# Visualization Project

By: Shreeman Gautam

# Requirements for Visualization

- Visualization must be able to show relative abundance of microbiomes in individuals.
- Two individuals can be easily compared on the type and abundance of microbiomes they have.
- Visualization can show clear differences in individuals with different diseases(compared to healthy individuals).

# Requirement for Visualization[L = Literature]

- Visualization should very likely be 2-D, not 3-D.
  - Most, if not all, papers I read used 2D.
  - Recurring arguments were made in papers that 3-D representations caused occlusion.

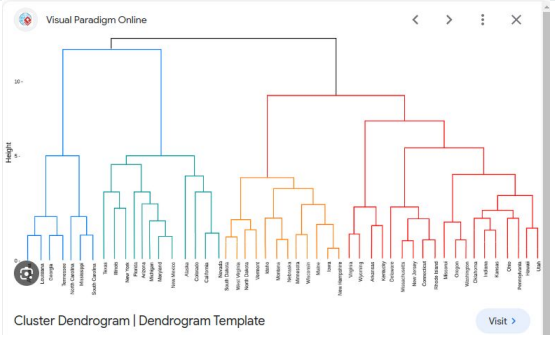
# Requirements for 2D Visualization - Hierarchical[L]

- Implicit vs Explicit, Axes vs Radial, 2 by 2 = 4 categories of visualizations of hierarchical data.[1]
  - Visualization should fall under one of these categories.

**Table 4:** *Types of visualization methods of hierarchical data.*

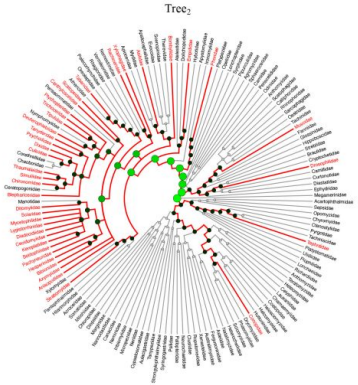
| Type | Description | Axes-Oriented Layout | Radial Layout |
|---|---|---|---|
| *Explicit* | Visualization method representing hierarchy as a node-link diagram [55] | Dendrogram, Intended tree (see Fig. 2, Fig. 4) | Circular tree (see Fig. 3) |
| *Implicit* | Visualization method representing hierarchy in a space-filling way [55] | Tree-maps (see Fig. 5) | Sunburst (see Fig. 6) |

# 2D Visualization – Explicit?[L]





- Arguments against:
  - Explicit Axes:
    - Examples: Dendrogram and Indented List.
    - Less scalable with lots of data.
  - Explicit Radial:
    - Example: Circular tree[**to the right of this**].
    - Not good with lots of data.[2]
- Arguments for:
  - Easier than implicit to Trace Ancestry of leaf or node.
  - General Structure of Hierarchy Apparent



# 2D Visualization – Implicit?[L]

- Arguments against:
  - Hierarchy not as apparent compared to explicit.
- Arguments for:
  - Implicit representation is more suitable with lots of hierarchical data.[3]
  - Treemap[**top right**] and Sunburst[**bottom right**] are examples.

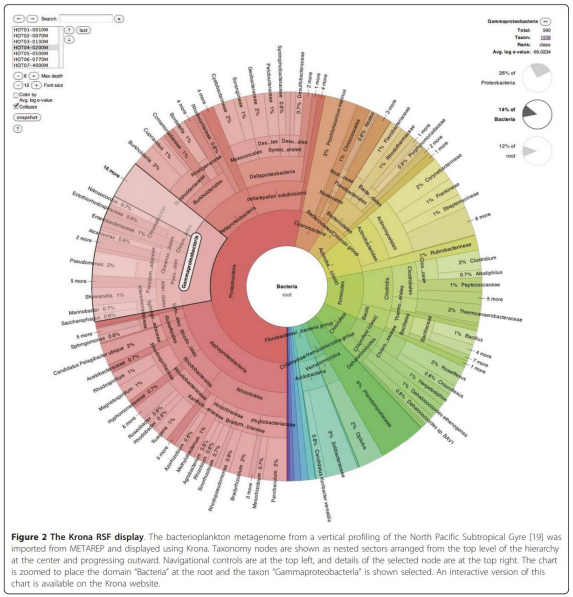


**Figure 2 The Krona RSF display**. The bacterioplankton metagenome from a vertical profiling of the North Pacific Subtropical Gyre [19] was imported from METAREP and displayed using Krona. Taxonomy nodes are shown as nested sectors arranged from the top level of the hierarchy at the center and progressing outward. Navigational controls are at the top left, and details of the selected node are at the top right. The chart is zoomed to place the domain "Bacteria" at the root and the taxon "Gammaproteobacteria" is shown selected. An interactive version of this chart is available on the Krona website.

# Implicit Axes Or Radial?[L]

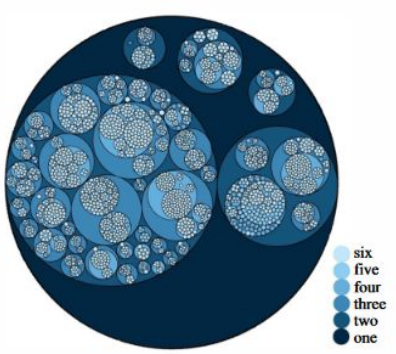
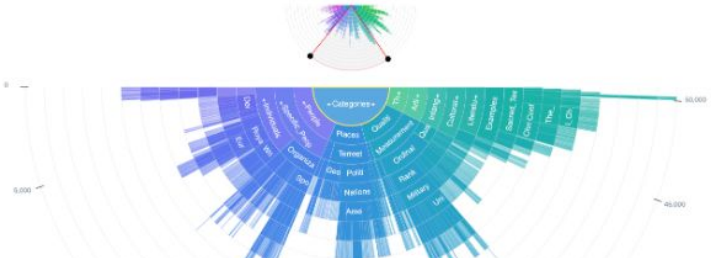



Figure 2: Circular treemaps automatically generated by our variational algorithm for the hierarchical data with 6 levels and 2516 nodes. The color legend indicates the level of the hierarchy.

- From this paper[4], 17 people found:
  - Sunburst and Circular Treemaps performed best for large hierarchies.
    - Large Hierarchy defined as 6 levels with 2-5 child nodes per level
  - Treemap, Icicle Plot[**bottom right**] and Circular Treemaps[**top right**] performed best for small hierarchies.
    - Small hierarchy defined as 2 levels with 10-20 child nodes per level
- From this paper[5], 12 people preferred a sundown chart and icicle plot over treemap for representation of hierarchical quantitative data.
  - Sundown chart[**bottom left**] = semicircular variant of sunburst



# Sunburst or Circular Treemap. Multiple Views?[L]

- Having different tree representations at the same time can be complementary.[6]
  - Each representation has its strengths and weaknesses.
  - Users can walk away getting a full picture of the data.

