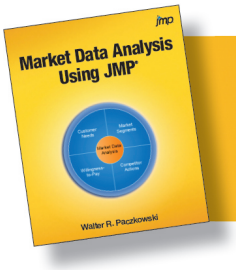
Market Data Analysis Using JMP®

Walter R. Paczkowski



Contents

1

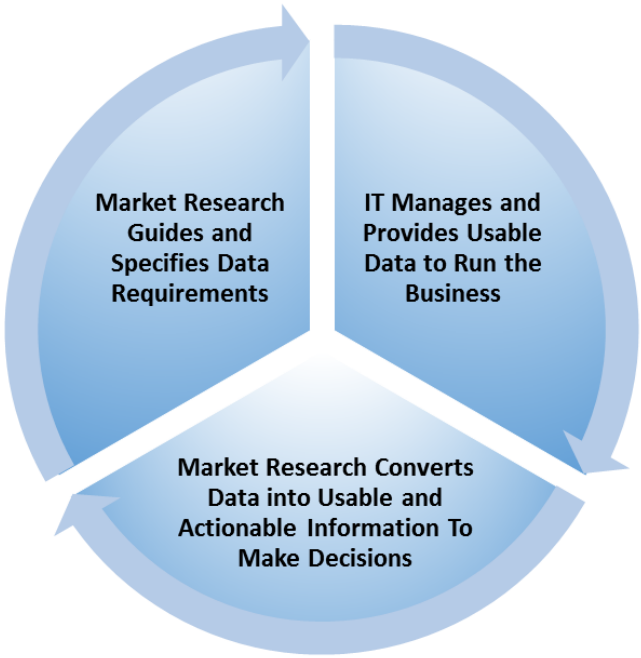
The State of Market Research

Introduction

The role and function of market research is to enable information-driven decision making in businesses. Business leaders need actionable market information, not just “data,” for making intelligent decisions in highly competitive, differentiated, and rapidly evolving markets fraught with uncertainties and risks. Information is insight into customer purchasing behaviors, product preferences, willingness-to-pay, and groupings (a.k.a., segments) that increases understanding by showing previously unknown trends, patterns, anomalies, and relationships. The uncertainties and risks of markets are greatly reduced, albeit not eliminated, because of this information, so that better decisions can be made and the best actions can be taken for a business problem. Data are the building blocks for information, almost like Lego bricks that haven't been assembled. As building blocks, they can be assembled in many different ways with rules specifying how to do the assembling to create information that leads to decisions.

Market researchers provide that information while others in the business, such as those in the IT department, provide data, not information. This is an important distinction because there’s a division of labor that, on the one hand, separates the two functional areas but, on the other hand, interconnects them in a complex, synergistic way. Market researchers analyze the data for trends, patterns, and relationships while the IT staff organize and maintain the data in *data warehouses* and make them accessible in a useful form to a wide audience usually through *data marts*. The market researchers, in turn, inform and guide the IT staff regarding the type of data they need and the form or organization of the data most convenient for them. This synergy is illustrated in Figure 1.1 on page 3.

Figure 1.1 *Market Research – IT Synergy*



This chapter discusses the major tools, statistical and software, that market researchers use and need to turn data into information through analysis. These tools, however, are often unknown or inadequate for the job. So, this chapter identifies problems with some tools and argues what needs to be done to better address market researchers' statistical and data handling problems.

This chapter is also a guide to succeeding chapters. In later chapters, I describe how the problems and issues I raise here are addressed with JMP, a fully functional, user-friendly statistical package with dynamic capabilities for performing penetrating analysis and extracting information that can serve as the focal point for handling the problems. Market researchers need software with three features:

1. powerful data handling methods to manipulate and shape data for meaningful analyses;
2. dynamic graphing and tabulation capabilities to drill down and link data for deep and penetrating insight;
3. statistical methods appropriate only for the type of data being analyzed.

JMP fills the bill perfectly. It provides appropriate tools and analysis reports tailored for those who are novice data analysts and for those who are more sophisticated in the tools-of-the-trade, while also providing rich analytical methods as well as scripting capabilities for creating custom reports and analyses.

So, although this chapter is about how I see the current state of market research, it's also about the role that JMP plays in converting data into information; that is, it is about its role in *data analysis*. Data analysis is not just looking at data and reporting what you see; it's about breaking raw data into parts and reassembling them to extract information. I expand on the role that JMP plays in this task in Chapter 2.

This chapter has six main sections:

1. The first introduces the role of market research in providing the information that key decision makers need to run a business.
2. The second outlines the sources of the data that are turned into information. Surveys are the main source that most researchers use, but it's not the only one. Data can also come from a data warehouse or data mart.
3. The third section is the first of three that discuss the main tools that most researchers rely on. This section focuses on "the tabs."
4. The fourth section adds the typical visual displays, pie and bar charts, to the discussion of tools. These two charts not only are the visual reporting tools, but they're also quite often the analytical tools.
5. The fifth section discusses the role of spreadsheets in data analysis and the issues they have for analyzing data.
6. The final main section argues that we can do better – and that JMP is the tool to use.

