Zebrafish Diet and Infection Targetted Analysis

Connor Draney, Sunni Patton, Kelly Shannon, Michael Sieler

05 March, 2022

Contents

Import Data

We obtained data in the form of a phyloseq data object from the Human Genome Project (HMP) that was previously processed for taxanomic identification of microbial organisms. We then cleaned and processed the data further for statistical analysis.

```
# Download Data
# temp_test = tempfile()
# test_url = "http://joey711.github.io/phyloseq-demo/HMPv35.RData"
# download.file(test_url, destfile = paste0(data.path, "/Raw/HMPv35.RData"))

load(paste0(data.path, "/Raw/HMPv35.RData"))
ps.unclean <- HMPv35
rm(HMPv35)  # remove this obj from global env
```

Clean sample data

The 4743 samples in the HMP were obtained from multiple individuals (2555 males and 2188 females), processed by several locations, and in some cases repeated samples were taken from the same individual. We subset the data to ensure independence by limiting samples from one location and only the first visit. Additionally, we removed samples noted as "Mislabeled" or "Contaminated". After removing samples using these criteria, we were left with 609 samples.

```
# Find the center with largest counts for the following conditions
sample.data.frame(ps.unclean) %>%
  count(RUNCENTER,
        Mislabeled,
        Contaminated,
        visitno) %>%
  arrange(-n)
```

```
##    RUNCENTER Mislabeled Contaminated visitno   n
## 1       JCVI      FALSE        FALSE       1 609
## 2        BCM      FALSE        FALSE       2 583
## 3         BI      FALSE        FALSE       1 538
## 4       WUGC      FALSE        FALSE       1 518
## 5       WUGC         NA           NA       1 513
## 6       WUGC      FALSE        FALSE       2 473
## 7       JCVI      FALSE        FALSE       2 400
## 8         BI      FALSE        FALSE       2 380
## 9         BI         NA           NA       1 160
## 10    BCM,BI         NA           NA       1  85
## 11 JCVI,WUGC         NA           NA       1  79
## 12   JCVI,BI         NA           NA       1  75
## 13       BCM      FALSE        FALSE       1  73
## 14  BCM,WUGC      FALSE        FALSE       1  49
## 15  BCM,WUGC         NA           NA       1  41
## 16      WUGC         NA           NA       2  26
## 17       BCM         NA           NA       1  25
```

```
## 18  BCM,JCVI      FALSE        FALSE       1  23
## 19  BCM,JCVI      FALSE        FALSE       2  17
## 20       BCM      FALSE        FALSE       3  15
## 21 JCVI,WUGC      FALSE        FALSE       1  14
## 22    BCM,BI      FALSE        FALSE       2  12
## 23      WUGC      FALSE        FALSE       3   9
## 24    BI,BCM         NA           NA       1   7
## 25    BCM,BI      FALSE        FALSE       1   6
## 26    BI,BCM      FALSE        FALSE       2   6
## 27  WUGC,BCM      FALSE        FALSE       1   5
## 28  WUGC,BCM         NA           NA       1   1
## 29 WUGC,JCVI         NA           NA       1   1
```

```
# Subset based on parameters
ps.cleaned <- subset_samples(ps.unclean,  # physeq object
                             Mislabeled == F &  # Remove mislabeled
                             Contaminated == F &  # remove contaminated
                             visitno == 1 & # first visit, so no repeat sampling of individuals, indep
                             RUNCENTER == "JCVI"  # Same center, independence
                             )


# Remove columns, if rownames are already sample names, no need to have an extra sample column
sample_data(ps.cleaned) <- sample_data(ps.cleaned)[, c(-1)]  # [rows, cols]

# Check to see how many samples dropped from original dataset
print(paste0("Dropped ",
             nrow(sample_data(ps.unclean)) - nrow(sample_data(ps.cleaned)),
             " samples after cleaning.")
            )
```

```
## [1] "Dropped 4134 samples after cleaning."
```

```
# head(sample.data.frame(ps.cleaned))
```

Rarefaction

We rarefied the data to control for uneven sampling efforts Sanders, H. L. (1968), Willis, A.D. (2019).

```
ps.norm <- ps.cleaned
rarefaction.minimum <- 1000
min.smpl.size <- min(sample_sums(ps.norm)[sample_sums(ps.norm) >= rarefaction.minimum])

summary(sample_sums(ps.norm))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       1    2679    3577    3863    4696   29969
```

```
# Rarefy

ps.rar <- {
```

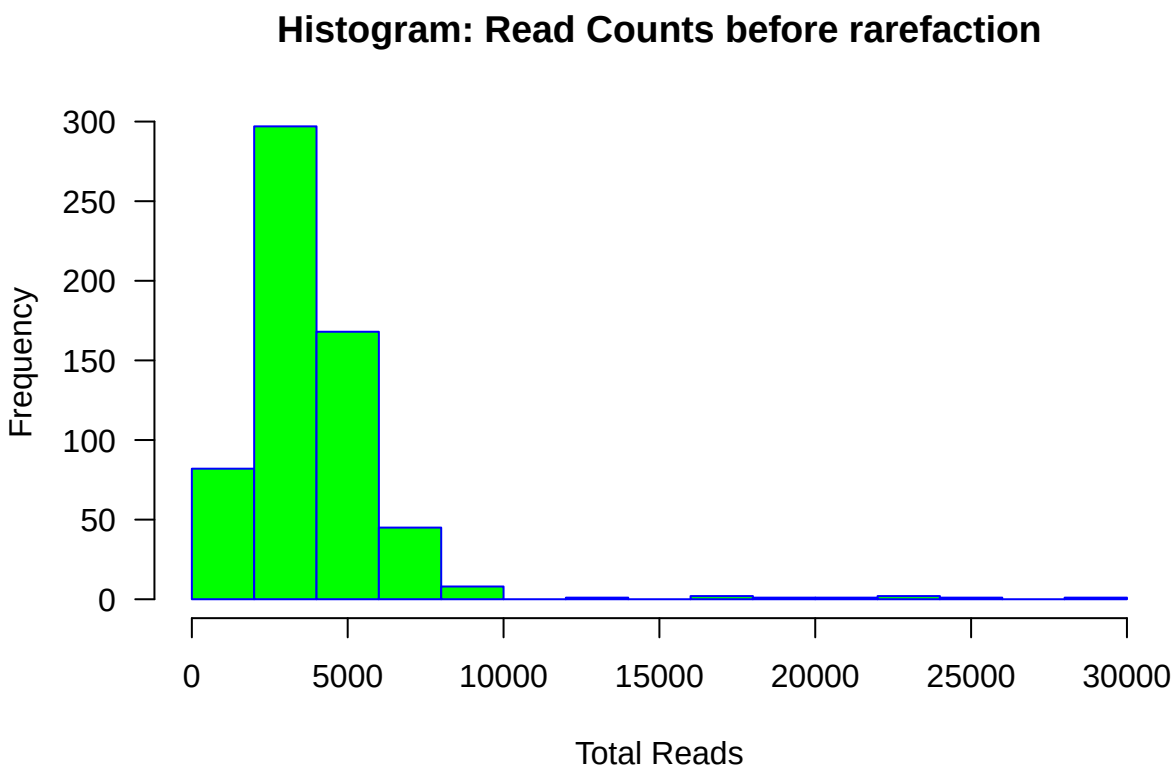
```
  ps.rar.tmp0 <- ps.norm
  ps.rar.tmp1 <- rarefy_even_depth(
      physeq = ps.rar.tmp0,
      sample.size = min.smpl.size,
      trimOTUs = TRUE,
      rngseed = 42
      )
  rename.NA.taxa(ps.rar.tmp1)
}

ps.rar
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 24839 taxa and 562 samples ]
## sample_data() Sample Data:       [ 562 samples by 8 sample variables ]
## tax_table()   Taxonomy Table:    [ 24839 taxa by 7 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 24839 tips and 24650 internal nodes ]
## refseq()      DNAStringSet:      [ 24839 reference sequences ]
```

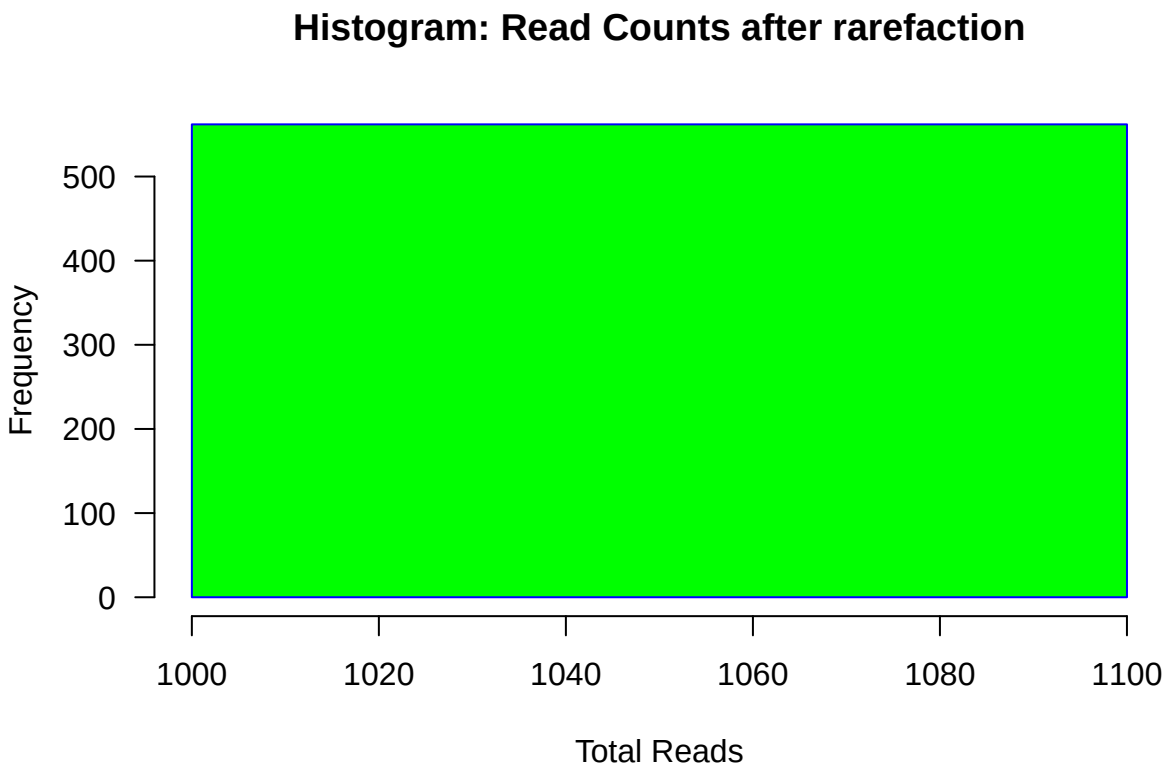
Plot pre- and post-rarefaction

Check the number of reads before and after rarefaction.

```
# rare.plot.preRare <- rarecurve(t(otu_table(ps.norm)), step=50, cex=0.5, label=F)
hist(sample_sums(ps.norm), main="Histogram: Read Counts before rarefaction", xlab="Total Reads",
     border="blue", col="green", las=1, breaks=12)
```




```
hist(sample_sums(ps.rar), main="Histogram: Read Counts after rarefaction", xlab="Total Reads",
     border="blue", col="green", las=1, breaks=12)
```




```
# Make a list of phyloseq objects, dataframes and datatables
ps.list <- list(RAR = list(ps.all = ps.rar,
                           df.all = sample.data.frame(ps.rar),
                           dt.all = sample.data.table(ps.rar)
                           ))

# Add observation numbers to data
ps.list[["RAR"]][["df.all"]] <- dplyr::mutate(ps.list[["RAR"]][["df.all"]], obs_num = row_number(), .be
ps.list[["RAR"]][["dt.all"]] <- dplyr::mutate(ps.list[["RAR"]][["dt.all"]], obs_num = row_number(), .be


# head(ps.list$RAR$df.all)
```

Alpha-diversity

We calculated alpha diversity scores, a measure of number of unique organisms within a single sample, using the indices Shannon, Simpson, Phylogenetic, and Richness (Whittaker 1960). Each calculates alpha-diversity using slightly different mathematical approaches to measure evenness (distribution of organisms) and/or richness (number of organismss). If differences in results are seen between indices, this can reveal insights into which kinds of organisms are present (e.g., common vs rare).

```
# Select which alpha measures we want to analyze
methods.alpha <- c("Shannon", "Simpson", # Non-phylogenetic measures, add additional measures here
                   "Phylogenetic", "Richness") %>%
```

```
                  purrr::set_names()  # Set's names list elements in "alpha.methods"

# Calculate alpha scores, save to list
ps.list$RAR[["alphaScore"]] <- alpha_base(ps.list$RAR$ps.all,  # Phyloseq object
                                methods.alpha,  # list of alpha methods
                                "Sample",
                                phylo.div = T
                                )
```

```
## [1] "Calculating non-phylogenetic alpha scores..."

## Warning in `[.data.table`(tmp.dt, , `:=`(se.chao1, NULL)): Column 'se.chao1'
## does not exist to remove

## [1] "Calculating phylogenetic alpha scores..."
```

```
# Add alpha scores to data table
ps.list$RAR[["dt.all.alpha"]] <- ps.list$RAR$dt.all[ps.list$RAR$alpha, on = "Sample"] %>% setkeyv("Samp

# Melt data table
ps.list$RAR[["dt.all.alpha.melt"]] <- melt_to_datatable_2(datatable1 = ps.list$RAR$dt.all,
                                                          datatable2 = ps.list$RAR$dt.all.alpha,
                                                          vars = methods.alpha,
                                                          var.name = "Alpha.Metric",
                                                          samp.name = "Sample",
                                                          val.name = "Alpha.Score"
                                                         )
```

Fit linear model

We fit the data using linear model to predict alpha-diversity score as a function of body sub-site, sex, or their interaction.

```
caseInfStats <- list()

data <- ps.list$RAR[["dt.all.alpha.melt"]]
# data <- dplyr::mutate(data, obs_num = row_number(), .before = 1)

caseInfStats[["mod.unref"]] <- lapply(methods.alpha, function(alpha){
  lm( formula = "Alpha.Score ~ sex*HMPbodysubsite",
       data = subset(data, Alpha.Metric == alpha)
      )
})
```

Check assumptions (unrefined model)

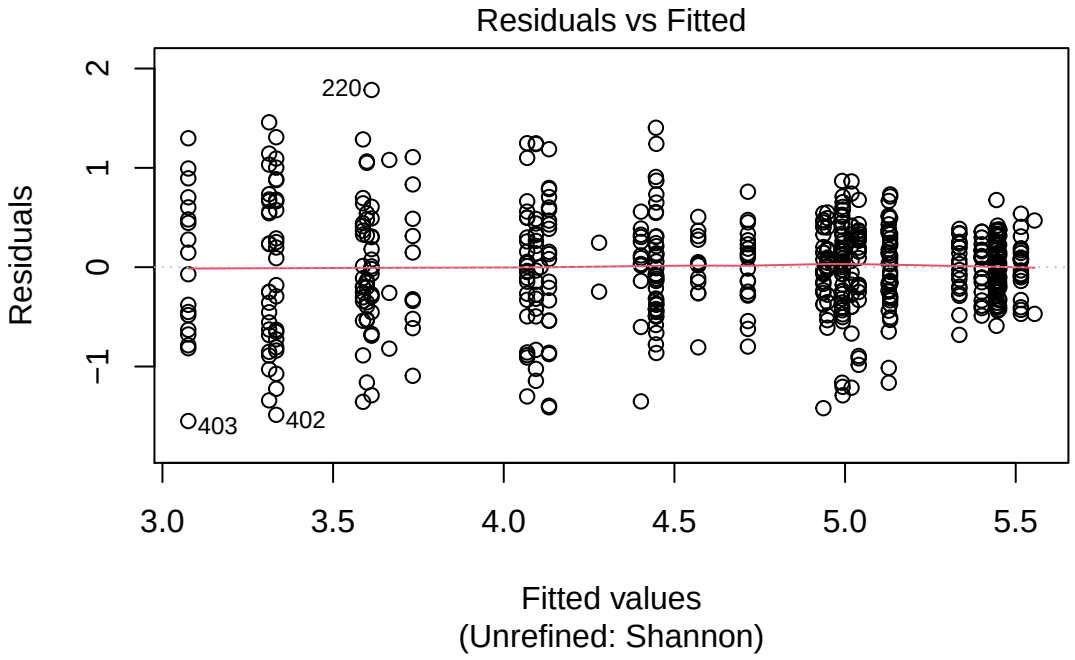
To assess if the assumptions of linear model regression were met, we visually inspected residuals vs fitted values, standardized residuals vs quantiles, standardized residuals vs fitted values and residuals vs leverage for each diversity index. In general, Residuals vs fitted appears slightly heteroscedasticity, Q-Q plots show that the data slightly deviates from the diagonal line indicating that the data may be non-normal, Scale-location

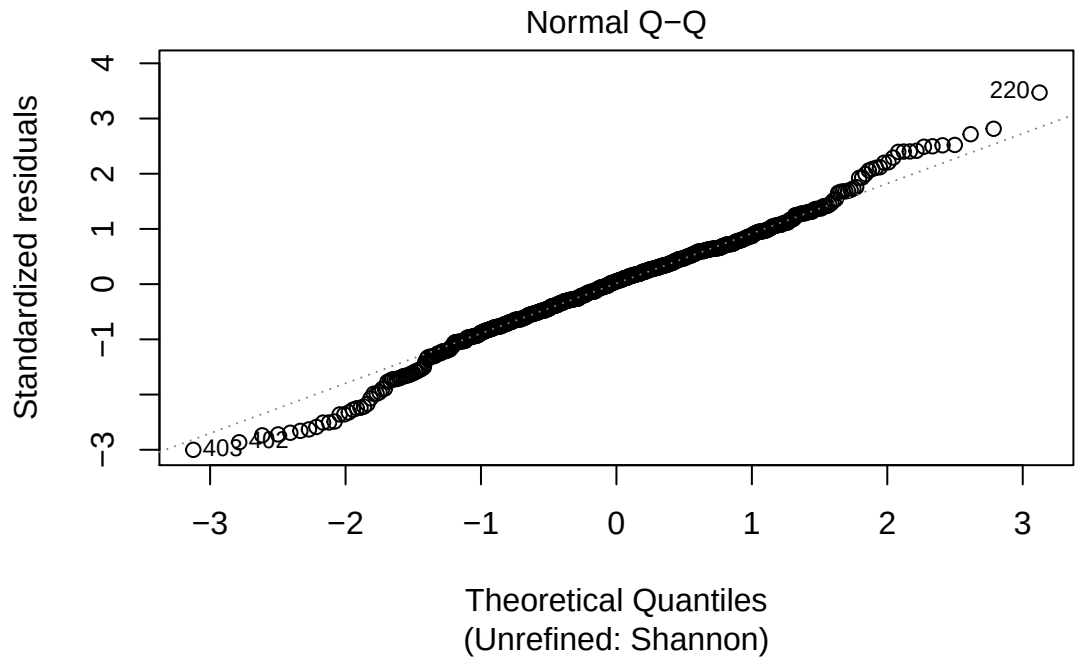
plots show that std. residuals are negatively associated with fitted values indicating heteroscedasticity, and there are several points with high leverage, but none that appear to have too high. The plots associated with the Simpson's index appear to be significantly heteroscedastic and non-normal.

