Designing Information:
Communicating Complex Knowledge in Tables and Graphs

Association for Business Communication European Conference

May 29, 2003

Barbara L. Shwom
Northwestern University
Karl P. Keller
Communication Partners

Tables and graphs are crucial for communicating complex information

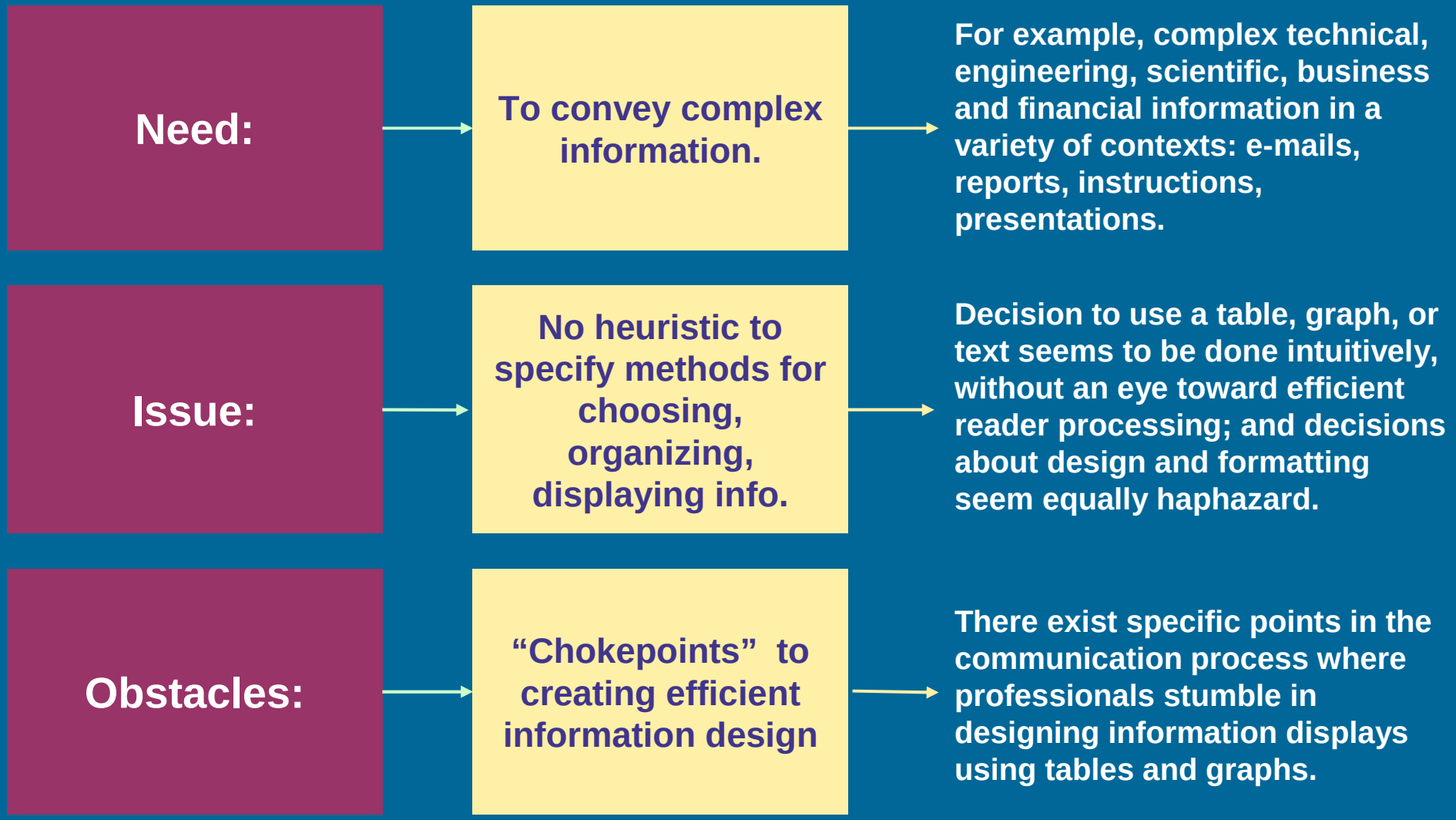
- **Create a "picture" for reasoning about and analyzing quantitative and conceptual information**
  - Makes cognitive processing easier
  - Provides "content/information rich" view at a glance
  - Directs attention toward the content rather than methodology
- **Describe, explore, summarize a set of numbers**
- **Convey messages about significance of data**

Complex information is easier to comprehend when it is "adjacent in space" rather than "stacked in time."

Edward Tufte

Envisioning Information

Technical/business professionals face tough challenges creating good information display



Our hypothesis about obstacles: four fundamental areas where business communicators go wrong

Must decide to select either table, text, or graph

Must know logical purpose behind use of table / graph. Must also make purpose immediately understandable, and the information transparent

Must create best appearance/ format for quick cognitive processing and effective message communication

Must use standard software packages – Word, Excel, PowerPoint – to execute best design -- controlling the software, and not having software control the outcome

Currently undertaking research with students at J.L Kellogg Graduate School of Management, and consulting clients, to survey and quantify these four "chokepoints."

The Decision Point – How do I select either text, table, or graph?

1. Decision Point
2. Logical Purpose
3. Design Principles
4. Software Tools

Real-world example: Complex technical information

- **Context:**

  U.S. Federal Aviation Administration Airworthiness Directive

- **Purpose:**

  To direct mechanics, who service airplanes with certain types of GE Aircraft turbo-prop engines, to replace a specific part on those engines -- various models of "GGT stage 2 forward cooling plates" -- based on new, reduced periods of time, or "cycles."