Market Research Role in Business

At the end of each day, every business must sell a product. If nothing is sold, nothing is earned and the business ceases to exist. What is sold must meet customer needs at the right price, be the best, be the first, and be easily accessible. To make sure their product is sold, business managers need to know four things:

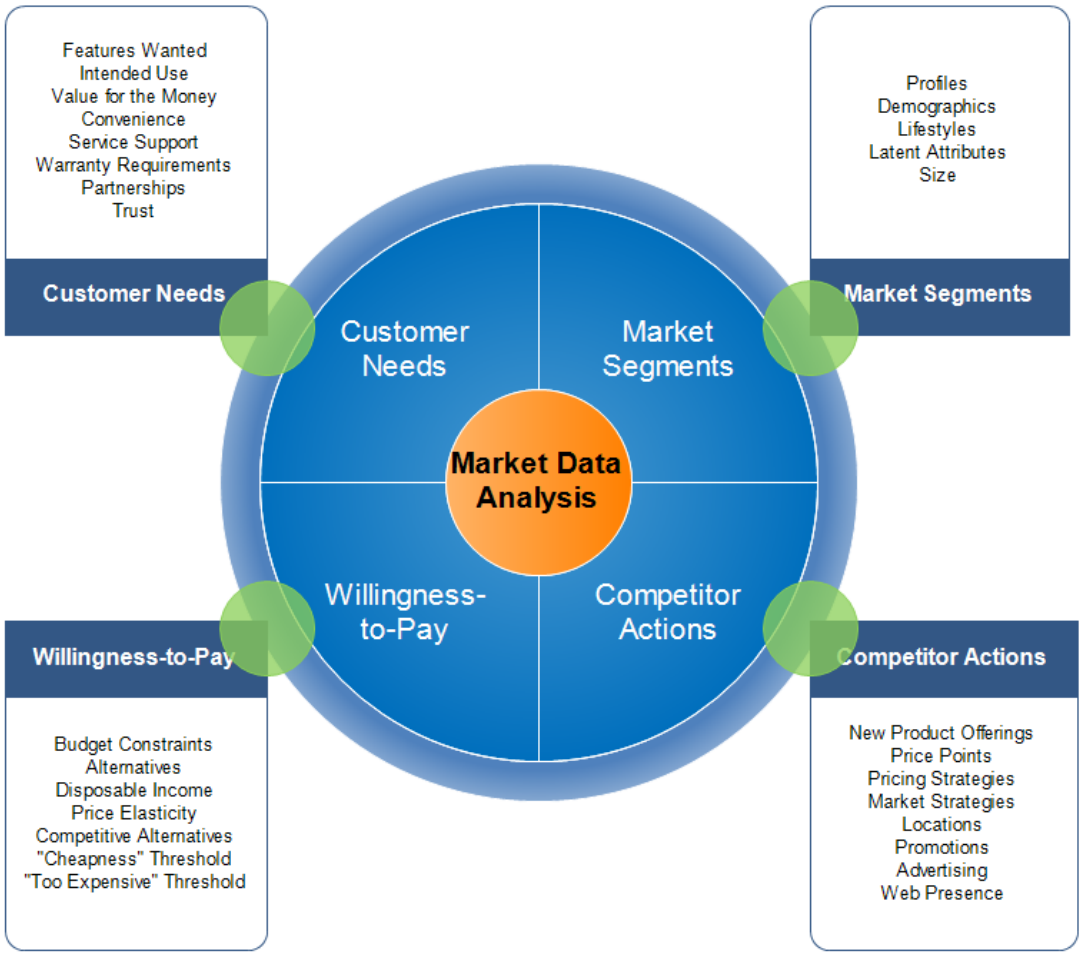
1. what customers want;
2. how much customers will pay;
3. what competitors are offering;
4. how consumers are grouped or segmented for optimal pricing, product development, and selling.

With this information, they can strategize and measure the effects of their strategic and tactical decisions on their key business metrics. A few key business metrics are profit,

revenue, contribution, and sales, but there are probably many more that any business manager can devise.

The market research department in any major business is the organization typically called upon to provide this information. This information-providing function is displayed in Figure 1.2 on page 5.

Figure 1.2 *Market Research Functions*



In order to fulfill their objectives, market researchers typically use four tools. Their use, of course, depends on the researchers' sophistication and the problem they're addressing. Nonetheless, these four tools are used more often than not. These are:

1 surveys as the prime data collection methodology;

2 cross-tabulations as the main analytical tool;

3 simple, static charts such as pie and bar charts for analysis and display;

4 spreadsheets for managing, organizing, analyzing, and displaying survey data.

I'll discuss each of these in detail in the following sections.

Surveys for Data Collection

A Focus on Surveys

Surveys are definitely the major source of data for market studies in modern businesses. They're unsurpassed for learning opinions and revealing preferences. Opinions are varied but are usually about the quality of service and performance on key measures or attributes of the product or business. This could include service rep responsiveness, pricing and payment options, delivery promptness, product quality, respect and courtesy, and so forth. Customer satisfaction studies, for example, try to isolate these opinions to measure how well the business is performing, especially relative to the competition.

Understanding preferences is important because they're usually about new product features, price points, messages, etc., but these preferences can only be revealed when people are exposed to the object in question (e.g., a new price point), which may be impossible to do in an actual market implementation. In a pricing study, for example, surveys reveal preferences for pricing programs, strategies, and willingness-to-pay without the firm actually changing prices in the market, which could alienate vital customers if not done correctly.

Surveys are used to collect *experimental data* as opposed to *transactional data*. I often refer to survey data as experimental data because they're collected under controlled conditions with an experimental design in at least one part of the questionnaire. An experimental design is also used to sample respondents and to show them questions or tasks. The data collection design could be a simple random sample (not typically used) or a stratified random sample (more typically used). Some examples of studies that use an experimental approach are conjoint, MaxDiff, and discrete choice. I'll show in Chapter 6 how these can be designed and analyzed using JMP. I discuss survey data, in general, in several chapters of this book where I'll also show how JMP is a powerful and robust tool for turning survey data into actionable and useful information.

