

# Exploratory Data Analysis with Power BI

## Overview

Students will practice turning clean, structured data models into visualizations to gain meaningful insights.

This lesson is mostly about the **how** of creating visuals.

In the next lesson (Communicating Insights) we'll explore the **why** behind formatting and enhancing choices.

## Learning Objectives

In this lesson, students will:

- Explore different types of analytics
- Leverage the power of data visualization
- Apply the right type of chart to a given question
- Use visualization to gain insights and tell the story

**Duration:** 3 hours



# Suggested Agenda - Exploratory Data Analysis (1/2)

Time		Activity	Topics
0:00–0:10	10 mins	<u>Opening</u>	Today's Agenda & Learning Objectives
0:10–0:15	5 mins	<u>Topic Introduction</u>	What is EDA?
0:15–0:30	15 mins	<u>Quick Review</u>	Descriptive Statistics
0:30–0:35	5 mins	<u>Topic Introduction</u>	Types of Visualization
0:35–0:40	5 mins	<u>Guided Practice</u>	Setting Up Our Data Model
0:40–0:50	10 mins	<u>Guided Practice</u>	Bar and Column Charts
0:50–1:20	30 mins	<u>Guided Practice</u>	Line and Area Charts, Hierarchies, Trendlines, Forecasts, Prediction vs. Forecast
1:20–1:30	10 mins	<u>BREAK</u>	



# Suggested Agenda - Exploratory Data Analysis (2/2)

Time		Activity	Topics
1:30–1:55	25 mins	<a href="#"><u>Guided Practice</u></a>	Scatter Charts, Bubble Charts, Play Axis, Reference Lines
1:55–2:10	15 mins	<a href="#"><u>Guided Practice</u></a>	Pie, Donut & Treemap Charts, Themes
2:10–2:25	15 mins	<a href="#"><u>Guided Practice</u></a>	Specialty Charts, Box and Whisker Charts
2:25–2:30	5 mins	<a href="#"><u>Guided Practice</u></a>	Conditional Formatting
2:30–2:35	5 mins	<a href="#"><u>Independent Practice</u></a>	Conditional Formatting
2:35–2:50	15 mins	<a href="#"><u>Independent Practice</u></a>	Histogram Activity
2:50–3:00	10 mins	<a href="#"><u>Review &amp; Wrap Up</u></a>	Reflection, Exit Tickets



Advanced Analytics

---

# Hello, Humana!

# Before we get into today's content... mid-program survey!

- Please take 5-10 minutes to carefully fill out the survey below. This survey goes more in-depth than Exit Tickets and we'll ask you provide detailed feedback on the:
  - Pacing
  - Content
  - Instructors (individually)
- Note: Please select “mid-program” for question 2

Survey link here: [https://bit.ly/HUMAA\\_PROG](https://bit.ly/HUMAA_PROG)



# What Have We Learned So Far?

Last lesson we:

- **Loaded data** from multiple sources
- **Cleaned and reshaped data** using the Power Query Editor
- **Merged and appended** data
- **Created and managed relationships** between tables
- **Calculated new fields** using DAX

# Today's Agenda

Here's what we'll cover today:

- Summary statistics
- Descriptive vs. Predictive Analysis
- Visualization as a Tool for Data Science
- Types of Visualization
  - Bar / Column Charts
  - Line / Area Charts
  - Scatter Charts
- Specialty Charts
- Conditional Formatting



# Learning Objectives

After this lesson, you'll be able to:

- Choose the best visual for a given inquiry
- Build common visualizations, including:
  - Bar and Column Charts
  - Line and Area Charts
  - Scatter Charts
  - Pie and Donut Charts
- Install and employ 3rd party specialty charts
- Apply conditional formatting to charts





Computers Out:

# Prepare for Class

## Before we begin:

Please be prepared to access the following files for this lesson:

1. **TableRelationships.pbix**



Advanced Analytics

---

# Exploratory Data Analysis

# Why Are We Doing This?

I have lots of data comparing two variables. If I know the value of one, can I predict the other?

How can I get a snapshot of the business?

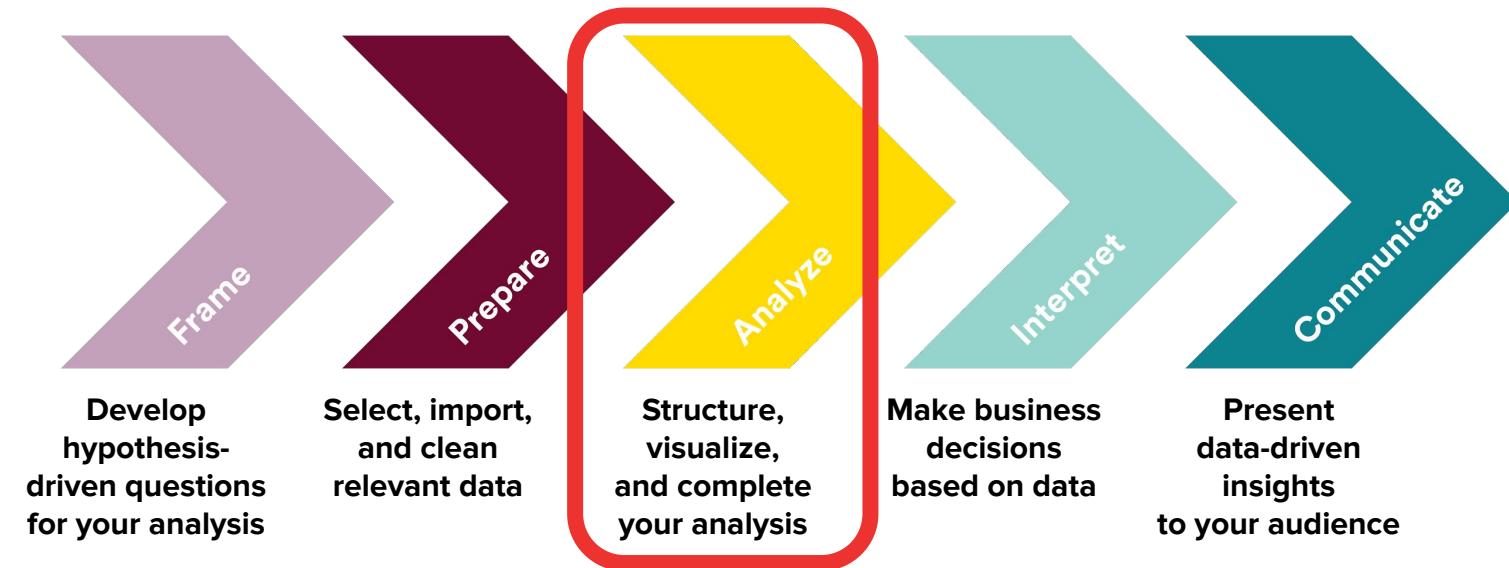
How do I show what I want to say?

I have LOTS of data. How can I summarize it?

I have data going back 5 years. Can I predict what will happen next month?

# Data Journey

Now that we've selected, cleaned and organized our data into a working model, it is time to analyze what we have with available tools.



# How does this lesson fit in?

## Power BI Unit Overview

Once you've obtained your data (via SQL or other remote data stores), you need to explore, statistically model, and visualize it in order to find and communicate meaningful insights.

	Lesson	Description
1	Why BI Tools?	Orient individuals to Power BI with an interactive session exploring pre-existing dashboards and the insights they reveal in Power BI. Evaluate what makes a dashboard/analysis effective or not from a communication and clarity perspective.
2	Wrangling + Exploring	Learn how to clean, join, and label data in Power BI in order to set it up for analysis. Then, begin exploring relationships between variables you've brought into an analysis to identify those with the most compelling insights.
3	Exploratory Data Analysis	Model and program relevant variables from exploration into dashboards in Power BI to make your dashboards live entities that others can explore. Begin introducing the concept of significance of findings in stats terms in order to identify the reliability of results.
4	Communicating Insights	Visualize datasets in Power BI to explore and communicate findings. Then, discuss how to use Power BI to share insights & templates.
5	Lab 2	Use Power BI to analyze sales data and answer prompts related to products, sales, and marketing.

# Lesson Context

For this lesson, we are **Product Analysts** at the multinational manufacturing company “Adventure Works Cycles”. We’ve been tasked with exploring purchasing habits of particular customers by year ...and we’ll be doing that in Power BI! Doing so will allow us to make better recommendations on marketing strategies, develop or discontinue products, and forecast next year’s sales! To do that, we will:

1. Review common methods of analysis, specifically focusing on descriptive statistics
2. Practice applying summary statistics to sample data in Power BI
3. Learn and practice set analysis to slice and filter our data
4. Apply DAX functions to aggregate data in sample dashboards
5. Assemble Scatter Plots in to explore/identify relationships between variables
6. Embed trendlines and reference lines in our charts to visualize relevant trends
7. Predict the value of one variable based on a second variable with simple linear regression

Advanced Analytics

---

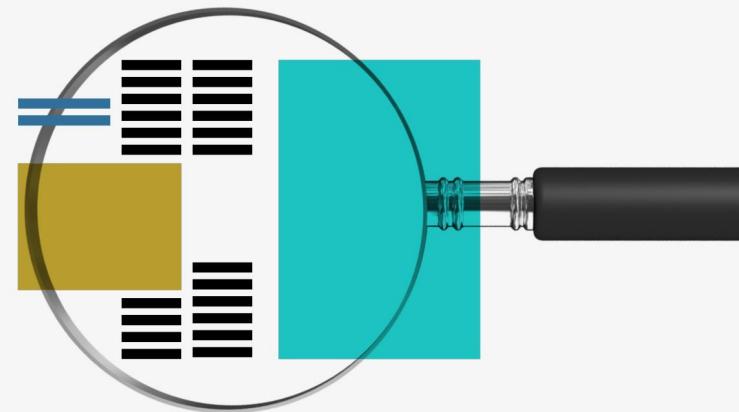
# What is EDA?

# Exploratory Data Analysis

...is exactly what it sounds like!

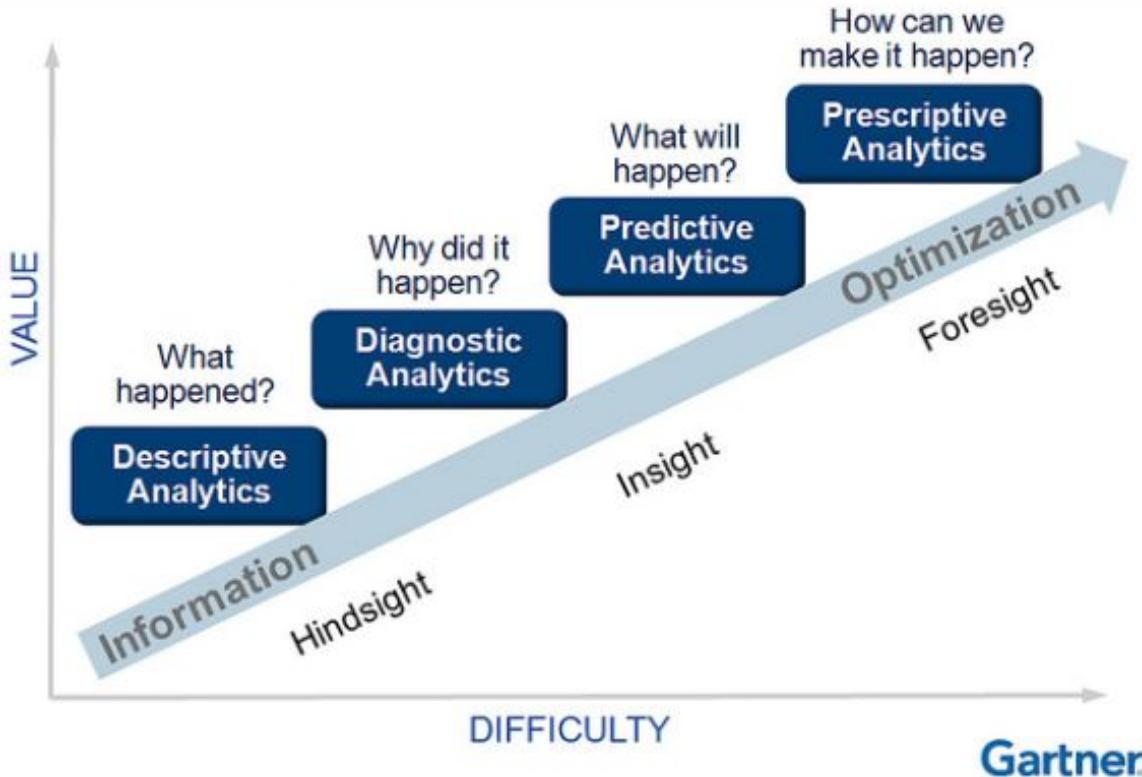
**EDA may involve:**

- descriptive vs. predictive analytics
- summary statistics
- visualization as a tool for further analysis



*When designing dashboards, think about  
how you can provide these as tools for your users.*

# Types of Analytics



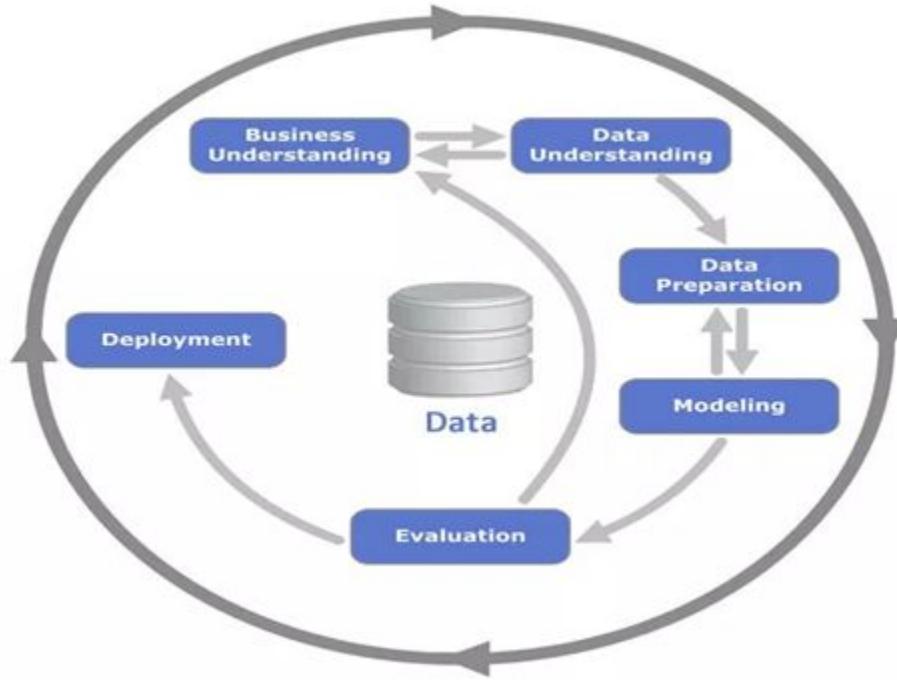
# Summary Statistics

Where line charts and scatter plots may uncover trends and correlations, summary statistics like **mean**, **median**, **max/min** and **sum** are also useful.

In Power BI these are called **measures**, and we can add them to reports and dashboards.

# Visualization as a Tool for Data Science

CRISP-DM  
Process  
Diagram



Advanced Analytics

---

# Quick Review: Descriptive Statistics

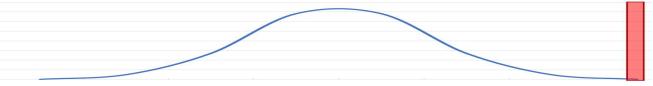
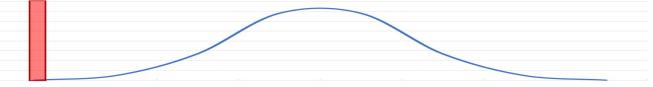
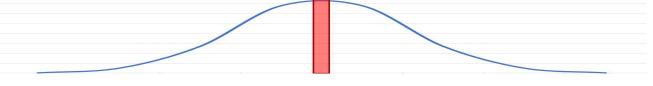
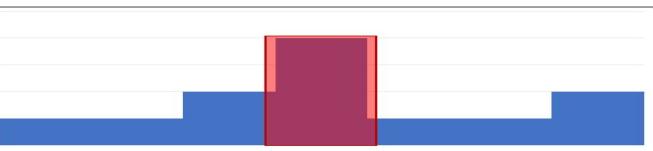
# Summary Statistics

**Measures of Central Tendency** tell us about the center of the data; at a basic level, what one might expect to happen again.

**Measures of Dispersion** tell us how "spread out" our data is, or how far individual values stray from the "center"

Question	Statistic
What was the most amount of money spent in one sale?	<b>Max</b> for a range of sales
How much do I usually spend on utilities?	<b>Mean</b> (average) of utility spend <b>Median</b> (midpoint) of utility spend
When I buy widgets, how many widgets do I usually buy at once?	<b>Mode</b> of widget purchase quantity

# Summary Statistics

Statistic	How to Calculate	Plot
<b>Max</b>	Maximum (largest) value in the range; identify by ordering the data set.	
<b>Min</b>	Minimum (smallest) value in the range; identify by ordering the data set.	
<b>Mean</b>	Mathematical midpoint in the range. Sum the range, then divide by the number of values in the range.	
<b>Median</b>	Presented midpoint in the range. Order the data set, then look for the middle value (or take the mean of middle two).	
<b>Mode</b>	Value most commonly found in the data set. Order the data set and count frequency of each value.	

