# Data Visualization (Iconographics) and Statistics 

## Microsoft Excel 2013

## Instructor: Don Bremer

Presented and co-sponsored by:


University of Minnesota Duluth
Driven to Discover

11 East Superior Street, Suite 210
Duluth, MN 55802
218-726-7298 (main)
888-387-4594 (toll free)
www.umdced.com
AMERICATS
ced@umdced.com
MINNESOTA

## Copyright® 2017 University of Minnesota Duluth Center for Economic Development. All Rights Reserved.

Reproduction of this printed material in whole or in part without permission is prohibited.
Microsoft and Access are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

The University of Minnesota Duluth is an equal opportunity employer and educator.

The Mission of the University of Minnesota Duluth Center for Economic Development is to assist entrepreneurs and businesses to grow and succeed.

The Minnesota Small Business Development Center program is funded in part through a Cooperative Agreement with the U.S. Small Business Administration, Minnesota Department of Employment and Economic Development and regional support partners. All opinions, conclusions or recommendations expressed are those of the author(s) and do not necessarily reflect the program sponsors. Programs are open to the public on a nondiscriminatory basis. Reasonable accommodations for persons with disabilities will be made if requested at least two weeks in advance.

## Data Visualization and Iconographics

## Contents

Data Visualization and Iconographics ..... 4
Why We Visualize ..... 4
Types of Data ..... 5
Qualitative (Attributes) ..... 5
Quantitative (Metrics) ..... 5
What different types of graphs and where to use them ..... 9
Bar Graphs ..... 9
Line Graphs ..... 11
Pie Charts ..... 13
Create your own Pie Chart ..... 13
Pie or Bar Graph? ..... 14
Pie Charts \#2 ..... 14
Histograms ..... 14
Create your own Histogram ..... 15
Scatterplot ..... 18
Create your own Scatterplot ..... 18
Bubble Graphs ..... 19
Create your own Bubble Graph ..... 19
Pareto Graph ..... 21
Tips on sprucing up graphs in Excel ..... 22
Sort bar graph data before designing. ..... 22
Remove background lines ..... 22
Stay away from 3D effects ..... 23
Pivot Tables ..... 24
Using Power Map ..... 26
Create a population and clustered bar graph ..... 29
Statistics \& Other Mind Blowing Items ..... 30
Correlation ..... 30
Descriptive Analysis ..... 31
Moving Average ..... 33
Regression ..... 35
How to read it ..... 36
Significance F and P -values ..... 36
Coefficients ..... 37
Residuals ..... 37
Goal Seek ..... 37
Data solver ..... 39

## Data Visualization and Iconographics

For good data visualization, it must be both be interesting (meaningful \& relevant) and have integrity (accuracy and consistency).

What Makes Good Information Design?



Things to consider for the right type of visualization:

- What kind of relationship are you exploring?
- How many variables?
- What can you simplify?
- Compare your data to similar visualizations


## Why We Visualize

Good Visualizations Should Make Data Actionable

| 4 | 4 | B | c | D | E | F | 6 | H | 1 | 1 | K |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1. | Fico Range | 660-678 |  | 679-713 |  | 714.749 |  | 750-779 |  | $780+$ |  |
| 21 | Loan Purpose | Avg Loan Amount | Avg interest Rate | Avg Loan Amount | Avg interest Rate A | Avg Loan Amount A | Avg finterst Rate | Avg Loan Amount | Avg interest hate | Avg Loan Amount | vEinterest Hate |
| 3 C | Car | \$5,826 | 15.4\% | S6,334 | 129\% | \$6,761 | \$.7\% | \$7,020 | 8.2\% | \$7,131 | 7.4\% |
| 4 | credit card | \$10,073 | 16.0\% | \$10,636 | 13.2\% | \$11,937 | 10.0\% | \$11,464 | 8.2\% | \$10.670 | 7.4\% |
| 5 | Debt Consolidation | \$10,143 | 16.5\% | \$11,557 | 13.8\% | \$13,211 | 10.6\% | \$13,276 | 8.6\% | \$12,899 | 8.085 |
| 6 | Education | \$3,734 | 14.8\% | \$5,153 | 12.7\% | \$7,185 | 10.4\% | \$7,689 | 9.5\% | \$4,913 | 8.6\% |
| 7 | Home limprovement | \$8,551 | 16.2\% | \$9,434 | 13.6\% | \$11,039 | 10.6\% | \$12,206 | 8.9\% | \$12,974 | 8.4\% |
| 8 | House | \$11,665 | 17.6\% | \$11,297 | 14.5\% | \$13,300 | 114\% | \$12,778 | 9.2\% | \$15,334 | 9.2\% |
| 3 | Major Purchase | \$6,442 | 15.7\% | 57,273 | 13.3\% | \$7,603 | 9.756 | \$8,536 | 8.1\% | 59,22a | 7.6\% |
| 10 | Medical | \$7,115 | 16.0\% | \$6,825 | 13.0\% | S8,281 | 9.9\% | S9,585 | 8.6\% | 99,925 | 7.8\% |
| 11 | Moving | \$5,124 | $16.0 \%$ | 55,492 | 127\% | \$6,999 | $9.7 \%$ | \$7,477 | 8.0\% | \$10,009 | 7.9\% |
| 12 | Other | \$6,151 | 15.9\% | \$7,082 | 132\% | 58,125 | 10.1\% | \$S.711 | 8.6\% | \$10,725 | 8.3\% |
| 13. | Renewable Energy | \$6,472 | 16.5\% | \$6,995 | 13.1\% | \$7,296 | 9.7\% | \$10,540 | 8.2\% | \$9,202 | 7.4\% |
| 14 | Small Business | \$10,071 | 17.4\% | \$11,091 | 14.8\% | \$13,108 | 120\% | \$13,898 | 10.2\% | \$14,099 | 9.6\% |
| 15 | Vacation | \$4,458 | 15.5\% | 54,544 | 12.1\% | \$5,324 | 9.15 | \$6,969 | 7.9\% | \$8,178 | 6.5\% |
| 16 | Wedding | \$8,002 | 15.9\% | \$9,028 | 13.4\% | \$9,589 | 10.1\% | \$11,219 | 8.6\% | \$11,414 | 7.9\% |



