IPSI

Module 2.3

Describing and communicating data effectively: Categorical data

Trainer:.....

Module objectives

At the end of this module, participants will be able to:

1. Describe categorical data using measures of frequency, central tendency and variation
2. Frame questions that can be answered by available district health systems data
3. Apply appropriate data visualization techniques to answer questions about district health system performance

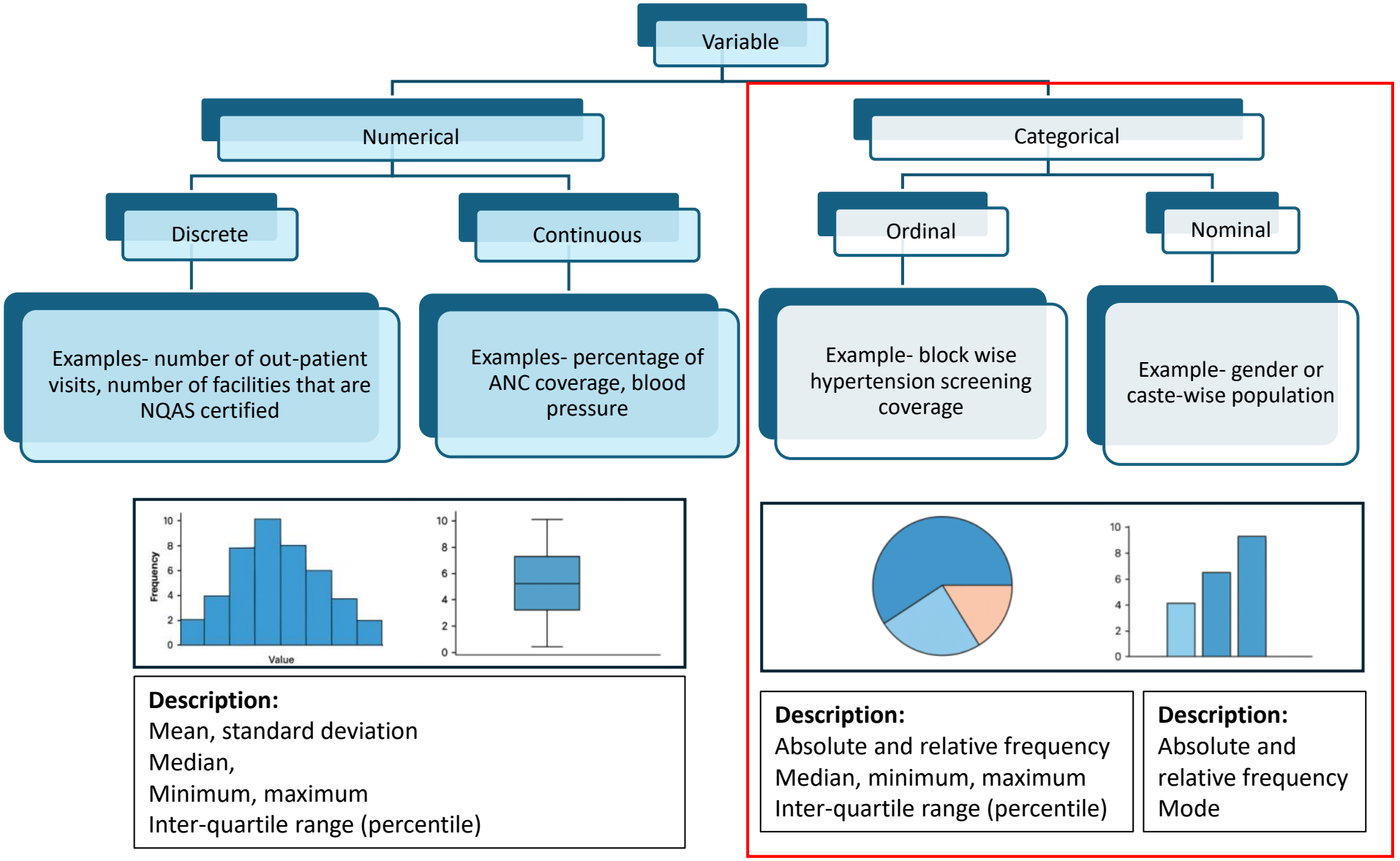
Module outline

Sr. No.	Session name	Session format	Time
1	Analysing and visualizing categorical data Data analysis and visualization techniques • Tables	Facilitator led exercise and discussion Group exercise	40 min
2	• Bar graphs • Pie charts		40 min

Session 1:
Analysing and visualizing categorical data- Tables

Duration: 40 min

Visualizations according the data variable type



Categorical data:

- Tables
- Bar charts
- Pie charts

Adapated from- Spriestersbach A, Röhrig B, du Prel JB, Gerhold-Ay A, Blettner M. Descriptive statistics: the specification of statistical measures and their presentation in tables and graphs. Part 7 of a series on evaluation of scientific publications.

