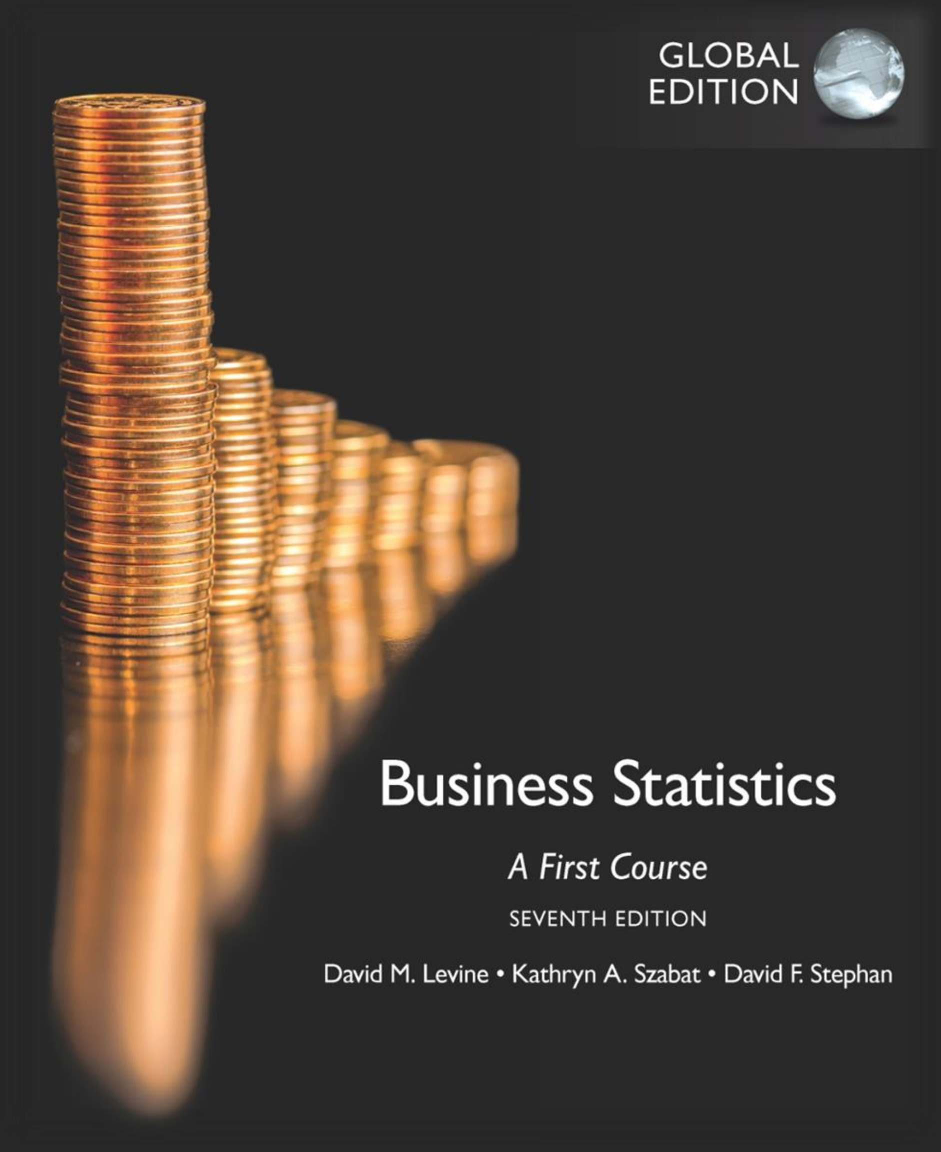


GLOBAL EDITION

Business Statistics

A First Course

SEVENTH EDITION

David M. Levine • Kathryn A. Szabat • David F. Stephan

Chapter 2

Organizing and Visualizing Variables

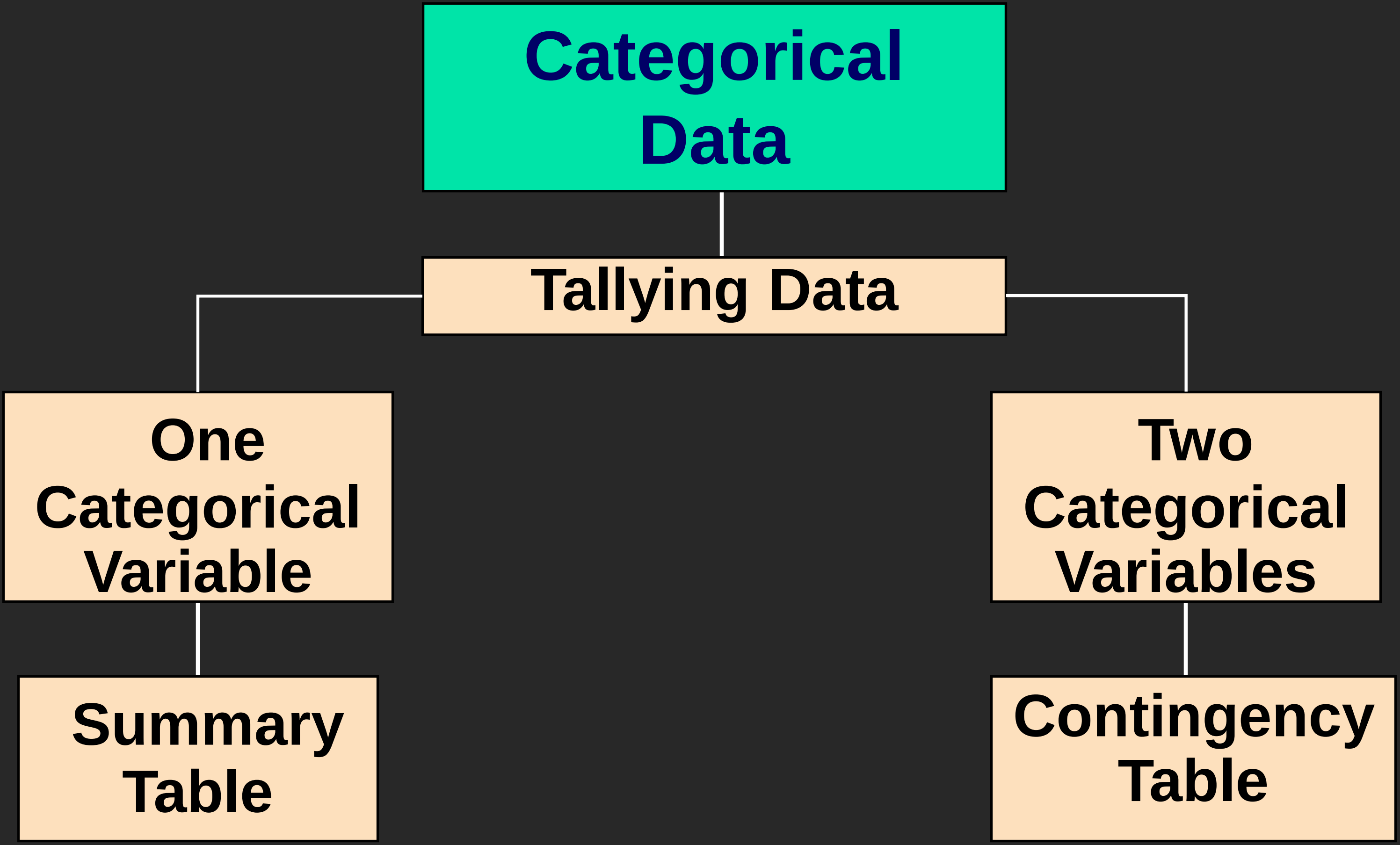
Objectives

In this chapter you learn:

- Methods to organize variables.
- Methods to visualize variables.
- Methods to organize or visualize more than one variable at the same time.
- Principles of proper visualizations.

Categorical Data Are Organized By Utilizing Tables

DCOVA



Categorical Data

Tallying Data

One Categorical Variable

Summary Table

Two Categorical Variables

Contingency Table

Organizing Categorical Data: Summary Table

DCOVA

- A **summary table** tallies the frequencies or percentages of items in a set of categories so that you can see differences between categories.

Main Reason Young Adults Shop Online

Reason For Shopping Online?	**Percent**
Better Prices	37%
Avoiding holiday crowds or hassles	29%
Convenience	18%
Better selection	13%
Ships directly	3%

Source: Data extracted and adapted from “Main Reason Young Adults Shop Online?”
USA Today, December 5, 2012, p. 1A.

A Contingency Table Helps Organize Two or More Categorical Variables

DCOVA

- Used to study patterns that may exist between the responses of two or more categorical variables More details of the definition

- Cross tabulates or tallies jointly the responses of the categorical variables

- For two variables the tallies for one variable are located in the rows and the tallies for the second variable are located in the columns

Contingency Table - Example

DCOVA

- A random sample of 400 invoices is drawn.
