

Graphical Descriptive Techniques 1

Chapter 2 - Graphical Descriptive Techniques 1



"I can prove it or disprove it! What do you want me to do?"

1

Data and Data Sets

➤ **Data** are the facts and figures collected, analyzed, and summarized for presentation and interpretation.

➤ All the data collected in a particular study are referred to as the **data set** for the study.

Elements, Variables, and Observations

➤ **Elements** are the entities on which data are collected.

➤ A **variable** is a characteristic of interest for the elements

➤ The set of measurements obtained for a particular element is called an **observation**.

➤ A data set with n elements contains n observations.

➤ The total number of data values in a complete data set is the number of elements multiplied by the number of variables.

2

2

Data, Data Sets, Elements, Variables, and Observations

Element Names Company	Variables		
	Stock Exchange	Annual Sales(\$M)	Earn/Share(\$)
Dataram	NQ	73.10	0.86
EnergySouth	N	74.00	1.67
Keystone	N	365.70	0.86
LandCare	NQ	111.40	0.33
Psychemedics	N	17.60	0.13

3

3

Scales of Measurement

The scale determines the amount of information contained in the data and the statistical analyses that are most appropriate.

■ Nominal

Data are **labels or names** used to identify an attribute of the element.

A **nonnumeric label** or **numeric code** may be used.

Example:

Students of a university are classified by the school in which they are enrolled using a nonnumeric label such as Business, Humanities, Education, and so on.

Alternatively, a numeric code could be used for the school variable (e.g. 1 denotes Business, 2 denotes Humanities, 3 denotes Education, and so on).

4

4

Graphical Descriptive Techniques 1

Scales of Measurement

■Ordinal

The data have the properties of nominal data and the order or rank of the data is meaningful.

A nonnumeric label or numeric code may be used.

Example:

Students of a university are classified by their class standing using a nonnumeric label such as Freshman, Sophomore, Junior, or Senior.

Alternatively, a numeric code could be used for the class standing variable (e.g. 1 denotes Freshman, 2 denotes Sophomore, and so on).

5

5

Scales of Measurement

■Interval

The data have the properties of ordinal data, and the interval between observations is expressed in terms of a fixed unit of measure.

Interval data are always numeric.

Example:

Melissa has an SAT score of 1885, while Kevin has an SAT score of 1780. Melissa scored 105 points more than Kevin.

6

6

Data can be further classified as being categorical or quantitative.

Categorical Data

Labels or names used to identify an attribute of each element

Often referred to as qualitative data

Use either the nominal or ordinal scale of measurement

Can be either numeric or nonnumeric

Appropriate statistical analyses are rather limited

7

7

Quantitative Data

Quantitative data indicate how many or how much

discrete, if measuring how many

continuous, if measuring how much

Quantitative data are always numeric.

Ordinary arithmetic operations are meaningful for quantitative data.

Example : *Latte*

Qualitative data: robust aroma frothy appearance strong taste burgundy cup

Example : *Latte*

Quantitative data: 12 ounces of latte serving temperature 150° F. serving cup 7 inches in height cost \$4.95

8

8

Graphical Descriptive Techniques 1

Summarizing Data

- Frequency Distribution
- Relative Frequency and Percent Frequency Distributions
- Bar Chart
- Pie Chart

9

Categorical data use labels or names to identify categories of like items.

Quantitative data are numerical values that indicate how much or how many.

Frequency Distribution

A **frequency distribution** is a tabular summary of data showing the frequency (or number) of items in each of several non-overlapping classes.

The objective is to provide insights about the data that cannot be quickly obtained by looking only at the original data.

