DS684
Cloud Computing
Week 09

Regarding Labs and Assignments

- Class participation means more than Zoom attendance. You must actively participate in the discussion and labs, and answer questions.

- Must hit Submit button, otherwise no grade
- If you need extension in time, must send written request (***email***). Otherwise no grade and no makeup. Requests sent over Zoom chat do not count.
- For any technical difficulty (installation, Azure access, etc), you must send written explanation (***email***) before the deadline. Otherwise no grade and no makeup.

Teaching Schedule

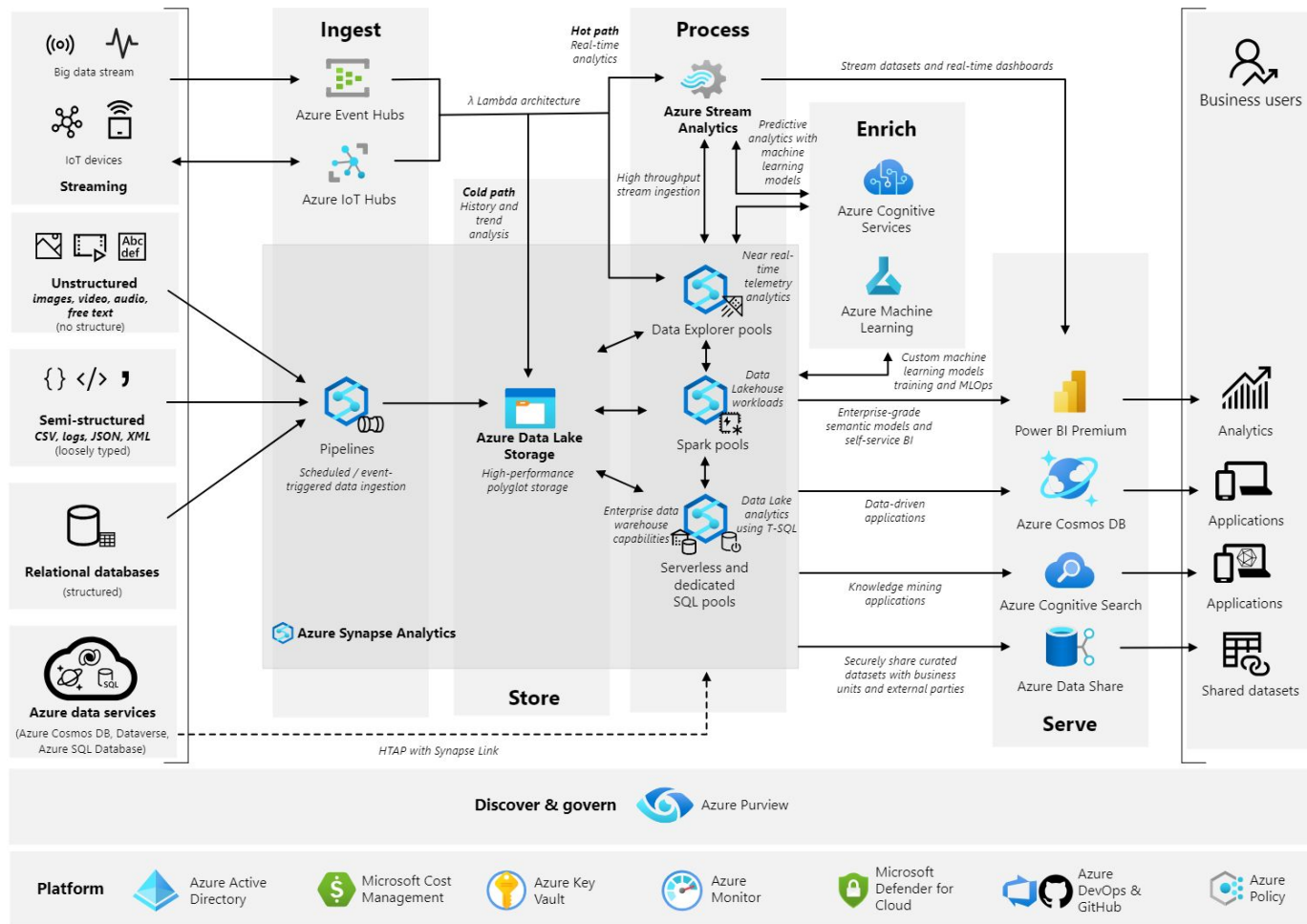Week 7: Azure Synapse Analytics Part I: Data Warehouse

Week 8: Azure Synapse Analytics Part II: Data Engineering

Week 9: Visualization using Power BI

Week 10: Azure Machine Learning

Week 11: Final project presentation

Azure Data Related Services

Ingest

Big data stream

Azure Event Hubs

IoT devices

Azure IoT Hubs

Streaming

Unstructured
images, video, audio, free text
(no structure)

Semi-structured
CSV, logs, JSON, XML
(loosely typed)

Relational databases
(structured)

Azure data services
(Azure Cosmos DB, Dataverse, Azure SQL Database)

Hot path
Real-time analytics

Process

Azure Stream Analytics

λ Lambda architecture

Cold path
History and trend analysis

High throughput stream ingestion

Pipelines
Scheduled / event-triggered data ingestion

Azure Data Lake Storage
High-performance polyglot storage

Near real-time telemetry analytics

Data Explorer pools

Data Lakehouse workloads

Spark pools

Enterprise data warehouse capabilities

Data Lake analytics using T-SQL

Serverless and dedicated SQL pools

Azure Synapse Analytics

Store

HTAP with Synapse Link

Enrich

Azure Cognitive Services

Azure Machine Learning

Predictive analytics with machine learning models

Custom machine learning models training and MLOps

Stream datasets and real-time dashboards

Business users

Power BI Premium

Enterprise-grade semantic models and self-service BI

Azure Cosmos DB

Data-driven applications

Azure Cognitive Search

Knowledge mining applications

Azure Data Share

Securely share curated datasets with business units and external parties

Serve

Analytics

Applications

Applications

Shared datasets

Discover & govern   Azure Purview

Platform   Azure Active Directory   Microsoft Cost Management   Azure Key Vault   Azure Monitor   Microsoft Defender for Cloud   Azure DevOps & GitHub   Azure Policy

Microsoft Azure

Agenda

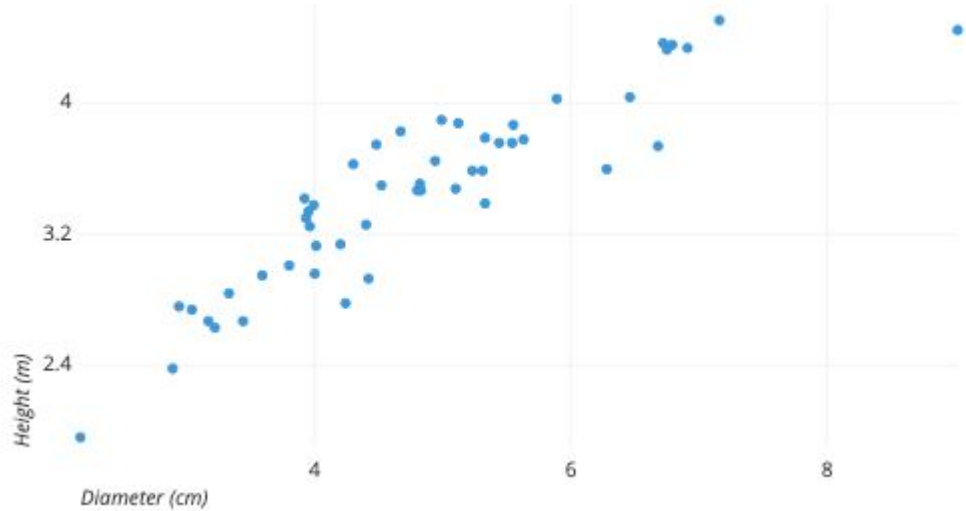- **Visualization**
    - Scattered Dot and Visual Characteristics
    - Visual Enhancements
    - Compare and Contrast
- The Good, the Bad, and the Ugly
    - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
    - Lab: Power BI

Levels of Analysis
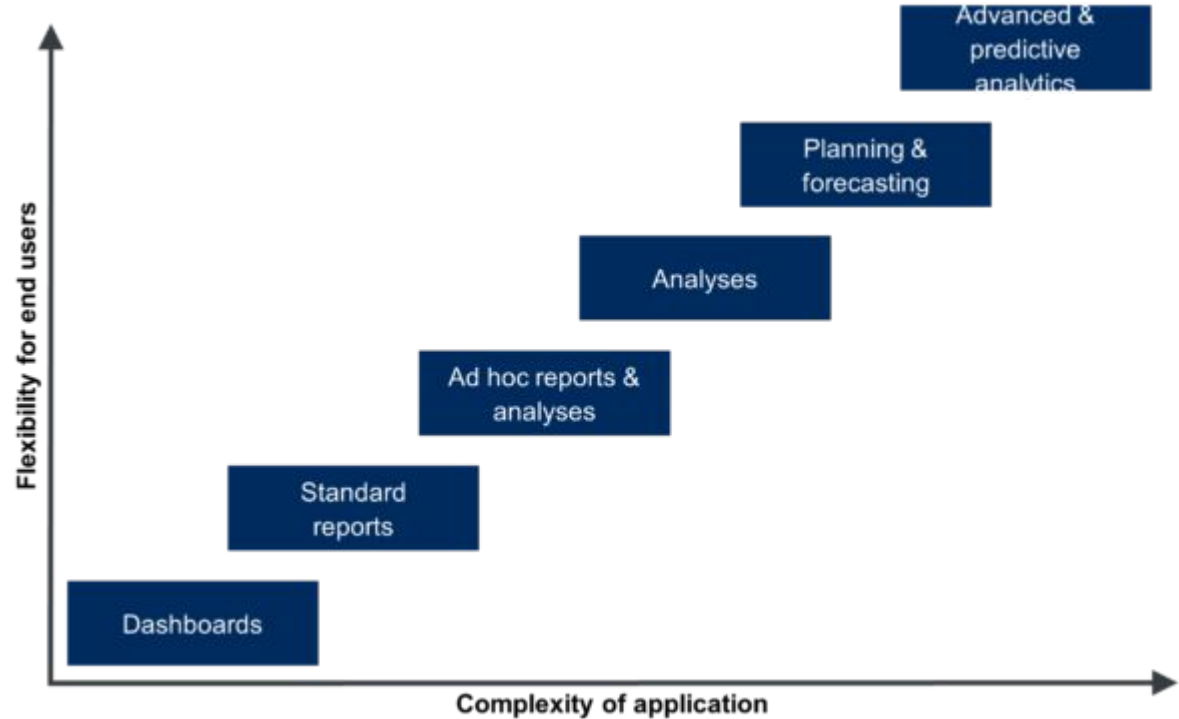
Types of Data Analysis

- Dashboard (summarized)
  - Data visualization
- Static (operational) report
- Ad hoc reporting and Online Analytical Processing (OLAP)
  - Based on cube or based on databases
- Data analysis
- Planning and forcasting
  - Data science
- Machine learning/Artificial Intelligence

Dashboard Visualization

Why do you need to visualize data?

What can these two numbers tell you?

```
[3, 6]
```

Visualization
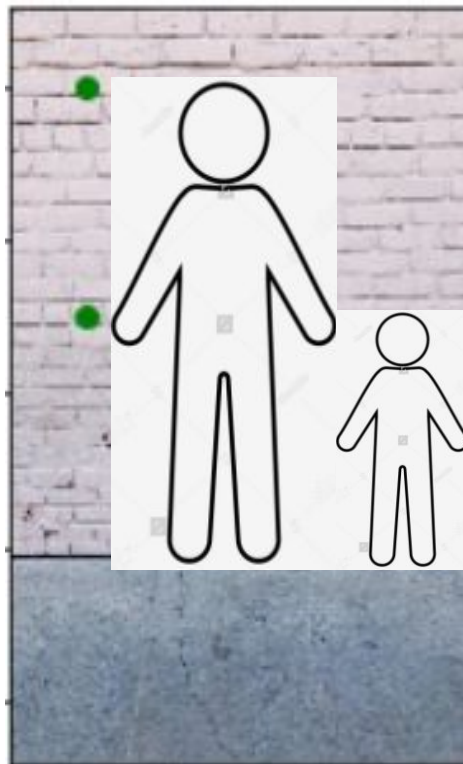
Trending
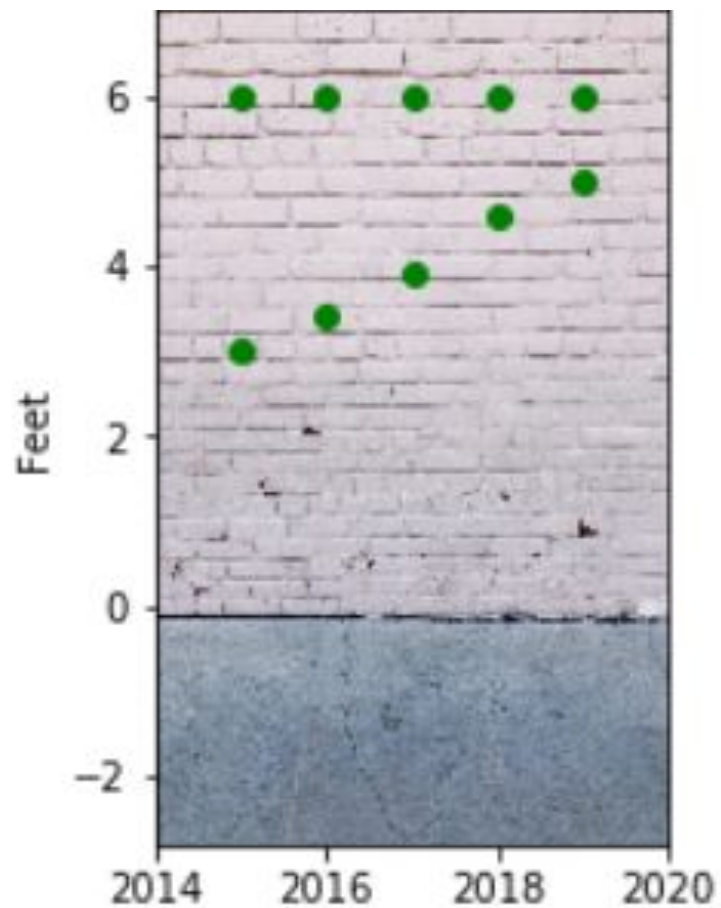
Story telling