- **Form:**

  Text

The text below was produced as part of this complex Airworthiness Directive

"This action establishes the following new, reduced cyclic life limits for affected GGT stage 2 forward cooling plates:

(1) 8,000 cycles since new (CSN) for GGT stage 2 forward cooling plates, P/N 6064T10P01, identified by serial numbers listed in Tables 1 and 2 of GE Aircraft Engines (CT7-TP Series) SB No. A772-381, dated January 17, 1996 for GE CT7-5A2, -7A, -9B, and –9C engine models

(2) 12,000 CSN for GGT state 2 forward cooling plates, P/N 6064T10P01 (not listed in (1) above) and P/N 5086T91P02, for GE CT7-5A2 and –7A engine models

(3) 9,000 CSN for GGT state 2 forward cooling plates, P/N P/N 6064T10P01 (not listed in (1) above) and P/N 5086T91P02, for GE CT7-9B/-9C engine models."

This passage is extremely dense, difficult to read, and replete with hard-to-follow numerical designations.

Is this the optimal form to convey this information?

Our hypothesis: the "decision" to use text driven by the writer's assumption about rhetorical form…

What is the rhetorical form of this text?

- ***Narrative?***
- ***Argument?***
- ***Instructions?***
- ***Process?***
- ***List?***

Clearly, a list.

But it appears that the writer, assuming that the Airworthiness Directive is offering instructions, applies the rhetoric of instructions: i.e., step 1, 2, and 3…

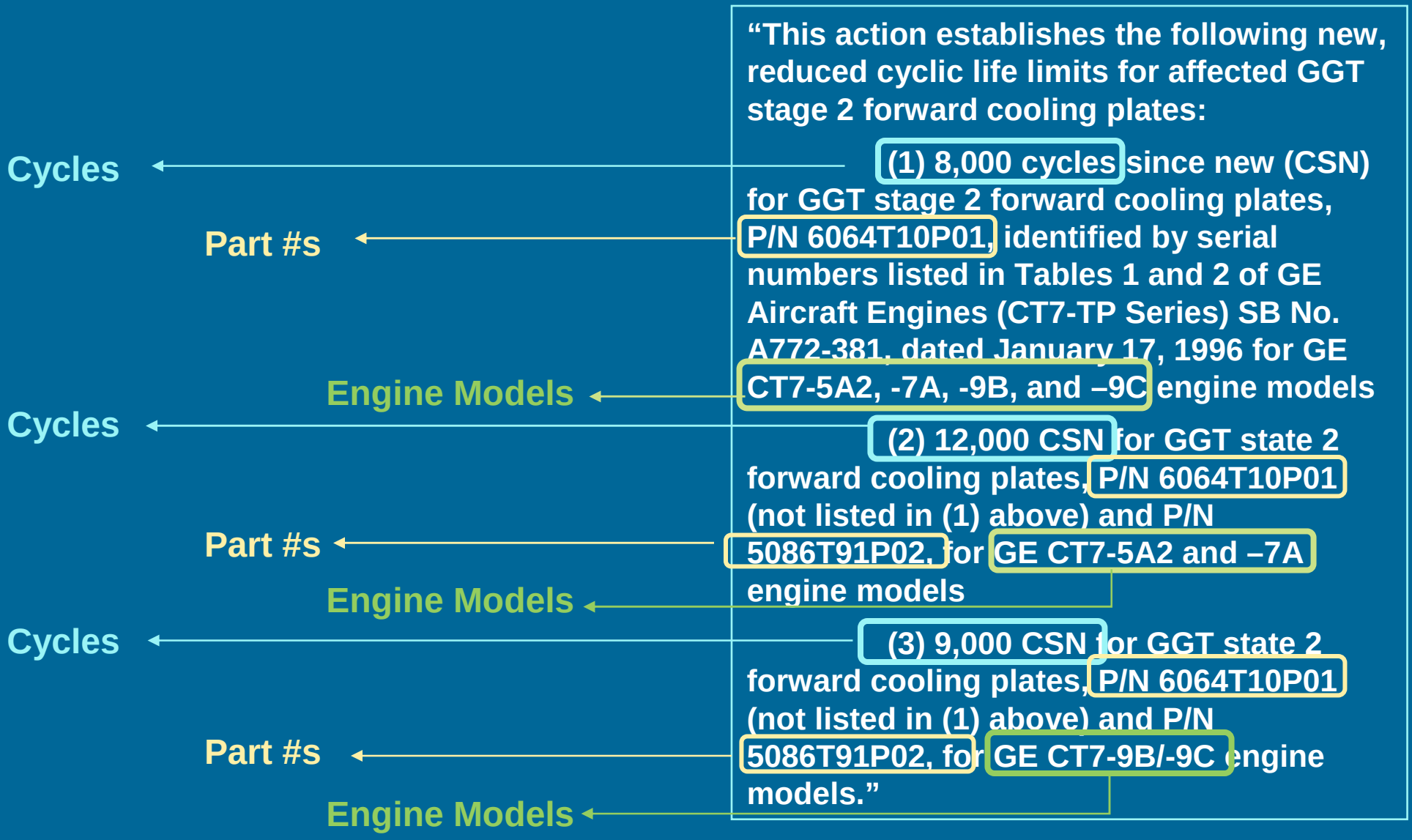
"This action establishes the following new, reduced cyclic life limits for affected GGT stage 2 forward cooling plates:

(1) 8,000 cycles since new (CSN) for GGT stage 2 forward cooling plates, P/N 6064T10P01, identified by serial numbers listed in Tables 1 and 2 of GE Aircraft Engines (CT7-TP Series) SB No. A772-381, dated January 17, 1996 for GE CT7-5A2, -7A, -9B, and –9C engine models

