Exploring Customer Behavior: Survey Analysis Best Practices

Walter R. Paczkowski, Ph.D.

Data Analytics Corp. and Rutgers University

Key Question

What are the Best Practices in survey data analysis?

Best Practices for Data Analysis

Data analysis is:

1. Dynamic
2. Complex Multivariate

for the sole purpose of providing key decision makers with useful, insightful, and actionable information.

Decision Makers Need Information

Private and public sector leaders need *information* to make decisions:

Private Sector

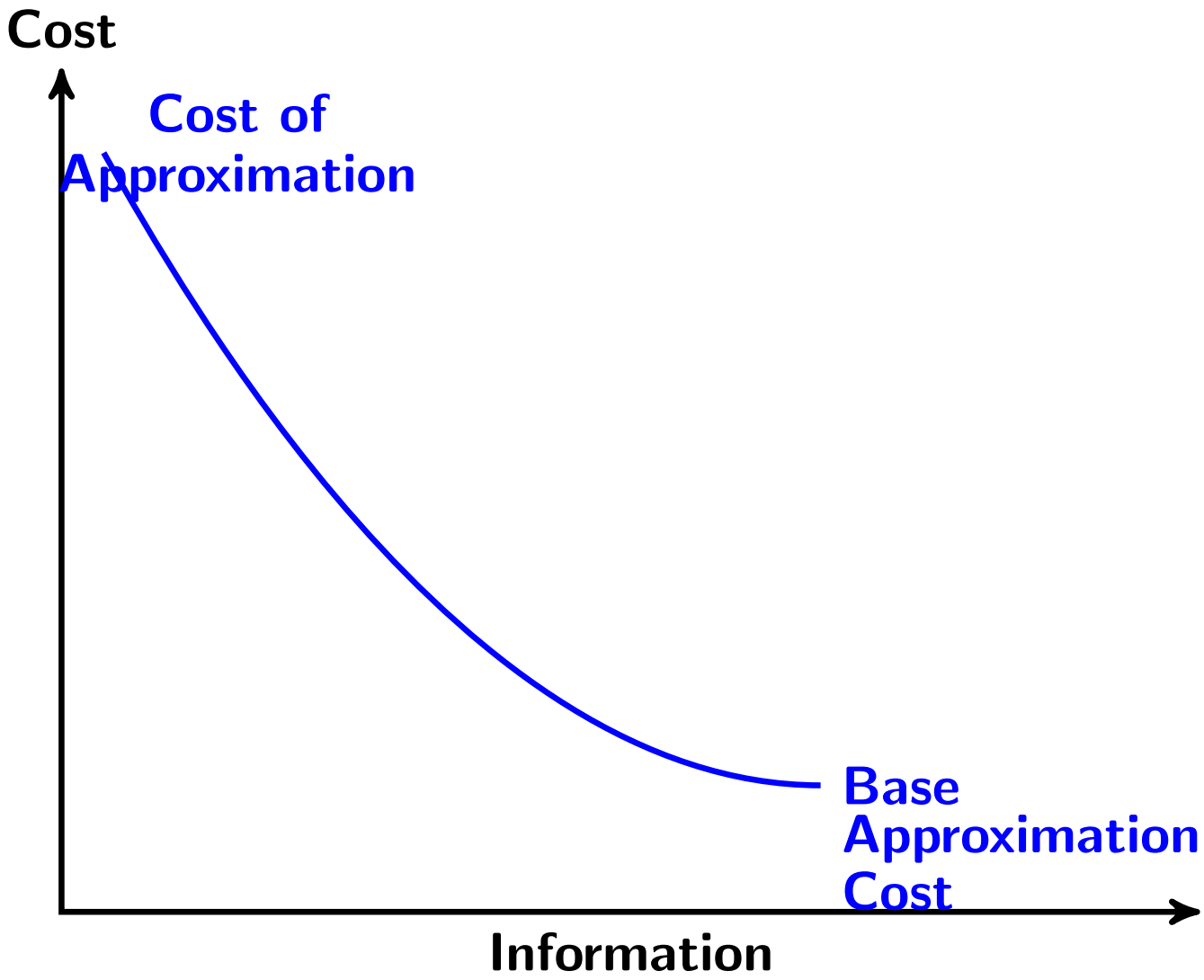
- What product to sell.
- How much to charge.
- How to promote the product.

Public Sector

- What services to provide.
- How much to increase/decrease taxes.
- What policies to implement.

