

TECH 646 Analysis of Research in Industry and Technology

PART IV

Analysis and Presentation of Data:

Data Presentation and Description; Exploring, Displaying, and Examining Data; Hypothesis Testing; Measures of Association; Multivariate Analysis; Presenting Insights and Findings

Ch. 16 Exploring, Displaying, and Examining Data

Lecture note based on the text book and supplemental materials:
Cooper, D.R., & Schindler, P.S., *Business Research Methods* (12th edition), McGraw-Hill/Irwin

Paul I-Hai Lin, Professor

<http://www.etcs.pfw.edu/~lin>

A Core Course for M.S. in Technology Program
Purdue University Fort Wayne

1

Exploring, Displaying, and Examining Data

Learning Objectives ... Understand

- That exploratory data analysis techniques provide insights and data diagnostics by emphasizing visual representations of the data.
- How cross-tabulation is used to examine relationships involving categorical variables, serves as a framework for later statistical testing, and makes an efficient tool for data visualization and later decision-making.

2

Research as Competitive Advantages

“As data availability continues to increase, the importance of identifying/filtering and analyzing relevant data can be a powerful way to gain an information advantage over our competition.”

**Tom H.C. Anderson
founder & managing partner
Anderson Analytics, LLC**

3

Research Skill Improves Data Discovery

- Radius Global Market Research, formerly DDW, www.radius-global.com

DDW is a global player in research services. As this ad proclaims, you can “push data into a template and get the job done,” but you are unlikely to make discoveries using that process.

It's not the clay.

It's the potter.

Pushing data into a template gets the job done. But delivering breakthrough marketing insights requires a special skill. DDW provides research on a global scale, yet is hands-on to shape and customize each research engagement. Our top talent is involved at every step, offering unique research and marketing perspectives. The result? Solutions for success.

DDW Data Development **Worldwide**

Know More. Is your online research feeling uninformative? DDW's Know More Internet Panel consists of over 40 million consumers, recruited and maintained with the highest standards. Call Managing Director Chris Lerner at 972.853.1100 to discuss the opportunities and solutions to your online research.

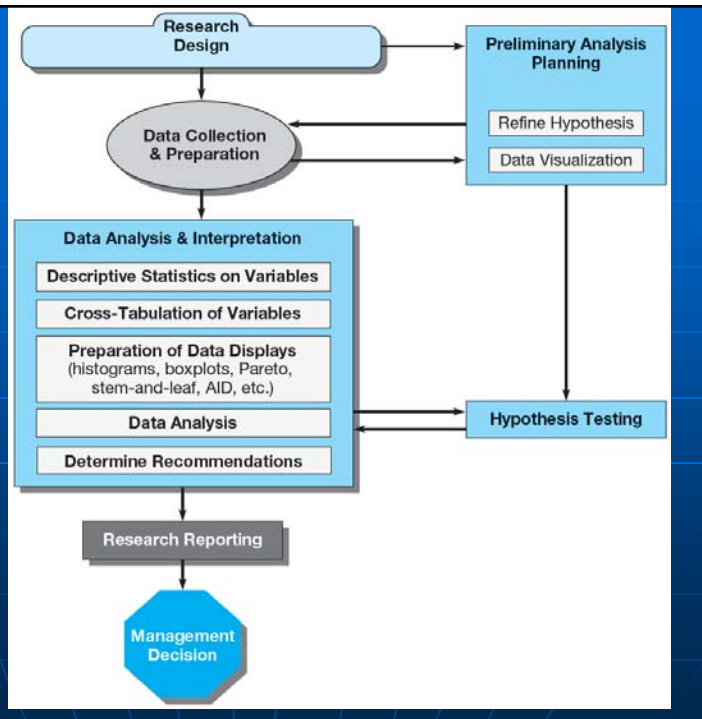
Headquarters: New York, Albany, Stamford, Chicago, Kansas City, San Francisco, Long Beach

Exploratory Data Analysis

- Data entry via
 - Spread sheet, Data editor of a statistical program, Optical mark recognition, etc.
- Exploratory Data Analysis
 - Frequency Tables (nominal variable, percentage)
 - Bar Charts, Pie Charts (nominal variable)
 - Histogram (interval-ratio data)
 - Stem-and-Leaf displays
 - Pareto Diagrams
 - Boxplots
 - Mapping

