

R Notebook

Food and housing insecurity is a social problem. This is an interesting research we loved from the get go. One of the problems we encountered was the missing values in the data, and how to handle them, especially a question that was a dependent variable and had about 95% of the data missing. We also faced some challenges when it came to questions with multiple choices and how to compound them. We began the data cleaning with Excel which was counter reproducible. Because of what we have learnt in this course, we began the data cleaning again using R programming language, which makes the data cleaning process reproducible[1].

Considering one of the dependent variables, we were caught between making another choice for the response variable and using machine learning algorithm for missing value imputation because about 95% of the data was missing. We had to schedule a meeting with the collaborator in other express our concerns and propose our solutions. This meeting helped us understand the best approach to handle the missing value in the data.

Also we exchanged our codes and data through outlook, which made the whole collaborative process strenuous. But after the class where we discuss collaborative web services like GitHub and binder, we began to save our codes on github and share it among ourselves which makes it easily editable therefore making collaboration much easier.[1]

Reference

1.[Roger D Peng]Reproducible research checklist

<https://www.youtube.com/watch?v=pjL2uP-fmPY&list=PLjTlxb-wKvXOU4WW4p3qc4VKWTI4gLNUf&index=4>

2.Link to github files https://github.com/fbiney/DS_Project4/network/members