Birth Canal Measurements in Primates

Team 7: Tianyue Wang, Shijie Cai, Elizabeth Morris, Matt Zewbert

TF: Ryan Gryder

Client: Natalie Laudicina

4/26/2018

1. Overview

1.1 Project Background

In this project, our client Natalie Laudician, a phd student conducted a research on the evolution of birth canals of primates species. She is interested in comparing birth canals among related primate species including Hylobates lar, Symphalangus syndactylus, Gorilla, Pan troglodytes (Pan t.) and Pongo.

Our client retrieved a variety of geometric measurements from pelvic bones collected from four museums using 3D scanning device. The goal of this project is to recognize the best measurements to use for comparing the birth canals of different species with a particular focus on identifying the “constraints” of the birthing process and how they differ across the species of interest.

1.2 Research Questions

Our client has following questions:

1. Comparing birth canals within primate species using measurements from 3D scanned pelvic bones from the species of interest.
2. Identifying the “constraints” of the birthing process and how they differ across the species of interest

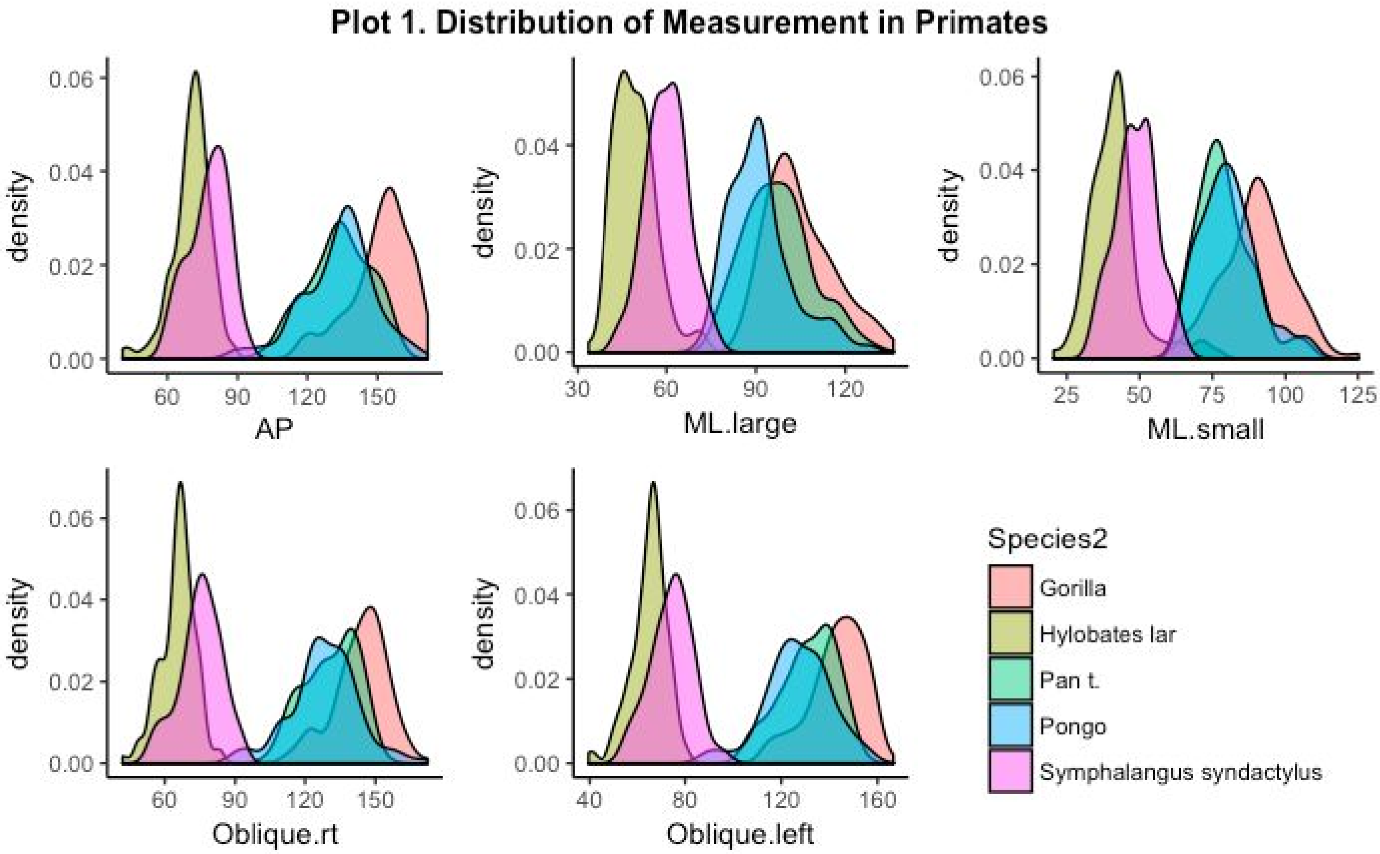
1.3 Data Description

The dataset is collected in a manner that we have:

- 13 levels of birth canal measurements
    - 11 rings and 2 Ischial tuberosities
    - 5 Measurements for each level:
        - AP (anterior-posterior diameter), ML.large(maximum mediolateral), ML.small(minimum mediolateral), Oblique.rt(oblique right: widest cross dimension), Oblique.left(widest cross dimension from other way).
- 5 primate species:
    - Hylobates lar, Symphalangus syndactylus, Gorilla, Pan t. and Pongo
- Museums:
    - 4 museums where pelvic bones were collected
- Specimen number:
    - 44 distinct ID for each pelvic bone measurements

- Fetal Measurements:
    - Fixed Neonatal Head Measurements and Shoulder Breadth for 5 species

