

Data crunchers, this session is for you

FOIAfest 2016 | March 12, 2016

Denise Linn, Program Analyst for the Smart Chicago Collaborative

smartchicagocollaborative.org

@DKLinn | @SmartChicago



Agenda

Introductions

Some Excel Tricks & Time Savers

Quick, Easy Mapping: Chicago Library Branches

Analysis in Excel: Lobbyist Compensation in Chicago

- Filtering, IFs, COUNTIFs, SUMIFs, Pivot Tables....

Resources

Introductions

Tricks & Time Savers

Some basics:

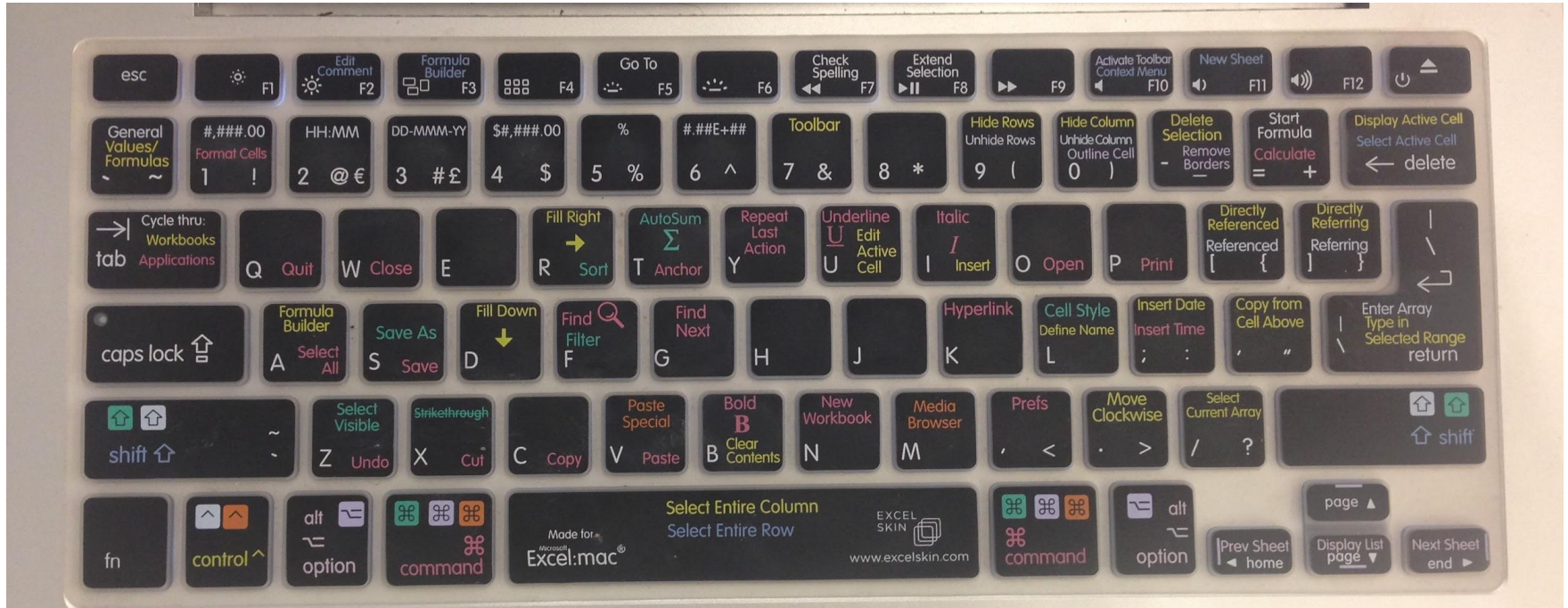
Operations = + * / ^

Order of Operations: BEDMAS

Basic formulas: =SUM(), =AVERAGE(), =COUNT(), =RANK()

Cheatsheets are your friend! The Internet has many to offer. Here is a good one: <http://www.collegedegreeseach.net/excel-cheats/>

Shortcuts for Mac:



Time savers

Shift + space → select an entire row

Control + space → selects an entire column

Control + a → selects an entire worksheet

Ctl/Command + up key → jump up to the top of the workbook

Ctl/Command + down key → jump up to the bottom of the workbook

Time Savers (cont.)

