

TRGN 527: Applied Data Science and Bioinformatics

UNIT I. Introduction and Basic Data Science

Week 1 – Lecture 2

Enrique I. Velazquez Villarreal, M.D., Ph.D., M.P.H., M.S. | Assistant Professor

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Leader of the USC Bioinformatics Core – *USC CaRE2 Health Equity Center*

David W. Craig, Ph.D. | Professor and Vice Chair

Dept. of Translational Genomics

USC | Keck School of Medicine | Norris Comprehensive Cancer Center

Co-Director, Institute of Translational Genomics

Topics

- Introduction to R, R studio, Bioconductor, open source software and Terminal. Basics of R, including install and configure software necessary for a statistical programming environment.

Introduction to R

<https://www.r-project.org>

Introduction to R

- R is a language and environment for statistical computing and graphics.
- It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues.
- R can be considered as a different implementation of S.
- There are some important differences, but much code written for S runs unaltered under R.



Introduction to R

- R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible.
 - The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.



Introduction to R

- One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed.
- Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.
- R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.



Terminal – version 3.5.X for this course

Last login: Tue Nov 27 17:24:12 on ttys000

M-200472:~ enriquevelazquez\$ R

R version 3.5.1 (2018-07-02) -- "Feather Spray"

Copyright (C) 2018 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.

>

Installing R

<https://www.r-project.org>

Installing R

Go to: <https://www.r-project.org>



The screenshot shows a Safari browser window with the URL <https://www.r-project.org> in the address bar. The page content is as follows:

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#)

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- [R version 3.4.2 \(Short Summer\)](#) has been released on Thursday 2017-09-28.
- [The R Journal Volume 9/1](#) is available.
- [R version 3.3.3 \(Another Canoe\)](#) has been released on Monday 2017-03-06.
- [The R Journal Volume 8/2](#) is available.
- [useR! 2017](#) (July 4 - 7 in Brussels) has opened registration and more at <http://user2017.brussels/>
- Tomas Kalibera has joined the R core team.
- The R Foundation welcomes five new ordinary members: Jennifer Bryan, Dianne Cook, Julie Josse, Tomas Kalibera, and Balasubramanian Narasimhan.
- [The R Journal Volume 8/1](#) is available.
- The [useR! 2017](#) conference will take place in Brussels, July 4 - 7, 2017.
- [R version 3.2.5 \(Very, Very Secure Dishes\)](#) has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- [Notice XQuartz users \(Mac OS X\)](#) A security issue has been detected with the Sparkle update

A large orange arrow points to the [download R](#) link, and the text "CLICK ON" is written in red to its right.

Installing R

Safari File Edit View History Bookmarks Window Help

CRAN – Mirrors

https://cran.r-project.org/mirrors.html

Apple Genome Repository UPCI Yahoo Blueprint Genetics Facebook Google iCloud Twitter Wikipedia News Popular Signalert University of Bristol

USA

<http://www.stats.bris.ac.uk/R/>
<https://mirrors.ebi.ac.uk/CRAN/>
<http://mirrors.ebi.ac.uk/CRAN/>
<https://cran.ma.imperial.ac.uk/>
<http://cran.ma.imperial.ac.uk/>

<https://cran.cnr.berkeley.edu/> **CLICK ON**
<http://cran.cnr.berkeley.edu/>
<http://cran.stat.ucla.edu/>
<https://mirror.las.iastate.edu/CRAN/>
<http://mirror.las.iastate.edu/CRAN/>
<https://ftp.ussg.iu.edu/CRAN/>
<http://ftp.ussg.iu.edu/CRAN/>
<https://rweb.crmda.ku.edu/cran/>
<http://rweb.crmda.ku.edu/cran/>
<https://cran.mtu.edu/>
<http://cran.mtu.edu/>
<http://cran.wustl.edu/>
<http://archive.linux.duke.edu/cran/>
<http://cran.case.edu/>
<http://iis.stat.wright.edu/CRAN/>
<https://ftp.osuosl.org/pub/cran/>
<http://ftp.osuosl.org/pub/cran/>
<http://lib.stat.cmu.edu/R/CRAN/>
<http://cran.mirrors.hoobly.com/>
<https://mirrors.nics.utk.edu/cran/>
<http://mirrors.nics.utk.edu/cran/>
<https://cran.revolutionanalytics.com/>
<http://cran.revolutionanalytics.com/>

Vietnam

<https://mirror.vinahost.vn/CRAN/>

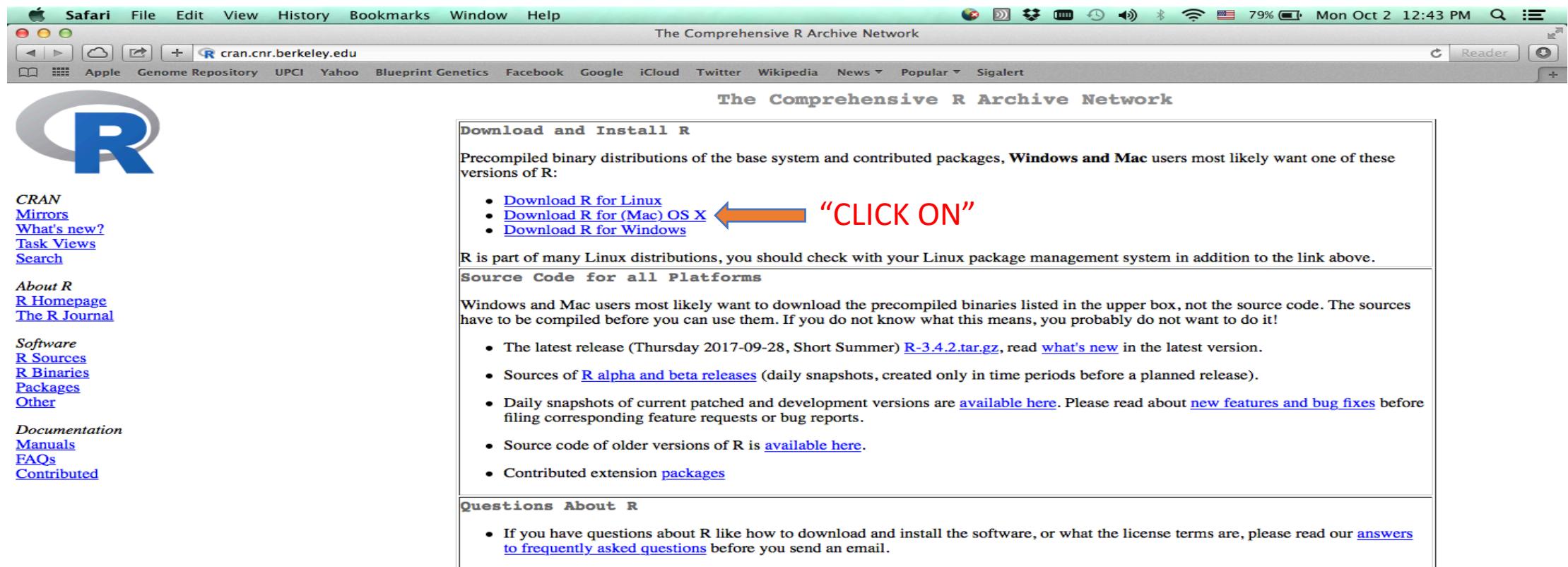
Venezuela

University of Bristol
EMBL-EBI (European Bioinformatics Institute)
EMBL-EBI (European Bioinformatics Institute)
Imperial College London
Imperial College London

University of California, Berkeley, CA
University of California, Berkeley, CA
University of California, Los Angeles, CA
Iowa State University, Ames, IA
Iowa State University, Ames, IA
Indiana University
Indiana University
University of Kansas, Lawrence, KS
University of Kansas, Lawrence, KS
Michigan Technological University, Houghton, MI
Michigan Technological University, Houghton, MI
Washington University, St. Louis, MO
Duke University, Durham, NC
Case Western Reserve University, Cleveland, OH
Wright State University, Dayton, OH
Oregon State University
Oregon State University
Statlib, Carnegie Mellon University, Pittsburgh, PA
Hoobly Classifieds, Pittsburgh, PA
National Institute for Computational Sciences, Oak Ridge, TN
National Institute for Computational Sciences, Oak Ridge, TN
Revolution Analytics, Dallas, TX
Revolution Analytics, Dallas, TX

VinaHost

Installing R



The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#) ← “CLICK ON”
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Thursday 2017-09-28, Short Summer) [R-3.4.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

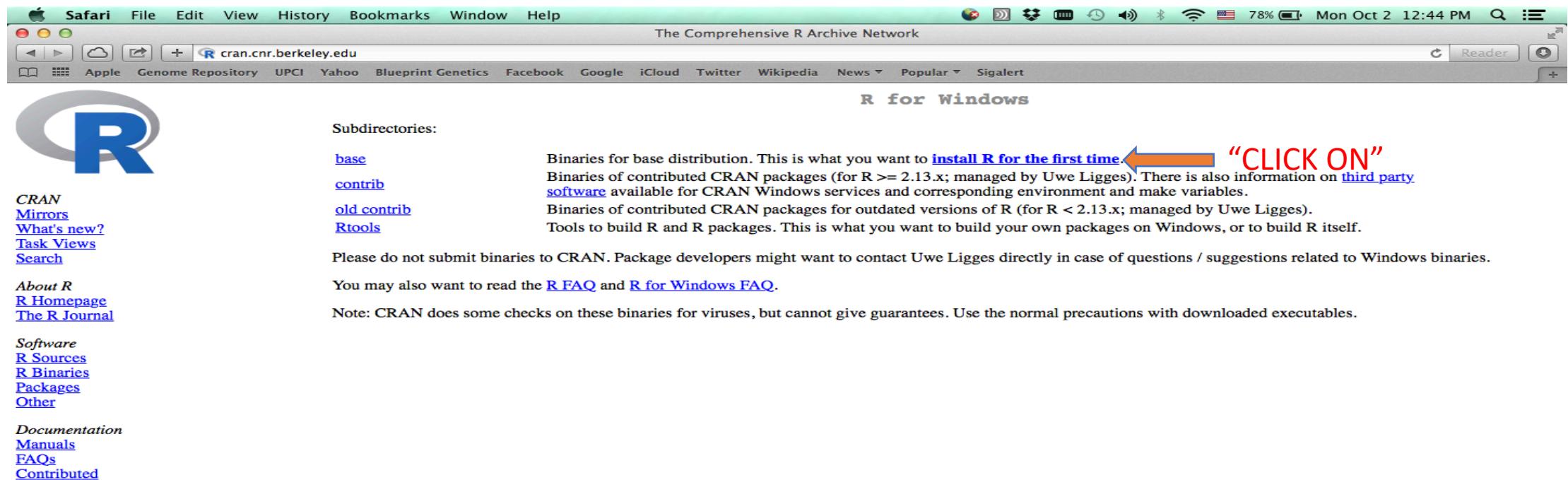
Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?

