

# Management of Large-Scale Omics Data

# Visualization assignment

Carbonez David (r0292212)

2/06/2017

---

## Design

Task 1: one visualization that shows the difference between STATE\_23000 and STATE\_0

### *Visual 1 (also see implementation part)*

Draw a hexagon with one species at each vertex. The dot in the middle is the center of the hexagon. Along the line from the center to the vertex, put a dot that represents the distance from the closest node to that species. Do this for each species and then connect the dots. The shape represents the distance of each species to their closest node. Make an overlay of the two shapes, showing the differences between the two states.

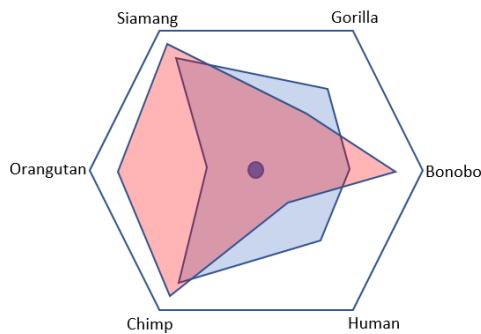


Figure 1: Digital mockup of visual 1, part 1.

### *Visual 2*

Create a Venn-diagram with the communal part depicting the parts of the tree that are identical. The separate parts of the Venn-diagram depict the differences. A colored dots shows where the communal part fits in in the separate parts.

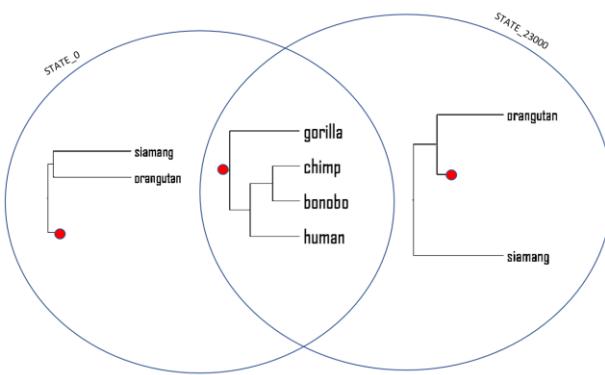


Figure 2: Digital mockup of visual 2, part 1.

### *Visual 3*

Create one tree that contains all possibilities in new branches. Color code the branches so that it's clear that they come from different trees. The original distances are preserved.

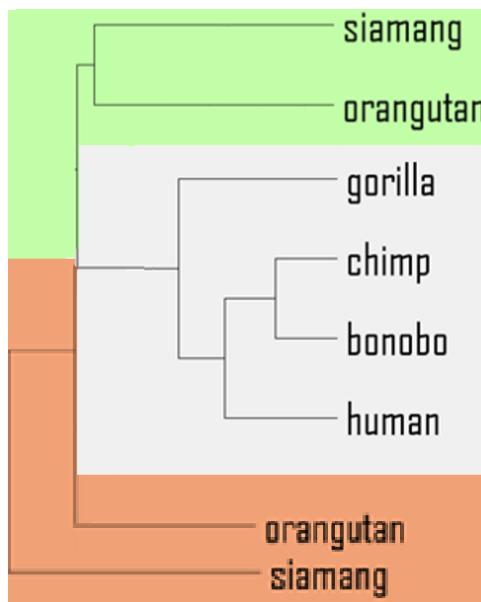


Figure 3: Digital mockup of visual 3, part 1.

Task 2: one visualization containing all trees and how they are similar/different

#### Visual 1

Order the trees in the nexus-file in a decreasing order (this can be done with FigTree). Then count the number of times the same order is present. (e.g. 83 trees have siamang as their most distant branch, 17 trees have orangutan as their most distant branch). Depict this order as depicted in the sketch below. The top row is the most distant branch, then follow the lines downward to find the entire path of the tree. The thickness of the lines is a measurement of the number of trees that followed this path. In the digital mockup below, it is clear that the most followed path is siamang → orangutan → gorilla → human → chimp & bonobo.

