CIS 9440 - Data Warehousing and Analytics

Class #10

Class note

Thanksgiving Eve class, 11/23/22, will be hosted remote via Zoom. No in-person class on this day.

Week 10 Class Overview:

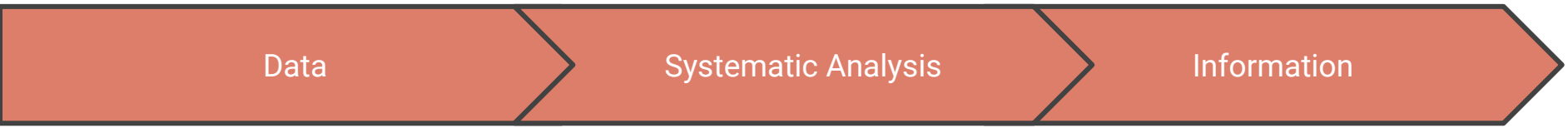
1. Introduction to Analytics
2. Introduction to Business Intelligence (BI)
3. Analyses by Data Type
4. BI Application Best Practices
5. BI Workshop #1

Week 10 Class Overview:

1. **Introduction to Analytics**
2. Introduction to Business Intelligence (BI)
3. Analyses by Data Type
4. BI Application Best Practices
5. BI Workshop #1

What are Analytics?

Analytics: information resulting from the systematic analysis of data or statistics.



