



Excel Tools: Summarizing Data

Instructor: Richard Dunks

Teaching Assistant: Mark Yarish

Datapolitan****

Data Solutions for the Modern Metropolis

Goals for the Course

- Provide hands-on experience using Excel to clean and summarize data, including useful tips and tricks to working with city data
- Cover best practices when analyzing and visualizing data
- Introduce advanced functionality within Excel as it relates to summarizing data

Key Takeaways for the Course

- You will be more proficient using Excel for cleaning, analyzing, and visualizing data
- You will have a better understanding of the analytics process
- You will be familiar with Excel functions and other advanced features of Excel for analyzing data
- You will be familiar with fundamental best practices for visualizing data in Excel

Goals for this Morning

- Explore the use of Excel for performing basic descriptive analysis
- Introduce functions and pivot tables for data analysis
- Introduce basic chart creation in Excel to communicate data

Benefits of Excel

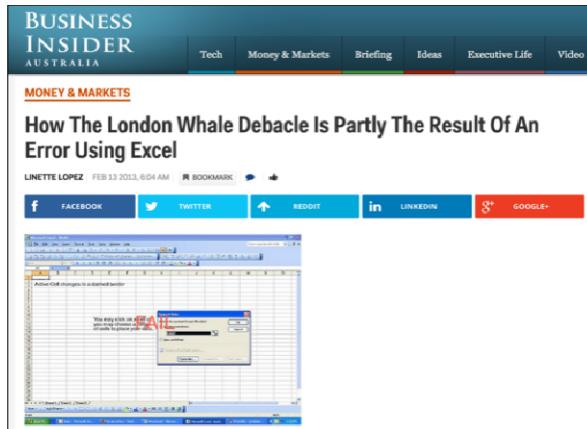
- Easy to use
- Very visual
- Lots of features and functions
- Easy to make charts
- Does a lot of formatting for you

Drawbacks of Excel

- Not very intuitive
- Hard to find what you're looking for (and they keep moving things around)
- Lots of features and functions
- Easy to make (bad) charts
- Does a lot of formatting for you

Cautionary Tale – London Whale

- \$6.2 billion lost by JP Morgan Chase & Co



<http://www.businessinsider.com.au/excel-partly-to-blame-for-trading-loss-2013-2>

Cautionary Tale – London Whale

- Caused largely by Excel mistakes:
 - Manual data errors
 - Manual copy and paste
 - Simple formula error that hid volatility
- Fined over \$1 billion for poor internal oversight of trading activities

MarketWatch    

 NEWS VIEWER MARKETS INVESTING TRADING DECK PERSONAL FINANCE RETIREMENT ECONOMY REAL ESTATE

Home > Personal Finance

88% of spreadsheets have errors

Published: Apr 20, 2013 7:39 a.m. ET

    139

Aa 

Research: Bad math rampant in family budgets and Harvard studies



By **JEREMY OLSAN**
EDITOR
 

Microsoft Excel makes it easy for anyone to do the kind of number crunching once reserved for accountants and statisticians. But the world's best-selling spreadsheet software has also contributed to the proliferation of bad math.

Close to 90% of spreadsheet documents contain errors, a 2008 analysis of multiple studies suggests. "Spreadsheets, even after careful development, contain errors in 1% or more of all formula cells," writes [Ray Panko](#), a professor of IT management at the University of Hawaii and an authority on bad spreadsheet practices. "In large spreadsheets with thousands of formulas, there will be dozens of undetected errors."

<http://www.marketwatch.com/story/88-of-spreadsheets-have-errors-2013-04-17>

Common Shortcuts in Excel

- Ctrl + arrow key
 - Go to the end of the text in that direction
- Ctrl + c
 - Copy the selection to clipboard
- Ctrl + v
 - Paste clipboard contents to the cell(s)
- More shortcuts at
 - <http://office.microsoft.com/en-us/excel-help/keyboard-shortcuts-in-excel-2010-HP010342494.aspx>

Common Shortcuts in Excel

PDF Quick Guides

- CTRL key shortcuts
 - <http://officeimg.vo.msecnd.net/en-us/files/680/886/AF102692107.pdf>
- FUNCTION key shortcuts
 - <http://officeimg.vo.msecnd.net/en-us/files/051/990/AF102692116.pdf>
- Miscellaneous key shortcuts
 - <http://officeimg.vo.msecnd.net/en-us/files/149/767/AF102692134.pdf>

GETTING STARTED

Open Inventory_Example.xlsx

Writing Simple Formulas

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	
3	Pens	30	\$ 6.99	
4	Copier Paper	5	\$ 15.99	
5	Paper clips	100	\$ 4.99	
6			Total	

Writing Simple Formulas

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	=B2*C2
3	Pens	30	\$ 6.99	
4	Copier Paper	5	\$ 15.99	
5	Paper clips	100	\$ 4.99	
6			Total	

Writing Simple Formulas

A	B	C	D	
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	
4	Copier Paper	5	\$ 15.99	
5	Paper clips	100	\$ 4.99	
6			Total	

Writing Simple Formulas

A	B	C	D	
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6			Total	

Writing Simple Formulas

A	B	C	D	E
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6			Total	=D2+D3+D4+D5

Writing Simple Formulas

A	B	C	D	
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6			Total	\$ 893.55

Writing Simple Formulas

	A	B	C	D	E
1	Description	Quantity	Unit Price	Total	
2	Binders	10	\$ 10.49	\$ 104.90	
3	Pens	30	\$ 6.99	\$ 209.70	
4	Copier Paper	5	\$ 15.99	\$ 79.95	
5	Paper clips	100	\$ 4.99	\$ 499.00	
6			Total	=SUM(D2:D5)	

Writing Simple Formulas

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6			Total	\$ 893.55

EXCEL FUNCTIONS

Full documentation is online at
<http://office.microsoft.com/en-us/excel-help/excel-functions-by-category-HP010342656.aspx>

Basic Syntax

=<FunctionName>(parameter1,parameter2,...)

Example:

=SUM(A5:A8)

- Will sum the values in cells A5 through A8

SUM Function

	A	B	C	D	E
1	Description	Quantity	Unit Price	Total	
2	Binders	10	\$ 10.49	\$ 104.90	
3	Pens	30	\$ 6.99	\$ 209.70	
4	Copier Paper	5	\$ 15.99	\$ 79.95	
5	Paper clips	100	\$ 4.99	\$ 499.00	
6			Total	=SUM(D2:D5)	

COUNT Function

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6	Number of Items	=COUNT(B2:B5)		\$ 893.55

COUNT Function

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6	Number of Items	4	Total	\$ 893.55

Calculating Discount

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6	Number of Items	4	Sub Total	\$ 893.55
7			Delivery Fee	\$ 15.00
8			10% Discount	=-(D6*0.1)

Calculating Discount

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6	Number of Items	4	Sub Total	\$ 893.55
7			Delivery Fee	\$ 15.00
8			10% Discount	\$ (89.36)

SUM Function with Negative Numbers

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6	Number of Items	4	Sub Total	\$ 893.55
7			Delivery Fee	\$ 15.00
8			10% Discount	\$ (89.36)
9			Total	=SUM(D6:D8)
10				

SUM Function with Negative Numbers

	A	B	C	D
1	Description	Quantity	Unit Price	Total
2	Binders	10	\$ 10.49	\$ 104.90
3	Pens	30	\$ 6.99	\$ 209.70
4	Copier Paper	5	\$ 15.99	\$ 79.95
5	Paper clips	100	\$ 4.99	\$ 499.00
6	Number of Items	4	Sub Total	\$ 893.55
7			Delivery Fee	\$ 15.00
8			10% Discount	\$ (89.36)
9			Total	\$ 819.20

Functions

- Allow you to do calculations in Excel
- Also allow you to manipulate text

Categories of Functions

- Compatibility functions
- Cube functions
- Database functions
- Date and time functions
- Engineering functions
- Financial functions
- Information functions
- Logical functions
- Lookup and reference functions
- Math and trigonometry functions
- Statistical functions
- Text functions
- User defined functions that are installed with add-ins

**CALCULATE THE TOTAL POPULATION OF
THE 5 BOROUGHS FOR EACH DECADE
BETWEEN 1970 AND 2010**

Open the file NYC_Population_1970-2010.xlsx

From NYC DCP Bytes of the Big Apple:
http://www.nyc.gov/html/dcp/html/census/demo_tables_2010.shtml

Calculating Total Borough Population

COMMUNITY DISTRICT NUMBER	COMMUNITY DISTRICT NAME	Total Population 1970	Total Population 1980
BRONX COMMUNITY DISTRICTS			
1	Melrose, Mott Haven, Port Morris	138,557	78,441
2	Hunts Point, Longwood	99,493	34,399
3	Morrisania, Crotona Park East	150,636	53,635
4	Highbridge, Concourse Village	144,207	114,312
5	University Hts., Fordham, Mt. Hope	121,807	107,995
6	East Tremont, Belmont	114,137	65,016
7	Bedford Park, Norwood, Fordham	113,764	116,827
8	Riverdale, Kingsbridge, Marble Hill	103,543	98,275
9	Soundview, Parkchester	166,442	167,627
10	Throgs Nk., Co-op City, Pelham Bay	84,948	106,516
11	Pelham Pkwy, Morris Park, Laconia	105,980	99,080
12	Wakefield, Williamsbridge	135,010	128,226
Total Borough Population		=SUM(C3:C14)	
Population Change			
Percentage Change			

Calculating Total Borough Population

COMMUNITY DISTRICT NUMBER	COMMUNITY DISTRICT NAME	Total Population 1970	Total Population 1980
BRONX COMMUNITY DISTRICTS			
1	Melrose, Mott Haven, Port Morris	138,557	78,441
2	Hunts Point, Longwood	99,493	34,399
3	Morrisania, Crotona Park East	150,636	53,635
4	Highbridge, Concourse Village	144,207	114,312
5	University Hts., Fordham, Mt. Hope	121,807	107,995
6	East Tremont, Belmont	114,137	65,016
7	Bedford Park, Norwood, Fordham	113,764	116,827
8	Riverdale, Kingsbridge, Marble Hill	103,543	98,275
9	Soundview, Parkchester	166,442	167,627
10	Throgs Nk., Co-op City, Pelham Bay	84,948	106,516
11	Pelham Pkwy, Morris Park, Laconia	105,980	99,080
12	Wakefield, Williamsbridge	135,010	128,226
Total Borough Population		1,478,524	1,170,349
Population Change			
Percentage Change			

Calculating Population Change

COMMUNITY DISTRICT NUMBER	COMMUNITY DISTRICT NAME	Total Population 1970	Total Population 1980
BRONX COMMUNITY DISTRICTS			
1	Melrose, Mott Haven, Port Morris	138,557	78,441
2	Hunts Point, Longwood	99,493	34,399
3	Morrisania, Crotona Park East	150,636	53,635
4	Highbridge, Concourse Village	144,207	114,312
5	University Hts., Fordham, Mt. Hope	121,807	107,995
6	East Tremont, Belmont	114,137	65,016
7	Bedford Park, Norwood, Fordham	113,764	116,827
8	Riverdale, Kingsbridge, Marble Hill	103,543	98,275
9	Soundview, Parkchester	166,442	167,627
10	Throgs Nk., Co-op City, Pelham Bay	84,948	106,516
11	Pelham Pkwy, Morris Park, Laconia	105,980	99,080
12	Wakefield, Williamsbridge	135,010	128,226
Total Borough Population		1,478,524	1,170,349
Population Change		$=D15-C15$	
Percentage Change			

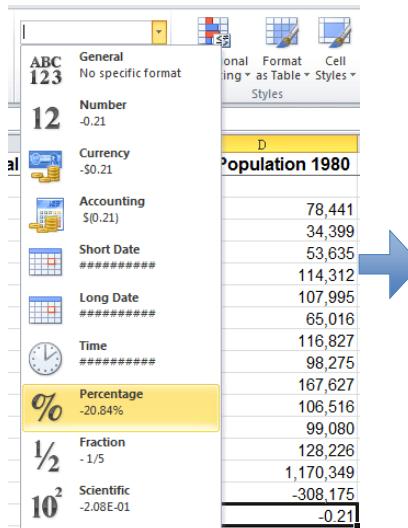
Calculating the Population Change

COMMUNITY DISTRICT NUMBER	COMMUNITY DISTRICT NAME	Total Population 1970	Total Population 1980
BRONX COMMUNITY DISTRICTS			
1	Melrose, Mott Haven, Port Morris	138,557	78,441
2	Hunts Point, Longwood	99,493	34,399
3	Morrisania, Crotona Park East	150,636	53,635
4	Highbridge, Concourse Village	144,207	114,312
5	University Hts., Fordham, Mt. Hope	121,807	107,995
6	East Tremont, Belmont	114,137	65,016
7	Bedford Park, Norwood, Fordham	113,764	116,827
8	Riverdale, Kingsbridge, Marble Hill	103,543	98,275
9	Soundview, Parkchester	166,442	167,627
10	Throgs Nk., Co-op City, Pelham Bay	84,948	106,516
11	Pelham Pkwy, Morris Park, Laconia	105,980	99,080
12	Wakefield, Williamsbridge	135,010	128,226
Total Borough Population		1,478,524	1,170,349
Population Change		-308,175	
Percentage Change			

Calculate Percentage Change

COMMUNITY DISTRICT NUMBER	COMMUNITY DISTRICT NAME	Total Population 1970	Total Population 1980
BRONX COMMUNITY DISTRICTS			
1	Melrose, Mott Haven, Port Morris	138,557	78,441
2	Hunts Point, Longwood	99,493	34,399
3	Morrisania, Crotona Park East	150,636	53,635
4	Highbridge, Concourse Village	144,207	114,312
5	University Hts., Fordham, Mt. Hope	121,807	107,995
6	East Tremont, Belmont	114,137	65,016
7	Bedford Park, Norwood, Fordham	113,764	116,827
8	Riverdale, Kingsbridge, Marble Hill	103,543	98,275
9	Soundview, Parkchester	166,442	167,627
10	Throgs Nk., Co-op City, Pelham Bay	84,948	106,516
11	Pelham Pkwy, Morris Park, Laconia	105,980	99,080
12	Wakefield, Williamsbridge	135,010	128,226
Total Borough Population		1,478,524	1,170,349
Population Change		-308,175	=D16/C15
Percentage Change			

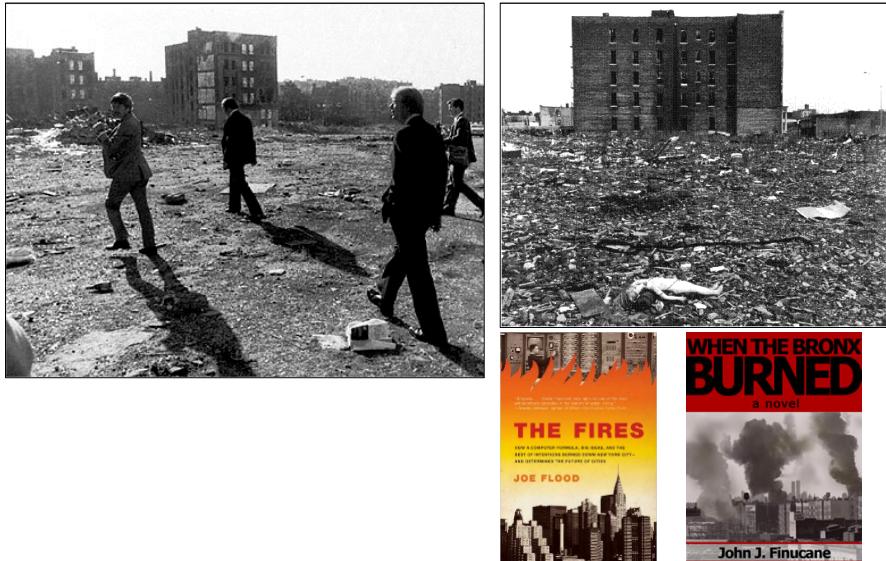
Calculate Percentage Change



The screenshot shows the 'Format Cells' dialog box for cell D16. The 'Number' tab is selected, showing 'Percentage' with '-0.21' decimal places. The formula bar above the dialog box shows the formula $=D16/C15$. The status bar at the bottom of the screen displays the percentage value -20.84%. To the right of the dialog box is a table of Bronx Community District population data.