# Why Sunburst?[L]

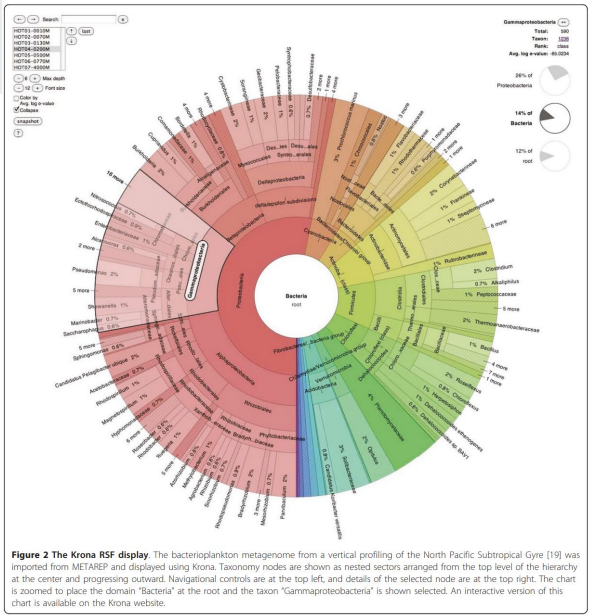


**Figure 2 The Krona RSF display**. The bacterioplankton metagenome from a vertical profiling of the North Pacific Subtropical Gyre [19] was imported from METAREP and displayed using Krona. Taxonomy nodes are shown as nested sectors arranged from the top level of the hierarchy at the center and progressing outward. Navigational controls are at the top left, and details of the selected node are at the top right. The chart is zoomed to place the domain "Bacteria" at the root and the taxon "Gammaproteobacteria" is shown selected. An interactive version of this chart is available on the Krona website.

- This paper[7] uses a sunburst(**Krona**) to show metagenomic hierarchy.
  - Not specifically a microbiome hierarchy.
  - Paper has 1500+ citations on google scholar.
  - Many papers(from what I gleaned) that cited this paper were/had
    - Biological in nature and not published at viz conferences.
    - Sunburst to represent microbiome hierarchy
    - This is a paper[8] that cited **Krona(there are many more like this):**
      - **https://www.sciencedirect.com/science/article/pii/S2001037023003859**
- Circular treemap is still an option(along with other representations) for visualization!

# Visualizations seen along the way[L]

Taken from this paper[9].

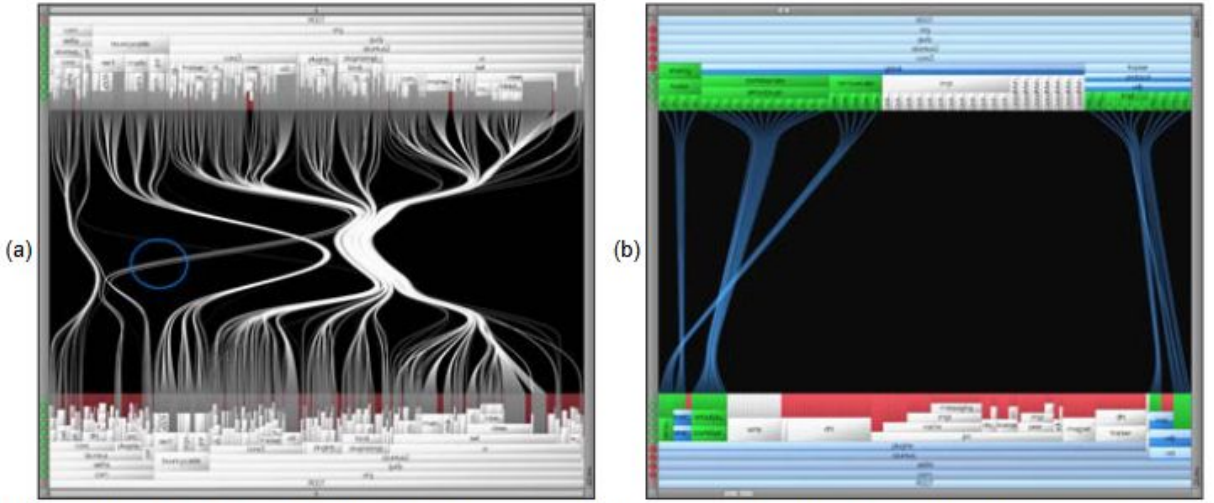



**Figure 7:** *Comparison of larger data sets; (a)* Azureus *v2.2 (top, 2,283 leafs) is compared to* Azureus *v2.3 (bottom, 3,179 leafs). Red shading is used to depict nodes that are unique to each hierarchy. The encircled area hints at relocation of part of the hierarchy; (b) an SE inspects the encircled area in more detail by crossing it. Zooming is subsequently performed to provide the SE with a more detailed view of the changes.*

# Visualizations seen along the way[L]

Taken from this paper[10].

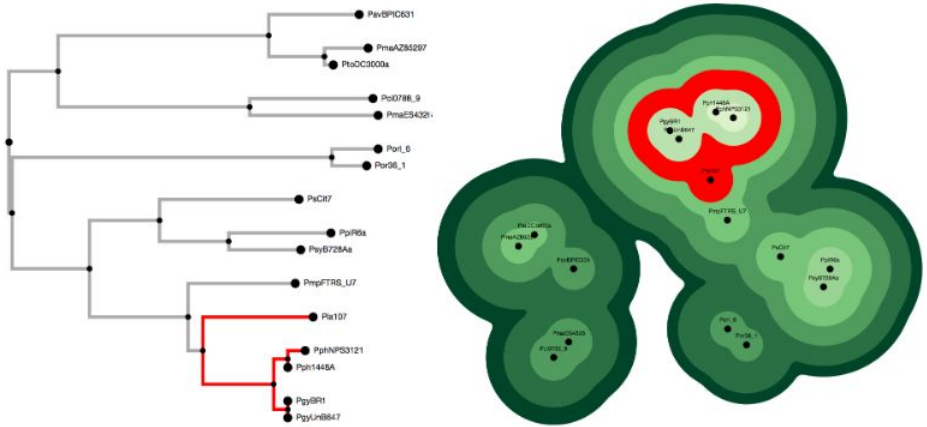



**Fig 2.** a) Phylogenetic tree with 16 OTUs. b) Topo-phylogeny chart based on the same structure. The equivalent branches are highlighted red in both figures.

# Visualizations seen along the way[L]

Taken from this paper[11].