Visualization is about storytelling

There are good visualizations and bad visualizations
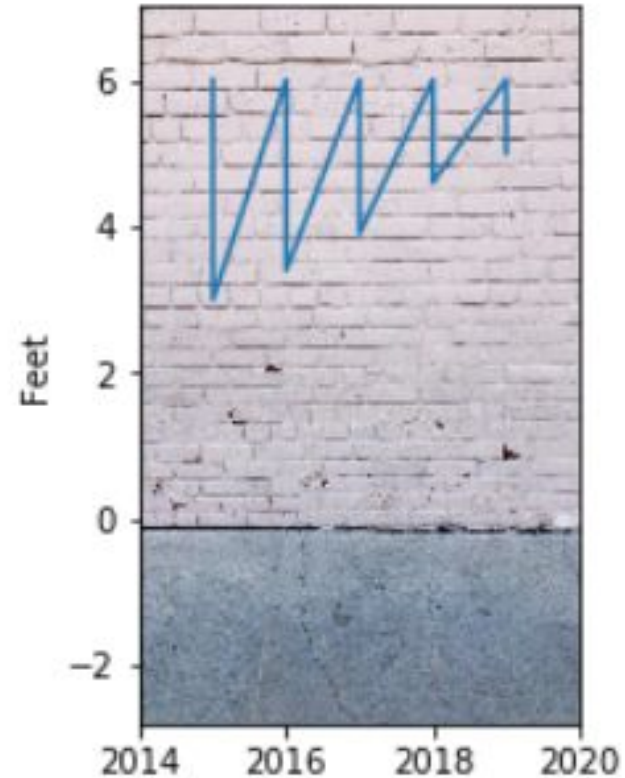
Agenda

- Visualization
  - Scattered Dot and Visual Characteristics
  - Visual Enhancements
  - Compare and Contrast
- The Good, the Bad, and the Ugly
  - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
  - Lab: Power BI

Scattered Dot

All visualization starts from scattered dots.

Use scattered dot first for pattern discovery

Do not assume you know the pattern (and what visualization works best)

Iris Dataset

Sometimes, the pattern is not so clear. Further analysis is needed

Iris dataset: measurement of 3 different species of Iris ===>

Iris Dataset
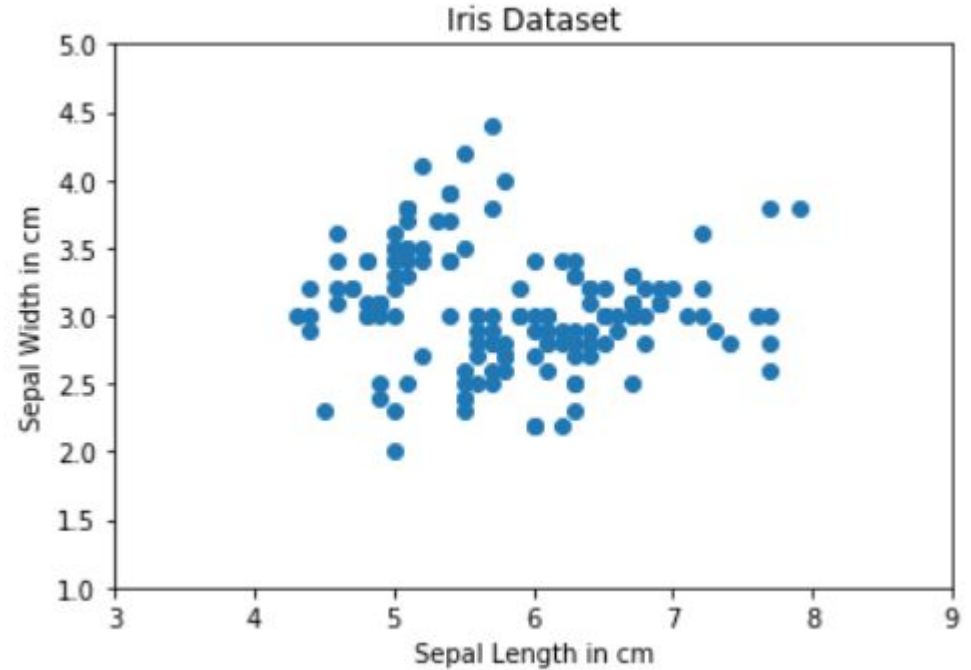
Visualization Characteristics

- Position: X axis and Y axis
- Color
- Size
- Shape (line, circle, rectangle)
- Background
- Placement

Position

Identify the most relevant
indicators for your analysis

Figure 1: Magic Quadrant for Cloud Infrastructure and Platform Services

Iris Dataset

Use color to group dots
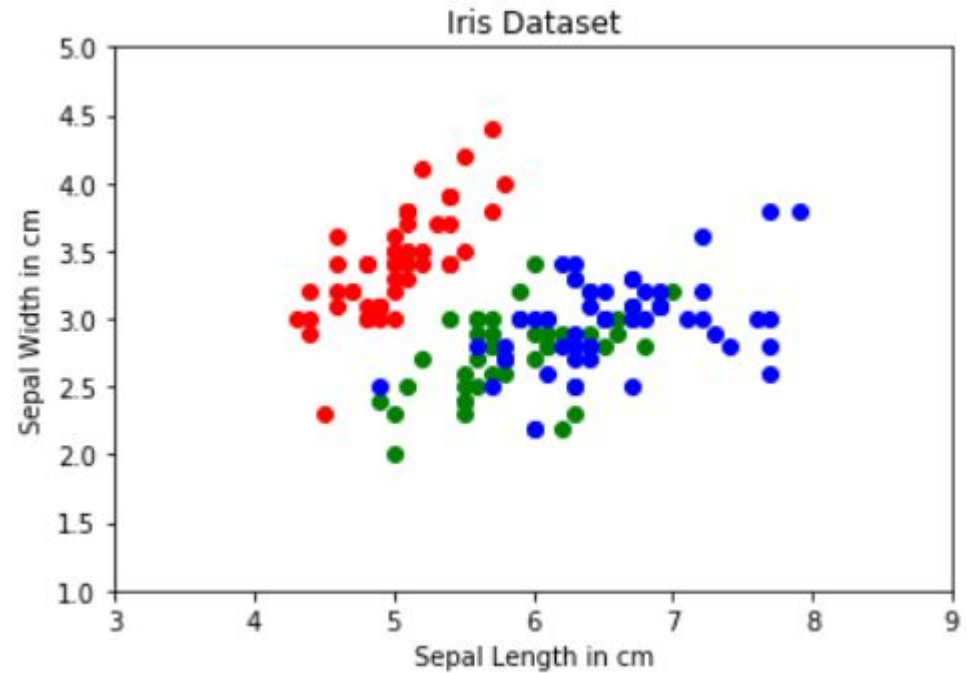
Size

Size adds an additional dimension.

But, does size deliver any information here? If not, we may not need it.
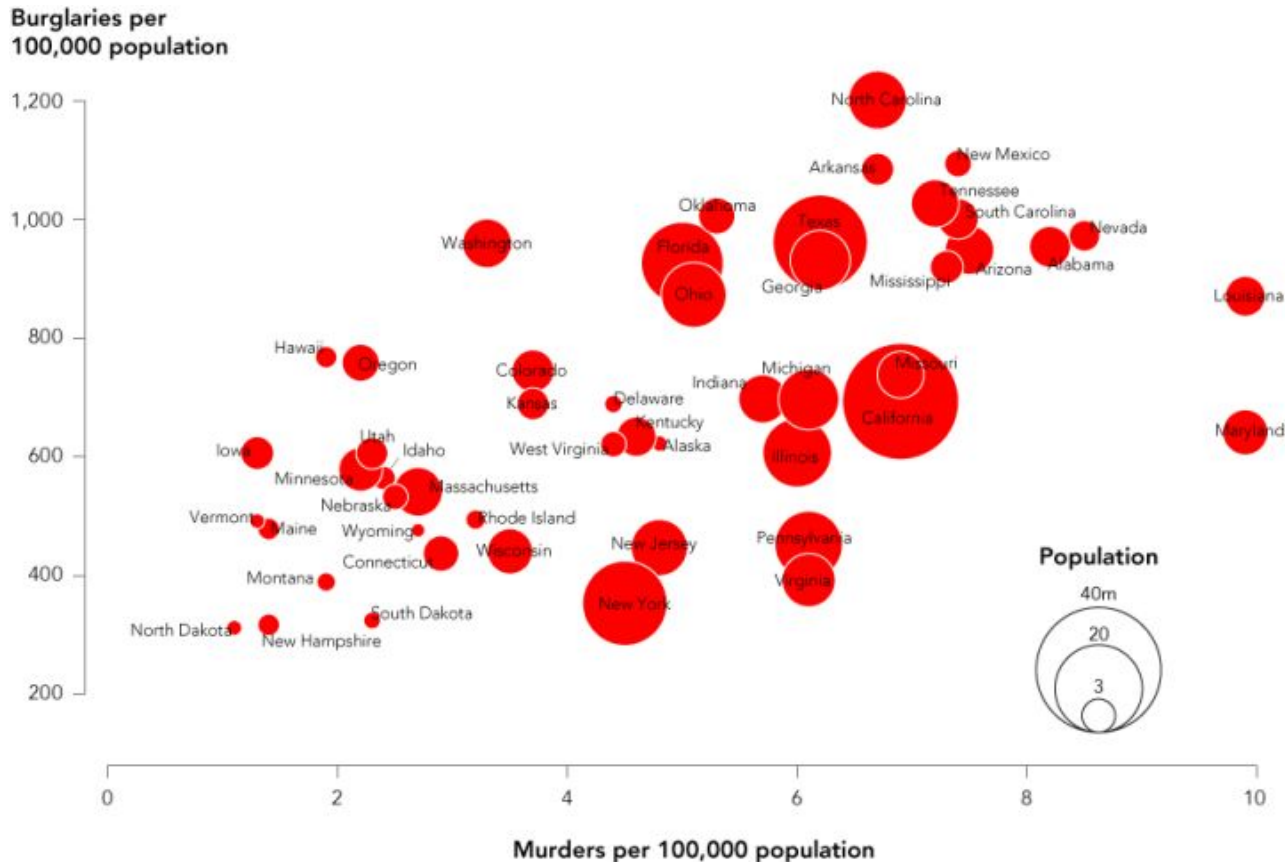
Color and Size

Characteristic Selection

Choose the characteristics (X-Y axial and color in this case) that can make the story more compelling

- First graph with color still can not separate two groups clearly
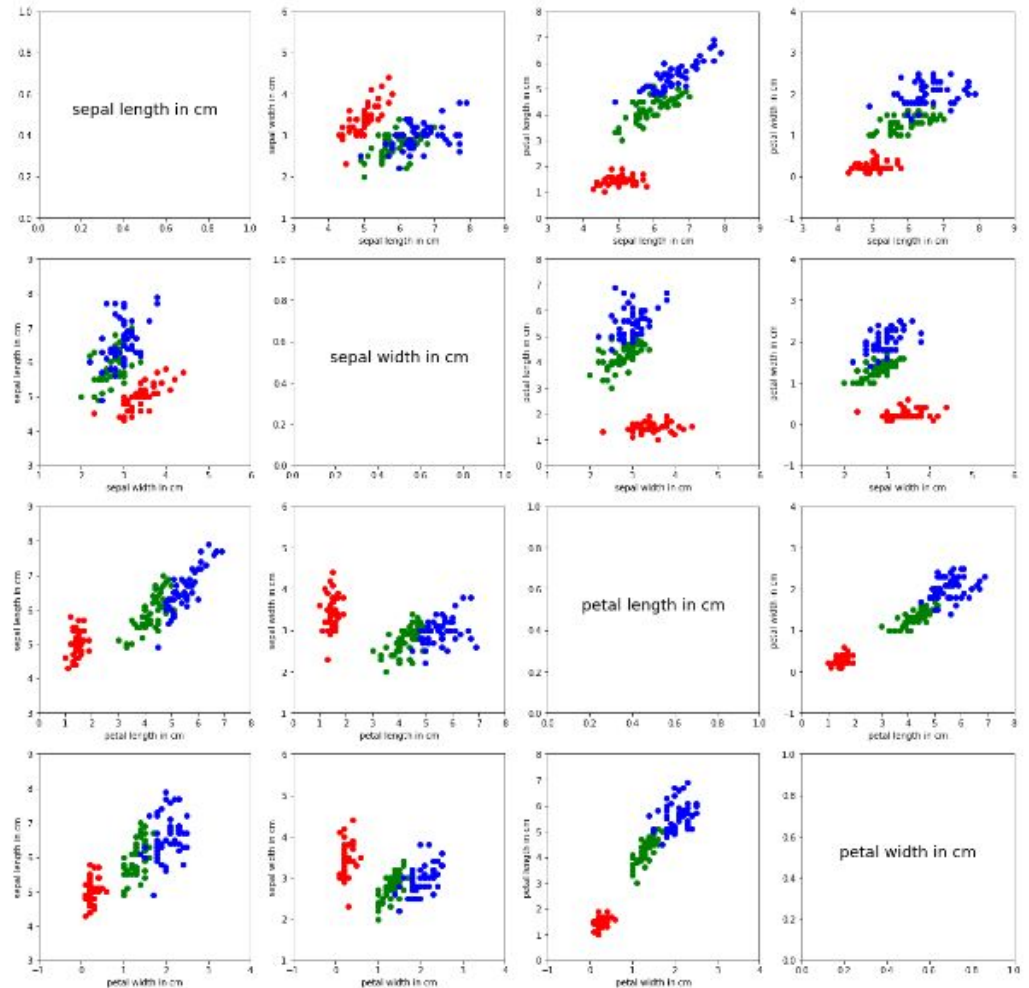
Agenda