R is ‘GNU S’, a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

Installing R



The Comprehensive R Archive Network

R for Windows

Subdirectories:

- [base](#) Binaries for base distribution. This is what you want to [install R for the first time](#). ← **"CLICK ON"**
- [contrib](#) Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.
- [old_contrib](#) Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
- [Rtools](#) Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

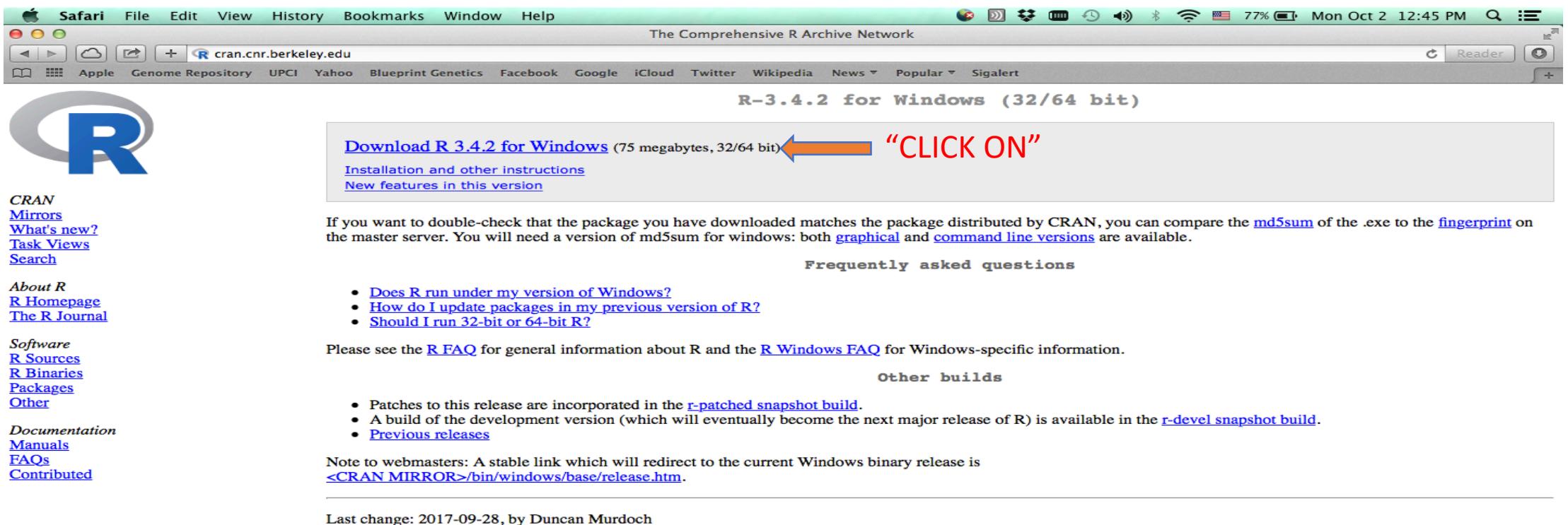
[About R](#)
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

Installing R

IF YOU HAVE WINDOWS SYSTEM:



The screenshot shows a Safari browser window on a Mac OS X desktop. The title bar reads "The Comprehensive R Archive Network". The address bar shows the URL "cran.cnr.berkeley.edu". The main content area displays the "R-3.4.2 for Windows (32/64 bit)" page. On the left, there's a large blue "R" logo and a sidebar with links like "CRAN", "About R", "Software", and "Documentation". The central content includes a link to "Download R 3.4.2 for Windows (75 megabytes, 32/64 bit)". A red arrow points to this link with the text "CLICK ON" written in red above it. Below the download link are sections for "Installation and other instructions" and "New features in this version". Further down, there's a note about md5sum and fingerprint, a "Frequently asked questions" section with links to "Does R run under my version of Windows?", "How do I update packages in my previous version of R?", and "Should I run 32-bit or 64-bit R?". There's also a note about the R FAQ and R Windows FAQ, and a "Other builds" section with links to "r-patched snapshot build", "r-devel snapshot build", and "Previous releases". At the bottom, there's a note for webmasters about a stable link redirecting to the current Windows binary release. The footer indicates the last change was on 2017-09-28 by Duncan Murdoch.

R-3.4.2 for Windows (32/64 bit)

[Download R 3.4.2 for Windows \(75 megabytes, 32/64 bit\)](#) ← “CLICK ON”

[Installation and other instructions](#)
[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is
<<CRAN MIRROR>/bin/windows/base/release.htm>.

Last change: 2017-09-28, by Duncan Murdoch

Installing R studio

<https://www.rstudio.com>

Installing R studio

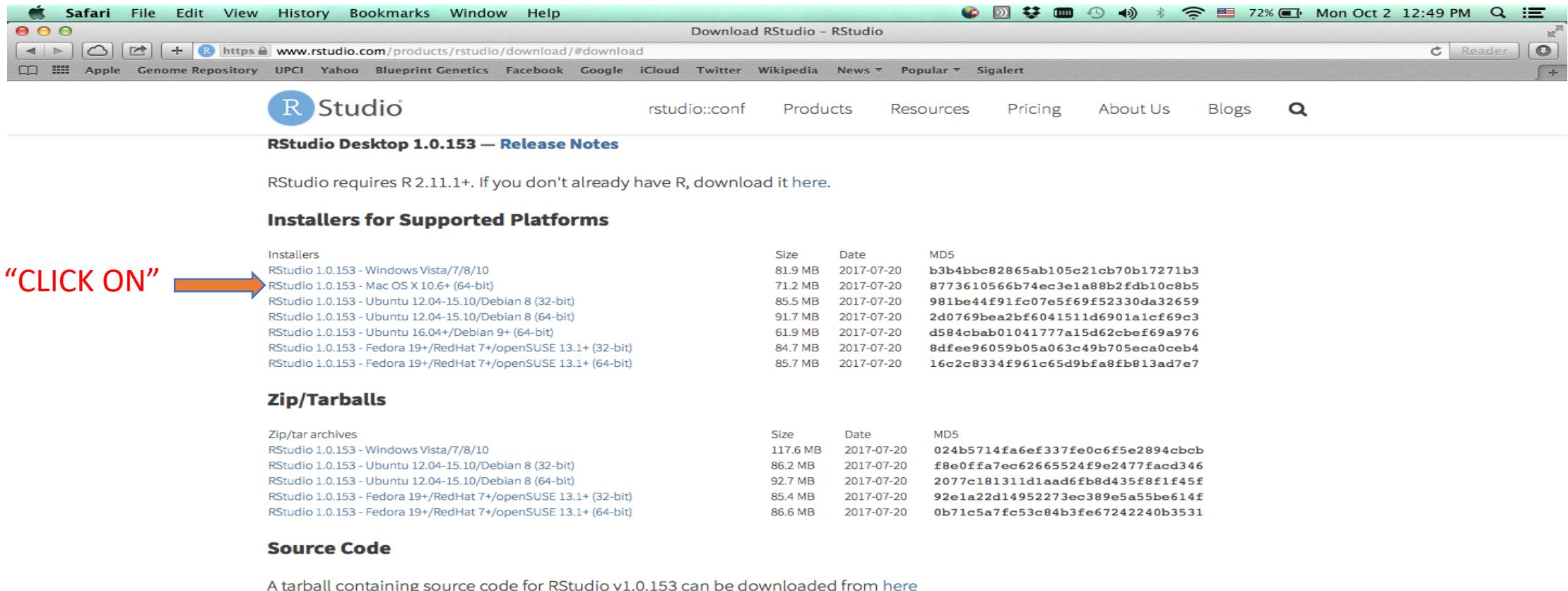
<https://www.rstudio.com/products/rstudio/download/>

The screenshot shows the RStudio download page on a Mac OS X system. The browser window title is "Download RStudio - RStudio". The page features the RStudio logo and navigation links for "rstudio::conf", "Products", "Resources", "Pricing", "About Us", and "Blogs". A search bar is also present. The main content area is titled "Choose Your Version of RStudio" and includes a screenshot of the RStudio IDE interface. Below this, five license options are listed:

Version	License	Price
RStudio Desktop	Open Source License	FREE
RStudio Desktop	Commercial License	\$995 per year
RStudio Server	Open Source License	FREE
RStudio Server Pro	Commercial License	\$9,995 per year
RStudio Server Pro + RStudio Connect	Commercial License	\$29,995 per year

Each row contains a "DOWNLOAD" button, a "Learn More" link, and a small green dot icon below it. A red arrow points to the "DOWNLOAD" button for the RStudio Desktop (Open Source License) row. The text "Integrated Tools for R" is visible at the bottom left.

Installing R studio



The screenshot shows a Mac OS X desktop with a Safari browser window open to the RStudio download page. The title bar reads "Download RStudio - RStudio". The address bar shows the URL "https://www.rstudio.com/products/rstudio/download/#download". The main content area displays the "RStudio Desktop 1.0.153 — Release Notes" page. A red arrow points to the "Installers for Supported Platforms" section, which lists various operating system and distribution combinations with their file sizes, dates, and MD5 checksums.

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.0.153 - Windows Vista/7/8/10	81.9 MB	2017-07-20	b3b4bbc82865ab105c21cb70b17271b3
RStudio 1.0.153 - Mac OS X 10.6+ (64-bit)	71.2 MB	2017-07-20	8773610566b74ec3e1a88b2fdb10c8b5
RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	85.5 MB	2017-07-20	981be44f91fc07e5f69f52330da32659
RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	91.7 MB	2017-07-20	2d0769bea2bf6041511d6901a1cf69c3
RStudio 1.0.153 - Ubuntu 16.04+/Debian 9+ (64-bit)	61.9 MB	2017-07-20	d584cbab01041777a15d62cbef69a976
RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	84.7 MB	2017-07-20	8dfee96059b05a063c49b705eca0ceb4
RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	85.7 MB	2017-07-20	16c2c8334f961c65d9bfa8fb813ad7e7

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.0.153 - Windows Vista/7/8/10	117.6 MB	2017-07-20	024b5714fa6ef337fe0c6f5e2894cbcb
RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	86.2 MB	2017-07-20	f8e0ffa7ec62665524f9e2477facd346
RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	92.7 MB	2017-07-20	2077c181311d1aad6fb8d435f8f1f45f
RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	85.4 MB	2017-07-20	92ela22d14952273ec389e5a55be614f
RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	86.6 MB	2017-07-20	0b71c5a7fc53c84b3fe67242240b3531

Source Code

A tarball containing source code for RStudio v1.0.153 can be downloaded from [here](#)

Installing R studio

Introduction to Bioconductor

<https://bioconductor.org>

Introduction to Bioconductor

- Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data.
- Bioconductor uses the R statistical programming language, and is open source and open development.
- It has two releases each year, and an active user community.
- Bioconductor is also available as an AMI (Amazon Machine Image) and a series of Docker images.

The screenshot displays the official Bioconductor website at bioconductor.org. The top navigation bar includes links for Home, Install, Help, Developers, and About. A search bar is also present. The main content area features sections for "EuroBioC 2018", "Install", "About Bioconductor", "Use", and "Develop". Below these are "Support" and "Events" sections. The "Support" section shows recent tweets from Bioconductor users, while the "Events" section lists BioCAsia 2018 and EuroBioC 2018. The footer contains copyright information and links to various Bioconductor resources.

