Reducing Data Complexity

Luis Francisco Gomez Lopez

FAEDIS

2024-02-25

Contents

- Please Read Me
- Purpose
- Consumer brand perception survey
- References

Please Read Me

- This presentation is based on (Chapman and Feit 2019, chap. 8)

Purpose

-

Consumer brand perception survey

- On a scale from 1 to 10, where 1 is least and 10 is most, how `<perceptual adjective>` is `<brand>`?

- 100 respondents rate 10 brands on 9 perceptual adjectives
  - **perform**: has strong performance $(1, 2, \dots, 10)$
  - **leader**: is a leader in the field $(1, 2, \dots, 10)$
  - **latest**: has the latest products $(1, 2, \dots, 10)$
  - **fun**: is fun $(1, 2, \dots, 10)$
  - **serious**: is serious $(1, 2, \dots, 10)$
  - **bargain**: products are a bargain $(1, 2, \dots, 10)$
  - **value**: products are a good value $(1, 2, \dots, 10)$
  - **trendy**: is trendy $(1, 2, \dots, 10)$
  - **rebuy**: I would buy from `<brand>` again $(1, 2, \dots, 10)$
  - **brand**: coffee brand rated by a consumer $(a, b, \dots, j)$

Consumer brand perception survey

- Import data

```
consumer_brand <- read_csv("http://goo.gl/IQl8nc")
consumer_brand |> head(n = 5)
```

```
# A tibble: 5 x 10
  perform leader latest   fun serious bargain value trendy rebuy brand
    <dbl>  <dbl>  <dbl> <dbl>   <dbl>   <dbl> <dbl>  <dbl> <dbl> <chr>
1       2      4      8     8       2       9     7      4     6 a
2       1      1      4     7       1       1     1      2     2 a
3       2      3      5     9       2       9     5      1     6 a
4       1      6     10     8       3       4     5      2     1 a
5       1      1      5     8       1       9     9      1     1 a
```

Consumer brand perception survey

- Transform data

```
consumer_brand_scale <- consumer_brand |>
  mutate(across(perform:rebuy,
               .fns = ~ scale(x = .x,
                              center = TRUE,
                              scale = TRUE)[,1]))

consumer_brand_scale |> head()
```

```
# A tibble: 6 x 10
  perform leader latest    fun serious bargain  value trendy  rebuy brand
    <dbl>  <dbl>  <dbl>  <dbl>   <dbl>   <dbl>  <dbl>  <dbl>  <dbl> <chr>
1  -0.777 -0.160  0.586  0.704  -0.836    1.78   1.11 -0.445  0.893 a
2  -1.09  -1.31  -0.713  0.340  -1.20    -1.22  -1.39  -1.17  -0.679 a
3  -0.777 -0.543 -0.388  1.07   -0.836    1.78   0.276 -1.54   0.893 a
4  -1.09   0.607  1.24   0.704  -0.476  -0.0971  0.276 -1.17  -1.07  a
5  -1.09  -1.31  -0.388  0.704  -1.20     1.78   1.94  -1.54  -1.07  a
6  -0.777  1.37   0.911 -0.389  -0.476    1.40   1.11  -1.54  -0.679 a
```

Consumer brand perception survey

- Summarize data
  - Ups the table is really big!!! Try it in your console to see the complete table

```
consumer_brand_scale |> skim()
```

Table 1: Data summary

Name	consumer_brand_scale
Number of rows	1000
Number of columns	10
Column type frequency:	
character	1
numeric	9
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
brand	0	1	1	1	0	10	0

Consumer brand perception survey

- Correlation matrices
    - Pearson correlation coefficients for samples in a tibble

```
correlation_matrix <- consumer_brand_scale |>
  select(perform:rebuy) |>
  correlate(use = "pairwise.complete.obs", # There are NA values
            method = "pearson",
            diagonal = NA)
correlation_matrix # Ups!!! The tibble is wide. Check out the tibble in your console
```

```
# A tibble: 9 x 10
  term    perform  leader   latest     fun  serious  bargain    value   trendy
  <chr>     <dbl>   <dbl>    <dbl>   <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1 perform NA        0.500  -0.122  -0.256    0.359   0.0571   0.102   0.00873
2 leader   0.500  NA        0.0269 -0.290    0.571   0.0331   0.118   0.0665
3 latest  -0.122   0.0269 NA        0.245    0.00995 -0.254  -0.343   0.628
4 fun     -0.256  -0.290   0.245   NA       -0.281  -0.0666  -0.145   0.128
5 serious  0.359   0.571   0.00995 -0.281   NA       -0.00266 0.0238  0.121
6 bargain  0.0571  0.0331 -0.254   -0.0666  -0.00266 NA        0.740  -0.351
7 value    0.102   0.118  -0.343   -0.145    0.0238   0.740  NA       -0.435
8 trendy   0.00873 0.0665  0.628    0.128    0.121   -0.351  -0.435  NA
9 rebuy    0.307   0.209  -0.397   -0.237    0.181    0.467   0.506  -0.298
# i 1 more variable: rebuy <dbl>
```
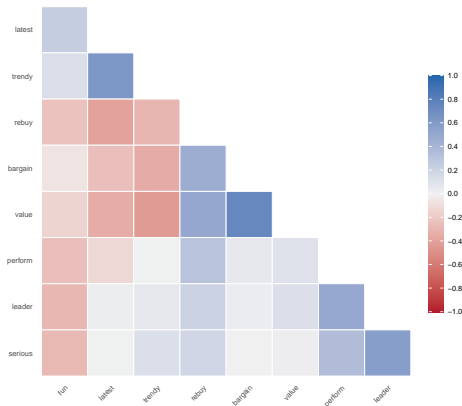
Consumer brand perception survey

