

Introducing R and the RStudio IDE

Alana Schick, IMC Bioinformatics, University of
Calgary

a.schick@ucalgary.ca, [@alana_schick](https://twitter.com/alana_schick)



UNIVERSITY OF CALGARY
CUMMING SCHOOL OF MEDICINE



UNIVERSITY OF
CALGARY

International
Microbiome
Centre

What is R?

- R (since 1995) is a programming language developed to teach statistics
- R is open source (ie. free), widely used, flexible, and powerful





Packages for everything

`car` – car's `Anova` function is popular for making type II and type III Anova tables

`mgcv` – Generalized Additive Models

`lme4/nlme` – Linear and non-linear mixed effects models

`randomForest` – Random forest methods from machine learning

`multcomp` – Tools for multiple comparison testing

`vcd` – Visualization tools and tests for categorical data

`glmnet` – Lasso and elastic-net regression methods with cross validation

`survival` – Tools for survival analysis

`caret` – Tools for training regression and classification models



Packages for everything

```
> install.packages("meme")
> library(meme)
> meme(img, "data science",
  "so hot right now")
```



What is R? What is RStudio?

- R (since 1995) is a programming language developed to teach statistics
- R is open source (ie. free), widely used, flexible, and powerful
- RStudio is an Integrated Development Environment (IDE) that allows users to run R in a more user-friendly way



What is R? What is RStudio?

R: Engine



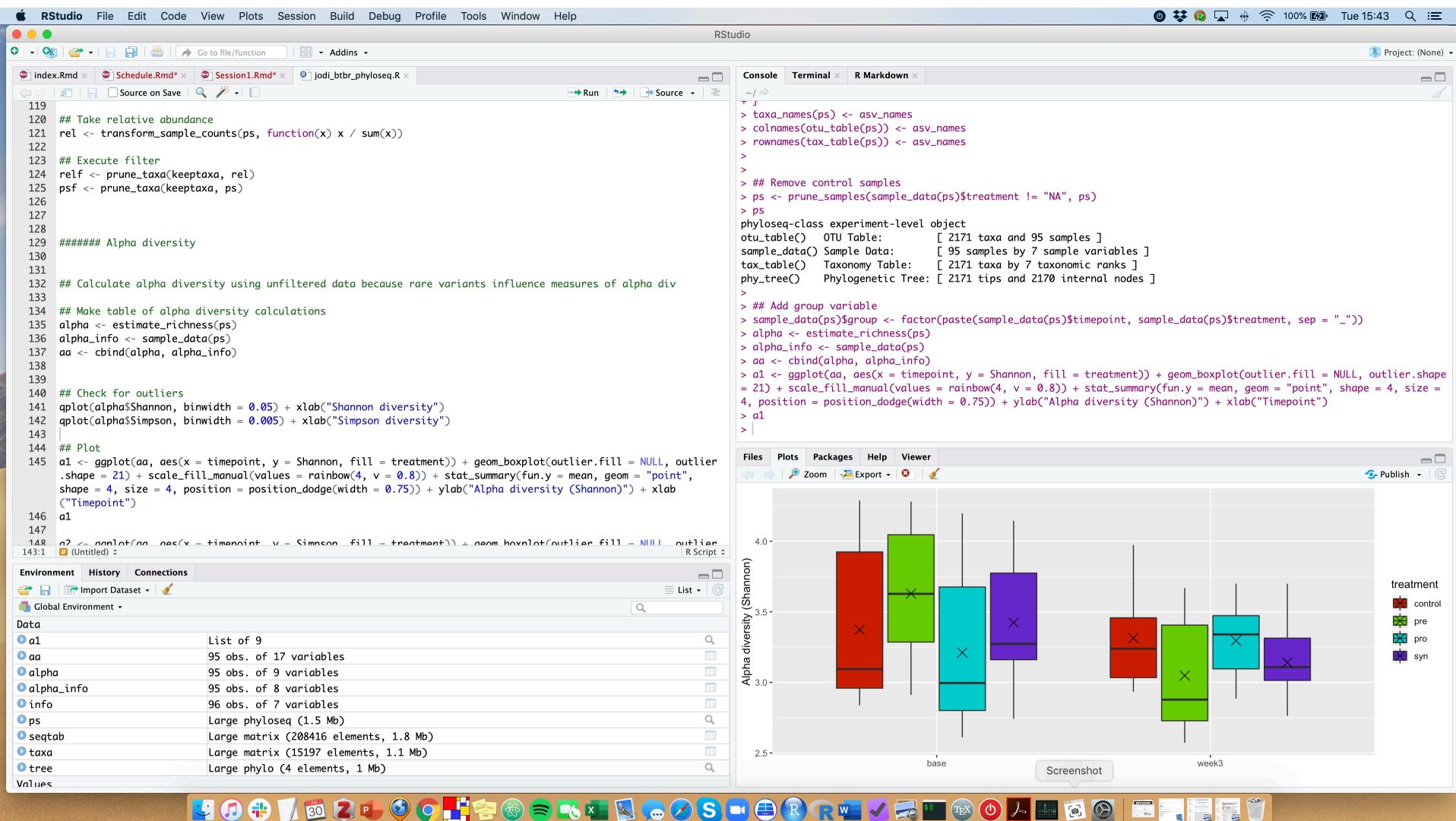
RStudio: Dashboard



<http://moderndive.com/>



RStudio looks like this



RStudio screen

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

Project: (None)

