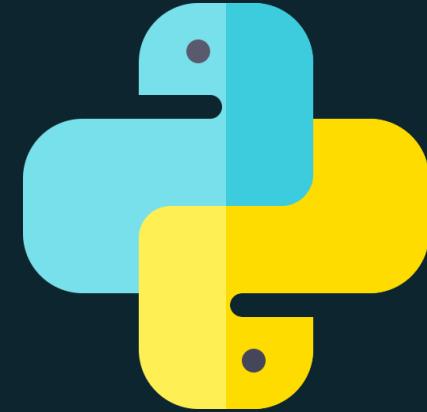


Brief Python

Python Course for Programmers

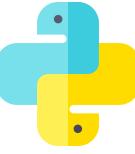


Learn Python Language for Data Science

CHAPTER 4: OVERVIEW OF DATA SCIENCE WITH EXCEL

DR. ERIC CHOU

IEEE SENIOR MEMBER



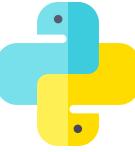
Objectives

- Excel Charts
- Type of Charts
- Edward Tufte's Data Graphic Principles
- Different types of charts/graphs in Excel



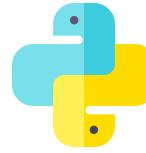
Excel Charts

LECTURE 1



Charts

- Charts provide a way of presenting and comparing data in graphical format.
- Embedded charts or chart sheets
 - Embedded charts are objects in the worksheet along with the data.
 - Chart sheets – Exist on a separate sheet by itself.
- All charts are linked to the charted data. The chart data appears in the plot area.



Terminology

Chart Types:

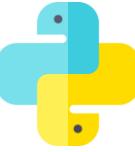
- Bar Graph
- Column Chart
- Line Graph
- Pie Chart
- XY Scatter Chart
- Histogram

Chart Properties:

- Data Labels
- Data Markers
- Data Series
- Plot Area
- Legends
- Data Point

Other Terms:

- Chart
- Chart Sheet
- Chart Wizard
- Embedded Chart
- Worksheet
- Category Axis
- Value Axis



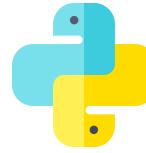
Charts and Data

A Chart is a graphical representation of Data.

The Chart area is called **data source**.

- It is usually a range of cells in the worksheet.
- You will need to highlight the data source before creating the chart.

Use the **chart wizard** to create your charts.



Data Series

| Community | Planned | Started | Completed |
|----------------|---------|---------|-----------|
| Oak Bend | 6 | 10 | 15 |
| River Knoll | 9 | 7 | 12 |
| Glenview North | 12 | 15 | 17 |

Data of a same
column is named a
data series

Q2 Home Construction

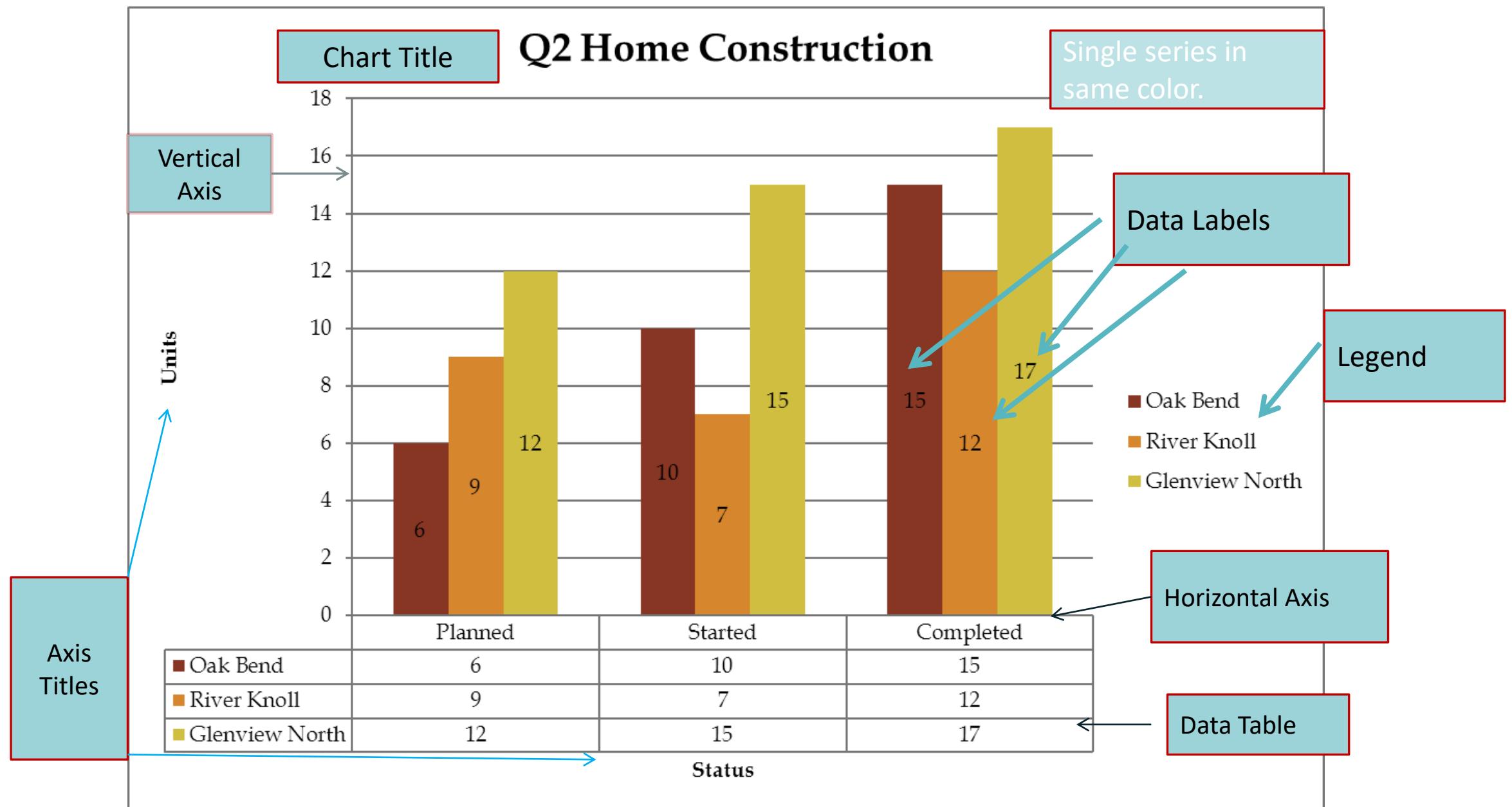
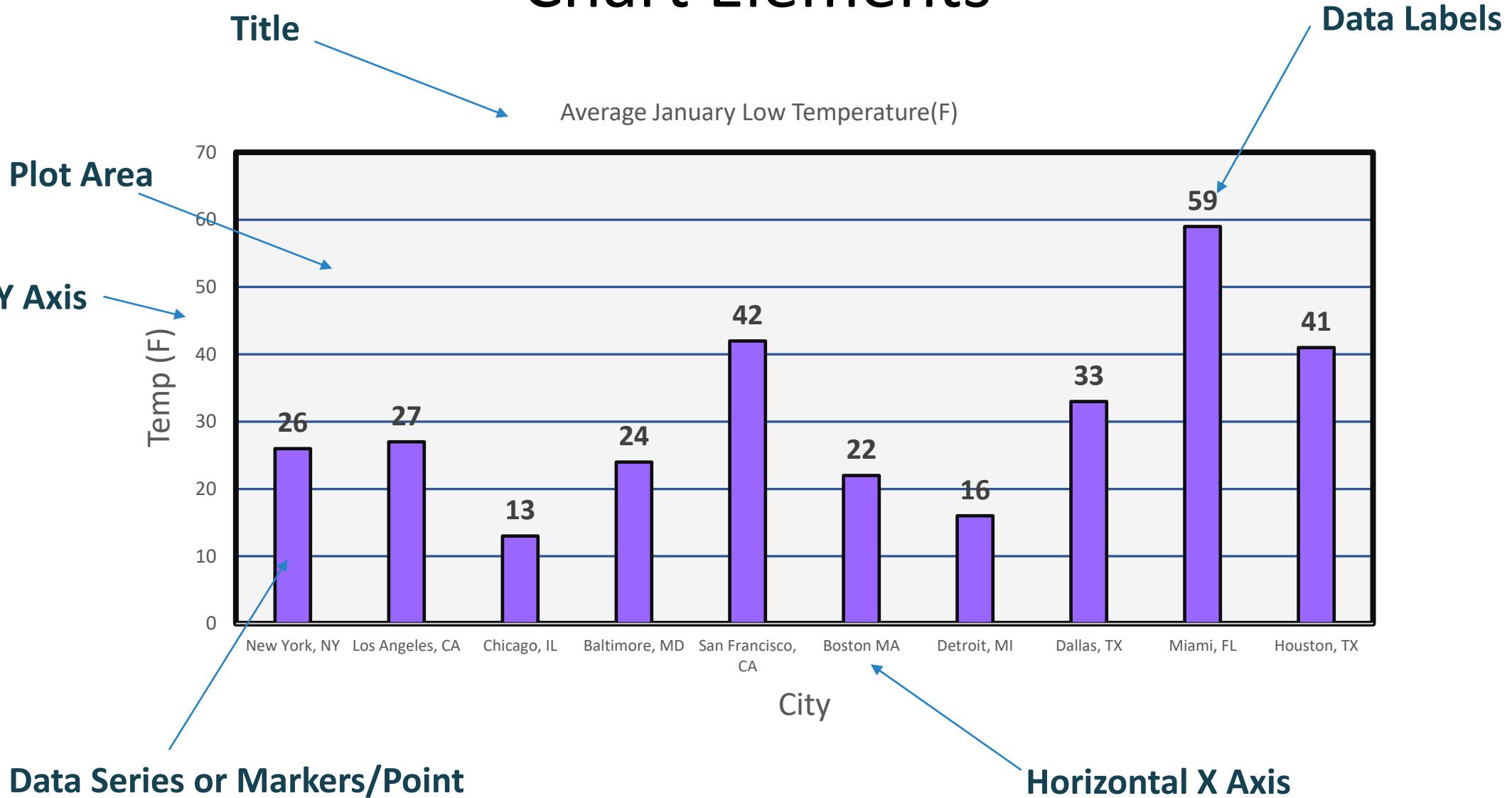


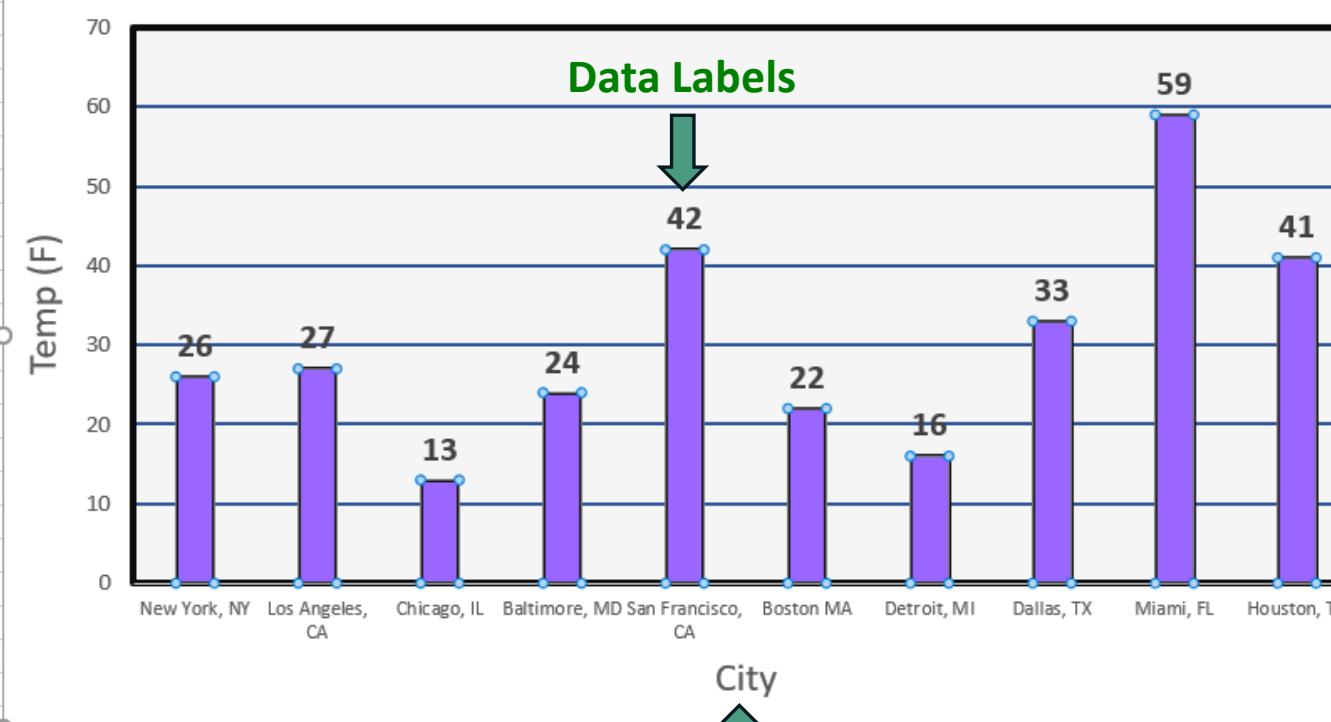
Chart Elements



| City | Temp(F) |
|-------------------|---------|
| New York, NY | 26 |
| Los Angeles, CA | 27 |
| Chicago, IL | 13 |
| Baltimore, MD | 16 |
| San Francisco, CA | 33 |
| Dallas, TX | 59 |
| Miami, FL | 41 |
| Houston, TX | 41 |

Category Labels

Data Series



Title

Data Labels

42

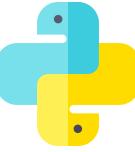
Axis Label

Chart Elements



Gridline

Legends



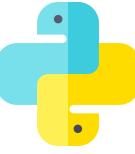
Vocabulary

Title: headings that identify the contents of the chart or the axes in the chart; most charts have a chart title and titles for each axis.

Axes: lines that establish a relationship between data in a chart.

Category Axis: is what is shown in a chart, created from row or column headings; the x and y axis.

Value Axis: The value axis is the horizontal or vertical grouping of values from the worksheet.



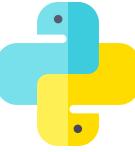
Vocabulary

Plot Area: rectangular area bound by the category and value axis.

Legend: describes all the data markers or points in a chart.

Gridlines: lines through a chart that relate the data in a chart to the axes.

Data Labels: text or numbers that identify the values depicted by the chart objects.



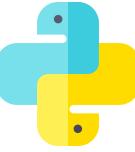
Vocabulary

Plot Area: rectangular area bound by the category and value axis.

Legend: describes all the data markers or points in a chart.

Gridlines: lines through a chart that relate the data in a chart to the axes.

Data Labels: text or numbers that identify the values depicted by the chart objects.

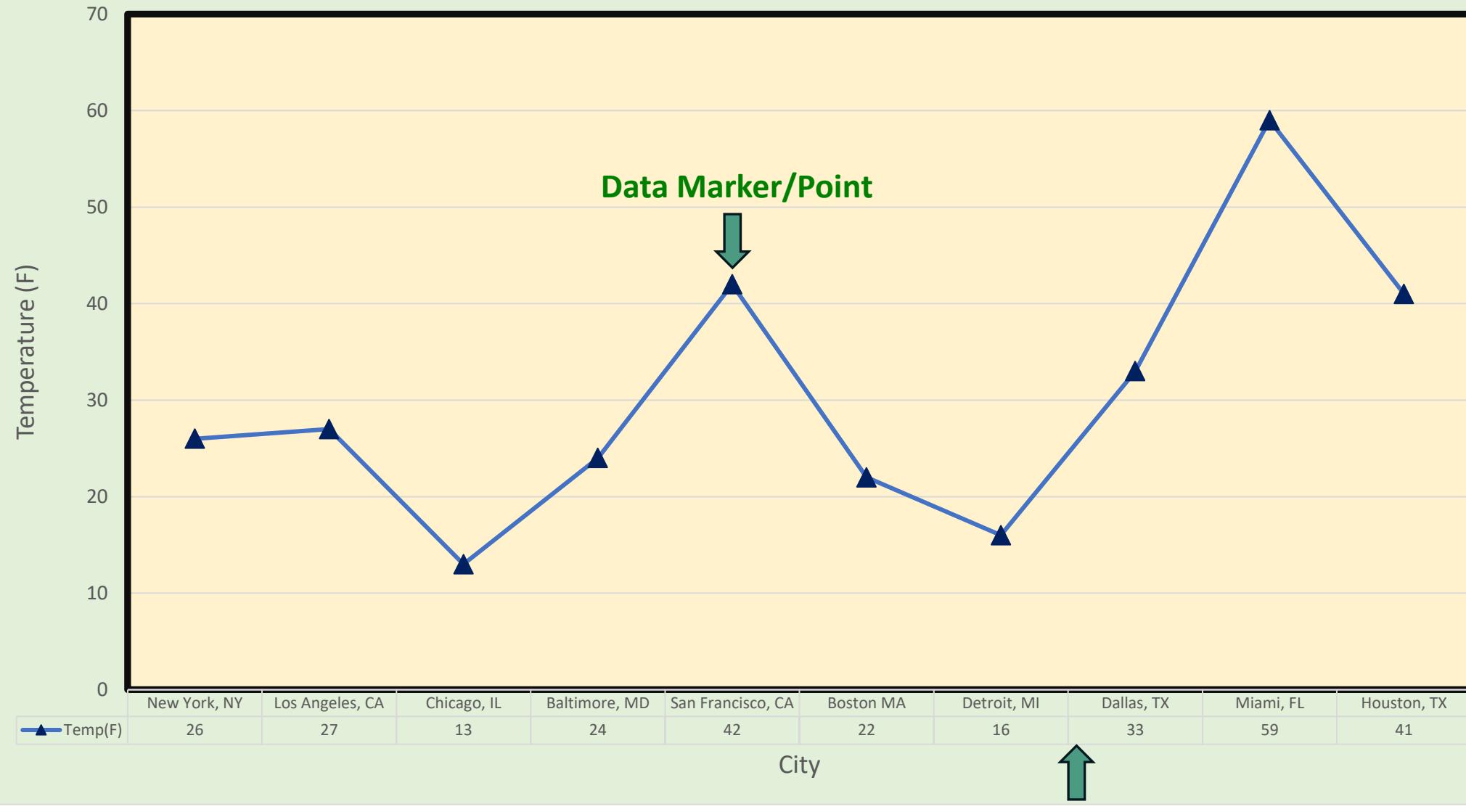


Vocabulary

Data Markers/Points: a chart symbol(such as a bar, line, dot, slice, and so forth) that represents a single data point of value from the corresponding worksheet cell.

Data Table: Data Series values displayed in a grid below the chart.

Average January Low Temperature

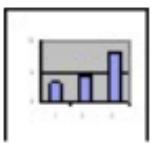




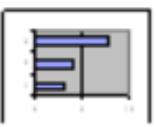
Types of Charts

LECTURE 2

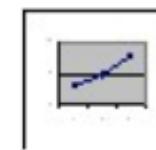
Chart Types



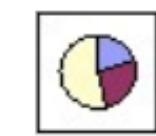
- **Column** —uses vertical bars of varying height



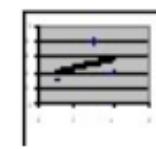
- **Bar** —uses horizontal bars of varying height



- **Line** —uses points connected by a line; good to show trends or growth.



- **Pie** —shows the relationship of a part to a whole

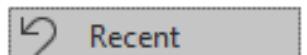


- **Scatter** —(XY chart) shows the relationship between two categories of data



Recommended Charts

All Charts



Recent



Templates



Column



Line



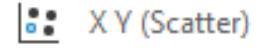
Pie



Bar



Area



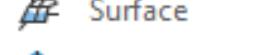
X Y (Scatter)



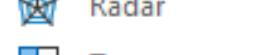
Map



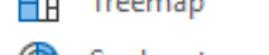
Stock



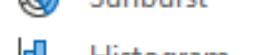
Surface



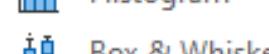
Radar



Treemap



Sunburst



Histogram



Box & Whisker



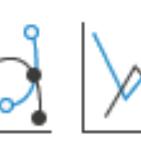
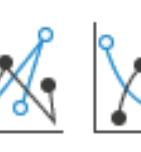
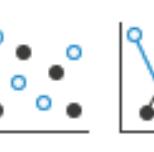
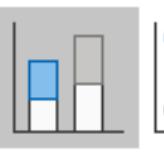
Waterfall



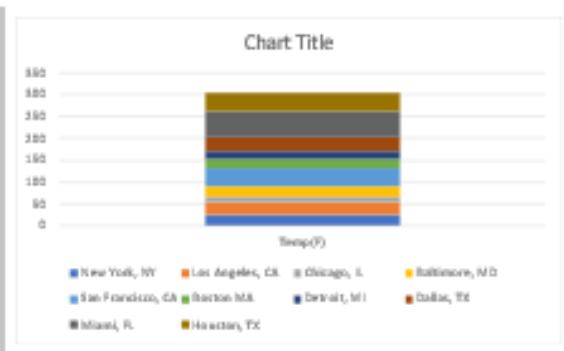
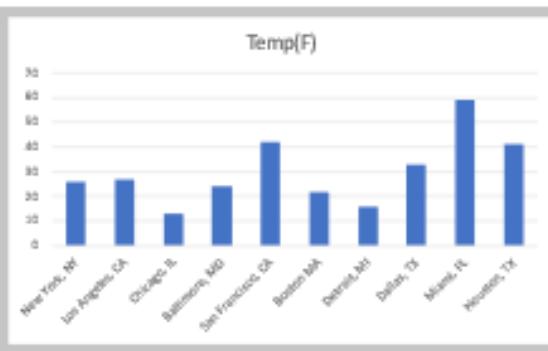
Funnel

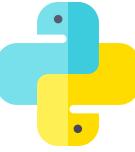


Combo



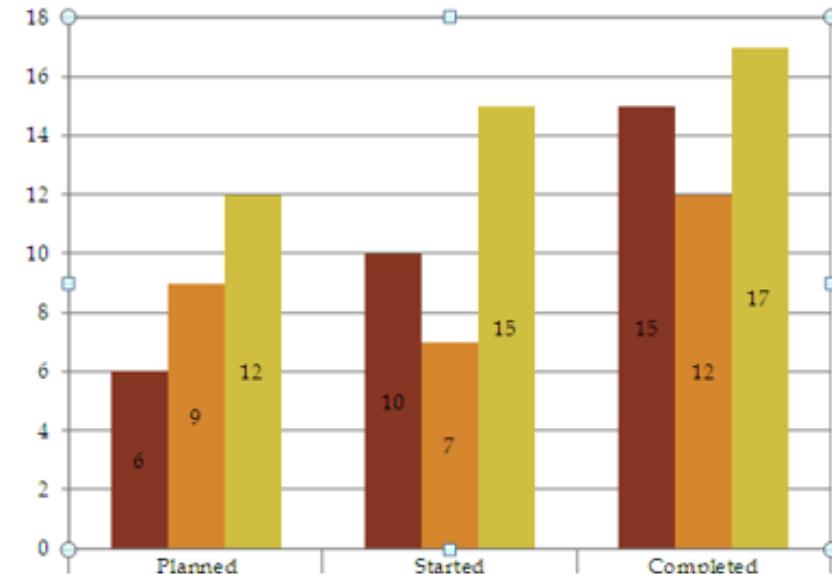
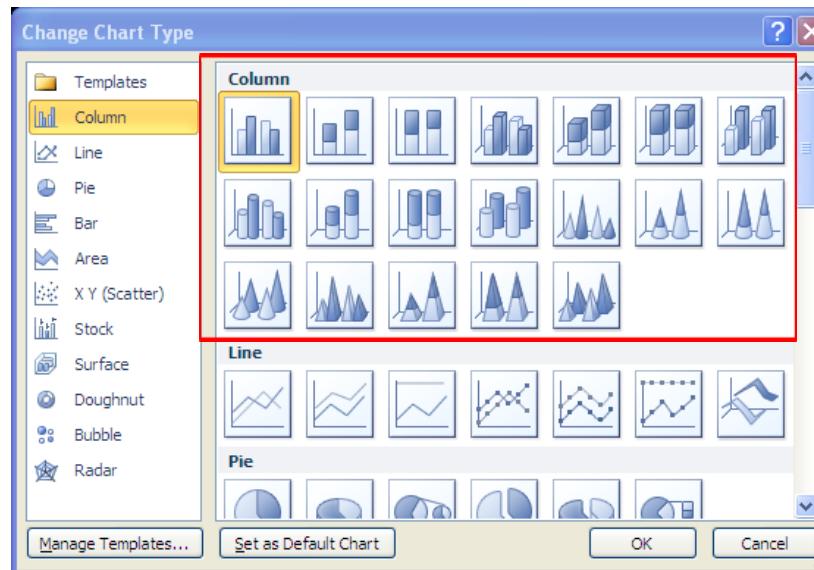
Stacked Column

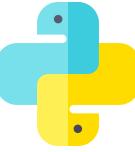




Column Chart

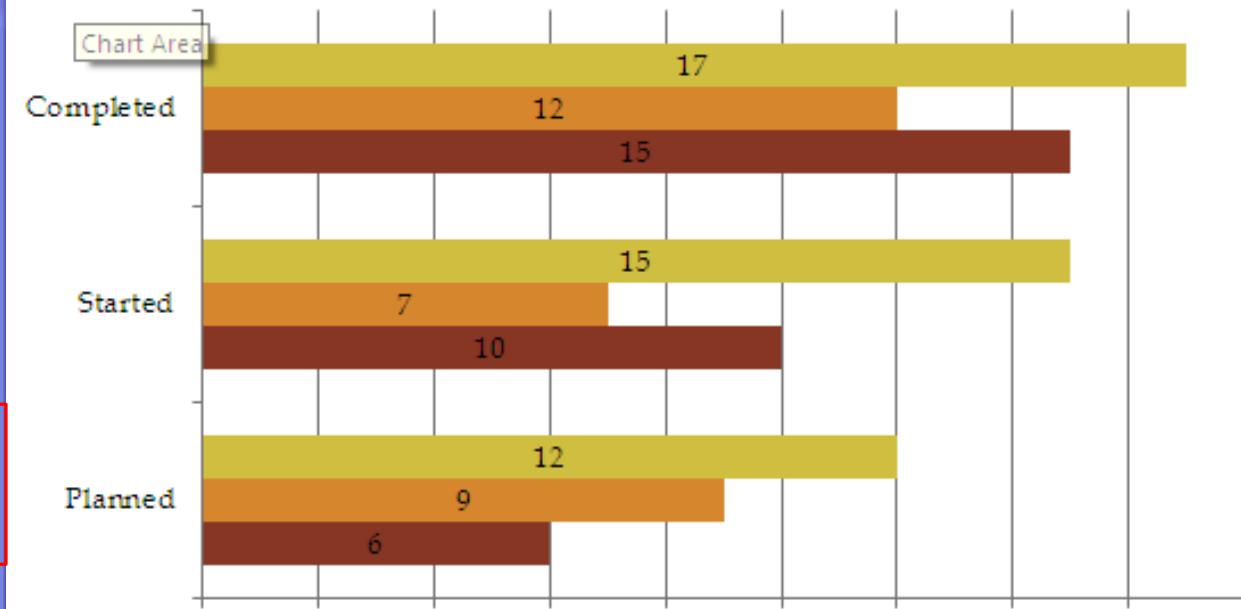
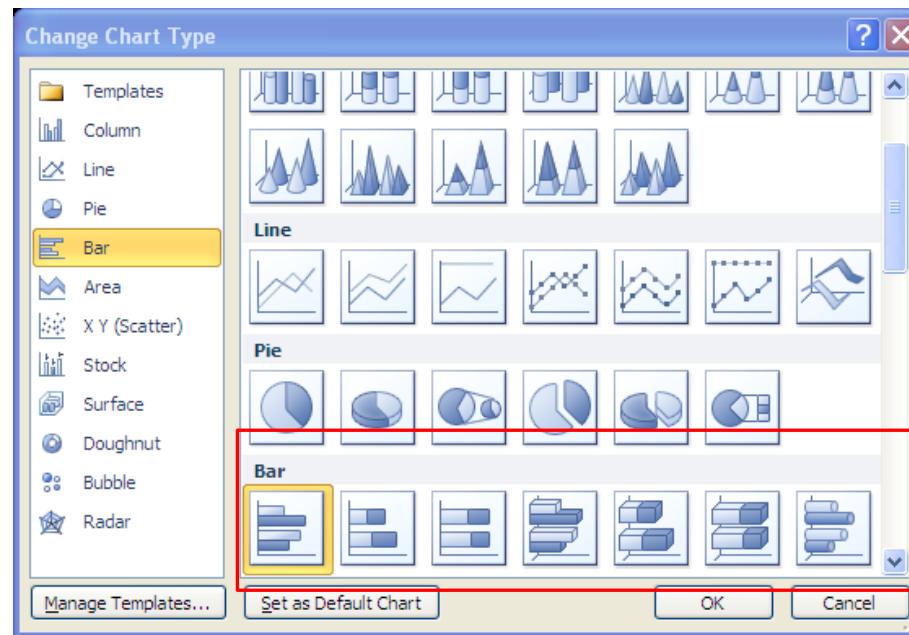
- Compare individual or sets of values.
- The height of each bar corresponds to its value

