

STA 111: DESCRIPTIVE STATISTICS

LECTURE 3

Topic: Data Presentation/Exploration

By: E. E. E. Akarawak



**SCHOOL OF
SCIENCE AND
TECHNOLOGY**

PAN-ATLANTIC UNIVERSITY

Lecture Outline

- ❖ Introductory notes
- ❖ Frequency distribution for both grouped and ungrouped data
- ❖ Charts for Univariate categorical data
 - Bar Charts
 - Pie Charts
- ❖ Tables and Charts for bivariate categorical data

Expected Learning Outcome

Learning Outcomes:

- ✉ At the end of this lecture, the students should be able to:
 - i. Understand the concept of data exploration.
 - ii. Represent both ungrouped and grouped data in tables
 - iii. Represent qualitative data using charts, like bar charts, pie charts

Introductory Notes

- ☞ Recall that Statistics has two main areas: Descriptive Statistics and Inferential Statistics
- ☞ Descriptive Statistics is that aspect that has to do with methods for data collection, presentation and summary.
- ☞ Analysis based on these methods is referred to as descriptive analysis
- ☞ Exploratory Data Analysis (EDA) is the first step in data analysis, and helps to pave way for further analysis.
- ☞ Data exploration refers to the different ways of organizing and summarizing data to make it to appear meaningful and interpretable for decision making.
- ☞ It comprises of data presentation and summary
- ☞ It can be done using tables and charts; then numerical summaries

Data Exploration

To explore means to:

examine

investigate

search out

verify

walk around

discover

diagnose

prepare

organize

Introductory notes: Why Explore

Before you begin any analyses, it is important to examine all your data

Why explore data?

- to catch mistakes
- discover anomalies
- to see patterns in the data
- to find violations of statistical assumptions
- to avoid future trouble
- to avoid erroneous conclusion

Exploratory Data Analysis (EDA) Tools

Tables:

- Frequencies
- Cross tabs

Charts/Graphs:

- Line graphs
- Bar charts
- Pie charts
- Histogram
- Stem and Leaf Display
- Box Plot (Box and whisker diagram)

Before you Begin

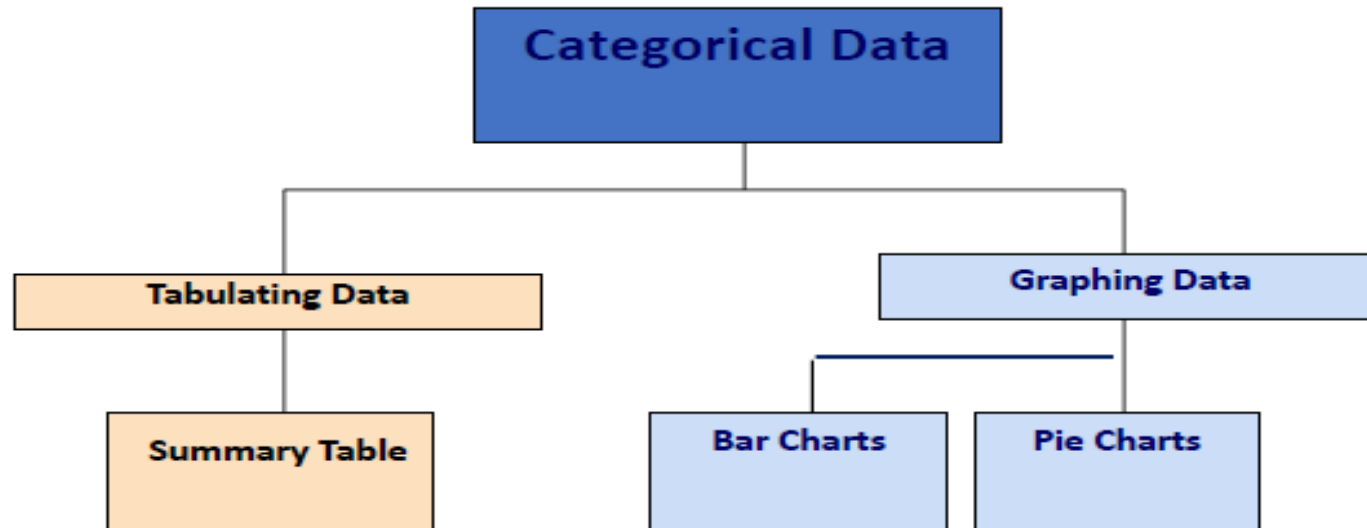
- ❖ You must understand your data
- ❖ Choices depend on your data type
- ❖ Also, remember your measurement scale