Tables

Sl No	Block	ANC Registered
1	A	2118
2	B	691
3	C	1417
4	D	627
5	E	1625
6	F	639
7	G	1701
8	H	1323
9	I	2169
10	J	1639
11	K	1453
12	L	134
13	M	1043
14	N	1589
15	O	1735
District		**19903**

Data source: HMIS Portal, District X

Questions:

- **What is the number of ANC registrations in each block of district X?**
- **Which blocks have the highest/lowest ANC registration in the district?**
  - Tables are one of the commonest methods of presenting data.
  - The table shown here shows the number of antenatal care (ANC) registrations over a nine-month period for each block in the district X.
  - In presentations, it is generally not a good idea to show tables because it draws the audience attention to the numbers and makes it difficult to quickly understand the overall pattern or point being made.

Tables

Sl No	Block	ANC Registered
1	A	2118
2	B	691
3	C	1417
4	D	627
5	E	1625
6	F	639
7	G	1701
8	H	1323
9	I	2169
10	J	1639
11	K	1453
12	L	134
13	M	1043
14	N	1589
15	O	1735
District X		**19903**

Data source: HMIS Portal, District X

Questions:

- **What is the number of ANC registrations in each block of district X?**
- **Which blocks have the highest/lowest ANC registration in the district?**

The way the information is presented here is not very meaningful - Why?

- It is not easy to see which block did the best or worst.
- Does not tell us how each block performed relative to another block.

Tables

Sl No	Block	ANC Registered
1	L	134
2	D	627
3	F	639
4	B	691
5	M	1043
6	H	1323
7	C	1417
8	K	1453
9	N	1589
10	E	1625
11	J	1639
12	G	1701
13	O	1735
14	A	2118
15	I	2169
District X		**19903**

Data source: HMIS Portal, District X

Questions:

- **What is the number of ANC registrations in each block of the district?**
- **Which blocks have the highest/lowest ANC registration in the district?**

One easy way to present this information more meaningfully is to organize the table in ascending or descending order.

- Now you can easily see which block did the best, and which did the worst in terms of total ANC registrations.

Tables

Sl No	Block	ANC Registered
1	L	134
2	D	627
3	F	639
4	B	691
5	M	1043
6	H	1323
7	C	1417
8	K	1453
9	N	1589
10	E	1625
11	J	1639
12	G	1701
13	O	1735
14	A	2118
15	I	2169
District X		**19903**

Data source: HMIS Portal, District X

Questions:

- **What is the number of ANC registrations in each block of the district?**
- **Which blocks have the highest/lowest ANC registration in the district?**

  - Another way to meaningfully show block performance is by creating a heat map. The colored column shows the ANC registrations where the lower values are in red and and higher ones in green – this makes it easier to see which blocks have high/low ANC registrations.

Tables

Sl No	Block	ANC Registered
1	L	134
2	D	627
3	F	639
4	B	691
5	M	1043
6	H	1323
7	C	1417
8	K	1453
9	N	1589
10	E	1625
11	J	1639
12	G	1701
13	O	1735
14	A	2118
15	I	2169
District X		**19903**

Data source: HMIS Portal, District X

Question:

- **Which are the good and poor performing blocks in the district?**

One way to assess block performance is by identifying blocks that performed higher and lower than average.

- The average number of ANC visits for this period is 1327. Here the blocks that performed below average are color coded in red.

Tables

Block	ANC Registrations
I	2169
A	2118
O	1735
G	1701
J	1639
E	1625
N	1589
K	1453
C	1417
H	1323
M	1043
B	691
F	639
D	627
L	134

Data source: HMIS Portal, District X

Question:

- **Which are the good and poor performing blocks in the district?**
  - A more informative way to examine block performance is to divide the data into quantiles and color code them.
  - In the example shown here, blocks are grouped in quartiles i.e. groups of 25% (remember the exercise on quartiles in the previous module)
  - Here the blocks in the lowest quartile i.e. lowest 25% of blocks in terms of ANC registrations are shaded in red, the next highest 25% in yellow, the next highest 25% in green, and the highest 25% of blocks are shaded in blue.
  - By grouping in quartiles one can see which group of blocks are the top performers, which are mid-level performers, and which are poor performers.

Tables

Sl No	Block	ANC Registrations	Total Estimated Pregnancies
1	A	2118	2176
2	B	691	840
3	C	1417	1488
4	D	627	670
5	E	1625	1658
6	F	639	644
7	G	1701	1760
8	H	1323	1372
9	I	2169	2280
10	J	1639	1642
11	K	1453	1482
12	L	134	140
13	M	1043	1060
14	N	1589	1690
15	O	1735	1774
District X		**19903**	**20676**

Data source: HMIS Portal, District X

Question:

- **Which blocks have the highest/lowest ANC registration in the district?**

  - Identifying high/low performing blocks based on absolute number of ANC registrations is not correct.
  - A more appropriate metric of performance is the percentage of pregnant women in each block who have had an ANC registration (using appropriate denominator)
  - This is because each block will have different populations and so have different numbers of women who are pregnant.
  - For this we will need information on the denominators i.e. the number of women who were pregnant in this time period.