- **Visualization**
  - Scattered Dot and Visual Characteristics
  - Visual Enhancements
  - Compare and Contrast
- The Good, the Bad, and the Ugly
  - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
  - Lab: Power BI

Turn Dots into Shapes

Scattered dot can show patterns. Patterns can be further visualized using shapes

- Line
- Bar
- Pie

Bar chart

Use shape to signify the dots
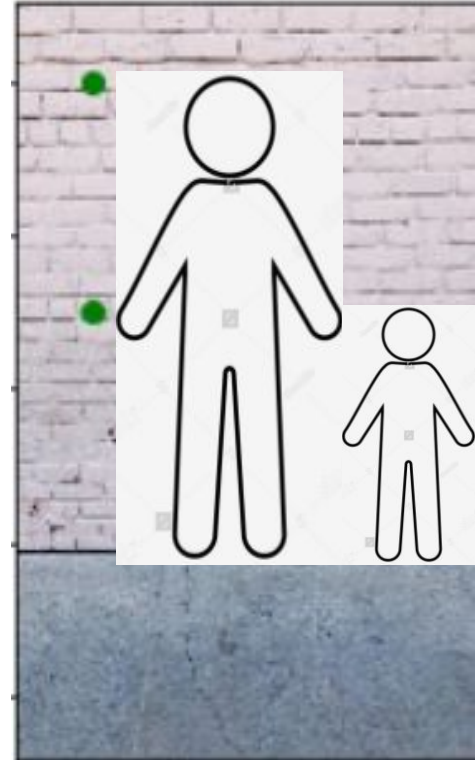
Bar chart

Comparing numbers for a discrete number of groups (and subgroups)
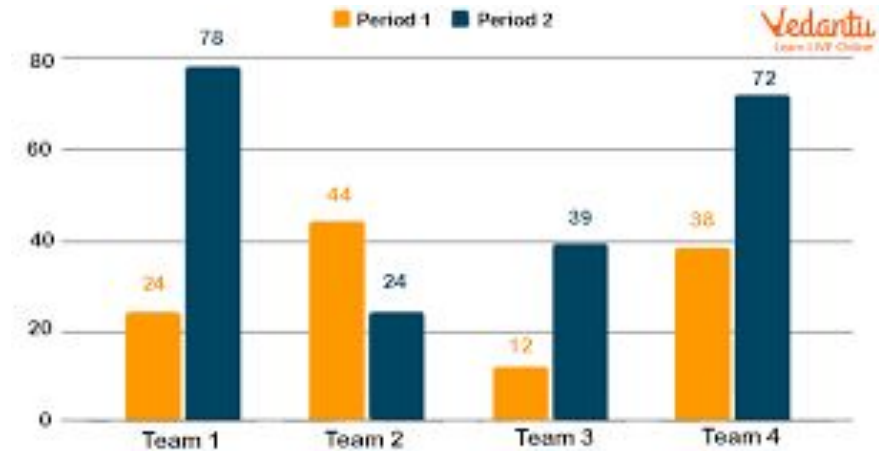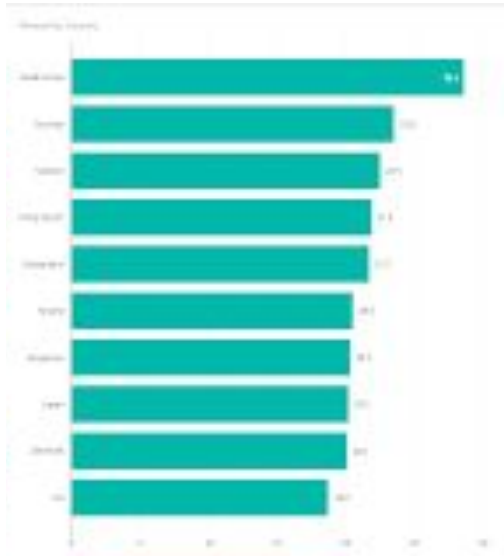
Bar chart

Not necessarily just bars. Length, width, area can all serve the purpose of comparison.

| PERSONNEL | 5.41M | 1.35M | | x 10,000 |

| ARMOURED UNITS | 144K | 60K | | x 1,000 |

| AIR FORCE | 20,723 | 4,173 | | x 200 |

| NAVAL FORCES | 2,049 | 605 | | x 50 |

| NUCLEAR WEAPONS | 6065 | 6255 | | x 200 |
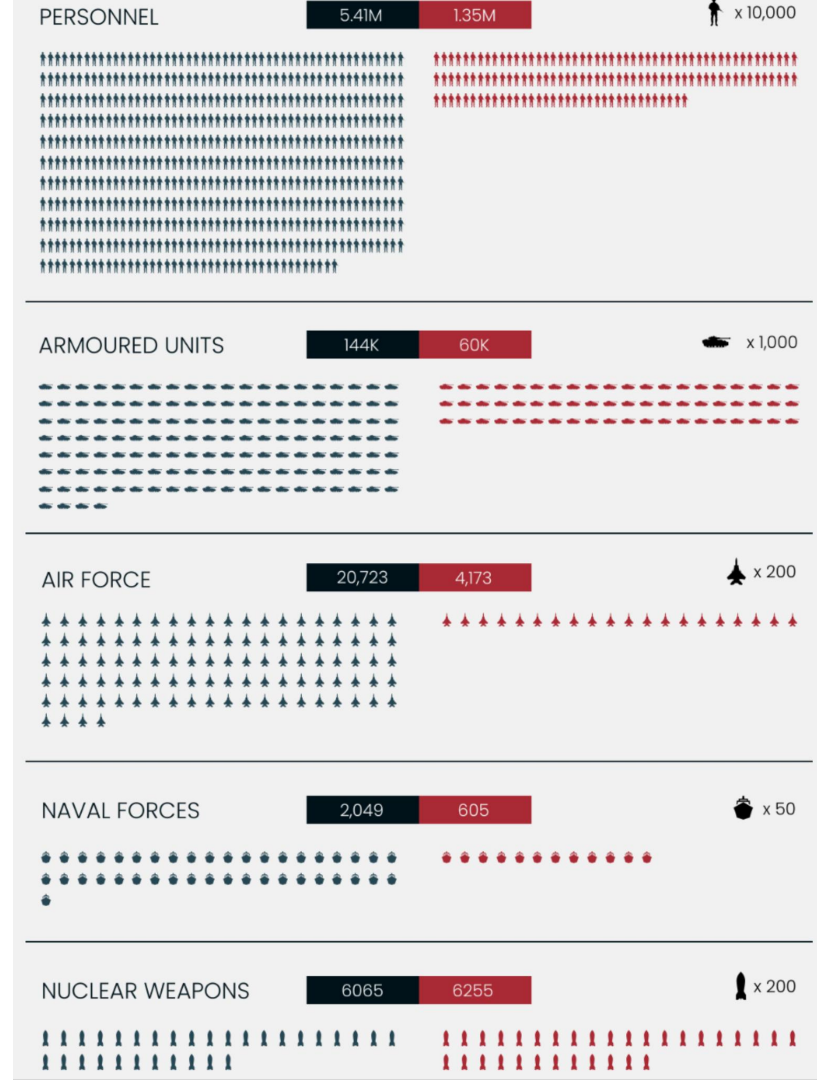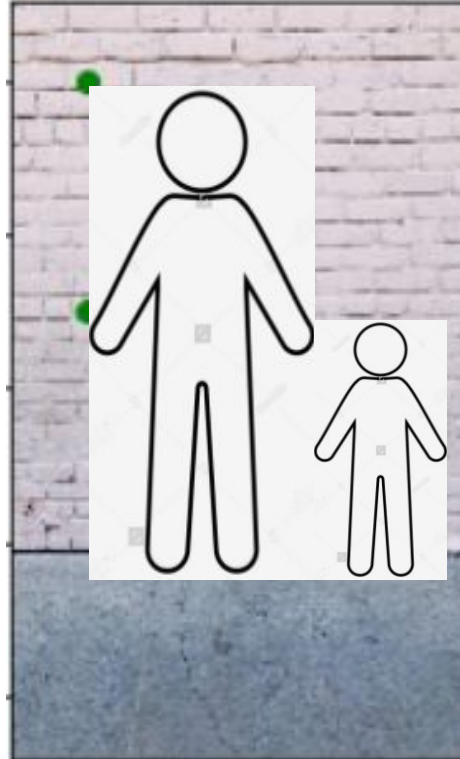
Line chart

Continuous. Usually used for continuous series, such as time series.
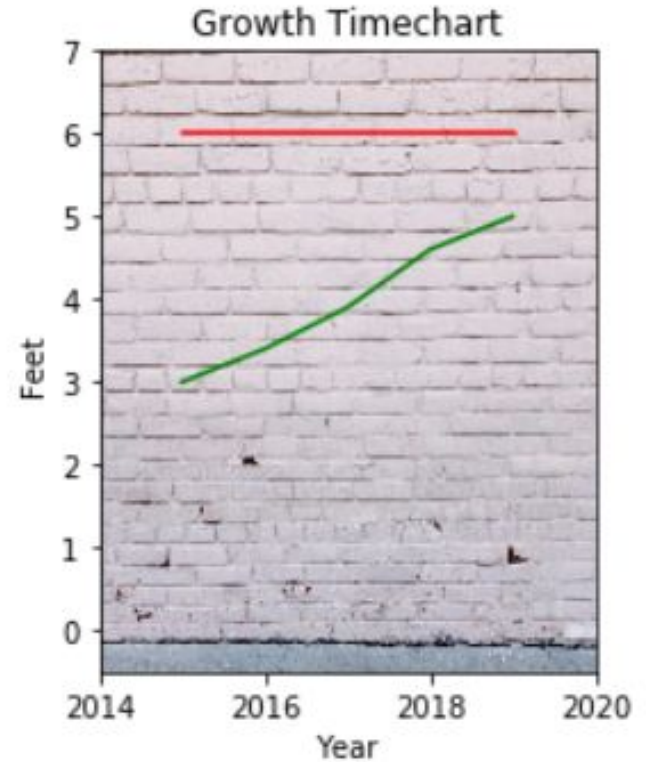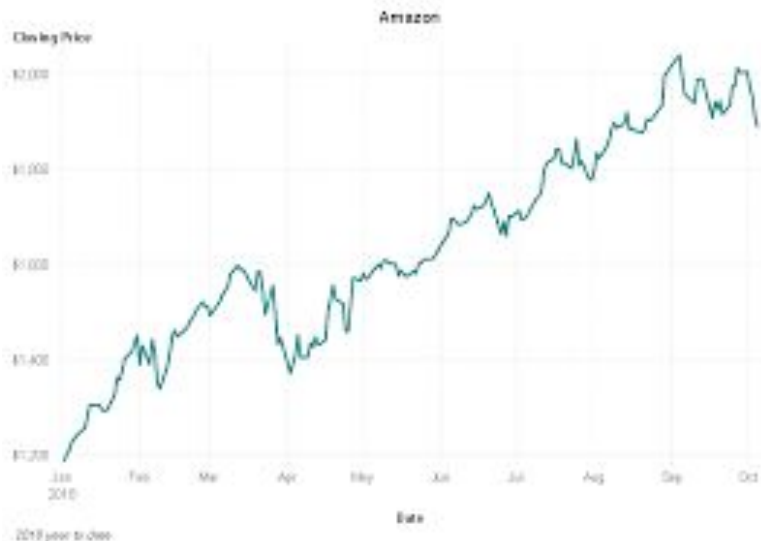
Trend Line

Use trend line to indicate
statistics characteristics
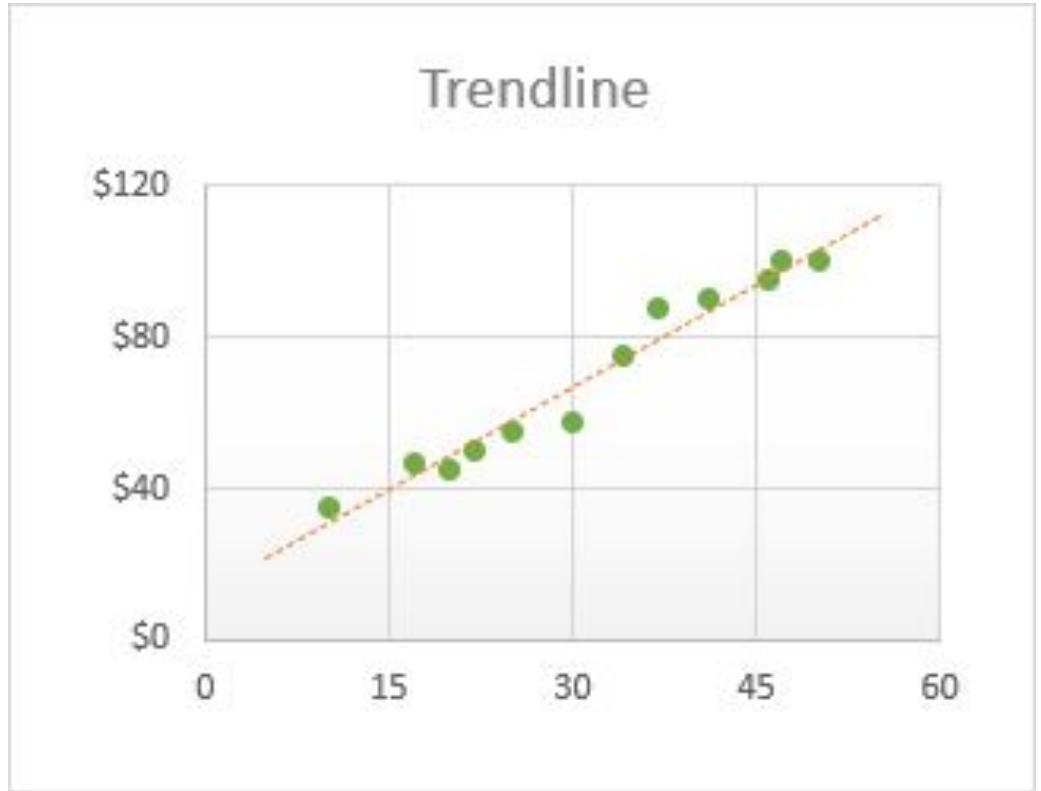
Pie chart