5

Data Exploration, Examination, and Analysis in the Research Process



Useful Measures

- **Measures of Performance**
 - Service time
 - Number of incorrect procedures
 - Number of warranty claims
 - Profit
 - Number of missed opportunities
- **Product Evaluation**
 - Number of sales
 - Number of return product
 - Count of reasons for return
 - Number of repeated customers

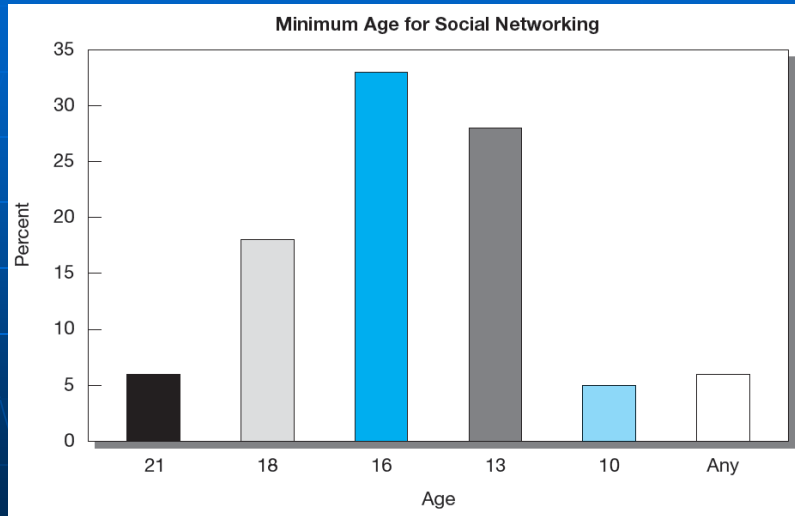
7

Exhibit 16-2 Appropriate Social Networking Age (Frequency Table)

Value Label	Value	Frequency	Percent	Valid Percent	Cumulative Percent
21 years old	1	60	6	6	6
18 years old min	2	180	18	18	24
16 years old min	3	330	33	33	57
13 years old min	4	280	28	28	85
10 years old min	5	50	5	5	90
Any age	6	60	6	6	96
No opinion	7	40	4	4	100
		1,000	100	100	

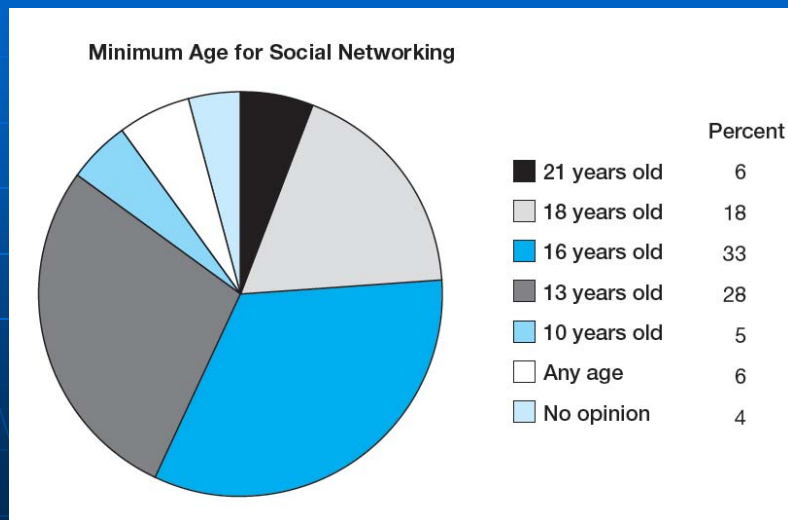
Valid Cases 1,000; Missing Cases 0

Exhibit 16-3 Bar Chart



9

Exhibit 16-3 Pie Chart



10

Exhibit 16-4 Average Annual Purchase of PrimeSell's Top 50 Customers (Frequency Table – with an ordered list of values)

