Class 1: Introductions to the Course, Surveys and RMarkdown

MAST5953: Creating Your Own Data

Dr. Miriam Sorace

www.miriamsorace.github.io

30 October 2023

Outline of Today's Class

Course Introduction

Surveys: An Introduction

R & RMarkdown

Course Introduction

MAST5953 Structure

Survey Design - Autumn Term

Mon - 4-6pm (except Nov 20 when we'll start at 5!!), room: KS15

1. **Week 13 - 30 October 2023**: Class 1: Introduction to the Course, Resources and to RMarkdown
2. **Week 14 - 6 November 2023:** Class 2: Principles in Survey Question Design
3. **Week 15 - 13 November 2023:** Class 3: Sampling & Data Collection Methods
4. **Week 16 - 20 November 2023:** Class 4: Survey Error & Bias Correction
5. **Week 17 - 27 November 2023:** Class 5: Ethics, Recap & Assessment Q&A ****start your survey!****
6. **Week 18 - 4 December 2023:** Class 6: Working Together on Assignment 1
7. **Week 19 - 11 December 2023:** Class 7: Data Science – Data Collection Strategies & Primer on Web Technologies

Web-Scraping & Text Mining - Spring Term

Time Slot and Location TBC - check your timetable in December/January

1. **Week 25 - 15-19 January 2024**: Class 8 Web Scraping & Regular Expressions
2. **Week 25 - 15-19 January 2024**: Class 9 Preparing for Assignment 2: Web Scraping Speeches
3. **Week 26 - 22-26 January 2024:** Class 10 Text Mining I - Text Pre-Processing
4. **Week 26 - 22-26 January 2024**: Class 11 Text Mining II - Topic Models

