

# Introduction to R

---

Elin Videvall

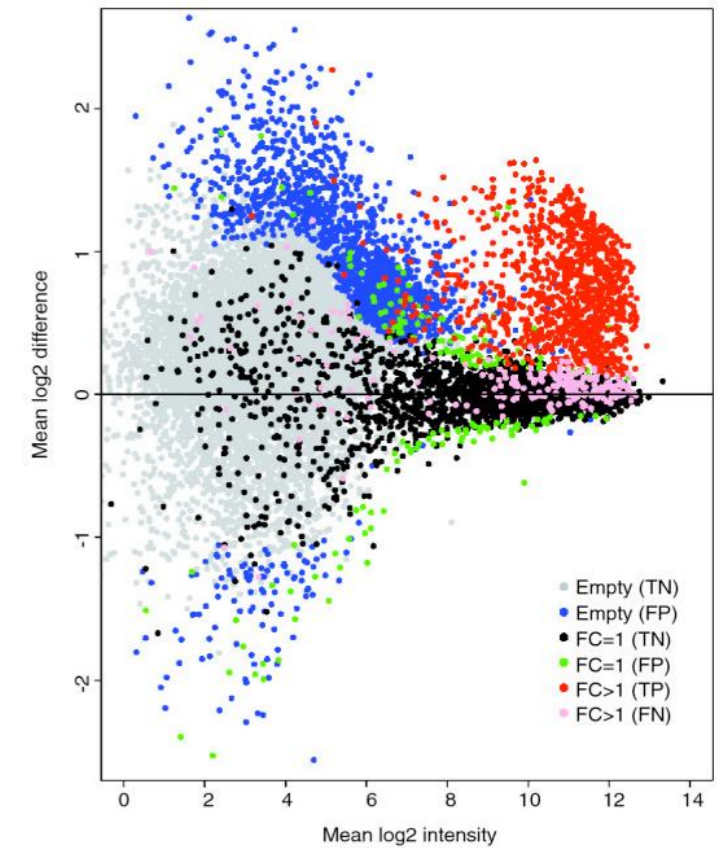
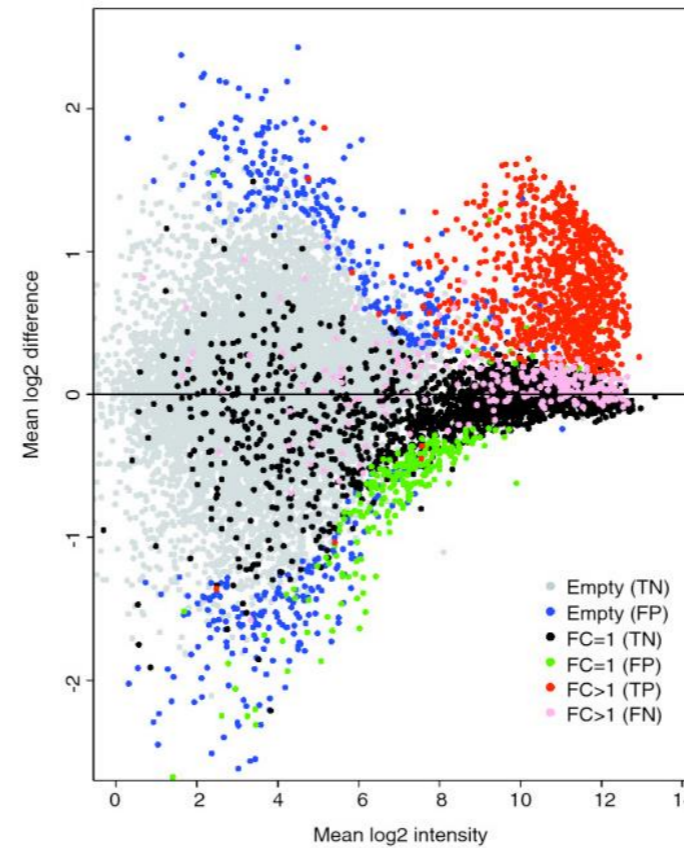
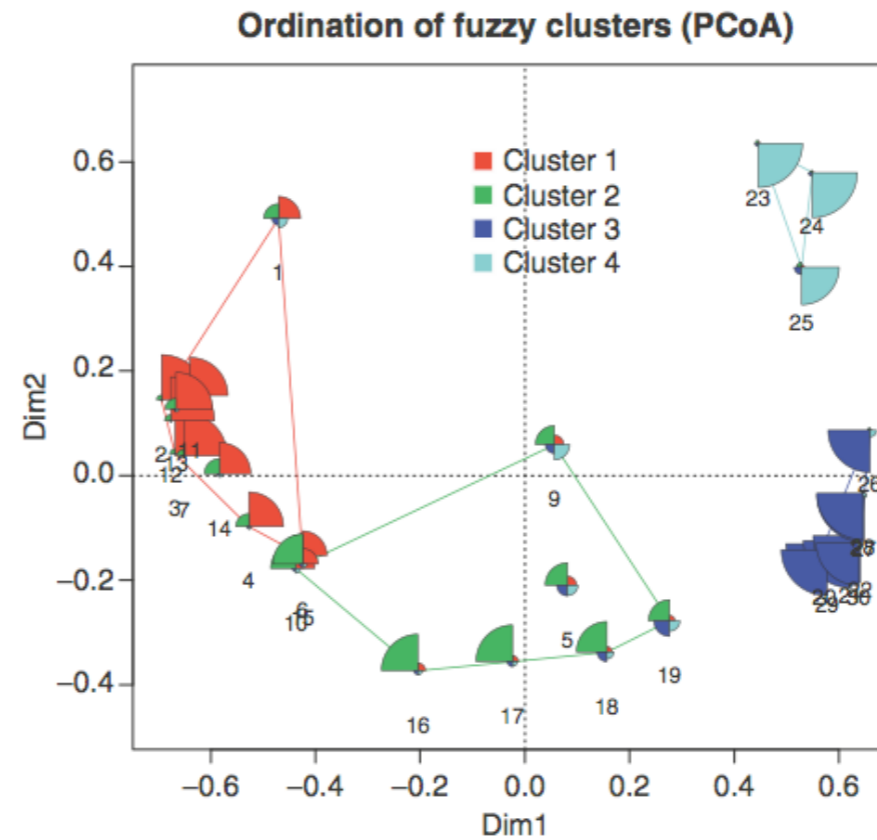
14 Jan 2017

 @ElinVidevall



# What is R?

A programming language and software environment for statistical computing and graphics



# History of R

---

- S programming language - created by John Chambers at Bell Labs, US
- R programming language - created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand
- Partly named after Ross and Robert's names and partly as a play on the name of S



# Why is R so useful (and popular)?

---

- Data management and manipulation
- Well established system of packages and documentation
- Support for rich statistical simulation and modeling
- High-level interpreted language to prototype new computational methods
- Active development and dedicated community
- Cutting-edge graphical data visualization
- Free!



