

Microbial Ecology Analysis Report

Powered by MicrobiomeStat (Ver 1.1.1)

2023-09-07

Contents

| | |
|---|----|
| 1. Data summary and preparation | 2 |
| 1.1 Parameter setting | 2 |
| 1.2 Summary statistics and data overview | 4 |
| 2. Alpha diversity analysis | 4 |
| 2.1 Data visualization | 4 |
| 2.1.1 Alpha diversity boxplot | 5 |
| 2.1.2 Alpha diversity spaghettiplot | 6 |
| 2.2 Trend test | 7 |
| 2.2.1 Shannon index | 7 |
| 2.2.2 Simpson index | 8 |
| 2.3 Volatility test | 9 |
| 2.3.1 Shannon index | 9 |
| 2.3.2 Simpson index | 9 |
| 3. Beta diversity analysis | 10 |
| 3.1 Data visualization | 10 |
| 3.1.1 Beta diversity ordinationplot | 10 |
| 3.1.2 Beta diversity principal coordinate spaghettiplot | 11 |
| 3.1.3 Beta diversity change spaghettiplot | 13 |
| 3.2 Trend test | 14 |
| 3.2.1 BC distance | 14 |
| 3.2.2 Jaccard distance | 15 |
| 3.3 Principal coordinate trend test | 16 |
| 3.3.1 BC distance | 16 |
| 3.3.2 Jaccard distance | 16 |
| 3.4 Volatility test | 18 |
| 3.4.1 BC distance | 19 |
| 3.4.2 Jaccard distance | 19 |
| 3.5 Principal coordinate volatility test | 19 |
| 3.5.1 BC distance: | 19 |
| 3.5.2 Jaccard distance: | 20 |
| 4. Feature-level Analysis | 21 |
| 4.1 Data visualization(overall) | 21 |
| 4.1.1 Feature areaplot | 21 |
| 4.1.2 Feature heatmap | 22 |
| 4.1.3 Feature change heatmap | 23 |
| 4.1.4 Feature barplot | 24 |
| 4.2 Trend test | 25 |
| 4.3 Volatility test | 32 |
| 4.4 Data visualization(significant features) | 36 |
| 4.4.1 Significant features boxplot | 37 |

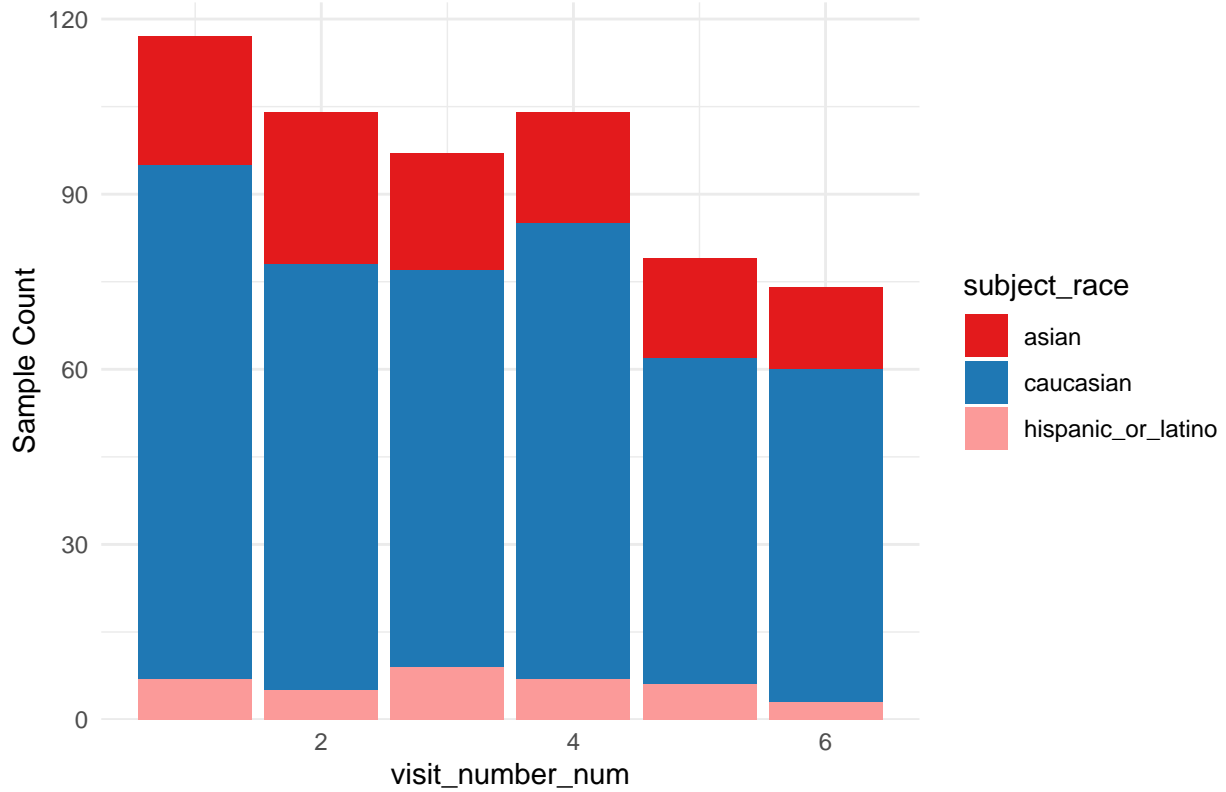
1. Data summary and preparation

1.1 Parameter setting

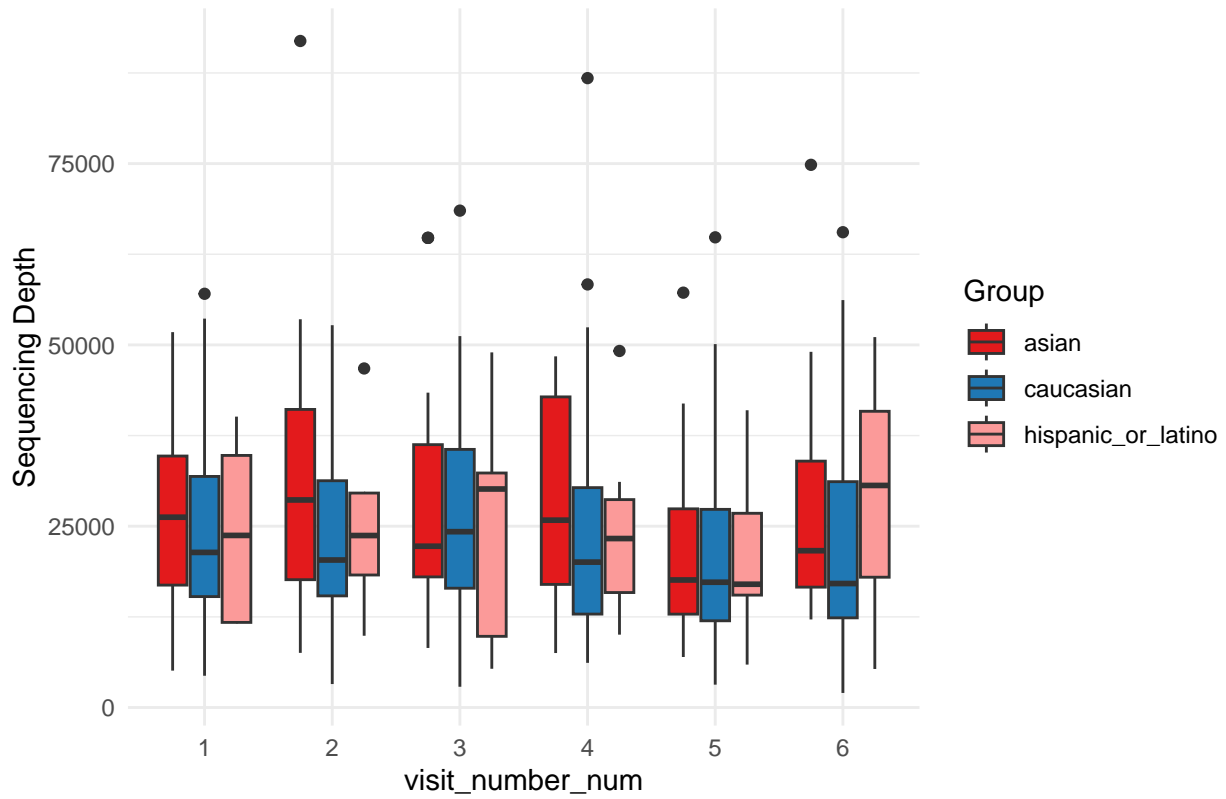
| Parameter | Value |
|---------------------|------------------|
| data.obj | subset_T2D.obj |
| feature.dat.type | count |
| group.var | subject_race |
| test.adj.vars | NULL |
| vis.adj.vars | NULL |
| strata.var | subject_gender |
| subject.var | subject_id |
| time.var | visit_number_num |
| t0.level | 1 |
| ts.levels | 2, 3, 4, 5, 6 |
| alpha.obj | NULL |
| alpha.name | shannon, simpson |
| depth | NULL |
| dist.obj | NULL |
| dist.name | BC, Jaccard |
| pc.obj | NULL |
| prev.filter | 0.1 |
| abund.filter | 1e-04 |
| feature.change.func | relative change |
| feature.number | 15 |
| feature.level | Family |
| feature.mt.method | none |
| feature.sig.level | 0.1 |
| transform | sqrt |
| base.size | 12 |
| theme.choice | bw |
| custom.theme | NULL |
| palette | NULL |
| pdf | TRUE |
| file.ann | NULL |
| pdf.wid | 11 |
| pdf.hei | 8.5 |

```
mStat_results <- mStat_summarize_data_obj(data.obj = data.obj,
                                           time.var = time.var,
                                           group.var = group.var,
                                           palette = palette)
```

Histogram of Sample Counts over Time



Boxplot of Sequencing Depth over Time



1.2 Summary statistics and data overview

| Category | Variable | Value |
|-------------------------------|--|----------|
| Basic Statistics | Number of samples | 575 |
| Basic Statistics | Number of features | 9533 |
| Basic Statistics | Min. reads per sample | 2007 |
| Basic Statistics | Max. reads per sample | 91908 |
| Basic Statistics | Total reads dplyr::across all samples | 14138179 |
| Basic Statistics | Average reads per sample | 1483.078 |
| Basic Statistics | Median reads per sample | 21062 |
| Basic Statistics | Proportion of zero counts | 0.963 |
| Basic Statistics | Count of features that only appear once | 1505 |
| Metadata | Number of metadata variables | 14 |
| Feature Annotations | Proportion of missing annotations in Kingdom | 0 |
| Feature Annotations | Proportion of missing annotations in Phylum | 0 |
| Feature Annotations | Proportion of missing annotations in Class | 0.002 |
| Feature Annotations | Proportion of missing annotations in Order | 0.012 |
| Feature Annotations | Proportion of missing annotations in Family | 0.128 |
| Feature Annotations | Proportion of missing annotations in Genus | 0.484 |
| Feature Annotations | Proportion of missing annotations in Species | 0.887 |
| Phylogenetic Tree | Exists in the dataset | No |
| Time-Series Information | Earliest sample time-point | 1 |
| Time-Series Information | Latest sample time-point | 6 |
| Distribution of sample counts | Sample Count at Time-point: 1 | 117 |
| Distribution of sample counts | Sample Count at Time-point: 2 | 104 |
| Distribution of sample counts | Sample Count at Time-point: 3 | 97 |
| Distribution of sample counts | Sample Count at Time-point: 4 | 104 |
| Distribution of sample counts | Sample Count at Time-point: 5 | 79 |
| Distribution of sample counts | Sample Count at Time-point: 6 | 74 |

2. Alpha diversity analysis

2.1 Data visualization

```
cat('
')
```

```
alpha_boxplot_results <- generate_alpha_boxplot_long(data.obj = data.obj,
  alpha.obj = alpha.obj,
  alpha.name = alpha.name,
  depth = depth,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  adj.vars = vis.adj.vars,
  base.size = base.size,
  theme.choice = theme.choice,
```

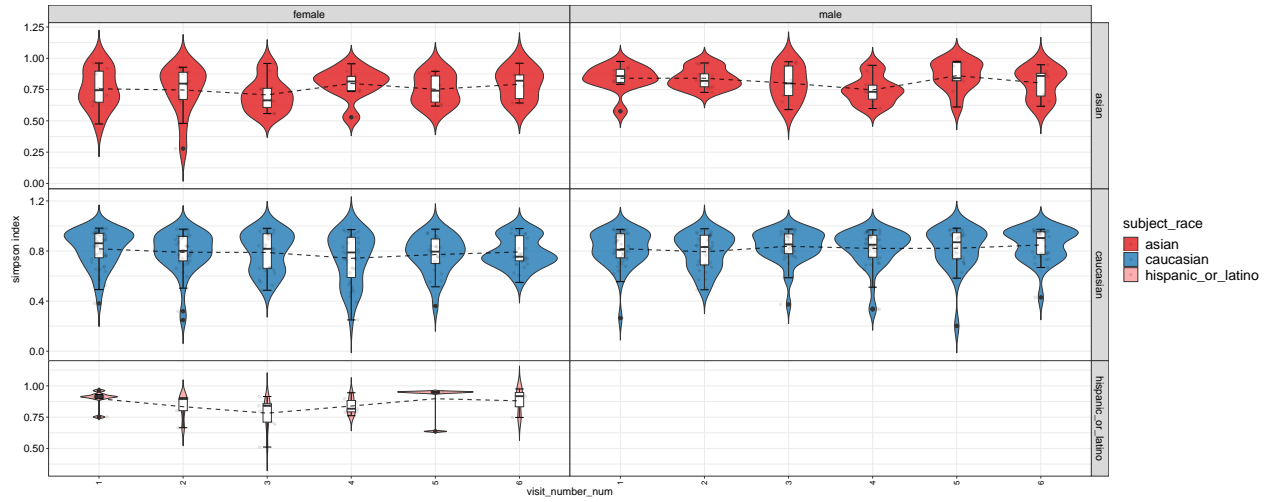
```

custom.theme = custom.theme,
palette = palette,
pdf = pdf,
file.ann = file.ann,
pdf.wid = pdf.wid,
pdf.hei = pdf.hei)

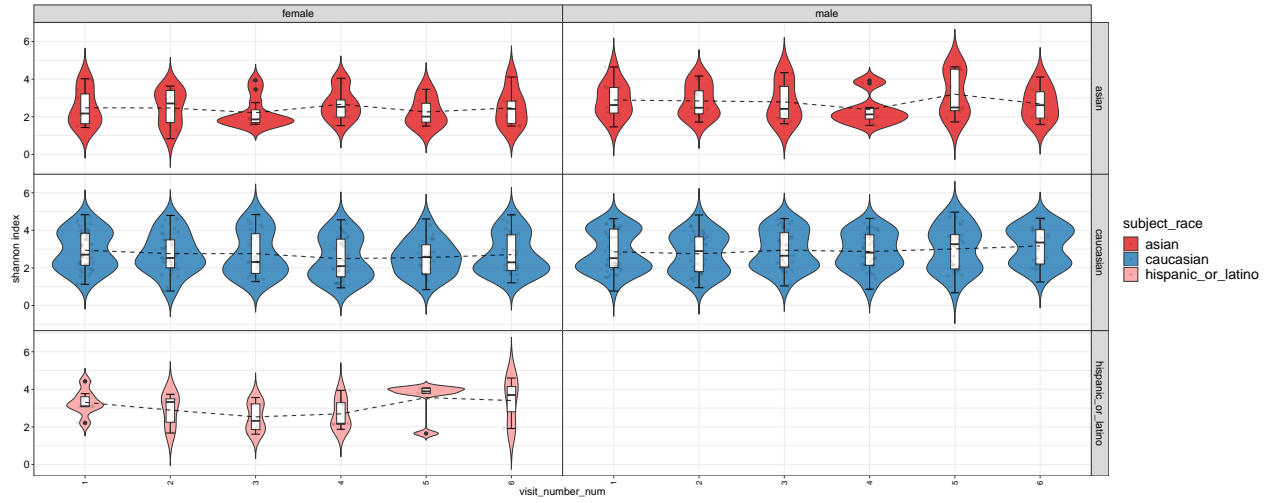
```

