DATASOCIETY:

Introduction to Tableau

Day 1

Who we are

Data Society's mission is to integrate Big Data and machine learning best practices across entire teams and empower professionals to identify new insights.

We provide:

- High-quality data science training programs
- Customized executive workshops
- Custom software solutions and consulting services

Since 2014, we've worked with thousands of professionals to make their data work for them.

About the course

- **Instructor introduction**

- **Schedule:**
  - 4 days
  - 11 am – 2 pm
  - 1 or 2 short breaks per session
  - Individual and group workshop activities and presentations

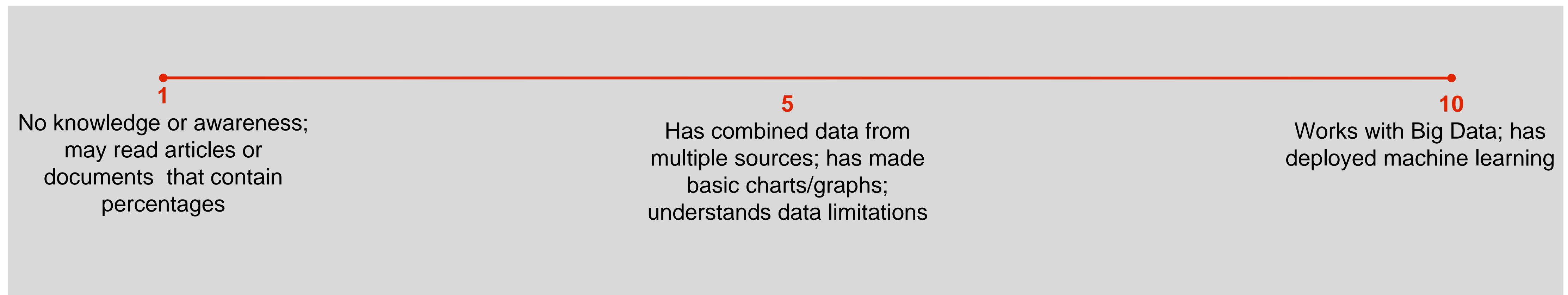# Best practices for virtual learning

1. Find a quiet place, free of as many distractions as possible. Headphones are recommended.

2. Stay on mute unless you are speaking.

3. Remove or silence alerts from cell phones, e-mail pop-ups, etc.

4. Participate in activities and ask questions. This will be interactive!

5. Give your honest feedback so we can troubleshoot problems and improve the course.
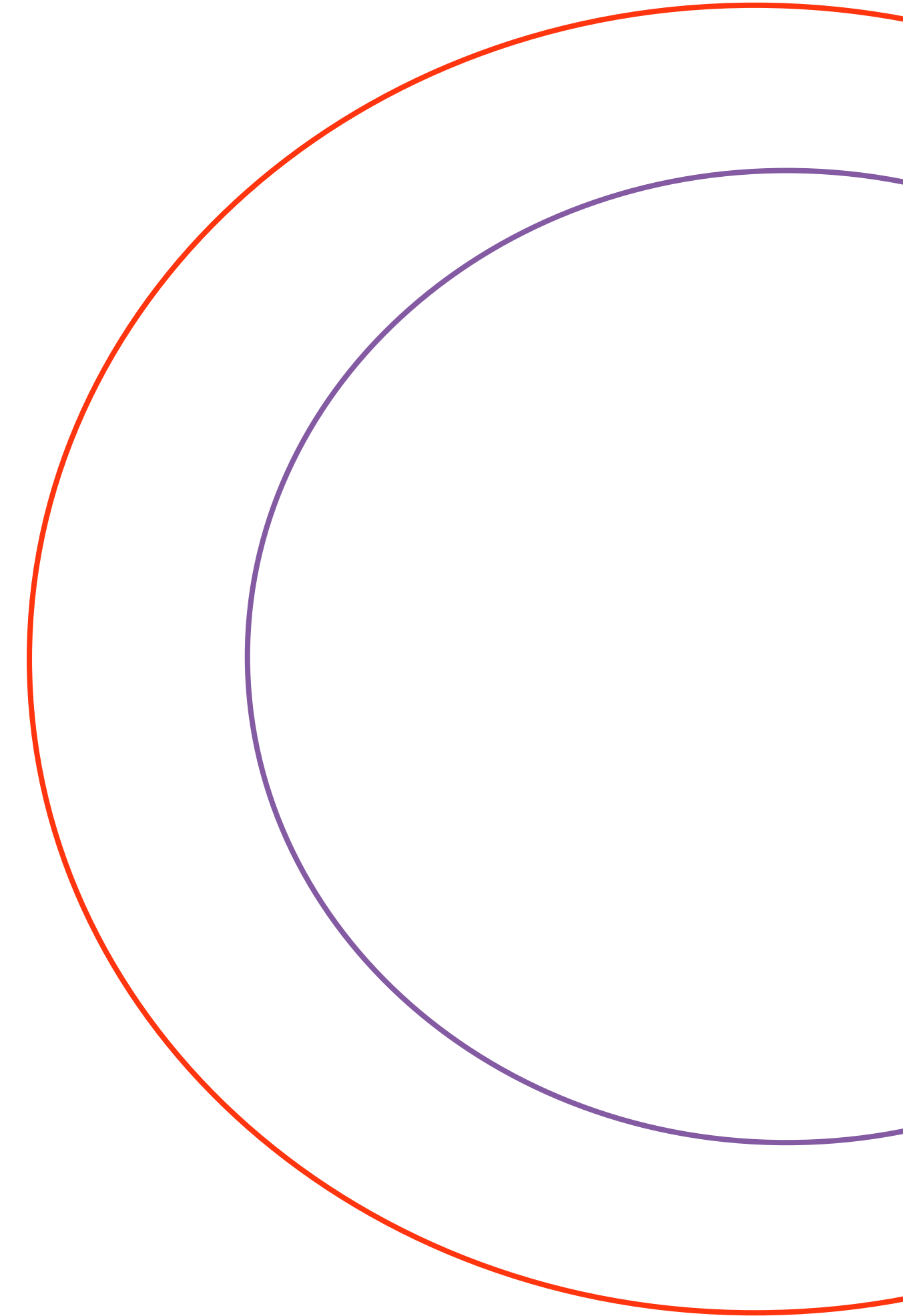
Chat: Introductions

- Introduce yourself with the following details:
    1. Name
    2. Job title
    3. What you hope to get out of the training
    4. What you rate your current data literacy level on a scale of 1-10

1
No knowledge or awareness; may read articles or documents that contain percentages

5
Has combined data from multiple sources; has made basic charts/graphs; understands data limitations

10
Works with Big Data; has deployed machine learning

Agenda

- Introduce data visualization

- Discuss chart types by data

- Discuss chart types by form

- Highlight visual building blocks

- Introduce color theory and best practices

- Explain the need for Tableau and describe its features

Agenda

- Introduce data visualization

- Discuss chart types by data

- Discuss chart types by form

- Highlight visual building blocks

- Introduce color theory and best practices

- Explain the need for Tableau and describe its features

What is data visualization?

- **Data visualization** is any attempt to make data more easily digestible by rendering it in a visual context (e.g., charting, graphing, etc.)

- We use data visualization to transform raw data into something compelling

- Data visualization is at the **intersection of art and science**
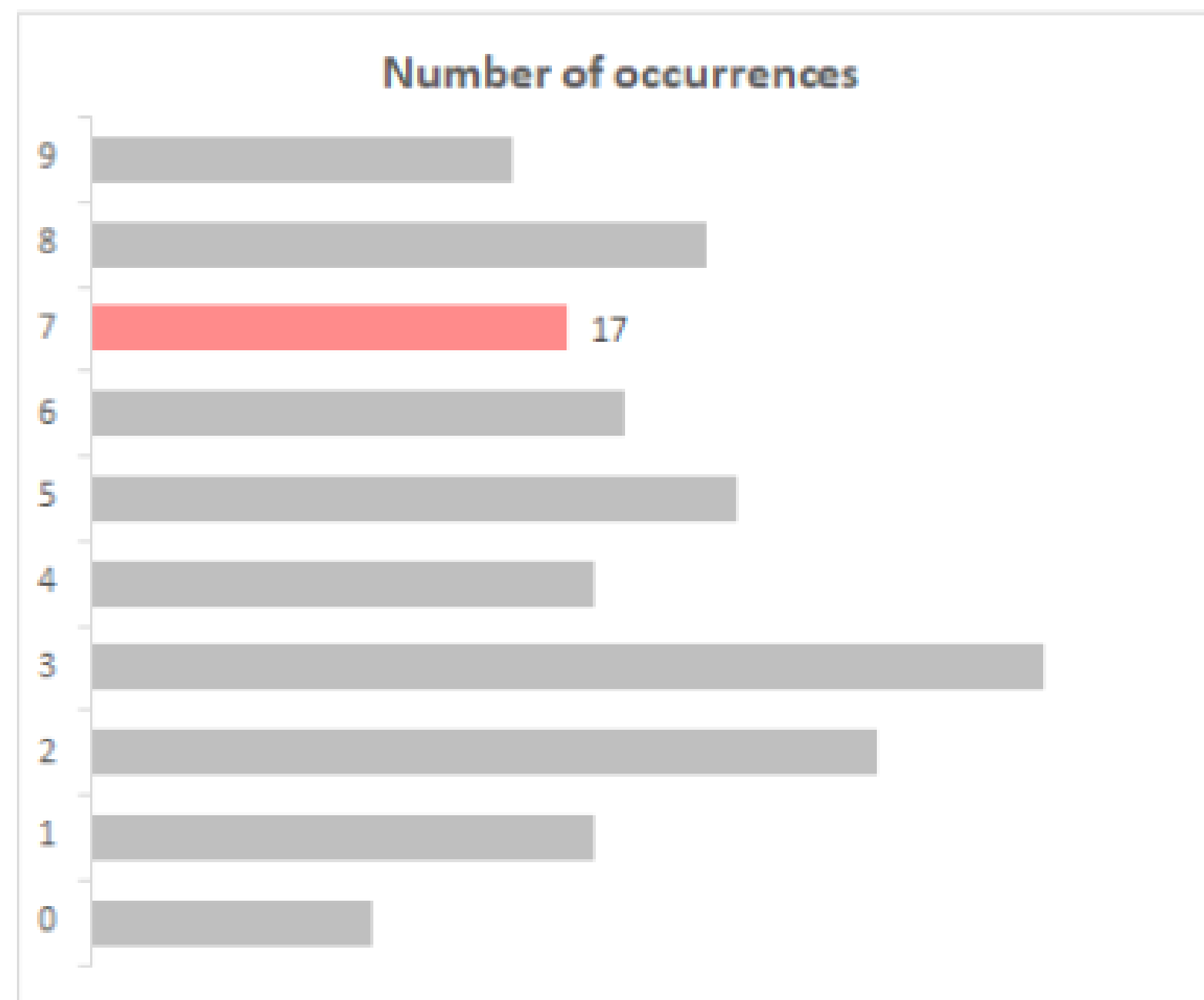
Data viz example

Count how many times the number 7 appears in the grid below.

```
5 2 8 3 6 1 9 3 6 2 5 3 7 4 3 8 3
8 5 8 9 6 2 1 4 4 3 9 3 6 5 2 4 9
1 0 2 7 5 2 8 3 6 1 6 2 9 3 8 3 8
5 8 4 7 2 0 3 7 3 5 4 7 1 8 2 0 1
2 5 3 6 4 3 9 1 0 8 9 5 7 3 4 5 3
2 7 5 2 8 3 6 1 6 2 9 3 8 3 8 5 8
4 7 2 0 3 7 3 5 4 7 1 8 2 0 1 9 6
2 1 4 4 3 9 3 6 5 2 4 9 1 0 2 7 5
2 8 3 6 1 6 2 9 3 8 3 8 5 8 4 7 2
0 3 7 3 5 4 7 1 8 2 0 1 2 5 3 6 4
3 9 1 8 9 5 0 7 3 4 5 3 2 7 5 2 8
3 6 1 6 2 4 6 2 7 5 9 1 5 2 6 3 6
```
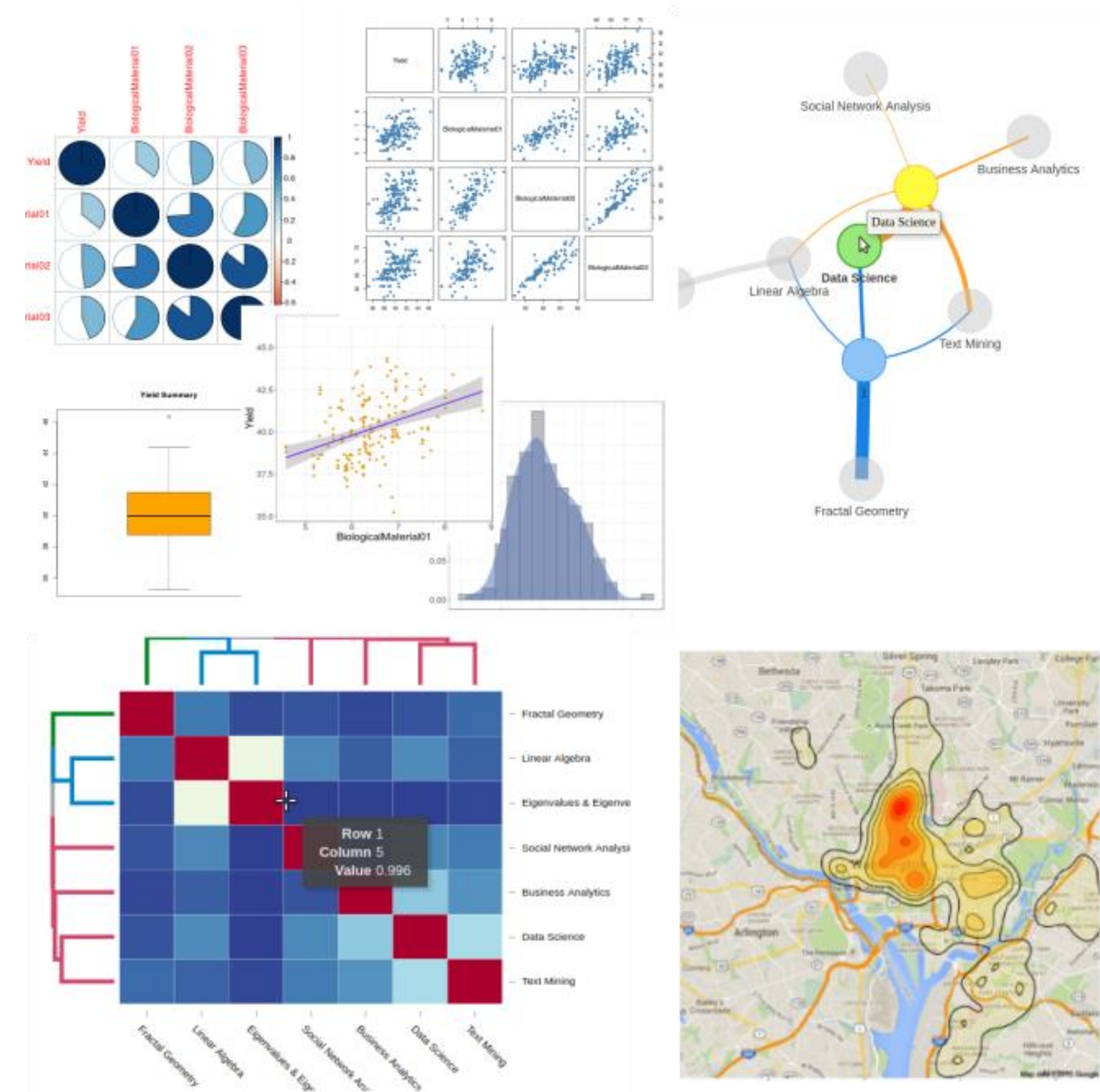
Data viz example

What about now? Is it easier to count?

Why build a visualization?