# Descriptive Statistics in Action

Apply summary statistics to sales data:

- The most sold in one month is  $\text{Max}(\text{Units Sold})$ : 2,500
- The least sold in one month is  $\text{Min}(\text{Units Sold})$ : 500
- The average quantity sold in one month is  $\text{Mean}(\text{Units Sold})$ : 1,500
- The midpoint quantity sold in one month is  $\text{Median}(\text{Units Sold})$ : 1,500
- The most common quantity sold in one month is  $\text{Mode}(\text{Units Sold})$ : 1,500

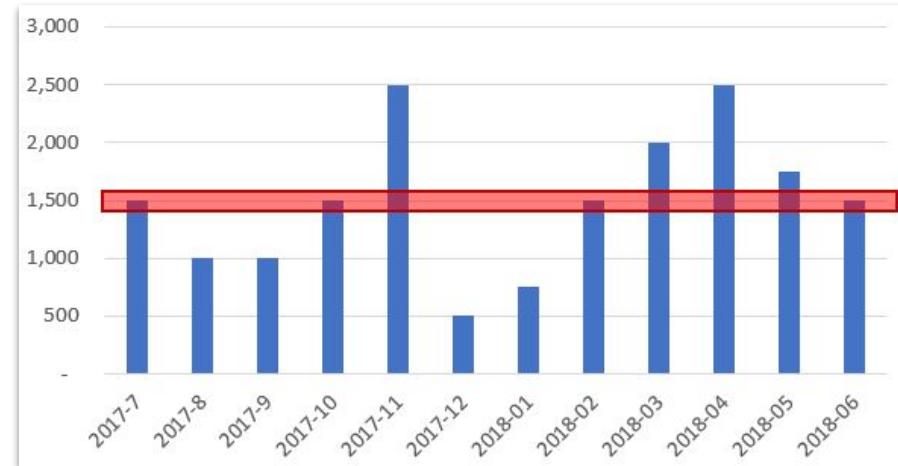
Acme Widget Company Sales Ledger					
Customer	Month	Units Sold	Discount %	Amount	
18493	2017-7	1,500	2.5%	\$ 7,312.50	
18493	2017-8	1,000	2.5%	\$ 4,875.00	
18493	2017-9	1,000	5.0%	\$ 4,750.00	
18493	2017-10	1,500	5.0%	\$ 7,125.00	
18493	2017-11	2,500	15.0%	\$10,625.00	
18493	2017-12	500	10.0%	\$ 2,250.00	
18493	2018-01	750	2.5%	\$ 3,656.25	
18493	2018-02	1,500	2.5%	\$ 7,312.50	
18493	2018-03	2,000	5.0%	\$ 9,500.00	
18493	2018-04	2,500	5.0%	\$11,875.00	
18493	2018-05	1,750	2.5%	\$ 8,531.25	
18493	2018-06	1,500	2.5%	\$ 7,312.50	

# Descriptive Statistics Have Limitations

Summary statistics are useful to get a directional “feel” for the data, but they are limited in their decision-making value.

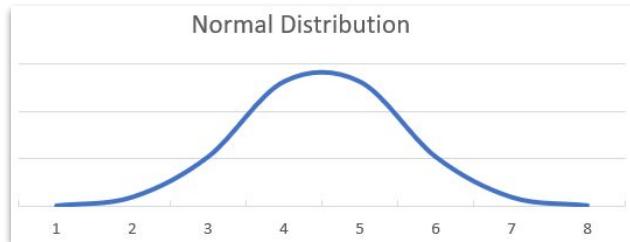
For example, if we created the same statistics for ALL customers, we could try to plan inventory or sales budget to the mean of 1,500 units per month.

However, we would be ill-prepared for eight of the 12 months, assuming that history repeats itself.



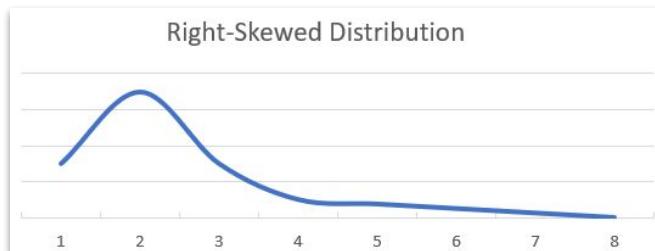
# Data Distributions

A **normal distribution** (a.k.a., bell curve) is dense in the center of the distribution (the mean) and has equal tails. The observed mean and median will have roughly the same value.



A **right-skewed** distribution is asymmetric, with a longer tail to the right. Observed **median** will be less than the **mean**.

Left-skewed has the opposite shape, with a mean less than the median.



# Measures of Dispersion

**Variance** is a measure of how far each value strays from the **mean** observed in the data set. It is a squared value, which makes it difficult to directly compare against values in the data set.

**Standard deviation** is more useful for the analyst, as it is expressed in the same units as the **mean** and other values in the data set. Formally, it is the square root of **variance**. Functionally, like variance, it is a measure of data dispersion from the **mean**. Practically, a standard deviation that is closer to zero indicates a tightly clustered sample; a large standard deviation indicates more variability in the data.

## POPULATION STANDARD DEVIATION

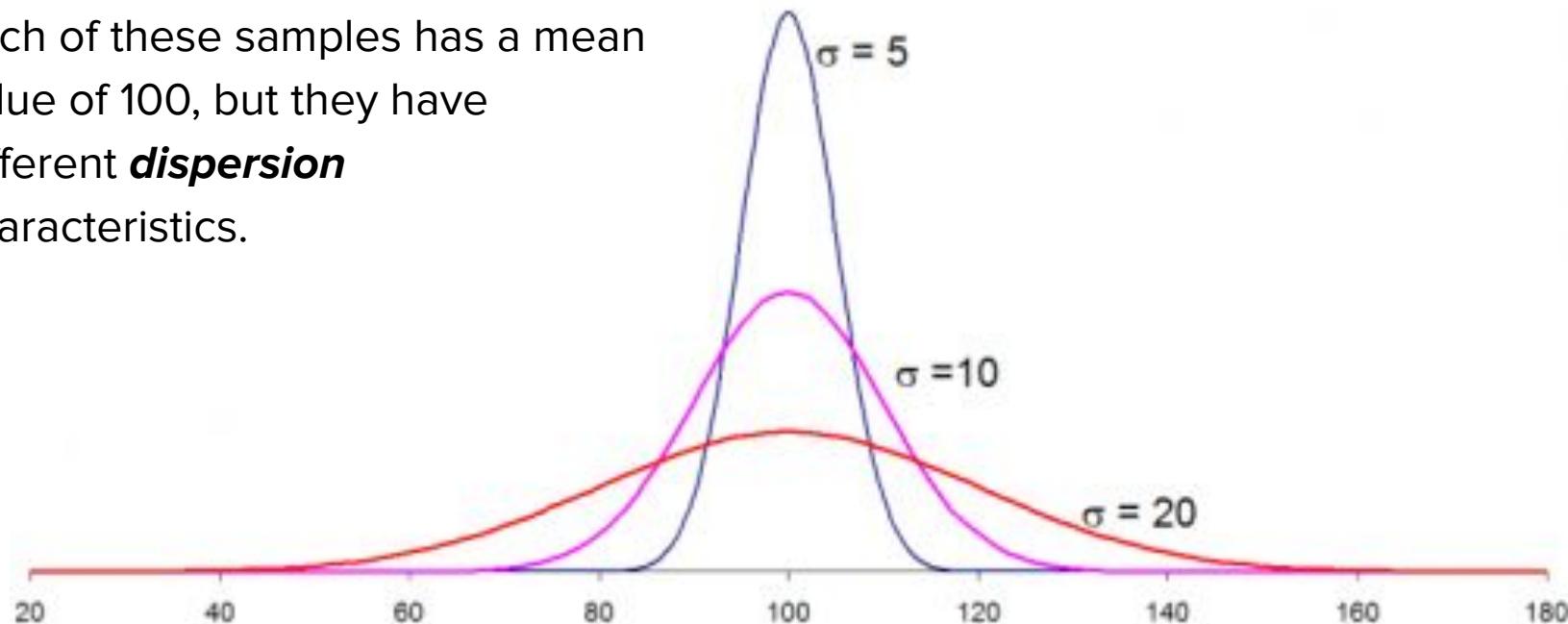
$$\sigma = \sqrt{\frac{\sum(X - \mu)^2}{N}}$$



# Measures of Dispersion in a Data Set

Consider three samples of data.

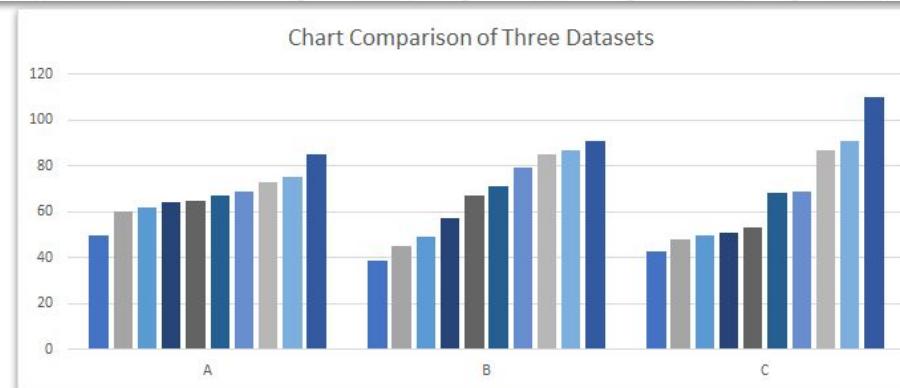
Each of these samples has a mean value of 100, but they have different ***dispersion*** characteristics.



# Measures of Dispersion in a Data Set

Consider three samples of data: Each of these samples has a mean value of 67, but they have different ***dispersion*** characteristics.

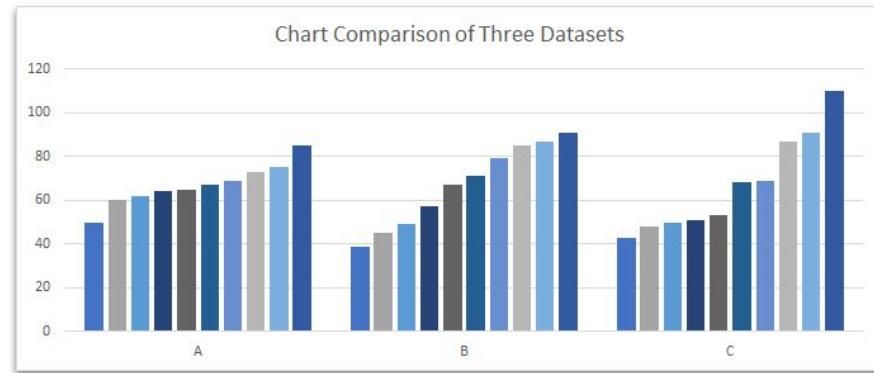
Sample	Mean	Median	Min	Max	Variance	SD	Confidence Interval
A (65,75,73,50,60,64,69,62,67,85)	67	66	50	85	80.4	8.97	67 +/- 6.41
B (85,79,57,39,45,71,67,87,91,49)	67	69	39	91	315.2	17.75	67 +/- 12.7
C (43,51,53,110,50,48,87,69,68,91)	67	60.5	43	110	450.8	21.23	67 +/- 15.19



# Measures of Dispersion in a Data Set

What does this mean?

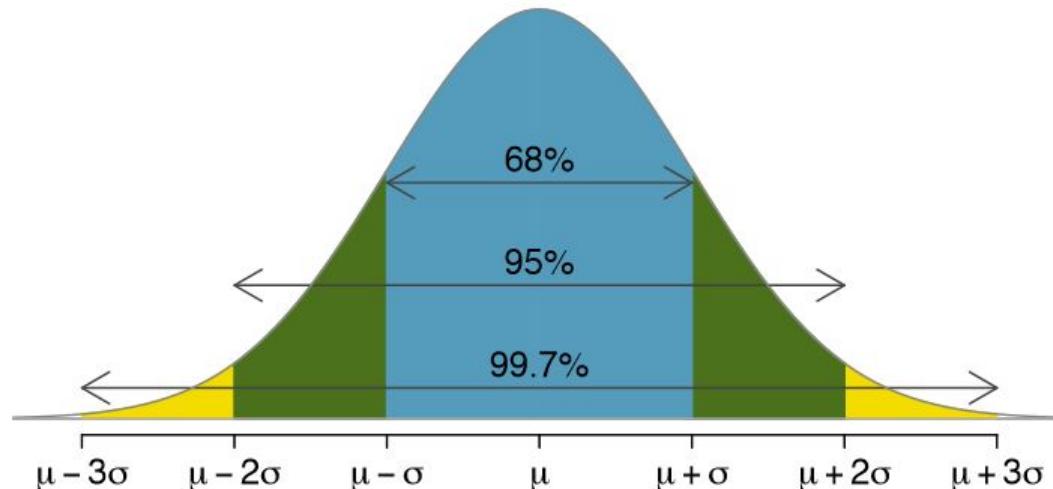
- **Min** and **max** can help us describe the ends of the data set.
- **Mean** and **median** identify the center and start to provide some information about shape.
- **Standard deviation** tells us more about dispersion within the distribution.



Sample	Mean	Median	Min	Max	Variance	SD	Confidence Interval
A (65,75,73,50,60,64,69,62,67,85)	67	66	50	85	80.4	8.97	67 +/- 6.41
B (85,79,57,39,45,71,67,87,91,49)	67	69	39	91	315.2	17.75	67 +/- 12.7
C (43,51,53,110,50,48,87,69,68,91)	67	60.5	43	110	450.8	21.23	67 +/- 15.19

# Normal Distribution

- A **normal distribution** will have 68% of the sample data population fall within one standard deviation of the mean, and 95% of sample data will fall within two standard deviations of the mean.
- Around 99.7% of samples will fall within three standard deviations of the mean.



# Wrapping Up Descriptive Statistics

- Descriptive statistics *describe* data and tell us about the past.
- These statistics may be based on evaluating the *range*, *central tendency*, or *dispersion* of the data set:
  - **Range:** Min and max
  - **Central Tendency:** Mean, median, and mode
  - **Dispersion:** Variance, standard deviation
- Data sets can present in many different *distributions*:
  - **Normal** distribution is centered around the mean
  - **Skewed** distributions rise sharply and then trail off



Questions?



## What can descriptive statistics do?

- A. Describe data and tell us about the past.
- B. Describe data and tell us about where we need to be.
- C. Describe the past and tell us about the future of our data.



## What can descriptive statistics do?

- A. Describe data and tell us about the past.
- B. Describe data and tell us about where we need to be.
- C. Describe the past and tell us about the future of our data.

# Quiz!



## Fill in the blanks:

Measures of Central Tendency tell us about the \_\_\_\_\_ of the data; at a basic level, what one might \_\_\_\_\_ to happen again.

Measures of \_\_\_\_\_ tell us how "spread out" our data is, or how far individual values stray from the "center". These include \_\_\_\_\_ and \_\_\_\_\_ .

# Quiz!



## Fill in the blanks:

Measures of Central Tendency tell us about the **CENTER** of the data; at a basic level, what one might **EXPECT** to happen again.

Measures of **DISPERSION** tell us how "spread out" our data is, or how far individual values stray from the "center". These include **VARIANCE** and **STANDARD DEVIATION**.

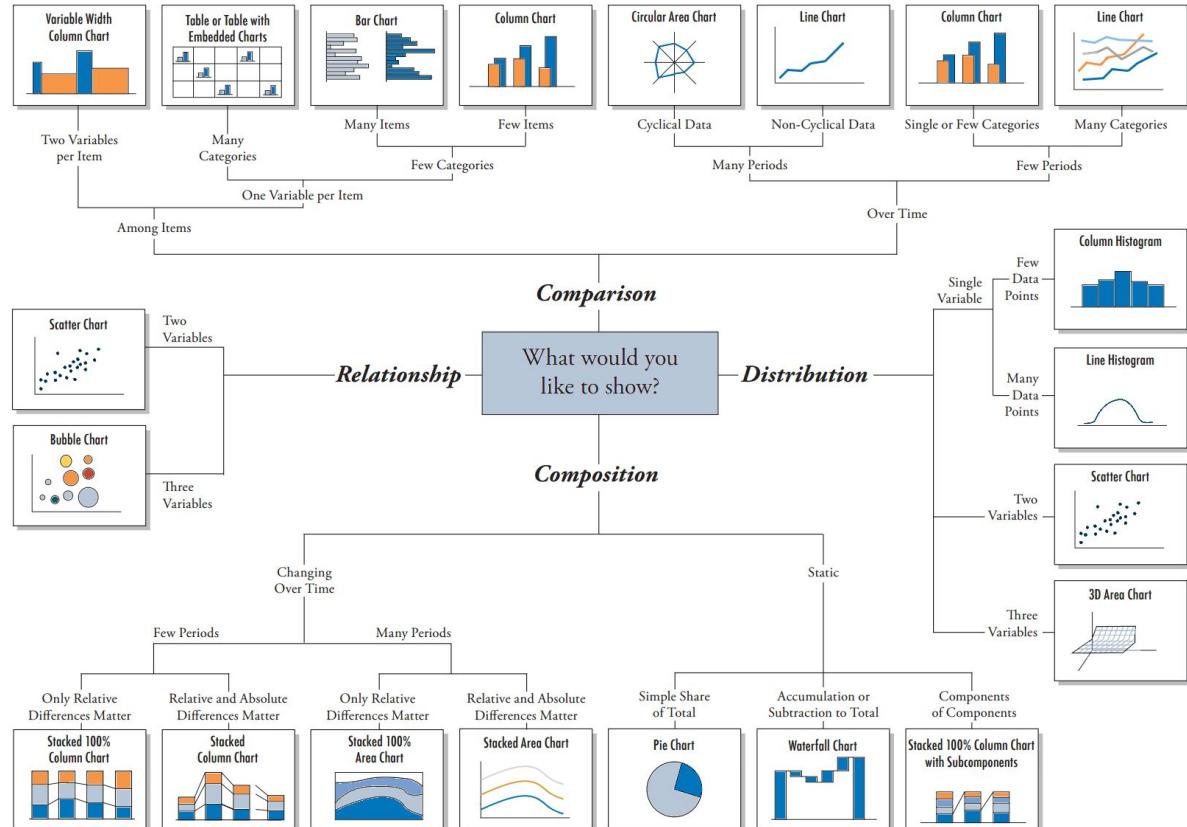
Advanced Analytics

---

# Types of Visualizations

# Chart Options

Choose based  
on the **variables**  
you are trying to  
visualize.