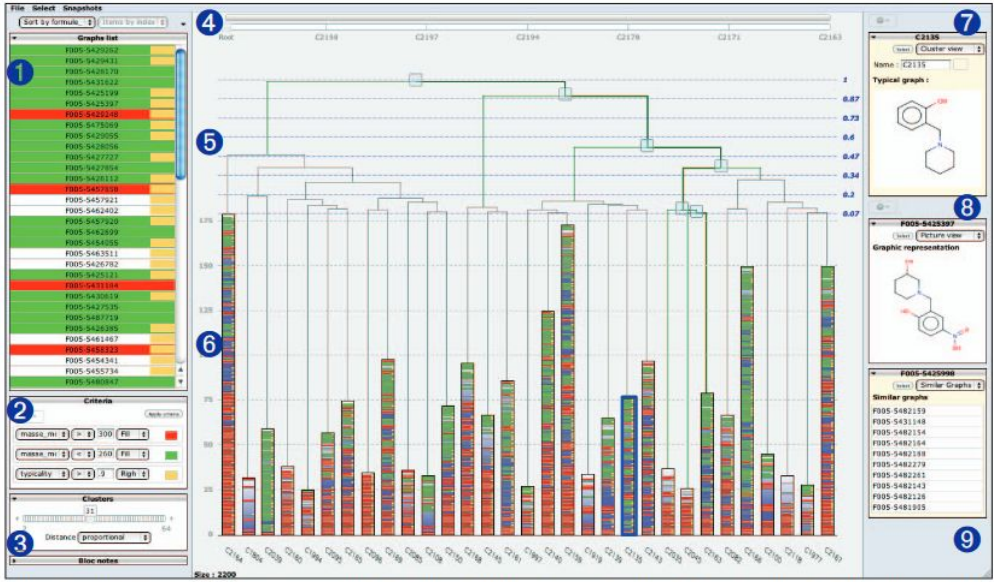


Figure 3. Here is a screenshot of the prototype. The blue numbers highlight the main parts of this interface whose roles are detailed in the paper. The central part contains a Stacked Tree (4,5,6). The left (1,2,3) and right (7,8,9) widgets allow to explore and to control this structure but are independent of the Stacked Trees. While this screenshot has been done on a small 13" screen, 2200 molecule are displayed with for each one up to three properties.

# Visualizations seen along the way[L]

Taken from this paper[12].

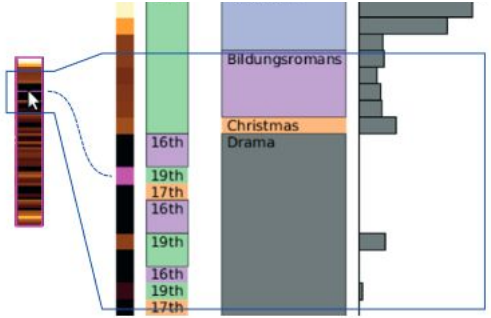



**Figure 4:** *Left: Heat map that is hovered over in the main view, it depicts the frequency of words that belong to the 'fuel' class per novel, where novels are ordered by genre. Right: Instance view, showing meta-data for two properties, and frequencies as bars. For illustrative purposes, the entire main view tree is not shown, and the blue lines have been added to show the relation between the low-detail heat map and high-detail instance view. Observe that drama novels make less mention of fuel than adventure novels.*

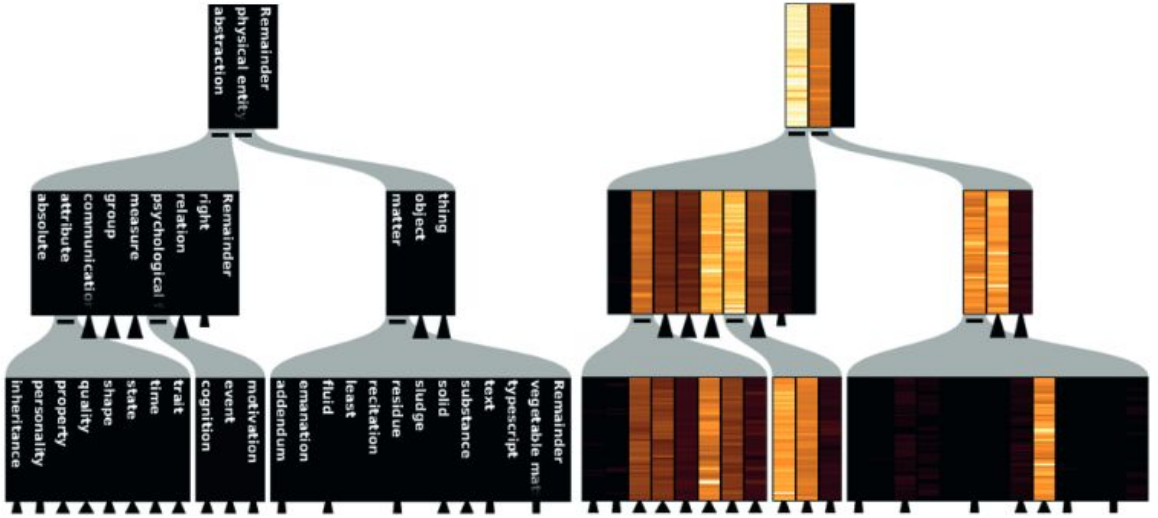
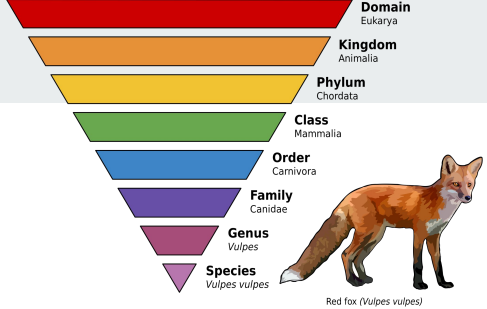



**Figure 3:** *Left: The hierarchy representation of the main view as applied to the novels data set, where cells display the names of their nodes (in this case entity classes). Right: The same representation, but where cells display instance weights with heat maps (every horizontal line represents a novel). The weights are scaled by groups of siblings, which in this case enables the identification of overall abundant words. For example, most novels contain less words that describe something physical than those that describe something abstract, and words that refer to some form of matter mostly refer to a substance.*




# Hierarchical Data That We Have

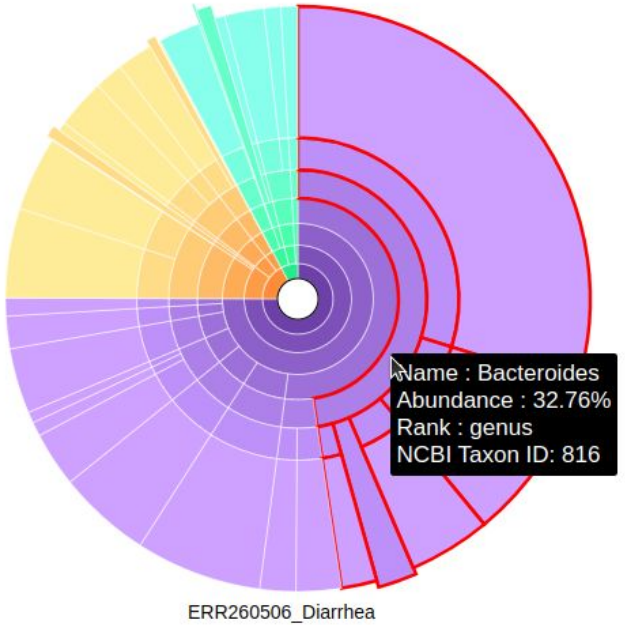
- We will use the ranks[plus strain, minus domain] from the above picture in our data.
- We have 17 **csv** files and each **csv** file is a hierarchical sample comprising an individual's microbiomes.
  - Below picture is a snippet of a sample.
- These **csv**s are converted to **json** files which store the hierarchical samples like a "tree".
  - Picture on the right is a "tree" version of below picture.
- Pictures taken from sample: **ERR260506_Diarrhea**


```
{
    "name": "sk__Bacteria__2",
    "children": [
        {
            "name": "p__Bacteroidota__976",
            "children": [
                {
                    "name": "c__Bacteroidia__200643",
                    "children": [
                        {
                            "name": "o__Bacteroidales__171549",
                            "children": [
                                {
                                    "name": "f__Bacteroidaceae__815",
                                    "children": [
                                        {
                                            "name": "g__Bacteroides__816",
                                            "children": [
                                                {
                                                    "name": "s__Bacteroides fragilis__817",
                                                    "children": [
                                                        {
                                                            "name": "Bacteroides fragilis?species?817",
                                                            "value": 0.205231
                                                        }
                                                    ]
                                                },
                                                {
                                                    "name": "s__Bacteroides caccae__47678",
                                                    "children": [
                                                        {
                                                            "name": "Bacteroides caccae ATCC 43185?strain?411901",
                                                            "value": 0.06312
                                                        }
                                                    ]
                                                },
                                                {
                                                    "name": "s__Bacteroides sp. 1_1_30__457387",
                                                    "children": [
                                                        {
                                                            "name": "Bacteroides sp. 1_1_30?species?457387",
                                                            "value": 0.031264
                                                        }
                                                    ]
                                                },
                                                {
                                                    "name": "s__Bacteroides xylanisolvens__371601",
                                                    "children": [
                                                        {
                                                            "name": "Bacteroides xylanisolvens XB1A?strain?657309",
                                                            "value": 0.013509
                                                        }
                                                    ]
                                                },
                                                {
                                                    "name": "Bacteroides?genus?816",
                                                    "value": 0.014524
                                                }
                                            ]
                                        }
                                    ]
                                }
                            ]
                        }
                    ]
                }
            ]
        },
```


| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | run_id | ncbi_taxon_id | taxon_rank_level | relative_abundance | name | lineage | CDF | | | | | | | | |
| 2 | ERR260506 | 817 | species | 0.205231 | Bacteroides fragilis | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | s__Bacteroides fragilis__817 | 0.989406697423317 | | |
| 3 | ERR260506 | 411901 | strain | 0.06312 | Bacteroides caccae ATCC 43185 | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | s__Bacteroides caccae__47678 | st__Bacteroides caccae ATCC 43185__411901 | 0.989300109930278 | |
| 4 | ERR260506 | 457387 | species | 0.031264 | Bacteroides sp. 1_1_30 | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | s__Bacteroides sp. 1_1_30__457387 | 0.99919349155781 | | |
| 5 | ERR260506 | 657309 | strain | 0.013509 | Bacteroides xylanisolvens XB1A | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | s__Bacteroides xylanisolvens__371601 | st__Bacteroides xylanisolvens XB1A__657309 | 0.93878075936513 | |
| 6 | ERR260506 | 816 | genus | 0.014524 | Bacteroides | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | 0.087832007470537 | | | |

# Sunburst Structure – Approach 1

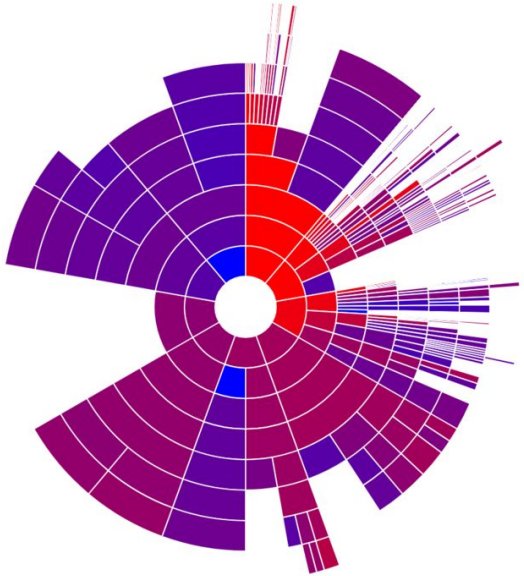
- Sample: **ERR260506_Diarrhea**
- The highlighted area on the sunburst[**on the right**] corresponds to the relative abundance.
- Approach: Area/Size of individual arc is determined by the relative_abundance value of that arc.
- Pitfall:
  - Organism that has relatively low abundance will not be seen very likely.
- Ignore Colors




| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | run_id | ncbi_taxon_id | taxon_rank_level | relative_abundance | name | lineage | CDF | | | | | | | | |
| 2 | ERR260506 | 817 | species | 0.205231 | Bacteroides fragilis | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | s__Bacteroides fragilis__817 | 0.989406697423317 | | |
| 3 | ERR260506 | 411901 | strain | 0.06312 | Bacteroides caccae ATCC 43185 | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | s__Bacteroides caccae__47678 | st__Bacteroides caccae ATCC 43185__411901 | 0.989300109930278 | |
| 4 | ERR260506 | 457387 | species | 0.031264 | Bacteroides sp. 1_1_30 | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | s__Bacteroides sp. 1_1_30__457387 | 0.99919349155781 | | |
| 5 | ERR260506 | 657309 | strain | 0.013509 | Bacteroides xylanisolvens XB1A | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | s__Bacteroides xylanisolvens__371601 | st__Bacteroides xylanisolvens XB1A__657309 | 0.93878075936513 | |
| 6 | ERR260506 | 816 | genus | 0.014524 | Bacteroides | sk__Bacteria__2 | p__Bacteroidota__976 | c__Bacteroidia__200643 | o__Bacteroidales__171549 | f__Bacteroidaceae__815 | g__Bacteroides__816 | 0.087832007470537 | | | |

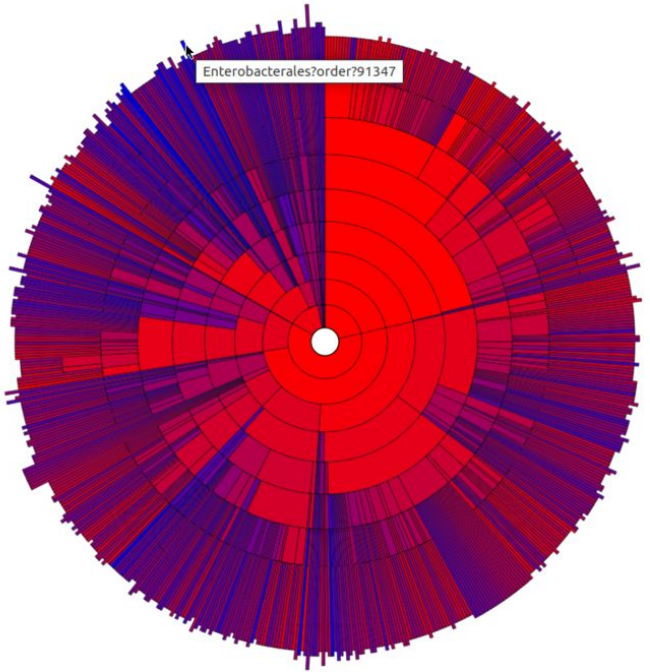
# Sunburst Structure – Approach 2

- Approach: Area/Size of individual arc is determined by:
  - How much area it's parent covers
  - How many siblings arc has
- Pitfall
  - Arc with many children may not get enough space to represent all its children.
- Ignore Colors

