ISA 401: Business Intelligence & Data Visualization

01: Introduction to BI and Data Viz

Fadel M. Megahed, PhD

Raymond E. Gloss Professor in Business
Farmer School of Business
Miami University

@FadelMegahed
fmegahed
fmegahed@miamioh.edu
Automated Scheduler for Office Hours

Fall 2025

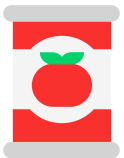
Learning Objectives for Today's Class

- Describe **course objectives** and **structure**.
- Define **data visualization** and describe its **main goals**.
- Describe the **BI methodology** and its **major concepts**.

Course Design, Expectations, and Overview

The Analytics Journey: Pre-Analytics [1]

- **Pre-Analytics/Data Management:** where one attempts to **extract** the needed *data* for analysis. Data can either be:



- Stale, uninteresting, convenient
- Highly processed and archived
- Example: `iris`, `mtcars`, `titanic`

- Fresh, interesting, challenging
- Impactful
- Examples: Cincinnati Open Data Portal, Ohio Data Portal, US Government's Open Data.

While the highly processed data can be useful in learning basic concepts, **real-world (often messy)** data real are much interesting to work with -- **e.g., we can make useful & meaningful decision from the data.** In this class, we will learn how to scrape, extract and clean messy data in addition to visualizing clean[ed] data.

Source: Slide inspired by Kia Ora's What I mean by "data".

The Analytics Journey: Pre-Analytics [2]

05:00

Non-Graded Class Activity #1

> *Take 5 minutes to discuss with your partner*

Activity | Your Solution | Fadel's Approach (No Solution Shown)

- Go to https://data.cincinnati-oh.gov/Safety/Traffic-Crash-Reports-CPD-/rvmt-pkmq/data
- Download the data utilizing the export column and answer the following questions:
  - How many **observations/rows** and **columns** do we have in the dataset?
  - How many **crashes** are reported in the dataset?

The Analytics Journey: Descriptive [1]

Descriptive Analytics: where one attempts to **understand** the data through **descriptive statistics** and **visualizations**.

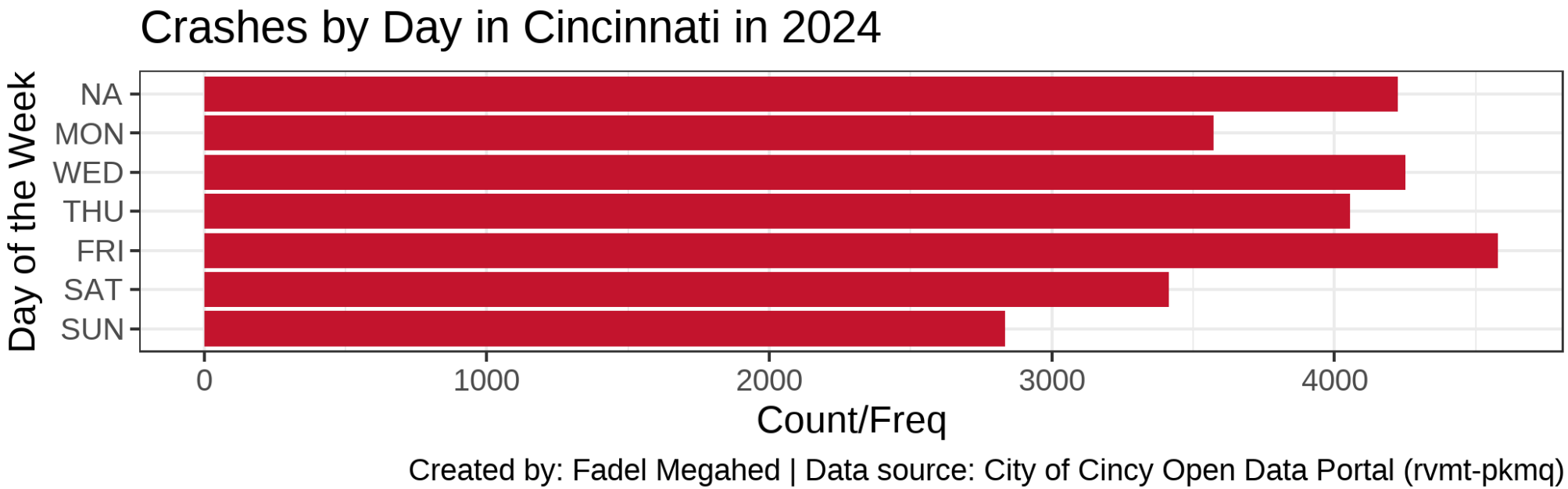
Descriptive Statistics for 2 Categorical Variables

```
## $dayofweek
##
##  MON  SUN  TUE  SAT  FRI  THU  WED
## 3572 2834 4225 3414 4579 4054 4250
##
## $weather
##
##                            2 - CLOUDY                                1 - CLEAR
##                                  3484                                    19331
##                              4 - RAIN                    3 - FOG, SMOG, SMOKE
##                                  3447                                       36
##                    99 - OTHER/UNKNOWN                                6 - SNOW
##                                   276                                      322
##                       5 - SLEET, HAIL    8 - BLOWING SAND, SOIL, DIRT, SNOW
##                                    16                                        7
## 9 - FREEZING RAIN OR FREEZING DRIZZLE                 7 - SEVERE CROSSWINDS
##                                     7                                        2
```

The Analytics Journey: Descriptive [2]

