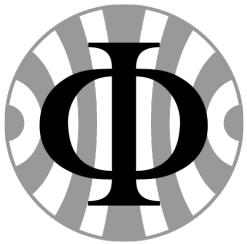
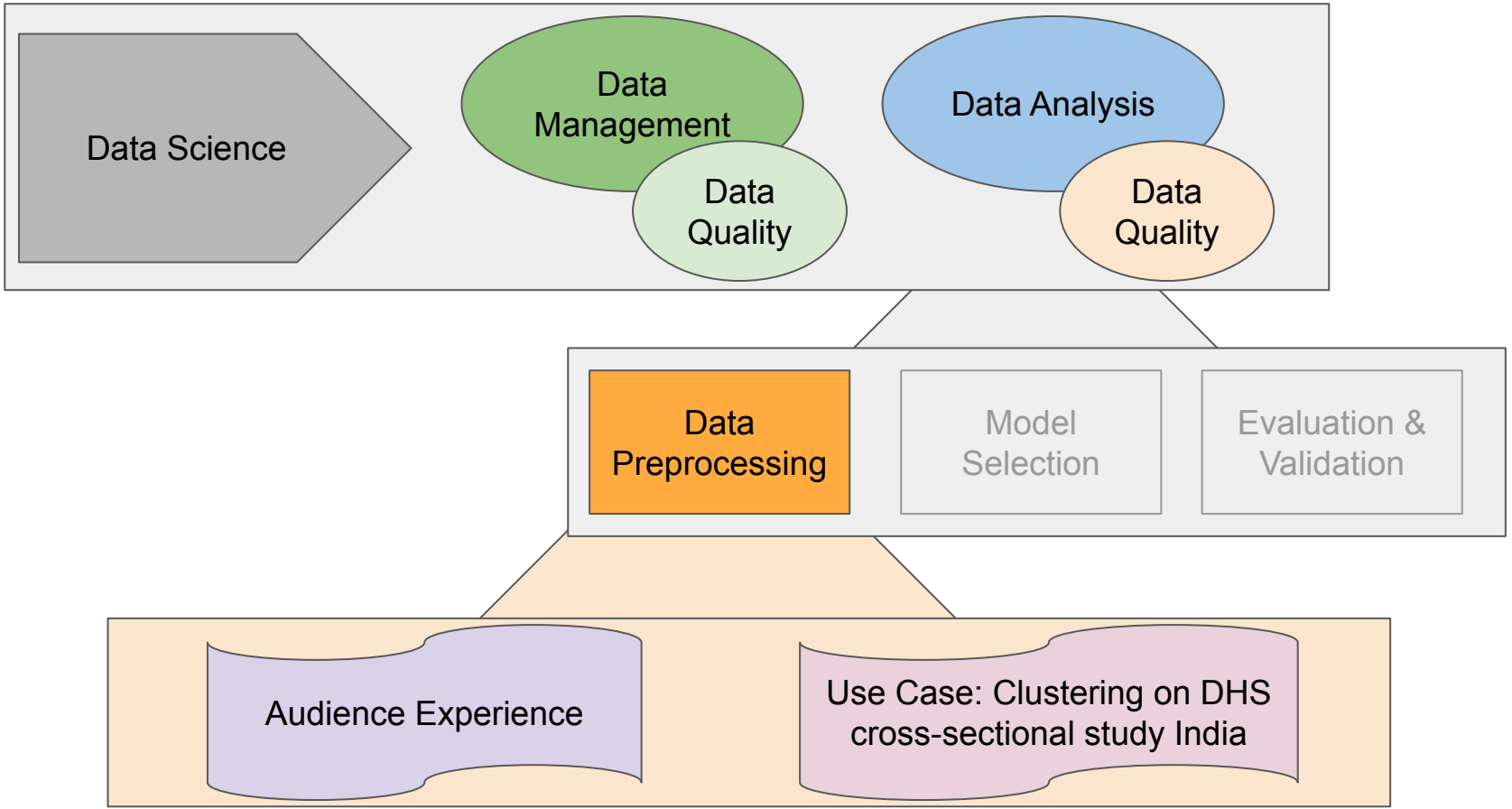
Data Quality and Data Preprocessing



Anna-Katharina Nitschke
& Carlos Brandl, Carola Behr, Fabian Egersdörfer and Prof. Matthias Weidemüller
Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany

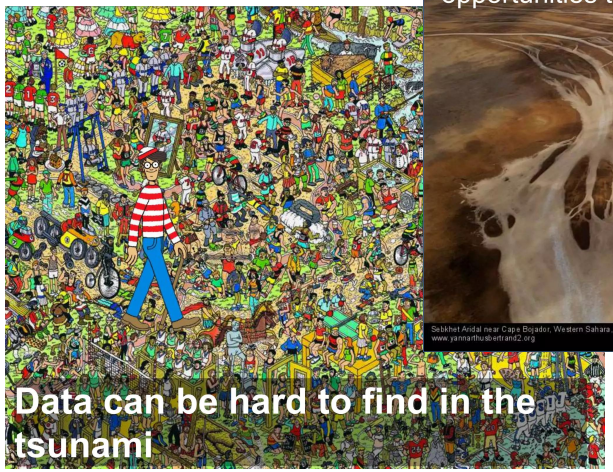
in collaboration with Prof. Till Bärnighausen
Global Health Institut Heidelberg

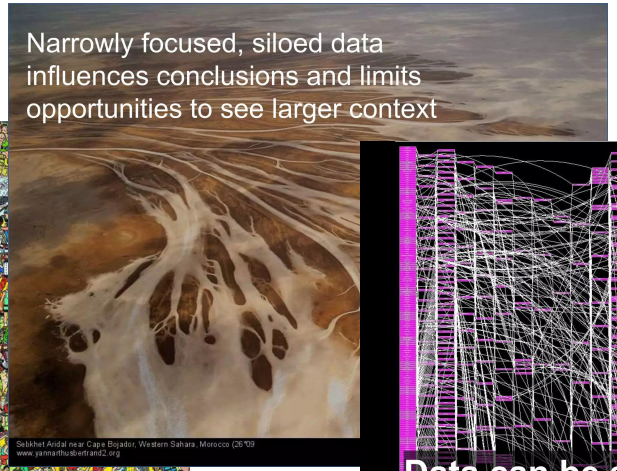


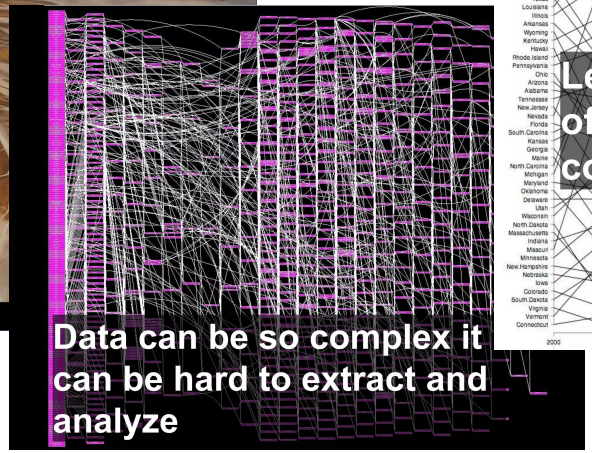
USAID; Measure Evaluation; 2016; Surfing a Great Wave: Data Science and Global Health https://de.slideshare.net/slideshow/surfing-a-great-wave-data-science-and-global-health/57022175

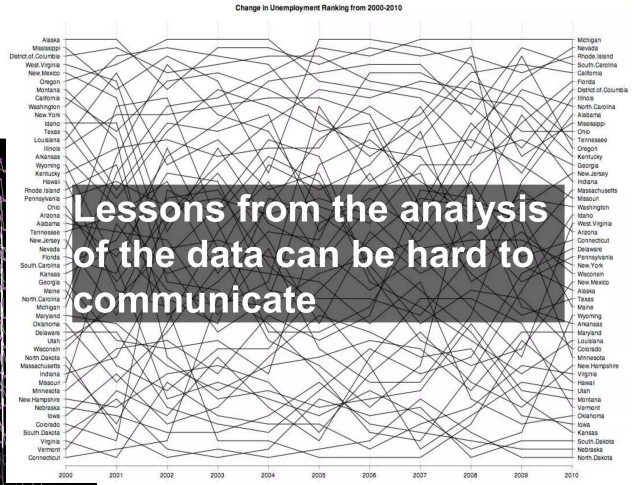
Data Science

There is considerable opportunity.
But also there are challenges.









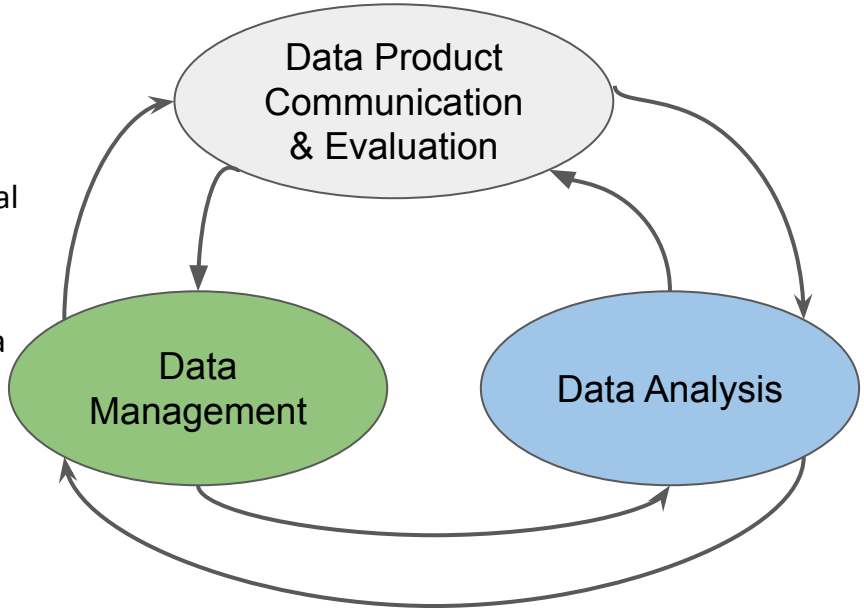
→ Importance of standards and good practice to assure quality of the conducted research

USAID; Measure Evaluation; 2016; Surfing a Great Wave: Data Science and Global Health https://de.slideshare.net/slideshow/surfing-a-great-wave-data-science-and-global-health/57022175

Data Science

Main objectives and tasks within the discipline of data science include:

1. **Locating and retrieving key data that can inform decision making**: Identifying and obtaining relevant data is a fundamental task in data science.
2. **Turning that data into actionable information**: Analyzing and processing data to generate insights that can drive decisions is a core function.
3. **Communicating that information in the most effective way**: Presenting findings clearly and effectively to stakeholders is crucial for data-driven decision making.
4. **Evaluating the impact of data-driven decisions**: Assessing the outcomes of decisions made based on data analysis helps to understand the effectiveness and refine future approaches.