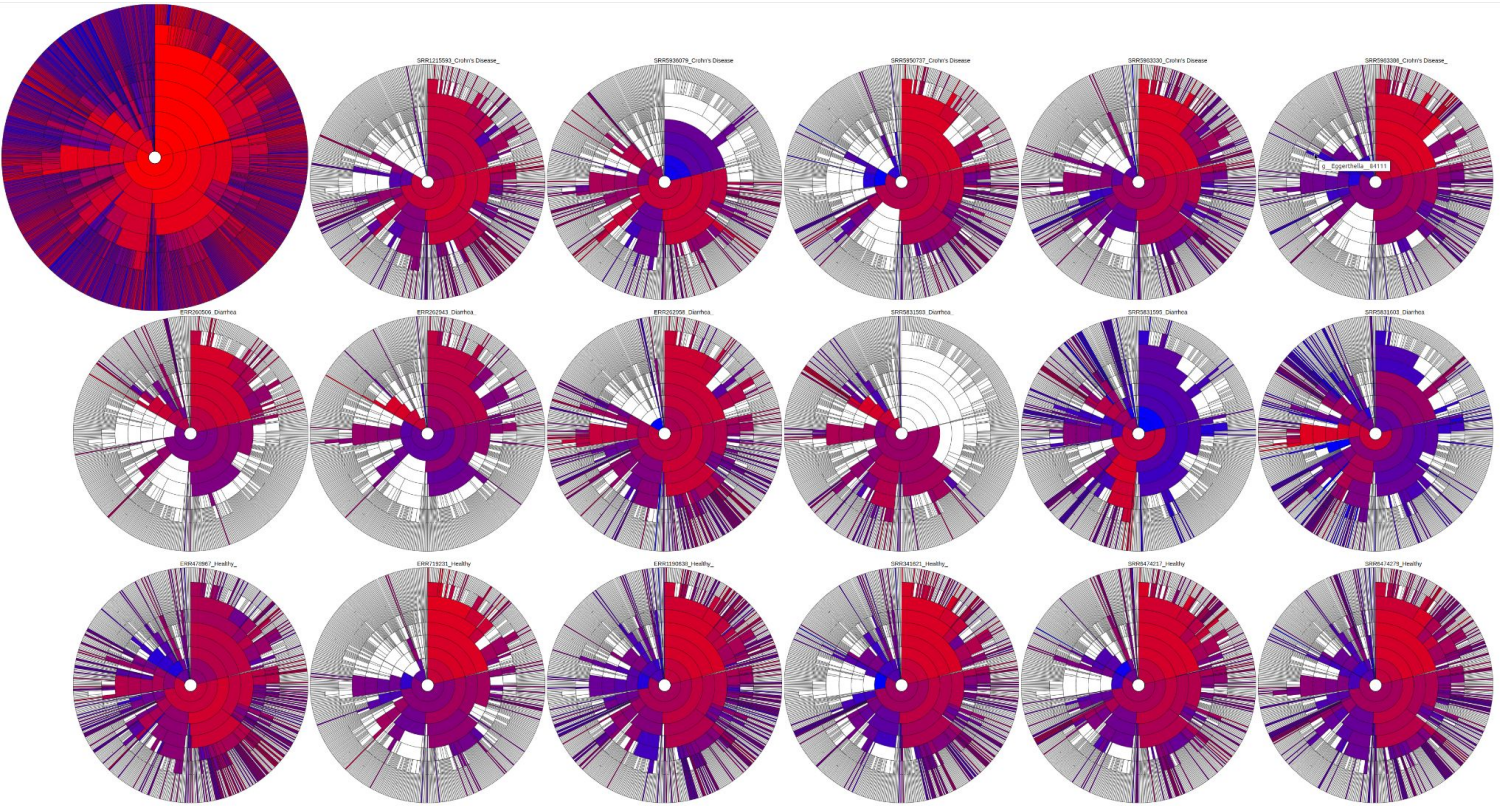


# Sunburst Structure – Approach 3

- Approach: Area/Size of individual arc is determined by:
  - How many children arc has
- If a leaf is found early in the hierarchy, it is pushed outward, therefore the jagged edges.
- Pitfall:
  - Organism with high relative abundance but with less children(not 0) will not be seen very likely.
- Ignore Colors



# Approach 3 applied to datasets



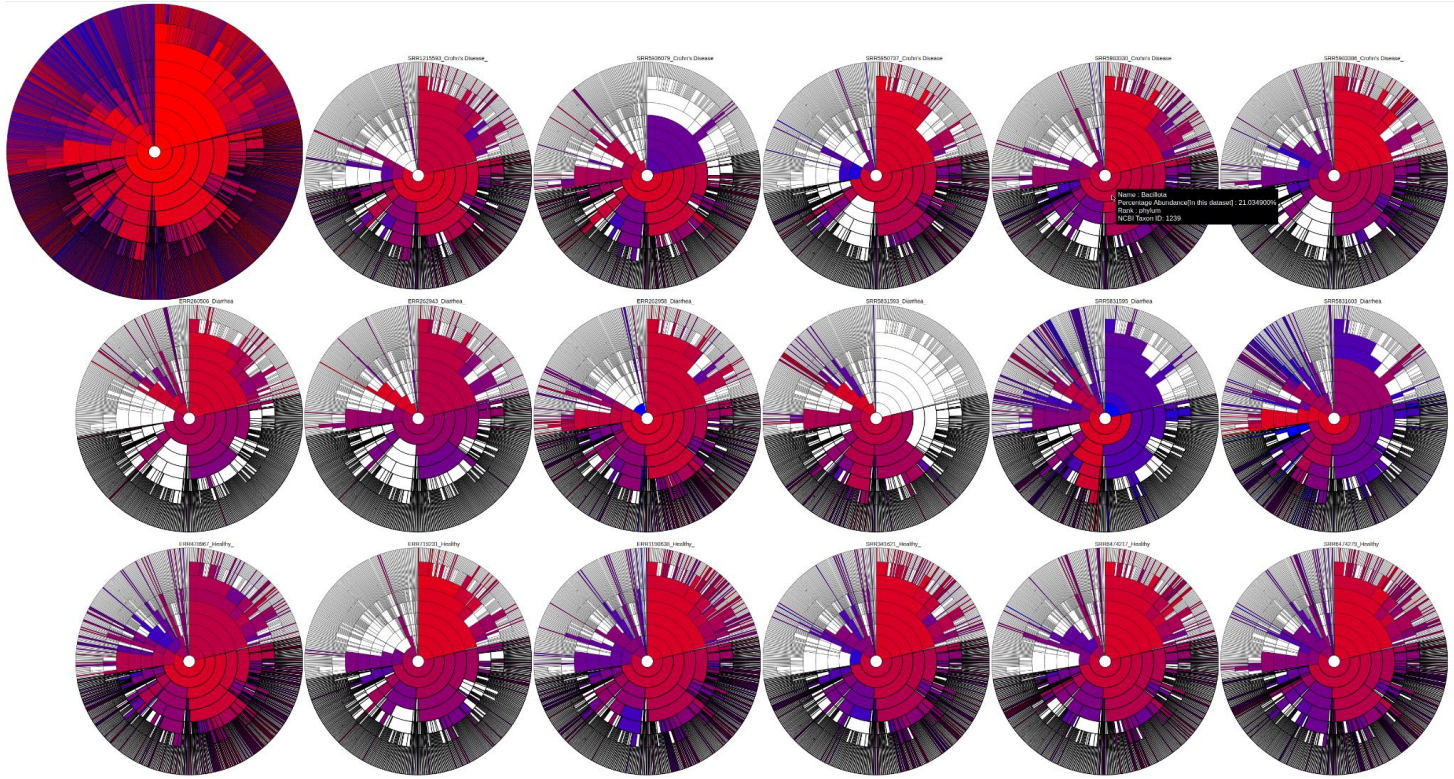
- Red = High Abundance
- Blue = Low Abundance
- White = Absence

