Final Project Part 1

36-617A: Applied Linear Models

Vinay Maruri

Carnegie Mellon University, 12/08/2023

Contents

1 Data and EDA **2**

1.1 Overview 2

1.2 Question 1 2

1.3 Question 2 3

1.4 Question 3 4

1.5 Instructions for further analysis 4

1 Data and EDA

1.1 Overview

In the first phase of this project, I intend to investigate three questions that were asked to us by the Assistant Minister of Health from Vietnam. First, I will summarize how the Vietnamese people rate the attractiveness, impressiveness, sufficiency, and popularity of the information they receive in annual health exams administered. Second, I will explore what factors make a person less likely to get a annual health checkup. Finally, I will assess the Deputy Assistant Minister's claim that the only factors that influence the probability of a Vietnamese person getting a checkup is if they have health issues or know someone in their family that has health issues.

To conduct this first phase, I will be working with data that was published by [1]. This dataset is comprised of 2068 observations from a total of 2479 people that were approached for the survey (Response rate: 83.3%, 2 rejected observations). Each of these people were randomly sampled from the population of Hanoi and Hung Yen, Vietnam for the survey, and then interviewed for 12 to 15 minutes by a survey team member. In total, there are 48 data variables and 2 identifying variables for each observation.

1.2 Question 1

In terms of how patients rate the attractiveness, impressiveness, sufficiency, and popularity of the information they receive in checkups, where 1 is the lowest rating and 5 is the highest rating, patients on average give these factors a below average rating (see Figure 1). The average sufficiency rating was 3.01 (sd = 1.16, 95% CI: 0.69-5), the average attractiveness rating

was 2.69 (sd = 1.15, 95% CI: 0.39-4.99), the average impressiveness rating was 2.82 (sd = 1.14, 95% CI: 0.54-5), and the average popularity rating was 2.79 (sd = 1.18, 95% CI: 0.43-5).

In terms of how Vietnamese doctors and nurses are doing at communicating this information to patients, it's appropriate to measure how reliable they sound, how responsive they are to the patient and whether the patient feels empathy and assurance from the medical provider when this information is discussed (see Figure 2).
The average reliability rating for doctors was 3.57 (sd = 1.08, 95% CI: 1.41-5), the average assurance rating for doctors was 3.69 (sd = 1.09, 95% CI: 1.51-5), and the average empathy rating was 3.47 (sd = 1.25, 95% CI: 0.97-5).

1.3 Question 2

In terms of the factors that make a person less likely to get a checkup every twelve months, they are as follows based on my visual analysis of the data. A person is less likely to get a checkup every twelve months if they don't have health insurance, are male, do not have stable job status, are unmarried, are not from Hanoi, listed worrying symptoms as the reason for their last checkup, believes that checkups are a waste of time or a waste of money, believes that checkups are not urgent or important, does not have their checkup subsidized by an employer or community, does not have a habit of getting regular checkups, does not follow constant updates on their health measures, never received long term medical treatment, their preferred way of dealing with new symptoms is not to go to a clinic, and their general opinion of public health is bad. I determined that these are important by seeing from the plot that the proportion of responses in this category were different between people that had less than

12 months since their last checkup and other people (see Figures 3-17).

1.4 Question 3

To understand whether the Deputy Assistant Minister's hypothesis held any merit, I repeated the analysis that I did in question 2, but added an additional layer of aggregation on the stable health status variable and whether the patient was having health issues. We know from question 2 that a person is less likely to get a checkup every twelve months if they have never received long term medical treatment, and from the available data this is the only available gauge of a person having ongoing health issues. In terms of whether someone in their family is having health issues has an effect on their likelihood to get a checkup every twelve months, this does not seem like an important factor (see Figure 18). Within the groups of people who get a annual checkup every 12 months, 12-24 months, and every more than 24 months, the proportion of people who said they and their family were in good health or not is relatively equal across each of the groups.

Without further statistical analysis, I must conclude that only the first part of the Deputy Assistant Minister's hypothesis holds merit (that these factors don't matter much unless you have health issues), and the second part about knowing that someone in their family has health issues does not hold merit.

1.5 Instructions for further analysis

The additional statistical analyses that should be conducted for question 2 are as follows:

- Compare the frequencies of these explanatory variables across each of the time since

last visit groups and test whether their proportions are different from expected using a chi-square test.

- To determine the most important factors, this could be established with binomial GLMs (in the case of categorical data) and linear models (in the case of numeric data) of the explanatory variable on the time since last visit groups.

- A simpler analysis would involve computing correlations between the explanatory variables (transformed into numeric factors) and the time since last visit groups (again transformed into a numeric factor)

The additional statistical analyses that should be conducted for question 3 are as follows:

- ANOVA should be conducted with a two-way interactive model to determine if there is an interactive effect of having health issues/knowing someone in the family has health issues and the explanatory variable on time since last visit groups.

- In a similar vein, it would also be useful to generate the described interaction terms and fit binomial GLMs/linear models to determine the significance of the interaction terms and the explanatory variables combined on the time since last visit groups.

	N	Mean	Standard Deviation	Median	Minimum	Maximum
Sufficiency	2068	3.012573	1.165371	3	1	5
Attractiveness	2068	2.690522	1.145010	3	1	5
Impressiveness	2068	2.818085	1.140770	3	1	5
Popularity	2068	2.797727	1.175705	3	1	5

Figure 1: Summary of patient feelings on information given in annual health checkups

	N	Mean	Standard Deviation	Median	Minimum	Maximum
Reliability	2068	3.571663	1.077370	4.0	1	5
Assurance	2068	3.694778	1.090284	4.0	1	5
Empathy	2068	3.469584	1.253828	3.5	1	5

Figure 2: Summary of patient feelings on how the doctor sounded in annual health checkups

References

[1] QH Vuong. "Survey data on Vietnamese propensity to attend periodic general health examinations". In: *Scientific Data* 4.170142 (2017). DOI: `https://doi.org/10.1038/sdata.2017.142`.

