

IIT KHARAGPUR



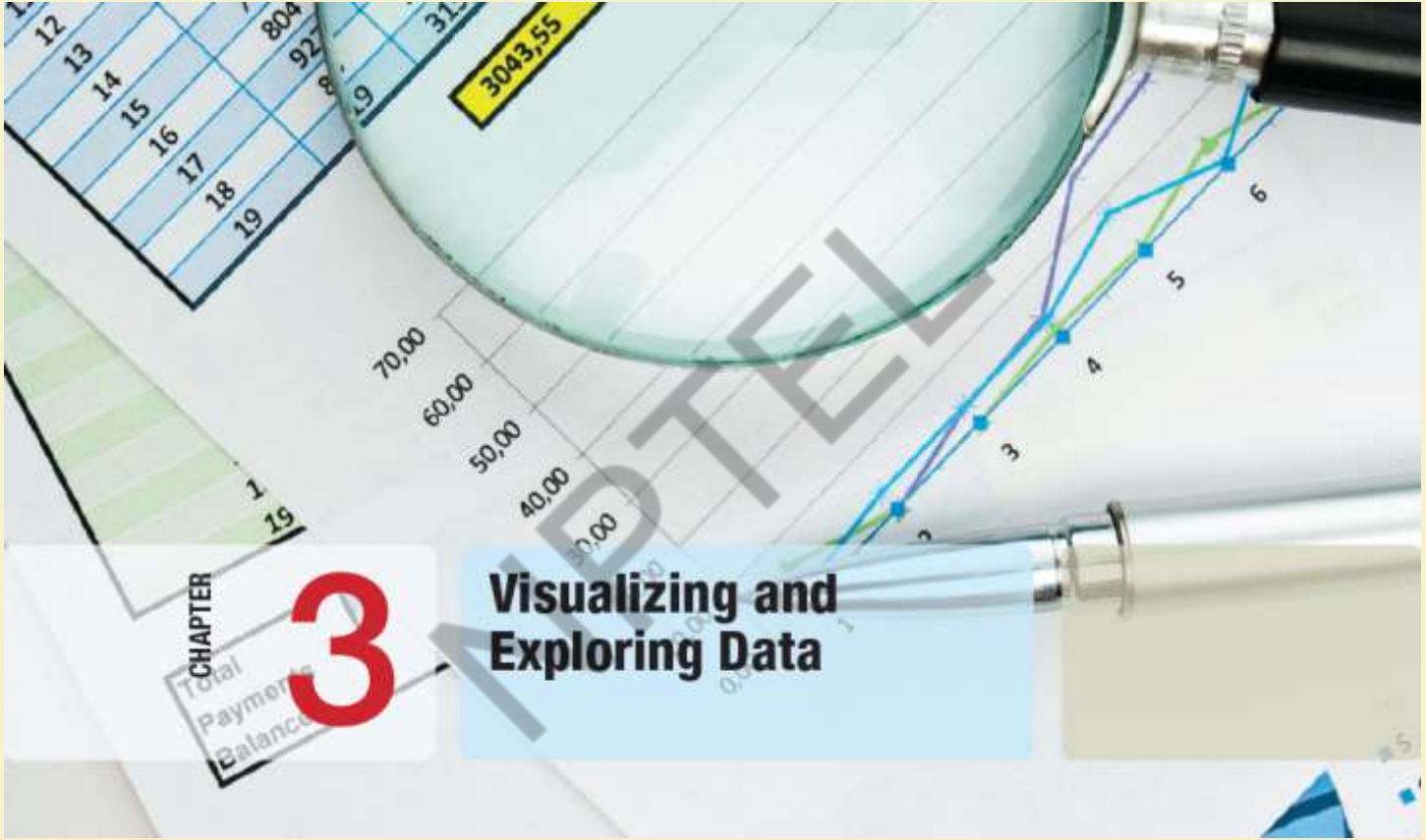
NPTEL ONLINE
CERTIFICATION COURSES

Business Analytics for Management Decision

Chapter 3 Visualizing and Exploring Data

Rudra P. Pradhan
Vinod Gupta School of Management
IIT KHARAGPUR

NPTEL



NPTEL

Chapter 3 Topics

- Data Visualization
- Data Queries: Using Sorting and Filtering
- Statistical Methods for Summarizing Data
- Exploring Data Using PivotTables

NPTEL

Data Visualization

Creating Charts in Microsoft Excel

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

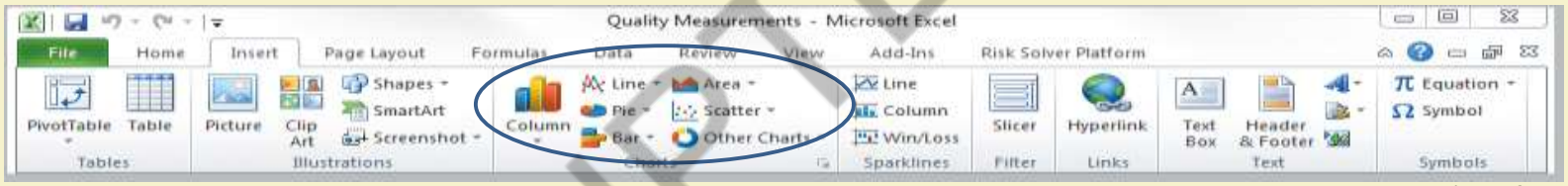


Figure 3.1

- Use chart tools to customize.

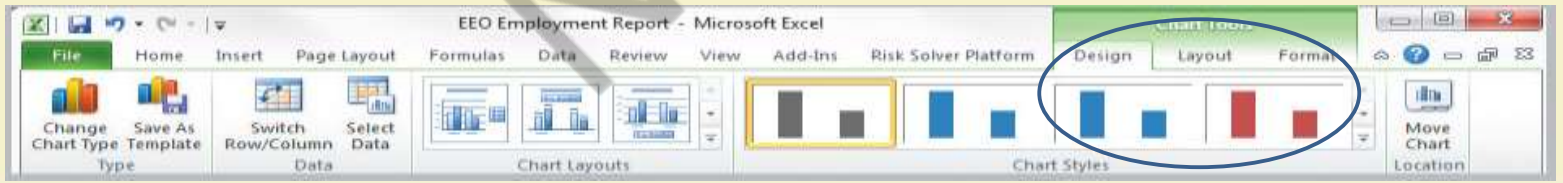


Figure 3.2

NPTEL

Data Visualization

Example 3.1 Creating a Column Chart

Highlighted Cells

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

NPTEL

Figure 3.3

Data Visualization

Example 3.1 (continued) Creating a Column Chart

Choose column chart (clustered or stacked).

Add chart title (Alabama Employment).

Rename Series1, Series2, and Series3 (ALL EMPLOYEES, Men, Women).