Exploring Qualitative Data

- ☛ If 1 ordinal or nominal – categorical univariate, then do
 - Frequencies distribution table
 - Simple bar chart
 - Pie chart
- ☛ If 2 ordinal or nominal – categorical bivariate, then do
 - Cross-tabulation (table)
 - Multiple or component bar chart

Exploring Qualitative Data

Categorical Data Can be Summarized By
Tables & Graphs



Exploring Quantitative Data

- Discrete, if ordinal or grouping in any form is possible, then frequencies
- Continuous, if grouped, then frequencies
- If not, no frequencies

- Charts for quantitative data
 - Line graph
 - Histogram
 - Stem and leaf
 - Box plots
- Summary Statistics

Frequency Distribution Table

X	f	Rf	%
0	2	0.083	8.3
1	6	0.25	25.0
2	6	0.25	25.0
3	4	0.167	16.7
4	3	0.125	12.5
5	3	0.125	12.5
Total	24	1	100.0

Table 1: X = No. of children in 24 families

1 2
5 3
2 2
1 1
3 2
4 0
2 4
4 5
1 1
0 5
1 2
3 3

Some Definitions

- ❖ X: the variable upon which values are taken
- ❖ Frequency (f): Number of times each value appears
- ❖ Relative Frequency (Rf): each frequency divided by total frequency
- ❖ Percentage (%): $Rf \times 100$
- ❖ Cumulative Frequency (Cf): Progressive total of frequency for each category
- ❖ Cumulative Percentage (C%): Progressive total of the percentage for each category

Cumulative Frequency Distribution Table

X	f	Rf	%	Cf	C%
0	2	0.083	8.3	2	8.3
1	6	0.25	25.0	8	33.3
2	6	0.25	25.0	14	58.3
3	4	0.167	16.7	18	75.0
4	3	0.125	12.5	21	87.5
5	3	0.125	12.5	24	100.0
Total	24	1	100.0		

Table 2: Extension of Table 1 to Cumulative frequencies

Grouped Data

Grouping is mostly used for continuous data or discrete data with large range. Below are some rules for grouping:

- ❦ (i) Determine the largest (maximum) and smallest (minimum) observations. The range is the difference, $R = \text{maximum} - \text{minimum}$
- ❦ (ii) A frequency distribution should have a minimum of 5 classes and a maximum of 20. For small data sets, use between 5 and 10 classes. For large data sets, use up to 20 classes.
- ❦ (iii) Divide the range by the chosen number of classes to get the class width
- ❦ (iv) Each data entry must fall into one and only one class.

Grouping (*contd*)

- ❦ (v) There should be no gaps. Moreover, if there are no entries for a particular class, that class must still be included with a frequency of 0.
- ❦ (vi) The first interval should begin about as much below the smallest value as the last interval ends above the largest.
- ❦ (vii) The intervals are called class intervals and the boundaries are called class boundaries.
- ❦ (viii) The class limits are the smallest (lower) and largest (upper) possible observed values in a class.
- ❦ (ix) The class mark is the midpoint of a class.

Example 1

55	60	61	35	41	43	50	78	72	83
45	70	76	31	49	65	79	83	41	86
53	62	52	47	38	57	64	78	47	54
43	73	85	48	66	48	85	86	82	48
56	84	37	57	57	45	95	45	73	39

- 👤 We set up a grouped frequency for the data, as below

Solution

Class	Class limits	Frequency	Class boundaries	Class center
1	30 – 39	5	29.5 – 39.5	34.5
2	40 – 49	13	39.5 – 49.5	44.5
3	50 – 59	9	49.5 – 59.5	54.5
4	60 – 69	6	59.5 – 69.5	64.5
5	70 – 79	8	69.5 – 79.5	74.5
6	80 – 89	8	79.5 – 89.5	84.5
7	90 – 99	1	89.5 – 99.5	94.5

