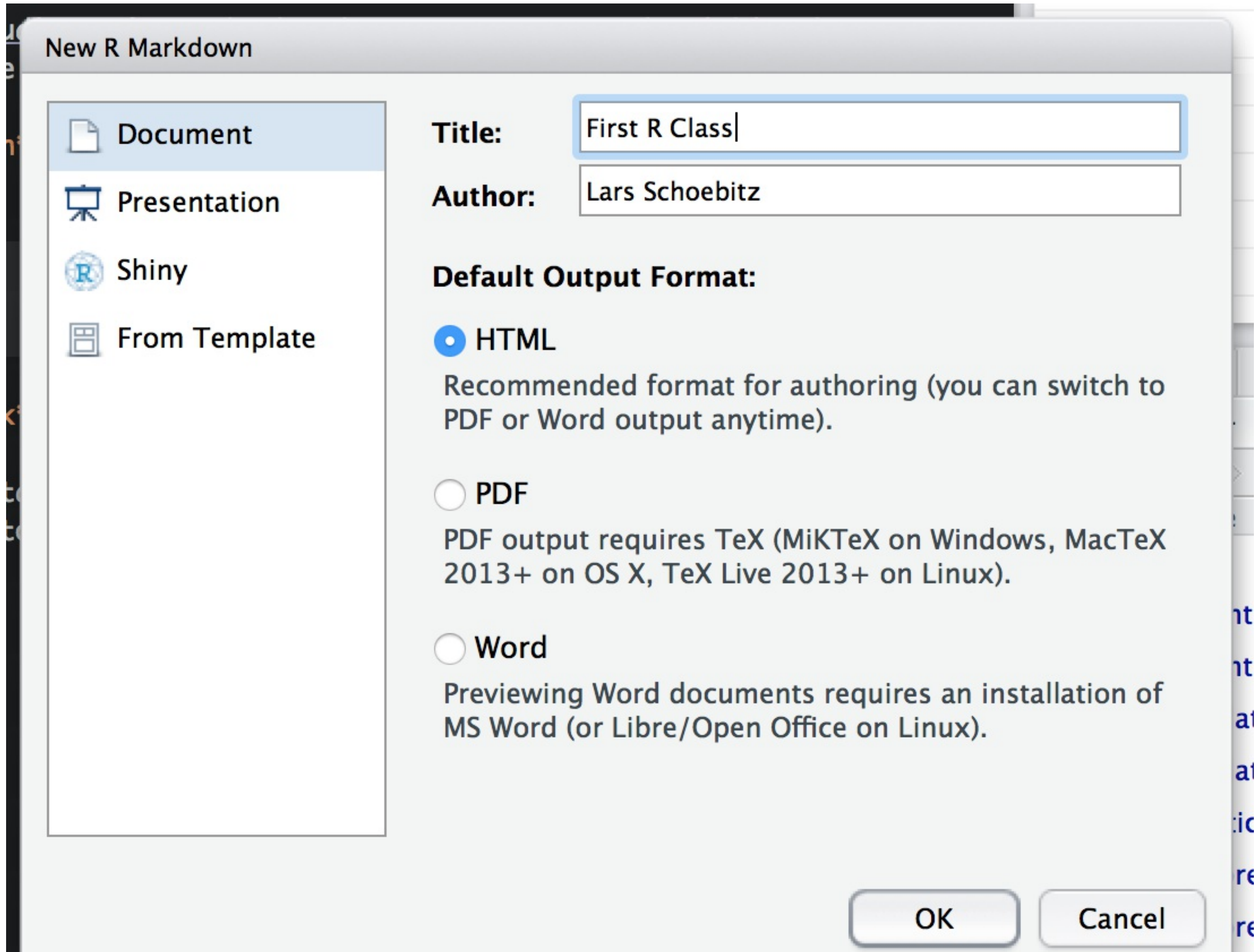
Install R Markdown

- type `install.packages("rmarkdown")` into the Console and hit ↵
- go to `File -> New File -> R Markdown...`

R Markdown

Install R Markdown

- type `install.packages("rmarkdown")` into the Console and hit ↵
- go to `File -> New File -> R Markdown...`



Title: First R Class

Author: Lars Schoebitz

Default Output Format:

● HTML

Recommended format for authoring (you can switch to PDF or Word output anytime).

○ PDF

PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

○ Word

Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

RMarkdown

Live Tutorial

...

Why use R?

Open...

- open access
- open data
- open research
- open science
- open government

Why use R?

Open...

