Week 3

Pedro Henrique Brant

2025-07-25

This report explores potential exposure variables related to dietary patterns in the NHANES dataset. All of the variables are categorical.

We begin by loading the necessary libraries:

Next, we read in the cleaned and merged dataset, which combines the variables chosen by Professor Fregni and the ones chosen by the students.

To make our dietary variables more readable and meaningful, we recode them from numeric codes to labeled factors. This includes whether the person is currently on a diet (`drqsdiet`) and several types of diets such as weight loss, low fat, low salt, etc.

To evaluate internal consistency, we created a derived variable `any_specific_diet_flag` that indicates whether participants reported following at least one specific diet. We then cross-tabulated this with the general question on whether the participant was currently on any diet. This allows us to identify potential discrepancies—such as individuals who reported a specific dietary pattern but did not indicate they were currently on a diet. The table below summarizes all combinations of these two variables:

To begin, we explore the general question about whether participants are currently following a special diet. The variable `currently_on_diet` captures this information using labeled categorical responses. The plot and table below display the distribution of responses, including those who responded “Yes”, “No”, “Don’t know”, or left the question unanswered.

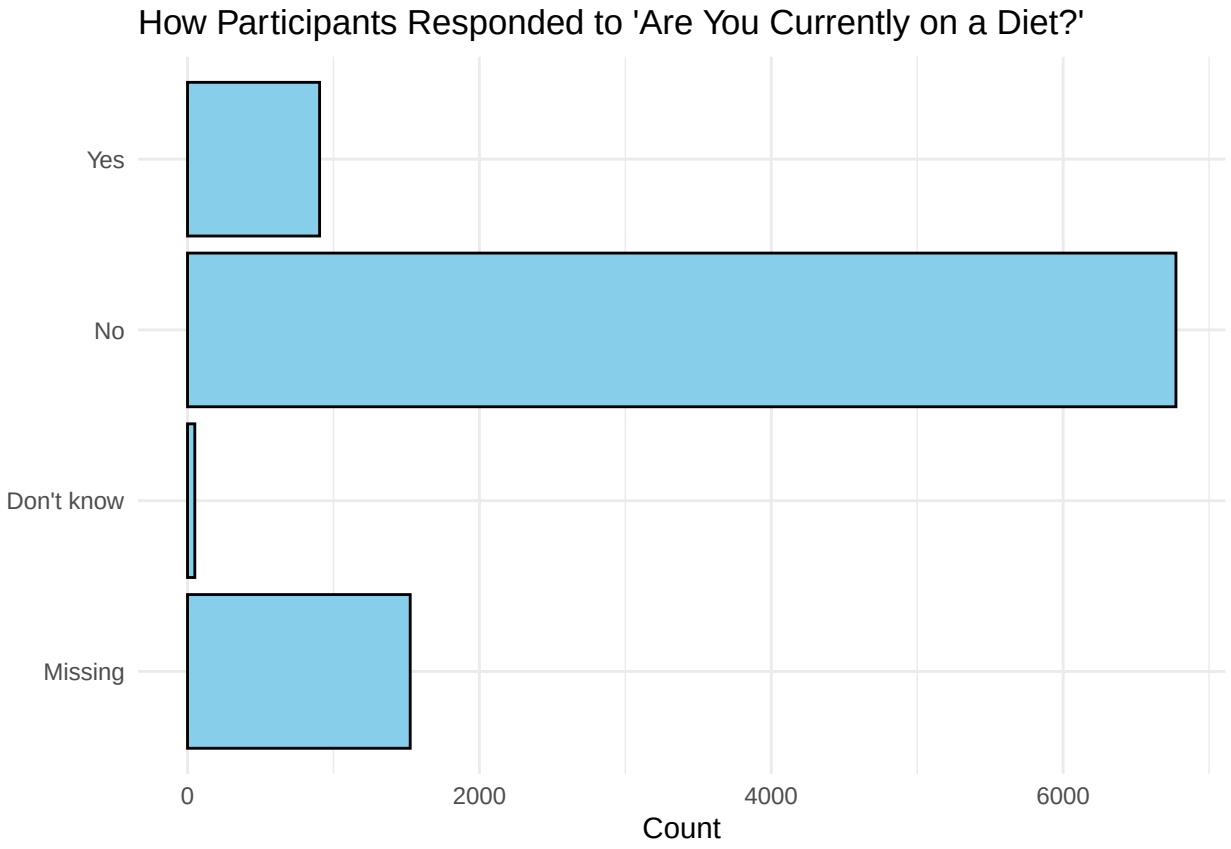
Consistency Between Diet Flags

Are there any patients flagged to be on a specific diet that didn't show up as being currently on a diet?

Currently on diet	Following any specific diet	Count	Percent
Yes	Yes	905	9.8
Yes	No	0	0.0
No	Yes	0	0.0
No	No	6773	73.2
Don't know	Yes	0	0.0
Don't know	No	50	0.5
Missing	Yes	0	0.0
Missing	No	1526	16.5

How Participants Responded to 'Are You Currently on a Diet?'