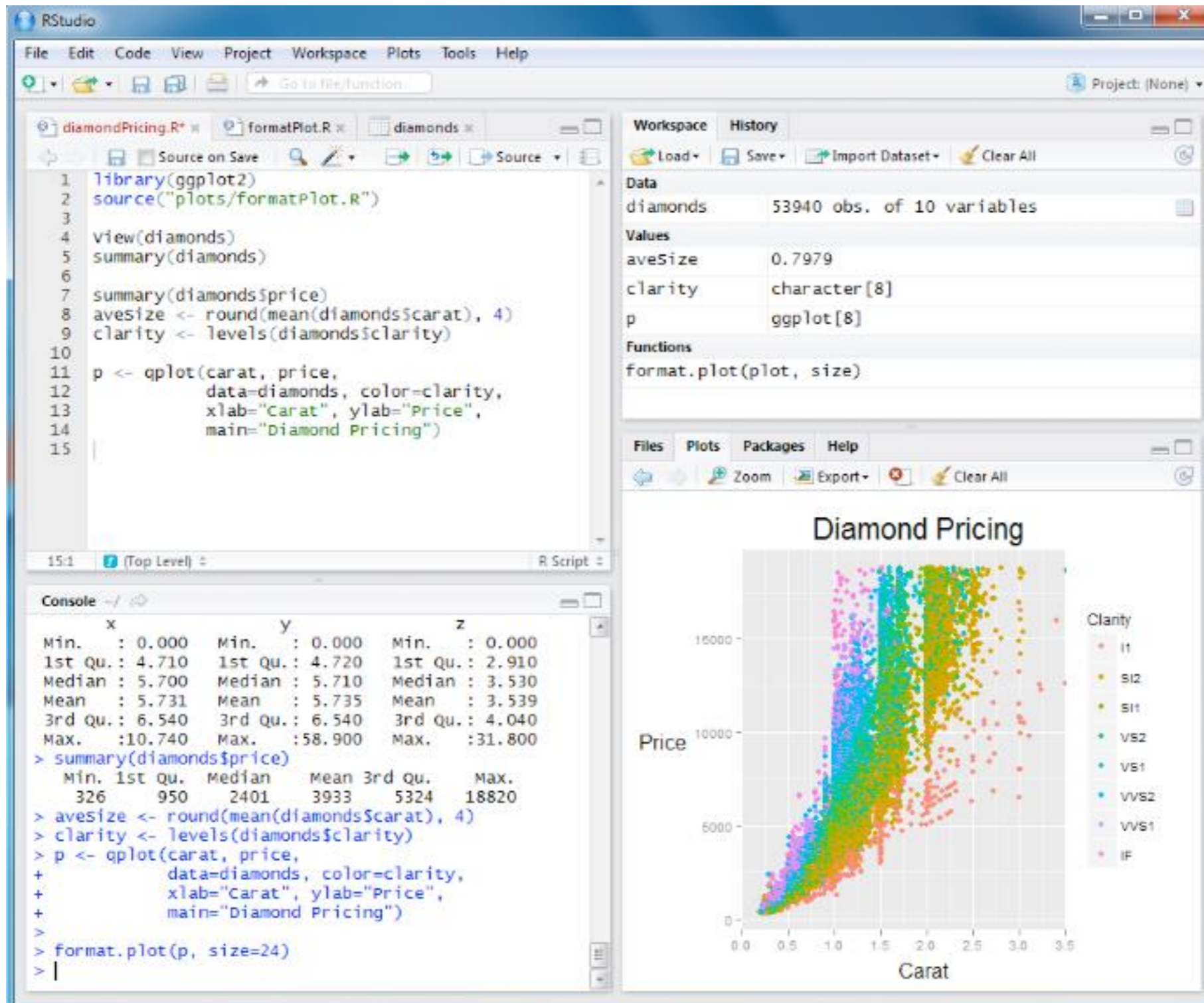
# Running R

---

- Install an R Integrated Development Environment (IDE)
  - RStudio: <http://www.rstudio.com>
  - Can make working with R much easier, particularly for a new user
  - Run on Windows, Mac or Linux OS
- Or from the command line, type R



# R Studio



# R Studio

Open up a tab in your own browser (outside X2Go)

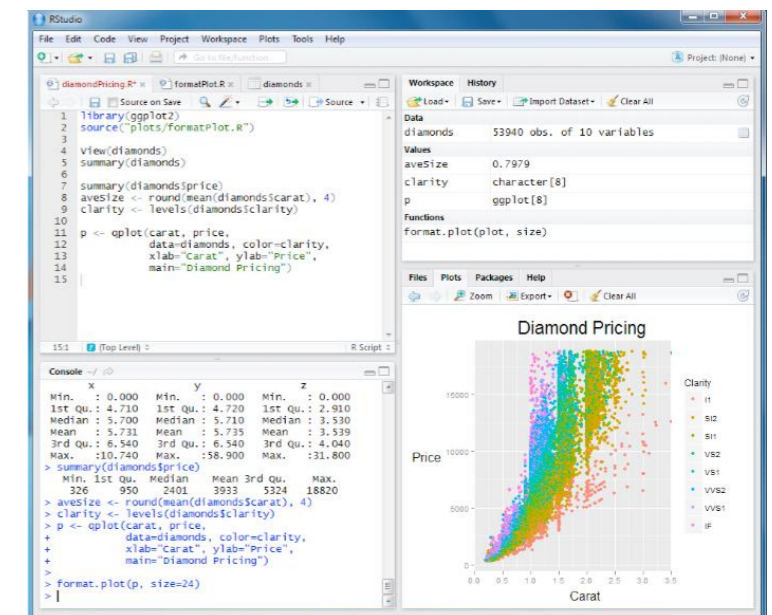
<YourDNSid>:8787

user: <retracted>

pass: <retracted>

File -> New file -> R script

File -> Save as -> "R\_intro\_exercise.R"



# Help

---

- How to remember all of the variables?
- ? is your friend
- ?plot

plot {graphics}

## Generic X-Y Plotting

### Description

Generic function for plotting of `R` objects. For more details about the graphical parameter arguments, see [par](#).

For simple scatter plots, [plot.default](#) will be used. However, there are `plot` methods for many `R` objects, including [functions](#), [data.frames](#), [density](#) objects, etc. Use `methods(plot)` and the documentation for these.

### Usage

```
plot(x, y, ...)
```

R Documentation

`type`

what type of plot should be drawn. Possible types are

- "p" for **p**oints,
- "l" for **l**ines,
- "b" for **b**oth,
- "c" for the lines part alone of "b",
- "o" for both 'o**ver**plotted',
- "h" for 'h**istogram**' like (or 'high-density') vertical lines,
- "s" for stair **s**teps,
- "S" for other **s**teps, see 'Details' below,
- "n" for no plotting.