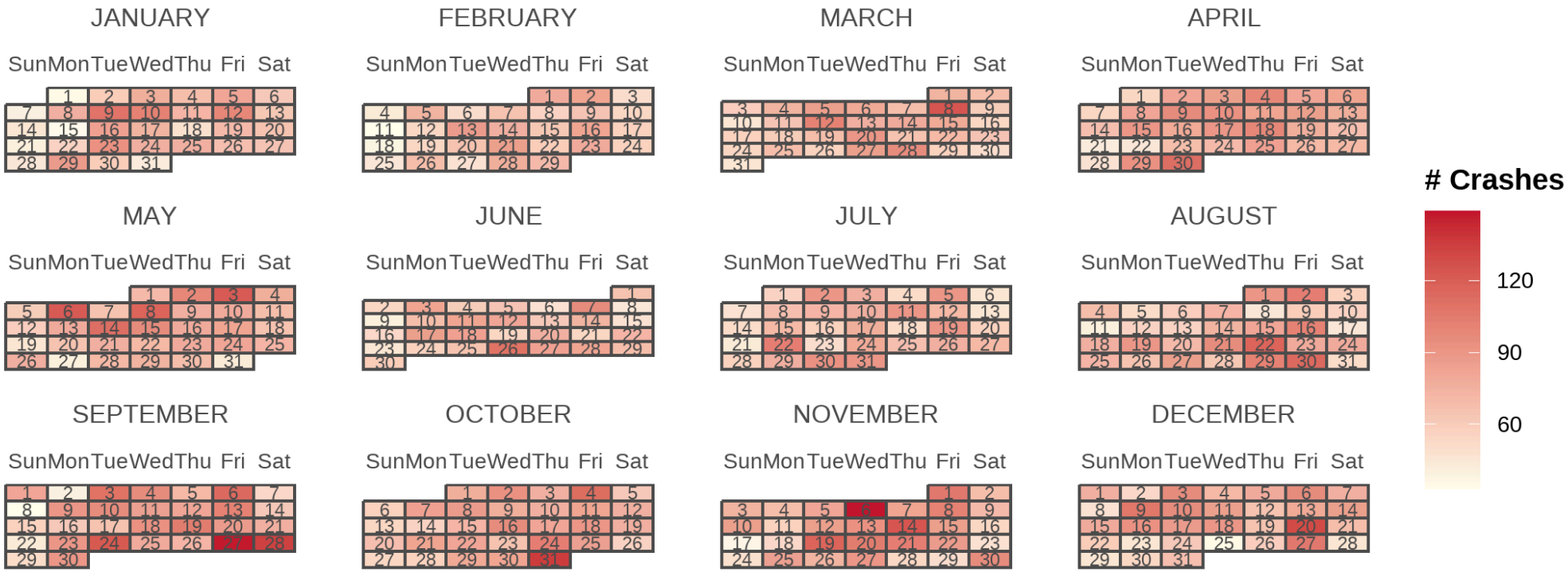
Descriptive Analytics: where one attempts to **understand** the data through **descriptive statistics** and **visualizations**.

A Simple Visualization - A Bar Chart of Crashes Per Day



The Analytics Journey: Descriptive [3]

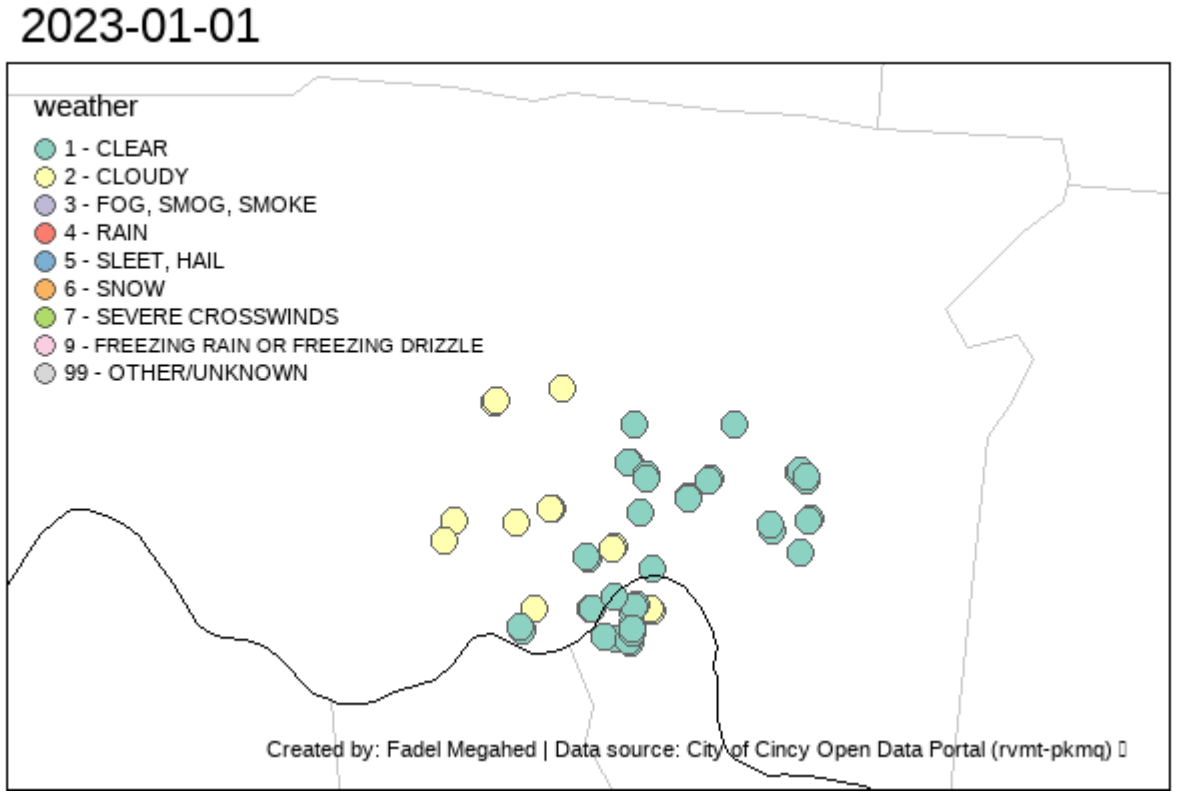
Descriptive Analytics: where one attempts to **understand** the data through **descriptive statistics** and **visualizations**.



Created by: Fadel Megahed | Data source: City of Cincy Open Data Portal (rvmt-pkmq)

The Analytics Journey: Descriptive [4]

Descriptive Analytics: where one attempts to **understand** the data through **descriptive statistics** and **visualizations**.



The Analytics Journey: Predictive [1]

Predictive Analytics: where **statistical** and **machine learning** models are used to help us utilize independent variable[s] to predict an outcome variable of choice.

- **Many** consider this component to be the 🍰 aspect of the analytics journey.
- IMO, this is not always true, but your success in this stage is **hinged on**:
  - **Correct** ✅ data, i.e.,
    - *Do you actually capture the important predictors?*
    - *Is your data aggregated to the right level?*
  - **Cleaned** 🛁 data, i.e.,
    - *Is your data tidy?*
    - *Is your data technically correct?*
    - *Is your data consistent?*

The Analytics Journey: Predictive [2]

Predictive Analytics: where **statistical** and **machine learning** models are used to help us utilize independent variable[s] to predict an outcome variable of choice.

