Chapter 2: Describing Data with Graphs
STAT 2601 – Business Statistics

Esam Mahdi

School of Mathematics and Statistics
Carleton University

January 15, 2026

Table of Contents

What is a Frequency Distribution?

Definition

A **frequency distribution** is a table that organizes data into classes or categories, showing the number of observations (frequency) in each class.

Real-world Example:
Survey of 50 Carleton University students' favorite study spots on campus:

- Library: 25 students
- Residence: 12 students
- Cafeteria: 8 students
- Classroom: 5 students

Study Spot	Frequency
Library	25
Residence	12
Cafeteria	8
Classroom	5
Total	50

Table: Frequency Distribution

Frequency, Relative Frequency, and Percentage

Measuring "How Often"

Three ways to present frequency information:

1. **Frequency:** Number of measurements in each category
2. **Relative Frequency:** Proportion of measurements

$$\text{Relative Frequency} = \frac{\text{Frequency}}{n},$$

where n = total number of measurements

3. **Percentage:** Percentage $=$ Relative frequency $\times\ 100\%$

Important Properties

- Sum of frequencies = n
- Sum of relative frequencies = 1
- Sum of percentages = 100%

Spot	Freq	Rel. Freq	Percent
Library	25	0.50	50%
Residence	12	0.24	24%
Cafeteria	8	0.16	16%
Classroom	5	0.10	10%
Total	50	1.00	100%

Graphically Summarizing Qualitative Data: Bar Chart

Features:

- Bars represent categories
- Height shows frequency
- Bars don't touch each other
- Categories can be in any order

Example: A bag of M&Ms contains 21 candies

- Red: 2 colors
- Blue: 5 colors
- Green: 3 colors
- Orange: 3 colors
- Brown: 6 colors
- Yellow: 2 colors

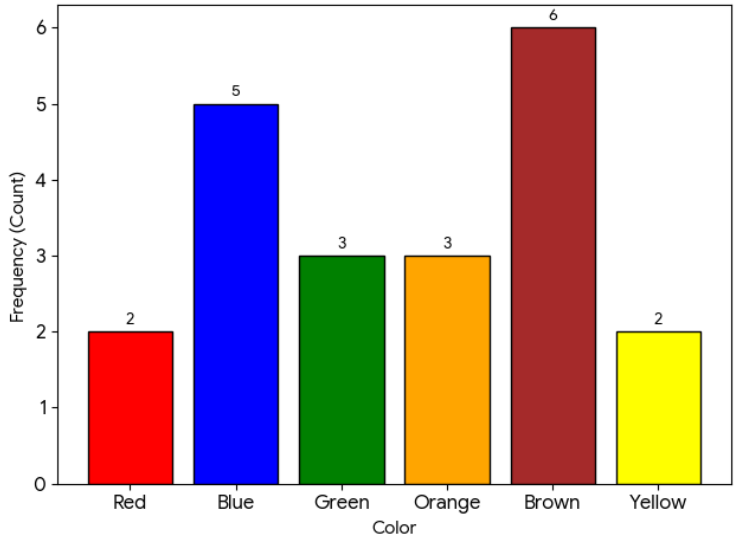

Figure: Bar Chart Example

Graphically Summarizing Qualitative Data: Pie Chart

Features:

- Circle represents whole (100%)
- Slices show proportions
- The size of each slice is proportional to the magnitude of the displayed variable associated with each category or class.

