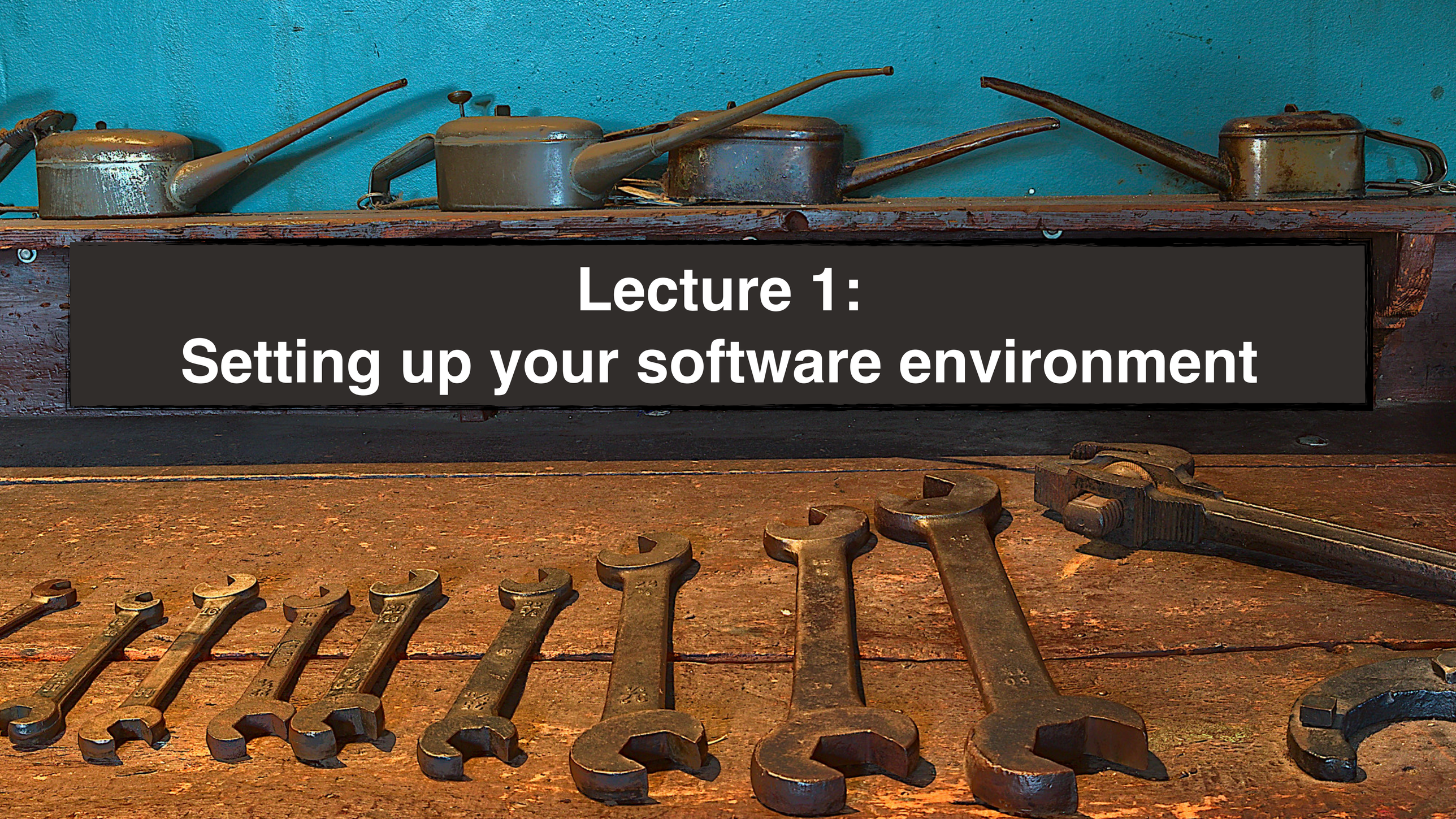
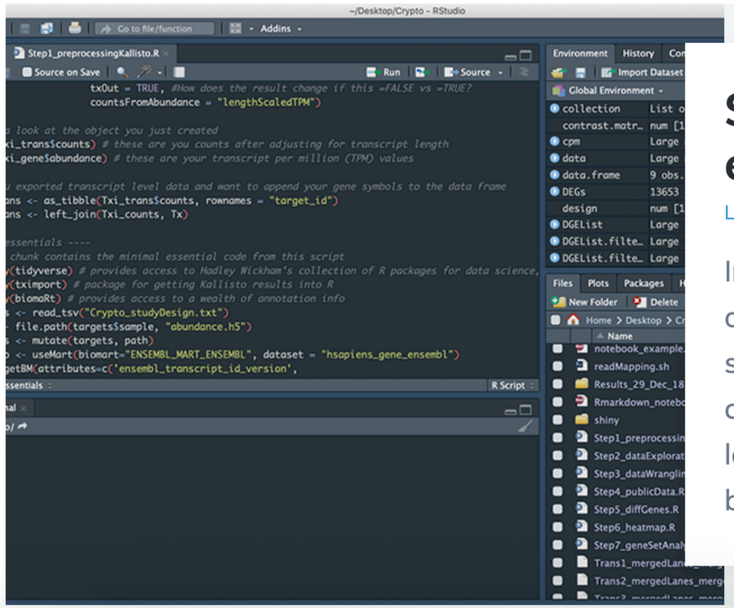


Lecture 1:
Setting up your software environment



Setting up your software environment

Lecture 1 • watch by September 1, 2021

In the first half of this lecture we'll discuss th
open-source, cross-platform R/bioconductor
software that we will use throughout the
course. Then each student will set-up their o
laptop to be a powerful, stand-alone
bioinformatics workstation.

Software installation and IT support

Lab 1 (optional) • September 1, 2021

Installing software can be a real headache, so let us help you! This lab will be focused on helping you with IT support and getting to know the software tools that we'll be using throughout the course.

Setting up your software environment

Lecture 1 • watch by September 1, 2021

Lecture slides on iCloud

Overview

In the first half of this lecture we'll discuss the open-source, cross-platform R/bioconductor software that we will use throughout the course. Then each student will set-up their own laptop to be a powerful, stand-alone bioinformatics workstation.

Learning objectives

- Brief overview of the tools we'll use throughout the course (R/Bioconductor, RStudio, etc.)
- *Your* laptop with all the software needed for the course

re

through installing the following software

ing Language - The only programming language we'll work with in class.

development environment for the R programming language.

editor - a simple but powerful text editor that is 'code aware'

o Code - an excellent choice for working with virtually any kind of code outside We'll use this later in the semester for connecting to and working with GitHub

oper tools - *Only download if running Mac OS on your laptop.*

ly download if running a Windows OS on your laptop. Cygwin will give your linux-like capabilities

s is the software we'll use for mapping raw reads to a reference transcriptome. Detailed instructions for installing and using Kallisto and other course-related software is available on my lab's protocols site here.

AutoSave OFF hmp2_metadata

Home Insert Draw Page Layout Formulas Data Review View Share Comments

Paste Calibri (Body) 12 General Conditional Formatting Format as Table Cell Styles Insert Delete Format Sort & Filter Find & Select Ideas Sensitivity

A1 Project

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Project	External ID	Participant I	site_sub_col	data_type	week_num	date_of_rec	interval_day	visit_num	Research Pro	PDO Numbe	GSSR IDs	Product	LCSET	Aggregated	WR ID	# Lanes in Ag	reads_raw	reads_filtere	reads_qc_fai
2	C3001CSC1_BP	206615	C3001	C3001CSC1	biopsy_16S	2			1	ibdmdb										
3	C3001CSC2_BP	206614	C3001	C3001CSC2	biopsy_16S	2			1	ibdmdb										
4	C3002CSC1_BP	206617	C3002	C3002CSC1	biopsy_16S	0			1	ibdmdb										
5	C3002CSC2_BP	206619	C3002	C3002CSC2	biopsy_16S	0			1	ibdmdb										
6	C3002CSC3_BP	206616	C3002	C3002CSC3	biopsy_16S	0			1	ibdmdb										
7	C3002CSC4_BP	206618	C3002	C3002CSC4	biopsy_16S	0			1	ibdmdb										
8	C3003CSC1_BP	206621	C3003	C3003CSC1	biopsy_16S	1			1	ibdmdb										
9	C3003CSC2_BP	206622	C3003	C3003CSC2	biopsy_16S	1			1	ibdmdb										
10	C3003CSC3_BP	206620	C3003	C3003CSC3	biopsy_16S	1			1	ibdmdb										
11	C3004CSC1_BP	206624	C3004	C3004CSC1	biopsy_16S	0			1	ibdmdb										
12	C3004CSC2_BP	206623	C3004	C3004CSC2	biopsy_16S	0			1	ibdmdb										
13	C3004CSC3_BP	206626	C3004	C3004CSC3	biopsy_16S	0			1	ibdmdb										
14	C3004CSC4_BP	206625	C3004	C3004CSC4	biopsy_16S	0			1	ibdmdb										
15	C3005CSC1_BP	206628	C3005	C3005CSC1	biopsy_16S	0			1	ibdmdb										
16	C3005CSC2_BP	206627	C3005	C3005CSC2	biopsy_16S	0			1	ibdmdb										
17	C3006CSC1_BP	206630	C3006	C3006CSC1	biopsy_16S	0			1	ibdmdb										
18	C3006CSC2_BP	206629	C3006	C3006CSC2	biopsy_16S	0			1	ibdmdb										
19	C3011CSC1_BP	206636	C3011	C3011CSC1	biopsy_16S	4			1	ibdmdb										
20	C3011CSC2_BP	206635	C3011	C3011CSC2	biopsy_16S	4			1	ibdmdb										
21	C3015CSC1_BP	206644	C3015	C3015CSC1	biopsy_16S	1			1	ibdmdb										
22	C3015CSC2_BP	206643	C3015	C3015CSC2	biopsy_16S	1			1	ibdmdb										
23	C3016CSC1_BP	206645	C3016	C3016CSC1	biopsy_16S	0			1	ibdmdb										
24	C3016CSC2_BP	206646	C3016	C3016CSC2	biopsy_16S	0			1	ibdmdb										
25	C3017CSC1_BP	206648	C3017	C3017CSC1	biopsy_16S	0			1	ibdmdb										
26	C3017CSC2_BP	206647	C3017	C3017CSC2	biopsy_16S	0			1	ibdmdb										
27	C3021CSC1_BP	206656	C3021	C3021CSC1	biopsy_16S	2			1	ibdmdb										
28	C3021CSC2_BP	206655	C3021	C3021CSC2	biopsy_16S	2			1	ibdmdb										
29	C3022CSC1_BP	206657	C3022	C3022CSC1	biopsy_16S	0			1	ibdmdb										
30	C3022CSC2_BP	206658	C3022	C3022CSC2	biopsy_16S	0			1	ibdmdb										
31	C3023CSC1_BP	206659	C3023	C3023CSC1	biopsy_16S	0			1	ibdmdb										
32	C3023CSC2_BP	206660	C3023	C3023CSC2	biopsy_16S	0			1	ibdmdb										
33	C3027CSC1_BP	206667	C3027	C3027CSC1	biopsy_16S	0			1	ibdmdb										
34	C3027CSC2_BP	206668	C3027	C3027CSC2	biopsy_16S	0			1	ibdmdb										
35	C3029CSC1_BP	206670	C3029	C3029CSC1	biopsy_16S	0			1	ibdmdb										
36	C3029CSC2_BP	206669	C3029	C3029CSC2	biopsy_16S	0			1	ibdmdb										
37	C3030CSC1_BP	206671	C3030	C3030CSC1	biopsy_16S	0			1	ibdmdb										

