La Guardia Community College

DATA 203
DATA VISUALIZATION USING TABLEAU

Class 6

LEARNING OBJECTIVES

BEST PRACTICES FOR VISUALIZATION

1. Types of Visualizations
2. Design Principles
3. Colors that Matter
4. Visual Perception

TYPE OF VISUALIZATIONS

BEST PRACTICES FOR VISUALIZATION

1. **Truthful** – based on thorough and objective research.
2. **Functional** – accurate and allows users to act upon your information
3. **Beautiful** - well-designed and draws the user's attention through an aesthetically pleasing display of information
4. **Insightful** – provides information like trends, insights, and inferences that would be difficult to see otherwise.
5. **Enlightening** – highlights evidence, tells a story and enlightens users with a way that is easy to understand.

TYPE OF VISUALIZATIONS

BEST PRACTICES FOR VISUALIZATION

1. Clearly indicate how values relate to one another
2. Represent the values accurately.
3. Make it easy to compare values.
4. Make it easy to rank and order the values.
5. Make it obvious how people should use this information, what they should use it to accomplish, and encourage them to follow through.

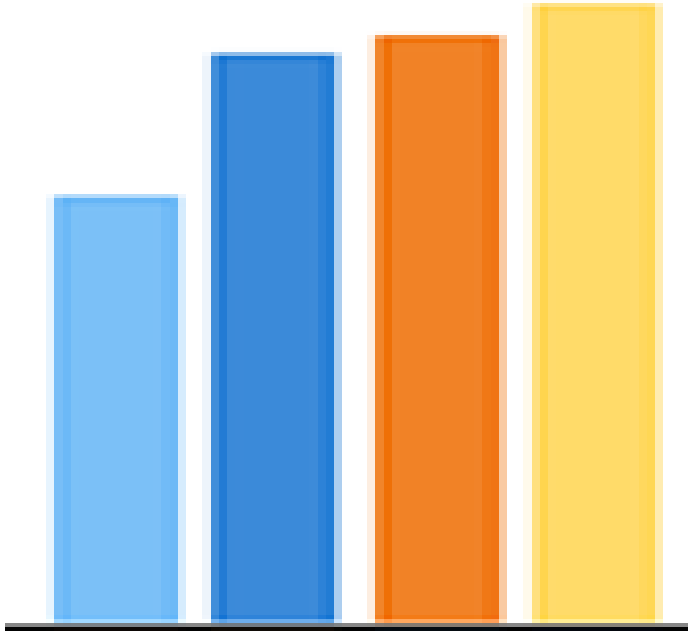
TYPE OF VISUALIZATION

- In a single day we could create as much as 2 Quintillion bytes of data.
- Know how to organize data so that it is easy to analyze and understand.

What type of visualization should you use to properly digest your data?

TYPE OF VISUALIZATION BAR CHART

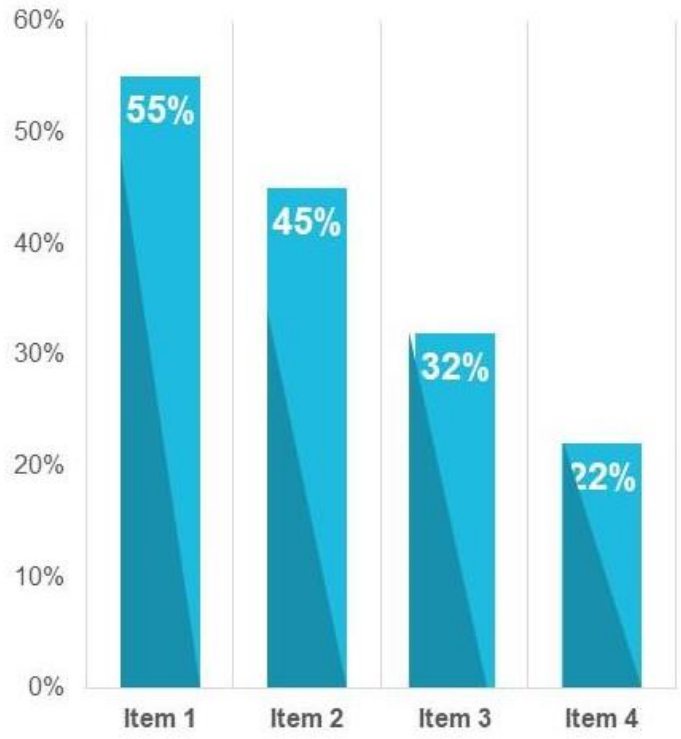
1. One of the most common chart
2. Categorical Data
3. You plot data into bars or columns that represent the number of observations for that category or series.

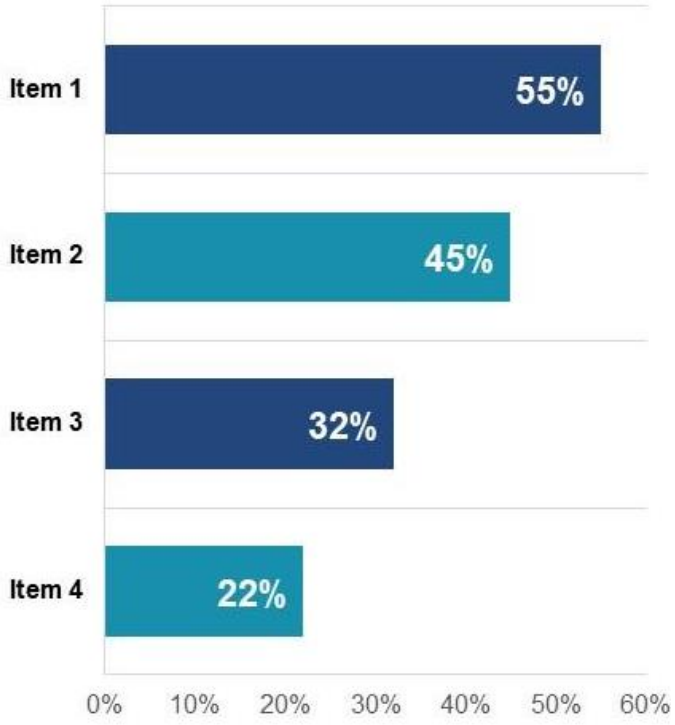


TYPE OF VISUALIZATION
BAR GRAPH

- **When to use Horizontal vs Vertical Bar Graph**

  - Readability of labels
  - Number of categories
  - Presence of negative values
  - Representation of time





TYPE OF VISUALIZATION - BAR GRAPH

Gaps

Makes sure that your graph does not look like a histogram.

Gap size should be ¼-½ the size of the bar itself.

Bar Length/Width

A minimum ratio of 10:1 is a good thumb of rule to follow.

Baseline

Always start your bar graph at 0

If you use truncated bar graph, make sure to call it out.

TYPE OF VISUALIZATION - BAR GRAPH

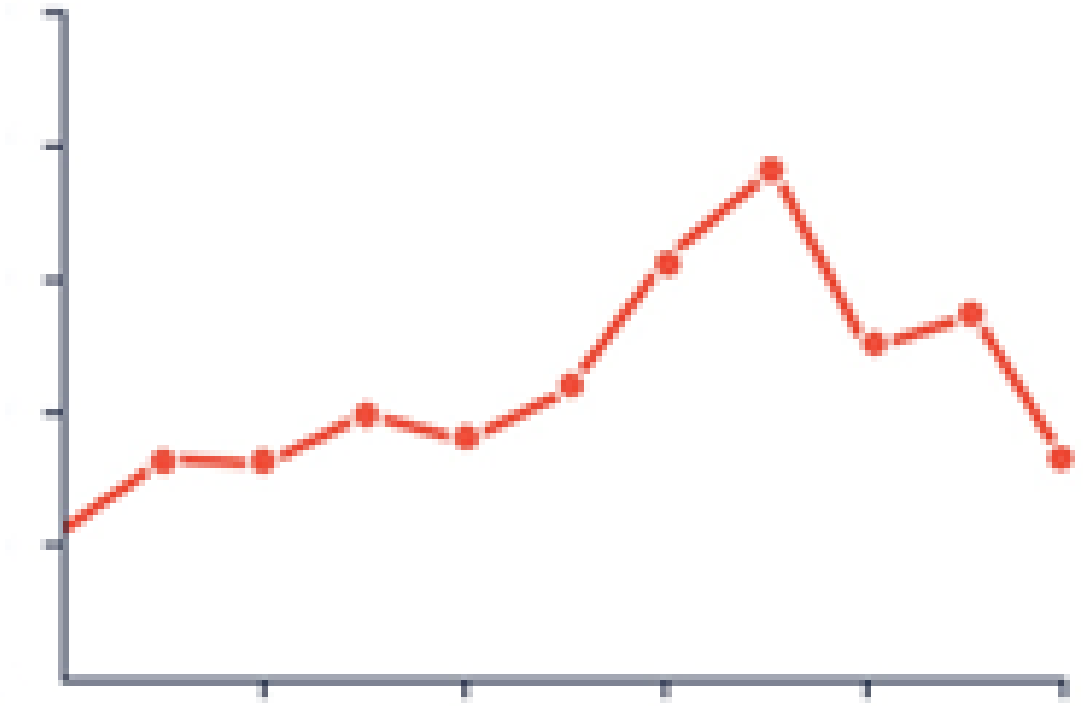
- **Labels**
  - Use Horizontal labeling.
  - If one label clashes with another, use abbreviations or acronyms, or a legend.
- **Trends**
  - Pre-sort from largest to smallest or vice – versa
  - Trends over time always use chronological order of your category.
- **Annotations**
  - Place them in the middle or at the end of the bars.

TYPE OF VISUALIZATION - BAR GRAPH

- **Colors**
  - Use color in moderation, especially if you have 5-7 categories
  - Use color with purpose
  - Sometimes there is not a need to color each category differently.
  - If you do not know what colors to use, use your organization's color.

TYPE OF VISUALIZATION - LINE GRAPH

Graph that shows the change of data over a continuous time interval or time span.