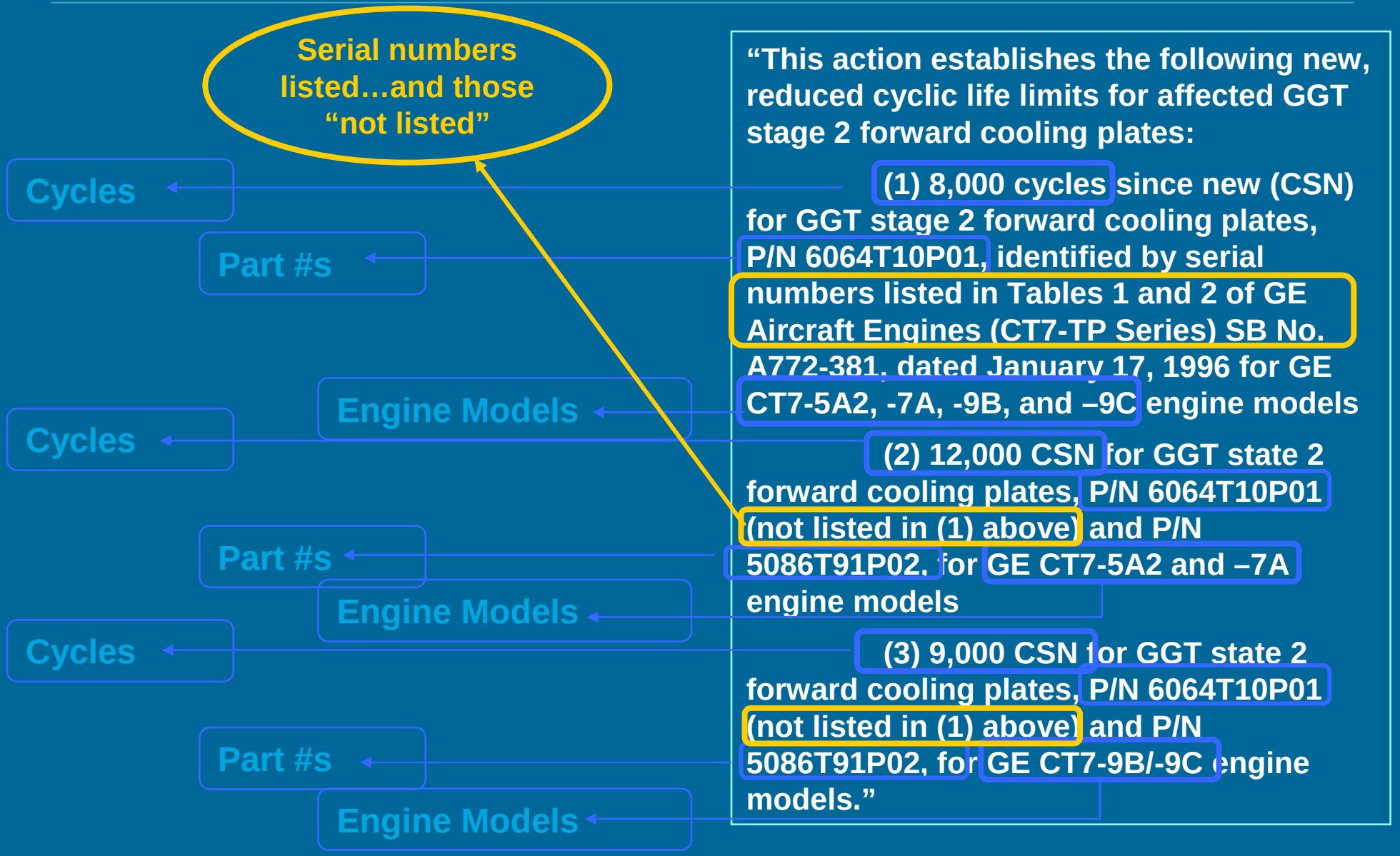
(2) 12,000 CSN for GGT state 2 forward cooling plates, P/N 6064T10P01 (not listed in (1) above) and P/N 5086T91P02, for GE CT7-5A2 and –7A engine models

(3) 9,000 CSN for GGT state 2 forward cooling plates, P/N P/N 6064T10P01 (not listed in (1) above) and P/N 5086T91P02, for GE CT7-9B/-9C engine models."

But it is not a list of sequential items...it is a list of relational items...



AND one other item separates these three "steps..."



Suppose writer recognized that four elements were related to a range of variables....

1. ***Cycle limits – 3 sets: 8,000; 12,000; 9,000***
2. ***Part numbers – 2 sets: P/N 606… and P/N 508***
3. ***Engine models – 4 sets: -5A2, -7A, -9B, -9C***
4. ***Serial Numbers – 2 sets: those identified, and those NOT identified***

Then a different decision might have been made...

Table 6: Revised cycle limits for GGT stage 2 forward cooling plates, Part numbers P/N 6064T10P01 and P/N 5086T91P02

Engine Model	Part Number	Serial Number	New Cycle Limits
GE CT7-5A2, -7A	6064T10P01	serial numbers identified in tables 1 and 2 of the specified SB	8,000 cycles since new
		serial numbers not identified in tables of the specified SB	12,000 cycles since new
GE CT7-9B, and -9C	6064T10P01	serial numbers identified in tables 1 and 2 of the specified SB	8,000 cycles since new
		serial numbers not identified in tables of the specified SB	9,000 cycles since new
	5086T91P02	all serial numbers	9,000 cycles since new

Logical Purpose – How can I make the purpose and message transparent?

Example: Statistics on academic journal acceptance rates

- **Context:**
  Web site of the Academy of Management Review:
  http://www.aom.pace.edu/amr/
  Statistics on acceptance rates for manuscripts from period 6/30/96 to 6/30/98
  http://www.aom.pace.edu/amr/stats96-98.htm#accept

- **Purpose:**
  To show how submitted articles were handled by the editorial board (accepted, accepted after revision, rejected, etc.) and percentages for each category. Note: low acceptance rate tends to convey impression that publication in a journal is highly competitive and therefore prestigious.

- **Form:**
  Table

Purpose appears to be conveying low levels of acceptance (thereby confirming "prestige")...

