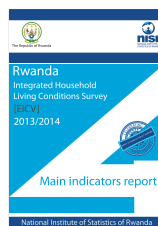


Describing data

LECTURE 15

The dissemination phase

- Last step of the journey
- Typically, after survey and data processing are concluded, a report is released, aimed at describing **main findings** from survey



Dissemination of main findings

- **Benefits**
 - Inform general public, researchers, specialists
 - Influence policy decisions
 - Promote NSO itself: credibility increases through transparency
- **Risks**
 - Exposure to criticism and contradiction
 - Loss of exclusivity
 - Lack of technical capacity to do the dissemination
- **Costs**
 - Creating and documenting data files
 - Creating access tools and safeguards
 - Responding to inquiries

The dissemination report

Varies according to topics of the survey, target audience, etc., but there are some common elements:

1. **Background information** on sampling
2. **Descriptive statistics** (roughly corresponding to survey modules)
3. In the case of income and expenditure surveys, measures of **inequality and poverty**

Next slides cover these 3 points, with tips for effective presentation and examples.

1. Background information

The need for background information

- Survey reports are usually designed to be accessible to a **non-technical audience**
- But **technical background information must still be present**, to inform more advanced readers and facilitate comparisons over time and across countries
- Background information can be presented separately from “core” results, as introductory chapters, appendices, or even a companion document, but must not be omitted

What not to miss

- Reports should document at least the following **survey design features** and **processing choices**:
 - a. Sampling design
Sample size, stratification, representativeness...
 - b. Data collection and processing
Fieldwork, outlier detection and treatment, data imputation...
 - c. Definitions of economic concepts and aggregates used
E.g. disposable income, total household consumption, imputed rent... May be presented as a glossary

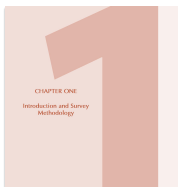
Documentation on sampling design

what to include

- Sampling design report with
 - Allocation of sample into strata and indication of **excluded strata**, if any
 - Estimation **formulas** (selection probabilities and weights)
- Household listings forms
- Sample frames
 - For the first sampling stage/s: list of all **sampling units**
 - For the last sampling stage: list of all **households** in each sample point
- Non-response rates
- On the survey datasets
 - Sampling **weights**

Kenya, 2015

Kenya Integrated Household Budget Survey (KIHBS)



Parameters	2015/16 KIHBS
Sample design	
Survey Domains	National, 47 Counties, Rural/Urban
Sampling Frame	NASSEP V (5,360 Clusters)
Sample Size & Allocation	
National	24,000 Households (2,400 Clusters)
Rural	14,120 Households (1,412 Clusters)
Urban	9,880 Households (988 Clusters)

Kenya, 2015

Kenya Integrated Household Budget Survey (KIHBS)

1.10 Survey response rates

The survey achieved high sample response rates. Nationally, 91 per cent of the sampled households participated and completed questionnaires. As shown in Table 1.2, from 23,852 households that were sampled for the survey, a total of 21,773 households were successfully interviewed. The response rate for rural households was higher (93.6%) compared to that of urban households (88.0%). Part of the non-response was due to non-coverage of 13 clusters spread across different counties occasioned by either insecurity or non-availability of households due to movement of populations in nomadic areas

Table 1.2: Response rates

Result	Residence		Total
	Urban	Rural	
Households selected	9,670	13,982	23,652
Households interviewed	6,681	13,092	19,773
Household response rate	88.0	93.6	91.3

Documentation on fieldwork

what to include

- Training
 - Calendar
 - Quizzes
 - Evaluation forms and selection procedures
- Composition and territorial deployment of the field teams
- Dates of field work
- Problems encountered
- Changes to field procedures
- Supervision forms
- Non-response rates, by interviewer

Uganda, 2016/17

National Household Survey

1.6.3 Fieldwork

A centralized approach to data collection was employed through which 13 mobile field teams grouped at the UBOS headquarters were deployed to the different sampled areas. Each team comprised one field supervisor, three or four enumerators and a driver. The field staff were recruited based on fluency of the local language spoken in the respective region of deployment while the supervisors were balanced between males and females. Prior to the deployment of fieldwork teams, ten listing teams each comprising of a team leader and two listers were constituted to update the number of households within the sampled EAs.