TYPE OF VISUALIZATION - LINE GRAPH

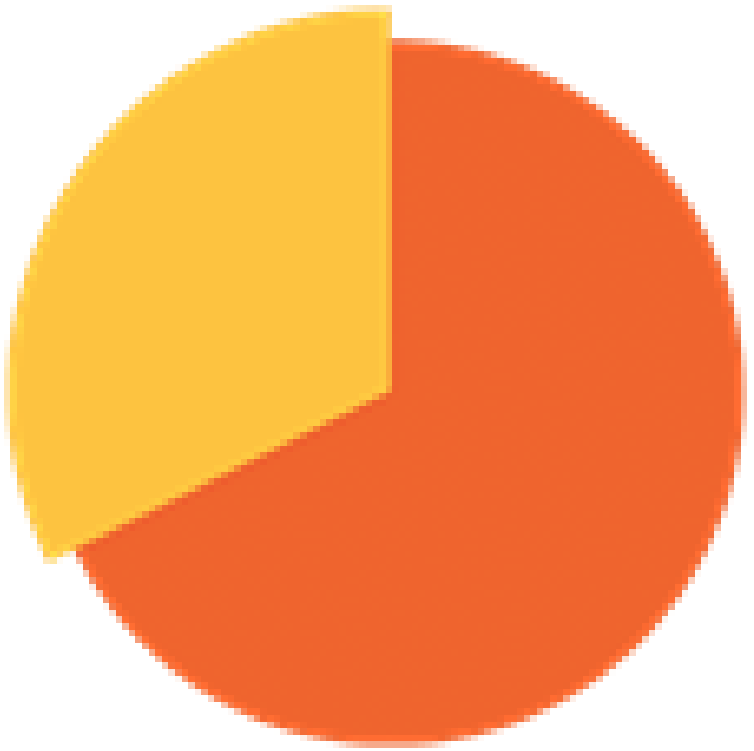
- **Colors**
  - Use color in moderation, especially if you have 5-7 categories
  - Use color with purpose
  - Sometimes there is not a need to color each category differently.
  - If you do not know what colors to use, use your organization's color.

TYPE OF VISUALIZATION - LINE GRAPH

- **Number of lines**
  - Five or fewer lines only
  - If lines can be well separated, then you can use more.
- **Legend**
  - Provide a legend if you using multiple data series.
  - Place the legend at the same height as the data series.
  - Label the lines individually
  - The color of the label should be the same as the color of the line.

TYPE OF VISUALIZATION - PIE CHART

- Show part to whole relationship in your data.
- The entire circle is a whole and each slice is a relevant section or the part.
- The entire pie is equal to 100%



TYPE OF VISUALIZATION - PIE CHART

How to Use

Size of Pie Slice - by angle, area or arc length

When To Use

- Show Part- to- Whole Relationship
- Show rough proportions of what the whole is split into.
- Show that one portion of the total is relatively large/small.

When NOT To Use

- Your data has negative values
- Have more than one data series
- Have to show the zero value
- Have more than five categories
- Compare parts to each other
- Parts do not make a meaningful whole
- Parts are not mutually exclusive

TYPE OF VISUALIZATION - PIE CHART

- **How to Use**
  Size of Pie Slice - by angle, area or arc length

- **When To Use**
  - Show Part- to- Whole Relationship
  - Show rough proportions of what the whole is split into.
  - Show that one portion of the total is relatively large/small.

When NOT To Use

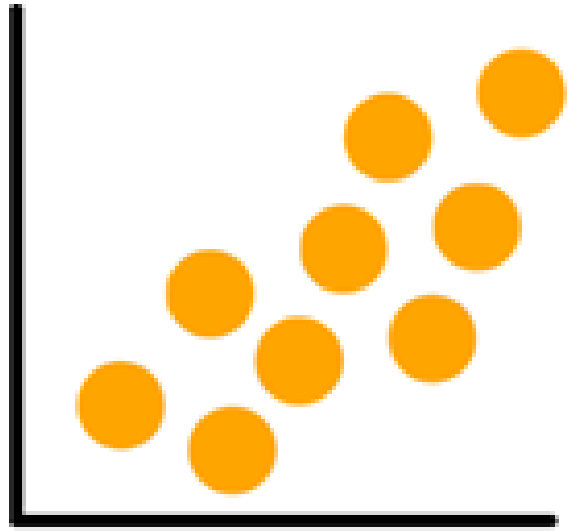
- Your data has negative values
- Have more than one data series
- Have to show the zero value
- Have more than five categories
- Compare parts to each other
- Parts do not make a meaningful whole
- Parts are not mutually exclusive

TYPE OF VISUALIZATION - PIE CHART

- **Color**
  - Pick a color and use its different shades for the pie slices, sans the slice you want to emphasize
  - Stay away from using rainbow colors
- **Labels**
  - Label smaller pie slices outside of the chart
  - You can include both the absolute values or their relative proportions

TYPE OF VISUALIZATION - SCATTER PLOT

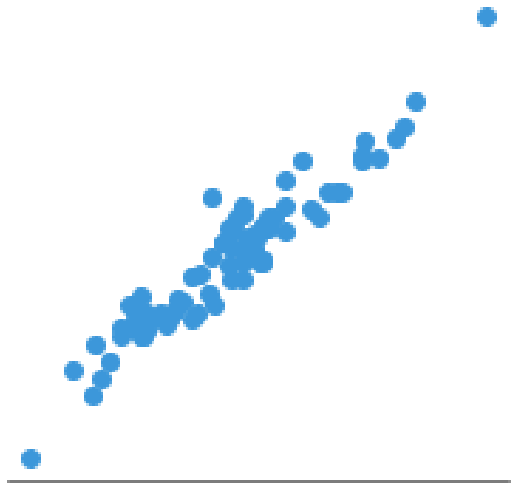
- Shows two variables (X,Y) in the form of points on a rectangular coordinate system.
- The position of the point is at the intersection of its two values.
- Displays the strength, the direction, and form of relationship between the variables.
- It measure correlation between two variables.



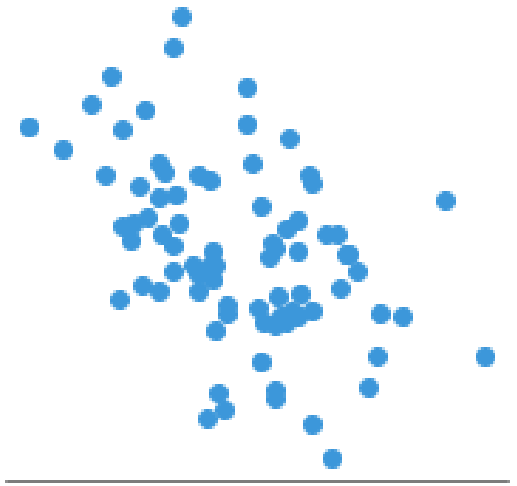
DATA VISUALIZATION - SCATTER PLOT

- **Correlation Types**
  - FORM - Linear vs. non-linear
  - DIRECTION - Positive vs. Negative
  - STRENGTH - Strong, moderate, weak

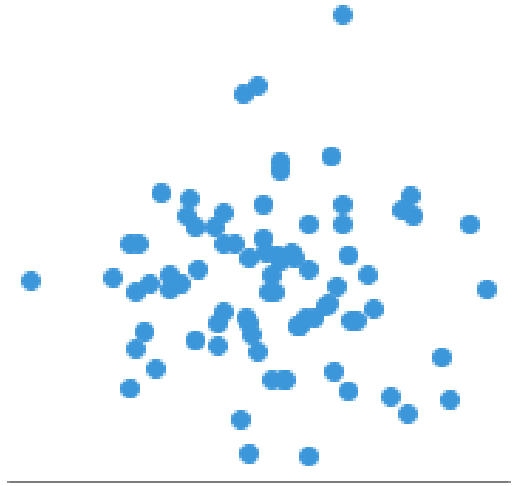
DATA VISUALIZATION - SCATTER PLOT



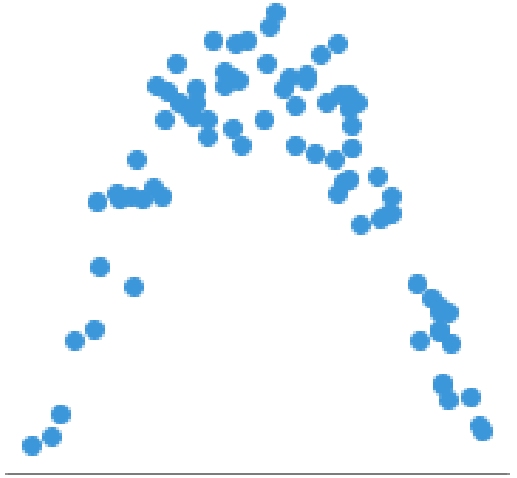
strong, positive, linear



moderate, negative, linear



null / no relationship



strong, non-linear

DATA VISUALIZATION - SCATTER PLOT

- **Trend Line**
  - Use it to help visualize the strength of the relationship between the variables.
  - Do not have more than two trend lines.
  - Helps identify outliers faster
- **Plot**
  - Use hollowed out or semi-transparent plots
  - Use different colors and shapes if you have more than one category/ group

TYPE OF VISUALIZATION - SCATTER PLOT

- **Avoid Over Plotting**
  - Reducing the dot size
  - Using transparency
  - Using only a sample of your data
- **Grouping**
  - Use different colors
  - Use symbols

CHOOSING THE RIGHT VISUAL

- There is **no golden rule** when it comes to choosing the right visual

- Choosing the right visual is **an art and a science**.

- Use the **type of data** you have to help guide you
  - Quantitative vs Qualitative
  - Discrete vs. continuous
  - Measurement scales: nominal, ordinal, interval, ratio

- Know who **your audience** is
  - Experience
  - Familiarity
  - Comfort level
  - Openness

CHOOSING THE RIGHT VISUAL

- What is our intention and what are we trying to show?
  - Comparison - compare actual values
  - Time-related - see change in value over time
  - Correlation - relationship between two variables.
  - Composition - part to whole or some type of hierarchy
  - Distribution - highlight the distribution of values
  - Spatial - quantitative values in spatial regions such as geographical maps.