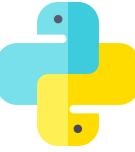




Bar Chart

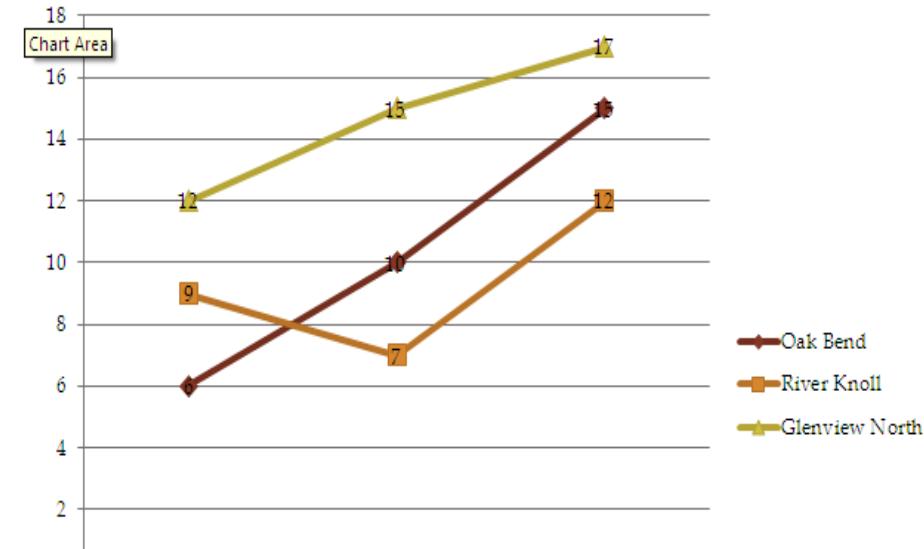
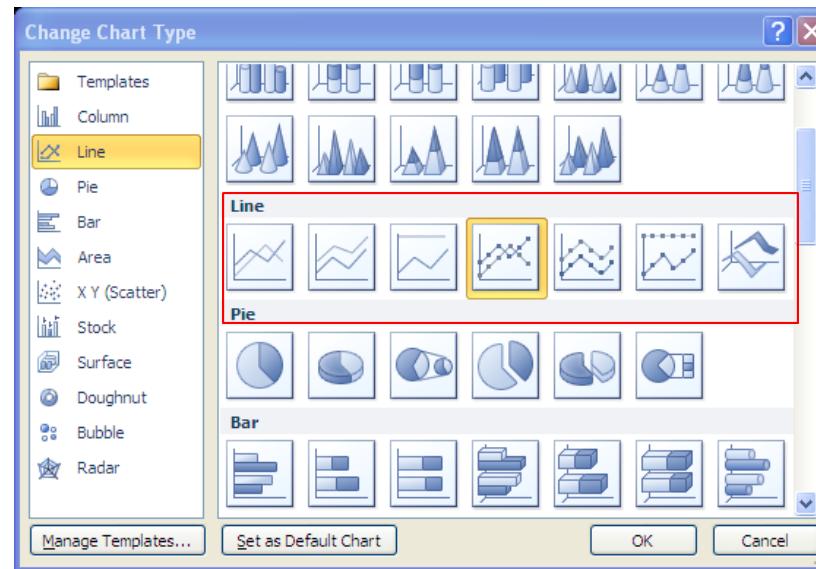
- Illustrates comparisons among individual items.

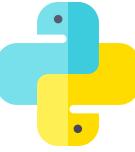




Line Chart

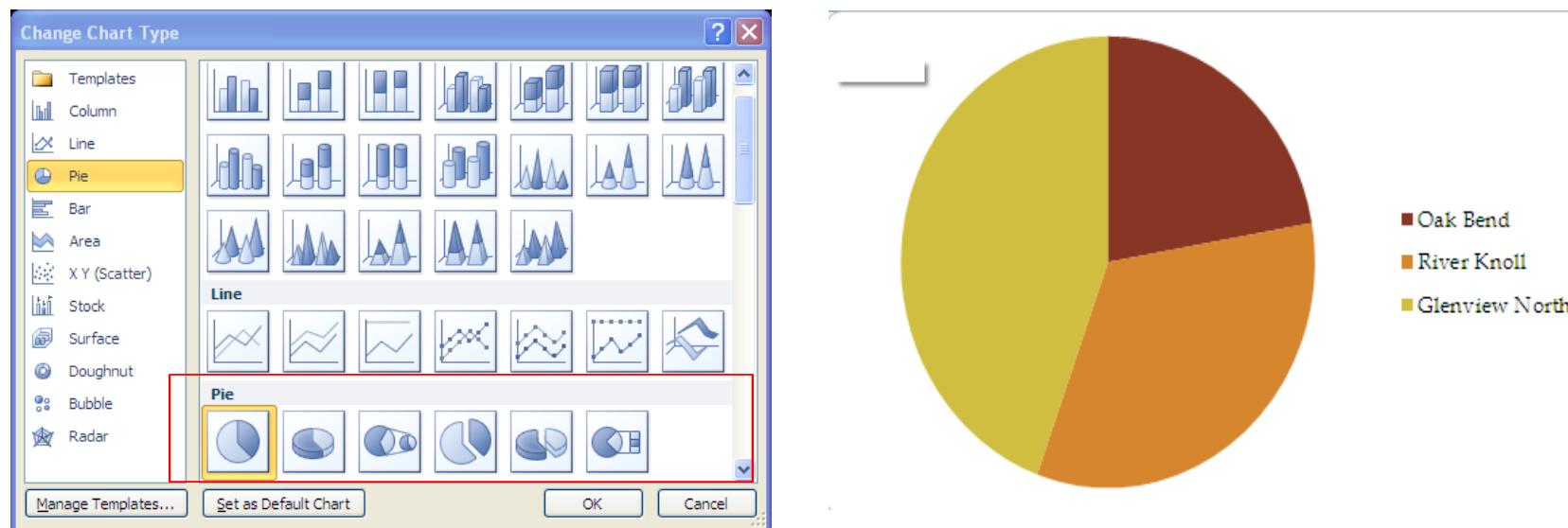
- Shows changes over time, such as trends
- Plotting trends.
- Two or more types of data.

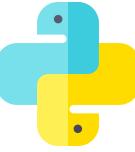




Circle Pie Chart

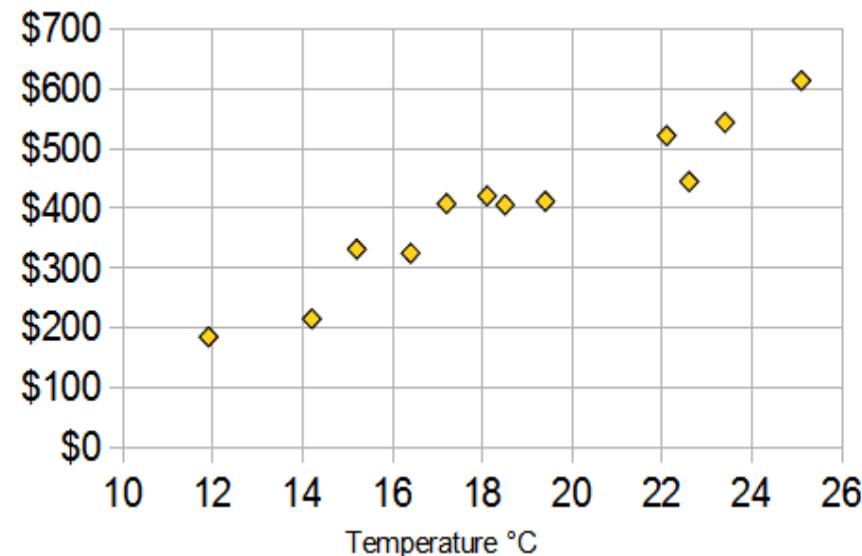
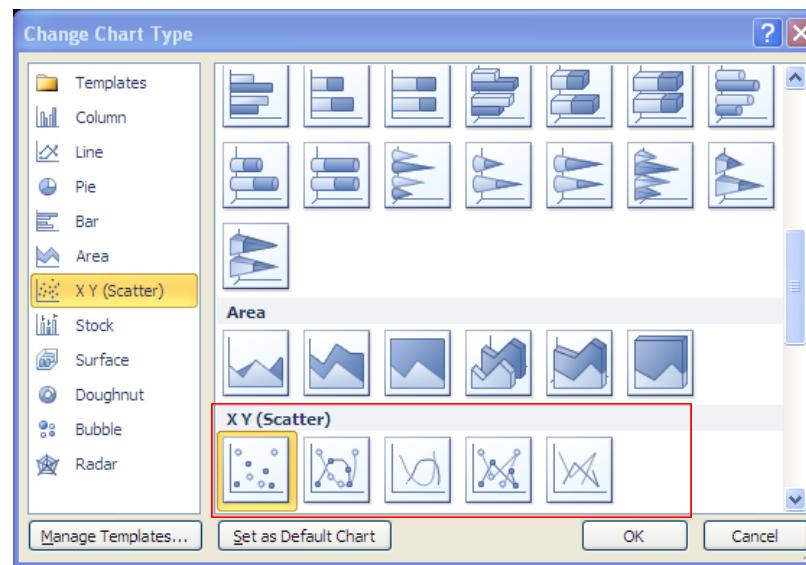
- Shows the relationships of each value in a single series to the whole
- The size of a pie wedge represents the percentage that value contributes to the total

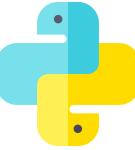




Scatter Chart

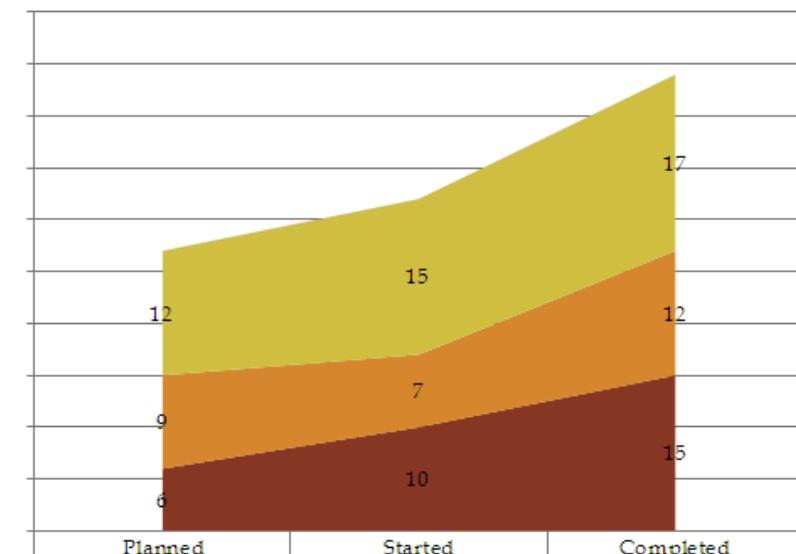
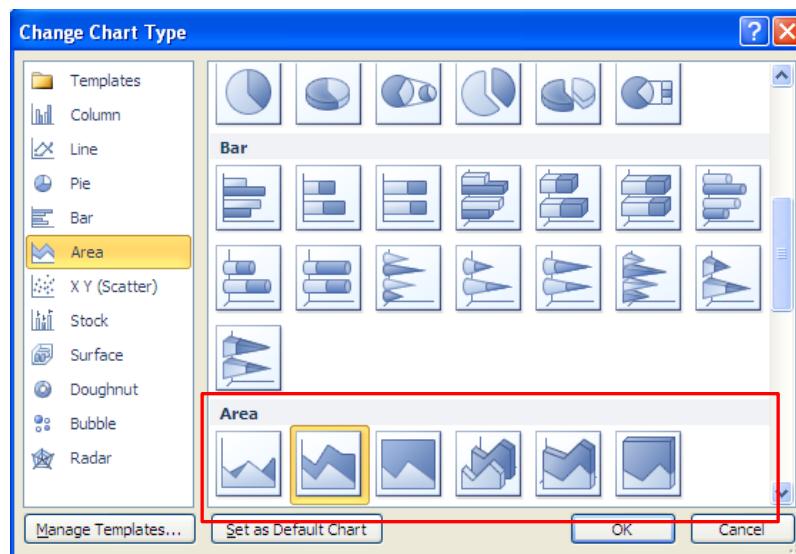
- Also known as XY charts
- Represent data points as dots. The dots for each series appear in different colors.
- The position of the dots reveals a trend.

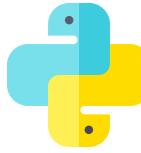




Area Chart

- Displays the magnitude of change over time
- Look like filled in line charts





More Chart Info

Printing

- Charts can be embedded (as part of the worksheet that contains the data).
- Charts can be placed on a separate chart sheet.

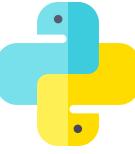
Changing the appearance of the data makers

- The data series, markers or points are usually displayed as solid color bar, slice, or column.
- The solid color can be replaced with gradients, patterns, and clip art.



Data Creation

LECTURE 3



What is a CSV file?

- A CSV file is a type of plain text file that uses specific structuring to arrange tabular data.
- CSV is a common format for data interchange as it's compact, simple and general. Many online services allow its users to export tabular data from the website into a CSV file.
- Files of CSV will open into Excel, and nearly all databases have a tool to allow import from CSV file. The standard format is defined by rows and columns data. Moreover, each row is terminated by a newline to begin the next row. Also within the row, each column is separated by a comma.

Free Comma Separating Tool

Do you often need to take a spreadsheet of data and convert to a comma-delimited list?

Be it for taking a list of zip codes or names to make an SQL query, or to take data from a CSV and be able to paste into an array. At delim.co we make that just a little easier.

Enter your non-delimited data on the left, hit the button, and boom, separated data on the right. Special configs are below if the defaults aren't what you need!

Column Data Here...

| | |
|----|---|
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |
| 6 | 6 |
| 7 | 7 |
| 8 | 8 |
| 9 | 9 |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | |
| 17 | |
| 18 | |
| 19 | |
| 20 | |



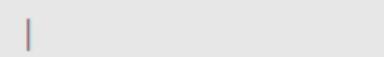
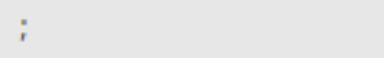
Converter Settings

Delimited Data Here...

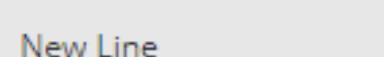
| | |
|----|-------------------|
| 1 | 1,2,3,4,5,6,7,8,9 |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | |
| 17 | |
| 18 | |
| 19 | |
| 20 | |



Delimiters



Spaces



<https://delim.co/#>



random csv generator world's simplest randomization tool

Free online random CSV file generator. Just press a button and get your random CSV file. There are no ads, popups or nonsense, just a random CSV data creator. Press a button, get a random CSV. Created by developers from team Browserling.



we wrote the curl cookbook!

Super exciting news – we just wrote the [Curl Cookbook](#) full of organic, nutritious, and completely irresistible curl recipes. Check it out!



random csv



```
"tried", "solid", "month", "pile", "end", "longer"  
"pine", "stay", "felt", "fix", "quiet", "silk"  
"something", "year", "sing", "lay", "battle", "zoo"  
"tales", "cost", "meet", "white", "vessels", "where"  
"bound", "separate", "climate", "card", "element", "shoe"  
"coach", "fuel", "time", "flame", "engineer", "clock"  
"deer", "growth", "lose", "volume", "church", "dinner"
```

[Chain with...](#)[Save as...](#)[Copy to clipboard](#)

<https://onlinerandomtools.com/generate-random-csv>



random csv generator options

CSV Options

“ 6

How many columns to generate?

“ 6

How many rows to generate?

First row counts as a header.

“ CSV Delimiter

Column-separating character

Uses comma if not specified.

Field Quotes

Quote all fields

Generator Options

Use random English words

Use random English words

for each CSV field

Use completely random strings

Use randomly generated strings

for each CSV field

“ 10

Maximum random string length?

(Only works with random strings)

“

!"#\$%&'0*+,./01234567
89;:<=>?
@ABCDEFGHIJKLMNP
QRSTUVWXYZ[N]^_abcd
efghijklmnopqrstuvwxyz
{})~

Alphabet to use for random strings



Copy of Explore example

File Edit View Insert Format Data Tools Add-ons Help Last edit was 2 minutes ago

Comments Share

| | A | B | C | D | E | F | G | H | I | J |
|----|------------------------------|------|---------------|-----------|----------------|--------------------|-------|---------|--------------|---------------|
| 1 | Edition | Year | Host Country | Winner | Runner up | Average attendance | Teams | Matches | Goals scored | Average goals |
| 2 | 1930 World Cup Uruguay | 1930 | Uruguay | Uruguay | Argentina | 32,808 | 13 | 18 | 70 | 3.9 |
| 3 | 1934 World Cup Italy | 1934 | Italy | Italy | Czechoslovakia | 21,353 | 16 | 17 | 70 | 4.1 |
| 4 | 1938 World Cup France | 1938 | France | Italy | Hungary | 20,872 | 15 | 18 | 84 | 4.7 |
| 5 | 1950 World Cup Brazil | 1950 | Brazil | Uruguay | Brazil | 47,511 | 13 | 22 | 88 | 4 |
| 6 | 1954 World Cup Switzerland | 1954 | Switzerland | Germany | Hungary | 29,562 | 16 | 26 | 140 | 5.4 |
| 7 | 1958 World Cup Sweden | 1958 | Sweden | Brazil | Sweden | 23,423 | 16 | 35 | 126 | 3.6 |
| 8 | 1962 World Cup Chile | 1962 | Chile | Brazil | Czechoslovakia | 27,912 | 16 | 32 | 89 | 2.8 |
| 9 | 1966 World Cup England | 1966 | England | England | Germany | 48,848 | 16 | 32 | 89 | 2.8 |
| 10 | 1970 World Cup Mexico | 1970 | Mexico | Brazil | Italy | 50,124 | 16 | 32 | 95 | 3 |
| 11 | 1974 World Cup Germany | 1974 | Germany | Germany | Netherlands | 49,099 | 16 | 38 | 97 | 2.6 |
| 12 | 1978 World Cup Argentina | 1978 | Argentina | Argentina | Netherlands | 40,679 | 16 | 38 | 102 | 2.7 |
| 13 | 1982 World Cup Spain | 1982 | Spain | Italy | Germany | 40,572 | 24 | 52 | 146 | 2.8 |
| 14 | 1986 World Cup Mexico | 1986 | Mexico | Argentina | Germany | 46,039 | 24 | 52 | 132 | 2.5 |
| 15 | 1990 World Cup Italy | 1990 | Italy | Germany | Argentina | 48,389 | 24 | 52 | 115 | 2.2 |
| 16 | 1994 World Cup United States | 1994 | United States | Brazil | Italy | 68,991 | 24 | 52 | 141 | 2.7 |
| 17 | 1998 World Cup France | 1998 | France | France | Brazil | 43,517 | 32 | 64 | 171 | 2.7 |
| 18 | 2002 World Cup Korea & Japan | 2002 | Korea & Japan | Brazil | Germany | 42,269 | 32 | 64 | 161 | 2.5 |
| 19 | 2006 World Cup Germany | 2006 | Germany | Italy | France | 52,491 | 32 | 64 | 147 | 2.3 |
| 20 | 2010 World Cup South Africa | 2010 | South Africa | Spain | Netherlands | 49,670 | 32 | 64 | 145 | 2.3 |
| 21 | 2014 World Cup Brazil | 2014 | Brazil | Germany | Argentina | 53,592 | 32 | 64 | 171 | 2.7 |
| 22 | | | | | | | | | | |
| 23 | | | | | | | | | | |

Source: https://en.wikipedia.org/wiki/FIFA_World_Cup

World Cup

Explore

ANSWERS

Ask a question about your data

- ② Distribution of Winner
- ② Correlation between Teams and Matches
- ② Average of Teams

FORMATTING

Alternating colors for A1:J21 EDIT

ANALYSIS

Teams, Matches and Goals scored

The chart displays three data series over time:

- Teams:** Represented by a blue line, starting around 20 in 1930 and rising steadily to approximately 32 by 2010.
- Matches:** Represented by a red line, starting around 15 in 1930 and rising steadily to approximately 64 by 2010.
- Goals scored:** Represented by a yellow line, starting around 70 in 1930 and rising steadily to approximately 171 by 2010.

Year

<https://www.google.com/sheets/about/>

Convert CSV to JSON

Use the tool on this page to convert CSV data to JSON

From CSV/Excel

- CSV To Delimited
- CSV To Flat File
- CSV To GeoJSON
- CSV To HTML Table
- CSV To JSON
- CSV To KML
- CSV To Markdown
- CSV To Multi-line Data
- CSV To PDF
- CSV To SQL
- CSV To XML
- CSV To YAML
- Pivot CSV
- Transpose CSV
- Query CSV with SQL

To CSV/Excel

- Flat File to CSV
- GeoJSON To CSV
- HTML Links To CSV
- HTML Table To CSV
- JSON To CSV
- KML To CSV
- SQL To CSV
- XML To CSV
- YAML To CSV

Data Tools

- CSV Escape Tool
- CSV Template Engine
- CSV Editor
- Generate Test Data
- Email Extractor
- Phone Extractor
- URL Extractor
- CSV Home

What can this tool do? ▾

What are my options? ▾

What else? ▾

Step 1: Select your input

Enter Data Choose File Enter URL

Enter or paste CSV here

Clear Input

Example

Step 2: Choose input options (optional) ▾

Step 3: Choose output options (optional) ▾

Step 4: Create Custom Output via Template (optional) ▾

Step 5: Generate output

Choose Conversion Type:

CSV To JSON

CSV To Keyed JSON

CSV To JSON Array

CSV To JSON Column Array

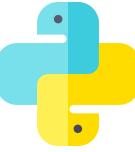
<http://www.convertcsv.com>





Python csv module

ACTIVITY



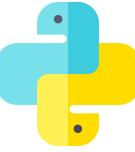
CSV Sample File

Demo Program: data.csv

- Data in the form of tables is also called CSV (comma separated values) - literally "comma-separated values." This is a text format intended for the presentation of tabular data. Each line of the file is one line of the table.
- The values of individual columns are separated by a separator symbol - a comma (,), a semicolon (;) or another symbol. CSV can be easily read and processed by Python.

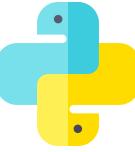
Table Data

| Programming language | Designed by | Appeared | Extension |
|----------------------|-------------------|----------|-----------|
| Python | Guido van Rossum | 1991 | .py |
| Java | James Gosling | 1995 | .java |
| C++ | Bjarne Stroustrup | 1983 | .cpp |



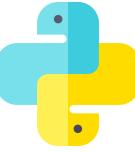
Python CSV Module

- Python provides a CSV module to handle CSV files. To read/write data, you need to loop through rows of the CSV. You need to use the split method to get data from specified columns.



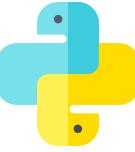
CSV Module Functions

- `csv.field_size_limit` – return maximum field size
- `csv.get_dialect` – get the dialect which is associated with the name
- `csv.list_dialects` – show all registered dialects
- `csv.reader` – read data from a csv file
- `csv.register_dialect` - associate dialect with name
- `csv.writer` – write data to a csv file
- `csv.unregister_dialect` - delete the dialect associated with the name the dialect registry



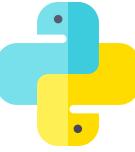
CSV Module Functions

- `csv.QUOTE_ALL` - Quote everything, regardless of type.
- `csv.QUOTE_MINIMAL` - Quote fields with special characters
- `csv.QUOTE_NONNUMERIC` - Quote all fields that aren't numbers value
- `csv.QUOTE_NONE` – Don't quote anything in output



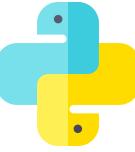
How to Read a CSV File

- To read data from CSV files, you must use the reader function to generate a reader object.
- The reader function is developed to take each row of the file and make a list of all columns. Then, you have to choose the column you want the variable data for.
- It sounds a lot more intricate than it is. Let's take a look at this example, and we will find out that working with csv file isn't so hard.



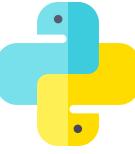
Read the File in as List of Lists

```
#import necessary modules  
import csv  
with open('X:\data.csv','rt')as f:  
    data = csv.reader(f)  
    for row in data:  
        print(row)
```



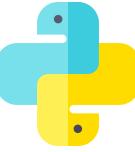
Output

```
[ 'Programming language; Designed by; Appeared; Extension' ]  
[ 'Python; Guido van Rossum; 1991; .py' ]  
[ 'Java; James Gosling; 1995; .java' ]  
[ 'C++; Bjarne Stroustrup; 1983; .cpp' ]
```



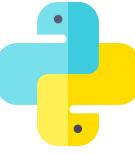
How to Read a CSV as a Dictionary

- You can also use **DictReader** to read CSV files. The results are interpreted as a dictionary where the header row is the key, and other rows are values.



Read CSV File into List of Dictionaries

```
#import necessary modules  
import csv  
  
reader = csv.DictReader(open("data.csv"))  
for raw in reader:  
    print(raw)
```

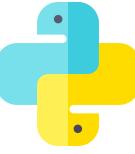


Output

```
OrderedDict([('Programming language', 'Python'), ('Designed by', 'Guido van Rossum'), ('Appeared', '1991'), ('Extension', '.py')])
```

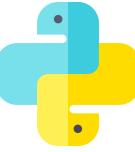
```
OrderedDict([('Programming language', 'Java'), ('Designed by', 'James Gosling'), ('Appeared', '1995'), ('Extension', '.java')])
```

```
OrderedDict([('Programming language', 'C++'), ('Designed by', 'Bjarne Stroustrup'), ('Appeared', '1985'), ('Extension', '.cpp')])
```



How to write CSV File

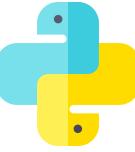
- When you have a set of data that you would like to store in a CSV file you have to use writer() function. To iterate the data over the rows(lines), you have to use the writerow() function.



Write List of Lists as CSV File

```
#import necessary modules
import csv
with open('X:\writeData.csv', mode='w') as file:
    writer = csv.writer(file, delimiter=',', quotechar='"', \
                        quoting=csv.QUOTE_MINIMAL)

#way to write to csv file
writer.writerow(['Programming language', 'Designed by', \
                 'Appeared', 'Extension'])
writer.writerow(['Python', 'Guido van Rossum', '1991', '.py'])
writer.writerow(['Java', 'James Gosling', '1995', '.java'])
writer.writerow(['C++', 'Bjarne Stroustrup', '1985', '.cpp'])
```



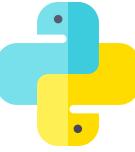
Output File: WriteData.csv

Programming language, Designed by, Appeared, Extension

Python, Guido van Rossum, 1991, .py

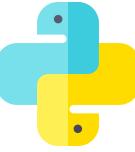
Java, James Gosling, 1995, .java

C++, Bjarne Stroustrup, 1983, .cpp



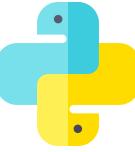
Writing to CSV Files with Pandas

- Writing to CSV file with Pandas is as easy as reading. Here you can convince in it. First you must create DataFrame based on the following code.