Transactional or sales data, also called *observational data*, are records of what people actually did in the market: what was purchased, how much was purchased, price paid, purchase date, location, and much more. The emphasis is on what was done -- *retrospective* -- not what will be or could be done -- *prospective*. In addition, the data are for customers for that business only; data on customers of like-products sold by competitors are not, and cannot be, included. This results in an incomplete view of the market. I discuss transactional

data, now called *Big Data,* in Chapter 8 and show how JMP will enable you to manage and analyze this data for information.

The Need for Speed

In our modern technology-driven economy, most surveys are done online with sophisticated programming that allows questionnaires to be quite complex. Paper and pencil and telephone surveys are still used, but online surveys are now the norm rather than the exception. Their use contributes to a subtle problem -- the need for speed. They can be quickly implemented with panels of respondents that can likewise be quickly assembled to meet pre-specified criteria (e.g., ethnic/racial profile, education level, product experience), thus shortening implementation time. In addition, many online survey vendors have templates for most standard questionnaires and question sets that just require users to modify wording for their particular problem, thus speeding up the process of writing and fielding a study.

This speed translates into clients requiring (almost demanding) a quick turn-around of reports, which are expected hours -- or less -- after a study leaves the field. The easiest way to meet this expectation is to ask online vendors to provide canned reports that have simple charts and tables for each question to go along with the questionnaire templates.

Surveys Are Overemphasized

Transaction data, maintained in either a data warehouse or data mart, can also be used for developing information about markets. A data warehouse is an all encompassing compilation of data on every aspect of the business and a data mart is a specific functional area (e.g., marketing, finance, logistics) subset of the data warehouse. Scanner data are a prime example of transaction data.

Market researchers leave working with databases to data scientists who are usually in the IT department, thus ignoring this data's richness for providing insight about market strengths, weaknesses, opportunities, and threats (*SWOT*). There are two reasons for this. The first is the sheer complexity of these databases, which is daunting to most. A very special skill set is needed to handle, let alone analyze, them. I touch on some issues in Chapter 8.

The second reason is the belief that only surveys can tell you what customers want, and only surveys can reveal their opinions and preferences. To a good extent this is true, especially regarding new products, enhancements to a product line based on modifications to product features, or new price points. These will never appear in a database, which, by its nature, must be historical. Showing what did happen is unimportant for many business decisions. Opinions and preferences for new products, features, and price points can only be obtained by asking.

What people actively did as reflected in a database is equally important for information-driven business decisions because you can learn how people behaved under different conditions (e.g., price points), what sold, when sales increased and decreased (e.g., seasonal patterns), what appealed to which group of consumers (i.e., segments), and so forth. So databases should not be ignored, but relying on surveys to the exclusion of databases can go too far.

Wheeler (2012) cites an anecdote about a data mining professional who was told by a client: "*The problem with your data … is that it's not the real data. We should use real data … data from the surveys we take, not data from the web*." This is misguided because an overreliance on surveys is dangerous. Quite often what people say they want and will buy is not always what they buy. People tend to be inconsistent. Because of this, discrete choice experiments, based on surveys, are sometimes combined with actual purchase data to gain better insight into preferences. I briefly discuss this in Chapter 6.

Tabs as an Analysis Tool

A Focus on Tabs

In addition to surveys being overworked as a data collection method, cross-tabulations of the data, or simply "the tabs," are equally overworked as the primary analysis tool. A client once told me that "*Everything you need to know is in the tabs*." In my opinion, this sums up the prevalent view of market researchers.

Tabs are quite often printed in books with voluminous pages and 8-point font. They're predefined once the survey is designed with a complex set of programming code generating them, code that becomes difficult to change so that the tabs are usually unchangeable; i.e., they're static. This has an important implication for data analysis: you can't look at rearrangements or reconfigurations of the data without having to reprogram the tab software, and maybe printing new books. There's a time cost, as well as a monetary cost, which plays to the "need for speed" issue. To minimize these costs and to quickly generate a report, only one set of static tabs will be produced. This definitely handicaps analysis.

Tabs Are Just Contingency Tables

The concept of a cross-tab is familiar to statisticians; they're just contingency tables created by crossing two discrete variables to create a single table. The cells of the table are the number of observations (the count or frequency) determined simultaneously by a particular combination of two categories of the two variables. Dividing by an appropriate base changes the cell frequencies into cell proportions or percentages. The data in the cells of the table can thus be presented in several ways:

- raw counts or frequencies;
- percentages of the total sample (base: total sample);
- percentages of the columns (base: column totals);
- percentages of the rows (base: row totals).

The columns are called *banner points* or *variables* and the rows *studs*. The banner points are interpreted as independent variables that determine the responses to survey questions. Demographics, consumer segments, and quota groups are typical banner points. The studs

are the question responses and so can be interpreted as dependent variables. A typical cross-tab page would have several banner points (usually at a very small point size) but only one stud variable. The process of creating a cross-tab table or tab is called “tabbing.”

A typical cross-tab is shown in Figure 1.3 on page 9. For this example, the two variables are brand of yogurt and where yogurt may be purchased. The *Brand* is the banner point and the *Store* where yogurt is typically purchased is the stud. There were 906 survey respondents. I produced this cross-tab using JMP; I explain how in Chapter 5. When continuous variables are involved (for example, dollars spent), sums, means, medians, and standard errors are also included.

