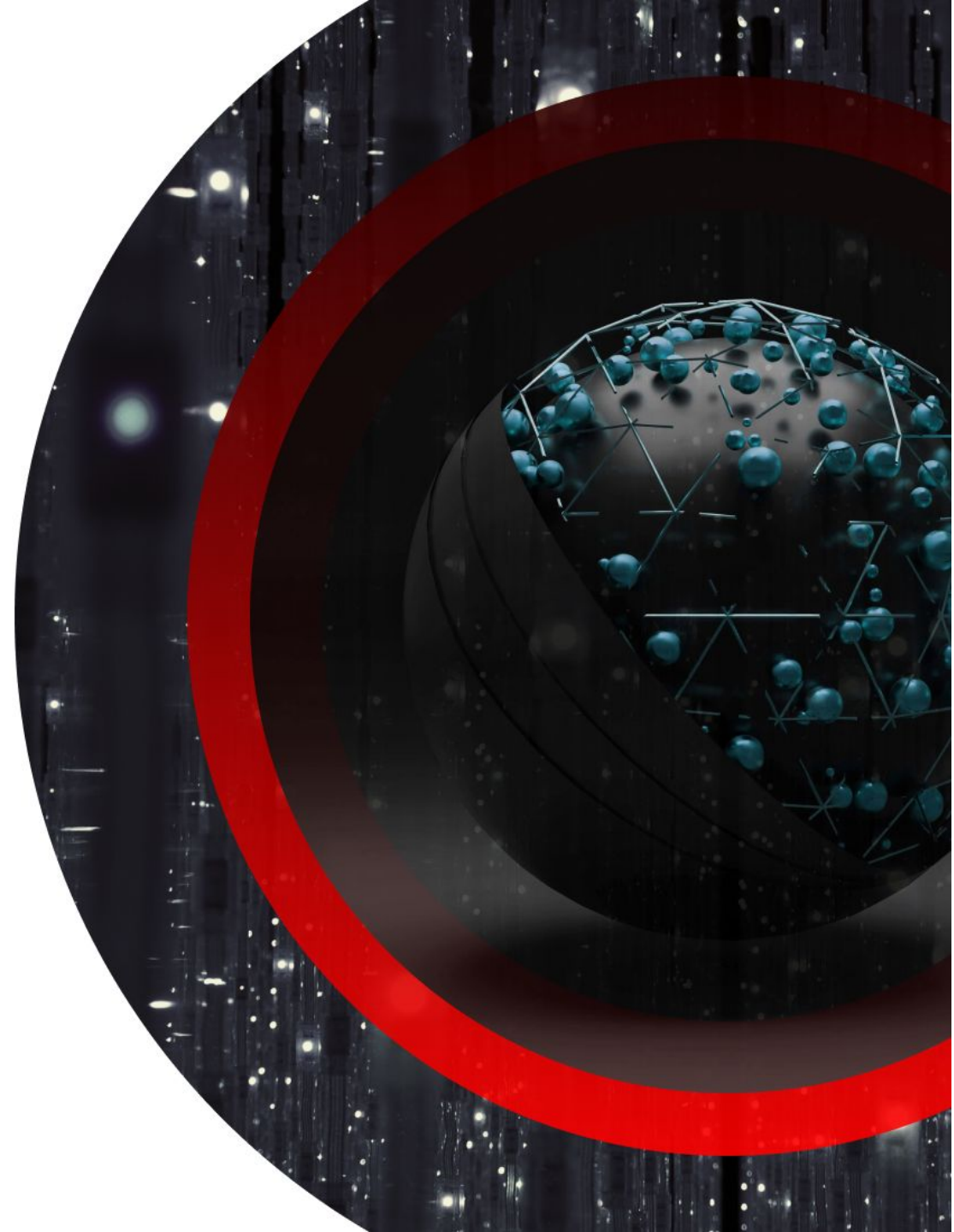


# DATA SCIENCE AND ANALYTICS

Introductory Course



# CLASS NORMS

## COURTESY IN CLASS

Remaining on mute unless called on, exercising courtesy during breakout rooms, and using the chat box for questions only

## ATTENDANCE

100% attendance is expected and contributes to success in passing the course and the program.