More on Grouping

- ❖ Class Boundaries: The true boundaries between classes
- ❖ Class boundaries can be obtained by subtracting 0.5 from the lower class limit and adding it to the upper class limit.
- ❖ E.g., the class boundary for class 5 is: 69.5 – 79.5

Simple Bar Charts and Pie Charts

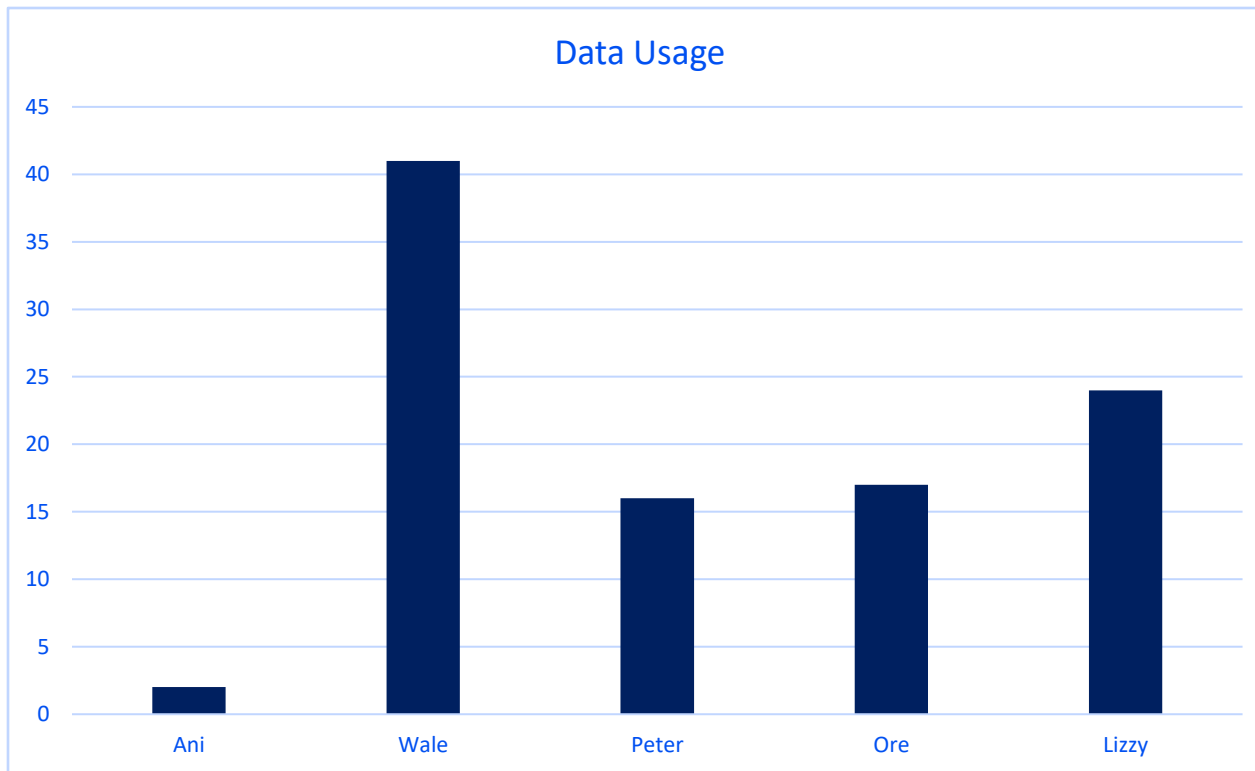
- ❖ Involves a single univariate variable with categories or distinct groups
- ❖ In bar charts, the bars represent frequencies and there must be equal spaces between bars.
- ❖ In pie charts, frequencies are represented by sectors of a circle.