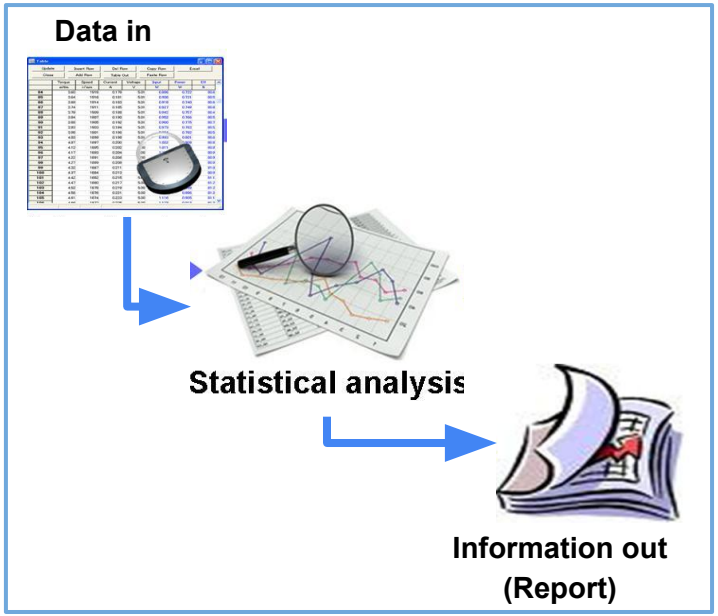


USAID; Measure Evaluation; 2016; Surfing a Great Wave: Data Science and Global Health https://de.slideshare.net/slideshow/surfing-a-great-wave-data-science-and-global-health/57022175

Data Analysis: From Data to Information

Be mindful of the data that you select for your research study!
Data Quality will impact your analysis strongly!

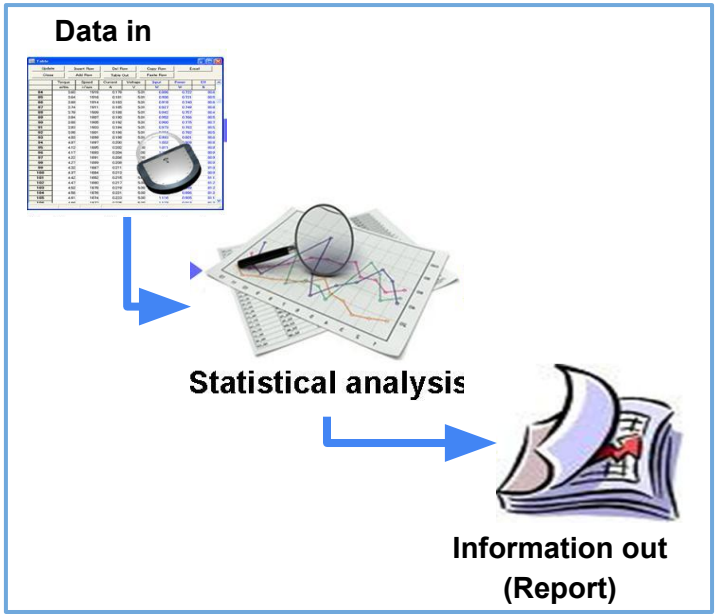
Someone wisely said "if we take care in the beginning, the end will take care of itself".



partly extracted from presentation of Kavita Singh from Public Health Foundation India
Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

Importance of Data Management for Final Study Report

OTHERWISE



The final study report, which is the product of sophisticated computer programs and a statistical analysis, is only as good as the collected data.

→ look at data generation processes within the study design

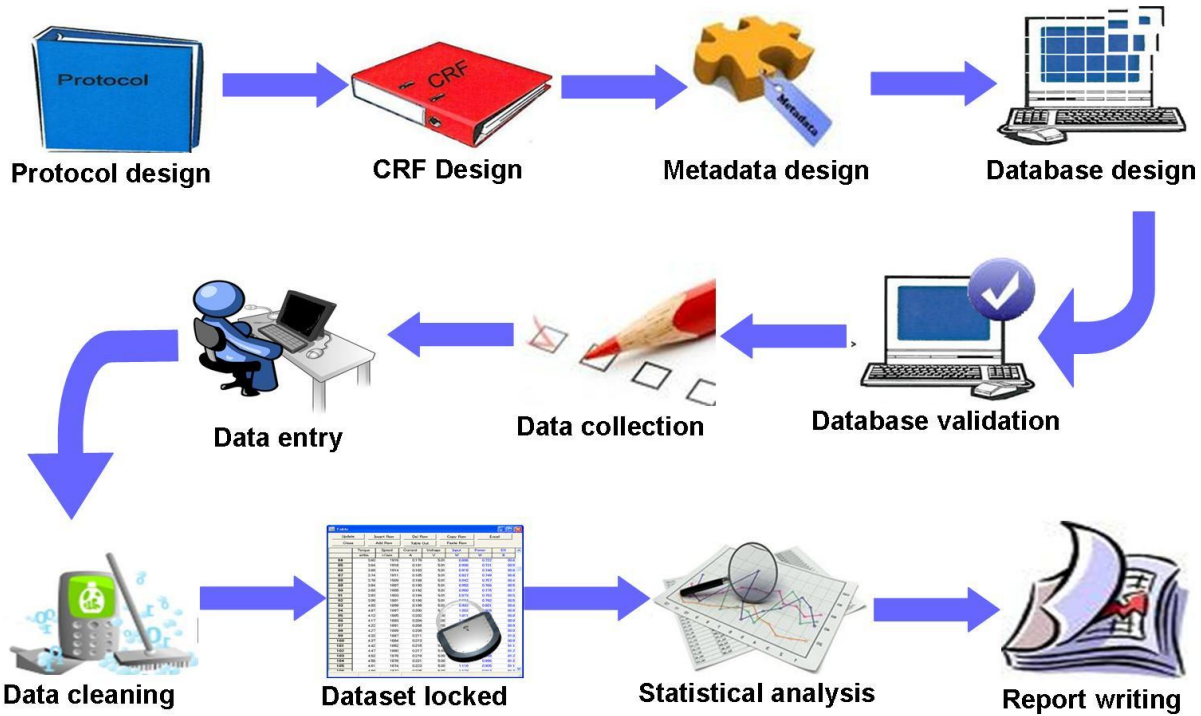
partly extracted from presentation of Kavita Singh from Public Health Foundation India

Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

Data Generation Process - Study Design

Protocol Design for Research Study

"... is a document that describes the background, rationale, objectives, design, enrollment criteria, methodology, data recording requirements, statistical considerations, and organization of a clinical research study."



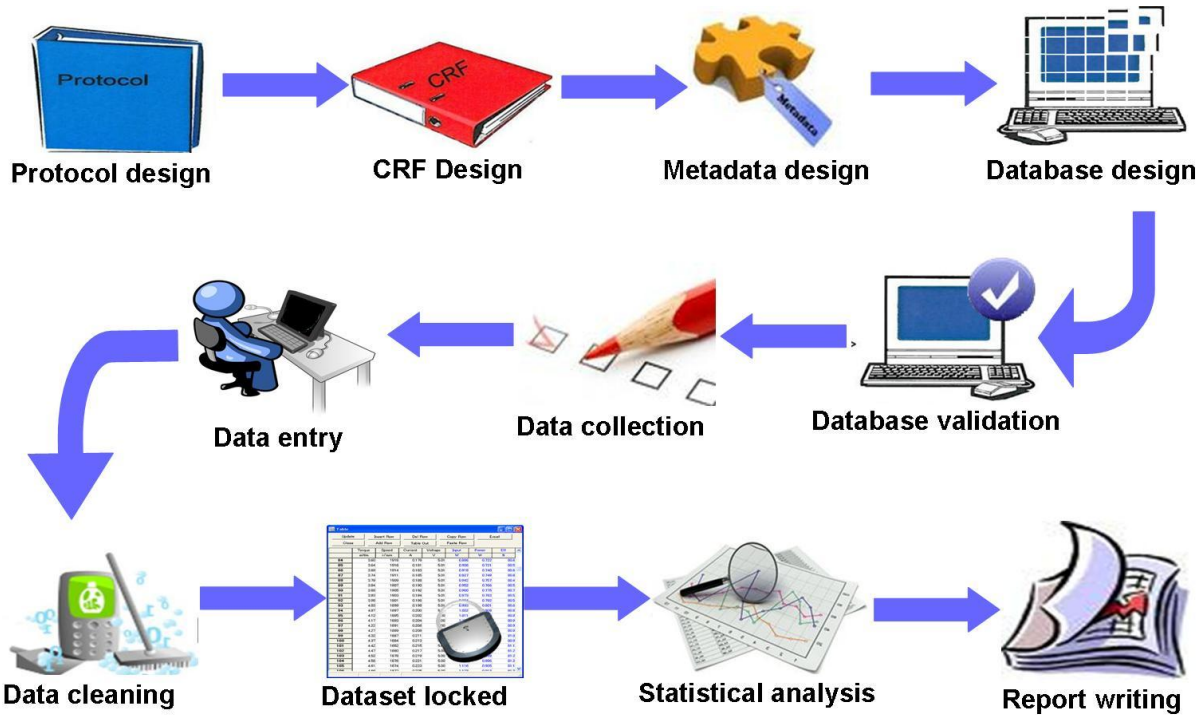
partly extracted from presentation of Kavita Singh from Public Health Foundation India

Data Generation Process - Study Design

Case Report Form (CRF)
"... is a printed or electronic document that is designed to collect required research, administrative, and/or regulatory data for a clinical trial."

→ it is important that the CRFs be designed with clarity and ease of use in mind
→ the design of CRFs has a direct impact on the quality of the data collected for a trial, so it is worthwhile to take time over the design and development of the forms