Percentage, portion: Only use it when the story is simple (or at the beginning of story)
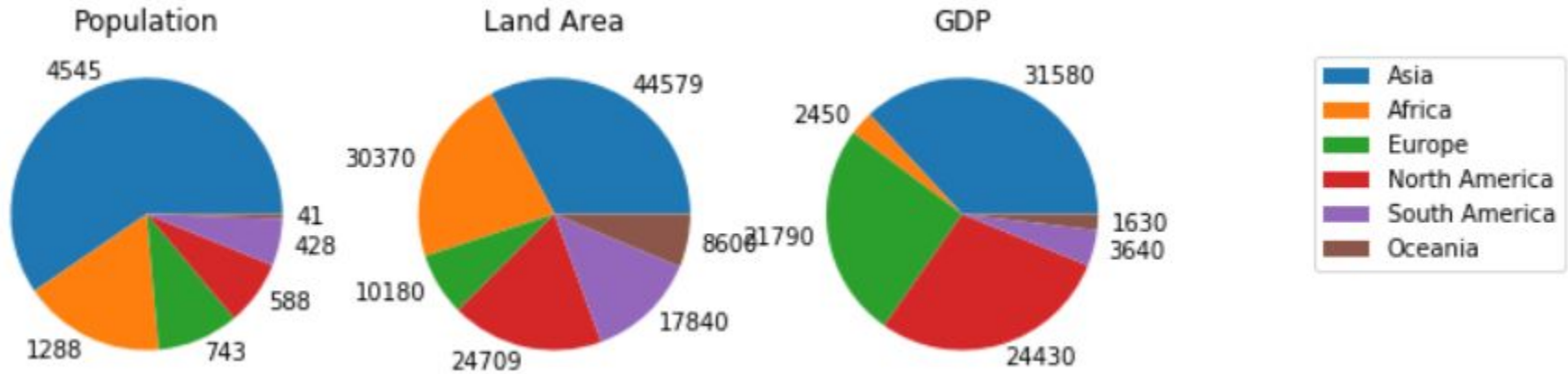
Pie chart

Avoid using pie charts when

- there are a large number of groups, or
- there are a number of groups with small percentages
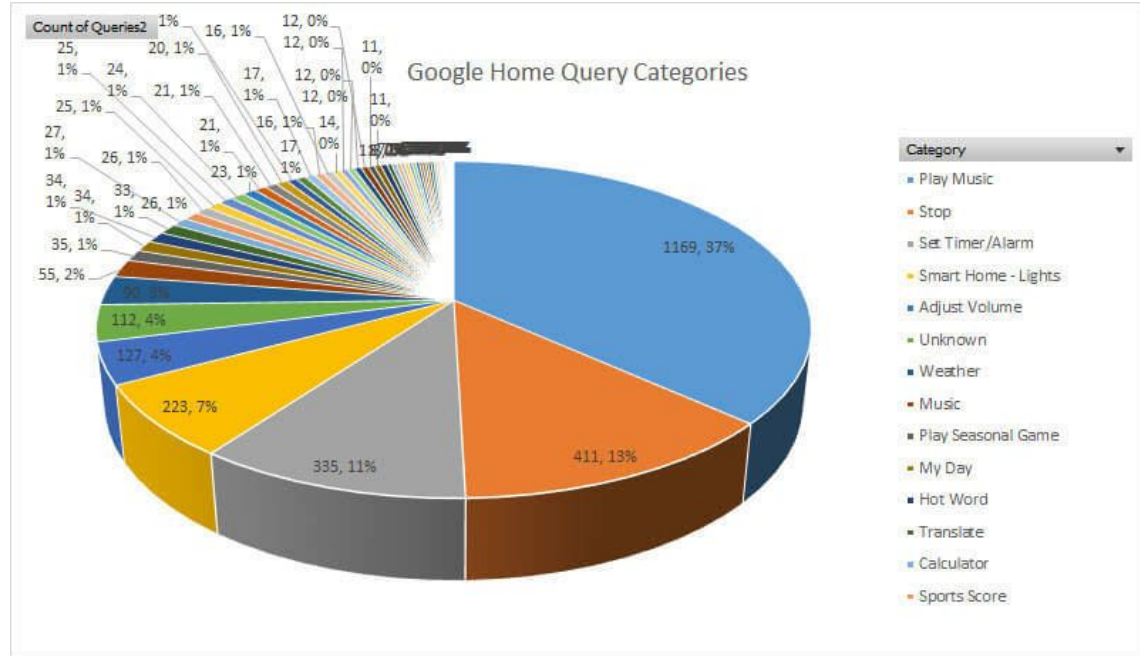
Background

Most common background: map

Scattered dot plotted based on longitude and latitude.
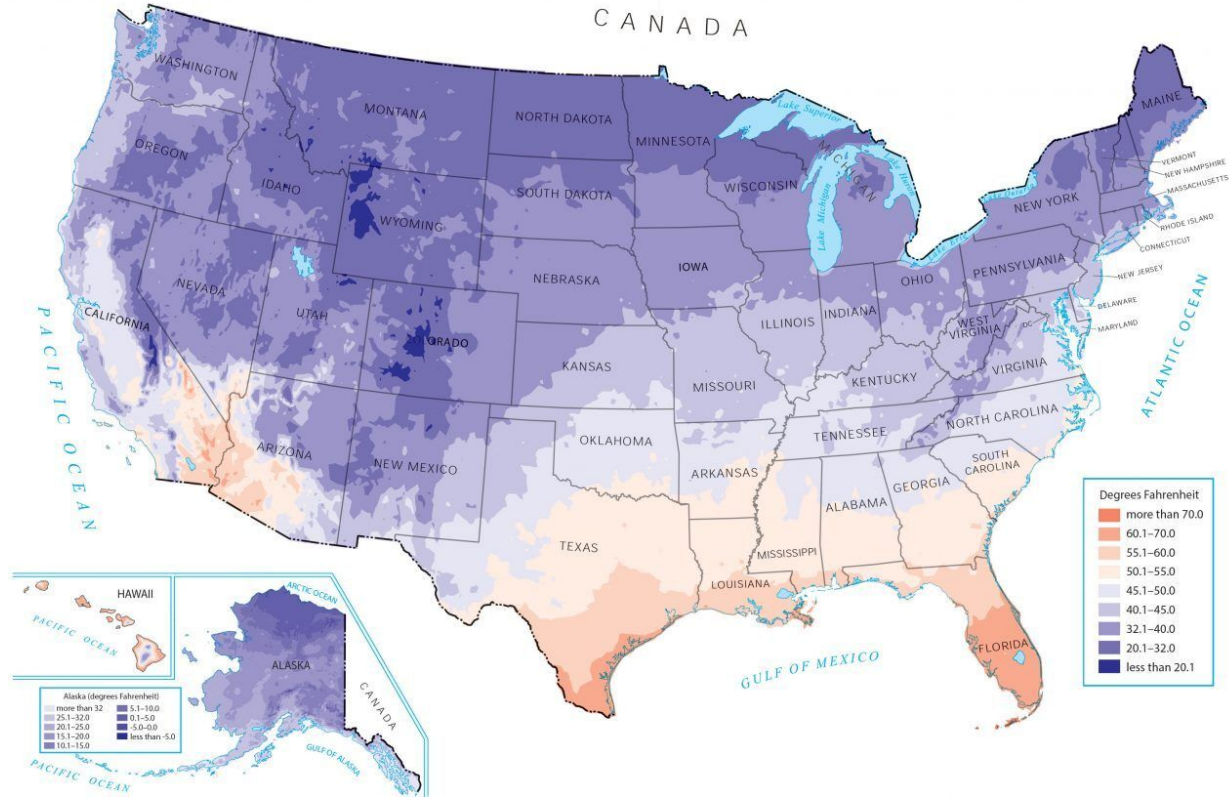
Agenda

- Visualization
  - Scattered Dot and Visual Characteristics
  - Visual Enhancements
  - Compare and Contrast
- The Good, the Bad, and the Ugly
  - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
  - Lab: Power BI

Compare and Contrast

Use same scale/background but different visualization characteristics (most common: color) to show the difference or similarity

Compare and Contrast

Subgroups within same chart

Good for contrast of discrete groupings and sub groupings
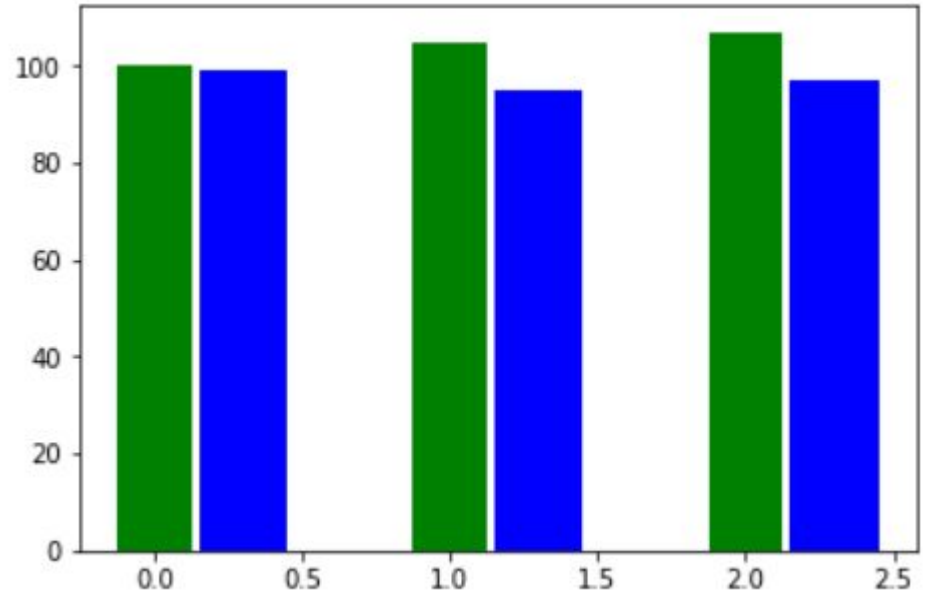
Compare and Contrast

Make sure you show clearly whether it is compare (similarity) or contrast (difference)

Note: the scale is different here, so that comparison is misleading.
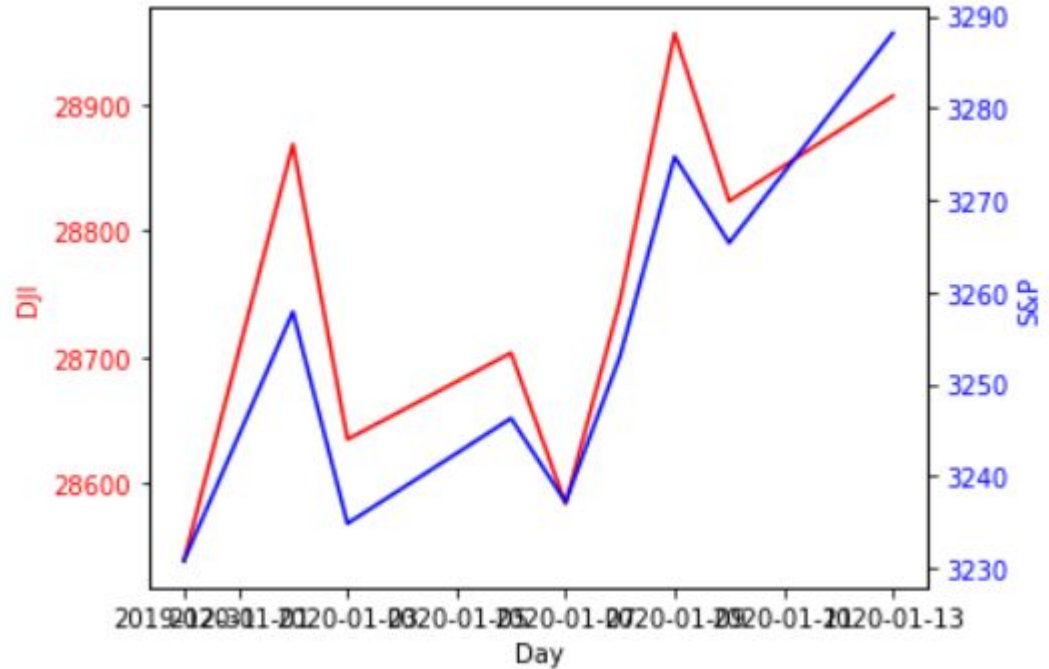
Heterogenous Chart

One data series is used as background for another

Bar-Line chart

- Bar: Actual numbers
- Line: thresholds, trends

Agenda

- Visualization
  - Scattered Dot and Visual Characteristics
  - Visual Enhancements
  - Compare and Contrast
- The Good, the Bad, and the Ugly
  - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
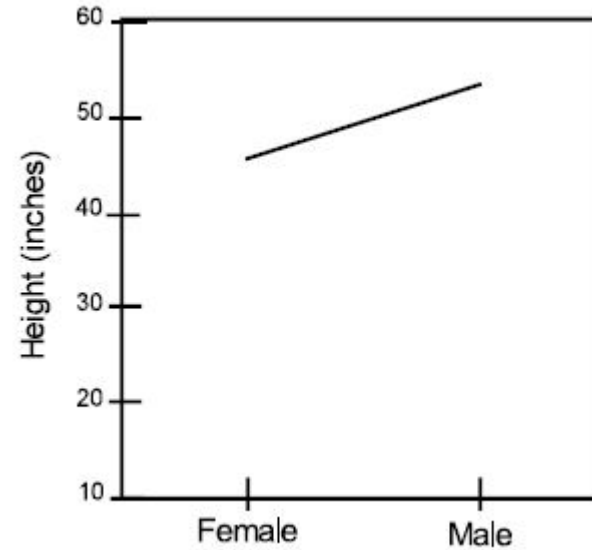 - Lab: Power BI

What is Data Visualization about?

Data visualization is about storytelling.