Write A Pandas Data Frame to a CSV File

```
from pandas import DataFrame  
  
C = {'Programming language': ['Python', 'Java', 'C++'],  
      'Designed by': ['Guido van Rossum', 'James Gosling', 'Bjarne Stroustrup'],  
      'Appeared': ['1991', '1995', '1985'],  
      'Extension': ['.py', '.java', '.cpp'],  
      }  
  
df = DataFrame(C, columns= ['Programming language', 'Designed by', \  
                           'Appeared', 'Extension'])  
  
export_csv = df.to_csv (r'pandaresult.csv', index = None, header=True)  
  
# here you have to write path, where result file will be stored  
  
print (df)
```



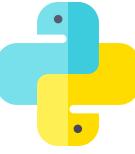
Output File:

Programming language, Designed by, Appeared, Extension

0 Python, Guido van Rossum, 1991, .py

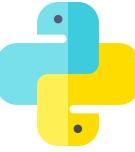
1 Java, James Gosling, 1995, .java

2 C++, Bjarne Stroustrup, 1983, .cpp



Summary

- So, now you know how to use method 'csv' and also read and write data in CSV format. CSV files are widely used in software applications because they are easy to read and manage, and their small size makes them relatively fast for processing and transmission.
- The csv module provides various functions and classes which allow you to read and write easily. You can look at the official Python documentation and find some more interesting tips and modules. CSV is the best way for saving, viewing, and sending data. Actually, it isn't so hard to learn as it seems at the beginning. But with a little practice, you'll master it.



Summary

- Pandas is a great alternative to read CSV files.
- Also, there are other ways to parse text files with libraries like ANTLR, PLY, and PlyPlus. They can all handle heavy-duty parsing, and if simple **String** manipulation doesn't work, there are regular expressions which you can use.



Design of Charts

LECTURE 4



Visualizing Data



Edward Tufte

Wrote ***The Visual Display of Quantitative Information***

- book is concerned with the design of statistical graphs
 - using points, lines, numbers, words, shading, color to present quantitative information
- **ink** = everything that is on the paper
- In Excel, ink = data, labels and chart effects

Data Graphics Principles



Above all else, show the **data**



Maximize the **data-ink** ratio, within reason



Erase non-data-ink, within reason



Erase redundant data-ink, within reason



Revise and edit

Figure 3.2: An example of “chart junk”

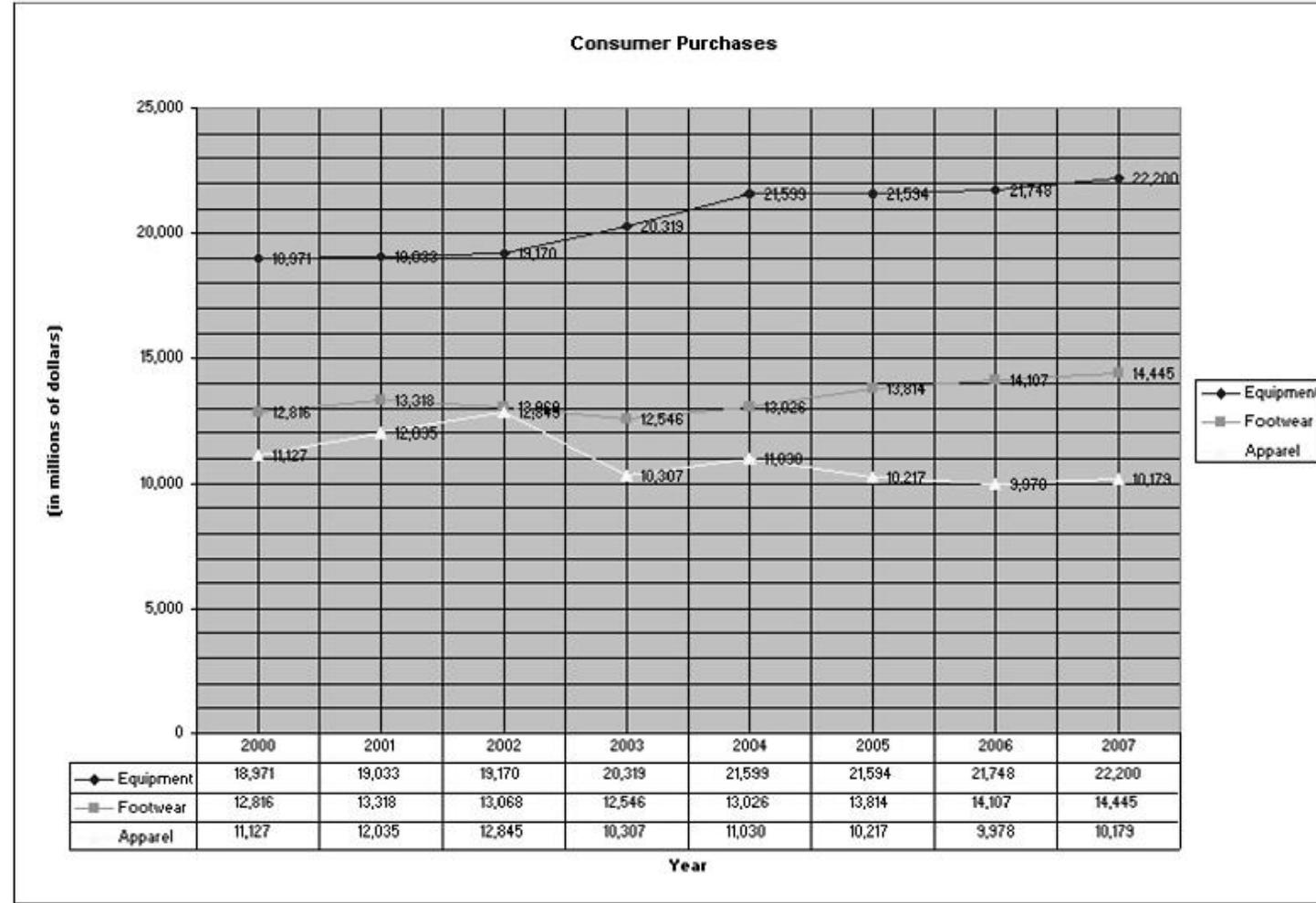


Chart Junk

Data Graphics Principles

“Above all else show the data”

- Don’t clutter a chart by adding unneeded illustration or decoration.

“Maximize the data-ink ratio”

- Refers to the portion of ink that is devoted to displaying the data vs. the portion of graphic that can be removed without losing the data.

Data Graphics Principles

“Erase non-data-ink”

- Non-data-ink is a part of the chart that decorates more than informs.

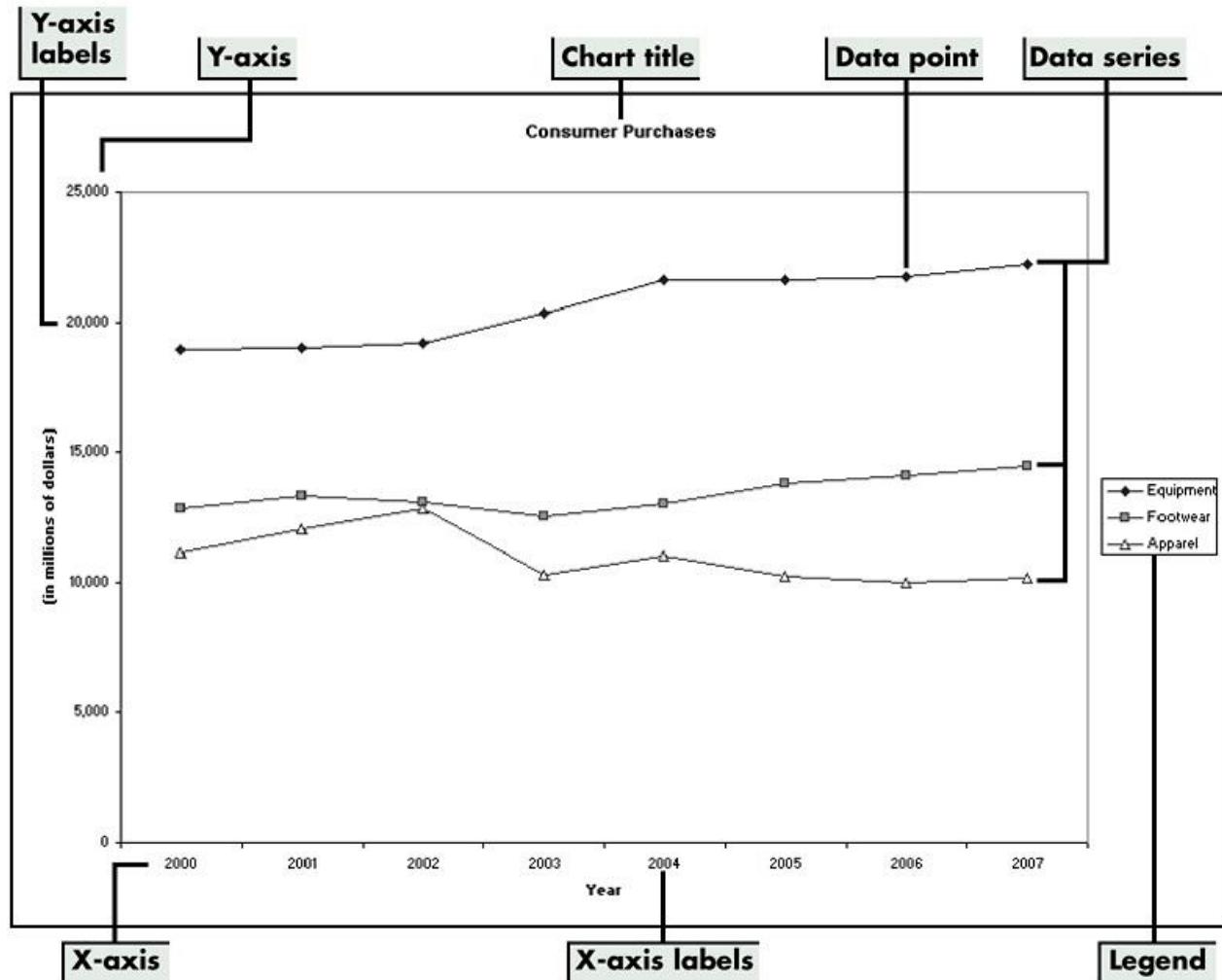
“Erase redundant data ink”

- Redundant data ink is ink that repeats information.

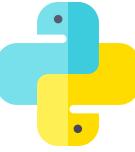
“Revise and edit”

- Revise and edit charts like you would a piece of writing.

Figure 3.3: “Chart junk” removed



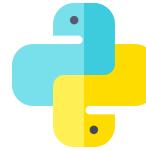
No Chart
Junk



Charts in Excel

Use Chart Wizard to:

- Select chart type
- Confirm source data
- Modify chart options
- Select location



Excel Chart Types

Table 3.1: Excel chart types

| Chart Type | Icon | Description |
|------------|------|---|
| Column | | Compares values across categories in a vertical orientation. Values are indicated by the height of the columns. |
| Bar | | Compares values across categories in a horizontal orientation. Values are indicated by the length of the bars. |
| Line | | Displays trends over time or by category. Values are indicated by the height of the lines. |
| Area | | Displays trends over time or by category. Values are indicated by the filled areas below the lines. |

Table 3.1: Excel chart types (cont.)

| Chart Type | Icon | Description |
|-------------------------|--|--|
| Pie |  | Compares the contribution each value in a single numeric data series makes to the whole, or 100%. Values are indicated by the size of the pie slices. |
| Doughnut |  | Compares the contribution each value in multiple numeric data series makes to the whole, or 100%. Values are indicated by the size of the doughnut segments. |
| XY (Scatter) |  | Compares pairs of numeric values on the X- and Y-axes, with the data points plotted proportionally to the values on the X-axis; can also be used to display a functional relationship, such as $y=mx+b$. Values are indicated by the position of the data points. |
| Stock |  | Displays stock price and volume trends over time. Plotted values can include volume, opening price, highest price, lowest price, and closing price. |
| Radar |  | Compares values across categories in a circular orientation. Values are indicated by the distance from a center point. |
| Bubble |  | Compares sets of three values. Values are indicated by the size of the bubbles (filled circles). |
| Surface |  | Displays value trends in three dimensions. Values are indicated by areas with colors or patterns on the surface of the chart. |
| Cylinder, Cone, Pyramid |  ,  and  | Compare values across categories, similar to column and bar charts, using 3-D shapes of cylinders, cones, or pyramids in the chart (respectively) in place of columns or bars. Values are indicated by the size of the cylinders, cones, or pyramids. |

Excel Chart Types



Line chart- displays trends over time or by category.



Column chart- compares values across categories in a vertical orientation.

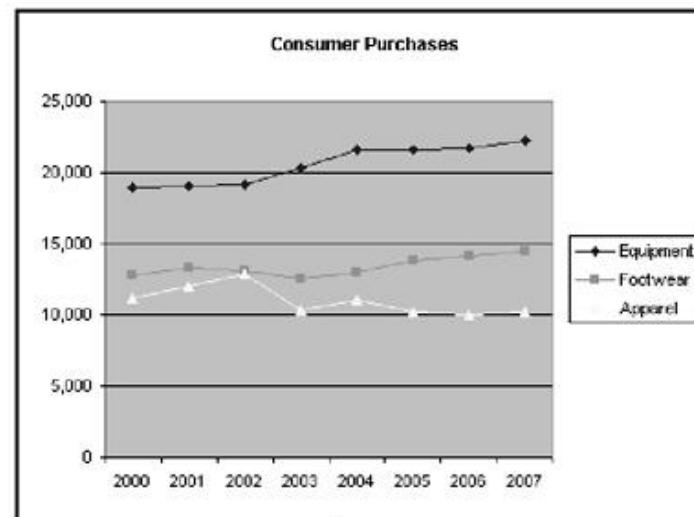
Line vs. Column Charts

Line vs Column

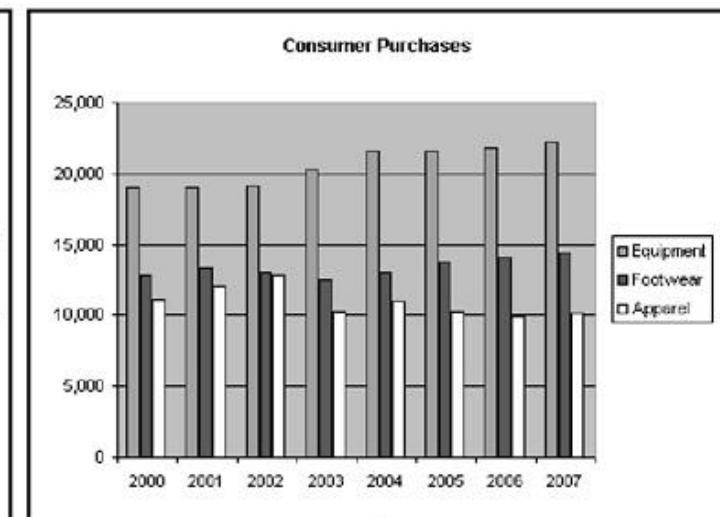
Line Chart shows the trends

Column Chart shows the comparison of different data groups.

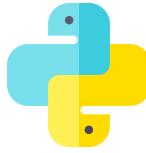
Figure 3.5: Line chart versus column chart



Line chart emphasizes the trend in each category over time



Column chart emphasizes the contribution that each category made each year



Line vs. Scatter Charts

XY (Scatter) charts –

- plot numeric values on *both* the X- and Y- axes based on the value of the data.
- **Line Chart** plots numeric values on one axis and category labels equidistantly on the other axis.

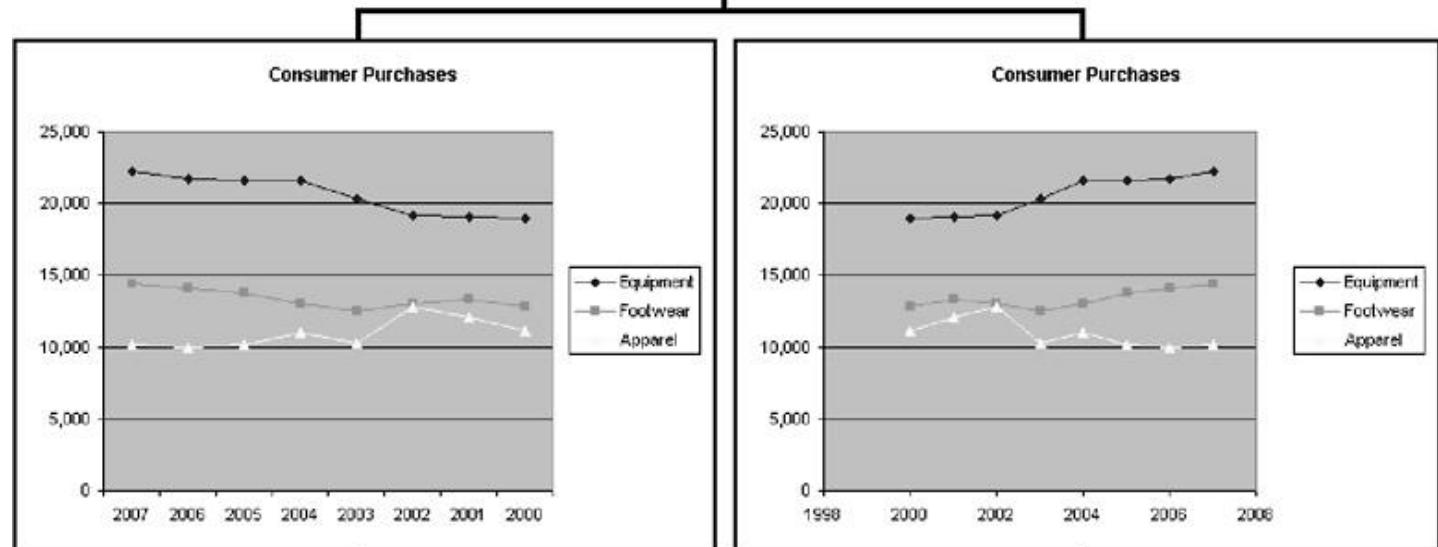
Line vs. Scatter Plot

Scatter Data Points or Markers

Line shows trends better

Figure 3.6: Line chart versus XY (Scatter) chart

| A | B | C | D | E | F | G | H | I | |
|---|--------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | Consumer Purchases by Category | | | | | | | | |
| 2 | Sales in millions of dollars | | | | | | | | |
| 3 | | | | | | | | | |
| 4 | | 2007 | 2006 | 2005 | 2004 | 2003 | 2002 | 2001 | 2000 |
| 5 | Equipment | 22,200 | 21,748 | 21,594 | 21,599 | 20,319 | 19,170 | 19,033 | 18,971 |
| 6 | Footwear | 14,445 | 14,107 | 13,814 | 13,026 | 12,546 | 13,068 | 13,318 | 12,816 |
| 7 | Apparel | 10,179 | 9,978 | 10,217 | 11,030 | 10,307 | 12,845 | 12,035 | 11,127 |
| 8 | Total | 46,824 | 45,835 | 45,625 | 45,655 | 43,196 | 45,083 | 44,387 | 42,914 |



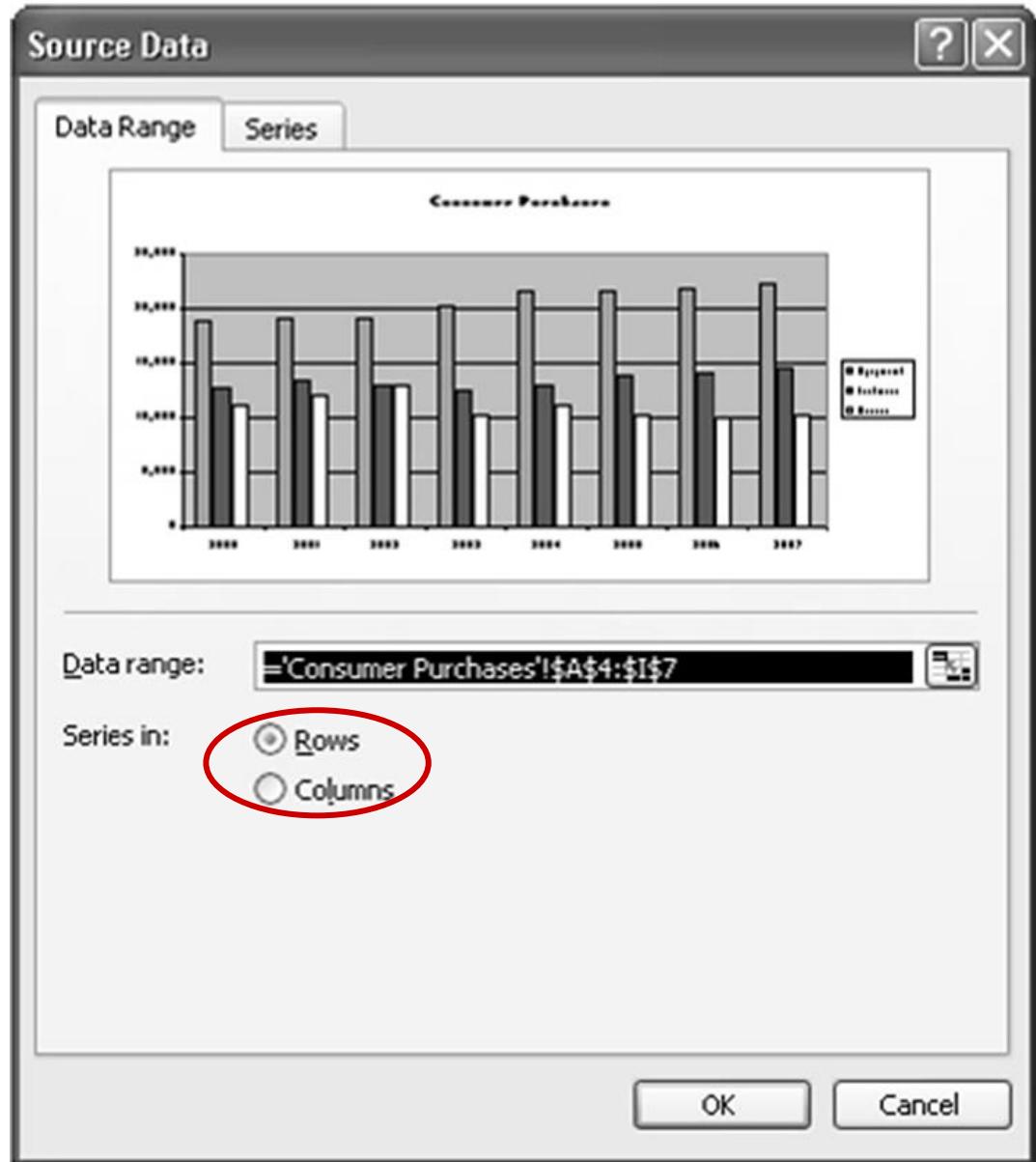
The line chart plots the X-axis based on the position of the categories in the data range

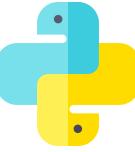
The XY (Scatter) chart plots the X-axis in numeric order based on the values in the data range

Data Source

Choose Data Series
from Row or Column?

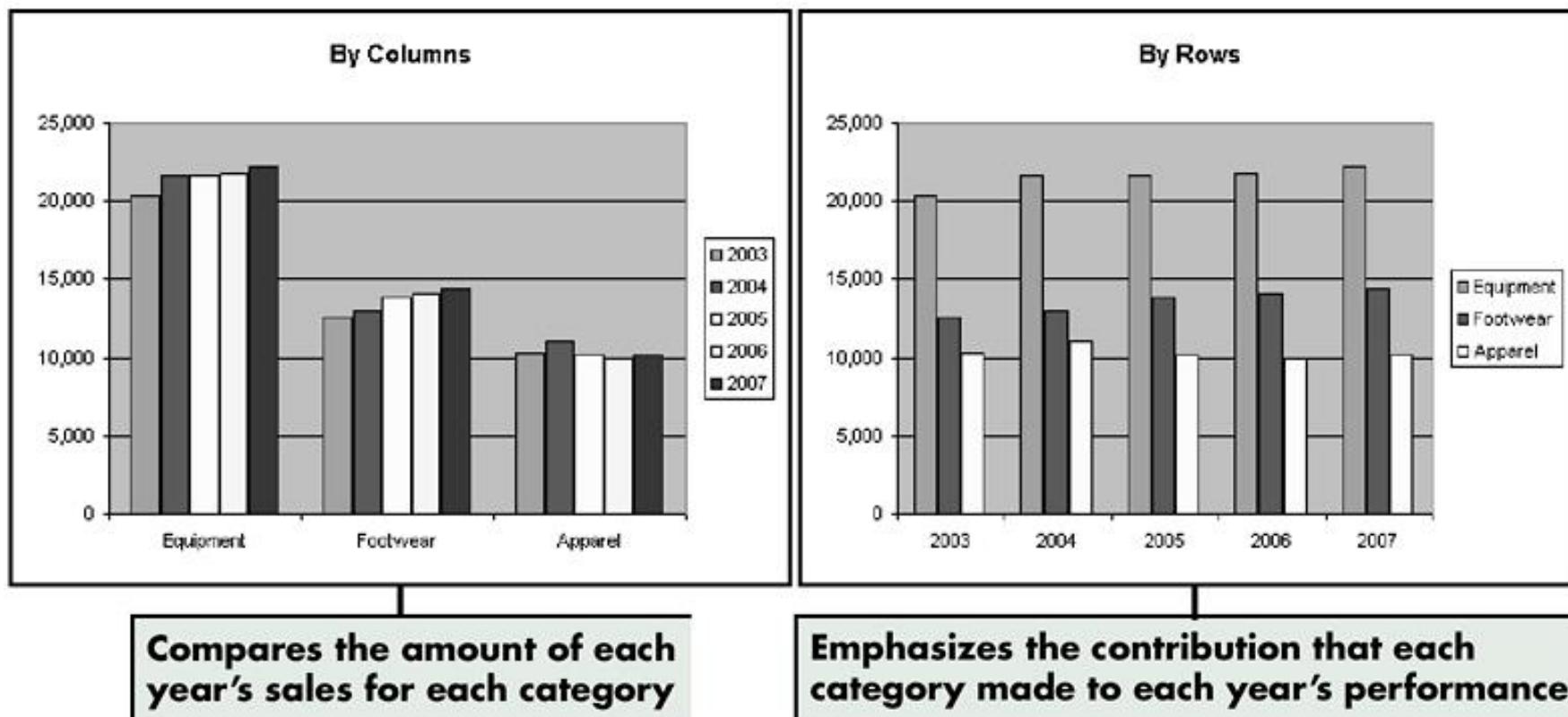
Figure 3.7: Source Data dialog box

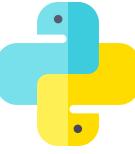




Rows vs. Columns

Figure 3.8: Displaying the data series by columns or by rows





Area vs Pie Charts

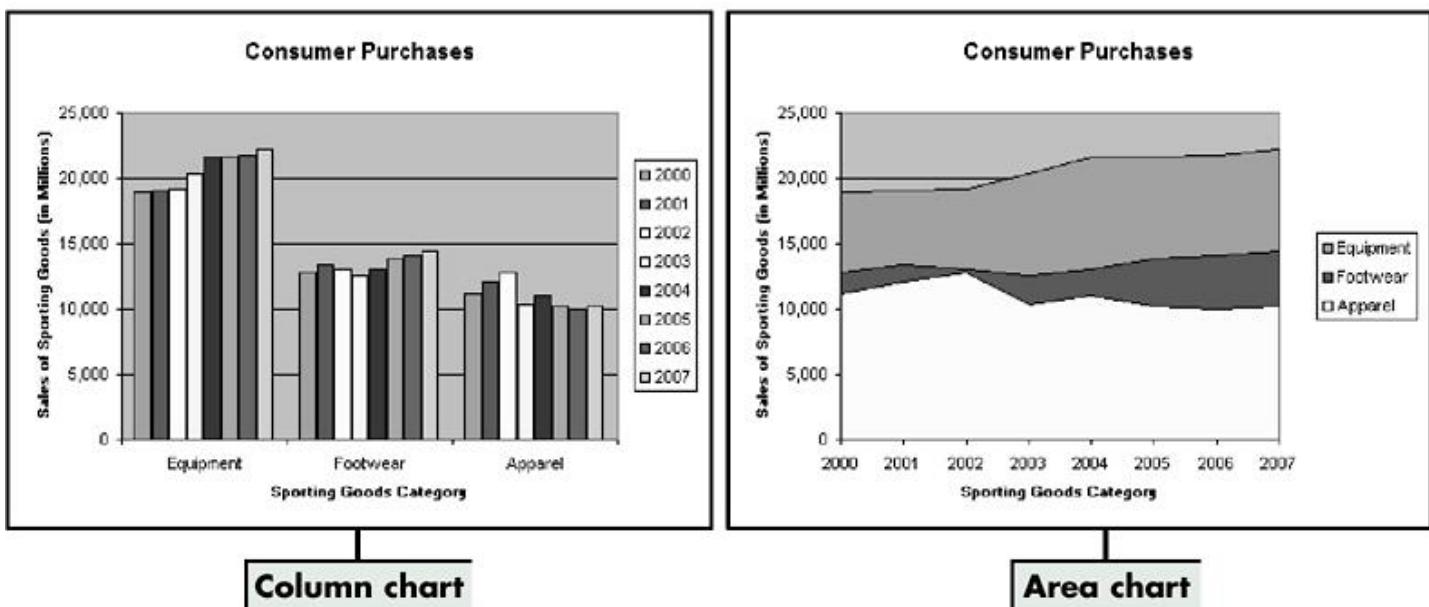
Area chart- combines the features of a line chart with a bar or column chart by filling in the area below the line, and displaying the trend values over time or categories.

