



- **Displaying Descriptive Statistics**

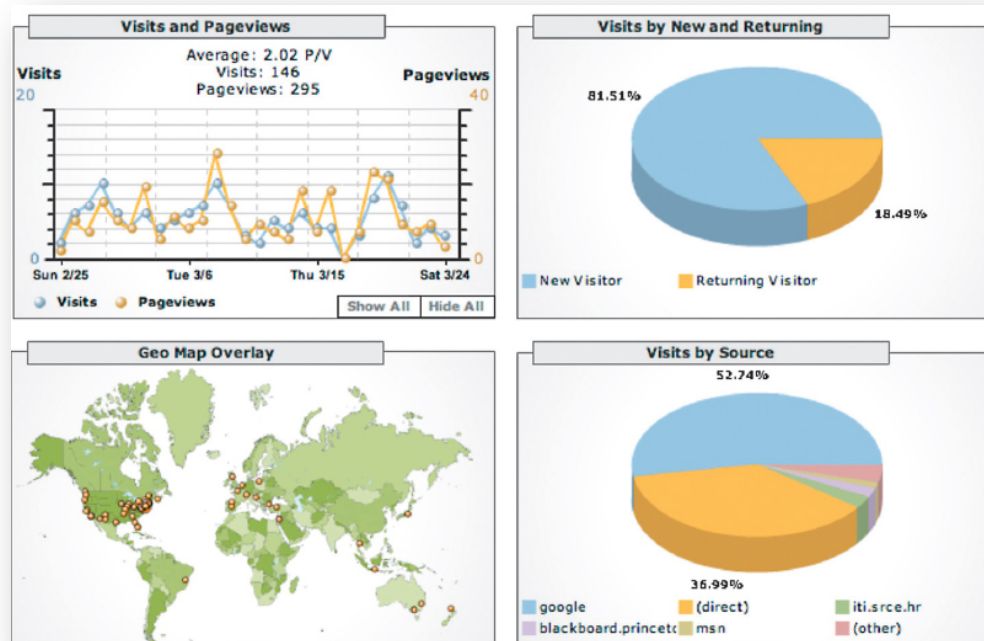
*Chapter 4*

# The Three Rules of Data Analysis

Rules 1, 2, and 3: Make a picture.

Pictures ...

1. Reveal things that can't be seen in a table of numbers.
2. Show important features and patterns in the data.
3. Provide an excellent means for reporting findings to others.



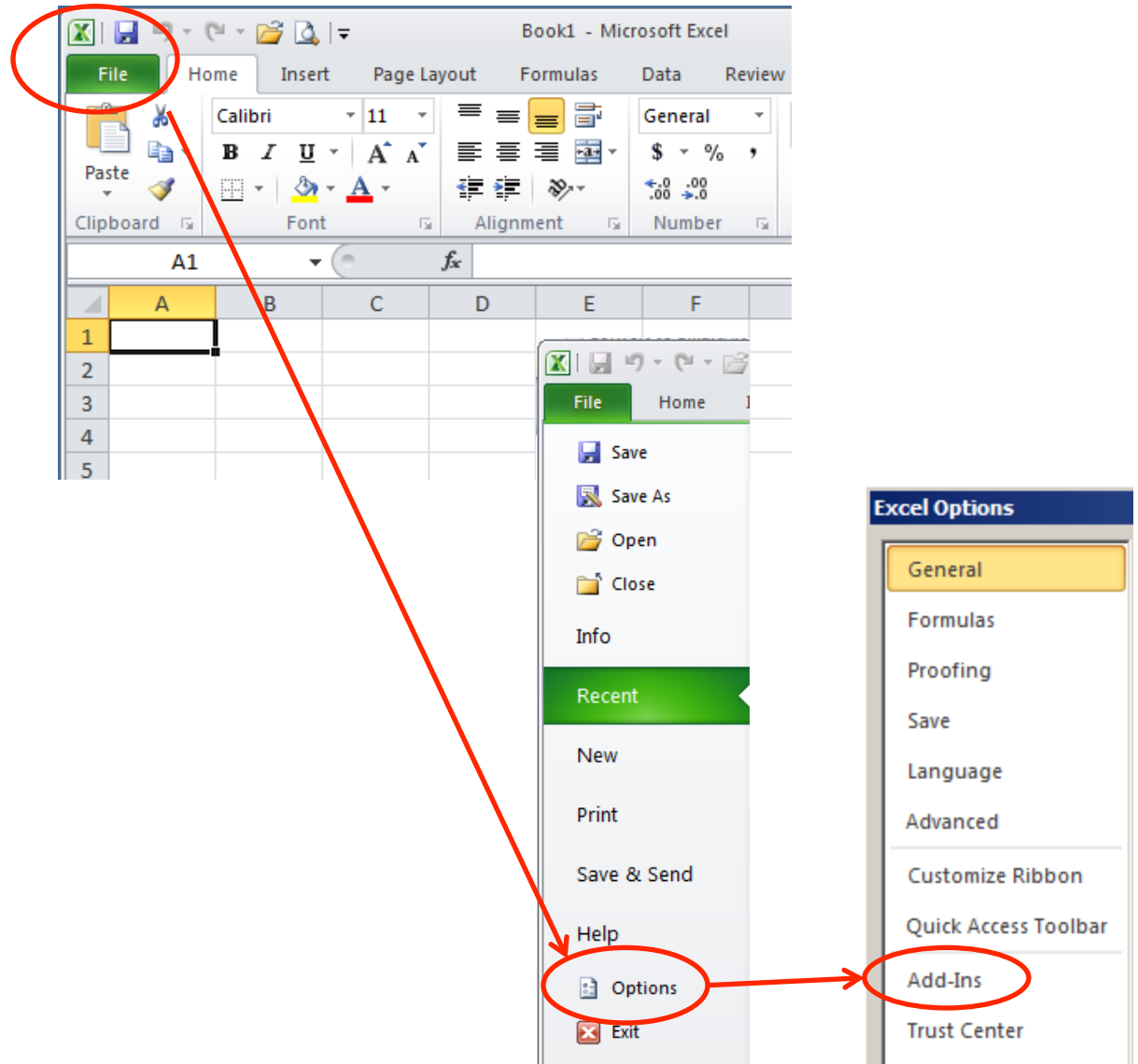
# Software for descriptive statistics

---

- You can use Minitab , Excel, SPSS or SAS to perform descriptive statistics analysis.
- **Microsoft Excel** has built-in options for data presentation and statistical analysis
- You may need to activate Excel's **Analysis Tool Pak Add-in** to see these options

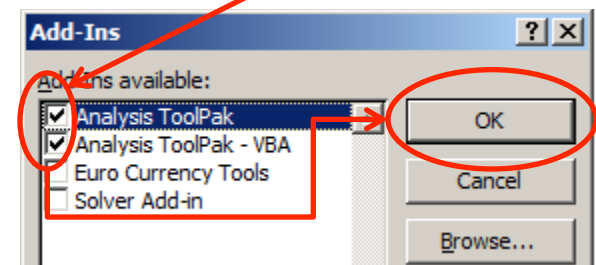
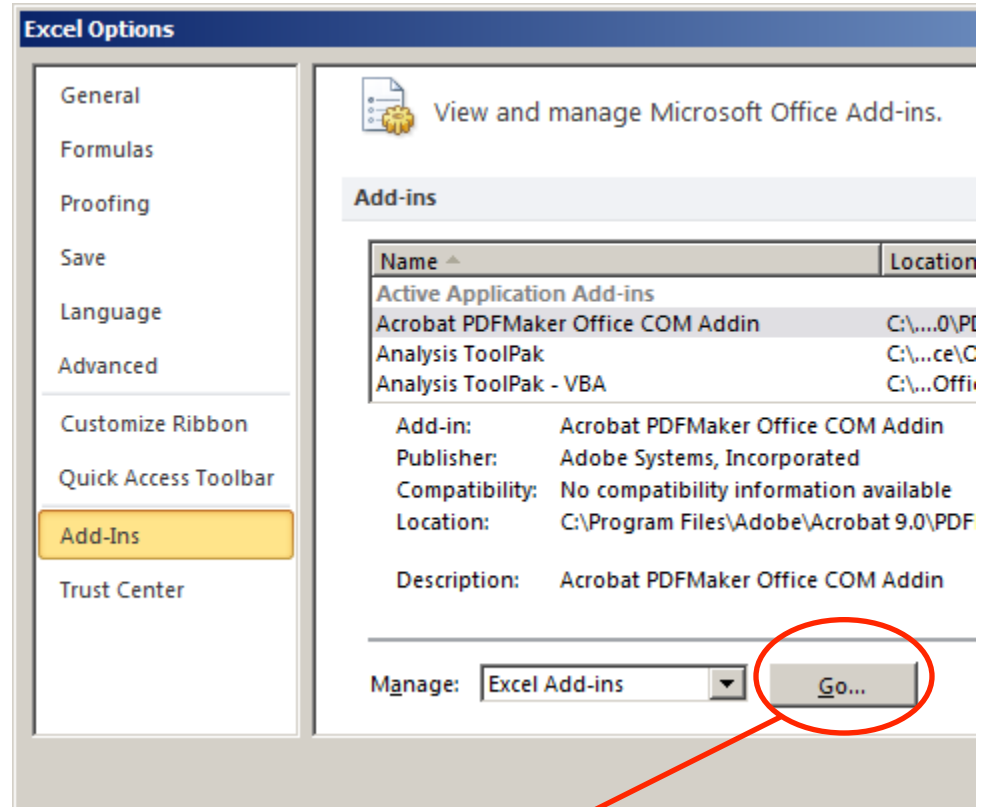
# How to activate excel data analysis tool pack

1. With Excel open, click on the **File** tab found in the upper left corner of your computer screen
2. Click **Options** shown in the drop down menu. This will open the **Excel Options** dialog box
3. Select **Add-Ins** in the left margin...



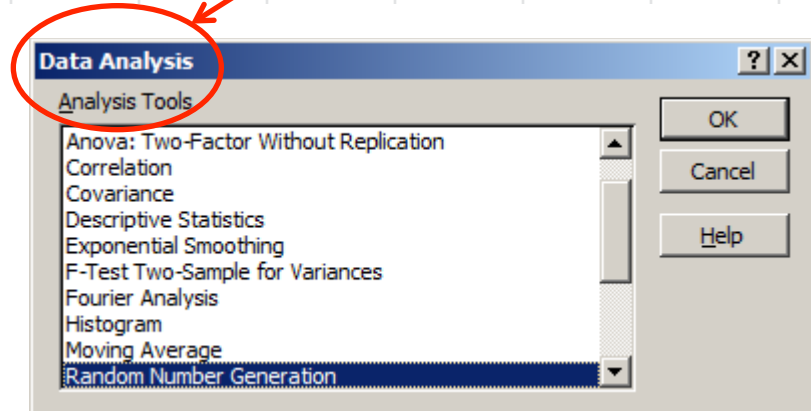
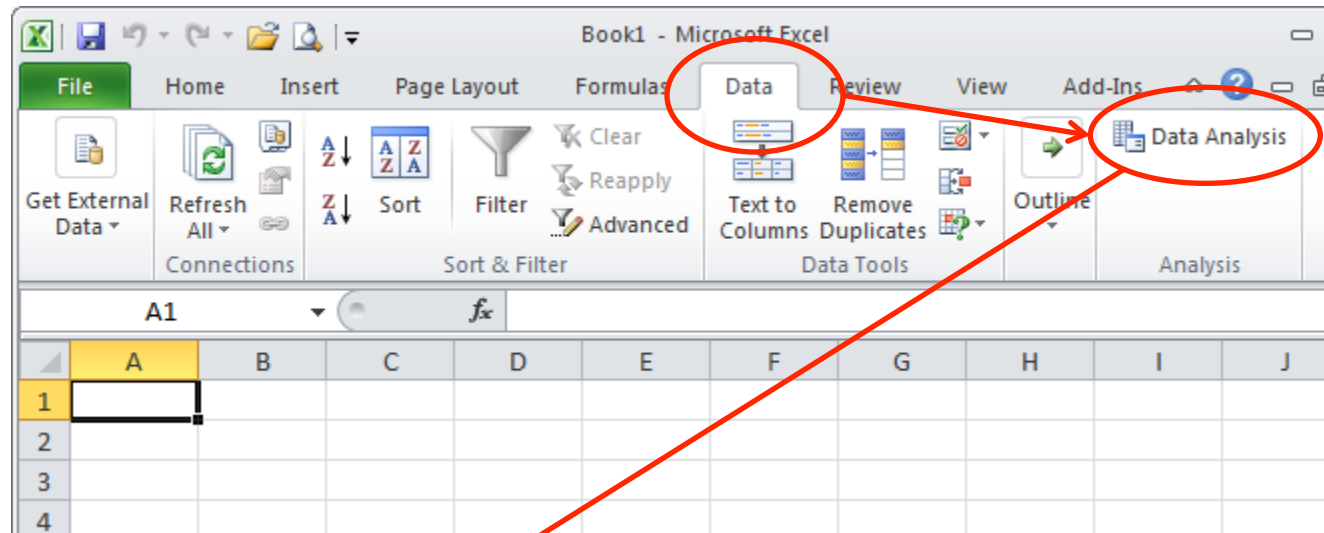
# How to activate excel data analysis tool pack

4. Click on **Go** at the bottom of the screen
5. Select the check boxes for **Analysis ToolPak** and **Analysis ToolPak - VBA** in the popup menu and click **OK**