Figure 1.3 *Example Cross-tab*

Contingency Analysis of Brand By Where Typically Purchase Yogurt

Which brand of yogurt do you typically buy? By Where do you typically buy yor yogurt?

Contingency Table

Where Typically Purchase Yogurt	Count / Total % / Col % / Row %	G	E	D	C	Major Competitor	Client	Total
		Brand						
National Grocery Chain Store		12	20	28	88	81	100	329
		1.32	2.21	3.09	9.71	8.94	11.04	36.31
		36.36	41.67	30.77	34.65	42.19	34.72	
		3.65	6.08	8.51	26.75	24.62	30.40	
Locally Owned Health Food Store		0	1	4	16	10	17	48
		0.00	0.11	0.44	1.77	1.10	1.88	5.30
		0.00	2.08	4.40	6.30	5.21	5.90	
		0.00	2.08	8.33	33.33	20.83	35.42	
National Health Food Store		18	22	42	115	74	142	413
		1.99	2.43	4.64	12.69	8.17	15.67	45.58
		54.55	45.83	46.15	45.28	38.54	49.31	
		4.36	5.33	10.17	27.85	17.92	34.38	
Convenience Store		3	5	17	35	27	29	116
		0.33	0.55	1.88	3.86	2.98	3.20	12.80
		9.09	10.42	18.68	13.78	14.06	10.07	
		2.59	4.31	14.66	30.17	23.28	25.00	
Total		33	48	91	254	192	288	906
		3.64	5.30	10.04	28.04	21.19	31.79	

Problems with Tabs

Too Much Reliance

Tabs, although widely used, have five problems that compromise their use:

1. they can quickly become voluminous thus hampering a search for information;
2. they have static rather than dynamic views of the data;
3. they largely contain just simple univariate descriptive statistics;
4. they don’t show or measure effects;

5 they don’t visualize relationships.

This is not to say you shouldn't use them; just that you shouldn't be so heavily reliant on them for analysis.

Problem 1: Tabs Can Be Voluminous

Cross-tabs can quickly become voluminous since tabs compare just two variables. Suppose a questionnaire has 100 variables and you tab every pair. This means you have 4,950 $(= [100 \times 99]/2)$ tables. Imagine looking through 4,950 tables trying to find actionable information! Where do you begin?

For the 4,950 tables, you need 990 sheets of paper, assuming that five tables fit on an 8.5 x 11 sheet of paper with an 8-point font. A ream of paper is 500 sheets (one standard package), so you need approximately two reams of paper. One ream is approximately 2.25 inches thick, so you would have a stack of paper almost 4.5 inches thick! And just for 100 variables. A typical questionnaire has more than 100 variables, so the size of the tabs has the potential to be very large.

Not long ago, the tabs were physically printed and bound into books (hence, this reams of paper issue). This made gleaning any insight from them challenging to say the least because of the sheer volume of pages, not to mention the fact that the pages are just static snapshots of the market based on people's responses to survey questions. The tabs are now in digital books, although physical books are still popular. Digital books are equally difficult to work with because the volume issue still remains. Plus, there's the added factor that paging from one part of a digital book to another makes it difficult to compare one page to another, unlike for a physical book.

Problem 2: Tabs Are Static Views

I mentioned that tabs are static, not providing a dynamic view of your data. *Dynamic* in this context means highly interactive. You can:

- drag and drop variables to a table or graph canvas to quickly create new views;
- click an icon or select from a menu to change views of the same data;
- drill-down and link tables and graphs to quickly see interrelationships.

“Dynamic interactive graphics *are graphics that can move smoothly and change in response to the data analyst's actions, the changes being computed and presented in real time*” (Young, Valero-Mora, and Friendly, 2006; emphasis in original). *Static* means non-interactive or fixed, so drilling down and linking are impossible. You can only see the one pattern or maybe two patterns in the view given by a static table or chart. The view can be changed, but not easily. If time is of the essence for most studies, then changing the tabs to create new views becomes a luxury many analysts can’t afford. The static tabs impose an unforeseen cost on doing data analysis as noted above.

JMP is designed to allow you to have dynamic views of your data, so that you can drill-down and link this data across several displays for comprehensive views. This is the strength of

JMP, its forte, for analyzing market data for information. I illustrate this in Chapters 4 and 5 and use this dynamic analytical capability in Chapters 6–8.

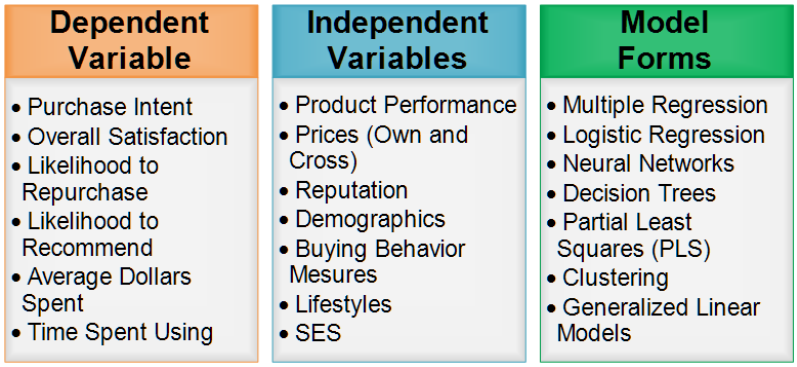
Problem 3: Tabs Have Descriptive Statistics

Tabs are just tables of descriptive statistics at best and these statistics are univariate. *Univariate* means that only one variable is summarized or analyzed. Means, medians, proportions, and standard errors are examples of univariate, descriptive statistics. Even though two variables are "crossed" in the table, just simple univariate summary statistics are calculated and reported. A simple correlation coefficient for a bivariate relationship, for example, is not calculated and shown.