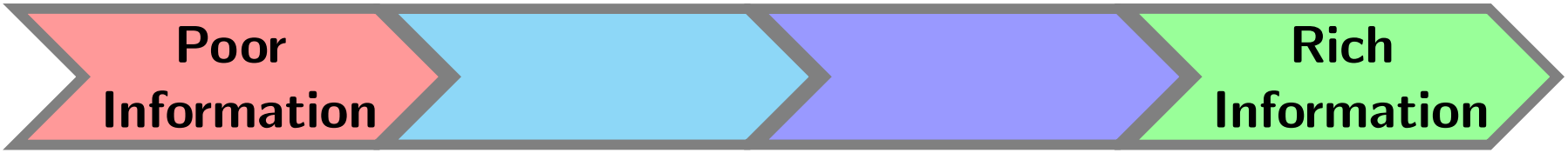
Without information on any of these, we have to approximate (i.e., guess) what would succeed – guessing is costly.

There's No Such Thing As A Free Lunch



Information At Any Stage Is Not Binary[1]

Information Continuum



Poor Information

- Raw
- Disorganized
- Fuzzy
- Unfiltered

Analytical Bridge

- Calculations
- Cleansing
- Modeling
- Visualization

Rich Information

- Insightful
- Organized
- Clear
- Filtered

[1] Based on Zahay, D., Griffin, A. & Fredicks, E. 2004. "Sources, uses, and forms of data in the new product development process". *Industrial Marketing Management*, 33, 657-666.

Data Are Poor Information

Data are raw, unfiltered, disorganized pieces of material (i.e., "stuff").

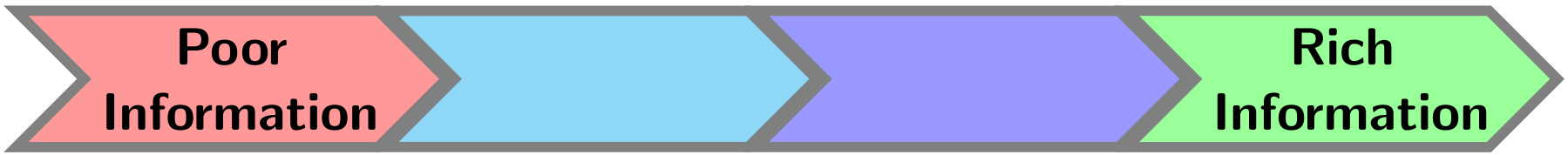
- "Stuff" is what we collect – and store in a closet.
  - Modern day data closets are computer files/folders, data marts, and data warehouses – and data lakes!
- Data are like Lego bricks that have to be assembled.
  - Like the bricks, they can be assembled in infinite ways reflecting our creativity and questions.
- *Rich Information* is insight built and extracted from data by creatively manipulating the data bricks.

The Analytical Bridge

Data Analytics is the process of taking raw data bricks/Poor Information and assembling/converting them into *Rich Information* using different data views.

