Power query PDQ! Demo notes
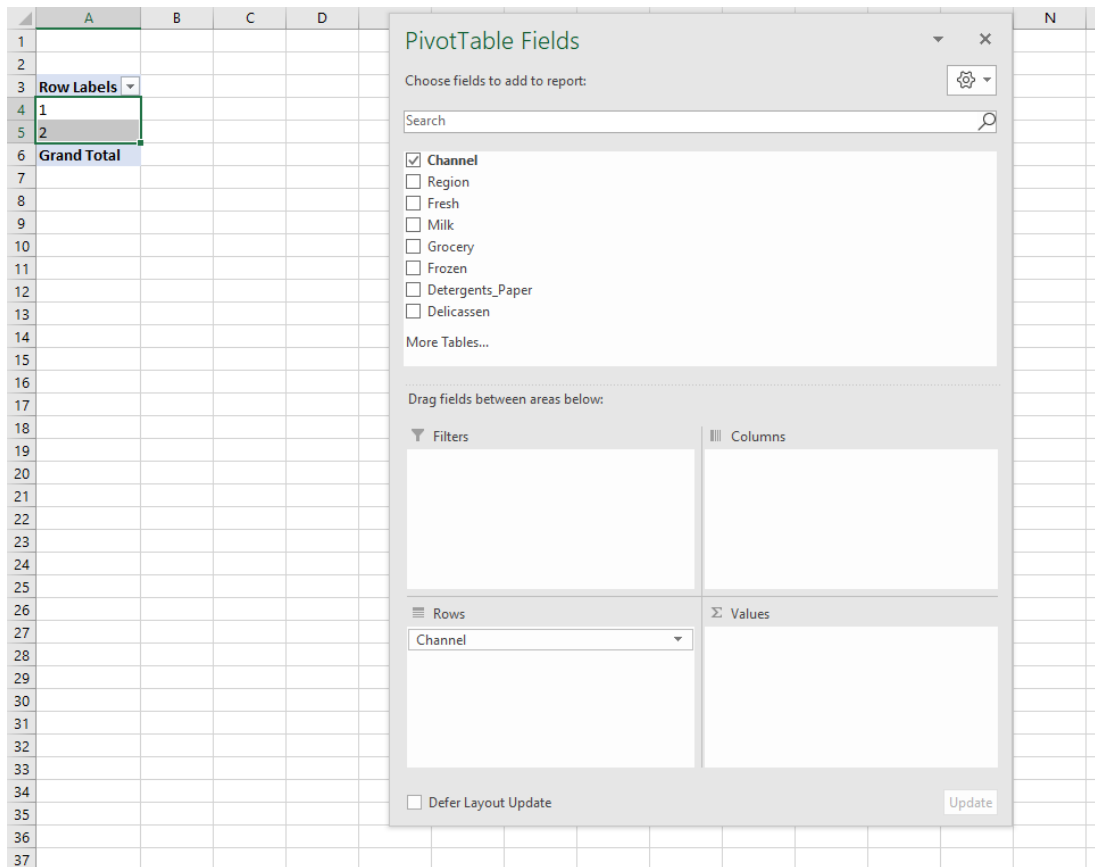
- I want to "pivot" on the file wholesale_customers.xlsx. What is the problem? .
  - There is no "sales" field to "pivot" on. I have several fields representing one attribute, category.
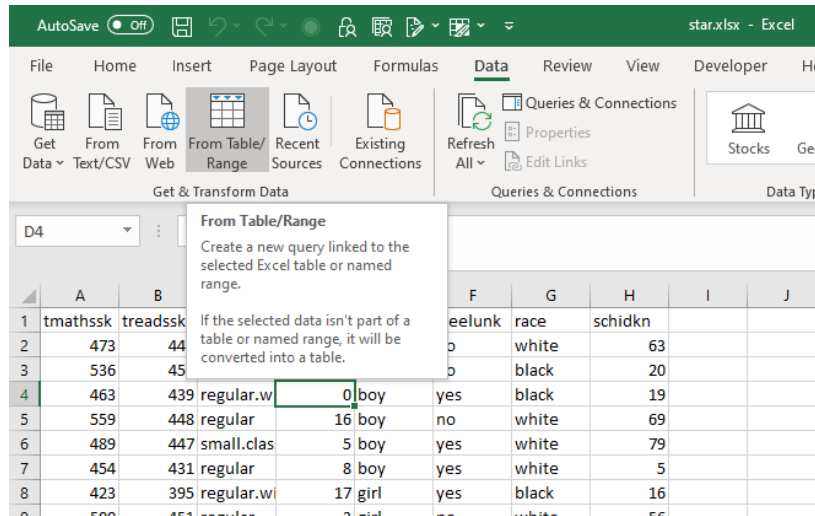