alpha_boxplot_results

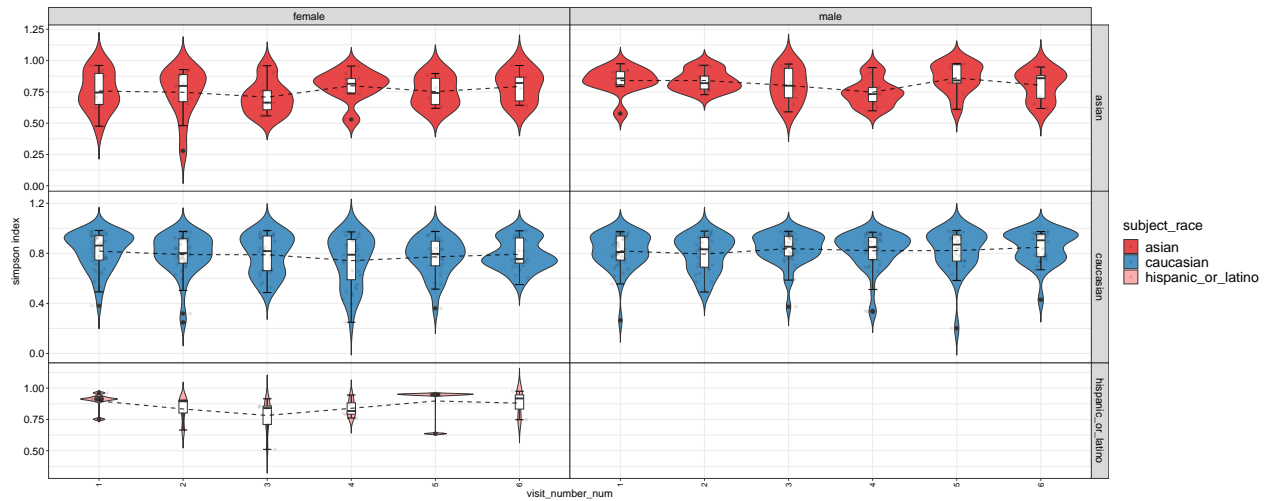
2.1.1 Alpha diversity boxplot



\$shannon



\$simpson

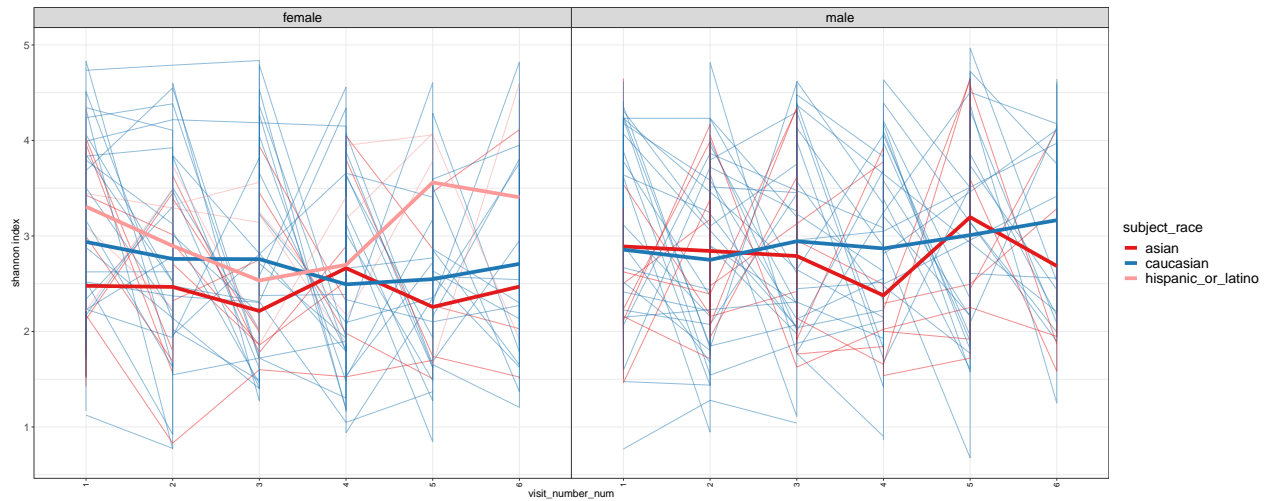


```
cat('
')
```

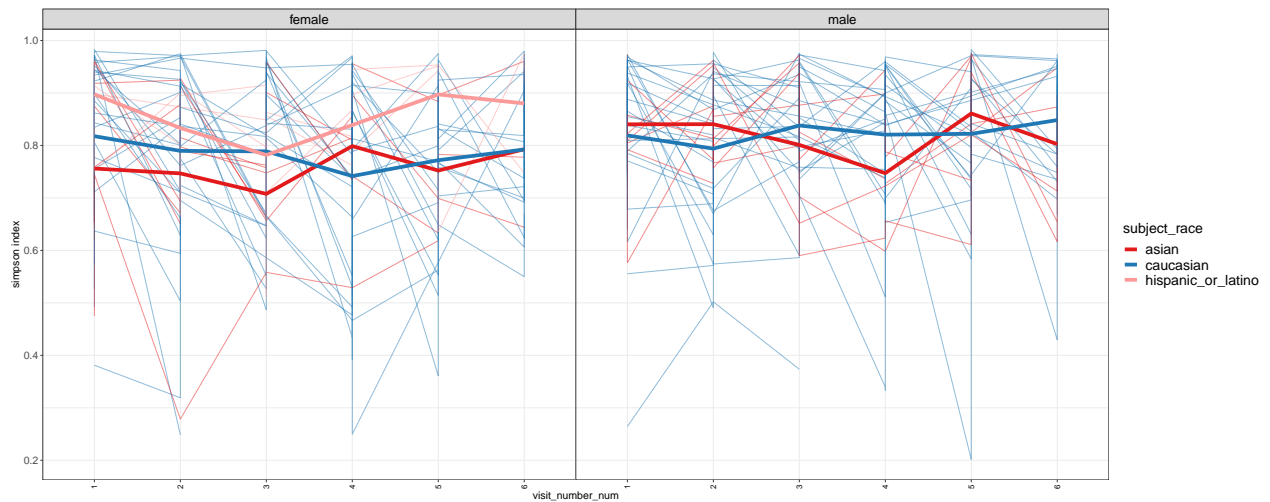
```
alpha_spaghettoplot_results <- generate_alpha_spaghettoplot_long(
  data.obj = data.obj,
  alpha.obj = alpha.obj,
  alpha.name = alpha.name,
  depth = depth,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  adj.vars = vis.adj.vars,
  base.size = base.size,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei)

alpha_spaghettoplot_results
```

2.1.2 Alpha diversity spaghettoplot \$shannon



Shannon



2.2 Trend test

```
alpha_trend_test_results <- generate_alpha_trend_test_long(
  data.obj = data.obj,
  alpha.obj = alpha.obj,
  alpha.name = alpha.name,
  depth = depth,
  time.var = time.var,
  subject.var = subject.var,
  group.var = group.var,
  adj.vars = test.adj.vars)
```

In this analysis, we utilized a linear mixed effects model with both random intercept and random slope to investigate potential interactions. Specifically, we tested the interaction between the variables `subject_race` and `visit_number_num`.

2.2.1 Shannon index Based on the linear mixed effects model, no significant interaction was detected between `visit_number_num` and the level `caucasian` of the variable `subject_race`, with a p-value of 0.835.

Based on the linear mixed effects model, no significant interaction was detected between `visit_number_num`

and the level hispanicorlatino of the variable subject_race, with a p-value of 0.650.

2.2.2 Simpson index Based on the linear mixed effects model, no significant interaction was detected between visit_number_num and the level caucasian of the variable subject_race, with a p-value of 0.717.

Based on the linear mixed effects model, no significant interaction was detected between visit_number_num and the level hispanicorlatino of the variable subject_race, with a p-value of 0.847.

- Detailed results for alpha diversity trend test:
 - shannon:

Table 3: Table continues below

| Term | Estimate | Std.Error |
|---|-----------|-----------|
| (Intercept) | 2.635 | 0.211 |
| subject_racecaucasian | 0.2271 | 0.2385 |
| subject_racehispanic_or_latino | 0.204 | 0.4553 |
| visit_number_num | -0.005673 | 0.05653 |
| subject_racecaucasian:visit_number_num | -0.01329 | 0.06376 |
| subject_racehispanic_or_latino:visit_number_num | 0.0545 | 0.1202 |
| subject_race:visit_number_num | NA | NA |

| Statistic | P.Value |
|-----------|-----------|
| 12.49 | 1.548e-26 |
| 0.9521 | 0.3423 |
| 0.4481 | 0.6548 |
| -0.1004 | 0.9201 |
| -0.2085 | 0.835 |
| 0.4536 | 0.6504 |
| 0.1986 | 0.82 |

- simpson:

Table 5: Table continues below

| Term | Estimate | Std.Error |
|---|-----------|-----------|
| (Intercept) | 0.7851 | 0.03328 |
| subject_racecaucasian | 0.02822 | 0.03759 |
| subject_racehispanic_or_latino | 0.05775 | 0.07364 |
| visit_number_num | -0.000609 | 0.007633 |
| subject_racecaucasian:visit_number_num | -0.003123 | 0.008622 |
| subject_racehispanic_or_latino:visit_number_num | 0.003134 | 0.01626 |
| subject_race:visit_number_num | NA | NA |

| Statistic | P.Value |
|-----------|-----------|
| 23.59 | 1.822e-36 |
| 0.7507 | 0.4551 |
| 0.7842 | 0.4361 |
| -0.07979 | 0.9364 |

| Statistic | P.Value |
|-----------|---------|
| -0.3622 | 0.7173 |
| 0.1928 | 0.8473 |
| 0.1368 | 0.8722 |

2.3 Volatility test

```
alpha_volatility_test_results <- generate_alpha_volatility_test_long(
  data.obj = data.obj,
  alpha.obj = alpha.obj,
  alpha.name = alpha.name,
  depth = depth,
  time.var = time.var,
  subject.var = subject.var,
  group.var = group.var,
  adj.vars = test.adj.vars)
```

In this analysis, we utilized a general linear model to examine the influence of the variable `subject_race` on volatility.

The alpha diversity volatility is calculated by averaging the rate of change in alpha diversity across consecutive time points. Specifically, for each pair of adjacent time points, we compute the difference in alpha diversity, normalize it by the time difference, and then take the average over all such pairs.

2.3.1 Shannon index Based on the general linear model, the level caucasian of the variable `subject_race` did not significantly influence the alpha diversity volatility, with a p-value of 0.425.

Based on the general linear model, the level hispanic_or_latino of the variable `subject_race` did not significantly influence the alpha diversity volatility, with a p-value of 0.790.

2.3.2 Simpson index Based on the general linear model, the level caucasian of the variable `subject_race` did not significantly influence the alpha diversity volatility, with a p-value of 0.522.

Based on the general linear model, the level hispanic_or_latino of the variable `subject_race` did not significantly influence the alpha diversity volatility, with a p-value of 0.696.

- Detailed results for alpha diversity volatility test:

– **shannon:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|----------|
| (Intercept) | 1.081 | 0.1788 | 6.047 | 9.76e-08 |
| subject_racecaucasian | 0.1619 | 0.2016 | 0.8034 | 0.4249 |
| subject_racehispanic_or_latino | 0.1106 | 0.413 | 0.2679 | 0.7897 |
| subject_race | NA | NA | 0.3237 | 0.7247 |
| Residuals | NA | NA | NA | NA |

– **simpson:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|-----------|
| (Intercept) | 0.1357 | 0.03015 | 4.501 | 3.107e-05 |
| subject_racecaucasian | 0.02187 | 0.03399 | 0.6436 | 0.5223 |
| subject_racehispanic_or_latino | -0.02731 | 0.06962 | -0.3923 | 0.6962 |

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------|----------|-----------|-----------|---------|
| subject_race | NA | NA | 0.4469 | 0.6417 |
| Residuals | NA | NA | NA | NA |

3. Beta diversity analysis

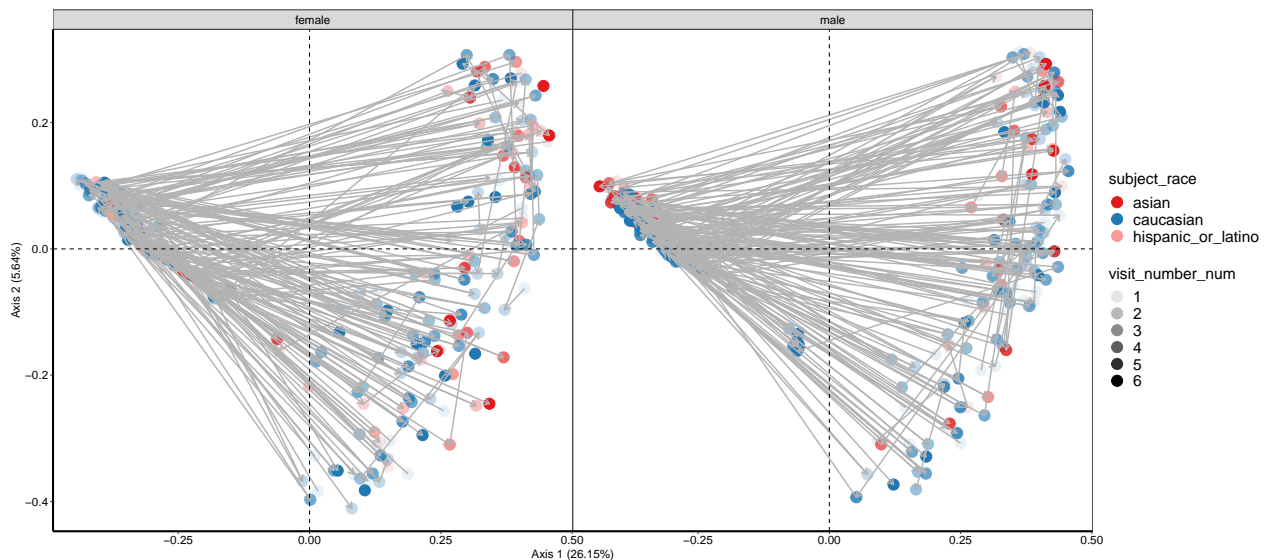
3.1 Data visualization

```
cat('
')
```