	Total Population 1970	Total Population 1980
138,557	78,441	
99,493	34,399	
150,636	53,635	
144,207	114,312	
121,807	107,995	
114,137	65,016	
113,764	116,827	
103,543	98,275	
166,442	167,627	
84,948	106,516	
105,980	99,080	
135,010	128,226	
1,478,524	1,170,349	
-308,175		
-0.21	-20.84%	

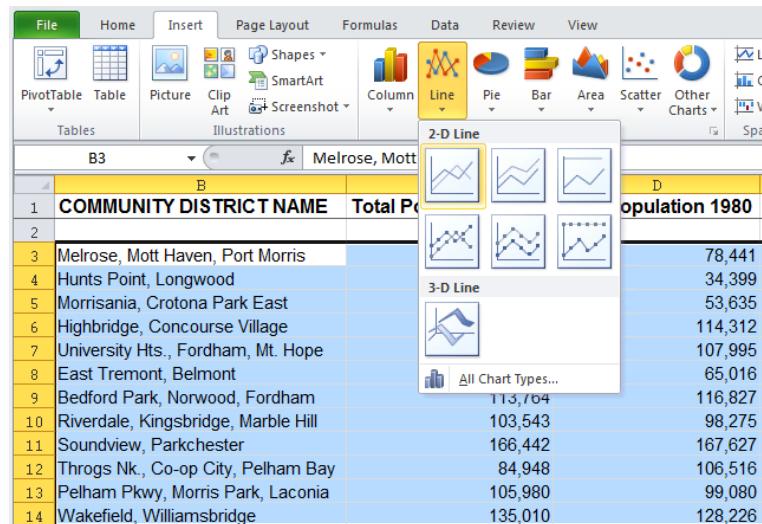
The Bronx in the 1970s



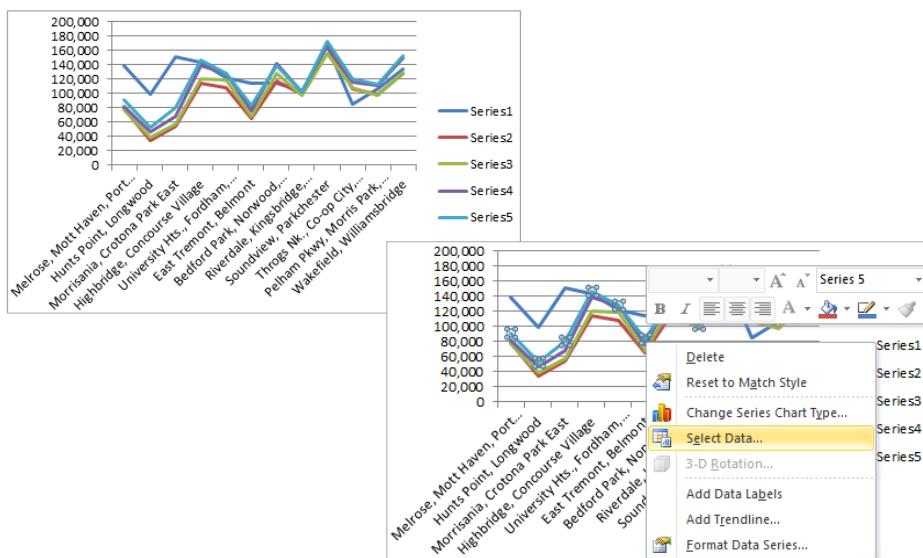
Create Graph of CD Population

COMMUNITY DISTRICT NAME	Total Population 1970	Total Population 1980	Total Population 1990	Total Population 2000	Total Population 2010
Battery Park City, Tribeca	7,706	15,918	25,366	34,420	60,978
Greenwich Village, Soho	84,337	87,069	94,105	93,119	90,016
Lower East Side, Chinatown	181,845	154,848	161,617	164,407	163,277
Chelsea, Clinton	83,601	82,164	84,431	87,479	103,245
Midtown Business District	31,076	39,544	43,507	44,028	51,673
Stuyvesant Town, Turtle Bay	122,465	127,554	133,748	136,152	142,745
West Side, Upper West Side	212,422	206,669	210,993	207,699	209,084
Upper East Side	200,851	204,305	210,880	217,063	219,920
Manhattanville, Hamilton Heights	113,606	103,038	106,978	111,724	110,193
Central Harlem	159,267	105,641	99,519	107,109	115,723
East Harlem	154,662	114,569	110,508	117,743	120,511
Washington Heights, Inwood	180,561	179,941	198,192	208,414	190,020
Total Borough Population	1,532,399	1,421,260	1,479,844	1,529,357	1,577,385
Population Change		-111,139	58,584	49,513	48,028
Percentage Change		-7.25	4.12	3.35	3.14

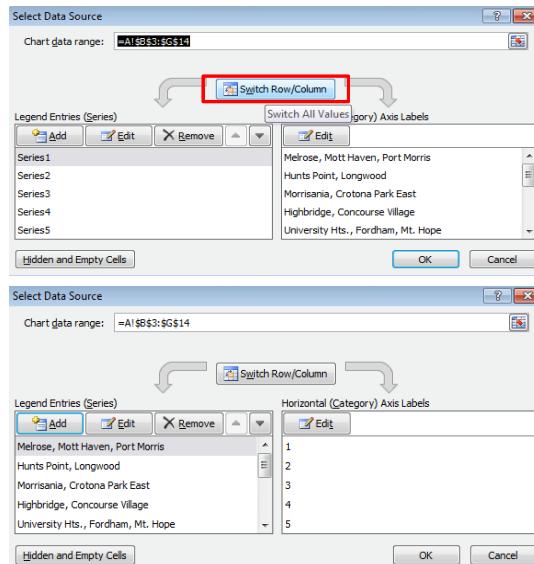
Create Graph of CD Population



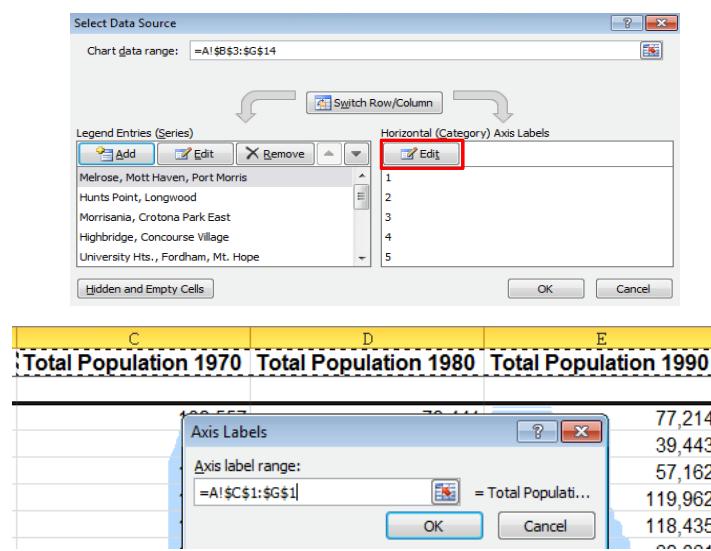
Create Graph of CD Population



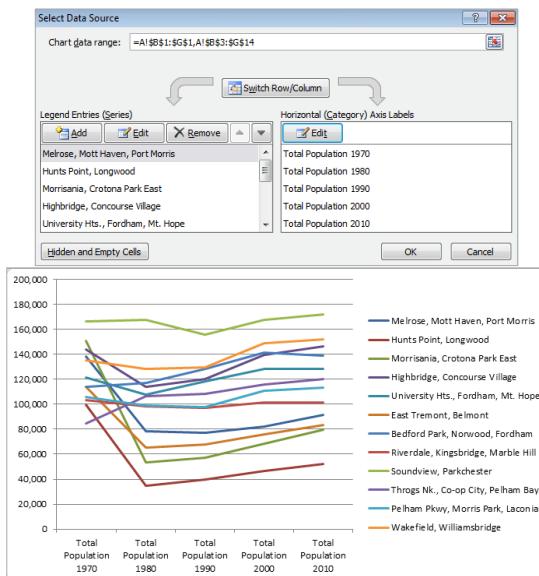
Create Graph of CD Population



Create Graph of CD Population



Create Graph of CD Population



Exploratory Data Analysis

- Goal -> Discover patterns in the data
- Approach
 - Understand the context
 - Summarize fields
 - Use graphical representations of the data
 - Explore outliers

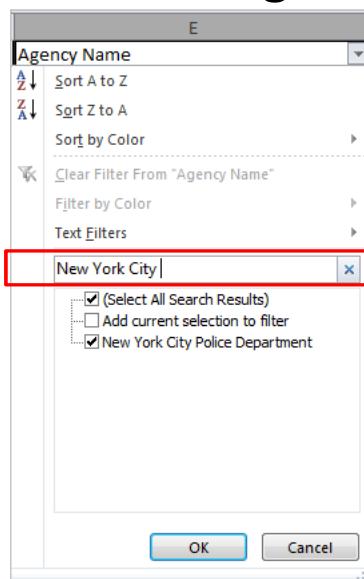
Tukey, J.W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

Filtering

The screenshot shows a Microsoft Excel spreadsheet titled "Filtering". The Data tab is selected in the ribbon. A tooltip for the "Filter" button is displayed, explaining that it enables filtering of selected cells. Below the ribbon, a table is shown with columns: Unique Key, Created Date, Closed Date, Agent, Agency Name, and Complaint Type. The "Agency Name" column header has a dropdown arrow indicating it is filtered. A red box highlights this dropdown arrow.

Unique Key	Created Date	Closed Date	Agent	Agency Name	Complaint Type
Netbook Pro	14-Aug-10	16-Aug-10			

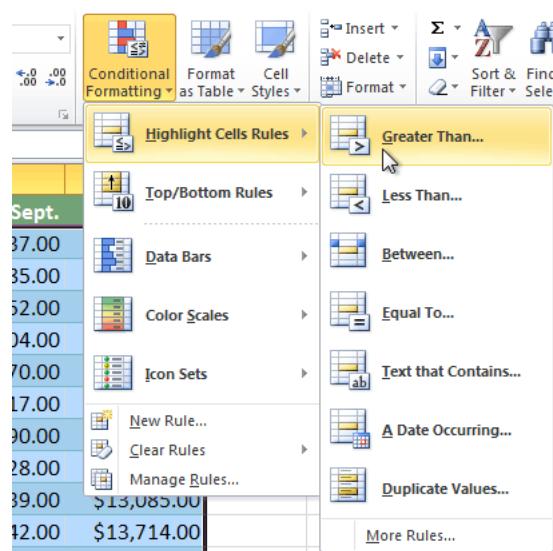
Filtering



Conditional Formatting

- Format cells based on value or add content to cells that visually describe the content
- Great for quickly visualizing data
- Makes tables more “presentation-ready”

Conditional Formatting



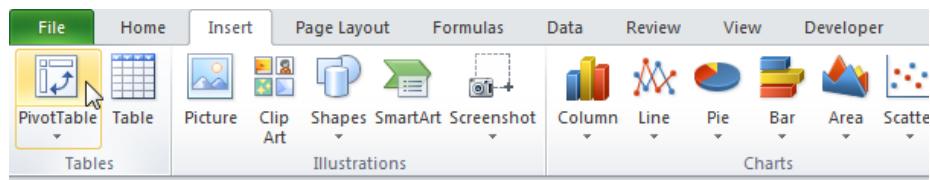
Conditional Formatting



Agency	Agency Name	Complaint Type
NYPD	New York City Police Department	Noise - Vehicle
NYPD	New York City Police Department	Noise - Commercial
NYPD	New York City Police Department	Noise - Street/Sidewalk
EDC	Economic Development Corporation	Noise - Helicopter
EDC	Economic Development Corporation	Noise - Helicopter
NYPD	New York City Police Department	Noise - Street/Sidewalk
EDC	Economic Development Corporation	Noise - Helicopter
NYPD	New York City Police Department	Noise - Street/Sidewalk
NYPD	New York City Police Department	Noise - Street/Sidewalk
NYPD	New York City Police Department	Noise - Street/Sidewalk
NYPD	New York City Police Department	Noise - Vehicle

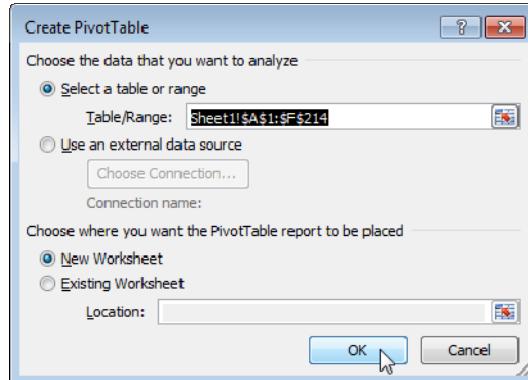
PivotTables

- What is a PivotTable?
 - A data summarization tool for quickly understanding and displaying the data you're analyzing
- How do I find it?