Pie chart- displays the percentage contribution that each category makes to a whole or 100%.

Column and Area Charts

When Column Chart goes too busy, Area Chart may be better.

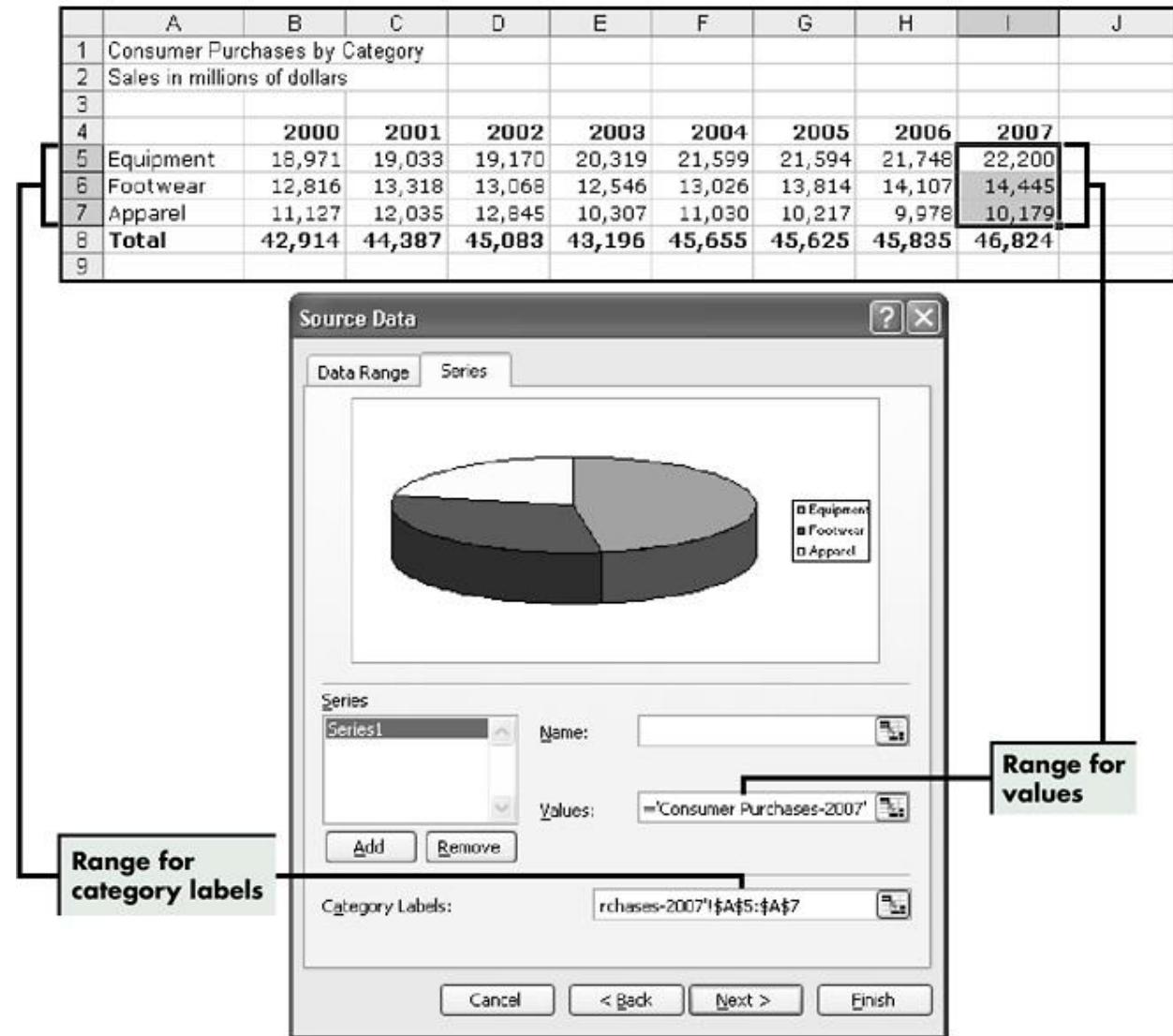
Figure 3.11: Column chart versus area chart



Pie Charts

How to Create Pie Charts?

Figure 3.12: Selecting the pie chart source data



Pie Charts

Modified or not Modified?

Emphasis

Borders

Legends

Labels

Figure 3.13: Default pie chart and modified pie chart

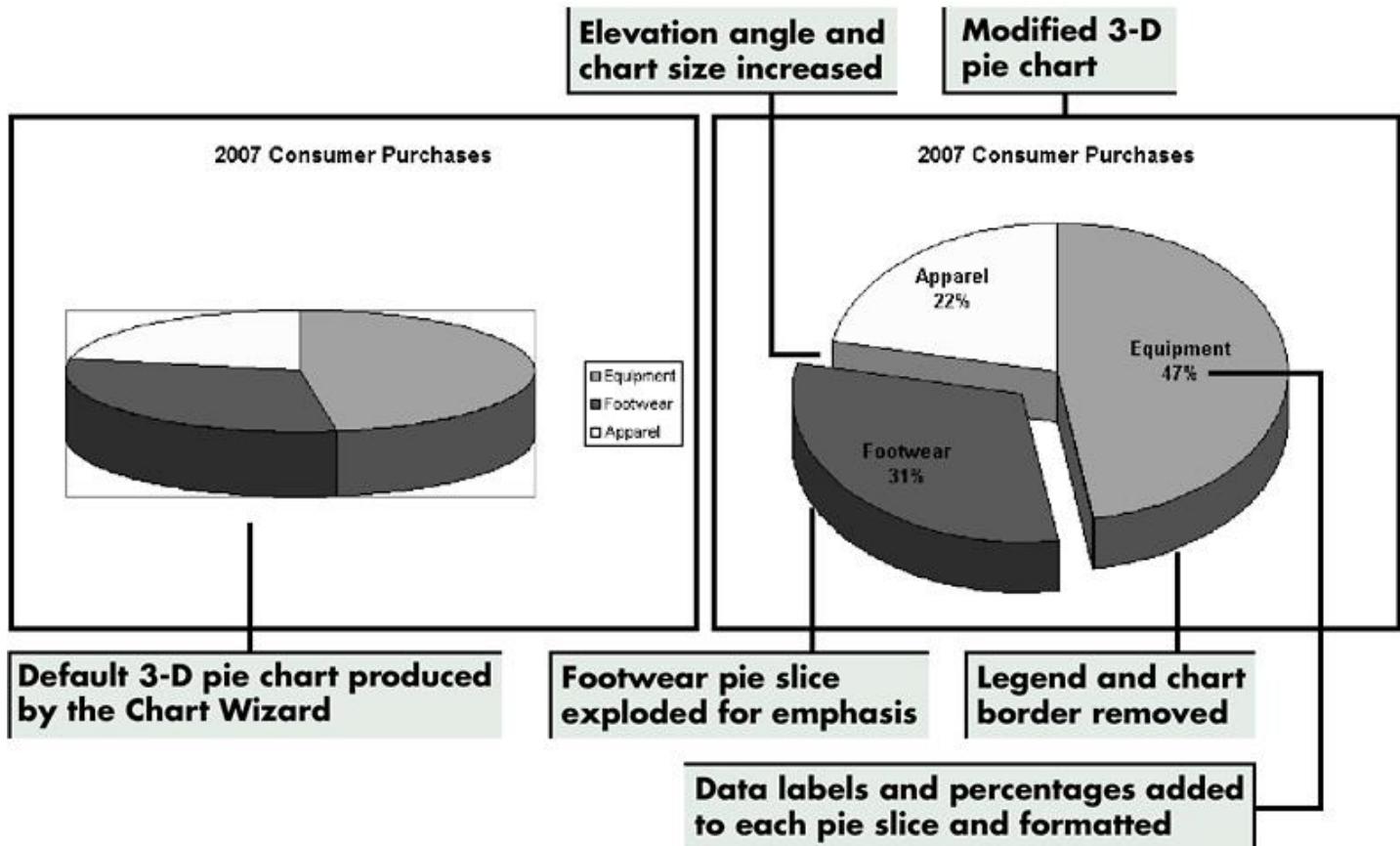


Figure 3.9: Selecting the right number of chart options

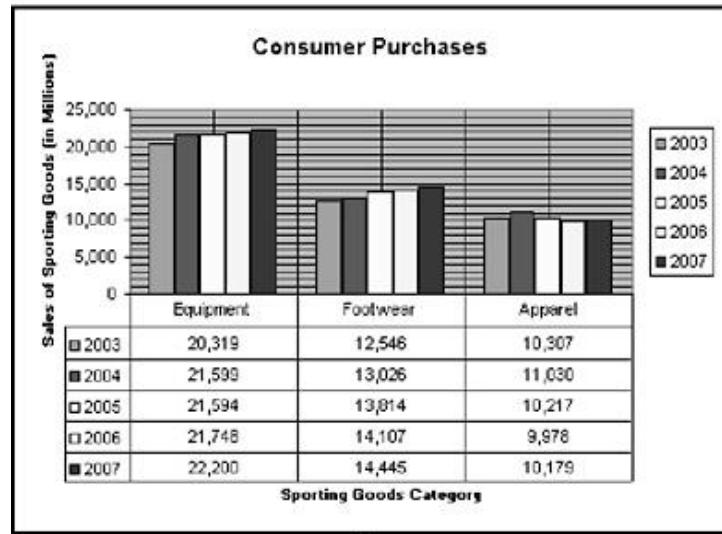
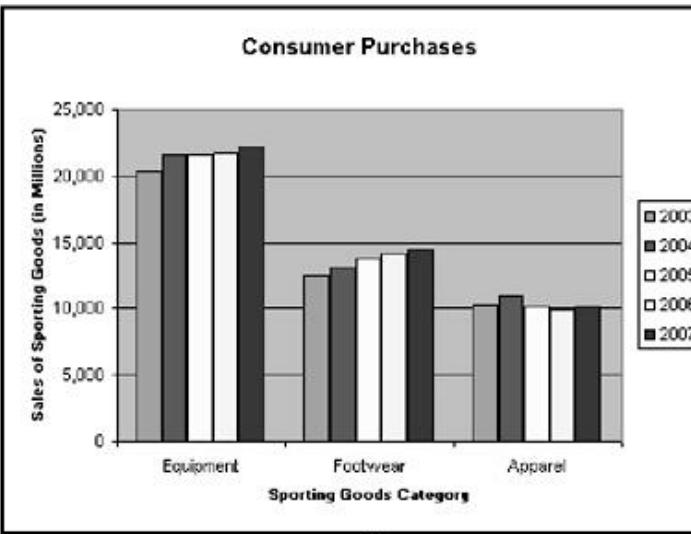
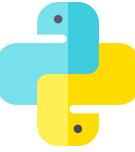


Chart with too many options



Final chart

Don't over-do
it with Chart
Options



Summary

- Tufte → “Above all else, show the data...” and other guidelines
- Different Chart Types in Excel
- Don’t overdo it with chart options – See bullet 1!!



Advanced Charts

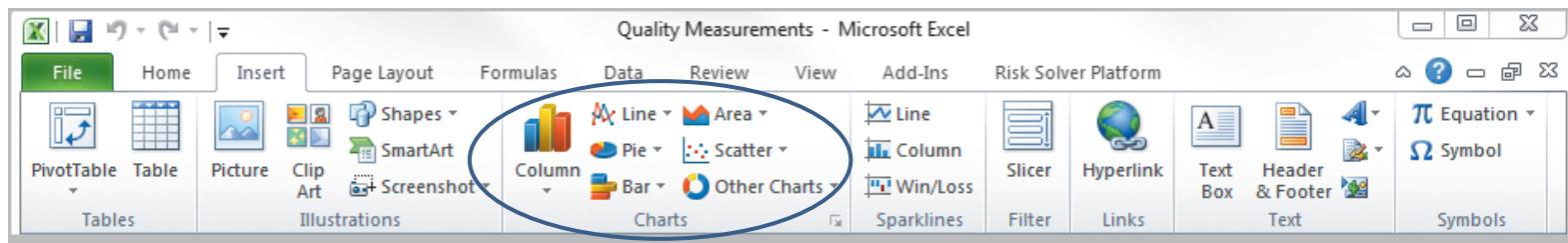
Business Analytics

LECTURE 5

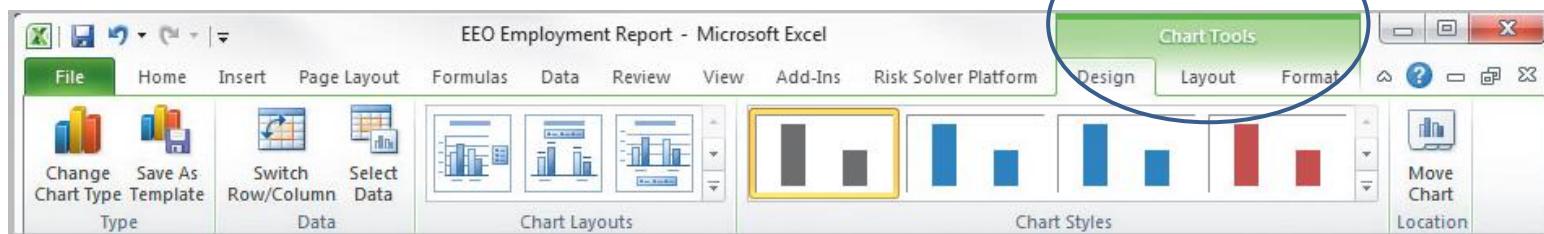
Data Visualization

Creating Charts in Microsoft Excel

- Select the *insert* tab.
- Highlight the data.
- Click on chart type, then subtype.



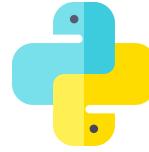
- Use *chart tools* to customize.





Lab

CREATE COLUMN CHARTS

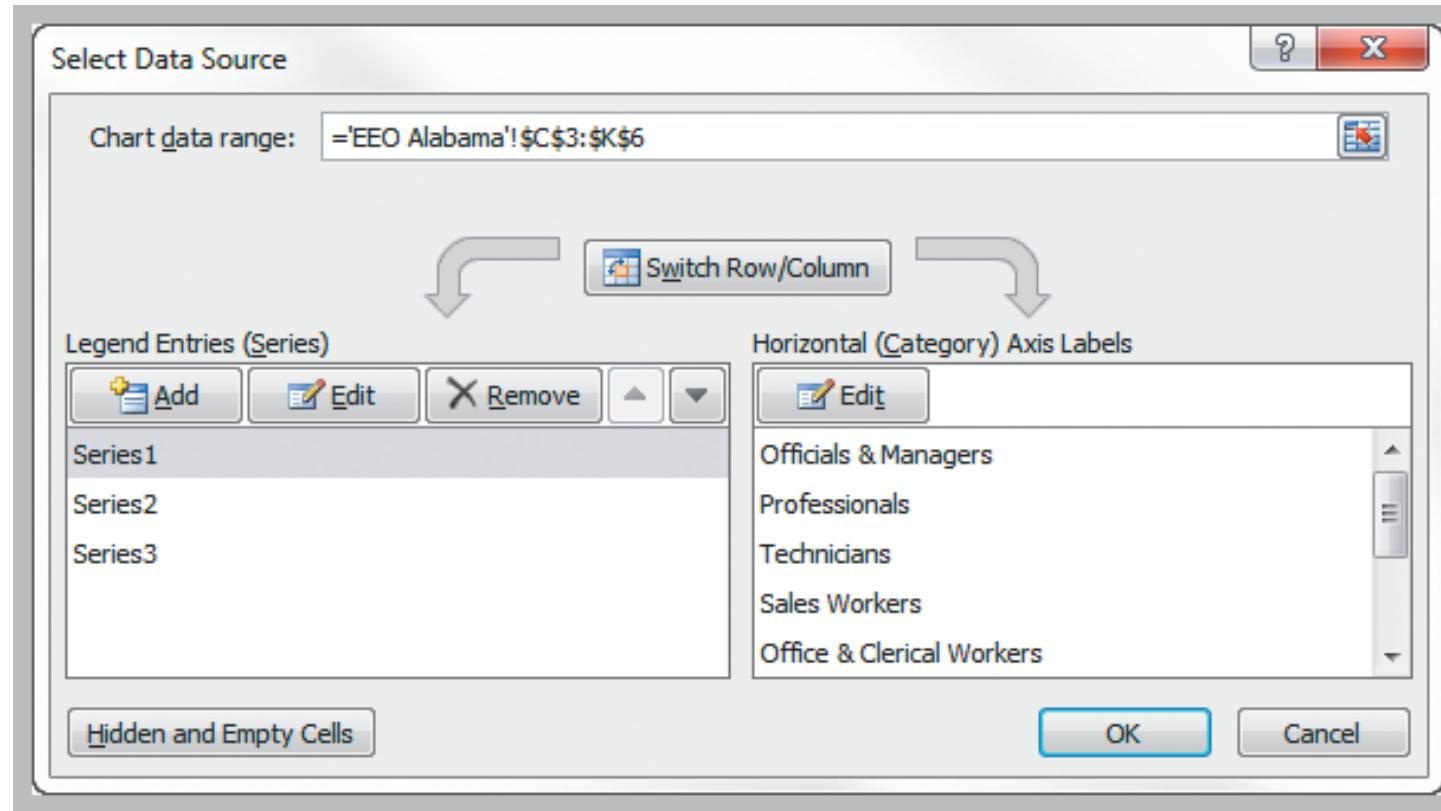


Creating a Column Chart

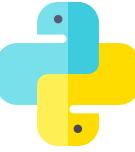
Highlighted Cells

| Equal Employment Opportunity Commission Report - Number Employed in State of Alabama, 2006 | | | | | | | | | | |
|--|------------------|----------------------|---------------|-------------|---------------|---------------------------|---------------|------------|----------|-----------------|
| Racial/Ethnic Group and Gender | Total Employment | Officials & Managers | Professionals | Technicians | Sales Workers | Office & Clerical Workers | Craft Workers | Operatives | Laborers | Service Workers |
| ALL EMPLOYEES | 632,329 | 60,258 | 80,733 | 39,868 | 62,019 | 67,014 | 61,322 | 120,810 | 68,752 | 71,553 |
| Men | 349,353 | 41,777 | 39,792 | 19,848 | 23,727 | 11,293 | 55,853 | 84,724 | 44,736 | 27,603 |
| Women | 282,976 | 18,481 | 40,941 | 20,020 | 38,292 | 55,721 | 5,469 | 36,086 | 24,016 | 43,950 |
| WHITE | 407,545 | 51,252 | 67,622 | 28,830 | 41,091 | 44,565 | 45,742 | 67,555 | 26,712 | 34,176 |
| Men | 237,516 | 36,536 | 34,842 | 16,004 | 17,756 | 7,656 | 42,699 | 50,537 | 17,802 | 13,684 |
| Women | 170,029 | 14,716 | 32,780 | 12,826 | 23,335 | 36,909 | 3,043 | 17,018 | 8,910 | 20,492 |
| MINORITY | 224,784 | 9,006 | 13,111 | 11,038 | 20,928 | 22,449 | 15,580 | 53,255 | 42,040 | 37,377 |
| Men | 111,837 | 5,241 | 4,950 | 3,844 | 5,971 | 3,637 | 13,154 | 34,187 | 26,934 | 13,919 |
| Women | 112,947 | 3,765 | 8,161 | 7,194 | 14,957 | 18,812 | 2,426 | 19,068 | 15,106 | 23,458 |
| BLACK | 187,419 | 7,516 | 9,914 | 9,974 | 19,389 | 21,107 | 12,232 | 45,709 | 27,459 | 34,119 |
| Men | 86,332 | 4,146 | 2,904 | 3,148 | 5,370 | 3,312 | 10,077 | 28,610 | 16,881 | 11,884 |
| Women | 101,087 | 3,370 | 7,010 | 6,826 | 14,019 | 17,795 | 2,155 | 17,099 | 10,578 | 22,235 |
| HISPANIC | 29,181 | 652 | 920 | 474 | 788 | 802 | 2,684 | 6,316 | 13,956 | 2,589 |
| Men | 20,777 | 486 | 614 | 326 | 334 | 206 | 2,496 | 4,856 | 9,676 | 1,783 |
| Women | 8,404 | 166 | 306 | 148 | 454 | 596 | 188 | 1,460 | 4,280 | 806 |
| ASIAN AMERICAN | 5,563 | 576 | 1,903 | 423 | 502 | 364 | 298 | 698 | 329 | 470 |
| Men | 3,091 | 419 | 1,214 | 265 | 181 | 84 | 254 | 349 | 152 | 173 |
| Women | 2,472 | 157 | 689 | 158 | 321 | 280 | 44 | 349 | 177 | 297 |

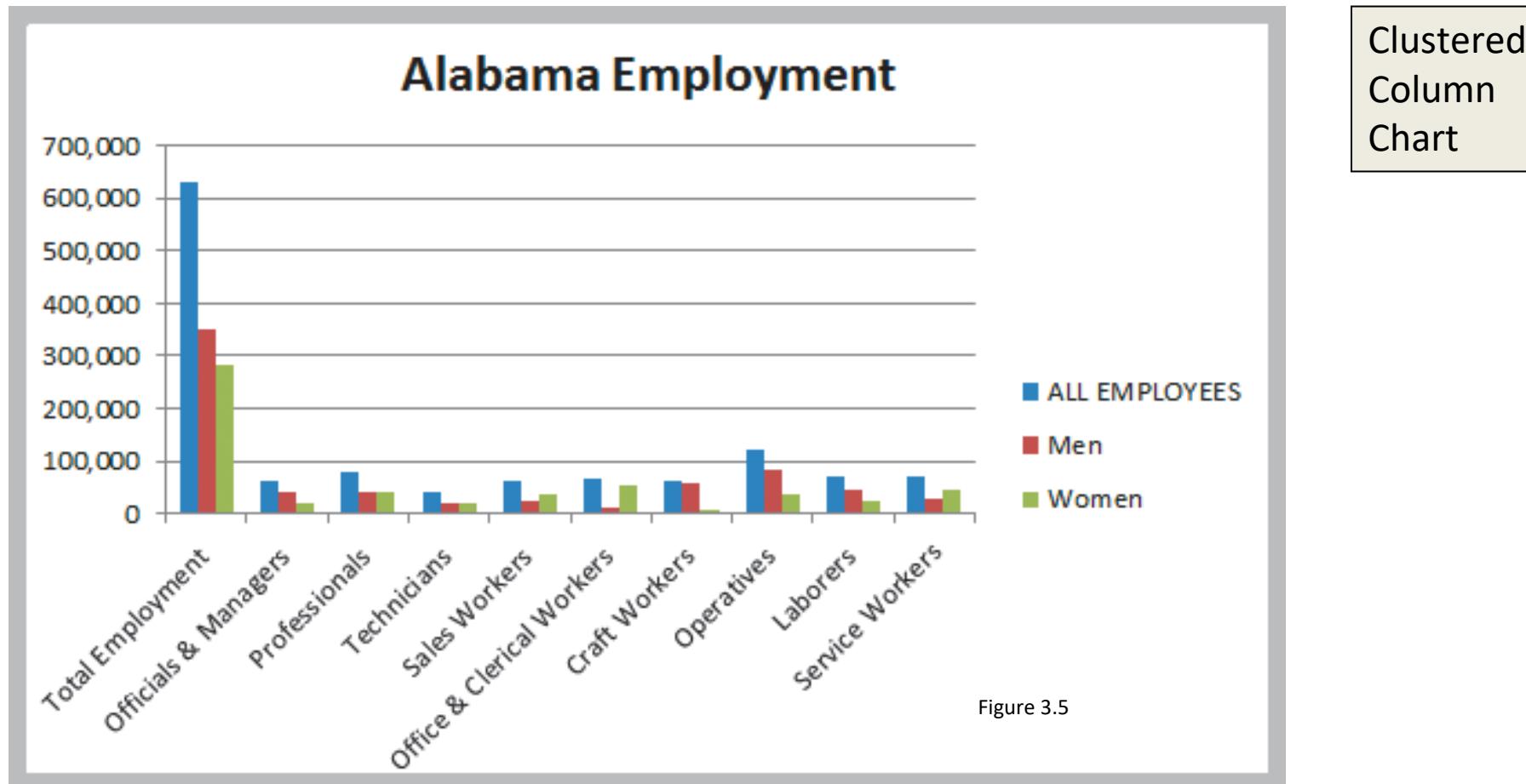
Creating a Column Chart

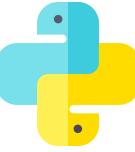


- Choose column chart (clustered or stacked).
- Add chart title (Alabama Employment).
- Rename Series1, Series2, and Series3
 - (ALL EMPLOYEES, Men, Women).

