How to Segment?

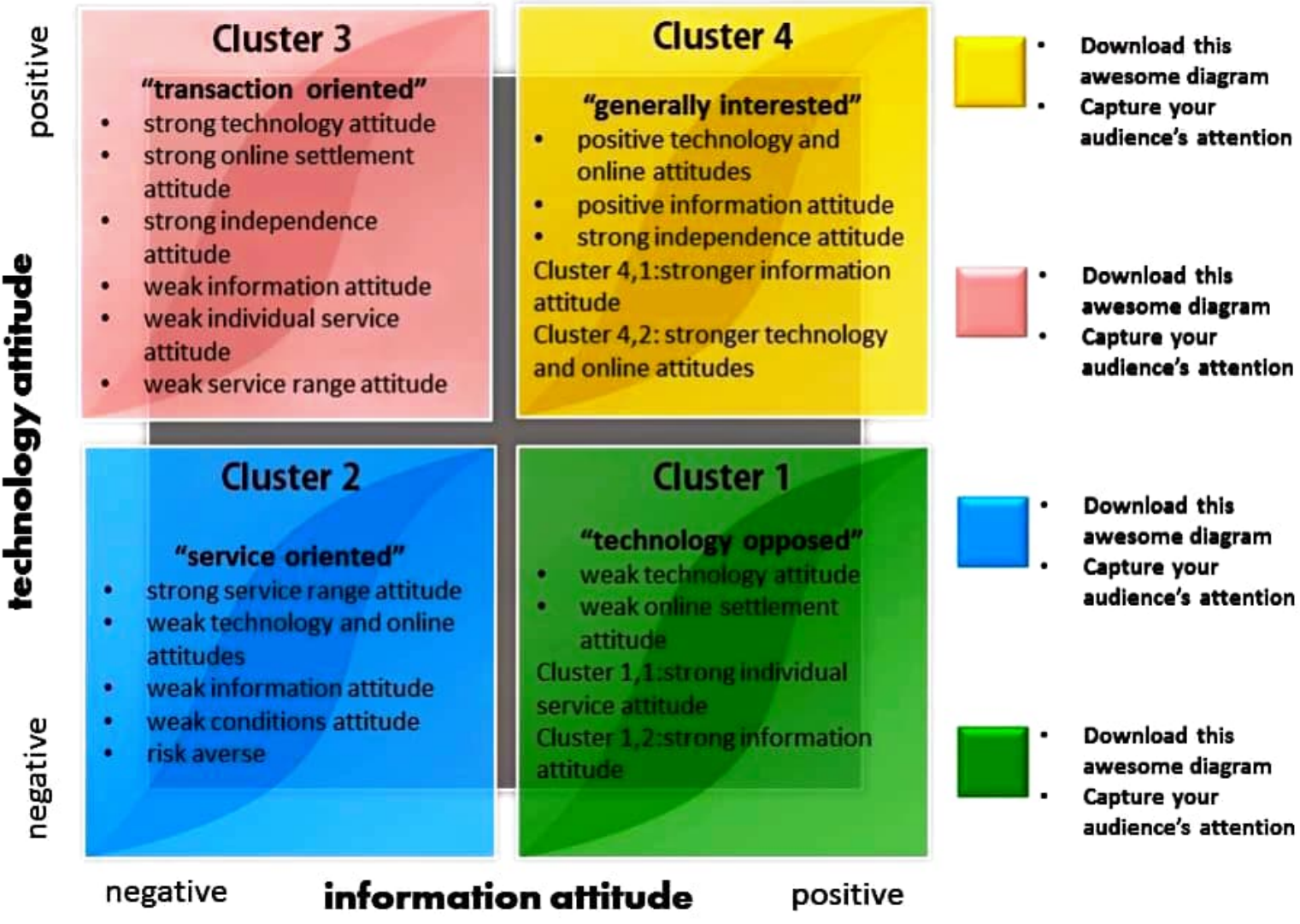
"We've broken your list into eighty-four subgroups. Our work here is done."

Many Segmentation Methods!

Today's Focus: Binary choice surveys

- Simplest of surveys to design & take.
- Cluster analysis is a great tool to understand how respondents fall into natural segments
- Methods also apply to any binary choice behavioral data sets.

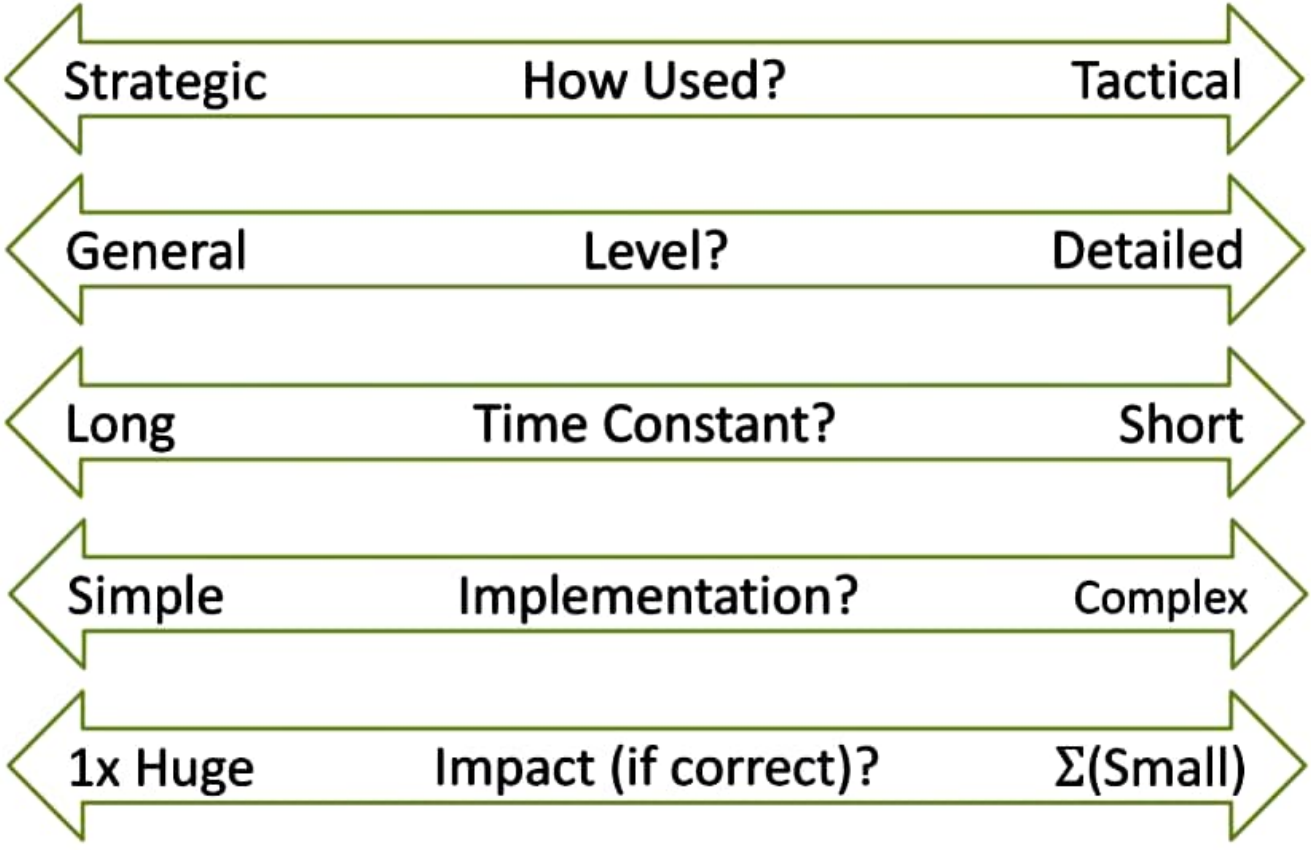
Customer Segmentation



Segmenting Binary Choice Data

- "Pick all that apply" type question.
  - Not picking is not the opposite of picking an attribute.
    - (item checked) <> NOT (item unchecked)
- *Totally unsupervised*. We only specify the number of clusters we want.
- Two *necessary criteria* for a "good" solution:
  1. The cluster solution is stable
     ~ Repeatable with different random starts
  2. The segments make sense to the business
     - Believable story AND is actionable AND has anticipated impact.