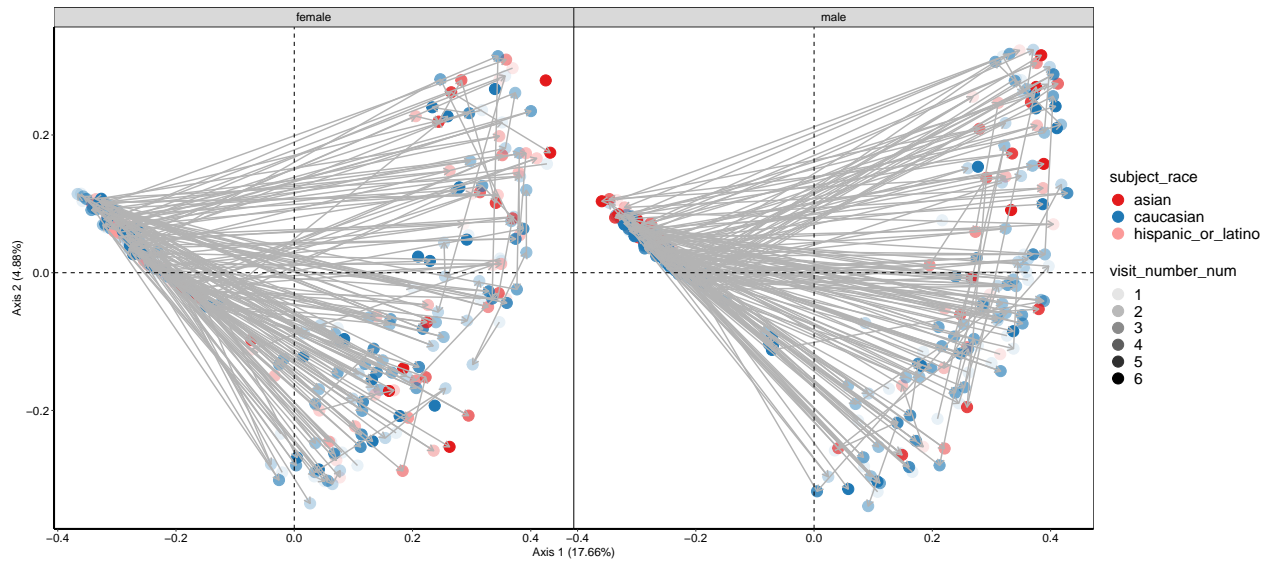
```
beta_ordination_results <- generate_beta_ordination_long(data.obj = data.obj,
  dist.obj = dist.obj,
  pc.obj = pc.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  adj.vars = vis.adj.vars,
  dist.name = dist.name,
  base.size = base.size,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei)
```

```
beta_ordination_results
```

3.1.1 Beta diversity ordinationplot \$BC



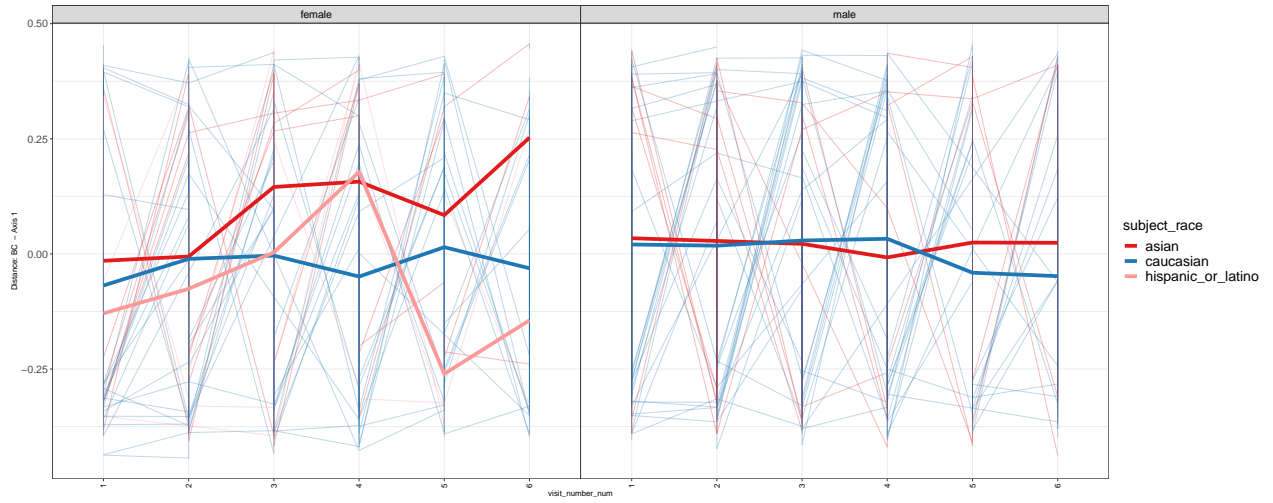
```
$Jaccard
```



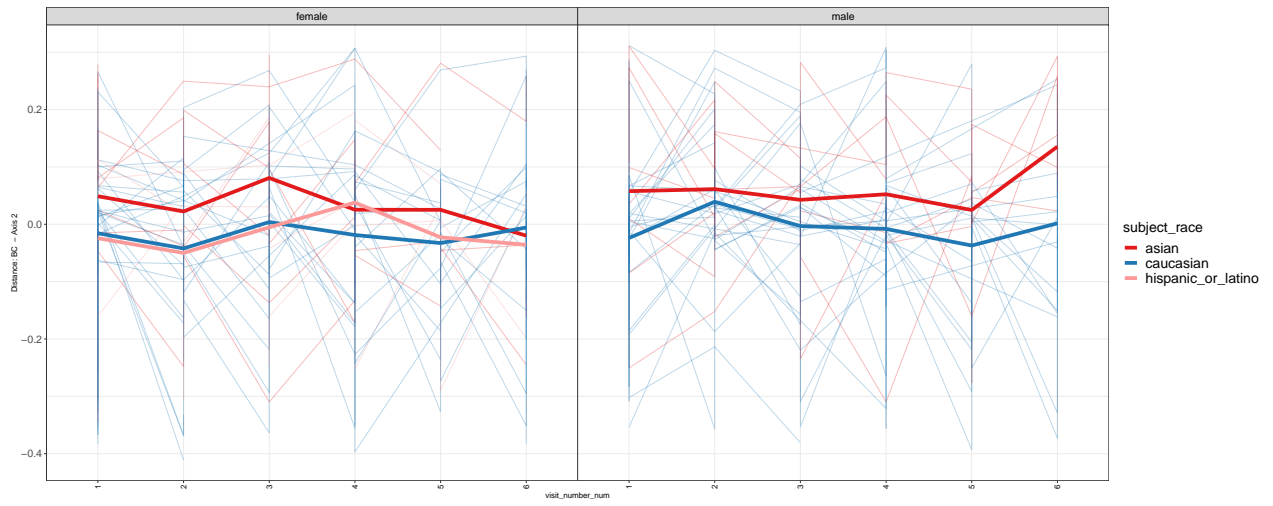
```
pc_boxplot_longitudinal_results <- generate_beta_pc_spaghettiplot_long(
  data.obj = data.obj,
  dist.obj = dist.obj,
  pc.obj = pc.obj,
  pc.ind = c(1, 2),
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  adj.vars = vis.adj.vars,
  dist.name = dist.name,
  base.size = base.size,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)
```

```
pc_boxplot_longitudinal_results
```

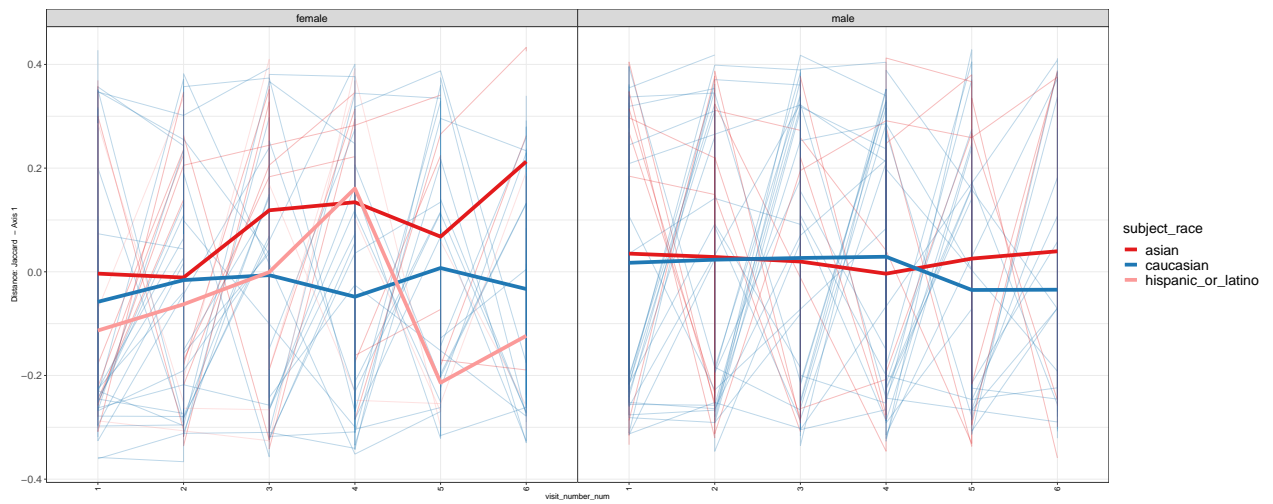
3.1.2 Beta diversity principal coordinate spaghettiplot \$BC BCPC1



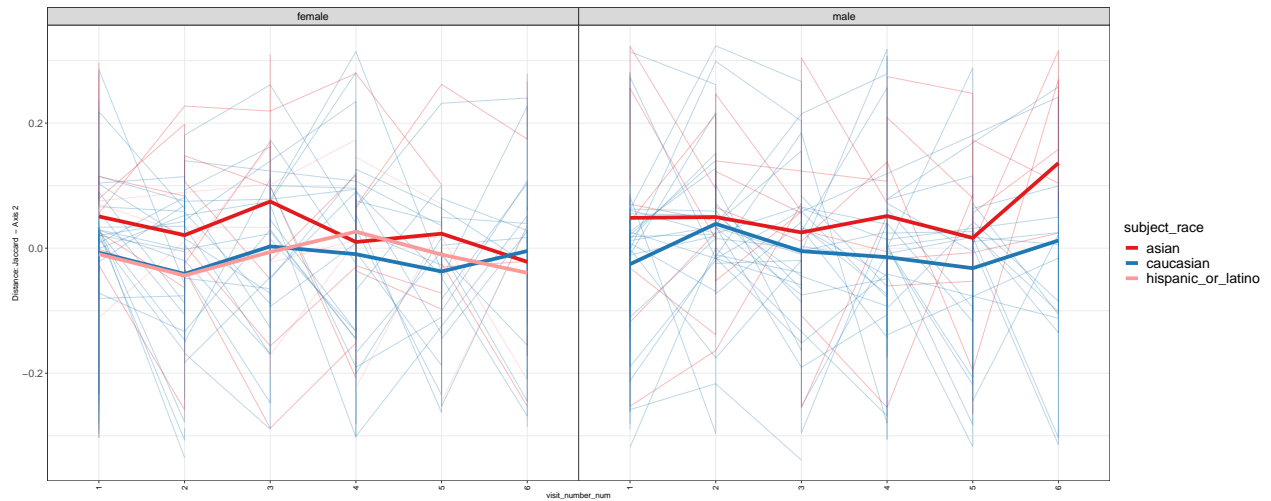
BCPC2



$\$$ Jaccard *JaccardPC1*



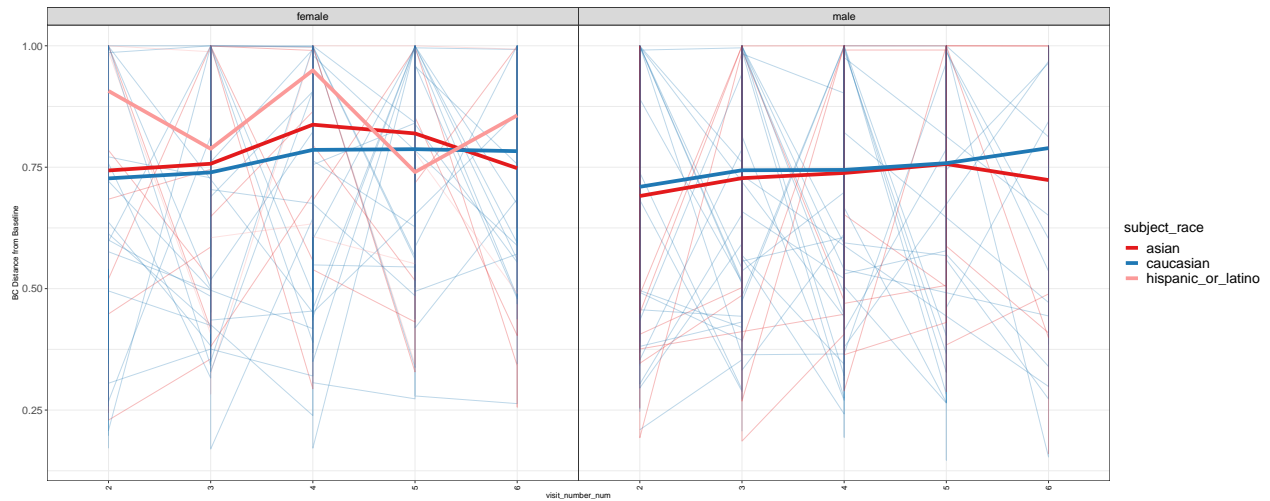
JaccardPC2



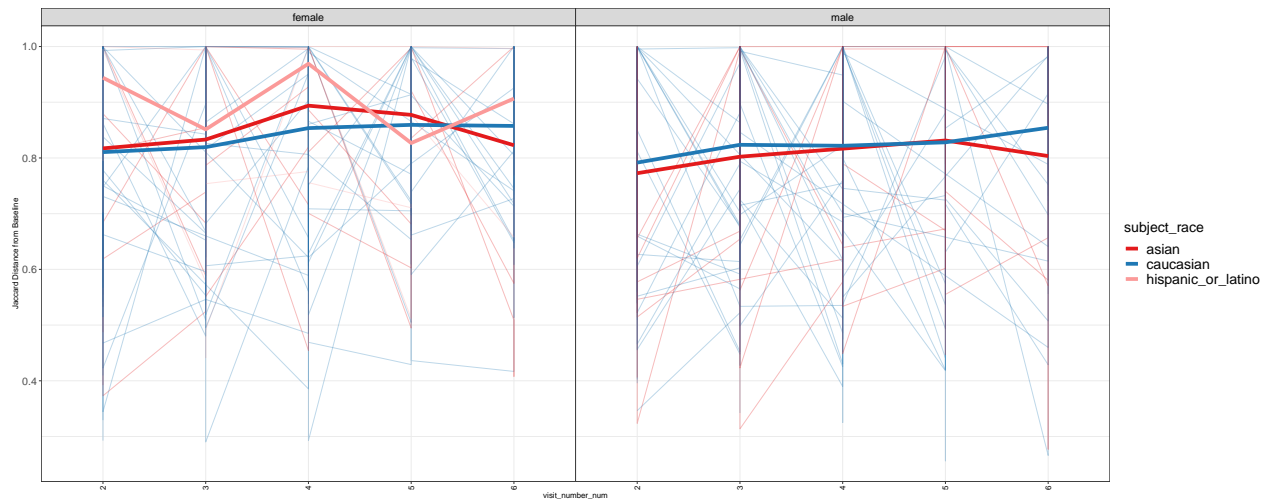
```
spaghettiplot_longitudinal_results <- generate_beta_change_spaghettiplot_long(
  data.obj = data.obj,
  dist.obj = dist.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  adj.vars = vis.adj.vars,
  dist.name = dist.name,
  base.size = base.size,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)
```

spaghettiplot_longitudinal_results

3.1.3 Beta diversity change spaghettiplot \$BC



\$Jaccard



3.2 Trend test

```
beta_trend_test_longitudinal_results <- generate_beta_trend_test_long(
  data.obj = data.obj,
  dist.obj = dist.obj,
  subject.var = subject.var,
  time.var = time.var,
  group.var = group.var,
  adj.vars = test.adj.vars,
  dist.name = dist.name)
```

In this analysis, we utilized a linear mixed effects model to investigate potential interactions. Specifically, we tested the interaction between the variables `subject_race` and `visit_number_num`, while considering the distances to the first/reference time point.