index.Rmd Schedule.Rmd* Session1.Rmd* jodi_btbr_phlyoseq.R

Source on Save Run Source

Script

```
119  
120 ## Take relative abundance  
121 rel <- transform_sample_counts(ps, function(x) x / sum(x))  
122  
123 ## Execute filter  
124 relf <- prune_taxa(keptaxa, rel)  
125 psf <- prune_taxa(keptaxa, ps)  
126  
127  
128 ##### Alpha diversity  
129  
130 ## Calculate alpha diversity using unfiltered data because rare variants influence measures of alpha div  
131  
132 ## Make table of alpha diversity calculations  
133 alpha <- estimate_richness(ps)  
134 alpha_info <- sample_data(ps)  
135 aa <- cbind(alpha, alpha_info)  
136  
137 ## Check for outliers  
138 qplot(alpha$Shannon, binwidth = 0.05) + xlab("Shannon diversity")  
139 qplot(alpha$Simpson, binwidth = 0.005) + xlab("Simpson diversity")  
140  
141 ## Plot  
142 a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")  
143 a1  
144  
145 a2 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Simpson)") + xlab("Timepoint")  
146 a2  
147  
148 (Untitled): R Script
```

Console Terminal R Markdown

```
> s  
> taxa_names(ps) <- asv_names  
> colnames(otu_table(ps)) <- asv_names  
> rownames(tax_table(ps)) <- asv_names  
>  
>  
> ## Remove control samples  
> ps <- prune_samples(sample_data(ps)$tre  
> ps  
phyloseq-class experiment-level object  
otu_table() OTU Table: [ 2171 taxa and 95 samples ]  
sample_data() Sample Data: [ 95 samples by 7 sample variables ]  
tax_table() Taxonomy Table: [ 2171 taxa by 7 taxonomic ranks ]  
phy_tree() Phylogenetic Tree: [ 2171 tips and 2170 internal nodes ]  
>  
> ## Add group variable  
> sample_data(ps)$group <- factor(paste(sample_data(ps)$timepoint, sample_data(ps)$treatment, sep = "-"))  
> alpha <- estimate_richness(ps)  
> alpha_info <- sample_data(ps)  
> aa <- cbind(alpha, alpha_info)  
> a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")  
> a1
```

Files Plots Packages Help Viewer

Plot

Environment

Import Dataset

Global Environment

Data

- a1 List of 9
- aa 95 obs. of 17 variables
- alpha 95 obs. of 9 variables
- alpha_info 95 obs. of 8 variables
- info 96 obs. of 7 variables
- ps Large phyloseq (1.5 Mb)
- seqtab Large matrix (208416 elements, 1.8 Mb)
- taxa Large matrix (15197 elements, 1.1 Mb)
- tree Large phylo (4 elements, 1 Mb)

Values

Screenshot

Console

Plots

treatment

- control
- pre
- pro
- syn

RStudio screen

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

Project: (None)

index.Rmd Schedule.Rmd Session1.Rmd jodi_btbr_phlyoseq.R

Run Source

Script

```
119  
120 ## Take relative abundance  
121 rel <- transform_sample_counts(ps, function(x) x / sum(x))  
122  
123 ## Execute filter  
124 relf <- prune_taxa(keuptaxa, rel)  
125 psf <- prune_taxa(keuptaxa, ps)  
126  
127  
128 ##### Alpha diversity  
129  
130 ## Calculate alpha diversity using unfiltered data because rare variants influence measures of diversity  
131  
132 ## Make table of alpha diversity calculations  
133 alpha <- estimate_richness(psf)  
134 alpha_info <- sample_data(psf)  
135 aa <- cbind(alpha, alpha_info)  
136  
137 ## Check for outliers  
138 qplot(alpha$Shannon, x = timepoint, y = Shanno, fill = treatment) + xlab("Shannon diversity")  
139 qplot(alpha$Simpson, x = timepoint, y = Simpson, fill = treatment) + xlab("Simpson diversity")  
140  
141 ## Plot  
142 a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")  
143 a1  
144 o1 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Simpson)") + xlab("Timepoint")  
145 o1  
146 o2 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Simpson)") + xlab("Timepoint")  
147 o2  
148 o3 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Simpson)") + xlab("Timepoint")  
149 o3  
150
```