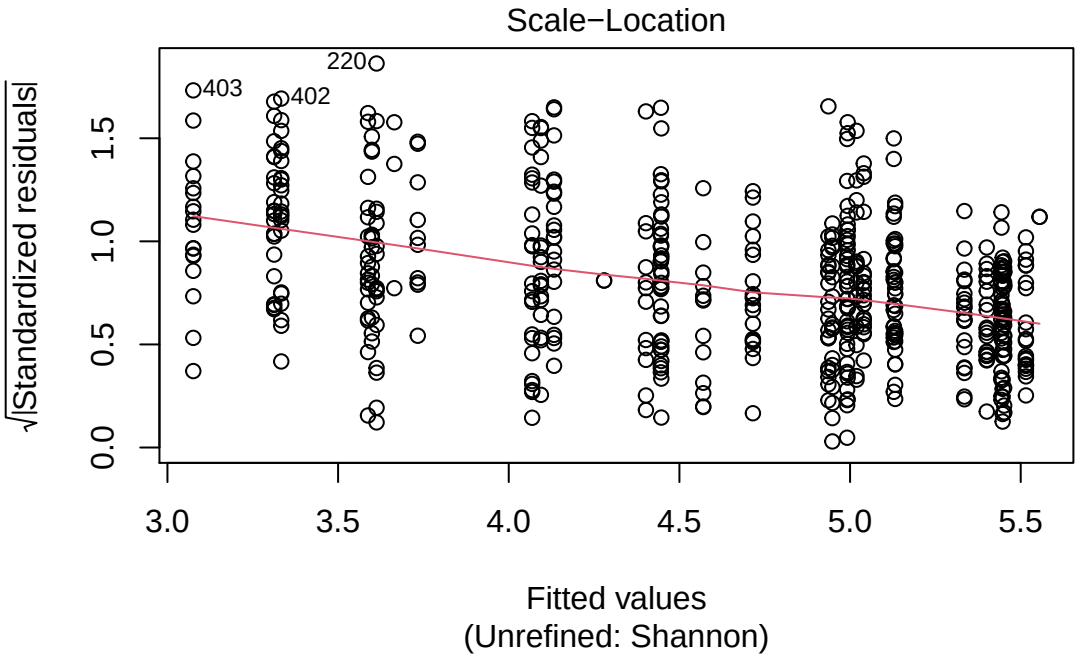
Shannon

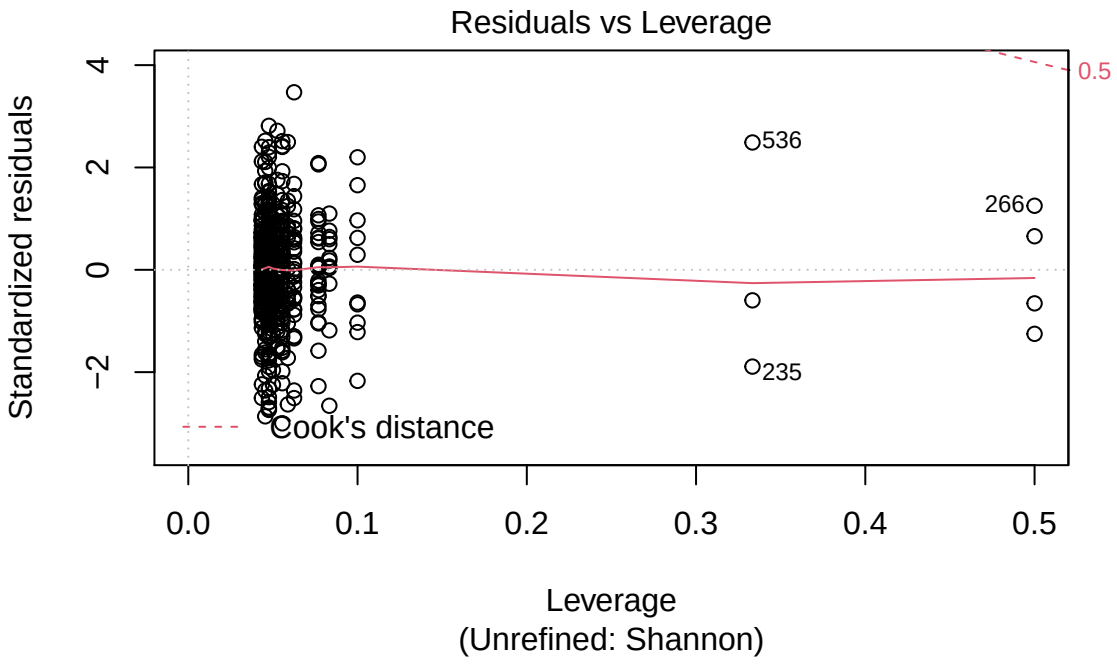
```
index <- "Shannon"
plot(caseInfStats[["mod.unref"]][[index]], sub = paste0("(Unrefined: ", methods.alpha[index],")") )
```

```
## Warning: not plotting observations with leverage one:
##   234
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
(Unrefined: Shannon)
220
403
402
−3
−2
−1
0
1
2
3
4

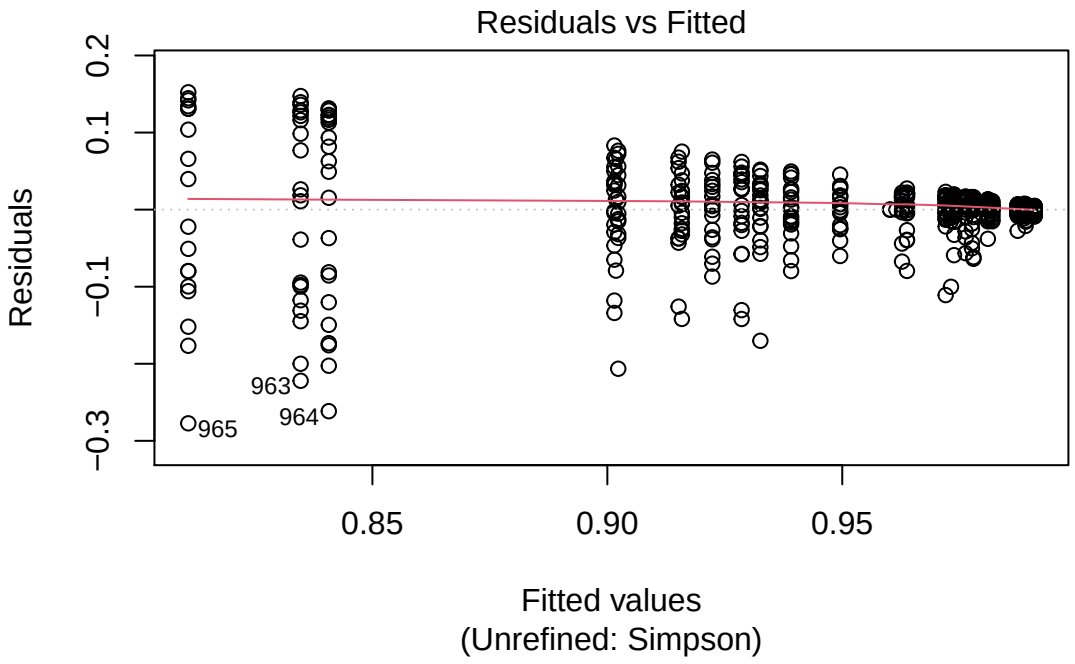

Scale−Location
√|Standardized residuals|
Fitted values
(Unrefined: Shannon)
220
403
402
0.0
0.5
1.0
1.5
3.0
3.5
4.0
4.5
5.0
5.5

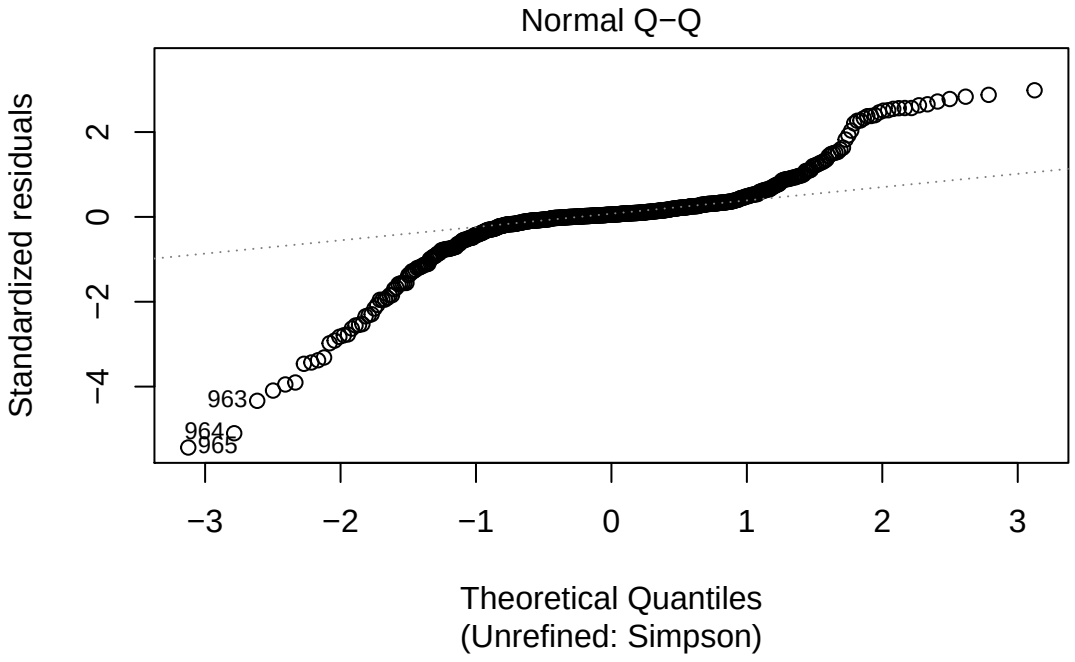


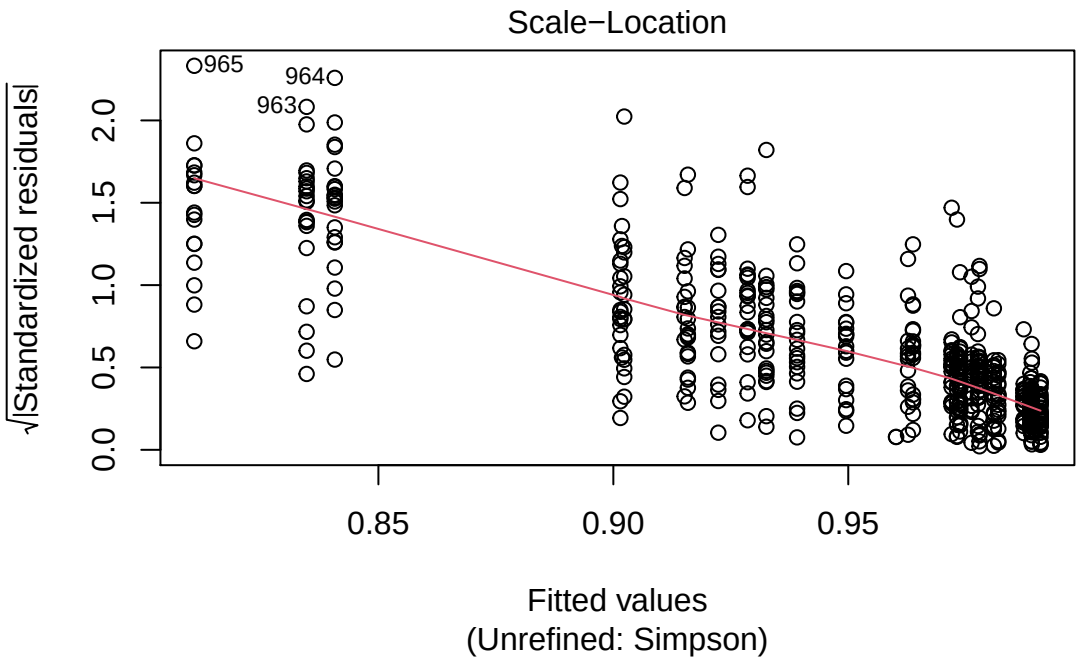
Simpson

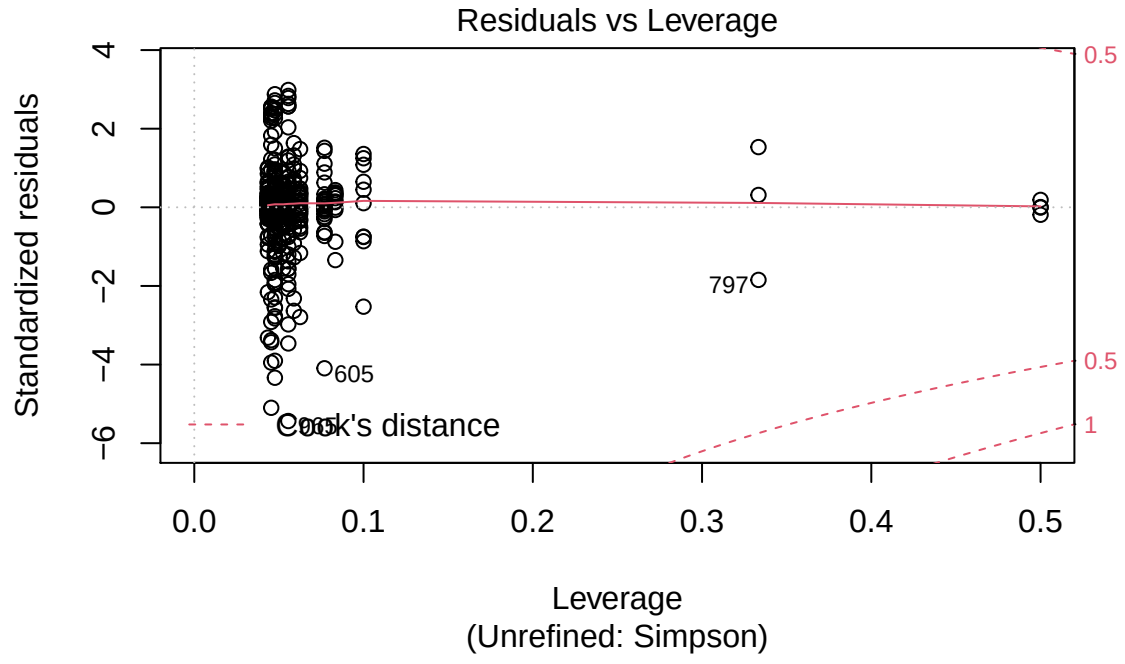
```
index <- "Simpson"
plot(caseInfStats[["mod.unref"]][[index]], sub = paste0("(Unrefined: ", methods.alpha[index],")") )
```

