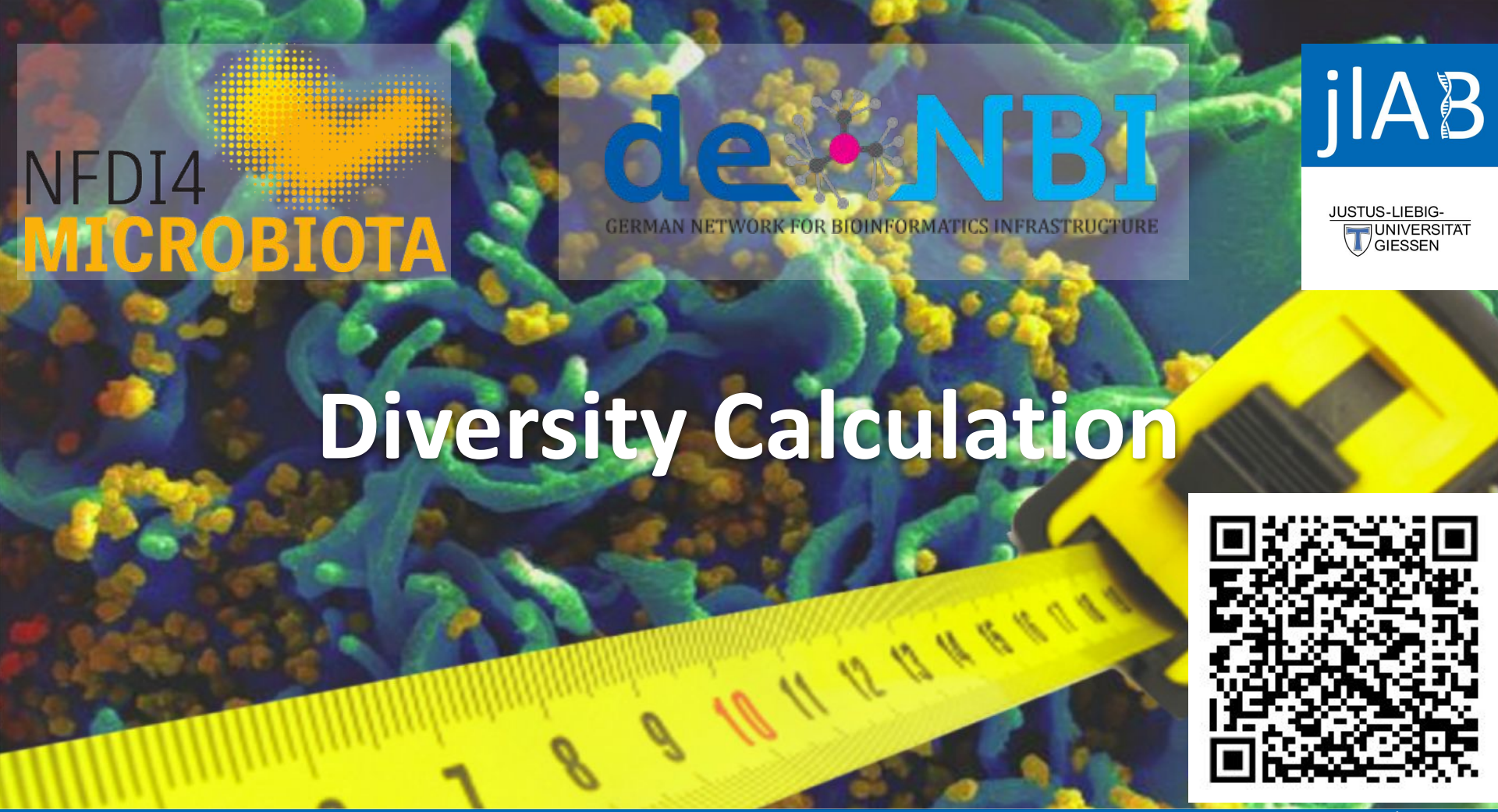
NFDI4 MICROBIOTA

de.NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

jlAB

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

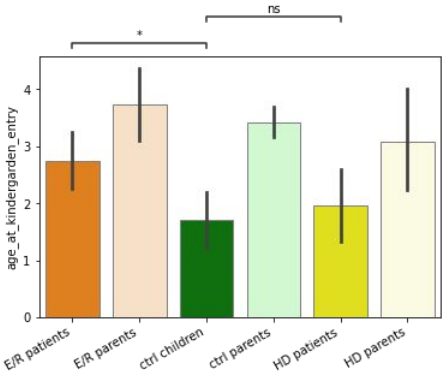
Diversity Calculation



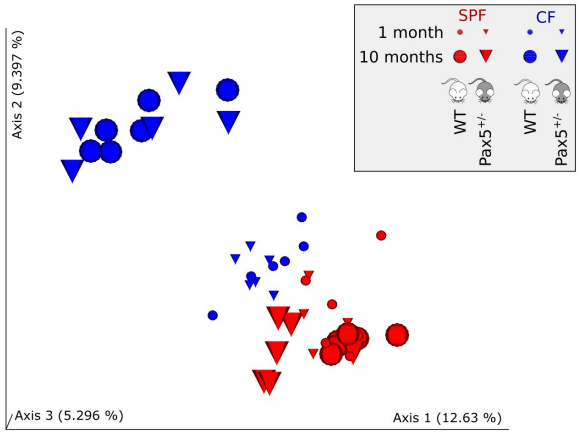
	Tuesday, Oct 8th	Wednesday, Oct 9th	Thursday, Oct 10th
10:00-10:30		QIIME2 and Qiita	Buffer for Hands-On, Coffee break in-between
10:30-11:00		Qiita Hands-On	
11:00-11:30	Welcome & Intro		
11:30-12:00	From raw data to feature tables	Coffee Break (~15min)	Coffee Break (~15min)
12:00-12:30		Qiita Hands-On	Buffer for Hands-On
12:30-13:00		Lunch Break	Closing remarks and Farewell
13:00-13:30	Lunch Break		
13:30-14:00		QIIME2 Hands-On	
14:00-14:30	Sequence quality control		
14:30-15:00	Coffee Break	Coffee Break	
15:00-15:30	QC Hands-on	Diversity Calculation	
15:30-16:00		QIIME2 Hands-On (cont'd.)	
16:00-16:30			

Contents

1. Alpha Diversity
2. Beta Diversity
3. Dimensionality Reduction (PCoA)



differences in diversity

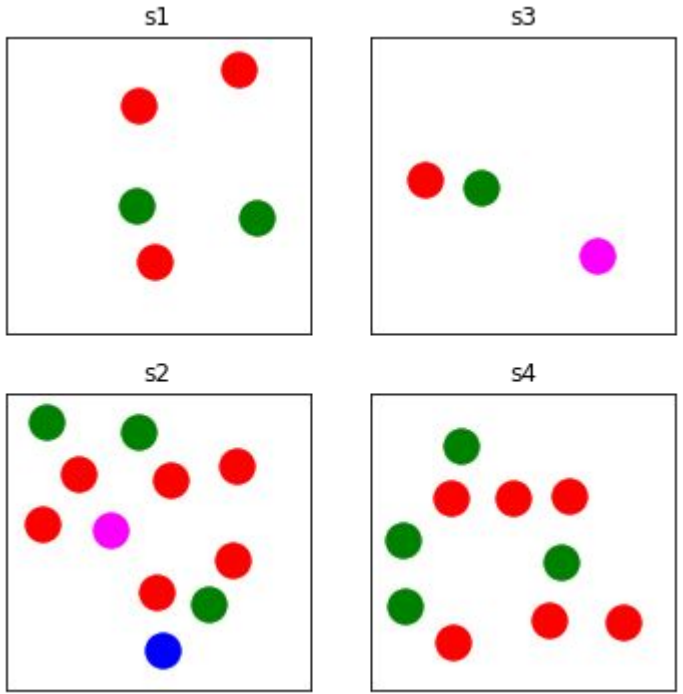


1. Alpha Diversity

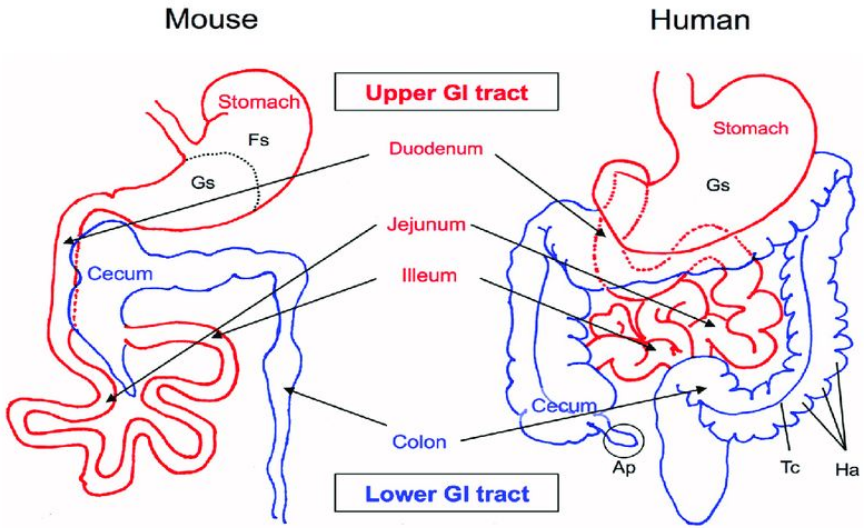
a. observed features
b. Chao1
c. Faith's PD