  Sector Angle: $= \frac{f}{n} \times 360°$
- Useful for showing parts of a whole
- Good for 5-7 categories max

Example: Monthly Expenses

- Rent: 40%
- Food: 25%
- Transportation: 15%
- Entertainment: 10%
- Savings: 10%

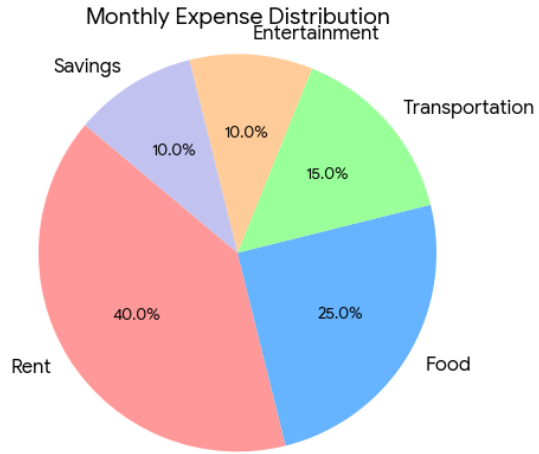

Figure: Pie Chart Example

Graphically Summarizing Qualitative Data: Pareto Chart

Features:

- Special bar chart
- Bars in descending order
- Line shows cumulative percentage
- Identifies "vital few" vs "trivial many"

Example: Customer Complaints

- Long wait time: 45 complaints
- Poor service: 30 complaints
- Price too high: 15 complaints
- Product quality: 10 complaints

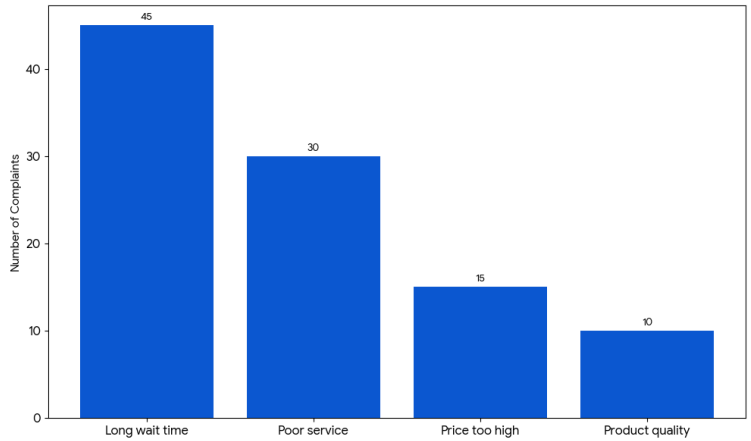

Figure: Pareto Chart Example

Graphically Summarizing Quantitative Data: Line Charts (Time Series)

Features:

- Shows data over time
- X-axis: Time intervals
- Y-axis: Variable of interest
- Reveals trends, cycles, patterns

Carleton University Example:

- Monthly enrollment applications
- Daily library visits
- Weekly campus bookstore sales

Example:

- Monthly sales data
- Daily stock prices
- Quarterly profits

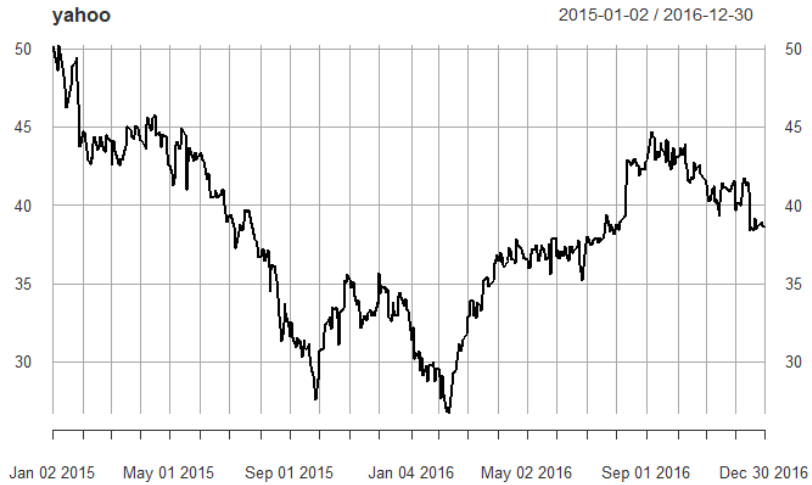

Figure: Time Series Example

Graphically Summarizing Quantitative Data: Histogram

What is a Histogram?