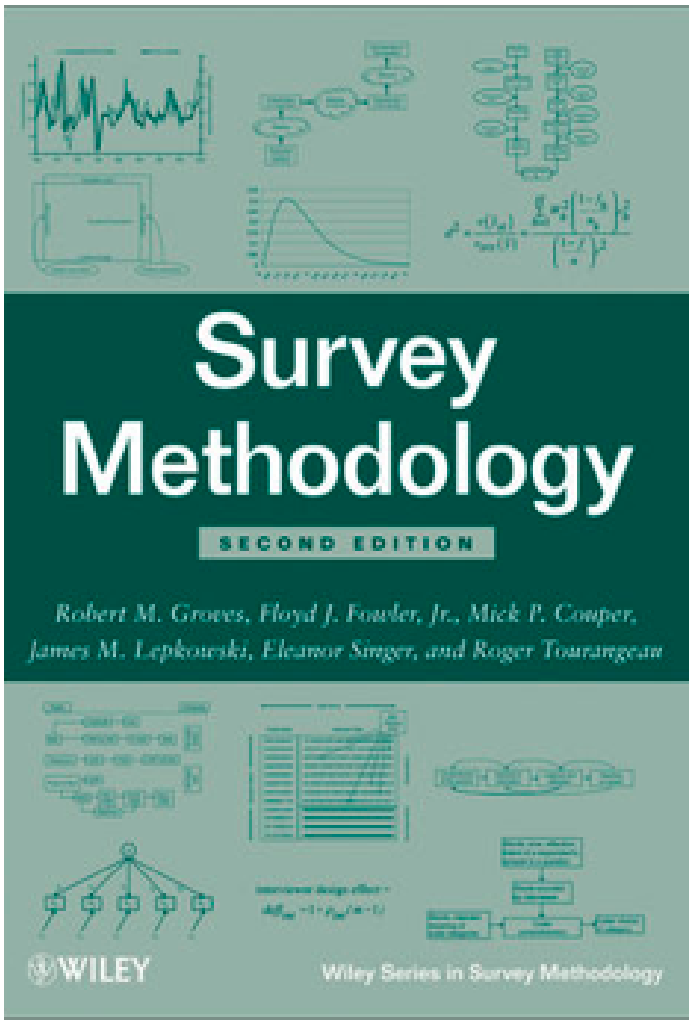
MAST5953 Structure

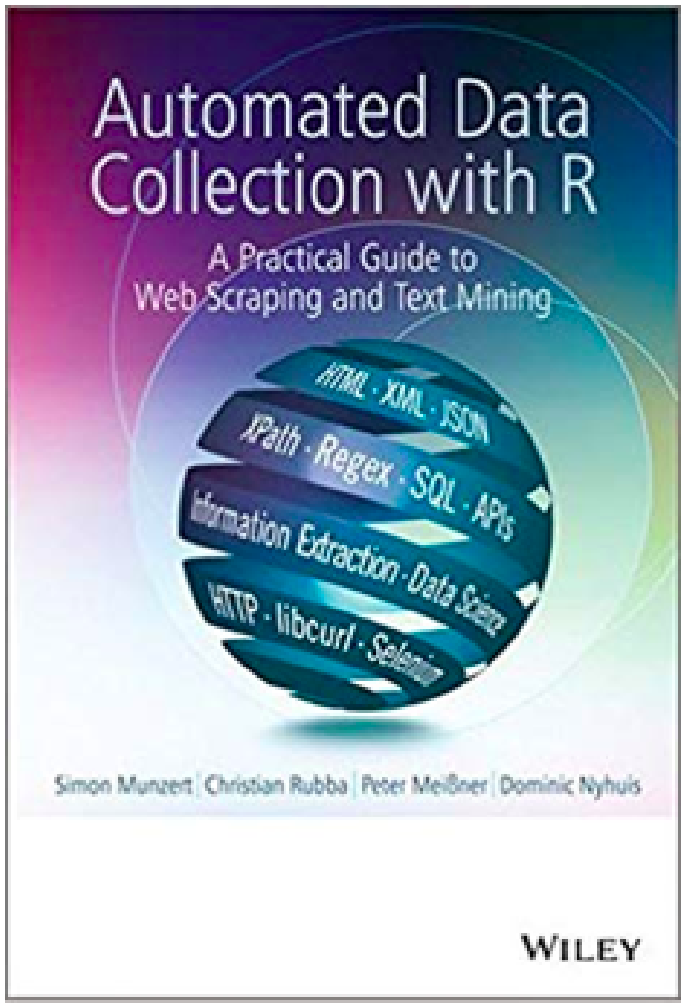
1. Support/Contacts
   - For all your questions use the Discussion Forums
     - But also: exchange contact info with each other and study together! There is no better way to learn than in teams!
   - For extenuating circumstances/mitigation/extensions contact CEMS Student Support: cemssupport@kent.ac.uk
   - Office Hours straight after class, no need to book just approach me after class ;)
2. Course Details & Materials
   - Course Moodle Page

MAST5953 Learning Outcomes

1. Design survey/text analysis protocols/codebooks
2. Collect web and survey data independently
   - Building Web Scrapers via R
   - Generating, administering and analysing survey data
3. Analyse text data via machine learning

MAST5953 Core Textbooks





MAST5953 Assessment & Feedback - Report 1

Details on Assessment tab on Moodle

- Survey design task - 50%, due: Mon, Jan 22 2024
- 1,000 words (excluding Appendix, Tables/Figures/References, R code)
- What you need to do:
  - Cognitive Interview (Qual) with 2,3 people
  - Questionnaire survey with 10,15 people
  - Qual. analysis of cogn. interview responses (using response process model)
  - Graphs/correlation analysis of questionnaire survey responses
- Start thinking about a topic
- Remember: a measurement exercise, not an answer to a research question!

Surveys: An Introduction

Discussion

Have you ever been interviewed? What was the interview like?

Survey Research

- "Obtaining information through asking questions" (Wolf et al. 2016: 4)
- "[...] a crucial building block in a modern information-based society" (Groves et al 2009: 3)

Survey Research

- Its importance has not been displaced in an era of big data
- Survey research is fundamental to describe a population accurately

Survey Research: Some Terminology ...