Excel

proceed with caution

BMC Bioinformatics

BioMed Central

Correspondence

Open Access

Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics

Barry R Zeeberg[†1], Joseph Riss[†2], David W Kane[3], Kimberly J Bussey[1], Edward Uchio[4], W Marston Linehan[4], J Carl Barrett[2] and John N Weinstein[*1]

Ziemann *et al. Genome Biology* (2016) 17:177
DOI 10.1186/s13059-016-1044-7

Genome Biology

COMMENT Open Access

Gene name errors are widespread in the scientific literature

Mark Ziemann[1], Yotam Eren[1,2] and Assam El-Osta[1,3*]

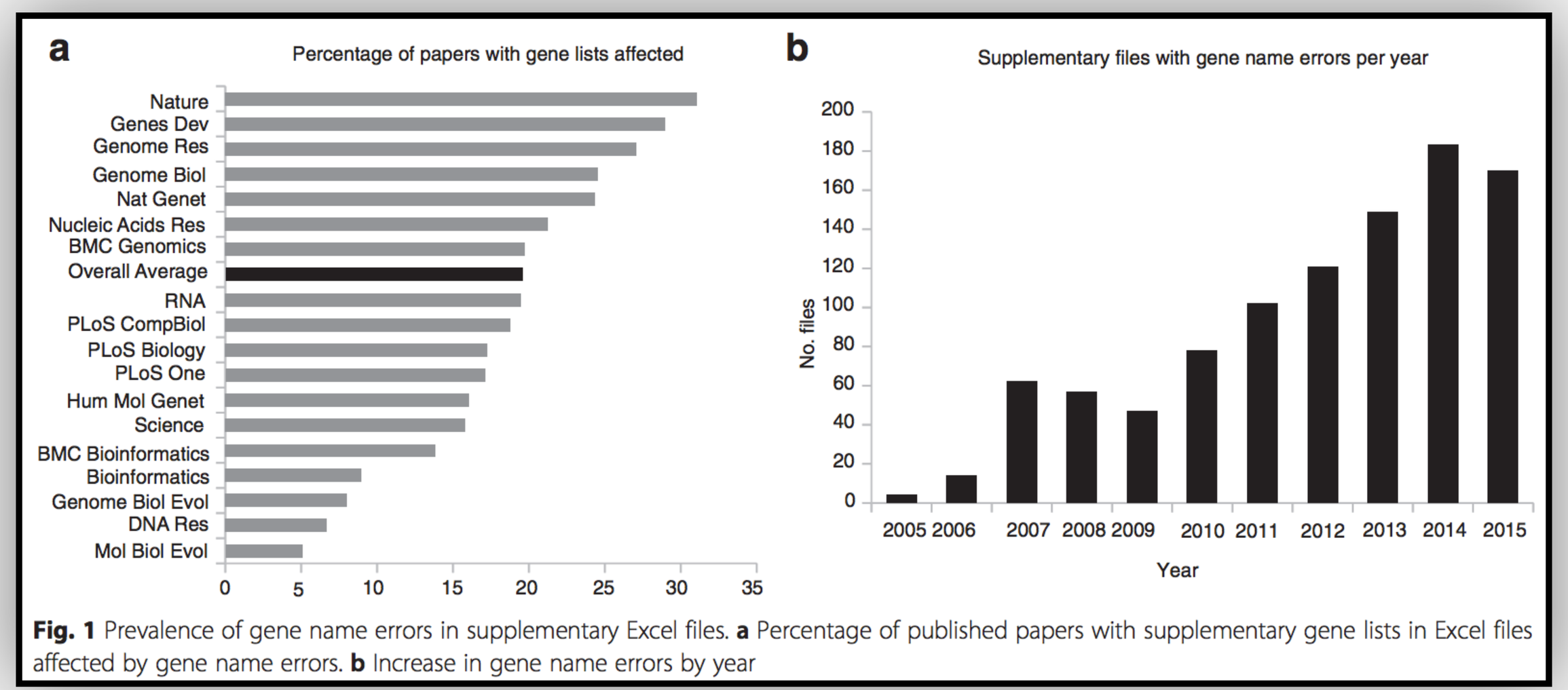
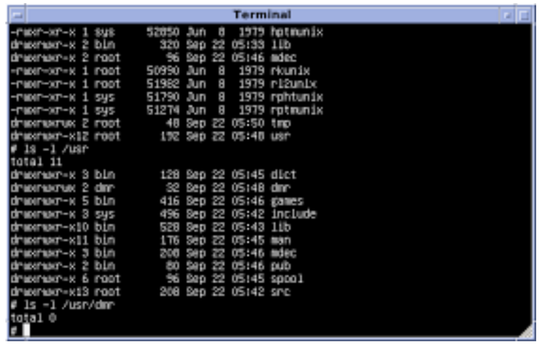


Fig. 1 Prevalence of gene name errors in supplementary Excel files. **a** Percentage of published papers with supplementary gene lists in Excel files affected by gene name errors. **b** Increase in gene name errors by year

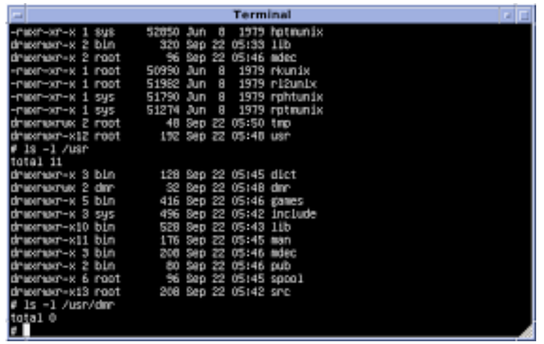
Ziemann et al., Genome Biology 2016

Unix (shell)



JavaScript

Unix (shell)



Bioconductor

OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

JavaScript

My laptop

Overview | Displays | Storage | Support | Service

macOS Catalina
Version 10.15.2

MacBook Pro (13-inch, 2017, Four Thunderbolt 3 Ports)
Processor 3.1 GHz Dual-Core Intel Core i5
Memory 8 GB 2133 MHz LPDDR3
Graphics Intel Iris Plus Graphics 650 1536 MB
Serial Number [redacted]

System Report... | Software Update...

Overview | Displays | Storage | Support | Service

macOS Big Sur
Version 11.3.1

MacBook Pro (13-inch, 2020, Four Thunderbolt 3 ports)
Processor 2.3 GHz Quad-Core Intel Core i7
Memory 32 GB 3733 MHz LPDDR4X
Graphics Intel Iris Plus Graphics 1536 MB
Serial Number [redacted]

System Report... | Software Update...

Other options for compute resources

Option 1
Lab workstation

Pros
- ✓ unrestricted access
- ✓ all the software you need
- ✓ secure

Cons

$15,000
12 core CPU
512 Gb RAM
10Tb RAID1 storage

Option 2
the ‘cloud’

Pros
- ✓ scalable based on needs
- ✓ easy set-up
- ✓ software installation made easy by docker and conda

Cons
- not secure
- data transfer
- Can accrue charges

Option 3
compute cluster

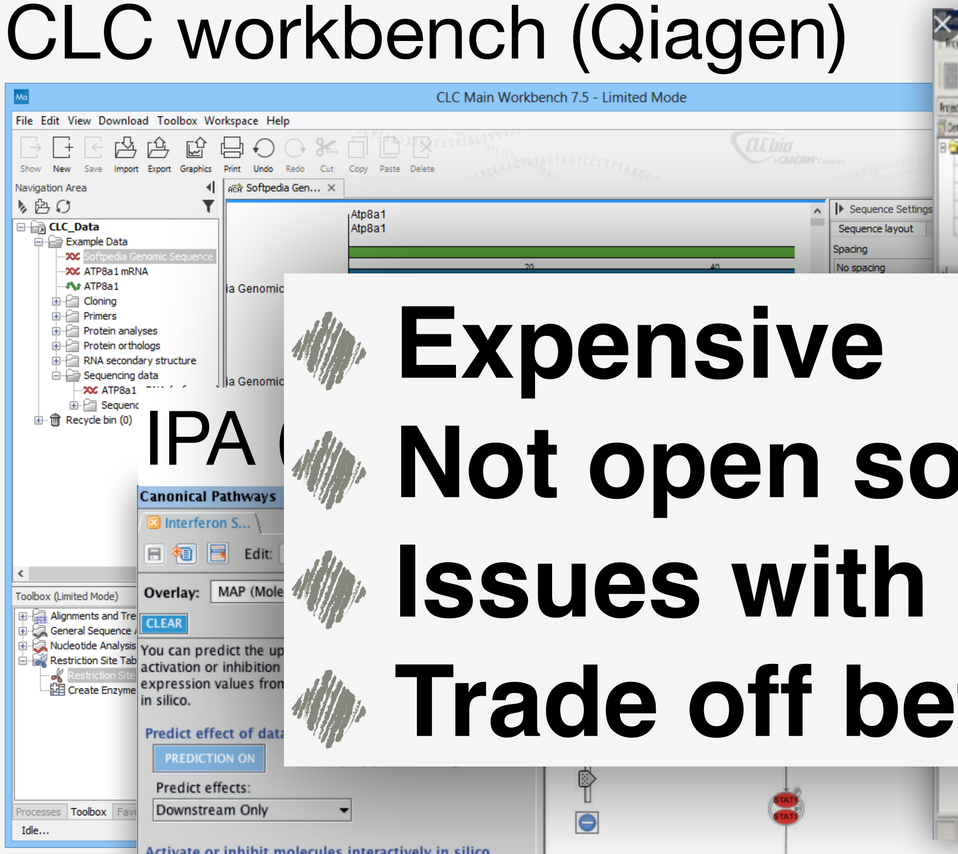
Pros
- ✓ scalable based on needs
- ✓ highly cost efficient
- ✓ SysAdmin taken care of
- ✓ secure