- To provide valuable, interpretable, and relevant **insights**

- To give a **visual or graphical representation** of data / concepts

- To **communicate** ideas

- To provide an **accessible** way to see and understand trends, outliers, and patterns in data

- To try to confirm a **hypothesis**

The Challenger explosion

- The 1986 **Space Shuttle Challenger explosion** is an emblematic case study in how data visualization can play an important role in decision-making

- The explosion happened due to low temperatures that affected shuttle parts

- **Edward Tufte**, a visualization expert, argues that the cause of this tragedy was an unreadable format of data given to launch decision makers
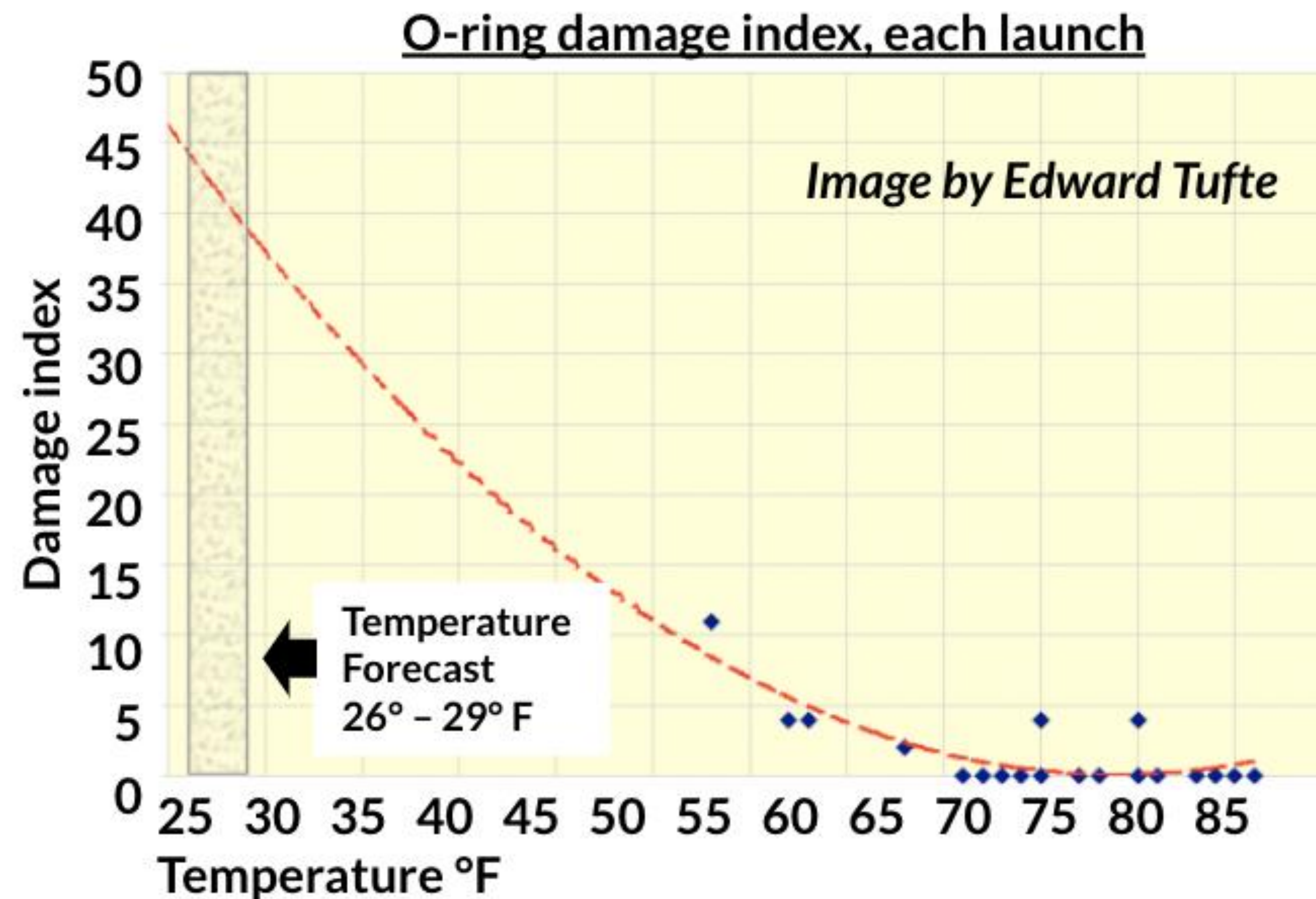
The original temperature visualization

- The chart below was presented to the experts at the time
- How easily interpretable do you think it is?
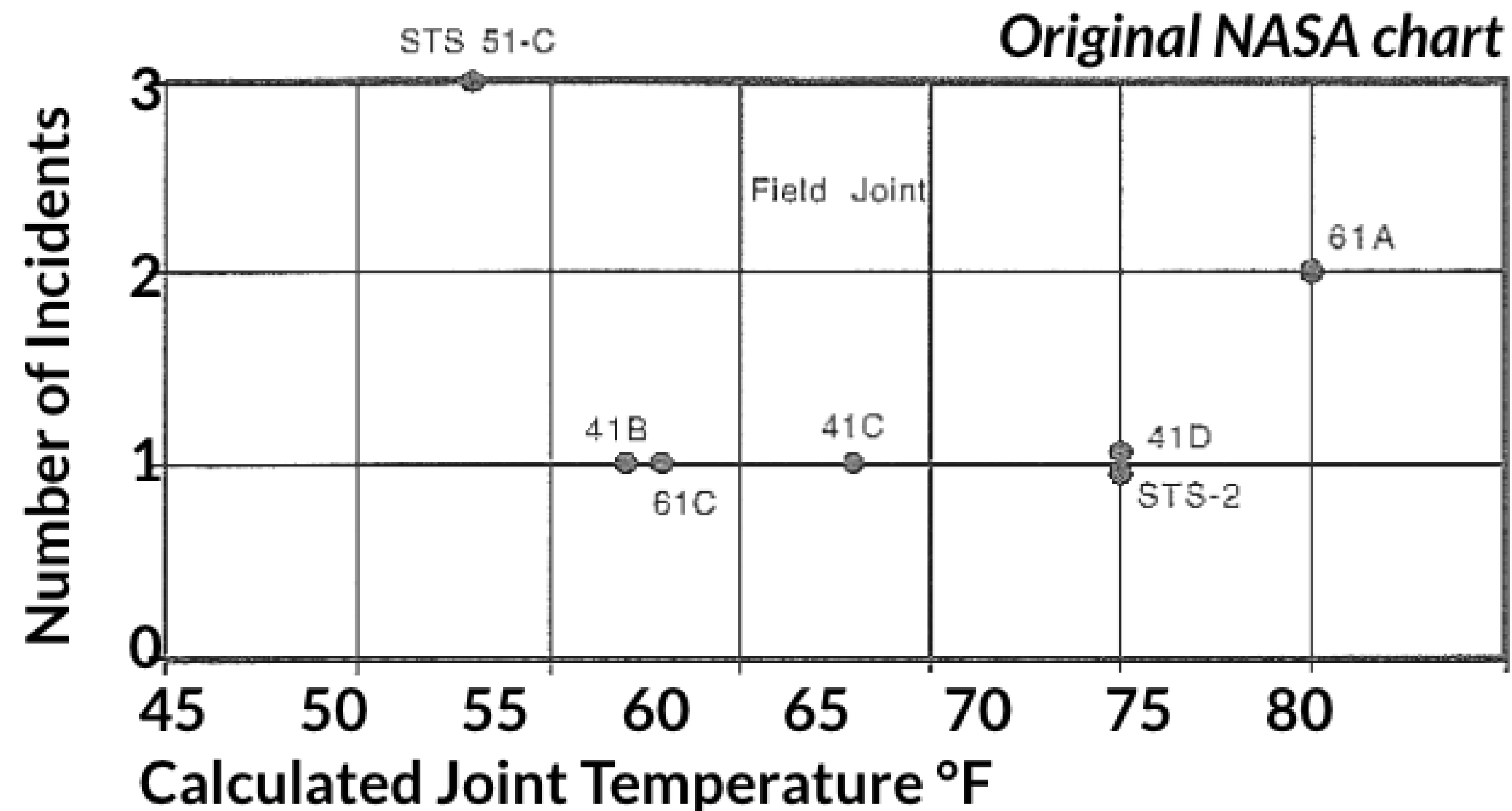
The revised temperature visualization

- Edward Tufte argues a better chart may have prevented disaster
- How easily interpretable do you think the revision he created is?

O-ring damage index, each launch

Chat questions: importance of data viz

- In breakout rooms, discuss the following questions:

  - What **types** of data visualizations does your organization produce? Who is the **audience**? What is the **purpose**?

  - What **improvements** would you like to see in the visualizations created or used by your organization? Why?

Agenda

- Introduce data visualization

- Discuss chart types by data

- Discuss chart types by form

- Highlight visual building blocks

- Introduce color theory and best practices

- Explain the need for Tableau and describe its features

Categorical data

- **Categorical data** is non-numeric or qualitative

- **Insight:** comparisons and proportions

- **Chart types:** vertical bar, column bar, horizontal bar, pie, bullet charts, stacked bar, and tree maps
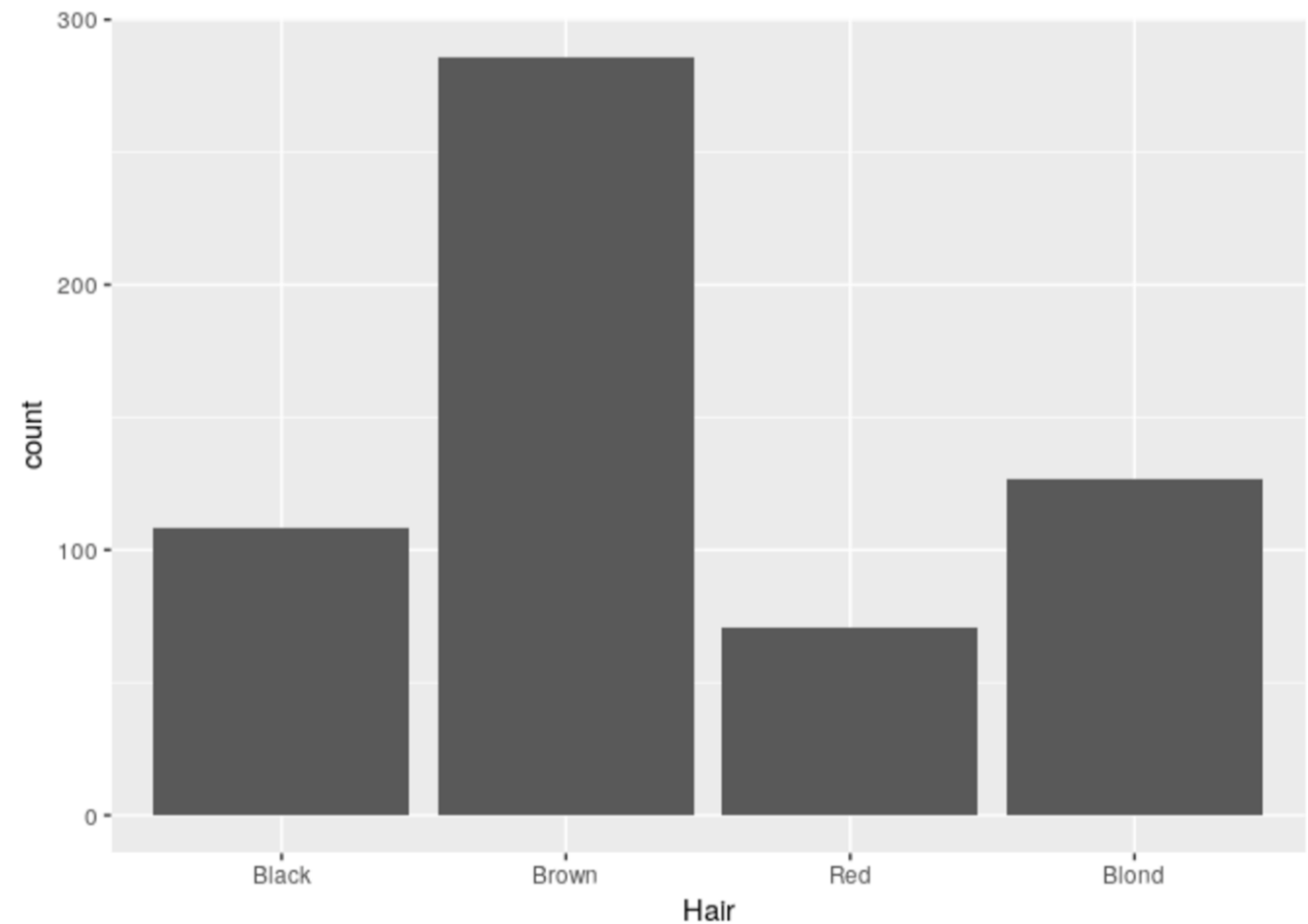
Univariate data

- **Univariate data** consists of a single numeric variable

- **Insight:** distributions, proportions, and frequencies

- **Chart types:** histogram, density, box plots
  - Note distribution shape and outliers – is the data normal?
 - These are the go-to for initial data exploration
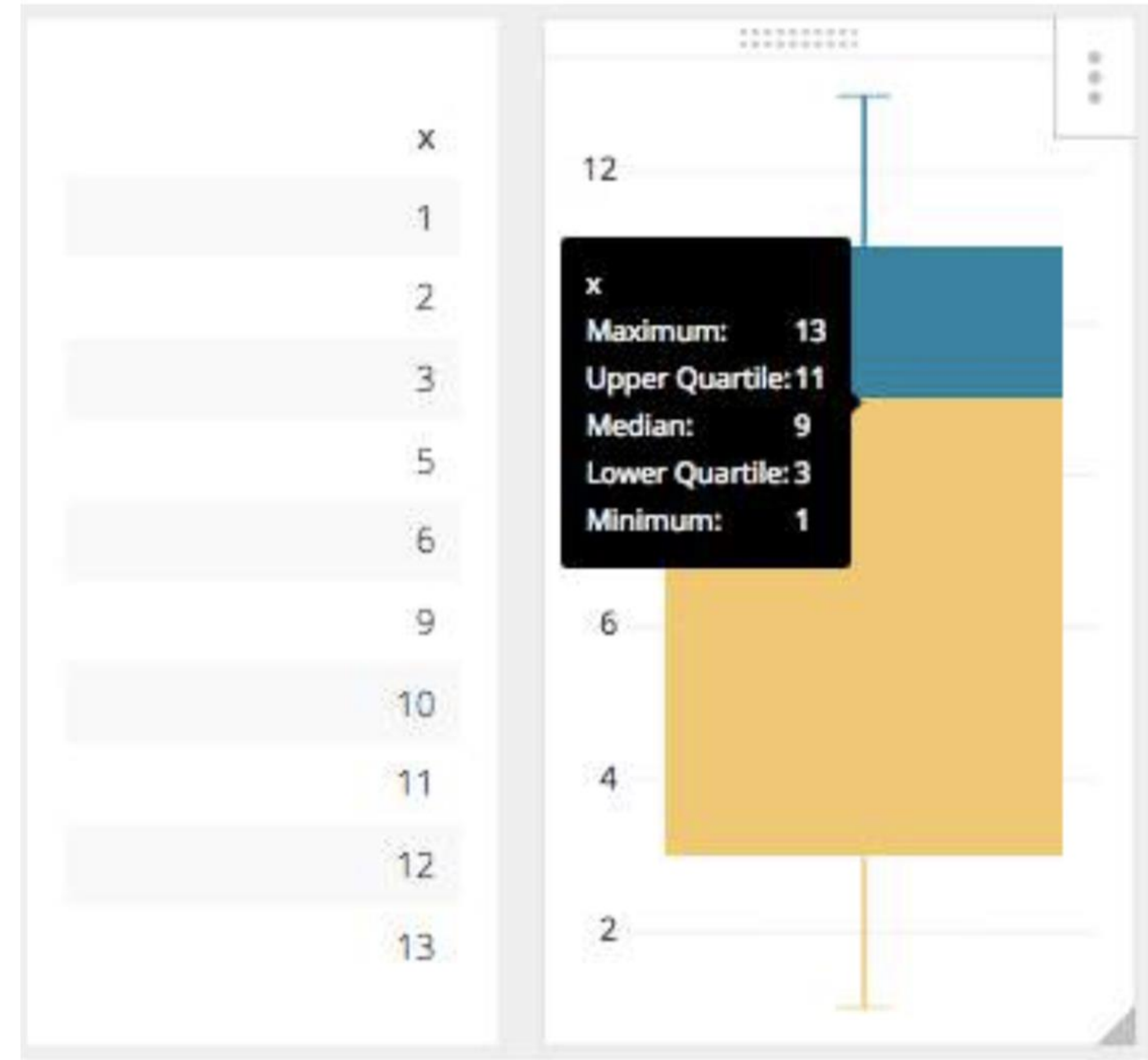
Bivariate data

