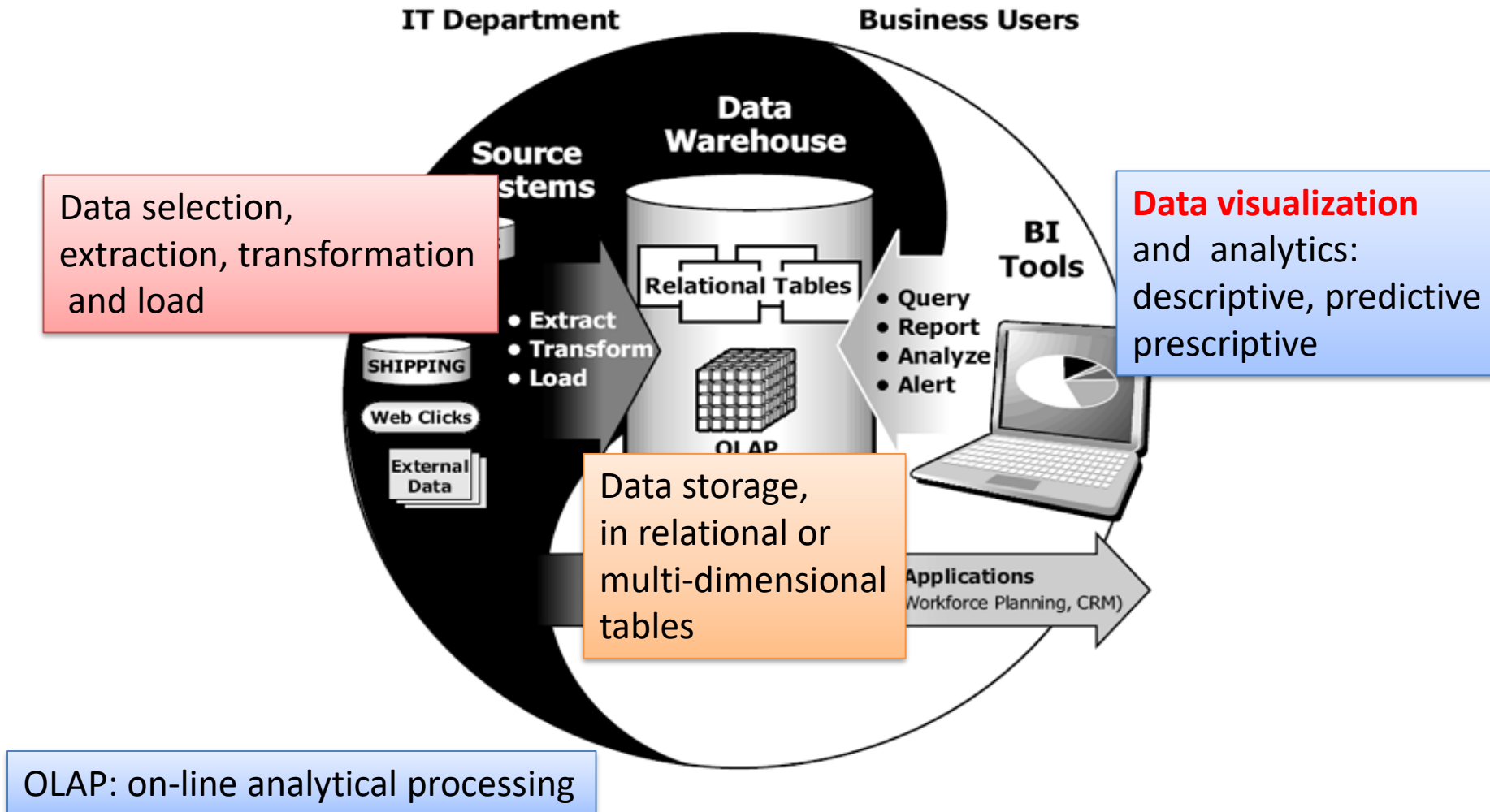


DATA VISUALIZATION



Architecture of a BI system



Outline


What is Visualization?

Why Visualize?

Data → Visualization

How to tell if a visualization is
“appropriate”

Guidelines



Visualization: what is?

- Visual **representations** of data that reinforce human **cognition**

Perhaps a
more helpful
question:

What are some ways
a “visualization” can
be **useful**?



Types of data

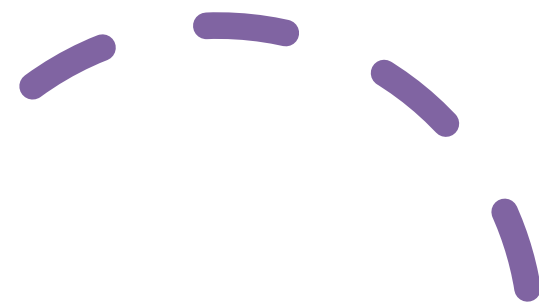
- Data can be classified in three groups
- Qualitative (Attributes)
 1. Nominal
 2. Ordinal
- Quantitative (Metrics)
 3. Numeric

Qualitative:Nominal data

- Data that be counted, but **not** ordered or aggregated.
- Examples:
 - Products – Books, Movies, Music
 - Gender – Male, Female
 - State – Virginia, Nevada, California



Qualitative: Ordinal data

- 
- 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

Metrics

- Quantitative data that can be **counted, ordered, and aggregated.**
- Examples:
 - Revenue, Cost, Profit
 - Number of Customers
 - Temperature
 - Time



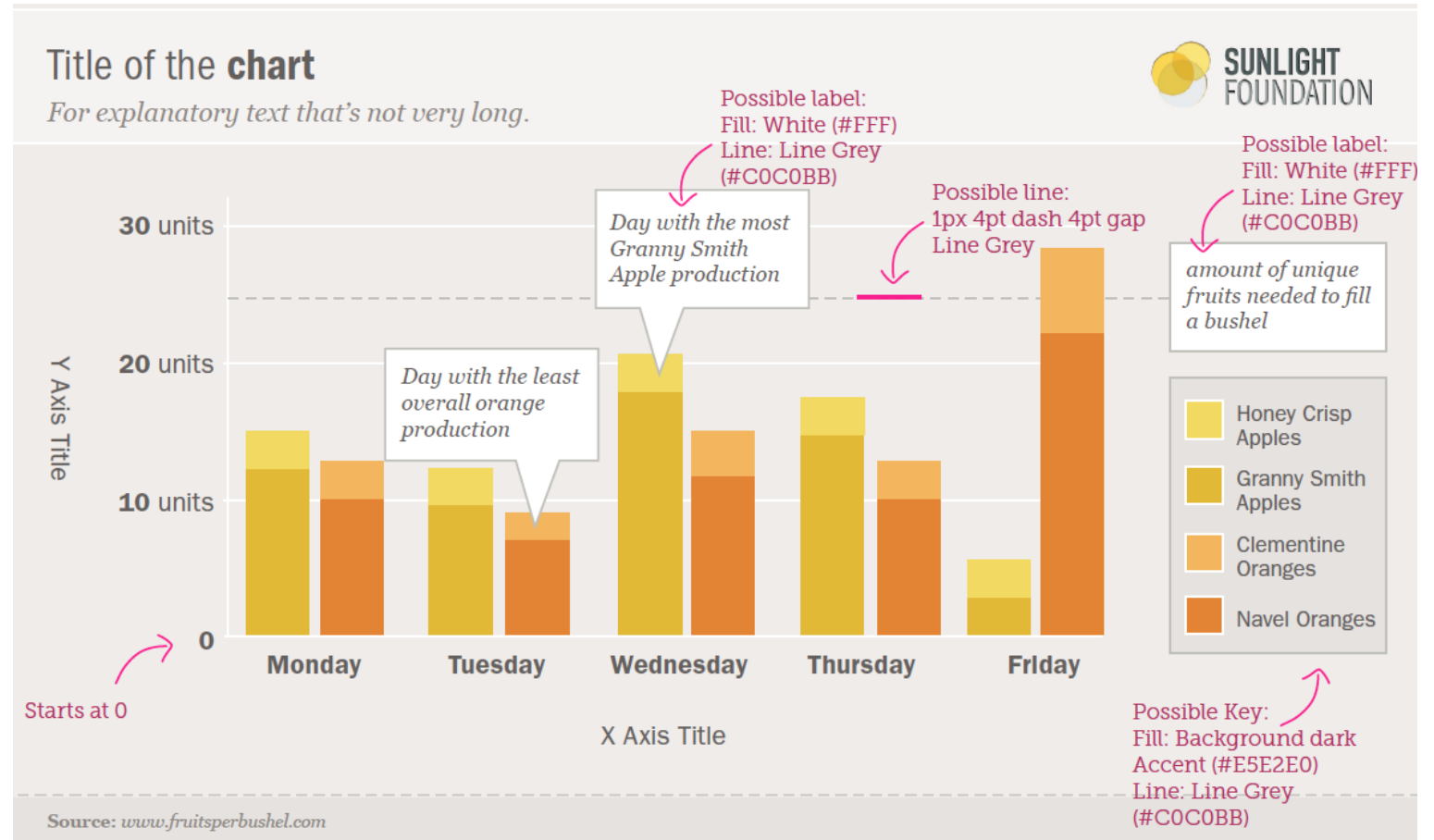
Ordinal Attributes and Metrics

- Some data can be used as **either ordinal or metrics**. Their classification is dependent on usage.
- Examples:
 - Age
 - Scores

Types of visualizations

- Bar charts (histograms)
- Line charts
- Scatterplots
- Maps
- Pie Charts
- Network (graphs)
- ...many others (will see creative examples)

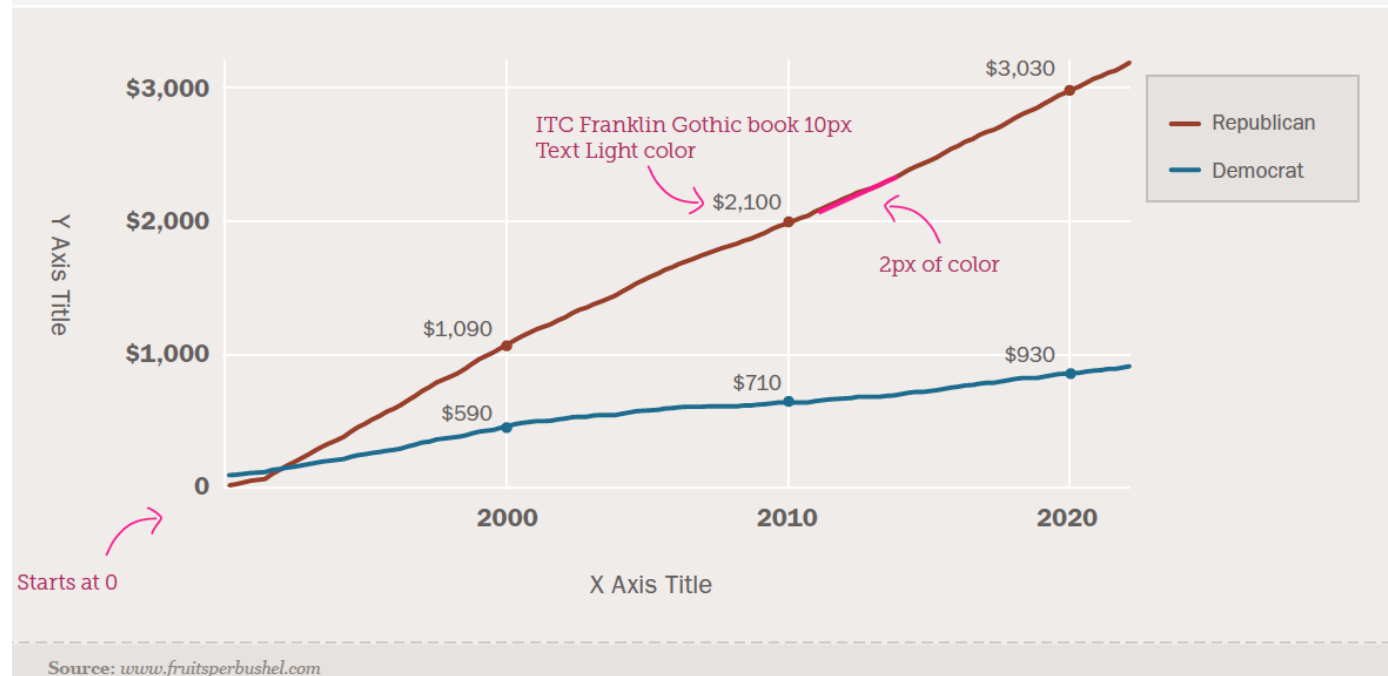
Bar Charts (histograms)



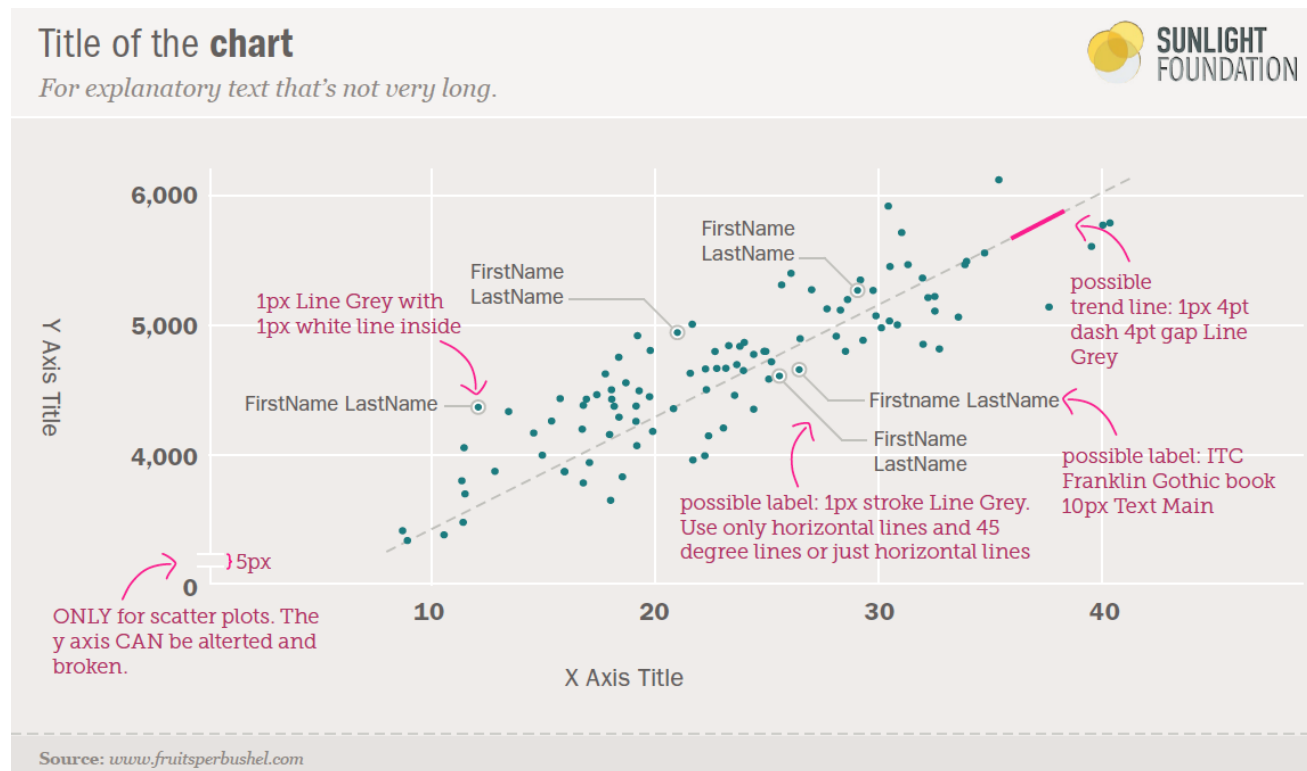
Line charts

Title of the **chart**

For explanatory text that's not very long.

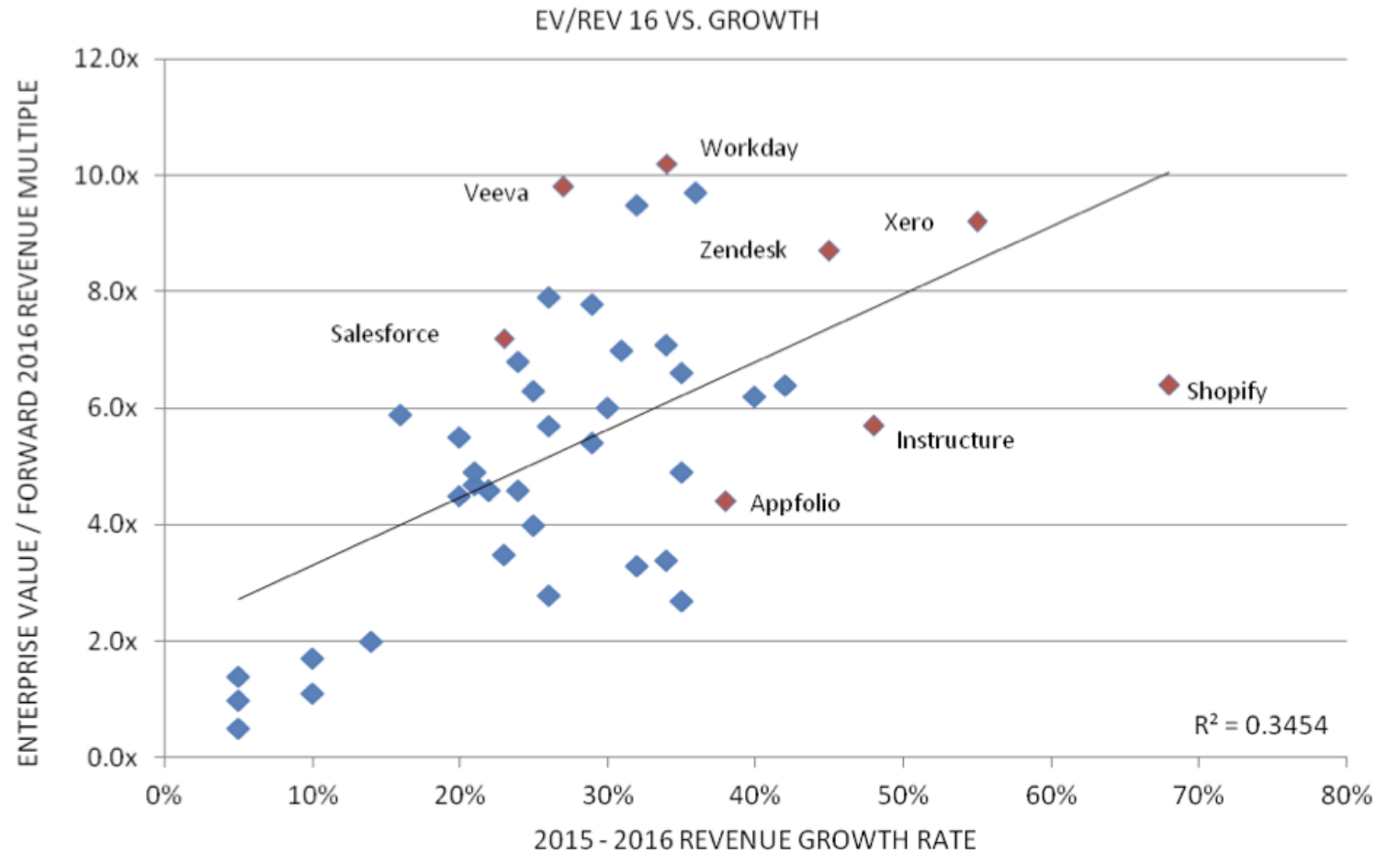


Scatterplots

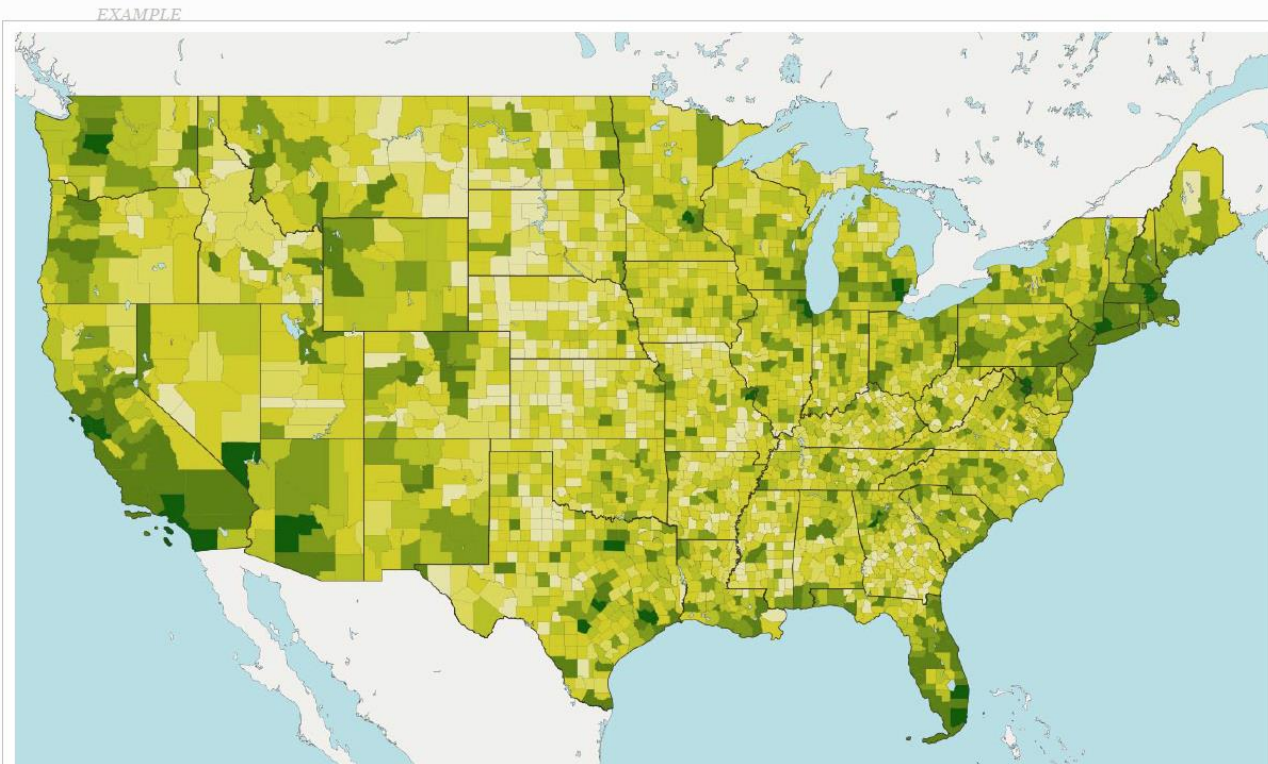


- Shows the relationship between two **continuous variables**
- Each point in the plot represents an observation
- You can change color or symbol to **highlight groups**
- Sometimes it is nice to show a trend line (regression)