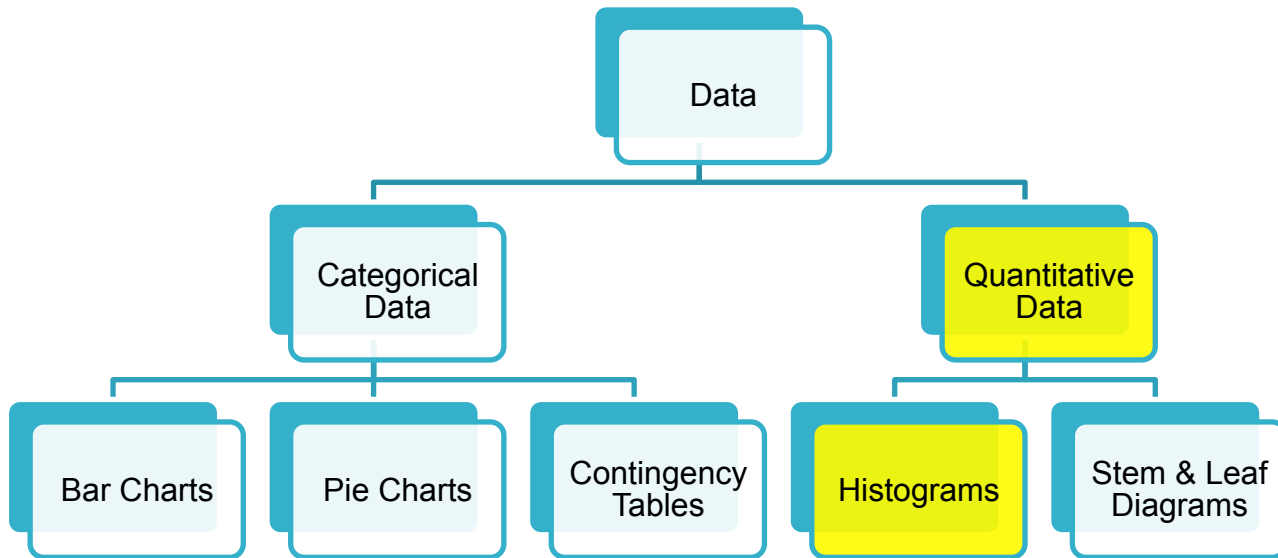
# How to activate excel data analysis tool pack

6. Select the Data tab. Click on **Data Analysis** on the right side of the application bar



# Outline

---



# Constructing a Frequency Distribution

---

- For quantitative data, **Frequency distribution** can be constructed that shows the number of data observations that fall into specific intervals
- Frequency distribution graphically summarizes information not readily observable by merely looking at data in a table
- It could be used for both discrete and continuous quantitative data. As you remember :
  - **Discrete data** are values based on observations that can be counted and are typically represented by whole numbers
    - represent something that has been counted
    - take on whole numbers such as 0, 1, 2, 3
  - **Continuous data** are values that can take on any real numbers, including numbers that contain decimal points
    - usually measured rather than counted
    - Examples are weight, time, and distance

# Constructing a Frequency Distribution

- **Example:** Number of iPads sold per day

**The Number of iPads  
Sold in Each of 50 Days**

---

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

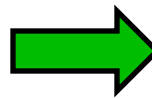
---

# Constructing a Frequency Distribution

- Example:** Frequency distribution of the number of iPads sold per day

**The Number of iPads Sold in Each of 50 Days**

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



**Frequency Distribution for the Number of iPads Sold in the Past 50 Days**

NUMBER SOLD PER DAY	FREQUENCY
0	5
1	8
2	14
3	13
4	6
5	4
<b>Total</b>	<b>50</b>

# Relative Frequency Distributions

- **Relative frequency distributions** display the **proportion** of observations of each class relative to the total number of observations
  - shows the fraction of observations in each class
  - found by dividing each frequency by the total number of observations
  - the fractions in a relative frequency distribution add up to 1.00
- **Example** : Frequency distribution of the number of iPads sold per day:

Relative Frequency Distribution for the Number of iPads Sold in the Past 50 Days

NUMBER SOLD PER DAY	FREQUENCY	RELATIVE FREQUENCY
0	5	$5/50 = 0.10$
1	8	$8/50 = 0.16$
2	14	$14/50 = 0.28$
3	13	$13/50 = 0.26$
4	6	$6/50 = 0.12$
5	4	$4/50 = 0.08$
Total	50	1.00

*Two iPads were sold on 28% of the days*

# Cumulative Relative Frequency Distributions

- A **cumulative relative frequency distribution** totals the proportion of observations that are *less than or equal to* the class at which you are looking
  - Shows the accumulated proportion as values vary from low to high
- **Example :**

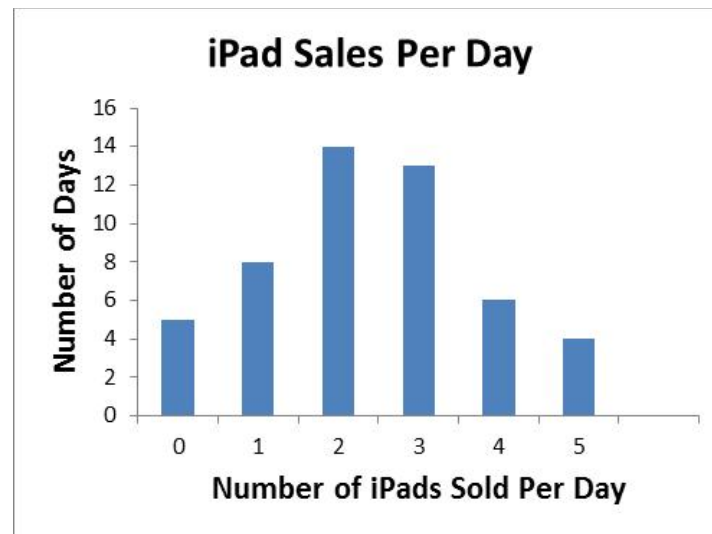
**Cumulative Relative Frequency Distribution for the Number of iPads Sold in the Past 50 Days**

NUMBER SOLD PER DAY	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
0	5	0.10	0.10
1	8	0.16	$0.10 + 0.16 = 0.26$
2	14	0.28	$0.26 + 0.28 = 0.54$
3	13	0.26	$0.54 + 0.26 = 0.80$
4	6	0.12	$0.80 + 0.12 = 0.92$
5	4	0.08	$0.92 + 0.08 = 1.00$
Total	50	1.00	

**Three iPads or less were sold on 80% of the business days**

# Using a Histogram to Graph a Frequency Distribution

- A **histogram** is a graph showing the number of observations in each class of a frequency distribution
- We usually use the term “bins” for the classes in the distribution



# Constructing a Histogram in Excel

---

**Step 1:** Determine the number of groups or classes to use.

**Step 2:** Establish the class width.

$$W = \frac{\text{Largest Value} - \text{Smallest Value}}{\text{Number of Classes}}$$

**Step 3:** Determine the class range (bin range) for each class.

**Step 4:** Count the number of values in each class using excel analysis tool pack :



# Constructing a Histogram in Excel

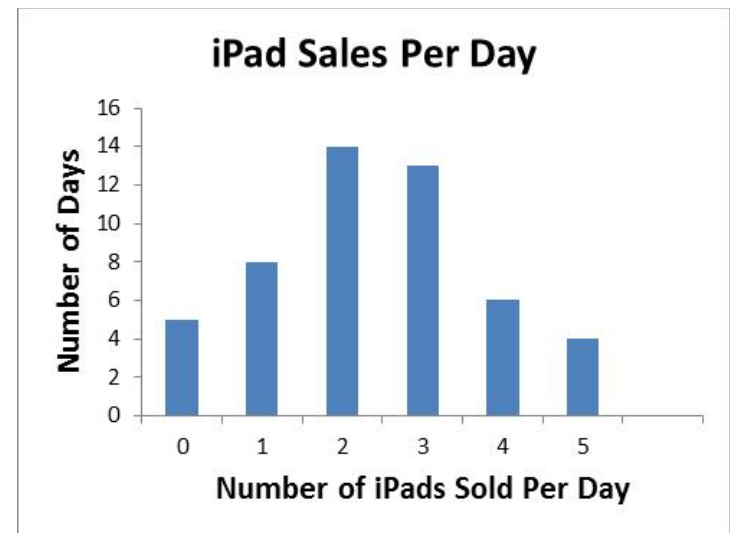
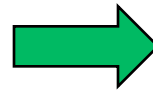
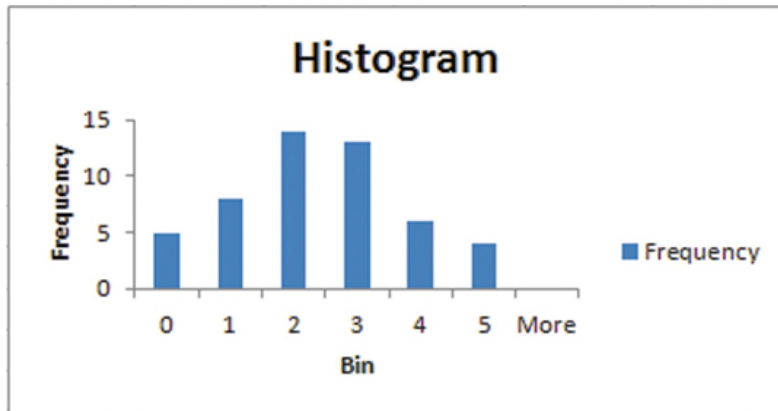
- In the **Input Range** text box, highlight the desired data
- In the **Bin Range** text box, highlight the bin values (create bins if not already created before)
- For **Output options**, select **New Worksheet Ply** and **Chart Output**
- Click **OK**

The screenshot shows the Microsoft Excel interface with the 'Histogram' dialog box open. The dialog box is titled 'Histogram' and has a 'Data Analysis' icon in the top right corner. The 'Input' section contains 'Input Range:' with the value '\$A\$1:\$E\$10' and 'Bin Range:' with the value '\$G\$2:\$G\$7'. The 'Output options' section has three radio buttons: 'Output Range:', 'New Worksheet Ply:', and 'New Workbook'. The 'New Worksheet Ply:' option is selected. Below these are two checkboxes: 'Pareto (sorted histogram)' and 'Cumulative Percentage', both of which are unchecked. At the bottom of the dialog box, there is a checkbox labeled 'Chart Output' which is checked. The 'OK', 'Cancel', and 'Help' buttons are on the right side of the dialog box. Red circles are drawn around the 'Input Range', 'Bin Range', 'New Worksheet Ply', and 'Chart Output' fields, with red arrows pointing from these circles to the 'OK' button. The background spreadsheet shows a grid with columns A through E and rows 1 through 12. The data in the spreadsheet is as follows:

	A	B	C	D	E
1	4	2	3	2	5
2	5	1	3	3	2
3	3	2	2	3	2
4	2	2	3	0	1
5	3	1	1	5	4
6	1	2	4	3	5
7	2	0	0	3	2
8	3	3	3	2	2
9	0	4	2	4	3
10	1	1	4	0	1
11					
12					

# Constructing a Histogram in Excel

- Customize the Excel graph to make it more attractive



- Stretch size to better proportion
- Eliminate "more" bin
- Modify the graph and axis labels
- Remove the redundant "Frequency" legend

# Constructing a Histogram in Excel

---

**Exercise:** (*constructing a histogram*)

**Step 1:** Open Capital.xls

**Step 2:**

Calculate Max and Min of data in B2:B301.

Divide into 10 classes

Class width =139. Lets use 150.

**Step 3:**

Using Min of 99 and class width of 150, we get class boundaries:

90 – 249; 249 – 399; ...