- **Bivariate data** consists of two (or more) numeric variables (i.e. weight and height)

- **Insight:** relationships, correlation, proportions, and frequencies

- **Chart types:** scatterplot, bubble, parallel, radar, bullet, and heat
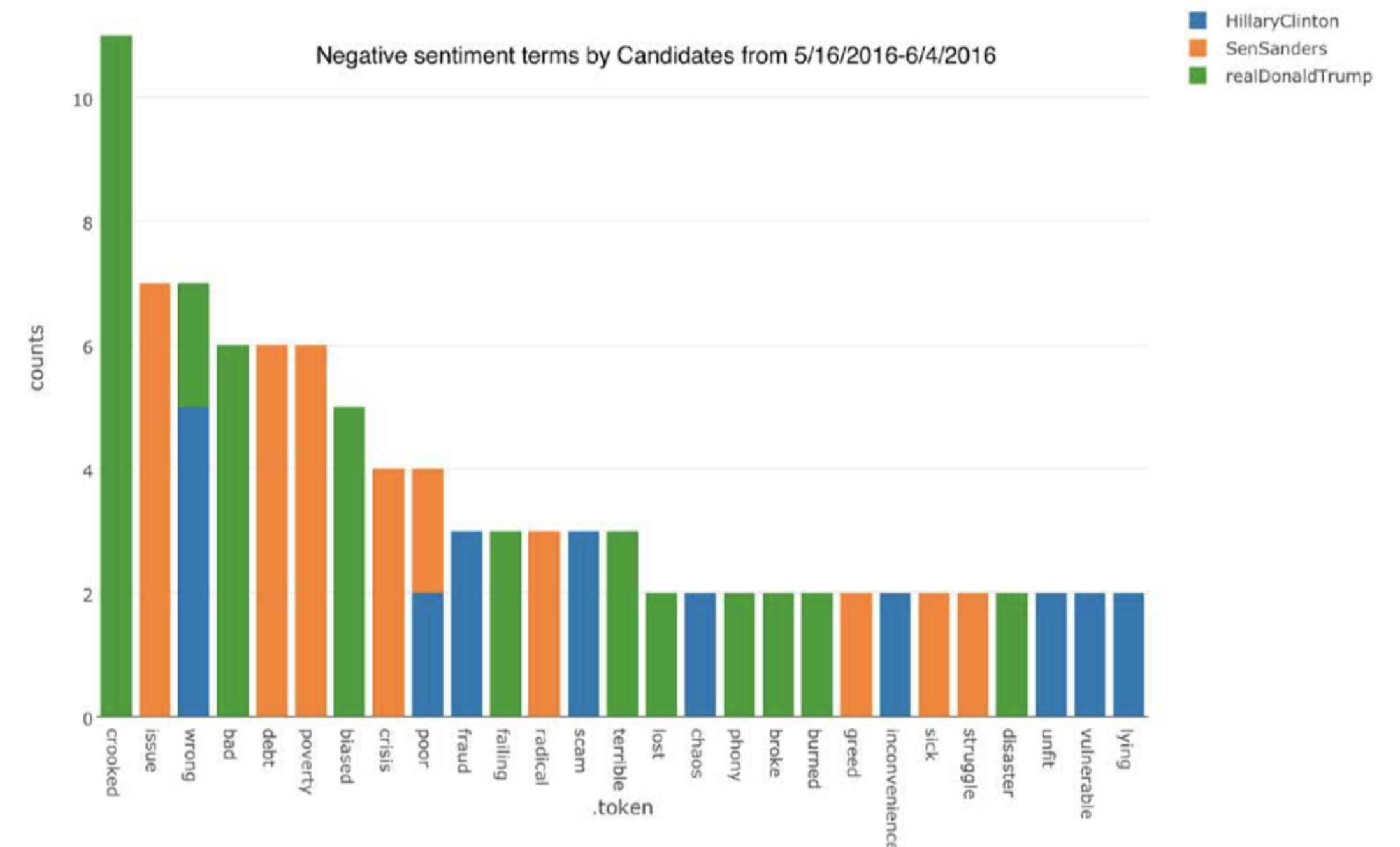

Weight and Height of Children
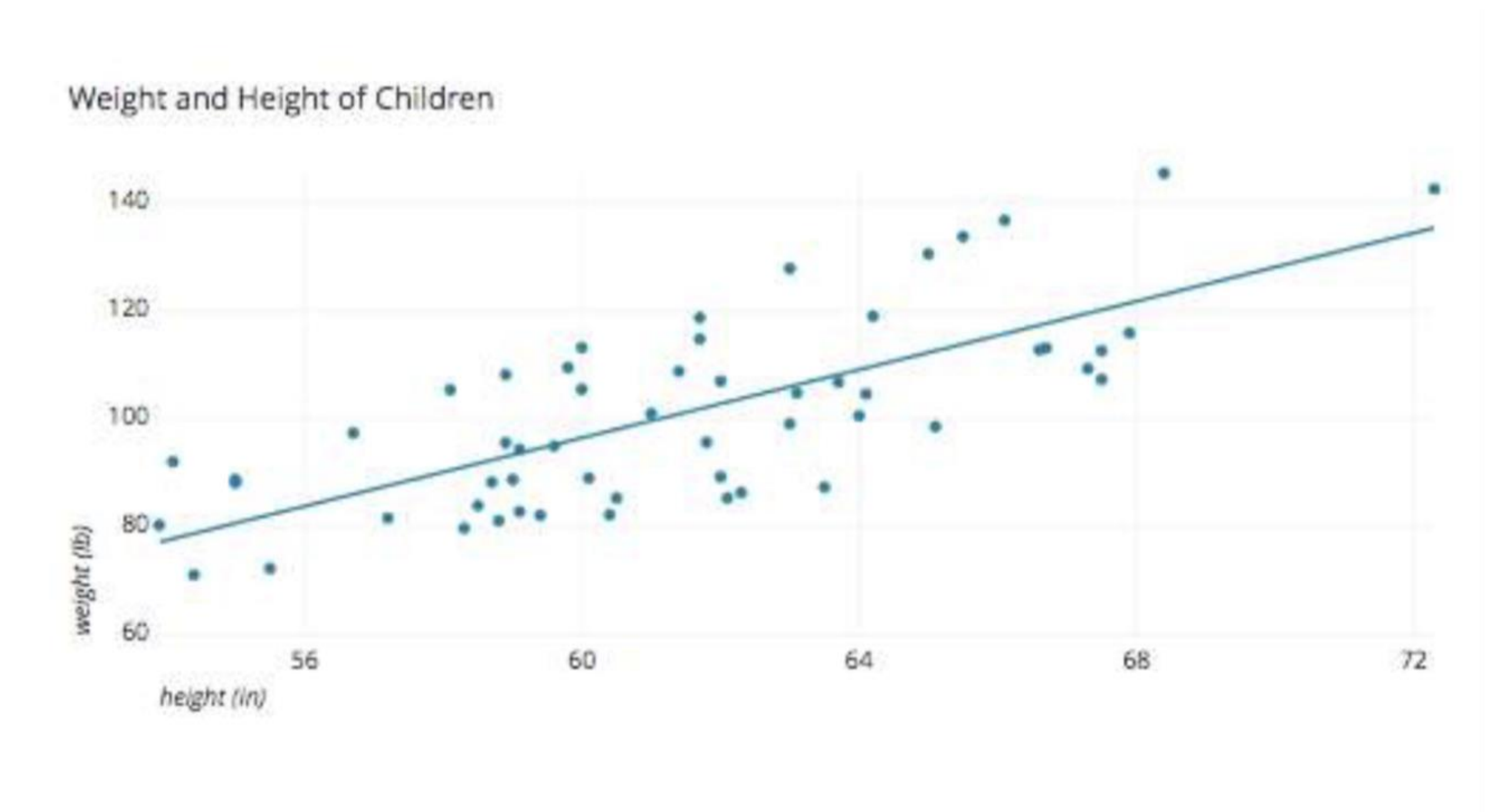
Trend data

- **Trend data** includes a time-based data (i.e., years, months, days, hours, etc.)

- **Insight:** trends, comparisons, and cycles

- **Chart types:** line, area, bubble, vertical bar

Annual Percent Change in North American Population
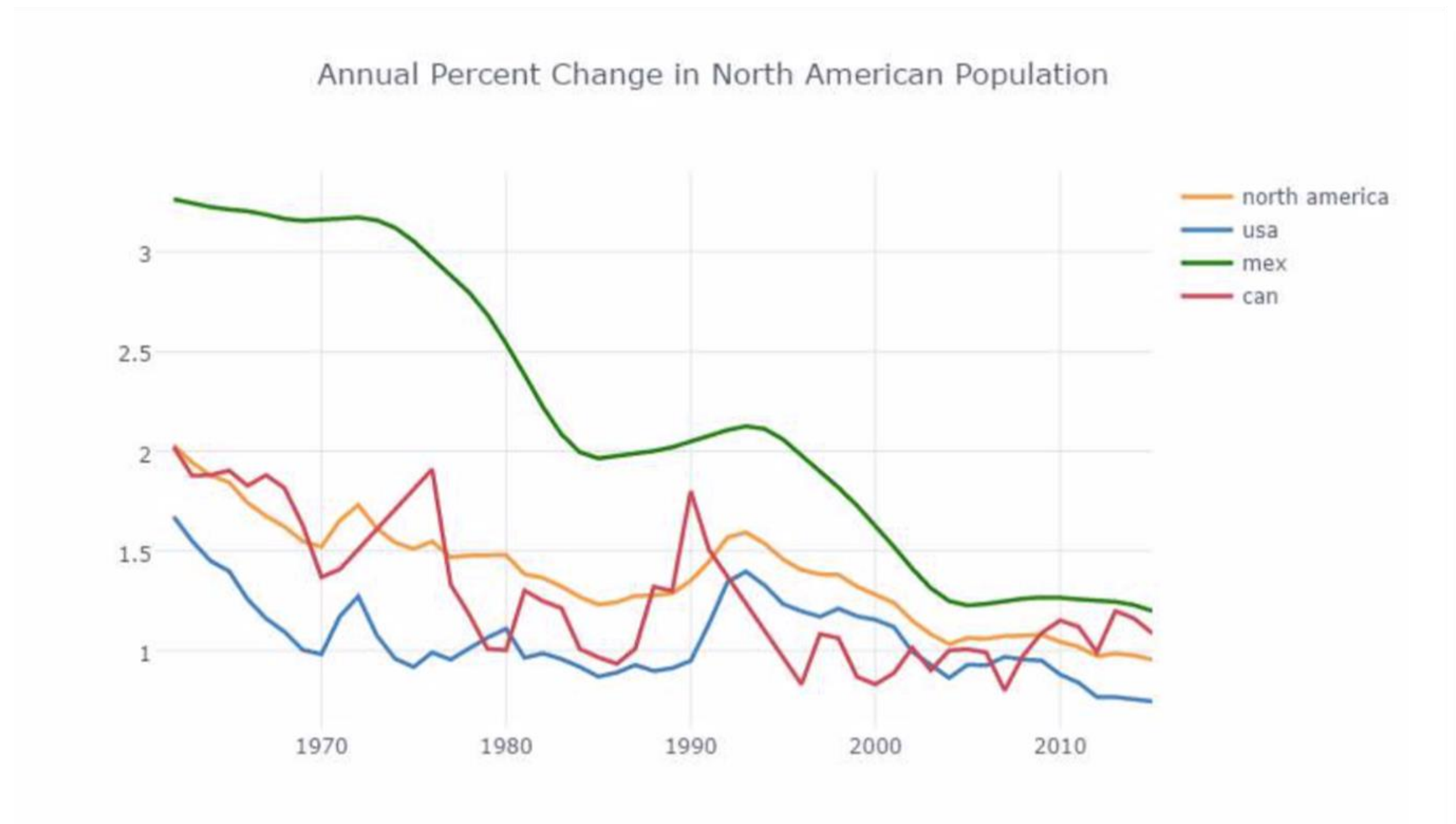
Text data

- **Text data** includes alphanumeric single words or phrases (keywords)

- **Insight:** sentiment, comparisons, and frequency

- **Chart types:** word cloud, histogram, stacked bar chart

- It's useful for analyzing frequency and making comparisons

Negative sentiment terms by Candidates from 5/16/2016-6/4/2016

Legend: HillaryClinton, SenSanders, realDonaldTrump

Geospatial data

- **Geospatial data** includes qualitative or quantitative information about specific locations

- **Insight:** locations, comparisons, and trends

- **Chart types:** choropleths filled map, point map, connection map, isopleth map

Smoke-free air law coverage by state (2017)
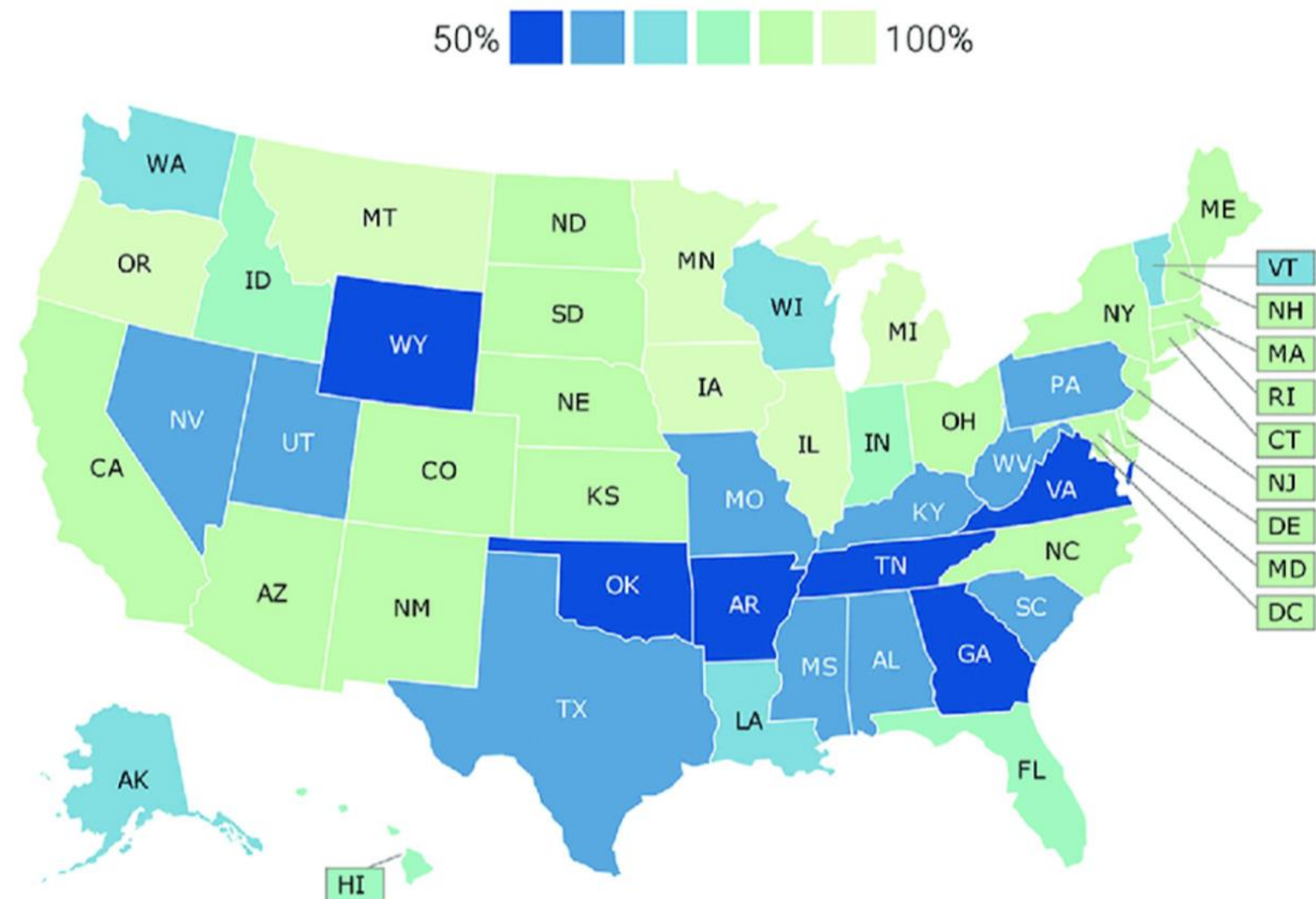
Agenda

