

# LDAによるホテルブランドマッピング

---

B5EB1106 酒井洋輔

# Contents

---

1.Motivation

2.Review previous research

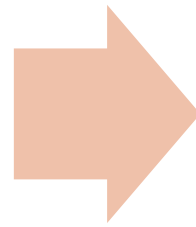
3.Mapping the brand of hotels

4.Challenges

# Motivation

---

ビジネスホテル  
の均質化



差別化の  
必要性

# Motivation

## Positioning

- 自社製品の差別化
- 競合力の向上
- 未開拓のポジションの発見
- **Positioning map**

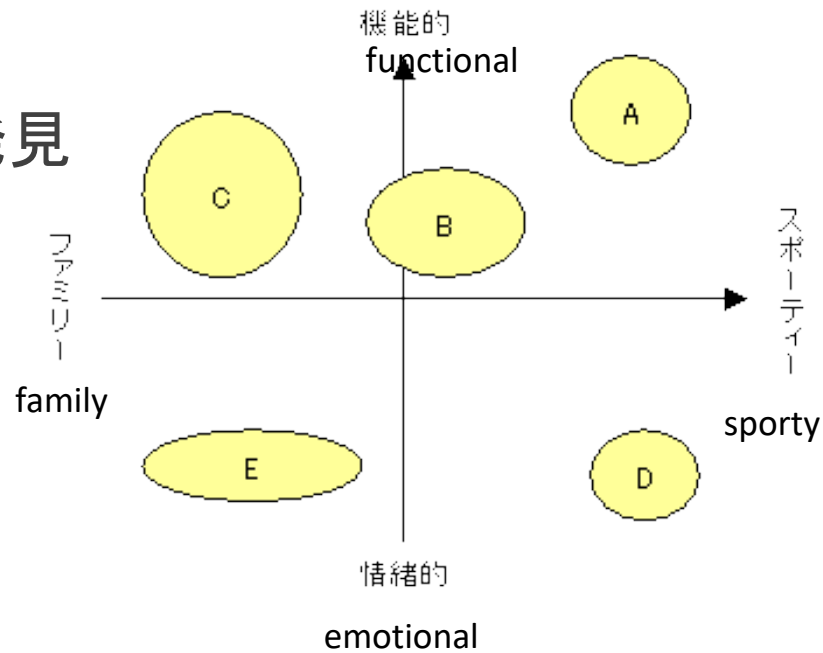


Image from  
[http://www.nsspirit-cashf.com/manage/positioning\\_analysis.html](http://www.nsspirit-cashf.com/manage/positioning_analysis.html)

# Motivation

## 消費者の知覚に基づき、ブランドをマッピングする

↓ questionnaire

| Q No | 質問文  | 中項目        | 列No | 項目名                 | 回答形態 | 回答番号 | 回答内容 |
|------|--|------------|-----|---------------------|------|------|------|
| Q4   | それぞれのアルコール飲料を飲むとどのような気分になると思いますか、あてはまると思うものを1つお選びください。※ご存知ない、飲んだことがない、場合でも、イメージでお答えください。 | アサヒスーパードライ | 39  | Q4-1 飲むと、幸せな気分になる   | 単一   | 同上   |      |
|      |  |            | 40  | Q4-1 飲むと、ほっとした気分になる | 単一   | 同上   |      |
|      |  |            | 41  | Q4-1 飲むと、爽やかな気分になる  | 単一   | 同上   |      |
|      |  | キリンラガービール  | 42  | Q4-2 飲むと、幸せな気分になる   | 単一   | 同上   |      |
|      |  |            | 43  | Q4-2 飲むと、ほっとした気分になる | 単一   | 同上   |      |
|      |  |            | 44  | Q4-2 飲むと、爽やかな気分になる  | 単一   | 同上   |      |
|      |  | エビスビール     | 45  | Q4-3 飲むと、幸せな気分になる   | 単一   | 同上   |      |
|      |  |            | 46  | Q4-3 飲むと、ほっとした気分になる | 単一   | 同上   |      |
|      |  |            | 47  | Q4-3 飲むと、爽やかな気分になる  | 単一   | 同上   |      |

Factor analysis

↓ brand mapping

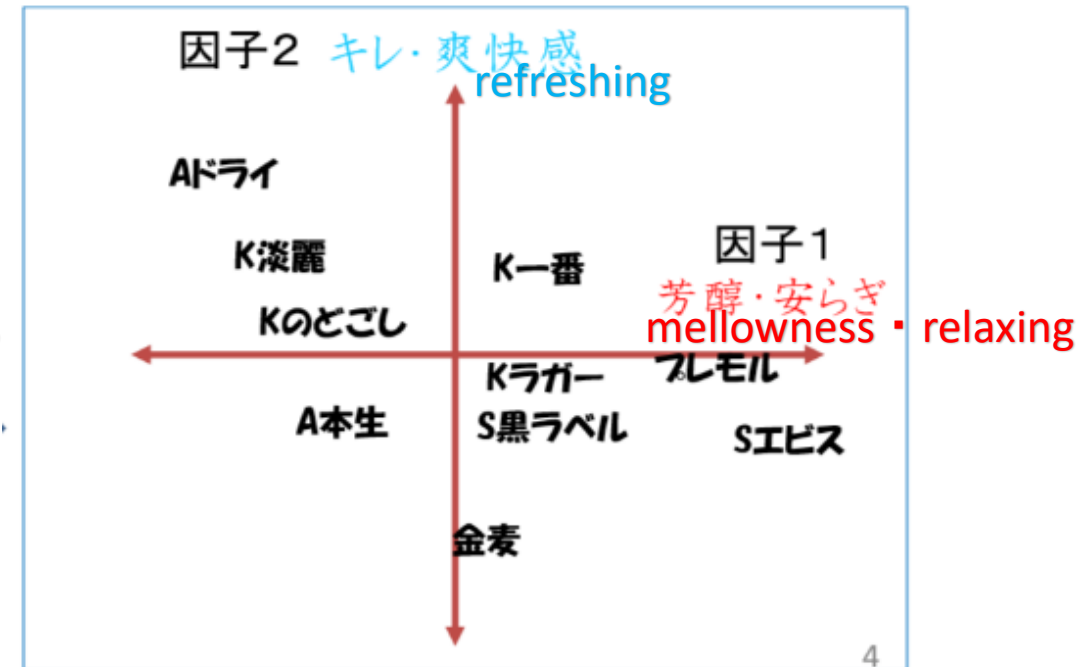


Image from:

<http://nakhirot.hatenablog.com/entry/20160724/1469337062>

# Motivation

---

一般的に、アンケートデータが使われる

しかし、アンケートデータには課題も多い・・・

- ・アンケート設計の難しさ
- ・外注のコストの高さ
- ・データを集めるのに時間がかかる                      etc...