- Graphical representation of **quantitative** data
- Bars represent frequency in intervals
- Bars touch each other
- Used for continuous data

Example: STAT 2601 Midterm Scores

- 40-50: 5 students
- 50-60: 12 students
- 60-70: 18 students
- 70-80: 22 students
- 80-90: 15 students
- 90-100: 8 students

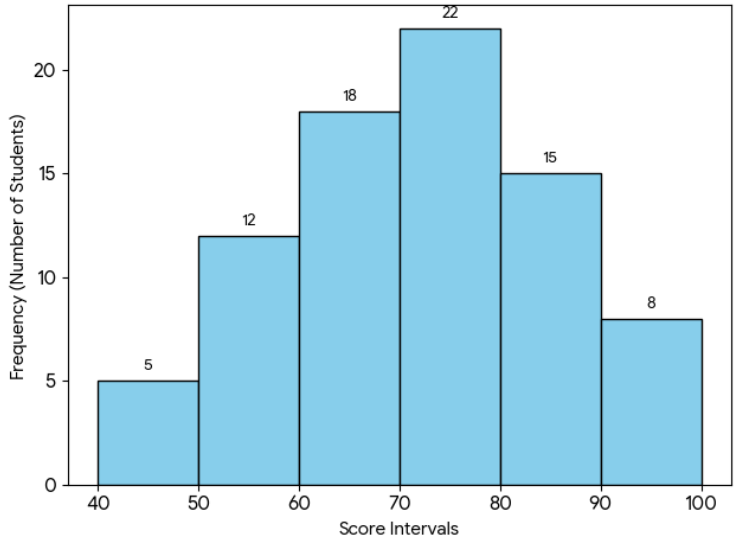

Figure: Histogram Example

Histogram Types

Frequency Histogram:

- Shows actual counts
- Height = frequency
- Sum of frequencies = n

Relative Frequency Histogram:

- Shows proportions
- Height = relative frequency
- Area = 1 (or 100%)
- sum of relative frequency = 1

Percentage Histogram:

- Shows percentages
- Height = percentage
- Sum of percentage = 100%

Same data, different perspectives:

- Frequency: "22 students scored 70-80"
- Relative: "0.275 of students scored 70-80"
- Percentage: "27.5% of students scored 70-80"