Example
scatterplot
(enterprise
value/revenue
vs growth)



Maps

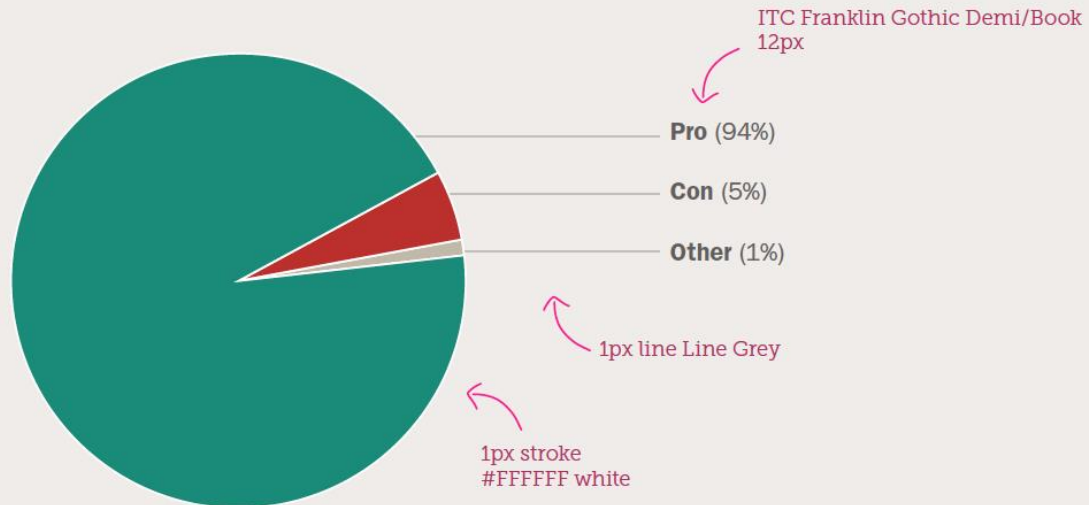


- Useful for analysis with a strong geographic component
- Remember: color scale comparisons are harder for humans than size comparisons. Keep this in mind as you choose between a map or another layout.

Pie Charts

Title of the **pie chart**

For explanatory text that's not very long.

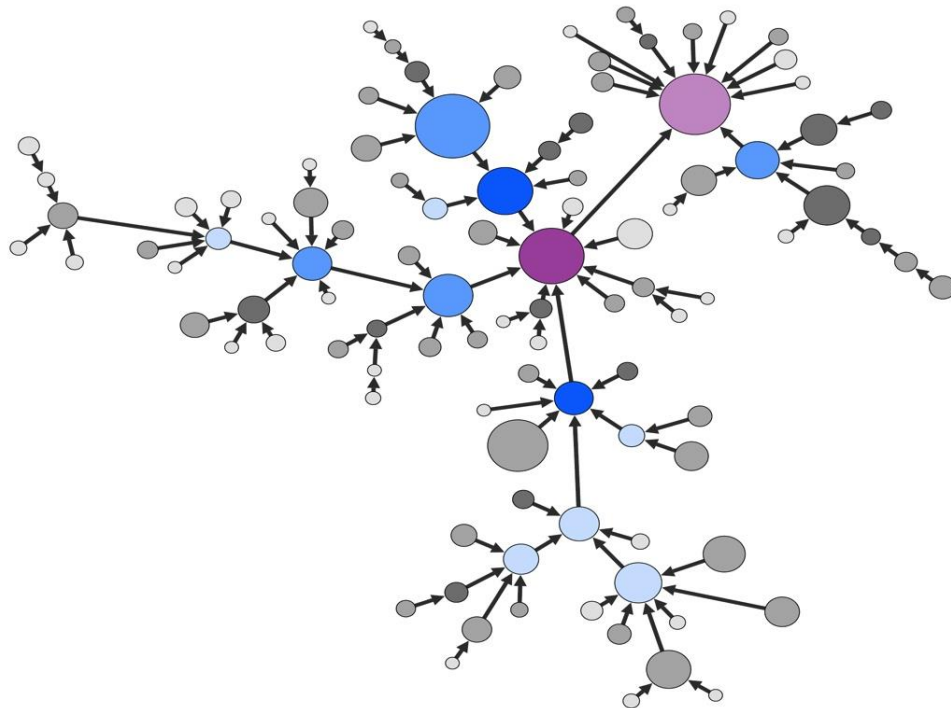


Source: www.fruitsperbushel.com

- Almost never the right choice (angular comparison is hard)
- Use only if the following 2 conditions are met:
 - You want to show **the relative relationship** between 2-3 things
 - They add up to 100%

Networks (graphs)

Network of company X sales representatives



This map shows the sales representatives (nodes) linked if one has recruited another one. Connected nodes are closer, disconnected nodes are farther. Node size shows the sales volume, colors shows the rank, text size shows the level of sponsoring activity.

100 sales representatives
99 connections

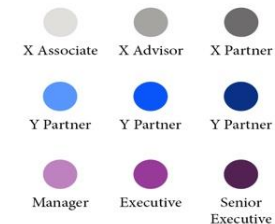
LEGEND



Sales Volume



Rank



© 2013 Sébastien Heymann, made with Gephi

- Useful for showing the **relationships** between entities
- Can use color, size, etc. to encode additional information about nodes/edges
- Caveat: network diagrams quickly become hairballs for large, dense data.

Connecting Data To Visualization

- Data have types
- Visualizations have types
- Reports have communication objectives
- **How do we map one onto the other?**
 1. **Depending on the information we want to convey**
 2. **Depending on the type of data we want to visualize**



SELECTING VISUALIZATIONS

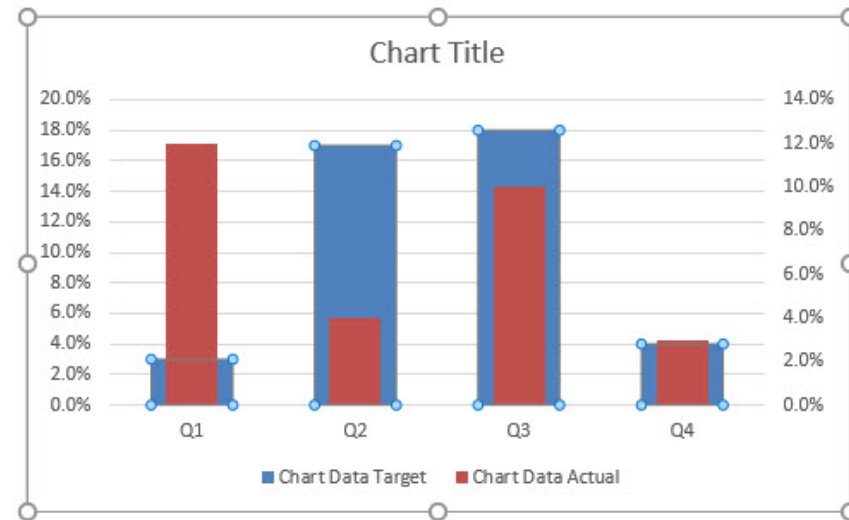
**Depending on the
information we want
to convey**

Do you want to **compare** values?

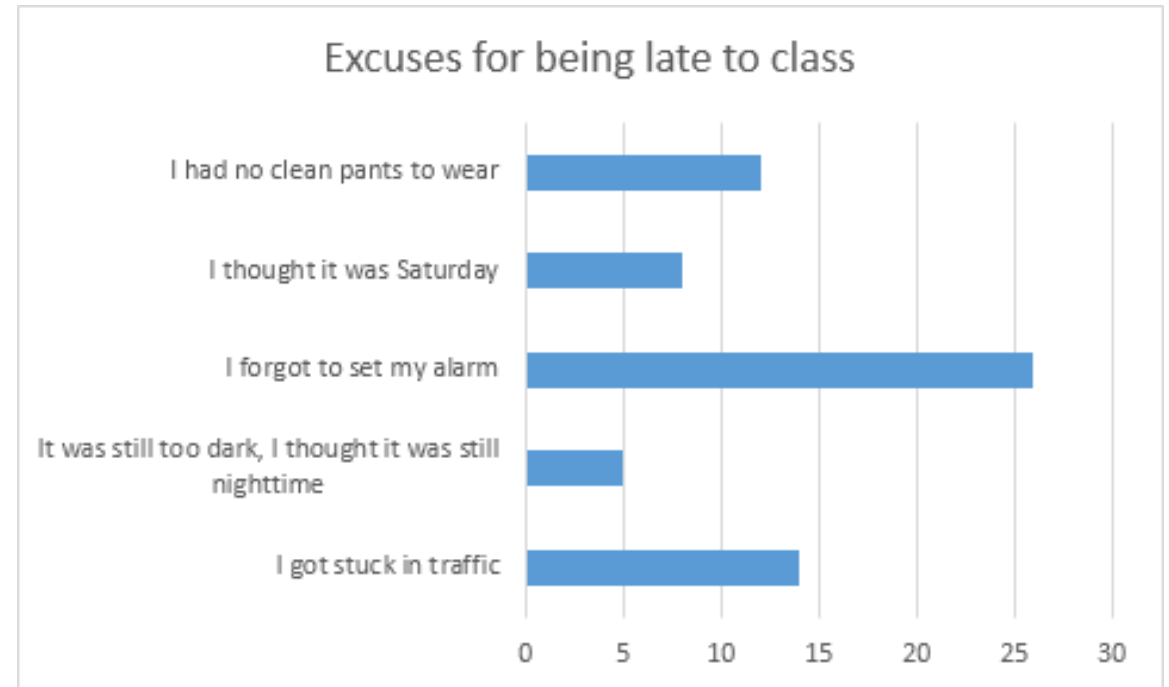
- Charts are perfect for comparing one or many value sets, and they can easily show the low and high values in the data sets.
- To create a **comparison chart**, use these types of graphs:
 - Column
 - Bar
 - Circular Area
 - Line
 - Scatter Plot
 - Bullet

A **column chart** is used to show a **comparison** among different items, or it can show a comparison of items over time.

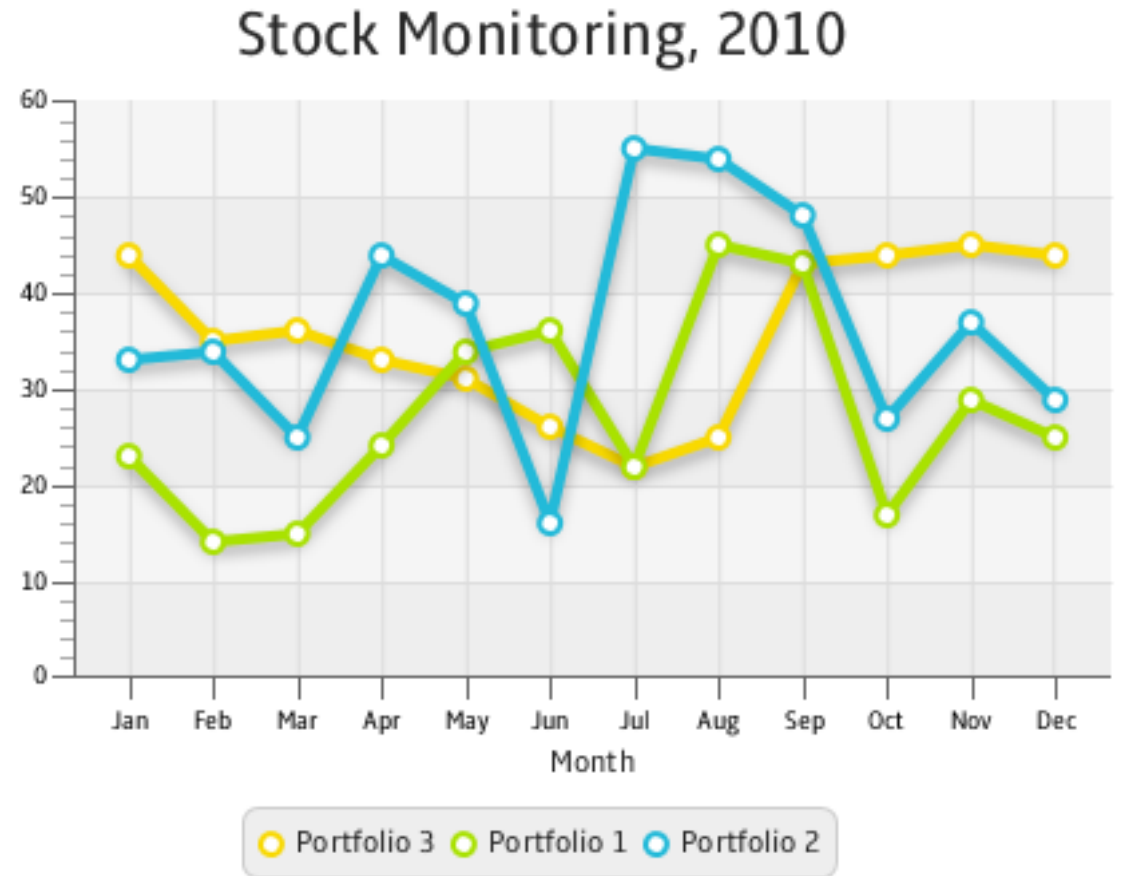
- Design Best Practices for Column Charts:
- Use consistent colors throughout the chart, selecting accent colors to highlight meaningful data points or changes over time.
- Use horizontal labels to improve readability.
- Start the y-axis at 0 to appropriately reflect the values in your graph.



A **bar chart**, basically a horizontal column chart, should be used to avoid clutter when one **data label is long** (e.g., “Individual contributors” or if you have more than 10 items to compare.



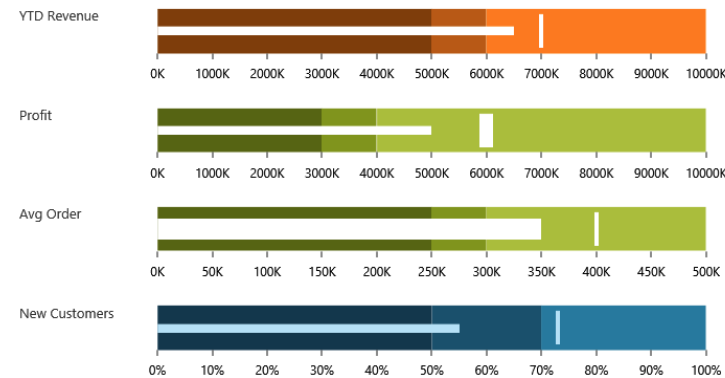
A **line chart** reveals trends or progress over time and can be used to show many different categories of data. You should use it when you chart a **continuous data set**.



A bullet graph reveals progress toward a goal, compares this to another measure, and provides context in the form of a rating or performance.

- In this example, vertical with bar is the target, horizontal white line is what has been achieved

Sunkost: Sales Target of 2010

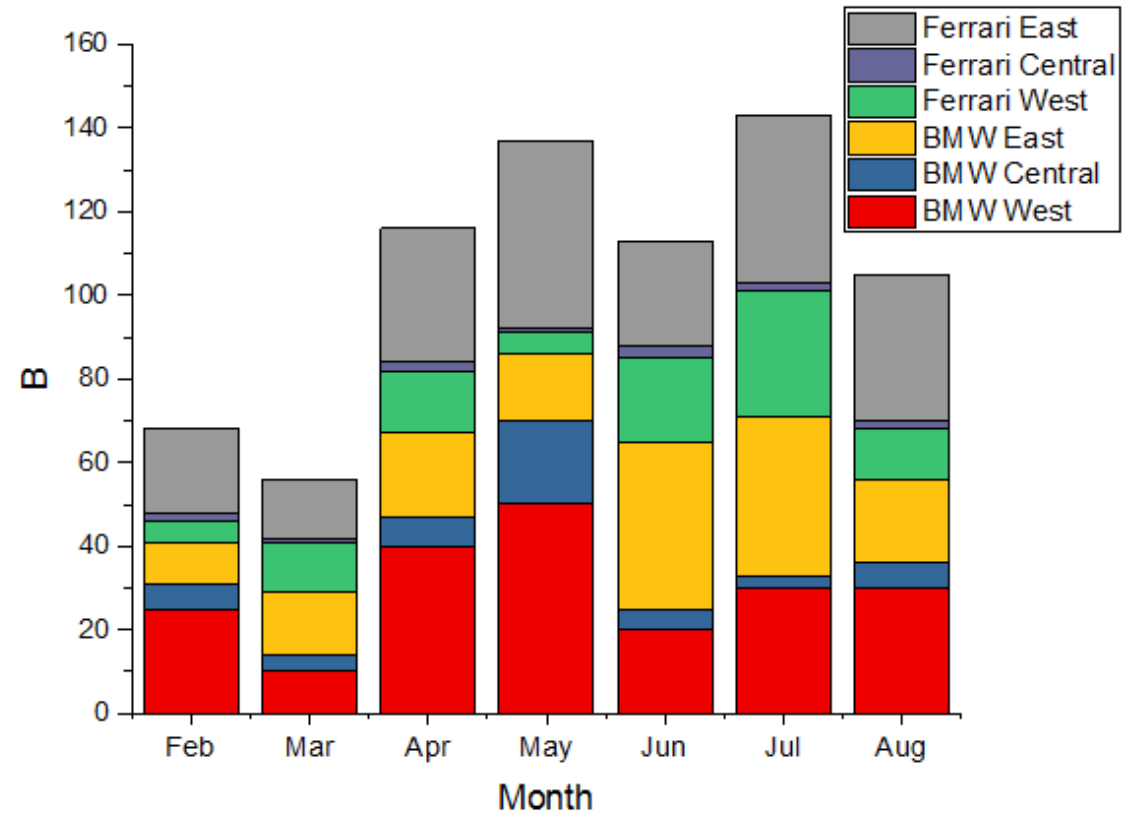


The sales totals to date for 2013 (white horizontal line), clearly exceed the total sales for all of 2012 (the beginning of the middle range). The 2013 sales numbers suggest that our new marketing campaign is successful, resulting in increased product penetration and a significant sales boost, working our way up to the targeted goals for the whole year (vertical white lines).

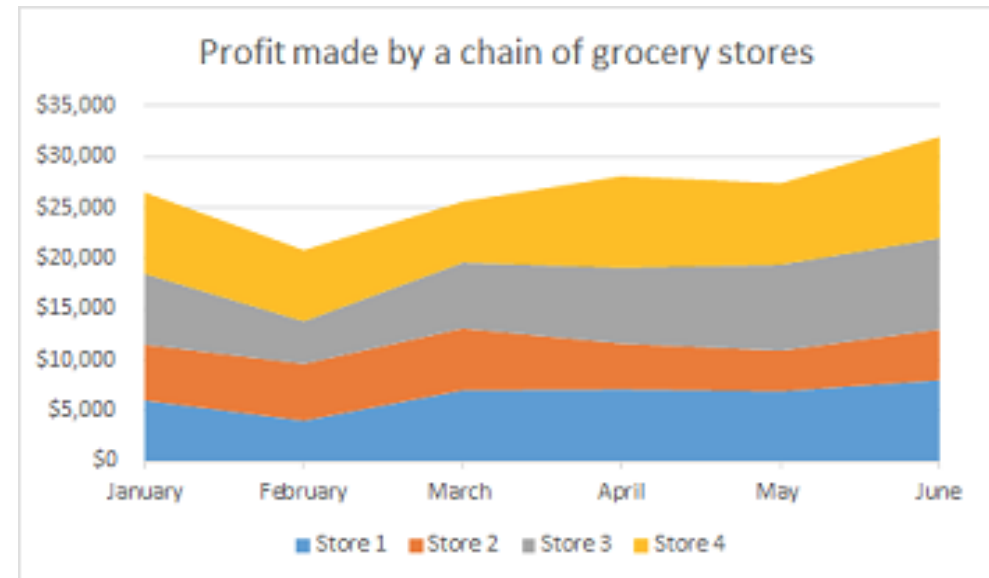
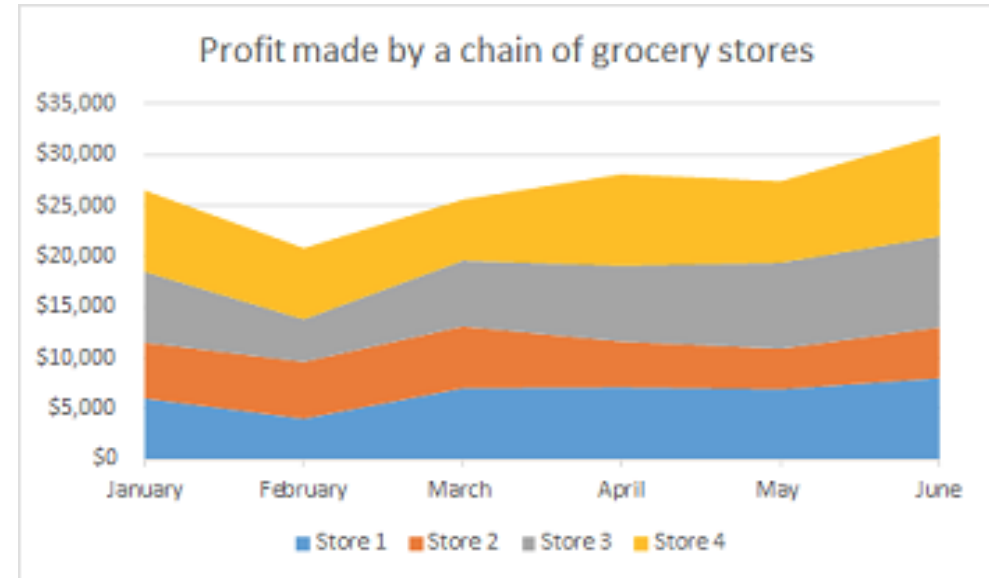
Do you want to show the composition of something?