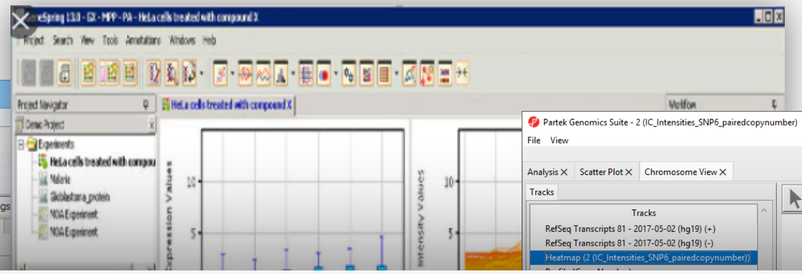
Cons
- may have to request software
- steep learning curve
- frequent server downtime

Commercial Solutions

GeneSpring (GeneSpring)

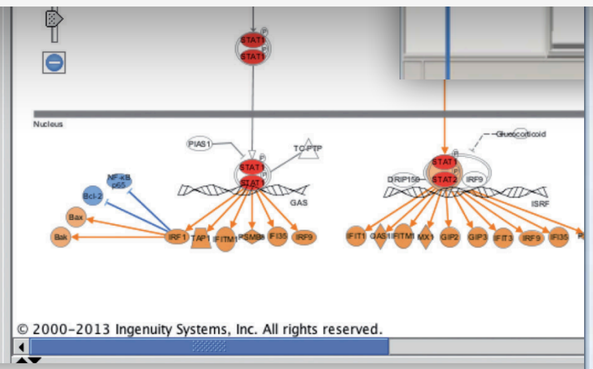
CLC workbench (Qiagen)



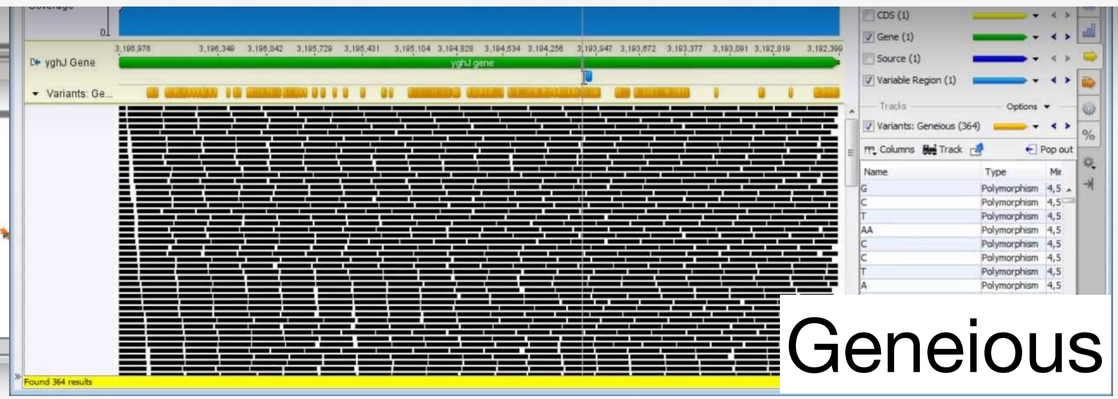


Partek Genomics Suite

IPA



- **Expensive**
- **Not open source**
- **Issues with reproducibility**
- **Trade off between power and simplicity**



Geneious

Why R/bioconductor?
a powerful tool for bioinformatics

Pros

- ☑ free and open-source software
- ☑ used for over a decade to analyze genomic data
- ☑ often don't need to write code *de novo*
- ☑ huge user community (important for learning/help)
- ☑ publication quality graphics
- ☑ reproducible results
- ☑ modular and standardized
- ☑ not just transcriptomics - ChIPseq, scRNAseq, genetic screens, interaction networks, and much more

Cons

- ☑ steep learning curve
- ☑ no single 'right way' to do anything
- ☑ historically, R was lacking in terms of user interface

Installing software

What is it?	What does it do?	How do I get it?	Free?	Cross platform?
R	**Programming language**	r-project.org	Yes	Yes
RStudio	IDE	rstudio.com	Yes*	Yes
Sublime	Text editor	sublimetext.com	Yes	Yes
Visual Studio Code	Text editor/IDE	code.visualstudio.com	Yes	Yes
Kallisto	Maps raw reads	protocols.hostmicrobe.org/conda	Yes	Yes
Kb-python	Single cell preprocessing		Yes	Yes
FastQC	Quality check reads		Yes	Yes
MultiQC	Summarize outputs		Yes	Yes
Sourmash			Yes	Yes
Centrifuge			Yes	Yes

I will demo how to install using Conda
(Kallisto is the most important)

Other great software, but not suitable for laptop

Installing software

What is it?	What does it do?	How do I get it?	Free?	Cross platform?
R	**Programming language**	r-project.org	Yes	Yes
RStudio	IDE	rstudio.com	Yes*	Yes
Sublime	Text editor	sublimetext.com	Yes	Yes
Visual Studio Code	Text editor/IDE	code.visualstudio.com	Yes	Yes
Kallisto	Maps raw reads	protocols.hostmicrobe.org/conda	Yes	Yes
Kb-python	Single cell preprocessing		Yes	Yes
FastQC	Quality check reads		Yes	Yes
MultiQC	Summarize outputs		Yes	Yes
Sourmash	Metagenomics		Yes	Yes
Centrifuge	Metagenomics		Yes	Yes

Other great software, but not suitable for laptop

Open R

RStudio as an IDE for R

"Integrated development environments [IDE] are designed to maximize programmer productivity by providing tight-knit components with similar user interfaces. IDEs present a single program in which all development is done. This program typically provides many features for authoring, modifying, compiling, deploying and debugging software" – *Wikipedia*