# R Mechanics - Installing & Loading Packages

---

Installing regular R packages:

```
install.packages("vegan")
```

Installing Bioconductor packages:

```
source("http://bioconductor.org/biocLite.R")
```

```
biocLite("DESeq2")
```

Loading packages in R:

```
library("vegan")
```

```
library("DESeq2")
```

# Basic R functionality

---

**TAB** completion

**Up-arrow** for last command

In RStudio:

**Ctrl + Enter** to run current line or selection

(**Cmd + Enter** in Mac OS)

# Where to learn more about R

---

- The R Project Homepage: <http://www.r-project.org>
- Quick R Homepage: <http://www.statmethods.net>
- Bioconductor: <http://www.bioconductor.org>
- An Introduction to R (long!): <http://cran.r-project.org/doc/manuals/R-intro.html>
- Google - there are tons of tutorials, guides, demos, packages and more

# R for Biologists

---

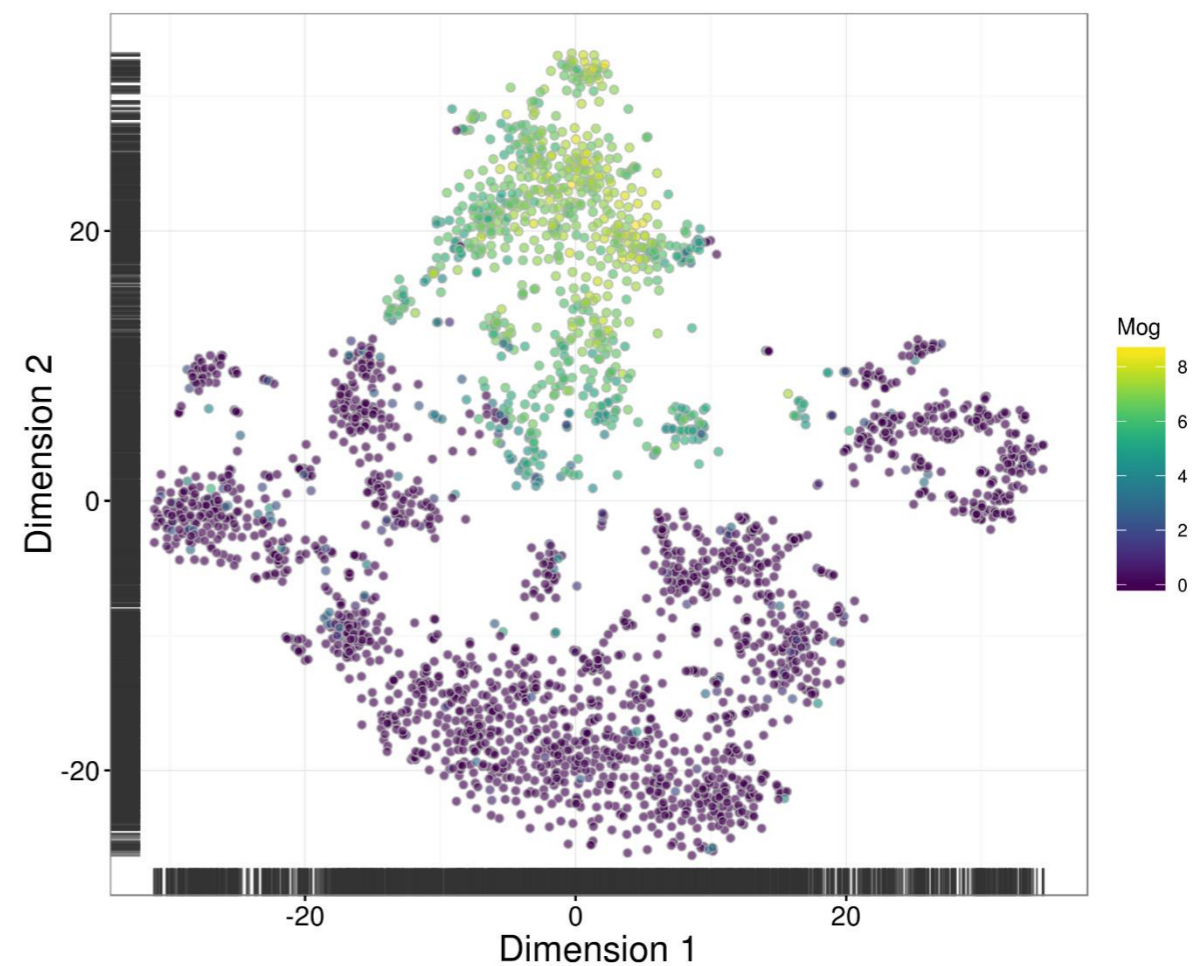
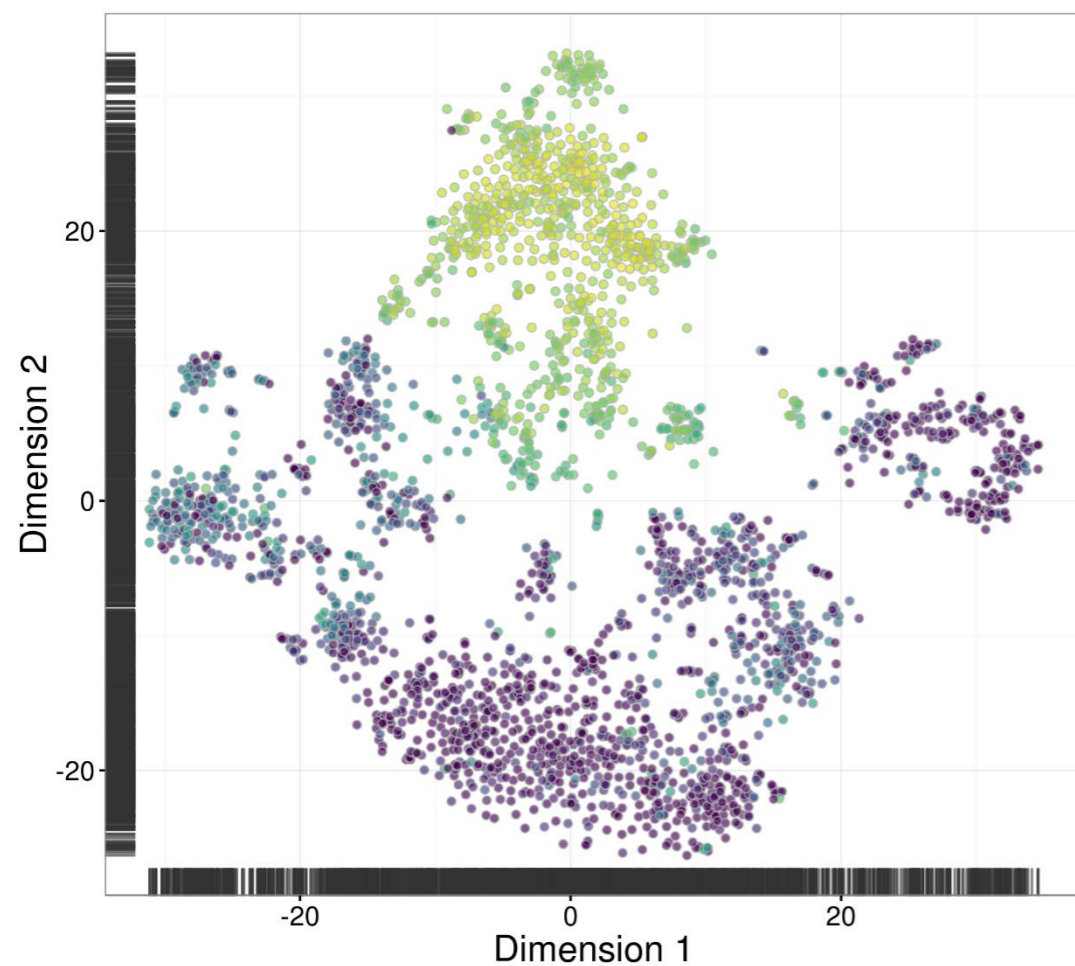
- Bioconductor (<http://bioconductor.org>)
  - 1,300 packages:
    - Variant detection: sequence analysis, PolyPhen database
    - Annotation: pathway analysis, BioMart, GO, KEGG, NCBI and many others
    - High-throughput assays: flow cytometry, mass spec, proteomics
    - Transcription factor binding, differential gene expression analysis
- Ecology (see: <http://cran.r-project.org/web/views/Environmetrics.html>)
  - Ordination
  - Cluster Analysis
  - Ecological Theory
  - Population Dynamics
  - Spatial Data Analysis
- Phylogenetics and Evolution (see: <http://cran.r-project.org/web/views/Phylogenetics.html>)
  - Ancestral State Reconstruction
  - Phylogenetic Inference
  - Trait Evolution