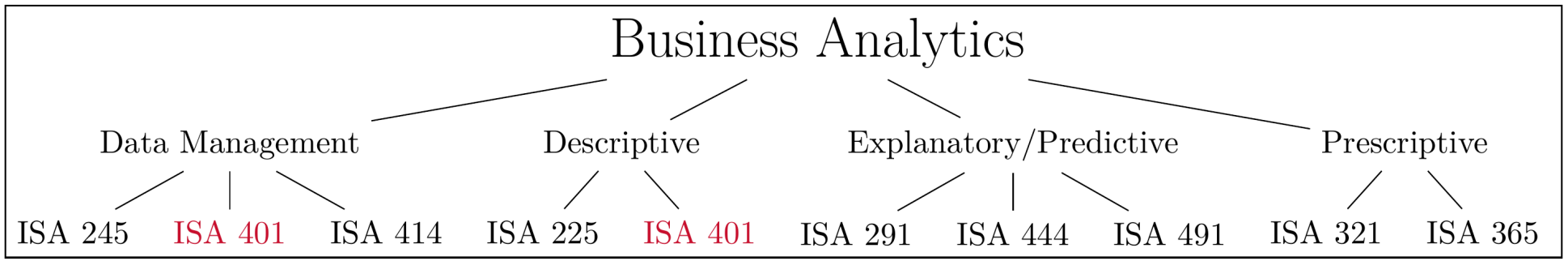
- With the aforementioned constraints/setup, now you can explore how to model the data using statistical and machine learning models?
- **Some recommendations:**
  - Start with the simplest (which is also often the most easy-to-explain) model first.
  - If you are happy with the predictive performance (i.e., no gains would be of practical benefit), you are done 👏.
  - If not, ↩️ and try other models.

The Analytics Journey: Prescriptive

Prescriptive Analytics: where **mathematical models** are used to make recommendations for business actions.

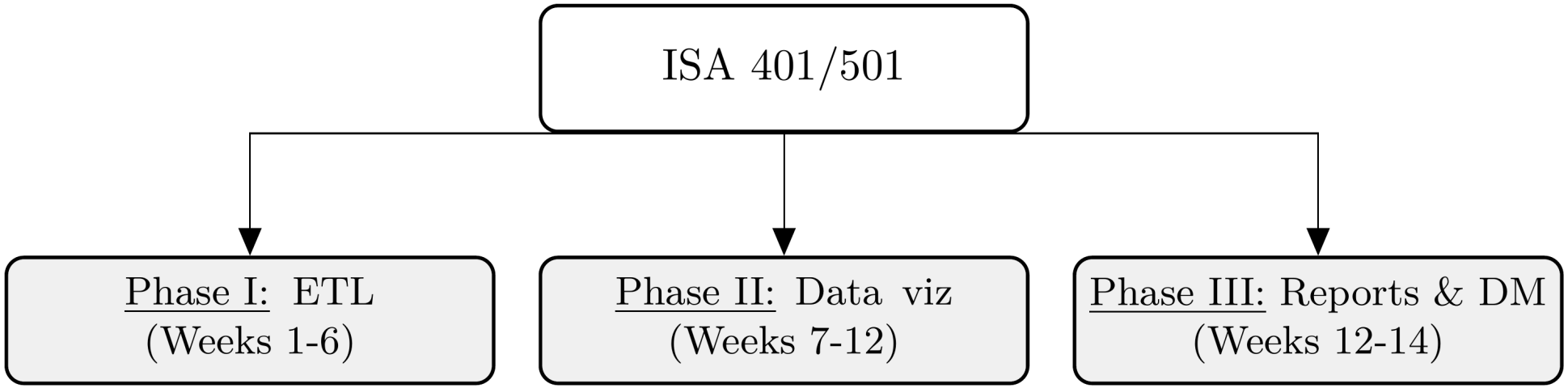
- Our **overarching goal** behind data/business analytics, is to **make informed decisions based on what we have learned from the data**. Hence, this stage is where we build on what we learned during the *descriptive* and *predictive* stages to make more informed decisions.
- Imagine that you are a large trucking company (e.g., Amazon, Fedex, JB Hunt), and you have models that show **both**:
  - Safety critical events that are associated with crashes.
  - The occurrence of safety critical events can be reasonably predicted as a function of: (a) driver characteristics, (b) weather conditions, and (c) traffic conditions.
- **As a business analyst, what two reasonable questions would you attempt to approach/optimize for?**

How does our Curriculum at Miami University Prepare you for this Journey?



My take on the courses within the business analytics major/minor at Miami University

ISA 401 Course: An Overview



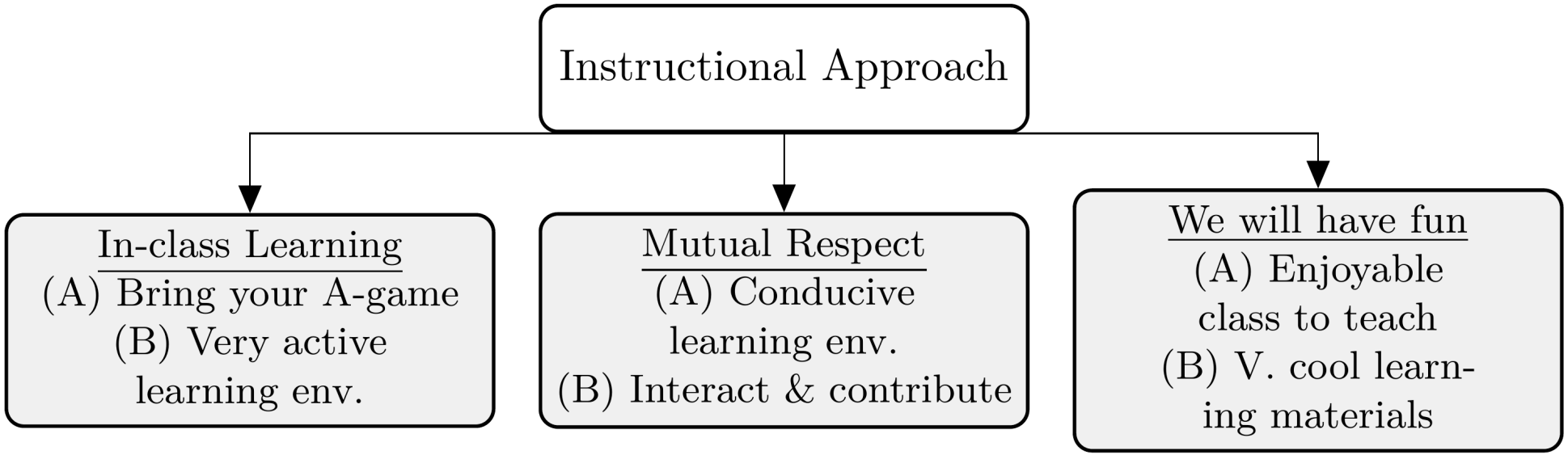
How the ISA 401 course is organized.