2. Alpha Diversity

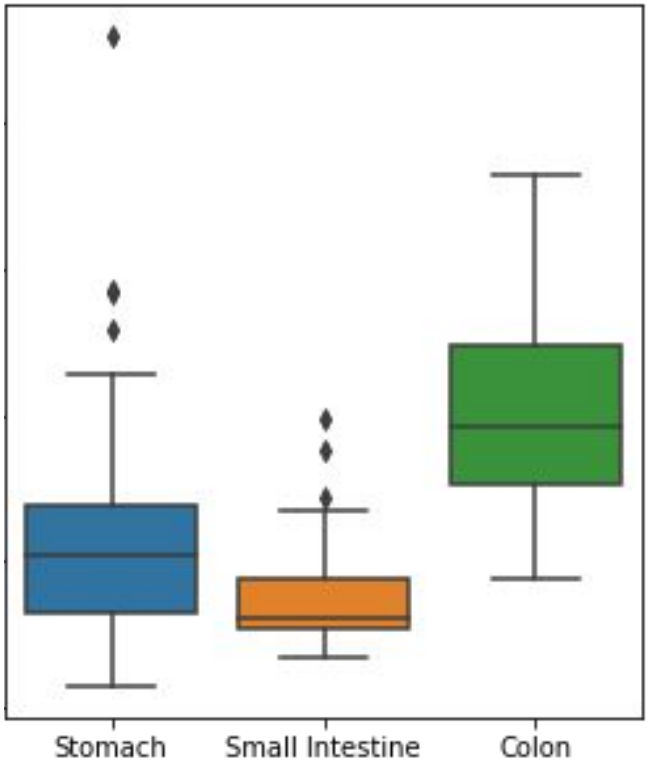
- one sample at a time
- "how complex is one sample?"
- alternative perspectives
  a. richness: observed_features
  b. richness: Chao1
  c. evenness: Shannon
  d. phylogenetic diversity: Faith's PD



Alpha Diversity: mouse gut microbiome

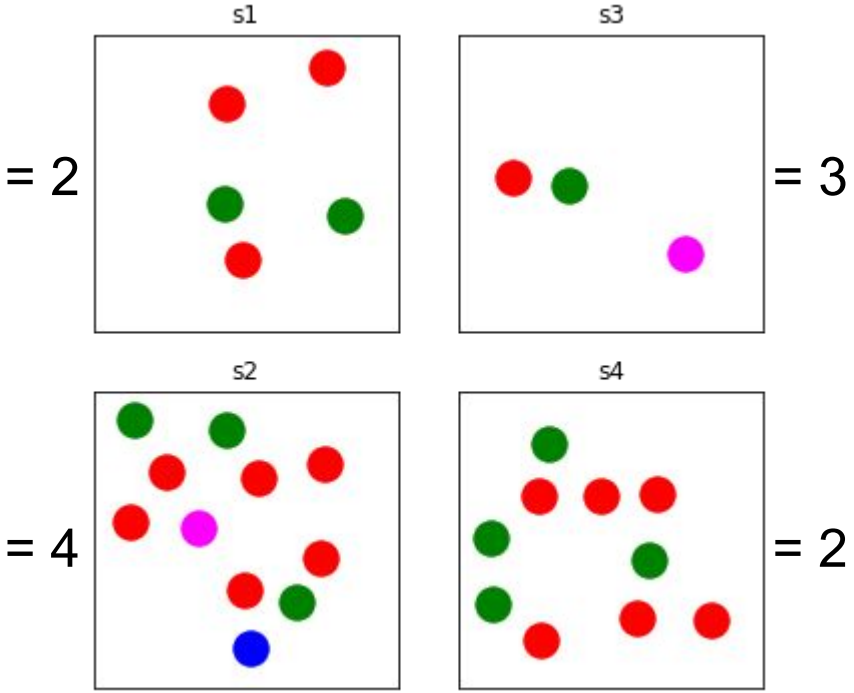


Nishiyama et al. (2016) *microorganisms*



2. Alpha Diversity - a: observed features

- richness
- simple:
  count number of features



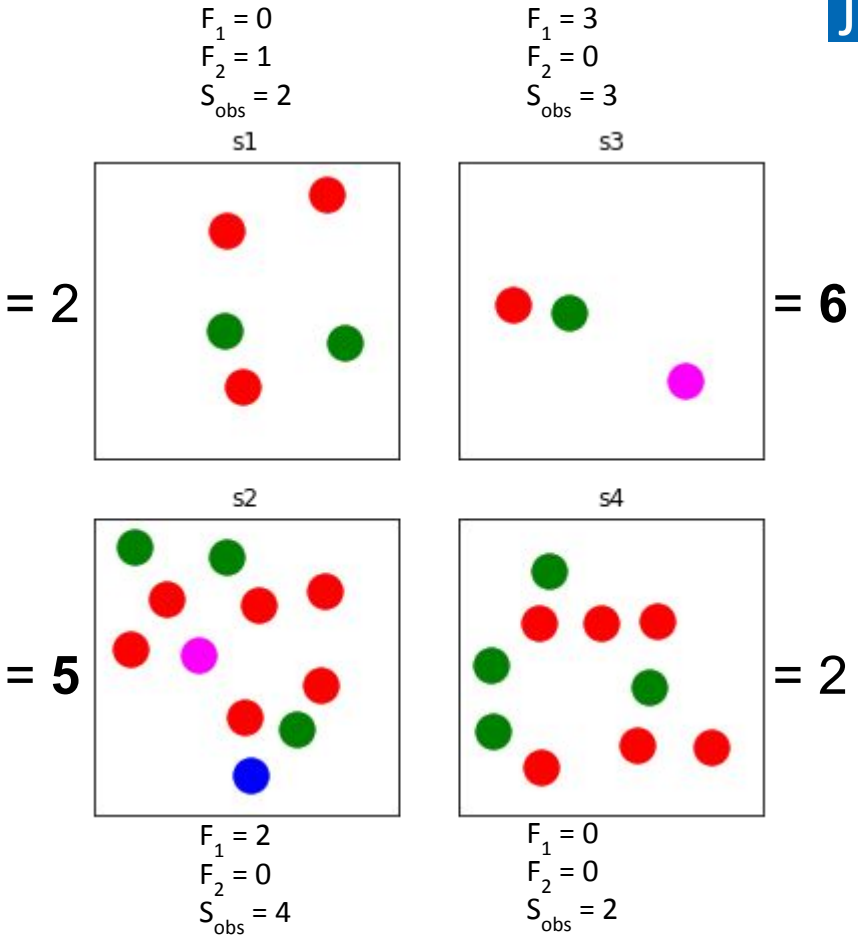
2. Alpha Diversity - b: Chao1

- richness
- simple:
  count number of features
- + emphasize singletons

$$chao1 = S_{obs} + \frac{F_1(F_1 - 1)}{2(F_2 + 1)}$$

Anne Chao

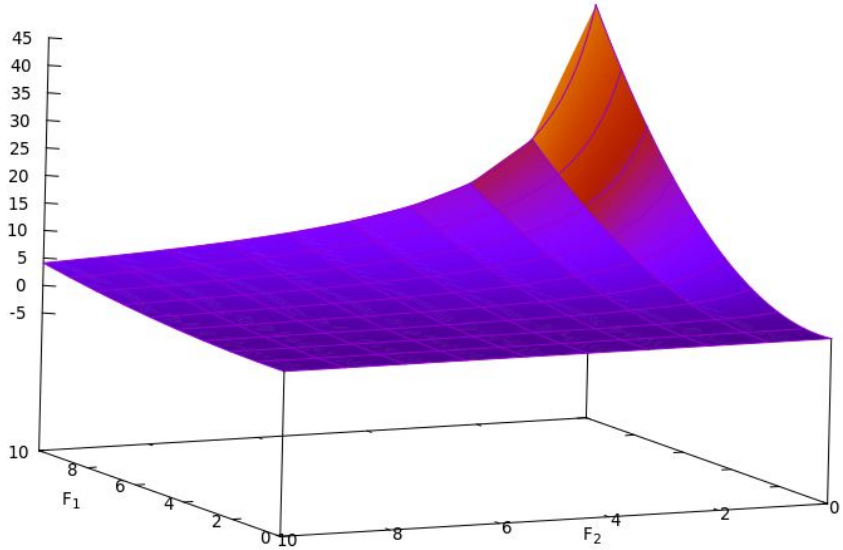
F_1 = features with count 1
F_2 = features with count 2
S_{obs} = observed features



2. Alpha Diversity - b: Chao1

- richness
- simple:
  count number of features
- + emphasize singletons

"when most of the information is concentrated on the low order occupancy numbers"



$$chao1 = S_{obs} + \frac{F_1(F_1 - 1)}{2(F_2 + 1)}$$

Anne Chao

F_1 = features with count 1
F_2 = features with count 2
S_{obs} = observed features

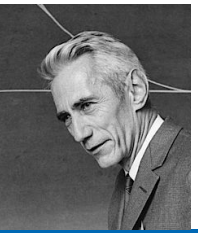
Anne Chao (1984) *Scand J Statist* " Nonparametric Estimation of the Number of Classes in a Population"

2. Alpha Diversity - c: Shannon

- evenness

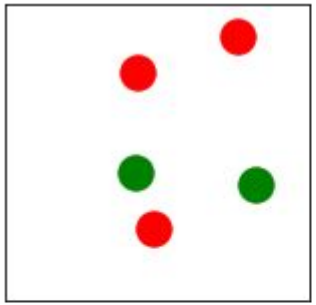
$$H = -\sum_{i=1}^{s} (p_i \log_2 p_i)$$