10

Frequency Distribution

■ Example: Marada Inn

Guests staying at Marada Inn were asked to rate the quality of their accommodations as being *excellent*, *above average*, *average*, *below average*, or *poor*. The ratings provided by a sample of 20 guests are:

Below Average	Average	Above Average
Above Average	Above Average	Above Average
Above Average	Below Average	Below Average
Average	Poor	Poor
Above Average	Excellent	Above Average
Average	Above Average	Average
Above Average	Average	

11

■ Excel Formula Worksheet

	A	B	C	D
1	Quality Rating		Quality Rating	Frequency
2	Above Average		Poor	=COUNTIF(\$A\$2:\$A\$21, C2)
3	Below Average		Below Average	=COUNTIF(\$A\$2:\$A\$21, C3)
4	Above Average		Average	=COUNTIF(\$A\$2:\$A\$21, C4)
5	Average		Above Average	=COUNTIF(\$A\$2:\$A\$21, C5)
6	Average		Excellent	=COUNTIF(\$A\$2:\$A\$21, C6)
7	Above Average		Total	=SUM(D2:D6)
8	Above Average			

	A	B	C	D
1	Quality Rating		Quality Rating	Frequency
2	Above Average		Poor	2
3	Below Average		Below Average	3
4	Above Average		Average	5
5	Average		Above Average	9
6	Average		Excellent	1
7	Above Average		Total	20
8	Above Average			

Note: Rows 9-21 are not shown.

12

Graphical Descriptive Techniques 1

Frequency Distribution

■ Example: Marada Inn

<u>Rating</u>	<u>Frequency</u>
Poor	2
Below Average	3
Average	5
Above Average	9
Excellent	<u>1</u>
Total	20

13

13

Relative Frequency Distribution

The relative frequency of a class is the fraction or proportion of the total number of data items belonging to the class.

A relative frequency distribution is a tabular summary of a set of data showing the relative frequency for each class.

Percent Frequency Distribution

The percent frequency of a class is the relative frequency multiplied by 100.

14

14

Relative Frequency and Percent Frequency Distributions

■ Example: Marada Inn

<u>Rating</u>	<u>Relative Frequency</u>	<u>Percent Frequency</u>
Poor	.10	10
Below Average	.15	15
Average	.25	25
Above Average	.45	45
Excellent	<u>.05</u>	<u>5</u>
Total	1.00	100

$.10(100) = 10$

$1/20 = .05$

15

15

But what if the sample size is much larger?

[Chapter2DataFiles\HotelRatings.xlsx](#)

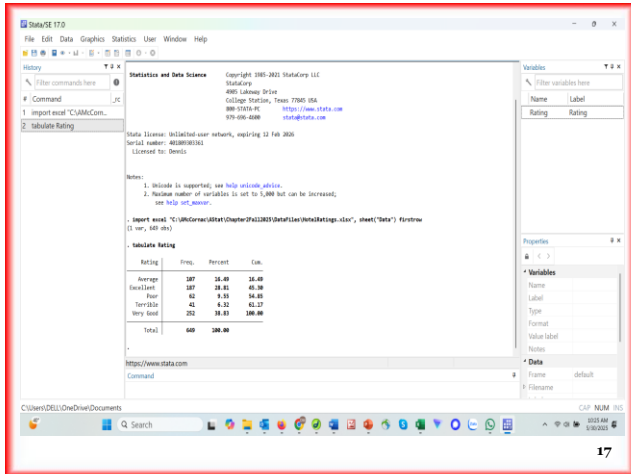
Looking at the Soft Drink Market

[Chapter2DataFiles\SoftDrink.xlsx](#)

16

16

Graphical Descriptive Techniques 1



17

Frequency Distribution Quantitative Data

Example

Sanderson and Clifford, a small public accounting firm wants to determine time in days required to complete year end audits. It takes a sample of 20 clients.

12	14	19	18
15	15	18	17
20	27	22	23
22	21	33	28
14	18	16	13

[DataFiles\Audit.xlsx](#)

18

Frequency Distribution

Example: Sanderson and Clifford

- If we choose five classes:
- Approximate Class Width = $(33 - 12)/5 = 4.2 \approx 4$

Time in days	Frequency
10-14	4
15-19	8
20-24	5
25-29	2
30-34	1
Total	20

19

Using Excel's PivotTable to Construct a Pivot Table

- **Step 1** – Select any cell in the data set
- **Step 2** – Click **Insert** on the Ribbon
- **Step 3** – In the **Tables** group click **Pivot Table**
- **Step 4** - When the **Create PivotTable** dialog box appears: Click
(a PivotTable and PivotTable Fields dialog box will appear in a new worksheet)