- Introduce data visualization

- Discuss chart types by data

- Discuss chart types by form

- Highlight visual building blocks

- Introduce color theory and best practices

- Explain the need for Tableau and describe its features
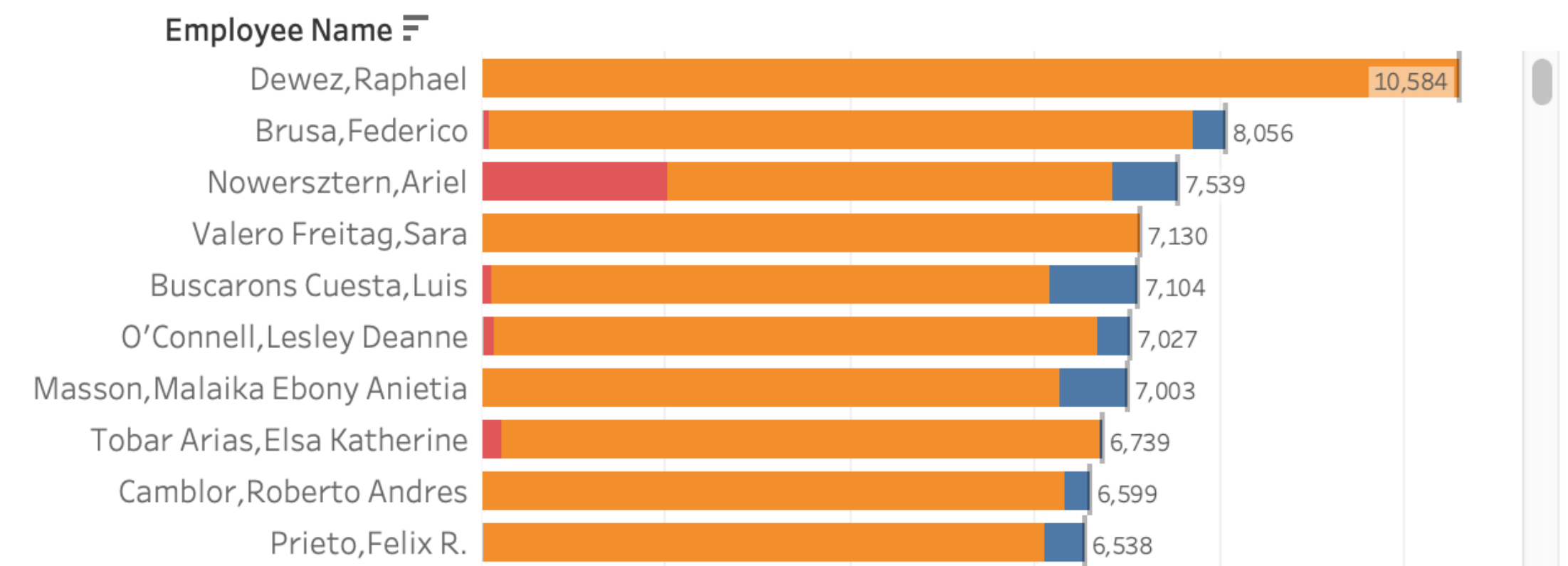
Simple text or table

- **Simple text:** when there is just a number or two to share, simple text can be a great way to communicate

  - "440 employees worked a total of 31,702 days, an average of 72.05 days per employee."

- **Tables:** useful when communicating to a mixed audience or to show a few different units of measure

Top Operations based on Work Days

| Index	Operation	Duration with the team	Work Days	Employee Count	Work Days/Employee
1	Portfolio Monitoring and Reporting	36	31,702	440	72.05
2	Support to Project Execution	36	21,315.375	385	55.364610390
3	Support to Project Preparation	36	21,270.25	422	50.403436019
4	Support to Fiduciary Work	36	16,639.5	204	81.566176471
5	Dialogue with public sector authoriti..	36	14,207.875	373	38.090817694
6	Technical Advisory and Quality Contr..	36	10,492	90	116.577777778
7	Economic Research (not product spe..	36	9,854.5	113	87.207964602
8	Country and Sector Programming	36	9,138.375	245	37.299489796
9	Trust Fund: Coordination and Corpor..	36	8,983.75	56	160.424107143
10	Macroeconomic monitoring	36	8,961.375	94	95.333776596
11	Strategic Outreach	36	7,495.875	249	30.103915663
12	Communications Planning and Client..	36	5,834.5	84	69.458333333
13	Support to TC (OS) Execution	36	5,731.5	161	35.599378882
14	Strengthening of Country Systems P..	36	5,293.375	111	47.688063063
15	INT Annual Research Report [Global ..	36	4,557.625	19	239.875

Bar chart

- **Bar charts** are used to express larger variations in data, how individual data points relate to a whole, comparisons, and ranking

- They express quantities through a bar's length, using a common baseline (=zero)

- **Note:** when the data has lengthy names, using a horizontal bar chart will make the data easier to read


Employee Names by number of hours

Employee Name	
Dewez,Raphael	10,584
Brusa,Federico	8,056
Nowersztern,Ariel	7,539
Valero Freitag,Sara	7,130
Buscarons Cuesta,Luis	7,104
O'Connell,Lesley Deanne	7,027
Masson,Malaika Ebony Anietia	7,003
Tobar Arias,Elsa Katherine	6,739
Camblor,Roberto Andres	6,599
Prieto,Felix R.	6,538

Line chart

- **Line charts** are used to plot continuous data in some unit of time: days, months, quarters or years

- They can also be used to show multiple series of data

- A line graph can also represent a summary statistic, like the average and confidence level range, or the point estimate of a forecast

Employee Ageing
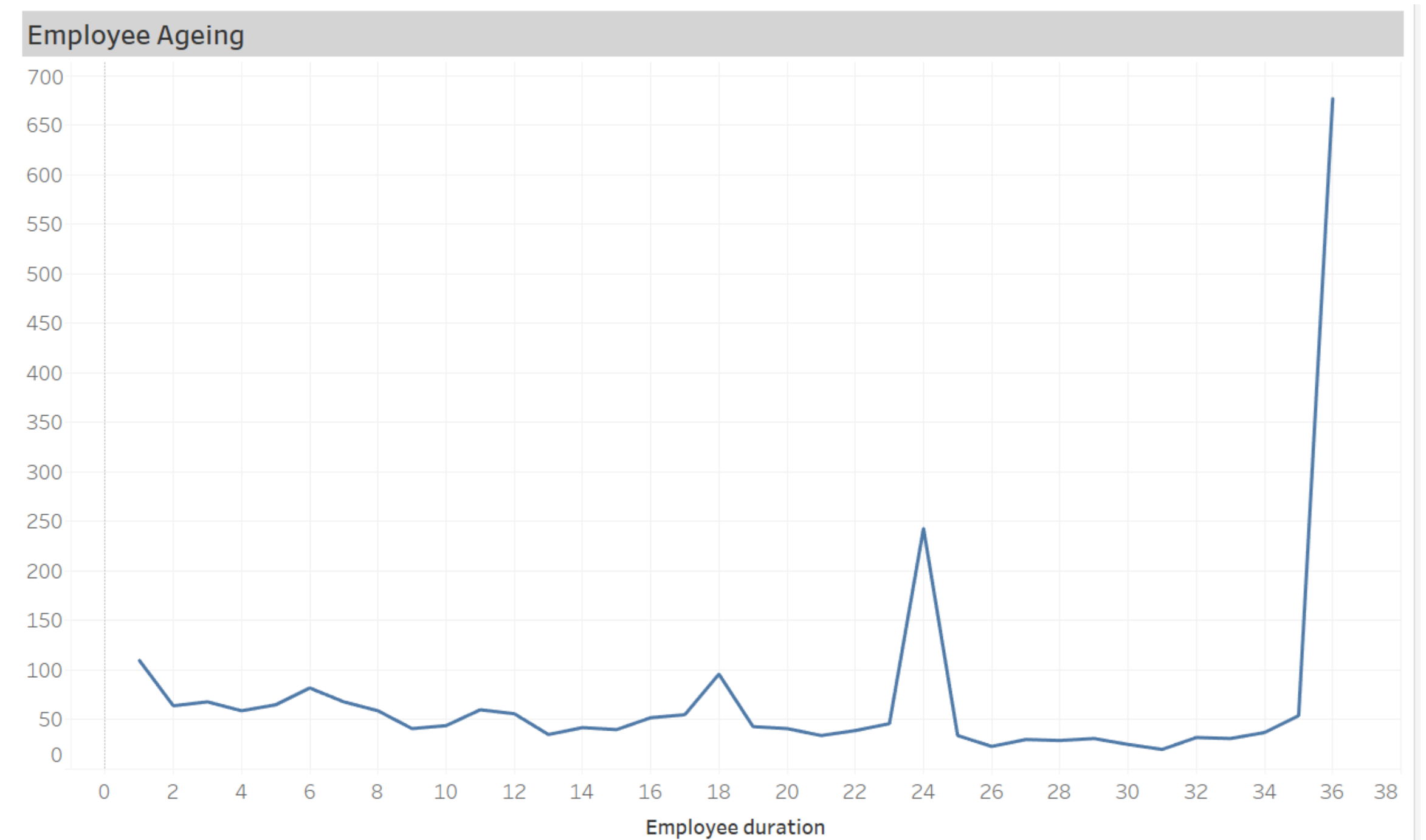
Area chart

- **Area charts** are used to summarize relationships between datasets, how individual data points relate to a whole

- The visual at right shows the monthly trend of active operations

  – **At a glance, how do you think this visual could be improved?**

Operations Count Trend
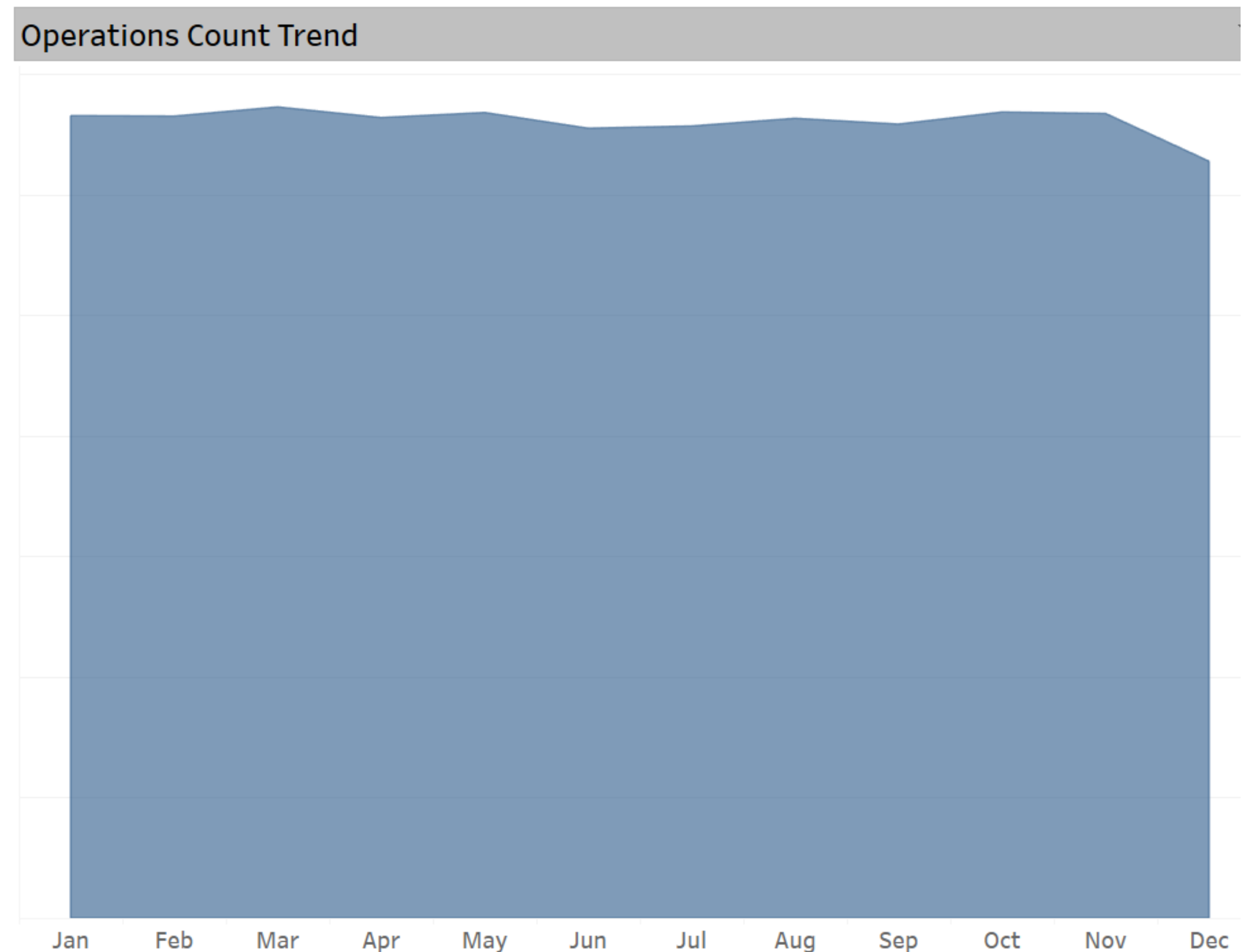
Heatmap

- **Heatmaps** visualize data in tabular format, using colored cells to show the relative magnitude of the numbers

  – When using a heatmap, it is helpful to restrict the number of different color gradations

- The visual at right shows the busiest months ranked by number of operations for each department

Depart..	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
CAN	1	5	2	4	5	9	3	6	10	7	8	11
CCB	6	5	8	1	2	7	5	4	3	9	10	11
CID	1	6	9	7	3	8	6	5	2	4	9	10
CSC	8	6	2	9	9	11	5	1	4	3	7	10
CSD	7	6	3	2	1	5	4	6	9	8	7	10
ESG	12	11	10	5	8	9	1	6	3	2	4	7
IFD	4	5	2	5	6	9	5	3	8	1	7	10
INE	9	4	2	5	1	6	6	7	8	3	5	10
INT	7	8	6	4	6	4	5	3	2	1	1	5
KIC	4	1	1	1	2	4	5	3	5	3	5	6
RES	7	4	8	3	7	3	5	6	2	1	4	3
SCL	2	7	3	5	1	10	8	9	6	4	6	11
VPC	9	8	6	3	1	11	2	7	5	4	6	10
VPS	8	9	7	6	5	5	3	4	1	3	2	3

Scatterplot

- **Scatterplots** show the type of relationship between two numeric variables

- Scatterplots are often used in scientific fields and are sometimes viewed as "complicated" to understand, but there are real-world uses as well

