

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 interface screenshot showing a workspace with multiple files, a console, a plot, and a menu bar.

File menu: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Window, Help.

Console:

```
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 a3 <- 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 a3
150
151 a4 <- 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")
152 a4
153
154 a5 <- 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")
155 a5
156
157 a6 <- 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")
158 a6
159
160 a7 <- 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")
161 a7
162
163 a8 <- 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")
164 a8
165
166 a9 <- 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")
167 a9
168
169 a10 <- 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")
170 a10
171
172 a11 <- 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")
173 a11
174
175 a12 <- 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")
176 a12
177
178 a13 <- 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")
179 a13
180
181 a14 <- 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")
182 a14
183
184 a15 <- 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")
185 a15
186
187 a16 <- 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")
188 a16
189
190 a17 <- 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")
191 a17
192
193 a18 <- 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")
194 a18
195
196 a19 <- 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")
197 a19
198
199 a20 <- 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")
200 a20
201
202 a21 <- 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")
203 a21
204
205 a22 <- 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")
206 a22
207
208 a23 <- 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")
209 a23
210
211 a24 <- 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")
212 a24
213
214 a25 <- 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")
215 a25
216
217 a26 <- 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")
216:1
```

Plots:

Environment:

- Import Dataset
- Global Environment
- Data
 - a1
 - aa
 - alpha
 - alpha_info
 - info
 - ps
 - seqtab
 - taxa
 - tree
- Values

Plots:

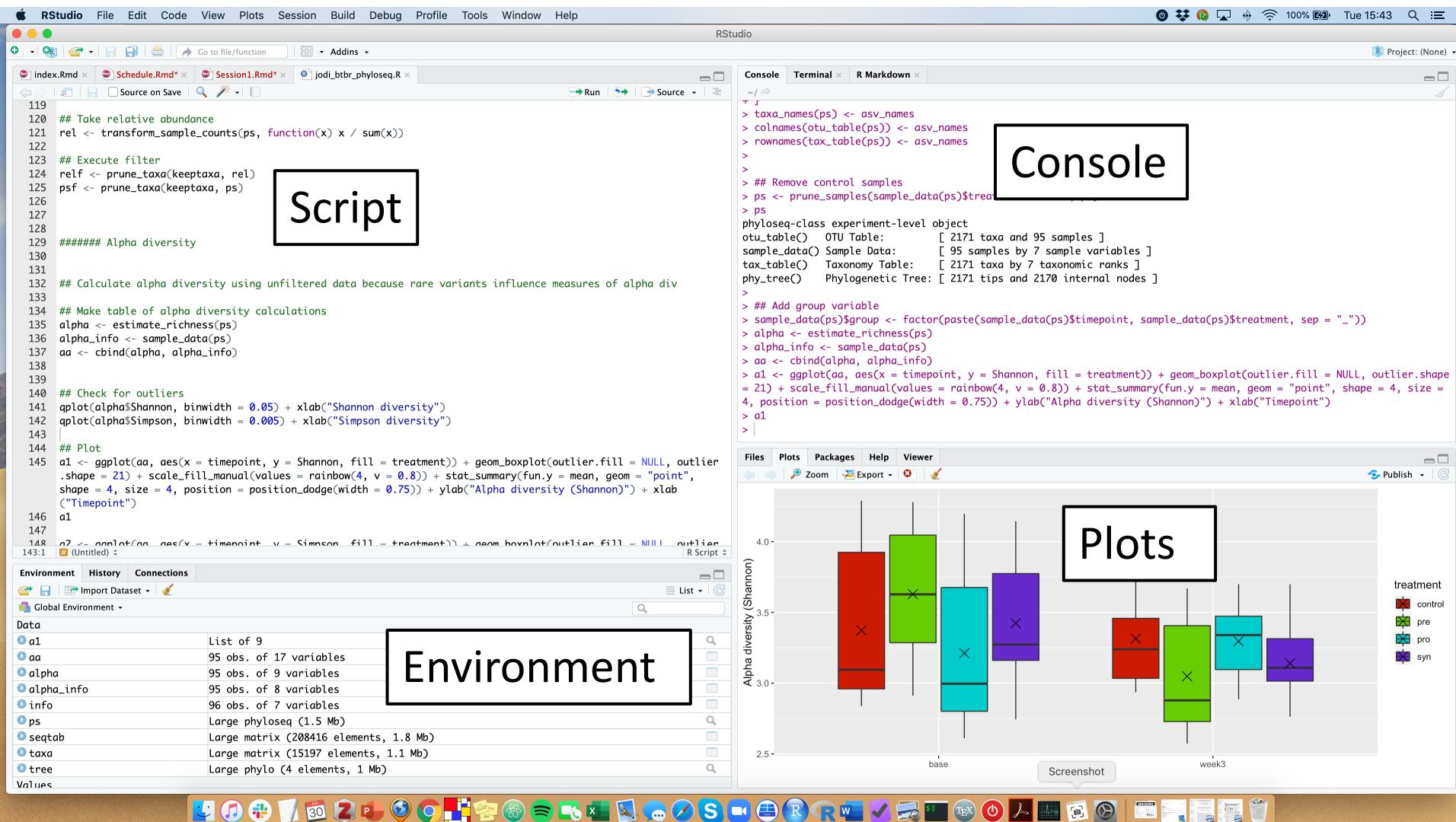
Terminal:

```
~/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)$treatment != "NA", ps)
> ps
phyloseq-class experiment-level object
otu_table() OTU Table: [ 2171 taxa and 95 samples ]
sample_data() Sample Data: [ 95 samples by 7 sample variables ]
tax_table() Taxonomy Table: [ 2171 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 2171 tips and 2170 internal nodes ]
>
> ## Add group variable
> sample_data(ps)$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
>
```