Which is better, all the numbers or the graphics? What story does this tell?
Types of Data
Qualitative (Attributes)

- Nominal
- Ordinal

Quantitative (Metrics)

- Numeric


## Nominal Attributes

Data that be counted, but not ordered or aggregated (grouped into classes or clusters).
Examples:

- Products - Books, Movies, Music
- Gender - Male, Female
- State - Virginia, Nevada, California

What are some for your data?

## Ordinal Attributes

Data that can be counted and ordered, but not aggregated
Examples:

- Date $-1 / 1 / 2014,1 / 2 / 2014 .$.
- Grades - A, B, C...
- Ranks - Like, Neutral, Dislike


## What are some for your data?

## Metrics

Quantitative data that can be counted, ordered, and aggregated.

## Examples:

- Revenue, Cost, Profit
- Number of Customers
- Temperature
- Time


## What are some for your data?

Ordinal Attributes and Metrics
Some data can be used as either attributes or metrics. Their classification is dependent on usage. Examples:

- Age
- Scores


## What are some for your data?

Visualizations

|  | Metric | Attribute <br> (Ordinal) |
| :--- | :--- | :--- |
| Attribute (Nominal) | Bar <br> Heatmap | Line (with Groups) <br> Bar (with Groups) |
| Attribute (Ordinal) | Column <br> Line | Scatter Grid |
| Metric | Scatter/ Bubble |  |

Appropriate Visual Enhancements

|  | Attribute (Nominal) | Attribute (Ordinal) | Metric |
| :--- | :---: | :---: | :---: |
| Color Hue | X | X | X |
| Color Saturation |  | X | X |
| Size | X | X |  |

Use the right color scheme and icons for the right situation. Which icons or colors are better in the graphics below:

Factory Production


City Population


Factory Production


City Population


## Use Opposing Colors for Comparisons




A combination of shading and different colors


Building Blocks of Design

- Visual Elements - Create a focal point
- Typography -- Avoid defaults, but keep it simple
- $\quad$ Space - Use negative space (around the edges of design)

Work with natural reading habits.

- Design flow: top to bottom, left to right
- Use font/graphic size/weight to indicate importance
- Use colors/bolding strategically

Limit the number of competing elements

- 2-3 typefaces - derive variety within these typefaces (for you, it will be one!)

Places to go for Data Visualization:

- Google Public Data Explorer


## What different types of graphs and where to use them

To paraphrase Scotty from Star Trek, "The right graph for the right application". Picking the right graph for the situation will do miles of work for you. What are you trying to accomplish? Showing growth? Showing a trend? Showing magnitude? There are graphs for that!

## Bar Graphs

The bars' heights are scaled according to their values and the bars can be compared to each other. Bar graphs can be drawn in a 3-dimensional way and compiled for data comparison about the same thing or location. So that more important categories are emphasized, bars in a bar graphs are arranged in order of frequency.


## Create your own Bar Graph

In this first exercise, we would like to see the magnitude of sales by date.
1.) Open the file Bar Charts
2.) Select from A2:D6
3.) Go to Insert->Insert Column Chart->2D Clustered Column

Boom! We have a chart. That was easy... But, my instructors always said a chart without a title is a graph. Add a chart title by selecting the Chart Layout with a Title on Top. Name the chart "Employee Sales by Month"

Now, it may be obvious what the $y$-axis is (money), but perhaps we should label it and use currency.

1.) Select the data that makes up the graph
2.) Go to the Home tab
3.) Select the \$ for money

To add a title on the axis:
1.) Select the Chart
2.) Select Layout under the Chart
3.) Select on Axis Titles
4.) Select Vertical

5.) Put in "Sales"

Now, this graph is minimal - meaning the minimal we can get away with. Let's make it better. Let's annotate why Peterson's numbers are so low, "He's rarely here"
1.) Select the text box under insert
2.) Draw out a box to put text
3.) Put in the text, "He's Rarely Here"

What if for every 5,000 sales, the sales people get a hat. We can put the hat every 5,000 sales.
1.) Select a data series (like March)
2.) Right-click and select Format Data Series
3.) Select Fill
4.) Select Picture or texture fill
5.) Insert a picture from File
6.) Select the hat
7.) Select Stack and Scale with 5000 units/ picture Other things to do:
1.) Remove Decimals from Sales
2.) Put a color or picture in for the background of the chart
3.) Put a line through the item showing minimum sales


## Line Graphs

Used to display comparisons between 2 variables, line graphs involve an $x$ axis horizontally and a $y$-axis vertically on a grid. Dot-connected and gridplotted lines are what comprise a line graph. These lines monitor and compare various data sets. Usually, the x-axis represents time measurements while the $y$-axis is a representative of measure or percentage of quantity. For this reason, a line graph is used often for tracking variables of one or more subjects in time.


## Create your own Line Graph

Line Graphs are great for showing trends in data. Is the item you are looking at going up or down? Is it staying consistent?

