Module 1: How to Navigate a World of Big Data

Background

This module serves two primary purposes:

1. **Orientation**: Help learners understand major concepts related to "Big Data," both within their organization and in the broader world, while identifying their role in the Big Data ecosystem.
2. **Workshop Introduction**: Introduce the workshop's purpose, motivation, and activity flow.

Note: Depending on the group composition, consider spending time before formally launching this module to build camaraderie and mutual understanding. Learners will not benefit from the workshop series if they are unable or unwilling to participate in open discussions and activities.

Additional Reading for Facilitators

- Barocas, Solon, and Andrew D. Selbst. *Big Data's Disparate Impact*. 2021, p. 63.
- boyd, danah, and Kate Crawford. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society*, vol. 15, no. 5, 2012, pp. 662–79.
- Dijck, Jose van. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance & Society*, vol. 12, no. 2, May 2014, pp. 197–208.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. First Mariner Books edition, Mariner Books, Houghton Mifflin Harcourt, 2014.
- Milan, Stefania, and Emiliano Treré. "Big Data from the South(s): Beyond Data Universalism." *Television & New Media*, vol. 20, no. 4, 2019, pp. 319–35.
- Pietsch, Wolfgang. *Big Data – The New Science of Complexity*. 2013.

Module Motivation

By the end of this module, learners will:

- Understand what "Big Data" is, how it is collected, and how it is used
- Engage in critical discussions about the ethics surrounding Big Data collection and usage
- Understand the importance of knowing big picture concepts when working with data
- Identify their role in the world of Big Data

Course Introduction

Key Points to Emphasize:

- Welcome learners and establish "ground rules" for the workshop series
- Acknowledge that learners are already strong critical thinkers with valuable lived experience and perspectives
- Explain that this course will help channel those skills into the world of data and "Big Data"

Workshop Series Structure:

- **Six modules** in total
- **Standard format** for each module:
    - Pre-activity to set the tone
    - Session roadmap
    - Primary activity with discussion
    - Journal entry for reflection and preview of next module

Learning Goals

Students will be able to:

1. Identify file tree structure and develop appropriate file-system schema for future projects
2. Practice thinking through potential uses of data, both positive and negative
3. Document (via logic chains) how data can be used to support a judgment and practice doing so in writing
4. Identify the data logic chain someone else has used

Activity 1: File Systems

Purpose:

Introduction to documentation systems used throughout the workshop series.

Instructions:

1. **For new learners**: Walk through well-organized file-system examples
2. **Introduce notation**: Use file-system notation (e.g., | and _ for nested files)
3. **Hands-on practice**: Have learners set up their own workspace schema for the workshop series

Real-World Application:

At DataWorks, this activity served as the foundation for restructuring our Team Drive and better organizing files across dozens of client projects in a cohesive, unified fashion.

Activity 2: Evidence-Based Assertions

Purpose:

Demonstrate how to build data-based arguments and examine others' arguments critically.

Key Discussion Points:

- What data are they using and how are they using it to support their argument?
- Does the data lend legitimacy to their argument?
- Is that the only reading of the data?
- How can we present findings respectfully, considering:
    - Dataset origins
    - Collection motivations
    - Potential limitations (sample size, included/excluded data points)

Facilitation Notes:

- Treat slides as discussion prompts rather than presentations
- Encourage group participation as community-building
- Understanding how each member builds arguments is crucial for future collaboration

Important Concepts to Address:

Correlation vs. Causation

- Often misunderstood as synonymous
- Critical distinction in data work where data is often viewed as "truth" or "objective"

Quantitative vs. Qualitative Investigation Styles

Example from DataWorks:

- **Qualitative approach**: Interviews with community members about pressing issues facing Atlanta, understanding why previous interventions failed

- **Quantitative approach**: Tracking rationale for pet surrenders to identify major causes for separation

Data Investigation Sub-Activity

Dataset: Environmental data from NOAA's Climate Data Online

Instructions:

- Learners investigate the dataset
- Create journal entries answering provided questions
- Remember: No right or wrong answers, many questions are subjective
- Goals:
    - Familiarize learners with data "cleanliness" concepts
    - Help team members understand each other's perceptions of data cleanliness
    - Facilitate future shared work

Reflection & Journal Entry

Complete this reflective exercise at the end of the module, paired with the data investigation sub-activity.

Post-Session Work for Facilitator

Choose one approach based on group makeup:

Option 1: Collect and reply to journals individually after the module

Option 2: Review submissions and use them as the basis for a small discussion or group-recollection activity at the beginning of the next module