- Each invoice is categorized as a small, medium, or large amount.
- Each invoice is also examined to identify if there are any errors.
- This data are then organized in the contingency table to the right.

Contingency Table Showing Frequency of Invoices Categorized By Size and The Presence Of Errors

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400

Contingency Table Based On Percentage Of Overall Total

DCOVA

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400

42.50% = 170 / 400
25.00% = 100 / 400
16.25% = 65 / 400

	No Errors	Errors	Total
Small Amount	42.50%	5.00%	47.50%
Medium Amount	25.00%	10.00%	35.00%
Large Amount	16.25%	1.25%	17.50%
Total	83.75%	16.25%	100.0%

83.75% of sampled invoices have no errors and 47.50% of sampled invoices are for small amounts.

Contingency Table Based On Percentage of Row Totals

DCOVA

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400

89.47% = 170 / 190
71.43% = 100 / 140
92.86% = 65 / 70

	No Errors	Errors	Total
Small Amount	89.47%	10.53%	100.0%
Medium Amount	71.43%	28.57%	100.0%
Large Amount	92.86%	7.14%	100.0%
Total	83.75%	16.25%	100.0%

Medium invoices have a larger chance (28.57%) of having errors than small (10.53%) or large (7.14%) invoices.

Contingency Table Based On Percentage Of Column Totals

DCOVA

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400

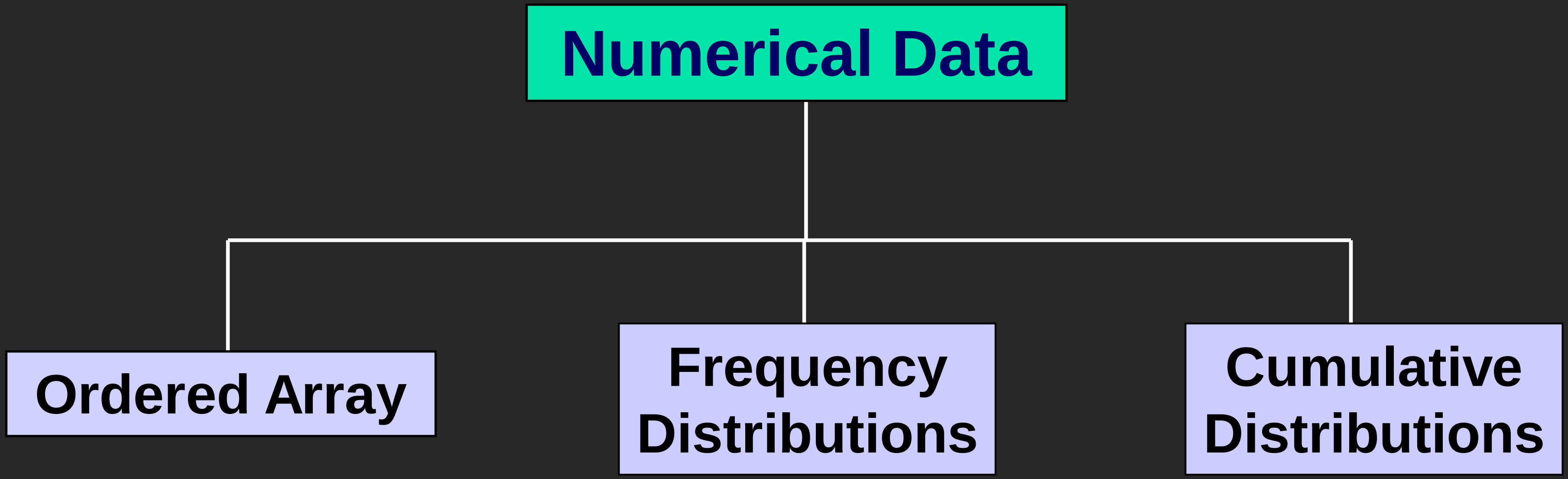
50.75% = 170 / 335
30.77% = 20 / 65

	No Errors	Errors	Total
Small Amount	50.75%	30.77%	47.50%
Medium Amount	29.85%	61.54%	35.00%
Large Amount	19.40%	7.69%	17.50%
Total	100.0%	100.0%	100.0%

There is a 61.54% chance that invoices with errors are of medium size.