  - **What relationship does this scatterplot represent?**

Hours Contributed vs Employee Count for Operation

28
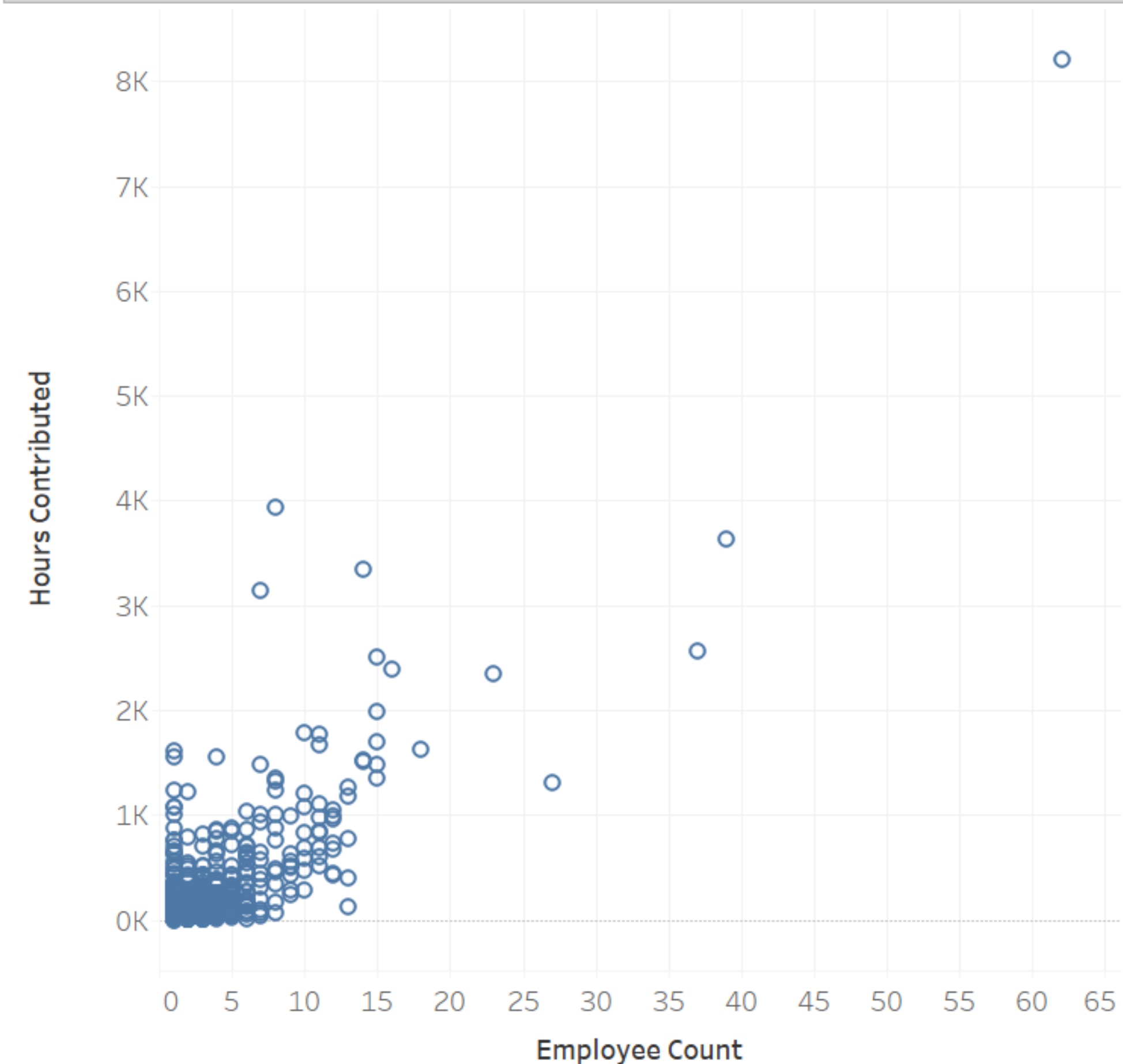
Knowledge check 1

Exercise 1
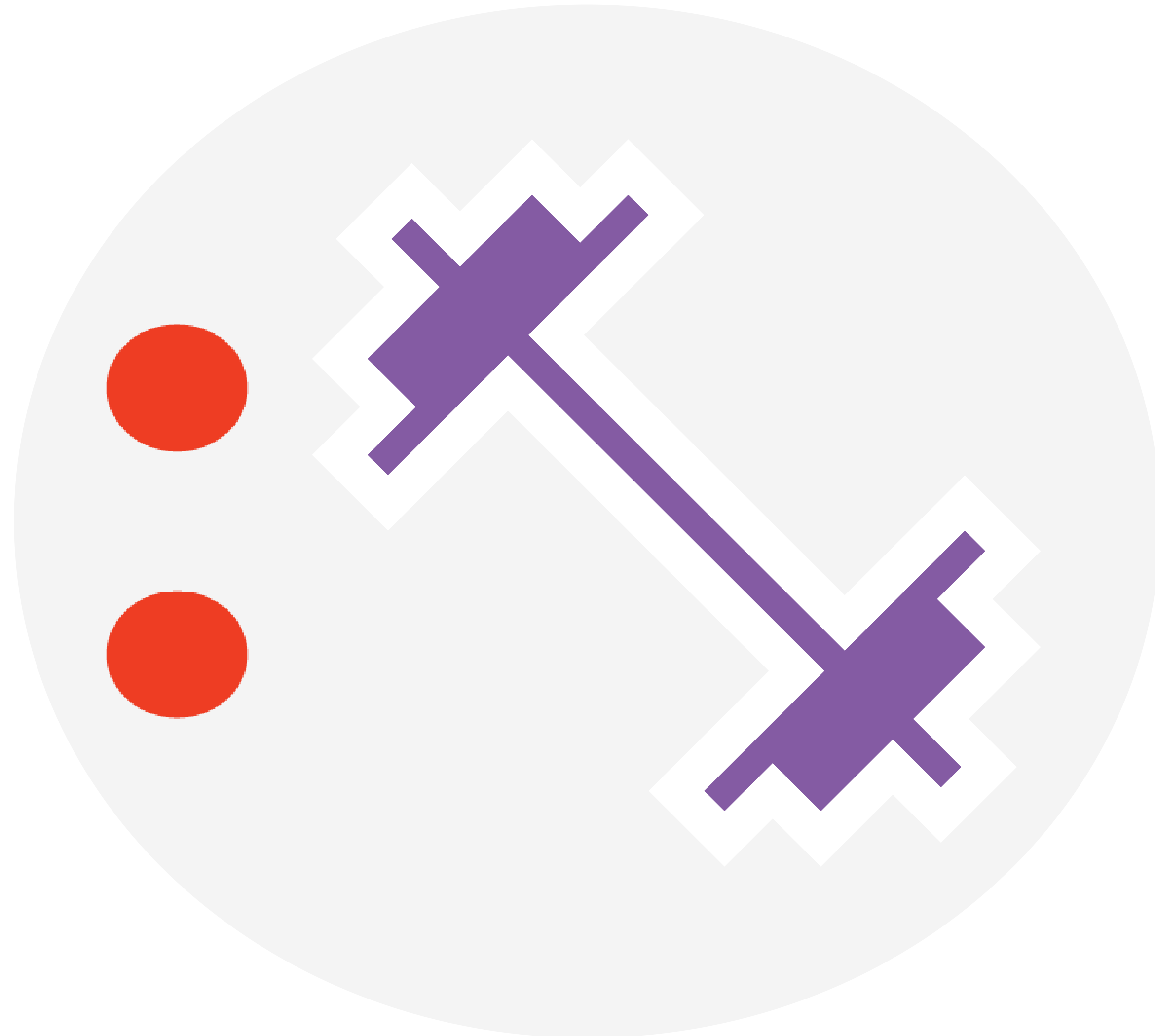
Break

Agenda

- Introduce data visualization

- Discuss chart types by data

- Discuss chart types by form

- Highlight visual building blocks

- Introduce color theory and best practices

- Explain the need for Tableau and describe its features

Visual building blocks

- Picking the right chart isn't enough; data visualization is an **art informed by science**

- We will be covering fundamental visual building blocks to help us build visuals that:

  – do not overwhelm our working memory

  – help the viewer focus on the main idea behind the visual

  – aid the viewer in recognizing familiar patterns
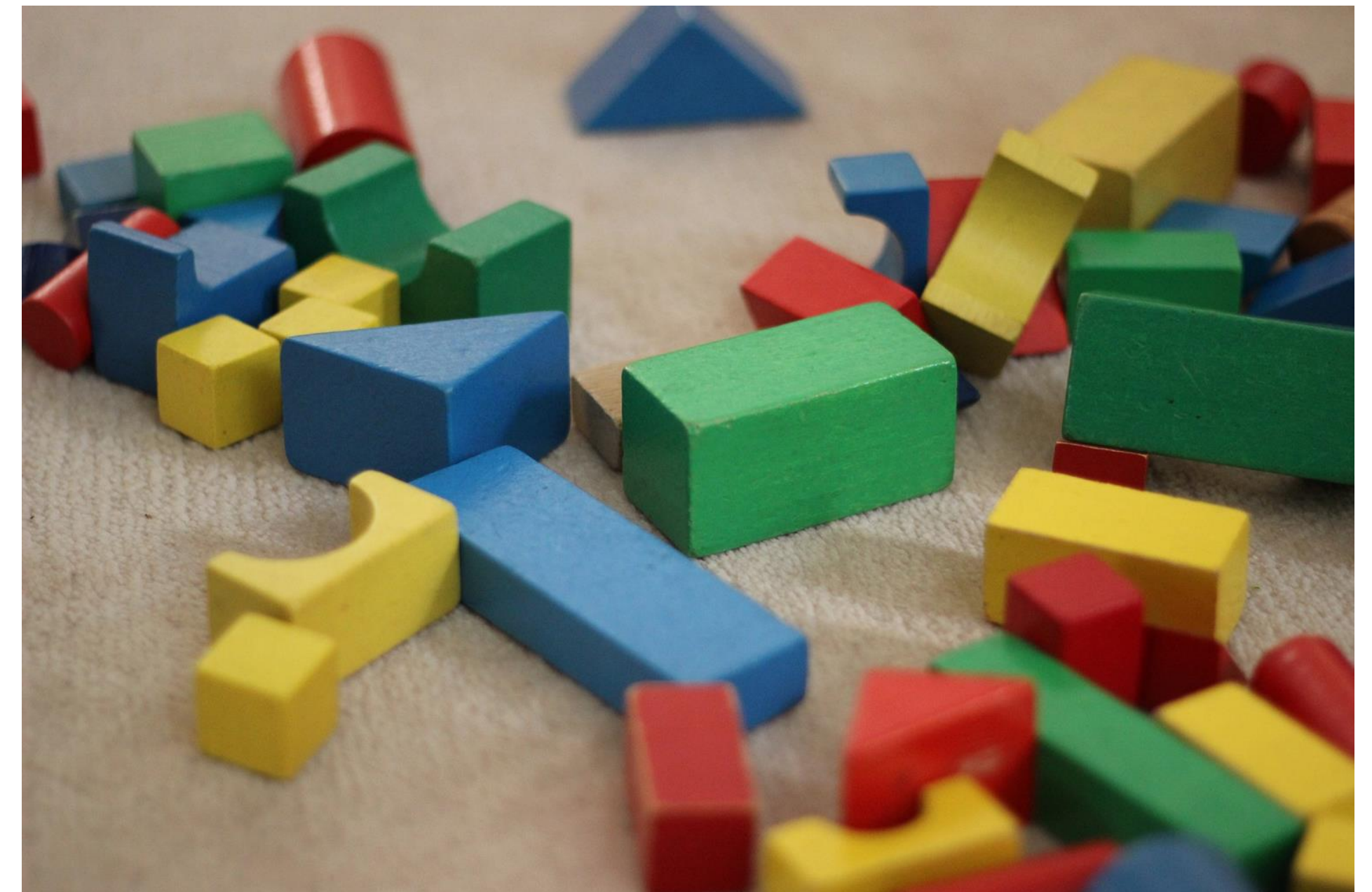
Visual design theory

- Our eyes "load" information while the brain "processes" it

- We give the most attention to what looks good and struggle when our working memory is overwhelmed

- For information to be effective, it should not provide more data than what the human brain can process

Example: buying oranges

- You want to buy oranges at a new supermarket
- Our eyes scan the layout of the supermarket, while the brain processes the various sections
- The brain then instructs the eyes to zone in on the fruit section by sending signals about how fruits look from memory
- The eyes then break the entire scanned area into parts and scan each part to spot the fruit section
- The process is repeated until oranges are located

Designing compelling visuals

- Our eyes and brains work the same way with data visualizations as they did in the oranges example

  - **Use visual clues** to make data visualizations easier for the audience

- However, every piece of information in a visualization also creates cognitive load on the viewer, asking them to use their brain power to process it

  - **Reduce visual clutter** to lower the cognitive load and help transmission of the message

Building block #1: shape

- **Shape:** Our brains group together similar-looking shapes to help us find patterns

- The scatter plot represents the hours worked and operations count for each employee

- The shapes represent all the categories of hours worked by an employee

Operation count vs. Hours

Building block #2: size

- **Size:** typically used used to communicate the intensity of the value

- The larger the size, the larger the value

- In this case, the bigger circles represent larger operation count by an employee

Operation count vs. Hours
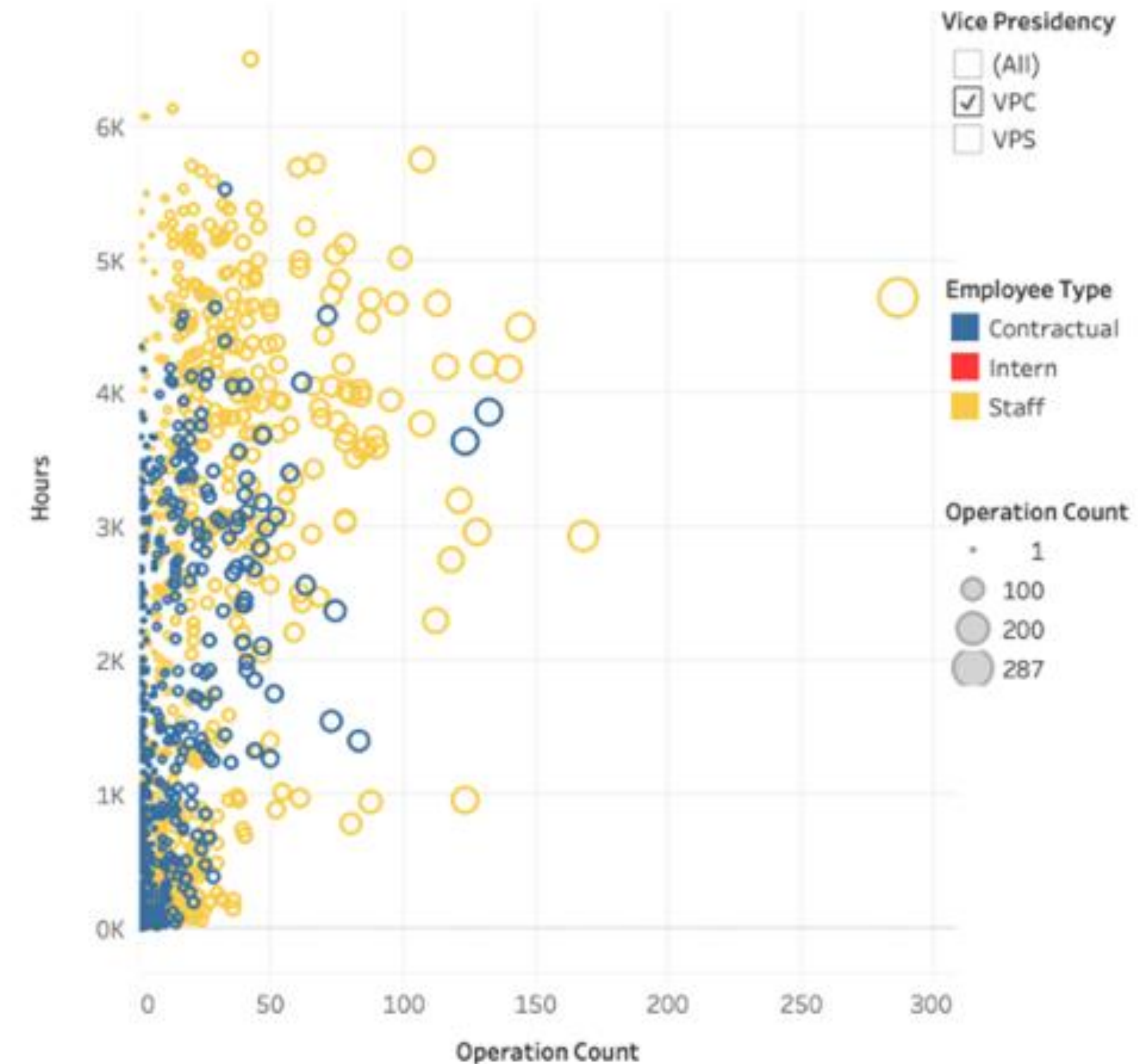
Building block #3: position