Customer Segmentation Themes



Tool we use: flexclust by Fritz Leisch

- Allows different distance measures
  - In particular, the Jaccard distance which is suited for binary survey data or optional properties lists.
  - 1 is a "yes" to the question - it is significant.
  - 0 is a "does not apply" – not opposite of "yes"
- Predict(kcca_object, newdata) to segment new customers.
- Additionally flexclust has very good diagnostic and visualization tools. As an R package, it leverages the rest of the R ecosystem.

Simple flexclust Run (1 of 2)

Set up input to flexclust:

```
library(flexclust)
data("volunteers")
vol_ch <- volunteers[-(1:2)]
vol.mat <- as.matrix(vol_ch)
```

Set up the parameters:

```
fc_cont <- new("flexclustControl") ## holds "hyperparameters"
fc_cont@tolerance <- 0.1
fc_cont@iter.max <- 30
fc_cont@verbose <- 1     ## verbose > 0 will show iterations
fc_family <- "ejaccard" ## Jaccard distance w/ centroid means
```

Invoke kcca(): “k-centroid cluster analysis”

```
fc_seed <- 577 ## Why we use this seed will become clear below
num_clusters <- 3 ## Simple example - only three clusters
set.seed(fc_seed)
vol.cl <- kcca(vol.mat, k = num_clusters, save.data = TRUE,
               control = fc_cont, family = kccaFamily(fc_family))
```

Simple flexclust Run (2 of 2)

First few iterations:

```
## 1 Changes / Distsum : 1415 / 951.9513
## 2 Changes / Distsum : 138 / 997.9507
## 3 Changes / Distsum : 39 / 998.6126
```

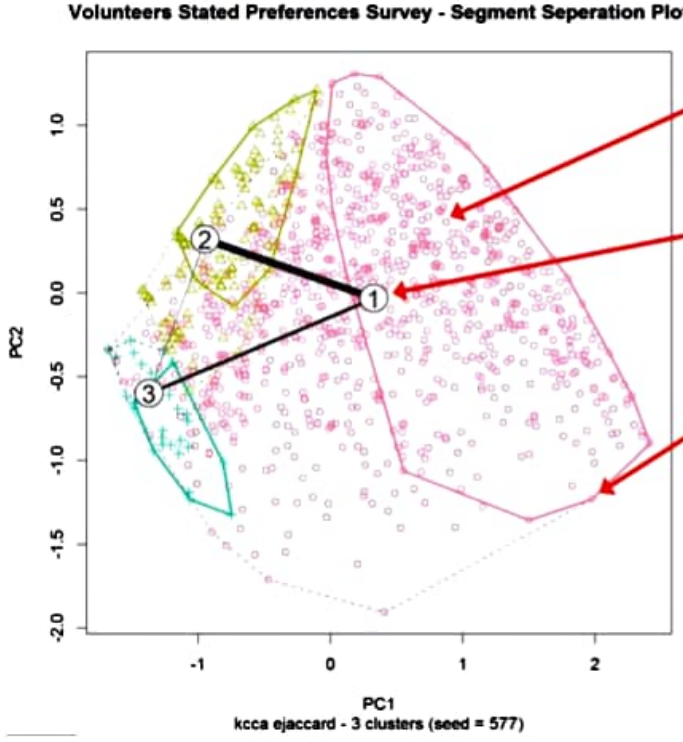
• • •

Results:

```
summary(vol.cl)
## kcca object of family 'ejaccard'
## call:
## kcca(x = vol.mat, k = num_clusters, family = kccaFamily(fc_family),
##     control = fc_cont, save.data = TRUE)
##
## cluster info:
##   size   av_dist  max_dist separation
## 1 1078 0.6663440 1.0000000  0.6455246
## 2  258 0.7388715 1.0000000  0.6568168
## 3   79 0.8962851 0.9569892  0.8284482
##
## no convergence after 30 iterations
## sum of within cluster distances: 979.7542
```

Segment Separation Plot

```
vol.pca <- prcomp(vol.mat) ## plot on first two principal components
plot(vol.cl, data = vol.mat, project = vol.pca, main = . . .)
```



Each respondent plotted against the first two principal components of data. Color is cluster assignment.

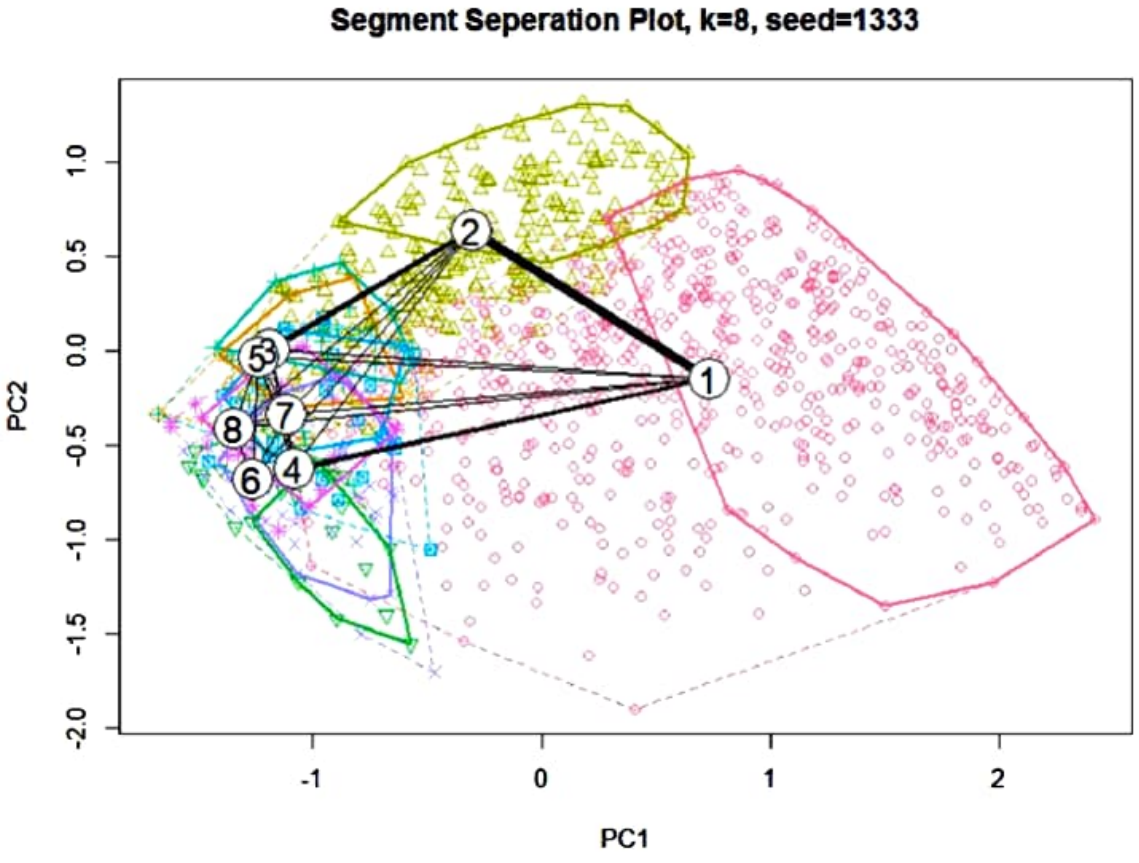
Centroid of each cluster. A thin line to other centroid indicates better separation (in real problem space)

Solid line encloses 50% of respondents in cluster; dotted 95%.

Purpose: Help business partners visualize clusters and how respondents fall within cluster boundaries. IOW, are clusters "real"?