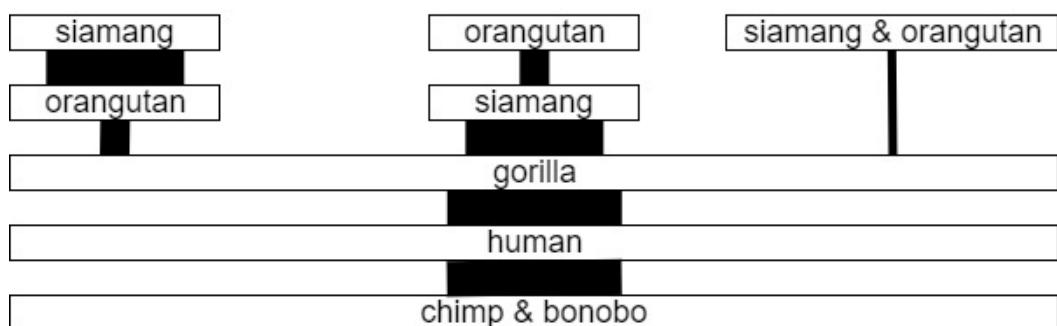


Figure 4: Digital mockup of visual 1, part 2.

#### Visual 2

Draw a hexagon with every possibility given in the nexus file. Inside the hexagon, draw an arrow for every branch from one species to the closest related other species. Outside of the hexagon, draw the most occurring events in bold arrows. Use different colors to clearly distinguish between the different connections.

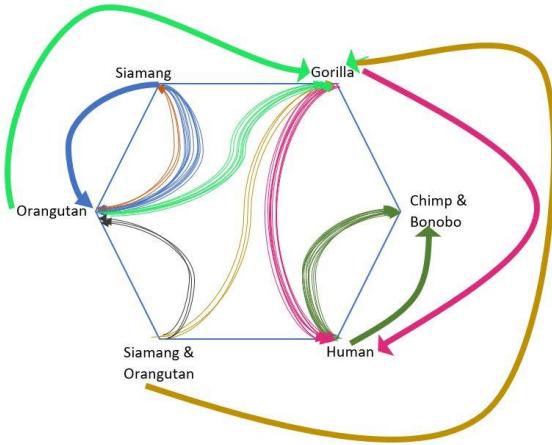


Figure 5: Digital mockup of visual 2, part 2.

### Visual 3 (also see implementation part)

Order the trees in the nexus-file in a decreasing order (this can be done with FigTree). Every branch tip can then be visualized as a dot in the graph below. The y-axis represents the distance of that branch tip to the closest node. The x-axis represents the divergence of the species. The branch that splits off first is the most diverged branch. The branch tips from one tree are connected between the dots with a line. The dots are color-coded according to the species. The screenshot of the implementation below shows that the animal that is the furthest away from all the others is in most cases the siamang (red), with a distance to the closest node varying from around 0.08 to 0.045. The second furthest away species in 83 cases the orangutan, and in 17 cases the siamang. The distance to the closest node varies between 0.069 and 0.042. (see figure 6)