```
## Warning: not plotting observations with leverage one:
##   234
```


Residuals vs Fitted
Residuals
0.2
0.1
−0.1
−0.3
963
964
965
0.85
0.90
0.95
Fitted values
(Unrefined: Simpson)


Normal Q−Q
Standardized residuals
2
0
−2
−4
963
964
965
−3
−2
−1
0
1
2
3
Theoretical Quantiles
(Unrefined: Simpson)

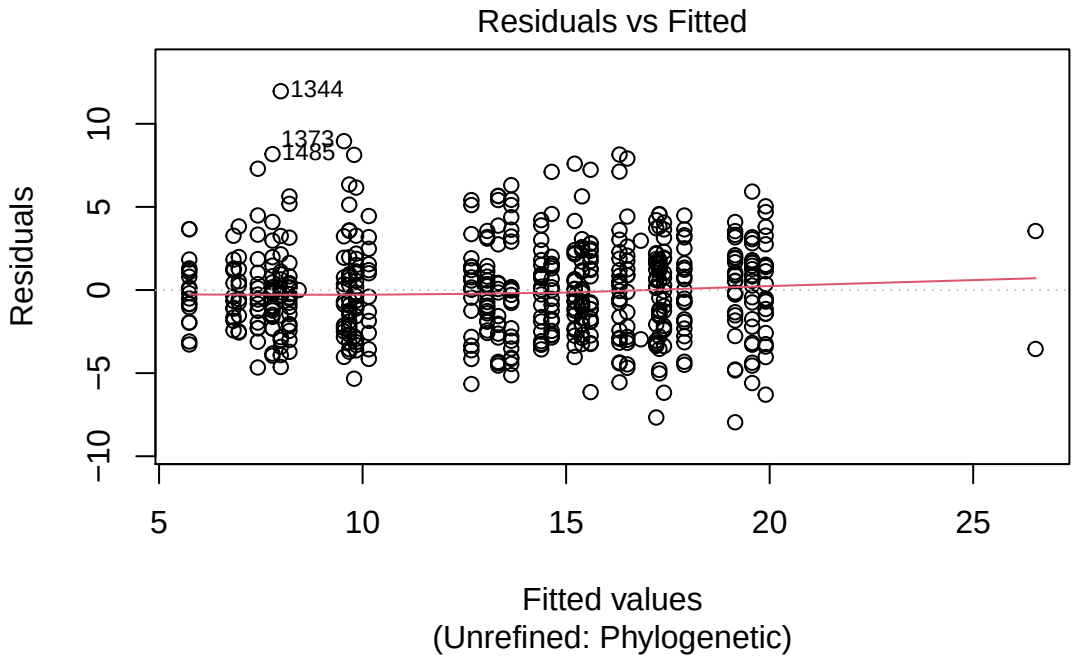

Scale−Location
965
964
963
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
0.85
0.90
0.95
Fitted values
(Unrefined: Simpson)

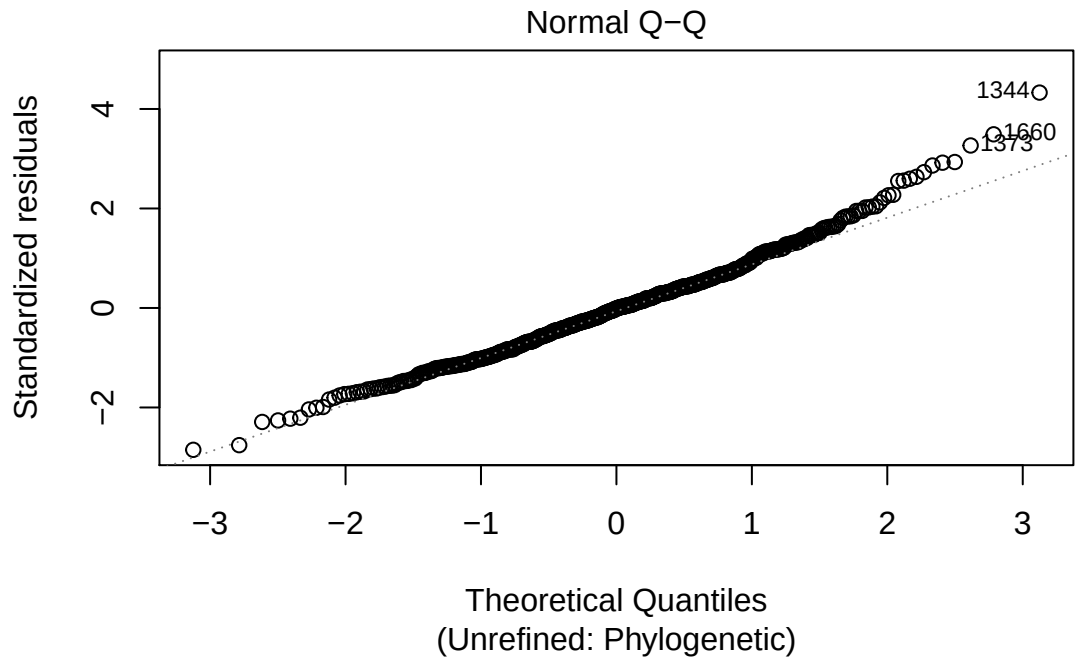

Residuals vs Leverage
Standardized residuals
4
2
0
−2
−4
−6
797
605
Cook's distance
0.5
0.5
1
0.0
0.1
0.2
0.3
0.4
0.5
Leverage
(Unrefined: Simpson)

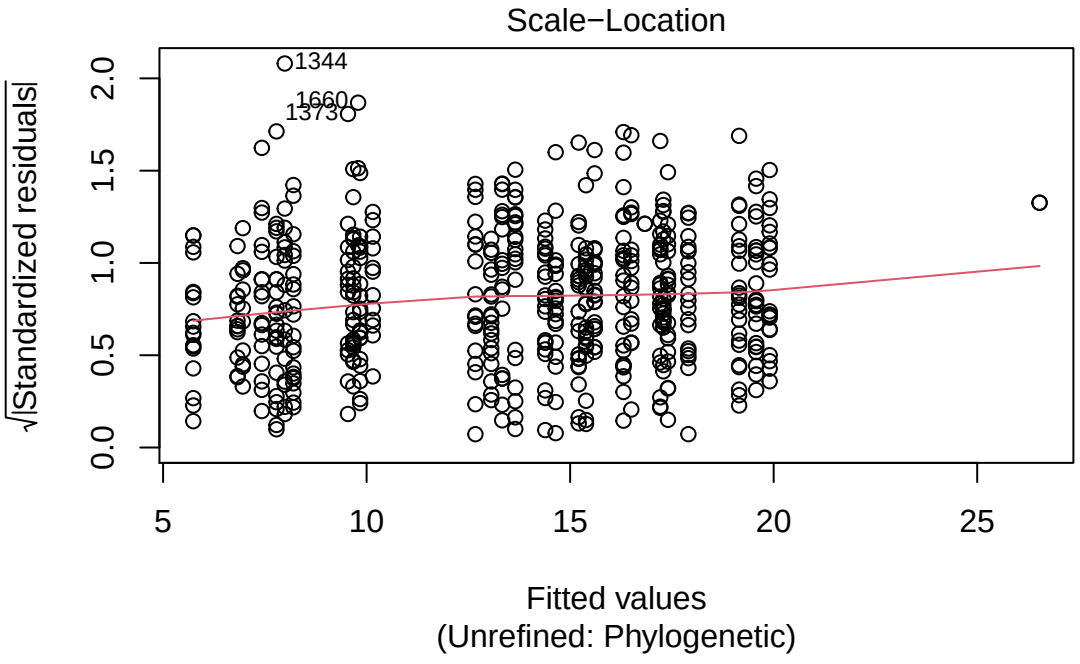
Phylogenetic

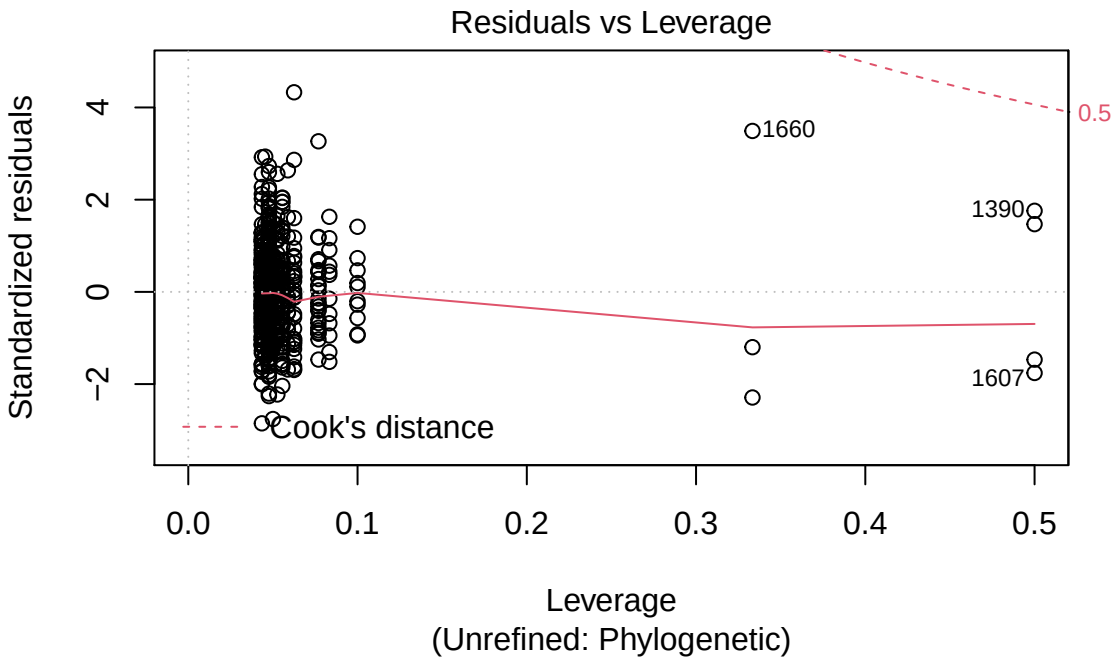
```
index <- "Phylogenetic"
plot(caseInfStats[["mod.unref"]][[index]], sub = paste0("(Unrefined: ", methods.alpha[index],")") )
```

```
## Warning: not plotting observations with leverage one:
##   234
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
(Unrefined: Phylogenetic)
1344
1660
1373
−3
−2
−1
0
1
2
3
−2
0
2
4

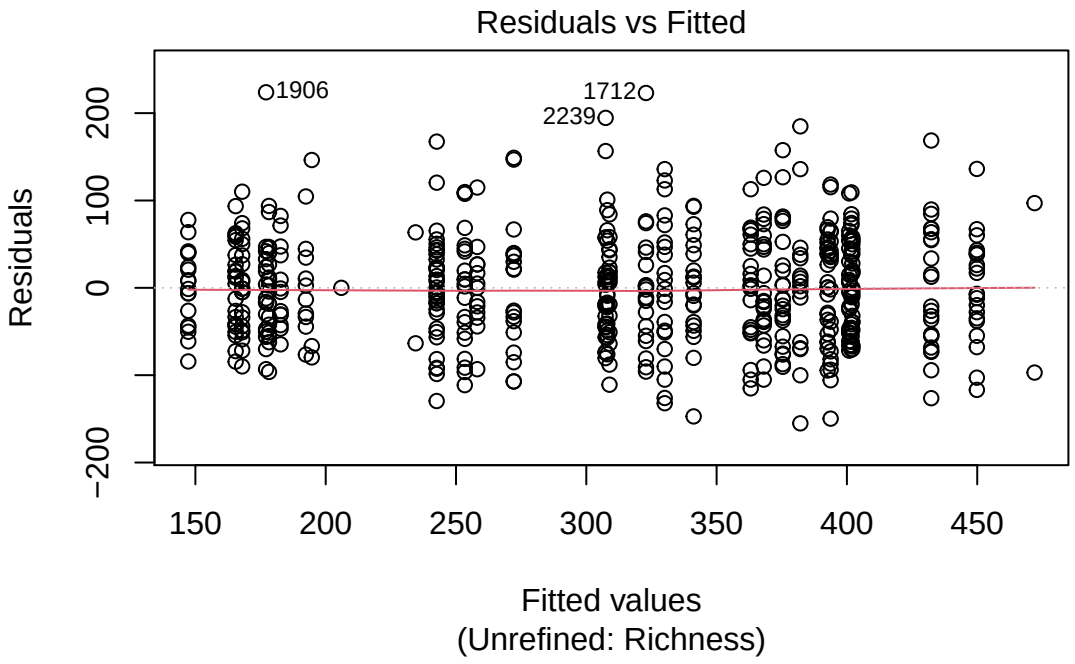

Scale−Location
√|Standardized residuals|
Fitted values
(Unrefined: Phylogenetic)
1344
1660
1373
5
10
15
20
25
0.0
0.5
1.0
1.5
2.0

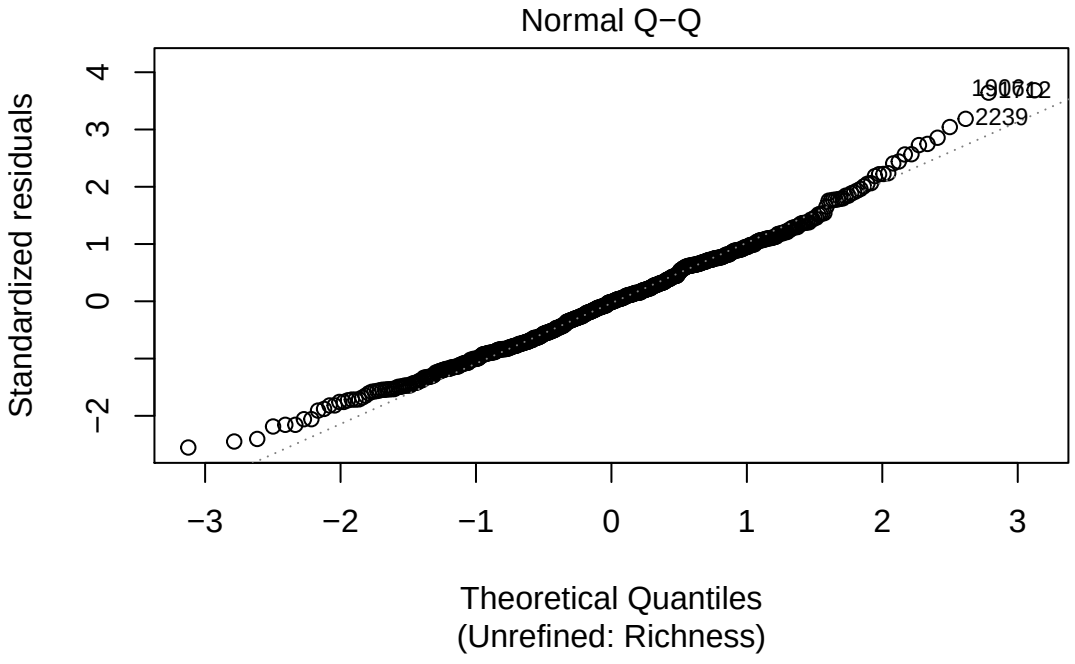


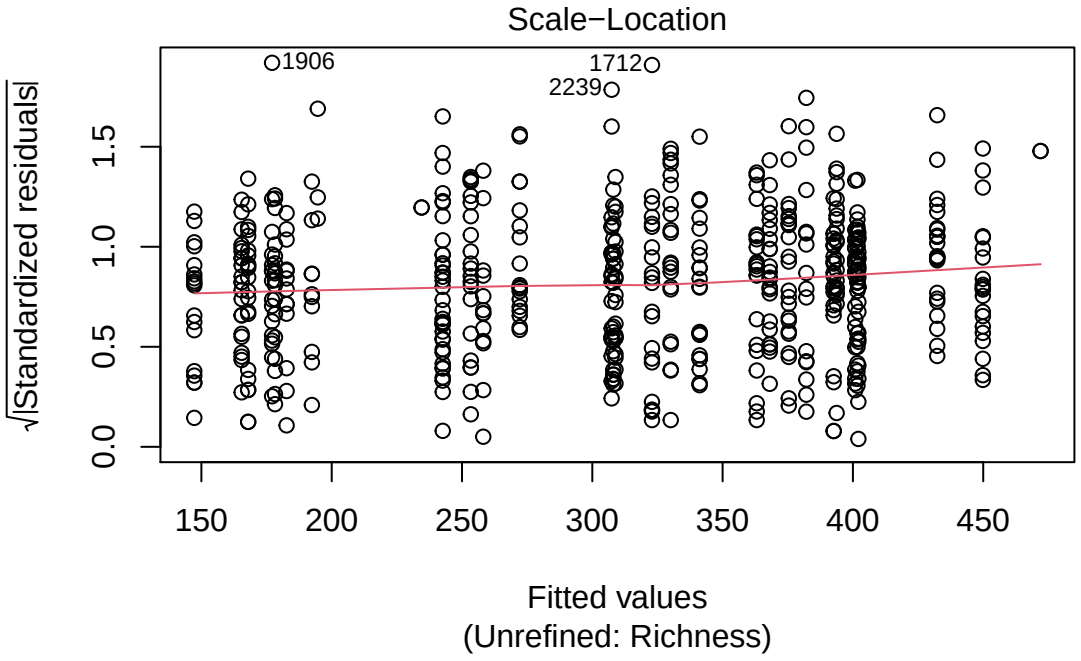
Richness

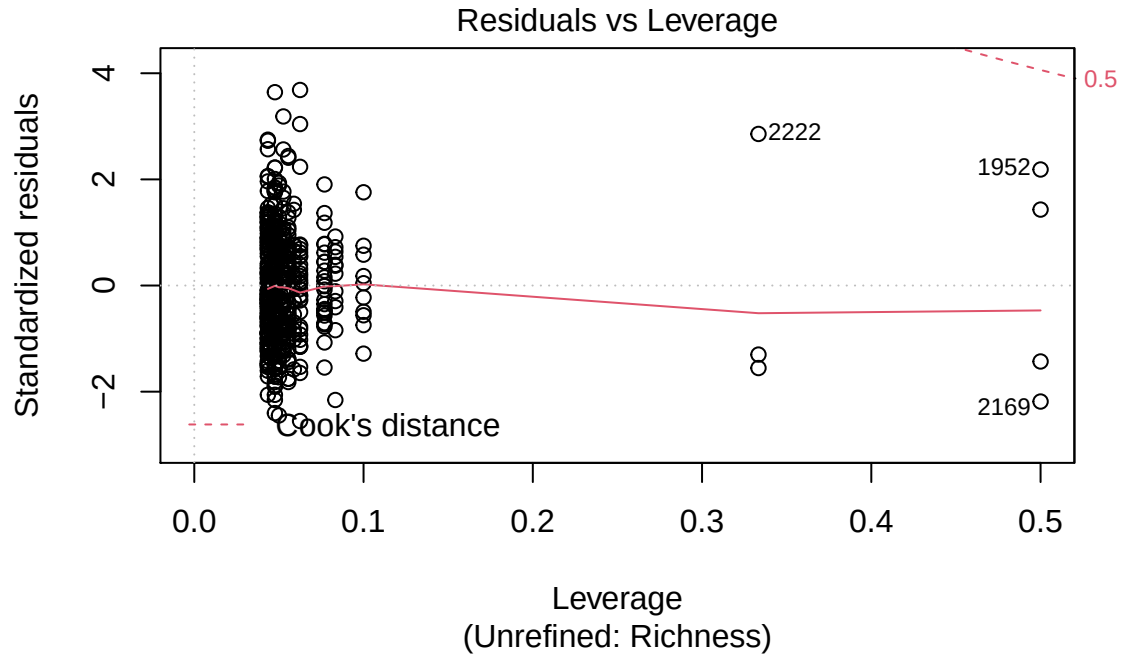
```
index <- "Richness"
plot(caseInfStats[["mod.unref"]][[index]], sub = paste0("(Unrefined: ", methods.alpha[index],")") )
```

```
## Warning: not plotting observations with leverage one:
##   234
```