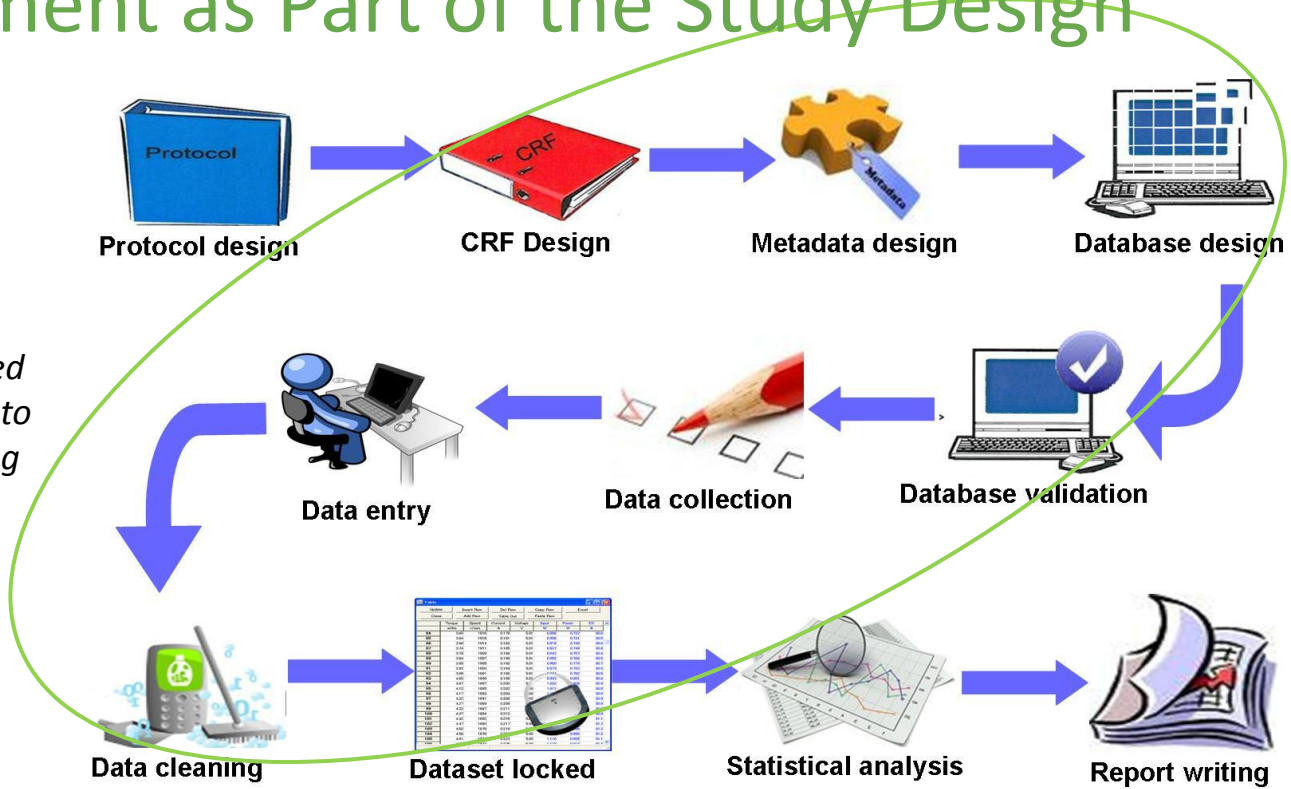
= first step of data management



Data Management as Part of the Study Design

Data Management
"...includes all aspects of data planning, handling, analysis, documentation and storage. The objective is to ensure the validity, *quality and integrity of data collected from subjects to a database system to create a reliable database containing high quality data."*

Data management is a too often neglected part of study design,

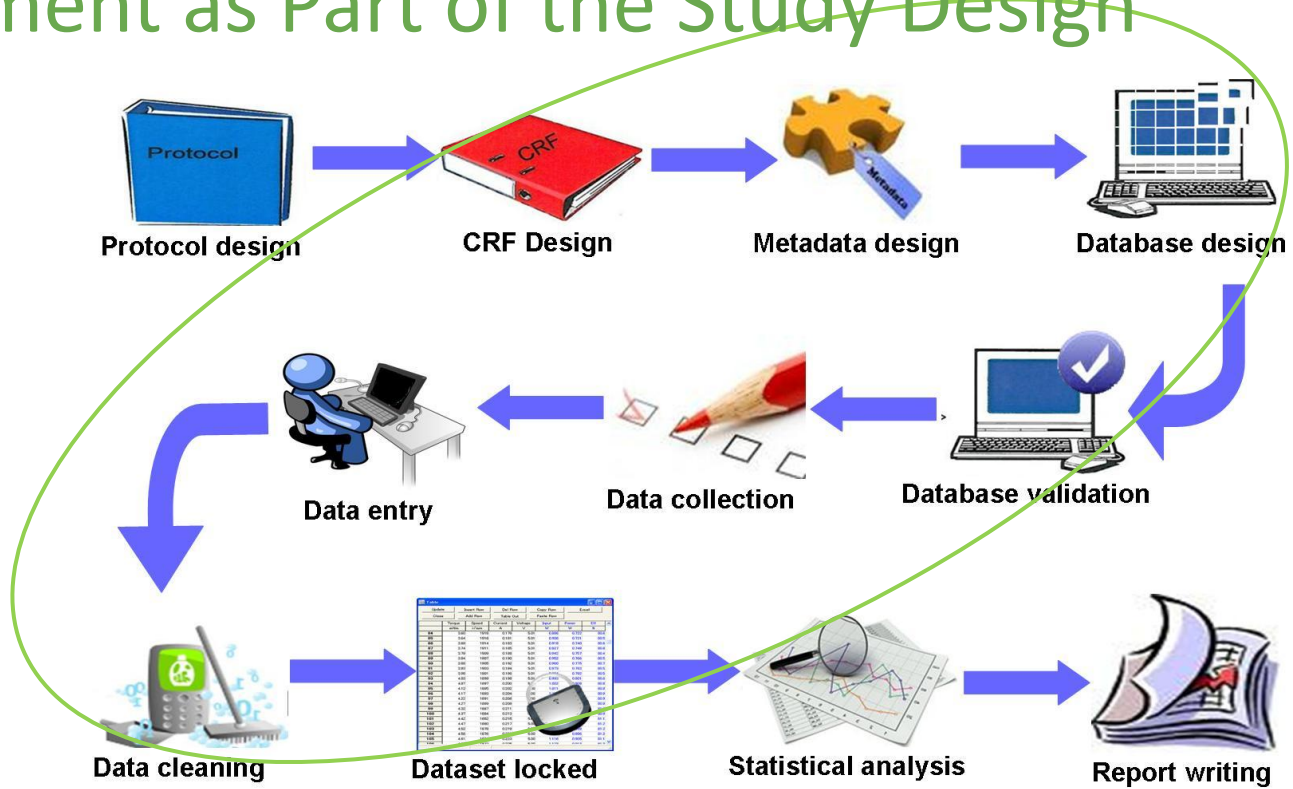


partly extracted from presentation of Kavita Singh from Public Health Foundation India

Data Management as Part of the Study Design

Data Management includes

1. Case report form(development)
2. Database development and validation
3. Data collection
4. Data entry, query and correction
5. data quality assurance
6. data lock, archive and transfer.



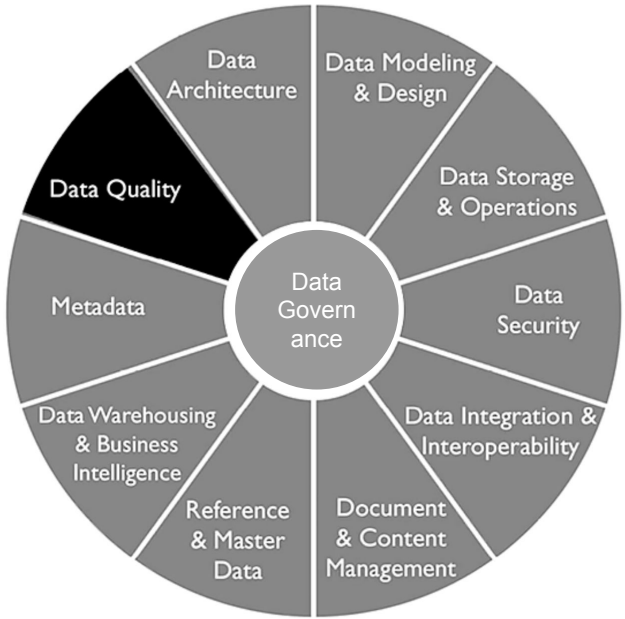
partly extracted from presentation of Kavita Singh from Public Health Foundation India

Data Quality as Part of Data Management

Data Quality Management

= one of the eleven data management categories identified by DAMA

"[consists of] the planning, implementation and control of the activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers."



DAMA-DMBOK2 Data Management Framework

Data management association international (DAMA International) 2017; presentation by LIGHTSONDATA; George Firican https://www.youtube.com/watch?v=4lFTRSBO4ss

Data Quality as Part of Data Management

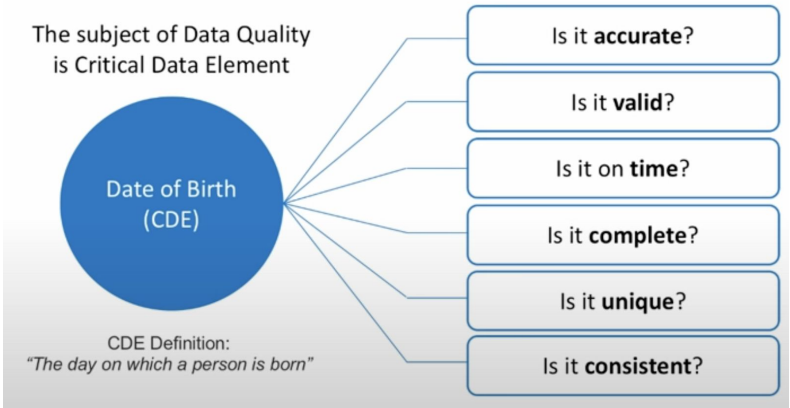
Data Quality Management

"... focused on ensuring the data adheres to our data quality dimension."

→ data fit for purpose

accuracy
validity
timeliness
completeness
uniqueness
consistency

"Data Quality Dimension refers to the aspect or feature of information that can be assessed and used to determine the quality of the data."



Data management association international (DAMA International) 2017; presentation by LIGHTSONDATA; George Firican https://www.youtube.com/watch?v=4lFTRSBO4ss

Data Quality as Part of Data Management

Data Quality Management

1. Accuracy: refers to how closely data in a clinical study reflects the true values or events it is meant to represent. Accurate data means that the recorded information correctly describes the participants' medical conditions, treatments, outcomes, and other relevant details. Inaccurate data can lead to incorrect conclusions, potentially compromising patient safety and the study's validity.

2. Validity: ensures that the data collected, measures what it is intended to measure. For example, a survey question designed to assess a patient's pain level should accurately capture the intensity of pain experienced by the patient. Valid data supports the reliability of study findings and ensures that the study's results are meaningful and applicable.

3. Timeliness: refers to the availability of data when it is needed. In clinical studies, this means that data should be collected, recorded, and made accessible in a timeframe that allows for effective monitoring, analysis, and decision-making. Delayed data can hinder the monitoring of adverse events, slow down the study progress, and impact decision-making processes.

Data management association international (DAMA International) 2017; presentation by LIGHTSONDATA; George Firican https://www.youtube.com/watch?v=4lFTRSBO4ss

Data Quality as Part of Data Management

Data Quality Management

4. Completeness: indicates that all necessary data has been collected for each study participant. This includes ensuring that no fields are left blank and that all required measurements and observations are recorded. Incomplete data can lead to bias, reduce the statistical power of the study, and limit the generalizability of the findings.

5. Uniqueness: ensures that each participant or event is recorded only once in the study dataset. There should be no duplicate records for the same individual or event. Duplicate records can distort statistical analyses and lead to erroneous conclusions.

6. Consistency: refers to the uniformity of the data across the study. This means that data values should not conflict with each other and should align with the study's protocols and expectations. Inconsistent data can cause confusion, reduce trust in the study's findings, and may indicate errors in data collection or entry.

