# Excel and Civic Data

This is the instructor’s guide for delivering the Excel and Civic Data workshop. It is modular, in that you can use each part independently. It is meant to be used along with the Excel files in this accompanying folder, but can also be used with any other open government data. It works best with data that includes location information, and some kind of date information (like an address and a completion date).

For this tutorial, you should use **Excel 2016 for Windows.** While most of this can be done in Excel 2013 for Windows, some of the menus and names have changed. Many of the tools are specific to Excel for Windows, so if you have attendees using a Mac, they should still use the Excel Windows version using either Basecamp or Parallels.

If you have students who are using Excel 2013, you will want to [download](https://www.microsoft.com/en-us/download/details.aspx?id=39379&CorrelationId=10f925dd-f2f6-4f2c-8d0b-c9978aad50eb) [Power Query](https://support.office.com/en-us/article/Introduction-to-Microsoft-Power-Query-for-Excel-6E92E2F4-2079-4E1F-BAD5-89F6269CD605) to access the functionality discussed in this tutorial. You will need to [download](https://www.microsoft.com/en-us/download/details.aspx?id=38395) [Power Map](http://labs.office.com/guides/Creating%20Cinematic%203-D%20Data%20Stories%20with%20Power%20Map%20for%20Excel.htm) if you are using Excel 2013 to access the mapping functionality. Some of the new chart types are specific to Excel 2016. It is not recommended for versions earlier than Excel 2013.

The workshop is divided into the following sections:

* Accessing data
* Analyzing data
* Mapping data (Windows only)
* Transforming data
* Advanced analytics (an optional module)

The conventions used in the script are as follows:

* + *In italics are things you type*
  + In non-italicized are things you say
  + **Note**: denotes an instructor note.

Getting Stared

Excel is a tool that most people are familiar with, and many people use every day. Excel can be leveraged to unlock the value of open data of all kinds, and it is particularly well-suited to query and visualize open government data from multiple sources. In this tutorial, you will learn how to access, transform, cleanse, query, and visualize data using tools that are easy to learn and easy to use.

Over the next two hours, I'm going to show you how to use Excel to access, manipulate, and visualize civic data, with an emphasis on Chicago’s city data. I'll give you some tips and tricks that I have learned based on experience.

And then I will show you some add-ons you can use to make mapping data and querying for data a lot easier

**Note for instructor:** this tutorial was built using Excel for Windows. Although most of these tools work on the Mac, mapping does not. If users are using Macs, it is suggested that they use Excel for Windows on the Mac using [Parallels](http://www.parallels.com/) or a similar technology.

**Note for instructor**: this tutorial assumes you have some datasets in mind that you would like to show your audience. All examples can be replaced with your own data. If you do not have data in mind, you can use the examples I have used here. The scenario I use is Chicago's open data, and the files are located here: <https://doc.co/smWQbx>

Accessing Data Sources in Excel

Before you look for, download, or use any data, it is very important that you define what problem you are trying to research, and/or what problem you are trying to solve. Being clear about this will make the rest of this much easier. Having said that, there is value to browsing the data that state, county, and city governments have made available, as it may inspire ideas for what you might be able to accomplish.

For the next 20 minutes (note: you can make this more or less), we can explore the data that is out there. (*use this section to show the data sources you plan to use. For example, if you are using a city data portal, explore the different data sets, how to find the right data, the data types, etc.). In this example, we use the City of Chicago's data portal at* [*http://data.cityofchicago.org*](http://data.cityofchicago.org)

There are many ways to bring data into Excel -- you can import data, you can connect to data, and you can subscribe to data. In this tutorial, you will learn how to do all three and see which approach is right for your needs.

*Open City Data Portal*

This is Chicago’s City Data Portal. It contains hundreds of data sets.

*Show the interface of the data portal*

*Show the categories and types of data (tables, maps, etc)*

*Click on Export and show (and explain) the export options, and describe when to use what*

*Show that it contains both datasets and maps. The maps can be downloaded and used later in the Mapping section*

Now the most popular data set is Salary info, as I can see when I am sorted by "Most Accessed". Let's get that one in Excel.

*For the Current Employee Names, Salaries, and Position Titles data set, open the Export and scroll down to and open the oData box.*

**Note for instructor**: this part requires you to find a dataset that uses the oData standard. It appears as a URL and allows you to subscribe to the data set in Excel If you are taking data from a Socrata portal, most of the data sets support oData. You can find the oData link in the Export options of the dataset.

Notice that I can export in a lot of formats. One particular format, called oData, allows you to subscribe to the data set as opposed to exporting it. That way, you can always be sure you are getting current data. Every time you refresh or open your workbook, it goes back and refreshes the data.

*Select and copy the oData URL*

*On the Data Tab, show the buttons on the Get External Data section.*

*Click on the From Other Sources arrow and choose "From oData Data Feed"*

*Paste the URL into the Link or File box*

Machine generated alternative text:
Data Connection Wizard 
Connect to a Data Feed 
Enter the information required to connect to a data feed. 
I. Location of the data feed 
Link or File: 
2. Log on credentials 
@ use the sign-in information for the person opening this file 
arm,vse... 
C) use this name and password 
user Name: 
Password: 
Cancel 
Back 
Next > 

**NOTE for instructor**: This can take a while depending on the size of the dataset and your wireless speed. I suggest you either use a small data set, or show this part of the way, but have the dataset downloaded already to work with so they can see how it is done without having to wait for it to actually be done.

*Click on Next*

If there are multiple tables, you can choose the one you want to work with

*Click on the only table there*

*Click Finish*

Now here, it offers me some shortcuts.

Machine generated alternative text:
Import Data 
Select hcn,v you want to view this data in your workbook. 
@Table 
C) PivotTable Report 
C) PivotChart 
C) Power View Report 
C) Only Create Connection 
Where do you want to put the data? 
@ Existing worksheet: 
C) New worksheet 
Add this data to the Data Model 
Pcoperties... 
Cancel 

For example, I can have the data converted right away into a pivot table or chart

But in this case, we are going to have it formatted as a table

*Click OK*

Note that the data is already formatted as a table.

Web Data

With the data access tools in Excel, it is easy to get data right off of tables that actually live on a web page, as opposed to in a data set. For example: Wikipedia has a list of Chicago’s community areas and neighborhoods. Chicago has 77 well defined community areas, and within those community areas, there are about 200 neighborhoods (less formally defined). I can query that page directly

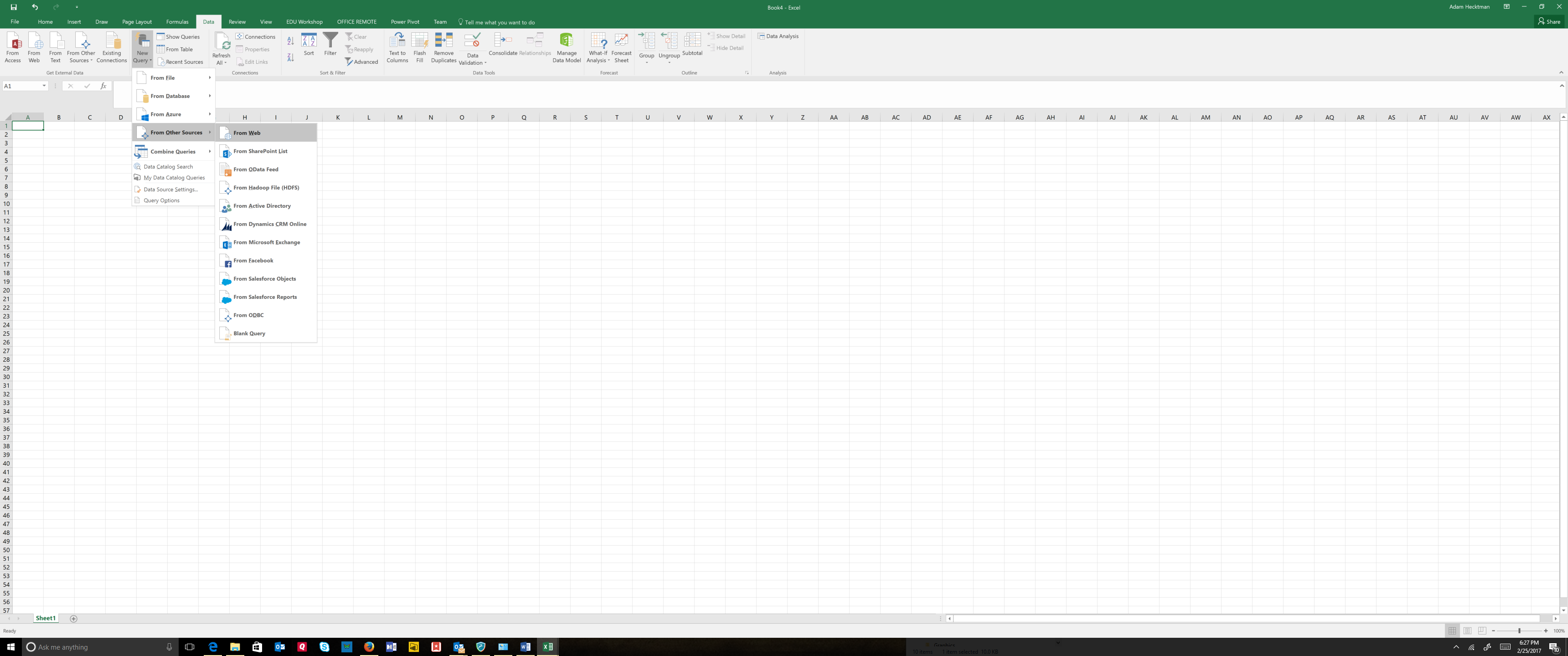
*Show the Wikipedia page: https://en.wikipedia.org/wiki/List\_of\_neighborhoods\_in\_Chicago*

