USAID
FROM THE AMERICAN PEOPLE

USAID Monitoring, Evaluation and Learning Activity
Quantitative Methods and Analysis

January 2023

Training Agenda

- Section 1: Quantitative Data and Methods
- Section 2: Sampling and Data Cleaning
- Section 3: Quantitative Analysis

- At the end of this training, you will:
  - Be familiar with the different quantitative data types and methods
  - Be able to choose an appropriate sampling strategy
  - Understanding the fundamentals of data cleaning
  - Be able to use descriptive statistical techniques
  - Be familiar with examples of the use of inferential statistical techniques
  - Be able to read quantitative outputs from select statistical outputs
  - Be familiar with data visualization approaches

Section 1

USAID
FROM THE AMERICAN PEOPLE

Key Ethical Principles

- Respect for persons/autonomy
- Beneficence
- Justice
- Risk

Ethics in Practice

- While U.S. federal regulations only cover specific practices for human subject research as opposed to project/activity monitoring and evaluation (see 22 CFR 225 for USAID) , it is expected practice to apply these standards across any data collection involving human participants as applicable.
- For privacy and protection of personal identifiable information (PII), research/evaluation studies should be compliant with USAID ADS 508.
- This includes:
  - Collecting informed consent/assent prior to data collection
  - Gathering only what data is needed
  - Protecting privacy, anonymity and confidentiality
  - Engaging with Institutional Review Boards/Ethics Review Committees (when appropriate)

In your own words:

- What is quantitative data?
- What questions can it answer?

What are Quantitative Data?

- What are quantitative data?
  - Things that can be counted or expressed numerically
  - Analyzed using statistical methods
  - Best used to answer what, when and who questions
  - Not well suited to how and why questions

Strengths	Limitations
Findings can be generalized, if selection process is well-designed and sample is representative of study population	Related secondary data sometimes not available, or accessing available data is difficult/impossible
Relatively easy to analyze	Difficult to understand context or program activities
Data can be very consistent, precise, reliable	Data may not be robust enough to explain complex issues
Cost and time	

Data/Variable Types

Variable Types	Categorical	Scale
	Nominal: values vary in categories; not possible to rank the categories created.	**Interval**: scale uses the same interval between one measurement and the next (but the zero point is arbitrary).
	Ordinal: possible to rank the categories or put them in an order. The intervals between the categories used are not defined.	**Ratio**: same as an interval variable but there is a true zero point

Discrete: Values that fall under integers or whole numbers	**Continuous:** Values that contain fractions or decimals

Data/Variable Types

Data Designs
Cross section: different subjects at the same time
Time series: the same subject at different times
Panel: many subjects at different times

Data/Variable Types

Primary (data collected by the evaluation team for a specific purpose)

- Survey
  - Self-administered or by someone else
  - Face-to-face, telephone, mail, web-based
- Structured observations

Secondary (data collected by someone else for some other purpose, but being utilized by the evaluation team)

- Indicator data
- Census data, national level survey data (LSMS), studies, criminal justice statistics, performance data, etc.
- Documents/texts

Data Collection Protocol

- Every study/evaluation should have a data collection protocol that should contain the following:
  - Purpose and questions to be answered
  - Description of the methodology/sampling approach
  - Instructions for how data will be protected, and informed consent will be gained
  - All data collection tools

Reducing Measurement Error

- Pilot instruments and incorporate feedback
  - Questionnaire bias
  - Interviewer bias
  - Question ordering
  - Social factors/ environment
- Thoroughly train enumerators and test them in the course of training
- Provide your enumerators with necessary tools for recording the data and assessing answers' accuracy
- Double check all data
- Include questions that allow ex post identification of different types of error of measurement
- Increasing the size of the sample will minimize the magnitude of the random error

Data Quality Standards

- Validity: Data should represent the intended result clearly and adequately.
- Reliability: Data should reflect stable and consistent data collection processes and analysis methods over time.
- Accuracy: Data should have a sufficient level of detail to permit informed management decision making.
- Integrity: Data should have safeguards to minimize risk of bias, transcription error, or data manipulation.
- Timeliness: Data should be available at a useful frequency, should be current, and should be timely enough to influence management decision making.

Group Activity: Case Study

- Please read through the case study and the proposed questions.
- Discuss the questions and proposed answers with your group.
- Please assign one person to report out what you discussed.

