Reanalyzing Shared Data about Brains and Behavior

Charmi S. Mehta (*csm5410@psu.edu*)
Major: *Psychology (Neuroscience Option)*
Supervisor: Rick O. Gilmore

BACKGROUND

There are many questions about humans being the most intelligent species. Scientists used to believe that humans have a billion neurons in their brain, and that humans had the highest number of neurons than any other species. This was, of course, proven false by Herculano-Houtzel [1]. Her paper shows that the number of neurons for humans according to their brain mass is similar to other species. In fact, animals such as Elephants are outliers. We were curious to analyze the data ourselves and see if we get the same result. Additionally, we tried to find new information from the shared data. The data shared had to be reformatted to perform smooth statistical analyses [2]. Therefore, datasharing and codesharing issues are also addressed with this research.

PROCEDURE

The link for the data was provided with the paper itself, and it was extremely easy to download the raw data-file. The data-file was in an Excel file with different excel sheets dedicated to different areas of the brain. We used 'Rstudio' software [2] provided by The Pennsylvania State University to manipulate the data. The tasks performed for the 'cleaning' process of the data were as follows:[1]

1. Excel file was converted to a CSV file (Comma Separated Value).
2. Empty spaces and extra characters such as '+/-' were removed.
3. Each column and variables were renamed.
4. Separate column called Brain_area was created to collapse different sheets of data in one CSV file.

A new Rmarkdown document was created to analyze the data by creating different types of graphs.[1] The figures provided in the original paper were also recreated. One such example is presented in Figure 2A (recreated graph) and 2B (original graph). The purpose of recreating the graphs was to determine if the data was reproducible. Some of the graphs that were used for analyses showed similar data trends as discussed in the paper. However, Figure 1 showed a new result that was not discussed in the paper.

[1] All the codes and figures (pdf and png format) produced for this project are openly shared on https://github.com/gilmore-lab/brain-behavior-data

RESULTS: FIGURE 2A AND 2B

Fig. 2A shows the three plots that were recreated from [1]. It shows the reproducibility of the data and its figures. Fig. 2B shows the original graphs printed in the article [1]. This graph is labelled 'Fig 3' in the paper.
Graph *a*: Shows a linear relationship between number of neurons and structure mass of Cerebral Cortex without using mouse and naked mole rat as the data with 95% confidence interval (see Discussion for more detail). **Graph *b*:** Shows a better fit for 95% confidence interval, but mouse in included in this graph. The naked mole rat is still excluded. The confidence interval changes a little with the inclusion of the mouse which is seen clearly in Fig. 2B than 2A. Here, code sharing would have been useful to recreate the exact plots presented in the paper. **Graph *c*:** Elephant is clearly an outlier, because the number of cerebellar neurons and the cerebral cortex neurons across all species is linear (except for Elephants).

REFERENCES

[1] S. Herculano-Houzel, K. Catania, P. R. Manger, and J. H. Kaas, "Mammalian brains are made of these: A dataset of the numbers and densities of neuronal and nonneuronal cells in the brain of glires, primates, scandentia, eulipotyphlans, afrotherians and artiodactyls, and their relationship with body mass," *Brain, Behavior and Evolution*, pp. 145–163, Sep 2015. [Online]. Available: http://dx.doi.org/10.1159/000437413