レビューデータをアンケートデータの代わりに使えないだろうか...？

# Motivation

---

レビューデータは多くの利点を持っている

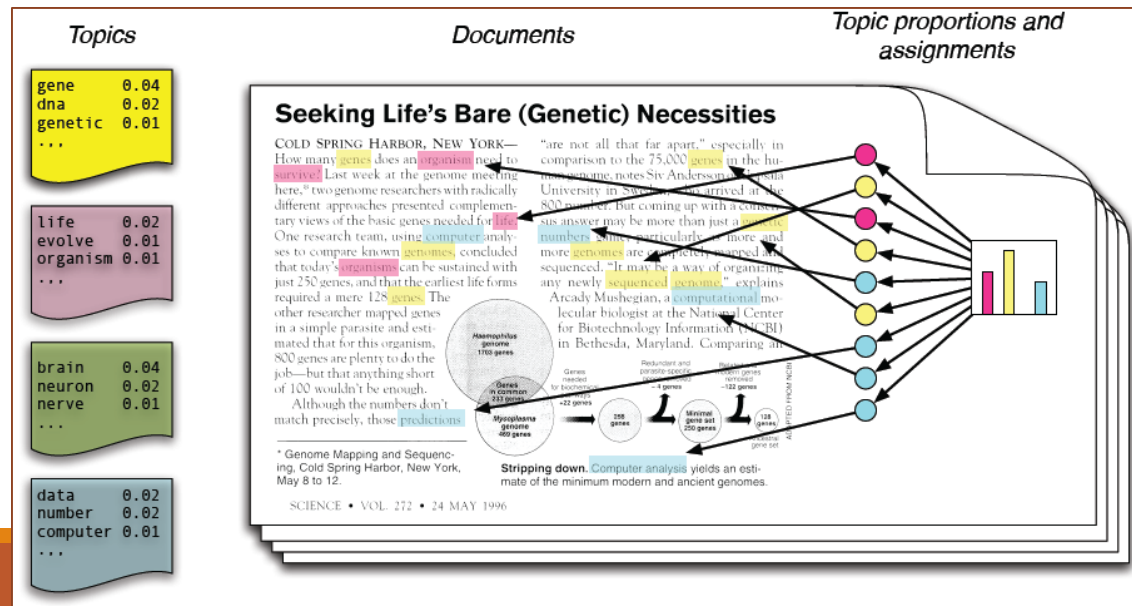
- ・膨大な量を、素早く集めることができる
- ・自動的に集積していく
- ・インターネットの普及により、投稿する人の偏りは小さくなりつつある

# Reviewing previous research

LDA (Latent Dirichlet Allocation )とは？

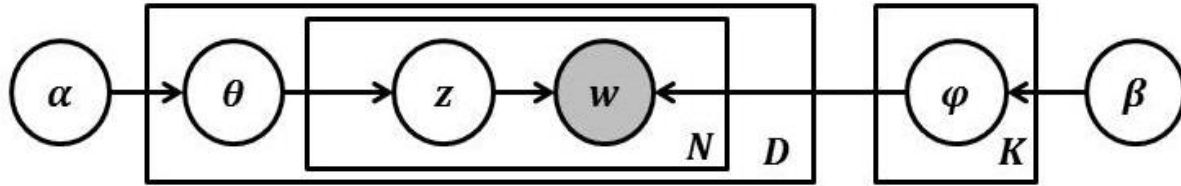
a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics.

—Latent Dirichlet Allocation (Blei & Ng & Jordan 2003)





# Latent Dirichlet Allocation



トピックモデル

$D$ : number of document

$N$ : number of word in the document

$W$ : word

$z$ : topic

$K$ : number of topic

$\alpha, \beta$ : hyper parameter

$\varphi$ : word distribution

$\theta$ : topic distribution

$$w_{d,i} \sim \text{Multi}(\Phi_{k,i})$$

$$z_{d,i} \sim \text{Multi}(\theta_{k,i})$$

$$\theta_d \sim \text{Dirichlet}(\alpha)$$

$$\Phi_k \sim \text{Dirichlet}(\beta)$$

# Reviewing previous research

---

## Why LDA?

- ー・高次元、かつスパース(疎)なデータに対応できる
  - ・文脈の情報なしで、ある単語の意味、極性(positive, negative)を解釈できる

high dimension and sparse data→

|         | review<br>1 | review<br>2 | review<br>3 | review<br>4 | review<br>5 | review<br>6 |
|---------|-------------|-------------|-------------|-------------|-------------|-------------|
| perfect | 0           | 0           | 0           | 0           | 0           | 0           |
| PM      | 0           | 0           | 0           | 0           | 0           | 0           |
| PM      | 0           | 0           | 0           | 0           | 0           | 0           |
| POLA    | 0           | 0           | 0           | 0           | 0           | 0           |
| Ponta   | 0           | 0           | 0           | 0           | 0           | 0           |
| PR      | 0           | 0           | 0           | 0           | 0           | 0           |

# Reviewing previous research

---

Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation

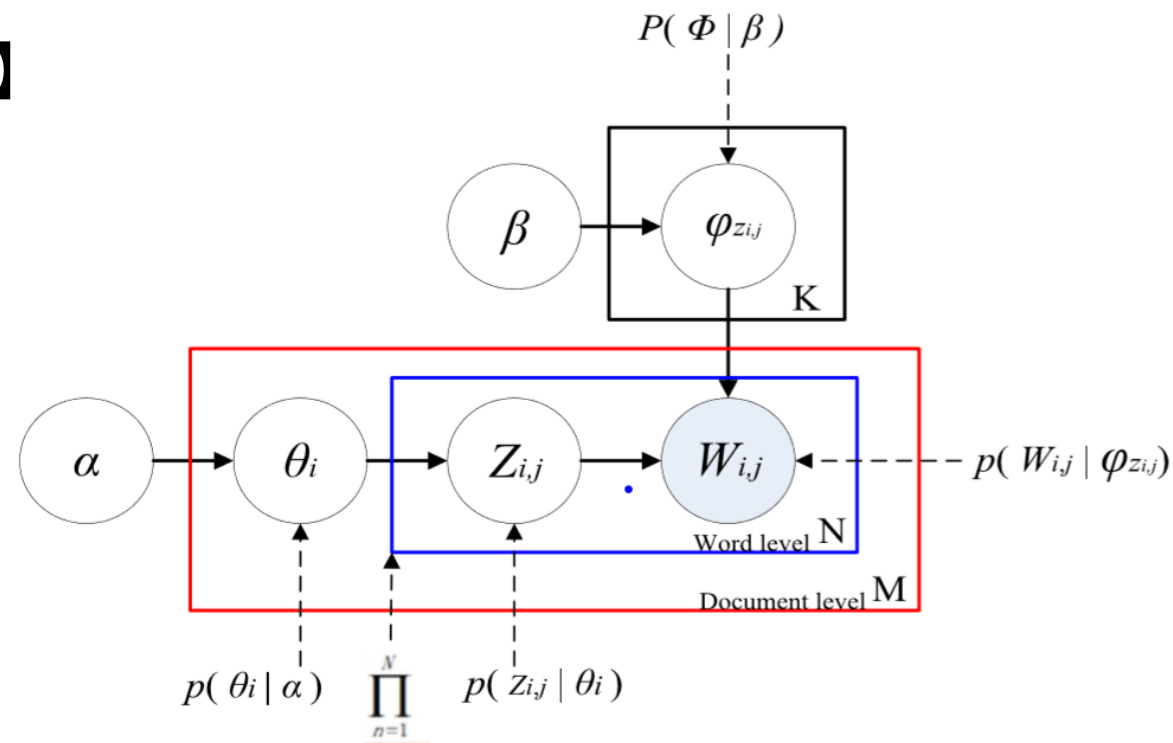
Yue Guo a, Stuart J. Barnes b, \*, Qiong Jia a (2017 Tourism Management)

