

The background is a blue-tinted financial candlestick chart. It features several technical indicators: a green parabolic SAR curve, a green trendline labeled '61.6 %: 99.19', and three price points highlighted in green boxes: '104.19' at a peak, '86.72' at a trough, and '72.48' at a lower peak. The chart is overlaid with a white grid.

# Data Visualization and Reporting

# Learning Outcome

- Learn the fundamentals of data visualization, including using different visualizations effectively.
- Develop expertise in designing impactful visualizations with best practices and tools.

# Data Visualization

Data visualization is the graphical representation of data using visual elements such as charts, graphs, and maps.

It helps to transform raw data into visual representations that are easier to understand and interpret.

Data visualization enables the identification of patterns, trends, and relationships within large datasets.

It facilitates the communication of complex information and data-driven insights in a concise and engaging manner.

Data visualization allows for interactive exploration and analysis, empowering users to gain deeper insights and make informed decisions.



# Why create visualizations?

- Answer questions (or discover them)
- Make decisions
- Visualize data in context
- Expand memory
- Support graphical calculations
- Identify patterns
- Present arguments or tell a story
- Inspire



# Principles for data visualization

- Determine your audience. What questions will they need answered?
- Choose the right kind of chart (or other visualization) to depict the type of information you have.
- Form follows function. Focus on how your audience needs to use the data and let that determine the presentation style.
- Provide the necessary context for data to be interpreted and acted upon appropriately.
- Keep it simple. Remove any non-essential information.

# Principles for data visualization

- Choose colors carefully to draw attention while also considering accessibility issues such as contrast.
- Seek balance in your visual elements, including texture, color, shape, and negative space.
- Use patterns (of chart types, colors, or other design elements) to identify similar types of information.
- Use proportion carefully so that differences in design size fairly represent differences in value.
- Be skeptical. Ask yourself questions about what data is *not* represented and what insights might therefore, be misinterpreted or missing.

This image displays a collection of 25 different infographic elements, all rendered in a dark blue color scheme. The elements are arranged in a grid-like fashion and include:

- Donut Charts:** Four donut charts showing percentages (80%, 75%, 50%, 25%) and two donut charts with labels (ONE, TWO, THREE, FOUR).
- Bar Charts:** A bar chart with four bars (ONE, TWO, THREE, FOUR) showing percentages (50%, 80%, 35%, 70%), a bar chart with six bars (JAN, FEB, MAR, APR, MAY, JUN) showing values (1, 2, 3, 4, 5, 6), a bar chart with six bars (JAN, FEB, MAR, APR, MAY, JUN) showing values (1, 2, 3, 4, 5, 6), and a bar chart with six bars (JAN, FEB, MAR, APR, MAY, JUN) showing values (1, 2, 3, 4, 5, 6).
- Line Graphs:** A line graph with six data points (10, 20, 30, 40, 50, 60) showing values (10, 15, 20, 25, 30, 35), a line graph with six data points (10, 20, 30, 40, 50, 60) showing values (10, 15, 20, 25, 30, 35), and a line graph with six data points (10, 20, 30, 40, 50, 60) showing values (10, 15, 20, 25, 30, 35).
- Maps:** A map of the United States and a map of Europe.
- Icons:** A set of icons representing people (men and women) and a set of icons representing different categories (ONE, TWO, THREE, FOUR).
- Other Elements:** A Venn diagram with three overlapping circles (A, B, C), a hexagon with a smaller hexagon inside, a pyramid with five levels (1, 2, 3, 4, 5), and a set of three horizontal bars (ONE, TWO, THREE).

# Data literacy

- Data literacy is the ability to read, work with, analyze and communicate with data.
- It's a skill that empowers all levels of workers to ask the right questions of data and machines, build knowledge, make decisions, and communicate meaning to others.



# Tools for data visualization

- Microsoft Power BI
- Python
- R Programming
- Tableau
- Qlik Sense
- Looker
- Zoho Analytics
- Google Charts
- SAS
- Etc.,

## Geospatial visualization

- These visualizations focus on the relationship between data and its physical location to create insight.
- Maps are the primary focus of geospatial visualizations.



Source: Tableau

# Text Visualization

- Text visualization is the technique of using graphs, charts, or word clouds to showcase written data in a visual manner.



Source: Tableau

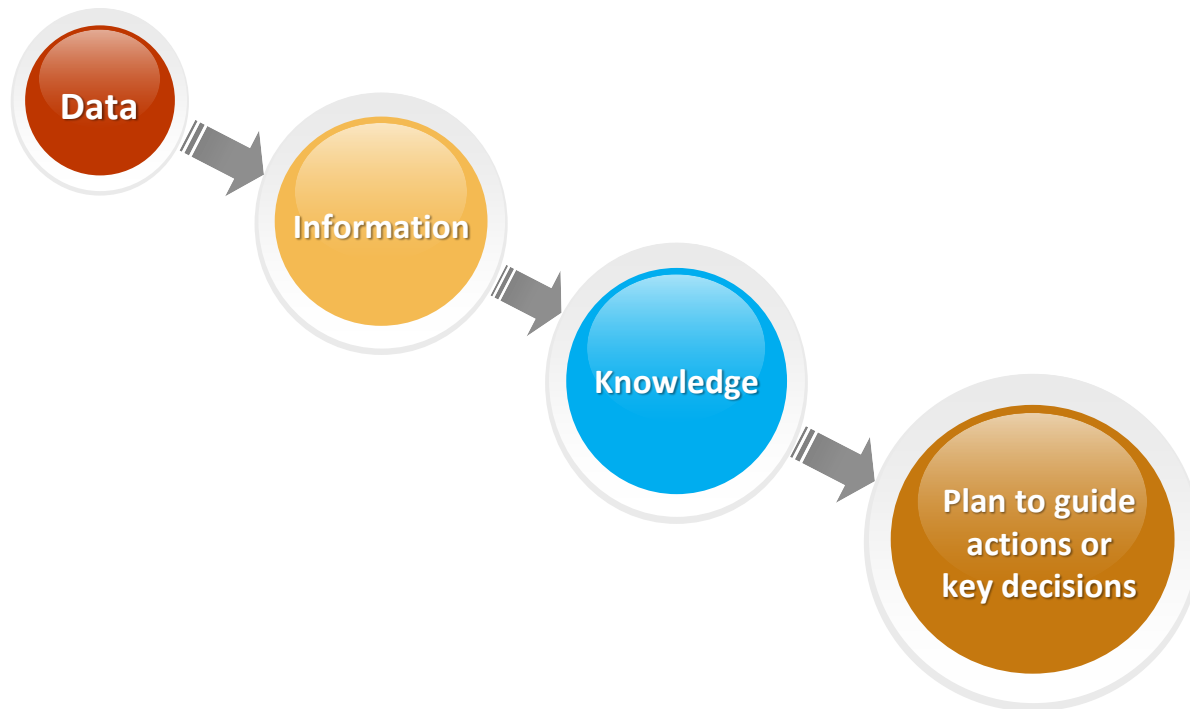
# Interactive visualization

Interactive data visualization refers to the use of software that enables direct actions to modify elements on a graphical plot.



# Business Intelligence

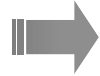
**Moving from data to action....**







Facts and statistics collected together for reference or analysis.



Facts provided or learned about something or someone.



Information and skills acquired through experience or education; the theoretical or practical understanding of a subject.



A plan of action or policy designed to achieve a major or overall goal.

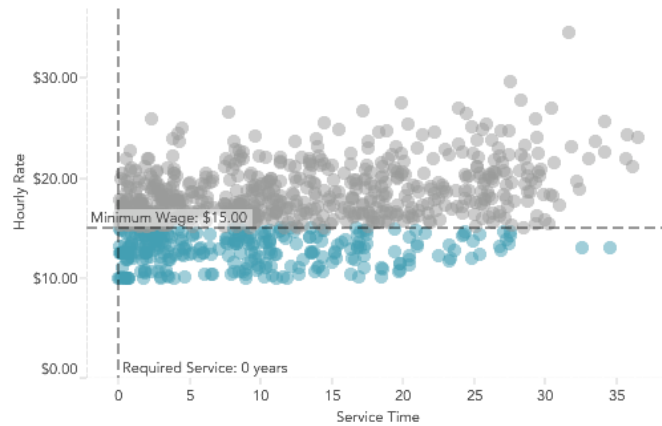
# Use of Color in Data Visualization

# Objectives

- Learn how to use of color in data visualization
- Understand color vision deficiency
- Learn options for designing colorblind-friendly data visualizations

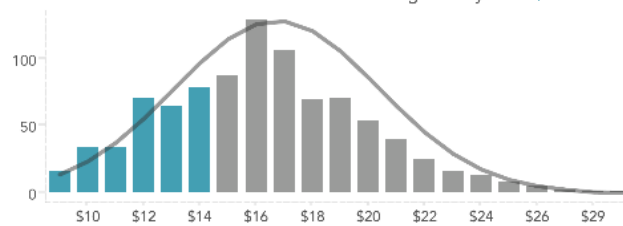
# What-If Analysis: Impact of Minimum Wage

Proposed Minimum Wage:  Required Service:

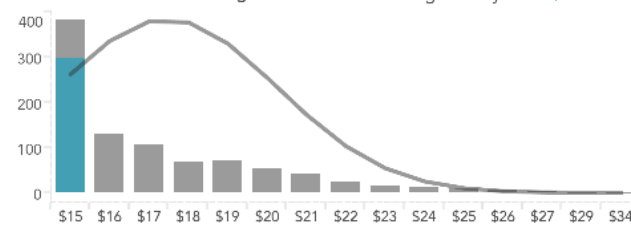


	Dollar Impact of Minimum Wage: \$1,788,805	Employees Below Minimum Wage: 296
Services	\$812,378	131
Facilities	\$730,562	93
Marketing	\$84,479	23
Logistics	\$50,630	15
Supply Chain	\$37,987	13
Legal	\$27,191	4
Operations	\$12,265	3
Customer Service	\$10,923	4
Information Technology	\$8,760	3
Human Resources	\$7,539	4
Purchasing	\$4,272	1
Research & Development	\$1,818	2

