|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| patient | Right maxillary score | Right anterior ethmoid score | Right posterior ethmoid score | Right sphenoid score | Right frontal score | Left maxillary score | Left anterior ethmoid score | Left posterior ethmoid score | Left sphenoid score | Left frontal score | Survey score | (Total CT score) |
| ID-1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 7 |
| ID-2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1.7 | 1 |
| ID-3 | 2 | 1 | 0 | 0 | 1 | 2 | 0 | 1 | 0 | 0 | 2.7 | 7 |

For patient 1

|  |  |  |  |
| --- | --- | --- | --- |
| Sinus | CT score (0/1/2) | Frequency | Survey score |
| Maxillary | 0 | 0 |  |
| Maxillary | 1 | 2 |  |
| Maxillary | 2 | 0 |  |
| Ethmoid | 0 | 0 |  |
| Ethmoid | 1 | 4 |  |
| Ethmoid | 2 | 0 |  |
| Sphenoid | 0 |  |  |
| Sphenoid | 1 |  |  |
| Sphenoid | 2 |  |  |
| Frontal | 0 |  |  |
| Frontal | 1 |  |  |
| Frontal | 2 |  |  |

Graph: <https://ggplot2.tidyverse.org/reference/theme.html>

DATA DESCRIPTION:

* The colored columns are CT score for sinuses – it is an objective evaluation of patients’ sinus condition, and it take on values 0, 1, or 2.
* The uncolored “survey score” column is a subjective evaluation of patients’ decrease of quality of life. The greater the number, the less happy they are.
* The uncolored “total CT score” column is the sum of the CT scores for each individual

GOAL: We want to see if there is a relationship between CT score (e.g. sinus disease severity) and survey score (e.g. quality of life)

Previous not-so-successful attempt:

* We previously looked at the sum of CT scores across all sinus regions but didn’t see a clear pattern. We figured that it’s because score 1 represents a WIDE spectrum of condition. Score 0 is completely healthy, while score 2 is severely ill. Score 1 is anything that’s not score 0 or score 2.
* People who have some mild sinus conditions across most sinuses, like patient ID-1, can have a semi-high total CT sinus score from adding a bunch of 1’s, but their quality of life might not be greatly affected because their sinus conditions are pretty mild.
* People who have a region with really severe sinus condition, like patient ID-3 with some score 2, can end up with a similar total CT sinus score like the previous case (both have total CT score = 7), but their quality of life can be very different (0 vs 2.7).

NEW IMPLEMENTATION PLANS:

Way 1: add **weight** to individual CT score before summing them

Way 2: count the frequency of each score value and correlate - we expect people with more score 2’s to have worse quality of life.

Examples on counting the frequency of each score value -

For patient ID-1,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Maxillary | Ethmoid | Sphenoid | Frontal |
| Score 0 | 0 | 0 | 2 | 1 |
| Score 1 | 2 | 4 | 0 | 1 |
| Score 2 | 0 | 0 | 0 | 0 |

For patient ID-2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Maxillary | Ethmoid | Sphenoid | Frontal |
| Score 0 | 1 | 4 | 2 | 2 |
| Score 1 | 1 | 0 | 0 | 0 |
| Score 2 | 0 | 0 | 0 | 0 |

For patient ID-3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Maxillary | Ethmoid | Sphenoid | Frontal |
| Score 0 | 0 | 2 | 2 | 1 |
| Score 1 | 0 | 2 | 0 | 1 |
| Score 2 | 2 | 0 | 0 | 0 |

My problems:

1. How to turn the original dataset into one with score frequency? How should the table be arranged?
2. How should I visualize the score frequency vs quality of life score? There are so many variables to keep track (e.g. sinus region, score level 0,1,2, and frequency) for each individual. What’s the most elegant way to present them on a graph?

* I have thought about using
  + 4 different colors to represent the 4 different sinus regions
  + 3 different shapes for score levels (0,1,2)
  + different color transparency for the score frequencies
* But then each individual would end up having 12 data points (4 colors, 3 shape) within a graph. How could the graph be clean and clear when this is the case?