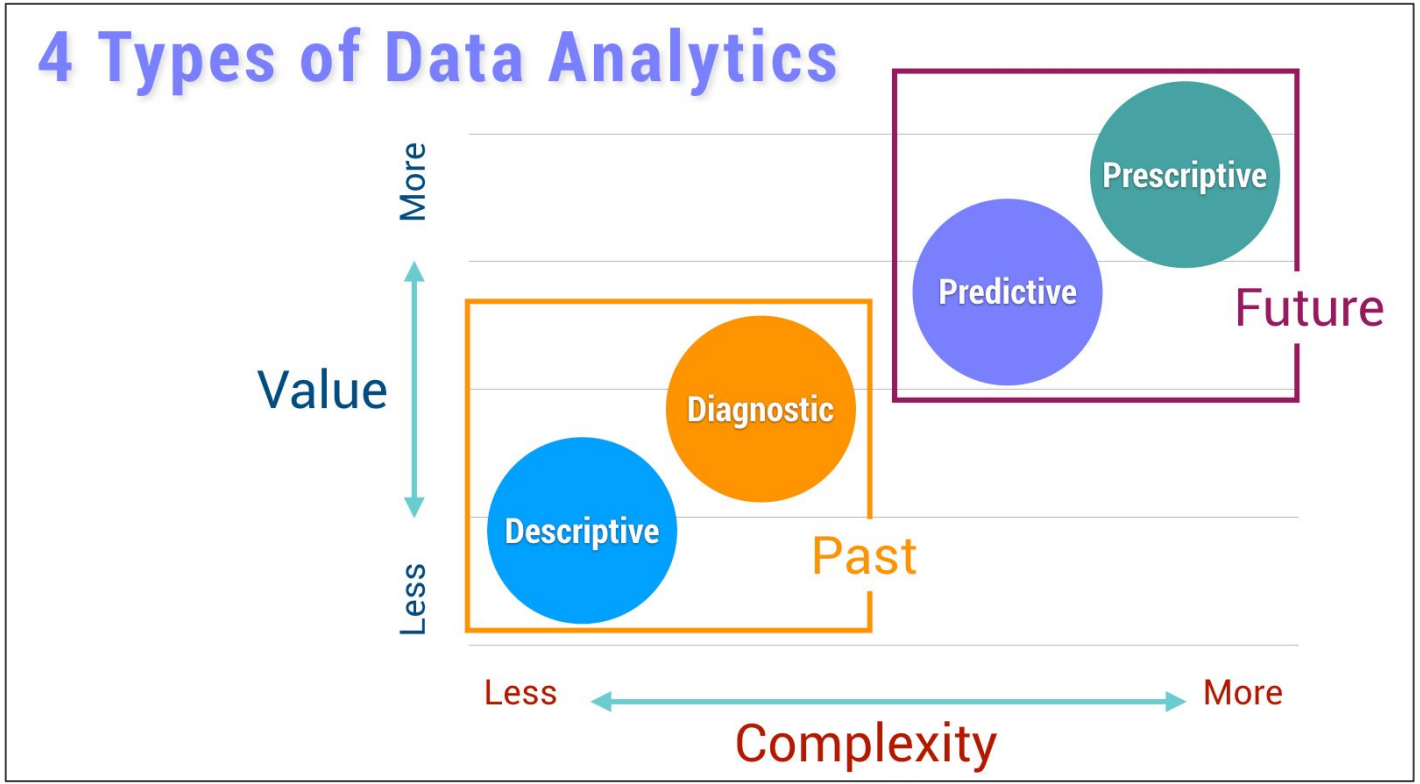
Example: NYC Ferry Ridership

KPI - Total Rides by Weekday

On average, Monday's are the most popular day to ride the NYC Ferry

- **But what type of analytics is this?**

Types of Analytics

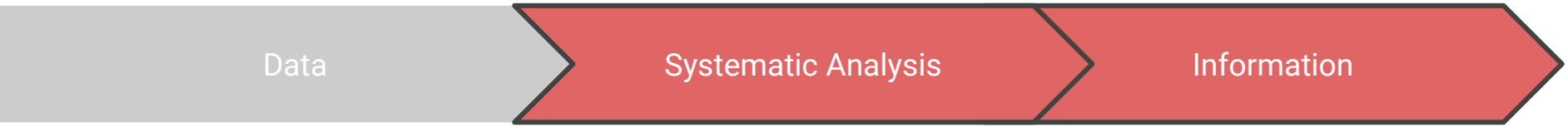


Audience by Type of Analytics

- **Descriptive:** operational audiences, “do these numbers indicate the operation is functioning as expected?”
- **Diagnostic:** questioning/business audiences
- **Predictive:** exploratory audience, “what would your prediction mean for the organization?”
- **Prescriptive:** decision-making audiences

Does Analytics relate to Business Intelligence (BI)?

Business Intelligence provides an **efficient** way to systematically analyze dimensionally modeled data and deliver the resulting information **effectively**.



NYC Ferry Ridership data
(Dimensional Model - ETL)

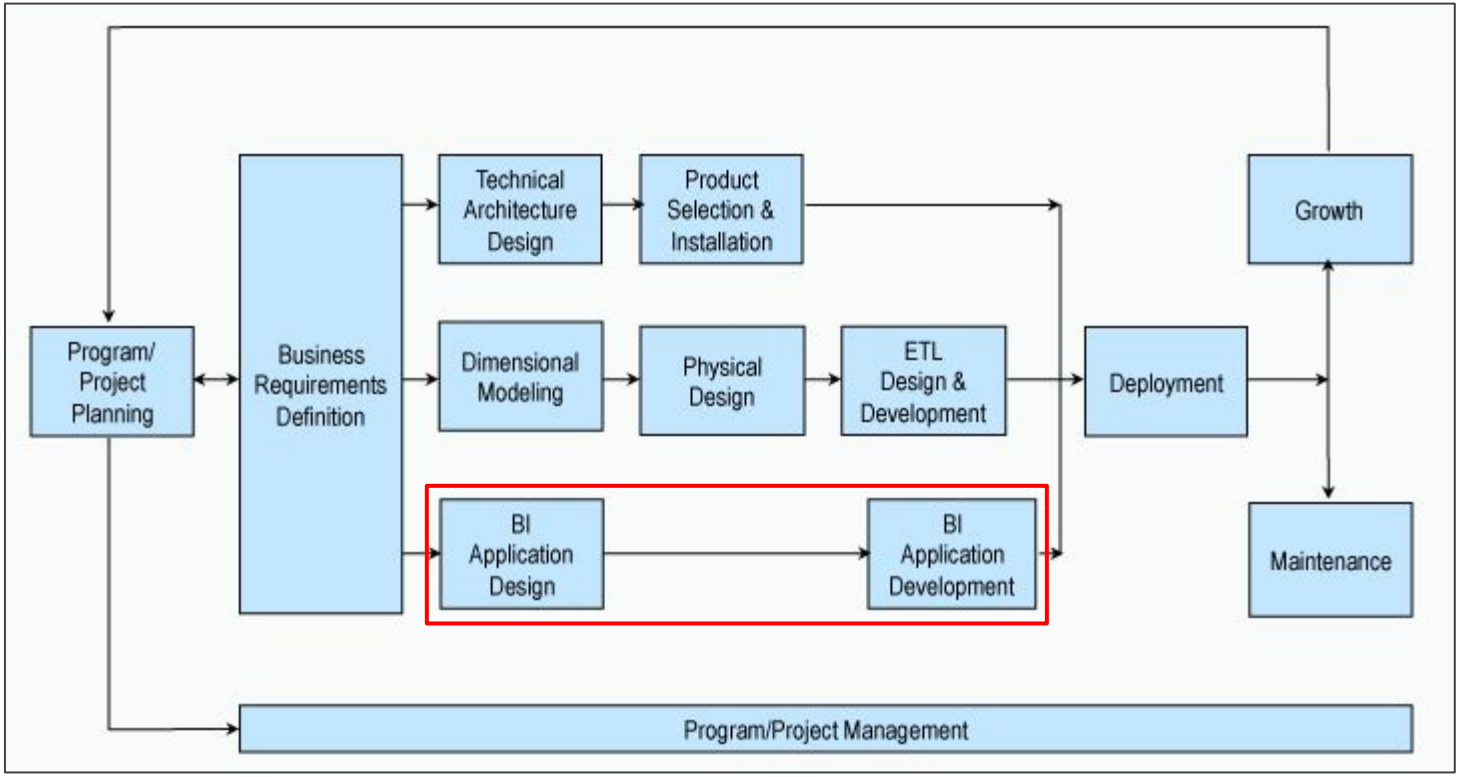
KPI - Total Rides by Weekday
(BI)

Bar Chart of Total Rides by Weekday
(BI)

Week 10 Class Overview:

1. Introduction to Analytics
2. **Introduction to Business Intelligence (BI)**
3. Analyses by Data Type
4. BI Application Best Practices
5. BI Workshop #1

Business Intelligence (BI) in the Kimball Lifecycle



Notice, you cannot start “BI Application Development” (actually building your BI Applications) until both the Middle and Top Kimball Lifecycle tracks are complete.

