Data Cheat Sheet: The Basics

|  |  |
| --- | --- |
| **Pros: Google Sheets** | **Pros: Excel** |
| Cloud storage (access from any device with Wifi) | Privacy and security |
| Stores up to 2 million cells | Can analyze larger amounts of data |
| Collaboration and sharing |  |
| Auto-save and Version history |  |
| Free |  |
| Looks the same as other programs (Docs) and on different computers |  |

**Levels of Data:** Workbook (the file) -> sheet (what you’re looking at on your screen) -> rows + columns -> cells.

**Cell Formatting:** Controls how a cell or column will be displayed. For example, how a date will be displayed (02/26/2018 or 02/26/18 or Feb. 26, etc.) or how many decimal places to show on a numeric value.

* Click on format in the toolbar, then number (the first selection in the dropdown) and choose a format in the next pop-out.

**Freeze Panes:** Allows you to lock a row or column (usually the first row) so that it remains in the screen as you scroll.

* Useful when navigating large data sets.
* View -> freeze -> 1 row, to freeze the header row.

**Searching:** Find a specific value in a spreadsheet.

* Use command + f, just like you do in other applications, OR do edit -> find and replace (from the toolbar).
* Will highlight those values and show you in the search bar how many exist in the spreadsheet. Use arrows to navigate among highlighted values.

**Sorting:** See the largest or smallest values in your dataset, or put rows in order by time or ABC.

* Click in upper left corner of spreadsheet to select all rows and columns.
  + Skipping this step could sort one column only and mix up your data.
* Data -> sort range …
* Check box for “Data has header row.”
* Choose a column to sort by.
* A-Z (will sort smallest to largest for numbers) or Z-A (largest at the top).
* Choose “then by” to sort by an additional column (useful when many rows have the same value).

**Filtering:** Filter data into a smaller sheet based on a column’s values.

* Filter is either on or off. Get to it with the funnel in the toolbar OR data -> filter.
* When filter is on, you’ll see lines next to the header rows. Click those to filter by that column’s values.
* When there’s an active column filter, there will be a filter icon next to the column’s title.
  + **Conditional Filtering:** Follow the same steps, but select conditional filtering once you click into the column you want to filter. This allows you to filter by ranges such as only rows with values under 10.
* If you need to come back often to a filter, save it. Data -> filter views -> save as filter view.

**Formulas:** An expression that calculates the value of a cell. For example, typing =2+2 in a cell would fill the cell with 4.

* There are hundreds of formulas. They all start with = and usually include a cell or range in between ().
* Example: =AVERAGE(C1:C22) would calculate the average of values in column c, rows 1-22.
* Some popular ones:
  + =MEDIAN() use in place of average when your dataset contains outliers.
  + =MIN() the min.
  + =MAX() the max in a dataset.
* Percent change: ((value2-value1)/value1)\*100
  + In GSheets, =((A2-A1)/A1)
  + Then, change the cell format to percent.
* Using a $ will anchor that value to make sure it doesn’t change as you apply formulas to later cells.
  + =$A1 locks A (the column) as you apply the formula to the right and left.
  + =A$1 locks 1 (the row) as you apply the formula up and down.
  + =$A$1 locks both, so you’re always referencing cell A1, even as you move up/down/right/left the sheet.

**Mean/Average and Median:** The mean adds all the data and then divides by the number of values. If all values were in a line from smallest to largest, the median would be the number directly in the middle. If outliers are present, in general, use the median. Otherwise, use mean.