- **Position:** proximity and position signify how the data is placed in the visual, and can highlight trends, patterns, or relationships

- Sorting employees **alphabetically** vs. sorting by **number of hours** are two different ways to plot the data above with very different positioning

Building block #4: labels and legends

- **Labels:** used to better read the data and show the value of specific datapoints

- **Legends:** used to identify the size, color, or other distinguishing feature

- For maximum impact, make sure labels and legends appear in proximity to what they describe
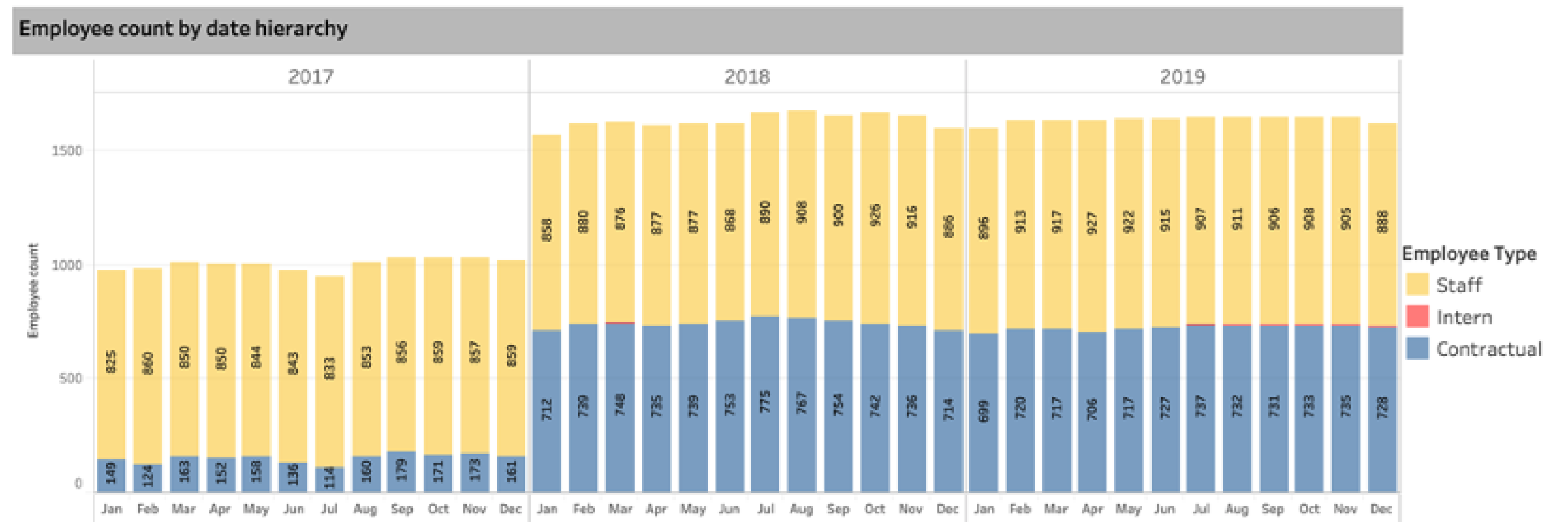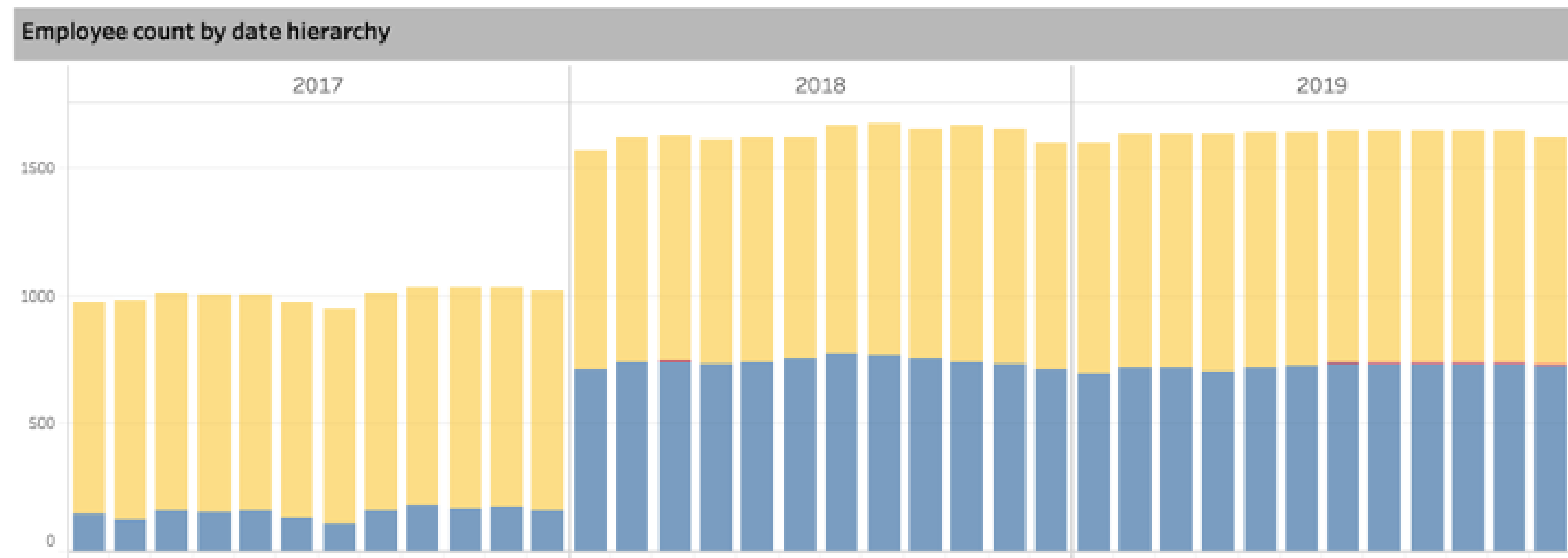
Building block #5: color

- **Color:** used to enhance and clarify a visual by highlighting, creating a scale or hierarchy, or otherwise establishing a focal point for the viewer's attention
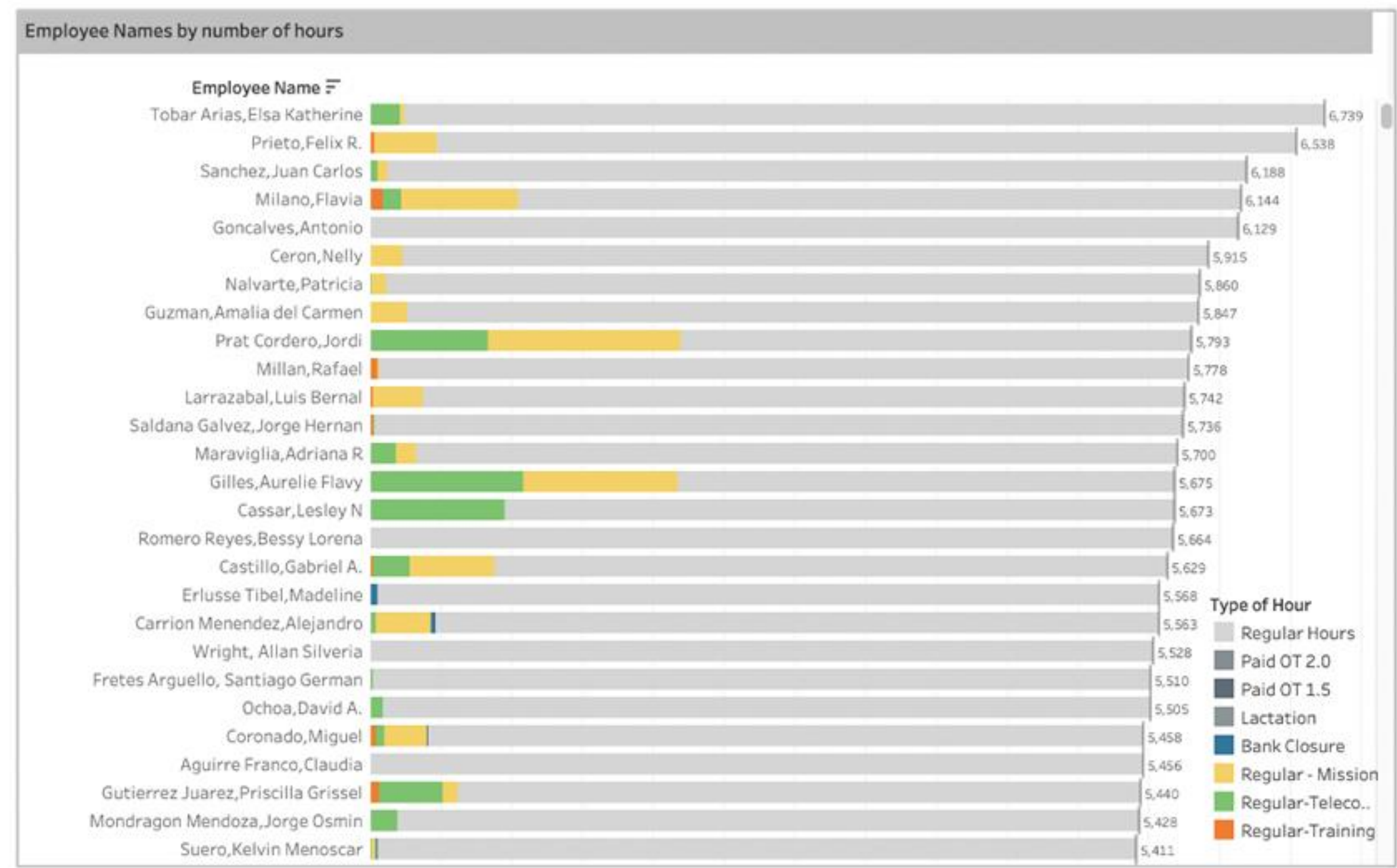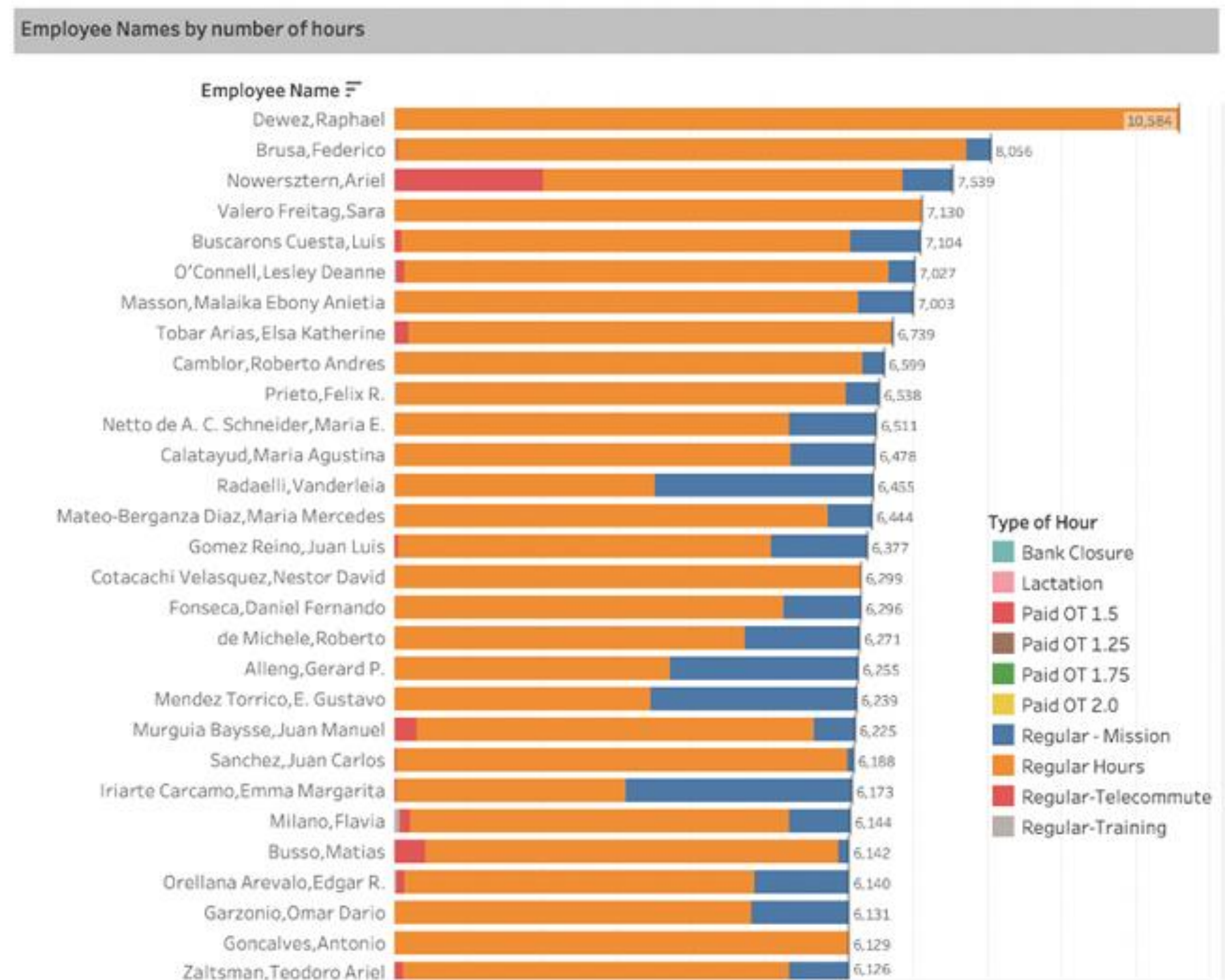
Agenda

- Introduce data visualization

- Discuss chart types by data

- Discuss chart types by form

- Highlight visual building blocks

- Introduce color theory and best practices

- Explain the need for Tableau and describe its features

Using color

- **Color** can have an outsize influence on how your visualization is interpreted:

  – Do not use too many colors; use colors sparingly

  – Strategically use color to highlight key points

  – Use a mild base color to make better contrasts

  – Use colors consistently

  – Check if you want to leverage brand colors

Arrests by Day of Week

But what is Highlighted?
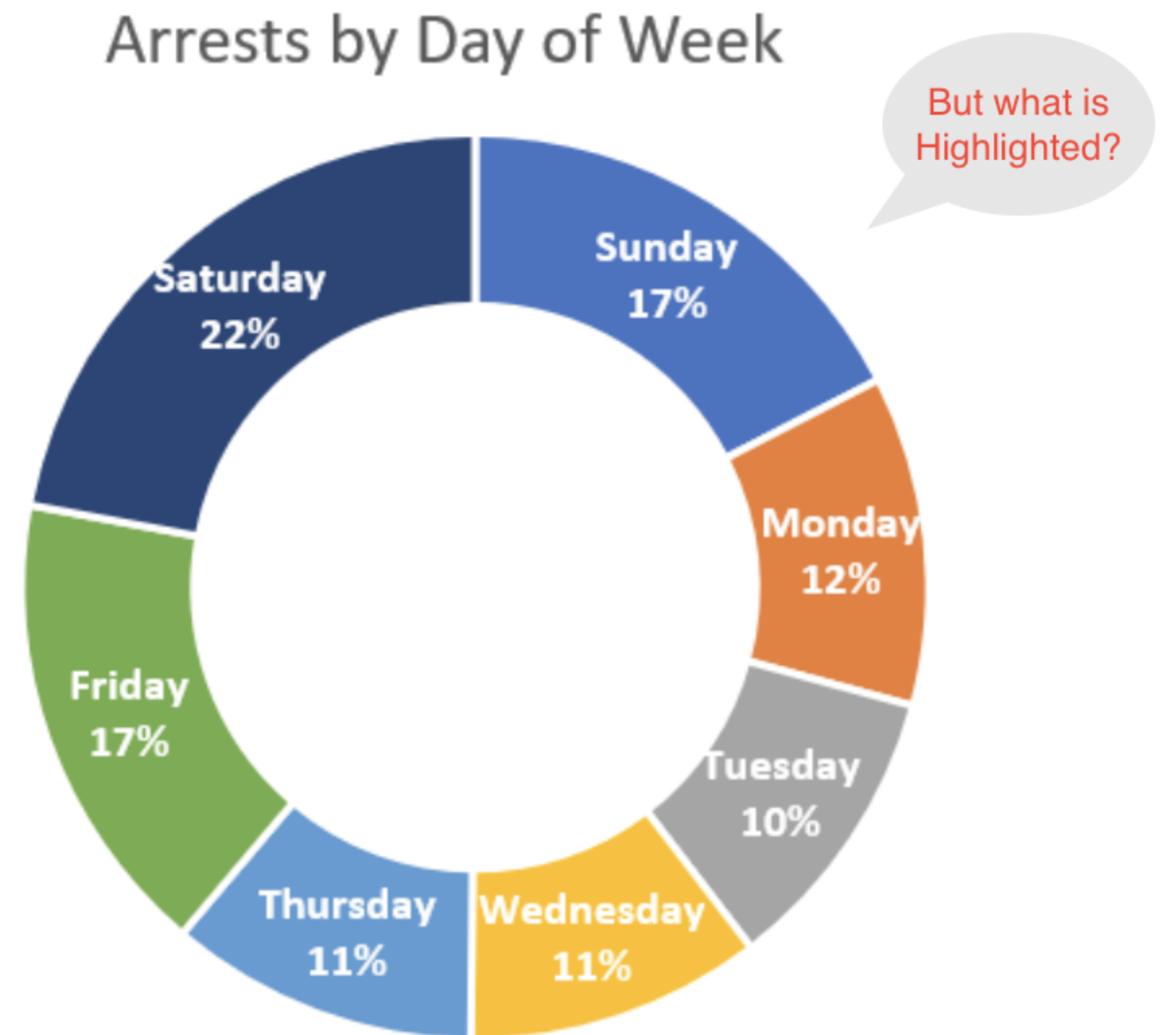
Basic color theory

- Using the **color wheel** is one way to leverage art and science to select meaningful combinations:

    - **Complementary:** two colors from the opposite side of the wheel

    - **Monochromatic:** various shades of the same color

    - **Analogous:** three similar but distinct colors

    - **Triadic:** three colors that are evenly spaced on the wheel

    - **Tetradic:** four colors that are evenly spaced
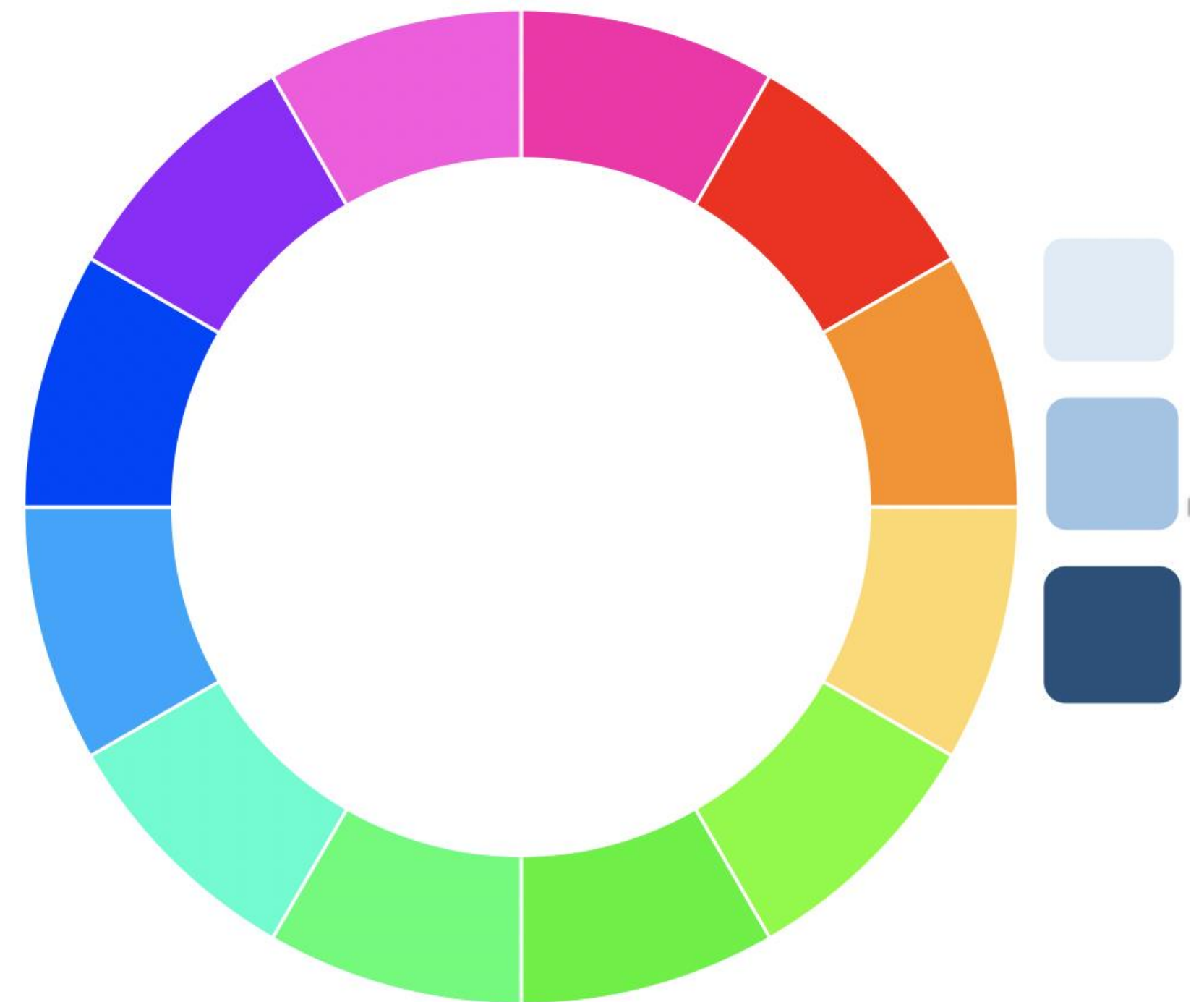
Advanced color theory

Warm and cool colors evoke different emotions
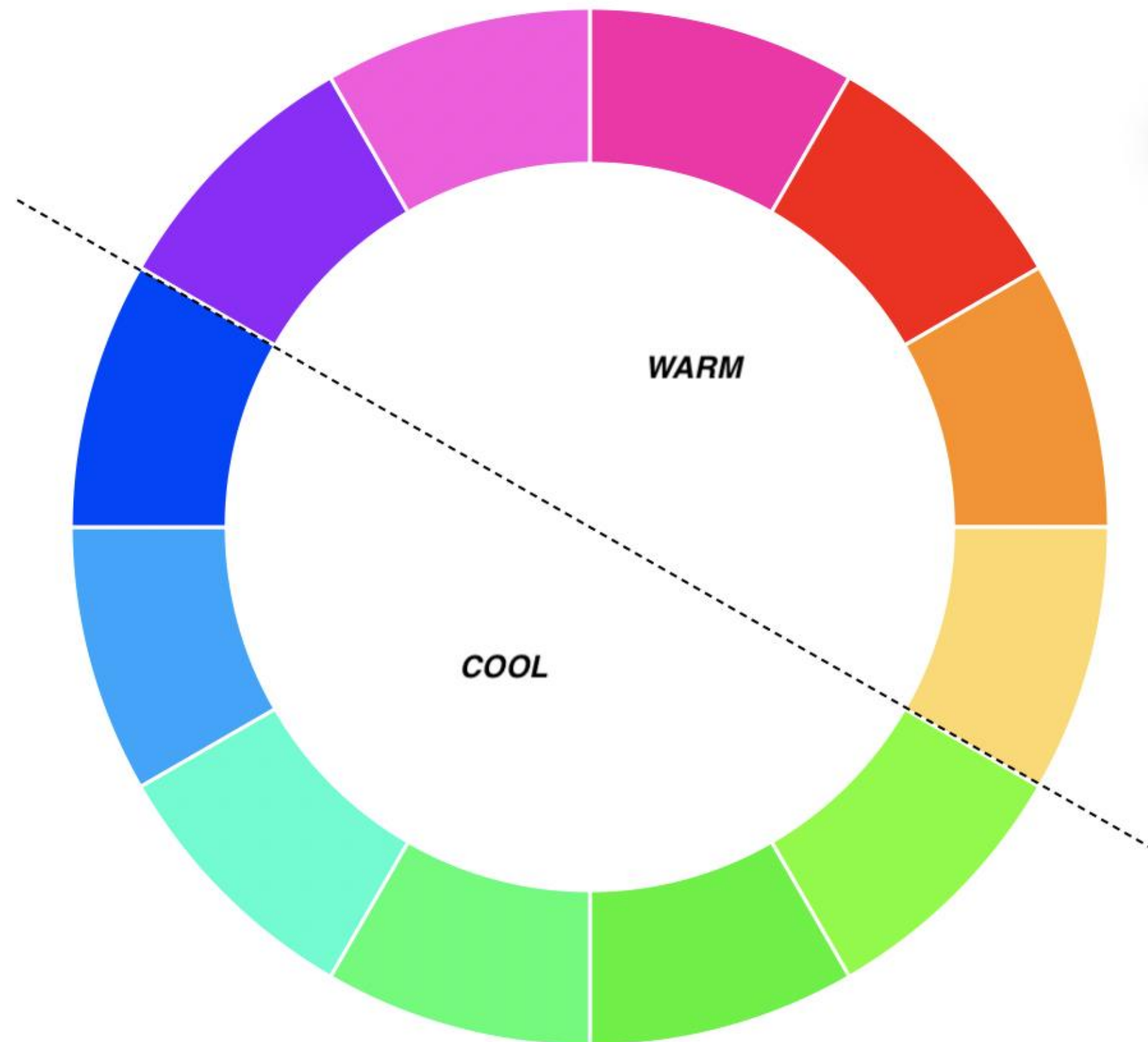
Create gradients using a base color to shade, tint, and tone
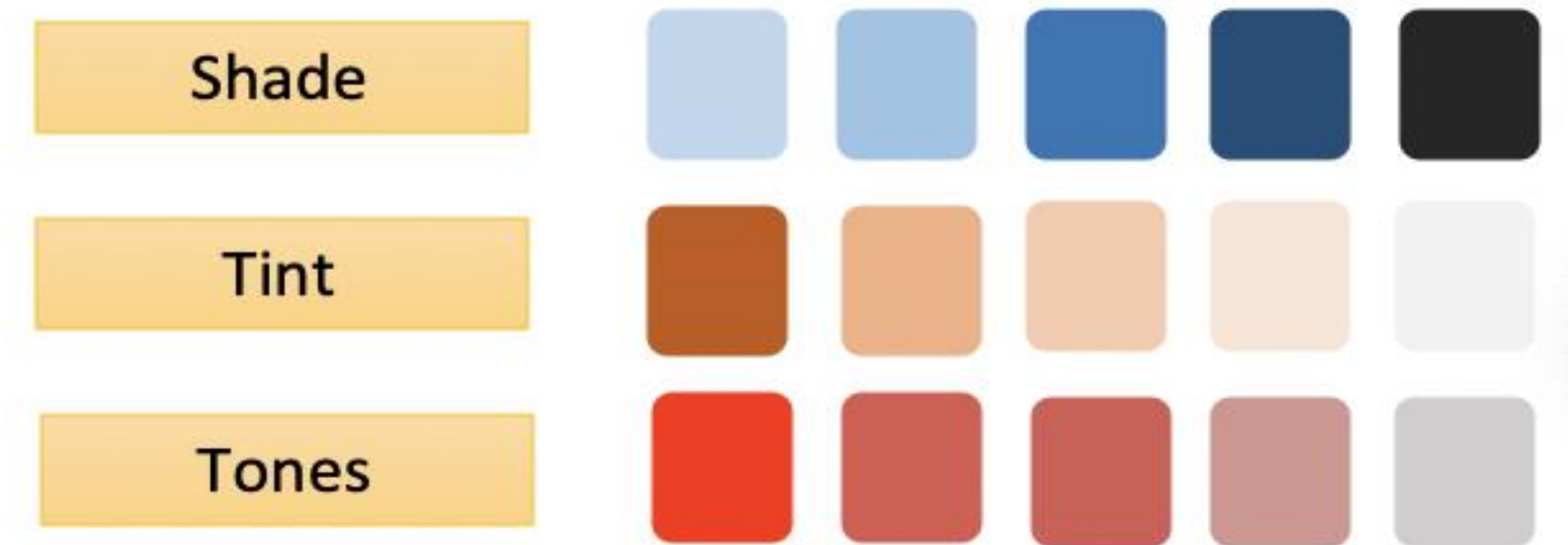
- **Shade:** *darkening* of colors by adding black to base hue
- **Tint:** *lightening* of colors by adding white to base hue
- **Tone:** created by combining black and white or using gray color

Encoding data with color

- Color schemes can even be used to represent different types of data

- There are three major ways to encode data with color in data visualization:

- **Sequential:** used to encode a value from low to high using a single color

- **Diverging:** used to encode a quantitative value with a midpoint
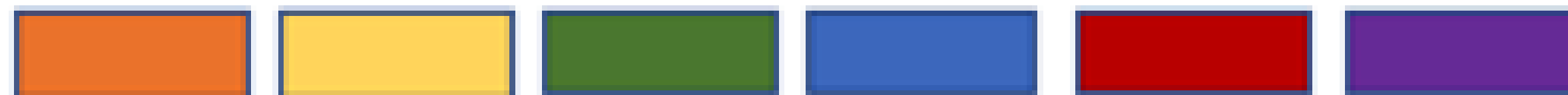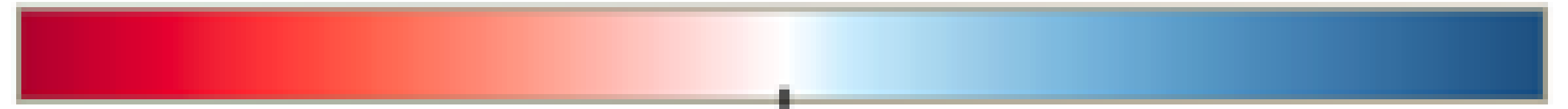
- **Categorical:** used to encode categories in no inherently quantitative order

Agenda

- Introduce data visualization

- Discuss chart types by data

- Discuss chart types by form

- Highlight visual building blocks

- Introduce color theory and best practices

- Explain the need for Tableau and describe its features

Tableau Walkthrough

- In the next few slides, we will discuss:

    - The benefits and overview of Tableau

    - How Tableau compares to Excel

    - Installation instructions

Why use Tableau?