Tables

Block	ANC Registered	Rank ANC Registered	Estimated Pregnancies	Pregnancies Registered (%)	Rank Pregnancies Registered
I	2169	1	2280	95	12
A	2118	2	2176	97	7
O	1735	3	1774	98	6
G	1701	4	1760	97	8
J	1639	5	1642	100	1
E	1625	6	1658	98	5
N	1589	7	1690	94	13
K	1453	8	1482	98	4
C	1417	9	1488	95	11
H	1323	10	1372	96	9
M	1043	11	1060	98	3
B	691	12	840	82	15
F	639	13	644	99	2
D	627	14	670	94	14
L	134	15	140	96	10
Total	**19903**		**20676**	**96**	

Data source: HMIS Portal, District X

Question:

- **Which blocks have the highest/lowest ANC registration in the district?**

This table shows:

- the number of pregnancies in each block
- The percentage of pregnancies registered for ANC
- The rankings of blocks by number of ANC registrations and by percentage of pregnancies registered.

Note how the rankings of block changes when ranked by the percentage of pregnancies registered vs total ANC registrations.

The appropriate way to judge block performance on ANC registrations is in terms of the proportions of pregnancies registered for ANC because this takes the population (i.e. denominator) into account.

Exercise 1: Tables
Block-wise performance on low birthweight babies

- Work in the same groups as before
- Open the Dataset_Module 2.3_Exercise 1-2
- Refer to the Instruction sheet_Module 2.3_Exercise 1

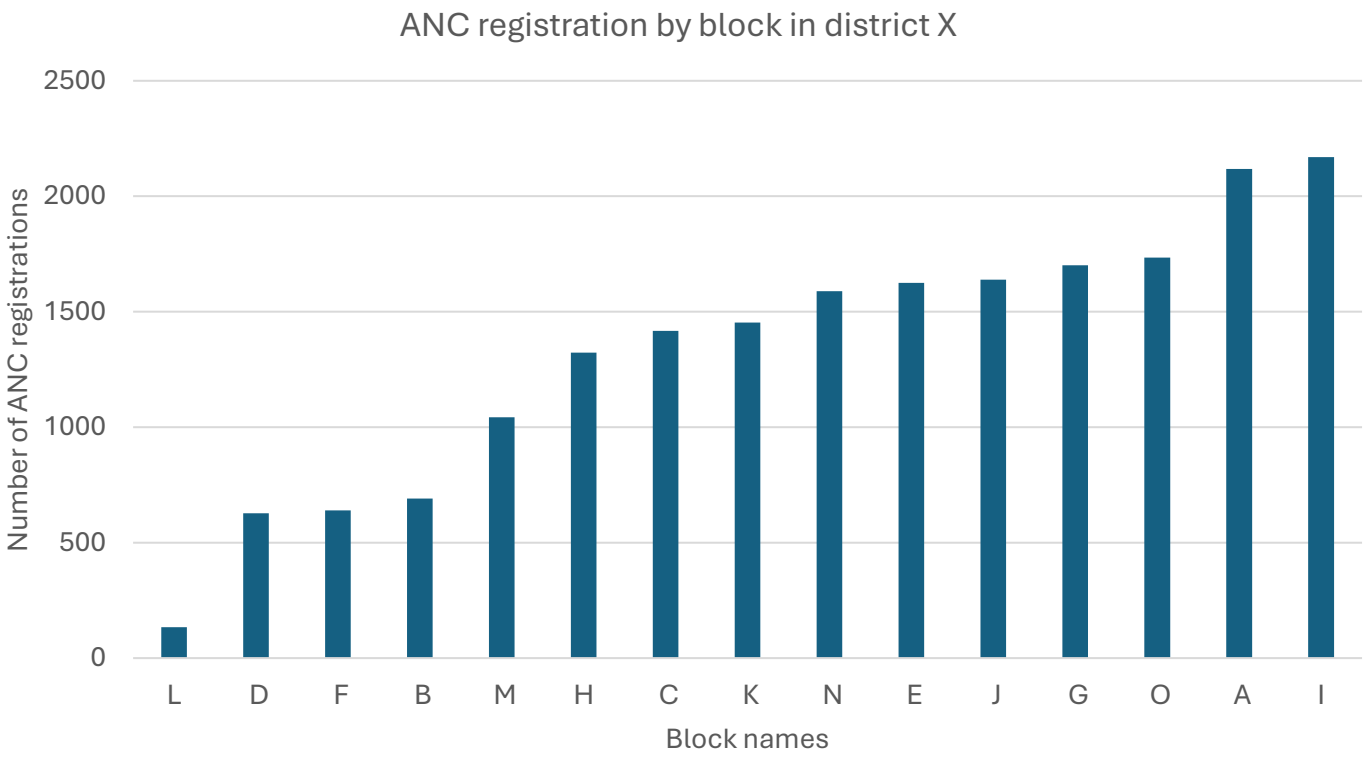
1. Let's start by exploring the dataset visually - are there any blocks with unusual/unexpected values e.g. very high/low number of low birth babies or with values that seem unusual with missing values?
2. Identify the best and worst performing blocks in terms of low-birth-weight babies?
   a. Sort the Blocks by LBW % (descending) and highlight the best/worst performing blocks.
   b. Use the conditional formatting option in Excel to create a heat map of LWB performance.
   c. Use the conditional formatting option in Excel to identify the top and bottom 4 performing blocks.
   d. Divide the blocks into quartiles (i.e. groups of 25%) of LBW and color code the LWB values in the table based on the 4 groups created.