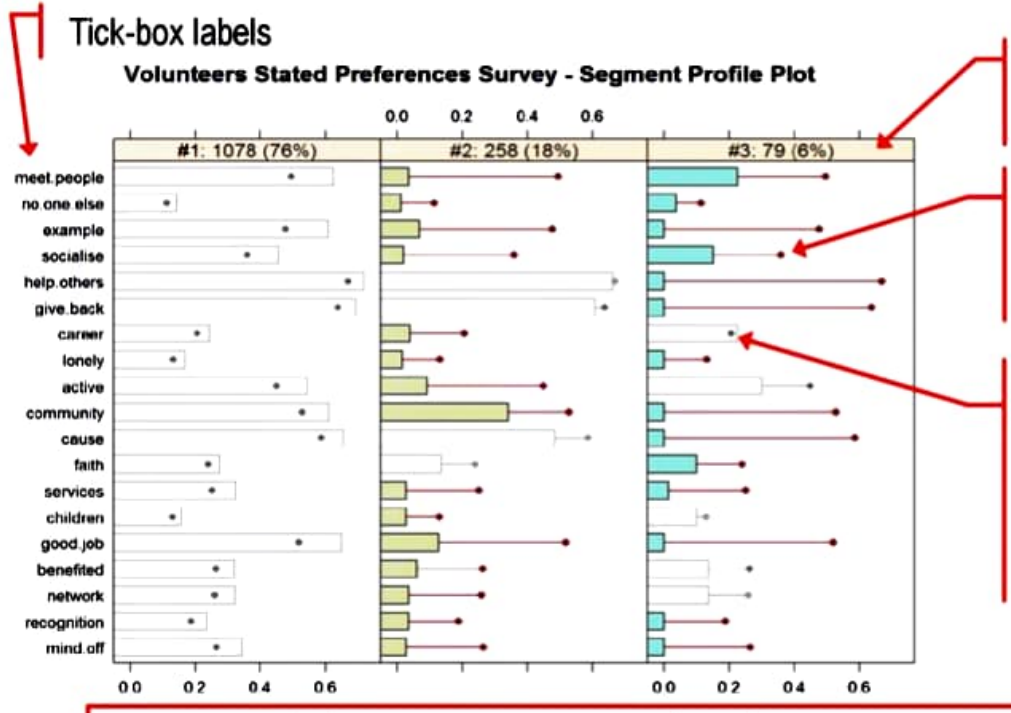
Also known as "neighborhood plot."

Segment Separation for “best” k = 8 (seed = 1333)



Segment Profile Plot

```
barchart(vol.cl, strip.prefix = "#", shade = TRUE,
         layout = c(vol.cl@k, 1), main = . . .)
```



Header: segment #, Count, & % total

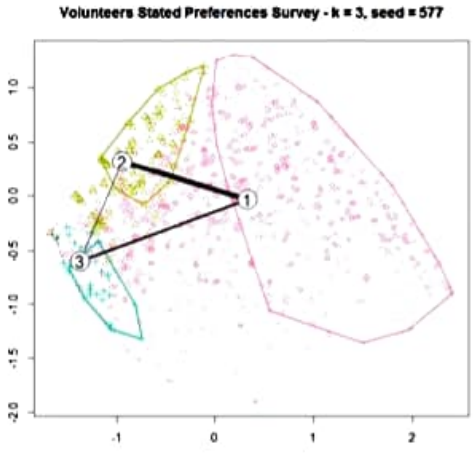
Bar: proportion of response in cluster.
Red line/dot: overall proportion

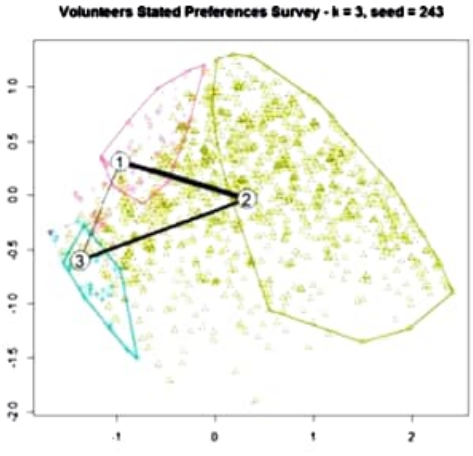
Greyed out when response not important to differentiate from other clusters.
BUT, can still be an important characteristic of cluster

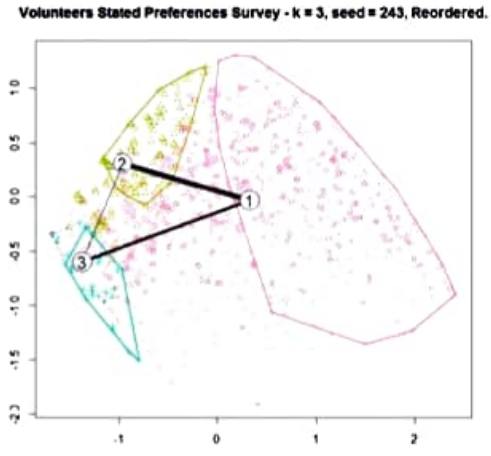
Purpose: Help business partners translate clusters into segment stories.
IOW, describe the clusters in business friendly terms.

The Numbering Problem

Two different seeds have nearly equal solutions, but are labeled differently:







fc_reorder {CustSegs}

Reorder clusters in a kcca object.

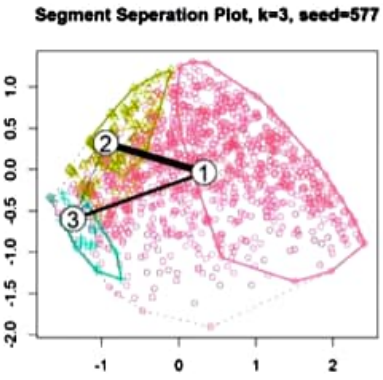
Usage: `fc_reorder(x, orderby = "decending size")`

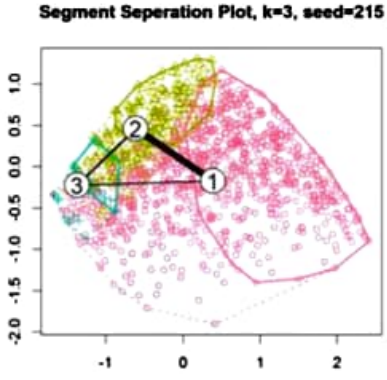
Simple Method to Explore Stability

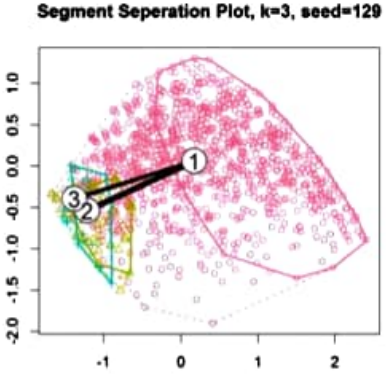
- For a given k, run a few hundred solutions (incrementing seed each time):
  - Re-order clusters in descending size order
  - Save: k, seed, cluster #, & count
- Call Size_1 the count for 1st cluster;
  Size_2 the count for 2nd cluster.
- Scatter plot w/ 2D density curves: Size_2 x Size_1
- Solve for peak location

The Stability Problem

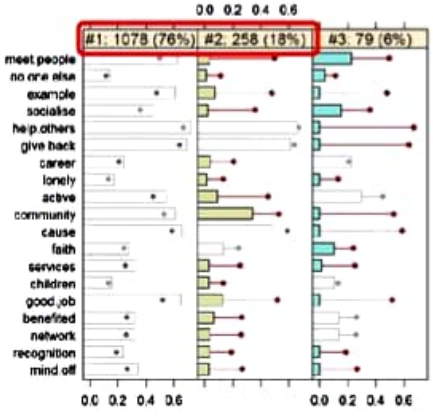
Three different seeds have quite different solutions:

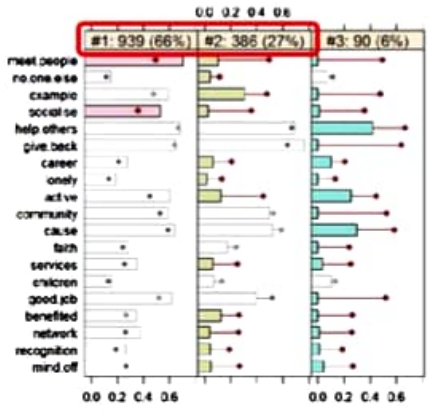


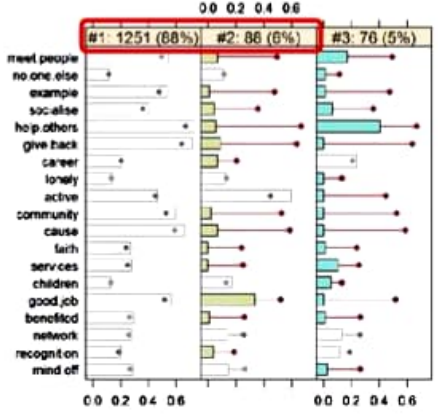




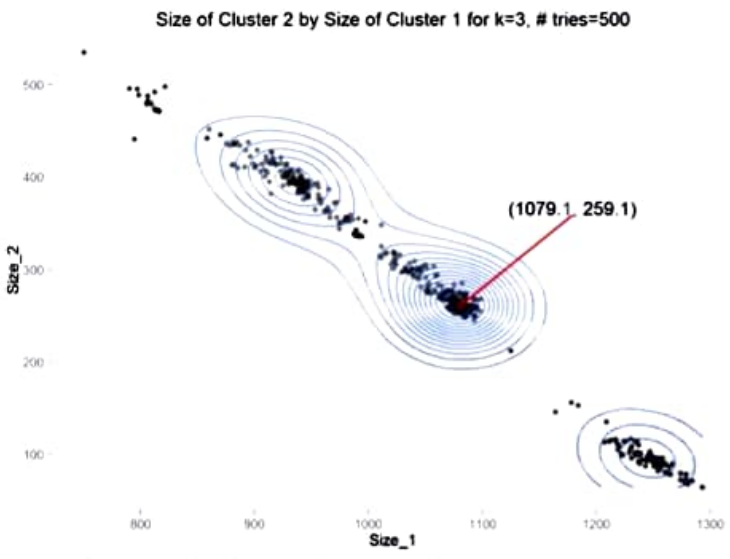
We need a simple way to classify each solution – just use sizes of two biggest clusters:

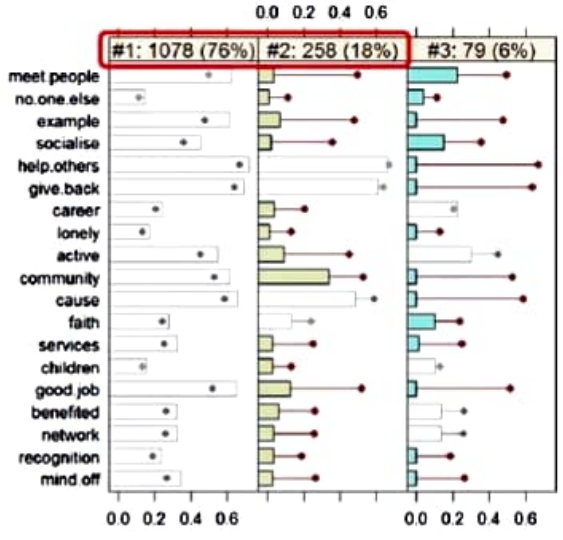






Stability Plot of kcca Solutions for k=3





fc_rclust {CustSegs}

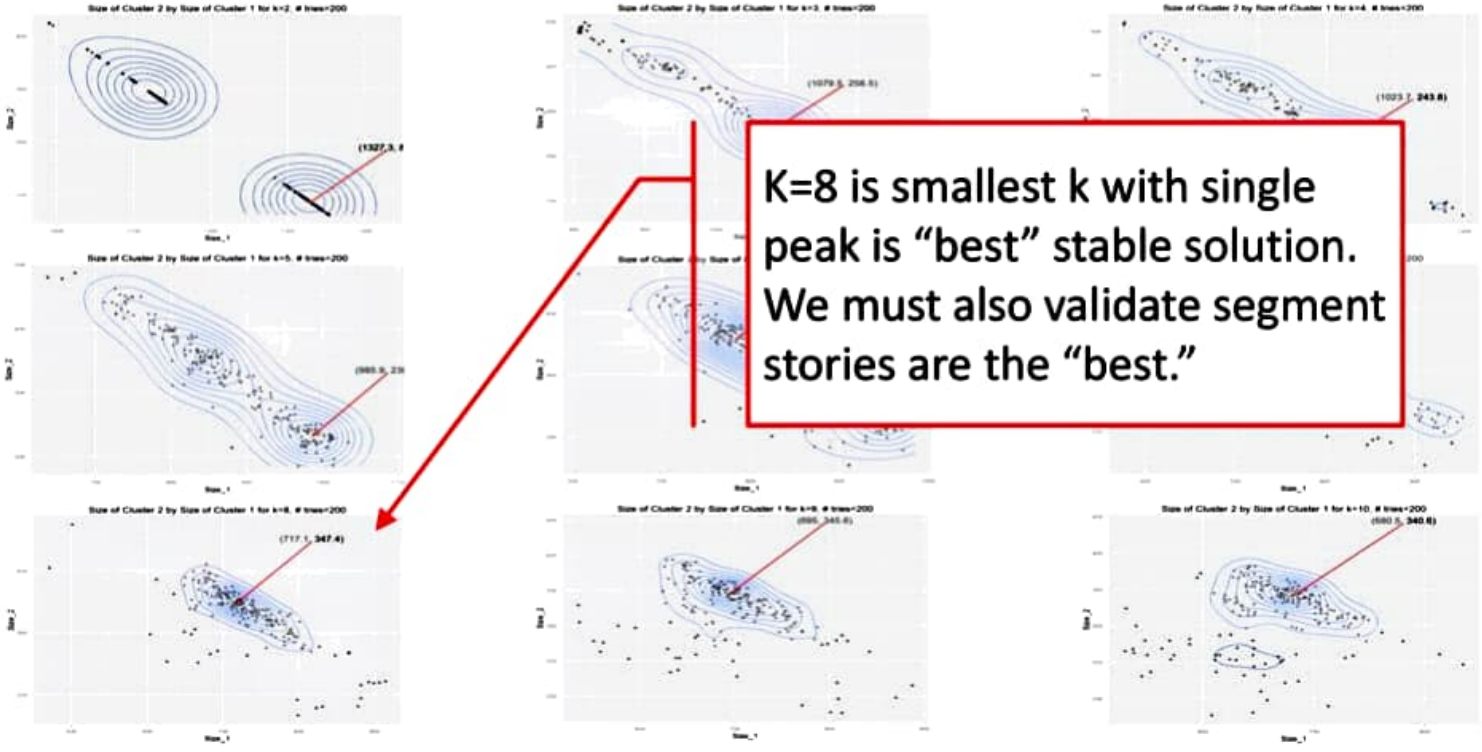
Generate a List of Random kcca Objects.

Usage:

```
fc_rclust(x, k, fc_cont, nrep = 100,
  fc_family, verbose = FALSE, FUN = kcca,
  seed = 1234, plotme = TRUE)
```