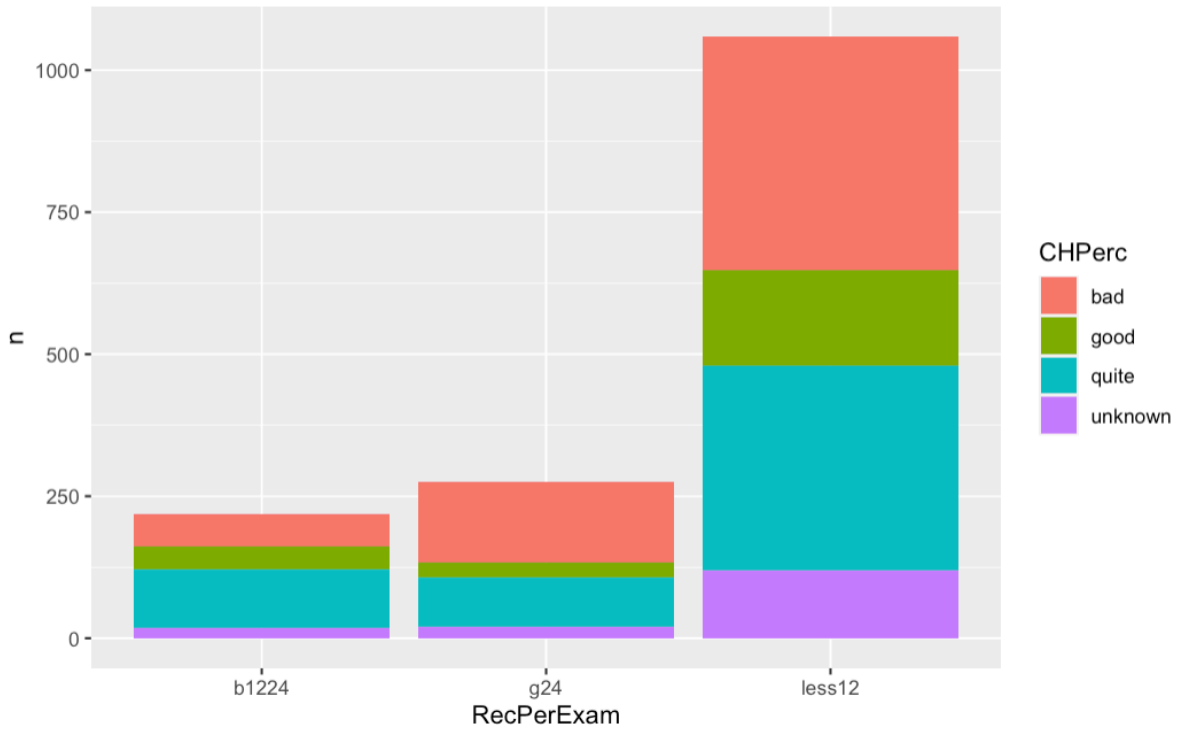

Figure 3: Summary of patient opinions on public health and annual checkup frequency

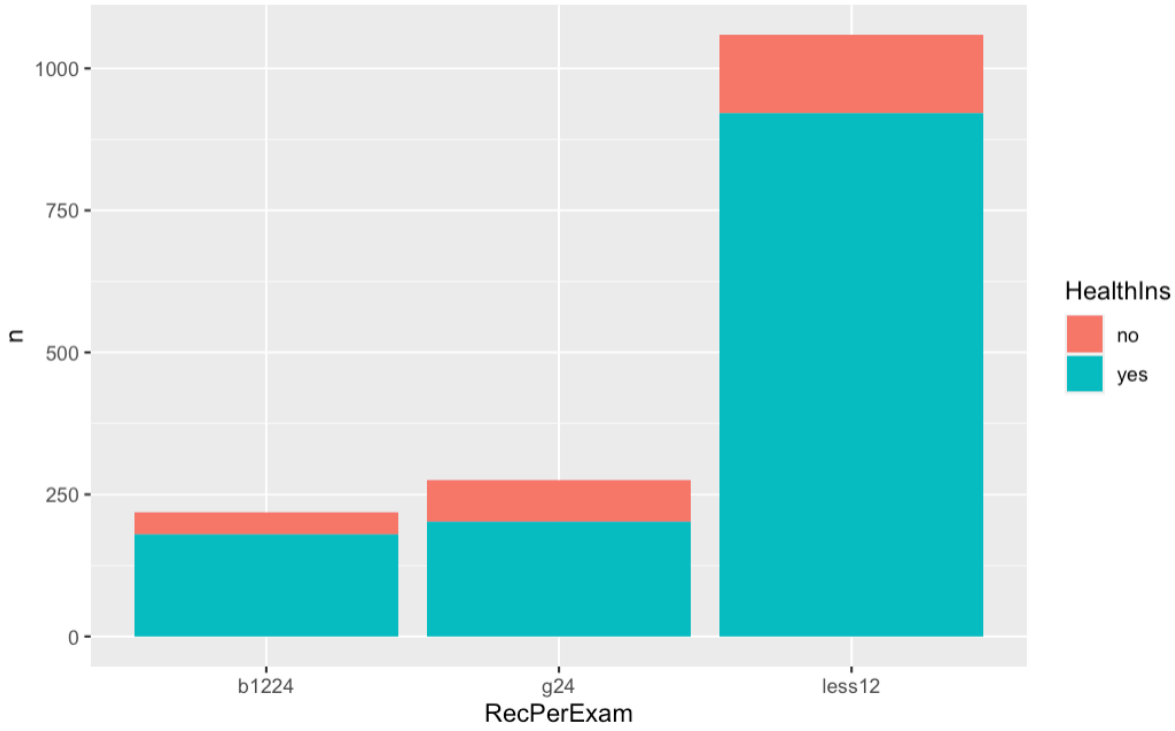

Figure 4: Summary of patient's health insurance status and annual checkup frequency