Console Terminal R Markdown

```
> s  
> taxa_names(ps) <- asv_names  
> colnames(otu_table(ps)) <- asv_names  
> rownames(tax_table(ps)) <- asv_names  
>  
>  
> ## Remove control samples  
> ps <- prune_samples(sample_data(ps)$tre  
> ps  
phyloseq-class experiment-level object  
otu_table() OTU Table: [ 2171 taxa and 95 samples ]  
sample_data() Sample Data: [ 95 samples by 7 sample variables ]  
tax_table() Taxonomy Table: [ 2171 taxa by 7 taxonomic ranks ]  
phy_tree() Phylogenetic Tree: [ 2171 tips and 2170 internal nodes ]  
>  
> ## Add group variable  
> sample_data(ps)$group <- factor(paste(sample_data(ps)$timepoint, sample_data(ps)$treatment, sep = "-"))  
> alpha <- estimate_richness(ps)  
> alpha_info <- sample_data(ps)  
> aa <- cbind(alpha, alpha_info)  
> a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")  
> a1  
>
```

Files Plots Packages Help Viewer

Plots

Environment

Global Environment

Data

- a1 List of 9
- aa 95 obs. of 17 variables
- alpha 95 obs. of 9 variables
- alpha_info 95 obs. of 8 variables
- info 96 obs. of 7 variables
- ps Large phyloseq (1.5 Mb)
- seqtab Large matrix (208416 elements, 1.8 Mb)
- taxa Large matrix (15197 elements, 1.1 Mb)
- tree Large phylo (4 elements, 1 Mb)

Values

Screenshot

control pre pro syn

RStudio screen

The screenshot shows the RStudio interface on a Mac OS X desktop. The window title is "RStudio". The menu bar includes "File", "Edit", "Code", "View", "Plots", "Session", "Build", "Debug", "Profile", "Tools", "Window", and "Help". The status bar at the bottom right shows "100% 16:27 Tue 16:27".

Script: The left pane displays an R script with code for phylogenetic analysis. It includes sections for relative abundance, filtering, alpha diversity calculations, and creating a phyloseq object. A box labeled "Script" highlights the script area.

```
119 ## Take relative abundance
120 rel <- transform_sample_counts(ps, function(x) x / sum(x))
121
122 ## Execute filter
123 relf <- prune_taxa(keptaxa, rel)
124 psf <- prune_taxa(keptaxa, ps)
125
126
127
128 ###### Alpha diversity
129
130
131
132 ## Calculate alpha diversity using unfiltered data because rare variants influence measures of alpha div
133
134 ## Make table of alpha diversity calculations
135 alpha <- estimate_richness(ps)
136 alpha_info <- sample_data(ps)
137 aa <- cbind(alpha, alpha_info)
138
138:1 (Untitled):1
```

Console: The top right pane shows the R console output. A box labeled "Console" highlights the output area.

```
> taxa_names(ps) <- asv_names
> colnames(otu_table(ps)) <- asv_names
> rownames(tax_table(ps)) <- asv_names
>
>
> ## Remove control samples
> ps <- prune_samples(sample_data(ps)$tre
> ps
phyloseq-class experiment-level object
otu_table() OTU Table: [ 2171 taxa and 95 samples ]
sample_data() Sample Data: [ 95 samples by 7 sample variables ]
tax_table() Taxonomy Table: [ 2171 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 2171 tips and 2170 internal nodes ]
>
> ## Add group variable
> sample_data(ps)$group <- factor(paste(sample_data(ps)$timepoint, sample_data(ps)$treatment, sep = "_"))
> alpha <- estimate_richness(ps)
> alpha_info <- sample_data(ps)
> aa <- cbind(alpha, alpha_info)
> a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape =
21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size =
4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")
> a1
> |
```

History: The bottom left pane shows the history of R commands run in the session. A box labeled "History" highlights the history area.

```
# MARCH Sample names
rownames(seqtab)
# Make a phyloseq object
ps <- phyloseq(otu_table(seqtab, taxa_are_rows=FALSE), sample_data(info), tax_table(taxa))
## Make a tree and add the tree to a new phyloseq object
tree <- rtree(ntaxa(ps), rooted = TRUE, tip.label = taxa_names(ps))
ps <- phyloseq(otu_table(seqtab, taxa_are_rows=FALSE), sample_data(info), tax_table(taxa), phy_tree(tree))
asv_names <- vector(dim(otu_table(ps))[2], mode = "character")
for (i in 1:dim(otu_table(ps))[2]){
asv_names[i] <- paste("ASV", i, sep = "_")
}
taxa_names(ps) <- asv_names
colnames(otu_table(ps)) <- asv_names
rownames(tax_table(ps)) <- asv_names
## Remove control samples
ps <- prune_samples(sample_data(ps)$treatment != "NA", ps)
ps
## Add group variable
sample_data(ps)$group <- factor(paste(sample_data(ps)$timepoint, sample_data(ps)$treatment, sep = "_"))
alpha <- estimate_richness(ps)
alpha_info <- sample_data(ps)
aa <- cbind(alpha, alpha_info)
a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape =
a1
```

Files: The bottom right pane shows the file browser. A box labeled "Files" highlights the file browser area.

Files	Plots	Packages	Help	Viewer
New Folder	Home	Applications	Help	Viewer
1.1 KB	Feb 11, 2019, 5:41 PM			

Screenshot

The dock at the bottom of the screen contains icons for various Mac OS X applications, including Mail, Safari, and Finder.