- *How would we have fixed this in Excel without Power Query*
- *What does this example tell us about how data should be shaped?*

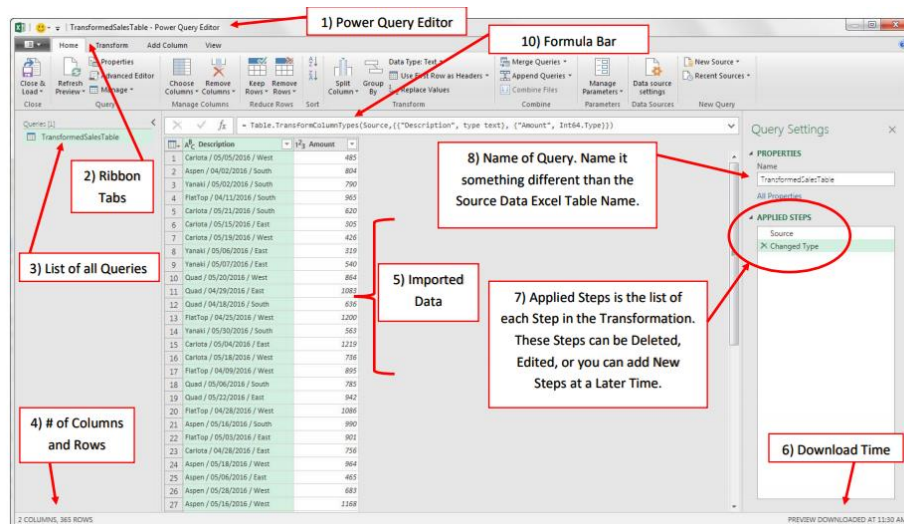**Import a Table into Power Query**

Demo: star.xlsx

1. Leave cursor anywhere inside the range you want to select
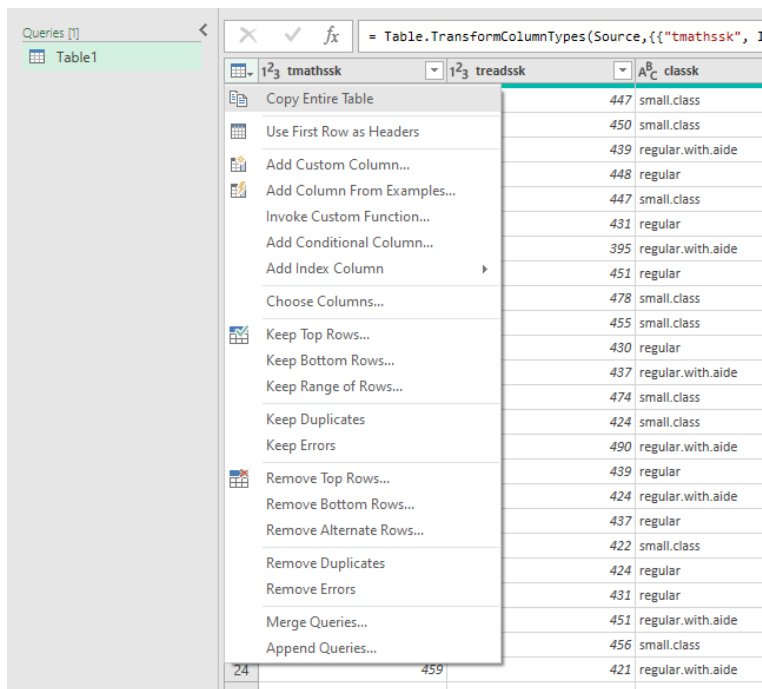
2. On the ribbon, select Data -> From Table/Range