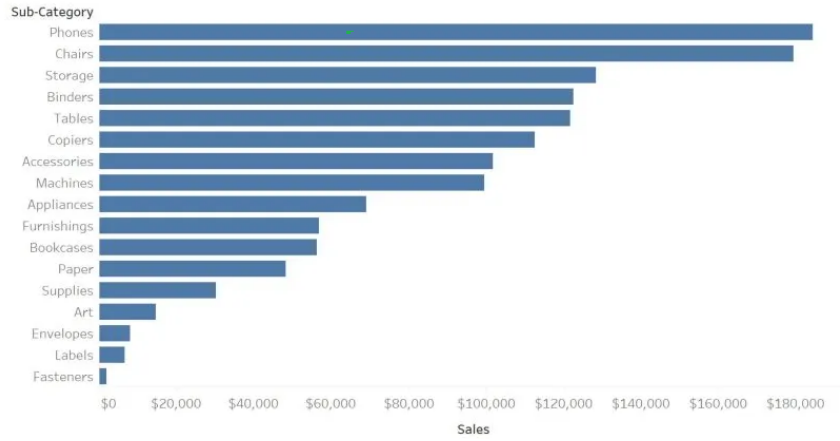
CHOOSING THE RIGHT VISUAL COMPARISON

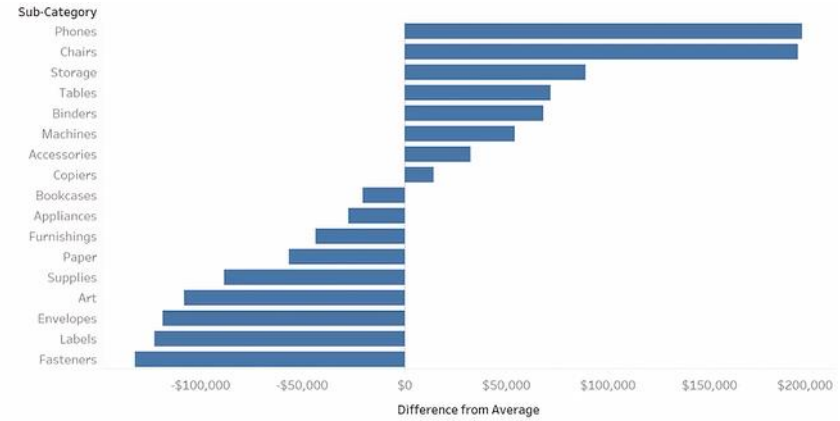
BAR CHART

- Commonly used for comparisons.
  - The length of the bar lets you compare quantitative numbers i.e. shorter vs longer bars.
- Rank subcategory by its value.

CHOOSING THE RIGHT VISUAL BAR CHART

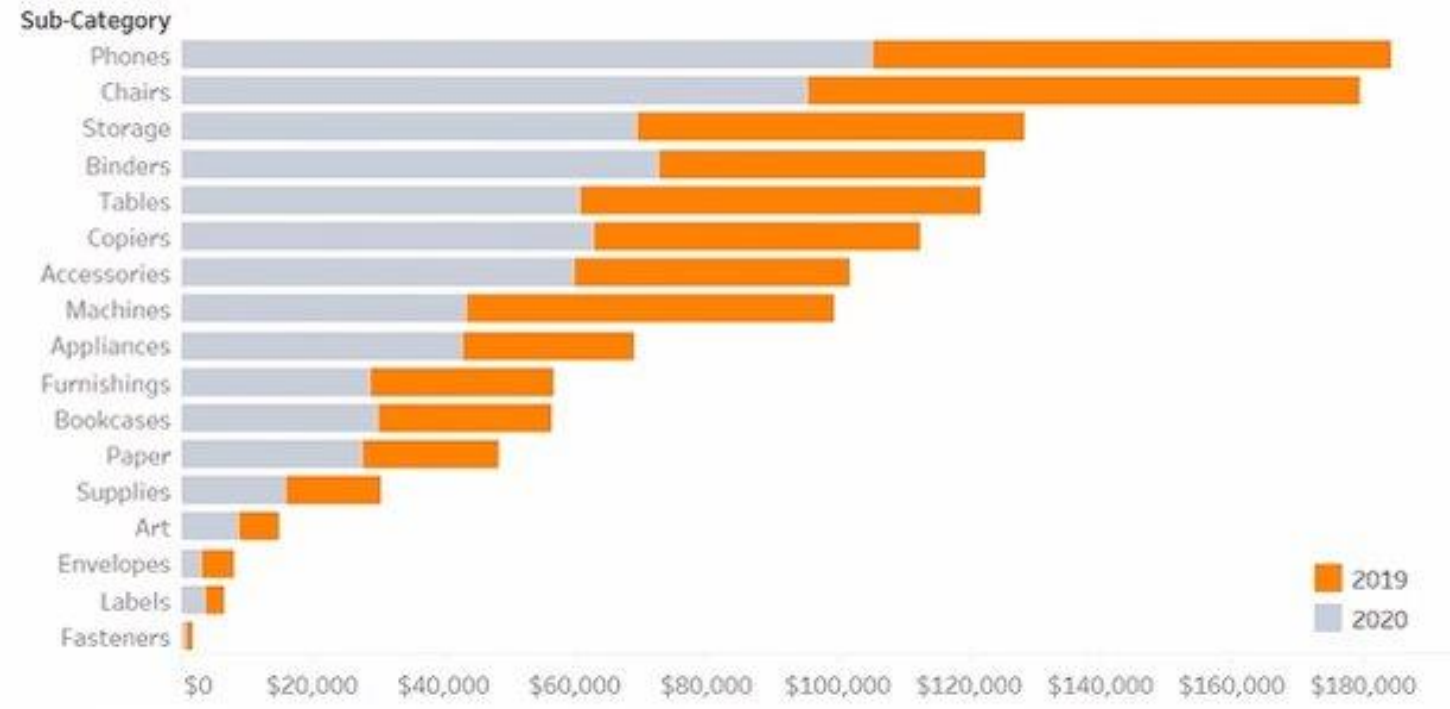
- Show differences in actual values
- Show magnitude of differences from a certain reference point





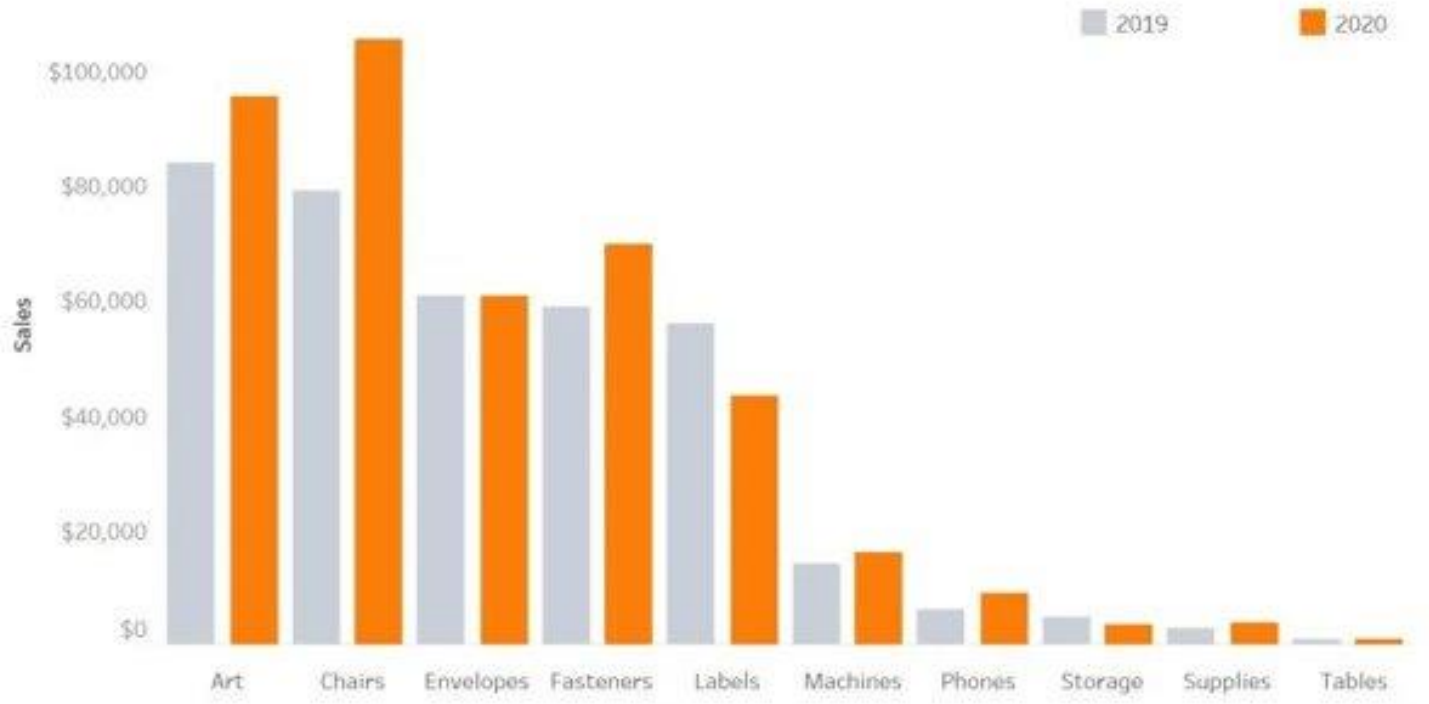
CHOOSING THE RIGHT VISUAL - BAR CHART

- Compare cumulative totals and what composes those totals
- You cannot easily compare the numbers of the individual parts



