



The **Market Research Event**

IIR's Yearly Forum on Best Practices, Key Trends and New Technologies Essential for Your Evolving Role in the Business

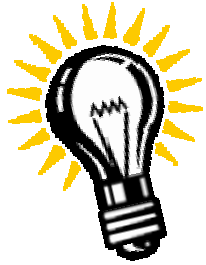
Best Segmentation Practices and Targeting Procedures that Provide the most Client-Actionable Strategy

Frank Wyman, Ph.D.
Director of Advanced Analytics
M/A/R/C[®] Research



Segmentation Defined

Segmentation: The dividing of a market's customers into subgroups in a way that optimizes the firm's ability to **profit** from the fact that customers have different needs, priorities, and economic levers.

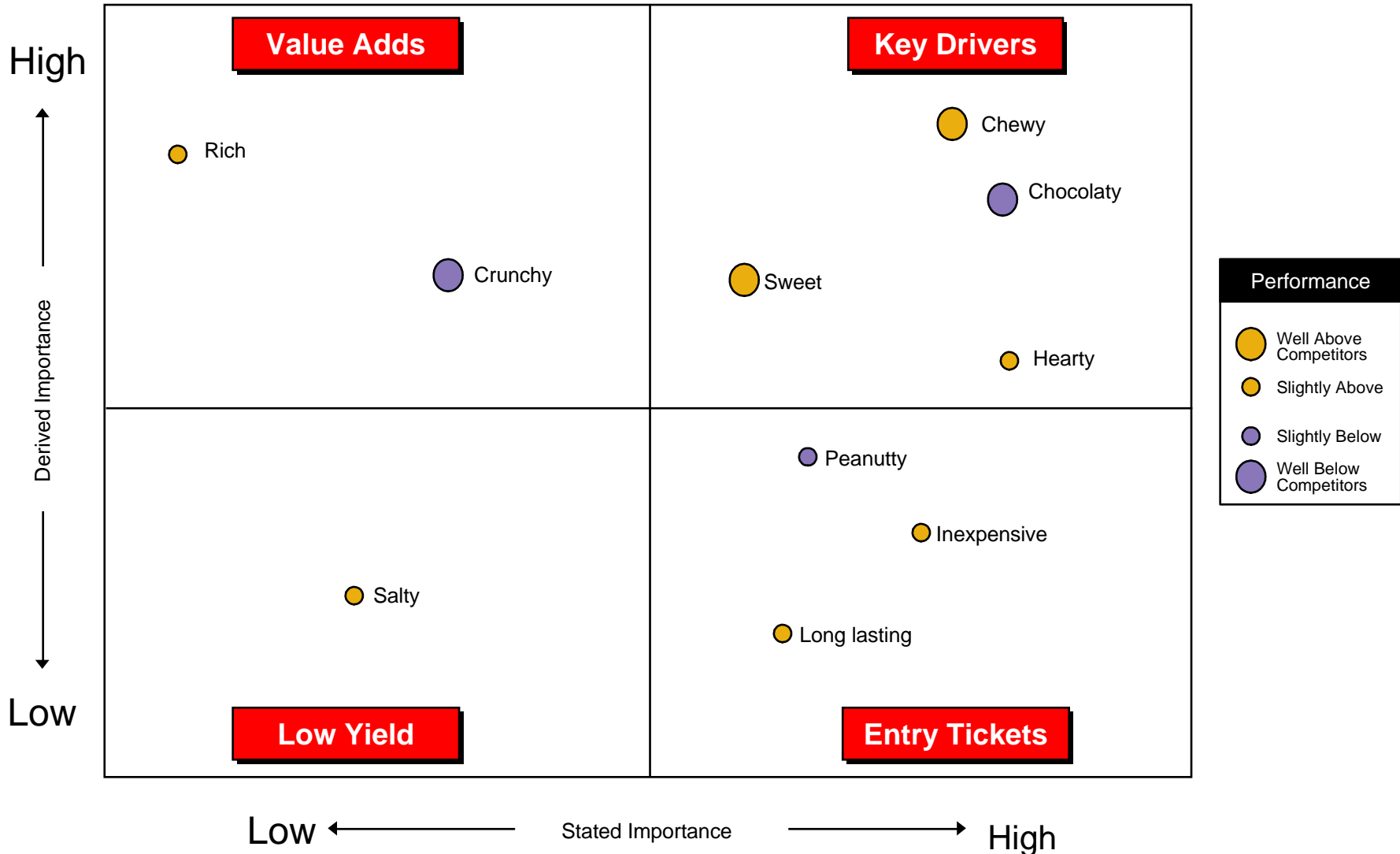


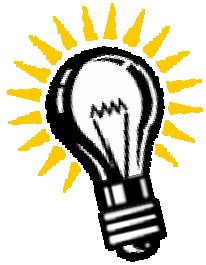
Best Practice #1

Keep in mind the end goal of enhancing **profitability**, as this can help increase the actionability of the segmentation. At each step ask “how can these results help improve profits?”