Value	Frequency	Percent	Cumulative Percent	Value	Frequency	Percent	Cumulative Percent
54.9	1	2	2	75.6	1	2	54
55.4	1	2	4	76.4	1	2	56
55.6	1	2	6	77.5	1	2	58
56.4	1	2	8	78.9	1	2	60
56.8	1	2	10	80.9	1	2	62
56.9	1	2	12	82.2	1	2	64
57.8	1	2	14	82.5	1	2	66
58.1	1	2	16	86.4	1	2	68
58.2	1	2	18	88.3	1	2	70
58.3	1	2	20	102.5	1	2	72
58.5	1	2	22	104.1	1	2	74
59.9	2	4	26	110.4	1	2	76
61.5	1	2	28	111.9	1	2	78
62.6	1	2	30	118.6	1	2	80
64.8	1	2	32	123.8	1	2	82
66.0	2	4	36	131.2	1	2	84
66.3	1	2	38	140.9	1	2	86
67.6	1	2	40	146.2	1	2	88
69.1	1	2	42	153.2	1	2	90
69.2	1	2	44	163.2	1	2	92
70.5	1	2	46	166.7	1	2	94
72.7	1	2	48	183.2	1	2	96
72.9	1	2	50	206.9	1	2	98
73.5	1	2	52	218.2	1	2	100
				Total	50	100	

11

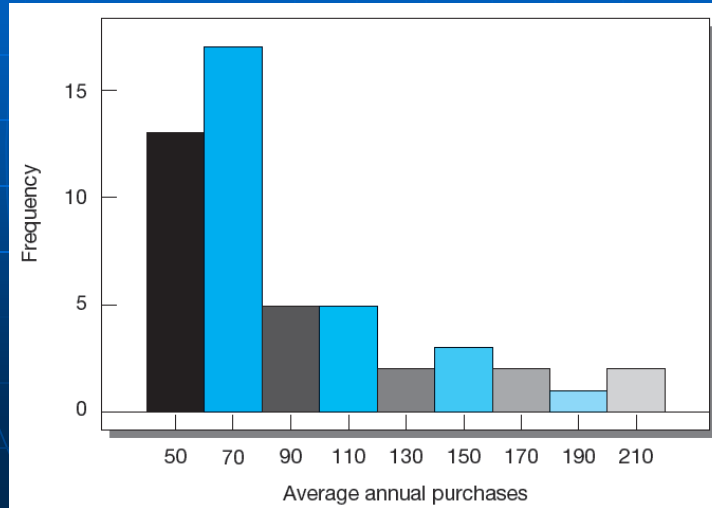
Histogram Display

- Measurement Scales:
 - Nominal
 - Ordinal
 - Interval
 - Ratio
- Display all intervals in a distribution for Skewness, Kurtosis, Modal pattern
- Single hum (a mode)
- Subgroups identifiable (multiple modes)
- Straggling/irregular data

12

Exhibit 16-5 Histogram of PrimeShell's Top Customer's Average Annual Purchases

- Ratio scale
- Continuous Interval, one bar for each interval



13

Stem-and-Leaf Displays

- Closely related to the histogram, but offers several unique advantages:
 - Presents actual data values that can be inspected directly
 - Visualization: range of values, shapes and spread, patterns
- Each row – a stem
- Each piece of information on the stem – a leaf
- Exhibit 16-6
 - 54, 55, 55, 56, 56, 56, 57, 58, 58, 58, 58, 59
 - 61, 62, 64, 66, 66, 67, 69, 69
 - etc

14

Stem-and-Leaf Displays

- Exhibit 16-6
 - 54, 55, 55, 56, 56, 56, 57, 58, 58, 58, 58, 59
 - 61, 62, 64, 66, 66, 67, 69, 69
 - First Line or Row
 - 5 || 4 5 5 6 6 6 7 8 8 8 8 9
 - Second Line or Row
 - 6 || 1 2 4 6 6 7 9 9

15

Exhibit 16-6 Stem-and-Leaf Display

5	455666788889
6	12466799
7	02235678
8	02268
9	
10	24
11	018
12	3
13	1
14	06
15	3
16	36
17	
18	3
19	
20	6
21	8