Acceptance Rates

	# MSS	% of New Submissions
Accepted Manuscripts	32	5.01%
Conditionally Accepted Manuscripts	36	5.63%
Accepted after submission	3	0.49%
Accepted after 1st revision	1	1.15% of revisions
Accepted after 2nd revision	17	44.74% of 2nd revisions
Accepted after 3d revision	11	78.57% of 3d revisions

Summary of Decisions

	# MSS	% of MSS	% of Decisions
Rejected after review	363	56.8%	58.8%
Returned without review	92	14.4%	14.9%
Revision requested	158	24.7%	25.6%
Accepted	3	0.5%	0.5%
Conditionally Accepted	1	0.2%	0.2%
Total Decisions Made	**617**		
Under Review	22	3.4%	
Total New Submissions	**639**		
Revisions Received	**171**		
Total Manuscripts Received	**810**		

Note: this is the order in which the two tables were presented on the web site.

The two tables, as a pair, create confusion; we can't see how numbers or categories relate

Is this the acceptance rate?

Or is this the acceptance rate?

Perhaps this number…

Equals this number…

	# MSS	% of New Submissions
Accepted Manuscripts	32	5.01%
Conditionally Accepted Manuscripts	36	5.63%
Accepted after submission	3	0.49%
Accepted after 1st revision	1	1.15% of revisions
Accepted after 2nd revision	17	44.74% of 2nd revisions
Accepted after 3d revision	11	78.57% of 3d revisions

	# MSS	% of MSS	% of Decisions
Rejected after review	363	56.8%	58.8%
Returned without review	92	14.4%	14.9%
Revision requested	158	24.7%	25.6%
Accepted	3	0.5%	0.5%
Conditionally Accepted	1	0.2%	0.2%
Total Decisions Made	**617**		
Under Review	22	3.4%	
Total New Submissions	**639**		
Revisions Received	**171**		
Total Manuscripts Received	**810**		

But the fact is, we have to work MUCH too hard to decipher this table…

The percentages sure look close

Also, reader has to work much too hard to derive percentages and totals...

32 ÷ 5.01% = 639

The 639 total is on the OTHER table…the top table doesn't tell us what total is used to derive the percentage acceptance rate

	# MSS	% of New Submissions
Accepted Manuscripts	32	5.01%
Conditionally Accepted Manuscripts	36	5.63%
Accepted after submission	3	0.49%
Accepted after 1st revision	1	1.15% of revisions
Accepted after 2nd revision	17	44.74% of 2nd revisions
Accepted after 3d revision	11	78.57% of 3d revisions

	# MSS	% of MSS	% of Decisions
Rejected after review	363	56.8%	58.8%
Returned without review	92	14.4%	14.9%
Revision requested	158	24.7%	25.6%
Accepted	3	0.5%	0.5%
Conditionally Accepted	1	0.2%	0.2%
Total Decisions Made	**617**		
Under Review	22	3.4%	
Total New Submissions	**639**		
Revisions Received	**171**		
Total Manuscripts Received	**810**		

Σ (3, 1, 17, 11) = 32, i.e, the "accepted manuscripts" number…but we have to deduce that the sum of these numbers gives us 32…

Why do these tables fail in their logical purpose?

- ***Categories are blurred –*** **distinction between total acceptances and acceptances on 1st submission not clear**
- ***Calculated results not transparent –*** **totals and components of those totals are on separate tables, and have to be derived by the reader**

	# MSS	% of New Submissions
Accepted Manuscripts	32	5.01%
Conditionally Accepted Manuscripts	36	5.63%
Accepted after submission	3	0.49%
Accepted after 1st revision	1	1.15% of revisions
Accepted after 2nd revision	17	44.74% of 2nd revisions
Accepted after 3d revision	11	78.57% of 3d revisions

	# MSS	% of MSS	% of Decisions
Rejected after review	363	56.8%	58.8%
Returned without review	92	14.4%	14.9%
Revision requested	158	24.7%	25.6%
Accepted	3	0.5%	0.5%
Conditionally Accepted	1	0.2%	0.2%
Total Decisions Made	**617**		
Under Review	22	3.4%	
Total New Submissions	**639**		
Revisions Received	**171**		
Total Manuscripts Received	**810**		

This revision presents distinct categories, shows calculations, and numbers that are consistent

Distinctions made between new submission types

Totals and % explicitly calculated

Numbers consistent

Lines numbered, and room for notes

Table 1: Submissions Breakdown

	New Submission type:	*#MSS*	*% of Total New Submissions*	*Notes:*
1.	Accepted immediately on submission	3	0.47%	
2.	Conditionally accepted manuscripts	1	0.16%	
2.	Revision Requested	158	24.73%	
4.	Under Review	22	3.44%	
5.	Rejected after review	363	56.81%	
6.	Returned without Review	92	14.40%	
7.	**Total New Submissions**	**639**	**100%**	**Sum of items 1 through 6**
8.	**+Revisions received, no review begun**	**171**		**Previously submitted MSS**
9.	**Total Manuscripts Received**	**810**		

Table 2: Decision Breakdown

	Decision type:	*#MSS*	*% of Total New Submissions*	
10.	Accepted immediately on submission	3	0.47%	
11.	Conditionally accepted manuscripts	1	0.16%	
12.	**Subtotal Accepted w/o revisions**	4	0.63%	
13.	Accepted after 1st revision	1	0.16%	
14.	Accepted after 2nd revision	17	2.66%	
15.	Accepted after 3d revision	11	1.72%	
16.	**Subtotal Revised Acceptances**	29	4.54%	
17.	**Overall Acceptances**	**33**	**5.16%**	**Sum of lines 12 and 16**
18.	MSS still out for revision	129	20.19%	
19.	MSS accepted after revision	29	4.54%	
20.	**Subtotal, MSS w/Revision Requested**	158	24.73%	
21.	Under Review	22	3.44%	
22.	Rejected after review	363	56.81%	
23.	Returned without Review	92	14.40%	
24.	**Total New Submissions**	639	100.00%	**Sum of lines 12, 20, 21-23**
25.	**+Revisions received, no review begun**	**171**		**Previously submitted MSS**
26.	**Total Manuscripts Received**	**810**		

Design principles: How do I design the information for easy cognitive processing?

