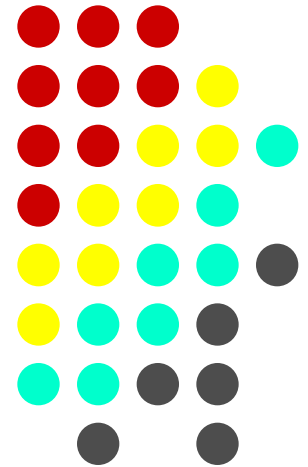
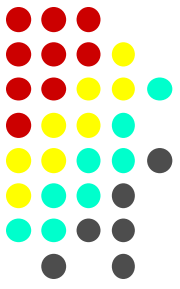


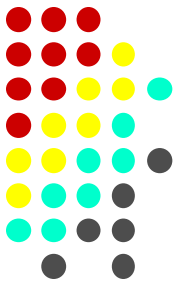
Introduction to Data Management





OBJECTIVES

- To introduce the basic concepts of Data Management
- To introduce the basic concepts of statistics
- To present the table and the tabulation plan
- INTRODUCTION TO STATISTICAL PACKAGES



WHAT IS DATA?

WHAT IS “INFORMATION”?

WHAT IS DATA MANAGEMENT

WHAT IS “DATA”?



- Data
 - raw information or facts that become useful information when organized in a meaningful way
 - information organized for analysis.
 - could be of qualitative or quantitative nature
 - individual measurements or individual records in a questionnaire

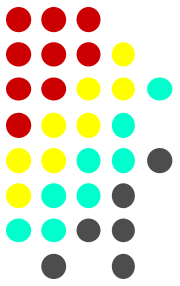
block	plot	cobwt	grainwt	carbon
1	1	17.7	811	2.6
1	2	19.9	136	2.54
1	3	28.3	172	2.83
2	1	26.63	148	2.41
2	2	24.88	170	2.39
2	3	25.46	158	2.47

WHAT IS “INFORMATION”?



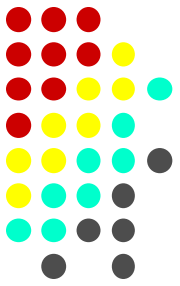
- Information is processed data from which conclusions can be drawn
- Information is a valuable resource for decision-making and for planning
- It is the results of processing, gathering, manipulating and organising data in a way that adds to the knowledge of the receiver

WHAT IS DATA MANAGEMENT?



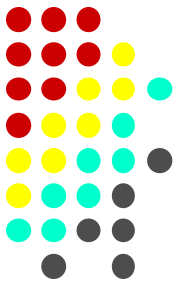
- DM is concerned with “looking after” and processing data – it involves:
 - Looking after field data sheets
 - Entering data into computer files
 - Checking and correcting the raw data
 - Preparing data for analysis
 - Documenting and archiving the data and meta-data
- DM is the consolidation of data (and meta-data) in a way that is easy to manipulate, retrieve and maintain

WHY IS DM IMPORTANT?



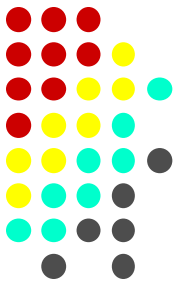
- Ensures data for analysis are of high quality so that conclusions are correct
- Good DM allows further use of the data in the future and enables efficient integration of results with other studies
- Good DM leads to :
 - Improved processing efficiency
 - Improved data quality
 - Improved meaningfulness of the data

DATA MANAGEMENT PROBLEMS



- Lack of skills – inability to use software or set up data checking procedures
- Multiple copies of files
- No one with responsibility for checking data
- No clear policy on archiving or making data available
- Lack of documentation
- Multiple entry of the same data
- Hand pre-processing of data

PLENARY DISCUSSION 1



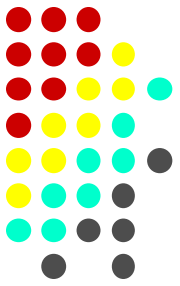
- Case Study Discussion
 - Have you encountered similar situations in your own workplace?

SOME STEPS IN DM



- Designing field data collection sheets
- Collecting data with appropriate quality control
- Checking raw data
- Data entry and organisation of computer files
- Backup of files
- Processing of data for analysis
- Checking of processed data
- Archiving data and meta-data for future use

NON-ELECTRONIC DM



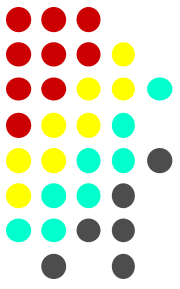
- Handling of questionnaires in the field (both completed and blank)
- Storage of questionnaires – protection from weather, termites, etc.
- Movement of the questionnaires – who has access to them
- Editing and coding