At the headquarters, a team of regional and senior supervisors undertook several other survey activities in line with the survey including data scrutiny, field monitoring, coordination and supervision among others. The field data collection commenced at the end of June 2016 and was completed in June 2017. Fieldwork was carried out in 12 separate trips, between which teams met at the headquarters for refresher training and debriefing sessions. During the meetings, the main issues discussed included logistical and data collection challenges which were resolved instantly.

Documentation on the construction of final databases

What to include

- Number of households included
 - Reasons for exclusions
 - Distribution of households
- Number of individuals included
- Unique identifiers
- How to merge files
- Problems encountered
- Methodology to construct aggregates

Egypt, 2015

Household Income, Expenditure and Consumption Survey (HIECS)

Expenditure Aggregates

Variable name	Variable label/ Expenditure item code	Variable content/ Expenditure item label
PREUDC	Expenses on pre-primary and primary education	Includes also expenses on literacy programs for students too old for primary school, including private tutoring and tutoring groups
	value: 1010101	Pre-school school fees
	value: 1010102	Primary school fees
	value: 1010104	Primary private tuition fees
	value: 1010105	Primary strengthening lessons fees

Other documentation

What to include

- Organizations included in preparation of work
- List of data sets and contents
- Codes not found in the questionnaire
 - Occupation codes
 - Industry codes
- Other information
 - Exchange rates
 - Consumer Price Index
- Supervisor control forms
- Questionnaire control forms
- Maps
- Abstract

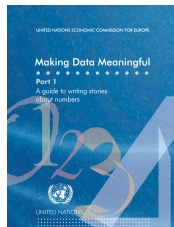
2. Descriptive statistics

How to describe data effectively

- Text
- Tables
- Graphs

Text

- **Effective writing** complements good tables and graphs
- This lecture will focus on the latter: writing deserves a separate discussion
- A useful reference



Tables

- Tables are **omnipresent** in data dissemination reports
- Often used when describing two variables jointly (two-way tables), e.g. income by region, population by age...

Table elements

UNECE (2009: 12)

Row stubs	Column headers
	Data
Notes	
Source	

What makes a good table

Golden rule #1

Express contents clearly

1. The **table title** should answer the questions “what”, “where” and “when”, but still be concise
2. Tables should be **self-contained**: use notes to clarify definitions, abbreviations, etc.
3. Percentage distributions of discrete variables should be clearly identified as either **percentages of households or percentages of the population**
4. **Row and column totals** should be reported, when they identify a marginal distribution

What makes a good table

Golden rule #2

Reduce clutter

1. Avoid unnecessary **colors, repetitions** (e.g. use % or \$ just once, in the title, rather than throughout the table)
2. **Precision of numbers**: do not present too many significant digits. Percentages: one decimal digit is usually enough. Numbers with four or more digits: no decimals at all. Large numbers: express them in thousands or millions
3. Be mindful of **spacing and alignment**

What's wrong with this table?

Final energy consumption by sector - Percentages

	1980	1985	1990	1995	2000	2002	2003
Transport	27.81	27.92	28.24	31.12	36.82	39.48	39.13
Residential	31.11	33.91	30.41	27.61	24.33	23.71	23.97
Industry	31.47	27.21	23.86	22.11	21.41	19.53	18.78
Agriculture	n/a	n/a	3.51	3.7	3.13	2.91	2.82
Services	9.61	10.96	13.98	15.46	14.33	14.37	15.3

Source: UNECE (2009: 12)