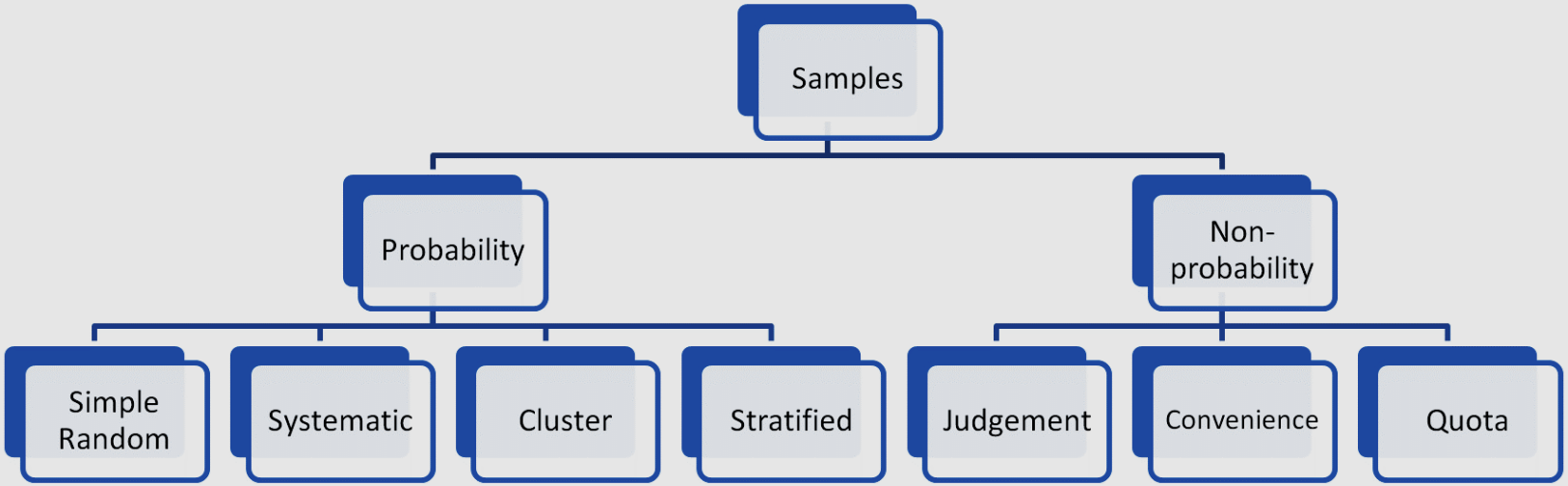
Section 2

When you think about sampling….what comes to mind?

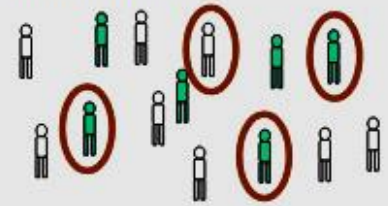
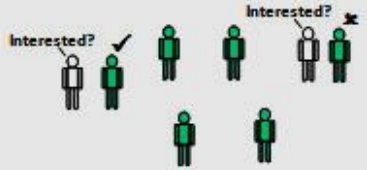
Sampling Strategy

- **Why sample?**
- Get information about large populations
- The primary goal of sampling is to get a representative sample, or a small collection of units or cases from a much larger collection or population, such that the researcher can study the smaller group and produce accurate generalizations about the larger group.
- Elements
  - Sampling technique
  - Sampling frame
  - Sample
- Sampling is an art as much as it is a science!

Sampling Techniques



Sampling Techniques

Random sampling 	Every member of a population has an equal chance of being selected E.g. Pulling names out of a hat	For very large samples it provides the best chance of an unbiased representative sample	For large populations it is time-consuming to create a list of every individual.
Stratified sampling 	Dividing the target population into important subcategories Selecting members in proportion that they occur in the population E.g. 2.5% of British are of Indian origin, so 2.5% of your sample should be of Indian origin... and so on	A deliberate effort is made to make the sample representative of the target population	It can be time consuming as the subcategories have to be identified and proportions calculated
Volunteer sampling 	Individuals who have chosen to be involved in a study. Also called self-selecting E.g. people who responded to an advert for participants	Relatively convenient and ethical if it leads to informed consent	Unrepresentative as it leads to bias on the part of the participant. E.g. a daytime TV advert would not attract full-time workers.
Opportunity sampling 	Simply selecting those people that are available at the time. E.g. going up to people in cafés and asking them to be interviewed	Quick, convenient and economical. A most common type of sampling in practice	Very unrepresentative samples and often biased by the researcher who will likely choose people who are 'helpful'

An activity team goes to a hospital to survey patients on the care they received. They survey the first fifty patients they see.