Viewer:

File menu: Files, Plots, Packages, Help, Viewer, Publish.

RStudio screen



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 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  
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 o1 <- cowplot::plot_grid(a1, a2, ncol = 2, labels = c("Shannon", "Simpson"))  
149 o1  
150  
151 o2 <- cowplot::plot_grid(o1, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
152 o2  
153  
154 o3 <- cowplot::plot_grid(o2, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
155 o3  
156  
157 o4 <- cowplot::plot_grid(o3, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
158 o4  
159  
160 o5 <- cowplot::plot_grid(o4, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
161 o5  
162  
163 o6 <- cowplot::plot_grid(o5, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
164 o6  
165  
166 o7 <- cowplot::plot_grid(o6, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
167 o7  
168  
169 o8 <- cowplot::plot_grid(o7, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
170 o8  
171  
172 o9 <- cowplot::plot_grid(o8, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
173 o9  
174  
175 o10 <- cowplot::plot_grid(o9, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
176 o10  
177  
178 o11 <- cowplot::plot_grid(o10, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
179 o11  
180  
181 o12 <- cowplot::plot_grid(o11, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
182 o12  
183  
184 o13 <- cowplot::plot_grid(o12, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
185 o13  
186  
187 o14 <- cowplot::plot_grid(o13, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
188 o14  
189  
190 o15 <- cowplot::plot_grid(o14, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
191 o15  
192  
193 o16 <- cowplot::plot_grid(o15, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
194 o16  
195  
196 o17 <- cowplot::plot_grid(o16, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
197 o17  
198  
199 o18 <- cowplot::plot_grid(o17, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
200 o18  
201  
202 o19 <- cowplot::plot_grid(o18, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
203 o19  
204  
205 o20 <- cowplot::plot_grid(o19, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
206 o20  
207  
208 o21 <- cowplot::plot_grid(o20, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
209 o21  
210  
211 o22 <- cowplot::plot_grid(o21, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
212 o22  
213  
214 o23 <- cowplot::plot_grid(o22, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
215 o23  
216  
217 o24 <- cowplot::plot_grid(o23, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
218 o24  
219  
220 o25 <- cowplot::plot_grid(o24, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
221 o25  
222  
223 o26 <- cowplot::plot_grid(o25, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
224 o26  
225  
226 o27 <- cowplot::plot_grid(o26, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
227 o27  
228  
229 o28 <- cowplot::plot_grid(o27, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
230 o28  
231  
232 o29 <- cowplot::plot_grid(o28, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
233 o29  
234  
235 o30 <- cowplot::plot_grid(o29, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
236 o30  
237  
238 o31 <- cowplot::plot_grid(o30, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
239 o31  
240  
241 o32 <- cowplot::plot_grid(o31, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
242 o32  
243  
244 o33 <- cowplot::plot_grid(o32, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
245 o33  
246  
247 o34 <- cowplot::plot_grid(o33, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
248 o34  
249  
250 o35 <- cowplot::plot_grid(o34, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
251 o35  
252  
253 o36 <- cowplot::plot_grid(o35, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
254 o36  
255  
256 o37 <- cowplot::plot_grid(o36, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
257 o37  
258  
259 o38 <- cowplot::plot_grid(o37, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
260 o38  
261  
262 o39 <- cowplot::plot_grid(o38, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
263 o39  
264  
265 o40 <- cowplot::plot_grid(o39, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
266 o40  
267  
268 o41 <- cowplot::plot_grid(o40, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
269 o41  
270  
271 o42 <- cowplot::plot_grid(o41, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
272 o42  
273  
274 o43 <- cowplot::plot_grid(o42, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
275 o43  
276  
277 o44 <- cowplot::plot_grid(o43, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
278 o44  
279  
280 o45 <- cowplot::plot_grid(o44, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
281 o45  
282  
283 o46 <- cowplot::plot_grid(o45, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
284 o46  
285  
286 o47 <- cowplot::plot_grid(o46, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
287 o47  
288  
289 o48 <- cowplot::plot_grid(o47, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
290 o48  
291  
292 o49 <- cowplot::plot_grid(o48, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
293 o49  
294  
295 o50 <- cowplot::plot_grid(o49, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