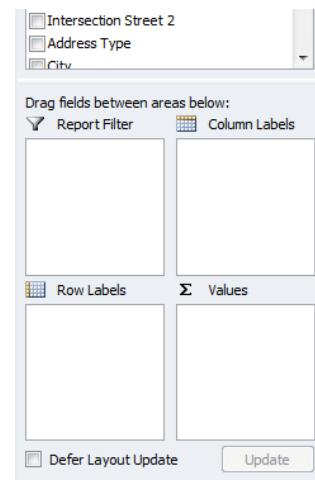
PivotTables

- Selecting range and destination

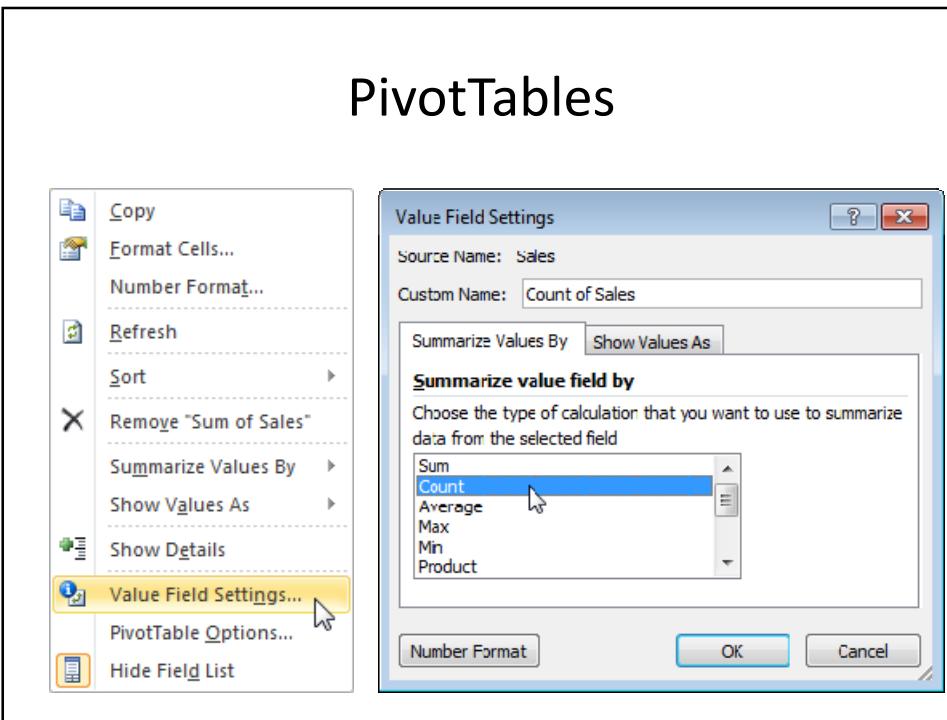


PivotTables

- Drag and drop fields to visualize
 - Row labels
 - Values
 - Filter
 - Column Labels



PivotTables

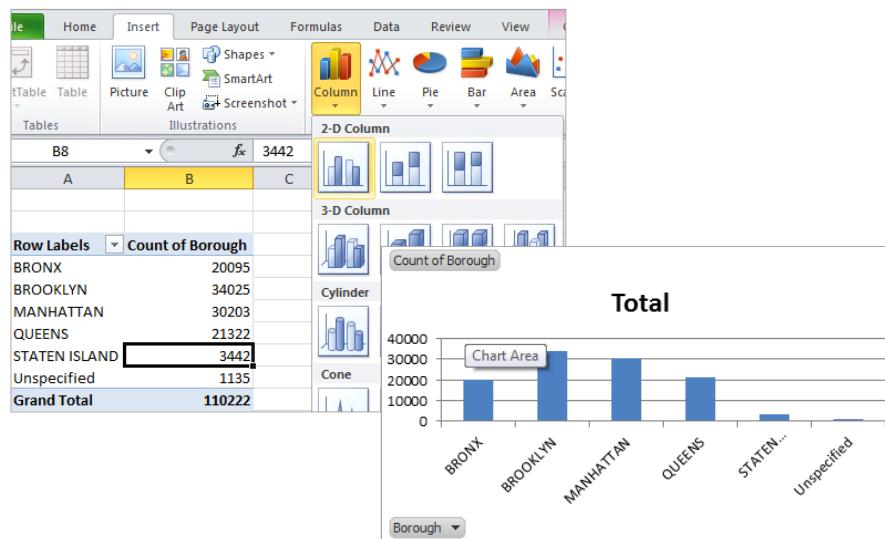


PivotTable of Complaints by Borough

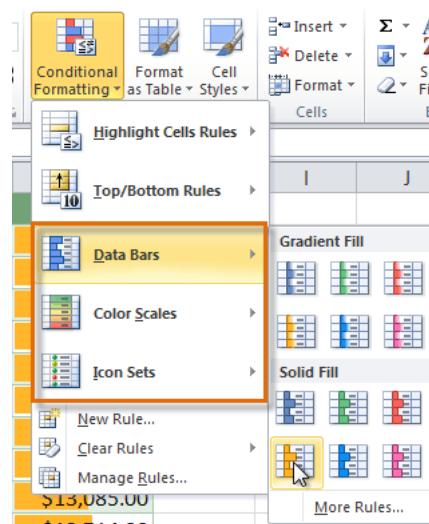
The screenshot shows the 'PivotTable Field List' window on the left and the resulting PivotTable report on the right. In the PivotTable Field List, the 'Borough' field is selected and being dragged into the Row Labels area. The resulting PivotTable report shows the count of complaints for each borough, with a total of 110,222 complaints across all boroughs.

Row Labels	Count of Borough
BRONX	20095
BROOKLYN	34025
MANHATTAN	30203
QUEENS	21322
STATEN ISLAND	3442
Unspecified	1135
Grand Total	110222

Chart Complaints by Borough



Conditional Formatting



Conditional Formatting

Row Labels	Count of Borough	Row Labels	Count of Borough
BRONX	20095	BRONX	20095
BROOKLYN	34025	BROOKLYN	34025
MANHATTAN	30203	MANHATTAN	30203
QUEENS	21322	QUEENS	21322
STATEN ISLAND	3442	STATEN ISLAND	3442
Unspecified	1135	Unspecified	1135
Grand Total	110222	Grand Total	110222

Conditional Formatting

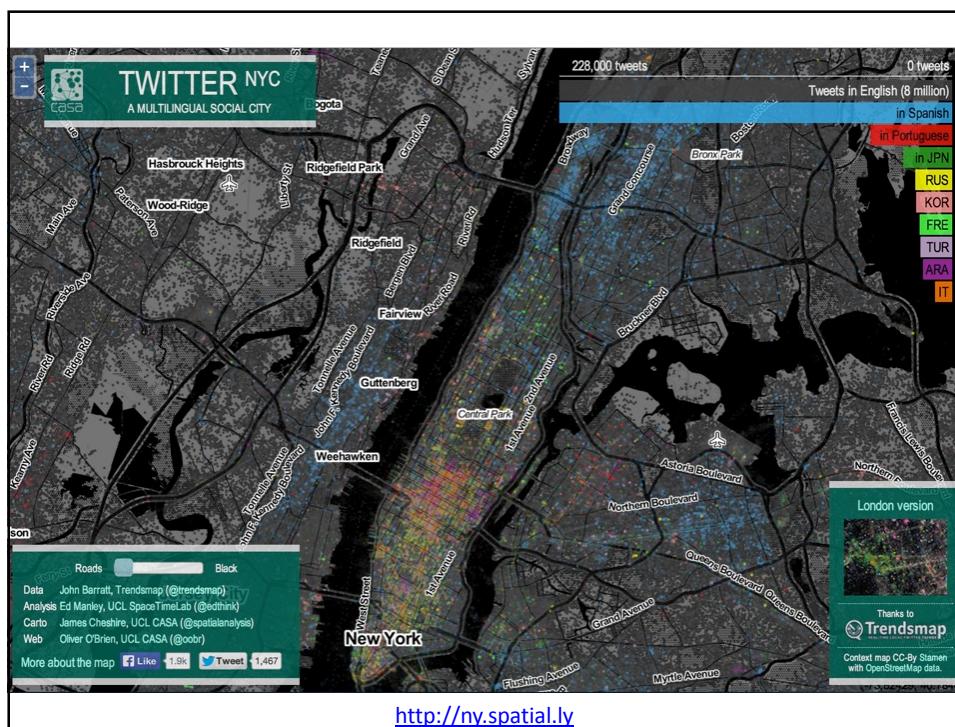
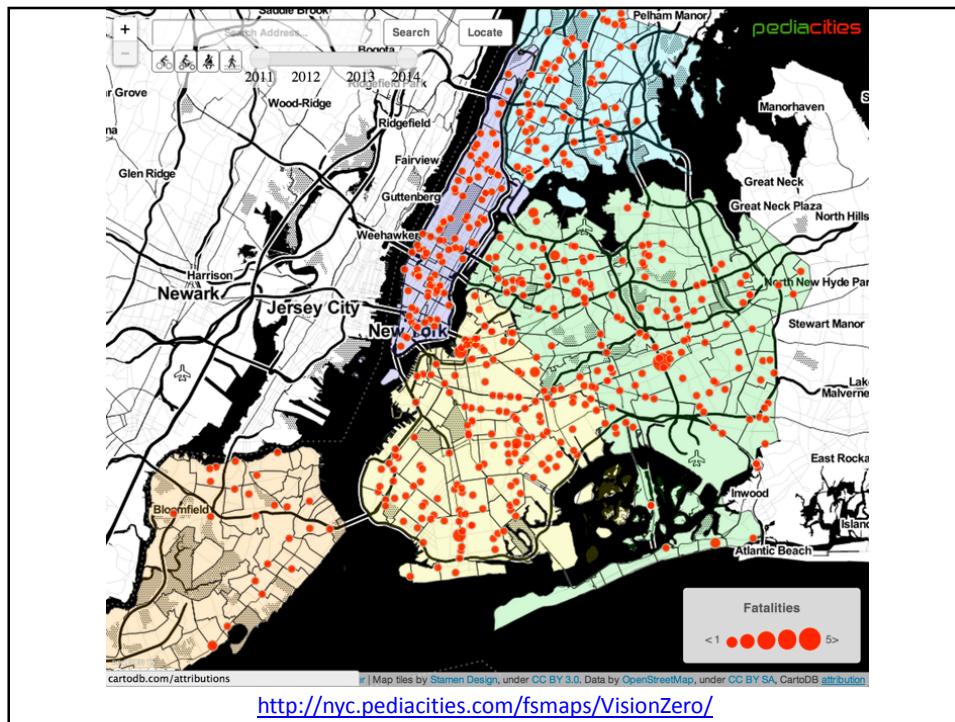
Row Labels	Count of Borough	Row Labels	Count of Borough
BRONX	20095	BRONX	20095
BROOKLYN	34025	BROOKLYN	34025
MANHATTAN	30203	MANHATTAN	30203
QUEENS	21322	QUEENS	21322
STATEN ISLAND	3442	STATEN ISLAND	3442
Unspecified	1135	Unspecified	1135
Grand Total	110222	Grand Total	110222

Exercise

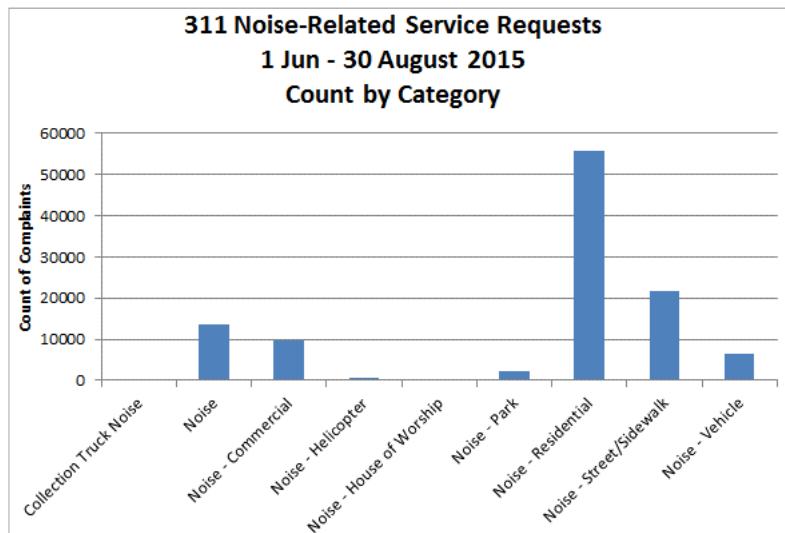
- In your groups, explore the complaints for your assigned borough
 - What are the common types of complaints?
 - Which Community Districts have the most complaints?
 - Identify any other meaningful patterns to the noise complaints

Goals for the Afternoon

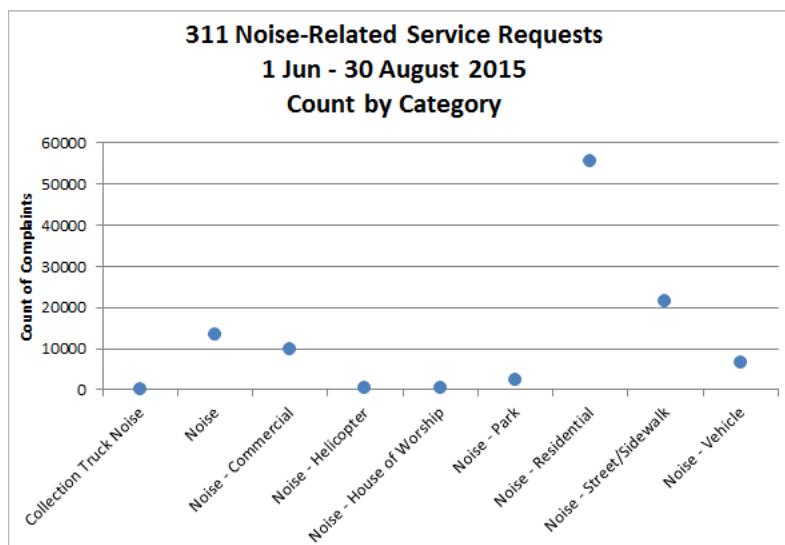
- Discuss the basic principles of visualization design
- Discuss visualizing data in Excel
- Apply advanced functions to summarizing data in Excel
- Practice communicating analytical findings



