国籍別・都道府県別宿泊者数の潜在パターン：行列分解してみよう

藤原 義久

兵庫県立大学大学院情報科学研究科

yoshi@gsis.u-hyogo.ac.jp

データ

日本の魅力を、
日本のチカラに。
JNTO
日本政府観光局

日本政府観光局(JNTO)

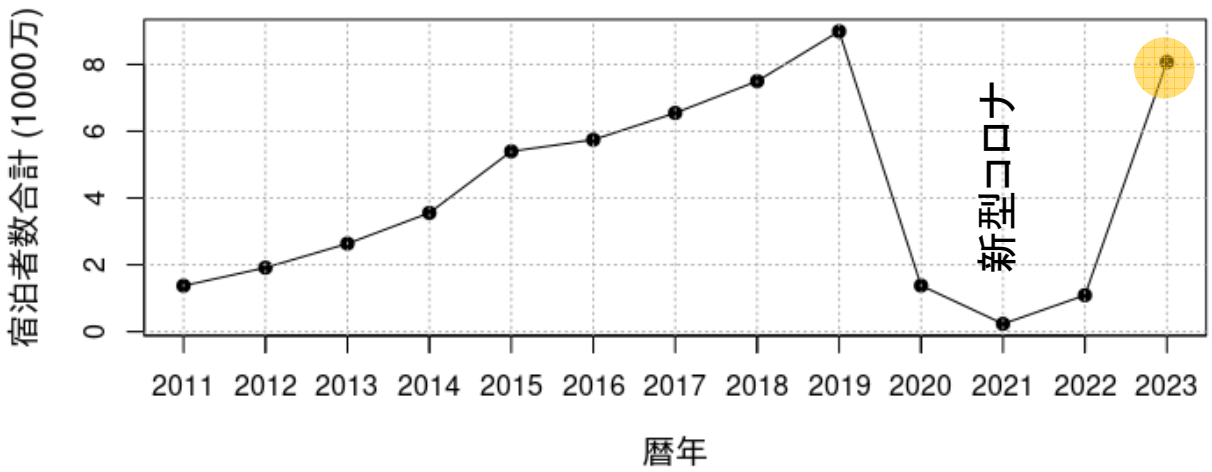
https://statistics.jnto.go.jp/

- **国籍別 都道府県別延べ宿泊者数**
  - ✓ 延べ宿泊者数は、調査対象となっている全宿泊施設からの回答に基づき集計
  - ✓ 観光庁「宿泊旅行統計調査（2011年～2024年）」より、日本政府観光局（JNTO）が作成