Introduction to Bioconductor

The screenshot shows the official Bioconductor website. At the top, there is a navigation bar with links for Home, Install, Help, Developers, and About. A search bar is also present. The main content area features several sections: "EuroBioC 2018" (with details about the meeting), "About Bioconductor" (with a brief overview and links to R, AMI, and Docker), "News" (with bullet points about software releases and opportunities), and four main functional sections: "Install", "Learn", "Use", and "Develop". Each section contains a brief description and a list of related links.

EuroBioC 2018

Join us at the [European Bioconductor meeting](#) on December 6 and 7, 2018, at the Technical University of Munich, Germany.

About Bioconductor

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, and an active user community. Bioconductor is also available as an [AMI](#) (Amazon Machine Image) and a series of [Docker](#) images.

News

- Bioconductor [3.8](#) is available.
- Core team job opportunities for [scientific programmer / analyst](#) and [senior programmer / analyst!](#)
- Bioconductor [F1000 Research Channel](#) available.
- Orchestrating high-throughput genomic analysis with [Bioconductor \(abstract\)](#) and other [recent literature](#).

Install »

- Discover [1649 software packages](#) available in *Bioconductor* release 3.8.

Get started with *Bioconductor*

- [Install *Bioconductor*](#)
- [Get support](#)
- [Latest newsletter](#)
- [Follow us on twitter](#)
- [Install R](#)

Learn »

Master *Bioconductor* tools

- [Courses](#)
- [Support site](#)
- [Package vignettes](#)
- [Literature citations](#)
- [Common work flows](#)
- [FAQ](#)
- [Community resources](#)
- [Videos](#)

Use »

Create bioinformatic solutions with *Bioconductor*

- [Software](#), [Annotation](#), and [Experiment](#) packages
- [Amazon Machine Image](#)
- [Latest release announcement](#)
- [Community Slack](#)
- [Support site](#)

Develop »

Contribute to *Bioconductor*

- [Developer resources](#)
- [Use Bioc 'devel'](#)
- ['Devel' packages](#)
- [Package guidelines](#)
- [New package submission](#)
- [Git source control](#)
- [Build reports](#)

Installing Bioconductor Packages

<https://bioconductor.org>

Bioconductor - Exercise

- Retrieve list of genes on Chromosome 1 from ENSEMBL
 - Use R package “biomaRt”
 - Source Ensembl
 - Dataset: hsapiens
 - Required Output:
 - Gene symbols on chromosome 1

Bioconductor

Bioconductor - Exercise

<https://bioconductor.org/help/search/index.html?q=biomaRt/>



Home Install Help Developers About

Search: biomaRt

[Home](#) » [Help](#) » [Search](#)



Search

Your search for *biomaRt* returned 5286 results.

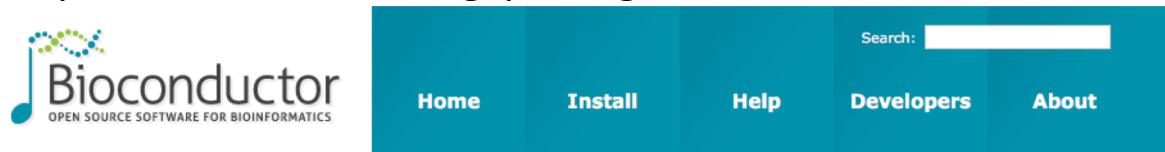
[Bioconductor - biomaRt](#) - /packages/release/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[biomaRt](#) - /packages/bioc/1.6/src/contrib/html/biomaRt.html
[biomaRt](#)
[biomaRt](#) - /packages/2.3/bioc/html/biomaRt.html
[biomaRt](#)
[biomaRt](#) - /packages/2.4/bioc/html/biomaRt.html
[biomaRt](#)
[biomaRt](#) - /packages/bioc/1.7/src/contrib/html/biomaRt.html
[biomaRt](#)
[Bioconductor - biomaRt \(development version\)](#) - /packages-devel/bioc/html/biomaRt.html
[Bioconductor - biomaRt \(development version\)](#)
[Bioconductor - biomaRt](#) - /packages/2.7/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.5/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.6/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.10/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.13/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.12/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.11/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/3.0/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.14/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/3.2/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/3.1/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.8/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/2.9/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)
[Bioconductor - biomaRt](#) - /packages/3.4/bioc/html/biomaRt.html
[Bioconductor - biomaRt](#)

[Next >](#)

Bioconductor

Bioconductor - Exercise

<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>



Home » Bioconductor 3.6 » Software Packages » biomaRt

biomaRt

platforms all downloads top 5% posts 31 / 1 / 2 / 11 in Bioc > 12.5 years

build ok

DOI: [10.18129/B9.bioc.biomaRt](https://doi.org/10.18129/B9.bioc.biomaRt) [f](#) [t](#)

Interface to BioMart databases (e.g. Ensembl, COSMIC, Wormbase and Gramene)

Bioconductor version: Release (3.6)

In recent years a wealth of biological data has become available in public data repositories. Easy access to these valuable data resources and firm integration with data analysis is needed for comprehensive bioinformatics data analysis. biomaRt provides an interface to a growing collection of databases implementing the BioMart software suite (<http://www.biomart.org>). The package enables retrieval of large amounts of data in a uniform way without the need to know the underlying database schemas or write complex SQL queries. Examples of BioMart databases are Ensembl, COSMIC, Uniprot, HGNC, Gramene, Wormbase and dbSNP mapped to Ensembl. These major databases give biomaRt users direct access to a diverse set of data and enable a wide range of powerful online queries from gene annotation to database mining.

Author: Steffen Durinck [aut], Wolfgang Huber [aut], Sean Davis [ctb], Francois Pepin [ctb], Vince S Buffalo [ctb], Mike Smith [ctb, cre]

Maintainer: Mike Smith <grimbough@gmail.com>

Citation (from within R, enter `citation("biomaRt")`):

Durinck S, Spellman P, Birney E and Huber W (2009). "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt." *Nature Protocols*, **4**, pp. 1184–1191.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A and Huber W (2005). "BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis." *Bioinformatics*, **21**, pp. 3439–3440.

Installation

To install this package, start R and enter:

```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("biomaRt")
```

Documentation

Documentation »

Bioconductor

- Package [vignettes](#) and manuals.
- [Workflows](#) for learning and use.
- [Course and conference](#) material.
- [Videos](#).
- Community [resources](#) and [tutorials](#).

R / [CRAN](#) packages and [documentation](#)

Support »

Please read the [posting guide](#). Post questions about Bioconductor to one of the following locations:

- [Support site](#) - for questions about Bioconductor packages
- [Bioc-devel](#) mailing list - for package developers

COPY the text

Bioconductor - Exercise

The screenshot shows the RStudio interface with the following components:

- Console Tab:** Displays the R code and output for installing Bioconductor packages. A blue arrow points from the text "PASTE in RStudio" to the first line of code: `> ## try http:// if https:// URLs are not supported`.
- Environment Tab:** Shows the Global Environment pane with the message "Environment is empty".
- Packages Tab:** Shows the System Library section of the Packages pane, listing various Bioconductor packages and their versions.

Code and Output (Console Tab):

```
> ## try http:// if https:// URLs are not supported
> source("https://bioconductor.org/biocLite.R")
trying URL 'https://bioconductor.org/packages/3.6/bioc/bin/macosx/el-capitan/contrib/3.4/BiocInstaller_1.28.0.tgz'
Content type 'application/x-gzip' length 57804 bytes (56 KB)
=====
downloaded 56 KB

The downloaded binary packages are in
  /var/folders/s2/6ymbdlgn1_xfkqmccsykms4r0000gn/T//RtmpvwuQk0/downloaded_packages
Bioconductor version 3.6 (BiocInstaller 1.28.0), ?biocLite for help
> biocLite("biomaRt")
BioC_mirror: https://bioconductor.org
Using Bioconductor 3.6 (BiocInstaller 1.28.0), R 3.4.2 (2017-09-28).
Installing package(s) 'biomaRt'
also installing the dependencies 'bit', 'bit64', 'blob', 'memoise', 'BiocGenerics', 'Biobase', 'IRanges', 'DBI', 'RSSQLite', 'S4Vectors', 'prettyunits', 'AnnotationDbi', 'progress'

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/bit_1.1-12.tgz'
Content type 'application/x-gzip' length 240695 bytes (235 KB)
=====
downloaded 235 KB

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/bit64_0.9-7.tgz'
Content type 'application/x-gzip' length 534016 bytes (521 KB)
=====
downloaded 521 KB

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/blob_1.1.0.tgz'
Content type 'application/x-gzip' length 12359 bytes (12 KB)
=====
downloaded 12 KB

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/memoise_1.1.0.tgz'
Content type 'application/x-gzip' length 27154 bytes (26 KB)
=====
downloaded 26 KB

trying URL 'https://bioconductor.org/packages/3.6/bioc/bin/macosx/el-capitan/contrib/3.4/BiocGenerics_0.24.0.tgz'
Content type 'application/x-gzip' length 571870 bytes (558 KB)
```

Packages Tab (System Library):

Name	Description	Version
assertthat	Easy Pre and Post Assertions	0.2.0
BH	Boost C++ Header Files	1.65.0-1
bindr	Parametrized Active Bindings	0.1
bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2
<input checked="" type="checkbox"/> BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.28.0
bitops	Bitwise Operations	1.0-6
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-20
class	Functions for Classification	7.3-14
cli	Helpers for Developing Command Line Interfaces	1.0.0
cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.6
codetools	Code Analysis Tools for R	0.2-15
compiler	The R Compiler Package	3.4.2
crayon	Colored Terminal Output	1.3.4
curl	A Modern and Flexible Web Client for R	3.1
data.table	Extension of 'data.frame'	1.10.4-3
<input checked="" type="checkbox"/> datasets	The R Datasets Package	3.4.2
digest	Create Compact Hash Digests of R Objects	0.6.14
downloader	Download Files over HTTP and HTTPS	0.4
dplyr	A Grammar of Data Manipulation	0.7.4

Bioconductor - Exercise

R version 3.4.2 (2017-09-28) -- "Short Summer"
Copyright (C) 2017 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.

```
> library(biomaRt)
```

Environment is empty

Name	Description	Version
AnnotationDbi	Annotation Database Interface	1.40.0
assertthat	Easy Pre and Post Assertions	0.2.0
BH	Boost C++ Header Files	1.65.0-1
bindr	Parametrized Active Bindings	0.1
bindrcpp	An 'Rcpp' Interface to Active Bindings	0.2
Biobase	Biobase: Base functions for Bioconductor	2.38.0
BiocGenerics	S4 generic functions for Bioconductor	0.24.0
BiocInstaller	Install/Update Bioconductor, CRAN, and github Packages	1.28.0
biomaRt	Interface to BioMart databases (e.g. Ensembl, COSMIC, Wormbase and Gramene)	2.34.2
bit	A class for vectors of 1-bit booleans	1.1-12
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')	1.1.0
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-20
class	Functions for Classification	7.3-14
cli	Helpers for Developing Command Line Interfaces	1.0.0
cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.6
codetools	Code Analysis Tools for R	0.2-15

Type:
library(biomaRt)

Bioconductor - Exercise

Type in R studio:

- library(biomaRt)
 - getwd()
 - setwd("/Users/enriquevelazquez/Documents/R_working_directory")
-
- listMarts()
 - ensembl=useMart("ensembl")
 - listDatasets(ensembl)

Output

Bioconductor - Exercise

> listDatasets(ensembl)	dataset	description	version
1	csabaeus_gene_ensembl	Vervet-AGM genes (Chlsab1.1)	Chlsab1.1
2	odegus_gene_ensembl	Degu genes (Octdeg1.0)	Octdeg1.0
3	ggallus_gene_ensembl	Chicken genes (Gallus_gallus-5.0)	Gallus_gallus-5.0
4	mfascicularis_gene_ensembl	Crab-eating macaque genes (Macaca_fascicularis-5.0)	Macaca_fascicularis-5.0
5	ptroglobutes_gene_ensembl	Chimpanzee genes (Pan_tro_3.0)	Pan_tro_3.0
6	amexicanus_gene_ensembl	Cave fish genes (AstMex102)	AstMex102
7	ccapucinus_gene_ensembl	Capuchin genes (Cebus_imitator-1.0)	Cebus_imitator-1.0
8	xmaculatus_gene_ensembl	Platyfish genes (Xipmac4.4.2)	Xipmac4.4.2
9	csavignyi_gene_ensembl	C.savignyi genes (CSAV 2.0)	CSAV 2.0
10	fabicollis_gene_ensembl	Flycatcher genes (FicAlb1.4)	FicAlb1.4
11	rnorvegicus_gene_ensembl	Rat genes (Rnor_6.0)	Rnor_6.0
12	psinensis_gene_ensembl	Chinese softshell turtle genes (Pelsin_1.0)	Pelsin_1.0
13	capercaillie_gene_ensembl	Brazilian guinea pig genes (CavAp1.0)	CavAp1.0
14	cporcellus_gene_ensembl	Guinea Pig genes (Cavpor3.0)	Cavpor3.0
15	jjaculus_gene_ensembl	Lesser Egyptian jerboa genes (JacJac1.0)	JacJac1.0
16	trubripes_gene_ensembl	Fugu genes (FUGU 4.0)	FUGU 4.0
17	pabelii_gene_ensembl	Orangutan genes (PPYG2)	PPYG2
18	ppaniscus_gene_ensembl	Bonobo genes (panpan1.1)	panpan1.1
19	cfamiliaris_gene_ensembl	Dog genes (CanFam3.1)	CanFam3.1
20	pmarinus_gene_ensembl	Lamprey genes (Pmarinus_7.0)	Pmarinus_7.0
21	cintestinalis_gene_ensembl	C.intestinalis genes (KH)	KH
22	saraneus_gene_ensembl	Shrew genes (sorAral1)	sorAral1
23	gmorhua_gene_ensembl	Cod genes (gadMor1)	gadMor1
24	oanatinus_gene_ensembl	Platypus genes (OANAS)	OANAS
25	csyrichta_gene_ensembl	Tarsier genes (Tarsius_syrichta-2.0.1)	Tarsius_syrichta-2.0.1
26	hmale_gene_ensembl	Naked mole-rat male genes (HetGla_1.0)	HetGla_1.0
27	mspreteij_gene_ensembl	Mouse SPRET/EIJ genes (SPRET_EIJ_v1)	SPRET_EIJ_v1
28	anancymmae_gene_ensembl	Ma's night monkey genes (Anan_2.0)	Anan_2.0
29	driero_gene_ensembl	Zebrafish genes (GRCz10)	GRCz10
30	rroxellana_gene_ensembl	Golden snub-nosed monkey genes (Rrox_v1)	Rrox_v1
31	mnemestrina_gene_ensembl	Pig-tailed macaque genes (Mmem_1.0)	Mmem_1.0
32	ccrigri_gene_ensembl	Chinese hamster CriGrI genes (CriGrI_1.0)	CriGrI_1.0
33	pcoquereli_gene_ensembl	Coquerel's sifaka genes (Pcoq_1.0)	Pcoq_1.0
34	hsapiens_gene_ensembl	Human genes (GRCh38.p10)	GRCh38.p10
35	sboliviensis_gene_ensembl	Bolivian squirrel monkey genes (SalBo1.0)	SalBo1.0
36	mlucifugus_gene_ensembl	Microbat genes (Myoluc2.0)	Myoluc2.0
37	ttruncatus_gene_ensembl	Dolphin genes (turTru1)	turTru1
38	lchalumnae_gene_ensembl	Coelacanth genes (LatChal1)	LatChal1
39	fdomarensis_gene_ensembl	Damara mole rat genes (DMR_v1.0)	DMR_v1.0
40	mleucophaeus_gene_ensembl	Drill genes (Mieu_le_1.0)	Mieu_le_1.0
41	etelfairi_gene_ensembl	Lesser hedgehog tenrec genes (TENREC)	TENREC
42	dmelanogaster_gene_ensembl	Fruitfly genes (BDGP6)	BDGP6
43	ecabolus_gene_ensembl	Horse genes (Equ_Cab_2)	Equ_Cab_2
44	sharrisii_gene_ensembl	Tasmanian devil genes (Devil_ref v7.0)	Devil_ref v7.0
45	mpohari_gene_ensembl	Shrew mouse genes (PAHARI_EIJ_v1.1)	PAHARI_EIJ_v1.1
46	hfemale_gene_ensembl	Naked mole-rat female genes (HetGla_female_1.0)	HetGla_female_1.0
47	mgallopavo_gene_ensembl	Turkey genes (Turkey_2.01)	Turkey_2.01
48	ocarolinensis_gene_ensembl	Anole lizard genes (AnoCar2.0)	AnoCar2.0
49	neugensi_gene_ensembl	Wallaby genes (Meug_1.0)	Meug_1.0
50	oniloticus_gene_ensembl	Tilapia genes (Orenil1.0)	Orenil1.0
51	btaurus_gene_ensembl	Cow genes (UMD3.1)	UMD3.1
52	mdomestica_gene_ensembl	Opossum genes (monDom5)	monDom5
53	dnomencinctus_gene_ensembl	Armadillo genes (Dasnov3.0)	Dasnov3.0
54	pformosa_gene_ensembl	Amazon Molly genes (Poecilia_formosa-5.1.2)	Poecilia_formosa-5.1.2
55	cjacchus_gene_ensembl	Marmoset genes (C_jacchus3.2.1)	C_jacchus3.2.1
56	panibus_gene_ensembl	Olive baboon genes (Panu_3.0)	Panu_3.0
57	olatipes_gene_ensembl	Medaka genes (HdrR)	HdrR
58	oprinceps_gene_ensembl	Pika genes (OchPrI2.0-Ens)	OchPrI2.0-Ens
59	amelanoleuca_gene_ensembl	Panda genes (ailMeli)	ailMeli
60	catys_gene_ensembl	Sooty mangabey genes (Caty_1.0)	Caty_1.0
61	mmusculus_gene_ensembl	Mouse genes (GRCh38.p5)	GRCh38.p5
62	tbelangeri_gene_ensembl	Tree Shrew genes (tupBell1)	tupBell1
63	tnigroviridis_gene_ensembl	Tetraodon genes (TETRAODON 8.0)	TETRAODON 8.0
64	tguttata_gene_ensembl	Zebra Finch genes (taeGut3.2.4)	taeGut3.2.4
65	nleucogenys_gene_ensembl	Gibbon genes (Nleu_3.0)	Nleu_3.0
66	mmurinus_gene_ensembl	Mouse Lemur genes (Mmur_3.0)	Mmur_3.0
67	rbietai_gene_ensembl	Black snub-nosed monkey genes (ASM169854v1)	ASM169854v1
68	pvampyrus_gene_ensembl	Megabat genes (pteVam1)	pteVam1
69	ggorilla_gene_ensembl	Gorilla genes (gorGor4)	gorGor4
70	cpalliatis_gene_ensembl	Angola colobus genes (Cong_pa_1.0)	Cong_pa_1.0
71	mcaroli_gene_ensembl	Ryukyu mouse genes (CAROLI_EIJ_v1.1)	CAROLI_EIJ_v1.1
72	xtropicalis_gene_ensembl	Xenopus genes (JGI 4.2)	JGI 4.2
73	oaries_gene_ensembl	Sheep genes (Oar_v3.1)	Oar_v3.1
74	locularius_gene_ensembl	Spotted gar genes (LepOcu1)	LepOcu1
75	itridescemlineatus_gene_ensembl	Squid genes (SpeTr12.0)	SpeTr12.0
76	gaculeatus_gene_ensembl	Stickleback genes (BROAD S1)	BROAD S1
77	dordii_gene_ensembl	Kangaroo rat genes (Dord_2.0)	Dord_2.0
78	ognarnettii_gene_ensembl	Rushlhahy genes (Otnogn3)	Otnogn3

Bioconductor - Exercise

Type in R studio

- `ensembl_mart <- useMart("ensembl", dataset = "hsapiens_gene_ensembl")`
- `ensembl_mart`

- `listAttributes(ensembl_mart)`
- `listFilters(ensembl_mart)`

Bioconductor - Exercise

Output:

```
> listAttributes(ensembl_mart)
  name
1  ensembl_gene_id
2  ensembl_gene_id_version
3  ensembl_transcript_id
4  ensembl_transcript_id_version
5    ensembl_peptide_id
6    ensembl_peptide_id_version
7    ensembl_exon_id
8      description
9      chromosome_name
10     start_position
11     end_position
12     strand
13     band
14     transcript_start
15     transcript_end
16     transcription_start_site
17     transcript_length
18     transcript tsl
19     transcript_gencode_basic
20     transcript_appris
21     external_gene_name
22     external_gene_source
23     external_transcript_name
24     external_transcript_source_name
25     transcript_count
26     percentage_gene_gc_content
27     gene_biotype
28     transcript_biotype
29     source
30     transcript_source
31     version
32     transcript_version
33     phenotype_description
34     source_name
35     study_external_id
36     strain_name
37     strain_gender
38     p_value
39     go_id
40     name_1006
41     definition_1006
42     go_linkage_type
43     namespace_1003
44     gosim_goa_accession
45     gosim_goa_description
46     ccds
47     chembl
48     clone_based_ensembl_gene
49     clone_based_ensembl_transcript
50       dbass3_name
51       dbass3_id
52       dbass5_name
53       dbass5_id
54     ens_hs_transcript
55     ens_hs_translation
56     entrezgene_trans_name
57     embl
58     arrayexpress
59     genedb
60     hgnc_id
61     hgnc_symbol
62     hgnc_trans_name
63     hpa
64     protein_id
65     kegg_enzyme
66     ens_lrg_gene
67     ens_lrg_transcript
68     merops
69     metacyc
70     mim_gene_description
71     mim_gene_accession
72     mim_morbid_description
73     mim_morbid_accession
74     mirbase_accession
75     mirbase_id
76     mirbase_trans_name
77     entrezgene
78     pdb
```