## PARTICIPATION

Keeping an open mind in discussions and sharing experiences, making contributions during team assignments, submitting assignments in Canvas, and participating in discussion boards

## USE OF CLASS RESOURCES

Follow along during the lecture with the lesson companion and download any in-class documents prior to class.





# PROGRAM PATH

**1 Introductory Course**

2 SQL and Databases

3 Statistics and Probability

4 Data Storytelling

**Milestone 1:** Building and  
Presenting Data Stories

5 Python Programming

6 Data Wrangling

7 Visual Communications

8 Advanced SQL Programming

**Milestone 2:** Data Integration,  
Preparation, Reporting, and  
Presentation

9 Business Intelligence

10 Big Data

11 Machine Learning

12 Applied AI

**Milestone 3:** Capstone Project:  
Delivering Insights and  
Presentations

# INTRODUCTORY COURSE PATH

1

INTRODUCTION TO  
DATA SCIENCE AND  
ANALYTICS

2

COMPUTING  
PRIMER

3

PROGRAMMING  
CONCEPTS

4

DISCOVERING AND  
CURATING DATA

5

STRUCTURING AND  
ANALYZING DATA

6

CLEANING AND  
ENRICHING DATA

7

**VALIDATING AND  
PRESENTING DATA**

8

INTRODUCTION TO  
DATA SCIENCE PROJECTS

9

ASSESSMENT NIGHT





# LESSON 1: VALIDATING DATA



# LESSON OUTLINE

- Data validation
- Types of data validation
- Accuracy vs. precision
- Tools for data validation





# BRAINSTORMING

Imagine you are a student who has taken notes throughout the semester in a variety of formats, such as:

- Handwritten notes
- iPad notes
- Shared Google Doc
- PowerPoint slides

**Think about how you might consolidate this information.**





LESSON GOALS



# WHAT ARE THE GOALS?

To learn to validate data



# WHY ARE THEY IMPORTANT?

Data validation is one of the most important tasks for data professionals.





# DATA VALIDATION

Ensuring data is consistent and accurate within a data set is known as **data validation**.

- Can be time-consuming
- Contributes to the most accurate results





# WHAT TO LOOK FOR

- Accuracy
- Quality
- Reliability
- Precision





# WHY DO WE VALIDATE DATA?

- To catch errors
- To confirm the data source and type compatibility
- To ensure the integrity and legitimacy of the results