- Use this type of chart to show how individual parts make up the whole of something, such as the device type used for mobile visitors to your website or total sales broken down by sales rep. Better suited to compare categories.
- To show composition, use these charts:
 - Pie
 - Stacked Bar
 - Stacked Column
 - Area
 - Waterfall

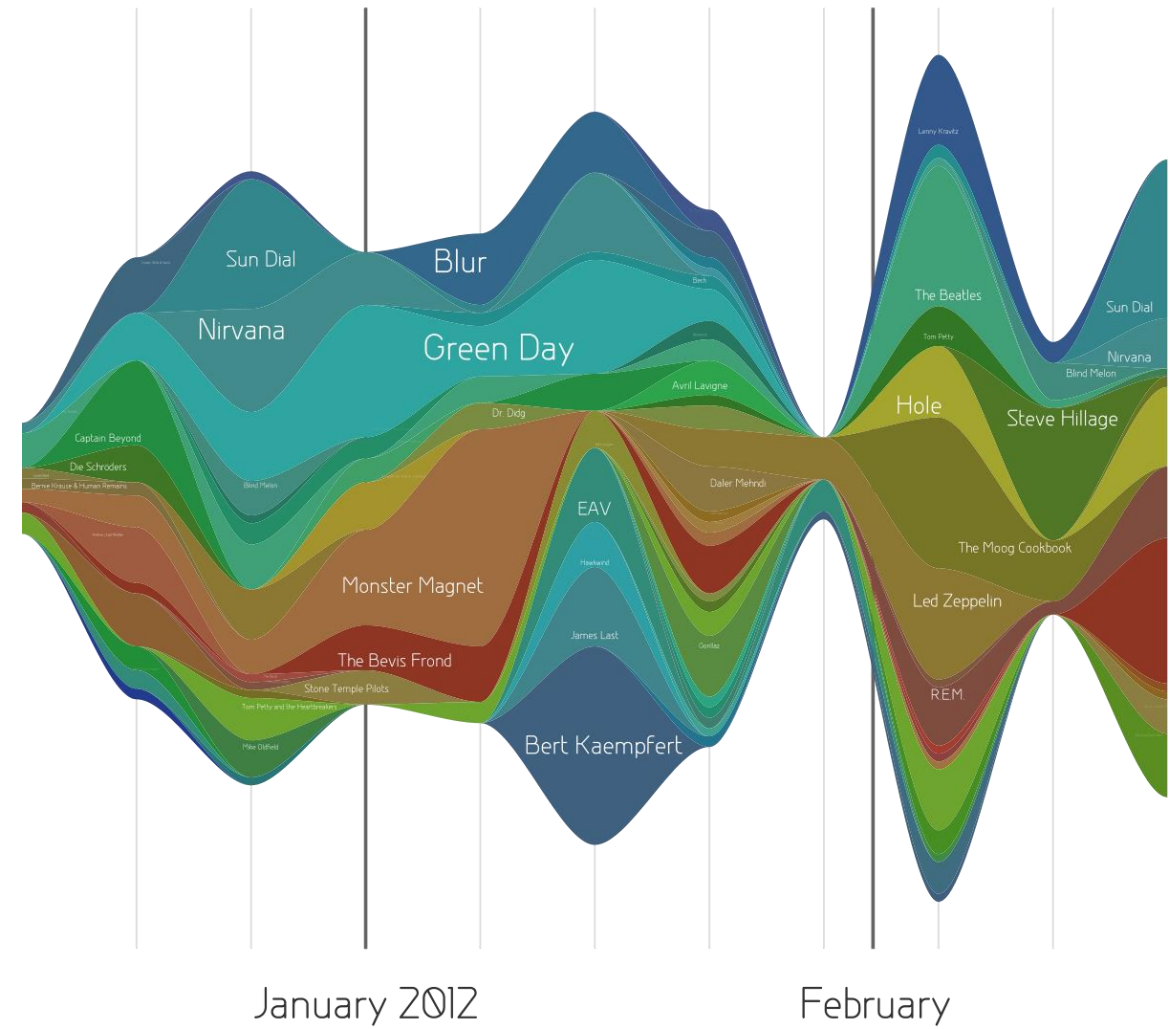
A **stacked** (bar or column) **chart** is used to break down and **compare parts of a whole**. Each bar in the **chart** represents a whole, and segments in the bar represent different parts or categories of that whole.



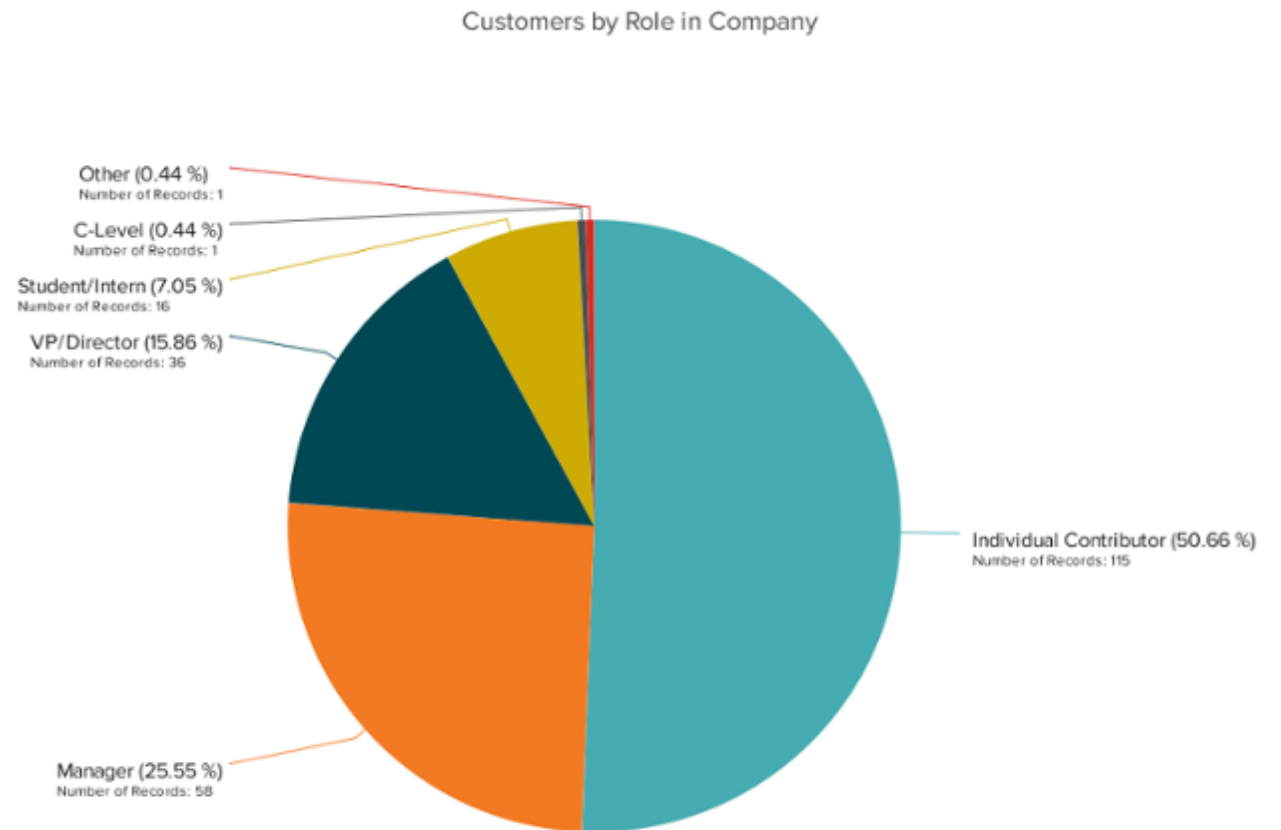
An **area chart** is basically a line chart, but the space between the x-axis and the line is filled with a color or pattern. It is useful for showing **part-to-whole relations**, such as showing individual sales reps' contribution to total sales for a year. It helps you analyze both overall and individual trend information.



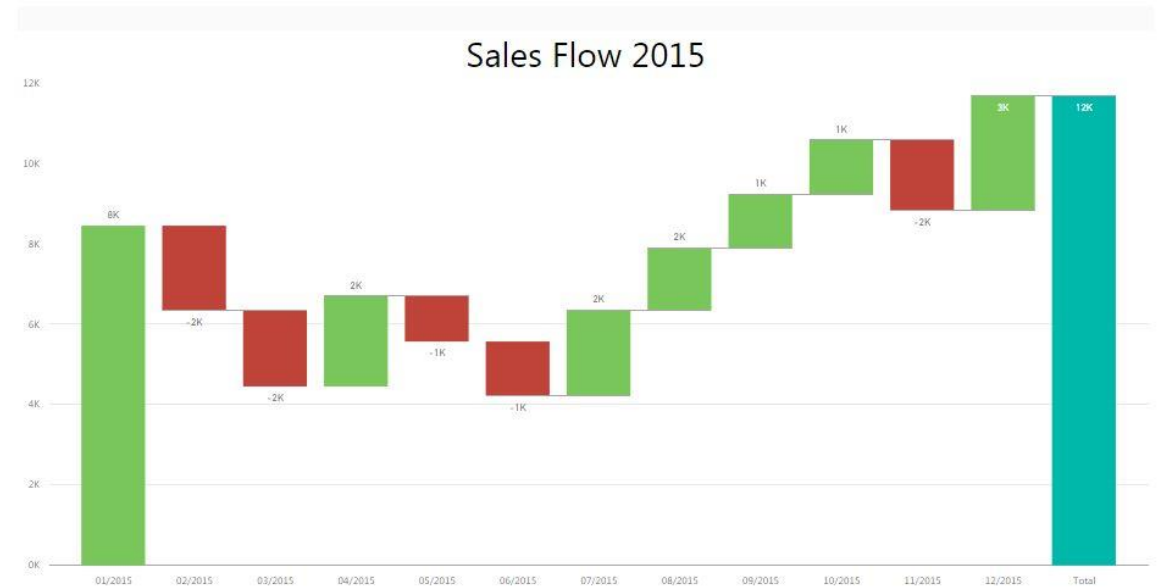
Stream charts (a.k.o. area charts)



A **pie chart** shows a static number and how **categories** represent part of a whole -- the composition of something. A pie chart represents numbers in percentages, and the **total sum of all segments** needs to equal 100%.



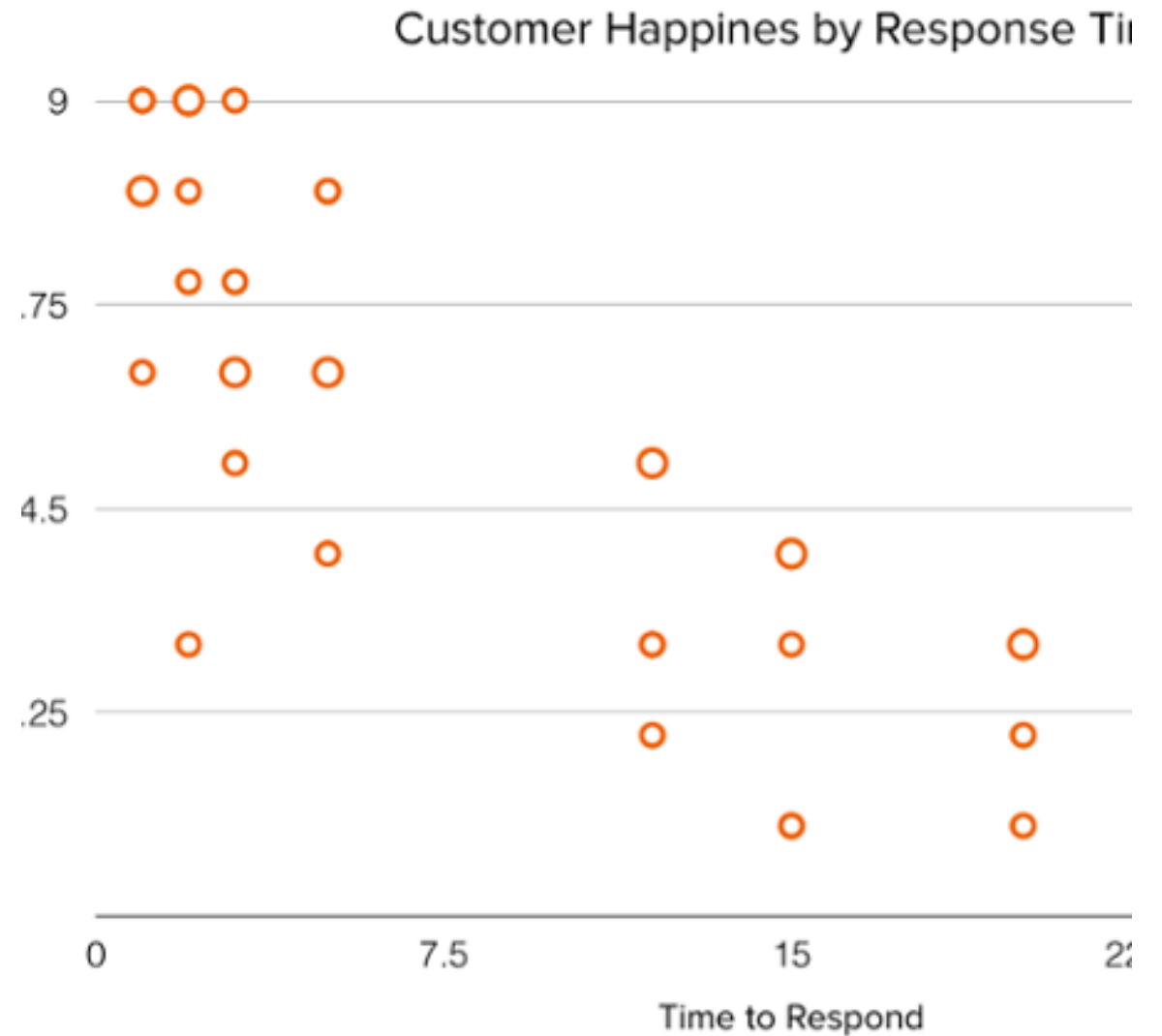
A **waterfall chart** should be used to show **how an initial value is affected by intermediate values** -- either positive or negative -- and **resulted in a final value**. This should be used to reveal the composition of a number. An example of this would be to showcase how subsequent movements affected the total balance (increases in value are coloured green and decreases are coloured red).



Do you want to understand the distribution of your data?

- Distribution charts help you to understand outliers, the normal tendency, and the range of information in your values.
- Use these charts to show distribution:
 - Scatter Plot
 - Line
 - Column
 - Bar

A **scatter chart** will show the relationship between two different variables or it can reveal the distribution trends. It should be used when there are many different data points, and you want to highlight similarities in the data set. This is useful when looking for outliers or for understanding the distribution of your data.

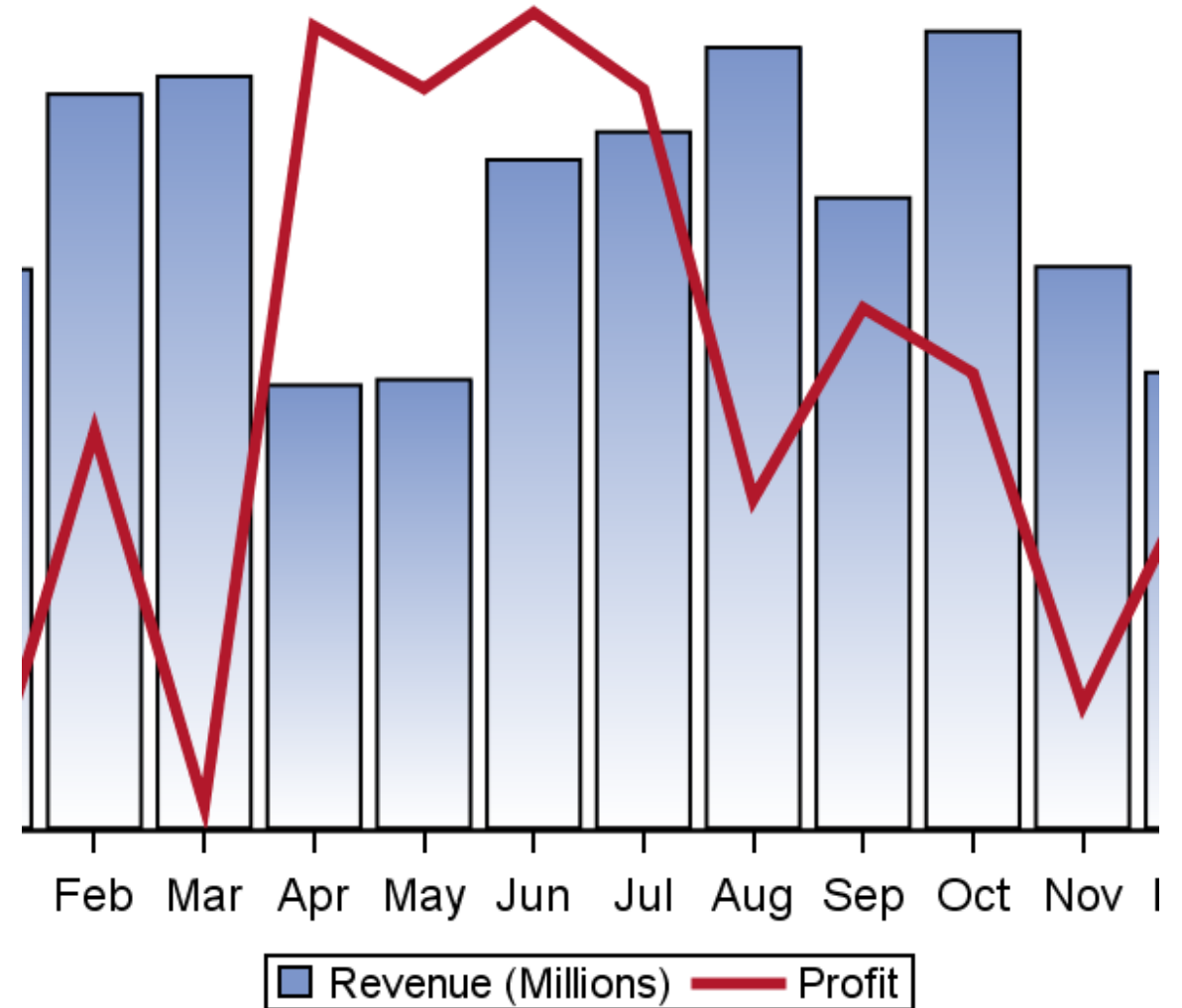


Are you interested in analyzing trends in your data set?

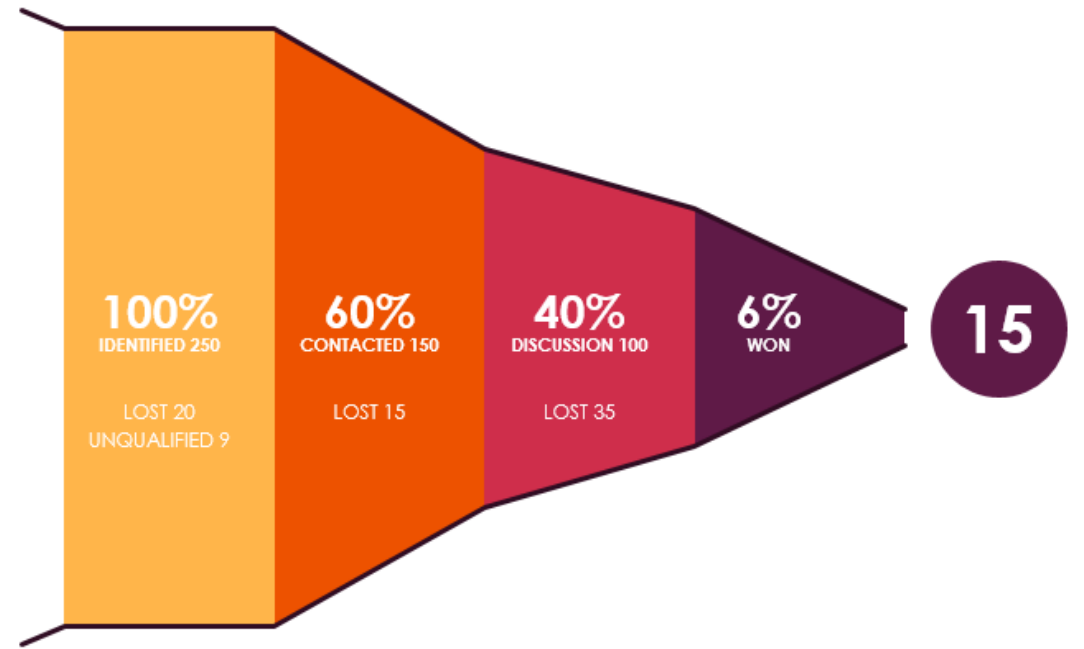
- If you want to know more information about how a data set performed during a specific time period, there are specific chart types that do extremely well.
- You should choose a:
 - Line
 - Dual-Axis Line
 - Funnels
 - Column

A **dual axis** chart allows you to plot data using **two y-axes** and a **shared x-axis**. It's used **with three data sets**, one of which is based on a **continuous** set of data and another which is better suited to being grouped by **category**. This should be used to **visualize a correlation or the lack thereof** between these three data sets.

