2 Tabular and Graphical Methods

2.1 Objectives

- Construct and interpret frequency distributions, pie charts, and bar graphs to summarize qualitative data.
- Construct and interpret grouped frequency distributions, cumulative frequency distributions, and histograms to summarize quantitative data.
- Construct side-by-side bar charts to summarize the relationship between two qualitative variables.
- Construct scatterplots to summarize the relationship between two quantitative variables.

2.2 Qualitative Data

The numerical sample statistics for summarizing one or more categorical variables is a proportion or a percent. Frequency distribution tables, pie charts, and bar charts are visual summarizations of qualitative data.

Frequency distribution tables, pie charts, and bar charts must divide a "whole" into categories. All frequencies must sum to the total number of observations, all relative frequencies must sum to 1, and all percent frequencies must sum to 100.

2.2.1 Frequency Distribution Tables

A frequency distribution summarizes qualitative data by grouping the data into categories and recording the number of observations in each category and the relative frequency (proportion) of observations in each category.

Frequency = count of elements

Relative frequency = (count of elements/total number of elements)

Example

Color	Frequency	Relative Frequency
Brown	112	0.0892
Yellow	105	0.0837
Red	109	0.0869
Orange	327	0.2606
Green	314	0.2502
Blue	288	0.2295
Total	1255	1

Table 1 Frequency and Relative Frequency Distribution for the Colors of Plain M&Ms

2.2.2 Pie Charts

A pie chart is a segmented circle whose segments portray the relative frequencies or percent frequencies of the categories of some qualitative variable. The size of the segments are proportional to the values depicted.