- open access
- open data
- open research
- open science
- open government
- open university
- open learning
- open education
- open source
- open everywhere

Open Data

Everywhere

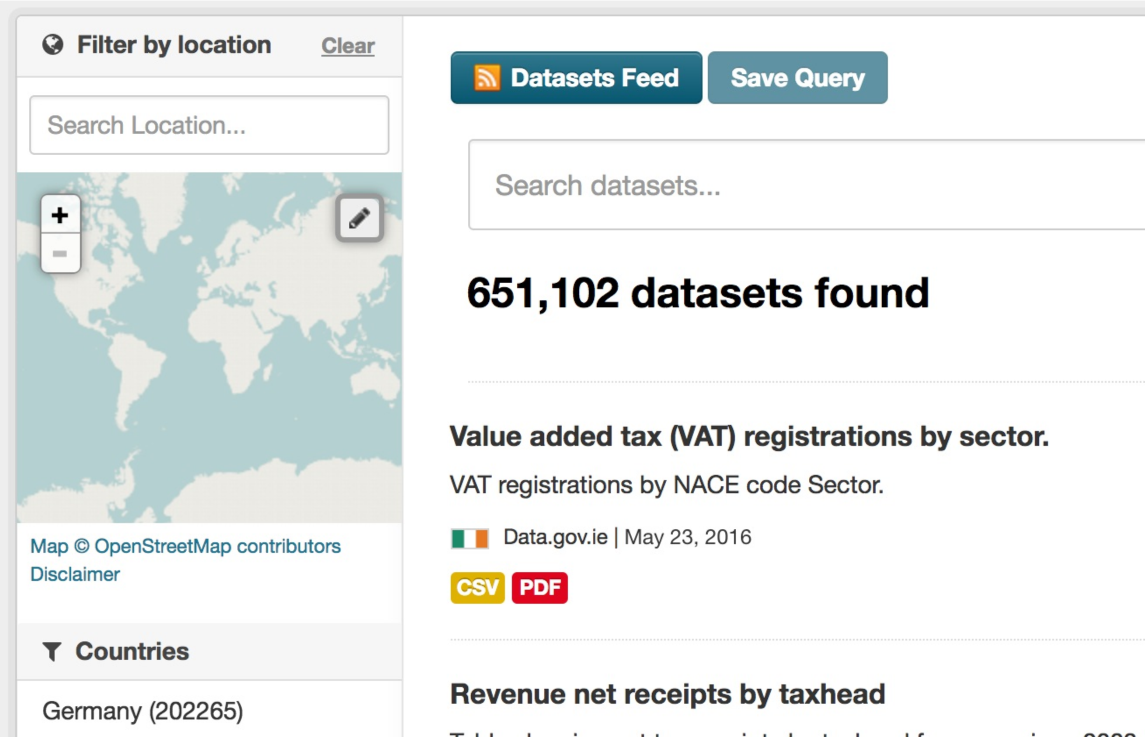
- European Data Portal
- Open Data Zurich
- World Bank DataBank
- World Bank Microdata Library
- rOpenSci Community
- Stats South Africa
- ArcGis opendata
- Open Knowledge International
- Gapminder

Open Data Twitter List

Open Data

Everywhere

- European Data Portal
- Open Data Zurich
- World Bank DataBank
- World Bank Microdata Library
- rOpenSci Community
- Stats South Africa
- ArcGis opendata
- Open Knowledge International
- Gapminder



Germany (202265)

651,102 datasets found

Value added tax (VAT) registrations by sector.

VAT registrations by NACE code Sector.