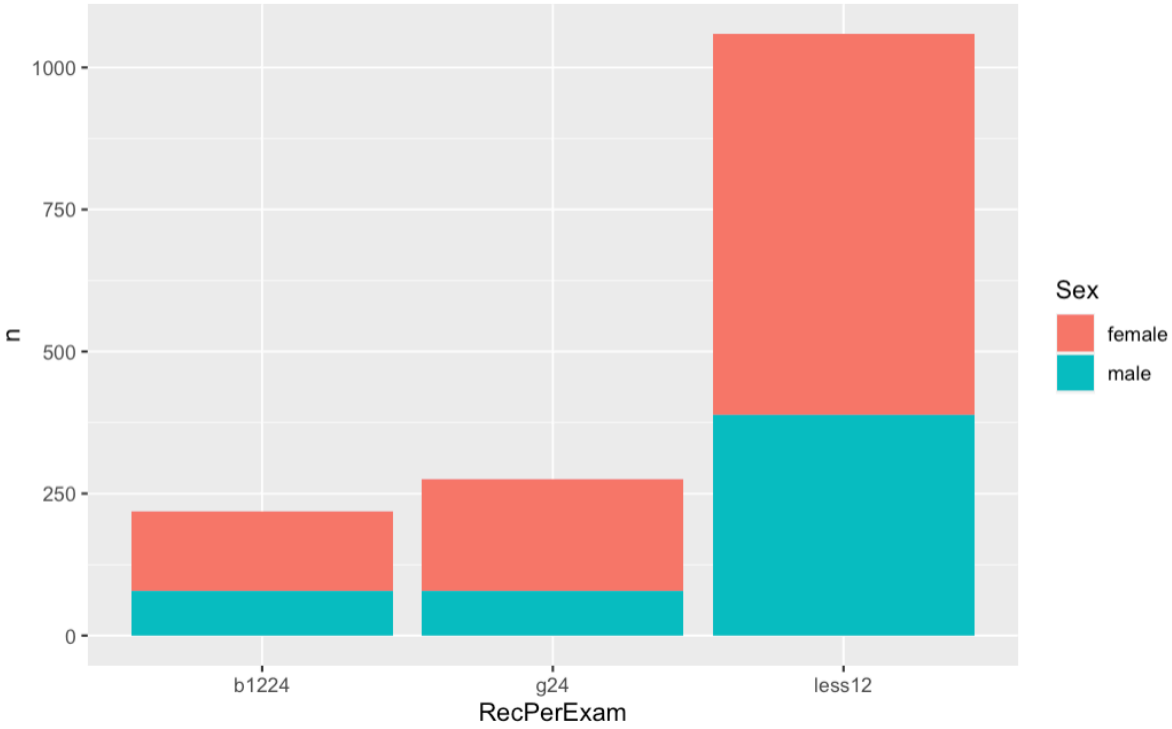

Figure 5: Summary of patient's sex and annual checkup frequency

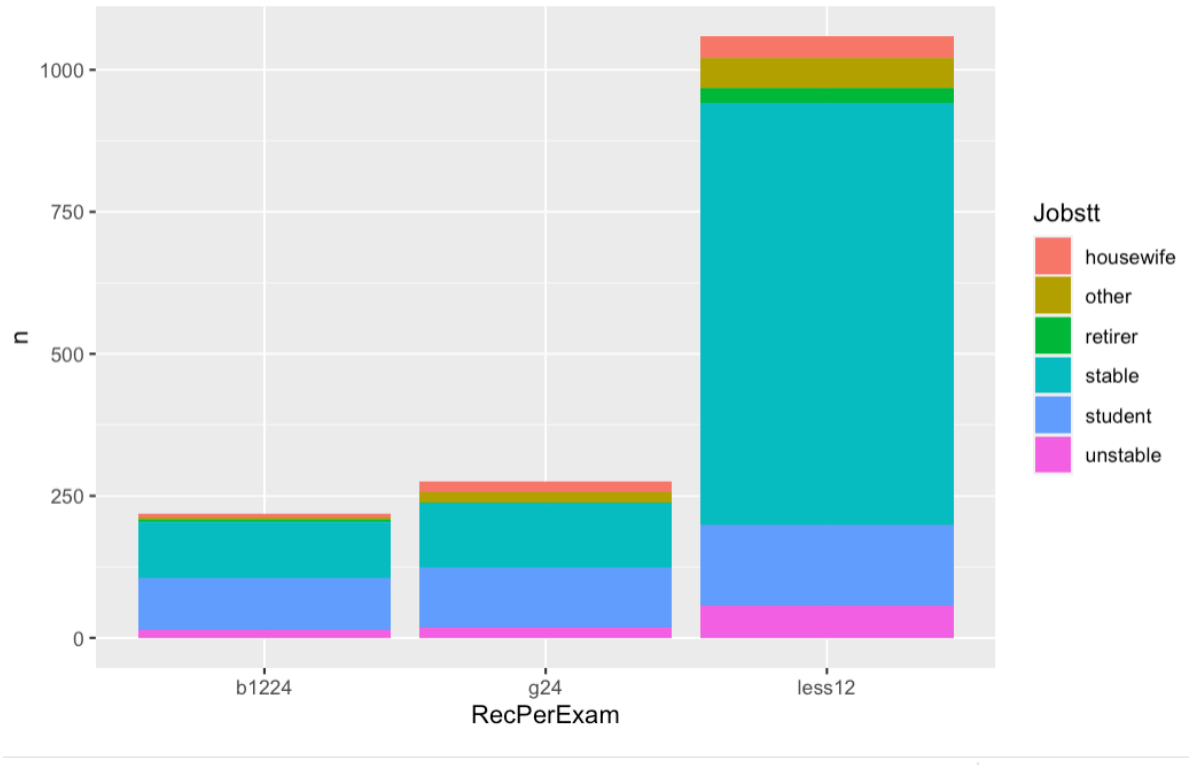

Figure 6: Summary of patient's job status and annual checkup frequency

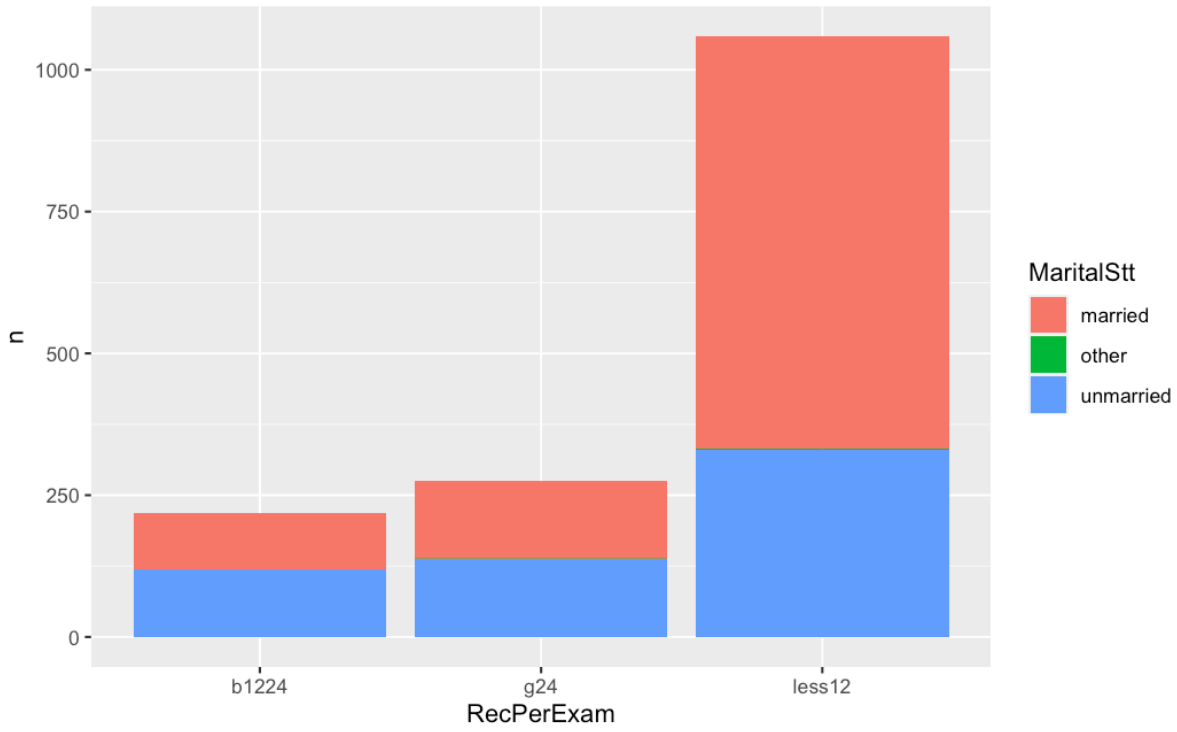