Excel will finish patterns in a row or column for you — days, dates, etc.

Double clicking on the bottom right box of a cell with a formula or pattern in it will continue that calculation or pattern for the entire column of work - no need to click, hold, and scroll.

Let's Map Chicago's Public Libraries

First - find the dataset

Go to Chicago's Open Data Portal

Search, "Library Locations Hours Contact". You'll find this Dataset: <https://data.cityofchicago.org/Education/Libraries-Locations-Hours-and-Contact-Information/x8fc-8rcq>

Click "Export." On the right hand side, choose "CSV for Excel"

Open the file.

We have enough to make a map

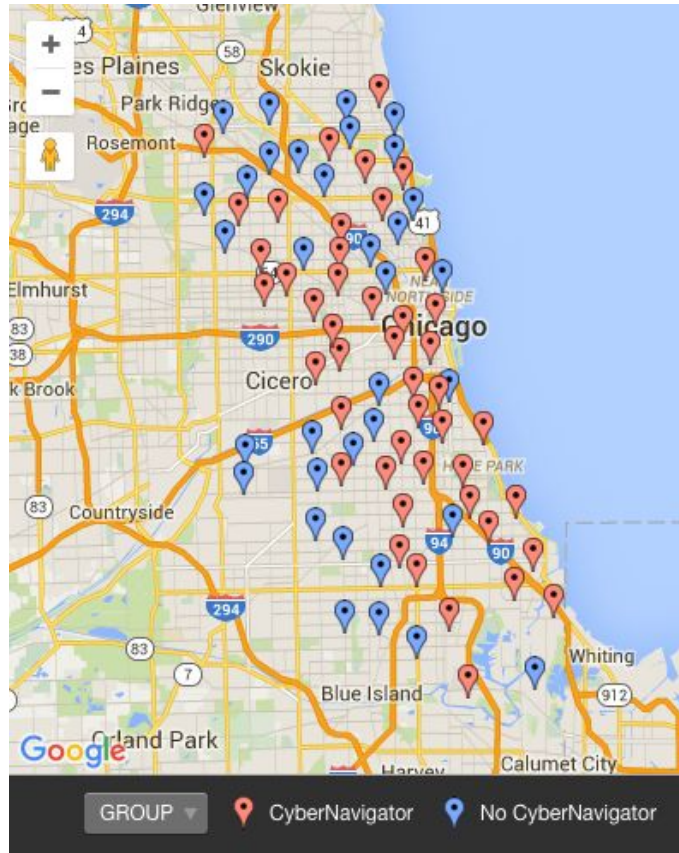
Insert a blank to the right of “CYBERNAVIGATOR.”

- In D2 type formula
 - =IF(C2="Yes","CyberNavigator","No CyberNavigator")
- Doubleclick the lower right box to continue the formula for the whole sheet.
- Label cell D1 “GROUP”

Now go to <http://batchgeo.com>. Copy the whole worksheet from Excel and paste it into the webpage. Click “Map Now” and wait.

Note the cool “advanced” options you have for mapping.

The Result:



Let's Use Excel to Analyze Chicago's Lobbyist Compensation Data

First, Let's Find the Data

Let's go to Chicago's Open Data Portal <https://data.cityofchicago.org>

Search, "Lobbyist Compensation"

Click "Export." On the right hand side, choose "CSV for Excel" (so we can capture all the rows)

Open the "Lobbyist Data Compensation" file. Save on your machine as an xlsx file.

Familiarize Yourself with the Data

Note that every row is a unique compensation ID

Note that clients, lobbyists, and lobbyist codes repeat.

Note this data is between 2012 - 2015.

Note the lobbyist codes & names


There are 24,391 compensation IDs!

What are the first questions that spring to your mind when you see this information? Has anyone worked with this before?

Make the Data Easier to Work With

Preserve the original data before you start working with it! Copy & paste the whole “Lobbyist_Data_-_Compensation.cs” worksheet onto a different sheet. Label that second sheet “Analysis”.

Click cell A2 on the second row of the “Analysis” sheet. Go to Window and select “Freeze Panes”

Click a cell on the top row, then click the filter icon. 