Data.gov.ie | May 23, 2016

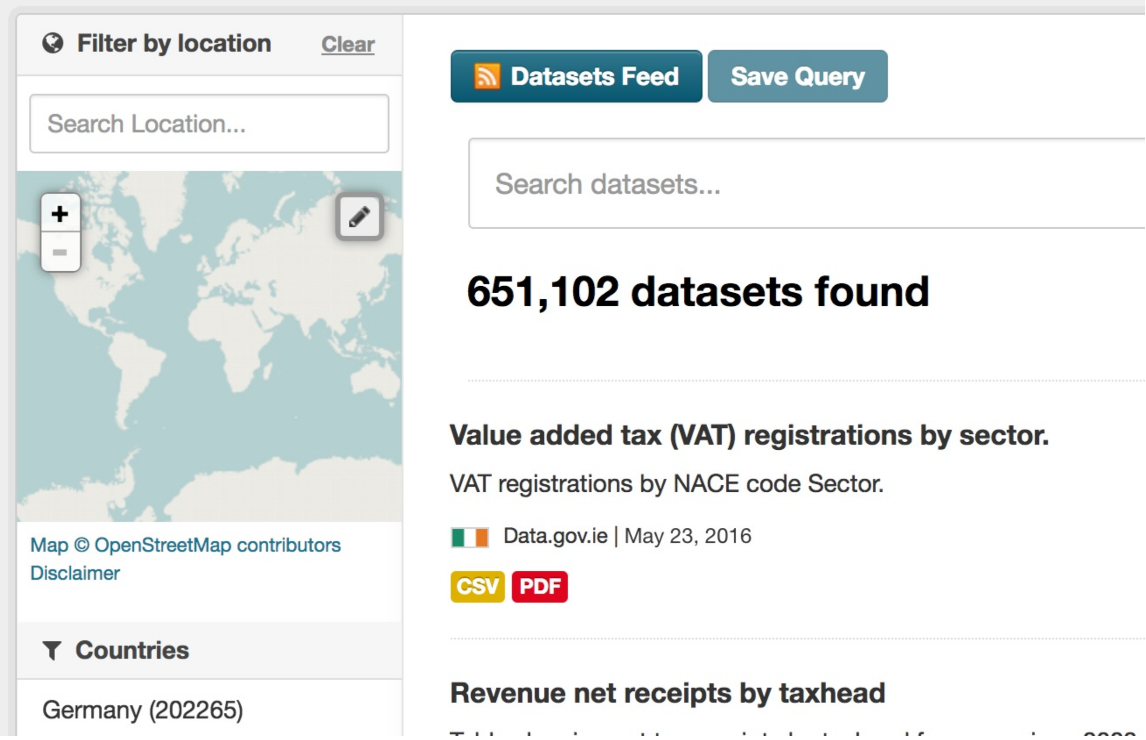
Revenue net receipts by taxhead

Open Data Twitter List

Open Data

Everywhere

- European Data Portal
- Open Data Zurich
- World Bank DataBank
- World Bank Microdata Library
- rOpenSci Community
- Stats South Africa
- ArcGis opendata
- Open Knowledge International
- Gapminder



Map © OpenStreetMap contributors
Disclaimer

Germany (202265)

651,102 datasets found

Value added tax (VAT) registrations by sector.

VAT registrations by NACE code Sector.

Data.gov.ie | May 23, 2016

Revenue net receipts by taxhead

But, what to do with all of this?

Open Data Twitter List

Data

Fuel Economy Data

- type `?mpg` into the Console

Data

Fuel Economy Data

- type `?mpg` into the Console

What is this dataset about?

```
## # A tibble: 6 <U+00D7> 11
##   manufacturer model displ  year   cyl      trans   drv
##          <chr> <chr> <dbl> <int> <int>      <chr> <chr>
## 1         audi    a4   1.8  1999     4   auto(l5)     f
## 2         audi    a4   1.8  1999     4 manual(m5)     f
## 3         audi    a4   2.0  2008     4 manual(m6)     f
## 4         audi    a4   2.0  2008     4   auto(av)     f
## 5         audi    a4   2.8  1999     6   auto(l5)     f
## 6         audi    a4   2.8  1999     6 manual(m5)     f
## # ... with 1 more variables: class <chr>
```

Data

Gapminder

- type `?gapminder` into the Console

Data

Gapminder

- type `?gapminder` into the Console
- type `??gapminder` into the Console

Data

Gapminder

- type `?gapminder` into the Console
- type `??gapminder` into the Console
- use `Google`

Data

Gapminder

- type `?gapminder` into the Console
- type `??gapminder` into the Console
- use `Google`
- type `install.packages("gapminder")` into the Console

Data

Gapminder

- type `?gapminder` into the Console
- type `??gapminder` into the Console
- use `Google`
- type `install.packages("gapminder")` into the Console
- type `?gapminder` into the Console

Gapminder Homepage. Please visit!

Data

Gapminder

- type `library(gapminder)` into the Console
- type `str(gapminder)` into the Console

Data

Gapminder

- type `library(gapminder)` into the Console
- type `str(gapminder)` into the Console

```
library(gapminder)

str(gapminder)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    1704 obs. o
##  $ country  : Factor w/ 142 levels "Afghanistan",..: 1 1
##  $ continent: Factor w/ 5 levels "Africa","Americas",..:
##  $ year     : int  1952 1957 1962 1967 1972 1977 1982 19
##  $ lifeExp  : num  28.8 30.3 32 34 36.1 ...
##  $ pop      : int  8425333 9240934 10267083 11537966 130
##  $ gdpPercap: num  779 821 853 836 740 ...
```

Data

Diamonds

- type `?diamonds` into the Console

Data

Diamonds

- type `?diamonds` into the Console
- what is the dataset about?

Data

Diamonds

- type `?diamonds` into the Console
- what is the dataset about?
- type **diamonds** into the Console

Data

Diamonds

- type `?diamonds` into the Console
- what is the dataset about?
- type **diamonds** into the Console

```
diamonds
```

```
## # A tibble: 53,940 <U+00D7> 10
##    carat       cut color clarity depth table price     x
##    <dbl>     <ord> <ord>   <ord> <dbl> <dbl> <int> <dbl>
## 1   0.23     Ideal     E     SI2  61.5    55   326  3.95
## 2   0.21   Premium     E     SI1  59.8    61   326  3.89
## 3   0.23      Good     E     VS1  56.9    65   327  4.05
## 4   0.29   Premium     I     VS2  62.4    58   334  4.20
## 5   0.31      Good     J     SI2  63.3    58   335  4.34
## 6   0.24 Very Good     J    VVS2  62.8    57   336  3.94
## 7   0.24 Very Good     I    VVS1  62.3    57   336  3.95
## 8   0.26 Very Good     H     SI1  61.9    55   337  4.07
```

R and RStudio

First steps with Software Carpentry

Please go to the following website and follow through the instructions until you reach the challenges at the bottom of the page.