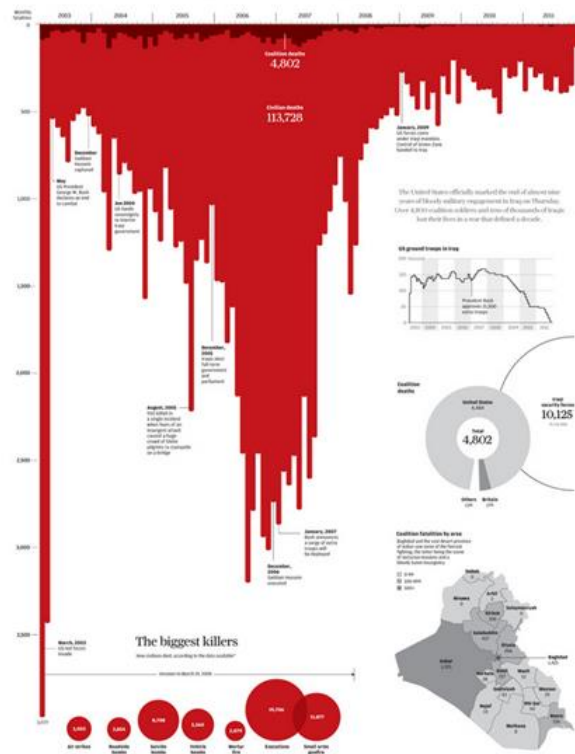
Current Distribution



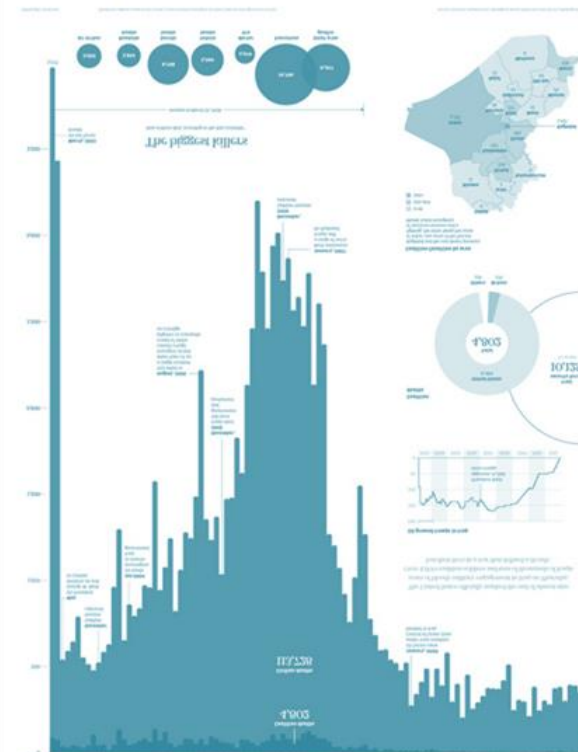
Distribution with Minimum Wage



## Iraq's bloody toll



## Iraq: Deaths on the Decline





# What meaning does color bring to the presentation?

## Earthtones

Gentle browns, blues. Calming, sinks into the page



## Cool

Soothing, restful, calm



## Unnatural colors

Alarming, unnerving, draws attention.



## Warm

Optimistic, active, vivid



## Increasing color intensity

Increasing saturation and brightness draws the eye and means the point is more important



Source: *Juice Analytics Whitepaper (part 3)*

# THE USE OF COLOR IN DATA VISUALIZATION

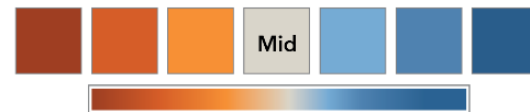
## SEQUENTIAL

color is ordered from low to high



## DIVERGING

two sequential colors with a neutral midpoint



## CATEGORICAL

contrasting colors for individual comparison



## HIGHLIGHT

color used to highlight something



## ALERT

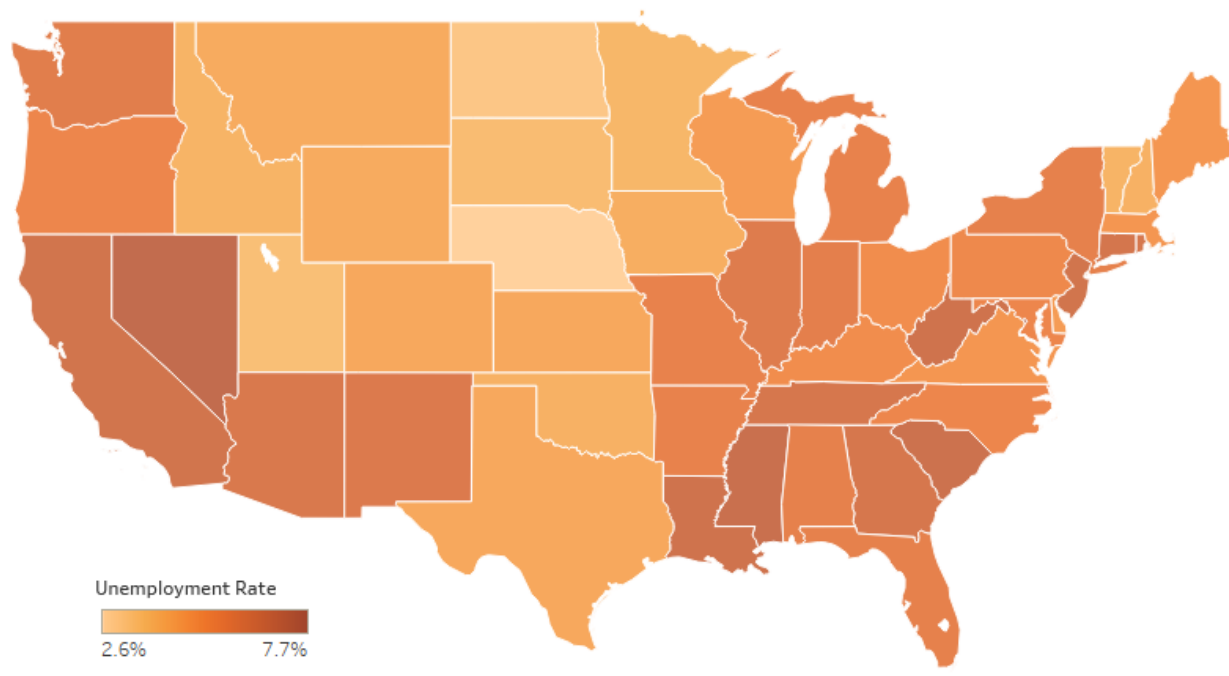
color used to get reader's attention



Source: *The Big Book of Dashboards* (Figure 1.16)

# Sequential Color

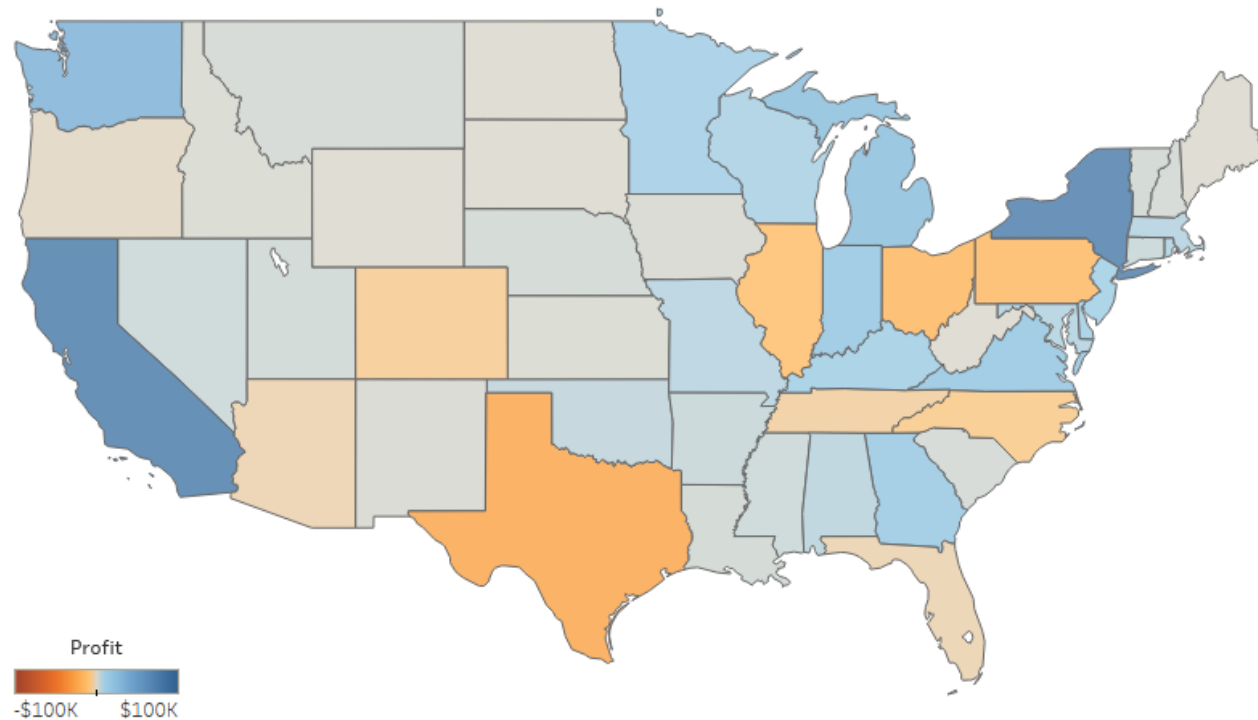
Unemployment Rate by State



Source: *The Big Book of Dashboards* (Figure 1.17)

# Diverging Color

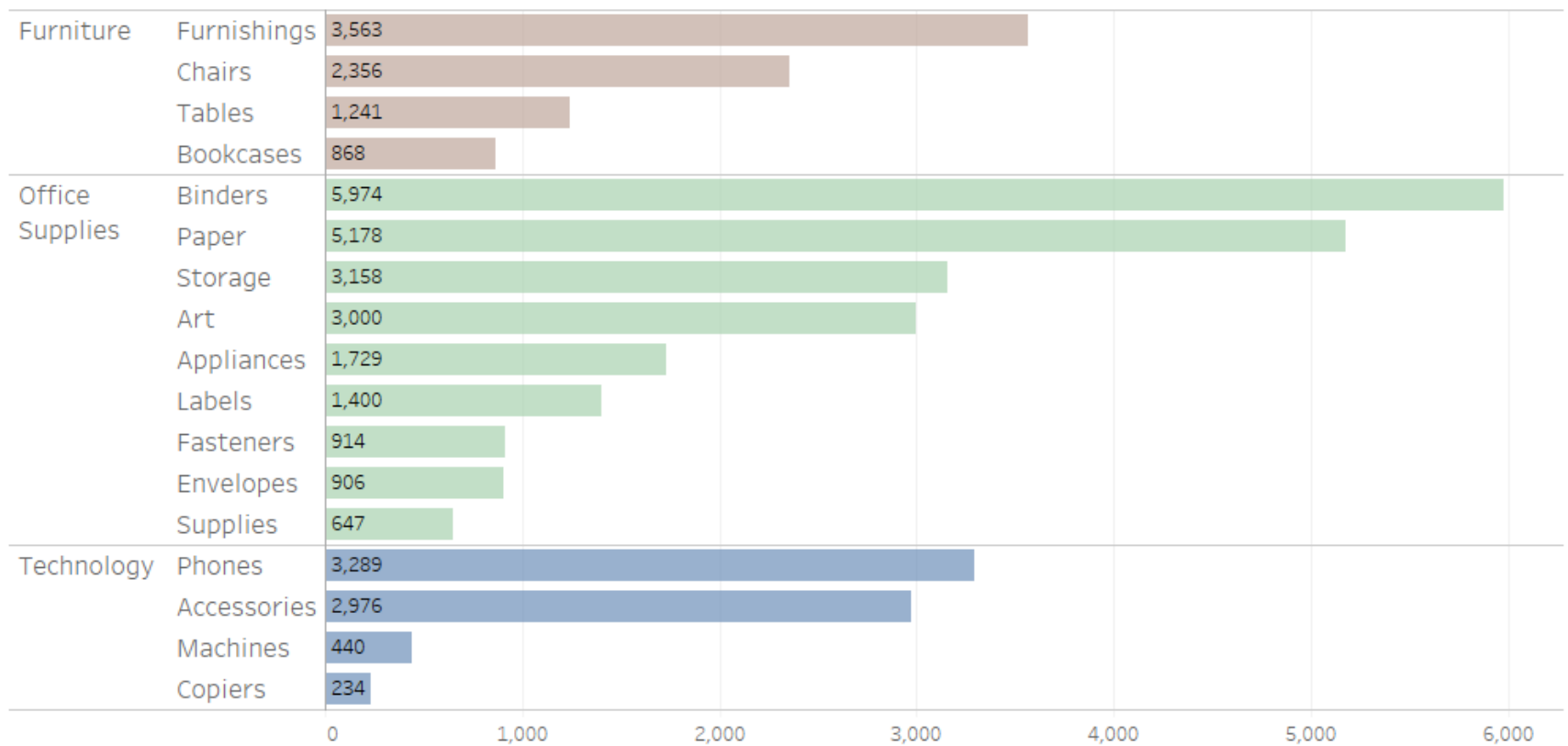
Profit by State



Source: *The Big Book of Dashboards* (Figure 1.19)