What’s the first thing you want to do?

The Extremes

To find the smallest compensation amount:

- =MIN(H2:H24392)

To find the largest compensation amount:

- =MAX(H2:H24392)

Click the filter button next to “COMPENSATION_AMOUNT” and select “descending.” Who is the client that paid out the largest compensation amount? Which lobbyist did it go to?

The “Average”

To find the average compensation amount:

- =AVERAGE(H2:H24392)

To find the median compensation amount:

- =MEDIAN(H2:H24392)

To find out how many compensation amounts have been over \$100K:

- =COUNTIF(H2:H24392, ">100000")

Takeaway: There are outliers — substantially larger compensation amounts that pull the average up.

Summing Compensations by Client

On your “Analysis” sheet, filter the COMPENSATION_AMOUNT in descending order. Look at the client names. Anything strike your interest?

Calculate the total amount of compensation given by any client by using =SUMIF(J2:J24392,"NAME",H2:H24392). For example:

- =SUMIF(J2:J24392,"Lucas Museum of Narrative Art",H2:H24392)
- =SUMIF(J2:J24392,"CVS Caremark Corp",H2:H24392)
- =SUMIF(J2:J24392,"Uber Technologies",H2:H24392)

Summing Compensations by Client (cont.)

Sort Client names Z → A. Find “Uber” under client name (CTL/Command + f). What else do you see?

Some clients are listed under several related names. This makes things difficult for us and our SUMIFs...

One way to tackle this issue: pivot tables!

A pivot table helps us group, regroup, and analyze our data in an interactive way.

Prepping Our Data for Pivot Table Making

Go to the original “Lobbyist_Data_-_Compensation.cs” worksheet. Select all, copy, & paste into a new sheet. Label this, “Pivot Table”

In the sheet “Pivot Table,” select the column “LOBBYIST_MIDDLE_INITIAL” and delete it.

- FYI, we’re only doing this because, if we want to make a pivot table, we can’t have blank cells

Create a new column “G” after “LOBBYIST_LAST_NAME.” Label G1 “LOBBYIST_FULL_NAME”. Into G2 type

- =CONCATENATE(“E1”, “ ”, “F1”)
- Double click the right bottom corner!

Making a Pivot Table

Create PivotTable

Choose the data that you want to analyze: _____

Use a table or a range in this workbook

Location:

Use an external data source

No data fields have been retrieved.

Choose where to place the PivotTable: _____

New worksheet

Existing worksheet

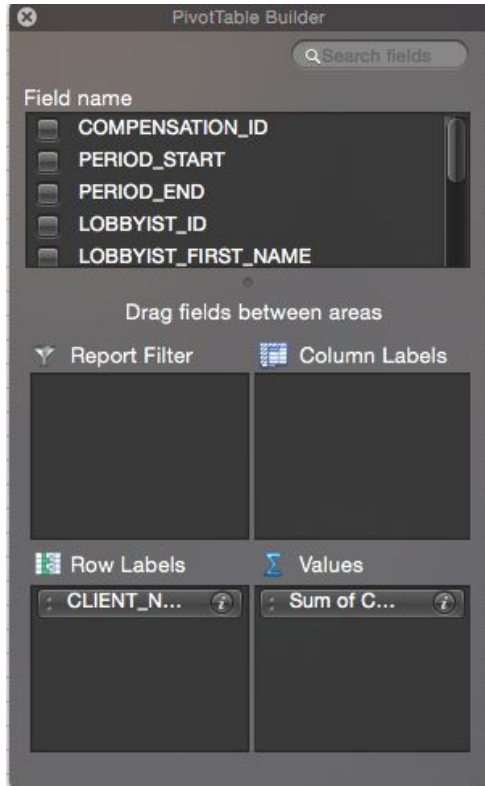
Location:

On the “Pivot Table” sheet, click “refresh data” button.

Select any cell, go to the Pivot Table icon under the “Data” section, and select “Create Manual Pivot Table...”

This window will pop up. Make sure you have the following selections chosen (see left)

Building Pivot Table #1



“Pivot Table Builder” will pop up. It will look like the image to the left.

- Drag and drop “CLIENT_NAME” under Row Labels.
- Drag and drop “Sum of Compensations” under Values.

