



# Data Visualization

---

Xccelerate - Data Science Immersive

# | DATA VISUALIZATION – KEY PRINCIPLES

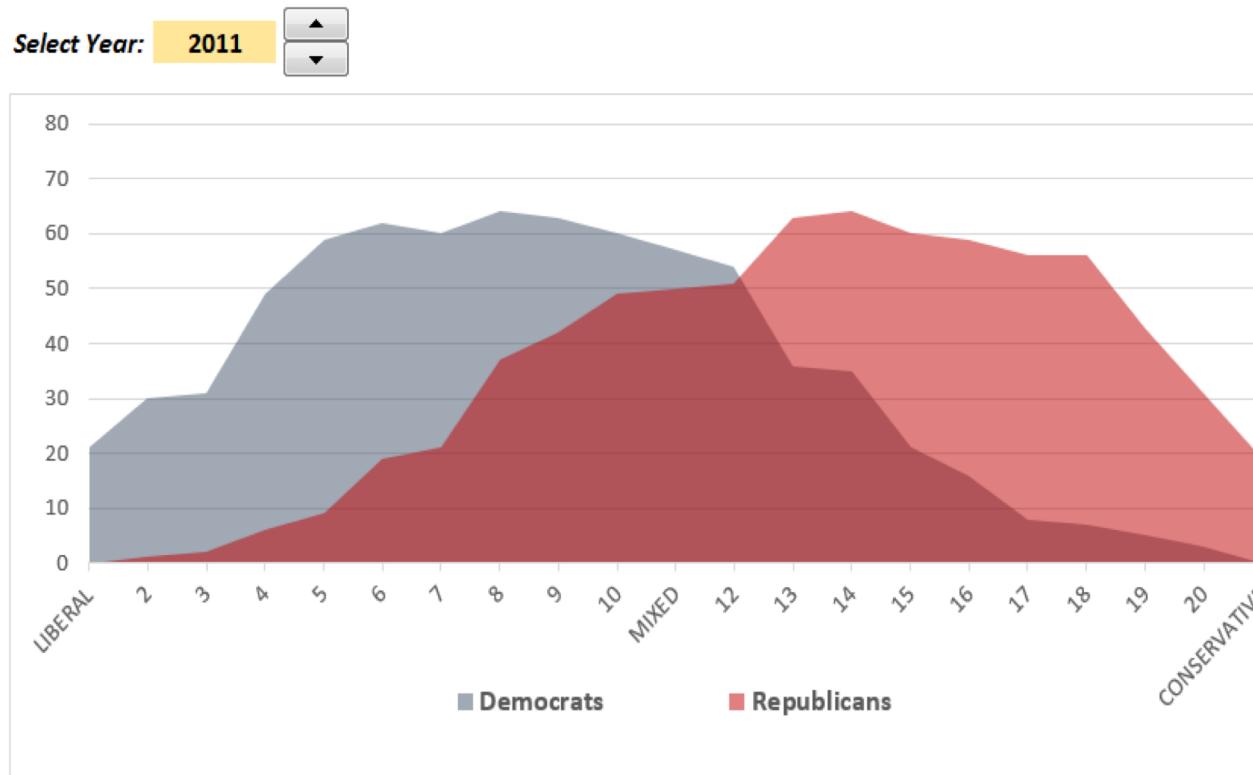


- Strive for **CLARITY & SIMPLICITY**
  - *Maximize impact, minimize noise*
  - *If it doesn't add value or serve a purpose, get rid of it*
- Focus on creating a **NARRATIVE**
  - *Don't just show data, tell a story about it*
  - *Communicate key insights clearly, quickly and powerfully*
- Strike a balance between **DESIGN & FUNCTION**
  - *Selecting the right type of chart is critical*
  - *Beautiful is good, functional is better, BOTH is ideal*