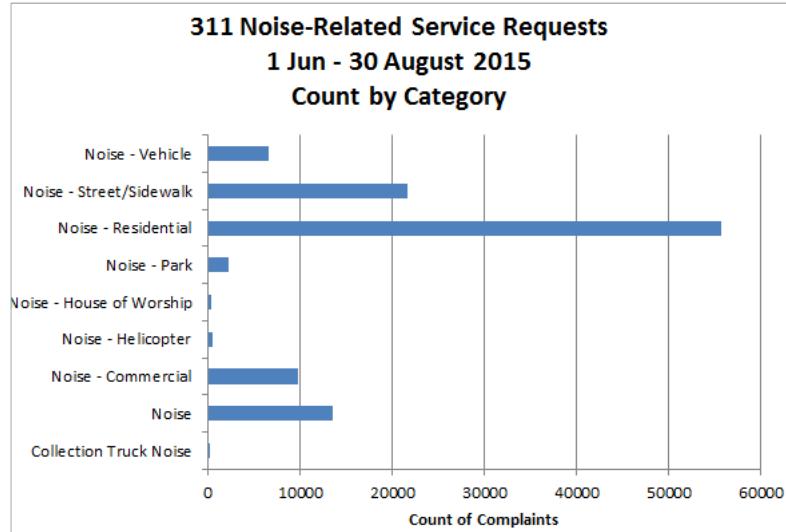
Bar Chart



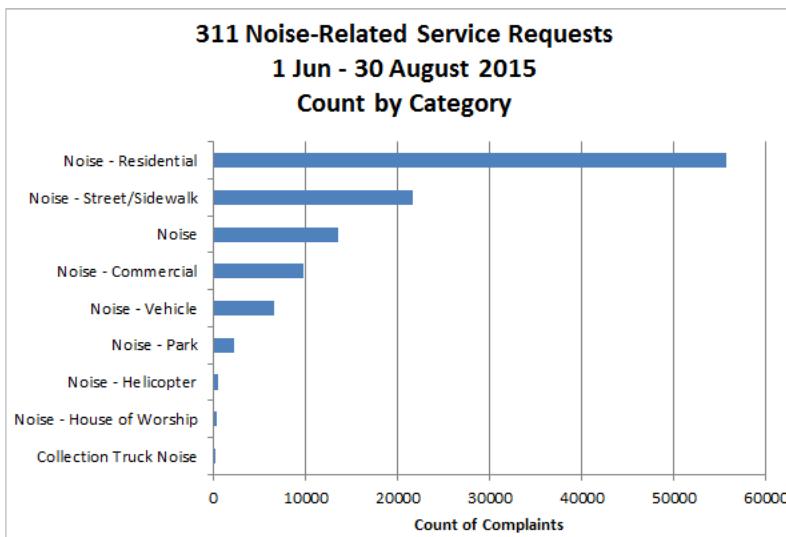
“Barless” Bar Chart



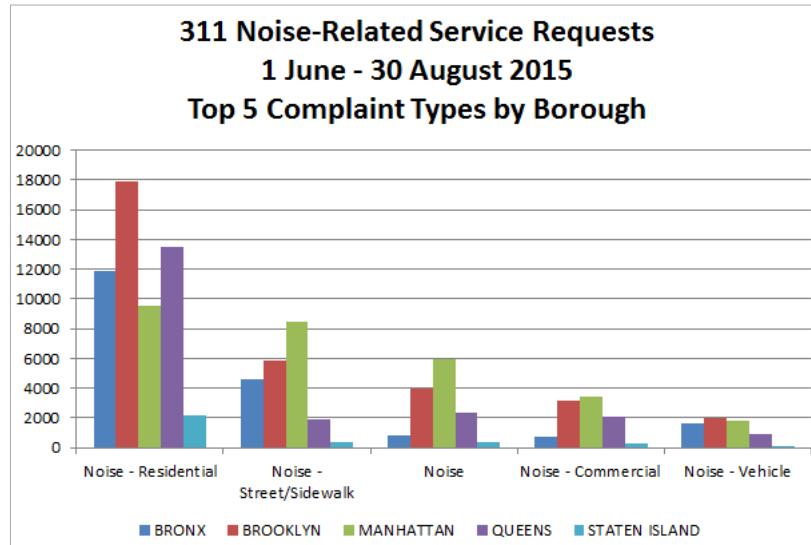
Horizontal Bar Chart



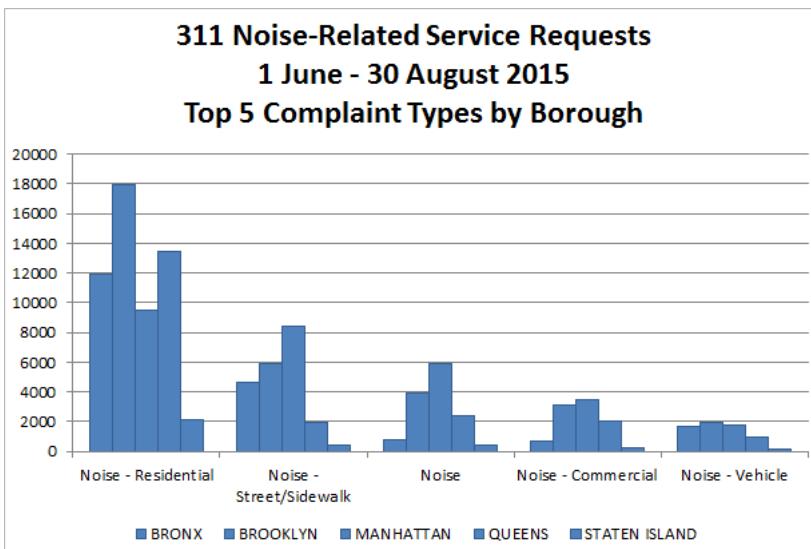
Ranked Bar Chart



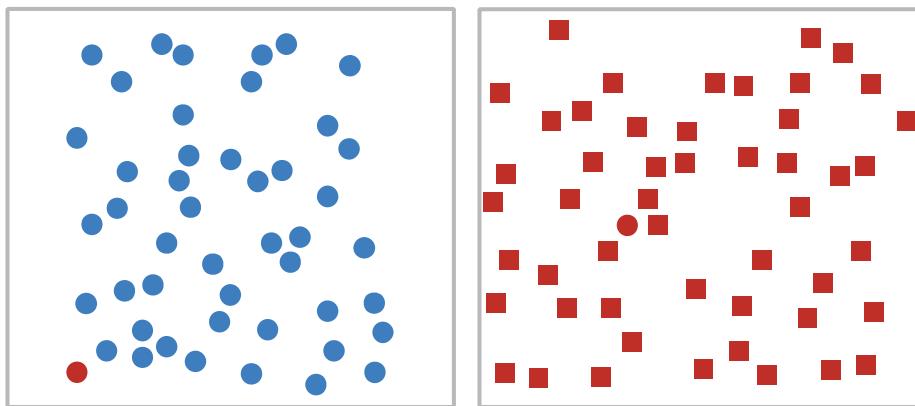
Grouped Bar Charts



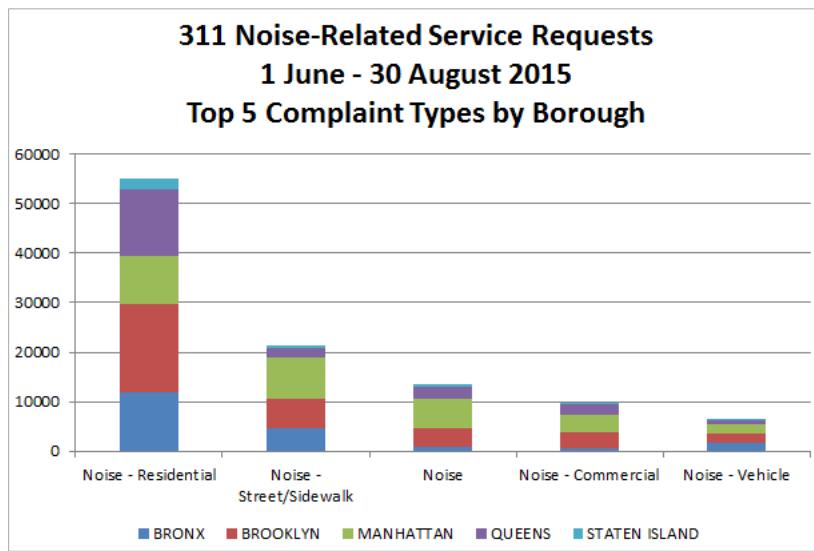
Grouped Bar Charts



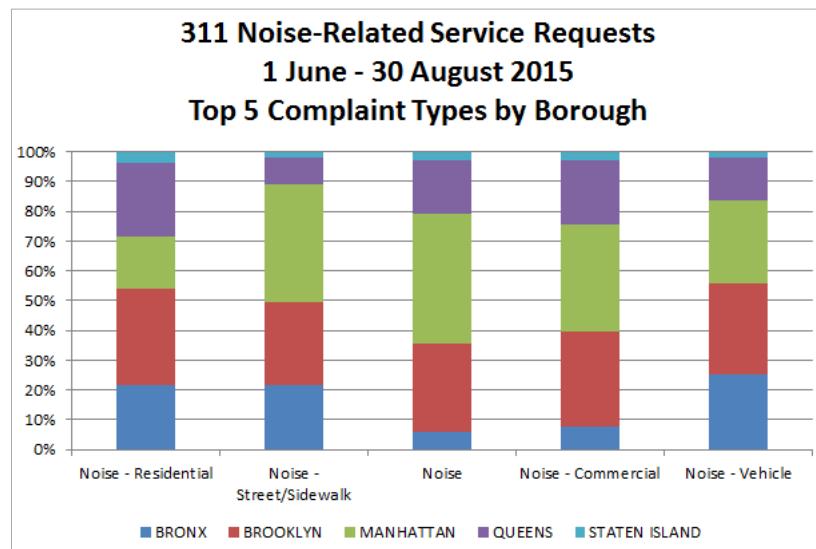
Precognitive Processing



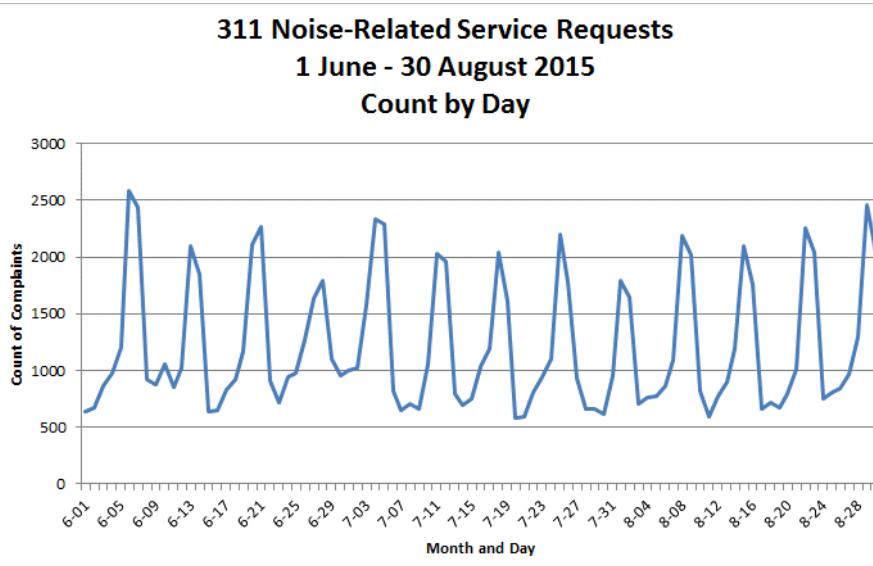
Stacked Bar Chart



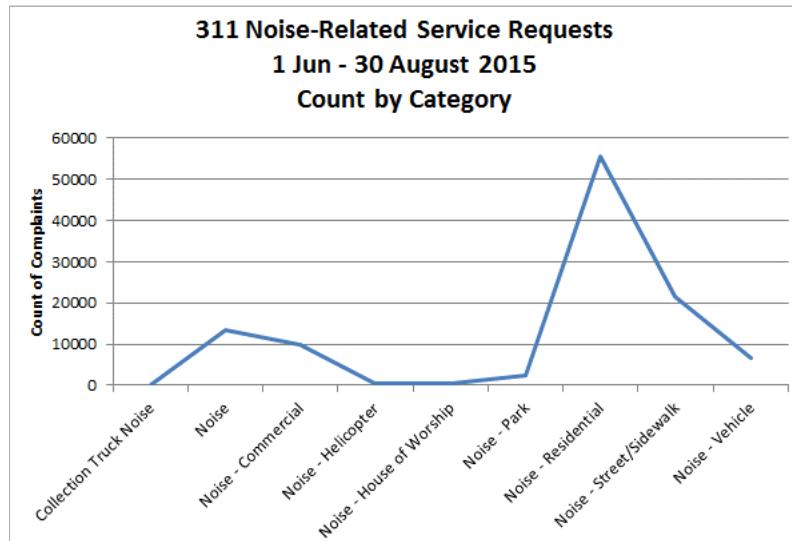
100% Stacked Bar Chart



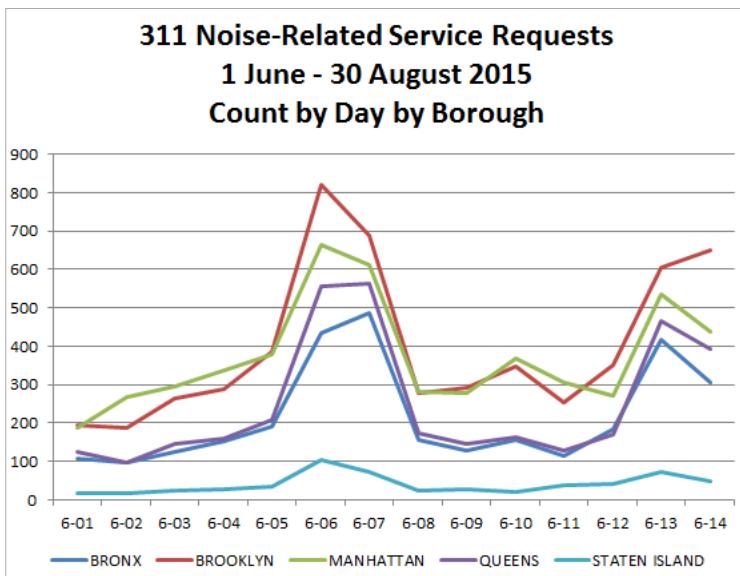
Line Chart

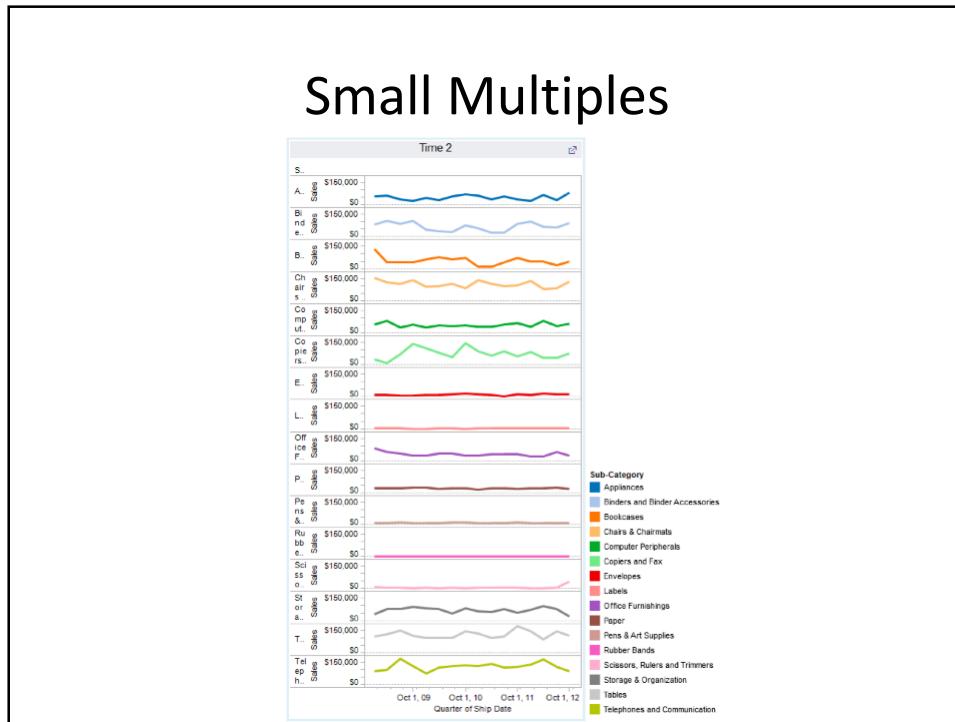
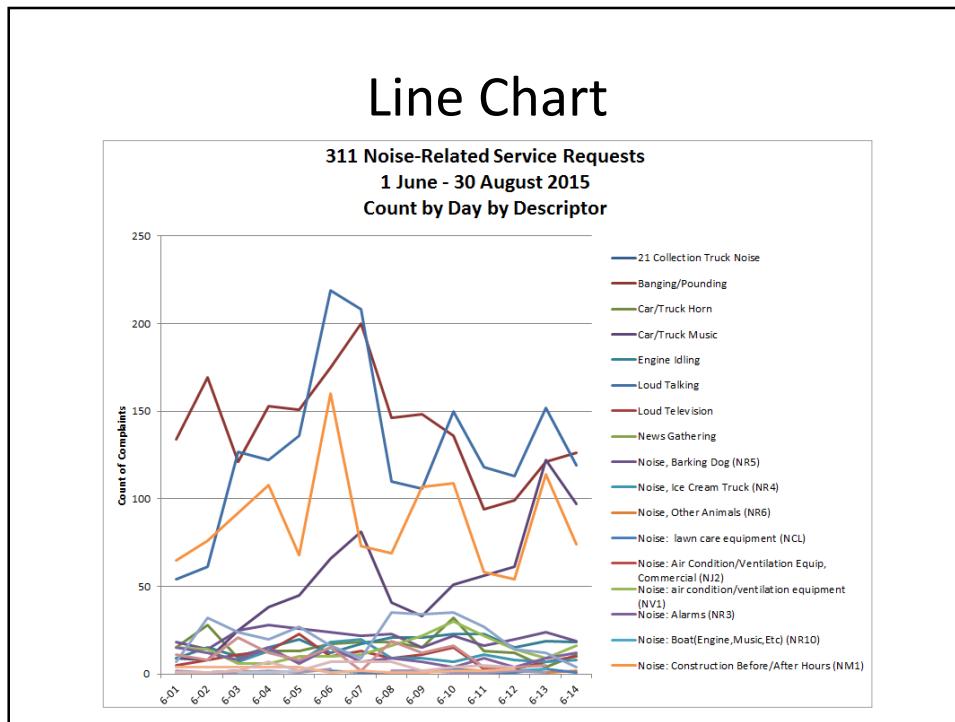


Line Chart with Categorical Data



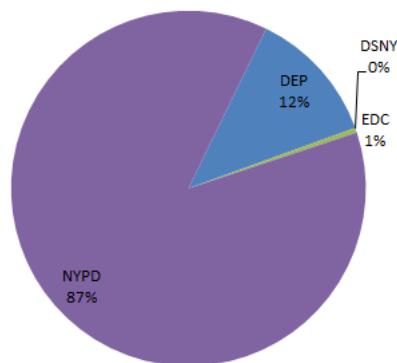
Line Chart





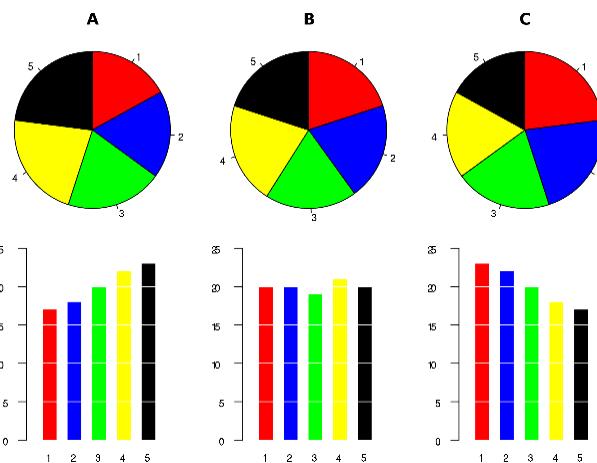
Pie Charts

**311 Noise-Related Service Requests
1 June - 30 August 2015
Complaints by Agency Responsible**



Pie Charts

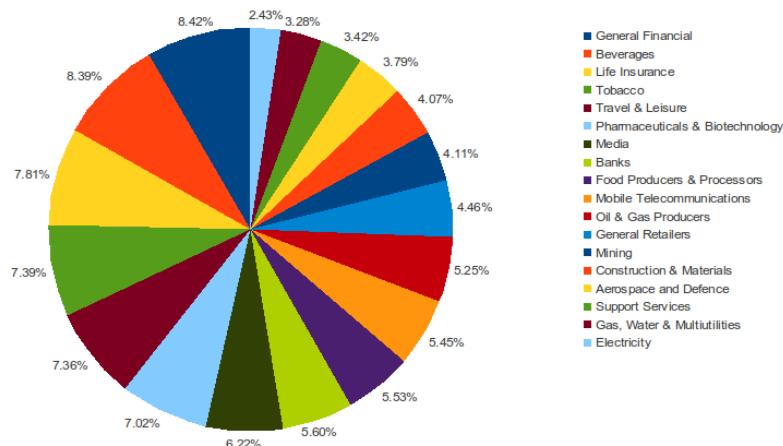
Which candidate got the most votes?



"Piecharts". Licensed under CC BY 1.0 via Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:Piecharts.svg#/media/File:Piecharts.svg>

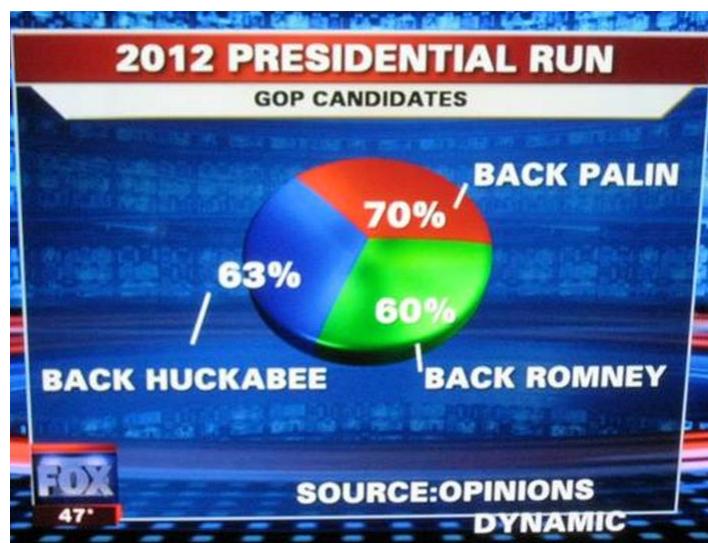
Pie Charts

Sector Weightings

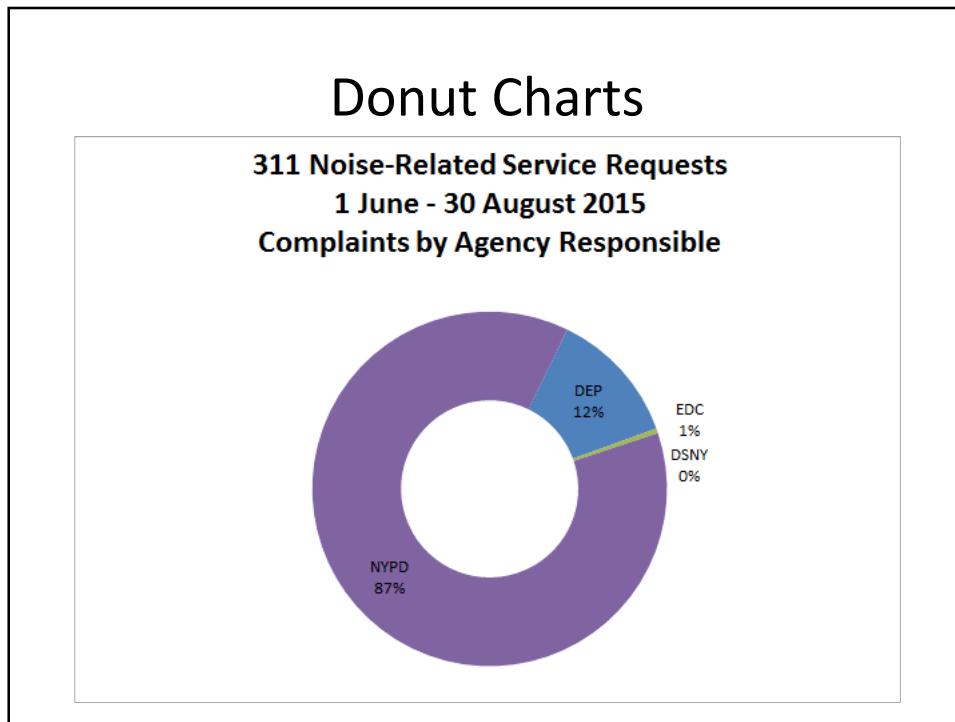


"Badpie" by Gilgongo - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:Badpie.png#/media/File:Badpie.png>

Pie Charts



<http://simplystatistics.org/2012/11/26/the-statisticians-at-fox-news-use-classic-and-novel-graphical-techniques-to-lead-with-data/>



Design Tip

How do you learn to make good charts?

...Make a lot of bad charts

Calculating the Hour from a Timestamp

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	=HOUR(B2)
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

HOUR Function

HOUR function

This article describes the formula syntax and usage of the HOUR function in Microsoft Excel.

Description

Returns the hour of a time value. The hour is given as an integer, ranging from 0 (12:00 A.M.) to 23 (11:00 P.M.).

Syntax

`HOUR(serial_number)`

The HOUR function syntax has the following arguments:

- **Serial_number** Required. The time that contains the hour you want to find. Times may be entered as text strings within quotation marks (for example, "6:45 PM"), as decimal numbers (for example, 0.78125, which represents 6:45 PM), or as results of other formulas or functions (for example, `TIMEVALUE("6:45 PM")`).

<https://support.office.com/en-us/article/HOUR-function-e2833b50-0db0-499b-acc5-e9ae03de8fb>