Claude Shannon



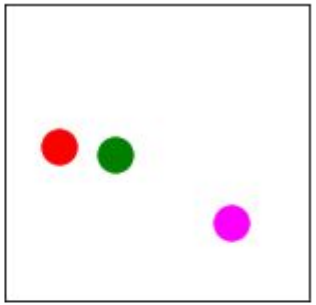
p_{red} = 3/5
p_{green} = 2/5

= 0.970951

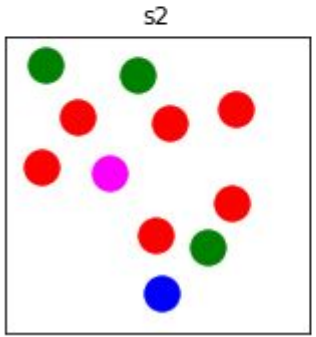


p_{red} = 1/3 $p_{magenta}$ = 1/3
p_{green} = 1/3

= 1.584963

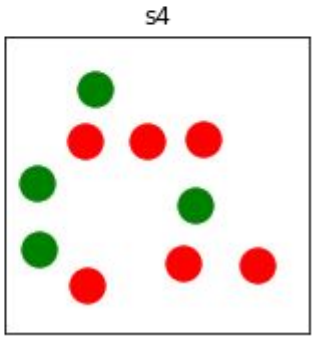


= 1.617190



p_{red} = 6/11 $p_{magenta}$ = 1/11
p_{green} = 3/11 p_{blue} = 1/11

= 0.970951

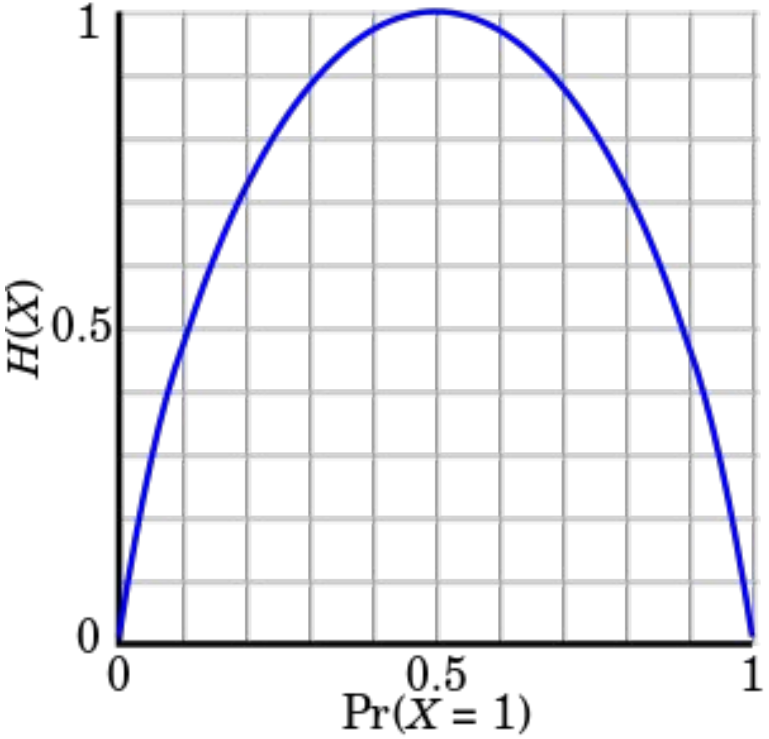


p_{red} = 6/10
p_{green} = 4/10

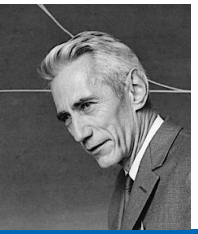
2. Alpha Diversity - c: Shannon

- evenness

$$H = -\sum_{i=1}^{s} (p_i \log_2 p_i)$$

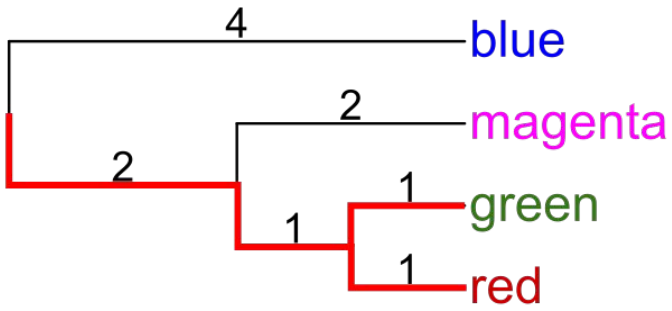


Claude Shannon



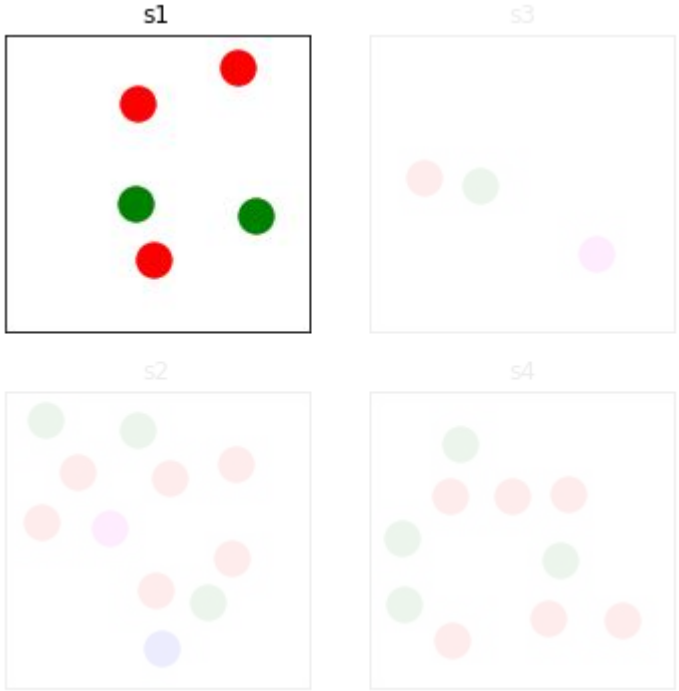
2. Alpha Diversity - d: Faith's PD

- captured evolutionary diversity
- needs a phylogenetic tree!



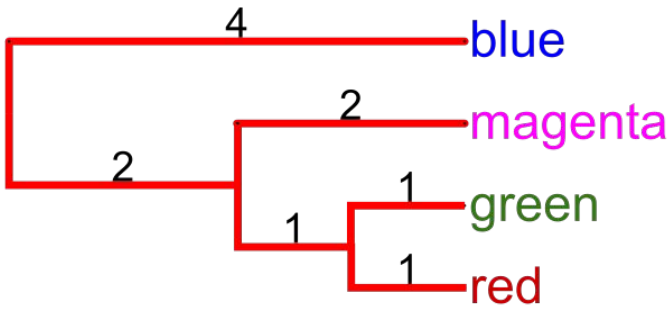
Daniel Faith

= 5

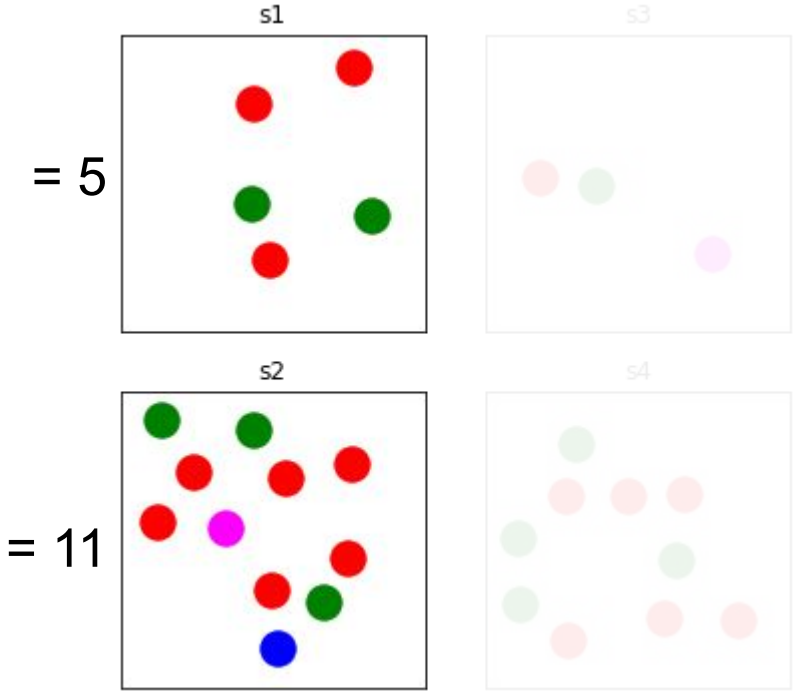


2. Alpha Diversity - d: Faith's PD

- captured evolutionary diversity
- needs a phylogenetic tree!

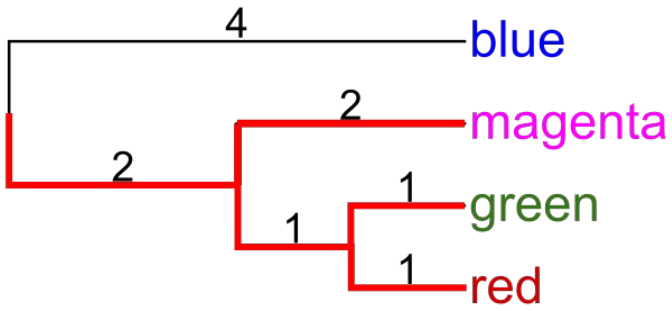


Daniel Faith

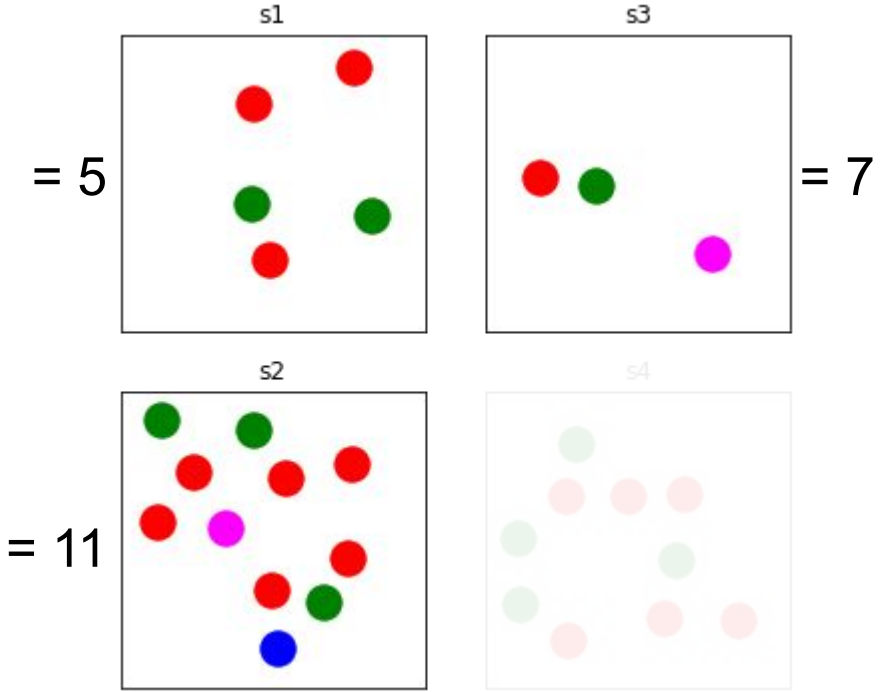


2. Alpha Diversity - d: Faith's PD

- captured evolutionary diversity
- needs a phylogenetic tree!