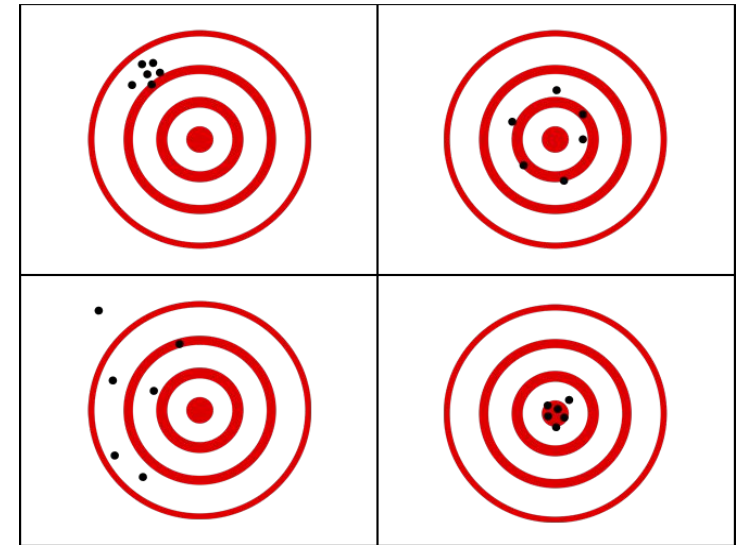


# PRECISION VS. ACCURACY

**Precision:** How close measurements are to each other

**Accuracy:** How close measurements are to the *true* value

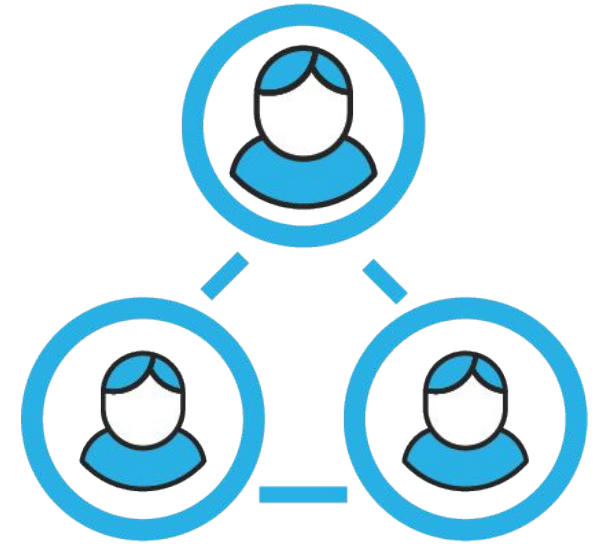
**Identify the level of precision and accuracy for each target.**





# METHODS FOR DATA VALIDATION

- Field level
- Form level
- Search criteria





## 1.7.1 Activity

At the beginning of the lesson, we imagined a scenario where we needed to consolidate notes taken throughout the semester in a variety of formats, such as:

- Handwritten notes
- iPad notes
- Shared Google Doc
- PowerPoint slides

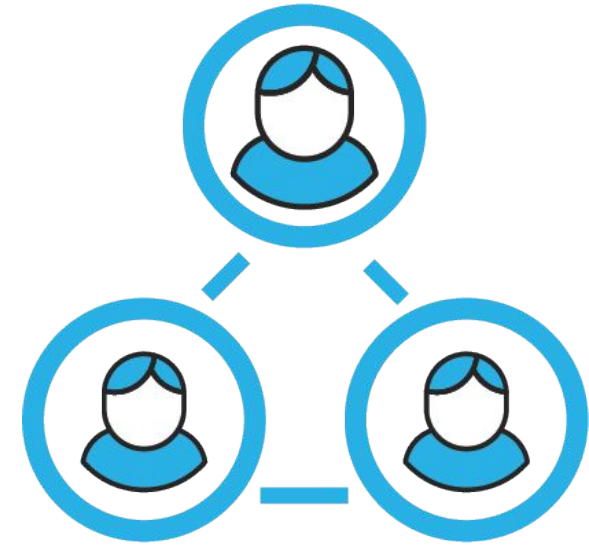
Answer the questions in the Activity 1.7.1 document to walk through the process of cleaning and validating the data in your notes.





# INTERVIEW TIME

What are four things you want to look for during data validation?





# REVIEW AND WRAP-UP

Take a few minutes to write about what you have learned.



# QUESTIONS?





**BREAK  
TIME**



# LESSON 2: VISUALIZING DATA



# LESSON OUTLINE

- The purpose of data visualization
- The parts of a graph
- Types of visualizations
- Tools





# DISCUSSION

Provide one example of a visualization. In your own words, describe its purpose.





LESSON GOALS



## WHAT ARE THE GOALS?

To explain the purpose of data visualization

To identify elements of a graph

To identify different types of visualizations and examples of how they are used



## WHY ARE THEY IMPORTANT?

Visualizations play an important role in the data life cycle. It is important to be able to read visualizations accurately and recognize the variety of ways data can be presented.