- Correlation matrices

  - Pearson correlation coefficients for samples in a tibble

```
correlation_matrix |>
  autoplot(method = "HC", # Hierarchical clustering: More details in Chapter 11
           triangular = "lower")
```

Consumer brand perception survey

- Mean ratings by brand

```
brand_mean <- consumer_brand_scale |>
  group_by(brand) |>
  summarise(across(everything(), .fns = mean))
brand_mean
```

```
# A tibble: 10 x 10
   brand perform leader  latest    fun serious bargain   value  trendy   rebuy
   <chr>   <dbl>  <dbl>   <dbl>  <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
 1 a     -0.886 -0.528  0.411   0.657 -0.919   0.214   0.185  -0.525  -0.596
 2 b      0.931  1.07   0.726  -0.972  1.18    0.0416  0.151   0.740   0.237
 3 c      0.650  1.16  -0.102  -0.845  1.22   -0.607  -0.441   0.0255 -0.132
 4 d     -0.680 -0.593  0.352   0.187 -0.692  -0.881  -0.933   0.737  -0.494
 5 e     -0.564  0.193  0.456   0.296  0.0421  0.552   0.418   0.139   0.0365
 6 f     -0.0587 0.270 -1.26   -0.218  0.589   0.874   1.02   -0.813   1.36
 7 g      0.918 -0.168 -1.28   -0.517 -0.534   0.897   1.26   -1.28    1.36
 8 h     -0.0150 -0.298 0.502   0.715 -0.141  -0.738  -0.783   0.864  -0.604
 9 i      0.335 -0.321  0.356   0.412 -0.149  -0.255  -0.803   0.591  -0.203
10 j     -0.630 -0.789 -0.154   0.285 -0.602  -0.0971 -0.0738 -0.481  -0.962
```

Consumer brand perception survey

- Mean ratings by brand

```
brand_mean_longer <- brand_mean |>
  pivot_longer(cols = perform:rebuy,
               names_to = "perceptual_adjectives",
               values_to = "value_mean") |>
  mutate(brand = fct_reorder(.f = brand, .x = value_mean),
         perceptual_adjectives = fct_reorder(.f = perceptual_adjectives, .x = value_mean))
brand_mean_longer
```

```
# A tibble: 90 x 3
   brand perceptual_adjectives value_mean
   <fct> <fct>                      <dbl>
 1 a     perform                   -0.886
 2 a     leader                    -0.528
 3 a     latest                     0.411
 4 a     fun                        0.657
 5 a     serious                   -0.919
 6 a     bargain                    0.214
 7 a     value                      0.185
 8 a     trendy                    -0.525
 9 a     rebuy                     -0.596
10 b     perform                    0.931
# i 80 more rows
```
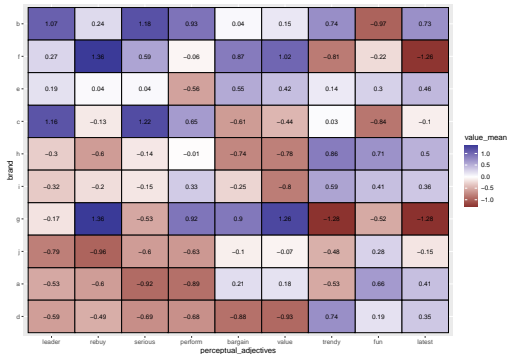
Consumer brand perception survey

- Heat map mean ratings by brand

```
brand_mean_longer |>
  ggplot(aes(x = perceptual_adjectives, y = brand)) +
  geom_tile(aes(fill = value_mean),
            color = "black") +
  geom_text(aes(label = round(value_mean, digits = 2)),
            size = 3) +
  scale_fill_gradient2()
```

Consumer brand perception survey

- Principal component analysis (PCA) and perceptual maps

References

Chapman, Chris, and Elea McDonnell Feit. 2019. *R For Marketing Research and Analytics*. 2nd ed. 2019. Use R! Cham: Springer International Publishing : Imprint: Springer. https://doi.org/10.1007/978-3-030-14316-9.