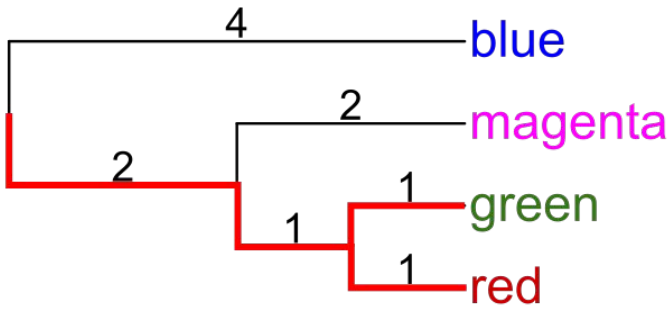


Daniel Faith

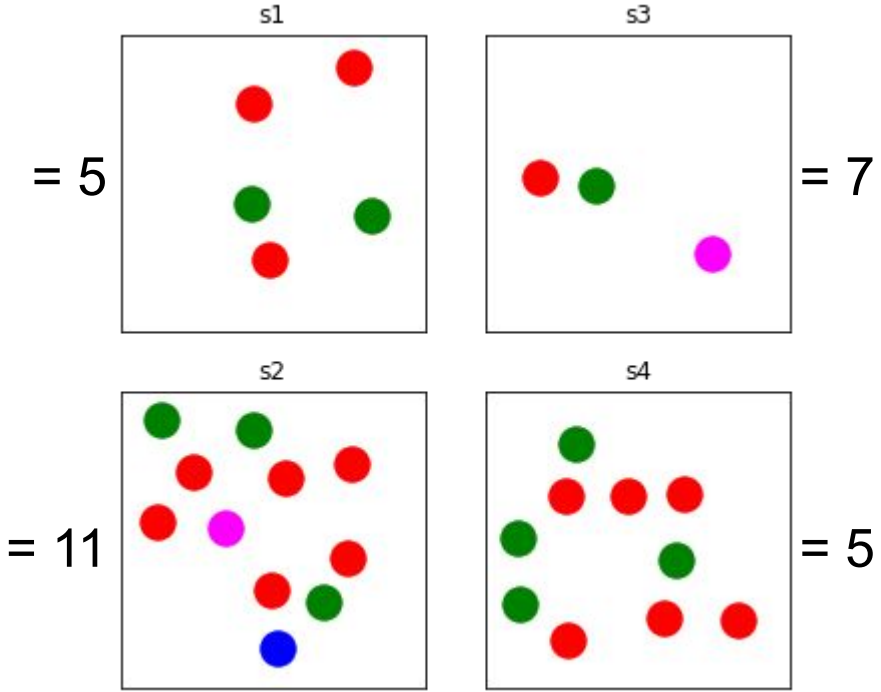


2. Alpha Diversity - d: Faith's PD

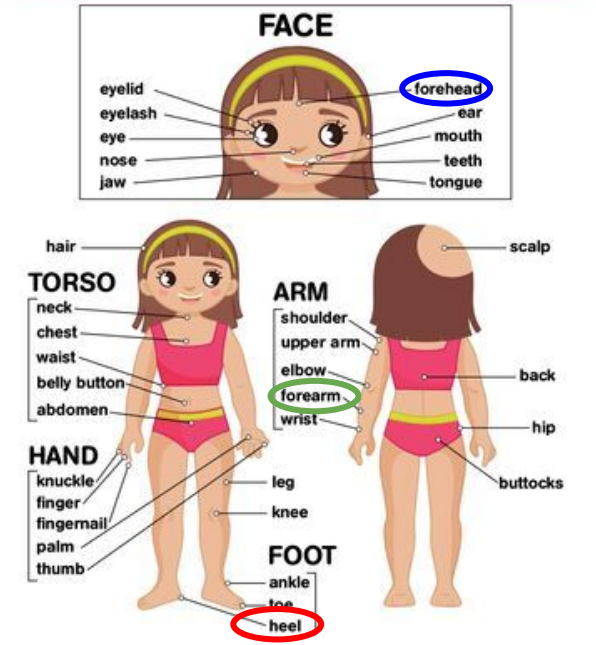
- captured evolutionary diversity
- needs a phylogenetic tree!



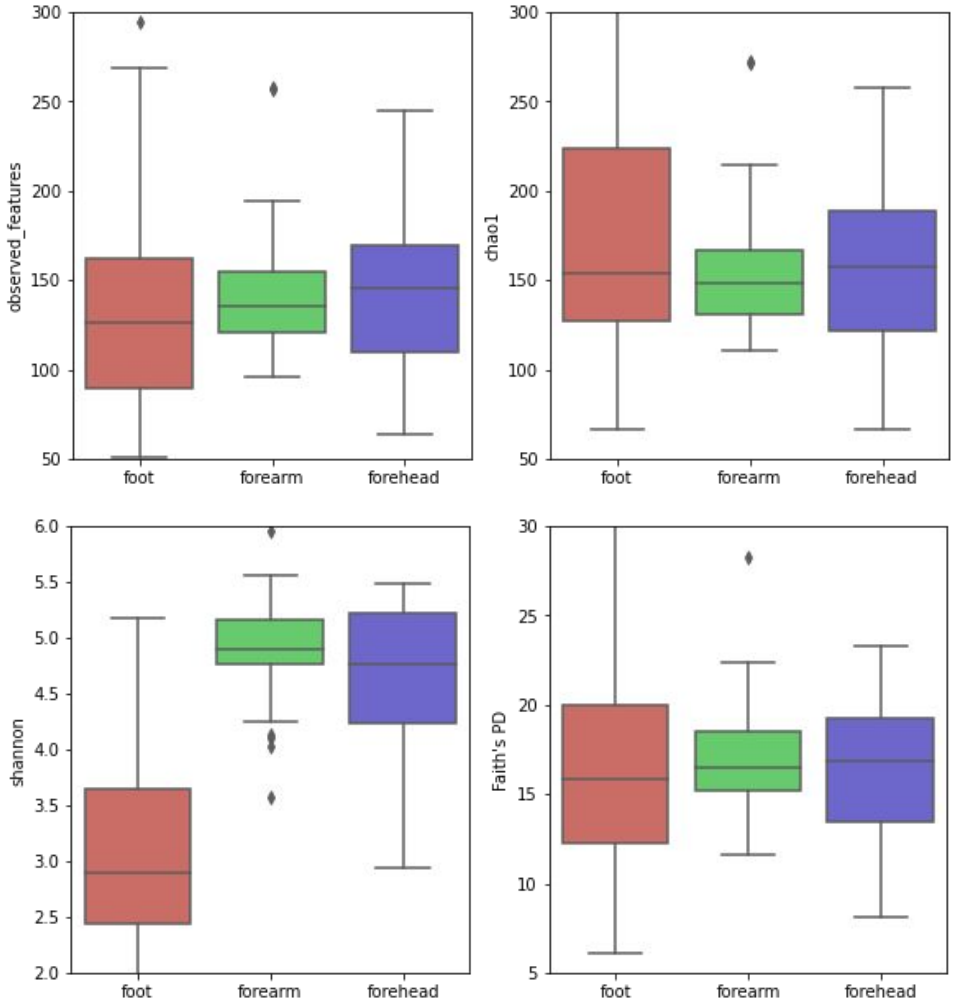
Daniel Faith



Alpha Diversity: skin microbiome



https://reference.yourdictionary.com/resources/parts-of-the-body-for-kids-names-and-basic-functions.html

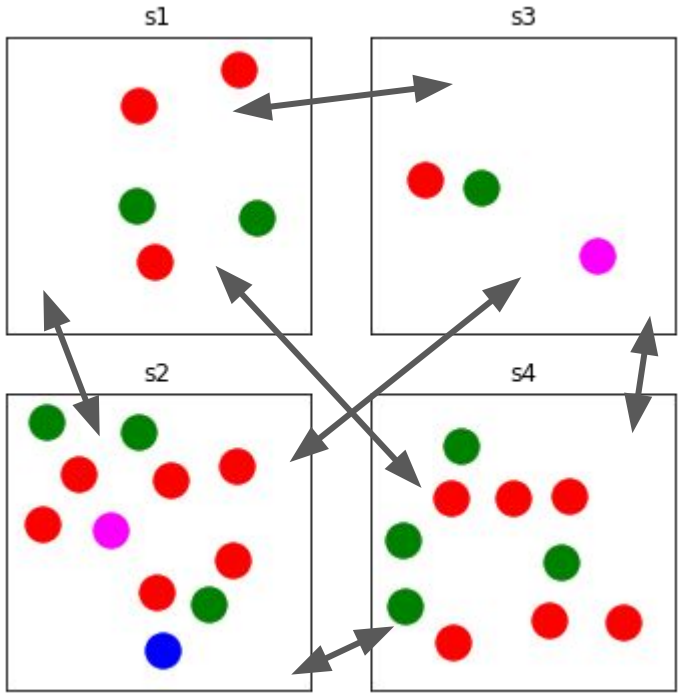


3. Beta Diversity

a. Jaccard
b. Bray-Curtis
c. UniFrac

3. Beta Diversity

- pair of samples at a time
- "how different are **two** samples?"
- alternative perspectives
  a. Jaccard
  b. Bray-Curtis
  c. UniFrac
- a "Distance"