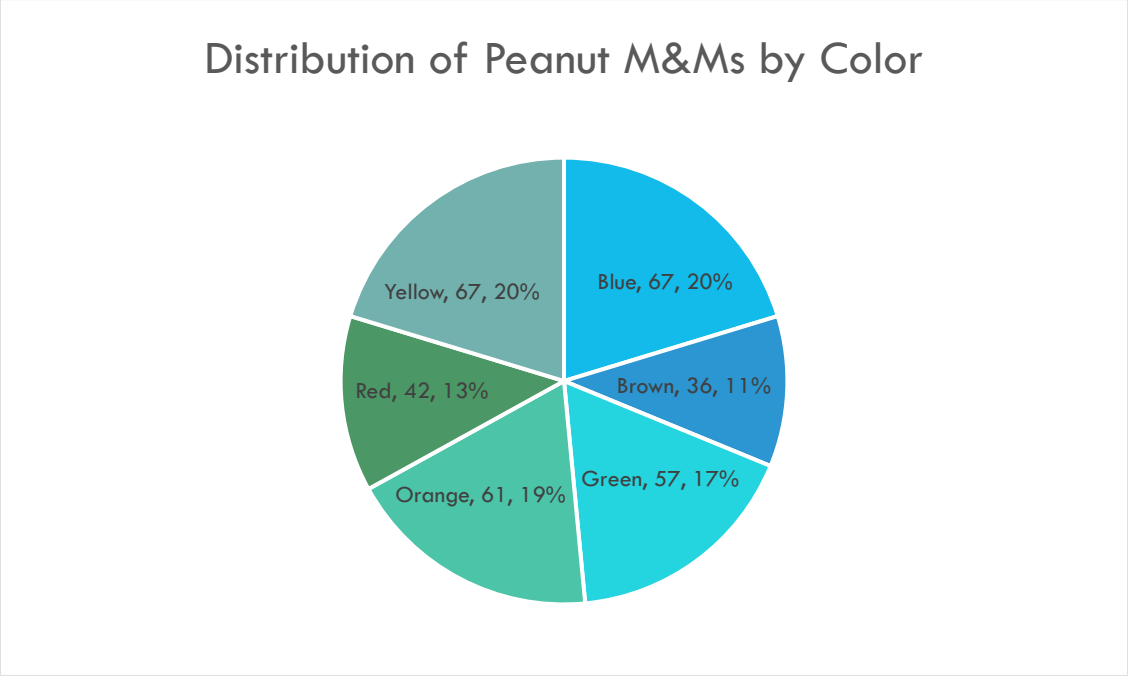

Figure 1 Pie Chart Illustrating the Distribution of Peanut M&Ms by Color

Important Considerations for Creating Pie Charts

- Colors must be distinguishable from one another.
- Legends or data labels (one per wedge) must be clear and easy to read.
- Use a bar chart if two or more categories are equal or differ only slightly. It is difficult to tell which pie slice is greatest or smallest or if they are equal. A bar chart better portrays slight differences in size and equality better than a pie chart.
- Use the simplest graph possible to convey the information in the data. Be clear, clean, and professional.
- Any graphical display of the distribution for a categorical variable should include two important statistics:
    - Percent or frequency for each category
    - Total sample size

  Sample size can be incorporated into a title or caption if it does not appear on the body of the chart.
- The chart must have a title.

2.2.3 Bar Charts

A bar chart is a series of horizontal or vertical bars where the bars portray the frequency or relative frequency for each category of some qualitative variable. The length of the bars are proportional to the values depicted.

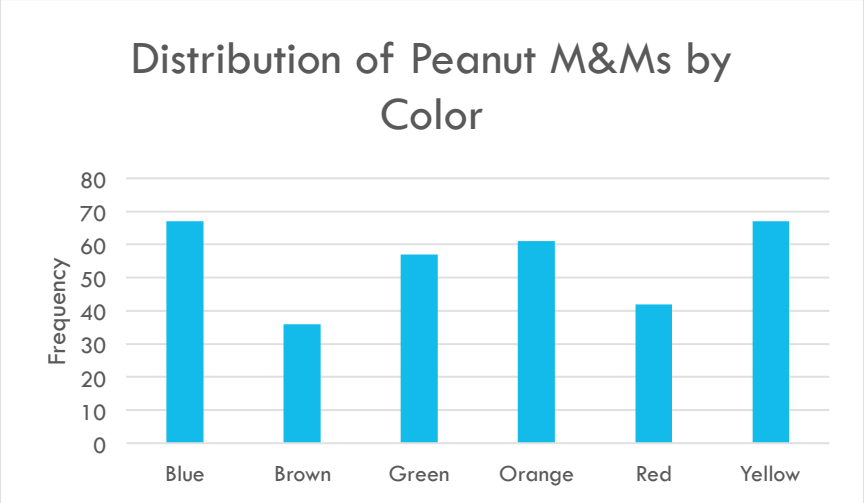

Figure 2 Bar Chart Illustrating the Distribution of Peanut M&Ms by Color

Important Considerations for Creating Bar Charts

- The axes must be marked clearly with numbers or category names. The axis with the frequency (proportion or percentage) measure has a title: Frequency, Relative Frequency (or Proportion), or Percent Frequency (or Percentage).
- The bars must have the same width. Visual distortions will distort the interpretation. Excel should do this automatically.
- The bars of a bar chart must be separated by spaces to create a visual separation of the categories because the items in each category are clearly separate from one another.
- The vertical axis must portray the differences between categories. A high upper limit of the vertical axis compresses the graph so that differences are not noticeable. A low upper limit that creates very small increments between marks stretches the graph so that differences are pronounced.
- Use the simplest graph possible to convey the information in the data. Be clear, clean, and professional.
- Any graphical display of the distribution for a categorical variable should include two important statistics:
  - Percent or frequency for each category
  - Total sample size

  Sample size can be incorporated into a title or caption if it does not appear on the body of the chart.
- The chart must have a title.

2.2.4 Contingency Tables (Cross-tabulation)

A contingency table shows the distribution of one variable in rows and another in columns and is used to study the relationship between the two variables.

When studying two categorical variables, the proportions or percentages should be represented as either row or column percentages to allow comparison of how one variable may influence the responses of the other variable.