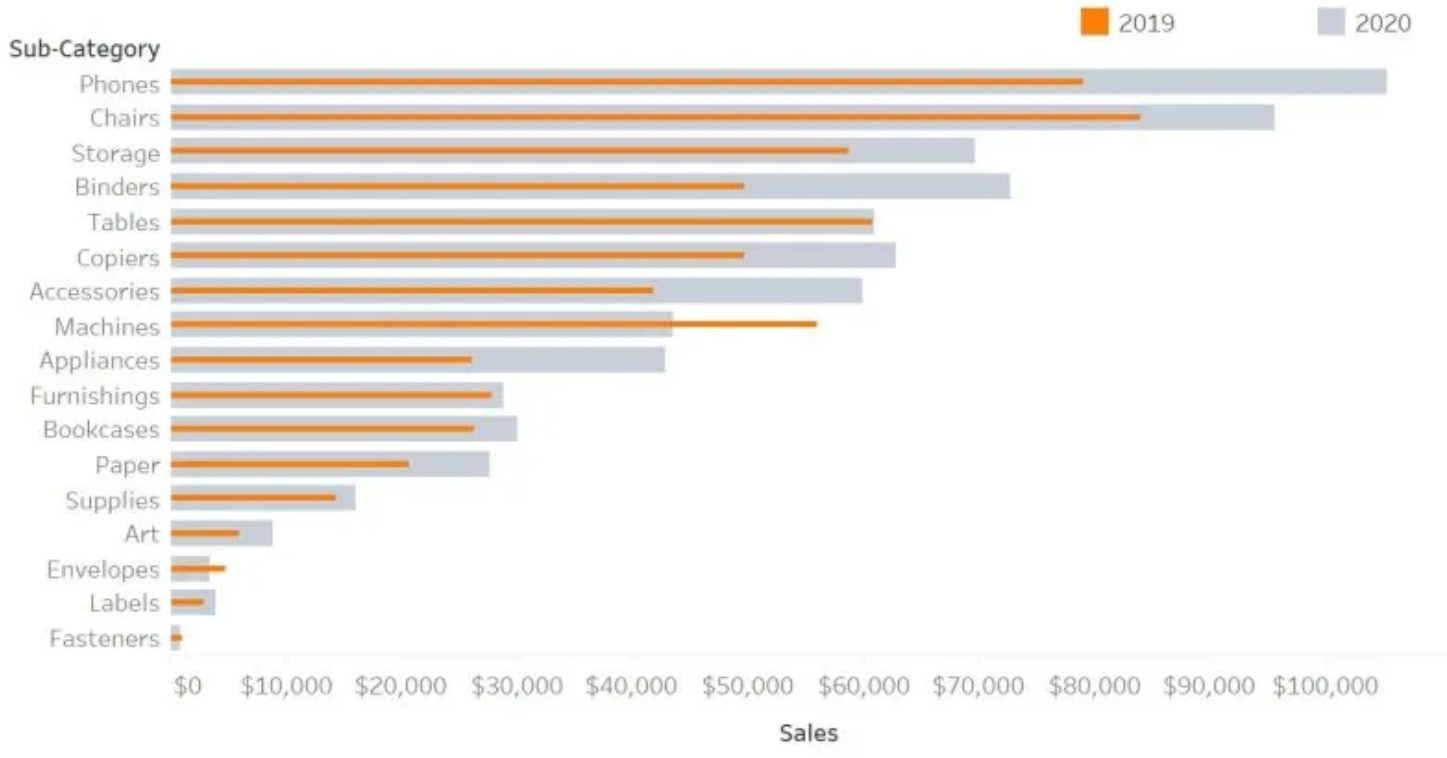
CHOOSING THE RIGHT VISUAL - BAR CHART

- Compare individual dimensions side by side
- Cluttered if you have too many dimensions



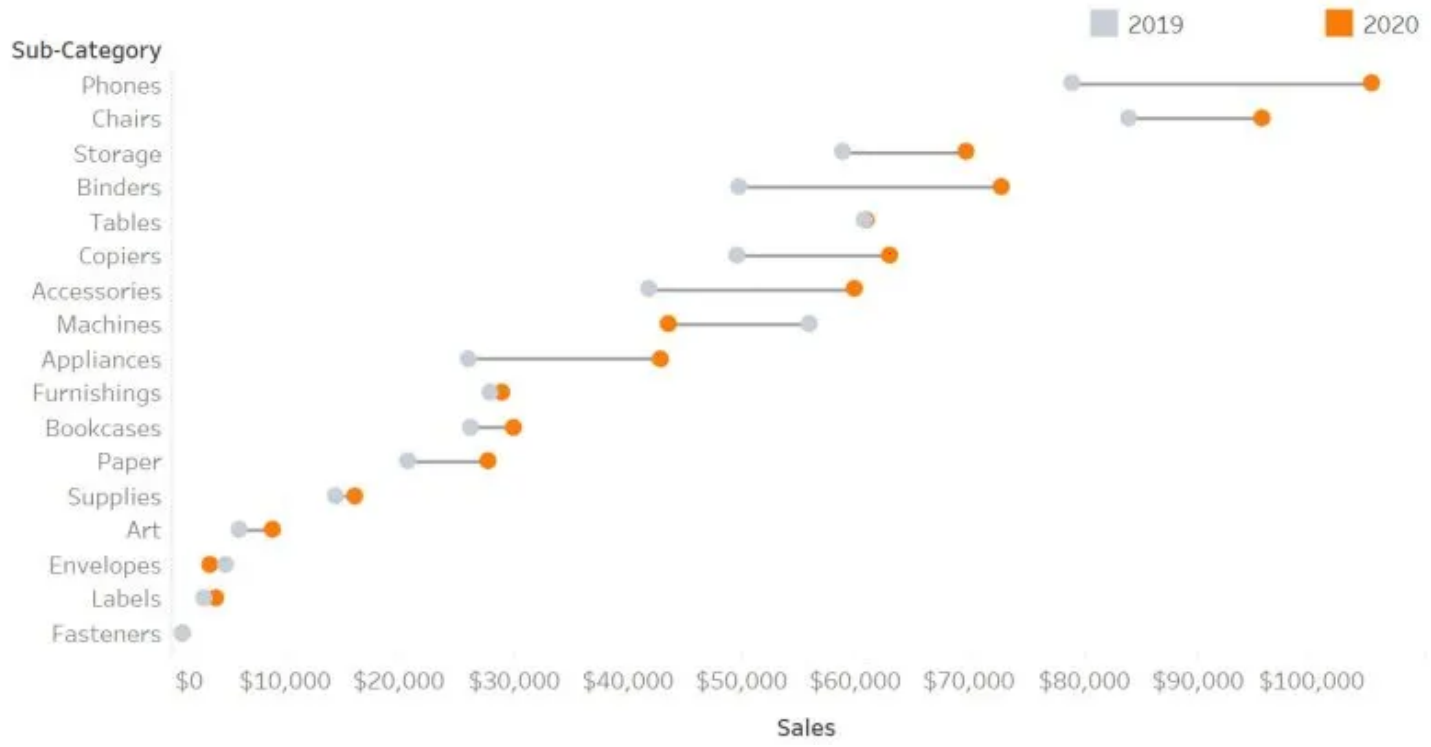
CHOOSING THE RIGHT VISUAL - BAR CHART

- Compare two quantitative values by varying the width and the color.
- Takes less space than a Clustered Bar Chart
- Width of the bar is not related to the value



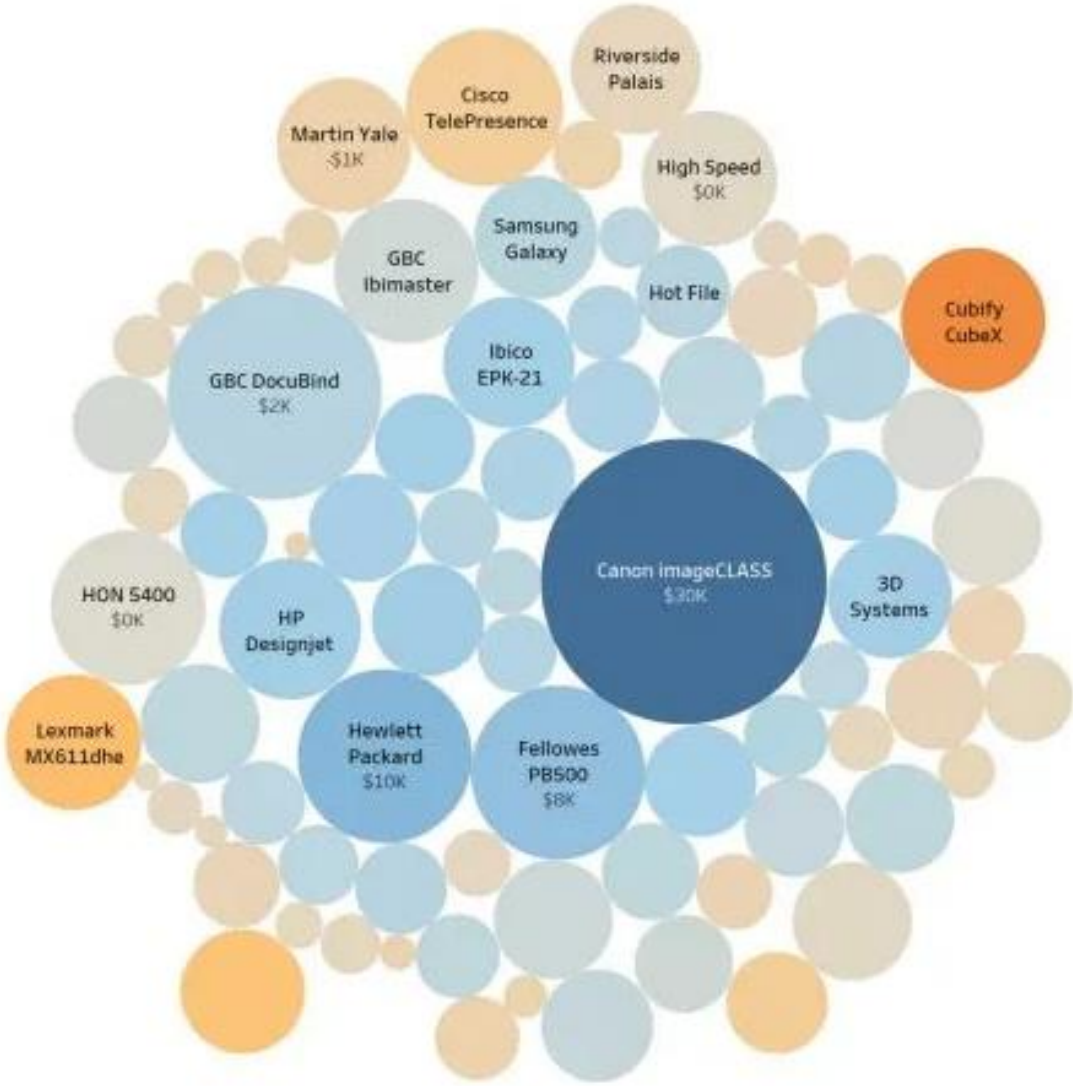
CHOOSING THE RIGHT VISUAL- BAR CHART

- Measure quantitative differences
- The length of the line is the difference between the two quantitative values.
- Does not work if the differences is too small.