- It offers a quick and easy way to create interactive visualizations and explore data

- It is easy to integrate with multiple data sources

- It is compatible with OS X, Windows, and Linux

- It integrates with R and Python for advanced analysis
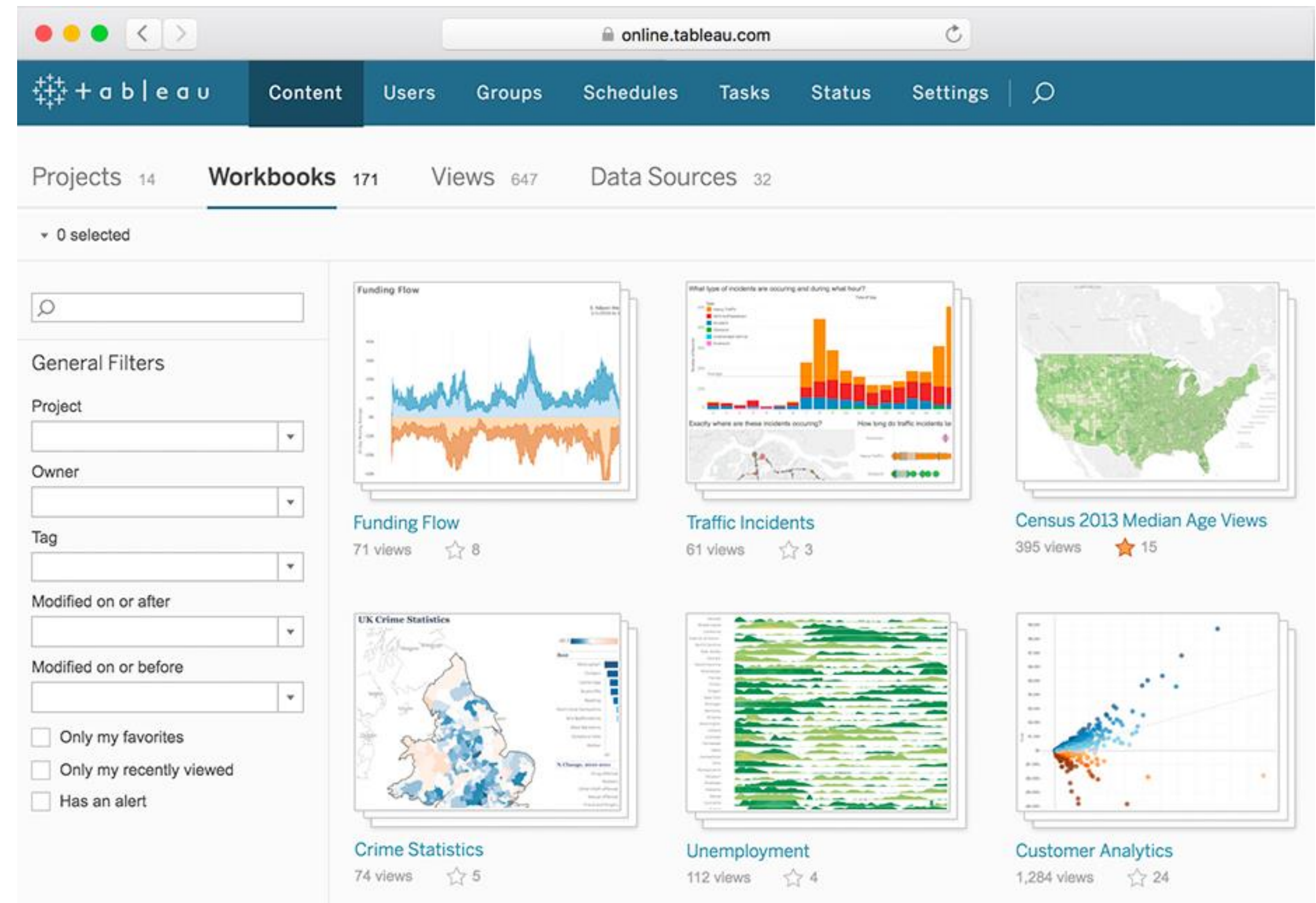
Visualizations in Tableau

- Powerful and fastest growing data visualization tool

- Interactive, reporting and analytic tool that can transform raw data into visualizations

- These visualizations are in the form of worksheets or dashboards that can provide actionable insights

- Enables to integrate and combine data from multiple sources

- Allows to easily analyze data with its intuitive drag and drop functionalities in its interface

Excel vs. Tableau

Parameters	Excel	Tableau
Purpose	Spreadsheet application used for manipulating data	Visualization tool used for data analysis
Usage	Most suitable for statistical analysis of structured data	Most suitable for quick and easy representation of large datasets, which helps resolve big data issues
Performance	Moderate speed, with no option to speed up	Moderate speed, with options to optimize and enhance the progress of an operation
Security	Relatively weak (compared with Tableau); needs regular updates	Extensive options to secure data without scripting; row-level security and permissions are built-in.

Excel vs. Tableau cont'd.

Parameters	Excel	Tableau
User Interface	Macro and Visual Basic scripting knowledge required to maximize tool potential	Tool can be used without any coding knowledge
Business Need	Best for preparing on-off reports with small datasets	Best while working with big data
Products	Bundled with MS Office Tools	Comes with different versions, such as Tableau server, cloud, and desktop
Integration	Integrates with ~60 applications	Integrates with ~250 applications
Real-time Data Exploration	You need to have an idea of where your data leads to generate insights	You are free to explore data without seeking concrete outcomes, especially with features like drill-down and data blending

Installation

- The Tableau Desktop version we have used during development is 2020.2, available here: **https://www.tableau.com/support/releases/desktop/2020.2**

Download 14-day trial of desktop

https://www.tableau.com/en-gb/products/trial
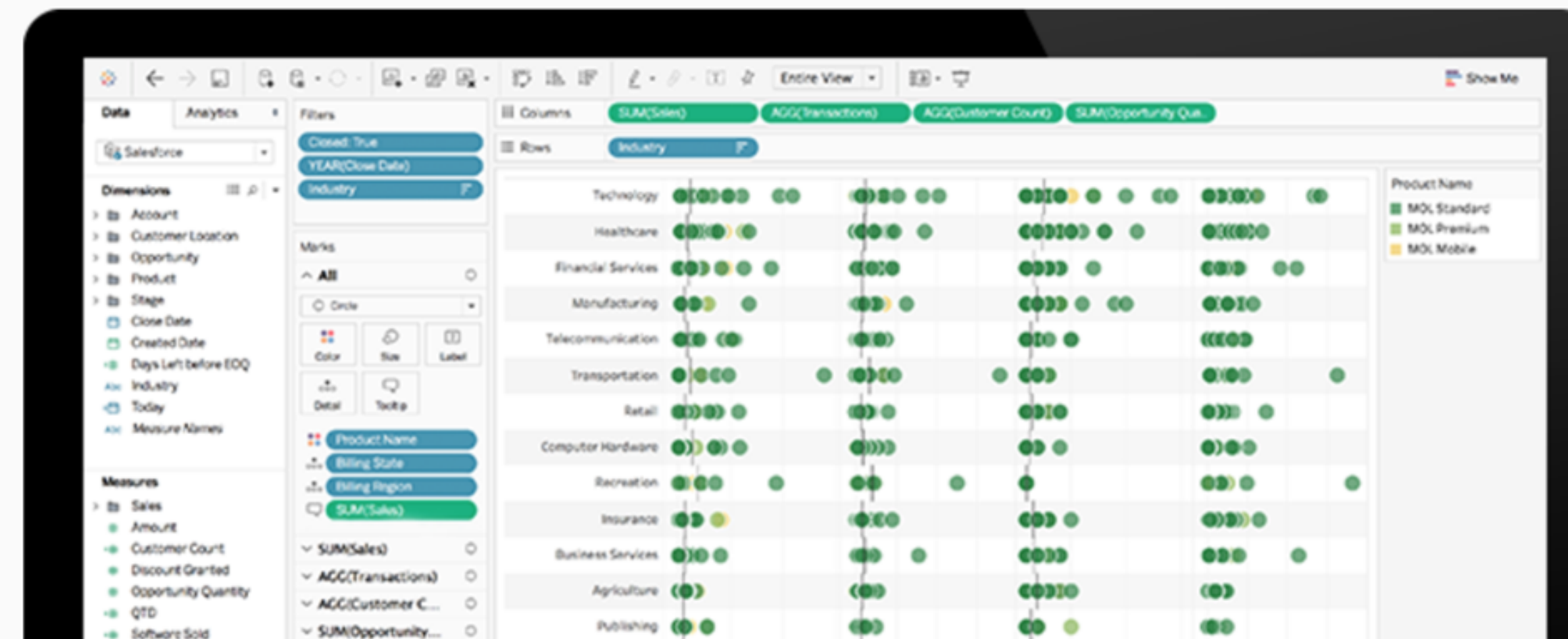
Download Tableau Public

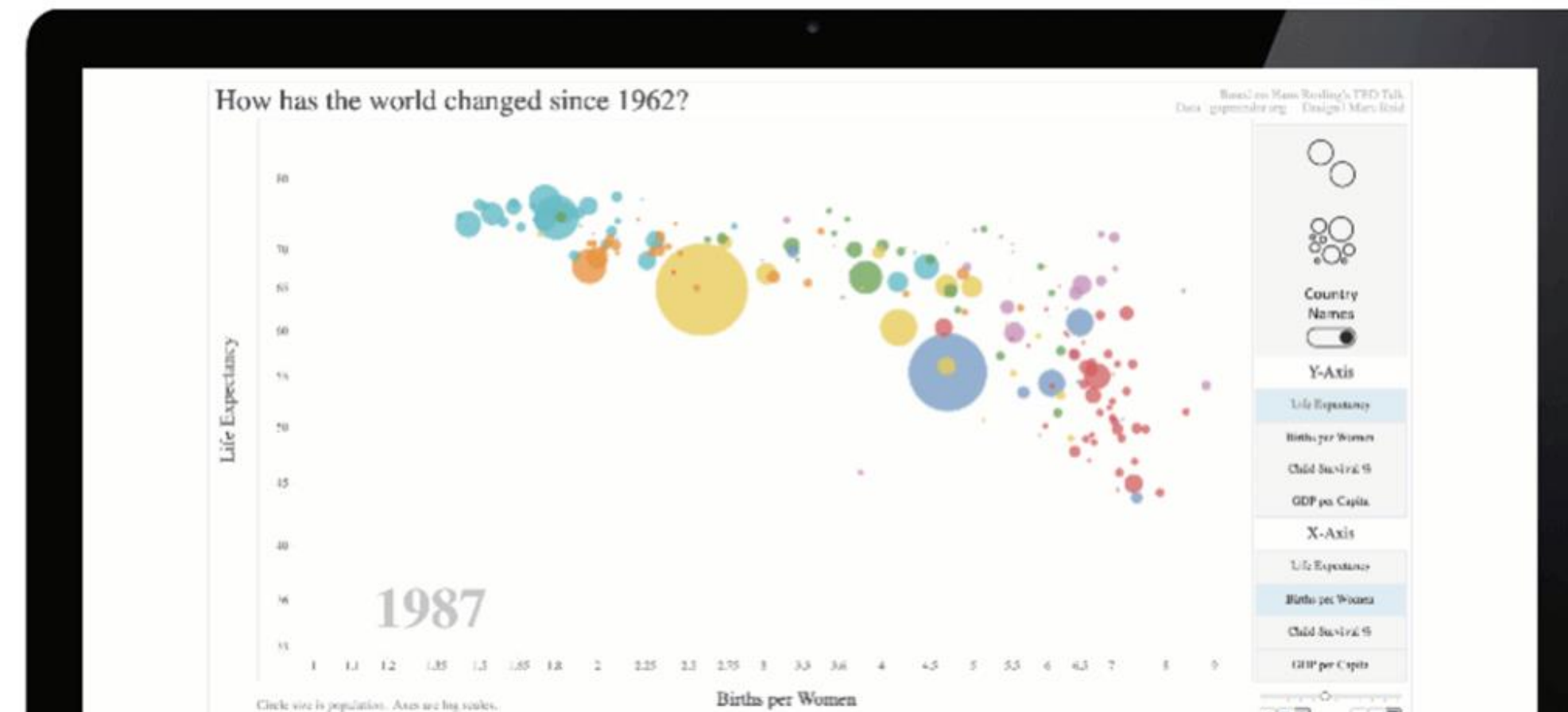- In addition to downloading Tableau, please visit the following link to download the freeware version of Tableau: **https://public.tableau.com/en-us/s/download**

Knowledge check 2

56

Exercise 2
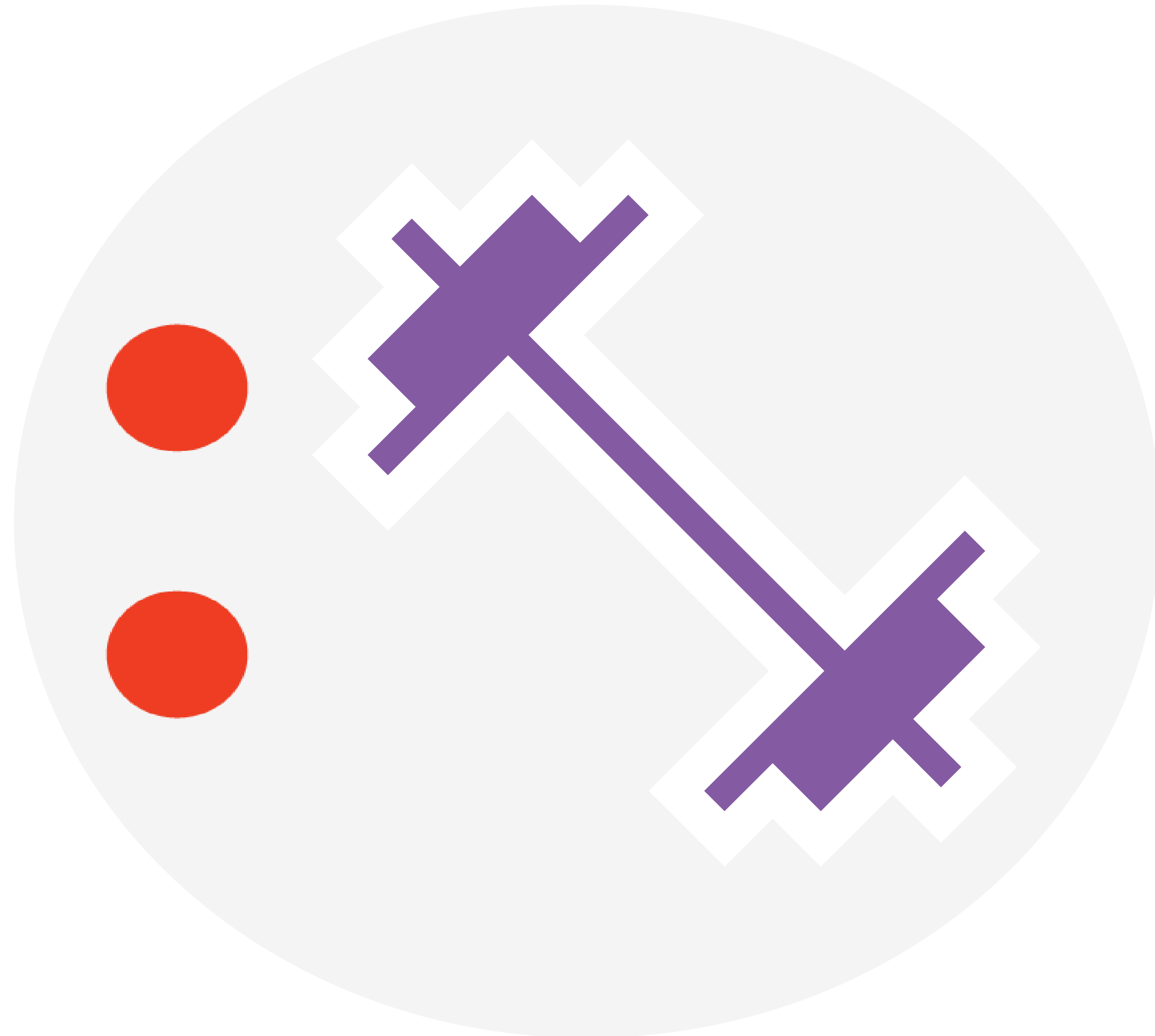
Recap: What we covered today

- Introduction to data visualization

- Discuss chart types by data and form

- Highlight visual building blocks

- Color theory and best practices

- Tableau and its uses

58

Upcoming module

In the next module, we will cover:

- Importing data into Tableau

- Key parts of the Tableau interface

- Creating charts and figures in Tableau

End of Day 1