RStudio®

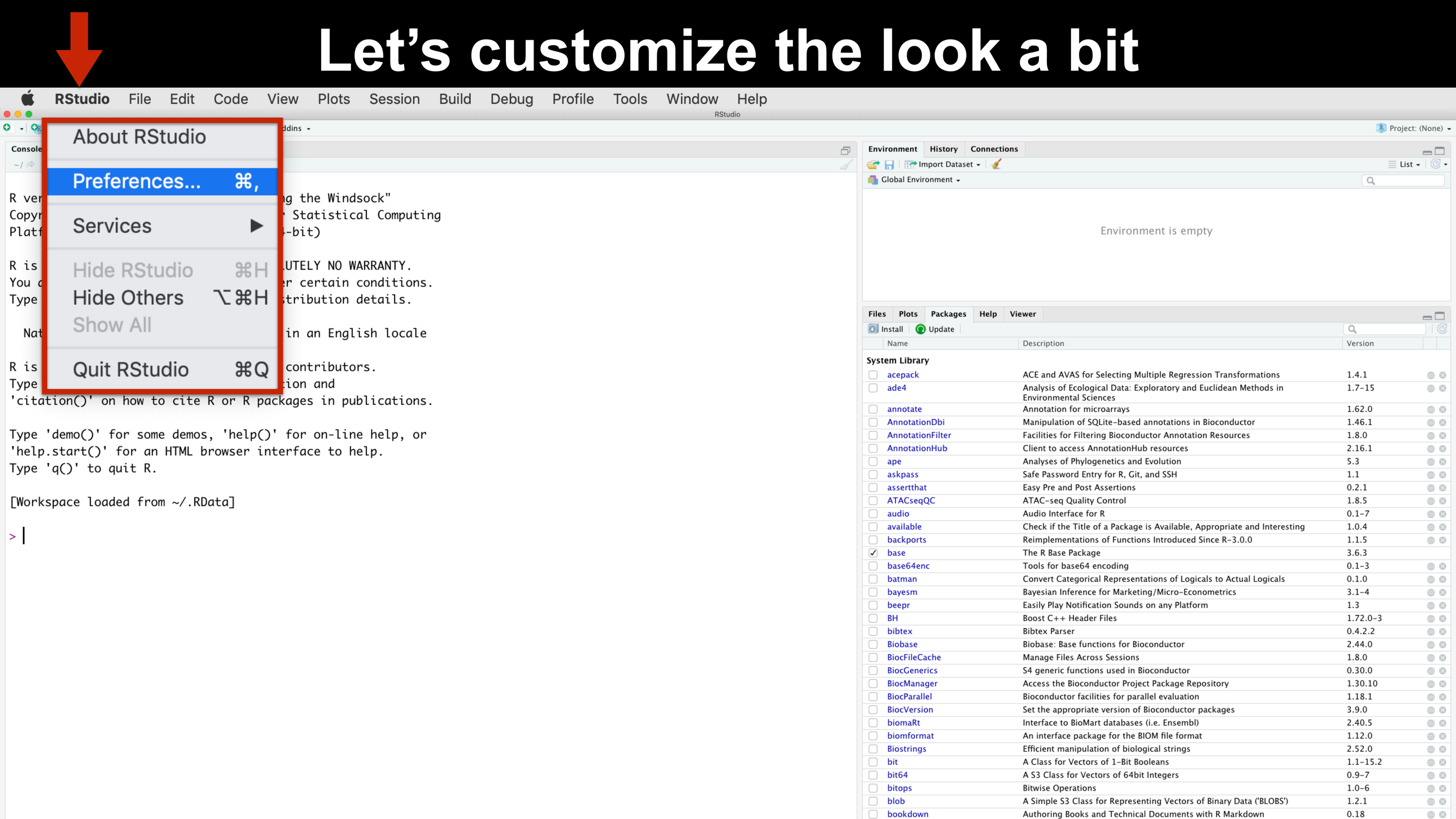
R

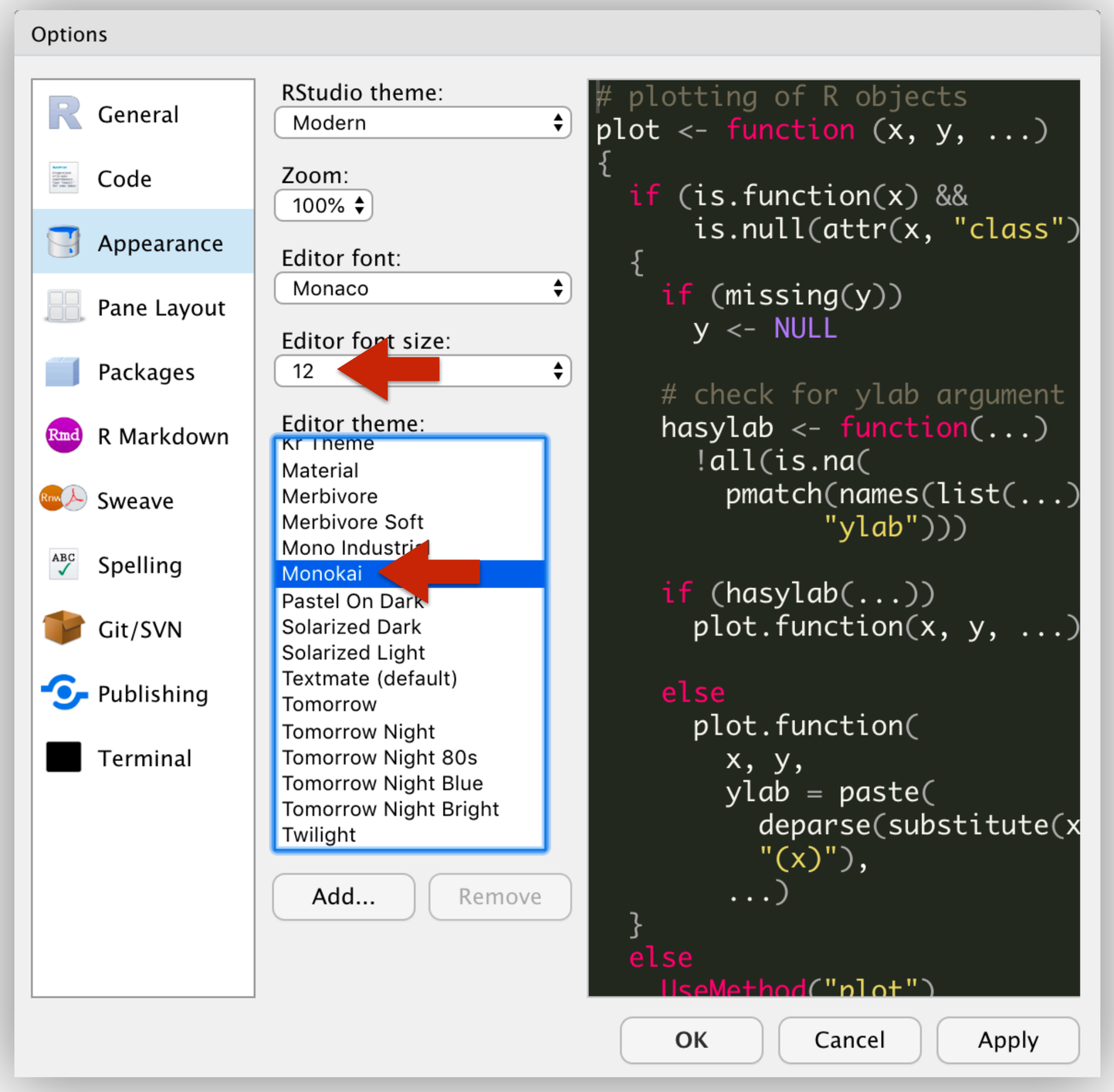
Terminal

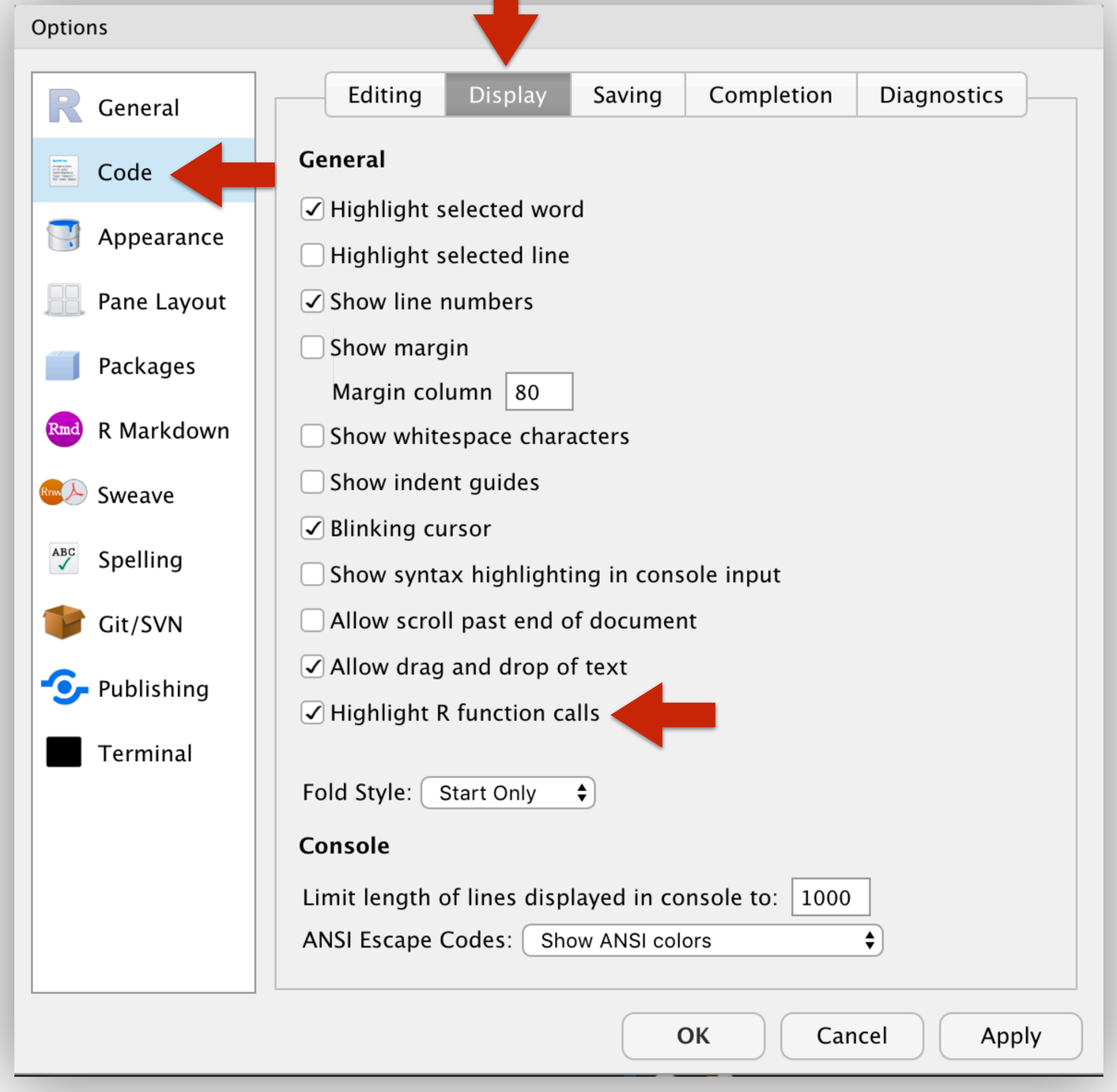
Text editor
(code aware)