Calculating the Hour from a Timestamp

Unique Key	Created Date	hour
30195273	3/18/15 2:12	1/3/04 0:00
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

Calculating the Hour from a Timestamp

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	1/3/04 0:00
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

Calculating the Hour from a Timestamp

The screenshot shows the Microsoft Excel ribbon at the top. The 'Home' tab is selected. In the 'Number' section of the ribbon, a dropdown menu is open, showing various format categories like General, Number, Currency, Time, Accounting, etc. The 'Number' category is highlighted with a red box. Below the dropdown, a table is displayed with columns A, B, and C. Column A is 'Unique Key', column B is 'Created Date', and column C is 'hour'. The formula in cell C2 is =B2*24, resulting in 2.00. The formula bar at the top also shows =B2*24.

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	2.00
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

Calculating the Hour from a Timestamp

This screenshot is similar to the one above, but the 'General' category in the 'Number' dropdown menu is highlighted with a red box. The table and formula are identical to the previous screenshot.

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	2
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

Calculating the Hour from a Timestamp

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	2
30203057	3/18/15 2:00	
30197320	3/18/15 1:58	
30194112	3/18/15 1:37	
30202379	3/18/15 1:36	
30199506	3/18/15 1:28	

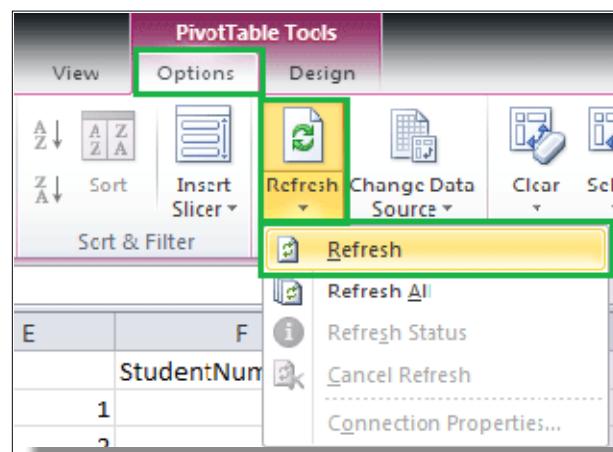
Double-click on lower right hand corner to expand formula to bottom of column

Calculating the Hour from a Timestamp

A	B	C
Unique Key	Created Date	hour
30195273	3/18/15 2:12	2
30203057	3/18/15 2:00	2
30197320	3/18/15 1:58	1
30194112	3/18/15 1:37	1
30202379	3/18/15 1:36	1
30199506	3/18/15 1:28	1

CREATE A PIVOTTABLE OF HOURS WHEN COMPLAINTS ARE RECEIVED

Refresh PivotTable to show Hours column



Refresh PivotTable to show Hours column

The image shows two identical 'PivotTable Field List' dialog boxes side-by-side. Both lists contain a long vertical scrollable list of fields: Unique Key, Created Date, Closed Date, Agency, Agency Name, Complaint Type, Descriptor, Location Type, Incident Zip, Incident Address, Street Name, Cross Street 1, Cross Street 2, Intersection Street 1, Intersection Street 2, and Address Type. In the right-hand dialog box, the 'Hour' field is highlighted with a red rectangular box.