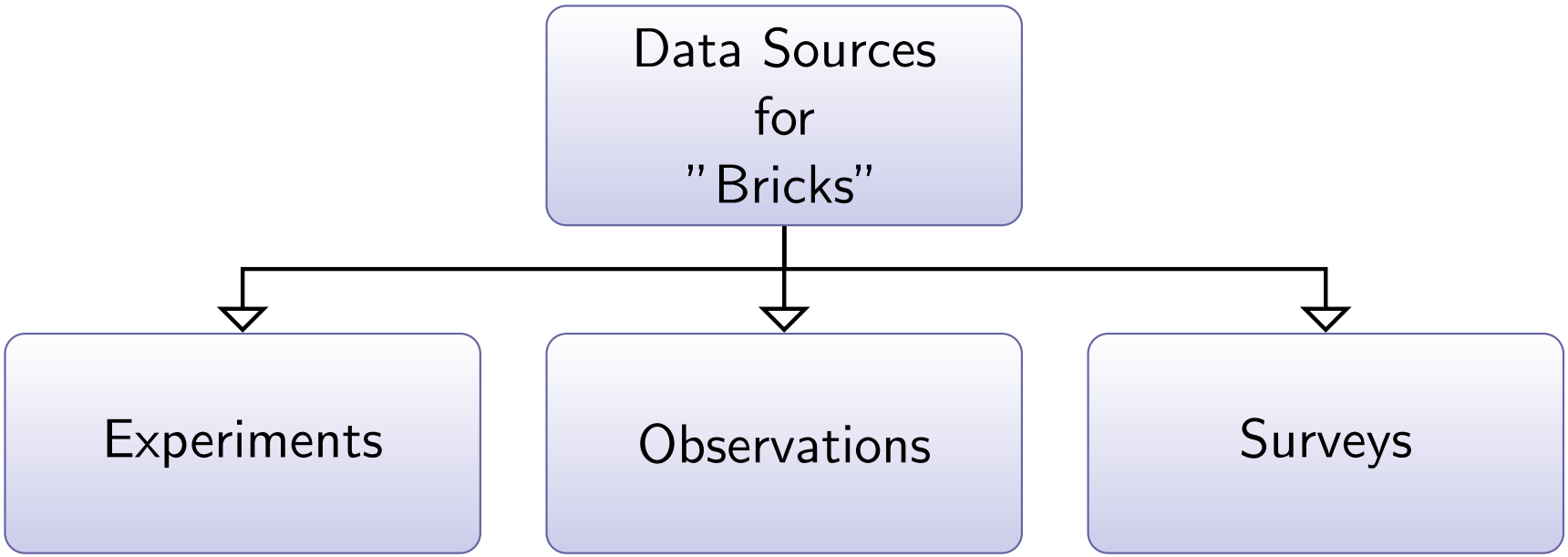
Yogurt Information Continuum

Information Continuum: Yogurt Purchases



- Simple Plot
- Demand Curve
- Outliers
- Conjecture
- Model
- Elasticities
- Segments
- Elasticities

The Data Must Come From Somewhere



- Controlled
- Possibilities
- New products

- Transactions
- Past Behavior
- Existing products

- Opinions
- Past/Current Behavior
- New concepts

My comments apply to any type of data, regardless of the source, but the focus is on survey data, the most common form in market research.

Guiding Principle

Survey data bricks have to be analyzed to assemble them into useful, insightful, and actionable *Rich Information*.

What Is Data Analysis?

To analyze means to break into parts.

Greek root: *analusis*

"a breaking up, a loosening, releasing." [a]

[a] Source: http://www.etymonline.com/index.php?term=analysis

Analyzing survey data means looking for

- relationships
- trends
- patterns
- anomalies

beneath the surface of the obvious and then to assemble the pieces from the analysis into a report. The report merely encapsulates the information.

Examples Of The Four Parts Of Analysis

Example

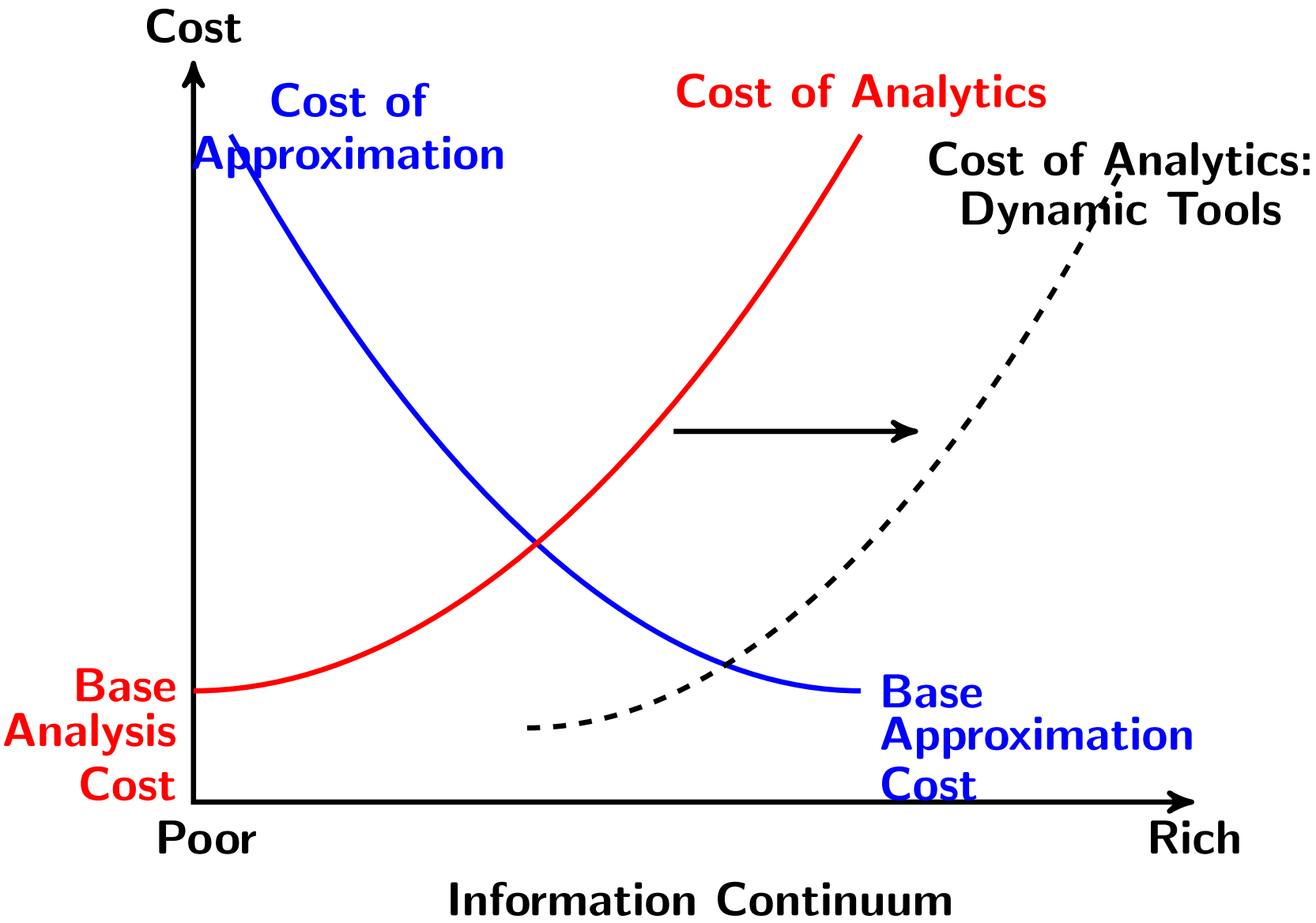
1. Relationships
   - Correlations between products purchased and distribution channels
   - Satisfaction or purchase intent key drivers
2. Trends
   - Developments or changes over time (e.g., a tracking study)
   - Before/after treatment
3. Patterns
   - Groupings of objects
   - Segments
   - Brands by segments
4. Anomalies
   - Outliers
     - Not all outliers are created equal
     - Some are innocuous; others are pernicious