Data management association international (DAMA International) 2017; presentation by LIGHTSONDATA; George Firican https://www.youtube.com/watch?v=4lFTRSBO4ss

Data Quality as Part of Data Management

Data Quality Management

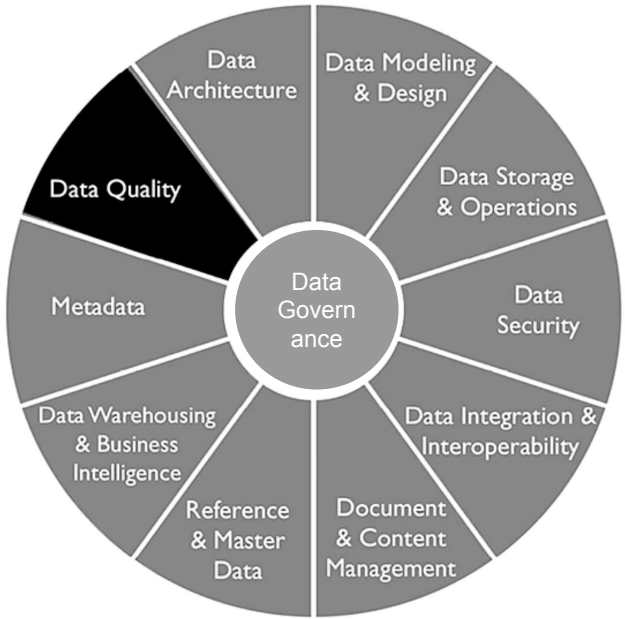
= one of the eleven data management categories identified by DAMA

"[consists of] the planning, implementation and control of the activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers."

Data Governance

= one of the eleven data management categories identified by DAMA

"The discipline which provides the necessary policies, processes, standards, roles and responsibilities needed to ensure that data is managed as an asset"

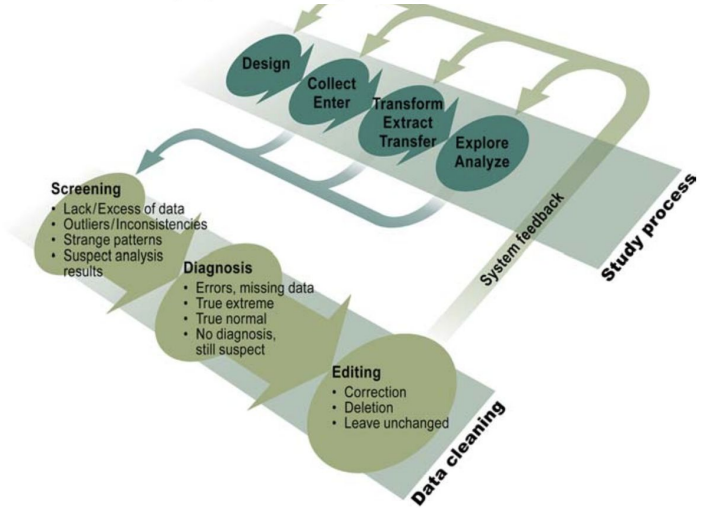


Data management association international (DAMA International) 2017; presentation by LIGHTSONDATA; George Firican https://www.youtube.com/watch?v=4lFTRSBO4ss

Data Quality as Part of Data Management

Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities

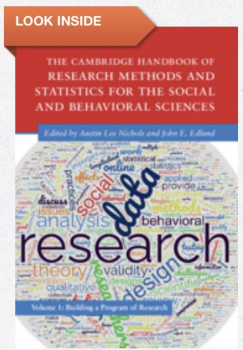
Jan Van den Broeck*, Solveig Argeseanu Cunningham, Roger Eeckels, Kobus Herbst



DOI: 10.1371/journal.pmed.0020267.g001

Figure 1. A Data-Cleaning Framework
(Illustration: Giovanni Maki)

Chapter 21. Data cleaning Solveig A. Cunningham and Jonathan A. Muir



The Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences

Volume 1: Building a Program of Research

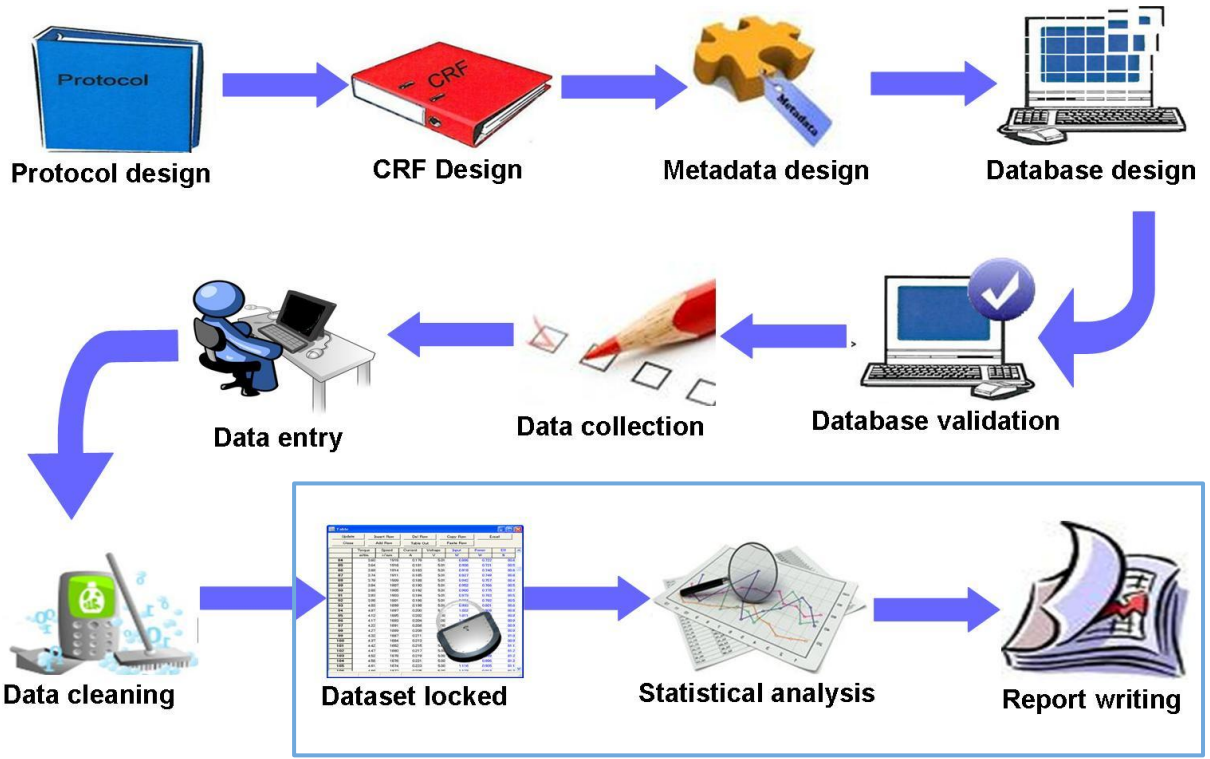
Part of Cambridge Handbooks in Psychology

EDITORS:
Austin Lee Nichols, Central European University, Vienna
John Edlund, Rochester Institute of Technology, New York
View all contributors

DATE PUBLISHED: June 2023
AVAILABILITY: Available
FORMAT: Paperback
ISBN: 9781108995245

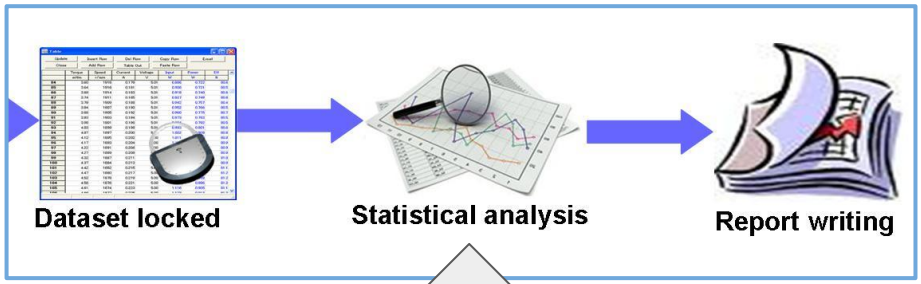
Data Cleaning within Study Design; Data Collection/ Data Entry; Data Management and Data Analysis