Chart of Complaints by Hour

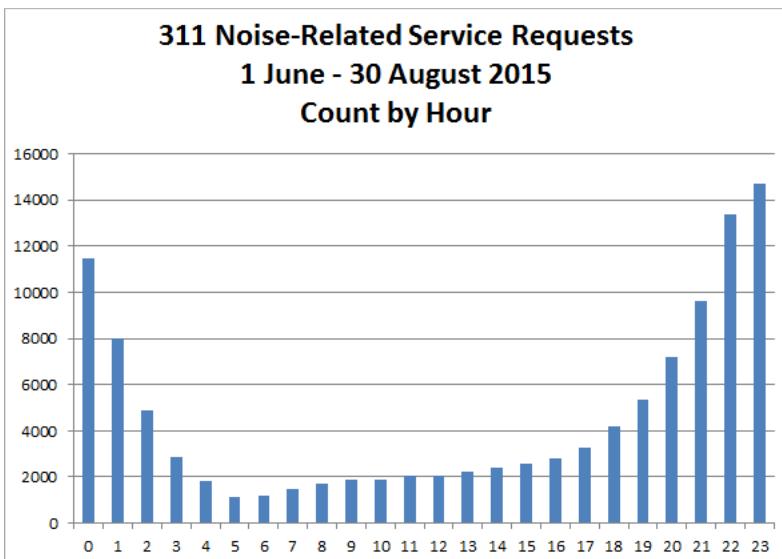
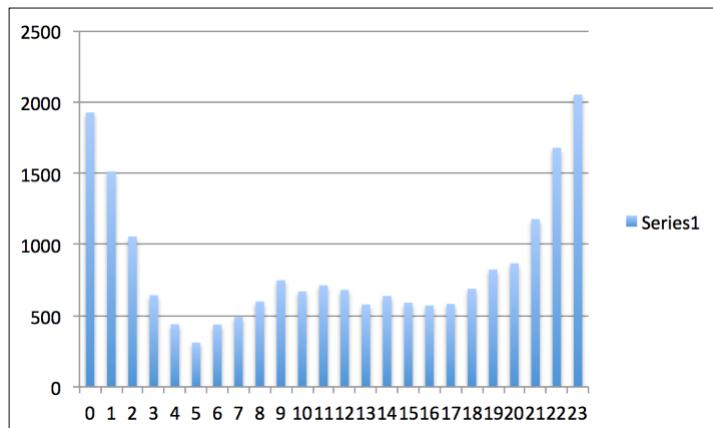
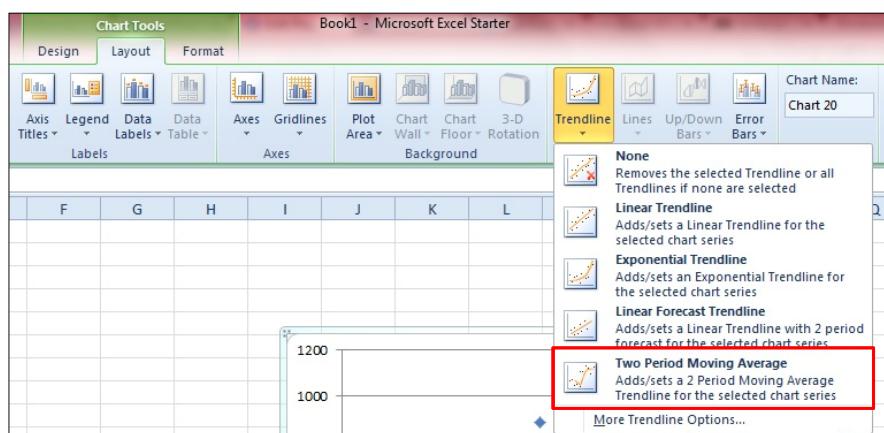


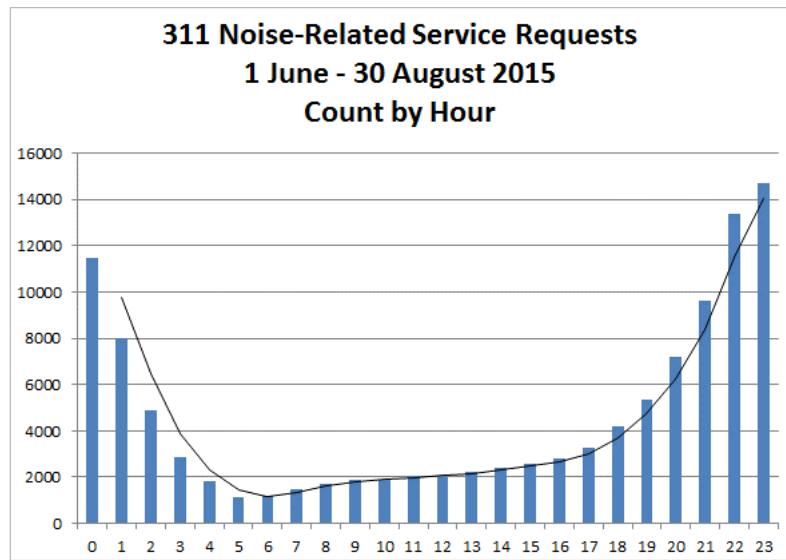
Chart of Complaints by Hour



Add a Trendline



Add a Trendline



Calculate the Day of the Week from Timestamp

A	B	C	D	E
Unique Key	Created Date	hour	dow	Closed Date
30195273	3/18/15 2:12		2	=WEEKDAY(B2)
30203057	3/18/15 2:00		2	
30197320	3/18/15 1:58		1	
30194112	3/18/15 1:37		1	3/18/15 3:09
30202379	3/18/15 1:36		1	3/18/15 2:06
30199506	3/18/15 1:28		1	

DAY Function

DAY function

This article describes the formula syntax and usage of the DAYfunction in Microsoft Excel.

Description

Returns the day of a date, represented by a serial number. The day is given as an integer ranging from 1 to 31.

Syntax

DAY(serial_number)

The DAY function syntax has the following arguments:

- **Serial_number** Required. The date of the day you are trying to find. Dates should be entered by using the DATE function, or as results of other formulas or functions. For example, use DATE(2008,5,23) for the 23rd day of May, 2008. Problems can occur if [dates are entered as text](#).

WEEKDAY Function

WEEKDAY function

This article describes the formula syntax and usage of the WEEKDAYfunction in Microsoft Excel.

Description

Returns the day of the week corresponding to a date. The day is given as an integer, ranging from 1 (Sunday) to 7 (Saturday), by default.

Syntax

WEEKDAY(serial_number,[return_type])

The WEEKDAY function syntax has the following arguments:

- **Serial_number** Required. A sequential number that represents the date of the day you are trying to find. Dates should be entered by using the DATE function, or as results of other formulas or functions. For example, use DATE(2008,5,23) for the 23rd day of May, 2008. Problems can occur if dates are entered as text.
- **Return_type** Optional. A number that determines the type of return value.

Return_type	Number returned
--------------------	------------------------

1 or omitted	Numbers 1 (Sunday) through 7 (Saturday). Behaves like previous versions of Microsoft Excel.
--------------	---

<https://support.office.com/en-us/article/WEEKDAY-function-e5dbebc7-5818-4818-90c0-7e7129e138d2>

Calculate the Day of the Week from Timestamp

A	B	C	D
Unique Key	Created Date	hour	dow
30195273	3/18/15 2:12	2	4
30203057	3/18/15 2:00	2	
30197320	3/18/15 1:58	1	
30194112	3/18/15 1:37	1	
30202379	3/18/15 1:36	1	
30199506	3/18/15 1:28	1	

Calculate the Day of the Week from Timestamp

A	B	C	D
Unique Key	Created Date	hour	dow
30195273	3/18/15 2:12	2	4
30203057	3/18/15 2:00	2	4
30197320	3/18/15 1:58	1	4
30194112	3/18/15 1:37	1	4
30202379	3/18/15 1:36	1	4
30199506	3/18/15 1:28	1	4

CHOOSE Function

CHOOSE function

This article describes the formula syntax and usage of the CHOOSEfunction in Microsoft Excel.

Description

Uses index_num to return a value from the list of value arguments. Use CHOOSE to select one of up to 254 values based on the index number. For example, if value1 through value7 are the days of the week, CHOOSE returns one of the days when a number between 1 and 7 is used as index_num.

Syntax

CHOOSE(index_num, value1, [value2], ...)

The CHOOSE function syntax has the following arguments:

- **Index_num** Required. Specifies which value argument is selected. Index_num must be a number between 1 and 254, or a formula or reference to a cell containing a number between 1 and 254.
 - If index_num is 1, CHOOSE returns value1; if it is 2, CHOOSE returns value2; and so on.
 - If index_num is less than 1 or greater than the number of the last value in the list, CHOOSE returns the #VALUE! error value.
 - If index_num is a fraction, it is truncated to the lowest integer before being used.
- **Value1, value2, ...** Value 1 is required, subsequent values are optional. 1 to 254 value arguments from which CHOOSE selects a value or an action to perform based on index_num. The arguments can be numbers, cell references, defined names, formulas, functions, or text.

CHOOSE Function

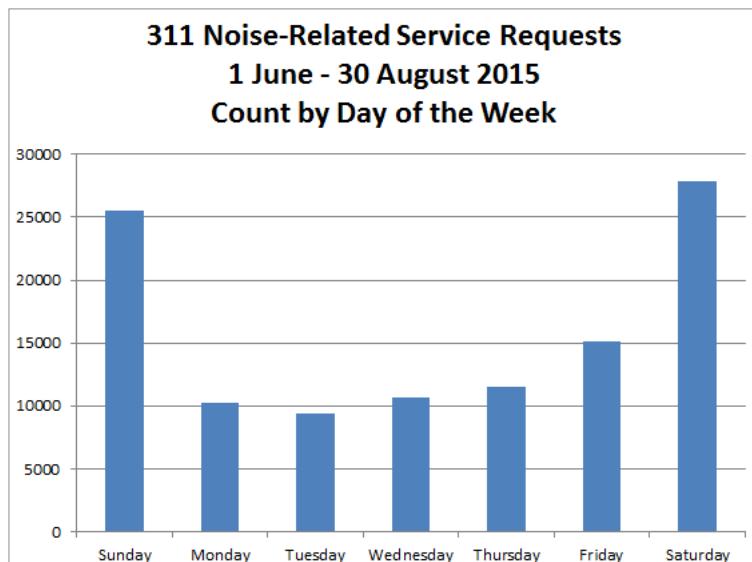
=CHOOSE(D2,"Sunday","Monday","Tuesday","Wednesday","Thursday","Friday","Saturday")

A	B	C	D	E
Unique Key	Created Date	hour	dow	weekday
30195273	3/18/15 2:12	2	4	Wednesday
30203057	3/18/15 2:00	2	4	
30197320	3/18/15 1:58	1	4	
30194112	3/18/15 1:37	1	4	
30202379	3/18/15 1:36	1	4	
30199506	3/18/15 1:28	1	4	

CHOOSE Function

A	B	C	D	E
Unique Key	Created Date	hour	dow	weekday
30195273	3/18/15 2:12		2	4 Wednesday
30203057	3/18/15 2:00		2	4 Wednesday
30197320	3/18/15 1:58		1	4 Wednesday
30194112	3/18/15 1:37		1	4 Wednesday
30202379	3/18/15 1:36		1	4 Wednesday
30199506	3/18/15 1:28		1	4 Wednesday

Complaints by Day of the Week

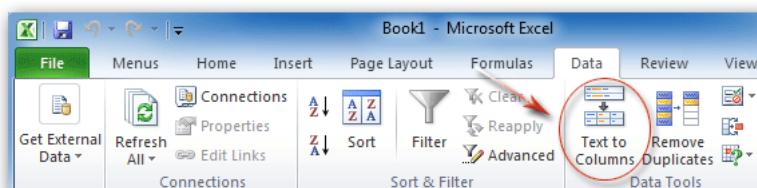


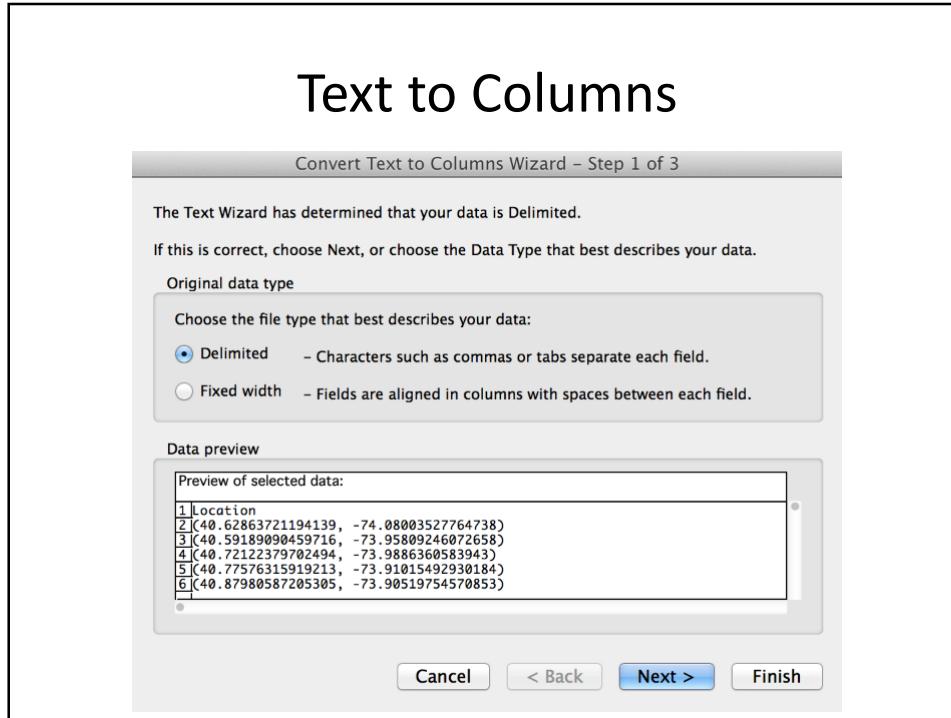
Splitting Data

	AX	AY	AZ
1	Latitude	Longitude	Location
2	40.62863721	-74.08003528	(40.62863721194139, -74.08003527764738)
3	40.5918909	-73.95809246	(40.59189090459716, -73.95809246072658)
4	40.7212238	-73.98863606	(40.72122379702494, -73.9886360583943)
5	40.77576316	-73.91015493	(40.77576315919213, -73.91015492930184)
6	40.87980587	-73.90519755	(40.87980587205305, -73.90519754570853)
7	40.57480362	-73.97352786	(40.57480361606634, -73.9735278635962)
8	40.68321658	-73.95387066	(40.68321657803109, -73.9538706637591)
9	40.6818709	-73.7660623	(40.68187089534303, -73.7660623046722)
10	40.70605964	-73.83143704	(40.70605964424251, -73.83143703998772)
11	40.74786736	-73.81847221	(40.74786736221416, -73.81847220968915)
12	40.64795817	-74.00018739	(40.64795817286927, -74.00018738922621)
13	40.6875289	-73.9717271	(40.68752889748776, -73.97172709811349)
14	40.86939583	-73.91661612	(40.86939583069894, -73.91661611865678)
15	40.71346027	-73.95874828	(40.7134602748696, -73.95874828450539)
16	40.86954934	-73.91634837	(40.86954933988228, -73.9163483681302)

What if we wanted to split the data in the Location column?