Markets, however, are *multivariate*, meaning that many variables or factors interact and work simultaneously to produce whatever result you're studying. It's never the case that only one variable explains another or one summary measure (e.g., a mean or proportion) reveals complex interrelationships. An example is purchase intent, which is frequently measured on a 5-point Likert Scale ranging from "*Extremely Unlikely*" to "*Extremely Likely*." A univariate statistic, such as the mean (assuming the measurement scale is at least interval), is frequently calculated, shown in the tabs, and reported to the client. This, however, ignores how it interacts with price, product features, types of stores where available, etc. -- and all simultaneously! These complex, simultaneous relationships can only be uncovered, and the direction and magnitude of the interactions determined, by a *regression model* (or an appropriate member of the regression family). Regressions are not in the tabs.

A model separate from the tabs is needed to estimate multivariate relationships. A model is a statistical representation, an abstraction, of what you believe is the relationship between a factor of interest, a dependent variable, and other key variables, the independent variables. There are many different types of models; the applicability of each depends on the nature of the data. This is summarized in Figure 1.4 on page 12. The type of model that can be used depends on the type of dependent and independent variables. The modeling types are *Continuous*, *Ordinal*, and *Nominal* scaled. JMP "knows" the type (and sometimes the role—e.g., dependent variable) and gives you appropriate modeling options in modeling *platforms*. For example, if the dependent variable is nominal, then logistic regression will be available as the recommended method or *personality* in the *Fit Model* platform, although you can choose something else. JMP allows you to estimate many different types of models as I illustrate in Chapters 5, 6, and 8. In fact, JMP helps the analysis process by providing only those model types appropriate for the data types being analyzed. I give a broad, high-level introduction to JMP platforms and personalities in Chapter 3.

Figure 1.4 *Statistical Model Highlights*



Problem 4: Tabs Don't Show Effects

Consider purchase intent again. A logical question to ask is: "*How much does purchase intent change because of a price change*?" Yes, intent will rise if price falls. That's obvious and simple economics, but the real issue is how much. This is given by an *elasticity*, something else that's not in the tabs. An elasticity shows the responsiveness of a change in one quantity (e.g., purchase intent) to a change in another (e.g., price). A key business metric, revenue, is intimately connected to a price elasticity. A model provides this measure of effect, which is another reason to go beyond tabs. I show you how to estimate elasticities for choice models in Chapter 6 and how you estimate them for sales data Chapter 8.

Problem 5: Tabs Don't Visualize Relationships

If a cross-tab table is large (i.e., bigger than a 2 x 2 table[1]), then it becomes very difficult to see relationships because people tend to have visual problems seeing such relationships in tables of numbers, especially large tables. For instance, in the table in Figure 1.3 on page 9 you can't tell which brands of yogurt are more closely associated (in a proximity sense) with the client's brand or which brands are more closely associated with each type of store. The best you can hope for is to spot a large number (frequency or percent) and then draw attention to it as if that number is the most important piece of information in the table. A graph of the table would be a tremendous help. This is what *correspondence analysis* produces. I discuss correspondence analysis and its associated graph in Chapter 5.

1 A 2 X 2 table has two rows and two columns so there are only four cells in the table.

Pie and Bar Charts for Data Analysis

Visualization Issues

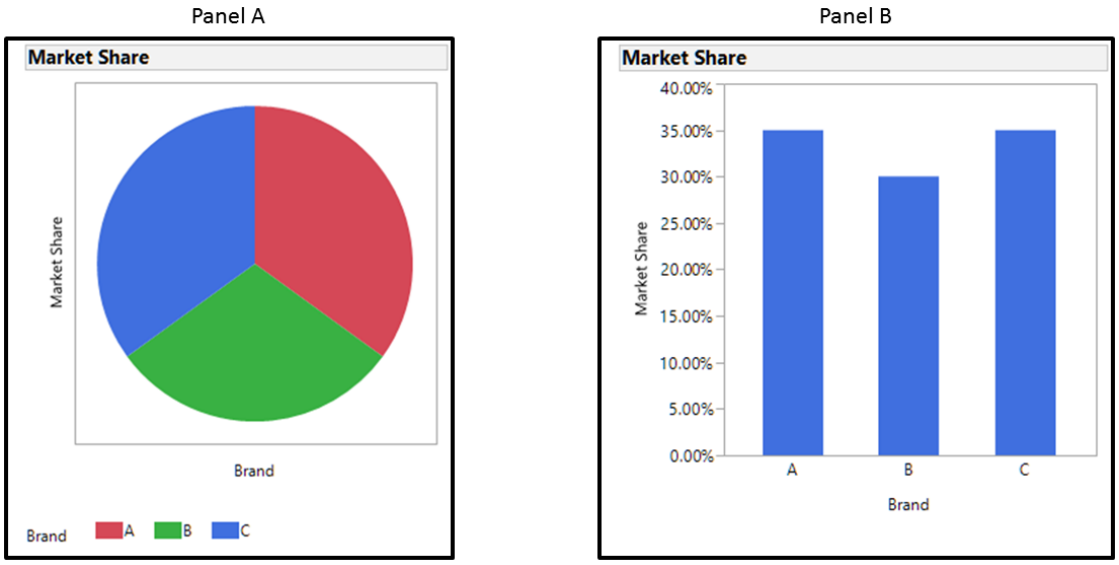
Pie and bar charts, used to display categorical data or data summaries such as proportions, are ubiquitous; everyone knows them. They're simple to produce and understand -- sometimes. Since most survey questions have categorical responses, these charts are liberally used. But they have problems.