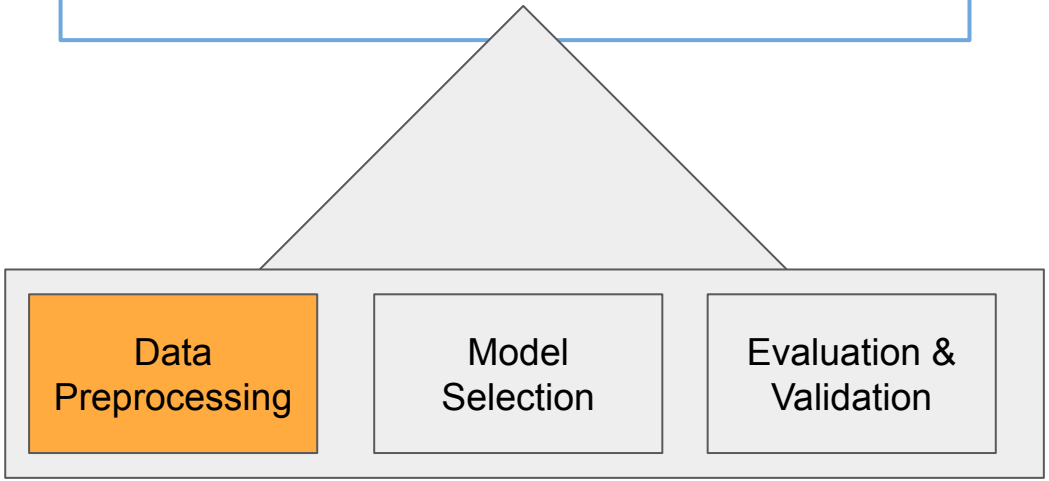
Data Quality as Part of Data Analysis



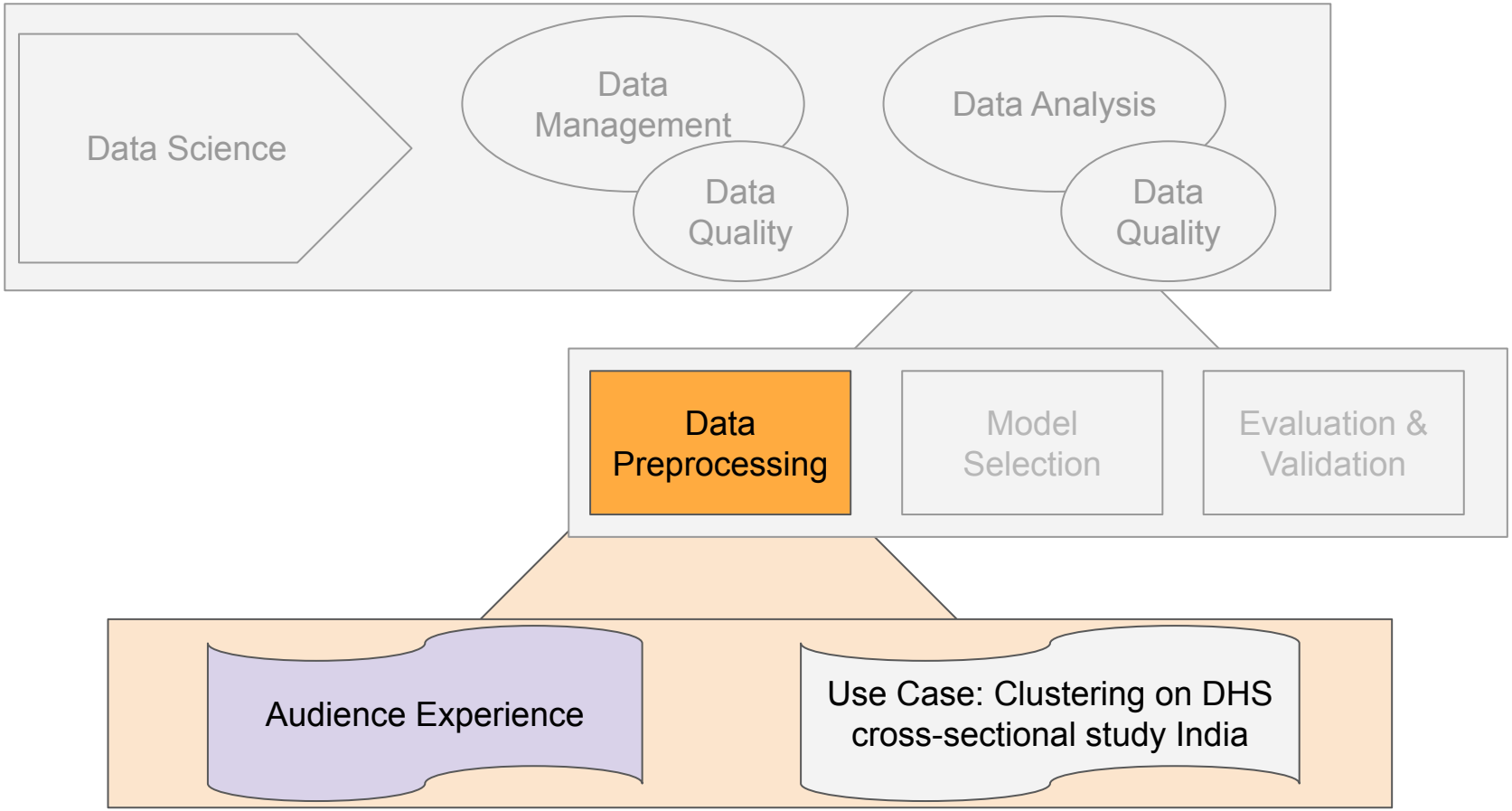
partly extracted from presentation of Kavita Singh from Public Health Foundation India

Data Analysis: From Data to Information

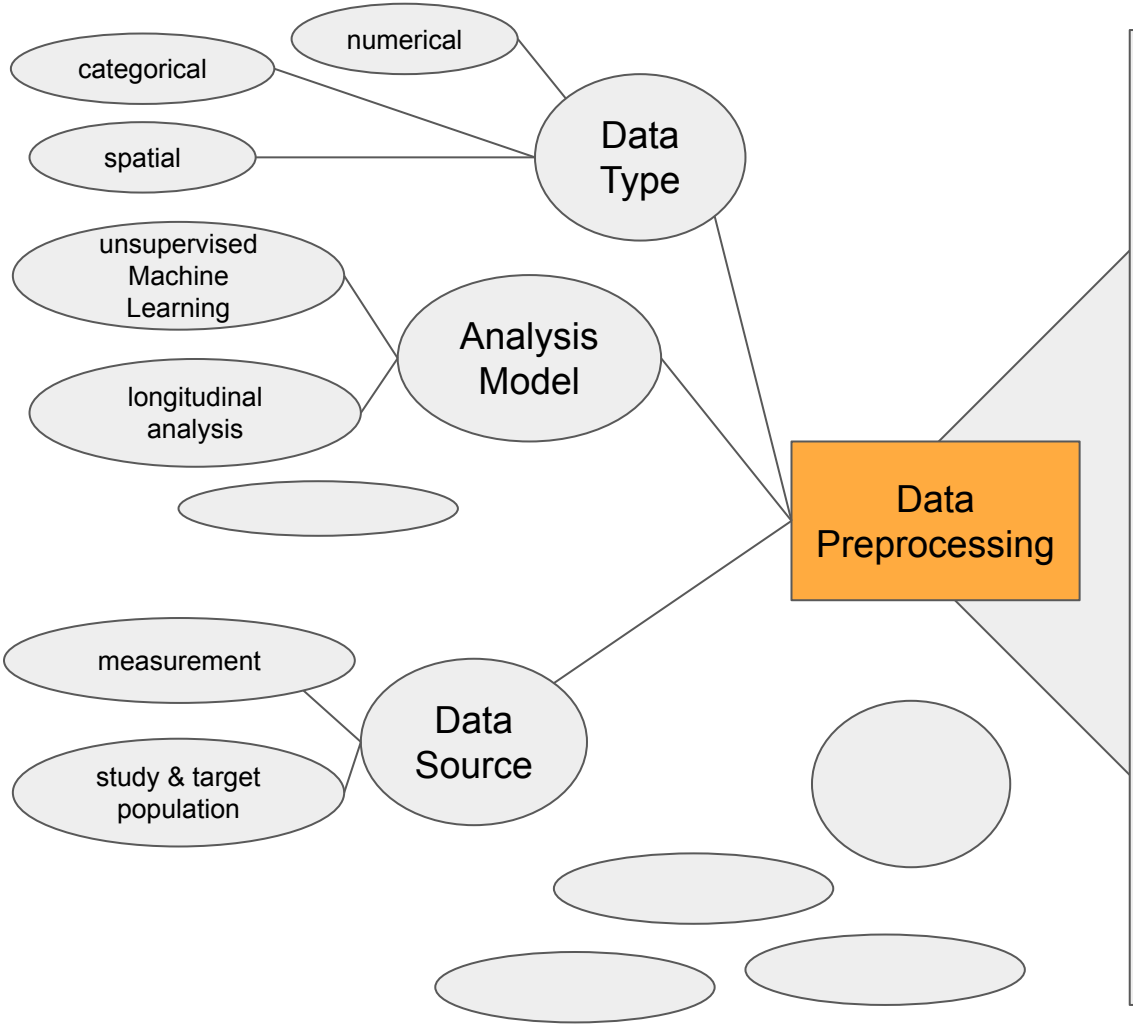




partly extracted from presentation of Kavita Singh from Public Health Foundation India



USAID; Measure Evaluation; 2016; Surfing a Great Wave: Data Science and Global Health https://de.slideshare.net/slideshow/surfing-a-great-wave-data-science-and-global-health/57022175



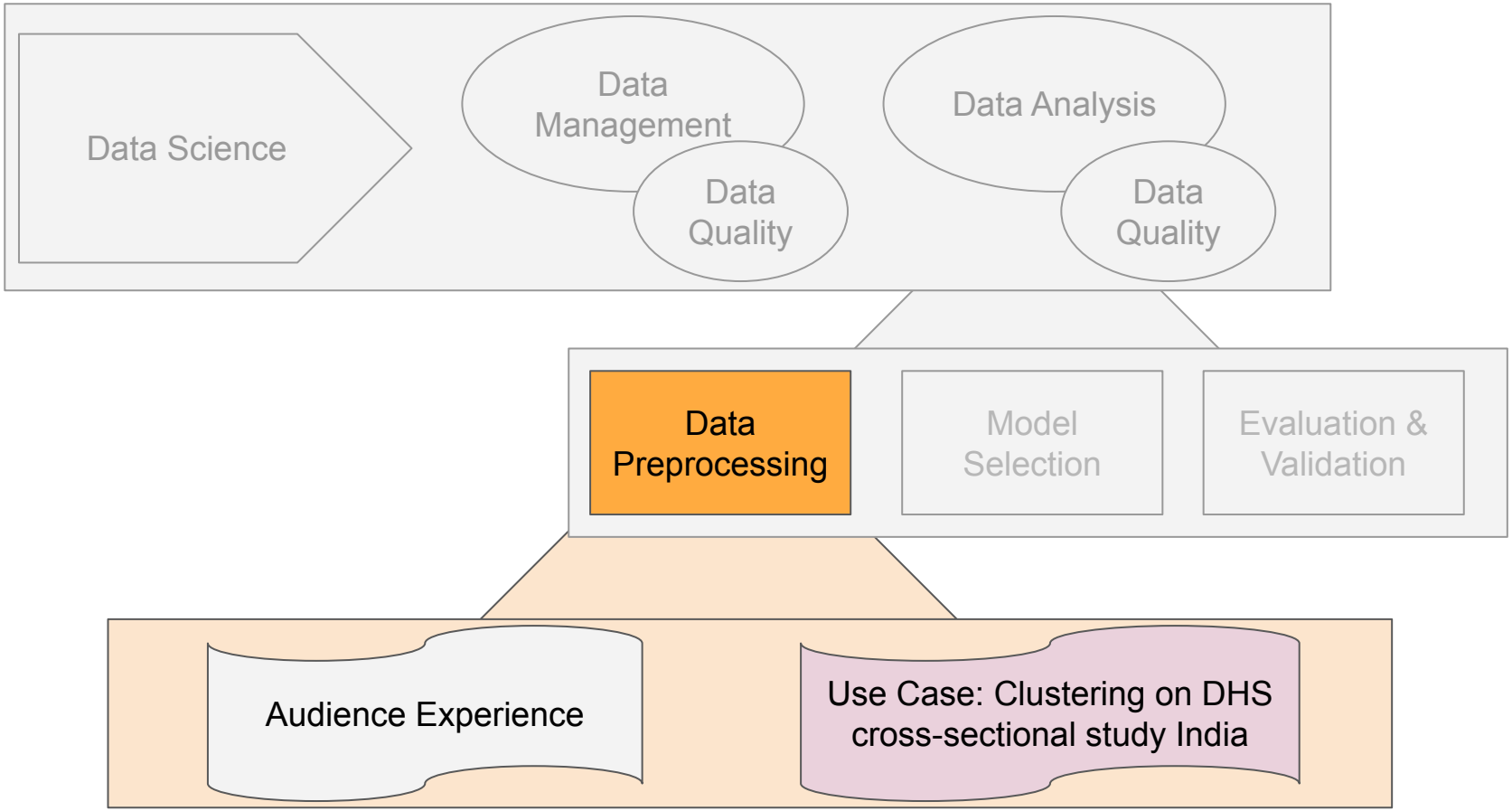
Audience Experience

Questions for audience:

- Collect experiences: What data analysis methods have you been using on what type of data? What data preprocessing was needed for this data analysis?
- Mark the preprocessing steps that are common across data sets & tasks.
- Select one person that can write down the discussion.
- Select one person to present the results.
- Share findings as PDF or PNG within the chat.

Enjoy the discussion!

names: .



USAID; Measure Evaluation; 2016; Surfing a Great Wave: Data Science and Global Health https://de.slideshare.net/slideshow/surfing-a-great-wave-data-science-and-global-health/57022175

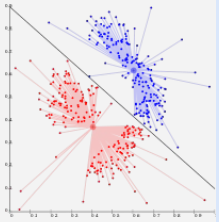
Objective: Address population health provision (i.e. health care delivery) for chronic non-communicable disease through identification of subpopulations with distinct prevalence patterns and their identifiers.

The DHS Program
Demographic and Health Surveys

DHS Program data are collected so they can be used to guide programs and policies to improve health and well-being.
NFHS-5 India: 625.000 women; 80.000 men

= **Cross-Sectional Study**
(data collection at one time point across study population with stratified sampling)

Clustering
k-means



Clustering is a procedure of organizing the objects in groups whose member exhibits some kind of similarity.