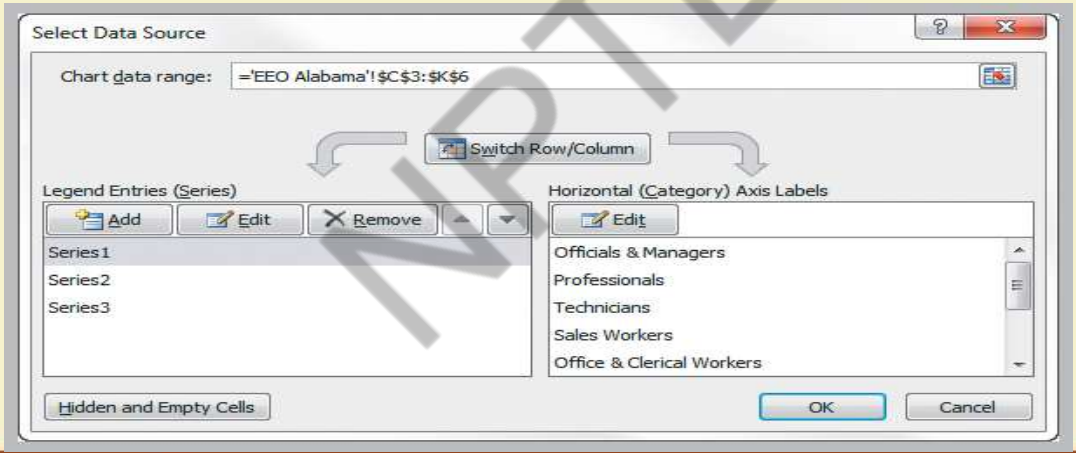
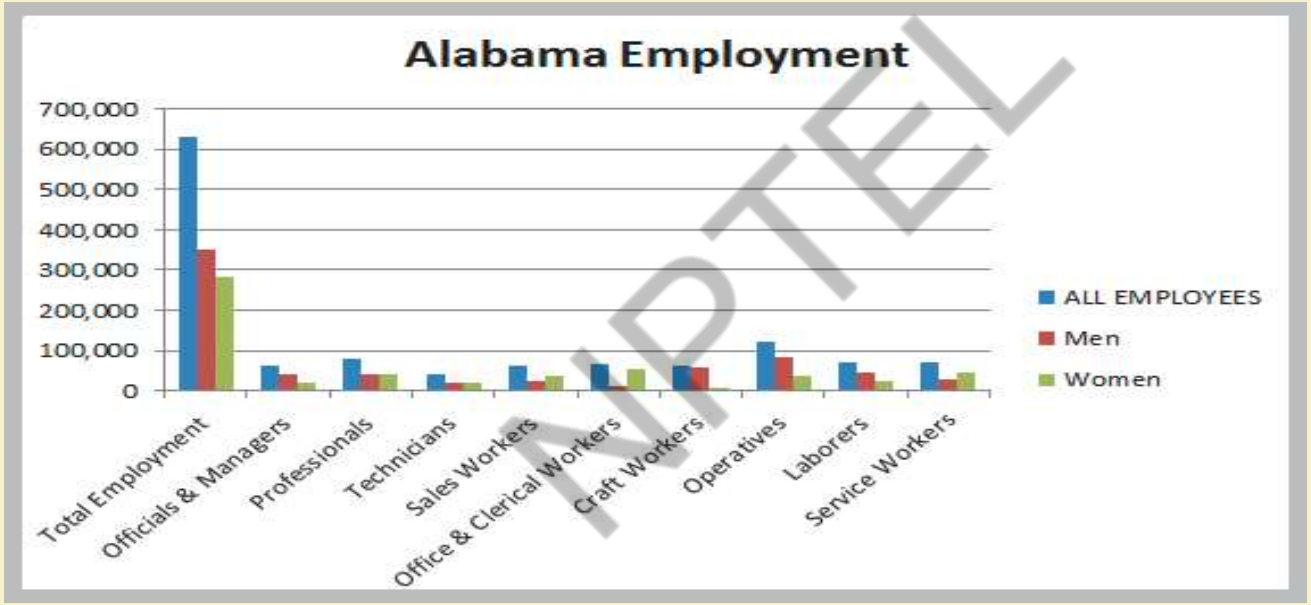