Click on the B5 cell. Go up to the Sort Icon and select “descending”.

What do we see?

Report Filter

Sum of COMPENSATION_AMOUNT

Row Labels

Total

American Beverage Association

4391322.17

Uber

3253500

Noresco LLC

3108000

CVS Caremark Corp

2990000

Academy for Urban School Leadership

2737500

Dell

2472000

Family Guidance Centers, Inc

2388000

Delaware North Companies

2365104

United Technologies Corporation

2250000

Hilton Worldwide

2190000

Interstate JCDecaux, LLC

1845475.33

Wal-Mart Stores, Inc.

1842111.13

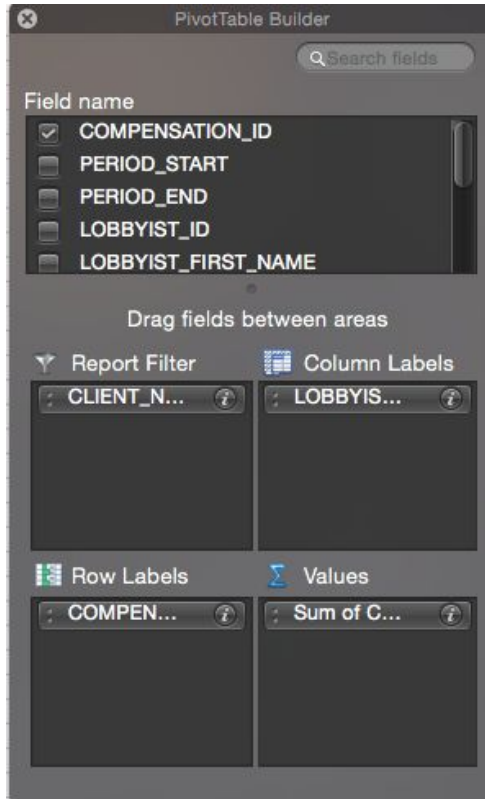
Presence Health System

1835559.5

Citibank, N.A.

1686000

Building Pivot Table #2



Make a new pivot table in another sheet.

“Pivot Table Builder” will pop up. It will look like the image to the right.

- Drag and drop “CLIENT_NAME” under Report Filter.
- Drag and drop “LOBBYIST_LAST_NAME” under Column Labels.
- Drag and drop “COMPENSATION_ID” under Row Labels
- Drag and drop “Sum of Compensations” under Values.

Pivot Tables V

Go to the filter icon in B1, search for “Uber,” and select Uber, Uber Technologies, and Uber Technologies, Inc. This is the pivot table you will see:

	A	B	C	D	E	F	G	H	I
1	CLIENT_NAME	(Multiple Items)							
2									
3	Sum of COMPENSATION_AMOUNT	Column Labels							
4	Row Labels	Alexander	Dring	Johnson	KASPER	Kellman	NOTTAGE	Rosenfeld	Grand Total
5	8414		27000						27000
6	8972				27000				27000
7	9045						27000		27000
8	9283				27000				27000
9	9309						27000		27000
10	9333		27000						27000
11	9817				27000				27000
12	9826						27000		27000

Play with this pivot tables more...

Build a pivot table displaying compensations sums by Lobbyist full name:

- Drag & drop “LOBBYIST_FULL_NAME” into “Row Labels”
- Drag & drop “COMPENSATION” into the summation box
- Sort descending!

To look up clients and money BY lobbyist....

- Drag & drop “LOBBYIST_FULL_NAME” into “reporter filter” category
- Drag & drop “CLIENT_NAME” into the “Row Labels” box
- Drag & drop “COMPENSATION” into the summation box

More resources & other free tools:

Visualization tools:

<http://timeline.knightlab.com/>

<http://timemapper.okfnlabs.org/>

<https://datawrapper.de/>

[Microsoft Power BI](#)

<https://storymap.knightlab.com/>

[Google sheets/charts](#)

Excel help:

<https://support.office.com/en-us/excel>

<https://support.office.com/en-us/article/Excel-functions-by-category-5F91F4E9-7B42-46D2-9BD1-63F26A86C0EB> (all functions by category)

Questions?