[3] https://doi.org/10.1145/2393216.2393327

Clusters are constructed though the analysis of the distribution of instance (people) within a parameter space.

disease markers
biological parameters
numerical
exposure markers
socio-economic factors
behavioural information
categorical

Data Preprocessing

Data Source
Analysis Model
Data Type

Questions for audience:

1. What are steps needed for data preprocessing for this use case?
2. In which ordering would you conduct the data preprocessing?

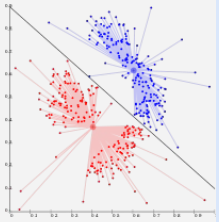
→ write down your own data preprocessing plan/ structure

The DHS Program
Demographic and Health Surveys

DHS Program data are collected so they can be used to guide programs and policies to improve health and well-being.
NFHS-5 India: 625.000 women; 80.000 men

= **Cross-Sectional Study**
(data collection at one time point across study population with stratified sampling)

Clustering
k-means



Clustering is a procedure of organizing the objects in groups whose member exhibits some kind of similarity.

[3] https://doi.org/10.1145/2393216.2393327

Clusters are constructed though the analysis of the distribution of instance (people) within a parameter space.

disease markers
biological parameters
numerical
exposure markers
socio-economic factors
behavioural information
categorical

Data Preprocessing

Data Source
Analysis Model
Data Type

Data Preprocessing

- Usually data is not handed in a format that is ready for straightaway analysis, for which you need to spend some time on data preprocessing.
- Results are highly dependent on decisions made during data preprocessing.
- You can not balance out poor data quality though optimizing the analysis model, your algorithm will not be able to learn meaningful information (better rethink your decisions).
  → It is better to have good data and a simple algorithm than poor data and a very complex algorithm.

1. Exploratory data analysis
2. Variable selection
3. Instance selection - Inclusion Criteria
4. Instance selection - Identification and exclusion of outliers
5. Additional Data preparation (method-dependent)

Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

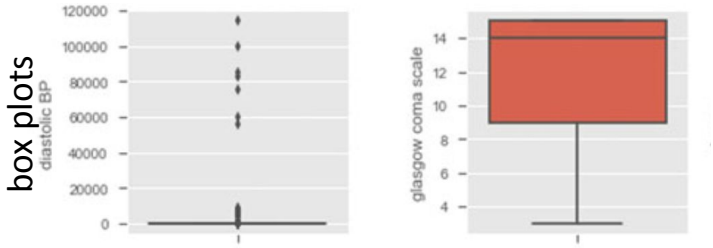
1. Exploratory Data Analysis / Data Visualisation

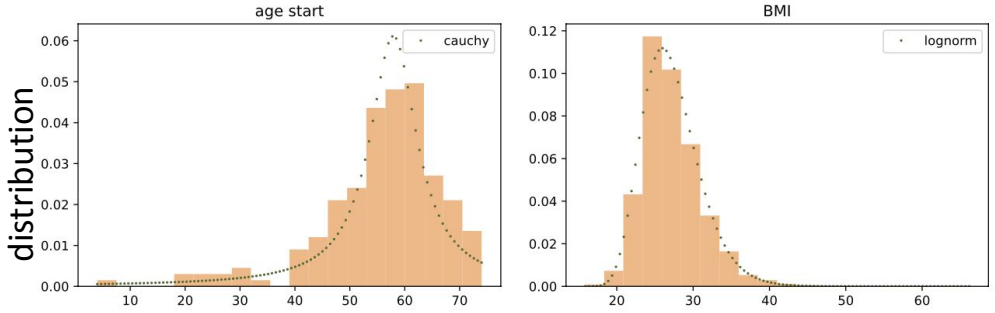
- **Tables**, like summary statistics (categorical values) → like the descriptive view on data yesterday
- **Plots**, like box plots or parameter distributions (numerical parameters)
- **Images**, through dimension reduction methods, like U-Map or PCA (numerical values)

Data Characteristics

	Female (N = 634 794)	Male (N = 84 004)
Haemoglobin [g/dl]	14·0 (1·8)	11·5 (1·6)
Random Blood Glucose [mmol/L]	6·3216 (1·7165)	6·1854 (1·5661)
Diastolic Blood Pressure [mmHg]	79·2 (9·1)	77·5 (8·8)
Systolic Blood Pressure [mmHg]	122·3 (12·2)	116·2 (13·2)
Waist Circumference [cm]	79·8 (11·2)	77·1 (11·6)
Hip Circumference [cm]	89·3 (9·8)	89·7 (10·8)
Height [cm]	162·7 (7·4)	152·1 (6·1)
Weight [kg]	59·2 (11·4)	51·6 (10·7)

Data are Mean (SD), reported separately for male and female population.

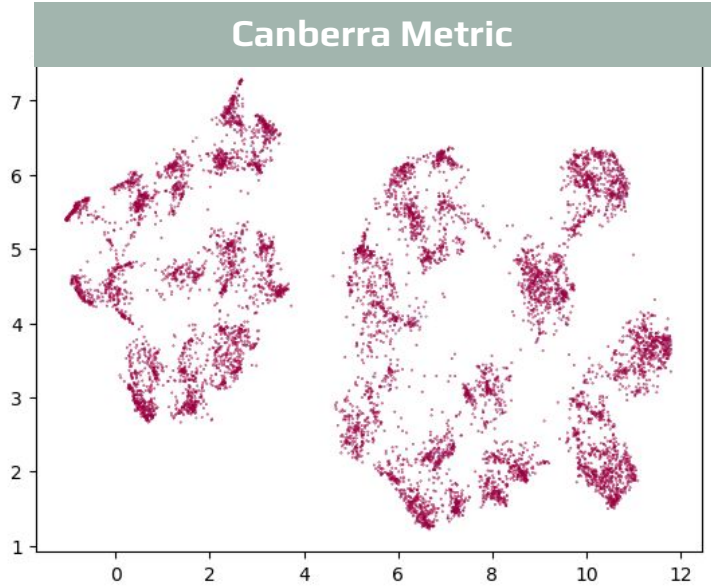




Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

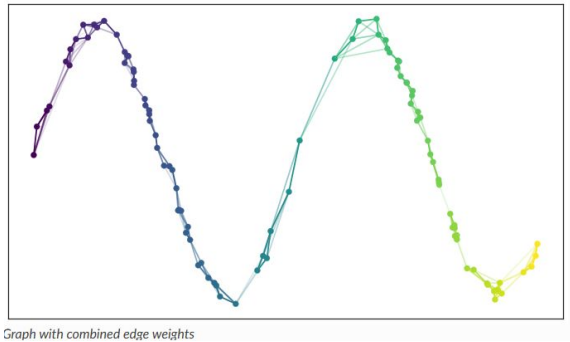
1. Exploratory Data Analysis / Data Visualisation

- **Visualisations though dimension reduction methods, like UMAP or Principal Component Analysis (PCA)**



Choice of Clustering Algorithm based on the data shape

Using UMAP / python package to identify low dimensional embedding of data topology



Introduction to UMAP: https://pair-code.github.io/understanding-umap/

Nitschke, Brandl, Bärnighausen et al.; 2024; preprint Machine learning-based Clustering of Chronic non-communicable Multimorbidity in India: A National Representative Study including 718811 Individuals

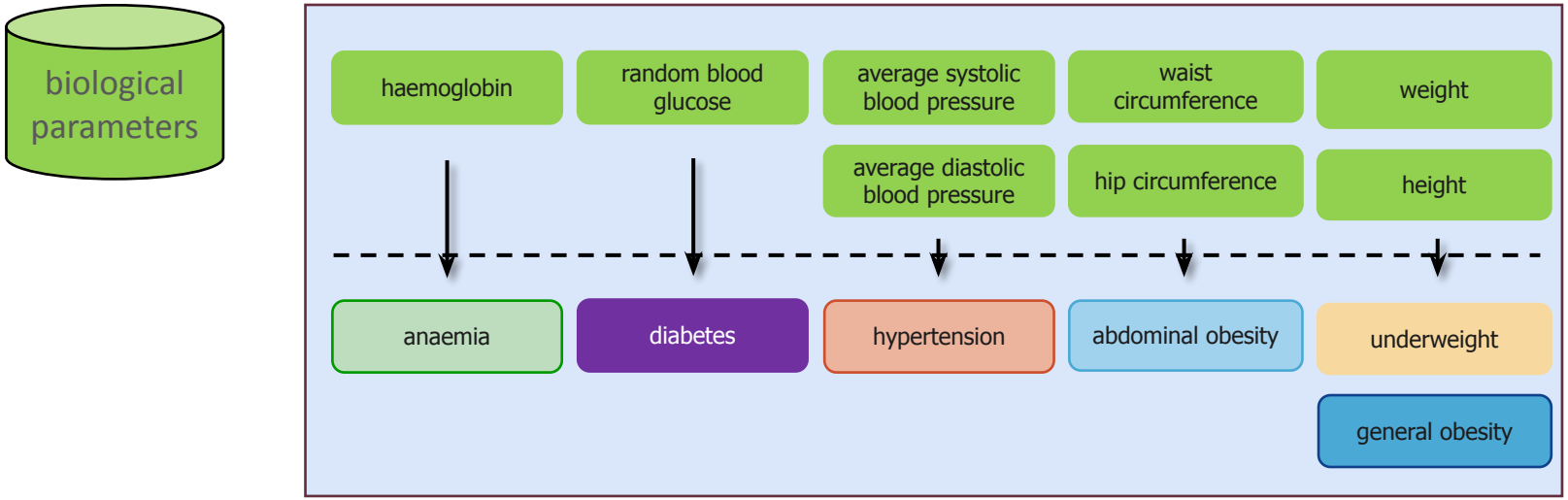
2. Variable Selection

First it is important to understand what the algorithm is supposed to learn so that you can select appropriate information/variables to describe the problem!

- adding more variables tends to decrease the sample size, because fewer patients are likely to have all of them collected at the same time;
- selecting a high number of variables might bias the dataset towards the selection of a specific group of patients whose characteristics required the measurement of those specific variables;
- variables should be independent with minimal correlation;
- the number of observations should be significantly higher than the number of variables, in order to avoid the curse of dimensionality.