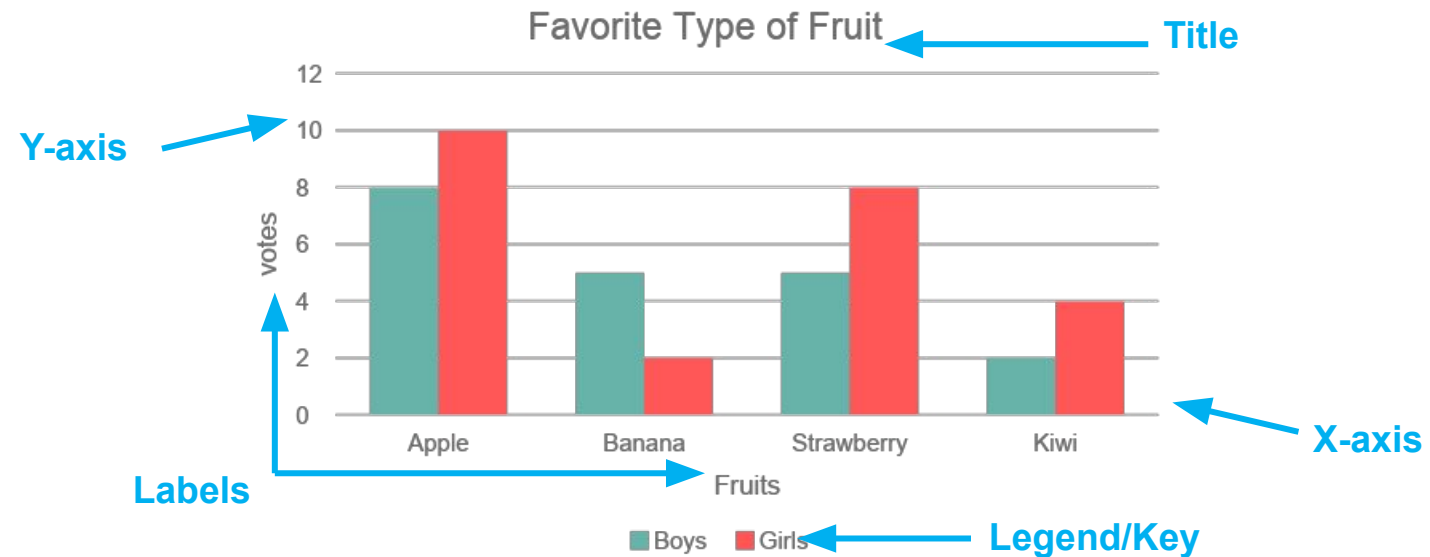
# DATA VISUALIZATIONS

**Data visualizations** are graphical representations that aid in understanding trends and patterns in data.

- Identify gaps and outliers.
- Analyze large amounts of data quickly.
- Make data-driven decisions.



# PARTS OF A GRAPH







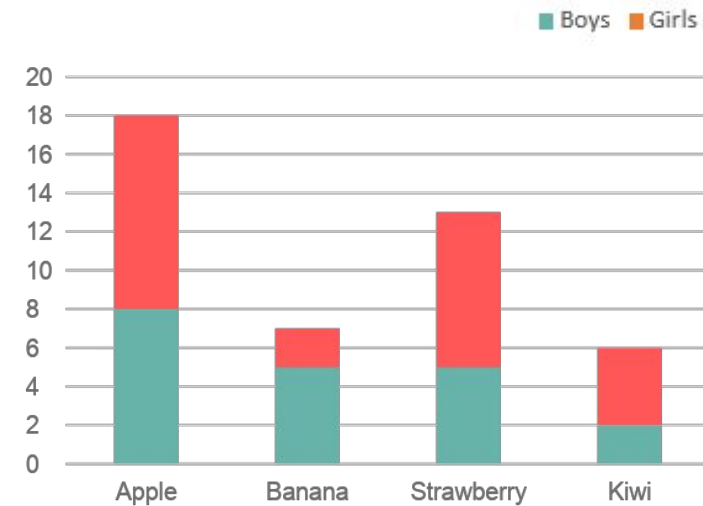
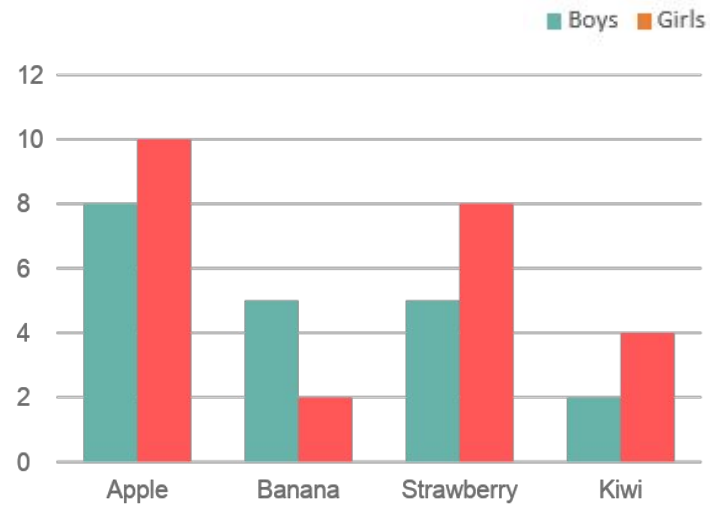
## 1.7.2 GUIDED ACTIVITY

How would you use a visualization in your workplace or industry?