- LDAと対応分析を利用してマッピングを行う
- TripAdvisorから、266,544件のホテルについてのレビューデータ

# Reviewing previous research

Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation

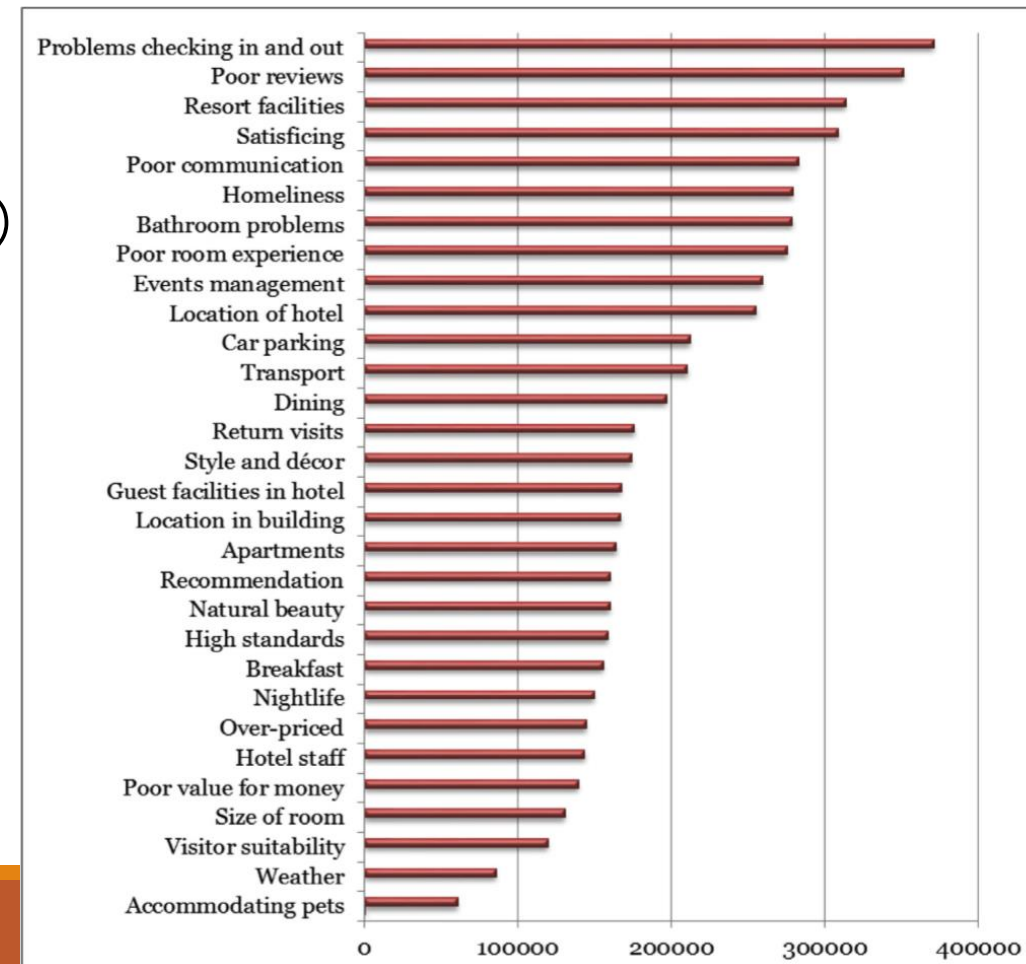
【graphical model】



# Reviewing previous research

## Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation

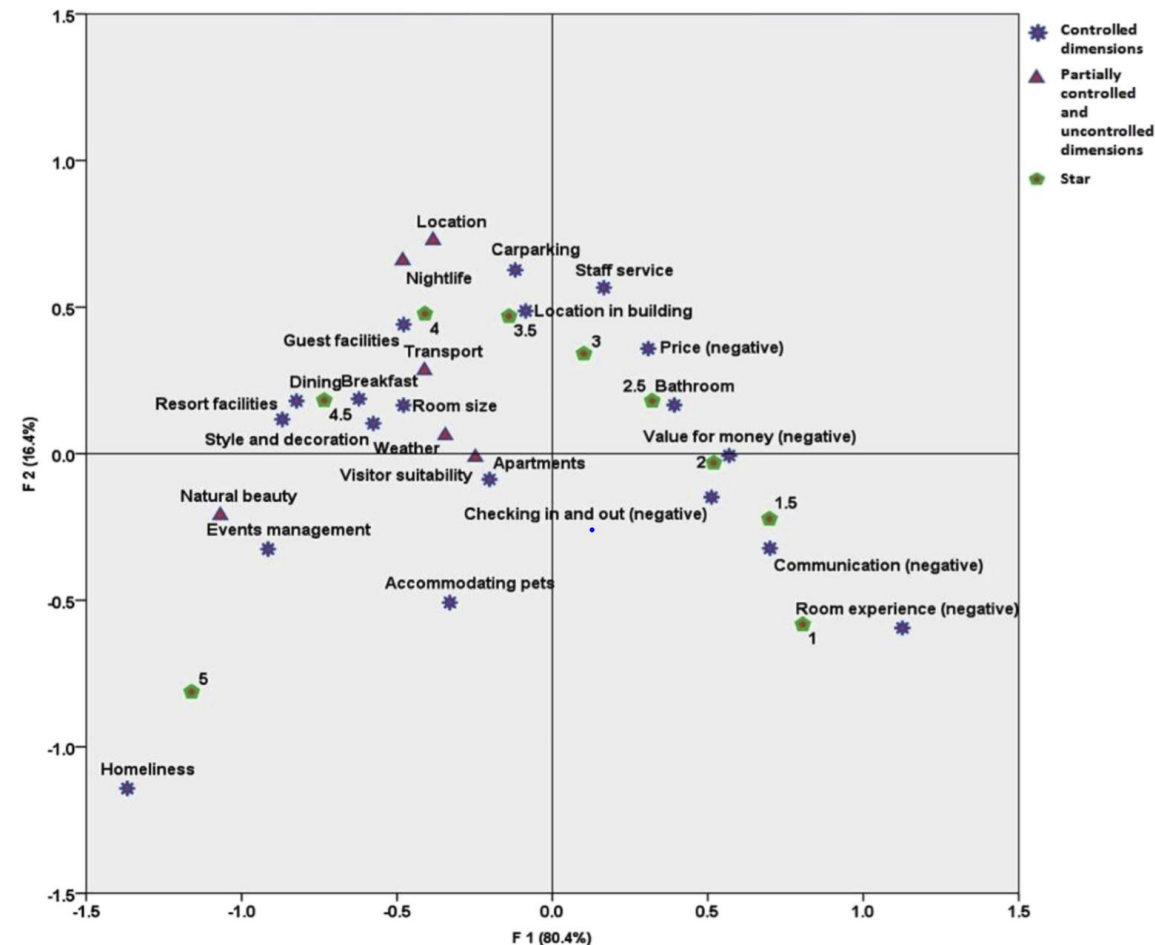
1. 単語分布φでトピックをラベリングする
2. ホテルを9段階に分ける(満足度? ランク?)
3. Generating a positioning matrix for hotel star ratings, counting the number of hotel star occurrences for each attribute.
4. 対応分析を行う