3. This will convert the range to a Table.

4. You will now see the Power Query Editor (source: https://people.highline.edu/mgirvin/AllClasses/348/MSPTDA/Content/PowerQuery/003-MSPTDA-IntroToPowerQuery.pdf)
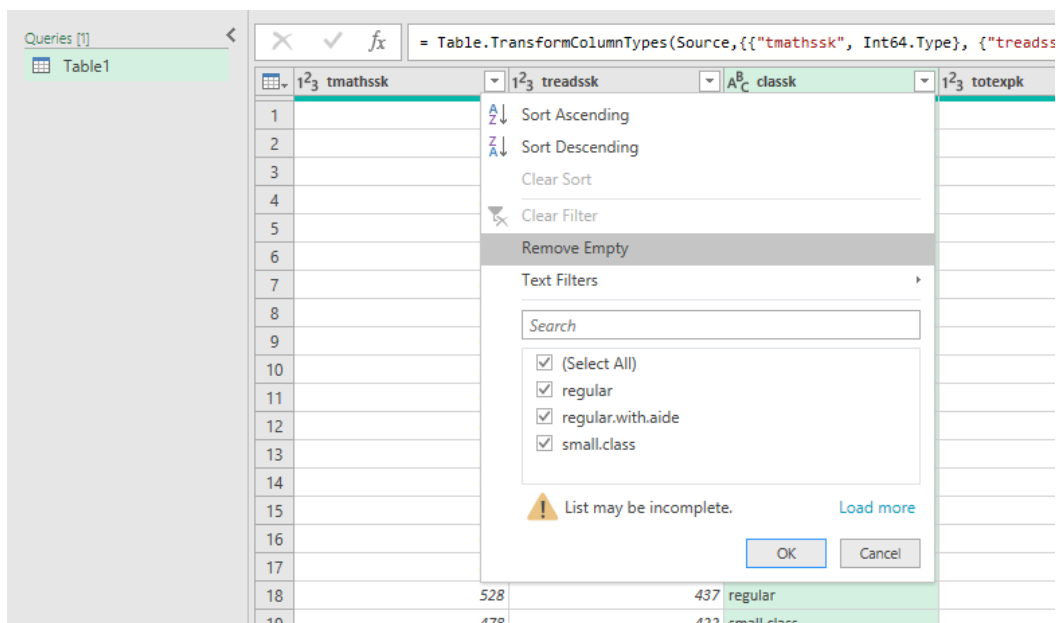

