Data Cheat Sheet: Separating Columns

*Separating columns is a common data cleaning task. Whether you’re trying to tease information out of a longer cell, separate a city and state or a first name from a last name, or strip out unnecessary or repetitive information in a cell.*

**LEFT/RIGHT:** These formulas work the same way. The difference, is that LEFT starts from the left-side of the cell and RIGHT starts at the right-side.

In its simplest form, we have add a list of addresses in column one and we want to get the zip code. The first argument is the cell you want to take the information from. The second argument is the number of characters you want to bring over. In this case, we know zip codes are always five characters.

**NYT ADDRESS ZIP CODE**

|  |  |
| --- | --- |
| 620 Eighth Avenue, New York, NY 10018 | =RIGHT (A1, 5) |
| 1627 I St NW, Washington, DC 20006 |  |
| 201 Spear St  San Francisco, CA 94105 |  |

**=SEARCH:** Often, the character length of what you’re searching for won’t be known (like it is in the zip code example). This formula looks for specific text in the cell and returns the *location* of that text. When combined with LEFT/RIGHT, search allows you to stop or start your substring based on the location of delimiter (like a comma).

It has three arguments:

1. What to search for. This will most likely be a string in quotes.
2. Where to search for it.
3. Optional: The character number you want to start the search on.

In the last name, first name example, the comma separating the names is our delimiter. To search for the location of the comma, the formula would be =SEARCH(“,”, A2). Below shows SEARCH combined with LEFT.

**COMPLETE NAME LAST**

|  |  |
| --- | --- |
| Cook, Lindsey | =LEFT(A2,SEARCH(",",A2)-1) |
| Chen, Elaine |  |
| Kim, Jin |  |

Without subtracting 1, your cell will include the comma.

**Other Useful Data Cleaning Formulas:**

* **MID:** For taking out something in the middle of a cell. =MID(column to pull from, character number to start on, length to extract).
* **LEN:** Tells you the length of any given cell. Often used in combination with search to find the number of characters to extract.
* **ISNUMBER:** Useful when paired with the search formula, because SEARCH will return #N/A if the characters you’re looking for aren’t present. Can also be used to extract numbers or identifiers from cells with lots of information.

**Hints:**

* Go one step at a time. If you strip one part of the cell in B, then another part in C, then another part in column D, you can have what you want in E. This is much easier than trying to do it all in one formula.
* Write out the cell on paper and number the characters. This will make it easier to debug your formula.
* Apply the formula down before moving to the next one. Just because it works on one cell doesn’t mean it will work on them all.