# Power BI Visuals Categories

The VISUALS REFERENCE  
FOR MICROSOFT POWER BI

SEP. 2018  
<http://sqlbi.com/visual-reference>

**COMPARISON**  
To compare the magnitude of measures

**CHANGE OVER TIME**  
To display the changing trend of measures

**RANKING**  
To rank measures in an order

**SPATIAL**  
To display measures over spatial maps

**FLOW**  
To display a flow or dynamic relations

**PART-TO-WHOLE**  
To identify the parts making up a measure total

**DISTRIBUTION**  
To display the distribution of values

**CORRELATION**  
To show correlations between measures

**SINGLE**  
To present single values

**FILTER**  
To control report filters

**NARRATIVE**  
To tell a story with data

**MISCELLANEOUS**

Legend:  
Recommended (light blue square)  
There is a better alternative (yellow square)  
Don't use in the category (grey square)  
Built-in visual (yellow circle with exclamation)  
Certified visual (blue circle with checkmark)  
R required (red circle with question mark)

2018 © SQLBI AND OKVIZ ARE TRADEMARKS OF SQLBI CORP / POWER BI IS A TRADEMARK OF MICROSOFT CORP

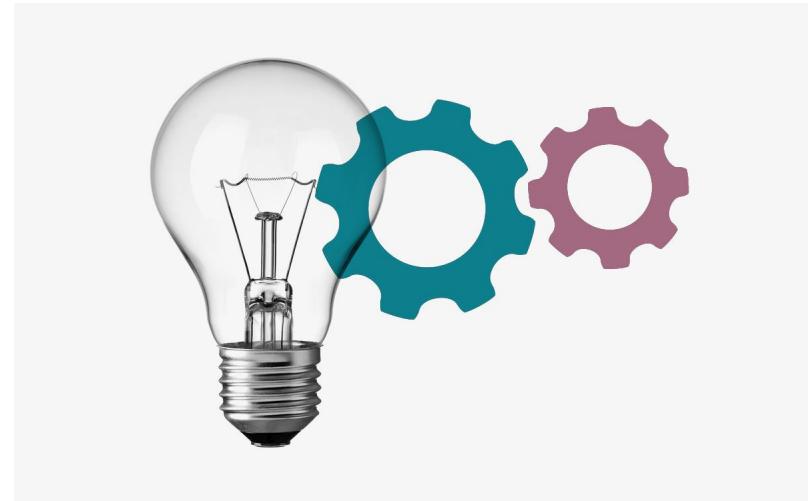


# In This Section...

We'll start by getting set up. We'll import our data and create the appropriate data models.

Then we'll build:

- Bar Charts and Column Charts
- Line Charts and Area Charts
- Scatter Charts
- Pie, Donut and Treemap Charts



Types of Visualizations

---

# Guided Walkthrough: Setting Up Our Data Model



# Guided Walk-Through: Setting Up Our Data Model

In this section we'll skip the Query Editor and move straight to the Data Model. Open Power BI Desktop, and open **TableRelationships.pbix**.

The screenshot shows the Power BI Desktop interface. On the left, there's a dark sidebar with 'Power BI Desktop' at the top, followed by 'Get data', 'Recent sources', and a highlighted 'Open other reports' button. A pink dashed arrow points from this button to the main workspace. The main workspace is titled 'TableRelationships - Power BI Desktop'. It has a ribbon with 'File', 'Home' (selected), 'Insert', 'Modeling', 'View', and 'Help'. The 'Home' tab has options like 'Get data', 'Clipboard', 'Data', 'Queries', 'Insert', 'Calculated columns', and 'Share'. Below the ribbon is a data grid titled 'SalesOrderID OrderQty ProductID UnitPrice StandardCost'. The grid contains several rows of data with totals at the bottom. To the right of the grid are three sections: 'Filters', 'Visualizations', and 'Fields'. The 'Fields' section lists various tables and columns such as 'Person Address', 'Production Product', and 'Sales SalesOrderDetail'. A red circular 'GA' logo is in the bottom right corner.

SalesOrderID	OrderQty	ProductID	UnitPrice	StandardCost
43659	5	708	\$7.0	2.40
43659	4	711	20.19	13.09
43659	2	712	5.19	6.92
43659	3	714	28.84	38.49
43659	2	716	20.19	13.09
43659	1	771	2,039.99	1,912.15
43659	1	772	2,039.99	1,912.15
43659	2	773	2,039.99	1,912.15
43659	2	774	2,039.99	1,912.15
43659	1	776	2,034.00	1,908.00
Total			56,423.747.61	130,335.89



## Guided Walk-Through:

# Setting Up Our Data Model

It may be a good idea to **Save As** something else.

This way you'll have a working file  
**(TableRelationships2.pbix** for example)  
as well as the original file in case  
something breaks.



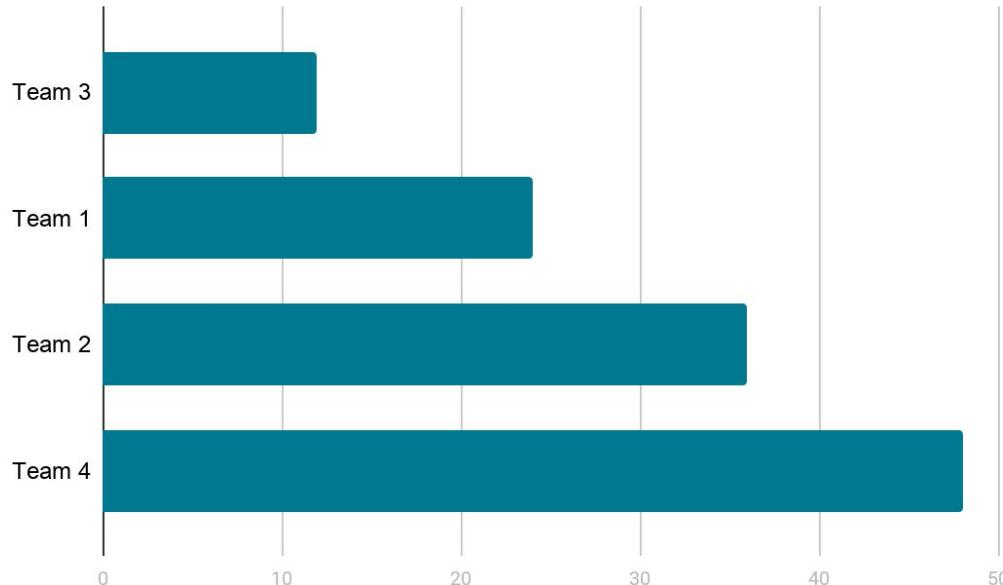
Types of Visualizations

---

# Guided Walkthrough: Bar and Column Charts

# Bar and Column Charts

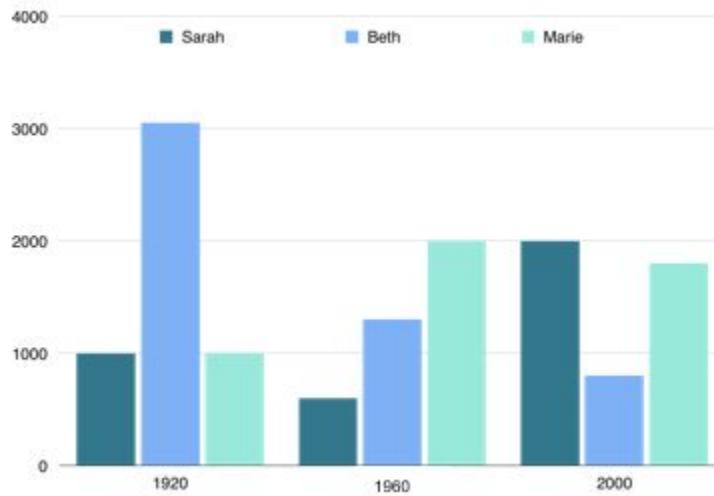
Points scored



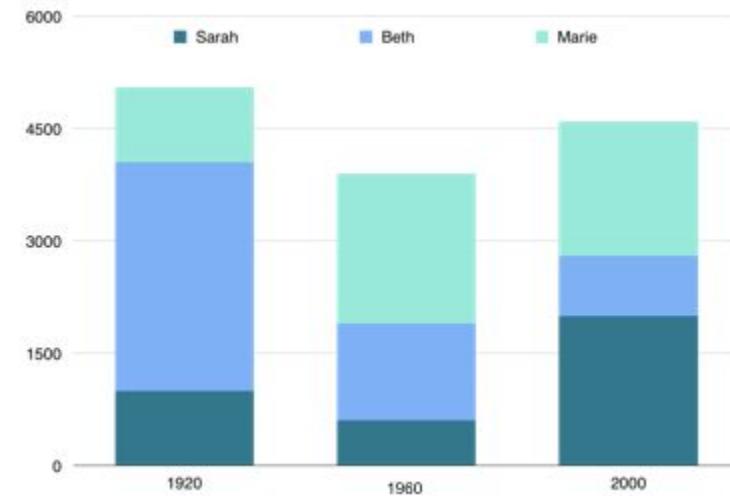
- Show relationships between **categorical** values
- Used to show **comparisons**
- Often ordered to show **rank**

# Clustered vs Stacked Charts

## Popularity of Baby Names



**CLUSTERED**



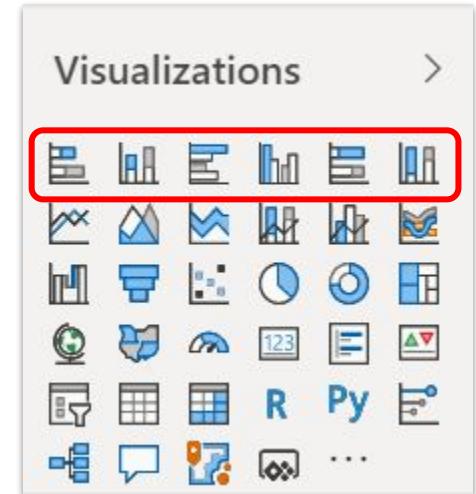
**STACKED**



# Bar and Column Charts

There are **6 types** of Bar and Column charts listed in the first line of the Visualization pane:

- **Stacked** (Bar and Column) breakdown the value of the Legend (optional) on the top of each other
- **Clustered** (Bar and Column) breakdown the value of the Legend (optional) next to each other
- **100% Stacked** (Bar and Column) breakdown the percentage of the Legend (optional) to the total Bar or Column





## Guided Walk-Through:

# Bar and Column Charts

In the Report view, open a new page and click **Stacked bar chart**.

Drag **Color** to *Axis*, **Class** to *Legend*, and **ProductID** to *Values*.

Filter **Color** to only display Black, Blue, Red and Silver.

The screenshot shows the Power BI Report view with a Stacked Bar Chart titled "Count of ProductID by Color and Class". The chart displays four categories: Black, Blue, Red, and Silver, each subdivided into three sub-categories: Class A, Class B, and Class C. The legend on the right side of the chart identifies the colors: Black (blue), Blue (orange), Red (red), and Silver (purple).

The Fields pane on the right side of the interface is open, showing the following configuration for the visualization:

- Visualizations:** A grid of icons representing different visualization types, with the Stacked Bar icon selected.
- Fields:** A list of fields categorized under "Production Product".
  - Color:** Checked, assigned to the **Axis** role.
  - Class:** Assigned to the **Legend** role.
  - ProductID:** Assigned to the **Values** role.

On the left side of the Fields pane, there are four numbered callouts highlighting specific settings:

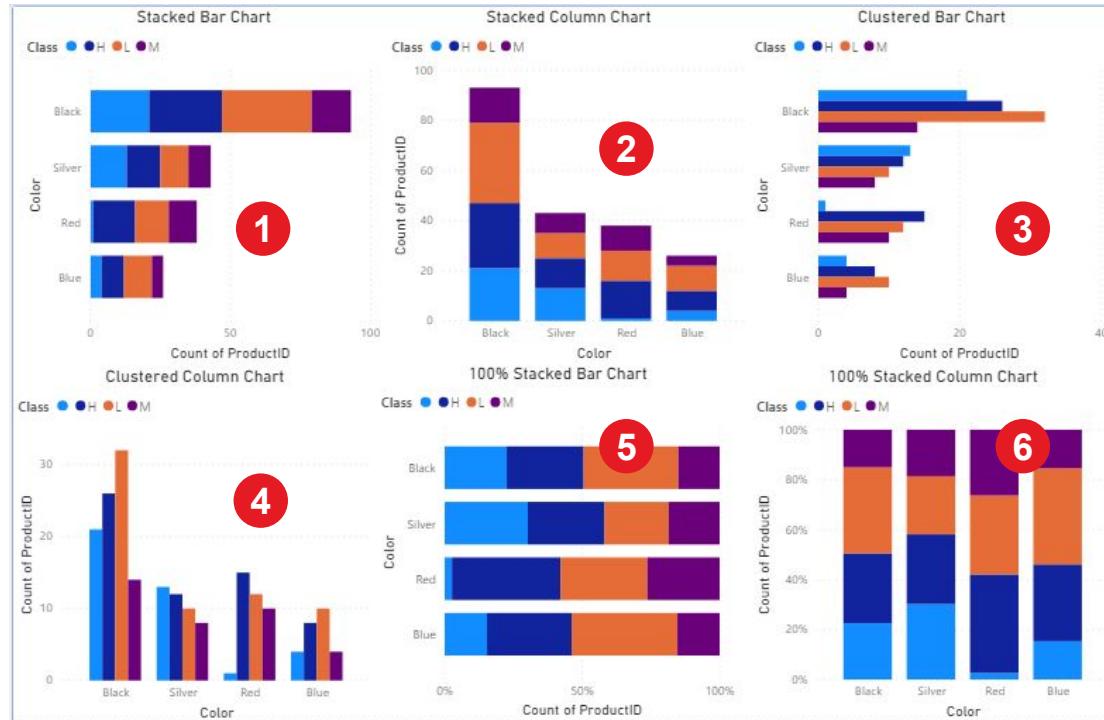
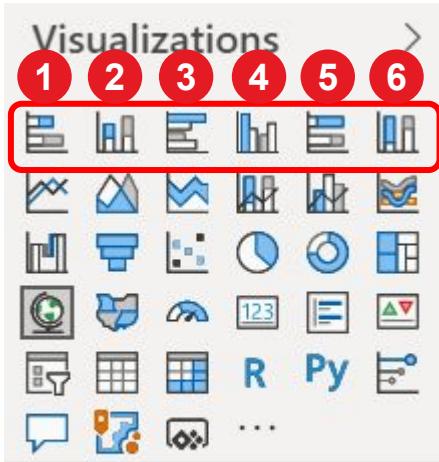
- ① A red box highlights the "Color" filter in the "Basic filtering" section of the Fields pane.
- ② A red box highlights the "Axis" role assignment for "Color".
- ③ A red box highlights the "Legend" role assignment for "Class".
- ④ A red box highlights the "Values" role assignment for "ProductID".



## Guided Walk-Through:

# Bar and Column Charts

Copy the chart five more times,  
and turn each one into a  
different bar or column chart.





## Guided Walk-Through:

# Bar and Column Charts

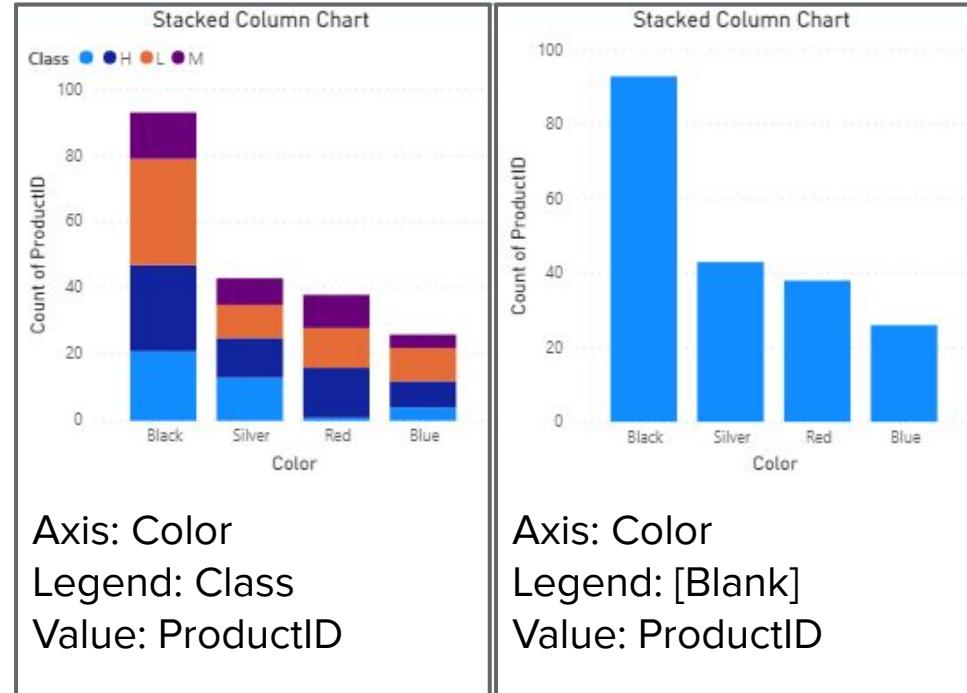
### Chart parameters:

**Axis** - the category being compared

**Legend** - optional subcategory

**Value** - the aspect of the category  
being counted/summed

*Select the Stacked column chart  
and remove Class from Legend to  
see this in action!*



Types of Visualizations

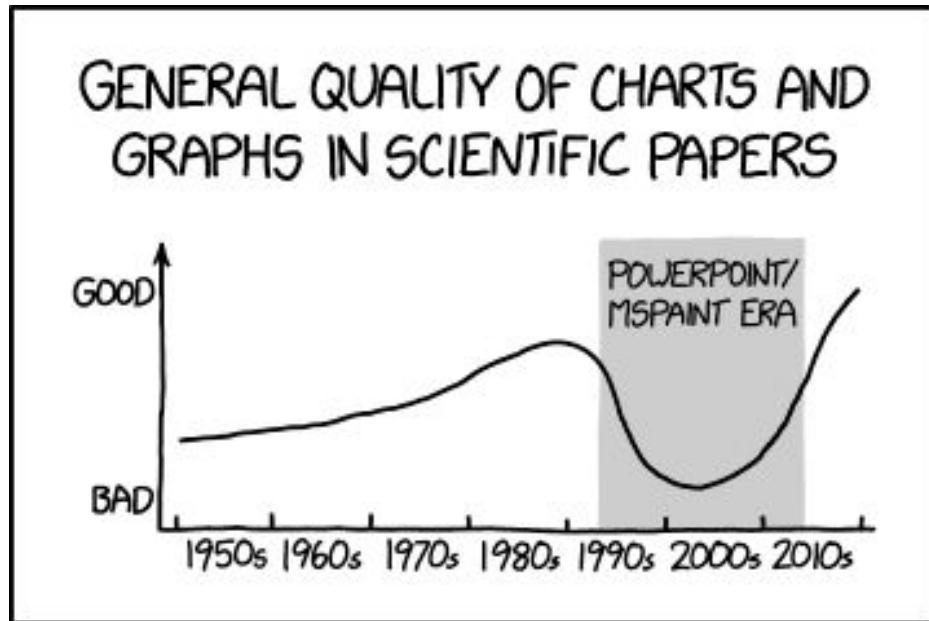
---

# Guided Walkthrough: Line and Area Charts

- Line, Area and Stacked area charts
- Hierarchies
- Trend Lines
- Forecasts



# Line Charts



- Show trends or change of one or several variables.
- Works well for a relatively small number of variables.
- Not to be used to show too many variables, as they become difficult to discern.

