

Q1 Responses for Quiz1

- A) Open Data is data that anyone can access, use and share to their liking. An example of open data that I could produce and others could use would be if I were to share daily weather observations in my city online. Researchers (or anyone) could use this data for climate studies or forecasting models.
- B) 1. Values could be missing because of data entry mistakes, like missed survey questions or sensor malfunction. Values could also be missing due to data corruption.
2. One method of dealing with missing values would be to outright delete them from the dataset. The assumption here is that the data is missing completely at random, meaning the values don't depend on any variable, and the risk is that if the assumption is wrong then the deletion of these values can result in bias by reducing the sample size. Another method of dealing with the values is imputation (replacing the value with the mean or median for example), where the assumption is that the missing values are missing at random, and the risk being that you may distort the relationships between variables.