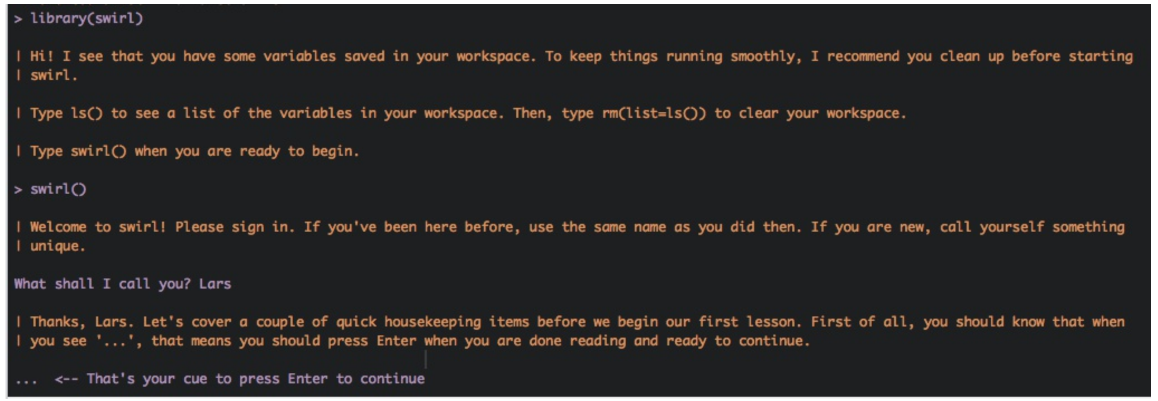
Software Carpentry

R and RStudio

First steps with swirl package

Learn R, in R.

- open the **swirl website**
- click on learn
    - Step 1: Done that
    - Step 2: Done that too
    - Step 3: Install swirl
    - Step 4: Start swirl
    - Step 5: ...



```
> library(swirl)

| Hi! I see that you have some variables saved in your workspace. To keep things running smoothly, I recommend you clean up before starting
| swirl.

| Type ls() to see a list of the variables in your workspace. Then, type rm(list=ls()) to clear your workspace.

| Type swirl() when you are ready to begin.

> swirl()

| Welcome to swirl! Please sign in. If you've been here before, use the same name as you did then. If you are new, call yourself something
| unique.

What shall I call you? Lars

| Thanks, Lars. Let's cover a couple of quick housekeeping items before we begin our first lesson. First of all, you should know that when
| you see '...', that means you should press Enter when you are done reading and ready to continue.

...  <-- That's your cue to press Enter to continue
```