2. Exploratory Data Analysis

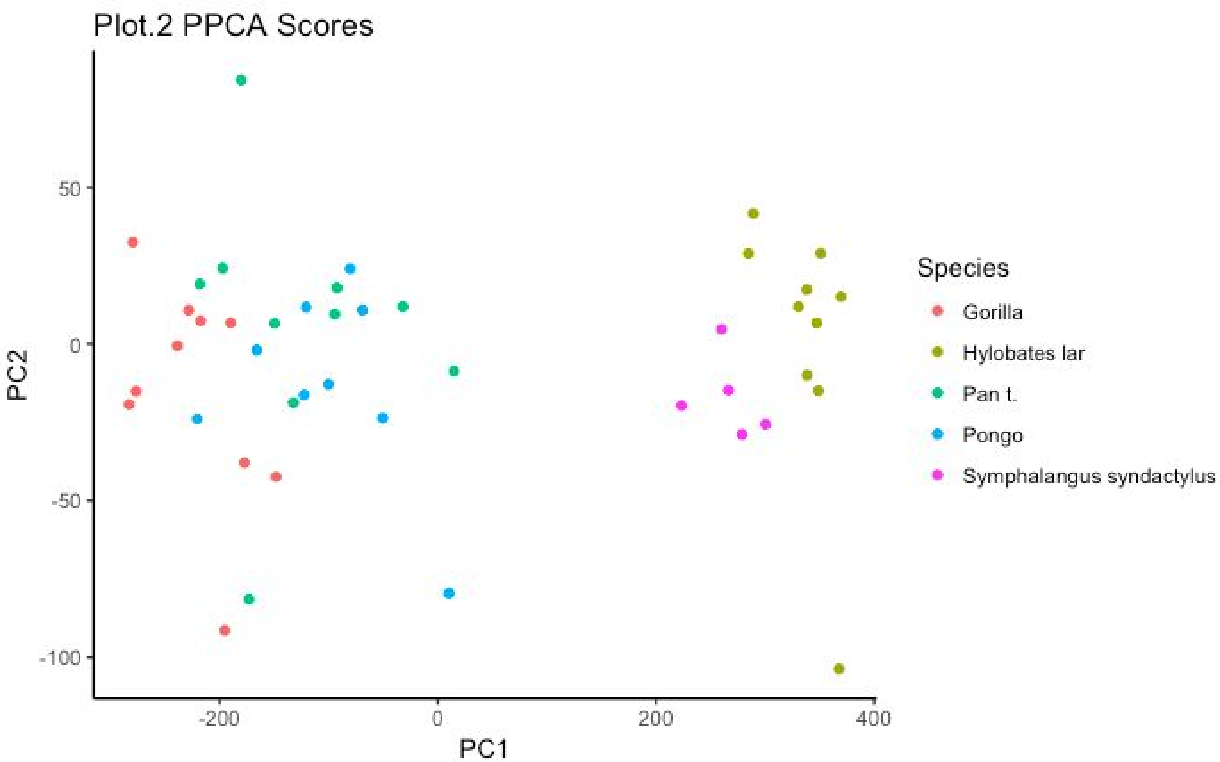


According to Plot.1, we can see a clear separation between Hylobates lar, Symphalangus syndactylus and the other three species in all five measurements. The former two have higher spikes and lower ranges in all measurements, while under the same measurements, the other three have lower spikes and approximately same lower ranges. Also, we observe normal distributed trends in measurements of all species. Other plots on more details are included in the appendix.

3. Methods

3.1 Probabilistic Principal Component Analysis (PPCA)

We exploit an unsupervised method to help locating the constraints of birth canals within the primate species. Since we have 7% missing observations in the dataset, the PPCA approach is appropriate in this case as it solves the missing values issue assuming a normal distribution in all measurements.