Distance Axioms

For all x_i, x_j and x_k

1. $d(x_i, x_j) \geq 0.$
2. $d(x_i, x_j) = 0$, iff x_i is equal to x_j.
3. $d(x_i, x_j) = d(x_j, x_i).$
4. $d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j).$

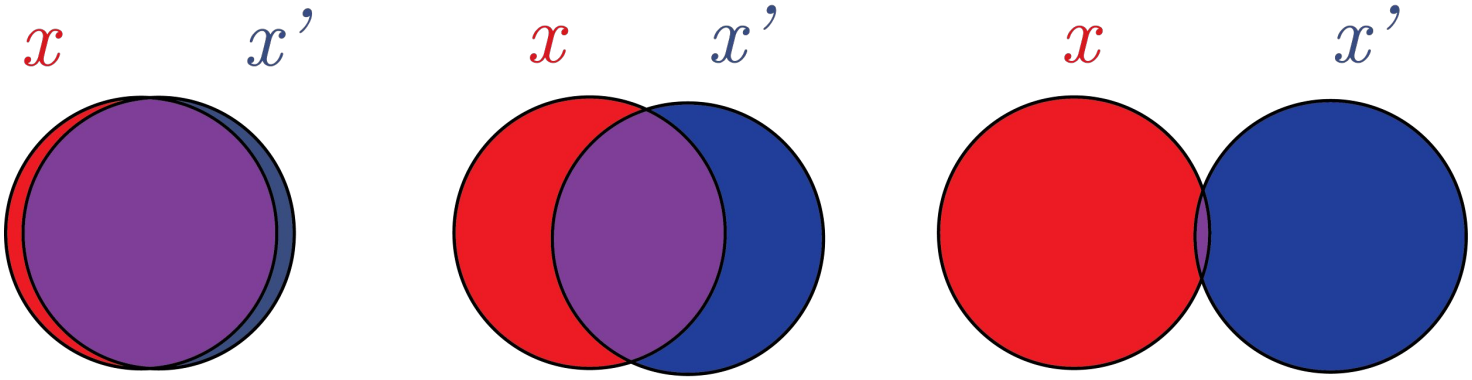
Distance Axioms

For all x_i, x_j and x_k

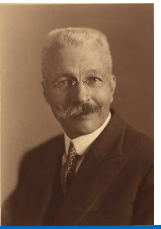
1. $d(x_i, x_j) \geq 0.$ **non-negativity**
2. $d(x_i, x_j) = 0$, iff x_i is equal to x_j. **identity**
3. $d(x_i, x_j) = d(x_j, x_i).$ **symmetry**
4. $d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j).$ **no wormholes**

3. Beta Diversity - a: Jaccard

Fraction of unique features,
regardless of abundance.



Paul Jaccard



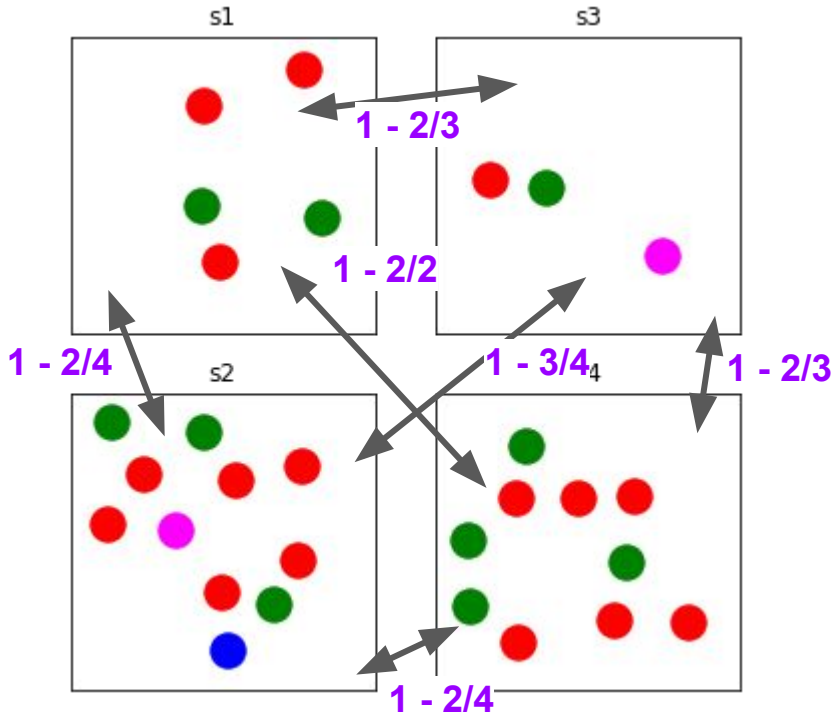
$$d_j \approx 0 \qquad d_j \approx 0.5 \qquad d_j \approx 1$$

📄 Jaccard, Paul. *Nouvelles recherches sur la distribution florale*. 1908.

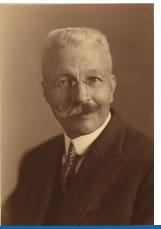
3. Beta Diversity - a: Jaccard

Fraction of unique features,
regardless of abundance.

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$



Paul Jaccard

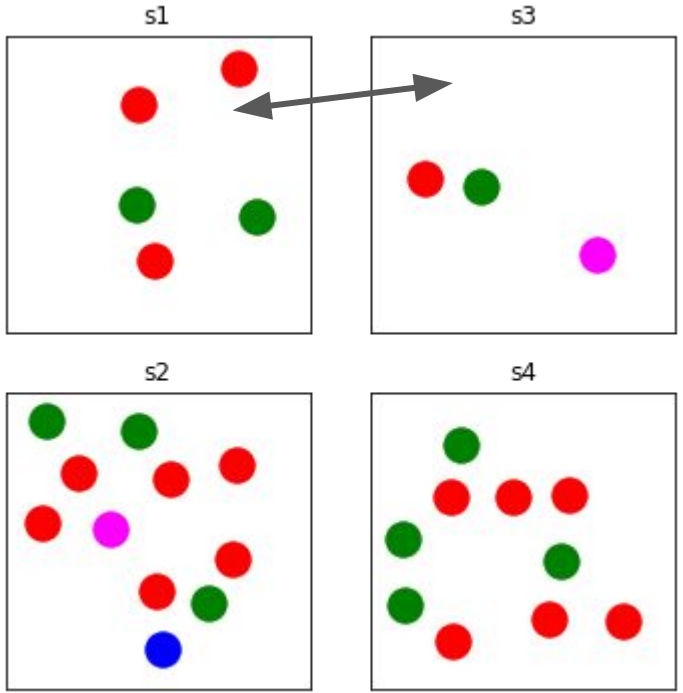


📄 Jaccard, Paul. *Nouvelles recherches sur la distribution florale*. 1908.