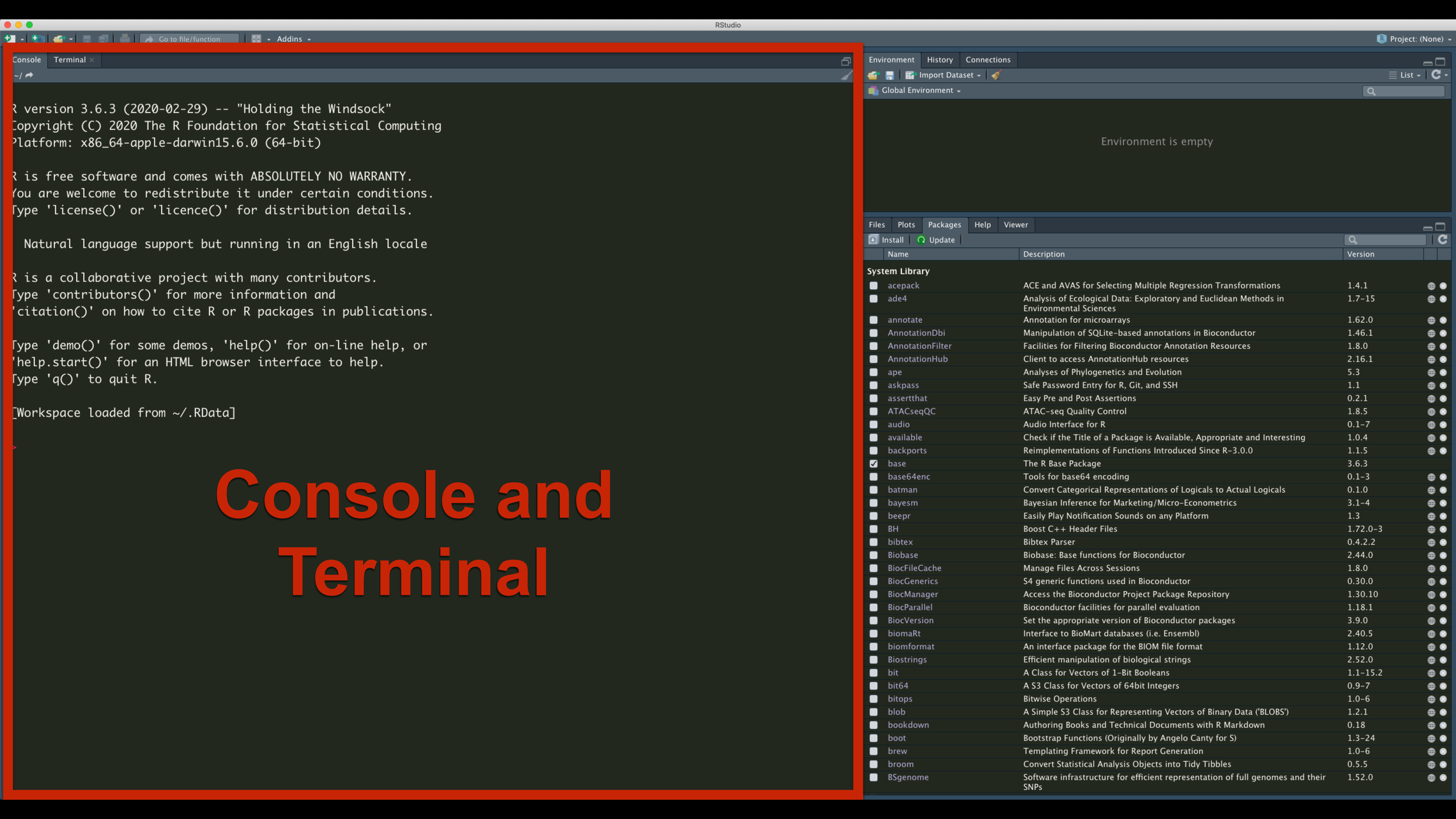
Open RStudio

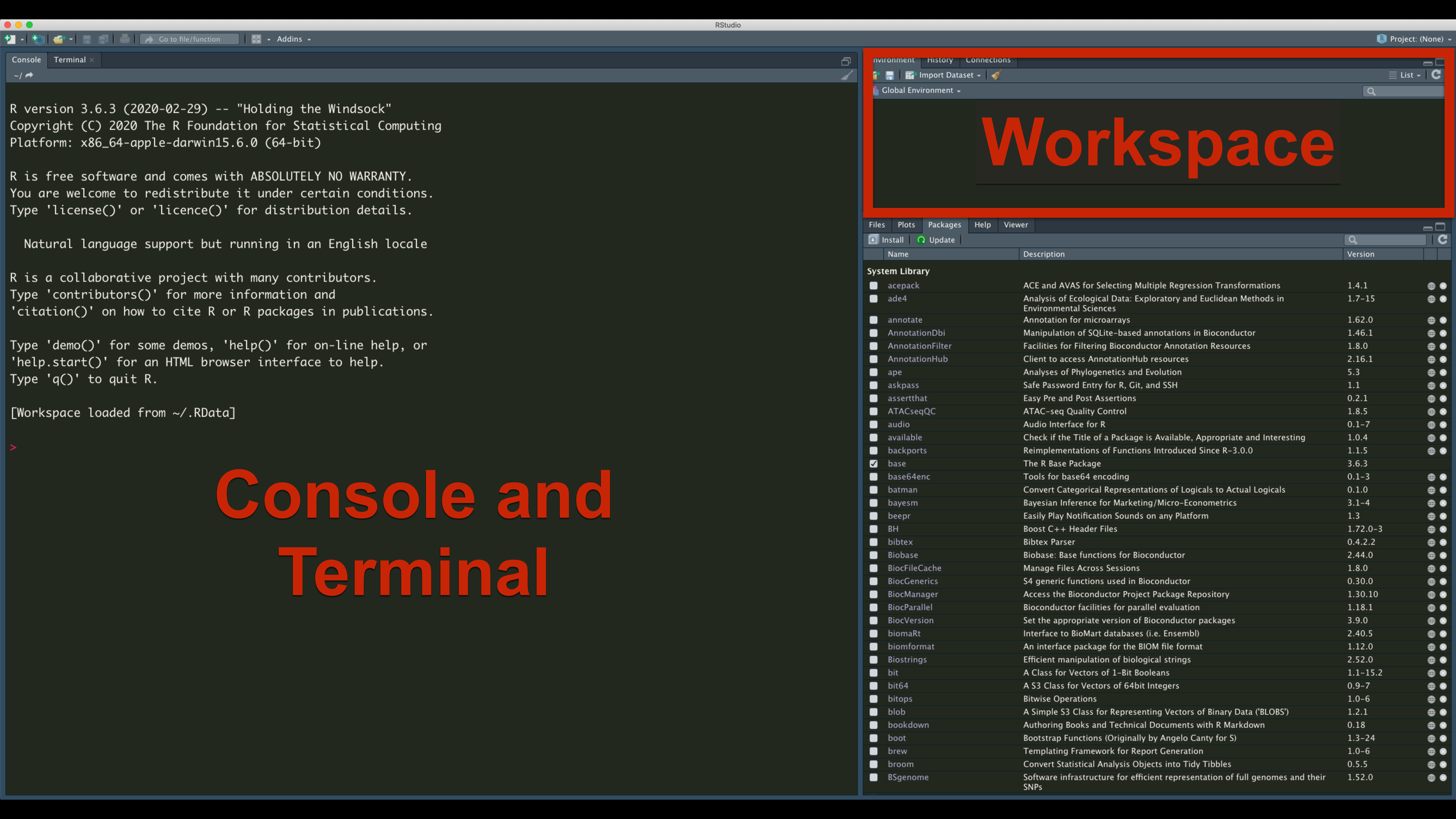
Let’s customize the look a bit

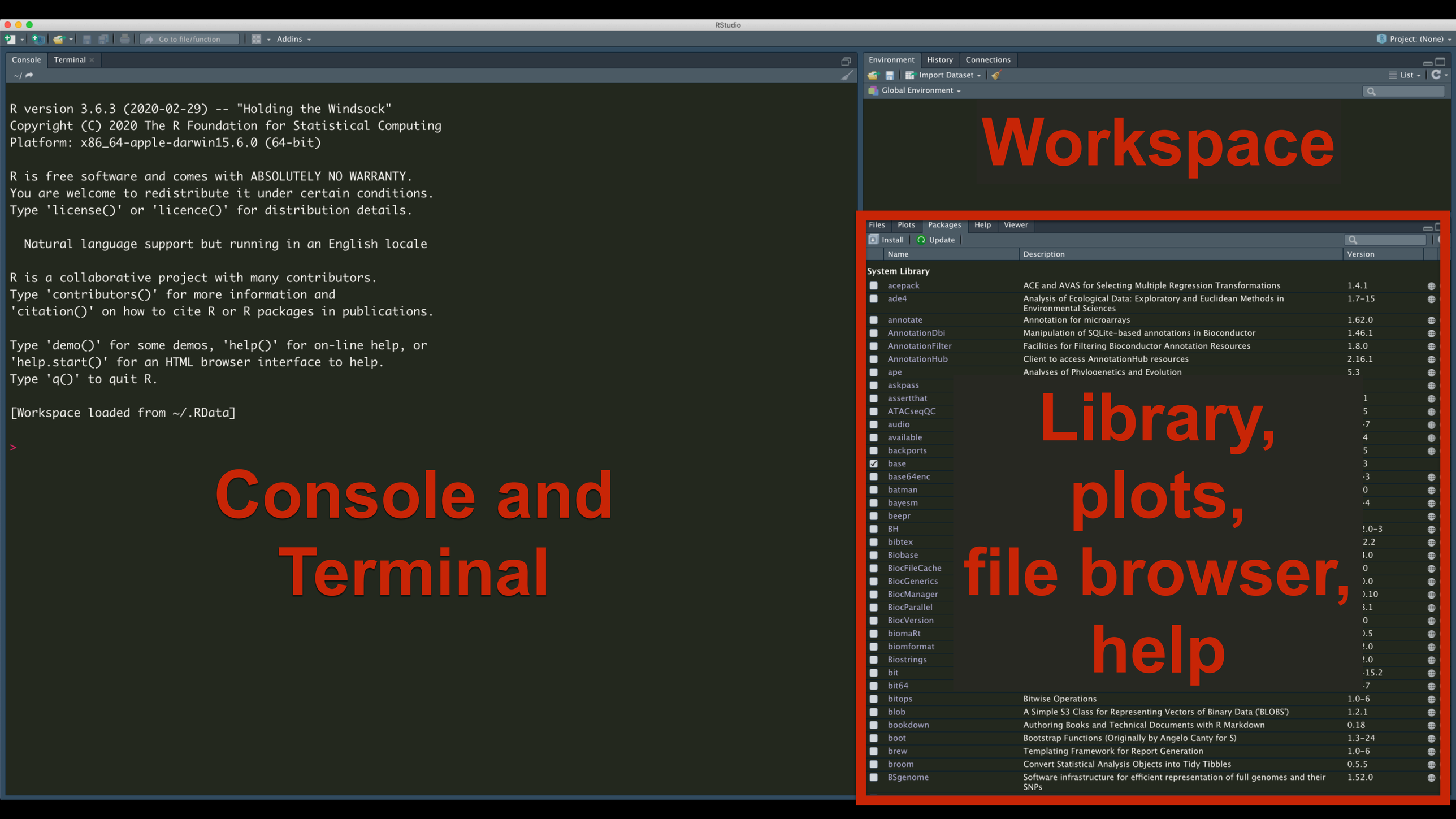



Options
General
Code
Appearance
Pane Layout
Packages
R Markdown
Sweave
Spelling
Git/SVN
Publishing
Terminal
RStudio theme:
Modern
Zoom:
100%
Editor font:
Monaco
Editor font size:
12
Editor theme:
Material
Merbivore
Merbivore Soft
Monokai
Pastel On Dark
Solarized Dark
Solarized Light
Textmate (default)
Tomorrow
Tomorrow Night
Tomorrow Night 80s
Tomorrow Night Blue
Tomorrow Night Bright
Twilight
Add...
Remove
# 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(substitute(x
"(x)"),
...)
}
else
OK
Cancel
Apply