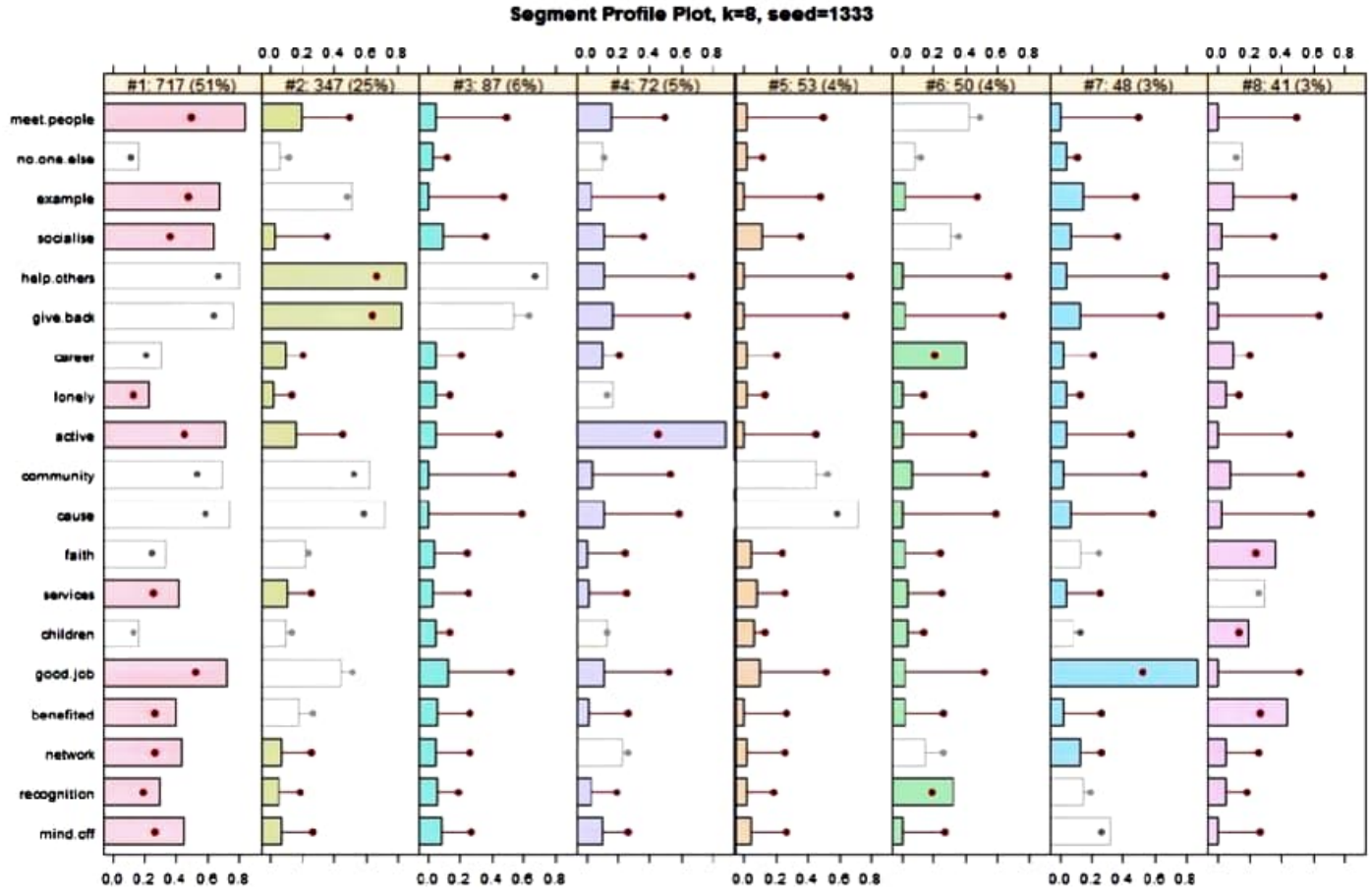
The "Best" k Problem

Generate stability plots for k = 2, 3, …, 10:



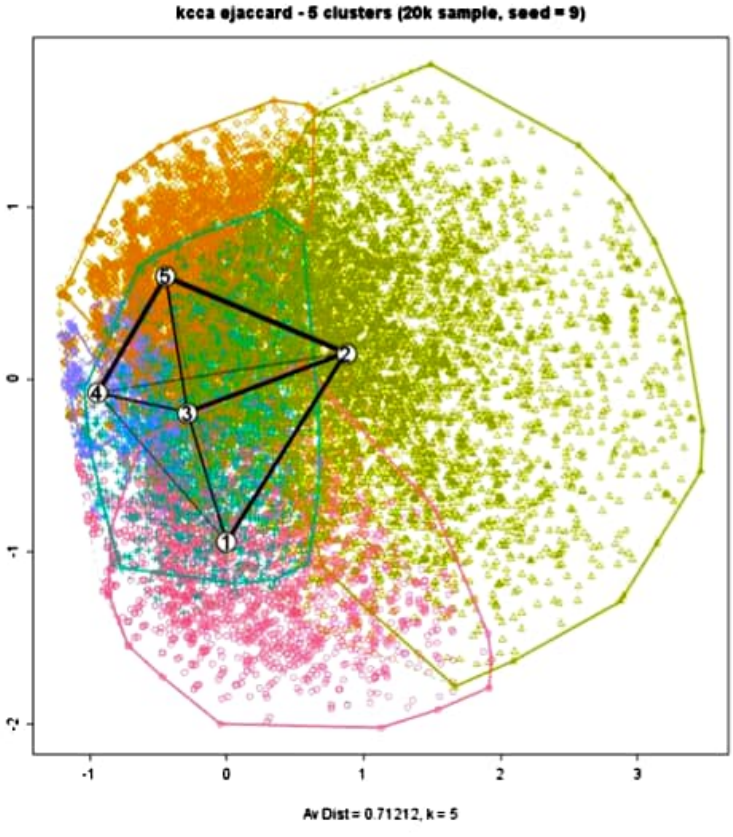
K=8 is smallest k with single peak is "best" stable solution. We must also validate segment stories are the "best."

Profile Plot for “best” k = 8 (seed = 1333)

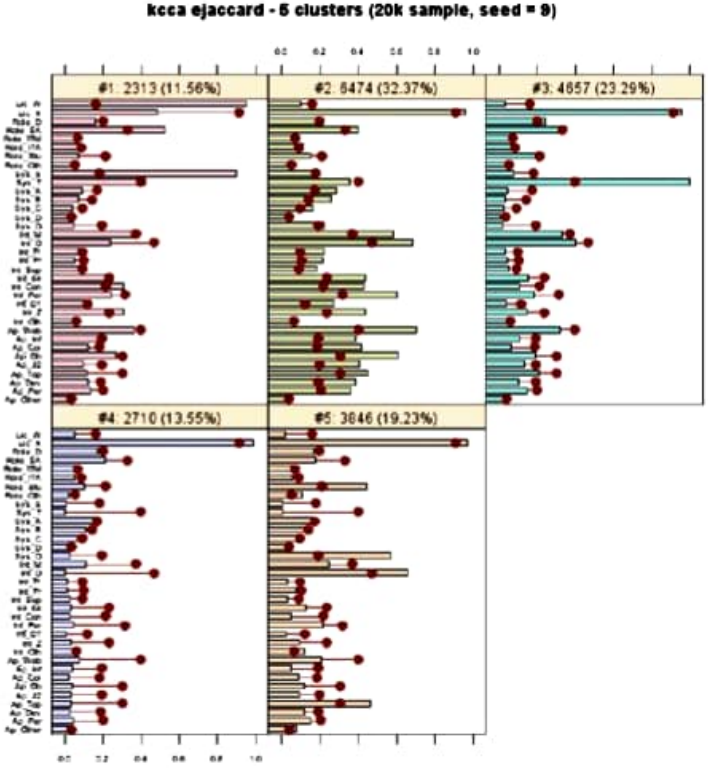


Example 1 - The 5-cluster solution

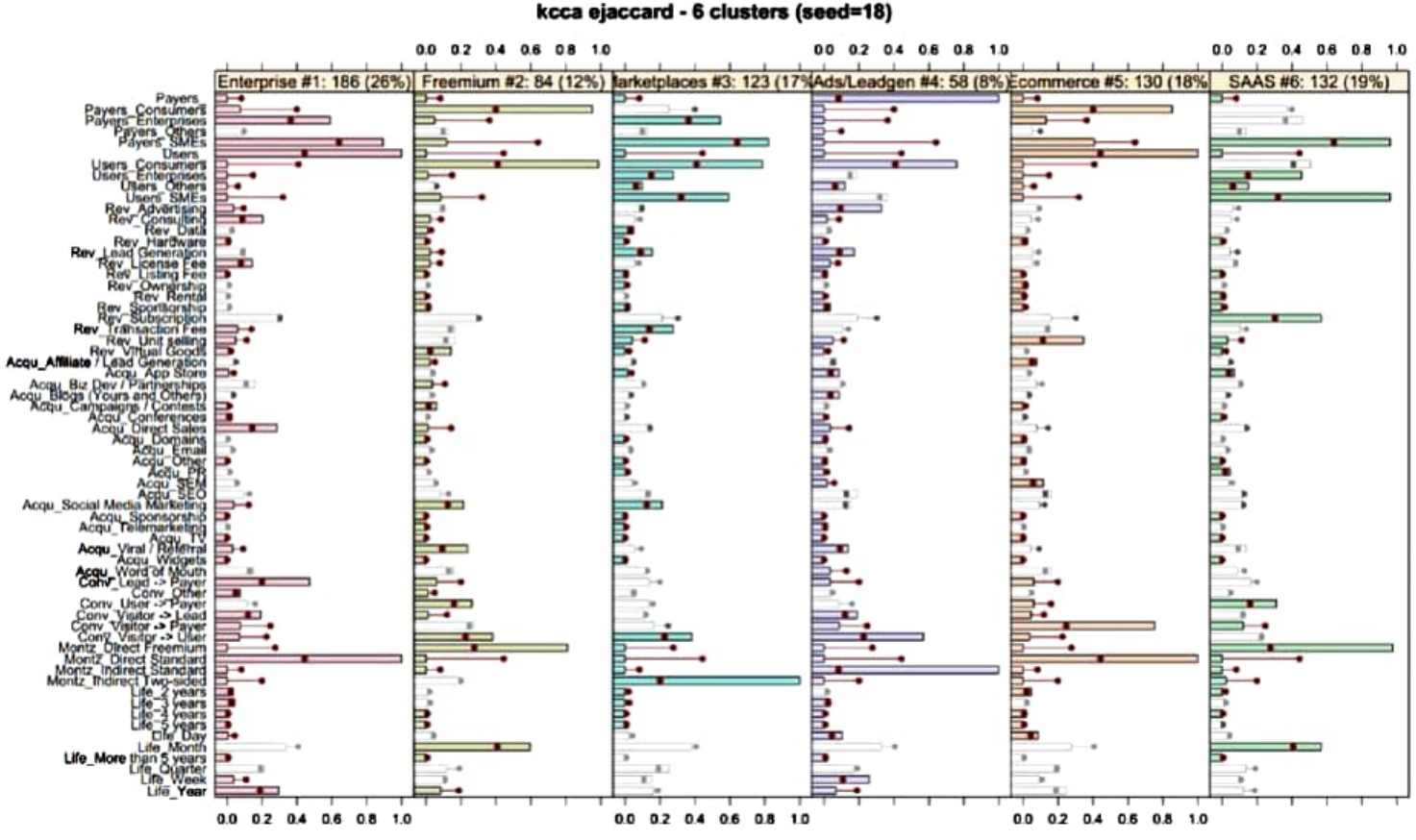
The 20k subjects plotted over the first two principal components:



The 5 clusters showing distribution of responses to each question:



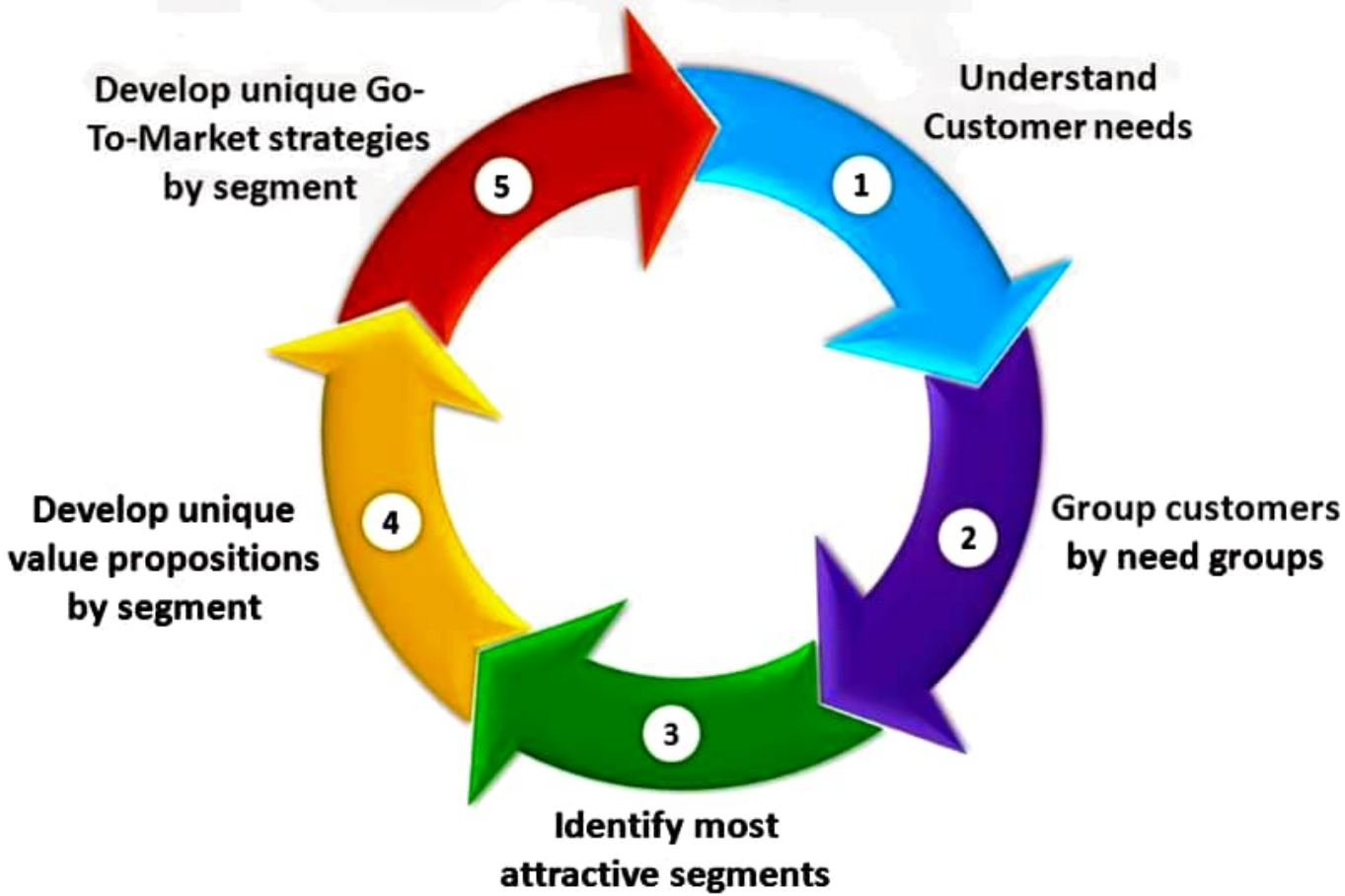
Example 2 – the 6-cluster solution



Example 2 – Business Attributes

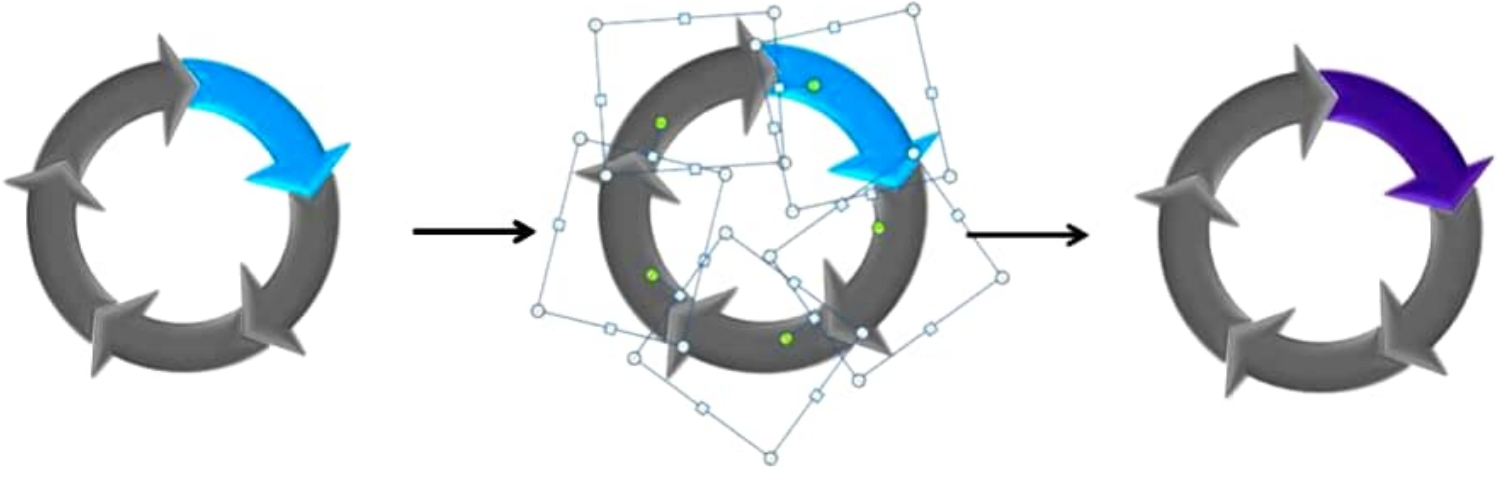
- ~1k respondents to "nature of your business" survey
- 62 check boxes or radio buttons
  - In six topics
  - Some are required
  - Coded as binary responses
- Goal: come up with "a few" segments to characterize the fundamental nature of the on-line business.
- 6-cluster solution: Enterprise, Freemium, Marketplace, Ads/Leadgen, Ecommerce, SAAS.

