Slide 1: How We Managed Complexity

- **Implemented a Logical Data Structure**
    - **/data/raw** for original, untouched source files.
        - Ensured a non-destructive and reproducible workflow.
    - **/data/curated** for final, analysis-ready datasets.
        - Simplified debugging and streamlined the pipeline.
- **Treated Documentation as a Core Asset**
    - Centralized all notes, data dictionaries, and source information.
    - Explicitly documented limitations and compromises to improve transparency.

Slide 2: Conclusion: Key Lessons

- **Plan for Imperfection**
    - Real-world data is messy and requires flexible planning.
- **Document Everything**
    - Recording limitations and compromises is crucial for transparency.
- **Discover Data First**
    - Preliminary data exploration is critical for setting realistic project goals.

So, how did we manage this complexity? We implemented logical structure to our data organization. On our GitHub repository, you'll see two main folders: /data/raw and /data/curated."

"The **/data/raw** folder contains all the original, untouched files exactly as we downloaded or fetched from API. This created a non-destructive workflow, meaning we could always go back to the original source. The **/data/curated** folder is where we stored our final, analysis-ready datasets after all the cleaning, merging, and aggregation. This separation made our entire pipeline reproducible and much easier to debug."

"Finally, we treated our documentation as a core asset. We centralized all our notes, data dictionaries, and source information in **README.md** and **metadata.json** files. Most importantly, we used these files to explicitly state the limitations and compromises we made during curation, which improve our data processing transparency

So, to wrap things up, what are the big takeaways from our project?"

(Move to Conclusion Slide)

This project taught us three key lessons:

- First, **plan for imperfection**; real-world data is messy and not always straight to the point we want.
- Second, **document everything**; record every limitation and compromisations we made to build transparency.
- And finally, **discover data first**; a preliminary data check is critical to setting realistic goals.

"Thank you. I'd be happy to answer any questions you might have."