[2] RStudio Team, *RStudio: Integrated Development Environment for R*, RStudio, Inc., Boston, MA, 2016. [Online]. Available: http://www.rstudio.com/
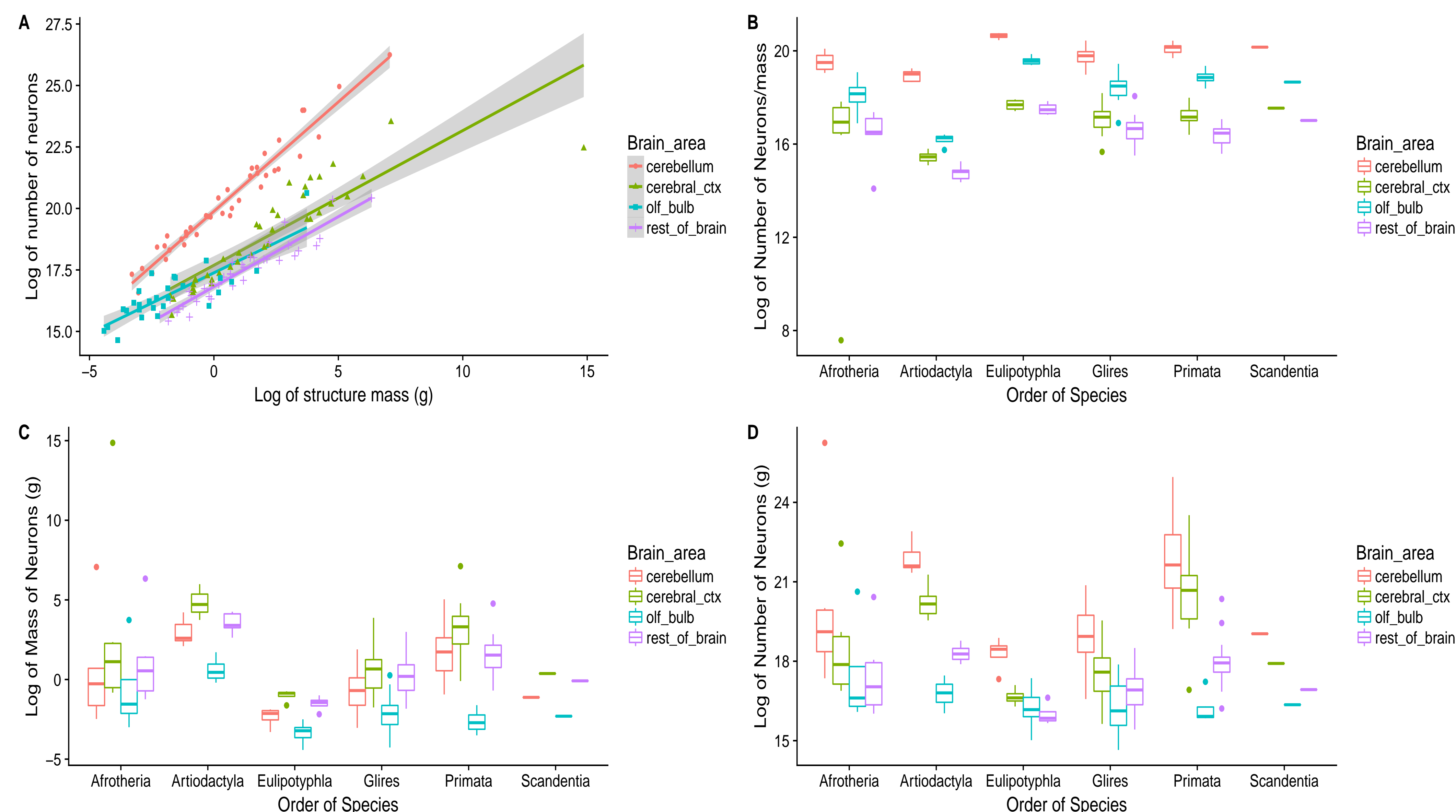
FIGURE 1. SELF-ANALYSIS FIGURES
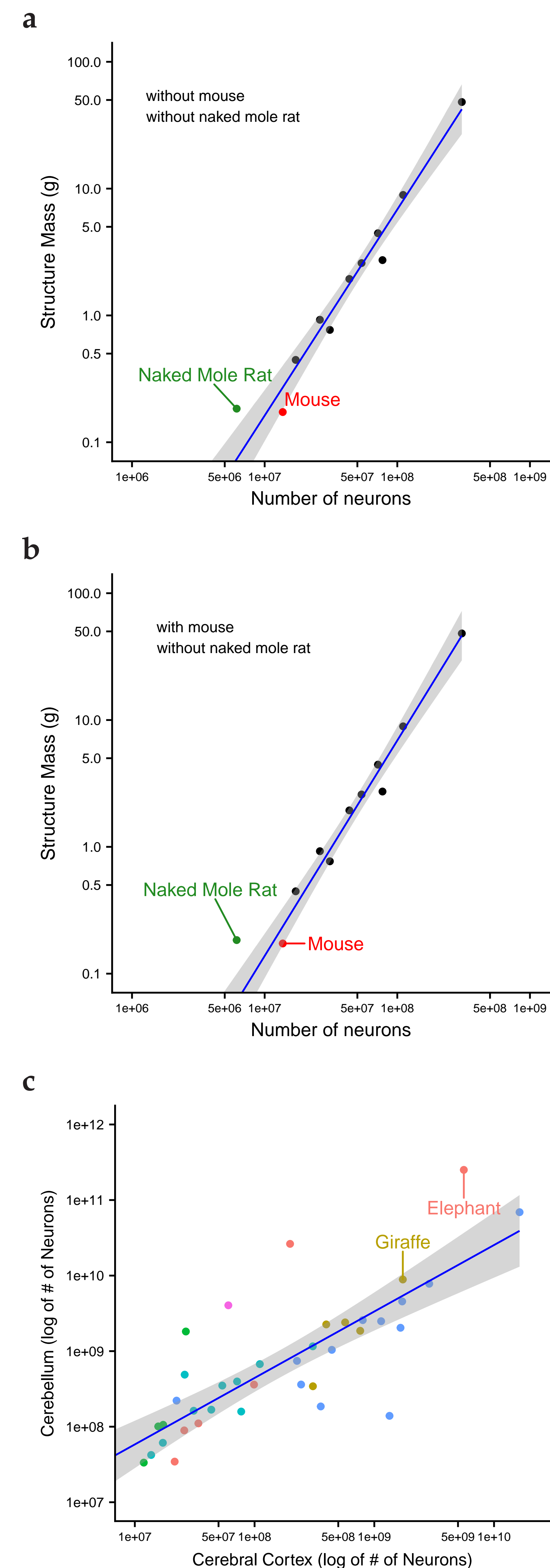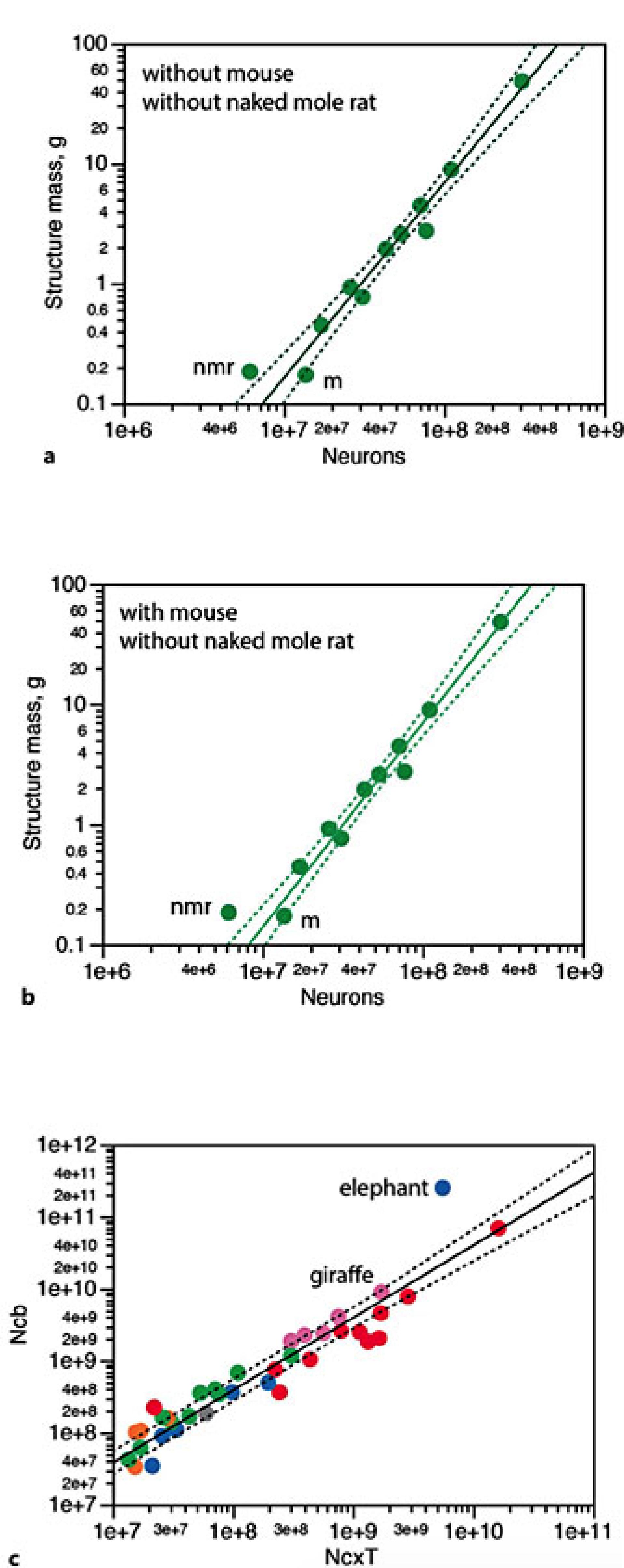

