

Cheat Sheet for Manipulating Voter Data in Excel

Here's a cheat sheet of excel tips, formulas, and functions that can be used to manipulate sample voter data that was discussed or mentioned during the **Power On Network Online Presentation: Surviving Voter Data with Henry Serrano from Community Voices Heard**.

You can practice the formulas and functions described here by using the **Sample Voter List.xls** file.

Note: *The instructions described here pertain to MS Excel 2003. All of the functions described here exist in later versions of MS Excel, but you may have to scour the menus because some of the commands / functions may reside in a different place as described here.*

Excel Tips

Most often, whether you've exported data from a voter database, such as the VAN or CEL, or you've been able to acquire an electronic voter file from the state or county board of elections, the format of the file may exist in a "comma separated value" formatted file, or ".CSV" file (file extension).

It's easiest to view and manipulate this type of file into an Excel spreadsheet, by doing the following:

- Open Excel
- Open the voter file with the **.csv** file extension – making sure you select "ALL FILES (*.*)" in the FILES OF TYPE selection drop down box.
- Then, save this file as copy of the file, but this time as an Excel formatted file (*.xls), this way you can maintain the original voter data file. You'll then be making changes only to the Excel formatted copy of file.

Review & Understand Your Data

Take to the time to really look at and understand your voter data.

- Decide what information you want to keep and what you do not.
- Get to know your organizational database and the fields you will want to import data into. Add any new appropriate fields in your database if you plan on capturing data from the voter file that currently doesn't exist in your organizational database.
- Delete any columns of information you've decided not to use or integrate into your organizational database
- Look for any inconsistencies in the data and "clean" it up as best you can.
- Add additional columns of information to help you identify the voters you will be integrating into your organizational database, for e.g. if you want to identify these voters as contacts you made in your public housing November 2008 canvass, then add a column and add a code to each voter's record such as "Public Housing Nov 2008 Canvass."
- If you decide to add new columns of information in your voter file, make sure to always give a Name / Title for that column. It will make importing your data into your database much easier, if you can identify what the column of data is intended for.

Tips about using Excel Formulas



- When using formulas in Excel, you can just click the cell or range of cells you want to include in the formula and excel will fill in the actual cell number(s) for you.
- If you don't know what formulas to use you can:
 - Click on the **fx** function on the data entry line and look for the formula you want. It will even help you with entering the syntax of the formula;
 - Use Excel's Help menu; or
 - (my personal favorite) use Google to find out what function you would need to use, e.g. in Google, search for: "parsing addresses in Excel"
- The formulas and data manipulation tricks you decide to use depend on what the data looks like and what you intend to do with it. What is explained here is only a tiny glimpse of what you can do with data, but is still plenty powerful! Understand what your data is, and then decide on the appropriate tactics to use. **Keep it CLEAN!**

Parsing Data (splitting up data into multiple columns)

A. Parsing a phone number into Area code and Home phone

- Review the data in the PHONE column, and notice that the majority of the phone numbers look like there's a 3 digit area code followed by a 7 digit phone number.
- Also notice that in some records there is no phone number listed and then there's an unusual number in cell V62. You can decide what to do (delete the number, or try to call / find the person and get a correct phone number). For now let's just delete the number in V62, since we don't have a good number for this person.
- Insert 2 new columns after the PHONE column and name one column "AREA CODE" (now Column W) and the other "HOME PHONE" (now Column X)
- To parse out the Area Code:
 - In cell W2, type: **=LEFT(V2,3)**
 - Press <enter> and you'll see you've extracted the area code
 - Copy cell W2 and paste into the rest of the cells in column W
 - To change the formulas into actual values, see the excel tip "Changing formulas into values" toward the end of this document
- To parse out the Home Phone:
 - In cell X2, type: **=RIGHT(V2,7)**
 - Press <enter> and you'll see you've extracted the home phone
 - Copy cell X2 and paste into the rest of the cells in column X
 - To change the formulas into actual values, see the excel tip "Changing formulas into values" toward the end of this document

B. Parsing a full zip code into Zip5 and Zip+4

- Review the data in the REGADDRZIP column (which is now column AL) and notice that most of the zip codes have a hyphen ("-") to separate the Zip5 from the Zip+4.
- Notice that there are also zip codes that only reflect the Zip5
- Insert 1 new column and name it "Zip+4" (now column AM), the Zip5 number will replace the full zip code in the REGADDRZIP column after we do the following manipulation:
 - Highlight all the zip codes in the REGADDRZIP column
 - Click DATA from the menu, then select TEXT TO COLUMNS, and follow the wizard