R and RStudio

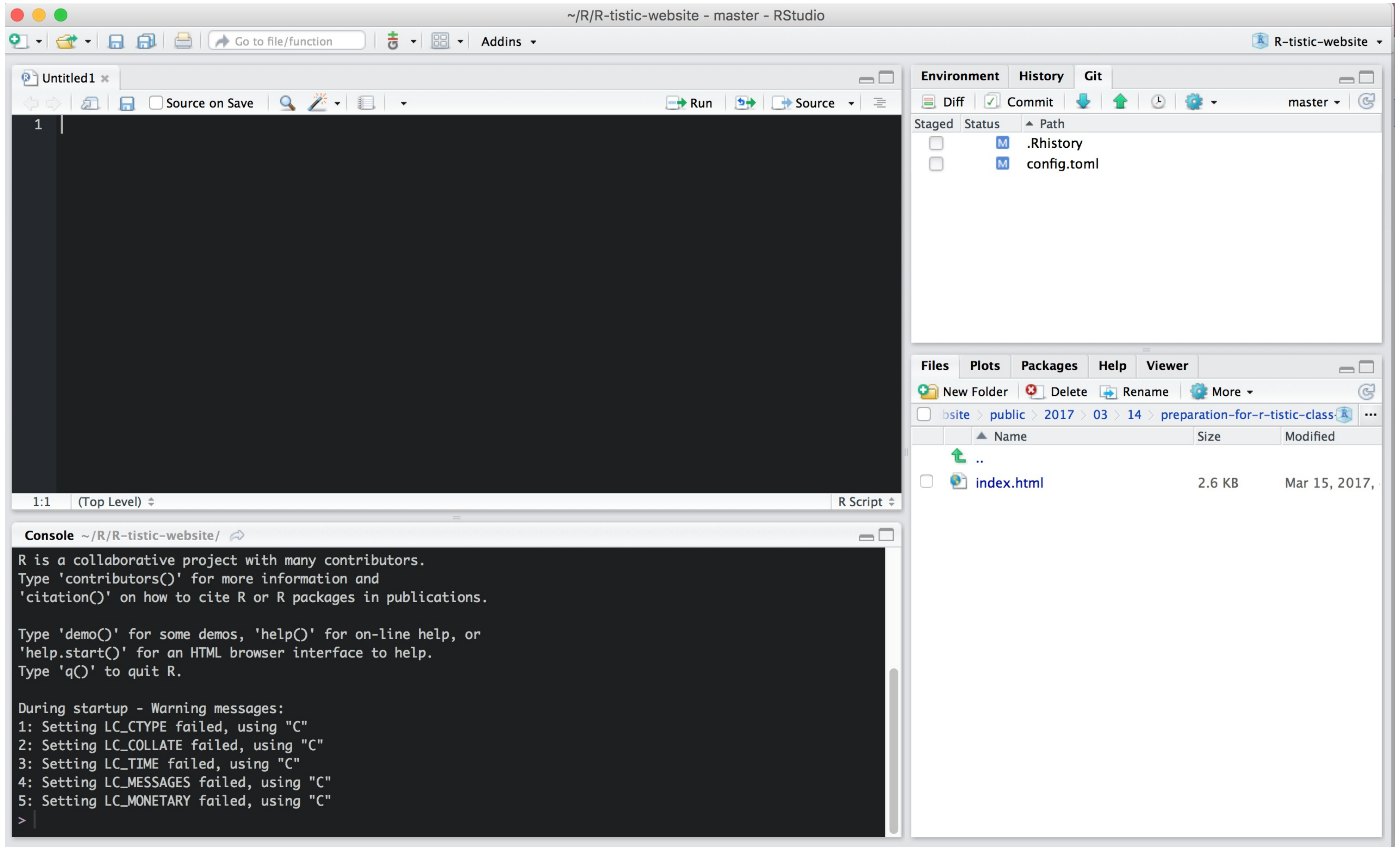
First steps with R Codeschool

Great way to learn R:
http://tryr.codeschool.com

RStudio

Start your first script

- go to `File -> New File -> R Script...`



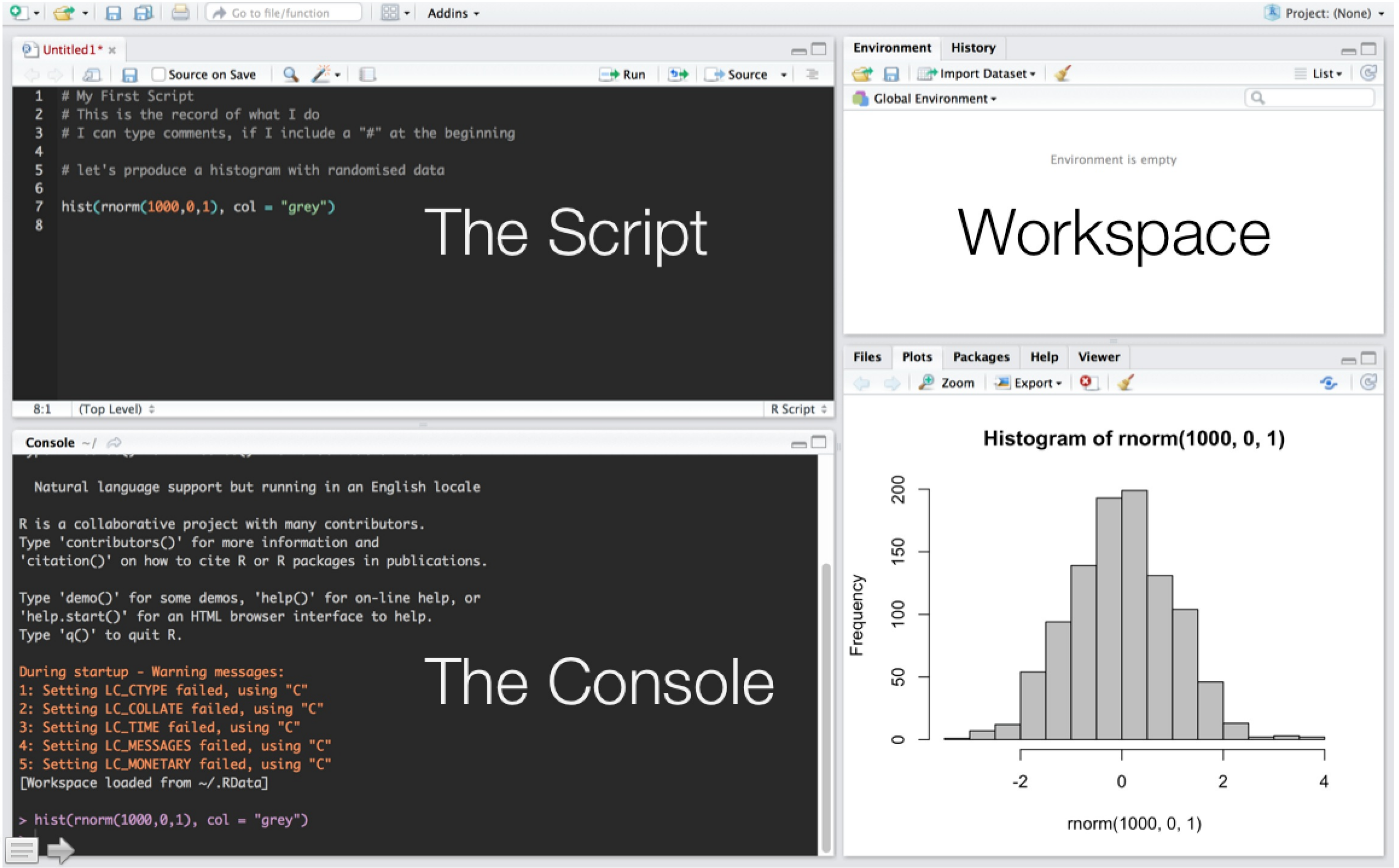
Staged	Status	Path
□	M	.Rhistory
□	M	config.toml