a. A Home ribbon is at the top, just like in Excel. The first three tabs are going to have data cleaning functionality.
b. The imported data is in the middle of the screen. We can click on rows and cells and see their values at the bottom of the screen.
c. There is a small table icon in the "corner" of the dataset. Click on that and there are some shortcuts to working with this data.
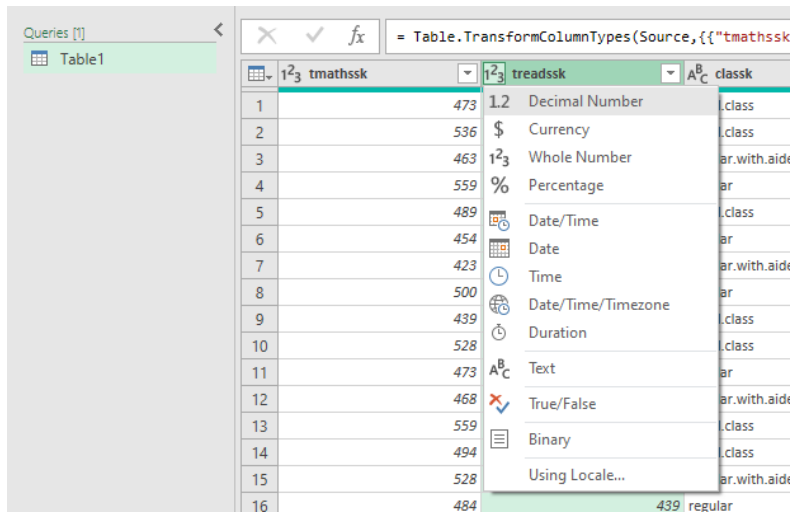
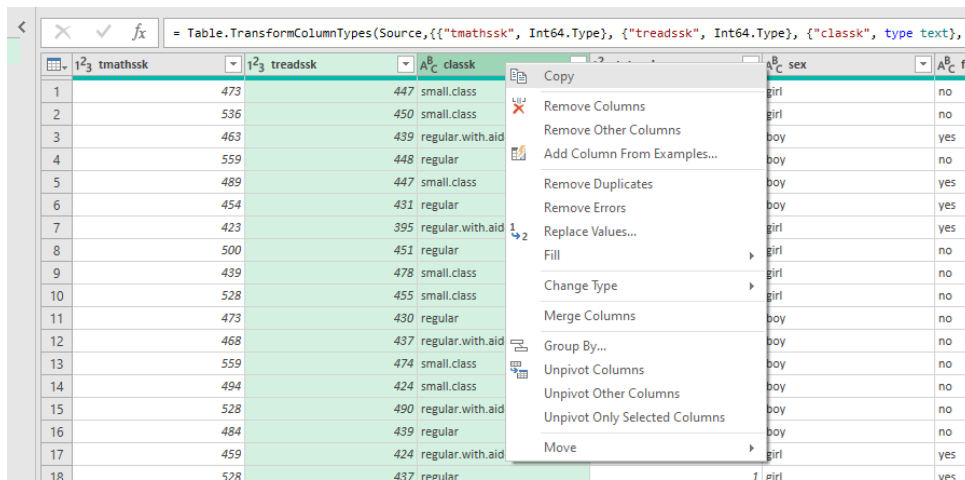d. Click on any column drop-down and you'll see you can filter it just like in native Excel.