Probabilistic or non-probabilistic?

An activity team is conducting interviews over the phone. In order to meet their required numbers, they use recommendations from participants to determine who to interview.

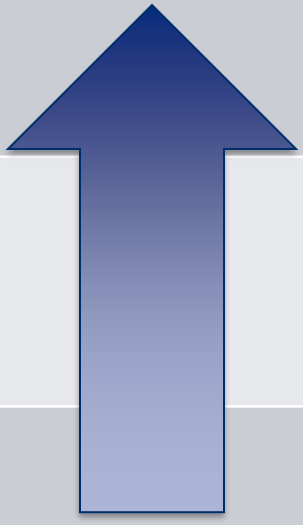
Probabilistic or non-probabilistic?

A team is surveying youth participating in a workforce development activity. They assign each youth a number using a random number generator and survey the first 50 youth identified.

Probabilistic or non-probabilistic?

Design Types

Type of Outcome Evaluation Design	Control or Comparison	Ability to produce causal evidence about a program
Experimental Design	• Control group • Random assignment	
Quasi-Experimental Design	• Control group • Non-random assignment	
Non-Experimental Design	• No control/ comparison group(s)	



Null Hypothesis

- In evaluations and some studies, we are trying to prove that an activity/intervention actually had its intended effect.
- The null hypothesis, typically noted as H0, is that nothing has changed.
- The statement that the evaluation achieved the desired effect is noted as H1.
- To do this, we used statistical methods to reject the null hypothesis.

Statistical Significance

- In order to reject the null hypothesis, we need a sample that is large enough to allow us to avoid two types of mistakes.

- The first is called a Type I error. This is when you conclude that there is a change or an effect when there actually is none. In other words, you reject the null hypothesis when you shouldn't have.

- To avoid these, statistical tests look for **significance**, a concept that measures the degree to which your results can be obtained due to chance.

- In social science/educational research the term $\alpha = .05$ is often used. This means there is a 5% or less chance that the results are due to chance.

- **A significance level of 5%** is the same as a **confidence level of 95%.**

Statistical Power

- The second kind of mistake is when you determine the program has no effect, when it actually does have an effect. You fail to reject the null hypothesis when you should have. This is called a **Type II error.**

- To avoid this, we also sample for statistical power or the probability that one will be able to find a significant effect.

- Commonly, power is set at 80%. This means that there is an 80% probability of correctly concluding that there is an effect/change when in fact, that change/effect does exist.

- The larger of the sample, the larger the power.

Effect Size

- The final element to calculate a sample size is the effect size. The effect size is essentially the size of the difference between groups.
- It can be calculated a number of different ways. One way is Cohen's d.
- Small=0.2, Medium= 0.5, Large= 0.8
- The smaller the effect size, the larger the necessary sample. Otherwise, the results will not be statistically significant.
- This comes from literature or previous studies/activities.

Source: Sullivan and Feinn. Using Effect Size—or Why the P Value Is Not Enough. Journal of Graduate Medical Education. September 2012.

Other factors in sampling

- Disaggregates/sub-populations
  - The more disaggregates the study/evaluation has where statistical significance is needed, the larger the sample needs to be
- Budget
- Attrition
  - Need to adjust for drop-outs from the study or evaluation.

Sampling Frame and Sample

- Once you have calculated your sample size, you should create a sampling frame. This is a list of all the potential participants who could be included in the sample.

- Apply your sampling strategy and technique accordingly.

Group Activity: Sampling Exercises

- Please complete the following exercise in small groups.

- We will come back together and discuss as one big group once you all have completed the exercise.

Source: USAID Learning Lab

Any remaining questions on sampling?