項目	データ数
国籍	20
都道府県	47
暦年	2011～2023

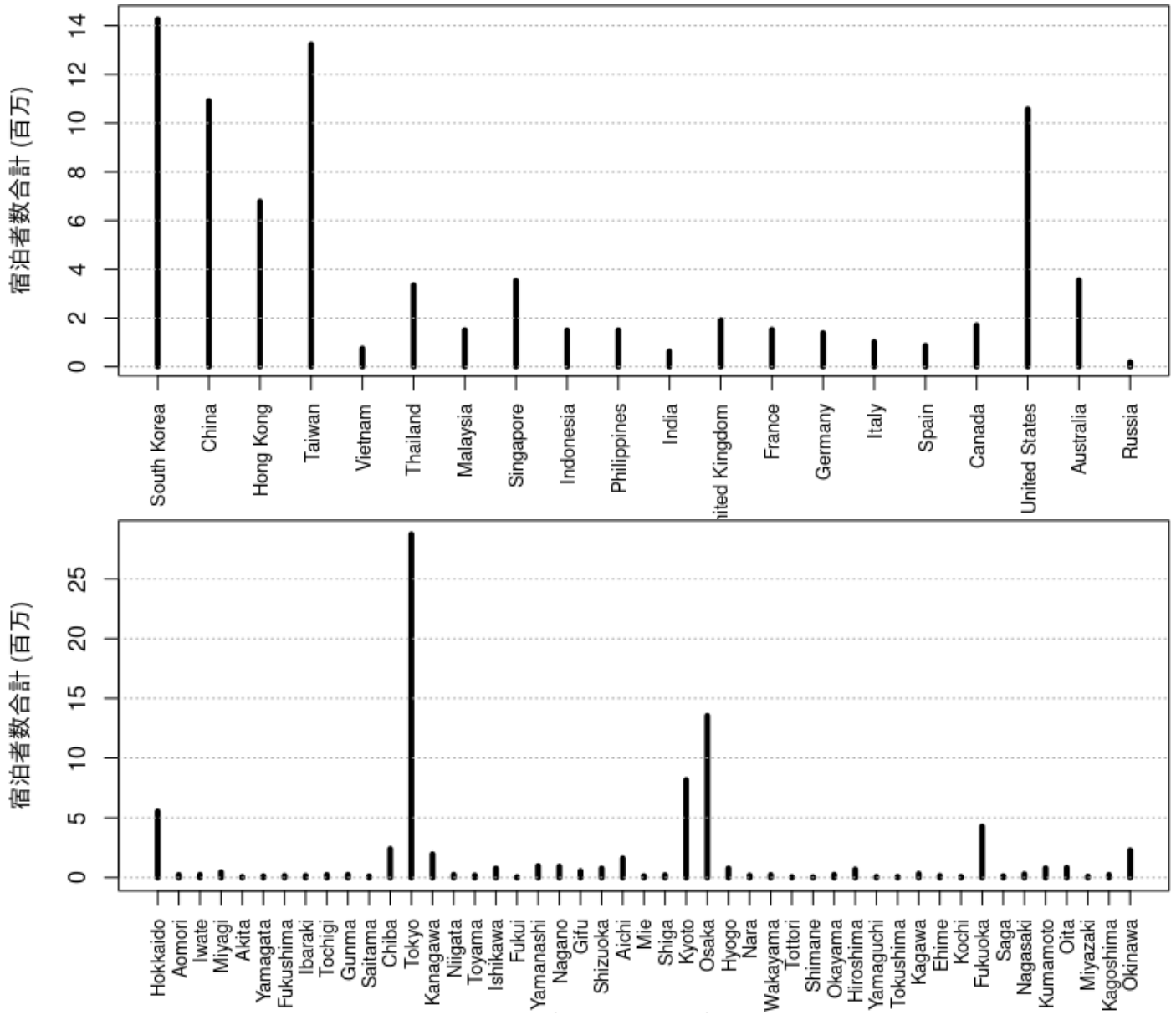
国籍別・都道府県別，暦年別の宿泊者数

```
> df
# A tibble: 11,562 × 4
   country   pref    year nstays
   <chr>     <chr>  <int>  <dbl>
 1 Australia Aichi   2011   5670
 2 Australia Aichi   2012   9600
 3 Australia Aichi   2013  10380
 4 Australia Aichi   2014  12710
 5 Australia Aichi   2015  14500
 6 Australia Aichi   2016  14190
 7 Australia Aichi   2017  22080
 8 Australia Aichi   2018  22350
 9 Australia Aichi   2019  28580
10 Australia Aichi   2020   5460
# ℹ 11,552 more rows
```