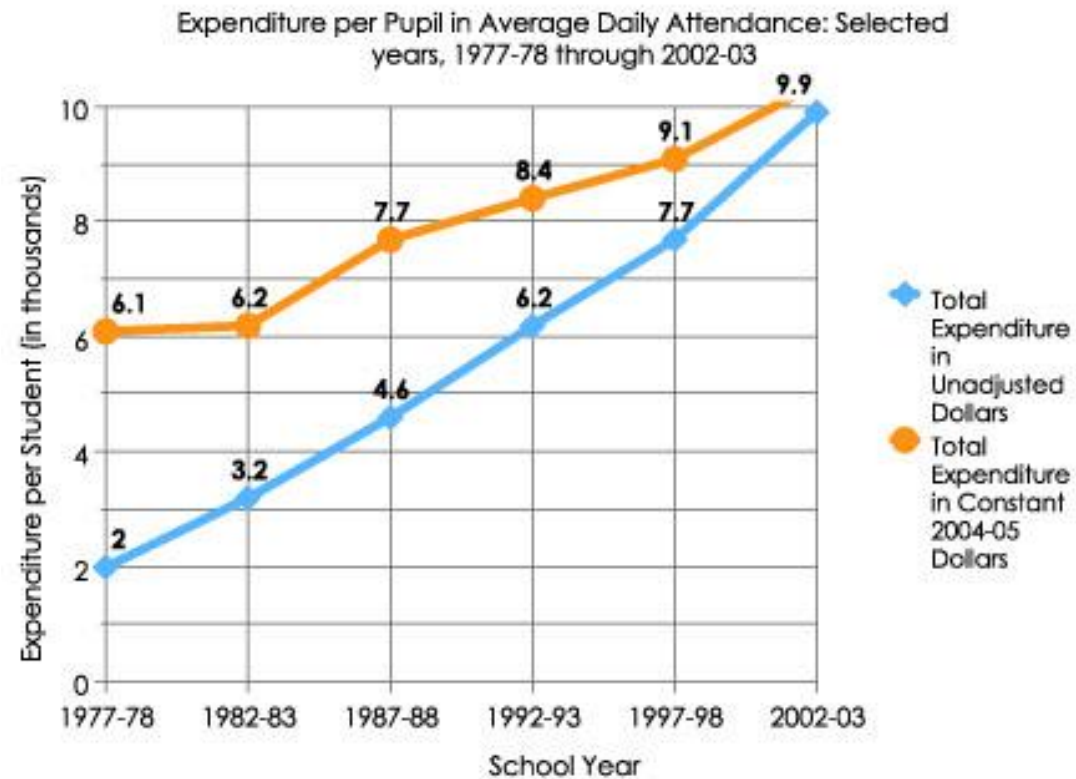




# BAR CHARTS



# LINE GRAPHS



The NCES Common Core of Data (CCD) 2004-2005

[National Center for Education Statistics](#) (accessed 10/6)

# SCATTERPLOTS



[Lumen](#) (accessed 10/6)



# GOOD VISUALIZATIONS

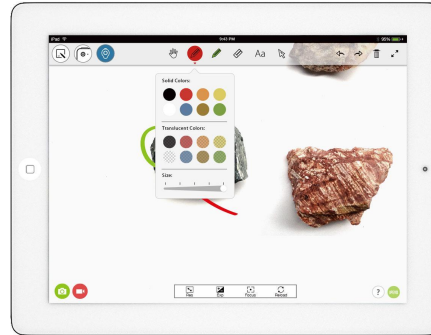
- Data positioning
- Chart choice
- Color choice
- Reduce *chart junk*