Tables Used For Organizing Numerical Data

DCOVA



Numerical Data

Ordered Array

Frequency Distributions

Cumulative Distributions

Organizing Numerical Data: Ordered Array

DCOVA

- An **ordered array** is a sequence of data, in rank order, from the smallest value to the largest value.
- Shows range (minimum value to maximum value)
- May help identify outliers (unusual observations)

Age of Surveyed College Students	**Day Students**					
	16	17	17	18	18	18
	19	19	20	20	21	22
	22	25	27	32	38	42
	Night Students					
	18	18	19	19	20	21
	23	28	32	33	41	45

Organizing Numerical Data: Frequency Distribution

DCOVA

- The **frequency distribution** is a summary table in which the data are arranged into numerically ordered classes.

- You must give attention to selecting the appropriate *number* of **class groupings** for the table, determining a suitable *width* of a class grouping, and establishing the *boundaries* of each class grouping to avoid overlapping.

- The number of classes depends on the number of values in the data. With a larger number of values, typically there are more classes. In general, a frequency distribution should have at least 5 but no more than 15 classes.

- To determine the **width of a class interval,** you divide the **range** (Highest value–Lowest value) of the data by the number of class groupings desired.

Organizing Numerical Data: Frequency Distribution Example

DCOVA

Example: A manufacturer of insulation randomly selects 20 winter days and records the daily high temperature

24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27

Organizing Numerical Data: Frequency Distribution Example

DCOVA

- Sort raw data in ascending order:
  12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
- Find range: **58 - 12 = 46**
- Select number of classes: **5 (usually between 5 and 15)**
- Compute class interval (width): **10 (46/5 then round up)**
- Determine class boundaries (limits):
  - **Class 1: 10 but less than 20**
  - **Class 2: 20 but less than 30**
  - **Class 3: 30 but less than 40**
  - **Class 4: 40 but less than 50**
  - **Class 5: 50 but less than 60**
- Compute class midpoints: **15, 25, 35, 45, 55**
- Count observations & assign to classes

Organizing Numerical Data: Frequency Distribution Example

DCOVA

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class	Midpoints	Frequency
10 but less than 20	15	3
20 but less than 30	25	6
30 but less than 40	35	5
40 but less than 50	45	4
50 but less than 60	55	2
Total		20

Organizing Numerical Data: Relative & Percent Frequency Distribution Example

DCOVA

Class	Frequency	Relative Frequency	Percentage
10 but less than 20	3	.15	15%
20 but less than 30	6	.30	30%
30 but less than 40	5	.25	25%
40 but less than 50	4	.20	20%
50 but less than 60	2	.10	10%
Total	20	1.00	100%

Relative Frequency = Frequency / Total, e.g. 0.10 = 2 / 20

Organizing Numerical Data: Cumulative Frequency Distribution Example

DCOVA

Class	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
10 but less than 20	3	15%	3	15%
20 but less than 30	6	30%	9	45%
30 but less than 40	5	25%	14	70%
40 but less than 50	4	20%	18	90%
50 but less than 60	2	10%	20	100%
Total	20	100	20	100%

Cumulative Percentage = Cumulative Frequency / Total * 100 e.g. 45% = 100*9/20

Why Use a Frequency Distribution?

DCOVA

- It condenses the raw data into a more useful form
- It allows for a quick visual interpretation of the data
- It enables the determination of the major characteristics of the data set including where the data are concentrated / clustered