Figure 7: Summary of patient's marital status and annual checkup frequency

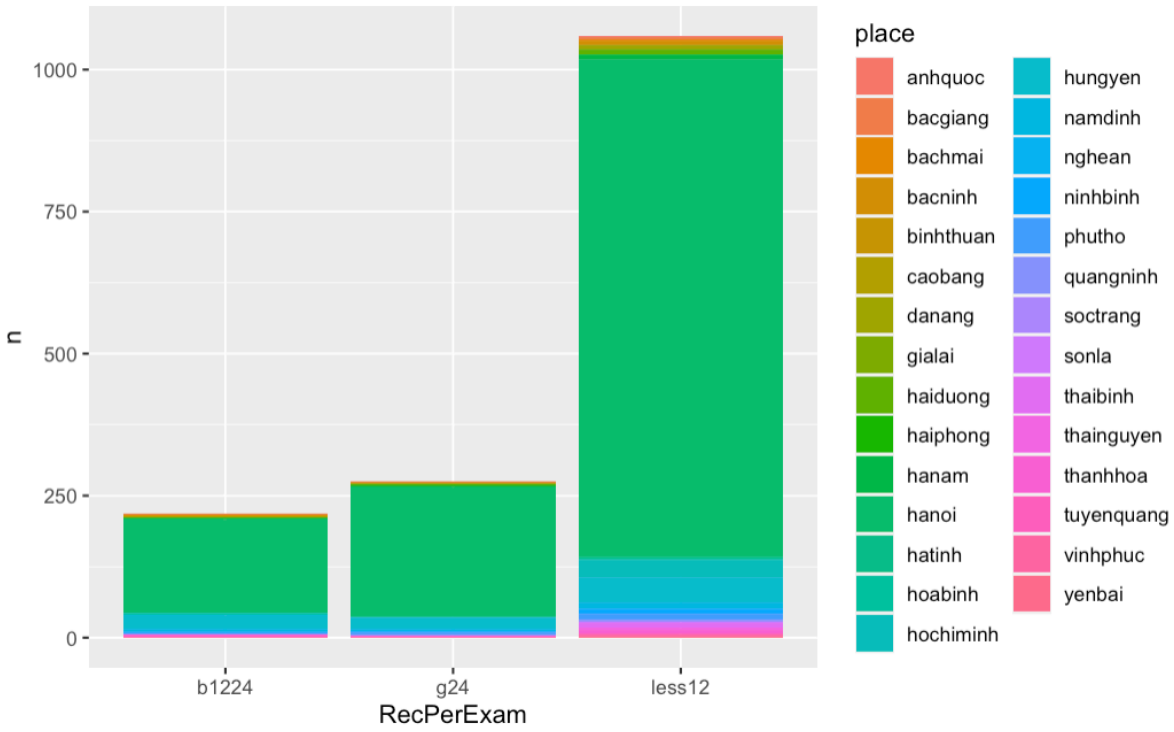

Figure 8: Summary of patient's interview location and annual checkup frequency

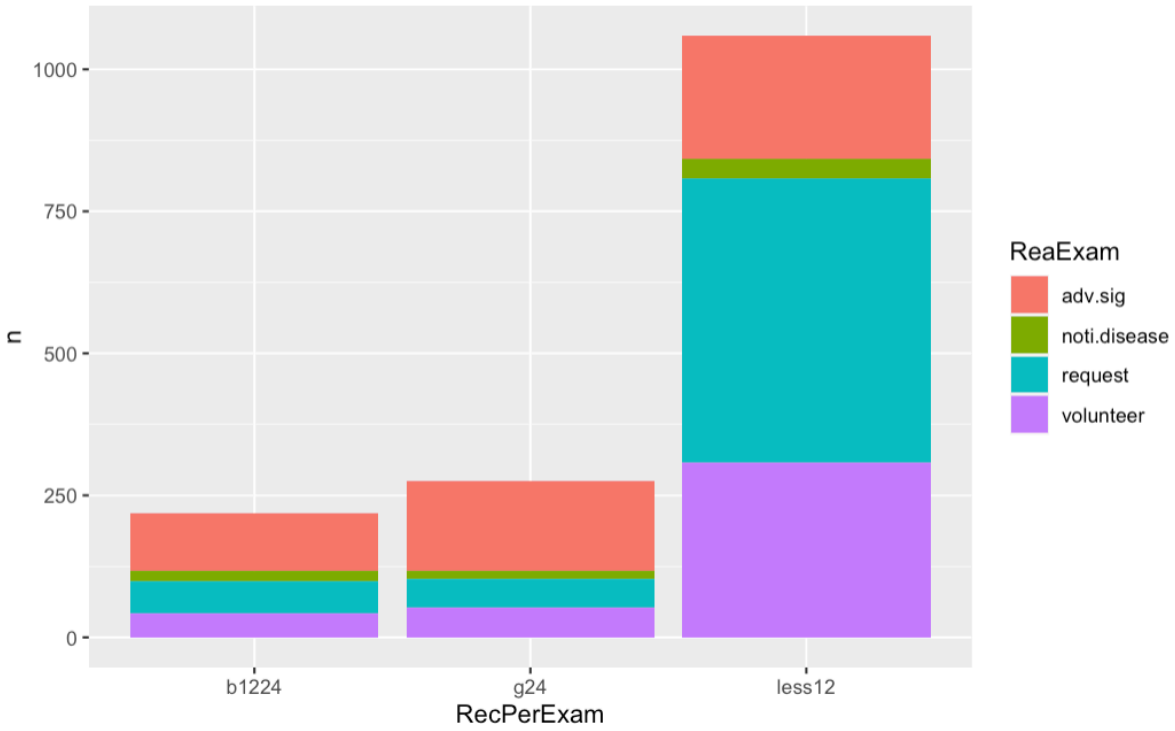

Figure 9: Summary of patient's reason for last exam and annual checkup frequency

