Data Driven Customer Segmentation

A whirl wind tour.

Jim Porzak
Data Science for Customer Intelligence (DS4CI.org)
Global Predictive Analytics Conference
Santa Clara, CA
May 5, 2015

Why Segment Customers?

- For strategic and tactical business decisions
- Strategic
  - Segments fairly stable over time
  - Conceptually high level
- Tactical
  - Reflect customers current state
  - Actionable today – in the weeds

How to Segment?

"We've broken your list into eighty-four subgroups. Our work here is done."

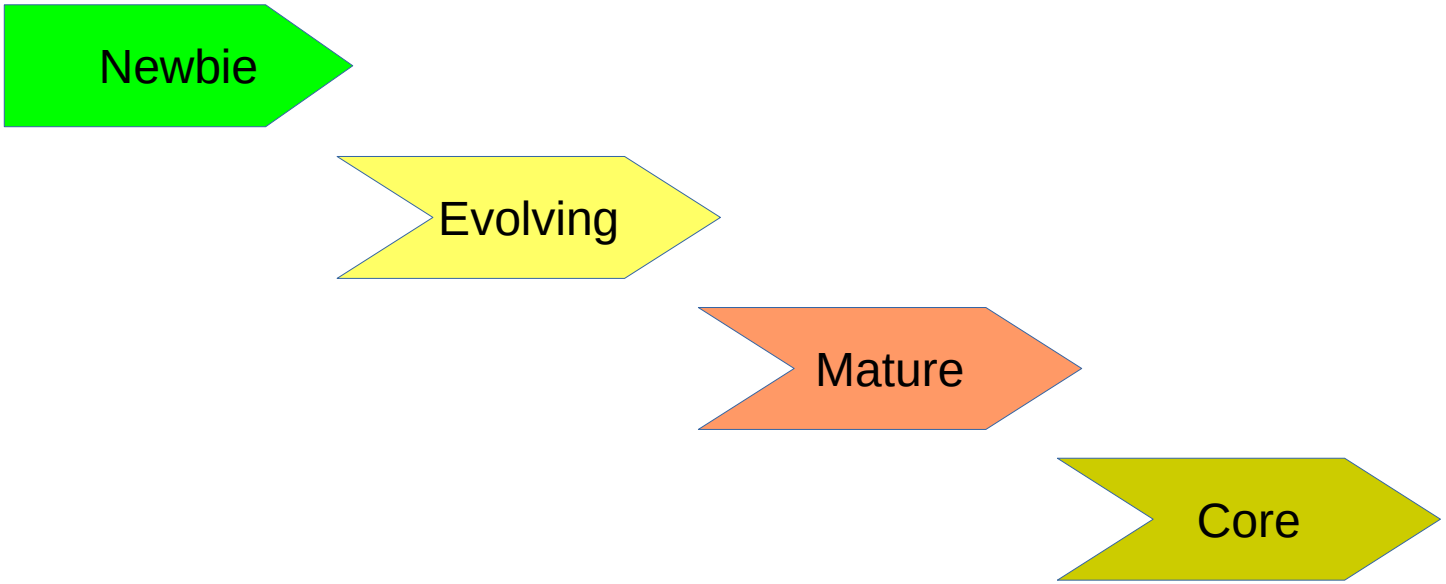
Segmentation Examples

From simple to more complex:

- Customer Tenure
  - A really simple concept – everyone “gets it”
  - Can have huge impact on an organization
- RFM Based Segmentation
  - Not your grandfather’s RFM!
  - Intuitive & actionable segments
- Customer Classification
  - Based on ~ static properties (preference or properties)
  - Use unsupervised clustering
- Customer Life-stage Segments (a preview)

