Bit by Bit:
Social Research in the Digital Age
Ch. 1 Introduction

Presented by Zachary Stine
February 1, 2019

Overview

- Author intro and SICSS

- 1.1 An ink blot

- 1.2 Welcome to the digital age

- 1.3 Research design

- 1.4 Themes of this book

- 1.5 Outline of this book

- Final notes

Author introduction: Matthew J Salganik

- Mathematics BA; Sociology MA and PhD

- Professor of Sociology at Princeton

- Co-facilitates the Summer Institute in Computational Social Science (SICSS)

- SICSS applications for 2019 due by **February 20**; apply here

- Teaching materials from SICSS 2018 are available here

- For more author information see https://www.princeton.edu/~mjs3/
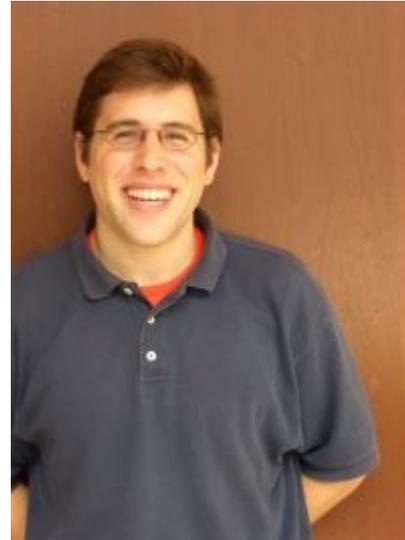
1.1 An ink blot

Blumenstock, Cadamuro, & On (2015)

- Study surveyed ~1k Rwandans by randomly sampling cell phone #s from provider database [link]

- Participants completed a survey (traditional social science)

- Researchers also utilized call records for 1.5 million customers in the provider database (not traditional social science)

- Trained model to predict someone's wealth from their call records based on interview data

- Used model to predict wealth for entire set of 1.5 million customers

- Produced high-resolution maps of wealth distribution across the country; accurate at smallest units of the country

- Results were similar to the Demographic and Health Surveys (gold standard) but 10x faster and 50x less expensive

A Rorschach inkblot test

- **Social scientists** may see a useful new measurement tool

- **Data scientists** may see a cool machine learning problem

- **Business people** may see a way to extract value from data they already have access to

- **Privacy advocates** may see scary potential for mass surveillance

- **Policy makers** may see a way for technology to help improve the world

1.2 Welcome to the digital age
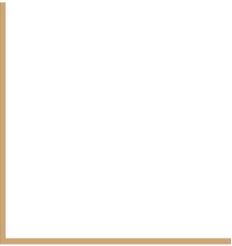
The digital age

- New opportunities for social research that were previously impossible

- New risks for social research to unintentionally cause harm

- Key feature of the digital age: computers everywhere

- Researchers are making a change analogous to going from photography to cinematography

Past and future

- Existing principles of social research from the past 100 years are still relevant and should continue to inform new research

- Challenge: to combine approaches of the past with present and future capabilities

- Example study relied on both traditional social science and data science

  - Model could not be trained without input from the surveys

  - Needed both surveys and phone records

- Social science and data science approaches should be combined

1.3 Research design

"Research design is about connecting questions and answers."

Two audiences

- **1) Social scientists** who have experience studying social behavior but are not familiar with opportunities of the digital age

- **2) "Data scientists"** who have backgrounds in comp. sci., statistics, info. sci., engineering, or physics

- Both audiences can learn from each other

- Goal of this book is to bring both communities together to make something greater than either could alone

Research design

- Best starting place is not abstract social theory or technical machine learning, but with **research design**

- Research = process of asking questions and answering them

- *Social* research = process of asking questions *about human behavior* and answering them

- Given a question, research design links that question to an answer

- Convincing research requires getting research design right

- Four approaches to research design: observing behavior, asking questions, running experiments, & collaborating with others

1.4 Themes of this book
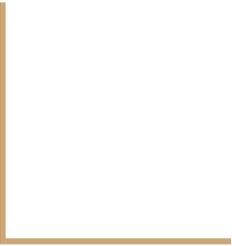
An art analogy

- Marcel Duchamp (1887 - 1968) created **readymades** by repurposing objects that already existed (e.g., *Fountain*)

- Michelangelo (1475 - 1564) created **custommades** by starting with an idea and then making it from scratch (e.g., *David*)

Readymade

Custommade

Photo source: https://commons.wikimedia.org/wiki/File:Duchamp_Fountaine.jpg; https://commons.wikimedia.org/wiki/File:%27David%27_by_Michelangelo_JBU0001.JPG

Mixing readymades and custommades

- Readymades in research - repurposing big data sources that have already been created (e.g., call records)

- Custommades in research - start with question and then create the data needed to answer it (e.g., survey data)

- Both styles are great; potentially even greater when combined

Second theme: ethics

- Researchers in the digital age will inevitably confront difficult ethical decisions

- Chapter 6 is wholly devoted to ethics, but will be considered in research design chapters as well