The Four Parts of Analysis: Yogurt Data

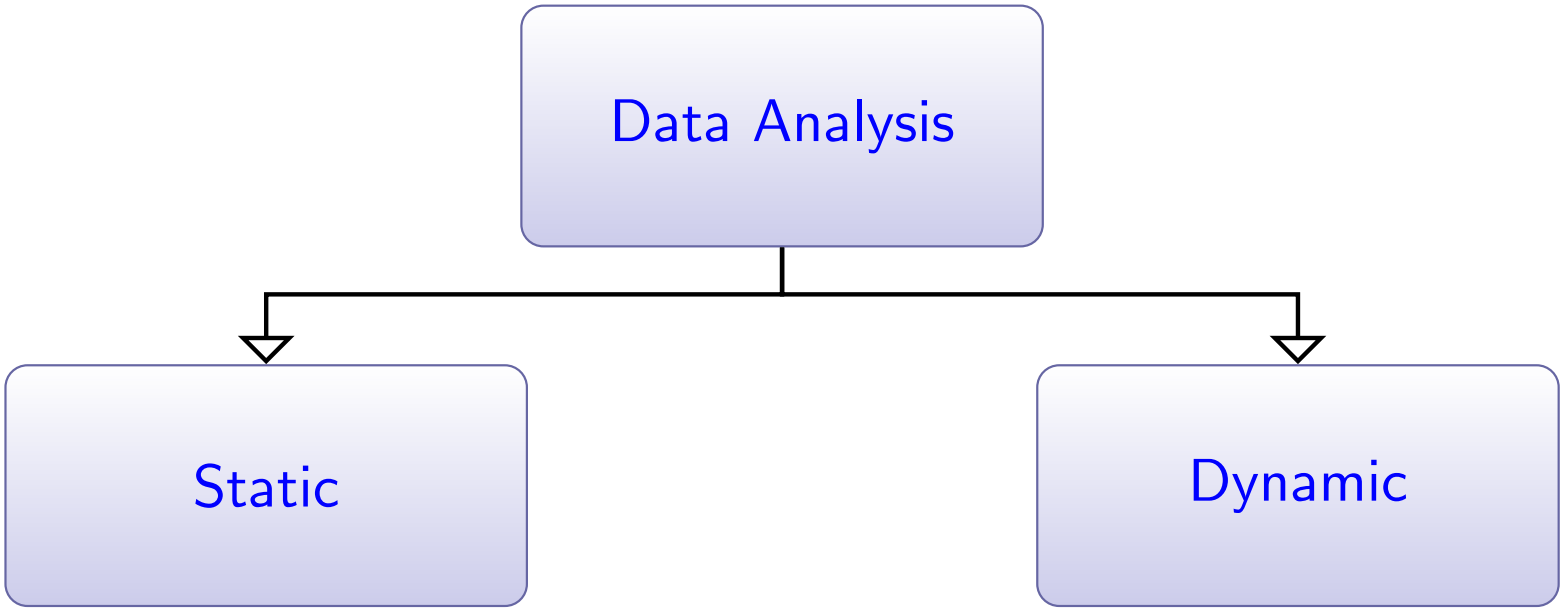
Example

1. Relationships
   - Negative relationship between price and quantity.
2. Trends
   - Not relevant since cross-sectional data.
3. Patterns
   - Groupings (albeit, hidden at first) of respondents in two segments.
4. Anomalies
   - Outliers in the tails of the distributions.