20

Graphical Descriptive Techniques 1

Using Excel's *PivotTable* to Construct a Pivot Table

- **Step 5** - In the PivotTable Fields dialog box:
 Drag **Audit time** to the **Rows** area
 Drag **Audit time** to the **Values** area
- **Step 6** - Click on **Sum of Audit time** in the **Values** area
- **Step 7** - Click **Value Field Settings** from the list of options
- **Step 8** - When the **Value Field Settings** dialog box appears:
 Under **Summarize value field by**, choose **Count** Click **OK**

21

21

Using *Excel's Pivot table* to construct a frequency distribution

A	B	C	D	E	F
3	Row Labels	Count of audit time			
4	10-14	4			
5	15-19	8			
6	20-24	5			
7	25-29	2			
8	30-34	1			
9	Grand total	20			

22

22

Using Excel's *PivotTable* to Construct a Frequency Distribution

To construct the frequency distribution, we must group the rows containing audit time.

- **Step 1** - Right click any cell in the PivotTable report containing an audit time.
- **Step 2** - Choose **Group** from the list of options that appears
- **Step 3** - When the Grouping dialog box appears:
 - ✓ Enter 10 in the **Starting at** box
 - ✓ Enter 34 in the **Ending at** box
 - ✓ Enter 5 in the **By** box
 - ✓ Click **OK**

23

23

Relative Frequency and Percent Frequency Distributions

Example: Sanderson and Clifford

Audit time (in days)	Relative Frequency	Percent Frequency
10 – 14	.20 (4/20)	20 (0.2 * 100)
15 – 19	.40	40
20 – 25	.25	25
25 – 29	.10	10
30 – 34	.05	5
	Total 1.00	100

24

24

Graphical Descriptive Techniques 1

Relative Frequency and Percent Frequency Distributions

Example: Sanderson and Clifford

Insights obtained from the Percent Frequency Distribution:

- ✓ 40% of the audits required from 15 to 19 days.
- ✓ Another 25% of the audits required 20 to 25 days.
- ✓ Only 5% of the audits required more than 30 days.

25

Example– Frequency Distribution

Work Status in the General Social Survey

GSS asked: "Last week were you working full time, part time, going to school, keeping house, or what?"

The responses were as follows:

1. Working full time
2. Working part time
3. Temporarily not working
4. Unemployed, laid off
5. Retired
6. School
7. Keeping house
8. Other

DATA: GSS2018

Variable: WRKSTAT (Column X)

WORK STATUS	CODE	FREQUENCY	RELATIVE FREQUENCY (%)
Working full time	1	1,134	48.34
Working part time	2	259	11.04
Temporarily not working	3	53	2.26
Unemployed, laid off	4	84	3.58
Retired	5	445	18.97
School	6	81	3.45
Keeping house	7	242	10.32
Other	8	48	2.05
Total		2,346	100

[Chapter2DataFiles\GSS2018.xlsx](#)

26

25

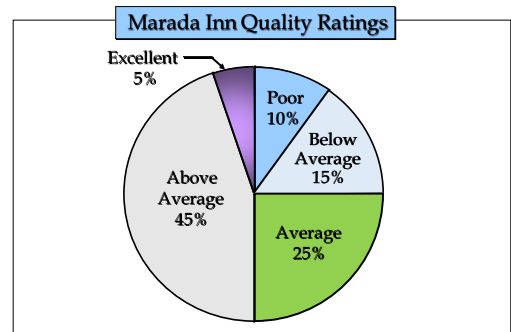
26

Pie Chart

- The pie chart is a commonly used graphical device for presenting relative frequency and percent frequency distributions for categorical data.
- First draw a circle; then use the relative frequencies to subdivide the circle into sectors that correspond to the relative frequency for each class.
- Since there are 360 degrees in a circle, a class with a relative frequency of .25 would consume $.25(360) = 90$ degrees of the circle.