# Reviewing previous research

Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation

【Result】



# 実際にやってみた

---

## 分析1

- 2014/09～2015/11
  - トピック数7
    - テキスト＋スコア (allだけ)
      - 対応分析
      - 主成分分析

## 分析2

- 2013/07～2015/09
  - トピック数8
    - テキスト＋スコア (allだけ)
      - 対応分析
      - 主成分分析

# Mapping of hotel brand

【Mapping using text data and score】

- ①モデルを推定し、トピックを抽出する
- ②あるブランドaの、レビューdに関するトピック分布  $\theta_{d^a_k}$  が、0.1を超えれば、そのレビューはトピックKに言及していると判断する。そのレビューに付与されている「all score」を、ブランドaのトピックKについての得点に加算する
- ③以下のような表を作成し、主成分分析と対応分析を行う

|            | topic<br>1 | topic<br>2 | topic<br>3 | topic<br>4 | topic<br>5 | topic<br>6 | topic<br>7 | topic<br>8 |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| apa        | 4442       | 4342       | 2855       | 2508       | 4793       | 4425       | 3419       | 2974       |
| routein    | 3911       | 4466       | 3990       | 4532       | 4307       | 3514       | 3503       | 4307       |
| touyokoin  | 3307       | 4539       | 2834       | 3954       | 4941       | 3287       | 4338       | 1376       |
| superhotel | 3323       | 4341       | 6175       | 4732       | 4433       | 4411       | 3837       | 3518       |
| domin      | 3591       | 4161       | 4508       | 4269       | 4027       | 4044       | 3620       | 9434       |



| Column     | Sample              |
|------------|---------------------|
| 投稿番号       | 99887766            |
| 投稿日時       | 2012/01/01 12:34:56 |
| ニックネーム     | user12345           |
| 目的         | レジャー                |
| 同伴者        | 家族                  |
| 評価 1（立地）   | 4                   |
| 評価 2（部屋）   | 4                   |
| 評価 3（食事）   | 0                   |
| 評価 4（風呂）   | 4                   |
| 評価 5（サービス） | 3                   |
| 評価 6（設備）   | 3                   |
| 評価 7（総合）   | 4                   |

↑score and user data

Review text, category→

| Column  | Sample  |
|---------|---|
| 施設ID    | 121212  |
| 投稿日時    | 2012/01/01 12:34:56   |
| ユーザ投稿本文 | 値段のわりには、きれいなホテルでした。以前泊まったときより、きれいに改装されたようで、入ったときにはびっくりしました。しかし、以前と同じくフロントでは丁寧な対応をして頂き、気分よく泊まることができました。接客には十分満足しましたが、駐車場が少し遠かったり、コンビニが近くになかったりしたところが少し残念です。朝食はバイキングでしたが、品数も多く、おいしかったです。レストランスタッフの方を呼んでもなかなか出てこられなかったのは残念でした。 |
| 投稿番号    | 99887766  |
| 分類      | 感情・情報   |
| プランID   | 242424  |
| プランタイトル | 新館、禁煙、朝食バイキング！平日お得プラン！  |
| 部屋種類    | 2   |
| 部屋名前    | 新館・禁煙・セミダブルルーム  |
| 施設回答本文  | この度は、ご利用頂きまして誠にありがとうございます。また、貴重なご意見をお寄せ頂きありがとうございました。今年初めに改装しましたが、お褒め頂きましてありがとうございます。レスト  |

# 分析1

---

## Setting

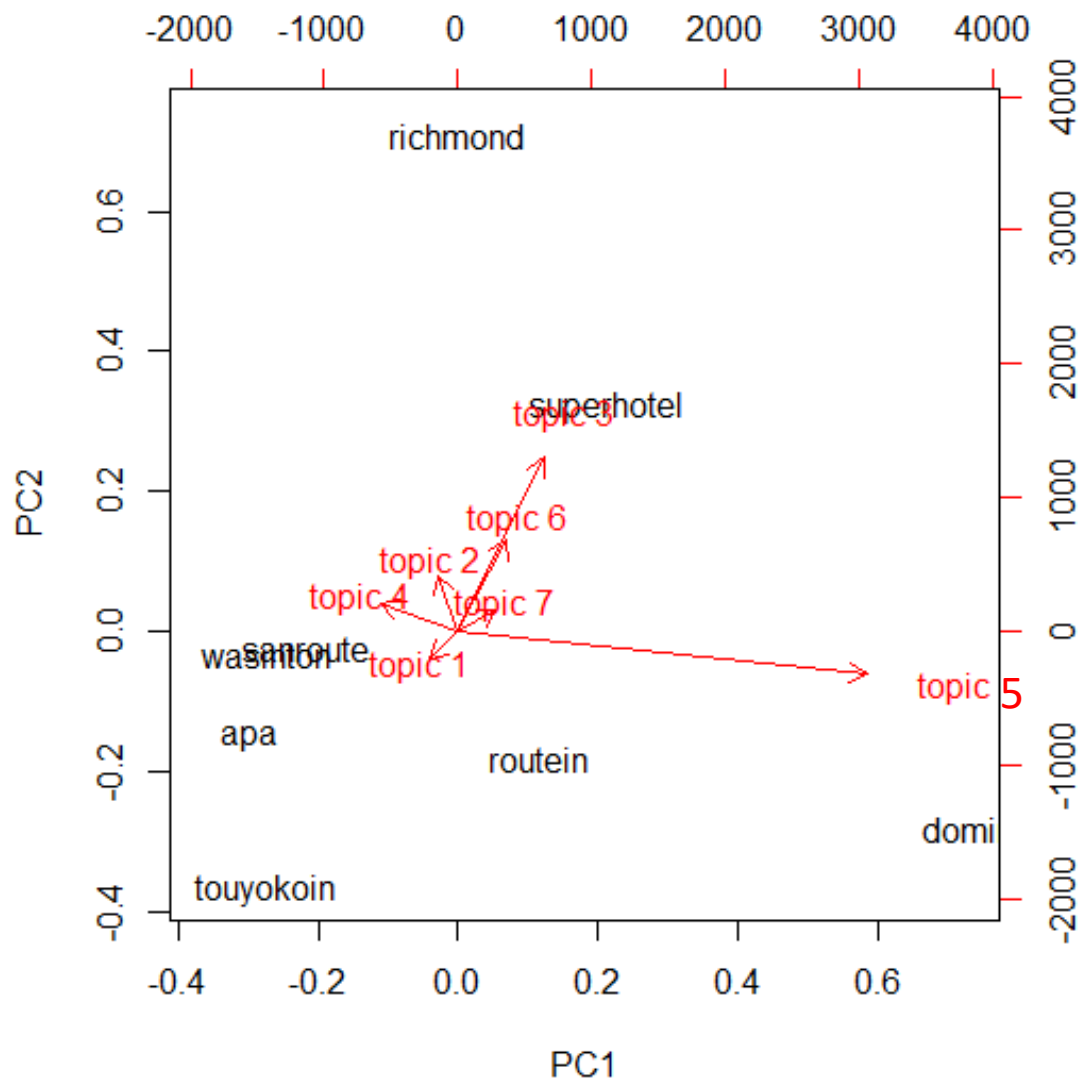
1. トピック数は7
2. Hyperparameter  $\alpha=0.5$   $\beta=0.1$
3. Using R package, “lda”, “ca”

