This sheet accompanies TechCamp Online: Excel 2008 Beyond the Basics for Mac users. This contains step-by-step instructions for the skills we covered in the online training.

Contents

Understanding Clean Data 1

Quick Fixes 2

Substitute Formula 2

Find & Replace 2

Review of filtering/sorting 2

Filtering 2

Sorting 3

Splitting Apart (or Parsing) Data 3

Text to Columns 3

Merging Data 4

Concatenate 4

“No name” Ampersand Formula 5

Formulas 5

If Formula 5

DATEDIF formula 5

Paste Special 6

Creating a Chart / Graph 6

Pivot Table 6

## Understanding Clean Data

Data has to be “clean” and accurate to be useful. Clean mostly means consistent.

Taking the accompanying spreadsheet, Practice File for Intermediate Excel, if we want to narrow our data down to get information about everyone in Washington, Texas, or everyone who paid a Full event fee, we can only do this if we give Excel consistent, correctly spelled and formatted data to manipulate.

“Dirty” data is missing info, is incorrect, or inconsistent – for example, when there are two spellings of a word, e.g. Bklyn vs. Brooklyn.

Spreadsheets & databases are not intelligent. Excel can’t know that Bklyn and Brooklyn refer to the same thing – and they really shouldn’t since they’re two different words. The data is only as good as the person’s attention is who enters it.

Look at data in Event Data and point out problem areas that are “dirty” – city, state, address.

## Quick Fixes

### Substitute Formula

Use the Substitute function to replace “Bklyn” with “Brooklyn”.

1. Insert a column after the **City State** column.
2. In cell to right of cell with “Bklyn” in it, type =SUBSTITUTE(E3, "Bklyn","Brooklyn") and hit Enter.
3. Copy the formula to remaining cells, select cell with formula in it and click the lower right-hand corner of the black outlining and drag to bottom of column. If it’s a long column, you can also double-click that corner and the formula will be applied to rest of column.

### Find & Replace

Find and replace does the same thing, but there is no need for a formula.

In order to limit the danger of using Replace, I always highlight the specific area of the spreadsheet to restrict the action to. Select the City column. Choose Edit, and click Replace. In Find, type Bklyn. In Replace, type Brooklyn. All strings found that match ‘Bklyn’ will be replaced with the new string.

For both this and **Find & Replace**, you have to be VERY careful. **+z** or **Command+z** for mistakes made!

## Review of filtering/sorting

### Filtering

Filtering is a way of narrowing down the data you’re looking at. An example is that you want to see only event attendees from Houston.

Click any cell in the document – no need to select anything.

Click **Data** > **Autofilter** to be able to apply a filter to any column in your spreadsheet. In the **City** column, click the arrow that appeared in the heading cell and you’ll see a dropdown of all the values in the menu. Click **Houston**. All records with *other* values in that column disappear (but don’t worry, they’re just hidden, not deleted).

When a filter has been applied, the arrow turns blue.

To remove the filter, click the arrow again in the header of that column and select **Show All**. If you have applied multiple filters, you can get rid of them all at once by returning to **Data** in the menu, and clicking **Autofilter** again, which deselects Autofilter.

### Sorting

Sorting is useful to automatically order data.

Click **Data** > **Sort…**. Select **Header Row** under **My data has** at the bottom, and now under **Sort by,** select the column that you want to sort (i.e. City). Finally, decide whether the order should be ascending or descending.

If it’s text, it will be put in alpha order; if numbers, in numerical order.

You can sort further by choosing a column under **Then by**.

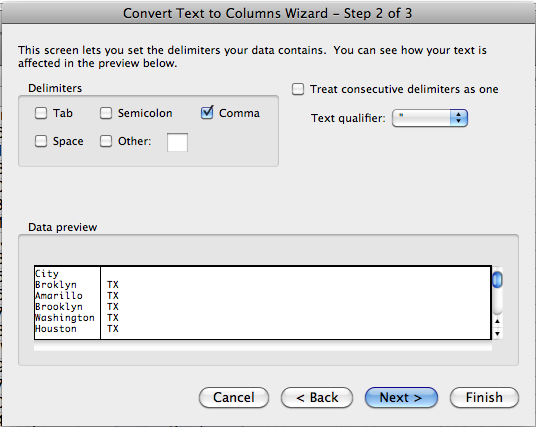
## Splitting Apart (or Parsing) Data

When trying to keep data clean, **it’s easier to have as little information as possible in each column**. This is also true if you’re doing something like an import into a database.

So rather than having **City** and **State** in the same column, you might decide to break them into two columns (for example, so that sorting is easier). This is also true of addresses – you’ll need to split street number and street name into separate columns to create a walk list for GOTV efforts.