Revenues and Profit by Month for 2014



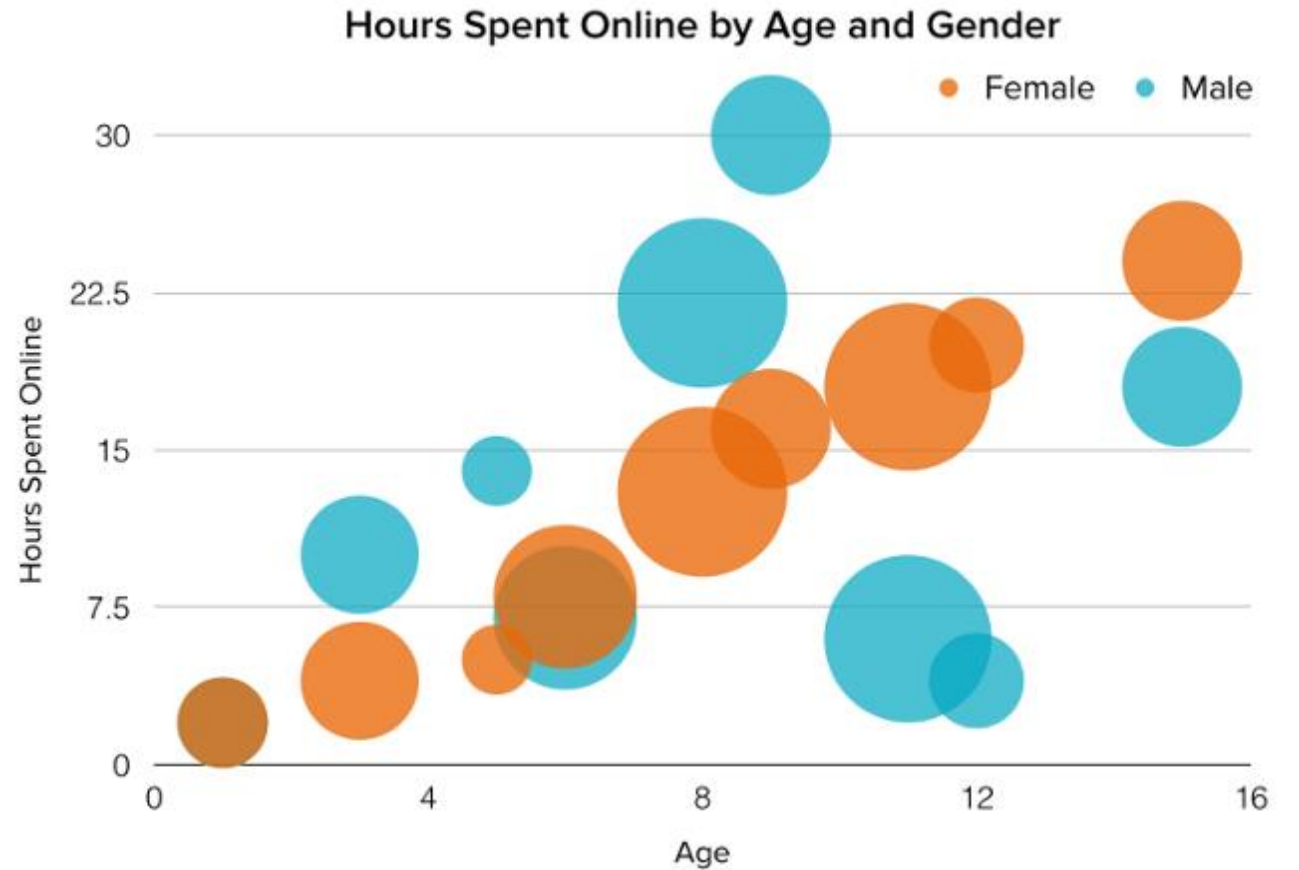
A **funnel chart** shows a series of steps and the **completion rate** for each step. This can be used to track the sales process or the conversion rate across a series of pages or steps (e.g. from contacts to contracts).



Do you want to better understand the relationship between value sets?

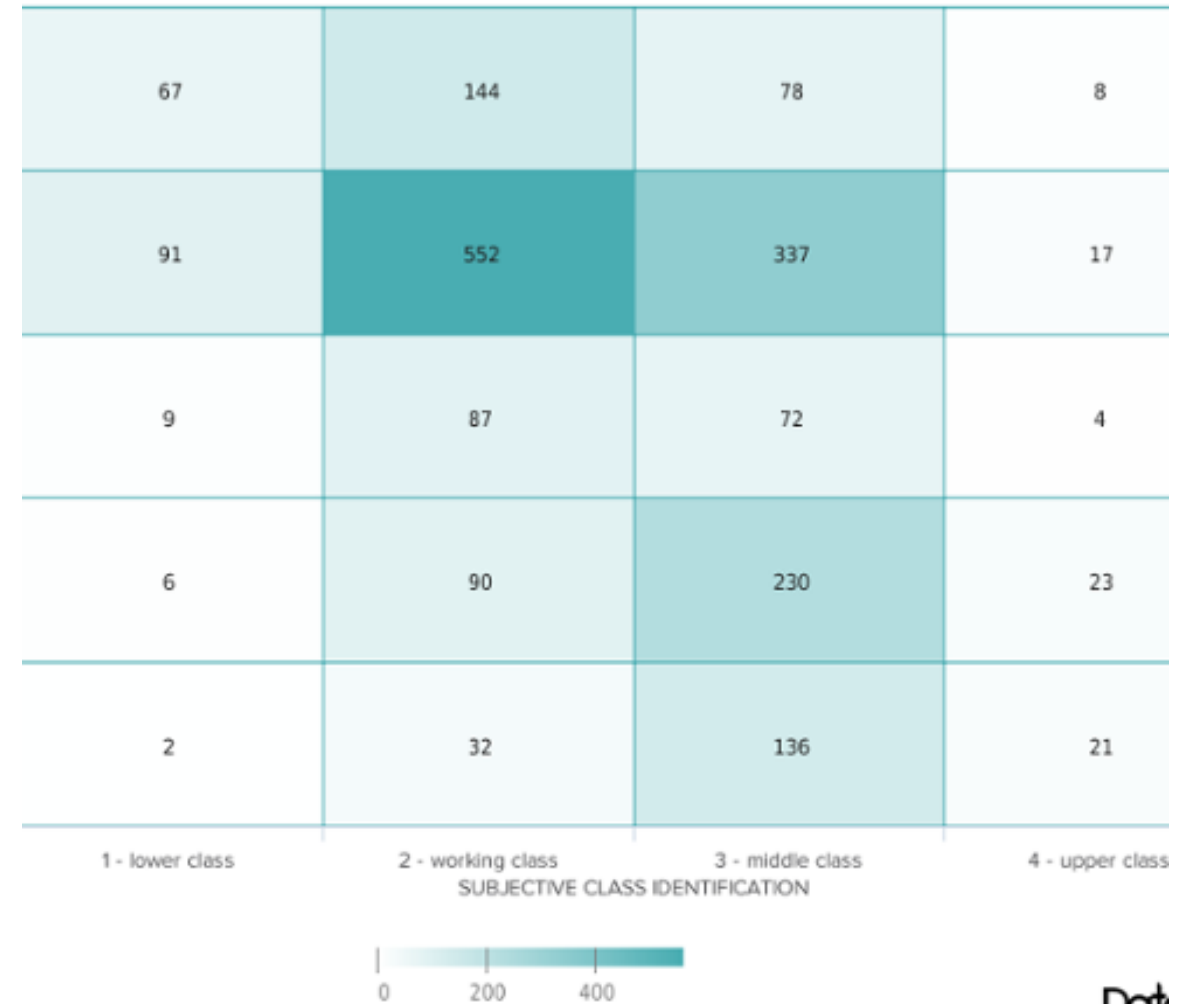
- Relationship charts are suited to showing **how one variable relates to one or numerous different variables**. You could use this to show how something positively effects, has no effect, or negatively effects another variable.
- When trying to establish the relationship between things, use these charts:
 - Scatter Plot
 - Heat maps
 - Bubble
 - Line
 - Networks
 - Spirals

A **bubble chart** is similar to a scatter plot in that it can show distribution or relationship. There is a **third dimension**, which is indicated by the size of the bubble or circle.

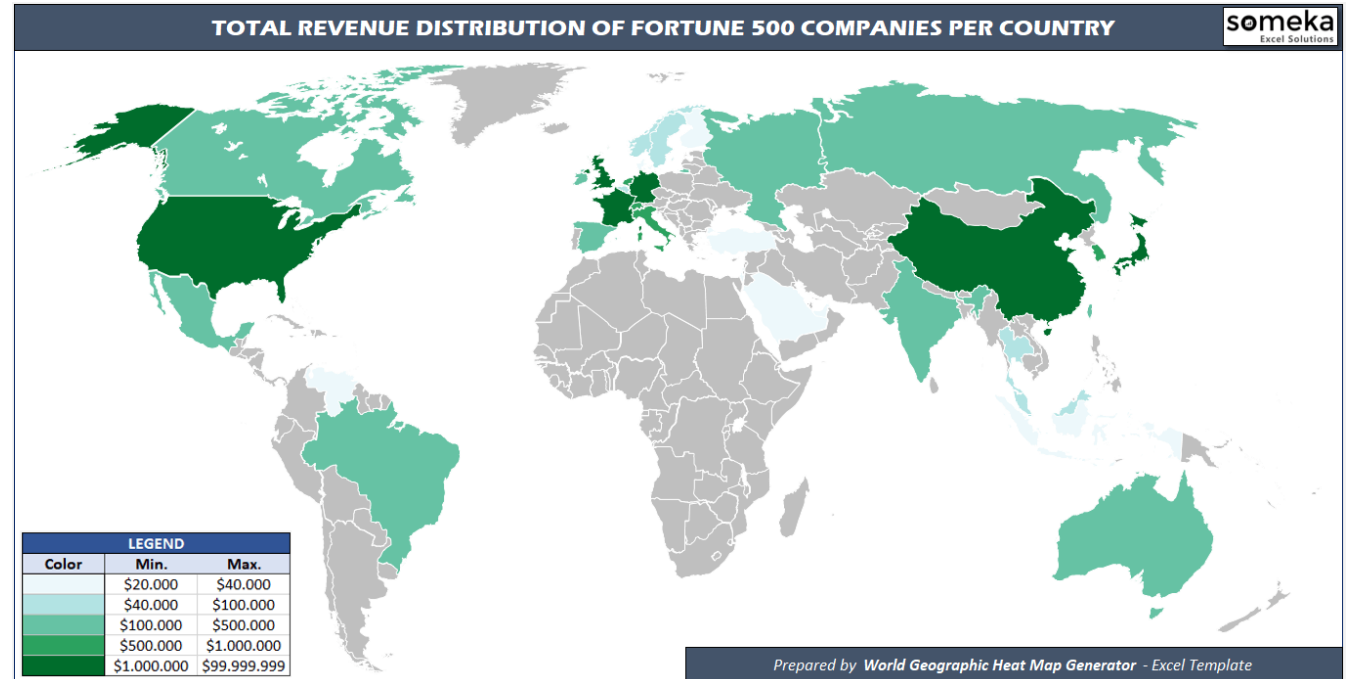


A **heat map** shows the relationship between two items and provides rating information, such as high to low or poor to excellent. The rating information is displayed using varying colors or saturation.

Highest Degree vs. Class Identification



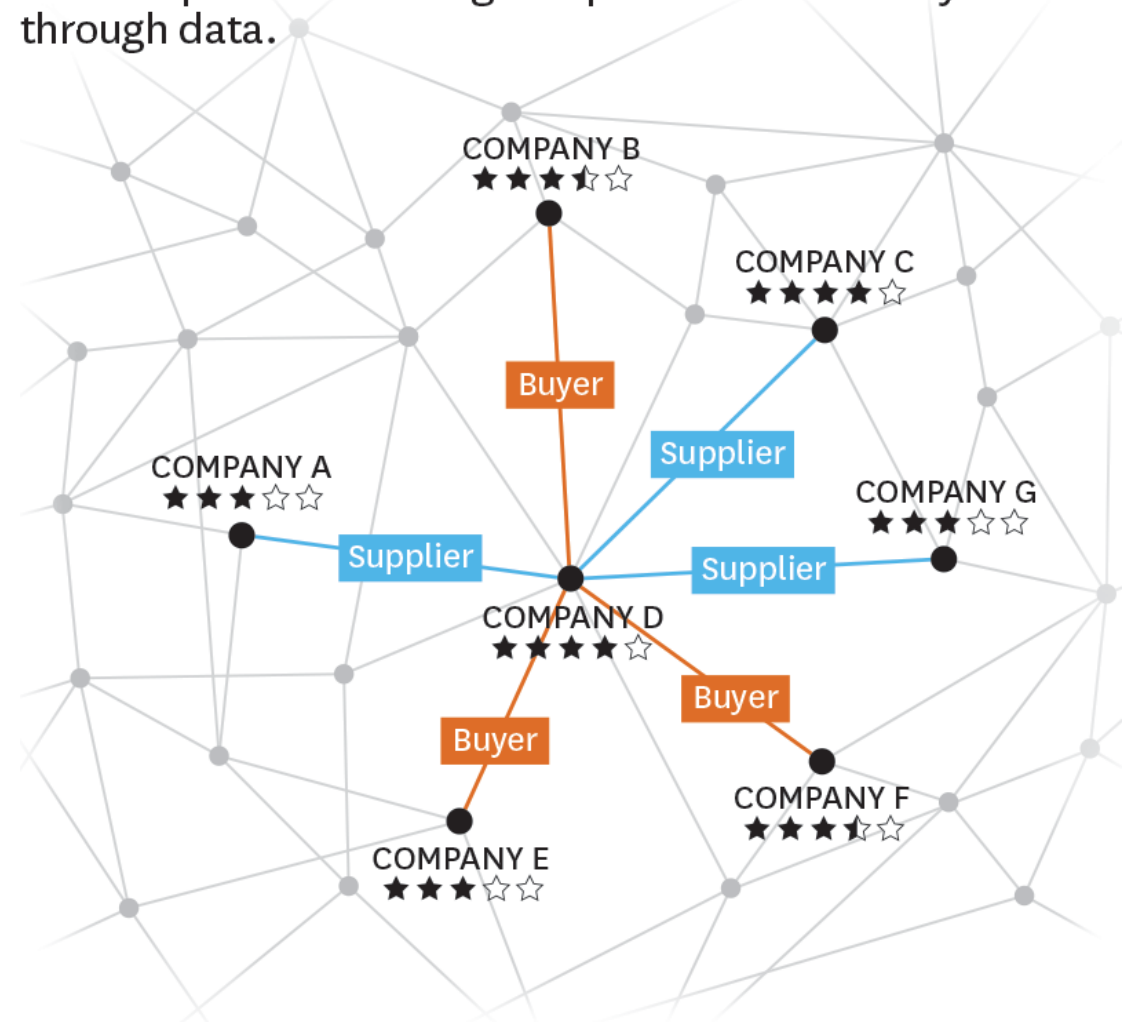
Heat maps can also be geographical maps



Network graphs
are useful to
show complex,
non-numerical
relations
between entities

THE COMMERCIAL GRAPH

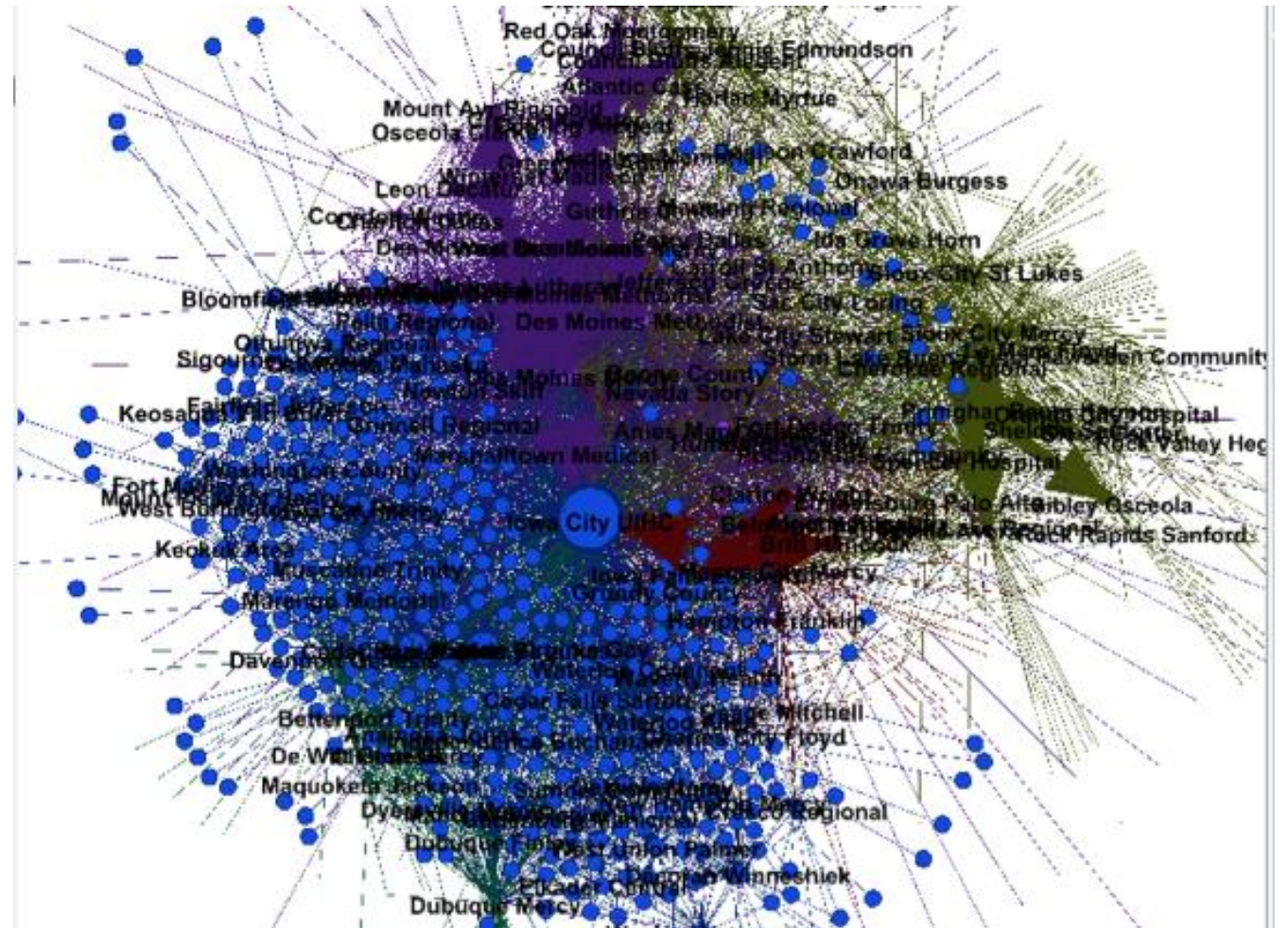
An example of visualizing complex business ecosystems through data.