Enter 249; 399;... in E2:E11

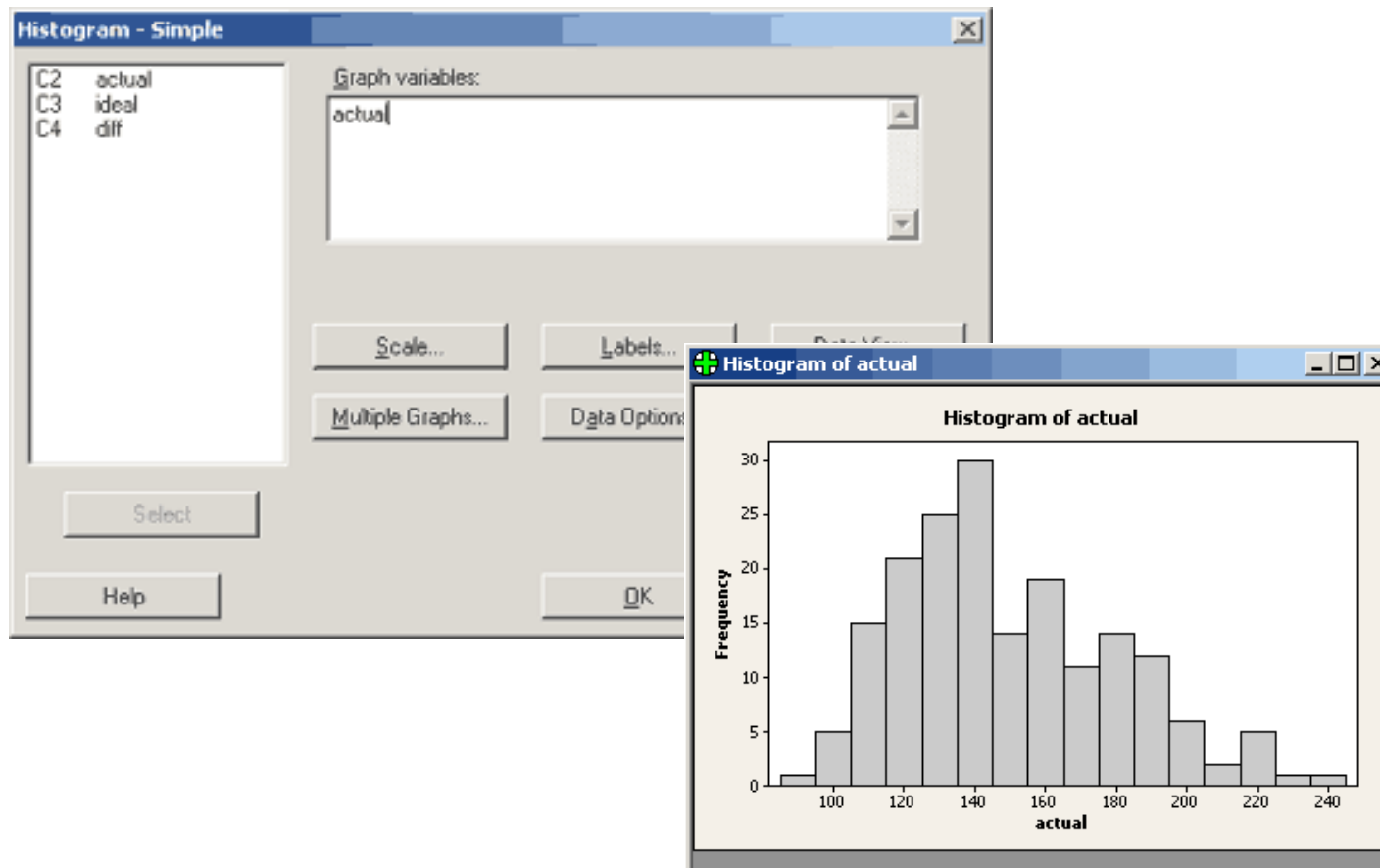
**Step 4:**

Tools/Data; Data Analysis; Histogram; Output range F1

Highlight G2:G11; Column graph

# Constructing a Histogram in Minitab

- For histogram in Minitab, Simply go to **graph** menu and select **Histogram**.

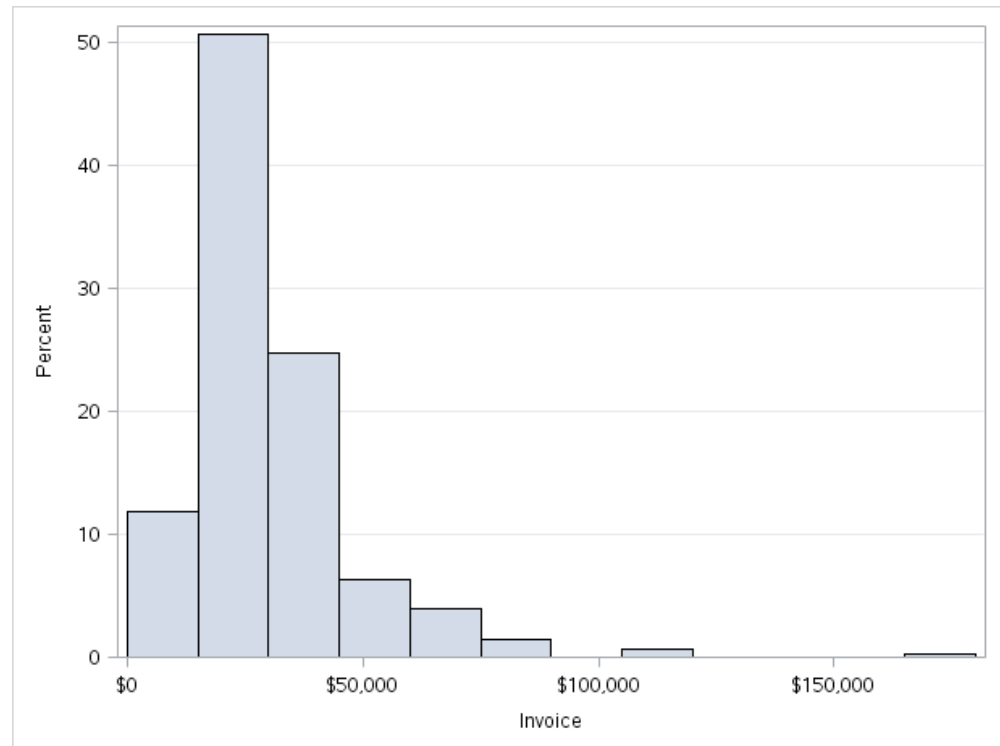


# Constructing a Histogram in SAS (Optional )

- Using the following command :

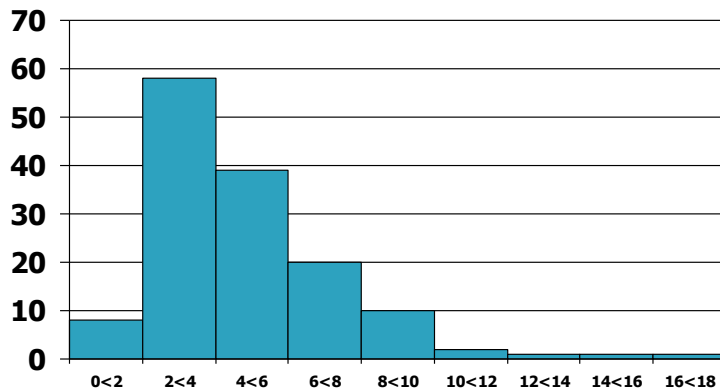
```
/* Option group 5 (GRAPH SIZE) parameters. */  
/*--Set Graph Size (in inches)--*/  
ods graphics / reset width=6.4in height=4.8in imagemap;
```

```
/*--SGPLOT proc statement--*/  
proc sgplot data=SASHELP.CARS;  
  /*--Histogram settings--*/  
  histogram Invoice /;  
  
  /*--Vertical or Response Axis--*/  
  yaxis grid;  
run;
```

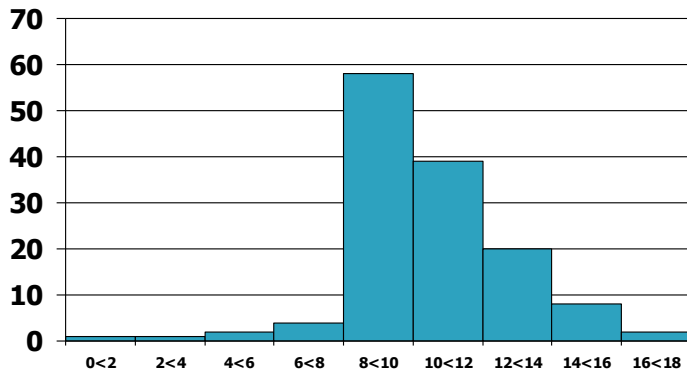


# Histogram Properties

- ① It provides visual indication of where the approximate center of the data is.

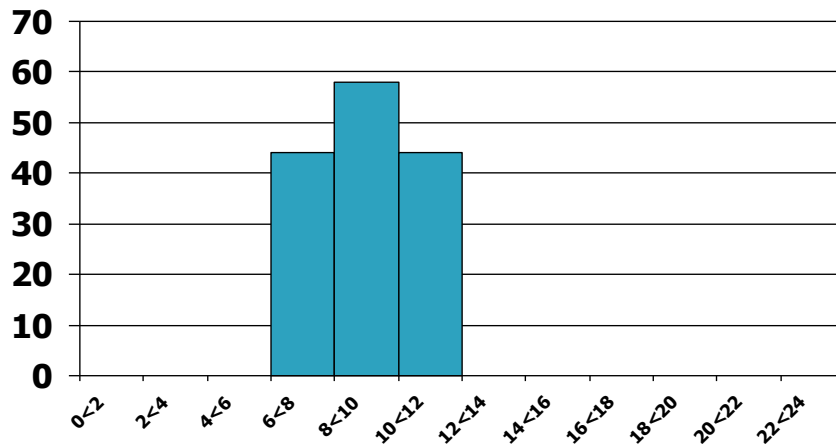
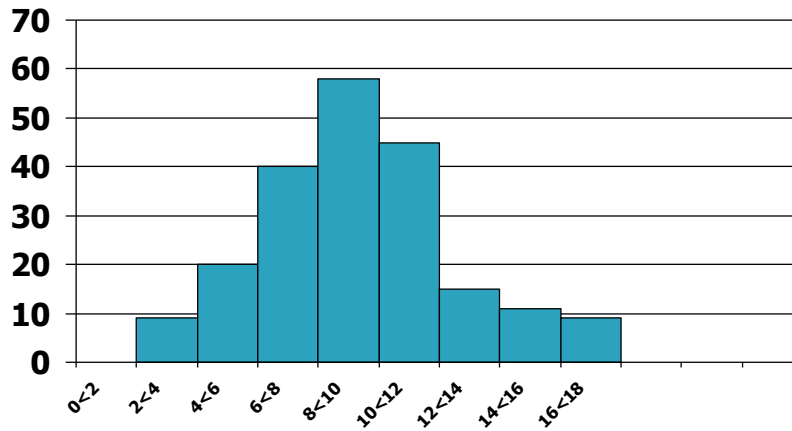


Different Centers



# Histogram Properties

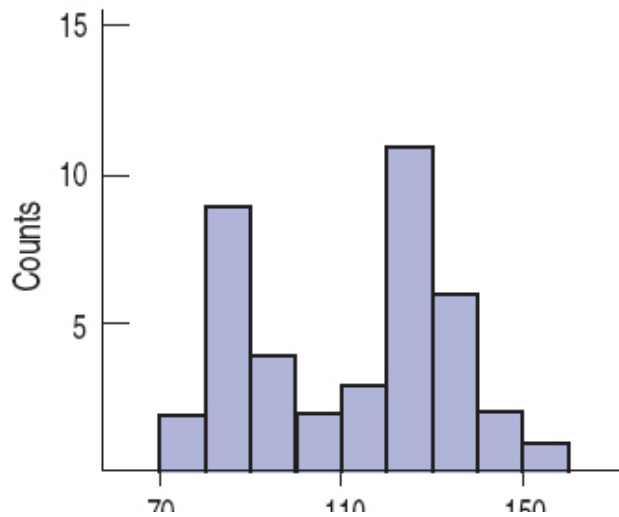
- ② It provides visual indication of spread of data from the Center.



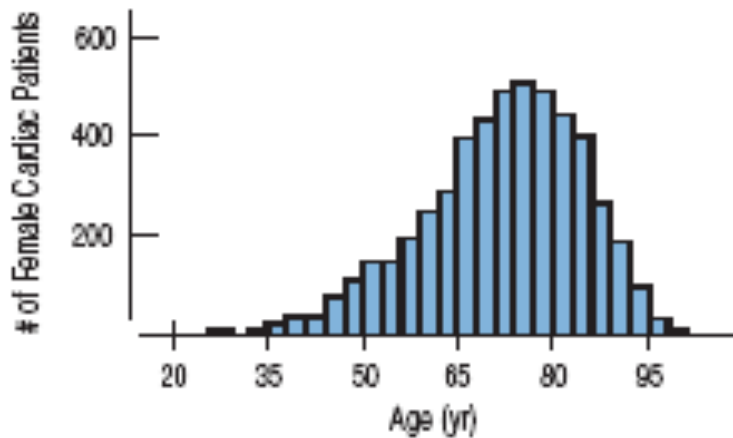
Same Center, Different Spread

# Histogram Properties

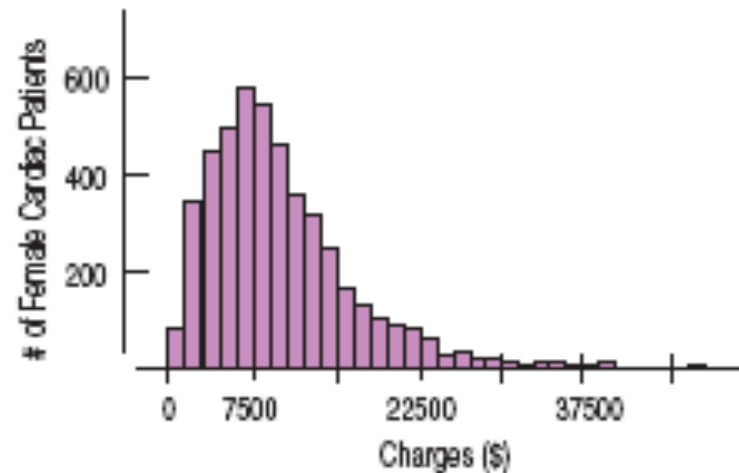
③ We can observe the shape of the distribution



***Bimodal Histogram***



***Skewed to the Left***



***Skewed to the Right***

# Rules for Classes for Grouped Data

---

- A histogram must satisfy these conditions. If it doesn't we call it a "bar chart"
1. **Equal-size classes.** All classes in the frequency distribution must be of equal width
  2. **Mutually exclusive classes.** Class boundaries cannot overlap
  3. **Include all data values.** Make sure all data values are accounted for in the total row of the frequency distribution
  4. **Avoid empty classes.** It is undesirable for a histogram to display a class so narrow that there are no observations in it

