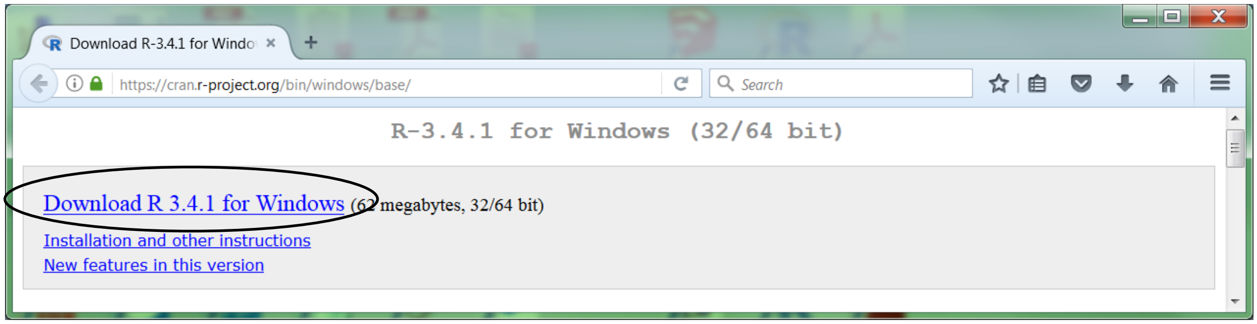
Introduction Lab for R

To download free copies of *R* and RStudio:

Install *R* onto your computer. Search "download R" and then select from the Mac or Windows platforms. Download *R* 3.3.X and follow the instructions to install. (Nothing to submit.)

You can also click one of the links below:

Windows®: https://cran.r-project.org/bin/windows/base/ select "Download R-X.x.x for Windows"


Download R-3.4.1 for Windo
https://cran.r-project.org/bin/windows/base/
Search

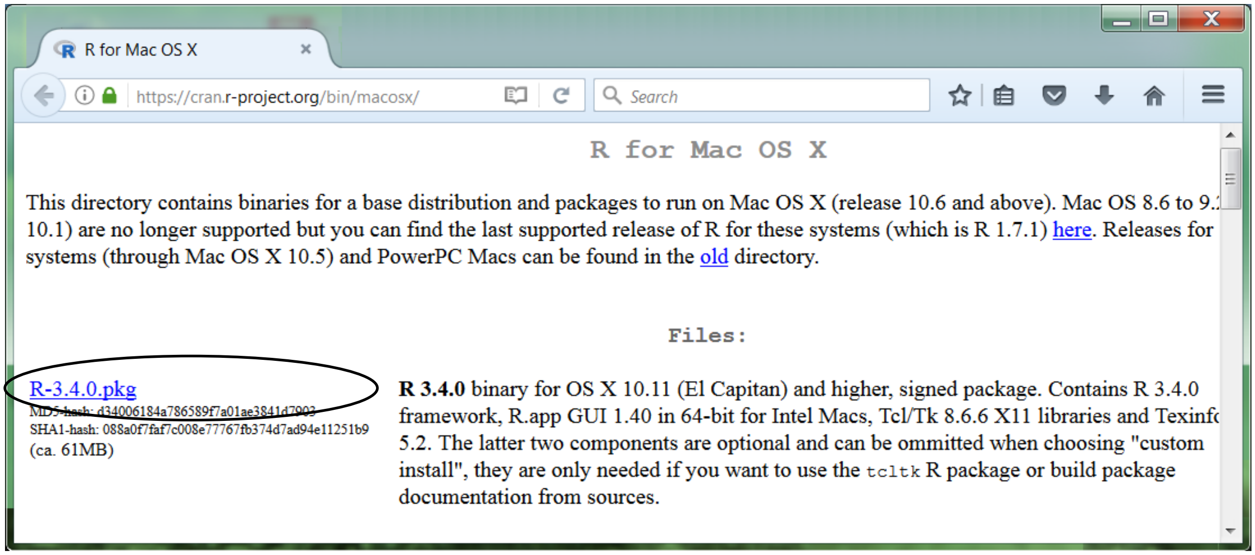
R-3.4.1 for Windows (32/64 bit)

Download R 3.4.1 for Windows (62 megabytes, 32/64 bit)

Installation and other instructions

New features in this version

Mac®: https://cran.r-project.org/bin/macosx/ Download the highest number "R-X.x.x.pkg"



R for Mac OS X

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9. 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) here. Releases for systems (through Mac OS X 10.5) and PowerPC Macs can be found in the old directory.