ISA 401 Course Objectives

Even though software will be extensively used, this is not a software class. **Instead, the focus is on understanding the underlying methods and mindset of how data should be approached.**

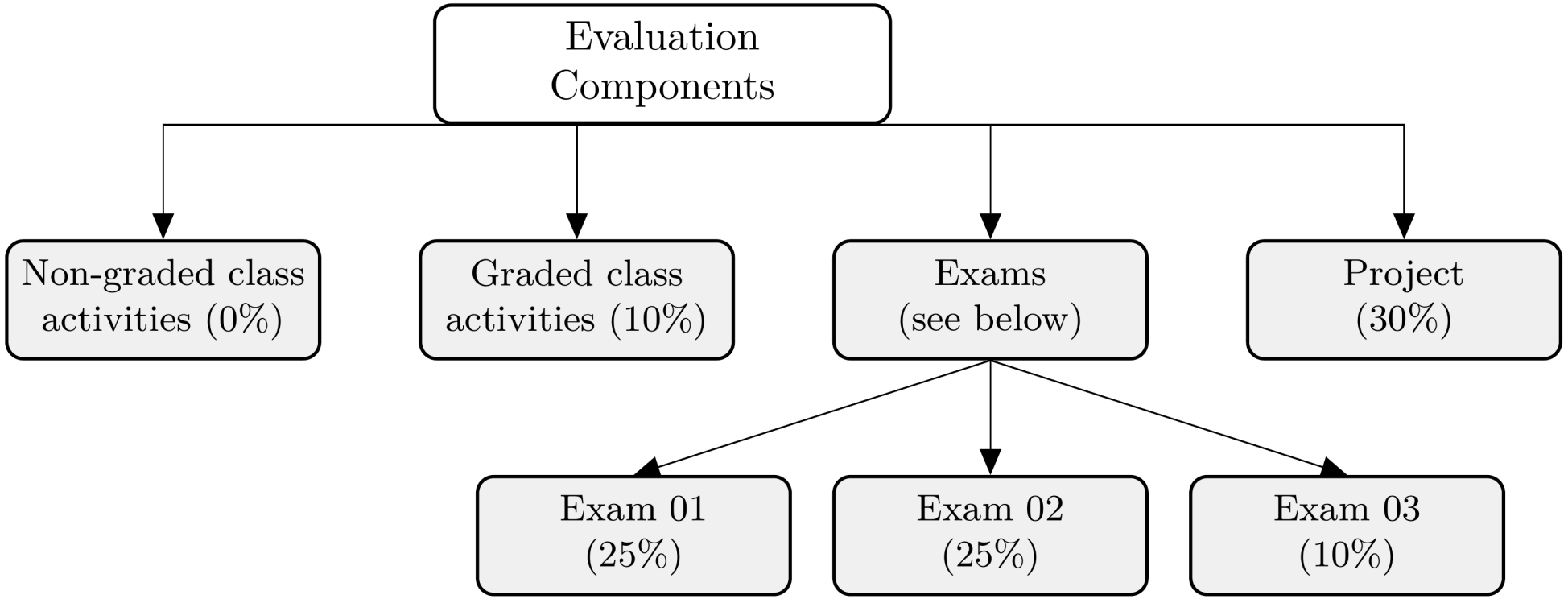
- Be capable of extracting, transforming and loading (ETL) data using multiple platforms (e.g.  , Power BI and/or Tableau).
- Write basic scripts to preprocess and clean the data.
- Explore the data using visualization approaches that are based on sound human factors.
- Understand how statistical/machine learning can capitalize on the insights generated from the data visualization process.
- Create interactive dashboards that can be used for business decision making, reporting and/or performance management.
- Be able to apply the skills from this class in your future career.

Instructional Approach



An overview of the instructional approach for ISA 401.

How will I Evaluate your Learning?



An overview of the evaluation components for ISA 401.

Introductions: Getting to Know Each Other

About Me – My route to Miami University

- Application of data-driven decisions (D3) in 3 continents.
- **Interests:** Applications in logistics, manufacturing, occupational safety & portfolios.
- **Collaborations with:** Aflac, GE Research, Gore, IBM Research, & Tennibot

Home 2 of 6

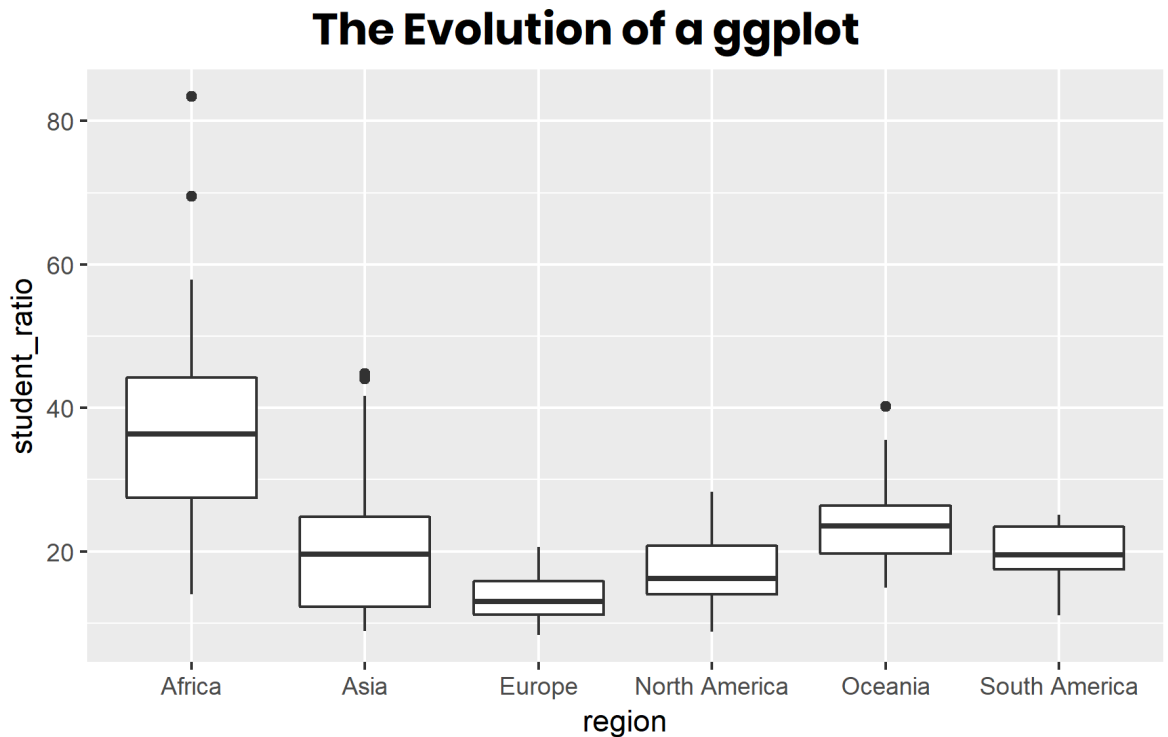
Getting to Know Your Learning Objectives