The first principal component (PC1) explains 93% of the variance capturing measurements mostly in AP and Oblique right, while the second principal component (PC2) explains about 3% of the variance capturing mostly ML large and ML small measurements.

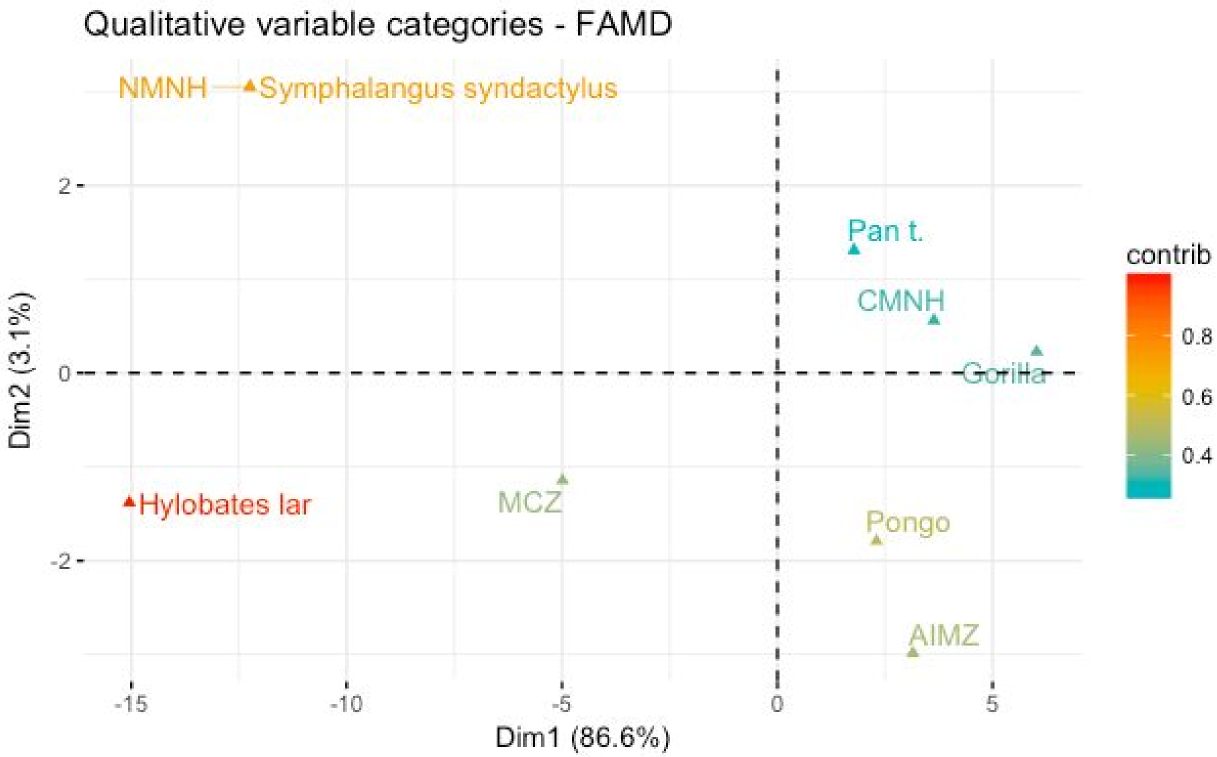
According to Plot.2, it is easily observed that Hylobates lar and Symphalangus syndactylus are very distinct from the other three species on the first principal component, which resonates with the separations we have seen in Plot.1. In PC1, the five measurements that makes most variations are AP measurements in ring 4,6,7,8 and Oblique left measurement in ring 9. On the opposite, the top five measurements making most variations are AP, Oblique.left, Oblique right measurements in the second Ischial tuberosities, ML small measurements in ring 1 and 2.