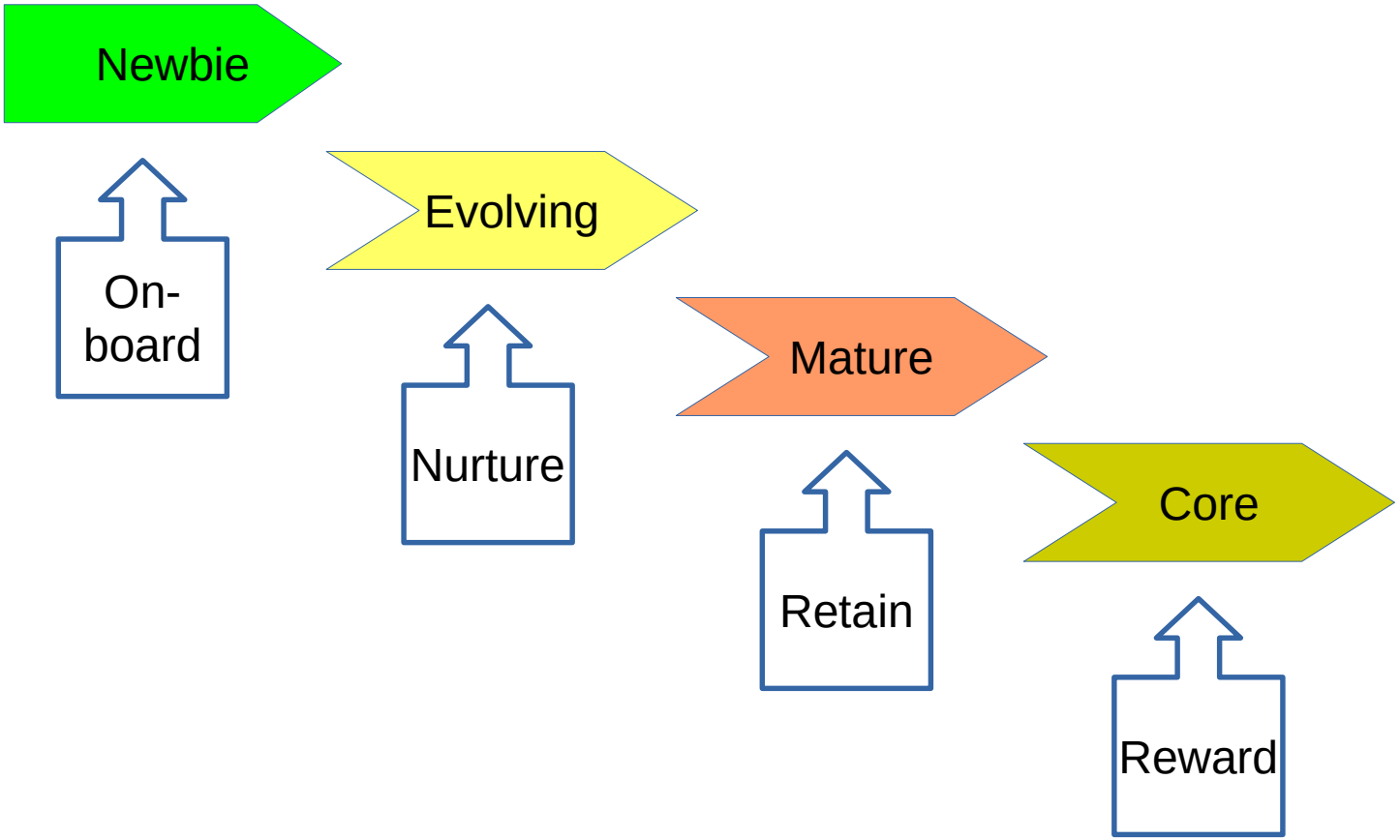
Customer Tenure Based Segments

Customer Tenure Segments



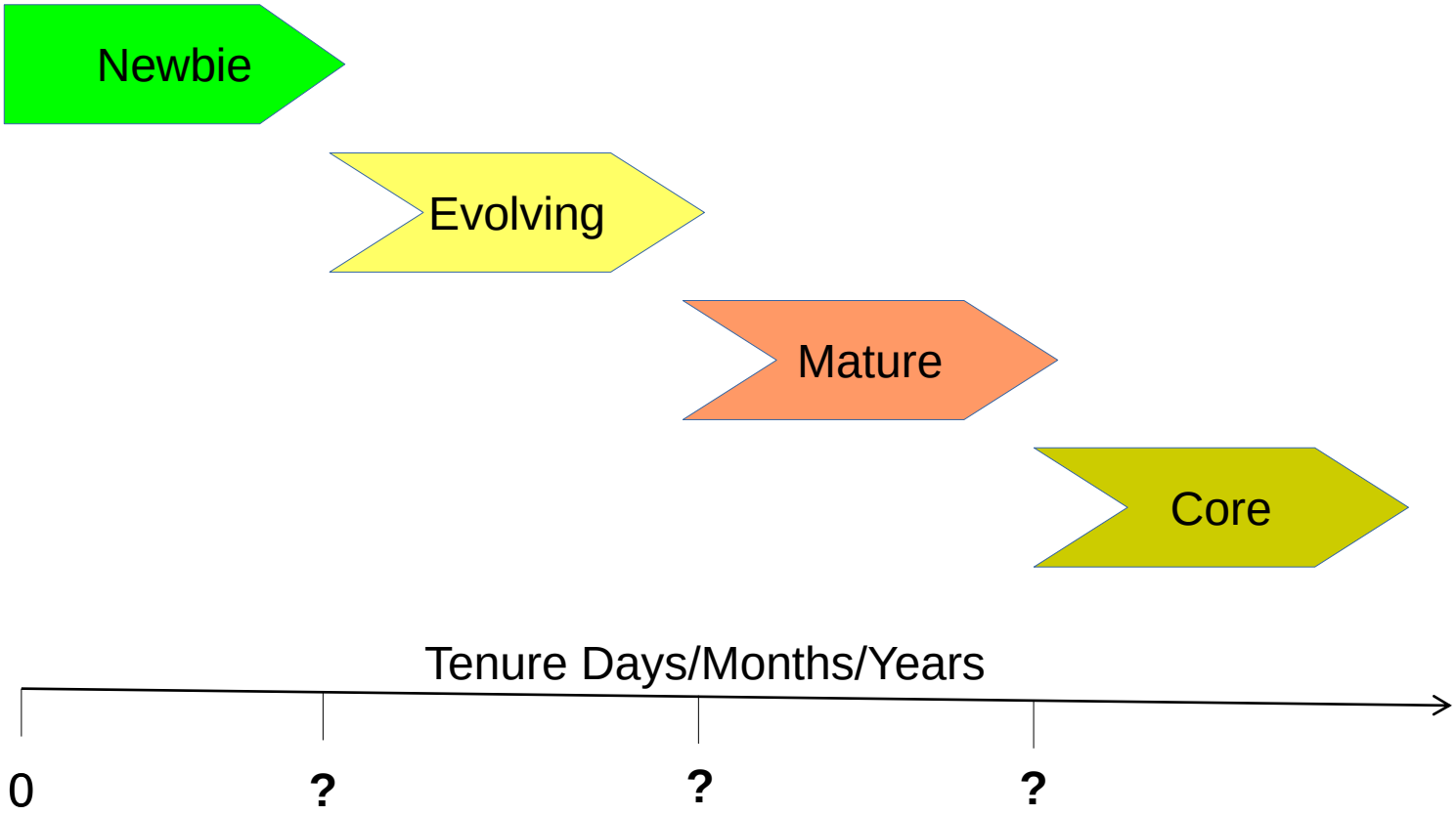
Customer Tenure Segments

With Marketing Actions



Customer Tenure Segments

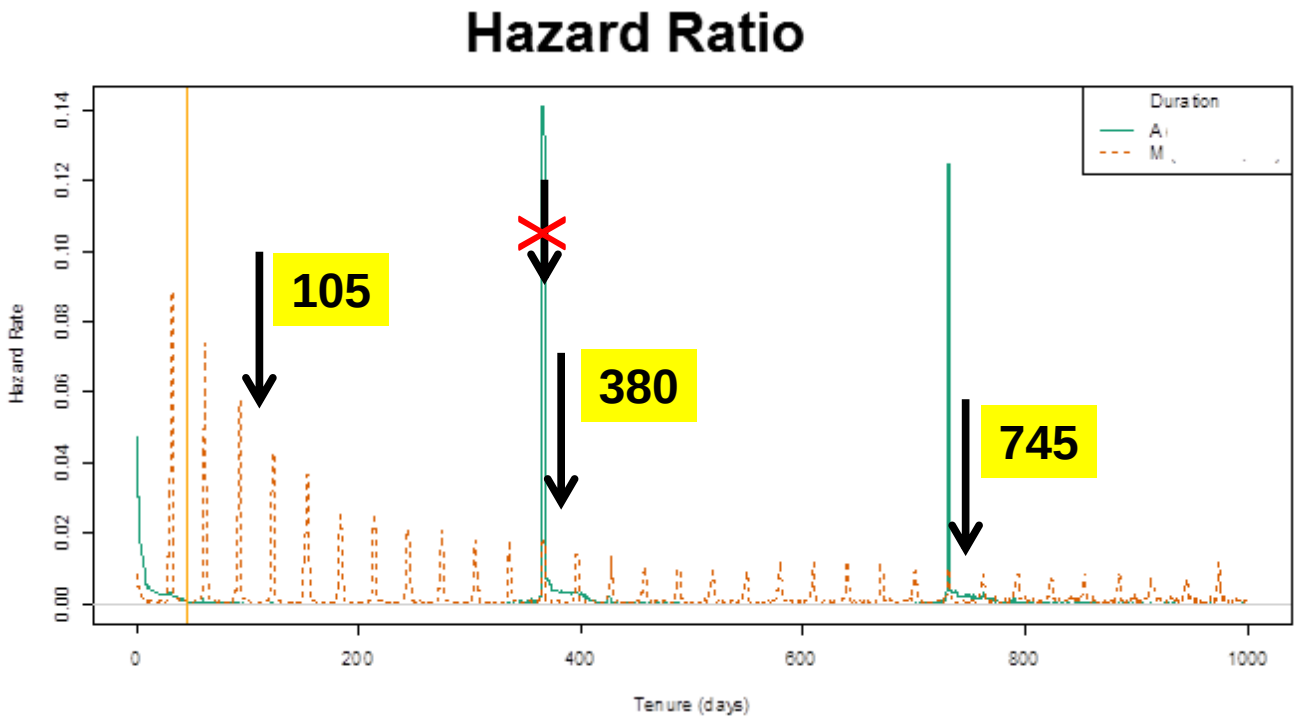
How to Determine Break Points?



Customer Tenure Segments

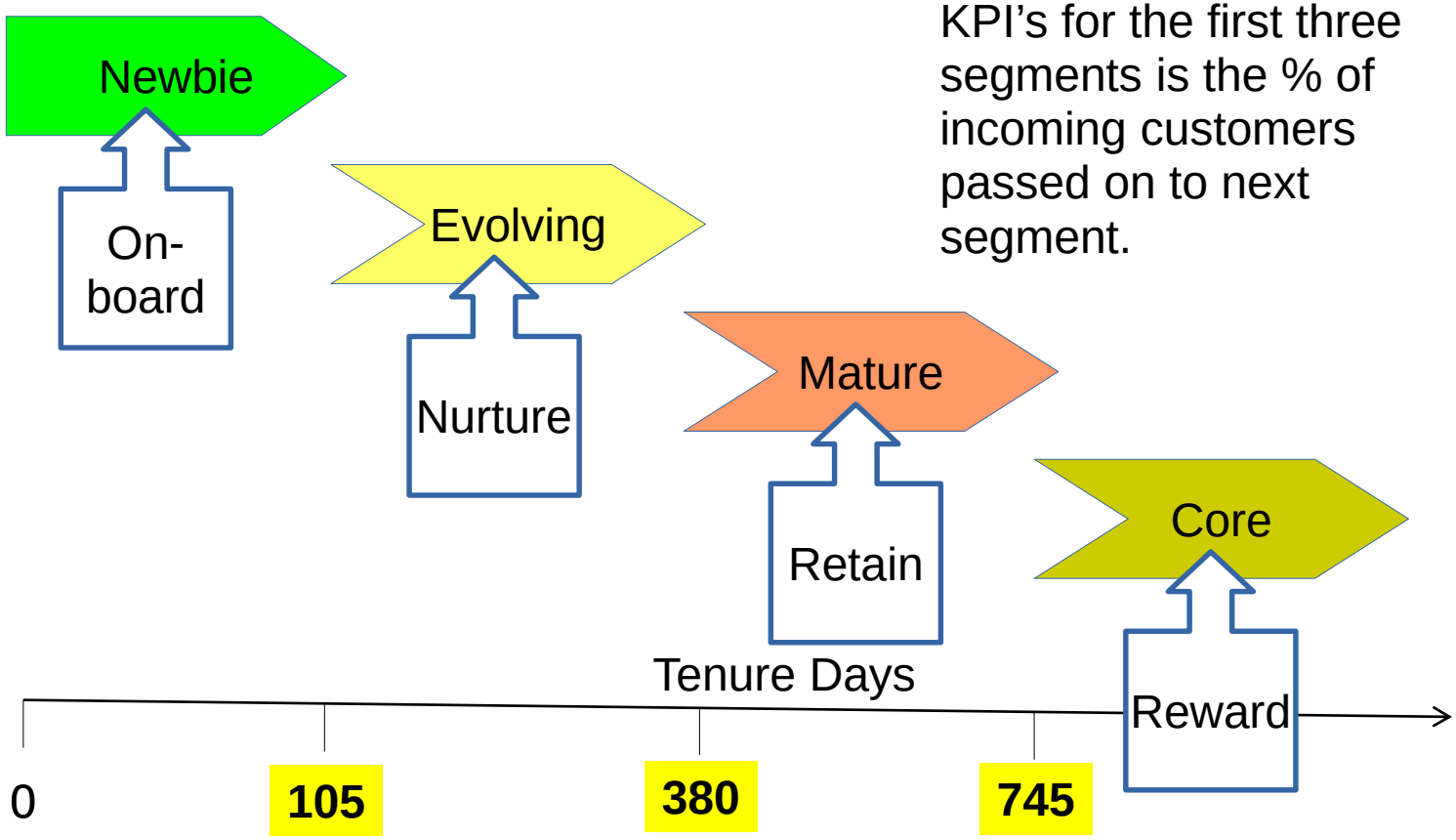
If subscription, look at hazard ratio curve.