# 分析1

## 【主成分分析】

topic1   topic2   topic3   topic4   topic5   topic6   topic7

|    | ▲    | ▲          | ▲     | ▲    | ▲     | ▲     | ▲      |
|----|------|------------|-------|------|-------|-------|--------|
| 1  | 音    | 立地         | ベッド   | 近い   | 美味しい  | アメニティ | スタッフ   |
| 2  | トイレ  | 料金         | スタッフ  | 駅    | 温泉    | ベッド   | 客      |
| 3  | 風呂   | 安い         | ツイン   | コンビニ | バイキング | 女性    | 女性     |
| 4  | バス   | 価格         | 風呂    | 立地   | 種類    | タオル   | 悪い     |
| 5  | 臭い   | 高い         | 広い    | 徒歩   | おいしい  | ルーム   | アウト    |
| 6  | 壁    | 値段         | 子供    | 遠い   | そば    | テレビ   | 団体     |
| 7  | エアコン | ビジネス       | ダブル   | バス   | メニュー  | カード   | 多い     |
| 8  | 古い   | プラン        | シングル  | 場所   | ご飯    | バス    | お客     |
| 9  | 窓    | 得          | 狭い    | 距離   | 風呂    | 大きい   | 外国     |
| 10 | タバコ  | 条件         | 家族    | 周辺   | パン    | 歯ブラシ  | 方々     |
| 11 | 浴槽   | コストパフォーマンス | 機会    | 繁華   | 無料    | キー    | 笑顔     |
| 12 | 空気   | 車          | ベット   | 駅前   | ごはん   | コンセント | 男性     |
| 13 | 温度   | コイン        | アメニティ | 空港   | おかず   | コーヒー  | 中国人    |
| 14 | 寒い   | 古い         | 嬉しい   | 地下鉄  | 夜鳴き   | 風呂    | エレベーター |
| 15 | カーテン | 無料         | 笑顔    | 圏内   | 味     | 無料    | お客様    |



# 分析1

---

## 【主成分分析】

### 解釈

- ・x軸が「食事、温泉」 y軸が「アメニティ、部屋」
- ・立地に関連するグループが見られる、しかしx軸でいうと負？
- ・ドーミーインは「食事、温泉」について飛びぬけていて、リッチモンドホテルは「部屋、アメニティ」に飛びぬけている。これはホテルの特徴とある程度合致している
- ・無料の食事を提供しているホテルが目立つ

# 分析 1

## 【主成分分析】

↓寄与率、累積寄与率

Importance of components:

|                        | PC1       | PC2      |
|------------------------|-----------|----------|
| Standard deviation     | 1424.6013 | 706.4988 |
| Proportion of Variance | 0.7159    | 0.1761   |
| Cumulative Proportion  | 0.7159    | 0.8919   |

↓主成分負荷量

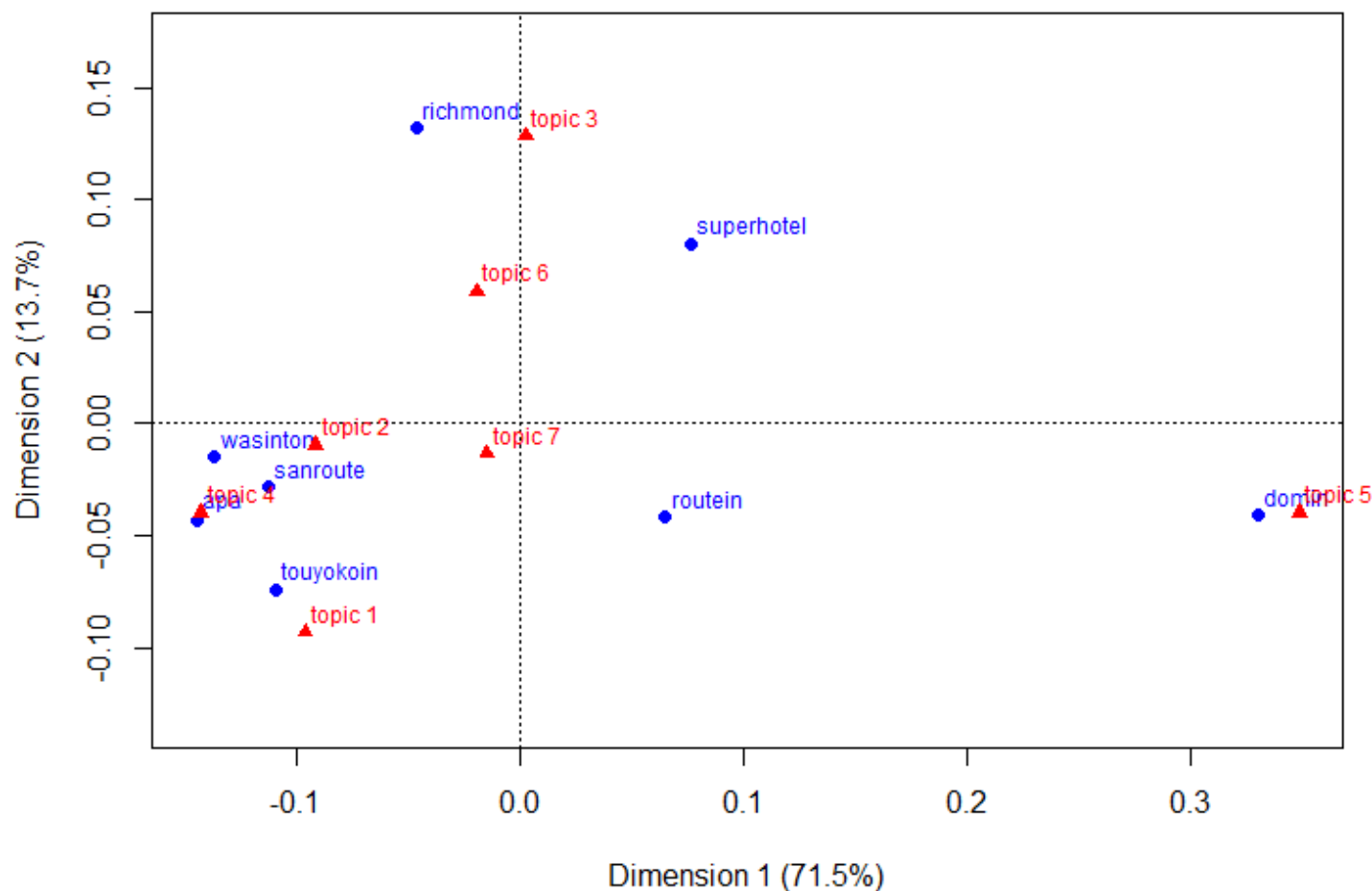
|     | topic 1       | topic 2     | topic 3      | topic 4     | topic 5      | topic 6    | topic 7     |
|-----|---------------|-------------|--------------|-------------|--------------|------------|-------------|
| PC1 | -2.986828e-01 | -0.21438914 | 0.4375749971 | -0.71299439 | 0.993613261  | 0.41236929 | 0.460871795 |
| PC2 | -2.728484e-01 | 0.57929265  | 0.8900752045 | 0.25966023  | -0.104535394 | 0.76856292 | 0.275486051 |

