Data Science Narratives
CMSE 830, Fall 2025

Step 1

In this ICA, you will not be coding at all. Instead, you will focus on crafting a narrative based on the data provided to you. You will break yourselves into two subgroups, and feel free to spread out in the large classroom we have. It is very important to keep track of time because your presentation is scheduled for 9:30.

Presentations will be informal "chalkboard talks" using the whiteboards, so take this opportunity to practice a visual, verbal presentation style. Each subgroup will work with one of the two frameworks below: the ABCDE Approach or the 5W1H Framework. After your presentations, you will compare the two approaches and itemize their pros and cons based on your experience.

Divide your group into two parts, as equal in size as you can; if your group is too small (<4), combine with another small group. One subgroup will follow the **ABCDE** approach, while the other will use the **5W1H** framework. (You will learn about both of them from the presentations, so only focus on one now.) As you work, focus on how the structure of the framework influences the way you analyze the data and craft your story. Afterward, you will present your findings between your two subgroups and reflect on how each framework shaped your storytelling experience.

Note that this is a *role playing exercise* so you will need to make some specific decisions about the setting and context. Be creative!

Framework 1: ABCDE Approach

- **A** - Analytics Type: Identify the type(s) of analytics you are conducting. Is your analysis descriptive (what happened), diagnostic (why it happened), predictive (what will happen), or prescriptive (what should be done)?
- **B** - Background & Audience: Consider who your audience is. What is their level of expertise, and what are their goals? Think about what context they will need to understand your data story.
- **C** - Context: What background or comparisons should you provide to give your audience a complete understanding of the data? This could involve industry trends, previous research, or comparative benchmarks.
- **D** - Data Narrative: Now that you've set the context, what is the main story your data is telling? What are the key insights or trends that stand out? Focus on crafting a narrative that is clear, concise, and compelling.

- **E** - Execution: Finally, think about how you will present your data. What are the most effective visualizations to convey your story? How can you package your final product for maximum impact?

Framework 2: 5W1H Data Storytelling Framework

- **W**hy: Why is this analysis important? What problem are you trying to solve, and why should your audience care?
- **W**ho: Who is your audience, and who are the stakeholders involved? How do their perspectives influence your narrative?
- **W**hat: What data do you have, and what are the key insights? How will you present these insights clearly?
- **W**hen: When did the events in your data occur, and when should action be taken based on these insights?
- **W**here: Where does the data come from, and where are the key effects or results located (e.g., in a particular region, department, or demographic)?
- **H**ow: How was the analysis conducted? While you don't need to go into technical detail, providing a high-level overview of your methodology will help build trust with your audience.

Guided Discussion (Before Group Work)

1. Context and Audience: Think about how the frameworks prioritize understanding the audience and context before diving into the data. Why is this step so important in data storytelling?
2. Narrative Structure: Consider how each framework structures the narrative. The **ABCDE** approach focuses on building a strong contextual foundation before diving into the data story, while the **5W1H** framework uses a question-driven approach. How do these differences impact the storytelling process?
3. Visualization and Delivery: Both frameworks end with considerations on how to best present the data. How do the visual aspects of storytelling help make your narrative more compelling? How will you choose visualizations that match the needs of your audience?

Step 2

Now that you are familiar with your framework, it's time to apply them to a real-world dataset. For this ICA, we will use a health-focused dataset on diabetes, which contains medical and lifestyle variables such as age, BMI, glucose levels, and insulin usage.

Dataset Overview: Diabetes Dataset

- **Source**: UCI Machine Learning Repository and Kaggle
- **Attributes**: The dataset contains information such as age, body mass index (BMI), blood pressure, glucose levels, insulin levels, and other health-related factors for patients, as well as an indicator of whether or not they have diabetes.
- **Objective**: Your goal is to craft a data story around diabetes risk factors and patient outcomes.

Instructions for Subgroups:

1. **Divide into Two Subgroups**: Each group will split into two subgroups. One subgroup will use the ABCDE framework, and the other will use the 5W1H framework to build their narrative around the diabetes dataset.
2. **Data Exploration**:
   - Spend time exploring the dataset. Focus on identifying patterns and trends that are worth storytelling.
   - Key aspects to consider:
     - How can the audience understand the relationship between different health factors and diabetes risk?
     - What are the most critical variables in predicting diabetes outcomes?
     - If you get stuck there are some guiding questions below to help you get started.
3. **Framework Application**:
   - Subgroup 1 (ABCDE): Follow the **ABCDE Approach** as outlined earlier. Focus on audience, context, and crafting a narrative around the analytics type and the final execution.
   - Subgroup 2 (5W1H): Follow the **5W1H Framework** to ask the essential questions about the data, and build your narrative around those answers.
4. **Time Management**: Remember that presentations will start promptly at 9:30. Keep track of time and ensure you have enough time to organize your thoughts and prepare for your presentation.
5. **Prepare a Chalkboard Talk**: You will be presenting your analysis using the whiteboards. Keep it simple and clear—this is a storytelling exercise, so focus on making the narrative engaging and easy to understand.

Step 3

Now that you've completed your exploration of the diabetes dataset and built your narrative using the assigned framework, it's time to present your findings. The goal of the presentation is to share your story in a clear and engaging way, focusing on how the framework shaped your analysis and narrative.