# Categorical Color

Quantity by Category and Subcategory

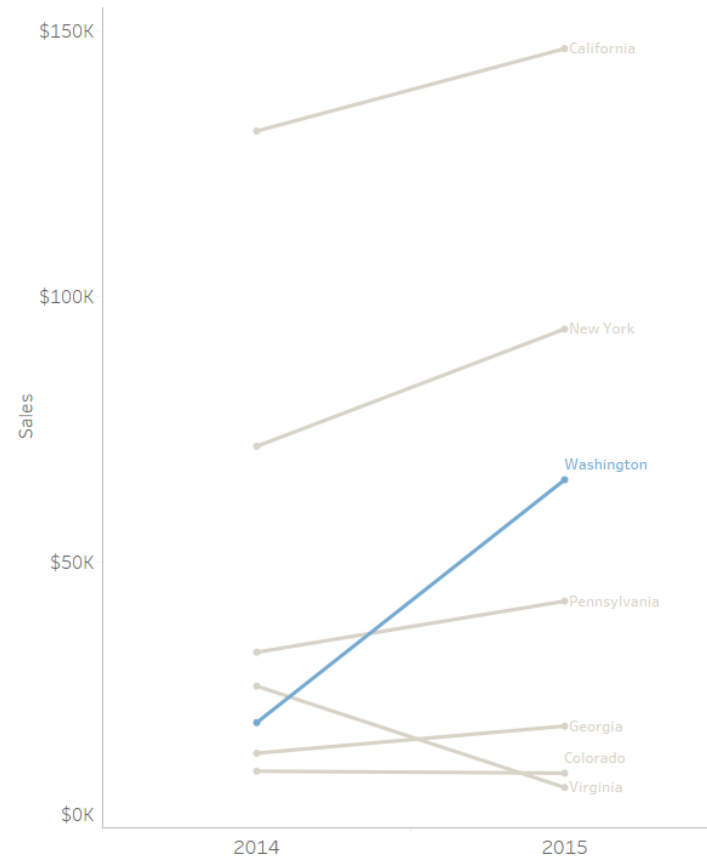


Source: *The Big Book of Dashboards* (Figure 1.20)



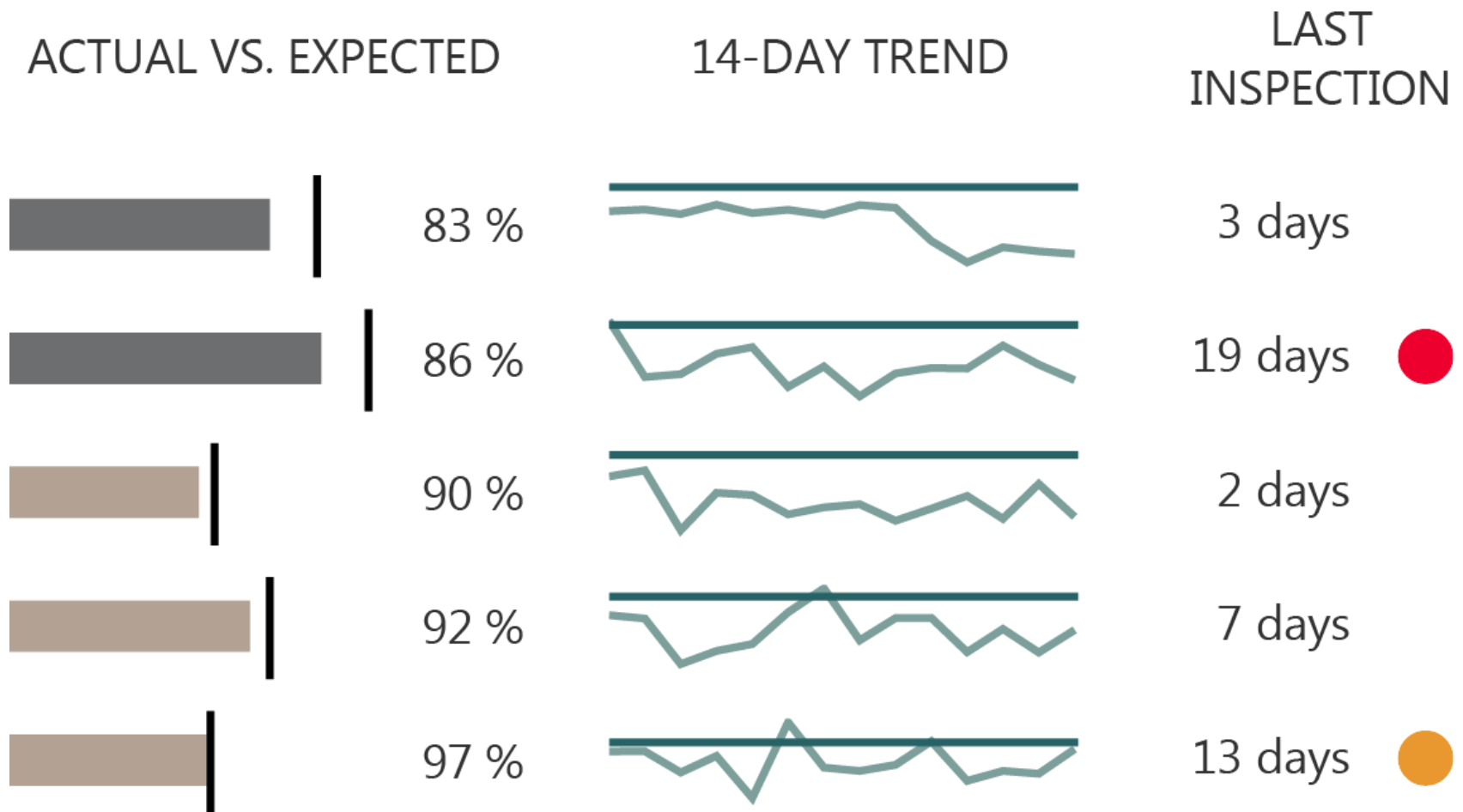
# Highlight Color

Sales by State, 2014-2015



Source: *The Big Book of Dashboards* (Figure 1.21)

# Alerting Color

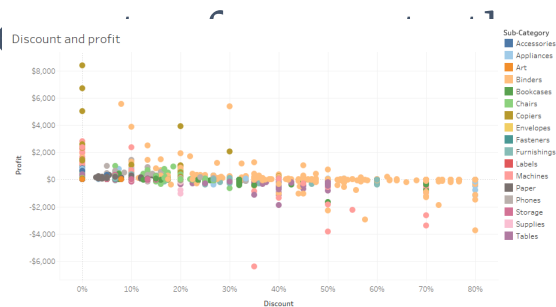


Source: *The Big Book of Dashboards* (Figure 1.22)

- Accessories
- Appliances
- Art
- Binders
- Bookcases
- Chairs
- Copiers
- Envelopes
- Fasteners
- Furnishings
- Labels
- Machines
- Paper
- Phones
- Storage
- Supplies
- Tables

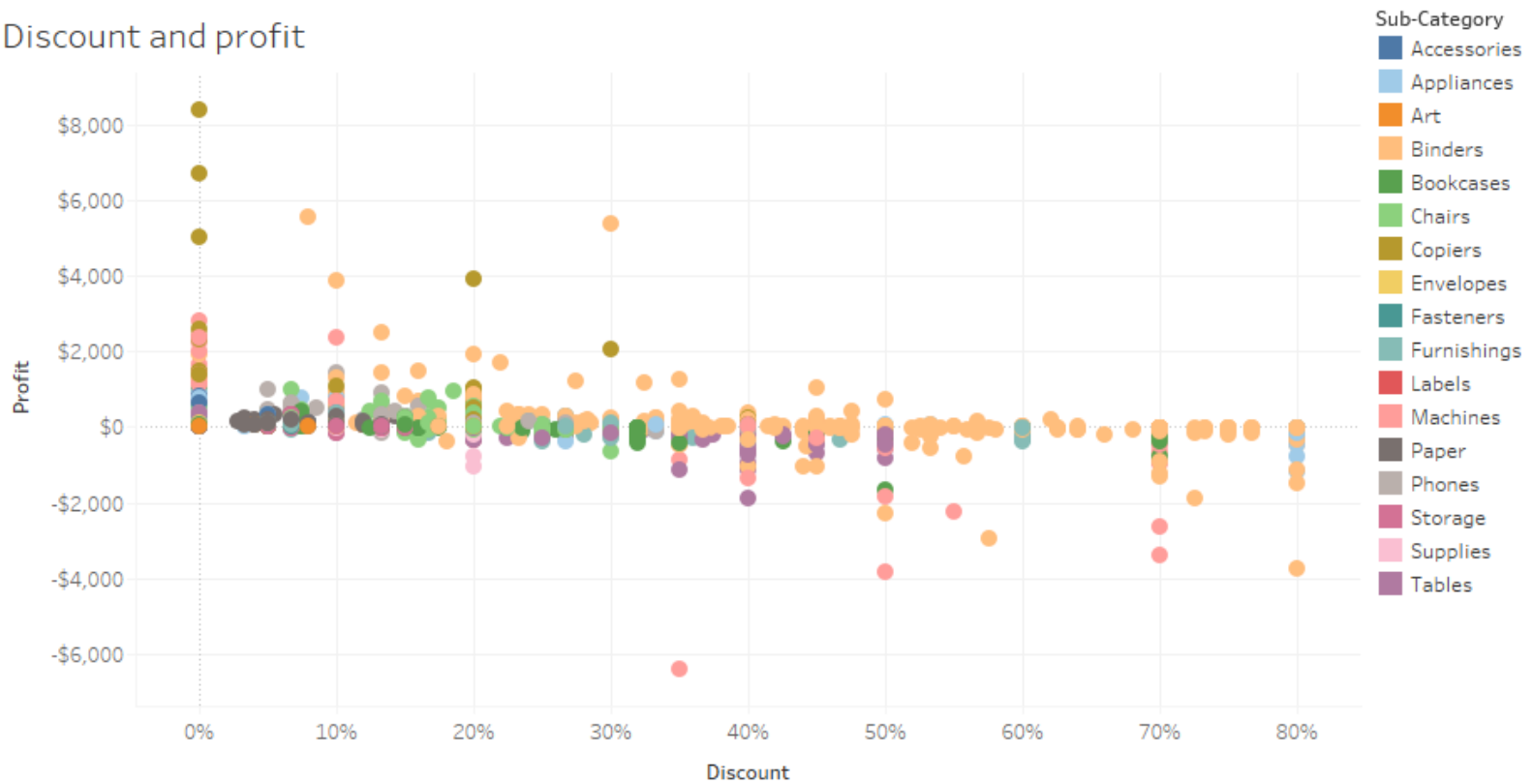
## Too Much Color

- Short-term Memory = “small chunks of information”
- Requires reusing the same or similar color
- Requires frequent color change

