

# Visualization assignment 2016-2017

## • Design:

### Task 1: the difference between "STATE\_0" (rename it as tree 1) and "STATE\_23000" (tree 2)

For comparing these two trees, I used R to plot the trees based on the given data file. They have a similar relationship on these four species: bonobo, chimp, human and gorilla. The distinct difference between them is the orangutan species belong to which group, whether in the 4 species group or together with siamang as another group. As shown in the Figure 1, there is a white line to indicate the groups, which is clear to see the imparity of tree 1 and tree 2.

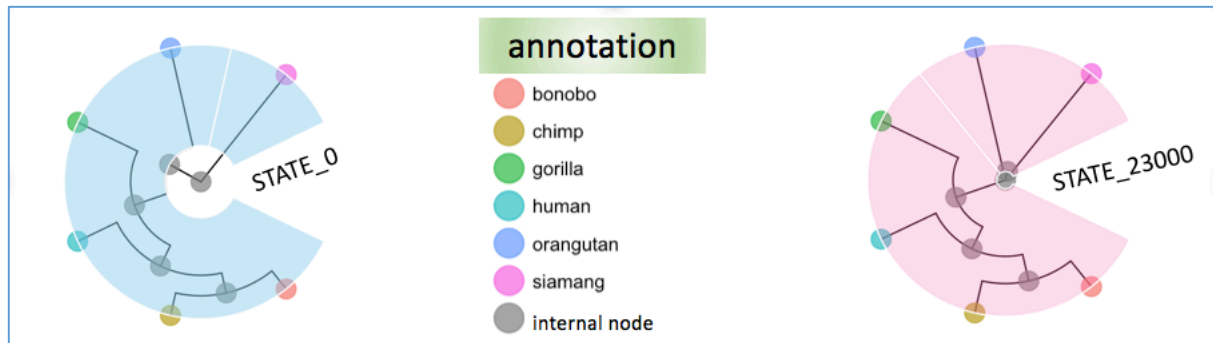


Figure 1. group classification of two trees (design 1)

However, design 1 doesn't show the branch length between species, so that I made design 2 to see the distance length between species intuitively. In Figure 2 left picture, each semicircle represents a specified species that labelled as a bigger point on the leftmost endpoint of the semicircle or as the species picture below semicircle. The small points on the semicircle are other species, which shows the distance between each of them and the specified species. And the outer circle expresses the information of tree 1 while inner circle is tree 2. In addition, the distance values from NEXUS file are transferred to the radian measurement, that is around  $8.78 \times 10^{-4}$  unit per radian. If someone is interested in a specified species, for instance, siamang, he can conclude that siamang is close to the orangutan than other species in tree 2 while it has same relationship among other species in tree 1.

Alternative way to show the distance is shown in design 3, but it zooms in the distance between groups not on the individual species as design 2. From the right picture of Figure 2, left side is tree 1 and right side is tree 2. Based on the Figure 1, five species except siamang can be aggregated as one group in tree 1, and the straight line between group species and siamang are quantitative (tree1: 0.1314; tree2: 0.1092 unit). It can show the difference between tree 1 and tree 2 simply by the length of straight line and the group classifier.