Data Cleaning

- This is a critical step. You cannot analyze data without a clean dataset.
- Dirty, messy data will cause you to have incorrect results in your analysis.
- Data cleaning takes significant time to do correctly and well.
- Under ADS 579, activities are required to submit cleaned datasets to the Data Development Library (DDL)

Wide vs Long Data Format

Wide Format

Team	Points	Assists	Rebounds
A	88	12	22
B	91	17	28
C	99	24	30
D	94	28	31

Long Format

Team	Variable	Value
A	Points	88
A	Assists	12
A	Rebounds	22
B	Points	91
B	Assists	17
B	Rebounds	28
C	Points	99
C	Assists	24
C	Rebounds	30
D	Points	94
D	Assists	28
D	Rebounds	31

Source: Statology.org/long-vs-wide-data/

Missing Data

- Important to distinguish between a true "0" value and a missing value
- Software like STATA and SPSS will often ignore missing values in certain functions which can warp analysis.
- Best to code with a value that would not be confused with an expected value, i.e. "-99"

Outliers

- Outliers can warp analysis, especially when they are a result of data entry error.
- If you see a value that is beyond the expected range of results, you should check with your data collectors to ensure that value is correct.
  - It may result in needing to go back to the respondent to get the correct value.
- If it is not possible to confirm the value, it may be justified in removing the value and replacing it as a missing value.

Data Format

- Having the correct and consistent data format is key to ensuring that you can conduct quantitative analysis.
- Data that appears to be numeric in nature could actually be stored as text data or string data.
- If data is text or string, software will not recognize it as a number and will not run any quantitative functions on it.
- In Excel, you can use the Convert to Number option or change the cell format to General/Number.
- STATA and SPSS have functions to convert string data to numeric data

Multiple Selection Questions

- When a survey question asks respondents to mark all that apply or pick a number choices, you will get a number of datapoints inside one cell.

- In these cases, you will need to recode each response as a separate column.

Group Activity

- I am going to share this raw dataset on the screen.

- We are going to walk through this as a group and point out the spots where we need to clean the data.

Section 3

What is quantitative data analysis?

- The search for patterns in raw data and for explanations for those patterns.
  - Descriptive: Describe aspects about a group of numbers
    - Consists of the collection, organization, classification, and presentation of data obtained from a sample.
    - do not allow us to make conclusions beyond the data we have analyzed or reach conclusions regarding any hypotheses we might have made
  - Inferential: techniques that allow us to use these samples to make generalizations about the populations
    - Associations/ relationships between variables.
    - Can be used to explain or predict these relationship

Frequency Tables

Class rank

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Freshman	147	33.8	36.2	36.2
	Sophomore	96	22.1	23.6	59.9
	Junior	98	22.5	24.1	84.0
	Senior	65	14.9	16.0	100.0
	Total	406	93.3	100.0	
Missing	System	29	6.7		
Total		435	100.0		

Source:
https://libguides.library.kent.edu/SPSS/FrequenciesCategorical

Crosstabs

```
. table sex grade, c(mean intrinsi)
```

Sex (0=F; 1=M)	Grade Level (6th, 7th, or 8th) 6	7	8
0	3.4221078	3.2818627	3.1474024
1	3.3608026	3.3354405	3.1539081

Source: www.reed.edu/psychology/stata/gs/tutorials/tables.html

Descriptive Metrics to measure the center value

- **Mode:** The most frequent value in a set of data.
  - Ex. If you take the data set, (3,4,5,6,3,3,3,3,7,7,8), the mode would be 3.
- **Median:** The middle value of the dataset.
  - If N is an odd number, you take the middle value.
    - Ex. If you take the data set, (3,3,3,3,3,4,5,6,7,7,8), the median would be 4. This is the 6^{th} value.)
  - If N is an even number, you sum the two middle values and divide by two.
    - Ex. If you take the data set, (3,3,3,3,3,3,4,5,6,7,7,8), the median would be 3.5 or (3+4)/2)

Descriptive Metrics-Center Value (Continued)

- Mean or Average: Sum of the dataset divided by the total number of value in the data
  - Ex. In the dataset (3,3,3,3,3,3,4,5,6,7,7,8), you sum all of the values and divide by total number. So it would be (3+3+3+3+3+3+4+5+6+7+7+8)/12= approx. 4.58

Descriptive Metrics- Variation

- Range: The difference between the largest and smallest values in a dataset.
  - Ex. In the dataset (3,3,3,3,3,3,4,5,6,7,7,8), the range would be 8-3=5.