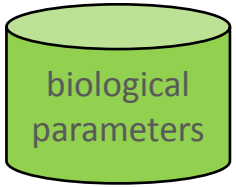
Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

2. Variable Selection



Nitschke, Brandl, Bärnighausen et al.; 2024; preprint Machine learning-based Clustering of Chronic non-communicable Multimorbidity in India: A National Representative Study including 718811 Individuals

2. Variable Selection

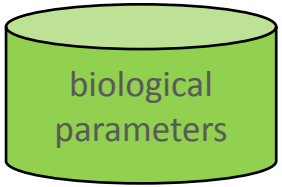


Supplemental Table 3: Disease Definition

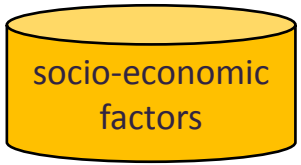
Disease	Parameter	Cut-off Screening	Cut-off Diagnostic
Anaemia	Haemoglobin [g/dl]	Female <12 Male < 13	Female <10 Male < 12
Diabetes, incl. corrected	Random Blood Glucose Level [mmol/L]	>7.77/>140	>11.1/>200
Hypertension, incl. corrected	Systolic/Diastolic Blood Pressure [mmHg]	>130/>80	≥140/≥90
Abdominal Obesity	Waist-to-Hip Ratio	Female > 0.85 Male > 0.9	Female > 0.85 Male > 0.9
Obesity	BMI [kg/m²]	>23.5	>27.5
Underweight	BMI [kg/m²]	<18.5	<18.5

Nitschke, Brandl, Bärnighausen et al.; 2024; preprint Machine learning-based Clustering of Chronic non-communicable Multimorbidity in India: A National Representative Study including 718811 Individuals

2. Variable Selection



- haemoglobin
- random blood glucose
- average systolic blood pressure
- waist circumference
- weight
- average diastolic blood pressure
- hip circumference
- height



- wealth index
- urban vs. rural
- state
- age
- caste
- working
- educational level

- smoking
- soft drinks
- fried food
- religion
- alcohol consumption

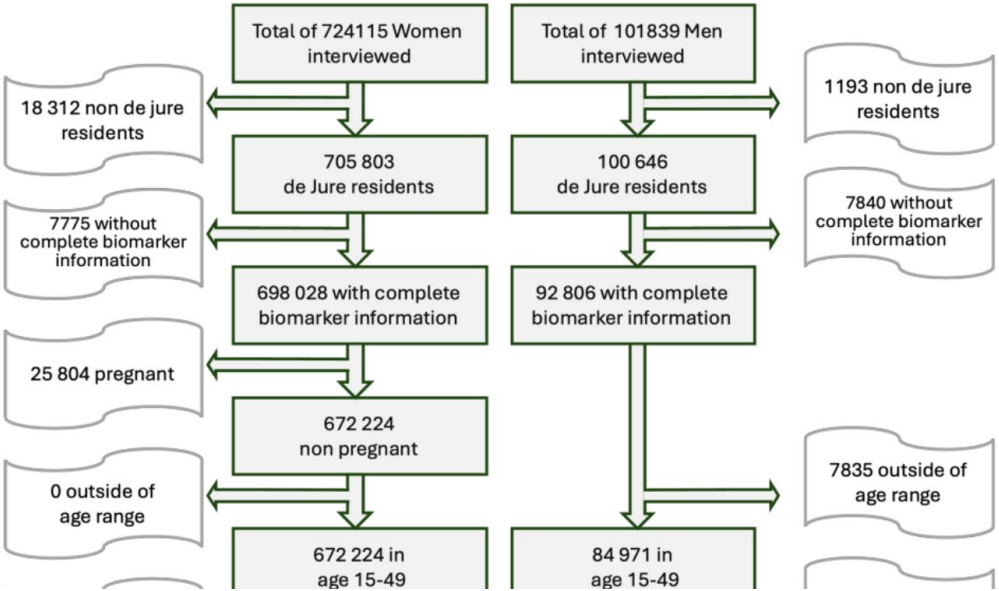
Nitschke, Brandl, Bärnighausen et al.; 2024; preprint Machine learning-based Clustering of Chronic non-communicable Multimorbidity in India: A National Representative Study including 718811 Individuals

3. Instance Selection - Inclusion Criteria

Analogously to case and control selection (discussed by Vidaisha Naidoo), the definition of inclusion criterias is essential for the quality of your analysis result!

- Analysis for men and women separately
- Only include resident people
- include only people without **missing values** (alternatively one can perform imputation)
- drop duplicated instances (redundancy)
- place age range

→ It is always important to keep track of the size of data while making decisions about inclusion, criteria



Nitschke, Brandl, Bärnighausen et al.; 2024; preprint Machine learning-based Clustering of Chronic non-communicable Multimorbidity in India: A National Representative Study including 718811 Individuals
Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

4. Instance Selection - Identification of Outliers

1. **Impossible values:** Values that are logically or physically impossible (e.g., a negative height).
    a. 1D-limits defined by expert knowledge
2. **Incorrect possible values**: Values that are within a valid range but incorrect in context
    a. parameter construction
    b. 2D-plots (Scatter Plots)

Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

4. Instance Selection - Identification of Outliers

1. Impossible values: Values that are logically or physically impossible
   a. 1D-limits defined by expert knowledge

Supplemental Table 1: Definition of non-plausible biological values

Parameter	Lower Limits	Upper Limits
Height [cm]	60	250
Weight [kg]	30	160
Waist circumference [cm]	50	160
Hip circumference [cm]	50	160
Random Blood Glucose [mmol/L]	2.2	33.3
Systolic blood pressure [mmHg]	70	240
Diastolic blood pressure [mmHg]	40	130
Haemoglobin [g/dl]	1	120

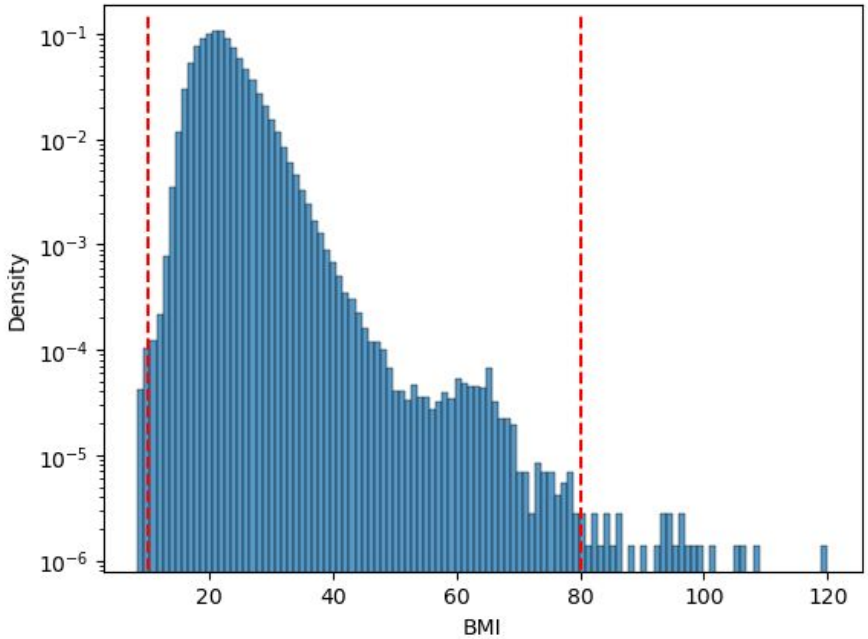
Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

4. Instance Selection - Identification of Outliers

3. Incorrect possible values: Values that are within a valid range but incorrect in context

→ a) parameter construction

- Visual, distribution based, density based
- Check via Adjusted Rand Index,
  that clustering stays stable



Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

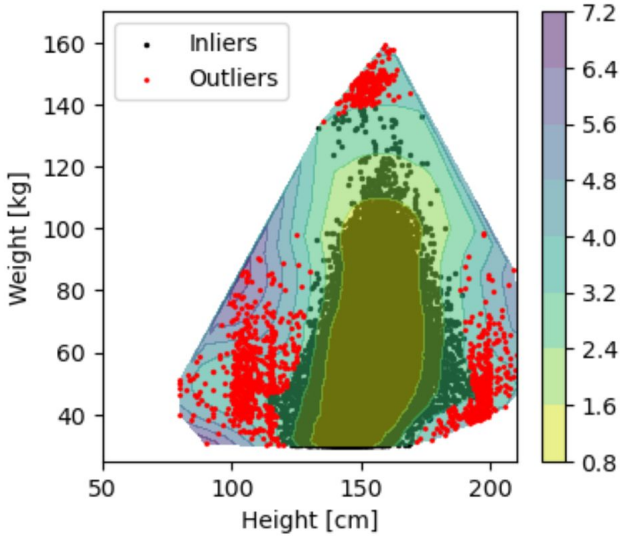
4. Instance Selection - Identification of Outliers

3. Incorrect possible values

→ 2D: **Local-Outlier Factor** for BMI

"The Local Outlier Factor (LOF) algorithm is an unsupervised anomaly detection method which computes the local density deviation of a given data point with respect to its neighbors. It considers as outliers the samples that have a substantially lower density than their neighbors."

https://medium.com/@pramodch/understanding-lof-local-outlier-factor-for-implementation-1f6d4ff13ab9

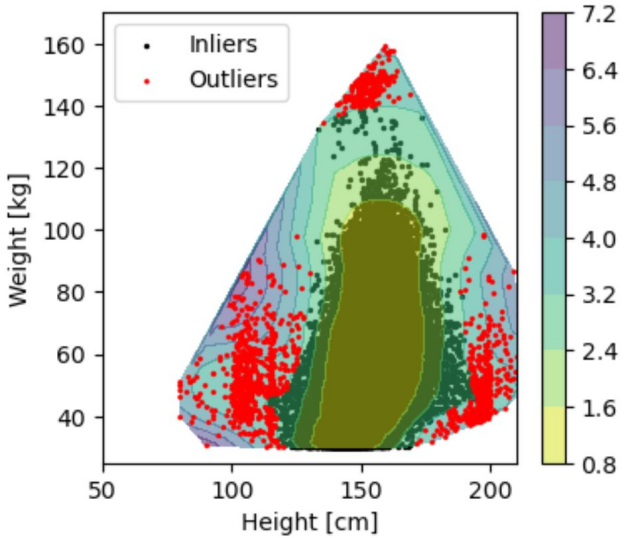


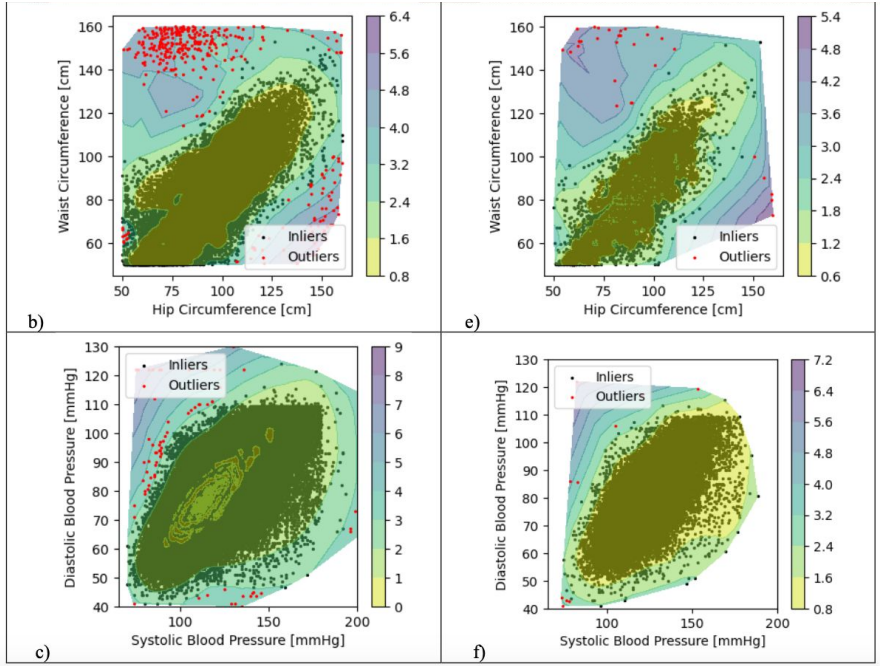
Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