Product Positioning Research



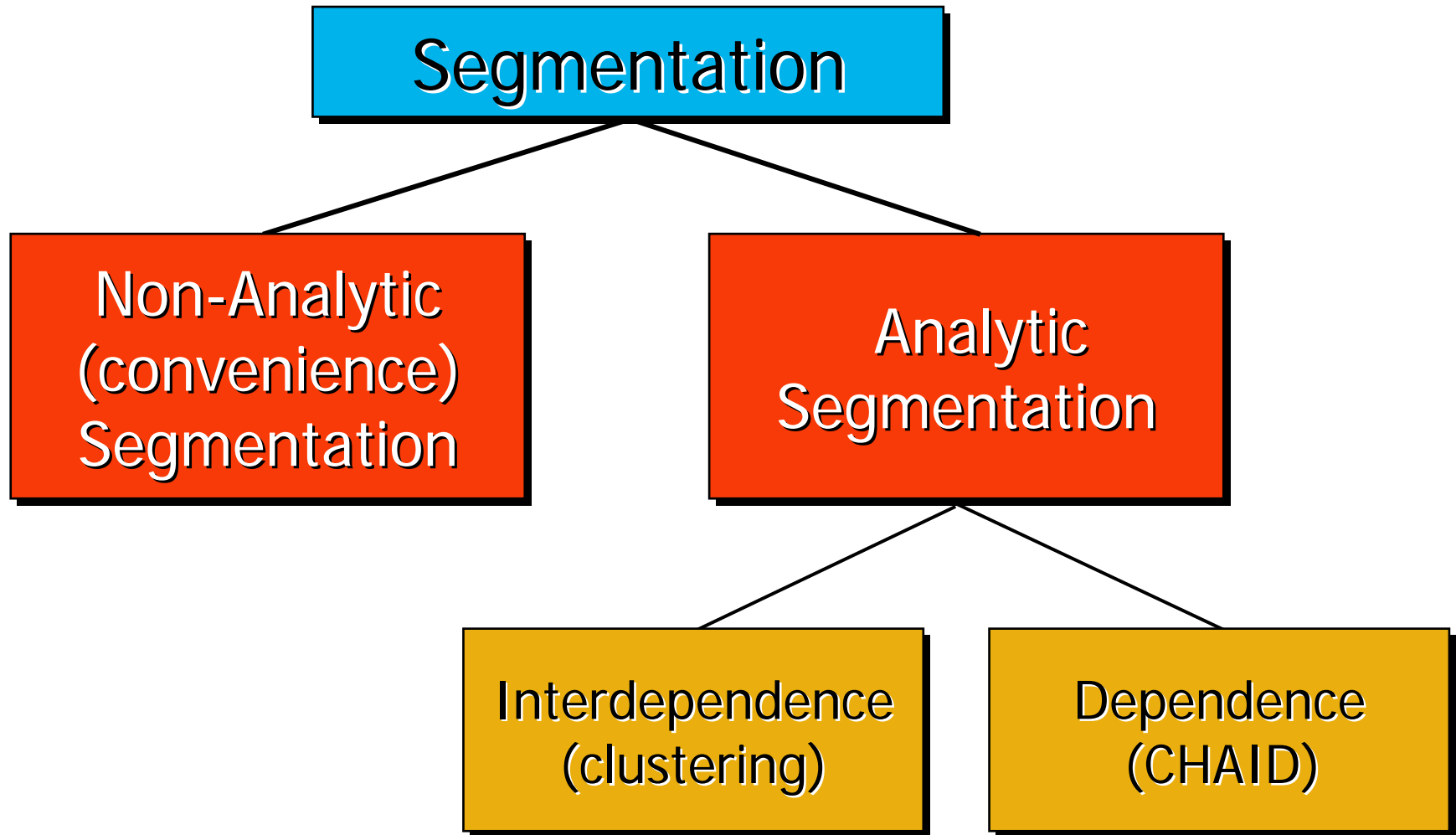


Best Practice #2

Conduct product positioning research around the same time as segmentation to improve the interpretability and enhance the actionability of the segmentation.