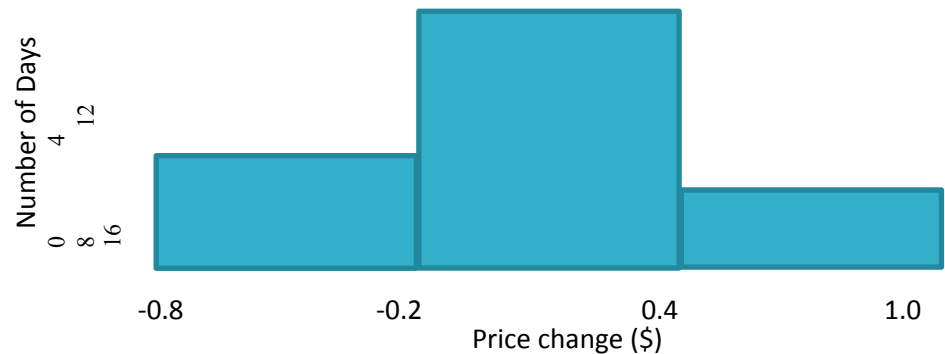
# The Consequences of Too Few or Too Many Classes

- Wide classes results in few class intervals
  - Can obscure important patterns
  - Gives a “blocky” distribution graph
  - Summarizes the data too much
  - Tells us little about the true distribution shape

- **Example :**

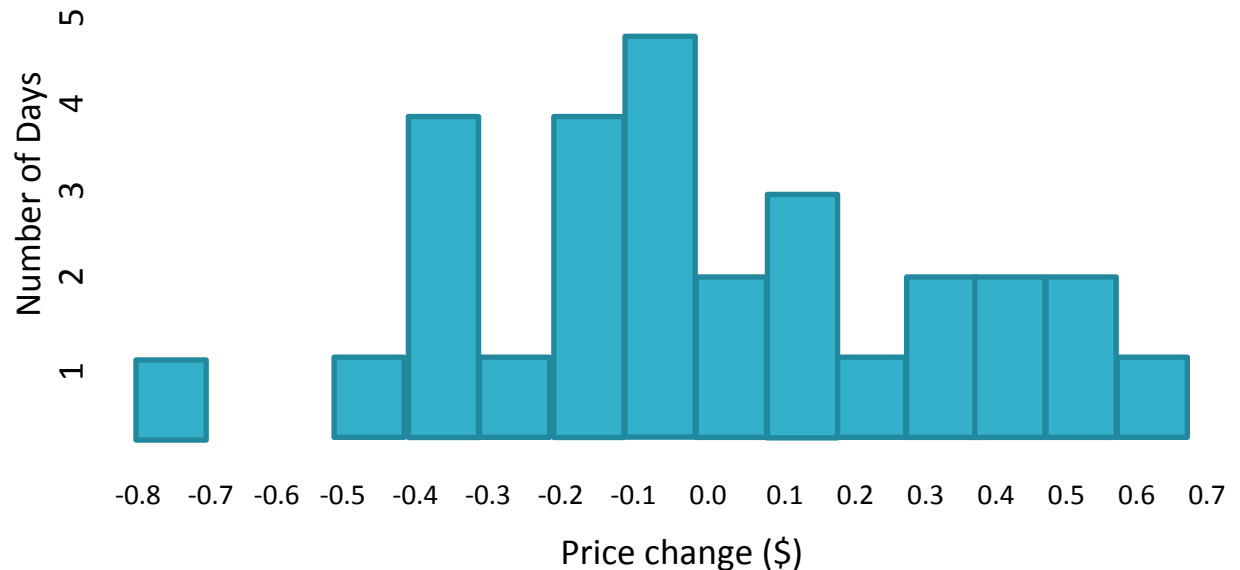
Bell Canada stock price change Sept 12 – Oct 21, 2011

*too few bins*



# The Consequences of Too Few or Too Many Classes

- Too many narrow classes in a histogram also has consequences
  - Results in a “jagged” histogram
  - Some classes may be empty
  - Does not summarize the data enough
- **Example :**  
Bell Canada stock price change Sept 12 – Oct 21, 2011

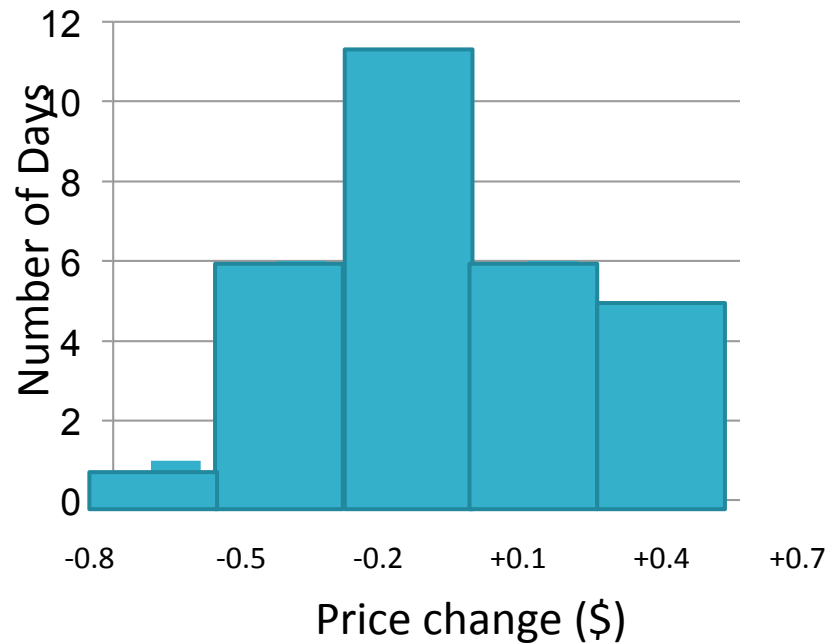


*too many bins*

# The Consequences of Too Few or Too Many Classes

---

- With good number of bins :



# Histograms – Good Number of Bins

## Class Interval

20-under 30

30-under 40

40-under 50

50-under 60

60-under 70

70-under 80

## Frequency

6

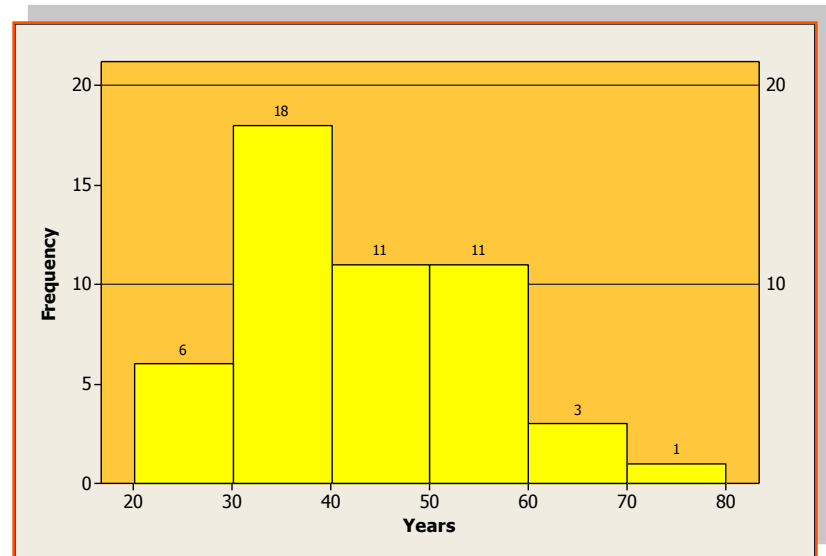
18

11

11

3

1



## Cumulative

## Frequency

## Class Interval

20-under 30

30-under 40

40-under 50

50-under 60

60-under 70

70-under 80

6

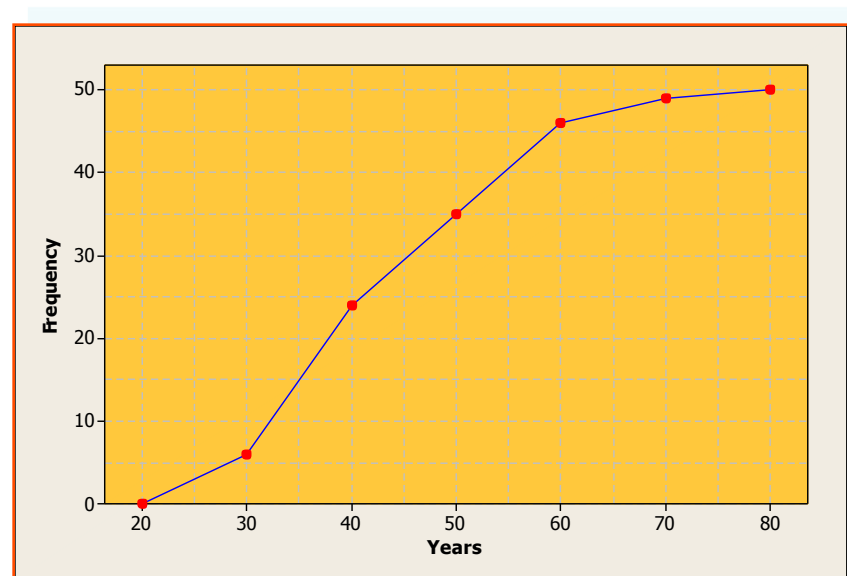
24

35

46

49

50



# Number of Classes (bins)

---

- One method to determine the number of classes in a frequency distribution is the rule

$$2^k > n$$

where  $k$  = Number of classes  
 $n$  = Number of data points

- Find the lowest value of  $k$  that satisfies the rule
- Suppose  $n = 50$ 
  - $2^5 = 32 < 50$  ( $k = 5$  is too small)
  - $2^6 = 64 > 50$  ( $k = 6$  is a good choice)

# Number of Classes (bins)

---

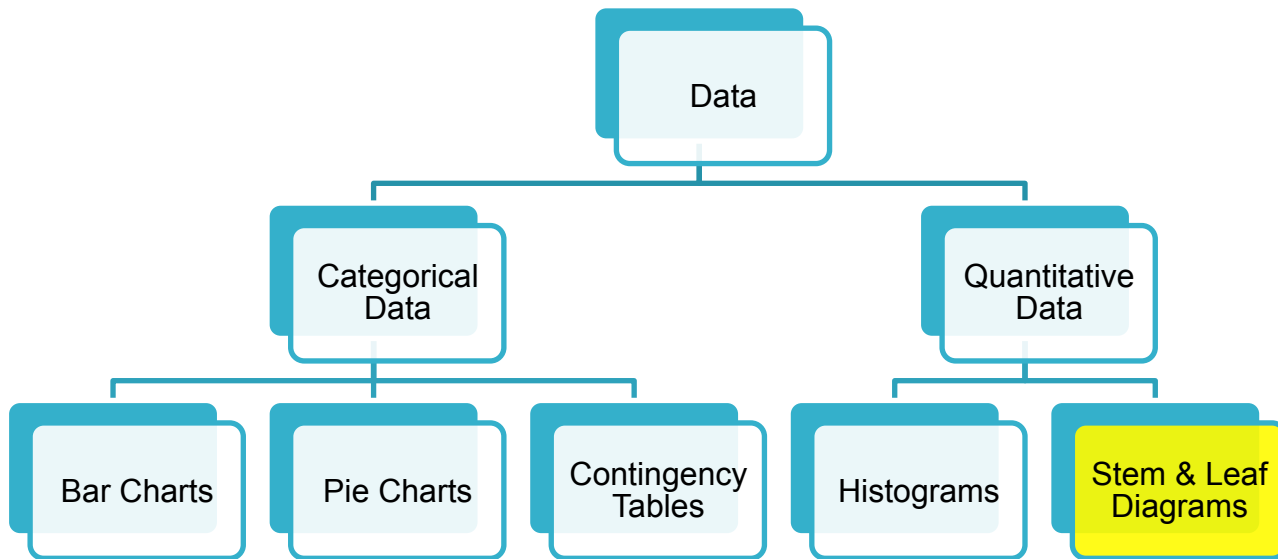
- Once  $k$  is known, the width of each class can be found
  - The width is the range of numbers to put into each class

$$W = \frac{\text{Largest Value} - \text{Smallest Value}}{\text{Number of Classes}}$$

- Round this estimate to a useful whole number that makes the frequency distribution more readable
- There is no one correct answer for the class width
- The goal is to create a histogram to clearly and usefully show the pattern in the data

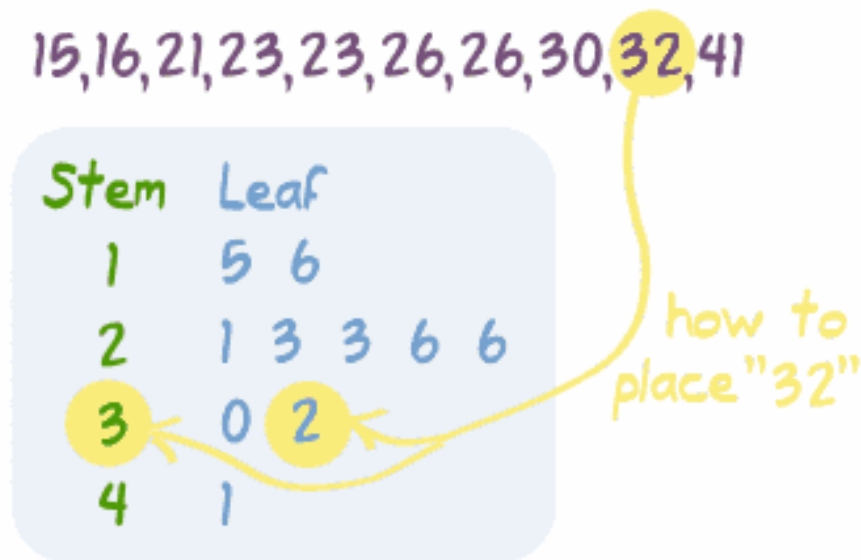
# Outline

---



# Stem and Leaf Diagrams

- A special table where each data value is split into a "leaf" (usually the last digit) and a "stem" (the other digits).
- The "stem" is used to group the scores and each "leaf" indicates the individual scores within each group.
- By listing all of the leaves to the right of each stem, we can graphically describe how the data are distributed. It provides a histogram-like view of the distribution.