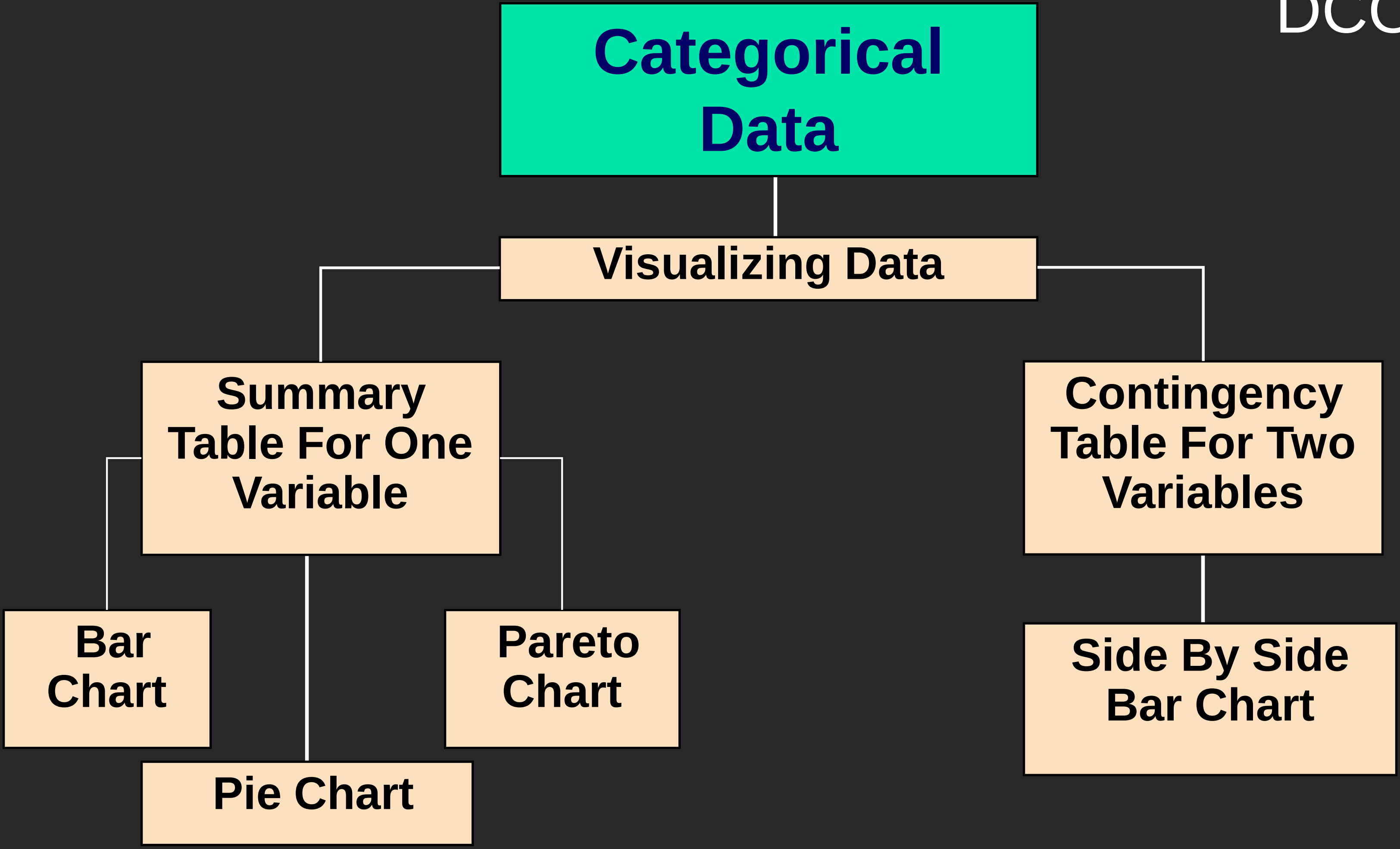
Frequency Distributions: Some Tips

DCOVA

- Different class boundaries may provide different pictures for the same data (especially for smaller data sets)
- Shifts in data concentration may show up when different class boundaries are chosen
- As the size of the data set increases, the impact of alterations in the selection of class boundaries is greatly reduced
- When comparing two or more groups with different sample sizes, you must use either a relative frequency or a percentage distribution

Visualizing Categorical Data Through Graphical Displays

DCOVA

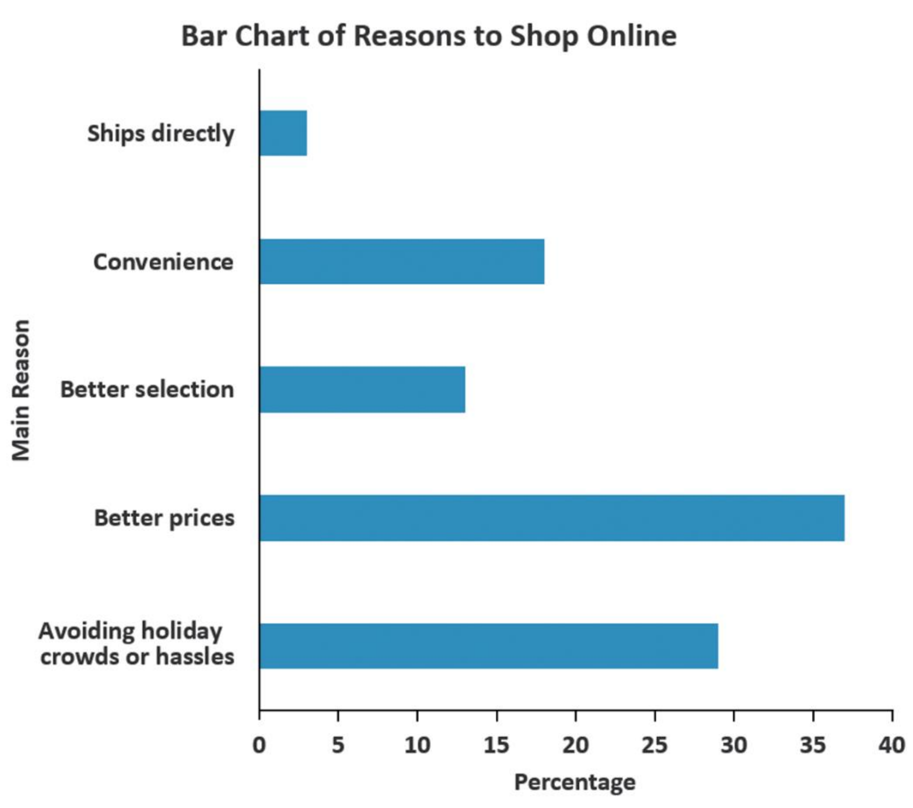


Visualizing Categorical Data: The Bar Chart

DCOVA

- The **bar chart** visualizes a categorical variable as a series of bars. The length of each bar represents either the frequency or percentage of values for each category. Each bar is separated by a space called a gap.

Reason For Shopping Online?	Percent
Better Prices	37%
Avoiding holiday crowds or hassles	29%
Convenience	18%
Better selection	13%
Ships directly	3%