Let's show the stock prices of Apple and Microsoft from Jan 1, 2000 to Jan 1, 2015 (weekly) against the Dow Jones Industrial Average

Open the file: Stock Prices.xls
Wow! That's a lot of data. Let's hide everything except the closing data for all items.
1.) Select the column that we want to hide (e.g. column $B$ )
2.) Right click on the selection giving you a context sensitive menu
3.) Select hide

Voilia! Repeat until it looks like this:
Now, to graph, select the date and the Apple close. We will add Microsoft close

|  | A | E | N | C |
| ---: | ---: | ---: | ---: | ---: |
| 1 |  | Apple | Microsoft |  |
| 2 | Date | Close | Close |  |
| 3 | $12 / 29 / 2014$ | 110.38 | 46.45 |  |
| 4 | $12 / 22 / 2014$ | 113.99 | 47.88 |  |
| 5 | $12 / 15 / 2014$ | 111.78 | 47.66 |  |
| 6 | $12 / 8 / 2014$ | 109.73 | 46.95 |  |
| 7 | $12 / 1 / 2014$ | 115 | 48.42 |  |
| 8 | $11 / 24 / 2014$ | 118.93 | 47.81 |  |
| 9 | $11 / 17 / 2014$ | 116.47 | 47.98 |  |
| 10 | $11 / 10 / 2014$ | 114.18 | 49.58 |  |
| 11 | $11 / 3 / 2014$ | 109.01 | 48.68 |  |
| 12 | $10 / 27 / 2014$ | 108 | 46.95 |  |
| 13 | $10 / 20 / 2014$ | 105.22 | 46.13 |  |
| 14 | $10 / 13 / 2014$ | 97.67 | 43.63 |  |
| 15 | $10 / 6 / 2014$ | 100.73 | 44.03 |  |
| 16 | $9 / 29 / 2014$ | 99.62 | 46.09 |  |
| 17 | $9 / 22 / 2014$ | 100.75 | 46.41 |  |
|  |  |  |  |  | later in the exercise. Use the CTRL+Shift+Down Arrow to grab everything quickly!



While selected, Insert->Graph->Line 2-D
Wow! That was easy! Let's add Microsoft.
1.) Select the chart
2.) Go to Chart Tools->Design-> And in the data section, choose Select Data.

3.) Click the Add button on the Legend Entries
4.) Type Microsoft into the Series name and select the series values

Awesome. Perhaps we should Title the chart and annotate some big dates:
Title: Microsoft vs. Apple
Dates:
iphone release: June 29, 2007
ipad release: April 3, 2010
windows 7 release: July 22, 2009
Not bad, I would call this a minimum chart.
Perhaps there is too much white space on top - tell the vertical axis to only go to 700 .
1.) Click the Axis
2.) Right-click and select Format Axis...
3.) Change maximum to 700


Now, let's add the DJIA information. When we do that, notice that everything is messed up. We are less than 100 for Microsoft and over 17,000 for the DJIA. So, let's put the DJIA on the secondary access. Some cleanup may be necessary.


Page 12 of 42

## Pie Charts

These charts represent the parts of a whole. Each 'section' or 'slice of the pie is a data percentage. From biggest to smallest, segments are arranged in a clockwise formation. This way, the pie chart features easy-to-compare subjects presented in a neat, easy-to-understand way.

## Create your own Pie Chart

1.) Open the file pie chart.
2.) Select the names in column $B$
3.) Using your CTRL key, select the gross pay in column E
4.) To insert the graph, select insert pie, 2 D


Again, not too hard. But, I don't like the color scheme. You can change that using the Gallery's selection on the top of the page.

Also, let's put the series name over the top of the pie pieces to make it easier to read along with their percentages.

OK, again I would call this the minimum acceptable chart. Perhaps we should draw our eyes to Smith since he is getting paid so much more than the rest of us. To do this:
1.) Click once on Smith's pie piece. They are now all selected

2.) Click again on Smith's pie piece. Now only Smith is selected.
3.) Drag Smith's pie piece out
4.) Re-color the pie piece to hot pink.

This time, let's leave the background white
A graph I found incorporating the colors and logo of Toys R Us.



## Pie or Bar Graph?

Remember this simple rule when deciding to use a pie chart or a bar graph:

- If you are comparing a percentage of a piece of the whole segment, use a pie chart.
- If you are comparing a fixed numbers or a trend over time, use a bar graph.

Consider the main point you are trying to convey. Your chart or graph should make one point, vividly.
Pie Charts \#2
Let's create some pie charts using the Business Indicators from the last report.
Using the BI from the previous 6 months, let's create a graph over the average hours worked:

Average hours worked, previous 6 months


■ Significantly Decreased ■ Moderately Decreased ■ No Change

- Moderately Increased $\quad$ Significantly Increased


When quantitative data is what you have, a histogram would be used to show it. This is a kind of graph that also uses bars. Ranges of values are listed at the bottom and these are called 'classes.' Taller bars represent the classes with greater frequencies.

## Create your own Histogram

Even though it seems that the makers of Excel put the kitchen sink in and ready to go - not everything is visible. There is another statistical package that is an Add-In if you want to go even further.

File->Options->Add-Ins->Manage Excel Add-ins (Go...)
Select on Analysis ToolPak and click OK.
A new item appears under data called "Data Analysis". When the item is selected, it shows the different things you can do:



## Open the file Histogram

Take for example, how many students in grade 9 are a certain height? How is this data distributed?

So, click on Data Analysis, then Histogram:

Data Visualization and Iconographics


The input range is the all of the student's heights. The bin range contains range of items want to throw them in. I made a container every 2 inches from below the minimum height.

The Output gives the possibility of putting it on the sheet, in a new worksheet, or a new workbook. I usually pick the new worksheet.