# Stem and Leaf Diagrams

## Example :

1. Sort the data from lowest to highest
2. Determine the unique stem values  
7, 8, 9 are the different stem values in this example
3. List the stems in a vertical column and then add the leaf values to the right of the appropriate stem, in ascending order

**Sorted Exam Scores**

78	78	79	79	79	80	80	80
80	81	81	82	83	83	84	84
84	85	86	87	88	90	92	95

7 | 8 8 9 9 9

8 | 0 0 0 0 1 1 2 3 3 4 4 4 5 6 7 8

9 | 0 2 5

# Stem and Leaf Diagrams

---

## Example (continued ):

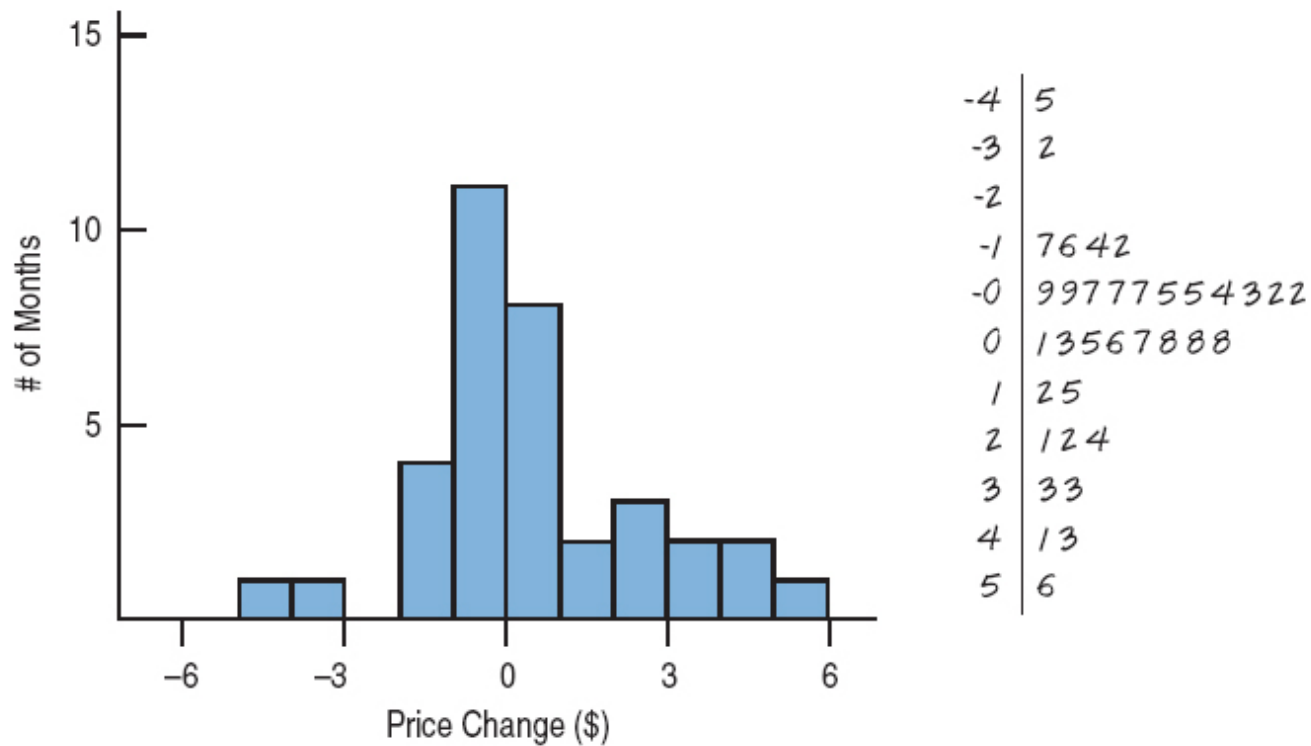
- To get more detail the stems can be split in half

```
7(5) | 8 8 9 9 9
8(0) | 0 0 0 0 1 1 2 3 3 4 4 4
8(5) | 5 6 7 8
9(0) | 0 2
9(5) | 5
```

- The stem labeled 7(5) stores all the scores between 75 and 79
- The stem 8(0) stores all the scores between 80 and 84

# Stem and Leaf Diagrams

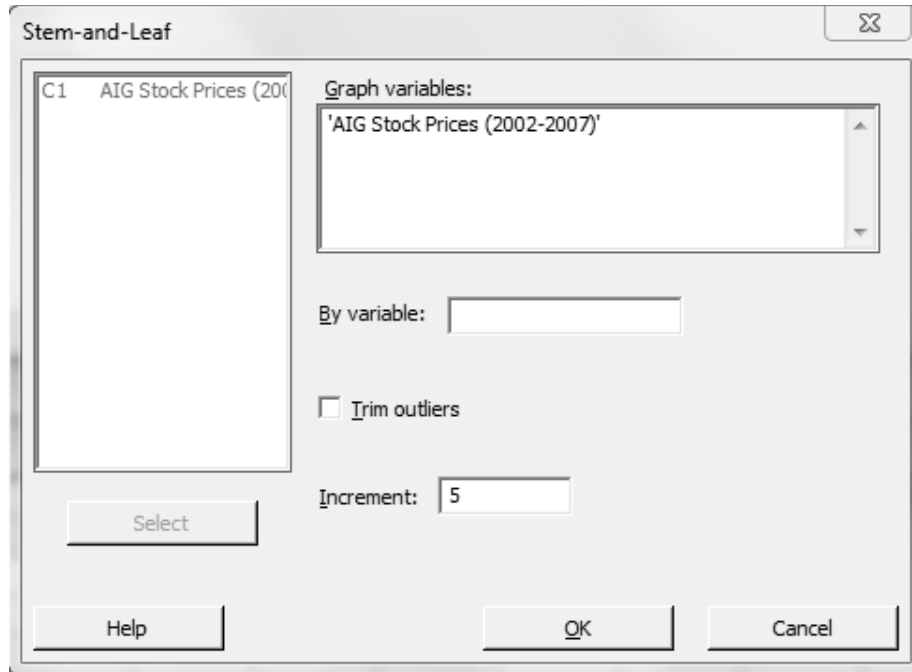
- All information is retained in a stem/leaf, whereas in a histogram, information is summarized



**Monthly Price Change in Enrons Stock 1997-2001**

# Constructing a stem and leaf diagram in Minitab

- For stem and leaf plot in Minitab, Simply go to **graph** menu and select **Stem and leaf**.

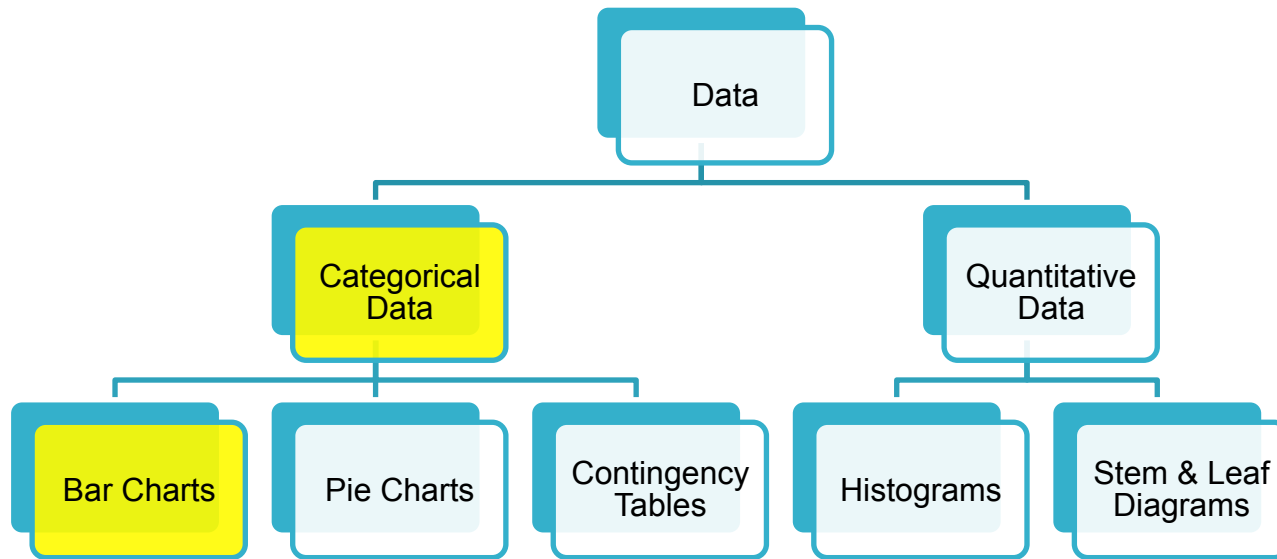


Stem-and-leaf of Prices N = 36  
Leaf Unit = 1.0

```
2    4    99
5    5   134
10   5   56799
13   6   134
(11) 6   56778888889
12   7   00111222334
1    7    7
```

# Outline

---



# Frequency Tables

---

- Here **Frequency table** can also help to display qualitative data by indicating the number of occurrences of various categories
- A **frequency table** organizes data by recording totals and category names as in the table below.
- The names of the categories label each row in the frequency table.

<b>Province</b>	<b>Corporate Stores</b>
Newfoundland and Labrador	13
Prince Edward Island	4
Nova Scotia	34
New Brunswick	21
Quebec	223
Ontario	155

*Number of Loblaws Stores in Eastern Canada*

# Relative Frequency Table

---

A **relative frequency table** displays the percentages that lie in each category rather than the counts.

<b>Province</b>	<b>Corporate Stores (%)</b>
Quebec	49.56
Ontario	34.44
Nova Scotia	7.56
Other	8.44
<b>Total</b>	<b>100</b>

*Percentage of Loblaws Stores in Eastern Canada*

# Frequency tables in Excel

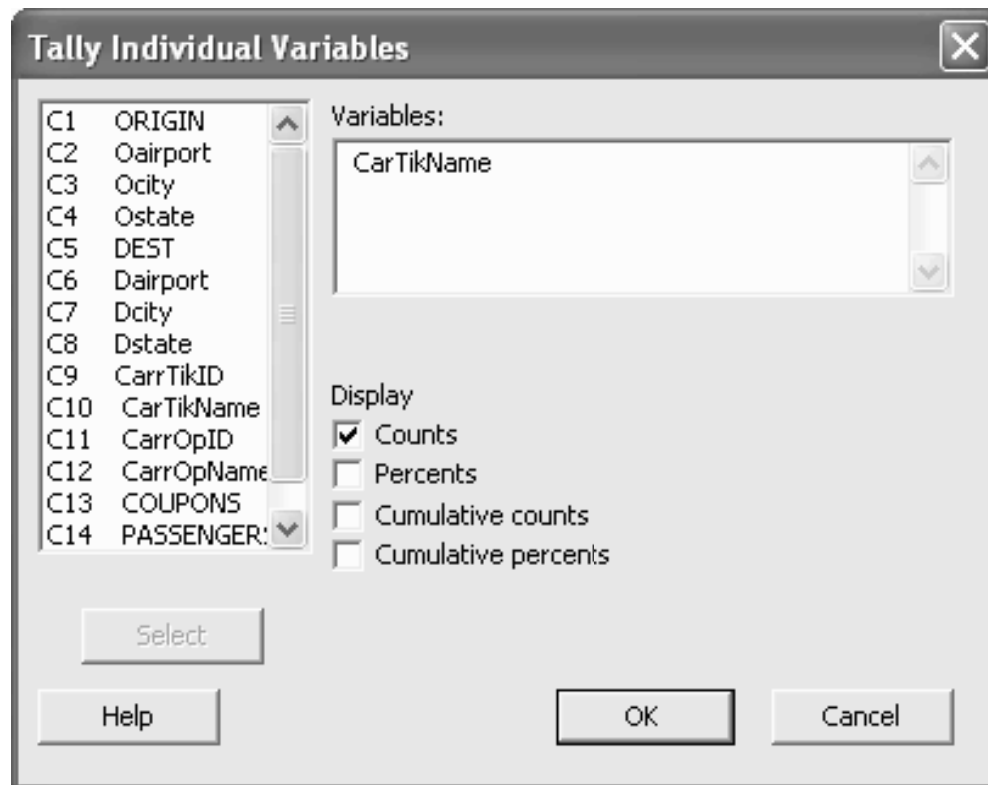
---

- Can use Excel's **COUNTIF** function to count the number of values matching a category label
- **FREQUENCY** function can also help you to create a frequency table.
- **Exercise :**
  - Open Trucks.xls
  - To calculate how many trucks were observed in each month...
  - Enter months 1,...,12 in M2:M13
  - Select N2.N13
  - Use **FREQUENCY** function
  - Data: A2:A201; Bin: M2:M13;
  - **Ctrl + Shift + Enter**

# Frequency tables in Minitab

- For creating a frequency tables in Minitab, go to:

**Stat . Tables . Tally Individual variables ...**

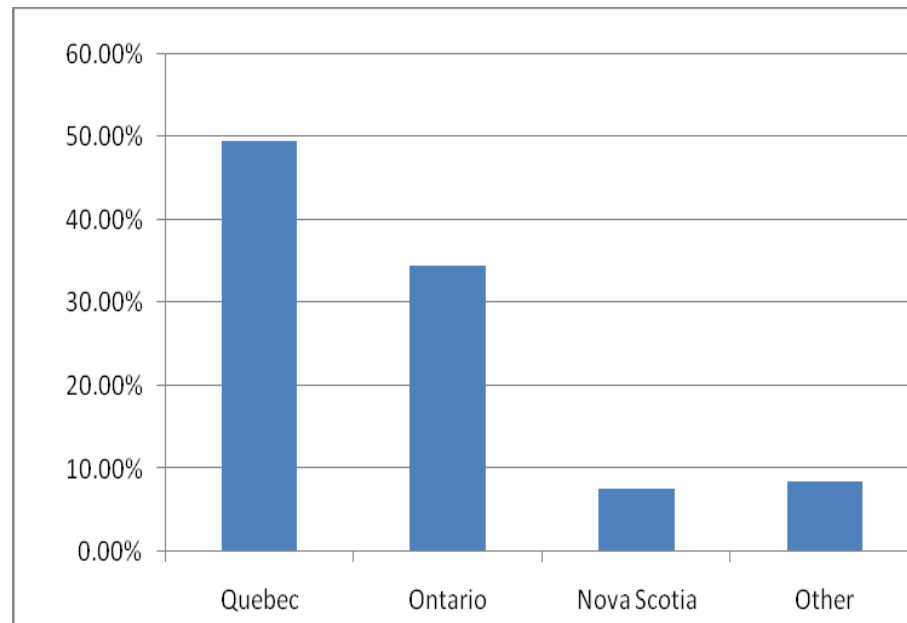


# Bar Charts

---

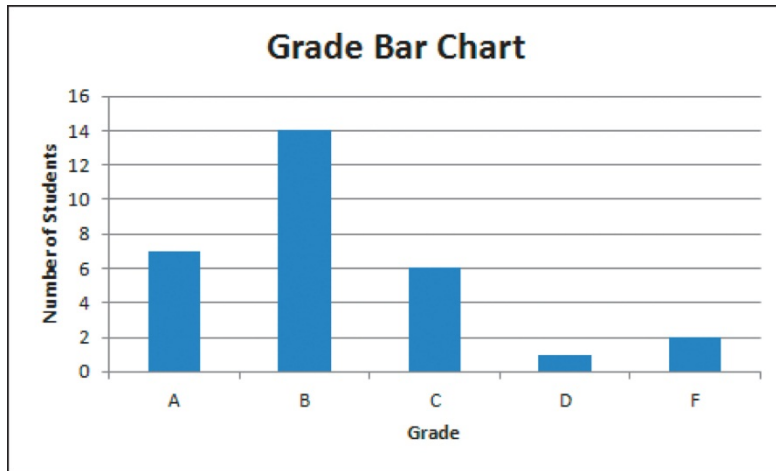
A *bar chart* displays the distribution of a categorical variable, showing the counts for each category next to each other for easy comparison.

The bar graph here gives a more *accurate* visual impression of the Loblaws data from pervious relative frequency table.

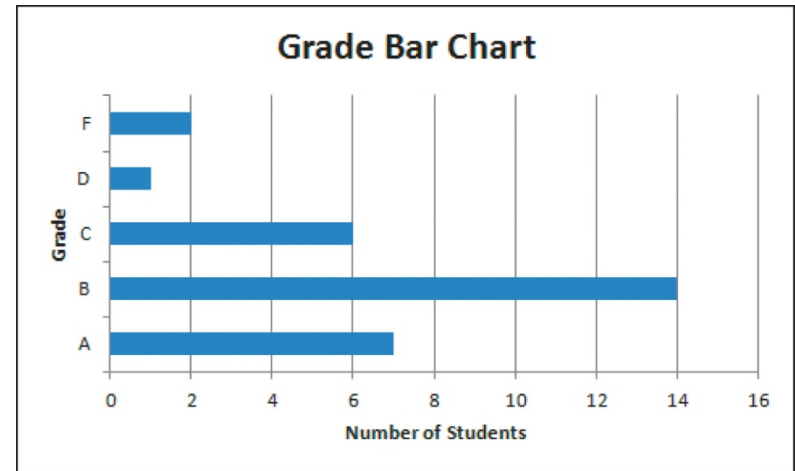


# Different types of Bar Charts

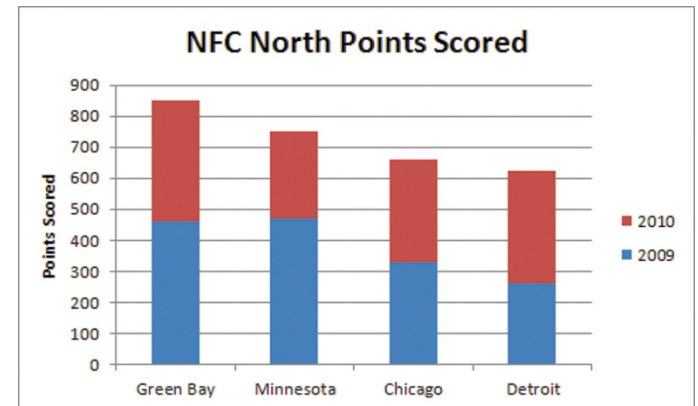
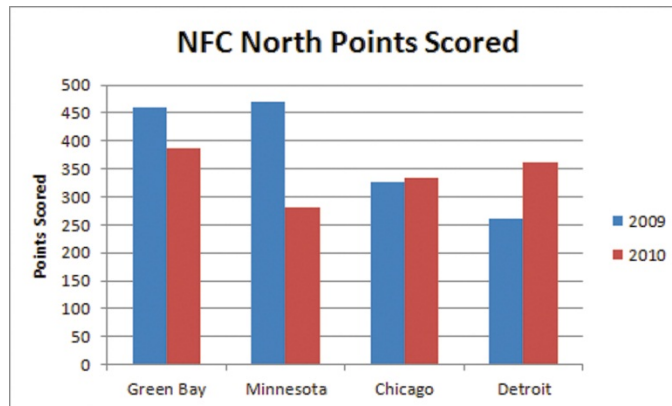
Horizontal bar chart



Vertical bar chart



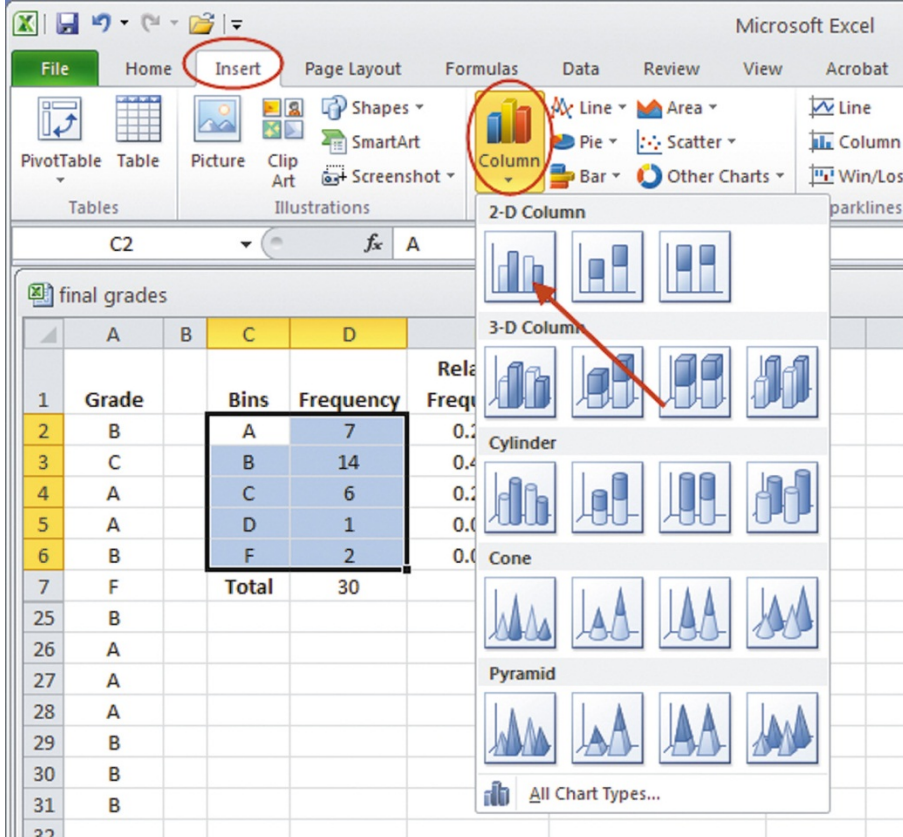
Can display multiple series with **clustered bar charts** or **stacked bar charts**:



# Bar Charts in Excel

## Exercise:

Create a bar chart for frequency distribution that was calculated for Trucks.xls



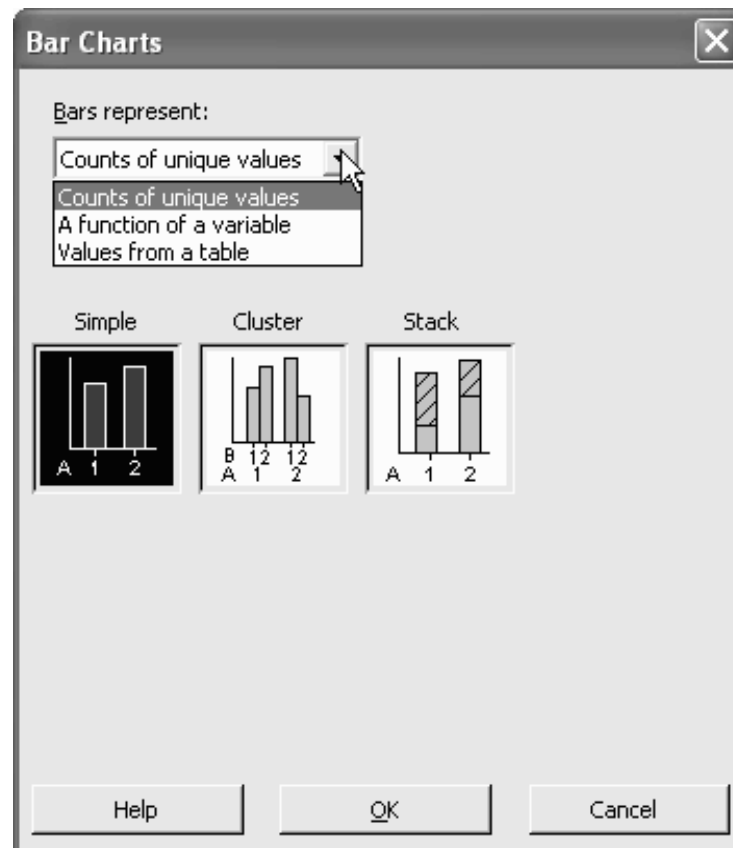
The screenshot shows the Microsoft Excel interface. The 'Insert' tab is selected and circled in red. The 'Column' chart type is highlighted in the 'Illustrations' group. A red arrow points to the '2-D Column' chart type. The background shows a data table with columns for Grade, Bins, and Frequency.

	A	B	C	D	
1	Grade		Bins	Frequency	Rela
2	B		A	7	0.2
3	C		B	14	0.4
4	A		C	6	0.2
5	A		D	1	0.0
6	B		F	2	0.0
7	F		Total	30	
25	B				
26	A				
27	A				
28	A				
29	B				
30	B				
31	B				
32					

# Bar Charts in Minitab

- For creating bar charts in Minitab, go to :

## Graph . Bar Chart ...



# Bar Charts in SAS (Optional)

- Using the following code :

```
/*--Set output size--*/
```

```
ods graphics / reset width=6.4in height=4.8in imagemap;
```

```
/*--SGPLOT proc statement--*/
```

```
proc sgplot data=SASHELP.CARS noautolegend;
```

```
/*--Bar chart settings--*/
```

```
vbar Make / transparency=0.00 dataskin=None name='Bar';
```

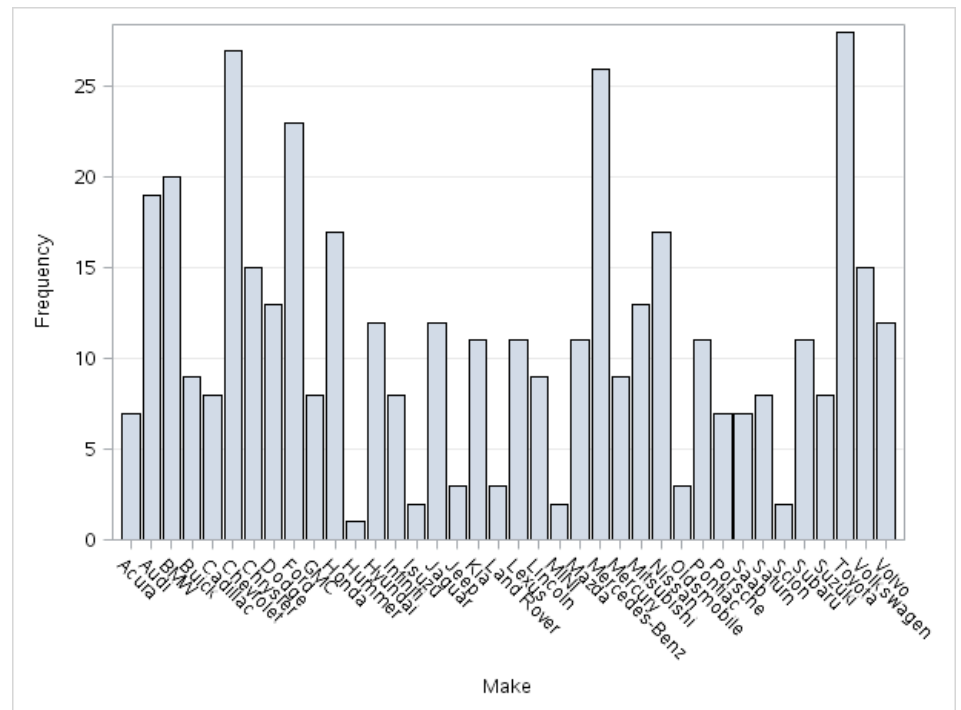
```
/*--Category Axis--*/
```

```
xaxis;
```

```
/*--Response Axis--*/
```