Row Labels	Peanut M&M	Plain M&M	Grand Total
Blue	67	288	355
Brown	36	112	148
Green	57	314	371
Orange	61	327	388
Red	42	109	151
Yellow	67	105	172
Grand Total	**330**	**1255**	**1585**

Table 2 Contingency Table Summarizing the Relationship Between M&M Type and Color

2.3 Practice Problems: Qualitative Data

The following exercises are also in the Excel file:
Qualitative Data Practice Problems.xlsx

Answers to the following exercises are in the Excel file:
Qualitative Data Practice Problems KEY.xlsx

2.3.1 Exercise 1: Auto Parts Chain

An auto parts chain asked customers to complete a survey rating the chain's customer service as average, above average, or below average. The following table shows the survey results.

Average	Below Average	Average
Above Average	Above Average	Above Average
Below Average	Average	Average
Above Average	Average	Below Average
Below Average	Below Average	Average

Rating	Frequency	Relative Frequency	Percent Frequency
Below Average	5	0.33	33
Average	6	0.40	40
Above Average	4	0.27	27

1. The proportion of customers who felt the customer service was Average is closest to _____.

2. A rating of Average or Above Average accounted for what number of responses to the survey?

2.3.2 Exercise 2: Professor Smith

Students in Professor Smith's business statistics course have evaluated the overall effectiveness of the professor's instruction on a five-point scale, where a score of 1 indicates very poor performance and a score of 5 indicates outstanding performance. The following table shows the results.

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

Rating	Frequency	Relative Frequency	Percent Frequency
1	1	0.03	3.33
2	5	0.17	16.67
3	7	0.23	23.33
4	8	0.27	26.67
5	9	0.30	30.00

1. What is the most common score given in the evaluations?

2. What percentage of students gave professor Smith an evaluation of at least 4?

3. What percentage of students gave Professor Smith an evaluation of 2 or less?

4. What is the relative frequency of the students who gave Professor Smith an evaluation of 3?

2.3.3 Exercise 3: US Poverty Level

The Statistical Abstract of the United States, 2010 provided the following frequency distribution of the number of people who live below the poverty level by region.

Region	Number of People (in 1000s)	Relative Frequency	Percent Frequency
Northeast	7,174	0.18	17.77
Midwest	8,137	0.20	20.16
South	16,457	0.41	40.77
West	8,593	0.21	21.29

1. What is the percentage of people who live below the poverty level in the West or Midwest?

2.3.4 Exercise 4: City Building Repair

A city in California spent $6 million repairing damage to its public buildings in 2010. The following table shows the categories where the money was directed.

Cause	Frequency	Relative Frequency	Percent Frequency
Termites	1,560,000	0.26	26%
Water Damage	480,000	0.08	8%
Mold	540,000	0.09	9%
Earthquake	1,320,000	0.22	22%
Other	2,100,000	0.35	35%

1. How much did the city spend to fix damage caused by mold?

2. How much more did the city spend to fix damage caused by termites compared to the damage caused by water?

2.3.5 Exercise 5: CBS News Survey

A survey conducted by CBS news asked 1,026 respondents: "What would you do with an unexpected tax refund?" The responses are summarized in the following table.

Category	Frequency	Relative Frequency	Percent Frequency
Pay off debts	523	0.51	51%
Put it in the bank	257	0.25	25%
Spend it	92	0.09	9%
I never get a refund	92	0.09	9%
Other	62	0.06	6%

1. How many people will either put it in the bank or spend it?

2.3.6 Exercise 6: Busy Airports

The world's busiest airports by passenger traffic for 2010…

Name	Location	# of Passengers (in millions)	Relative Frequency	Percent Frequency
Hartsfield-Jackson	Atlanta, Georgia, United States	93	0.26	25.55
Capital International	Beijing, China	76	0.21	20.88
London Heathrow	London, United Kingdom	70	0.19	19.23
O'Hare	Chicago, Illinois, United States	65	0.18	17.86
Tokyo	Tokyo, Japan	60	0.16	16.48

