



The Stories behind Missing Data: Policy and Practice in Social Services

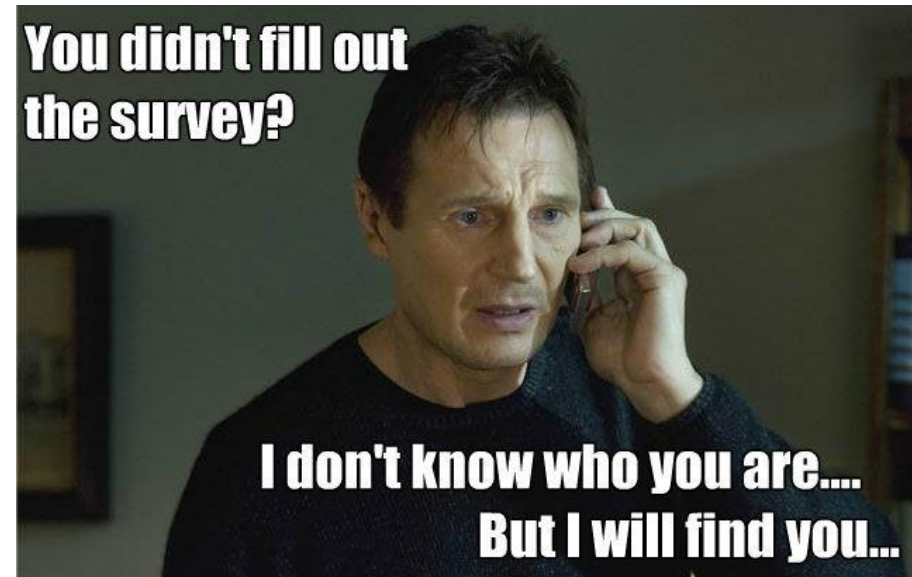
Rika Gorn

Who am I?

- @Rika Gorn, R-stats N00b and enthusiast! 😊
- 10 years working in social service organizations in immigration, mental health, homelessness and housing sectors
- Will discuss a few stories behind missing data in social services programs
 - **Why** do they exist?
 - **What** do they tell us about program design and outcomes?
 - **How** do we use policy, operational, and R solutions to improve services for vulnerable and underserved communities?

Reasons for Missing Data

- Not engaged audience
- Poor data collection tools
- Lack of organizational/technical infrastructure
- Poor data quality
- Challenges in data sharing/collaborating
- Self-censorship or self-preservation
- Lack of trust
- Dependent on other variables
 - MAR vs. MCAR vs. MNAR



Example 1: Where did the time go?

- Program serving adults with serious mental illness transitioning from the State Psychiatric hospital back to the community.
- Goal to reduce hospital readmissions, shorten length of stays, and reduce emergency department use.
- Clients show up in multiple hospitals at the same time and day!!

Policy	Organizational/Infrastructure	Analytical
<ul style="list-style-type: none">• Never underestimate the breadth of the network and lack of communication between players• Does the program model need a change? (e.g. Housing First)• What are the implications of security and privacy on the program? (e.g. HIPAA)	<ul style="list-style-type: none">• Train staff on data entry in EHR• Determine if staff have the right tools<ul style="list-style-type: none">• Case management vs. data entry in EHR• Diagnose staff turnover, motivation, and burnout <div data-bbox="886 1229 2384 1365" style="border: 1px solid black; padding: 5px; color: red;"><p>Result: Were able to determine that most hospital re-admissions occurred within the first 30 days of initial community entry</p></div>	<ul style="list-style-type: none">• Provide clarity on “definitions” for variables• Use R package for imputing values for univariate time series: ImputeTS, Amelia

Example 2: Sharing Dangerous Information

- Program working to house and protect vulnerable youth across 3 countries
- Enormous amounts of missing data on where youth go after program

Policy	Organizational/Infrastructure	Analytical
<ul style="list-style-type: none">• Determine if different outcomes are desired by network players or organizations• Understand relationships between civic actors and government agencies	<ul style="list-style-type: none">• Minimize mandatory collection of data that can jeopardize the safety of your clients• Provide safety and privacy precautions across multiple time junctions during your program• Understand your audience<ul style="list-style-type: none">• Board members vs. Program staff	<ul style="list-style-type: none">• Create new variables (e.g. “Refused to Answer”, Don’t Know vs. Missing)• <code>Forcats::fct_collapse::fct_lump</code>• Visualize missing data using <code>VIM</code>, <code>visdat</code>, <code>ggplot2</code>, <code>naniar</code>

Result: Outcome metrics revised to account for cultural and sector context

Thanks R-Ladies NYC!



@RikaGorn



Rikagx@gmail.com



@RikaGorn

(mostly pictures of my dog Zoe!)