SOURCE PLATFORM THINKING LABS

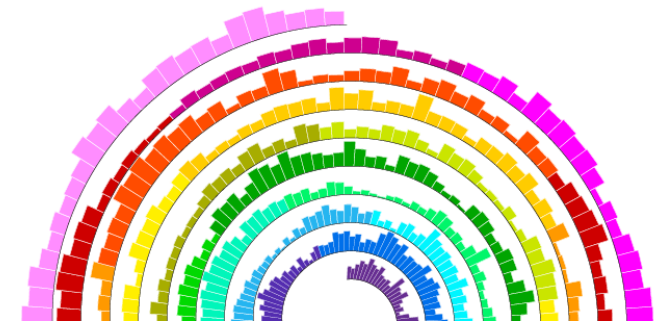
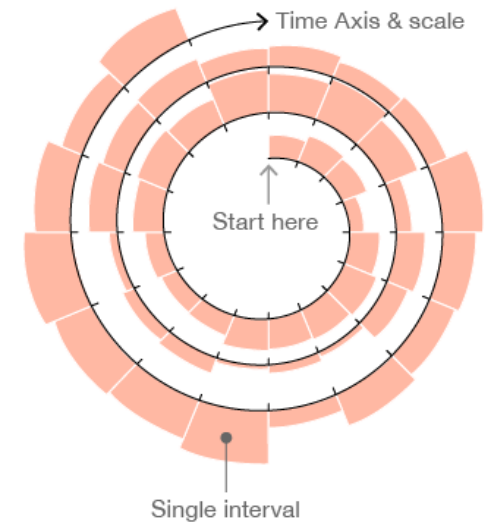
HBR.ORG

Although they quickly become unreadable

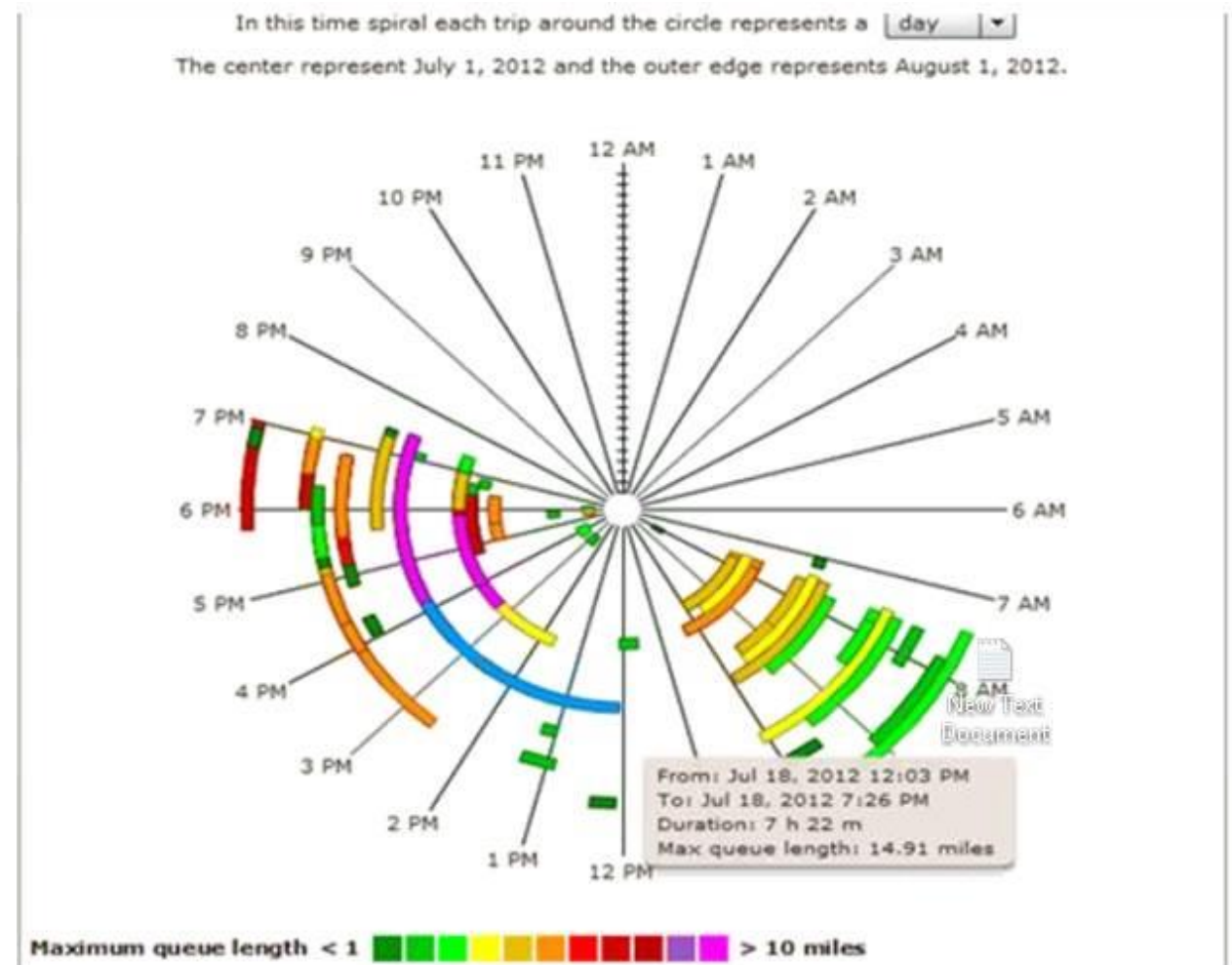


Spirals

- Spiral Plots are ideal for showing large data sets, usually to show **trends over a large time period**. This makes Spiral Plots great for displaying **periodic patterns**. Colour can be assigned to each period to break them up and to allow some comparison between each period.
- So for example, if we were to show data over a year, we could assign a colour for each month on the graph.



Spirals uses



Here each trip around the circle represent a day, colors represent queue lengths in streets

Spiral uses

- Here the spiral center is a “target” attribute, and dots represent attributes (single or sets of attributes) that influence the target. Strength of influence decrease as we move far from the centre along the spiral line



SELECTING VISUALIZATIONS

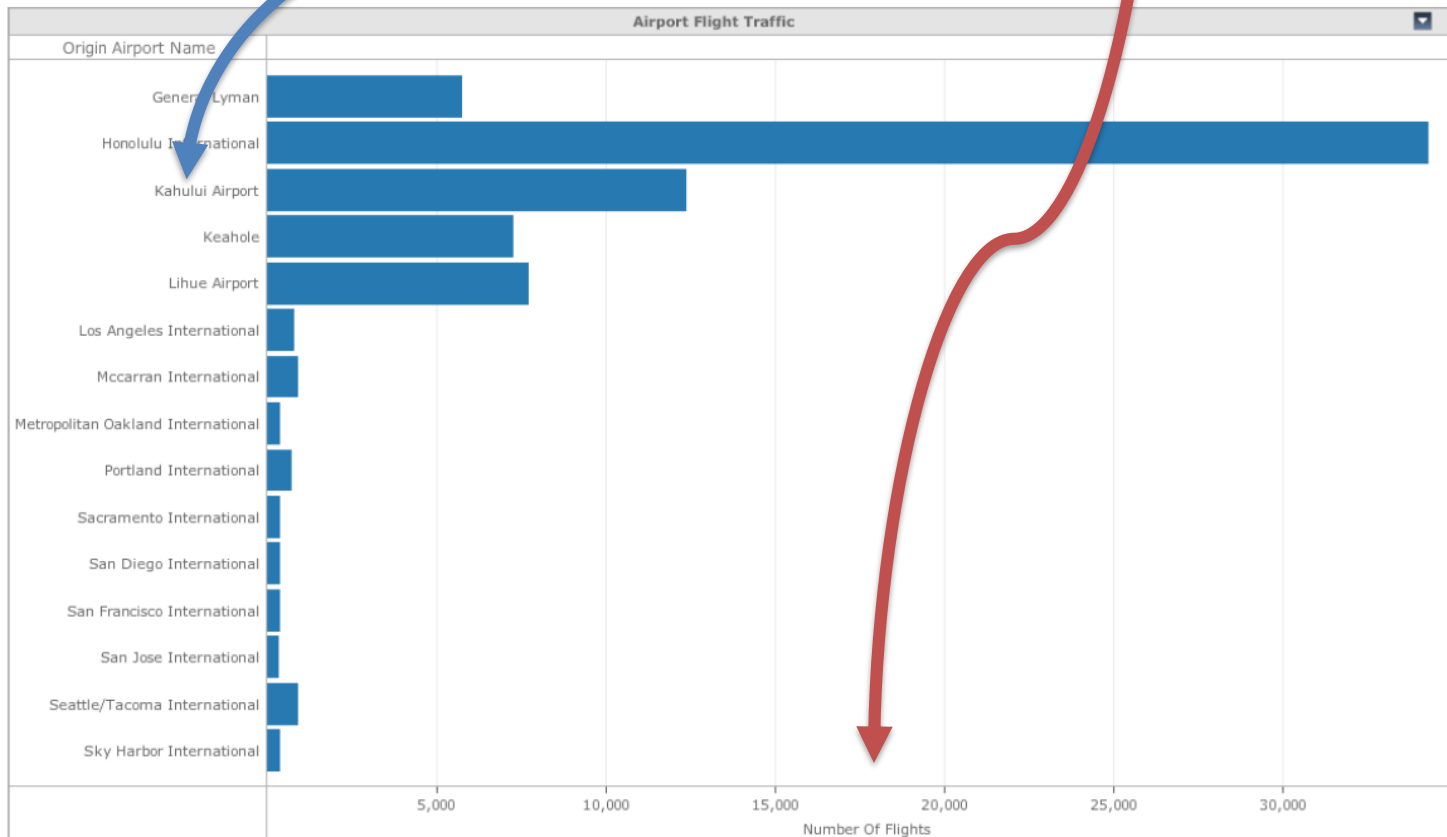
Depending on the type of data we want to visualize

Certain visualizations are better for specific types of data

	Metric	Attribute (Nominal)	Attribute (Ordinal)
Attribute (Nominal)	Bar Heatmap	Network	Line w/ Break-By Bar w/ Break-By
Attribute (Ordinal)	Column Line		Scatter Grid
Metric	Scatter/Bubble		

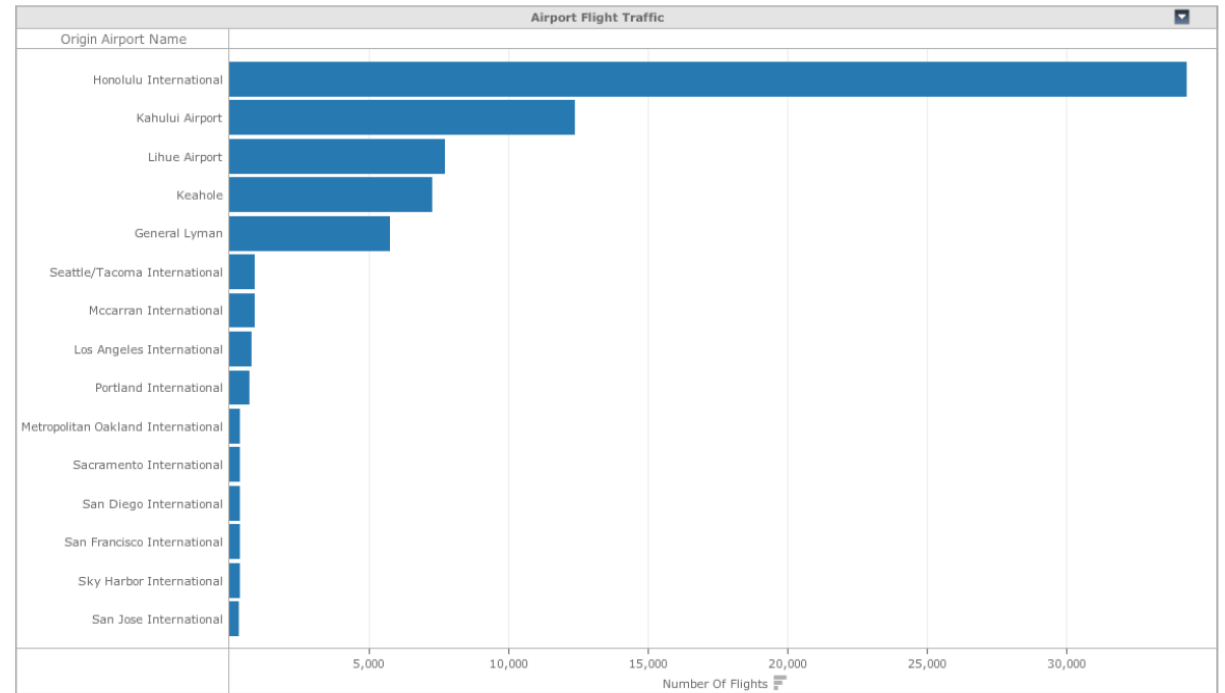
Attribute (Nominal) and Metric

Comparative Analysis Bar Chart - Unsorted



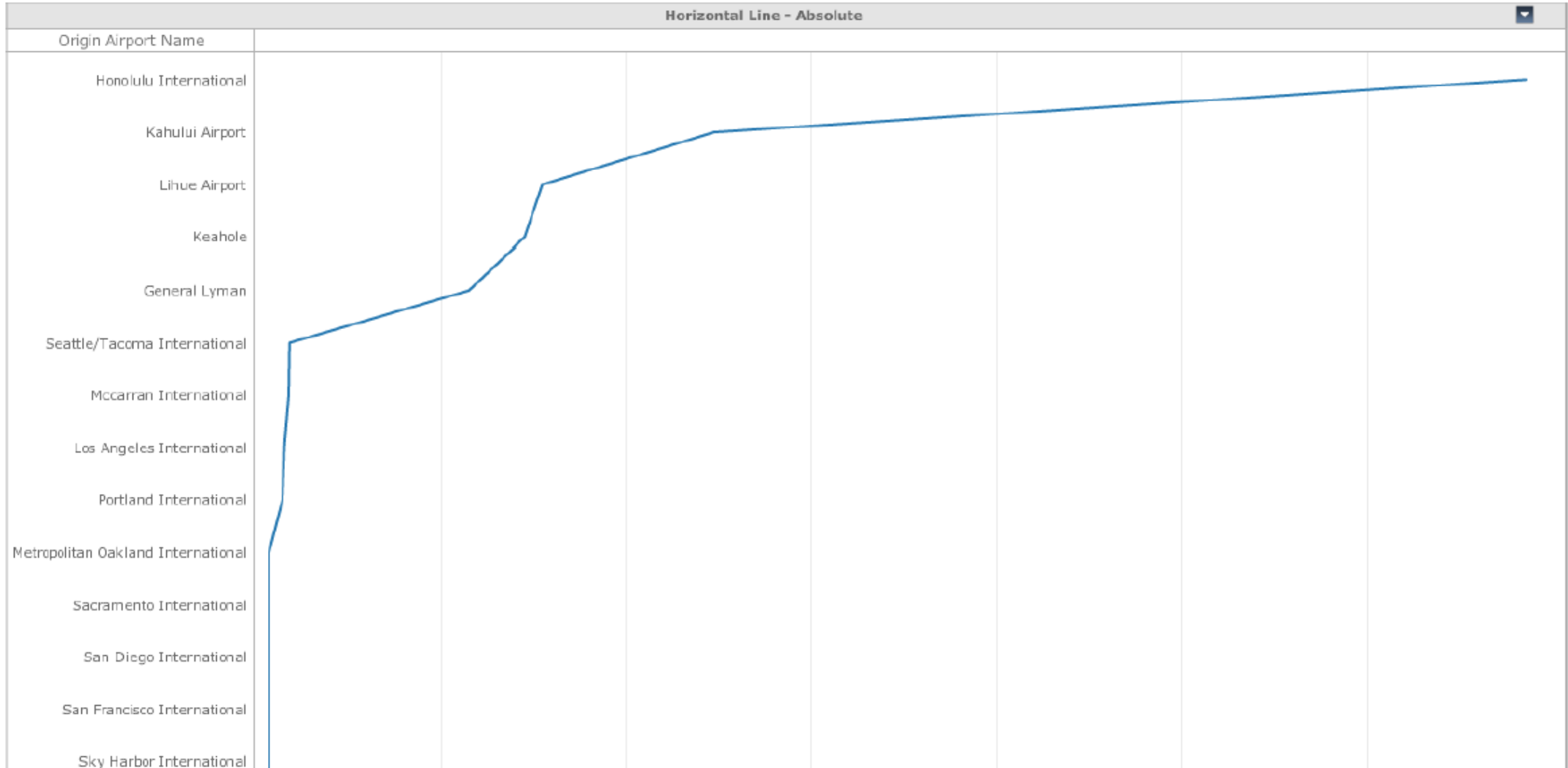
Attributes
(nominal) and
metrics

Comparative Analysis Bar Chart - Sorted



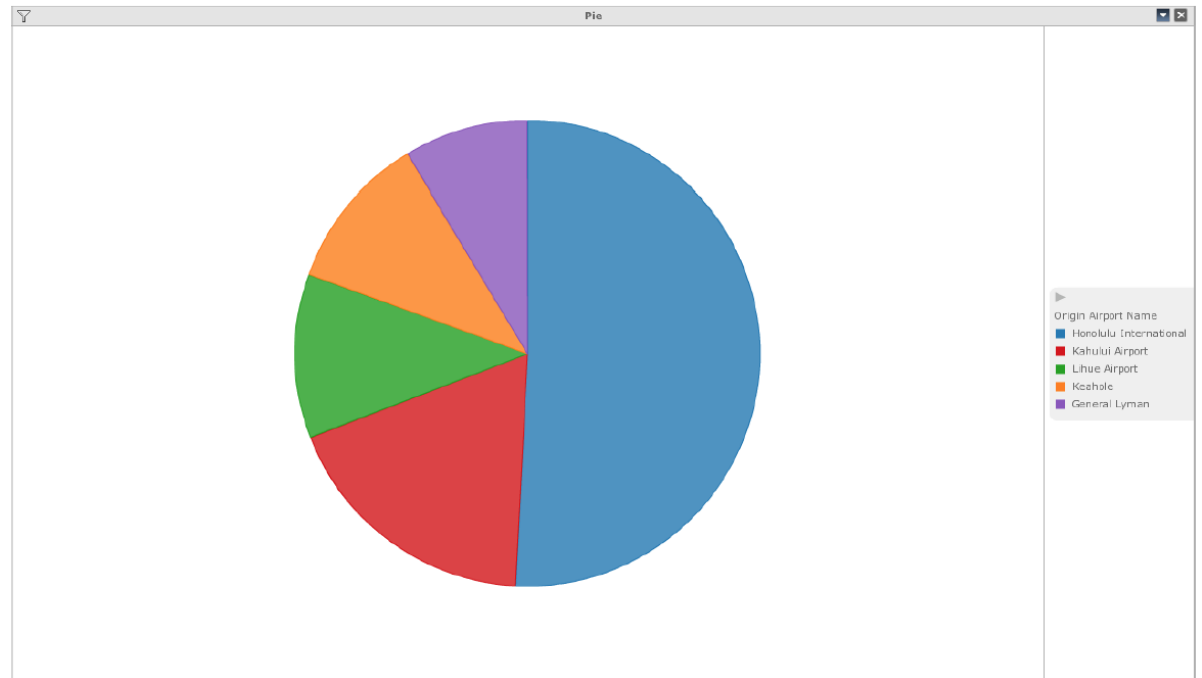
Comparative Analysis

Avoid: Line Chart – Implies continuity between points



Attribute
(Nominal) and
Metric

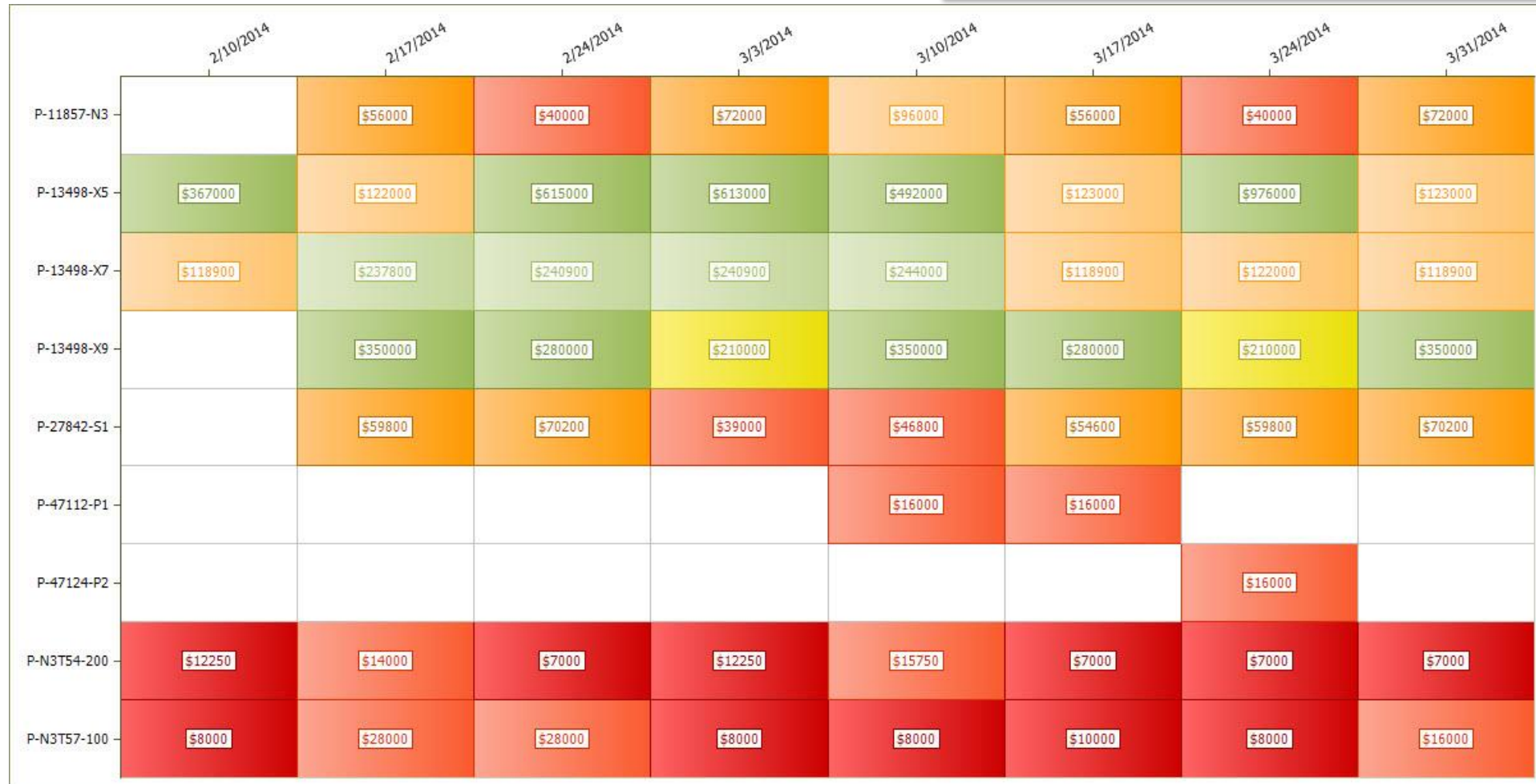
Contribution Analysis – Few Elements Pie Chart