Table redesign

Share of total energy consumption, by sector (in percent)
Ireland, 1980-2003

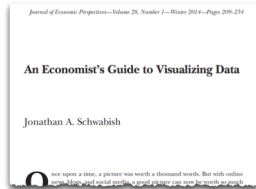
	1980	1985	1990	2000	1995	2002	2003
Transport	27.8	27.9	28.2	31.1	36.8	39.5	39.1
Residential	31.1	33.9	30.4	27.6	24.3	23.7	24.0
Industry	31.5	27.2	23.9	22.1	21.4	19.5	18.8
Agriculture	n/a ¹	n/a ¹	3.5	3.7	3.1	2.9	2.8
Services	9.6	11.0	14.0	15.5	14.4	14.4	15.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

¹ Data on energy consumption for the agricultural sector was not collected until 1990.
Source: Department of Public Enterprise, Ireland

Source: UNECE (2009: 12)

Graphs

- In many cases, presentation of data can be made more interesting and intuitive by using **graphs** or charts rather than **tables**
- Many of the “golden rules” that help make better tables also apply to graphs



What makes a good graph

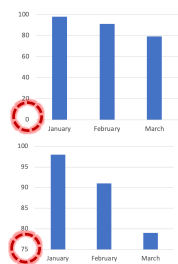
Golden rule #3

Express contents clearly

1. A good **graph title** answers the same questions as a good table title
2. Graphs should be **self-contained** too (use notes)
3. **Explain encoding**: always label axes and data series clearly
4. **Avoid visualizations that mislead the eye**: two notorious “sins” are bar charts with a nonzero baseline, and 3D pie charts

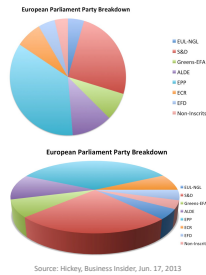
Bar charts with nonzero baseline

- Bar charts rely on bar **length** to show data: compare lengths to compare values
- Shifting the baseline **distorts the visual**: a value twice as high no longer corresponds to a bar twice as long
- Graphs on the right show the same data, but appear very different



3D pie charts

- Pie charts encode data in the **area** of each slice: larger slice equals higher share
- A 3D pie chart **distorts angles**, making the slice that is “closer” to the viewer appear larger than it actually is
- This visualization can **mislead** viewers, and should be avoided



What makes a good graph

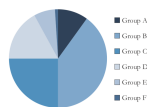
Golden rule #4

Reduce clutter

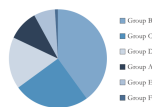
1. Again, avoid unnecessary **colors and decorative elements** that obfuscate the message of the graph
2. **Precision of numbers:** same recommendations as for tables
3. Do not crowd graph with **too many data points**: viewer should be able to understand the message of the graph easily, without having to parse too much visual information (if that is the issue, select a subset of relevant values, or consider using a table instead)

On pie charts

A Pie Chart



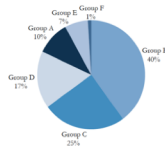
B: A Pie Chart, Rotated



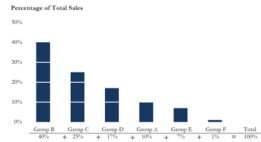
“Because **pie charts force readers to make comparisons using the areas** of the slices or the angles formed by the slices—something that our visual perception does not accurately support — they are not an effective way to communicate information” Schwabish (2014: 223)

Graph redesign

A pie chart, labeled



Pie chart alternative:
a bar or column chart

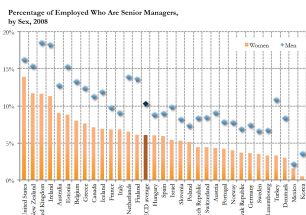


Source: Schwabish (2014: 223)



31

What's wrong with this graph?



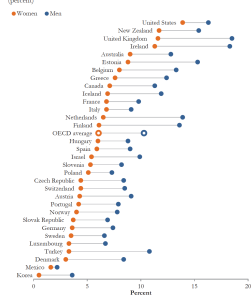
Source: Schwabish (2014: 218)



32

Graph redesign

Percentage of Employed Who Are Senior Managers, by Gender, 2008 (percent)



Source: Schwabish (2014: 220)

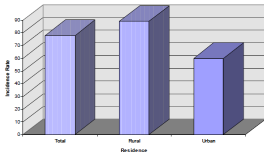


33

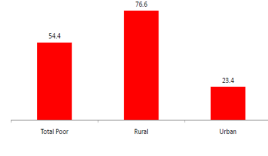
What's wrong with this graph?