e. You'll also see a symbol to the left of the column. This indicates the column's type. You can click on that to change the data type.
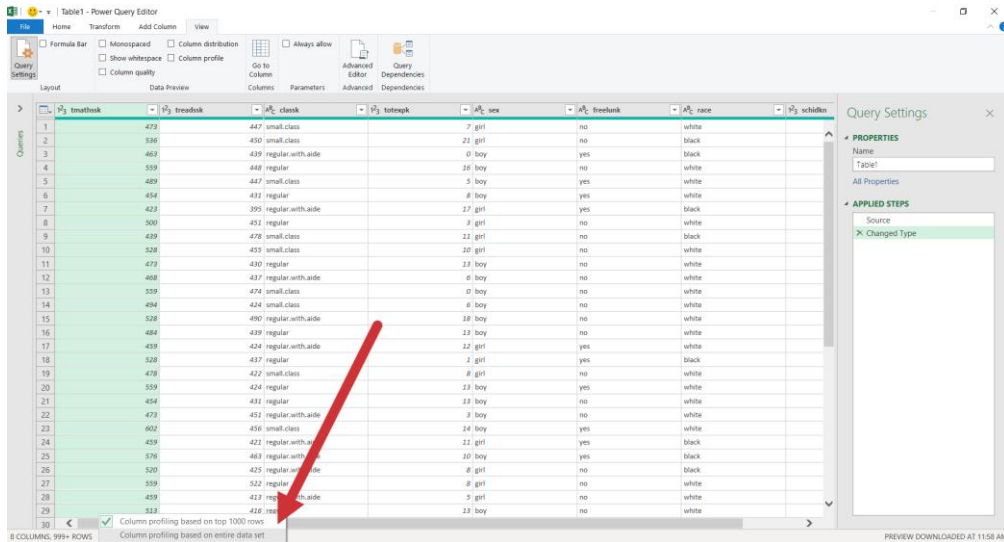
f. You can also right-click on a column to operate on it. Hold down Ctrl and click
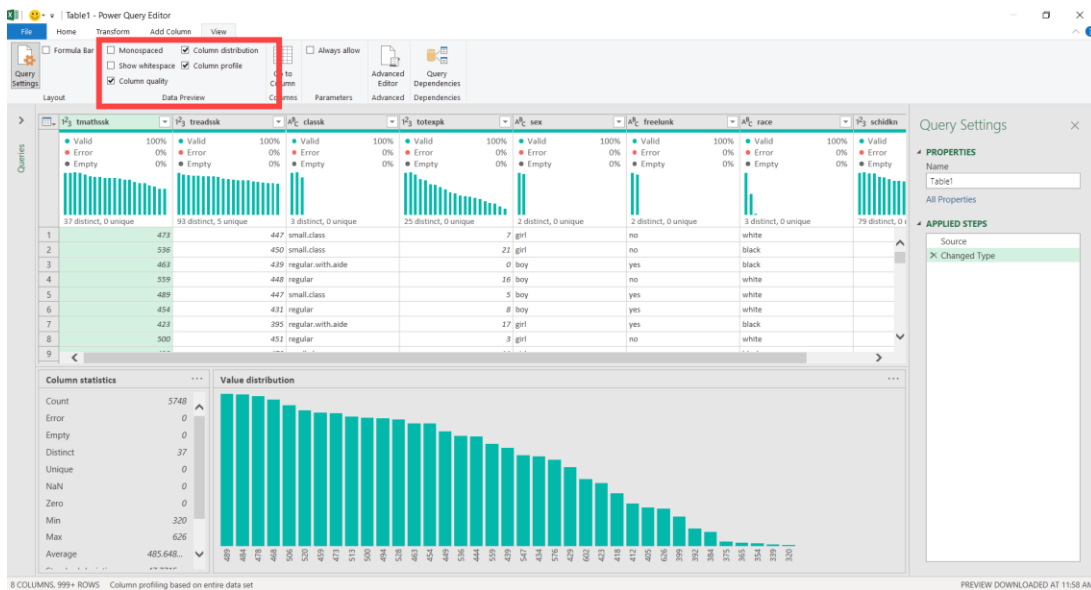   multiple columns to operate on multiple columns.

g. Now, go to the View tab on the home ribbon.
    i. Initially, what you are seeing in the Power Query editor is based on the
        first 1,000 rows.
    ii. To include all data in the Data Preview, click the message at the bottom
        which says `Column profiling based on top 1000 rows`. Change
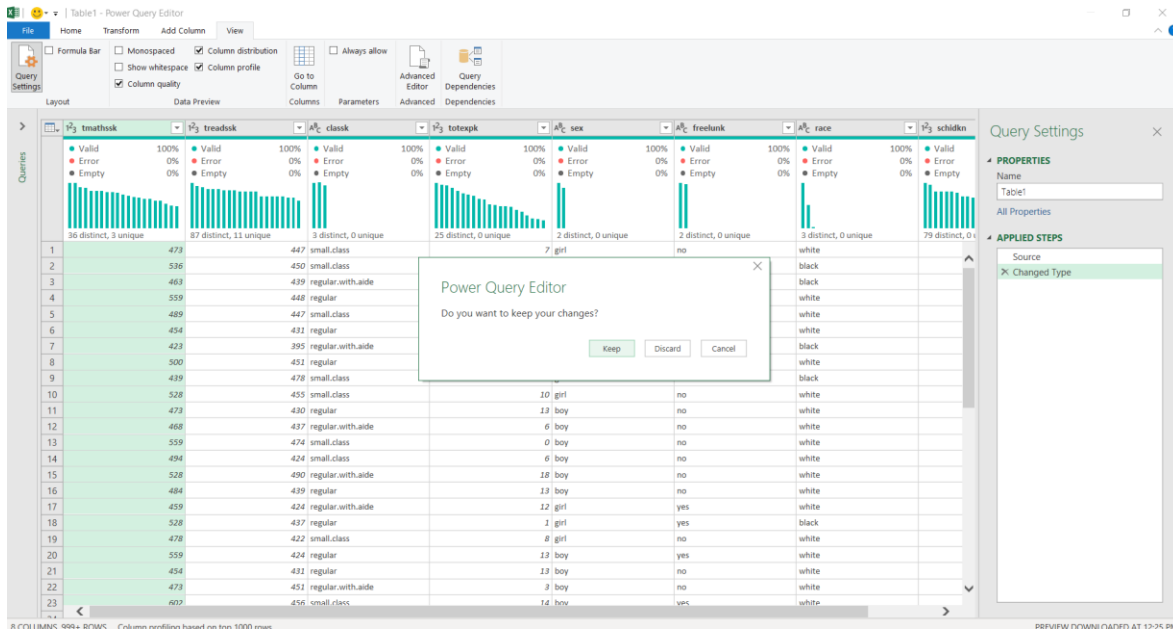   to `Column profiling based on entire data set`.