Customer Segmentation Process

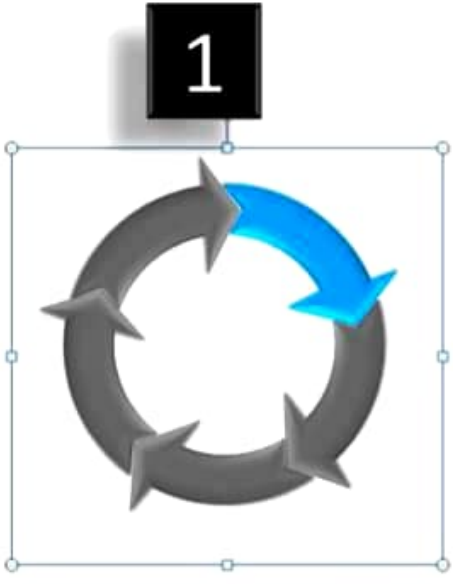


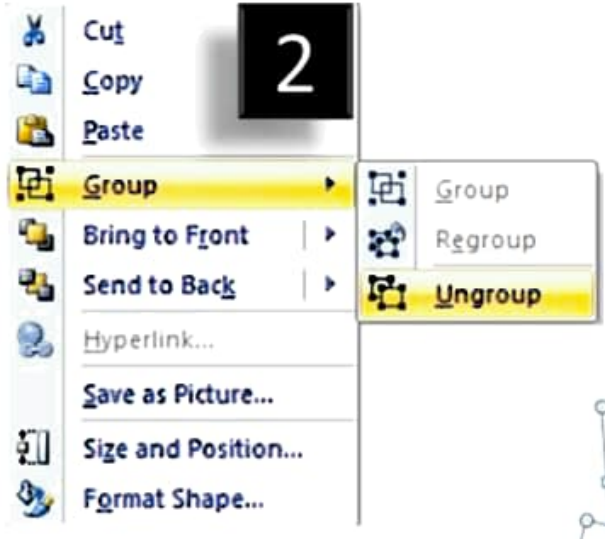
All images are 100% editable in Powerpoint

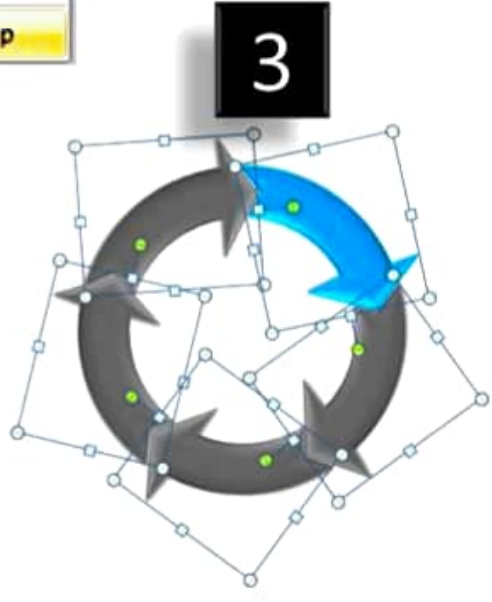
"Change color, size and orientation of any icon to your liking"



Ungrouping the object

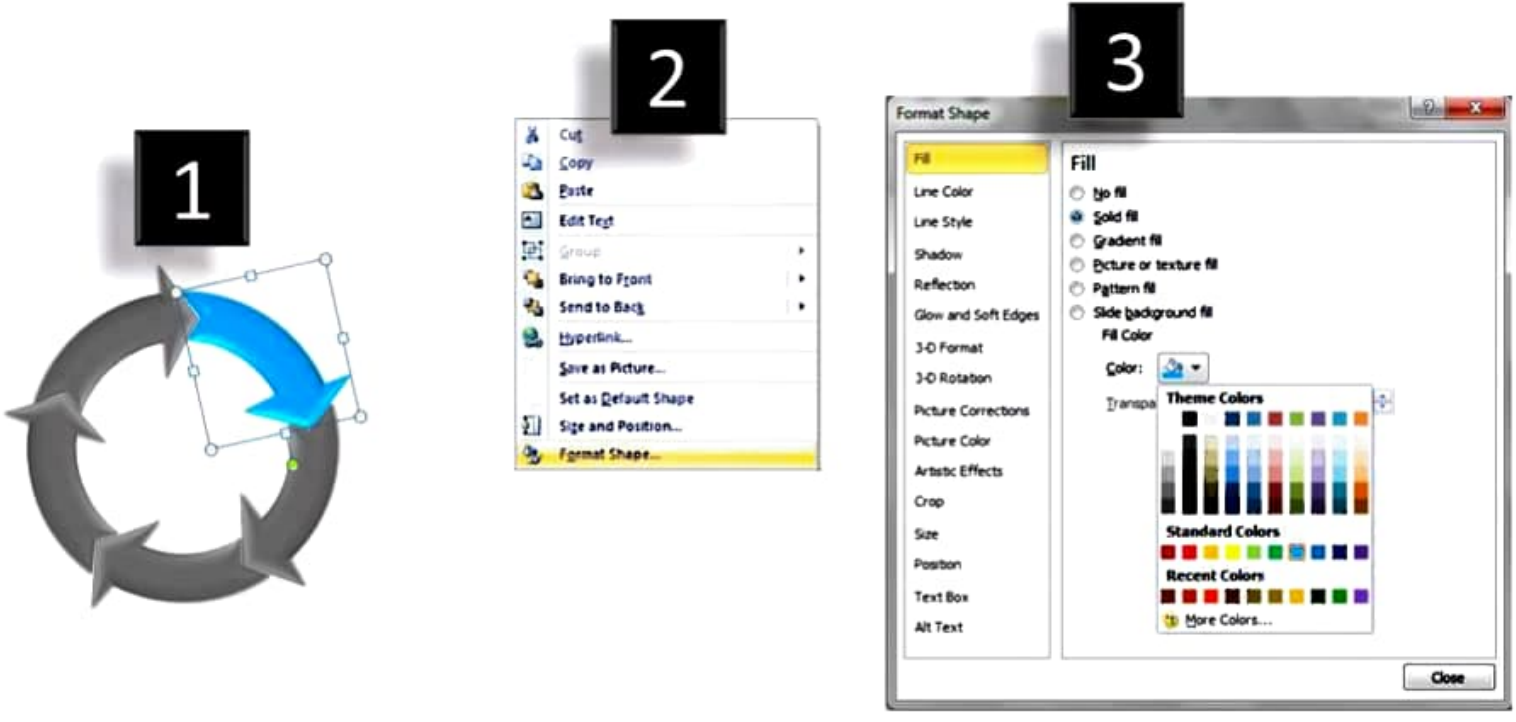






1. Right click the object.
2. Choose Group and then Ungroup.
3. Click beside the object and drag the arrow over it.

Edit Color



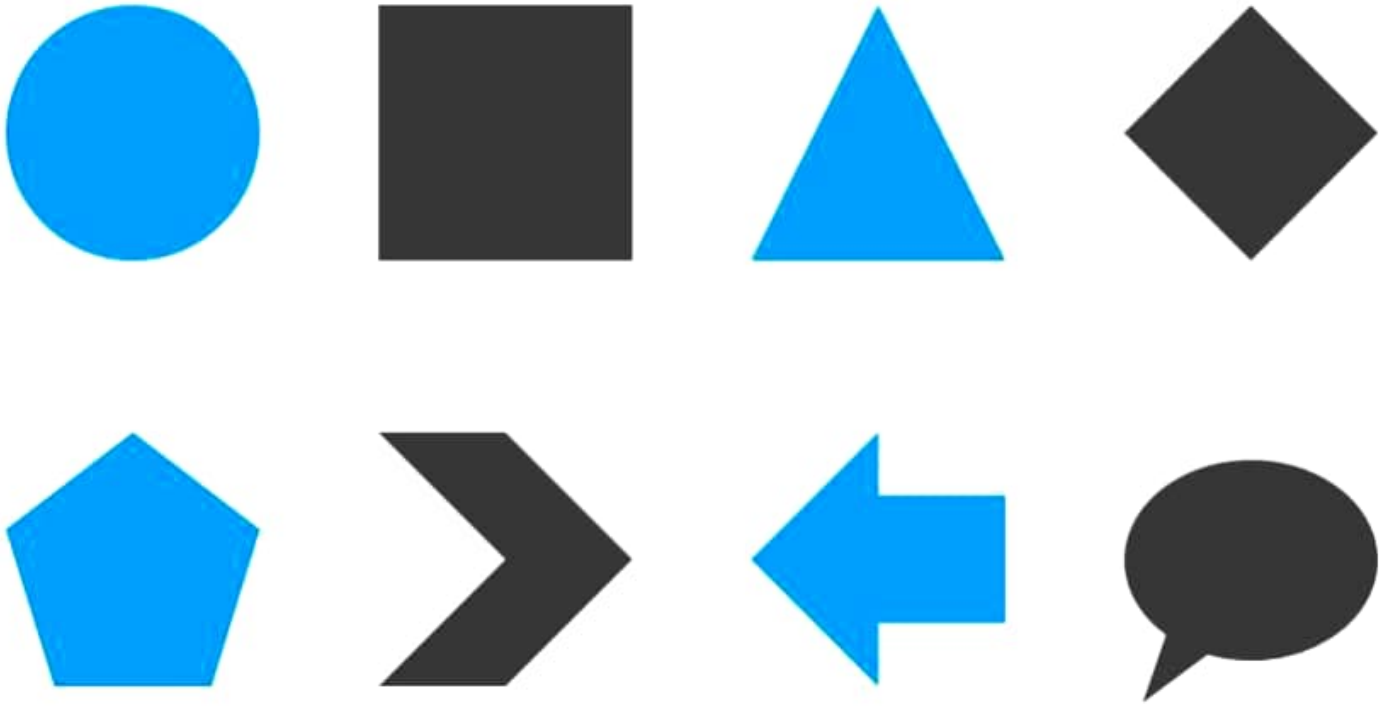
1. Select the shape to change the color and Right click the object(click any object which you want to change color)
2. Choose Format Shape in the dialog box.
3. Choose “Fill” in the Format Shape box then “Solid” or “Gradient” depending on the appearance of the object. Change colour as shown in the picture.