4. Instance Selection - Identification of Outliers

3. Incorrect possible values

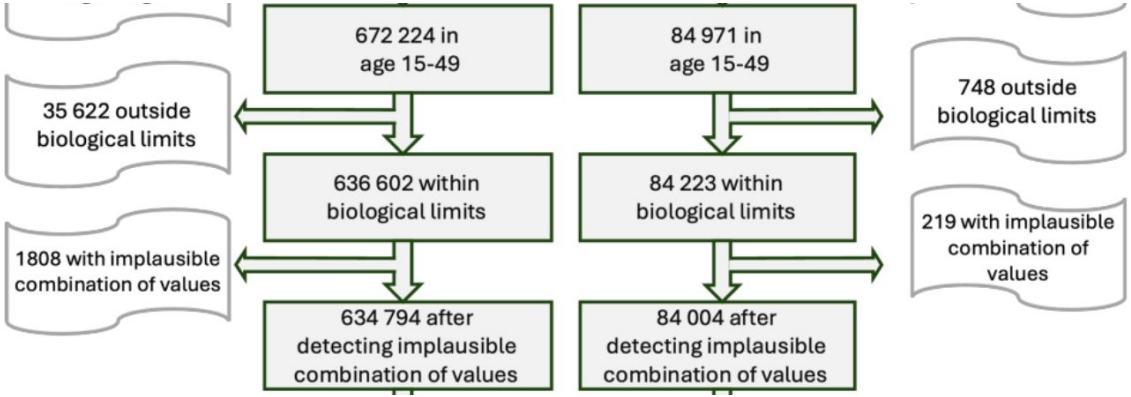
→ 2D: **Local-Outlier Factor** for BMI





Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

4. Instance Selection - Identification of Outliers



Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

5. Additional Data Preparation (method-dependent)

Additional factors influencing the results retrieved from clustering analysis and therefore need to be addressed during data preprocessing:

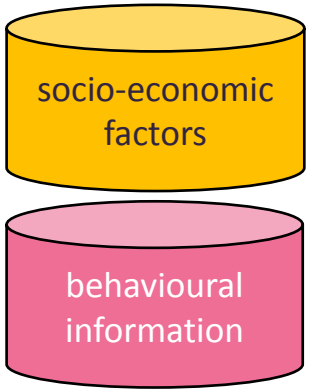
1. **Normalization or scaling of numerical values**: is used to ensure that all features lie between a given minimum and maximum value, often between zero and one. The maximum and minimum values of each feature should be determined during the training phase and the same values should be applied during the test phase.
   ⟶ cluster analysis is done on scaled values centred to a mean of 0 and a standard deviation of 1
2. **Ordering of ordinal parameters:** For ordinal variables, the order of categories matters. Proper encoding ensures that clustering algorithms interpret the ordinal nature correctly.
3. **Encoding nominal parameters** (one-hot encoding): One-hot encoding is used for categorical variables without intrinsic ordering. Proper encoding is important for accurate clustering.
4. **Bias-inducing behaviors** such as eating, drinking, medication, and acute stress
5. **Sampling weights of individuals:** Sampling weights can affect the representation of data points in clustering.

Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

5. Additional Data Preparation

2. Ordering of ordinal parameters: For ordinal variables, the order of categories matters. Proper encoding ensures that clustering algorithms interpret the ordinal nature correctly.

3. Encoding nominal parameters (one-hot encoding): One-hot encoding is used for categorical variables without intrinsic ordering. Proper encoding is important for accurate clustering.



Examples:

‘Marital’ Encoding
1: Never married
2: Currently married
3: Separated
4: Divorced
5: Widowed

→ One–hot encoding
‘Marital1’, ‘Marital2’…

How often do you drink alcohol:
1 "Almost every day"
2 "About once a week"
3 "Less than once a week“

→ 1 "Less than once a week“
2 "About once a week“
3 "Almost every day"

Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

5. Additional Data Preparation

4. Approaches used for dealing with bias-inducing behaviors:

- **Exclusion for single people**: Excluding individuals due to their bias-inducing behaviors is possible, but generally not a recommended approach because it could lead to loss of valuable data and introduce further biases
  → **like pregnant women**

- **Computation of counterfactuals for entire subpopulations**:
  → **like fasting and medication on blood glucose, or medication on blood pressure**

Supplemental Table 2: Individuals affected by imputation

	Female (N = 634 794)	Male (N = 84 004)
Hypertension Medication	17097 (2·7)	1496 (1·8)
Diabetes Medication	7584 (1·2)	1097 (1·3)
Fasting	9632 (1·5)	1533 (1·8)
Total	32374 (5·1)	3895 (4·6)

Data are N (%).

Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

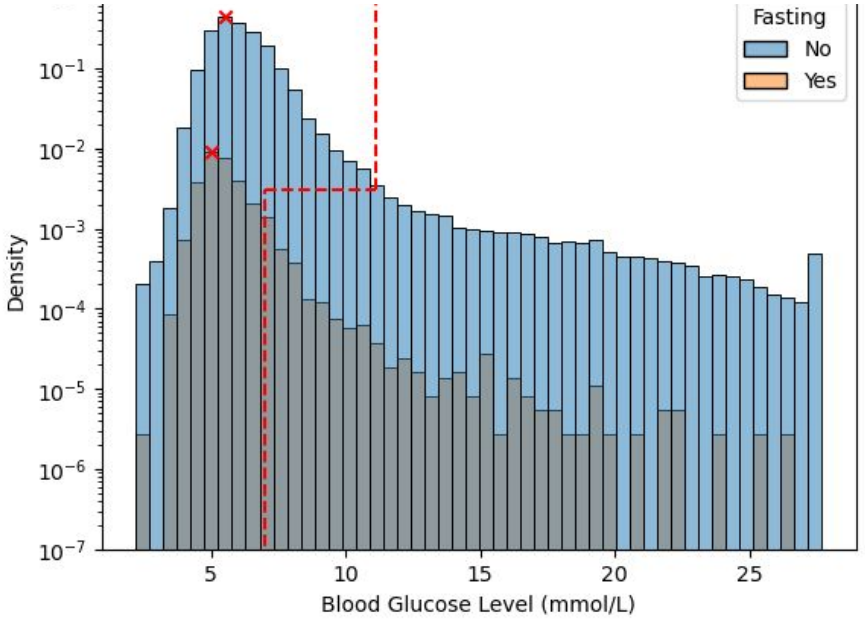
5. Additional Data Preparation

Example Fasted Diabetes:

How to treat people that

(1) have been fasting, to be equivalently viewed as

(2) people without fasting, by the algorithm



A. Shift by constant

B. Fit of Function/ Distribution of non-fasting people with blood glucose values and draw randomly

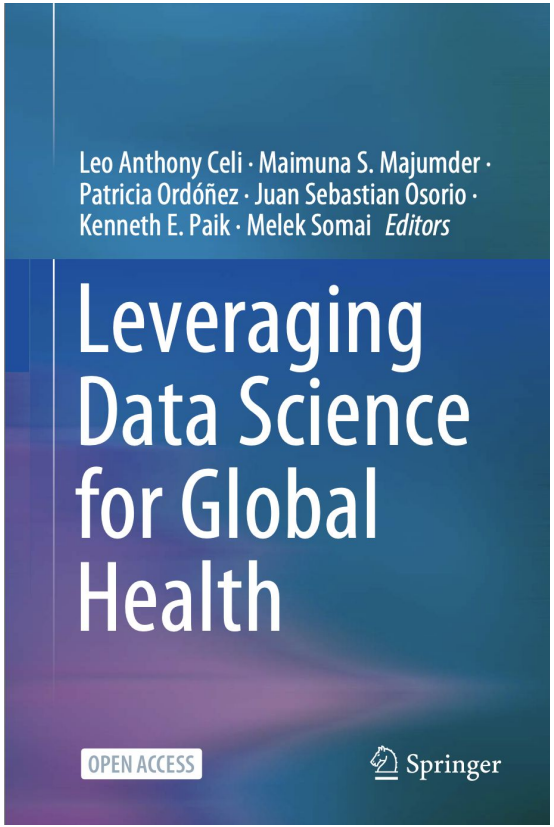
C. Imputation methods: prediction by other values → XGBoost

(Introduction to XGBoost: https://medium.com/analytics-vidhya/introduction-to-xgboost-algorithm-d2e7fad76b04)

Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

Data Preprocessing

Getting started with Data Science in Python. Including different categories of Machine Learning.

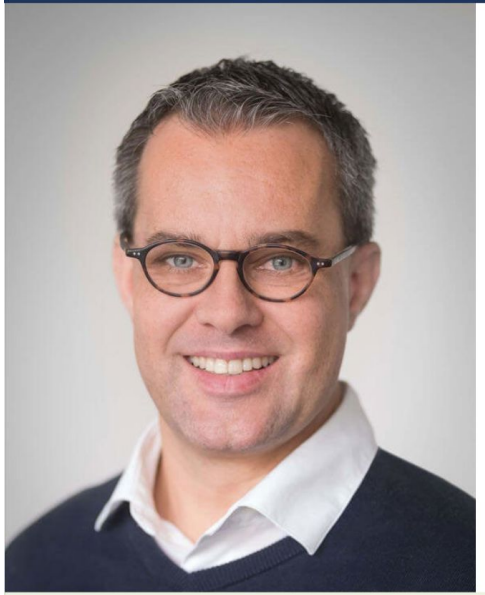


Calvin J. Chiew, 2020; *Leveraging Data Science for Global Health;* Chapter 8: Applied Statistical Learning in Python https://link.springer.com/book/10.1007/978-3-030-47994-7

DSI Monthly Seminar Series

WASHA Takwimu (Swahili for "Ignite Data")

Working on Applications for Data Science and Health in Africa



Professor Till Bärnighausen

Alexander von Humboldt University Professor & Director, Heidelberg Institute of Global Health

Senior Faculty, Africa Health Research Institute

Adjunct Professor of Global Health, Harvard T.H. Chan School of Public Health

August 21st, 2024

3-4pm SAST (South Africa) / 9-10am ET (US)

Join via Zoom link here

TEAM AND SUPPORT

Anna-Katharina Nitschke, Carlos Brandl, Jonathan Elias Berthold, Jannis Demel, Carola Behr Fabian Egersdörfer and Prof. Matthias Weidemüller

Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany

in collaboration with Prof. Till Bärnighausen, Kavita Singh, Prof. Ullrich Köthe, Michaela Theilmann, Jen Manne-Goehler, Prof. Martin Siegel, Sujata

THANK YOU!

Stay updated - publication will follow soon

STRUCTURES CLUSTER OF EXCELLENCE

DFG Deutsche Forschungsgemeinschaft

bwForCluster Helix

The DHS Program Demographic and Health Surveys

SPONSORED BY THE Federal Ministry of Education and Research

Baden-Württemberg MINISTRY OF SCIENCE, RESEARCH AND ARTS