27

Pie Chart



28

27

28

Graphical Descriptive Techniques 1

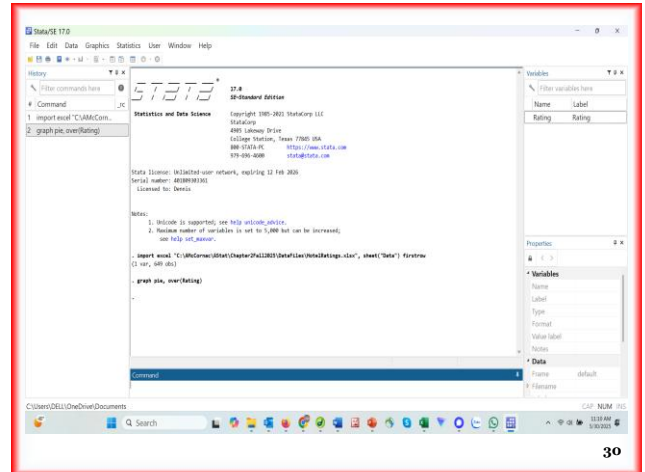
But what if the sample size is much larger?

[DataFiles\HotelRatings.xlsx](#)

Looking at the Soft Drink Market

[DataFiles\SoftDrink.xlsx](#)

29



30

Pie Chart

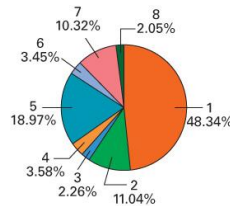
Work Status in the General Social Survey

A pie chart is used to show relative frequencies. The pie chart is created by segmenting a circle into several wedges, one for each category.

The angle on each wedge is proportional to the respective relative frequency. The size of the circle is arbitrary.

Interpret:

Only 48.3% of respondents were working full time. Among the rest, 19% were retired, 11% working part time, and 10.3% keeping house. The remaining were in one of the other four categories.



[Chapter2DataFiles\GSS2018.xlsx](#)

31

Energy Consumption in the USA in 2019

Use an appropriate graphical technique to depict the total energy consumption of the USA from all sources in 2019.

DATA: Xm02-02

Solution:

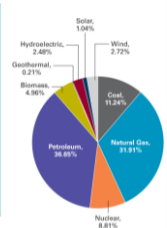
The pie chart is the appropriate method to describe the proportion of total energy consumption from each source.

Interpret:

The USA depends heavily on the fossil fuels (80%). Renewables amount to about 11% of the total consumption.

ENERGY SOURCES	QUADRILLIONS OF BTUs
Nonrenewable Energy Sources	
Coal	11.31
Natural Gas	32.10
Nuclear	0.862
Petroleum	36.67
Renewable Energy Sources	
Biomass	4.985
Geothermal	2.093
Hydroelectric	2.492
Solar	1.043
Wind	2.732
Total	100.6

Source: U.S. Energy Information Administration



[Chapter2DataFiles\Xm02-02.xlsx](#)

32

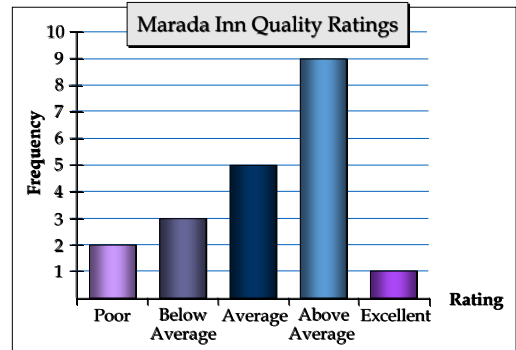
Graphical Descriptive Techniques 1

Bar Chart

- A bar chart is a graphical device for depicting qualitative data.
- On one axis (usually the horizontal axis), we specify the labels that are used for each of the classes.
- A frequency, relative frequency, or percent frequency scale can be used for the other axis (usually the vertical axis).
- Using a bar of fixed width drawn above each class label, we extend the height appropriately.
- The bars are separated to emphasize the fact that each class is a separate category.

33

Bar Chart



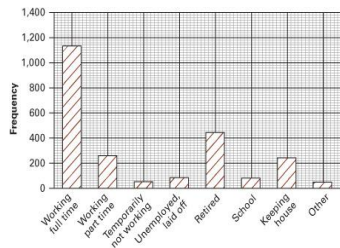
34

Bar Chart

Work Status in the General Social Survey

Graphical techniques generally catch a reader's eye more quickly than a table of numbers does. A bar chart is used to display frequencies.