- Recall Blumenstock et al. (2015) used call records

- Mayer, Mutchler, & Mitchell (2016) showed that even anonymized call records could be used to identify specific people and sensitive information about them when combined with public information [link]

Balancing risks and opportunities

- Researchers in the digital age have an increasing amount of power over subjects & participants (often in collaboration with companies or governments)

- Power = ability to do things to people without their awareness or consent

- Example: can enroll millions of people in online experiments without their knowledge (probably happens every time you use Amazon)

- Power of researchers has increased, but no equivalent increase in guidelines about how to responsibly use that power

- Rules, laws, and norms may be inconsistent and overlapping

1.5 Outline of this book
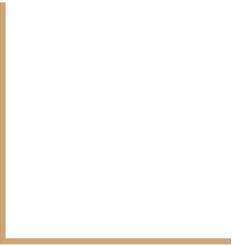
Four approaches to research design

- Chapters devoted to each of the broad approaches to research design: observing behavior, asking questions, running experiments, and creating mass collaboration

- Each requires different relationship between researchers and participants

- Each enables us to learn different things
  - Asking people questions might tell us something we could not learn from observation

  - Running experiments might tell us something we could not learn from observation/questions

  - Collaborating with participants mights tell us something we could not learn from observation/questions/experiments

Chapter 2: Observing behavior

- What and how researchers can learn from observation

- Focus on big data sources

- Ten common features of big data sources and how each affects researchers' ability to use those sources

- Three research strategies for big data sources

Chapter 3: Asking questions

- Explore what can be learned by asking questions that cannot be learned from observation alone

- Review of the traditional total survey error framework

- Show how the digital age enables new approaches to sampling and interviewing

- Two strategies for combining survey data and big data sources

Chapter 4: Running experiments

- Randomized controlled experiments and learning about causal relationships

- Comparison of experiments of the past with new opportunities in the digital age

- Description of main strategies for conducting digital experiments and their trade-offs

- Design advice for taking advantage of the power of digital experiments

- Consideration of responsibilities that come with this power

Chapter 5: Creating mass collaboration

- Explore how researchers can create mass collaborations for social research

- E.g. crowdsourcing and citizen science

- Review examples of successful mass collaboration projects

- Key organizing principles for successful mass collaborations

- Argues two claims
    - 1) mass collaboration can be harnessed for social research
    - 2) researchers using mass collaboration will be able to solve seemingly impossible problems

Chapter 6: Ethics

- Capabilities are changing faster than ethical guidelines

- Argues for a *principles-based* approach to ethical decisions

- Four established principles

- Two ethical frameworks

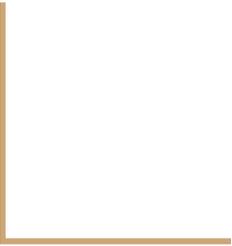- Practical tips for working in areas with unsettled ethics

Chapter 7: The future

- Review of the book's main themes

- Speculation about themes that will remain important in the future
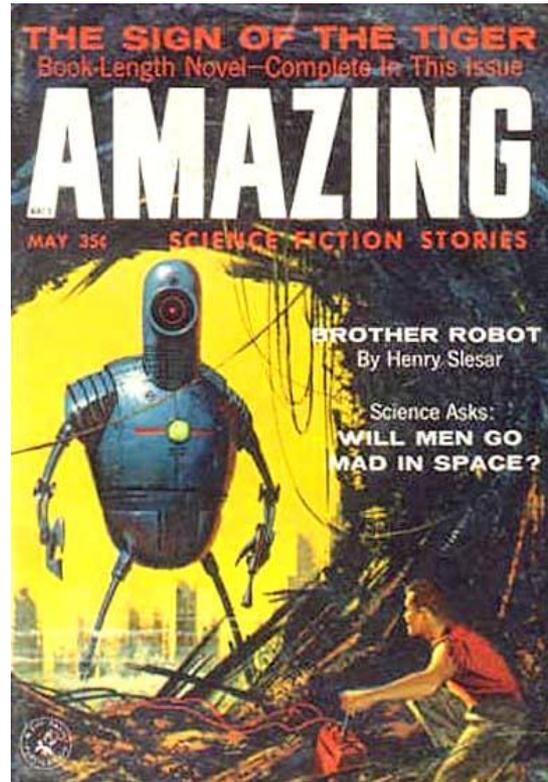
Final notes

How to use the book

- Chapters are written informally and use many examples to illustrate concepts

- "Mathematical notes" sections provide more formal and technical descriptions of concepts in the chapter

- "What to read next" sections provide literature reviews for more in-depth exploration

- "Activities" sections include opportunities for experiential learning with difficulty ratings for each activity

Helpful links

- 1st edition of the book available to read online at https://www.bitbybitbook.com/en/1st-ed/preface/

- Syllabus: http://www.princeton.edu/~mjs3/soc596_f2016/

- Slides: https://github.com/msalganik/soc596_f2016/tree/master/slides

- Videos: https://www.youtube.com/channel/UCkRV9I1xz2KwlgvLQ8OadKw

Thoughts?