Business Intelligence Requirements

Two required components:

1. **Transformed Data**
2. Business Intelligence Platform

Business Intelligence Platform

A Business Intelligence platform is **technology** that *(when used correctly)* **makes analytics of your organizations data highly efficient**.

Business Intelligence Platform Examples

- Microsoft Power BI
- Tableau
  - Tableau Viz of the Day
  - Tableau yearly conference
- Looker
  - GCP Integration
- Qlik Sense
- ThoughtSpot
  - “Life without dashboards”
- Oracle Analytics
  - “Enterprise” marketed

Poll! Favorite BI Platform

How are BI Platforms ranked? Who is Gartner?

Gartner Inc, is a global research firm providing information, advice, and tools for leaders in IT, finance, HR, and many other verticals.

Gartner publishes the **Gartner Magic Quadrant for BI Platforms** every year, and it is taken very seriously in the Data Warehousing space.

Gartner BI Platforms Ratings

- Varying costs for each Platform
- Each Platform has strengths and weaknesses; best option depends on your project

Source: Gartner (March 2022)

2020

2022

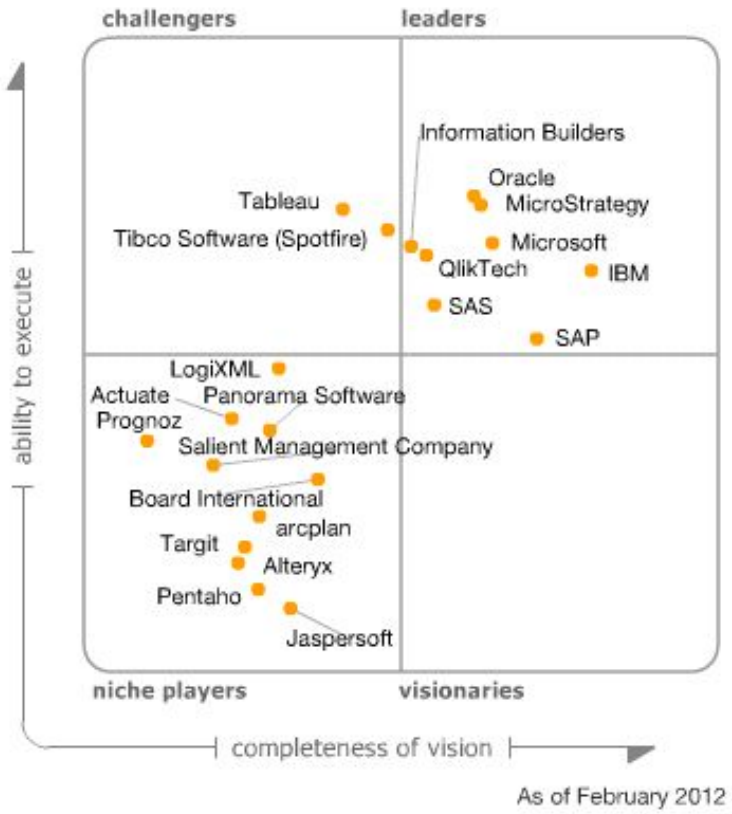
Source: Gartner (March 2022)

2016

2022

Source: Gartner (March 2022)

2012



2022

Source: Gartner (March 2022)

BI Platforms, things to learn:

- Types of Analyses by Data Type: distribution analysis, contribution analysis, etc
- BI Application best practices: consistency, colors, etc.

- Types of BI Applications: dashboard, report, etc
- Types of BI Users: Casual Consumers, Analysts, etc

Week 10 Class Overview:

1. Introduction to Analytics
2. Introduction to Business Intelligence (BI)
3. **Analyses by Data Type**
4. BI Application Best Practices
5. BI Workshop #1

Types of Analysis

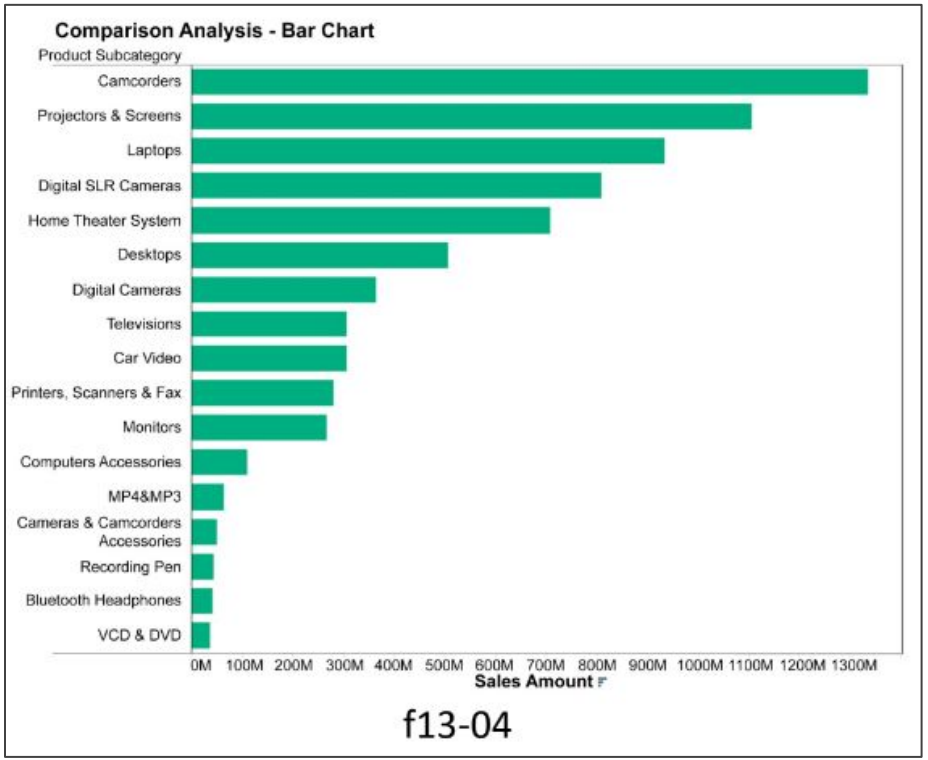
Pretend we are again building a Data Warehouse for the NYC Public Library system. We may run the following types of analyses:

- **Comparative Analysis** - Total check-outs by Book
- **Time-Series Analysis -** Popularity of Genres by Year
- **Contribution Analysis -** Total members by borough
- **Correlation Analysis** - Library Visits vs Books Checked Out by Member
- **Geographic Analysis** - Total libraries by Zip Code
- **Distribution Analysis** - Books checked out by Member

Comparative Analysis - Bar Chart

For a Bar Chart, you need:

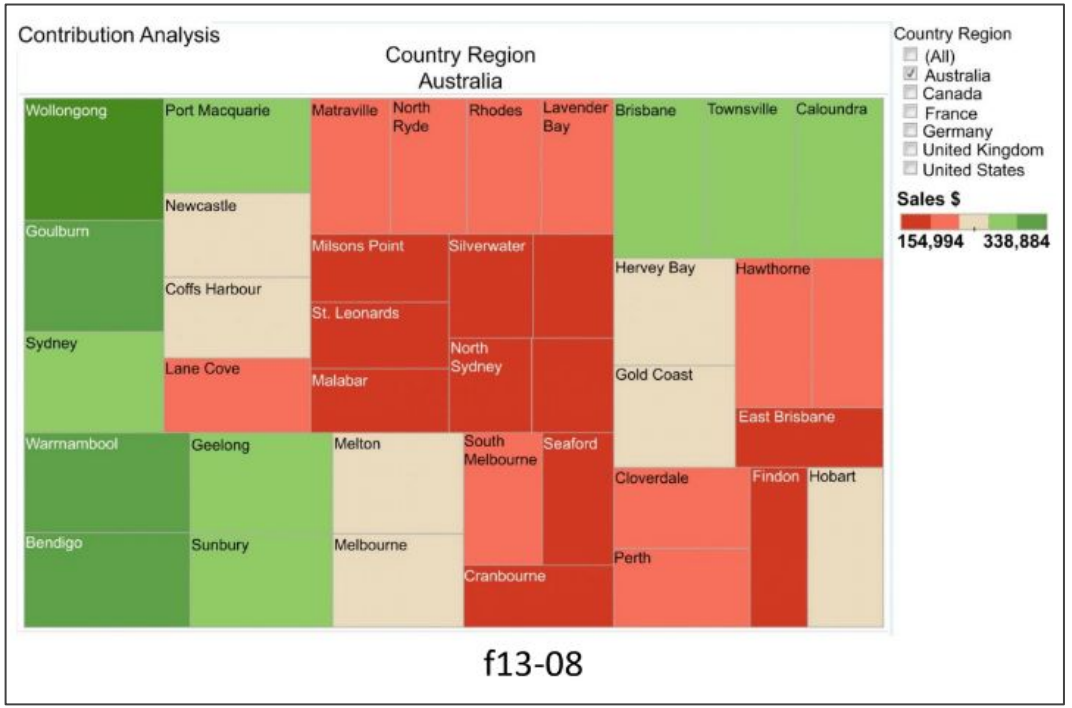
1. To note the discrete time period of the chart
2. Not too many segments
3. Sort bars by the measure value
4. Use one color for the bars when possible, this makes it easiest to interpret



Contribution Analysis - Heat Map

For a Heat Map, you need:

1. More than 4 segments, else use a Pie Chart
2. A color gradient that easily differentiates positive and negative
3. A legend to define the color gradient