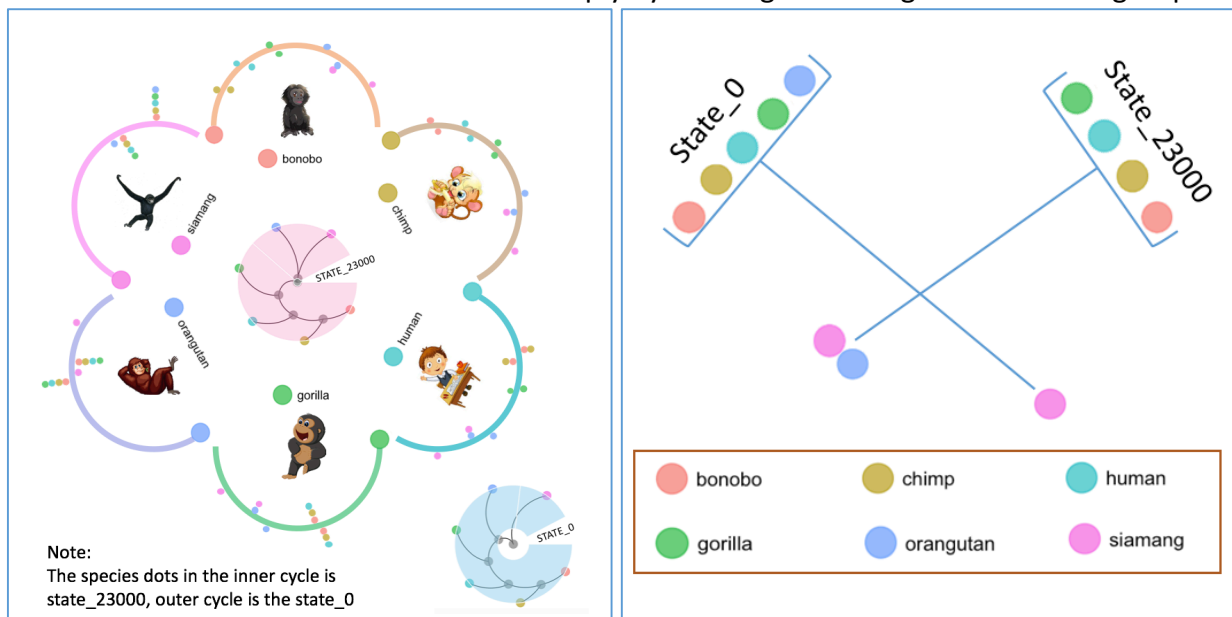
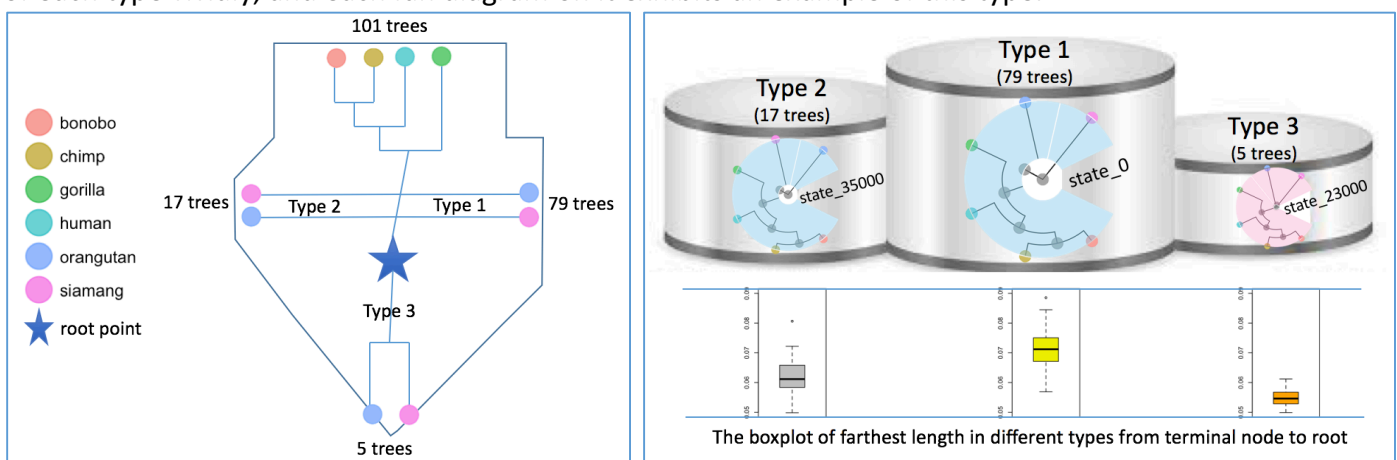


Figure 2. (Left, design 2) the evolution distance between other species by given one species  
(Right, design 3) the distance between groups of two trees

## Task2: containing all trees and how they are similar/different

Among 101 trees, they have the same pattern in the bonobo, chimp, gorilla and human, which is the common group shown in the top side of left picture of Figure 3. On the different part, there are 79 trees belong to type 1 that orangutan is closer than siamang to the common group, which is displayed in the right part of the picture. And 17 trees shown in the left side are type 2 that change the order of orangutan and siamang relative to the type 1. The minor amount is type 3 that orangutan and siamang can be regarded as the same, so that they are included in one group and shown in the bottom side.

