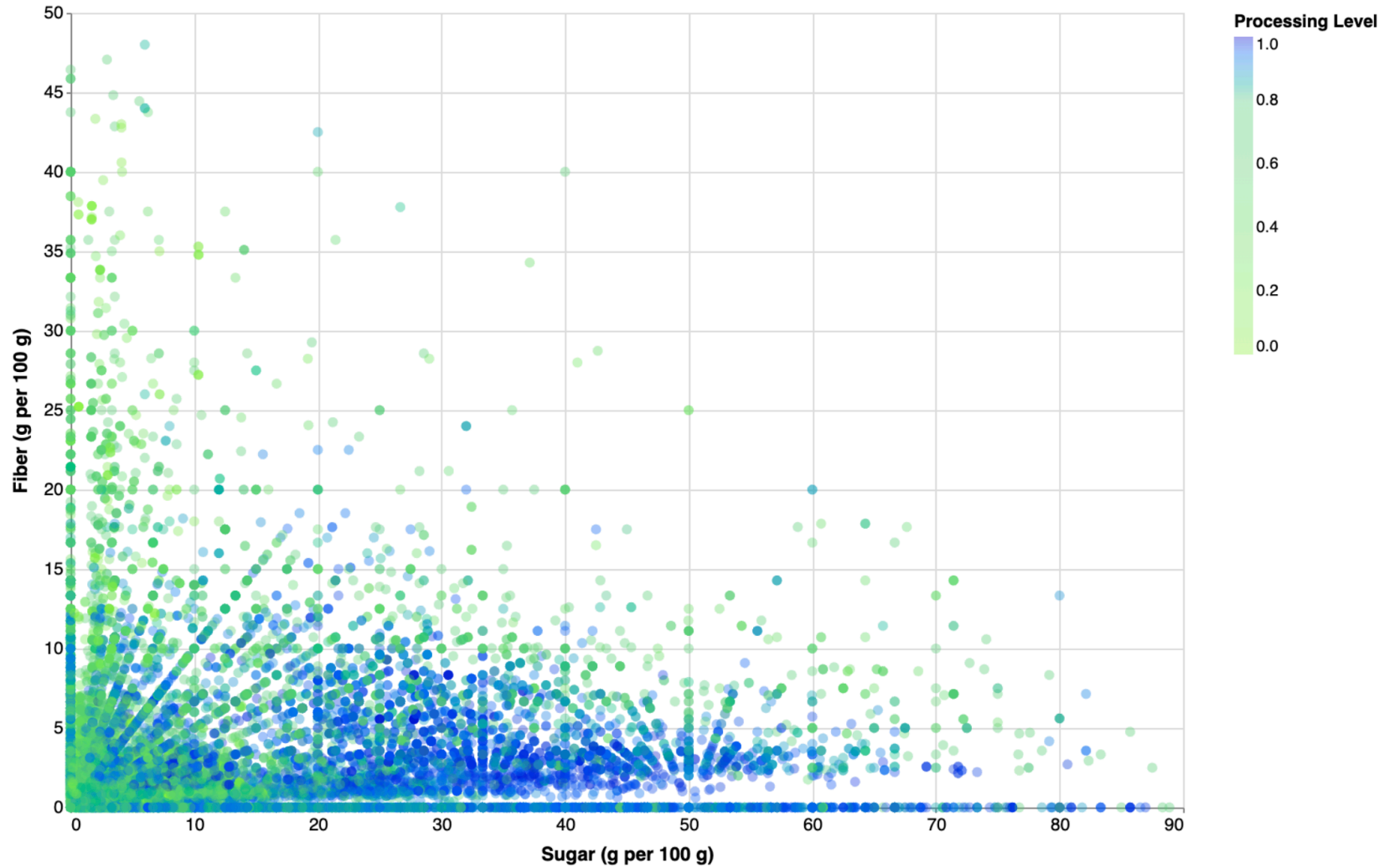


## The Nutrition Trade-Off: Why Processing Level Predicts Nutritional Quality

From fresh vegetables to packaged snacks, each level of processing adds sugar while stripping away dietary fiber



\* n = 25,670 • Removed Implausible, Missing & Extreme Data • Exclude 0-kcal beverages

Source: GroceryDB, 2025

This visualization shows that ultra-processed foods tend to have high sugar and low fiber, while minimally processed foods tend to have higher fiber and lower sugar. The scatterplot shows sugar (g per 100 g) on the x-axis and dietary fiber (g per 100 g) on the y-axis, with each point colored by processing level (FPro) using a green-to-blue gradient, which lighter greens for minimally processed foods and darker blues for ultra-processed products.

For the Data Preparation, I removed rows with missing key nutrients, filtered out zero-calorie beverages (water and unsweetened drinks with all macros equal to zero), and excluded implausible records in which sugar or fiber exceeded carbohydrates, or in which the sum of protein, fat, and carbohydrates exceeded 110 g per 100 g. To improve readability in the dense regions where most foods cluster, I trimmed only extreme outliers (keeping sugar below 90 g and fiber below 50 g), which affects fewer than 1% of records and does not alter the underlying trend. The final sample is 25,670 items.

I tried the square-root scales, but ultimately chose the linear axes because sugar and fiber share the same natural units across all products, making distance comparisons clear and maintaining the visual slope of the relationship. Processing level is mapped continuously to color instead of grouped into categories. In the original concept, I considered adding a fourth encoding dimension (point size for calorie density) and overlay elements such as smoothing curves, but I removed them to keep the visualization simple and avoid requiring viewers to interpret additional visual channels or statistical models.

Also, I selected the  $FPro \geq 0.8$  as the onset of the ‘ultra-processed’ region (starting of the transition from green to blue) because, in published applications, truly ultra-processed items cluster near the high end of the FPro spectrum (for example,  $FPro \approx 0.96, 0.995$  for snack and bread items). The continuous scale of FPro is defined from 0 (unprocessed) to 1 (ultra-processed), and examples of food items already considered ultra-processed in the literature tend to lie well above 0.8. Using 0.8 gives a conservative buffer, capturing the upper tail of processing intensity while avoiding misclassification of moderately processed foods as “ultra-processed.”