***Crohn's disease***

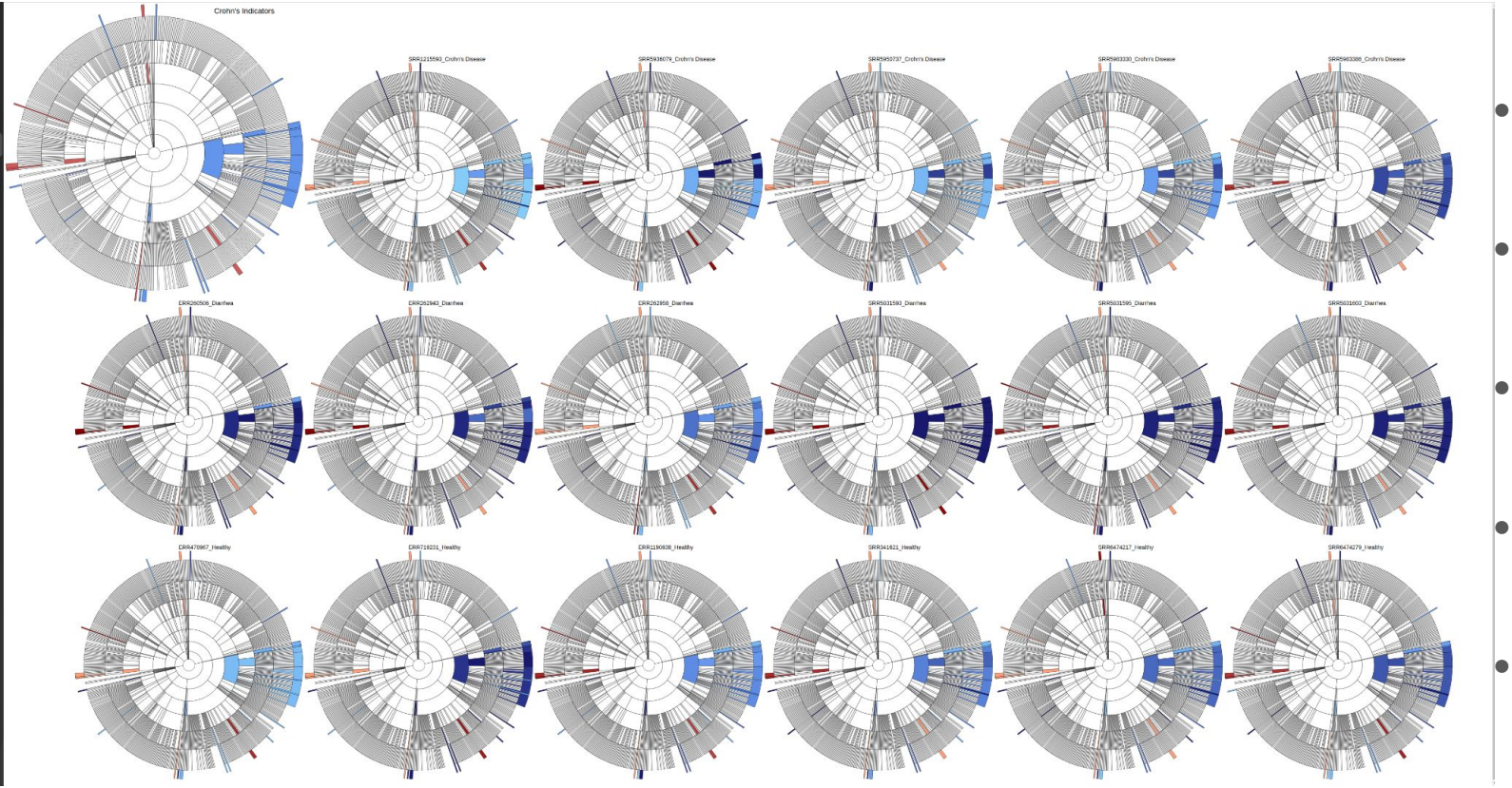
***Diarrhea***

***Healthy***

# Interaction applied to previous slide

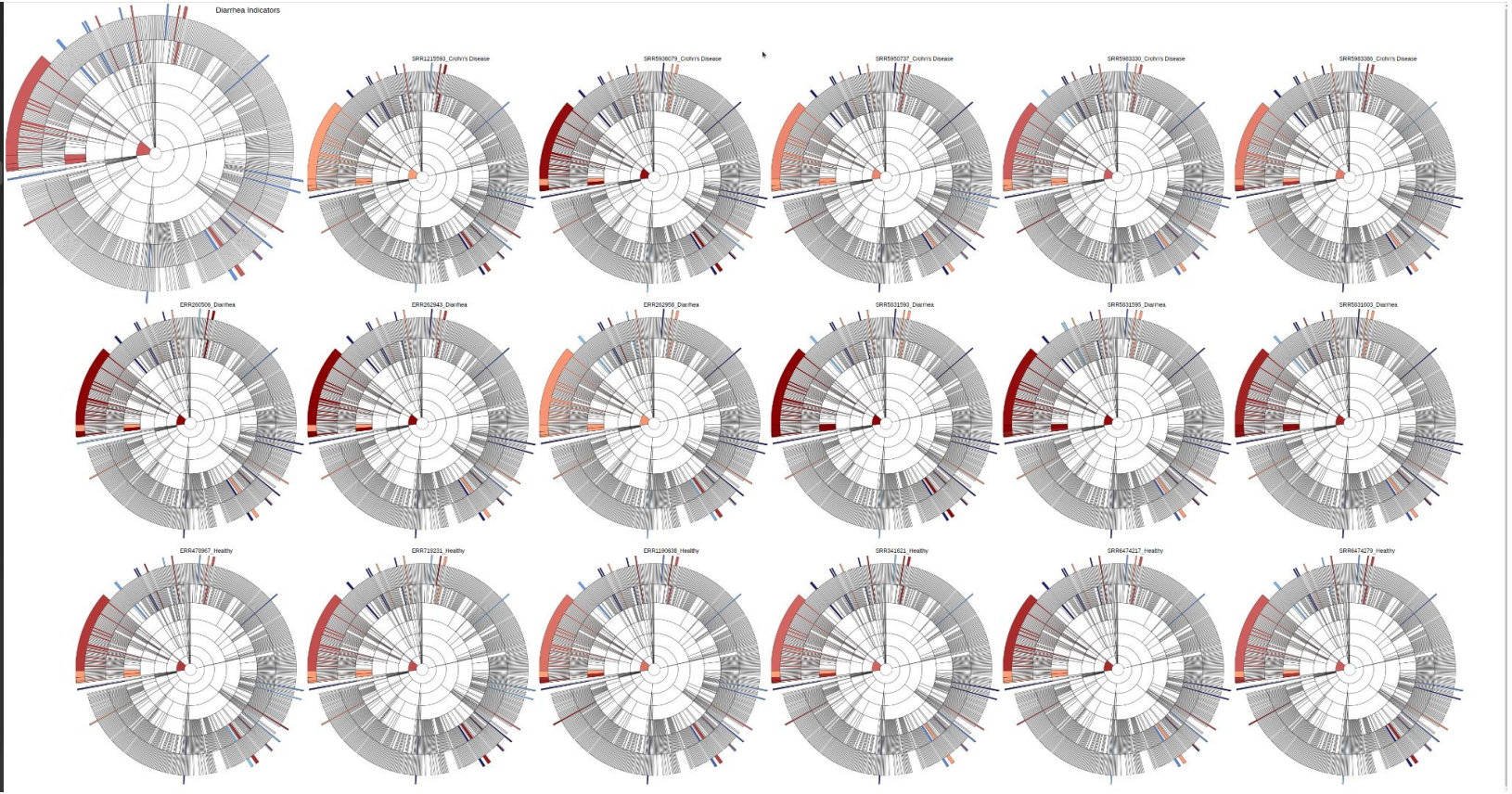


# Approach 3 applied to Crohn's indicators




- Leftmost topmost sunburst is the aggregation of the other 17 samples.
- Dark Red = High Abundance of High Abundance Organism
- Light Red = Low Abundance of High Abundance Organism
- Light Blue = High Abundance of Low Abundance Organism
- Dark Blue = Low Abundance of Low Abundance organism
- White = Non-indicator organisms

# Approach 3 applied to Diarrhea indicators



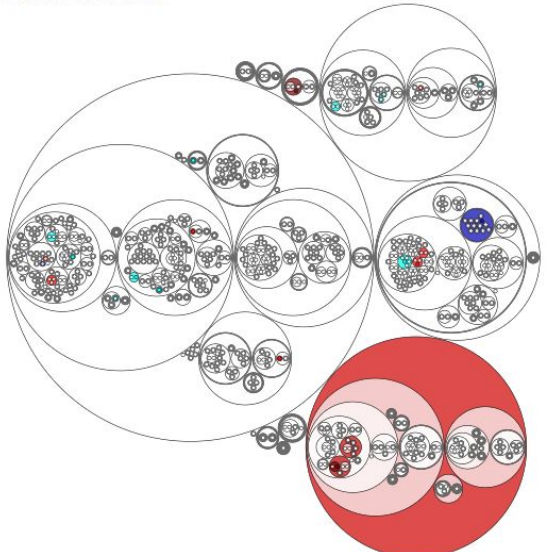

- Leftmost topmost sunburst is the aggregation of the other 17 samples.
- Dark Red = High Abundance of High Abundance Organism
- Light Red = Low Abundance of High Abundance Organism