The bar chart is created by drawing a rectangle representing each category. The height of the rectangle represents the frequency. The width of the base is arbitrary.



[Chapter2DataFiles\GSS2018.xlsx](#)

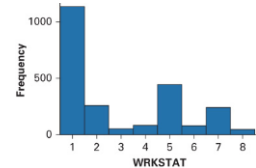
35

Frequency Distribution and Bar Chart

Example 2.1 – Stata Output and Instructions

[Chapter2Stata instructions.pdf](#)

WRKSTAT	Freq.	Percent	Cum.
1	1,134	48.34	48.34
2	259	11.04	59.38
3	53	2.26	61.64
4	84	3.58	65.22
5	445	18.97	84.19
6	81	3.45	87.64
7	242	10.32	97.95
8	48	2.05	100.00
Total	2,346	100.00	



36

Graphical Descriptive Techniques 1

Cumulative Distributions

Cumulative frequency distribution – shows the *number* of items with values less than or equal to the upper limit of each class..

Cumulative relative frequency distribution – shows the *proportion* of items with values less than or equal to the upper limit of each class.

Cumulative percent frequency distribution – shows the *percentage* of items with values less than or equal to the upper limit of each class.

37

37

Cumulative Distributions

Example: Sanderson and Cliffords

Audit time (Days)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percent Frequency
≤ 14	4	.20	20
≤ 19	12	.60	60
≤ 24	17	.85	85
≤ 29	19	.95	95
≤ 34	20	1.00	100

38

38

Crosstabulation

- A **crosstabulation** is a tabular summary of data for two variables.
- Crosstabulation can be used when:
 - one variable is qualitative and the other is quantitative,
 - both variables are qualitative, or
 - both variables are quantitative.
- The left and top margin labels define the classes for the two variables.

39

39

Crosstabulation

■ Example: Finger Lakes Homes

The number of Finger Lakes homes sold for each style and price for the past two years is shown below.

Price Range	Home Style				Total
	Colonial	Log	Split	A-Frame	
< \$200,000	18	6	19	12	55
≥ \$200,000	12	14	16	3	45
Total	30	20	35	15	100

40

40

Graphical Descriptive Techniques 1

Crosstabulation

- Example: Finger Lakes Homes

Price Range	Home Style				Total
	Colonial	Log	Split	A-Frame	
< \$200,000	18	6	19	12	55
≥ \$200,000	12	14	16	3	45
Total	30	20	35	15	100

Frequency distribution for the price range variable

Frequency distribution for the home style variable

41

Crosstabulation: Row Percentages

- Example: Finger Lakes Homes

Price Range	Home Style				Total
	Colonial	Log	Split	A-Frame	
< \$200,000	32.73	10.91	34.55	21.82	100
≥ \$200,000	26.67	31.11	35.56	6.67	100

Note: row totals are actually 100.01 due to rounding.

(Colonial and ≥ \$200K) / (All ≥ \$200K) x 100 = (12/45) x 100

42

Crosstabulation: Column Percentages

- Example: Finger Lakes Homes

Price Range	Home Style			
	Colonial	Log	Split	A-Frame
< \$200,000	60.00	30.00	54.29	80.00
≥ \$200,000	40.00	70.00	45.71	20.00
Total	100	100	100	100

(Colonial and ≥ \$200K) / (All Colonial) x 100 = (12/30) x 100

- Converting the entries in the table into row percentages or column percentages can provide additional insight about the relationship between the two variables.

43

Newspaper Readership Survey

A sample of newspaper readers was asked to report which newspaper they read: Globe and Mail (1) Post (2), Star (3), Sun (4), and to indicate whether they were blue-collar worker (1), white-collar worker (2), or professional (3).

DATA: Xm02-04

Determine if the two variables are related.

Solution:

By counting the number of times each of the 12 combinations occurs, we produce the cross-classification table of frequencies.

To investigate if occupation and newspaper are related, we convert the frequencies in each row to relative frequencies as shown in the cross-classification table of row proportions.