- learning goals for ISA 401?

So What is Data Visualization?

What is Data Visualization?

Data visualization involves **presenting data in a graphical format**. It is really a process that starts by getting data, creating initial plot(s) and modifying them to answer questions of interest (and possibly making the plot aesthetically pleasing). For example, see Cedric Scherer's visualization of the UNESCO data on global student to teacher ratios.



The Goals of Data Visualization

- **Record** information
- **Analyze** data to support reasoning
  - Develop and assess hypotheses (EDA)
  - Reveal patterns
  - Discover errors in data
- **Communicate** ideas to others
  - Infographics
  - Statistic charts
  - Interactive charts
  - Dashboards
- **Interact with the data (which supports all the above)**

Record Information

Tanya Shapiro

PELOTON STATS

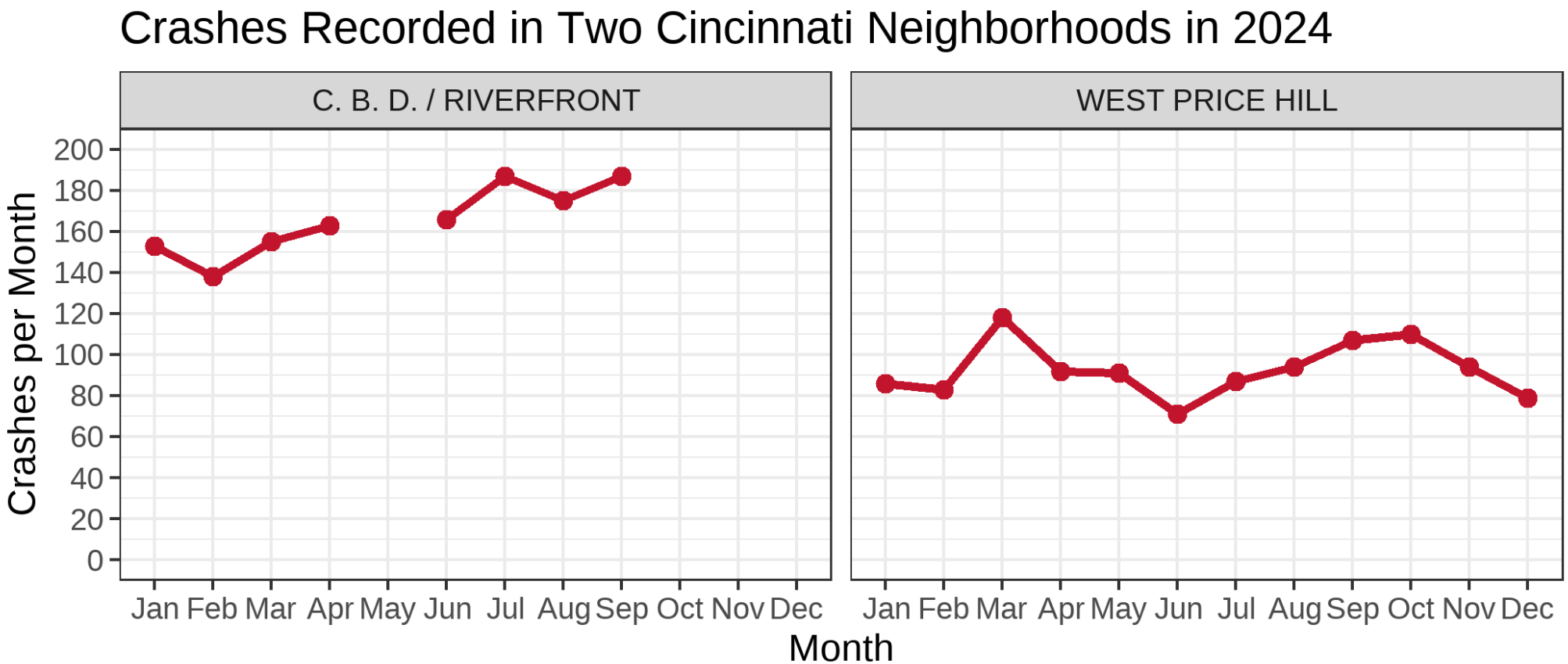
This project was influenced, in part, by Peloton's classic leaderboard: on a Peloton ride, users can see their ranking compared to other users based on output. I used this concept to build my own instructor leaderboard, ranking each instructor based on the amount of time I've personally invested with each one.

Of course I had more questions - how many workouts did I do? what was the most popular workout type? how did my numbers look last month or last year? To satiate my curiosity, I built a dynamic application with R Shiny to explore my Peloton metrics. This application includes a dynamic group by option to see metrics aggregated per instructor or discipline. More details nested per aggregation to reveal sub aggregations (i.e. instructor -> discipline or discipline -> instructor).

About Project

Source: Tanya Shapiro (2023). Peloton Stats. The project can be accessed here.