3. Beta Diversity - b: Bray & Curtis

compares relative abundances

$$\sum |u_i - v_i| / \sum |u_i + v_i|$$



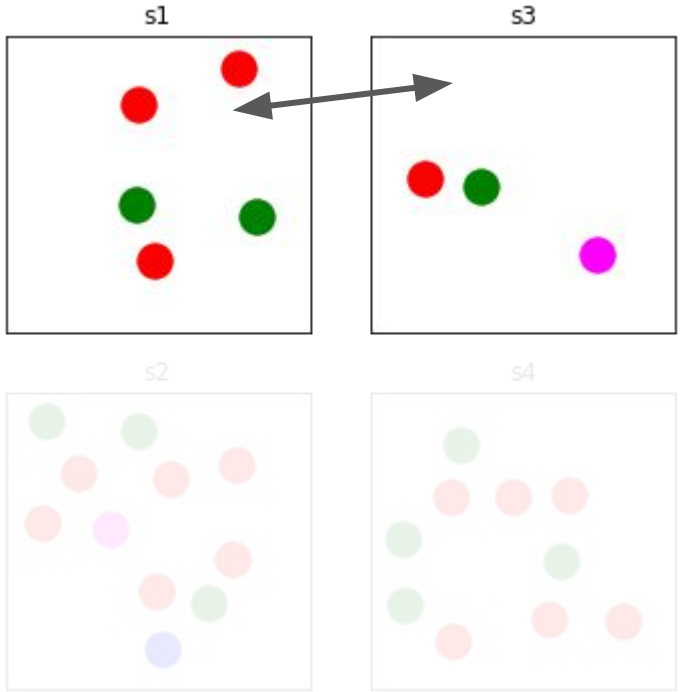
J. Roger Bray & John T. Curtis

3. Beta Diversity - b: Bray & Curtis

compares relative abundances

$$\sum |u_i - v_i| / \sum |u_i + v_i|$$

i = red, green, magenta, blue
u = s1
v = s2



J. Roger Bray & John T. Curtis

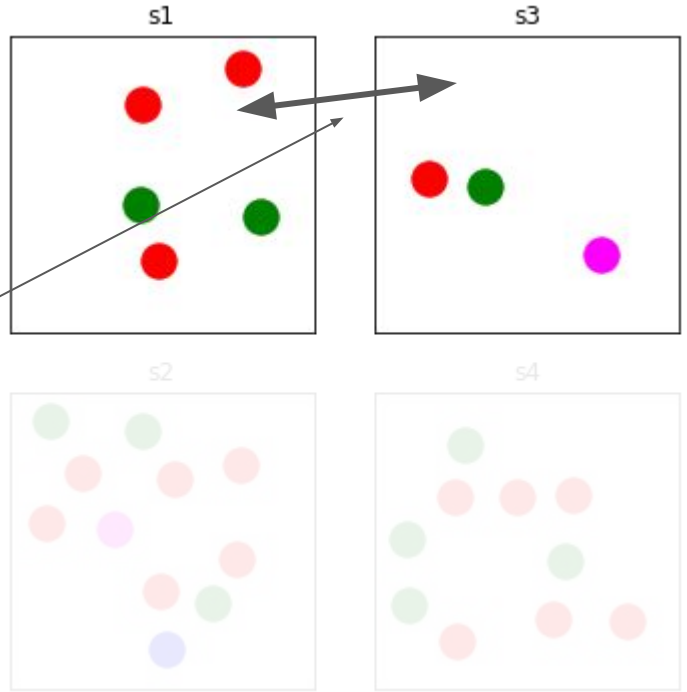
3. Beta Diversity - b: Bray & Curtis

compares relative abundances

$$\sum |u_i - v_i| / \sum |u_i + v_i|$$

i = red, green, magenta, blue
u = s1
v = s2

$$\frac{|3-1| + |2-1| + |0-1| + |0-0|}{|3+1| + |2+1| + |0+1| + |0+0|}$$



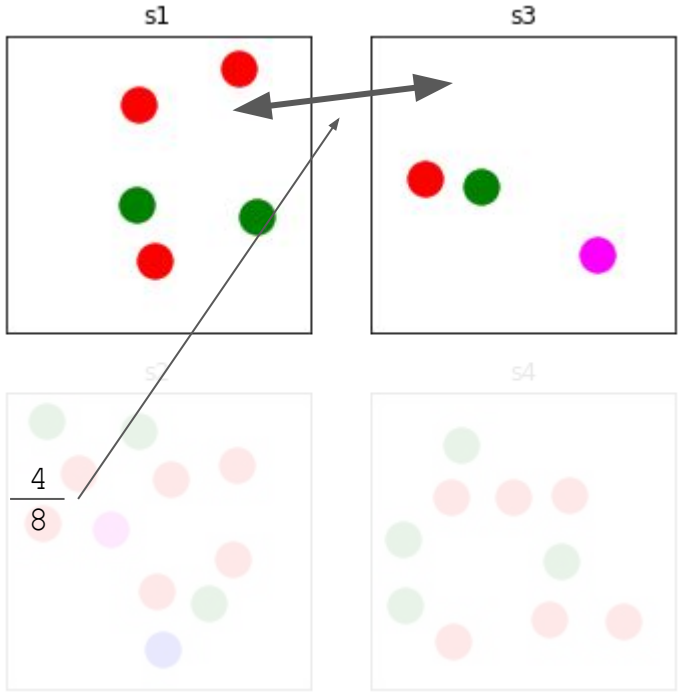
3. Beta Diversity - b: Bray & Curtis

compares relative abundances

$$\sum |u_i - v_i| / \sum |u_i + v_i|$$

i = red, green, magenta, blue
u = s1
v = s2

$$\frac{|3-1| + |2-1| + |0-1| + |0-0|}{|3+1| + |2+1| + |0+1| + |0+0|} = \frac{2 + 1 + 1 + 0}{4 + 3 + 1 + 0} = \frac{4}{8}$$



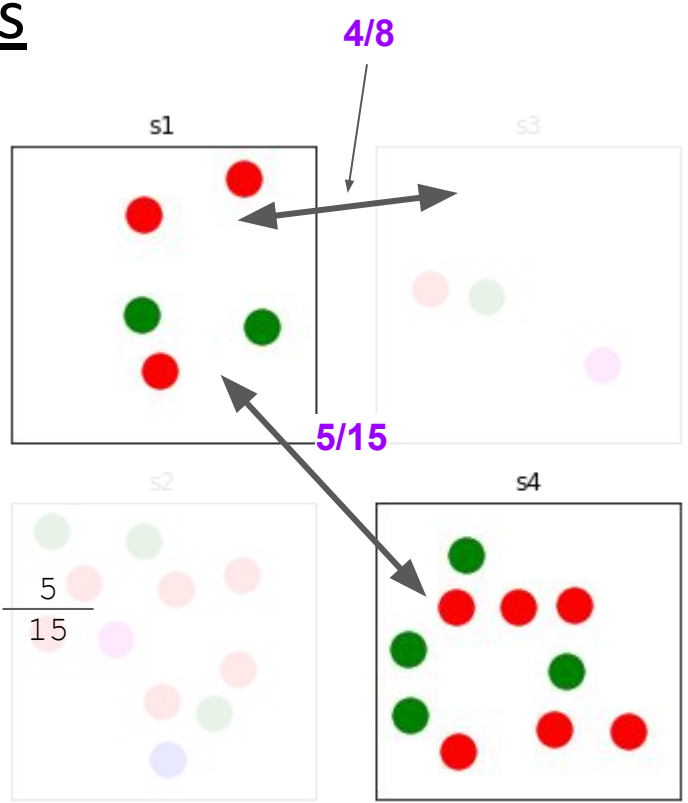
3. Beta Diversity - b: Bray & Curtis

compares relative abundances

$$\sum |u_i - v_i| / \sum |u_i + v_i|$$

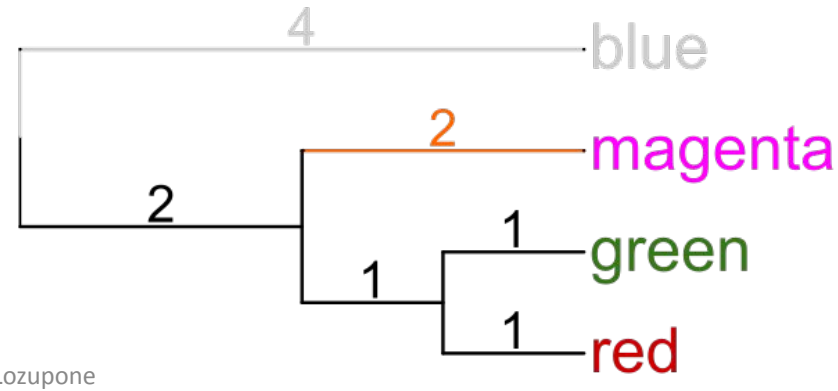
i = red, green, magenta, blue
u = s1
v = s4

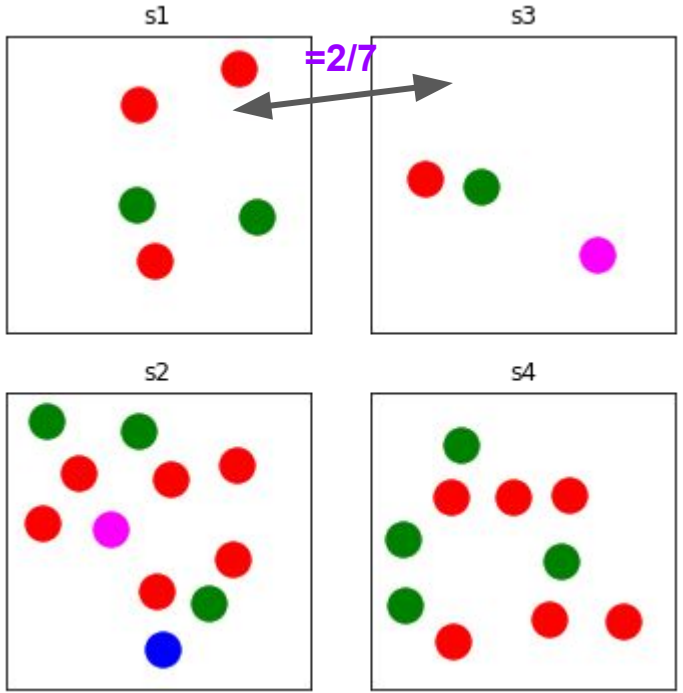
$$\frac{|3-6| + |2-4| + |0-0| + |0-0|}{|3+6| + |2+4| + |0+0| + |0+0|} = \frac{3 + 2 + 0 + 0}{9 + 6 + 0 + 0} = \frac{5}{15}$$



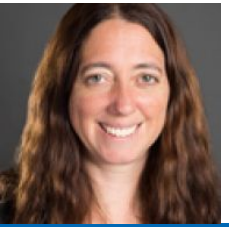
3. Beta Diversity - c: UniFrac

- Measures the fraction of unique branch length
- needs a phylogenetic tree!