It also gives the ability to add:

- Pareto charts
- Cumulative Percentages
- Chart Output

| Bin | Frequency |
| ---: | ---: |
| 56 | 0 |
| 58 | 0 |
| 60 | 2 |
| 62 | 2 |
| 64 | 5 |
| 66 | 5 |
| 68 | 4 |
| 70 | 4 |
| 72 | 2 |
| 74 | 1 |
| More | 0 |



Which, when graphed, makes it much easier to understand what is going on:


## Scatterplot

Scatterplots display paired data using the vertical or the $y$ axis and a horizontal axis or the x axis. The tools for statistics called correlation and regression are then used for showing trends on this type of graph.

## Create your own Scatterplot

Open the file scatterplot

1. Select the data from C4:D15
2. Go to Insert->Scatter-> 2D plot
3. Add Titles and Labels
4. Profit!


Make sure the $x$-axis values are on the right of the table. This is default in Excel.


An awesome feature in excel is to get the trendline!

1. Select the chart
2. Select Chart Tools->Layout->Trendline
3. You, as the expert, will have to say what type of trendline the data is representing.
4. I also recommend getting the equation on the chart


## Bubble Graphs

A bubble chart is a variation of a scatter chart in which the data points are replaced with bubbles, and an additional dimension of the data is represented in the size of the bubbles. Just like a scatter chart, a bubble chart does not use a category axis both horizontal and vertical axes are value axes. In addition to the $x$ values and $y$ values that are plotted in a scatter chart, a bubble chart plots $x$ values, $y$ values, and $z$ (size) values.


## Create your own Bubble Graph

Bubble charts are often used to present financial data. Different bubble sizes are useful to visually emphasize specific values.

To create a bubble chart, arrange your data in rows or columns on a worksheet so that x values are listed in the first row or column and corresponding y values and bubble size (z) values are listed in adjacent rows or columns.

Open the file bubblechart:

| Number of products | Sales | Market Share $\%$ |
| :---: | :---: | :---: |
| 14 | $\$ 12,200$ | $15 \%$ |
| 20 | $\$ 60,000$ | $33 \%$ |
| 18 | $\$ 24,400$ | $10 \%$ |
| 22 | $\$ 32,000$ | $42 \%$ |

In this bubble chart, the number of products is displayed along the horizontal axis, the sales amounts are displayed along the vertical axis, and the market share percentages are represented by the size of the bubbles.

Consider using a bubble chart when your data includes the following:

- Three values per data point : Three values are required for each bubble. These values can be in rows or columns on the worksheet, but they must be in the following order: $x$ value, $y$ value, and
 then z value.
- Multiple data series : Plotting multiple data series in a bubble chart (multiple bubble series) is similar to plotting multiple data series in a scatter chart (multiple scatter series). Scatter charts use sets of $x$ values and $y$ values, but bubble charts use sets of $x$ values, $y$ values, and $z$ values.

When you create a bubble chart, you can choose one of the following bubble chart subtypes.

- Bubble or bubble with 3-D effect : Both bubble chart types compare sets of three values instead of two. The third value determines the size of the bubble marker. You can choose to display bubbles in 2-D format or with a 3-D effect.
1.) Select only the numbers of products, sales, and market share
2.) Select Insert->Other Charts-> Bubble

Update accordingly!

## Pareto Graph

Many people have heard of this as the " $80-20$ rule". That is, that doing $20 \%$ of the work can generate $80 \%$ of the advantage of doing the entire job. In the idea of quality improvement, we can say that a large majority of the problems (80\%) are produced by a few key issues (20\%).

## Open the File Pareto

1.) Sort rows in decreasing order of importance of the causes (i.e., the most important cause first)
2.) Sum the numbers together

Add a cumulative percentage column to the table
Add a new column to the table and get the percentages of each of the complaints. This is done by taking the quantity of each complaint, dividing it by the total number of complaints, and multiplying by 100 .

For example, hospital cafeteria food (bad) has 159 complaints. So $159 / 557^{*} 100=28.5 \%$

| Patient Complaints for the Month | Quantity | Percent |
| :--- | :---: | ---: |
| Hospital cafeteria food bad | 159 | $28.5 \%$ |
| Waiting room overcrowded | 108 | $19.4 \%$ |
| Walk-up clinic not open Saturday morning | 75 | $13.5 \%$ |
| No parking available in the parking ramp | 56 | $10.1 \%$ |
| Coffee cold in waiting room | 44 | $7.9 \%$ |
| Appointment scheduled for different day | 34 | $6.1 \%$ |
| Pharmacy orders take too long to be filled | 28 | $5.0 \%$ |
| Doctor unfriendly during appointment | 17 | $3.1 \%$ |
| Medical file errors | 16 | $2.9 \%$ |
| Unfriendly receptionist | 9 | $1.6 \%$ |
| Wheel chair unavailable | 8 | $1.4 \%$ |
| Old magazines in waiting room | 3 | $0.5 \%$ |
| Name misspelled on medical chart | 0 | $0.0 \%$ |

3.) Plot with numbers on the left and cumulative percentages on the secondary $y$ axis.


## Tips on sprucing up graphs in Excel

## Sort bar graph data before designing.

If you're using a bar graph to display your data, this tip can make a big difference. Most bar graphs look like the one below.


They're kinda random. You spend just a fraction of a second too long figuring out which ones are outliers. Instead, you should reorder your data points to go from largest to smallest. Here's what that looks like:


If your bar graph is horizontal, larger values should be at the top. If your bar graph is vertical, order them from left to right. Why? That's how people read English (if you're presenting this data in another language where that isn't true, change up your order to better reflect reading patterns).

To order the graphs in Excel, you'll need to sort the data from largest to smallest. Click 'Data,' choose 'Sort,' and select how you'd like to sort everything.