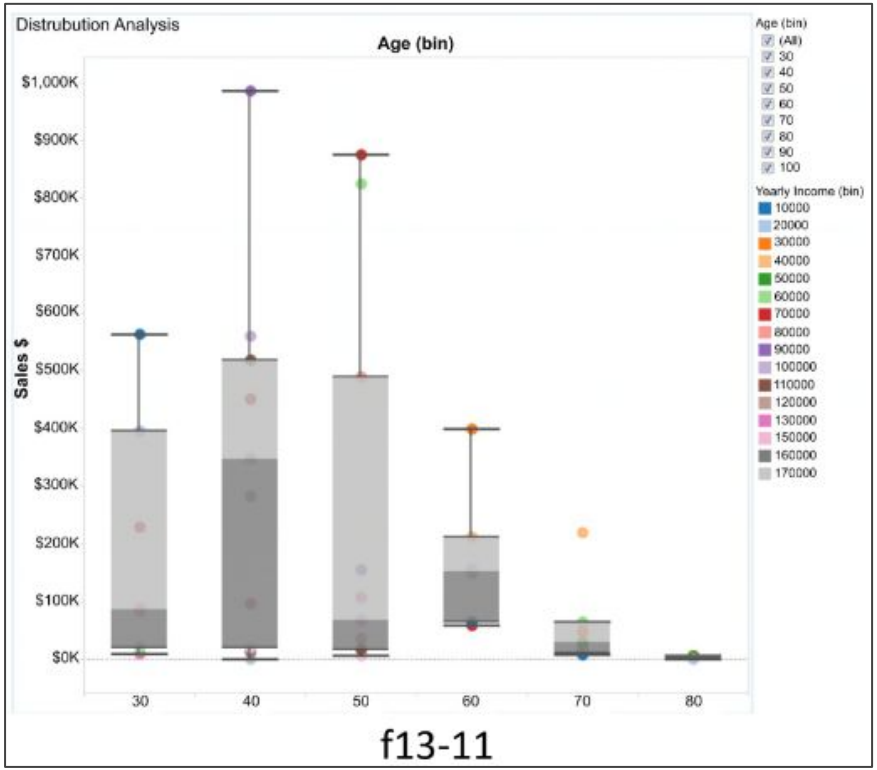


Distribution Analysis - Box Plot

For a Box Plot, you need:

1. Many values across a quantitative range
2. Break categories into columns (multiple boxes)
3. Consider how to represent outliers

For distribution analyses with one category, strongly consider Histograms or Violin Plots

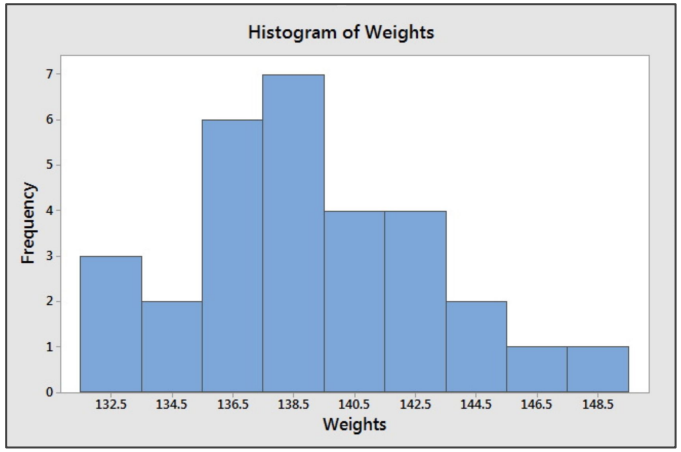


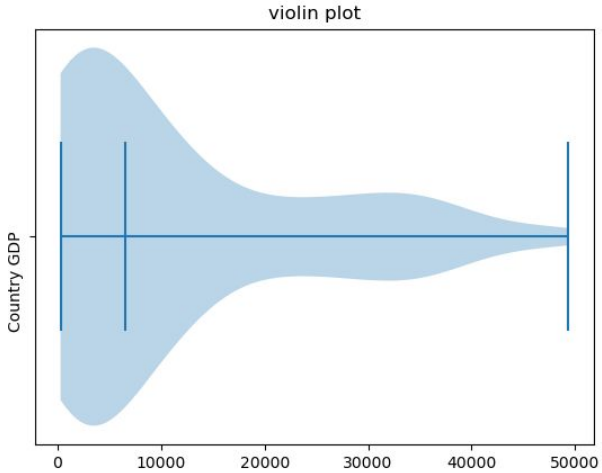
f13-11

Distribution Analysis - Violin Plot or Histogram

For distribution analyses with one category, strongly consider:

1. Histogram
2. Violin Plot

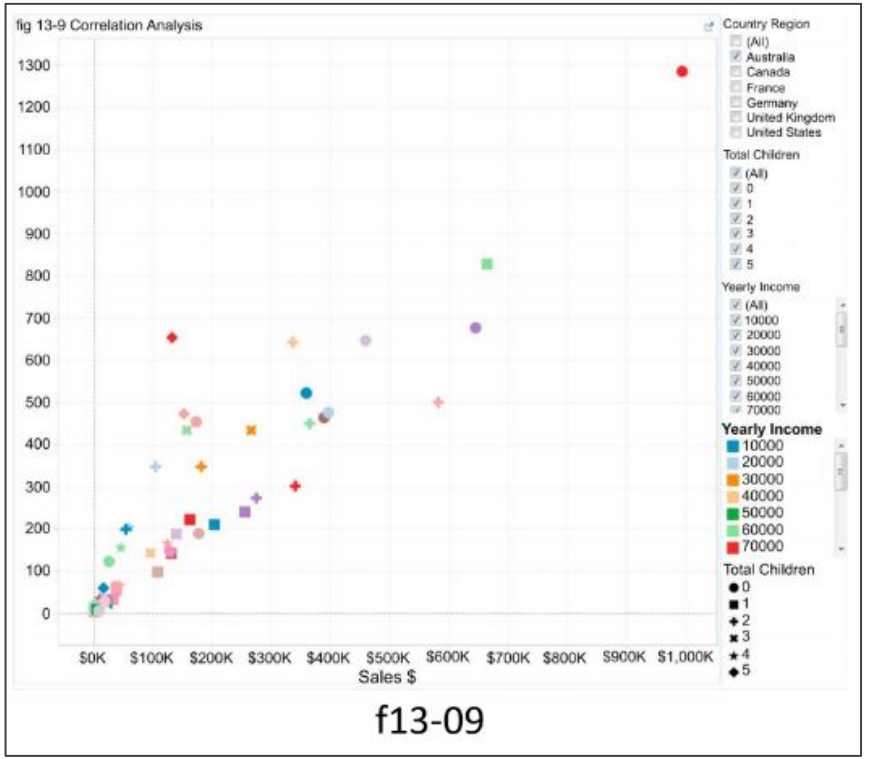




Correlation Analysis - Scatter Plot

For a Scatter Plot, you need:

1. Many data points
2. Use size and color to differentiate categories or amplitude
3. More evidence than just the Scatter Plot to determine true causation (always beware of lurking variables!)