IPSI

Session 2:
Bar chart and Pie chart

Duration: 30 min

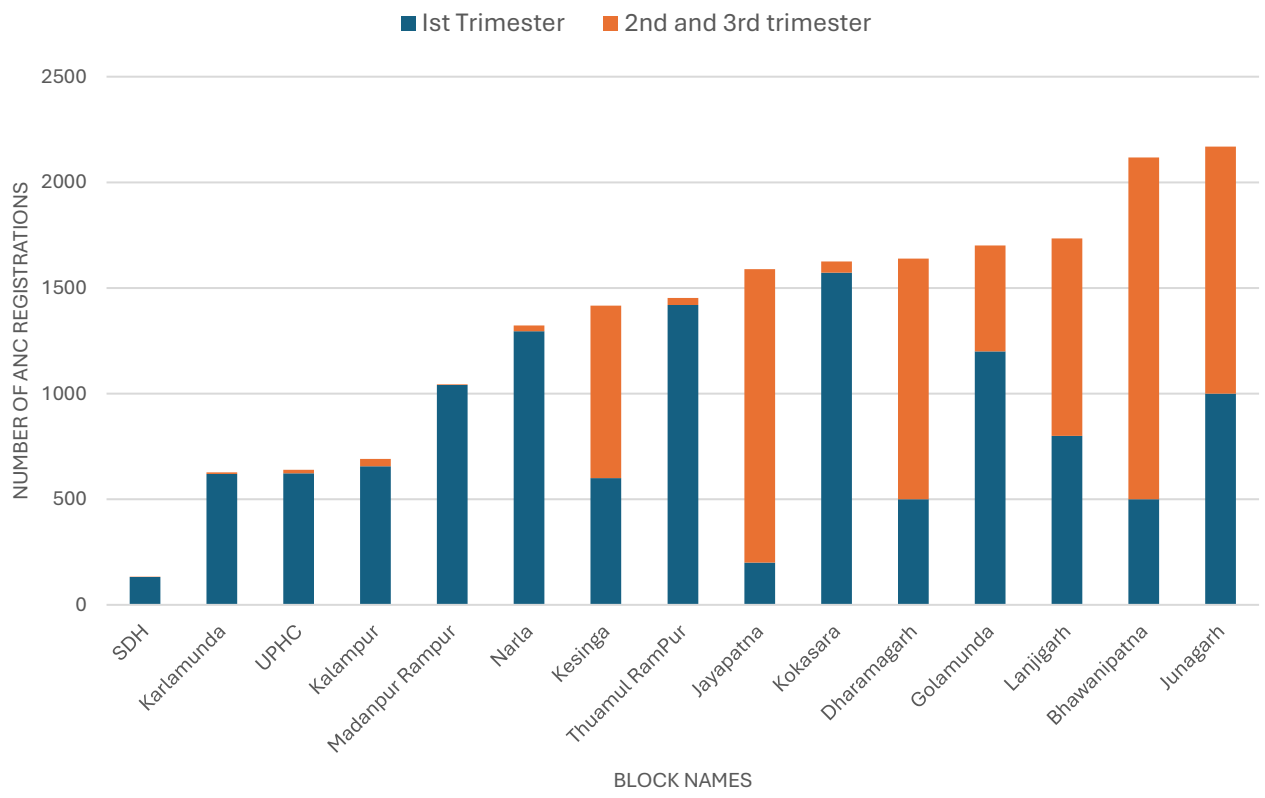
Categorial Data - Bar charts



- **Question: Which blocks have the highest/lowest levels of ANC registrations?**
- Bar graphs offer an easy way to show categorical data i.e. data points are associated with each category
- In the bar graph shown here, notice how quickly you can pick out the block with the highest/lowest ANC registrations.