# Bioconductor workflow examples

---

A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor

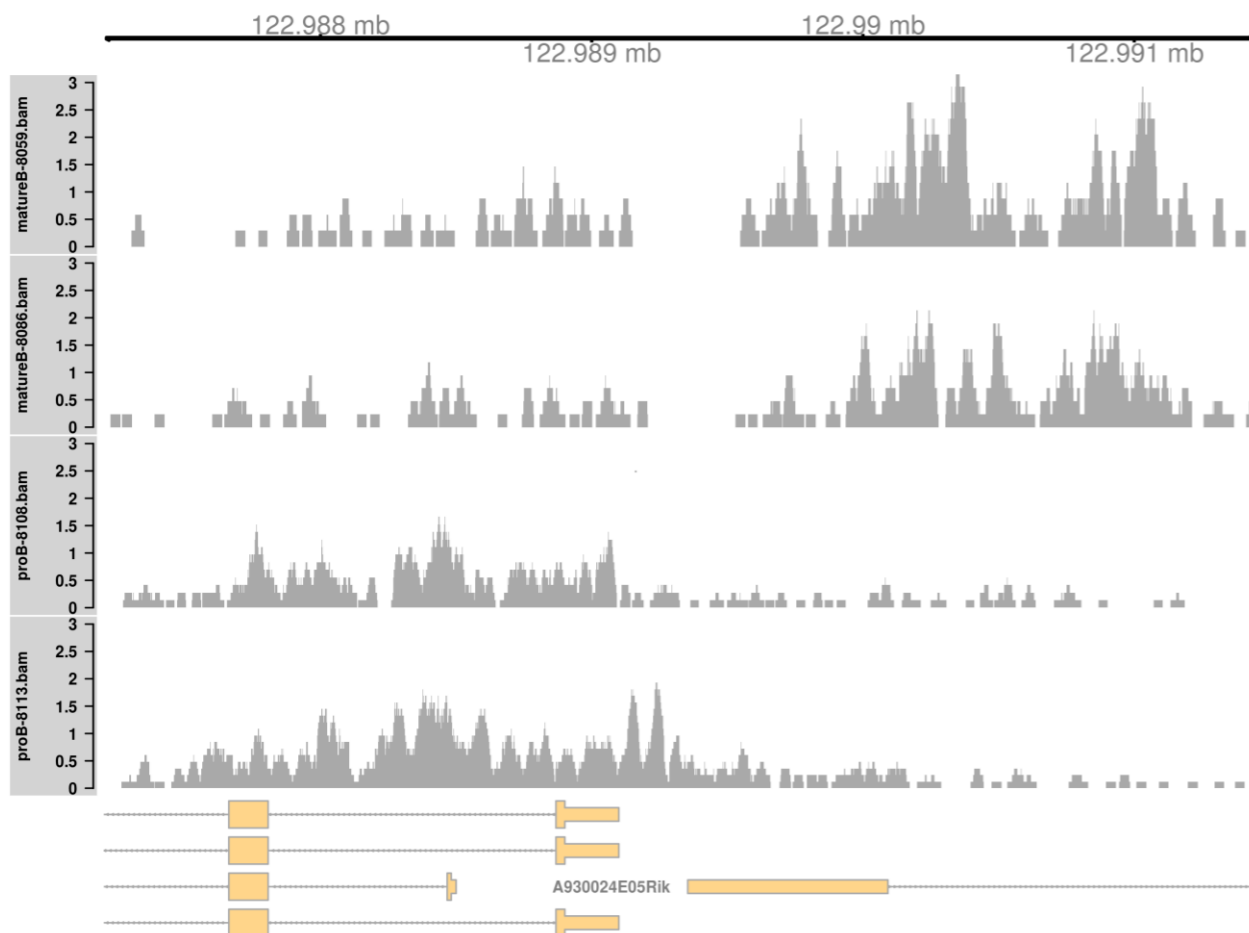
*Aaron T. L. Lun<sup>1</sup>, Davis J. McCarthy<sup>2</sup> and John C. Marioni<sup>3</sup>*



# Bioconductor workflow examples

From reads to regions: a Bioconductor workflow to detect differential binding in ChIP-seq data