Text to Columns





This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

- Tab
- Semicolon
- Comma
- Space
- Other:

Treat consecutive delimiters as one

Text qualifier: &

Data preview

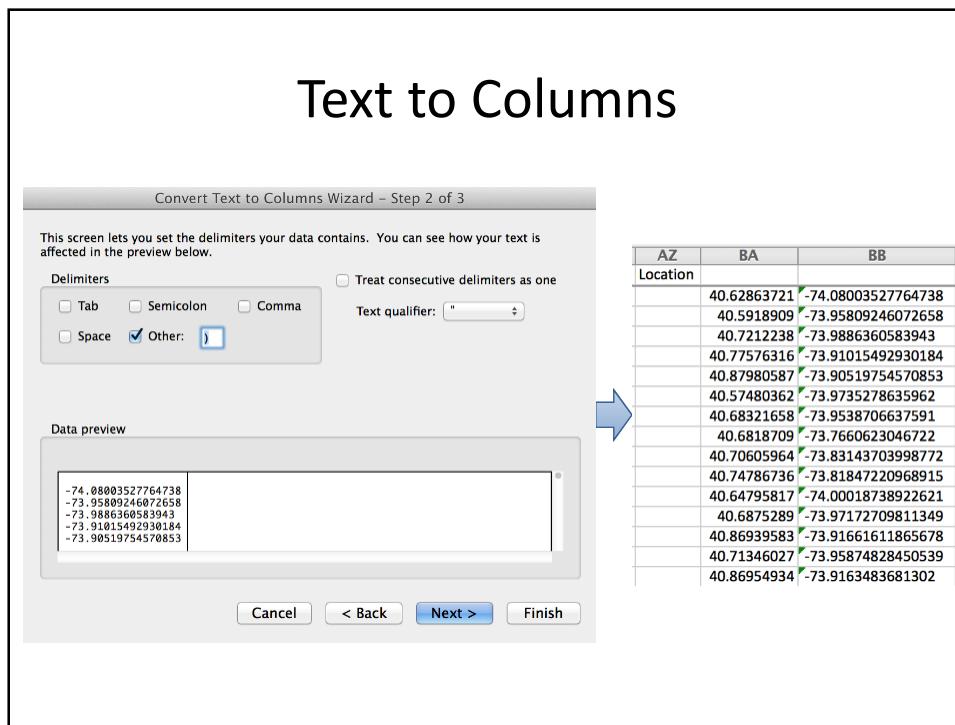
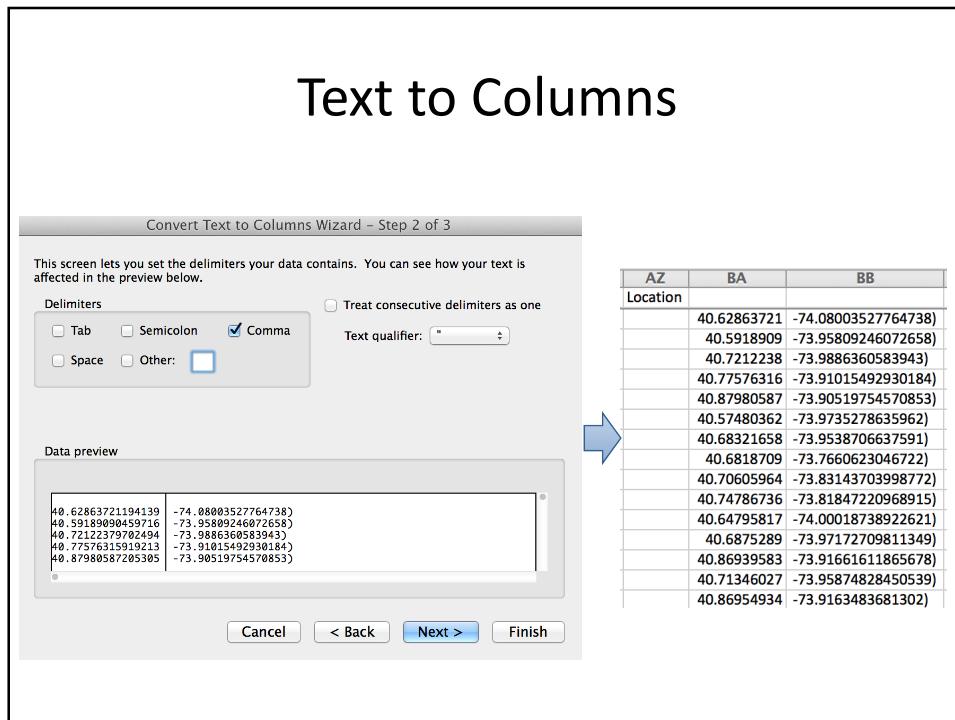
Location

40.62863721194139, -74.08003527764738
40.59189090459716, -73.95809246072658
40.72122379702494, -73.9886360583943
40.77576315919213, -73.91015492930184
40.87980587205305, -73.90519754570853
40.87980587205305, -73.90519754570853

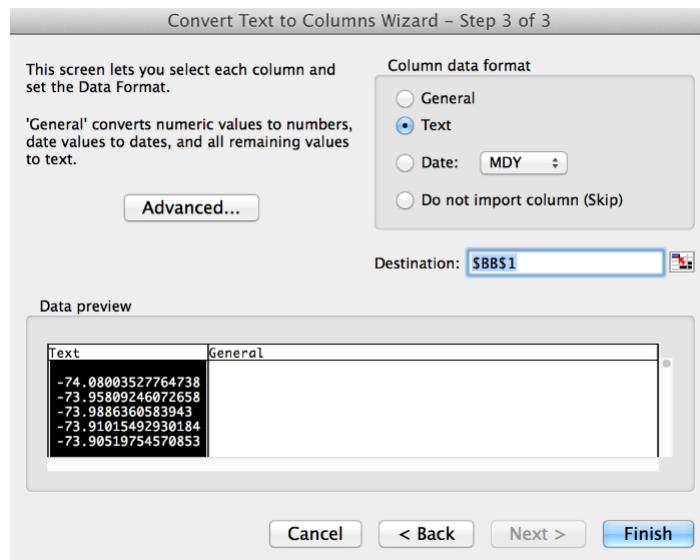
AZ BA

Location	40.62863721194139, -74.08003527764738
	40.59189090459716, -73.95809246072658
	40.72122379702494, -73.9886360583943
	40.77576315919213, -73.91015492930184
	40.87980587205305, -73.90519754570853
	40.87980587205305, -73.90519754570853

Cancel < Back Next > Finish



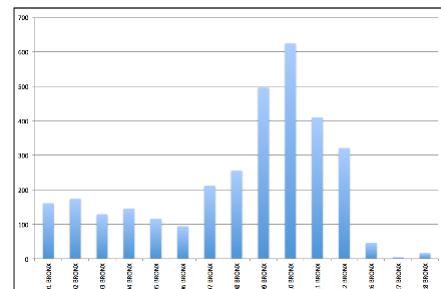
Text to Columns – Format as Text



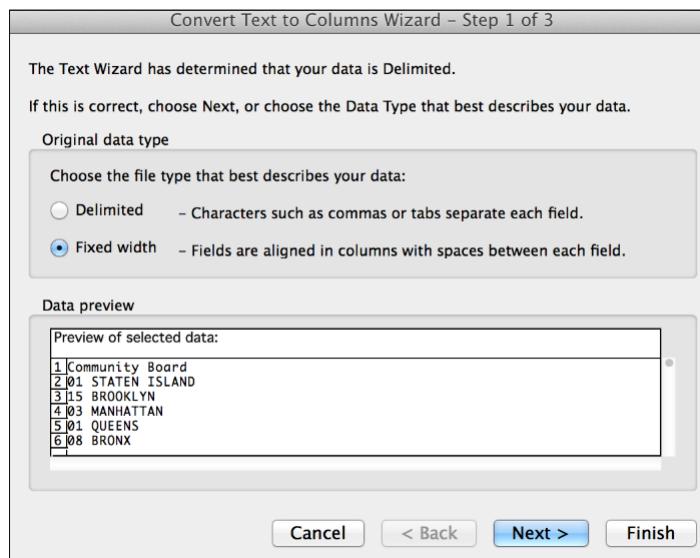
Splitting Data

Z	AA
Community Board	Borough
01 STATEN ISLAND	STATEN ISLAND
15 BROOKLYN	BROOKLYN
03 MANHATTAN	MANHATTAN
01 QUEENS	QUEENS
08 BRONX	BRONX
13 BROOKLYN	BROOKLYN
03 BROOKLYN	BROOKLYN
12 QUEENS	QUEENS
09 QUEENS	QUEENS
07 QUEENS	QUEENS
07 BROOKLYN	BROOKLYN
02 BROOKLYN	BROOKLYN
12 MANHATTAN	MANHATTAN
01 BROOKLYN	BROOKLYN
12 MANHATTAN	MANHATTAN
02 MANHATTAN	MANHATTAN
07 MANHATTAN	MANHATTAN

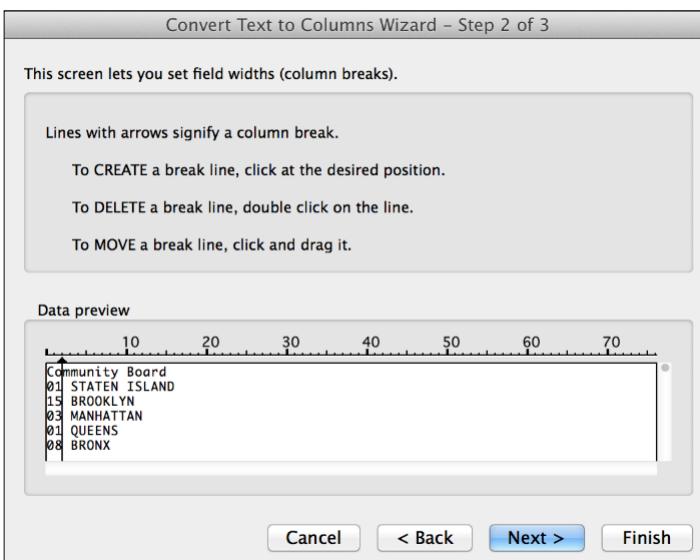
- What if we just wanted the Community Board numbers?



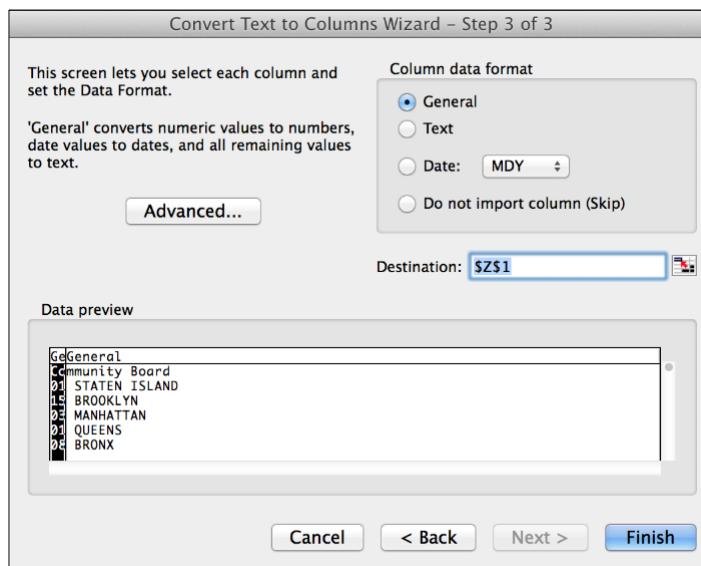
Text to Columns



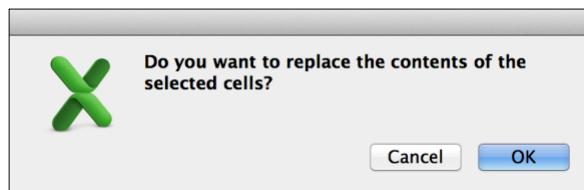
Text to Columns



Text to Columns



Text to Columns



MID Function

MID, MIDB functions

This article describes the formula syntax and usage of the **MID** and **MIDB**function in Microsoft Excel.

Description

MID returns a specific number of characters from a text string, starting at the position you specify, based on the number of characters you specify.

MIDB returns a specific number of characters from a text string, starting at the position you specify, based on the number of bytes you specify.

IMPORTANT MID is intended for use with languages that use the single-byte character set (SBCS), whereas MIDB is intended for use with languages that use the double-byte character set (DBCS). The default language setting on your computer affects the return value in the following way:

- MID always counts each character, whether single-byte or double-byte, as 1, no matter what the default language setting is.
- MIDB counts each double-byte character as 2 when you have enabled the editing of a language that supports DBCS and then set it as the default language. Otherwise, MIDB counts each character as 1.

The languages that support DBCS include Japanese, Chinese (Simplified), Chinese (Traditional), and Korean.

<https://support.office.com/en-in/article/MID-MIDB-functions-61cf6b06-bdf0-4ff8-9319-60e31f3f5884>

MID Function

Z	AA
Community Board	CB_Number
01 STATEN ISLAND	=MID(Z2,1,2)
15 BROOKLYN	
03 MANHATTAN	
01 QUEENS	
08 BRONX	
13 BROOKLYN	

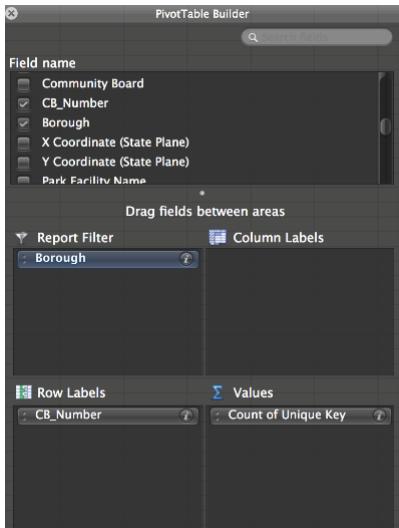
MID Function

Z	AA
Community Board	CB_Number
01 STATEN ISLAND	01
15 BROOKLYN	
03 MANHATTAN	
01 QUEENS	
08 BRONX	
13 BROOKLYN	

MID Function

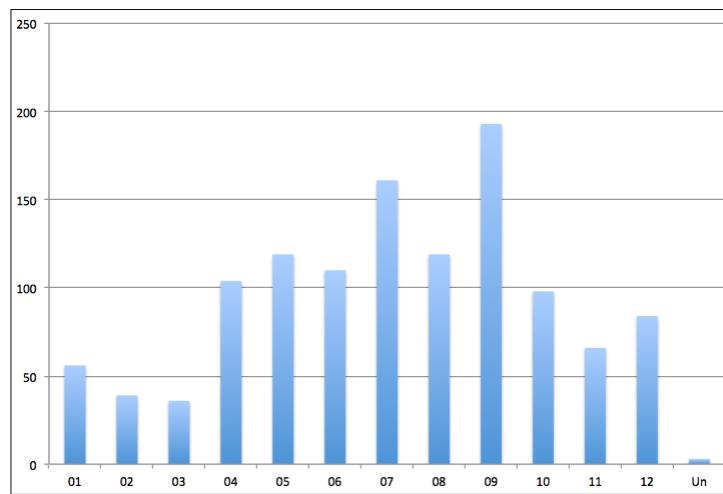
Z	AA
Community Board	CB_Number
01 STATEN ISLAND	01
15 BROOKLYN	15
03 MANHATTAN	03
01 QUEENS	01
08 BRONX	08
13 BROOKLYN	13

