Lecture 3
Data Literacy

Wan Fang

Southern University of Science and Technology

Agenda

- Design with Analytics
  - Analytical Thinking vs. Creative Intuition
  - Common Patterns of Design + Analytics

- The Importance of Context
  - 6 basic problem types
  - Who, What, and How
  - Questions to ask

"In this world of big data, basic data literacy—the ability to analyze, interpret, and even question data—is an increasingly valuable skill," says Harvard Business School Professor Janice Hammond
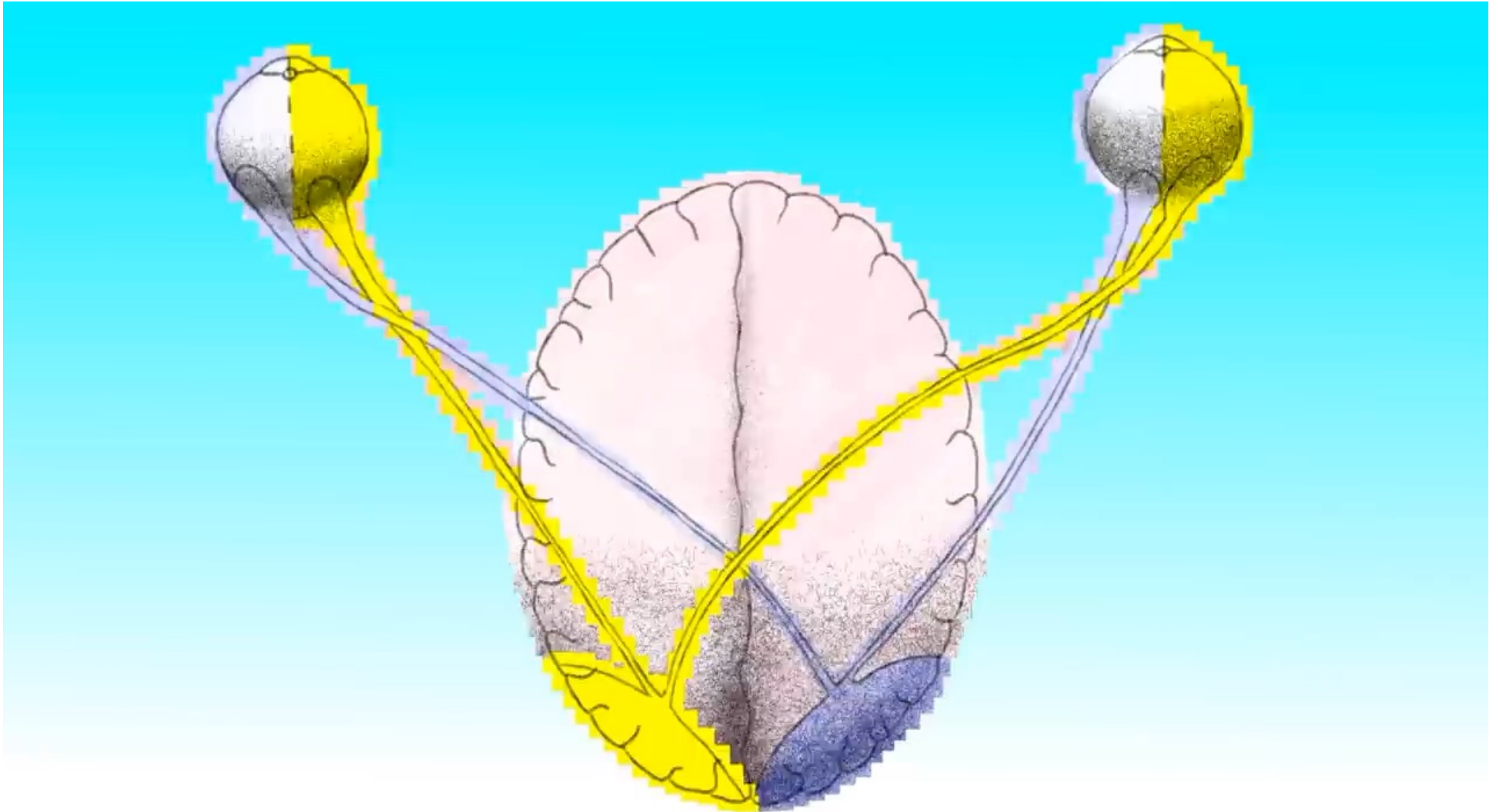
Design with Analytics