*Aaron T. L. Lun<sup>1</sup> and Gordon K. Smyth<sup>2</sup>*

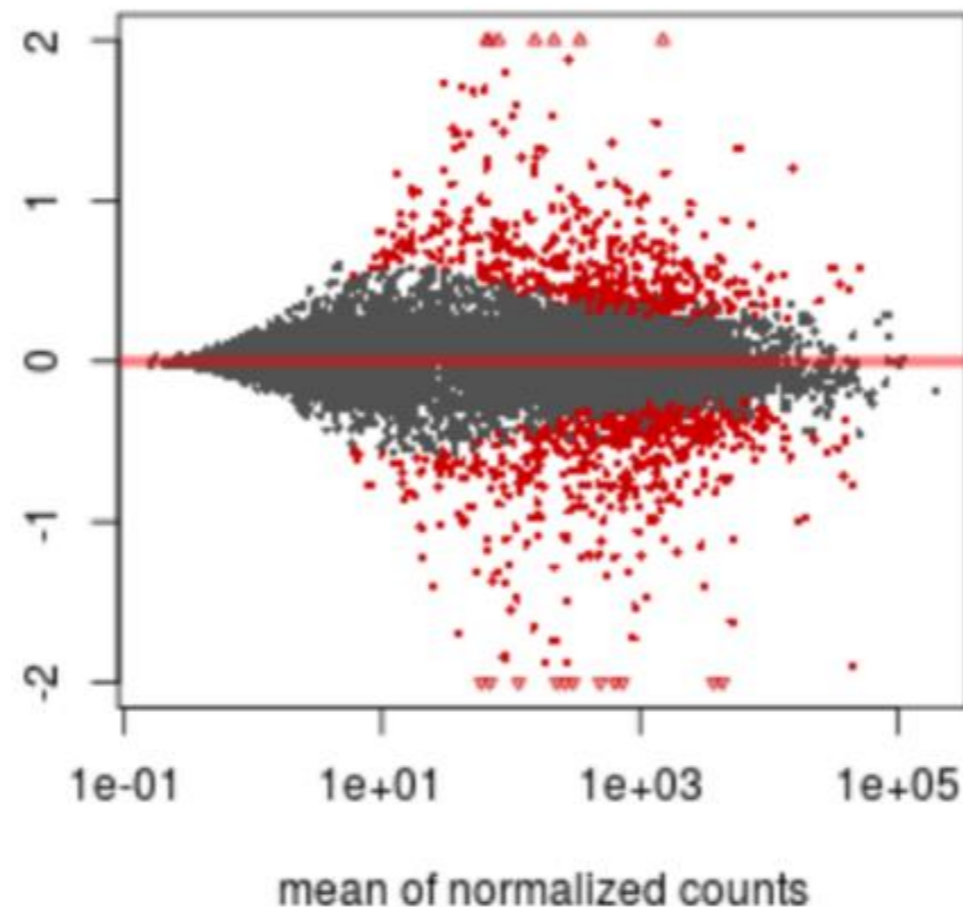


# Bioconductor package example

---

## Differential analysis of count data – the DESeq2 package

*Michael I. Love<sup>1</sup>, Simon Anders<sup>2</sup>, and Wolfgang Huber<sup>3</sup>*



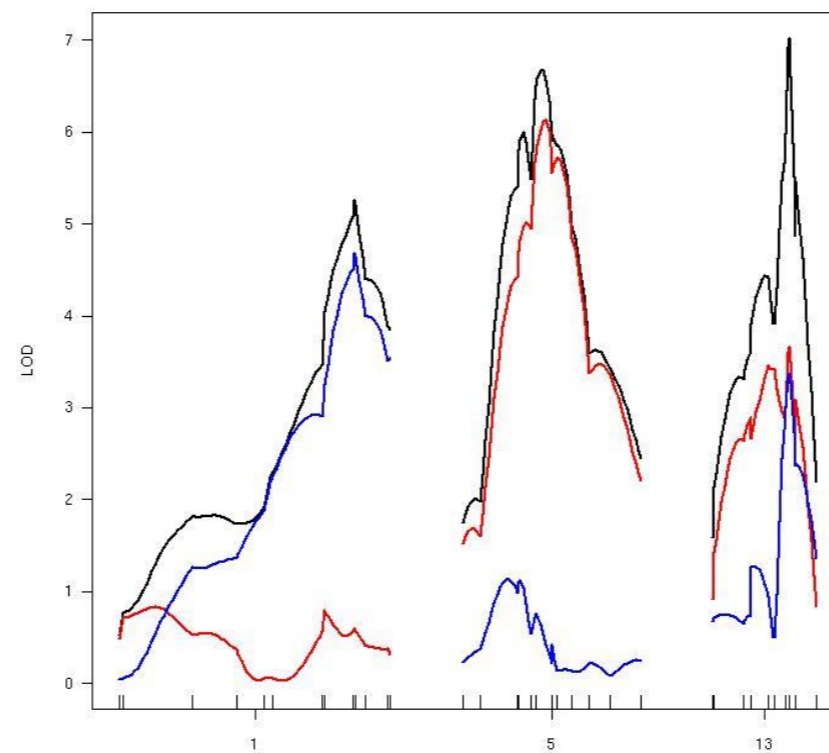
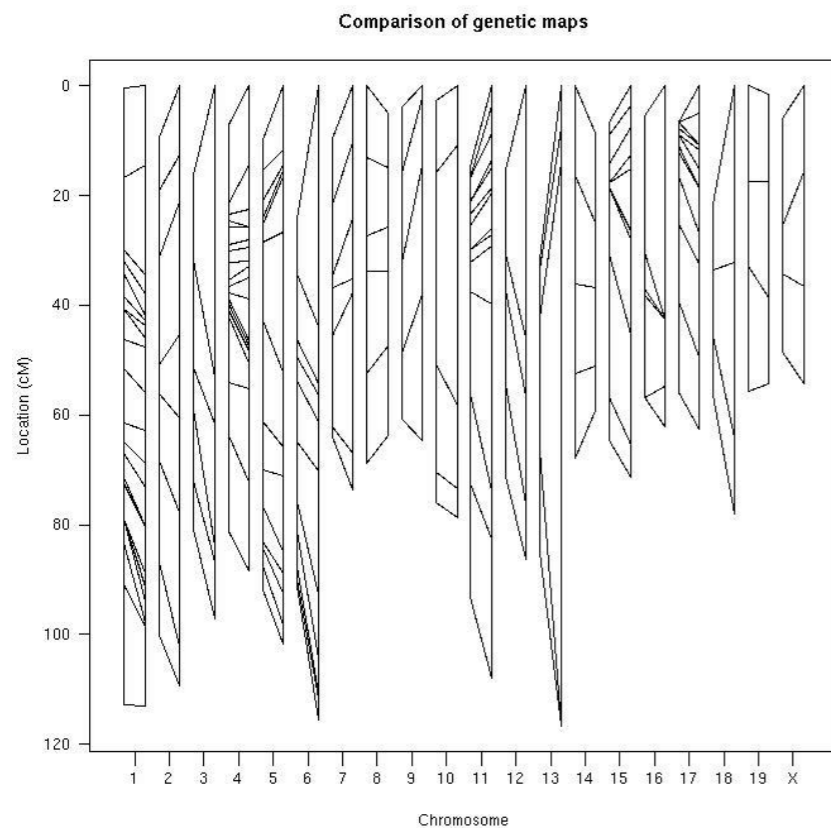
# R package example

---

## R/qtl: A QTL mapping environment

Software for mapping quantitative trait loci in experimental crosses

[www.rqtl.org](http://www.rqtl.org)