$$HR_i = \frac{\text{(\# terminated during day i)}}{\text{(\# active at beginning of day i)}}$$



Customer Tenure Segments

How to Determine Break Points?



KPI's for the first three segments is the % of incoming customers passed on to next segment.

RFM Based Segments

What's RFM?

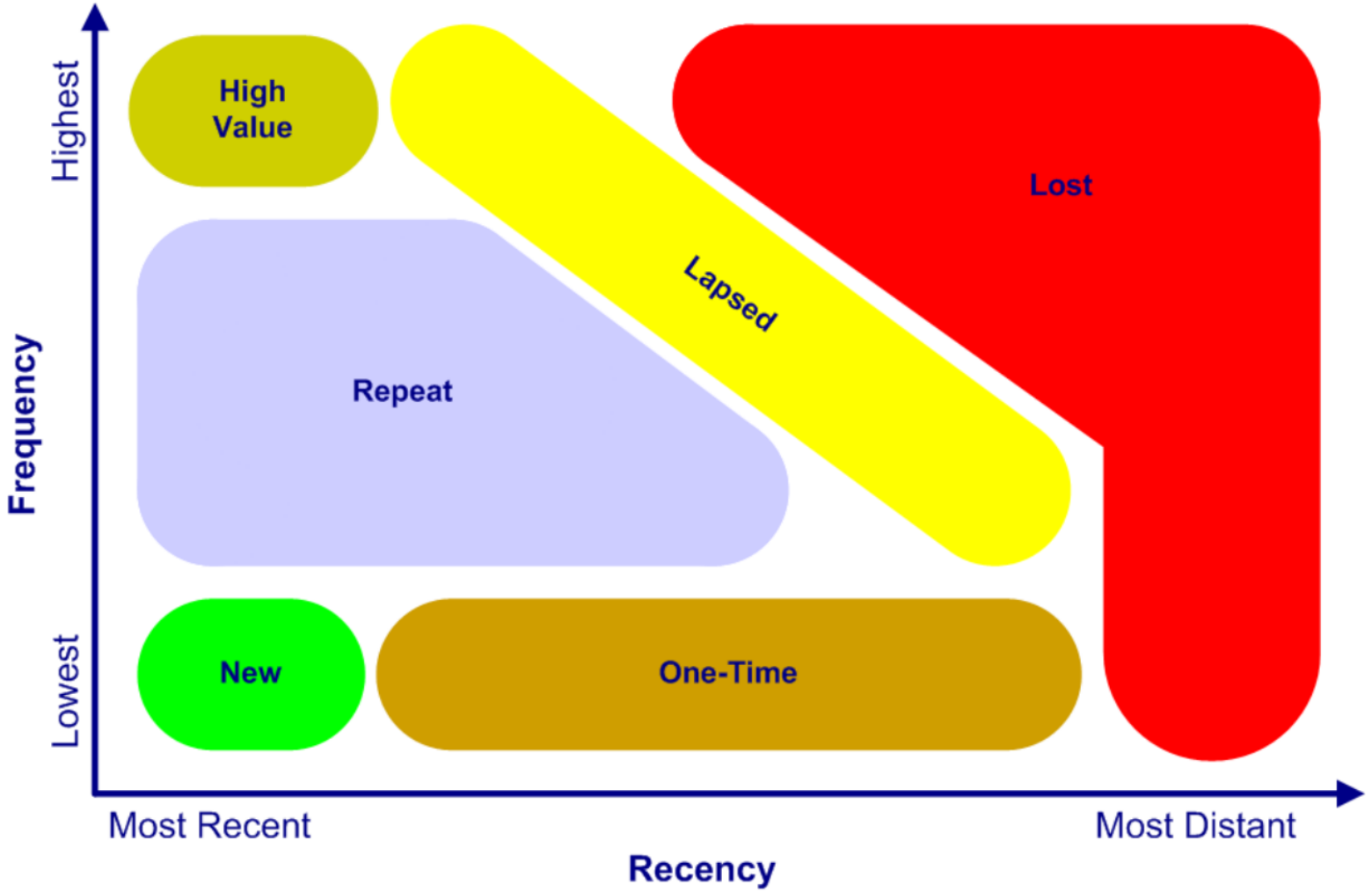
It's all about your customer!

- **R**ecency – How long since they did it?
- **F**requency – How often have they done it?
- **M**onetary – What have they paid you to do it?

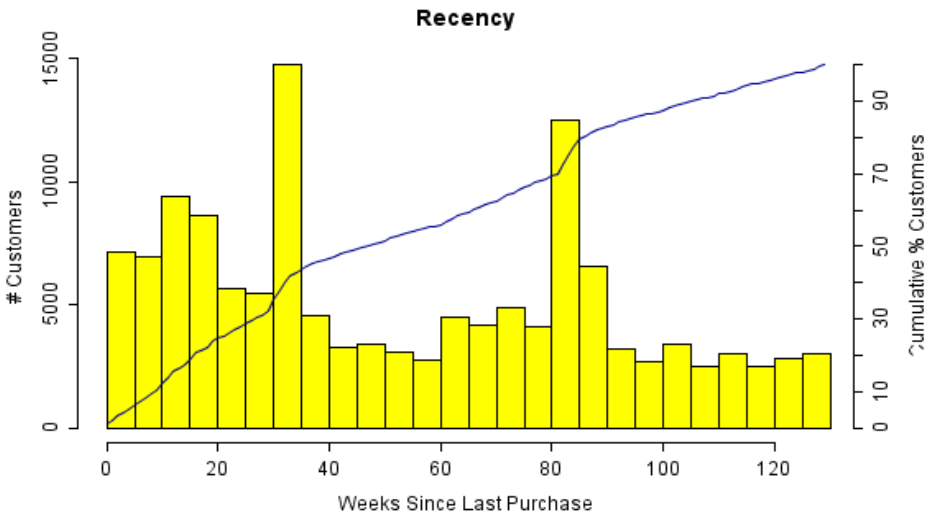
"it" depends on what you do & are measuring…

  - Selling things => event is purchasing (classical RFM)
  - Communicating => event is reading, clicking
  - Delivering => event is using

