How to Ask Your Analysts about Equity in your Model

The Basic Idea:

A data analyst uses variables to create a model that is the best possible representation of the world as they see it. They combine all of their knowledge, experience, opinions and other information to build this model.

The variables they choose (and don't choose) reflect a particular perspective on what's important/ relevant and there is no purely mathematical way to determine which variables to use. Additionally, the way those variables are used in the model - whether the variable is having a moderating, mediating or confounding effect - is mutually exclusive and subjective. It's not a bad thing, it's just reality.

Because expectations, assumptions and pre-existing ideas are part of how analysts create data models, it's important to at least be transparent about what choices were made with the variables and ideally adjust those choices to reflect the equity that the project is aiming for.

Tactful Opening Lines:

"Could you help me understand what the different parts of your model mean?"

"Hey, I'm trying to explain to someone what's being controlled for in our model and how; can you give me a hand?"

"Would you be able to walk me through the assumptions we're embedding in this model?"

"Hi, can we have a chat about how our model is accounting for moderators, mediators, and confounders?"

What this is *not* about:

P-hacking.
A crisis in reproducibility.
Placing blame.

Further Reading:

www.nature.com/articles/s41586-020-2314-9

https://hbr.org/2015/03/what-to-do-when-people-draw-different-conclusions-from-the-same-data

WE ALL COUNT

3 Steps towards Distributing Data Products Equitably

The Basic Idea:

Often, very equitable data projects trip at the finish line by packaging and distributing their information in a way that prioritizes the most privileged audience. What you say, how you say it and how your audience interacts with it matters a lot.

Step 1:

Choose your target audience in a way that reflects the equity you aimed for in the project. Note that "everyone" is not an effective audience target and when people aim for "universal" they usually default to the preferences of the most powerful stakeholders in the project.

Step 2:

Once you've identified your target (or targets, if you can make multiple versions of your data product, by all means go for it!) go through the checklists (on the back!)

Step 3:

Making the choices in these checklists can be done in a variety of ways. The best is to test a variety of styles, mediums, and narratives directly with your target audience.

The next best is to get advice from lived-experience experts that can represent some of your target audience.

Still worthwhile is just imagining what might work best for your audience, empathy is a powerful tool.

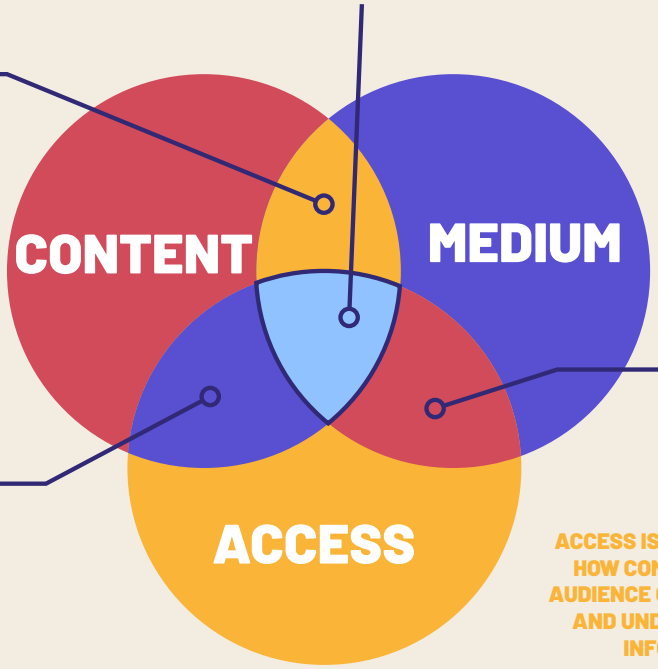
Real-world example:

A large non-profit organization (unnamed for privacy reasons) recently changed the content and format of its Annual General Report to prioritize the communication style and format that's most comfortable for its most marginalized clients (they tested this) rather than its wealthiest donors.

They were quite worried about making the donors mad, but instead, the donors loved the idea, some preferred the new style, *and* the donors shared this AGR more widely and proudly than any others the staff could remember.

WE ALL COUNT

WORKS FOR 'THIS' AUDIENCE



The content and medium work together to increase clarity and offer better depth.

CONTENT IS WHAT YOU INCLUDE IN YOUR COMMUNICATION AND HOW YOU CHOOSE TO SAY IT.

The content is created considering accessibility, including comprehension and cultural appropriateness.

MEDIUM IS WHAT FORM YOUR COMMUNICATION TAKES, LIKE THE DIFFERENCE BETWEEN A PAPER, AN ANIMATED VIDEO AND A SEARCHABLE DATABASE.

The medium of distribution supports the ideal accessibility of the information for each audience, including feedback and engagement.

ACCESS IS HOW EASILY AND HOW COMPLETELY YOUR AUDIENCE CAN ENGAGE WITH AND UNDERSTAND YOUR INFORMATION.

WE ALL COUNT

- **Subjects** - what should be included or excluded?
- **Language** - what type of words are best understood?
- **Length** - what is the ideal length?
- **Depth** - how deep into the data and method?
- **Tone** - What is the most equitable emotional undertone?
- **Perspective** - Whose POV is the information presented in?
- **Narrative** - What story type are you telling?
- **Clarity** - Who can understand the information?
- **Complexity** - Is it too complicated? Or too simple?
- **Relevance** - Is there meaning for the audience?
- **Cultural Translation** - Use of appropriate terminology, structure, stories and symbols for the audience.