There's No Such Thing As A Free Lunch



Not All Analyses Are The Same



- Perfectly non-interactive
- No drill-down
- Just snapshots

- Change views
- Highly interactive
- Reveals hidden

Advantage of Dynamic Analysis

Static analyses leave untapped information.
Dynamic analyses reveal *Rich Information*.

Analyze Data Using Dynamic Tools

You should dynamically analyze data by:

- *Linking* tables and graphs for drill-down.
- *Filtering* data to subset them based on other variables.
- *Dragging* & *Dropping* variables onto and around a canvas, either tabular or graphic.
- *Building* models and calculating univariate/complex multivariate statistics.
- *Profiling* to study or test scenarios.

Most Analysis Tools Are Static

There are two typical static analysis tools:

1. Tables
   - Simple, and also extensive, tables combined with graphs are often the sole forms of "analyses."
     - The reports are created at the same time as the "analysis" – they are the same.
   - Tables do not provide dynamic views of data.
     - Most reports mimic "the tabs" which are simple pre-programmed cross-tabulations at best.

Former Client

"*Everything you need to know is in the tabs.*"

2. Graphs
   - The static nature of graphs makes it difficult to quickly explore and test ideas.
     - No links back to the original data for drill-down or filtering.

Static Analysis Is Almost Always Univariate

Univariate analysis involves one variable, question, or question type (e.g., *CATA*)

- Calculate simple means and proportions.
- Also calculate "simple multivariate" summaries (e.g., correlations).