h. You can now change column appearance and add some statistics about each column using the Data Preview group of the View tab.

i. To exit the Power Query editor, hit the X on the upper-right. You can discard your changes for now.
   i. This will return you to "classic" Excel.
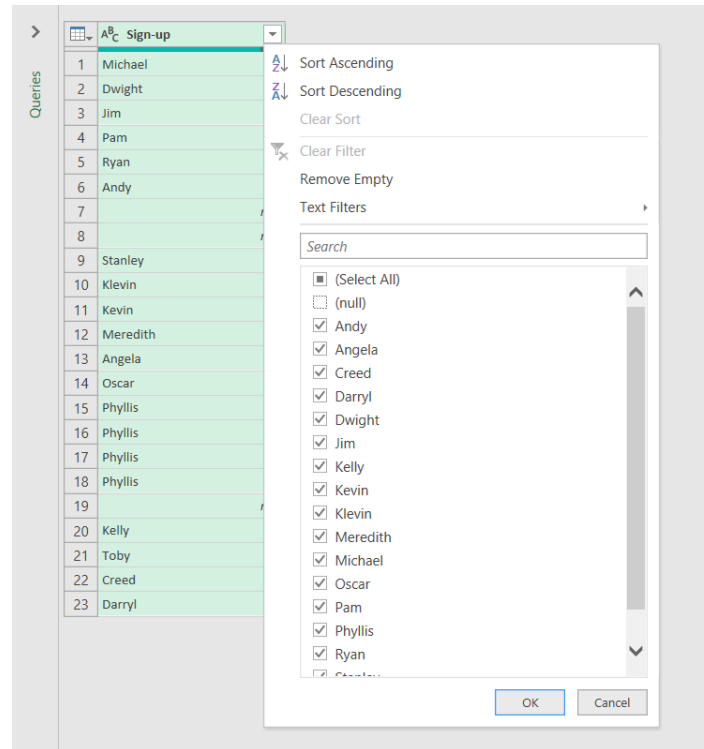
Demo: `office-rsvps.xlsx`