Wan Fang

Southern University of Science and Technology

[Adapted from Data Analytics for Designers by Tak Yeon Lee]

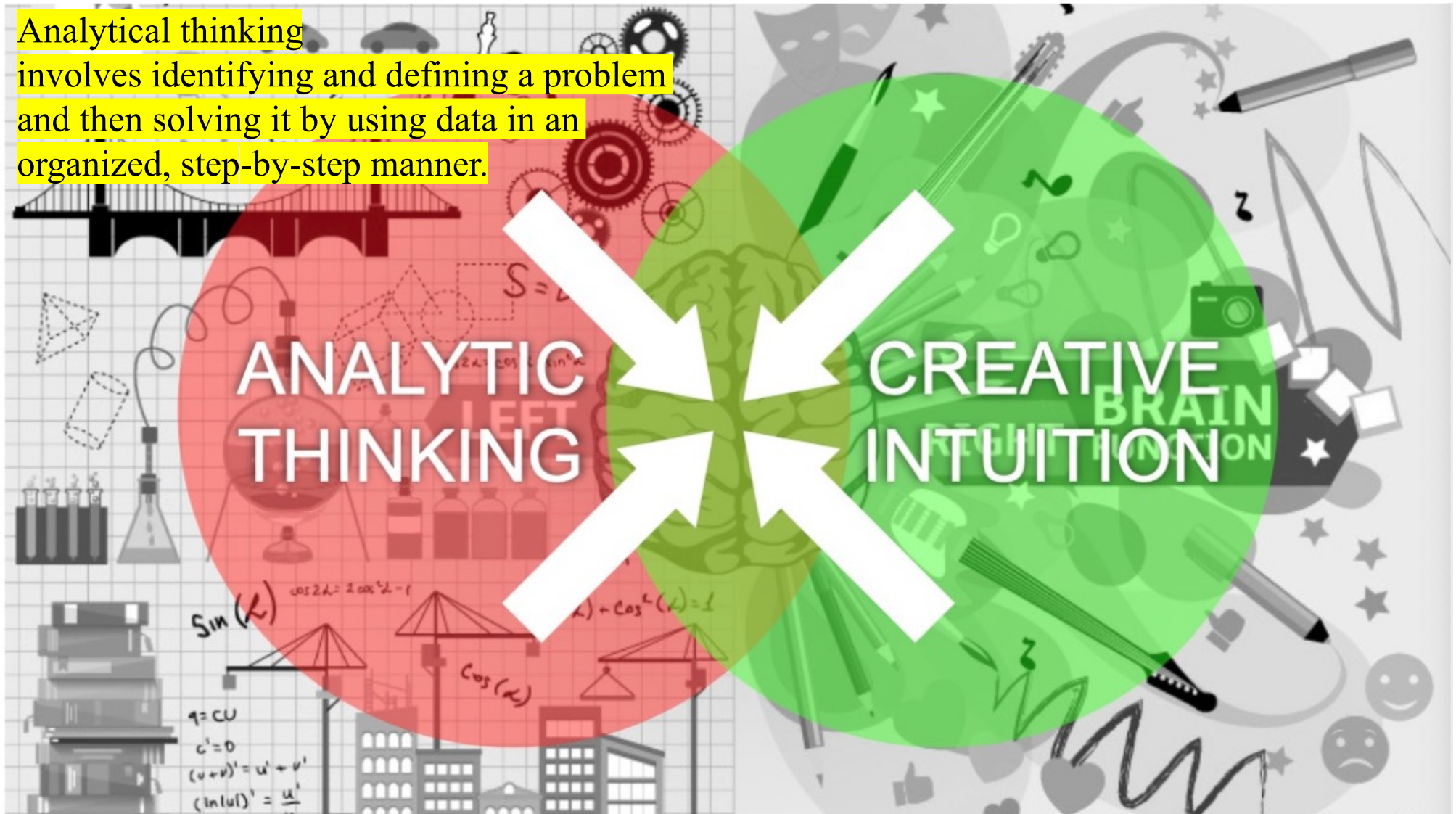# The Left Brain vs. Right Brain Myth

The Left Brain vs. Right Brain Myth

Both sides of the brain are needed!