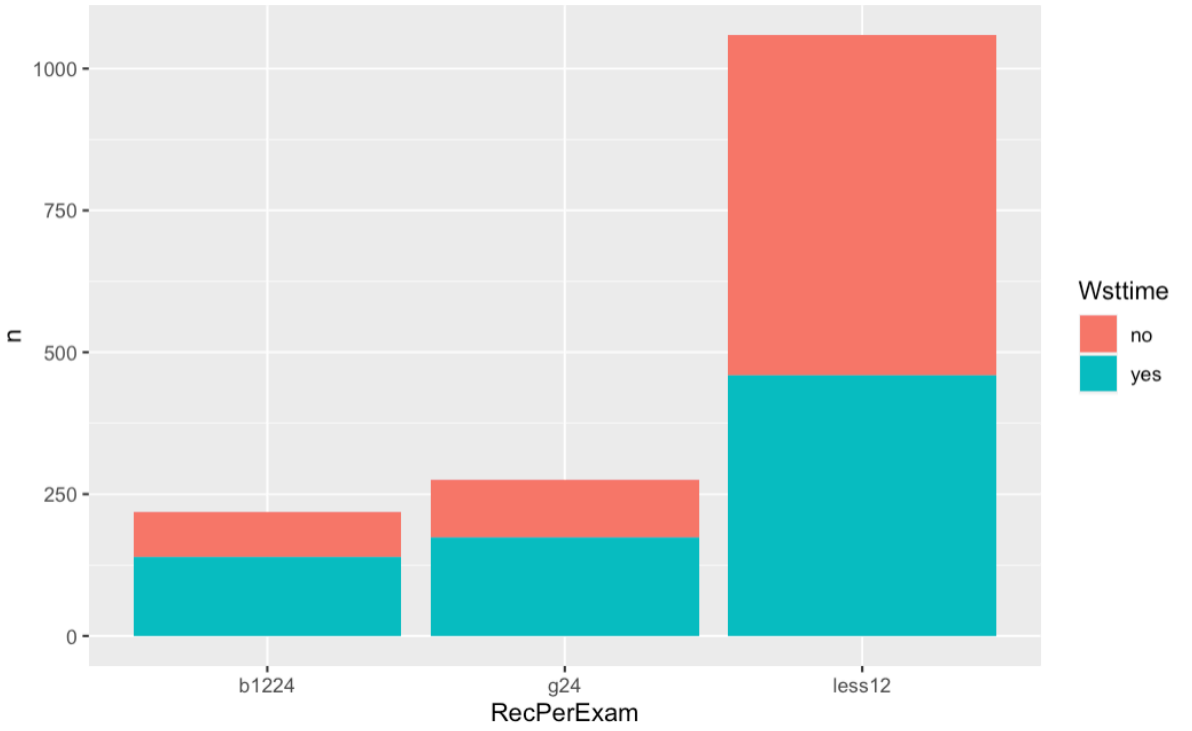

Figure 10: Summary of patient feelings on whether checkups are a waste of time and annual checkup frequency

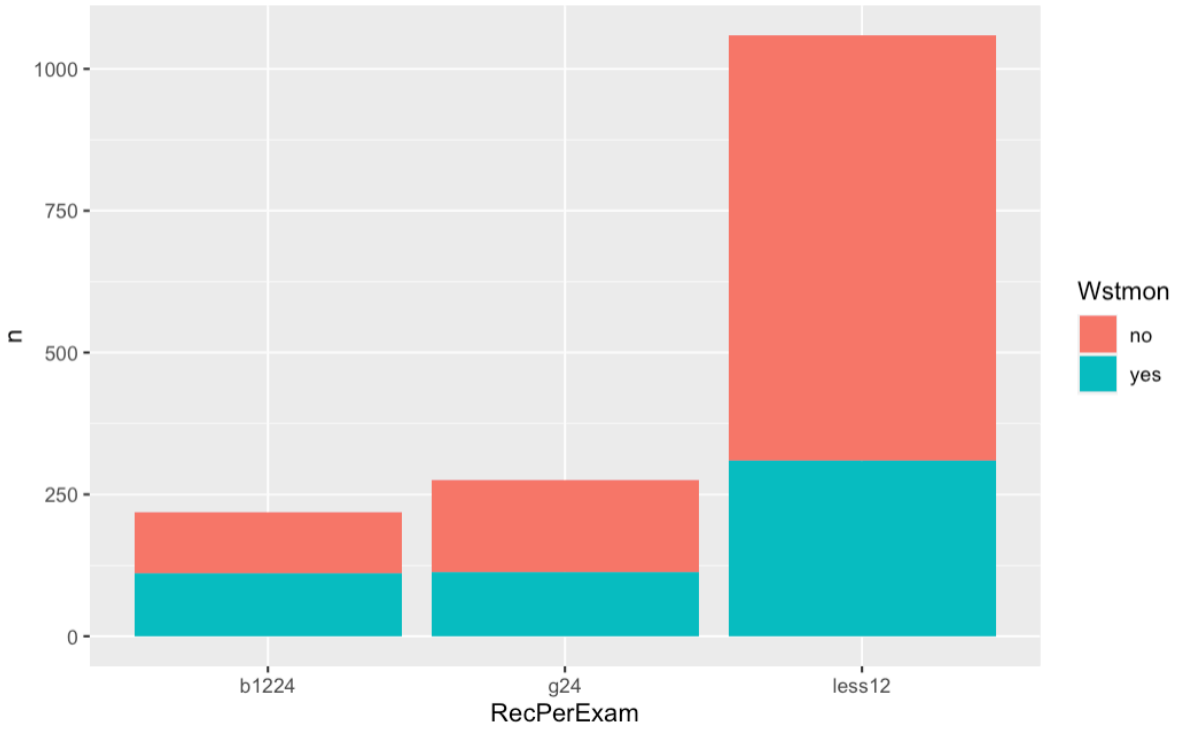

Figure 11: Summary of patient feelings on whether checkups are a waste of money and annual checkup frequency