- Light Blue = High Abundance of Low Abundance Organism
- Dark Blue = Low Abundance of Low Abundance organism
- White = Non-indicator organisms

# Abundance + Indicator Website

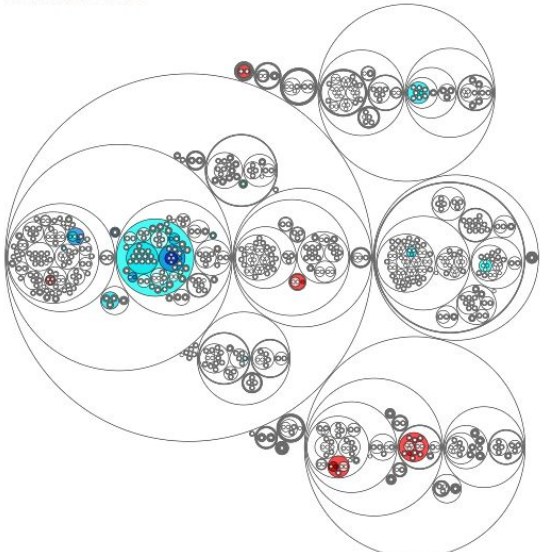
Website: https://shreeman5.github.io/MicrobiomeSunburst/sunburst.html

# Circular Treemap Indicators

DiarrheaIndicators

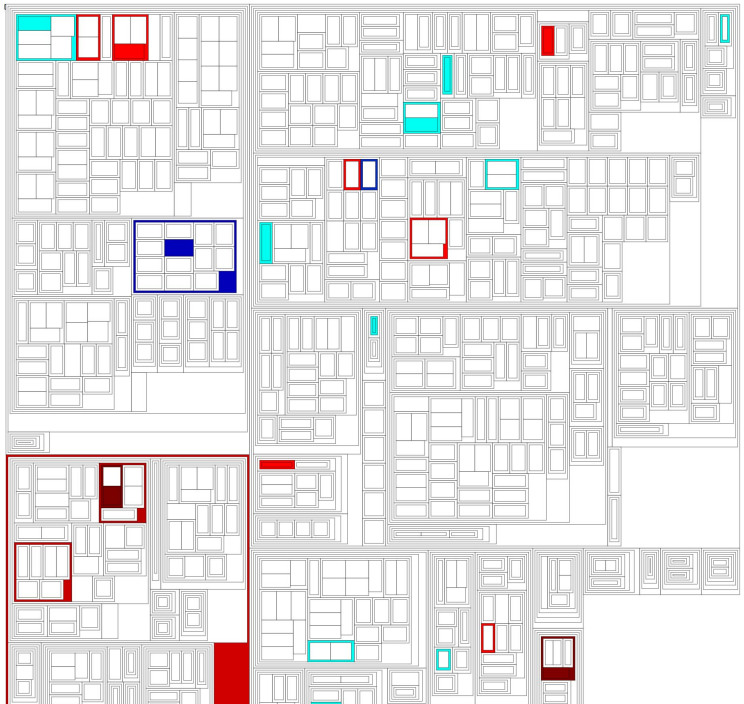


CrohnIndicators



- Dark Red = High Literature Weight of High Abundance Organism
- Light Red = Low Literature Weight of Low Abundance Organism
- Light Blue = Low Literature Weight of Low Abundance Organism
- Dark Blue = High Literature Weight of Low Abundance organism
- White = Non-indicator organisms
- Showing gallery view(like slide 18) with circular treemap was harder to parse visually.