ELECTRONIC DATA MANAGEMENT



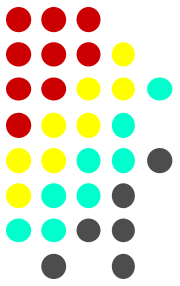
- Designing data entry system
- Data entry – including double entry
- Data cleaning – consistency checks
- Data security – regular backups – where are backups stored?
- Storage – how safe is your data?
- Documentation

DOCUMENTATION



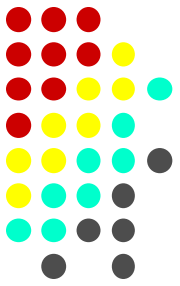
- Documentation should be part of the DM project planning:
 - **File Structure** – how will the data be organised
 - **Naming conventions** – for files and variables
 - **Data integrity** – what checks are in place
 - **Dataset documentation** – how will this be produced
 - **Variable construction** – what variables will be constructed following data collection; how will these be documented
 - **Project documentation** – how will you document decisions taken on field procedures, coding, etc.

ARCHIVING



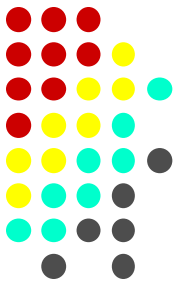
- Plans for archiving should be included in the project proposal and must be fully costed
- Proposal should include:
 - Schedule for data sharing
 - Formatting of final dataset
 - Documentation to be provided
 - Analytical tools to be provided if any
 - Mode of data sharing

BUILDING A DM STRATEGY



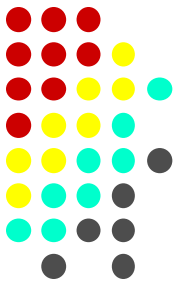
- Good DM is not something that will “look after itself” or evolve when left for long enough
- A DM strategy requires:
 - Commitment
 - Skills
 - Time
 - Money

DATA MANAGEMENT PLAN



- A DM plan will include:
 - Clearly defined roles for staff
 - A regular backup procedure
 - Details of data quality checks
 - Including DM on the agenda of project meetings
 - Procedure for upgrading software
 - Details of how archive is to be produced
 - Details of how the archive is to be maintained
 - For example can you still read 5¼” floppy disks?!
 - Bernoulli disks

ROLES AND RESPONSIBILITIES



- **Organisers** – handle raw data on a daily basis. They set up data filing systems, enter and check data and maintain data banks
- **Analysers** – analyse and interpret data, reducing raw observations to useful information
- **Managers** – responsible for providing an enabling environment for the first two groups and ensuring all commitments to stakeholders are met



GENERAL STATISTICS CONCEPTS

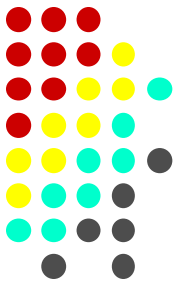


TABLE DESIGN

STATISTICAL TABLES



- **A statistical table is a collection and arrangement of data (numbers, derived measures, etc.) In which each item is uniquely identified.**

STATISTICAL TABLES



Mainly classified into three;

- Leader work
- Text tabulations
- Formal statistical tables

STATISTICAL TABLES

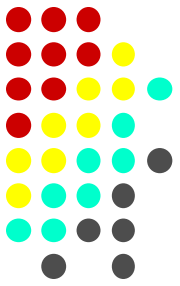


- Leader work;
 - Usually part of the text
 - Simple table without title, and boxhead
 - Presents only one or two columns of figures
 - Cannot stand alone
 - Difficult to cite
 - Example.

The unemployment rates among young adults are listed below:

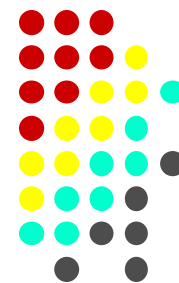
12 to 14 years.....	14.1
15 to 19 years.....	12.0
20 to 24 years.....	10.8

STATISTICAL TABLES



- Text tabulation;
 - Usually appears as part of the text of a report
 - Simple ruled table without a table number or title
 - Has no identification and is introduced by an explanatory statement
 - It is difficult to cite
 - Cannot stand alone

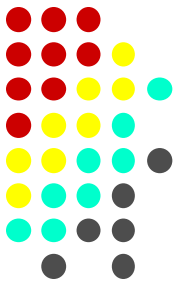
Example of a text tabulation



Unemployment rates among young people are as follows:

Age	Total Labour force	Unemployed	
		Number	Percent
12 to 14 years	300,000	28,955	9.7
15 to 19 years	550,000	65,545	11.9
20 to 24 years	710,000	96,567	13.6