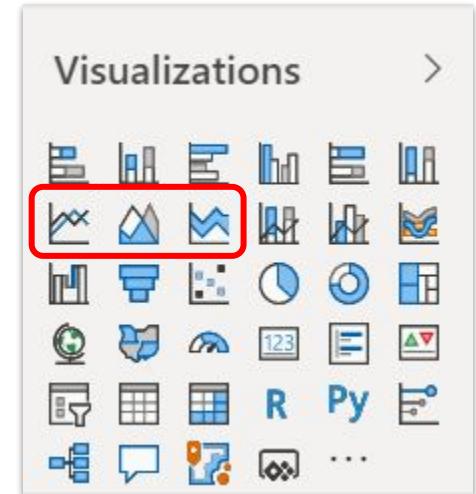




# Line and Area Charts

There are **3 types** of Line and Area charts listed in the second line of the Visualization pane:

- **Line chart** Each category is assigned its own line.
- **Area chart** Lines are shaded underneath. Relative transparency highlights areas of overlap.
- **Stacked area chart** Values are additive, and each category retains its own color without overlap.



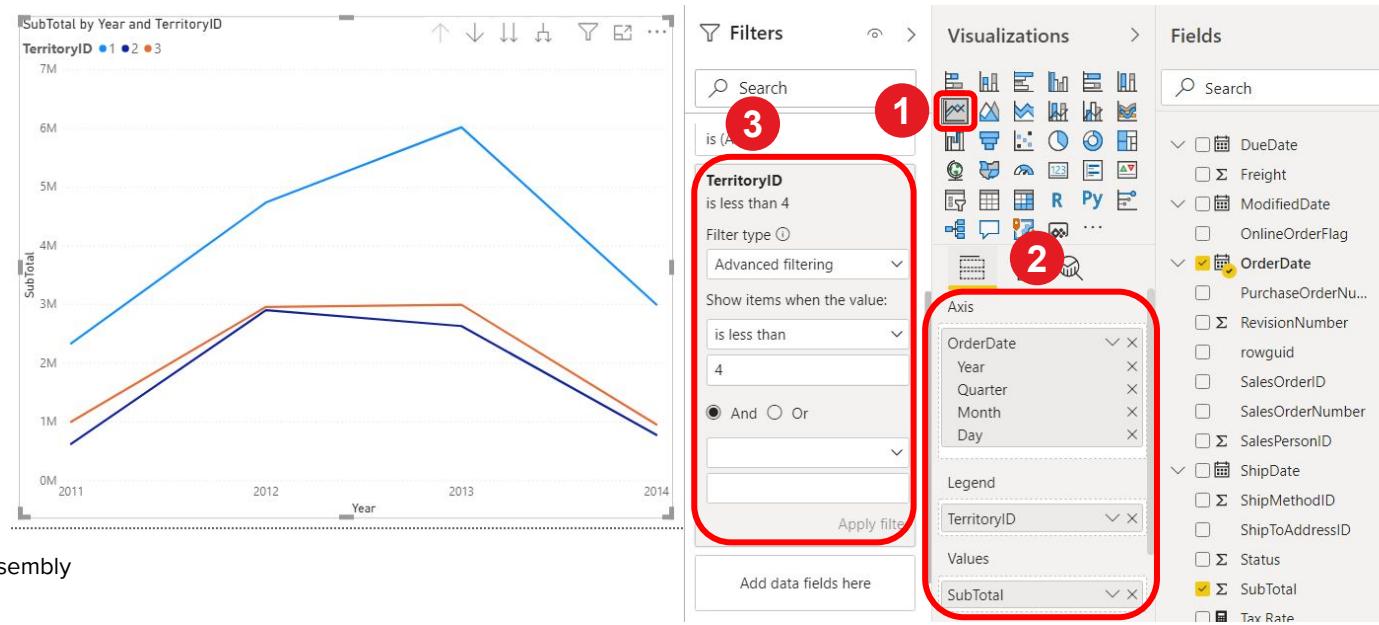


# Guided Walk-Through: Line and Area Charts

In the Report view, open a new page and click **Line chart**.

Drag **OrderDate** to *Axis*, **TerritoryID** to *Legend*, and **SubTotal** to *Values*.

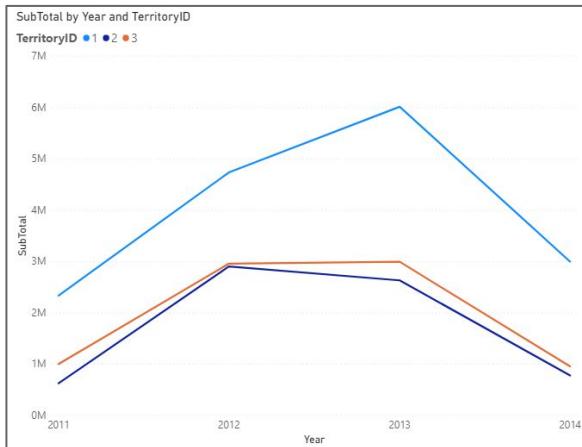
Filter **TerritoryID** to only values less than 4.



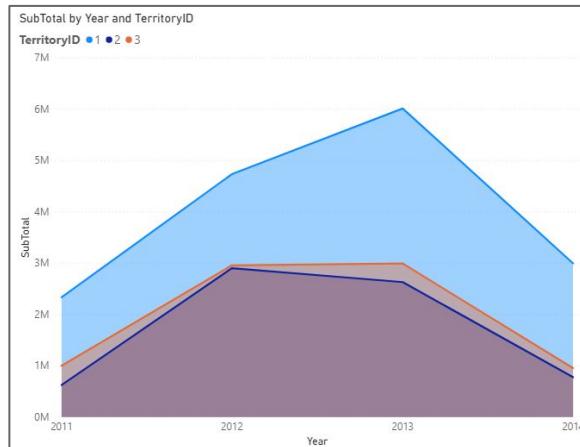


# Guided Walk-Through: Line and Area Charts

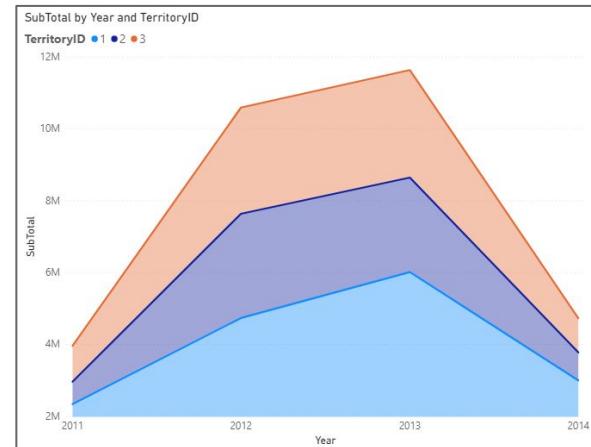
Alternately click between **Line chart**, **Area chart** and **Stacked area chart**.



LINE CHART



AREA CHART



STACKED AREA CHART

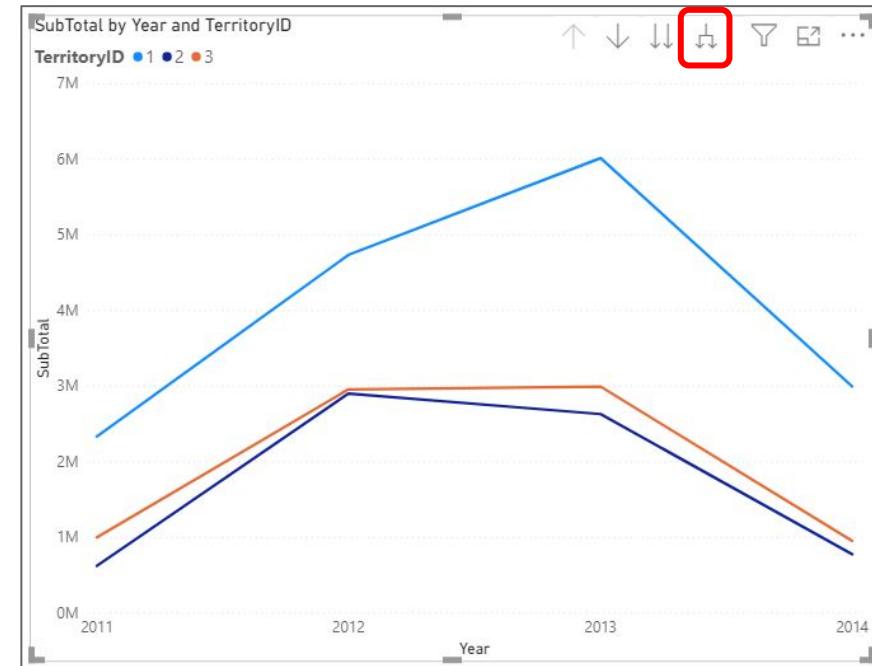


## Guided Walk-Through: Hierarchies

According to our chart, Territory 3 (in orange) consistently outsells Territory 2. However, we're only observing four data points (2011, 2012, 2013, 2014).

Fortunately we can obtain more data with just one click!

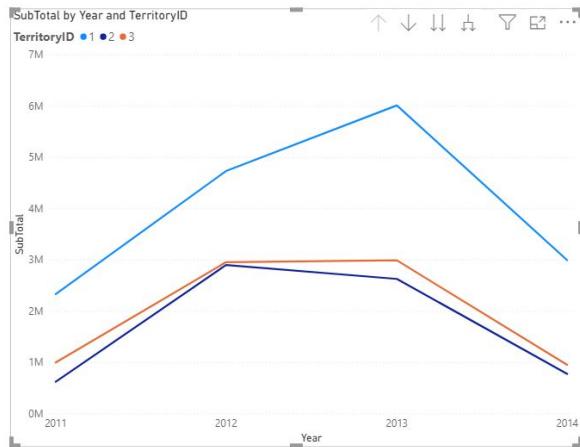
With the chart selected, click the forked down-arrow icon. This lets you drill down the hierarchy from year to quarter. Click it again to drill down to month.





# Guided Walk-Through: Hierarchies

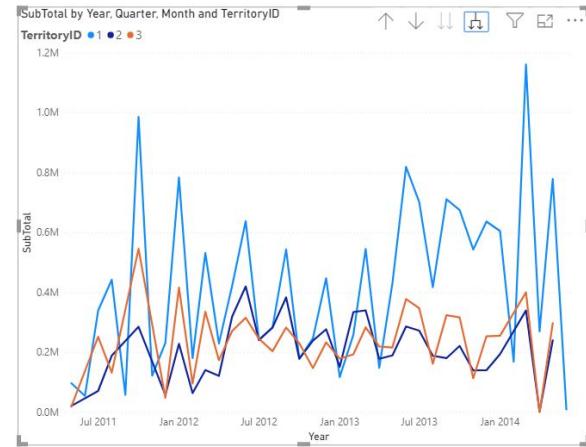
Click the drill down icon to descend through hierarchies. The up-arrow reverses.



YEAR



QUARTER



MONTH

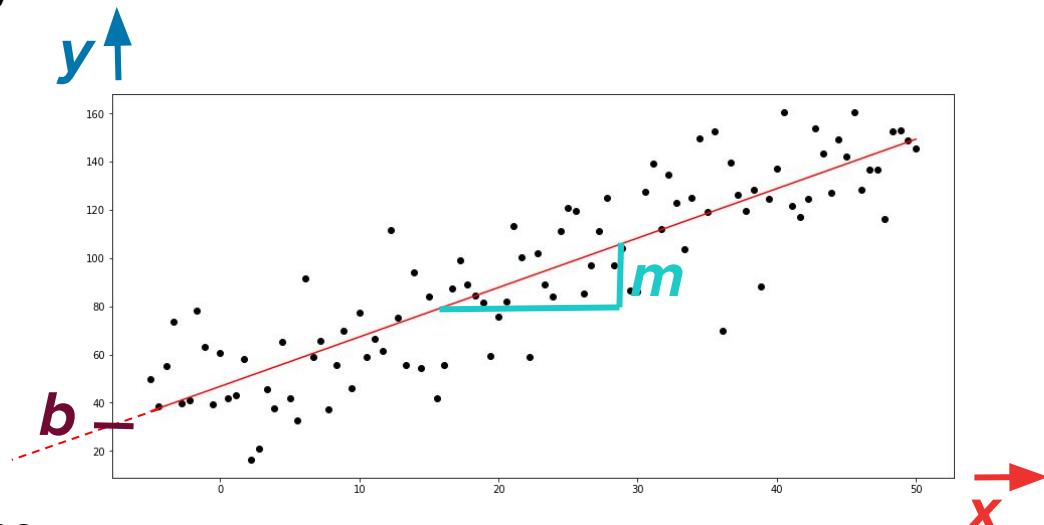
# Quick Review: Linear Regression

A regression line is represented by:

$$y = mx + b$$

**Key:**

- $y$  = dependent variable (thing being measured)
- $m$  = slope of the red line
- $x$  = independent variable
- $b$  = y-intercept, or the point on the y-axis where the red line will cross

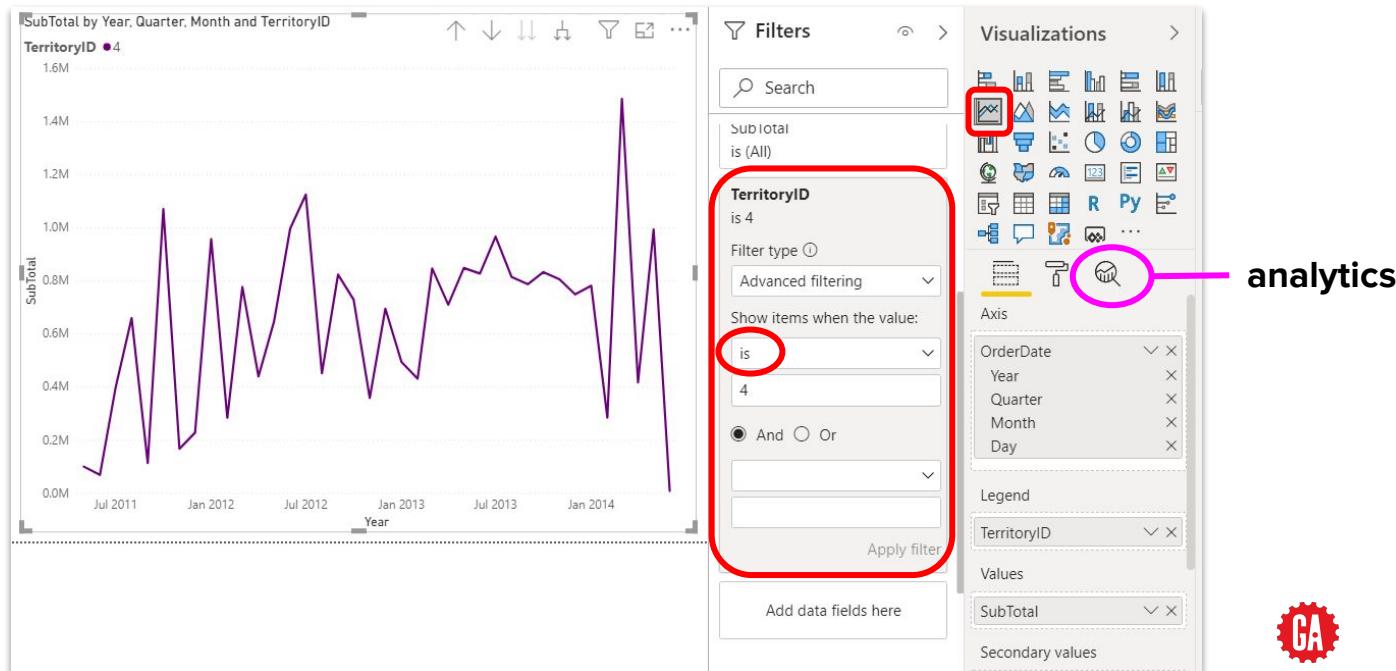




## Guided Walk-Through: Trendlines

Select our chart as a line chart drilled down to month.  
Edit the filter to include *only* Territory 4.