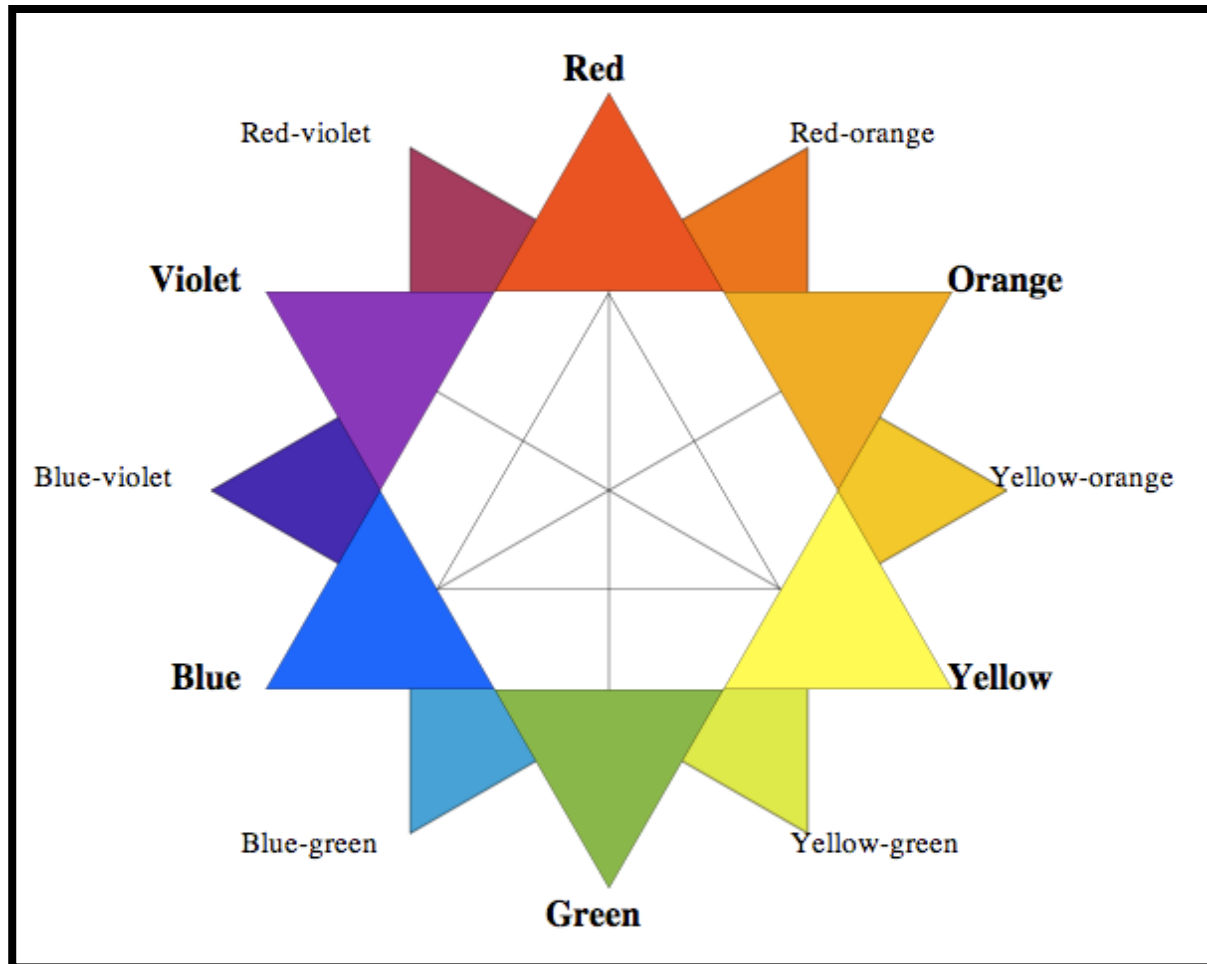


# Too Much Colors

Discount and profit

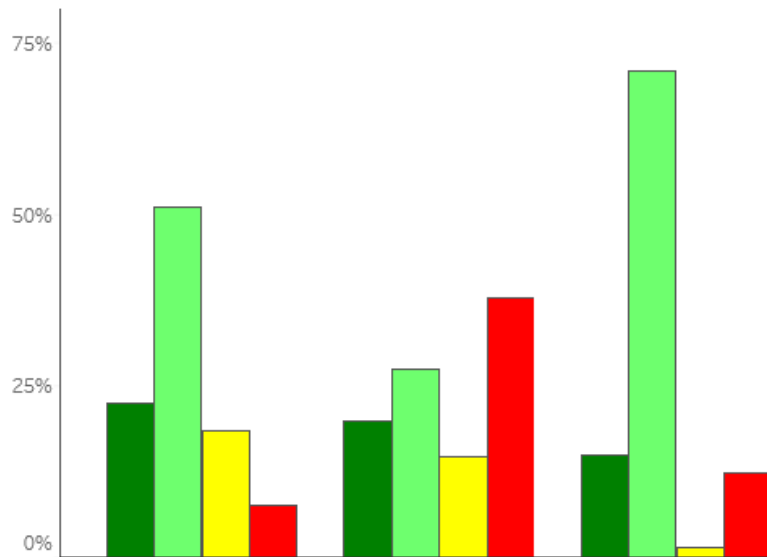


# Color

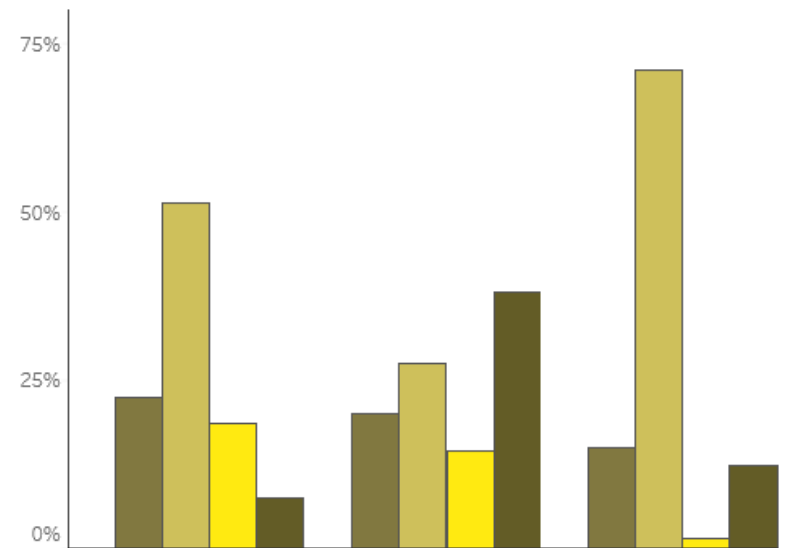




Traffic Light Colors

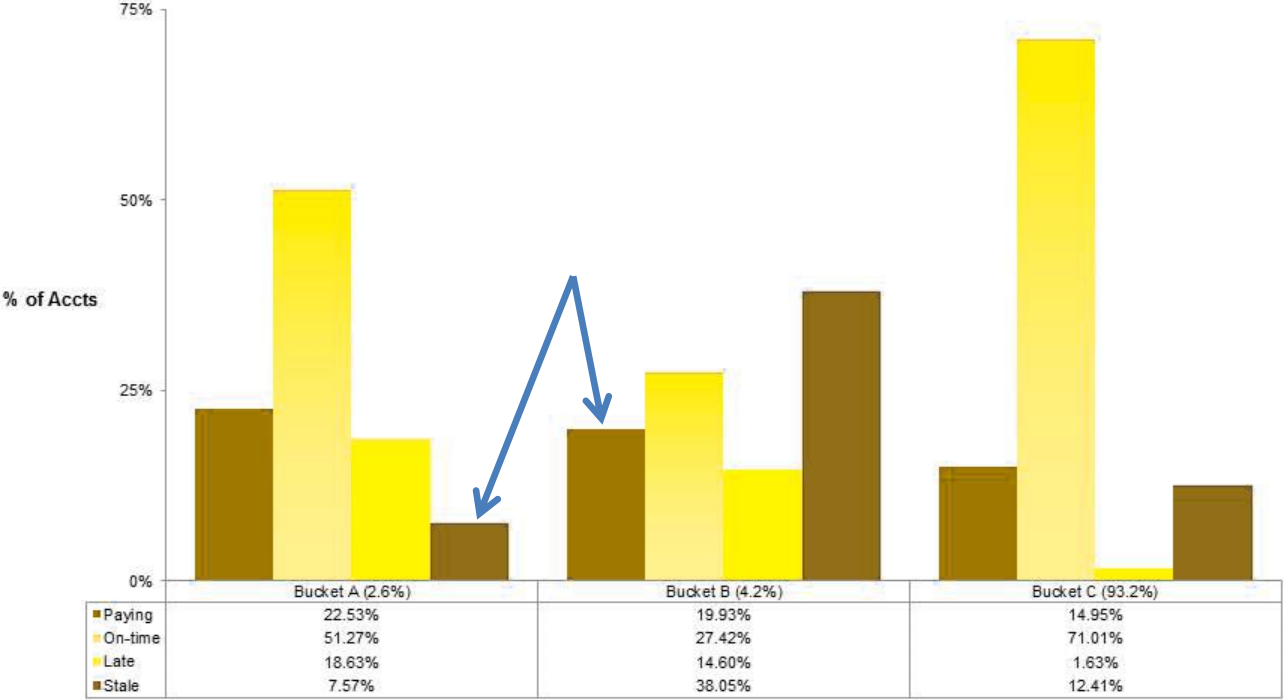


Protanopia Simulation



Source: *The Big Book of Dashboards* (Figure 1.24)

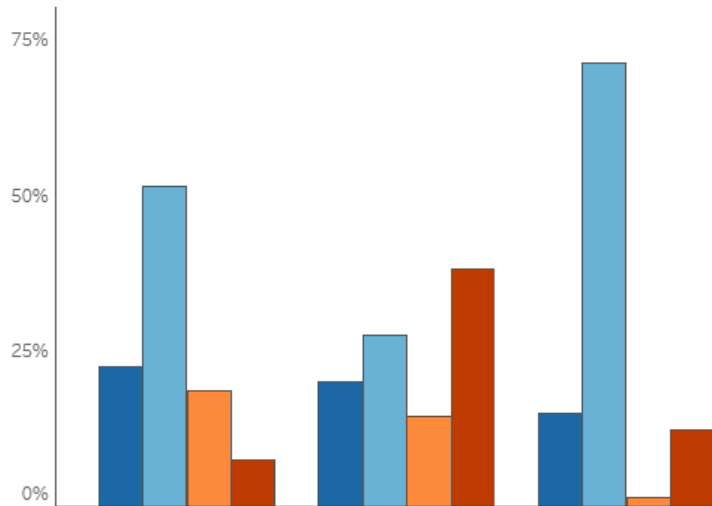
# Protanope Simulation



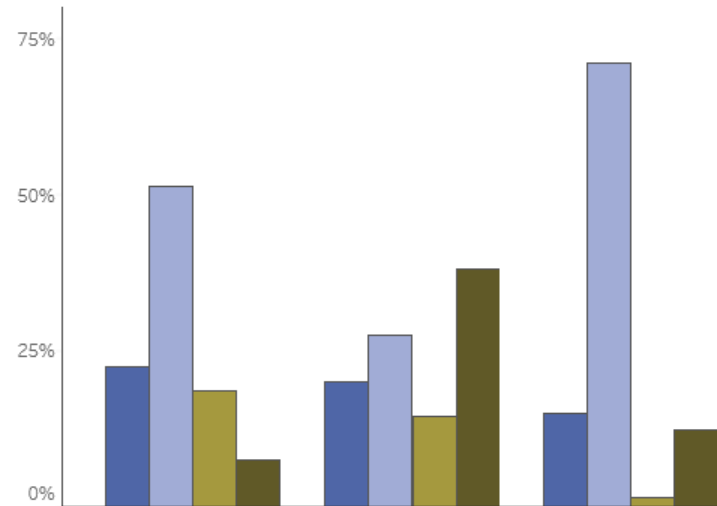
# VisCheck

[www.vischeck.com](http://www.vischeck.com)