Figure 3.4

NPTEL

Data Visualization

Example 3.1 (continued) Creating a Column Chart



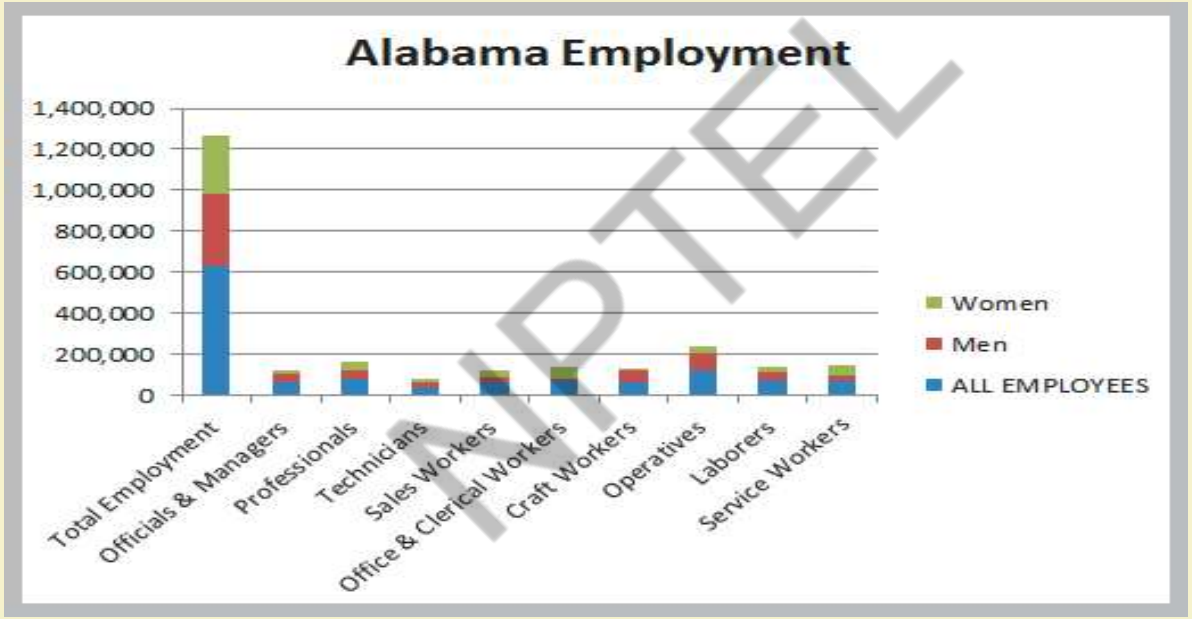
Clustered Column Chart

Figure 3.5

NPTEL

Data Visualization

Example 3.1 (continued) Creating a Column Chart



Stacked Column Chart

Figure 3.6

NPTEL

Data Visualization

Example 3.2 Line Chart for U.S. Exports to China

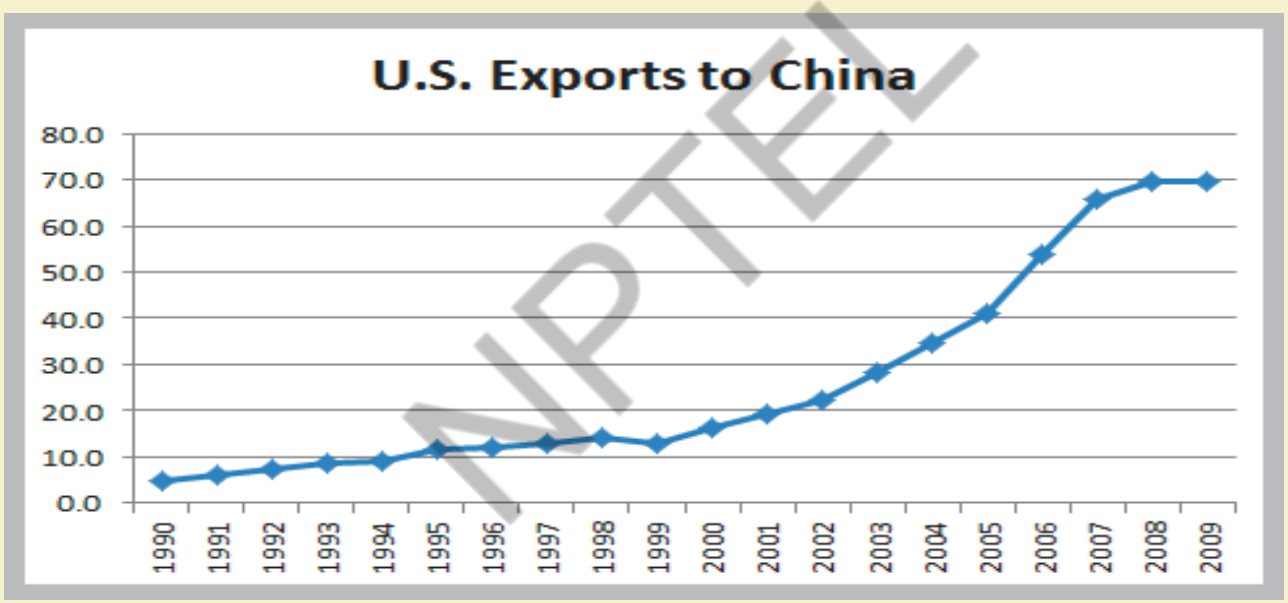


Figure 3.7

NPTEL

Data Visualization

Example 3.3 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

Figure 3.8

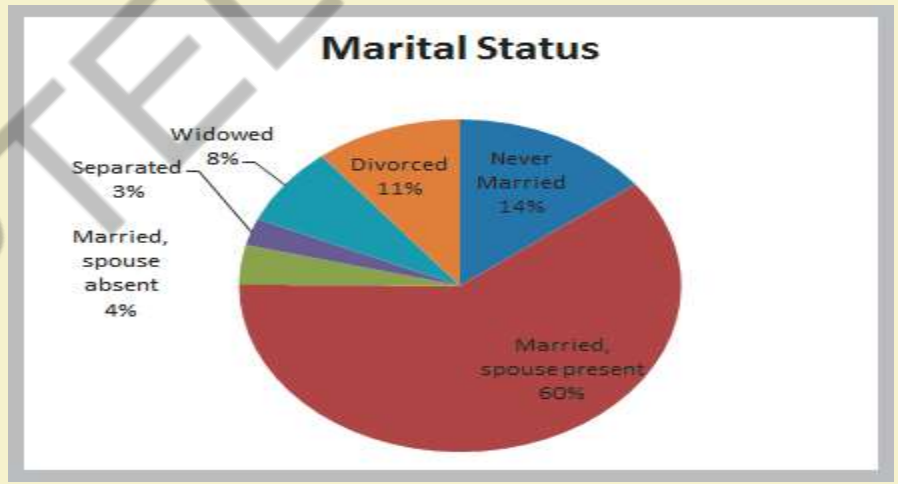


Figure 3.9

NPTEL

Data Visualization

Example 3.4 Area Chart for Energy Consumption

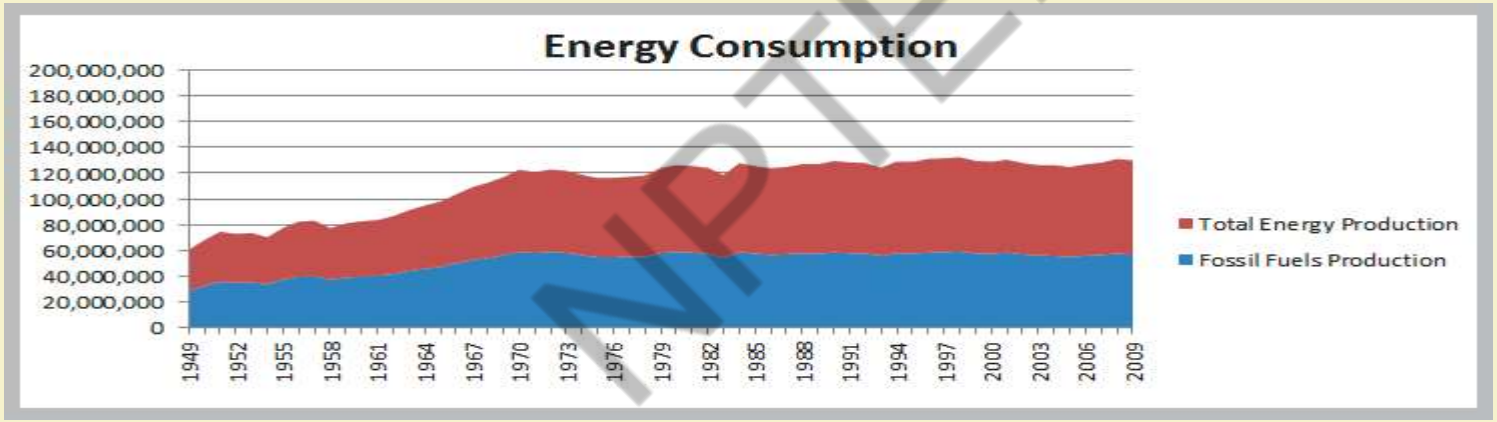


Figure 3.10

NPTEL

Data Visualization

Example 3.5 Scatter Chart for Real Estate Data

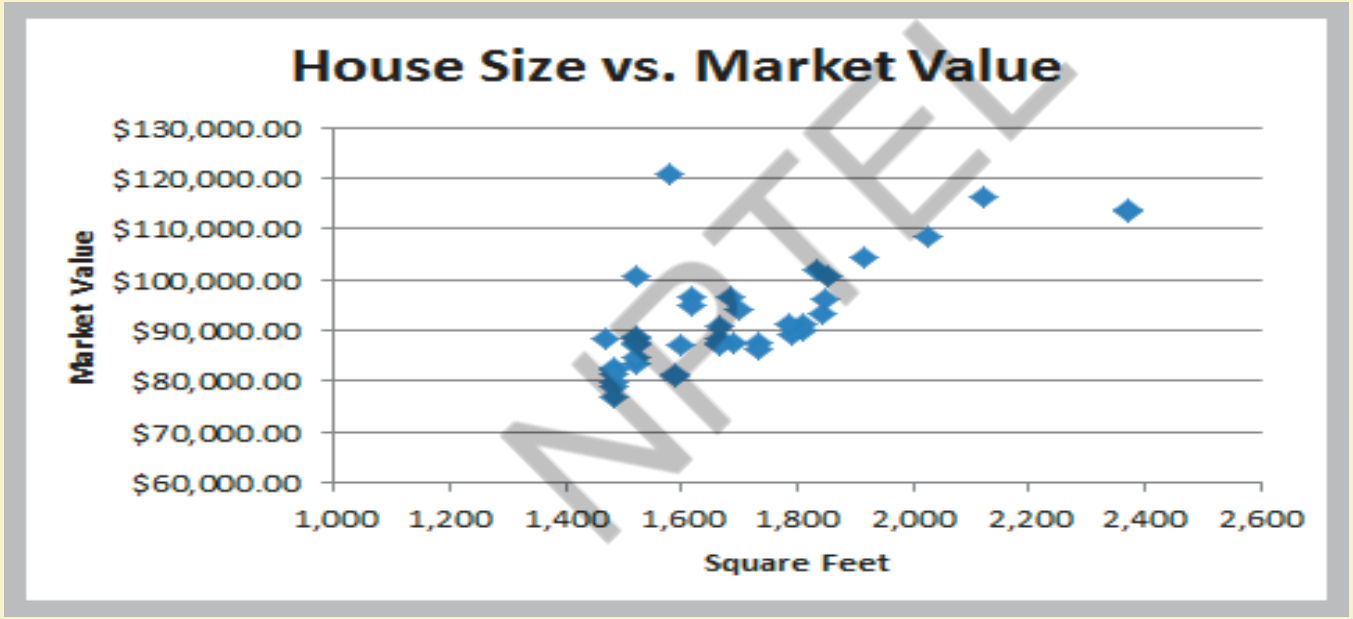


Figure 3.11

Data Visualization

Example 3.6

Bubble Chart for Comparing Stock Characteristics



Figure 3.12

NPTEL

Data Visualization

Miscellaneous Excel Charts

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

NPTEL

Data Queries: Using Sorting and Filtering

Example 3.7

Sorting Data in the Purchase Orders Database



Sort by Supplier

Figure 3.13

	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	Durrable Products	A1234	9399	Gasket	$ 3.65	1,250	$ 4,562.50	45	10/01/11	10/06/11
13	Durrable Products	A1235	9399	Gasket	$ 3.65	1,450	$ 5,292.50	45	10/03/11	10/08/11
14	Durrable Products	A1344	5454	Control Panel	$ 220.00	550	$ 121,000.00	45	10/09/11	10/14/11
15	Durrable Products	A1345	9399	Gasket	$ 3.65	1,470	$ 5,365.50	45	10/07/11	10/12/11
16	Durrable Products	A1346	9399	Gasket	$ 3.65	1,985	$ 7,245.25	45	10/05/11	10/11/11

Figure 3.14

NPTEL

Data Queries: Using Sorting and Filtering

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.