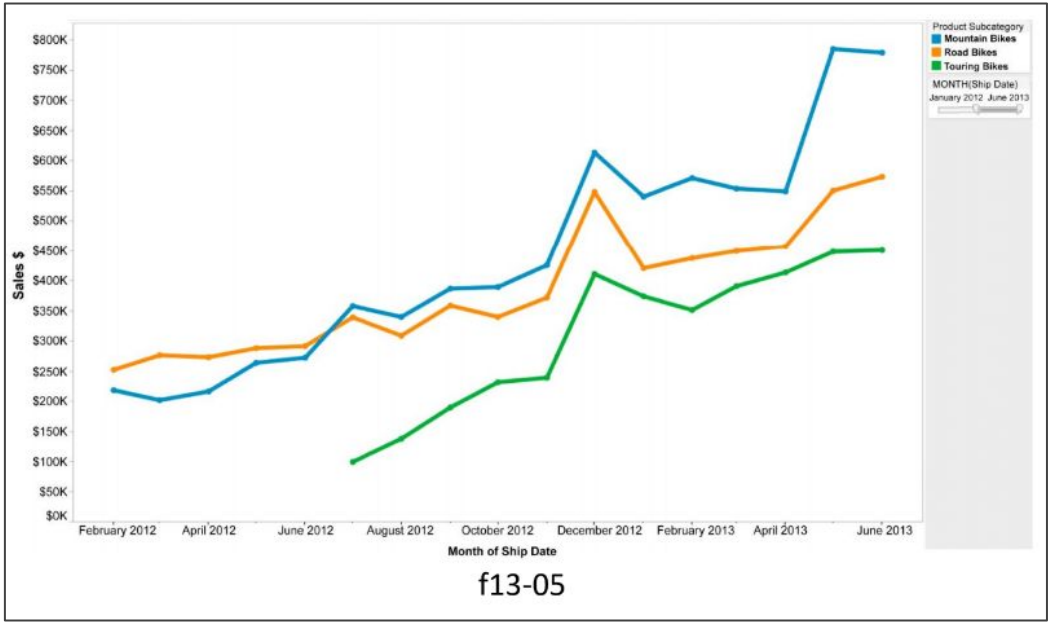


f13-09

Time-Series Analysis - Line Chart

For a Line Chart, you need:

1. Time as the x-axis
2. Different colors for different lines (optional)
    a. Legend to display color definitions

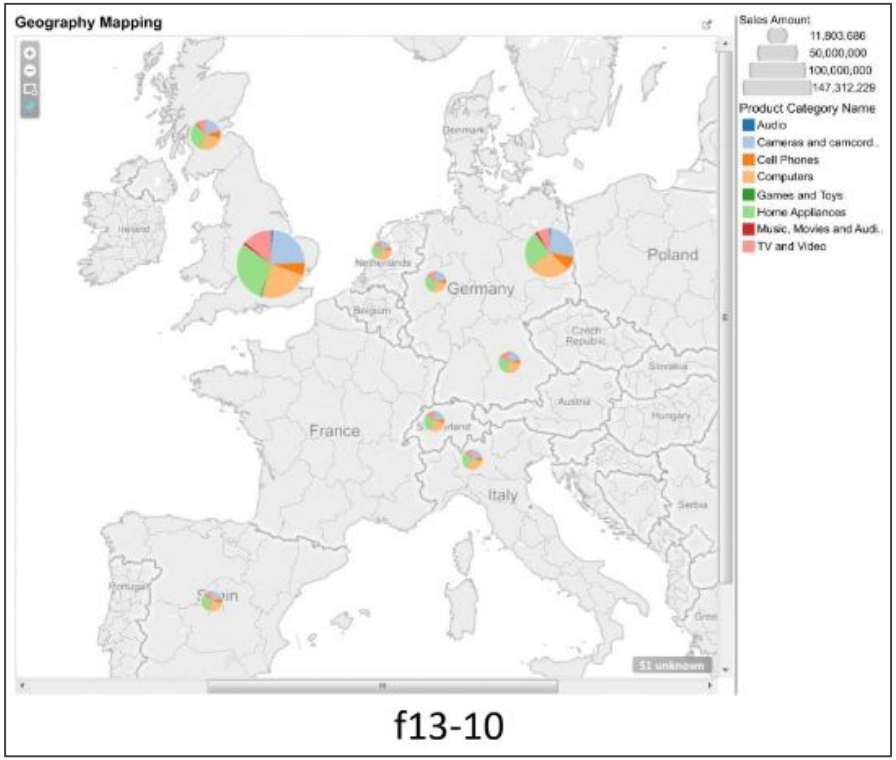


f13-05

Geographic Analysis - Geography Map

For a Map, you need:

1. Data by location
2. Consider incorporating color to distinguish categories



f13-10

Poll!