Data visualization experts abhor pie charts. Tufte, probably the most dominant and best known data visualization expert, once wrote: "*The only thing worse than a pie chart is several of them*." (Tufte, 1983). Pie charts are not effective data displays because we have problems with angles, especially comparing them, just as we have problems with large tables. Lengths are easier for us to see and compare because they're linear; angles are not. This is illustrated in Figure 1.5 on page 14. The pie chart in Panel *A*, which shows market shares for three brands, illustrates the main problem with this type of display. Which has the smallest share? The actual shares are:

- A: 35%
- B: 30%
- C: 35%.

The bar chart in Panel *B* shows the same data, but now the difference in shares is clear. The bar chart shows the same data in a much clearer fashion. See Few (2007) for similar comparisons.

Figure 1.5 *Comparison of Pie and Bar Charts*



In addition to the angle issue, pie charts are frequently drawn with too many "slices," which becomes a cognitive challenge for deciphering key messages. This is compounded when you're asked to compare across several pies in a single slide or, worse yet, across multiple slides in a report.

Because of these problems, some visualization experts recommend replacing pie charts with bar charts as in Panel *B* of Figure 1.5 on page 14, another very popular, static chart that appears in most market research reports. Bar charts are viewed as "unrolled" pie charts—they contain the same information, but you can discern differences much faster and more accurately because of the linear nature of the bars. See Few (2007) for interesting comparisons. Also, bar charts can show more quantities than simple pie charts. For instance, they can show:

- counts or frequencies;
- means;
- proportions;
- sums;
- trends.

3-D effects are often added to bar charts (as well as to pie charts) to show depth, but this just counters their advantage because people have depth perception problems as well as angle problems.

Other Chart Forms Are Ignored

There are many other chart forms that can be used to display data. Some are:

- Box plots
- Bubble plots—to display three quantities at once. Colors add a fourth dimension—but not for color-challenged people!
- Dot plots
- Heat maps
- Line plots
- Mosaic plots
- Scatter plots
- Treemaps
- Panel (also called trellis) displays

JMP has a great graphing platform, *Graph Builder*, that not only enables you to create all these graph types but also enables you to create them dynamically, so that you can see your data from many different perspectives. In short, it actually allows you to analyze your data. I discuss *Graph Builder* in Chapter 5 and illustrate its use throughout the remainder of this book. See Cleveland and McGill (1984) for a more technical discussion of graphical displays.

Ignoring visual issues, pie and bar charts frequently replace analysis of market data. Many reports have a pie followed by a bar followed by a pie followed by—well, it keeps going on and on, so that there's one chart for each question. This is just reporting at worst and simple descriptive "analysis" at best so that the analysis and reporting of data are the same; the same tools are used to study and report the data. Information, however, is not extracted because a connection is not made between (or among) questions (variables) to show relationships, trends, patterns, and anomalies (i.e., outliers). This is the analysis of data.

The Chartjunk Issue

I would be remiss if I didn't comment on *chartjunk*, a term coined by Tufte (1983):

The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new. The purpose of decoration varies -- to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills. Regardless of its cause, it is all non-data-ink or redundant data-ink, and it is often chartjunk.

Microsoft PowerPoint and spreadsheet software (discussed next) make it very easy to create elaborate charts that look impressive but are not really statistical and don't add to the extraction of information from data. The chartjunk graphs pass the burden of analysis onto the report reader (i.e., the client or manager) by forcing him/her to sift through the junk to find the market information the study was designed to deliver in the first place. Statistical charts

and graphs should be used. See Su (2008) for a discussion of spreadsheet graphics that don't adhere to basic principles of statistical graphics and which thus overburden the reader. These principles have been well stated by others. See the references in Su (2008) and a slightly different perspective on chartjunk by Few (2011).

Spreadsheets As Data Managers and Analysis Tools

Overreliance on Spreadsheets

Spreadsheets are very heavily relied upon for analyses and database management. There are several major issues associated with their use. They're easy to use in general, but this is not a reason to use them as database managers. They're not database managers.

Figure 1.6 on page 16 shows an example of a section of a large spreadsheet typical of what you might receive from an online survey vendor. There are 18,492 rows or records, which represent that number of people, and 81 columns or variables on each person. There's no indication of what *Q1*, *Q2*, and so on are, no indication of their values, and no definitions for the *Code* variable. Sometimes documentation is provided, perhaps as another worksheet in the same workbook as the data, but often it's not. You have to refer to the questionnaire to find the definition for the variables, such as *Q1*, and even then, it's not clear what they are.