Zambia, Living conditions monitoring survey report 1996 and 2015

Incidence of poverty rural/urban, 1996



Incidence of poverty rural/urban, 2015



3. Inequality and poverty

Overview

- Tips for presentation of generic summary statistics still apply
- There are a few additional points to be made specifically about presenting results on poverty and inequality:
 - a. Popular **measures** and graphics (from lectures 13 and 14)
 - b. Best practices for making **comparisons**

Suggested inequality measures

Malawi poverty assessment IHS2, IHS3, Inequality indices

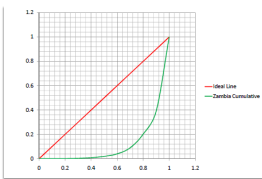
- **Gini** is so prevalent internationally that it **cannot be omitted**, the rest is extra credit

	Malawi		Urban		Rural	
	2004	2010	2004	2010	2004	2010
GE (1)	0.28	0.41	0.48	0.54	0.21	0.28
Theil L (GE(2))	0.25	0.34	0.39	0.41	0.19	0.23
Theil index (GE(1))	0.31	0.42	0.44	0.47	0.20	0.25
GE (2)	0.58	0.96	0.73	0.88	0.29	0.38
Gini	0.39	0.45	0.48	0.49	0.34	0.38

Suggested inequality charts

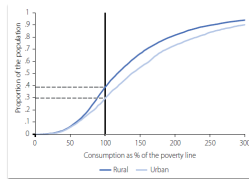
- **Lorenz curve**

Zambia, Living conditions monitoring survey report 2015



- **Overlaid CDFs**

Kenya gender and poverty assessment 2015/16



Suggested poverty measures

2015/16 Kenya Integrated Household Budget Survey (KIHBS)

- **FGT**, the rest is extra credit

Table 4.3: Overall Poverty Estimates (Individual) by Residence and County, 2015/16

Residence / County	Headcount Rate (%)	Distribution of the Poor (%)	Poverty Gap (%)	Severity of Poverty (%)	Population ('000)	Number of Poor ('000)
National	36.1	100.0	10.4	4.5	45,371	16,401
Rural	40.1	71.3	11.5	5.0	29,127	11,687
Peri-Urban	27.5	5.6	6.9	2.6	3,340	920
Core-Urban	29.4	23.1	8.9	3.9	12,905	3,795

Making comparisons

- Many audiences (policy makers, general public) are especially interested in comparisons of poverty and inequality, over time or across regions
- Poverty and inequality **trends** are among the most visible and impactful results to emerge during dissemination
- **Comparability** of underlying data and methods is key: if processes that led up to estimates differ, comparison is **invalid**
- Being **transparent** on comparability is key!

Changes in data and methodology

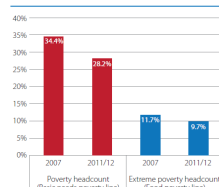
- **Comparability** of data and methods underlying the estimates that are being presented is key
- If processes that led up to estimates differ, comparison is **invalid** and readers may be misled
- **Minimize** incomparability
- If some discrepancies remain, be fully **transparent**

Tanzania, 2012

Poverty Assessment

The analysis of the poverty trend is challenged by changes in the HBS design, but the adjustments made to counter the change in design support the decline of poverty. Assessing the changes in poverty levels over time is subject to issues of comparability stemming from changes in the survey design and methodological improvements implemented during the 2011/12 HBS. These issues were addressed using different methods, including the reevaluation of the consumption aggregates for HBS 2007 using the same approach as in 2011/12, as well as nonparametric and parametric imputation procedures. The different

Figure ES.1 Poverty and Extreme Poverty Incidence



Source: HBS 2007 and 2011/12.