### Text to Columns

1. Insert column after **City State** column.
2. **Insert** > **Columns**.
3. Highlight the **City State** column.
4. Click **Data** > **Text to Columns**.
5. Follow the wizard through these steps.
6. Select **Delimited** if the data you want to split up is divided by something consistent like a comma, space, or hyphen.
7. Click Next.
8. Click on the appropriate **Delimiter** (comma in this case).



1. Preview the data to make sure it works and click **Next**.
2. Choose the data format for the columns that will result. **Important note**: choose **Text** for zip codes and phone numbers so Excel doesn’t treat them like numbers\*.
3. Click **Finish**.

\*Otherwise, if a zip code has leading zero, the zero will be dropped after **Text to Columns** is applied. By default, Excel makes data type to General. Zip codes in the northeastern U.S. and many +4 digit zip codes start with zeros, and these zeros will be deleted unless you are careful about formatting cells to **Text**.

## Merging Data

The opposite of parsing or splitting data apart is merging data.

### Concatenate

Insert column after **Last Name** column.

Type **=CONCATENATE(B2, “ “, C2)** in a blank cell in your new column and hit **Enter**.

This should give you the contents of B2 and C2 with a space in the middle.

Double-click the lower right corner (known as the **Fill Handle**) of the cell you just typed the formula into to apply it to the entire column.

### “No name” Ampersand Formula

Insert column after **Last Name** column.

Type **=B2&” “&C2** in a blank cell in your new column and hit **Enter**. The ‘&’ between each of the items you want to put together acts as an unusual concatenate formula.

## Formulas

### If Formula

**I want to do a targeted direct mailing to everyone over 55.** I’ll need to divide the data into two groups – those older than 55 and those younger than 55.

Right now, if you used filter to do that, you would have to **manually deselect** all the ages that are younger than 55. Instead, I can use IF to say “if person is older than 55, put yes, if not, put no”.

1. **Insert column** after **Age** in **Voter Data** spreadsheet (should be column Z). We’re going to apply a calculation to Age. This is why it’s **important that age is treated like a number.** Our if formula must follow this structure: **=if(logical test, value if true, value if false).**
2. In second cell of new column (should be AA), type **“=if(Z2 > 55, “Yes”, “No”). Note:** Copying and pasting from this Word doc into Excel usually doesn’t work.
3. Hit **Enter**.
4. Double-click the lower right corner (known as the **Fill Handle**) of the cell you just typed the formula into to apply it to the entire column.

**Note:** 55 is being treated like a number, but ‘yes’ and ‘no’ are strings of text, so when we put strings in formulas, they always need quotations around them.

Now, you can easily filter the data for all of the people older than 55 by filtering for Yes. You could now copy and paste the results to a new document and do a mail merge to send your mailing just to those people.

### DATEDIF formula

This formula is not covered in the recorded webinar.

Notice that the ages are out of date. What if we want this spreadsheet to stay up to date and calculate ages given today’s date? We could recalculate ages based on the date of birth (DOB) column. Make sure DOB column has cells formatted as DATE.

1. **Insert column** after **Age** in **Voter Data** spreadsheet (should be column Z). This is going to be the Updated Age column.
2. Type **=DATEDIF(AB2, TODAY(), "Y")** and hit **Enter**. **Today()** is just a formula on its own that results in whatever today’s date is. **“Y”** results in years. You could also use “M” or “D” for months or days.
3. Double-click the lower right corner (known as the **Fill Handle**) of the cell you just typed the formula into to apply it to the entire column.
4. Insert a blank column. Follow the steps in the **If Formula** section to *accurately* calculate age based on today’s date and the DOB column. Now you know who’s actually 55 today. Your spreadsheet will be up to date today, in a week, and in a year’s time as more people on the list fall into the over 55 category.

## 

## Paste Special

### Paste Values

Paste Special allows you to pick and choose what you copy and paste in Excel. By default, you paste everything. You may instead want a cell to contain a value rather than the formula that results in that value. (I really want 4, not =SUM(2+2)).

To get values from formulas:

Highlight range of cells to copy.

**Edit** > **Copy** or on the keyboard, **+c**.

Click where you want to paste. **Edit** > **Paste Special**, click **Values** (rather than **All**).