Noise Complaints by Community Board



Borough	BRONX
Count of Unique Key	
Row Labels	Total
01	56
02	39
03	36
04	104
05	119
06	110
07	161
08	119
09	193
10	98
11	66
12	84
Un	3
Grand Total	1188

Noise Complaints in the Bronx by Community Board



Question Driven Analysis

Vision Zero (dB)

Tasks:

- Given 311 noise complaint data, assist enforcement efforts by identifying community districts that have a high volume of noise complaints and the time frame enforcement resources should be deployed to combat the noise issue at its peak
- Identify the prevalent types of noise complaints in these areas to guide enforcement in each community district

NYC Community Districts



CONCATENATE Function

- Useful for combining text fields
- Basic Syntax
=CONCATENATE(text1, [text2], ...)
- Example

A	B	C	D
First Name	Last Name	Full Name	
Jane	Doe	=CONCATENATE(A2,B2)	
A	B	C	D
First Name	Last Name	Full Name	
Jane	Doe	JaneDoe	

CONCATENATE Function

- Useful for combining text fields
- Basic Syntax
=CONCATENATE(text1, [text2], ...)
- Example

A	B	C	D
First Name	Last Name	Full Name	
Jane	Doe	=CONCATENATE(A2," ",B2)	
A	B	C	D
First Name	Last Name	Full Name	
Jane	Doe	Jane Doe	

Preparing the Population Data

	A	B
1	COMMUNITY DISTRICT NUMBER	
2	BRONX COMMUNITY DISTRICTS	
3	1	=CONCATENATE("0",A3," BRONX")
4	2	
5	3	
6	4	
7	5	

Preparing the Population Data

	A	B
1	COMMUNITY DISTRICT NUMBER	
2	BRONX COMMUNITY DISTRICTS	
3	1	
4	2	01 BRONX
5	3	
6	4	
7	5	

Preparing the Population Data

1	COMMUNITY DISTRICT NUMBER
2	BRONX COMMUNITY DISTRICTS
3	1
4	2
5	3
6	4
7	5
8	6
9	7
10	8
11	9
12	10
13	11
14	12

Preparing the Population Data

12	10	=CONCATENATE(A12," BRONX")
13	11	011 BRONX
14	12	012 BRONX
12	10	10 BRONX
13	11	011 BRONX
14	12	012 BRONX
12	10	10 BRONX
13	11	11 BRONX
14	12	12 BRONX

Nested Functions



=CONCATENATE(IF(LEN(A3)<2,"0","",A3," BRONX"))

Test whether
the length of
the number in
A3 is less than 2
(example “1”)

```
=CONCATENATE(  
    IF(  
        LEN(A3)<2,  
        "0",  
        "..."  
    ),  
    A3,  
    " BRONX"  
)
```

Nested Functions



=CONCATENATE(IF(LEN(A3)<2,"0","",A3," BRONX"))

If the length of
the number is
less than 2, add
a zero
If it isn’t, add an
empty space

```
=CONCATENATE(  
    IF(  
        LEN(A3)<2,  
        "0",  
        "..."  
    ),  
    A3,  
    " BRONX"  
)
```

Nested Functions

```
=CONCATENATE(IF(LEN(A3)<2,"0","",A3," BRONX"))
```

- This is an example of nesting functions
- How many functions do we use?

```
=CONCATENATE(  
    IF(  
        LEN(A3)<2,  
        "0",  
        "")  
    ),  
    A3,  
    " BRONX"  
)
```

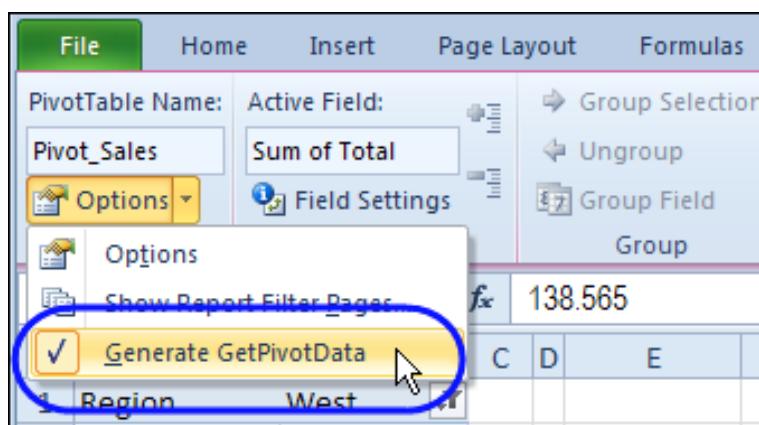
VLOOKUP

- Useful for referencing data from another sheet (reference tables)
- Basic Syntax
`=VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])`
- Example
`=VLOOKUP(A2,SHEET2$A1:$D500,4,TRUE)`

Adding the VLOOKUP

	A	B	C
Row Labels	Total	CB Population	
0 Unspecified	60		
01 BRONX	56		
01 BROOKLYN	1023		
01 MANHATTAN	567		
01 QUEENS	666		
01 STATEN ISLAND	242		

Turn Off the GetPivotData “Feature”



Adding the VLOOKUP

```
(fx) =VLOOKUP(A6,'[NYC_Population_1970-2010.xlsx]A'!$C$1:$I$81,7, FALSE)
```

```
1 =VLOOKUP(A6,  
2   ' [NYC_Population_1970-2010.xlsx]A'!$C$1:$I$81,  
3   7,  
4   FALSE)
```

- The value to be looked up (the Community Board Number)

Adding the VLOOKUP

```
(fx) =VLOOKUP(A6,'[NYC_Population_1970-2010.xlsx]A'!$C$1:$I$81,7, FALSE)
```

```
1 =VLOOKUP(A6,  
2   ' [NYC_Population_1970-2010.xlsx]A'!$C$1:$I$81,  
3   7,  
4   FALSE)
```

- The value to be looked up (the Community Board Number)
- The lookup range
 - File -> “NYC_Population_1970-2010.xlsx”
 - Sheet name -> “A”
 - Separator -> “!”
 - Cell Range -> “\$C\$1:\$I\$81”

Adding the VLOOKUP

```
(fx) =VLOOKUP(A6,'[NYC_Population_1970-2010.xlsx]A'!$C$1:$I$81,7, FALSE)
```

```
1 =VLOOKUP(A6,  
2   ' [NYC_Population_1970-2010.xlsx]A'!$C$1:$I$81,  
3     7,  
4   FALSE)
```

- The value to be looked up (the Community Board Number)
- The lookup range
- The column to be returned (“Total Population 2010”)

Adding the VLOOKUP

```
(fx) =VLOOKUP(A6,'[NYC_Population_1970-2010.xlsx]A'!$C$1:$I$81,7, FALSE)
```

```
1 =VLOOKUP(A6,  
2   ' [NYC_Population_1970-2010.xlsx]A'!$C$1:$I$81,  
3     7,  
4   FALSE)
```

- The value to be looked up (the Community Board Number)
- The lookup range
- The column to be returned (“Total Population 2010”)
- Whether to use fuzzy matching (No)

Adding the VLOOKUP

Count of Unique Key	Total	CB Population
Row Labels		
0 Unspecified	60	
01 BRONX	6	91497
01 BROOKLYN	1023	
01 MANHATTAN	567	
01 QUEENS	666	
01 STATEN ISLAND	242	

Adding the VLOOKUP

Count of Unique Key	Total	CB Population
Row Labels		
0 Unspecified	60	
01 BRONX	6	91497
01 BROOKLYN	1023	173083
01 MANHATTAN	567	60978
01 QUEENS	666	191105
01 STATEN ISLAND	242	175756

Calculating Complaints Per 1000 People

Count of Unique Key	Total	CB Population	Complaints_per_1000
Row Labels			
0 Unspecified	60		
01 BRONX	56	91497	=B6/(C6/1000)
01 BROOKLYN	1023	173083	
01 MANHATTAN	567	60978	
01 QUEENS	666	191105	
01 STATEN ISLAND	242	175756	

Calculating Complaints Per 1000 People

Count of Unique Key	Total	CB Population	Complaints_per_1000
Row Labels			
0 Unspecified	60		
01 BRONX	56	91497	0.612041925
01 BROOKLYN	1023	173083	
01 MANHATTAN	567	60978	
01 QUEENS	666	191105	
01 STATEN ISLAND	242	175756	

Calculating Complaints Per 1000 People

Count of Unique Key Row Labels	Total	CB Population	Complaints_per_1000
0 Unspecified	60		
01 BRONX	56	91497	0.612041925
01 BROOKLYN	1023	173083	5.910459144
01 MANHATTAN	567	60978	9.298435501
01 QUEENS	666	191105	3.48499516
01 STATEN ISLAND	242	175756	1.376908896

Complaints Per 1000 People

Count of Unique Key Row Labels	Total	CB Population	Complaints_per_1000
0 Unspecified	60		
01 BRONX	56	91497	0.612041925
01 BROOKLYN	1023	173083	5.910459144
01 MANHATTAN	567	60978	9.298435501
01 QUEENS	666	191105	3.48499516
01 STATEN ISLAND	242	175756	1.376908896
02 BRONX	39	52246	0.746468629
02 BROOKLYN	513	99617	5.149723441
02 MANHATTAN	1139	90016	12.65330608
02 QUEENS	344	113200	3.038869258
02 STATEN ISLAND	68	132003	0.515139807
03 BRONX	36	79762	0.451342745
03 BROOKLYN	541	152985	3.536294408
03 MANHATTAN	1143	163277	7.000373598
03 QUEENS	240	171576	1.398797035
03 STATEN ISLAND	102	160209	0.636668352
04 BRONX	104	146441	0.710183623
04 BROOKLYN	325	112634	2.885451995
04 MANHATTAN	1096	103245	10.61552618
04 QUEENS	85	172598	0.492473841

Macros

- A way to record frequently used operations in order
- Automates data processing tasks
- Saves time and increases efficiency
- Easy to record and use a macro
- For more, see <http://www.excelfunctions.net/Record-A-Macro.html>

Review

- Using basic mathematical functions in Excel
- Creating charts
- Conditional formatting
- PivotTables
- Designing visualizations
- Using date/time functions
- Splitting and concatenating text
- VLOOKUPs

Bytes of the Big Apple Population Data

The screenshot shows the NYC Planning Department's website under the "Office of the Mayor". The main page title is "Reference > Population > 2010 Census". Below it, a section titled "POPULATION 2010 Demographic Tables" is displayed. A table titled "Demographic Tables" lists four categories: "NYC & Boroughs", "Census Tracts", "Community Districts", and "NTAs* / Other Geographies". Under each category, there are two download links: PL-P1 NYC, PL-P1 CT, PL-P1 NTA, and PL-P1 CD. The "PL-P1 CD" link in the "Community Districts" row is highlighted with a red box.

http://www.nyc.gov/html/dcp/html/census/demo_tables_2010.shtml

NYC Open Data Portal

The screenshot shows the NYC Open Data Portal. At the top, it displays "1100+ Datasets Available". Below the header, there are filters for "All Types" and "All Categories", and a "Hide Tiles" button. The main area is titled "Data Catalog" and contains a search bar with the placeholder "Search" and a magnifying glass icon. To the right of the search bar are icons for sorting by popularity and relevance. A list of datasets is shown, each with a checkbox, name, popularity (number of views), and type. The datasets listed are:

Name	Popularity	Type
1. Wifi Hotspot Locations	70,224 views	Map
2. 311 Service Requests from 2010 to Present	42,221 views	Table
3. Subway Entrances	40,868 views	Map
4. MTA Data	14,794 views	Table
5. Restaurant Inspection Results	25,724 views	Table

At the bottom of the catalog area, there is a link: <https://nycopendata.socrata.com/data>

NYC Open Data Portal – 311 Data

NYC OpenData 1100+ Datasets Available

311 Service Requests from 2010 to Present
All 311 Service Requests from 2010 to present. This information is ...

Unique Key	Created Date	Closed Date
1 29097371	10/19/2014 02:57:13 AM	
2 29098073	10/19/2014 02:29:15 AM	
3 29096227	10/19/2014 02:13:44 AM	
4 29096249	10/19/2014 02:13:30 AM	
5 29094817	10/19/2014 02:13:26 AM	10/19/2014 02:13:26 AM
6 29093575	10/19/2014 02:09:32 AM	
7 29095900	10/19/2014 02:06:57 AM	10/19/2014 02:06:57 AM
8 29094006	10/19/2014 02:06:00 AM	
9 29097010	10/19/2014 02:05:21 AM	
10 29097687	10/19/2014 02:05:20 AM	
11 29094997	10/19/2014 02:04:52 AM	
12 29096621	10/19/2014 02:04:31 AM	10/19/2014 02:04:31 AM
13 29094357	10/19/2014 02:02:08 AM	10/19/2014 02:02:08 AM

Filter

Conditional Formatting

Sort & Roll-Up

Filter

Filter this dataset based on contents.

Unique Key is

+ Add a New Filter Condition

Never created a filter before? Watch a short tutorial video [here](#).

Download 311 Data

Filter

Filter this dataset based on contents.

Created Date is between
 06/01/2015 12:00:00 AM and
 08/30/2015 11:59:59 PM
 and
 options

Complaint Type contains
 noise
 and
 options

Export

SODA API

OData

Print

Download

Download a copy of this dataset in a static format

Download As

CSV

CSV for Excel

JSON

PDF

RDF

RSS

XLS

XLSX

XML

OpenOffice and LibreOffice

- Free and open-source office suites of software
- Fully featured with similar functionality to Microsoft Office
- And they're free
- <http://www.openoffice.org/>
- <https://www.libreoffice.org/>

OpenRefine (Google Refine)

- Free and open-source data cleaning tool
- Works in the browser
- Works with lots of data types
 - Great for converting between Excel files, CSV, and JSON
- Facets are powerful (facet on anything)
- Meant specifically to clean data
- Available at: <http://openrefine.org/>

Python

- Computer language useful for working with data
- Easy to learn syntax for simple operations
- Lots of mathematical and scientific packages for advanced analysis
- Visualization packages for creating charts and graphs
- Can install on city computers
- Distribution available at:
<https://www.continuum.io/why-anaconda>

Contact Information

Instructor

- Name: Richard Dunks
- Email: richard@datapolitan.com
- Website: <http://www.datapolitan.com>
- Blog: <http://blog.datapolitan.com>
- Twitter: @datapolitan

TA

- Name: Mark Yarish