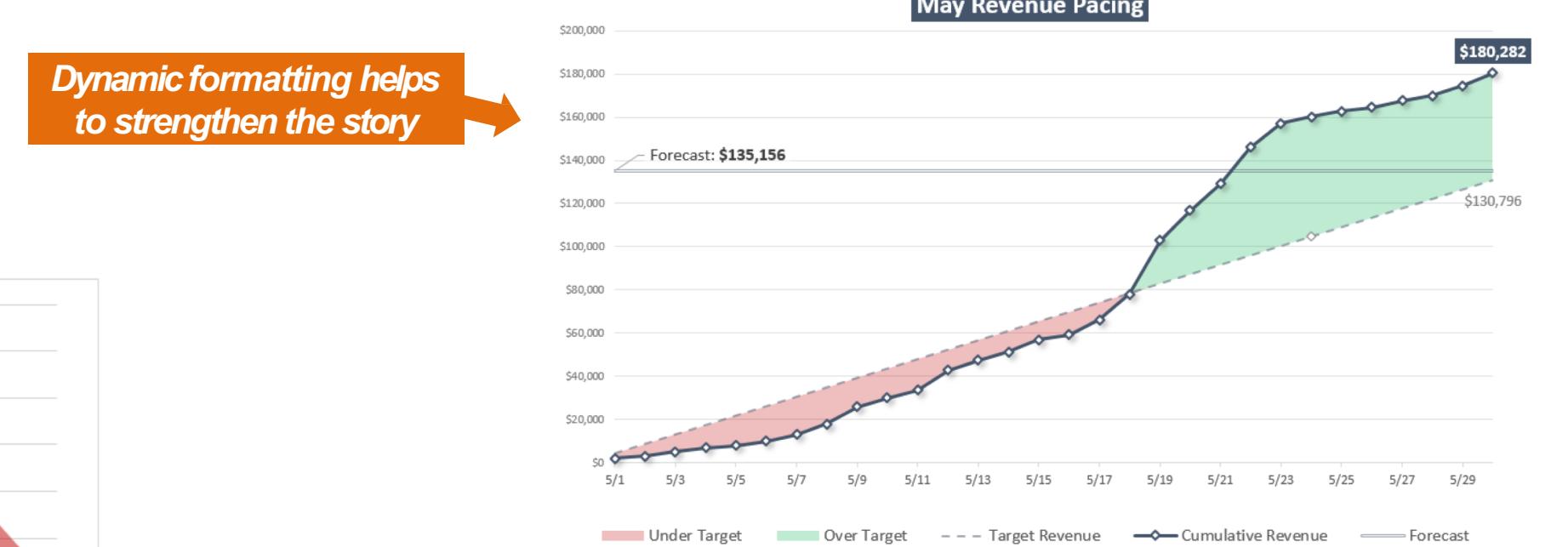


# THE GOOD, THE BAD, AND THE UGLY

## THE GOOD:



Clean, simple visualization with animation over time



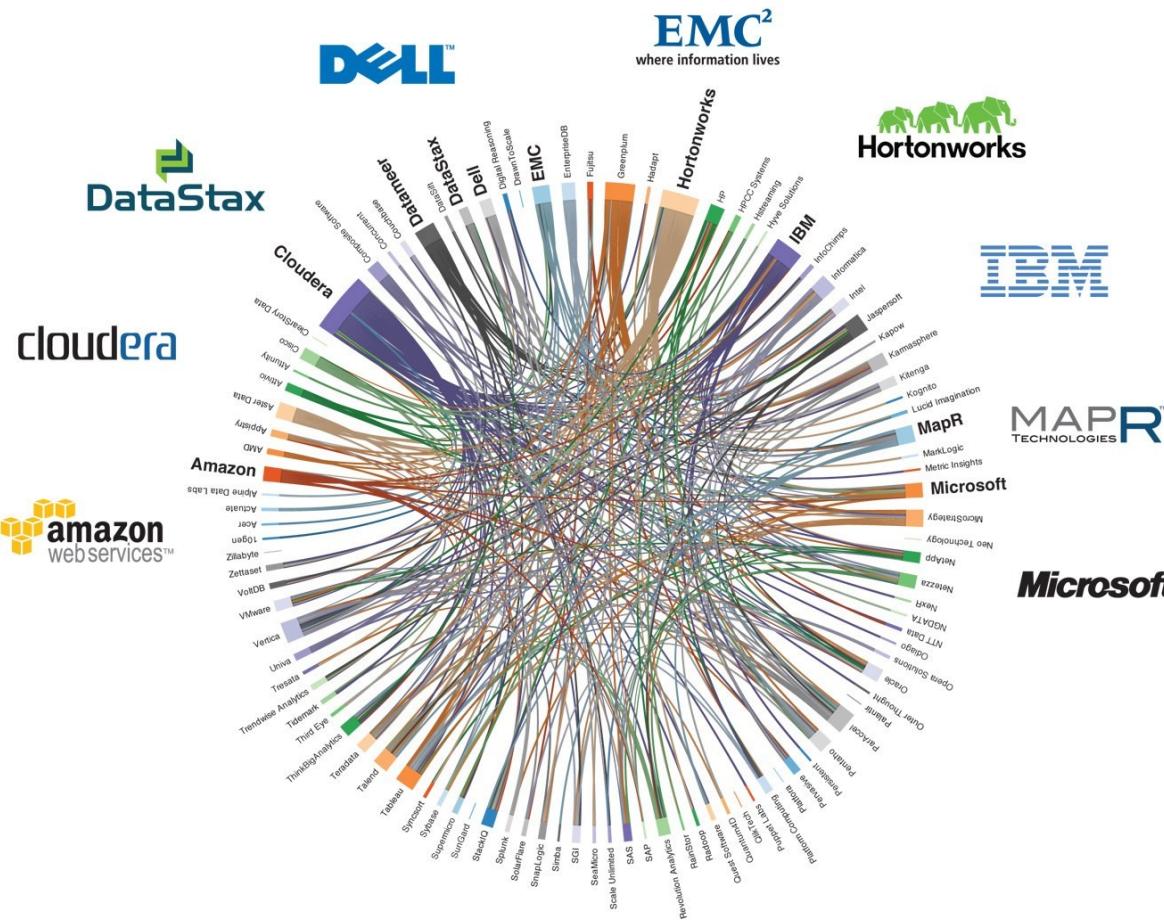
Intuitive custom chart design

# THE GOOD, THE BAD, AND THE UGLY

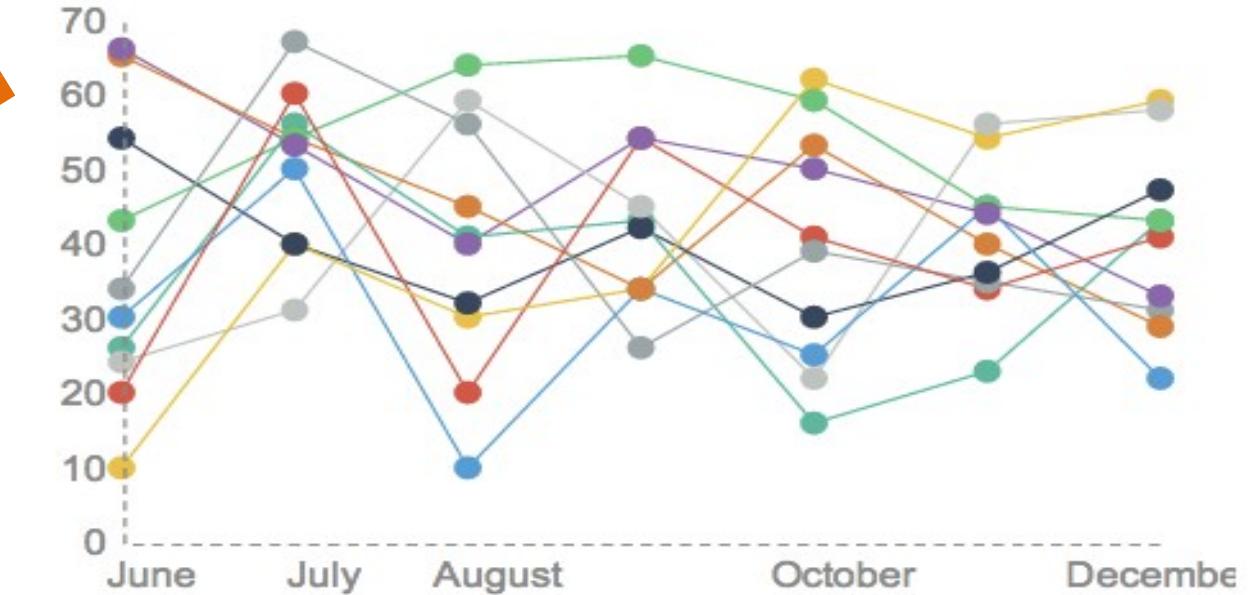


## THE BAD:

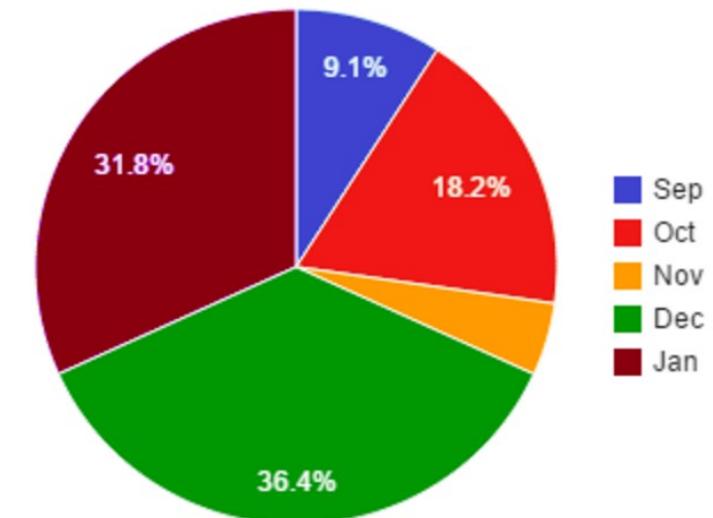
*Only Design, No Function, Useless*



*Busy, no clear narrative*



Monthly Sales

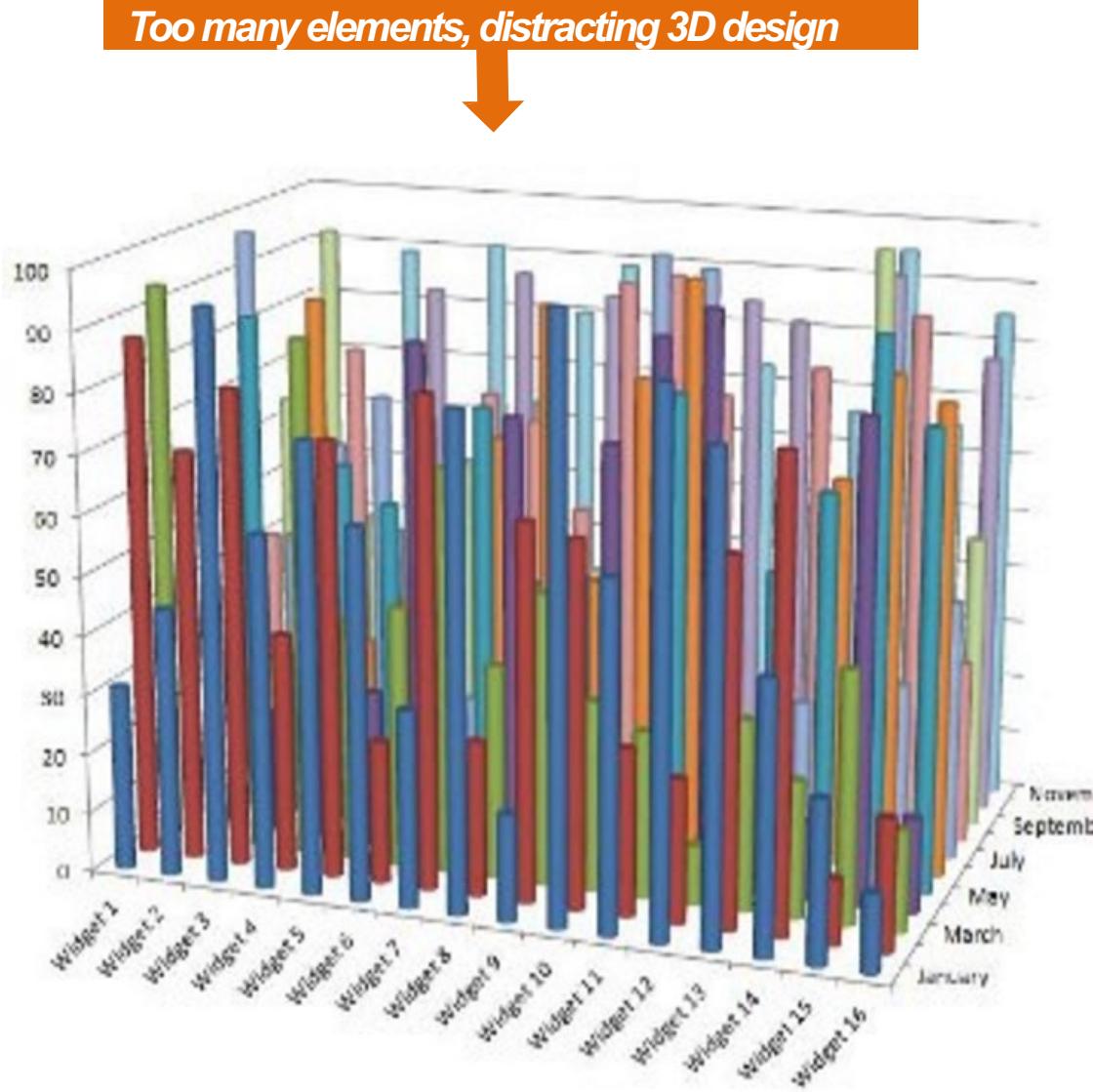


*Misleading Chart Type*

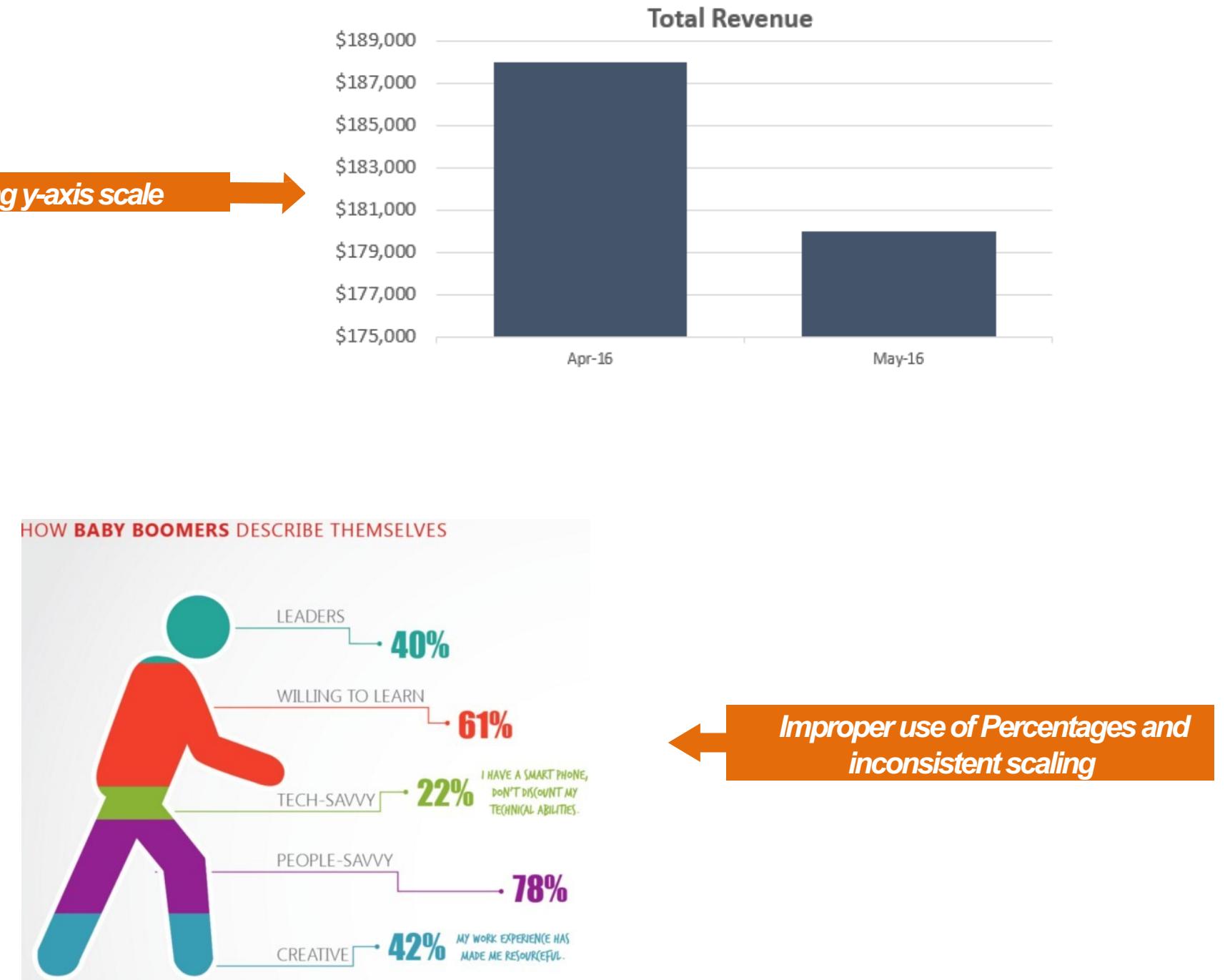


# THE GOOD, THE BAD, AND THE UGLY

## THE UGLY:



*Misleading y-axis scale*





# | THE 3 KEY QUESTIONS

1

- What **type of data** are you working with?
  - *Integer, real number, categorical, time-series, geo-spatial, etc.*