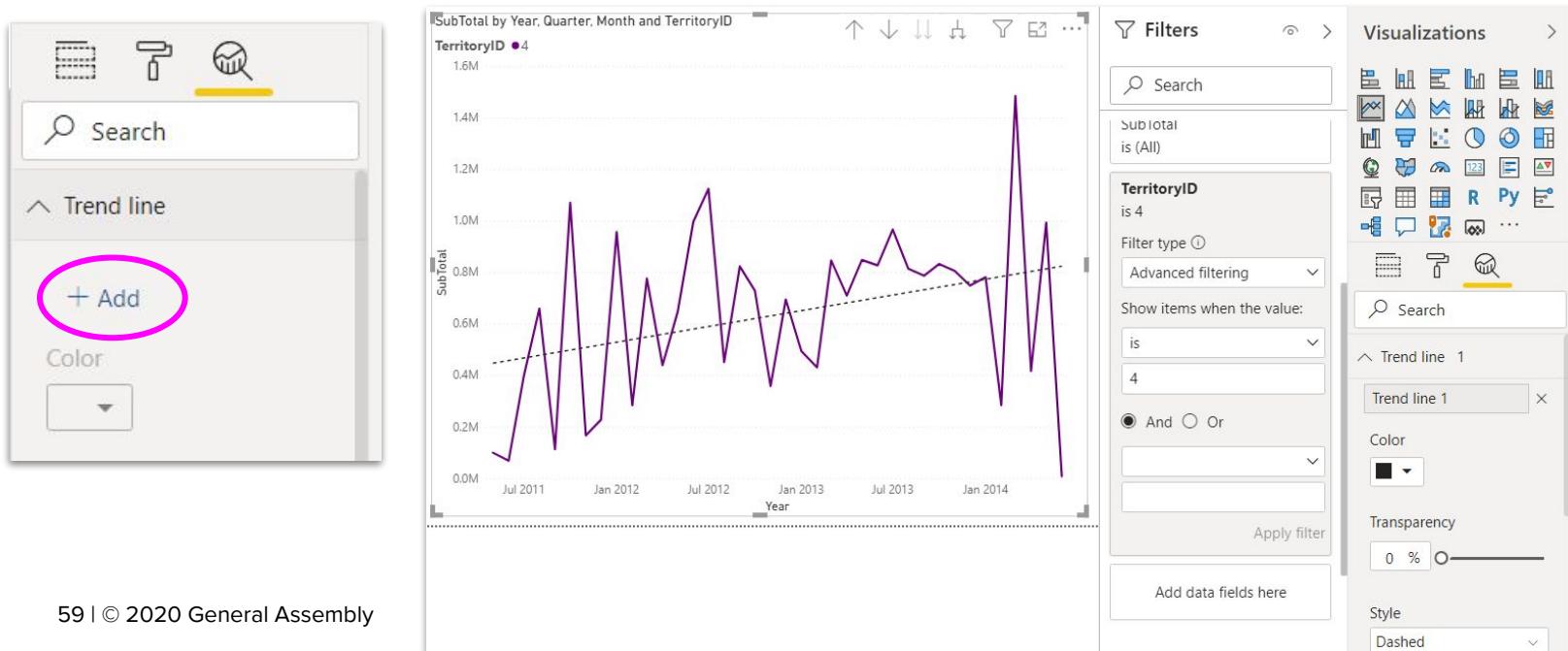
Next, click the  
**analytics** icon.





# Guided Walk-Through: Trendlines

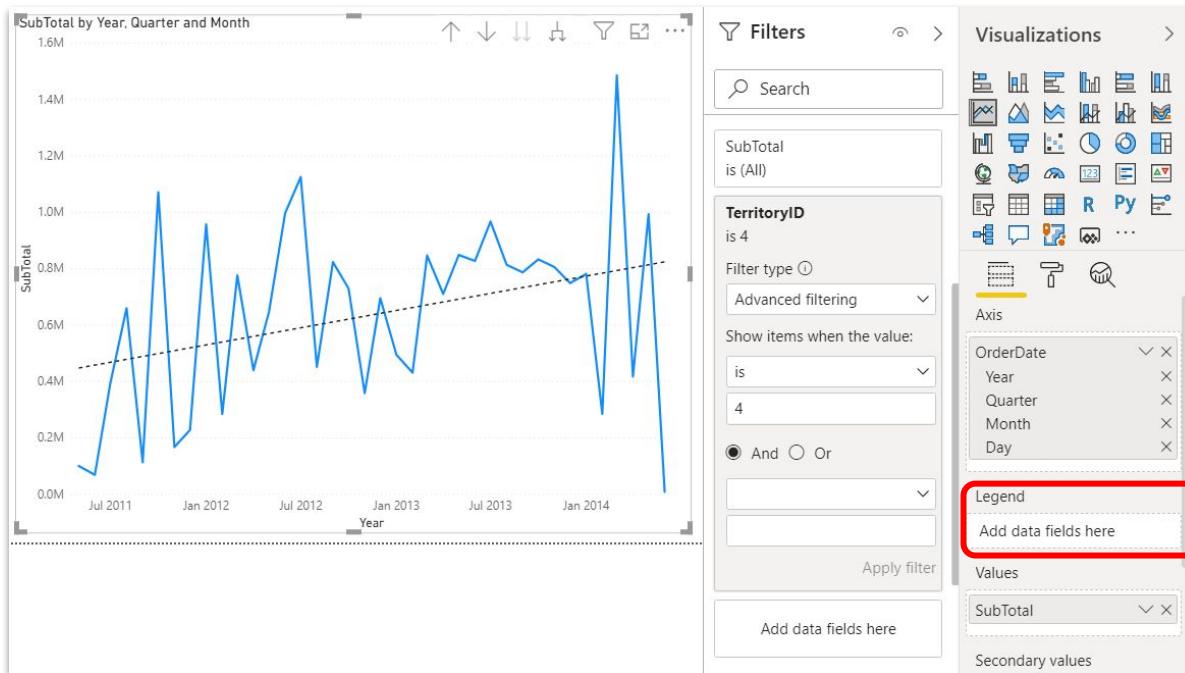
Expand **Trend line** and click **+ Add** to insert a trendline.  
*Feel free to experiment with formatting!*





## Guided Walk-Through: Forecasts

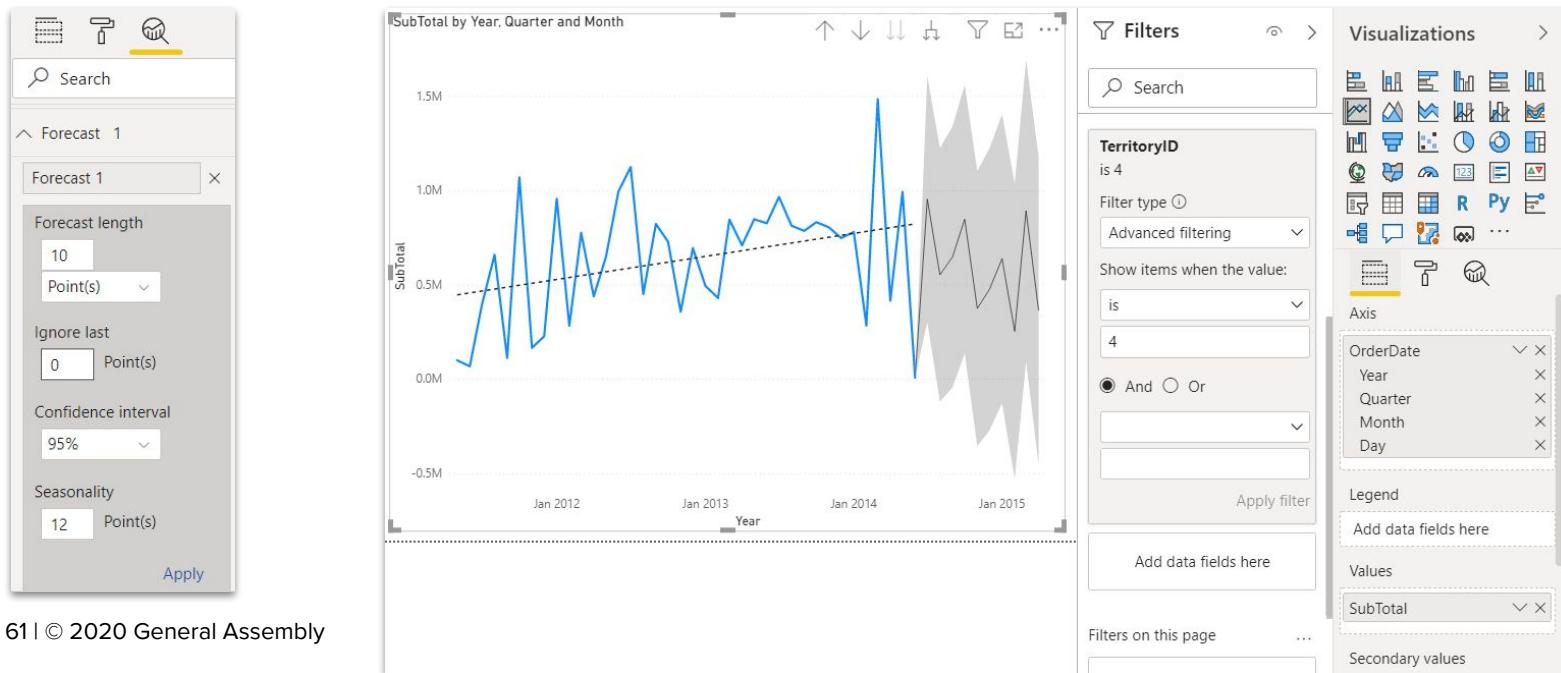
To add a forecast we first need to remove TerritoryID from Legend.  
Next, open the analytics area and add a Forecast.



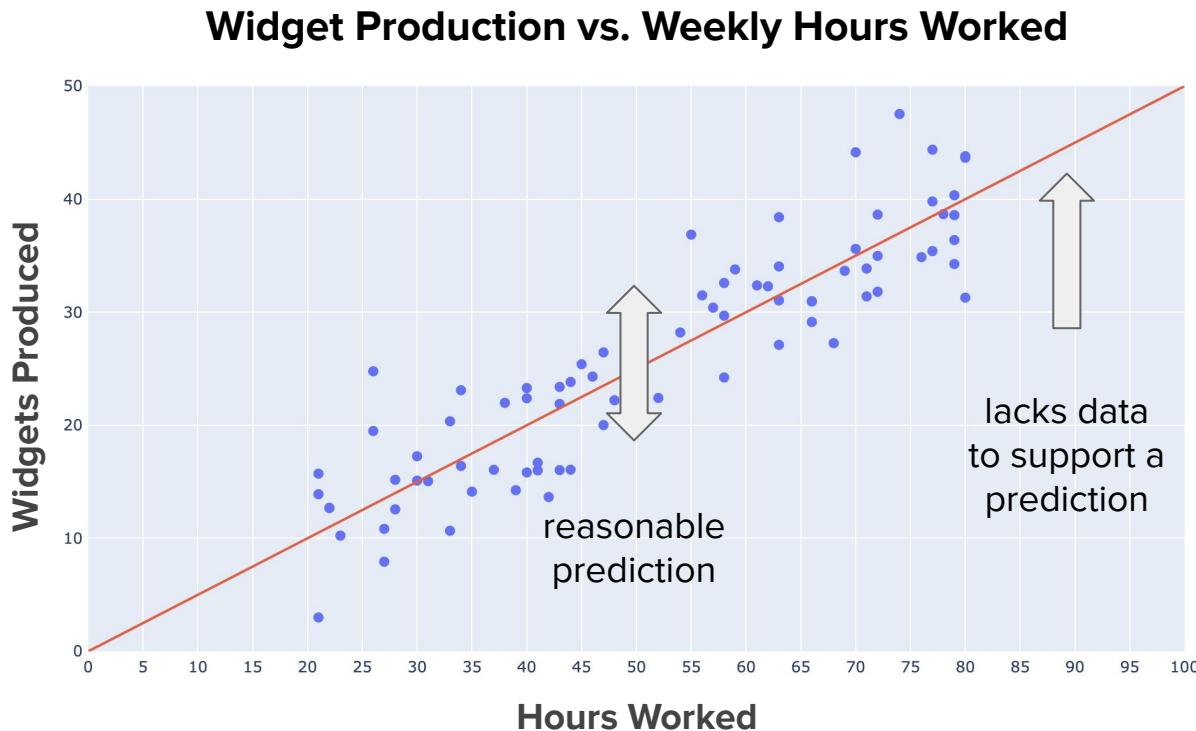


## Guided Walk-Through: Forecasts

Set the Forecast length to 10 data points, the Confidence interval to 95%, and the seasonality to 12 points. Then hit Apply.



# Prediction vs. Forecast



**A regression analysis between two or more variables does NOT support predictions outside the range of data.**

# Stretch Break



I USED TO THINK  
CORRELATION IMPLIED  
CAUSATION.



THEN I TOOK A  
STATISTICS CLASS.  
NOW I DON'T.



SOUNDS LIKE THE  
CLASS HELPED.  
WELL, MAYBE.



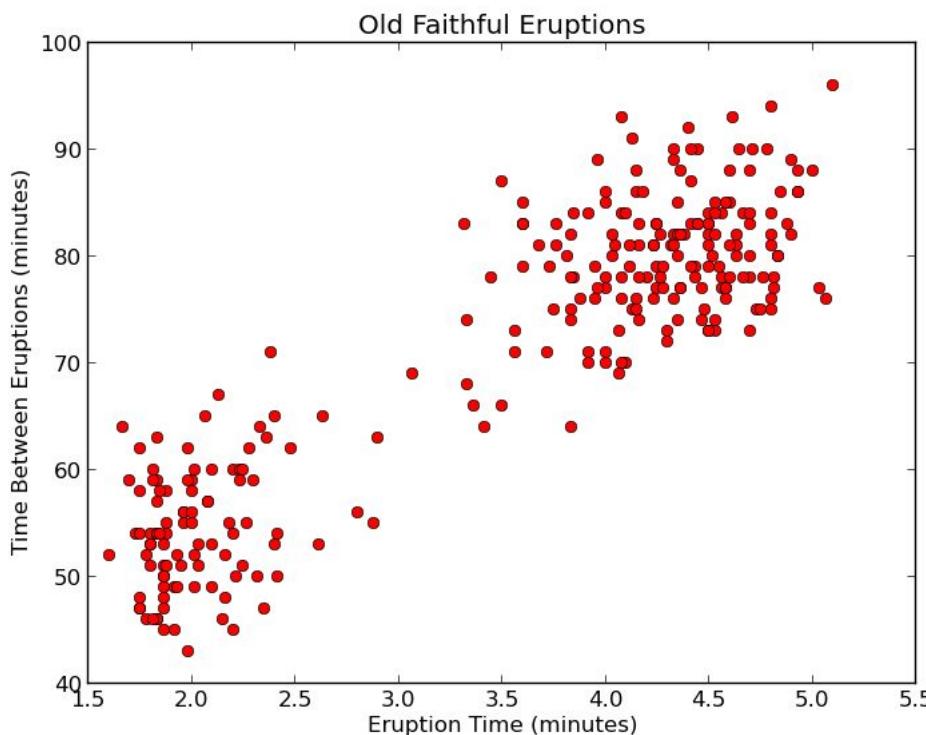
Types of Visualizations

---

# Guided Walkthrough: Scatter Charts

- **Bubble Charts**
- **Play Axis**
- **Reference Lines**

# Scatter Charts

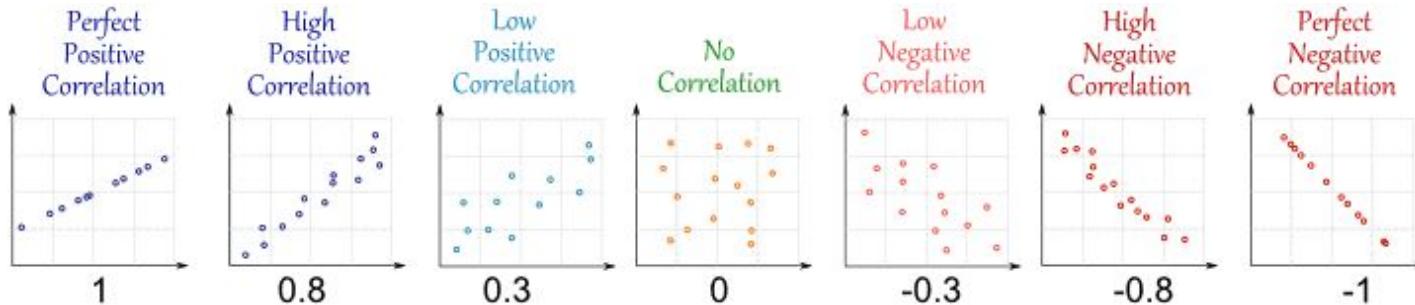


- may uncover **correlations** between variables
- may express a third dimension using the **size** of each dot (Bubble Charts)
- may represent different categories using different **colors** of dots

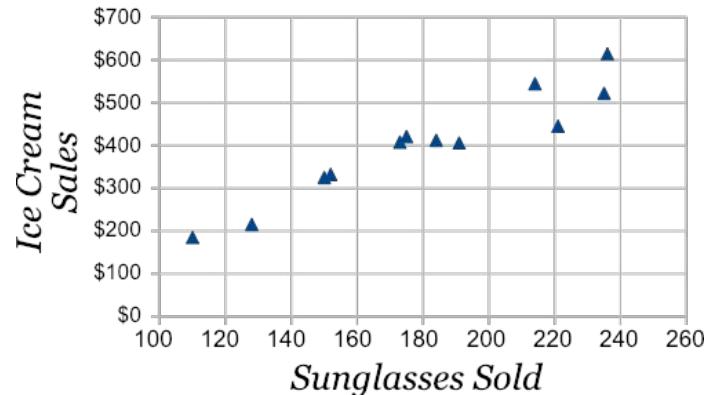
# Correlation

**Correlation** is a measure of how well two variables are related.

Values are bound between -1 (perfect negative) and 1 (perfect positive correlation).



**Analysts beware:** “Correlation does not imply causation.” Many things may be correlated mathematically but have no true bearing on one another!