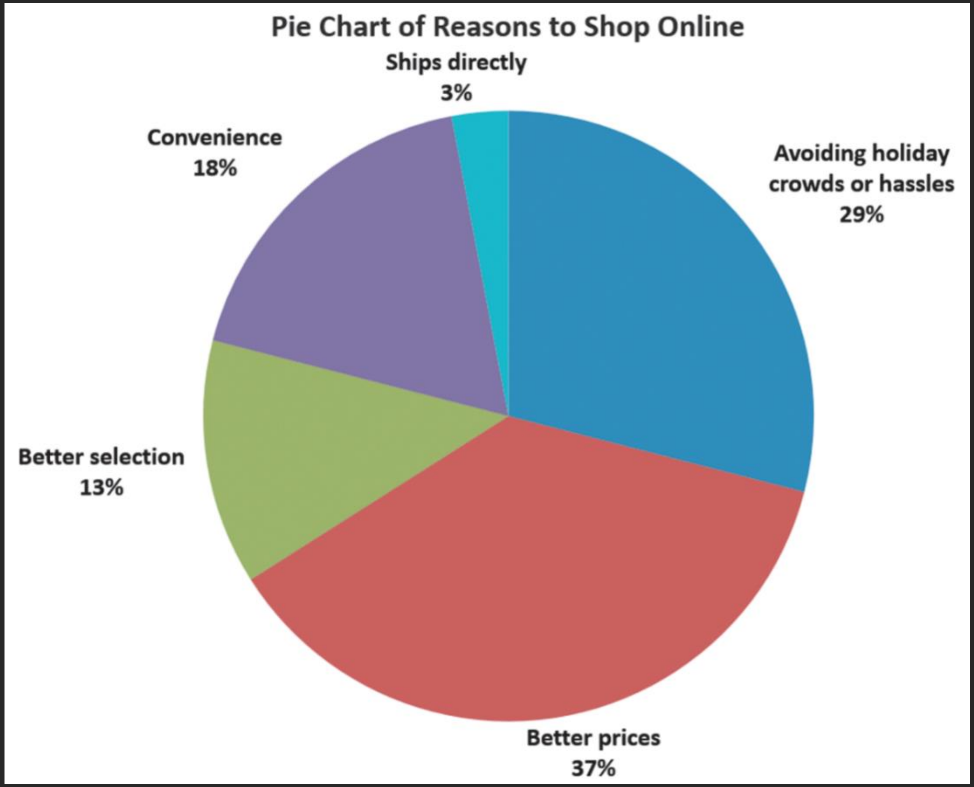


Visualizing Categorical Data: The Pie Chart

DCOVA

- The **pie chart** is a circle broken up into slices that represent categories. The size of each slice of the pie varies according to the percentage in each category.

Reason For Shopping Online?	Percent
Better Prices	37%
Avoiding holiday crowds or hassles	29%
Convenience	18%
Better selection	13%
Ships directly	3%



Visualizing Categorical Data: The Pareto Chart

DCOVA

- Used to portray categorical data
- A vertical bar chart, where categories are shown in descending order of frequency
- A cumulative polygon is shown in the same graph
- Used to separate the “vital few” from the “trivial many”

Visualizing Categorical Data: The Pareto Chart (con't)

DCOVA

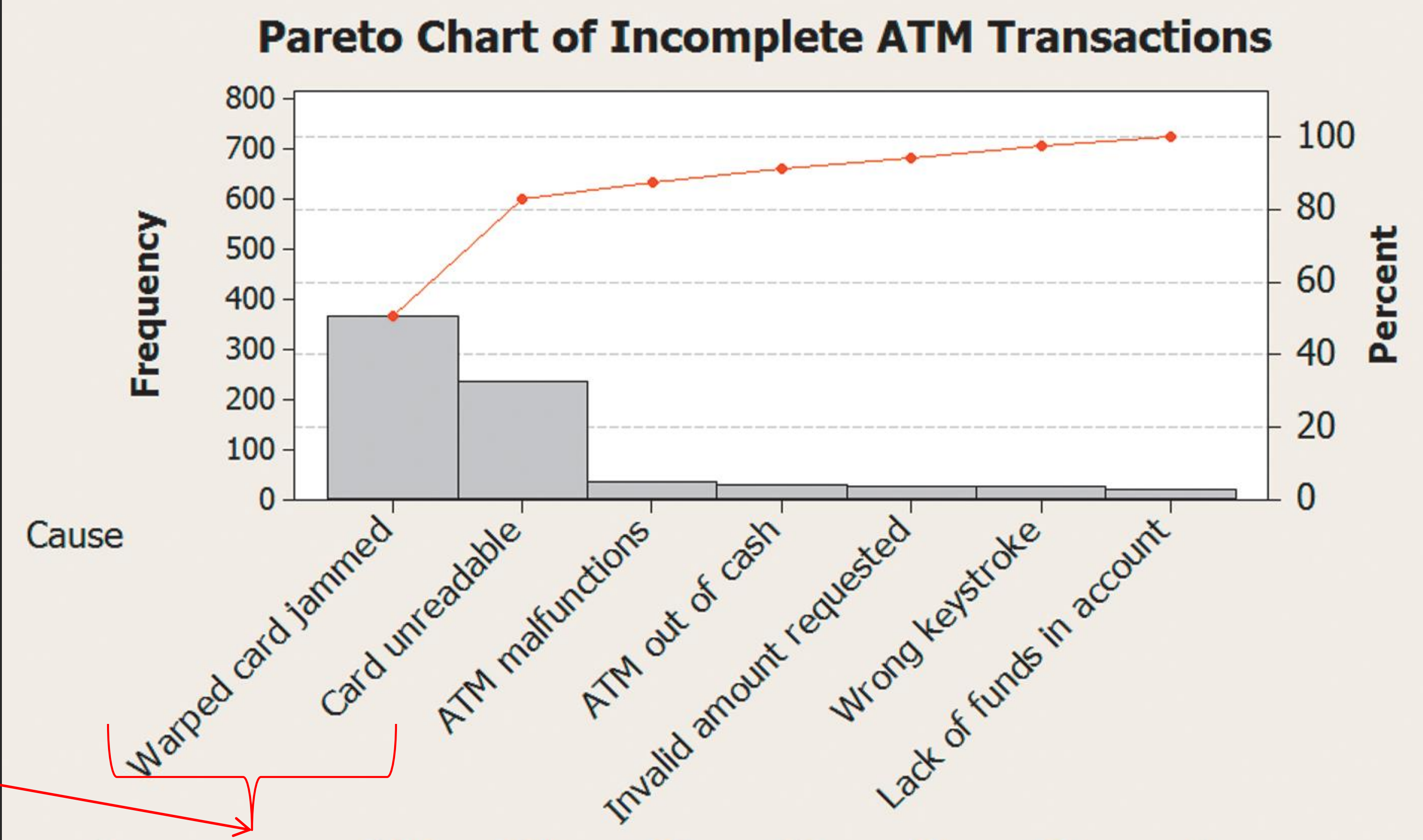
Ordered Summary Table For Causes Of Incomplete ATM Transactions

Cause	Frequency	Percent	Cumulative Percent
Warped card jammed	365	50.41%	50.41%
Card unreadable	234	32.32%	82.73%
ATM malfunctions	32	4.42%	87.15%
ATM out of cash	28	3.87%	91.02%
Invalid amount requested	23	3.18%	94.20%
Wrong keystroke	23	3.18%	97.38%
Lack of funds in account	19	2.62%	100.00%
Total	**724**	**100.00%**	

Source: Data extracted from A. Bhalla, "Don't Misuse the Pareto Principle," *Six Sigma Forum Magazine, May 2009, pp. 15–18.*

Visualizing Categorical Data: The Pareto Chart (con’t)

DCOVA



The “Vital Few”