Which of the following are BI Applications?

10 Minute Break

Week 10 Class Overview:

1. Introduction to Analytics
2. Introduction to Business Intelligence (BI)
3. Analyses by Data Type
4. **BI Application Best Practices**
5. BI Workshop #1

BI Application Best Practices

- **Consolidate** BI Application List during Design phase
  - The Project's KPIs drive the BI Applications required
  - Ensure to limit BI Applications to as few as possible
    - This minimizes chance for errors
    - Less overlap means users are looking at same applications, speaking same data language
    - Many requirements can be satisfied with one BI Application if you **enable filtering and parameters**

Data Applications, Best Practices

- **Consistency** is key
  - Keep similar styles, backgrounds, layouts, fonts, color schemes, and legends for all BI Applications.
  - Users will become accustomed to the design and will develop an eye for consistent styles.

Data Applications, Best Practices

- **Simple** is always better
  - This is not a preference, simplicity is always best for BI Applications.
  - Try to use as little as needed to portray the intended KPI’s/data
    - No extra colors, lines, textboxes, etc.

Data Applications, Best Practices

- Think about **Location**
  - On a Dashboard, place the most important information at the top
    - (or even better, top-left)
  - Independent variables on the x-axis, dependent on the y-axis

Data Applications, Best Practices

- Only use **colors when necessary**
  - Colors are effective, but also distracting.
    - So, only use colors when necessary to make an important distinction.
  - Think about other ways to show differences as well, such as size and markers.

Data Applications, Best Practices

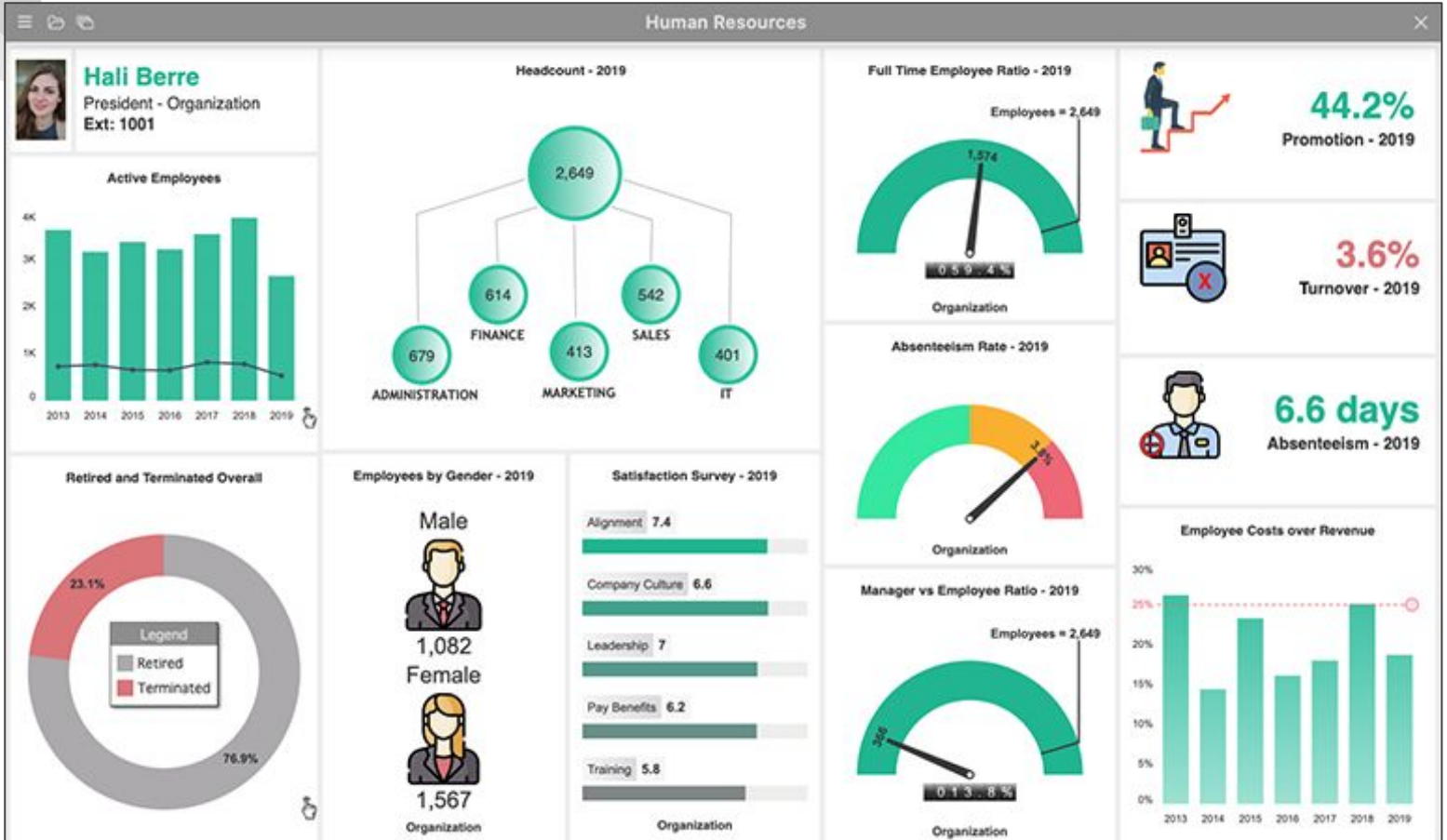
- Flag Outliers
- Include Labels
  - Assume your viewer has **never seen this data before**
  - Labels will guide them to understanding