Figure 1.6 *Example of Spreadsheet Data*

id	Location	Code	Q1	Q2	Q3	Q4_1	Q4_2	Q4_3	Q4_4	Q4_5	Q4_6	Q4_7
001	CA	115	4	5	2	4	4	4	4	4	4	4
001	CA	124	4	5	2	5	5	4	4	4	4	4
001	CA	130	5	8	2	6	3	6	5	5	6	5
001	CA	139	3	3	3	3	4	3	3	3	3	3
001	CA	145	5	6	3	4	5	5	4	4	4	4
001	CA	149	5	6	2	5	5	6	5	5	5	5
001	CA	175	5	7	3	5	5	6	5	5	5	5
001	CA	184	4	5	3	4	5	4	4	4	5	4
001	CA	186	6	8	3	5	5	6	5	5	5	5
001	CA	199	5	6	3	5	5	5	5	5	5	4
001	CA	201	4	7	3	5	4	5	4	4	5	4
001	CA	206	2	4	3	3	4	2	3	3	4	3
001	CA	233	3	2	3	4	4	2	3	3	4	3

Spreadsheets are good as simple data structures such as *flat files*. Flat files have just a rectangular array of data with no links to other data; they have just simple rows and columns.

They're inadequate for complex data structures involving several tables, which are, effectively, a 3-D data cube. They lack:

1. **Variable documentation**, except by creating yet another spreadsheet, preferably in the same workbook.
2. **Value mapping** from numerics to descriptive labels.
3. **Variable grouping** for quick data location and management.
4. **Table operations** such as joining/splitting/stacking.
5. **Programming capabilities**, aside from *Visual Basic for Applications* (*VBA*), which is not a statistical programming language.
6. **Sophisticated statistical operations** beyond arithmetic operations and simple regression analysis. Add-on packages help, but they tend to lack depth and rely on the spreadsheet engine.

Cell-centric Problem

Spreadsheets are notorious for making it difficult to track formulas and catch errors because they're cell-centric. Each cell could have a separate formula; even cells in the same column for a single variable could have different variables. As an example, for the sample spreadsheet in Figure 1.6 on page 16, creating a new variable that is *Q1* + *Q2* requires creating 18,492 formulas, one for each cell. The chance for error is, of course, huge.

Auditing Problem

The cell-centric formula issue leads to an auditing problem. The cells in a spreadsheet are often linked to other cells, either across spreadsheets in the same workbook or across workbooks, and often with no clear pattern. Tracing and reproducing these links is difficult or impossible for very large spreadsheets so that auditing calculations becomes difficult. The chance of an undetected error is higher the more complex the spreadsheet.

Easy Graphs

Spreadsheets provide the ability to easily graph data. The static nature of the graphs, however, makes it difficult to quickly explore and test ideas. Often what is presented is the first and only analysis. There are no links back to the original data for further analyses. As an example, there's no ability to click a bar in a bar chart to identify or subset the data in that bar. The static graphs are limited. These simple charts, combined with the tabs, are often the sole forms of (descriptive) analyses.

You Can Do Better

The result of focusing almost exclusively on surveys, static univariate tabs, and spreadsheets with static pie and bar charts is that the wealth of information that could be gained about market forces is left untapped. Important relationships are left uncovered and unexplored, resulting in simple conclusions and recommendations. But you should and can do better.

Your goal should be to provide actionable information about market forces based on dynamic tools and sophisticated analyses, so your client or other business leaders can make the best informed information-driven decisions. You need to go beyond a cult of surveys, tabs, and static pie/bar charts, and move beyond just market research to researching the market with more sophisticated, dynamic analysis tools. So, an agenda item for this book is to promote more sophisticated research and analysis of markets with broader data and enhanced, dynamic tools.

This is where JMP enters the picture. I focus in Chapter 2 on how JMP can help you in this area.

Points to Remember

There are several main points to remember from this chapter:

1. Market researchers provide information about market operations, organization, needs, and performance.
2. Market researchers rely on four tools:
    a. **surveys**: overworked to the exclusion of databases;
    b. **tabs**: provide static snapshots of survey results that are difficult to change and are univariate;
    c. **simple charts**: static and simplistic;
    d. **spreadsheets**: error-prone and used for analysis and as database managers.
3. The goal of market researchers should be to provide actionable information about market forces with dynamic tools and sophisticated analyses.

Exercises

Research Concepts

1. What is the role of market research in business?
2. Distinguish between data and information.
3. What are the major data sources used by market researchers?
4. Compare and contrast the major data sources used by market researchers.
5. Describe some implications of a “need for speed” attitude in market research.
6. What is a “tab”? How are tabs used in market research?
7. List and discuss five problems with tabs.
8. How are the “need for speed” and tabs related?
9. What are two popular graphic displays of data? Summarize some issues with each one.
10. Compare and contrast dynamic and static graphs and tables.
11. Compare and contrast univariate and multivariate statistics. Give a marketing example of each.
12. How prevalent are spreadsheets? How are they generally used by market researchers? Summarize issues with using spreadsheets.

JMP Concepts

1. Name three data modeling types recognized by JMP.
2. Briefly state how JMP can aid you in gaining a comprehensive view of your data.
3. Describe how JMP can help you extract information from raw data.

Works Cited

Cleveland, William S., and Robert McGill. (September 1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, Vol. 79, No. 387, pp. 531 - 554.

Few, Stephen. (2007, August). *Save the Pies for Dessert.* Retrieved from Perceptual Edge: http://perceptualedge.com/library.php

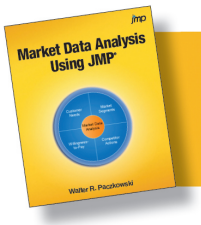
Few, Stephen. (2011, April/May/June). *The Chartjunk Debate: A Close Examination of Recent Findings.* Retrieved from Perceptual Edge: http://perceptualedge.com/library.php

Su, Yu-Sung. (June 15, 2008). It's easy to produce chartjunk using Microsoft Excel 2007 but hard to make good graphs. *Computational Statistics & Data Analysis*, Vol. 52, No.10, pp. 4594–4601.