2

- What are you trying to **communicate**?
  - *Relationship, comparison, composition, distribution, trending, etc.*

3

- Who is the **end user** consuming this information?
  - *Analyst, CEO, client, intern, etc.*

# BAR & COLUMN CHARTS



## COMMONLY USED FOR:

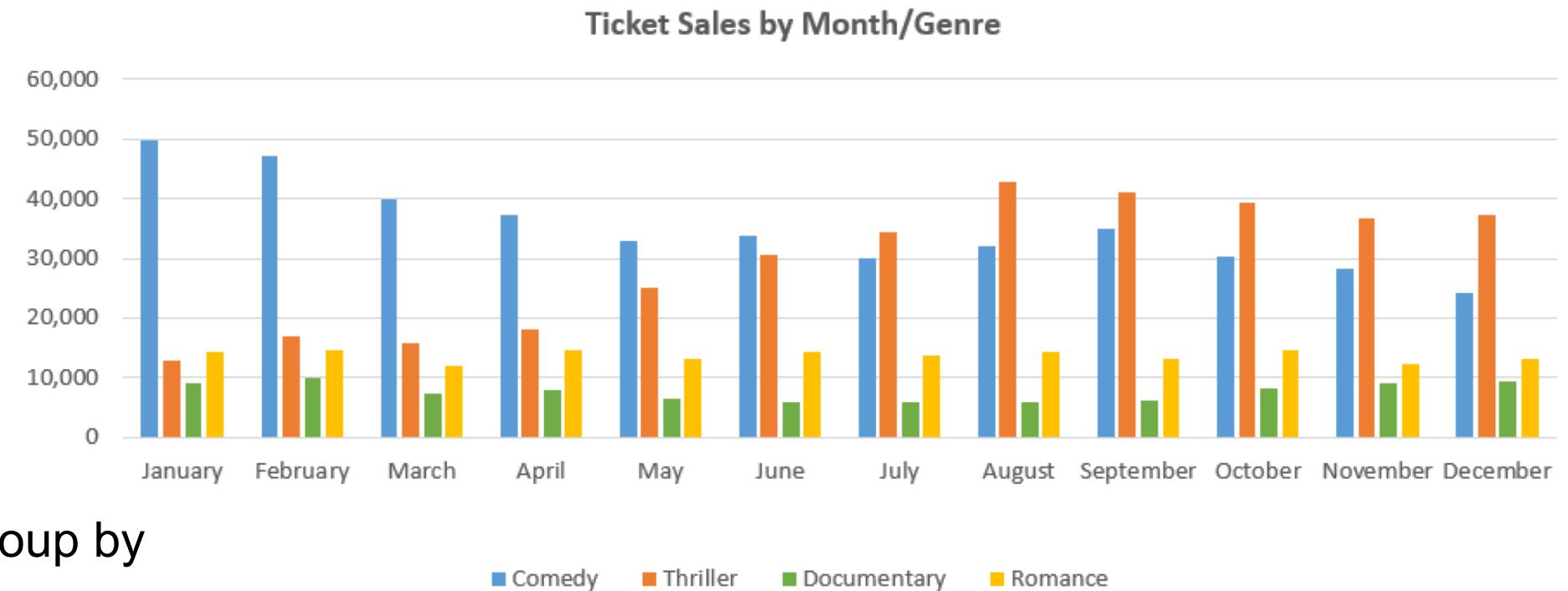
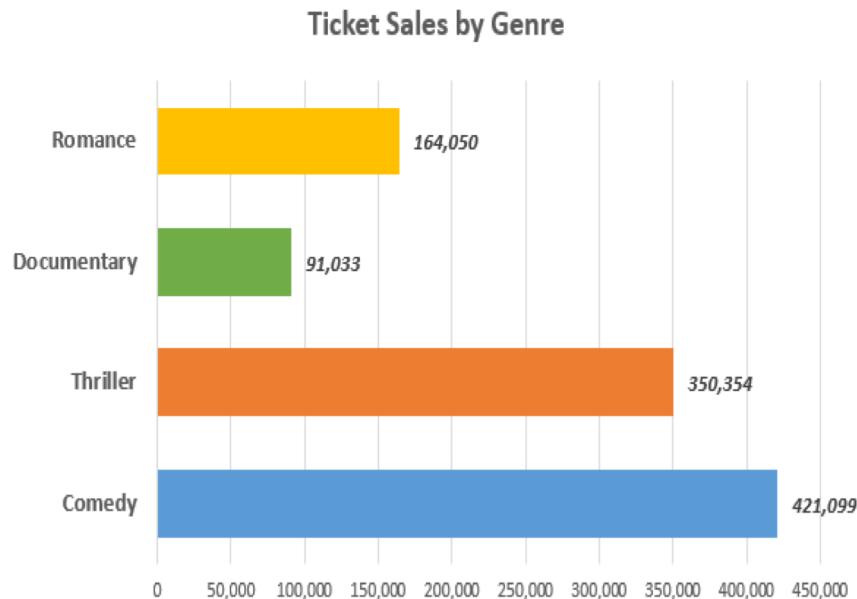
- Comparing numerical data across categories

## EXAMPLES:

- *Total sales by product type*
- *Population by country*
- *Revenue by department, by quarter*

## TIPS:

- Use **stacked** or **clustered** bars/columns to group by subcategory or compare multiple metrics



# HISTOGRAMS & PARETO CHARTS



## COMMONLY USED FOR:

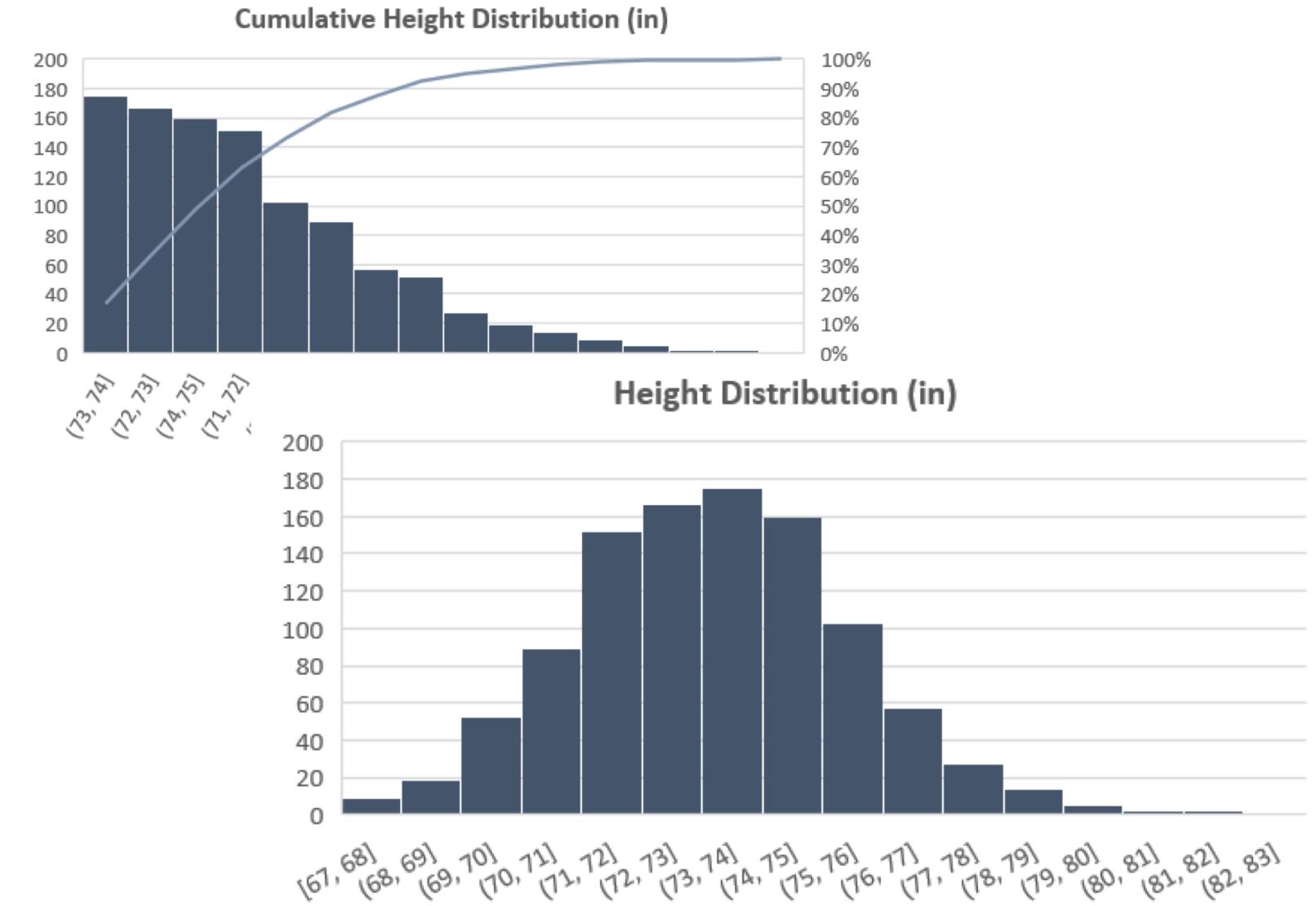
- **Showing the distribution of a continuous data set**

## EXAMPLES:

- *Frequency of test scores among students*
- *Distribution of population by age group*
- *Distribution of heights or weights*

## TIPS:

- Adjust the bin size to customize the grouping of values
- Use Pareto Charts to show the cumulative impact of each bin, ordered by significance