Worksheet: `signups`

1. You are consulting with the Party Planning Committee to clean up a list of RSVP's to a party. We would like to have the list sorted alphabetically, with duplicates, blanks and misprints removed.
2. This could be accomplished easily enough in classic Excel, but we would like to track each step of the data cleaning process, and we would like a solution that continues to work as more people RSVP to the list. These requirements make Power Query an excellent choice.
3. Create the connection from the range. Your data will be converted into a table.
4. You will see that blank values have been populated as `null` in Power Query. This is a special value indicating a missing value. It's not the same thing as zero!
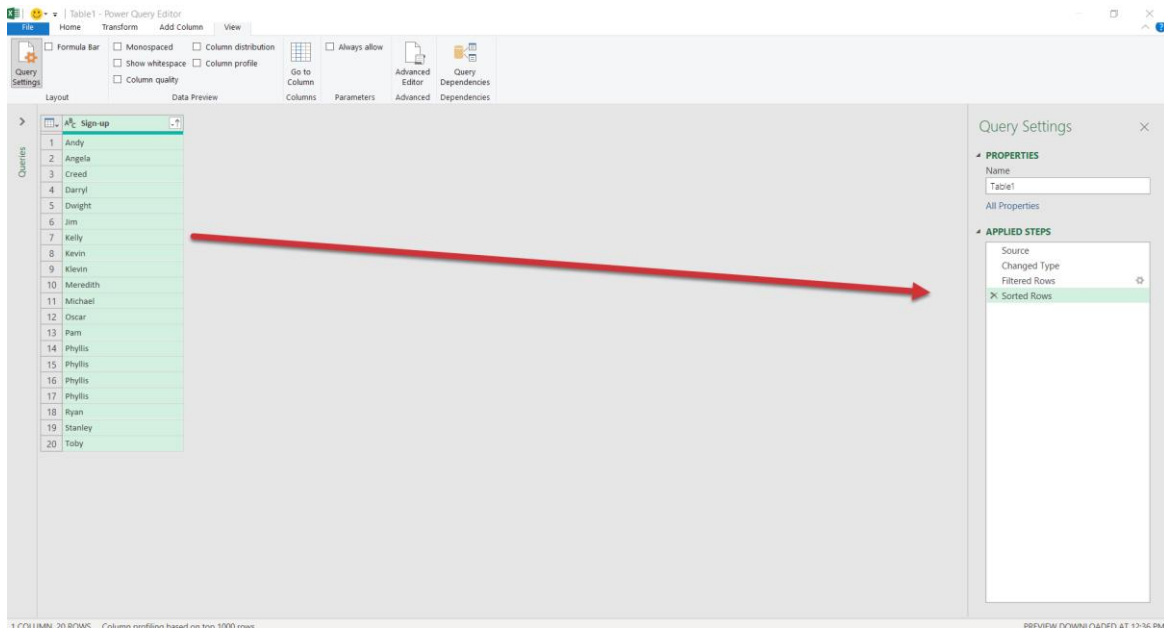

5. We want to filter out missing records, so select the drop-down on the column label and de-select `null`. This will remove them.
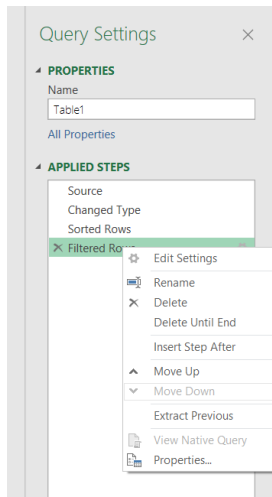
6. We can also sort the list A-Z with the same menu.
7. You will begin to see a running list of the steps we have taken on the right-hand side of the editor (Applied Steps).
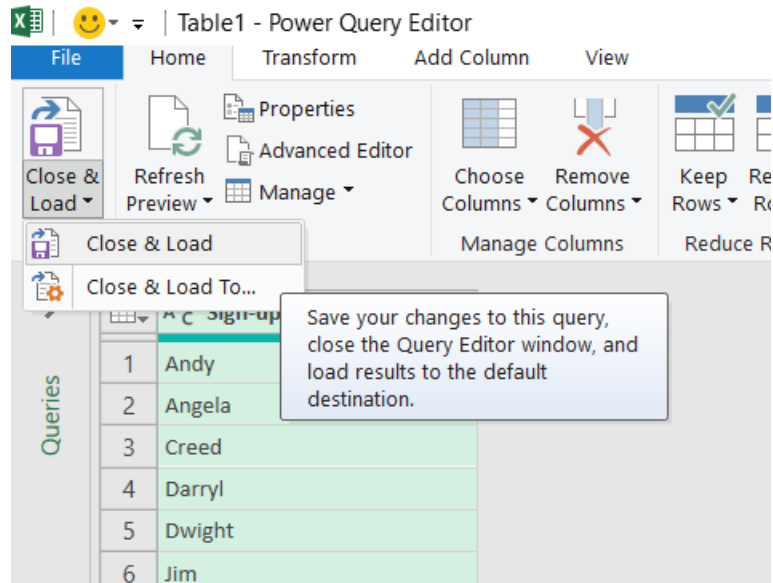
8. Let's remove the third step, Filtered Rows. Our dataset remains sorted A-Z, but nulls are no longer filtered out.
   a. **Careful: There is no "undo" for removing an applied step!**
9. Go ahead and re-filter the nulls from the data. You will see that becomes the last Applied Step.
10. You can modify the ordering of an Applied Step by right-clicking it.