Colorblind-Friendly Blue and Orange

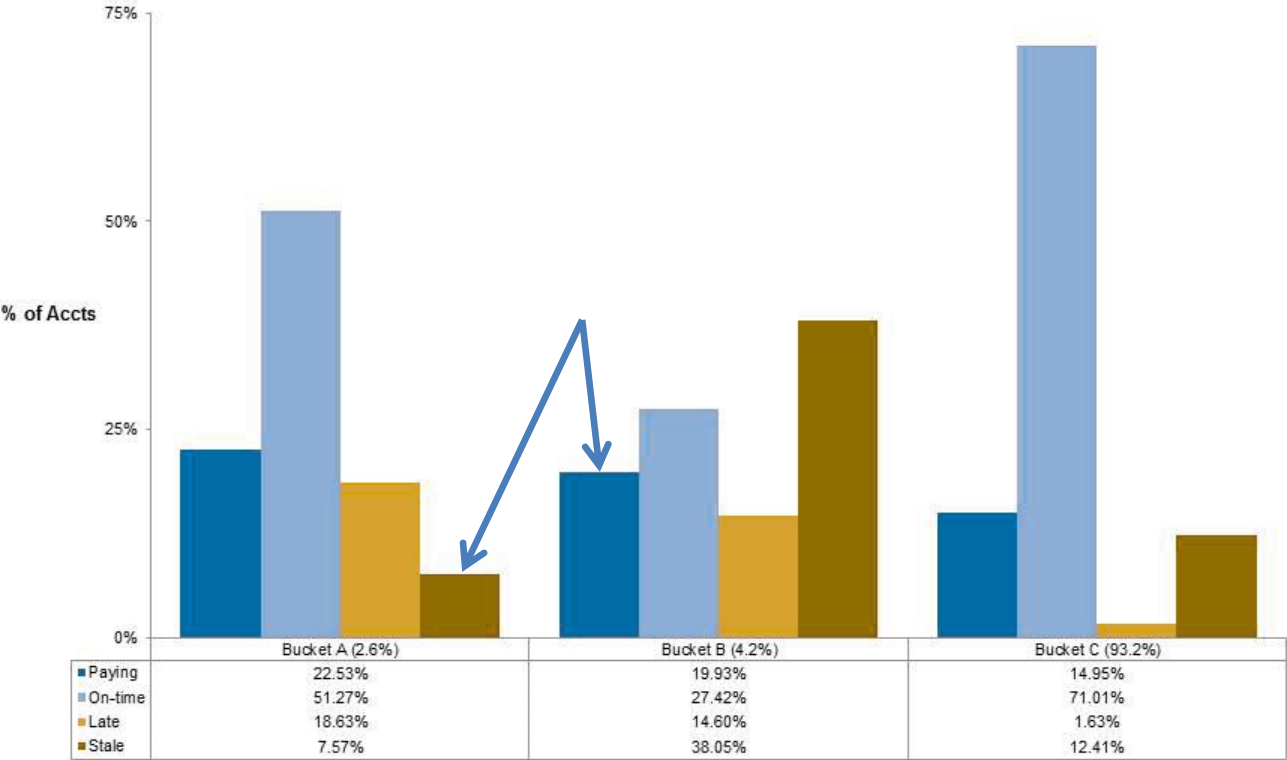


Protanopia Simulation

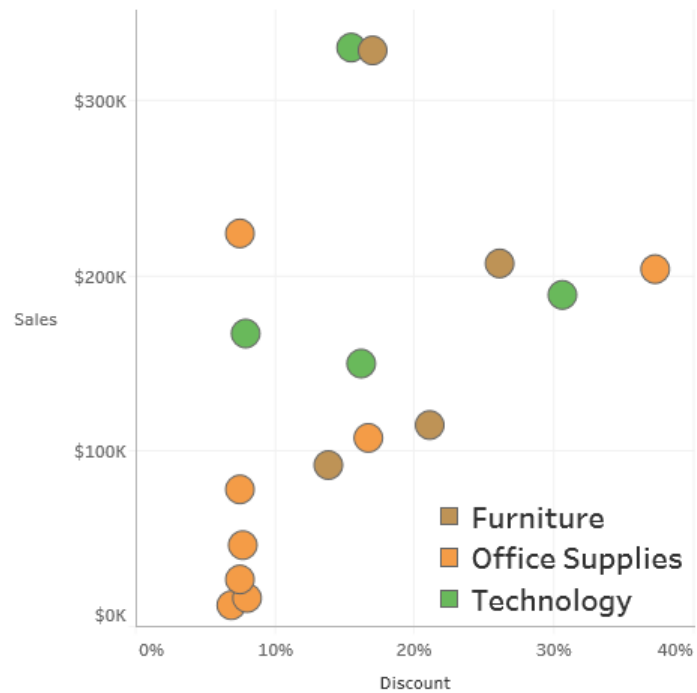


Source: *The Big Book of Dashboards* (Figure 1.25)

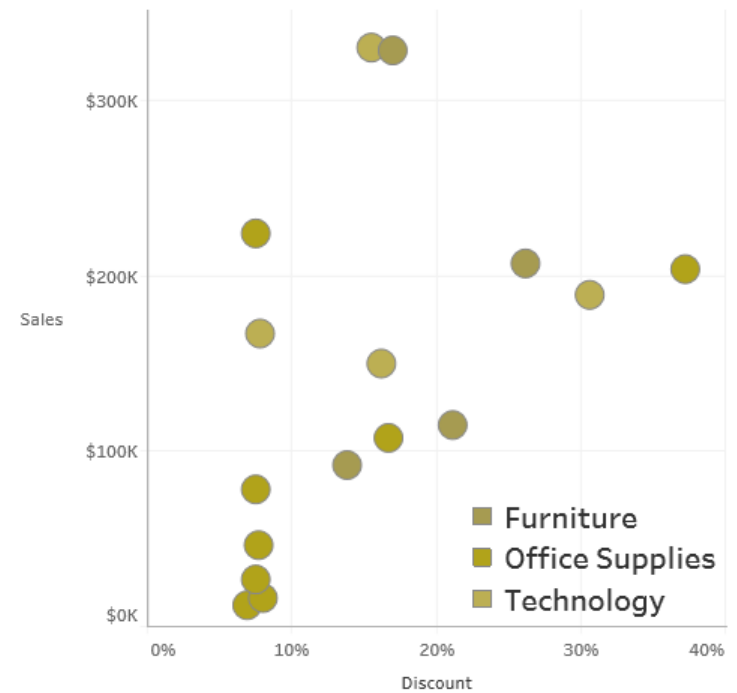
# Protanope Simulation



## Normal Color

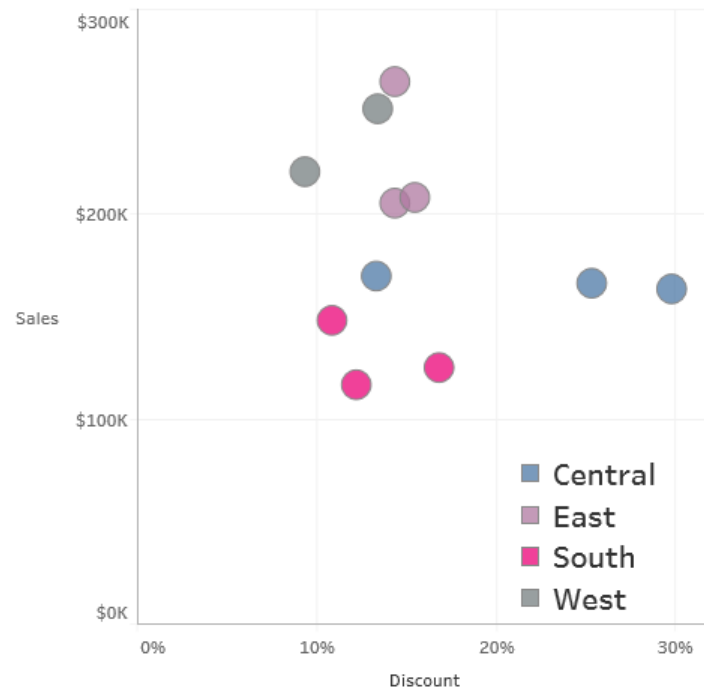


## Protanopia CVD Simulation

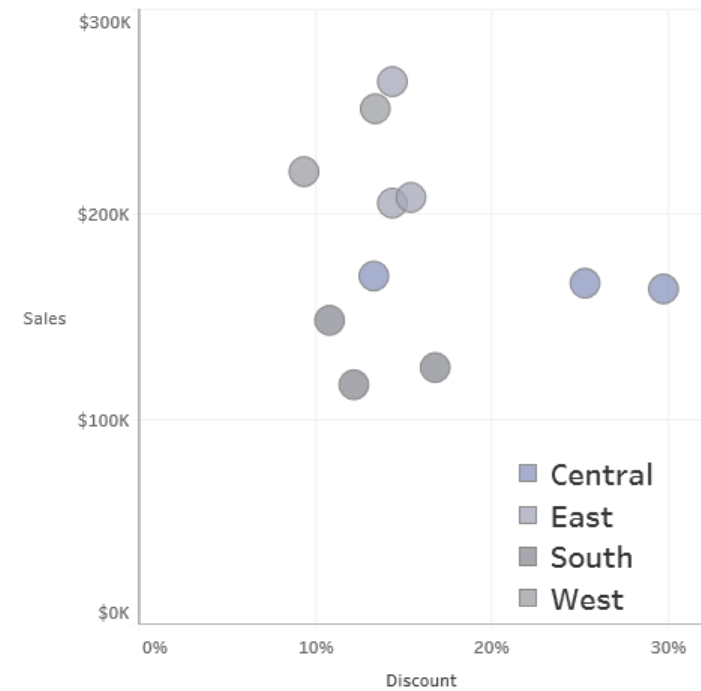


Source: *The Big Book of Dashboards* (Figure 1.26)

## Normal Color



## Deuteranopia CVD Simulation



Source: *The Big Book of Dashboards* (Figure 1.27)

## Traffic Light Color Palette

#E22049



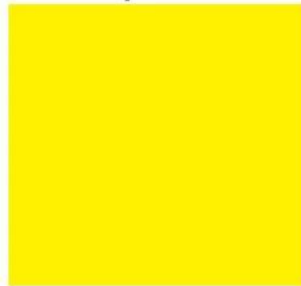
#FFF200



#0D9E49



## Deuteranopia Simulation



Source: *The Big Book of Dashboards* (Figure 33.8)