Notice how the relative frequencies in rows 2 and 3 are similar and how there are large differences between row 1 and rows 2 and 3. This tells us that the variables occupation and newspaper are related.

Cross-tabulation of Frequencies

OCCUPATION	NEWSPAPER				TOTAL
	GAM	POST	STAR	SUN	
Blue collar	27	18	38	37	120
White collar	29	43	21	15	108
Professional	13	51	22	39	125
Total	69	112	81	72	334

Cross-tabulation of Row Proportions

OCCUPATION	NEWSPAPER				TOTAL
	GAM	POST	STAR	SUN	
Blue collar	.23	.15	.32	.31	1.00
White collar	.27	.40	.29	.04	1.00
Professional	.26	.48	.27	.06	1.00
Total	.23	.32	.23	.20	1.00

Chapter2DataFiles\Xm02-04.xlsx

44

Graphical Descriptive Techniques 1

Cross-classification Tables with Excel

Excel can produce a cross-classification table using several methods.

We will use and describe the PivotTable to (1) create the cross-classification table featuring the counts and (2) create a table showing the row relative frequencies.

Instructions:

1. Click **Insert** and **PivotTable**.
2. Select the whole **Table/Range**. Click **OK**.
3. In the **Pivot Table Fields**, check **Reader** and right-click **Add to Values**.
4. Check **Occupation** and right-click **Add to Row Labels**.
5. Check **Newspaper** and right-click **Add to Column Labels**.
6. Place the cursor in the table and right-click **Summarize Values by** and click **Count**.
7. To convert to row percentages, right click any number, click **Show Values As** and click **% of Row Total**.

PivotTable of Frequencies

	A	B	C	D	E	F
1	Count of Reader	Column Labels				
2	Row Labels		1	2	3	4 Grand Total
3	1		27	18	38	37
4	2		29	43	21	15
5	3		33	51	22	20
6	Grand Total		89	112	81	72

PivotTable of Row Proportions

A		B		C	D	E	F
Count of Reader	Column Labels						
Row Labels		1	2	3	4	Grand Total	
1		0.23	0.15	0.32	0.31	1.00	
2		0.27	0.40	0.19	0.14	1.00	
3		0.25	0.40	0.17	0.15	1.00	
Grand Total		0.25	0.32	0.23	0.20	1.00	

45

Crosstabulation Examples

Refer to Restaurant Example - [Chapter2DataFiles\Restaurant.xlsx](#)



A bit difficult to fully digest, but very nutritional and packed with vitamins a, p-hat, x-bar, and especially mu and sigma!

46

Graphing the Relationship Between Two Nominal Variables with Excel

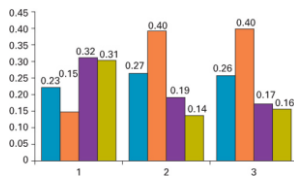
There are several ways to graphically display the relationship between two nominal variables. We have chosen two-dimensional bar charts for each of the three occupations.

The charts can be created from the output of the PivotTable either with counts or with row proportions.

Instructions:
From the cross-classification table, click **Insert** and **Column**. You can do the same from any completed cross-classification table.

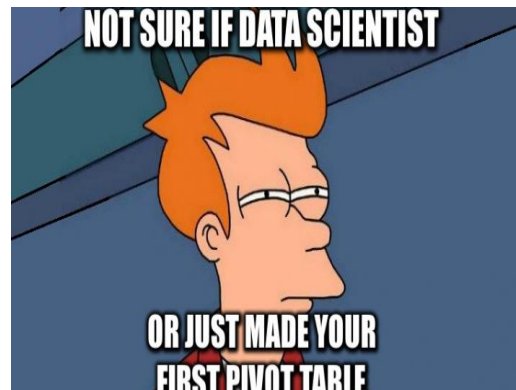
Interpret:

The graphs tell us the same story as the table did. The shapes of the bar charts for occupations 2 and 3 (white collar and professional) are very similar. Both differ considerably from the bar chart for occupation 1 (blue collar).



[DataFiles\Xmo2-04.xlsx](#)

47



48