Spoiler Alert: Final Segments



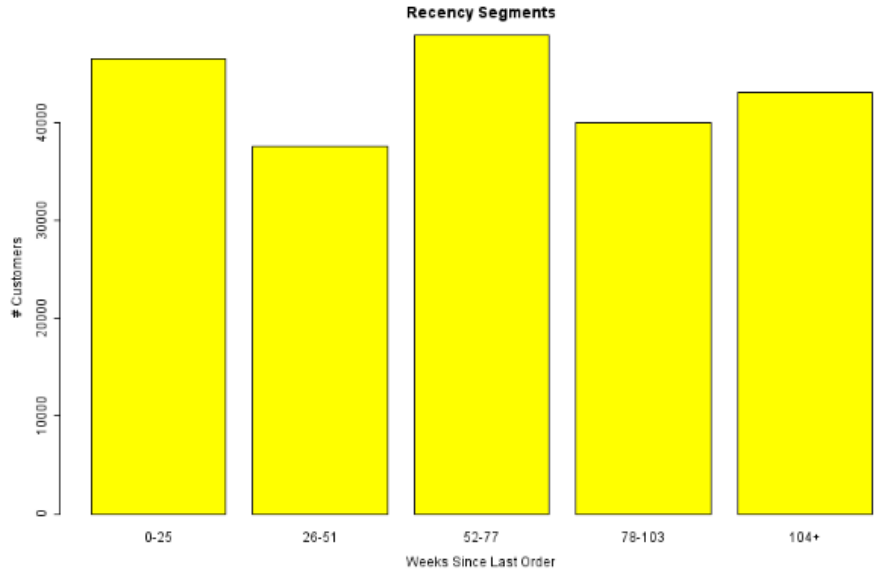
Recency



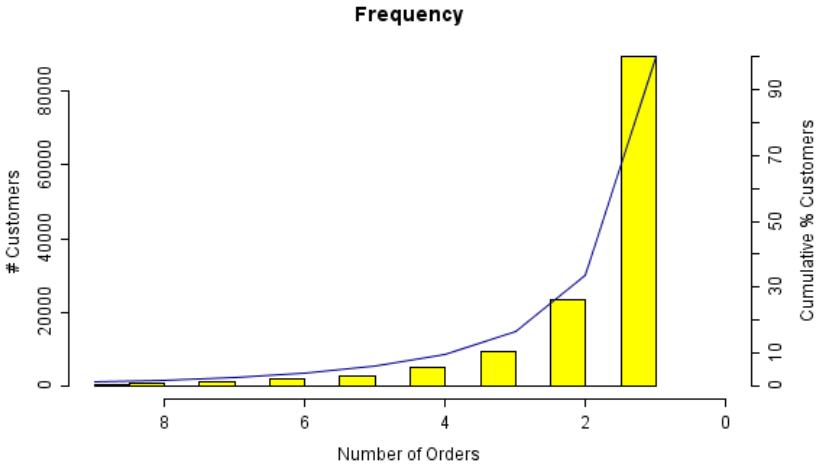
Breaks (weeks <=): 25, 51, 77, 103, <else>

levels = c("0-5", "6-11", "12-17", "18-23", "24-29"))

Note levels labeled in months, not weeks



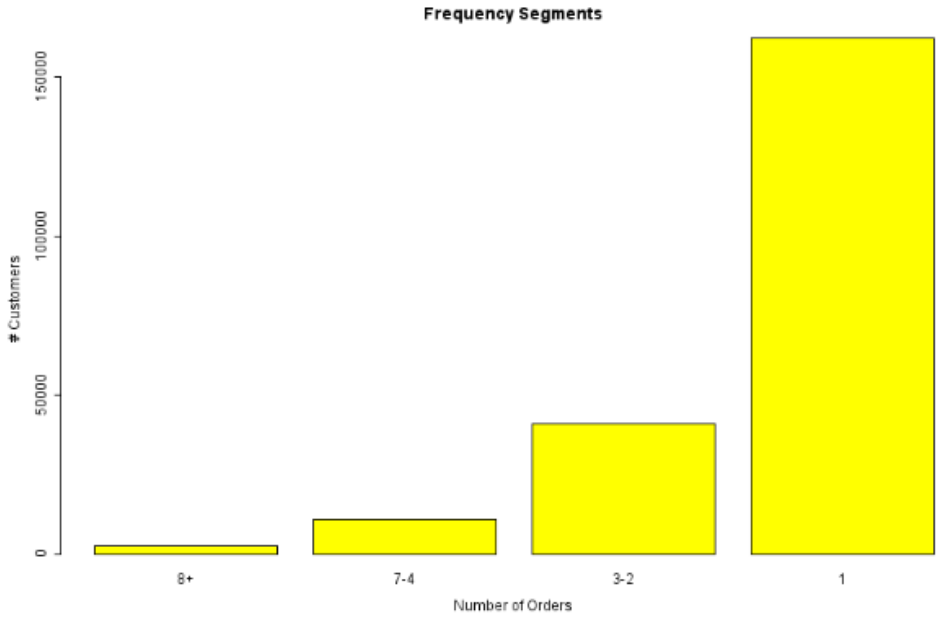
Frequency



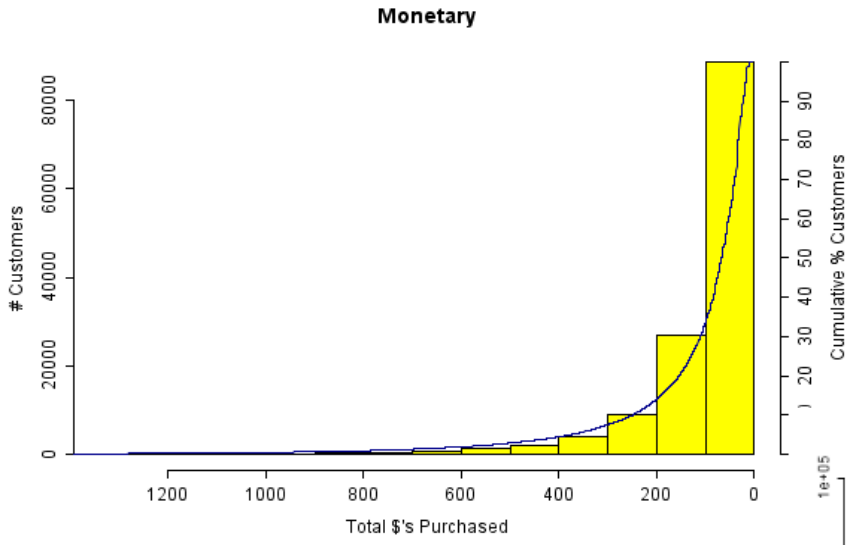
Breaks (count <=): 1, 3, 7, <else>

levels = c("8+", "7-4", "3-2", "1"))

Note ordering for best is left.