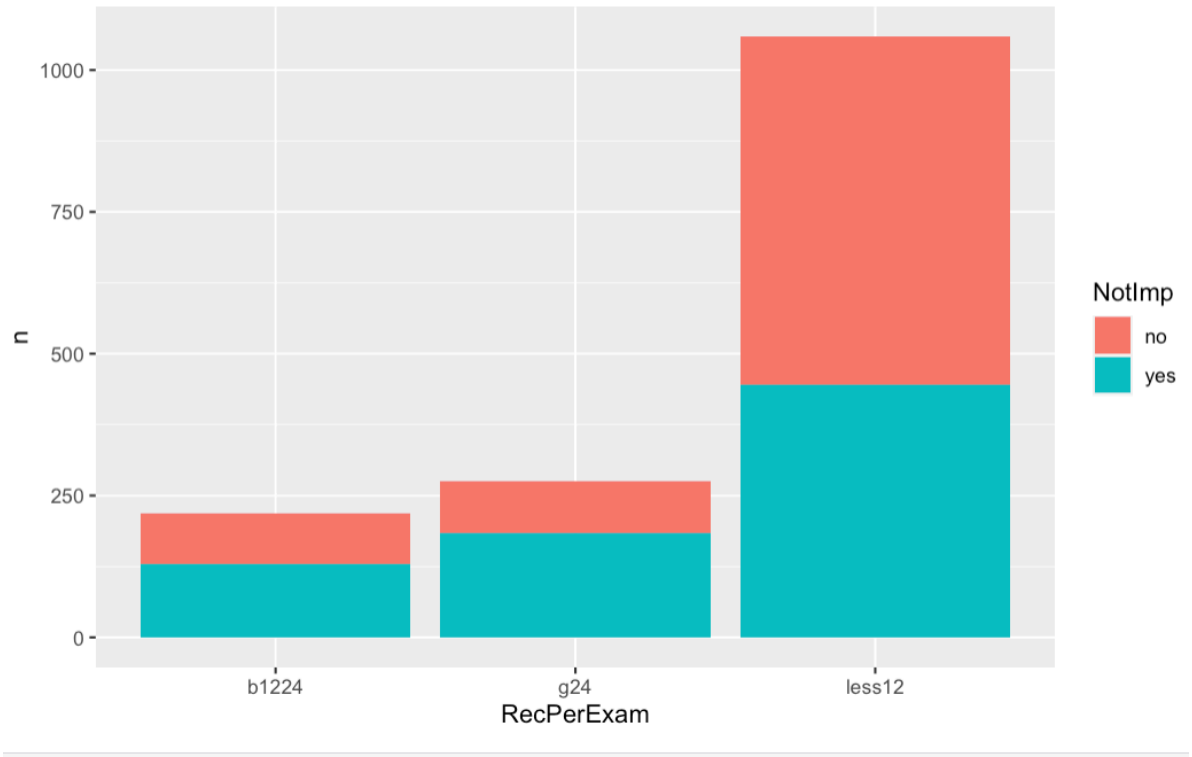

Figure 12: Summary of patient feelings on whether checkups are not important and annual checkup frequency

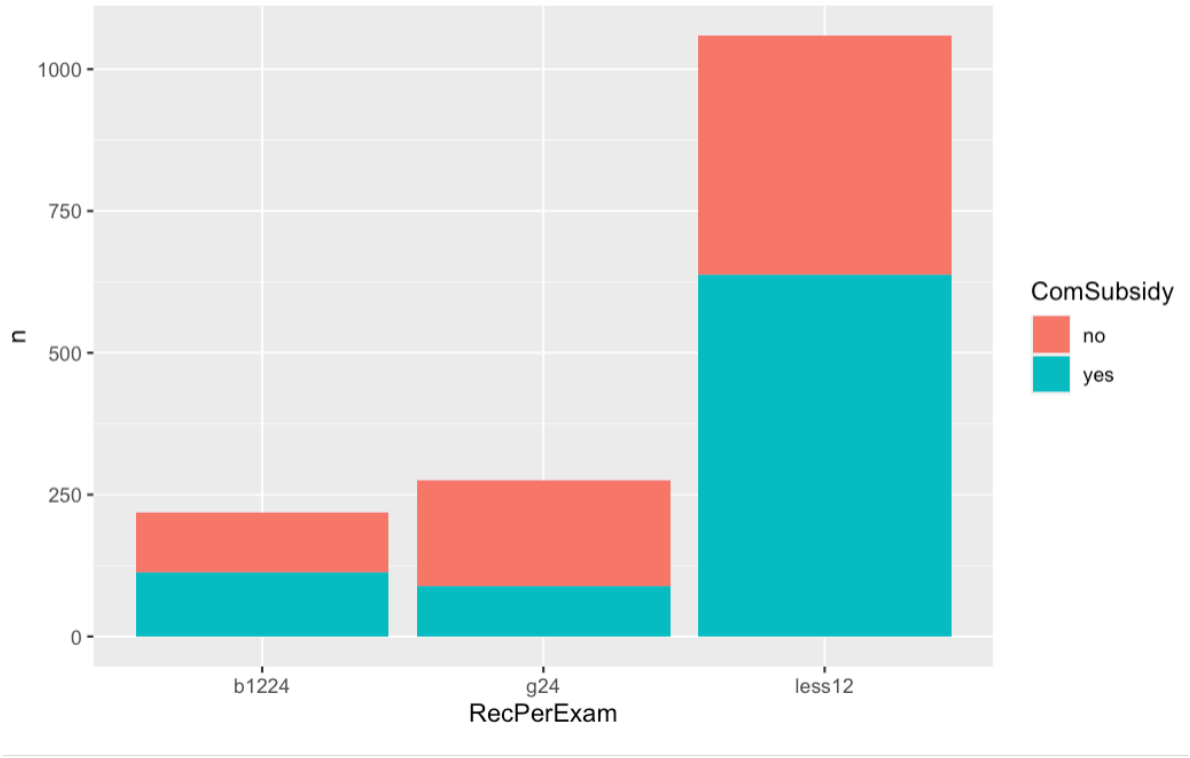

Figure 13: Summary of whether patients' checkups are subsidized and annual checkup frequency

