

RNA-Seq Session Organization

- Introduction to RNA-Seq Experiment and Design
- RNA-Seq Alignment and Normalization
- RNA-Seq Differential Expression and Functional Analysis
- [Hands-On] RNA-Seq Differential Expression Analysis in R walkthrough

R Primer with Genomic Data

Why learn programming?

Data size and scope is constantly growing.

Saving time- automating analysis

Saving face- reproducible analysis and findings

- (Hopefully!) less human error

Most things have tradeoffs. Questions to ask yourself:

- Is this something I will have to do often?
- Is it tedious?
- Do I already know how to do this?
 - For small tasks, probably easier/quicker to use what you are comfortable with

Writing the actual code

Do NOT use a word processing program (e.g. MS Word) to write your code

- Introduces special/unseen characters which break things!
- Also be careful of copy/paste from PDFs or from the web for similar reasons

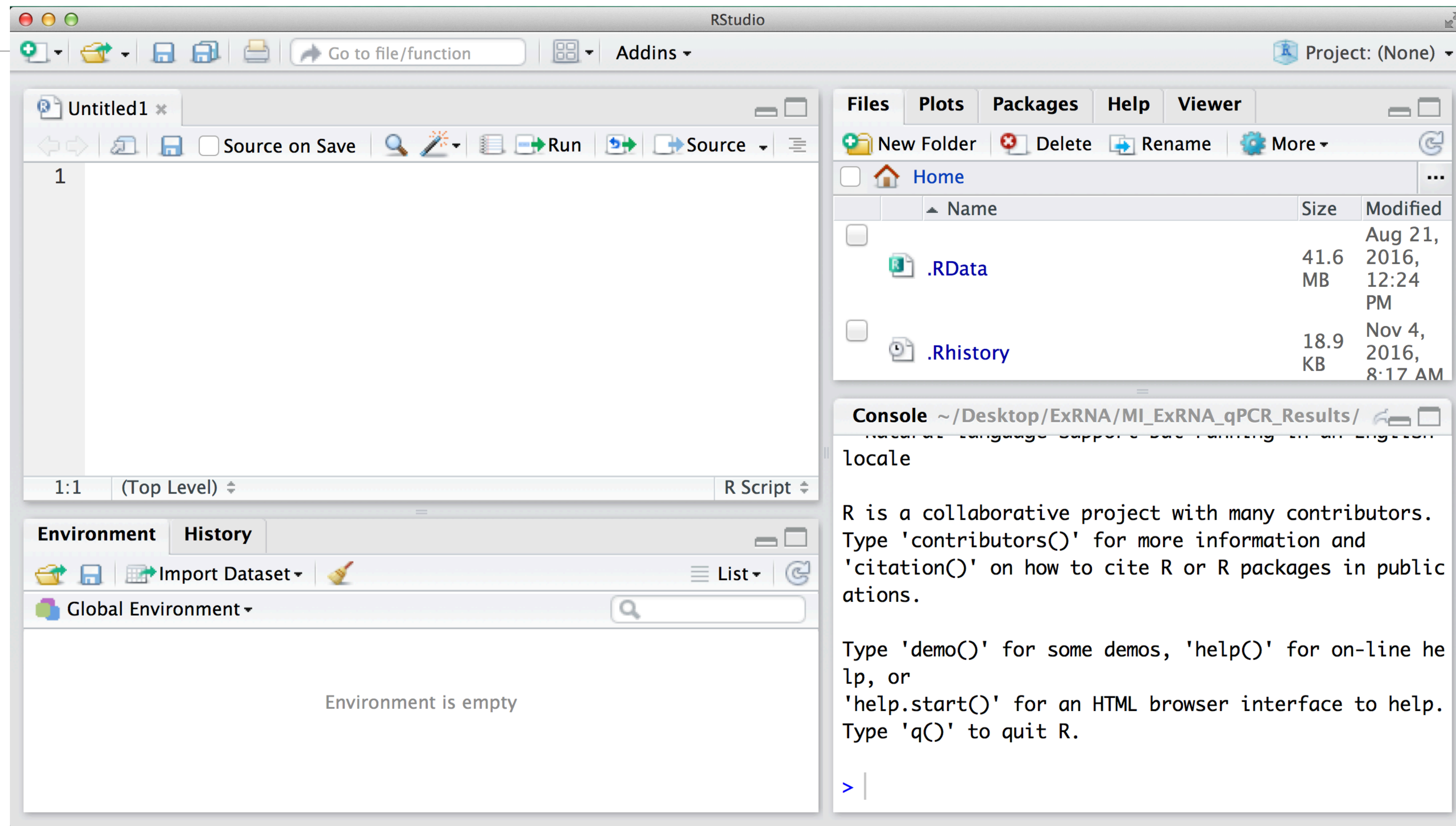
Use a “code aware” editor-- offers highlighting and coloring which can be helpful

- Notepad++ for Windows
- TextMate, etc. for Mac

Even better-- use a **I**ntegrated **D**evelopment **E**nvironment (IDE)

- Offers improved syntax highlighting, auto-completion, debugging, etc.
- RStudio is very popular for R development (<https://www.rstudio.com/products/rstudio2/>)

Getting familiar with your R installation and Rstudio



Coding- the importance of being exact

```
> f <-2  
> F + 1  
Error: object 'F' not found
```

Case matters!

Try not to use spaces

- Sometimes you simply cannot: variable names cannot contain spaces
- Even if it is valid (such as in file names) it just makes things more difficult

```
> FA <-2  
> F A + 1  
Error: unexpected symbol in "F A"
```


R relies on its rich libraries



CRAN (The Comprehensive R Archived Networks)

- The primary R library archive
- Large number of statistical and machine learning packages available



Bioconductor

- R library archive specifically for Open Source Bioinformatics Software
- Largest collection of any Bioinformatics Library

Note: you will see the term ‘package’ and ‘library’ when talk about R. Package is the collection of function and library is where packages are archived.

Example Analysis:

Input Data

- 3 WT vs 3 Treated Cell line RNA-Seq data
- Single End 75bp RNA-Seq, STAR aligned, HTSeq quantified raw count

We will perform:

1. Install and load libraries from CRAN and Bioconductor
2. Load Data
3. PCA on raw count data
4. Hierarchical Clustering Tree
5. DESeq2 to perform differential gene expression analysis
6. Heatmap
7. Volcano plot