3.2 Factor Analysis of Mixed Data (FAMD)

As the previous section discussed variance in the birth canal measurements using the Probabilistic Principal Component Analysis, this section will discuss the correlation among the 2 categorical variables: museum and species, along with all the birth canal measurements.

Similar to Probabilistic PCA, we selected 2 dimensions of factors based on optimal percentage of variances explained. The first dimension explained 86.6% of the variance by capturing most birth canal measurements, while the second dimension explained 3.1% of the variance by capturing Species, Museums and ML large in the second ring.

Plot. 3 Qualitative variable categories in Mixed Factor Analysis



As indicating from Plot. 3, we can see two distinct primate groups in the first dimension (Dim1). The first group consists of Hylobates lar and Symphalangus syndactylus, while the other three species are included in the other group. We can observe from the plot that the museums(NMNH, MCZ, CMNH and AIMZ) also have distinct separations in both dimensions.

While the factor analysis in mixed data offers a new perspective by involving two more categorical variables, however, it neglects the missing values issue.

3.3 Decision trees

To help locating constraints where rotations are needed during primates' labor and delivery, we introduced fetal measurements including fetal head measurements (FO length) and neonatal shoulder breadth.

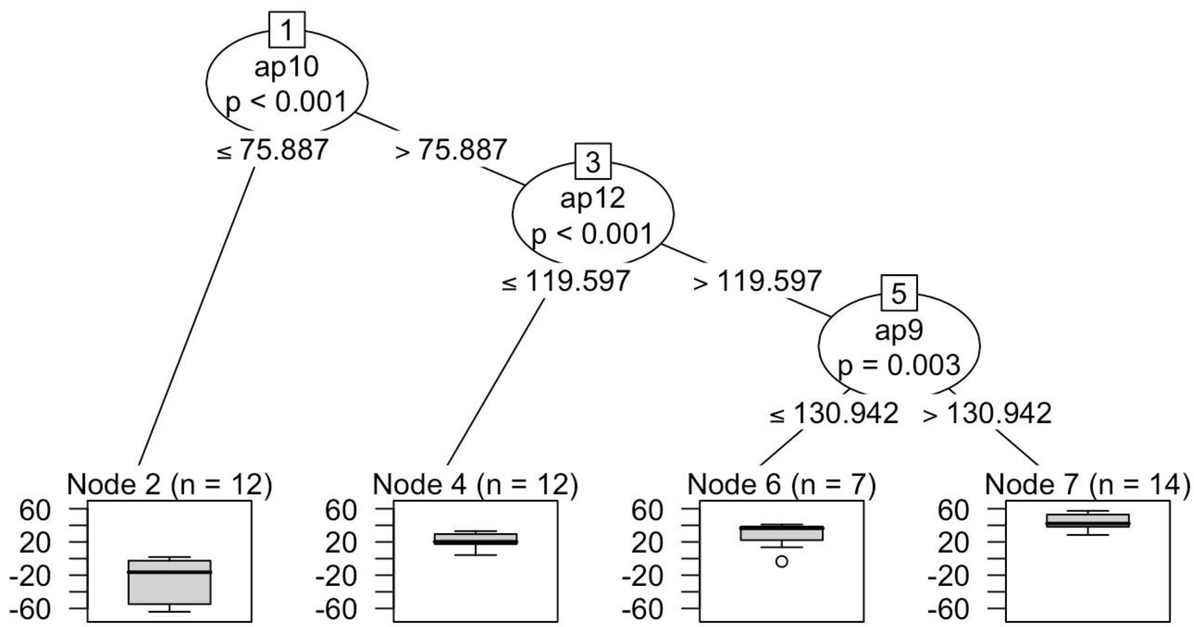
We define the following situations as constraints:

1. **FO length > AP birth canal measurement:** when the fetal head is larger than the AP measurement, a rotation in the shoulder is mandatory.

2. **Neonatal shoulder breadth > ML large measurement**: when the should breadth of the fetus is longer than the ML large measurement of the birth canal, a further rotation is in need.

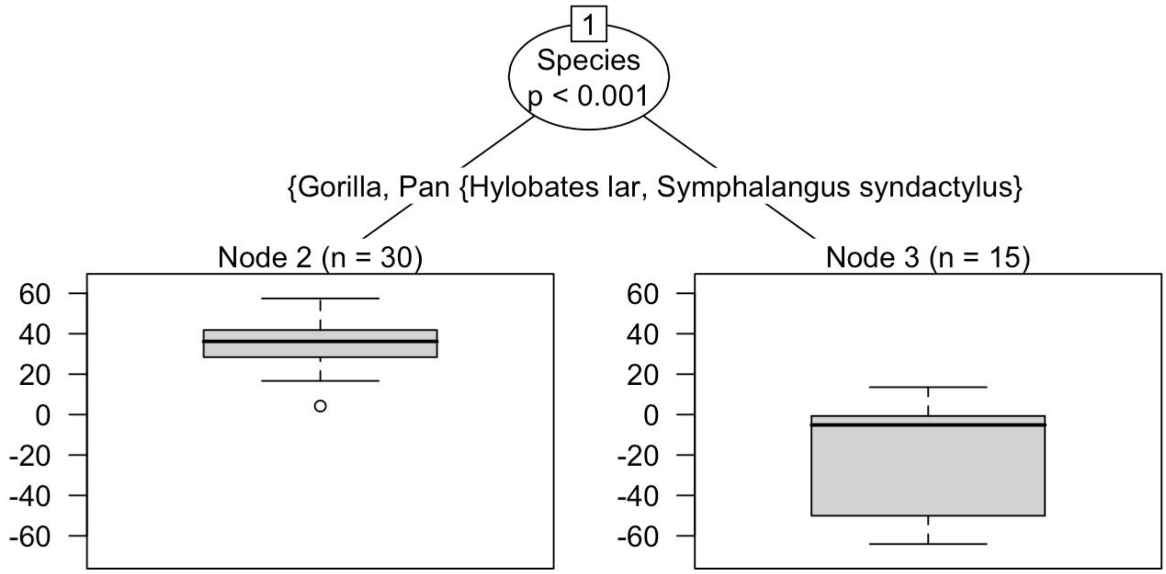
In order to identify these situations statistically, we first created a response value for each observation by subtracting FO length from the minimum value of AP, where a negative difference indicates a rotation in shoulder is necessary. From there, we fit a decision tree indicating the response values corresponding to the AP measurements. In a similar case, we fit another tree indicating the differences between fetal shoulder breadth and ML large measurements where constraints are most recognized.