Response	Frequency	Percent (%)
Yes	905	9.8
No	6773	73.2
Don't know	50	0.5
Missing	1526	16.5



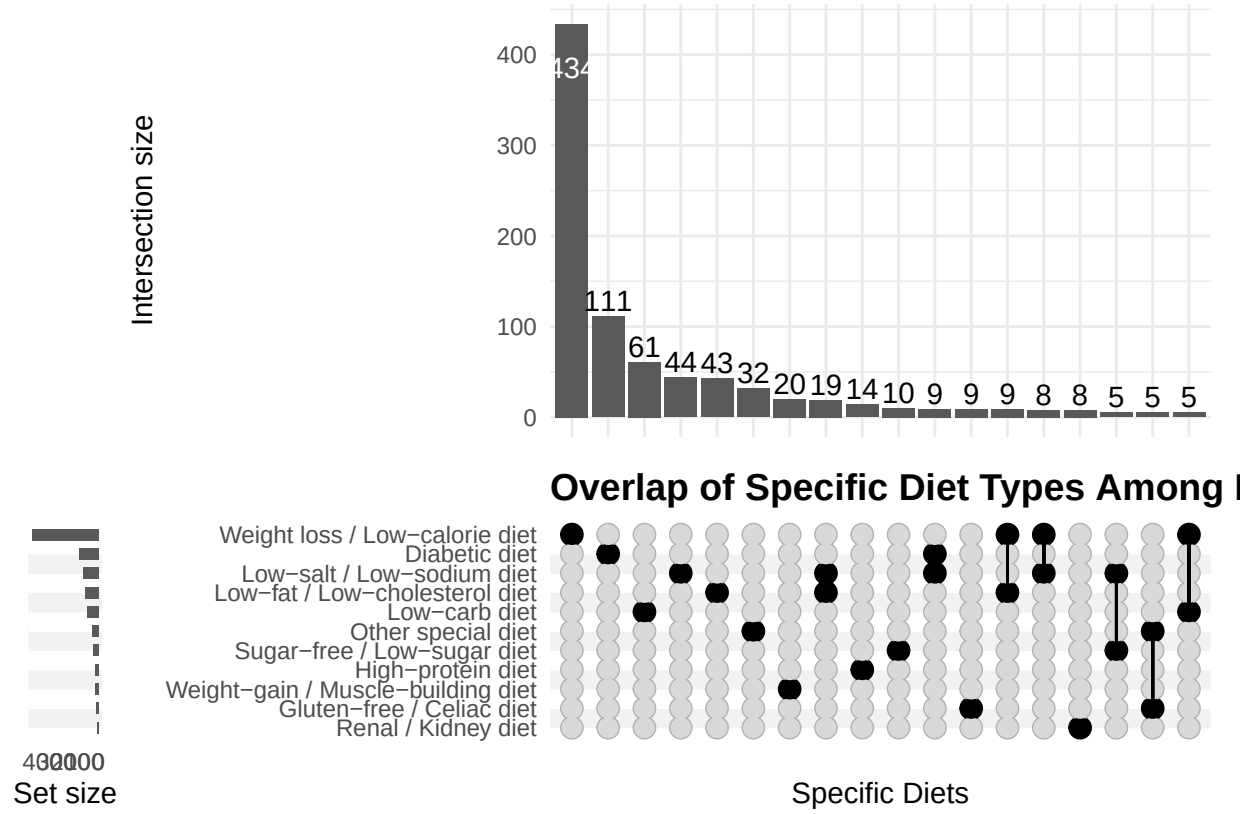
To explore the distribution of special diets among respondents, we first calculate the total number of

Absolute and relative frequencies of specific diets

Following a diet?	Frequency	Percent (%)
Yes	905	9.8
• *Weight loss/Low calorie diet*	475	5.1
• *Diabetic diet*	139	1.5
• *Low salt/Low sodium diet*	109	1.2
• *Low fat/Low cholesterol diet*	95	1.0
• *Low carbohydrate diet*	85	0.9
• *Other special diet*	44	0.5
• *Sugar free/Low sugar diet*	40	0.4
• *High protein diet*	26	0.3
• *Weight gain/Muscle building diet*	23	0.2
• *Gluten-free/Celiac diet*	16	0.2
• *Renal/Kidney diet*	10	0.1
• *High fiber diet*	4	0.0
• *Low fiber diet*	2	0.0
No	6773	73.2
Don't know	50	0.5
Missing	1526	16.5

individuals currently following a diet, followed by a breakdown of the specific types of diets they report. We also summarize responses from those not currently on a diet, those who answered "Don't know," and those with missing data. The resulting table presents both absolute and relative frequencies, formatted for publication using the `gt` package. Specific diets are displayed as indented, italicized subcategories beneath the "Yes" group for clarity.

To visualize the overlap and combinations of different specific diets among participants who reported following a special diet, we used an UpSet plot. This type of plot offers a clear summary of how frequently participants reported one or more particular dietary patterns, as well as which combinations are most common.



This plot shows the most common individual diets on the left, and the set intersections (i.e. combinations of diets) along the bottom, with bar heights indicating the number of participants in each intersection.