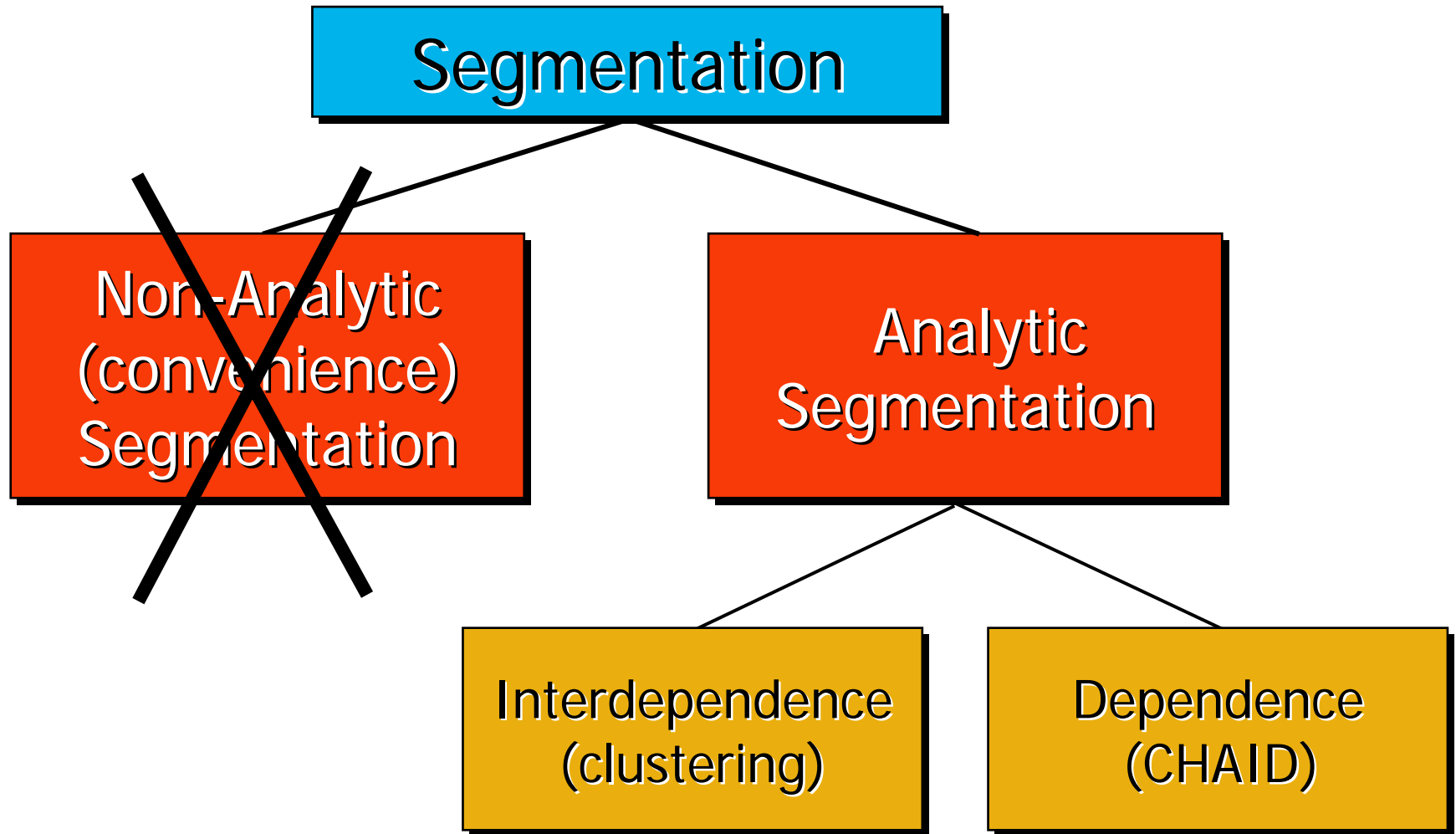


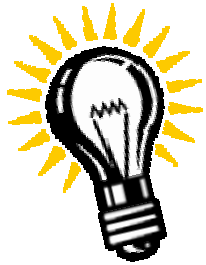
Three basic approaches to segmentation





Three basic approaches to segmentation



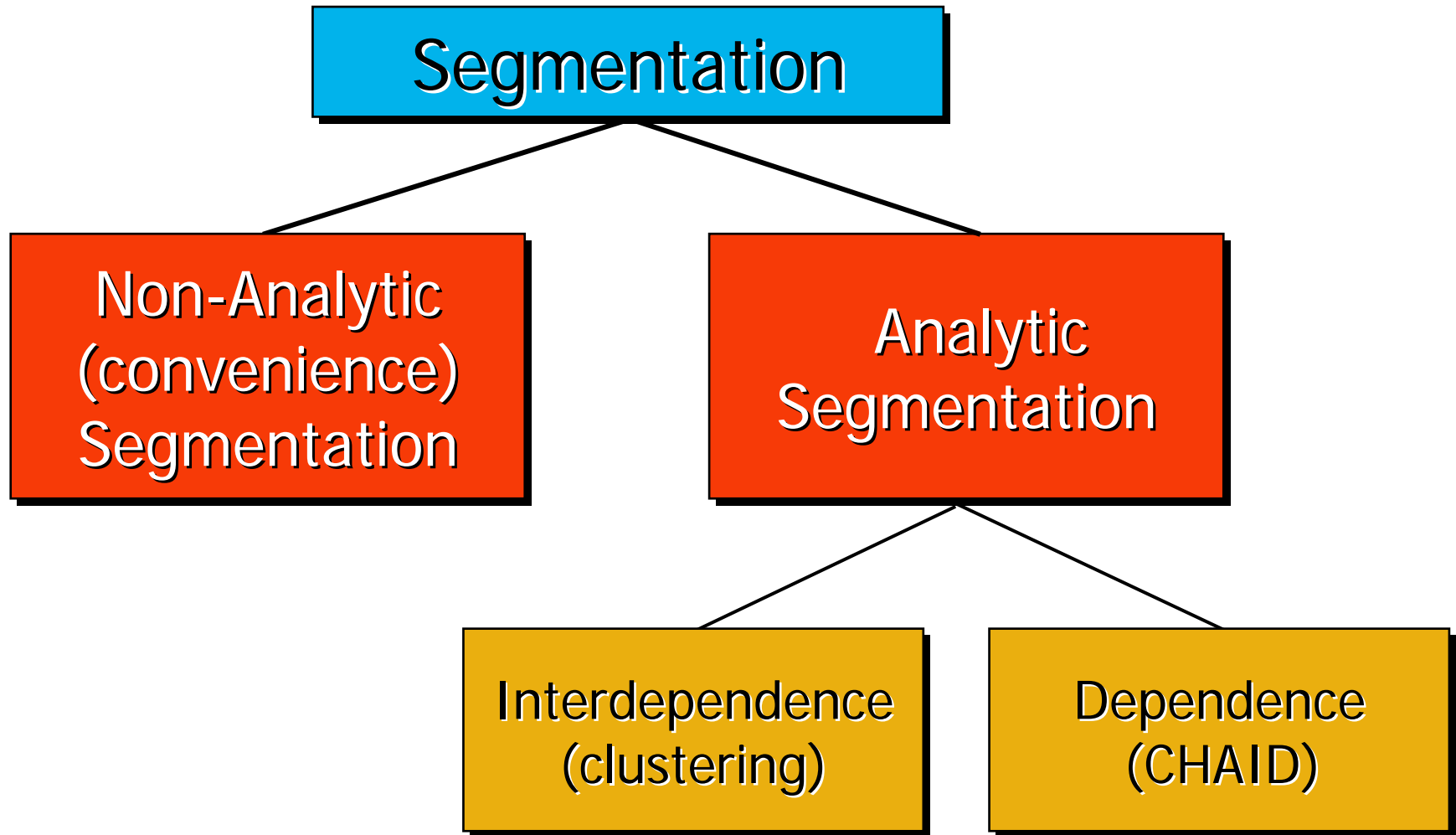


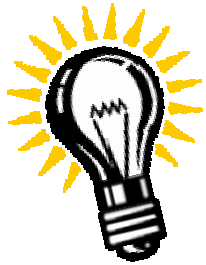
Best Practice #3

Choose to take an analytic approach to segmentation.



Three basic approaches to segmentation





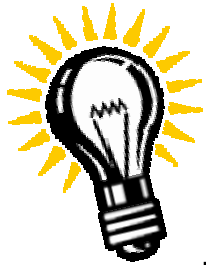
Best Practice #4

- Strategy => Interdependence (clustering)
- Tactics => Dependence (CHAID)



Considerations in deciding between an interdependence or dependence approach, or both

- Is a primary objective “to know the entire market” (what are the needs of customers, how many needs groups exist, and how large is each)?
- Is a primary objective “to know how to target those who will buy my offering/product/service”?
- How old is the market?
- To what degree are customer needs being met?
- How old is your product (existing, new launch, line extension)?
- How crowded is the market?
- How much room for differentiation is there?
- How dynamic is the market (how much has the market changed since your last segmentation)?