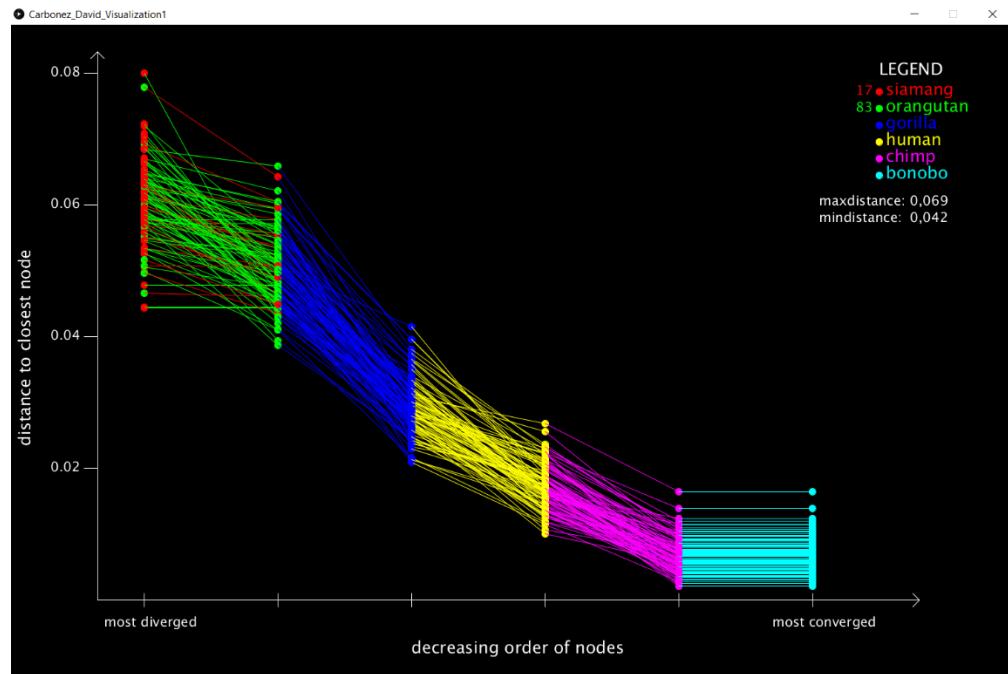


Figure 6: Digital mockup (screenshot) of visual 3, part 2.

## Implementation

Both implementations were performed using the Processing sketchbook with Java.

### First implementation: design 1.1

Upon starting, it shows the hexagon with the species, a set of instructions and a legend on the right (figure 7A). Upon interacting (clicking the boxes next to the legend), the program displays the shape that corresponds to the specific state or both states at the same time and the corresponding color in the legend. (figure 7B: one state, figure 7C: both states at the same time).

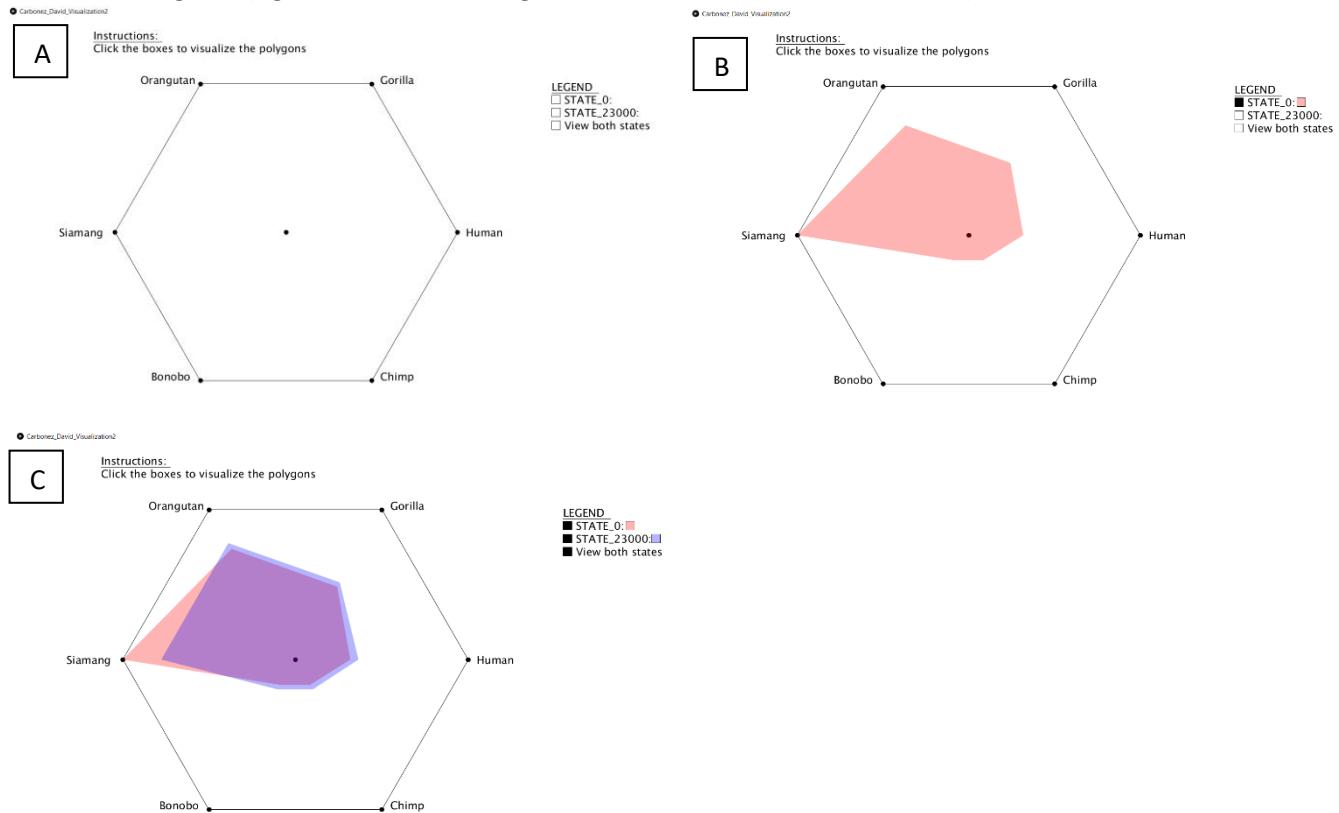


Figure 7: Screenshots of implementation 1. A: Start screen, B: one state selected, C: Both states selected