Bioconductor - Exercise

Output:

```
> listFilters(ensembl Mart)
      name                                     description
1  chromosome_name                         Chromosome/scaffold name
2          start                           Start
3          end                            End
4      band_start                         Band Start
5      band_end                          Band End
6 marker_start                         Marker Start
7 marker_end                          Marker End
8 encode_region                        Encode region
9      strand                           Strand
10 chromosomal_region                   e.g. 1:10000:-1, 1:100000:200000:1
11      with_ccds                         With CCDS ID(s)
12      with_chembl                        With ChEMBL ID(s)
13      with_clone_based_ensembl_gene      With Clone-based (Ensembl) gene ID(s)
14      with_clone_based_ensembl_transcript  With Clone-based (Ensembl) transcript ID(s)
15      with_dbass3                         With DataBase of Aberrant 3' Splice Sites ID(s)
16      with_dbass5                         With DataBase of Aberrant 5' Splice Sites ID(s)
17      with_ens_hs_transcript                With Ensembl Human Transcript ID(s)
18      with_ens_hs_translation               With Ensembl Human Translation ID(s)
19      with_entrezgene_trans_name          With EntrezGene transcript name ID(s)
20      with_embl                           With European Nucleotide Archive ID(s)
21      with_arrayexpress                  With Expression Atlas ID(s)
22      with_genedb                         With GeneDB ID(s)
23      with_go                             With GO ID(s)
24      with_goslim_goa                     With GOSlim GOA ID(s)
25      with_hgnc                           With HGNC Symbol ID(s)
26      with_hgnc_trans_name                With HGNC transcript name ID(s)
27      with_hpa                            With Human Protein Atlas ID(s)
28      with_protein_id                    With INSDC protein ID ID(s)
29      with_kegg_enzyme                   With KEGG Pathway and Enzyme ID(s)
30      with_ens_lrg_gene                   With LRG display in Ensembl gene ID(s)
31      with_ens_lrg_transcript              With LRG display in Ensembl transcript ID(s)
32      with_merops                          With MEROPS - the Peptidase Database ID(s)
33      with_metacyc                        With MetaCyc ID(s)
34      with_mim_gene                       With MIM gene ID(s)
35      with_mim_morbid                     With MIM morbid ID(s)
36      with_miRbase                        With miRBase ID(s)
37      with_miRbase_trans_name             With miRBase transcript name ID(s)
38      with_entrezgene                     With NCBI gene ID(s)
39      with_pdb                            With PDB ID(s)
40      with_reactome                       With Reactome ID(s)
41      with_reactome_gene                 With Reactome gene ID(s)
42      with_reactome_transcript            With Reactome transcript ID(s)
43      with_refseq_mrna_predicted         With RefSeq mRNA predicted ID(s)
44      with_refseq_ncrna_predicted        With RefSeq ncRNA predicted ID(s)
45      with_refseq_ncrna                 With RefSeq ncRNA ID(s)
46      with_refseq_peptide_predicted      With RefSeq peptide predicted ID(s)
47      with_rfam                           With RFAM ID(s)
48      with_rfam_trans_name               With RFAM transcript name ID(s)
49      with_rnacentral                     With RNACentral ID(s)
50      with_ucsc                           With UCSC Stable ID ID(s)
51      with_unigene                        With UniGene ID(s)
52      with_uniparc                        With UniParc ID(s)
53      with_uniprot_gn                     With UniProtKB Gene Name ID(s)
54      with_uniprotswissprot              With UniProtKB/Swiss-Prot ID(s)
55      with_uniprotstrembl               With UniProtKB/TrEMBL ID(s)
56      with_wikigene                       With WikiGene ID(s)
57      ensembl_gene_id                   Gene stable ID(s) [e.g. ENSG000000000003]
58      ensembl_gene_id_version           Gene stable ID(s) with version [e.g. ENSG000000000003.14]
59      ensembl_transcript_id             Transcript stable ID(s) [e.g. ENST000000000233]
60      ensembl_transcript_id_version     Transcript stable ID(s) with version [e.g. ENST000000000233.9]
61      ensembl_peptide_id               Protein stable ID(s) [e.g. ENSP000000000233]
62      ensembl_peptide_id_version       Protein stable ID(s) with version [e.g. ENSP000000000233.5]
63      ensembl_exon_id                  Exon ID(s) [e.g. ENSE00000327880]
64      external_gene_name               Gene Name(s) [e.g. MIR628]
65      external_transcript_name         Transcript Name(s) [e.g. MIR628-201]
66      ccds                            CCDS ID(s) [e.g. CCDS10]
67      chembl                           ChEMBL ID(s) [e.g. CHEMBL1075092]
68      clone_based_ensembl_gene         Clone-based (Ensembl) gene ID(s) [e.g. ABO15752.1]
69      clone_based_ensembl_transcript   Clone-based (Ensembl) transcript ID(s) [e.g. AB015752.1-201]
70      dbass3_name                      DataBase of Aberrant 3' Splice Sites name(s) [e.g. ABC8]
71      dbass3_id                         DataBase of Aberrant 3' Splice Sites ID(s) [e.g. 1]
72      dbass5_name                      DataBase of Aberrant 5' Splice Sites name(s) [e.g. ABCD1]
73      dbass5_id                         DataBase of Aberrant 5' Splice Sites ID(s) [e.g. 1]
74      ens_hs_transcript                 Ensembl Human Transcript ID(s) [e.g. ENST000000075430]
75      ens_hs_translation                Ensembl Human Translation ID(s) [e.g. ENSP000000075430]
```

Bioconductor - Exercise

Constructing a Query on R studio (bio)

- chromosome1 <- getBM(attributes = ("hgnc_symbol"), filters = "chromosome_name", values = "1", mart = ensembl_mart)
- chromosome1

Bioconductor - Exercise

Output:

```
> chromosome1
1      DDX11L1
2      WASH7P
3      MIR6859-1
4      MIR1302-2HG
5      MIR1302-2
6      FAM138A
7      OR4G4P
8      OR4G11P
9      OR4F5
10     CICP27
11     RNU6-1100P
12     MIR6859-2
13     RPL23AP24
14     WBP1LP7
15     OR4F29
16     CICP7
17     MTND1P23
18     MTND2P28
19     MTCO1P12
20     MIR6723
21     MTCO2P12
22     MTATP8P1
23     MTATP6P1
24     MTCO3P12
25     WBP1LP6
26     OR4F16
27     CICP3
28     RNU6-1199P
29     FAM87B
30     LINC01128
31     LINC00115
32     FAM41C
33     TUBB8P11
34     SAMD11
35     NOC2L
36     KLHL17
37     PLEXHN1
38     PERM1
39     HES4
40     ISG15
41     AGRN
42     RNF223
43     C1orf159
44     LINC01342
45     MIR200B
46     MIR200A
47     MIR429
48     TTLL10-AS1
49     TTLL10
50     TNFRSF18
51     TNFRSF4
52     SDF4
53     B3GALT6
54     C10TNF12
55     UBE2J2
56     LINC01786
57     SCN11D
58     ACAP3
59     MIR6726
60     PUSL1
61     INTS11
62     MIR6727
63     CPTP
64     TAS1R3
65     DVL1
66     MIR6808
67     MXRA8
68     AURKAIP1
69     NDUFB4P8
70     CCNL2
71     MRPL20
72     RN7SL657P
73     ANKRD65
74     TMEM88B
75     LINC01770
76     VWA1
77     ATAD3C
78     ATAD3B
```

Bioconductor - Exercise

- `chromosome1 <- getBM(attributes = c("hgnc_symbol", "ensembl_gene_id"), filters = "chromosome_name", values = "1", mart = ensembl_mart)`
- `chromosome1`

Bioconductor - Exercise

Output:

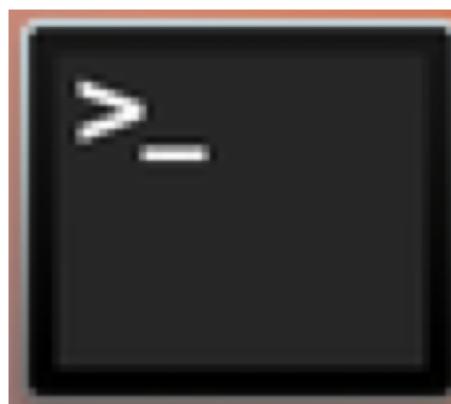
```
> chromosome1 <- getBM(attributes = c("hgnc_symbol", "ensembl_gene_id"), filters = "chromosome_name", values = "1", mart = ensembl_mart)
> chromosome1
  hgnc_symbol ensembl_gene_id
 1          DDX11L1 ENSG00000223972
 2          WASH7P ENSG00000227232
 3          MIR6859-1 ENSG00000278267
 4          MIR1302-2HNG ENSG00000243485
 5          MIR1302-2 ENSG00000284332
 6          FAM138A ENSG00000237613
 7          OR4G4P ENSG00000268020
 8          OR4G11P ENSG00000240361
 9          OR4F5 ENSG00000186092
10          ENSG00000238009
11          ENSG00000239945
12          C1CP27 ENSG00000233750
13          ENSG00000268903
14          ENSG00000269981
15          ENSG00000239906
16          ENSG00000241860
17          RNU6-1100P ENSG00000222623
18          ENSG00000241599
19          ENSG00000279928
20          ENSG00000279457
21          MIR6859-2 ENSG00000273874
22          ENSG00000228463
23          RPL23AP24 ENSG00000236679
24          ENSG00000236601
25          ENSG00000237094
26          WBP1LP7 ENSG00000269732
27          OR4F29 ENSG00000284733
28          C1CP7 ENSG00000233653
29          ENSG00000250575
30          ENSG00000278757
31          ENSG00000230021
32          ENSG00000235146
33          MTND1P23 ENSG00000225972
34          MTND2P28 ENSG00000225630
35          MTCO1P12 ENSG00000237973
36          MTR6723 ENSG00000278791
37          MTCO2P12 ENSG00000229344
38          MTATP8P1 ENSG00000240409
39          MTATP6P1 ENSG00000248527
40          MTCO3P12 ENSG00000198744
41          WBP1LP6 ENSG00000268663
42          OR4F16 ENSG00000284662
43          C1CP3 ENSG00000229376
44          ENSG00000228327
45          RNU6-1199P ENSG00000223181
46          ENSG00000229905
47          ENSG00000237491
48          ENSG00000230092
49          FAM87B ENSG00000177757
50          LINC01128 ENSG00000228794
51          LINC00115 ENSG00000225880
52          FAM41C ENSG00000230368
53          TUBB8P11 ENSG00000234711
54          ENSG00000283040
55          ENSG00000272438
56          ENSG00000230699
57          ENSG00000241180
58          ENSG00000223764
59          SAMD11 ENSG00000187634
60          NOCL ENSG00000188976
61          KLHL17 ENSG00000187961
62          PLEKH1 ENSG00000187583
63          PERM1 ENSG00000187642
64          ENSG00000272512
65          HES4 ENSG00000188290
66          ISG15 ENSG00000187608
67          ENSG00000231702
68          ENSG00000224969
69          AGRN ENSG00000188157
70          ENSG00000242590
71          ENSG00000217801
72          ENSG000002273443
73          RNFL223 ENSG00000237330
74          C1orf159 ENSG00000131591
75          LINC01342 ENSG00000223823
76          MIR200B ENSG00000207730
77          MIR200A ENSG00000207607
```