CHOOSING THE RIGHT VISUA - BUBBLE CHART

- Highlight minimum and maximums
- A good chart to grab your audience's attention



CHOOSING THE RIGHT VISUAL TIME-RELATED

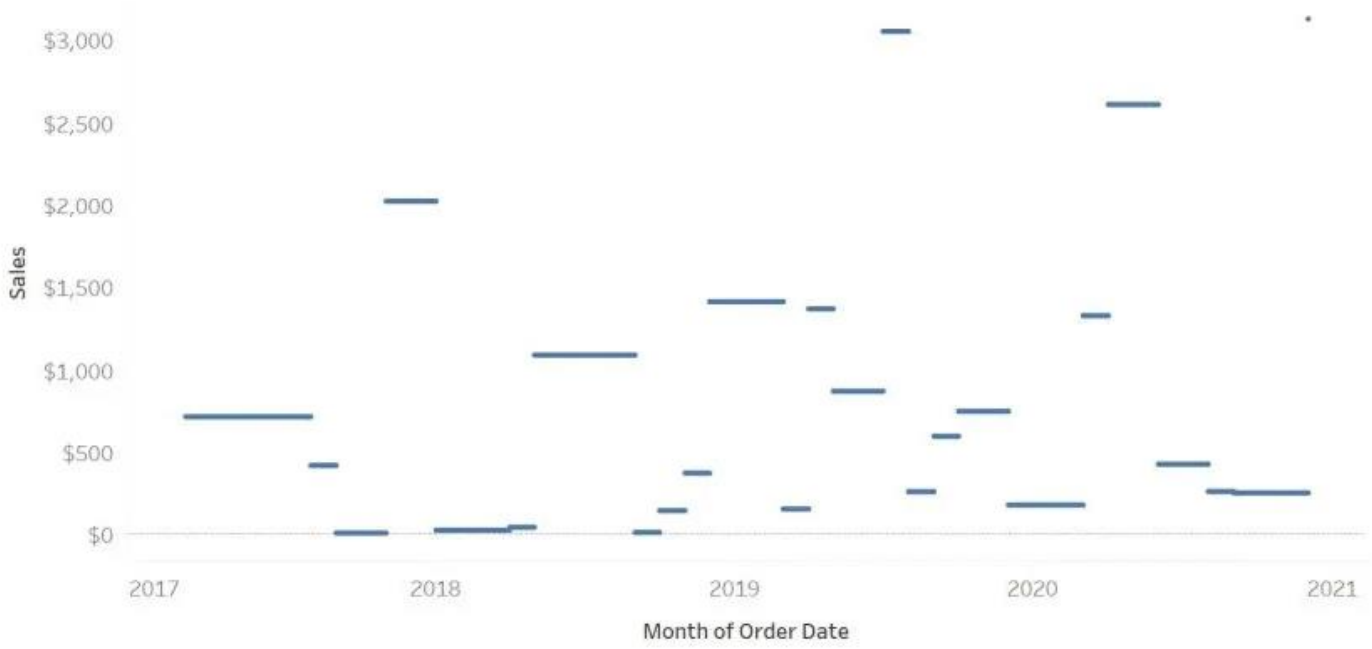
- **LINE CHART**
  - Shows trend over time to track changes over short or long time periods.
  - When tracking changes over short period, it's better to use a line chart rather than a bar chart.
  - There is an assumption that lines are continuous and there is a gradual transition from one point to the next.

CHOOSING THE RIGHT VISUAL - STEP LINE CHART

- Same as a line chart but uses straight lines.
- Show the magnitude of change between quantitative values
- Show that values are not continuous
- Show changes happen at irregular intervals

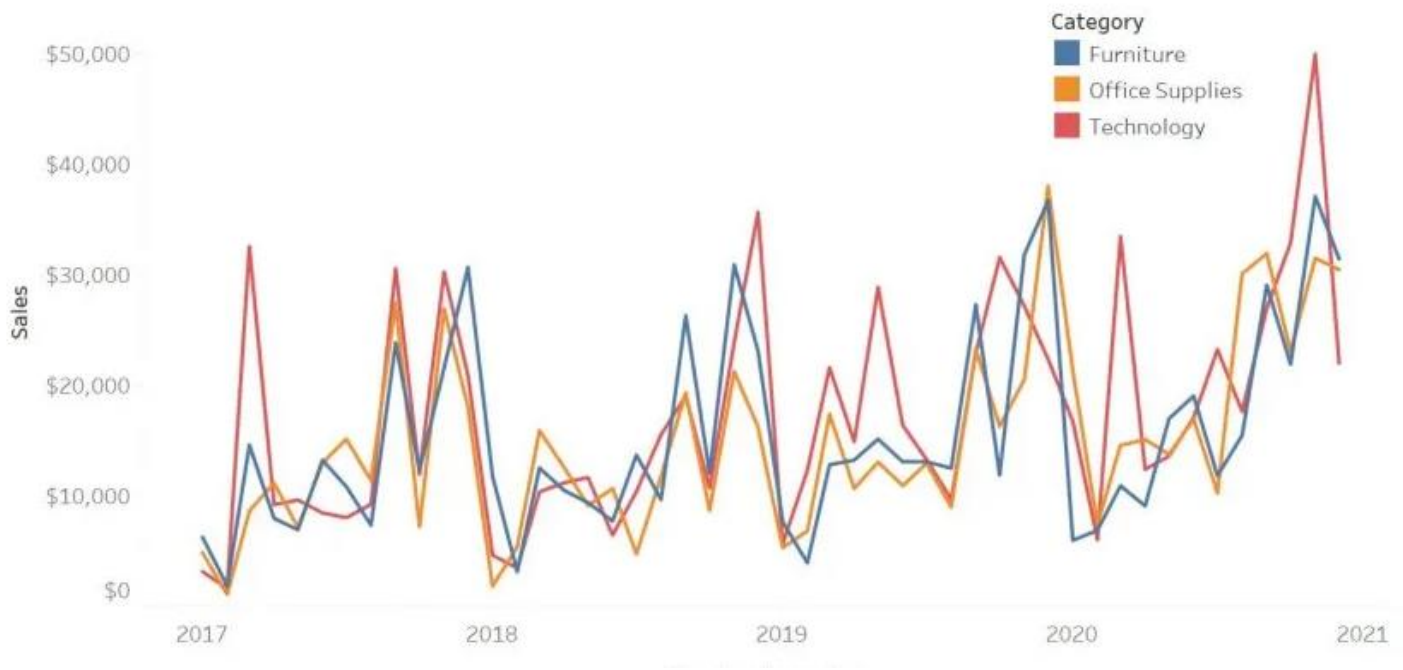
CHOOSING THE RIGHT VISUAL- JUMP CHART

- Same as step line but does not use the vertical lines.
- Shows the length of time between data points
- Show the length of time over which change happens
- Not a good chart to show the magnitude of change



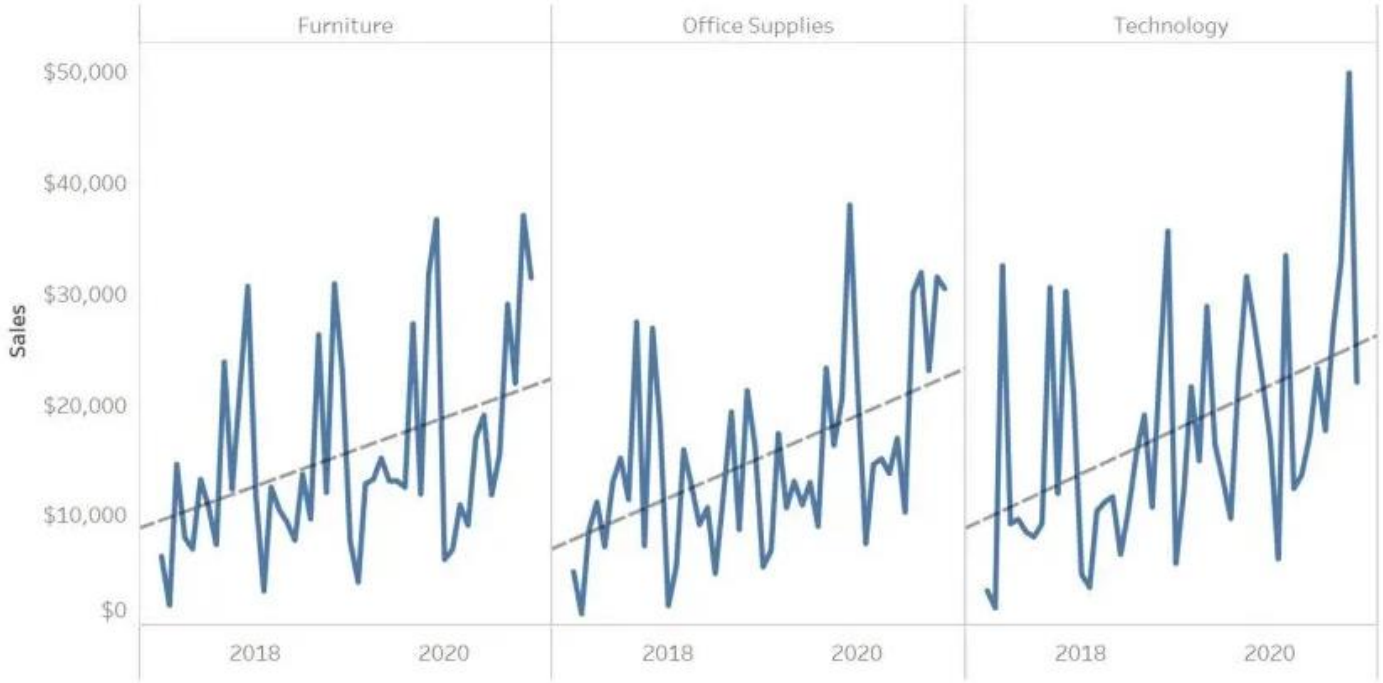
CHOOSING THE RIGHT VISUAL - MULTIPLE LINE CHART

- Same as line chart but with multiple categories
- Shows trend over time and seasonality
- Get confusing when you have too many lines.
- Its best to show similarity in trend for all categories