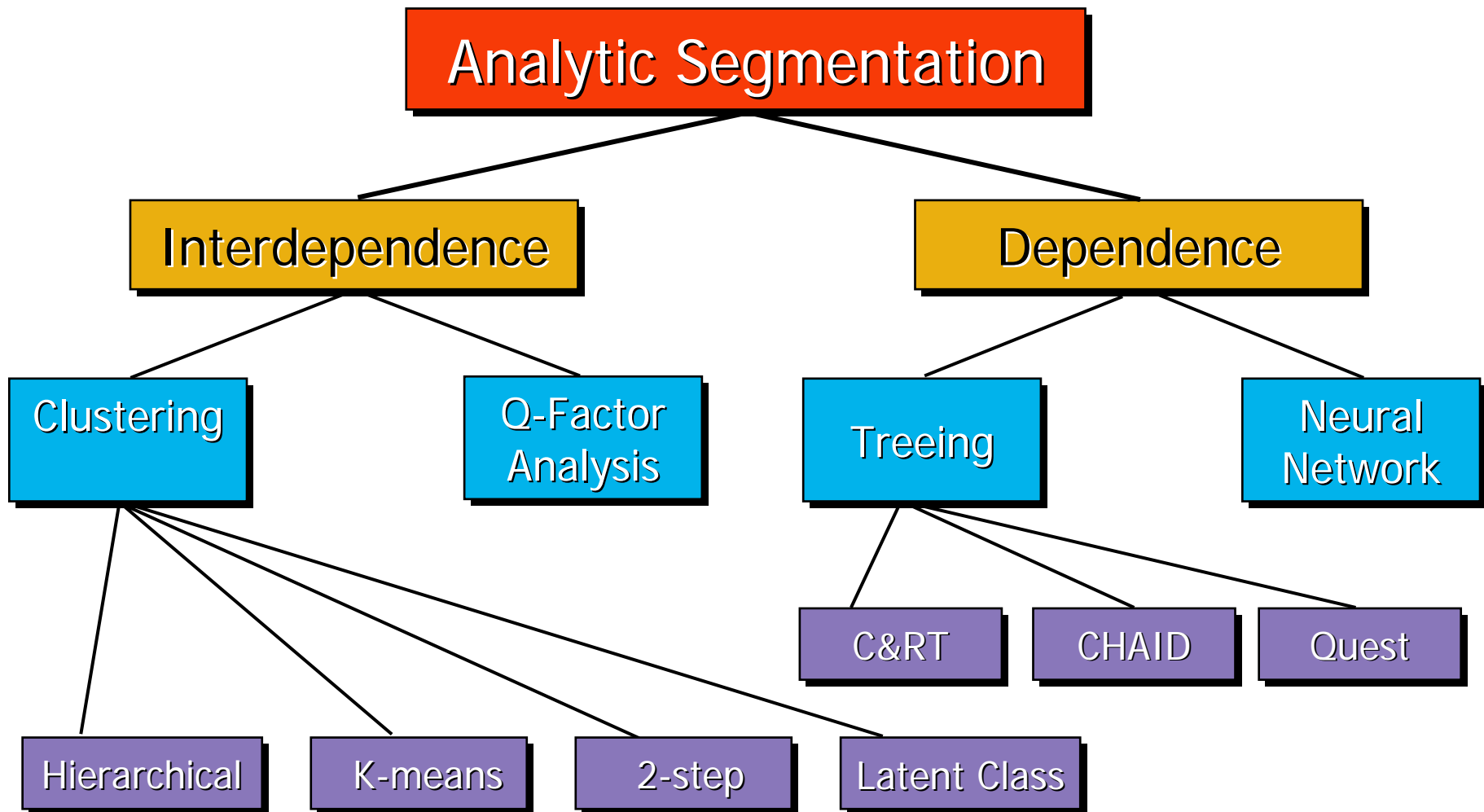


Best Practice #5

Let the parameters of your market and product offering determine which of the 2 analytic approaches to take.

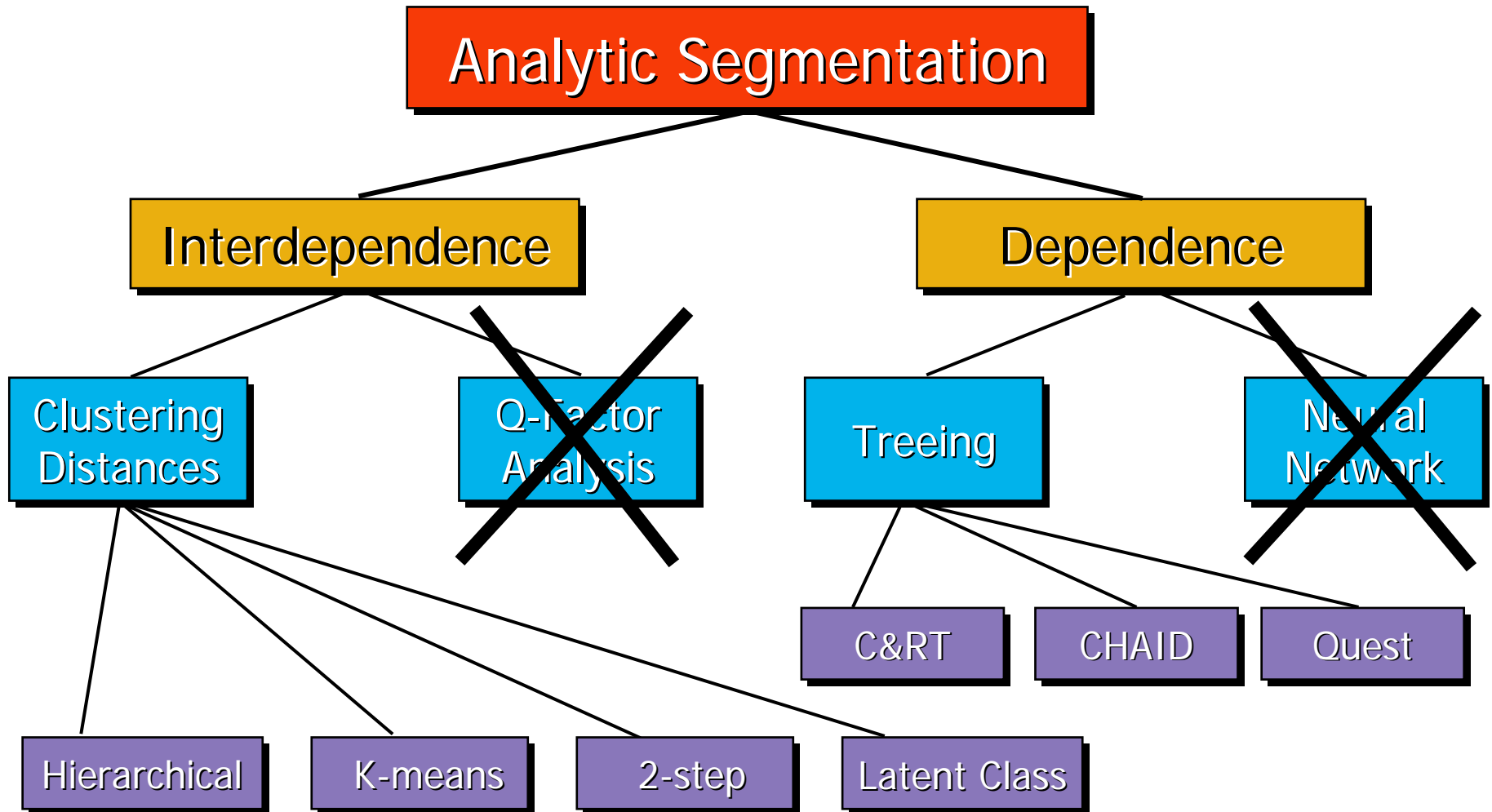


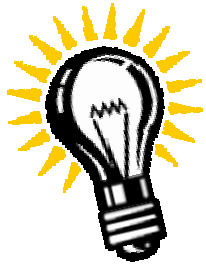
A general taxonomy of analytical segmentation methods