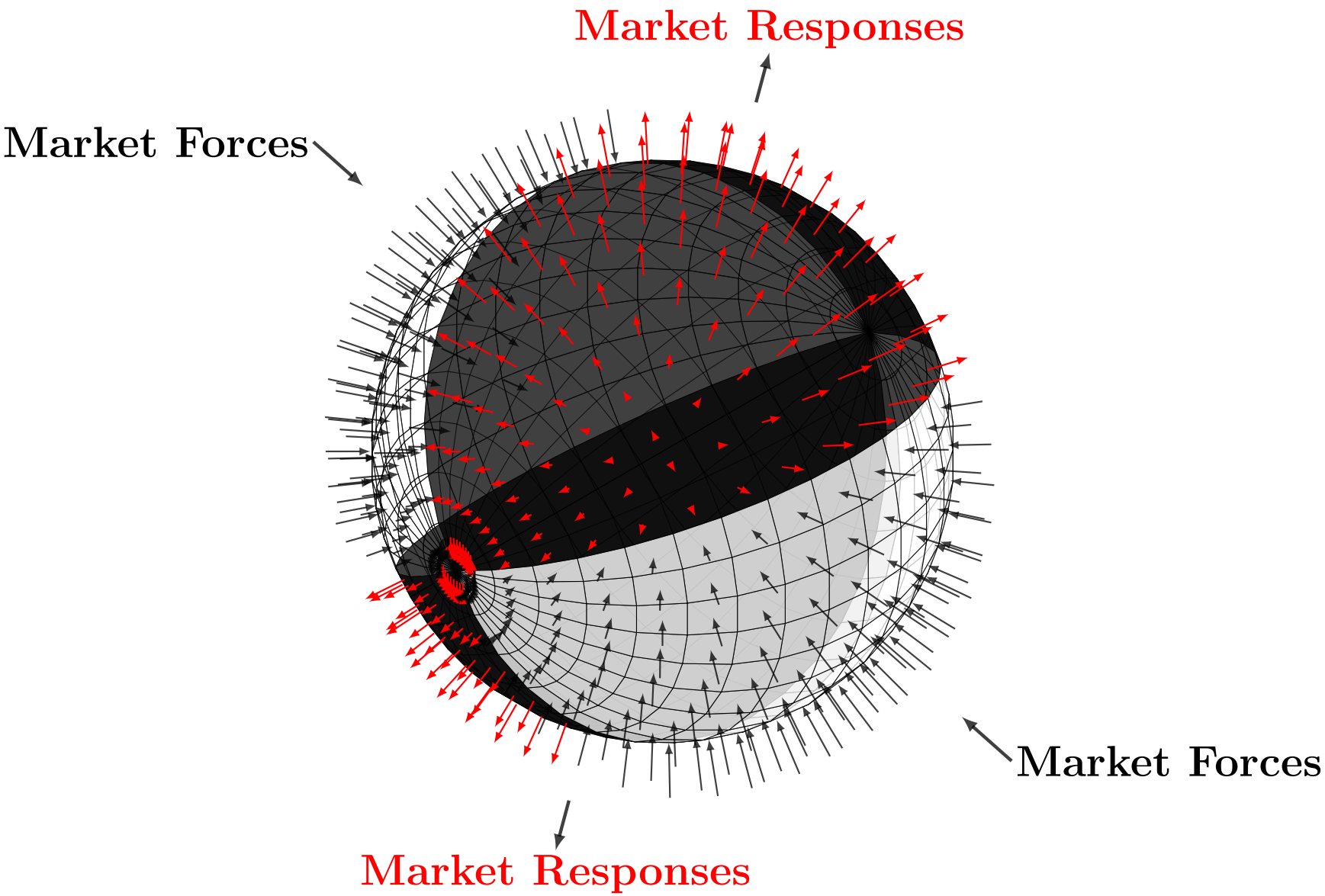
Result Is Oversimplification

A univariate or "simple multivariate" approach simplifies the world (i.e., markets).

Problem

The world is "complex multivariate."

The Economy Is Complex With Many Opposing Forces



The Result: Understanding Is Lost

Focusing almost exclusively on static univariate or simple multivariate analyses results in a wealth of untapped understanding.

- The myriad of market forces is ignored.
- Poor Information is presented.
- Important relationships remain uncovered, unexplored, buried beneath the surface of the data ocean.
- Simple or naive conclusions and recommendations result.

Analysis Goal

The goal of researchers should be to provide decision makers with

- useful
- insightful
- actionable

Rich Information based on dynamic tools and sophisticated, complex multivariate analyses.

Key Messages

Best Practices for Data Analysis

Dynamic, complex multivariate data analysis provides Rich Information.

Important
- Information is important for decision making

Continuum
- Information is a continuum, not binary

Analyze
- Data are converted to *Rich Information* through analyses

Costs
- Reduce costs with dynamic, complex multivariate analyses

Best Practices Case Study

Survey[a] of yogurt:

- consumption patterns;
- flavor and brand preferences;
- brand loyalty;
- competitive landscape.

[a] Based on fictional data for illustrative purposes only.

Study Objectives

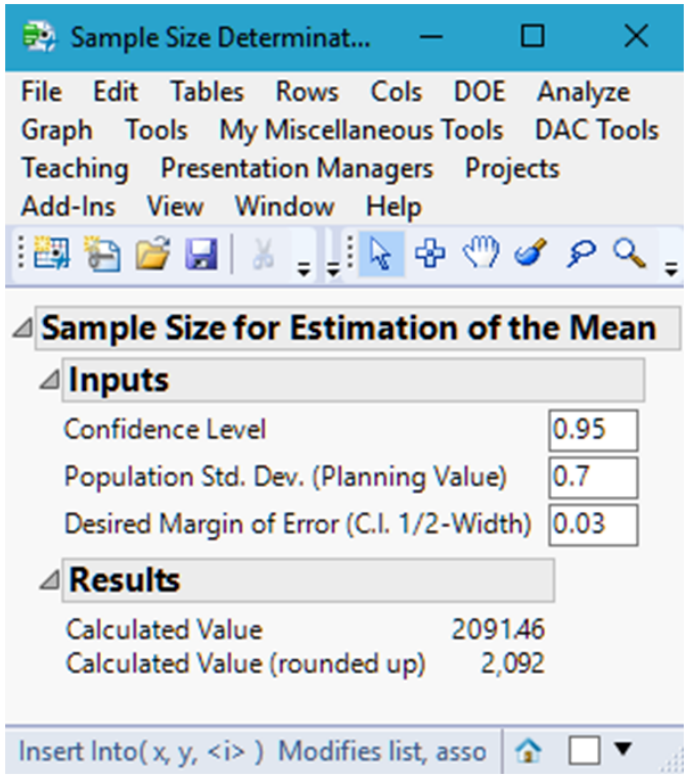
The yogurt company's marketing team set several objectives:

1. Understand shopping behavior:
   - price paid
   - amount purchased
   - where shop
   - brands purchased.
2. Understand yogurt preferences: flavors.
3. Understand consumer segments.
4. Understand shopping behavior by segments.
5. Understand key drivers for buying client brand.
6. Understand brand loyalty.

Sample Size

Sample size was determined using a JMP script[2]:

- A 95% confidence interval with a 3% margin of error.
- $N = 240,780,262$ market consumers from industry reports.
- $\sigma = 0.7$ for the price paid for yogurt from prior studies.



Recommended: $n = 2,092$. Final usable: $n = 2,000$.