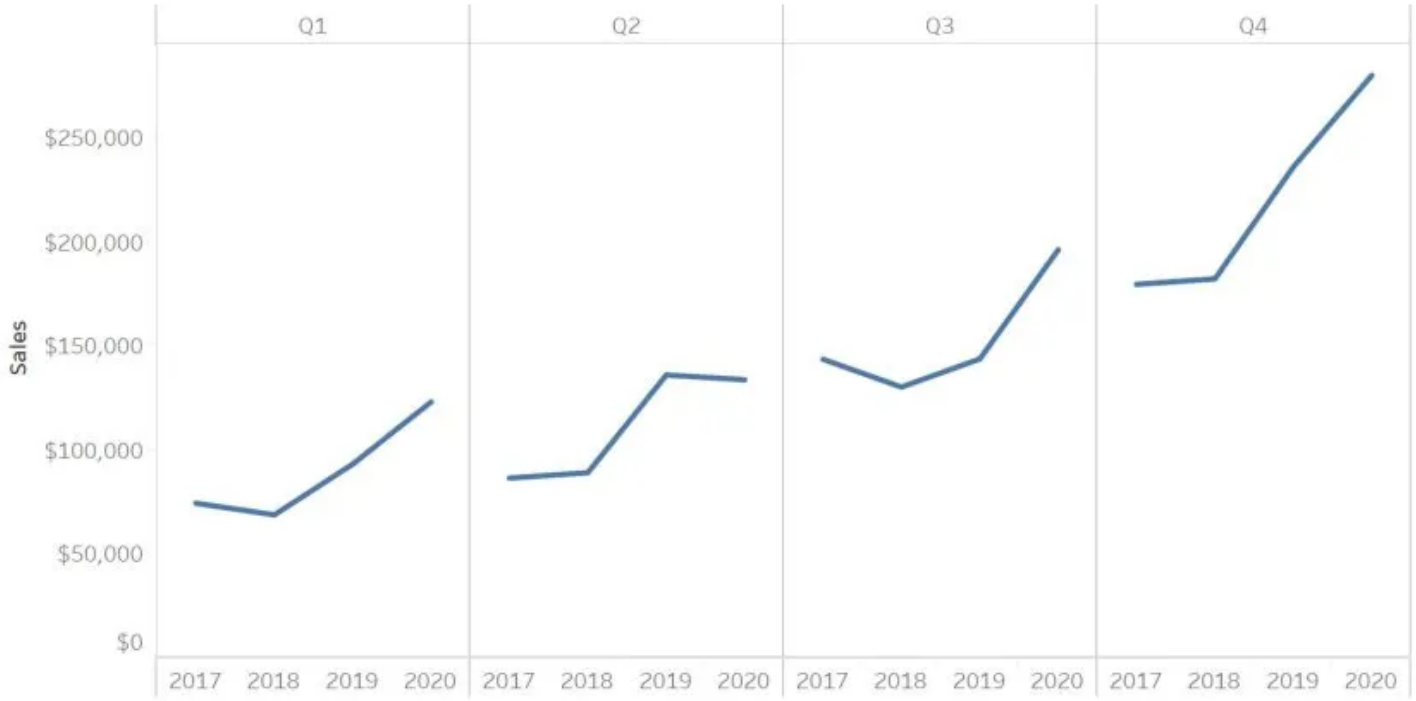
CHOOSING THE RIGHT VISUAL - SMALL MULTIPLE CHART

- Compare multiple trends side by side
- Details that you see in the multiple line charts will not be seen here.
- You see each individual category more clearly and you can still compare



CHOOSING THE RIGHT VISUAL - CYCLE PLOT CHART

- Show trend and seasonality in time-series data
- Trend for
  - day-of-week
  - hour-of-day
  - month-of-year
  - quarter-of-year



CHOOSING THE RIGHT VISUAL - HEAT MAP

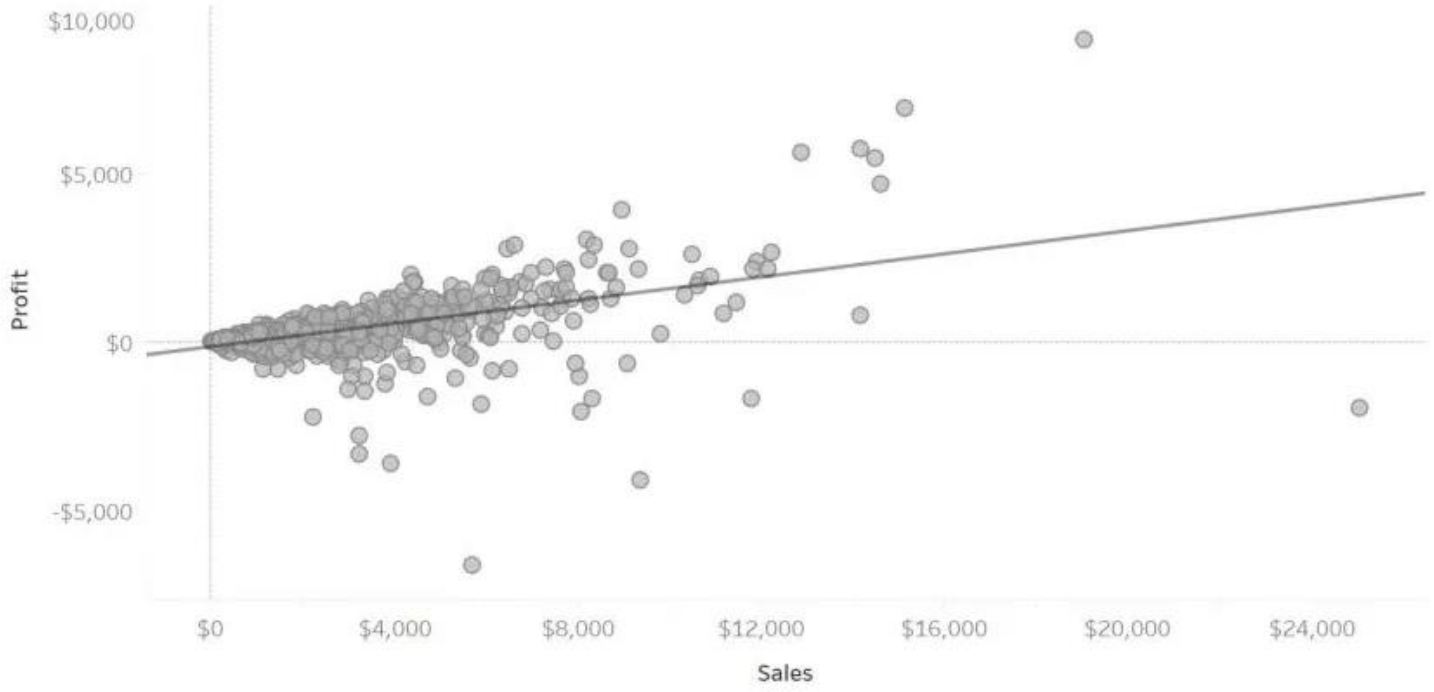
- Show trend and seasonality in time-series data
- Matrix of colors with different intensity.
- It is great to view outliers based on the intensity of the color.

CHOOSING THE RIGHT VISUAL - CORRELATION

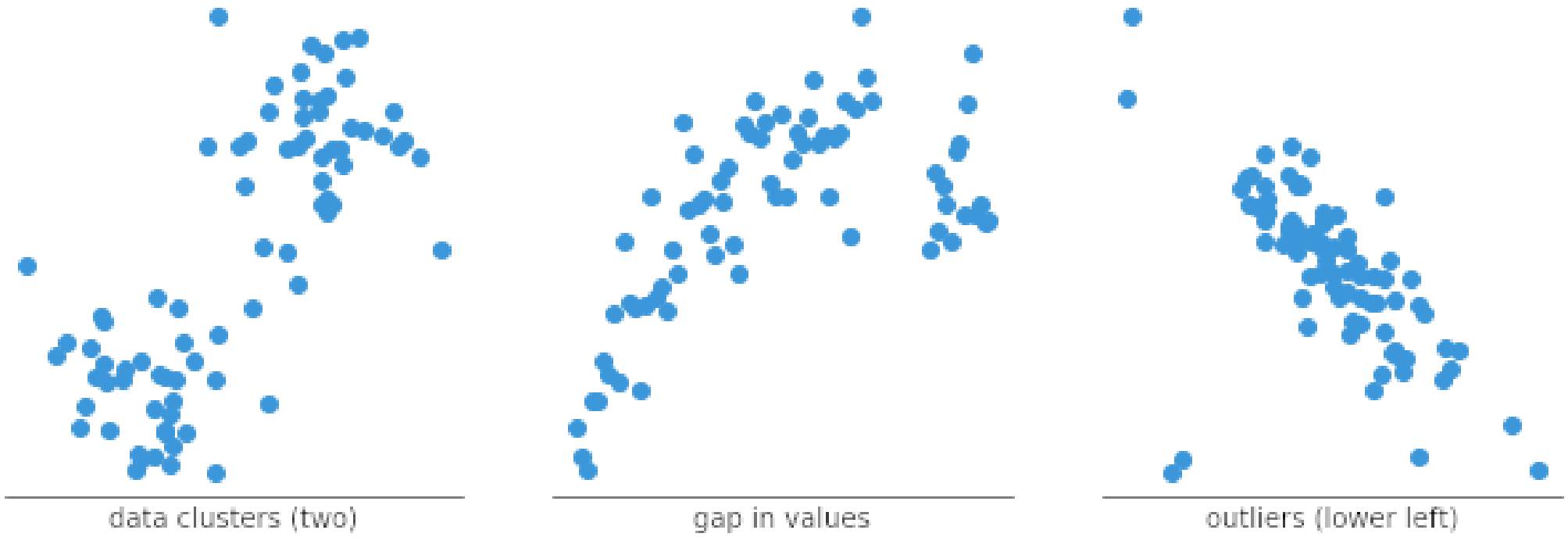
- When you want to show relationships between your data points.
- Even when there is a correlation between two variables, we can only say that there is an **observable relationship** and not conclude that one variable causes a change in the other.
- The correlation could be coincidental or there could be other factors that could be contributing to the correlation.