Example: Analysis of distribution channels

- **Context:**

  Consulting presentation to hardware product manufacturer on issues of product distribution and logistics

- **Purpose:**

  To compare number of store locations by distribution category to company product sales

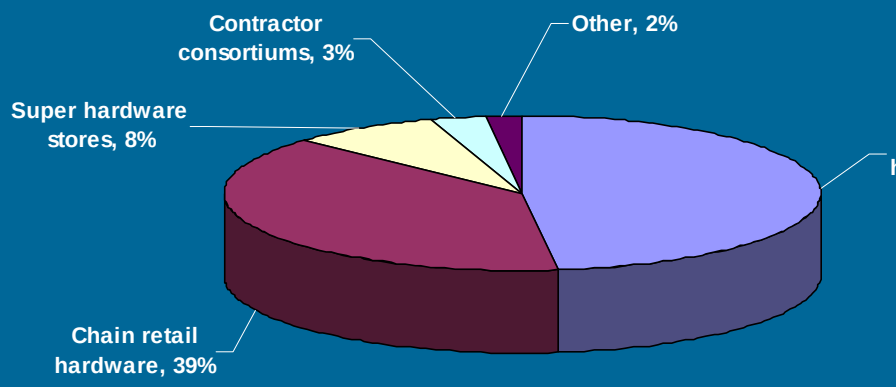
- **Form:**

  Two side-by-side pie charts

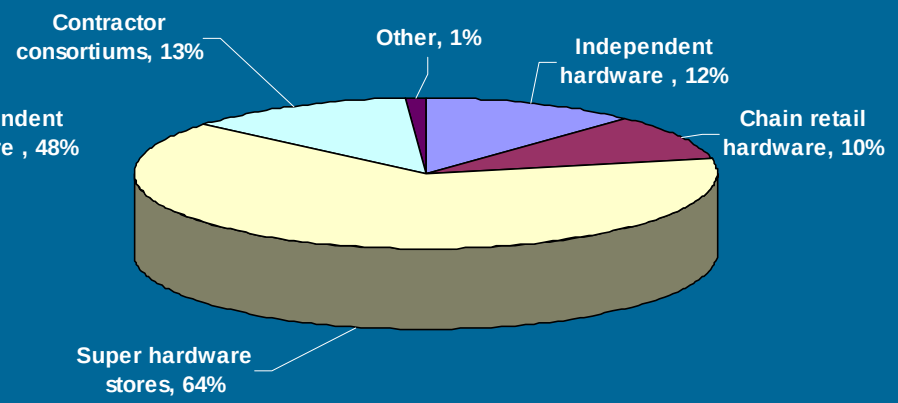
Deciding to chart is a good decision, and data is straightforward, but is this optimal design?

Super hardware stores account for 9% of store locations, but responsible for 64% of our product sales