Name	Size	Modified
..		
index.html	2.6 KB	Mar 15, 2017,

```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
>
```

RStudio

Overview

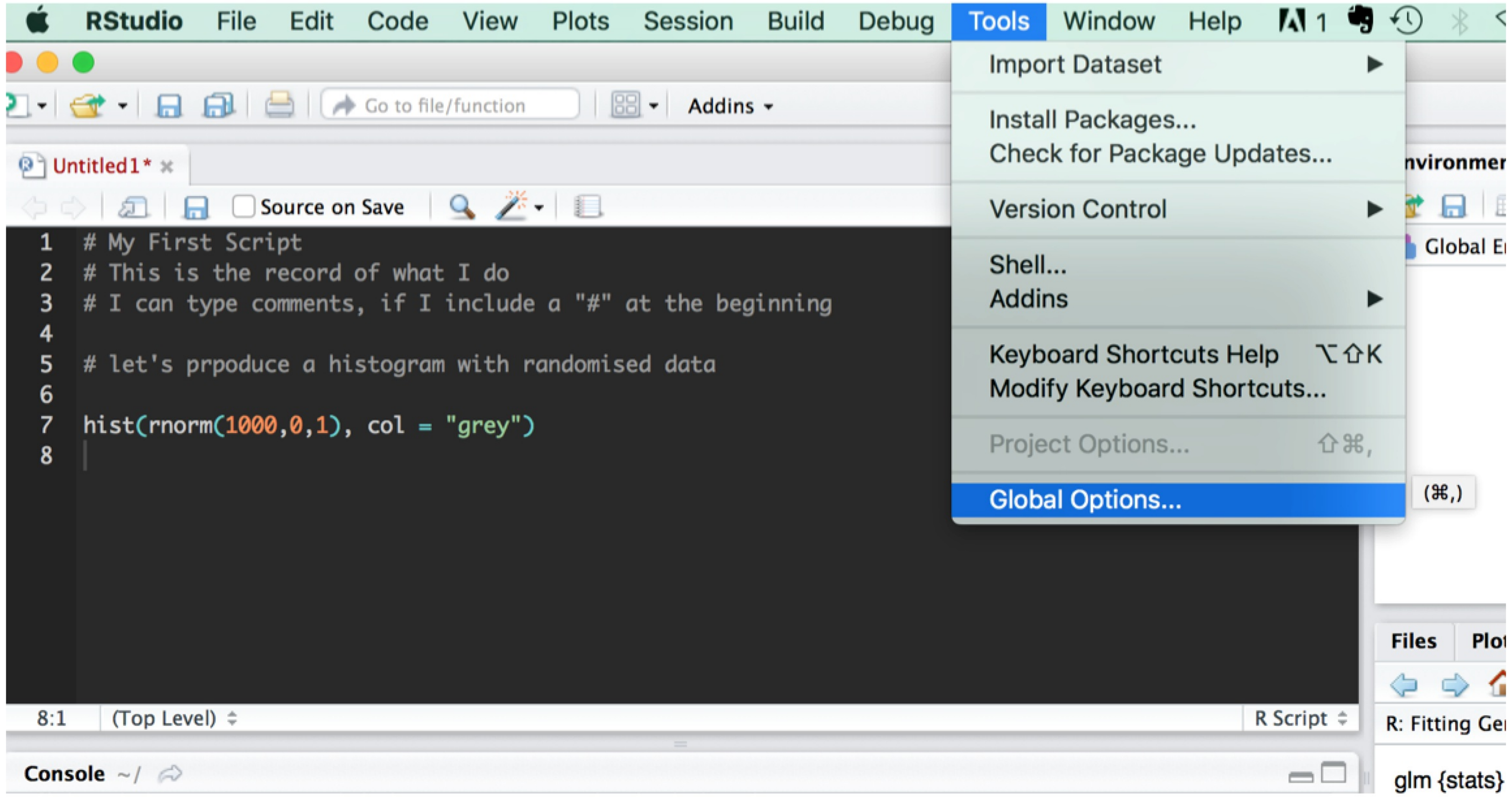


```
# My First Script
# This is the record of what I do
# I can type comments, if I include a "#" at the beginning

# let's prpoduce a histogram with randomised data

hist(rnorm(1000,0,1), col = "grey")
```

```
  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
5: Setting LC_MONETARY failed, using "C"
[Workspace loaded from ~/.RData]

> hist(rnorm(1000,0,1), col = "grey")
```