296 o50  
297  
298 o51 <- cowplot::plot_grid(o50, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
299 o51  
300  
301 o52 <- cowplot::plot_grid(o51, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
302 o52  
303  
304 o53 <- cowplot::plot_grid(o52, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
305 o53  
306  
307 o54 <- cowplot::plot_grid(o53, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
308 o54  
309  
310 o55 <- cowplot::plot_grid(o54, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
311 o55  
312  
313 o56 <- cowplot::plot_grid(o55, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
314 o56  
315  
316 o57 <- cowplot::plot_grid(o56, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
317 o57  
318  
319 o58 <- cowplot::plot_grid(o57, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
320 o58  
321  
322 o59 <- cowplot::plot_grid(o58, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
323 o59  
324  
325 o60 <- cowplot::plot_grid(o59, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
326 o60  
327  
328 o61 <- cowplot::plot_grid(o60, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
329 o61  
330  
331 o62 <- cowplot::plot_grid(o61, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
332 o62  
333  
334 o63 <- cowplot::plot_grid(o62, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
335 o63  
336  
337 o64 <- cowplot::plot_grid(o63, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
338 o64  
339  
340 o65 <- cowplot::plot_grid(o64, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
341 o65  
342  
343 o66 <- cowplot::plot_grid(o65, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
344 o66  
345  
346 o67 <- cowplot::plot_grid(o66, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
347 o67  
348  
349 o68 <- cowplot::plot_grid(o67, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
350 o68  
351  
352 o69 <- cowplot::plot_grid(o68, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
353 o69  
354  
355 o70 <- cowplot::plot_grid(o69, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
356 o70  
357  
358 o71 <- cowplot::plot_grid(o70, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
359 o71  
360  
361 o72 <- cowplot::plot_grid(o71, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
362 o72  
363  
364 o73 <- cowplot::plot_grid(o72, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
365 o73  
366  
367 o74 <- cowplot::plot_grid(o73, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
368 o74  
369  
370 o75 <- cowplot::plot_grid(o74, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
371 o75  
372  
373 o76 <- cowplot::plot_grid(o75, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
374 o76  
375  
376 o77 <- cowplot::plot_grid(o76, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
377 o77  
378  
379 o78 <- cowplot::plot_grid(o77, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
380 o78  
381  
382 o79 <- cowplot::plot_grid(o78, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
383 o79  
384  
385 o80 <- cowplot::plot_grid(o79, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
386 o80  
387  
388 o81 <- cowplot::plot_grid(o80, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
389 o81  
390  
391 o82 <- cowplot::plot_grid(o81, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
392 o82  
393  
394 o83 <- cowplot::plot_grid(o82, ps, ncol = 2, labels = c("Shannon", "Simpson"))  
395 o83  
396  
397 o84 <- cowplot::plot_grid(o83, psf, ncol = 2, labels = c("Shannon", "Simpson"))  
398 o84  
399  
399
```

Console

Plots

Environment

The screenshot shows the RStudio interface with several windows open. The 'Script' pane on the left contains a large block of R code for alpha diversity analysis. The 'Console' pane in the top right shows the output of this code, including phylogenetic objects and boxplots. The 'Plots' pane below the console displays two boxplots of alpha diversity (Shannon and Simpson) across different time points and treatments. The 'Environment' pane at the bottom left lists various R objects and their types. A blue arrow points from the 'Script' pane towards the 'Environment' pane, indicating a flow from code execution to variable inspection.

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 top right shows "100% battery", "Tue 16:27", and a search icon.

The main workspace consists of several panes:

- Script**: The leftmost pane contains R code for phylogenetic analysis. It includes sections for relative abundance calculations, filtering, alpha diversity estimation, and phylogenetic tree construction. A callout box labeled "Script" highlights this area.
- Console**: The top-right pane displays the R console output. It shows the creation of various objects like `taxa_names`, `otu_table`, `sample_data`, `tax_table`, and `phy_tree`. A callout box labeled "Console" highlights this area.
- History**: The bottom-left pane shows the history of R commands run in the session. A callout box labeled "History" highlights this area.
- Files**: The bottom-right pane shows the file system structure under the "Home" directory. It includes subfolders for ".Rhistory", "Applications", "Desktop", "Documents", "Downloads", "Dropbox", "Library", "Movies", "Music", "Pictures", "Public", and "Zotero". A callout box labeled "Files" highlights this area.

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