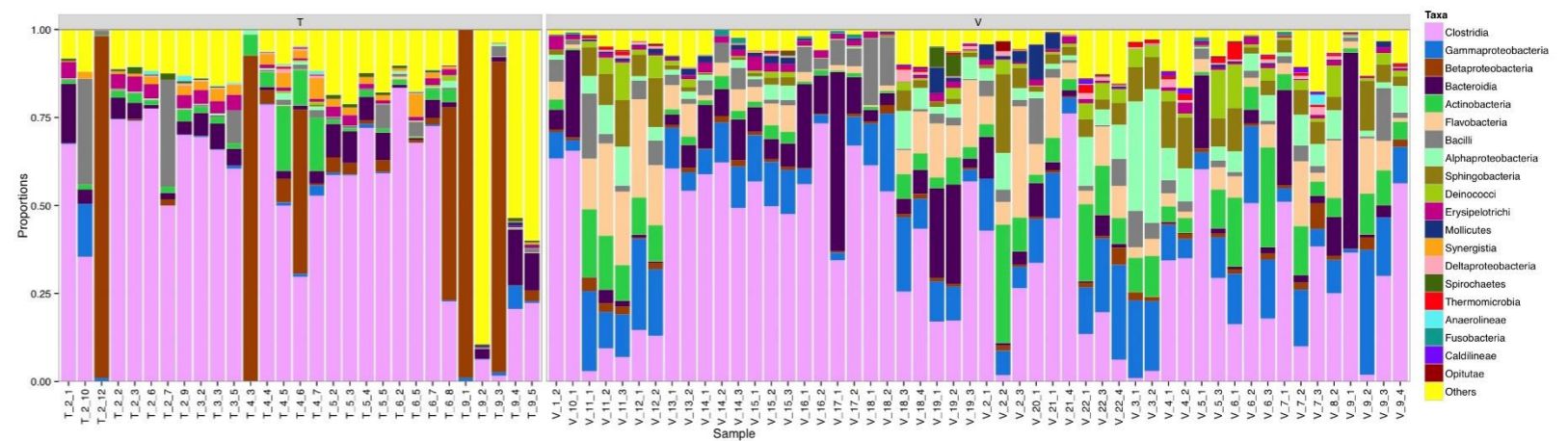
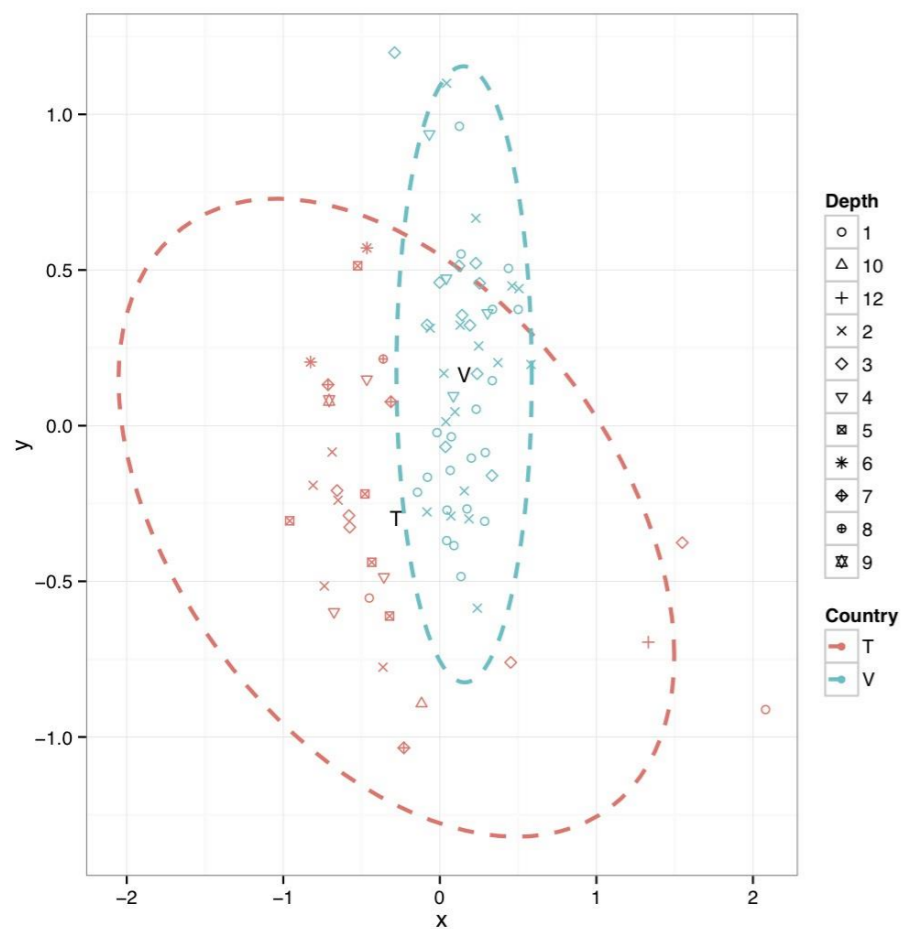