# 分析1

## 【対応分析】

topic1   topic2   topic3   topic4   topic5   topic6   topic7

|    | ▲    | ▲          | ▲     | ▲    | ▲     | ▲     | ▲      |
|----|------|------------|-------|------|-------|-------|--------|
| 1  | 音    | 立地         | ベッド   | 近い   | 美味しい  | アメニティ | スタッフ   |
| 2  | トイレ  | 料金         | スタッフ  | 駅    | 温泉    | ベッド   | 客      |
| 3  | 風呂   | 安い         | ツイン   | コンビニ | バイキング | 女性    | 女性     |
| 4  | バス   | 価格         | 風呂    | 立地   | 種類    | タオル   | 悪い     |
| 5  | 臭い   | 高い         | 広い    | 徒歩   | おいしい  | ルーム   | アウト    |
| 6  | 壁    | 値段         | 子供    | 遠い   | そば    | テレビ   | 団体     |
| 7  | エアコン | ビジネス       | ダブル   | バス   | メニュー  | カード   | 多い     |
| 8  | 古い   | プラン        | シングル  | 場所   | ご飯    | バス    | お客     |
| 9  | 窓    | 得          | 狭い    | 距離   | 風呂    | 大きい   | 外国     |
| 10 | タバコ  | 条件         | 家族    | 周辺   | パン    | 歯ブラシ  | 方々     |
| 11 | 浴槽   | コストパフォーマンス | 機会    | 繁華   | 無料    | キー    | 笑顔     |
| 12 | 空気   | 車          | ベット   | 駅前   | ごはん   | コンセント | 男性     |
| 13 | 温度   | コイン        | アメニティ | 空港   | おかず   | コーヒー  | 中国人    |
| 14 | 寒い   | 古い         | 嬉しい   | 地下鉄  | 夜鳴き   | 風呂    | エレベーター |
| 15 | カーテン | 無料         | 笑顔    | 圏内   | 味     | 無料    | お客様    |

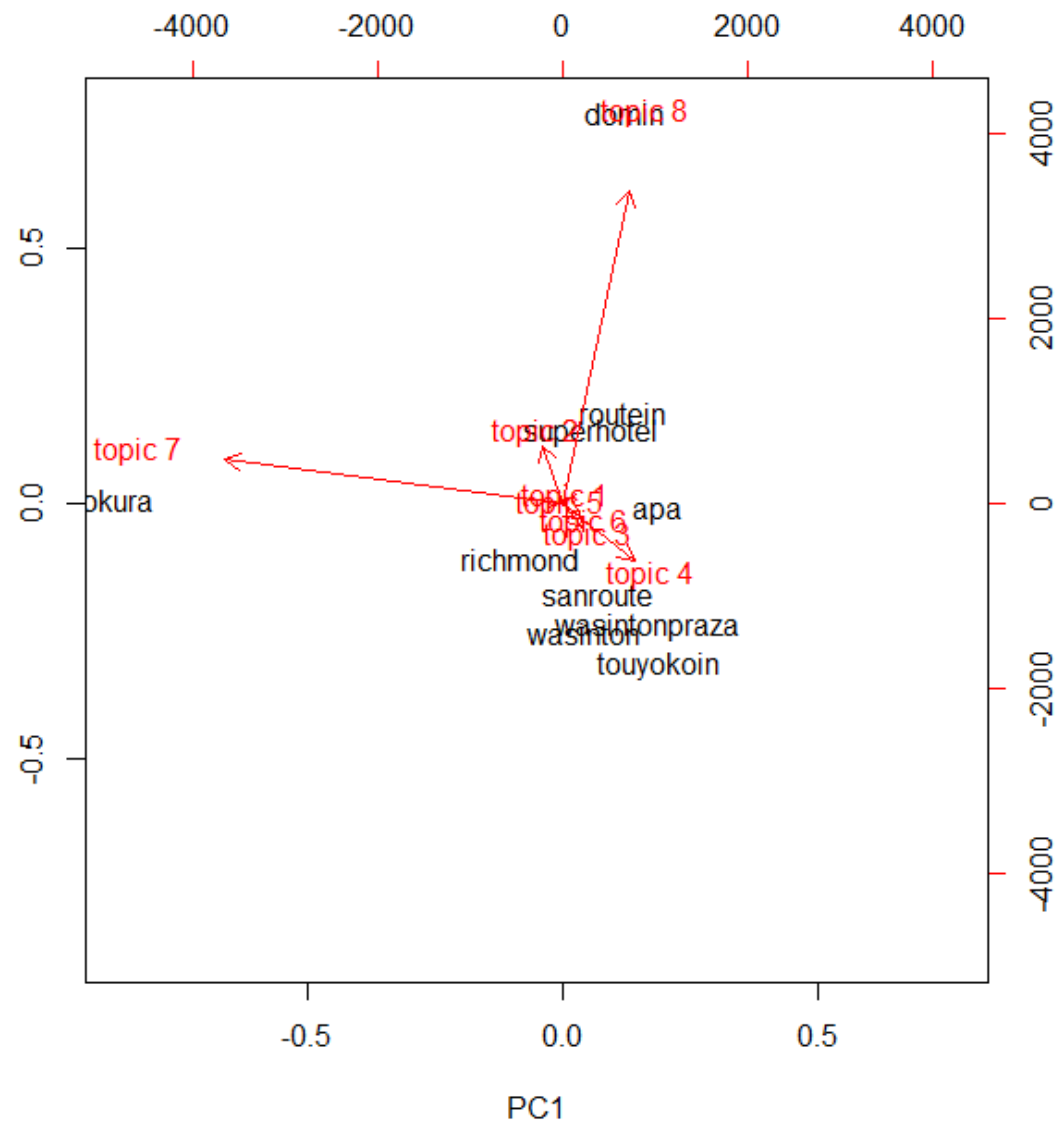


# 分析2 トピック数8

## 【主成分分析】

topic1 topic2 topic3 topic4 topic5 topic6 topic7 topic8

|    | topic1 | topic2 | topic3     | topic4 | topic5 | topic6 | topic7  | topic8 |
|----|--------|--------|------------|--------|--------|--------|---------|--------|
| 1  | 客      | 美味しい   | 安い         | 駅      | ベッド    | 音      | 子供      | 温泉     |
| 2  | スタッフ   | 種類     | 立地         | 近い     | 広い     | トイレ    | スタッフ    | 浴場     |
| 3  | 女性     | バイキング  | 価格         | コンビニ   | 狭い     | 臭い     | 家族      | 風呂     |
| 4  | 悪い     | パン     | 高い         | 遠い     | シングル   | エアコン   | 美味しい    | そば     |
| 5  | カード    | 無料     | 値段         | 徒歩     | ルーム    | ルーム    | 夜景      | 疲れ     |
| 6  | アウト    | おいしい   | 料金         | 立地     | ツイン    | 悪い     | 笑顔      | 最高     |
| 7  | 多い     | コーヒー   | ビジネス       | バス     | バス     | 壁      | グレード    | 夜鳴き    |
| 8  | キー     | 多い     | 古い         | 繁華     | アメニティ  | 古い     | 素晴らしい   | 気持ちよい  |
| 9  | エレベーター | 品数     | 条件         | 距離     | タオル    | 窓      | 娘       | 露天風呂   |
| 10 | 男性     | ご飯     | コストパフォーマンス | 周辺     | ダブル    | 隣      | ハウステンボス | 天然     |
| 11 | 荷物     | メニュー   | プラン        | 車      | ベット    | 空調     | ディズニー   | サウナ    |
| 12 | 外国     | ごはん    | 無い         | 場所     | 枕      | バス     | 神戸      | おいしい   |
| 13 | 無い     | 和食     | 最高         | 東京     | 大きい    | タバコ    | おいしい    | 無料     |
| 14 | 団体     | 味      | 内容         | 空港     | 歯ブラシ   | 廊下     | 皆さん     | 湯      |
| 15 | お客     | おかず    | 狭い         | 地下鉄    | 風呂     | 寒い     | 楽しい     | 美味しい   |