Options
General
Code
Appearance
Pane Layout
Packages
R Markdown
Sweave
Spelling
Git/SVN
Publishing
Terminal
Editing
Display
Saving
Completion
Diagnostics
General
Highlight selected word
Highlight selected line
Show line numbers
Show margin
Margin column
80
Show whitespace characters
Show indent guides
Blinking cursor
Show syntax highlighting in console input
Allow scroll past end of document
Allow drag and drop of text
Highlight R function calls
Fold Style:
Start Only
Console
Limit length of lines displayed in console to:
1000
ANSI Escape Codes:
Show ANSI colors
OK
Cancel
Apply


RStudio
Go to file/function
Addins
Project: (None)
Console
Terminal
~/
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)
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.
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.
[Workspace loaded from ~/.RData]
Console and Terminal
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
System Library
acepack ACE and AVAS for Selecting Multiple Regression Transformations 1.4.1
ade4 Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences 1.7-15
annotate Annotation for microarrays 1.62.0
AnnotationDbi Manipulation of SQLite-based annotations in Bioconductor 1.46.1
AnnotationFilter Facilities for Filtering Bioconductor Annotation Resources 1.8.0
AnnotationHub Client to access AnnotationHub resources 2.16.1
ape Analyses of Phylogenetics and Evolution 5.3
askpass Safe Password Entry for R, Git, and SSH 1.1
assertthat Easy Pre and Post Assertions 0.2.1
ATACseqQC ATAC-seq Quality Control 1.8.5
audio Audio Interface for R 0.1-7
available Check if the Title of a Package is Available, Appropriate and Interesting 1.0.4
backports Reimplementations of Functions Introduced Since R-3.0.0 1.1.5
base The R Base Package 3.6.3
base64enc Tools for base64 encoding 0.1-3
batman Convert Categorical Representations of Logicals to Actual Logicals 0.1.0
bayesm Bayesian Inference for Marketing/Micro-Econometrics 3.1-4
beepr Easily Play Notification Sounds on any Platform 1.3
BH Boost C++ Header Files 1.72.0-3
bibtex Bibtex Parser 0.4.2.2
Biobase Biobase: Base functions for Bioconductor 2.44.0
BiocFileCache Manage Files Across Sessions 1.8.0
BiocGenerics S4 generic functions used in Bioconductor 0.30.0
BiocManager Access the Bioconductor Project Package Repository 1.30.10
BiocParallel Bioconductor facilities for parallel evaluation 1.18.1
BiocVersion Set the appropriate version of Bioconductor packages 3.9.0
biomaRt Interface to BioMart databases (i.e. Ensembl) 2.40.5
biomformat An interface package for the BIOM file format 1.12.0
Biostrings Efficient manipulation of biological strings 2.52.0
bit A Class for Vectors of 1-Bit Booleans 1.1-15.2
bit64 A S3 Class for Vectors of 64bit Integers 0.9-7
bitops Bitwise Operations 1.0-6
blob A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') 1.2.1
bookdown Authoring Books and Technical Documents with R Markdown 0.18
boot Bootstrap Functions (Originally by Angelo Canty for S) 1.3-24
brew Templating Framework for Report Generation 1.0-6
broom Convert Statistical Analysis Objects into Tidy Tibbles 0.5.5
BSgenome Software infrastructure for efficient representation of full genomes and their SNPs 1.52.0


RStudio
Go to file/function
Addins
Project: (None)
Console
Terminal
~/
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)
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.
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.
[Workspace loaded from ~/.RData]
>
Console and Terminal
History
Connections
Import Dataset
List
Global Environment
Workspace
Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
System Library
acepack ACE and AVAS for Selecting Multiple Regression Transformations 1.4.1
ade4 Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences 1.7-15
annotate Annotation for microarrays 1.62.0
AnnotationDbi Manipulation of SQLite-based annotations in Bioconductor 1.46.1
AnnotationFilter Facilities for Filtering Bioconductor Annotation Resources 1.8.0
AnnotationHub Client to access AnnotationHub resources 2.16.1
ape Analyses of Phylogenetics and Evolution 5.3
askpass Safe Password Entry for R, Git, and SSH 1.1
assertthat Easy Pre and Post Assertions 0.2.1
ATACseqQC ATAC-seq Quality Control 1.8.5
audio Audio Interface for R 0.1-7
available Check if the Title of a Package is Available, Appropriate and Interesting 1.0.4
backports Reimplementations of Functions Introduced Since R-3.0.0 1.1.5
base The R Base Package 3.6.3
base64enc Tools for base64 encoding 0.1-3
batman Convert Categorical Representations of Logicals to Actual Logicals 0.1.0
bayesm Bayesian Inference for Marketing/Micro-Econometrics 3.1-4
beepr Easily Play Notification Sounds on any Platform 1.3
BH Boost C++ Header Files 1.72.0-3
bibtex Bibtex Parser 0.4.2.2
Biobase Biobase: Base functions for Bioconductor 2.44.0
BiocFileCache Manage Files Across Sessions 1.8.0
BiocGenerics S4 generic functions used in Bioconductor 0.30.0
BiocManager Access the Bioconductor Project Package Repository 1.30.10
BiocParallel Bioconductor facilities for parallel evaluation 1.18.1
BiocVersion Set the appropriate version of Bioconductor packages 3.9.0
biomaRt Interface to BioMart databases (i.e. Ensembl) 2.40.5
biomformat An interface package for the BIOM file format 1.12.0
Biostrings Efficient manipulation of biological strings 2.52.0
bit A Class for Vectors of 1-Bit Booleans 1.1-15.2
bit64 A S3 Class for Vectors of 64bit Integers 0.9-7
bitops Bitwise Operations 1.0-6
blob A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') 1.2.1
bookdown Authoring Books and Technical Documents with R Markdown 0.18
boot Bootstrap Functions (Originally by Angelo Canty for S) 1.3-24
brew Templating Framework for Report Generation 1.0-6
broom Convert Statistical Analysis Objects into Tidy Tibbles 0.5.5
BSgenome Software infrastructure for efficient representation of full genomes and their SNPs 1.52.0

