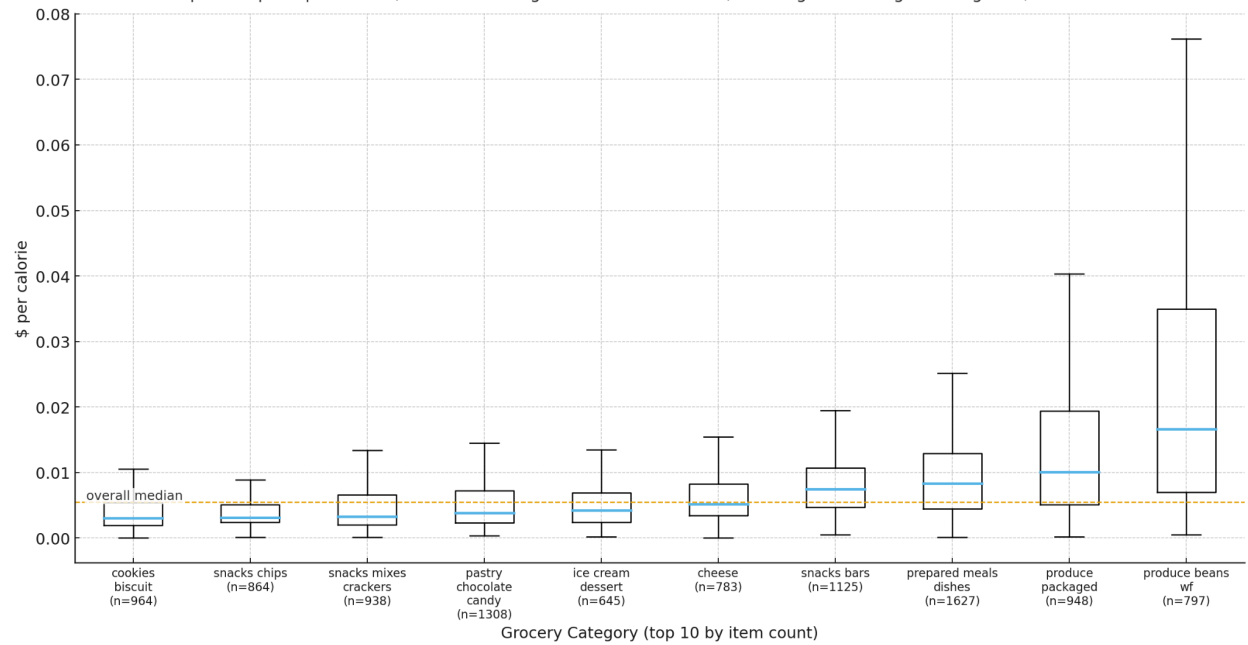


Which grocery categories offer the cheapest calories?

Boxplots of price per calorie; filtered to categories with ≥ 400 items; showing the 10 largest categories; outliers hidden



Visualization Write-Up

Motivation and Question

This visualization explores the question: *Which grocery categories offer the cheapest calories?* The goal was to identify which types of food provide the most calorie value per dollar to help us discern how prices differ across processed snacks, dairy, and fresh produce. I went with a boxplot because it efficiently summarizes the distribution across categorical groups while showing variation, medians, and potential skew.

Data Transformations and Rationale

The raw dataset contained over 26,000 grocery items from multiple stores and categories. To ensure meaningful and stable comparisons:

1. **Selecting the top 10 categories by item count:** Restricting to the 10 largest groups helped me keep the visualization legible and prevented overcrowding of category labels. The top 10 still captures a broad selection of the grocery choices.
 2. **Sorting by median price per calorie:** I ordered the categories left-to-right by their median values so the reader can quickly identify which categories are cheaper or more expensive.
 3. **Suppressing individual outliers:** Outliers were hidden to reduce overplotting and focus on general trends.
 4. **Overall median reference line:** I added a dashed line for the global median of price per calorie to provide a consistent benchmark across all categories.
-

Design Decisions

My visualization uses position along a common scale to encode price per calorie to allow direct quantitative comparison. The boxes and whiskers convey distribution shape, while I included *n* values beneath each category to communicate the sample size. The overall median line helps provide additional orientation and context. Finally, outlier suppression, consistent axis scaling, and tick labels improve readability and help make the plot more readable.

Takeaway

From the final chart, it becomes clear that heavily processed “snack-like” categories (like cookies, chips, crackers, and candy) tend to deliver calories at lower cost compared to fresh produce or packaged beans, which occupy the higher end of the price-per-calorie spectrum.