A general taxonomy of analytical segmentation methods



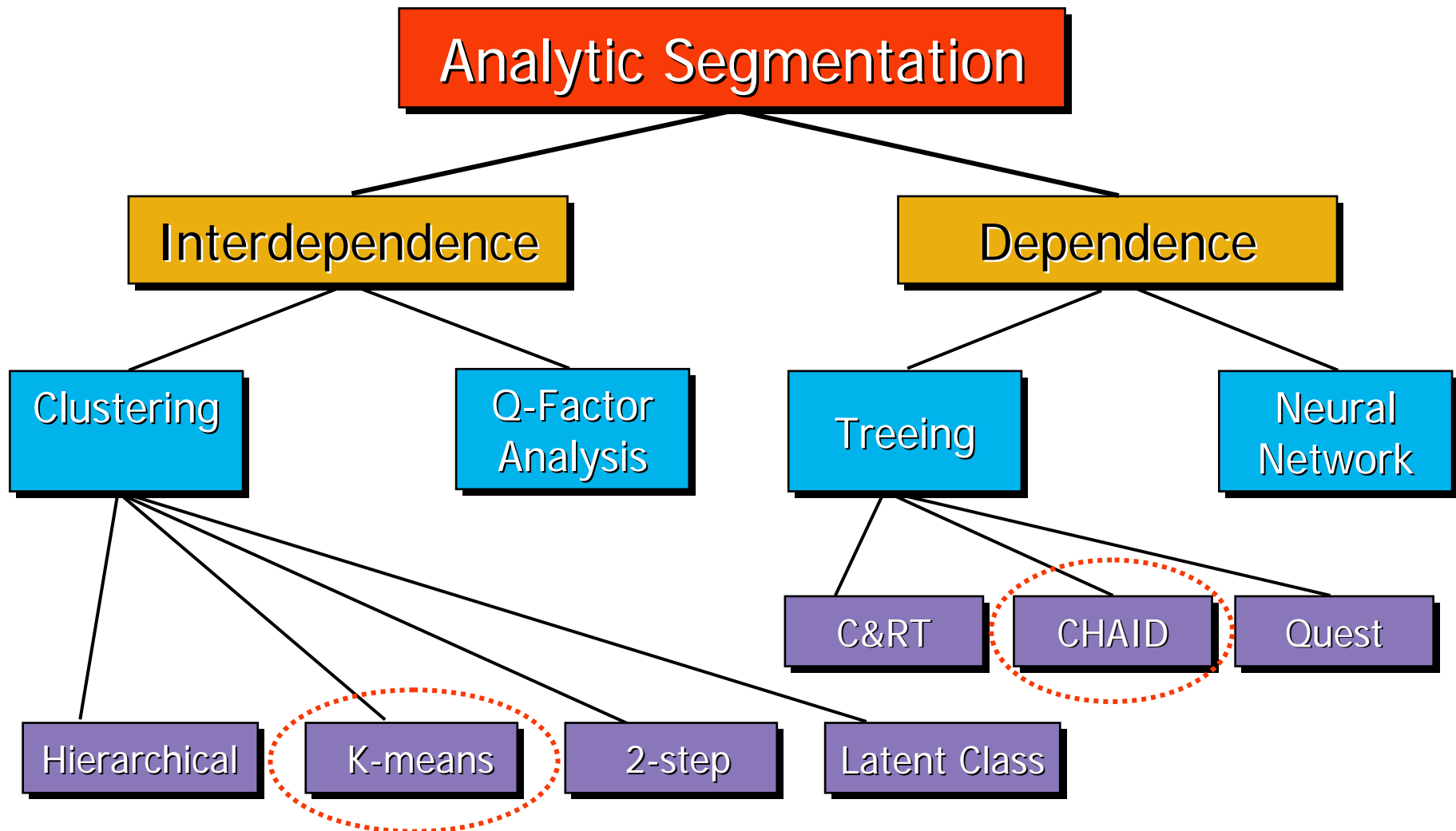


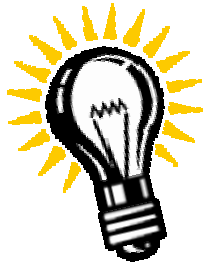
Best Practice #6

Friends don't let friends use Q-factor analysis and Neural net for segmentation!



A general taxonomy of analytical segmentation methods





Best Practice #7

Best all-around approaches are k-means clustering (for interdependence) and CHAID (for dependence).



Objectives of the 2 analytic approaches to segmentation

■ Interdependence (clustering)

- What are the existing “camps” of needs among consumers?
 1. How many segments (“natural camps”) are there?
 2. How is each defined w.r.t needs/beliefs/behaviors?
 3. What is the size of each segment?
 4. What are the “markers” of each segment?
 5. Given their needs and my product’s features, which segments are good targets ?

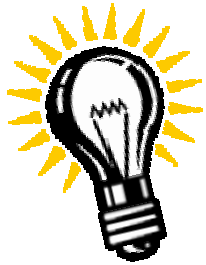
■ Dependence (CHAID)

- Who will buy my product and who will not?
 1. How many segments are there with distinctly different propensities toward buying my product?
 2. What are the key drivers that define the segments who will buy (and not buy) my product?
 3. How do I best reach those most likely to buy my offering?
 4. Given propensities and sizes, which segments are good targets?



Data Considerations for Segmentation in General

- **Sample size: 400-600 minimum; 1000+ typical.** Figure that a 5% segment can hold promise and that you do not want to make inferences from subgroup (segment) sample sizes of less than 30 ... thus, $30/5\% = 600$ minimum.
- **Representative (random) sample; no over-sampling.**



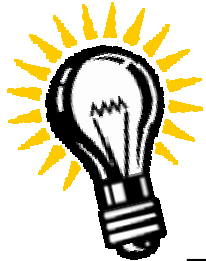
Best Practice #8

For analytic segmentation, gather a large (600+), random sample.



Questionnaire Considerations for Segmentation in General

- Include (towards end) lots of demographics, as well as media and channel use items
- Include at least some competitive product usage and perceived performance items



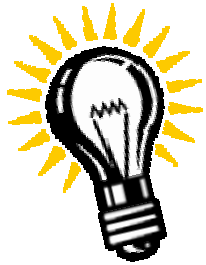
Best Practice #9

Questionnaire should include lots of demographics, media use, and channel use items so that you can discern how to best reach target segments. Also, including at least some items regarding competitive product use and perceived performance is a good idea.



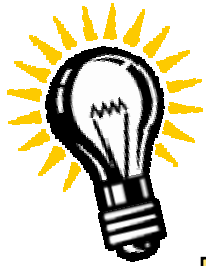
Data Considerations for Clustering

- Preceded by good qualitative work (focus groups) in order to have solid knowledge of all the needs dimensions currently in effect in the market
- Shorter scale (5-point) better than longer (11-point) since,
 - Battery may/should be long
 - Helps avoid **differential scale use**
- Place ***distinct*** anchors (labels) on each point of scale to help avoid **differential scale use**, e.g.,
 - 1 totally agree, 2 somewhat agree, 3 neutral, 4 somewhat disagree, 5 totally disagree
- Needs statements/questions should themselves be “extreme.”
A types are better than B:
 - A. I like my candy to be extremely crunchy. ... I absolutely love chocolate.
 - B. I like my candy to be crunchy. ... I like chocolate.