## Alternate Traffic Light Color Palette

#8E191C



#FFDE17



#83C775



Deuteranopia Simulation



Source: *The Big Book of Dashboards* (Figure 33.9)

## Alternate Traffic Light Color Palette

#8D1D1C



#FFDE17



#83C79B



Deuteranopia Simulation



Source: *The Big Book of Dashboards* (Figure 33.10)

# Reasons Why I hate Pie Charts



■ Because I am color blind

# Color Vision Deficiency (aka Colorblind)

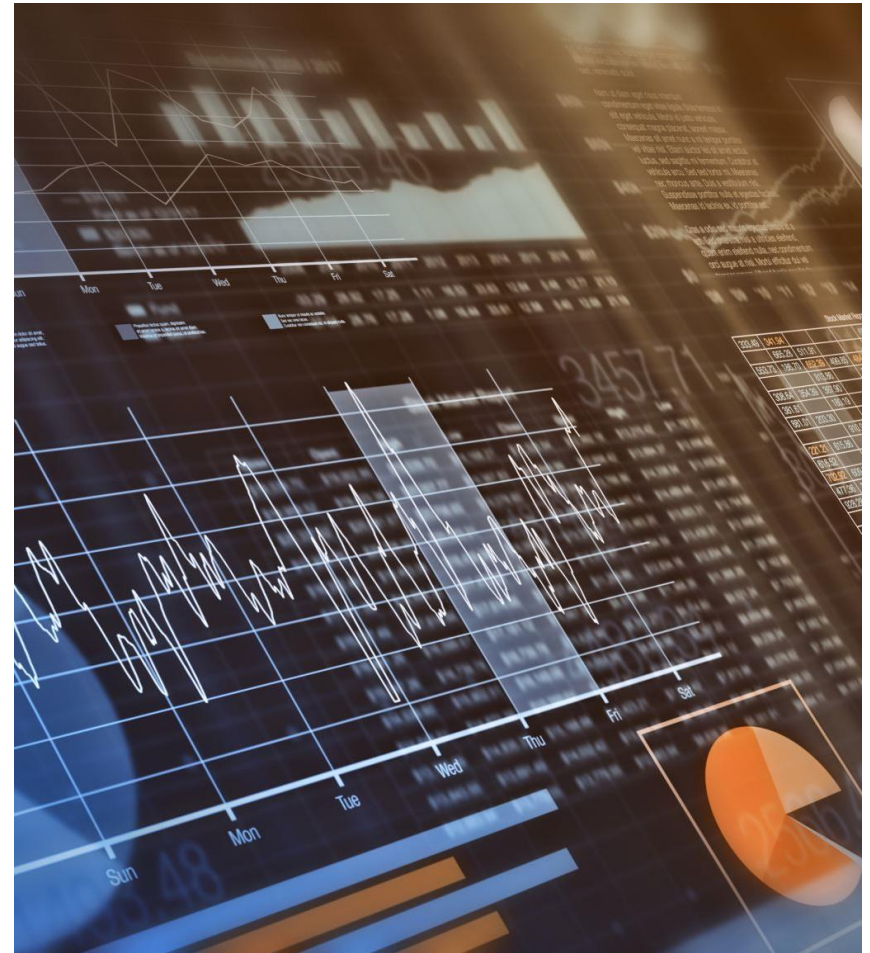
Color  
Normal



Color  
Vision  
Deficiency

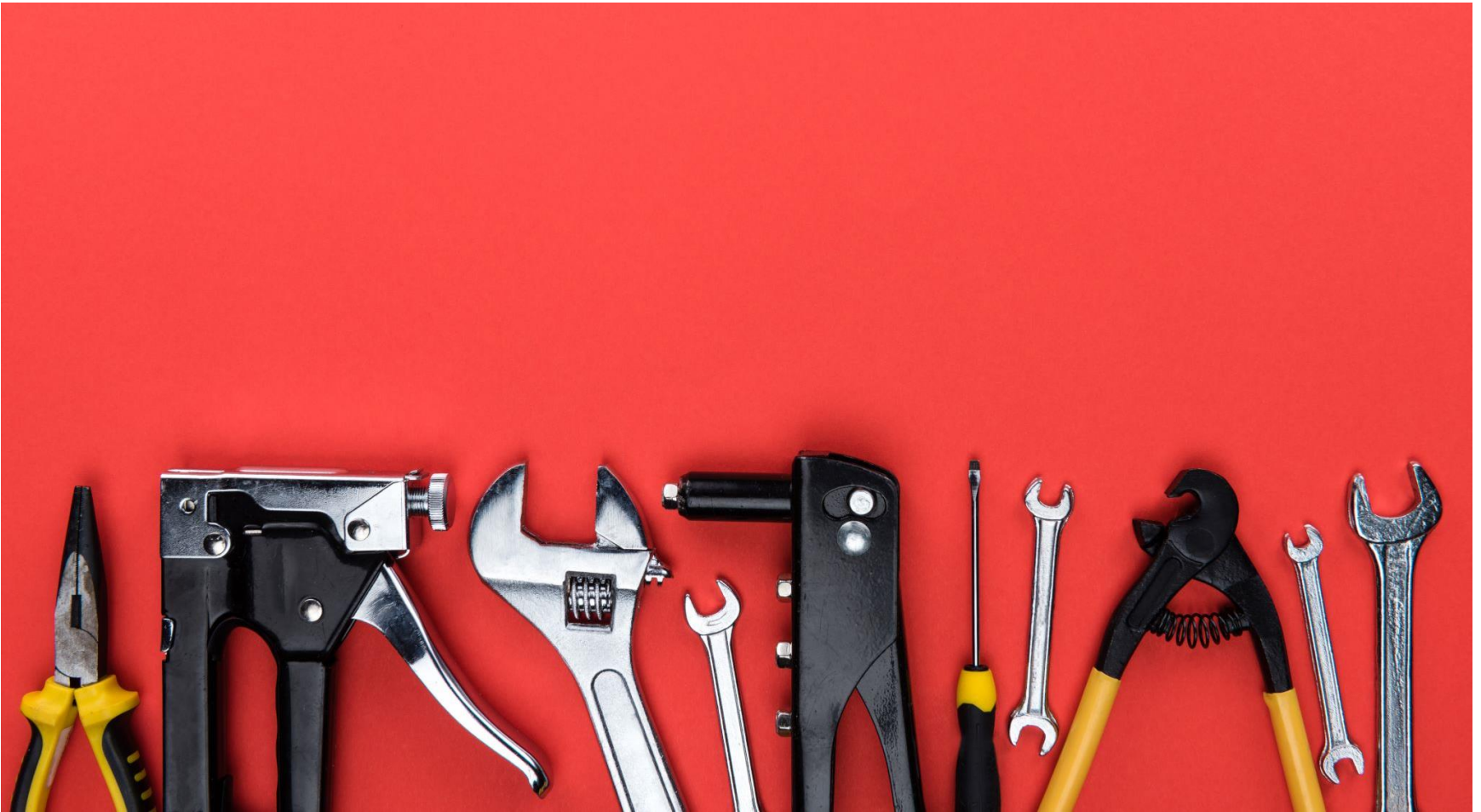
## How to choose the right data visualization tools for your application

- Choosing the best data visualization software tool can be tricky.
- The key point is to understand your project needs.
- Working with the right tools from the beginning can save you time, energy and money.





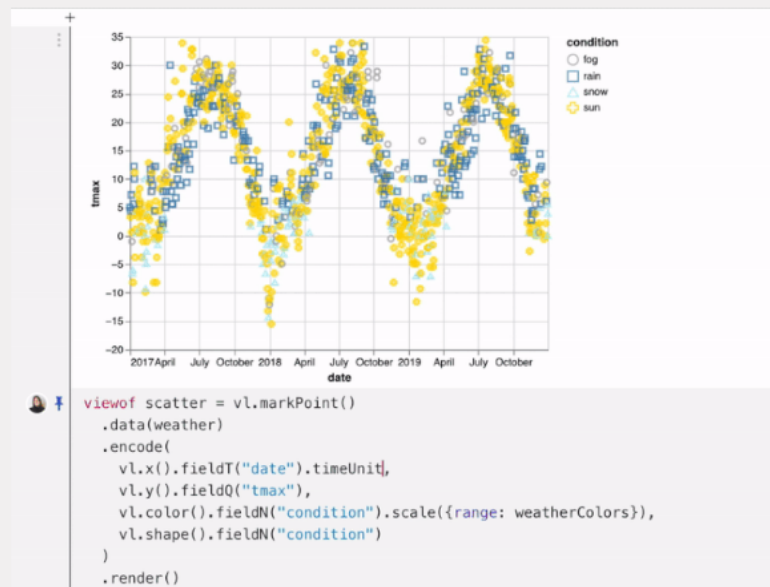
## How to choose the right data visualization tools for your application



# Flexibility vs. Easy to Use

Visualization tools focused on **flexibility** have robust feature sets that provide complete control for configuration, fine-grained adjustments of visualization design, and more advanced analytical capabilities.

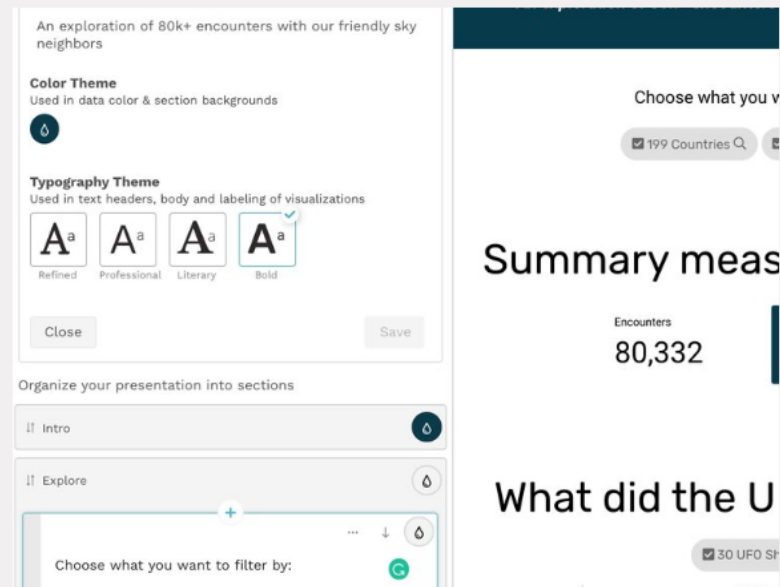
Best for: Experienced data analysts, developers



Coding-based configuration in Observable

A focus on **easy-of-use** allows non-technical users to get started quickly. However, there may be a sacrifice in control over details, customization, and functionality. The best tools make good default choices.