How to R – 2 ways

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

Project: (None)

index.Rmd Schedule.Rmd* Session1.Rmd* jodi_btbr_phlyoseq.R

Source on Save Run Source

Script

Console

1. Typing and executing commands interactively here.

2. Using and saving scripts (ie. plain text files of code).

```
## Take relative abundance
rel <- transform_sample_counts(ps, function(x) x / sum(x))

## Execute filter
relf <- prune_taxa(keptaxa, rel)
psf <- prune_taxa(keptaxa, ps)

##### Alpha diversity

## Calculate alpha diversity using Shannon index
alpha <- estimate_richness(ps)
alpha_info <- sample_data(ps)
aa <- cbind(alpha, alpha_info)

## Check for outliers
qplot(alpha$Shannon, binwidth = 0.05)
qplot(alpha$Simpson, binwidth = 0.05)

## Plot
a1 <- ggplot(aa, aes(x = timenpoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timenpoint")
a1
a2 <- ggplot(aa, aes(x = timenpoint, y = Simpson, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Simpson)") + xlab("Timenpoint")
a2
```

Files Plots Packages Help Viewer

Alpha diversity (Shannon)

treatment

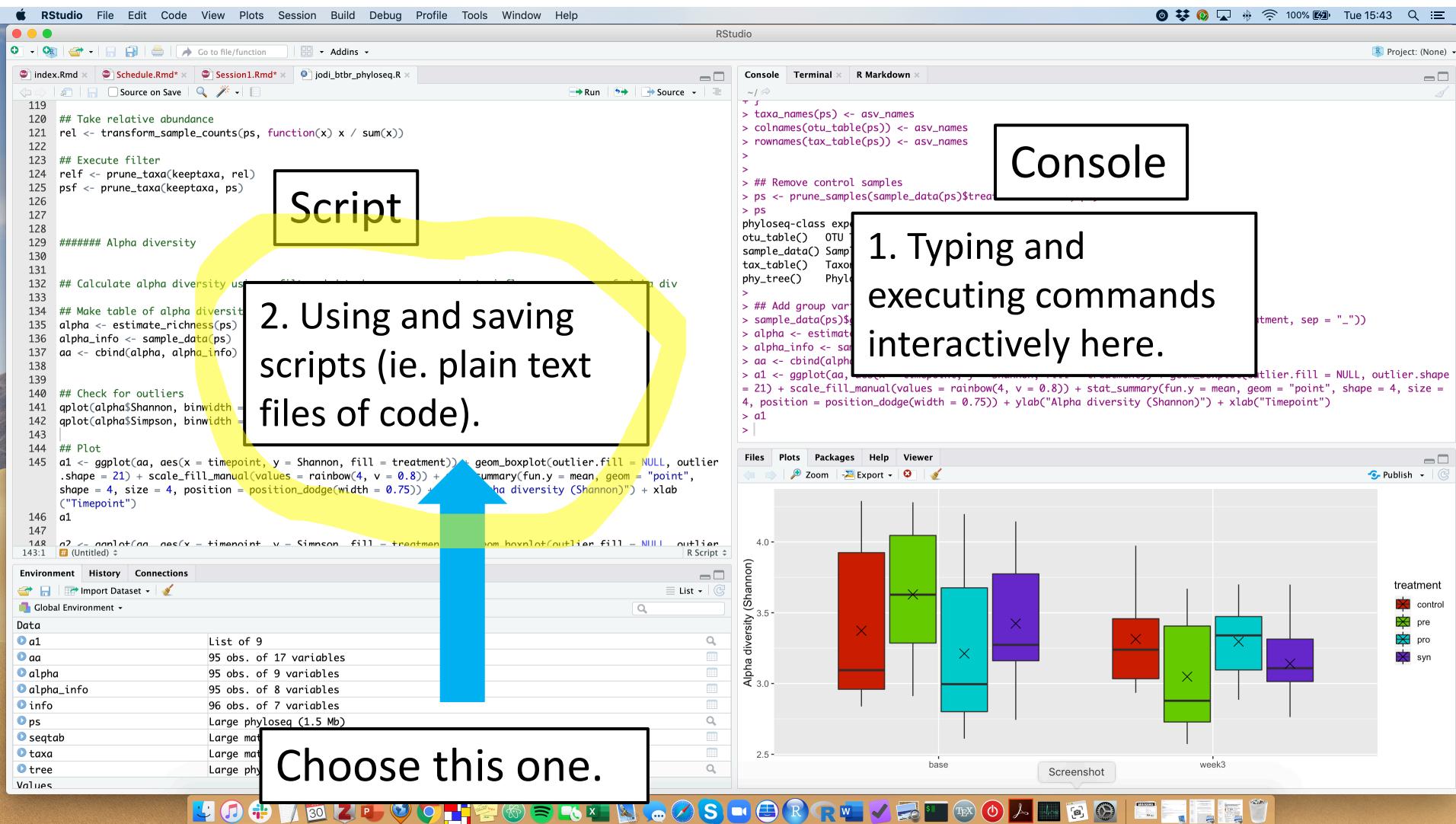
- control
- pre
- pro
- syn

base week3

Screenshot

100% Tue 15:43

How to R – 2 ways



Scripts

```
1 # jodi_btbr project, Alana Schick, April 2019
2 # This is a script to analyze the output tables of the DADA2 workflow in phyloseq
3 # Have two output files from dada2 - a sequence table and a taxonomy table, read them into R using the readRDS() function
4 # The formatted sample metadata is in a table called "jodi_btbr_metadata.txt"
5
6 library(phyloseq)
7 #packageVersion("phyloseq")
8 library(ggplot2)
9 #packageVersion("ggplot2")
10 library(ape)
11 library(viridis)
12 library(grid)
13 library(gridExtra)
14 library(reshape2)
15 library(DESeq2)
16 library(fields)
17 library(vegan)
18 library(ggpubr)
19 library(plyr)
20 library(RColorBrewer)
21
22 path_to_project <- "/Users/alanaschick/Dropbox/Jodi_BTBR"
23
24 # Read in files
25 seqtab <- readRDS(file.path(path_to_project, "seqtab.rds"))
26 taxa <- readRDS(file.path(path_to_project, "taxa.rds"))
27 info <- read.table(file.path(path_to_project, "jodi_btbr_metadata.txt"), header = TRUE)
28
29 # Match sample names
30 rownames(info) <- rownames(seqtab)
31
32 # Make a phyloseq object
```

Everything in the console will be forgotten when you close the session.

Scripts are saved, keeping a complete record of the commands you ran so you can run them again (ie. completely reproducible).

Can execute parts of this or the entire script.

Scripts - commenting

```
1 # jodi_btbr project, Alana Schick, April 2019