Introduction to Terminal

<https://github.com/0nn0/terminal-mac-cheatsheet>

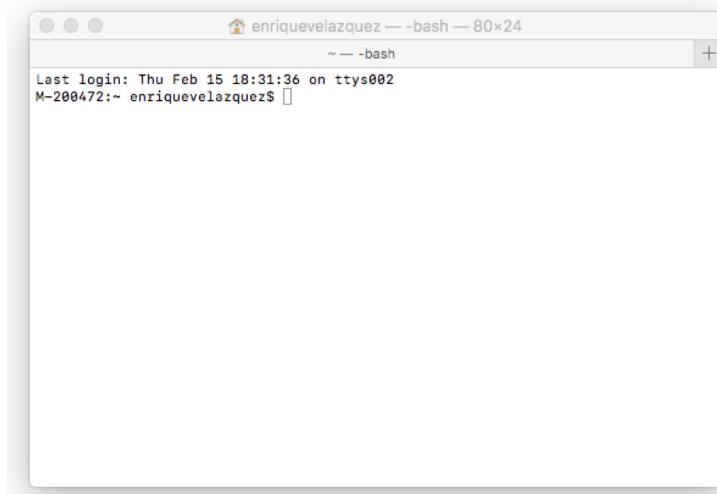
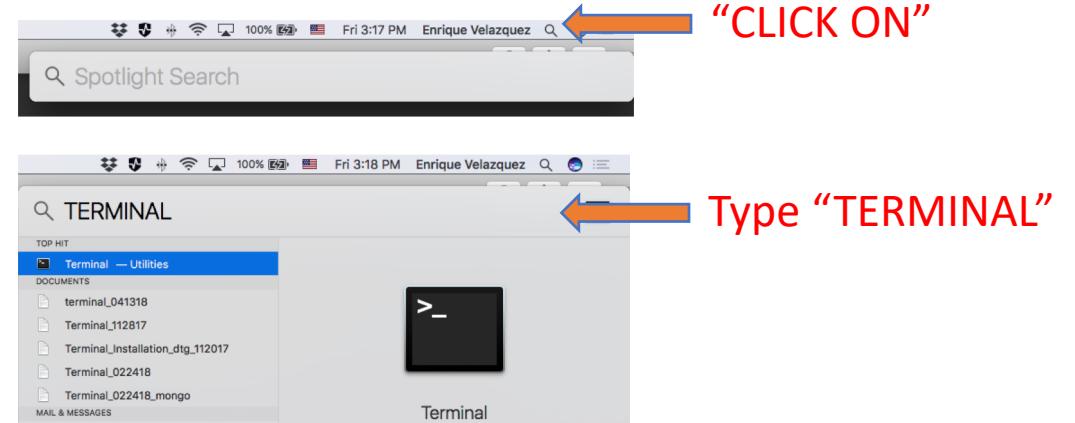
Introduction to Terminal

- Before graphical user interfaces (GUIs) became, the only way to use a computer was through a command line interface.
- Commands can be simple word or words entries (such as the “date command) or more complex.
- Most commands accept arguments.
- The general format for commands is:
 - Command option(s) filename(s)



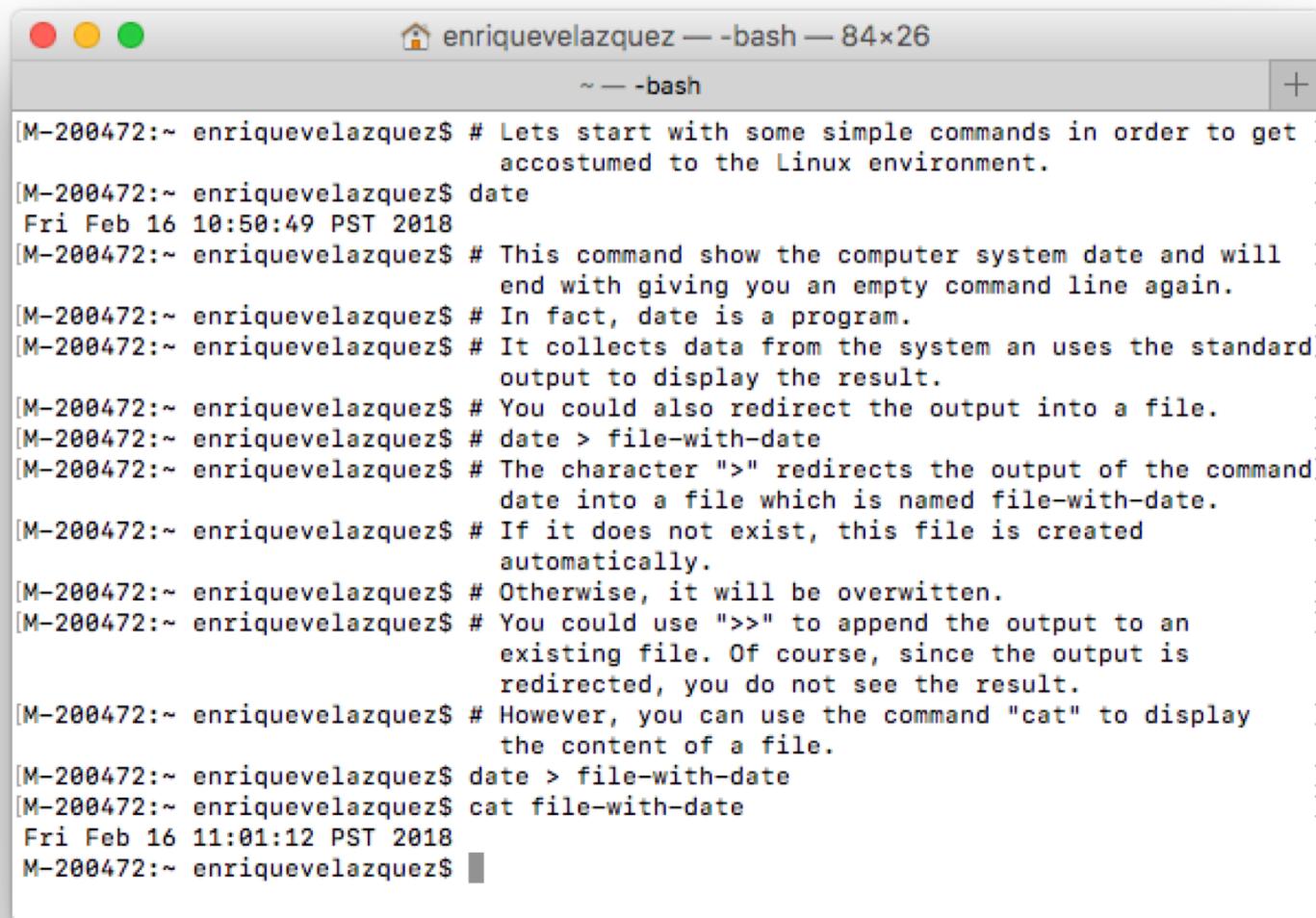
Opening Terminal

- Type “Terminal” in your Mac-Finder



Exercise in Terminal

Type the following Commands:

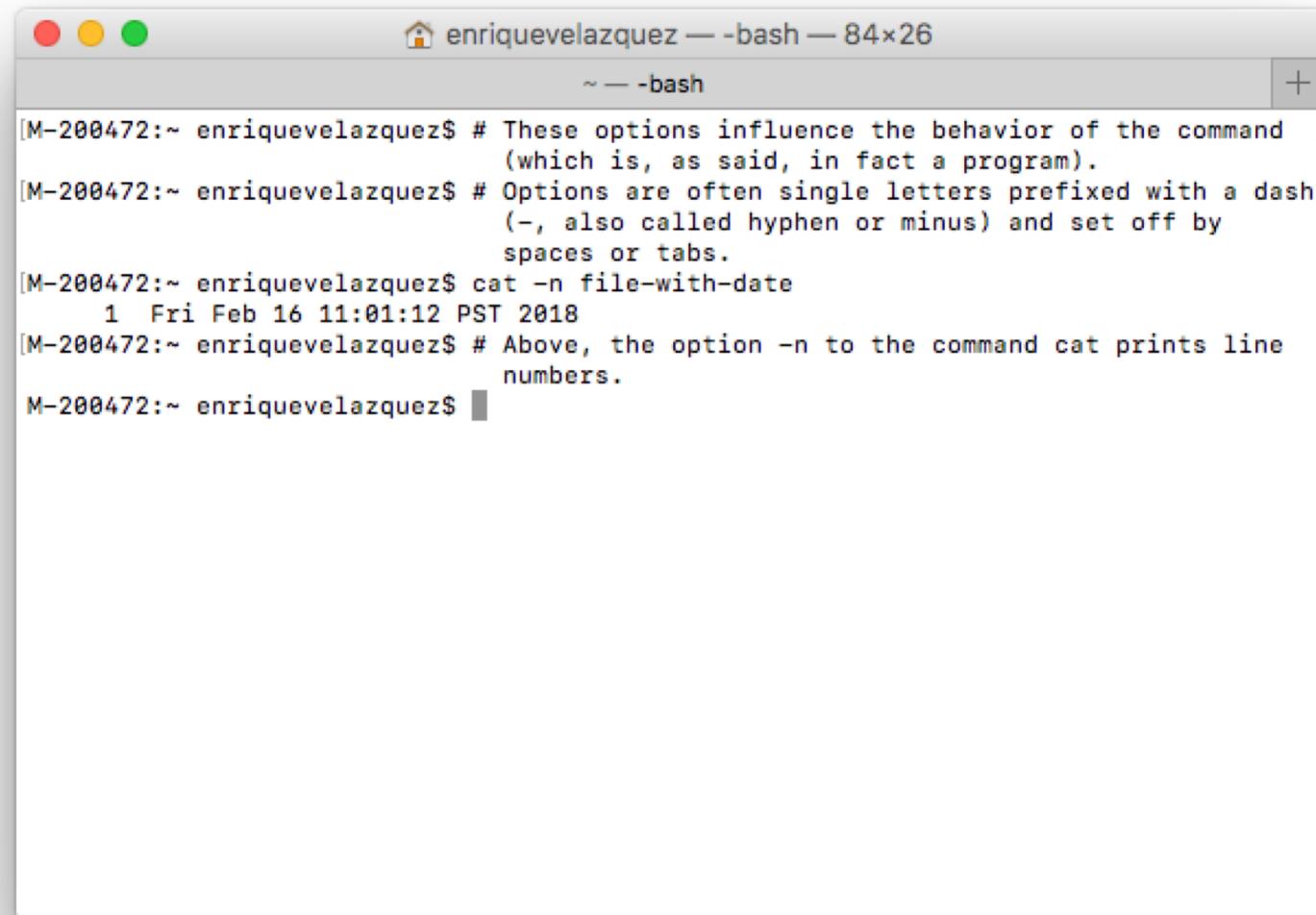


```
[M-200472:~ enriquevelazquez$ # Lets start with some simple commands in order to get ]
      accostomed to the Linux environment.
[M-200472:~ enriquevelazquez$ date
Fri Feb 16 10:50:49 PST 2018
[M-200472:~ enriquevelazquez$ # This command show the computer system date and will ]
      end with giving you an empty command line again.
[M-200472:~ enriquevelazquez$ # In fact, date is a program.
[M-200472:~ enriquevelazquez$ # It collects data from the system an uses the standard]
      output to display the result.
[M-200472:~ enriquevelazquez$ # You could also redirect the output into a file.
[M-200472:~ enriquevelazquez$ # date > file-with-date
[M-200472:~ enriquevelazquez$ # The character ">" redirects the output of the command]
      date into a file which is named file-with-date.
[M-200472:~ enriquevelazquez$ # If it does not exist, this file is created
      automatically.
[M-200472:~ enriquevelazquez$ # Otherwise, it will be overwritten.
[M-200472:~ enriquevelazquez$ # You could use ">>" to append the output to an
      existing file. Of course, since the output is
      redirected, you do not see the result.
[M-200472:~ enriquevelazquez$ # However, you can use the command "cat" to display
      the content of a file.
[M-200472:~ enriquevelazquez$ date > file-with-date
[M-200472:~ enriquevelazquez$ cat file-with-date
Fri Feb 16 11:01:12 PST 2018
M-200472:~ enriquevelazquez$ ]]
```