Percentage of Store Locations



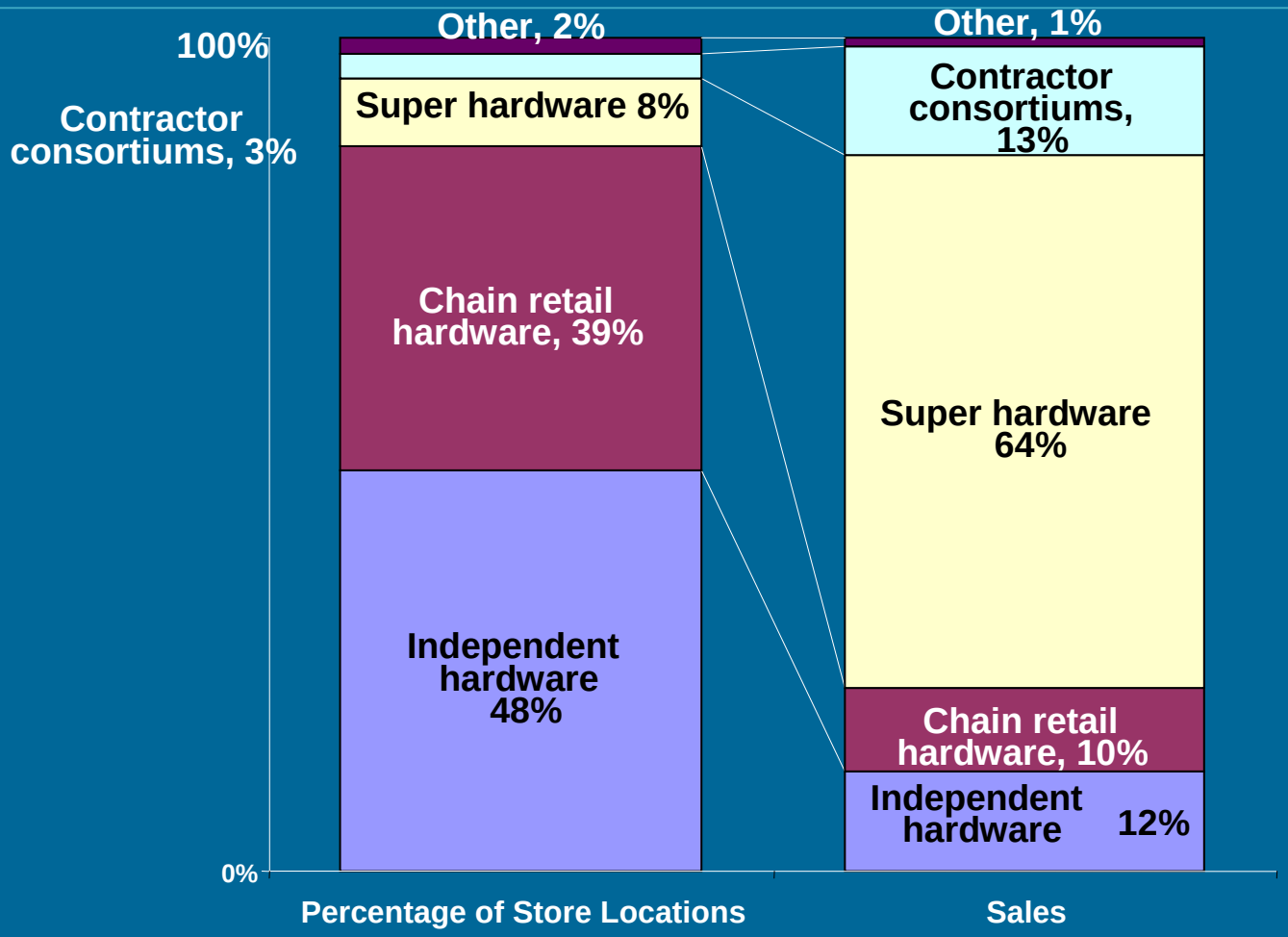
Product Sales



- **3 dimensional renderings bad practice: distort visual representation of percentages**
- **Side-by-side pie charts increase difficulty of cognitive processing: eyes must move back and forth to see important comparisons**

Instead, column comparison chart is actually the best design choice

Super hardware stores account for 9% of store locations, but responsible for 64% of our product sales

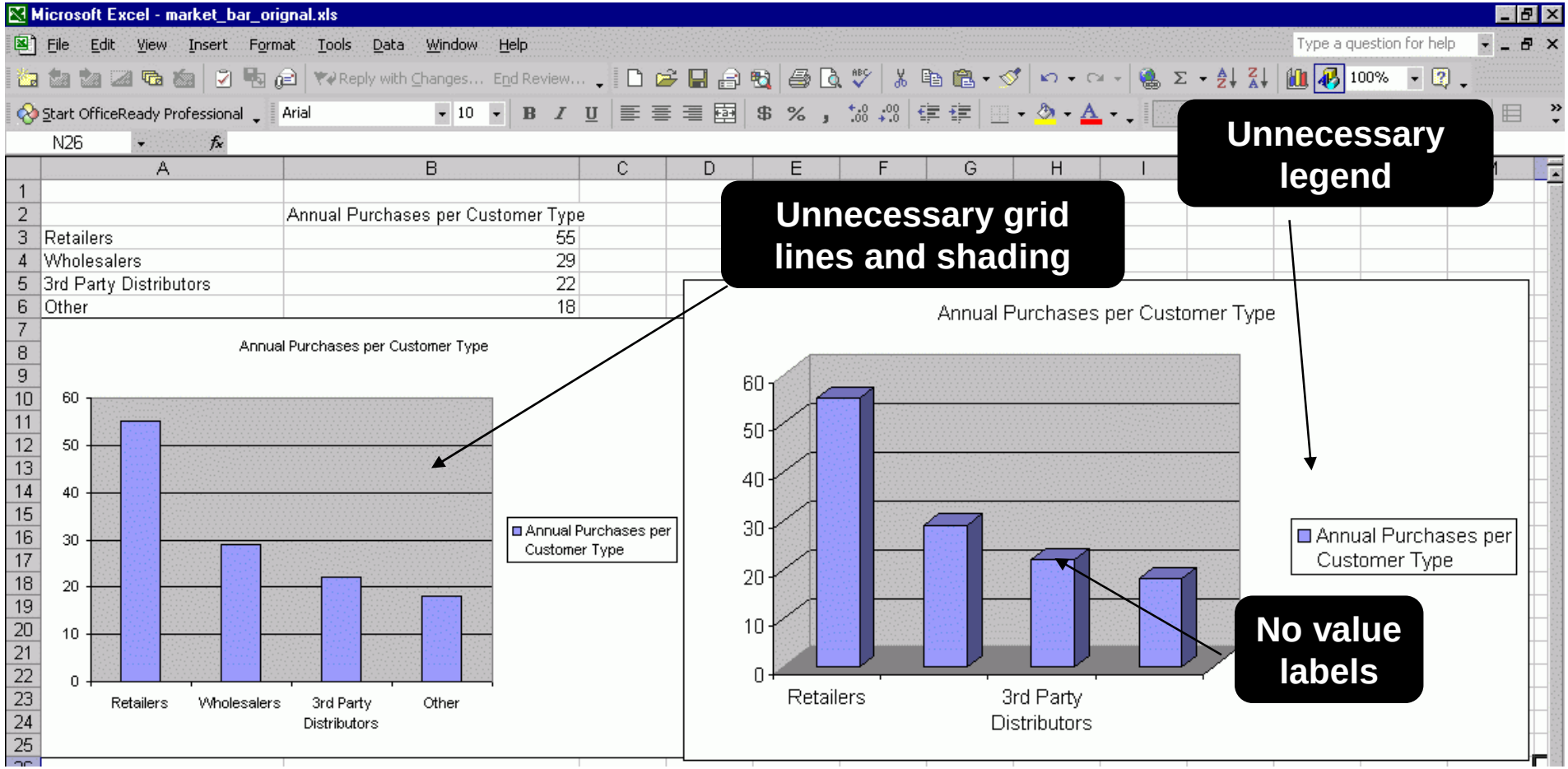


Software Tools: How can I create best design using the software available?