3.2.1 BC distance Based on the linear mixed effects model, no significant interaction was detected between `visit_number_num` and the level `caucasian` of the variable `subject_race`, in terms of the distances to the first/reference time point, with a p-value of 0.647.

Based on the linear mixed effects model, no significant interaction was detected between `visit_number_num`

and the level hispanicorlatino of the variable subject_race, in terms of the distances to the first/reference time point, with a p-value of 0.445.

3.2.2 Jaccard distance Based on the linear mixed effects model, no significant interaction was detected between visit_number_num and the level caucasian of the variable subject_race, in terms of the distances to the first/reference time point, with a p-value of 0.640.

Based on the linear mixed effects model, no significant interaction was detected between visit_number_num and the level hispanicorlatino of the variable subject_race, in terms of the distances to the first/reference time point, with a p-value of 0.484.

- Detailed Results for Beta Diversity Trend Test:
 - **BC:**

Table 9: Table continues below

| Term | Estimate | Std.Error |
|---|----------|-----------|
| (Intercept) | 0.7168 | 0.05948 |
| subject_racecaucasian | -0.02828 | 0.06817 |
| subject_racehispanic_or_latino | 0.1778 | 0.1193 |
| visit_number_num | 0.009211 | 0.01475 |
| subject_racecaucasian:visit_number_num | 0.007723 | 0.01684 |
| subject_racehispanic_or_latino:visit_number_num | -0.02279 | 0.02984 |
| subject_race:visit_number_num | NA | NA |

| Statistic | P.Value |
|-----------|-----------|
| 12.05 | 9.375e-30 |
| -0.4149 | 0.6784 |
| 1.491 | 0.1369 |
| 0.6246 | 0.5324 |
| 0.4587 | 0.6466 |
| -0.7638 | 0.4452 |
| 0.6751 | 0.5094 |

- **Jaccard:**

Table 11: Table continues below

| Term | Estimate | Std.Error |
|---|----------|-----------|
| (Intercept) | 0.796 | 0.04424 |
| subject_racecaucasian | -0.01945 | 0.05072 |
| subject_racehispanic_or_latino | 0.1274 | 0.08892 |
| visit_number_num | 0.007687 | 0.0109 |
| subject_racecaucasian:visit_number_num | 0.005825 | 0.01245 |
| subject_racehispanic_or_latino:visit_number_num | -0.01542 | 0.02201 |
| subject_race:visit_number_num | NA | NA |

| Statistic | P.Value |
|-----------|-----------|
| 17.99 | 1.533e-52 |

| Statistic | P.Value |
|-----------|---------|
| -0.3835 | 0.7015 |
| 1.433 | 0.1532 |
| 0.7051 | 0.4809 |
| 0.468 | 0.6399 |
| -0.7007 | 0.4837 |
| 0.6125 | 0.5423 |

3.3 Principal coordinate trend test

```
beta_pc_trend_test_longitudinal_results <- generate_beta_pc_trend_test_long(
  data.obj = data.obj,
  dist.obj = dist.obj,
  pc.obj = pc.obj,
  subject.var = subject.var,
  time.var = time.var,
  group.var = group.var,
  adj.vars = test.adj.vars,
  dist.name = dist.name)
```

In this analysis, we utilized a linear mixed effects model to investigate potential interactions. Specifically, we tested the interaction between the variables `subject_race` and `visit_number_num`, while considering individual principal components.

3.3.1 BC distance

3.3.1.1 Results for principal component: PC1 No significant interaction detected between `visit_number_num` and the level `caucasian` of the variable `subject_race`, p-value = 0.287.

No significant interaction detected between `visit_number_num` and the level `hispanicorlatino` of the variable `subject_race`, p-value = 0.530.

3.3.1.2 Results for principal component: PC2 No significant interaction detected between `visit_number_num` and the level `caucasian` of the variable `subject_race`, p-value = 0.652.

No significant interaction detected between `visit_number_num` and the level `hispanicorlatino` of the variable `subject_race`, p-value = 0.942.

3.3.2 Jaccard distance

3.3.2.1 Results for principal component: PC1 No significant interaction detected between `visit_number_num` and the level `caucasian` of the variable `subject_race`, p-value = 0.251.

No significant interaction detected between `visit_number_num` and the level `hispanicorlatino` of the variable `subject_race`, p-value = 0.548.

3.3.2.2 Results for principal component: PC2 No significant interaction detected between `visit_number_num` and the level `caucasian` of the variable `subject_race`, p-value = 0.641.

No significant interaction detected between `visit_number_num` and the level `hispanicorlatino` of the variable `subject_race`, p-value = 0.840.

- Detailed Results for Beta Diversity Principal Coordinate Trend Test:

- BC:

* **PC1:**

Table 13: Table continues below

| Term | Estimate | Std.Error |
|---|-----------|-----------|
| (Intercept) | -0.008022 | 0.06509 |
| subject_racecaucasian | 0.007754 | 0.07353 |
| subject_racehispanic_or_latino | -0.02764 | 0.1394 |
| visit_number_num | 0.01866 | 0.01802 |
| subject_racecaucasian:visit_number_num | -0.02159 | 0.02028 |
| subject_racehispanic_or_latino:visit_number_num | -0.02434 | 0.03875 |

| Statistic | P.Value |
|-----------|---------|
| -0.1233 | 0.9019 |
| 0.1055 | 0.9161 |
| -0.1983 | 0.8429 |
| 1.036 | 0.3008 |
| -1.065 | 0.2875 |
| -0.6282 | 0.5301 |

* **PC2:**

Table 15: Table continues below

| Term | Estimate | Std.Error |
|---|-----------|-----------|
| (Intercept) | 0.03478 | 0.03508 |
| subject_racecaucasian | -0.04282 | 0.03958 |
| subject_racehispanic_or_latino | -0.04457 | 0.07887 |
| visit_number_num | 0.002307 | 0.007029 |
| subject_racecaucasian:visit_number_num | -0.003589 | 0.00796 |
| subject_racehispanic_or_latino:visit_number_num | -0.001075 | 0.01482 |

| Statistic | P.Value |
|-----------|---------|
| 0.9915 | 0.3246 |
| -1.082 | 0.2827 |
| -0.5652 | 0.574 |
| 0.3283 | 0.7428 |
| -0.4508 | 0.6523 |
| -0.07253 | 0.9422 |

– **Jaccard:**

* **PC1:**

Table 17: Table continues below

| Term | Estimate | Std.Error |
|-------------|-----------|-----------|
| (Intercept) | -0.005599 | 0.05344 |

| Term | Estimate | Std.Error |
|---|----------|-----------|
| subject_racecaucasian | 0.005722 | 0.06037 |
| subject_racehispanic_or_latino | -0.02946 | 0.1145 |
| visit_number_num | 0.01628 | 0.0148 |
| subject_racecaucasian:visit_number_num | -0.01913 | 0.01665 |
| subject_racehispanic_or_latino:visit_number_num | -0.01913 | 0.03181 |

| Statistic | P.Value |
|-----------|---------|
| -0.1048 | 0.9166 |
| 0.09478 | 0.9245 |
| -0.2573 | 0.797 |
| 1.1 | 0.2718 |
| -1.149 | 0.251 |
| -0.6014 | 0.5478 |

* **PC2:**

Table 19: Table continues below

| Term | Estimate | Std.Error |
|---|-----------|-----------|
| (Intercept) | 0.02976 | 0.0324 |
| subject_racecaucasian | -0.03701 | 0.03656 |
| subject_racehispanic_or_latino | -0.03063 | 0.07267 |
| visit_number_num | 0.002049 | 0.006636 |
| subject_racecaucasian:visit_number_num | -0.003508 | 0.007513 |
| subject_racehispanic_or_latino:visit_number_num | -0.00282 | 0.014 |

| Statistic | P.Value |
|-----------|---------|
| 0.9187 | 0.3611 |
| -1.012 | 0.3146 |
| -0.4215 | 0.6748 |
| 0.3088 | 0.7576 |
| -0.4669 | 0.6408 |
| -0.2014 | 0.8404 |

3.4 Volatility test

```
beta_volatility_test_longitudinal_results <- generate_beta_volatility_test_long(
  data.obj = data.obj,
  dist.obj = dist.obj,
  subject.var = subject.var,
  time.var = time.var,
  group.var = group.var,
  adj.vars = test.adj.vars,
  dist.name = dist.name)
```

In this analysis, we utilized a general linear model to examine the influence of the variable `subject_race` on beta diversity volatility.

The beta diversity volatility is calculated by averaging the rate of change in beta diversity across consecutive time points. Specifically, for each pair of adjacent time points, we compute the difference in beta diversity, normalize it by the time difference, and then take the average over all such pairs.

3.4.1 BC distance Based on the general linear model, the level caucasian of the variable `subject_race` did not significantly influence the beta diversity volatility, with a p-value of 0.891.

Based on the general linear model, the level hispanic_or_latino of the variable `subject_race` did not significantly influence the beta diversity volatility, with a p-value of 0.233.

3.4.2 Jaccard distance Based on the general linear model, the level caucasian of the variable `subject_race` did not significantly influence the beta diversity volatility, with a p-value of 0.814.

Based on the general linear model, the level hispanic_or_latino of the variable `subject_race` did not significantly influence the beta diversity volatility, with a p-value of 0.203.

- **BC:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|-----------|
| (Intercept) | 0.3403 | 0.06433 | 5.289 | 1.817e-06 |
| subject_racecaucasian | -0.01005 | 0.07268 | -0.1382 | 0.8905 |
| subject_racehispanic_or_latino | -0.1791 | 0.1486 | -1.205 | 0.2328 |
| subject_race | NA | NA | 0.7878 | 0.4595 |
| Residuals | NA | NA | NA | NA |

- **Jaccard:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|----------|
| (Intercept) | 0.3955 | 0.07006 | 5.644 | 4.78e-07 |
| subject_racecaucasian | -0.01874 | 0.07916 | -0.2367 | 0.8137 |
| subject_racehispanic_or_latino | -0.2083 | 0.1618 | -1.287 | 0.203 |
| subject_race | NA | NA | 0.8671 | 0.4254 |
| Residuals | NA | NA | NA | NA |

3.5 Principal coordinate volatility test

```
beta_pc_volatility_test_longitudinal_results <- generate_beta_pc_volatility_test_long(
  data.obj = data.obj,
  dist.obj = dist.obj,
  pc.obj = pc.obj,
  subject.var = subject.var,
  time.var = time.var,
  group.var = group.var,
  adj.vars = test.adj.vars,
  dist.name = dist.name)
```

In this analysis, we utilized a general linear model to examine the influence of the variable `subject_race` on beta diversity PC volatility. The beta diversity PC volatility is calculated by averaging the rate of change in principal components of beta diversity across consecutive time points. Specifically, for each pair of adjacent time points, we compute the difference in principal components, normalize it by the time difference, and then take the average over all such pairs.

3.5.1 BC distance:

Results for PC1 PC: Based on the general linear model, the level caucasian of the variable subject_race did not significantly influence the beta diversity PC volatility, with a p-value of 0.256.

Based on the general linear model, the level hispanic_or_latino of the variable subject_race did not significantly influence the beta diversity PC volatility, with a p-value of 0.701.

Results for PC2 PC: Based on the general linear model, the level caucasian of the variable subject_race did not significantly influence the beta diversity PC volatility, with a p-value of 0.668.

Based on the general linear model, the level hispanic_or_latino of the variable subject_race did not significantly influence the beta diversity PC volatility, with a p-value of 0.896.

3.5.2 Jaccard distance:

Results for PC1 PC: Based on the general linear model, the level caucasian of the variable subject_race did not significantly influence the beta diversity PC volatility, with a p-value of 0.159.

Based on the general linear model, the level hispanic_or_latino of the variable subject_race did not significantly influence the beta diversity PC volatility, with a p-value of 0.712.

Results for PC2 PC: Based on the general linear model, the level caucasian of the variable subject_race did not significantly influence the beta diversity PC volatility, with a p-value of 0.774.

Based on the general linear model, the level hispanic_or_latino of the variable subject_race did not significantly influence the beta diversity PC volatility, with a p-value of 0.595.