1. The percentage of passenger traffic in the five busiest airports that occurred in Asia is the closest to _____.

2. How many more millions of passengers flew out of Atlanta than flew out of Chicago?

2.3.7 Exercise 7: Children's Library

	Number of Unique Titles	Relative Frequency
Rick Riordan	6	0.08
CS Lewis	12	0.16
J.K. Rolling	8	0.10
Orson Scott Card	9	0.12
Erin Hunter	32	0.42
Lois Lowry	4	0.05
Suzanne Collins	3	0.04
Veronica Roth	3	0.04

1. Create pie and bar charts for the data.

2. What is the scale of measurement?

3. Which author authored the most unique titles in the library?

2.3.8 Exercise 8: Marital Status

	1960	2010
Married	0.71	0.52
Single	0.15	0.28
Divorced	0.05	0.14
Widowed	0.09	0.06

1. Create pie and bar charts for the data.

2. What percentage of adults were married in 1960?

3. What percentage of adults were married in 2010?

2.4 Quantitative (Numeric) Data

Frequency distribution tables, histograms, and box plots are visual summarizations of quantitative data.

2.4.1 Grouped Frequency Distribution Tables

A frequency distribution summarizes quantitative data by grouping the data into user specified **classes** and recording the number of observations that fall within the range of each **class** and the relative frequency of observations in each **class**.

A frequency distribution for quantitative data also shows the cumulative frequency and cumulative relative frequency for the ordered classes.

Class and Data Grouping Requirements

- All classes are mutually exclusive - no data point can be in more than one class.
- All classes are collectively exhaustive - all data points must fit into one of the class designations.
- The total number of classes range between 5 and 20.

Width of a class = (Max value - Min value) / (Number of Classes)

All bins must be the same width. It may be necessary to start the first bin with a number lower than the lowest in the dataset or to end the last bin with a number that is greater than the largest number in the dataset.

Example

Weights	Frequency	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
0.35 up to 0.4	12	12	0.0992	0.099173554
0.4 up to 0.45	33	45	0.2727	0.371900826
0.45 up to 0.475	37	82	0.3058	0.67768595
0.475 up to 0.5	4	86	0.0331	0.710743802
>0.5	35	121	0.2893	1
Grand Total	**121**		**1**	

Table 3 Grouped Frequency, Cumulative Frequency, Relative Frequency, and Cumulative Relative Frequency Distribution for the Weight of 121 Packages of M&Ms

2.4.2 Histograms

A **histogram** is a series of adjacent rectangles (no spaces between) where the width and height of each rectangle represent the class width and frequency (or relative frequency) of the respective class. A histogram displays the overall distribution of the data not the distribution of individual values.

A histogram is the best graphical tool for displaying the relative frequency of grouped quantitative data.

There is **no space** between bars because the items in each class are not that distinct from one another. For example, define the classes "6.0-10.9" and "11.0-15.9." There is not much difference between 10.9 and 11.0.