Exercise in Terminal

Type the following Commands:

Most Commands
Accept Options:

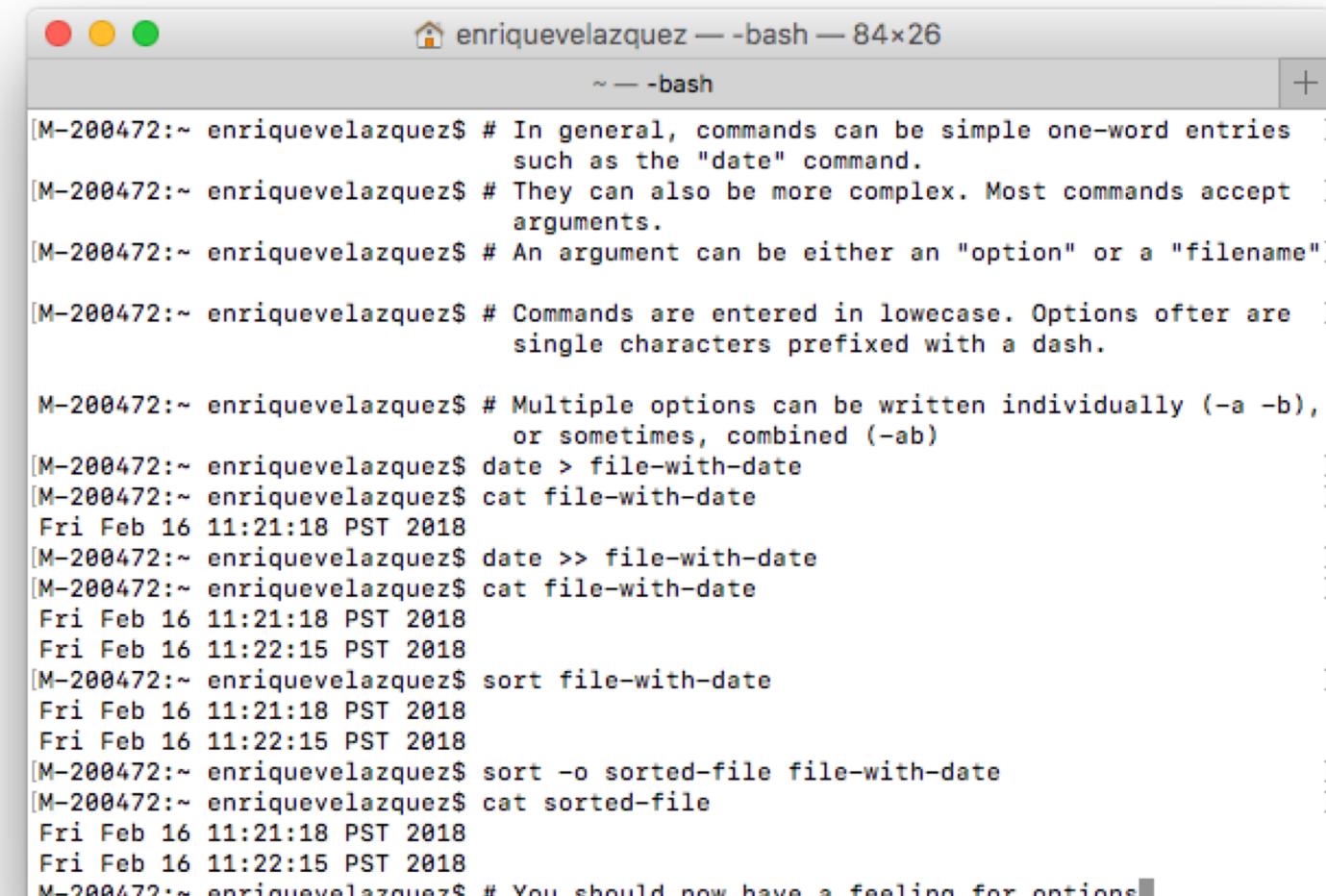


```
[M-200472:~ enriquevelazquez$ # These options influence the behavior of the command ]
      (which is, as said, in fact a program).
[M-200472:~ enriquevelazquez$ # Options are often single letters prefixed with a dash]
      (-, also called hyphen or minus) and set off by
      spaces or tabs.
[M-200472:~ enriquevelazquez$ cat -n file-with-date
  1  Fri Feb 16 11:01:12 PST 2018
[M-200472:~ enriquevelazquez$ # Above, the option -n to the command cat prints line
      numbers.
M-200472:~ enriquevelazquez$ ]
```

Exercise in Terminal

Type the following Commands:

Syntax of Commands:

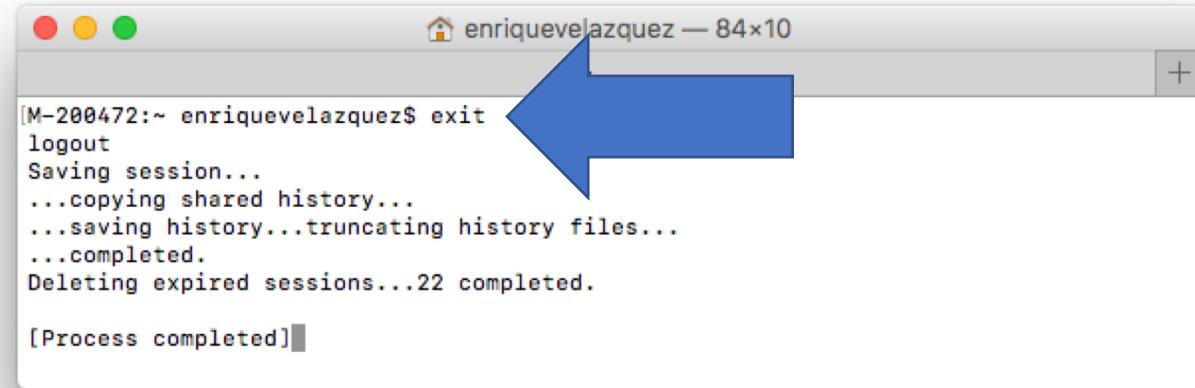


The screenshot shows a terminal window titled "enriquevelazquez — bash — 84x26". The window contains a series of commands demonstrating various aspects of command-line syntax. The commands include:

- # In general, commands can be simple one-word entries such as the "date" command.
- # They can also be more complex. Most commands accept arguments.
- # An argument can be either an "option" or a "filename"
- # Commands are entered in lowercase. Options often are single characters prefixed with a dash.
- # Multiple options can be written individually (-a -b), or sometimes, combined (-ab)
- date > file-with-date
- cat file-with-date
- Fri Feb 16 11:21:18 PST 2018
- date >> file-with-date
- cat file-with-date
- Fri Feb 16 11:21:18 PST 2018
- Fri Feb 16 11:22:15 PST 2018
- sort file-with-date
- Fri Feb 16 11:21:18 PST 2018
- Fri Feb 16 11:22:15 PST 2018
- sort -o sorted-file file-with-date
- cat sorted-file
- Fri Feb 16 11:21:18 PST 2018
- Fri Feb 16 11:22:15 PST 2018
- # You should now have a feeling for options

Exercise in Terminal

Type the following Commands:



```
[M-200472:~ enriquevelazquez$ exit
logout
Saving session...
...copying shared history...
...saving history...truncating history files...
...completed.
Deleting expired sessions...22 completed.

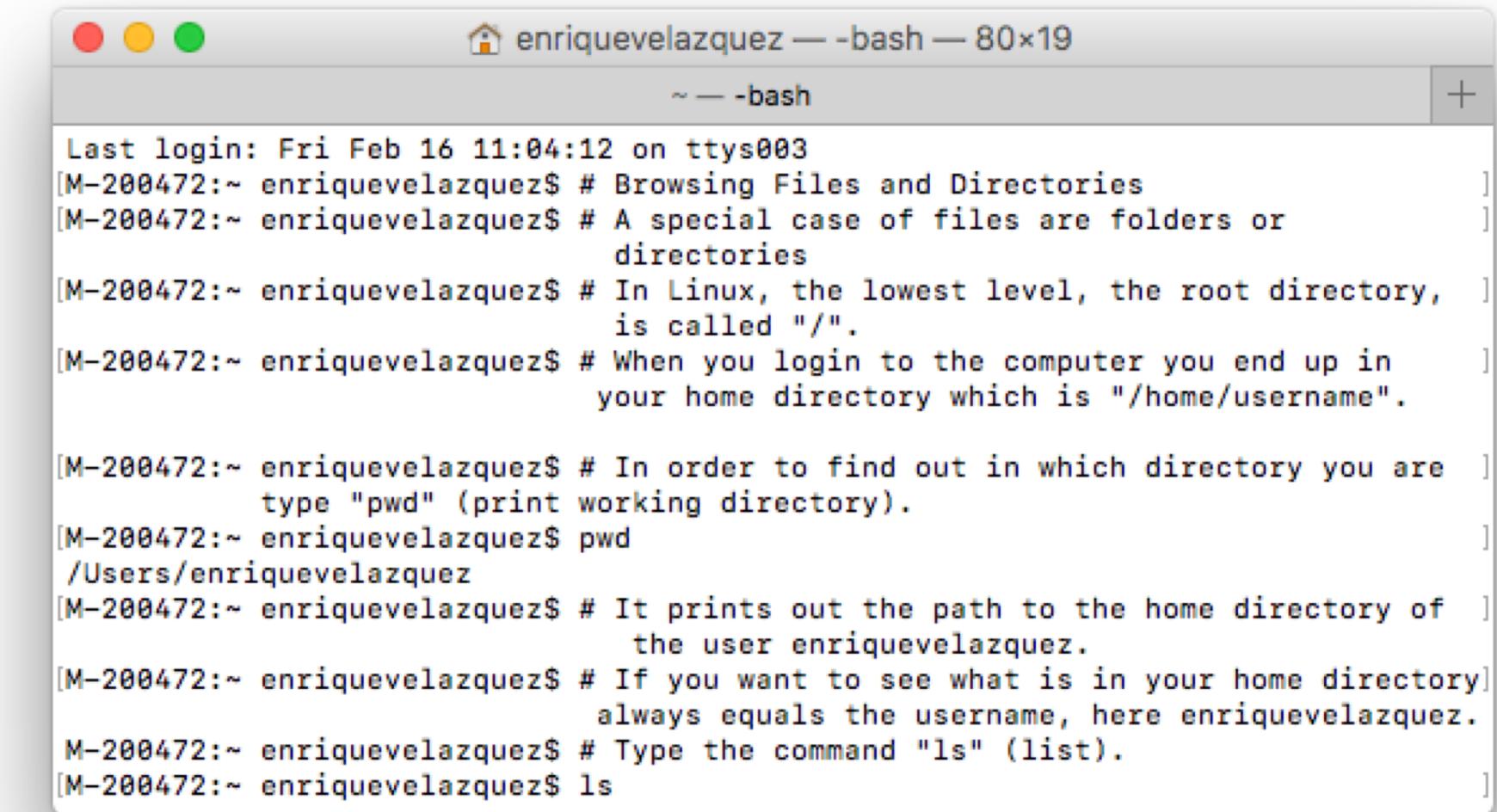
[Process completed]
```

Don't forget logout from your current session!