Analytical thinking involves identifying and defining a problem and then solving it by using data in an organized, step-by-step manner.

Five essential aspects of analytical skills

- **Curiosity:** a desire to know more about something, asking the right questions

- **Understanding context:** understanding where information fits into the "big picture"

- **Having a technical mindset:** breaking big things into smaller steps

- **Data design:** thinking about how to organize data and information

- **Data strategy:** thinking about the people, processes, and tools used in data analysis

Why Analytic Thinking is important for designers?

- Analytic thinking gives designers new business opportunities
  - Digital Transformation has created a lot data-centric businesses.
  - Analytic Thinking is a good way for designers to find new opportunities.
  - In addition, data literacy (a basic AT skill) is the key to communicate with engineers and business experts

- Analytic thinking help designers improve creative intuition
  - While being a strong asset of designers, creative intuition could be biased or fixated based on prior knowledge.
  - Analytic thinking is one of the best ways to test and fix biases in designer's creative intuition.

Why Creative Intuition is important for data analytics

- Knowing <u>what the data is telling</u> / not telling us
- Knowing <u>where the look</u> next
- Knowing <u>when to stop</u> looking and take action
- Knowing <u>who needs to hear</u> and how to get through to them
- Knowing <u>why any of it matters</u> in the first place

While analytic thinking is hands-on tactical skills,
Intuition helps us make **high-level strategic decisions**

Common Patterns of Design + Analytics

Paper: Understanding User Behavior Through Log Data and Analysis

	Observational	Experimental
Lab Studies *Controlled interpretation of behavior with detailed instrumentation*	In-lab behavior observations	In-lab controlled tasks, comparison of systems
Field Studies *In the wild, ability to probe for detail*	Ethnography, case studies, panels (e.g., Nielsen)	Clinical trials and field tests
Log Studies *In the wild, little explicit feedback but lots of implicit signals*	Logs from a single system	A/B testing of alternative systems or algorithms

Table 1. Different types of user data in HCI research.

- Log Analysis
- Designing the entire data pipeline
- Before and After AI Modeling

Source

Log Analysis

- Log analysis is a common method to understand user behavior.

- While off-the-shelf analytic tools allows (e.g. Google Analytics) to get basic insights, designers want much more:
  - (1) Exploratory analysis,
  - (2) Gain insights via custom metrics and experiments at scale, and
  - (3) Data storytelling.

- As results of log analysis, designers usually create reports or dashboard to be shared within the organizations.
- Going further, insights gained from log analysis can guide designers toward intelligent data-driven services and features.
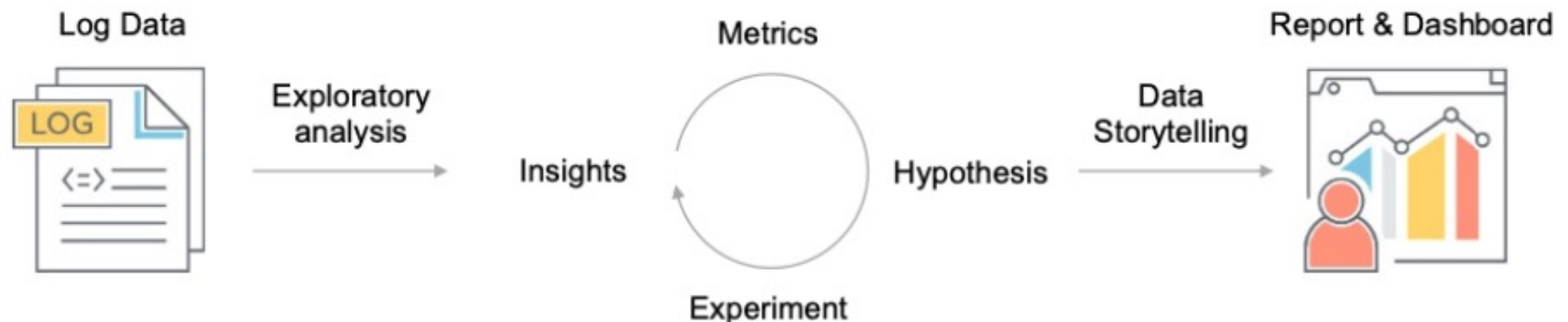
Log Analysis Tools