Analyze Data



Reveal Patterns: The 1854 Cholera Outbreak

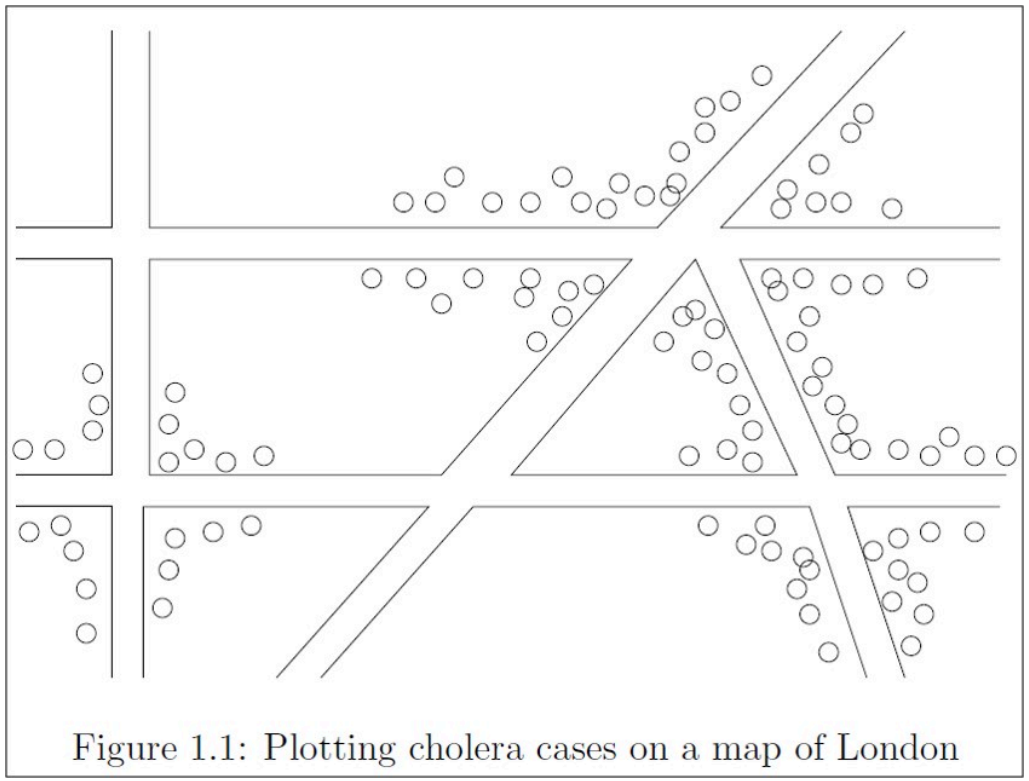
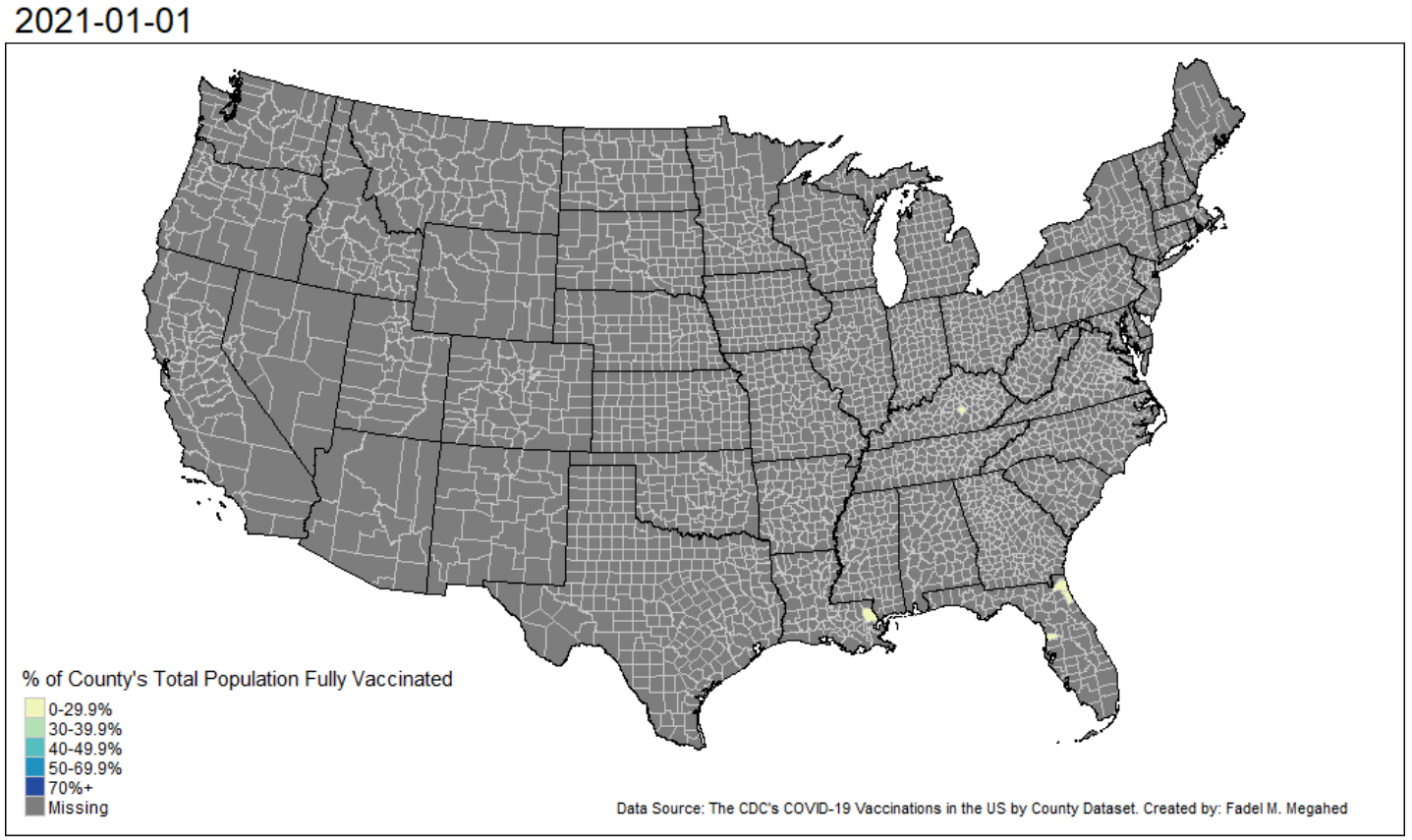