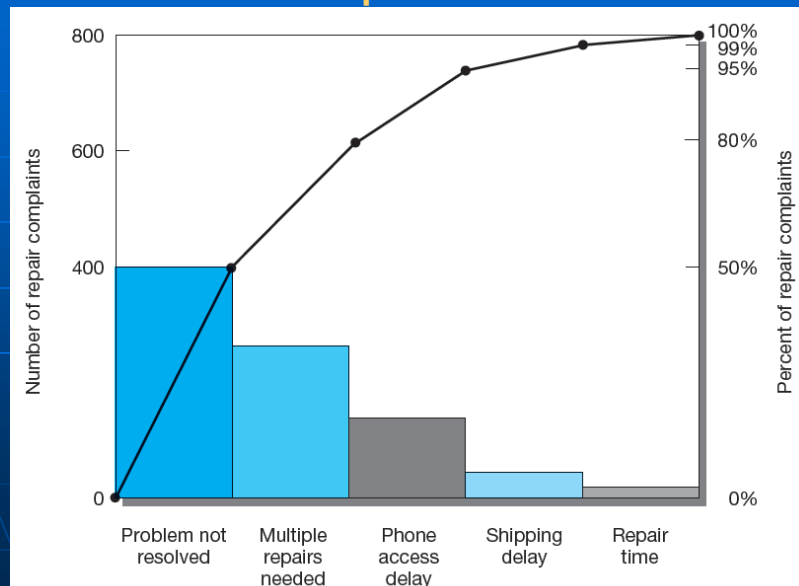
16

Pareto Diagram

- **The Pareto principle (80/20 rule)**
 - Few causes (the vital few 20%) are responsible for the majority (80%) of problems
 - Improvement benefits can be leveraged by focusing attention on the key issues (the 20%)
- **MindWriter Customer Complains – Pareto Chart or diagram (percentages sum to 100%): Occurrence Categories**
 - Problem not resolved
 - Multiple repairs needed
 - Phone access delay
 - Shipping delay
 - Repair time

17

Exhibit 16-7 Pareto Diagram of WindWriter Repair Complains



Boxplots (Box-and-Whisker)

- A graphical display that describes several features of a data set
 - Median
 - Upper and Lower quartiles
 - Largest and Smallest Observations, and
 - Outliers
- Basic Components
 - Outliers (Extreme or far outside values)
 - The rectangular plot (50% of the data values)
 - A center line (marking median)
 - The two edges – Hinge (upper and lower quartiles)
 - Smallest and Largest observed values

19

Boxplot Components

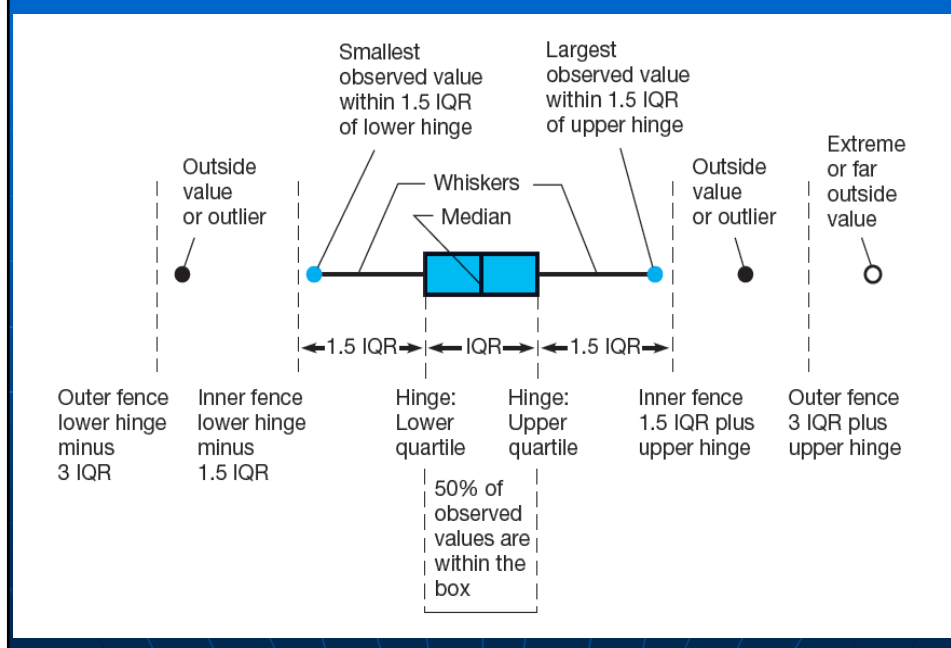
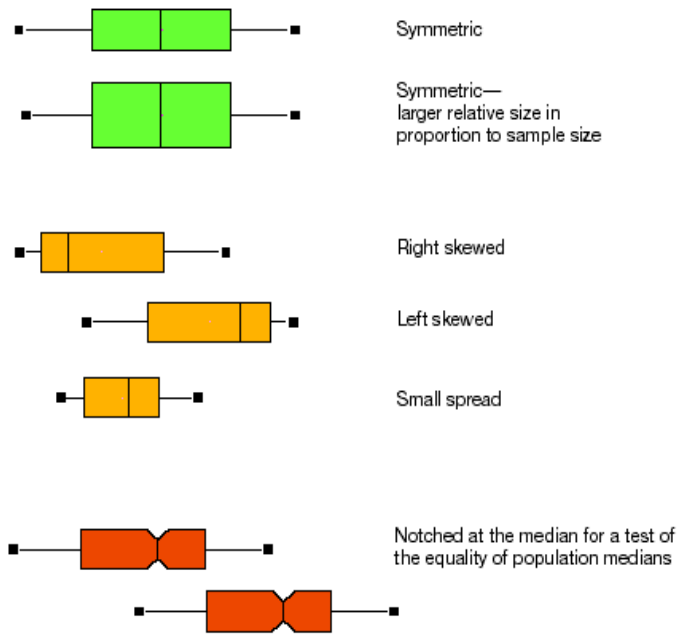
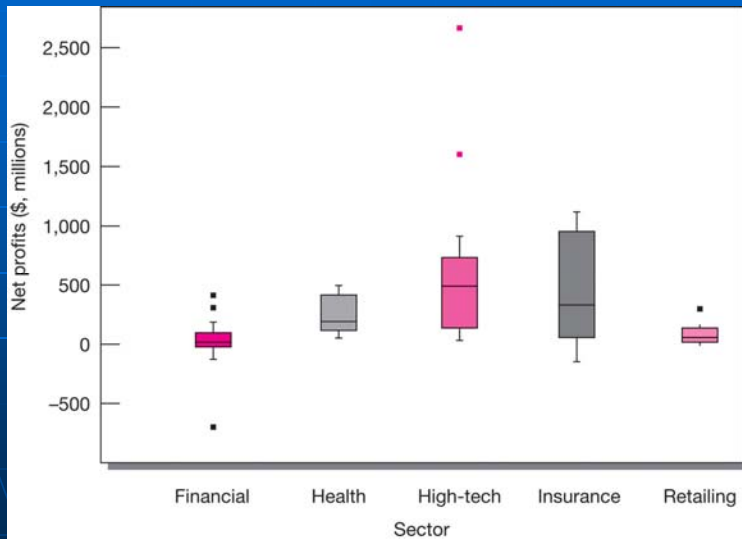