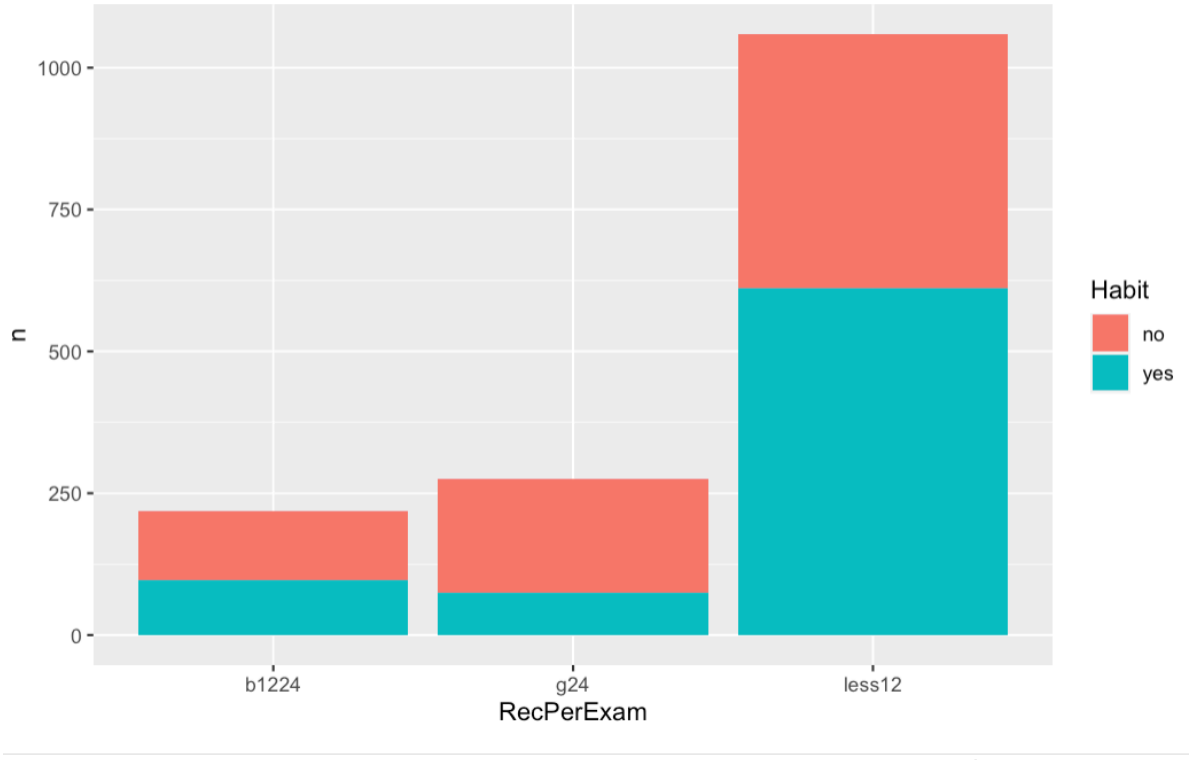

Figure 14: Summary of whether a patient has the habit of getting a checkup and annual checkup frequency

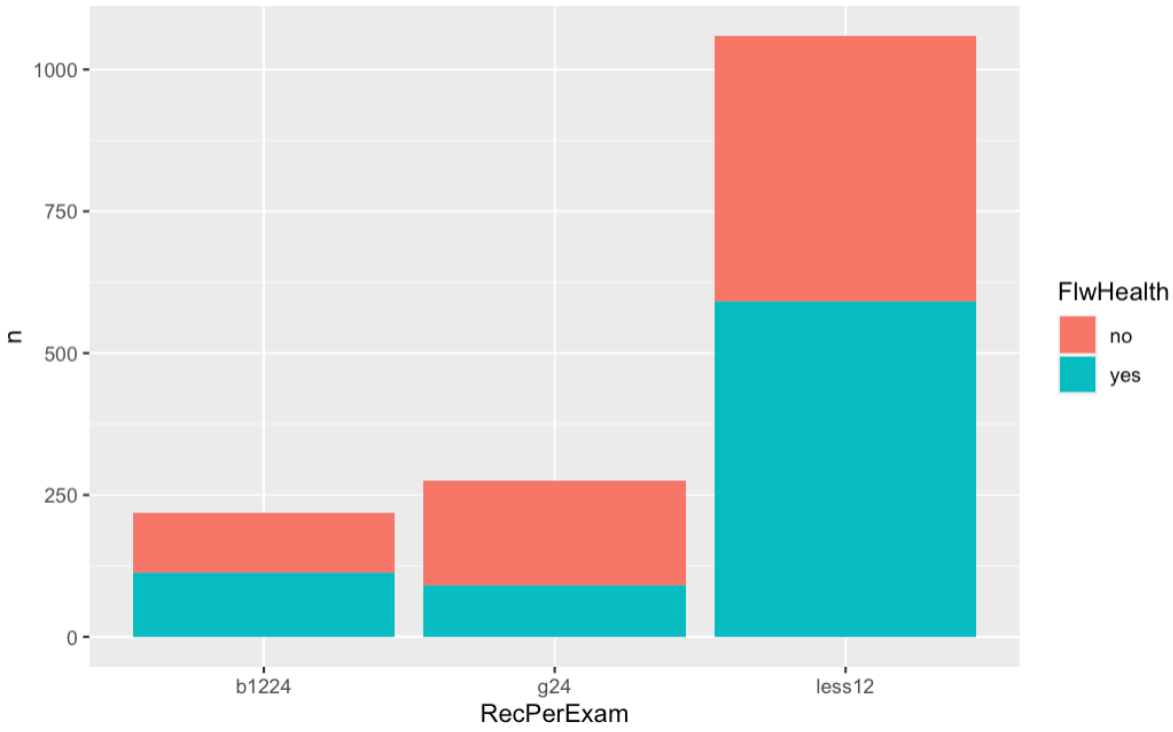

Figure 15: Summary of whether a patient follows updates on their health measures and annual checkup frequency

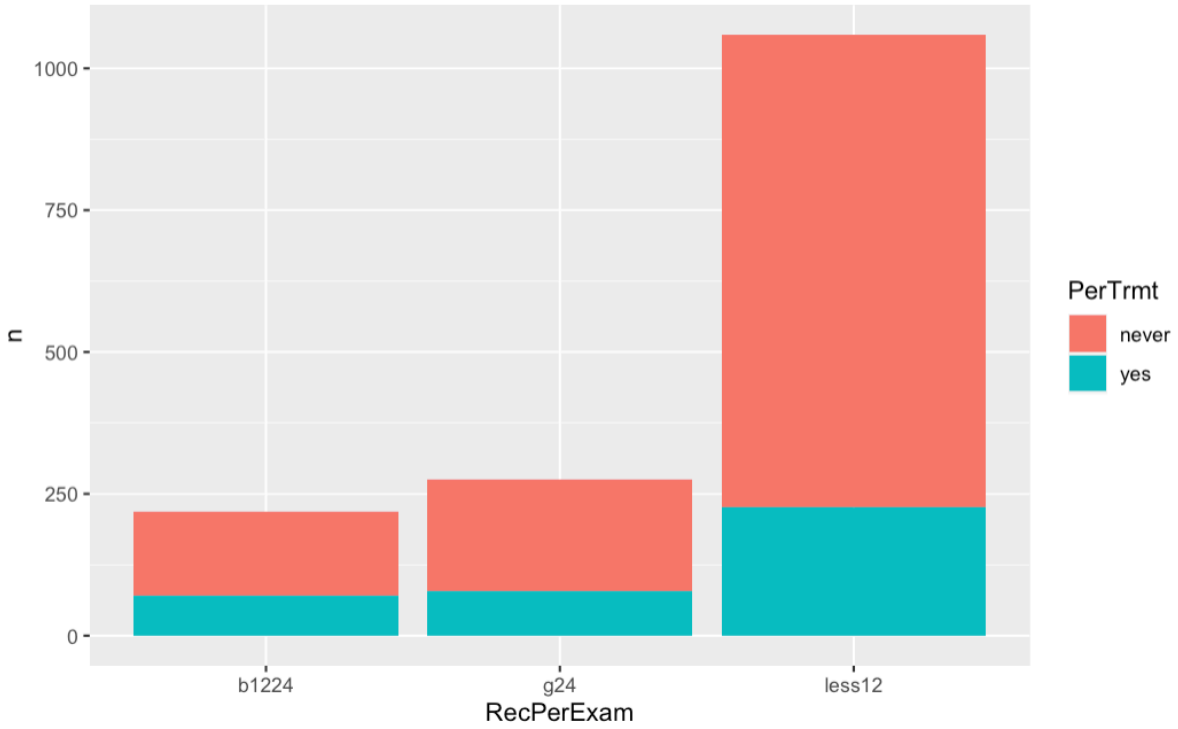

Figure 16: Summary of whether a patient is receiving long-term medical treatment and annual checkup frequency