### Transpose

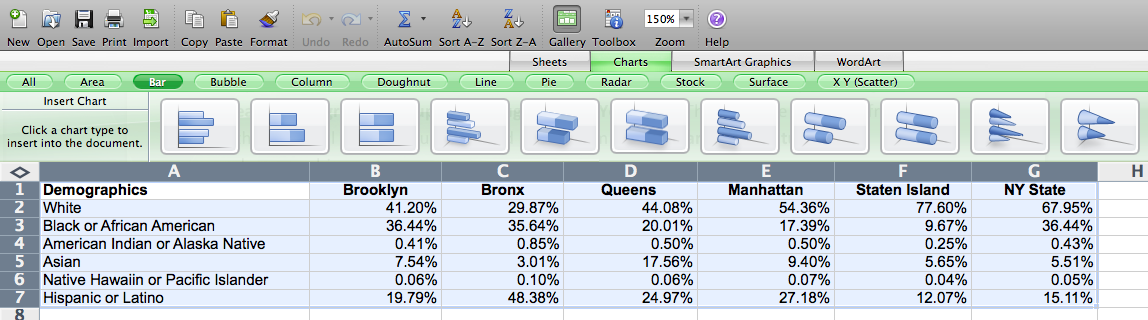
Switch to the **Charts** tab. Let’s say I get some new info about each of these demographic groups and I’d like to make a new chart where my headers (the boroughs of New York City) become the row headers. In other words, I want to preserve the order of the boroughs but make stack them vertically instead of horizontally.

1. Click and drag to select the headers.
2. Click **Edit > Copy**.
3. Click a cell where you want your “stack” of row headings to start.
4. **Edit > Paste Special**.
5. Select **Transpose** at the bottom. Transpose makes a column paste as a row, or a row paste as a column.
6. Click **Ok**.

## Creating a Chart / Graph

Move to a the worksheet called **Charts**. In order to make a useful chart, your data has to be set up in a way that will make it easy to make a chart out of. Notice that we’ve set up the data on the **Charts** worksheet so that each column contains the **percentages of each demographic group by borough in New York City**. The chart we generate from this data will take the numbers and make them into a **clear visual picture** of the demographics.

1. Click **Charts** immediately above the worksheet you’re in.
2. Alternately, click **Insert** in the menu and select **Charts**.
3. Highlight **A1-A7** and drag across the percentages, including headings.
4. Click the type of chart. Here, **Bar** is selected.



1. Click on a bar chart that will make it easy to understand your data.
2. Your chart will appear on your spreadsheet.
3. Resize the chart if needed by clicking a corner and dragging to make bigger or smaller.

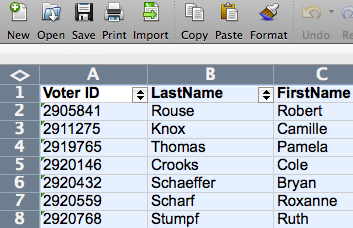
## Pivot Table

**Pivot Tables** allow you to manipulate multiple subsets of information and analyze them.

Using the Voter Data worksheet, we want to see how many voters of a certain age voted in which precinct. This will give me a picture of voting by age and precinct that can lead to more questions about a door-knocking campaign that my organization wants to do.

**Remove blank rows or columns.** Check that all columns have headings. Format data type correctly.

1. **Highlight the entire table** with column headings. Do this by clicking the upper left-hand diamond to the left of column A rather than by clicking and dragging.



1. Click **Data** > **Pivot Table Report.**
2. By default, Excel recognizes your data range.
3. Click Next.
4. Also leave **New Worksheet** selected, and click **Ok.**
5. Click **Layout**.

This is where you’ll set up what info goes into the chart from your data.

1. Click **Layout** tosee options based on the headings of your data.
2. Now, you can make different sets of information show up in the chart by dragging headings from your highlighted spreadsheet to the image that represents your chart. In this box, you’ll place row labels, column labels, and values.
3. Select **Precinct Name** and **Age** from the headings to the right**,** and drag them into **ROW** (We might even choose our **Updated Age** column that we created in the **DATEDIF Formula** section of this document).
4. We’ll also select **Party** and drag it into **Column.**
5. Take **Party** again and drag it into **Data**. This becomes **COUNT** of **Party**.
6. Click **Next** and then **Finish**.

Notice on the pivot table you’ve created that you can see how many voted per precinct. You can also see how many voted by which party (Dems/Repubs/Independent).

We may want to go back to our list and do another sort/if formula to divide voters into a couple of groups by age. Then we could recreate our pivot table based on those age groups.

Our chart appears when we click **Finish,** as does a box with all the headings, in case you want to add more data to your chart. To make changes, just click and drag a category into (or out of) the row, column, or count areas of the pivot table from the pop-up that appears.

You also have at your command the drop-down filters that we saw during the first session.

1. For example, you can use **drop-down filters** to view the data for only one precinct.
2. Right-click on **Sex, select “add to column labels”** to see the sex of people from each party. Now, notice that the information is broken down further by sex.