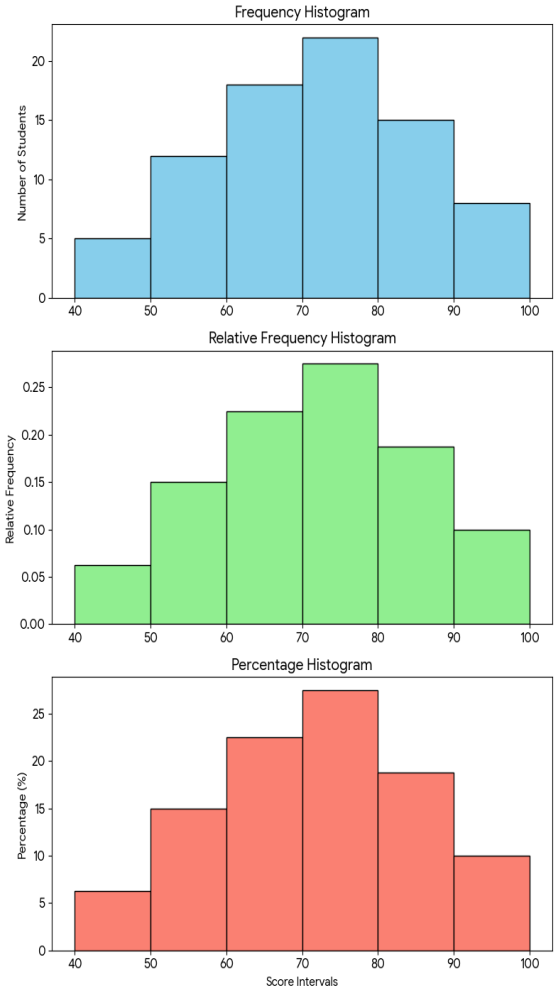

Figure: Three Types of Histograms

Relative Frequency Histograms

Construction Steps

1. Divide data range into 5-12 subintervals of equal length
2. Calculate approximate width: $\frac{\text{Range}}{\text{\# subintervals}}$
3. Round width up to convenient value
4. Use **left inclusion** method
5. Create frequency table
6. Draw histogram with subintervals on x-axis, relative frequencies on y-axis

Interpretation

- Height = proportion in that class = probability that a random measurement belongs to that class
- Total area = 1 (or 100%)

Distribution Shapes

- **Mound-shaped:** Symmetric, bell-shaped (Mean = Median)
- **Skewed right:** Long tail on right (Mean > Median)
- **Skewed left:** Long tail on left (Mean < Median)
- **Bimodal:** Two peaks
- **Uniform:** Equal frequencies

Outlier

A data point that is significantly differs from the rest of the observations in a dataset, lying an unusually large distance from the overall pattern, often at the extremes.

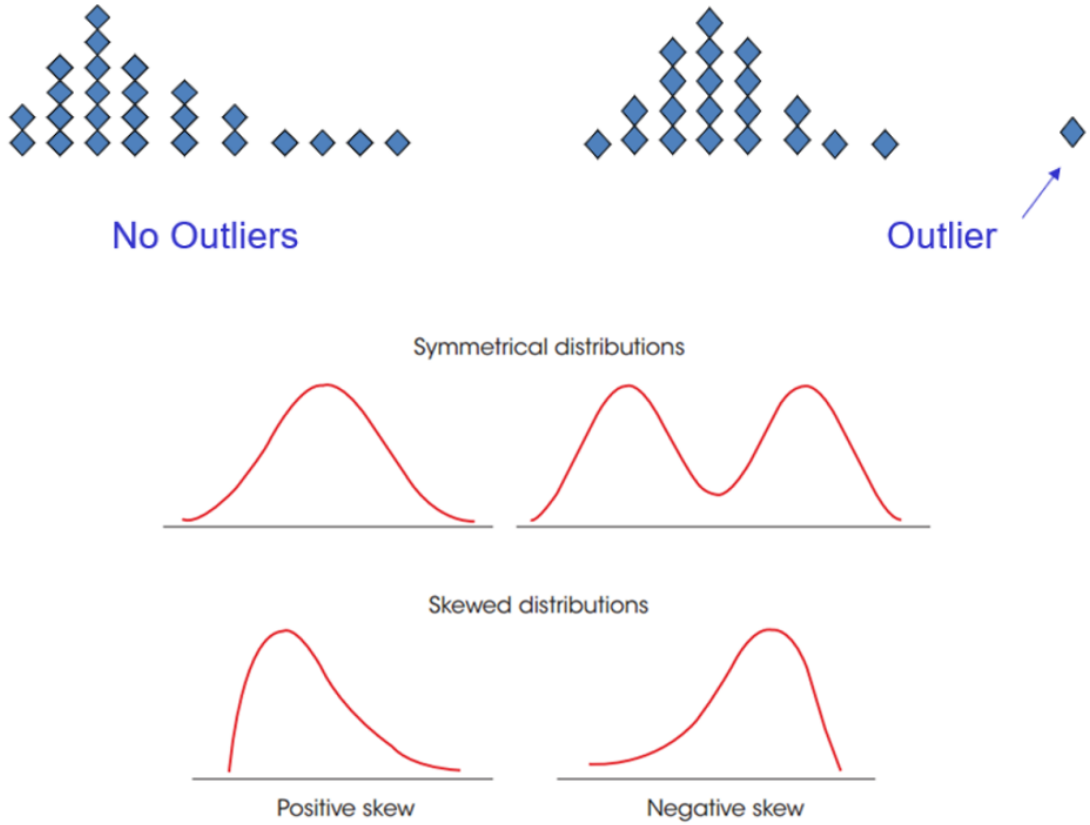

Figure: Histogram Showing Outliers

Detecting Outliers with Histograms

What are outliers?

