Data integrity

inte-gritty

5 questions

 -- 5 questions you should ask every time:
    -- Are there missing or truncated values?
    -- Are there out of range values?
    -- look at all the values that appear in a column/row - do they make sense?
    -- Are there similar repeated values?
    -- Is the data format what you'd expect for the data type?

Truncation

TRUNC: Truncates a number to a certain number of significant digits by omitting less significant digits.

ROUNDUP: Rounds a number to a certain number of decimal places, always rounding up to the next valid increment.

ROUNDDOWN: Rounds a number to a certain number of decimal places, always rounding down to the next valid increment.

MROUND: Rounds one number to the nearest integer multiple of another.

INT: Rounds a number down to the nearest integer that is less than or equal to it.

FLOOR: Rounds a number down to the nearest integer multiple of specified significance.

CEILING: Rounds a number up to the nearest integer multiple of specified significance.

Strategies

- Consciously define reasonable range for each field
- Consciously define appropriate data type for range
- What are the number of digits you expect?
- Sort
- Use filter dialogue box
- Look for close cousins
- Pivot
- Look for outliers (graphing can help)

Know when to walk away

Sometimes your data is simply not robust, detailed, accurate, or large enough to do what you want with it

http://bit.ly/2glYmJ4