Practice: Checking the Data Table for Issues

In this practice, you use the file **Messy Data.jmp** to identify data quality issues. This data table has seven variables: **acceptable?**, **lot number**, **unit number**, **location**, **humidity**, **voltage**, and **density**.

1. Look at the data table for evidence of the following issues:

   - incorrect formatting
   - incomplete data,
   - missing data, and
   - dirty or messy data.

2. For each variable, what do you learn?

acceptable?: This variable is missing three values.
lot number: This is a label, so it should be coded as nominal or none.
unit number: This is also a label, and the categories are unordered. This should be coded as nominal.
location: The labels for the countries are messy.
humidity: This is missing four values.
voltage: Most of these values are zero. If it is impossible for voltage to be zero, then these might actually be missing values. You would need to confirm this.
density: These are all the same value, 40.

Hide Solution

Close