LogAI: A Library for Log Analytics and Intelligence

Designing the entire data pipeline

Real World   **Data collection**   **Collected Data**   **Data-driven Features**
(e.g. Report, Dashboard, Recommendation, …)
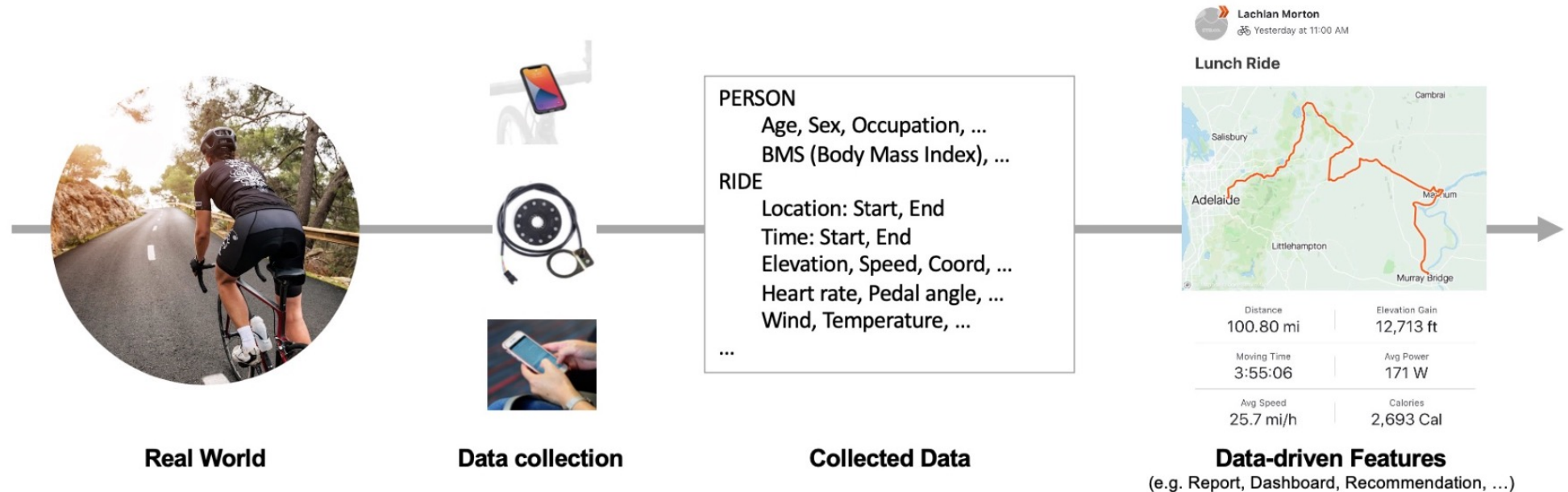
- Notice that the entire pipeline can be designed.
  - CASE 1. If the activity does not exist, designers must envision every step from scratch.
  - CASE 2. If there's a collected dataset, analysts would first examine collected data to develop **data-driven features**. If the dataset is not satisfactory, they would try different collection methods. If cyclists' behavior does not have signal of interest, they can even redesign cycling.
- Both cases require analytic thinking and creative intuition at the same time.