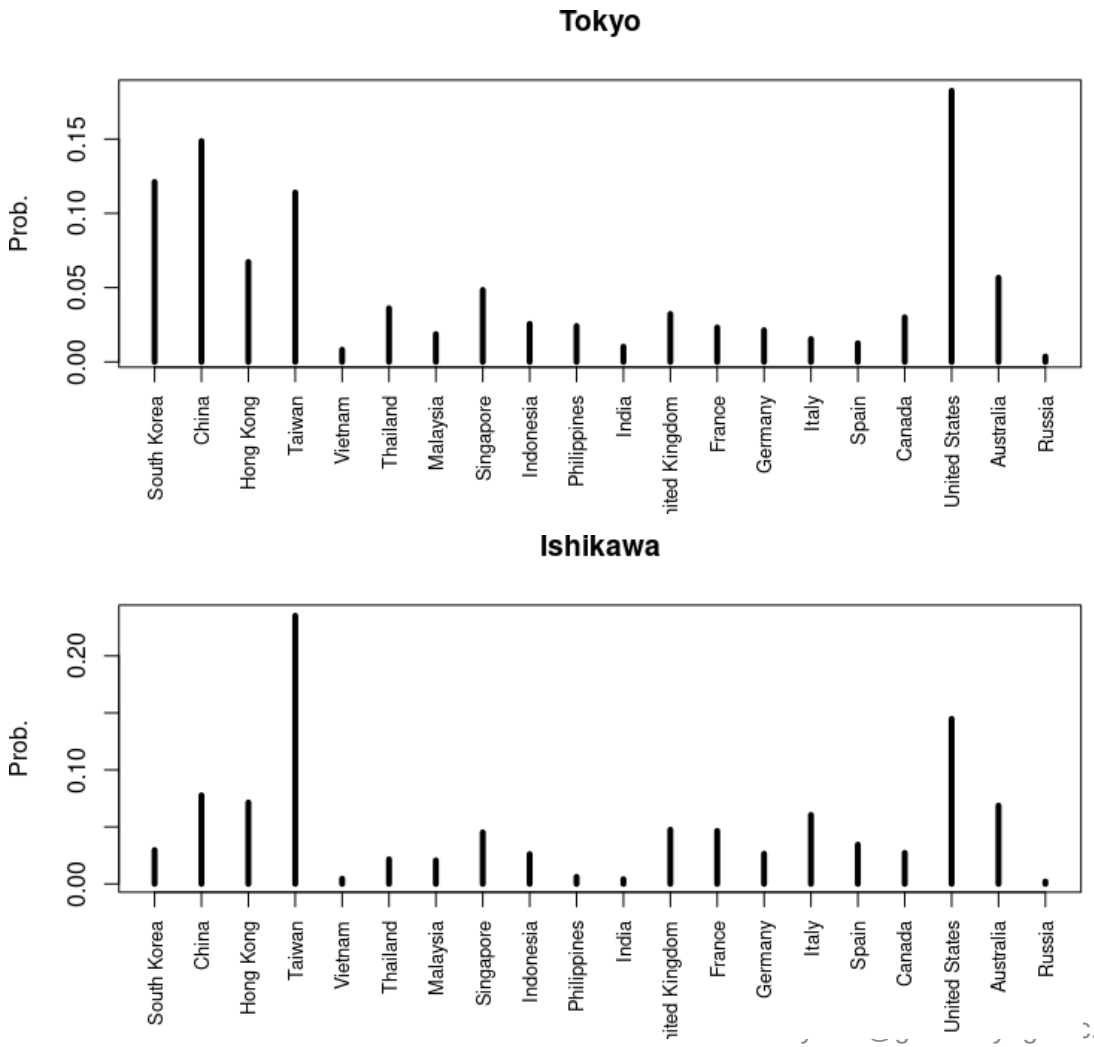
→ 以下では直近の2023年を対象にする

宿泊者数の絶対値

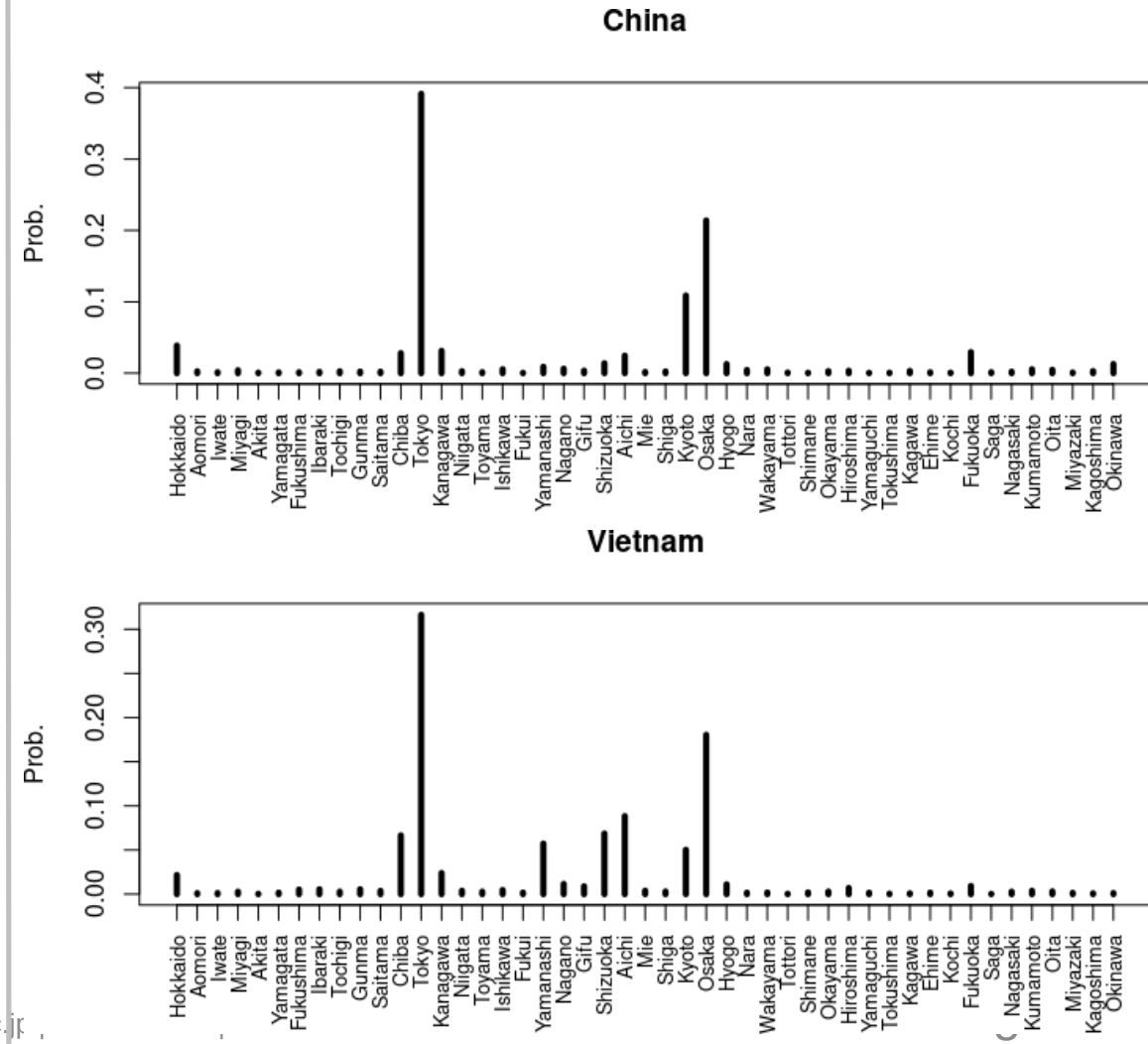

宿泊者数合計 (百万)
0
2
4
6
8
10
12
14
South Korea
China
Hong Kong
Taiwan
Vietnam
Thailand
Malaysia
Singapore
Indonesia
Philippines
India
France
Germany
Italy
Spain
Canada
United States
Australia
Russia
宿泊者数合計 (百万)
0
5
10
15
20
25
Hokkaido
Aomori
Iwate
Miyagi
Akita
Yamagata
Fukushima
Ibaraki
Tochigi
Gunma
Saitama
Chiba
Tokyo
Kanagawa
Niigata
Toyama
Ishikawa
Fukui
Yamanashi
Nagano
Gifu
Shizuoka
Aichi
Mie
Shiga
Kyoto
Osaka
Hyogo
Nara
Wakayama
Tottori
Shimane
Okayama
Hiroshima
Yamaguchi
Tokushima
Kagawa
Ehime
Kochi
Fukuoka
Saga
Nagasaki
Kumamoto
Oita
Miyazaki
Kagoshima
Okinawa