STATISTICAL TABLES

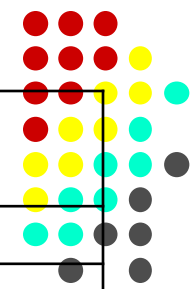


- **Formal Statistical table;**
 - FINISHED FORM OF TABULAR PRESENTATION
 - SELF-SUFFICIENT
 - INCLUDES A NUMBER AND A TITLE
 - INCLUDES HEADNOTES AND FOOTNOTES
 - STUB AND BOXHEAD ARE PRECISELY WORDED

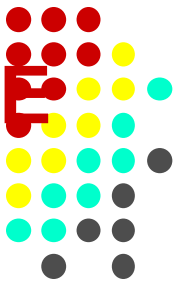
Table 2.1: Household Population by Age, Sex and Residence
(Percentage Distribution of De facto Population according to 2006 UDHS)

Age Group	Urban-Rural Residence		Total
	Urban	Rural	
0 to 4 years	16.6	20.5	20.0
5 to 9 years	14.2	17.6	17.2
10 to 14 years	13.2	15.2	15.0
15 to 19 years	12.5	8.9	9.4
20 to 24 years	11.9	6.9	7.5
25 to 29 years	10.1	6.5	6.9
30 to 34 years	6.9	5.4	5.6
35 to 39 years	5.0	4.2	4.3
40 to 44 years	2.8	3.1	3.0
45 to 49 years	2.2	2.4	2.4
50 to 54 years	1.6	2.3	2.3
55 to 59 years	0.7	1.8	1.7
60 to 64 years	0.8	1.8	1.7
65 to 69 years	0.6	1.2	1.1
70 to 74 years	0.4	1.1	1.0
75 to 79 years	0.2	0.5	0.5
80+ years	0.2	0.6	0.6
Missing/ Don't know	0.1	0.0	0.0
Total	100.0	100.0	100.0
Number	4,674	31,855	36,528

Note: De facto population, are persons who stayed in the household the night before the interview



STRUCTURAL PARTS OF A FORMAL TABLE



- **TABLE NUMBER**

- IDENTIFYING DEVICE
- POSITIONS A TABLE WITHIN A SERIES
- IN A SINGLE REPORT NO TABLE SHOULD BEAR SAME NUMBER

- **TABLE TITLE**

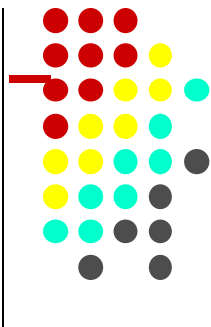
- BRIEF STATEMENT OF THE NATURE, CLASSIFICATION, UNIVERSE, GEOGRAPHICAL AREA AND TIME REFERENCE
- IT IS THE WHAT, HOW, WHERE AND WHEN OF THE TABLE
- IT DESCRIBES THE TABLE CONTENTS
- SERVES AS A REFERENCE GUIDE
- DISTINGUISHES BETWEEN TABLES IN THE SAME REPORT.
- WRITTEN IN TELEGRAPHIC STYLE, NOT IN COMPLETE SENTENCES
- VERBS AND UNNECESSARY WORDS ARE OMITTED
- HEADNOTE USED TO QUALIFY AN OTHERWISE A LONG TITLE
- NEVER BE FOOTNOTED
- SHOULDN'T HAVE A PERIOD/ FULL-STOP AT THE END
- MAKE THE TABLE TITLE PROMINENT.



- **EXAMPLES OF TABLE TITLE:**
 - ***TABLE 2.1: RELATIONSHIP BY AGE AND SEX, UGANDA: 1991***
 - ***TABLE 2.1: PERCENTAGE DISTRIBUTION OF HOUSEHOLD POPULATION BY RELATIONSHIP, AGE AND SEX, UGANDA: 1991***
 - ***TABLE 2.1: TABLE SHOWING THE PERCENTAGE DISTRIBUTION OF POPULATION ENUMERATED FROM HOUSEHOLDS IN UGANDA CATEGORIZED ACCORDING TO RELATIONSHIP, AGE AND SEX OF INDIVIDUALS ACCORDING TO THE POPULATION AND HOUSEHOLD CENSES OF 1991***

STRUCTURAL PARTS OF A FORMAL TABLE

CONT...



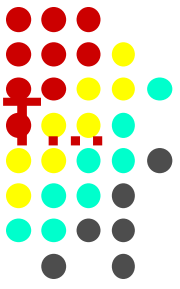
- Headnote

- Statement placed below the table title
- Presents information which cannot be presented elsewhere
- Refer to the table as a whole
- Enclosed in parentheses (brackets)
- Follows the same rules of the table title.

