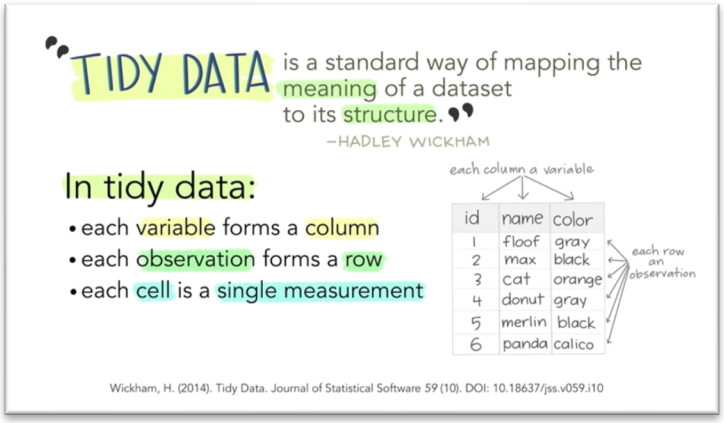
Supplemental Materials

To accompany: Majewska. AA. 2026. Statistics and Data Visualization in CUREs. Extended abstract #29 in: Cole M and O'Brien J, eds. *Advances in biology laboratory education*. Volume 46. Publication of the 46th Conference of the Association for Biology Laboratory Education (ABLE). DOI: https://doi.org/10.37590/able.v46.extabs29

Correspondence to: Ania A. Majewska, majewska@uga.edu

Data formatting

To analyze data using statistical software R and DataClassroom.com, our data will have to be in a **tidy** format. Please use Excel to input and store your data. You can use Google Sheets or Microsoft Excel.



An example of tidy data format:

SiteID	Date	Time of day	Dissolved oxygen	pH	Water color	Notes
A1	2025-05-29	morning	0.33	0.55	yellow	Rain day
B1	2025-05-29	afternoon	0.02	NA	brown	Missing pH, meterX not working

Tips for data entry in Excel:

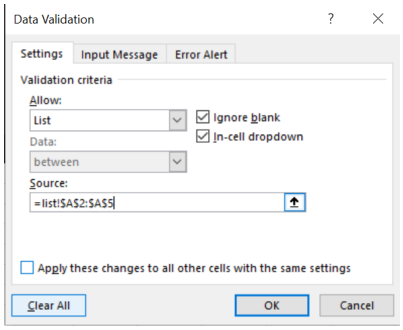
Name columns consistently and without spaces or special characters.

Use consistent formats, codes, spellings, etc. and do not mix data types in a single column. Don't use Colors, Fonts, *Italics*, or visuals as data

Use the same formats for dates: use the date format YYYY-MM-DD or use separate columns for Year, Month, and Day.

Use 'Data Validation' in Excel to help you keep formatting consistent:

- Visit the menu ribbon, select DATA, and select the DATA VALIDATION option in the DATA TOOLS box.
- In the dialog box, select LIST from the ALLOW dropdown menu.
- Enter your cursor in SOURCE box and visit your list worksheet. Select the terms you want to use as standardized values for your column.
- Once you select OK, your column cells should include dropdown menus featuring the values in your list.



Fill in all Cells: use NA ("Not Applicable" or "Not Available") to make clear the data is missing rather than unintentionally left blank. Make a separate column for notes if you need to explain why data is missing.

Use a descriptive file name. For instance, a file named *SEV_SmallMammalData_20200525.csv* indicates the project the data is associated with a project (SEV), the theme of the data (SmallMammalData) and also when this version of the data was created (20200525). This name is much more helpful than a file named *mydata.xls.*

How to make an existing datasheet into tidy format?

1. Remove empty, non-data rows
2. Complete incomplete rows and headers (for example, by repeating labels that appear only in the first row of a group)
3. Remove sub-totals and totals
4. Use 'transpose' function in Excel to copy and 'transform' the data from wide to long format. What does that mean?
   Say you have a dataset where you have a measurement that is taken every year for many years, like CO2 Emissions. If you have one column for every year of the measurement, the table is in wide format (see below left). If, instead, you have a column that says "Year" and a column that says "CO2 Emissions," the table is in long format (see below right).

Wide Format:

State	Sector	1990	1991	1992	1993
Alabama	Commercial	2.43	2.00	2.10	2.05
Alabama	Electric Power	50.28	54.18	56.93	62.82
Alabama	Industrial	25.15	25.36	28.36	26.18
Alabama	Residential	3.09	3.02	3.20	3.42
Alabama	Transportation	28.13	28.70	29.39	29.51
Alaska	Commercial	2.20	2.23	2.51	2.56
Alaska	Electric Power	2.61	2.46	2.29	2.32
Alaska	Industrial	15.83	17.44	18.87	18.51
Alaska	Residential	1.58	1.60	1.74	1.73
Alaska	Transportation	12.09	11.17	10.86	11.03
Arizona	Commercial	1.90	1.83	1.78	1.74

Long Format:

State	Sector	Year	Emissions
Alabama	Commercial	1990	2.42906
Alabama	Commercial	1991	1.999039
Alabama	Commercial	1992	2.10271
Alabama	Commercial	1993	2.05046
Alabama	Commercial	1994	2.054954
Alabama	Commercial	1995	1.962636
Alabama	Commercial	1996	2.16117
Alabama	Commercial	1997	2.417254
Alabama	Commercial	1998	1.903054
Alabama	Commercial	1999	2.187003

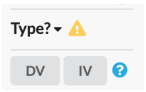
Analysis and Visualization in DataClassroom.com

1. Check that your data follows the tidy format

2. Import your data

   i. Click green button "Import data"

   ii. Select "Upload a spreadsheet from your computer" or "Fetch data from a spreadsheet on Google Drive"

   iii. Specify data types for each column by clicking down arrow next to Type? 

      a. Numeric: measurements (temperature, pH, decibel readings)

      b. Categorical: site names, habitat types

      c. Date/time: collection dates and times

iv. Specify if the data is a dependent (DV) or independent (IV) variable by clicking one of the two options under the variable: 

3. Visualization

   i. Consider graph type you need:

      a. Scatter plots: relationships between two numeric variables

      b. Box plots: distribution of data across categories

      c. Bar charts: comparing categorical data

      d. Line graphs: changes over time

   ii. Select "Graph" from the dashboard and “Switch to Graph tab” or use the “Graph Wizard” (it will walk you through graph type selection)

iii. Choose variables for your x and y axes. If you see your data table, you can choose by clicking on “Show” under variable.

iv. You can change x and y variables under “Selected columns” and by selecting X, Y, or Z:

v. To change graph type, click on 

vi. To customize your plot,

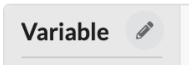
a. click on “Appearance” to:

i. Change point size, type, and colors

ii. Add a title

b. Change your axis labels:

i. Click on pencil symbol on the table and change the name of the column.

Be sure to include units you want to display 

4. Statistical analysis

i. Choose "Descriptive stats" in the graph tab, it will add standard deviation if you have multiple data points, and click “values”, it will add a table with descriptive statistics.

ii. Select  "Graph-driven test" next to the “Appearance” button for guided statistical analysis

- An appropriate test will be automatically selected for you. Click “Calculate”

Handout contains text adapted from the following resources:
Data validation: https://libguides.tulane.edu/excel_clean_analyze
Excel best practices: https://libraryguides.unh.edu/excel/bestpractices
How to make data tidy: https://www.lib.uiowa.edu/data/manage/data-structure/
Filling in blanks: https://www.extendoffice.com/documents/excel/771-excel-fill-blank-cells-with-value-above.html