## Remove background lines

Graphs allow you to roughly compare data within a set, not dig into it. No one's looking at your graph to see incremental differences between data points -- they want to see general, overarching trends.

To help people focus on those trends, remove the lines in the background of your graph or chart. These lines are superfluous, unhelpful, and distracting -- cut them from your graph to help people focus on the big takeaways.

Stay away from 3D effects
This comes from my time teaching 5S/Lean/6 Sigma things. I found out that (occasionally), making items 3D may give a different impression than the graph is actually portraying (especially on pie charts) because of the parallax making the front items look bigger.


If you actually look at the area each section takes up on the screen, you'll see why it's easy to misinterpret 3D graphs:


## Pivot Tables

A PivotTable report is an interactive table that you can use to quickly summarize large amounts of data. You can rotate its rows and columns to see different summaries of the source data, filter the data by displaying different pages, or display the details for areas of interest You can start a Pivot table by opening the Pivot Table toolbar and clicking on the wizard.

Open Sorting Lists.xls
Independent Variable - a variable (often denoted by $x$ ) whose variation does not depend on that of another.

Dependent Variable - a variable (often denoted by y) whose value depends on that of another.

1. Click anywhere in the data
2. Insert->Pivot Table
3. Click OK


Drag and drop Independent fields on "Drop Column Fields" and "Drop Row Fields"
Drop Dependent fields on Drop Value Fields Here

| Sum of Quantity | Month |  |  |  |
| :--- | ---: | ---: | ---: | ---: |
| County | April-00 | May-00 | September-00 | Grand Total |
| Cook |  | 22000 |  | 22000 |
| ltasca | 66500 | 42500 |  | 109000 |
| Lake |  | 60000 | 10600 | 70600 |
| St. Louis | 58500 |  | 26500 | 85000 |
| Grand Total | 125000 | 124500 | 37100 | 286600 |

Check the data, is it correct? What does this tell you?
Perhaps we should re-arrange the data. Instead of County and Month - try Type and month. Do this just like playing solitaire - drag and drop the fields to the place you want.

| Sum of Quantity | Month |  |  |  |
| :--- | ---: | :--- | :--- | :--- | :--- |
| Type | April-00 | May-00 | September-00 | Grand Total |
| Blue Spruce | 37500 | 56000 | 3100 | 96600 |
| Concolor Fir | 13500 | 22000 |  | 35500 |
| Frazier Fir | 6500 | 14500 | 7500 | 28500 |
| Scotch Pine | 15000 |  |  | 15000 |
| White Pine | 52500 | 32000 | 26500 | 111000 |
| Grand Total | 125000 | 124500 | 37100 | 286600 |

What does this tell you?
Make a graph of it using Pivot Charts:



## Using Power Map

Power Map is a new 3D visualization add-in for Excel for mapping, exploring, and interacting with geographical and temporal data, enabling people to discover and share new insights.

Go to Google and search for Power Map or go to this link:
http://www.microsoft.com/en-us/download/details.aspx?id=29074 (Power Pivot - 2010)
https://www.microsoft.com/en-us/download/details.aspx?id=38395 (Power Map - 2013)
After install - you must enable the addin.

1. Go to File > Options > Add-Ins.
2. In the Manage box, click COM Add-ins> Go.
3. Check the Microsoft Office Power Pivot in Microsoft Excel box, and then click OK. If you have other versions of the Power Pivot add-in installed, those versions are also listed in the COM Add-ins list.

The ribbon now has a Power Pivot tab.


Now, we just need some data! Open the file PerCapitaByCounty.xlsx
When using Powermap, we have to be very precise in location. The best way for getting precise geographic data is Longitude and Latitude. But, that won't work when we are talking about counties! So, we have to be precise in the counties we are talking about. Did you know there are St. Louis counties in MN and MI? Or Lake counties in MN, IN, IL? So, states are important!

1. Select the data from A1:G9. Notice that the headers of the data were selected.
2. Click Insert->Map->Launch Power Map
3. A new Panel will open saying "Launch Powermap" and at the bottom, there will be a + sign with the label "New Tour"
4. Click "New Tour"
5. A new window opens with the world on it. On the right, a panel shows up that has the headings of our data.
6. Select County

Items appear on the map with County in the Geography and Map Level
7. Click Next

variable to graph. Select Per Cap 2007.
We now have columns of the money on each county. This is great - but not very useful.
9. Above Height on the right, there are different ways to look at the data

## 

## HEIGHT

a. Stacked Column
b. Clustered Column
c. Bubble Graph
d. Heat Map
e. Visualization by Region-select this
10. Now, each of the counties are colored. They do differ slightly between them, but we may need to increase the amount of contrast. Select the Cog on the top of the panel for Layer Options.
11. This formatting allows for changes in color and contrast.


## LAYER OPTIONS $\mid$ SCENE OPTIONS

1 mume 1 $-1$


With this new chart, lots of changes can be made.

1. Themes - how should the geography be shown?
2. Map Labels
3. Flat Map
4. Charts
5. Annotations

Let's add some Annotations on what the major industries are

1. Right click on the county
2. Select Add annotation
3. Type in the information for the county


Create a population and clustered bar graph
Open the file CountryAndWorkingPop.xls

- Select all the data and push it into PowerMap
- Call this layer Population of County
- Create another layer
- Select County names again (Click Next)
- Select the remaining data and make it a clustered column

| GEOGRAPHY |
| :--- |
| Map by County (County) |
| $\perp$ Range |
| $\perp$ Range1 |
| 4 Range2 |
| $\square$ Both |
| $\square$ County |
| $\square$ Employed Elsewhere |
| $\square$ Living Elsewhere |
| $\square$ Population |


HEIGHT


## Statistics \& Other Mind Blowing Items

## Correlation

The correlation coefficient (a value between -1 and +1 ) tells you how strongly two variables are related to each other. We can use the CORREL function or the Analysis Toolpak add-in in Excel to find the correlation coefficient between two variables.