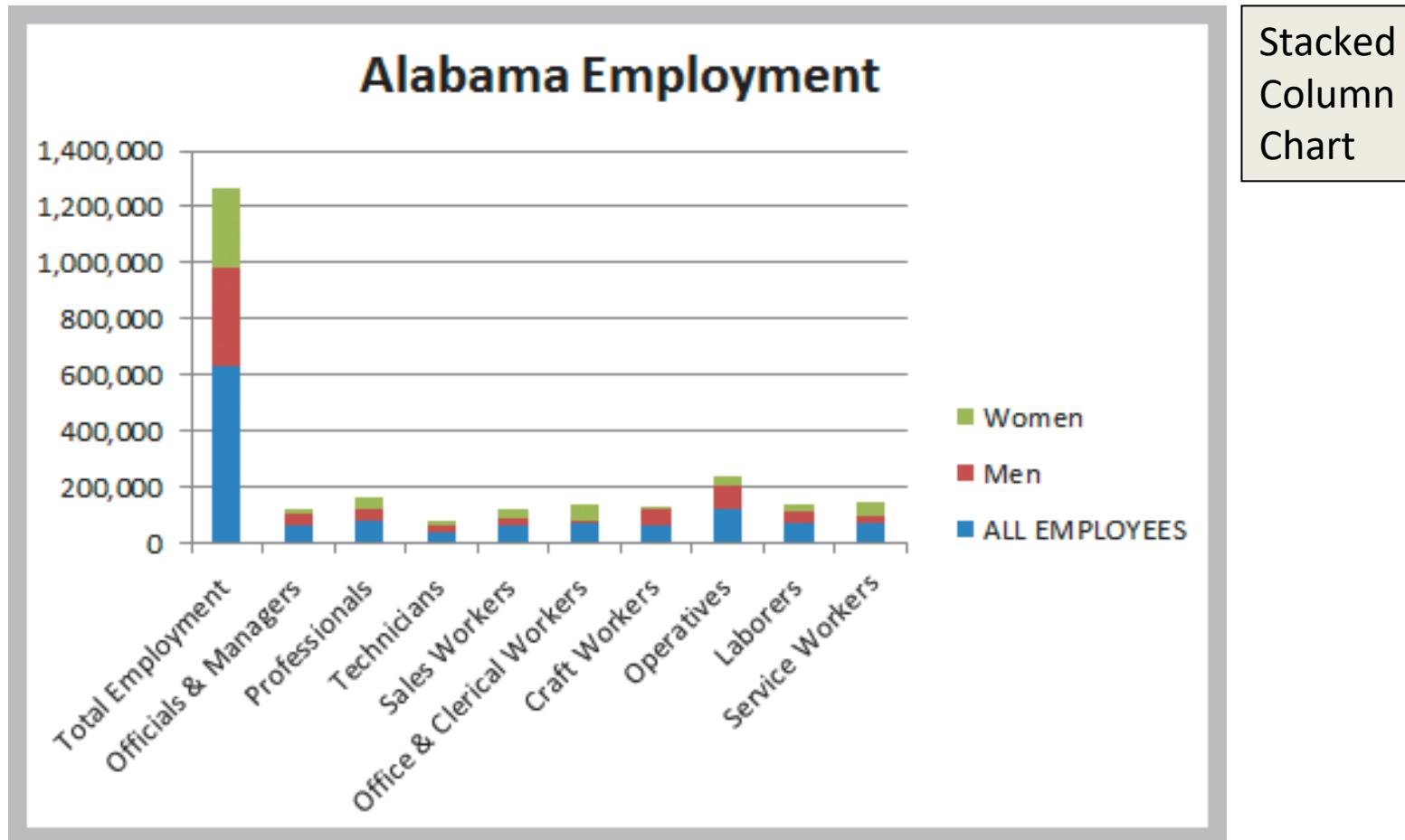


Creating a Column Chart

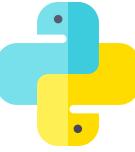




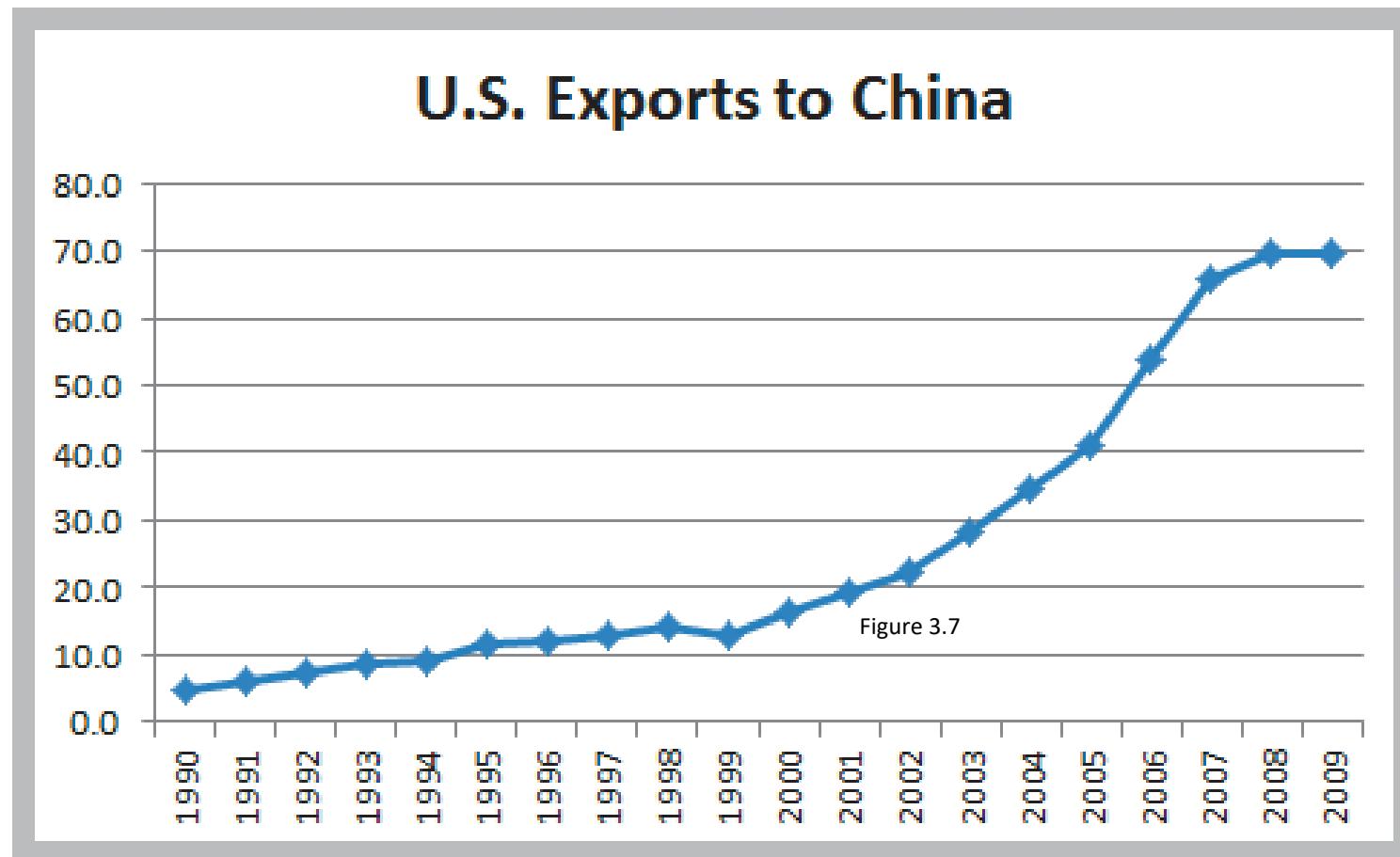
Creating a Column Chart

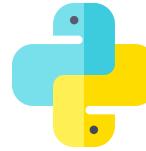


Stacked
Column
Chart



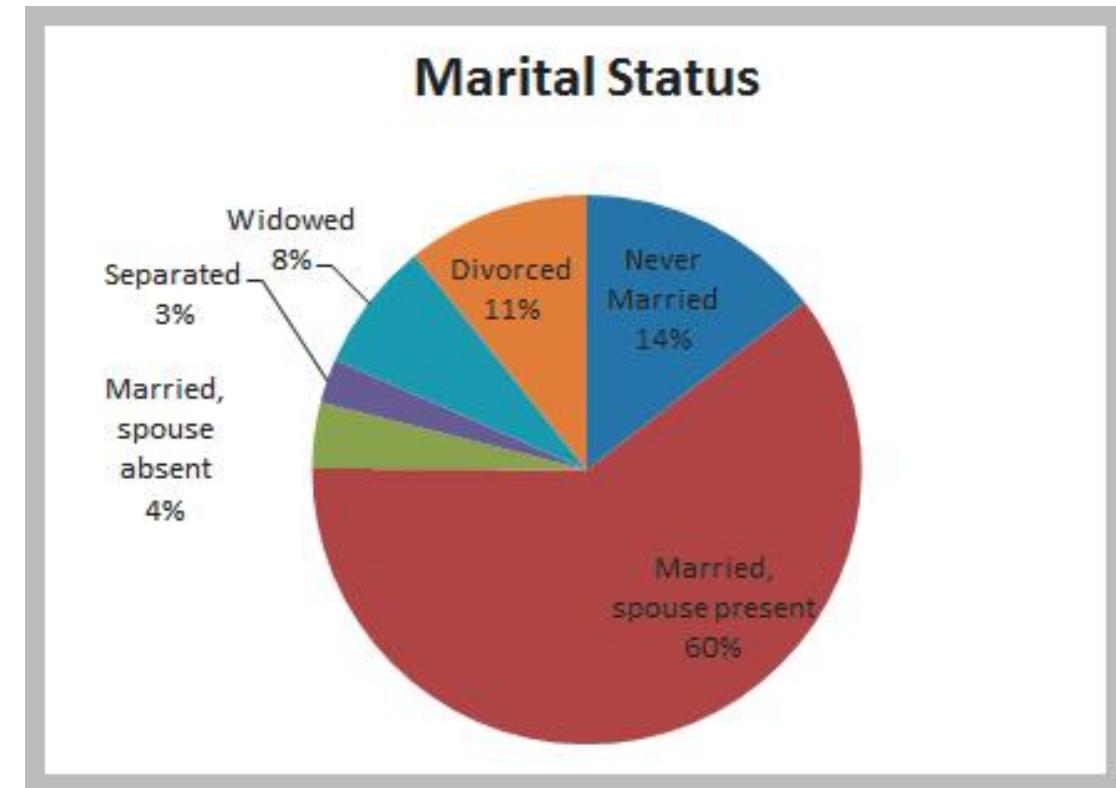
Line Chart for U.S. Exports to China

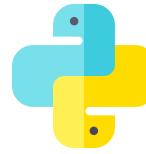




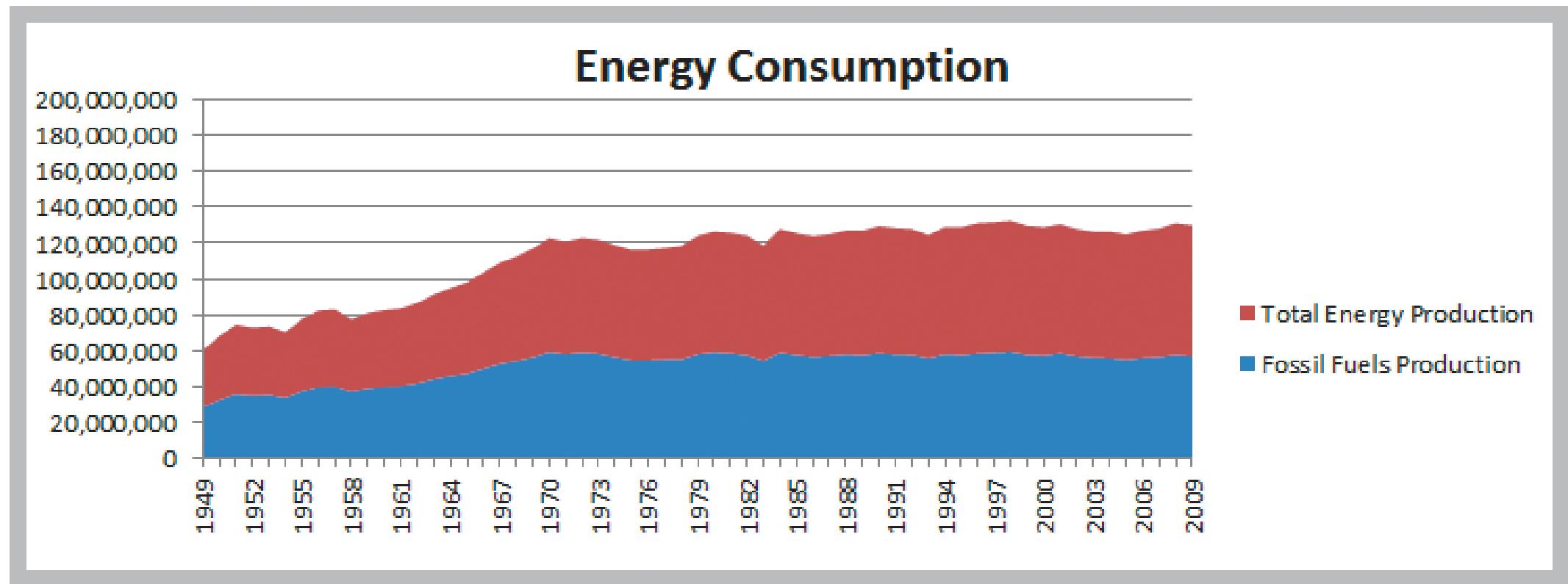
Pie Chart for Census Data

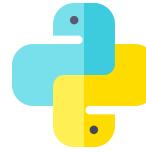
| | A | B |
|----|-------------------------|-------------|
| 18 | Marital Status | |
| 19 | Never Married | 25,752,000 |
| 20 | Married, spouse present | 107,008,000 |
| 21 | Married, spouse absent | 6,844,000 |
| 22 | Separated | 4,605,000 |
| 23 | Widowed | 13,577,000 |
| 24 | Divorced | 19,030,000 |



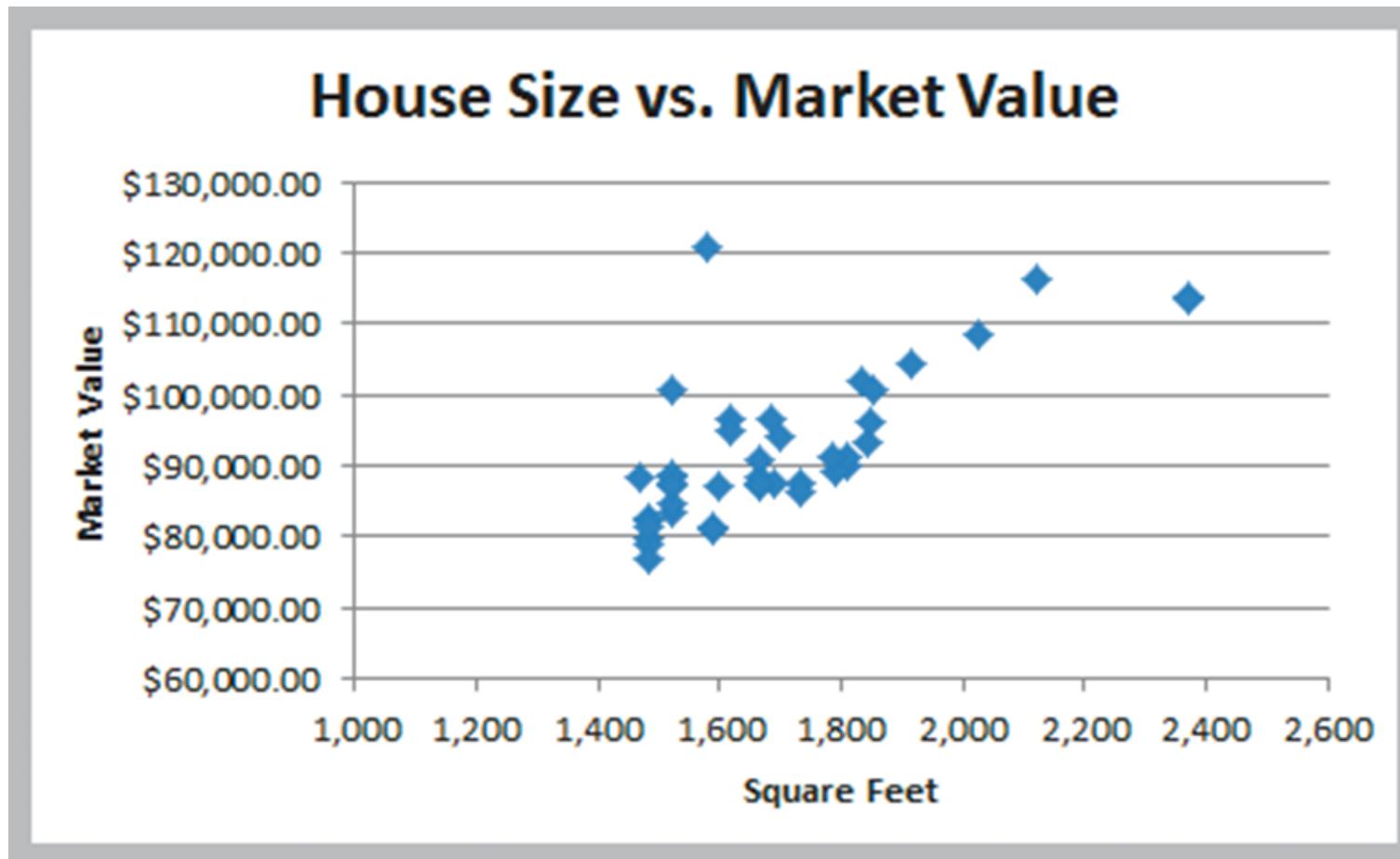


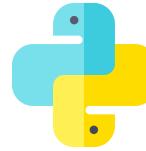
Area Chart for Energy Consumption



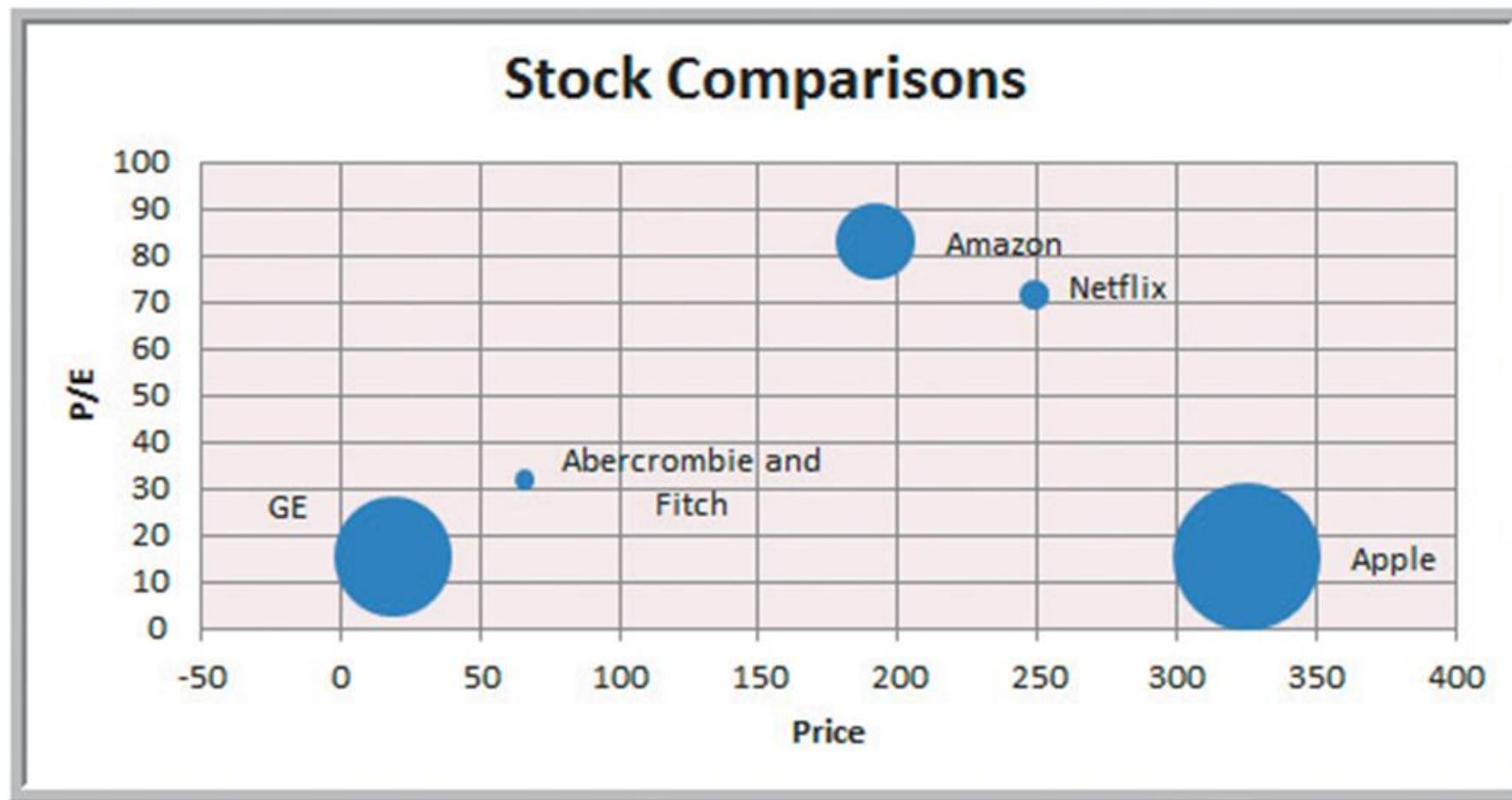


Scatter Chart for Real Estate Data





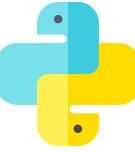
Bubble Chart for Comparing Stock Characteristics





Miscellaneous Excel Charts

ACTIVITY



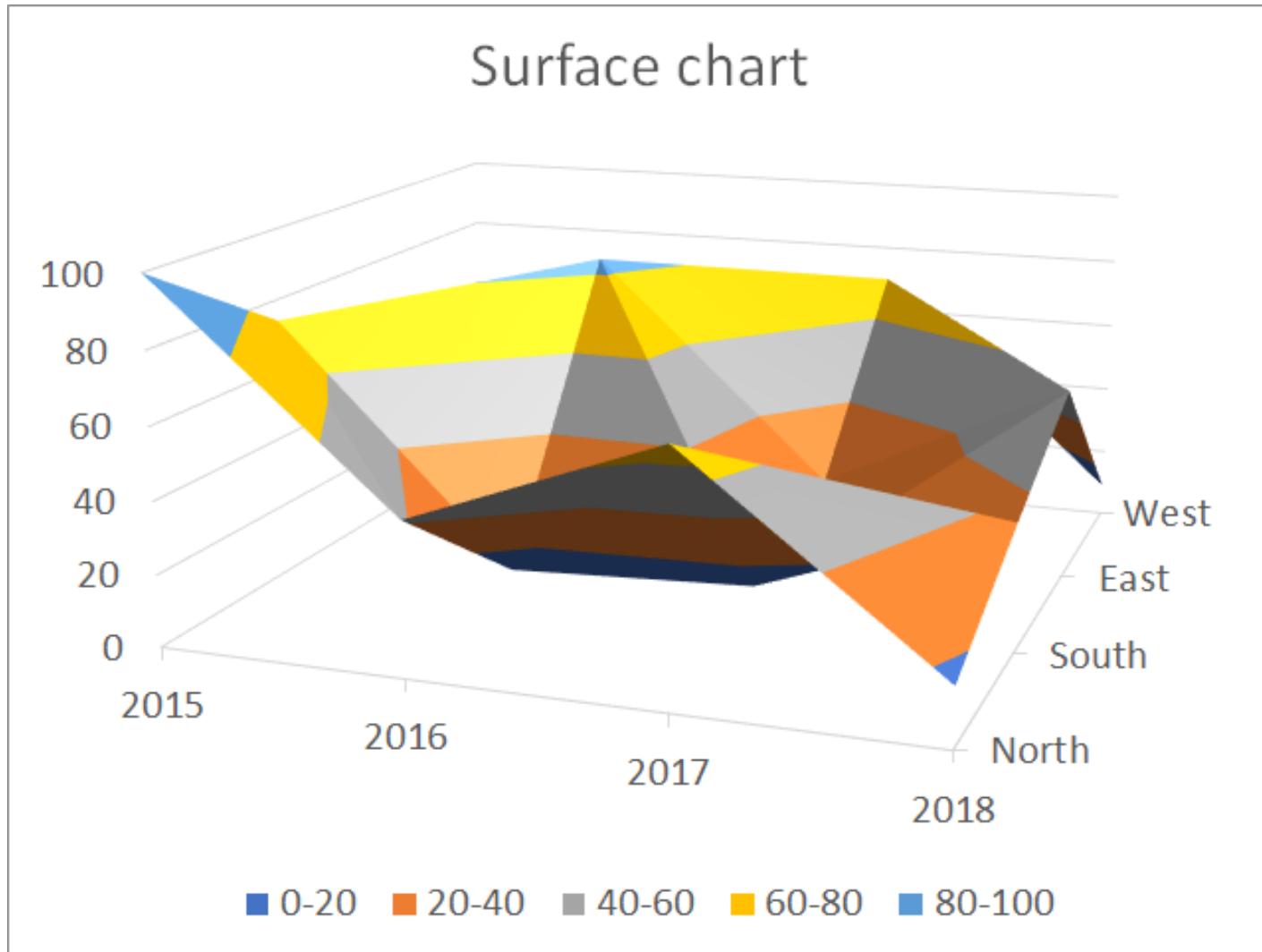
Miscellaneous Excel Charts

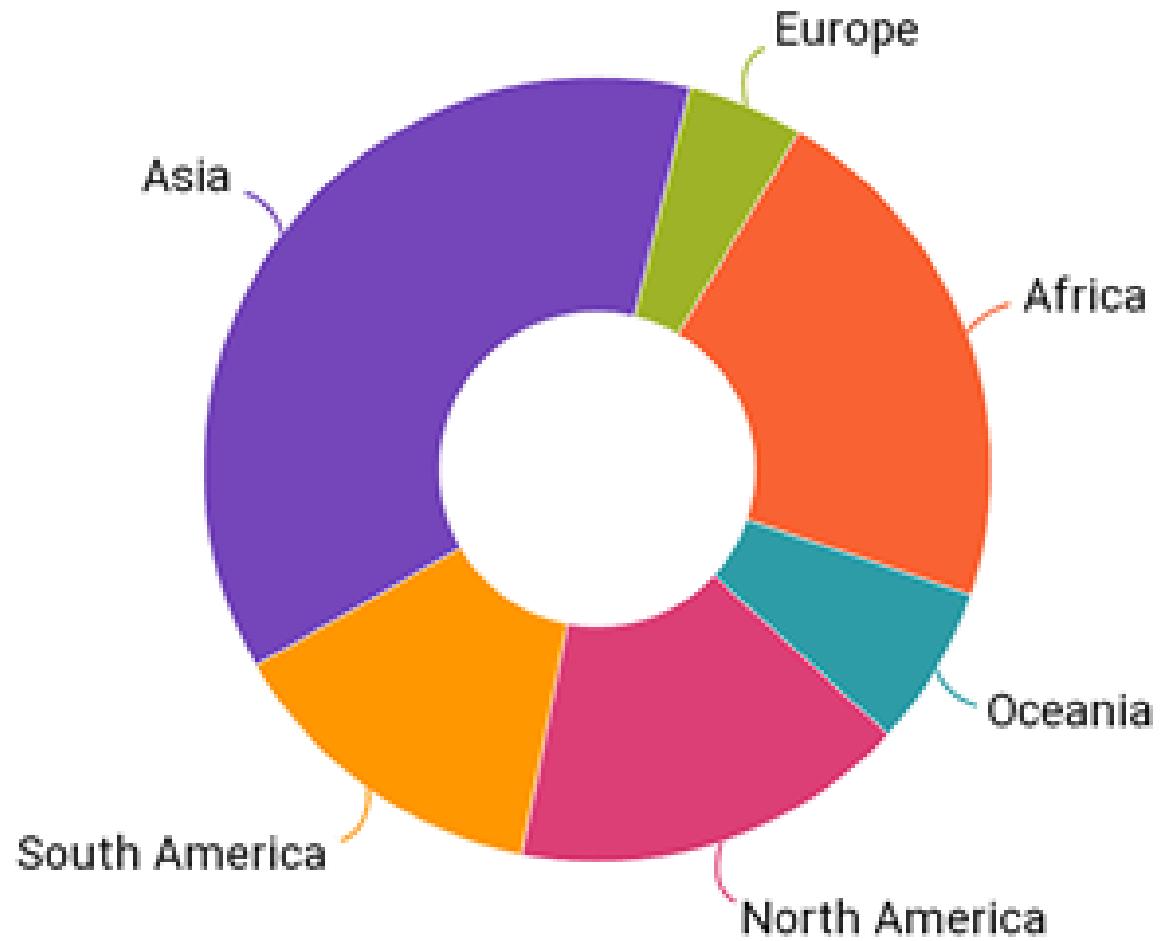
- Stock chart
- Surface chart
- Doughnut chart
- Radar chart
- Geographic mapping



