

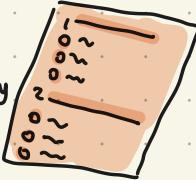
Week 1: Data Exploration

protecting data

Ethical data analysis

Data can be collected by

1. Interviews
2. Observations → mostly used by scientists
3. forms
4. Questionnaires
5. Surveys
6. Cookies → track people online's activities and interests



First-party data

↳ data collected by an individual or group using their own resources.

Second-party data

↳ data collected by a group directly from its audience and then sold.



Third-party data (not reliable)

↳ outsources not directly collected

Data Collection Considerations

③ How the data will be collected?

④ choose data sources

④ Decide what data to use

④ How much data to collect?

① Select the right data type

② Determine the time frame

Population → entire sample → some groups

Data Formats

Quantitative data → e.g. numbers, range, price

Qualitative data → cannot be counted, measured
e.g. text, description, heading

Need answer immediately → Historical Data

Discrete → counted & limited number

e.g. price, rating, stats (no decimal accepted version)
have decimal limited

Continuous → unspecified number of possible points

Nominal → categorized without a set order (no sequence)

e.g. Yes/No

Ordinal → within an order

e.g. 1, 2, 3, 4, 5

Must know!

lives inside a company's own systems

outside company

Internal

External

collected by a researcher from first-hand sources

Primary

Secondary
by other people

* subjective and explanatory
measures of qualities and characteristics
e.g. - Exercise activity most enjoyed

* specific and objective measures of numerical facts
e.g. - Percentage of alcohol in beverages

1001 110010
00110 101011

0101

almost numerical value

Continuous Weight, Height 52.4
- Temperature 25.0
- Runtime in video 18:34

Discrete limited number of values

e.g. - Number who visit a hospital on Dec 2022
- Room's max capacity

- Tickets sold 50, 79
records

Structured in a certain format (rows, columns)

e.g. - Expense reports
fields

- Tax returns

Have "Data Model"
how data is organized and structured

e.g. - EER (Entity Relationship Diagram)
- UML (Unified Modeling Language)

high-level view of the data structure
e.g. how data interacts across an org.

Nominal not categorized with a set order

Ordinal → categorized with a set order / scale

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101

1001
00110
0101

0101

00110
101011

0101