Example: Using Excel to create graphs and tables

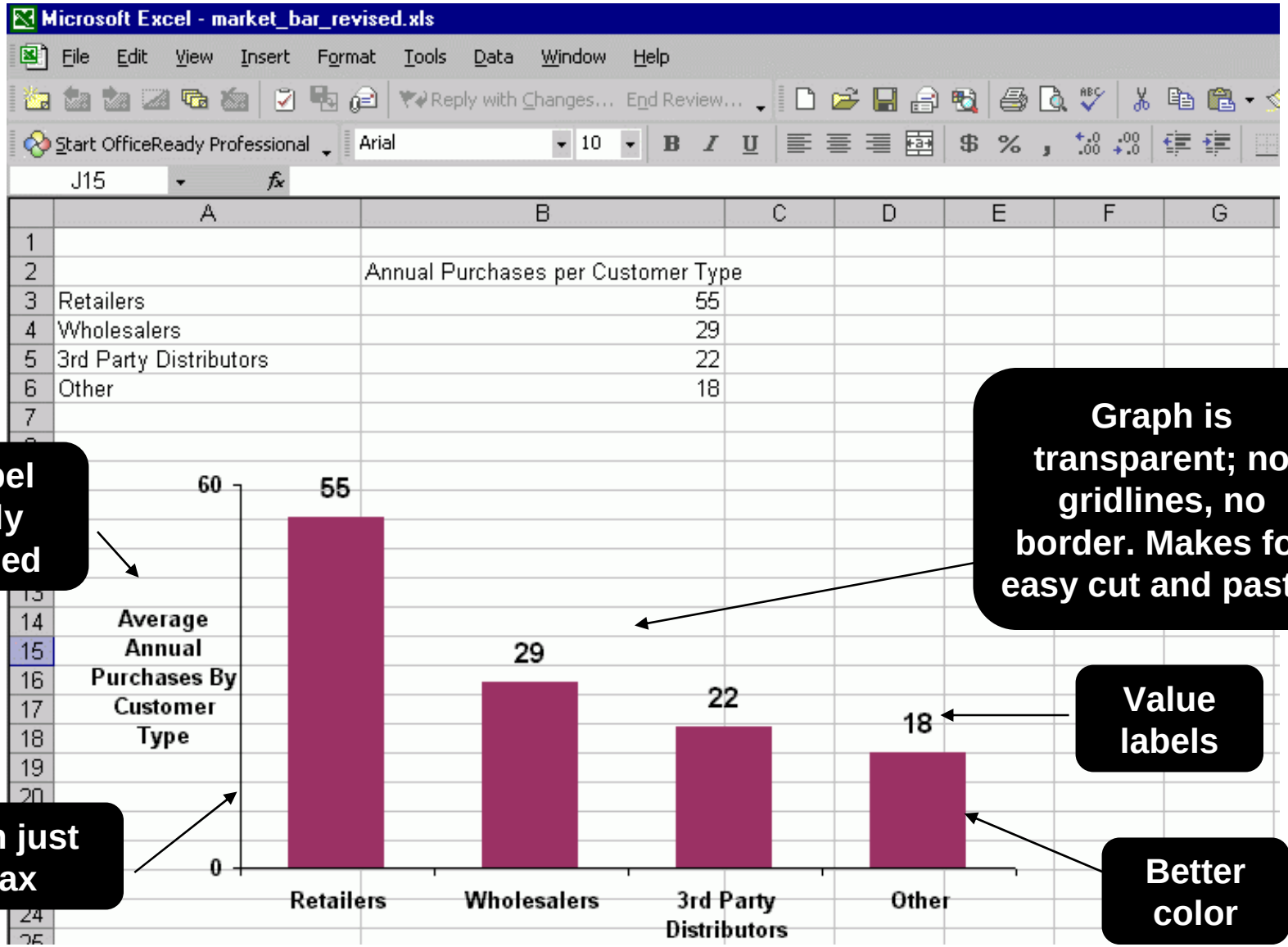
- **Context:**
  Company presentation about sales activities in specific distribution channel
- **Purpose:**
  To show that one category is the largest purchaser of products
- **Form:**
  Column graph

Excel wizard defaults to graphs that violate best practices in graph design



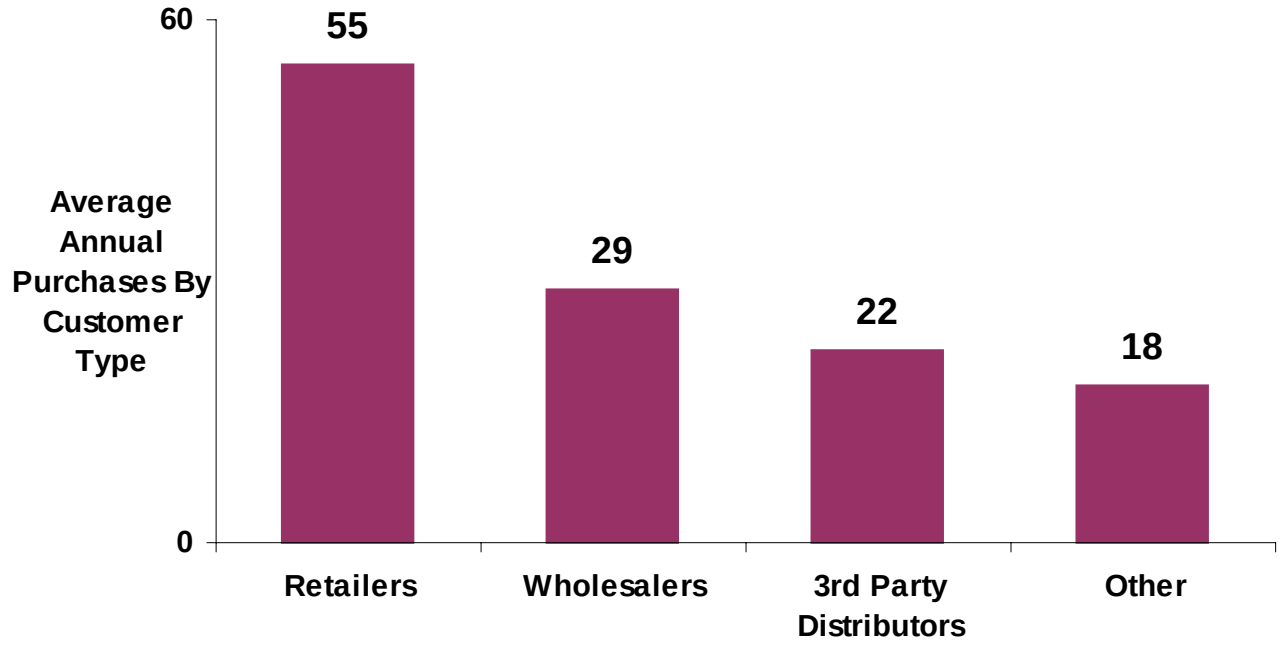
Some users erroneously try to create a better look by choosing 3-D