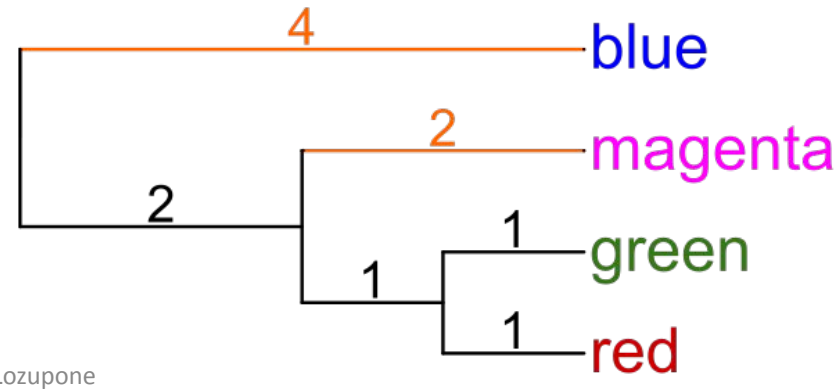


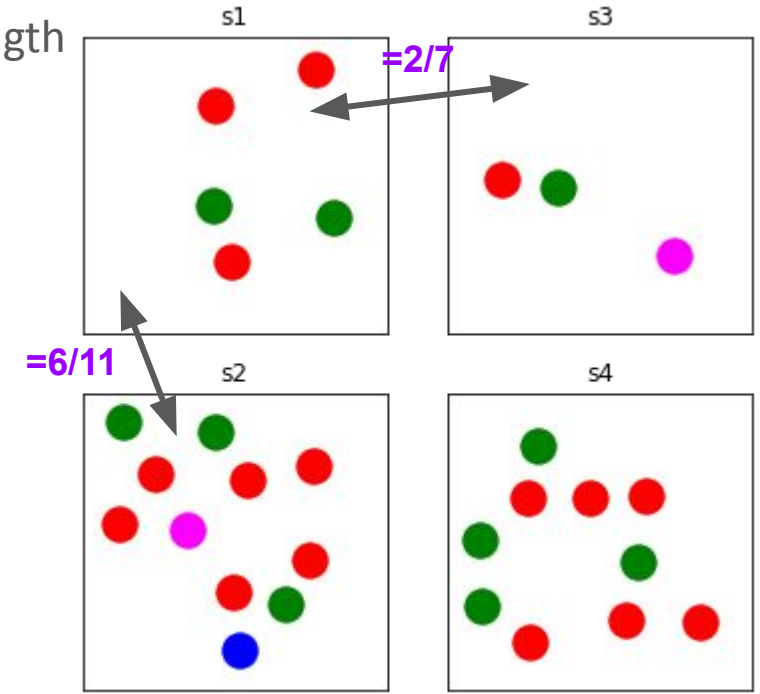
Catherine Lozupone



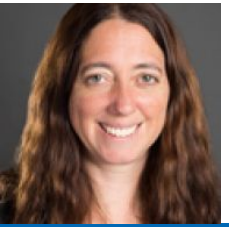
3. Beta Diversity - c: UniFrac

- Measures the fraction of unique branch length
- needs a phylogenetic tree!



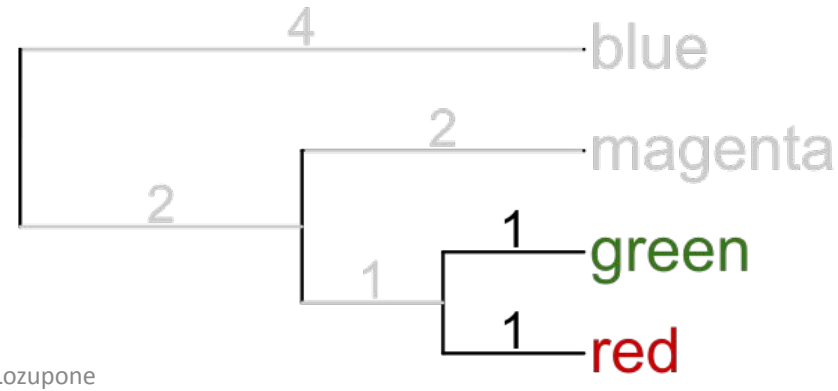


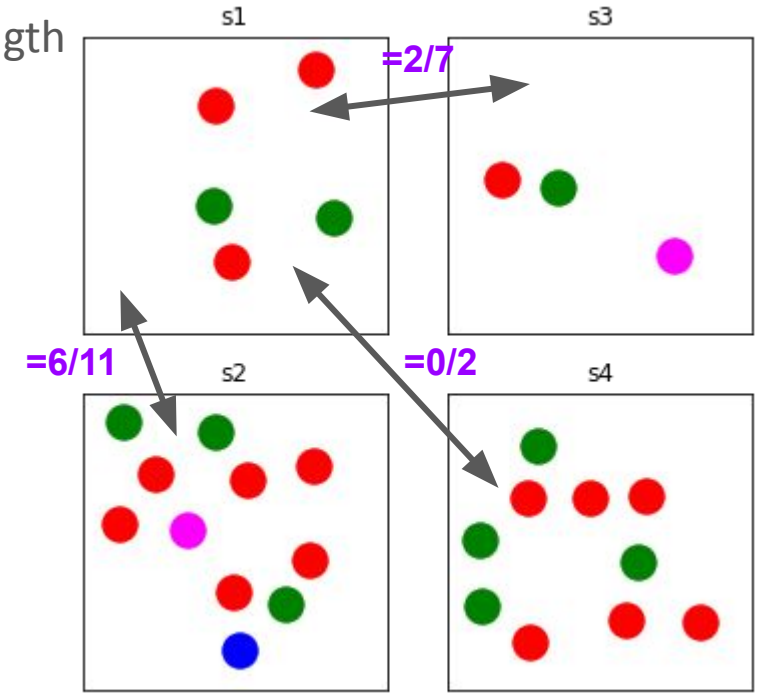
Catherine Lozupone



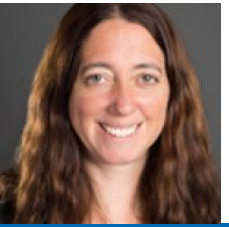
3. Beta Diversity - c: UniFrac

- Measures the fraction of unique branch length
- needs a phylogenetic tree!



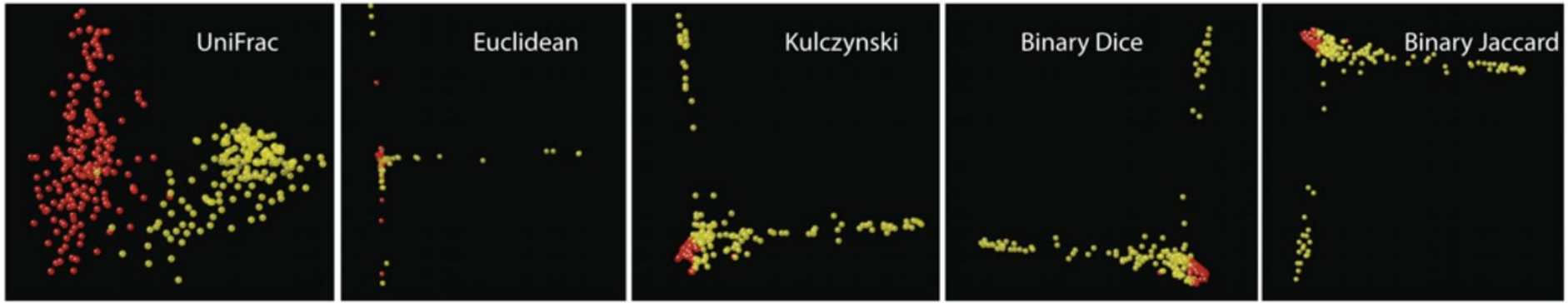


Catherine Lozupone



Which metric to use?

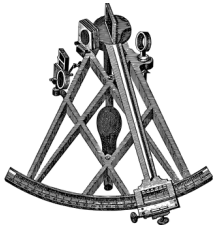
Selecting a metric

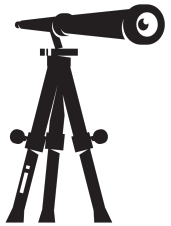


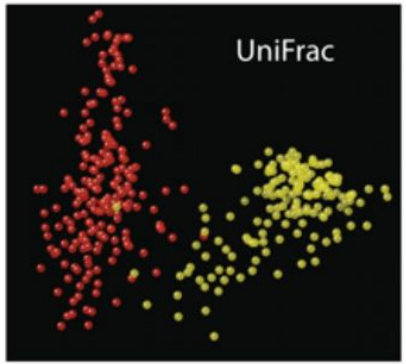
- Vertebrate Gut
- Free living

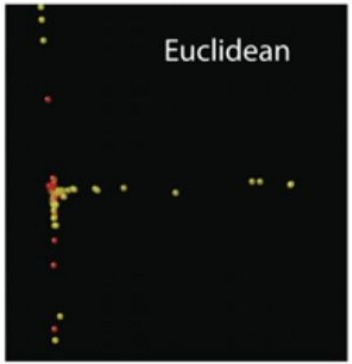
Hamady, Micah, and Rob Knight. "Microbial community profiling for human microbiome projects: Tools, techniques, and challenges." *Genome research* 19.7 (2009): 1141-1152.