Note: A correlation coefficient of +1 indicates a perfect positive correlation. As variable $X$ increases, variable $Y$ increases. As variable $X$ decreases, variable $Y$ decreases.

Please type in the values from A1:B6 below on your spreadsheet.

| 4 | A | B | C | D |
| :---: | :---: | :---: | :---: | :---: |
| 1 | A | B | C |  |
| 2 | 0 | 2 | 2 |  |
| 3 | 14 | 6 | 11 |  |
| 4 | 1 | 8 | 3 |  |
| 5 | 10 | 5 | 13 |  |
| 6 | 5 | 6 | 4 |  |
| 7 |  |  |  |  |
| 8 |  |  |  |  |
| ก |  |  |  |  |



1. On the Data tab, click Data Analysis.
2. Select Correlation and click OK

| Data Analysis |  |
| :--- | :--- |
| Analysis Tools |  |
| Anova: Single Factor  <br> Anova: Two-Factor With Replication  <br> Anova: Two-Factor Without Replication  <br> Correlation  <br> Covariance  <br> Descriptive Statistics  <br> Exponential Smoothing  <br> F-Test Two-Sample for Variances  <br> Fourier Analysis  <br> Histogram  |  |

3. Select the range $\mathrm{A} 2: \mathrm{C} 6$ as the Input Range

4. Check Labels in first row.
5. Select cell A9 as the Output Range.
6. Click OK.

|  | $A$ | $B$ | $C$ |
| ---: | ---: | ---: | ---: |
| A | 1 |  |  |
| B | 0.191516 | 1 |  |
| C | 0.909268 | 0.108893 |  |

Conclusion: variables $A$ and $C$ are positively correlated ( 0.91 ). Variables $A$ and $B$ are not correlated (0.19). Variables B and C are also not correlated (0.11)

## Descriptive Analysis

Back in my high school math days, my teacher used to grade on a curve. When the grades came out, he gave us every possible statistical way of looking at the data. Now, I believe he was using a specific type of program to do this. Now, we have excel!

1. Select Data Analysis
2. Select Descriptive Statistics
3. Select the range A2:A15 as the Input Range.
4. Select cell C1 as the Output Range.
5. Make sure Summary statistics is checked.


Result:

Scores
Column1
82
Mean
81.21429

Standard Error
4.045318

Median 85

Mode 93
Standard
Deviation 15.13619
Sample Variance 229.1044
Kurtosis -1.42605
Skewness -0.40211
Range 42
Minimum 58
Maximum 100
Sum 1137
Count 14

## Some definitions

- Mean - What we would call the average
- Median - The middle value if all the numbers were in order
- Mode - The number most often repeated
- Standard Error - A measure of the variability of the set. The deviation over number of items.
- Standard Deviation - used to find the amount of variation in a group of items
- Sample Variance - The average of the squared differences from the Mean.
- Kurtosis - A descriptor of the distribution. A normal distribution has a kurtosis of 0 .
- Skewness - The measure of asymmetry in the distribution. A normal distribution has no skew (0).
- Range - The difference between the largest and smallest values.


## Moving Average

The following illustration shows a summary of Contoso Pharmaceutical's inventory for their 10 bestselling products from last year. This report shows a large variance in the ending inventory quantities from month to month, indicating both shortages and stagnant product - neither of which is good for business. With a few simple steps, you can use this information to manage your inventory levels more precisely this year.

| Prod_ID | Aug | Sep | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun | Jul |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1176 | 96 | 10 | 10 | 72 | 72 | 60 | 60 | 14 | 2 | 93 | 94 | 26 |
| 401 | 19 | 11 | 4 | 47 | 47 | 69 | 52 | 13 | 13 | 56 | 64 | 75 |
| 1482 | 78 | 11 | 7 | 46 | 5 | 30 | 30 | 19 | 9 | 100 | 90 | 74 |
| 1548 | 96 | 11 | 0 | 99 | 99 | 74 | 74 | 18 | 1 | 73 | 18 | 74 |
| 1406 | 48 | 13 | 65 | 99 | 99 | 46 | 46 | 16 | 16 | 94 | 33 | 58 |
| 1517 | 3 | 13 | 13 | 26 | 26 | 92 | 92 | 10 | 1 | 44 | 18 | 47 |
| 301 | 15 | 15 | 32 | 55 | 55 | 17 | 8 | 46 | 8 | 59 | 69 | 84 |
| 303 | 32 | 18 | 41 | 65 | 65 | 11 | 11 | 50 | 1 | 72 | 43 | 51 |
| 688 | 46 | 18 | 0 | -20 | 26 | 75 | 75 | 15 | 1 | 23 | 99 | 49 |

Open the file Moving Average

The Moving Average analysis tool projects values in the forecast period, based on the average value of the variable over a specific number of preceding periods. A moving average provides trend information that a simple average of all historical data would mask. This example uses the data for Contoso product 1176 to predict a target inventory level for the new fiscal year.

1. On the Tools menu, click Data Analysis.
2. In the Data Analysis dialog box, click Moving Average, and then click OK.
3. The Moving Average dialog box opens.

4. In the Input Range box, enter a single row or column of data. This example uses the row of data from product 1176 on the Contoso top-10 products report.
5. In the Interval box, enter the number of values that you want to include in the moving average. In this example, enter 3, the default interval.