Data Applications, Books

1. Data Story by Nancy Duarte
2. Storytelling with Data
3. Many more to pick from and get inspiration!

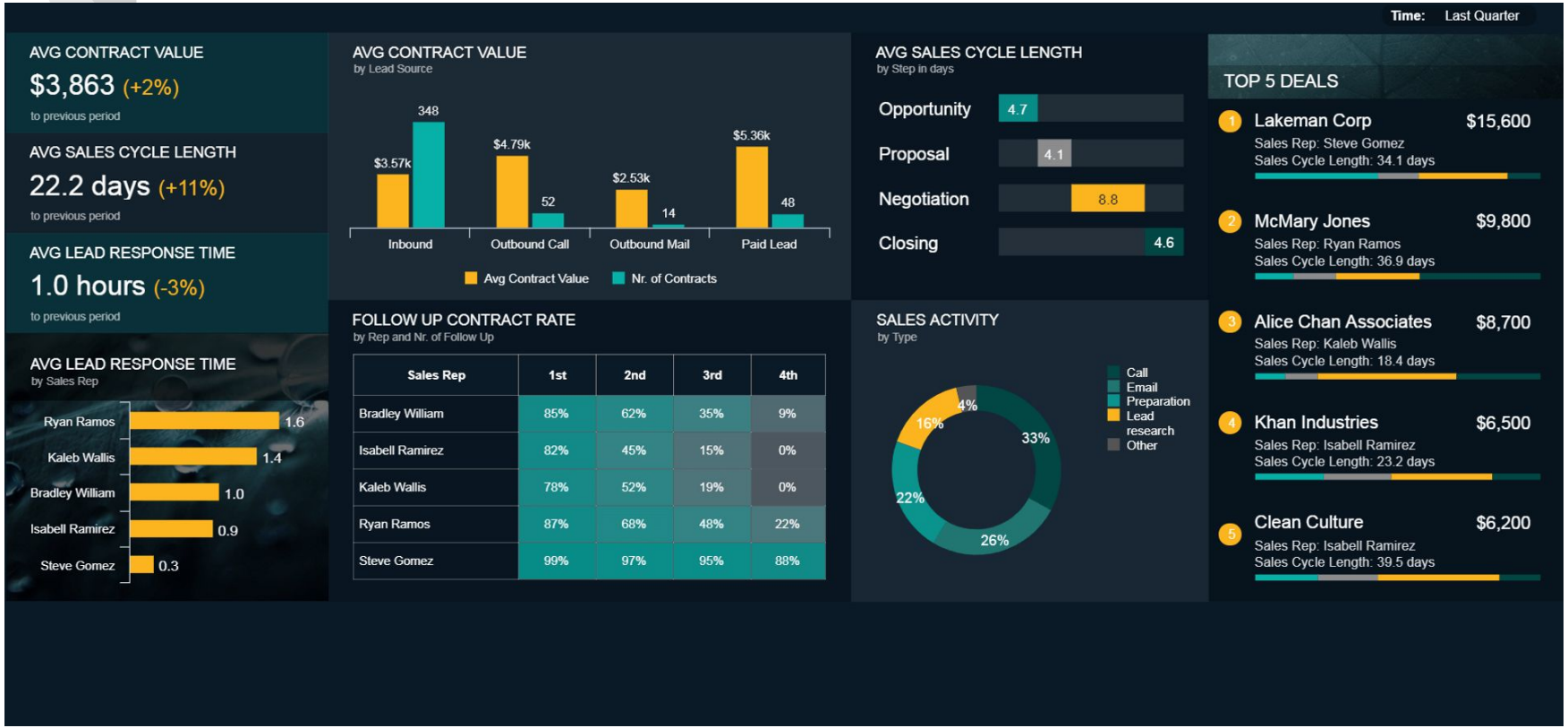
Let’s critique 3 BI Dashboards

Let' Critique:
Consistency, Simplicity, Location, Colors, Outliers, Labels?



Let' Critique:
Consistency, Simplicity, Location, Colors, Outliers, Labels?

Let' Critique:
Consistency, Simplicity, Location, Colors, Outliers, Labels?



Week 10 Class Overview:

1. Introduction to Analytics
2. Introduction to Business Intelligence (BI)
3. Analyses by Data Type
4. BI Application Best Practices
5. **BI Workshop #1**

Sidewalk Cafe Data Warehouse

1. **Comparative Analyses**
    a. (Together) Average Sidewalk Square Footage by Borough
    b. Pending Applications by Borough
2. **Contribution Analysis** - Total Sidewalk Tables by Street
3. **Geographic Analysis** - Total Sidewalk Chairs by Zip Code
4. **Distribution Analysis** - Sidewalk Square Footage by Business

Homework:

1. Final Project Milestone #3 on Blackboard, due Monday, 11/14/22
1. Final Project Milestone #4 on Blackboard, due Friday, 11/25/22