Exhibit 16-9 Diagnostics with Boxplots



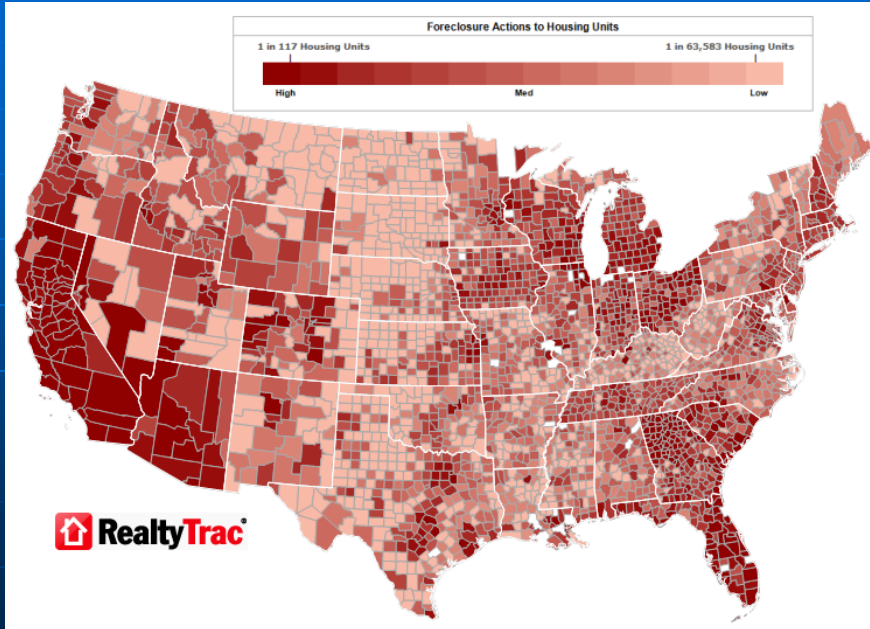
21

Exhibit 16-10 Boxplot Comparison of Customer Sectors



22

Mapping: Geographic (home foreclosure for May 2012)