It's perceptual to see the common and diversity among all trees, however, it doesn't show the branch length. Since it's impossible to draw the distance between species in every tree like design 2 of task 1, I focused on the variability with different types as described in design 1. By plotting all trees in one picture via R ggtree function, I found the length between the terminal species to root was slightly different. Therefore, I calculated the farthest length of each type and plotted them as a boxplot (shown in the bottom of right picture). From these boxplots, type 1 has the largest average value, that is around 0.071 unit. Second one is type 2 with mean 0.062 unit. They both have one tree in the outlier. In terms of type 3, it contains only 5 trees and the average farthest length is approximate 0.055. Moreover, the podium mimics the total amount of each type vividly, and each fan diagram on it exhibits an example of this type.



**Figure 3. (Left, design 1) three different types of group classification**  
**(Right, design 2) the statistic value about the farthest length from species to the root among three types**

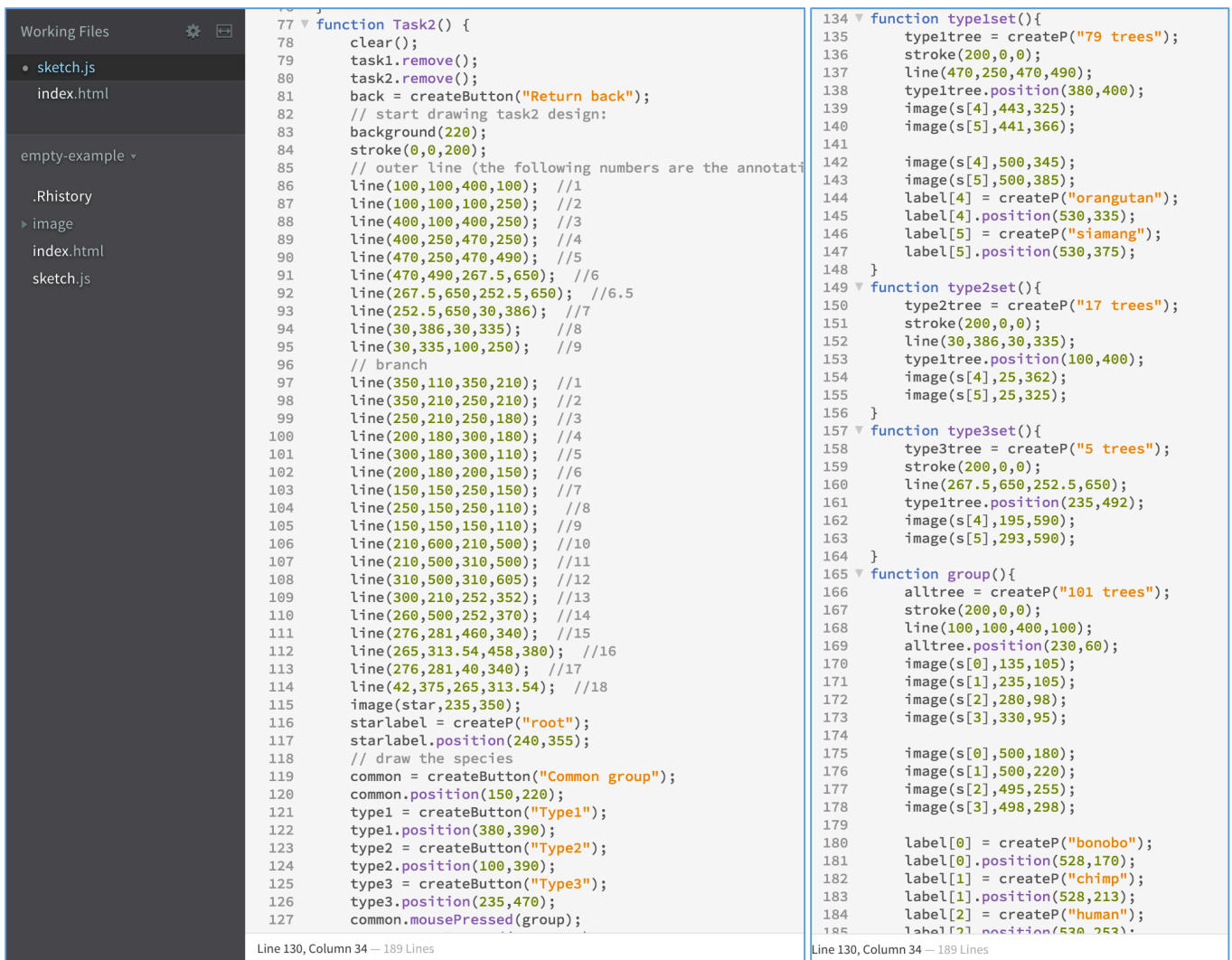
### • implementation:

I used p5 to do the implementation part for the design 2 of task1 and design1 of task2. In the whole setup, you can choose which task you want to see in the homepage. For the task1, I only show the design picture I have done in the first part. To be honest, I should choose design 3 to show due to its simpler format and I believe I can. I really like the design 2, although it looks complicated. However, it can directly tell us the distance between other species and the given one. Therefore, I inserted this design picture which was also precisely measure by tools based on the given branch length data. The following pictures display the declaration of the variables, drawing data of task2 and the interactive parts of code.

```
Working Files
• sketch.js
index.html

empty-example
.Rhistory
image
index.html
sketch.js

1 var canvas;
2 var task1; //text to show 'task1', so that can click it then show the task1 design
3 var task2; //text to show 'task2', so that can click it then show the task2 design
4 var back; //button object back to homepage for choosing task
5 var img_tree; //the picture of homepage
6 var img_task1; //task1 design picture (I'm sorry here I think I'm a bit cheating since I didn't use the p5 to generate this design,
7 // instead of using the draw picture to show. It's hard for me to finish it on the implementation part. But the design picture was also
8 // relative precisely to measure by the given branch length data)
9 // species points and label; star is the root label
10 var s=[];
11 var label=[];
12 var star;
13 var starLabel;
14 var common; // the common part of all the trees
15 var alltree; // text to show the total amount to trees, i.e. 101 trees
16 var type1; // button to show trees belong to type 1 look like, same with var 'type2 and 3'
17 var type2;
18 var type3;
19 var type1tree; // text to show the number of this type trees, i.e 79 trees, same as below
20 var type2tree;
21 var type3tree;
22 // load some picture that I want to show.
23 function preload(){
24   img_task1 = loadImage('image/p1.png');
25   img_tree = loadImage('image/tree.png');
26   star = loadImage('image/star.png');
27   for (var i = 0; i < 6; i++){
28     s[i] = loadImage('image/s'+ i + '.png');
29   }
30 }
31 function setup() {
```



**Figure 4. (Top) the declaration of all variables; (Below) left picture displays the overviews of how to generate the task2 design, and the interactive action data are shown in the right side part.**

For the interactive parts, I created 4 buttons, one is 'common group' for the similarity, and three types for the differences of all the trees. In the right picture of Figure4, function group is to show the similar relationship pattern when you click the 'common group' button. And other three functions are to display the difference part of this type. Meantime, it shows the amount of trees belongs to this part by the length of red lines. This part should be more clear in the screencast video.

I chose some parts of my implementation code, while I attached the original code in the 'p5code' file in case the discrete screenshots are not clear.

### • Insights:

After doing the visualization of comparing all the trees, they both have a same relationship pattern with species bonobo, chimp, human and gorilla. The difference is in the orangutan and siamang, either one of them closer or they are same to the 4 species group. From the design 2 of second task, there is an interesting result in the boxplot, which is obvious to see the more trees this type has, the longer distance from species to the root is. I'm not sure if my hypothesis makes sense, but according to this discovery I suppose that the accuracy of building a phylogenetic tree may be related to the evolutionary time defined.

The following thinking is the reason that I have this hypothesis. Firstly, I assume this tree pattern is right if there are many trees have the same relationship pattern with it, which might be wrong in some cases. Secondly, type1 contains 79 trees, which is a large proportion over 101 trees. Based on the preceding assumption, type1 trees are better than type2 and type3. In addition, the farthest branch length is type1 followed by type2 and type3. In my personal view, whatever the alignment method, maybe DNA or protein sequence, these species may have more diversity on their sequences if the scientists define a longer evolutionary time. So the more diversity between species will give more information to make a true

relationship between species. Finally, I guess the right decision of drawing a phylogenetic tree may have a longer evolutionary time.

I feel bad that due to my limited ability and imagination, I can't come up with the third design for task 2. And I'm really sorry about that I only have 2 designs of task2.

- **Screencast:**

The below link is to show my implementation result on the website by using p5. I introduced how it works and description about my visualization in the video. However, it took 3 minutes and a half since my speaking speed was not fast. (You can change the speed on youtube to save time 😊) I really appreciate your listening and watching.

Screencast link: <https://www.youtube.com/watch?v=jMWmtw-1oG8>