# VISUALIZATION TOOLS



Excel



Chartio and Miro

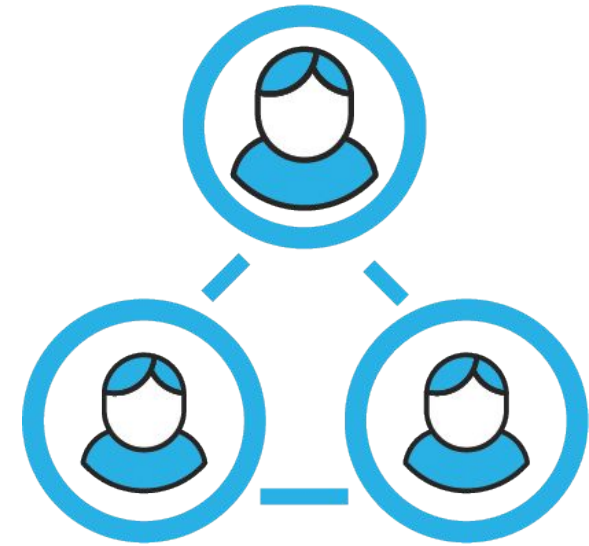


From: [Creative Commons](#)



# INTERVIEW TIME

What are the criteria for a good visualization?





# REVIEW AND WRAP-UP

Take a few minutes to write about what you have learned.



# QUESTIONS?





**BREAK  
TIME**



# LESSON 3: PRESENTING DATA



# LESSON OUTLINE

- Formats for presenting data
- Presentation pitfalls and mistakes
- Soft skills development





# DISCUSSION

“The greatest value of a picture is when it forces us to notice what we never expected to see.”  
—John Tukey, mathematician





LESSON GOALS



# WHAT ARE THE GOALS?

To identify different ways to present data

To recognize common pitfalls of data presentation

To practice soft skills such as communication, listening, and giving and receiving criticism



# WHY ARE THEY IMPORTANT?

We want to be able to communicate data stories effectively, avoiding mistakes and developing our presentation skills.



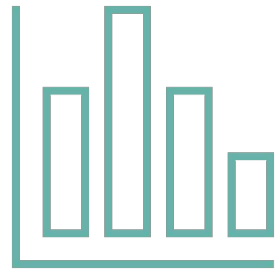
# EFFECTIVE PRESENTATIONS

An **effective presentation** communicates data clearly and generates action from the audience.