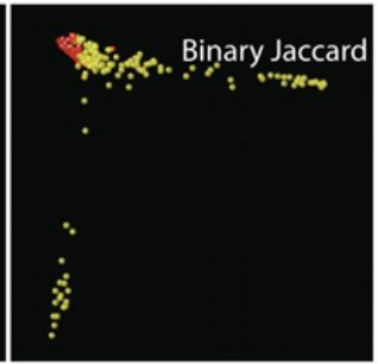
Selecting a metric



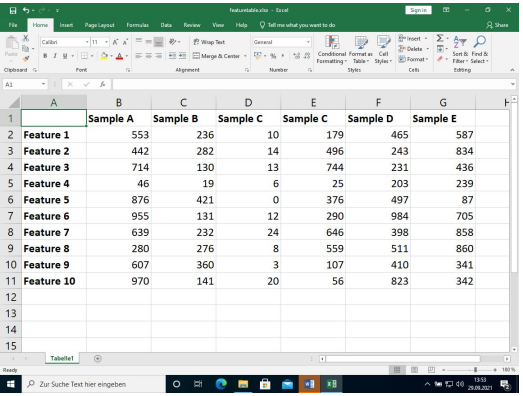




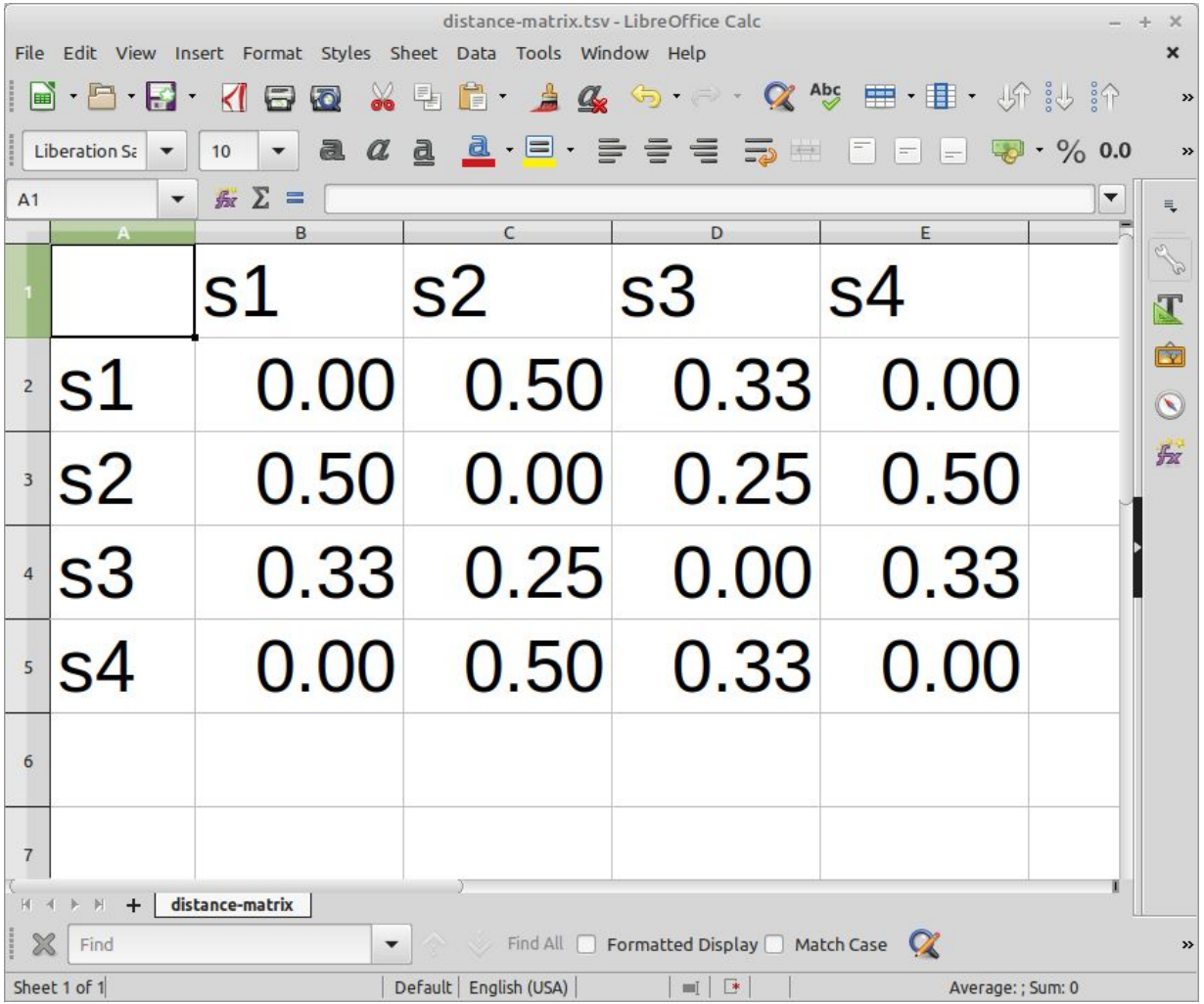
- Vertebrate Gut
- Free living

Hamady, Micah, and Rob Knight. "Microbial community profiling for human microbiome projects: Tools, techniques, and challenges." *Genome research* 19.7 (2009): 1141-1152.

4. Dimensionality Reduction (PCoA)

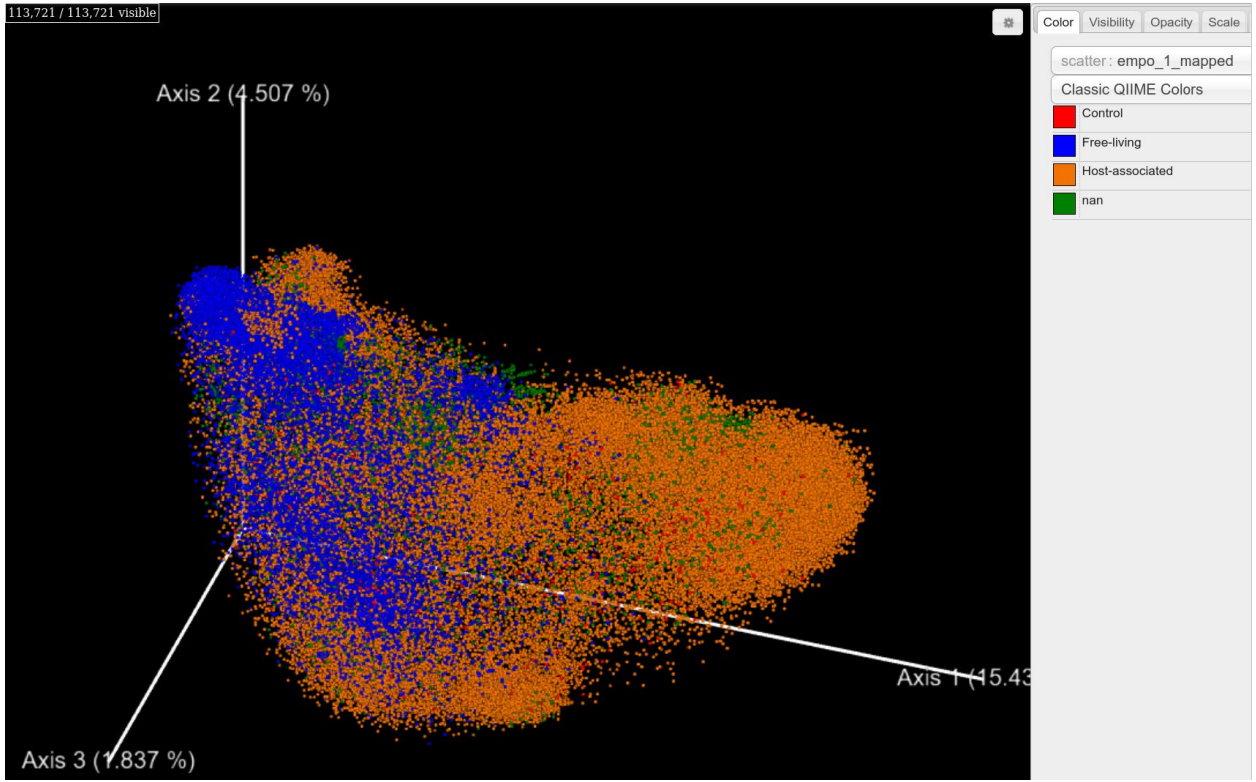


	A	B	C	D	E	F	G
1		Sample A	Sample B	Sample C	Sample C	Sample D	Sample E
2	Feature 1	553	236	10	179	465	587
3	Feature 2	442	282	14	496	243	834
4	Feature 3	714	130	13	744	231	436
5	Feature 4	46	19	6	25	203	239
6	Feature 5	876	421	0	376	497	87
7	Feature 6	955	131	12	290	984	705
8	Feature 7	639	232	24	646	398	858
9	Feature 8	280	276	8	559	511	860
10	Feature 9	607	360	3	107	410	341
11	Feature 10	970	141	20	56	823	342
12							
13							
14							
15							

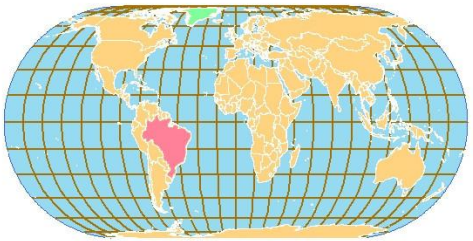


	A	B	C	D	E
1		s1	s2	s3	s4
2	s1	0.00	0.50	0.33	0.00
3	s2	0.50	0.00	0.25	0.50
4	s3	0.33	0.25	0.00	0.33
5	s4	0.00	0.50	0.33	0.00
6					
7					

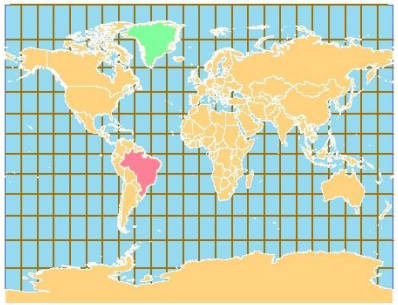
https://bit.ly/2LHMDFC



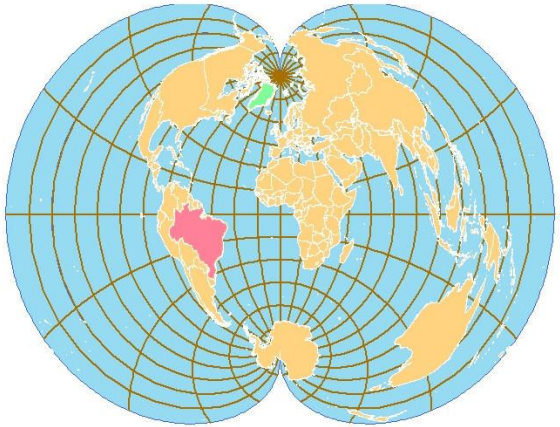
Four World Map Projections



Eckert IV



Gall Stereographic



Polyconic

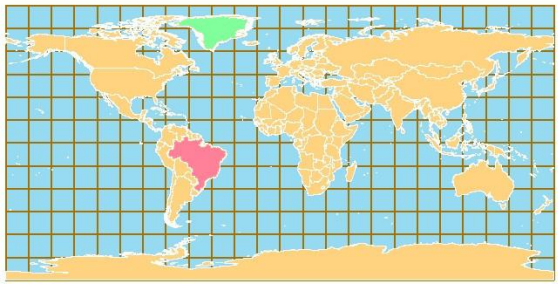


Plate Carree

Principal Scale 1:375,000,000
Created by Matt Sandee 02/10
Data Source: ESRI 2008

area in km²

Brazil: 8,515,767

Greenland: 2,166,086

https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_area