*Click on the Data tab on the ribbon*

*Under the Get External Data section, click New Query and on from other sources*

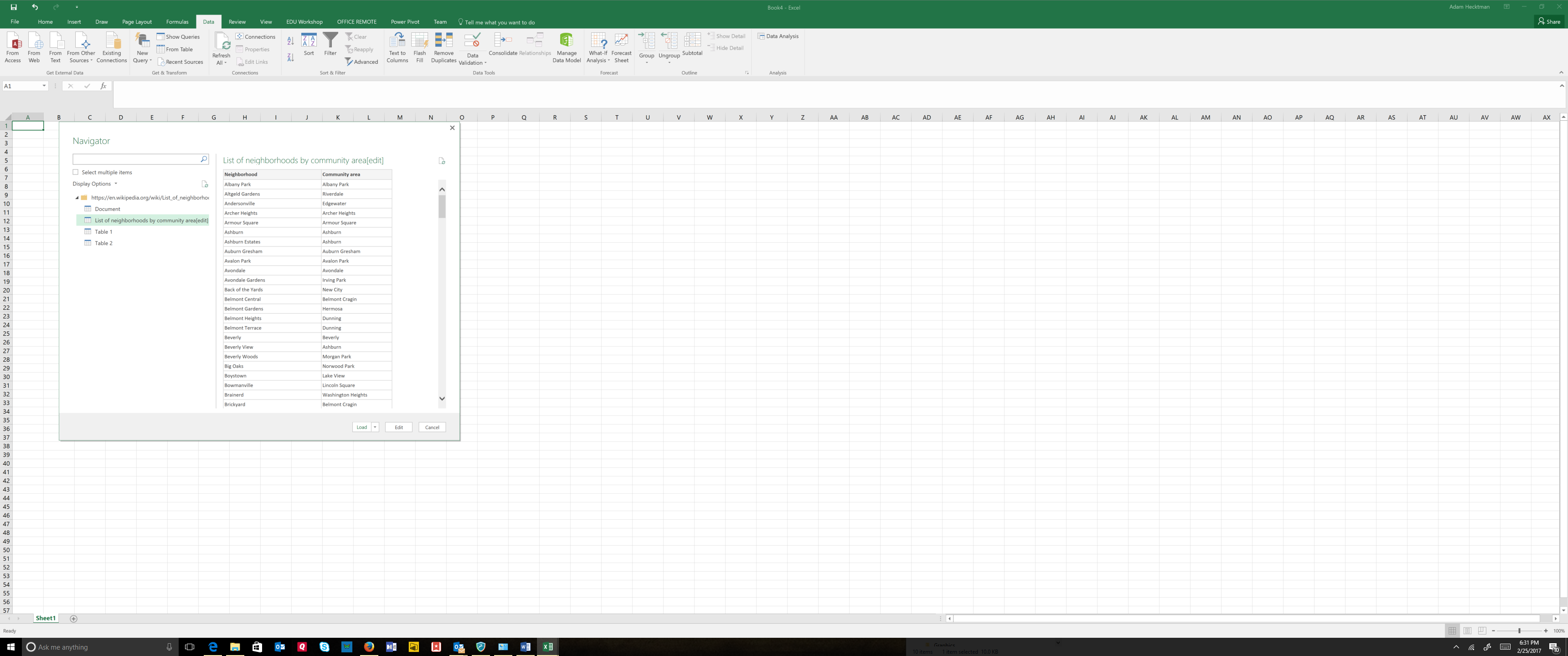
*Choose From Web*

*For the URL, give it the same URL as the web page you opened (*[*https://en.wikipedia.org/wiki/List\_of\_neighborhoods\_in\_Chicago*](https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago)*)*



You will see that it lists the tables it finds on that page. I’ll choose the list of neighborhoods and community areas and preview it to make sure it is what I want. If it needs to be edited, I can do that from here, but for now we will simply load it.

*Choose the List of Neighborhoods…table and click on the Load button*



This created a pre-formatted table for me, complete with filtering options. If I want to filter in on a specific community area, I can click on the filter and find it.

*Click on the filter arrow next to Community Area*

*Start typing Lakeview*

*Show that eight neighborhoods show up*

*Clear the filter*

Now this is good, but what I really want to do is enter my data by seeing each community area in one column, and the corresponding neighborhoods within those areas in another. This is an ideal scenario for a Pivot Table.

*Click inside the table*

*Click Pivot Table from the Insert Tab (in the Tables section)*

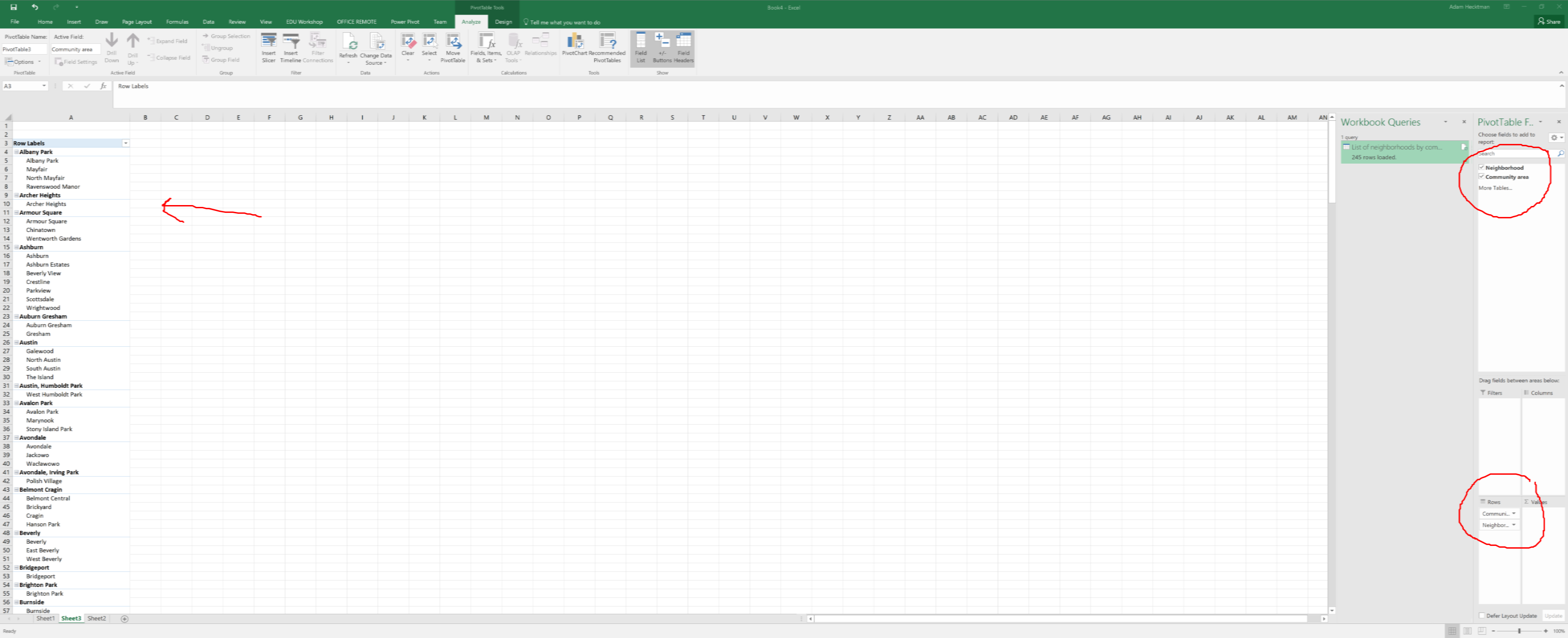
*We will use the defaults, so choose OK*

This will give you a very visual way to create what you want to see. To build the pivot table, simply choose the fields that you want from the field list, and where you want them to appear. In this case, we want rows for community area, and sub-rows for neighborhoods.

Do you want to see this with Cleveland data? Do it again with <https://en.wikipedia.org/wiki/Cleveland#Demographics>

Let’s go back to the Chicago neighborhood data.

*Drag Community Area to Rows, and Neighborhood below Community Area*



Now I can start entering my data by neighborhood.

Analyzing Data

Now, because it came in as a table, I have some features already lit up for me. For one, it gives you the ability to filter.

*Show filters by filtering to a department - for example, filter just the Aviation department.*

*Then clear the filter.*

Next I can do some of the standard formatting things like:

*Take salary and format the entire column as $*

*Show how you can use the Reduce the Decimal button so that it is only dollar values, no cents*

Now, let's take a look at some features that help you visualize the data quickly.

Let's say that everyone gets a bump next year in their salary of 5%

*In the empty column to the right of employee salary type a name first row - something like "New Salary"*

Note what happened here. Excel automatically it formats the column in the style of the table for me, and it made it actually part of the table. Let's get that data in there for salaries.