# R package example

## Multivariate Analysis of Ecological Communities in R: vegan tutorial

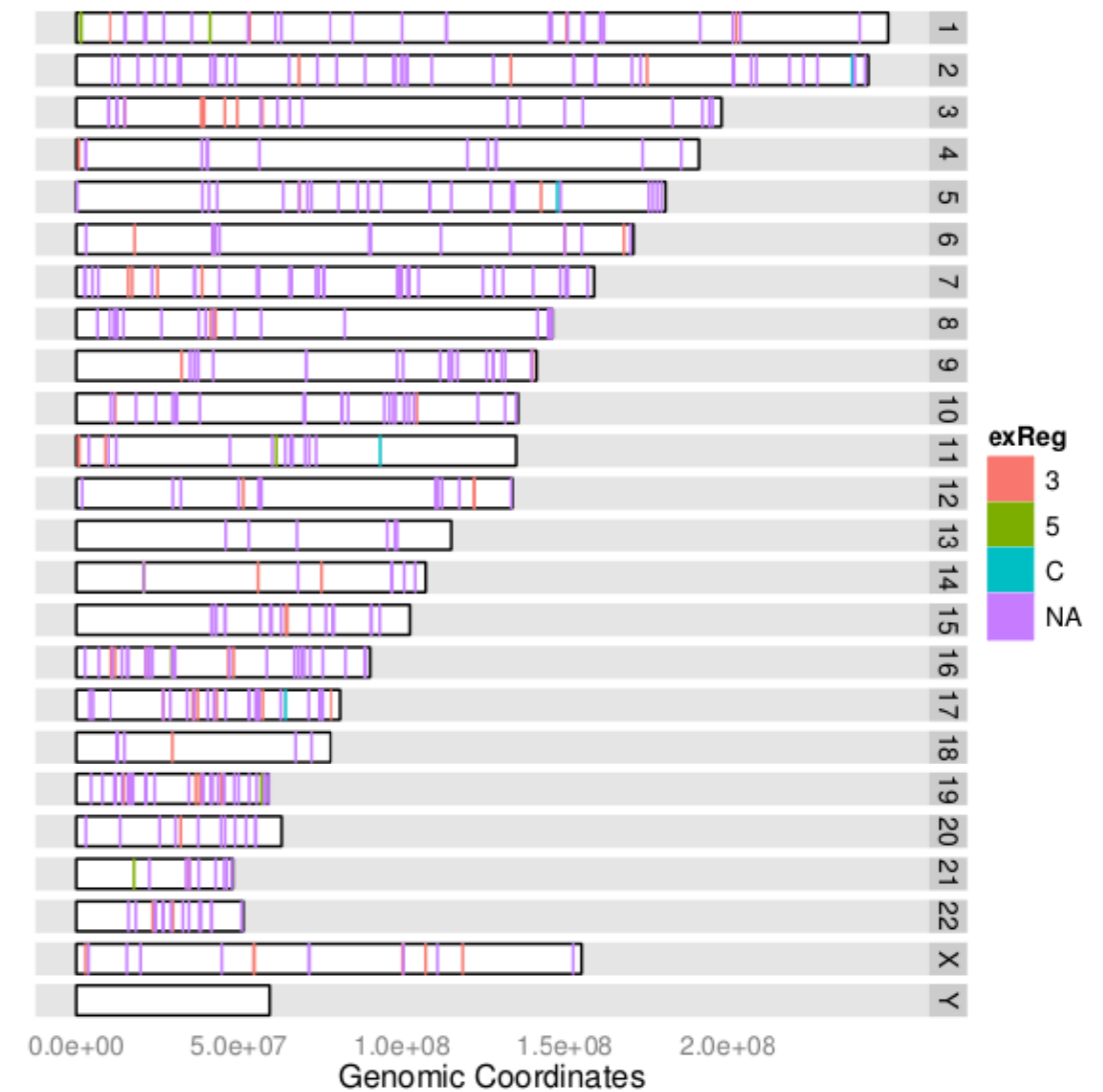
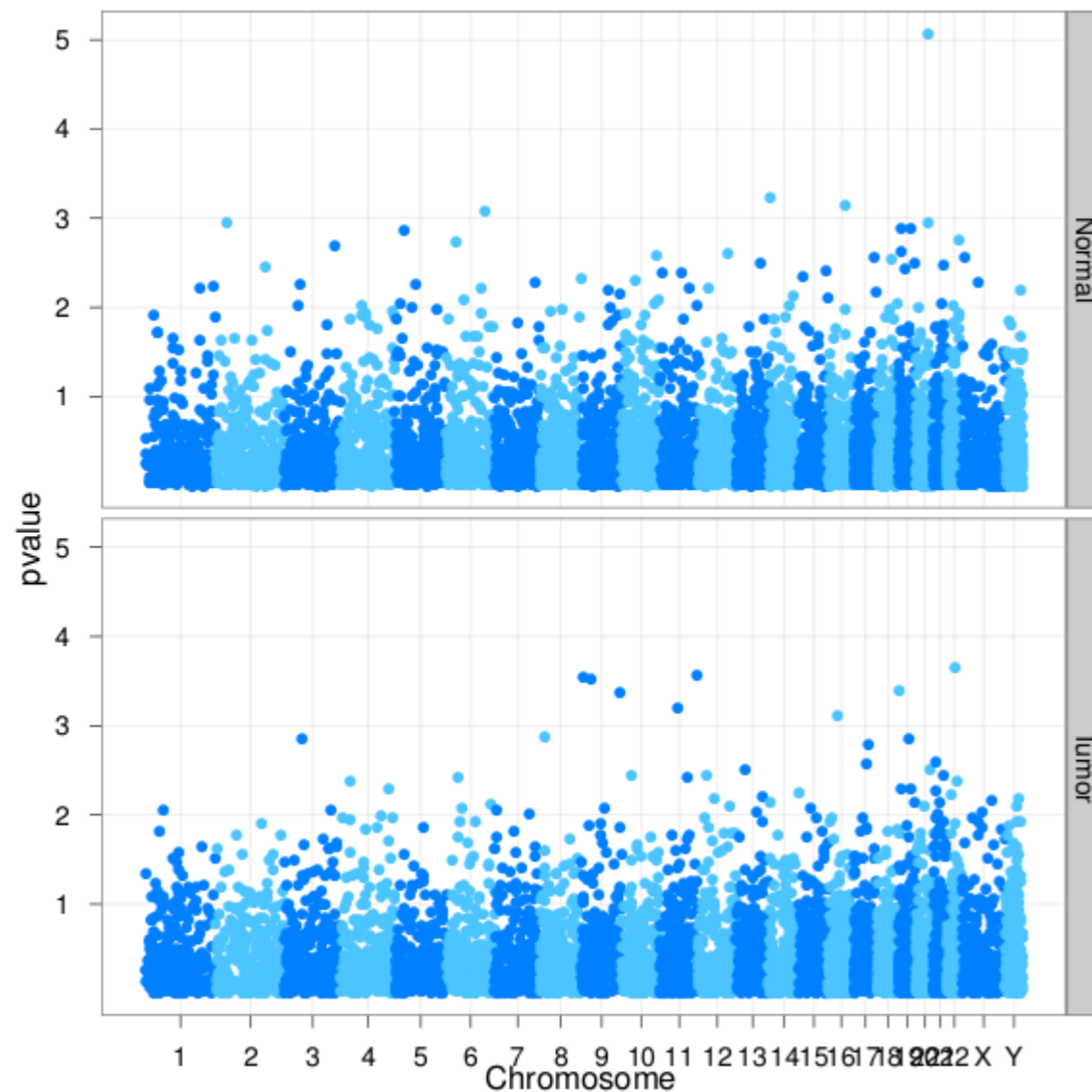
Jari Oksanen