SPSS Cross-Tabulation

		OVERSEAS ASSIGNMENT		Row Total
		Yes	No	
GENDER		1	2	
	Male	22 35.5 78.6 22.0	40 64.5 55.6 40.0	62 62.0
Female	2	6 15.8 21.4 6.0	32 84.2 44.4 32.0	38 38.0
Column Total		28 28.0	72 72.0	100 100.0

Cell content

Cell 2, 1 (row 2, column 1)

Marginals

Percentage in Cross-Tabulation

Study 1					Study 2								
OVERSEAS ASSIGNMENT					OVERSEAS ASSIGNMENT								
	GENDER		Count	Row Total	Count	Row Pct	No	Row Total	Count	Row Pct			
			Yes								No	Yes	No
			1								2	1	2
			Col Pct								Col Pct	Col Pct	Col Pct
			Tot Pct				Tot Pct			Tot Pct			
Male	1		22	62	225	25.0	675	900	225	25.0			
			35.5	64.5	62.5	75.0	60.0	62.5	75.0	60.0			
			78.6	55.6	15.0	59.2	45.0	15.0	59.2	45.0			
			22.0	40.0		45.0			45.0				
Female	2		6	38	135	22.5	465	600	135	22.5			
			15.8	84.2	37.5	77.5	40.0	37.5	77.5	40.0			
			21.4	44.4	9.0	40.8	31.0	9.0	40.8	31.0			
			6.0	32.0		31.0			31.0				
Column Total			28	100	360	24.0	1140	1500	360	24.0			
Total			28.0	100.0	24.0	76.0	100.0	24.0	76.0	100.0			

Guidelines for Using Percentages

Don't Average percentages

Don't use too large of a percentage

Don't use too small a base

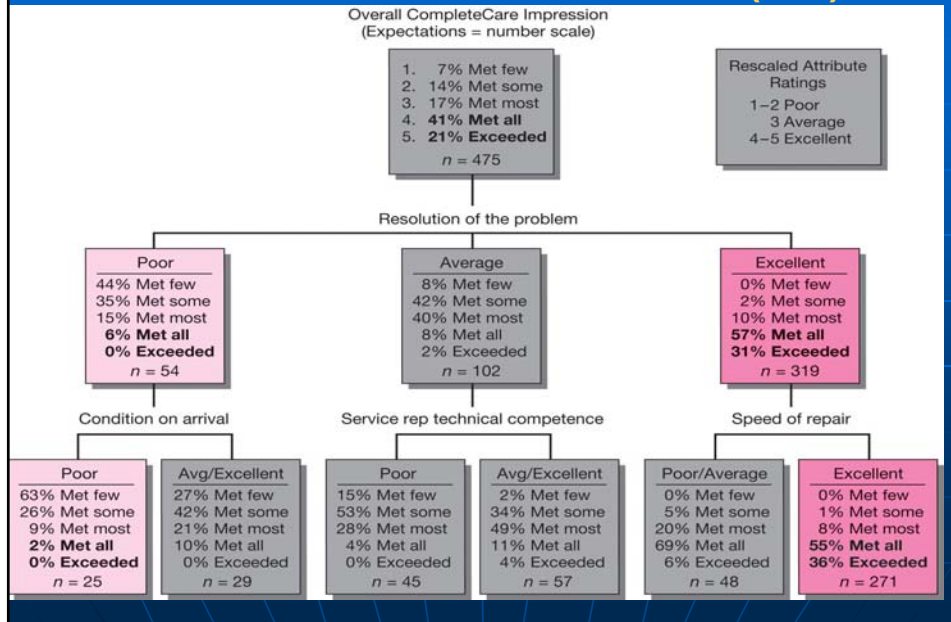
Changes should never exceed 100%

Cross-Tabulation with Control and Nested Variables

	Control Variable					
	Category 1			Category 2		
	Nested Variable		cat 3	Nested Variable		cat 3
cat 1	cat 2	cat 1		cat 2		
Stub...	Cells...					

EMPLOYMENT CATEGORY	SEX OF EMPLOYEE			
	MALES		FEMALES	
	MINORITY CLASSIFICATION		MINORITY CLASSIFICATION	
	WHITE	NONWHITE	WHITE	NONWHITE
CLERICAL	16%	7%	18%	7%
OFFICE TRAINEE	7%	3%	17%	2%
SECURITY OFFICER	3%	3%		
COLLEGE TRAINEE	7%	0%	1%	
EXEMPT EMPLOYEE	6%	0%	0%	
MBA TRAINEE	1%	0%	0%	
TECHNICAL	1%			

Automatic Interaction Detection (AID)



Exploratory Data Analysis

We have 610 pages of research that all lead to one conclusion.