2 # This is a script to analyze the output tables of the DADA2 workflow in phyloseq
3 # Have two output files from dada2 - a sequence table and a taxonomy table, read them into R using the readRDS()
4 # The formatted sample metadata is in a table called "jodi_btbr_metadata.txt"
5
6 library(phyloseq)
7 #packageVersion("phyloseq")
8 library(ggplot2)
9 #packageVersion("ggplot2")
10 library(ape)
11 library(viridis)
12 library(grid)
13 library(gridExtra)
14 library(reshape2)
15 library(DESeq2)
16 library(fields)
17 library(vegan)
18 library(ggpubr)
19 library(plyr)
20 library(RColorBrewer)
21
22 path_to_project <- "/Users/alanaschick/Dropbox/time/projects/jodi_btbr"
23
24 # Read in files
25 seqtab <- readRDS(file.path(path_to_project, "results/seqtan_final.rds"))
26 taxa <- readRDS(file.path(path_to_project, "results/taxa_final.rds"))
27 info <- read.table(file.path(path_to_project, "jodi_btbr_metadata2.txt"), header = TRUE)
28
29 # Match sample names
30 rownames(info) <- rownames(seqtan)
31
32 # Make a phyloseq object
```

Comment out lines of your scripts by using the `#` symbol. R will not run these.

Be descriptive. You will not remember what you did a year later.

Packages

```
1 # jodi_btbr project, Alana Schick, April 2019
2 # This is a script to analyze the output tables of the DADA2 workflow in phyloseq
3 # Have two output files from dada2 - a sequence table and a taxonomy table, read them into R using the readRDS() function
4 # The formatted sample metadata is in a table called "jodi_btbr_metadata.txt"
5
6 library(phyloseq)
7 #packageVersion("phyloseq")
8 library(ggplot2)
9 #packageVersion("ggplot2")
10 library(ape)
11 library(viridis)
12 library(grid)
13 library(gridExtra)
14 library(reshape2)
15 library(DESeq2)
16 library(fields)
17 library(vegan)
18 library(ggpubr)
19 library(plyr)
20 library(RColorBrewer)
21
22 path_to_project <- "/Users/alanaschick/"
23
24 # Read in files
25 seqtab <- readRDS(file.path(path_to_project, "results/seqtan_final.rds"))
26 taxa <- readRDS(file.path(path_to_project, "results/taxa_final.rds"))
27 info <- read.table(file.path(path_to_project, "jodi_btbr_metadata2.txt"), header = TRUE)
28
29 # Match sample names
30 rownames(info) <- rownames(seqtan)
31
32 # Make a phyloseq object
```

Packages are collections of R functions developed for a specific task.

Packages need to first be installed on your computer.

After installed, `library()` is the command used to load a package.

Packages

A screenshot of the RStudio interface showing a code editor, console, and plot area. A large blue arrow points from the top left towards the 'Tools' menu.