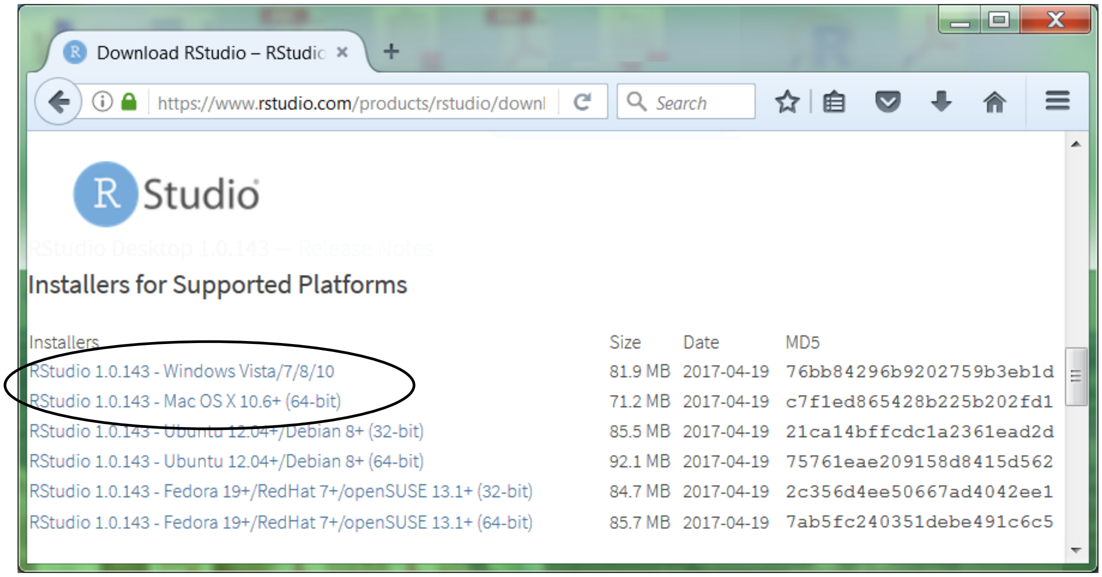
Files:

R-3.4.0.pkg
MD5-hash: d34006184a786589f7a01ae3841d7903
SHA1-hash: 088a0f7faf7c008e77767fb374d7ad94e11251b9
(ca. 61MB)

R 3.4.0 binary for OS X 10.11 (El Capitan) and higher, signed package. Contains R 3.4.0 framework, R.app GUI 1.40 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinf 5.2. The latter two components are optional and can be ommitted when choosing "custom install", they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

Install *R*Studio onto your computer. Search "download *RStudio*" and select the appropriate operating system, Mac® or Windows®. Download and follow the instructions to install. *RStudio* is just a friendlier interface for organizing your work in *R*. (Nothing to submit.)

You can also click the link https://www.rstudio.com/products/rstudio/download/


Download RStudio – RStudic ×
https://www.rstudio.com/products/rstudio/downl
Search
R Studio