Residuals vs Fitted
1906
1712
2239
Residuals
200
100
0
−200
150
200
250
300
350
400
450
Fitted values
(Unrefined: Richness)


Normal Q−Q
1906
1712
2239
Standardized residuals
4
3
2
1
0
−2
−3
−2
−1
0
1
2
3
Theoretical Quantiles
(Unrefined: Richness)





Remove unusual observations

The previous plots revealed that the data is largely consistent, but there are some inconsistencies in variance and non-normality for some of the alpha indices. To correct for this, we applied case influence statistics to

remove any unusually influential and leveraged samples. We removed any samples using cutoffs of 2p/n (p is the number of unknown parameters in the model) for leverage, 1 for Cook's Distance and above 2 or below -2 for standardized residuals.

```
caseInfStats <- list()

data <- ps.list$RAR[["dt.all.alpha.melt"]]
# data <- dplyr::mutate(data, obs_num = row_number(), .before = 1)

caseInfStats[["mod.unref"]] <- lapply(methods.alpha, function(alpha){
  lm( formula = "Alpha.Score ~ sex*HMPbodysubsite",
       data = subset(data, Alpha.Metric == alpha)
      )
})

# Fortify data for plotting
caseInfStats[["dataFort.unref"]] <- lapply(methods.alpha, function(alpha){
  fortify(caseInfStats$mod.unref[[alpha]], subset(data, Alpha.Metric == alpha))
})

# Rename some column names
lapply(methods.alpha, function(alpha){
  setnames(caseInfStats$dataFort.unref[[alpha]],
           old=c(".hat", ".cooksd", ".stdresid"),
           new=c("Lev", "CooksD", "StdResid"))
})


# Unrefinded model

## Plot

caseInfStats[["unref.plot"]] <- lapply(names(methods.alpha), function(alpha){
  qplot(obs_num, value, data = reshape::melt(caseInfStats$dataFort.unref[[alpha]][, c("obs_num","Lev",
    id.vars = "obs_num")) +
    geom_point(aes(color = variable)) +
    facet_grid(variable ~ ., scale = "free_y") +
    labs(title = paste0("Case-influence statistics plot: Unrefined model (", alpha,")")) +
    scale_color_brewer(palette = "Dark2") +
    theme(legend.position = "none") + scale_x_continuous(breaks = scales::breaks_pretty(10))

})

names(caseInfStats[["unref.plot"]]) <- names(methods.alpha)
```

From the original 562 samples, we removed 45, 54, 33, and 31 unusual observations in the Shannon, Simpson, Phylogenetic and Richness indices, respectively.

Fit refined model

After removing the unusual samples, we refit the data to a linear model.

```r
# Refined model

### Statisticians use rough cutoffs of 2p/n (p is the number of unknown parameters in the model) for le
#     1 for Cook's Distance and above 2 or below -2 for standardized residuals.
#     Observations falling outside these ranges warrant further attention.

caseInfStats[["cutoff.lev"]] <- lapply(methods.alpha, function(alpha){
  cutOff.lev <- (2 * length(caseInfStats$mod.unref[[alpha]][["coefficients"]]) /
                   nrow(caseInfStats$dataFort.unref[[alpha]]))
})


caseInfStats[["dataFort.sub"]] <- lapply(methods.alpha, function(alpha){
  caseInfStats$dataFort.unref[[alpha]]  %>%
    dplyr::filter(Lev < caseInfStats[["cutoff.lev"]][[alpha]]) %>%  # Leverage Cut off
    dplyr::filter(StdResid < 2 & StdResid > -2 ) %>%  # StdResid Cut Off
    dplyr::select(obs_num:Alpha.Score)  # Removes the old case statistic influence data
})


# make refined model
caseInfStats[["mod.ref"]] <- lapply(methods.alpha, function(alpha){
  lm( formula = "Alpha.Score ~ sex*HMPbodysubsite",
       data = caseInfStats[["dataFort.sub"]][[alpha]]
      )
})


# Fortify data for plotting
caseInfStats[["dataFort.ref"]] <- lapply(methods.alpha, function(alpha){
  fortify(caseInfStats$mod.ref[[alpha]], caseInfStats[["dataFort.sub"]][[alpha]])
})

# Rename some column names
lapply(methods.alpha, function(alpha){
  setnames(caseInfStats$dataFort.ref[[alpha]],
           old=c(".hat", ".cooksd", ".stdresid"),
           new=c("Lev", "CooksD", "StdResid"))
})


# Refinded model Plot

caseInfStats[["ref.plot"]] <- lapply(names(methods.alpha), function(alpha){
  tmp.data <- caseInfStats[["dataFort.ref"]]
  qplot(obs_num, value, data = reshape::melt(tmp.data[[alpha]][, c("obs_num","Lev", "CooksD", "StdResid
    id.vars = "obs_num")) +
    geom_point(aes(color = variable)) +
    facet_grid(variable ~ ., scale = "free_y") +
    labs(title = paste0("Case-influence statistics plot: Refined model (", alpha,")")) +
    scale_color_brewer(palette = "Dark2") +
    theme(legend.position = "none") + scale_x_continuous(breaks = scales::breaks_pretty(10))
```

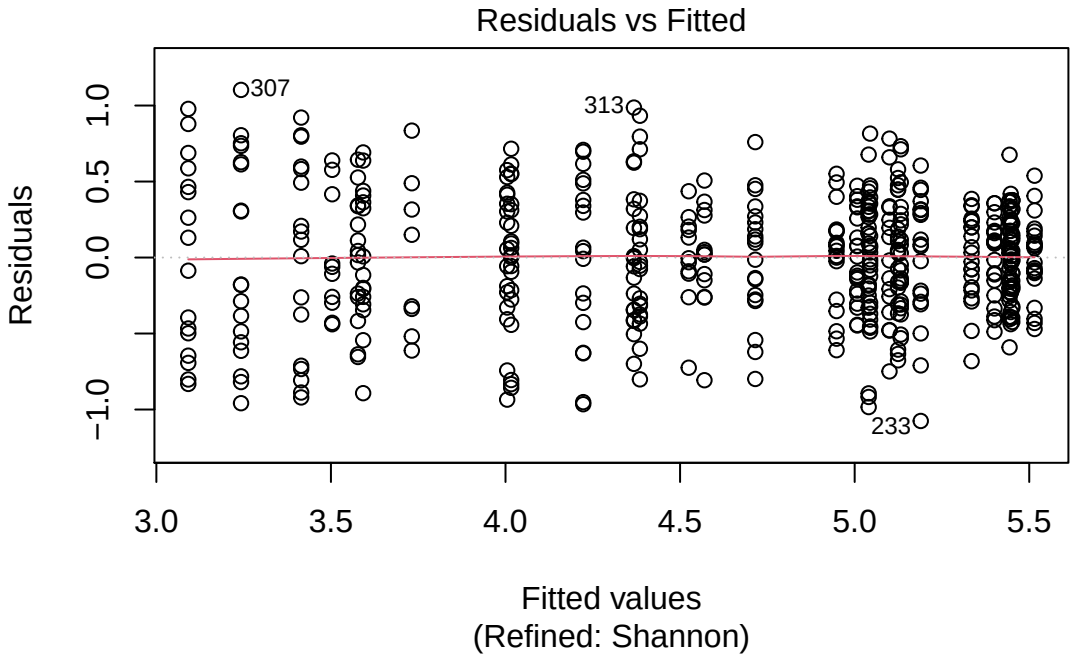
```
})

names(caseInfStats[["ref.plot"]]) <- names(methods.alpha)
```

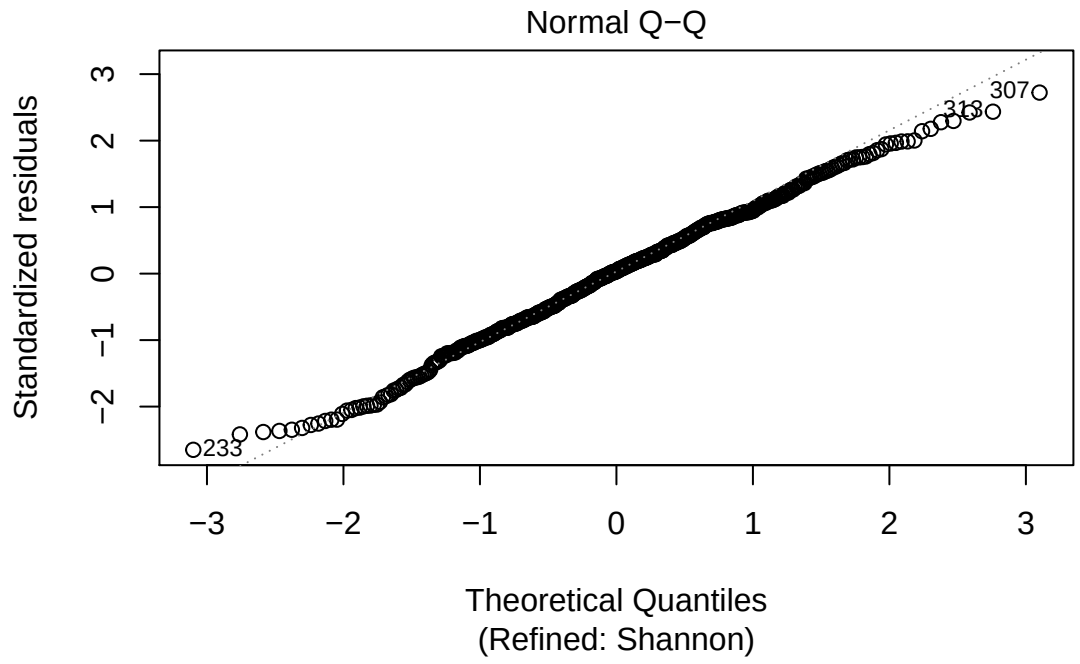
Check assumptions (refined model)

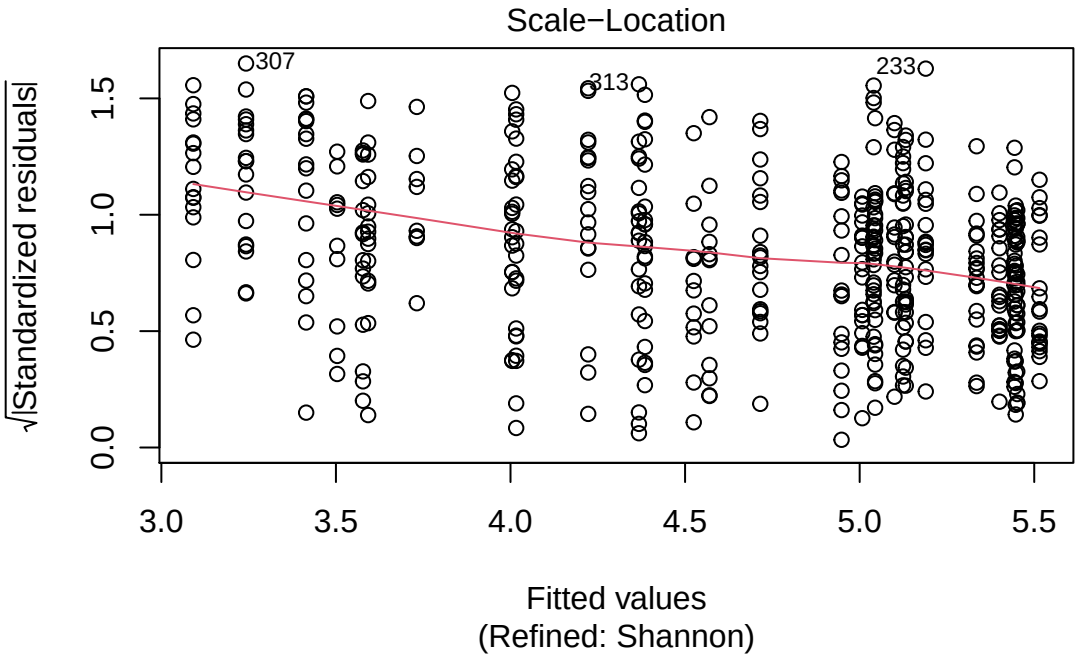
After removing the unusual observations, the plots for each indice appear to have been improved. However, the Simpson's data appears to not meet the assumptions of normality.

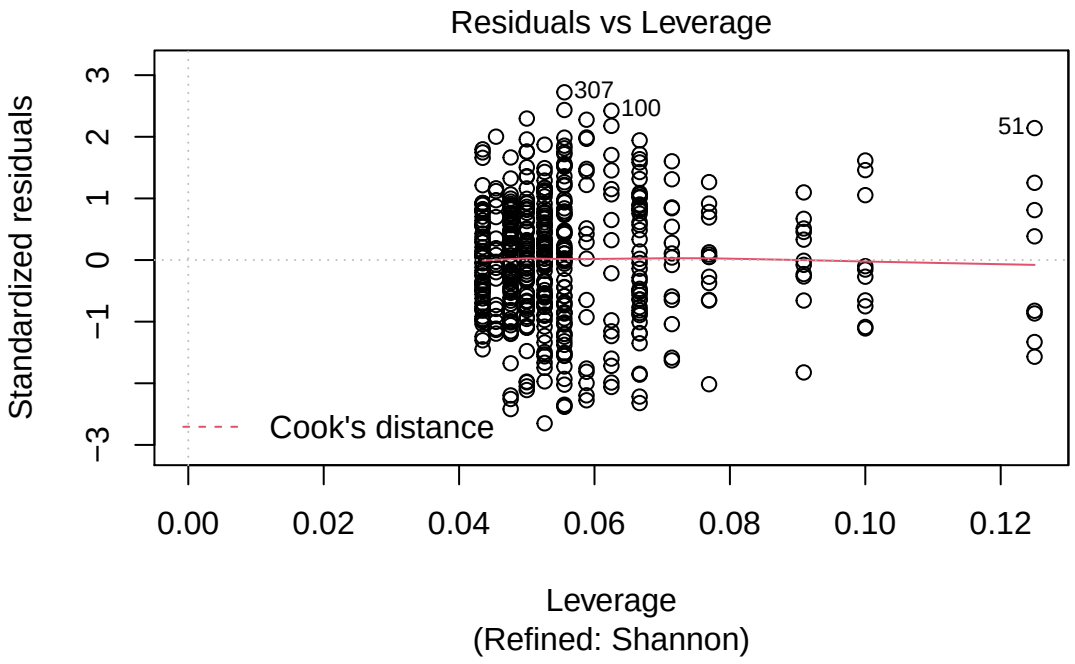
Shannon

```
index <- "Shannon"
plot(caseInfStats[["mod.ref"]][[index]], sub = paste0("(Refined: ", methods.alpha[index],")") )
```



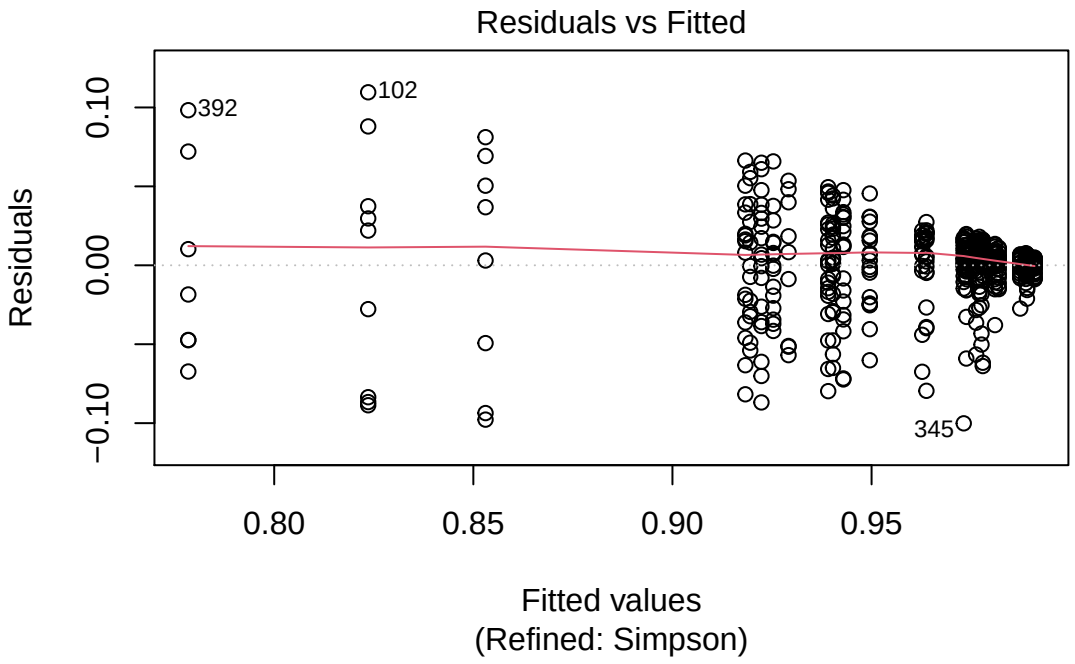

Normal Q−Q
Standardized residuals
Theoretical Quantiles
(Refined: Shannon)
233
313
307
−3
−2
−1
0
1
2
3