Example 2

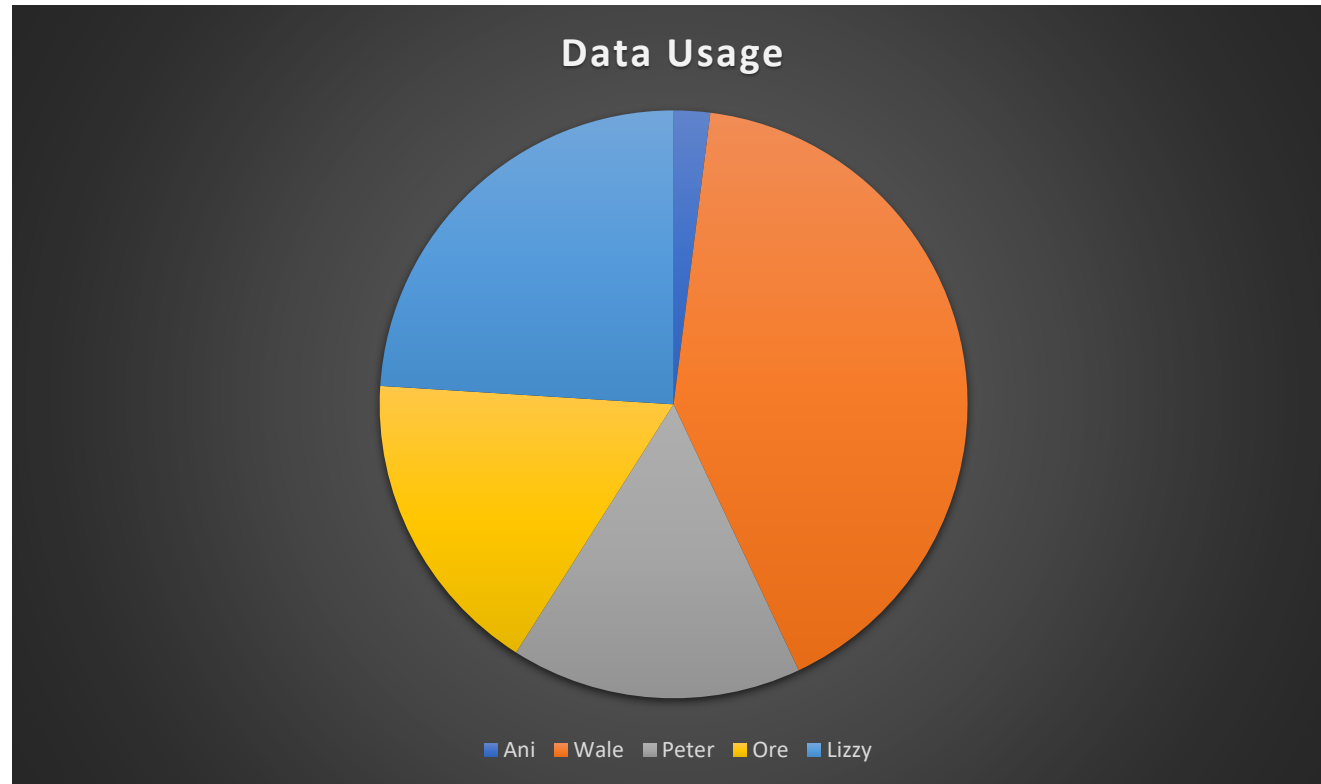
- ✿ The data below represent average weekly data usage by 5 students (Ani, Wale, Peter, Ore and Lizzy) in certain university. Obtain a bar chart and pie chart for the data (result from Excel chart)