Before ~~and After~~ AI Modeling

Common Issues

Irrelevant or Underspecified problem definition

Biased, incomplete, insufficient dataset

No validation on usability issues of the model in the context of real usage

Getting lots of feedback from users, but it's too late to fix the model

Less than 10% of AI projects are actually used

Current

Problem Definition

Data Preparation

Train

Test

Product Development

Usability Test

Deploy

Roles of Designers with Analytic Skills

Understanding Problem — Through interview, observation, and survey, designers gain in-depth understanding of the problem domain and existing dataset

Exploratory Data Analysis — Designers look into the existing dataset, and find issues hidden underneath, and propose how to fix them

Data Prototyping — Even if no dataset is available, designers can prototype a data schema based on their problem and data understanding

Wizard-of-Oz Prot. — AI features are ideal for Wizard-of-Oz prototyping if designers have clear understanding of the dataset and the model's capability

This class teaches skills for Exploratory Data Analysis (EDA). However, the other tasks also require data understanding which is gained via EDA.

Keep in mind how creative intuition helps data analytics

1. Knowing **what the data is telling us** and not telling us
2. Knowing **where the look next**
3. Knowing **when to stop looking** and take action
4. Knowing **who needs to hear** and how to get through to them
5. Knowing **why any of it matters** in the first place

~~Before and~~ After AI Modeling

Common Issues

Irrelevant or Underspecified problem definition

Biased, incomplete, insufficient dataset

No validation on usability issues of the model in the context of real usage

Getting lots of feedback from users, but it's too late to fix the model

Less than 10% of AI projects are actually used

Current

Problem Definition

Data Preparation

Train

Test

Product Development

Usability Test

Deploy

Roles of Designers with Analytic Skills

Understanding Model Performance	Most AI models make mistakes, designers must have in-depth understanding of when / how AI models fail.
Make the worst case scenario tangible	Not only the best but also the worst-case scenarios (i.e. model returning wrong (or even no) results) should be designed
Gatekeeping data-driven features	If a data-driven feature does not have significant benefits for users, designers would drop / redesign it.
Envision Iterative Cycles	In usual, data-driven AI-infused systems are not complete at their first versions. Designers propose a roadmap (i.e. iterative cycles of development) rather than finalized systems.

The Importance of Context

cole nussbaumer knaflic

storytelling
with
data

a data
visualization
guide for
business
professionals

WILEY

Source

6 basic problem types

Discovering Connections

Interdisciplinary Links

Comparative Studies

Finding Patterns

Statistical Analysis

Behavioral Patterns

Making Predictions

Forecasting Trends

Risk Assessment

6 basic problem types

Categorizing Things

Data Organization

Classifying Information

Spotting Unusual Things

Anomaly Detection

Error Identification

Identifying Themes

Literary Analysis

Art Interpretation

Who, What, and How

- ***To whom are you communicating?***
    - It is important to have a good understanding of <u>who your audience is</u> and <u>how they perceive you</u>. This can help you to **identify common ground** that will help you <u>ensure they hear your message</u>.

- ***What do you want your audience to know or do?***
    - You should be clear <u>how you want your audience to act</u> and take into account <u>how you will communicate to them</u> and <u>the overall tone that you want to set for your communication</u>.

- ***How can you use data to help make your point?***
    - It's **only after** you can <u>concisely answer these first two questions</u> that you're ready to move forward with the third.

Who: Your audience

- **The more specific you can be about who your audience is, the better position you will be in for successful communication.**

    - *Avoid general audiences*, such as "internal and external stakeholders" or "anyone who might be interested"—by trying to communicate to too many different people with disparate needs at once, you put yourself in a position where you can't communicate to any one of them as effectively as you could if you narrowed your target audience.

    - Sometimes this means *creating different communications for different audiences*.

- **Identifying the decision maker** is one way of narrowing your audience.

    - Understand how to **resonate** with them and form a communication that will meet their needs and yours.

Who: You

- **Think about the relationship that you have with your audience** and **how you expect that they will perceive you**.

  - *Will you be encountering each other for the first time through this communication?*
  - *Do you have an established relationship?*
  - *Do they already trust you as an expert, or do you need to work to establish credibility?*

- These are important considerations when it comes to determining how to structure your communication and whether and when to use data, and may impact the order and flow of the overall story you aim to tell.