Attribute (nominal) and Metric

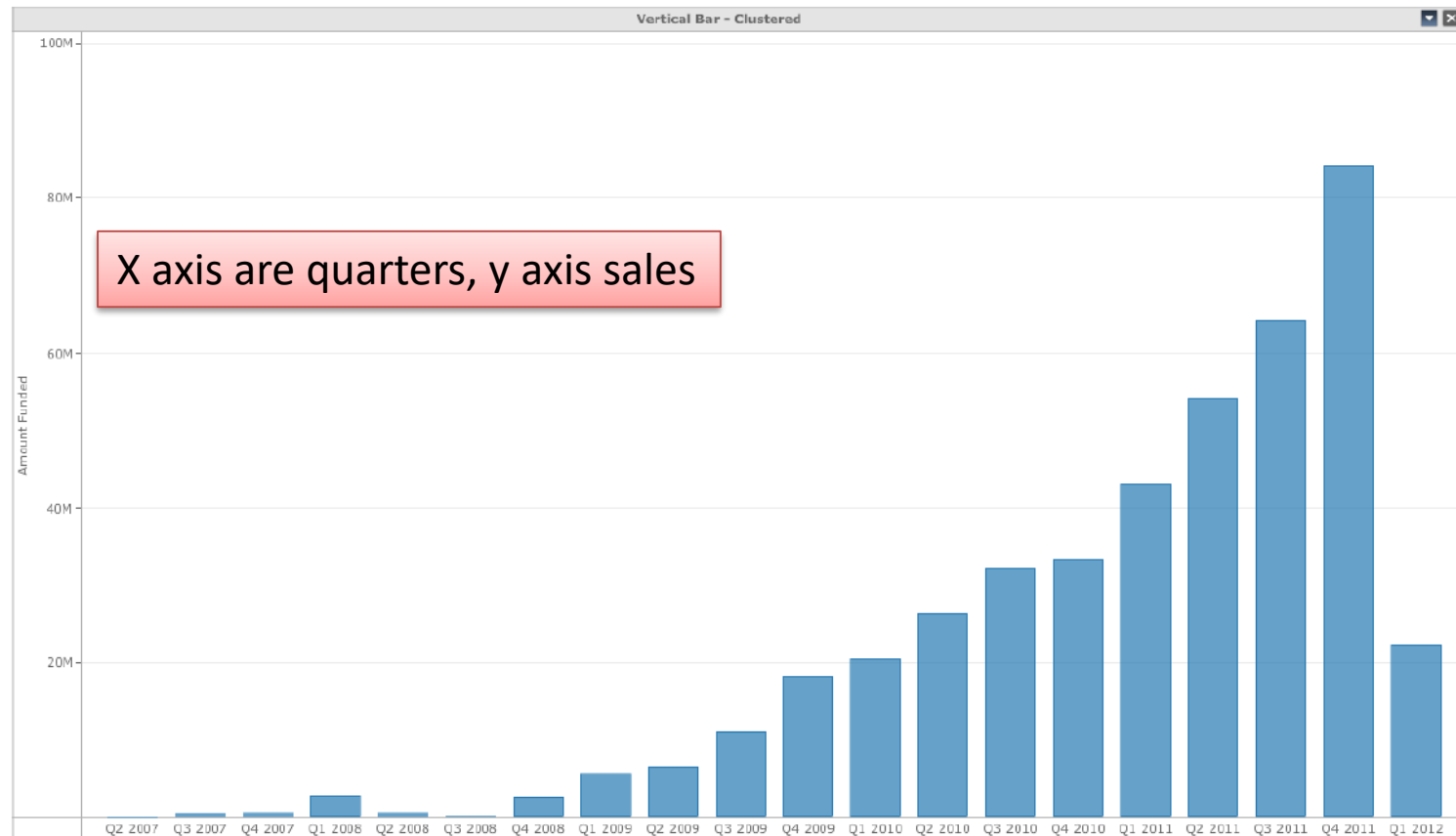
Many elements: heat map

Here 3 elements, data, products and sales



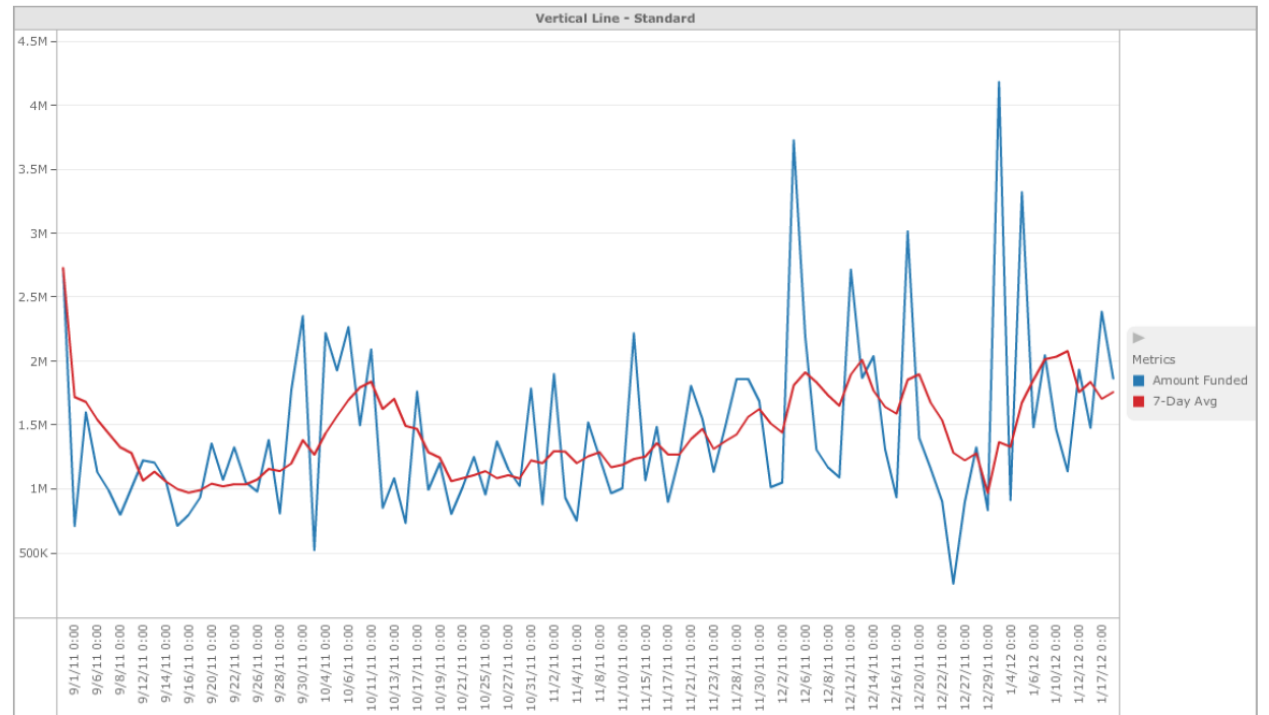
Attribute (Ordinal) and Metric

Time-Series Analysis – Few Elements Column Chart



Attribute (Ordinal) and Metric

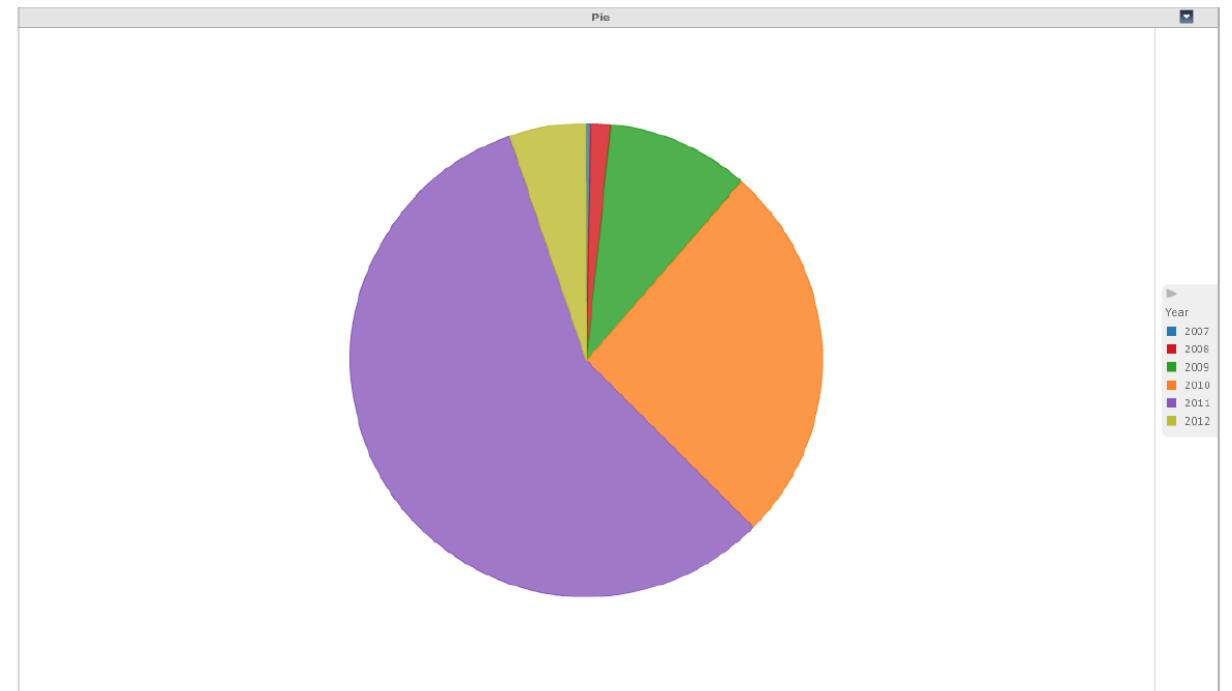
Time-Series Analysis – Many Elements Line Chart



Time series

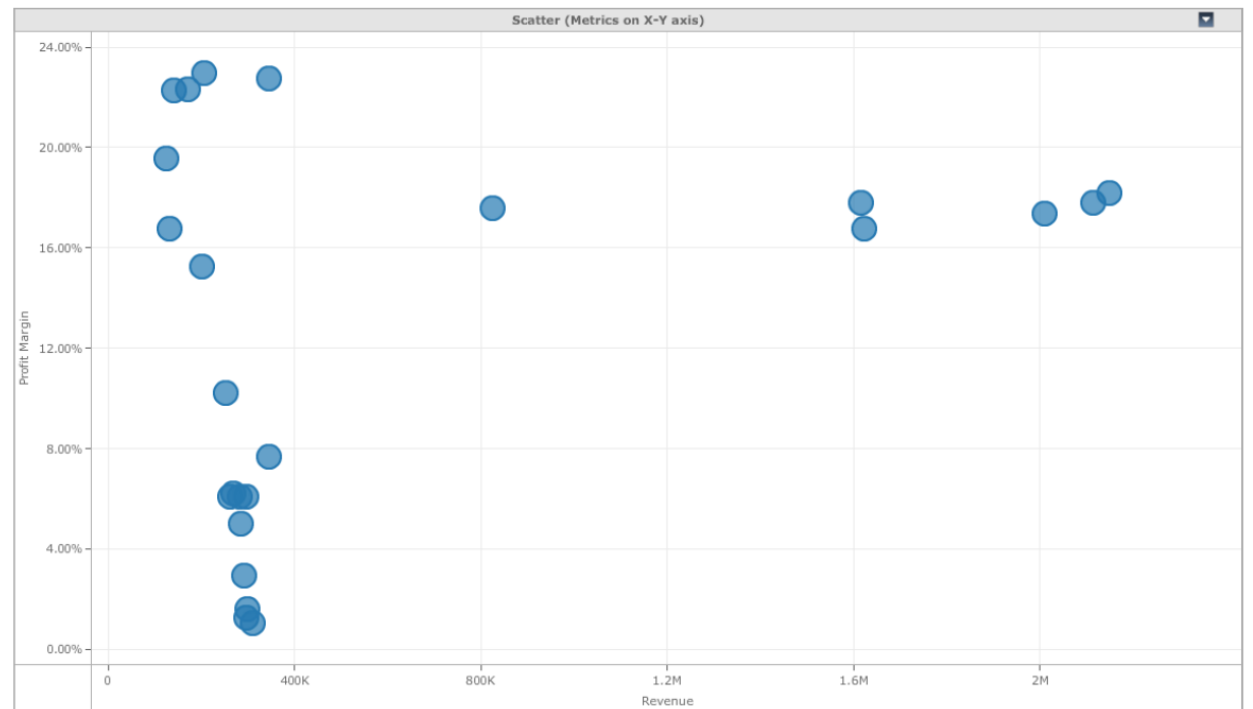
Time-Series Analysis

Avoid: Pie Chart – Removes ordinality



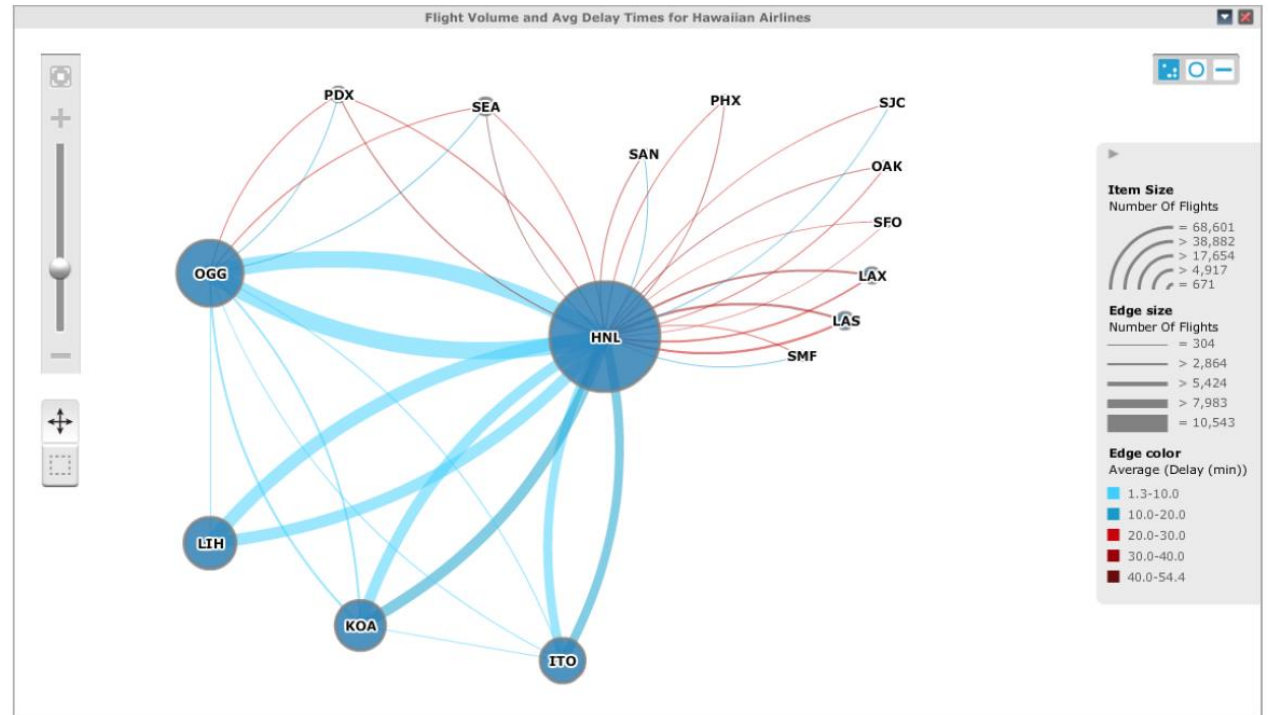
Metric and Metric

Correlation Analysis Scatter Plot



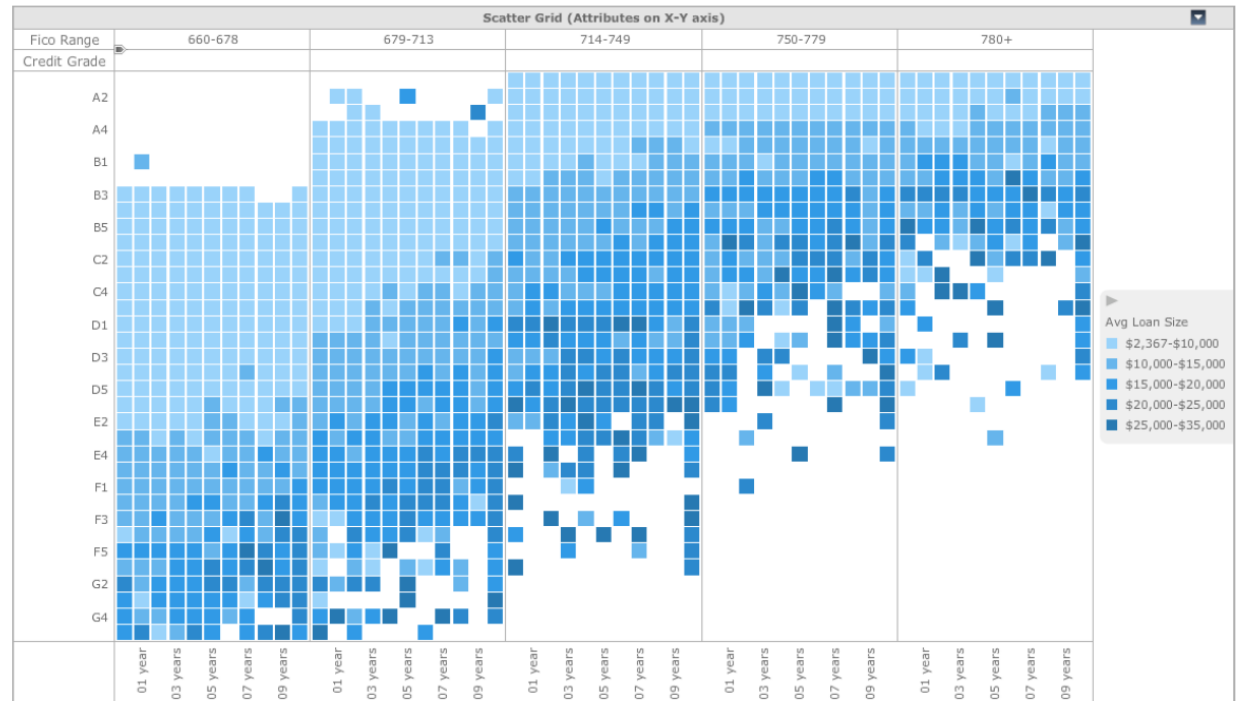
Attribute
(Nominal) and
Attribute
(Nominal)

Market Basket or Network Analysis Network Visualization



Attribute (Ordinal) and Attribute (Ordinal)

Cluster or Heatmap Analysis Scatter Grid



How to **improve** a visualization

- Colors
- Saturations
- Size
- Interpretability
- Performance
- Layout
- Interactivity