Data source: HMIS Portal, District X

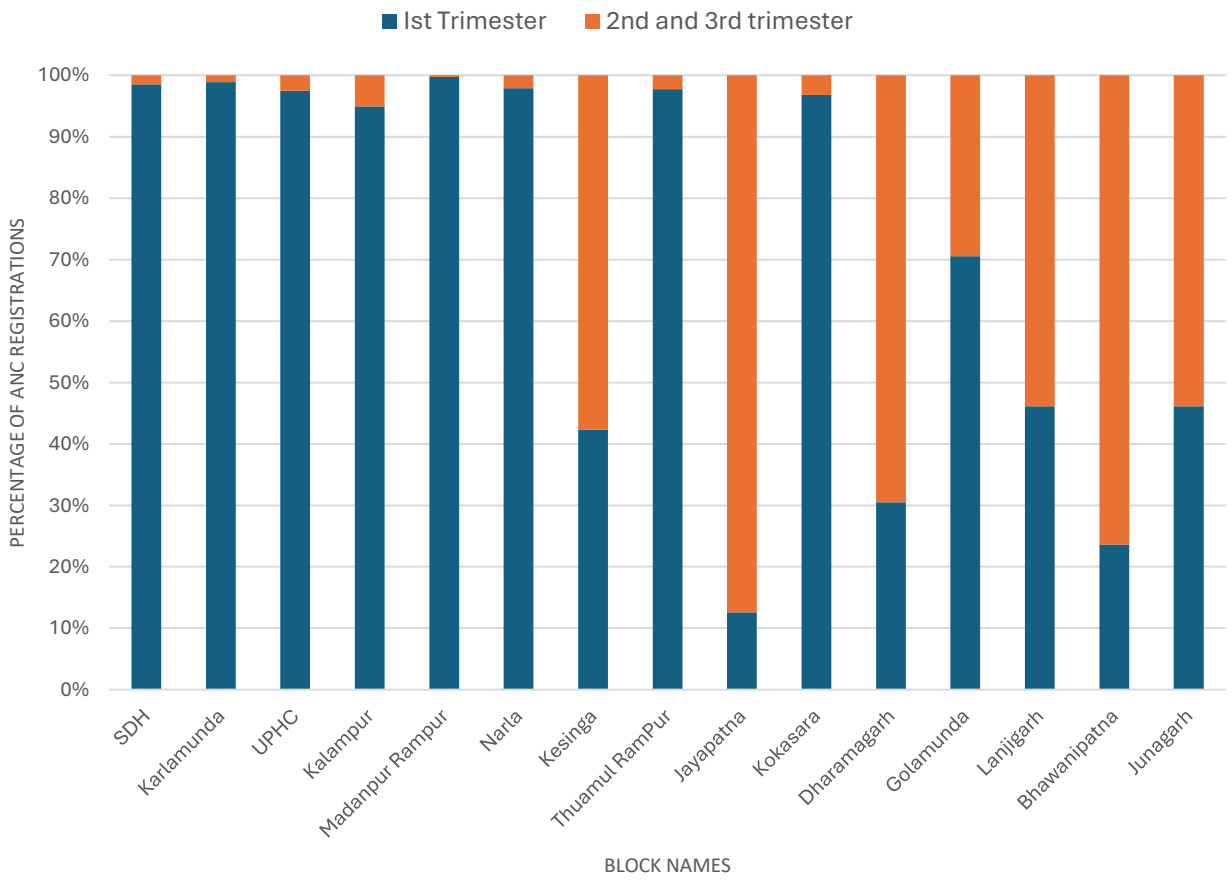
Bar charts – sub-categories



Data source: HMIS Portal, District X

- **Question: What share of ANC registrations are in the first trimester?**
- Stacked bar graphs are useful for visualizing the contribution of sub-categories. In bar graph shown here notice that we can easily see:
  - Blocks with the highest/lowest ANC registrations
  - The share of first trimester registrations in total registrations.
  - **Insight:** Blocks with the highest ANC registrations have a low share of first trimester ANC registrations, which is a concern.