What: Action

- ***What do you need your audience to know or do?***
  - This is the point where you think through how to make what you communicate **relevant** for your audience and form a clear understanding of why they should care about what you say
  - You should always want your audience to know or do something. If you can't concisely articulate that, you should revisit whether you need to communicate in the first place

The audience knows better than the presenter and therefore should choose whether and how to act on the information presented.

What: Action

- ***What do you need your audience to know or do?***
  - This is the point where you think through how to make what you communicate **relevant** for your audience and form a clear understanding of why they should care about what you say.
  - You should always want your audience to know or do something. If you can't concisely articulate that, you should revisit whether you need to communicate in the first place

~~The audience knows better than the presenter and therefore should choose whether and how to act on the information presented.~~

> This assumption is false. If you are the one analyzing and communicating the data, you likely know it best.

When it really isn't appropriate to recommend an action explicitly, <u>encourage discussion toward one.</u>

> Suggesting possible next steps can be a great way to get the conversation going because it gives your audience something to react to rather than starting with a blank slate.
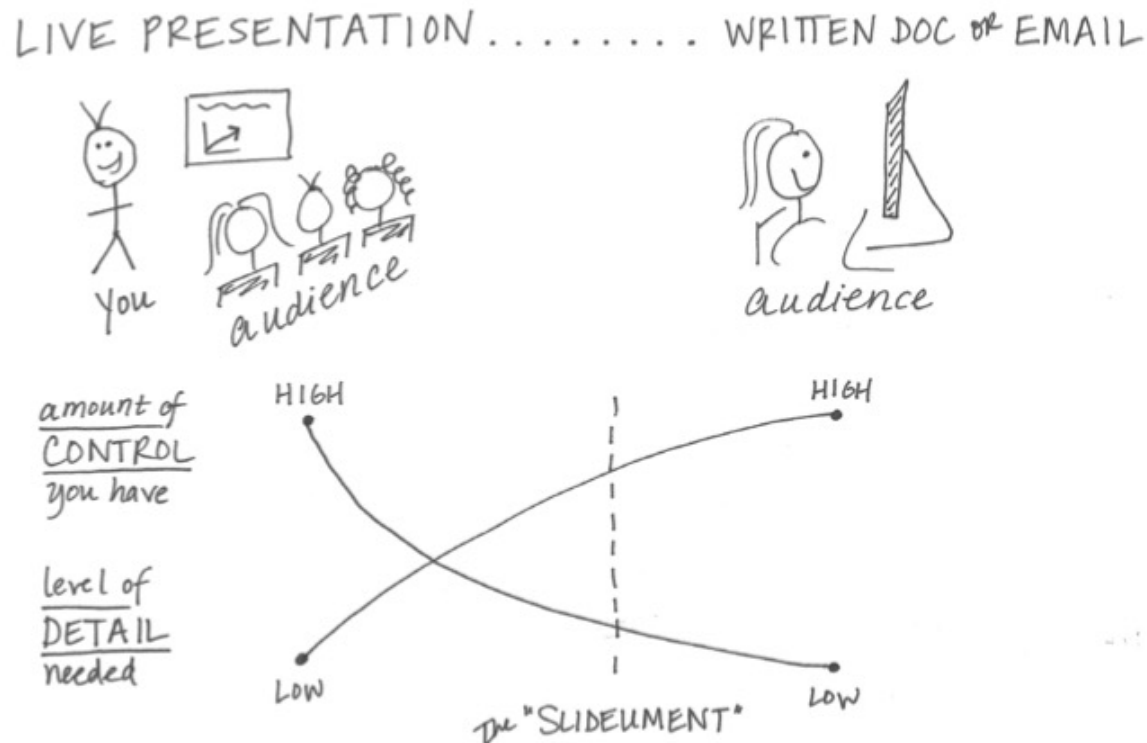
Prompting action