CHOOSING THE RIGHT VISUAL - SCATTERPLOT

- Shows relationships of two continuous variables.
- The trend line shows the direction in which the group of points follow
- Highlights outliers and clusters of data



CHOOSING THE RIGHT VISUAL- SCATTERPLOT

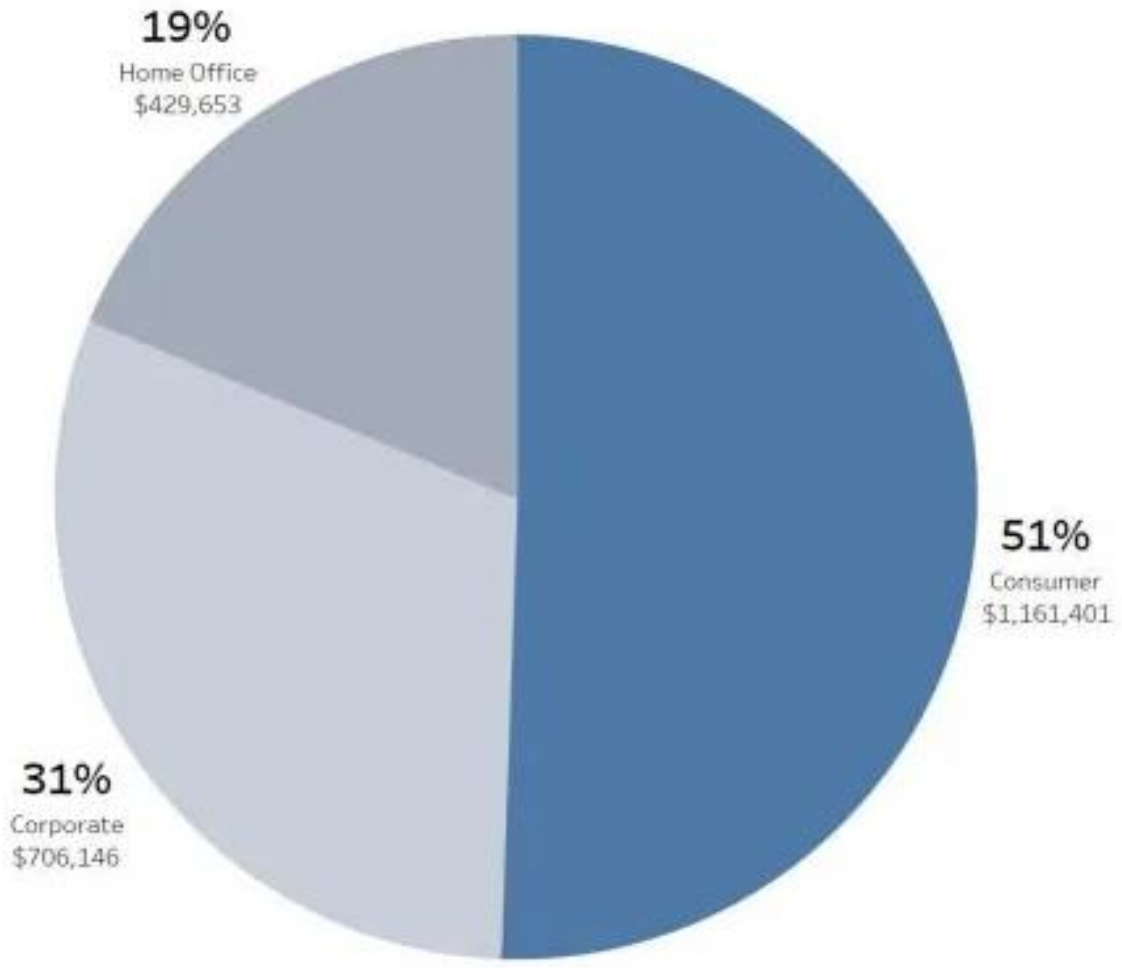


CHOOSING THE RIGHT VISUAL COMPOSITION

- When you want to show composition of your **qualitative data**.
- This usually that takes into account **part to whole relationships**

CHOOSING THE RIGHT VISUAL - PIE CHART

- When you want to show **rough proportions** to give your audience a rough idea of what composes a quantitative value.
- Use labels to provide numeric equivalent of the slices.
- It's not a good choice when you need accurate measurements

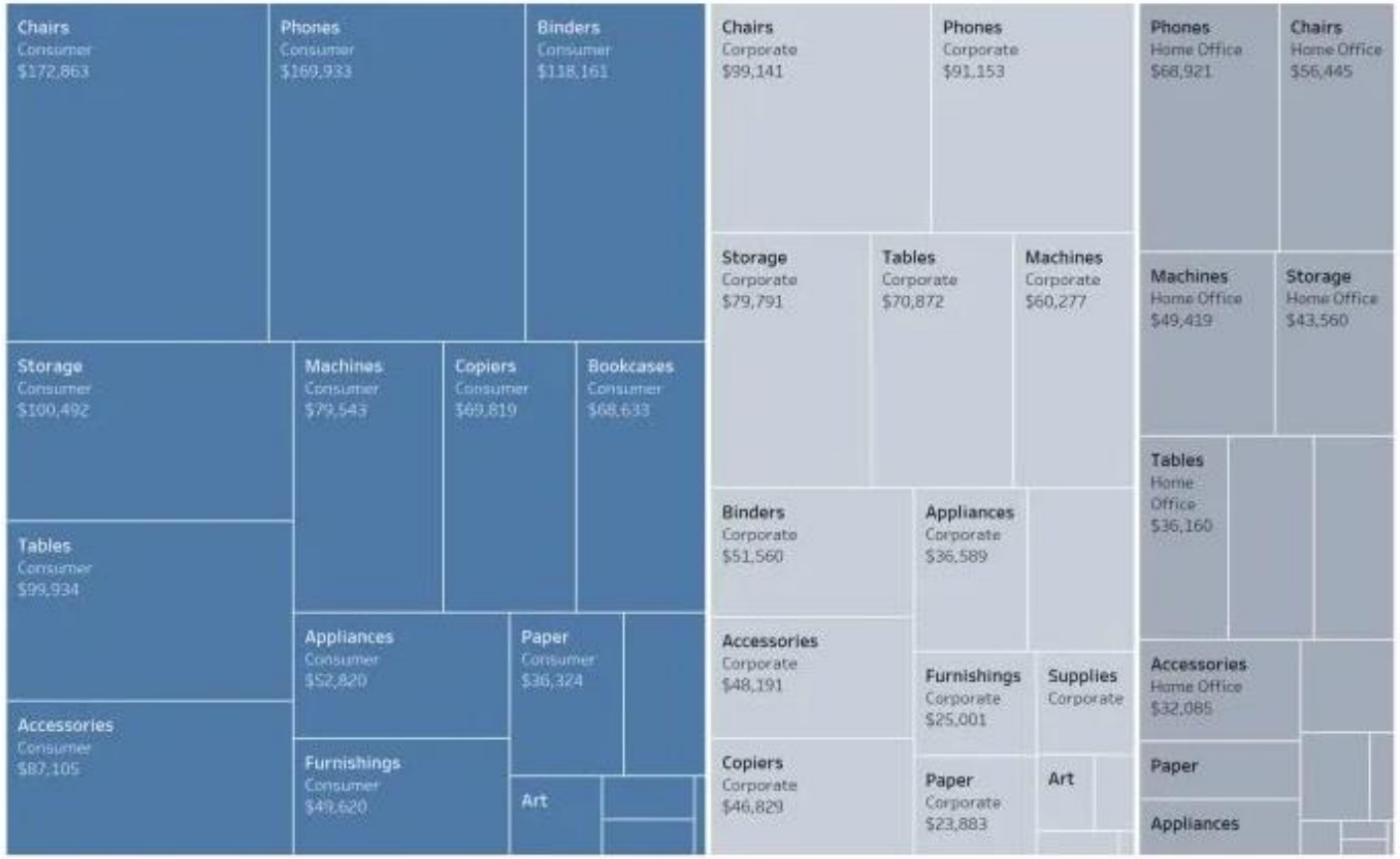


CHOOSING THE RIGHT VISUAL - STACKED BAR

- Show the **relative proportion** rather than the absolute value.
- The numbers are converted to represent their proportion to the whole (100%)
- Used when it's important to highlight the relative proportion

CHOOSING THE RIGHT VISUAL - TREE MAP

- Show **part to whole relationship** just like the pie chart.
- Show proportions using area not angles and **within the hierarchy**.
- Best used as an interactive chart