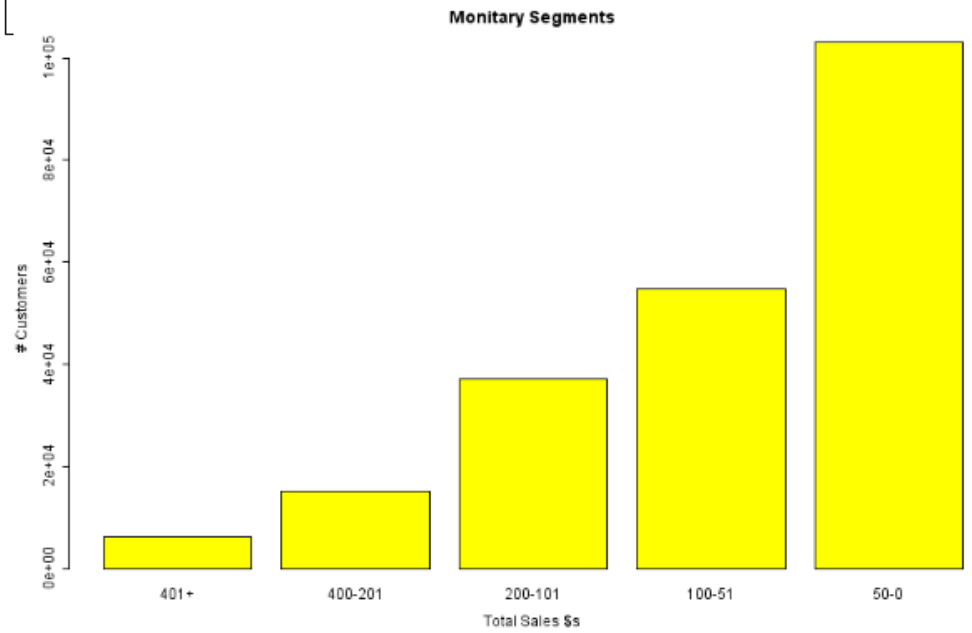
Monetary



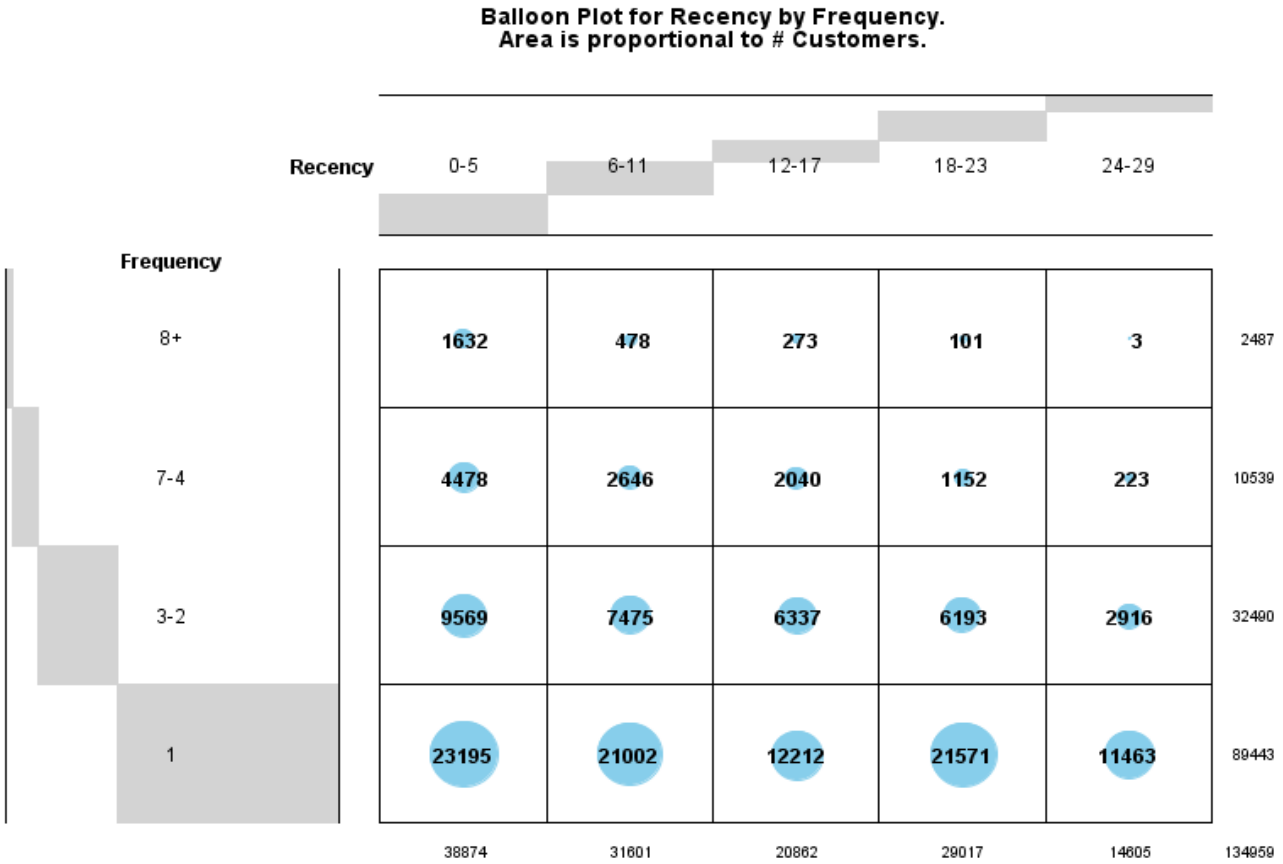
Breaks (value <=): 50, 100, 200, 400, <else>

levels = c("401+", "400-201", "200-101", "100-51", "50-0"))

Again ordering is best is left.