Plot.4 Decision tree for AP measurements



According to Plot.4, AP measurements in ring 9, 10 and 11 are the main locations where major differences between fetal head measurements and AP birth canal measurements are observed. The boxplots under each node indicate the distribution of observations falling into each tree split. Since only the first boxplot on the left side has a medium below 0 with most of values falling in the negative range, ring 10 is most likely to be the location of constraints when a rotation is shoulder is necessary.

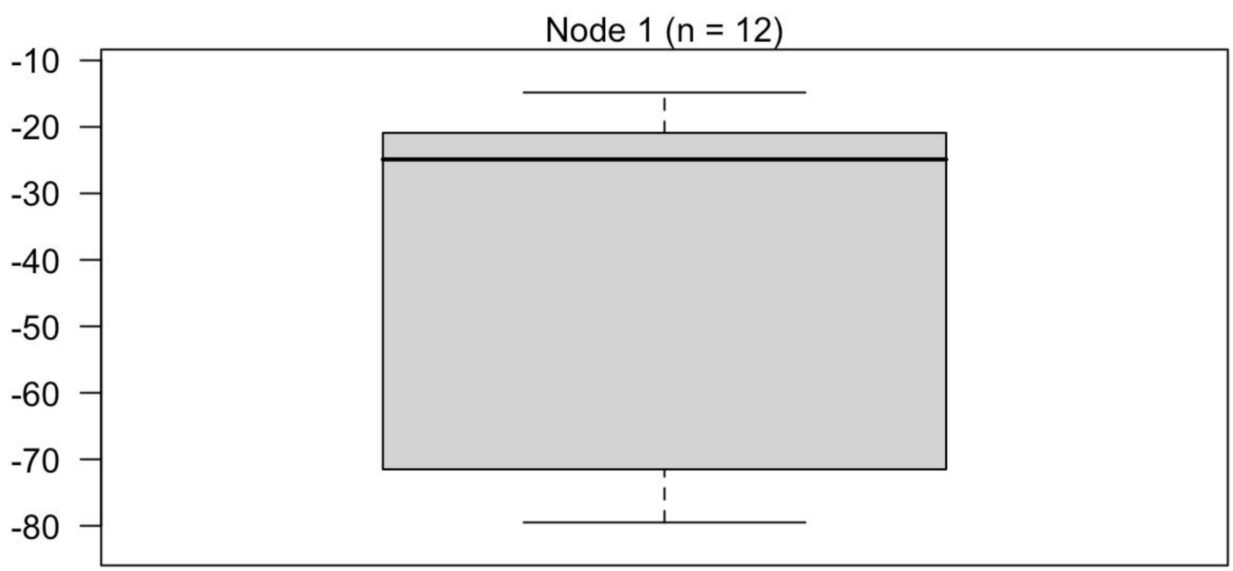
Plot.5 Decision tree for species



As showing above, Plot.5 demonstrates a classification for all primate species while indicating whether there would be constraint in the birth canal. Hylobates lar and Symphalangus syndactylus are classified into the same group, which is consistent with the PPCA results in plot 2. Also, these two species are more likely to have constraints in their birth canals.