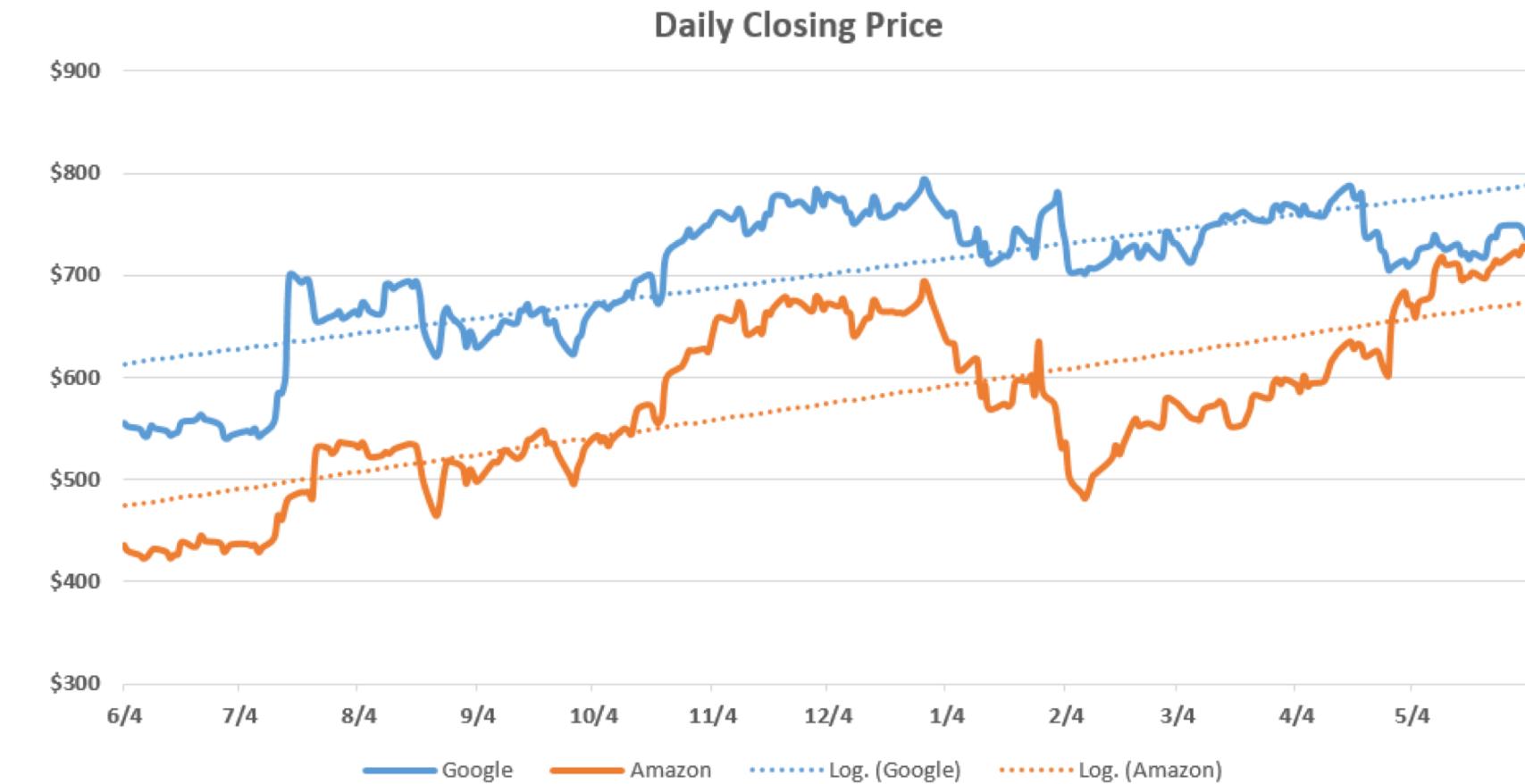
# LINE CHARTS

## COMMONLY USED FOR:

- Visualizing trends over time

## EXAMPLES:

- Stock price per hour
- Average temperature by month
- Profit by quarter



## TIPS:

- Use linear or polynomial trendlines to visualize patterns or forecast future periods



# AREA CHARTS

## COMMONLY USED FOR:

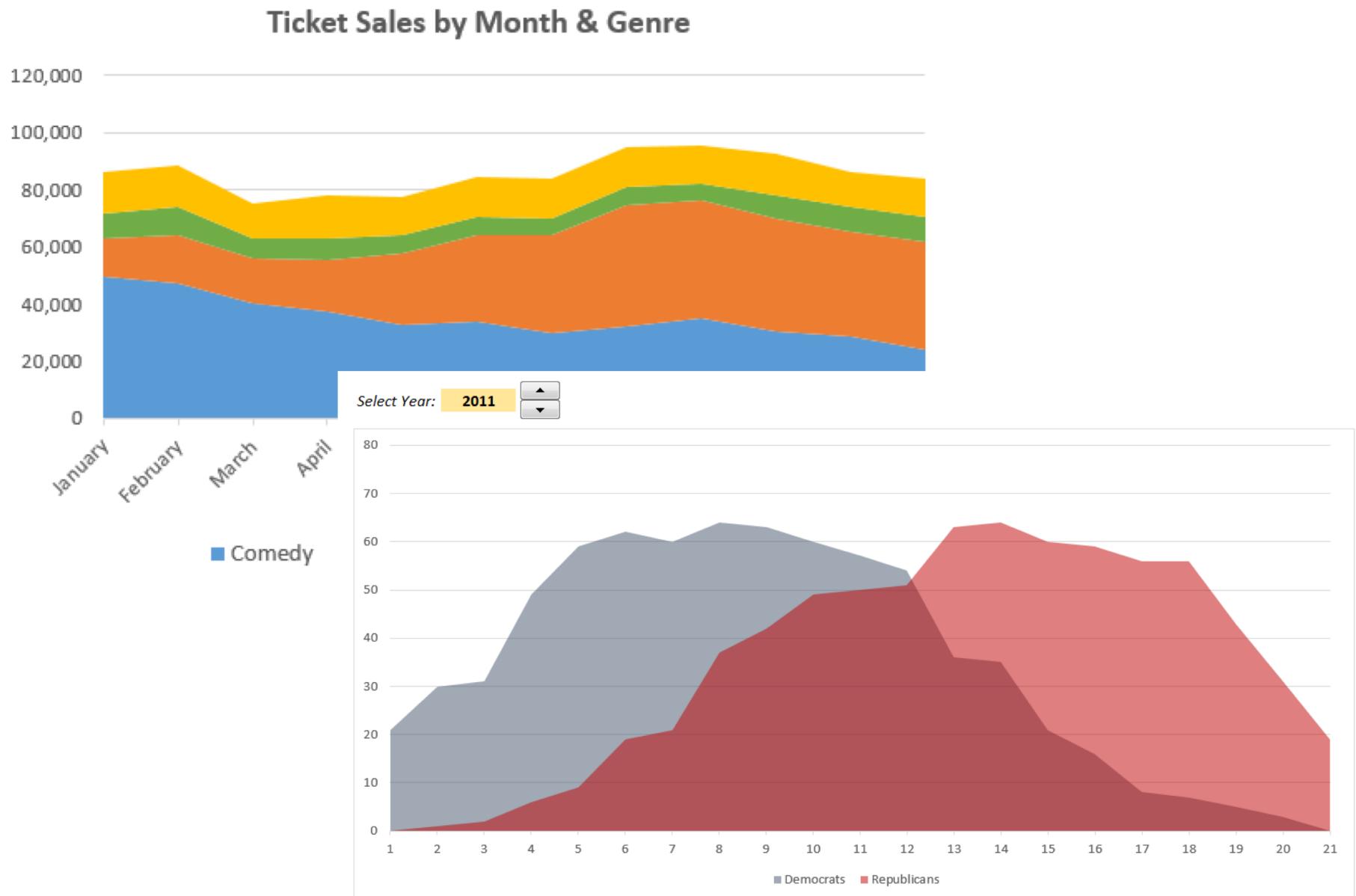
- Showing changes in data composition over time

## EXAMPLES:

- Sales by department, by month
- % of total downloads by browser, by week
- Population by continent, by decade

## TIPS:

- Keep the number of unique categories relatively low (<6) to maintain clarity





# PIE AND DONUT CHARTS

## COMMONLY USED FOR:

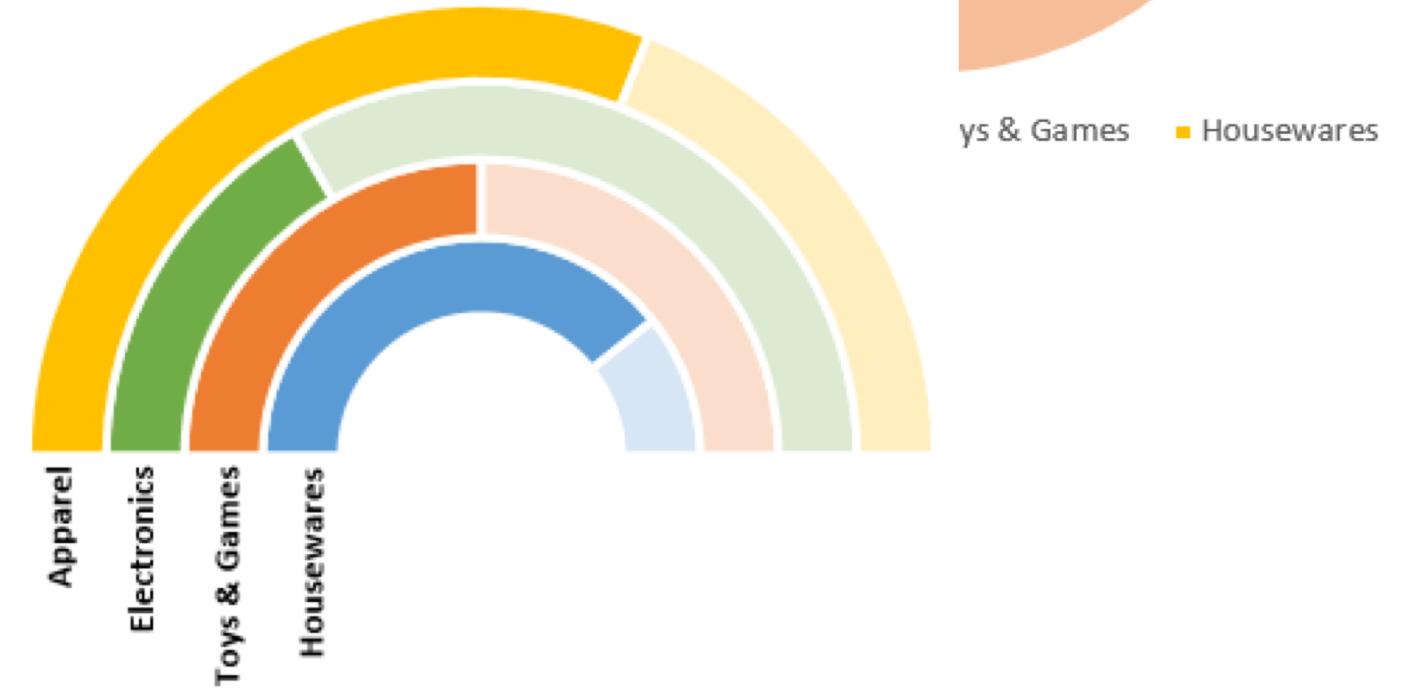
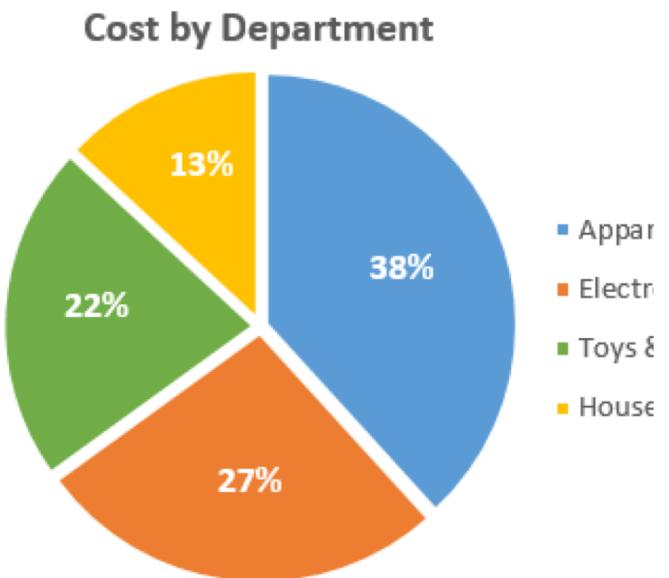
- Comparing proportions totaling 100%