Best Practice #10

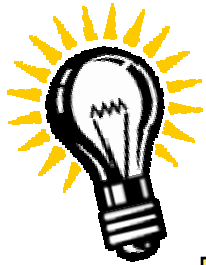
Precede interdependence (clustering) segmentation with fresh qualitative research (focus groups).



Best Practice #11

Slay differential scale use!

Avoid this evil bias by designing a "needs" battery based on a short (1-5 agree-disagree) scale with clearly differentiated anchors on every point and items worded in the "extreme."



Best Practice #12

Consider using conjoint- or discrete-choice-based utilities as the basis of clustering.

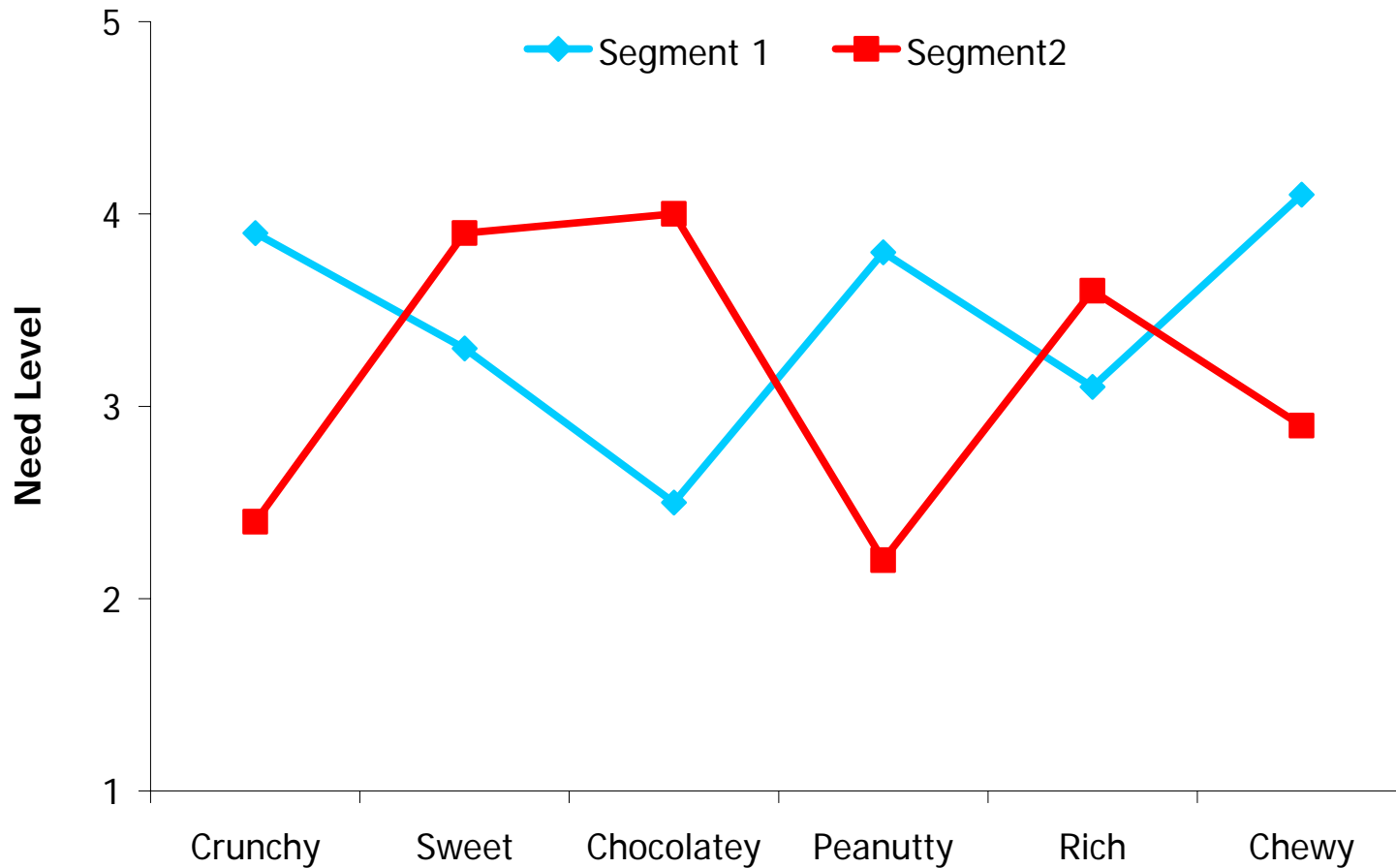


Data Cleansing for Clustering

- Biggest problem in clustering is differential scale use (response style bias).
- Test for differential scale use via clustering 2 segments. If, across attributes, the 2 profiles are strongly correlated and different only in general level of response, then a differential scale use problem exists in the data and needs to be fixed.