- **BC:**

- **PC1:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|-----------|
| (Intercept) | 0.4429 | 0.0478 | 9.266 | 3.026e-13 |
| subject_racecaucasian | -0.06172 | 0.05389 | -1.145 | 0.2565 |
| subject_racehispanic_or_latino | 0.04262 | 0.1104 | 0.3861 | 0.7007 |
| subject_race | NA | NA | 1.056 | 0.354 |
| Residuals | NA | NA | NA | NA |

- **PC2:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|-----------|-----------|-----------|-----------|
| (Intercept) | 0.1362 | 0.02313 | 5.887 | 1.811e-07 |
| subject_racecaucasian | 0.01125 | 0.02608 | 0.4314 | 0.6677 |
| subject_racehispanic_or_latino | -0.007046 | 0.05342 | -0.1319 | 0.8955 |
| subject_race | NA | NA | 0.145 | 0.8653 |
| Residuals | NA | NA | NA | NA |

- **Jaccard:**

- **PC1:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|-----------------------|----------|-----------|-----------|-----------|
| (Intercept) | 0.3674 | 0.03874 | 9.483 | 1.303e-13 |
| subject_racecaucasian | -0.06222 | 0.04367 | -1.425 | 0.1593 |

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|---------|
| subject_racehispanic_or_latino | 0.03322 | 0.08946 | 0.3713 | 0.7117 |
| subject_race | NA | NA | 1.51 | 0.2291 |
| Residuals | NA | NA | NA | NA |

– PC2:

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|-----------|
| (Intercept) | 0.1323 | 0.02044 | 6.472 | 1.857e-08 |
| subject_racecaucasian | 0.006649 | 0.02304 | 0.2885 | 0.7739 |
| subject_racehispanic_or_latino | -0.02526 | 0.04721 | -0.535 | 0.5946 |
| subject_race | NA | NA | 0.2863 | 0.7521 |
| Residuals | NA | NA | NA | NA |

4. Feature-level Analysis

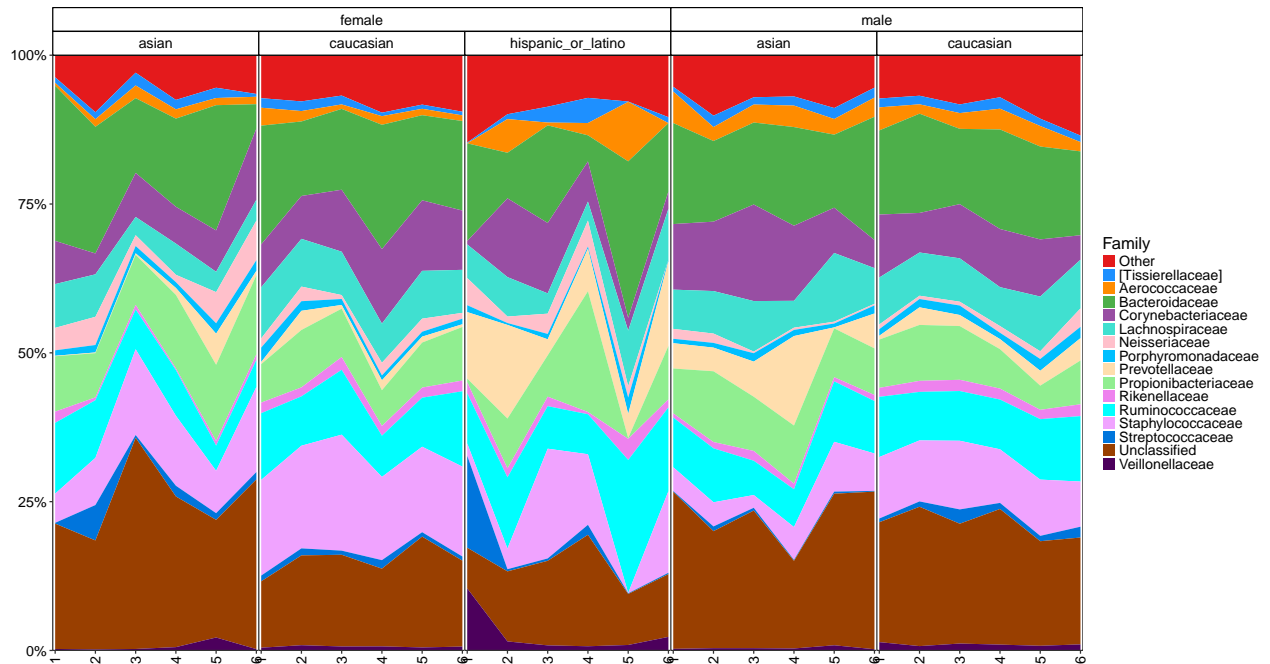
4.1 Data visualization(overall)

```
taxa_areaplot_long_results <- generate_taxa_areaplot_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  feature.level = feature.level,
  feature.dat.type = feature.dat.type,
  feature.number = feature.number,
  base.size = base.size,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)
```

4.1.1 Feature areaplot

- Average Version: This plot displays the average proportions for each time point, group, and strata.

\$Family



```

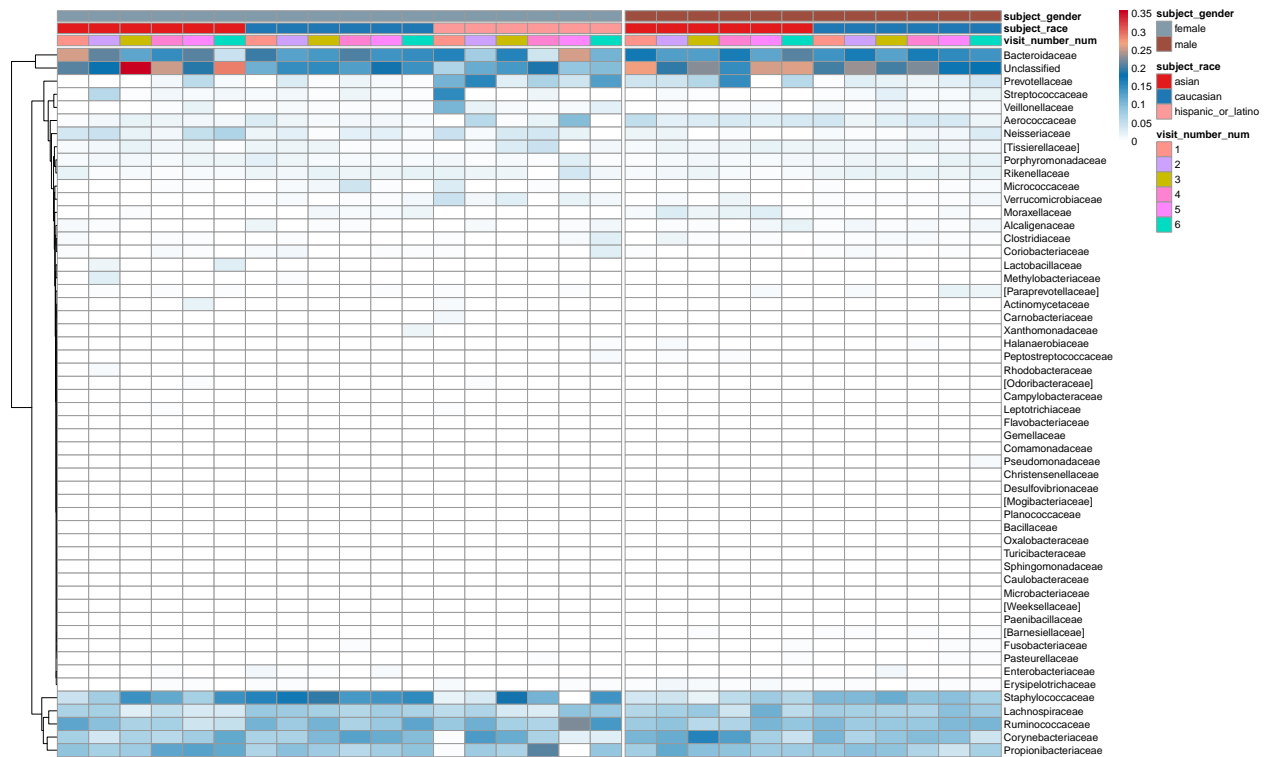
taxa_heatmap_long_results <- generate_taxa_heatmap_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  feature.level = feature.level,
  feature.dat.type = feature.dat.type,
  features.plot = NULL,
  top.k.plot = NULL,
  top.k.func = NULL,
  prev.filter = prev.filter,
  abund.filter = abund.filter,
  base.size = base.size,
  palette = palette,
  cluster.cols = NULL,
  cluster.rows = NULL,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)

```

4.1.2 Feature heatmap

- Average Version: This plot displays the average proportions for each time point, group, and strata.

\$Family



```

taxa_change_heatmap_long_results <- generate_taxa_change_heatmap_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  feature.level = feature.level,
  feature.dat.type = feature.dat.type,
  features.plot = NULL,
  top.k.plot = NULL,
  top.k.func = NULL,
  feature.change.func = feature.change.func,
  prev.filter = prev.filter,
  abund.filter = abund.filter,
  base.size = base.size,
  palette = palette,
  cluster.cols = NULL,
  cluster.rows = NULL,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)

```

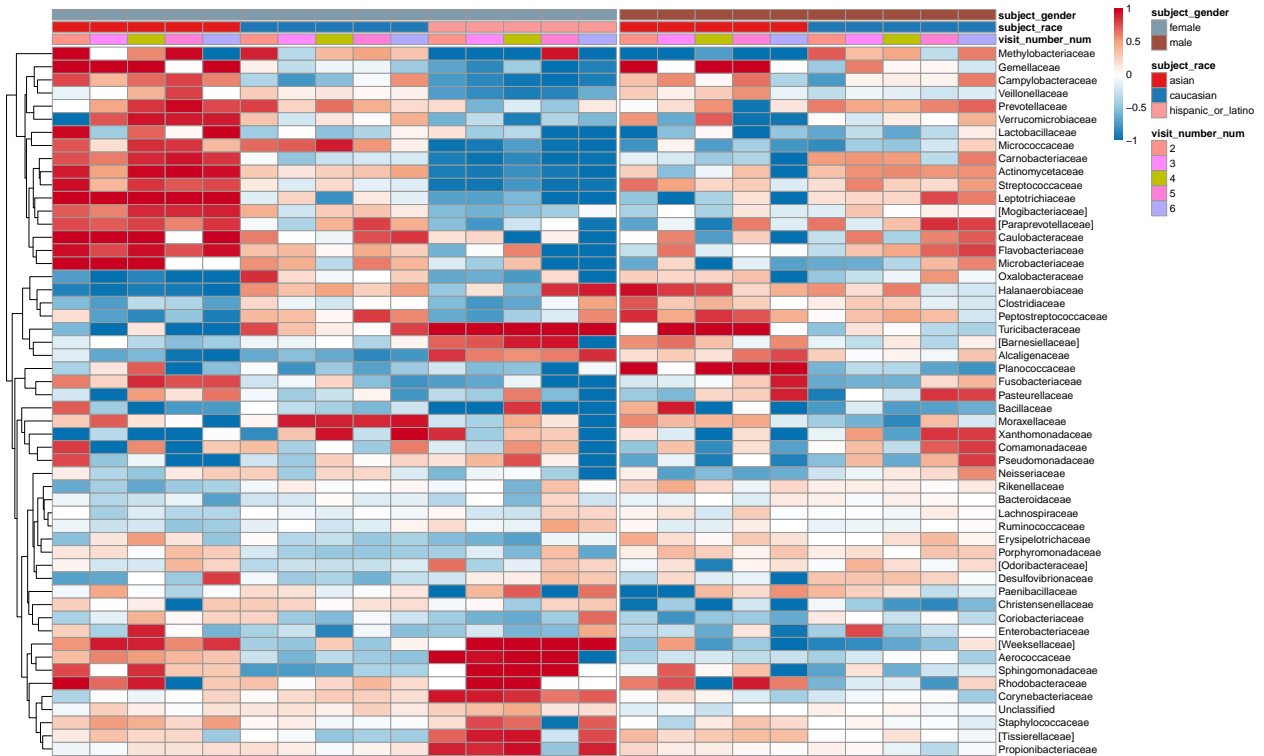
4.1.3 Feature change heatmap

- Change Calculation: Relative Change The changes from t0.level were computed as the difference

between the current value and t0.level divided by the sum of the two.

- Average Change: This plot displays the average changes for each time point, group, and strata.

\$Family

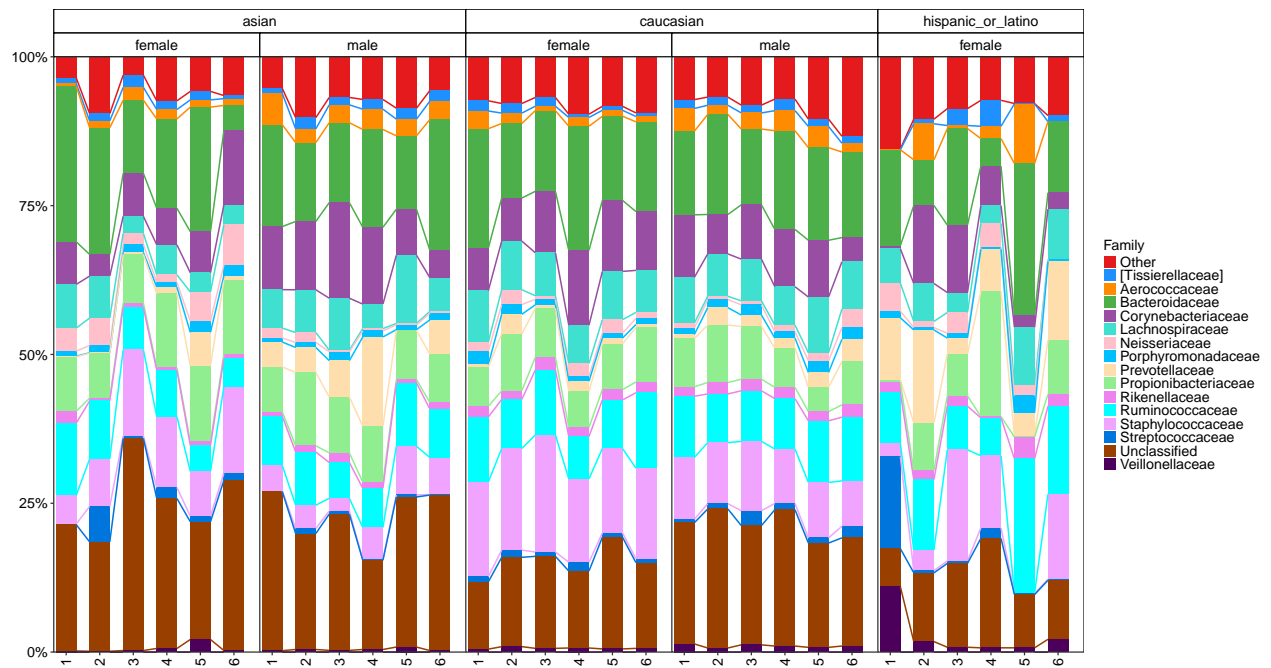


```
taxa_barplot_long_results <- generate_taxa_barplot_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  group.var = group.var,
  strata.var = strata.var,
  feature.level = feature.level,
  feature.dat.type = feature.dat.type,
  feature.number = feature.number,
  t0.level = t0.level,
  ts.levels = ts.levels,
  base.size = base.size,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)
```