## EXAMPLES:

- Percentage of budget spent by department
- Proportion of internet users by age range
- Breakdown of site traffic by source

## TIPS:

- Keep the number of slices small (<6) to maximize readability
- Use a donut chart to visualize more than one series at once, or a custom “race track” visualization



# SCATTER PLOTS



## COMMONLY USED FOR:

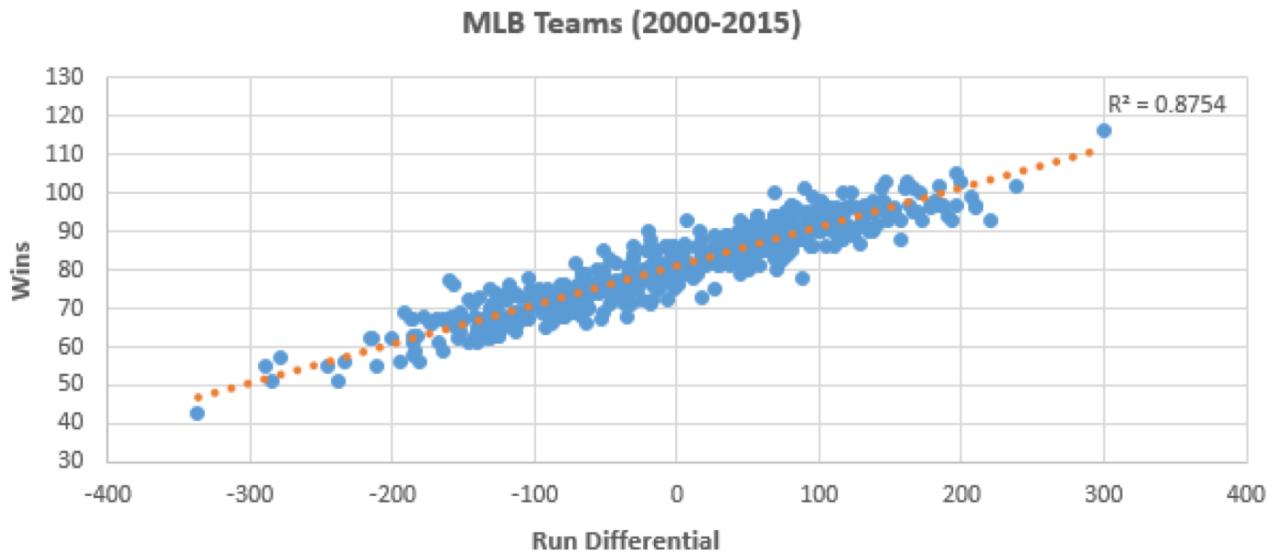
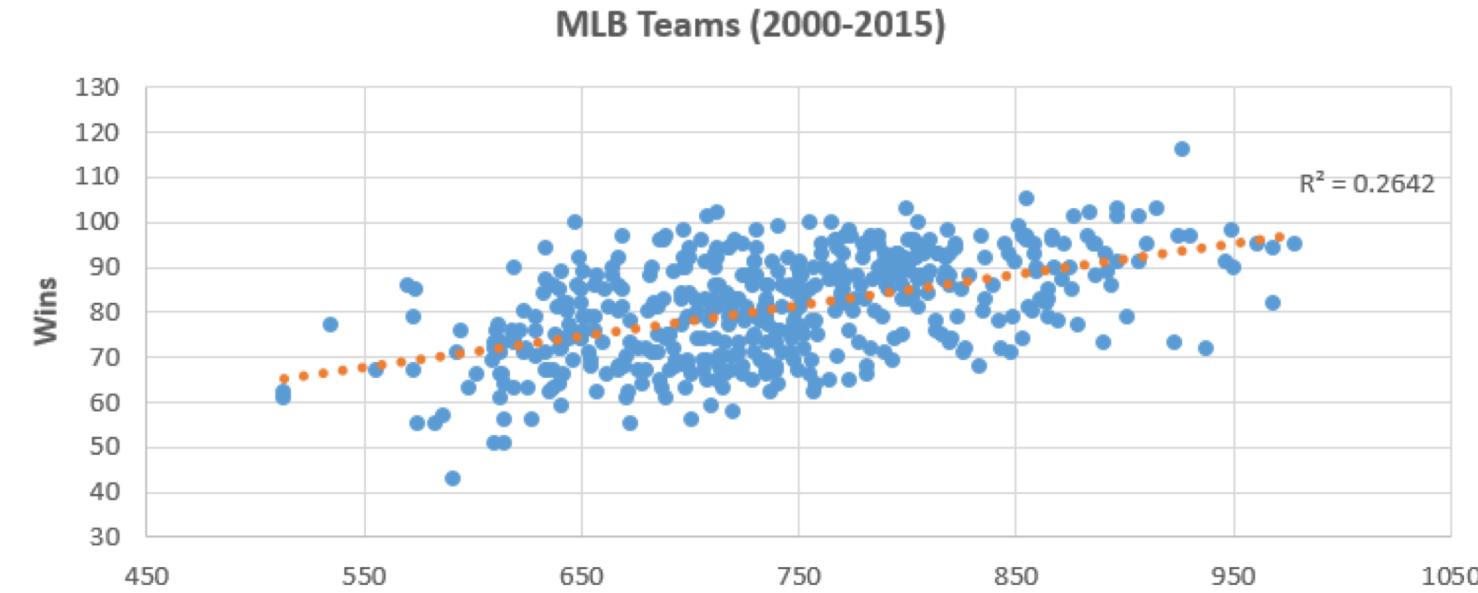
- Exploring correlations or relationships between series

## EXAMPLES:

- Number of home runs and salary by player
- Ice cream sales and average temperature by day
- Hours of television watched by age

## TIPS:

- Add a **trendline** or line of best fit to quantify the correlation between variable
- Remember that **correlation does not imply causation**





# BUBBLE CHARTS

## COMMONLY USED FOR:

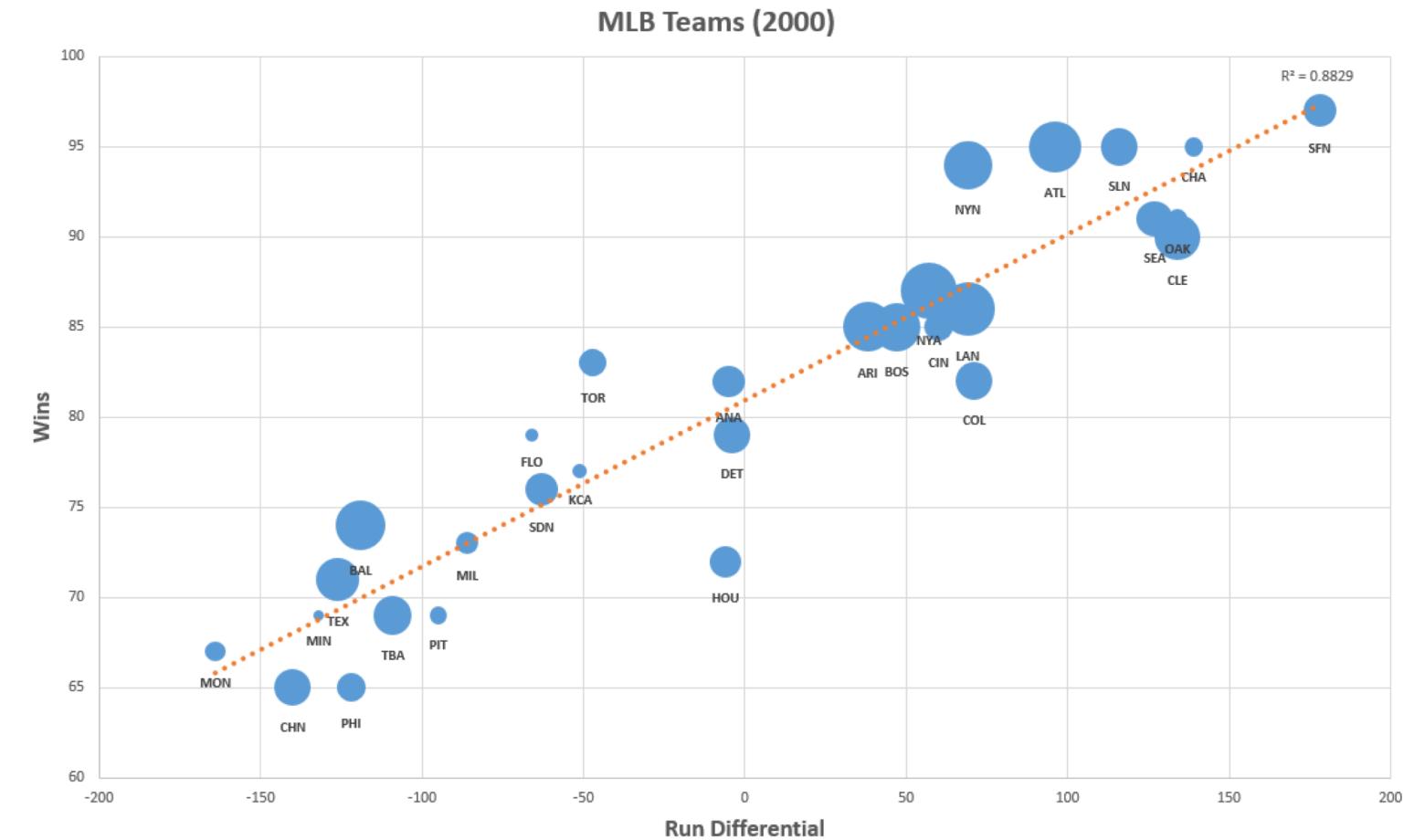
- Adding a third dimension (size) to a scatter plot format

## EXAMPLES:

- *Product sales (X), Revenue (Y), and Market Share (size) by Company*
- *Income per Capita(X), Life Expectancy (Y) and Population (size) by Country*