# Guided Walk-Through: Scatter Charts

Let's try to determine what products drive the highest sales revenue - is it high-priced items, or large quantities of lower-priced items?

In the Report view click on Scatter chart.  
From the **SalesOrderDetail** table, add  
**UnitPrice** to *X Axis*,  
**LineTotal** to *Y Axis*, and  
**ProductID** to *Details*.

The screenshot shows the Power BI Fields pane with the following configuration:

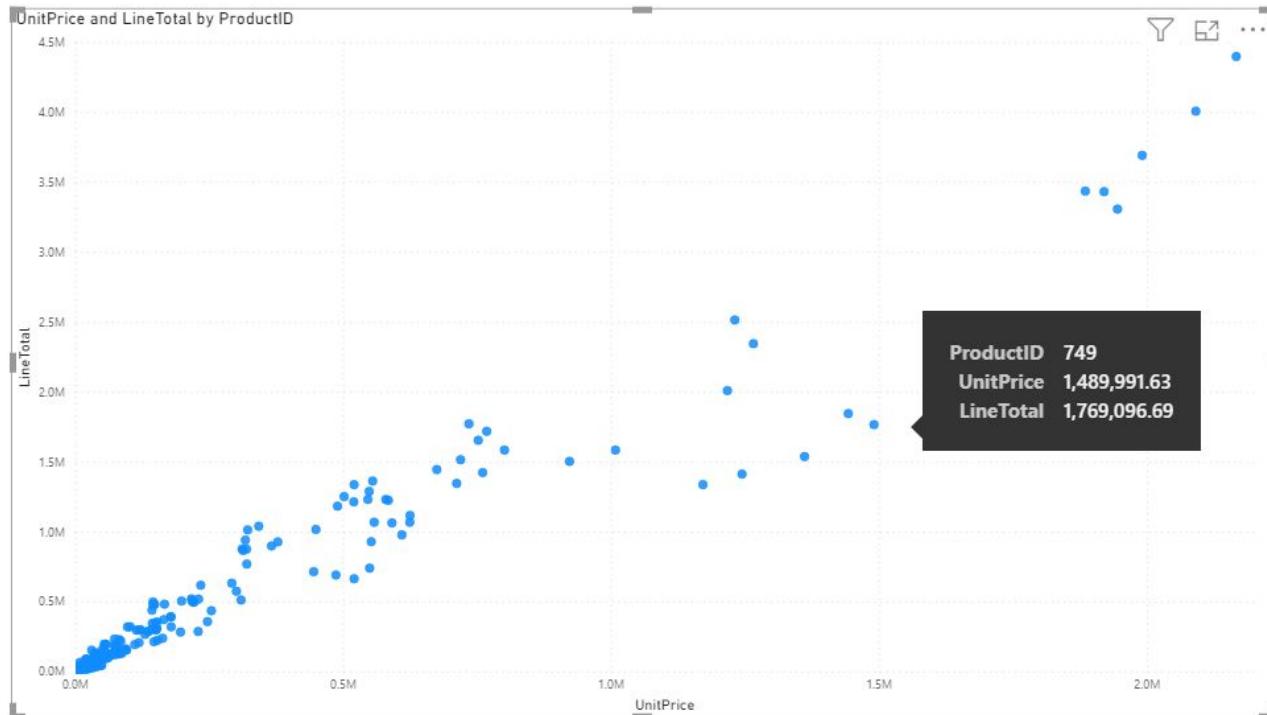
- Visualizations** icon selected in the left pane.
- Fields** pane on the right with a search bar.
- SalesOrderDetail** table expanded in the Fields list.
- ProductID** selected and highlighted in the **Details** section of the Fields pane.
- UnitPrice** selected and highlighted in the **X Axis** section of the Fields pane.
- LineTotal** selected and highlighted in the **Y Axis** section of the Fields pane.



# Guided Walk-Through: Scatter Charts

This doesn't tell us much!

Both UnitPrice and  
LineTotal are shown as  
**sums** for each ProductID





# Guided Walk-Through: Scatter Charts

We want to see LineTotal sums based on *individual* unit price.

Expand the **UnitPrice** dropdown and select Average instead.

The screenshot shows the Power BI Fields pane. On the left, there are two sections: "Details" and "Legend". In the "Details" section, there is a dropdown menu for "ProductID". Below it is a "Legend" section with a placeholder "Add data fields here". Under "X Axis", there is a dropdown menu for "UnitPrice" which is currently set to "UnitPrice". Below "X Axis" is a "Y Axis" section with a dropdown menu for "LineTotal". A red box highlights the "UnitPrice" dropdown, and a red arrow points from this box to the "Fields" pane on the right.

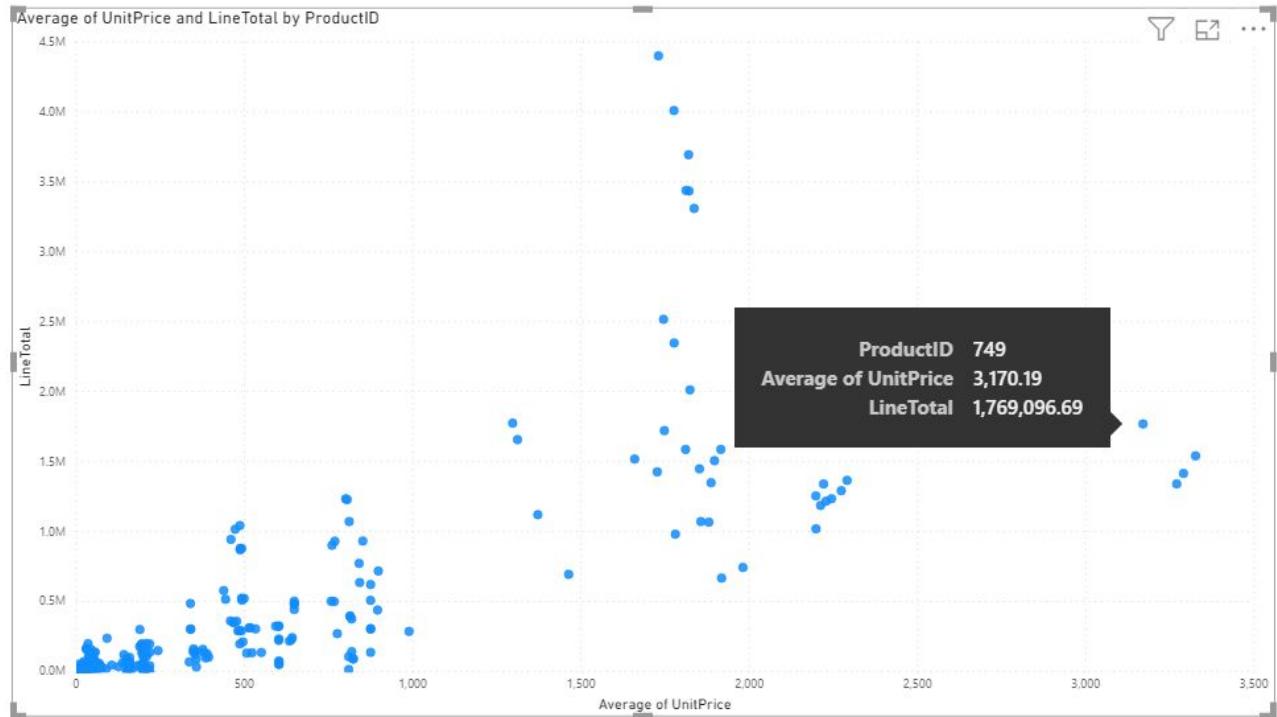
The screenshot shows the Power BI Fields pane with a context menu open over the "UnitPrice" dropdown. The menu has several options: "Remove field", "Rename", "Move to", "Don't summarize", "Sum", "Average" (which is checked with a green checkmark), "Minimum", "Maximum", "Count (Distinct)", "Count", "Standard deviation", "Variance", "Median", "Show value as", "New quick measure", and "UnitPrice" (with a checked checkbox). A red box highlights the "Average" option.



# Guided Walk-Through: Scatter Charts

Now we see a pattern that products with mid-range prices have the largest total sales volumes.

Can we get a sense of how Order Quantity contributes?





# Guided Walk-Through: Bubble Charts

Drag **OrderQty** to **Size** to turn this into a Bubble Chart

The screenshot shows the Power BI editor interface. On the left, the **Fields** pane lists various data fields from the **Sales** and **SalesOrderDetail** tables. In the center, the visualization configuration area displays the current state of a chart:

- Visualizations**: Shows icons for various chart types.
- Add data fields here**: A placeholder for dragging fields.
- X Axis**: Set to "Average of UnitPrice".
- Y Axis**: Set to "LineTotal".
- Size**: Set to "OrderQty".
- Play Axis**: No selection.

A red box highlights the "Size" field in the configuration area, indicating it is the target for the guided walk-through action.

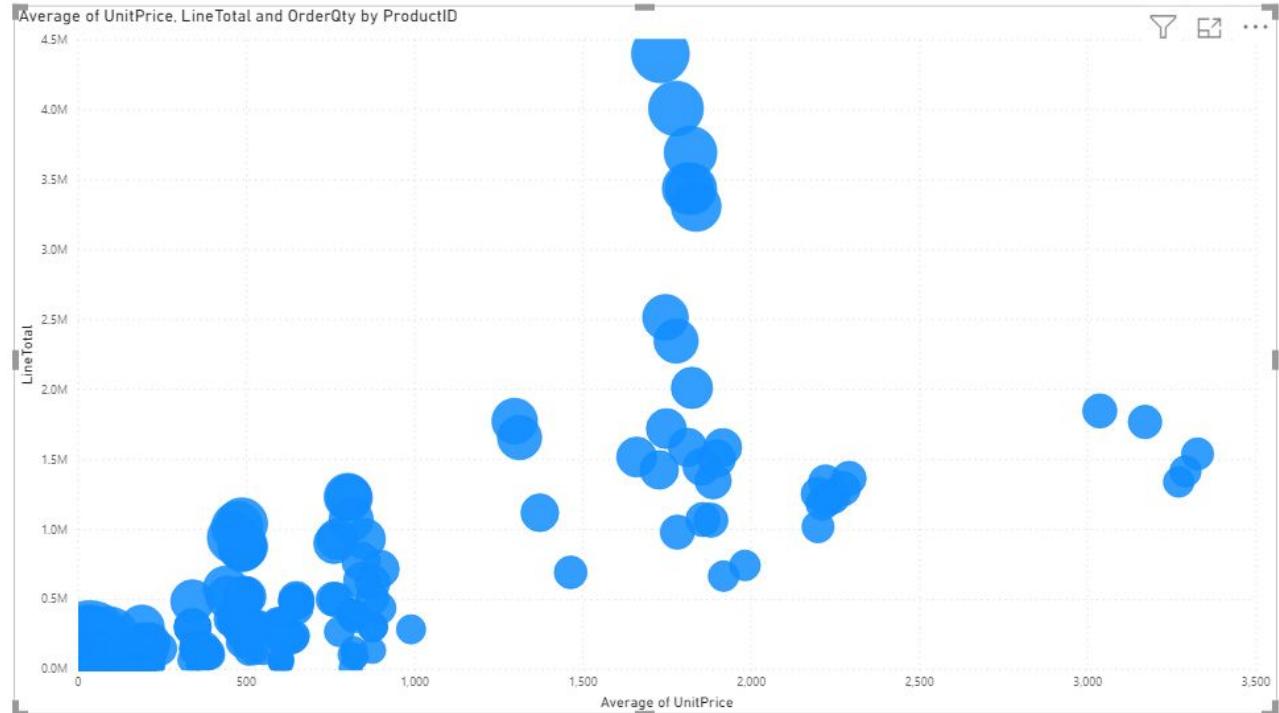
Table	Field	Status
Sales	weight	<input type="checkbox"/>
	WeightUnit	<input type="checkbox"/>
	WeightUnitMeasur...	<input type="checkbox"/>
	CarrierTrackingNum...	<input type="checkbox"/>
SalesOrderDetail	LineTotal	<input checked="" type="checkbox"/>
	ModifiedDate	<input type="checkbox"/>
	OrderQty	<input checked="" type="checkbox"/>
	ProductId	<input checked="" type="checkbox"/>
	rowguid	<input type="checkbox"/>
	SalesOrderDetailID	<input type="checkbox"/>
	SalesOrderID	<input type="checkbox"/>
	SpecialOfferID	<input type="checkbox"/>
	UnitPrice	<input checked="" type="checkbox"/>
	UnitPriceDiscount	<input type="checkbox"/>
SalesOrderHeader	Play Axis	<input type="checkbox"/>



## Guided Walk-Through: Bubble Charts

As expected, dot size increases as you move upward, and for smaller priced items that contribute to revenue.

Bubble charts are not very useful when dots are densely clustered.





## Guided Walk-Through: Bubble Charts

Let's add another dimension to see how different product lines contribute to revenue.

From Production.Product,  
drag **ProductLine** to *Legend*.

The screenshot shows the Power BI Fields pane. On the left, under 'Visualizations', there is a grid of icons representing various chart types. Below the grid, the 'Details' section shows fields: 'ProductID' (selected), 'Legend', and 'ProductLine'. The 'Legend' field is highlighted with a red box. An arrow points from this red box to the 'ProductLine' field in the 'Fields' list on the right. The 'Fields' list contains a search bar and a list of fields categorized by table. The 'Production Product' table is expanded, showing fields like 'Class', 'Color', 'DaysToManufacture', etc., with 'ProductLine' checked (indicated by a yellow checkmark). Another red box highlights the 'ProductLine' field in the 'Fields' list.

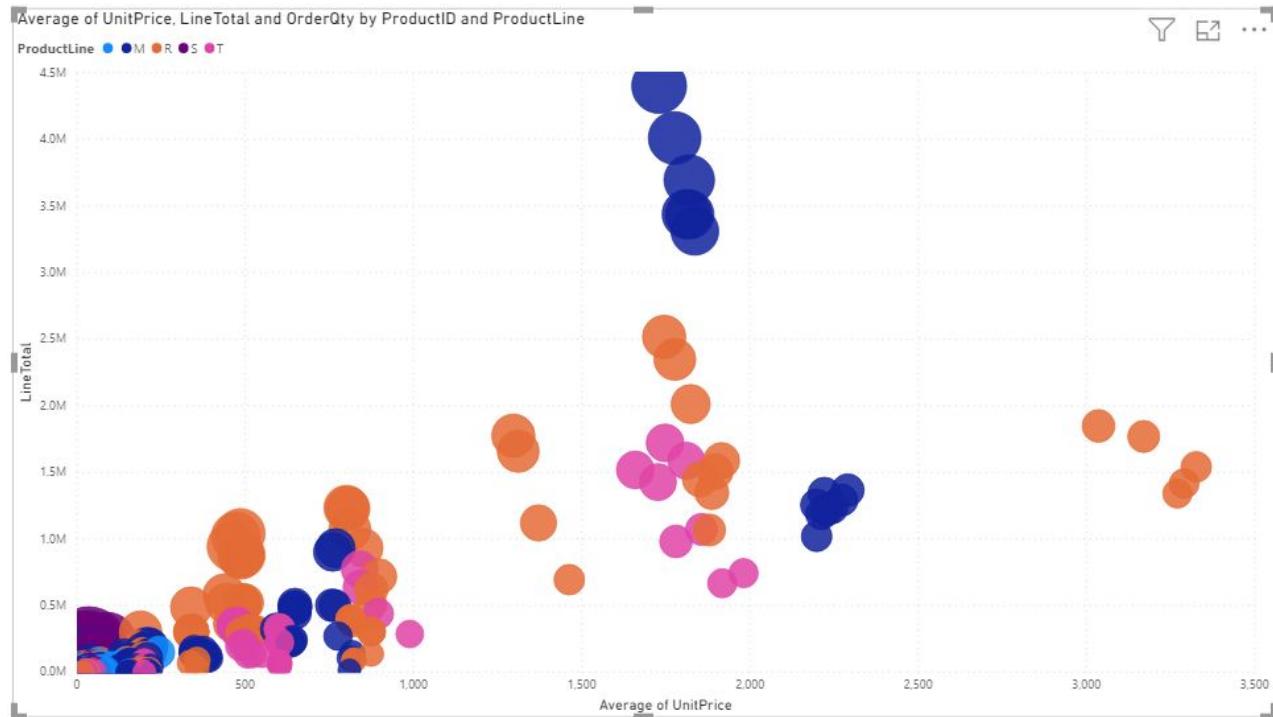
Table	Field	Status
Production Product	Class	<input type="checkbox"/>
	Color	<input type="checkbox"/>
	Σ DaysToManufacture	<input type="checkbox"/>
	DiscontinuedDate	<input type="checkbox"/>
	FinishedGoodsFlag	<input type="checkbox"/>
	Σ ListPrice	<input type="checkbox"/>
	MakeFlag	<input type="checkbox"/>
	ModifiedDate	<input type="checkbox"/>
	Name	<input type="checkbox"/>
	ProductID	<input checked="" type="checkbox"/>
ProductModel	ProductLine	<input checked="" type="checkbox"/>
	ProductModelID	<input type="checkbox"/>



# Guided Walk-Through: Bubble Charts

Here we learn that  
Mountain Bikes are the  
highest individual  
revenue generators.

- M = Mountain
- R = Road
- S = Standard
- T = Touring
- [none]





## Guided Walk-Through: Play Axis

Another powerful tool is Play Axis.

From **Sales.SalesOrderHeader**,  
expand **OrderDate** to display the  
hierarchy, and drag **Year** to *Play Axis*.