You'd rather not go through 610 pages of research to get a conclusion.



Cut to the chase.

This Booth Research Services ad suggests that the researcher's role is to make sense of data displays.

Great data exploration and analysis delivers insight from data.

BRS

Booth Research Services
1-800-927-2577 / www.brsresearch.com

Snapshot: Novation



No standard vocabulary across companies

Serve variety of users

Ad hoc analysis with sophisticated visualizations

Big data with sophisticated analytical tool.

30

Snapshot: Digital Natives vs. Digital Immigrants

30 subjects = 15 natives,
15 immigrants

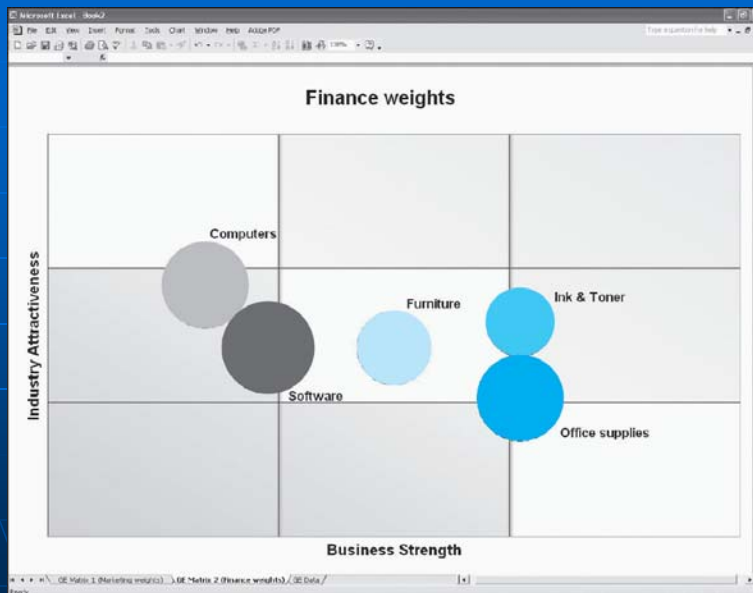
Monitored media
behaviors

300 hours of real-time
data

Biometric Monitoring:
emotional engagement

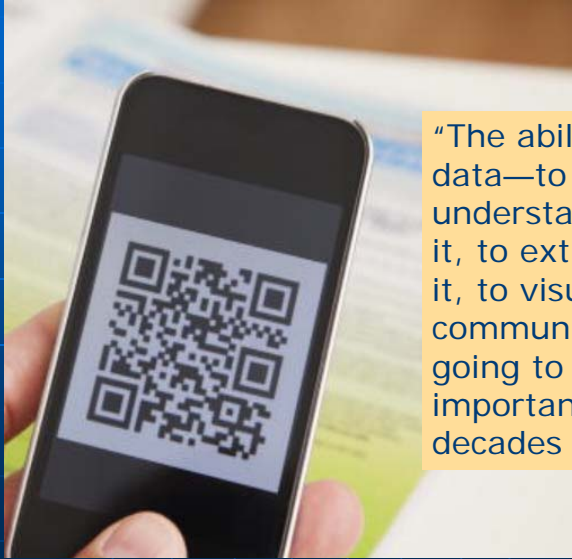
31

Snapshot: Empowering Excel



32

Snapshot: Internet-age Research



"The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades"

33

Research Thoughts Leader

"As data availability continues to increase, the importance of identifying/filtering and analyzing relevant data can be a powerful way to gain an information advantage over our competition."