Presentation Guidelines:

1. **Chalkboard Talk Format**: Your presentation will be delivered as a "chalkboard talk," meaning you will use the whiteboards in the classroom to visually communicate your story. This is your opportunity to practice conveying insights without relying on slides—focus on clarity, flow, and engagement.
2. **Framework Explanation**:
   ○ Briefly explain which framework you used (**ABCDE** or **5W1H**).
   ○ Describe how the framework guided your analysis and storytelling. For example:
      ▪ Did the **ABCDE** framework help you think more about the audience and context?
      ▪ Did the **5W1H** framework help you organize your thoughts logically?
3. **Key Insights**:
   ○ Present the key findings from your analysis. Make sure to highlight the most important insights from the dataset.
   ○ If relevant, provide any recommendations or actions based on the data (e.g., health interventions, policy suggestions, etc.).
4. **Visualization**:
   ○ Use the whiteboard to sketch out any visualizations that support your story. These could include:
      ▪ Trends over time
      ▪ Comparisons between groups (e.g., age, BMI, glucose levels)
      ▪ Key relationships between variables (e.g., glucose and diabetes outcomes)
   ○ Keep your visuals simple and clear, ensuring they help communicate your narrative effectively.

Time Management:

● Each subgroup will have **6-8 minutes** to present their story, with an additional **2 minutes for Q&A**. Stay focused on the key points and avoid getting into technical details that aren't essential to the narrative.
● Ensure you leave time for the Q&A, where your peers will ask questions about your story and framework.

Comparison and Discussion:

After both subgroups have presented, you will compare your experiences with the frameworks. This is a chance to reflect on the pros and cons of each approach:

- **Which framework felt more intuitive** to use in the context of this dataset?
- **Did one framework** help clarify the narrative better than the other?
- **What were the challenges** in using each framework? Did either framework limit your storytelling in any way?
- **What would you change** in how you approached the data or the storytelling process?

The goal of this discussion is to help you understand how different storytelling frameworks can shape the way you analyze and present data. By comparing the two approaches, you'll gain insight into how to choose the right storytelling method for different types of data and audiences.

Step 4

Now that you've completed your presentations and compared the two frameworks, it's time to reflect on your experience. This step is designed to help you think about what you learned from the exercise and how you can apply data storytelling techniques in future data science projects.

Individual Reflection:

- **What did you learn** about data storytelling? How did working with a specific framework change the way you thought about the dataset?
- **Which framework** felt more natural to you? Would you choose a particular framework again for a different type of data or audience?
- **What was challenging** about using these frameworks? How might you improve your approach to data storytelling in the future?

Write down a few sentences summarizing your reflections. These reflections will help you connect the lessons from this ICA to your broader learning in data science.

Group Feedback:

- As a group, discuss what worked well and what could be improved in your presentations. Did you feel your narrative was clear and engaging?
- How could the **presentation format** (chalkboard talks) be improved? What aspects of the discussion and comparison were most valuable?

Next Steps:

- Think about how you can apply these data storytelling frameworks to future assignments or projects. When analyzing new datasets, consider which framework might help you best communicate the story behind the data.
- Keep these frameworks in mind as you move through the course, as data storytelling is an essential skill in data science, whether you're presenting to stakeholders, writing reports, or crafting visualizations.

Guiding Questions for Data Exploration

General Questions (Applicable to Both Frameworks):

1. **What patterns** do you notice in the data? Are there clear relationships between any variables and diabetes outcomes (e.g., BMI, glucose levels)?
2. **What are the key variables** in the dataset? Which factors seem to have the strongest influence on predicting diabetes?
3. **How does the data change** across different demographic groups (e.g., age, gender)? Do you see any disparities in outcomes?
4. **Are there any surprising insights** or trends that stand out from the data?
5. **What potential actions** could be taken based on your findings? (e.g., public health recommendations, lifestyle changes for at-risk individuals)
6. **What are the limitations** of the data? Are there missing variables, biases, or issues that could affect the accuracy of the story?

For Subgroup 1 (ABCDE Framework):

- **A - Analytics Type**: Is your analysis primarily descriptive (what happened), diagnostic (why it happened), or predictive (what will happen)? How do you want to frame your story?
- **B - Background & Audience**: Who is your audience, and how familiar are they with medical or health data? What do they care about?
- **C - Context**: What background information do you need to provide for your audience to understand the trends in the dataset?
- **D - Data Narrative**: What's the main takeaway or story you want to convey? How does the data support your narrative?
- **E - Execution**: What are the best ways to visualize the data for your audience? How can you present your story in a clear and compelling manner?

For Subgroup 2 (5W1H Framework):

- **Why**: Why is the analysis important? What problem are you solving, and why should the audience care about diabetes risk factors?
- **Who**: Who is impacted by the findings in this dataset? Who should take action based on this data (e.g., healthcare providers, individuals, policymakers)?
- **What**: What are the critical data points or trends you've discovered? What story do the data tell?

- **When**: When does the data show these effects? Are these trends consistent over time, or is there a specific period where the effects are more pronounced?
- **Where**: Where do the trends in the data apply? Are there specific regions, demographics, or subgroups that are most affected?
- **How**: How was the analysis conducted, and how can your insights be used in a real-world context (e.g., policy recommendations, individual health interventions)?