



Data Readiness Checklist

- ☐ Assign a unique ID to each observation or participant.
- ☐ Keep variable names clear, short, and consistent.
- ☐ Ensure dates are in a standard format (YYYY-MM-DD).
- ☐ Document units of measurement (e.g., minutes, dollars, percent).
- ☐ Avoid merged cells, hidden columns, or multiple header rows.
- ☐ Check for missing values and decide how to handle them.
- ☐ Create a simple codebook that explains each variable.
- ☐ Store data in a clean format (CSV preferred over XLSX).

Long vs. Wide Format

Although I'm happy to handle data preparation, it's always best to keep your data organized in a format that's easy to analyze. In most cases, this means using long format — each row is a single observation, and each column describes a variable. Long format is the most flexible for analysis and visualization, especially when you have repeated measures or time points.

Wide format can still be useful for quick summaries or presentations, but if you keep the underlying dataset in long format, you'll save time and reduce errors when it comes to real analysis. Think of wide format as a snapshot for people, and long format as the structure your tools — and your analyst — prefer.

Wide Format (easier to read)

| Student | Score_Time1 | Score_Time2 |
|---------|-------------|-------------|
| Alice | 85 | 90 |
| Ben | 78 | 82 |
| Carla | 92 | 95 |

Long Format (better for analysis)

| Student | Time | Score |
|---------|------|-------|
| Alice | 1 | 85 |
| Alice | 2 | 90 |
| Ben | 1 | 78 |
| Ben | 2 | 82 |
| Carla | 1 | 92 |
| Carla | 2 | 95 |