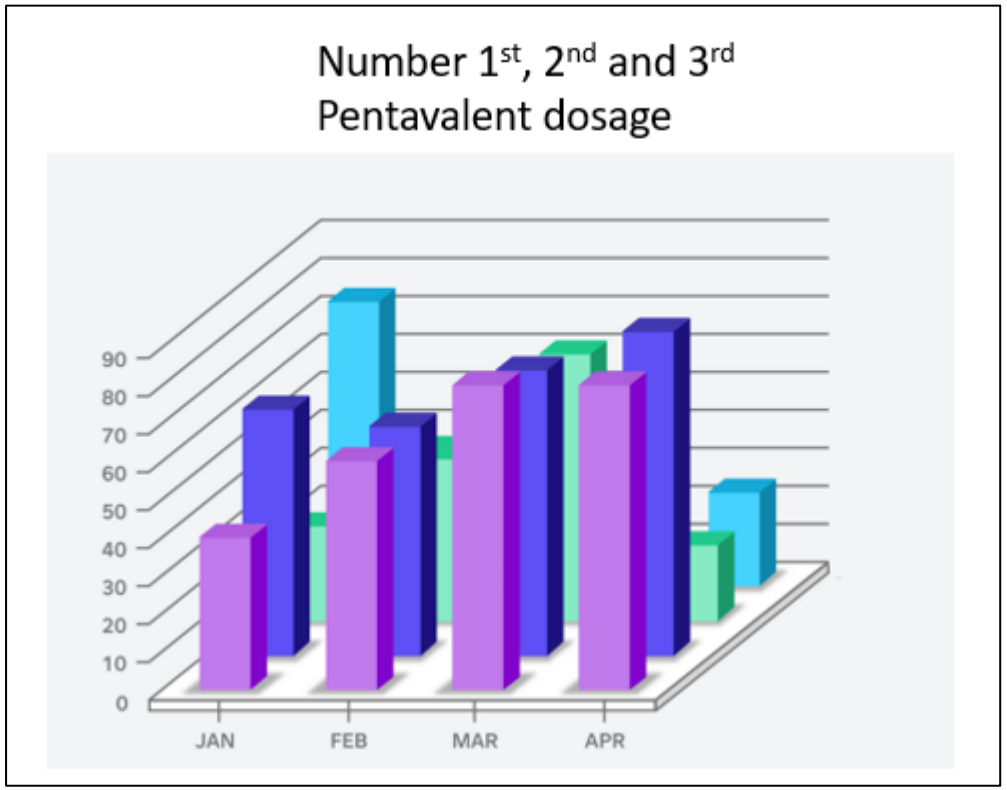
Bar charts – sub-categories



- **Question: What share of ANC registrations are in the first trimester?**
- 100% bar graphs are useful for visualizing the contribution of sub-categories. In bar graph shown here notice that we can easily see:
  - the share of first trimester registrations in total registrations.
  - **Insight:** Blocks with the highest ANC registrations have a low share of first trimester ANC registrations, which is a concern.

Data source: HMIS Portal, District X

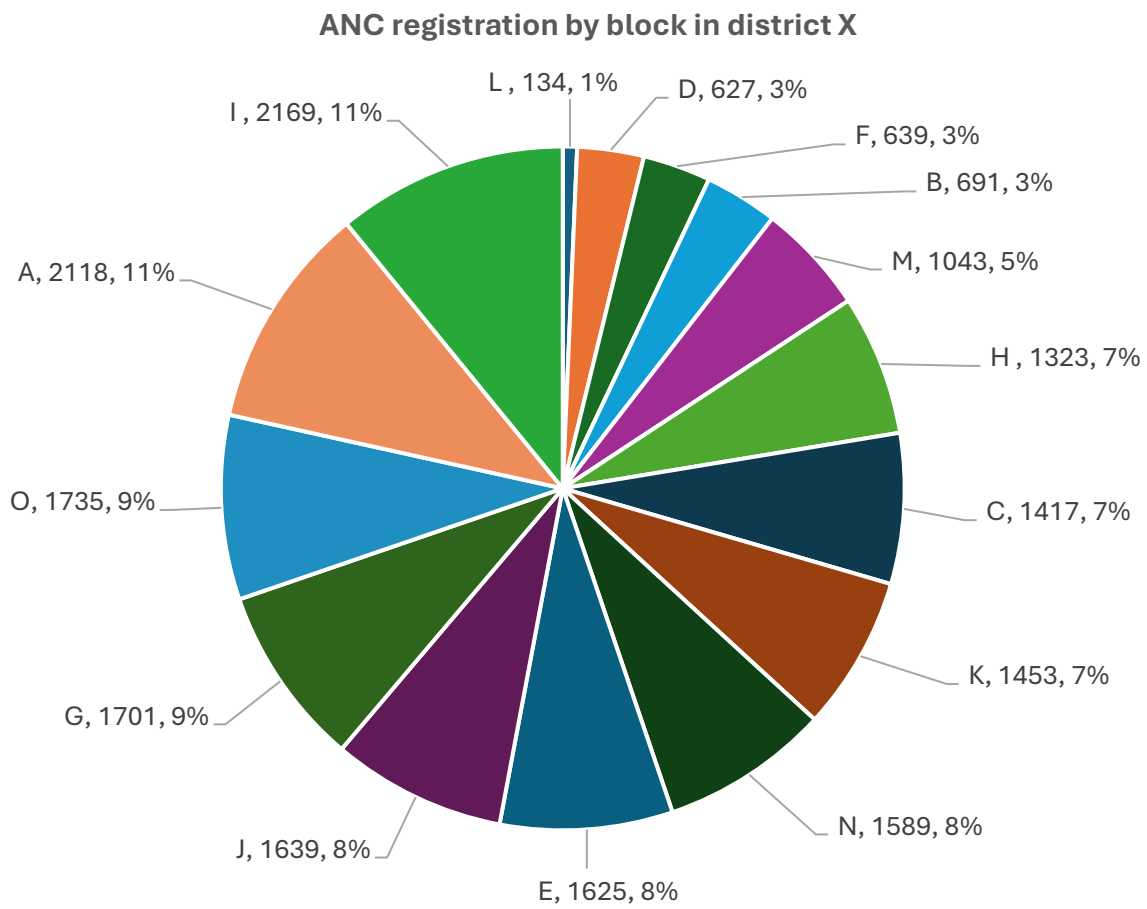
Spotting poor visualization practices: 3D Bar Chart – *Number of 1st, 2nd, and 3rd Pentavalent Dosages*



Data source: HMIS Portal, District X

- **Problems:**
  - 3D design distorts data perception and makes it harder to compare bar heights accurately.
  - Perspective makes some data points appear smaller/larger than they are.
  - Gridlines and background clutter distract from the data.
- **Better Approach:**
  - Use a 2D grouped bar chart.
  - Remove unnecessary 3D effects and background elements.
  - Add clear axis labels and legends.