NPTEL

Data Queries: Using Sorting and Filtering

Example 3.8 Applying the Pareto Principle

Sort by

	A	B	C	D	E	F	G	H	I
1	Bicycle Inventory								
2									
3	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		

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

Figure 3.15

NPTEL

Data Queries: Using Sorting and Filtering

Example 3.9 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.

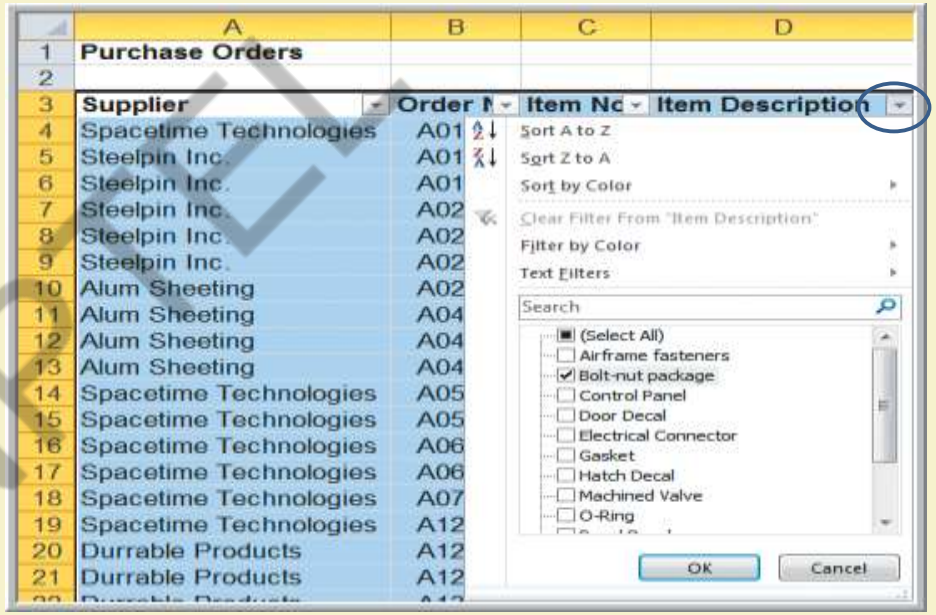


Figure 3.16

Data Queries: Using Sorting and Filtering

Example 3.9 (continued)

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										
3	Supplier	Order N	Item Nc	Item Description	Item Cc	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	Durrable 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	09/20/11
43	Steelpin Inc.	B0445	4312	Bolt-nut package	$ 3.75	4,150	$ 15,562.50	30	09/03/11	09/11/11

Figure 3.17

Data Queries: Using Sorting and Filtering

Example 3.10 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…

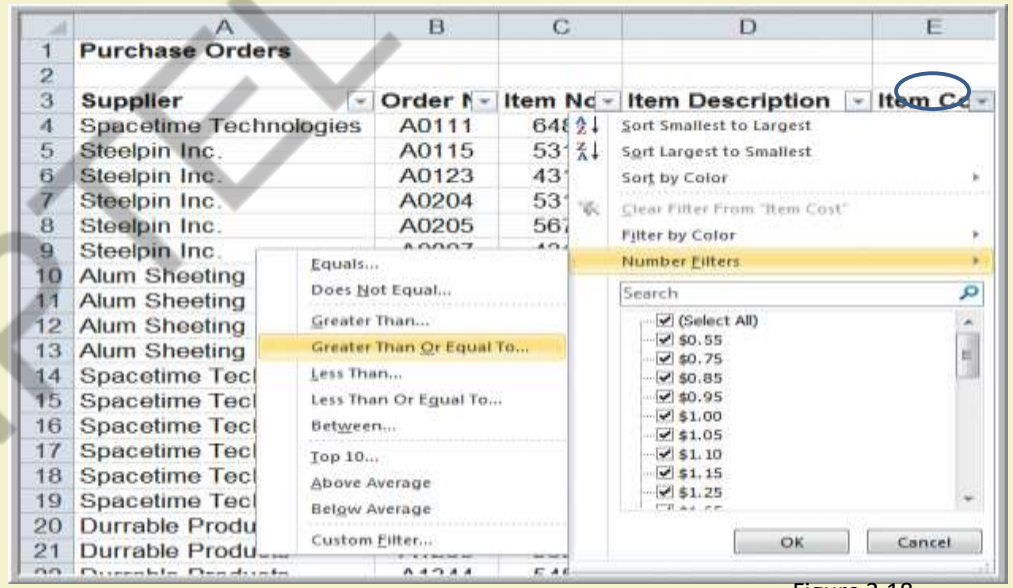


Figure 3.18

NPTEL

Data Queries: Using Sorting and Filtering

Example 3.10 (continued) Filtering by Item Cost

Custom *AutoFilter* dialog box

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

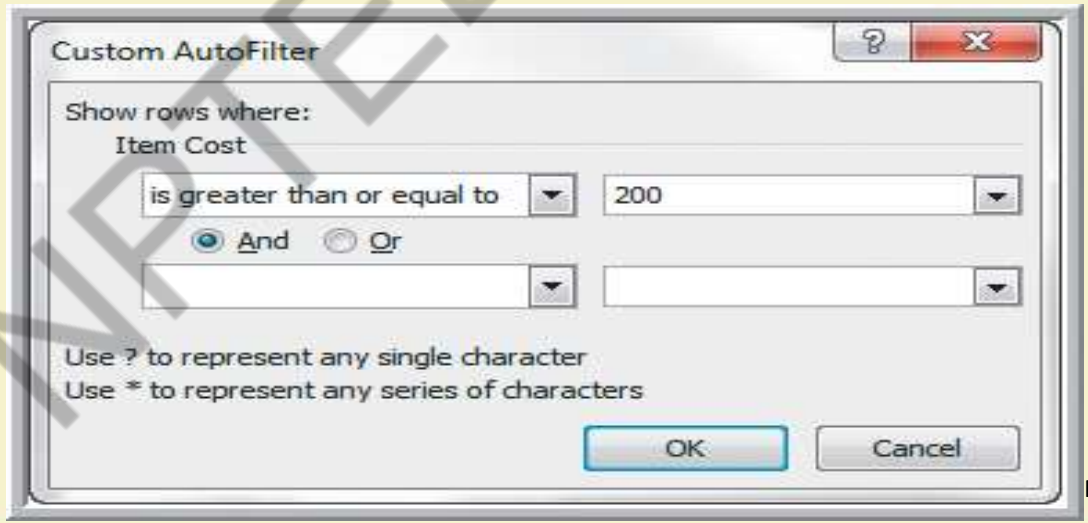


Figure 3.19

NPTEL

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.

NPTEL

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.