# 分析2

---

## 【主成分分析】

### 解釈

- ・x軸が「？」 y軸が「食事、温泉」
- ・リッチモンドはビジネスホテル中でも価格帯が高いホテル、x軸は価格に係している？
- ・立地を売りにするホテルは、分析1と同様に密集している



# 分析2

## 【主成分分析】

### ↓寄与率、累積寄与率

Importance of components:

|                        | PC1       | PC2       |
|------------------------|-----------|-----------|
| Standard deviation     | 1512.8159 | 1402.2628 |
| Proportion of Variance | 0.4846    | 0.4164    |
| Cumulative Proportion  | 0.4846    | 0.9010    |

### ↓主成分負荷量

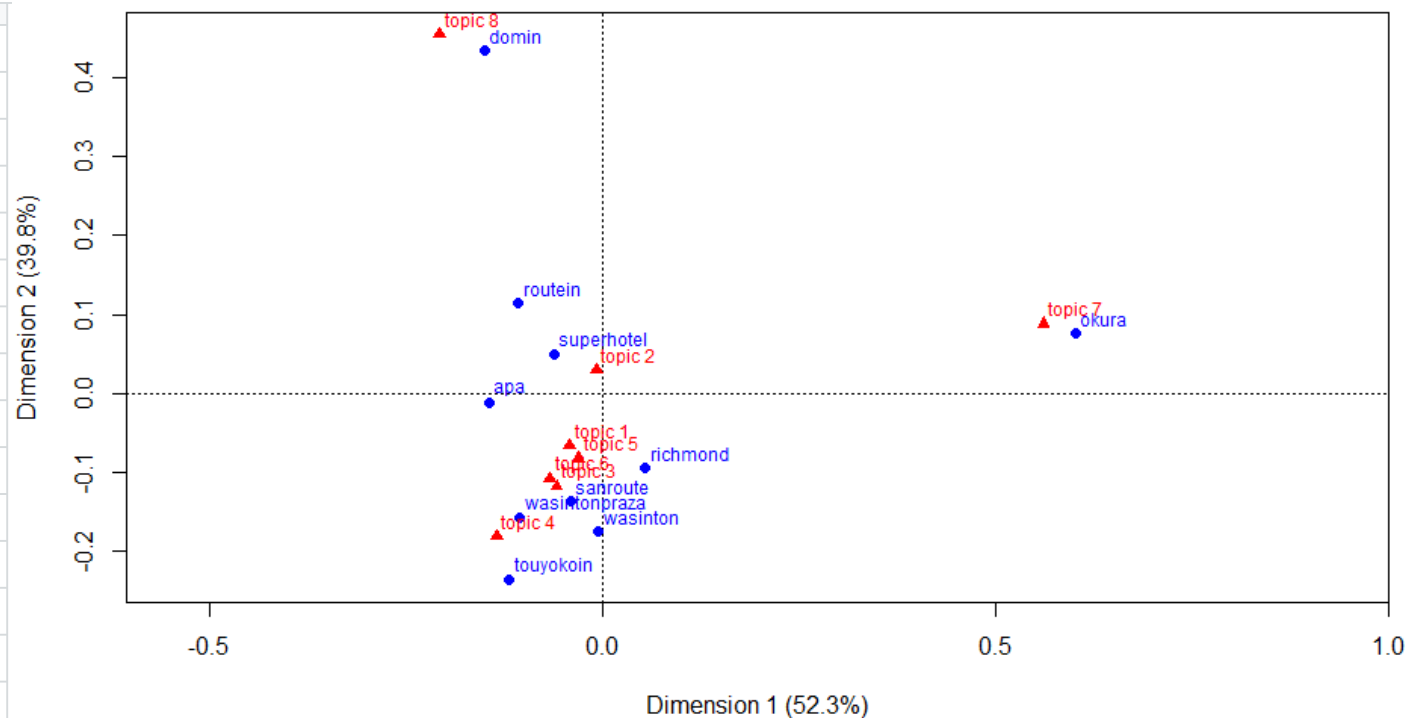
|     | topic 1     | topic 2       | topic 3     | topic 4      | topic 5       | topic 6     | topic 7      | topic 8      |
|-----|-------------|---------------|-------------|--------------|---------------|-------------|--------------|--------------|
| PC1 | 0.083829783 | -0.2013196951 | 0.31418206  | 0.669203798  | -0.0157885187 | 0.42175400  | -0.990809411 | 0.2070507306 |
| PC2 | 0.138868688 | 0.5835931128  | -0.38278963 | -0.524475099 | 0.0001144248  | -0.34650434 | 0.128385024  | 0.9771852950 |

# 分析2

## 【対応分析】

topic1 topic2 topic3 topic4 topic5 topic6 topic7 topic8

|    | ▲      | ▲     | ▲          | ▲    | ▲     | ▲    | ▲       | ▲     |
|----|--------|-------|------------|------|-------|------|---------|-------|
| 1  | 客      | 美味しい  | 安い         | 駅    | ベッド   | 音    | 子供      | 温泉    |
| 2  | スタッフ   | 種類    | 立地         | 近い   | 広い    | トイレ  | スタッフ    | 浴場    |
| 3  | 女性     | バイキング | 価格         | コンビニ | 狭い    | 臭い   | 家族      | 風呂    |
| 4  | 悪い     | パン    | 高い         | 遠い   | シングル  | エアコン | 美味しい    | そば    |
| 5  | カード    | 無料    | 値段         | 徒歩   | ルーム   | ルーム  | 夜景      | 疲れ    |
| 6  | アウト    | おいしい  | 料金         | 立地   | ツイン   | 悪い   | 笑顔      | 最高    |
| 7  | 多い     | コーヒー  | ビジネス       | バス   | バス    | 壁    | グレード    | 夜鳴き   |
| 8  | キー     | 多い    | 古い         | 繁華   | アメニティ | 古い   | 素晴らしい   | 気持ちよい |
| 9  | エレベーター | 品数    | 条件         | 距離   | タオル   | 窓    | 娘       | 露天風呂  |
| 10 | 男性     | ご飯    | コストパフォーマンス | 周辺   | ダブル   | 隣    | ハウステンボス | 天然    |
| 11 | 荷物     | メニュー  | プラン        | 車    | ベット   | 空調   | ディズニー   | サウナ   |
| 12 | 外国     | ごはん   | 無い         | 場所   | 枕     | バス   | 神戸      | おいしい  |
| 13 | 無い     | 和食    | 最高         | 東京   | 大きい   | タバコ  | おいしい    | 無料    |
| 14 | 団体     | 味     | 内容         | 空港   | 歯ブラシ  | 廊下   | 皆さん     | 湯     |
| 15 | お客     | おかず   | 狭い         | 地下鉄  | 風呂    | 寒い   | 楽しい     | 美味しい  |