NOTE The interval is the number of data points used to calculate the moving average. The larger the interval, the smoother the moving average line; the smaller the interval, the more the moving average is affected by individual data point fluctuations.
In the Output Range box, enter the cell address where you want the results to start.
3. Select the Chart Output check box to see a graph comparing the actual and forecasted inventory levels.
4. Click OK.


## Regression

Regression is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.

Open up the file Regression.xlsx
The big question is: is there a relation between Quantity Sold (Output) and Price and Advertising (Input). In other words: can we predict Quantity Sold if we know Price and Advertising?

|  | A | B | C | D |  |
| :---: | ---: | ---: | ---: | ---: | ---: |
| 1 | Quantity Sold | Price |  | Advertising |  |
| 2 | 8500 | $\$ 2$ | $\$ 2,800$ |  |  |
| 3 | 4700 | $\$ 5$ | $\$ 200$ |  |  |
| 4 | 5800 | $\$ 3$ | $\$ 400$ |  |  |
| 5 | 7400 | $\$ 2$ | $\$ 500$ |  |  |
| 6 | 6200 | $\$ 5$ | $\$ 3,200$ |  |  |
| 7 | 7300 | $\$ 3$ | $\$ 1,800$ |  |  |
| 8 | 5600 | $\$ 4$ | $\$ 900$ |  |  |
|  |  | $\$$ |  |  |  |

1. Open the data tab and select Data Analysis
2. Select Regression and click OK
3. Select the $Y$ Range (A1:A8). This is the predictor variable (also called dependent variable).
4. Select the $X$ Range(B1:C8). These are the explanatory variables (also called independent variables). These columns must be adjacent to each other.
5. Check Labels
6. Select an Output Range
7. Check Residuals
8. Click OK


How to read it
R Square equals 0.962 , which is a very good fit. $96 \%$ of the variation in Quantity Sold is explained by the independent variables Price and Advertising. The closer to 1, the better the regression line (read on) fits the data.

| Regression Statistics |  |
| :--- | ---: |
| Multiple R | 0.980681 |
| R Square | 0.961736 |
| Adjusted R |  |
| Square | 0.942604 |
| Standard Error | 310.5239 |
| Observations | 7 |

Significance F and P-values
ANOVA

|  |  |  |  |  | Significance |  |
| :--- | ---: | ---: | ---: | :---: | :---: | :---: |
|  | $d f$ | SS | MS | $F$ | $F$ |  |
| Regression | 2 | 9694300 | 4847150 | 50.26854 | 0.001464128 |  |
| Residual | 4 | 385700.4 | 96425.11 |  |  |  |
| Total | 6 | 10080000 |  |  |  |  |

To check if your results are reliable (statistically significant), look at Significance $F$ ( 0.001 ). If this value is less than 0.05 , you're OK. If Significance $F$ is greater than 0.05 , it's probably better to stop using this set of independent variables. Delete a variable with a high P-value (greater than 0.05 ) and rerun the regression until Significance F drops below 0.05 .

Coefficients

|  | Coefficients | Standard Error | $t$ Stat | P-value | Lower 95\% | Upper 95\% | Lower 95.0\% |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| Intercept | 8536.213882 | 386.9117478 | 22.06243 | $2.5 \mathrm{E}-05$ | 7461.974654 | 9610.453111 | 7461.974654 |
| Price | -835.7223514 | 99.65304469 | -8.38632 | 0.001106 | -1112.40356 | -559.0411432 | -1112.40356 |
| Advertising | 0.592228496 | 0.104346803 | 5.675579 | 0.004755 | 0.302515325 | 0.881941666 | 0.302515325 |

The regression line is: $y=$ Quantity Sold $=8536.214-835.722$ * Price +0.592 * Advertising. In other words, for each unit increase in price, Quantity Sold decreases with 835.722 units. For each unit increase in Advertising, Quantity Sold increases with 0.592 units. This is valuable information! I smell a raise!

You can also use these coefficients to do a forecast. For example, if price equals \$4 and Advertising equals $\$ 3000$, you might be able to achieve a Quantity Sold of $8536.214-835.722$ * $4+0.592$ * $3000=$ 6970.

Residuals

| Observation | Predicted Quantity Sold | Residuals |
| ---: | ---: | ---: |
| 1 | 8523.008967 | -23.00896712 |
| 2 | 4476.047825 | 223.9521754 |
| 3 | 6265.938227 | -465.9382265 |
| 4 | 7160.883427 | 239.1165726 |
| 5 | 6252.733311 | -52.73331119 |
| 6 | 7095.05812 | 204.9418798 |
| 7 | 5726.330123 | -126.3301229 |

The residuals show you how far away the actual data points are fom the predicted data points (using the equation). For example, the first data point equals 8500 . Using the equation, the predicted data point equals $8536.214-835.722 * 2+0.592 * 2800=8523.009$, giving a residual of $8500-8523.009=-23.009$.

## Goal Seek

Even though you know the formula to something, sometimes you would like to adjust the numbers until you reach a desired outcome. Take for example the payment of a home mortgage. You may know what type of loan you want and the payment - yet, you would need to figure the percentage necessary to get that type of loan.

Create a new worksheet and type:

1. In cell A1, type Loan Amount.
2. In cell A2, type Term in \# of Payments.
3. In cell A3, type Interest Rate.
4. In cell A4, type Payment.

Next, add the values you know

1. In B1, type 100000 - this is the amount you want to borrow
2. In B2, type 180 - the number of months to pay off the loan

We know the equation for finding a mortgage payment is PMT. In this case, we want to type this into B4:
$=P M T(B 3 / 12, B 2, B 1)$
In this example, we want to pay $\$ 900 /$ month. We won't enter that amount here, because you want to use Goal Seek to determine the interest rate, and Goal Seek requires that you start with a formula.