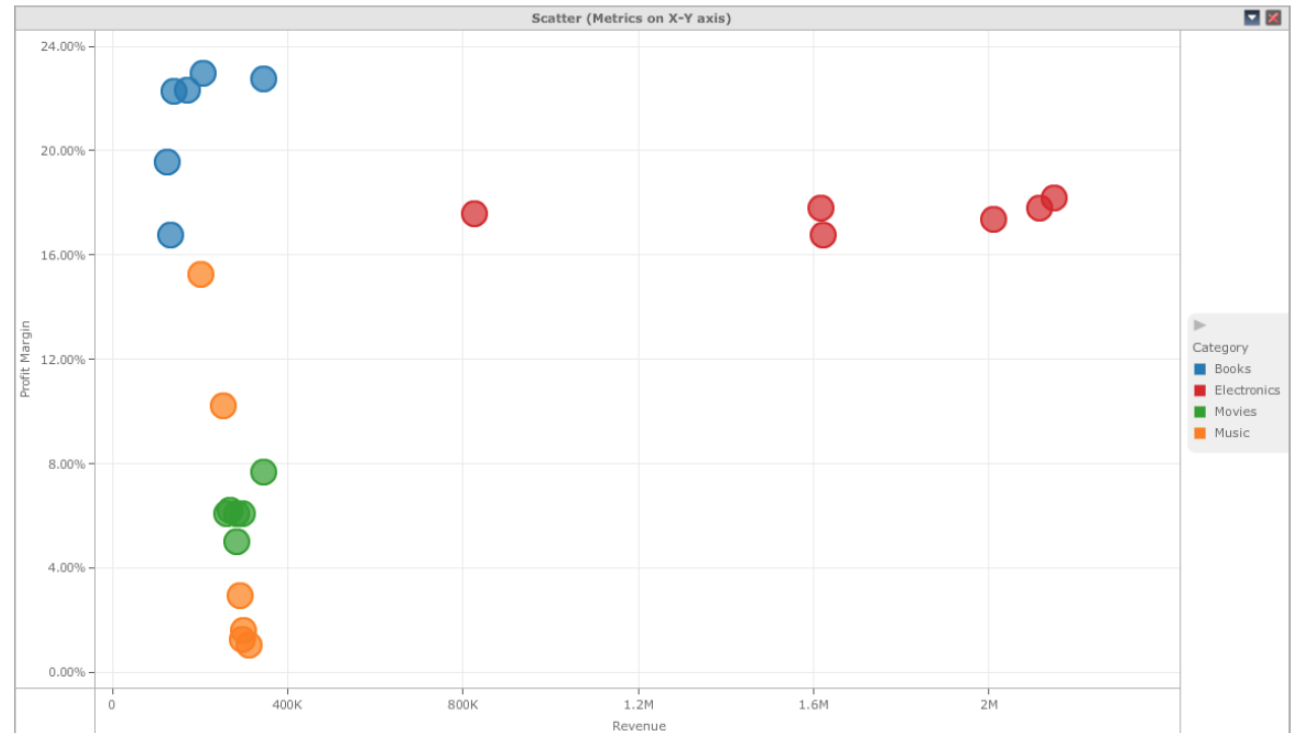


Enhancing Visualizations for Additional Insights

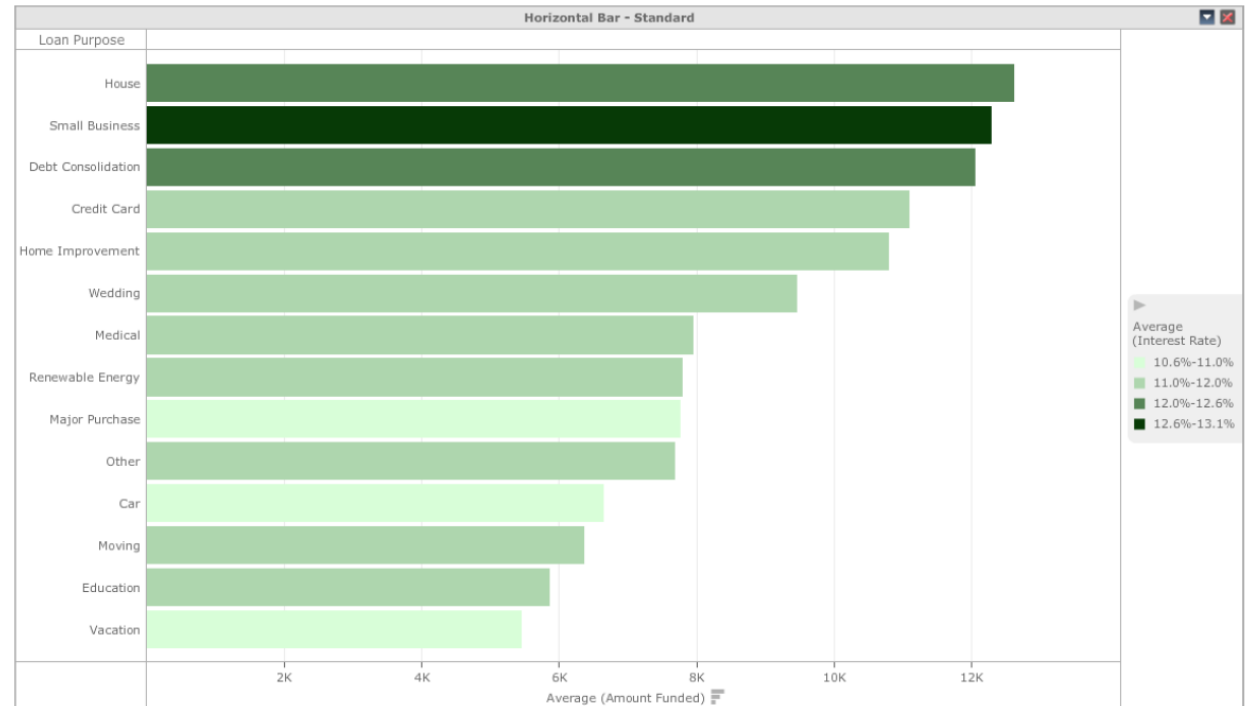
Appropriate Visual Enhancements

	Attribute (Nominal)	Attribute (Ordinal)	Metric
Color Hue	x	x	x
Color Saturation		x	x
Size		x	x

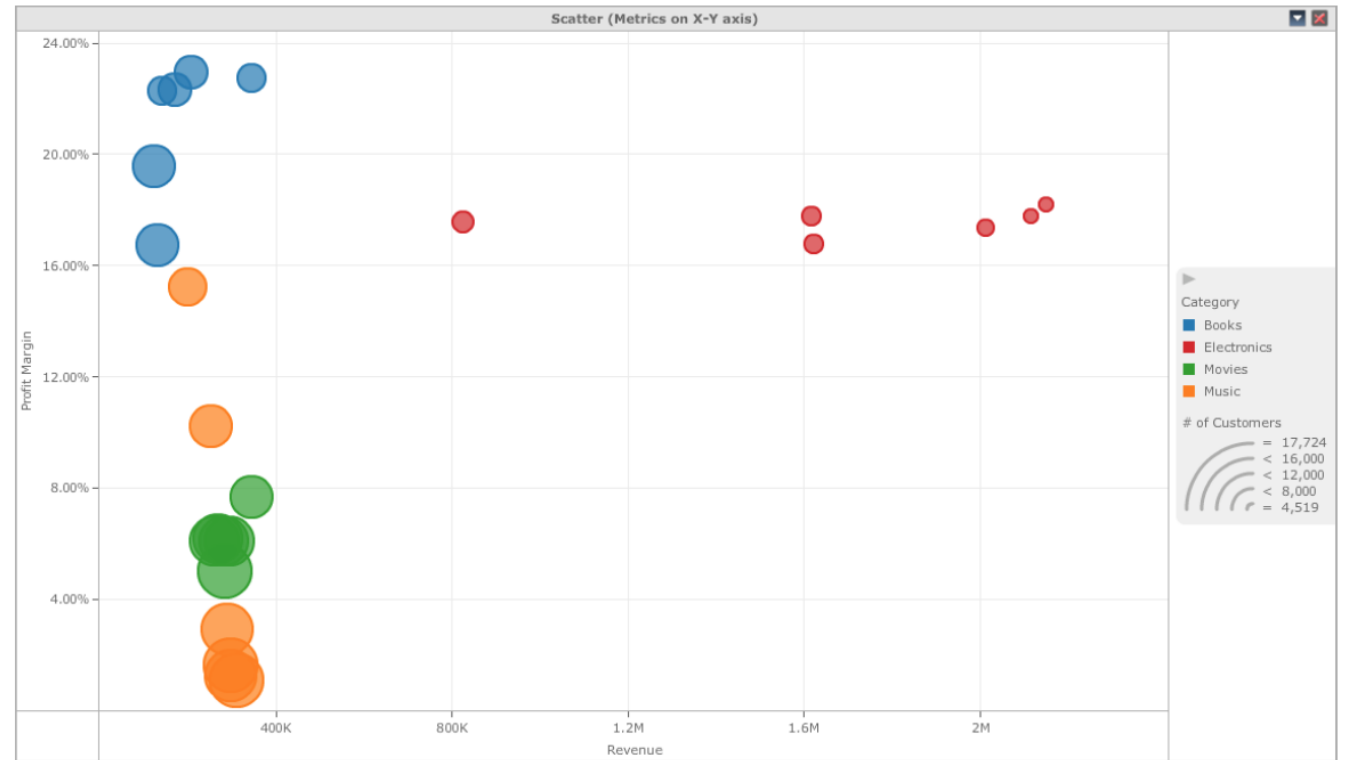
Color (Hue) to Identify Groups of Attributes



Color (Saturation) to Highlight Metric Patterns



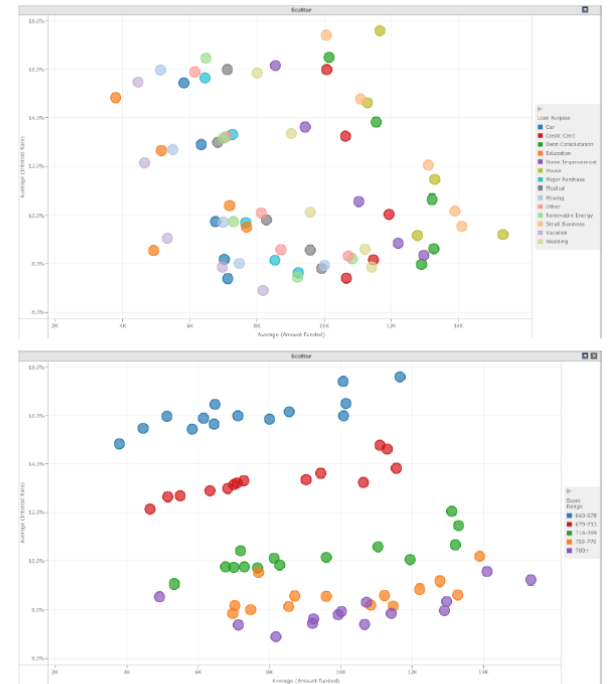
Adding Size to Emphasize Metric Trends



More hints on colors

Colors Should Enhance Data Comprehension, Not Distract

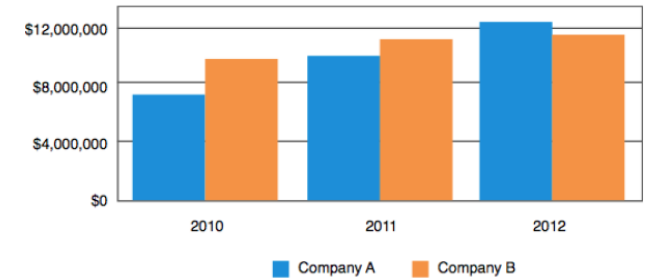
Use Fewer Than 6 Colors



More hints on
colors

Use Colors to Emphasize Comparisons

Use Opposing Colors for Comparisons



More hints on colors

Avoid color confusion!

Why should we both change from our home kits?

Home Jerseys

Wales



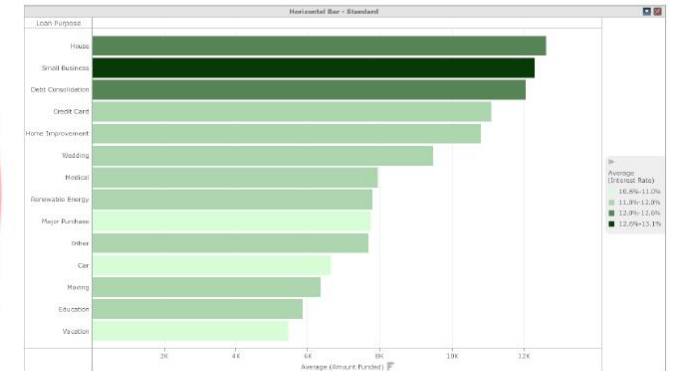
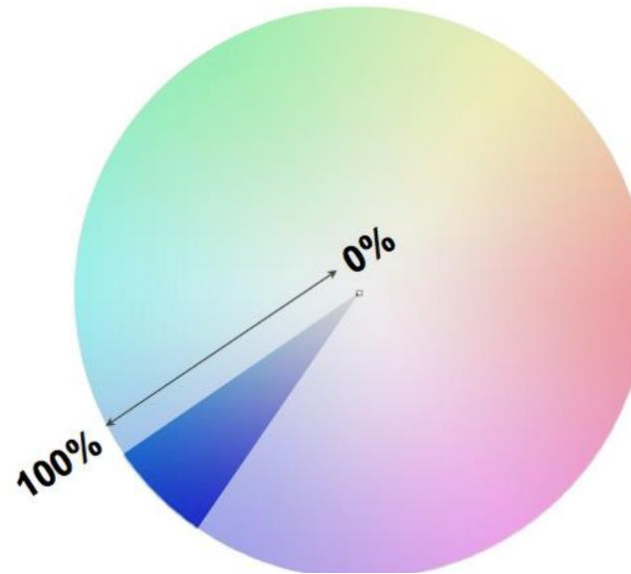
Portugal



More hints on colors

Use Color Saturation Correctly

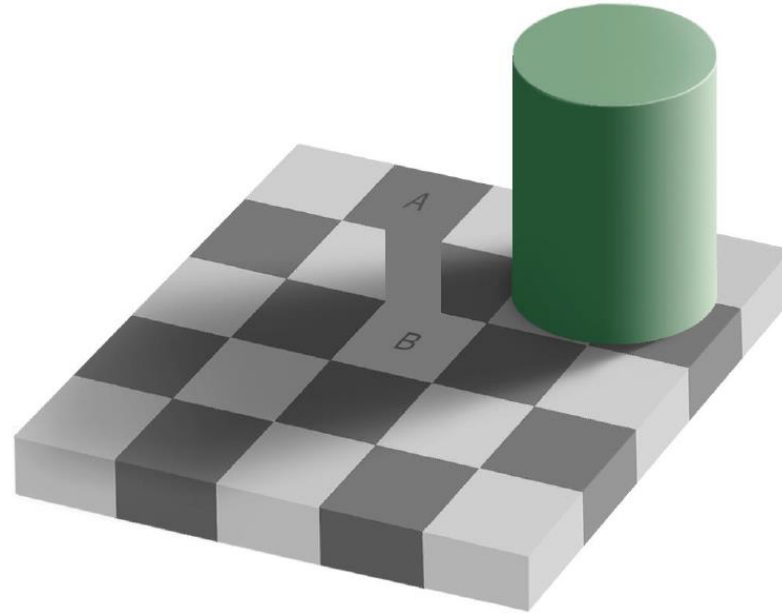
Less Saturation: Smaller Values
More Saturation: Greater Values



More hints on
colors

Color Constancy Can Confound Data Comprehension

Avoid Color Gradients for Backgrounds



Interpretability: avoid users do the math

- Which months were below target?
- Which months were above target?
- And by how much?

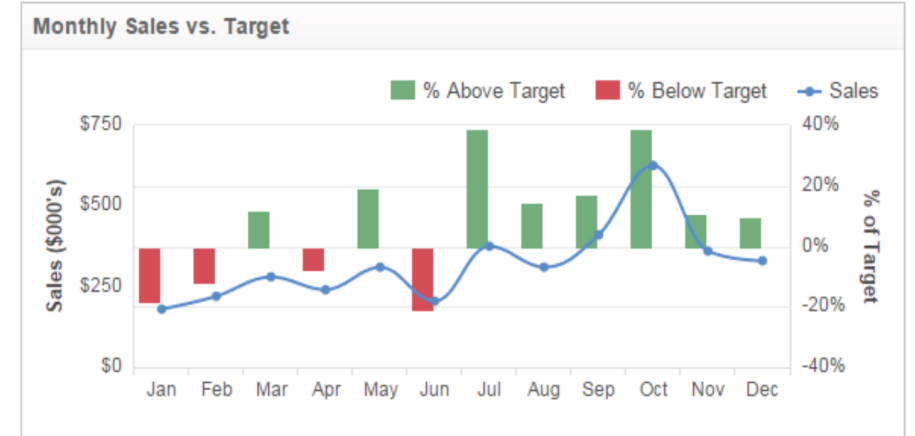
Looks typical right?



Interpretability: avoid users do the math

- Shows Actuals
- Shows Good and Bad months
- Quantifies good and bad
- Uses and Overlay and dual Y Axis.

Go the extra mile to make it great



Interpretability: sloppy labelling

Make it idiot proof, I mean “self explanatory”!

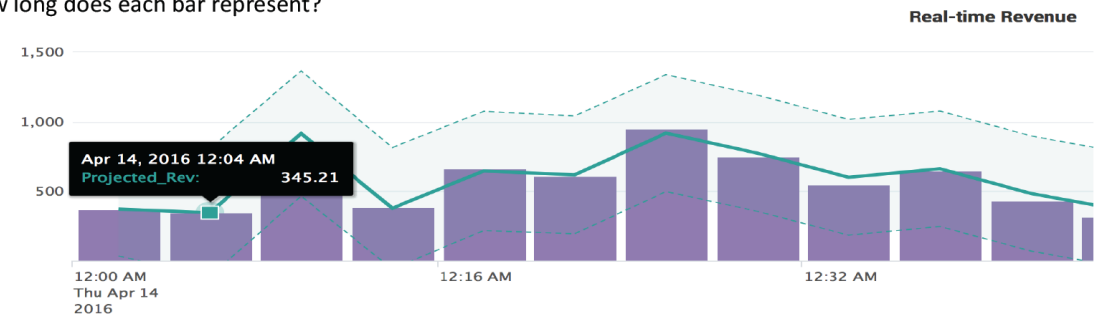
- What is the unit of measure on the Y axis?
- Are you using a log scale? (If so, mention it)
- Are the numbers shown in K's, M's or B's?
- If Currency, which currency?
- Net Revenue or Gross Revenue?
- How long is the rolling window being shown?
- How long does each bar represent?



Real Time Revenue

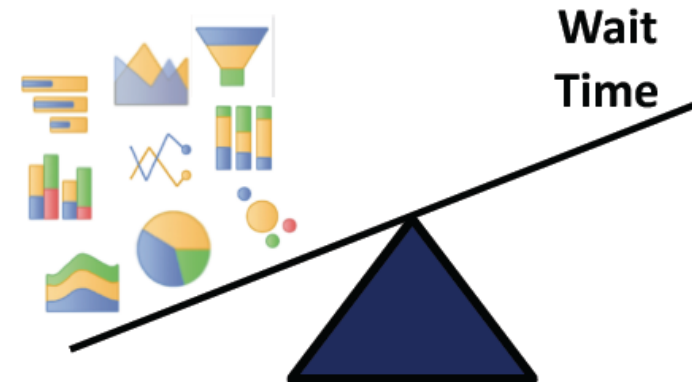


Real Time Gross Revenue (\$) last 1 hour in 4 minute buckets



Performance: reduce load times

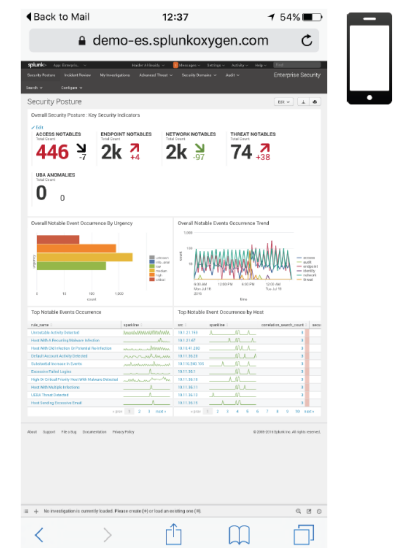
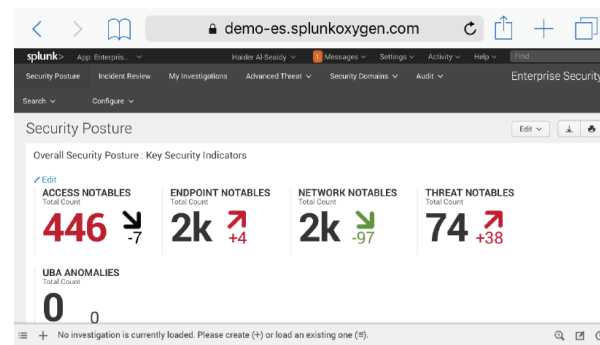
- **Limit the number of objects on a single screen**
- Limit real-time searches
- Specify filters to reduce the data
- Look for long-running searches and schedule them in advance (e.g. do not load twitter data at at query time)
- Use **summary indexes** to reduce the search load of the dashboard
- **Look for commonality between searches. When found, combine several searches into one**



Layout: size to the right width

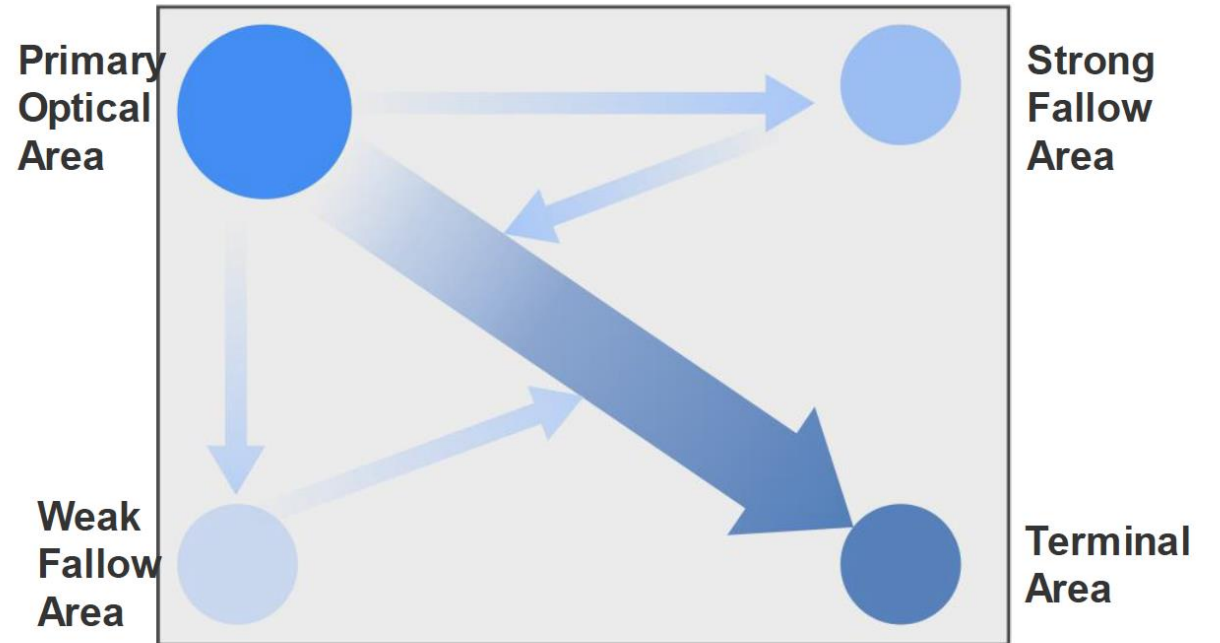
Orientation

- Landscape mode on a mobile will require scrolling to see content
- Are your key metrics at the top?
- Sizes to the width of the app