- Unusually high or low values
- Far from other observations
- May indicate errors or special cases

Example: Ottawa House Prices (2023)

- Most houses: $\$400,000 - \$800,000$
- Some outliers: $\$3,000,000+$
- These appear as isolated bars

In Histograms:

- Isolated bars at extremes
- Gaps in the distribution
- Bars far from main cluster

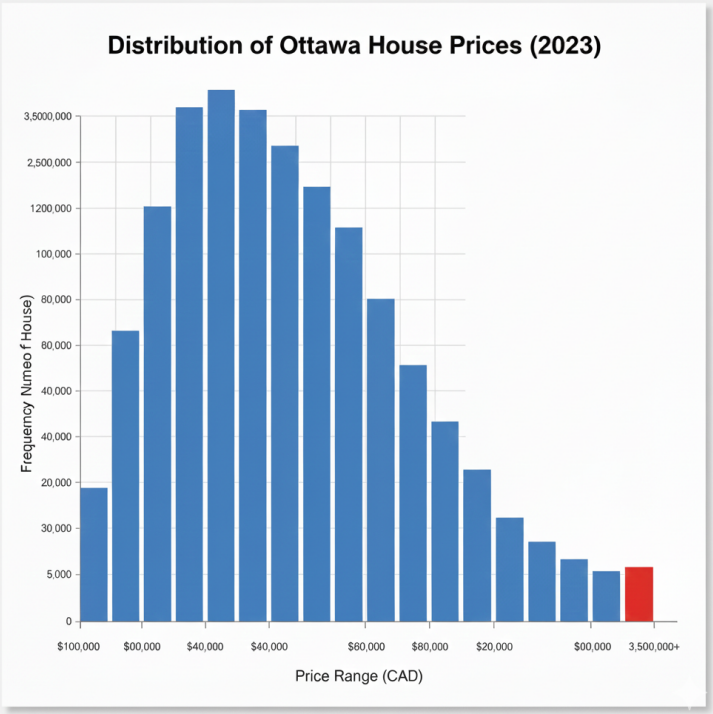

Figure: Histogram Showing Outliers

Example: Faculty Ages

Data: Ages of 50 tenured faculty

34, 48, 70, 63, 52, 52, 35, 50, 37, 43, 53, 43, 52, 44, 42, 31, 36, 48, 43, 26, 58, 62, 49, 34, 48, 53, 39, 45, 34, 59, 34, 66, 40, 59, 36, 41, 35, 36, 62, 34, 38, 28, 43, 47, 51, 59, 35, 47, 41, 53

Age	Tally	Freq.	Rel. Freq.	Percent
25 to < 33	𝍸	5	0.10	10%
33to < 41	𝍸||||	14	0.28	28%
41to < 49	𝍸|||	13	0.26	26%
49to < 57	𝍸||||	9	0.18	18%
57to < 65	𝍸||	7	0.14	14%
65to < 73	||	2	0.04	4%
Total		50	1.00	100%

Calculations

- Min = 26, Max = 70
- Range = 70 - 26 = 44
- Choose 6 intervals
- Width = 44/6 = 7.33
- Use width = 8
- Start at 25

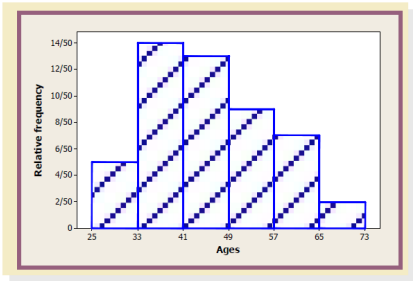

Figure: Three Types of Histograms

Shape Analysis

- **Shape:** Skewed right
- **Outliers:** None apparent

Graphically Summarizing Quantitative Data: Stem-and-Leaf Plot

What is a Stem-and-Leaf Display?