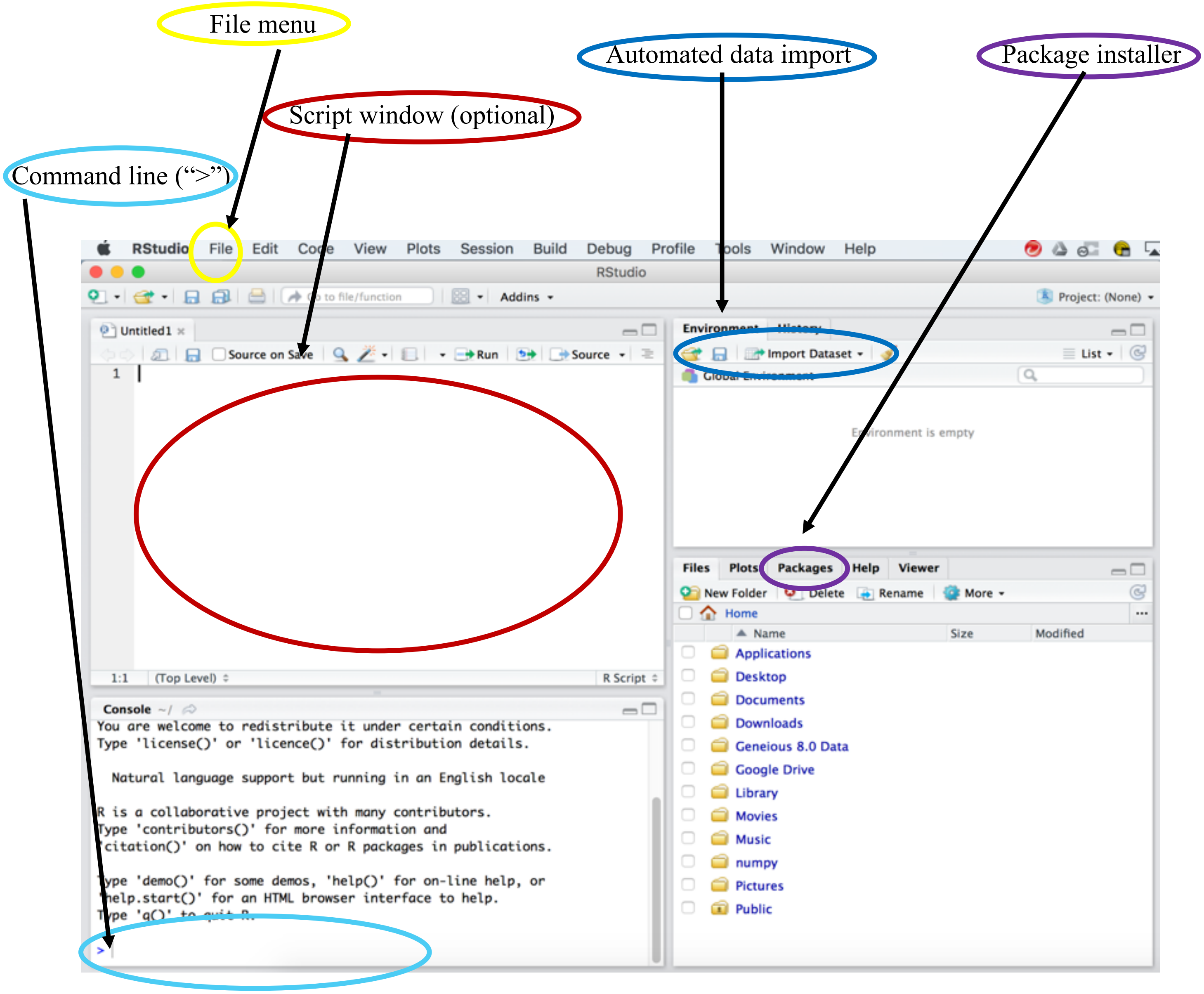
RStudio Desktop 1.0.143 — Release Notes

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.0.143 - Windows Vista/7/8/10	81.9 MB	2017-04-19	76bb84296b9202759b3eb1d
RStudio 1.0.143 - Mac OS X 10.6+ (64-bit)	71.2 MB	2017-04-19	c7f1ed865428b225b202fd1
RStudio 1.0.143 - Ubuntu 12.04+/Debian 8+ (32-bit)	85.5 MB	2017-04-19	21ca14bffcdc1a2361ead2d
RStudio 1.0.143 - Ubuntu 12.04+/Debian 8+ (64-bit)	92.1 MB	2017-04-19	75761eae209158d8415d562
RStudio 1.0.143 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	84.7 MB	2017-04-19	2c356d4ee50667ad4042ee1
RStudio 1.0.143 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	85.7 MB	2017-04-19	7ab5fc240351debe491c6c5

Tip for using *R* on a Mac®: When downloading the data files for the lab, you will want to use Google Chrome as your web browser. Some versions of Safari do not download the lab data properly.