11. Remove duplicates by going to Home on the ribbon, then Remove Rows -> Remove Duplicates.
    a. You'll also see there is an option here to remove blank rows, this would have been another way to filter out nulls.
12. Last but not least, there is a misprint in the data: a `Klevin` in here. We don't want that either, so filter it out.
13. On the upper left of the Home tab, there is a Close & Load menu. Click the drop-down and select Close & Load.
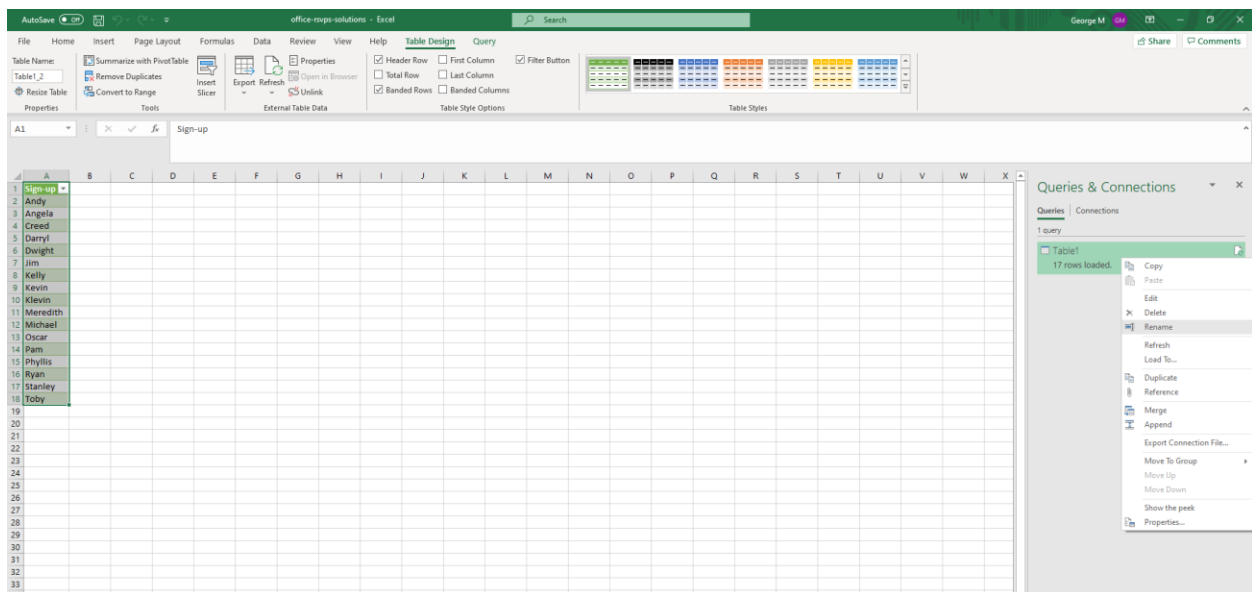
14. The result of our query has been *loaded* back into Excel (the L part of ETL!).
15. To the right of our table is a Queries & Connections menu. Our query is named `Table1`.
    That's not a very descriptive name, so let's rename it to `party_rsvp`.
    a. If you want to close out this menu, you can open it again under Data -> Queries &
       Connections.

16. Now, any changes made to our source data will be re-loaded into Power Query, go through each step of the data-cleaning process, and be loaded into this new table upon refresh.

17. For an example, I am going to insert two lines into my table, Roy and a blank row.

18. Go back to the loaded query, right-click and select Refresh.

19. Roy made it into the RSVP, the blank didn't and the results remain sorted alphabetically!