[2] *SampleSizeForConfidenceIntervals.jsl* which is bundled with JMP installation.

Data Collected

The 2,000 consumers were asked:

- the brand of yogurt they purchase;
- where they buy it;
- how much they pay;
- how much they buy;
- their favorite flavors;
- attribute importance and satisfaction;
- overall brand satisfaction;
- demographic questions (e.g., gender, age, household income) to profile these respondents.

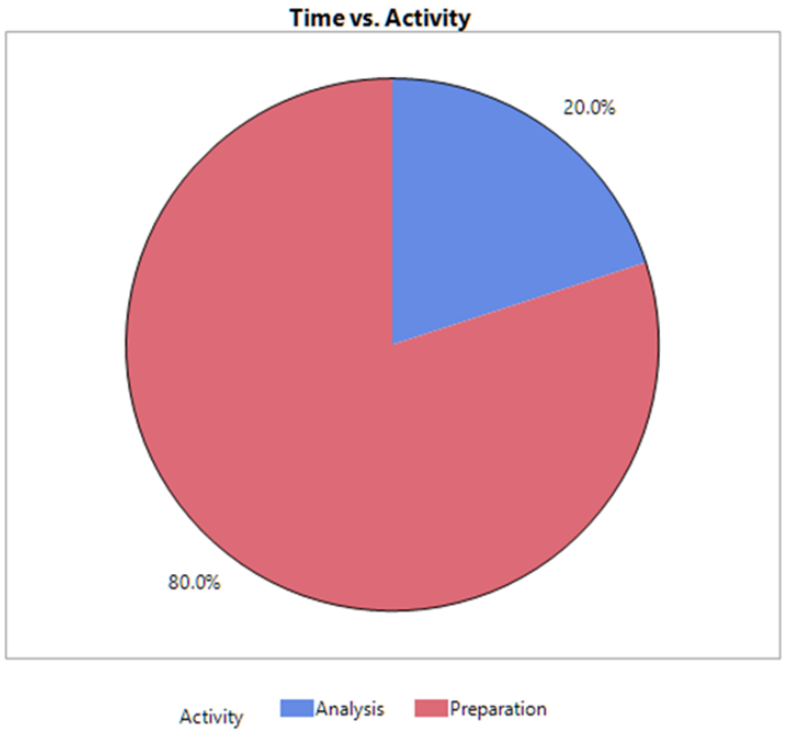
There are six brands: the client's brand, a major competitor, and four others.

Time Spent On Data Preparation

Most of your time is spent on data preparation, not on planning, collecting, analysis, reporting.[3]

Preparation involves:

- Importing.
- Documenting.
- Grouping.
- Wrangling.
- Modifying and Transforming.



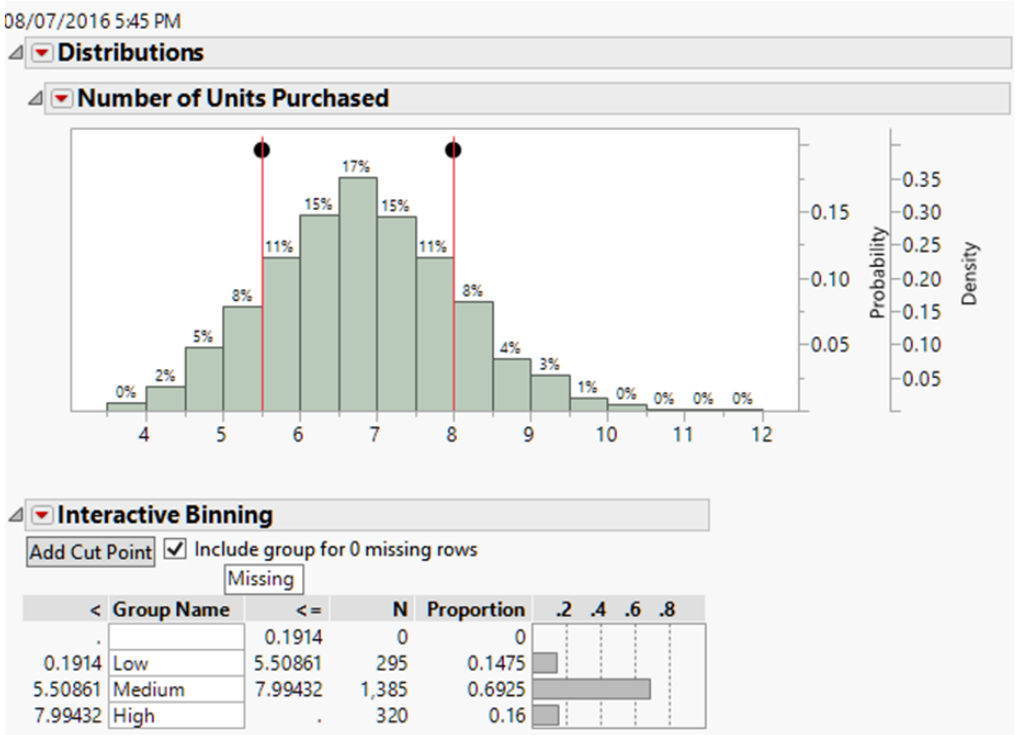
This 80/20 rule held here.