● Use a meaningful order. Avoid alphabetical order.
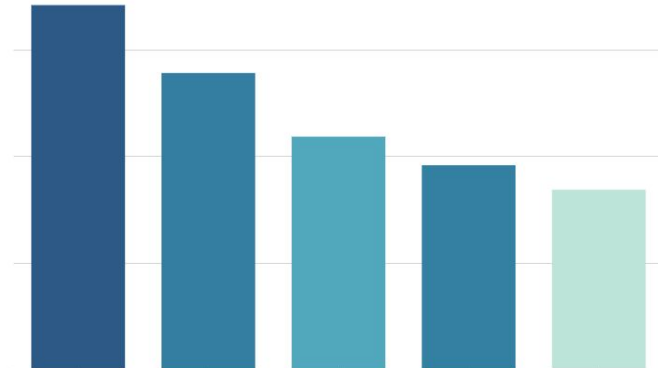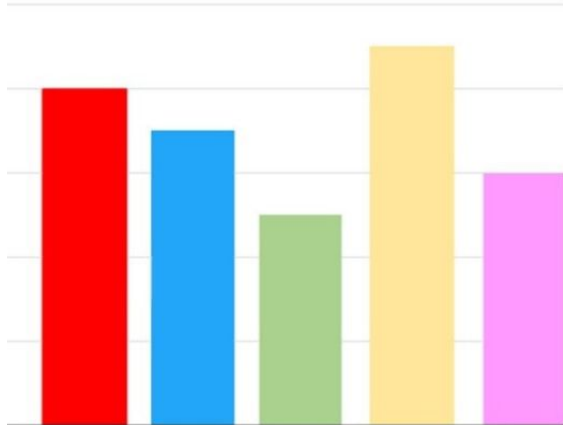
Discrete vs Continuous

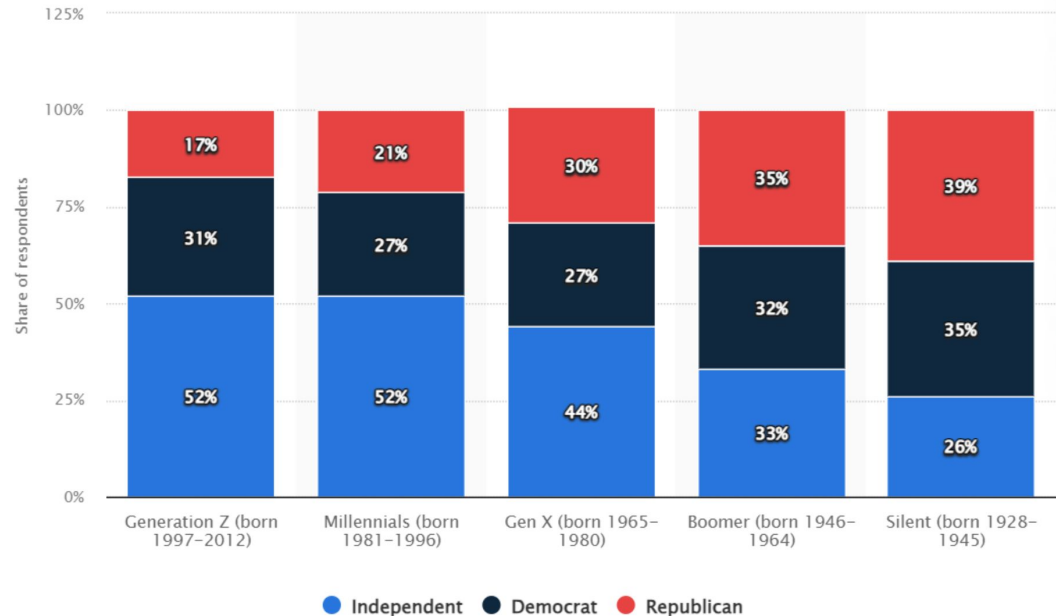Bar chart vs Line chart: Discrete or continuous

Follow Conventions

Color: Blue vs Red

Placement:
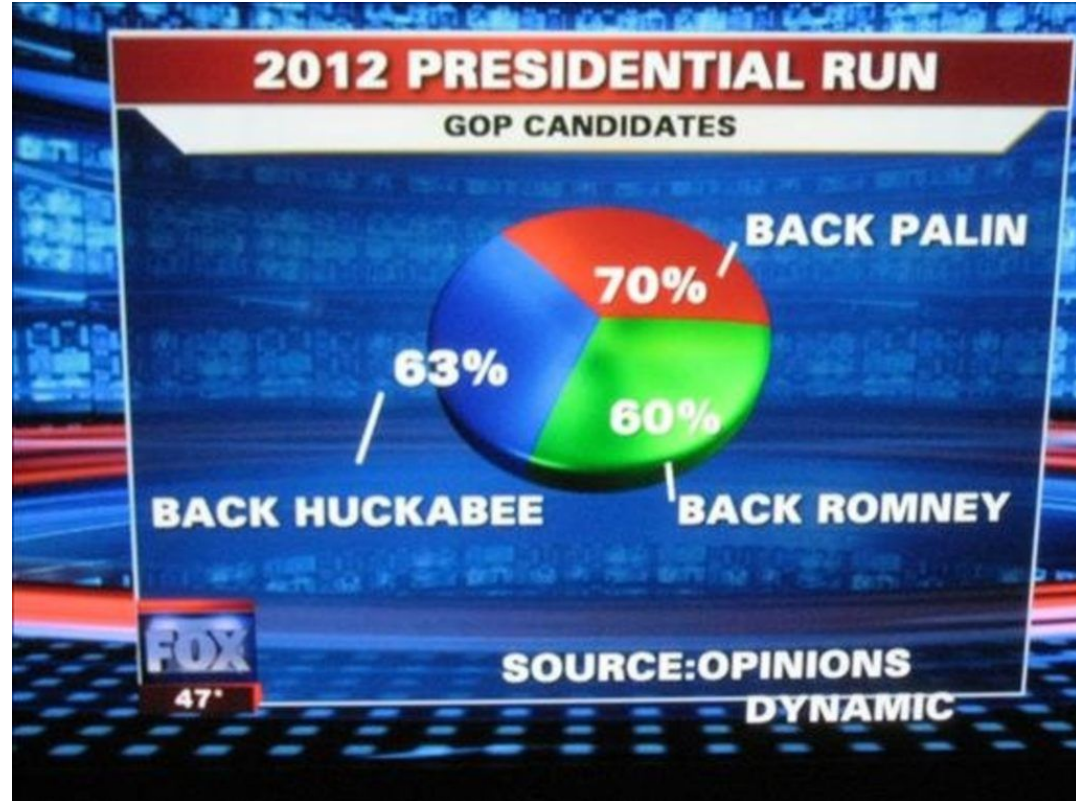independent in the
middle

Axis: Limit to 100%

Party identification in the United States in 2022, by generation

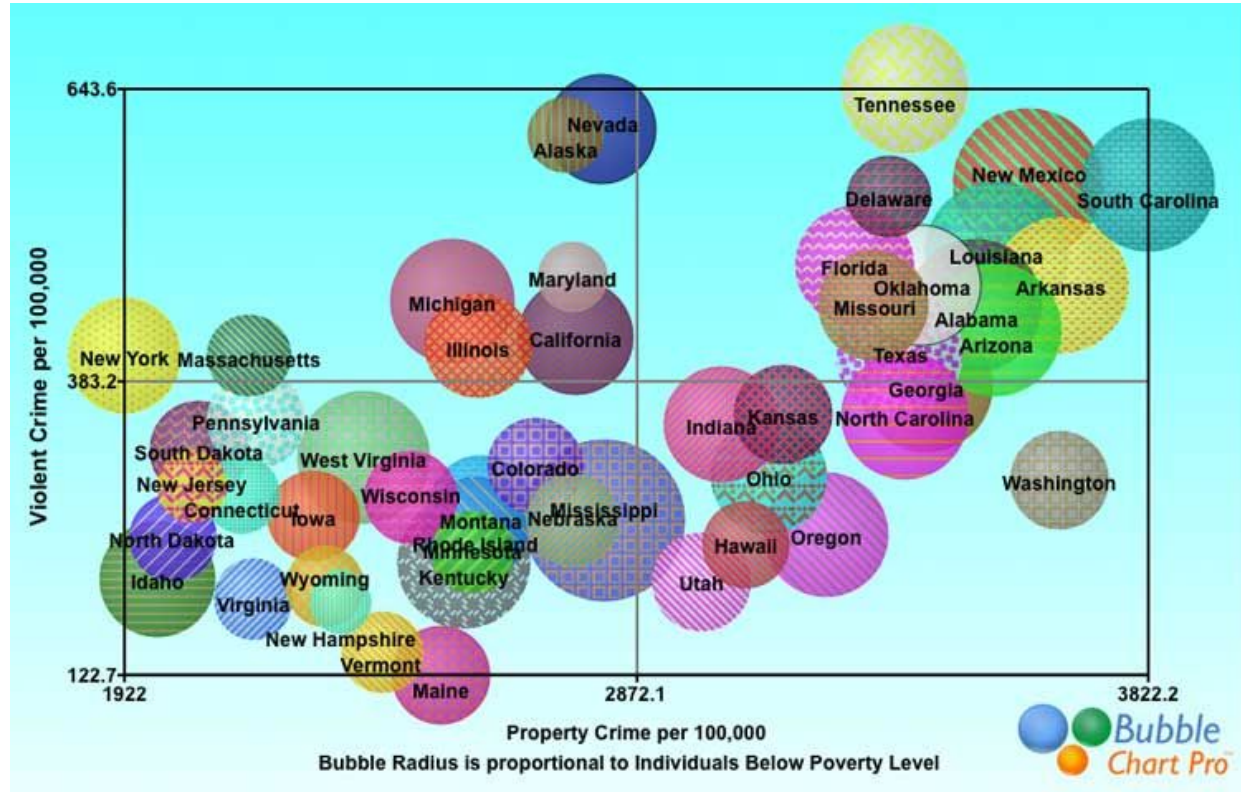# Choose the Right Type

63 % + 70% + 60% = 193%

- Pie charts should always add up to 100%
- If you use actual numbers, make sure to explain the meaning of total pie value

Colorful != Helpful

Color should be used for grouping or as an indicator

Do not apply one color for each scattered dot

Colorful != Helpful

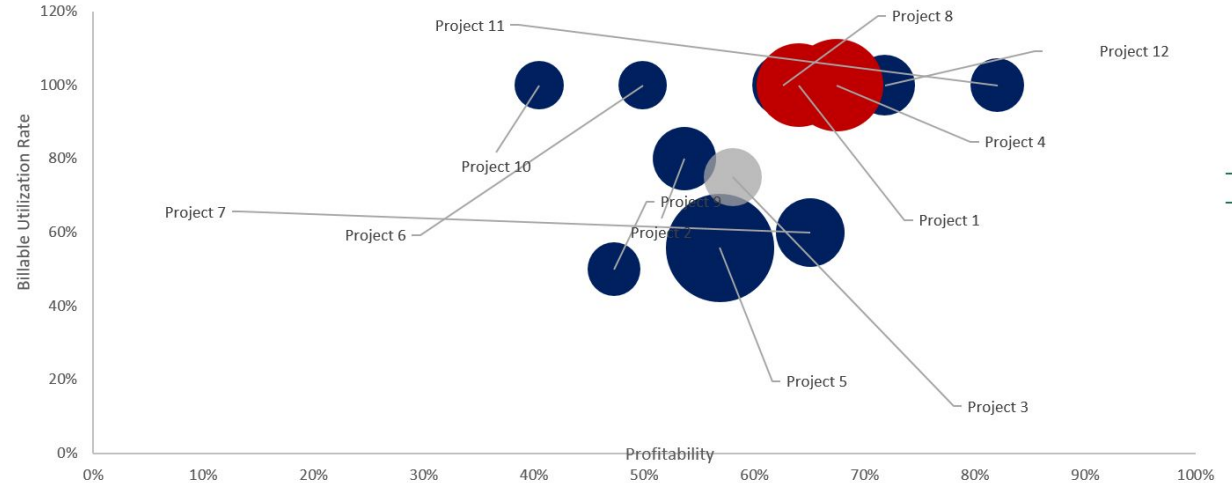Does red and gray mean anything special here?

Follow conventions:

red: alert, exception; grey: inactive

Project evaluation by BUT and Profitability
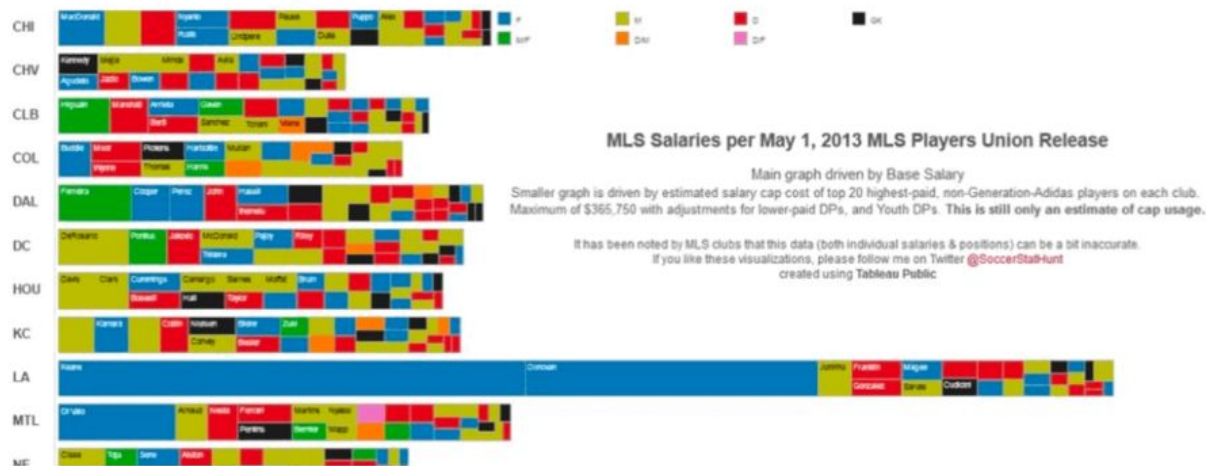All Customers by entity | 2021-W46 | Budget >400 hours

Rule of 3 and Rule of 7

Unless there is a simple and clear pattern,

- Do not show more than 3 types of visualization hints
- Do not show more than 7 data groupings

What is this visualization

trying to tell us?