Stock Chart

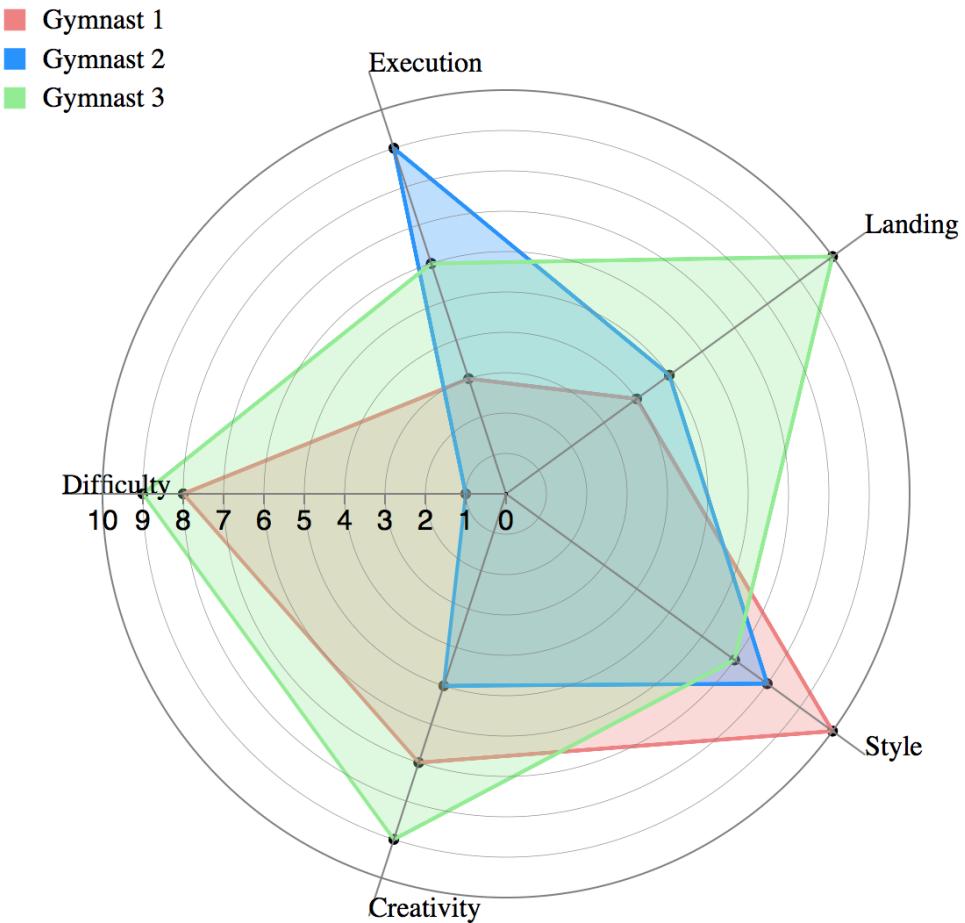
Surface Chart



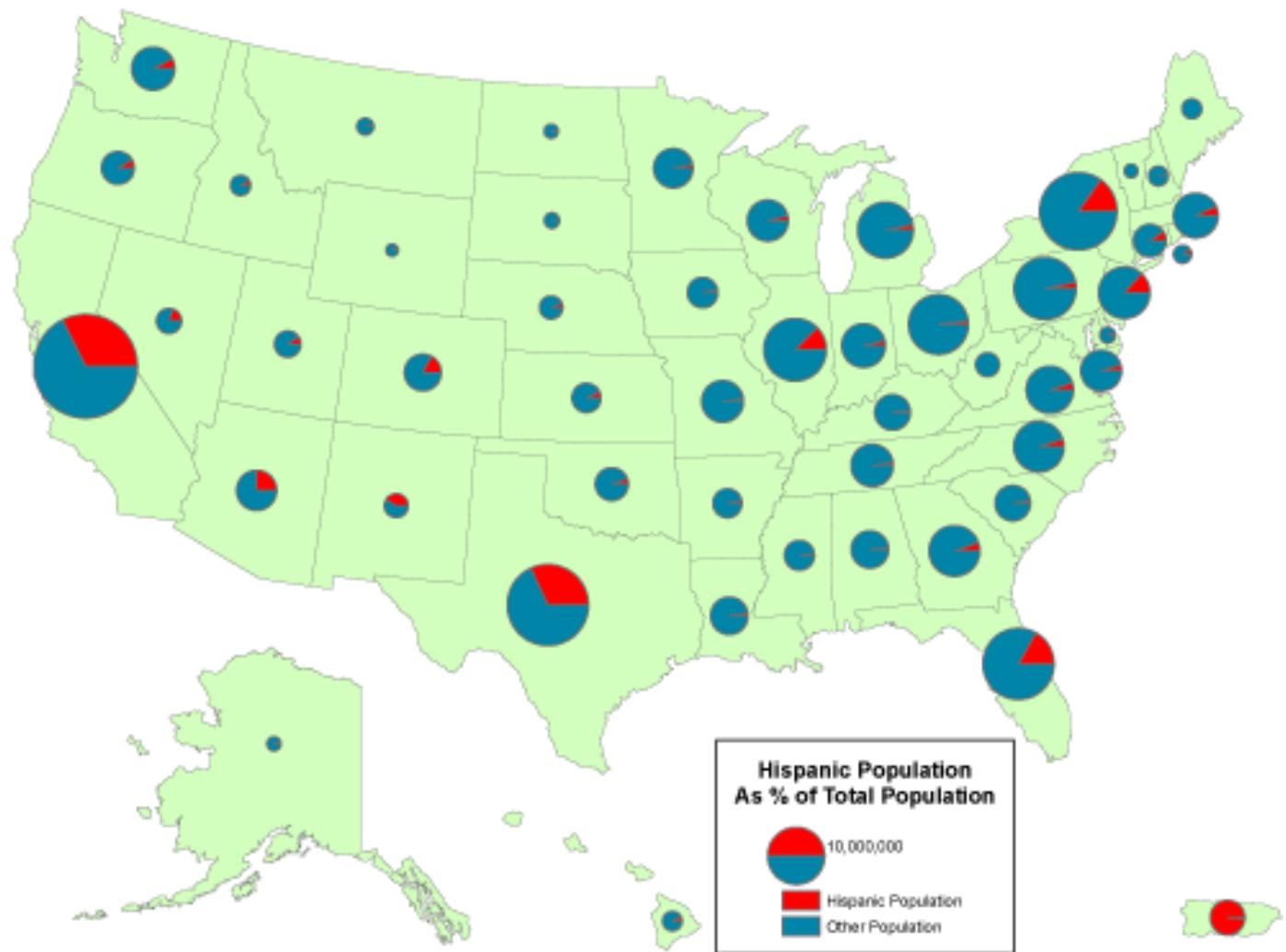


Doughnut Chart

Gymnast Scoring Radar Chart



Radar Chart

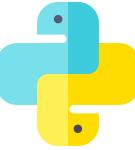


Geographic Mapping



Data Queries: Using Sorting and Filtering Business Analytics

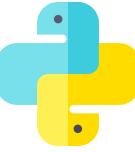
LECTURE 5



Sorting Data in the Purchase Orders Database

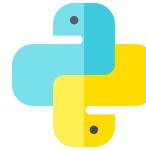
The screenshot shows the Microsoft Excel interface with the Data ribbon selected. A blue circle highlights the 'Sort' button under the 'Sort & Filter' section. Below the ribbon, a table titled 'Purchase Orders' is displayed. The first column is labeled 'Supplier'. A blue arrow points from the text 'Sort by Supplier' to the 'Supplier' column header in the table.

| | A | B | C | D | E | F | G | H | I | J |
|----|------------------|-----------|----------|--------------------|-----------|----------|----------------|--------------------|------------|--------------|
| 1 | Purchase Orders | | | | | | | | | |
| 2 | | | | | | | | | | |
| 3 | Supplier | Order No. | Item No. | Item Description | Item Cost | Quantity | Cost per order | A/P Terms (Months) | Order Date | Arrival Date |
| 4 | Alum Sheeting | A0223 | 4224 | Bolt-nut package | \$ 3.95 | 4,500 | \$ 17,775.00 | 30 | 10/15/11 | 10/20/11 |
| 5 | Alum Sheeting | A0433 | 5417 | Control Panel | \$ 255.00 | 500 | \$ 127,500.00 | 30 | 10/20/11 | 10/27/11 |
| 6 | Alum Sheeting | A0443 | 1243 | Airframe fasteners | \$ 4.25 | 10,000 | \$ 42,500.00 | 30 | 08/08/11 | 08/14/11 |
| 7 | Alum Sheeting | A0446 | 5417 | Control Panel | \$ 255.00 | 406 | \$ 103,530.00 | 30 | 09/01/11 | 09/10/11 |
| 8 | Alum Sheeting | B0247 | 1243 | Airframe fasteners | \$ 4.25 | 9,000 | \$ 38,250.00 | 30 | 09/05/11 | 09/12/11 |
| 9 | Alum Sheeting | B0447 | 5634 | Side Panel | \$ 185.00 | 150 | \$ 27,750.00 | 30 | 10/25/11 | 11/03/11 |
| 10 | Alum Sheeting | B0479 | 5634 | Side Panel | \$ 185.00 | 140 | \$ 25,900.00 | 30 | 10/29/11 | 11/04/11 |
| 11 | Alum Sheeting | B0567 | 1243 | Airframe fasteners | \$ 4.25 | 10,500 | \$ 44,625.00 | 30 | 10/10/11 | 10/17/11 |
| 12 | Durable Products | A1234 | 9399 | Gasket | \$ 3.65 | 1,250 | \$ 4,562.50 | 45 | 10/01/11 | 10/06/11 |
| 13 | Durable Products | A1235 | 9399 | Gasket | \$ 3.65 | 1,450 | \$ 5,292.50 | 45 | 10/03/11 | 10/08/11 |
| 14 | Durable Products | A1344 | 5454 | Control Panel | \$ 220.00 | 550 | \$ 121,000.00 | 45 | 10/09/11 | 10/14/11 |
| 15 | Durable Products | A1345 | 9399 | Gasket | \$ 3.65 | 1,470 | \$ 5,365.50 | 45 | 10/07/11 | 10/12/11 |
| 16 | Durable Products | A1346 | 9399 | Gasket | \$ 3.65 | 1,985 | \$ 7,245.25 | 45 | 10/05/11 | 10/11/11 |



Pareto Analysis

- An Italian economist, Vilfredo Pareto, observed in 1906 that a large proportion of the wealth in Italy was owned by a small proportion of the people.
- Similarly, businesses often find that a large proportion of sales come from a small proportion of customers.
- A Pareto analysis involves sorting data and calculating cumulative proportions.



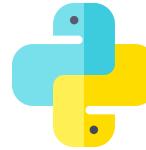
Applying the Pareto Principle

Sort by

| A | B | C | D | E | F | G | H | I | |
|----|-------------------|---------------|---------------|---------------|-----------------------|------------------|-----------------|------------|--------------|
| 1 | Bicycle Inventory | | | | | | | | |
| 2 | Product Category | Product Name | Purchase Cost | Selling Price | Supplier | Quantity on Hand | Inventory Value | Percentage | Cumulative % |
| 4 | Road | Runroad 5000 | \$450.95 | \$599.99 | Run-Up Bikes | 5 | \$ 2,254.75 | 11.2% | 11.2% |
| 5 | Road | Runroad 1000 | \$250.95 | \$350.99 | Run-Up Bikes | 8 | \$ 2,007.60 | 10.0% | 21.1% |
| 6 | Road | Elegant 210 | \$281.52 | \$394.13 | Bicyclist's Choice | 7 | \$ 1,970.64 | 9.8% | 30.9% |
| 7 | Road | Runroad 4000 | \$390.95 | \$495.99 | Run-Up Bikes | 5 | \$ 1,954.75 | 9.7% | 40.6% |
| 8 | Mtn. | Eagle 3 | \$350.52 | \$490.73 | Bike-One | 5 | \$ 1,752.60 | 8.7% | 49.3% |
| 9 | Road | Classic 109 | \$207.49 | \$290.49 | Bicyclist's Choice | 7 | \$ 1,452.43 | 7.2% | 56.5% |
| 10 | Hybrid | Eagle 7 | \$150.89 | \$211.46 | Bike-One | 9 | \$ 1,358.01 | 6.7% | 63.3% |
| 11 | Hybrid | Tea for Two | \$429.02 | \$609.00 | Simpson's Bike Supply | 3 | \$ 1,287.06 | 6.4% | 69.7% |
| 12 | Mtn. | Bluff Breaker | \$375.00 | \$495.00 | The Bike Path | 3 | \$ 1,125.00 | 5.6% | 75.2% |
| 13 | Mtn. | Eagle 2 | \$401.11 | \$561.54 | Bike-One | 2 | \$ 802.22 | 4.0% | 79.2% |
| 14 | Leisure | Breeze LE | \$109.95 | \$149.95 | The Bike Path | 5 | \$ 549.75 | 2.7% | 81.9% |
| 15 | Children | Runkidder 100 | \$50.95 | \$75.99 | Run-Up Bikes | 10 | \$ 509.50 | 2.5% | 84.5% |
| 16 | Mtn. | Jetty Breaker | \$455.95 | \$649.95 | The Bike Path | 1 | \$ 455.95 | 2.3% | 86.7% |
| 17 | Leisure | Runcool 3000 | \$85.95 | \$135.99 | Run-Up Bikes | 5 | \$ 429.75 | 2.1% | 88.9% |
| 18 | Children | Coolest 100 | \$69.99 | \$97.98 | Bicyclist's Choice | 6 | \$ 419.94 | 2.1% | 91.0% |
| 19 | Mtn. | Eagle 1 | \$410.01 | \$574.01 | Bike-One | 1 | \$ 410.01 | 2.0% | 93.0% |
| 20 | Children | Green Rider | \$95.47 | \$133.66 | Simpson's Bike Supply | 4 | \$ 381.88 | 1.9% | 94.9% |
| 21 | Leisure | Breeze | \$89.95 | \$130.95 | The Bike Path | 4 | \$ 359.80 | 1.8% | 96.7% |
| 22 | Leisure | Blue Moon | \$75.29 | \$105.41 | Simpson's Bike Supply | 4 | \$ 301.16 | 1.5% | 98.2% |
| 23 | Leisure | Supreme 350 | \$50.00 | \$70.00 | Bicyclist's Choice | 3 | \$ 150.00 | 0.7% | 98.9% |
| 24 | Children | Red Rider | \$15.00 | \$25.50 | Simpson's Bike Supply | 8 | \$ 120.00 | 0.6% | 99.5% |
| 25 | Leisure | Starlight | \$100.47 | \$140.66 | Simpson's Bike Supply | 1 | \$ 100.47 | 0.5% | 100.0% |
| 26 | Hybrid | Runblend 2000 | \$180.95 | \$255.99 | Run-Up Bikes | 0 | \$ - | 0.0% | 100.0% |
| 27 | Road | Twist & Shout | \$490.50 | \$635.70 | Simpson's Bike Supply | 0 | \$ - | 0.0% | 100.0% |
| 28 | | | | | Total | | \$ 20,153.27 | | |

Figure 3.15

75% of the bicycle inventory value comes from 40% (9/24) of items.



Filtering Records by Item Description

Highlight A3:J97

Data tab

Sort & Filter group

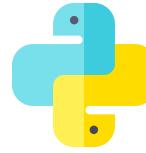
Filter

Click on the D3
dropdown arrow.

Select Bolt-nut
package to filter out
all other items.

| | A | B | C | D |
|----|------------------------|---------|---------|------------------|
| 1 | Purchase Orders | | | |
| 2 | | | | |
| 3 | Supplier | Order N | Item Nc | Item Description |
| 4 | Spacetime Technologies | A01 | | |
| 5 | Steelpin Inc. | A01 | | |
| 6 | Steelpin Inc. | A01 | | |
| 7 | Steelpin Inc. | A02 | | |
| 8 | Steelpin Inc. | A02 | | |
| 9 | Steelpin Inc. | A02 | | |
| 10 | Alum Sheeting | A02 | | |
| 11 | Alum Sheeting | A04 | | |
| 12 | Alum Sheeting | A04 | | |
| 13 | Alum Sheeting | A04 | | |
| 14 | Spacetime Technologies | A05 | | |
| 15 | Spacetime Technologies | A05 | | |
| 16 | Spacetime Technologies | A06 | | |
| 17 | Spacetime Technologies | A06 | | |
| 18 | Spacetime Technologies | A07 | | |
| 19 | Spacetime Technologies | A12 | | |
| 20 | Durable Products | A12 | | |
| 21 | Durable Products | A12 | | |
| 22 | Durable Products | A12 | | |

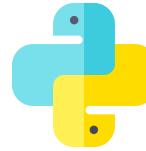
The screenshot shows a Microsoft Excel spreadsheet titled "Purchase Orders". The data is organized into columns: Supplier, Order N, Item Nc, and Item Description. The "Item Description" column header has a dropdown arrow, which is circled in blue. A context menu is open over this dropdown, showing options like "Sort A to Z", "Sort Z to A", "Sort by Color", "Clear Filter From 'Item Description'", "Filter by Color", and "Text Filters". The "Text Filters" section is expanded, showing a search bar and a list of items with checkboxes. The checkbox for "Bolt-nut package" is checked. At the bottom right of the filter dialog are "OK" and "Cancel" buttons.



Filtering Records by Item Description

Filter results for the bolt-nut package

| | A | B | C | D | E | F | G | H | I | J |
|----|------------------------|-----------|----------|------------------|----------|---------|--------------|-----------------|----------|-------------|
| 1 | Purchase Orders | | | | | | | | | |
| 2 | Supplier | Order No. | Item No. | Item Description | Item Cn. | Quantit | Cost per ord | A/P Terms (Mon) | Order Da | Arrival Dat |
| 6 | Steelpin Inc. | A0123 | 4312 | Bolt-nut package | \$ 3.75 | 4,250 | \$ 15,937.50 | 30 | 08/25/11 | 09/01/11 |
| 9 | Steelpin Inc. | A0207 | 4312 | Bolt-nut package | \$ 3.75 | 4,200 | \$ 15,750.00 | 30 | 09/01/11 | 09/10/11 |
| 10 | Alum Sheeting | A0223 | 4224 | Bolt-nut package | \$ 3.95 | 4,500 | \$ 17,775.00 | 30 | 10/15/11 | 10/20/11 |
| 19 | Spacetime Technologies | A1222 | 4111 | Bolt-nut package | \$ 3.55 | 4,200 | \$ 14,910.00 | 25 | 09/15/11 | 10/15/11 |
| 25 | Spacetime Technologies | A1444 | 4111 | Bolt-nut package | \$ 3.55 | 4,250 | \$ 15,087.50 | 25 | 09/20/11 | 10/10/11 |
| 26 | Spacetime Technologies | A1445 | 4111 | Bolt-nut package | \$ 3.55 | 4,200 | \$ 14,910.00 | 25 | 09/25/11 | 10/25/11 |
| 27 | Spacetime Technologies | A1449 | 4111 | Bolt-nut package | \$ 3.55 | 4,600 | \$ 16,330.00 | 25 | 10/05/11 | 10/19/11 |
| 29 | Durable Products | A1457 | 4569 | Bolt-nut package | \$ 3.50 | 3,900 | \$ 13,650.00 | 45 | 10/05/11 | 10/10/11 |
| 35 | Spacetime Technologies | A3467 | 4111 | Bolt-nut package | \$ 3.55 | 4,800 | \$ 17,040.00 | 25 | 09/05/11 | 09/20/11 |
| 36 | Spacetime Technologies | A5689 | 4111 | Bolt-nut package | \$ 3.55 | 4,585 | \$ 16,276.75 | 25 | 09/10/11 | 09/30/11 |
| 43 | Steelpin Inc. | B0445 | 4312 | Bolt-nut package | \$ 3.75 | 4,150 | \$ 15,562.50 | 30 | 09/03/11 | 09/11/11 |

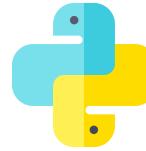


Filtering Records by Item Cost

To identify items that cost at least \$200

- Click on dropdown arrow for item cost
- *Number Filters*
Greater Than Or Equal To...

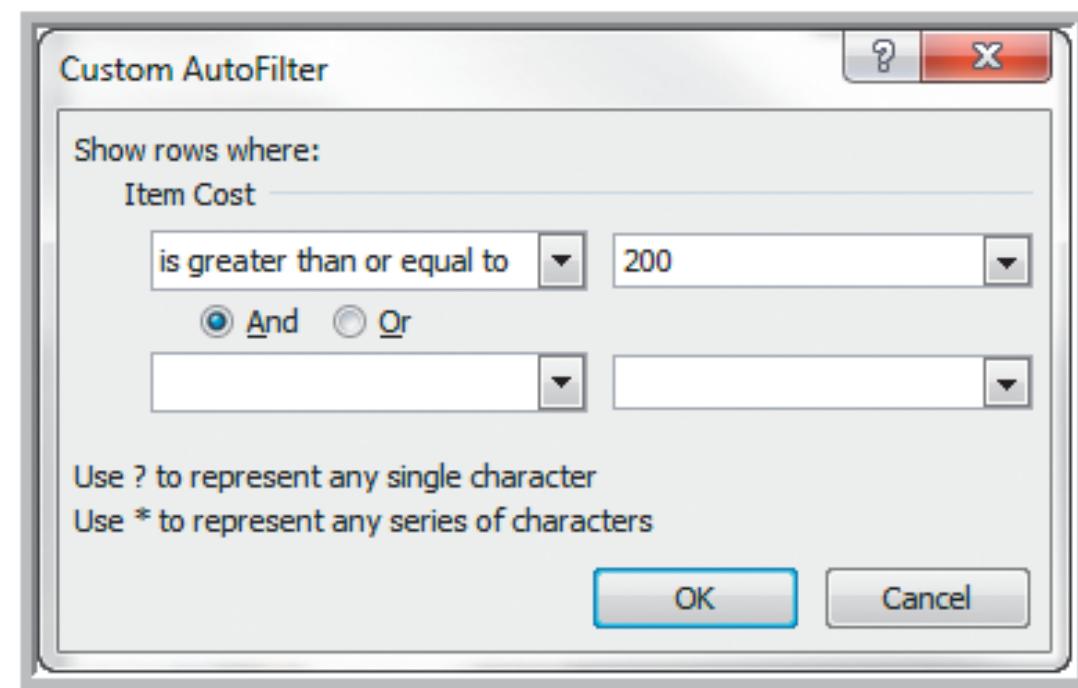
| A | B | C | D | E |
|--------------------------|---------|--------|-------------------------------|---------|
| 1 Purchase Orders | | | | |
| 2 | | | | |
| 3 Supplier | Order N | Item N | Item Description | Item Co |
| 4 Spacetime Technologies | A0111 | 648 | Sort Smallest to Largest | |
| 5 Steelpin Inc. | A0115 | 531 | Sort Largest to Smallest | |
| 6 Steelpin Inc. | A0123 | 431 | Sort by Color | |
| 7 Steelpin Inc. | A0204 | 531 | Clear Filter From "Item Cost" | |
| 8 Steelpin Inc. | A0205 | 561 | Filter by Color | |
| 9 Steelpin Inc. | A0207 | 421 | Number Filters | |
| 10 Alum Sheeting | | | Search | |
| 11 Alum Sheeting | | | (Select All) | |
| 12 Alum Sheeting | | | \$0.55 | |
| 13 Alum Sheeting | | | \$0.75 | |
| 14 Spacetime Tech | | | \$0.85 | |
| 15 Spacetime Tech | | | \$0.95 | |
| 16 Spacetime Tech | | | \$1.00 | |
| 17 Spacetime Tech | | | \$1.05 | |
| 18 Spacetime Tech | | | \$1.10 | |
| 19 Spacetime Tech | | | \$1.15 | |
| 20 Durable Prod | | | \$1.25 | |
| 21 Durable Prod | | | OK | |
| 22 Durable Prod | | | Cancel | |



Filtering by Item Cost

Custom AutoFilter
dialog box

- Click OK
- Only items costing at least \$200 is then displayed.



Data Queries: Using Sorting and Filtering

AutoFilter criteria is based on the data type.

- *Number Filters* includes numerical criteria.
- *Date Filters* include tomorrow, next week, etc.

AutoFilter can be used sequentially.

- First filter by one variable.
- Then filter those data by another variable.



Data Queries: Using Sorting and Filtering

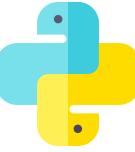
Analytics in Practice: Discovering Value of Data Analysis at Alders International

- ▶ Duty free operations at airports, seaports, etc.
- ▶ Maintain a data warehouse to track point-of-sale information and inventory levels.
- ▶ Pareto analysis revealed that 80% of profits were generated from 20% of their product lines.
- ▶ Allows selective elimination of less profitable items.



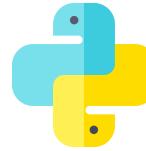
Statistical Methods for Summarizing Data Business Analytics

LECTURE 5



Statistical Methods for Summarizing Data

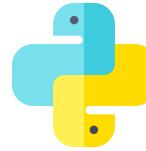
- A statistic is a summary measure of data.
- Descriptive statistics are methods that describe and summarize data.
- Microsoft Excel supports statistical analysis in two ways:
 1. Statistical functions
 2. *Analysis Toolpak* add-in for PCs
(for Macs, StatPlus is similar)



Constructing a Frequency Distribution for Items in the Purchase Order Database

Copy Column D (Item Description) to Column A in a new worksheet

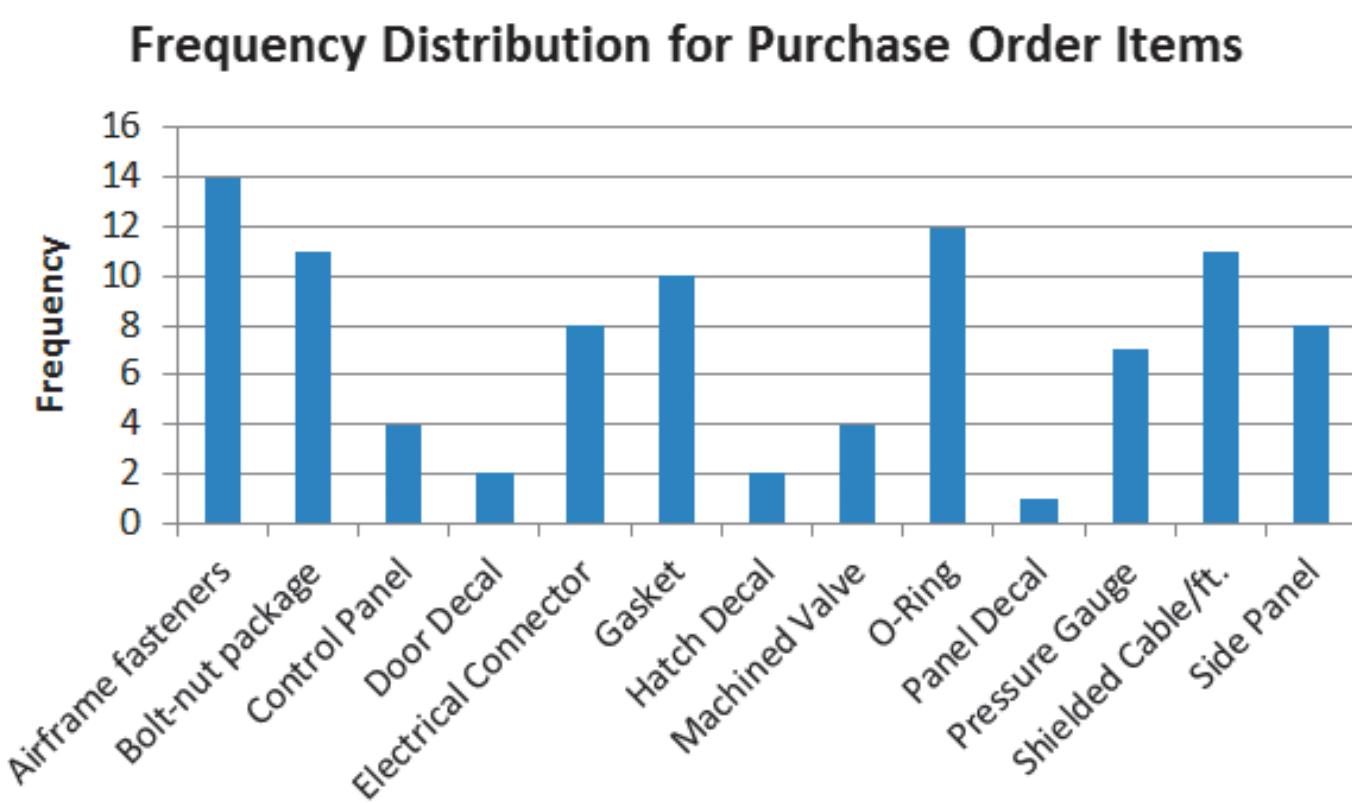
| | A | B | C | D | E | F | G | H | I | J |
|----|------------------------|----------|----------|--------------------|-----------|----------|----------------|--------------------|------------|--------------|
| 1 | Purchase Orders | | | | | | | | | |
| 2 | | | | | | | | | | |
| 3 | Supplier | Order No | Item No. | Item Description | Item Cost | Quantity | Cost per order | A/P Terms (Months) | Order Date | Arrival Date |
| 4 | Spacetime Technologies | A0111 | 6489 | O-Ring | \$ 3.00 | 900 | \$ 2,700.00 | 25 | 10/10/11 | 10/18/11 |
| 5 | Steelpin Inc. | A0115 | 5319 | Shielded Cable/ft. | \$ 1.10 | 17,500 | \$ 19,250.00 | 30 | 08/20/11 | 08/31/11 |
| 6 | Steelpin Inc. | A0123 | 4312 | Bolt-nut package | \$ 3.75 | 4,250 | \$ 15,937.50 | 30 | 08/25/11 | 09/01/11 |
| 7 | Steelpin Inc. | A0204 | 5319 | Shielded Cable/ft. | \$ 1.10 | 16,500 | \$ 18,150.00 | 30 | 09/15/11 | 10/05/11 |
| 8 | Steelpin Inc. | A0205 | 5677 | Side Panel | \$195.00 | 120 | \$ 23,400.00 | 30 | 11/02/11 | 11/13/11 |
| 9 | Steelpin Inc. | A0207 | 4312 | Bolt-nut package | \$ 3.75 | 4,200 | \$ 15,750.00 | 30 | 09/01/11 | 09/10/11 |
| 10 | Alum Sheeting | A0223 | 4224 | Bolt-nut package | \$ 3.95 | 4,500 | \$ 17,775.00 | 30 | 10/15/11 | 10/20/11 |
| 11 | Alum Sheeting | A0433 | 5417 | Control Panel | \$255.00 | 500 | \$ 127,500.00 | 30 | 10/20/11 | 10/27/11 |
| 12 | Alum Sheeting | A0443 | 1243 | Airframe fasteners | \$ 4.25 | 10,000 | \$ 42,500.00 | 30 | 08/08/11 | 08/14/11 |
| 13 | Alum Sheeting | A0446 | 5417 | Control Panel | \$255.00 | 406 | \$ 103,530.00 | 30 | 09/01/11 | 09/10/11 |
| 14 | Spacetime Technologies | A0533 | 9752 | Gasket | \$ 4.05 | 1,500 | \$ 6,075.00 | 25 | 09/20/11 | 09/25/11 |
| 15 | Spacetime Technologies | A0555 | 6489 | O-Ring | \$ 3.00 | 1,100 | \$ 3,300.00 | 25 | 10/05/11 | 10/10/11 |