The code editor shows R script content:

```
119
120 ## Take relative abundance
121 rel <- transform_sample_counts(ps, function(x) x / sum(x))
122
123 ## Execute filter
124 relf <- prune_taxa(keptaxa, rel)
125 psf <- prune_taxa(keptaxa, ps)
126
127
128 ###### Alpha diversity
129
130 ## Calculate alpha diversity using unfiltered data because
131
132 ## Make table of alpha diversity calculations
133 alpha <- estimate_richness(ps)
134 alpha.info <- sample_data(ps)
135 aa <- cbind(alpha, alpha.info)
136
137 ## Check for outliers
138 qplot(alpha$Shannon, binwidth = 0.05) + xlab("Shannon diversity")
139 qplot(alpha$Simpson, binwidth = 0.005) + xlab("Simpson diversity")
140
141 ## Plot
142 o1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment,
143 .shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) +
144 shape = 4, size = 4, position = position_dodge(width = 0.75)) +
145 ("Timepoint")
146 o1
147
148 o2 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment,
149 .shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) +
150 shape = 4, size = 4, position = position_dodge(width = 0.75)) +
151 ("Timepoint")
152 o2
```

The console shows R session output:

```
119
120 ## Take relative abundance
121 rel <- transform_sample_counts(ps, function(x) x / sum(x))
122
123 ## Execute filter
124 relf <- prune_taxa(keptaxa, rel)
125 psf <- prune_taxa(keptaxa, ps)
126
127
128 ###### Alpha diversity
129
130 ## Calculate alpha diversity using unfiltered data because
131
132 ## Make table of alpha diversity calculations
133 alpha <- estimate_richness(ps)
134 alpha.info <- sample_data(ps)
135 aa <- cbind(alpha, alpha.info)
136
137 ## Check for outliers
138 qplot(alpha$Shannon, binwidth = 0.05) + xlab("Shannon diversity")
139 qplot(alpha$Simpson, binwidth = 0.005) + xlab("Simpson diversity")
140
141 ## Plot
142 o1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment,
143 .shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) +
144 shape = 4, size = 4, position = position_dodge(width = 0.75)) +
145 ("Timepoint")
146 o1
147
148 o2 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment,
149 .shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) +
150 shape = 4, size = 4, position = position_dodge(width = 0.75)) +
151 ("Timepoint")
```

The plot area displays two boxplots side-by-side. The x-axis categories are 'base' and 'week3'. The y-axis ranges from 2.5 to 4.0. The legend indicates four treatment groups: 'control' (red), 'pre' (green), 'pro' (cyan), and 'syn' (purple). Each boxplot shows the median, quartiles, and whiskers, with individual data points overlaid as 'x' marks.

A modal dialog box titled 'Install Packages' is open in the foreground:

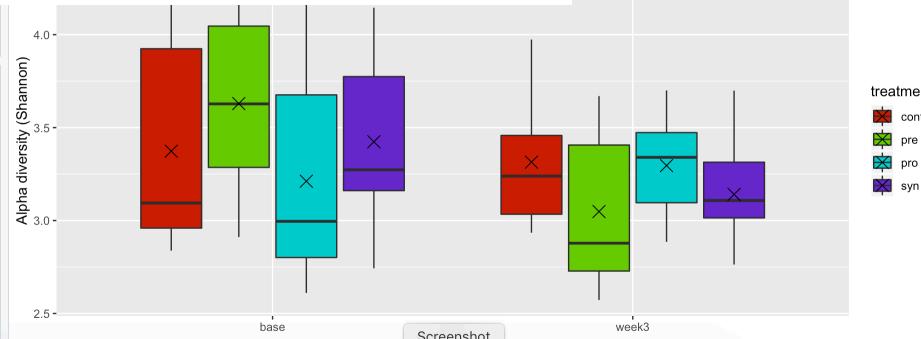
- Install from:** Repository (CRAN)
- Packages (separate multiple with space or comma):** (empty input field)
- Install to Library:** /Library/Frameworks/R.framework/Versions/3.6/Resources/library
- Install dependencies:**
- Buttons:** Install, Cancel

Working directory

Every time you open RStudio, it goes to a default directory, usually your home directory.

You can use the command **setwd()** to change the working directory.

```
setwd("home/aschick/projects/workshop")
```



Working directory

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help RStudio Addins Go to file/function Addins ~ Run Source ~ j Project: (None) 100% Tue 15:43

```
119 ## Take relative abundance
120 rel <- transform_sample_counts(ps, function(x)
121
122 ## Execute filter
123 relf <- prune_taxa(keptaxa, rel)
124 psf <- prune_taxa(keptaxa, ps)
125
126
127
128 ###### Alpha diversity
129
130 ## Calculate alpha diversity using unfiltered data
131
132 ## Make table of alpha diversity calculations
133 alpha <- estimate_richness(psf)
134 alpha_info <- sample_data(psf)
135 aa <- cbind(alpha, alpha_info)
136
137
138 ## Check for outliers
139 qplot(alpha$Shannon, binwidth = 0.05) + xlab("Timepoint")
140 qplot(alpha$Simpson, binwidth = 0.005) + xlab("Timepoint")
141
142 ## Plot
143 o1 <- ggplot(aa, aes(x = timepoint, y = Shannon))
144 o1 + geom_boxplot(outlier.size = 0.75) + stat_summary(fun.y = mean, geom = "point",
145 .shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point",
146 shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab(
147 "Timepoint")
148 o2 <- ggplot(aa, aes(x = timepoint, y = Simpson))
149 o2 + geom_boxplot(outlier.size = 0.75) + stat_summary(fun.y = mean, geom = "point",
150 .shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point",
151 shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Simpson)") + xlab(
152 "Timepoint")
153
```

However: you may want to run your script on a different computer with a different directory structure where that directory does not exist.