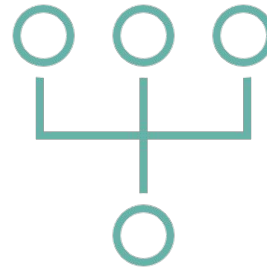




# CHOOSE THE APPROPRIATE VISUAL



**Charts**



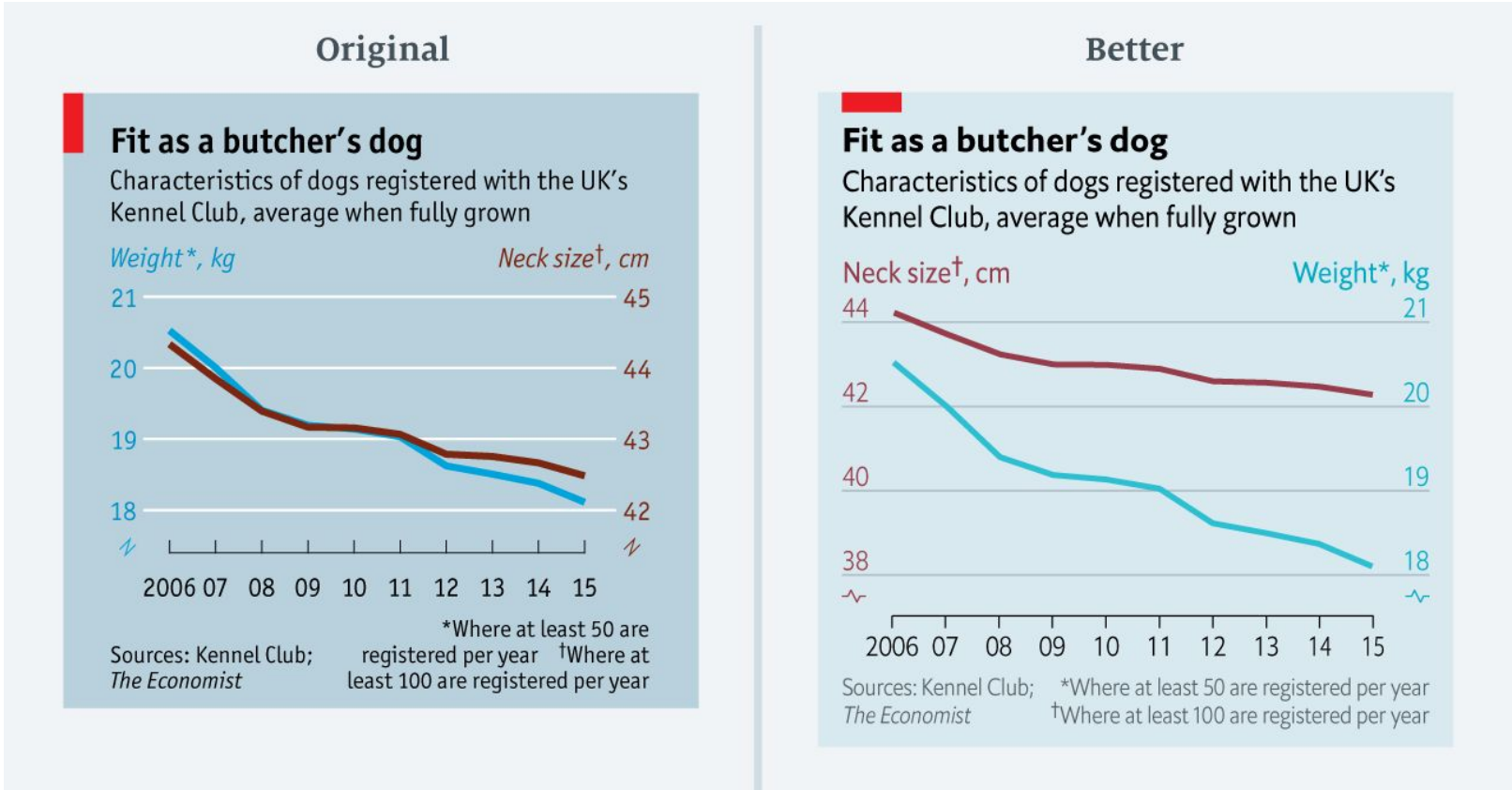
**Diagrams**



**Maps**



# SIMPLE FORMATTING



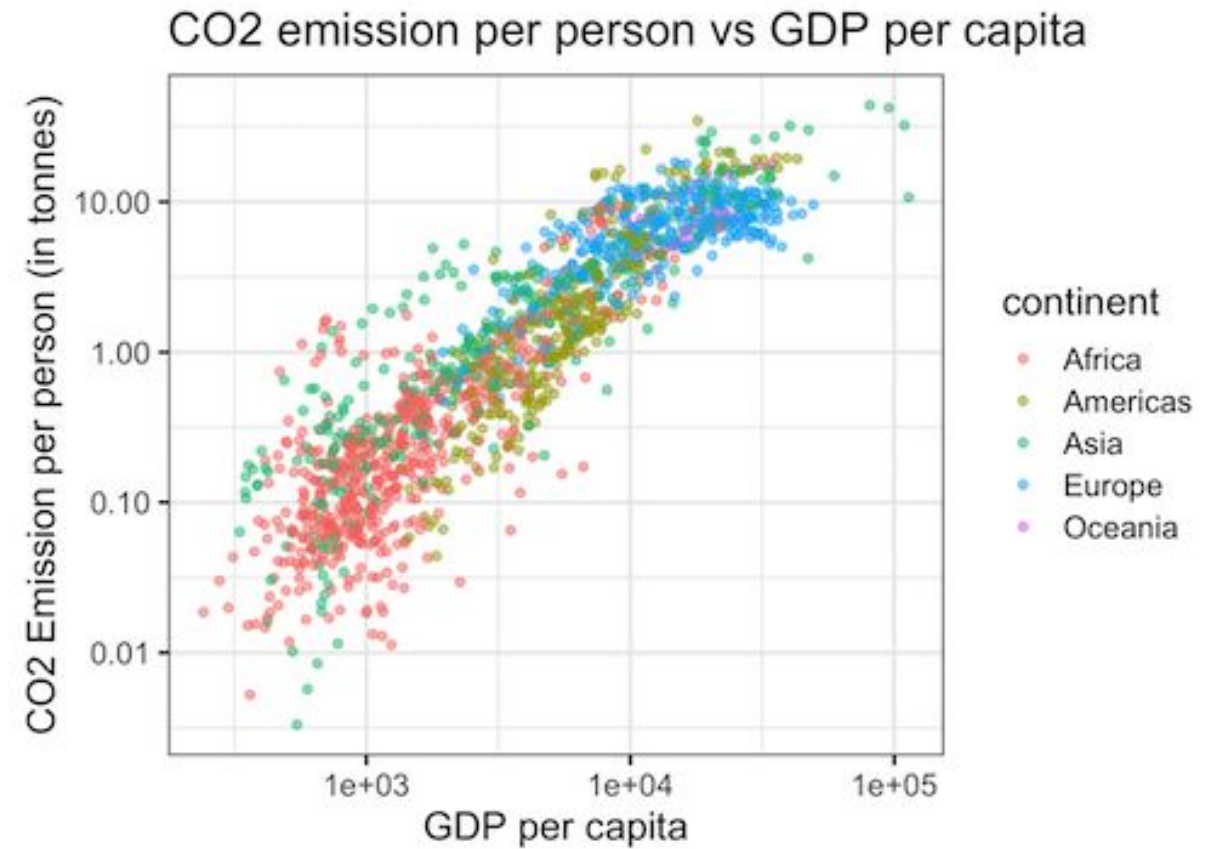
[Medium](#) (accessed 10/6)



# REALIZATION ZONES

A **realization zone** is a number or range of data that reveals the most important or crucial point of the presentation.

- Circle or shading
- Purposeful slide title
- One major point





## 1.7.3 Guided Activity

Complete the activity found in the 1.7.3 Guided Activity document to walk through making decisions about presenting sales data in a real-world scenario.



# PRESENTATION PITFALLS

1. Irrelevant or incomplete data
2. Lack of imagination