## TIPS:

- Use color as a fourth dimension to differentiate between categories



# BOX AND WHISKER CHARTS

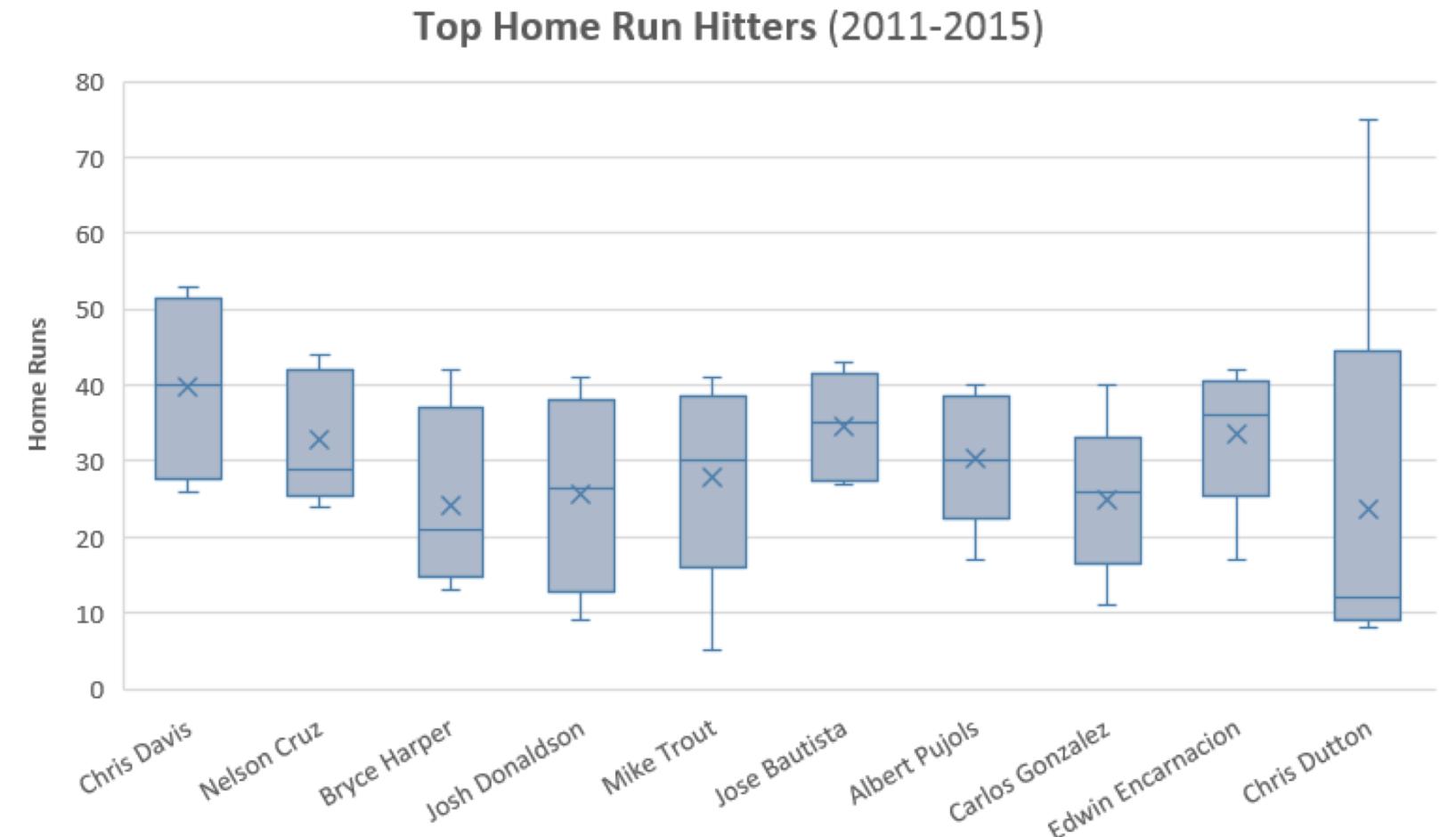


## COMMONLY USED FOR:

- **Visualizing statistical characteristics across data series**

## EXAMPLES:

- *Comparing historical annual rainfall across cities*
- *Analysing distributions of values and identifying outliers*
- *Comparing mean and median height/weight by country*



# TREE MAPS & SUNBURST CHARTS



## COMMONLY USED FOR:

- **Visualizing hierarchical data with natural groups/subgroups**

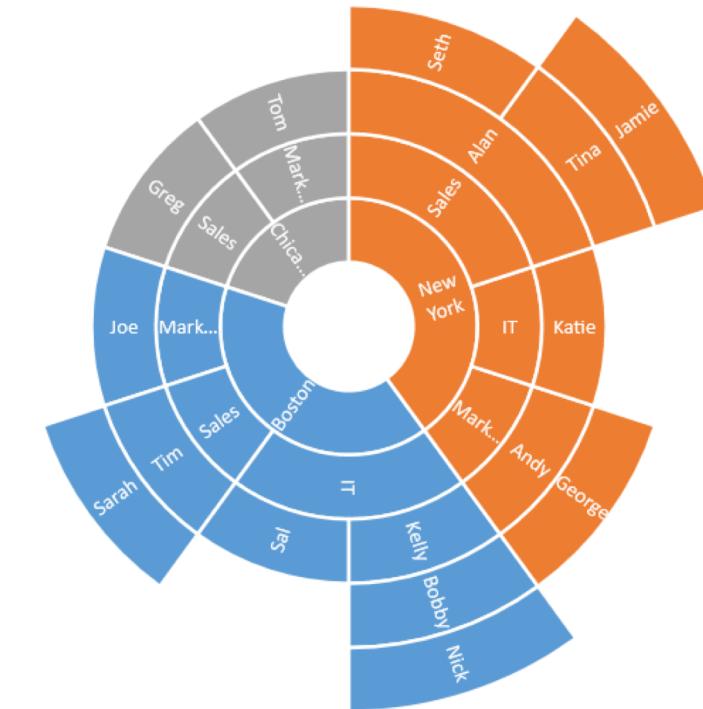
## EXAMPLES:

- *Revenue by Book Title, Sub-Genre, and Genre*
- *Number of Employees by Department and Office*
- *Population by City, State, and Region*

## TIPS:

- Use **Tree Maps** when you are only visualizing 1 or 2 hierarchical levels (i.e topic & sub-topic) or when relative sizes are important, and **Sunburst charts** to visualize the depth of multiple hierarchical levels

Comedy Drama Horror Sci-Fi





# WATERFALL CHARTS

## COMMONLY USED FOR:

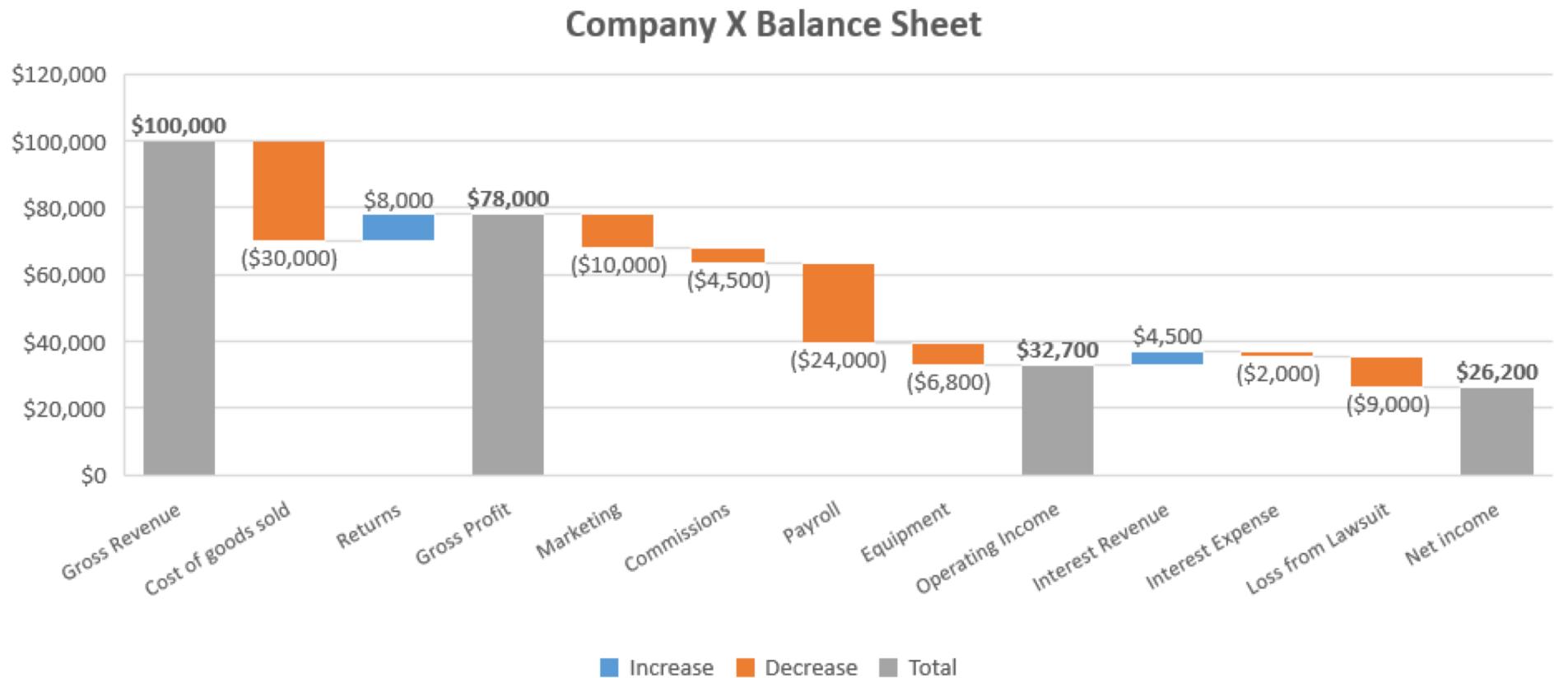
- Showing the net value after a series of positive and negative contributions

## EXAMPLES:

- Corporate balance sheet analysis
- Personal income and spending

## TIPS:

- Use **sub-totals** to create “checkpoints” and split up certain types of gains/losses (i.e. **Gross Revenue - Cost of Goods Sold = Gross Profit**, **Gross Profit – Operating Expenses = Operating Income**, etc.)