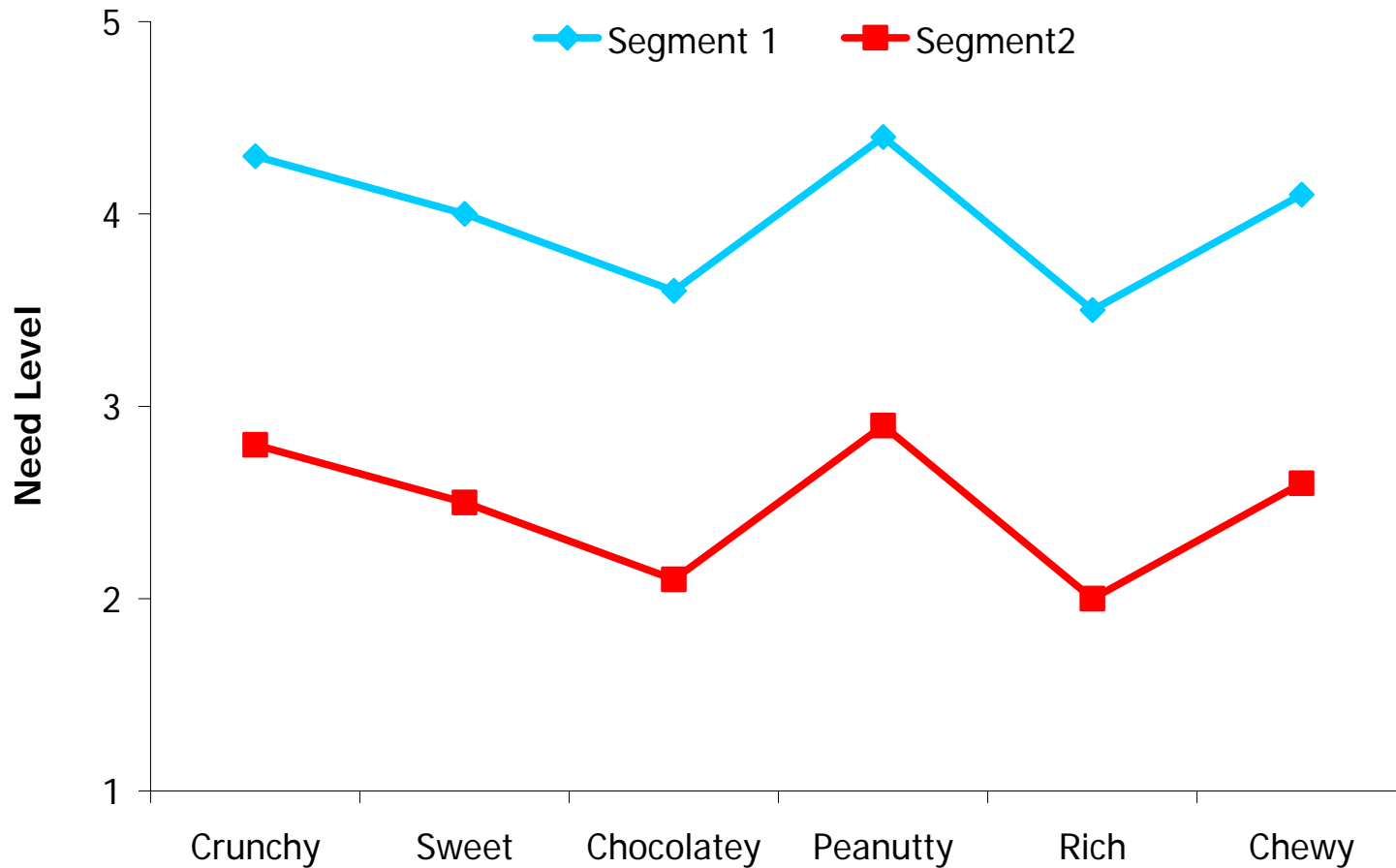


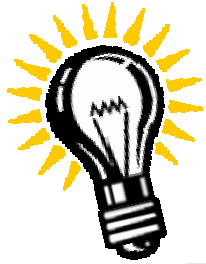
Data Cleansing for Clustering





Data Cleansing for Clustering





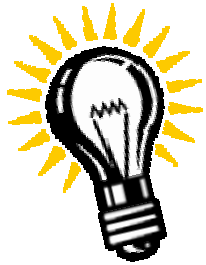
Best Practice #13

Before clustering in earnest, always first test for response style (differential scale use) bias by examining the “parallel-ness” of profiles for 2-3 clusters.



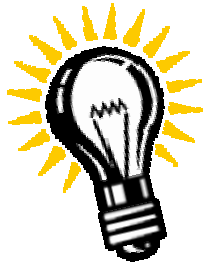
Data Cleansing for Clustering

- Semi and full ipsatization (re-centering and re-dispersing)
- Outliers (ok – leave them in)
- Missing data presents a problem – so fill in. Use mean of other attributes for the respondent adjusted by a sample-wide factor for that attribute. Or if only a few missing data points then just throw out cases via listwise deletion.
- Will need to standardize (e.g. Z scores) if basis variables arise from different scales (why I strongly urge just one single same-scaled “needs” battery)



Best Practice #14

If differential scale use (response style) bias exists, then fix it via ipsatization!



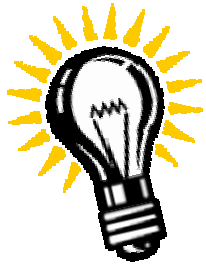
Best Practice #15

When clustering data, outliers are OK (leave 'em alone)
missings are *not* (fill 'em in).



Precede Clustering Analysis with Factor Analysis

<u>Original item</u>	<u>Factor 1</u>	<u>Factor 2</u>	<u>Factor 3</u>	<u>Factor 4</u>
Crunchy		+		
Sweet			+	
Chocolaty	+			
Peanutty		+		
Rich			+	
Chewy				+



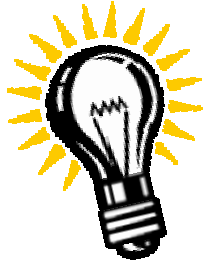
Best Practice #16

Cluster only unique dimensions;
factor analyze original items to
get the unique dimensions.



Data Analysis in Clustering

1. Determine the best number of cluster segments:
 - Via k-means, examine the decomposition of segment sizes in higher and higher order solutions (2-12); crosstab prior-run results with next higher-order solution. Stop once the larger segments stabilize and/or new segments pull from too *many* prior segments and/or some segments begin to substantially *grow* in size, then a “too many segments” point has been reached.
 - Use hierarchical (dendrograms), latent class, and two-step cluster analysis to help determine final answer for # of segments.
 - Client input.
2. Obtain final cluster solution via two runs of k-means:
 - Use first pass results (centroids) as starting points for 2nd, final k-means run.
3. ANOVA of basis variable means across segments, while opportunistic, gives a basic sense of how much differentiation exists between final segments.



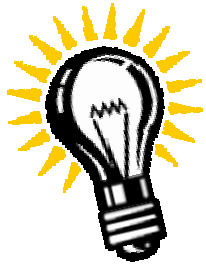
Best Practice #17

In clustering, use numerous algorithms to solidify answer to “How many segments?” Once the number is decided, obtain final cluster segments using a two-step k-means process.



Data Considerations for CHAID

- Stated purchase intent
- Exaggeration-corrected, derived purchase intent (Assessor[®])
- Conjoint/choice derived
- May be something different than purchase intent:
 - Retention
 - Advertising response
 - Targetability = difference in predicted usage and actual usage
- The long list of demographics, media use, and channel use become potential drivers.



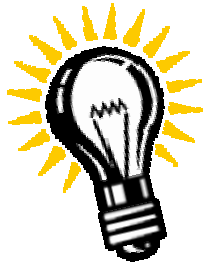
Best Practice #18

In CHAID, the dependent variable should be as reliable and accurate as possible, usually better than a simple stated intention item.



No special data cleansing needed for CHAID...

- No need for factor analysis
- No need for ipsatizing or standardizing
- Outliers usually ok
- Missing data ok (as it is treated as diff value)



Best Practice #19

In CHAID, there is no real need for heavy data cleansing.



