Data Quality 101

- Have you checked your data for quality issues?
- Has anyone questioned the quality of your data?
- Have you worried about including your data in a presentation?
- Have you recently changed the way your data are collected
  (e.g., new system/contractor/data elements)?

For questions, contact ACL's Office of Performance and Evaluation.

	Inaccuracy	Inconsistency	Missing Data	Bias	Outliers
1 Identify Potential Issues **2** Ask Yourself the Following	Are the data values the correct values, and are the data represented in a consistent and unambiguous form?	Are there inconsistencies or contradictions in data across variables and/or time (e.g., duplicate data or discrepancies in data definitions between data sources)?	Are all available data elements or data points included in the data set? How many data are missing? Missing data can occur because of nonresponse or attrition.	Are the data systematically different from the population parameter being estimated? Do respondents have different characteristics from non-respondents?	Are there data points with characteristics that are considerably different from most of the other data points in the data set? Outliers could indicate "interesting" cases or errors in the data.
3 ACL Examples	Information about OAA Title III services are usually overre-ported or underreported. For example, a state would report 100,000 services for several years and then increase to 500,000 the next year, or vice versa.	For OAA Title VI data, a tribe was erroneously listed under anoth-er state. Though it was recently corrected on AGID, because it was listed erroneously in other places it has to be listed that way on AGID for those past years.	Missing data in data sets such as the SPR and CIL PPR leads to underreporting and gaps in data. (Missing data is not posted on AGID).	NSOAAP suffers from nonre-sponse bias, so the respondents who answered the survey may not be representative of the tar-get population. Among surveys targeting the aging population, the individual may no longer receive the service or may have moved into assisted living or a nursing home.	For variables with a small overall total, outliers can have a large impact. One state or grantee's number can impact the entire region or U.S. total and can skew a data set. ACL is not able to force corrections even if the error is an obvious typo. In one case the reported number spiked from 3 to 11,000 and back down to 3.
4 What to Do Next	- Use trend data to verify if data are inaccurate (over- or underreported). - Follow up with data provider. - Verify with stakeholder as appropriate.	- Use historical data to verify changes, possibly reviewing data from before changes were made (e.g., PPR updates, new grant cycles). - Verify with stakeholder as appropriate.	- Use filtering to identify missing data points; calcu-late and report percentage of data missing. - Identify columns that have no data reported. - Verify with stakeholder as appropriate.	- Intentionally explore alterna-tive explanations for the data. - Conduct subgroup analysis where possible. - Verify with stakeholder as appropriate.	- Identify outliers by analysis (e.g., at-a-glance, filtering, Min/Max functions in Excel). - Are there extenuating circumstances that cause outliers (e.g., change in program delivery, COVID-19 reporting, change in services)? - Verify with stakeholder as appropriate.