***Demographic Data Analysis***

A graph with numbers and a number of missing in periodontal disease risk

AI-generated content may be incorrect.We first look at the difference between total research\_id counts between the 2022 set and the 2023 extension, including comparing the number of missing periodontitis disease risk classifications:

A graph of red and blue bars

AI-generated content may be incorrect.We look at the differences for each category between years:

A graph with numbers and a number of exam results

AI-generated content may be incorrect.We want to see the transition of risk classes between the two years:

This is how the final demograhpics dataset is determined

* We have a list of demographic\_data [x, y, …, z].
  + This represents the dumping of files into the demographic folder, and all of x, y, …, z is a demographic dataset.
* We determine which of the data files is the biggest, as this will give us the most “information” about research\_ids that we can use. We will use this as the “base” dataset.
  + From the other datasets, we will see if there are new research\_ids there that aren’t in base dataset, and add them.
* The periodontal disease risk is determined as follows:
  + The classification from the latest dataset is always used
  + If it is missing, then it will use the next latest classification
  + If that is also missing, then it will use the third latest classification, etc.
  + From this method, we have:
    - 9072 total research\_ids
    - 546 pdr values missing in base df but available in latest df
    - 1245 pdr values that were changed to reflect newest data
    - 1791 total pdr values changed to reflect latest data