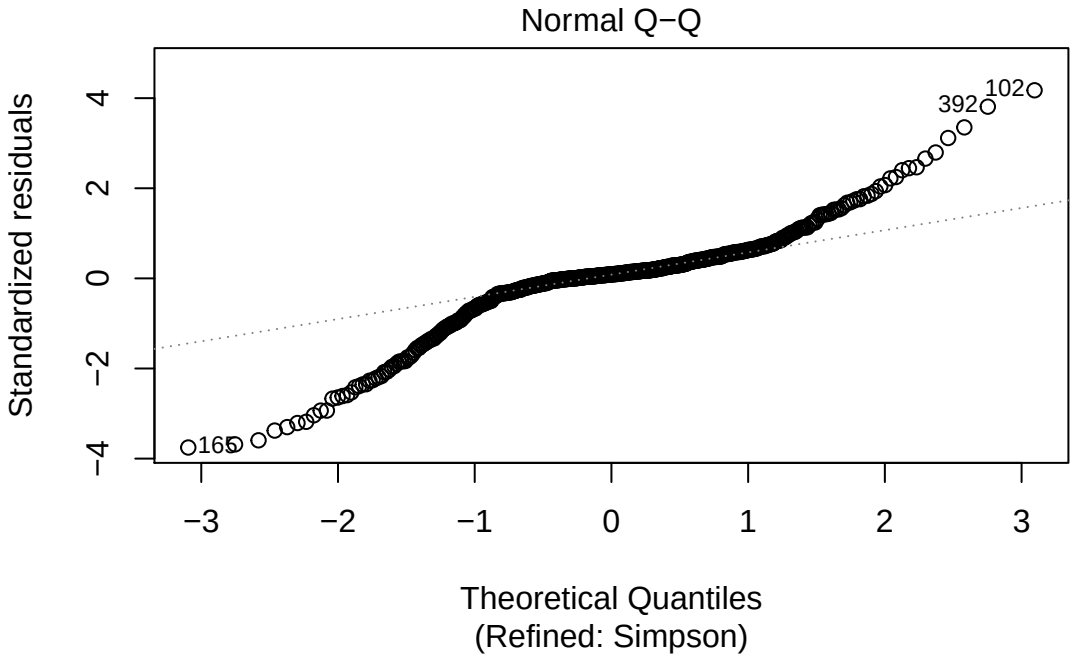

Scale−Location
√|Standardized residuals|
Fitted values
(Refined: Shannon)
307
313
233
0.0
0.5
1.0
1.5
3.0
3.5
4.0
4.5
5.0
5.5

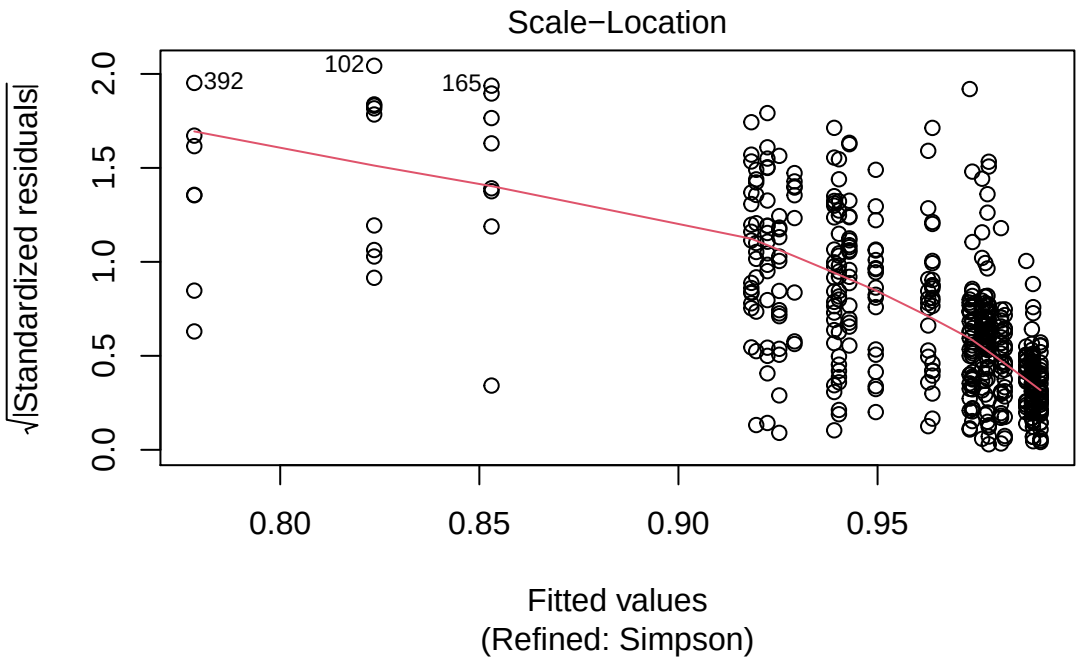


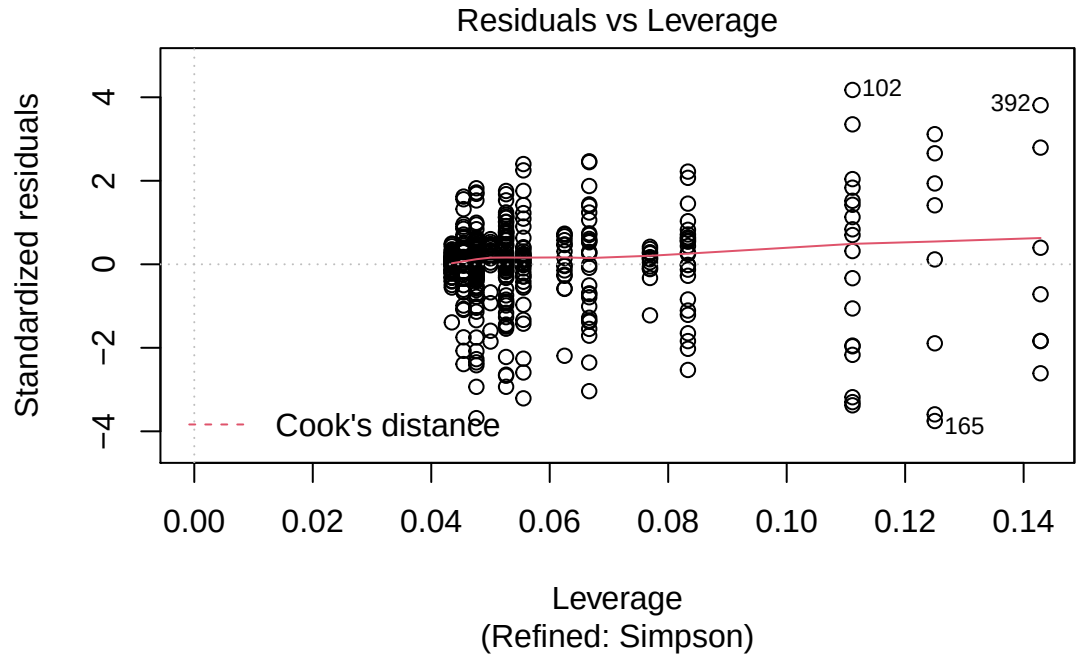
Simpson

```
index <- "Simpson"
plot(caseInfStats[["mod.ref"]][[index]], sub = paste0("(Refined: ", methods.alpha[index],")") )
```


Residuals vs Fitted
Residuals
0.10
0.00
−0.10
392
102
345
0.80
0.85
0.90
0.95
Fitted values
(Refined: Simpson)

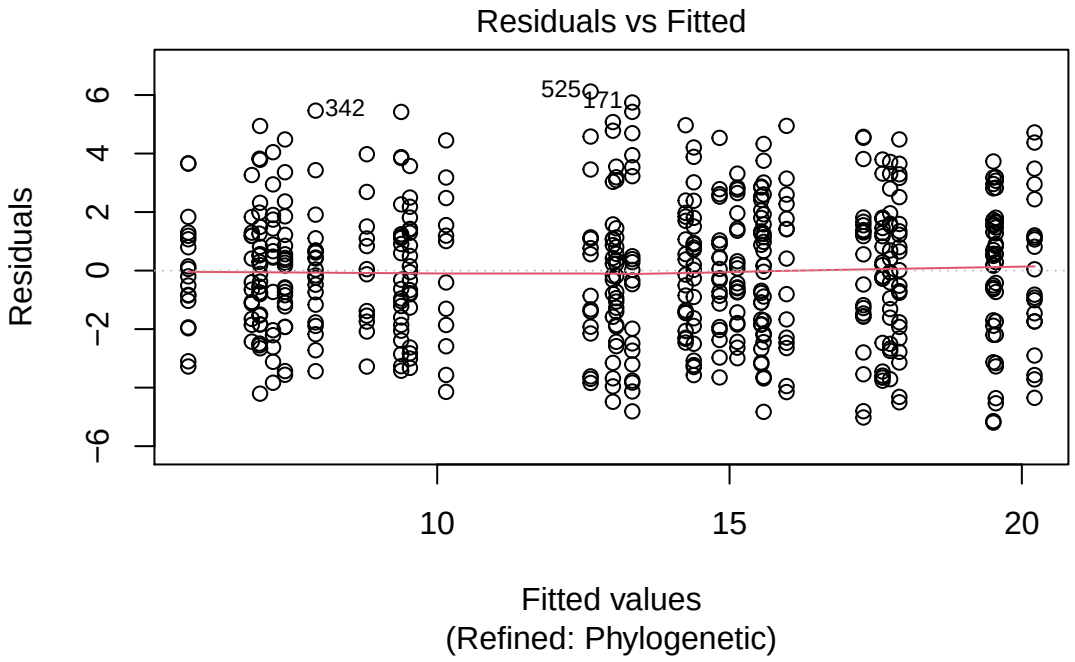

Normal Q−Q
Standardized residuals
4
2
0
−2
−4
392
102
165
−3
−2
−1
0
1
2
3
Theoretical Quantiles
(Refined: Simpson)

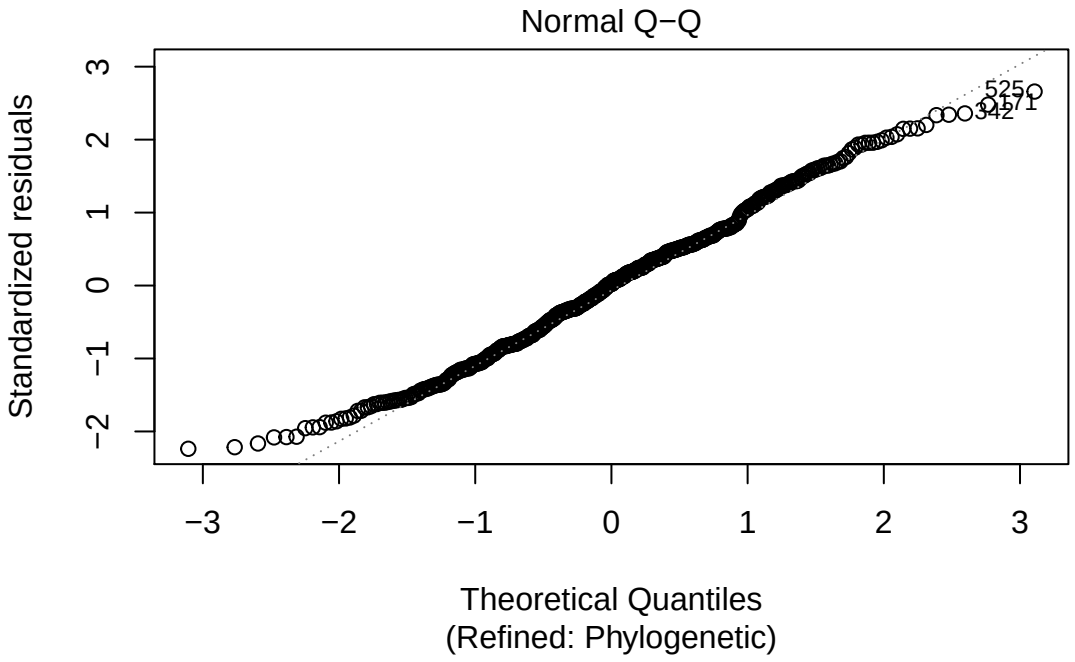

Scale−Location
√|Standardized residuals|
392
102
165
0.0
0.5
1.0
1.5
2.0
0.80
0.85
0.90
0.95
Fitted values
(Refined: Simpson)

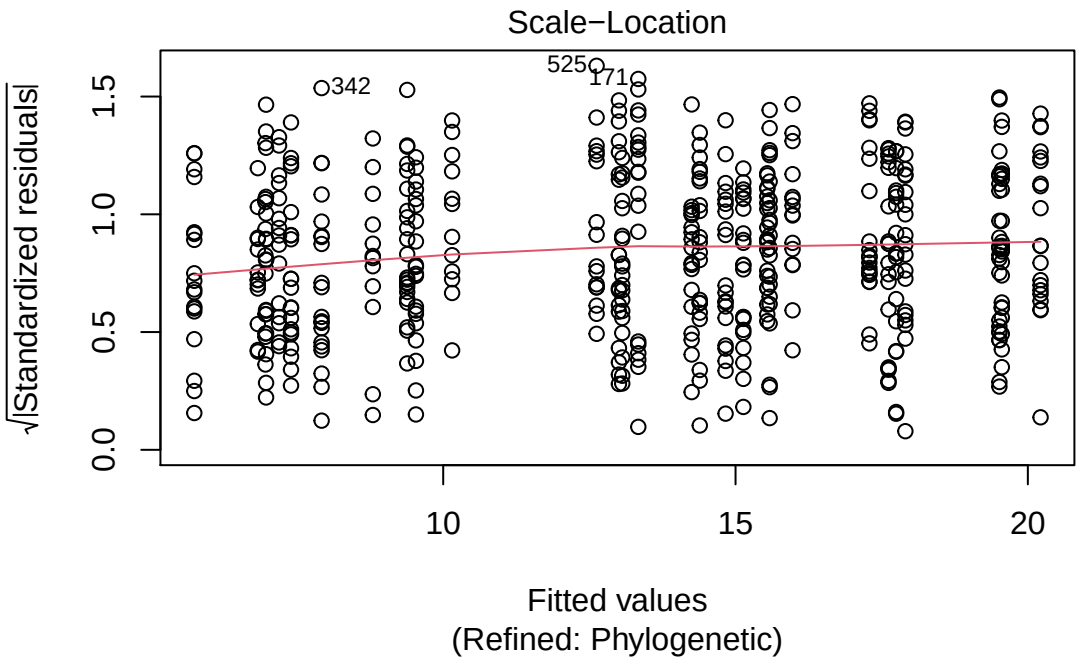

Residuals vs Leverage
Standardized residuals
102
392
165
Cook's distance
−4
−2
0
2
4
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Leverage
(Refined: Simpson)

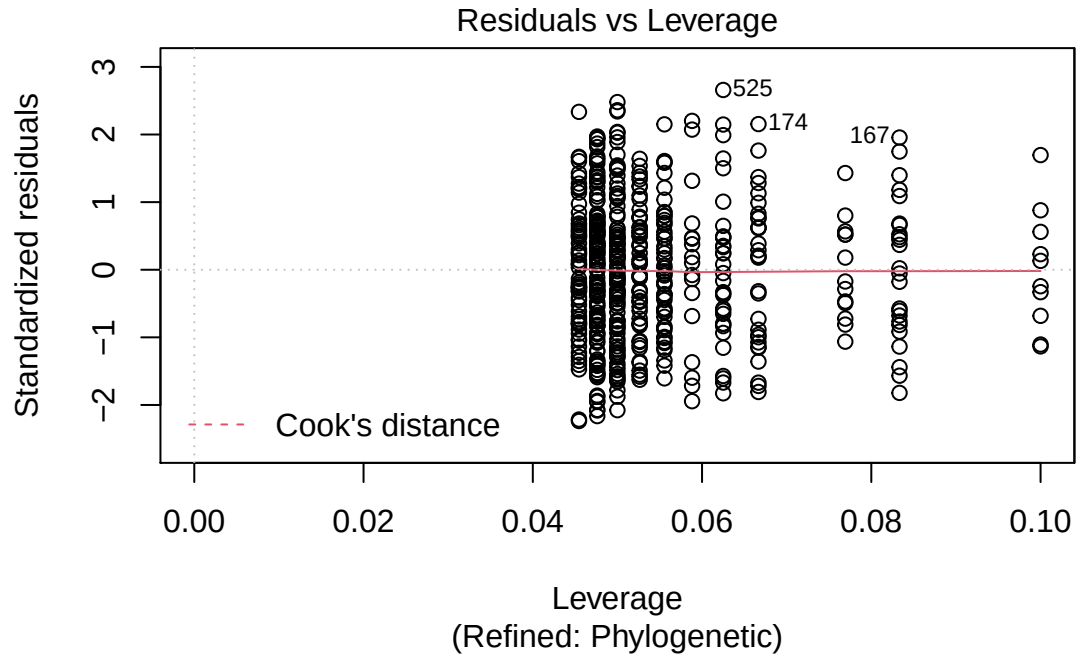
Phylogenetic

```
index <- "Phylogenetic"
plot(caseInfStats[["mod.ref"]][[index]], sub = paste0("(Refined: ", methods.alpha[index],")") )
```