NPTEL

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)

Statistical methods are essential to Business Analytics

NPTEL

Statistical Methods for Summarizing Data

Example 3.11 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

Figure 3.20

Statistical Methods for Summarizing Data

Example 3.11 (continued) 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)

Figure 3.21

	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

Figure 3.22

Statistical Methods for Summarizing Data

Example 3.11 (continued) Constructing a Frequency Distribution for Items in the Purchase Order Database

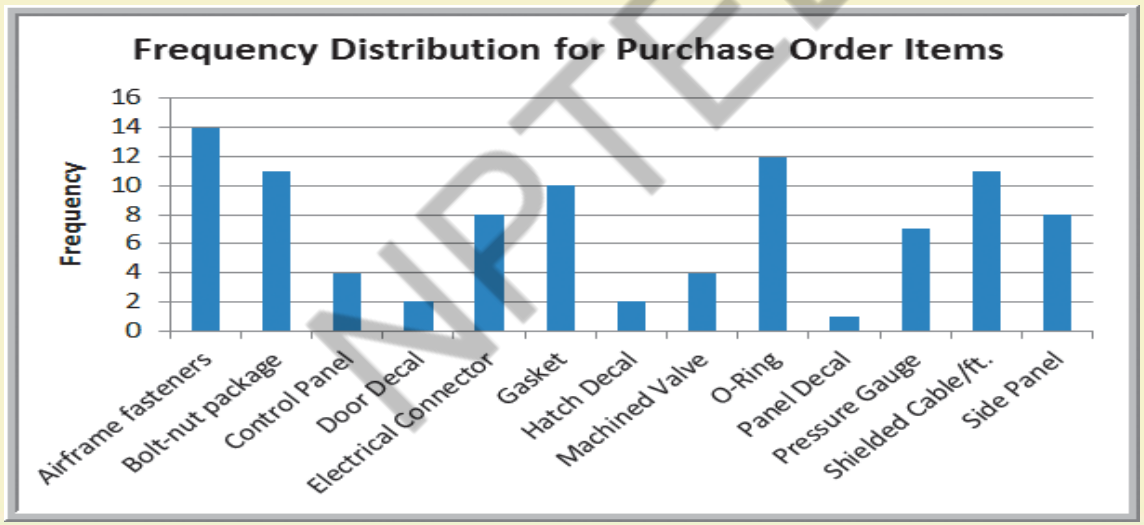


Figure 3.23

NPTEL

Statistical Methods for Summarizing Data

Example 3.12 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

Figure 3.24

NPTEL

Statistical Methods for Summarizing Data

Example 3.13 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

Figure 3.25

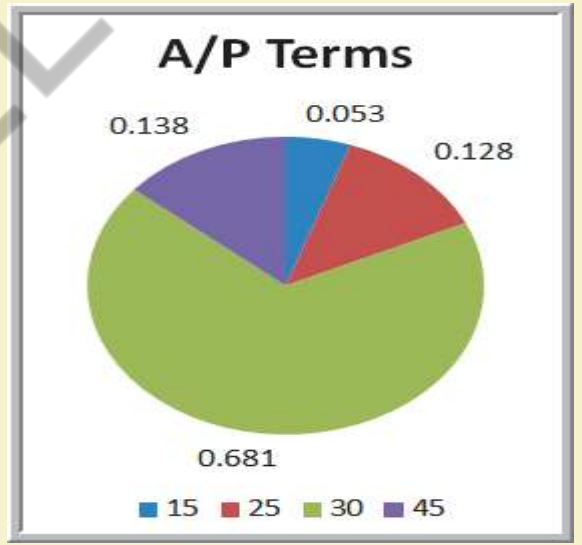


Figure 3.26

NPTEL

Statistical Methods for Summarizing Data

Excel's Histogram Tool

Using the *Analysis Toolpak*

Data

Data Analysis

Histogram

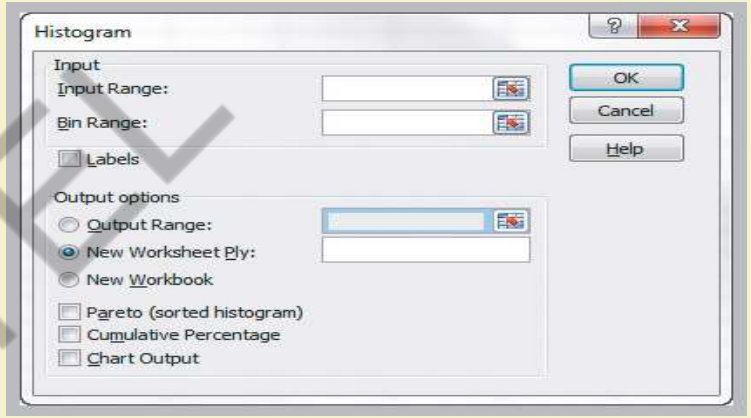


Figure 3.27

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

Choose Labels if columns have headers rows.

Choose *Chart Output.*

NPTEL

Statistical Methods for Summarizing Data

Example 3.14
Using the Histogram Tool for A/P Terms

A/P data in H3:H97

Bins below in H99:H103

Month
15
25
30
45



Figure 3.28

NPTEL

Statistical Methods for Summarizing Data

Example 3.14 (continued)

Using the Histogram Tool for A/P Terms

	A	B
1	*Months*	*Frequency*
2	15	5
3	25	12
4	30	64
5	45	13
6	More	0

Table above is not linked to chart.

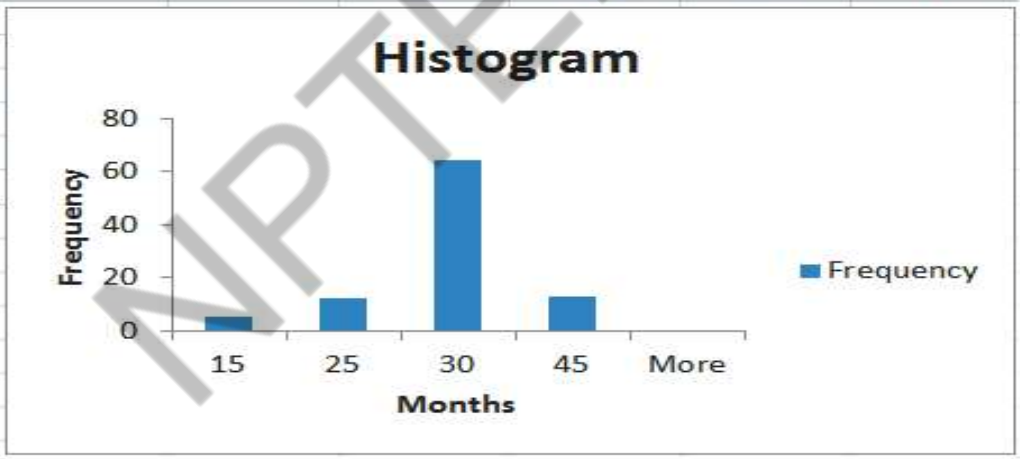


Figure 3.29

NPTEL

Statistical Methods for Summarizing Data

Example 3.15 Constructing a Frequency Distribution and Histogram for Cost Per Order

	A	B
1	*Upper Group Limit*	*Frequency*
2	0	0
3	26000	68
4	52000	8
5	78000	11
6	104000	4
7	130000	3
8	More	0