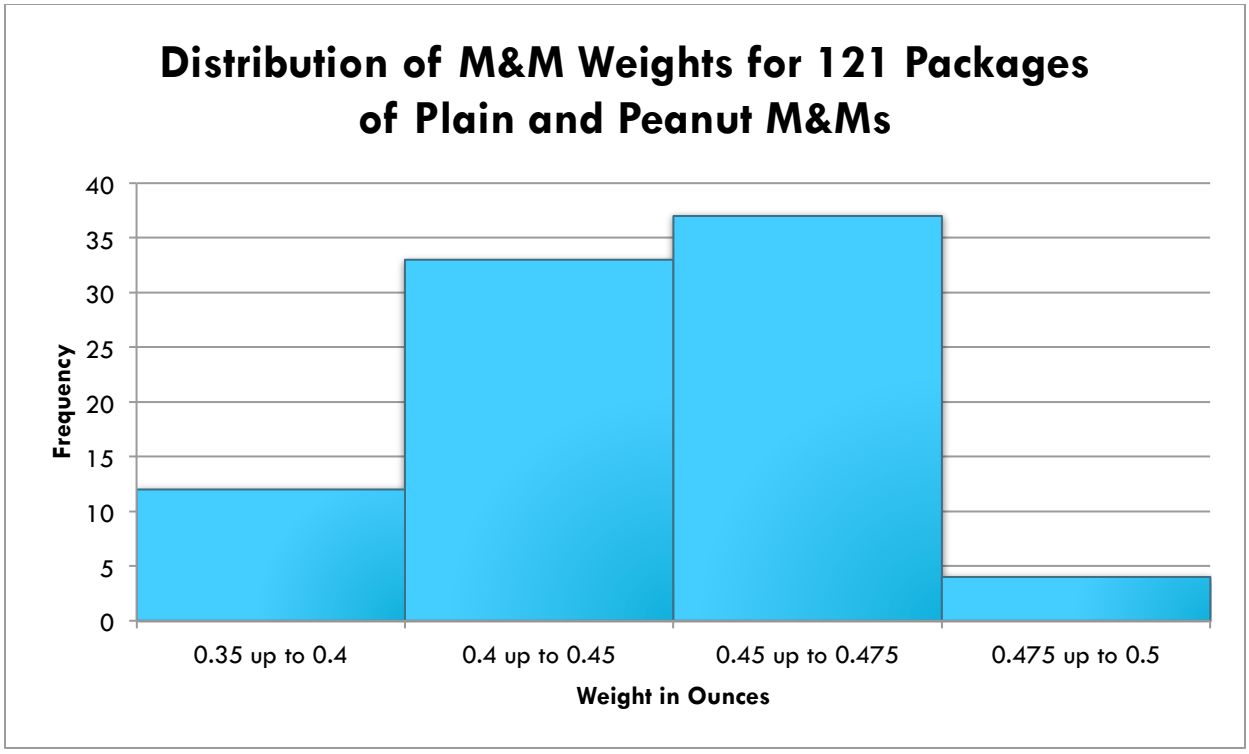

Figure 3 Histogram Illustrating the Distribution of Weight for 121 Packages of Plain and Peanut M&Ms

2.4.3 Scatterplots

A scatterplot is a graphical tool that illustrates how two variables are related, i.e., how one variable affects another. The two variables are graphed as coordinate pairs (x, y).

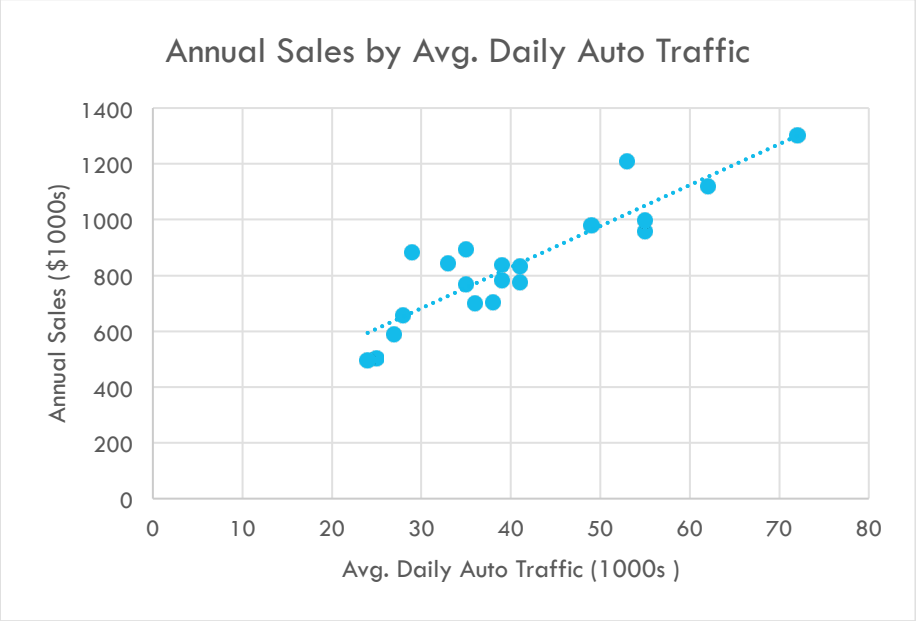

Figure 4 Scatterplot Illustrating the Relationship Between Annual Sales and Daily Automobile Traffic