### Second implementation: design 2.3

The second design draws shows one dot per species per tree. All the dots are immediately represented. Interaction is possible by moving the mouse over a group of dots representing on specific step of divergence. If you mouse over a group, the legend will show how many dots of a specific color are contained in that group and will show the minimum and maximum distance to the closest node. Figure 8 shows the case where the mouse is hovering over the first (most diverged) group.

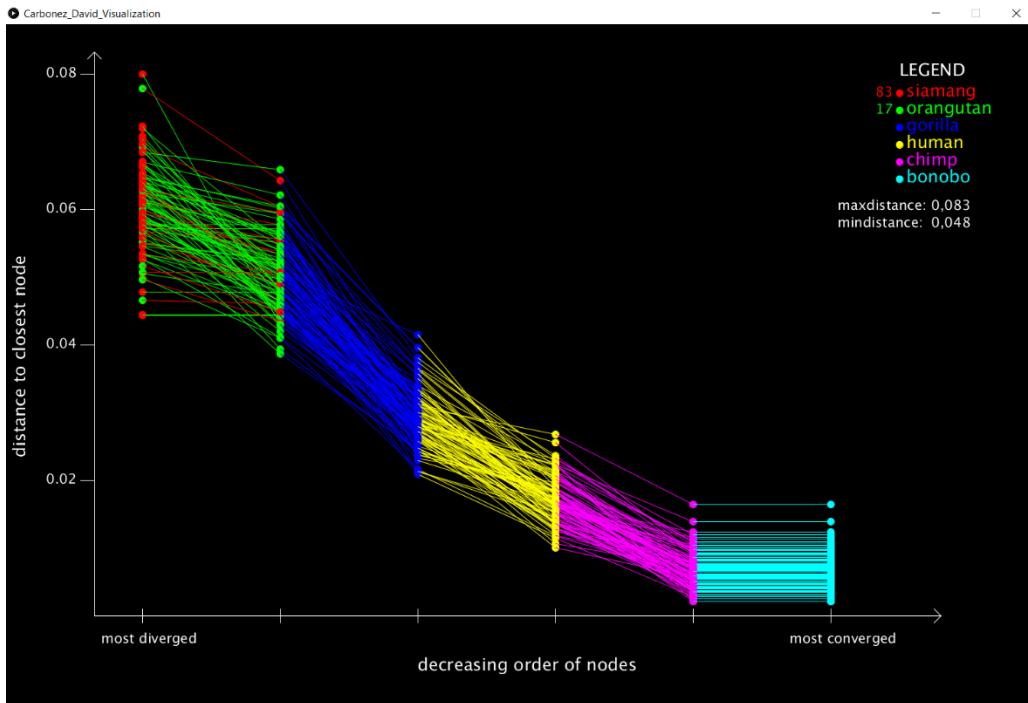


Figure 8: screenshot of implementation of visual 2.3

## Insights

The first visualization showed that between STATE\_0 and STATE\_23000, there is not a big difference. The measurements were normalized so that the largest value present (siamang in STATE\_0) would be equal to the maximum value. The distances for every species to the closest node follow a similar trend between the two states.

The second visualization shows that in 83 out of 100 trees, the siamang is the most diverged species. 17 out of 100 trees show that the orangutan is the most diverged species. The second most diverged animal is most commonly the orangutan. Going further in the direction of less divergence, all models agree that the order is gorilla, human, chimp & bonobo. Moreover, this visualization shows that the ranges over which the distances vary decreases as the species diverge less. A last conclusion shown in the second visualization is that in every tree, the bonobo and the chimp have an equal distance to the closest node, shown by the horizontal lines between the dots.

## Screencast

A screencast, in which I explain how my visualizations work and show they interactivity can be found at: <https://www.youtube.com/watch?v=LMRNdqDdGMI>