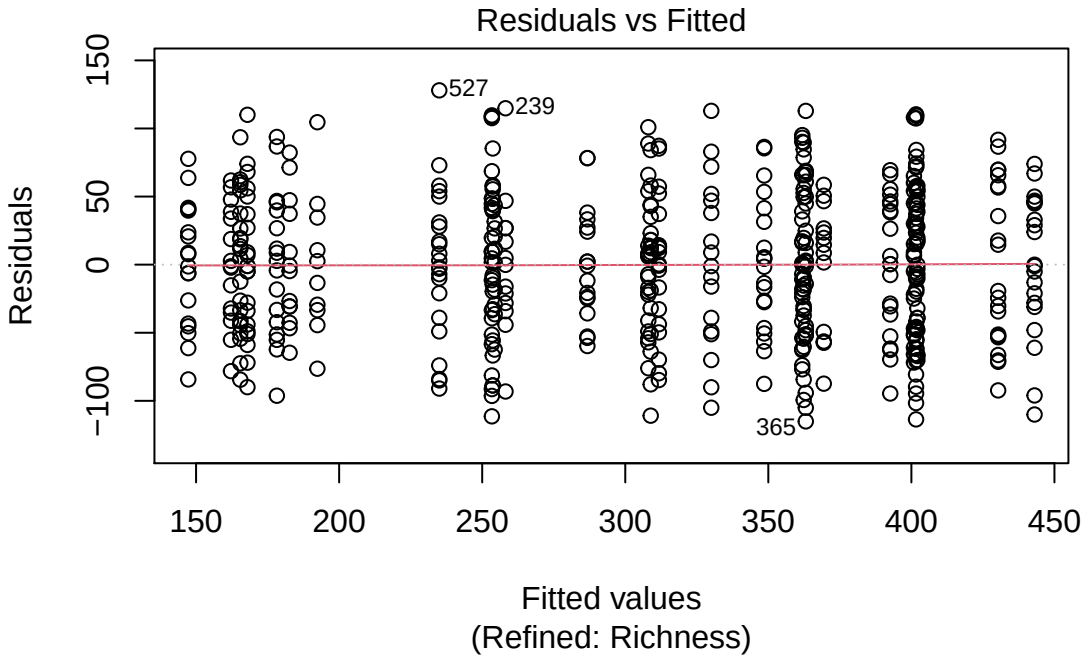


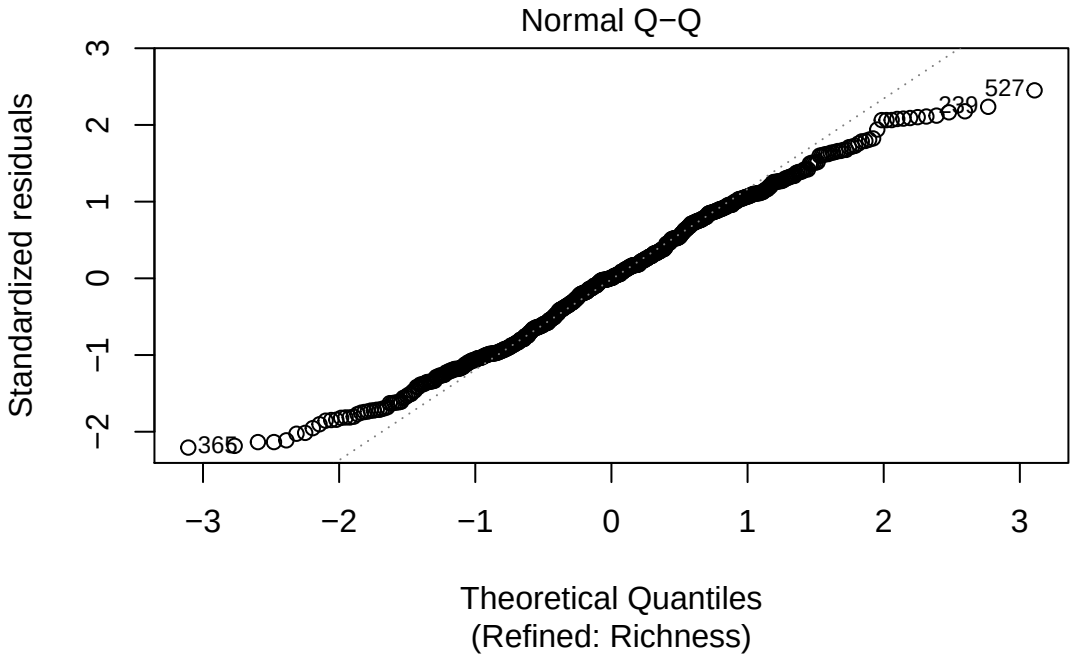

Scale−Location
√|Standardized residuals|
342
525
171
0.0
0.5
1.0
1.5
10
15
20
Fitted values
(Refined: Phylogenetic)

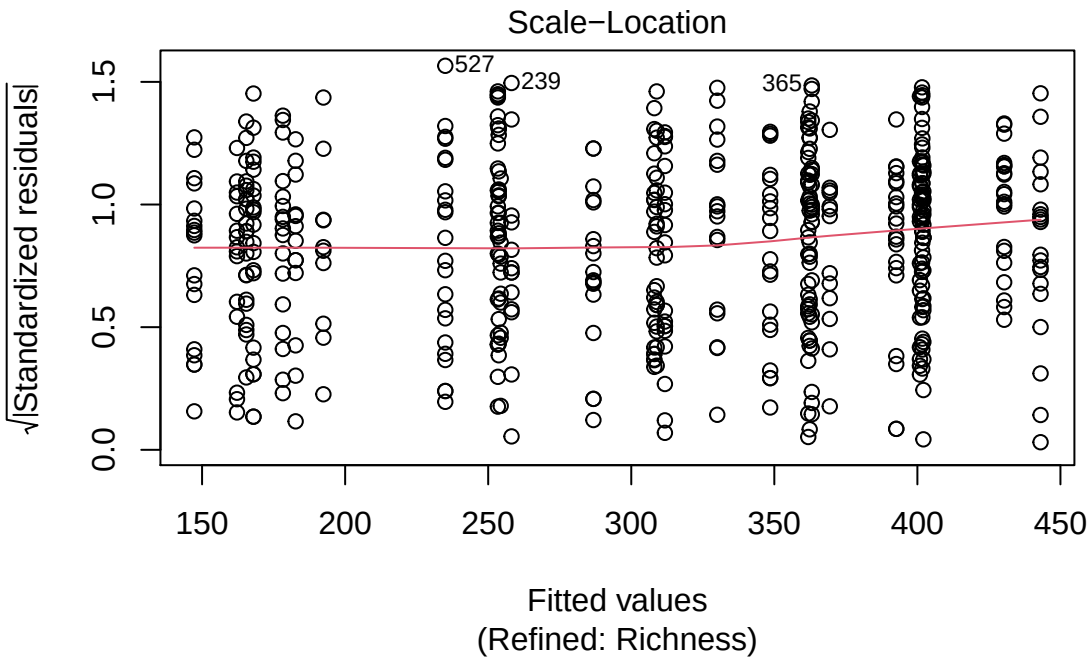

Residuals vs Leverage
Standardized residuals
525
174
167
Cook's distance
−2
−1
0
1
2
3
0.00
0.02
0.04
0.06
0.08
0.10
Leverage
(Refined: Phylogenetic)

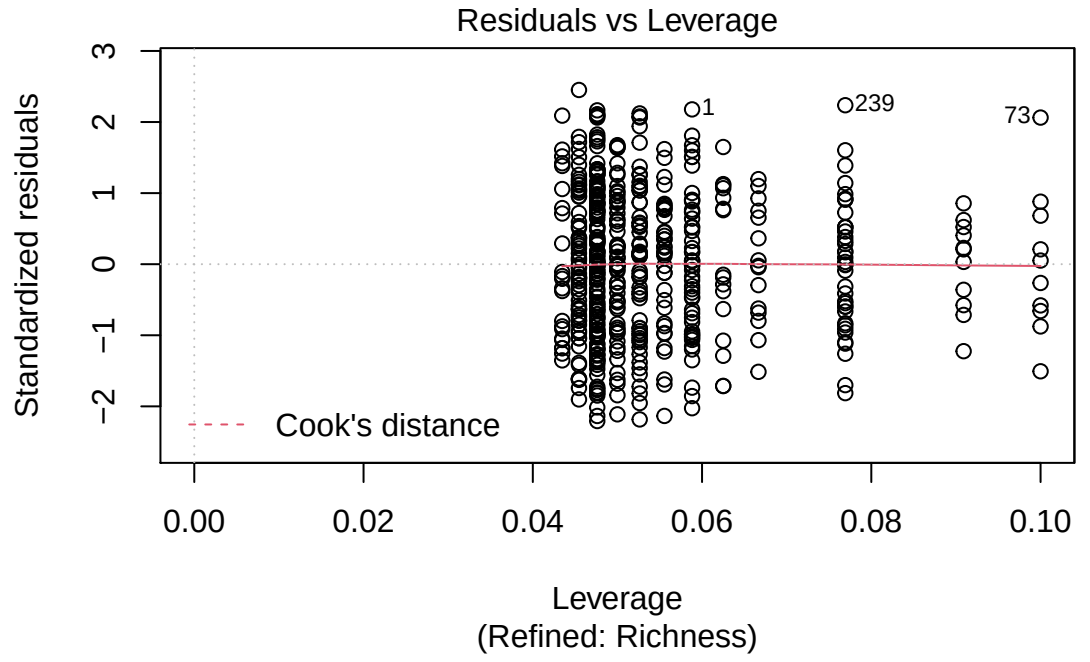
Richness

```
index <- "Richness"
plot(caseInfStats[["mod.ref"]][[index]], sub = paste0("(Refined: ", methods.alpha[index],")") )
```



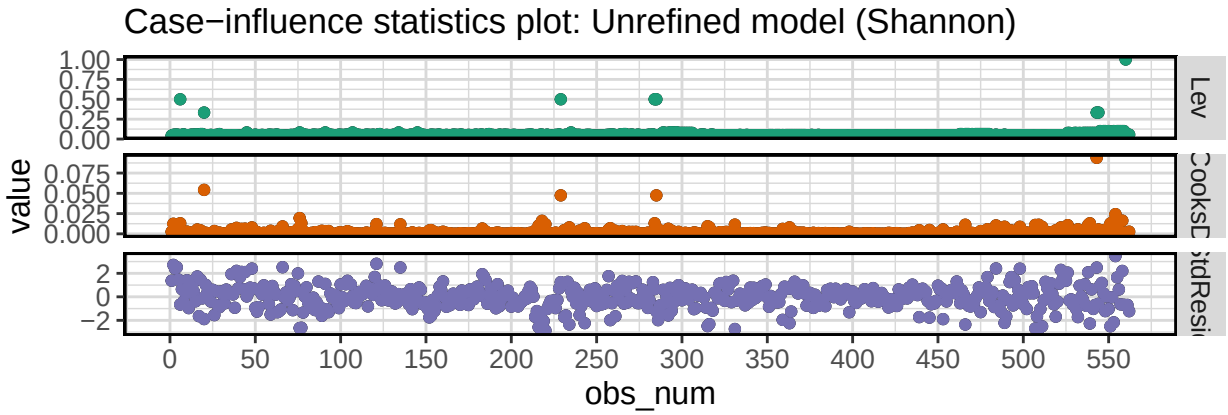


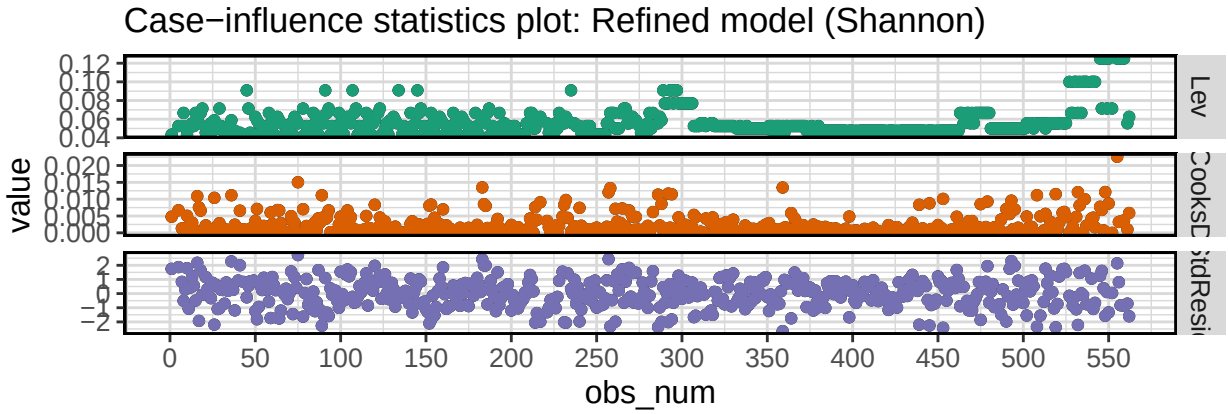




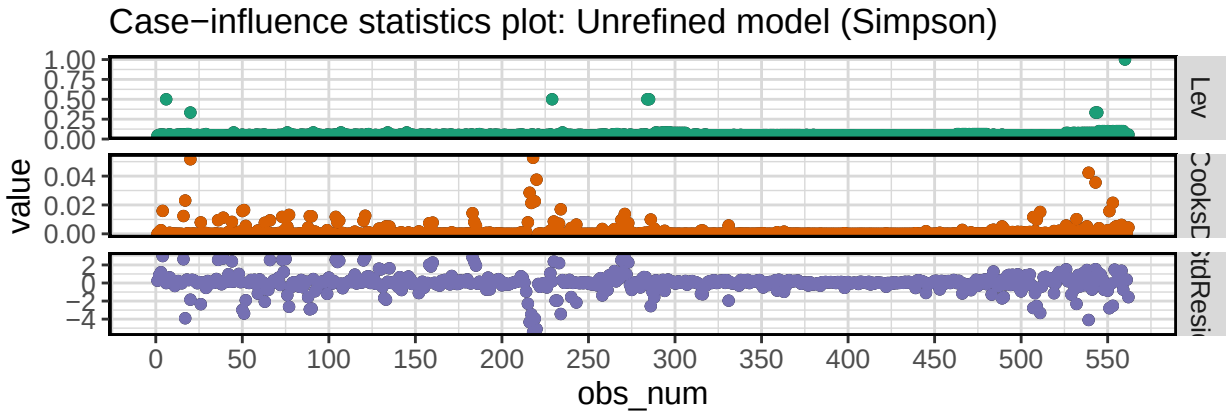
Plot unrefined and refined models

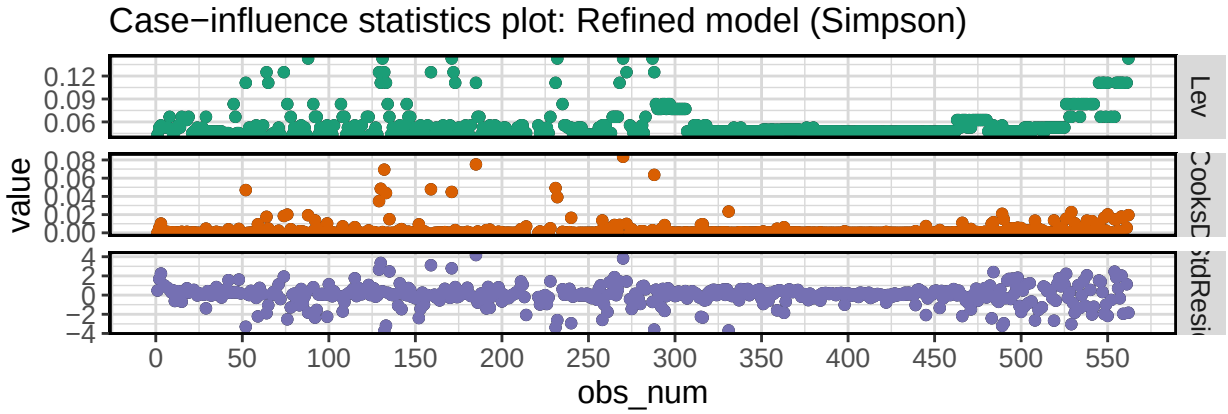
```
## $Shannon
```


Case−influence statistics plot: Unrefined model (Shannon)
value
obs_num
Lev
CooksD
StdResi

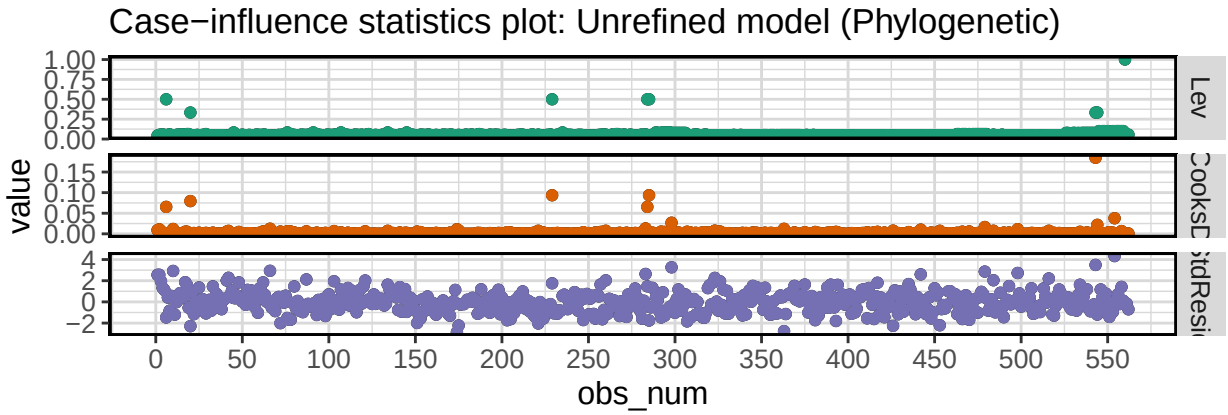

Case−influence statistics plot: Refined model (Shannon)
value
obs_num
Lev
CooksD
StdResi