Figure 1.1: Plotting cholera cases on a map of London

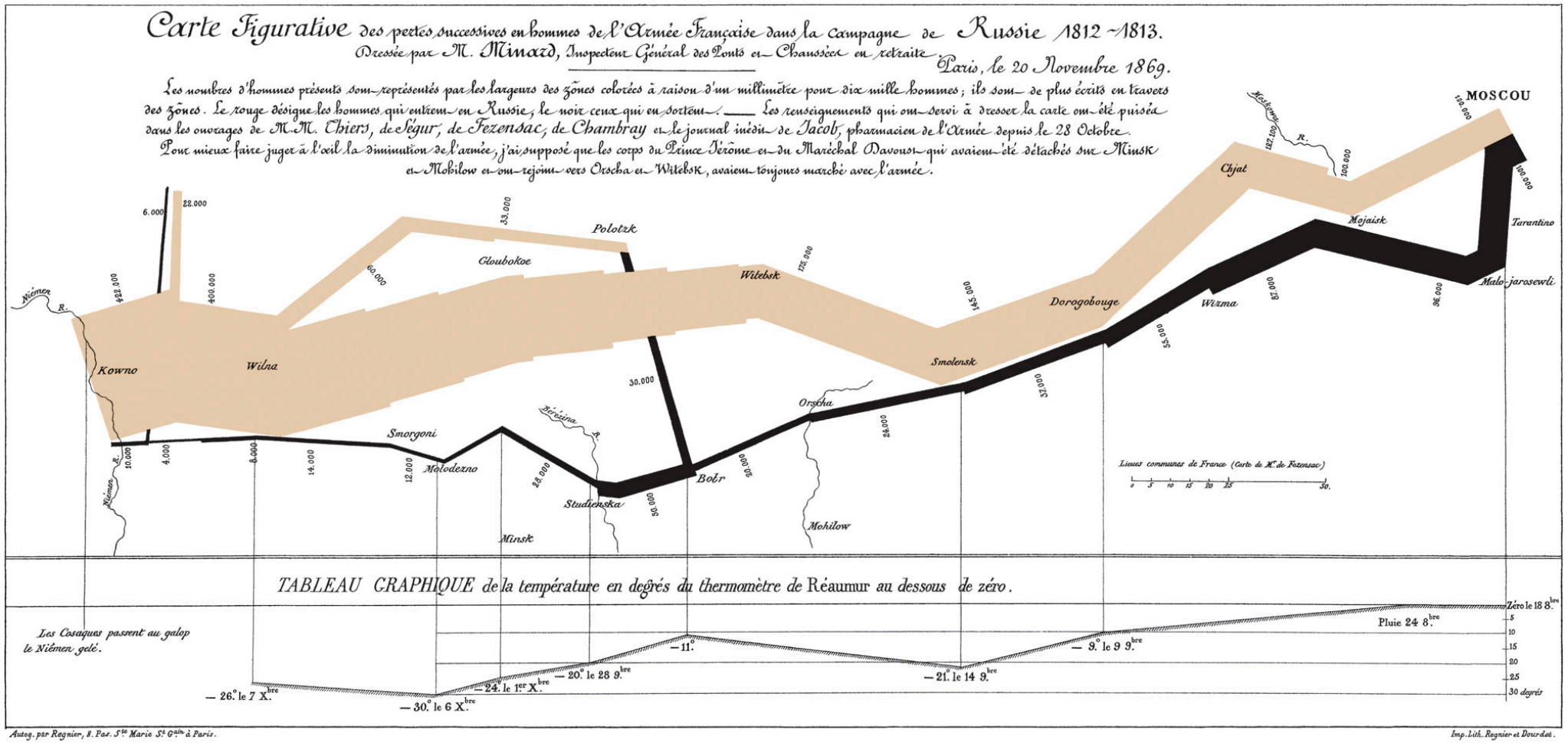
The physician John Snow, dealing with a Cholera outbreak plotted the cases on a map of the city (see schematic above).

Source: Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). Mining of Massive Data Sets (Third Edition). Cambridge University Press. Image is from Chapter 1, which can be accessed here.

Reveal Patterns: COVID-19 Vaccination Rates



Communicate Ideas: C.J Minard 1869



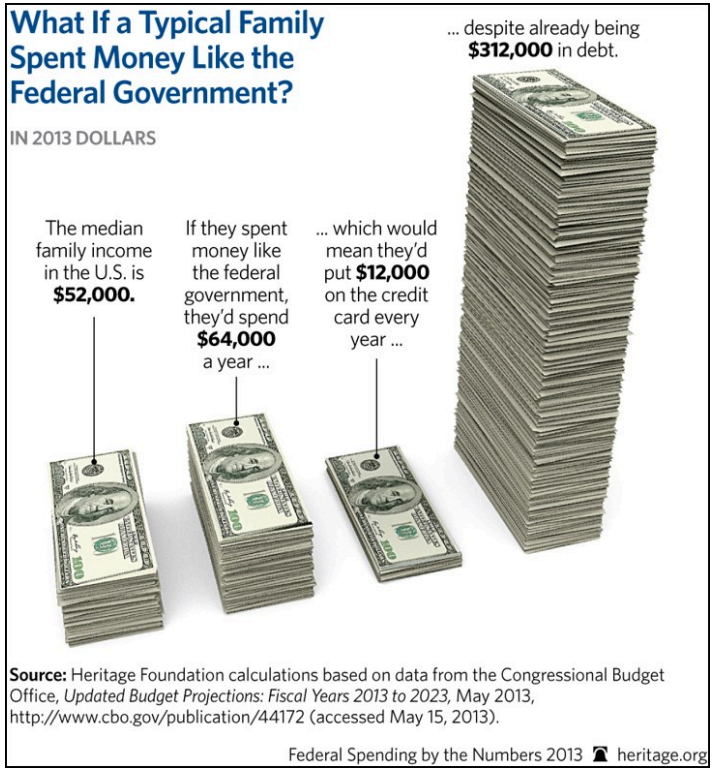
Communicate Ideas

05:00

Activity | Your Solution

Non-Graded Class Activity #2

- Who is the target audience?
- What is the data represented in this visualization? Be Specific.
- How is the data visually encoded?
- Do you like/dislike this visualization? Why?
- Would you do visualization like this for a similar dataset? Why? Why not?