Constructing a Frequency Distribution for Items in the Purchase Order Database

| | A | B |
|-----|-------------------------|-------------------------------|
| 100 | Item Description | Frequency |
| 101 | Airframe fasteners | =COUNTIF(\$D\$4:\$D\$97,A101) |
| 102 | Bolt-nut package | =COUNTIF(\$D\$4:\$D\$97,A102) |
| 103 | Control Panel | =COUNTIF(\$D\$4:\$D\$97,A103) |
| 104 | Door Decal | =COUNTIF(\$D\$4:\$D\$97,A104) |
| 105 | Electrical Connector | =COUNTIF(\$D\$4:\$D\$97,A105) |
| 106 | Gasket | =COUNTIF(\$D\$4:\$D\$97,A106) |
| 107 | Hatch Decal | =COUNTIF(\$D\$4:\$D\$97,A107) |
| 108 | Machined Valve | =COUNTIF(\$D\$4:\$D\$97,A108) |
| 109 | O-Ring | =COUNTIF(\$D\$4:\$D\$97,A109) |
| 110 | Panel Decal | =COUNTIF(\$D\$4:\$D\$97,A110) |
| 111 | Pressure Gauge | =COUNTIF(\$D\$4:\$D\$97,A111) |
| 112 | Shielded Cable/ft. | =COUNTIF(\$D\$4:\$D\$97,A112) |
| 113 | Side Panel | =COUNTIF(\$D\$4:\$D\$97,A113) |

| | A | B |
|-----|-------------------------|------------------|
| 100 | Item Description | Frequency |
| 101 | Airframe fasteners | 14 |
| 102 | Bolt-nut package | 11 |
| 103 | Control Panel | 4 |
| 104 | Door Decal | 2 |
| 105 | Electrical Connector | 8 |
| 106 | Gasket | 10 |
| 107 | Hatch Decal | 2 |
| 108 | Machined Valve | 4 |
| 109 | O-Ring | 12 |
| 110 | Panel Decal | 1 |
| 111 | Pressure Gauge | 7 |
| 112 | Shielded Cable/ft. | 11 |
| 113 | Side Panel | 8 |



Constructing a
Frequency
Distribution for
Items in the
Purchase Order
Database

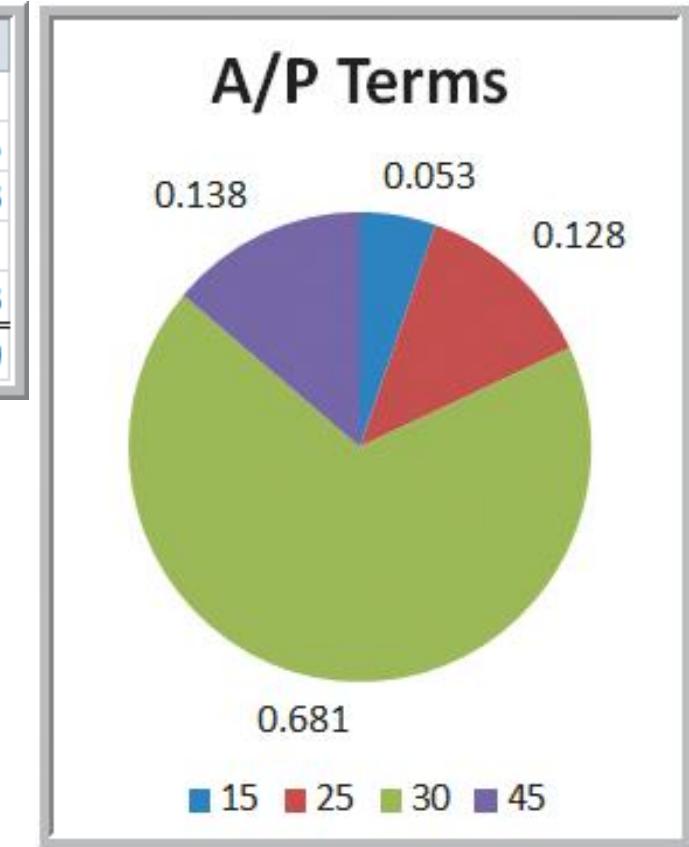
Constructing a Relative Frequency Distribution for Items Purchased

Compute relative
frequencies by
dividing each
frequency by 94.

| | A | B | C |
|-----|----------------------|-----------|--------------------|
| 100 | Item Description | Frequency | Relative Frequency |
| 101 | Airframe fasteners | 14 | 0.149 |
| 102 | Bolt-nut package | 11 | 0.117 |
| 103 | Control Panel | 4 | 0.043 |
| 104 | Door Decal | 2 | 0.021 |
| 105 | Electrical Connector | 8 | 0.085 |
| 106 | Gasket | 10 | 0.106 |
| 107 | Hatch Decal | 2 | 0.021 |
| 108 | Machined Valve | 4 | 0.043 |
| 109 | O-Ring | 12 | 0.128 |
| 110 | Panel Decal | 1 | 0.011 |
| 111 | Pressure Gauge | 7 | 0.074 |
| 112 | Shielded Cable/ft. | 11 | 0.117 |
| 113 | Side Panel | 8 | 0.085 |
| 114 | Total | 94 | 1.000 |

Frequency and Relative Frequency Distribution for A/P Terms

| | A | B | C |
|-----|-----------|-----------|--------------------|
| 117 | A/P Terms | Frequency | Relative Frequency |
| 118 | 15 | 5 | 0.053 |
| 119 | 25 | 12 | 0.128 |
| 120 | 30 | 64 | 0.681 |
| 121 | 45 | 13 | 0.138 |
| 122 | Total | 94 | 1.000 |



Excel's Histogram Tool

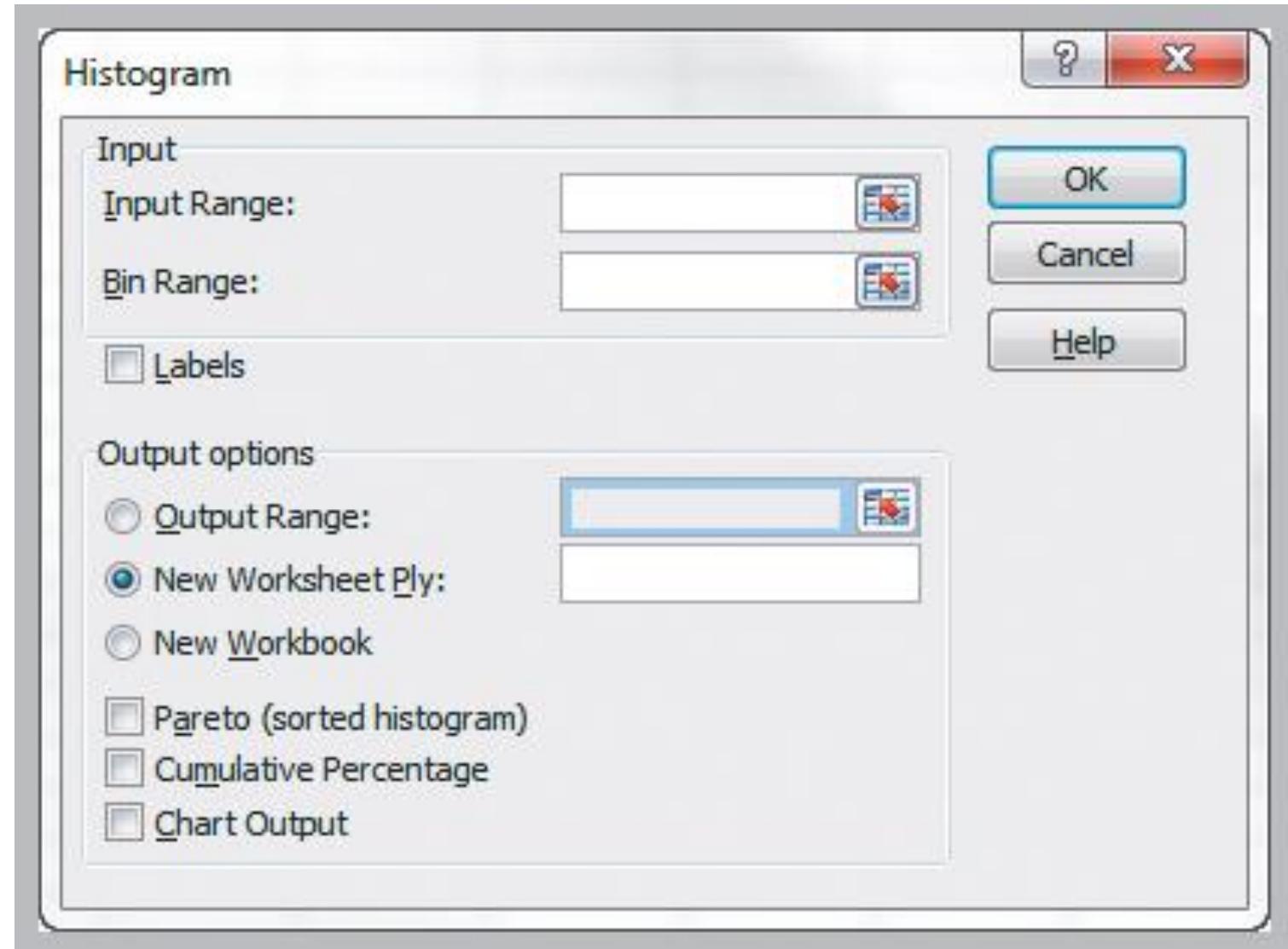
Using the Analysis Toolpak

Data
Data Analysis
Histogram

Fill in the *Input Range* and *Bin Range* (optional).

Choose *Labels* if columns have headers rows.

Choose *Chart Output*.



Using the Histogram Tool for A/P Terms

A/P data in H3:H97

Bins below in H99:H103

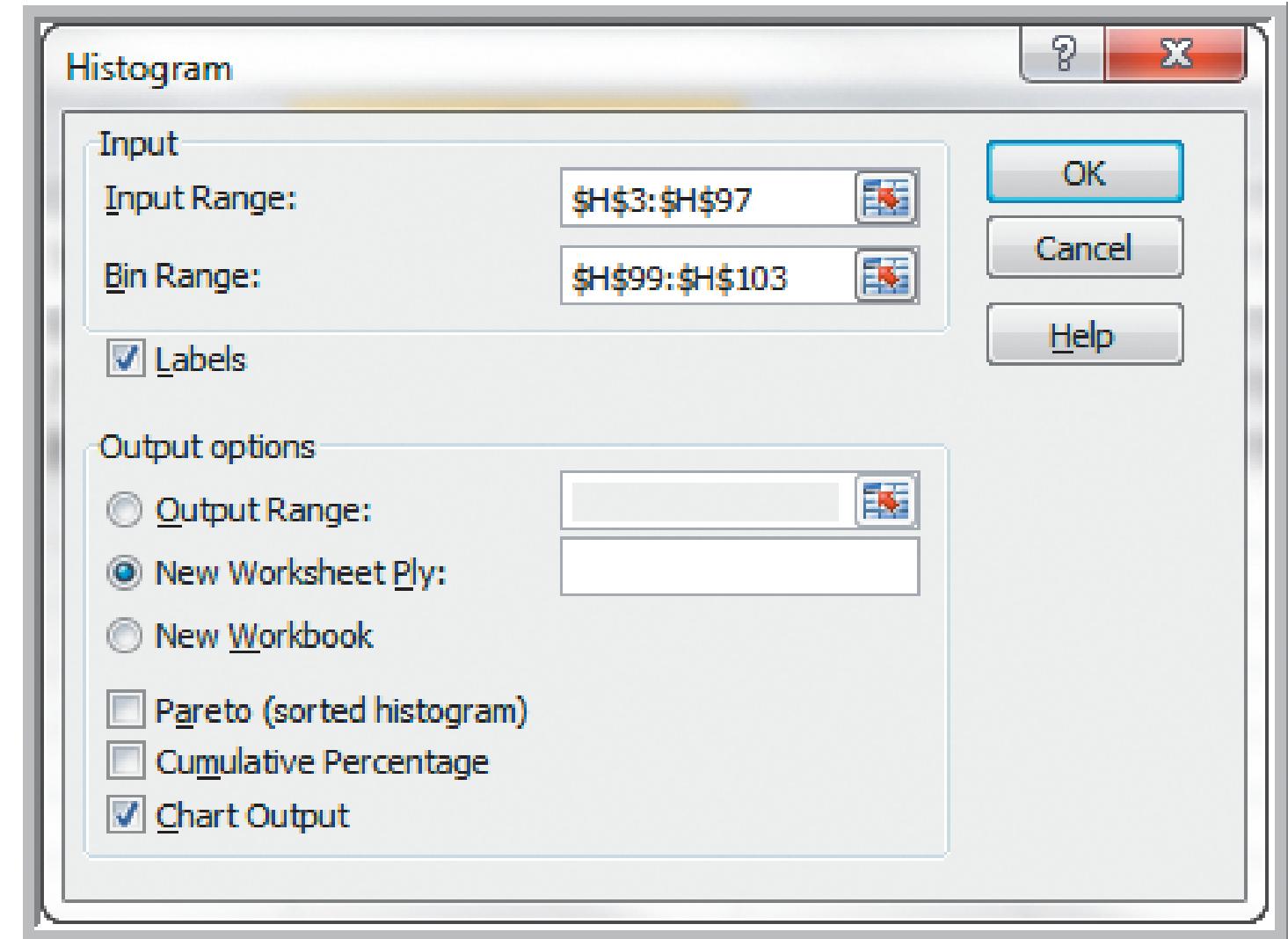
Month

15

25

30

45



Using the Histogram Tool for A/P Terms

Some program needed to create the data for histogram.

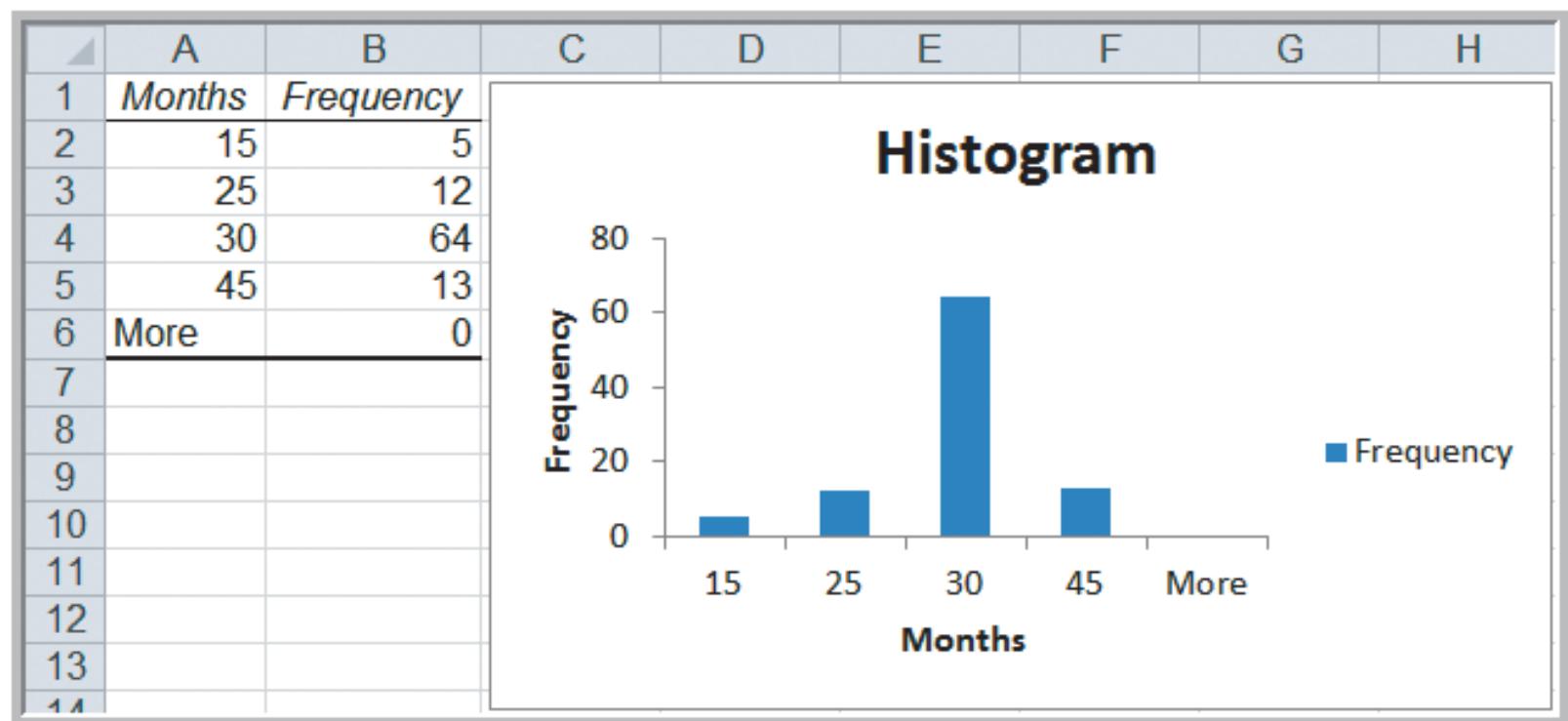
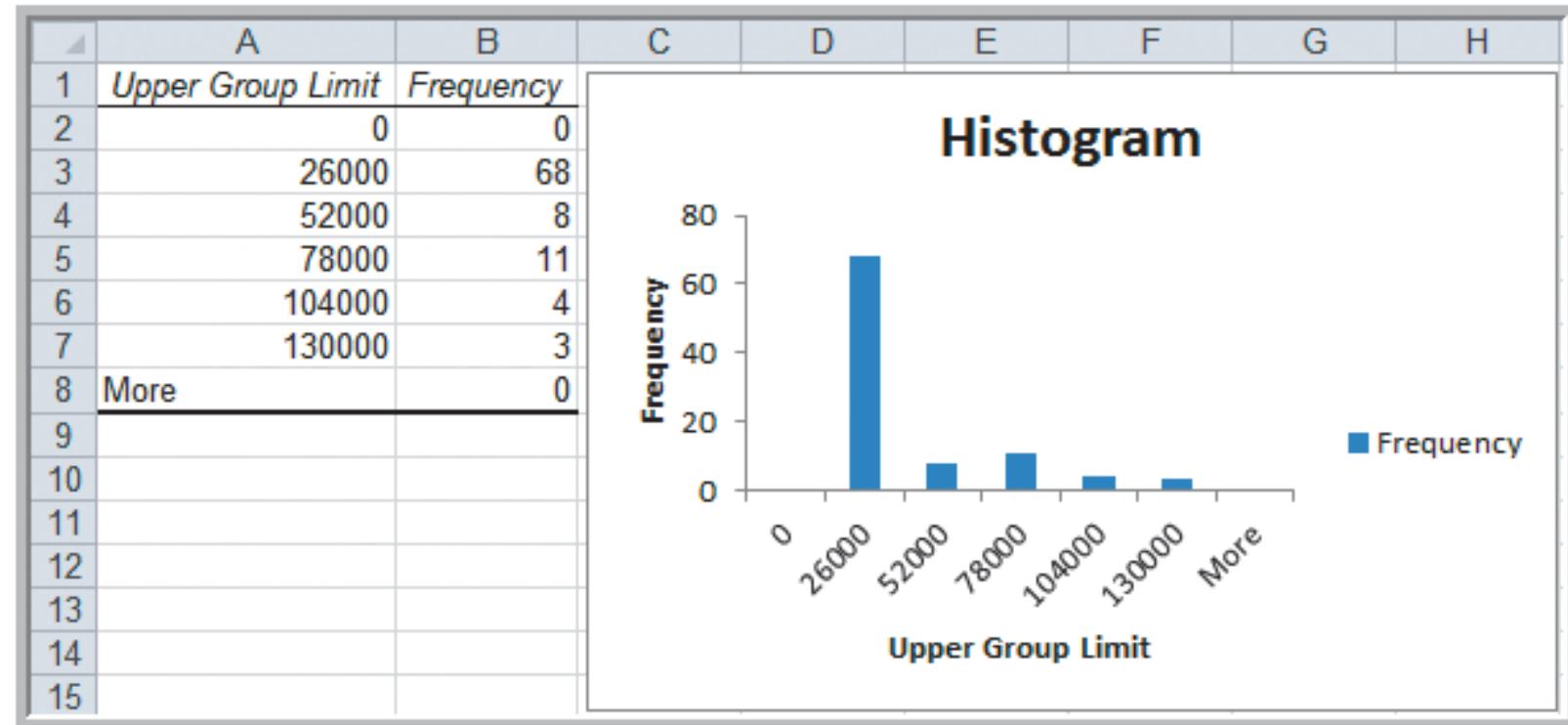


Table above is
not linked to
chart.

Constructing a Frequency Distribution and Histogram for Cost Per Order

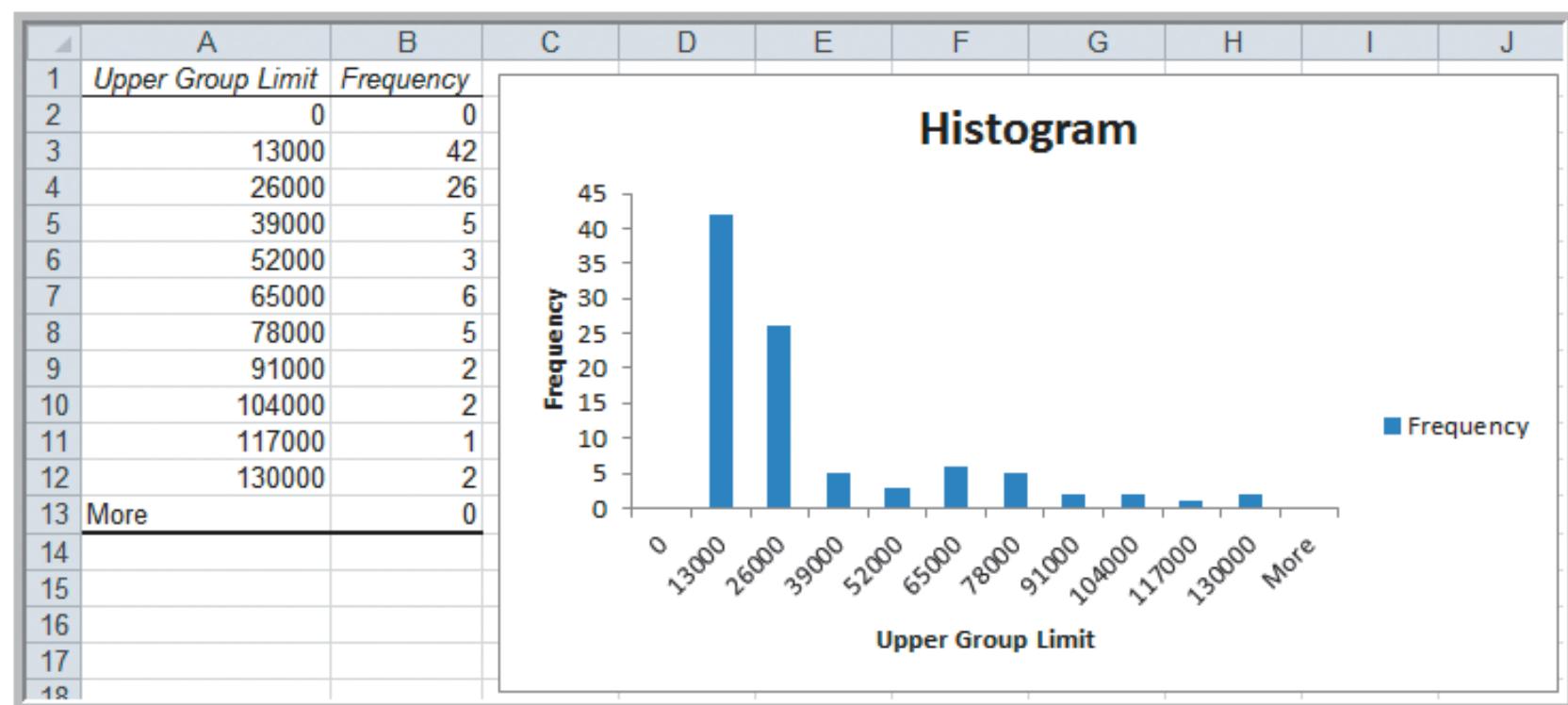
Some program needed to create the data for histogram.



5 groups with a \$26,000 group width

Constructing a Frequency Distribution and Histogram for Cost Per Order

Some program needed to create the data for histogram.



10 groups with a \$13,000 group width

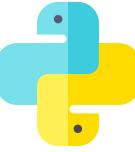
Computing Cumulative Relative Frequencies for the Cost Per Order Data

Some program needed
to create the data.

| A | B | C | D | |
|----|-------------------|-----------|--------------------|-------------------------------|
| 1 | Upper Group Limit | Frequency | Relative Frequency | Cumulative Relative Frequency |
| 2 | 0 | 0 | 0.000 | 0.000 |
| 3 | 13000 | 42 | 0.447 | 0.447 |
| 4 | 26000 | 26 | 0.277 | 0.723 |
| 5 | 39000 | 5 | 0.053 | 0.777 |
| 6 | 52000 | 3 | 0.032 | 0.809 |
| 7 | 65000 | 6 | 0.064 | 0.872 |
| 8 | 78000 | 5 | 0.053 | 0.926 |
| 9 | 91000 | 2 | 0.021 | 0.947 |
| 10 | 104000 | 2 | 0.021 | 0.968 |
| 11 | 117000 | 1 | 0.011 | 0.979 |
| 12 | 130000 | 2 | 0.021 | 1.000 |
| 13 | More | 0 | 0.000 | 1.000 |
| 14 | | 94 | | |



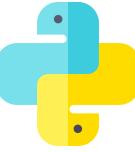
Ogive



Computing Percentiles

Compute the 90th percentile for *cost per order* in the Purchase Orders Data.