*In the first empty row (2) start typing "=" and then click on the column to the left.*

*Show that in the formula bar, instead of a cryptic row and column, it gives you a meaningful name*

*Now type \* 1.05 and hit enter.*

Machine generated alternative text:
UEL 
'HER 
Font 
=[@[employee_annual_salary]]*1.05 
job_titles 
WATER RATE TAKER 
POLICE OFFICER 
POLICE OFFICER 
CHIEF CONTRACT EXPEDITER 
CIVIL ENGINEER IV 
ASST TO THE ALDERMAN 
GENERAL LABORER - DSS 
TRAFFIC CONTROL AIDE-HOURLY 
STAFF ASST TO THE ALDERMAN 
Alignment 
department 
WATER MGMNT 
POLICE 
POLICE 
GENERAL SERVICES 
WATER MGMNT 
CITY COUNCIL 
STREETS & SAN 
OEMC 
cm COUNCIL 
Number 
New 
90,744 —[@[employee_annual_salaryll 
*1.05 
84,450 
84,450 
89,880 
106,836 
70,764 
41,850 
20,051 
49,452 

*Hit enter*

Notice that Excel finished the formatting of the formula for me. This is the autocorrect feature at work, so that you have "=[@[Employee Annual Salary]]\*1.05" in the formula bar

Excel also automatically predicts that you want the entire column to look like this and does a Flash Fill for you. To make it look like the old salary, you can do the formatting again

*Format as dollars by clicking $ on the toolbar.*

**Quick Analysis Tools**

It can be tough to extract meaning out of a large set of numbers in a worksheet. Excel provides a way to quickly analyze your data using conditional formatting. For example, you can use a color scale to differentiate high, medium, and low housing values. Called the Quick Analysis tool, it will quickly analyze and visualize data for you. Excel has Quick Analysis tools that can take rows or columns of data and add visuals.

*Click on a cell in the table*

*Ctl+Q to bring up the quick analysis tools*

*Hover over the data bars, the color scale, the icons*

*Choose Data Bars and how they provide a visualization within the cell*

Machine generated alternative text:
Font 
job_titles 
WATER RATE TAKER 
POLICE OFFICER 
POLICE OFFICER 
CHIEF CONTRACT EXPEDITER 
CIVIL ENGINEER IV 
ASST TO THE ALDERMAN 
GENERAL LABORER - DSS 
TRAFFIC CONTROL AIDE-HOURLY 
STAFF ASST TO THE ALDERMAN 
ELECTRICAL MECHANIC 
FIRE ENGINEER-EMT 
105,336 
POLICE OFFICER 
FOSTER GRANDPARENT 
CLERK Ill 
INVESTIGATOR - 'PRA Il 
POLICE OFFICER 
POLICE OFFICER 
FIREFIGHTER (PER ARBITRATORS AWARD) 
POLICE OFFICER 
FIREFIGHTER/PARAMEDIC 
ENGINEERING TECHNICIAN VI 
FIREFIGHTER-EMT 
LIEUTENANT 
CROSSING GUARD 
Alignment 
-PARAMEDIC 
ELECTRICAL MECHANIC-AUTO-POLICE MTR MNT 
CIVIL ENGINEER Il 
GENERAL LABORER - DSS 
FOSTER GRANDPARENT 
POLICE OFFICER 
PARAMEDIC 
MOTOR TRUCK DRIVER 
LIBRARY ASSOCIATE- HOURLY 
POLICE OFFICER 
SENIOR PROGRAMMER/ANALYST 
CIVIL ENGINEER IV 
ENGINEERING TECHNICIAN V 
POLICE OFFICER 
SUPERVISING TRAFFIC CONTROL AIDE 
SANITATION LABORER 
POLICE OFFICER 
AMBULANCE COMMANDER 
POOL MOTOR TRUCK DRIVER 
FIREFIGHTER-EMT 
SEWER BRICKLAYER 
TREE TRIMMER 
TRAFFIC SIGNAL REPAIRMAN 
PERSONAL COMPUTER OPERATOR Ill 
SENIOR COMPANION 
department 
WATER MGMNT 
POLICE 
POLICE 
GENERAL SERVICES 
WATER MGMNT 
CITY COUNCIL 
STREETS & SAN 
OEMC 
cm COUNCIL 
AVIATION 
FIRE 
POLICE 
FAMILY & SUPPORT 
POLICE 
1 PRA 
POLICE 
POLICE 
FIRE 
POLICE 
FIRE 
WATER MGMNT 
FIRE 
FIRE 
OEMC 
GENERALSERVICES 
WATER MGMNT 
STREETS & SAN 
FAMILY & SUPPORT 
POLICE 
FIRE 
STREETS & SAN 
PUBLIC LIBRARY 
POLICE 
FAMILY & SUPPORT 
WATER MGMNT 
BUSINESS AFFAIRS 
POLICE 
OEMC 
STREETS & SAN 
POLICE 
FIRE 
STREETS & SAN 
FIRE 
WATER MGMNT 
STREETS & SAN 
TRANSPORTN 
HEALTH 
FAMILY & SUPPORT 
Number 
90,744 
84,450 
84,450 
89,880 
106,836 
70,764 
41,850 
20,051 
49,452 
93,600 
100,320 
90,618 
2,756 
46,896 
73,920 
74,028 
81,588 
99,228 
81,588 
88,596 
108,228 
92,682 
111,474 
17,846 
93,600 
58,536 
41,850 
2,756 
87,384 
70,380 
72,862 
25,334 
46,668 
106,836 
106,836 
98,616 
70,380 
59,652 
Fo rmattinq 
N ew salary 
95,281 
88,673 
88,673 
94,374 
112,178 
74,302 
43,942 
21,054 
51,925 
98,280 
95,149 
2,894 
49,241 
77,616 
77,729 
85,667 
104,189 
85,667 
93,026 
113,639 
97,316 
117,048 
18,739 
98,280 
61,463 
43,942 
2,894 
91,753 
73,899 
76,506 
26,601 
49,001 
Charts 
Data Bars 
Color... 
112,178 
112,178 
103,547 
73,899 
62,635 
Totals 
Icon Set 
Tables 
Greater... 
Sparklines 
Clear... 
Conditional Formatting 
2,756 $ 
uses rules to highlight interesting data. 
2,894 

Visualizations like data bars make it very easy to visualize values in a range of cells. A longer bar represents a higher value.

*Click on a cell inside the table*

*Ctl+Q to bring up the quick analysis tools*

*Show Data Bars*

*Show icons*

Now I can access Sparklines from here and it will give me a quick and easy view of trends. But sparklines are really useful for civic data, which tends to show trends over time, so let’s take a deeper look:

**Sparklines**

[Note to presenter: you should change this spreadsheet to make it relevant to your local region]

Here I have a list of housing counseling centers in Chicago (which are real), and the number of clients they took in on each day of the week for a month (which is made up data).

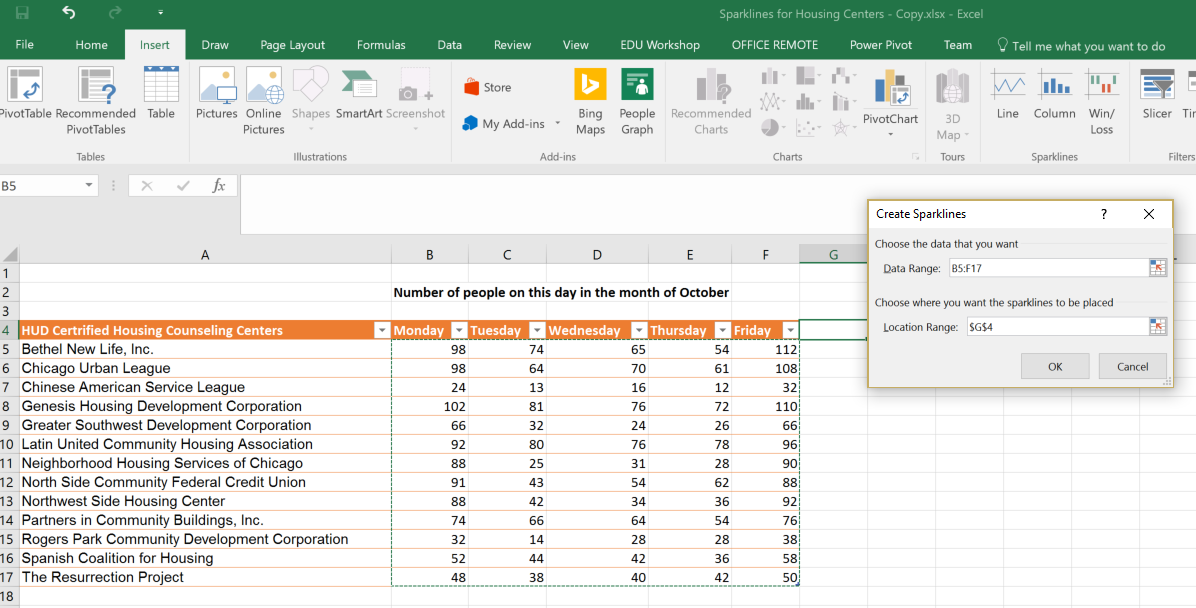
*Open the Sparklines for Housing Centers workbook*

*Click on the Insert tab on the ribbon*

I have a number of sparklines that I can use to show the trend. And I am going to choose the Line sparkline and apply it to the days of the week:

*Click on Line*

*Select the range of numbers for the data range*

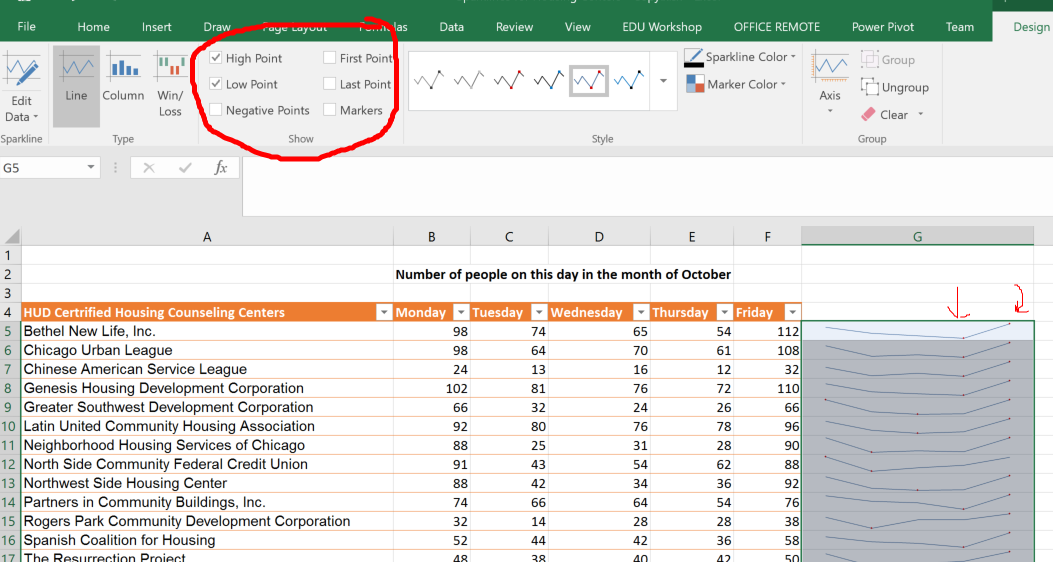


*And the corresponding area in column G for the location range (in this one, it is G5:G17)*

*Make the column with the sparklines wider*

First, you will notice that it put the sparklines right in the cells corresponding to the columns. So I can see that Mondays and Fridays are the big days for client visits. Next, you will notice that you are automatically on the Design tab (and if you are not, click on the Design tab on the ribbon). This will let you customize the sparklines inside the cell. For example, we can call out the highs and the lows*,*

*In the Show section, check High Point and Low Point*



Then, I can differentiate the markers for the high verses the low:

*Click on the Marker Color*

*Choose a separate color for the High Point and the Low Point*

Now, the default scale for these mini-charts is to scale across the numbers for each row. So let’s see what happens when we change one of the rows to have very small numbers.

*Change the numbers in row 5 (Bethel New Life in the example) to be single digit numbers like 1, 2, 3, 4,5.*

You can see that it doesn’t reflect the fact that these are relatively low numbers. You can change this so that it scales across the rows:

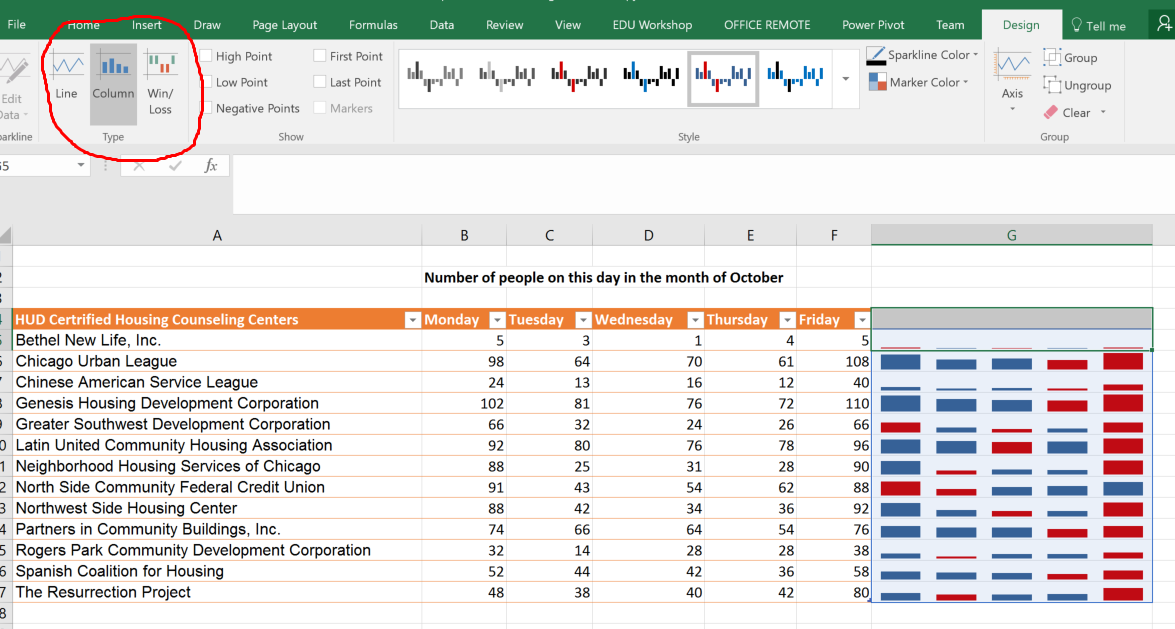
*Click on the Design tab*

*Click on the arrow under Axis*

*Change both the Vertical Axis Minimum and Maximum to be the Same for All Sparklines*

Now you can see that the row with lower numbers appears almost flat. Line graphs are only one type of sparklines. You can also choose the Column Graph and this will make the difference even more dramatic

*Click on Column in the Type section of the Design tab.*



Bar Charts, Line Charts, and Scatterplots

I am going to be showing you some charts and talking about when to use them. This is also a good place to take a look at some of the recommended charts. Recommended charts gives you some choices that Excel has picked out for you by looking at your data and looking at recommended chart types.

*Select the entire chart including headers.*

*Open recommended charts.*

*Choose Stacked Bar chart*

This is a good chart when you want to make comparisons of similar data across categories.

*Another good chart is the combo chart, which allows you to create a bar chart for one series and a line chart for another.*

*Select the Visitors and Temperature columns*

*Select Insert Chart*

*From the All Charts tab, choose Combo*

*Change Visitors to a Column or Clustered Column chart.*

*Make Temperature a Line Chart*

*Look at the suggested settings on the top and choose the second one (Clustered Column, Line on Secondary Axis).*

*Give the chart a title and move it out of the way.*

You can see that there seems to be a trend. To get a better view of the trend, we might want to look at a scatterplot with a trendline. Excel has many options for the trendline, including giving you the exact regression equation and R squared values.

*With the two columns selected, click Insert and click on the scatterplot button.*

*Choose the first option.*

*Click on the plus button and add a trendline*

*Right click on the trendline and you will see that it is set to linear.*

*Also add the equation by checking the Display Equation on chart button.*

*You can also choose to add the R2 . This is the % of the variability in the visits that can be explained by the temperature.*

*Close the Sparklines for Housing Centers workbook*

Hierarchy Charts

Let’s go back to our community area and neighborhood workbook. This time, I have added some data to it so that we can see how to visualize categorical data when there is a hierarchy and groupings. We will create both a Treemap and a Sunburst chart, both useful for categorized data where there are different levels of categories and sub-categories.

*Open Neighborhood Centers workbook*

Let’s say I work for an agency or a social services organization. I want to summarize the data of how many clients I serviced by community area and then by neighborhood. Since this is a hierarchy, I will use a hierarchy chart. Hierarchy charts are new to Excel 2016.

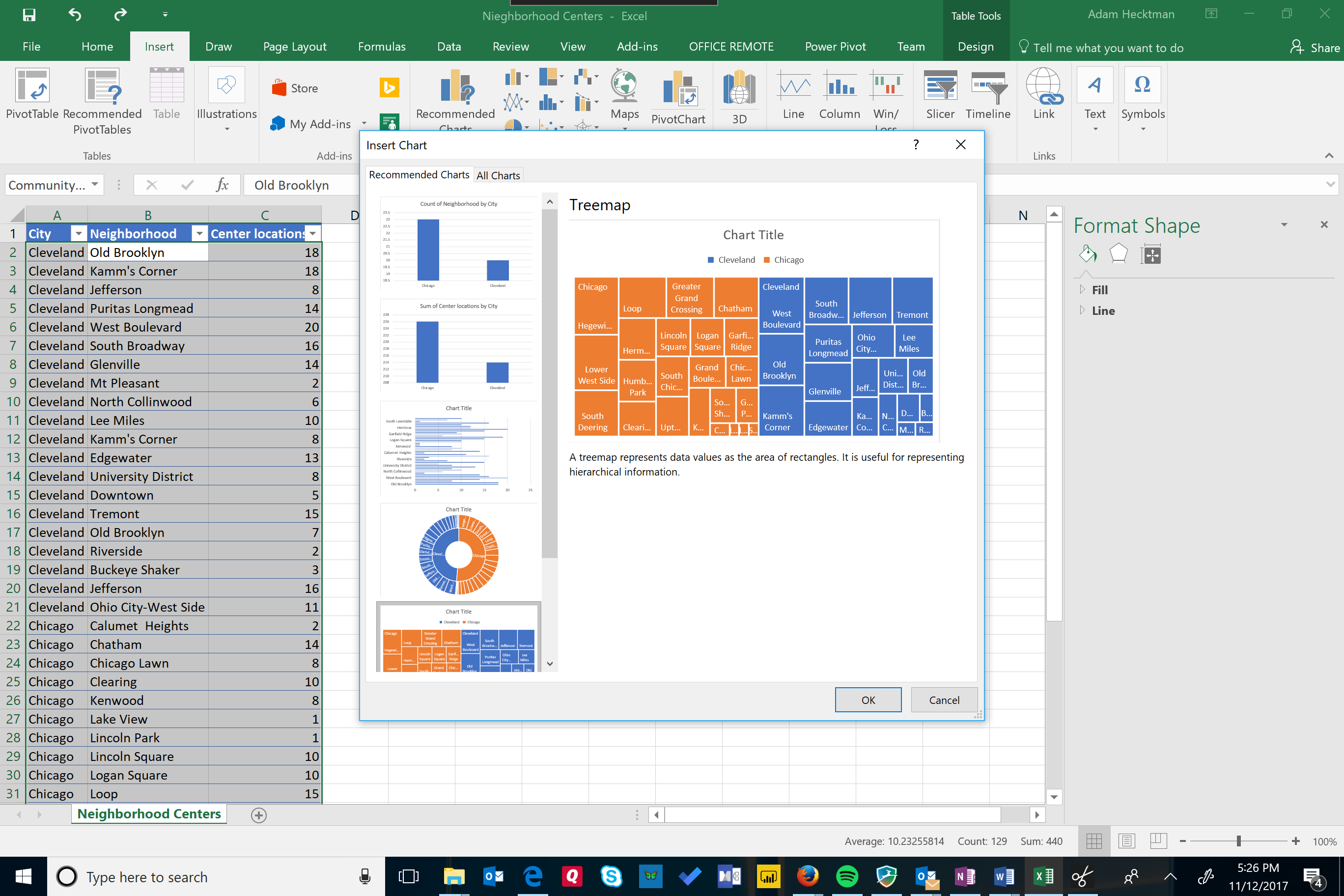
Now, to use a hierarchy chart, the data has to be in a regular table, not a Pivot Table. And it has to be structured like a hierarchy (i.e. within community areas, neighborhoods). Notice that I have data for some neighborhoods, none for others. That is ok. I will insert a Treemap to visualize how this data breaks down.