- Make sure DELIMITED is selected under ORIGINAL DATA TYPE , and click NEXT
- In DELIMITERS, uncheck TAB
- Check the OTHER box, and type "-" (without the quotes)
- Look at the DATA PREVIEW to make sure the data looks the way you want, then click NEXT
- Since we are dealing with zip codes, and zip5 codes could have leading zeroes, make sure you change the COLUMN DATA FORMAT to TEXT, by clicking on the zip5 column under DATA PREVIEW, and selecting TEXT,
- Click FINISH
- Rename the REGADDRZIP column to "ZIP5"

C. Combining data into a full street address

If your voter file comes with the address separated into House number, street, apartment, etc., but your organizational database requires that the address needs this info in one (1) field, then you'll need to combine the fields. Let's suppose that there does a combined address field does not currently exist in the voter file:

- In some voter data files, you may see info for registration address and mailing address. You'll need to decide which address you want to keep (or both). In this case, let's work with the registration address.
- Best Practice tip: refer to the **Postal Service's Publication 28** for standard address formats: <http://pe.usps.com/cpim/ftp/pubs/Pub28/pub28.pdf>. Standardizing addresses, such as street, apartment, etc., is part of keeping your data "clean."
- If you have any blank columns in the registration address, you can ignore those.
- Look at REGADDRUNITTYPE and notice that some use "#". According to Publication 28, the "#" is not a standard format, so make sure to change all the "#" to "APT". Quick and dirty way is to Filter by "#", then use copy & paste to change it to "APT"
- First format the following columns as TEXT: REGADDRSTREETNUM, REGADDRSTREETNAME, REGADDRSTREETTYPE, REGADDRUNITTYPE, REGADDRSUNITNUMBER.
- Add one column before REGADDRSTREETNUM (now column AB), and name it "FULLREGADDR" or something like that, and leave this column formatted as GENERAL (the default).
- Place your cursor in cell AB2, then use either of the following formulas to combine the address fields. (Note: when using the formulas, you can just click the cell you want to include in the formula and excel will fill in the actual cell number for you.)
 - =CONCATENATE(AC2, " ", AF2, " ", AG2, " ", AI2, " ", AJ2) or
 - = AC2&" "&AF2&" "&AG2&" "&AI2&" "&AJ2
- Then copy and paste the formula to the rest of the cells in the FULLREGADDR column.

D. Parsing a full street address into street number, street, street type, apartment

Let's reverse engineer the FULLREGADDR column to learn how to parse out the individual address components.

- Before you can continue parsing, you will see (if you click on any cell in the FULLREGADDR column) that the cell holds the formulas. We need to change the Formulas into actual values before we do any further calculations or parsing (also see "CHANGING FORMULAS INTO VALUES" later in this document).
 - Insert a new column after FULLREGADDR and name it FULLREGADDR NEW (now column AC)

- Highlight all the cells under FULLREGADDR and hit CTRL+C (or click EDIT, select COPY)
 - Click in cell AC2 and click on EDIT -> PASTE SPECIAL, and select VALUES, then OK. You can now delete the column FULLREGADDR that has just the formulas, and FULLREGADDR NEW is now column AB.
- To Parse out the street number:
- Insert a column after FULLREGADDR NEW and name it STREETNUMBER (now column AC)
 - In cell AC2, type: **=LEFT(AB2,FIND(" ",AB2)-1)**
 - then copy and paste the formula in the rest of the cells in the STREETNUM.
 - Remember that the cells now in STREETNUMBER are only formulas, so you'll need to change these into values later before you import, or do any other calculations, as we did above. *See also "CHANGING FORMULAS INTO VALUES" later in this document.*
- To Parse out the street name, street type, apt, and apt number:
- Insert a new column and call it STREETAPTNUMBER after STREETNUMBER, (now column AD)
 - In cell AD2, type: **=RIGHT(AB2,LEN(AB2)-FIND(" ",AB2))** - this formula finds everything after the space after the street number
 - Then copy and paste the formula to the rest of the cells in the STREETAPTNUMBER column
 - Remember that the cells now in STREETNUMBER are only formulas, so you'll need to change these into values later before you import, or do any other calculations, as we did earlier. *See also "CHANGING FORMULAS INTO VALUES" later in this document.*
- To Parse out the street name and street type:
- Make sure you've changed the formulas into values in the STREETAPTNUMBER column, we'll refer to this column as also column AD.
- Insert a new column after STREETAPTNUMBER column and call it STREETNAME, now column AE
 - In Cell AE2, type: **=LEFT(AD2,FIND(" APT ",AD2))** - (note the space before and after the letters APT) this formula looks for the street name and street type before " APT "
 - Then copy and paste the formula to the rest of the cells in the STREETAPTNUMBER column
 - You'll notice there are some cells with "#VALUE1" because there no " APT " was found. For these, you can manually just copy the street name over.
 - Remember that the cells now in STREETNAME are only formulas, so you'll need to change these into values later before you import, or do any other calculations, as we did earlier. *See also "CHANGING FORMULAS INTO VALUES" later in this document.*
- Parse out the Street type:

In this case we can Text to Columns, since a space separates that street name from street type

- Insert a new column after STREETNAME column and call it STREETTYPE, now column AF
- Highlight all the data in the cells below STREETNAME
- Click on DATA on the menu, then select TEXT TO COLUMNS
- Follow the wizard, make sure DELIMITED is selected,
- Click NEXT and check the SPACE checkbox under Delimiter
- Click NEXT, and leave the columns formatted as GENERAL and click FINISH.

➤ Parse out the apt type and apt number:

Since we know that what we're looking for are only "APT" followed by the apartment number we can do the following:

- Insert a new column after STREETTYPE and call it TYPEAPTNUMBER after (now column AG)
- In cell AG2, type: `=RIGHT(AD2,LEN(AD2)-FIND(" APT ",AD2))` - (note the space before and after the letters APT)
- Then copy and paste the formula to the rest of the cells in the TYPEAPTNUMBER column
- You'll notice there are some cells with "#VALUE1" because there are no apartment types or numbers for these records. For these, you can manually enter a blank.
- Remember that the cells now in TYPEAPTNUMBER are only formulas, so you'll need to change these into values later before you import, or do any other calculations, as we did earlier. *See also "CHANGING FORMULAS INTO VALUES" later in this document.*

➤ Parse out the apt number:

Look at our data in TYPEAPTNUMBER, we can see that a space separates "APT" and the apartment number, so we can then do the following:

- Insert a new column after TYPEAPTNUMBER column and call it APTNUMBER, now column AH
- Rename column TYPEAPTNUMBER to APTTYPE (column AG)
- Highlight all the data in the cells below APTTYPE
- Click on DATA on the menu, then select TEXT TO COLUMNS
- Follow the wizard, make sure DELIMITED is selected,
- Click NEXT and check the SPACE checkbox under Delimiter
- Click NEXT, and leave the columns formatted as GENERAL and click FINISH.

E. Using the IF function to set a "YES" or "NO" value based on if a phone number exists or not

Let's say we want to track a YES / NO field in our organizational database indicating if a phone number exists or not. Sometimes it's easier to set this field in the excel spread sheet. We've already parsed and cleaned up the phone numbers earlier:

- Remember that the data in the cells under HOME PHONE (column X) are only formulas, so you'll need to change these into values later before you do any other calculations, as we did earlier. *See also "CHANGING FORMULAS INTO VALUES" later in this document.*
- Insert a column after HOME PHONE and name it "PHONE?" (now column Y)
- In cell Y2, type: **=IF(X2="", "NO", "YES")** - this formula will enter a "NO" if there is no phone number, and a "YES" if a phone number exists.
- Copy and paste the formula to the rest of the cells in the PHONE? Column.
- Remember that the data in the cells now in PHONE? are only formulas, so you'll need to change these into values later before you import, or do any other calculations, as we did earlier. *See also "CHANGING FORMULAS INTO VALUES" later in this document.*

F. Changing formulas into values

- Before you can do further calculations, or importing into your organizational database, you'll need to check your data to make sure that you're working with actual values and not formulas.
- For example, if you click on any cell in the AREA CODE column (column W) that the cell holds the formulas. We need to change this into actual values before we do any further calculations or parsing.
- Insert a new column after AREA CODE and name it AREA CODE NEW (now column X)
- Highlight all the cells under AREA CODE and hit CTRL+C (or click EDIT from the menu, then select COPY)
- Click in cell X2 and click on EDIT on the menu, then select PASTE SPECIAL
- Select VALUES, then OK.
- You can now delete the column AREA CODE that had the formulas