- Tabular display of quantitative data
- Preserves individual data values
- Shows shape like histogram
- Easy to construct by hand

Construction Steps

1. Divide each measurement into **stem** and **leaf**
2. List stems in a column with vertical line
3. Record leaf portion in matching row
4. Order leaves from lowest to highest
5. Provide a key for coding

Parts:

- Stem: Leading digit(s)
- Leaf: Trailing digit(s)
- Key: How to read it

Stem	Leaf
6	5 7 8
7	1 2 2 4 5 8 9
8	1 3 5 8 9
9	2 5 8

Table: Stem-and-Leaf - **Unit Leaf:** 1

Example: STAT 2601 Assignment 1 Scores

Data: $65, 67, 68, 71, 72, 72, 74, 75, 78, 79, 81, 83, 85, 88, 89, 92, 95, 98$

Creating a Stem-and-Leaf Display

Step-by-Step Example:

Carleton bookstore daily coffee sales (in dollars) for 20 days:

$$23, 34, 27, 45, 38, 29, 41, 36, 52, 47, 31, 44, 39, 42, 55, 48, 33, 46, 40, 50$$

Steps:

1. Sort data: 23, 27, 29, 31, 33, 34, 36, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 50, 52, 55
2. Choose stems (tens digit): 2, 3, 4, 5
3. Add leaves (ones digit)

Stem	Leaf	
2	3 7 9	
3	1 3 4 6 8 9	Unit Leaf = 1
4	0 1 2 4 5 6 7 8	
5	0 2 5	

Interpretation:

- Most sales: \$40-\$48 range
- Shape: Approximately symmetric (though it may be argued that the distribution is slightly left-skewed).
- No obvious outliers

Detecting Outliers with Stem-and-Leaf

Example: Employee Salaries (in $000s)

Data : 42, 45, 48, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 63, 65, 68, 71, 85, 92

Stem	Leaf	
4	2 5 8	
5	1 2 3 4 5 6 8 9	
6	1 2 3 5 8	Unit leaf = 1
7	1	
8	5	
9	2	

Outliers Detected:

- Most salaries: $\$51,000 - \$68,000$
- Possible outliers:
  - $\$85,000$ (senior manager?)
  - $\$92,000$ (executive?)
- Isolated stems 8 and 9

Business Insight:

- Two highly paid employees
- May need separate analysis
- Could skew average salary

Identifying Distribution Shapes with Stem-and-Leaf

Stem	Leaf
3	4 5 6
4	1 2 3 4 5 6
5	0 1 2 3 4 5 6 7
6	0 1 2 3 4 5 6 7 8 9
7	1 1 2 3 3 3 6 7
8	2 2 3 4 4
9	1 6 9

Shape: Symmetric (mound shape), centered around the 60s

Stem	Leaf
0	2 3 4 5 6 7 8 8 9
1	0 1 2 3 4 4 5 6 7 8 9
2	0 1 2 3
3	5
4	
5	8

Shape: Skewed right with a long tail toward higher values

Stem	Leaf
2	1
3	0 2
4	1 2 3
5	4 5 6 7 8 9
6	2 3 4 5 6 7 8 9
7	1 2 3 4 5 6 7 8 9

Shape: Skewed left with a tail toward lower values

Stem	Leaf
3	1 2 2 3
4	
5	0 1 2 3
6	
7	8 8 9
8	0 1 2 3

Shape: Multimodal (more than two modes)

Stem	Leaf
5	0 0 1 2 3
6	
7	8 8 9 9
8	0 0 1 2 3

Shape: Bimodal (two clear clusters (50s and 80s))

Stem	Leaf
2	0 1 2 3 4 5 6 7 8 9
3	0 1 2 3 4 5 6 7 8 9
4	0 1 2 3 4 5 6 7 8 9

Shape: Uniform (evenly distributed)