[3]Source: Dasu T and Johnson T. *Exploratory Data Mining and Data Cleaning*. New York: John Wiley & Sons, Inc, 2003.

Data Preparation

Created two variables for later use.

1. Number of units purchased groups.
   - Used JMP Add-in: *Interactive Binning*.
     - Low: 5 or less
     - Medium: 6 - 8
     - High: 9+
2. Buyer/Non-Buyer of the client's brand.

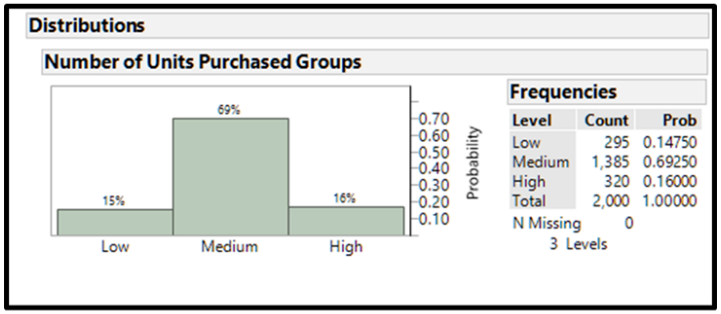


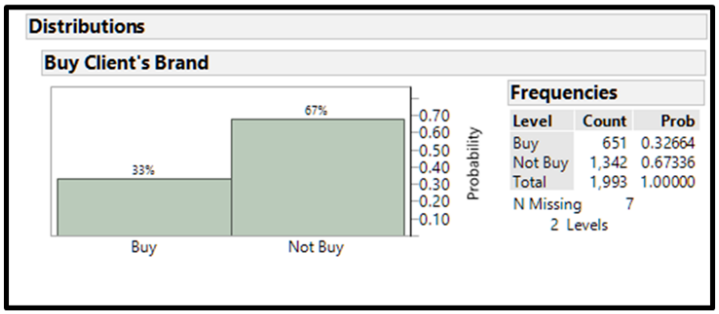
Segments were also added based on a prior segmentation study.[4]

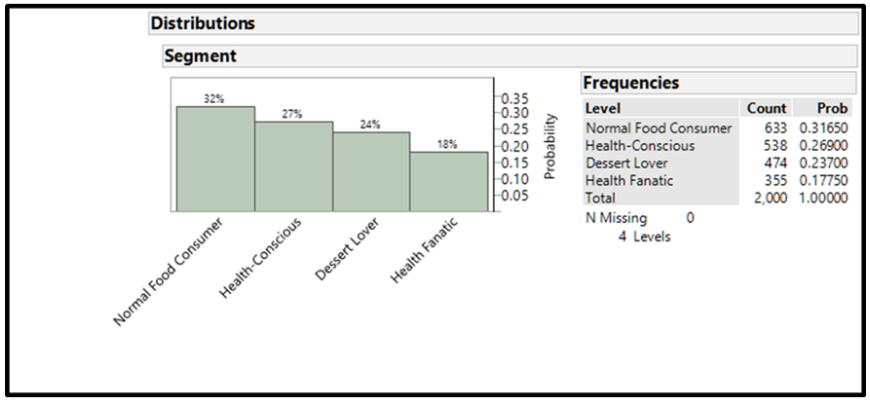
[4] Binning add-in available at `https://community.jmp.com/docs/DOC-6237`

Data Preparation - Distributions

Some key distributions







Structure For This Section

I'll look at four sections of the survey with respect to four study objectives:

1. Flavors selected
2. Brand Preferences by Segments
3. Brand Selection
4. Loyalty
    - Derived Importance
    - Net Promoter Score (NPS)
    - Loyalty Indexes

Live Demonstrations

Best Practices Key Message

Best practices use dynamic, complex multivariate analyses of data with statistical, tabulating, and graphing capabilities supported by *LDF*s.

Data Analytics Corp.

Walter R. Paczkowski, Ph.D.
President
Walt@DataAnalyticsCorp.com

44 Hamilton Lane, Plainsboro, NJ 08536
Voice: 609-936-8999
Fax: 609-936-3733