Or you may want to work in multiple directories.

Alpha diversity (Shannon)

control pre pro syn

Screenshot

RStudio Project

File > New Project...

Clicking on New Directory will create an RStudio Project.

This directory will have all the data, files, plots, etc. for that project as well as a .Rproj file.

The screenshot shows the RStudio interface with a blue arrow pointing to the 'File' menu. A callout box highlights 'File > New Project...'. Below it, another callout box highlights the text 'Clicking on New Directory will create an RStudio Project.' and 'This directory will have all the data, files, plots, etc. for that project as well as a .Rproj file.' In the center, a 'New Project' dialog box is open, showing three options: 'New Directory' (Start a project in a brand new working directory), 'Existing Directory' (Associate a project with an existing working directory), and 'Version Control' (Checkout a project from a version control repository). At the bottom right of the dialog, there is a 'Screenshot' button. The background shows the RStudio environment with code in the editor and a plot in the viewer.

Error messages

```
Console Terminal x R Markdown x ~ / >
> ## Remove control samples
> ps <- prune_samples(sample_data(ps)$treatment != "NA", ps)
> ps
phyloseq-class experiment-level object
otu_table() OTU Table: [ 2171 taxa and 95 samples ]
sample_data() Sample Data: [ 95 samples by 7 sample variables ]
tax_table() Taxonomy Table: [ 2171 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 2171 tips and 2170 internal nodes ]
>
> ## Add group variable
> sample_data(ps)$group <- factor(paste(sample_data(ps)$timepoint, sample_data(ps)$treatment, sep = "_"))
> alpha <- estimate_richness(ps)
> alpha_info <- sample_data(ps)
> aa <- cbind(alpha, alpha_info)
> a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")
> a1
> ord1 <- ordinate(relf, method = "NMDS", distance = "bray")
Error in ordinate(relf, method = "NMDS", distance = "bray") :
  object 'relef' not found
> b1 <- plot_ordination(relf, ord1, color = "timepoint", shape = "treatment", title = "NMDS - Bray") + scale_colour_manual(values = viridis(3))
Error in plot_ordination(relf, ord1, color = "timepoint", shape = "treatment", :
  object 'relef' not found
> b1
Error: object 'b1' not found
>
```

Error messages

Console Terminal × R Markdown ×

```
>  
> ## Remove control samples  
> ps <- prune_samples(sample_data(ps)$tr  
> ps  
phyloseq-class experiment-level object  
otu_table() OTU Table: [ 2171  
sample_data() Sample Data: [ 95 sa  
tax_table() Taxonomy Table: [ 2171  
phy_tree() Phylogenetic Tree: [ 2171  
>  
> ## Add group variable  
> sample_data(ps)$group <- factor(paste(  
> alpha <- estimate_richness(ps)  
> alpha_info <- sample_data(ps)  
> aa <- cbind(alpha, alpha_info)  
> a1 <- ggplot(aa, aes(x = timepoint, y  
= 21) + scale_fill_manual(values = rainb  
4, position = position_dodge(width = 0.7  
> a1  
> ord1 <- ordinate(relf, method = "NMDS"  
Error in ordinate(relf, method = "NMDS",  
object 'relf' not found  
> b1 <- plot_ordination(relf, ord1, col  
ual(values = viridis(3))  
Error in plot_ordination(relf, ord1, col  
object 'relf' not found  
> b1  
Error: object 'b1' not found  
>
```



```
tment, sep = "_"))  
  
tlier.fill = NULL, outlier.shape  
om = "point", shape = 4, size =  
"Timepoint")  
  
NMDS - Bray") + scale_colour_man
```

Getting help

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

Project: (None)

```
index.Rmd * Schedule.Rmd * Session1.Rmd * jodi_btbr_phlyoseq.R
```

Source on Save Run Source

119
120 ## Take relative abundance
121 rel <- transform_sample_counts(ps, function(x) x / sum(x))
122
123 ## Execute filter
124 relf <- prune_taxa(keptaxa, rel)
125 psf <- prune_taxa(keptaxa, ps)
126
127
128 ##### Alpha diversity
129
130
132 ## Calculate alpha diversity using unfiltered data because rare variants influence measures of alpha div
133
134 ## Make table of alpha diversity calculations
135 alpha <- estimate_richness(ps)
136 alpha_info <- sample_data(ps)
137 aa <- cbind(alpha, alpha_info)
138
139 ## Check for outliers
140 qplot(alpha\$Shannon, binwidth = 0.05) + xlab("Shannon diversity")
141 qplot(alpha\$Simpson, binwidth = 0.005) + xlab("Simpson diversity")
143
144 ## Plot
145 a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")
146 a1
147
148 a2 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Simpson)") + xlab("Timepoint")
143:1 (Untitled): R Script