FIGURE 2A. RECREATED

FIGURE 2B. ORIGINAL

RESULTS: FIGURE 1

Afrotheria: belong to the African continent (e.g. golden moles). **Artiodactyla:** hoofed animals (e.g. Deer). **Eulipotyphla:** examples are moles and hedgehogs. **Glires:** include *rodents* (e.g. Rats) and *lagomorphs* (e.g. Rabbits). **Primata:** include primates (e.g. gorillas and human beings). **Scandetia:** include species like *treeshrews*.

(A) Figure A shows the number of neurons (NoN) per mass of each brain area of all the species. Cerebellum seems to have higher density neurons per mass than cerebral cortex which was discussed in the paper.

(B) Density of each brain area (NoN/mass) of different Order of Species (OS) is shown here. Trends of individual brain areas seem consistent among different OS. Density of each brain area from highest to lowest are: Cerebellum, Olfactory Bulb (Olf), Cerebral Cortex (Ctx), and Rest of Brain (RoB). *Graph (A)* shows ctx having more density than olf which contradicts *Graph (B)*. Finally, density of Eulipotyphla seem higher than Primata.

(C) This graph only looks at the mass of neurons of each brain area arranged by the OS. This graph also shows consistency among different OS. The rank from largest mass to the lowest is: Cerebral Cortex, RoB, Cerebellum, and Olfactory Bulb. Additionally, Artiodactlya show higher mass than Primata.

(D) This graph solely looks at the NoN on each brain area arranged by the OS. This graph does not show consistency among different OS. NoN from highest to lowest are: Cerebellum, Cerebral Cortex, RoB, and Olfactory Bulb (Eulipotyphla is an exception). Overall, Primata seems have equal NoN to that of Artiodactyla, but Primata has greater variation than Artiodactyla.

DISCUSSION

Addressing the issue of reproducible data, researchers should keep in mind about sharing their data in an easily manipulable format. Herculano-Houtzel has shared all her data along with her paper which is extremely helpful for other researchers to perform their personal analyses. However, the format she shared the data was in excel file which was not easily manipulable. As seen in the procedure section, data-cleaning was an essential task before performing the main statistical analyses.

Differences between Fig. 2A and 2B show that R codes used to create the figures should also be shared. Removing 'mouse' from the dataset still shows 'naked mole rat' as an outlier, therefore the data points of the naked mole rat should be used with caution in comparative studies. Similar conclusion goes for elephants, as shown in *Graph 'c'* of Fig. 2A and 2B.

The contradicting results between figures 1A and 1B can be further analyzed by the sensitivity of functioning between those brain areas. Herculano-Houtzel shows that Primates do have higher number of neurons than other OS (as shown in Fig. 1D). This might mean that the density of each brain area does not matter as much as the NoN. Perhaps comparison of task-based performance and NoN, mass, and its density of brain areas of different species can help in determining if the NoN of each species are more predictable of their performance. This can also help to distinguish if Artiodactyla are actually more intelligent than Primates.

CONCLUSION

In conclusion, it can be said that, for the time-being, NoN are considered the most predictable way to measure intelligence across different OS. However, task-based performance of each species and its comparison with different brain areas might lead to a different conclusion in the future. Finally, due to Dr. Herculano-Houtzel's shared data, we could further our statistical analyses which shows the importance of data sharing. Additionally, it shows the importance of sharing R codes. These data could also be used for meta-analysis with task-based performance and maybe their diet preferences as well.