4.1.4 Feature barplot

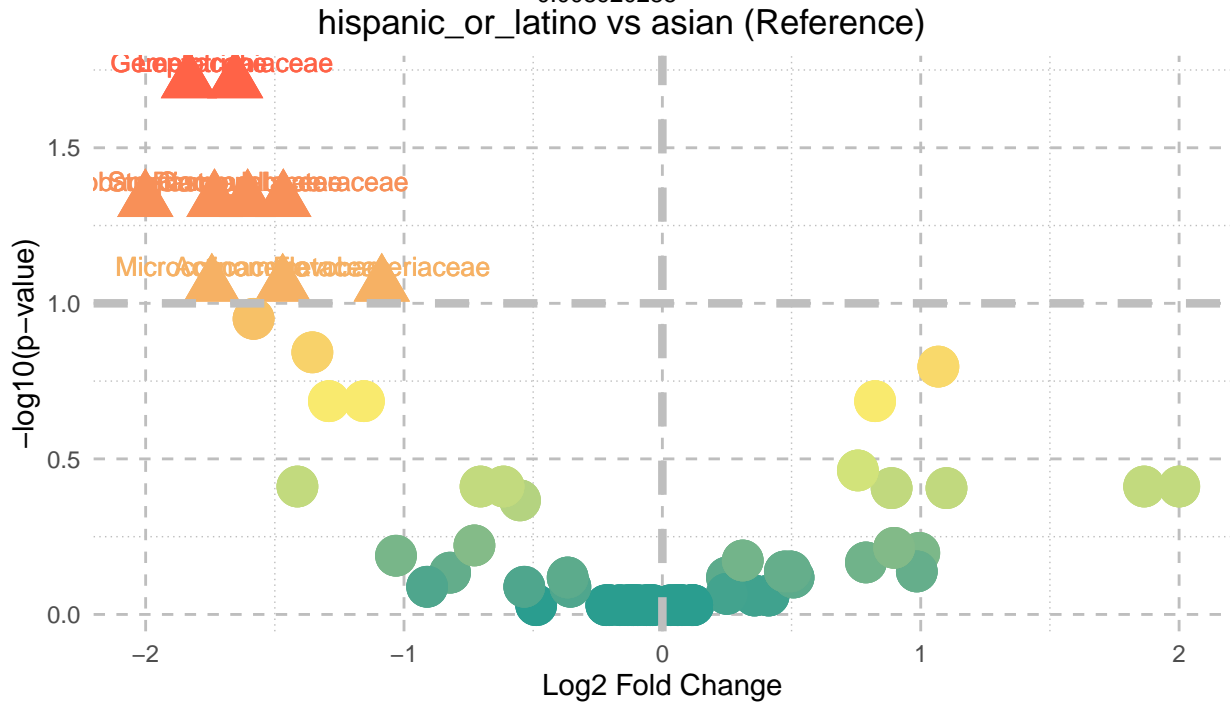
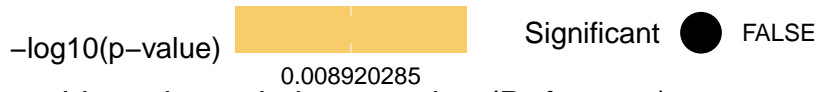
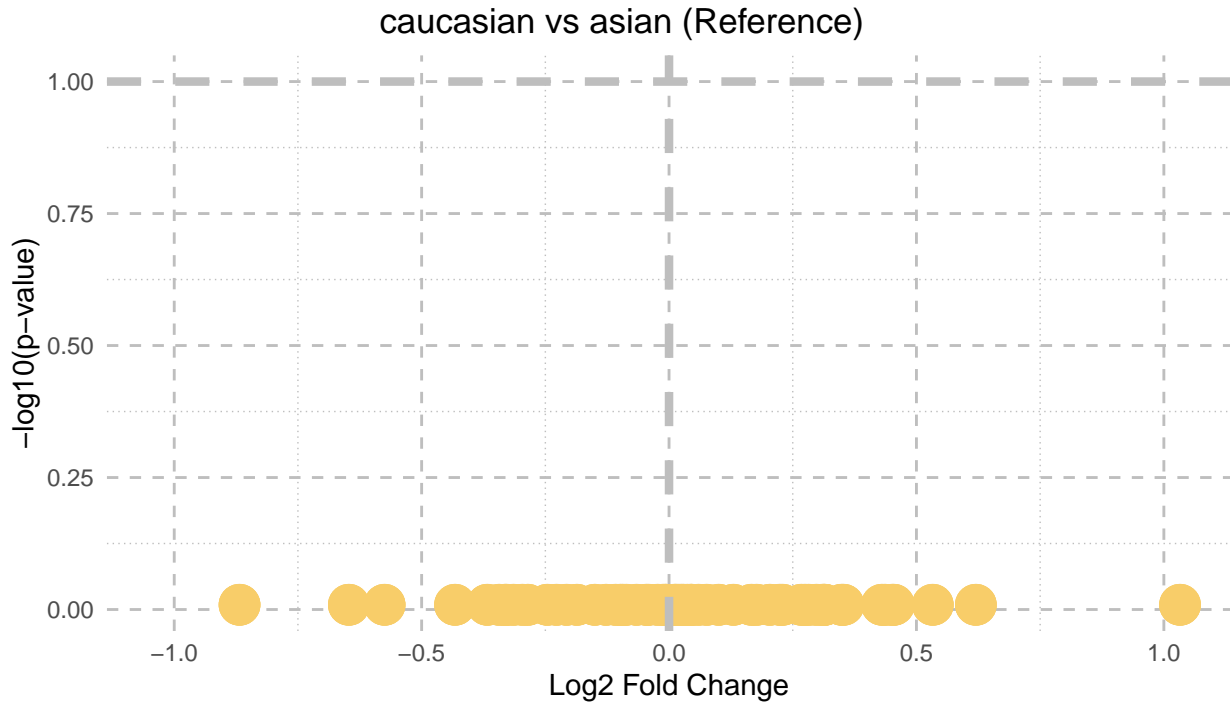
- Average Version: This plot displays the average proportions for each time point, group, and strata.

\$Family

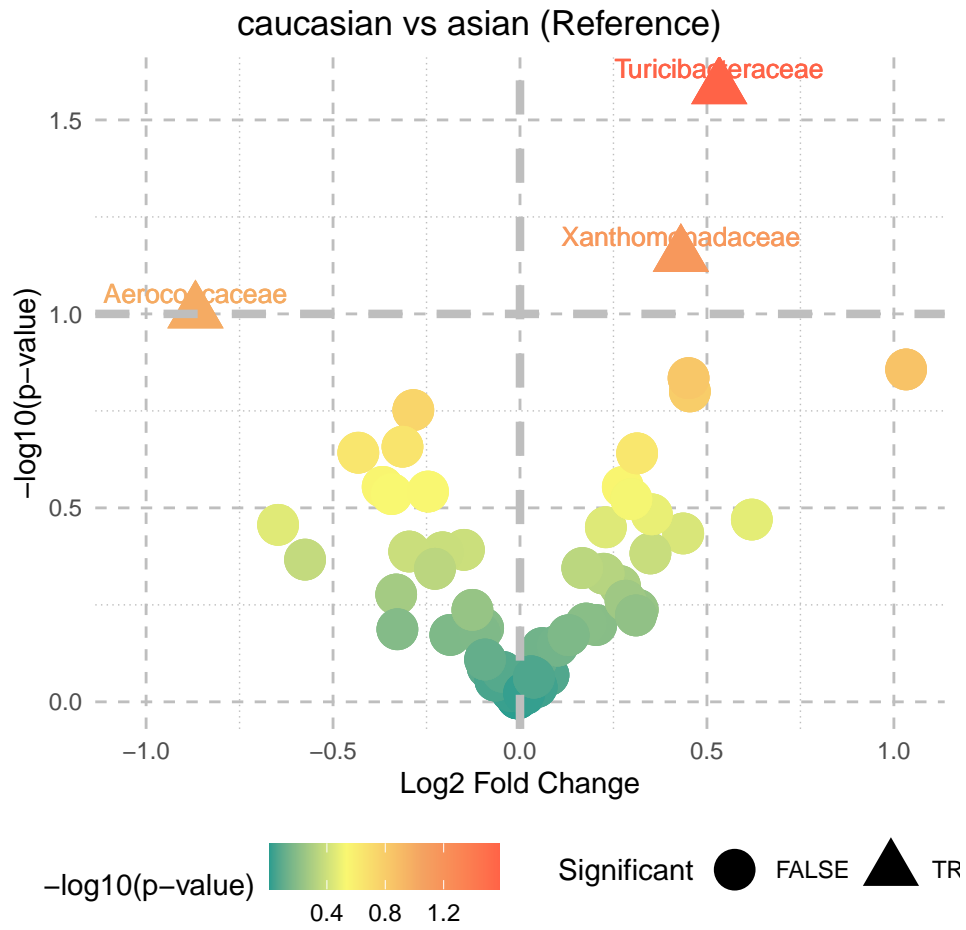


4.2 Trend test

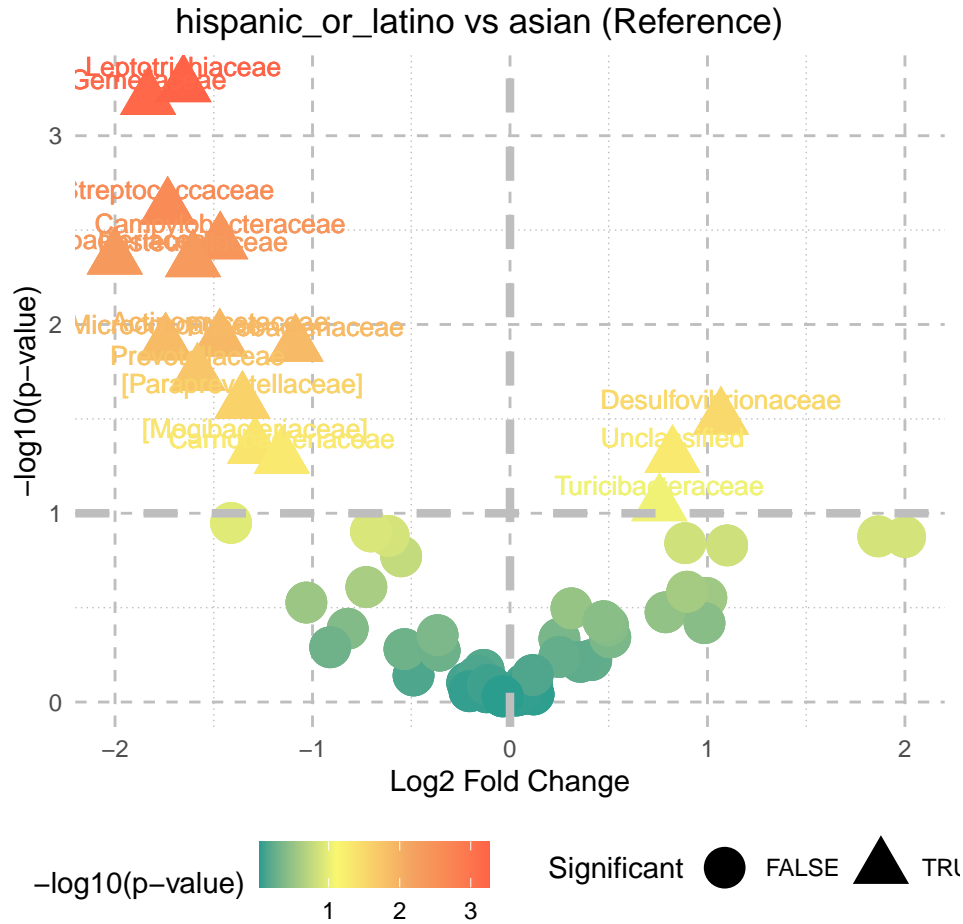
```
taxa_trend_test_results <- generate_taxa_trend_test_long(  
  data.obj = data.obj,  
  subject.var = subject.var,  
  time.var = time.var,  
  group.var = group.var,  
  adj.vars = test.adj.vars,  
  prev.filter = prev.filter,  
  abund.filter = abund.filter,  
  feature.level = feature.level,  
  feature.dat.type = feature.dat.type,  
  ...)
```



\$Family *Family*caucasian vs asian (Reference)



*Family*hispanic_or_latino vs asian (Reference)



In this analysis, we utilized the LinDA linear mixed effects model to investigate potential interactions in the context of Taxa Trend Test. Specifically, we tested the interaction between the variables subject_race and visit_number_num, for different taxa, while adjusting for other covariates.

- Significant features in trend test results

For the taxon Family, significant interactions were identified in the Taxa Trend Test Results using the none method for p-value adjustment, based on a threshold of 0.1:

Table 27: Table continues below

| Variable | Group | Base.Mean | Log2.Fold.Change |
|----------------------|--------------------|-----------|------------------|
| Aerococcaceae | hispanic_or_latino | 1413 | -0.868 |
| Aerococcaceae | asian | 1413 | -0.868 |
| Aerococcaceae | caucasian | 1413 | -0.868 |
| Turicibacteraceae | hispanic_or_latino | 149.4 | 0.533 |
| Turicibacteraceae | asian | 149.4 | 0.533 |
| Turicibacteraceae | caucasian | 149.4 | 0.533 |
| Xanthomonadaceae | hispanic_or_latino | 147.3 | 0.4303 |
| Xanthomonadaceae | asian | 147.3 | 0.4303 |
| Xanthomonadaceae | caucasian | 147.3 | 0.4303 |
| [Mogibacteriaceae] | hispanic_or_latino | 297.6 | -1.291 |
| [Mogibacteriaceae] | asian | 297.6 | -1.291 |
| [Mogibacteriaceae] | caucasian | 297.6 | -1.291 |
| [Paraprevotellaceae] | hispanic_or_latino | 285.3 | -1.355 |

| Variable | Group | Base.Mean | Log2.Fold.Change |
|----------------------|--------------------|-----------|------------------|
| [Paraprevotellaceae] | asian | 285.3 | -1.355 |
| [Paraprevotellaceae] | caucasian | 285.3 | -1.355 |
| Actinomycetaceae | hispanic_or_latino | 258.5 | -1.47 |
| Actinomycetaceae | asian | 258.5 | -1.47 |
| Actinomycetaceae | caucasian | 258.5 | -1.47 |
| Campylobacteraceae | hispanic_or_latino | 292.8 | -1.467 |
| Campylobacteraceae | asian | 292.8 | -1.467 |
| Campylobacteraceae | caucasian | 292.8 | -1.467 |
| Carnobacteriaceae | hispanic_or_latino | 284.7 | -1.155 |
| Carnobacteriaceae | asian | 284.7 | -1.155 |
| Carnobacteriaceae | caucasian | 284.7 | -1.155 |
| Desulfovibrionaceae | hispanic_or_latino | 258 | 1.068 |
| Desulfovibrionaceae | asian | 258 | 1.068 |
| Desulfovibrionaceae | caucasian | 258 | 1.068 |
| Flavobacteriaceae | hispanic_or_latino | 252.6 | -1.086 |
| Flavobacteriaceae | asian | 252.6 | -1.086 |
| Flavobacteriaceae | caucasian | 252.6 | -1.086 |
| Fusobacteriaceae | hispanic_or_latino | 280.2 | -2.001 |
| Fusobacteriaceae | asian | 280.2 | -2.001 |
| Fusobacteriaceae | caucasian | 280.2 | -2.001 |
| Gemellaceae | hispanic_or_latino | 210.2 | -1.833 |
| Gemellaceae | asian | 210.2 | -1.833 |
| Gemellaceae | caucasian | 210.2 | -1.833 |
| Leptotrichiaceae | hispanic_or_latino | 173.2 | -1.654 |
| Leptotrichiaceae | asian | 173.2 | -1.654 |
| Leptotrichiaceae | caucasian | 173.2 | -1.654 |
| Micrococcaceae | hispanic_or_latino | 669.2 | -1.744 |
| Micrococcaceae | asian | 669.2 | -1.744 |
| Micrococcaceae | caucasian | 669.2 | -1.744 |
| Pasteurellaceae | hispanic_or_latino | 277.5 | -1.605 |
| Pasteurellaceae | asian | 277.5 | -1.605 |
| Pasteurellaceae | caucasian | 277.5 | -1.605 |
| Prevotellaceae | hispanic_or_latino | 2399 | -1.582 |
| Prevotellaceae | asian | 2399 | -1.582 |
| Prevotellaceae | caucasian | 2399 | -1.582 |
| Streptococcaceae | hispanic_or_latino | 7473 | -1.734 |
| Streptococcaceae | asian | 7473 | -1.734 |
| Streptococcaceae | caucasian | 7473 | -1.734 |
| Turicibacteraceae | hispanic_or_latino | 149.4 | 0.7567 |
| Turicibacteraceae | asian | 149.4 | 0.7567 |
| Turicibacteraceae | caucasian | 149.4 | 0.7567 |
| Unclassified | hispanic_or_latino | 873824 | 0.8234 |
| Unclassified | asian | 873824 | 0.8234 |
| Unclassified | caucasian | 873824 | 0.8234 |