- **Survey**: Information-gathering method that works via administration of a list of questions
- **Interview**: Generally synonymous with survey but the interview implies a person-to-person 'conversation'
- **Poll**: private-sector surveys, mostly in relation to vote-intention/prediction

Types of Surveys/Interviewing

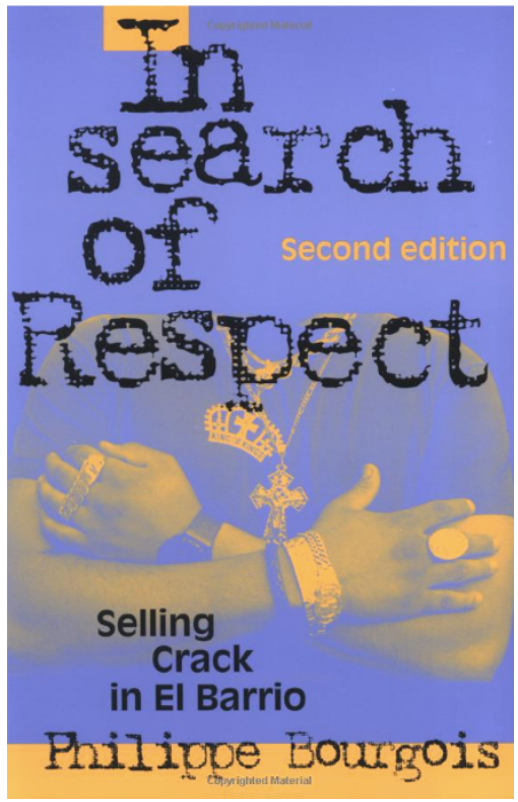
Structured vs. Semi-Structured vs. Unstructured

Q10 We have a number of parties in <country> each of which would like to get your vote. How probable is it that you will ever vote for the following parties? Please answer on a scale where 0 means "not at all probable" and 10 means "very probable".
[PARTY LIST C, PTV list up to 10 parties]
<Source: EES2014 QPP12>

	0 not at all probable	1	2	3	4	5	6	7	8	9	10 very proba-ble	Don't know the party
Party 1	0	1	2	3	4	5	6	7	8	9	10	98
Party 2	0	1	2	3	4	5	6	7	8	9	10	98
Party 3	0	1	2	3	4	5	6	7	8	9	10	98
Party 4	0	1	2	3	4	5	6	7	8	9	10	98
Party 5	0	1	2	3	4	5	6	7	8	9	10	98
Party 6	0	1	2	3	4	5	6	7	8	9	10	98
Party 7	0	1	2	3	4	5	6	7	8	9	10	98
Party 8	0	1	2	3	4	5	6	7	8	9	10	98
Party 9	0	1	2	3	4	5	6	7	8	9	10	98
Party 10	0	1	2	3	4	5	6	7	8	9	10	98

Q11 In political matters people talk of "the left" and "the right". What is your position? Please indicate your views using any number on an 11-point-scale. On this scale, where 0 means "left" and 10 means "right," which number best describes your position?
<Source: EES 2014 QPP13>

	0 Left	1	2	3	4	5	6	7	8	9	10 Right	DK
	0	1	2	3	4	5	6	7	8	9	10	98



Quantitative Interviewing

Strengths and Weaknesses

Strengths:

- Less expensive quick results
- Strong assurance of anonymity (especially self-administered ones)
- Lower interviewer-induced bias
- Consistent measures: higher comparability
- High reliability and replicability

Weaknesses:

- Clarifications/re-formulation difficult
- Partial responses more likely (especially in self-administered ones)
- Assumes complete knowledge on the phenomenon

Qualitative Interviewing

Strengths and Weaknesses

Strengths:

- Allow theory and conceptual exploration / conceptual refinement
- Allow to gain missing knowledge, respondent is the expert
- Allow direct evaluation of the role of socio-cultural context on attitudes/behaviours

Weaknesses:

- Interviewer effects are expected to be very strong
- Can lead to digressions from research aims
- Cannot be used for causal inference / hypothesis testing
- Sampling usually by convenience, cannot be used for hypothesis testing
- Reliability & replicability difficult

Which Type Should I Use?

1. What is the target population?
   - Elite/experts: qualitative
   - Public/large group: quantitative

2. What is the nature of the phenomenon/concept of interest?
   - Unknown: qualitative
   - Established/measurable: quantitative

Types of Survey Mode

Mode Effects

- Face-to-Face (PI: personal interviewing)
  - **Pros:** they boost response rates and allow for meaningful clarification
  - **Cons:** require costly interviewer training to ensure consistency and similar levels of experience + not appropriate when sensitive questions are asked + interviewer demographic characteristics might impact responses (say on race or gender).
- Telephone (CATI)
- Mail or Web-based (Self-Administered)
  - **Pros:** they guarantee privacy + reduce costs and increase response timeliness
  - **Cons:** suffer from higher non-response and item non-response rates + web-based surveys face the additional problem of building adequate sampling frames

Applied Example
The European Election Study

- Check out the questionnaire

R & RMarkdown

Course Requirements: R

Particularly for the Spring Term Sessions!