The screenshot shows the Power BI Fields pane with the following configuration:

- Visualizations** icon is selected in the left sidebar.
- Fields** section on the right lists various fields:
  - CurrencyRateID (Σ)
  - CustomerID (Σ)
  - DueDate (Due Date)
  - Freight (Σ)
  - ModifiedDate (Modified Date)
  - OnlineOrderFlag
  - Quarter (Quarter)
  - Month (Month)
  - Day (Day)
  - PurchaseOrderNumber (Purchase Order Number)
  - RevisionNumber (Revision Number)
  - rowguid
  - SalesOrderID
- Play Axis** section in the bottom-left:
  - OrderDate (Order Date) is expanded, showing its hierarchy.
  - Year (Year) is selected and dragged into the Play Axis area.
- Tooltips** section: Add data fields here.
- Drill through** section: None.

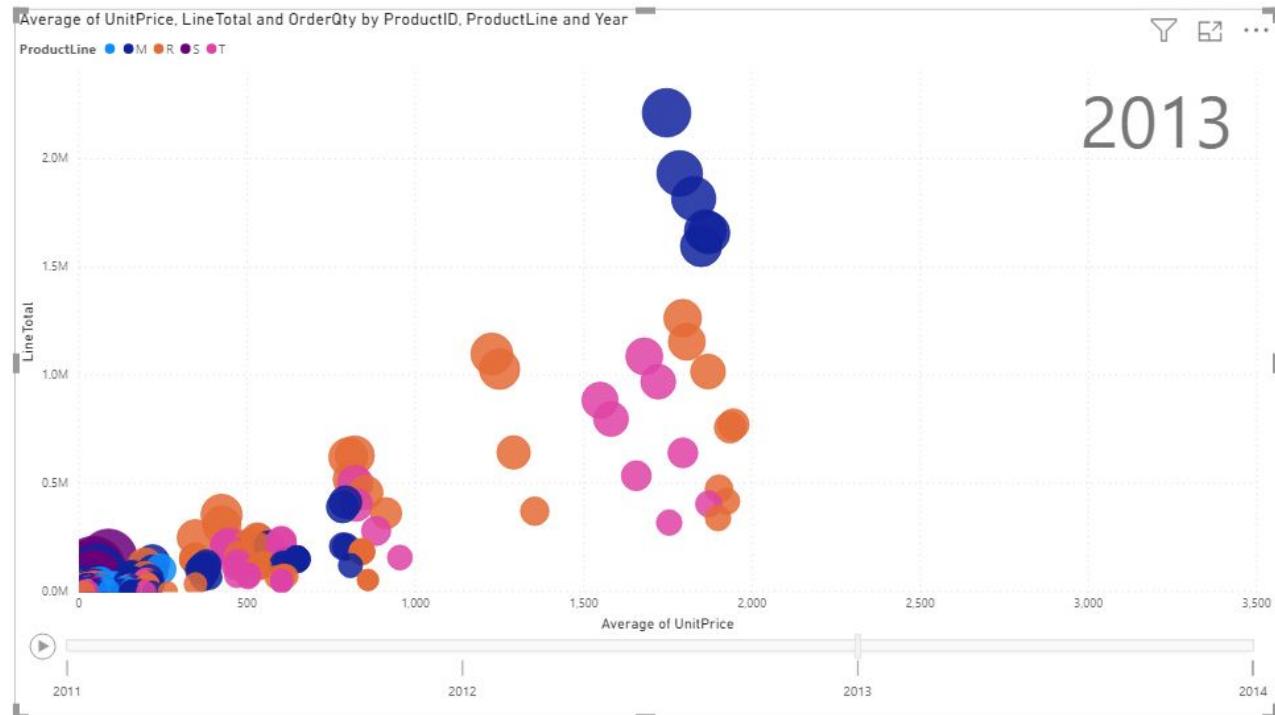


## Guided Walk-Through: Play Axis

Move the slider to see different years.

Put the slider on 2011 and hit the Play button.

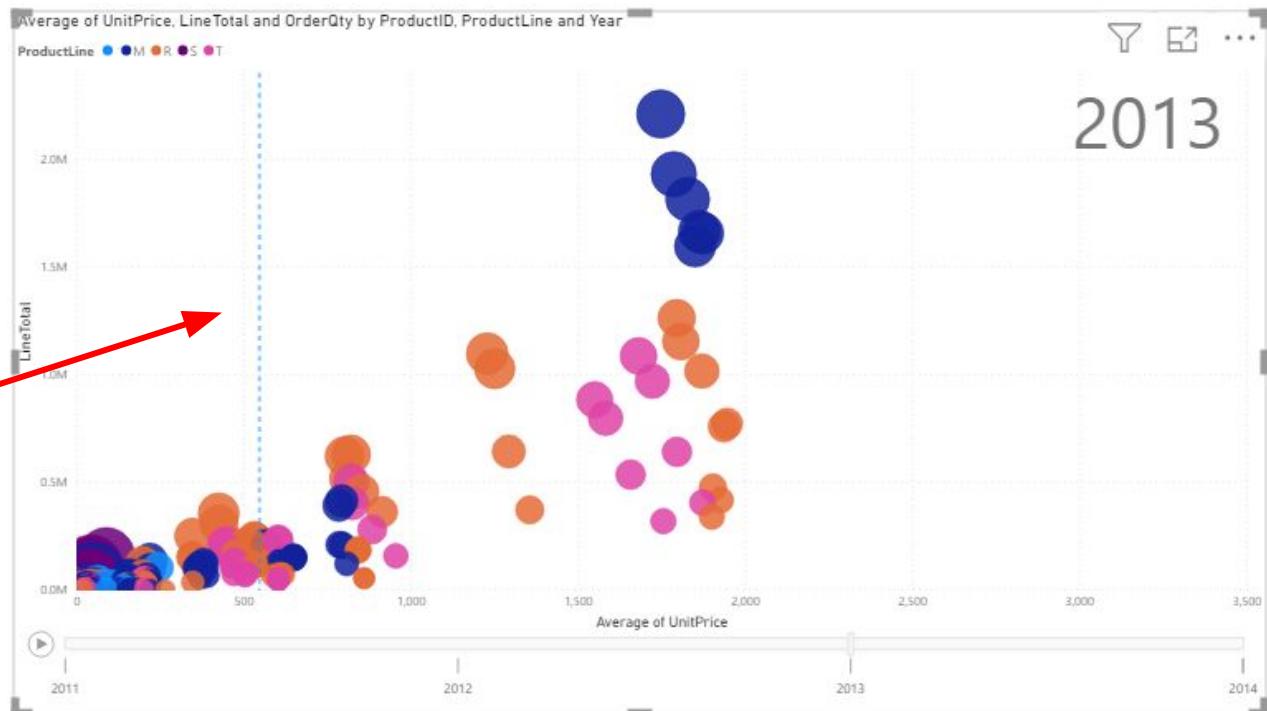
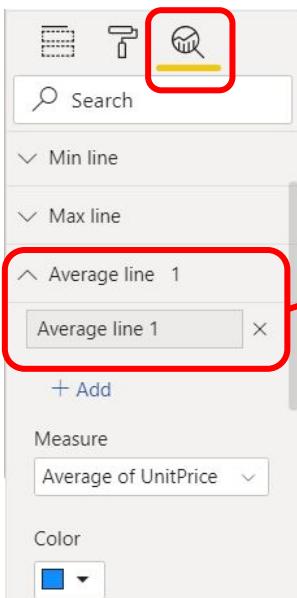
What observations can you make?





# Guided Walk-Through: Reference Lines

Open **analytics** and  
add an *Average line*.



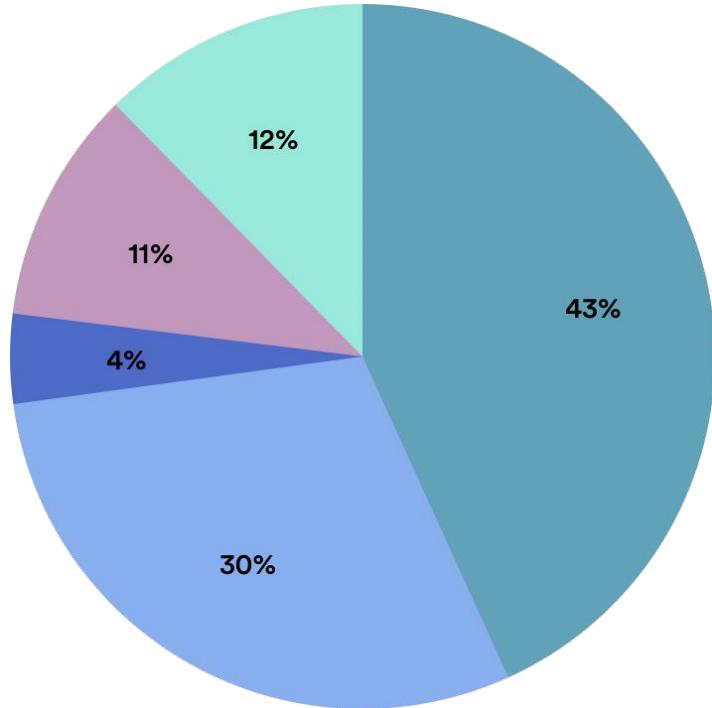
Types of Visualizations

---

# Guided Walkthrough: Pie, Donut and Treemap Charts

- Themes

# Pie Charts



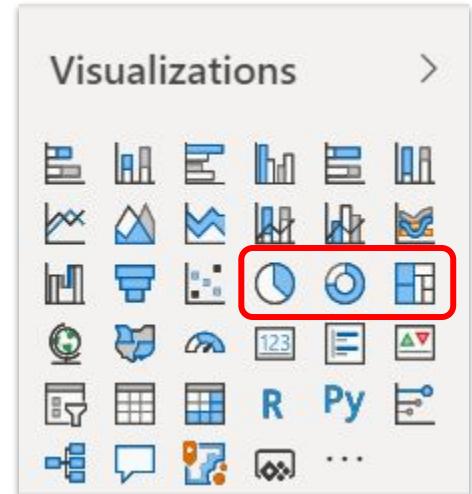
- Show a proportion of segments that comprise a whole.
- Segments must add up to 100%
- Work best for values around 25%, 30% or 75%.
- Easier for readers to spot these percentages in a pie chart than in a stacked bar or column chart.
- Difficult to clearly recognize relative size of segments.





# Pie, Donut and Treemap Charts

- These compare the **relative percentage** of each category towards the whole, or 100%
- Good for comparing a handful of categories
- Too many categories make them difficult to read and interpret





## Guided Walk-Through:

# Pie, Donut and Treemap Charts

Let's improve our understanding of which products contribute most toward revenue.

In the Report view click on Pie chart.

From **SalesOrderDetail**, add  
**LineTotal** to *Values*.

From **Production.Product**, add  
**ProductLine** to *Legend*.

The screenshot shows the Power BI visualization editor interface. On the left, there is a 'Visualizations' pane with various chart icons, one of which is a pie chart icon, highlighted with a red box. To the right is a 'Fields' pane with a search bar and a list of fields from the 'Sales SalesOrderDetail' table. The 'LineTotal' field is checked and highlighted with a red box. Below these panes is the main editing area. It contains a 'Legend' section with a dropdown set to 'ProductLine', also highlighted with a red box. Below it is a 'Details' section with a placeholder 'Add data fields here'. At the bottom is a 'Values' section with a dropdown set to 'LineTotal', also highlighted with a red box. A red arrow points from the checked 'LineTotal' in the Fields pane down to its corresponding dropdown in the Values section.

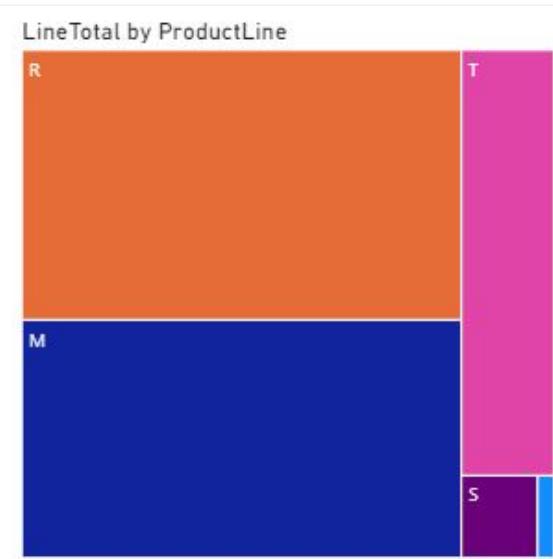
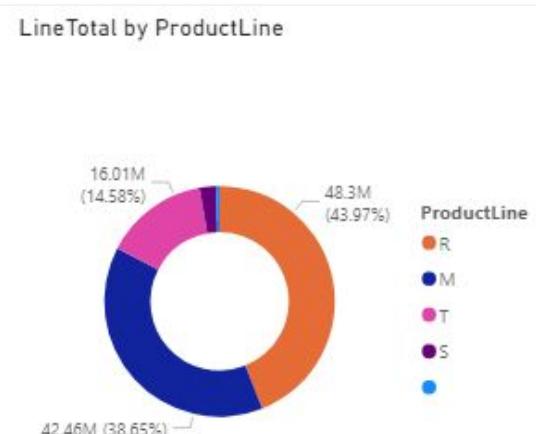
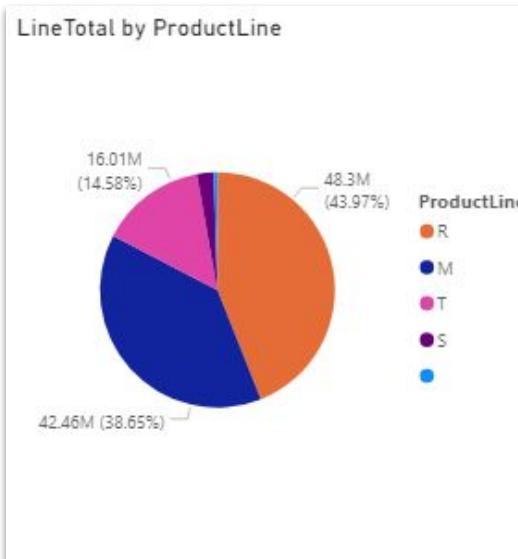
Fields
SizeUnitMeasureCo...
<input checked="" type="checkbox"/> Σ StandardCost
<input type="checkbox"/> Style
<input checked="" type="checkbox"/> Σ Weight
<input type="checkbox"/> WeightUnit
<input type="checkbox"/> WeightUnitMeasur...
<input checked="" type="checkbox"/> Sales SalesOrderDetail
CarrierTrackingNum...
<input checked="" type="checkbox"/> Σ LineTotal
<input type="checkbox"/> ModifiedDate
<input checked="" type="checkbox"/> Σ OrderQty
<input type="checkbox"/> ProductID
<input type="checkbox"/> rowguid
<input checked="" type="checkbox"/> Σ SalesOrderDetailID
<input type="checkbox"/> SalesOrderID



## Guided Walk-Through:

# Pie, Donut and Treemap Charts

Duplicate the pie chart and turn the others into a donut chart and a treemap.  
Here we see that Road bikes contributed the most to overall revenue.





# Guided Walk-Through: Themes

You can change  
the **theme** from  
the View ribbon

The screenshot shows the Microsoft Power BI ribbon with the 'View' tab selected. Below the ribbon, the 'Themes' section displays three different theme variations: 'Bloom', 'LineTotal by ProductLine', and 'LineTotal by ProductLine'. Each theme is represented by a pie chart and a corresponding bar chart. The 'Bloom' theme has a purple and teal color scheme, while the other two themes have a purple, teal, blue, and orange/red color scheme.

ProductLine	Bloom	LineTotal by ProductLine (1)	LineTotal by ProductLine (2)
R	48.3M (43.97%)	16.01M (14.58%)	42.46M (38.65%)
M	42.46M (38.65%)	48.3M (43.97%)	16.01M (14.58%)
T	16.01M (14.58%)	48.3M (43.97%)	42.46M (38.65%)
S	1.6M (1.4%)	1.6M (1.4%)	1.6M (1.4%)

Advanced Analytics

---

# Specialty Charts



# Specialty Charts

**We've only scratched the surface of what's available.**

Included with Power BI Desktop we have choropleths, combo charts, funnels, gauges, ribbon charts, etc. See [\*\*Visual types in Power BI\*\*](#) for full details.

**But there's more!**

Additional visuals, including those developed by third parties, are available from [\*\*Microsoft AppSource\*\*](#).

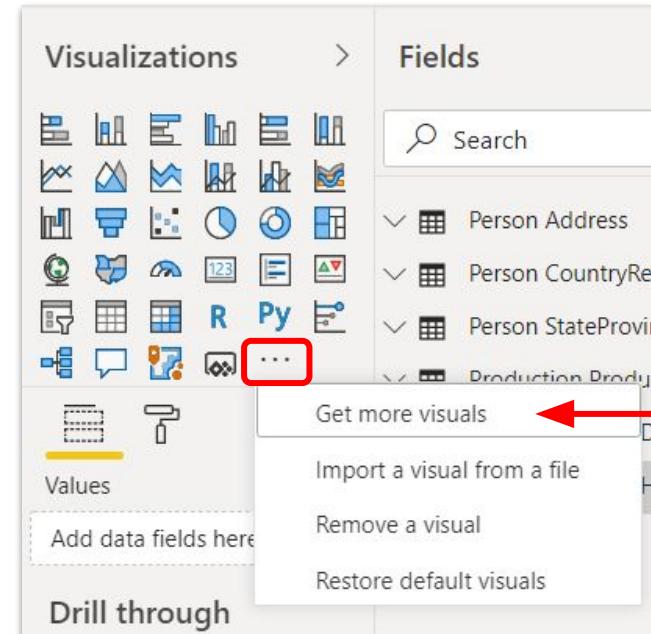
...and Power BI makes it very easy to get them!



## Guided Walk-Through: Get More Visuals

Open a new page in Report view, and click the **ellipsis** under Visualizations

Select *Get more visuals* to open Microsoft AppSource





# Guided Walk-Through: Get More Visuals

From here you can **search** for  
visuals or **browse** by category.

Power BI Visuals

AppSource | My organization

Add-ins may access personal and document information. By using an add-in, you agree to its Permissions, License Terms and Privacy Policy.