Table 28: Table continues below

| LFC.SE | Stat | P.Value | Adjusted.P.Value | Mean.Abundance |
|--------|--------|---------|------------------|----------------|
| 0.5251 | -1.653 | 0.09892 | 0.9797 | 0.01752 |
| 0.5251 | -1.653 | 0.09892 | 0.9797 | 0.02287 |
| 0.5251 | -1.653 | 0.09892 | 0.9797 | 0.01985 |

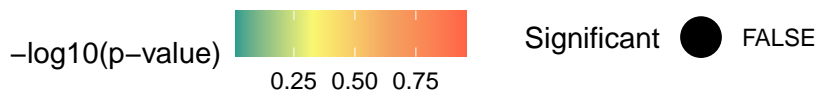
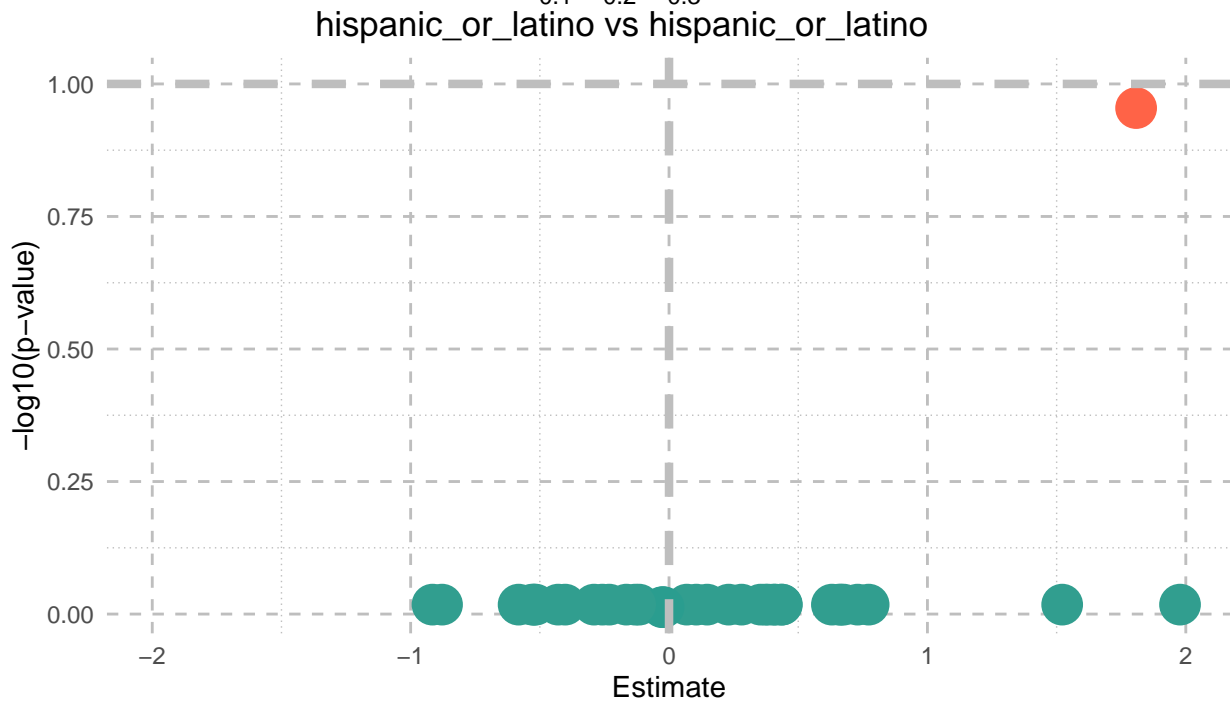
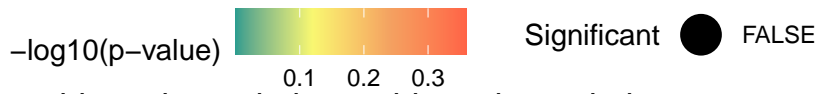
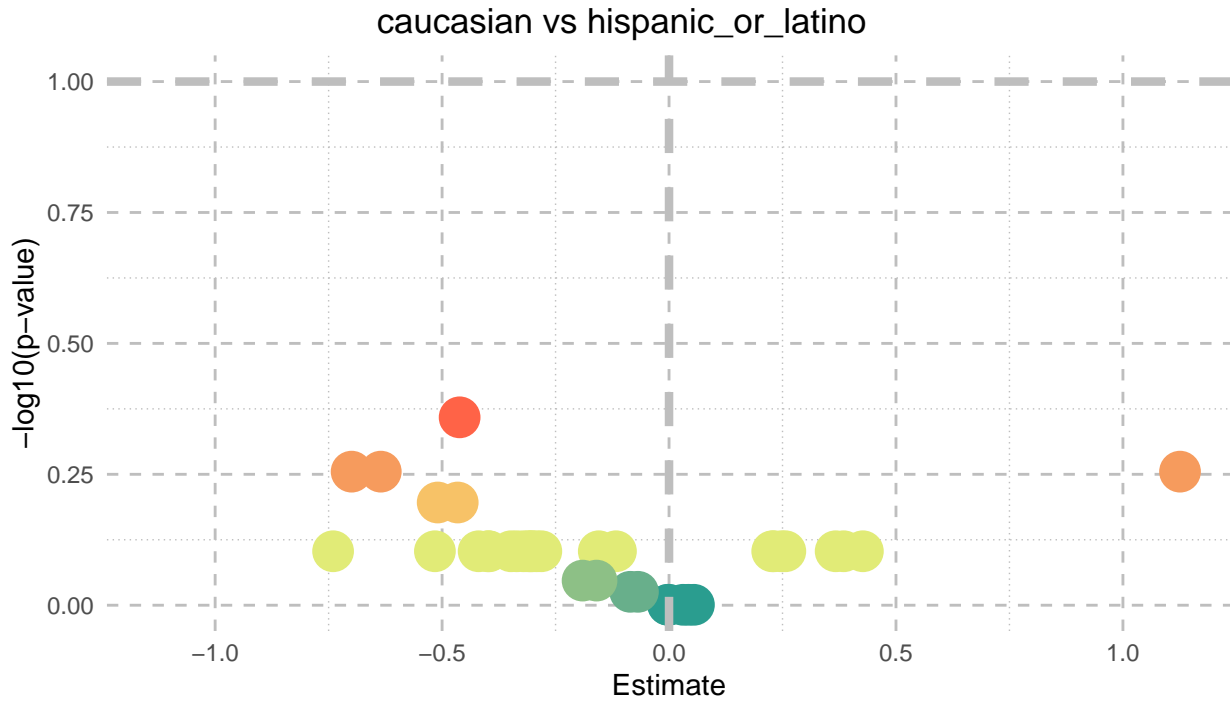
| LFC.SE | Stat | P.Value | Adjusted.P.Value | Mean.Abandance |
|--------|--------|-----------|------------------|----------------|
| 0.2383 | 2.237 | 0.02615 | 0.9797 | 0.000145 |
| 0.2383 | 2.237 | 0.02615 | 0.9797 | 0.000167 |
| 0.2383 | 2.237 | 0.02615 | 0.9797 | 0.0001694 |
| 0.2331 | 1.846 | 0.07078 | 0.9797 | 0.0002512 |
| 0.2331 | 1.846 | 0.07078 | 0.9797 | 0.0001199 |
| 0.2331 | 1.846 | 0.07078 | 0.9797 | 0.000117 |
| 0.6413 | -2.013 | 0.04488 | 0.2063 | 0.0008564 |
| 0.6413 | -2.013 | 0.04488 | 0.2063 | 0.0003626 |
| 0.6413 | -2.013 | 0.04488 | 0.2063 | 0.0003651 |
| 0.6064 | -2.234 | 0.02592 | 0.1437 | 0.002614 |
| 0.6064 | -2.234 | 0.02592 | 0.1437 | 0.002182 |
| 0.6064 | -2.234 | 0.02592 | 0.1437 | 0.001534 |
| 0.5637 | -2.607 | 0.01212 | 0.0875 | 0.0008029 |
| 0.5637 | -2.607 | 0.01212 | 0.0875 | 0.0004126 |
| 0.5637 | -2.607 | 0.01212 | 0.0875 | 0.0004979 |
| 0.5031 | -2.917 | 0.003702 | 0.04682 | 0.000645 |
| 0.5031 | -2.917 | 0.003702 | 0.04682 | 0.0005537 |
| 0.5031 | -2.917 | 0.003702 | 0.04682 | 0.0004001 |
| 0.5878 | -1.965 | 0.05073 | 0.2063 | 0.0008896 |
| 0.5878 | -1.965 | 0.05073 | 0.2063 | 0.0006612 |
| 0.5878 | -1.965 | 0.05073 | 0.2063 | 0.0005388 |
| 0.495 | 2.158 | 0.03143 | 0.1597 | 0.0006967 |
| 0.495 | 2.158 | 0.03143 | 0.1597 | 0.000293 |
| 0.495 | 2.158 | 0.03143 | 0.1597 | 0.0001973 |
| 0.4352 | -2.495 | 0.01291 | 0.0875 | 0.0002975 |
| 0.4352 | -2.495 | 0.01291 | 0.0875 | 0.0002975 |
| 0.4352 | -2.495 | 0.01291 | 0.0875 | 0.0002261 |
| 0.6731 | -2.973 | 0.00446 | 0.04682 | 0.0008422 |
| 0.6731 | -2.973 | 0.00446 | 0.04682 | 0.0007432 |
| 0.6731 | -2.973 | 0.00446 | 0.04682 | 0.0008845 |
| 0.5315 | -3.448 | 0.0006369 | 0.01943 | 0.0007965 |
| 0.5315 | -3.448 | 0.0006369 | 0.01943 | 0.000385 |
| 0.5315 | -3.448 | 0.0006369 | 0.01943 | 0.0004621 |
| 0.4751 | -3.481 | 0.0005449 | 0.01943 | 0.0007454 |
| 0.4751 | -3.481 | 0.0005449 | 0.01943 | 0.0003274 |
| 0.4751 | -3.481 | 0.0005449 | 0.01943 | 0.0002468 |
| 0.6971 | -2.501 | 0.0127 | 0.0875 | 0.004551 |
| 0.6971 | -2.501 | 0.0127 | 0.0875 | 0.001211 |
| 0.6971 | -2.501 | 0.0127 | 0.0875 | 0.002661 |
| 0.5581 | -2.876 | 0.004605 | 0.04682 | 0.0007082 |
| 0.5581 | -2.876 | 0.004605 | 0.04682 | 0.0003343 |
| 0.5581 | -2.876 | 0.004605 | 0.04682 | 0.0003255 |
| 0.669 | -2.365 | 0.01839 | 0.1122 | 0.07008 |
| 0.669 | -2.365 | 0.01839 | 0.1122 | 0.02082 |
| 0.669 | -2.365 | 0.01839 | 0.1122 | 0.01324 |
| 0.5687 | -3.048 | 0.002425 | 0.04682 | 0.01432 |
| 0.5687 | -3.048 | 0.002425 | 0.04682 | 0.006119 |
| 0.5687 | -3.048 | 0.002425 | 0.04682 | 0.009239 |
| 0.4454 | 1.699 | 0.09035 | 0.3445 | 0.000145 |
| 0.4454 | 1.699 | 0.09035 | 0.3445 | 0.000167 |
| 0.4454 | 1.699 | 0.09035 | 0.3445 | 0.0001694 |
| 0.4102 | 2.007 | 0.05016 | 0.2063 | 0.1207 |

| Mean.Prevalence | Output.Element |
|-----------------|---|
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |
| 1 | subject_racehispanic_or_latino:visit_number_num |

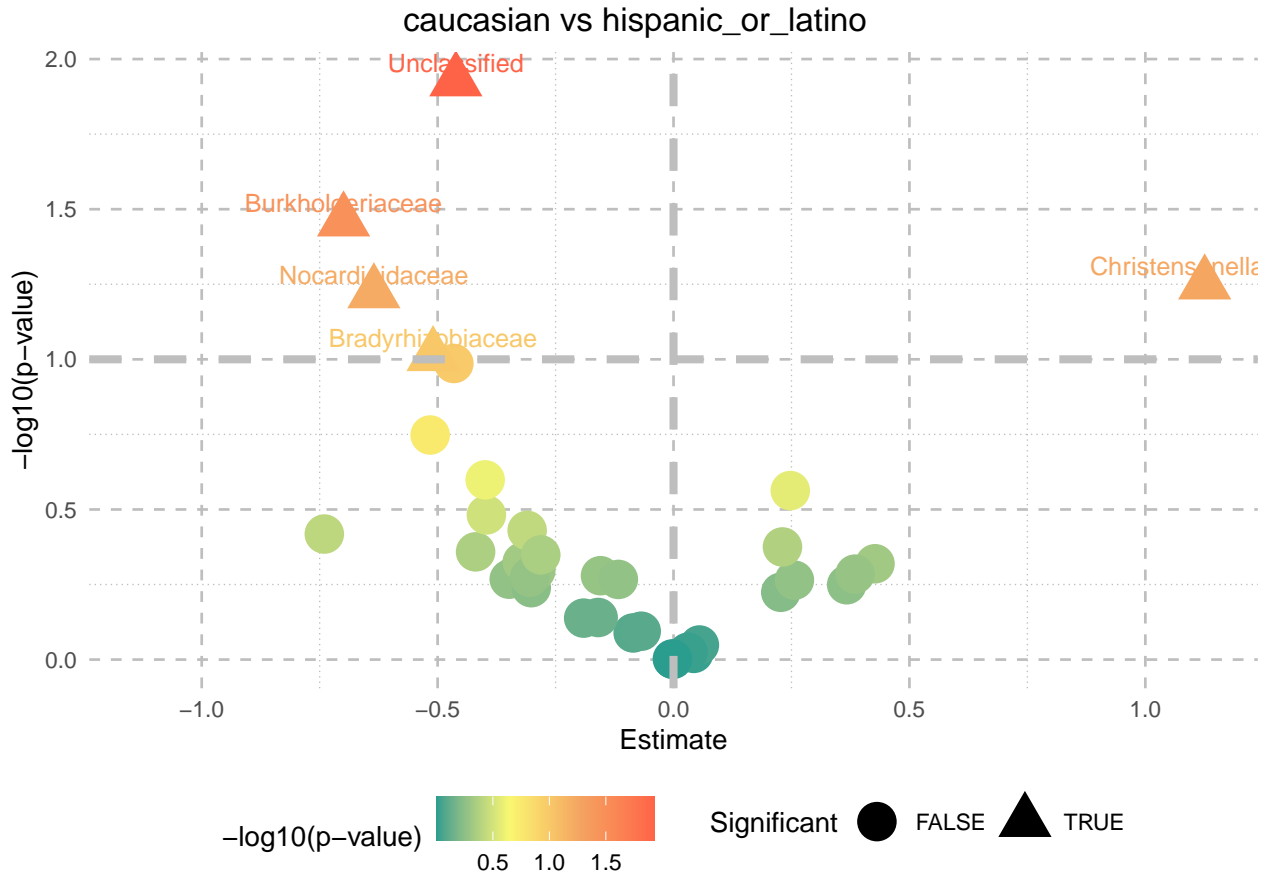
The taxa trend test results for individual taxa or features have been saved in the current working directory. Each taxa rank has its own file named with the prefix: `taxa_trend_test_results_` followed by the taxon rank and the file extension `.csv`. Please refer to these files for more detailed data.