Getting familiar with the *RStudio* environment

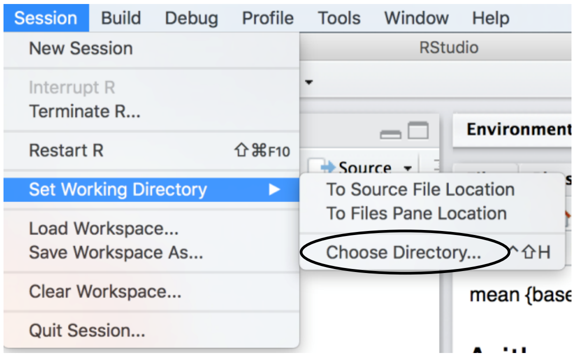


```
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  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.

>
```

Each of the labs in this course will consist guided instructions through the tools and a series of application questions which utilize the tools you just walked through. You must submit a written lab report in a Word or PDF document clearly documenting the code you used, the output to perform ALL of the steps in the lab, along with written answers to questions that ask for it. Occasionally there are steps that don't actually require any input, but are used in a later step, you need not provide output for these. If you are unsure of what to document, remember that the purpose is to document that you have completed every step in the lab. Your labs should be submitted in Canvas.

The following work may be completed after you have downloaded R and RStudio.

1. **Loading data**. We begin this lab by first loading our data into *R* for analysis. There are multiple ways to do this and you may run across others online. Here we choose an approach that tends to be fairly robust across different operating systems and other downstream analysis, and is also of practical use. Begin by downloading the file *FakeData.csv* and saving it in a folder where you can find it. We will load the data in *R* using the *read.csv()* command. Prior to this, we need to tell *R* which folder the data is stored in. From the *Session* menu select *Set Working Directory* and *then Choose Directory.*



Once your list of folders appears, select the folder where you saved the *FakeData.csv* file. Then click "Open." Now, to open the file enter the command at the command line (where the ">" is).

Fakedata <- read.csv("FakeData.csv")

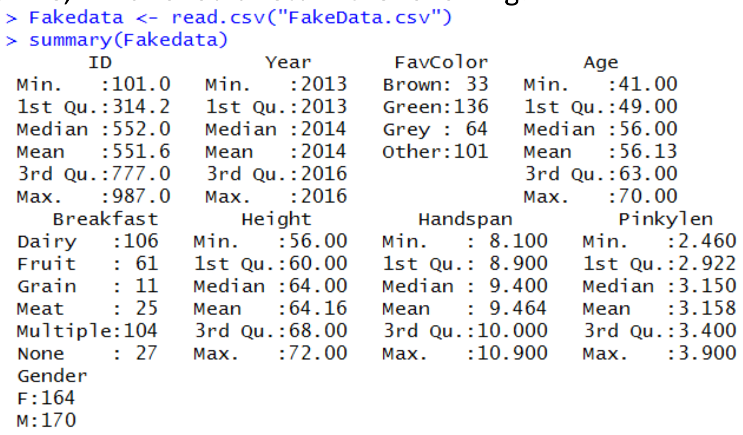
End then hit enter. This command reads the data in the file called FakeData.csv and saves it under the variable *Fakedata*, so that we can call the table of data for analysis later. Note that you can change the name of the variable to anything you want, and in particular you want to name your dataset something easy to remember and easy to type.

Pro Tip: If you get an error message, be careful that the name of the data file is exact, including upper and lower case. Secondly, often copying and pasting commands will result in errors as text editors will change the quote marks to fancy curved quotes, which *R* cannot interpret.