*Select some or all of the Community areas, neighborhoods, and the data.*

*You must also select the column headings*

*Click on Insert, and choose the Hierarchy Charts from the Charts section.*

*Choose a Treemap*



*Expand the Treemap so that you can easily see the individual blocks.*

Note that you will already be in the Design tab, and from here, you can experiment with different Treemap chart designs.

*Explore the different designs*

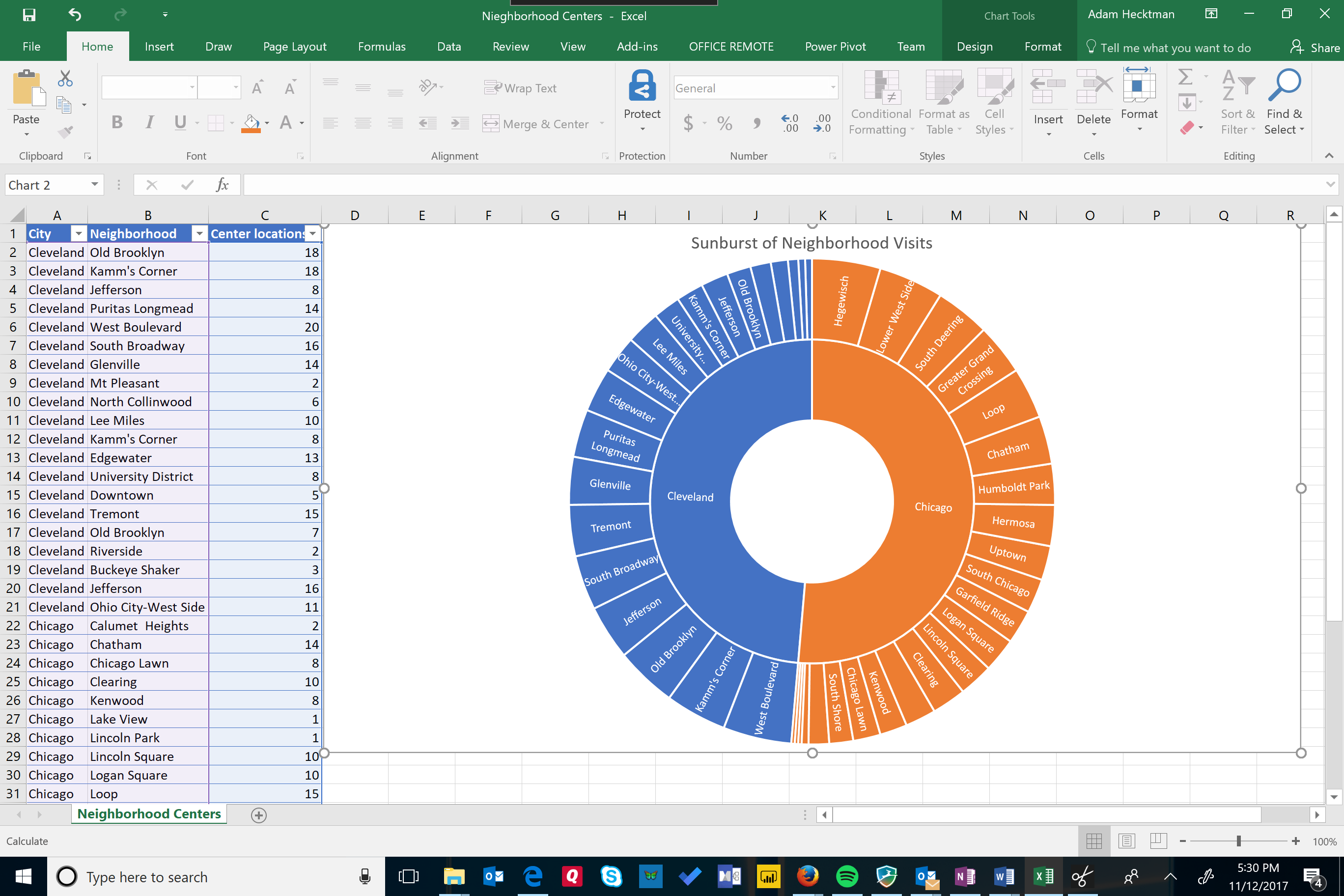
*Add a chart title.*

Treemaps are good when you want to compare relative sizes. For example, if you want to see what areas make up most of your work. A good portion of my clients live in Lake View and Auburn community areas. Look deeper at Lake View. Graceland West and Southeast Ravenswood make up a lot of my client locations in this community, while Boystown makes up a small number of my clients.

Another useful chart type for categorical data where there are sub-categories is called the Sunburst. We can create the sunburst without having to do anything to this chart beyond changing the type.

*Click on Change Chart Type*

*Click on Sunburst*



With the sunburst chart, each level of the hierarchy is represented by one ring with the innermost ring as the top of the hierarchy. A sunburst chart is similar to a doughnut chart. However, a sunburst chart with multiple levels of categories shows how the outer rings relate to the inner rings. The sunburst chart is most effective at showing how one ring is broken into its contributing pieces.

*Close the workbook*

Mapping Data

**Note for instructor**: For this one, if you are not using my sample data sets, you should have a data set already downloaded. I used the Chicago 311 Graffiti Requests dataset.

It needs to have a column that connotes time (such as creation date) and location columns (such as Latitude and Longitudes)

*Open the data set.*

Now here is a tip. If you are offered up an option to download it in XLS or XLSX format, either skip it or test it. It seems like you are saving time, but data sometimes gets truncated, depending on the portal you are using.

After you download it, save it as an Excel workbook right away - you will have access to many more features that CSV alone can provide.

You should also Format it as a Table right away for readability. This is simple to do, as you have a button that does just that right on the home tab.

Machine generated alternative text:
Formatted - Excel 
Normal 
Check Cell 
Neutral 
Linked Cell 
Calculation 
Note 
Conditional Format as 
Formatting • Table • 
Light 
Medium 
Bad 
Explanatory 
Input 
- 
- 
Dark 
New Table Style... 
New PivotTable Style... 
- 
- 

Now, not only is it easier to read, you also get those table features like filtering and sorting.

*This is a good place to Show filters and sort options, like:*

* + *Filtering multiple values*
  + *Searching in the filter search box for values to filter*
  + *Sorting by various criteria*
  + *Etc.*

Tip: You should give your tables names. Excel gives generic names (Table1, Table2, etc.). When you do the format as table, you will already be on the Table Tools Design tab.

*On the left, click in the Table Name box and replace Table1 with Graffiti (or whatever your data set is)*