```
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
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) R Script
# MARCH Sample names
rownames(info) <- 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 = ..
```

How to R – 2 ways

The screenshot shows the RStudio interface with several panels:

- Code Editor:** Displays R code for phylogenetic analysis, including filtering samples and calculating alpha diversity.
- Script:** A box containing the text "2. Using and saving scripts (ie. plain text files of code)."
- Console:** A box containing the text "1. Typing and executing commands interactively here."
- Plots:** A box showing a box plot of Alpha diversity (Shannon) across timepoints (base, week3) for four treatment groups: control (red), pre (green), pro (cyan), and syn (purple).

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

```
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
131
132 ## Make table of alpha diversity
133 alpha <- estimate_richness(ps)
134 alpha_info <- sample_data(ps)
135 aa <- cbind(alpha, alpha_info)
136
137 ## Check for outliers
138 aplot(alpha$Shannon, binwidth = 1)
139 aplot(alpha$Simpson, binwidth = 1)
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)) + labs(y = "Alpha diversity (Shannon)", x = "Timepoint")
143 o1 <- a1
144 o2 <- a1 + geom_boxplot(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)) + labs(y = "Alpha diversity (Simpson)", x = "Timepoint")
145 o3 <- a1 + geom_boxplot(aes(x = timepoint, y = richness, 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)) + labs(y = "Alpha diversity (Richness)", x = "Timepoint")
146 o4 <- a1 + geom_boxplot(aes(x = timepoint, y = Chao1, 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)) + labs(y = "Alpha diversity (Chao1)", x = "Timepoint")
147 o5 <- a1 + geom_boxplot(aes(x = timepoint, y = H, 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)) + labs(y = "Alpha diversity (H)", x = "Timepoint")
148 o6 <- a1 + geom_boxplot(aes(x = timepoint, y = PD_whole, 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)) + labs(y = "Alpha diversity (PD whole)", x = "Timepoint")
149 o7 <- a1 + geom_boxplot(aes(x = timepoint, y = PDRarefaction, 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)) + labs(y = "Alpha diversity (PD rarefaction)", x = "Timepoint")
150 o8 <- a1 + geom_boxplot(aes(x = timepoint, y = SimpsonEven, 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)) + labs(y = "Alpha diversity (Simpson even)", x = "Timepoint")
151 o9 <- a1 + geom_boxplot(aes(x = timepoint, y = ChaoEven, 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)) + labs(y = "Alpha diversity (Chao even)", x = "Timepoint")
152 o10 <- a1 + geom_boxplot(aes(x = timepoint, y = RichEven, 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)) + labs(y = "Alpha diversity (Rich even)", x = "Timepoint")
153 o11 <- a1 + geom_boxplot(aes(x = timepoint, y = HEven, 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)) + labs(y = "Alpha diversity (HEven)", x = "Timepoint")
154 o12 <- a1 + geom_boxplot(aes(x = timepoint, y = PDEven, 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)) + labs(y = "Alpha diversity (PDEven)", x = "Timepoint")
```

Console Terminal R Markdown

Script

Console

1. Typing and executing commands interactively here.

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

Choose this one.

Files Plots Packages Help Viewer

Screenshot

Alpha diversity (Shannon)

treatment

- control
- pre
- pro
- syn

base week3

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.

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

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

Project: (None)

```
index.Rmd * Schedule.Rmd * Session1.Rmd * jodi_btbr_phylseq.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 ## 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 o2
```

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

Run Terminal R Markdown

```
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
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 ]
2171 tips and 2170 internal nodes ]
```

Tools > Install packages...

Install Packages

Install from: Repository (CRAN)

Configuring Repositories

Packages (separate multiple with space or comma):

Install to Library: /Library/Frameworks/R.framework/Versions/3.6/Resources/library

Install dependencies

Install Cancel

Screenshot

treatment

- control
- pre
- pro
- syn

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.

Working directory

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help RStudio Addins Go to file/function Addins ~/ jodi_btbr_phyloseq.R

```
119
120 ## Take relative abundance
121 rel <- transform_sample_counts(ps, function(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
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))
143 o1 + geom_boxplot(outlier.shape = 21) + scale_fill_manual(values = c("red", "green", "cyan", "purple"))
144 o1 + geom_point(outlier.size = 4, position = position_jitter())
145 o1 + stat_boxplot(outlier.color = "black", outlier.fill = "white", outlier.shape = 4, outlier.size = 4, outlier.v = 0.8) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4) + ylab("Alpha diversity (Shannon)") + xlab("Timepoint")
146
147 o2 <- ggplot(aa, aes(x = timenpoint, y = Simpson, fill = treatment)) + geom_boxplot(outlier.fill = NULL, outlier.shape = 4, v = 0.8) + stat_summary(fun.y = mean, geom = "point", shape = 4, size = 4) + ylab("Alpha diversity (Simpson)") + xlab("Timepoint")
148
149 # (Untitled):1
```

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")
```

Files Plots Packages Help Viewer Publish

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

Working directory

The figure shows a screenshot of the RStudio interface. On the left, the code editor displays a script for phylogenetic analysis, including sections for relative abundance, alpha diversity calculations, outlier detection, and plotting. The right side of the interface shows a text box containing two bullet points: "However: you may want to run your script on a different computer with a different directory structure where that directory does not exist." and "Or you may want to work in multiple directories." Below this text box is a box plot titled "Screenshot" comparing "Alpha diversity (Shannon)" across different time points ("base" and "week3") and treatments ("control", "pre", "pro", "syn"). The y-axis ranges from 2.5 to 4.0. The legend indicates the following color mapping for treatments:

- control (red)
- pre (green)
- pro (cyan)
- syn (purple)

The box plot shows that "pre" treatment generally has the highest alpha diversity, followed by "pro", "syn", and "control".