DATA

Avg. Daily Auto Traffic (000s)	Annual Sales ($000s)
62	1121
35	766
36	701
72	1304
41	832
39	782
49	977
25	503
41	773
39	839
35	893
27	588
55	957
38	703
24	497
28	657
53	1209
55	997
33	844
29	883

2.5 Practice Problems: Quantitative Data

The following exercises are also in the Excel file:
Quantitative Data Practice Problems.xlsx

Answers to the following exercises are in the Excel file:
Quantitative Data Practice Problems KEY.xlsx

2.5.1 Exercise 1: Midwestern Homes

The following data represent the recent sales price (in $1,000s) of 24 homes in a Midwestern city.

187	125	165	170	230	139	195	229
239	135	188	210	228	172	127	139
122	181	196	237	115	199	170	239

1. Sort the data in ascending order.
2. Determine the number of classes. Usually 5 to 20 classes is the rule.
3. Count and record the number of data points that fall into each class (frequency).
4. Calculate and record the cumulative frequency.
5. Calculate and record the relative frequency and cumulative relative frequency.

Grouped Sales Prices	Frequency	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency

1. Suppose the data on house prices will be grouped into five classes. The width of the classes for a frequency distribution or histogram is the closest to _____.
2. Suppose the data are grouped into five classes, and one of them will be "115 up to 140." - that is, $\{x; 115 \leq x < 140\}$. The relative frequency of this class is _____.
3. Suppose the data are grouped into five classes, and one of them will be "165 up to 190." - that is, $\{x; 165 \leq x < 190\}$. The frequency of this class is _____.

2.5.2 Exercise 2: Statistics Quiz

The following data represent scores on a pop quiz in a statistics section.

16	16	17	32	32	33	37	44	45	47
55	56	56	62	66	70	72	74	82	84

Note: The data are already sorted.

Instructions

1. Create a frequency distribution for this data.
2. Answer the questions below.

Grouped Scores	**Frequency**	**Cumulative Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**

1. Suppose the data on quiz scores will be grouped into five classes. The width of the classes for a frequency distribution or histogram is _____.

2. Suppose the data are grouped into five classes, and one of them will be "30 up to 44." that is, $\{x; 30 \leq x < 44\}$. The frequency of this class is _____.

3. Suppose the data are grouped into five classes, and one of them will be "30 up to 44" —that is, $\{x; 30 \leq x < 44\}$. The relative frequency of this class is _____.

2.5.3 Exercise 3: Eastside HS

Thirty students at Eastside High School took the SAT on the same Saturday. Their raw scores are given next.

1,450	1,480	1,490	1,530	1,590	1,620	1,620	1,640	1,650	1,710
1,740	1,780	1,800	1,800	1,820	1,830	1,830	1,840	1,870	1,900
1,910	1,950	1,950	1,980	2,000	2,010	2,100	2,260	2,350	2,390

Note: The data are already sorted.

Instructions

1. Create a frequency distribution for this data.
2. Answer the questions below.

Grouped Scores	Frequency	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency

1. Consider a frequency distribution of the data that groups the data in classes of 1400 up to 1600, 1600 up to 1800, 1800 up to 2000, and so on. How many students scored at least 1800 but less than 2000?

2. Consider a frequency distribution of the data that groups the data in classes of 1400 up to 1600, 1600 up to 1800, 1800 up to 2000, and so on. What percent of students scored less than 2200?

3. Consider a frequency distribution of the data that groups the data in classes of 1400 up to 1600, 1600 up to 1800, 1800 up to 2000, and so on. What is the approximate relative frequency of students who scored more than 1600 but less than 1800?