MLS Salaries per May 1, 2013 MLS Players Union Release

Main graph driven by Base Salary
Smaller graph is driven by estimated salary cap cost of top 20 highest-paid, non-Generation-Adidas players on each club.
Maximum of $365,750 with adjustments for lower-paid DPs, and Youth DPs. This is still only an estimate of cap usage.

It has been noted by MLS clubs that this data (both individual salaries & positions) can be a bit inaccurate.
If you like these visualizations, please follow me on Twitter @SoccerStatHunt
created using Tableau Public

Use Map with Caution

Map background can be both a blessing and a curse

- Conveniently show geography related data
- Limited by the nature of geography
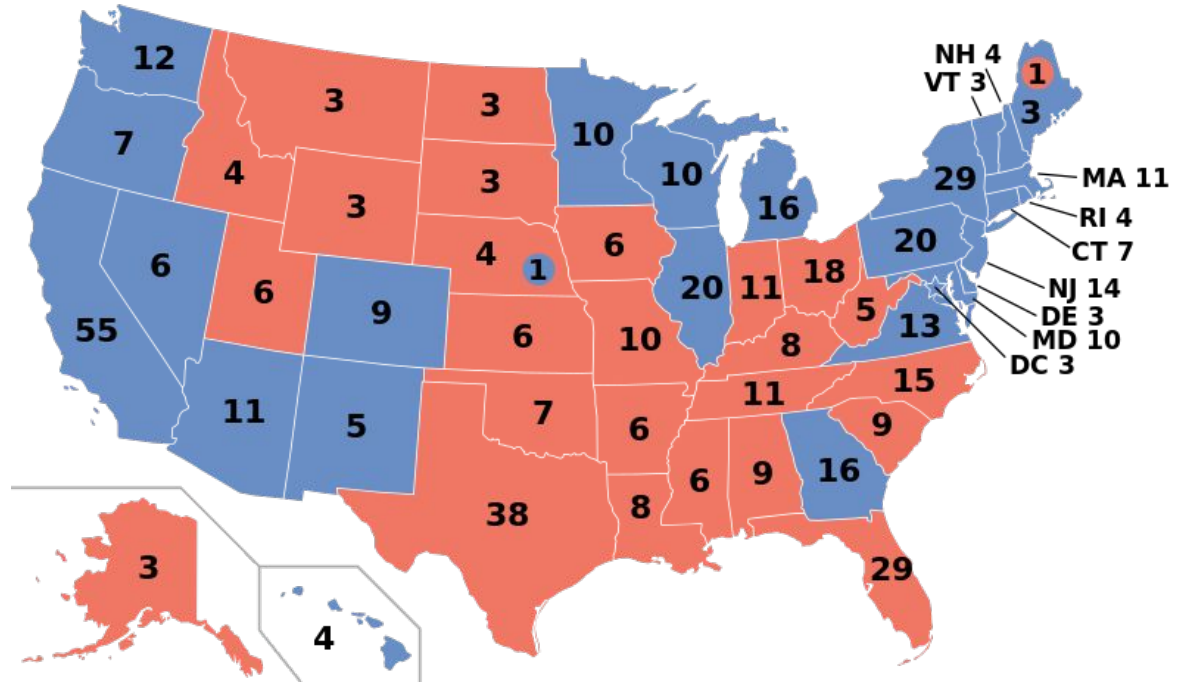
Agenda

- Visualization
  - Scattered Dot and Visual Characteristics
  - Visual Enhancements
  - Compare and Contrast
- The Good, the Bad, and the Ugly
  - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
  - Lab: Power BI

The Greatest Visualization in Military History

When we understand that slide, we'll have won the war

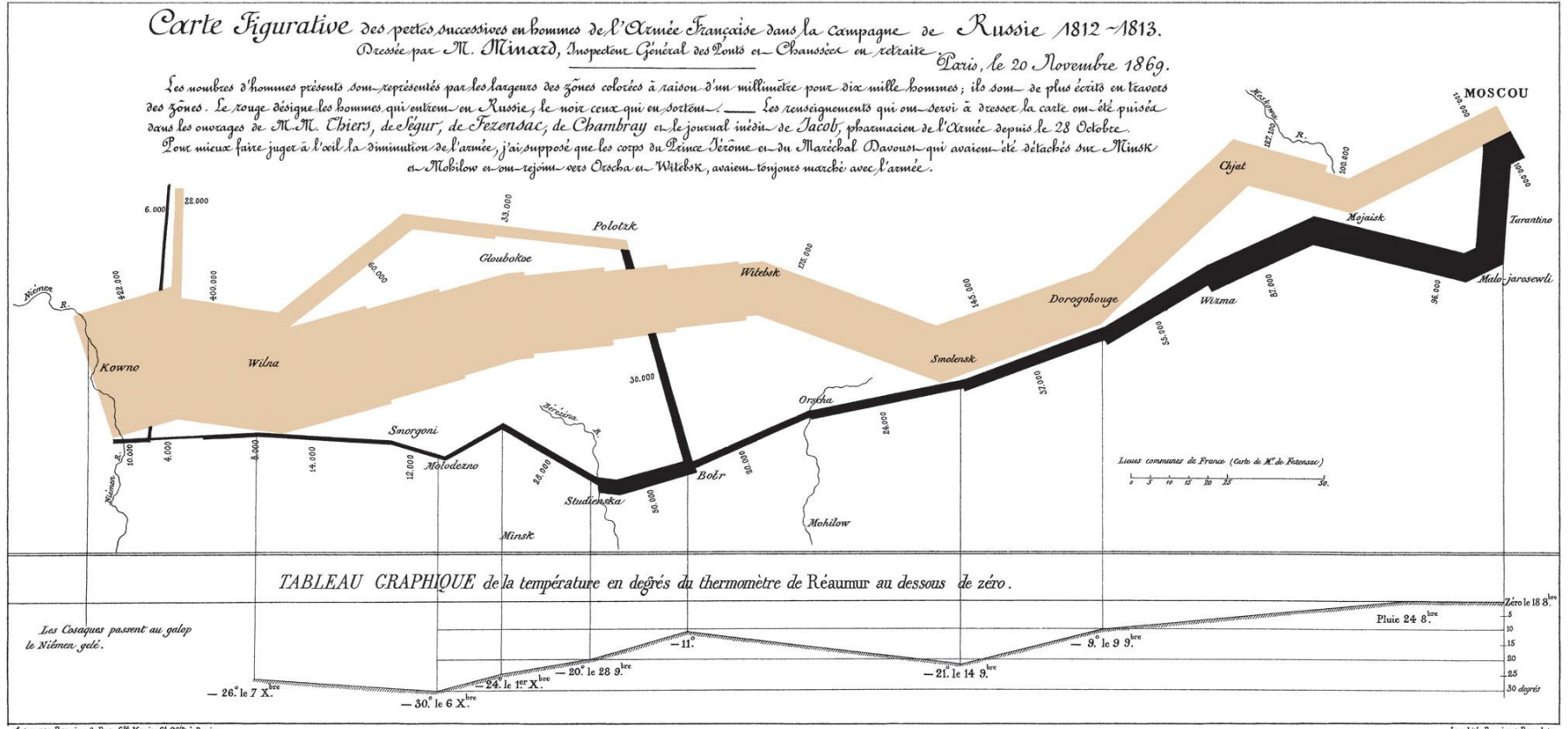– US military General Stanley McChrystal

Afghanistan Stability / COIN Dynamics

WORKING DRAFT – V3

PA Consulting Group
© PA Knowledge Limited 2009
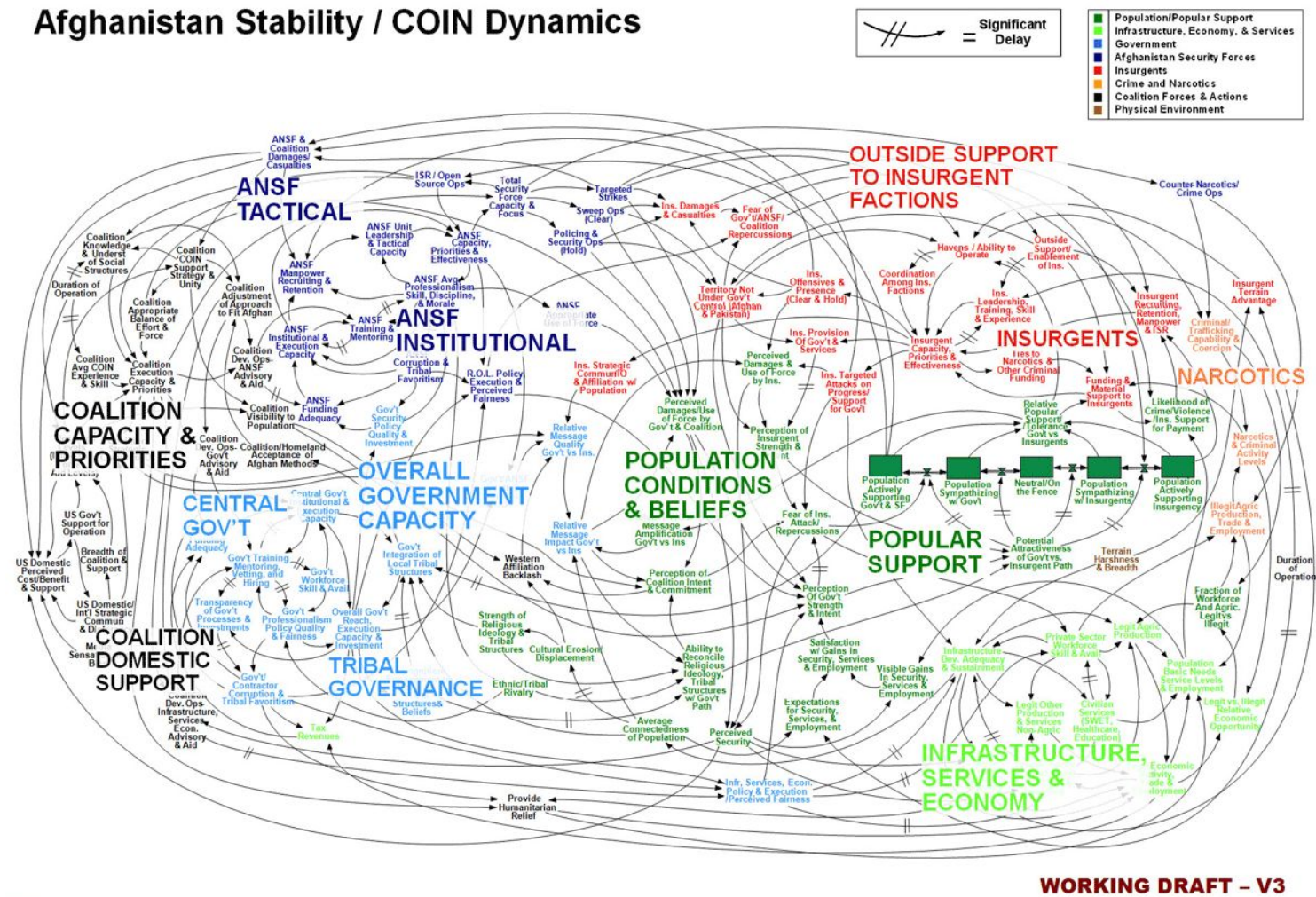
The Greatest Visualization in Military History - The Real One

"Carte figurative des pertes successives en hommes de l'Armée française dans la campagne de Russie en 1812-1813", is a graphical depiction of the losses of French Army during Napoleon's ruinous invasion of Russia of 1812-13, drawn in 1869 by Charles Joseph Minard, a former French civil engineer.

The Greatest Data Visualization in History

The Greatest Data Visualization in History

Why is it Great?

- How to evaluate a military operation?
  - Land (geographical measurement)
  - Size of force
- How are they visualized?
  - Map (and implicitly, time)
  - Size (width)
  - Color (grouping into 2; advance and retreat)

Agenda

- Visualization
  - Scattered Dot and Visual Characteristics
  - Visual Enhancements
  - Compare and Contrast
- The Good, the Bad, and the Ugly
  - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
  - Lab: Power BI

How to Tell a Complex Story

If the story is beyond one visualization

- Define a story line with multiple episodes
- Create one visualization for each episode
- Arrange in a logical order
- Arrange in separate pages/tabs and use links to connect the story line

Element Placement

In societies where writing starts from left to right, human eyes usually

- start at the upper left corner,
- then horizontally move to the right,
- then move back to upper left corner and
- then moves down,
- then take a glance at the bulk page
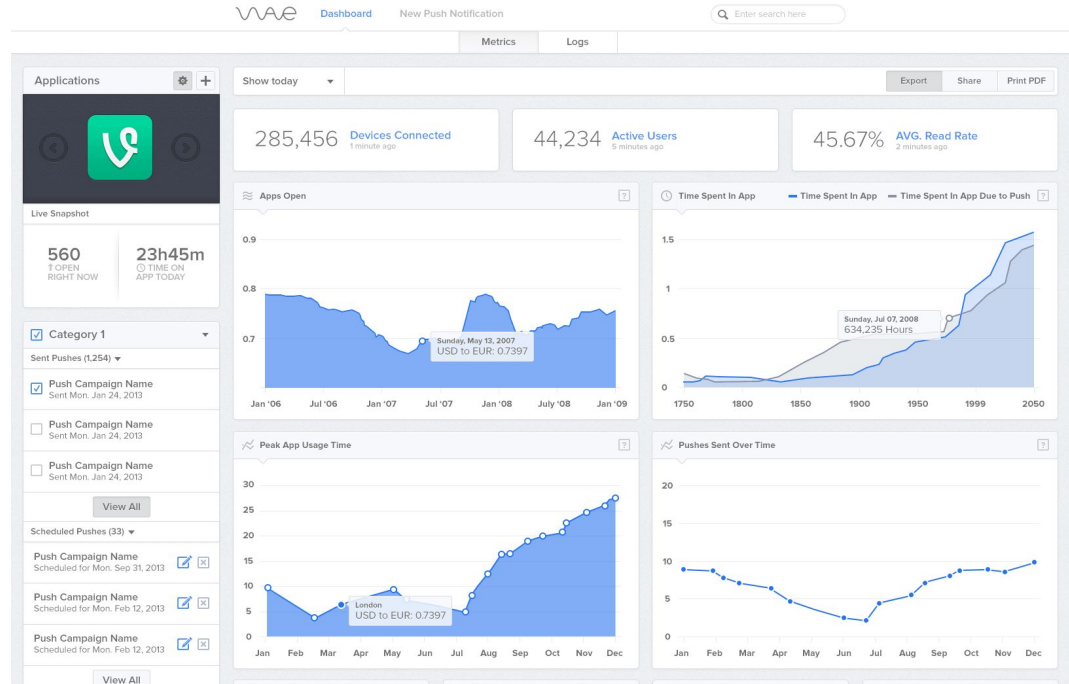
Element Placement