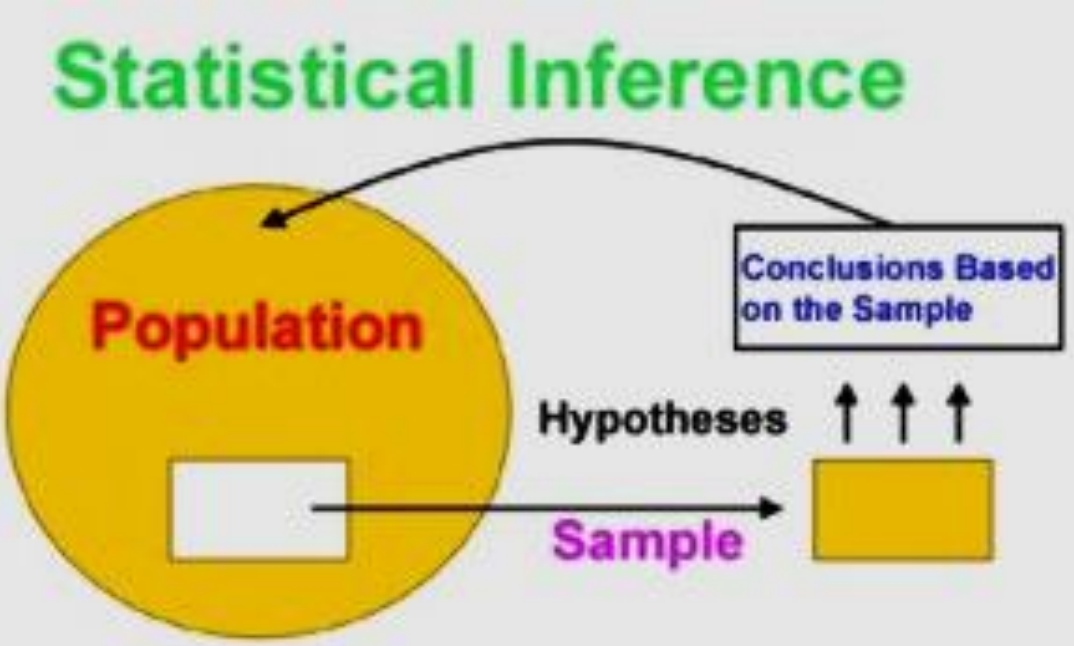
- Standard variation: A measure of the amount of variation of a set of values.
  - A lower standard deviation means that values are close to the mean. A higher standard deviation means that values are spread out.

Descriptive Statistics in Excel

- Mode: =MODE.MULT(A:A)
- Median: = MEDIAN(A:A)
- Mean/Average: =AVERAGE(A:A)
- Range: = MAX(A:A)-MIN(A:A)
- Standard Deviation: = STDEV.S(A:A)

Any questions on descriptive statistics?

Inferential Statistics



Key to using Inferential Statistics

- Understanding the question you want to answer and using the appropriate statistic test.
- Understanding what is possible given your data.

What statistical test would be useful?

- You are conducting an evaluation with a treatment and control group. The evaluation is looking at whether the treatment group has a greater amount of income after participating in the activity. You want to know if the treatment group does indeed have higher income at endline.

T-Tests

- Two sample student t-test
  - used to determine whether or not two populations means are equal
  - Null hypothesis
    - The two population means are equal
  - Alternative hypothesis
    - Two tailed: the two population means are not equal
    - One tailed: The first population mean is greater than or less than the second population mean.