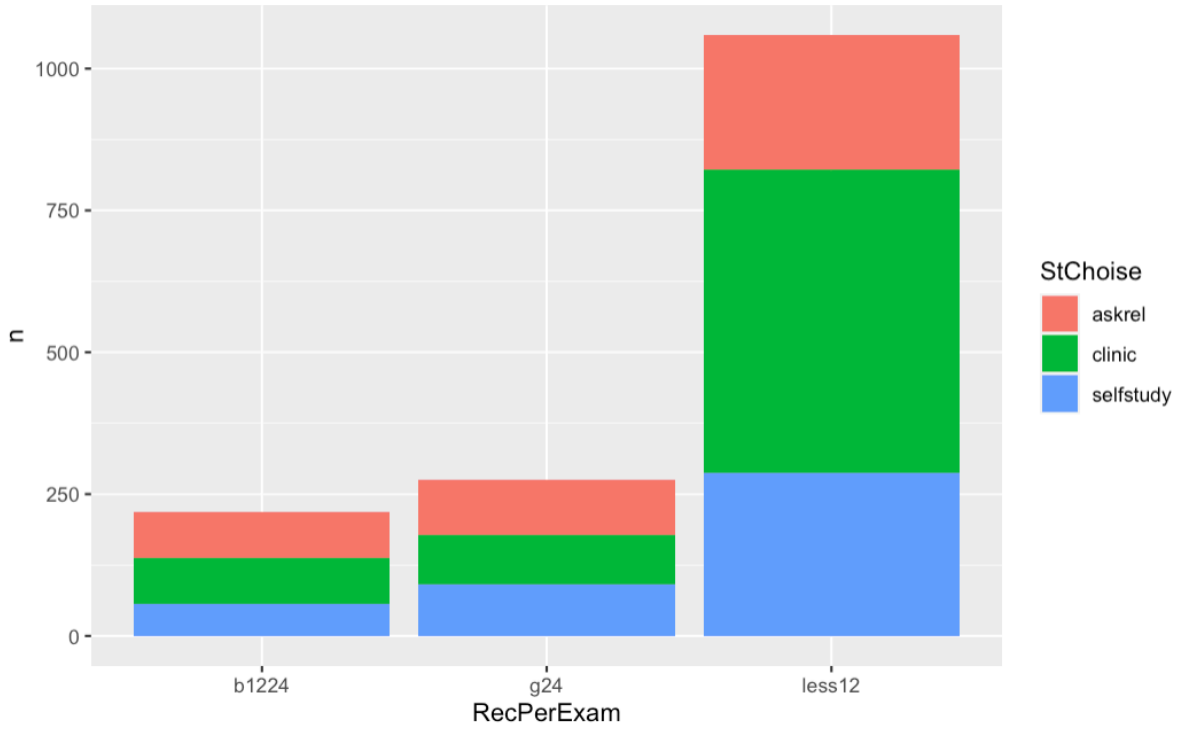

Figure 17: Summary of a patient's preferred way of dealing with new symptoms and annual checkup frequency

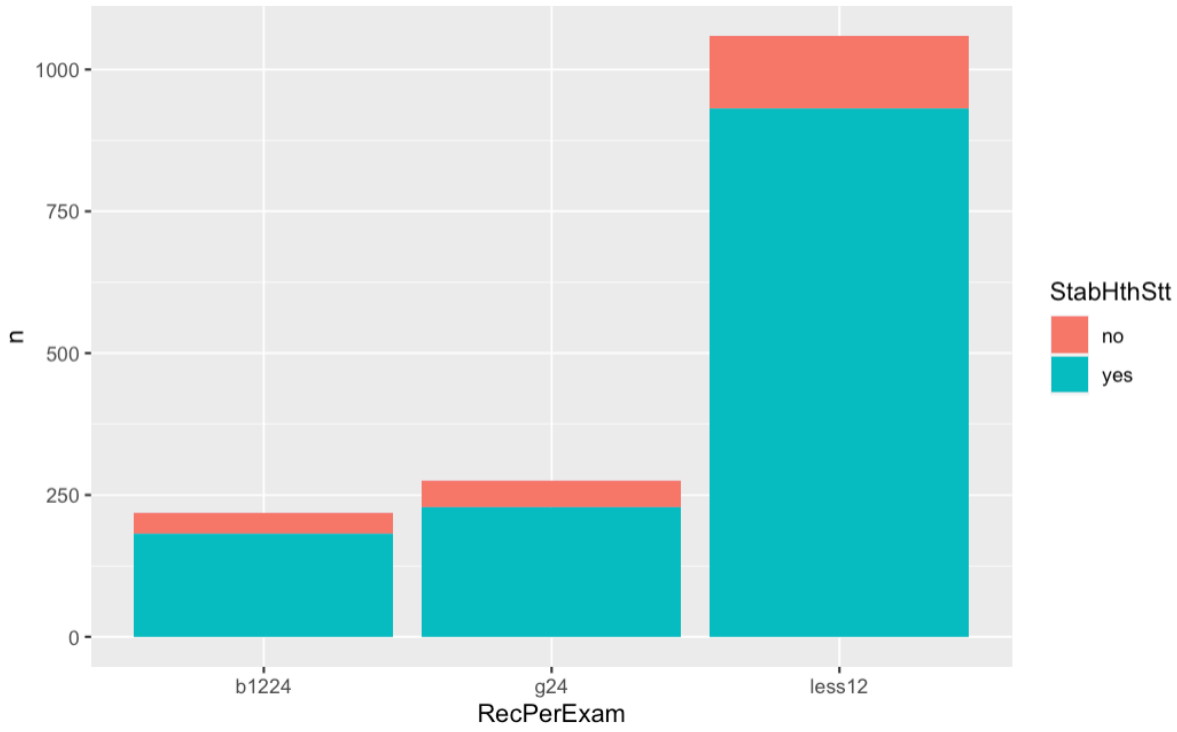

Figure 18: Summary of a patient's preferred way of dealing with new symptoms and annual checkup frequency