Best for: Non-technical users, business users



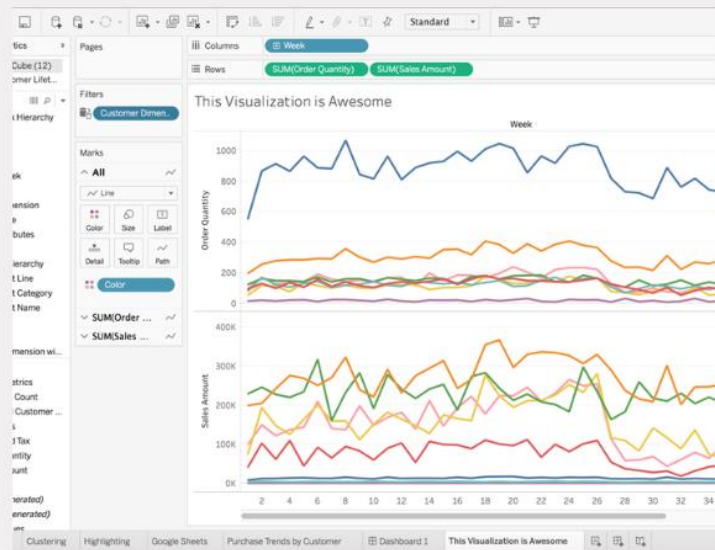
Drag-and-drop configuration in Juicebox



# Visual Analytics/Exploration vs. Data Storytelling/Explanation

**Visual analytics** tools improve the speed and capabilities of data analysis. The visualizations reveal patterns so the analyst can pursue additional and deeper insights into the data.

Best for: Data analysts and data scientists working multiple data sets.



*Tableau is a leader in visual analytics*

**Data storytelling** tools focus on the communication of data between people. The user wants to convey a message or insights to a target audience. The data can be combined with data and images to tell the full story.

Best for: Business users, consultants, subject matter experts.

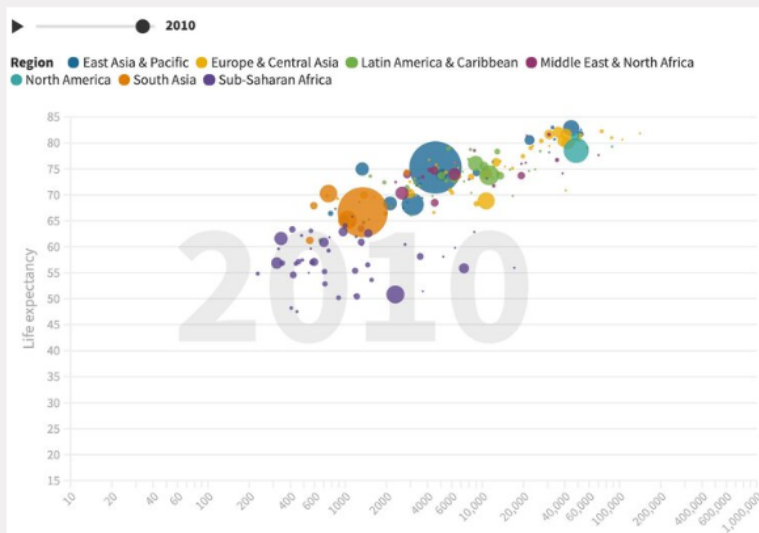


*Juicebox is focused on data storytelling*

# Independent Visualizations vs. Applications/Dashboards

**Independent visualization** tools are designed to create one chart at a time. Each chart is an independent “island” with a single data set. Often these charts are embedded as part of a separate website.

Best for: Data journalists, students, public relations and marketing



*Flourish generates stand-alone advanced visualizations*

**Application and dashboard creation** tools create an application or dashboard as the basic unit. In this case, you are creating a collection of visual components that function together to cover a broader array of information.

Best for: Internal reporting/dashboards, consultants presenting results

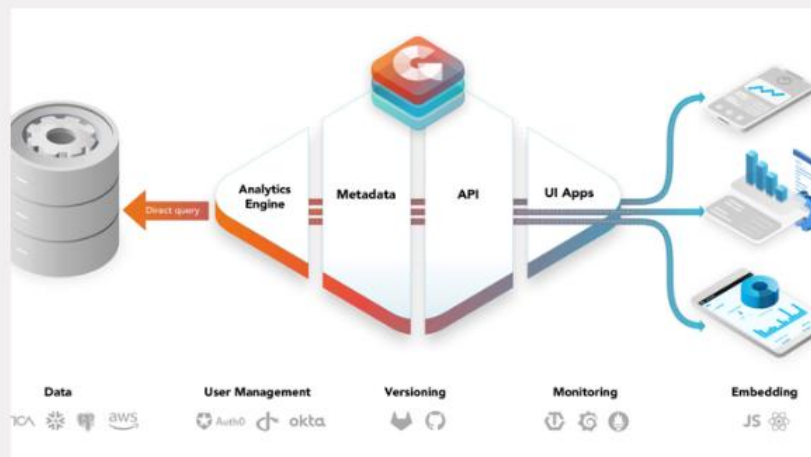


*Datapine is a tool for creating dashboards*

# Broad Data Platforms vs. Focused Visualization Tools

Complex **data platforms** will emphasize visualization capabilities as an entry point for new users. These platforms may include a wide array of capabilities for data management, governance, and modeling.

Best for: Technology teams looking for a single solution to cover all their data needs.



Solutions **focused on visualization** tend to stick to what they are good at. In particular, these solutions don't attempt to be the “single source of truth” for data in an organization.

Best for: Individuals and teams with access to data sources.



# Cost to Get Started

**Free with limitations** on volume of use or advanced features.

Best for: Users interested in exploring capabilities using data that may be sensitive.

**Free for public visualizations** and may also have functional limitations.

Best for: Users who aren't concerned about sharing their data publicly (e.g. students)

**Trial access** is full-featured but for a limited time.

Best for: Users looking to test the solution with expectations to move to a paid license.

For starters looking to improve their visual communication

**\$0**

Free forever

## Basic

Get started for free to see how easy it is to design beautiful data visualizations with Infogram.

**Free**

Forever

## Pro

For those who want to download, share privately, and access premium templates and images.

**\$19** / month

Billed yearly

## Start your 15-day Free Trial

Build your first data story in 10 minutes

# Beautiful Trash

a week of trash routes in Cincinnati

# BEAUTIFUL TRASH

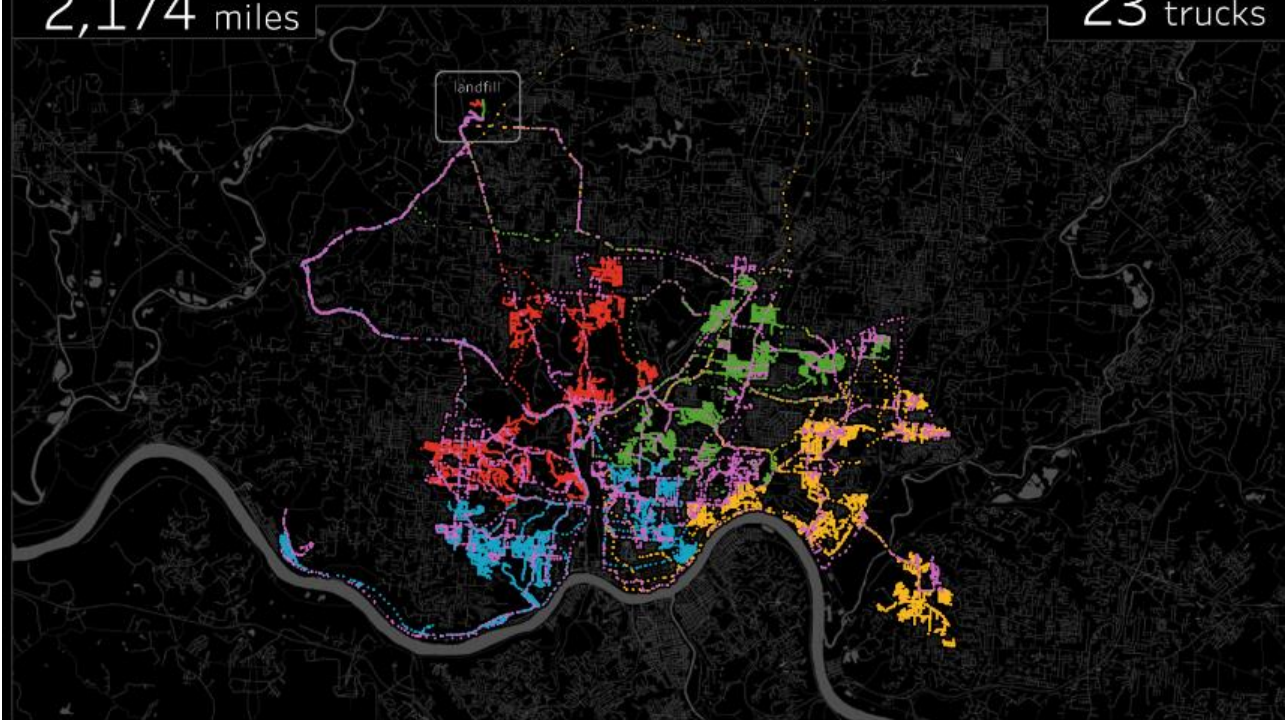
a week of trash routes in Cincinnati



2,174 miles

05:46:45 to 20:59:27 [hh:mm:ss]

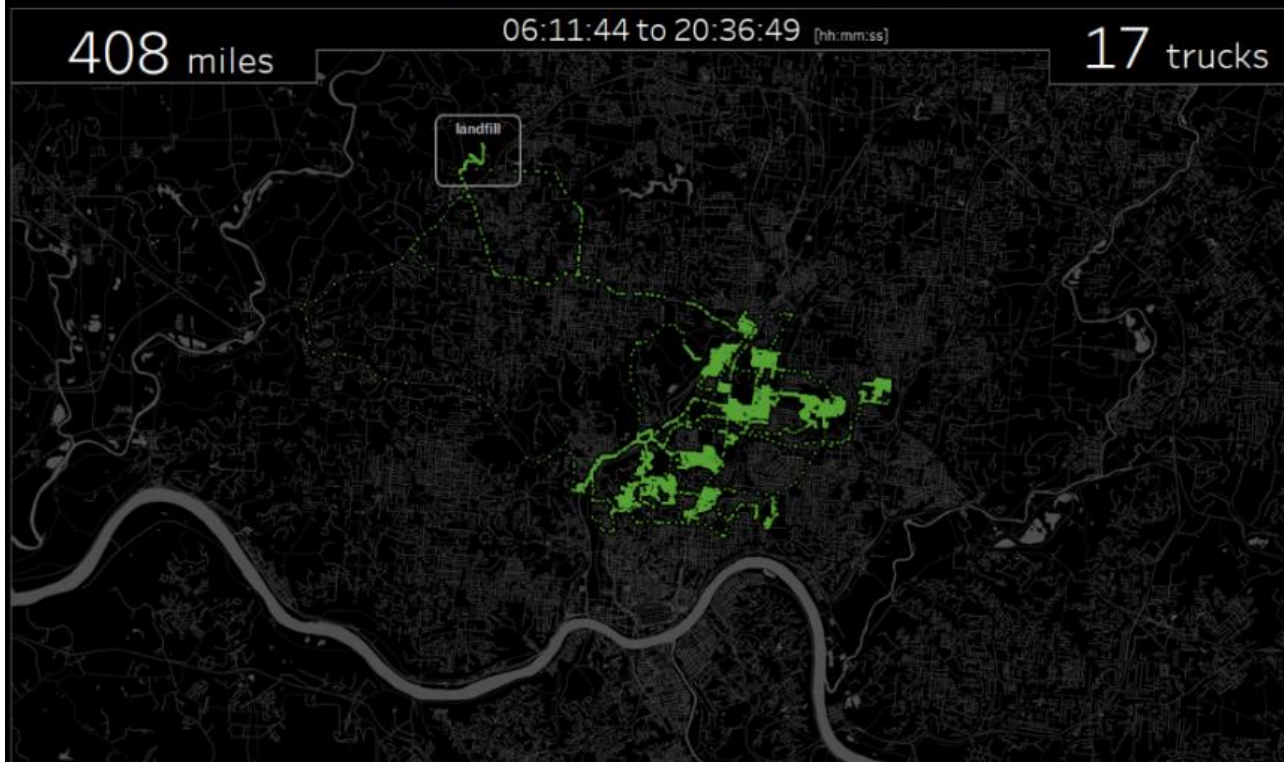
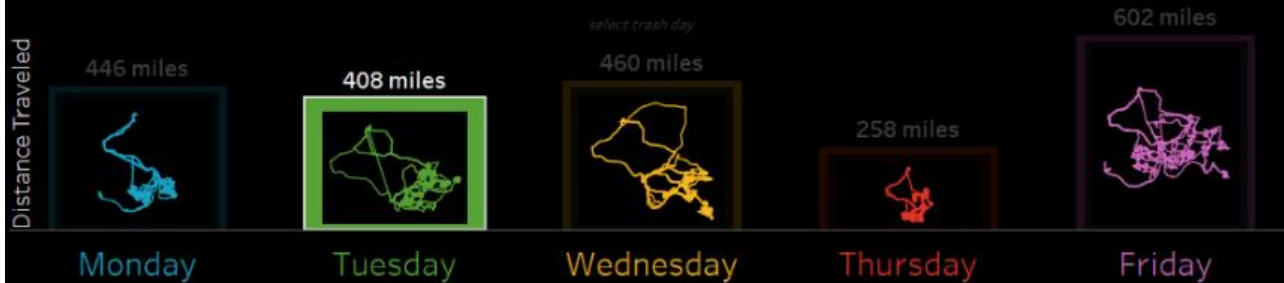
23 trucks