4.3 Volatility test

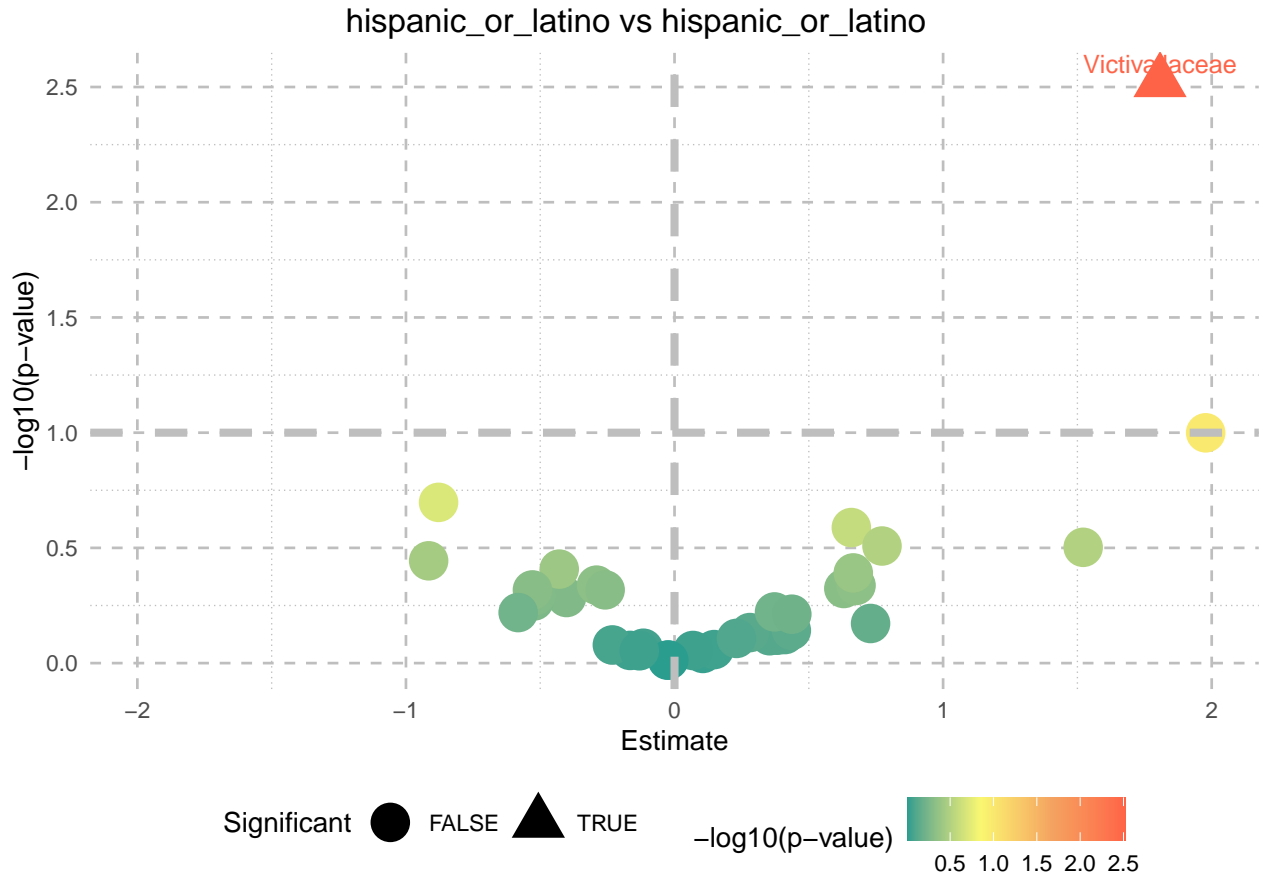
```
taxa_volatility_test_results <- generate_taxa_volatility_test_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  group.var = group.var,
  adj.vars = test.adj.vars,
  prev.filter = prev.filter,
  abund.filter = abund.filter,
  feature.level = feature.level,
  feature.dat.type = feature.dat.type
)
```

\$Family *Family*subject_racecaucasian



*Family*subject_racehispanic_or_latino



In this analysis, a general linear model was utilized to investigate the influence of the variable `subject_race` on the volatility of various taxa abundances.

Taxa abundances were transformed using the centered log-ratio (CLR) transformation. For count data, 0.5 was added to all counts before performing the CLR. For proportion data, zeros were replaced by half the minimum non-zero proportion for each taxon.

- Significant features in volatility test results

– **Bifidobacteriaceae:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|-----------|
| (Intercept) | 0.9004 | 0.2497 | 3.606 | 0.0006272 |
| subject_racecaucasian | -0.4656 | 0.2815 | -1.654 | 0.1032 |
| subject_racehispanic_or_latino | 0.6583 | 0.5766 | 1.142 | 0.2581 |
| subject_race | NA | NA | 3.22 | 0.04682 |
| Residuals | NA | NA | NA | NA |

– **Burkholderiaceae:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|-----------|
| (Intercept) | 1.466 | 0.2871 | 5.105 | 3.484e-06 |
| subject_racecaucasian | -0.6993 | 0.3236 | -2.161 | 0.03464 |
| subject_racehispanic_or_latino | -0.1158 | 0.663 | -0.1747 | 0.8619 |
| subject_race | NA | NA | 2.587 | 0.08351 |
| Residuals | NA | NA | NA | NA |

| Term | Estimate | Std.Error | Statistic | P.Value |
|------|----------|-----------|-----------|---------|
|------|----------|-----------|-----------|---------|

– **Christensenellaceae:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|----------|
| (Intercept) | 1.477 | 0.5129 | 2.879 | 0.005492 |
| subject_racecaucasian | 1.126 | 0.5782 | 1.947 | 0.05615 |
| subject_racehispanic_or_latino | 1.978 | 1.185 | 1.669 | 0.1001 |
| subject_race | NA | NA | 2.394 | 0.09984 |
| Residuals | NA | NA | NA | NA |

– **Victivallaceae:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|----------|
| (Intercept) | 0.1722 | 0.2532 | 0.6801 | 0.499 |
| subject_racecaucasian | 0.2309 | 0.2854 | 0.809 | 0.4216 |
| subject_racehispanic_or_latino | 1.808 | 0.5847 | 3.091 | 0.003001 |
| subject_race | NA | NA | 4.863 | 0.01098 |
| Residuals | NA | NA | NA | NA |

– **Unclassified:**

| Term | Estimate | Std.Error | Statistic | P.Value |
|--------------------------------|----------|-----------|-----------|-----------|
| (Intercept) | 1.458 | 0.1577 | 9.244 | 3.305e-13 |
| subject_racecaucasian | -0.4614 | 0.1778 | -2.595 | 0.01183 |
| subject_racehispanic_or_latino | -0.2582 | 0.3642 | -0.709 | 0.481 |
| subject_race | NA | NA | 3.416 | 0.03926 |
| Residuals | NA | NA | NA | NA |

The taxa volatility test results for individual taxa or features have been saved in the current working directory. Each taxa rank and sub-rank combination has its own file named with the prefix: taxa_volatility_test_results_ followed by the main taxon, sub-taxon, and the file extension .csv. Please refer to these files for more detailed data.

4.4 Data visualization(significant features)

```
if (!is.null(combined_significant_taxa)){
  taxa_boxplot_results <- generate_taxa_boxplot_long(
    data.obj = data.obj,
    subject.var = subject.var,
    time.var = time.var,
    t0.level = t0.level,
    ts.levels = ts.levels,
    group.var = group.var,
    strata.var = strata.var,
    feature.level = feature.level,
    feature.dat.type = feature.dat.type,
    features.plot = combined_significant_taxa,
```

```

transform = transform,
top.k.plot = top.k.plot,
top.k.func = top.k.func,
prev.filter = prev.filter,
abund.filter = abund.filter,
base.size = 10,
theme.choice = theme.choice,
custom.theme = custom.theme,
palette = palette,
pdf = pdf,
file.ann = file.ann,
pdf.wid = pdf.wid,
pdf.hei = pdf.hei)

taxa_indiv_boxplot_results <- generate_taxa_indiv_boxplot_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  feature.level = feature.level,
  features.plot = combined_significant_taxa,
  transform = transform,
  feature.dat.type = feature.dat.type,
  top.k.plot = top.k.plot,
  top.k.func = top.k.func,
  prev.filter = prev.filter,
  abund.filter = abund.filter,
  base.size = 10,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = TRUE,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei)
}

```

4.4.1 Significant features boxplot The boxplot results for individual taxa or features can be found in the current working directory. The relevant file is named: `taxa_indiv_boxplot_long_subject_subject_id_time_visit_number_num_feature_level_Family_transform_sqrt_prev_filter_0.1_abund_filter_1e-04_group_subject_race_strata_subject_gender_Family.pdf`. Please refer to this file for more detailed visualizations.

```

cat('
')

```

```

if (!is.null(combined_significant_taxa)){
  taxa_spaghettoplot_results <- generate_taxa_spaghettoplot_long(
    data.obj = data.obj,

```

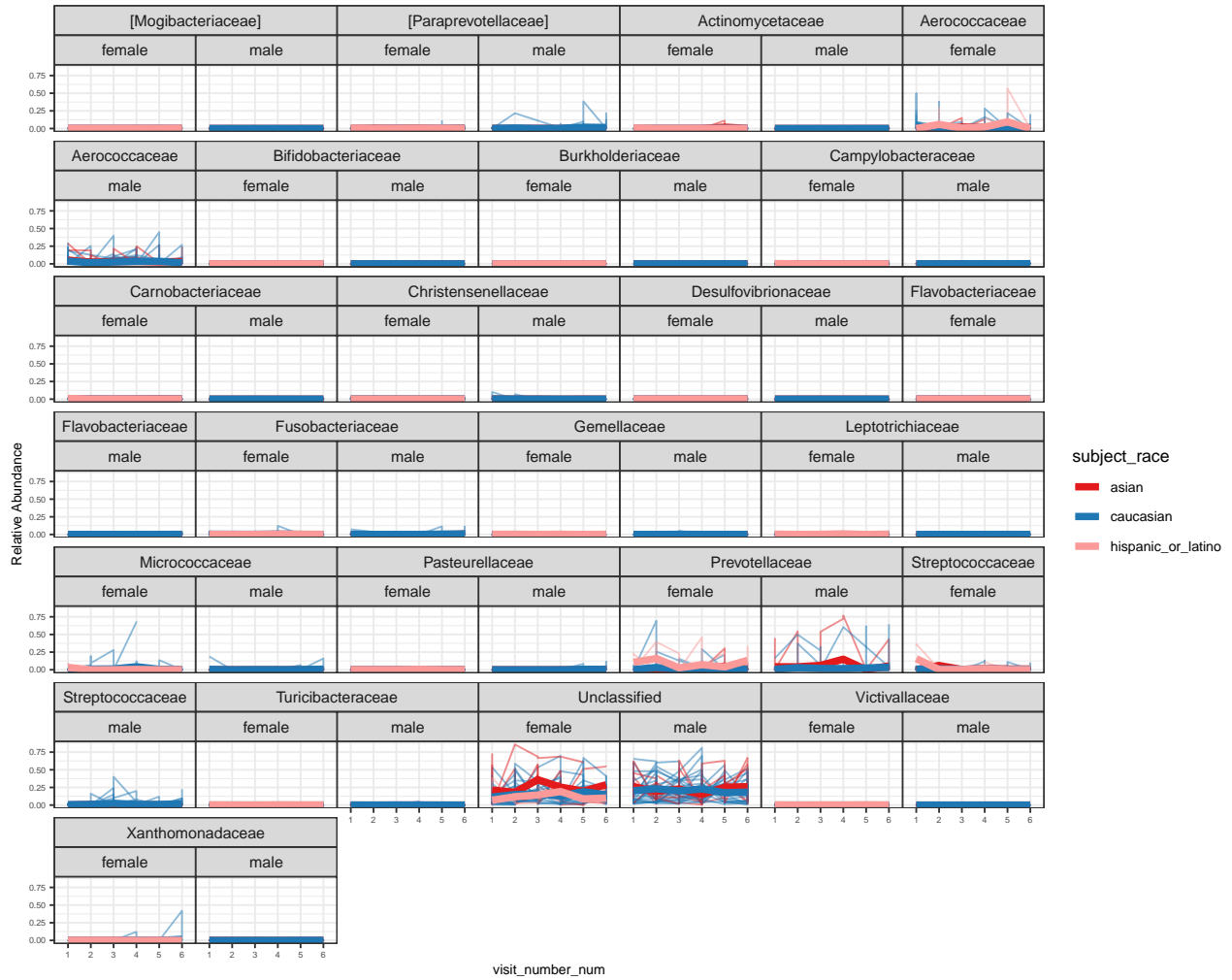
```

        subject.var = subject.var,
        time.var = time.var,
        group.var = group.var,
        strata.var = strata.var,
        t0.level = t0.level,
        ts.levels = ts.levels,
        feature.level = feature.level,
        feature.dat.type = feature.dat.type,
        features.plot = combined_significant_taxa,
        top.k.plot = top.k.plot,
        top.k.func = top.k.func,
        prev.filter = prev.filter,
        abund.filter = abund.filter,
        base.size = 10,
        theme.choice = theme.choice,
        custom.theme = custom.theme,
        palette = palette,
        pdf = pdf,
        file.ann = file.ann,
        pdf.wid = pdf.wid,
        pdf.hei = pdf.hei)

taxa_indiv_spaghettoplot_results <- generate_taxa_indiv_spaghettoplot_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  change.base = change.base,
  feature.change.func = feature.change.func,
  feature.level = feature.level,
  features.plot = combined_significant_taxa,
  feature.dat.type = feature.dat.type,
  top.k.plot = top.k.plot,
  top.k.func = top.k.func,
  prev.filter = prev.filter,
  abund.filter = abund.filter,
  base.size = 10,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = TRUE,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei)
}

```

4.4.2 Significant features spaghettoplot `[[1]]`



The spaghettiplot results for individual taxa or features can be found in the current working directory. The relevant file is named: `taxa_indiv_spaghettiplot_long_subject_subject_id_time_visit_number_num_group_subject_race_strata_subject_gender_feature_level_Family_prev_filter_0.1_abund_filter_1e-04_base_size_12_theme_choice_bw_pdf_wid_11_pdf_hei_8.5.pdf`. Please refer to this file for more detailed visualizations.