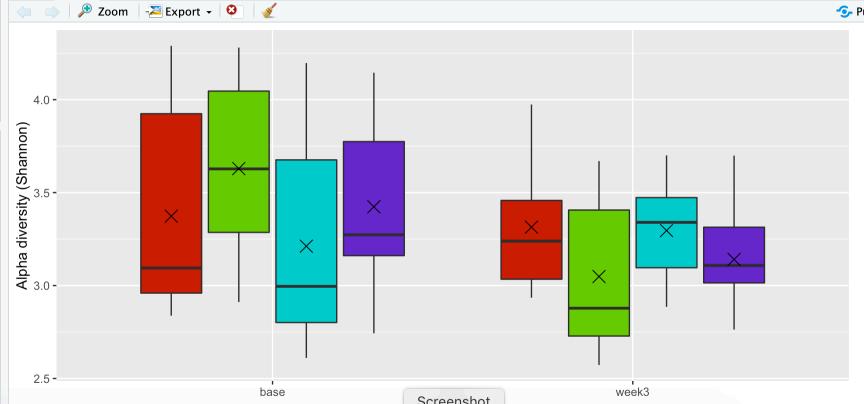
Console Terminal R Markdown

```
> s  
> taxa_names(ps) <- asv_names  
> colnames(otu_table(ps)) <- asv_names  
...$treatment != "NA", ps)  
  
> ps  
phyloseq-class experiment-level object  
@OTU_table() OTU Table: [ 2171 taxa and 95 samples ]  
@Sample_data() Sample Data: [ 95 samples by 7 sample variables ]  
@Table() Taxonomy Table: [ 2171 taxa by 7 taxonomic ranks ]  
@Phylo() Phylogenetic Tree: [ 2171 tips and 2170 internal nodes ]  
  
> sample_data(ps)$group <- factor(paste(sample_data(ps)$timepoint, sample_data(ps)$treatment, sep = "-"))  
> alpha <- estimate_richness(ps)  
> alpha_info <- sample_data(ps)  
> aa <- cbind(alpha, alpha_info)  
> a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")  
> a1
```

Files Plots Packages Help Viewer

Zoom Export

Screenshot



Alpha diversity (Shannon)

treatment

- control
- pre
- pro
- syn

base week3

Environment History Connections

Import Dataset

Global Environment

Data

- a1 List of 9
- aa 95 obs. of 17 variables
- alpha 95 obs. of 9 variables
- alpha_info 95 obs. of 8 variables
- info 96 obs. of 7 variables
- ps Large phyloseq (1.5 Mb)
- seqtab Large matrix (208416 elements, 1.8 Mb)
- taxa Large matrix (15197 elements, 1.1 Mb)
- tree Large phylo (4 elements, 1 Mb)

Values

Mac OS X Dock icons: Finder, Mail, Safari, Google Chrome, Spotify, iMovie, iPhoto, iCal, iWork, R, RStudio, Word, Tex, Arrows, etc.

Getting help

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help RStudio 100% Tue 15:43 Project: (None)

```
index.Rmd * Schedule.Rmd * Session1.Rmd * jodi_btbr_phyloseq.R * Go to file/function Addins * Run Source ~ / s > taxa_names(ps) <- asv_names > colnames(otu_table(ps)) <- asv_names
```

119
120 ## Take relative abundance
121 rel <- transform_sample_counts(ps, function(x) x / sum(x))
122
123 ## Execute filter
124 relf <- prune_taxa(keptaxa, rel)
125 psf <- prune_taxa(keptaxa, ps)
126
127
128
129 ##### Alpha diversity
130
131
132 ## Calculate alpha diversity using unfiltered data because rare vo
133
134 ## Make table of alpha diversity calculations
135 alpha <- estimate_richness(psf)
136 alpha_info <- sample_data(psf)
137 aa <- cbind(alpha, alpha_info)
138
139
140 ## Check for outliers
141 qplot(alpha\$Shannon, binwidth = 0.05) + xlab("Shannon diversity")
142 qplot(alpha\$Simpson, binwidth = 0.005) + xlab("Simpson diversity")
143
144 ## Plot
145 a1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment))
.shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + s
shape = 4, size = 4, position = position_dodge(width = 0.75)) + yl
("Timepoint")
146
147
148 a2 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 21) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4) + xlab("Timepoint")
149
150
151 (Untitled):

1) Search in Help tab

2) Type ? followed by the function name in the console (or ?? for installed packages)

> ?barplot
> ??geom_point

Getting help

- 1) Search in Help tab
- 2) Type ? followed by the function name in the console (or ?? for installed packages)
- 3) Google the error message



See website for tips and resources!

The internet will make those bad words go away



Essential

Googling the
Error Message

ORLY?

*The Practical Developer
@ThePracticalDev*

Summary and best practices

- Always save your code in R scripts
- Load packages using `library()` at the top of your script
- Write clear, readable code with comments*
- Be mindful of your working directory or location of files
- Use RStudio projects to organize scripts, data, and output

*See <http://adv-r.had.co.nz/Style.html> for tips.