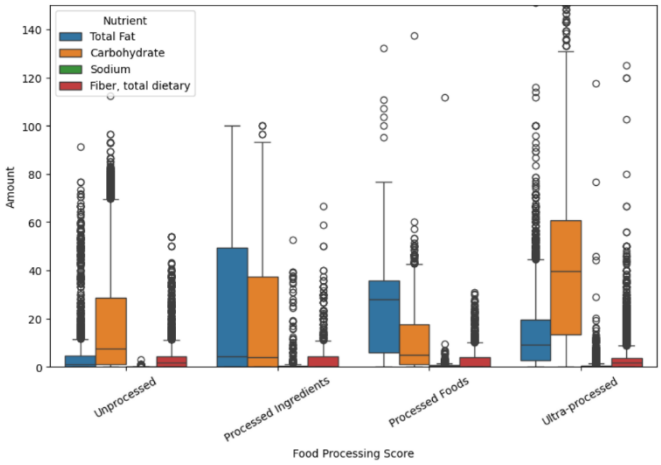
Khanh Phan

DSC 209R

8 October 2025

Project 1: Expository Visualization- Final Submission

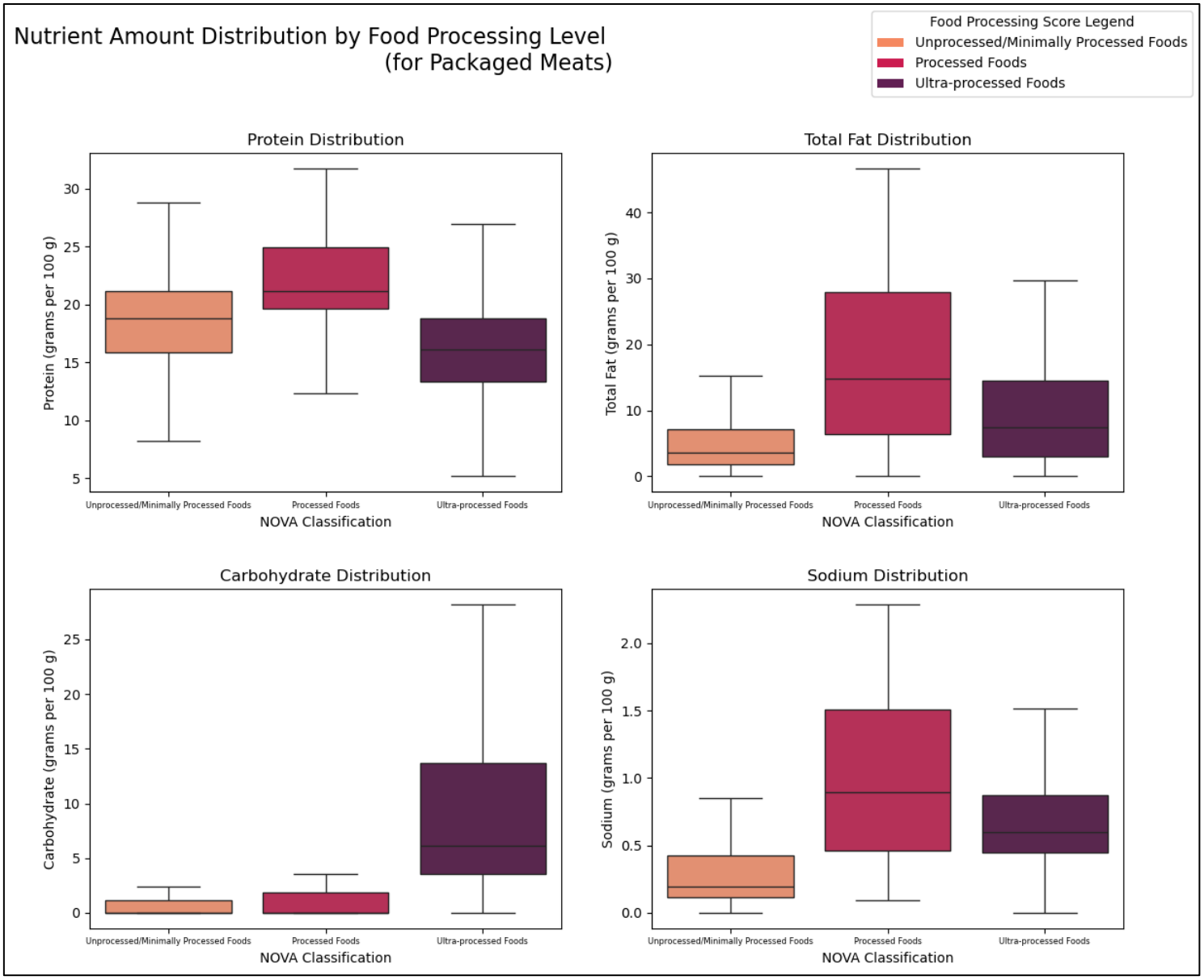
My main goal going into the project was to show the audience different distributions of nutritional content depending on how processed the food is; notably choosing a boxplot in understanding my prospective audience would be one with an academic background. While I initially created boxplots as below:



I realized that it wasn't a fair representation of the data, it was a comparison across all food categories which then resulted in wildly distributed boxplots. Additionally given all the different nutrients to be shown and their different scales, it was busy to look at. I then looked to hone in on only packaged meats (so that I could compare less processed ones to their alternative option) and separated the box plots. I also chose protein, fat, carbohydrate, and sodium as key nutrient markers to analyze, though it is scalable for more.

I removed duplicates from the data and wanted to focus on the IQR/median region by using a boxplot to which was more robust against outliers. The data was then pivoted in pandas and plotted across subplots using seaborn and matplotlib. I chose a palette that implied low-to-high for the amount of processing a food category was (darker being more processed). I think my personal design choice would be to omit the x-axis labels and ticks and using only the legend for reference for better at a glance viewing.

The resultant set of plots with the chosen nutrients show that processed foods and ultra-processed foods (in this case being packaged meats) tend to have elevated fat, carbohydrate, and sodium than those unprocessed. I chose those as they are often the ones that are mentioned in media to be needed to be controlled in moderation, with the protein chart to serve as a sort of control. The choices I made in portraying this plot, while correlative and observational, are able to push the narrative towards food that is more processed being more excessive in nutrients and perhaps unhealthy.



Dataset Source: TrueFood.Tech (grocerydb.csv)