Example

Construct a stem-and-leaf plot for the following dataset. Describe its distribution and determine whether any outliers are present.

$$1.3, 1.6, 1.7, 1.7, 2.0, 2.2, 2.4, 2.6, 2.8, 2.9, 3.0, 3.2, 4.1, 9.4$$

Solution: First, arrange the data in ascending order (already sorted):

$$1.3,\ 1.6,\ 1.7,\ 1.7,\ 2.0,\ 2.2,\ 2.4,\ 2.6,\ 2.8,\ 2.9,\ 3.0,\ 3.2,\ 4.1,\ 9.4$$

Using the integer part as the stem and the first decimal place as the leaf, the stem-and-leaf plot is:

Stem	Leaf
1	3 6 7 7
2	0 2 4 6 8 9
3	0 2
4	1
5	
6	
7	
8	
9	4

Unit Leaf: 0.1

The distribution is positively skewed (right-skewed), with most observations clustered between 1.3 and 3.2 and a long tail extending to the right. The value 9.4 is clearly an outlier.

Dot Plots

When to Use Dot-Plots

- Summarize qualitative (categorical) data
- Quantitative data with few repeated values
- Small to moderate dataset sizes
- Shows individual data points
- Reveals outliers and patterns

Dot plot Construction

1. Draw a number line covering data range
2. Place a dot above the number line for each observation
3. Stack dots for repeated values

Quality Control Example: Gear diameters (cm) from a machine

```
1.991  1.891  1.991  1.988  1.993  1.989  1.990  1.988  1.988
1.993  1.991  1.989  1.989  1.993  1.990  1.994
```

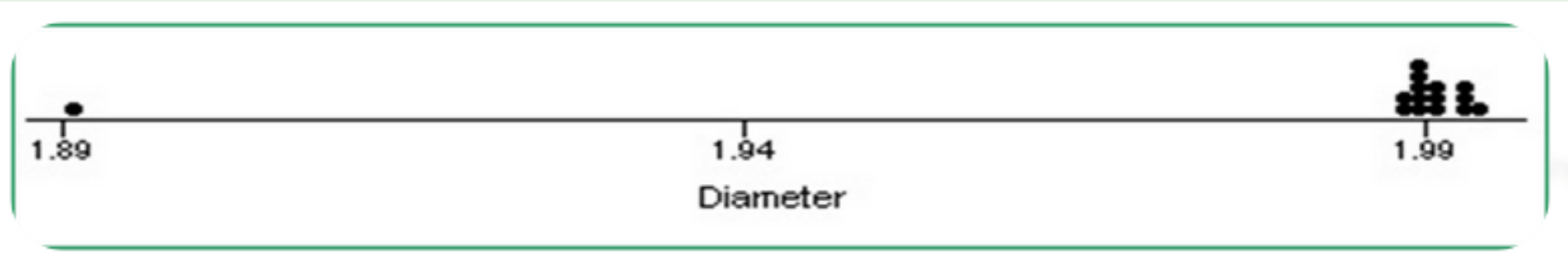

Figure: dot-plot Example

Note: 1.891 is likely a typing error (outlier).

Summary

Chapter 2 Summary

1. **Frequency Distributions:** Organize data into meaningful categories
2. **Graphical Methods for Qualitative Data:**
   - Bar Charts: Compare categories
   - Pie Charts: Show parts of a whole
   - Pareto Charts: Identify most important factors
   - Time Series: Analyze patterns over time
3. **Histograms:** Visualize quantitative data distribution
   - Detect outliers
   - Identify shape (symmetric, skewed)
   - Determine modality (uni/bi/multimodal)
4. **Stem-and-Leaf Displays:**
   - Alternative to histograms
   - Preserve individual data values
   - Quick exploratory analysis

Remember

The goal of descriptive statistics is to summarize and understand data before making inferences.