Tufte, Edward R. (1983). *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

Wheeler, Schaun. (2012, April 2). *Surveys, Assumptions, and the Need for Data Collection Alternatives.* Retrieved May 22, 2015, from R-Bloggers: https://www.r-bloggers.com/surveys-assumptions-and-the-need-for-data-collection-alternatives/

Young, Forrest W., Pedro M. Valero-Mora, and Michael Friendly. (2006). *Visual Statistics: Seeing Data with Dynamic Interactive Graphics.* Hoboken, NJ: John Wiley & Sons.



About This Book

Purpose

Businesses are drowning in data. Market data, whether from surveys or a data warehouse or data mart, must be analyzed dynamically to reveal patterns and relationships. Simple static tables and pie and bar charts don't suffice. To convert market data into actionable, data-driven decisions, you need software that offers three key features: powerful data handling and scripting tools; and dynamic graphing and tabulation capabilities to drill down and link data; market-oriented statistical methods, such as choice models.

JMP® gives you access to the right data-handling tools, a rich array of analytical methods, and dynamic views of your data so that you can drill-down and link the data across several displays for comprehensive views. Regardless of your level of experience, it offers tools and analysis reports that are accessible to novices and sophisticates alike.

The goal of this book is to introduce JMP to the market research community and to explain how to use JMP to implement some statistical techniques that can increase the power and sophistication of the data analysis process.

Is This Book for You?

This book is for data analysts, who, like me, are charged with analyzing market data. These economists, statisticians, and market researchers must estimate models, look for relationships, and produce insight from data to help business leaders make data-driven decisions.

It's also for students and professors in quantitative methods courses that deal with market data. Students who focus on quantitative methods can learn about the high-powered statistical methods in JMP and can discover how the graphing and tabulation capabilities in JMP can help them easily and dynamically gain insight into their data.

This book is slightly technical at times, but not greatly so. An undergraduate background in statistics (through regression analysis) and perhaps an econometrics course are helpful. The

focus is on showing how to use JMP with market data. Therefore, it's assumed that basic statistics, probability theory, and regression analysis are understood.

Scope of This Book

Here's the general outline of the book:

- Chapters 1–3 set the stage. They cover the state of market research, describe how JMP can help, and introduce JMP, for those unfamiliar with it.
- Chapters 4 and 5 describe the JMP features that are useful for preparing data for analysis. They discuss analyzing survey data with JMP.
- Chapters 6 and 7 apply JMP to specific business and marketing problems. They cover using choice models to study consumer behavior and using survey data to segment a market.
- Chapter 8 shifts the focus from survey data to sales data from a data warehouse or data mart. Emphasis is placed on data visualization using the *JMP Graph Builder*.

Exercises are included at the end of each chapter. Solutions are available at the author page at http://support.sas.com/paczkowski. The exercises are divided into three groups:

- Research concepts covering the technicalities introduced in the chapter.
- JMP concepts covering the main tools of JMP that can be used for analyzing market data.
- JMP problems that use the research concepts and JMP concepts to address data problems.

Some of the examples in this book come from my work with market data related to consumer products, but the material and methods described can be used with other types of market data. With one exception, the data used in this book is simulated. In Chapter 3, I use the Big Class data set that comes with JMP.

Throughout, JMP commands are capitalized and italicized. Menu commands, also capitalized and italicized, are run together and separated by a forward slash. For example, *File/Preferences* refers to the File menu on the main menu bar and the Preferences submenu. Data table column names are also italicized.

In this book, as in any human endeavor, you might find a mistake. Please let me know about anything I missed by contacting me at walt@dataanalyticscorp.com.

Keep in Touch

We look forward to hearing from you. We invite questions, comments, and concerns. If you want to contact us about a specific book, please include the book title in your correspondence.

To Contact the Author through SAS Press

By e-mail: saspress@sas.com

Via the Web: http://support.sas.com/author_feedback

SAS Books

For a complete list of books available through SAS, visit http://support.sas.com/bookstore.

Phone: 1-800-727-3228

Fax: 1-919-677-8166

E-mail: sasbook@sas.com

SAS Book Report

Receive up-to-date information about all new SAS publications via e-mail by subscribing to the SAS Book Report monthly eNewsletter. Visit http://support.sas.com/sbr.

Publish with SAS

SAS is recruiting authors! Are you interested in writing a book? Visit http://support.sas.com/saspress for more information.

About the Author

Walter R. Paczkowski is a market research consultant focused on helping companies turn their market data into actionable market information in a wide range of industries, such as telecommunications, pharmaceuticals, jewelry, food and beverage, and automotive. With nearly 40 years of extensive quantitative experience as an analyst in AT&T's Analytical Support Center, a member of the technical staff at AT&T Bell Labs, head of pricing research at AT&T's Computer Systems Division, and founder of his consulting company Data Analytics Corporation, he brings a wealth of knowledge to share about data analysis. He is currently an adjunct professor in the Department of Economics at Rutgers University, and is also an adjunct professor in the Department of Mathematics and Statistics at The College of New Jersey. He received his PhD in Economics from Texas A&M University.

Learn more about this author by visiting his author page at http://support.sas.com/paczkowski. There you can download free book excerpts, access example code and data, read the latest reviews, get updates, and more.