Here is what you can create in Excel if you know how to manipulate the options in the wizard

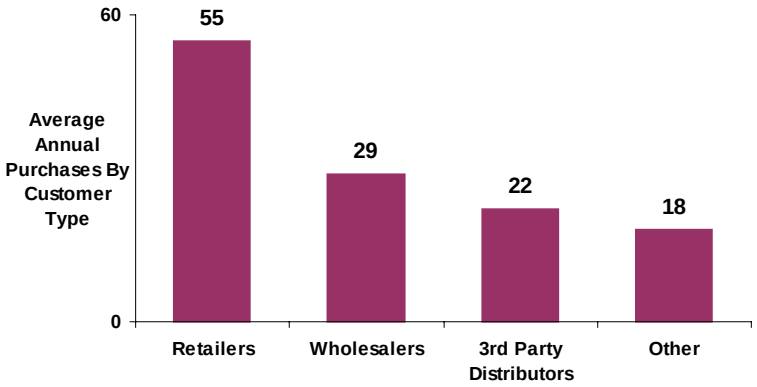


Final chart, with clean uncluttered look, and good message headline

Retailer channel is biggest purchaser of our products



What are the steps to create this look?



Step	Inside Wizard Steps
1	Select shaded cells B3:C7 above and column chart type, first option, then press next
2	View chart in wizard--data souce correct?looks correct?, then press next
3	In chart options, select following tabs:
3a	Gridlines -- uncheck all boxes (must do this HERE, else maximum gridline will remain)
3b	Legend -- uncheck legend box, want no legend
3c	Data labels -- check "show values", click next,
4	Select chart location "as object in", so it stays in spreadsheet, press finish
	Outside Wizard Steps
5	Select entire chart area, Format | Selected Chart Area|
	Patterns | Border | None, Area | None, Press OK
6	Select any shaded area, Format | Selected Plot Area|
	Patterns | Border | None, Area | None, Press OK
7	Select Y axis, Format | Selected Axis| Press Scale Tab, Uncheck
	Major Unit, Minor Unit, insert maximum unit number into
	Major Unit AND Minor Unit boxes, Press OK
8	Select data series, Format | Selected Data Series | Border
	Custom, and make the border of the columns the same color as the fill on
	the graph -- Note:unless you do this you will see no outline in your B&W print
9	Select entire chart area, bold your text
10	Graph now ready to be "pasted as picture" inside PowerPoint

The good news is that you only need to do this once, because you can save this graph design as a "custom graph" and select it next time you are in Excel

Excel spreadsheets must also be manipulated to create well-designed tables

A Excel spreadsheet with minimal modifications

All caps for header rows, items?

Mis-aligned items

Unnecessary cell padding

Too many $

Inconsistent fonts

Excessive grid lines, column dividers

Table not numbered or titled

	CALLS/DAYS	DAYS/YEAR	CALLS/YEAR
WIDGETS	8	200	1,600
DOOHICKEYS	5	200	1,000
TOOTOOS	5.5	200	1,100
INGOTS	5	200	1,000
STITCHES	5	200	1,000
CURLS	5	200	1,000
PEEWEES	8	200	1,600

COST/ REP	REP COST / CALL						
1998	1999	2000	2001	2002	2003	2004	2005
$ 155,056	$ 99.82	102.81	105.90	109.07	112.35	115.72	119.19
$ 154,634	$ 159.27	164.05	168.97	174.04	179.26	184.64	190.18
$ 144,620	$ 135.42	139.48	143.66	147.97	152.41	156.99	161.69
$ 169,191	$ 174.27	179.49	184.88	190.43	196.14	202.02	208.08
$ 169,191	$ 174.27	179.49	184.88	190.43	196.14	202.02	208.08
$ 215,456	$ 221.92	228.58	235.43	242.5	249.77	257.27	264.98
$ 161,308	$ 103.84	106.96	110.17	113.47	116.68	120.38	123.99

A better designed version

Sentence case--consider italics to set apart items, header rows

Tables numbered, titled for easy reference

Simple, single vertical column dividers

Consistent fonts

One line, one row

One row of $ all that's needed

Simple column spanner

Table 1 -- Call Totals by Product

	Calls/Day	*Days/Year*	*Calls/Year*
Widgets	8.0	200	1,600
Doohickeys	5.0	200	1,000
Tootoos	5.5	200	1,100
Ingots	5.0	200	1,000
Stitches	5.0	200	1,000
Curls	5.0	200	1,000
Peewees	8.0	200	1,600

Table 2 -- Cost Per Rep and Rep Call

Total Rep Cost	Rep Cost / Call						
1998	*1999*	*2000*	*2001*	*2002*	*2003*	*2004*	*2005*
$ 155,056	$ 99.82	$ 102.81	$ 105.90	$ 109.07	$ 112.35	$ 115.72	$ 119.19
154,634	159.27	164.05	168.97	174.04	179.26	184.64	190.18
144,620	135.42	139.48	143.66	147.97	152.41	156.99	161.69
169,191	174.27	179.49	184.88	190.43	196.14	202.02	208.08
169,191	174.27	179.49	184.88	190.43	196.14	202.02	208.08
215,456	221.92	228.58	235.43	242.50	249.77	257.27	264.98
161,308	103.84	106.96	110.17	113.47	116.68	120.38	123.99

Good alignment, consistent decimalization

Eliminating grid gives clean but readable look

Increasing awareness of 4 “chokepoints” can help us design good displays of complex information

Select proper type

Know purpose and message, and make them transparent

Create best appearance

Know software capabilities to execute good design