Settlement	Data Usage
Ani	2
Wale	41
Peter	16
Ore	17
Lizzy	24
	100

A Simple Bar Chart for Consumption Data



A Pie Chart for Consumption Data



Cross tabulation

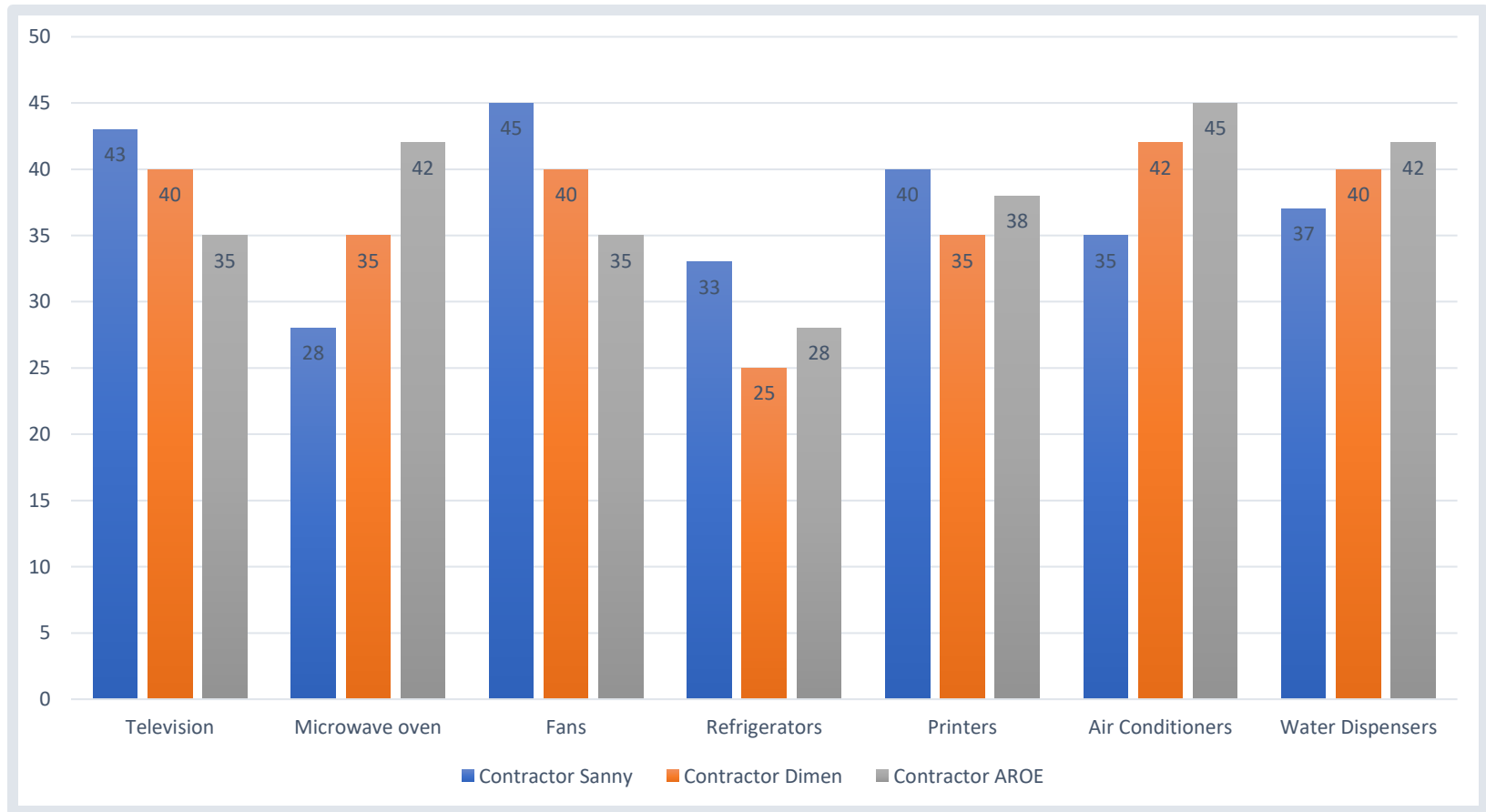
- It refers to a two-way table of rows and columns representing two categorical variables, say A and B, in which the rows represent categories of variable A and the columns represent the categories of variable B. The space where A and B intersects is called a cell and contains the frequency or count of A and B.

- Example:** Table 1:

Table 1 shows the number of electrical appliances supplied by 3 contractors (Sanny, Dimen and AROE) to a small company for its members of staff. Draw multiple and component bar charts.

Appliances	Contractor		
	Sanny	Dimen	AROE
Television	43	40	35
Microwave oven	28	35	42
Fans	45	40	35
Refrigerators	33	25	28
Printers	40	35	38
Air Conditioners	35	42	45
Water Dispensers	37	40	42
Total	261	257	265

A Multiple Bar Chart



A Component Bar Chat