Customer Segmentation Powerpoint Presentation

This slide is 100% editable. Adapt it to your needs and capture your audience's attention

Instructions To Edit This Slide

This slide has been made in PowerPoint using basic PowerPoint shapes, and all graphics are 100% editable in PowerPoint.
You can change color, size and orientation of any element to your liking, and add/edit any piece of text anywhere.

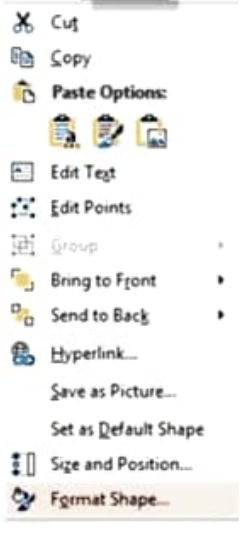


Resize Any Shape And Edit Color

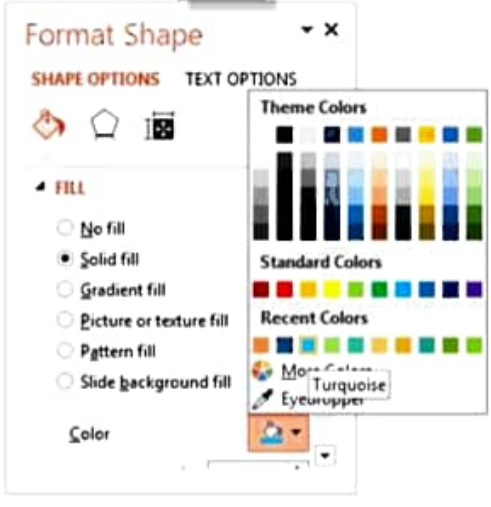
The entire slide has been made in PowerPoint itself, and you can edit the color of any component in this graphic. To change the color, just right click the object to highlight it, and then follow instructions -

1

2



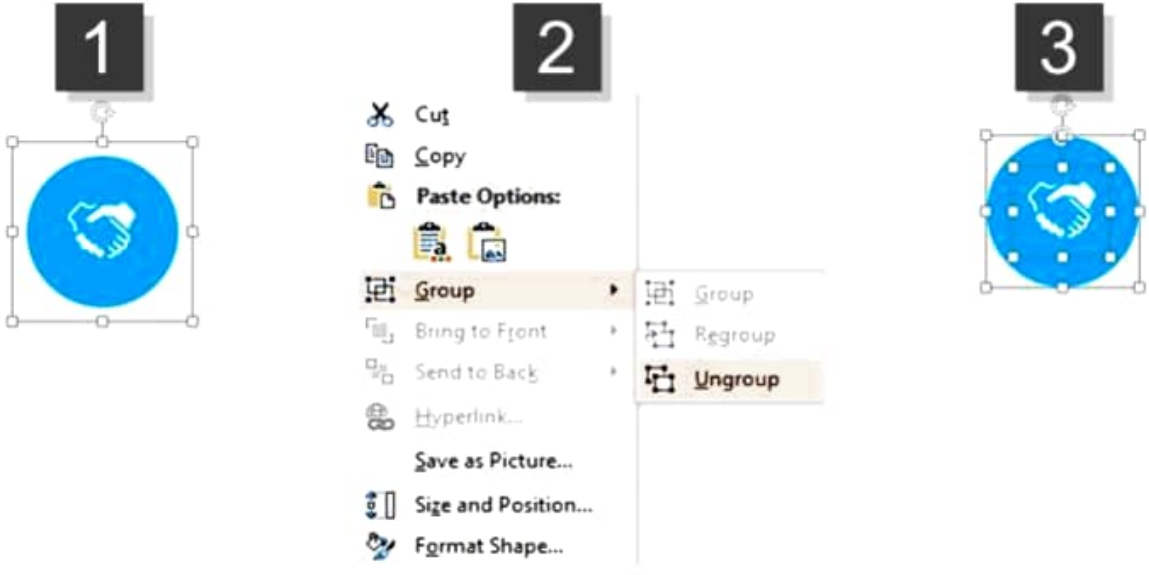
3



1. Right click on the object or shape that you want to change the color of.
2. Choose Format Shape in the dialog box.
3. Choose "Fill" in the Format Shape box then "Solid" or "Gradient" depending on the appearance of the object. Change colour as shown in the picture.

Use Your Own Icons In Any Slide

You can replace the icons in this slide with any icons you like from our site, or completely remove them if required -



The icons we have shown are just for demo purposes. You can replace these icons with the correct icons to suit your message. We have a large collection of icons to choose from.

1. To replace any icon, double click on it and select it, and then right click it to ungroup it.
2. Delete the icon and replace it with any icon or graphic from our site.

Select From Thousands Of Editable Icons On Every Topic

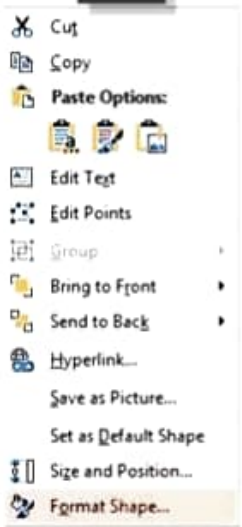
These icons are available at the Icons section on www.slideteam.net

Replace Image In This Slide With Your Own

The image we have shown in this slide is for demo purposes only. You can replace this image with your own -

1

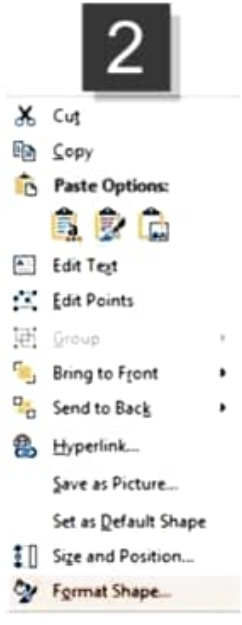
2

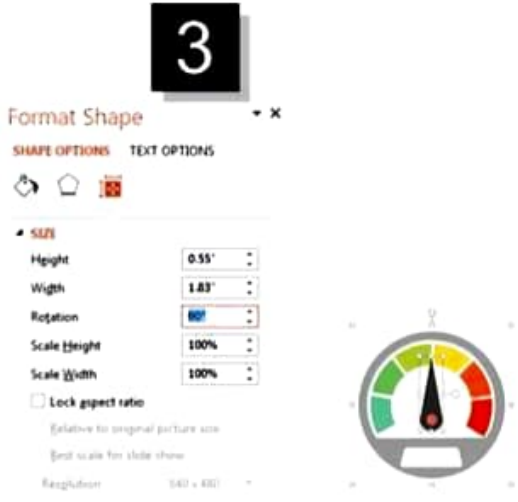


3

Change Orientation Of Any Graphic

You can change the angle of any graphic in this slide e.g. if the slide has a needle, or any other shape, you can just select it and change the angle to suit your requirement.





1. For example, you can select the Dashboard Needle by double clicking on it, and then right click it once to highlight it.
2. Choose Format Shape in the dialog box.
3. Choose "Size" in the Format Shape box then "Rotation" and enter your preferred rotation angle. Once the angle is to your liking, adjust the position of the needle back to the centre. You can also just rotate it using the rotate icon on the top of the shape itself once it is highlighted.