3. Inappropriate choice for visual
4. Repetitive presentation
5. Too many visuals





# PRESENT TO YOUR AUDIENCE

- Look at the group.
- Talk with your audience.
- Use slides as a reference.



From: [Austin Distel on Unsplash](#)



## 1.7.3 INDEPENDENT ACTIVITY

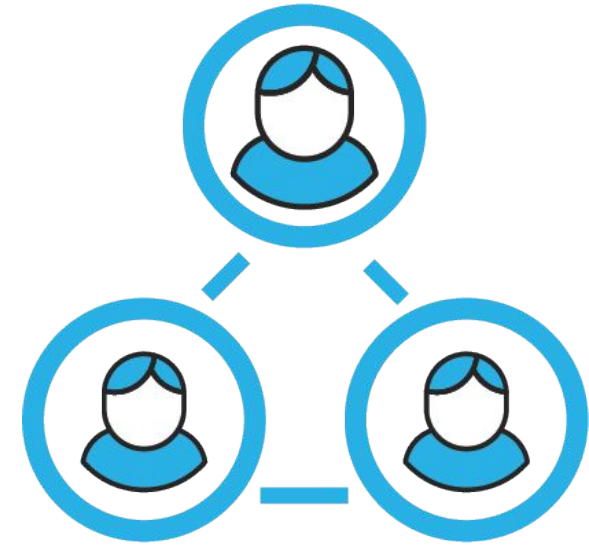
Open the 1.7.3 Independent Activity file and examine the visualizations included in it. Choose a graph and make a 2- to 5-minute presentation describing the results it shows.





# INTERVIEW TIME

How would you prepare for delivering a data presentation?





# REVIEW AND WRAP-UP

Take a few minutes to write about what you have learned.





# QUESTIONS?





# NEXT STEPS



Assigned Activities



Reminders

