**Data Vis Complaints project clean up documentation:**

Removed words and words following:

Forms of

Other part of

With

Without

Of degree of

Lateral end of

Part of

Following

Due to

Removed words:

Unspecified

Specified

Other

Lower/ upper (except referring to quadrant)

First, second, third, fourth, fifth, sixth

Left/right

Not otherwise listed

Not otherwise specified

Removed y and s at the end of words like dependence (y) and disorder (s)

Lengthened some common abbreviations such as:

Inj, obst, poisn, prt, disord, musc, infrc, unil, disp, fx, pulmon

Uppered the first letter of each word, lowered the rest.

Removed punctuation

Trimmed white spaces

Removed double spaces

This reduced the diagnosis categories by ~1000 for the full data set.