1. On the Data tab, in the Data Tools group, click What-If Analysis, and then click Goal Seek.
2. In the Set cell box, enter the reference for the cell that contains the formula that you want to resolve. In the example, this reference is cell B4.
3. In the To value box, type the formula result that you want. In the example, this is -900 . Note that this number is negative because it represents a payment.
4. In the By changing cell box, enter the reference for the cell that contains the value that you want to adjust. In the example, this reference is cell B3.
5. Click OK

Remember to change the Number look into a so that it makes sense.

|  |  | A | B |
| :--- | :--- | ---: | ---: |

## Data solver

The idea behind the solver is to optimize a process or question. To do this, we usually have multiple equations and multiple unknowns. It is the solver's duty to take these equations and optimize the results to get the best overall answer that you define. That's right - you have to tell it what it is looking for before it can find it!

An optimization model has three parts: the target cell, the changing cells, and the constraints.

## Target cell

The target cell represents the objective or goal. We want to either minimize or maximize the target cell.

## Changing cells

Changing cells are the spreadsheet cells that we can change or adjust to optimize the target cell.

## Constraints

Constraints are restrictions you place on the changing cells

## Installing and Running the Solver

To install the Solver, go to the Options Panel->Add-Ins->Go...
This is the same as getting the analysis Toolpak - but in this case, we want the Solver Add-in. When it is installed, we will see a new "Solver" button in the Analysis section of the Data tab.

I've set up a spreadsheet that we will use to find the maximum profit while creating a certain amount of pallets using the solver.

Open the file Pallet_Solver.xlsx

| 4 | 1 A | B | c | D | E | F | G | H | I |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 |  |  |  |  |  |  |  |  |  |
| 2 |  |  | Panel | Type |  |  |  |  |  |
| 3 |  | Tahoe | Pacific | Savannah | Aspen |  |  |  |  |
| 4 | Pallets | 0 | 0 | 0 | 0 | Total Profit |  |  |  |
| 5 | Profit | \$ 450.00 | \$1,150.00 | \$ 800.00 | \$400.00 | \$ |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |
| 7 |  | Resour | ces Required | d per Pallet | Type | Used | Available |  |  |
| 8 | Glue | 50 | 50 | 100 | 50 | 0 | 5800 | quarts |  |
| 9 | Pressing | 5 | 15 | 10 | 5 | 0 | 730 | hours |  |
| 10 | Pine Chips | 500 | 400 | 300 | 200 | 0 | 29200 | pounds |  |
| 11 | Oak Chips | 500 | 750 | 250 | 500 | 0 | 60500 | pounds |  |
| 12 |  |  |  |  |  |  |  |  |  |
| 13 |  |  |  |  |  |  |  |  |  |

Creating your own worksheet should be done in a similar fashion - where the target cells and the changing cells are arranged in some logical fashion and are marked - making it easier to see what you are doing.

In this worksheet, B4:E4 are the changing cells which represent our decision variables representing the number of pallets of each type of panel to produce. The solver will determine the optimal values for these cells.

Notice that the profit for each pallet of panels ( $\$ 450, \$ 1,150, \$ 800$ and $\$ 400$ ) was entered in cells B5, C5, D5 and E5, respectively. This allows us to compute the target cell in F5 as:

```
=B5*B4+C5*C4+D5*D4+E5*E4
```

Or equivalently,
=SUMPRODUCT(B5:E5,B4:E4)
In cells B8:E11, we've entered the amount of resources needed to produce a pallet of each type of panel. For example, the value 15 in cell C9 means that 15 hours of pressing is required to produce a pallet of Pacific style panels.

With these values in place, we can enter a formula in cell F8 to compute the total amount of glue used for any number of pallets produced:

Formula for cell F8: =SUMPRODUCT(B8:E8,\$B\$4:\$E\$4)
With these values in place, we can enter a formula in cell F8 to compute the total amount of glue used for any number of pallets produced:

Formula for cell F8: =SUMPRODUCT(B8:E8,\$B\$4:\$E\$4)
We can copy this formula to cells F9:F11 to compute the total amount of pressing, pine chips, and oaks chips used. (The dollar signs in $\$ \mathrm{~B} \$ 4: \$ \mathrm{E} 4$ specify that this cell range stays constant, while the cell range B8:E8 becomes B9:E9, B10:E10, and B11:E11 in the copied formulas.) The formulas in cells F8:F11 correspond to the left hand side values of the constraints.

In cells G8:G11, we've entered the available amount of each type of resource (corresponding to the right hand side values of the constraints). This allows us to express the constraints as:

## \$F\$8:\$F\$11<=\$G\$8:\$G\$11

This is equivalent to the four constraints: $\mathrm{F} 8<=\mathrm{G} 8, \mathrm{~F} 9<=\mathrm{G} 9, \mathrm{~F} 10<=\mathrm{G} 10$, and $\mathrm{F} 11<=\mathrm{G} 11$. We can enter this set of constraints directly in the Solver dialogs along with the non-negativity conditions:

```
$B$4:$E$4>= 0
```

To let the Solver know which cells on the worksheet represent the decision variables, constraints and objective function, we click Solver button on the Data tab, which displays the Solver Parameters dialog. In the Set Objective (or Set Target Cell) edit box, we type or click on cell F5, the objective function. In the By Changing Variable Cells edit box, we type B4:E4 or select these cells with the mouse.

To find the optimal solution, we simply click on the Solve button!


At this point, it asks to keep the solution or restore the values. You can also get a report which describes the solution.

Select the Outline Reports and click OK. In Excel 2013, you get new tabs with all the information