Source: www.statology.org/two-sample-t-test/

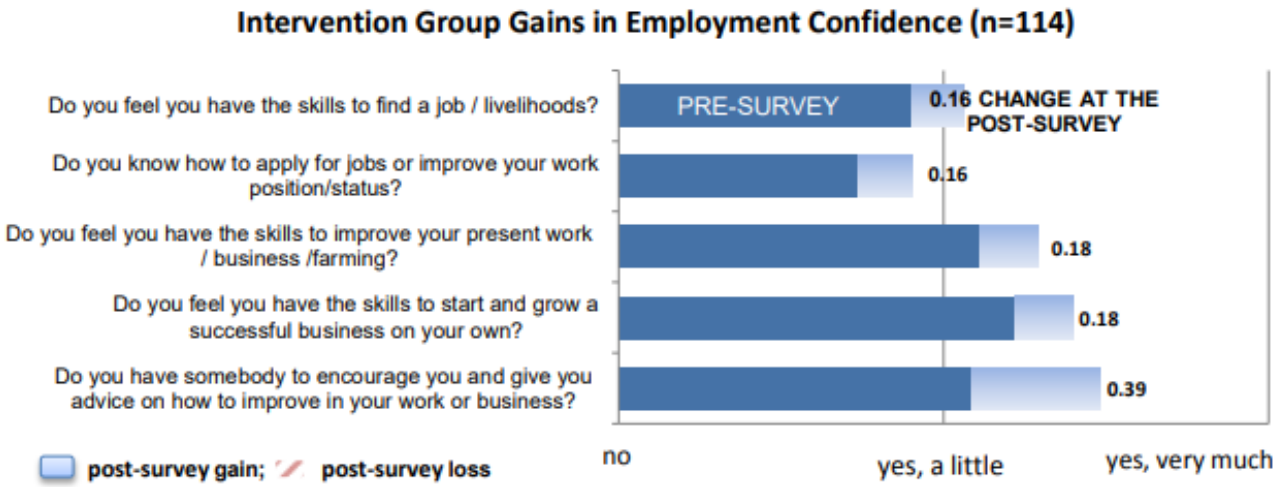
T-Tests

- Assumptions
  - Observations in one sample should be independent of the observations in the other sample.
  - Data should be approximately normally distributed.
  - Data in both samples should have been obtained using random sampling.
  - The two samples should have roughly the same variance.

Source: www.statology.org/two-sample-t-test/

T-Tests

Figure 8. Intervention group shows significantly larger gains on confidence questions



Source: USAID. USAID Advancing Youth Project: Livelihoods and Work Readiness Project. 2014.

T-Tests

- Weaknesses
  - Must meet assumptions in order for the analysis to work
  - Need to interpret the difference in means
  - Heavily reliant on means
  - Does not control for anything else

What statistical test would be useful?

- In evaluating an agricultural activity, you are interested in determining the relationship between the amount of household income and the number of hectares a household owns.

Pearson Correlation

- Assumptions
  - Two or more continuous variables
  - Linear relationship between the variables
  - Independent cases
  - Normally distributed variables
  - Random sample

Source: libguides.library.kent.edu/spss/pearsoncorr

Pearson Correlation

Table 10. Pearson correlations for EGMA subtasks

	Number identification	Quantity discrimination	Missing number	Addition fluency	Addition level 2	Subtraction fluency	Subtraction level 2	Word problems
Number identification	1.00							
Quantity discrimination	0.57***	1.00						
Missing number	0.55***	0.65***	1.00					
Addition fluency	0.57***	0.55***	0.58***	1.00				
Addition level 2	0.44***	0.45***	0.51***	0.59***	1.00			
Subtraction fluency	0.51***	0.50***	0.53***	0.69***	0.53***	1.00		
Subtraction level 2	0.36***	0.34***	0.41***	0.44***	0.63***	0.51***	1.00	
Word problems	0.30***	0.36***	0.40***	0.42***	0.43***	0.45***	0.42***	1.00

*$p < .05$, **$p < 0.01$, ***$p < .001$.

Source: RTI International: The Primary Math and Reading (PRIMR) Initiative: Endline Impact Evaluation. 2014.

Pearson Correlation

- Weaknesses
  - Not controlling for any other variables
  - Requires normally distributed data

What statistical test would be useful?

- You are evaluating an education activity focused on early grade reading. You want to understand the relationship between test performance and a number of different variables including participation in the education activity.

Regression

- Method to understand the relationship between variables.
  - One variable, noted as y, is the response or dependent variable. This would be your variable of interest like test performance.
  - Another variable, x, is the predictor or independent variable.
  - Depending on the relationship between x and y, you can use a regression formula to quantify the relationship between the two variables.
    - In the case of a linear relationship, the basic regression formula looks like this:
      - $Y=B_0+B_1X$
    - If you have multiple variables that have linear relationship, the regression formula looks like:
      - $Y=B_0+B_1X_1+B_2X_2+\ldots\ldots+B_pX_P + e$