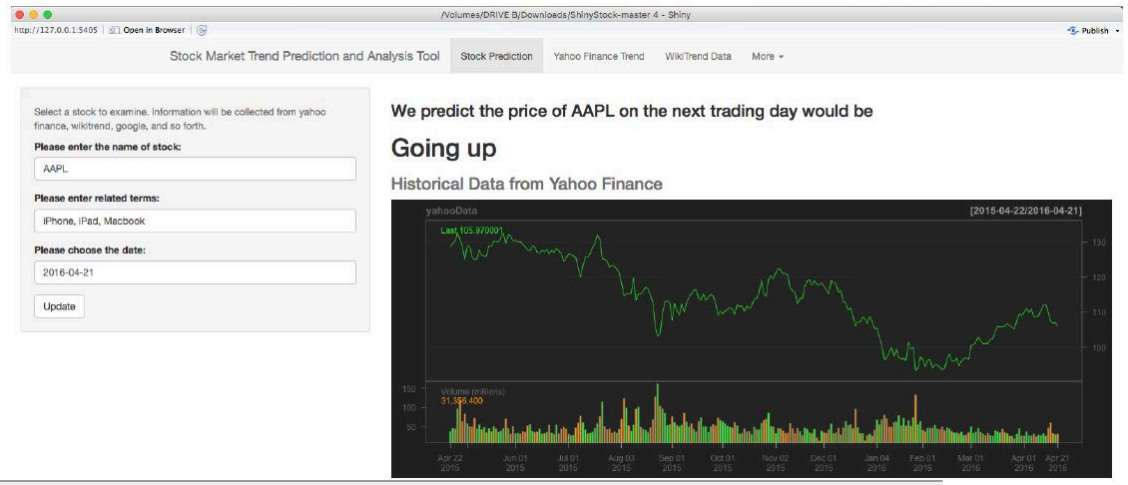


Interact: GapMinder/ Hans Rosling Example

Business Intelligence: From Visualizations to Dashboards to Insights

What is Business Intelligence?

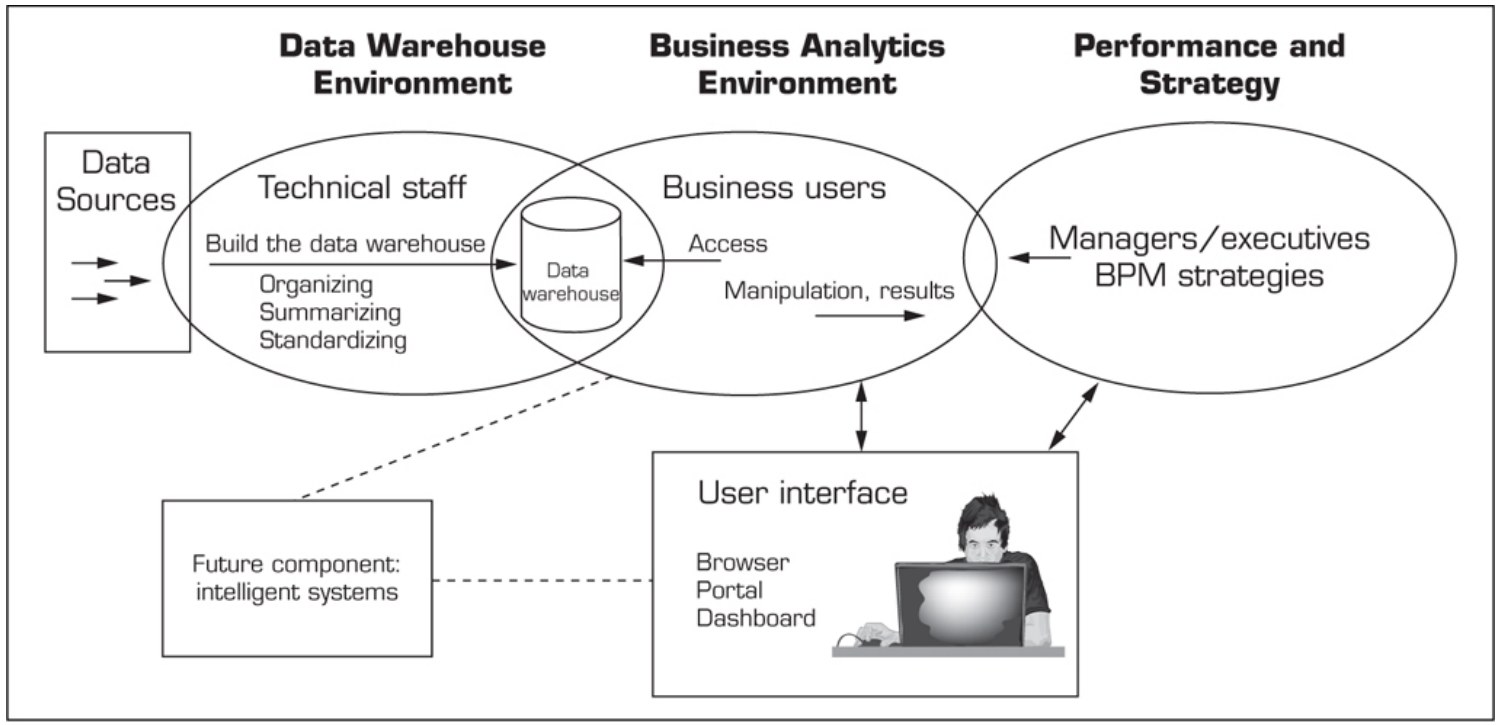
> "... to enable **interactive access (sometimes in real time)** to data, to enable manipulation of data, and to give business managers and analysts the ability to conduct appropriate analysis. By analyzing ... data, situations, and performances, decision makers get valuable insights that enable them to **make more informed and better decisions** ... BI is based on the **transformation of data to information, then to decisions, and finally to actions.**"



Quote from Sharda, R., Delen, D., & Turban, E. (2013). Business Intelligence: A managerial perspective on analytics. Prentice Hall Press.

Image Credit: Joint work with Bin Weng.

The BI Process



Source: Based on W. Eckerson, Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions. The Data Warehousing Institute, Seattle, WA 2003, p. 32, Illustration 5.

Image Credit: Sharda, R., Delen, D., & Turban, E. (2013). Business Intelligence: A managerial perspective on analytics. Prentice Hall Press.

Recap

Summary of Main Points

By now, you should be able to do the following:

- Describe **course objectives** and **structure**.
- Define **data visualization** and describe its **main goals**.
- Describe the **BI methodology** and its **major concepts**.

 Review and Clarification 

- **Class Notes**: Take some time to revisit your class notes for key insights and concepts.
- **Zoom Recording**: The recording of today's class will be made available on Canvas approximately 3-4 hours after the session ends.
- **Questions**: Please don't hesitate to ask for clarification on any topics discussed in class. It's crucial not to let questions accumulate.

📖 Required Readings 📖

📈 R Prep

- Workflow: Basics
- Names and Values
- Vectors
- Subsetting

🤖 LLM: Prep

- A Very Gentle Introduction to Large Language Models without the Hype

🎯 Assignment 🎯

- Complete Assignment 01 on Canvas to reinforce your understanding and application of the topics covered today as well as the assigned readings.