- Boxhead

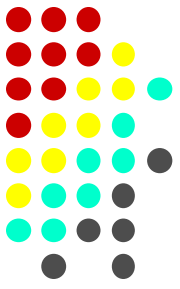
- Contains the individual column heads, spanner heads, and banner heads
- Column head is a descriptive title placed above the column
- Column head is the basic unit of boxhead
- Spanner is a classifying, descriptive, or qualifying caption
- Banner is the head that extends across the entire boxhead
- Wording in the boxhead should be brief but clear

STRUCTURAL PARTS OF A FORMAL TABLE –CONT



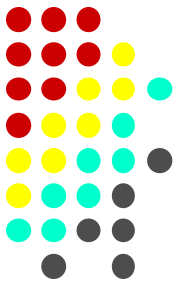
- Stub
 - Usually the first column on the left
 - Contains the centerheads and stubheads
 - Describes the specific data in the table and makes clear the relationships among the items
 - Center head describes a group of line captions until another center head is reached
 - Line caption describes the data on a given line of the cells
 - A group of related line captions with their heads constitute a block

STRUCTURAL PARTS OF A FORMAL TABLE –CONT...

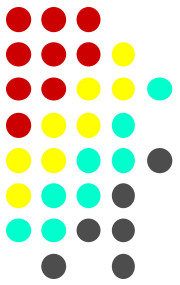


- Field
 - Depository of statistical information
 - Extends from the bottom rule of the boxhead to the bottom of the table, and from the column rule at the right of the stub to the right-hand edge of the table
 - A cell is the basic unit of presentation
- Footnote
 - A statement or note inserted at the foot (bottom) of a table
 - Serves as a specific reference, source note of the data or a general note referring to the table as a whole or a major portion of it

DATA TABULATION PLAN

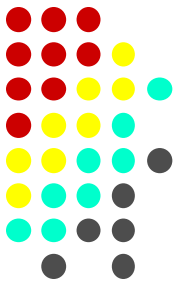


- Provides model tables to present the major findings of a survey in a manner that will be useful to policy makers and program managers
- helps to provide guidance concerning the most important indicators that should be derived from the data to be analysed
- presents the level of analysis expected



INTRODUCTION TO STATISTICAL PACKAGES

REASON FOR STATISTICAL ANALYSIS



- Transform raw data into information
- The general purpose of statistical analysis is to provide meaning to what otherwise would be a collection of numbers and/or values.
- Provide a way of drawing inductive inferences from data and distinguishing the signal (the phenomenon of interest) from the noise (statistical fluctuations) present in the data
- Statistical analysis procedures are categorized according to the type of statistics generated; i.e descriptive, associative, and inferential.

REASON FOR STATISTICAL ANALYSIS- Cont..



- 1. *Descriptive statistics*** portray individuals or events in terms of some predefined characteristics, like measure of central tendency and dispersion –Mean, Median, Range, Standard Deviation, etc.
- 2. *Associative or relative statistics*** seek to identify meaningful interrelationships between or among data. Such statistics include; univariate, bivariate and multivariate analysis. For instance, "Is there a relationship between salt intake and diastolic blood pressure among middle-age women?" is a problem definition suitable for analysis by associative statistics.

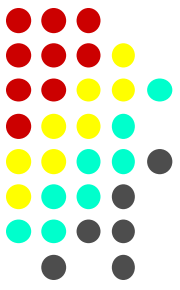
REASON FOR STATISTICAL ANALYSIS-Cont..



3. ***Inferential statistics*** seek to assess the characteristics of a sample in order to make more general statements about the parent population, or about the relationship between different samples or populations.

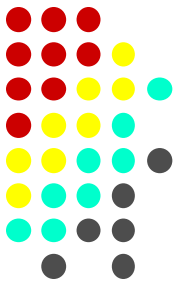
 - Measures of differences of the means and measures of statistical significance
 - For Example; "Does a low sodium diet lower the diastolic blood pressure of middle-age women?" represents a problem definition suitable for inferential statistics.

ISSUES TO CONSIDER IN ANALYSIS



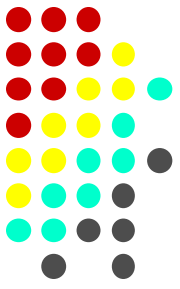
- There are a number of issues to consider with respect to data analysis. These include:
 - Having the necessary skills to analyze
 - Following acceptable norms for data analysis and presentation
 - Choosing the appropriate statistical software
 - Providing honest and accurate analysis
 - Manner of presenting data
 - Extent of data analysis

WHAT IS A STATISTICS PACKAGE?

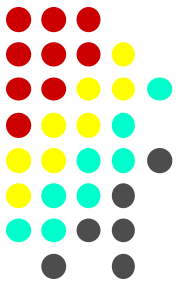


- It is essentially a column calculator
 - like Excel when the data are in “list format”
- A Statistical package is a computer programme that specializes in statistical data analysis.

WHAT STATISTICAL SOFTWARES CAN DO IN RELATION TO DATA ANALYSIS



- Input data into the computer
- Organise data
- Compare data
- Manage data
- Create new variables
- Summarise data (transform raw data)
- Generate tables and graphs



END