**Tom H.C. Anderson
founder & managing partner
Anderson Analytics, LLC**

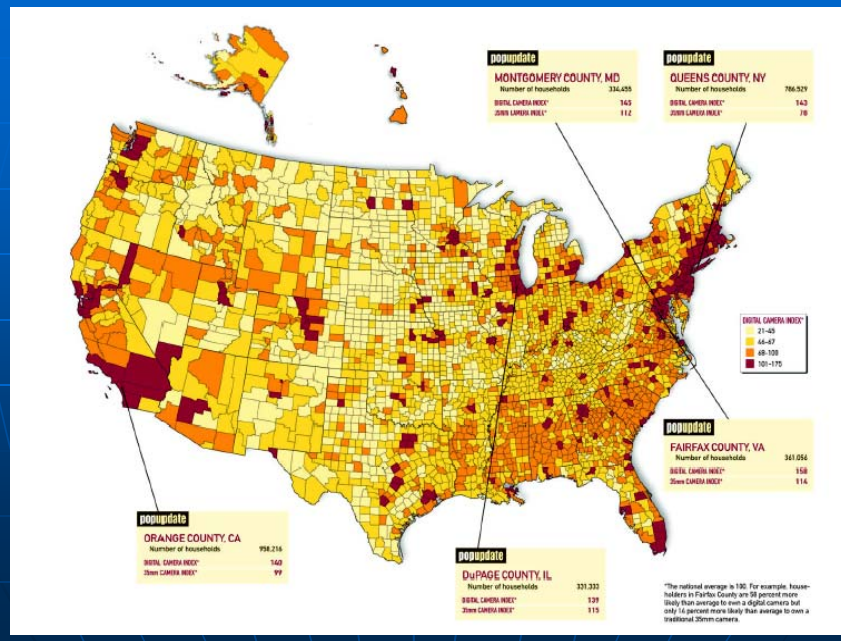
34

PulsePoiny: Research Revelation

65 The percent boost in company revenue created by best practices in data quality.

35

Geograph: Digital Camera Ownership



CloseUp: Original Data Table

Table 1 Spending by Internet Users in Selected Western European Countries 2010 (EUROs in Billions)

	Annual Spending	Annual Purchases
France	Euro 664.5	16
Germany	Euro 658.0	20
Italy	Euro 345.5	14
Spain	Euro 560.1	10
United Kingdom	Euro 2284.9	36

CloseUp: Arranged by Spending Most to Least

Table 2 E5 Per Capita One-Year Online Spending (2010)

	Annual Spending (EUROs)	Average Annual Purchases	Annual Spending (US\$)
United Kingdom	2284.9	36	1736.2
Germany	658.0	20	500.0
France	664.5	16	505.0
Italy	345.5	14	262.6
Spain	560.1	10	425.6

Currency Exchange Rate: 1 US\$ = 1.316 EURO

CloseUp: Arranged by Average Annual Purchase, Most to Least

Table 3 E5 Per Capita One-Year Online Spending (2010)

	Annual Spending (EURO)	Average Annual Purchases	Annual Spending (US\$)	
United Kingdom	2284.9	36	1736.24	Above Average Frequency Buyers
Germany	658.0	20	500.00	Above Average Frequency Buyers
France	664.5	16	504.97	Below Average Frequency Buyers
Italy	345.5	14	262.57	Below Average Frequency Buyers
Spain	560.1	10	425.61	Below Average Frequency Buyers
Average	902.6	19.2	685.88	

Currency Exchange Rate: 1 US\$ = 1.316 EURO

39

CloseUp: Arranged by Average Transaction, Most to Least

Table 4 ES Per Capita One-Year Online Spending (2010)

	Annual Spending (EUROs)	Annual Spending (US\$)	Average Annual Purchases	Average Transaction (EURO)	Average Transaction (US\$)
United Kingdom	2284.9	1736.24	36	63.5	48.23
Spain	560.1	425.61	10	56.0	42.56
France	664.5	504.97	16	41.5	31.56
Germany	658.0	500.00	20	32.9	25.00
Italy	345.5	262.57	14	24.7	18.75
Average	902.6	685.88	19.2	47.0	35.72

Currency Exchange Rate: 1 US\$ = 1.316 EURO

40

CloseUp: Arranged by Estimated Average Transaction, Least to Most

Table 5 E5 Per Capita One-Year Online Spending (2010 vs. 2015)

	Estimated Percent Internet Users 2015	Estimated Internet Users 2015 (millions)	Estimated Average Transaction 2015 (EURO)	Estimated Average Transaction 2015 (US\$)
Germany	89.9	72.2	51.5	37.06
Italy	78.3	45.3	58.6	42.14
France	87.3	56.9	63.8	45.92
United Kingdom	94.9	58.2	95.2	68.48
Spain	77.7	34.5	106.4	76.55
Average	78.3	45.3	58.6	42.14

41

Summary

42