RStudio

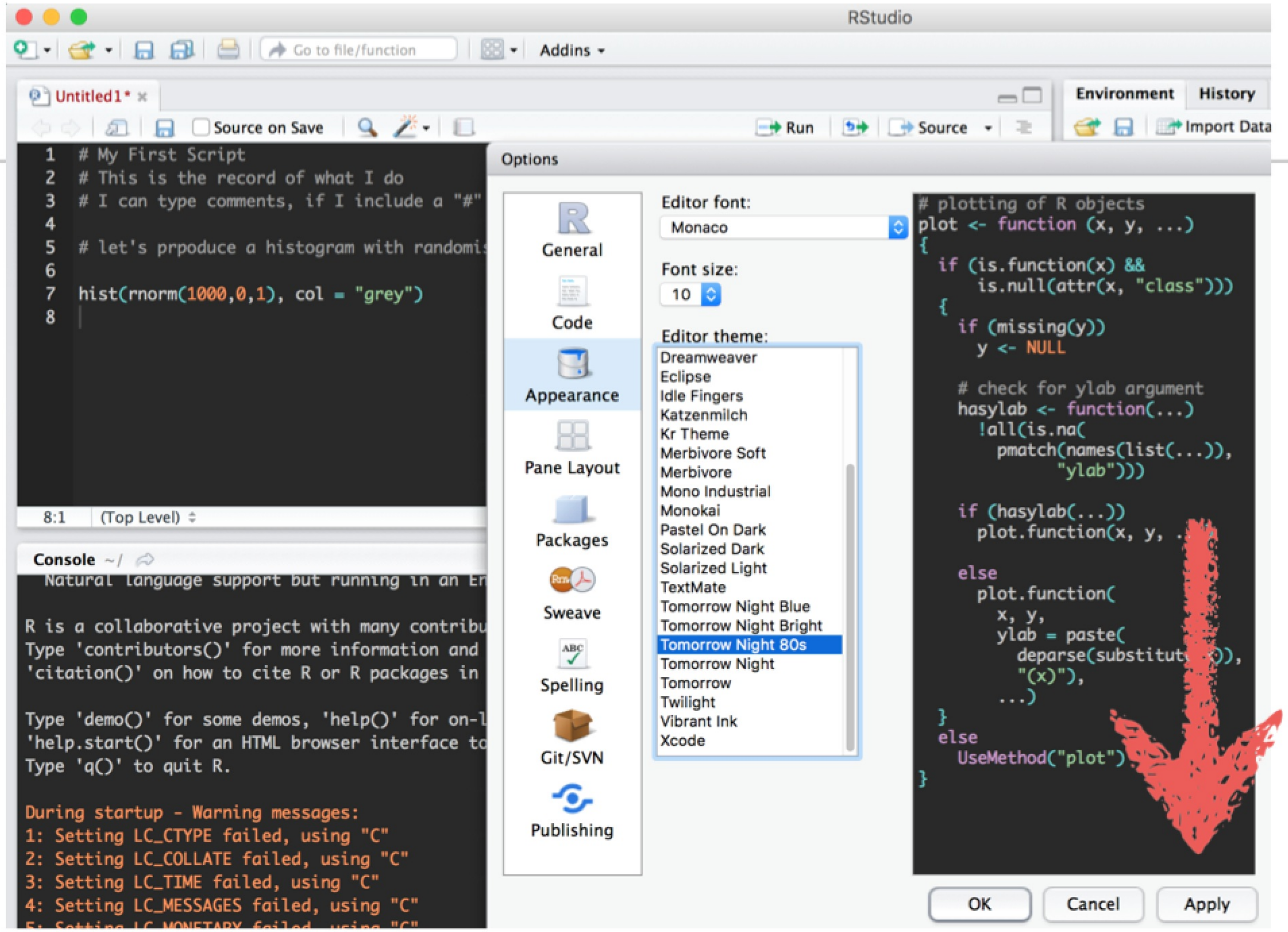
Themes



```
# My First Script
# This is the record of what I do
# I can type comments, if I include a "#" at the beginning

# let's prpoduce a histogram with randomised data

hist(rnorm(1000,0,1), col = "grey")

```

RStudio

Themes



```
# My First Script
# This is the record of what I do
# I can type comments, if I include a "#"

# let's prpoduce a histogram with randomi

hist(rnorm(1000,0,1), col = "grey")
```

```
  Natural language support but running in an En

R is a collaborative project with many contribu
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in

Type 'demo()' for some demos, 'help()' for on-l
'help.start()' for an HTML browser interface to
Type 'q()' to quit R.

During startup - Warning messages:
1: Setting LC_CTYPE failed, using "C"
2: Setting LC_COLLATE failed, using "C"
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
```