5 groups with a $26,000 group width

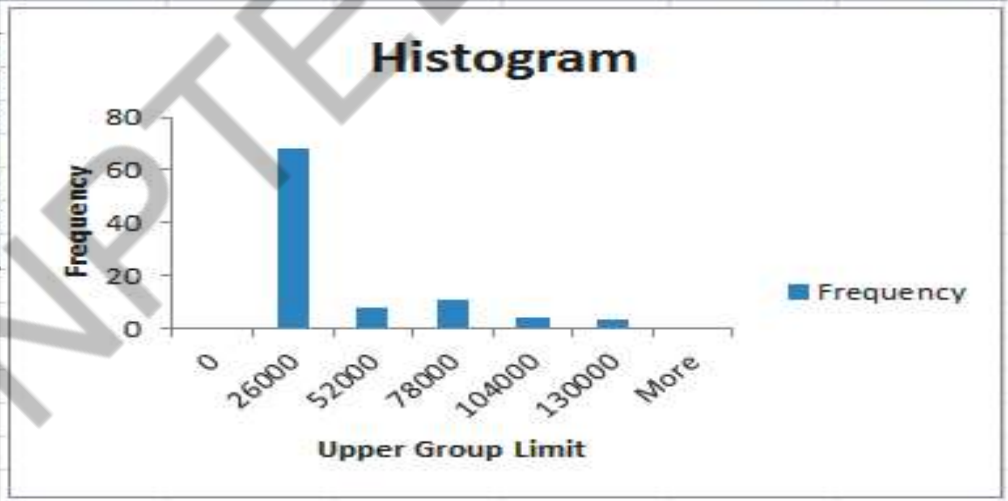


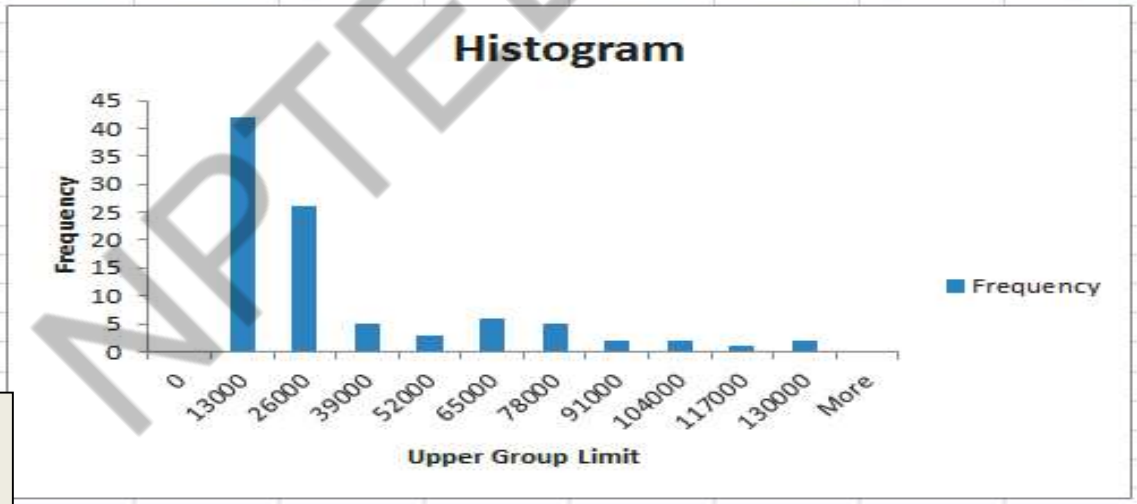
Figure 3.30

NPTEL

Statistical Methods for Summarizing Data

Example 3.15 (continued) Constructing a Frequency Distribution and Histogram for Cost Per Order

	A	B
1	*Upper Group Limit*	*Frequency*
2	0	0
3	13000	42
4	26000	26
5	39000	5
6	52000	3
7	65000	6
8	78000	5
9	91000	2
10	104000	2
11	117000	1
12	130000	2
13	More	0



10 groups with a $13,000 group width

NPTEL

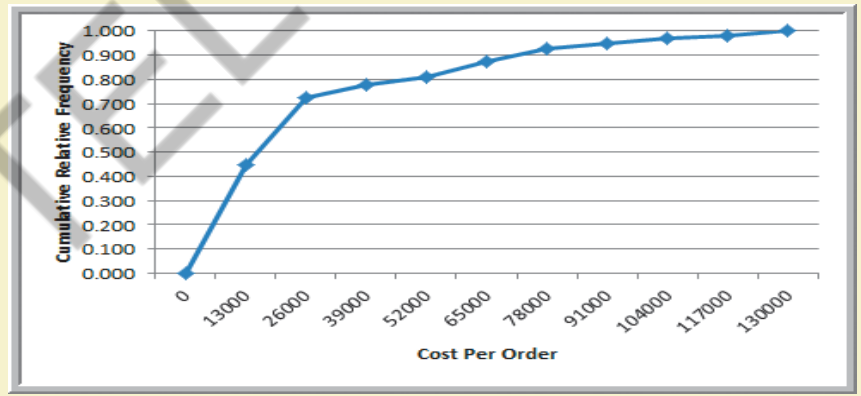
Figure 3.31

Statistical Methods for Summarizing Data

Example 3.16 Computing Cumulative Relative Frequencies for the Cost Per Order 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		

Figure 3.32



Ogive

Figure 3.33

NPTEL

Statistical Methods for Summarizing Data

Example 3.17 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

NPTEL

Statistical Methods for Summarizing Data

Example 3.18 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.

NPTEL

Statistical Methods for Summarizing Data

Example 3.19 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%

Figure 3.34

NPTEL

Statistical Methods for Summarizing Data

Example 3.20 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

NPTEL

Statistical Methods for Summarizing Data

Example 3.21 Constructing a Cross-Tabulation

- *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

Figure 3.35

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

NPTEL

Statistical Methods for Summarizing Data

- Example 3.21 (continued) Constructing a Cross-Tabulation

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 3.1

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%

Table 3.2

NPTEL

Statistical Methods for Summarizing Data

- Example 3.21 (continued) Constructing a Cross-Tabulation



Figure 3.36

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 3.1

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

NPTEL

Exploring Data Using PivotTables

Data
Tables
PivotTable

Follow wizard steps.

PivotTables allow:

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

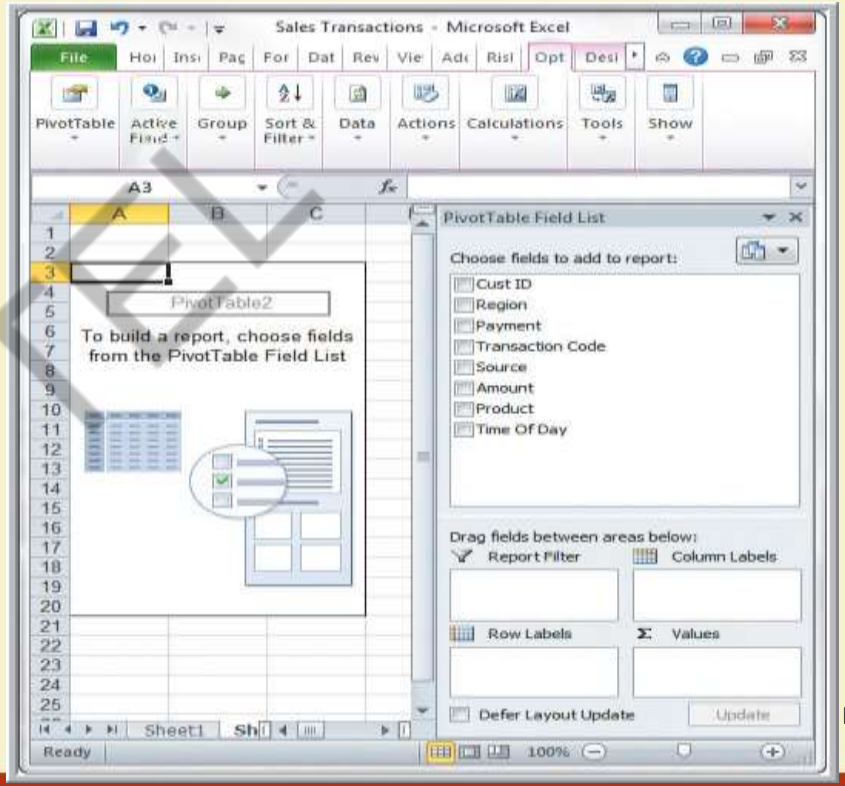


Figure 3.37

NPTEL

Exploring Data Using PivotTables

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.