- Upper left corner: titles, icons
- Top horizontal bar: key indicators
- Left vertical bar: break downs, links
- Main area: visualizations, maps
- Right vertical bar: filters, notes
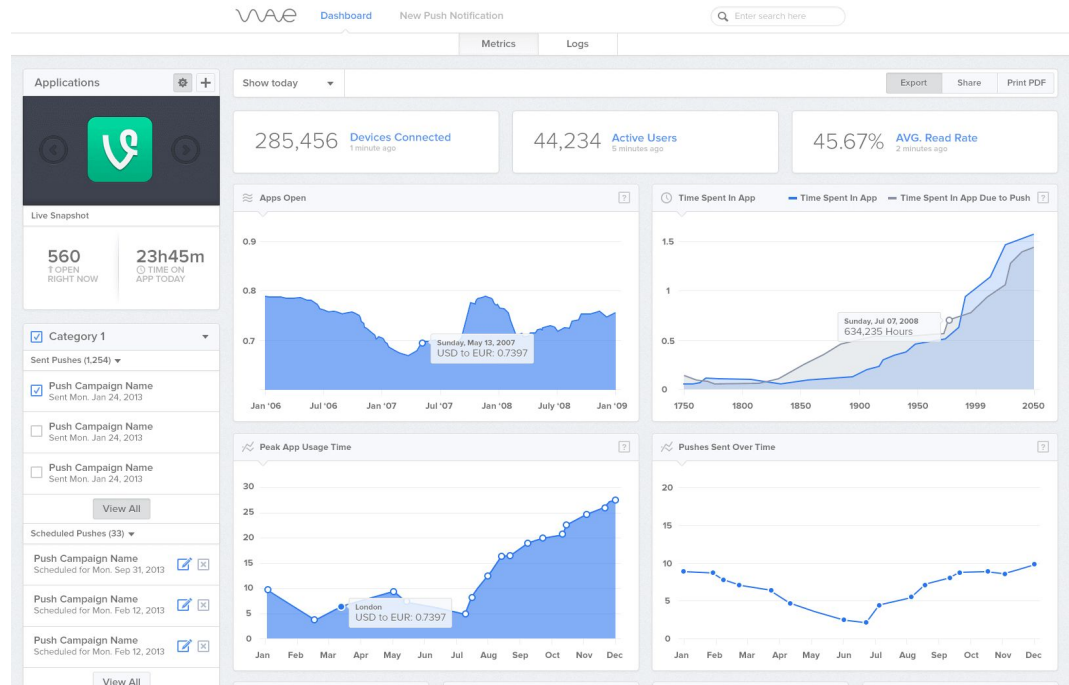
Agenda

- Visualization
    - Scattered Dot and Visual Characteristics
    - Visual Enhancements
    - Compare and Contrast
- The Good, the Bad, and the Ugly
    - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
    - Lab: Power BI

Microsoft Power BI

Microsoft's flag business intelligence tool, capable of different levels of business intelligence reporting

Dashboard using Power BI
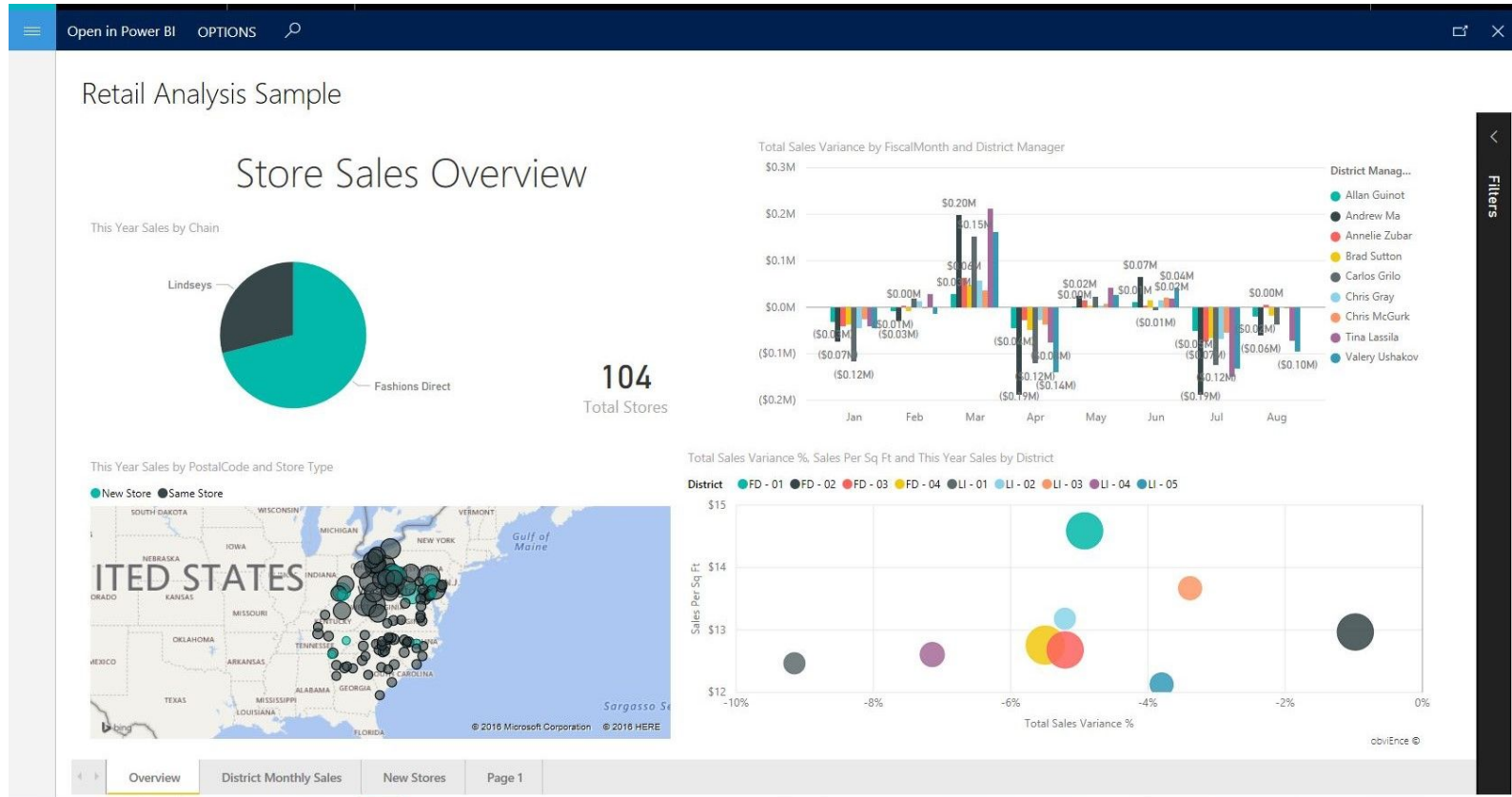
Power BI Operational Report Example

| Category	Status	Avg Price	Last Year	This Year	Goal
100-Groceries	🟢	$1.36	$810,176	$829,776	$810,176
090-Home	🟢	$3.28	$2,913,647	$3,053,326	$2,913,647
080-Accessories	🟢	$4.22	$1,273,096	$1,379,259	$1,273,096
070-Hosiery	🔴	$3.57	$573,604	$486,106	$573,604
060-Intimate	🔴	$4.02	$955,370	$852,329	$955,370
050-Shoes	🟡	$13.73	$3,640,471	$3,574,900	$3,640,471
040-Juniors	🔴	$7.06	$3,105,550	$2,930,385	$3,105,550
030-Kids	🟡	$5.20	$2,726,892	$2,705,490	$2,726,892
020-Mens	🟡	$6.89	$4,453,133	$4,452,421	$4,453,133
010-Womens	🔴	$6.70	$2,680,662	$1,787,958	$2,680,662
Total	🟡	**$5.19**	**$23,132,601**	**$22,051,952**	**$23,132,601**

Microsoft Power BI

Navigate through multiple tabs

Agenda

- Visualization
  - Scattered Dot and Visual Characteristics
  - Visual Enhancements
  - Compare and Contrast
- The Good, the Bad, and the Ugly
  - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
  - Lab: Power BI

Microsoft Power BI

Interactive data visualization software product developed by Microsoft with a primary focus on business intelligence

Power BI

Find insights, track progress, and make decisions faster using rich visualizations.
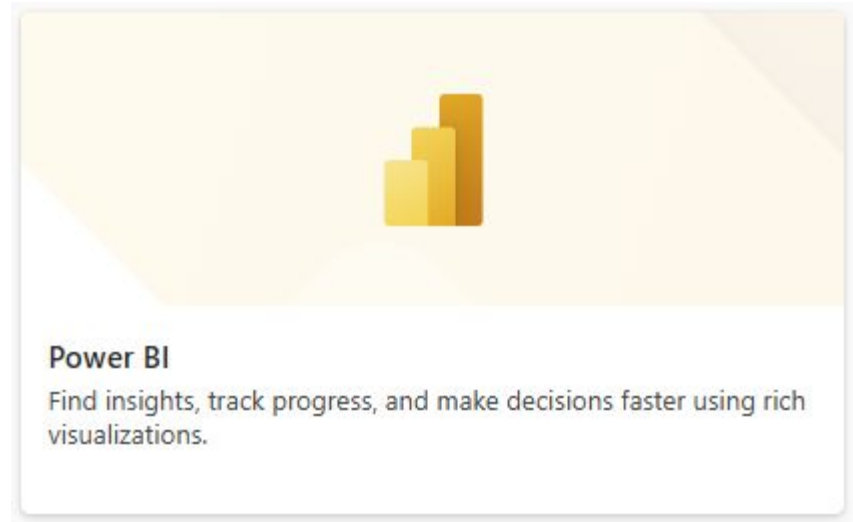
Microsoft Power BI Dashboard

Microsoft Learn
Intro to dashboards for Power BI ...

phData
Web Traffic Power BI Dashboard Example ...

Reddit
Made a modern UI looking kind of ...

Microsoft Learn
Dashboards for business users of the ...

Coupler.io Blog
24 best Power BI dashboard examples in ...

Coupler.io Blog
24 best Power BI dashboard example...

webdashboard.com
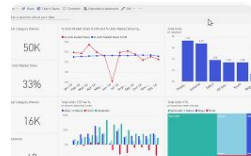Power BI Sharing - Webdashboard

Reddit
Made a bunch of Pow...

Simplilearn.com
Power BI Dashboard [Updted ...

Cobit Solutions
Top 12 Power BI Dashboard Reports ...

Zebra BI
Power BI Dashboard Design: Avoid The...
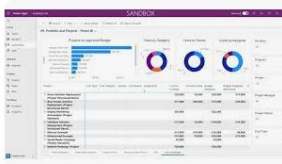
Microsoft Learn
View a Power BI dashboard - Power ...

Numerro
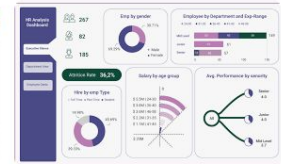Top 10 Power BI Dashboard Exampl...

Windsor.ai
Paid Media Overview Dashboard Template

Alpha Serve
Power BI Dashboard Templates ...

BrightWork.com
Power BI Project Management Dashboards ...

Mammoth Analytics
Power BI Templates: The 9 Best Ones ...

ZoomCharts
Financial Dashboards in Power BI

BrightWork.com
Power BI Project Management Dashboards...

Data for Microsoft Power BI

Power BI can get data directly from databases, data warehouses, and/or data lakehouses.

However, the original storage format (star-schema and relational) may not directly fits into the analysis needs.