Machine generated alternative text:
Removal.xlsx - Excel 
Draw 
Page Layout 
Table Name 
Table 1 
•3 Resize Table 
Creation Date— Status 
FormuLæ 
nsert 
Slicer 
Properties 
Open in Browser 
Ref resh 
Unlink 
Zble Data 
Power Pivot 
HeaderRo..v 
I Row 
Banded Rows 
Team 
First Column 
Last Column 
nded Columns 
Style Options 
Tell me what you want to do 
Filter Button 
Summarize with PivotTable 
Remove Duplicates 
Convert to Range 
Export 
Zble Styles 
2746 W 50TH ST 
Completion Date— Service Request Numbe— Type of Service Reques— What Type of Surface is the Graffiti on — Where is the Graffiti locate* Street Address— ZIP Code— X Coordinate' Y Coordinate— Ward 
Police Distric— Community Area— SSA 
Latitude. Longitude. 
3 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 completed 
3/3/2016 16-01292850 
3/3/2016 16-01290692 
3/3/2016 16-01293583 
3/3/2016 16-01290517 
3/3/2016 16-01290723 
3/3/2016 16-01290510 
3/3/2016 16-01290515 
3/3/2016 16-01287321 
3/3/2016 16-01290677 
3/3/2016 16-01290543 
3/3/2016 16-01290534 
3/3/2016 16-01290525 
3/3/2016 16-01290547 
Graffiti Remova I 
Graffiti Remova I 
Graffiti Removal 
Graffiti Remova I 
Graffiti Removal 
Graffiti Remova I 
Graffiti Removal 
Graffiti Remova I 
Graffiti Removal 
Graffiti Remova I 
Graffiti Removal 
Graffiti Remova I 
Graffiti Remova I 
Metal- Painted 
Brick - unpainted 
Other / unknown Surface 
Metal - unpainted 
Brick - Painted 
Brick - Painted 
Brick - unpainted 
Wood - unpainted 
Brick - Painted 
Brick - Painted 
Metal - Painted 
Metal 
unpainted 
Garbage Cart 
Front 
Mail Box 
847 W BELMONT 
4401 W DIVISION 
5819 W FULLERTC 
1640 W 34TH ST 
5306 S PULASKI R 
2155 W 21ST PL 
2059 W 19TH ST 
2751 W BLOOMI,' 
1132 N MONITOF 
1016 W CULLERT( 
1165 W 18TH ST 
2100 W 19TH ST 
1636 W 34TH ST 
60651 
60639 
60632 
60608 
60647 
60651 
1169936.596 
1146814.312 
1137104.706 
1165855.68 
1150602.188 
1162330.344 
1162942.02 
1157652.986 
1137151.089 
169815.28 
168890.38 
1162672.82 
1165874.98 
1921450.184 
1907631.165 
1915356.767 
1882192.302 
1869112.07 
1889714.362 
1890732.481 
1911864.17 
1906949.595 
1890583.17 
1891562.497 
1890726.081 
1882192.702 
17 41.93985 
41.9024 
41.92378 
41.83243 
41.79659 
41.85292 
41.85569 
41.9138 
41.90122 
41.85537 
41.85785 
41.85592 
41.83243 
-87.651105 
-87.736264 
-87.771732 
-87.667232 
-87.723404 
-87.680488 
-87.678275 
-87.696717 
-87.771842 
-87.652191 
-87.655761 
-87.678512 
-87.667075 

Now we are going to create a 3D Map to show where graffiti has been reported over time. 3D Maps allow you to take geographical and temporal data and turn them into maps, and even create videos to share and post those maps.

*Click 3D Map from the Insert tab*

First time you do this with your data, it will say on the bottom "loading model" and this may take a while, but it only happens once. When it says loading model, it is referring to a data model.

A Data Model is an approach for integrating data from multiple tables. Effectively, what you are doing is building a relational database inside an Excel workbook, without having to have database expertise. Within Excel, Data Models are used pretty much without you having to know anything about them. They are used in Maps, PivotTables, Pivot Charts, and Power View reports, etc.

But today, we are just going to build a map.

Machine generated alternative text:
Text 
Insert 
Legend 
Time Date & 
Line Time 
Time 
our 
Editor 
Layer 
Pane 
View 
Field 
Field List 
Drag fields to the Layer Pane. 
Completion Date 
Creation Date 
Latitude 
Longitude 
Police District 
Service Request Number 
SSA 
Status 
Street Address 
Type of Seo/ice Request 
What Type of Surface is the Graffiti o... 
Where is the Graffiti located? 
Add Layer 
Layer 1 
Location 
Add Field 
Height 
Add Field 
Category 
Add Field 
Time 
Add Field 
Filters 
Layer Options 

First, let's talk about some terminology. On the right, you will see options for the layer. You create layers that show specific data on your map (like the number of graffiti requests). You can combine layers on top of layers, and then build scenes from those layers. Scenes are like scenes in a movie, with transition effects and all. Then, you can combine scenes into a tour. We are going to do all of that here.

**Layers**

Your first decision you need to make is which fields in your table you want to use to denote location. This is an important question to consider. It relates back to the question you want your analysis to answer. For example, if you need precise locations to answer your question, longitude and latitude (lat/long) make good options. However, let's say you want to do some analysis at the block level in a neighborhood, or at the zip code level. Lat/long will be too granular.

In this case, we are going to use Lat/Long.

*In the Location box of the Layer options, choose +Add Field, and add Latitude and Longitude*

Machine generated alternative text:
Zield List 
Drag fields to the Layer Pane. 
Completion Date 
Creation Date 
Latitude 
Longitude 
Police District 
Service Request Number 
SSA 
Status 
Street Address 
Type of Seo/ice Request 
What Type of Surface is the Graffiti o... 
Add Layer 
Layer 1 
Location 
O 
Latitude 
O 
GDngitude 
Add Field 
Height 
Add Field 
Select One.„ 
Select One.„ 

Now these are just field names from your data set. You have to tell the map what they refer to.

*Where it says "select one", choose latitude and longitude respectively. If there are other fields that automatically came up in your location box, you can delete them by clicking on the X on the right.*

Right above the location box, you will see that you have options for visualizing your data: you can choose stacked column, clustered column, bubble charts, heat map, and shaded regions. (For many, many more visualizations for your data, you can also use Power BI, a tool for working with data that does not require Excel) We will stick with the default, a stacked column chart.

Next, we need to determine which field will represent the height of our column. Let's say that we want to know how many graffiti removal requests were made for each location. You should look for a field that is unique to every request. Most open government datasets have one. In this case, each request is assigned a Service Request Number, so counting the number of Service Request Numbers for each location is a good option

*Under Height, click + Add Field and choose Service Request Number. It will default to Count .*

Now, it is one thing to see how many requests you have in a location, but you have an opportunity to show more data. For example, you can have each bar (column) on the graph broken down into different colors to represent different things. In this case, let's use the colors to represent where the graffiti is located (or what type of surface it is on)

*Under Category, click + Add Field and choose Where is the graffiti located.*

Now, before we go and play with the map, let's do one more housekeeping item. This layer is called Layer 1. Rename it to something meaningful.

Machine generated alternative text:


*Click on the pencil and rename the layer to "Graffiti by Location"*

Notice that this also creates a legend. If you re-size it, you can see all of the colors of the bars and what they represent.

Take a few minutes to explore this map.

*Zoom in using the mouse wheel or using the + and - buttons*

*Tilt up and down using the arrows*

*Generally explore the map with the audience*

Now I can zoom around, tilt, etc. but since there is just so much data, I find it hard to know where I am. And when using civic data, there are a couple of important things for your users to understand: location and specific data points. Two things will help:

*First, Click on the Layer Options and reduce the thickness of the column to something around 50% (you can see it change in real time so you will know what makes sense for your visualization).*

Machine generated alternative text:
x 
.47- 
Express Way Job 
Fence 
Front 
Garage 
Garbage Cart 
Hydrant 
Mail Box 
Newspaper Box 
Overpass 
Phone 
Pole 
Rear 
Side 
Sign 
Traffic Control Box 
Viaduct 
Add Field 
Height 
Service Request Number (Count - Not...' 
Add Field 
Category 
Where is the Graffiti located? 
Time 
Add Field 
Layer Options 
Height 
Thic kness 
Guk current scale 
Opacity 
Show Values 
Zeroes 
Negatives 
Nulls 
Data Card 
Custom ize 

*On the toolbar, click on Map Labels. This will go to Bing Maps and pull in the labels and put them on your map. It is important to do this for viewers who may not know your city as well as you do.*

**Scenes**

Now, one reason why Excel is such a great tool for mapping data is that it is both spatial and temporal. Let's add a time dimension.

*In the Layer Pane, choose Time and + Add Field. Choose creation date. If there is a lot of data, this may take a bit. You will see the status bar on the bottom left say "Processing…"*

Machine generated alternative text:
Add Layer 
Graffiti By Location 
Location 
@ Latitude 
O 
GDngitude 
Add Field 
Height 
Latitude 
Longitude 
Service Request Number (Count - 
Add Field 
Category 
Where is the Graffiti located? 
Time 
Add Field 
Tablel 
Completion Date 
Creation Date 
Not...' 
x 

Once you add a time dimension, you will see something new: a media strip.

Machine generated alternative text:
ell 

*If you click the "play" button, you will see this data grow over time.*

But it is not very interesting. For one thing, you don't really need to know the requests every minute of every day.

*Right click on the time box on the upper left and change the format to show just days, or months depending on your data set.*

*Next click on Scene Options and experiment with the various transition effects*.

*Click on the Play Tour button from the toolbar.*

For datasets that cover a large geography, it is useful to have Excel move over the entire map using these transitions.

Now you can create scenes with this mapping data just like you would be creating a movie.

You can build layers on top of layers

*Click Add Layer to the Layer pane and give this new layer a name*

*Hide the first layer by clicking on the eye so that you can focus on this new layer*

*Use a heat map for this layer*

*Assign the location to Zip Code (or, if your dataset does not have Zip, use some other geography larger in scope than lat/long) and map it to Postal Code*

*Choose Service Request Number (Count) for the Value*

*Choose Creation Date for the time*

*Experiment with Layer Options, showing the impact directly on the map*

Machine generated alternative text:
Add Layer 
Graffiti By Zip 
Location 
@ ZIP Code 
Add Field 
Value 
Postal Code 
Graffiti By Zip 
Service Request Number (Count - Not Blank) 
or Less 
Not...' 
80,686m 
or More 
Lyo s 
Forest View - 
Service Request Number (Count - 
Add Field 
Time 
Creation Date (None) 
Filters 
Layer Options 
Color Scale 
Radius of Influence 
Guk current scale 
Opacity 
Visual Aggregation 
Default 
Colors 
Show Values 
Zeroes 
Negatives 
Nulls 
Data Card 
Custom ize 
C h i c a 90 
Midway 
Airport 
earing 

Then you can build scenes by hiding and exposing those layers

*Create a scene that shows the heat map alone by Hiding Graffiti by Location and give the scene a meaningful name*

Machine generated alternative text:
Play 
Tour • 
Create Capture 
Video Screen 
Tour 
New 
Scene • 
Themes 
Scene 
cene 
Refresh 
Shapes 
Layer 
Scene Options 
Map 
Tour 1 
Requests by Zip 
o (10 sec) 
Scene duration (sec) 
Scene Name 
Requests by Zip 
Effects 
Transition duration (sec) 
Flat Find Custom 
Map Location Regions 
Map 
10.00 
3m 
Effect 
Effect Speed 
Time 
Start date 
End date 
Speed 
Map Type 
Change 
Circle 
5/26/2004 AM 
3/3/2016 AM 

*Create a second scene by clicking on the New Scene button at the toolbar and chose to make a copy of the first*

*Change this scene so that it hides the Heat Map and shows the Requests by Location*

|  |  |
| --- | --- |
| Machine generated alternative text: Tour  Tour 1  Requests by Zip  2  Requests by Locati  Scene  0(10 sec)  Layer  Scene Options  Scene duration (sec)  Scene Name  Requests by Location  Effects  Transition duration (sec)  Ions  Map  10.00  3m  Circle  Effect  Effect Speed  Time  Start date  End date  Speed  Map Type  Change  5/26/2004 AM  3/3/2016 AM | Machine generated alternative text: Add Layer  > Graffiti By Location |

*If you like, create a third scene with the two together.*

**Tours**

Now, you have a tour made up of all the scenes you created.