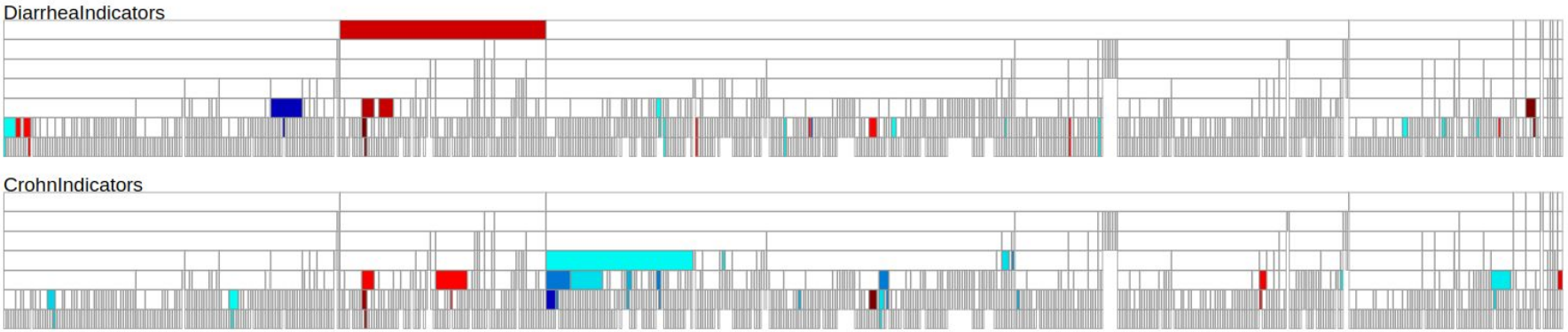
# Rectangular Treemap Indicators



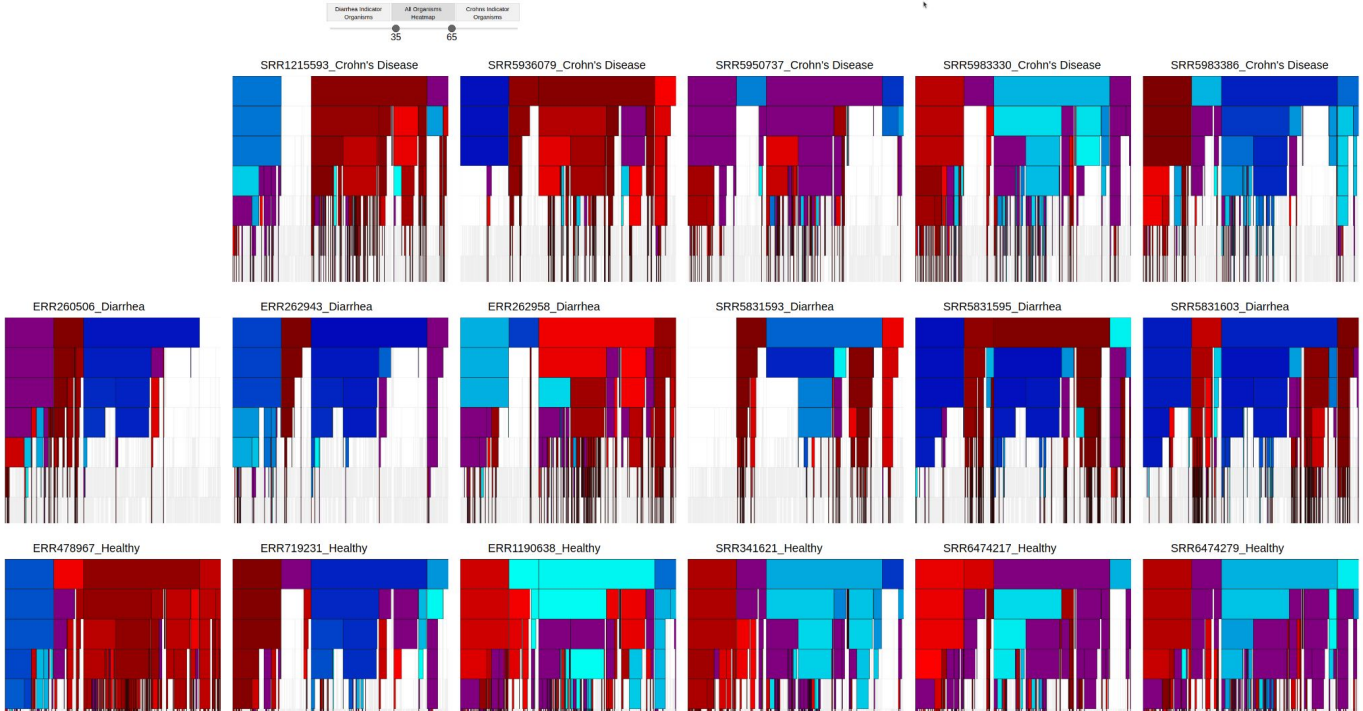
- Colors mean same thing as previous page
- Negatives:
  - A rectangular treemap will use more than double screen size of circular treemap to show the **same** information that the circular treemap is showing.
  - If we used the same dimensions that we used to make a circular treemap, rectangular treemap won't even be able to show all the indicators.
  - This is consistent with literature where we learn that treemap, icicle plot and circular treemap are good with small hierarchy(2 levels with 10-20 child nodes) and sunburst and circular treemap are good with big hierarchy(6 levels with 2-5 child nodes)
    - For this reason, no gallery view(like slide 18) using treemap.

# Icicle Plot Indicators

- Colors mean same thing as previous page
- Positives:
  - It uses **less** screen space than the circular treemap to show the same information.
  - I think it's easier to trace hierarchies in this more than circular treemap.
  - How about gallery view(like slide 18)?



# Icicle Plot Gallery View




- At the lower levels, colors start to blur.
- May not be a good fit to show gallery view.

# References


[1]:https://acris.aalto.fi/ws/portalfiles/portal/28853528/283_1_717_2_10_20180201.pdf

[2]:https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9373115&casa_token=-HI7Z8uzFBUAAAAA:KFKb12-c9gwQpcrAugCJXsnB7oWf3XzZbiAU-5Zyt6cA23mw2Ci2CnOR9APtmeDdx8R3H9bHNA

[3]:https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9373115&casa_token=-HI7Z8uzFBUAAAAA:KFKb12-c9gwQpcrAugCJXsnB7oWf3XzZbiAU-5Zyt6cA23mw2Ci2CnOR9APtmeDdx8R3H9bHNA

[4]:https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9373115&casa_token=-HI7Z8uzFBUAAAAA:KFKb12-c9gwQpcrAugCJXsnB7oWf3XzZbiAU-5Zyt6cA23mw2Ci2CnOR9APtmeDdx8R3H9bHNA

[5]:https://arxiv.org/pdf/1908.01277

[6]:https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f56e5596b634afab4d89c56dc26b4f69a3344f67

[7]:https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-385

[8]:**https://www.sciencedirect.com/science/article/pii/S2001037023003859**

[9]:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2008.01205.x?casa_token=USq6oahVzwIAAAAA%3ADSmWQmlEn8O7Cbi6VXoVt9_Jl4jrBSMRfnispAPjsiwHGH_0065HjUQ7v-T_PTvEsA8e6zudqQ0Jugw

[10]:https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0175895&type=printable

[11]:https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6295817

[12]:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2011.01963.x?casa_token=RWRlgXnVrg0AAAAA:85depYRf-dDjKjfS4qkKkQHc_aLQ2XK2Ull6eV6a5Y_qv5IrHHnR_Kft5tB_7RLuRbCTyO5H37DcHF4