# Bioconductor package example

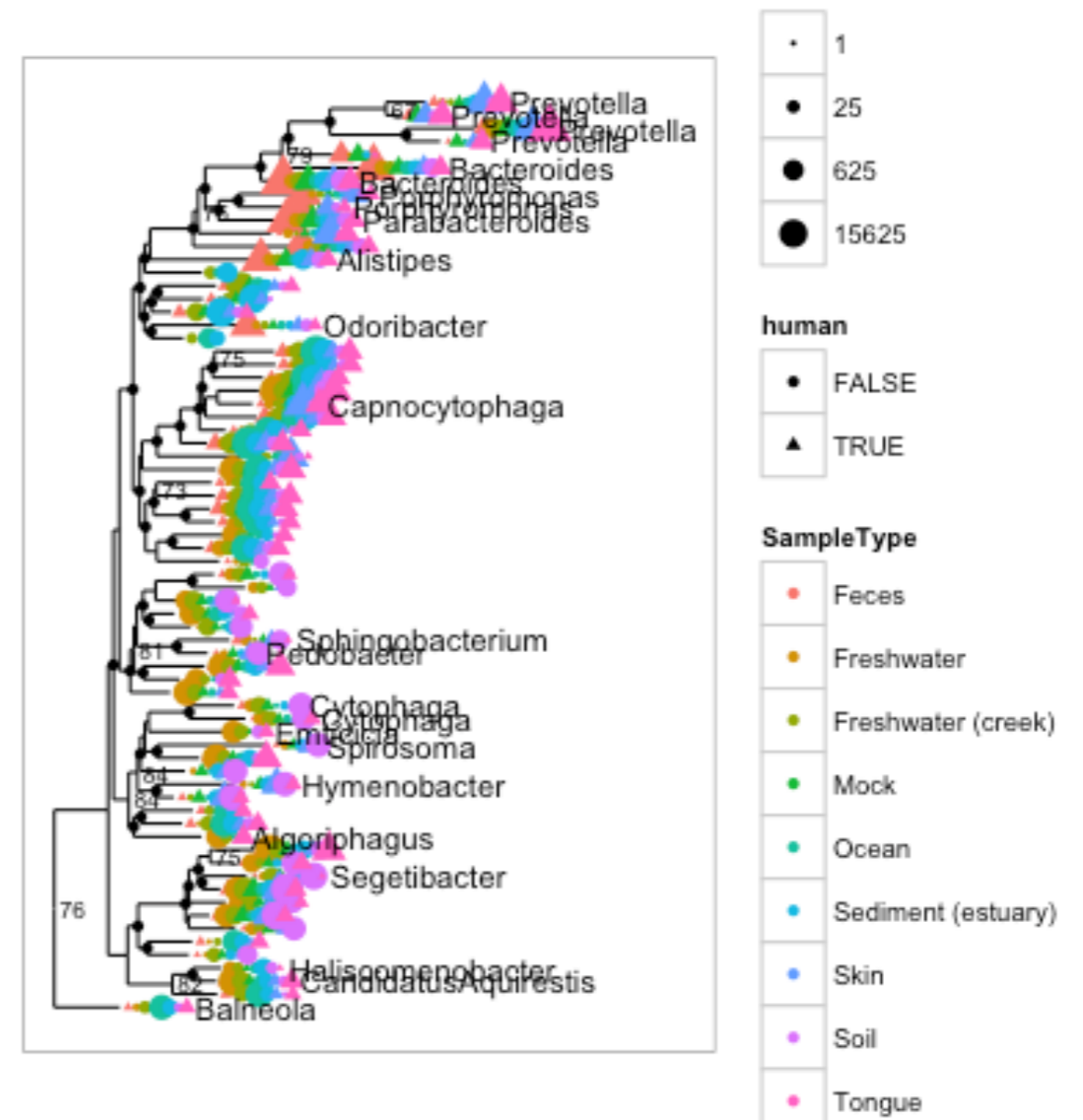
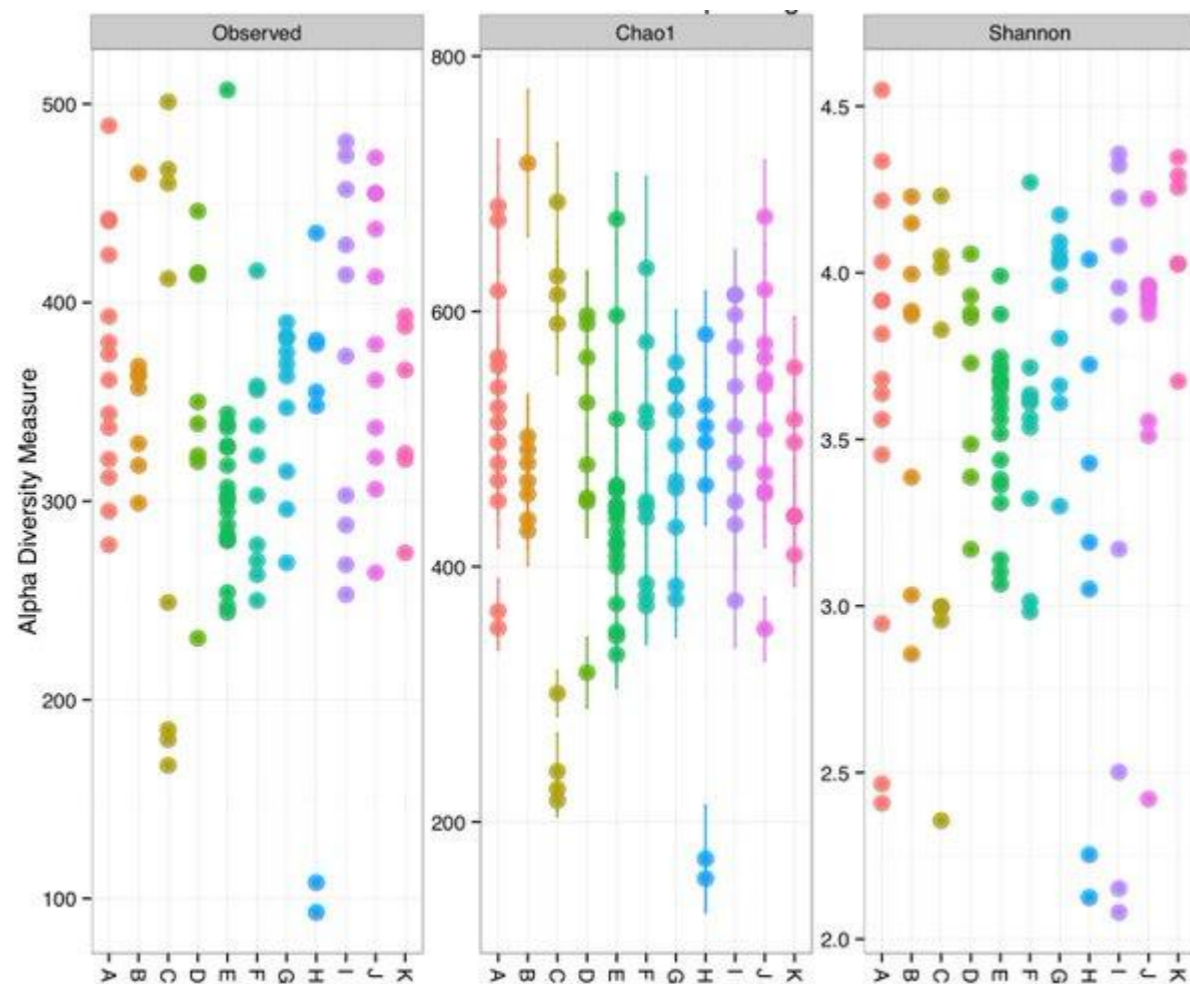
*ggbio*: visualization toolkits for genomic data

Tengfei Yin<sup>1</sup>



# Bioconductor package example

## phyloseq



<http://joey711.github.io/phyloseq/>

# Exercises

1. Introduction to R  
– done
2. Introduction to ggplot2 and RMarkdown
3. Introduction to plotting with PhyloSeq