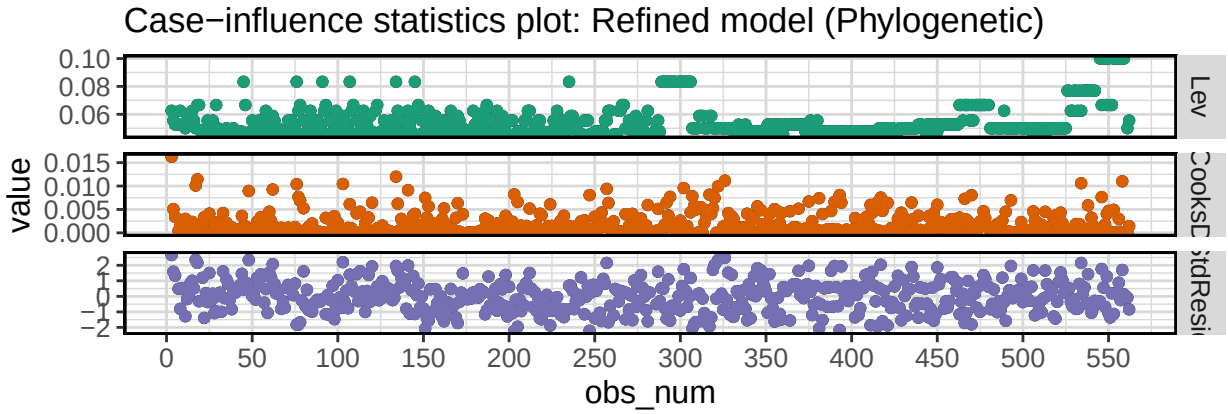
```
##
## $Simpson
```



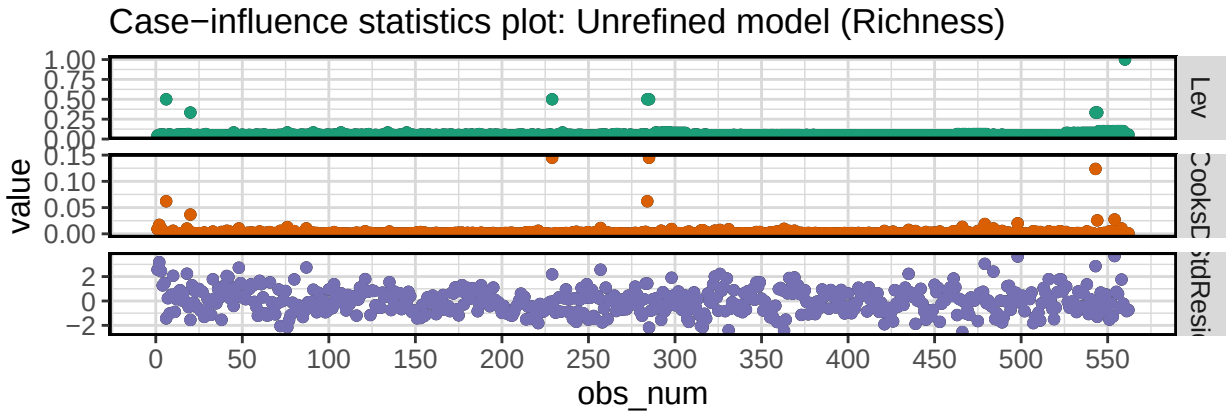


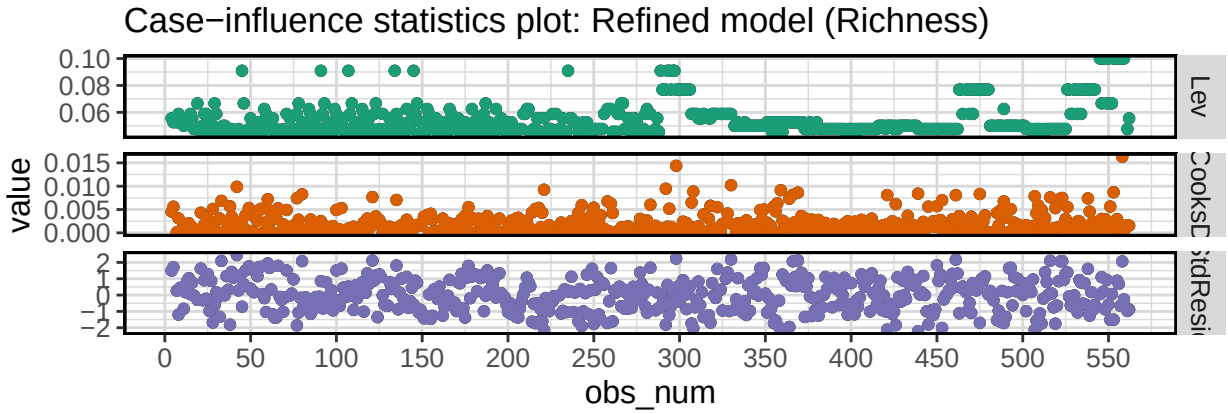

```
##
## $Phylogenetic
```


Case−influence statistics plot: Unrefined model (Phylogenetic)
value
obs_num
Lev
CooksD
StdResi


Case−influence statistics plot: Refined model (Phylogenetic)
value
obs_num
Lev
CooksD
StdResi

```
##
## $Richness
```





Linear model results

Shannon

```
index <- "Shannon"

Anova(caseInfStats[["mod.ref"]][[index]], type = 2)
```

```
## Anova Table (Type II tests)
##
## Response: Alpha.Score
##                       Sum Sq  Df F value    Pr(>F)
## sex                    2.139   1 12.3088 0.0004925 ***
## HMPbodysubsite       252.107  15 96.7150 < 2.2e-16 ***
## sex:HMPbodysubsite     5.719  12  2.7425 0.0013018 **
## Residuals             84.804 488
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
caseInfStats[["mod.ref"]][[index]] %>% report()
```

```
## We fitted a linear model (estimated using OLS) to predict Alpha.Score with sex and HMPbodysubsite (f
```

```
##
##   - The effect of sex [male] is statistically non-significant and positive (beta = 0.05, 95% CI [-0.1
##   - The effect of HMPbodysubsite [Attached Keratinized gingiva] is statistically significant and nega
##   - The effect of HMPbodysubsite [Buccal mucosa] is statistically significant and negative (beta = -0
##   - The effect of HMPbodysubsite [Hard palate] is statistically non-significant and negative (beta =
##   - The effect of HMPbodysubsite [Left Retroauricular crease] is statistically significant and negati
##   - The effect of HMPbodysubsite [Mid vagina] is statistically significant and negative (beta = -1.11
##   - The effect of HMPbodysubsite [Palatine Tonsils] is statistically significant and positive (beta =
##   - The effect of HMPbodysubsite [Posterior fornix] is statistically significant and negative (beta =
##   - The effect of HMPbodysubsite [Right Retroauricular crease] is statistically significant and negat
##   - The effect of HMPbodysubsite [Saliva] is statistically significant and positive (beta = 0.92, 95%
##   - The effect of HMPbodysubsite [Stool] is statistically significant and positive (beta = 0.42, 95%
##   - The effect of HMPbodysubsite [Subgingival plaque] is statistically significant and positive (beta
##   - The effect of HMPbodysubsite [Supragingival plaque] is statistically significant and positive (be
##   - The effect of HMPbodysubsite [Throat] is statistically significant and positive (beta = 0.52, 95%
##   - The effect of HMPbodysubsite [Tongue dorsum] is statistically non-significant and positive (beta
##   - The effect of HMPbodysubsite [Vaginal introitus] is statistically significant and negative (beta
##   - The interaction effect of HMPbodysubsite [Attached Keratinized gingiva] on sex [male] is statisti
##   - The interaction effect of HMPbodysubsite [Buccal mucosa] on sex [male] is statistically non-signi
##   - The interaction effect of HMPbodysubsite [Hard palate] on sex [male] is statistically significant
##   - The interaction effect of HMPbodysubsite [Left Retroauricular crease] on sex [male] is statistica
##   - The interaction effect of HMPbodysubsite [Mid vagina] on sex [male] is statistically non-signifi
##   - The interaction effect of HMPbodysubsite [Palatine Tonsils] on sex [male] is statistically non-si
##   - The interaction effect of HMPbodysubsite [Posterior fornix] on sex [male] is statistically non-si
##   - The interaction effect of HMPbodysubsite [Right Retroauricular crease] on sex [male] is statistic
##   - The interaction effect of HMPbodysubsite [Saliva] on sex [male] is statistically non-significant
##   - The interaction effect of HMPbodysubsite [Stool] on sex [male] is statistically non-significant a
##   - The interaction effect of HMPbodysubsite [Subgingival plaque] on sex [male] is statistically non-
##   - The interaction effect of HMPbodysubsite [Supragingival plaque] on sex [male] is statistically no
##   - The interaction effect of HMPbodysubsite [Tongue dorsum] on sex [male] is statistically non-signi
##   - The interaction effect of HMPbodysubsite [Vaginal introitus] on sex [male] is statistically signi
##
## Standardized parameters were obtained by fitting the model on a standardized version of the dataset.
```

```
Anova(caseInfStats[["mod.ref"]][[index]], type = 2) %>% report()
```

```
## The ANOVA suggests that:
##
##   - The main effect of sex is statistically significant and small (F(1, 488) = 12.31, p < .001; Eta2
##   - The main effect of HMPbodysubsite is statistically significant and large (F(15, 488) = 96.72, p <
##   - The interaction between sex and HMPbodysubsite is statistically significant and medium (F(12, 488
##
## Effect sizes were labelled following Field's (2013) recommendations.
```

Simpson

```
index <- "Simpson"

Anova(caseInfStats[["mod.ref"]][[index]], type = 2)
```

```
## Anova Table (Type II tests)
```

```
##
## Response: Alpha.Score
##                      Sum Sq  Df F value Pr(>F)
## sex                 0.00170   1  2.1934 0.1393
## HMPbodysubsite      0.72782  15 62.5740 <2e-16 ***
## sex:HMPbodysubsite 0.00851  12  0.9150 0.5315
## Residuals           0.37143 479
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
caseInfStats[["mod.ref"]][[index]] %>% report()
```

```
## We fitted a linear model (estimated using OLS) to predict Alpha.Score with sex and HMPbodysubsite (f
##
##   - The effect of sex [male] is statistically non-significant and positive (beta = 0.01, 95% CI [-0.
##   - The effect of HMPbodysubsite [Attached Keratinized gingiva] is statistically significant and nega
##   - The effect of HMPbodysubsite [Buccal mucosa] is statistically significant and negative (beta = -
##   - The effect of HMPbodysubsite [Hard palate] is statistically non-significant and negative (beta =
##   - The effect of HMPbodysubsite [Left Retroauricular crease] is statistically significant and negati
##   - The effect of HMPbodysubsite [Mid vagina] is statistically significant and negative (beta = -0.1
##   - The effect of HMPbodysubsite [Palatine Tonsils] is statistically non-significant and positive (be
##   - The effect of HMPbodysubsite [Posterior fornix] is statistically significant and negative (beta =
##   - The effect of HMPbodysubsite [Right Retroauricular crease] is statistically significant and negat
##   - The effect of HMPbodysubsite [Saliva] is statistically significant and positive (beta = 0.03, 95%
##   - The effect of HMPbodysubsite [Stool] is statistically non-significant and positive (beta = 0.02,
##   - The effect of HMPbodysubsite [Subgingival plaque] is statistically significant and positive (beta
##   - The effect of HMPbodysubsite [Supragingival plaque] is statistically significant and positive (be
##   - The effect of HMPbodysubsite [Throat] is statistically non-significant and positive (beta = 0.01
##   - The effect of HMPbodysubsite [Tongue dorsum] is statistically non-significant and positive (beta
##   - The effect of HMPbodysubsite [Vaginal introitus] is statistically significant and negative (beta
##   - The interaction effect of HMPbodysubsite [Attached Keratinized gingiva] on sex [male] is statisti
##   - The interaction effect of HMPbodysubsite [Buccal mucosa] on sex [male] is statistically non-signi
##   - The interaction effect of HMPbodysubsite [Hard palate] on sex [male] is statistically non-signifi
##   - The interaction effect of HMPbodysubsite [Left Retroauricular crease] on sex [male] is statistica
##   - The interaction effect of HMPbodysubsite [Mid vagina] on sex [male] is statistically non-signifi
##   - The interaction effect of HMPbodysubsite [Palatine Tonsils] on sex [male] is statistically non-si
##   - The interaction effect of HMPbodysubsite [Posterior fornix] on sex [male] is statistically non-si
##   - The interaction effect of HMPbodysubsite [Right Retroauricular crease] on sex [male] is statistic
##   - The interaction effect of HMPbodysubsite [Saliva] on sex [male] is statistically non-significant
##   - The interaction effect of HMPbodysubsite [Stool] on sex [male] is statistically non-significant a
##   - The interaction effect of HMPbodysubsite [Subgingival plaque] on sex [male] is statistically non-
##   - The interaction effect of HMPbodysubsite [Supragingival plaque] on sex [male] is statistically no
##   - The interaction effect of HMPbodysubsite [Tongue dorsum] on sex [male] is statistically non-signi
##   - The interaction effect of HMPbodysubsite [Vaginal introitus] on sex [male] is statistically signi
##
## Standardized parameters were obtained by fitting the model on a standardized version of the dataset.
```

```
Anova(caseInfStats[["mod.ref"]][[index]], type = 2) %>% report()
```

```
## The ANOVA suggests that:
##
##   - The main effect of sex is statistically not significant and very small (F(1, 479) = 2.19, p = 0.1
##   - The main effect of HMPbodysubsite is statistically significant and large (F(15, 479) = 62.57, p <
```

```
##   - The interaction between sex and HMPbodysubsite is statistically not significant and small (F(12,
##
## Effect sizes were labelled following Field's (2013) recommendations.
```

Phylogenetic

```
index <- "Phylogenetic"

Anova(caseInfStats[["mod.ref"]][[index]], type = 2)
```

```
## Anova Table (Type II tests)
##
## Response: Alpha.Score
##                     Sum Sq  Df  F value    Pr(>F)
## sex                   22.3   1   3.9495  0.047430 *
## HMPbodysubsite     9391.5  15 110.8586 < 2.2e-16 ***
## sex:HMPbodysubsite  172.8  12   2.5503  0.002794 **
## Residuals          2823.9 500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
caseInfStats[["mod.ref"]][[index]] %>% report()
```

```
## We fitted a linear model (estimated using OLS) to predict Alpha.Score with sex and HMPbodysubsite (f
##
##   - The effect of sex [male] is statistically non-significant and negative (beta = -1.36, 95% CI [-3
##   - The effect of HMPbodysubsite [Attached Keratinized gingiva] is statistically non-significant and
##   - The effect of HMPbodysubsite [Buccal mucosa] is statistically significant and positive (beta = 2
##   - The effect of HMPbodysubsite [Hard palate] is statistically significant and positive (beta = 4.1
##   - The effect of HMPbodysubsite [Left Retroauricular crease] is statistically significant and negati
##   - The effect of HMPbodysubsite [Mid vagina] is statistically significant and negative (beta = -2.7
##   - The effect of HMPbodysubsite [Palatine Tonsils] is statistically significant and positive (beta =
##   - The effect of HMPbodysubsite [Posterior fornix] is statistically significant and negative (beta =
##   - The effect of HMPbodysubsite [Right Retroauricular crease] is statistically significant and negat
##   - The effect of HMPbodysubsite [Saliva] is statistically significant and positive (beta = 9.40, 95%
##   - The effect of HMPbodysubsite [Stool] is statistically significant and positive (beta = 2.85, 95%
##   - The effect of HMPbodysubsite [Subgingival plaque] is statistically significant and positive (beta
##   - The effect of HMPbodysubsite [Supragingival plaque] is statistically significant and positive (be
##   - The effect of HMPbodysubsite [Throat] is statistically significant and positive (beta = 4.98, 95%
##   - The effect of HMPbodysubsite [Tongue dorsum] is statistically significant and positive (beta = 2
##   - The effect of HMPbodysubsite [Vaginal introitus] is statistically significant and negative (beta
##   - The interaction effect of HMPbodysubsite [Attached Keratinized gingiva] on sex [male] is statisti
##   - The interaction effect of HMPbodysubsite [Buccal mucosa] on sex [male] is statistically signifi
##   - The interaction effect of HMPbodysubsite [Hard palate] on sex [male] is statistically significant
##   - The interaction effect of HMPbodysubsite [Left Retroauricular crease] on sex [male] is statistica
##   - The interaction effect of HMPbodysubsite [Mid vagina] on sex [male] is statistically non-signifi
##   - The interaction effect of HMPbodysubsite [Palatine Tonsils] on sex [male] is statistically non-si
##   - The interaction effect of HMPbodysubsite [Posterior fornix] on sex [male] is statistically non-si
##   - The interaction effect of HMPbodysubsite [Right Retroauricular crease] on sex [male] is statistic
##   - The interaction effect of HMPbodysubsite [Saliva] on sex [male] is statistically non-significant
##   - The interaction effect of HMPbodysubsite [Stool] on sex [male] is statistically non-significant a
```

```
##   - The interaction effect of HMPbodysubsite [Subgingival plaque] on sex [male] is statistically sig
##   - The interaction effect of HMPbodysubsite [Supragingival plaque] on sex [male] is statistically s
##   - The interaction effect of HMPbodysubsite [Tongue dorsum] on sex [male] is statistically non-sign
##   - The interaction effect of HMPbodysubsite [Vaginal introitus] on sex [male] is statistically non-
##
## Standardized parameters were obtained by fitting the model on a standardized version of the dataset.
```

```
Anova(caseInfStats[["mod.ref"]][[index]], type = 2) %>% report()
```

```
## The ANOVA suggests that:
##
##   - The main effect of sex is statistically significant and very small (F(1, 500) = 3.95, p = 0.047;
##   - The main effect of HMPbodysubsite is statistically significant and large (F(15, 500) = 110.86, p
##   - The interaction between sex and HMPbodysubsite is statistically significant and small (F(12, 500
##
## Effect sizes were labelled following Field's (2013) recommendations.
```