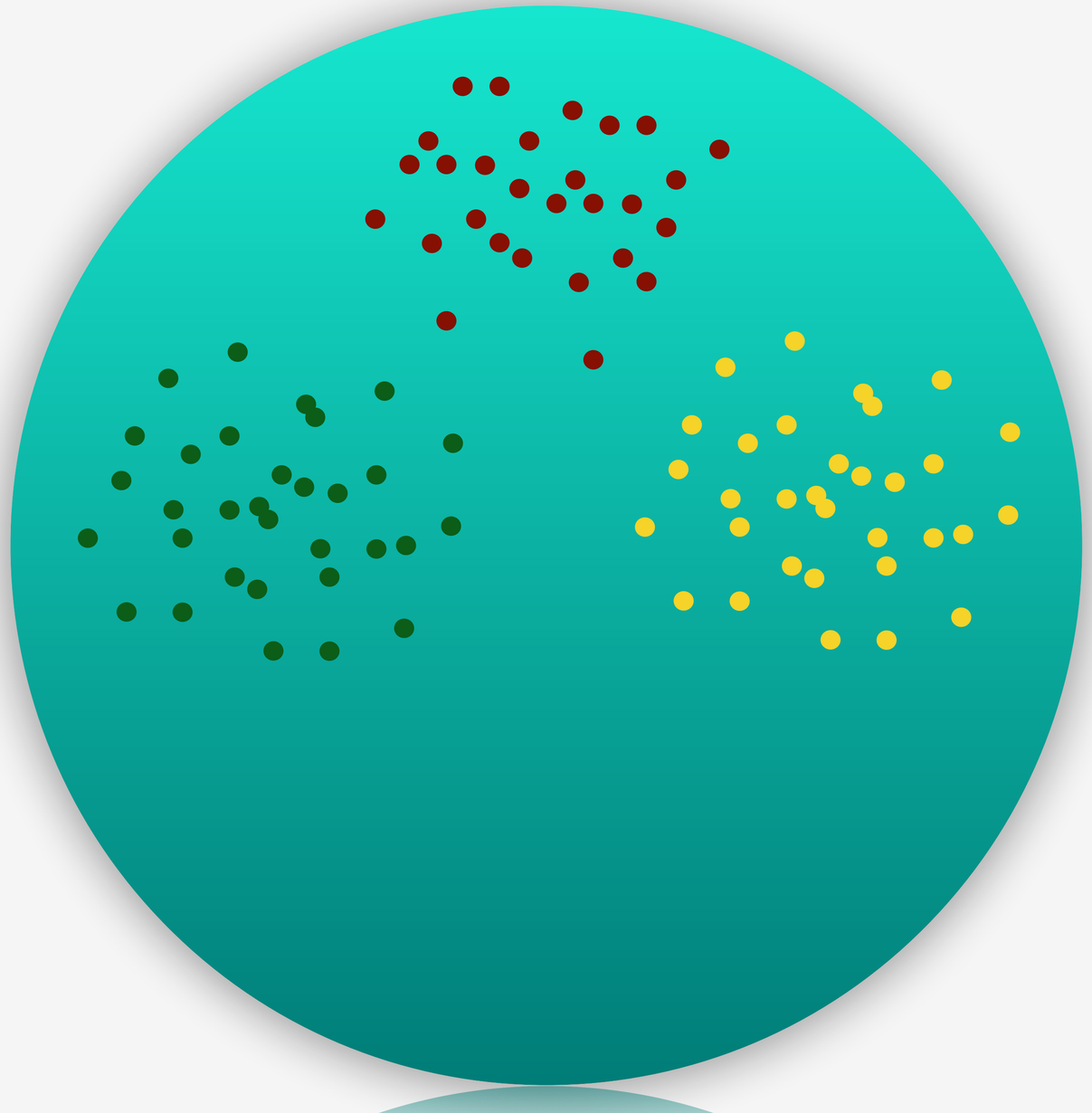

RStudio
Go to file/function
Addins
Project: (None)
Console
Terminal
~/
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
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.
[Workspace loaded from ~/.RData]
>
Console and Terminal
Environment
History
Connections
Import Dataset
List
Global Environment
Workspace
Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
System Library
acepack ACE and AVAS for Selecting Multiple Regression Transformations 1.4.1
ade4 Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences 1.7-15
annotate Annotation for microarrays 1.62.0
AnnotationDbi Manipulation of SQLite-based annotations in Bioconductor 1.46.1
AnnotationFilter Facilities for Filtering Bioconductor Annotation Resources 1.8.0
AnnotationHub Client to access AnnotationHub resources 2.16.1
ape Analyses of Phylogenetics and Evolution 5.3
askpass
assertthat 1
ATACseqQC 5
audio -7
available 4
backports 5
base 3
base64enc -3
batman 0
bayesm -4
beepr
BH .0-3
bibtex 2.2
Biobase .0
BiocFileCache 0
BiocGenerics .0
BiocManager .10
BiocParallel .1
BiocVersion 0
biomaRt .5
biomformat .0
Biostrings .0
bit -15.2
bit64 -7
bitops Bitwise Operations 1.0-6
blob A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') 1.2.1
bookdown Authoring Books and Technical Documents with R Markdown 0.18
boot Bootstrap Functions (Originally by Angelo Canty for S) 1.3-24
brew Templating Framework for Report Generation 1.0-6
broom Convert Statistical Analysis Objects into Tidy Tibbles 0.5.5
BSgenome Software infrastructure for efficient representation of full genomes and their SNPs 1.52.0
Library, plots, file browser, help

An interesting time for R

Open-source

Commercial

The wonderful world of R



Package type

- statistics
- graphing
- modeling

>18,000 packages on CRAN

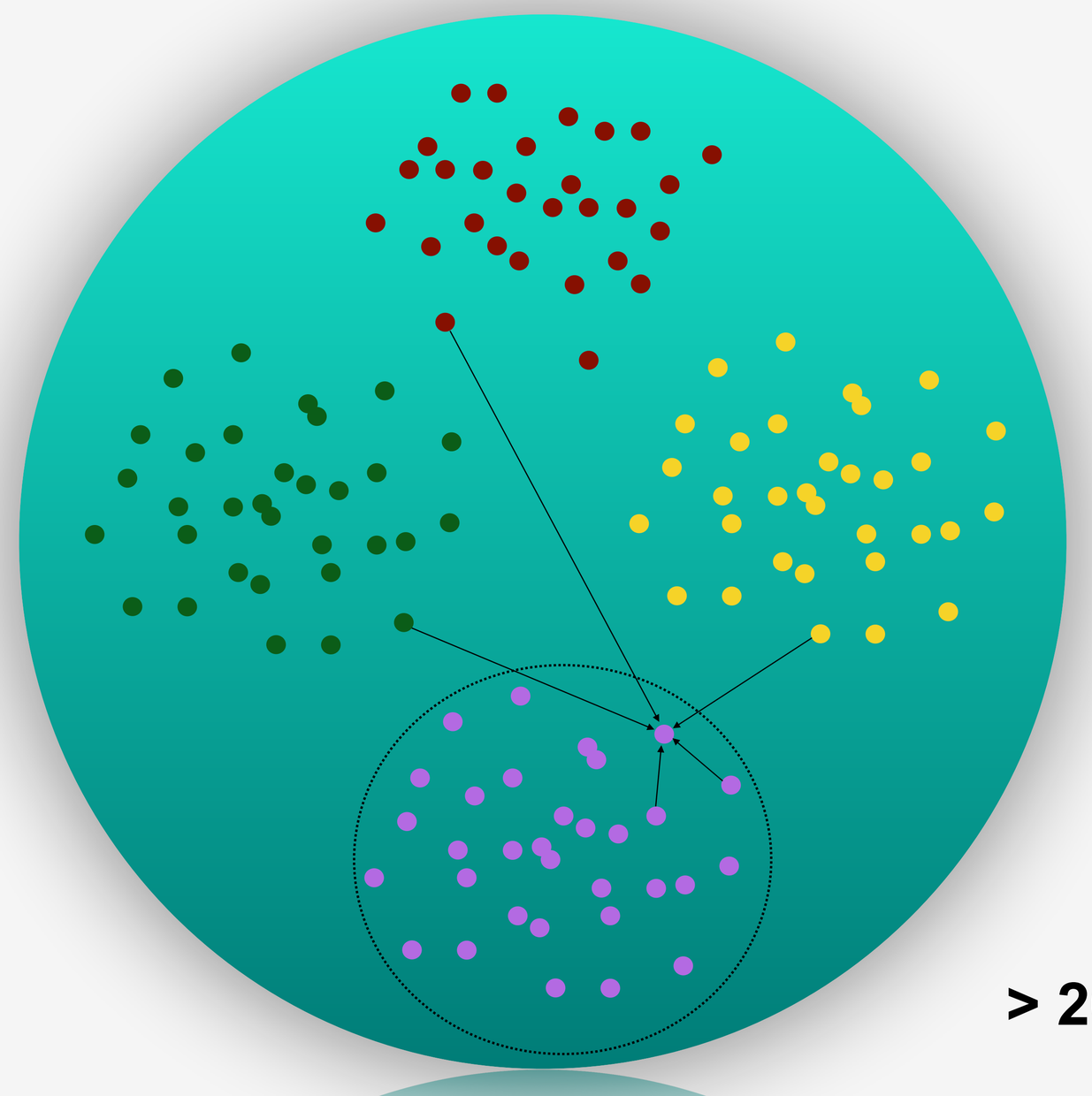
Orchestrating high-throughput genomic analysis with Bioconductor

Wolfgang Huber[1], Vincent J Carey[2,3], Robert Gentleman[4], Simon Anders[1], Marc Carlson[5], Benilton S Carvalho[6], Hector Corrada Bravo[7], Sean Davis[8], Laurent Gatto[9], Thomas Girke[10], Raphael Gottardo[11], Florian Hahne[12], Kasper D Hansen[13,14], Rafael A Irizarry[3,15], Michael Lawrence[4], Michael I Love[3,15], James MacDonald[16], Valerie Obenchain[5], Andrzej K Oleś[1], Hervé Pagès[5], Alejandro Reyes[1], Paul Shannon[5], Gordon K Smyth[17,18], Dan Tenenbaum[5], Levi Waldron[19] & Martin Morgan[5]

NATURE METHODS | VOL.12 NO.2 | FEBRUARY 2015 | 115

"We have embraced R for its scientific and statistical computing capabilities, for its graphics facilities and for the convenience of an interpreted language. R also interfaces with low-level languages including C and C++ for computationally intensive operations, Java for integration with enterprise software and JavaScript for interactive web-based applications and reports."

Packages are modular and can work together



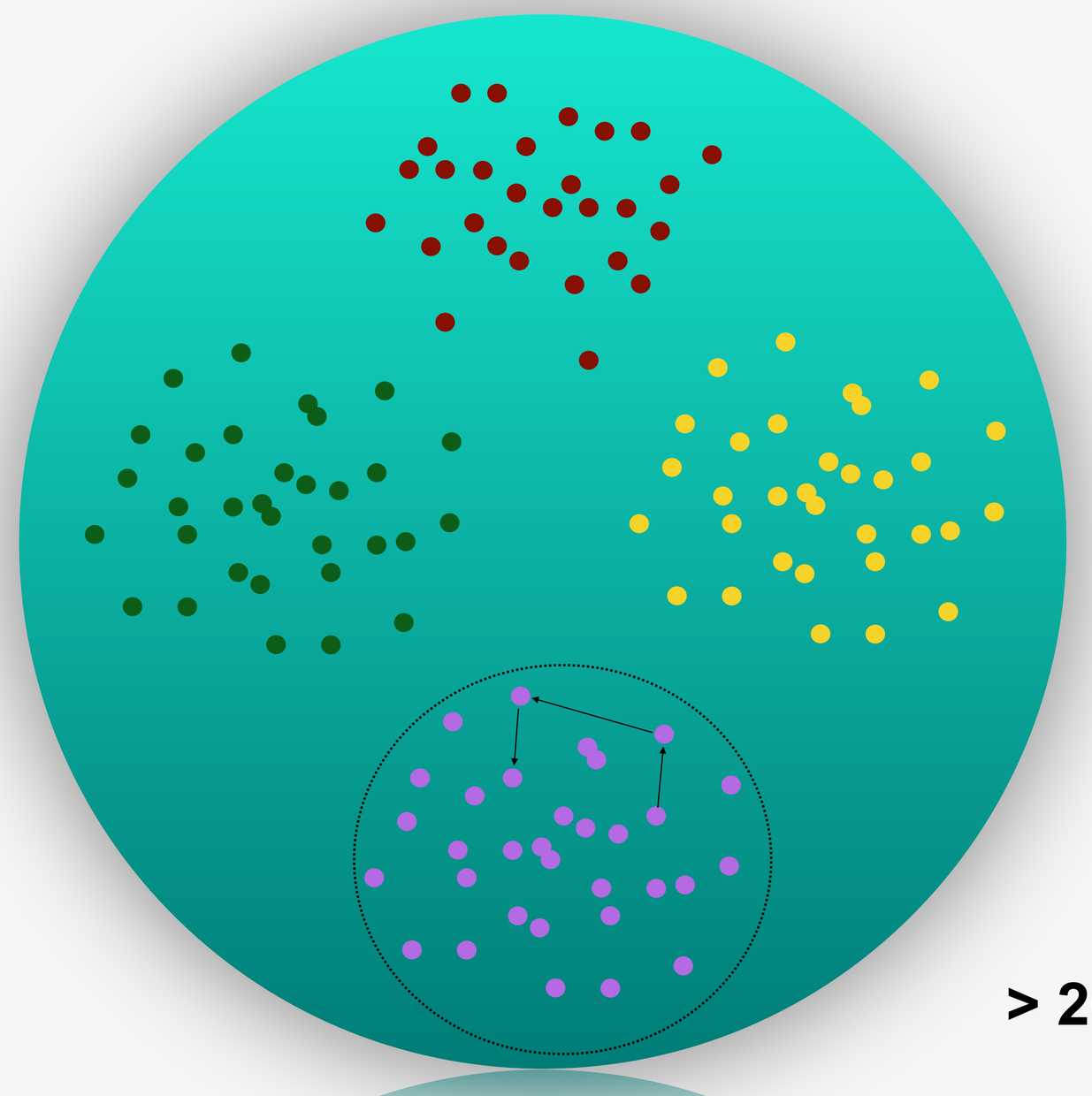
Package type

- statistics
- graphing
- modeling
- bioconductor

>18,000 packages on CRAN

Bioconductor
> 2042 packages (v3.13)