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). The background shows R code in the console and a boxplot plot.

```
119
120 ## Take relative abundance
121 rel <- transform(ps, rel = counts(ps, function(x) x / sum(x)))
122
123 ## Execute filter
124 relf <- prune_taxa(keep = !rel, ps)
125 psf <- prune_taxa(keep = !ps)
126
127
128 ###### Alpha diversity
129
130 ## Calculate alpha diversity
131
132 ## Make table
133 alpha <- estim
134 alpha_info <- sample_data(ps)
135 alpha <- cbind(alpha, alpha_info)
136
137
138 ## Check for outliers
139 qplot(alpha$Shannon, binwidth = 0.05) + xlab("Shannon diversity")
140 qplot(alpha$Simpson, binwidth = 0.005) + xlab("Simpson diversity")
141
142 ## Plot
143 o1 <- 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, shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon) ("Timepoint")")
144 o1
145 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, shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Simpson) ("Timepoint")")
146 o2
147 o3 <- 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, shape = 4, size = 4, position = position_dodge(width = 0.75)) + ylab("Alpha diversity (Shannon) ("Timepoint")")
148 o3
149
```

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)$treatment != "NA", ps)
```

New Project

Create Project

New Directory Start a project in a brand new working directory

Existing Directory Associate a project with an existing working directory

Version Control Checkout a project from a version control repository

Cancel

Screenshot

Alpha div

base week3

treatment

- control
- pre
- pro
- syn

Error messages

Console Terminal × R Markdown ×