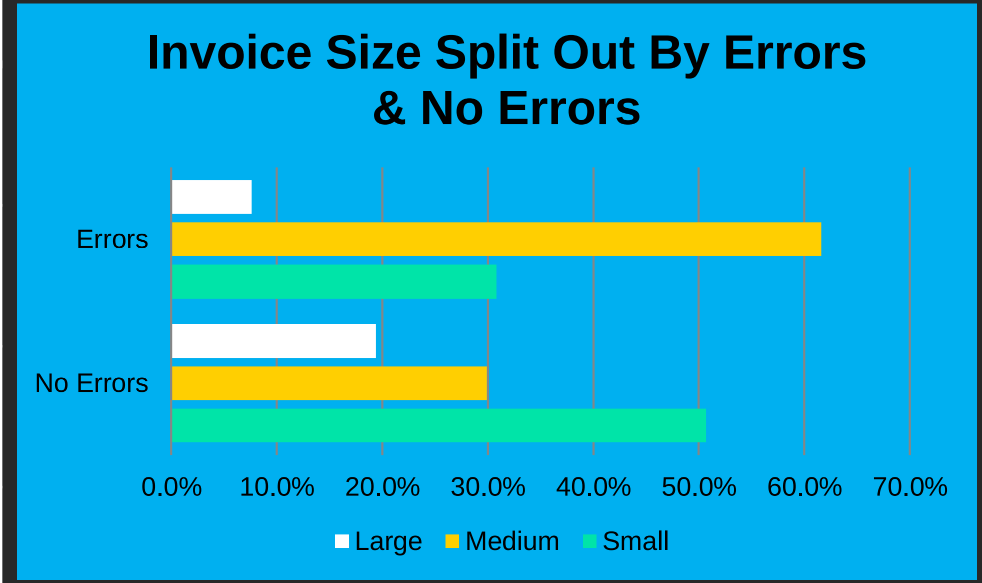
Cause	Warped card jammed	Card unreadable	ATM malfunctions	ATM out of cash	Invalid amount requested	Wrong keystroke	Lack of funds in account
Frequency	365	234	32	28	23	23	19
Percent	50.4	32.3	4.4	3.9	3.2	3.2	2.6
Cum %	50.4	82.7	87.2	91.0	94.2	97.4	100.0

Visualizing Categorical Data: Side By Side Bar Charts

DCOVA

- The **side by side bar chart** represents the data from a contingency table.

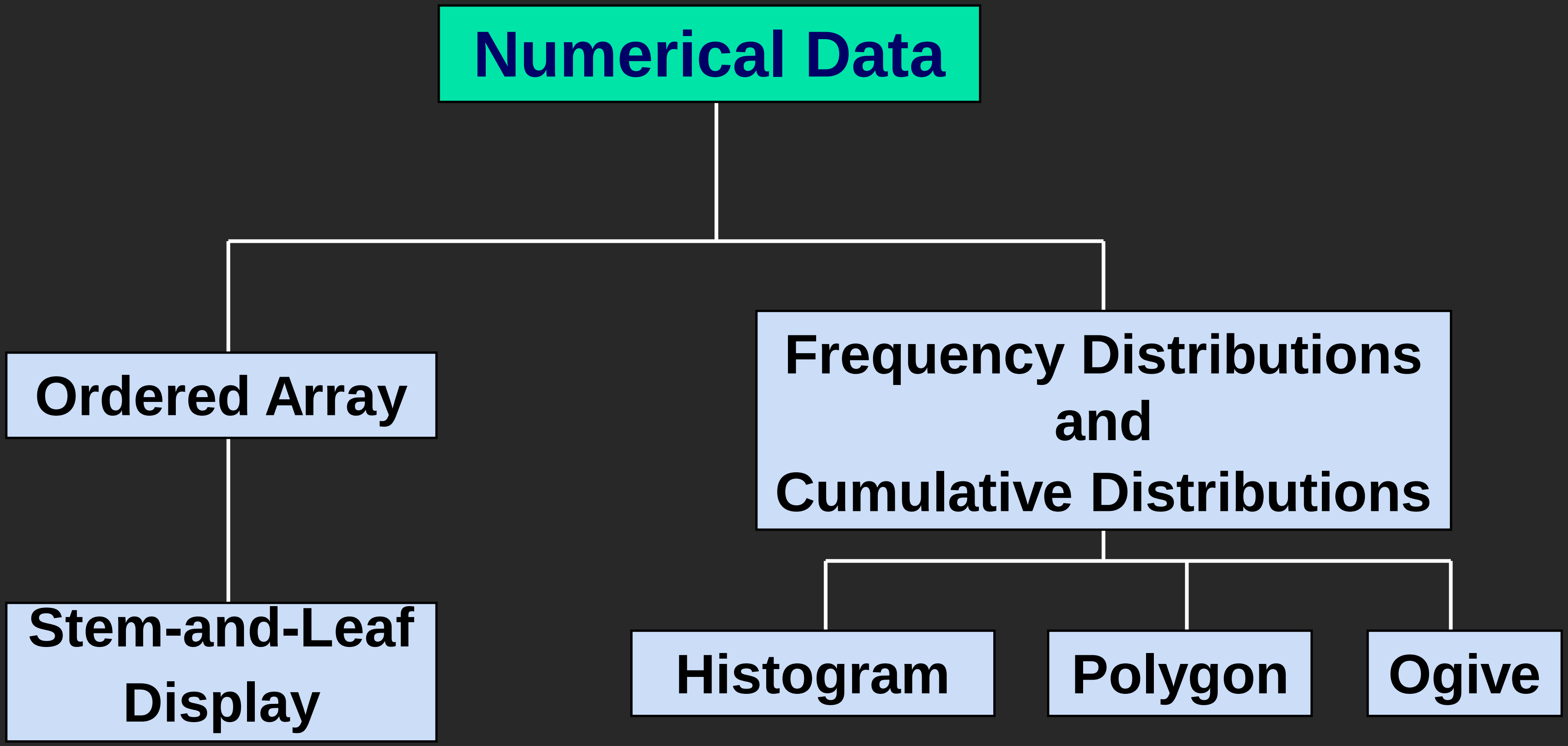
	No Errors	Errors	Total
Small Amount	50.75%	30.77%	47.50%
Medium Amount	29.85%	61.54%	35.00%
Large Amount	19.40%	7.69%	17.50%
Total	100.0%	100.0%	100.0%



Invoices with errors are much more likely to be of medium size (61.54% vs 30.77% and 7.69%)

Visualizing Numerical Data By Using Graphical Displays

DCOVA



Numerical Data

Ordered Array

Stem-and-Leaf Display

Frequency Distributions and Cumulative Distributions

Histogram

Polygon

Ogive

Stem-and-Leaf Display

DCOVA

- A simple way to see how the data are distributed and where concentrations of data exist

METHOD: Separate the sorted data series into leading digits (the **stems**) and the trailing digits (the **leaves**)

Organizing Numerical Data: Stem and Leaf Display

DCOVA

- A **stem-and-leaf display** organizes data into groups (called stems) so that the values within each group (the leaves) branch out to the right on each row.