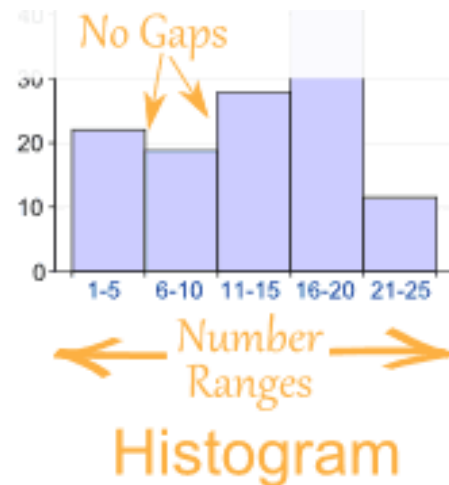
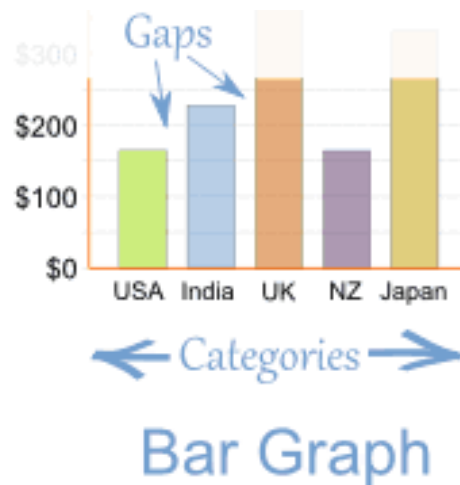
```
yaxis grid;
```

```
run;
```



# Histograms Vs. Bar Graphs.

- When do you use **Bar chart** and when **Histogram** ?

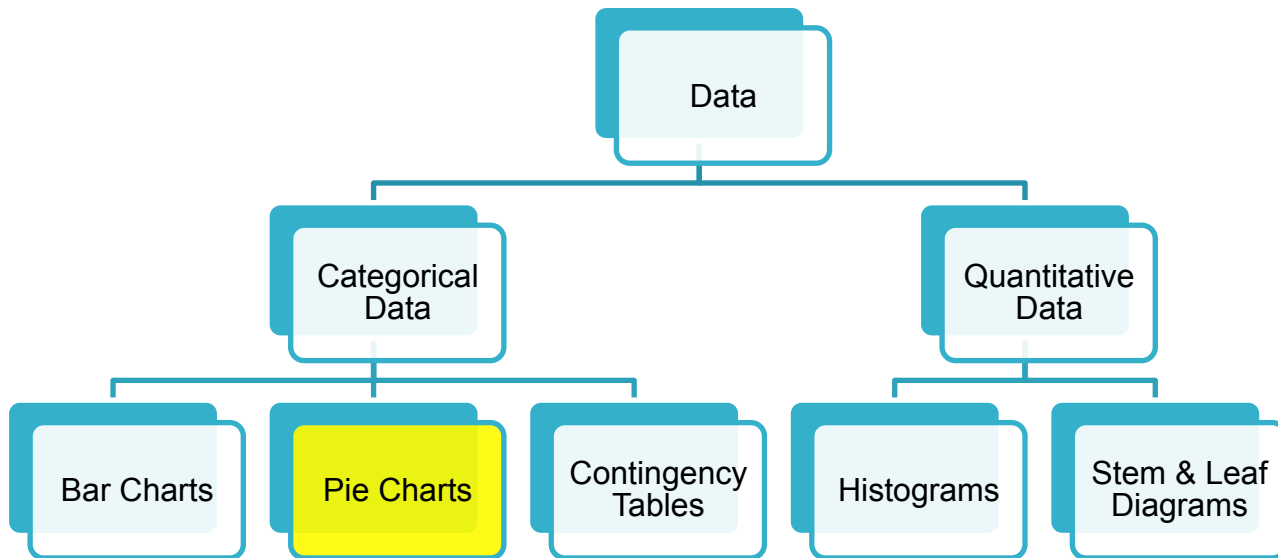


Bar Graphs are good when your data is in **categories** (such as "Comedy", "Drama", etc).

But when you have continuous data (such as a person's height) then use a Histogram.

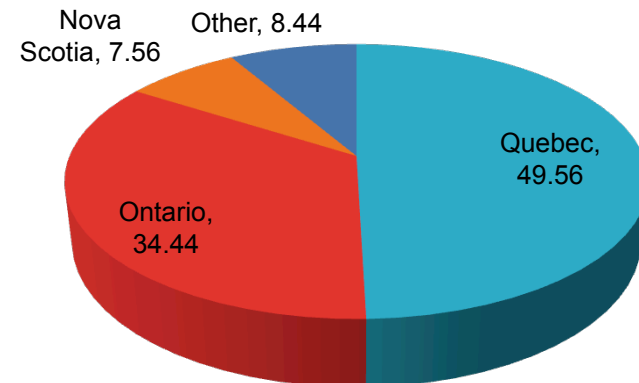
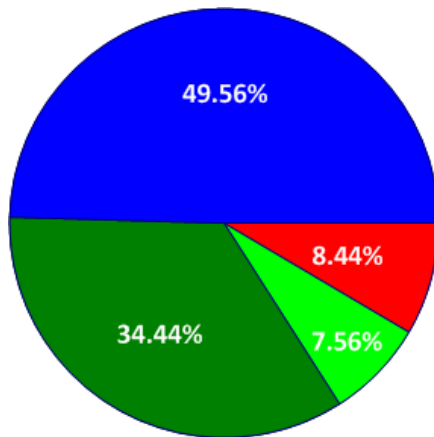
# Outline

---



# Pie Charts

- **Pie charts** show the whole group of cases as a circle sliced into pieces with sizes proportional to the fraction of the whole in each category.
- Each segment of the pie represents the relative frequency of one category
- The Loblaw's data is displayed below.

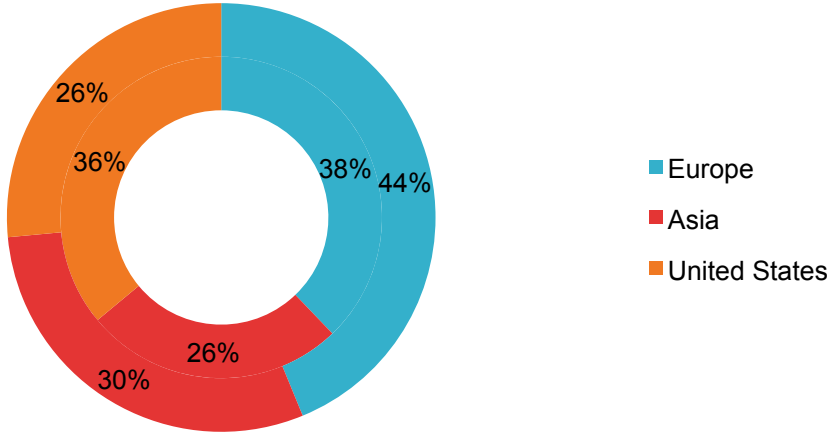


- Use a pie chart to compare the relative sizes of all possible categories

# Donut Charts

	2005	2006
Europe	\$12,704,714.00	\$17,987,034.00
Asia	\$8,774,099.00	\$12,214,447.00
United States	\$12,094,215.00	\$10,873,099.00

Sales by Region



# Pie Charts in Excel

## Constructing a Pie Chart in Excel

The screenshot shows the Microsoft Excel interface with the 'Insert' tab selected. The 'Pie' icon in the Charts group is highlighted with a red circle. A dropdown menu is open, showing various pie chart options. The first option under '2-D Pie' is highlighted with a yellow border and a red arrow points to it. The data table below shows the source data for the chart.

	A	B	C	D	E	H
1	Company	Units Shipped				
2	HP	4228				
3	Dell	3996				
4	Acer	2076				
5	Apple	1135				
6	Toshiba	1005				
7	Others	2837				
8						
9						

# Pie Charts in Excel

## Constructing a Pie Chart in Excel (continued)

The screenshot shows the Microsoft Excel interface with the 'Chart Tools' ribbon selected. The 'Data Labels' dropdown menu is open, showing options for displaying data labels on the pie chart. The 'Best Fit' option is highlighted, and a red arrow points to it. The pie chart is titled 'Units Shipped' and is based on the data in the table below.

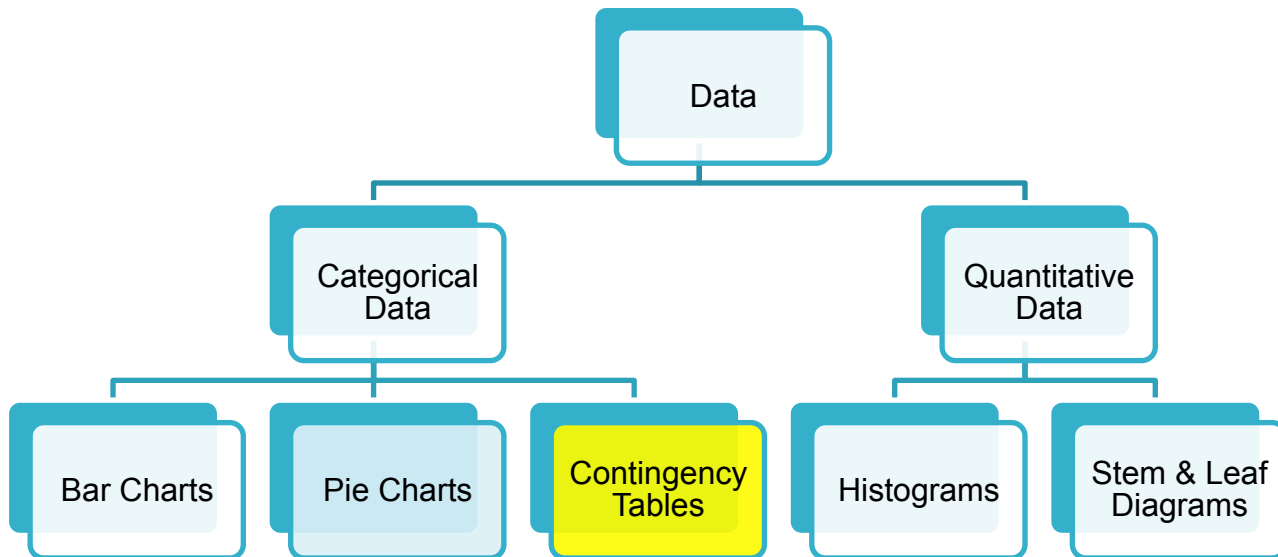
Company	Units Shipped
HP	4228
Dell	3996
Acer	2076
Apple	1135
Toshiba	1005
Others	2837

**Units Shipped**

- HP
- Dell
- Acer
- Apple
- Toshiba
- Others

# Outline

---



# Contingency Table

- **Contingency tables** provide a format to display observations that have more than one value associated with them
- **Example** : Survey question: “I have a strong preference for regional or traditional products and dishes from where I come from.”

		Regional Preference					Total	
		Agree Completely	Agree Somewhat	Neither Disagree nor Agree	Disagree Somewhat	Disagree Completely		Don't Know
Country	China	518	576	251	117	33	7	1502
	France	347	475	400	208	94	15	1539
	India	960	282	129	65	95	4	1535
	U.K.	214	407	504	229	175	28	1557
	U.S.	307	477	454	192	101	26	1557
Total		2346	2217	1738	811	498	80	7690

**Table 4.4** Contingency table of Regional Preference and Country. The bottom-line “totals” are the values that were in Table 4.3.

Count

Marginal Distribution

# Contingency Table

---

## Continued :

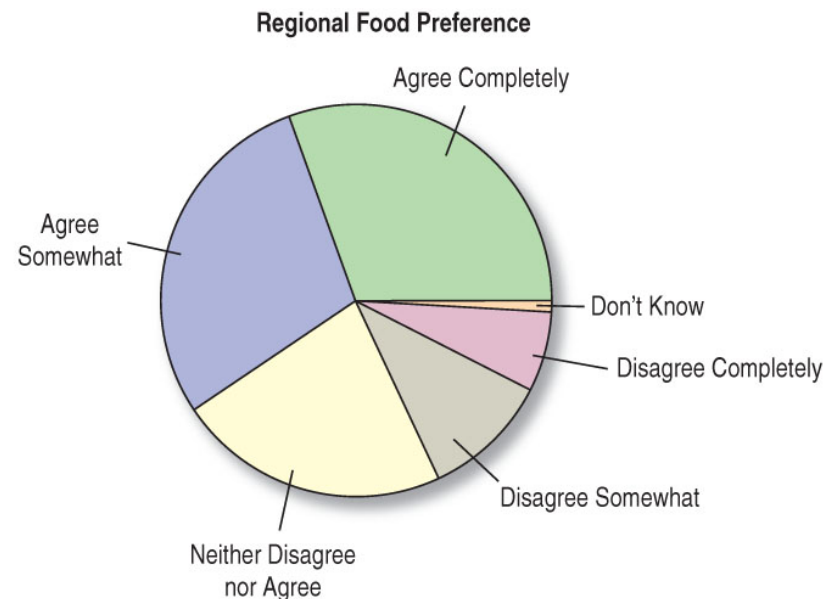
- To show how opinions on regional foods varied by countries, we can display the data in a *contingency table*.
- We have added introduced more detail by adding the countries as new variable.
- Each *cell* of a contingency table gives the count for a combination of values of both variables (e.g. Country, and Regional Preference).
- Here the indicated cell shows that 4 respondents from India (variable: *Country*) Don't Know (variable: *Regional Preference*) how they feel.
- The *marginal distribution* of a variable in a contingency table is the total count that occurs when the value of that variable is held constant.
- Here the marginal distribution shows the number of respondents from China was 1502, from France was 1539,

# Contingency Table

**Example:** Data was collected on the strength of consumer preferences for regional foods in their country. The data is displayed in both a frequency table and a pie chart.