In order to check that the data read correctly we will produce a quick summary of what is in the data to check that things have been read in correctly. Enter the command

summary(Fakedata)

at the command line, which should return the following:



```
> Fakedata <- read.csv("FakeData.csv")
> summary(Fakedata)
       ID             Year        FavColor        Age       
 Min.   :101.0   Min.   :2013   Brown: 33   Min.   :41.00  
 1st Qu.:314.2   1st Qu.:2013   Green:136   1st Qu.:49.00  
 Median :552.0   Median :2014   Grey : 64   Median :56.00  
 Mean   :551.6   Mean   :2014   Other:101   Mean   :56.13  
 3rd Qu.:777.0   3rd Qu.:2016               3rd Qu.:63.00  
 Max.   :987.0   Max.   :2016               Max.   :70.00  
    Breakfast       Height         Handspan        Pinkylen    
 Dairy   :106   Min.   :56.00   Min.   : 8.100   Min.   :2.460  
 Fruit   : 61   1st Qu.:60.00   1st Qu.: 8.900   1st Qu.:2.922  
 Grain   : 11   Median :64.00   Median : 9.400   Median :3.150  
 Meat    : 25   Mean   :64.16   Mean   : 9.464   Mean   :3.158  
 Multiple:104   3rd Qu.:68.00   3rd Qu.:10.000   3rd Qu.:3.400  
 None    : 27   Max.   :72.00   Max.   :10.900   Max.   :3.900  
 Gender 
 F:164  
 M:170  
```

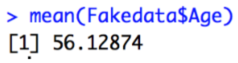
If you wish to view the contents of the data set, enter the command *Fakedata* (the name of the variable it is stored under) and you will see all of the entries. You may also notice that under the Environment tab in the upper right window of RStudio, there is a listing and description of the variable *Fakedata*. If you click on this, you will see the data in tabular form.

2. **Basic computations with data** (You may wish to use these commands on your homework). Suppose that you wish to compute the mean of a column of data. Now that your data is loaded you can begin the analysis, which will be much easier. Suppose you want to know the mean age of subjects in the dataset. We will have to specify which column of data we are referring to. *R* uses the syntax *Fakedata$Age,* to refer to this column. The part in front of the '$' specifies the name of the data set, and the part of the command after the '$' refers to the name of the column we are interested in. Again, make sure that you have typed the name of the column or data set exactly as it appears. You may also notice that when you enter the '$' *R* provides a list of names, all of columns in the data set. You may just click on one of these and have the name automatically entered to avoid typing mistakes.

   In order to compute the mean just enter the command

   mean(Fakedata$Age)

   Most commands follow a similar structure where *mean()* is the operation we want *R* to perform, and the thing inside the () is the data we want the operation performed on. Enter this command and verify that you obtain



```
> mean(Fakedata$Age)
[1] 56.12874
```

   Find the max and min of the age variable. What would you guess the command would be to compute the median of the Height variable from this data set?

3. **Computing on your own data:** Sometimes you may have a list of data you would like to summarize and it is isn’t already in a spreadsheet (say in a homework problem)? There are two approaches, the first is to put the data into Excel, Numbers or Google sheets (with labels) and save as a .csv. Once you do this save it to the location you want and modify the *read.csv()* command above to read the data in. This involves several steps so another way is to save the list of numbers directly and use it directly. Suppose you had five heights, such as 61,64,67,72,63 and want to compute the mean. You can save them under the variable name height using the command

   ```
   Height <- c(61,64,67,72,63)
   ```

   Note that then instead of using the $ notation, like Fakedata$Age to refer to the age column in the Fakedata set, you can now just refer to the variable *Height.* You can do this for any list of numbers, just wrap them with a c(). Try computing the mean for this data you just entered by entering the previous command and then

   ```
   mean(Height)
   ```

   Verify that you obtain a mean of 65.4 on this data using R.

Application questions

Answer the questions in this portion of the lab using the tools you used previously. Frequently this section will use a dataset different from the Fakedata.

1. What is the median Handspan?
2. What is the mean Pinkylen?
3. Compute the mean of the numbers 3, 9, 11, 201 using R.

Submit answers to these questions in a Word or PDF document including output and commands used for each step in the assignment. Prior to submitting double check that you have the output for each of the steps.