Categorical data - Pie charts



Data source: HMIS Portal, District X

- **Question: What share of each block in overall ANC registrations?**
- Pie charts are another way of showing categorical data when we are interested in communicating/understanding the relative share of each category
- In the bar graph shown the size of the pie represents the share of each block in total ANC registrations.
- Note that if we have too many categories, it becomes unwieldy to read pie charts.

Spotting poor visualization practices: Pie Chart – *Month of Year*

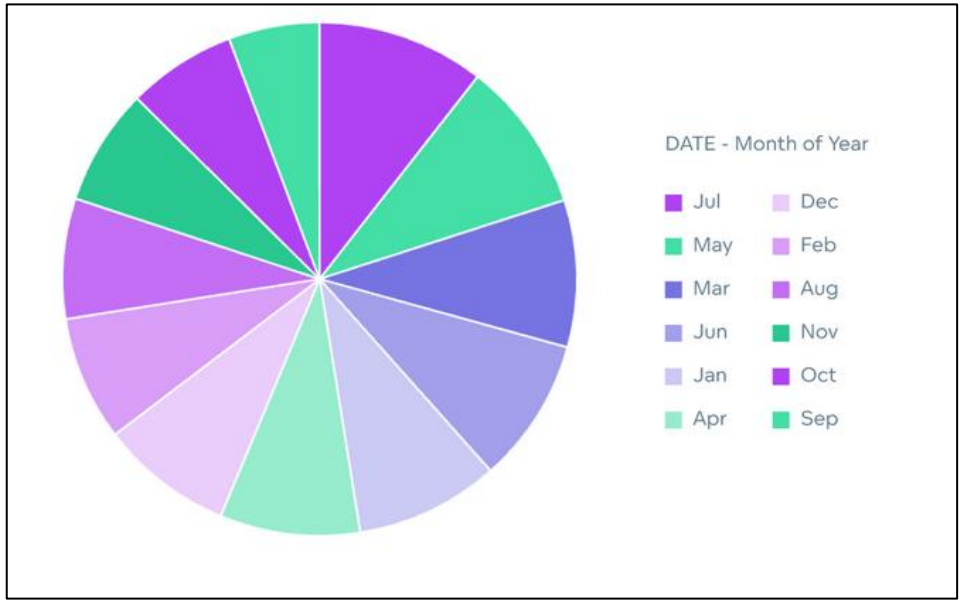


Image source: Storytelling with Data, Knaflic

Problems:

- Too many slices (12 categories), making it difficult to distinguish and compare sizes.
- Colors are too similar, reducing readability (especially for color-blind audiences).
- No clear ordering (months are scattered, not sequential).

Better Approach:

- Use a bar chart sorted chronologically or by value for easier comparison.
- Limit categories or group smaller ones into “Other.”
- Use a color palette with more contrast and accessibility in mind.

Exercise 2: Categorical data
Block-wise performance on low birthweight babies

- Work in the same groups as before
- Open the Dataset_Module 2.3_Exercise 1-2
- Refer to the Instruction sheet_Module 2.3_Exercise 2

In this exercise we will use different types of bar graphs to examine block performance.

You've already looked at data on low birth weight (LBW) across different administrative Blocks. In this exercise we will visualize:

1. Total births by type of Blocks
2. LBW burden by type of Blocks
3. Contribution of each Block to the total LBW burden in the district

Key Takeaways (1/2)

Data visualization transforms analysis into clear, actionable insights for decision-making

Tables:

- Useful for presenting exact numbers and detailed comparisons.
- Should be sorted, highlighted, or color coded (heat maps/quantiles) to reveal patterns.
- Avoid raw unsorted tables in presentations, they are hard to interpret quickly.

Bar Charts: Best for comparing categories; can use stacked or 100% stacked bars to show sub-categories and proportions.

Pie Charts: Show relative share of categories, but only effective with a few (≤5–6) categories.

- Always add clear axis labels, legends, and order categories logically.
- Avoid 3D effects and overcrowding these distort understanding.
- Use color contrast meaningfully (highlight key categories).

Key Takeaways (2/2)

Good visualizations: Choose the right chart type (tables, bar charts, pie charts, box plots, line plots, scatter plots) based on your question and data.

Poor visualizations can mislead, reduce credibility, and cause decision errors — avoid clutter, misleading scales, and inappropriate chart types.

Design principles:

- Simplify and focus on the message.
- Use clear labels, legends, and color contrast (consider accessibility).
- Avoid unnecessary 3D effects and decorative elements.