Editor font: Monaco

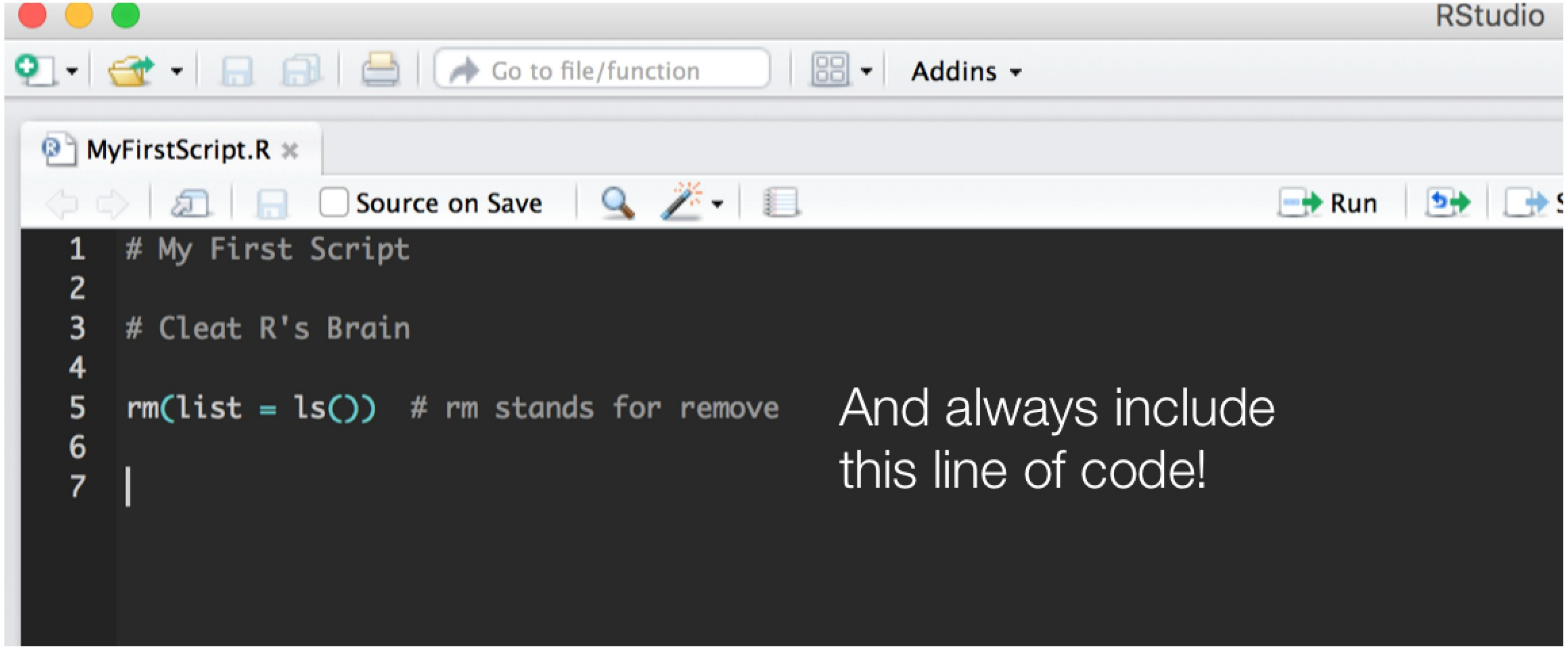
Font size: 10

Editor theme:
Dreamweaver
Eclipse
Idle Fingers
Katzenmilch
Kr Theme
Merbivore Soft
Merbivore
Mono Industrial
Monokai
Pastel On Dark
Solarized Dark
Solarized Light
TextMate
Tomorrow Night Blue
Tomorrow Night Bright
Tomorrow Night 80s
Tomorrow Night
Tomorrow
Twilight
Vibrant Ink
Xcode

```
# plotting of R objects
plot <- function (x, y, ...)
{
  if (is.function(x) &&
      is.null(attr(x, "class")))
  {
    if (missing(y))
      y <- NULL

    # check for ylab argument
    hasylab <- function(...)
      !all(is.na(
        pmatch(names(list(...)),
               "ylab")))

    if (hasylab(...))
      plot.function(x, y, .

    else
      plot.function(
        x, y,
        ylab = paste(
          deparse(substitut  )),
          "(x)"),
        ...)
  }
  else
    UseMethod("plot")
}
```

RStudio

The # key and rm(list = ls())



```
# My First Script

# Cleat R's Brain

rm(list = ls())  # rm stands for remove

```

And always include this line of code!

R Scripts

Why using scripts?

- we could do everything in the console, so why bother using a script?
- record of what you did and why (includes comments after #)
- allows you to quickly repeat the analysis and make changes
- the code in the console will not be saved, but you can save the script

R Scripts

Workflow

- type code into the Script
- add notes after # to remember what you are doing
- run the code (select it and then Ctrl + R or cmd + enter)
- if it's not working, edit the Script and run it again
- save the final Script including comments
- the script can be reused at any time

R Scripts

Write your first script

- write the preparation code into a script

R Scripts

Write your first script

- write the preparation code into a script

keep in mind:

R Scripts

Write your first script

- write the preparation code into a script

keep in mind:

- make comments
- load necessary libraries

R Scripts

Write your first script

- write the preparation code into a script

keep in mind:

- make comments
- load necessary libraries
- save your script file
- make sure you remember where and the name
- close RStudio (do not save the workspace)
- find and open your Script
- remember keyboard shortcuts:
- ctrl + A / cmd + A
- ctrl + R / cmd + enter