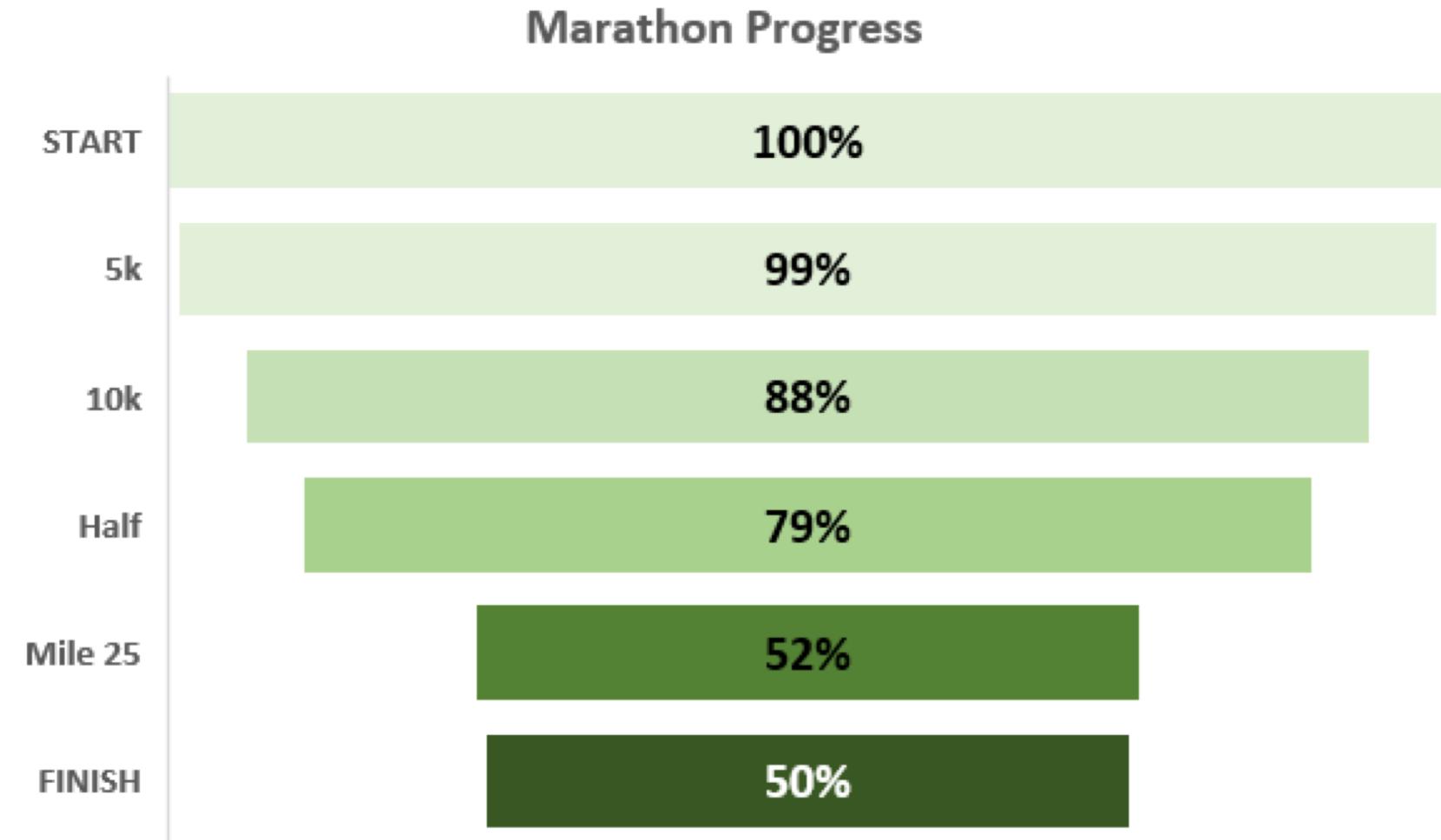
# FUNNEL CHARTS

## COMMONLY USED FOR:

- **Showing progress through the stages of a funnel**

## EXAMPLES:

- *Volume of views, clicks, and sales on an Ecom site*
- *Number of runners who reach each checkpoint in a marathon (5k, 10k, half, etc.)*



## TIPS:

- Use “**percent of total**” calculation to show the % of users (rather than #) at each funnel stage
- **Customize colors** to emphasize progression toward an end goal

# RADAR CHARTS



## COMMONLY USED FOR:

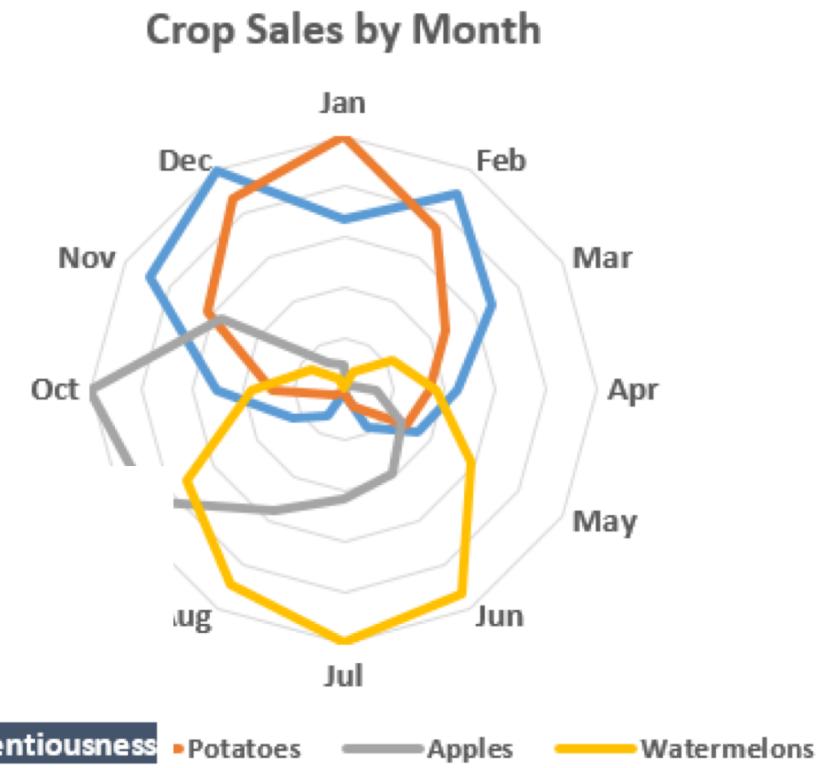
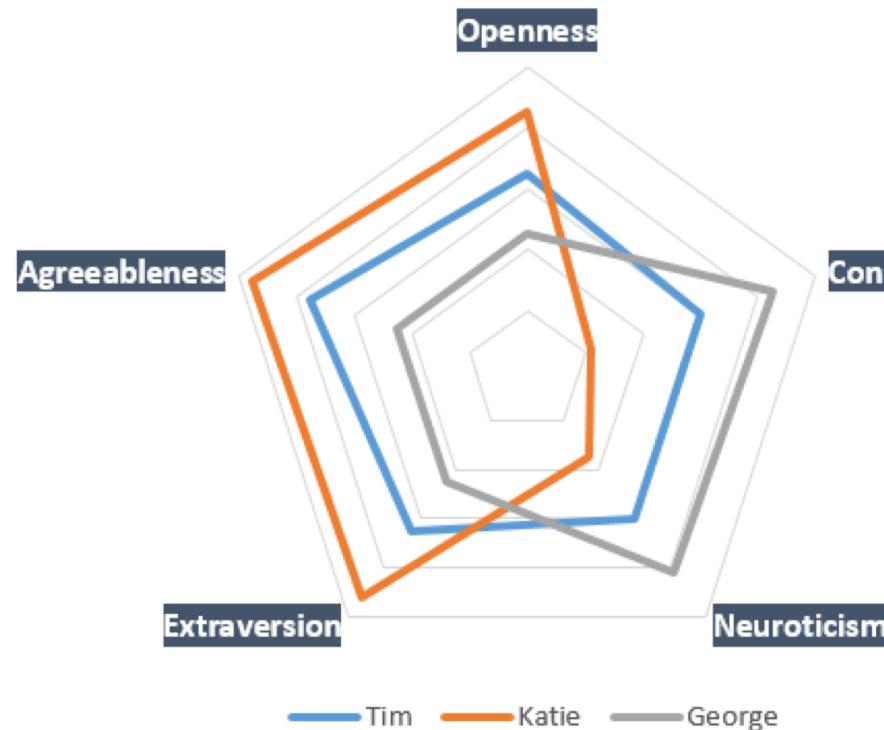
- Plotting three or more quantitative variables on a two-dimensional chart, relative to a central point

## EXAMPLES:

- Comparing test scores across multiple subjects
- Sales of different types of vegetables, by month
- Visualizing test results across subjects

## TIPS:

- Limit the number of categories or data series to minimize noise and maximize impact



# SURFACE AND CONTOUR CHARTS



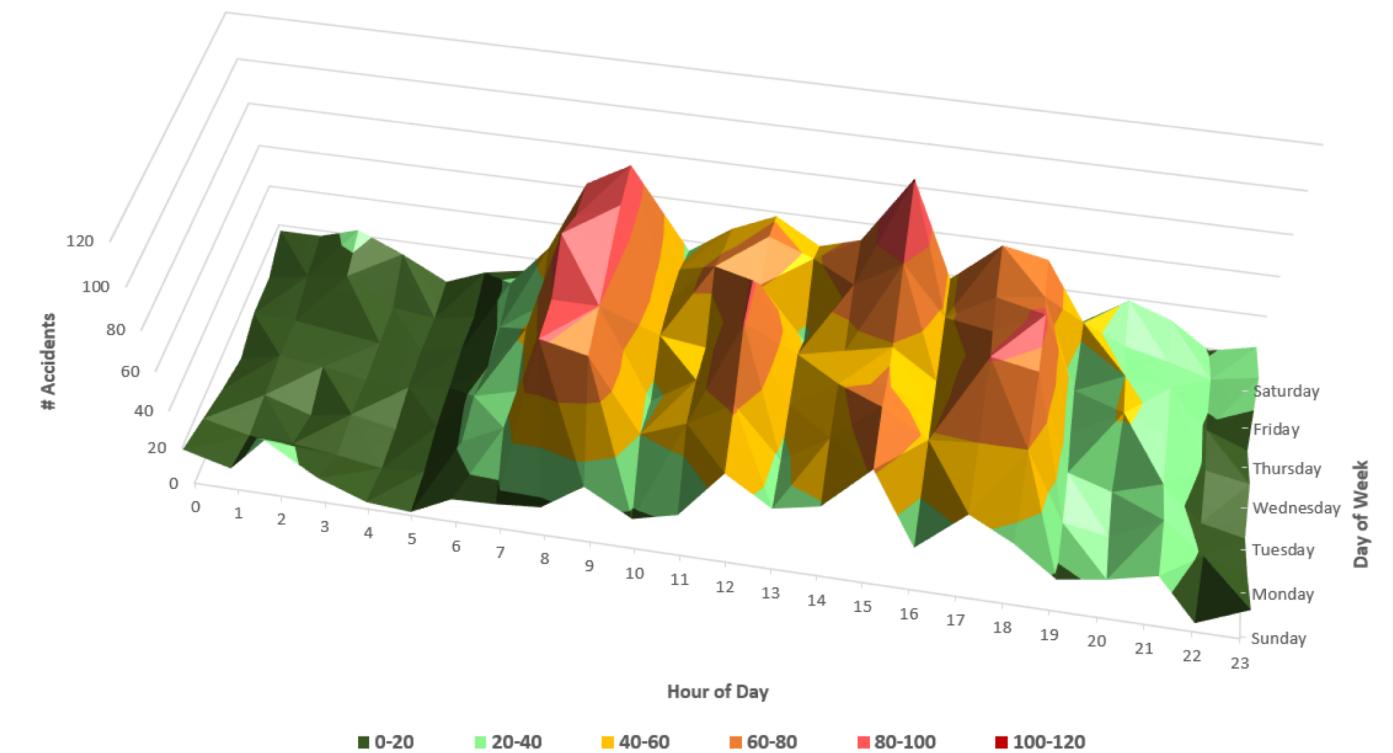
## COMMONLY USED FOR:

- Plotting data in three dimensions to find optimum combinations of values



## EXAMPLES:

- Accident rates by hour of day and day of week
- Elevation by latitude and longitude
- Cookie deliciousness by oven temp and baking time



## TIPS:

- Don't use surface charts if a simple **heat map** will tell the same story

# STOCK CHARTS



## COMMONLY USED FOR:

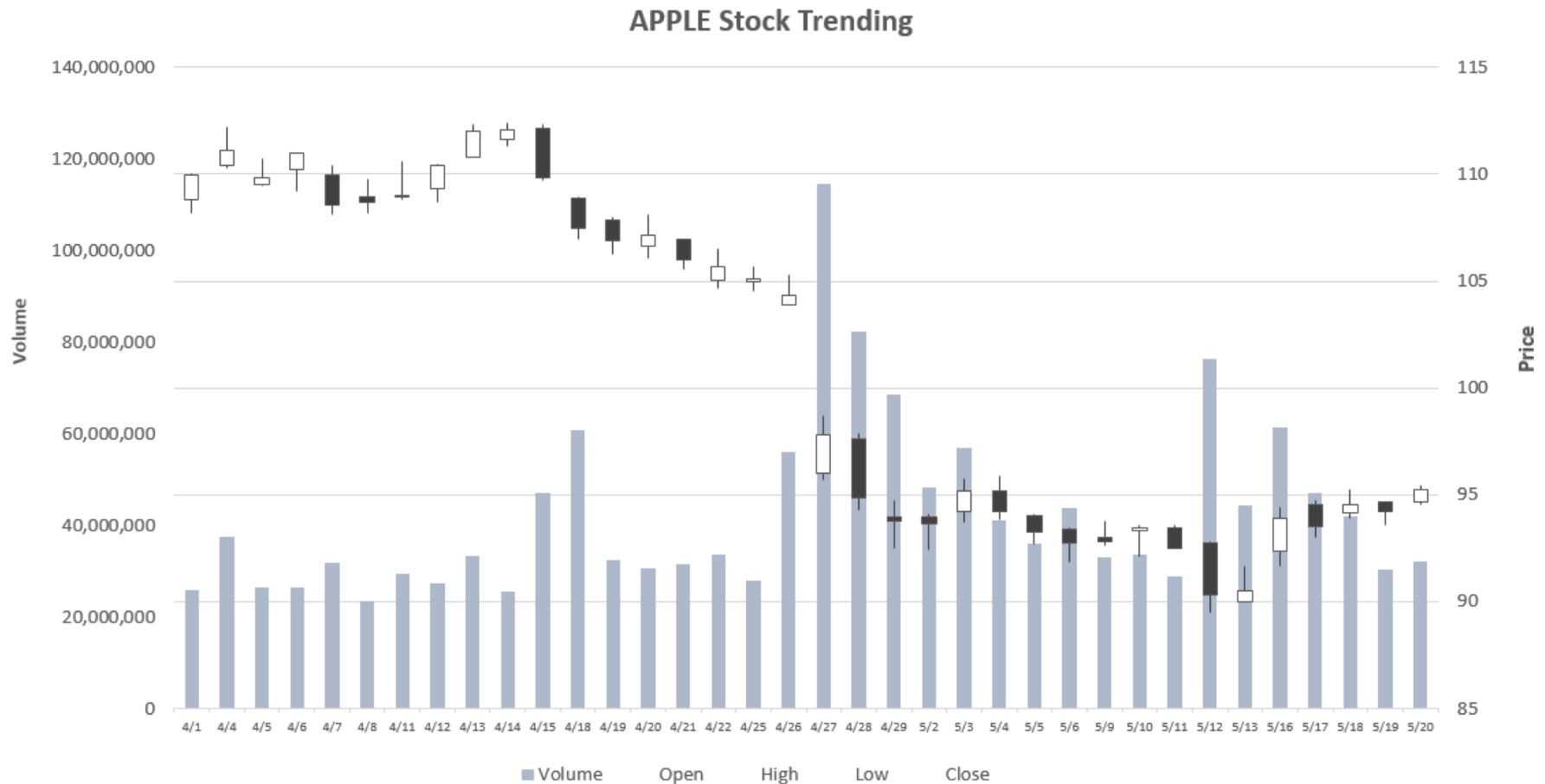
- Visualizing stock market data, including volume, high, low, open, and closing prices

## EXAMPLES:

- Facebook's daily stock performance in 2017
- High, low, and closing prices for Google in Q1
- Relative performance across multiple stocks

## TIPS:

- Manually set axis minimum/maximum values to enhance readability





# HEAT MAPS

## COMMONLY USED FOR:

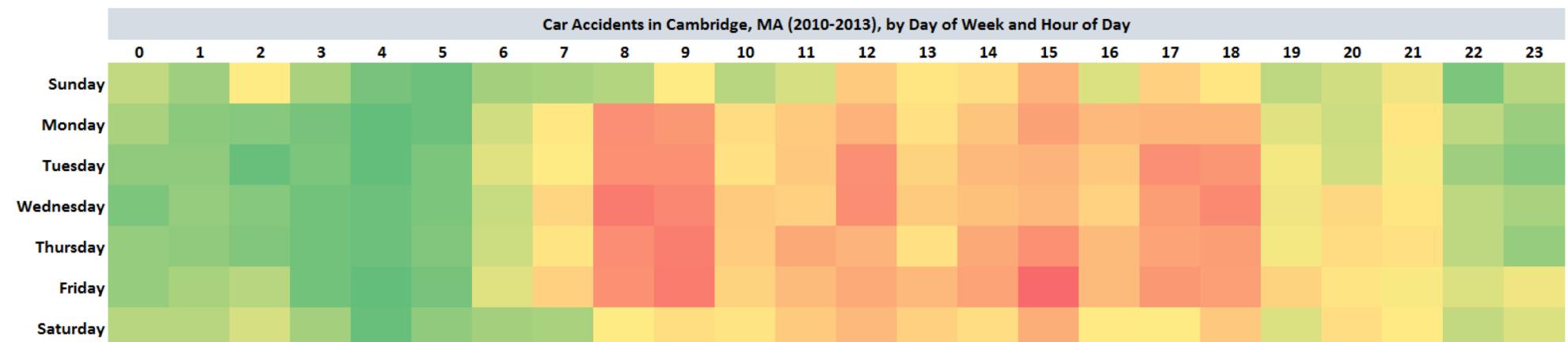
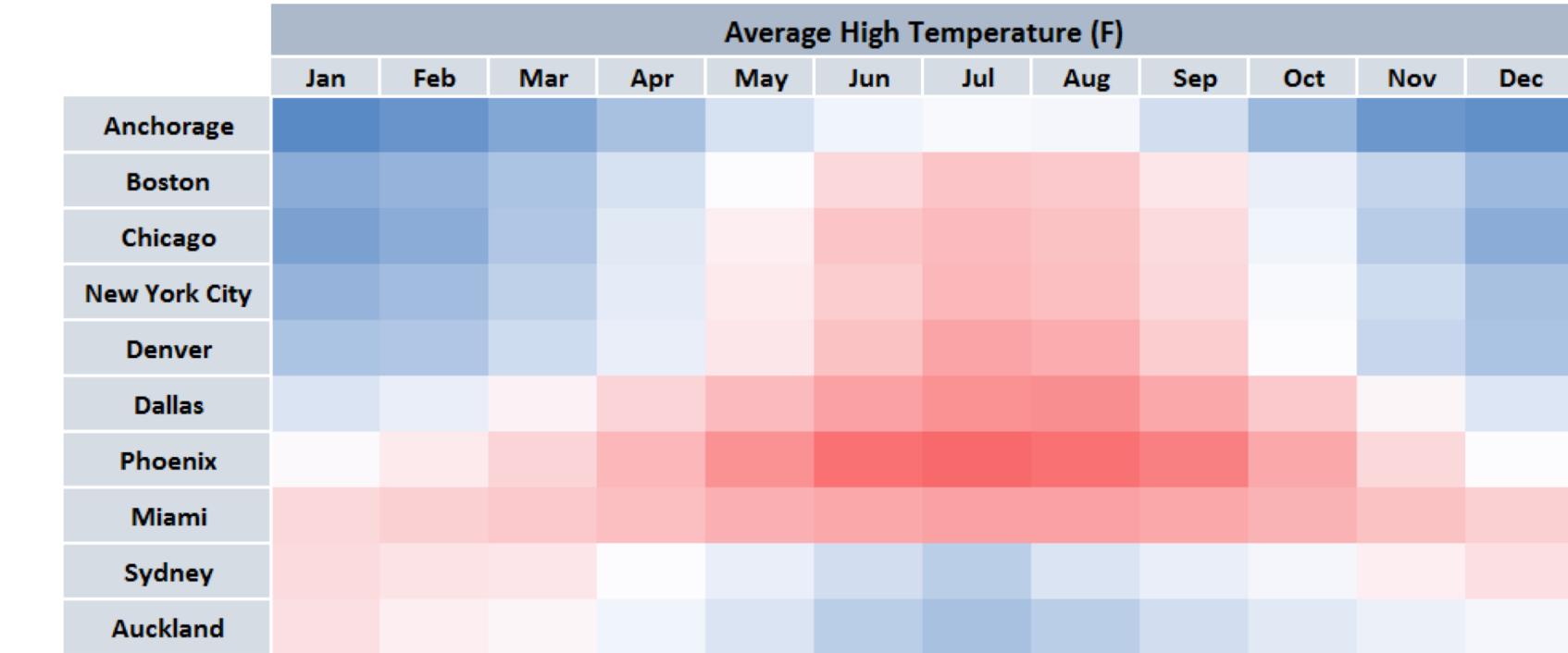
- Visualizing trends or relationships using color scales

## EXAMPLES:

- Accident rates by time of day and day of week
- Average temperature by city, by month
- Average sentiment by hashtag

## TIPS:

- Use intuitive color scales (i.e. red to green).



# GEOSPATIAL/CHOROPLETH MAP



## COMMONLY USED FOR:

- Visualizing location based data

## EXAMPLES:

- *Frequency of accidents by street address*
- *Unemployment rate by country*
- *Average rainfall by state*