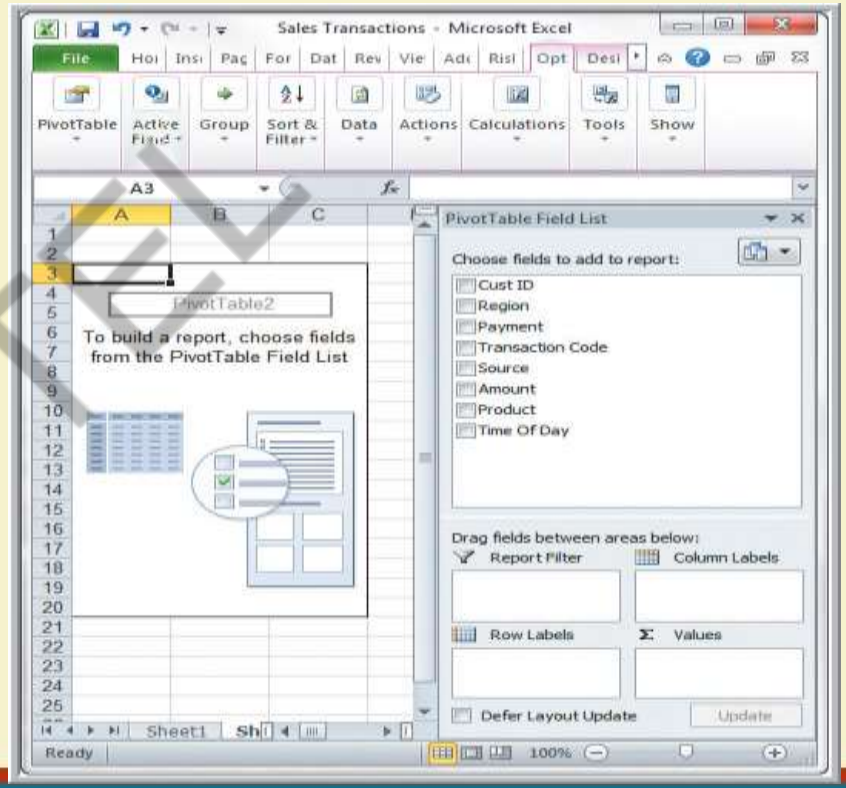


Figure 3.37

NPTEL

Exploring Data Using PivotTables

Example 3.22

Creating a PivotTable

Default PivotTable for Regional Sales by Product

(sum of CustID is meaningless)

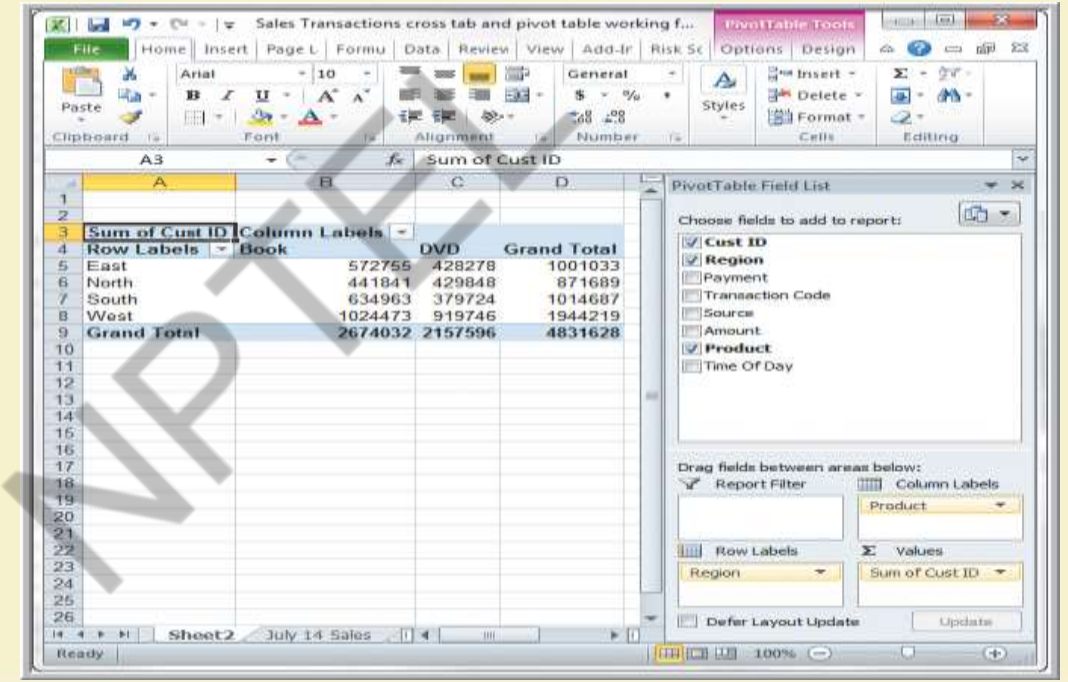


Figure 3.38

Exploring Data Using PivotTables

Example 3.22 (continued) Creating a PivotTable

Pivot Table Tools
Options
Active Field
Field Settings

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

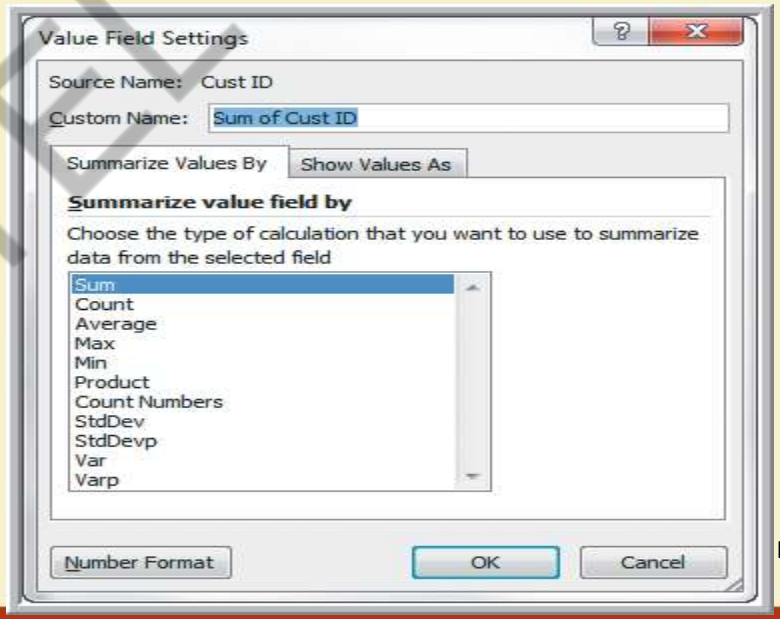


Figure 3.39

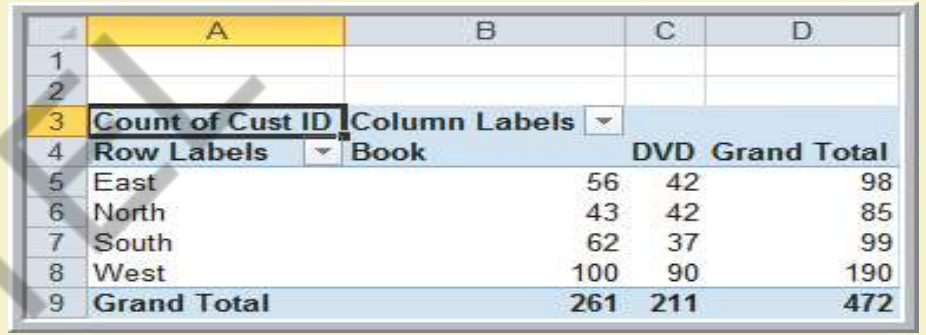
NPTEL

Exploring Data Using PivotTables

Example 3.22 (continued) Creating a PivotTable

PivotTable for Count of Regional Sales by Product

PivotTable results match those shown earlier in Table



	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

Figure 3.40

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

NPTEL

Exploring Data Using PivotTables

Example 3.22 (continued)

Creating a PivotTable

Drag *Source* into the *Row Labels* box.