*Click Play the tour*

And from there, you can turn this into a video that can be shared and posted

*Click create video*

*If you have music you can use, click Use Soundtrack and add it*

*If you decide to actually go ahead and make the movie, choose a low res option so that it doesn't take a long time*

Transforming data

*Open Address.xlsx or your own dataset that contains addresses*

This section is about how to use Excel's functions and tools to perform transformations on your data that are often required when working with open and civic data. For example, you often will get names in a single field, when you really just want last name. Or you will get full addresses, when you really need to analyze your data at the block level.

Let's start with breaking up addresses. You will notice that we have the full address in one column in the Addresses.xlsx workbook. But what if we either want to break it up so that the street address is in one column, the direction in another, the street in another, the suffix in another.

Or - what if instead of looking at this data at the address level, you want it at the block level. How much crime took place on the 1600 block of Halsted in 2014, for example. Excel has a built-in Text to Columns feature that makes this very easy.

*First, make some space by creating 5 columns next to the address. You can name them Block, Address, Direction, Street, and Suffix. The "Block" column will contain the block-level address (e.g. the 1600 Block vs. the 1612 Address).*

*From the Data tab on the toolbar, choose Text to Columns*

This is a very versatile feature. It is useful for not only transforming this kind of data, but taking data that may come in a non-delimited form and turning it into a delimited spreadsheet.

*In this case, we will choose Delimited and next.*

*Now we tell Excel what character is the delimiter. Notice that we can preview what the data will look like when we are finished.*

*Check "Space" and click Next.*

Machine generated alternative text:
suffix 
Address2 
3825 W 63RD ST 
1351 W ADDISON ST 
2222 ARCHER AVE 
City State 
CHICAGO 
CHICAGO 
CHICAGO 
60629 
60613 
60616 
Convert Text to Columns Wizard Step 2 of 3 
This screen lets you set the delimiters your data contains. You can see hcn,v your text is affected 
in the preview below. 
Delimiters 
z Tab 
Semicolon 
Comma 
Z] Space 
Other: 
Data preview 
Treat consecutive delimiters as one 
Text qualifier: 
Finish 

You can even tell it how the new columns should be formatted when you are finished

*For destination, either select columns that you want to put the data into with the Wizard collapsed, or simply type in the name of the range. We want to start breaking things up and placing them starting with Address, so it will look like this:*

Machine generated alternative text:
to 
Columns 
Address 
Reapply 
Advan ced 
Street 
Flash 
suffix 
Remove 
Duplicates 
Consolidate 
Data 
Relationsh ips Manage 
Data Model 
City State 
What-If 
Analysis • 
Forecast 
Sheet 
Sort & Filter 
Direction 
Validation • 
Data Tools 
Address2 
13825 W 63RD ST 
11351 W ADDISON ST 
12222 ARCHER AVE 
Convert Text to Columns Wizard - Step of 3 
{4635 W 63RD ST 
14759 W FULLERTON AVE 
14820 W Walton st. 
17242 W TOUHY AVE 
1159 N WABASH AVE 
15600 W FULLERTON AVE 
111718 S WESTERN AVE 
5010 W CHICAGO AVE 
14359 N PULASKI RD 
CHICAGO 
CHICAGO 
CHICAGO 
chicago 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
Forecast 
Zip Inspec 
60629 
60613 
60616 
628 
60629 
60639 
60651 
60631 
60601 
60639 
60643 
60651 
60641 

*Click Finish and it should look like this:*

Machine generated alternative text:
Block 
Address 
Direction 
3825 w 
1351 w 
2222 ARCHER 
11147 s 
46 E 
11334 s 
4635 w 
4759 w 
4820 w 
7242 w 
159 N 
5600 w 
11718 s 
5010 w 
4359 N 
Street 
63RD 
ADDISON 
AVE 
MICHIGAN 
CHICAGO 
MICHIGAN 
63RD 
FULLERTON 
Walton 
TOUHY 
WABASH 
FULLERTON 
WESTERN 
CHICAGO 
PULASKI 
suffix 
AVE 
AVE 
AVE 
AVE 
AVE 
AVE 
AVE 
AVE 
AVE 
Address2 
3825 W 63RD ST 
1351 W ADDISON ST 
2222 ARCHER AVE 
11147 S MICHIGAN AVE 
46 E CHICAGO AVE 
11334 S MICHIGAN AVE 
4635 W 63RD ST 
4759 W FULLERTON AVE 
4820 W Walton st. 
7242 W TOUHY AVE 
159 N WABASH AVE 
5600 W FULLERTON AVE 
11718 S WESTERN AVE 
5010 W CHICAGO AVE 
4359 N PULASKI RD 
City State 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
chicago 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
CHICAGO 
Zip 

Now, let's take that address and round it off to block level data so that we can put it in the Block column before recombining our address.

The Block column will contain the rounded off address number in the Address column. Excel has a number of rounding functions. An easy way to find the function you are looking for is to:

*Simply click on the function button and then type in plain language what it is you are looking for.*

*For example, type Round in the search for function, then G*o

Machine generated alternative text:
Connections 
ATI 
Get & Transform 
Insert Function 
Search for a function: 
ypeab 
lick Gol 
Or selecta category: Most Recently Used 
Select a function: 
CT 
MID 
YEAR 
SUM 
AVERAGE 
RO U ND DOWN (num 
Rounds a number down, tm,vard zero. 
Helo on this function 
U8MAN YOUTH MINISTRI 
Daycare (2- 6 Years) Risk 1 (High) 

It is important to always remember the question you ultimately want answered every step of the way. For example, if we used the Round function here, an address like 1758 would be rounded up to 1800. But really, the address 1758 is in the 1700 block of the street. So you need to select the ROUNDDOWN function.

*Select ROUNDDOWN.*

ROUNDDOWN only takes two arguments. The first is the number you want to round down. You can simply select the first address right on the spreadsheet, no need to type in the row/column:

Machine generated alternative text:
Block 
[@Address]) 
Address 
Direction 
3825 w 
1351 w 
Street 
63RD 
ADDISON 
2222 AR 
11147 s 
46 E 
11334 s 
4635 w 
4759 w 
4820 w 
7242 w 
159 
N 
5600 w 
Function Arguments 
ROUND 
Num ber 
Num digits 
suffix 
[@Address] 
Address2 
3825 W 63RD ST 
1351 W ADDISON ST 
3825 
number 
Rounds a number to a specified number of digits. 
Number is the number you want to round. 
Formula result = 
Helo on this function 
CHICAGO IL 
CHICAGO IL 
Cancel 

The 2nd argument is the number of digits you want to round the number to. This is the number of decimal places to the right or left of the decimal. If you are rounding to the right, use a positive number. In this case, we are rounding to the left (remember, 1758 to 1700).

*Select -2 for two decimal places to the left.*

*When you are done, the function will look like this: =RoundDown (B1,-2)*

*Notice that Excel predicted that you would want the entire column this way and executed a Flash Fill.*

Machine generated alternative text:
Add-ins 
Connections 
Properties 
Edit Links 
Connections 
h) 
h) 
h) 
h) 
h) 
h) 
h) 
h) 
h) 
h) 
h) 
h) 
h) 
dium) 
Power Pivot 
Address 
2200 
11100 
11300 
7200 
11700 
Tea m 
Filter 
Design 
Clear 
Reapply 
Adva nced 
Sort & Filter 
Direction 
3825 w 
1351 w 
2222 ARCHER 
11147 s 
46 E 
11334 s 
4635 w 
4759 w 
4820 w 
7242 w 
159 N 
5600 w 
11718 s 
5010 w 
4359 N 
Stree 
63RD 
ADDu 
AVE 
MICH 
CHIC' 
MICH 
63RD 
Wa Itc 
TOUP 
CHIC' 

*If you like, you can delete the address column.*

Now, it is time to bring the address together with the block-level data.

*Insert a column to the left of block. Call it Block Address*

We will use another function, Concatenate, to bring all of these columns (Block, Direction, Street, and Suffix) together, with a space in between each.