Search

Suggested for you ▾

Category

All

Advanced Analytics

Data Visualizations

Editor's Picks

Filters

Gauges

Infographics

KPIs

Maps

Power BI Certified

Time

Bullet Chart A bar chart with extra visual elements to provide additional context. Useful for tracking goals  
★★★★★

Word Cloud Create a fun visual from frequent text in your data  
★★★★★

Infographic Designer Beautify your reports with easy-to-create infographics  
★★★★★

Visio Visual Bring your business activities to life in ways that only Microsoft Visio diagrams can visualize

Add Add Add Add



# Guided Walk-Through: Get More Visuals

Search for **BoxandWhisker.pbviz** and click Add

The screenshot shows the Power BI Visuals AppSource page. A red arrow points from the search bar to the first result, which is highlighted with a red border. The result is titled "Box and Whisker chart by MAQ Software..." and includes a description: "Display data distribution quartiles in a box plot." Below the title are four star ratings and a yellow "Add" button. The search bar contains the text "Box and Whisker". The page also features a sidebar with categories like Category, Editor's Picks, All, Advanced Analytics, Data Visualizations, Filters, Gauges, Infographics, KPIs, Maps, Power BI Certified, and Time.

Power BI Visuals

AppSource | My organization

Add-ins may access personal and document information. By using an add-in, you agree to its Permissions, License Terms and Privacy Policy.

Manage Cookies

Sort by: Recommended ▾

Box and Whisker

Category

Editor's Picks

All

Advanced Analytics

Data Visualizations

Filters

Gauges

Infographics

KPIs

Maps

Power BI Certified

Time

Box and Whisker chart by MAQ Software...

Display data distribution quartiles in a box plot.  
★★★★★

Box and Whisker chart

Visualize a dataset in an effective way as a five-number summary (mean, median, quartiles, min/max).  
★★★★★

DataText Box

Apply to data labels, summary reports, data templates, and so on, once and for all, speed up data tr  
★★★★★

Strip Plot

Get to know your data without hiding critical details: the Strip Plot



## Guided Walk-Through: Get More Visuals

Now we have a new visual.

Note: this is only available to the current Data Model.

If you want to add a Box Plot to a different .pbix file, you would have to import it there.

The screenshot shows the Power BI interface with the 'Filters' and 'Visualizations' panes open. The 'Visualizations' pane on the right contains a grid of icons representing various chart types. One specific icon, which is a box plot, is highlighted with a pink square border. Below the grid, there are sections for 'Filters on this page' and 'Filters on all pages', each with an 'Add data fields here' button. At the bottom of the pane, there are buttons for 'Values' and a 'P' icon.

Specialty Charts

---

# Guided Walkthrough: Box and Whisker Chart

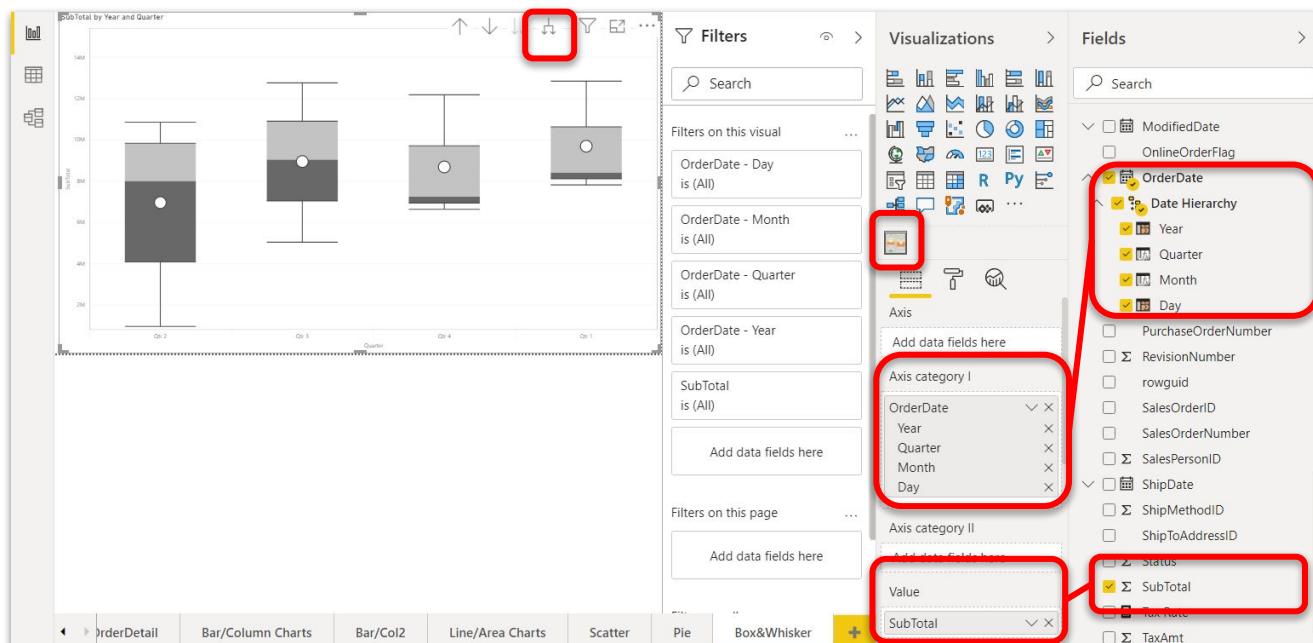


## Guided Walk-Through:

# Box and Whisker Chart

Open a new report page and click **Box and Whisker**. From Sales.SalesOrderHeader, drag **OrderDate** to *Axis category I*, and **SubTotal** to *Value*.

In the report pane, click the forked arrow to drill down to Quarter.



Advanced Analytics

---

# Conditional Formatting for Tables

# Conditional Formatting for Tables

- You can specify **customized cell colors**, including color gradients, based on field values.
- You can also represent cell values with data bars or KPI icons, or as active web links
- You can apply conditional formatting to any text or data field, as long as you base the formatting on a field that has numeric, color name or hex code, or web URL values.



## Guided Walk-Through:

# Conditional Formatting for Tables

Create a new report page  
and add a **Table**.

Add these fields to *Values*:

From Person.CountryRegion  
add **CountryRegionCode**

From Sales.SalesOrderDetail  
add **LineTotal** and **OrderQty**

The screenshot shows the Power BI desktop environment. On the left, there is a table visualization displaying data from the Person.CountryRegion and Sales.SalesOrderDetail tables. The table has columns: CountryRegionCode, LineTotal, and OrderQty. The last row is a summary row labeled 'Total' with values 109,846,381.40 and 274914 respectively. In the center, there is a 'Visualizations' pane with various chart icons. To the right, there is a 'Fields' pane. The 'Values' section of the Fields pane is highlighted with a red box and contains three entries: 'CountryRegionCode', 'LineTotal', and 'OrderQty'. The 'CountryRegionCode' entry has a yellow checkmark next to it. The 'LineTotal' and 'OrderQty' entries have a yellow checkmark and a summation symbol ( $\Sigma$ ) next to them. The 'Filters' section of the Fields pane is also highlighted with a red box and shows a grid icon with a yellow checkmark.

CountryRegionCode	LineTotal	OrderQty
AU	10,655,335.96	18293
CA	16,355,770.45	49381
DE	4,915,407.60	13143
FR	7,251,555.65	19906
GB	7,670,721.04	20099
US	62,997,590.71	154092
<b>Total</b>	<b>109,846,381.40</b>	<b>274914</b>



## Guided Walk-Through: Conditional Formatting - Icons

In the *Values* list, click the down-arrow next to **LineTotal**.

Select *Conditional formatting*, then *Icons*

The screenshot shows the Power BI ribbon with the 'Values' list selected. A context menu is open over the 'LineTotal' item in the 'Values' list. The menu items include: Remove field, Rename, Move, Conditional formatting (with a red arrow pointing to it), Remove conditional formatting, Don't summarize, Sum (selected with a checkmark), Average, Minimum, Maximum, Count (Distinct), Count, Standard deviation, Variance, Median, Show value as, New quick measure, and a checked checkbox for LineTotal.

Background color  
Font color  
Data bars  
Icons  
Web URL

Values

CountryRegionCode  
LineTotal  
OrderQty

Remove field  
Rename  
Move  
Conditional formatting  
Remove conditional formatting  
Don't summarize  
✓ Sum  
Average  
Minimum  
Maximum  
Count (Distinct)  
Count  
Standard deviation  
Variance  
Median  
Show value as  
New quick measure  
Σ LineTotal



# Guided Walk-Through: Conditional Formatting - Icons

Select a *Style*,  
and set parameters  
(we set Percent values  
at 0%, 10% and 50%+)

Hit OK

Icons - *LineTotal*

Format by: Rules      Apply to: Values only

Based on field: Sum of LineTotal      Summarization: Sum

Icon layout: Left of data      Icon alignment: Top

Style:

Rules:

If value is greater than or equal to 0 Percent and is less than 10 Percent then	
If value is greater than or equal to 10 Percent and is less than 50 Percent then	
If value is greater than or equal to 50 Percent and is less than or equal to 100 Percent then	

[Learn more](#) OK Cancel



## Guided Walk-Through:

# Conditional Formatting - Icons

You should see something like this:

CountryRegionCode	LineTotal	OrderQty
AU	✖ 10,655,335.96	18293
CA	⚠ 16,355,770.45	49381
DE	✖ 4,915,407.60	13143
FR	✖ 7,251,555.65	19906
GB	✖ 7,670,721.04	20099
US	✓ 62,997,590.71	154092
<b>Total</b>	<b>109,846,381.40</b>	<b>274914</b>



# Conditional Formatting - Colors

Now see if you can add conditional formatting to **OrderQty**.

Set three different background colors (your choice) based on:

- fewer than 15,000 units
- between 15,000 and 60,000
- greater than 60,000 units

CountryRegionCode	LineTotal	OrderQty
AU	✖ 10,655,335.96	18293
CA	⚠ 16,355,770.45	49381
DE	✖ 4,915,407.60	13143
FR	✖ 7,251,555.65	19906
GB	✖ 7,670,721.04	20099
US	✓ 62,997,590.71	154092
<b>Total</b>	<b>109,846,381.40</b>	<b>274914</b>



# Conditional Formatting - Colors

The result should look something like this:

CountryRegionCode	LineTotal	OrderQty
AU	✖ 10,655,335.96	18293
CA	⚠ 16,355,770.45	49381
DE	✖ 4,915,407.60	13143
FR	✖ 7,251,555.65	19906
GB	✖ 7,670,721.04	20099
US	✓ 62,997,590.71	154092
<b>Total</b>	<b>109,846,381.40</b>	<b>274914</b>

Advanced Analytics

---

# Activity





Solo Exercise:

## Build a Histogram

15 minutes



You've been asked to build a **histogram** that shows the relative frequencies by **SubTotal** of all the orders taken in September, 2013.

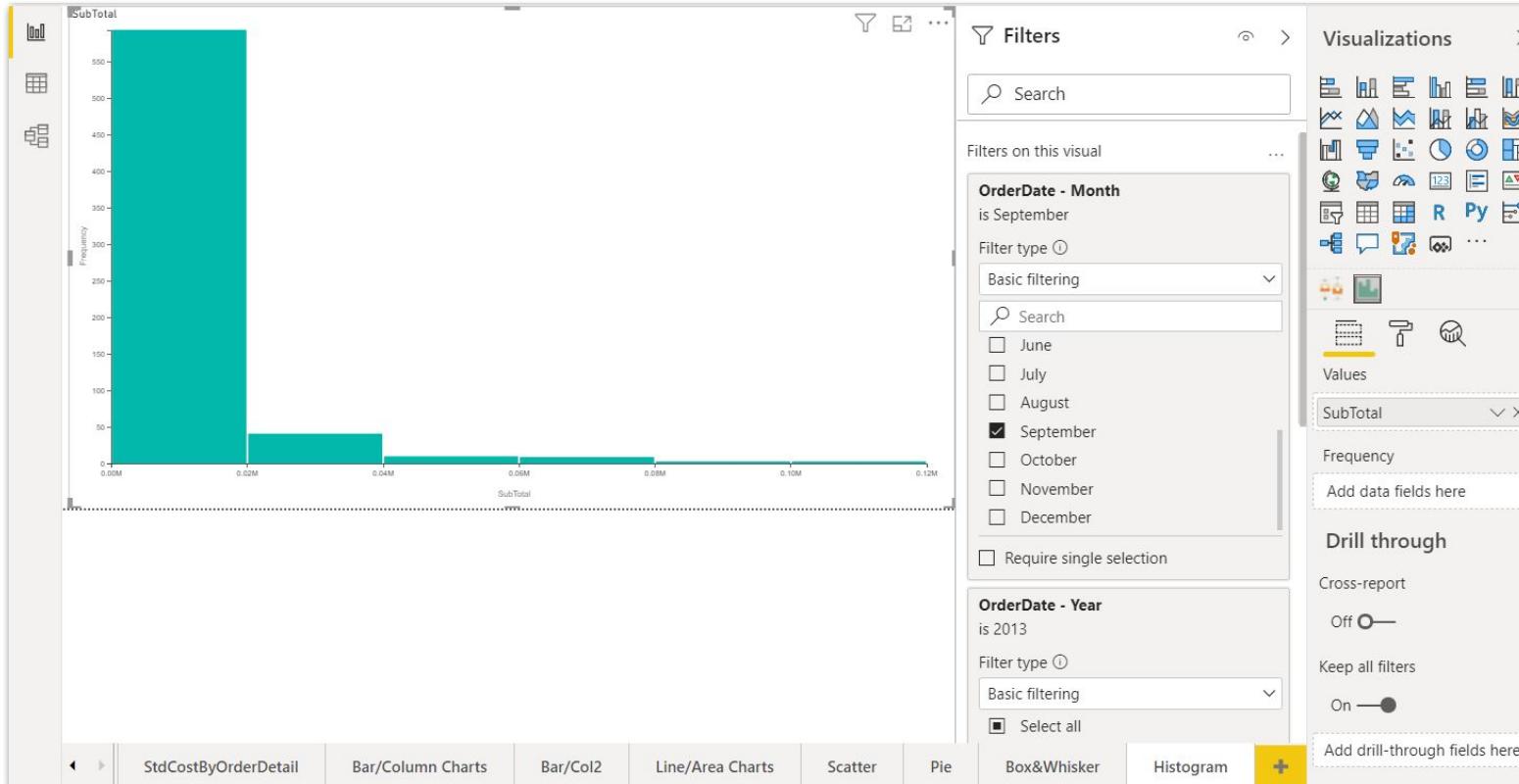
1. Install the Histogram visual from outside sources  
(either AppSource, or from the provided **Histogram.pbviz** file)
2. Add Sales.SalesOrderHeader **SubTotal** to *Value*
3. Filter based on **OrderDate Year** and **Month**
4. Reduce the histogram to 6 bins
5. Reduce the y-axis values to 0 decimal places



Solo Exercise:

# Build a Histogram - Solution

15 minutes





## Discussion:

# What Did You Think?

What charts did you find useful?

What would you like see more of?

There's a LOT more out there.

Feel free to investigate heatmaps, violin plots, and others!

Advanced Analytics

---

Let's Review!



# What Did We Just Cover?

In this lesson, we:

- Reviewed descriptive statistics
- Chose the best visual for a given inquiry
- Built common visualizations, including:
  - Bar and Column Charts
  - Line and Area Charts
  - Scatter Charts
  - Pie and Donut Charts
- Loaded new visualizations from outside sources
- Applied conditional formatting to charts



# Finish That Sentence

What are your biggest takeaways from today?



“Something that really got me thinking is...”

“The best thing I got out of today is...”

“I discovered...”

“I still want to learn about...”

“I was surprised that...”

# Ask Me Anything!



---

**Don't Forget Your  
Exit Tickets!**



# A Few Good References

[Exploratory Data Analysis](#) (Wikipedia)

[Bar Plots](#) (Data-to-Viz)

[Line Charts](#) (Data-to-Viz)

[Scatter, bubble charts in Power BI](#) (Microsoft)

[Visual types in Power BI](#)

[Microsoft AppSource](#)



See you next time!

---

**Thank you!**

Advanced Analytics

---

# Filters



## Guided Walk-Through: Filters

- Make sure the Filters pane is not collapsed
- There are 3 levels of filters
  - Visual level
  - Page level
  - Report level (all pages)
- **Visual level** will only show when you select a visual and by default has all the values, axes, legend in this visual that can be filtered

The screenshot shows the 'Filters' pane in Power BI. It contains three main sections, each with a red border around it:

- Filters on this visual**: Contains a search bar and a button labeled "Add data fields here".
- Filters on this page**: Contains a button labeled "Add data fields here".
- Filters on all pages**: Contains a button labeled "Add data fields here".



## Guided Walk-Through: Filters

- **Page level** filter is empty by default and you can drag and drop any field there to filter all the visuals in this page.
- **Report level** is similar to the page level but it works all the pages in the report
- To add a filter at any level you can just drag a field and drop in the level you want (e.g. **Product.Name**) default **Filter Type** is **Basic Filtering** where you can select individual values using a check box

The screenshot shows the 'Filters' pane in Power BI. It displays three levels of filtering:

- Filters on this visual**: An empty box with a placeholder 'Add data fields here'.
- Filters on this page**: An empty box with a placeholder 'Add data fields here'.
- Filters on all pages**: An expanded section showing a list of products with checkboxes. The list includes:
  - Name is (All)
  - Filter type ⓘ Basic filtering
  - Search
  - Fork End 1
  - Freewheel 1
  - Front Brakes 1



## Guided Walk-Through: Filters

- You can change the Basic filtering to **Advanced filtering** to handle more complex filter using one or multiple operators.
- Choose the operator (e.g. start with) and add the word Front then click apply.
- This will filter the product name and include the products that only start with the word Front on the its name
- You can add more conditions using And/Or if needed

Name  
is (All)

Filter type ⓘ

Advanced filtering

Show items when the value:

contains

Name  
is (All)

Filter type ⓘ

Advanced filtering

Show items when the value:

starts with

Front

And Or

Apply filter