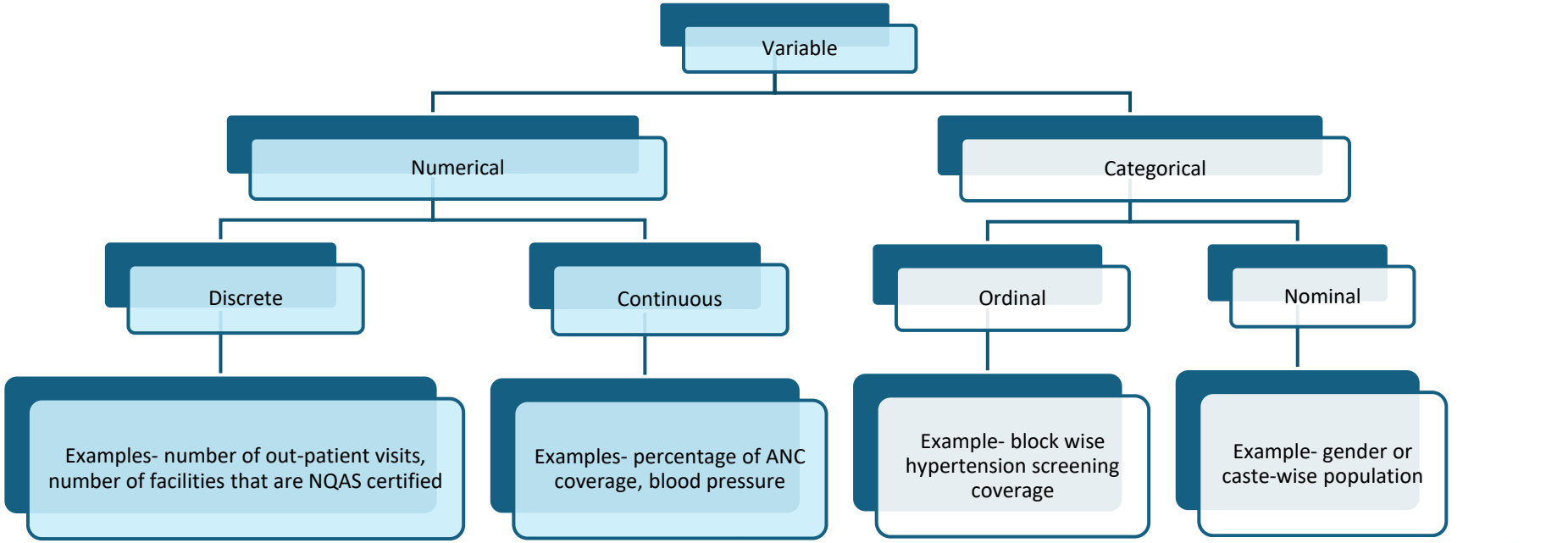
Useful Reading Resources

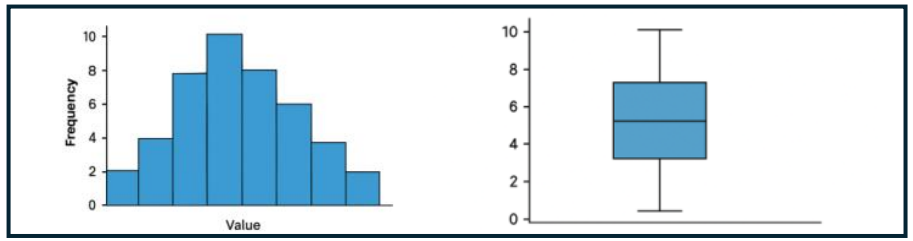
- Storytelling with Data, Knaflic
- The Visual Display of Quantitative Information by Edward R. Tufte.
- Beautiful Evidence by Edward R. Tufte.
- WHO Guidelines on Communicating Data Effectively
- NFHS and HMIS Reports

Summary- Section 2: Analysing and communicating data

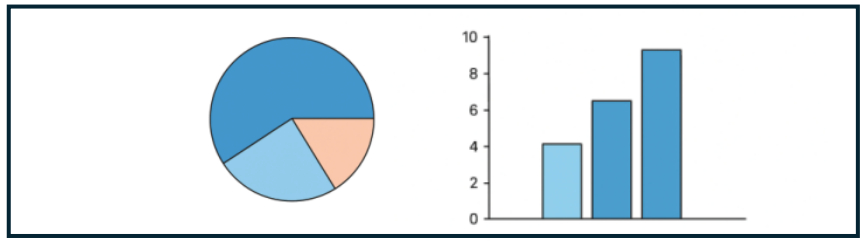
- Ask these questions as you start working with data-
  1. What data do you have?
  2. What questions do you want your data to answer?
  3. What visualization methods should you use to answer your questions?
  4. Did the findings from your analysis answer your question?
- Data/variable type guides the appropriate statistical tool selection and data visualization approach.
- Data visualization transforms analysis into clear, actionable insights for decision-making

Visualizations according the data variable type





Description:
Mean, standard deviation
Median,
Minimum, maximum
Inter-quartile range (percentile)



Description:
Absolute and relative frequency
Median, minimum, maximum
Inter-quartile range (percentile)

Description:
Absolute and relative frequency
Mode

Adapated from- Spriestersbach A, Röhrig B, du Prel JB, Gerhold-Ay A, Blettner M. Descriptive statistics: the specification of statistical measures and their presentation in tables and graphs. Part 7 of a series on evaluation of scientific publications.

Thank You