Layout:
People Have a
Bias in How
They Read and
Scan Content

Reading Gravity



Layout: place most relevant content in primary optical area

Reading Gravity

Primary Optical Area

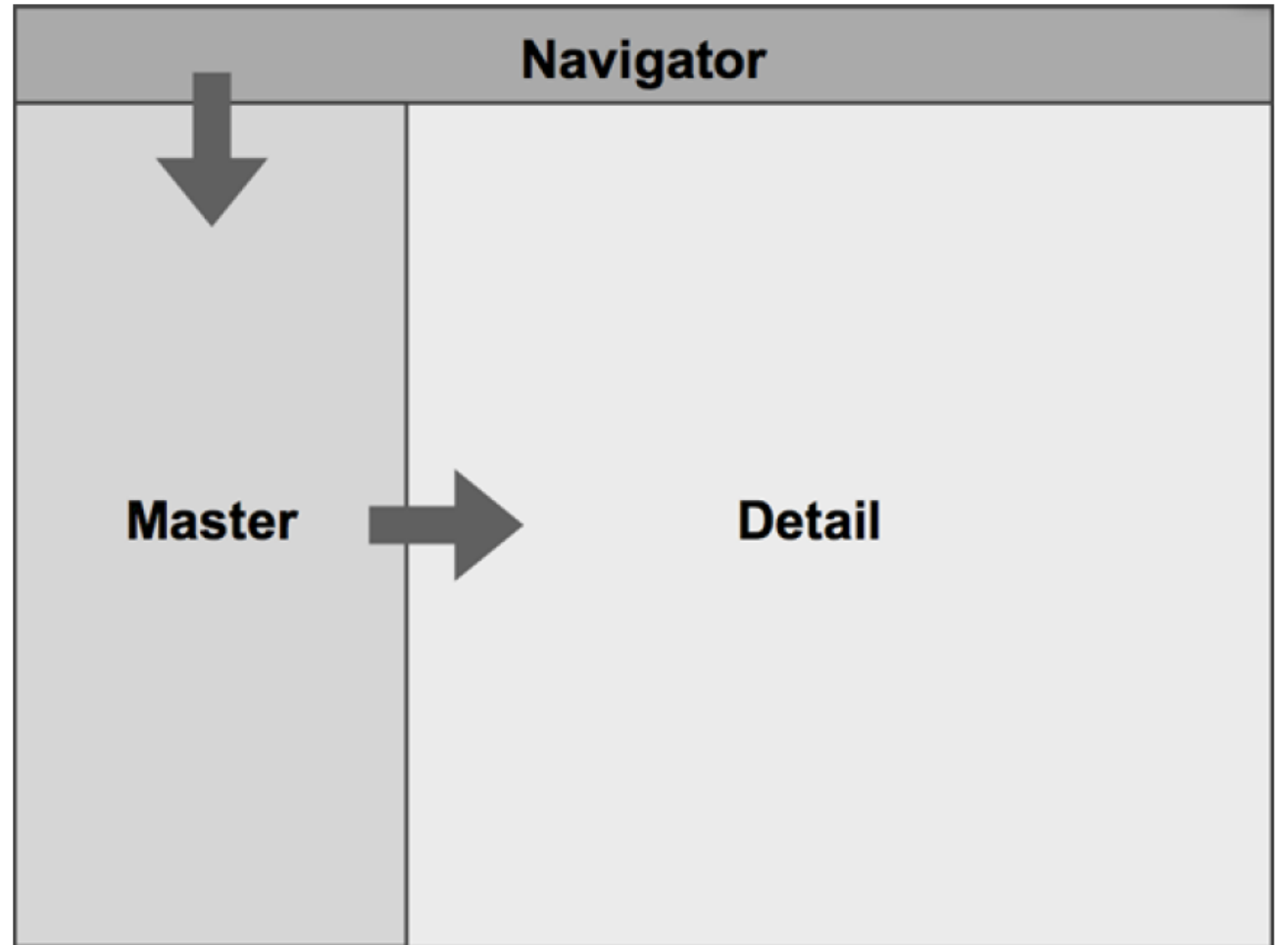


Strong Fallow Area

Weak Fallow Area

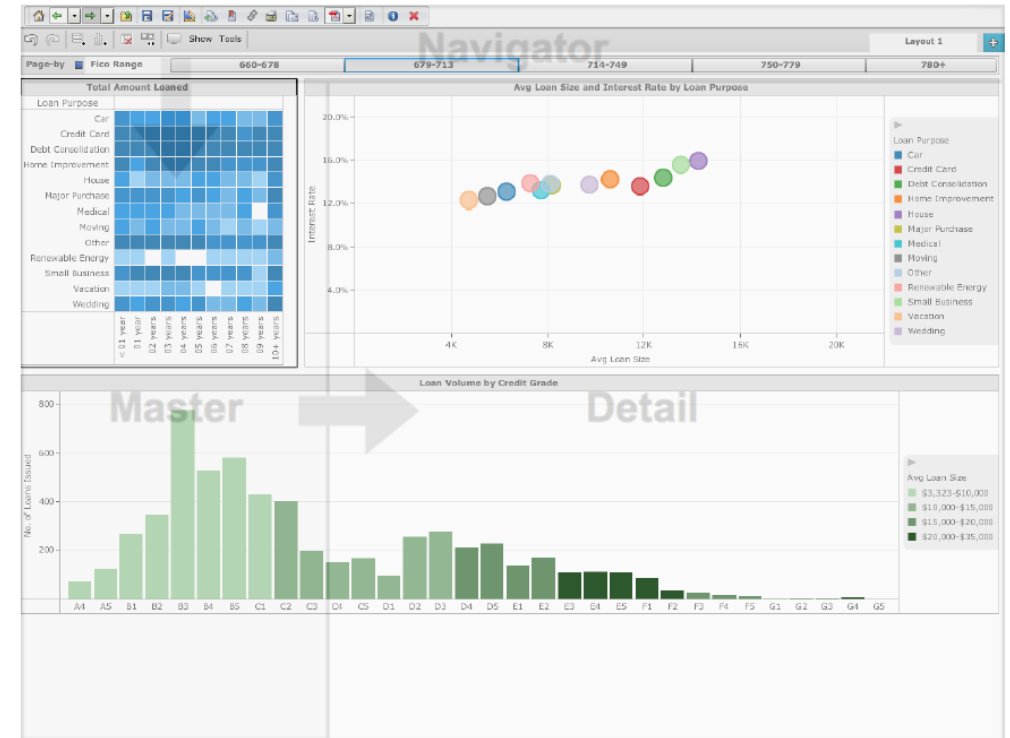
Terminal Area

Layout: present
content
hierarchically



Layout: present content hierarchically

Present Data Hierarchically

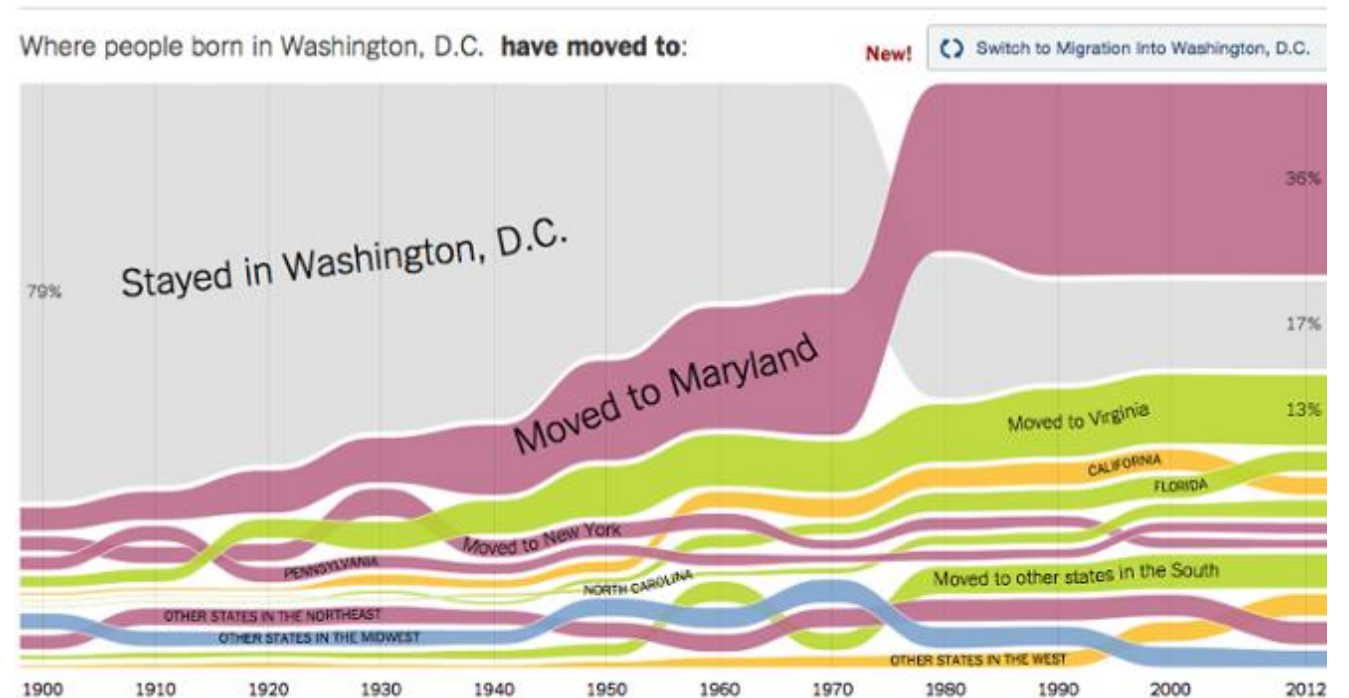


Interactivity & animation

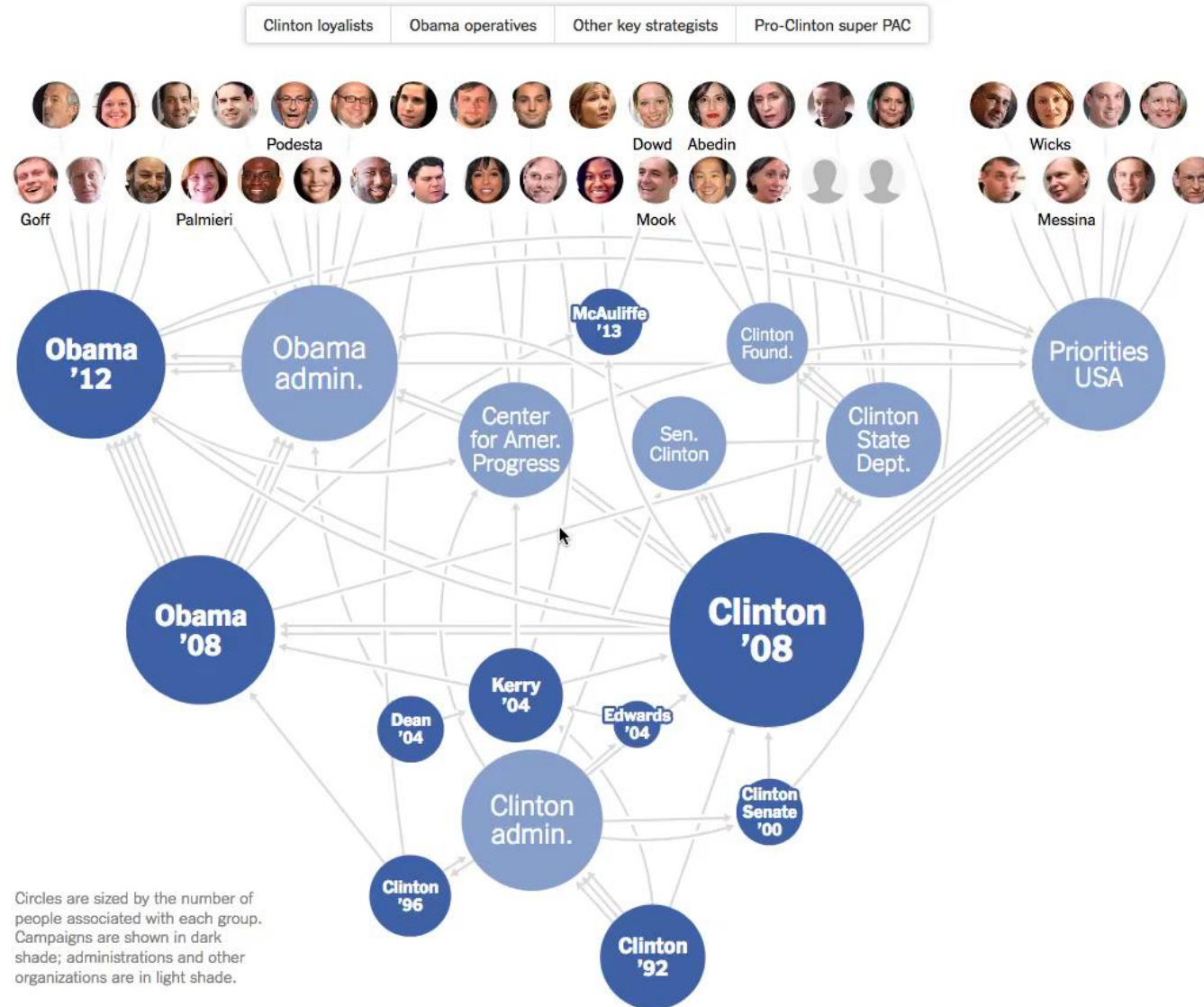
- Interactivity and animation are the latest and coolest features for presenting information
- User can adapt a visualization to his/her own needs and curiosity, interacting with the map
- Interactive maps –when well designed – greatly improve the efficacy of an interaction
- Some very cool visualizations: (see more on <http://blog.visme.co/best-information-graphics-2016/>)

[The New York Times' project](#) on where people born in a state move to

- It visualizes a large amount of data accumulated during more than
- 100 years. Yet, it is easy to understand, and it clearly highlights
- interesting trends.

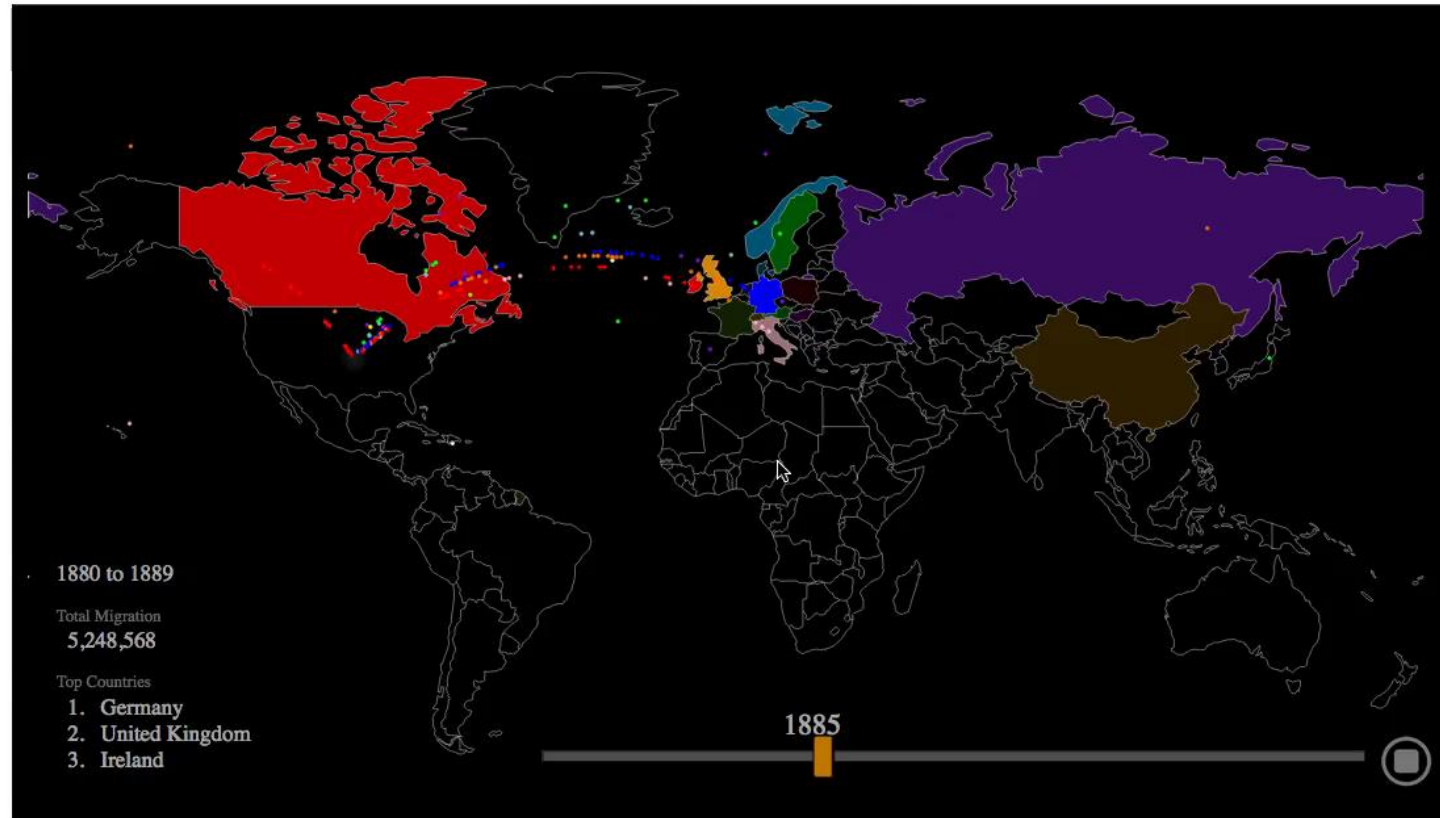


Politicians & Political campaigns



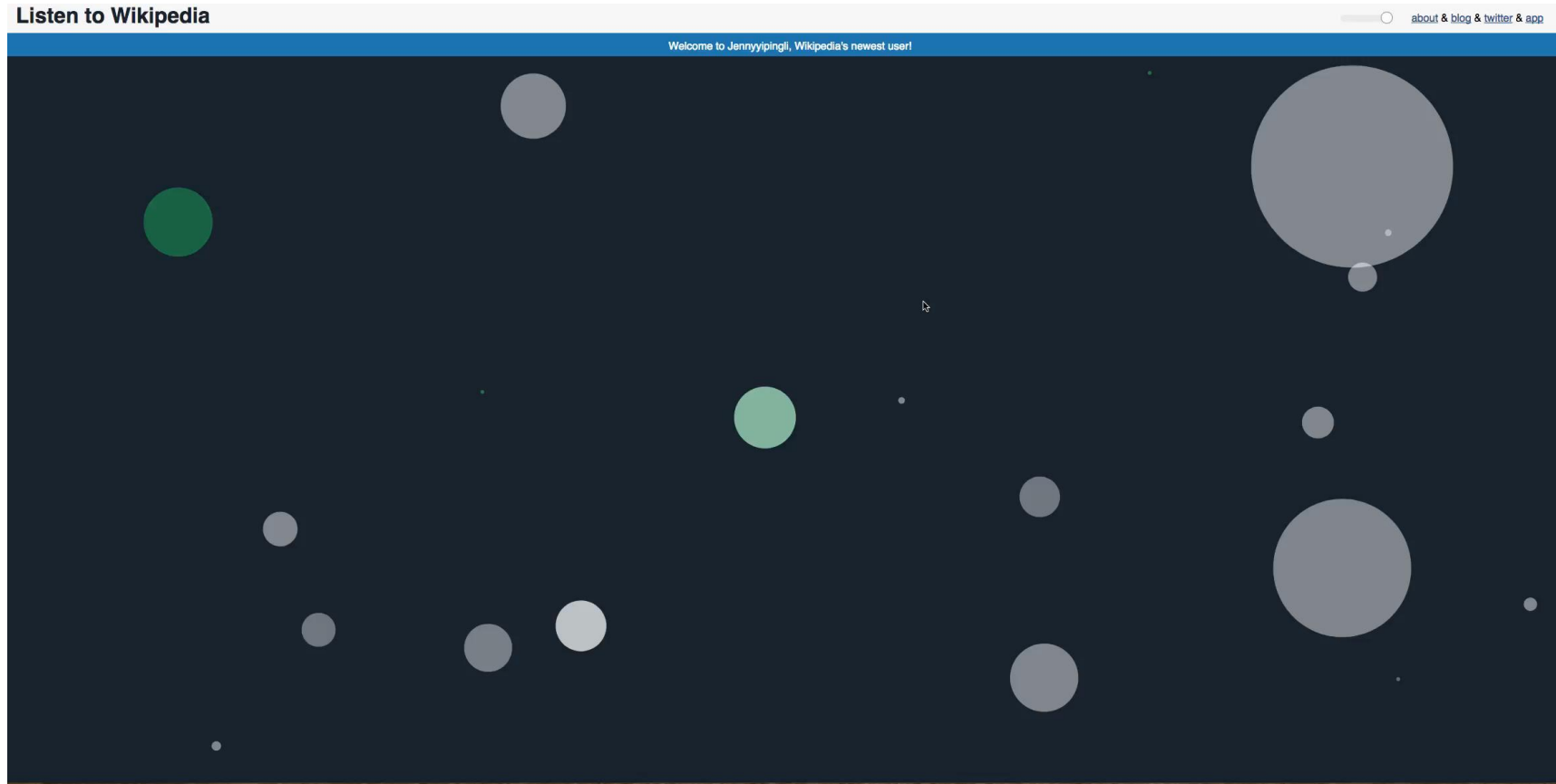
Immigration

Two Centuries of U.S. Immigration (1 dot = 10,000 people)



Full screen interactive map / HD video

Lissen to Wikipedia



International Trades

Click on a country to see its share of trade alone, or spin/navigate the globe by using your mouse.



Interactive: Mapping the Flow of International Trade

SPONSORED INFOGRAPHICS



HIGHCHARTS

- <https://www.highcharts.com/demo>

Homework

- Search on the web for fancy and new visualizations
- Can you imagine how a given visualization that you found can be useful for some business application?
- In your final project, when designing your report with Watson, **pay attention to indications provided in this presentation when designing the report layout and visualizations.**