- **Digital vs. Print** - what is the audience's preference?
- **Static vs. Dynamic** - video, animation, transitions?
- **Interactive vs. Demonstrative** - dashboard vs. slideshow, etc.
- **Branding** - are there secondary messages?
- **Institutional Presence** - Who vouches for the information?
- **Live vs. Standalone** - Will the information change?
- **Isolated vs. Network** - Is the info in an ecosystem of other info? What is the context?
- **Private vs. Public** - Will the information be experienced collectively or individually?
- **Senses vs. Brain** - Does it engage the eyes? The ears? Does it effectively provoke thought?
- **Appeal** - Is it beautiful? Pleasant to experience?

- **Digital vs. Print** - what tech do you need?
- **Paywall** - do you need to pay to access?
- **Ownership** - who 'owns' the information?
- **Training** - do you need a specific education to understand it?
- **Permanence** - Will it always be available?
- **Depth** - Can you see the methodology? Raw data? Collection tools? Data Biography?
- **Copyright** - What can you use the information for?
- **Ableism** - Providing alternatives for differing ability?
- **Feedback** - How can the audience respond to the content and the creators?

How to Ask about Using a Different Methodology

Questions to ask yourselves:

Did we pick the methodology or was it assigned to us?

Did we pick the research questions first, or the methodology first?

Do we feel in control over how the methodology is being used? Are we being steamrolled?

Does this methodology allow for the equity we want to focus on in this project?

Questions to ask your methodology expert:

What other methodologies should we consider?

Can that methodology answer these kinds of research questions?

What do you like about this methodology?

This Methodology Matrix* seems to suggest that there are several methodologies that might answer our questions, what do you think?

What methodologies are you most comfortable with? Are there any you're excited about?

* weallcount.com/methodology-matrix

Icebreakers for your methodology requirers:

"I think we've got an opportunity to use a really exciting methodology!"

"This other methodology looks equally robust and more applicable!"

"We agree that that methodology is proven, so is this methodology!"

"We think this methodology will do a better job of answering our actual research question!"

"We think it would be great to supplement that methodology with this methodology!"

WE ALL COUNT
project for equity in data science

How to Suggest that you Should Test your Data Viz

Ideas to (gently) bring up with your designer:

1. The data viz "best practices" that we've relied on in the past are not universal; they were mostly based on the comprehension and preference of western college students.

2. It's more equitable for us to supply data visualizations that are effective for our audience than to expect our audience to train themselves in our preferences.

3. The symbolism, structure, and language of data visualizations are not culturally universal. They can appear confusing or even oppressive to some audiences.

4. We tested every other part of the project, why stop now? We can easily test how effective our design is for the audiences we care about. Let's put our data viz in front of some of them and have them identify what each of the elements (colors, shapes, scale, axis, symbols, icons, orientation, patterns, units, etc.) means to them (or doesn't mean!). We All Count calls this the **Reverse-Engineered Legend.** *Note that you're not testing them on how well they understand the information, you're testing how effective your design is for them.

5. BONUS: We'll get to think outside the box and created more interesting, unique, impactful and effective designs, with the testing to back up their efficacy *in this specific situation.*

P.S. If we're stuck trying to improve our design for a specific audience, what kind of designs are they already using? What structures of visual storytelling are they already conditioned to understand and respond to? How do they see things?

Further reading:

https://weallcount.com/2020/07/30/reverse-engineering-data-viz-for-equity/

https://medium.com/@kennelliott/39-studies-about-human-perception-in-30-minutes-4728f9e31a73

https://aea365.org/blog/gedi-week-cultural-considerations-in-data-visualization-a-ge-di-take-by-mike-osiemo-mwirigi-and-glen-acheampong/

WE ALL COUNT
project for equity in data science

How to bring up that RCTs aren't the "Gold Standard" of all data.

*even for impact evaluations!

"RCTs are not the 'gold standard of evaluation'. They are the absolute best at just *one thing* and terrible at almost everything else."

- Heather Krause

5 REASONS TO CONSIDER USING A PROJECT DESIGN OTHER THAN AN RCT FOR YOUR RESEARCH:

You can get much better information with a variety of methods.

The return on investment is often better with other methods.

Your work will have greater emphasis on equity and ethics.

Your methods will likely be better for the people you're trying to support.

Funders are growing interested in innovative designs that are more efficient and effective.

Real solutions for your research questions are out there as new best practices emerge.

Unless you specifically want to know the average treatment effect for the entire population you can get better information for less money.

WHAT RCTS CANNOT TYPICALLY TELL YOU:

Will this scale?

Who does this work for? (It's possible for the treatment to be very good for the population but very bad for lots of individual people within the population)

Why does this work?

How long does the effect last?

Where does this work?

WHAT COULD YOU USE INSTEAD OF AN RCT?

"What effect is our project having?" (network models or matching)

"Why is our project causing impact?" (hierarchical bayesian analysis or acyclic graphs)

"How and for whom is our project working best?" (decomposition or structural models)

Try the Methodology Matrix and explore alternative designs!

https://weallcount.com/methodology-matrix/

WE ALL COUNT
project for equity in data science