The potential next step is to create another tree in order to locate constraints causing by situations where the minimum ML value is smaller than the FO length. However, the tree results

Plot.6 Decision tree for ML measurements



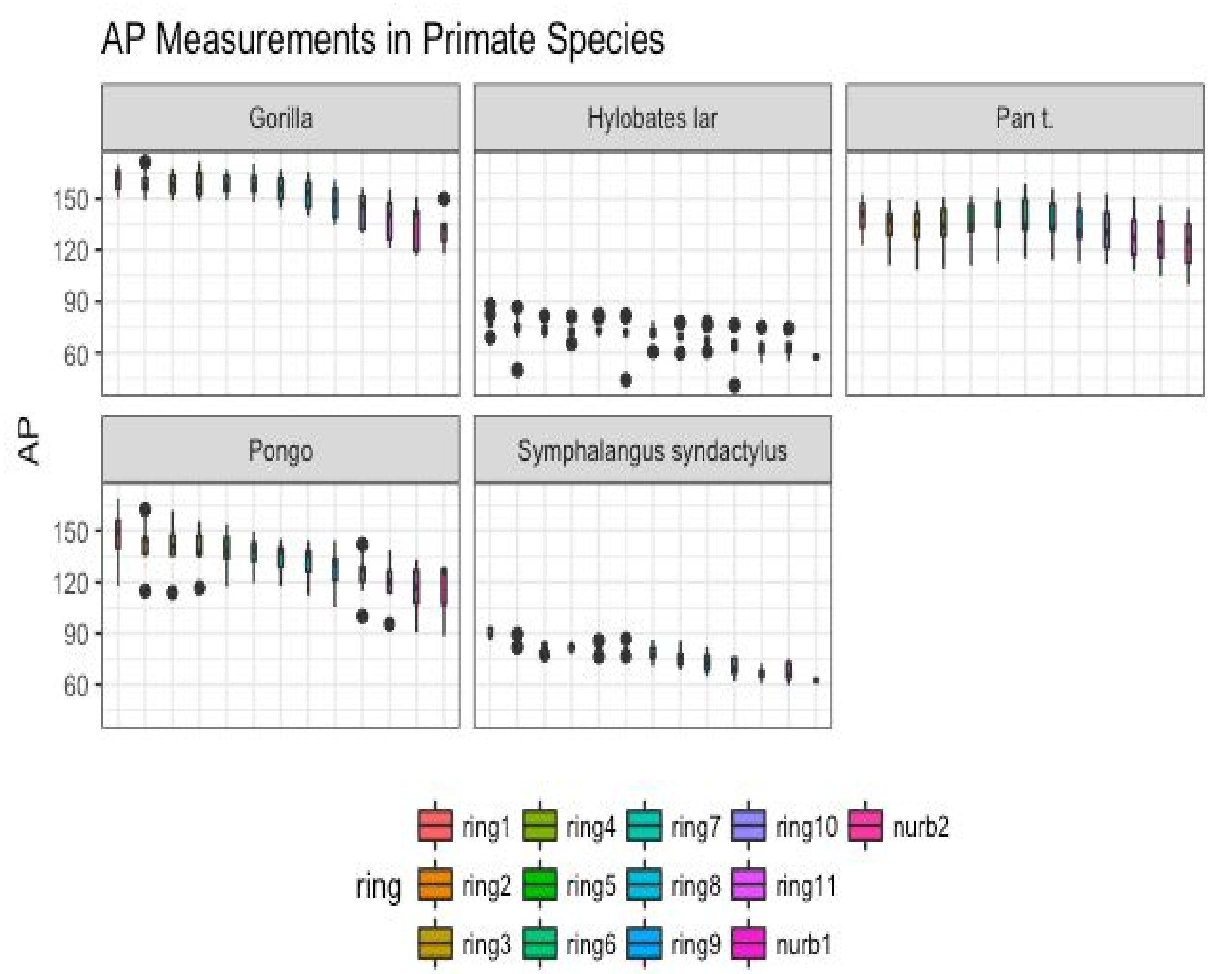
in having only one node, which means all ML measurements are classified into one single group, as showing in Plot.6. It can be derived that for all observations whose AP birth canal measurement is smaller than the FO length, their minimum ML values are very likely to be smaller than fetal shoulder breadth.