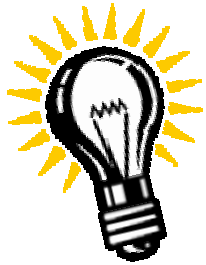
Data Analysis in CHAID

- Usually want to look at perhaps 3 or 4 different levels of solutions
 - Just the demographics
 - Just the media and channel use items
 - Some “mid” level (e.g., demographics and media/channel use)
 - Kitchen sink = everything in the questionnaire!



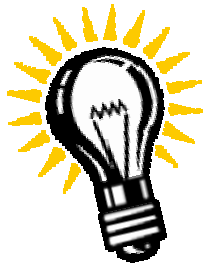
Typical CHAID Settings

- Drivers must be defined correctly as to level of measurement since will be treated differently
- Smallest child node at 5% of N, parent at twice child
- Grow trees usually to around 4-5 branch levels
- Alpha=.05
- Turn Bonferroni off



Best Practice #20

In CHAID, set the minimum size of child nodes at 5% of total N , and parent nodes at twice child node size.



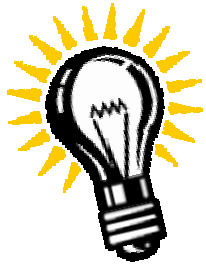
Best Practice #21

In CHAID, grow trees to a depth of about 4-5 branches with alpha set at .05 with no Bonferroni adjustment.



CHAID Requires Human Judgment Overlay

- Merging
- Redefining splits
- Cutting whole branches
- Eliminating some drivers

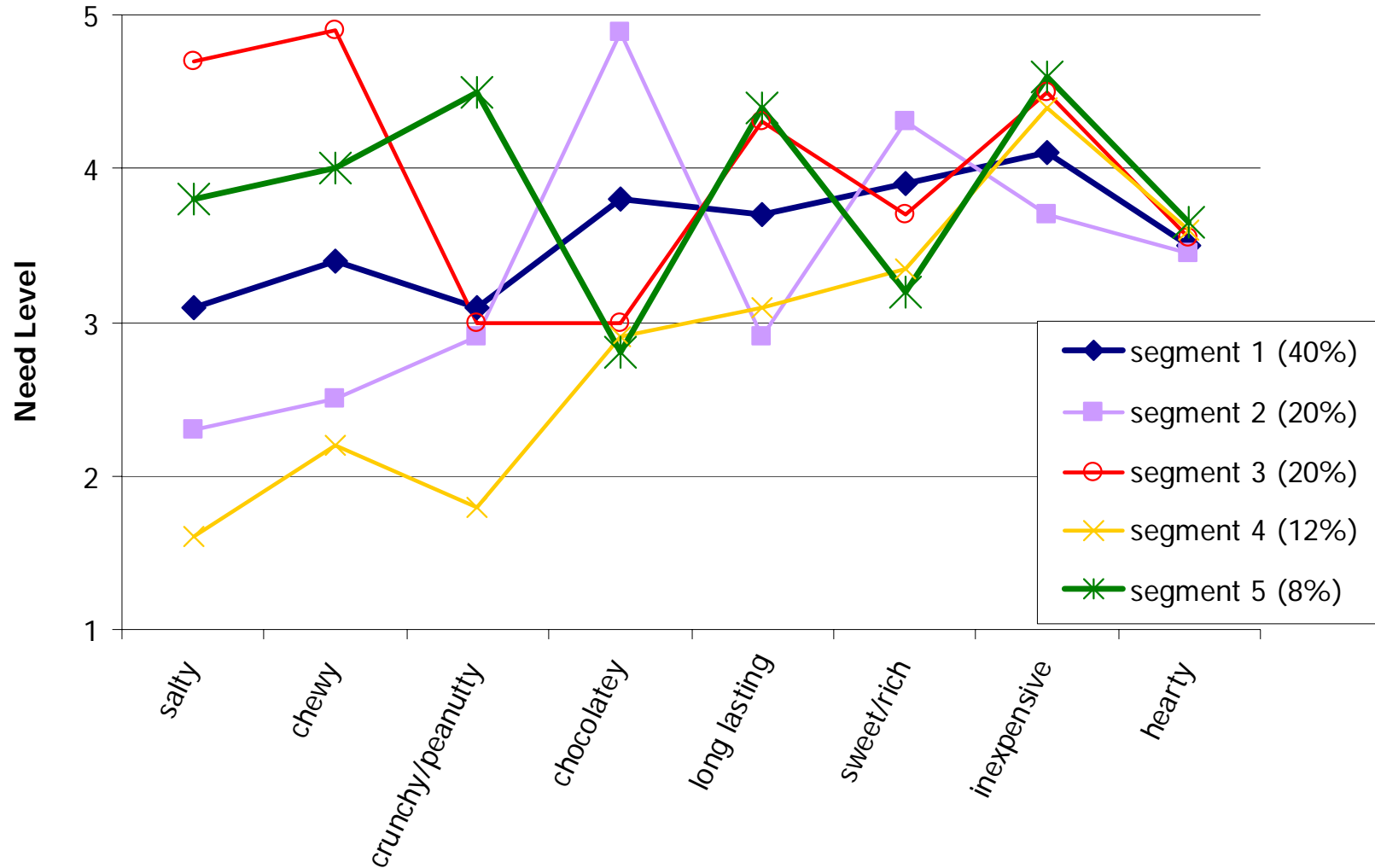


Best Practice #22

In CHAID, expect a fair amount of required human judgment overlay.



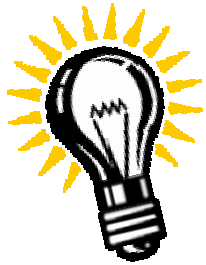
Presentation of Interdependence (Cluster) Segments





Presentation of Dependence (CHAID) Segments

Segment	Industry	Company Size	Position	Segment Size	Average Propensity
★ 1	Real Estate	---N/A---	---N/A---	4%	96%
★ 2	Services other than Real Estate	0-50 employees	---N/A---	22%	90%
★ 3	Services other than Real Estate	50+ employees	Marketing Executive	8%	82%
★ 4	Services other than Real Estate	50+ employees	Not a Marketing Executive	18%	70%
★ 5	Non-Service Industries	---N/A---	Director	18%	60%
6	Non-Service Industries	---N/A---	Higher than director	4%	38%
7	Non-Service Industries	---N/A---	Lower than Director	26%	14%
				Sum =	100%



Best Practice #23

The best way to present cluster segments is with profile line charts; the best way to present CHAID segments is with tables.



For more information:

Frank Wyman

Director of Advanced Analytics

M/A/R/C[®] Research

(864) 938-0282

frank.wyman@marcresearch.com

Copyright © 2005 M/A/R/C Research