Tanzania, 2012

Poverty Assessment, HBS 2007 and 2011/12 recall modules

Consumption and expenditure categories	HBS 2011/12 Recall period (months)			HBS 2007 Recall period (months)		
	1	3	12	1	3	12
Clothing and footwear (COICOP 3)			X			X
Housing and utilities (COICOP 04 + selected other)						
Rents		X		X		
Utilities	X					X
Energy		X				X
Building maintenance			X			
Housing equipment (COICOP 05)						
Household durables, furniture and furnishings			X			X
Small household appliances		X				X
Expenditures on domestic workers	X					X
Health expenditures (COICOP 06)		X				X

The importance of uncertainty

- Poverty calculations are based on a **sample** of households, and samples carry a margin of error in representing the population
- **Standard errors** should always be estimated along with poverty point estimates
- Crucial when making **comparisons** (over time, across regions): poverty changes should not be taken at the face value
- **Note:** probability weighting, clustering, and stratification, are **survey design features** which must be taken into account when estimating standard errors.

Ethiopia, 2015

Household Income Consumption & Expenditure

Table 9: Poverty indices in 2015/16

	Estimate	Std. Err.	[95% Conf. Interval]
Poverty head count index	0.235	0.008	0.220 0.250
Poverty gap index	0.067	0.003	0.061 0.073
Poverty severity index	0.028	0.002	0.024 0.031
Food poverty head count index	0.248	0.008	0.233 0.263
Food poverty gap index	0.067	0.003	0.061 0.073
Food poverty severity index	0.027	0.002	0.024 0.030

Source: computed from the 2015/16 HICE survey data

Sensitivity Analysis

Bosnia and Herzegovina, 2003 (vol. II)

Report No. 25143-BIH

Bosnia and Herzegovina Poverty Assessment

(In Two Volumes) Volume II: Data on Poverty

November 21, 2003

Poverty Reduction and Economic Management Unit
Europe and Central Asia Region



Sensitivity Analysis

Excerpt from the table of contents

VI. CHECKS FOR ROBUSTNESS OF POVERTY FINDINGS.....	45
A. Robustness Checks with Respect to Equivalence Scales.....	45
(i) Location and Poverty.....	46
(ii) Poverty by Displacement Status.....	47
(iii) Education of the Household Head.....	47
(iv) Employment Status.....	48
(v) Household Size.....	51
B. Robustness Checks Using Alternative Poverty Lines.....	53
(i) Location and Poverty.....	53
(ii) Poverty by Displacement Status.....	54
(iii) Education of the Household Head.....	54
(iv) Employment Status of Adults.....	55
(v) Household Size.....	58
C. Robustness Checks Using Alternative Definitions of Well-Being.....	58
(i) Location and Poverty.....	60
(ii) Poverty by Displacement Status.....	61
(iii) Education of the Household Head.....	61
(iv) Employment Status of Adults.....	62
(v) Household Size.....	64
D. Conclusions.....	64

Sensitivity Analysis

Conclusions

Table 6.2 Key characteristics of poverty and its robustness to measurement assumptions.

Characteristics of poverty	Baseline consumption per capita	OECD I scale	OECD II scale	Higher poverty line	Lower poverty line	Expenditure per capita
Mixed (semi-urban) municipalities in RS	yes	yes	yes	yes	yes	yes
Rural municipalities in FBiH	yes	yes	no	no	yes	no
IDPs and Refugees	yes	yes	yes	yes	yes	yes
Households headed by persons with low education (primary or less)	yes	yes	yes	yes	yes	yes
Households headed by persons with education above secondary	no	no	no	no	no	no
Unemployed (ILO) and inactive adults	yes	yes	yes	yes	yes	yes
Employed according to registration	no	no	no	no	no	no
Registered/unemployed	yes	yes	yes	yes	yes	yes
Household headed by elderly	no	no	no	no	no	no
Larger households	yes	yes	no	yes	yes	yes

Source: Staff estimates based on BIH-LSMS 2001.



