We've put together a brief data dictionary to explain the contents of the two data files.

[**cost.csv**](https://drive.google.com/file/d/1XMBy5lK_g8b8vmOh0dj9mWTAoXE3yktk/view?usp=share_link)

This file contains the vast majority of the data that underlies the project. Here's an overview of the columns included:

* *geo\_level* - The geography for which the row contains prices. Can either be national (i.e. "USA"), or an individual state (i.e. "AK" for Alaska)
* *id* - The unique ID for the procedure. You can use this ID to look up the name of a procedure in the labels.csv file.
* *insurance\_type* - The type of insurance a patient had. The options are "Medicare" or "Insurance" (aka private insurance).
* *source* - The source that the pricing data came from. The options are "CMS" (aka Centers for Medicare & Medicaid Services), "AHRQ" (aka Agency for Healthcare Research and Quality), and "BLS" (aka Bureau of Labor Statistics).
* *year* - The year the prices are for.
* *percent25, percent50, percent75* - The 25th, 50th, and 75th percentile price for the procedure. These are used for the big number and range plot in the designs.

[**labels.csv**](https://drive.google.com/file/d/1aDJ7iAQY_hcn9VOMzR0YZQC2gXxH5g3J/view?usp=share_link)

This file contains the names of the procedures. Here's an overview of the columns included:

* *id* - The unique ID for the procedure
* *label* - The name of the procedure.
* *label\_detailed* - A description of the procedure.
* *slug* - A slugged name of the procedure, which is unique.