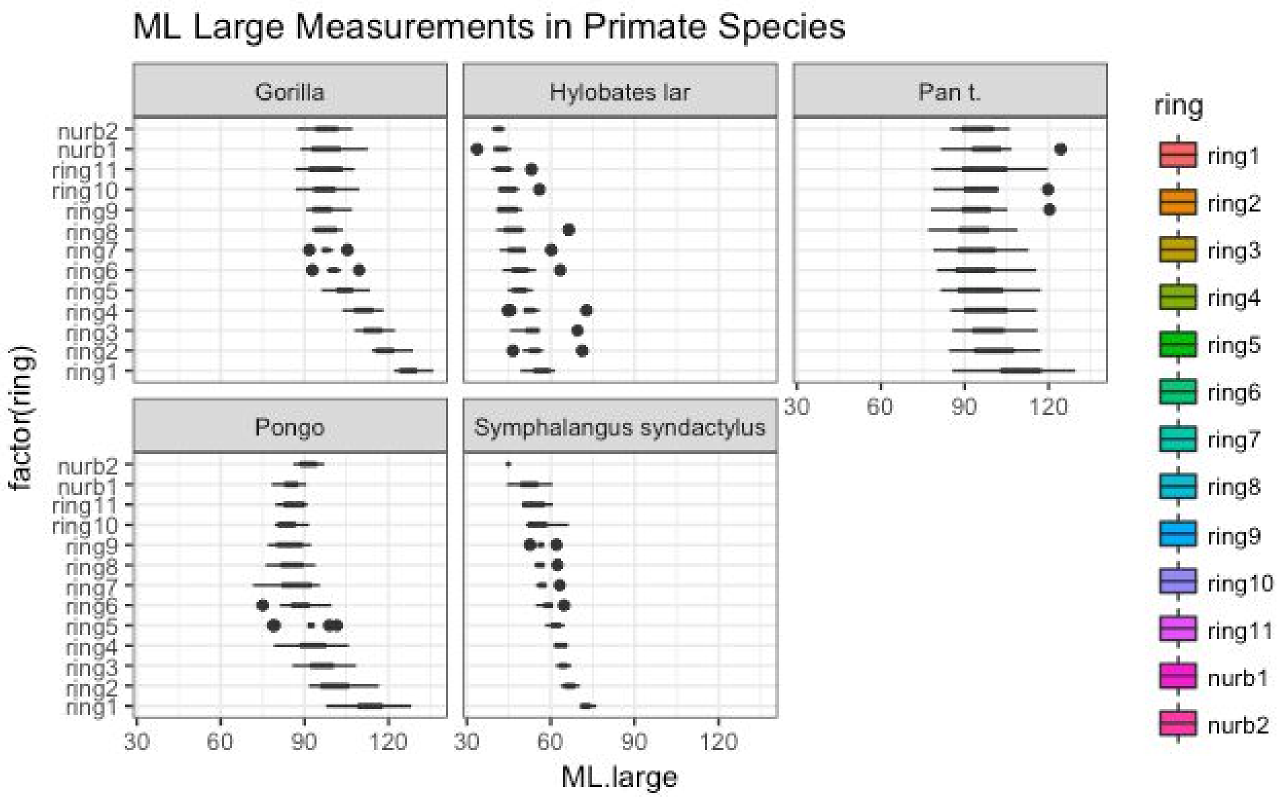
4. Discussion

The limitation of this study is that, we are using mean values for all the fetal measurements. If individual distinct values for fetal measurements are accessible, we can locate constraints in a more accurate way so that the limitation in Plot.6 can also be avoided.

5. Appendix

EDA




ML Large Measurements in Primate Species
Gorilla
Hylobates lar
Pan t.
Pongo
Symphalangus syndactylus
factor(ring)
nurb2
nurb1
ring11
ring10
ring9
ring8
ring7
ring6
ring5
ring4
ring3
ring2
ring1
30
60
90
120
ML.large
ring

PPCA

Mixed Factor Analysis