Age of Surveyed College Students	**Day Students**					
	16	17	17	18	18	18
	19	19	20	20	21	22
	22	25	27	32	38	42
	Night Students					
	18	18	19	19	20	21
	23	28	32	33	41	45

Age of College Students

Day Students

Stem	**Leaf**
1	67788899
2	0012257
3	28
4	2

Night Students

Stem	**Leaf**
1	8899
2	0138
3	23
4	15

Visualizing Numerical Data: The Histogram

DCOVA

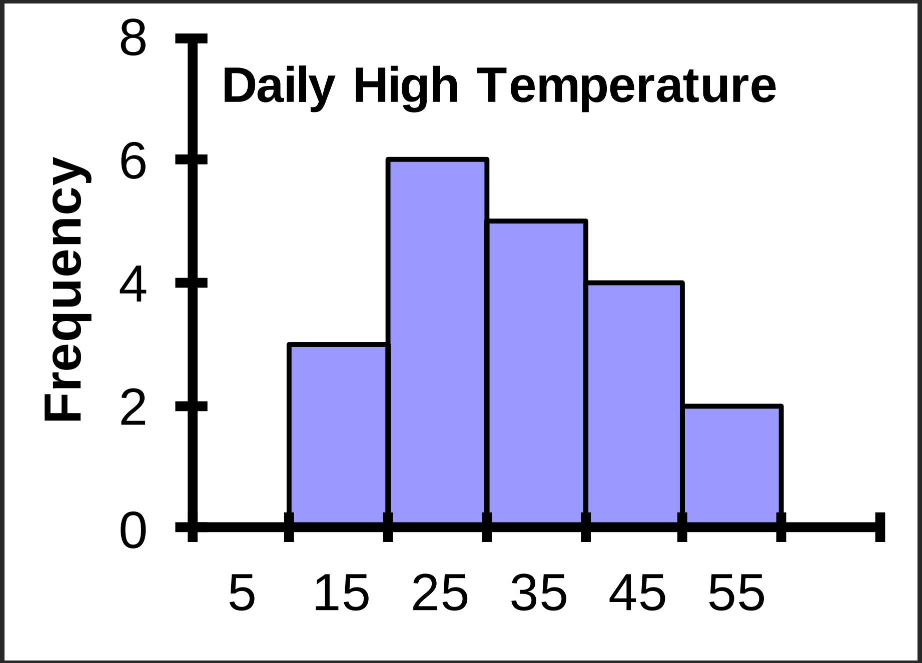
- A vertical bar chart of the data in a frequency distribution is called a **histogram.**
- In a histogram there are no gaps between adjacent bars.
- The **class boundaries** (or **class midpoints**) are shown on the horizontal axis.
- The vertical axis is either **frequency**, **relative frequency**, or **percentage**.
- The height of the bars represent the frequency, relative frequency, or percentage.

Visualizing Numerical Data: The Histogram

DCOVA

Class	Frequency	Relative Frequency	Percentage
10 but less than 20	3	.15	15
20 but less than 30	6	.30	30
30 but less than 40	5	.25	25
40 but less than 50	4	.20	20
50 but less than 60	2	.10	10
Total	20	1.00	100

(In a percentage histogram the vertical axis would be defined to show the percentage of observations per class)



Visualizing Numerical Data: The Polygon

DCOVA

- A **percentage polygon** is formed by having the midpoint of each class represent the data in that class and then connecting the sequence of midpoints at their respective class percentages.

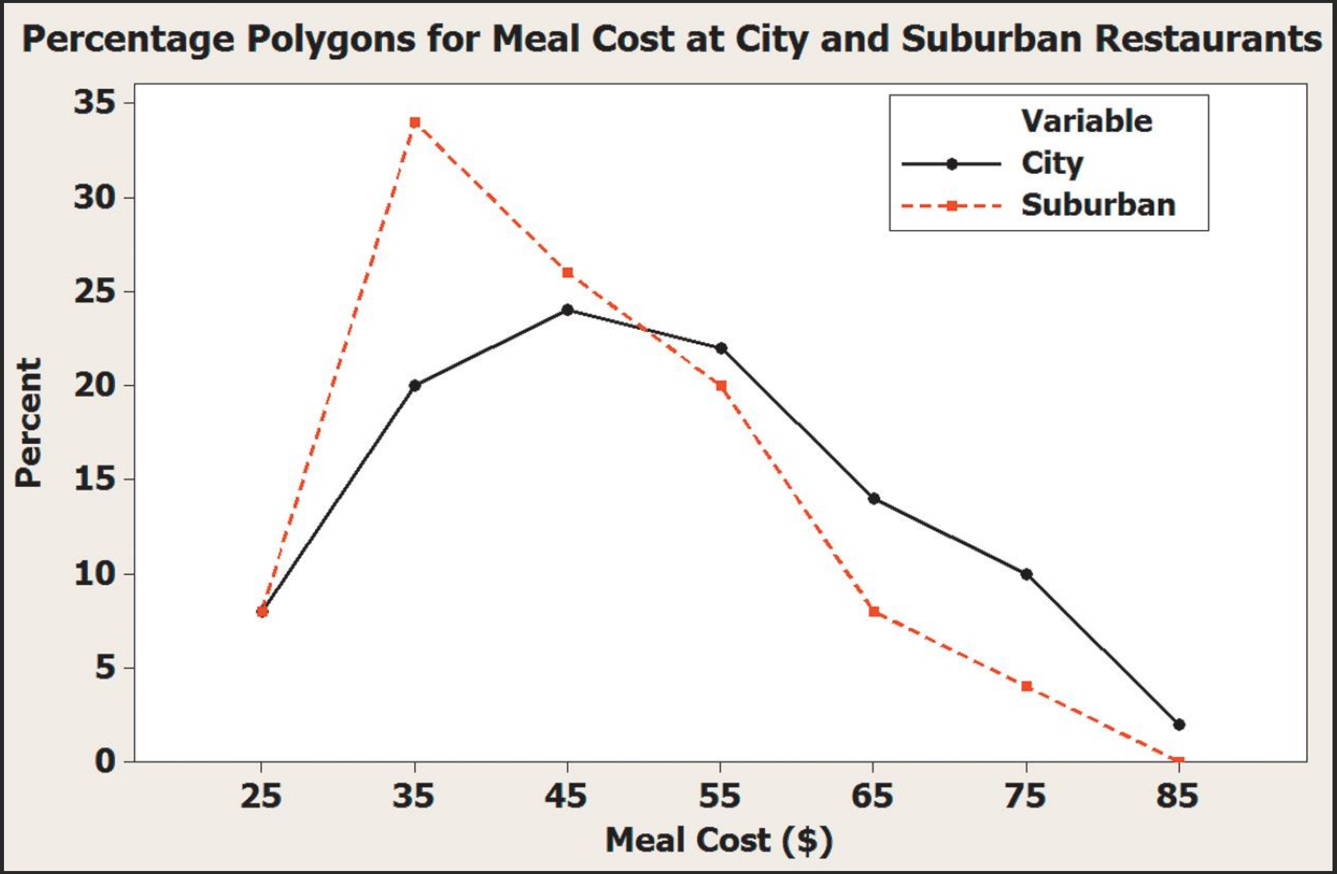
- The **cumulative percentage polygon,** or **ogive,** displays the variable of interest along the X axis, and the cumulative percentages along the Y axis.