PivotTable for Sales by Region, Product, and Order Source

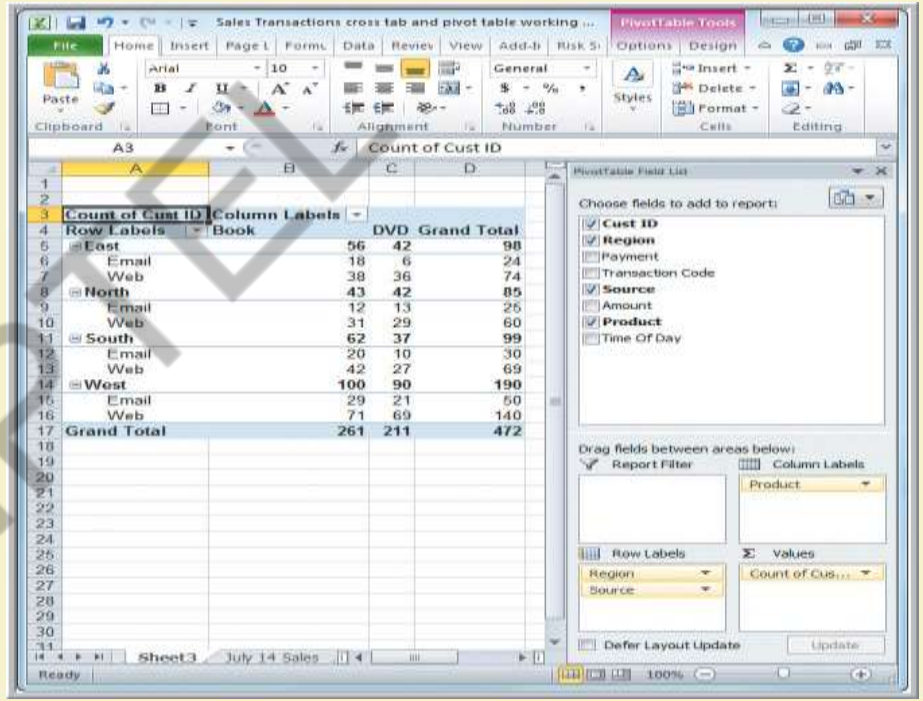


Figure 3.41

NPTEL

Exploring Data Using PivotTables

Example 3.23
Using the Pivot Table Report Filter

Drag *Payment* into *Report Filter* box.

PivotTable Filtered by Payment Type.

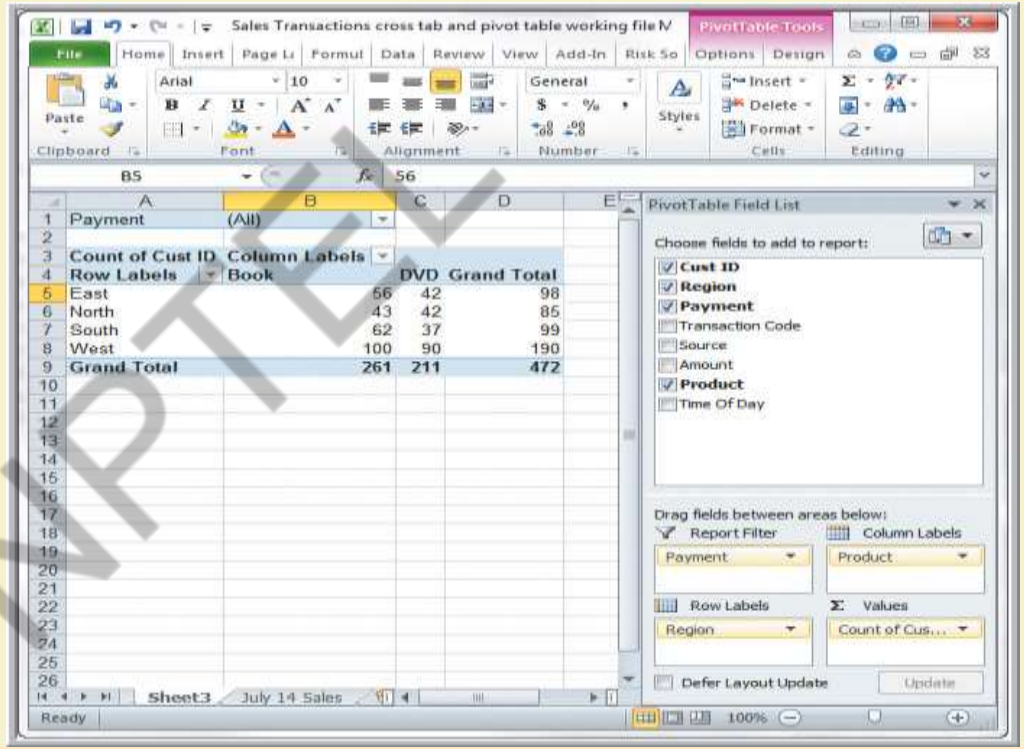


Figure 3.42

NPTEL

Exploring Data Using PivotTables

Example 3.23 (continued)

Using the PivotTable Report Filter

Click on the drop-down arrow in row 1.

Choose Credit-Card.

Obtain this cross-tabulation PivotTable for credit card transactions.

	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

Figure 3.43

NPTEL

Exploring Data Using PivotTables

Example 3.24 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.

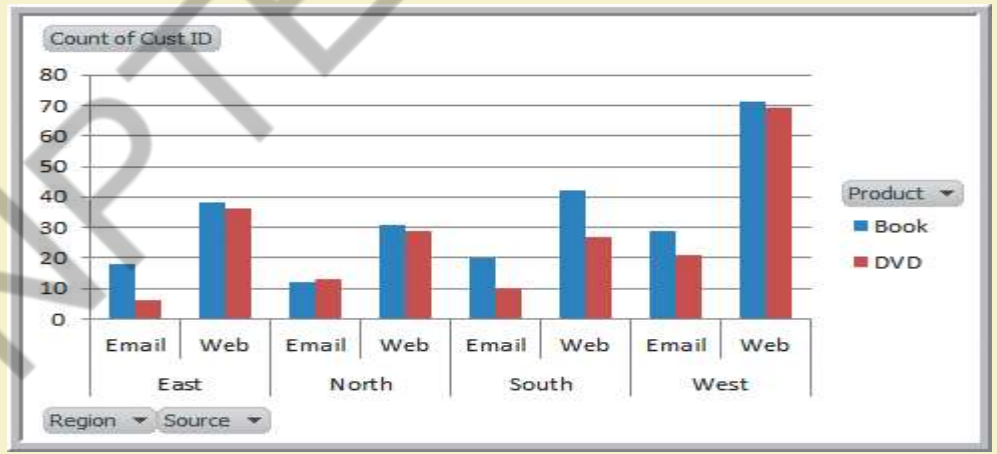


Figure 3.44

NPTEL

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

NPTEL

Chapter 3 - Key Terms (continued)

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

NPTEL

Case Study Performance Lawn Equipment (3)

- Recall that PLE produces lawnmowers and a medium size diesel power lawn tractor.
- Create charts of the satisfaction data, sales data, delivery time data, and other variables of interest.
- Compare shipping costs for existing and proposed plant locations.
- Examine customer attributes by region and write a formal report summarizing your results.

NPTEL