Plot # Customers by R x F

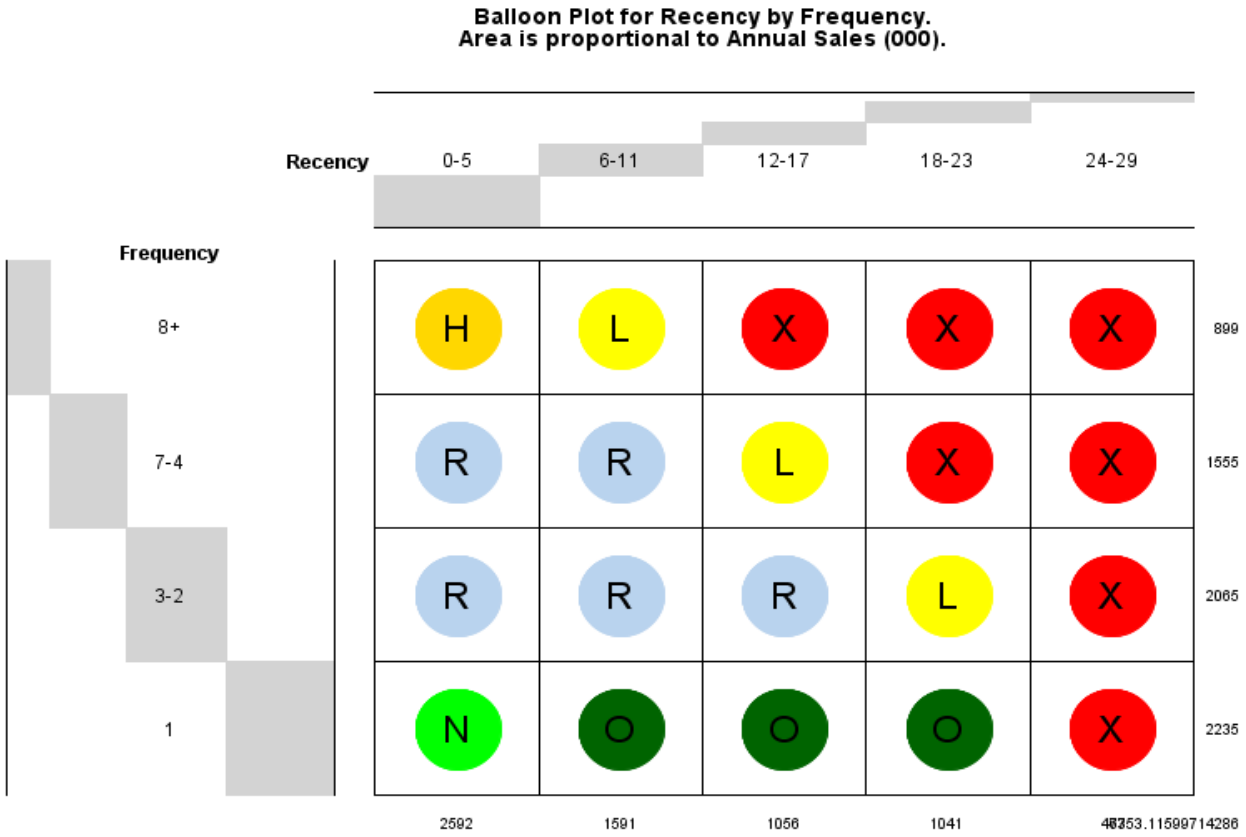


Now Bake in Monetary

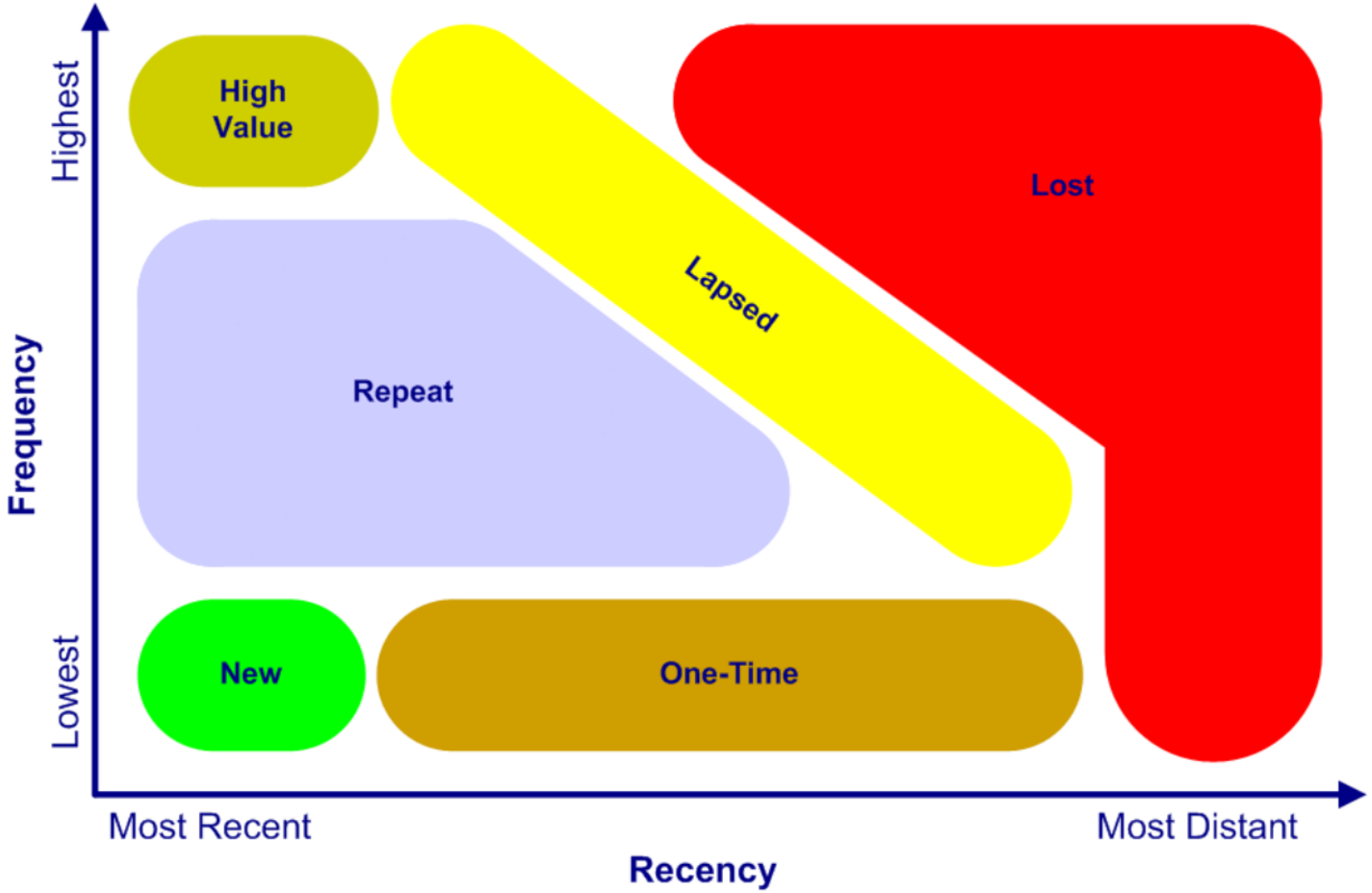
Too Complicated. Simplify!

- What cells can be combined?
  - Must make sense for marketing actions.
  - Way to many cells!
  - What is the story?
- How about two basic concepts?
  - Lifestage: New, Active, Lapsed, Lost
  - Value: Gold, Silver, Bronze
- Combine as:
  - New, Lapsed, & Lost
  - High Value, Repeat, & One-time

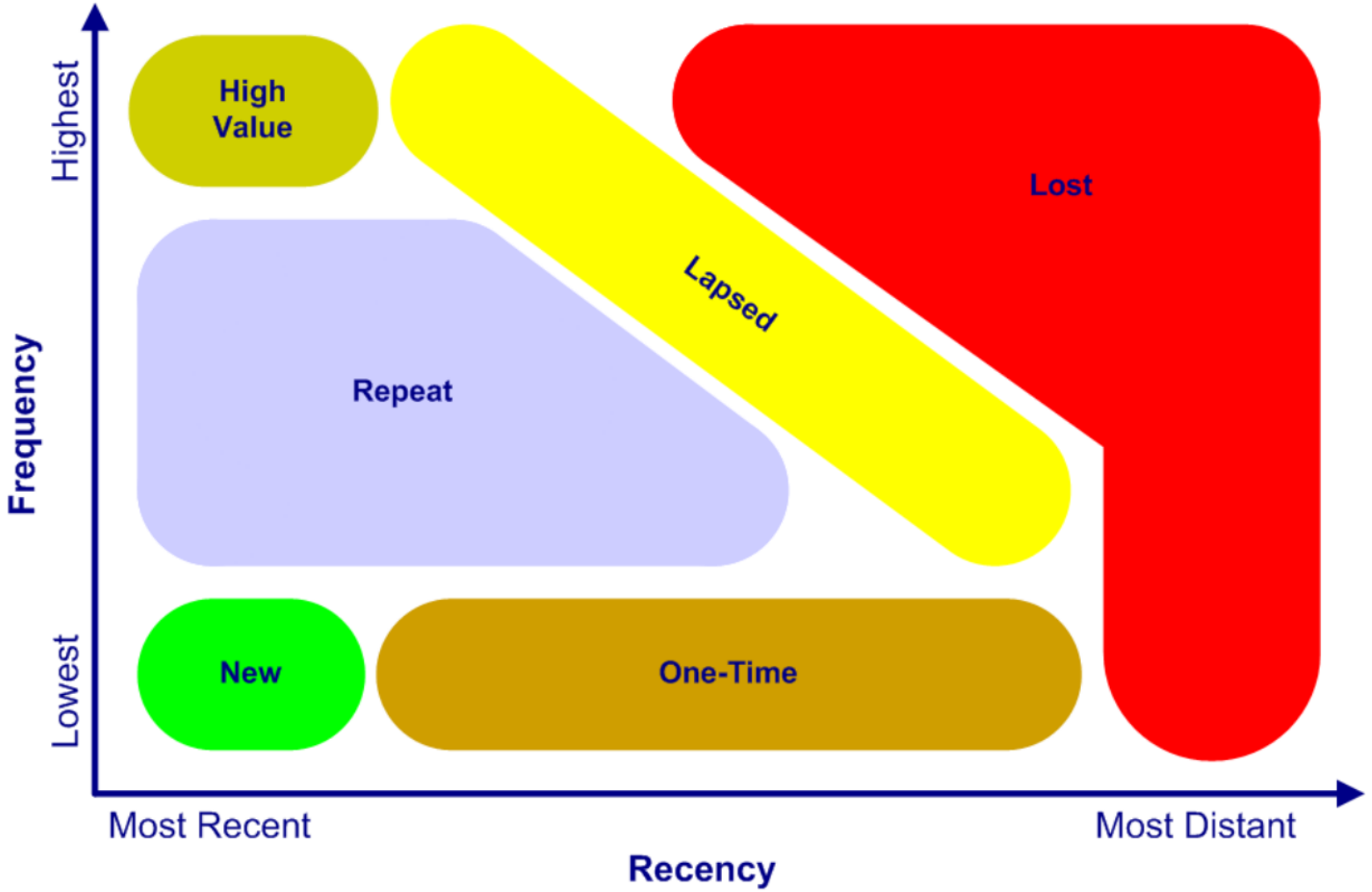
These Segments on Balloon Plot



Smoothing Out Adjacent Cells



Clear Marketing Actions per Segment



Cluster Based Segments

Cluster Based Segments

- Given a data set of non-required preferences and/or properties individual entities, find clusters to segment the entities.
- Totally unsupervised. We only specify the number of clusters we want.
- Two criteria for a “good” solution:
  - The cluster solution is stable
    - Repeatable with different random starts
  - The segments make sense
    - Our business partners can tell a story about each segments

Tool: flexclust by Fritz Leish

- Allows different distance measures
  - In particular, the Jaccard distance which is suited for optional response type survey or optional properties lists
  - 1 is a “yes” to the question - it is significant.
  - 0 is a didn't answer or does not apply. 0 is not the opposite of “yes.”
- Additionally flexclust had very good diagnostic and visualization tools. As an R package, it leverages the rest of the R ecosystem.