- Prior knowledge of `R` is a must, make sure you understand the language basics (packages, objects/vectors, core functions and vector + data management operations), and that you know how to trouble-shoot errors and install packages
  - Great `R` Intro Resources:
    - Adler, Joseph. 2009. *R in a Nutshell. A Desktop Quick Reference*. O'Reilly
    - Teetor, Paul. 2011. *R Cookbook*. O'Reilly.
    - I recommend the following website for revisions: `https://stats.idre.ucla.edu/r/`

Course Requirements: R

- Make sure you have installed `R` and `RStudio` and everything is up-to-date.
  - `R`:
    - Follow the instructions to download R here: R Installer
- `RStudio`:
  - Updates needed: Go to «Help» - «Check for Updates» - & follow instructions in the pop-up
  - To download RStudio: RStudio Installer

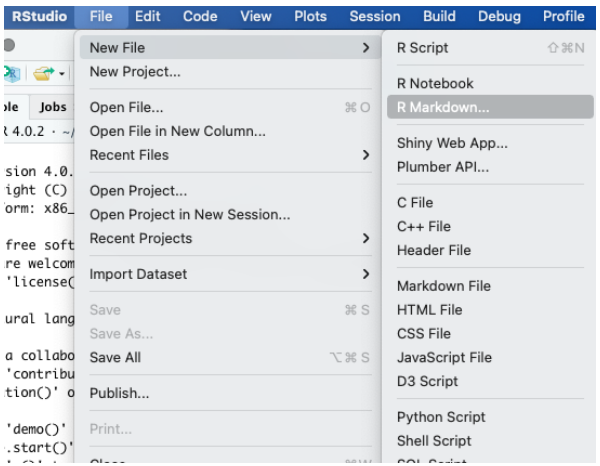
R debugging

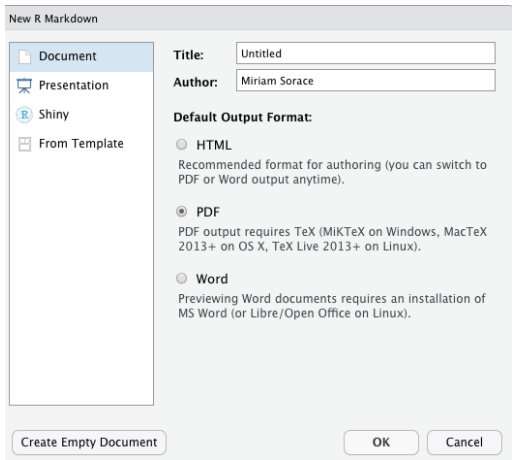
Examples of some common errors

- If it says: there is no package called 'xxx' you just need to type:
  - install.packages("xxx")
- If it says: could not find function 'xxx', just call the library by typing:
  - `require(xxx)` or `library(xxx)`
- If it says: cannot open the connection, it means you have not specified the correct working directory (i.e. folder) where the file is located
  - `setwd()`
- Great Resources to Trouble-Shoot Errors:
  - ChatGPT (or Phind)!!!
  - StackOverflow/Stack Exchange
  - **Googling/using AI to build your code is not cheating!!** :)

Creating an RMarkdown File

- From scratch:





- Opening an existing file/template: «File» - «Open File» - Navigate to the folder where you saved the .Rmd file & select the file.

RMarkdown Exercise

- Practical Exercise with RMarkdown Template.Rmd file in Class 1 Section on Moodle (see Course Moodle Page - Class 1)
- Also: Check Out the RMarkdown Cheat Sheet

For Next Time

- Is there some phenomenon on which you'd like to measure public opinion/attitudes (support/behaviour frequency/importance)? Which is it?
- Think of a survey question that you might want to ask ...

What did we learn today?

- Course structure and website
- Different types of surveys
  - Structured
  - Semi-structured
  - Unstructured
- How to write a report using RMarkdown