*Put the cursor in the first empty row of Block Address (row 2)*

*Start typing in the formula bar =CONCATENATE*

*Notice that Autocomplete starts suggesting concatenate, and when you double click on it, you get a guide for how to use the function*

Machine generated alternative text:
Insert 
Draw 
Page Layout 
Formulas 
Add-ins 
Connections 
Properties 
Refresh 
Connections 
Power Pivot 
From 
From Other Existing 
Sources • Connections 
From Table 
Query • 
Recent Sources 
Get & r 
Get External Data 
DBA Name 
-CONCATENATE( 
Facility Typ 
rant 
Restaurant 
Restaurant 
Restaurant 
Restaurant 
Restaurant 
Restaurant 
Risk 
Risk 1 (High) 
Risk 1 (High) 
Risk 1 (High) 
Risk 1 (High) 
Risk 1 (High) 
Risk 1 (High) 
Risk 1 (High) 
Block Address Bloc 
-CONCATENATE( 
58GORDlTAS SANTIAGO 
5 D'AGOSTINOS PIZZA 
3 DOLO GARDEN RESTAURANT & BAR 
27 THE NEW RANCH STEAK HOUSE 
2 FLACO'S TACOS 
47 M & R FOODS 
KING'S CHOP SUEY 
Table 
Team 
Sort & Filter 
Addres: 
2200 
11100 
11300 

The function format is very easy: = CONCATENATE(text, text, text). And that text can be a selected cell or column. For example:

*After the "(", click on the first Block address (F2 in the screenshot).*

*Note that @Block shows up. Then type ," " in order to insert a space*

*Click on Direction (H2) followed by ," ",*

*Click on Street (I2) followed by ," " ,*

*Click on Suffix and hit enter*

The full formula will be **=CONCATENATE([@Block]," ",[@Direction]," ",[@Street]," ", [@suffix])**

Notice a couple of interesting things:

* + Flash Fill has copied this formula all the way down for you
  + You now have a column with the block address as opposed to a specific location!

Many civic projects use block level data and may compare that to neighborhood data, city data, regional data, census data, etc. And, of course, it is easy to map this data out as well.

Advanced Analytics: Using Data Analysis Tools

The vast majority of time, Excel as it provides you with what you will need to access, combine, query, massage, transform, visualize, and otherwise play with your data.

Even for most statistics you might need, you have a lot of functions in Excel. For example, let’s look at a simple correlation.

There are, however, advanced tools for statistical analysis if you turn on the add-in. You do not need to download anything, you just need to turn it on. Doing so will give you tools for

* + Regression testing
  + Histograms
  + Anova, correlation, and covariance
  + Exponential smoothing
  + f-tests, z-tests, and t-tests
  + Fourier analysis
  + Moving averages, rank and percentile summaries
  + Sampling and a more sophisticated random number generator than the standard function

**Enabling the Add-in**

Going through all of these is beyond the scope of this training, so I will just give you a basic idea of how it can be used. To demonstrate how this data analysis tool pack can be used, I created a simple spreadsheet with a million rows, each row simulating the toss of a fair die using the function =RANDBETWEEN(1,6), which gives you a random number between 1 and 6 inclusive. You can either download this (A Million Rolls Of A Die.xlsx) or simply recreate it.

*In your spreadsheet, click File, Options, and choose Add-ins.*

*If you do not see the Analysis Tool Pack already in the list of add-ins, choose Excel-Add-ins from the Manage drop down in the bottom and hit Go*

Machine generated alternative text:
Excel Options 
General 
Formulas 
Proofing 
Language 
Advanced 
Customize Ribbon 
Quick Access Toolbar 
Add- ins 
Trust Center 
View and manage Microsoft Office Add-ins. 
Name 
Active Application Add-ins 
Analysis ToolPak 
Microsoft Power Map for Excel 
Microsoft Power Pivot for Excel 
Microsoft Power View for Excel 
Mindjet MindManager Add-In 
Team Foundation Add-in 
Inactive Application Add-ins 
Analysis ToolPak - VBA 
Date (XML) 
Euro Currency Tools 
Inquire 
Microsoft Actions Pane 3 
Solver Add-in 
Document Related Add-ins 
No Document Related Add-ins 
Disabled Application Add-ins 
No Disabled Application Add-ins 
Location 
Files\Microsoft 
Files\Microsoft Map Excel Add-inlEXCELPLUGINSHELL.DLL 
Files\Microsoft Excel Add-in\PowerPlvotExceICIientAddIn.dII 
Files\Microsoft View Excel Add-in\AdHocReportingExceICIient.dII 
16\Mm16ExceILinker.dII 
Files\Common Files\Microsoft Shared\Team Foundation 
Files\Microsoft 
Files\Common Files\Microsoft Shared\Smart TaglMOFL.DLL 
Files\Microsoft 
Files 
Files\Microsoft 
Excel Add- in 
COM Add-In 
COM Add-in 
COM Add-in 
COM Add-in 
COM Add-in 
Excel Add-in 
Action 
Excel Add-in 
COM Add-in 
XML Expansion Pack 
Excel Add-in 
Add-in: 
Publisher: 
ocation: 
Description: 
Analysis ToolPak 
Microsoft Corporation 
rmation available 
Files\Microsoft 
Provides data analysis tools for statistical an 
eering analysis 
Menage: Excel Add- ins 

*You will see a list of available add-ins. Check Analysis Tool Pack and click OK.*

*Now, when you click on the Data tab, you will see the Analysis Tool Pack.*

Now let's start doing some analysis on these million roles of a die. First, It’s a good idea to add bin numbers to your spreadsheet because they may be very useful for your analysis. Bin numbers represent intervals that you want to analyze. For example, you might want to create a histogram for how many 1's you role, how many 2's, etc. The 1, 2, 3,4, 5, and 6 are the bins in this case. If you don't create bin numbers, tools like the Histogram tool will create evenly distributed bin intervals by using the minimum and maximum values in the input range as start and end points.

**Histogram**

This tool generates data for the number of occurrences of a value in a data set. For example, in a city of 400,000 students, you can determine the distribution of scores in letter-grade categories. A histogram table presents the letter-grade boundaries and the number of scores between the lowest bound and the current bound. The single most-frequent score is the mode of the data. Back to the rolls of the die:

*Somewhere in the spreadsheet, enter in separate rows on a column 1, 2, 3, 4, 5, 6. If you are using my sample spreadsheet, it is already there.*

Machine generated alternative text:
ΙΙΙ ΙΙΙ ΙΙΙΙΙΙΙΙΙΙΙΙΙ: 
ΙΙ ΙΙΙΙΙΙΙΙΙΙΙΙΙΙΙΙΕ 

*Now, go to the data tab, and click on Data Analysis Tools.*

*We will start with a simple histogram. Click Histogram and the following dialog box will appear:*

Machine generated alternative text:
Itnnnnnnnnnnnnł 

*Select column A for the range*

*Select your Bins for the Bin Range (including the label, assuming you have Labels checked)*

*Make sure you check the Labels checkbox (otherwise it will think your label is non-numeric data)*

*Keep the New Worksheet Ply as your output option (you can give it a name)*

*Click ok. That is all there is to it and it quickly analyzed a million rows and gave you the histogram*

And that is it. You now have a histogram. And given your large population size, it looks about like what you would expect:

Machine generated alternative text:
Bins 
4 
6 
More 
Frequency 
174599 
175019 
174558 
174817 
174853 
174729 
Histogram 
200000 
150000 
100000 
50000 
6 
• Frequency 
More 

**Sampling**

The Data Analysis tools make it super simple to get samples from a population. Simply go to the spreadsheet that has your population data and:

*Choose Data Analysis Tools from the Data tab*

*Choose Sampling.*

*Give it the input range (if you are using the die spreadsheet, make sure you check Labels).*

*Tell it how big your sample size should be in Number of Samples*

*Tell it where you want it to show the output*

Machine generated alternative text:
ΙΙΙΙΙΙΙΙΙΙΕ 
ΙΙΙΙΙΙΙΙΙΙΙΙΙΙΙΙΙ! 
ΙΙΙΙΙΙΙΙΙΙΙΙΙΙΙΙΙ: 
ΟΙ Z Z 
ο οοο 
ΙΧΙΙΙΙΙΙΙΙΙΙΙΕ 

**Descriptive Statistics**

These are the basic statistics you would generally want to see if you are doing more sophisticated analysis of data. For example, if you are looking for information about the central tendency and variability of your data, this lets you do it in a few clicks.

*Choose Data Analysis Tools from the Data tab*

*Choose Descriptive Statistics.*

*Give it the input range (if you are using the die spreadsheet, make sure you check Labels).*

*For this example, have it group the data by column*

*Output the data to a New Worksheet Ply and give it a name*

*Check Summary Statistics*

*Check Confidence Level for Mean (you can keep it at 95%, which is standard)*

Machine generated alternative text:
Die Rolls 
41 
41 
31 
61 
31 
11 
41 
41 
61 
61 
31 
61 
51 
41 
3, 
31 
21 
Descriptive Statistics 
Input 
Input Range: 
Grouped By: 
Z] Labels in first rcn,v 
Output options 
C) Output Range: 
@ New Worksheet Ely: 
C) New Workbook 
Z] Summary statistics 
SA:SA 
@ Columns 
C) Rows 
Descriptive Stats 
Cancel 
Help 
Z] Confidence Level for Mean: 
Kth Largest: 
Kth Smallest: 

And now you will have a report with all of the basic descriptive statistics, without having to look up tables or use complex formulae:

Machine generated alternative text:
Die Rolls 
Mean 
Standard Error 
Median 
Mode 
Standard Deviation 
Sample Variance 
Kurtosis 
Skewness 
Range 
Minimum 
Maximum 
Sum 
Count 
Confidence 
3.501661 
0.001667 
1.707169 
2.914426 
-1.26658 
-0.00178 
3671754 
1048575 
0.003268 