Workflows utilize multiple packages to step through a complex process



Package type

- statistics
- graphing
- modeling
- bioconductor

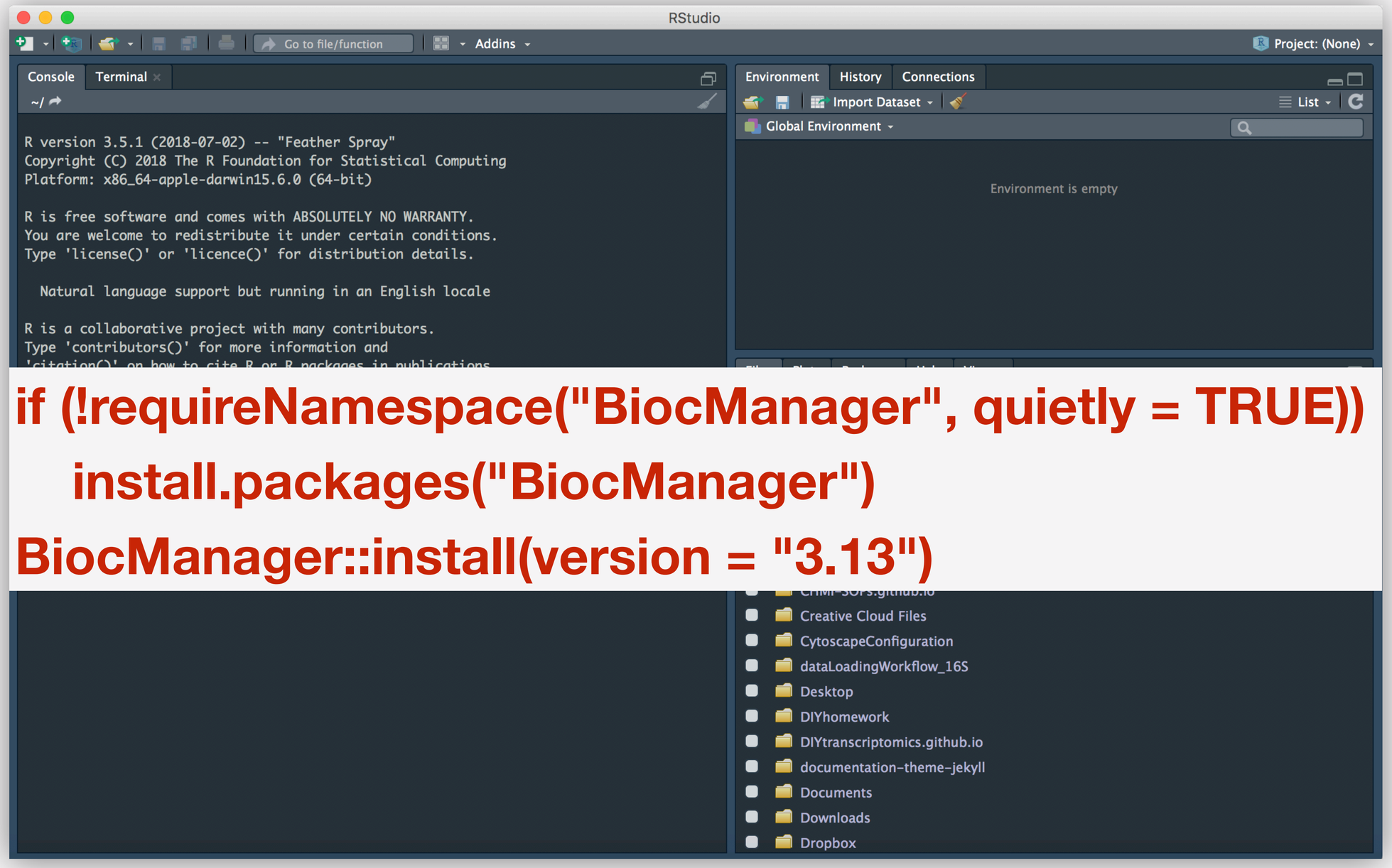
>15,500 packages on CRAN

Bioconductor
> 2042 packages (v3.13)

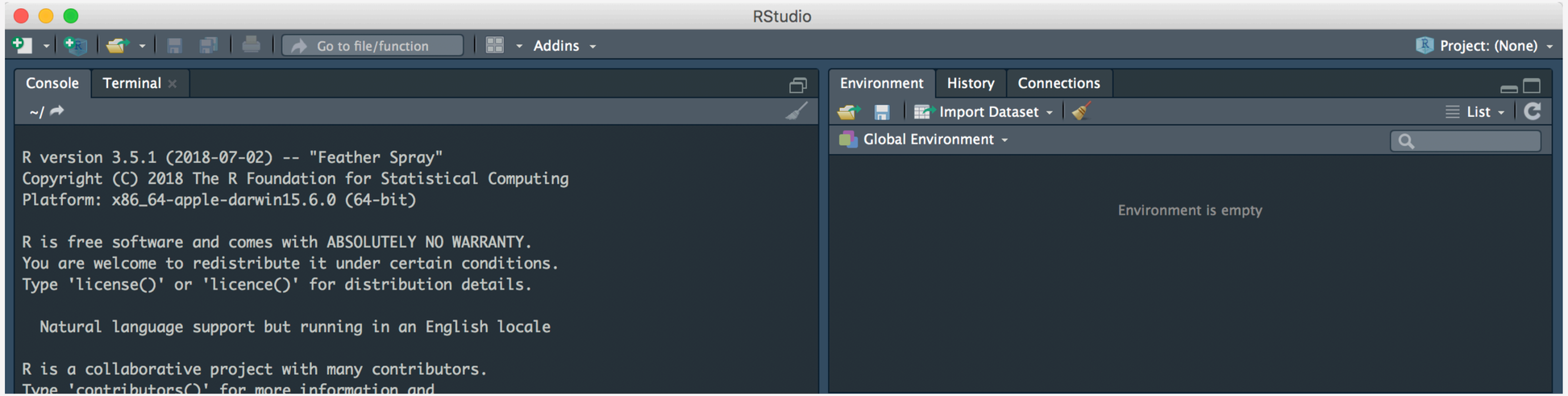
Think of R packages as the software equivalent of a scientific paper

- The *de facto* standard for how to communicate
- a formal and widely accepted structure
- are stand-alone, but build on a knowledge base
- target audience will vary
- content and quality will vary
- new packages are often the basis for a publication

how do I actually *get* bioconductor?

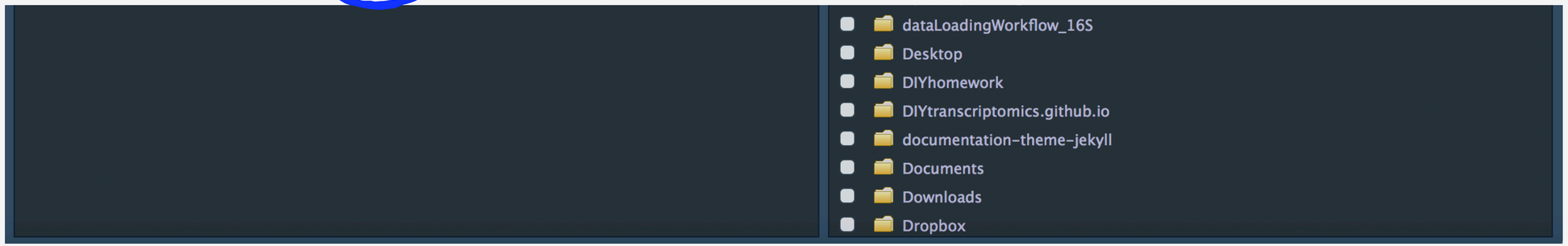


You can always update Bioconductor using the exact same bit of code



Upgrade ___ packages to Bioconductor version '3.13'? [y/n]:

Do you want to install from sources the packages which need compilation? (Yes/no/cancel)



Installing Kallisto

Installing software

What is it?	What does it do?	How do I get it?	Free?	Cross platform?
R	**Programming language**	r-project.org	Yes	Yes
RStudio	IDE	rstudio.com	Yes*	Yes
Sublime	Text editor	sublimetext.com	Yes	Yes
Visual Studio Code	Text editor/IDE	code.visualstudio.com	Yes	Yes
Kallisto	Maps raw reads	protocols.hostmicrobe.org/conda	Yes	Yes
Kb-python	Single cell preprocessing		Yes	Yes
FastQC	Quality check reads		Yes	Yes
MultiQC	Summarize outputs		Yes	Yes
Sourmash			Yes	Yes
Centrifuge			Yes	Yes

I will demo how to install using Conda
(Kallisto is the most important)

Other great software, but not suitable for laptop