# 分析 + only score

【Using only score】

レビューには「立地」、「食事」、「部屋」、「サービス」、「アメニティ」、「風呂」、「合計」について、それぞれ5段階で評価が付与されている

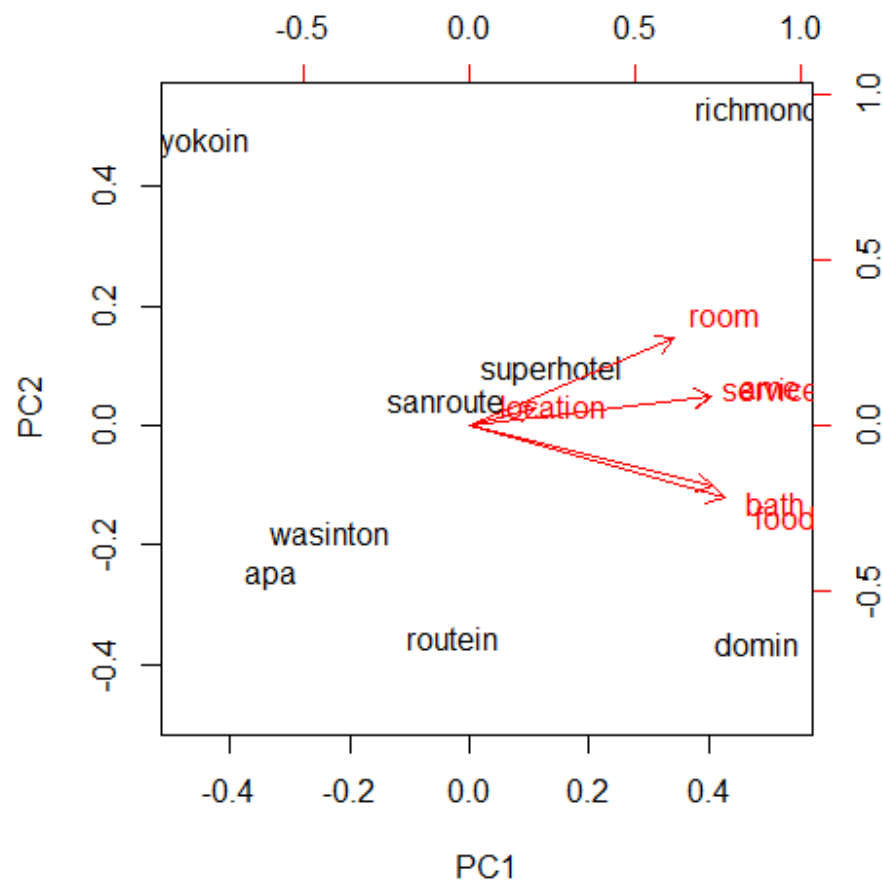
以下の表を作成し、対応分析を行う。パッケージは"ca"を用いる。

↓平均得点 表

|            | location | room     | food     | bath     | service  | ame      |
|------------|----------|----------|----------|----------|----------|----------|
| apa        | 4.089888 | 3.575441 | 3.678973 | 3.549759 | 3.642055 | 3.568218 |
| routein    | 4.018532 | 3.867733 | 3.896439 | 3.919331 | 3.909157 | 3.708576 |
| touyokoin  | 4.085840 | 3.825110 | 3.314079 | 3.436422 | 3.532290 | 3.429603 |
| superhotel | 4.226366 | 4.072009 | 3.991507 | 3.707903 | 4.158789 | 3.931315 |
| domin      | 4.335494 | 4.271070 | 4.371961 | 4.436386 | 4.250810 | 4.179092 |

# 分析 + only score

【主成分分析】



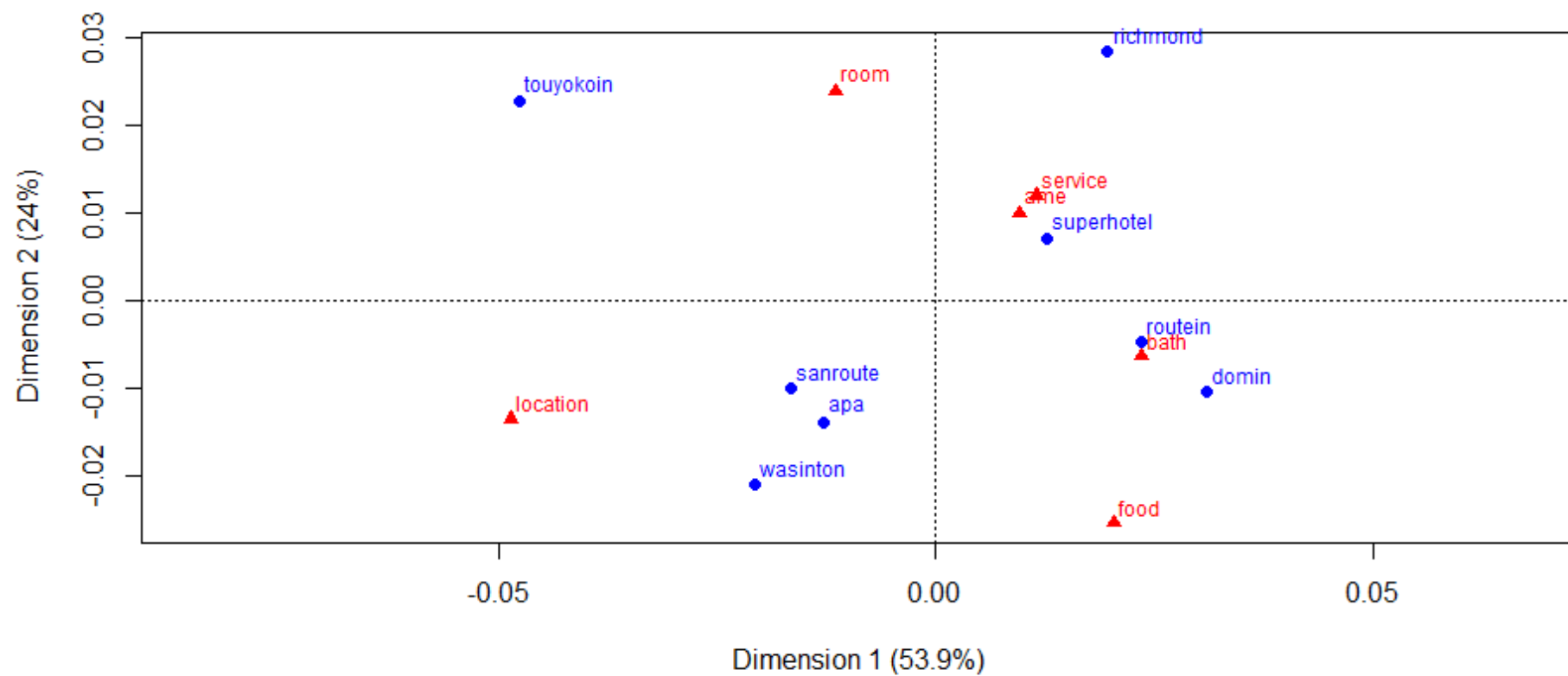
↓ 寄与率、累積寄与率

Importance of components:

|                        | PC1    | PC2    |
|------------------------|--------|--------|
| Standard deviation     | 0.7132 | 0.1815 |
| Proportion of Variance | 0.8756 | 0.0567 |
| cumulative Proportion  | 0.8756 | 0.9323 |

# 分析 + only score

【対応分析】



# Challenges

---

- ・そのトピックの重要性はわからない
- ・日本語のステミングができていない
- ・トピック数、パラメータを主観で決めている
- ・スコアだけの結果との比較、その解釈
- ・展開方式、その他の調査
- ・この結果とマーケティング論を裏付けとした提案

# Reference

---

- Deloitte Tohmatsu website:

<https://www2.deloitte.com/jp/ja/pages/deloitte-analytics/articles/analytics-plsa.html#>

- All about旅行

<https://allabout.co.jp/gm/gc/434435/>

- Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation

Yue Guo a, Stuart J. Barnes b, \*, Qiong Jia a (2017 Tourism Management)

- 『トピックモデルによる統計的潜在意味解析』 佐藤一誠