- Useful when there are two or more groups to compare.

Visualizing Numerical Data: The Percentage Polygon

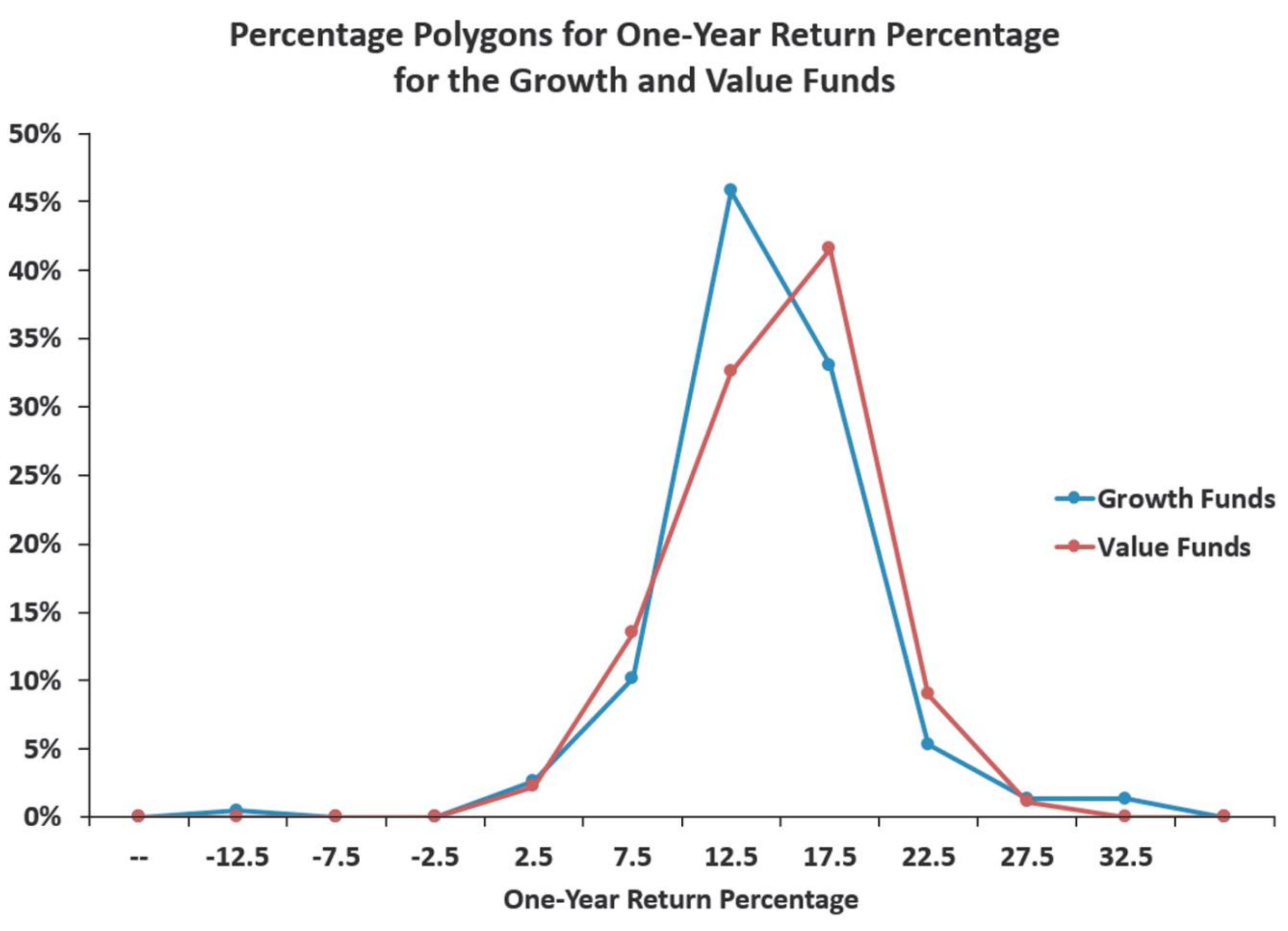
DCOVA

Useful When Comparing Two or More Groups



Visualizing Numerical Data: The Percentage Polygon

DCOVA



Chapter Summary

In this chapter we covered:

- Methods to organize variables.
- Methods to visualize variables.
- Methods to organize or visualize more than one variable at the same time.
- Principles of proper visualizations.