Response to Regional Food Preference Question	Counts	Relative Frequency
Agree completely	2346	30.51%
Agree somewhat	2217	28.83%
Neither disagree nor agree	1738	22.60%
Disagree somewhat	811	10.55%
Disagree completely	498	6.48%
Don't know	80	1.04%
<b>Total</b>	<b>7690</b>	<b>100.01%</b>

**Table 4.3** A combined frequency and relative frequency table for the responses (from all five countries represented: China, France, India, the U.K., and the U.S.) to the statement "I have a strong preference for regional or traditional products and dishes from where I come from."



**Figure 4.5** It's clear from the pie chart that the majority of respondents identify with their local foods.

Copyright © 2012 Pearson Canada Inc.

# Relative Contingency Table

- We may display the data as a percentage – as a **total percent**, **row percent**, or **column percent**, which show percentages with respect to the total count, row count, or column count, respectively.

		Regional Preference					Total	
		Agree Completely	Agree Somewhat	Neither Disagree Nor Agree	Disagree Somewhat	Disagree Completely		Don't Know
Country	China	6.74	7.49	3.26	1.52	0.43	0.09	19.53
	France	4.51	6.18	5.20	2.70	1.22	0.20	20.01
	India	12.48	3.67	1.68	0.85	1.24	0.05	19.96
	U.K.	2.78	5.29	6.55	2.98	2.28	0.36	20.25
	U.S.	3.99	6.20	5.90	2.50	1.31	0.34	20.25
	Total	30.51	28.83	22.60	10.55	6.48	1.04	100.00

Corresponds to the AND probability.

# Contingency Table - Conditional Distributions

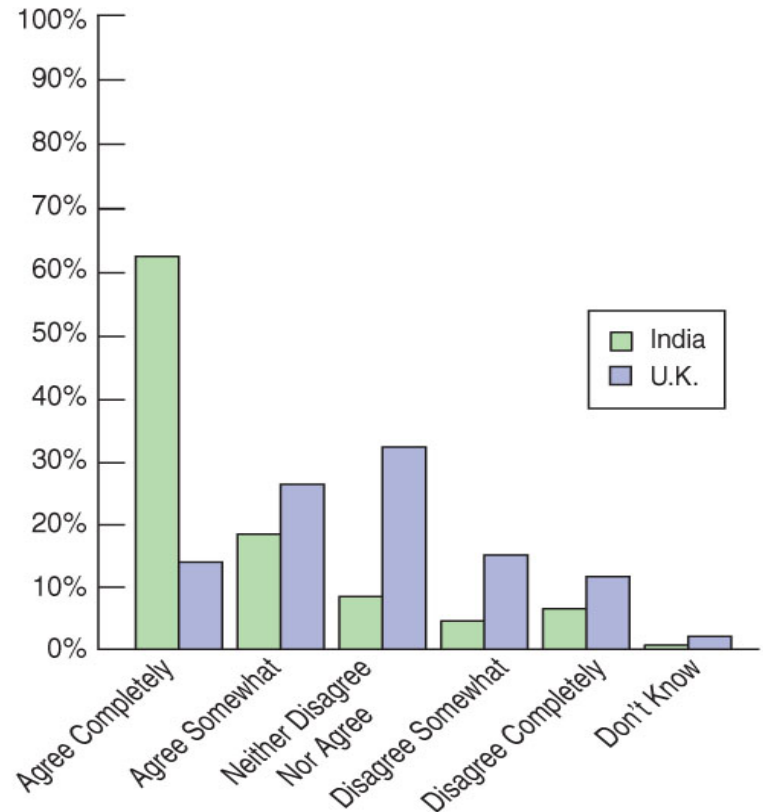
- We may want to restrict variables in a distribution to show the distribution for just those cases that satisfy a specified condition. This is called a *conditional distribution*.
- Here we consider only the preferences of the respondents from India and the U.K.

		Regional Preference						Total
		Agree Completely	Agree Somewhat	Neither Disagree nor Agree	Disagree Somewhat	Disagree Completely	Don't Know	
Country	India	960 62.54	282 18.37	129 8.40	65 4.23	95 6.19	4 0.26	<b>1535</b> <b>100%</b>
	U.K.	214 13.74	407 26.14	504 32.37	229 14.71	175 11.24	28 1.80	<b>1557</b> <b>100%</b>

**Table 4.7** The conditional distribution of Regional Preference conditioned on two values of Country: India and the United Kingdom. This table shows the row percentages.

# Contingency Tables - Bar Graph Side by Side Comparison

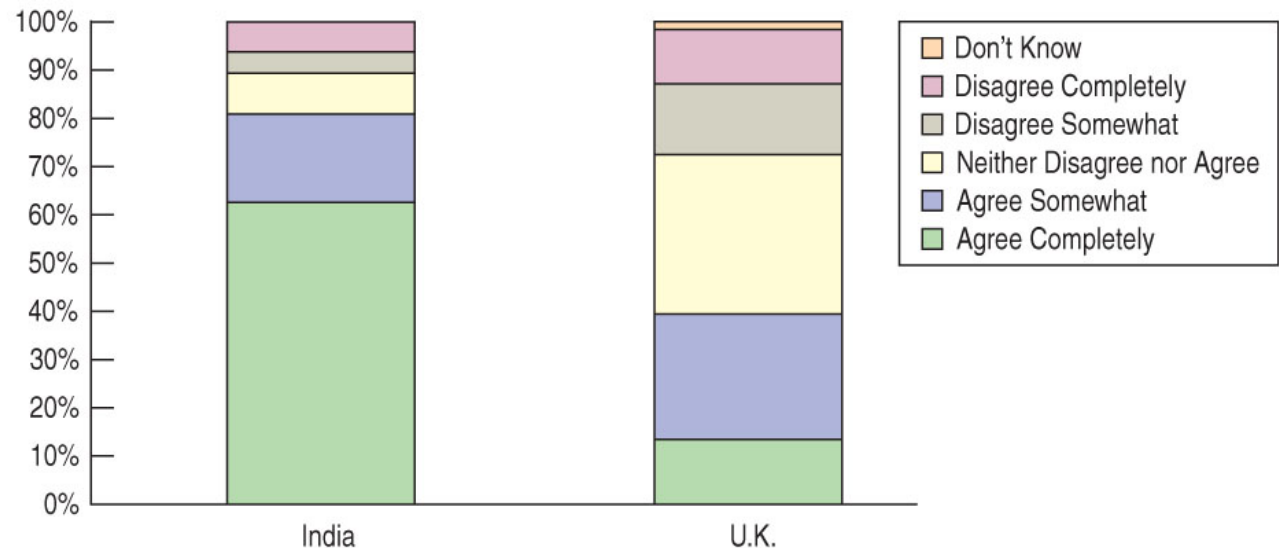
- We may display the results of a conditional distribution as a pie chart or as a bar graph.
- The data from the previous table is displayed here as a side-by-side bar chart.
- In a contingency table, when the distribution of one variable is the same for all categories of another variable, we say that the variables are *independent*. (That is, they are not associated.)
- If in a side-by-side bar chart, each pair of bars was the same height, then we can conclude the variables are independent.



**Figure 4.7** Side-by-side bar charts showing the conditional distribution of Regional Food Preference for both India and the United Kingdom. It's easier to compare percentages within each country with side-by-side bar charts than with pie charts.

# Contingency Tables - Segmented (Stacked) Bar Chart

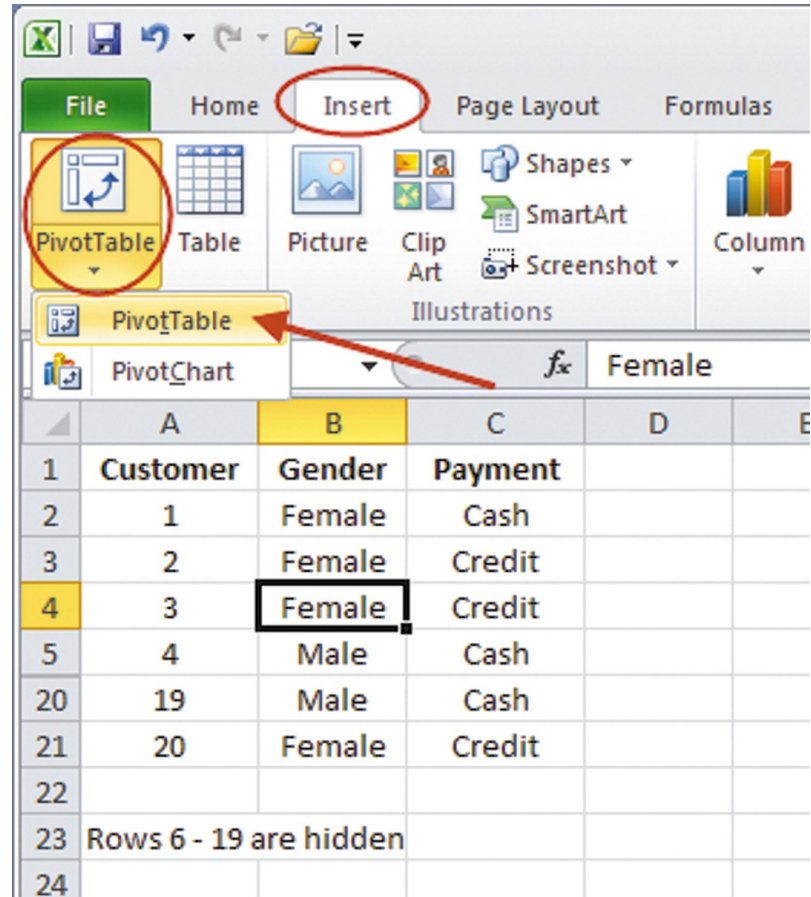
- Data can be displayed by dividing up bars rather circles. A *segmented bar chart* divides a bar proportionally into segments corresponding to the percentage in each group.
- The data from the conditional distribution pertaining to India and the U.K. are displayed here as segmented bar charts.



**Figure 4.8** Although the totals for India and the United Kingdom are different, the bars are the same height because we have converted the numbers to percentages. Compare this display with the side-by-side pie charts of the same data in Figure 4.6.

# Constructing a Contingency Table in Excel

1. Click on any cell within your data
2. Choose the **Insert** tab
3. Click on the **Pivot Table** icon
4. Click on **Pivot Table** in the drop-down menu
5. A **Create Pivot Table** dialog box will appear. Click **OK...**



# Constructing a Contingency Table in Excel

- 6. A new worksheet will be created for your pivot table

- From the **Pivot Table Field List**,

The screenshot shows an Excel PivotTable and its corresponding PivotTable Field List task pane. The PivotTable is located in the range A3:D8 and has the following data:

Row Labels	Female	Male	Grand Total
Cash	14	56	70
Credit	66	26	92
Debit	43	5	48
<b>Grand Total</b>	<b>123</b>	<b>87</b>	<b>210</b>

The PivotTable Field List task pane is on the right, titled "PivotTable Field List". It contains the following fields:

- Customer** (checked)
- Gender** (checked)
- Payment** (checked)

The task pane is divided into four areas:

- Report Filter:** Empty.
- Column Labels:** Gender.
- Row Labels:** Payment.
- Values:** Sum of Customer.

Annotations in the image:

- a:** A red arrow points from the "Customer" field in the list to the "Values" area.
- b:** A red arrow points from the "Gender" field in the list to the "Column Labels" area.
- c:** A red arrow points from the "Payment" field in the list to the "Row Labels" area.

**a & b.** Drag the desired variable names down into the **Column or Row Labels** boxes, as desired

**c.** Drag the variable name to be summarized down into the **Values** box

# Constructing a Contingency Table in Excel

- Resulting pivot table:

Creating a Pivot Table in Excel (Final Result)

	A	B	C	D	
1	Count of Customer	Column Labels			
2	Row Labels	Female	Male	Grand Total	
3	Cash	2	5	7	
4	Credit	7	2	9	
5	Debit	3	1	4	
6	Grand Total	12	8	20	
7					

# Constructing a Contingency Table in Excel

---

## Exercise :

- Open file **high-desert.xls**
  - Click a cell within data
  - Insert; Pivot table; ; existing worksheet
  - Click and drag:
    - Row: type;
    - Column: region
    - Data: sum of loan amount;
  - Try a few other options
- Open file **Capital.xls**
  - Click a cell within data
  - Insert; Pivot table ; existing worksheet
  - Row: credit card; Column: Gender
  - Data: credit card
  - To group rows in table, right click in column A; group;
  - Click on “Sigma Values” “Value Field Settings” to get other options.

# Constructing a Contingency Table in Minitab

- For constructing contingency tables in Minitab, go to :

**Stat . Tables . Cross Tabulations and Chi squares ...**