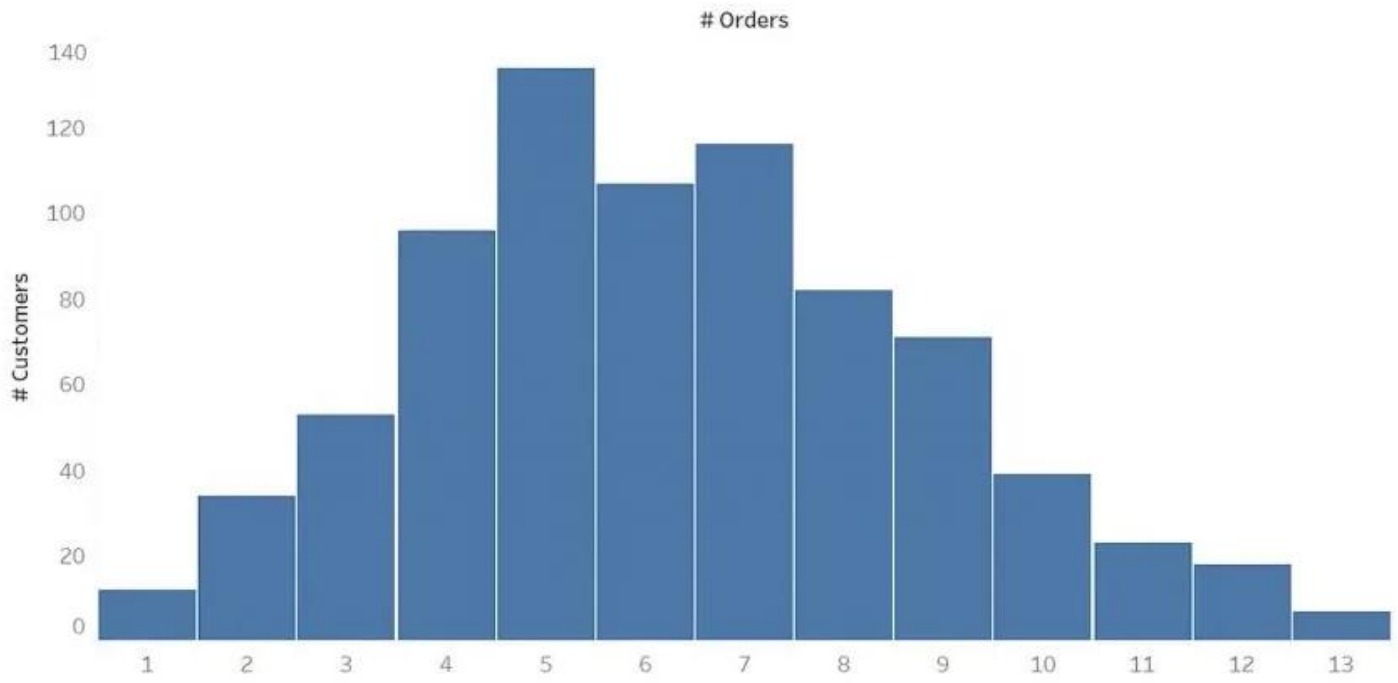


CHOOSING THE RIGHT VISUAL - DISTRIBUTION

- When you want to show distribution of your **qualitative data**.
- The attribute for which you want to display value is on the X axis.
- By default, the metric displayed on the Y axis is the record count of the attribute on the X axis.

CHOOSING THE RIGHT VISUAL - HISTOGRAM

- Shows the frequency distribution.
- Shows **how frequently** data occurs in certain intervals called bins
- The length of the bin corresponds to the quantitative value of the bin.

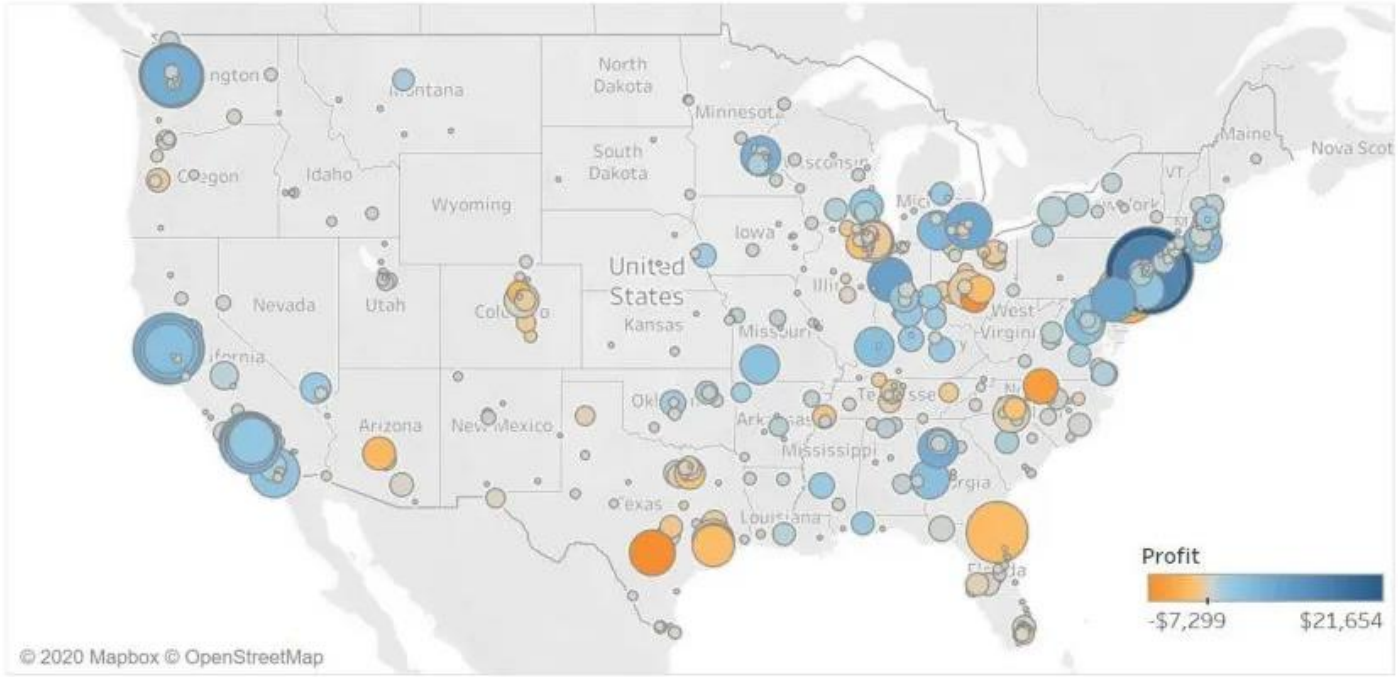


CHOOSING THE RIGHT VISUAL - SPATIAL

- Want to show cluster of information in a spatial area
- When data points are plotted on a map, it can provide more meaningful context and insights.
- Using spatial charts can help with strategizing and planning.
- Have to be careful about what the numbers represent, and if it skews interpretation.
- **Always use "per capita" means per population.**

CHOOSING THE RIGHT VISUAL - PROPORTIONAL SYMBOL MAP

- Uses symbols that can vary in size based on the quantitative value it represents.
- We can also use color alongside size to represent the same or different measures.
- Useful when we want to see cluster or concentration of circles in an area.
- Be careful about how to interpret the symbols and
- **use standardize measures.**



CHOOSING THE RIGHT VISUAL - FILLED MAP

- Instead of symbols we use color to represent quantitative values.
- Clear boundaries - at state, at city, or at zip code level
- Make it as granular or as aggregated as needed

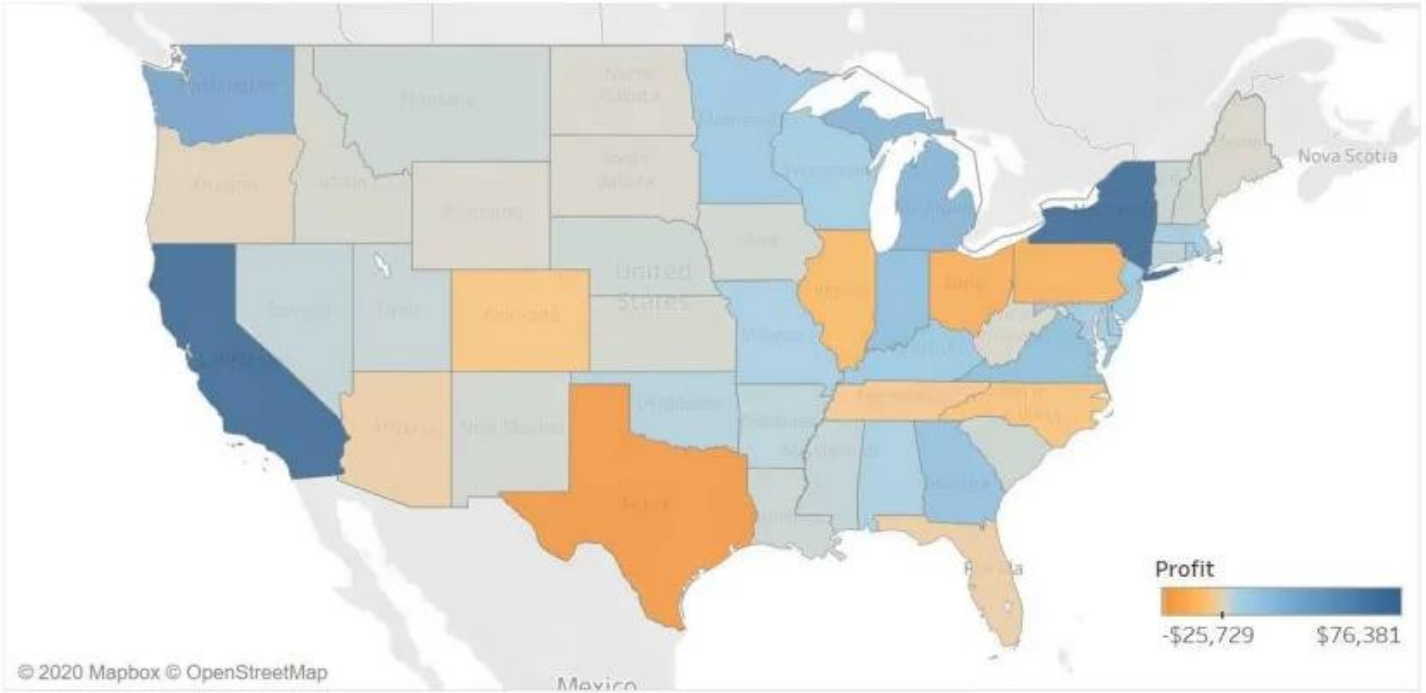


CHART SELECTION TOOL

Dr Andrew Abela created this Chart Selection Diagram that will help you pick the right chart for your analysis.

You can download the chart at

Chart Selection Diagram