Here are some action words to help act as thought starters as you determine what you are asking of your audience:

accept | agree | begin | believe | change | collaborate | commence | create | defend | desire | differentiate | do | empathize | empower | encourage | engage | establish | examine | facilitate | familiarize | form | implement | include | influence | invest | invigorate | know | learn | like | persuade | plan | promote | pursue | recommend | receive | remember | report | respond | secure | support | simplify | start | try | understand | validate

What: Mechanism

- ***How will you communicate to your audience?***
  - The method you will use to communicate to your audience has implications on a number of factors, including the ==amount of control== you will have over how the audience takes in the information and the <mark>level of detail</mark> that needs to be explicit.
  - We can think of the communication mechanism along a continuum, with live presentation at the left and a written document or email at the right

LIVE PRESENTATION . . . . WRITTEN DOC OR EMAIL

You   audience   audience

amount of **CONTROL** you have

level of **DETAIL** needed

HIGH   HIGH

LOW   LOW

The "SLIDEUMENT"

What: Tone

- ***What tone do you want your communication to set?***

  - Are you celebrating a success? Trying to light a fire to drive action? Is the topic lighthearted or serious?

  - The tone you desire for your communication will have implications on the design choices.

  - For now, think about and specify the general tone that you want to establish when you set out on the data visualization path.

How

- … **only after** we can clearly articulate *who our audience is* and *what we need them to know or do*—we can turn to the data and ask the question

- *What data is available that will help make my point?*
  - Data becomes supporting evidence of the story you will build and tell.

Ignore the nonsupporting data?

You might assume that showing only the data that backs up your point and ignoring the rest will make for a stronger case. I do not recommend this. Beyond being misleading by painting a one-sided story, this is very risky. A discerning audience will poke holes in a story that doesn't hold up or data that shows one aspect but ignores the rest. The right amount of context and supporting and opposing data will vary depending on the situation, the level of trust you have with your audience, and other factors.

Let's consider a specific example

- **Imagine you are a fourth-grade science teacher.**

- You just wrapped up <u>an experimental pilot summer learning program</u> on science that was aimed at giving kids exposure to the unpopular subject.

- You <u>surveyed the children</u> at the onset and end of the program to understand whether and how perceptions toward science changed.

- You <u>believe</u> the data shows a great success story.

- You <u>would like to continue</u> to offer the summer learning program on science going forward.

Please identify the *Who*, *What*, and *How*

- *who we have identified as our audience,*
- *what we need them to know and do, and*
- *the data that will help us make our case*

Identify the *Who*, *What*, and *How*

- **Who**: The budget committee that can approve funding for continuation of the summer learning program.

- **What**: The summer learning program on science was a success; please approve budget of $X to continue.

- **How**: Illustrate success with data collected through the survey conducted before and after the pilot program.

The *Who*, *What*, and *How* identified
- *who we have identified as our audience,*
- *what we need them to know and do, and*
- *the data that will help us make our case*

Questions to ask [for context]

- *What background information is relevant or essential?*

- *Who is the audience or decision maker? What do we know about them?*

- *What biases does our audience have that might make them supportive of or resistant to our message?*

- *What data is available that would strengthen our case? Is our audience familiar with this data, or is it new?*

- *Where are the risks: what factors could weaken our case and do we need to proactively address them?*

- *What would a successful outcome look like?*

- ***If you only had a limited amount of time or a single sentence to tell your audience what they need to know, what would you say?***

Craft effective questions

- SMART methodology

  - Specific — does the question address the problem? Does it have a context?

  - Measurable — does it give the answer that can be measured?

  - Action-oriented — will the info that we get help us devise an action plan?

  - Relevant — is it about a particular problem we are trying to solve?

  - Time-bounded — are the answers relevant to the specific time being studied?

Example of SMART questions

- What features do people look for when buying a new car?

  - Specific: Does the question focus on a particular car feature?

  - Measurable: Does the question include a feature rating system?

  - Action-oriented: Does the question influence creation of different or new feature packages?

  - Relevant: Does the question identify which features make or break a potential car purchase?

  - Time-bound: Does the question validate data on the most popular features from the last three years?

Thank you~

Wan Fang
Southern University of Science and Technology