Data can be stored into a cube, which is implemented based on analytical needs.
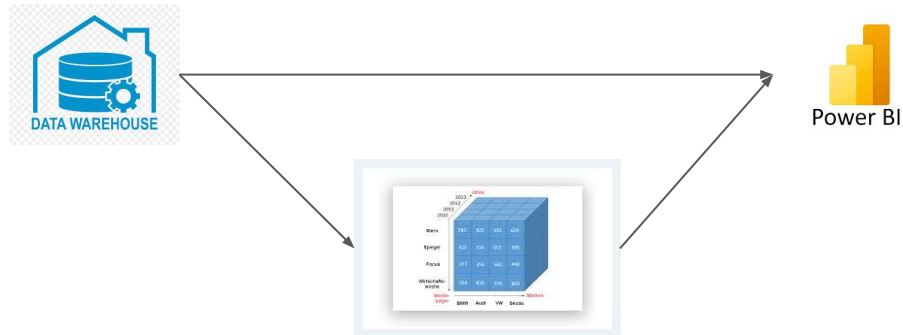
Online Analytical Processing (OLAP)

Online analytical processing (OLAP) is software technology you can use to analyze business data from different points of view. Organizations collect and store data from multiple data sources, such as websites, applications, smart meters, and internal systems.
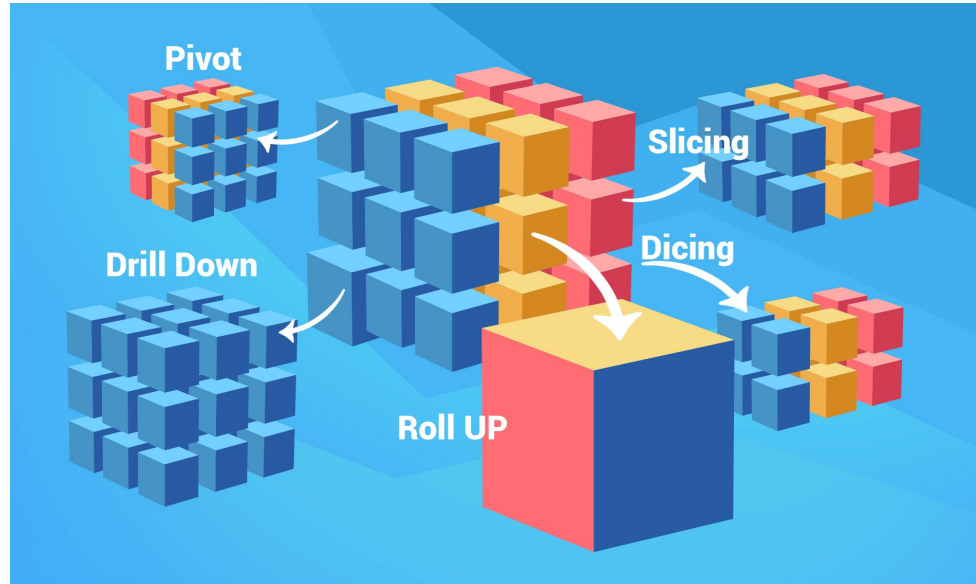
OLAP Cube

A data structure that overcomes the limitations of relational databases by providing rapid analysis of data. Cubes can display and sum large amounts of data while also providing users with searchable access to any data points. This way, the data can be rolled up, sliced, and diced as needed to handle the widest variety of questions that are relevant to a user's area of interest.
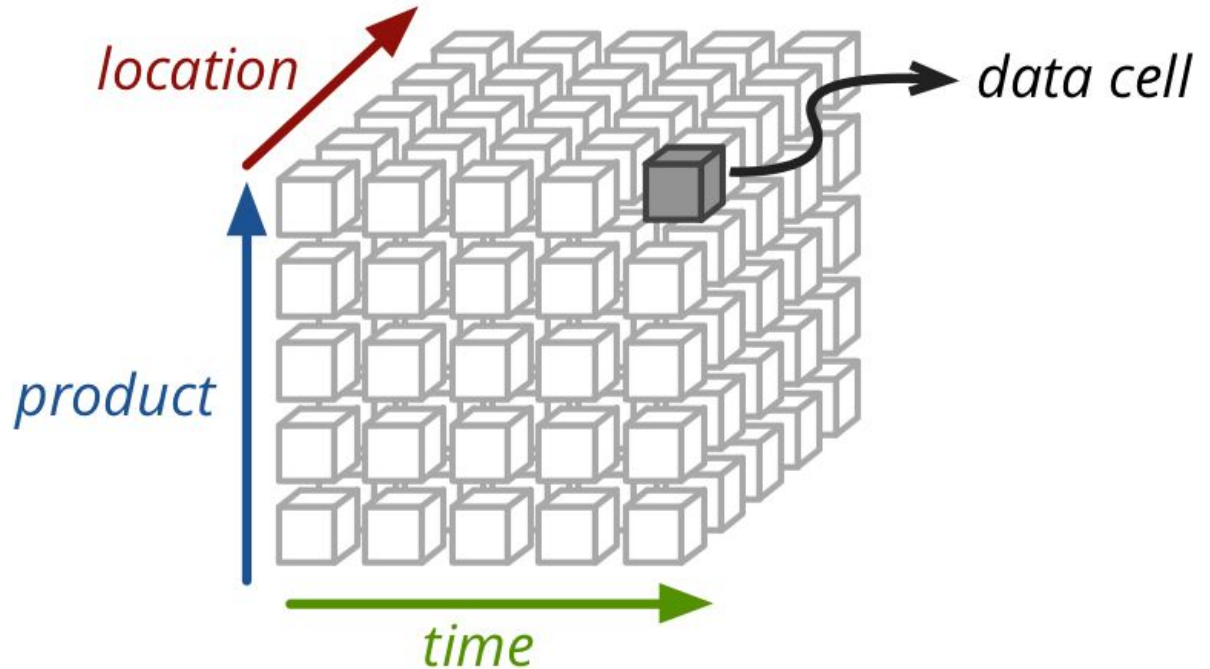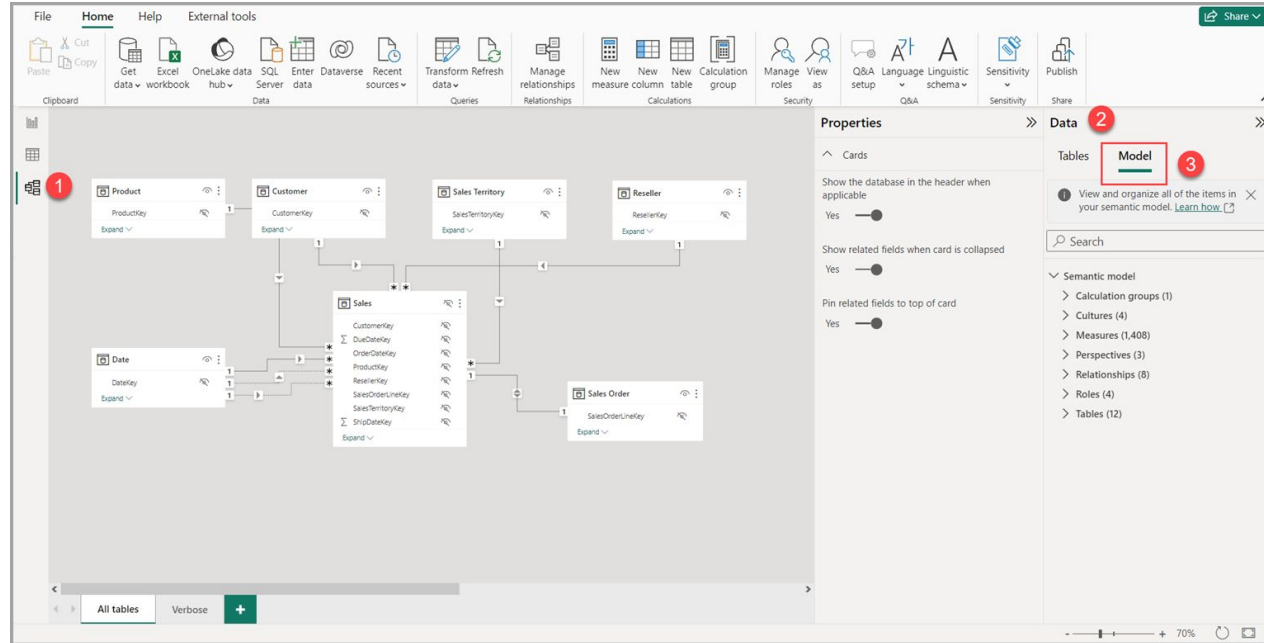
Building OLAP Cube

Data in OLAP cubes eventually comes from the data warehouse/lakehouse.

They are combined into one chunk of data, with indexes and labels
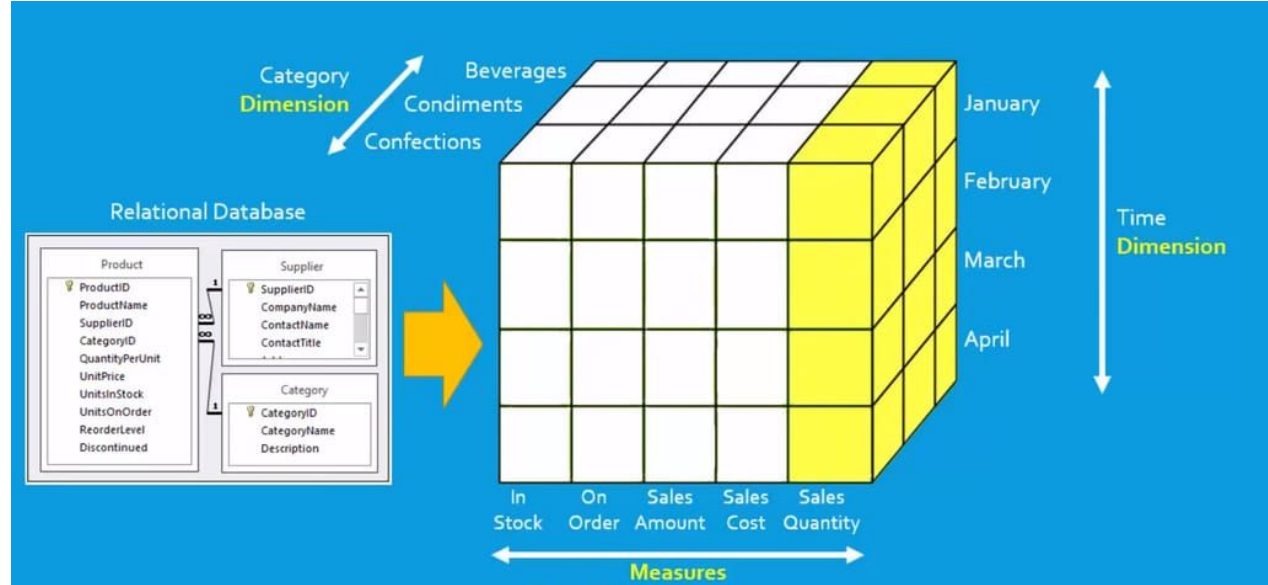
Cube in Power BI

Cube in Power BI is called Semantic Model

Different OLAPs

Depending on where the cube stores its data, there are 3 different OLAPs

- MOLAP (Multidimensional OLAP): Pre-computes and stores data in a multidimensional cube format, leading to fast query responses but requiring more storage space.
- ROLAP (Relational OLAP): Stores data in traditional relational databases and computes the results dynamically, allowing for larger datasets but with slower performance compared to MOLAP.
- HOLAP (Hybrid OLAP): Combines the benefits of MOLAP and ROLAP by storing part of the data in cubes and part in relational databases to balance performance and storage needs.

Storage Mode for Power BI Semantic Model

Power BI Semantic Model supports all OLAPs with Data Storage Mode setting

- Import Mode (MOLAP): Data is imported into the Fabric Power BI system and is stored on disk, but must be fully loaded into memory when queried or refreshed
- DirectQuery Mode (ROLAP): Stores only metadata that defines the model's structure. When queries are made, data is retrieved directly from the underlying source using native queries.
- DirectLake Mode (ROLAP): Bring data back whenever requested, instead of saving it in the Power BI system. It fetches data directly from the delta files instead of going through a relational query interface

Allows fallback between modes (HOLAP)

Current Status of Power BI Semantic Model

Power BI Semantic Model in Fabric is still under development

- Storage Mode can only be specified in Power BI Desktop, not in browser
- When created from scratch, Semantic Model can only get data from files
  - To use lakehouse tables, semantic models must be created from lakehouse interface
- Each semantic model can only come from one lakehouse
  - If you need data from tables in multiple lakehouses, create a lakehouse for this purpose, and use shortcuts to map to those tables – Which is the goal of building gold layer in Medallion Architecture
- Other minor usage discrepancies

What is a Power BI Semantic Model

- Definition
  - A collection of tables and their relationships
- Data
  - Data fetched from these tables (MOLAP), or
  - How to fetch the data (ROLAP)
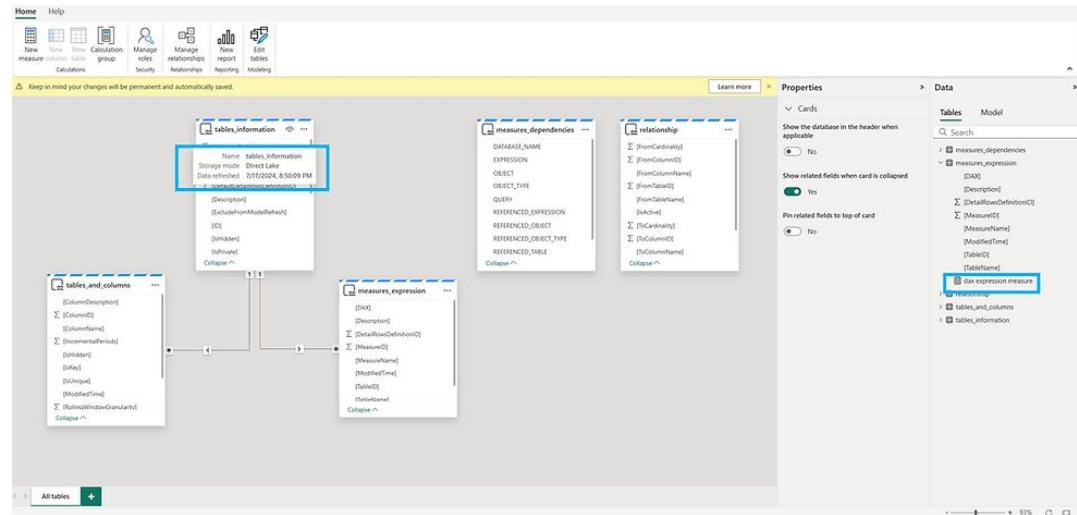
Using Power BI Semantic Model

- Create Semantic Model from lakehouse
  - Tables are introduced into the model at this step
- Setup relationship among tables
- Use this semantic model in Power BI

Agenda

- Visualization
  - Scattered Dot and Visual Characteristics
  - Visual Enhancements
  - Compare and Contrast
- The Good, the Bad, and the Ugly
  - The Greatest Visualization in Military History
- Telling a Complex Story
- Microsoft Power BI
  - Lab: Power BI

Power BI Lab

- Open Power BI in Fabric
- Create Semantic Model
- Create grid report
- Create scattered dot chart
- Create bar chart
- Create line chart

Final Project

Review Assignment 08

- Table creation

Assignment 09

- Visualize the data you got from previous assignments

Final Thoughts about Data Visualization

It is all about storytelling!

Design your visualization based on your story, not the other way around!

The Greatest Visualization in Military History

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

What did Napoleon need from a visualization

10000

412000

Assignment