Richness

```
index <- "Richness"

Anova(caseInfStats[["mod.ref"]][[index]], type = 2)
```

```
## Anova Table (Type II tests)
##
## Response: Alpha.Score
##                     Sum Sq  Df F value    Pr(>F)
## sex                  57658   1  20.181 8.752e-06 ***
## HMPbodysubsite     3907814  15  91.185 < 2.2e-16 ***
## sex:HMPbodysubsite   70147  12   2.046   0.01906 *
## Residuals          1434248 502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
caseInfStats[["mod.ref"]][[index]] %>% report()
```

```
## We fitted a linear model (estimated using OLS) to predict Alpha.Score with sex and HMPbodysubsite (f
##
##   - The effect of sex [male] is statistically non-significant and positive (beta = 3.79, 95% CI [-39
##   - The effect of HMPbodysubsite [Attached Keratinized gingiva] is statistically non-significant and
##   - The effect of HMPbodysubsite [Buccal mucosa] is statistically non-significant and negative (beta
##   - The effect of HMPbodysubsite [Hard palate] is statistically non-significant and positive (beta =
##   - The effect of HMPbodysubsite [Left Retroauricular crease] is statistically significant and negati
##   - The effect of HMPbodysubsite [Mid vagina] is statistically significant and negative (beta = -88.9
##   - The effect of HMPbodysubsite [Palatine Tonsils] is statistically significant and positive (beta =
##   - The effect of HMPbodysubsite [Posterior fornix] is statistically significant and negative (beta =
##   - The effect of HMPbodysubsite [Right Retroauricular crease] is statistically significant and negat
##   - The effect of HMPbodysubsite [Saliva] is statistically significant and positive (beta = 175.97, 9
##   - The effect of HMPbodysubsite [Stool] is statistically significant and positive (beta = 54.53, 95%
##   - The effect of HMPbodysubsite [Subgingival plaque] is statistically significant and positive (beta
```

```
##   - The effect of HMPbodysubsite [Supragingival plaque] is statistically significant and positive (be
##   - The effect of HMPbodysubsite [Throat] is statistically significant and positive (beta = 108.00, 9
##   - The effect of HMPbodysubsite [Tongue dorsum] is statistically significant and positive (beta = 53
##   - The effect of HMPbodysubsite [Vaginal introitus] is statistically significant and negative (beta
##   - The interaction effect of HMPbodysubsite [Attached Keratinized gingiva] on sex [male] is statisti
##   - The interaction effect of HMPbodysubsite [Buccal mucosa] on sex [male] is statistically non-signi
##   - The interaction effect of HMPbodysubsite [Hard palate] on sex [male] is statistically significant
##   - The interaction effect of HMPbodysubsite [Left Retroauricular crease] on sex [male] is statistica
##   - The interaction effect of HMPbodysubsite [Mid vagina] on sex [male] is statistically non-signifi
##   - The interaction effect of HMPbodysubsite [Palatine Tonsils] on sex [male] is statistically non-si
##   - The interaction effect of HMPbodysubsite [Posterior fornix] on sex [male] is statistically non-si
##   - The interaction effect of HMPbodysubsite [Right Retroauricular crease] on sex [male] is statistic
##   - The interaction effect of HMPbodysubsite [Saliva] on sex [male] is statistically non-significant
##   - The interaction effect of HMPbodysubsite [Stool] on sex [male] is statistically non-significant a
##   - The interaction effect of HMPbodysubsite [Subgingival plaque] on sex [male] is statistically non-
##   - The interaction effect of HMPbodysubsite [Supragingival plaque] on sex [male] is statistically no
##   - The interaction effect of HMPbodysubsite [Tongue dorsum] on sex [male] is statistically non-signi
##   - The interaction effect of HMPbodysubsite [Vaginal introitus] on sex [male] is statistically non-s
##
## Standardized parameters were obtained by fitting the model on a standardized version of the dataset.
```

```
Anova(caseInfStats[["mod.ref"]][[index]], type = 2) %>% report()
```

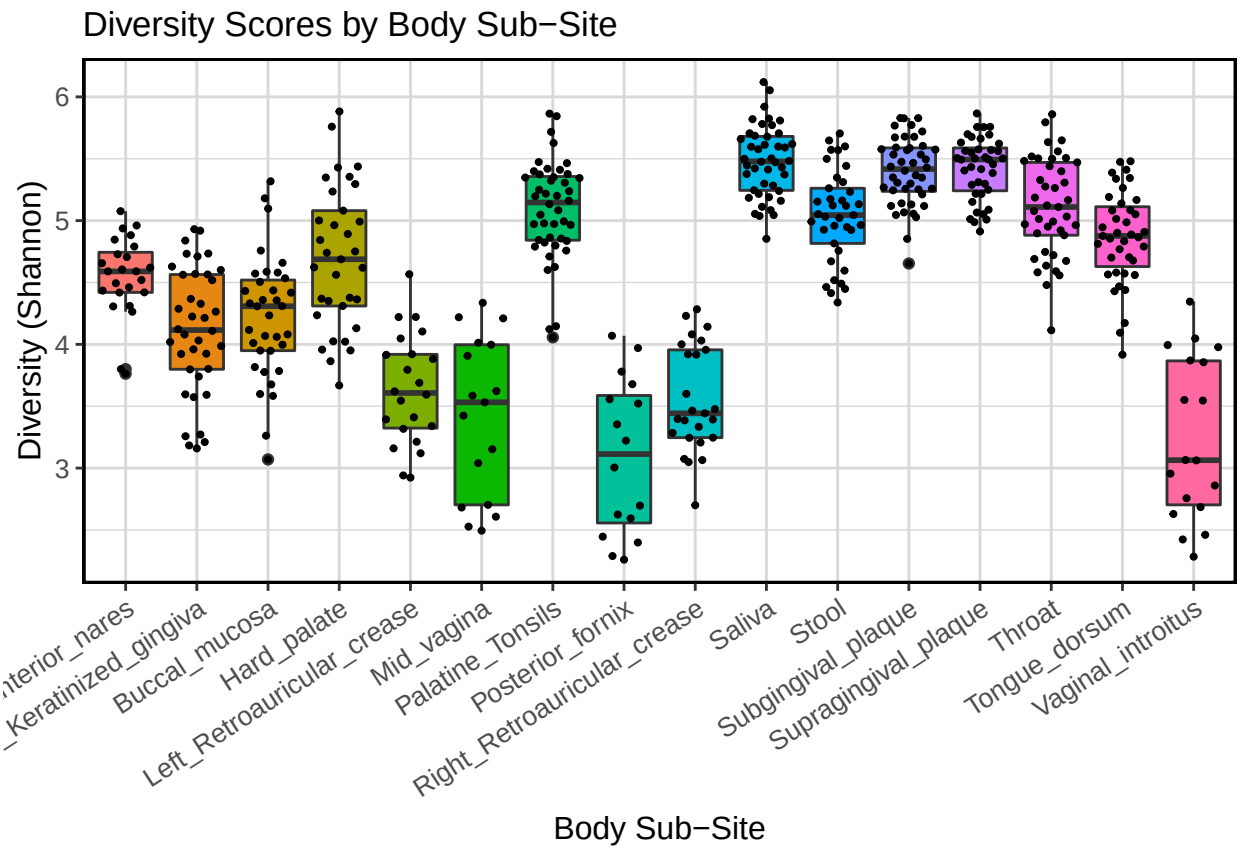
```
## The ANOVA suggests that:
##
##   - The main effect of sex is statistically significant and small (F(1, 502) = 20.18, p < .001; Eta2
##   - The main effect of HMPbodysubsite is statistically significant and large (F(15, 502) = 91.18, p <
##   - The interaction between sex and HMPbodysubsite is statistically significant and small (F(12, 502)
##
## Effect sizes were labelled following Field's (2013) recommendations.
```

Plots

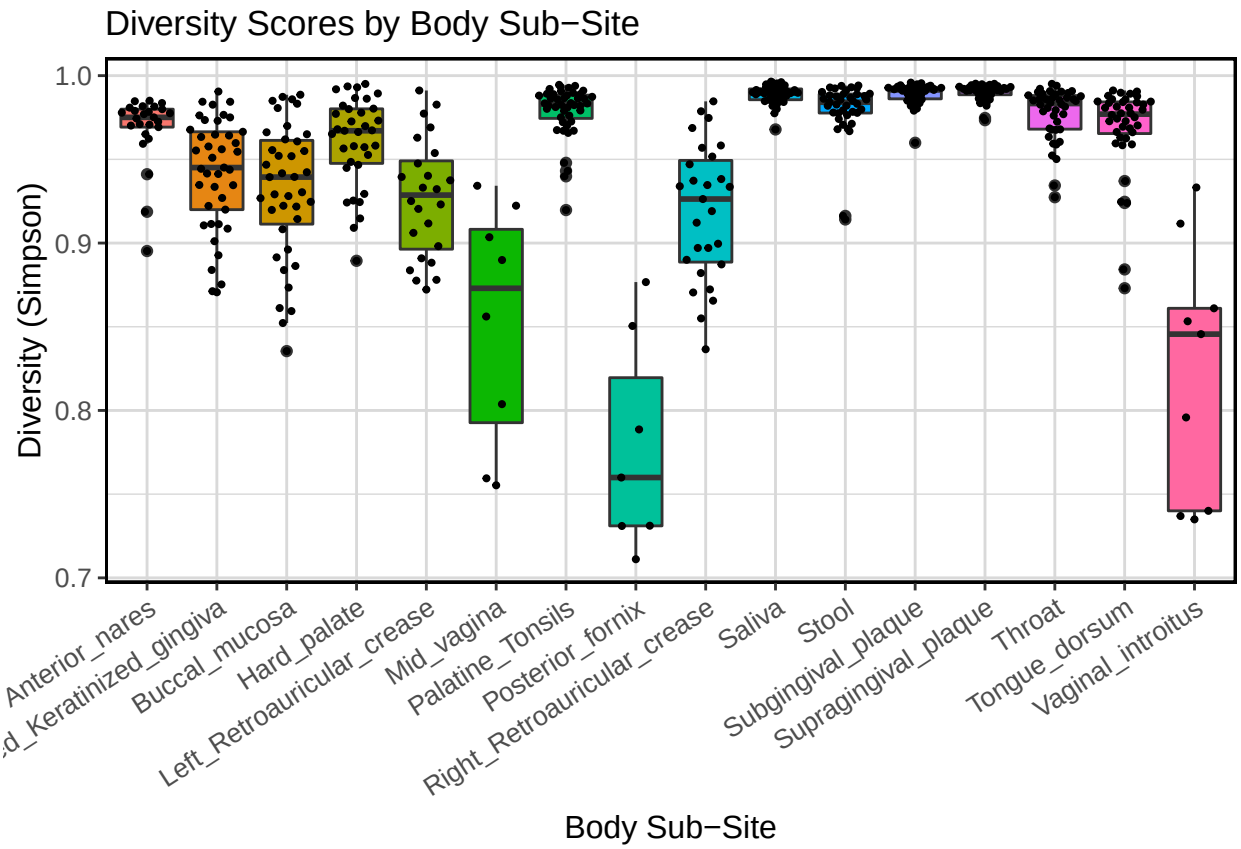
```
ps.list[["RAR"]][["Plots"]] <- list()
ps.list[["RAR"]][["Plots"]] <- lapply(names(methods.alpha), function(alpha){
  tmp.data <- na.omit(as.data.frame(caseInfStats[["dataFort.ref"]][[alpha]]) )
  ggplot(tmp.data, aes(x = HMPbodysubsite, y=Alpha.Score)) +
      geom_boxplot(aes(fill = HMPbodysubsite)) +
      ggbeeswarm::geom_quasirandom(size = 0.75) +  # spaces the dots out nicely
      facet_grid(. ~ .) +  # (Y-axis ~ X-axis)
      theme(legend.position = "none",
            axis.text.x = element_text(angle = 33, hjust = 1, vjust=1)
            ) +
      labs(
        title = "Diversity Scores by Body Sub-Site",
        # caption = "",
        y = paste0("Diversity (", alpha,")"),
        x = "Body Sub-Site"
      )
})
```

```
names(ps.list[["RAR"]][["Plots"]]) <- names(methods.alpha)

ps.list[["RAR"]][["Plots"]]
```

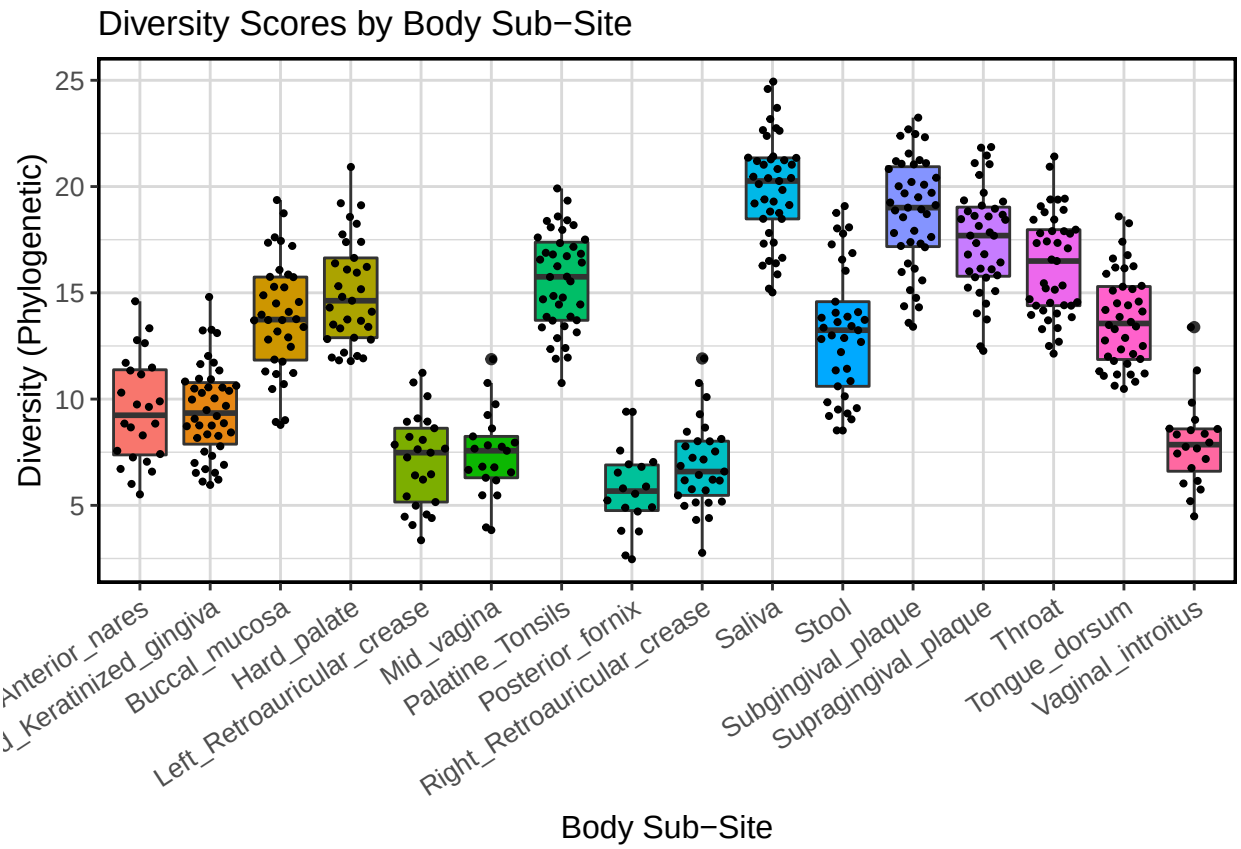
```
## $Shannon
```




```
##
## $Simpson
```




```
##
## $Phylogenetic
```


Diversity Scores by Body Sub−Site
Diversity (Phylogenetic)
25
20
15
10
5
Anterior_nares
J_Keratinized_gingiva
Buccal_mucosa
Hard_palate
Left_Retroauricular_crease
Mid_vagina
Palatine_Tonsils
Posterior_fornix
Right_Retroauricular_crease
Saliva
Stool
Subgingival_plaque
Supragingival_plaque
Throat
Tongue_dorsum
Vaginal_introitus
Body Sub−Site

```
##
## $Richness
```


Diversity Scores by Body Sub−Site
Diversity (Richness)
500
400
300
200
100
Anterior_nares
ed_Keratinized_gingiva
Buccal_mucosa
Hard_palate
Left_Retroauricular_crease
Mid_vagina
Palatine_Tonsils
Posterior_fornix
Right_Retroauricular_crease
Saliva
Stool
Subgingival_plaque
Supragingival_plaque
Throat
Tongue_dorsum
Vaginal_introitus
Body Sub−Site