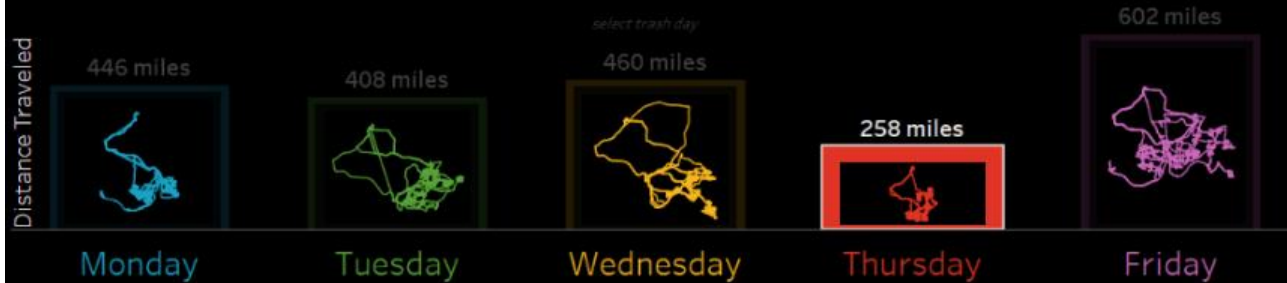
# BEAUTIFUL TRASH

a week of trash routes in Cincinnati



# BEAUTIFUL TRASH

a week of trash routes in Cincinnati



258 miles

06:24:04 to 20:20:53 [hh:mm:ss]

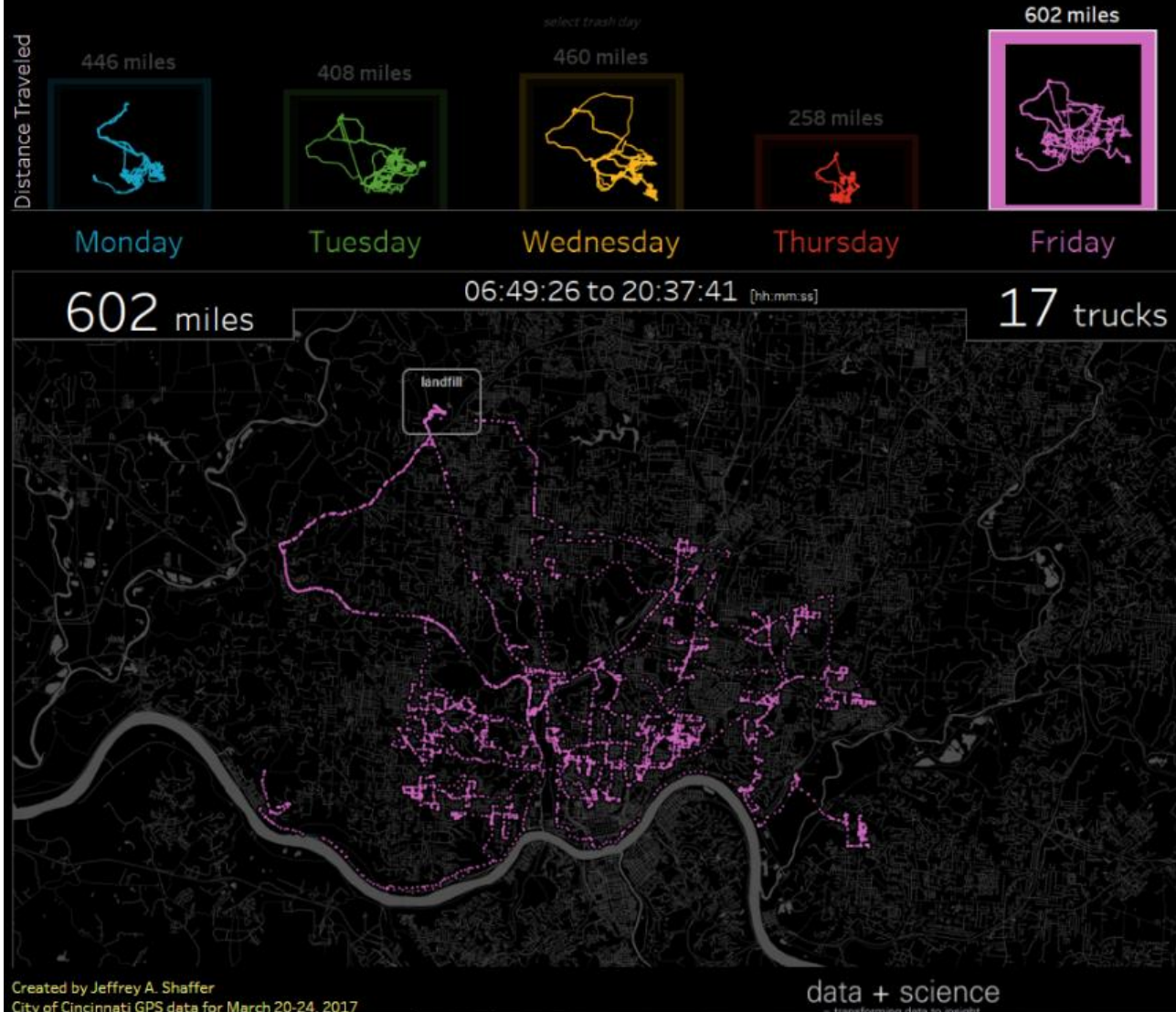
21 trucks





# BEAUTIFUL TRASH

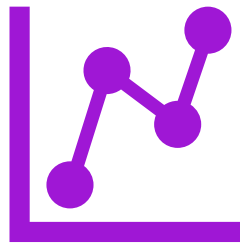
a week of trash routes in Cincinnati



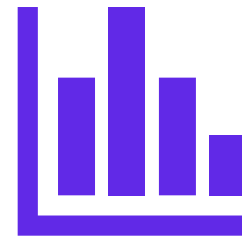
# Multivariate visualization



# Multivariate Analysis



Many statistical techniques focus on just one or two variables.



Multivariate analysis (MVA) techniques allow more than two variables to be analysed at once.

# What is Multivariate Data?

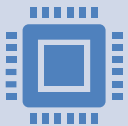


Each data point has  $N$  variables or observations



Each observation can be:

nominal or ordinal  
discrete or continuous  
scalar, vector, or tensor




May or may not have spatial, temporal, or other connectivity attribute

# Characteristics of a Variable


Order: grades have an order, brand names do not.



Distance metric: for income, distance equals difference.  
For rankings, difference is not a distance metric.



Absolute zero: temperature has an absolute zero, bank account balances do not.

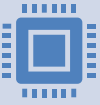


A variable can be classified by these three attributes, called *Scale*.



Effective visualizations attempt to match the scale of the data dimension with the graphical attribute conveying it.

# Sources of Multivariate Data



Sensors (e.g.,  
images, gauges)



Simulations



Census or other  
surveys



Commerce (e.g.,  
stock market)



Communication  
systems



Spreadsheets and  
databases

# Issues in Visualizing Multivariate Data

---

How many variables?

---

How many records?

---

Types of variables?

---

User task (exploration, confirmation, presentation)

---

Data feature of interest (clusters, anomalies, trends, patterns, ....)

---

Background of user (domain expert, visualization specialist, decision-maker, ....)

# Methods for Visualizing Multivariate Data

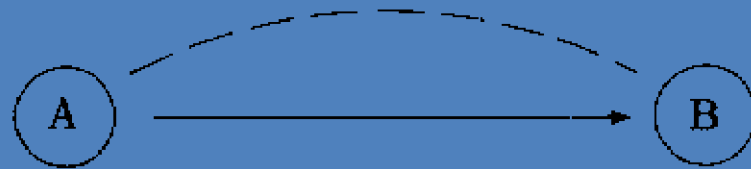
Dimensional Subsetting

Dimensional Reorganization

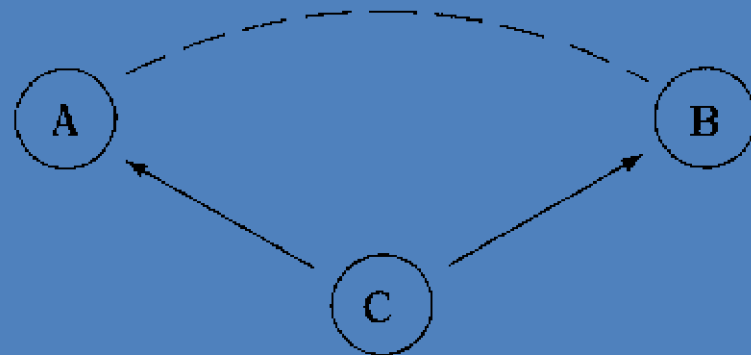
Dimensional Embedding

Dimensional Reduction

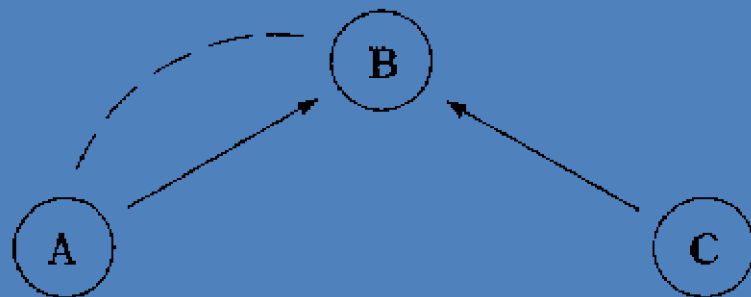




**CAUSATION**—Changes in A cause changes in B.



**COMMON RESPONSE**—Changes in both A and B are caused by changes in a third variable, C.



**CONFOUNDING**—Changes in B are caused both by changes in A and by changes in third variable C.

# Summary