Displaying Linkage Success Statistics to Identify Systemic Errors

Simpson, Mike[1], Yip, Harold[1], and Hills, Brent[1*]

[1]Population Data BC

Objective

The primary objective is to create a method for displaying linkage statistics to researchers, data stewards, and linkage specialists in an informative and meaningful way. The method must visually display the linkage summary data and highlight drops in the linkage success rate.

Approach

We created a web interface which shows linkage statistics by age and geography in calendar/service years. Each cell contains both the percentage of linked values along with the percentage of successfully linked data. The interface is filterable by gender, data-type, and whether to display the number of successful or unsuccessful linkages. Due to the high volume of data which will appear on the screen at one time, we use a heat map to highlight cells which have unusually high or low values. Totals are displayed with their own heat maps to compare easily years across ages group or age groups across years. We mask small cell sizes to preserve privacy.

Results

This approach allows people to easily spot drops in linkage success. If a particular year's data or age group has a lower linkage rate than the rest of the dataset, the heat map can clearly highlight that discrepancy. Displaying the number of linkages along with the rate helps us determine if the sample size is playing a role in a low linkage success rate.

Conclusion

Data quality issues can silently cause linkage success rates to drop in certain years, geographies, age groups, or genders. Displaying linkage statistics on a single page with a heat map allows people to quickly spot inconsistencies in linkages.

*Corresponding Author:
Email Address: brent.hills@popdata.bc.ca (B. Hills)