Lessons learned

1. **Explain clearly** what your table or graph is showing (titles, labels, notes...)
2. Only if point 1 is checked, **reduce clutter**: keep frills to a minimum
3. When showing results on **inequality and poverty**, include the shortlist of key measures and graphs indicated in this lecture, which the international community has come to expect
4. Comparisons: **document** changes in data and methodology, and include measures of **uncertainty** of estimates whenever possible

References

Required readings

Giovine, P., and Levin, M. (2005). Presenting simple descriptive statistics from household survey data. In UN, Household Sample Surveys in Developing and Transition Countries. Studies in Methods Series F No. 96

Schwabish, J. A. (2014). An economist's guide to visualizing data. *Journal of Economic Perspectives*, 28(1), 209-34.

Suggested readings

Dupriez, O., & Boyko, E. (2010). Dissemination of Microdata Files: Principles, Procedures and Practices. International Household Survey Network.

UNICEF (2009). Making Data Meaningful, Part 1: A Guide to writing stories about numbers. United Nations, Geneva.

UNICEF (2009). Making Data Meaningful, Part 2: A Guide to presenting statistics. United Nations, Geneva.

Suggested websites and blogs

(tutorials, tips, data visualization examples)

Alberto Cairo: thefunctionalart.com

Stephanie Evergreen: stephanieevergreen.com

Nathan Yau: flowingdata.com

Jonathan Schwabish: policyviz.com

Thank you for your attention

Homework

EX. 1 – Engaging with the literature

- The dissemination of microdata often (but not always) accompanies the dissemination of findings and summary statistics from a survey
- Summarize the discussion of the pros and cons of data dissemination in Dupriez et al. (2010) p. 16-23
<https://libsonline.org/default/files/resources/HSN-WP005.pdf>



Exercise 2 - Standard Errors

Poverty analysis of the integrated household survey in The Gambia 2003 (p49-50)

- Compare point estimates with interval estimates, assuming a 95% confidence level, and briefly comment on results

Table 5: Poverty by area

Poverty	Area	Estimate	Std. Err.
Head count index	Urban	33.4	3.0
	Rural	60.6	2.8

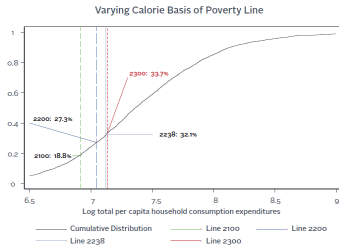
Table 7: Poverty by strata

Poverty	Strata	Estimate	Std. Err.
Head count index	Banjul Urban	6.6	4.0
	KMC Urban	32.1	4.0
	Brikama Urban	41.9	9.4
	Brikama Rural	50.1	4.8
	Manakonia Urban	65.6	12.3
	Manakonia Rural	55.9	8.3
	Kerewan Urban	42.7	10.3
	Kerewan Rural	67.1	5.9
	Kuntaur Urban	57.1	9.2
	Kuntaur Rural	91.9	3.6
	Jarijangbureh Urban	53.0	22.4
	Jarijangbureh Rural	63.0	8.4
	Basse Urban	44.4	7.9
	Basse Rural	63.3	8.0

Exercise 3 - Sensitivity analysis

Myanmar Poverty and Living Conditions Survey 2015

- Briefly comment on the robustness of the poverty line to different calorie norms.



Acknowledgments

The course was prepared by Giovanni Vecchi (U. of Rome Tor Vergata) and Giulia Mancini (U. of Rome Tor Vergata), with a core team comprising Nicola Amendola and Sédi Anne-Boukaka (U. of Rome Tor Vergata).

The work was conducted under the supervision of Gero Carletto, Michelle Jouvenal, Shelton Kanyanda, and Alberto Zezza (World Bank), who provided comments throughout the drafting process.

The team relied on advice and comments from many colleagues, who also contributed their own teaching material – Giovanni D'Alessio, Romina Gambacorta, Giuseppe Ilardi, Valentina Michelangeli, Andrea Neri (Bank of Italy), Gero Carletto, Dean Jolliffe, Shelton Kanyanda, Talip Kilic, Heather Moylan, Diane Steele, Alberto Zezza (World Bank), Federico Polidoro, Paolo Consolini, Valeria De Martino, Gabriella Donatiello (Istat), Piero Conforti (FAO).