      Source: www.statology.org

Regression

Item	Model	Measure	Coefficient	*T*	*p*-value	R^2
	Total teacher visits LCPS	Estimate	-.004 (.016)	-0.23	.82	.000
		Constant	61.33	4.86	<.001	
	Average visits per teacher Public	Estimate	-1.69 (1.68)	-1.01	.32	.003
		Constant	39.20 (4.86)	8.07	<.001	
	Average visits per teacher LCPS	Estimate	.853 (.466)	1.83	.07	.004
		Constant	50.17 (5.84)	8.59	<.001	
Kiswahili fluency	Total teacher visits Public	Estimate	.052 (.022)	2.38	.02	.009
		Constant	16.55 (2.57)	6.44	<.001	
	Total teacher visits LCPS	Estimate	-.006 (.008)	-0.74	.46	.001
		Constant	34.41 (2.44)	14.08	<.001	
	Average visits per teacher Public	Estimate	1.35 (.89)	1.51	.14	.004
		Constant	19.21 (2.37)	8.10	<.001	
	Average visits per teacher LCPS	Estimate	.372 (.214)	1.74	.09	.003
		Constant	28.25 (2.75)	10.26	<.001	

Source: RTI International: The Primary Math and Reading (PRIMR) Initiative: Endline Impact Evaluation. 2014.

Regression

- Regressions create models. If you do not have enough information to create an effective model, you may have a lot of error which will hamper your analysis.

Remaining questions on inferential statistics?

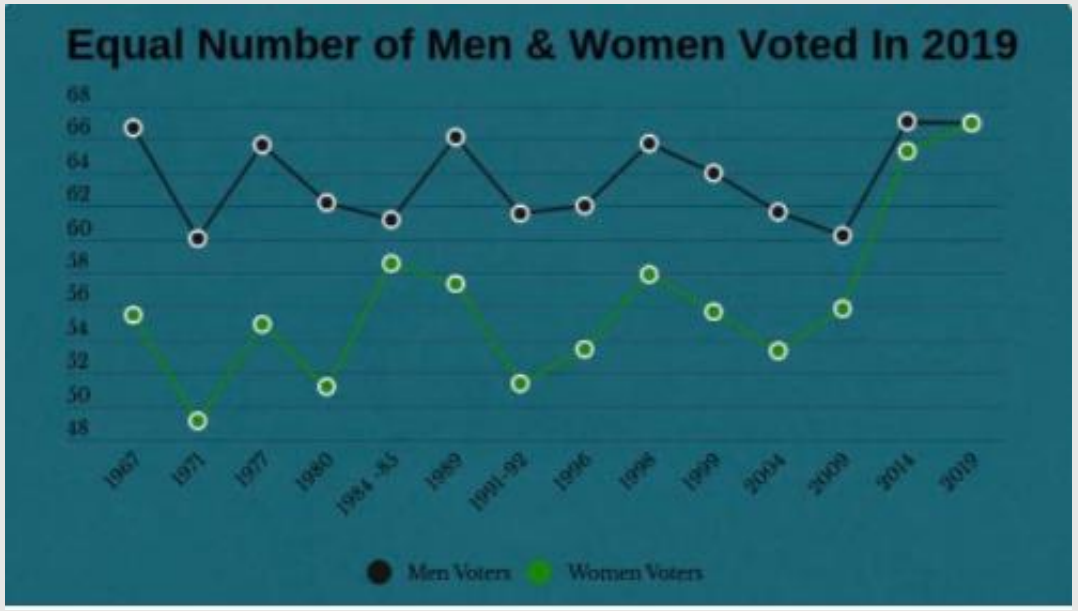
Data Visualization

What is wrong with this data visualization?

Source: www.espncricinfo.com/story/which-top-cricket-city-would-win-the-world-cup-1196522