- Rank of k^{th} percentile = $\frac{nk}{100} + 0.5$
- $n = 94$ observations
- $k = 90$
- Rank of 90th percentile = $94(90)/100+0.5$
= 85.1 (round to 85)
- Value of the 85th observation = \$74,375



Computing Percentiles in Excel

Compute the 90th percentile for *cost per order*.

- Excel function for the k^{th} percentile:

=PERCENTILE.INC(array, k)

=PERCENTILE.INC(G4:G97, 0.90)

= \$73,737.50

- Excel does not use the formula on previous slide.



Excel's Rank and Percentile Tool

Data

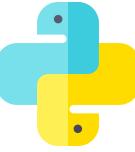
Data Analysis

Rank and Percentile

90.3rd percentile
= \$74,375

(same result as
manually computing
the 90th percentile)

| | A | B | C | D |
|----|-------|----------------|------|---------|
| 1 | Point | Cost per order | Rank | Percent |
| 2 | 94 | \$ 127,500.00 | 1 | 100.00% |
| 3 | 93 | \$ 121,000.00 | 2 | 98.90% |
| 4 | 92 | \$ 110,000.00 | 3 | 97.80% |
| 5 | 91 | \$ 103,530.00 | 4 | 96.70% |
| 6 | 90 | \$ 96,750.00 | 5 | 95.60% |
| 7 | 89 | \$ 82,875.00 | 6 | 94.60% |
| 8 | 88 | \$ 81,937.50 | 7 | 93.50% |
| 9 | 87 | \$ 77,400.00 | 8 | 92.40% |
| 10 | 86 | \$ 76,500.00 | 9 | 91.30% |
| 11 | 85 | \$ 74,375.00 | 10 | 90.30% |
| 12 | 84 | \$ 72,250.00 | 11 | 89.20% |
| 13 | 83 | \$ 65,875.00 | 12 | 88.10% |
| 14 | 82 | \$ 64,500.00 | 13 | 87.00% |
| 15 | 81 | \$ 63,750.00 | 14 | 86.00% |



Computing Quartiles in Excel

Compute the Quartiles of the Cost per Order data

- Excel function for quartiles: =QUARTILE.INC(*array, quart*)
- =QUARTILE.INC(G4:G97, 1) = \$6,757.81
- =QUARTILE.INC(G4:G97, 2) = \$15,656.25
- =QUARTILE.INC(G4:G97, 3) = \$27,593.75
- =QUARTILE.INC(G4:G97, 4) = \$127,500.00

Constructing a Cross-Tabulation

Excel's PivotTable
(covered next)
makes this easy.

- *Sales Transactions* database

| A | B | C | D | E | F | G | H | |
|----|-----------------------------|--------|---------|------------------|--------|----------|---------|-------------|
| 1 | Sales Transactions: July 14 | | | | | | | |
| 2 | | | | | | | | |
| 3 | Cust ID | Region | Payment | Transaction Code | Source | Amount | Product | Time Of Day |
| 4 | 10001 | East | Paypal | 93816545 | Web | \$20.19 | DVD | 22:19 |
| 5 | 10002 | West | Credit | 74083490 | Web | \$17.85 | DVD | 13:27 |
| 6 | 10003 | North | Credit | 64942368 | Web | \$23.98 | DVD | 14:27 |
| 7 | 10004 | West | Paypal | 70560957 | Email | \$23.51 | Book | 15:38 |
| 8 | 10005 | South | Credit | 35208817 | Web | \$15.33 | Book | 15:21 |
| 9 | 10006 | West | Paypal | 20978903 | Email | \$17.30 | DVD | 13:11 |
| 10 | 10007 | East | Credit | 80103311 | Web | \$177.72 | Book | 21:59 |
| 11 | 10008 | West | Credit | 14132683 | Web | \$21.76 | Book | 4:04 |
| 12 | 10009 | West | Paypal | 40128225 | Web | \$15.92 | DVD | 19:35 |
| 13 | 10010 | South | Paypal | 49073721 | Web | \$23.39 | DVD | 13:26 |

- Identify the number (and percentage) of books and DVDs ordered by region.

Constructing a Cross-Tabulation

Excel's PivotTable (covered next) makes this easy.

Table 3.1

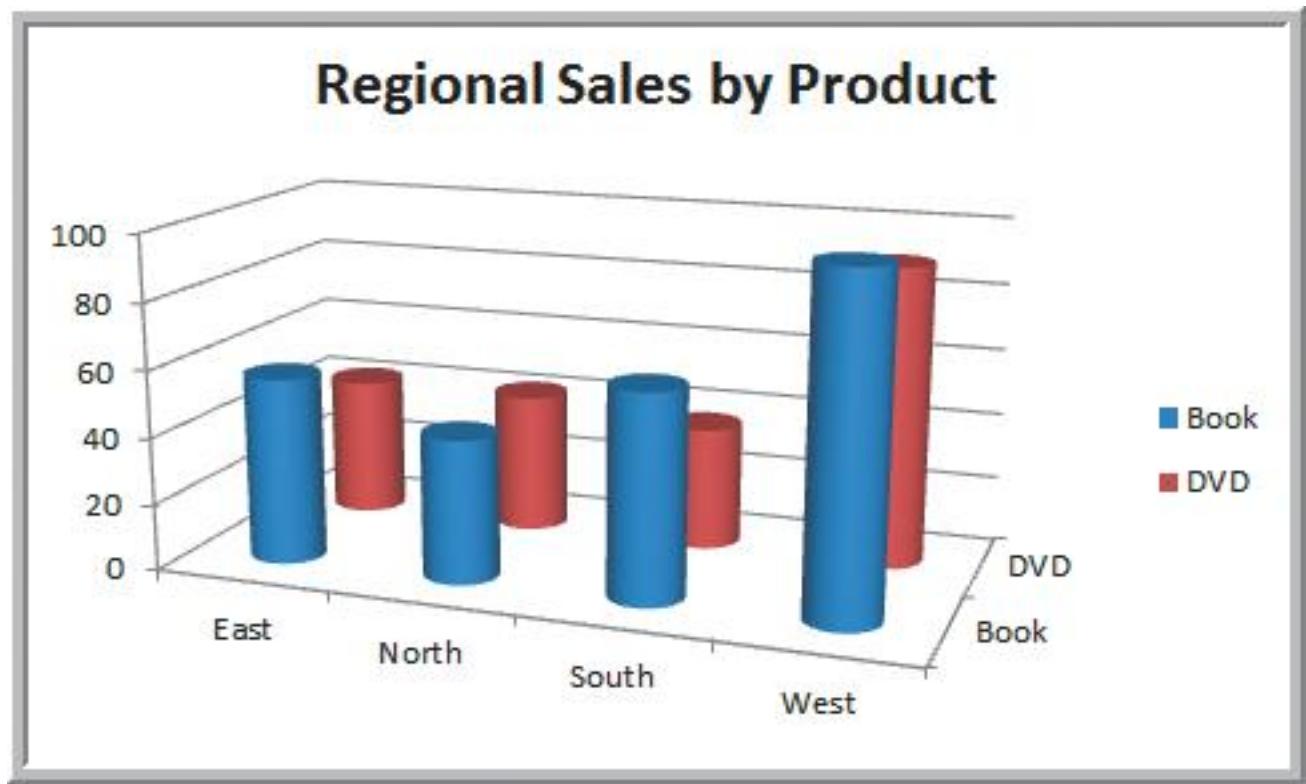
| Region | Book | DVD | Total |
|--------|------|-----|-------|
| East | 56 | 42 | 98 |
| North | 43 | 42 | 85 |
| South | 62 | 37 | 99 |
| West | 100 | 90 | 190 |
| Total | 261 | 211 | 472 |

Table 3.2

| Region | Book | DVD | Total |
|--------|-------|-------|--------|
| East | 57.1% | 42.9% | 100.0% |
| North | 50.6% | 49.4% | 100.0% |
| South | 62.6% | 37.4% | 100.0% |
| West | 52.6% | 47.4% | 100.0% |

Constructing a
Cross-Tabulation

Excel's PivotTable
(covered next)
makes this easy.

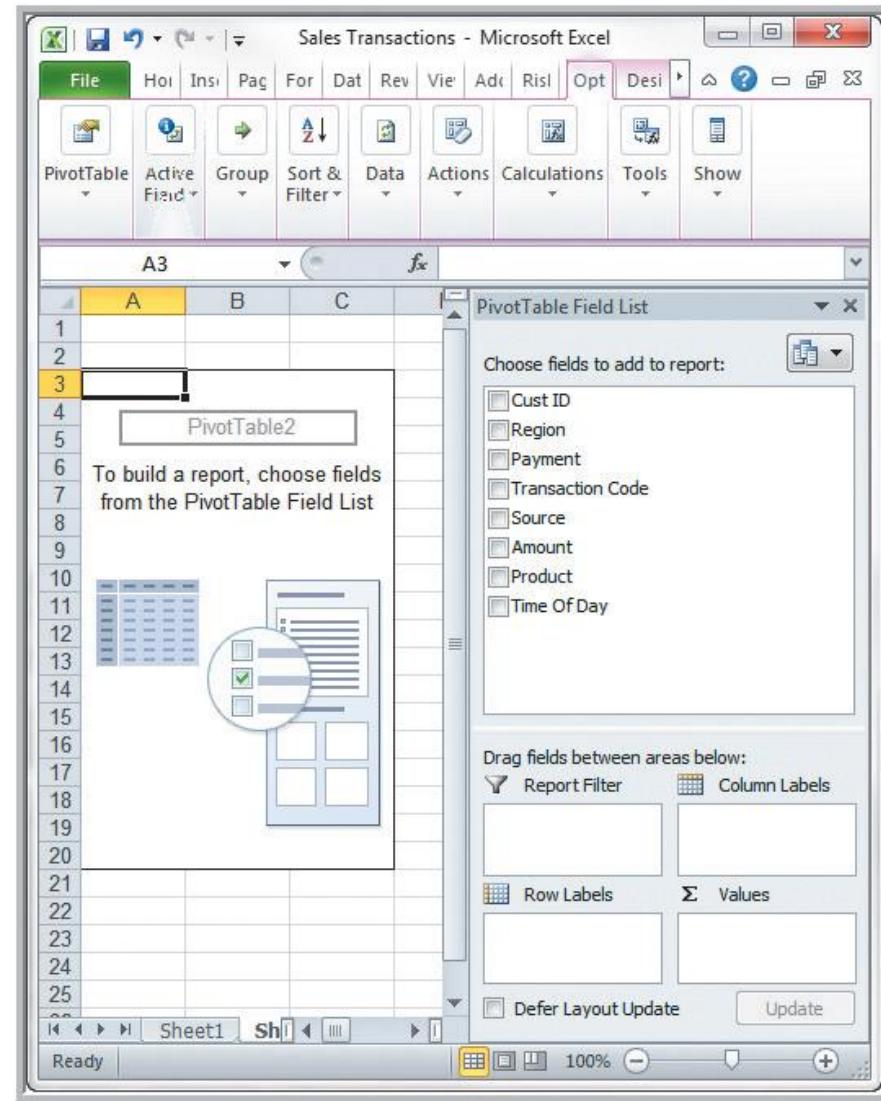


| Region | Book | DVD | Total |
|--------|------|-----|-------|
| East | 56 | 42 | 98 |
| North | 43 | 42 | 85 |
| South | 62 | 37 | 99 |
| West | 100 | 90 | 190 |
| Total | 261 | 211 | 472 |



Exploring Data Using PivotTables Business Analytics

LECTURE 5

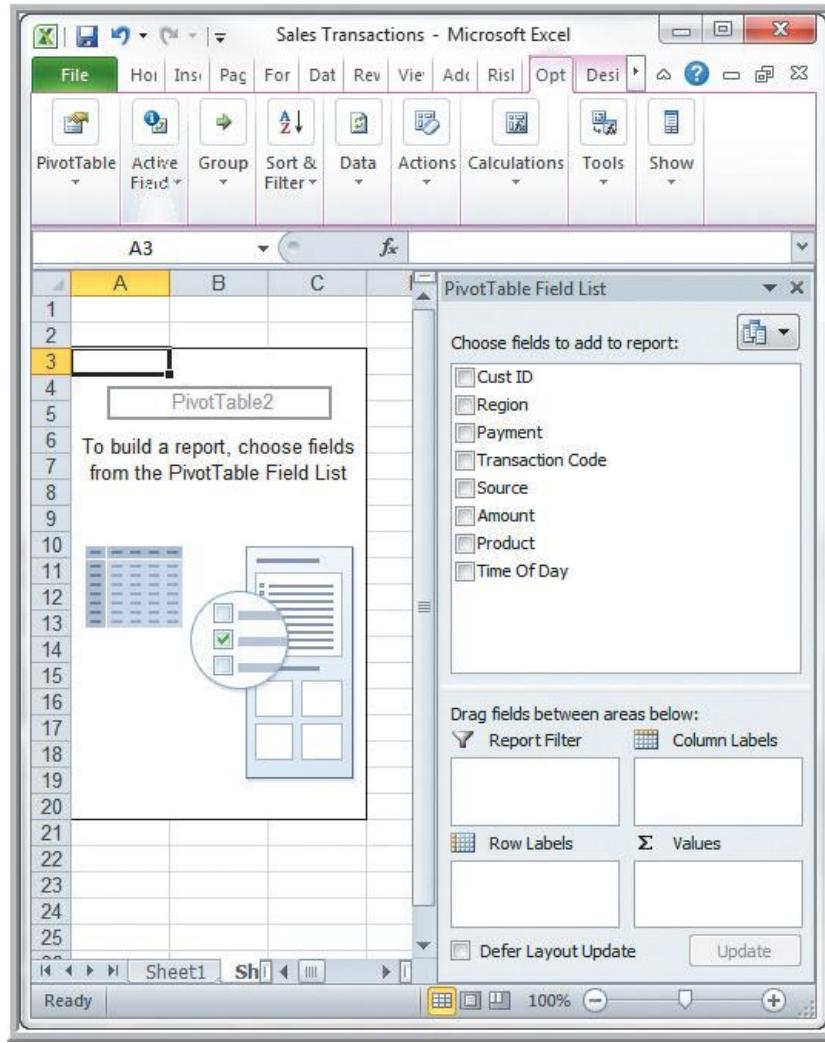


Data Tables: PivotTable

Follow wizard steps.

PivotTables allow:

- Quick creation of cross tabulations
- Numerous custom-made summary tables and charts



PivotTable Field List

Select the fields for:

- *Report Filter*
- *Column Labels*
- *Row Labels*
- Σ *Values*

Or, before choosing *PivotTable*, you can select a cell in the data and let Excel prepare a default PivotTable.

Sales Transactions cross tab and pivot table working f...

PivotTable Tools

A3 Sum of Cust ID

| | A | B | C | D |
|----|----------------|---------------|---------|-------------|
| 1 | | | | |
| 2 | | | | |
| 3 | Sum of Cust ID | Column Labels | DVD | Grand Total |
| 4 | Row Labels | Book | | |
| 5 | East | 572755 | 428278 | 1001033 |
| 6 | North | 441841 | 429848 | 871689 |
| 7 | South | 634963 | 379724 | 1014687 |
| 8 | West | 1024473 | 919746 | 1944219 |
| 9 | Grand Total | 2674032 | 2157596 | 4831628 |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |
| 17 | | | | |
| 18 | | | | |
| 19 | | | | |
| 20 | | | | |
| 21 | | | | |
| 22 | | | | |
| 23 | | | | |
| 24 | | | | |
| 25 | | | | |
| 26 | | | | |

PivotTable Field List

Choose fields to add to report:

- Cust ID
- Region
- Payment
- Transaction Code
- Source
- Amount
- Product
- Time Of Day

Drag fields between areas below:

Report Filter

Column Labels

Product

Row Labels

Σ Values

Region

Sum of Cust ID

Defer Layout Update

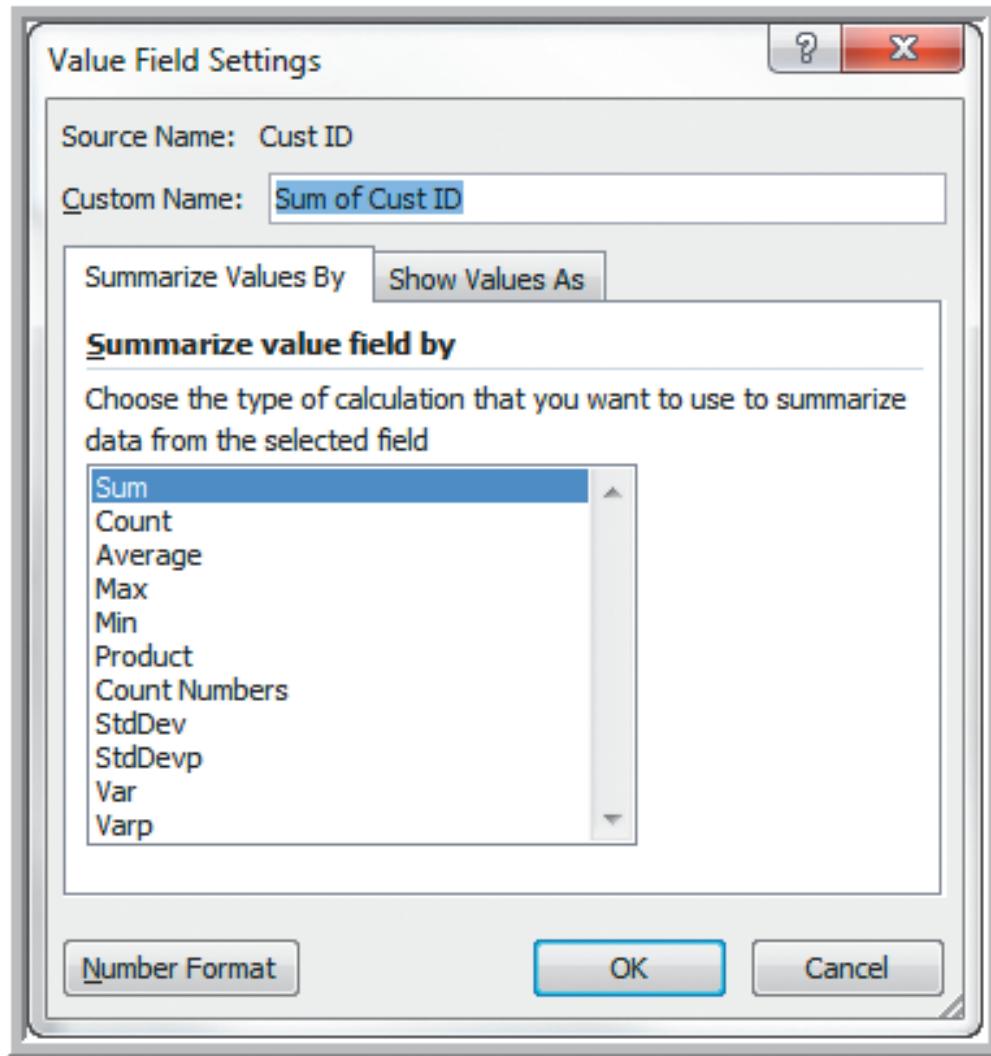
Update

Sheet2 July 14 Sales

Ready

Creating a PivotTable

- Default PivotTable for Regional Sales by Product
(sum of CustID is meaningless)



Creating a PivotTable

Pivot Table Tools

Options

Active Field

Field Settings

- Change summarization method in *Value Field Settings* dialog box
- Select Count

| | A | B | C | D |
|---|------------------|---------------|-----|-------------|
| 1 | | | | |
| 2 | | | | |
| 3 | Count of Cust ID | Column Labels | | |
| 4 | Row Labels | Book | DVD | Grand Total |
| 5 | East | 56 | 42 | 98 |
| 6 | North | 43 | 42 | 85 |
| 7 | South | 62 | 37 | 99 |
| 8 | West | 100 | 90 | 190 |
| 9 | Grand Total | 261 | 211 | 472 |

| Region | Book | DVD | Total |
|--------|------|-----|-------|
| East | 56 | 42 | 98 |
| North | 43 | 42 | 85 |
| South | 62 | 37 | 99 |
| West | 100 | 90 | 190 |
| Total | 261 | 211 | 472 |

Creating a PivotTable

- PivotTable for Count of Regional Sales by Product
- PivotTable results match those shown earlier in Table 3.1

The screenshot shows a Microsoft Excel spreadsheet titled "Sales Transactions cross tab and pivot table working ...". The PivotTable Tools ribbon tab is selected. The PivotTable Field List pane on the right lists fields: Cust ID, Region, Payment, Transaction Code, Source, Amount, Product, and Time Of Day. Under "Choose fields to add to report:", Cust ID, Region, Source, and Product are checked. The Report Filter pane below shows Product selected under Column Labels. The Row Labels pane shows Region and Source selected. The Values pane shows Count of Cust ID selected. The main table on the left displays sales data categorized by Region (East, North, South, West) and Order Source (Email, Web). The table includes sub-totals for each source within a region and a grand total for each region.

| | A | B | C | D |
|----|------------------|---------------|-----|-------------|
| 1 | | | | |
| 2 | | | | |
| 3 | Count of Cust ID | Column Labels | | |
| 4 | Row Labels | Book | DVD | Grand Total |
| 5 | East | 56 | 42 | 98 |
| 6 | Email | 18 | 6 | 24 |
| 7 | Web | 38 | 36 | 74 |
| 8 | North | 43 | 42 | 85 |
| 9 | Email | 12 | 13 | 25 |
| 10 | Web | 31 | 29 | 60 |
| 11 | South | 62 | 37 | 99 |
| 12 | Email | 20 | 10 | 30 |
| 13 | Web | 42 | 27 | 69 |
| 14 | West | 100 | 90 | 190 |
| 15 | Email | 29 | 21 | 50 |
| 16 | Web | 71 | 69 | 140 |
| 17 | Grand Total | 261 | 211 | 472 |
| 18 | | | | |
| 19 | | | | |
| 20 | | | | |
| 21 | | | | |
| 22 | | | | |
| 23 | | | | |
| 24 | | | | |
| 25 | | | | |
| 26 | | | | |
| 27 | | | | |
| 28 | | | | |
| 29 | | | | |
| 30 | | | | |
| 31 | | | | |

Creating a PivotTable

- Drag Source into the Row Labels box.
- PivotTable for Sales by Region, Product, and Order Source

The screenshot shows a Microsoft Excel spreadsheet titled "Sales Transactions cross tab and pivot table working file M". The PivotTable Tools ribbon tab is selected. The PivotTable Field List pane is open, showing fields available to add to the report: Cust ID, Region, Payment, Transaction Code, Source, Amount, Product, and Time Of Day. The "Payment" field is checked and highlighted. The main PivotTable area displays sales data categorized by Region (East, North, South, West) and Product (Book, DVD). The data includes Row Labels, Column Labels, and Values (Grand Total). The formula bar shows "B5" and "56". The status bar at the bottom indicates "Ready".

| | A | B | C | D | E |
|----|------------------|---------------|-----|-------------|-----|
| 1 | Payment | (All) | | | |
| 2 | | | | | |
| 3 | Count of Cust ID | Column Labels | | | |
| 4 | Row Labels | Book | DVD | Grand Total | |
| 5 | East | | 56 | 42 | 98 |
| 6 | North | | 43 | 42 | 85 |
| 7 | South | | 62 | 37 | 99 |
| 8 | West | | 100 | 90 | 190 |
| 9 | Grand Total | | 261 | 211 | 472 |
| 10 | | | | | |
| 11 | | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |
| 16 | | | | | |
| 17 | | | | | |
| 18 | | | | | |
| 19 | | | | | |
| 20 | | | | | |
| 21 | | | | | |
| 22 | | | | | |
| 23 | | | | | |
| 24 | | | | | |
| 25 | | | | | |
| 26 | | | | | |

Using the Pivot Table Report Filter

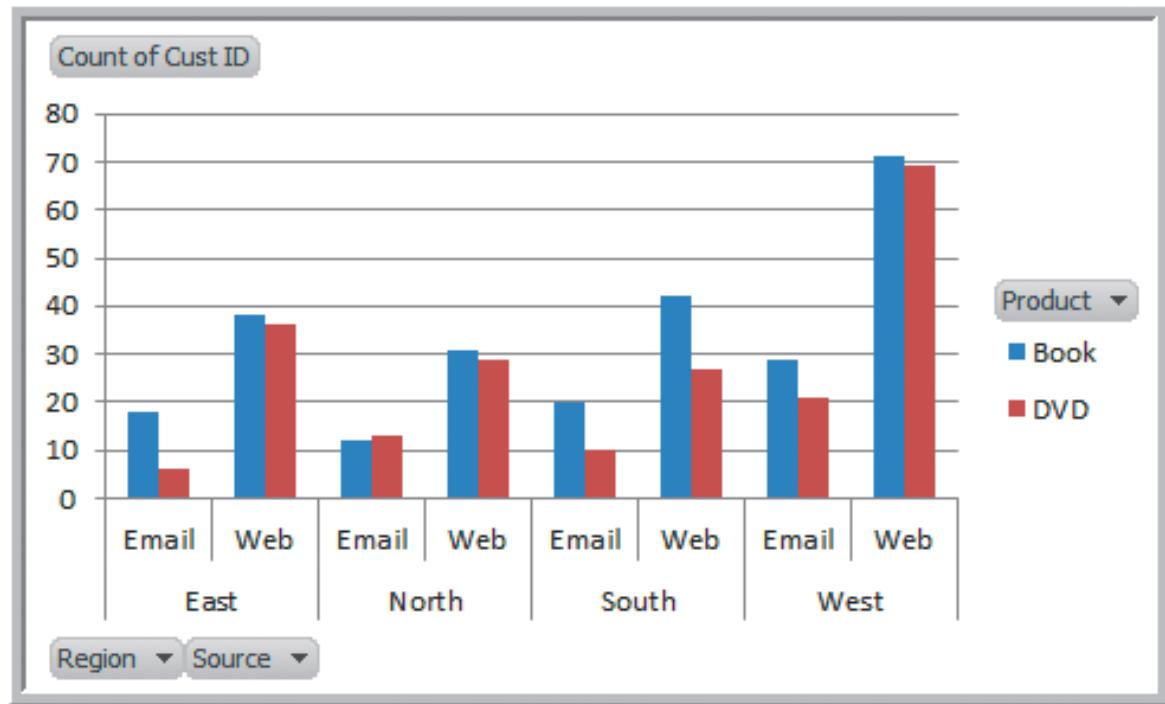
- Drag Payment into Report Filter box.
- PivotTable Filtered by Payment Type.

Using the PivotTable Report Filter

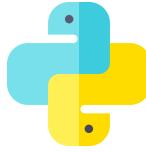
| | A | B | C | D |
|---|------------------|---------------|-----|-------------|
| 1 | Payment | Credit | | |
| 2 | | | | |
| 3 | Count of Cust ID | Column Labels | | |
| 4 | Row Labels | Book | DVD | Grand Total |
| 5 | East | 40 | 34 | 74 |
| 6 | North | 21 | 29 | 50 |
| 7 | South | 44 | 17 | 61 |
| 8 | West | 54 | 60 | 114 |
| 9 | Grand Total | 159 | 140 | 299 |

- Click on the drop-down arrow in row 1.
- Choose Credit-Card.
- Obtain this cross-tabulation PivotTable for credit card transactions.

A PivotChart for Sales Data

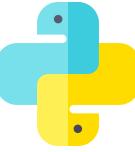


- Create a chart using the PivotTable for Sales by Region, Product, and Order Source.
 - *Insert*
 - *Column Chart*
- To display only Book data, click on the *Product* button and deselect DVD.



Key Terms

- Area chart
- Bar chart
- Bubble chart
- Column chart
- Contingency table
- Cross-tabulation
- Cumulative relative frequency
- Cumulative relative frequency distribution
- Data profile (fractile)
- Descriptive statistics
- Doughnut chart
- Frequency distribution
- Histogram
- k^{th} percentile
- Line chart
- Ogive
- Pareto analysis
- Pie chart



Key Terms

- PivotChart
- PivotTable
- Quartile
- Radar chart
- Relative frequency
- Relative frequency distribution
- Scatter chart
- Statistic
- Statistics
- Stock chart
- Surface chart