Using flexclust

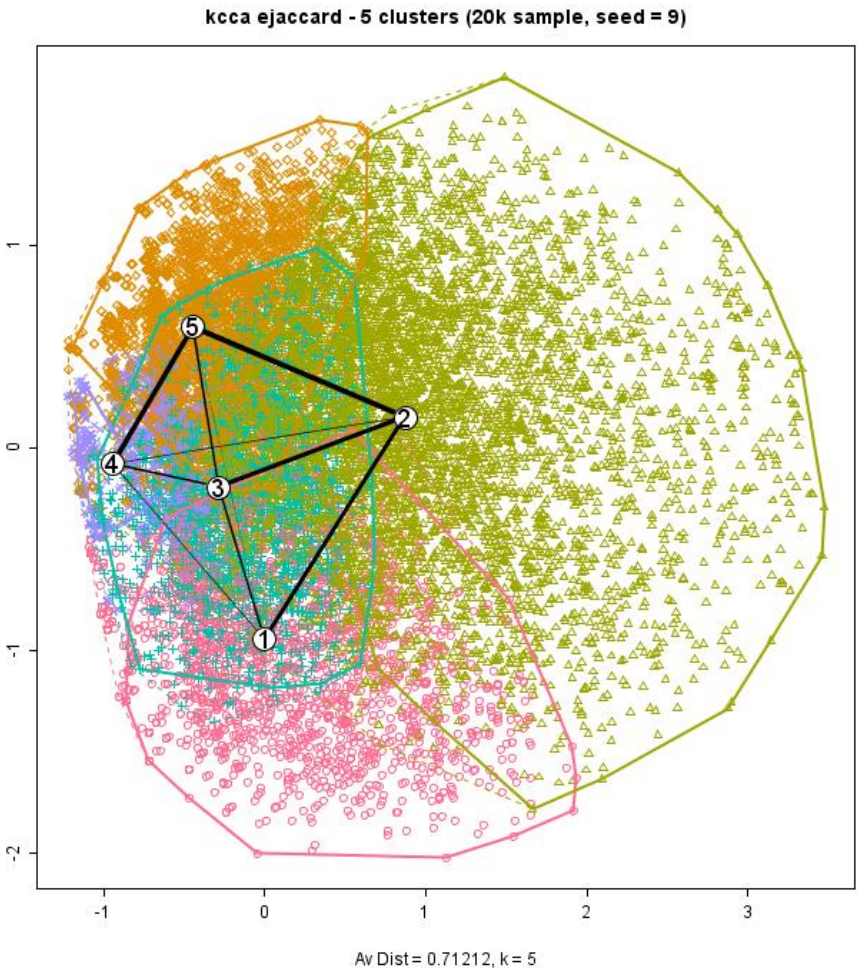
- Generate cluster solutions for
  - A range of desired # of clusters (k), say 3 to 8
    - Over a set of random seeds (~ 6-10)
- Discard cluster solutions for k's which do not have stable solutions
- Examine variable plots by segment for remaining k's
  - Pick the one which makes most sense to your business partners.

Example 1 – Survey Responses

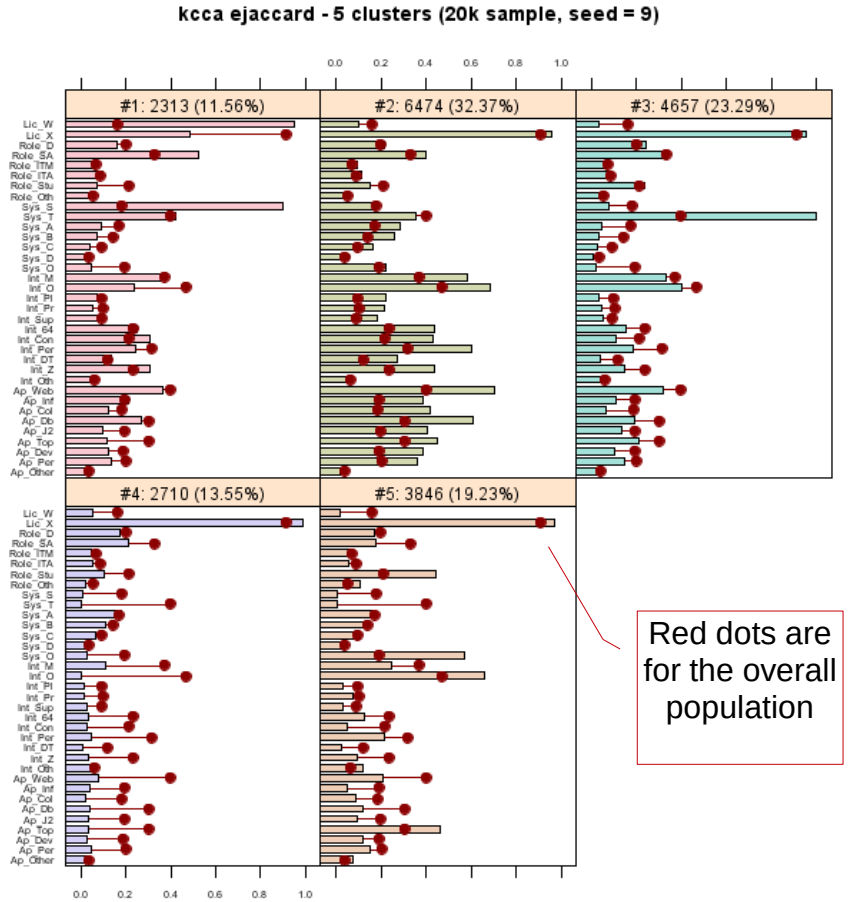
- 20k respondents to technical product use survey
- 35 check boxes or radio buttons
  - None are required
  - Coded as binary responses
- Goal: come up with “a few” segments which can be used to segment new respondents for follow up sales actions.

Example 1 - The 5-cluster solution

The 20k subjects plotted over the first two principal components:



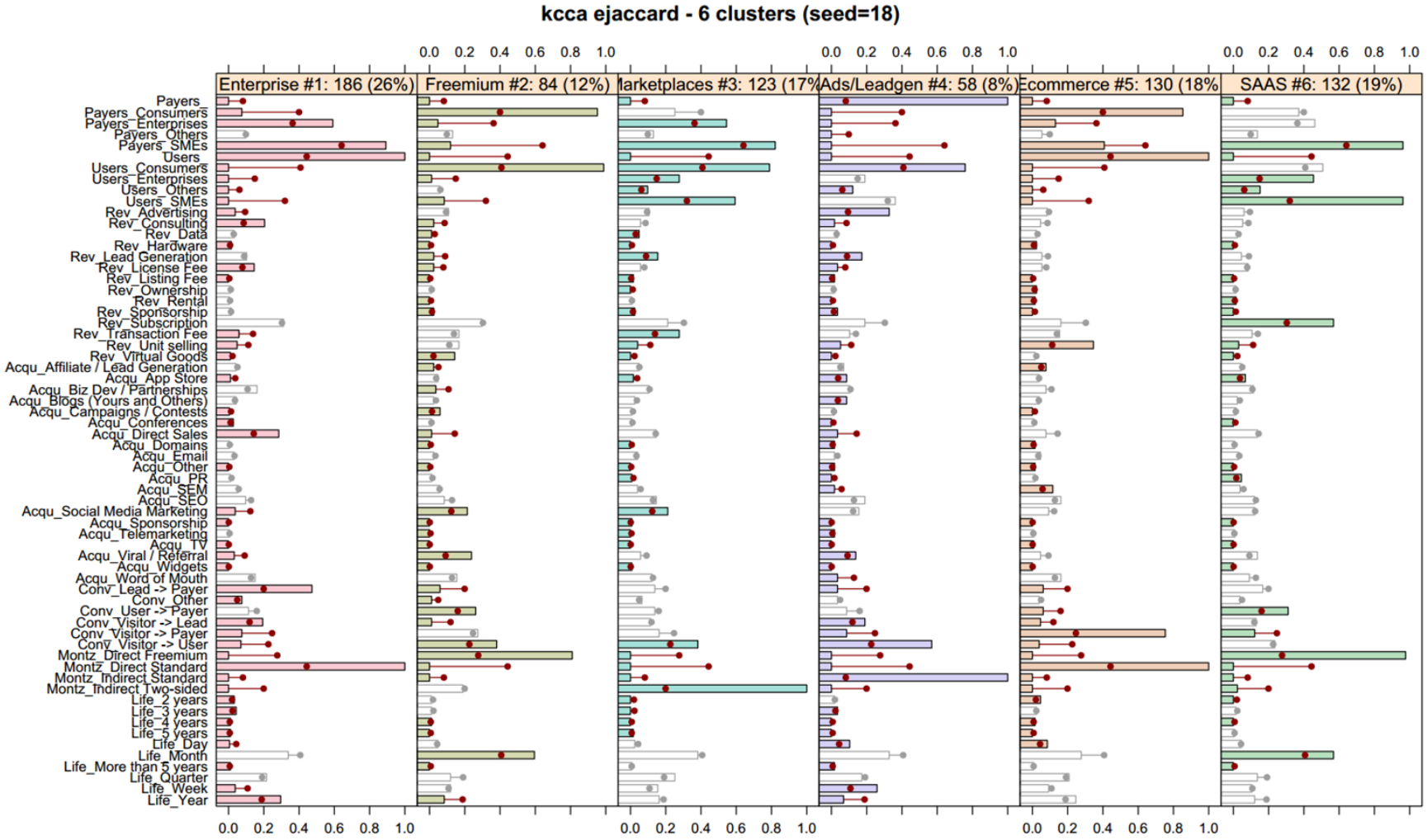
The 5 clusters showing distribution of responses to each question:



Example 2 – Business Attributes

- ~1k respondents to “nature of your business” survey
- 62 check boxes or radio buttons
  - In six topics
  - Some are required
  - Coded as binary responses
- Goal: come up with “a few” segments to characterize the fundamental natrue of the on-line business

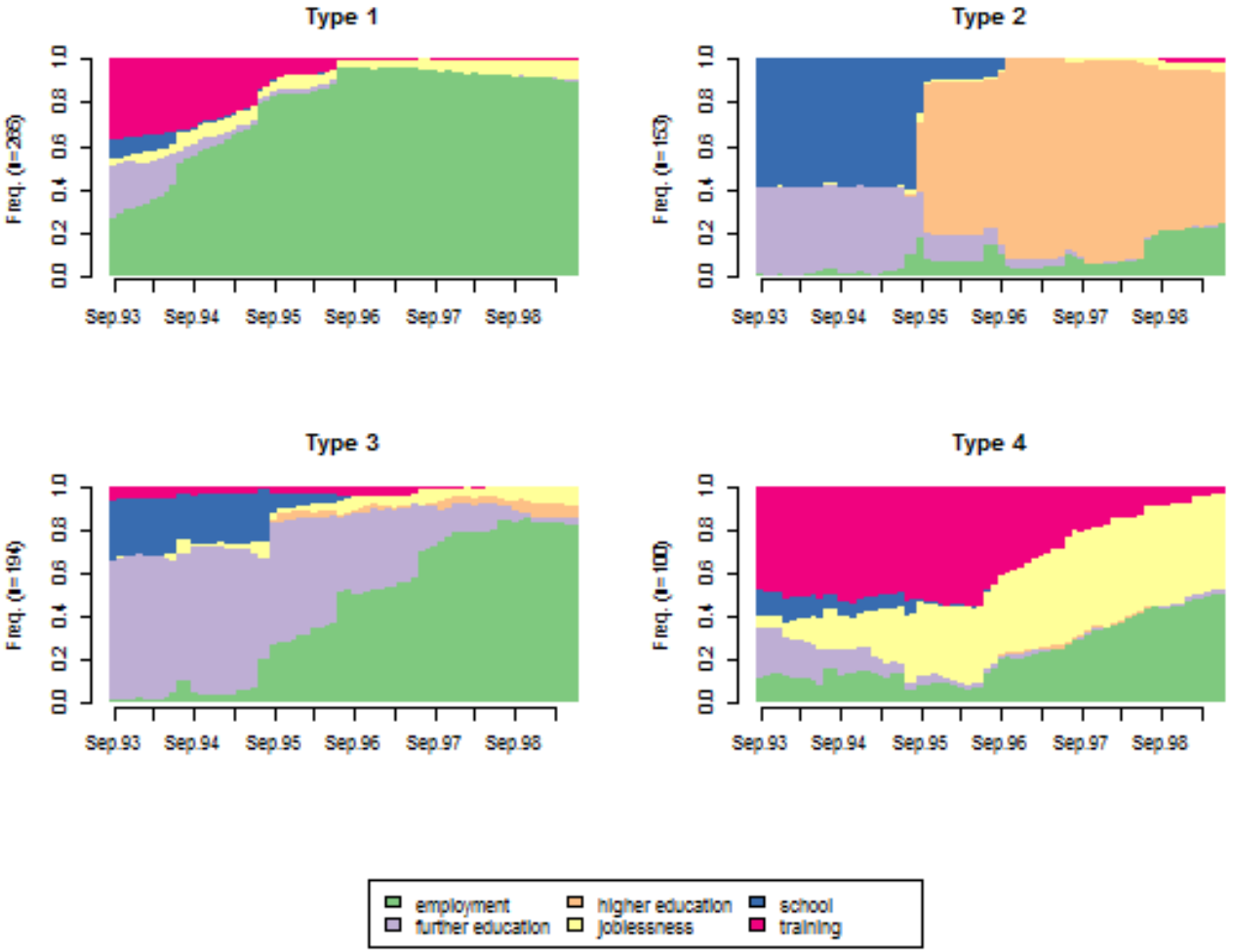
Example 2 – the 6-cluster solution



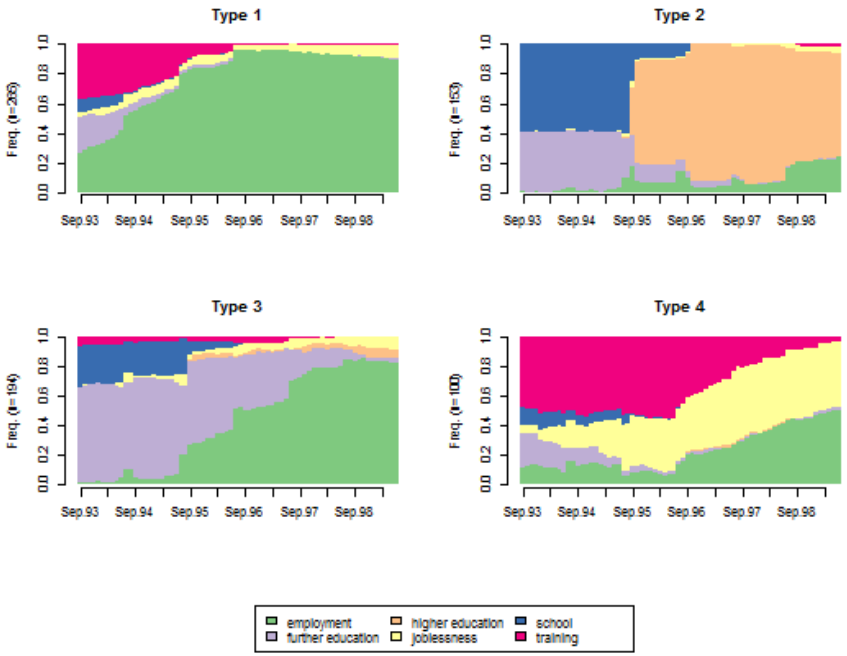
Customer Lifestage Segments
Preview

Sequence Analysis & Clustering

Example from TraMineR site – education / employment life stage clusters:



Can we find similar paths by our customers?



Stay tuned – we're working on that now!

What We Covered

- Why & how to segment you customers
- Three examples & one preview:
  - Customer Tenure Segments
  - RFM Based Segments
  - Cluster Based Segments
  - Life Stage Cluster Segments

Questions?
Comments?
Now is the time!

Learning More

- Tenure based segmentation & subscription survival
  - Subscription Survival for Fun & Profit: https://ds4ci.files.wordpress.com/2013/05/paw_sf2012_subscriptionsurvivalforfunandprofit.pdf
- RFM based segmentation
  - Workshop at N Cal DMA lunch group https://ds4ci.files.wordpress.com/2015/03/rfmb_dmanc_200905201.pdf
  - Using R for Customer Segmentation workshop at useR! 2008 Dortmund https://ds4ci.files.wordpress.com/2015/03/rfmb_dmanc_200905201.pdf
    - Also has sample data set & flexclust example
- Customer Classification
  - See above useR! 2008 workshop for details on flexclust
- Lifestage with TraMinerR http://traminer.unige.ch/
- Fader & Hardy http://www.brucehardie.com/talks.html
- Jim's Archives www.ds4ci.org/archives
- Contact: Jim@DS4CI.org