Data Visualization

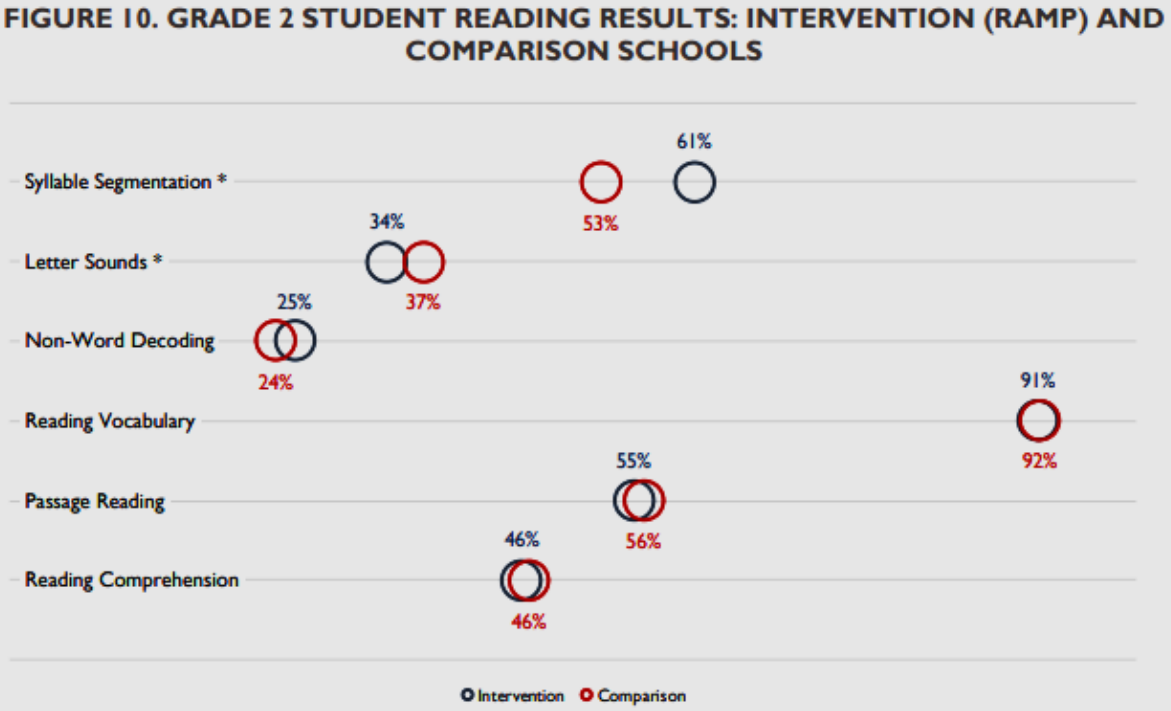
What is wrong with this data visualization?



Source: https://badvisualisations.tumblr.com/post/185213502001/if-we-could-see-past-the-colour-scheme-this-might

Data Visualization

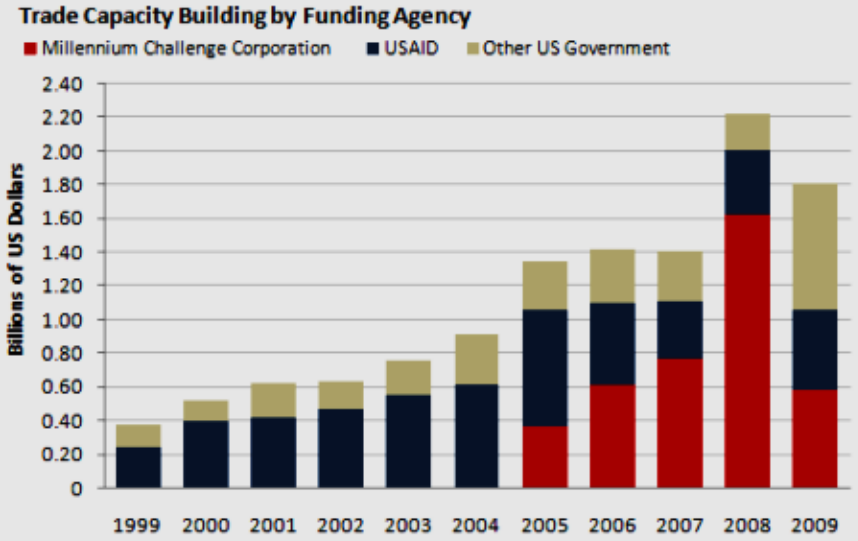
What works with this visualization?



Source: USAID. RAMP Impact Evaluation Final Report, Sept. 2019)

Data Visualization

What works with this visualization?



Source: U.S. Trade Capacity Building Database

Source: USAID. From Aid To Trade: Delivering Results: A Cross-Country Evaluation of USAID Trade Capacity Building. November 2010.

Group Exercise

- Discuss the sample graphs in small groups and discuss potential changes to them.

- We will come back and share the suggestions that you have made.

Remaining questions?

Thanks all!

- If further questions, please send them to Rasha Akkad at rakkad@jordanmela.com