```
>
> ## 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

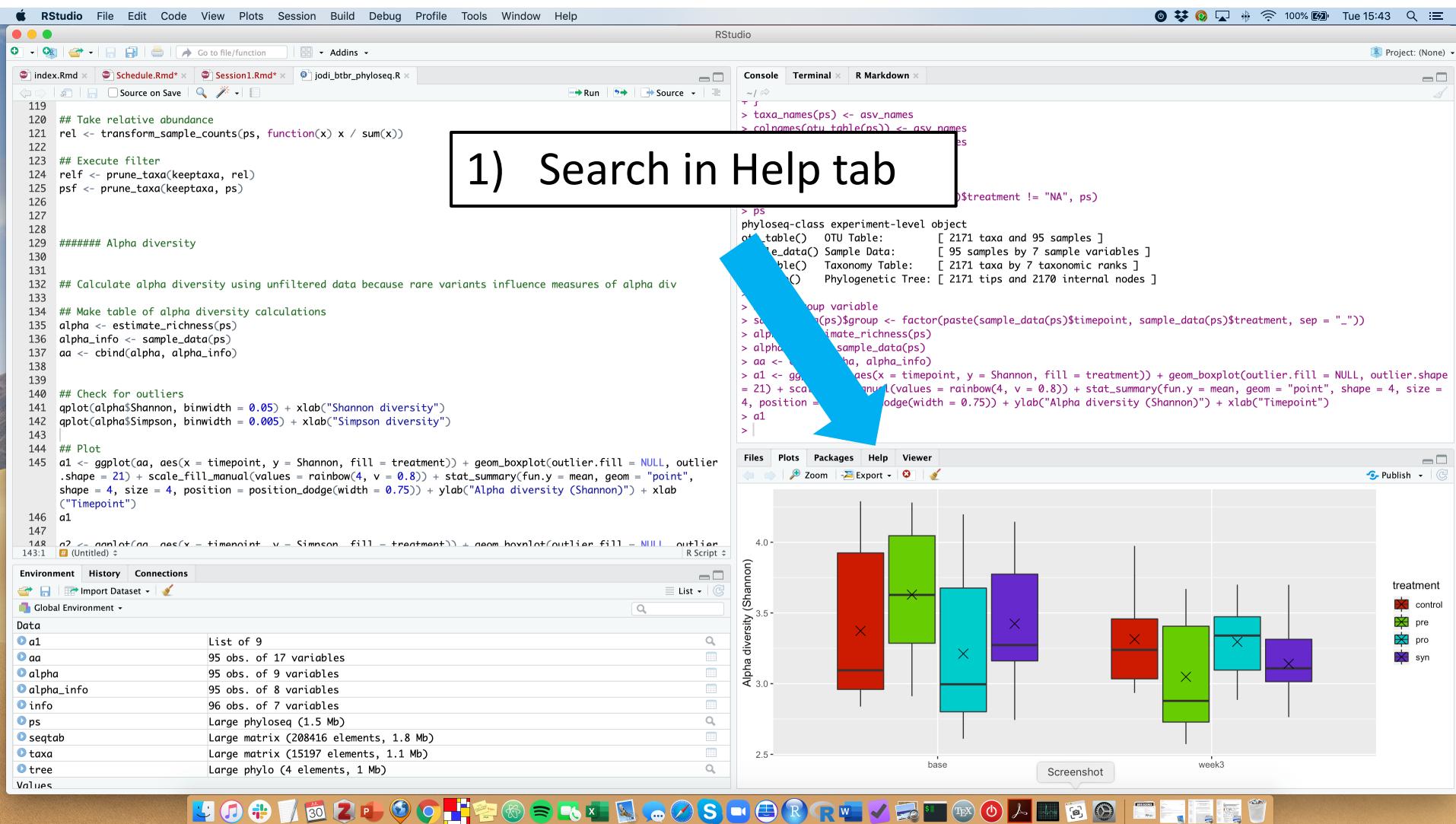
Console Terminal R Markdown

```
>  
> ## 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



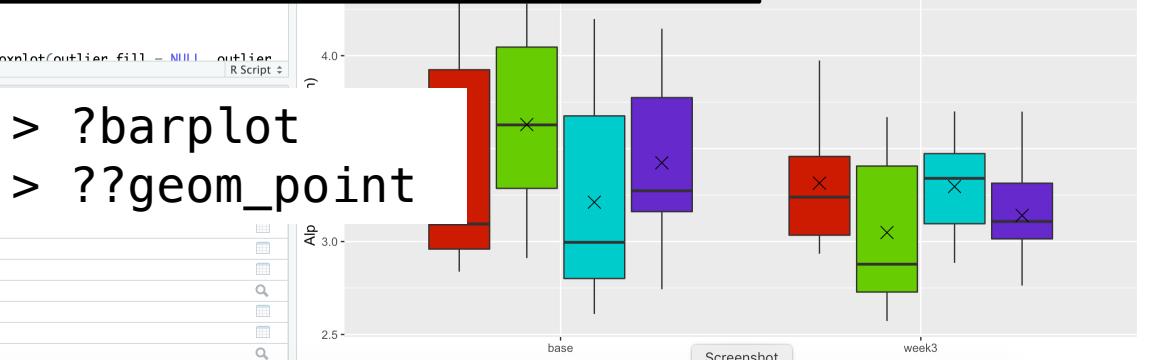
Getting help

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
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 ## Calculate alpha diversity using unfiltered data because rare vo
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
138 ## Check for outliers
139 qplot(alpha$Shannon, binwidth = 0.05) + xlab("Shannon diversity")
140 qplot(alpha$Simpson, binwidth = 0.005) + xlab("Simpson diversity")
141
142 ## Plot
143 o1 <- ggplot(aa, aes(x = timepoint, y = Shannon, fill = treatment))
144 .shape = 21) + scale_fill_manual(values = rainbow(4, v = 0.8)) + s
145 shape = 4, size = 4, position = position_dodge(width = 0.75)) + yl
146 ("Timepoint")
147
148 o2 <- ggplot(aa, aes(x = timepoint, y = Simpson, fill = treatment)) +
149 geom_boxplot(outlier.fill = NULL, outlier.shape = 16, outlier.size = 2) +
150 stat_summary(fun.y = mean, geom = "point", shape = 4, size =
151 3) + xlab("Timepoint") + ylab("Simpson diversity (Shannon)") + xlab("Timepoint")
152
153 (Untitled):1
```

1) Search in Help tab

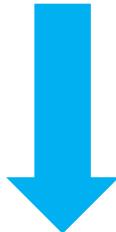
2) Type ? followed by the function name in the console (or ?? for installed packages)



> ?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.