Exercise in Terminal

Type the following Commands:

Working with
Files:

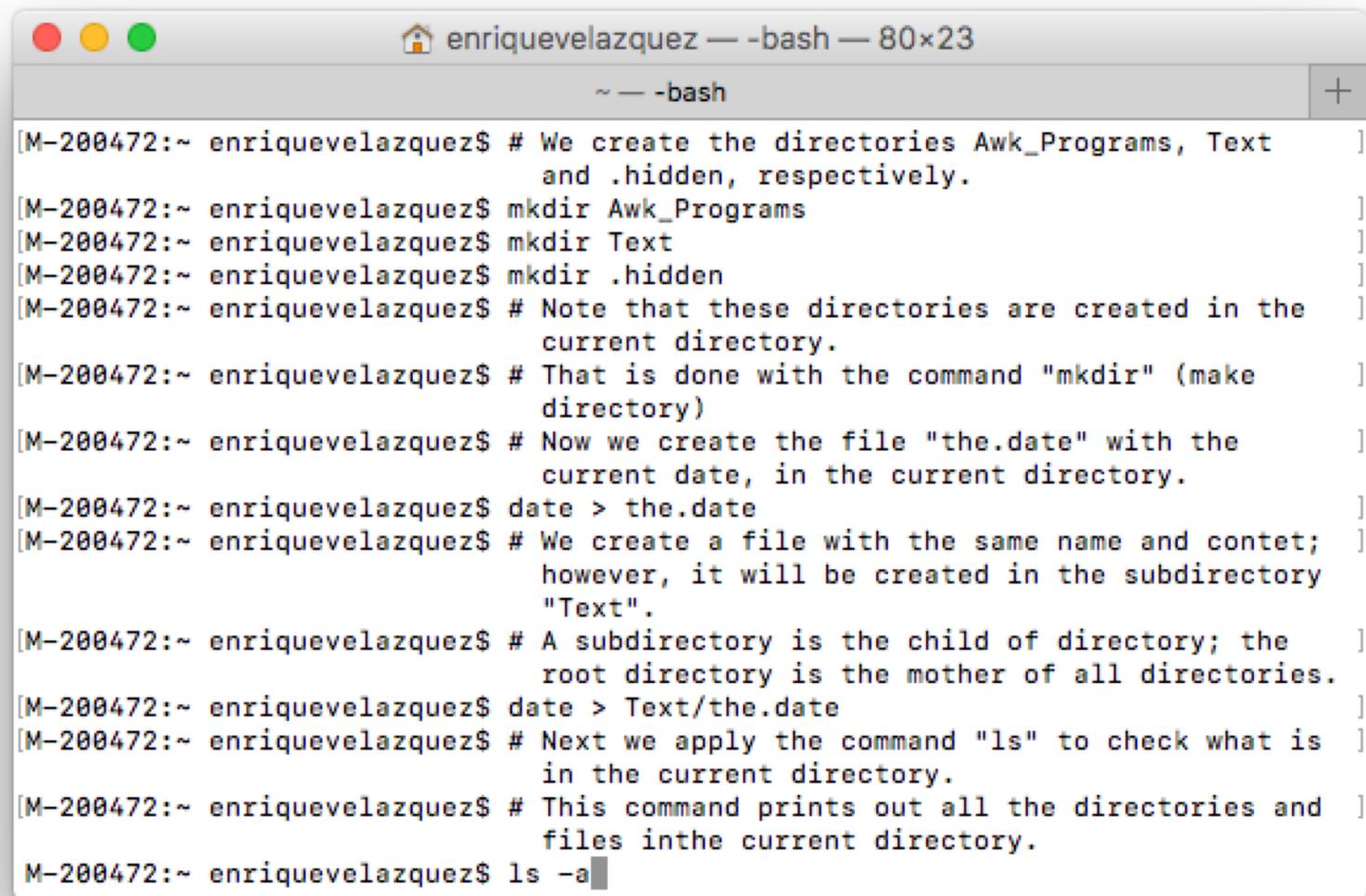


```
Last login: Fri Feb 16 11:04:12 on ttys003
[M-200472:~ enriquevelazquez$ # Browsing Files and Directories
[M-200472:~ enriquevelazquez$ # A special case of files are folders or
                               directories
[M-200472:~ enriquevelazquez$ # In Linux, the lowest level, the root directory,
                               is called "/".
[M-200472:~ enriquevelazquez$ # When you login to the computer you end up in
                               your home directory which is "/home/username".
[M-200472:~ enriquevelazquez$ # In order to find out in which directory you are
                               type "pwd" (print working directory).
[M-200472:~ enriquevelazquez$ pwd
/Users/enriquevelazquez
[M-200472:~ enriquevelazquez$ # It prints out the path to the home directory of
                               the user enriquevelazquez.
[M-200472:~ enriquevelazquez$ # If you want to see what is in your home directory
                               always equals the username, here enriquevelazquez.
[M-200472:~ enriquevelazquez$ # Type the command "ls" (list).
[M-200472:~ enriquevelazquez$ ls
```

Exercise in Terminal

Type the following Commands:

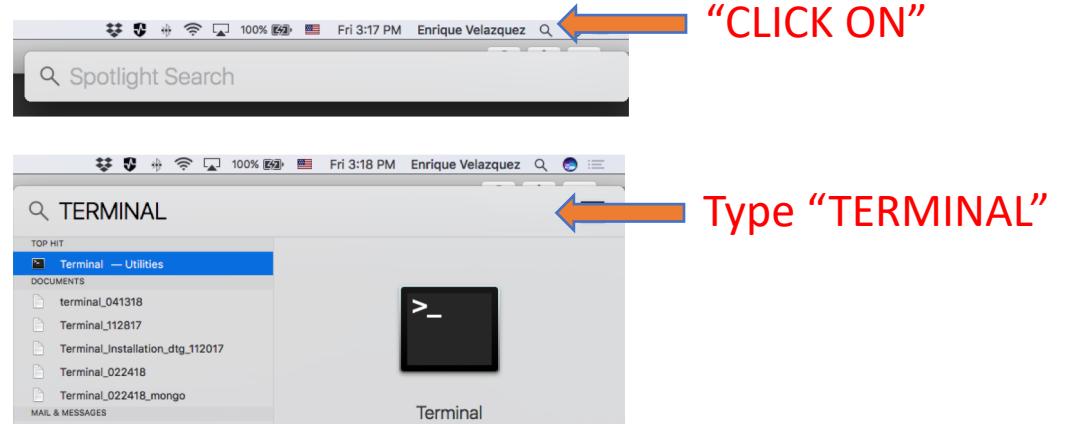
Browsing Files and Directories:



```
[M-200472:~ enriquevelazquez$ # We create the directories Awk_Programs, Text  
and .hidden, respectively.  
[M-200472:~ enriquevelazquez$ mkdir Awk_Programs  
[M-200472:~ enriquevelazquez$ mkdir Text  
[M-200472:~ enriquevelazquez$ mkdir .hidden  
[M-200472:~ enriquevelazquez$ # Note that these directories are created in the  
current directory.  
[M-200472:~ enriquevelazquez$ # That is done with the command "mkdir" (make  
directory)  
[M-200472:~ enriquevelazquez$ # Now we create the file "the.date" with the  
current date, in the current directory.  
[M-200472:~ enriquevelazquez$ date > the.date  
[M-200472:~ enriquevelazquez$ # We create a file with the same name and content;  
however, it will be created in the subdirectory  
"Text".  
[M-200472:~ enriquevelazquez$ # A subdirectory is the child of directory; the  
root directory is the mother of all directories.  
[M-200472:~ enriquevelazquez$ date > Text/the.date  
[M-200472:~ enriquevelazquez$ # Next we apply the command "ls" to check what is  
in the current directory.  
[M-200472:~ enriquevelazquez$ # This command prints out all the directories and  
files in the current directory.  
M-200472:~ enriquevelazquez$ ls -a
```

Opening R version in Terminal

- Type “Terminal” in your Mac-Finder



- Type “R”

Terminal – R version 3.5.1

Last login: Tue Nov 27 17:24:12 on ttys000

M-200472:~ enriquevelazquez\$ R

R version 3.5.1 (2018-07-02) -- "Feather Spray"

Copyright (C) 2018 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.

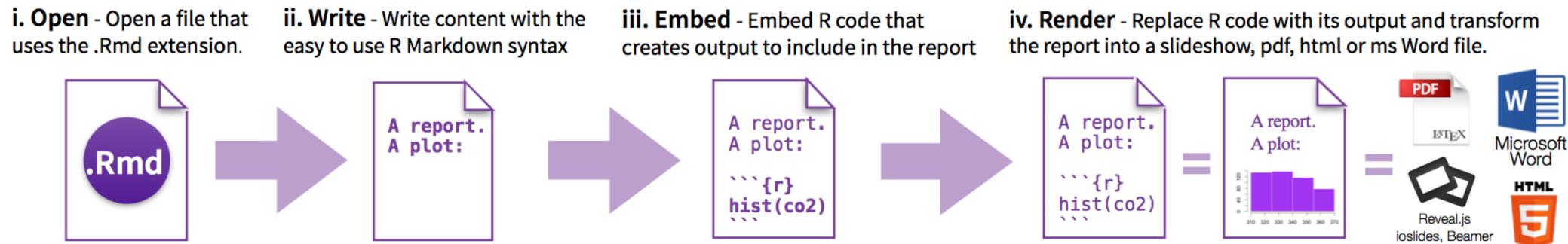
Type 'q()' to quit R.

>

R Markdown

- R Markdown is a foreman for writing reproducible, dynamic reports using R.
- It is used to embed R code and results into documents (pdfs, word, html...)
- Useful to make reports
- This course will use R Markdown for different exercises and Homeworks

Workflow:



R Markdown

- Exercise:
 - Review the “R markdown cheat sheet” from BlackBoard/Week 1/Lectures
 - Create a file using R Markdown
 - Follow the Workflow below and save as HTML file

Workflow:

- i. **Open** - Open a file that uses the .Rmd extension.
- ii. **Write** - Write content with the easy to use R Markdown syntax
- iii. **Embed** - Embed R code that creates output to include in the report
- iv. **Render** - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.