→ 絶対数から割合＝パターンへ変換

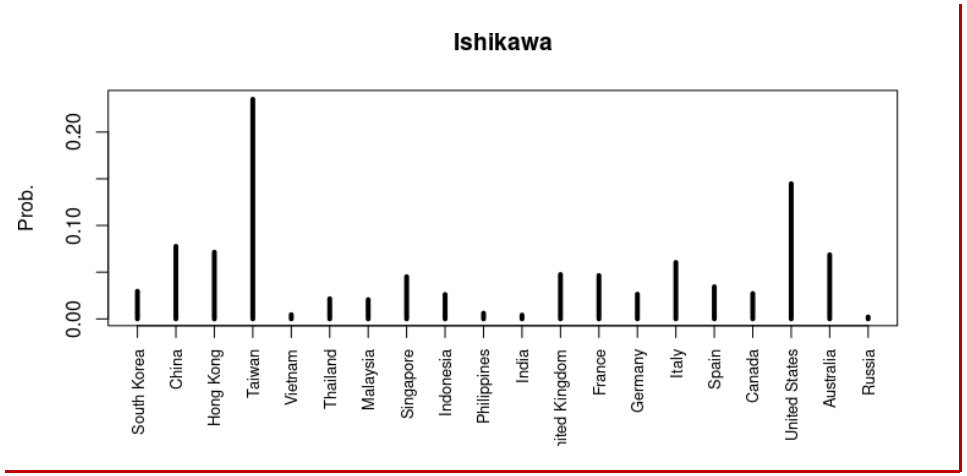
都道府県ごとの国籍パターン（例）



国籍ごとの都道府県パターン（例）

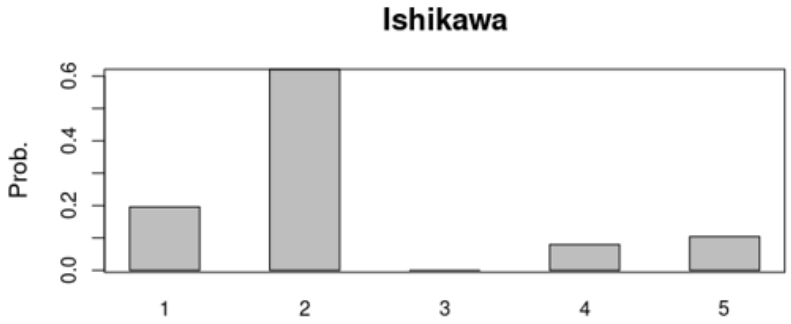


統計モデル～からくりを妄想



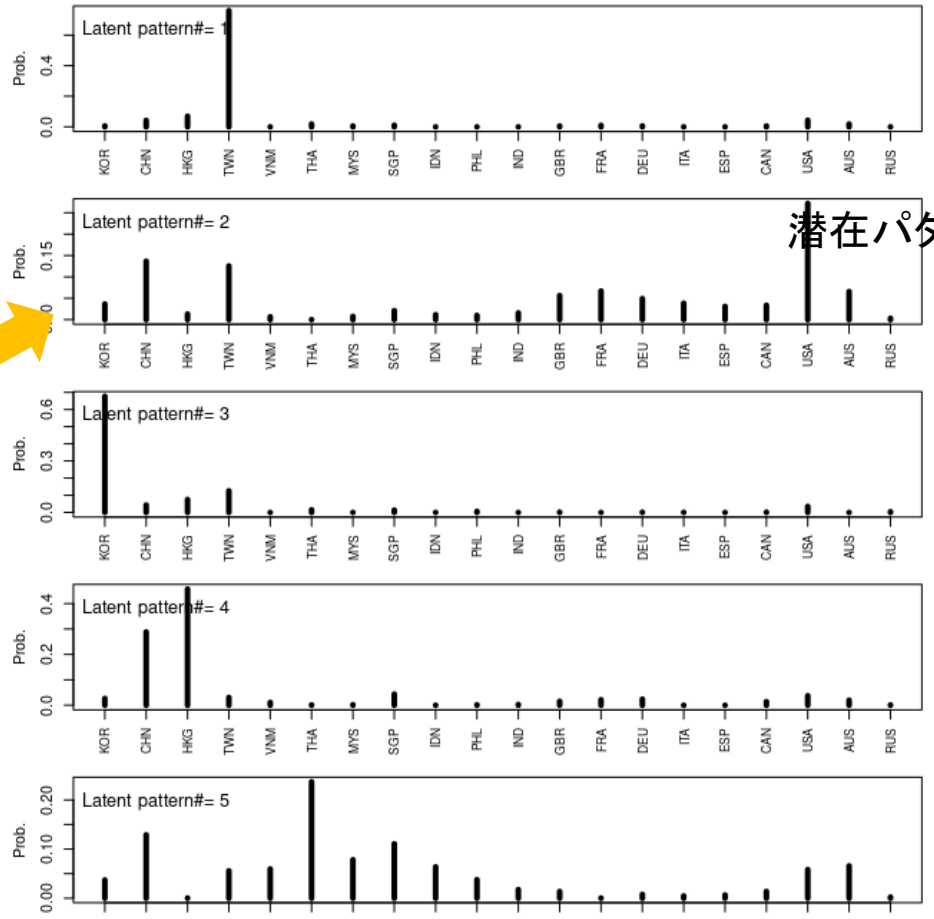
石川県の「サイコロ」 

「面2が出た」



サイコロ5面＝潜在パターン5つ

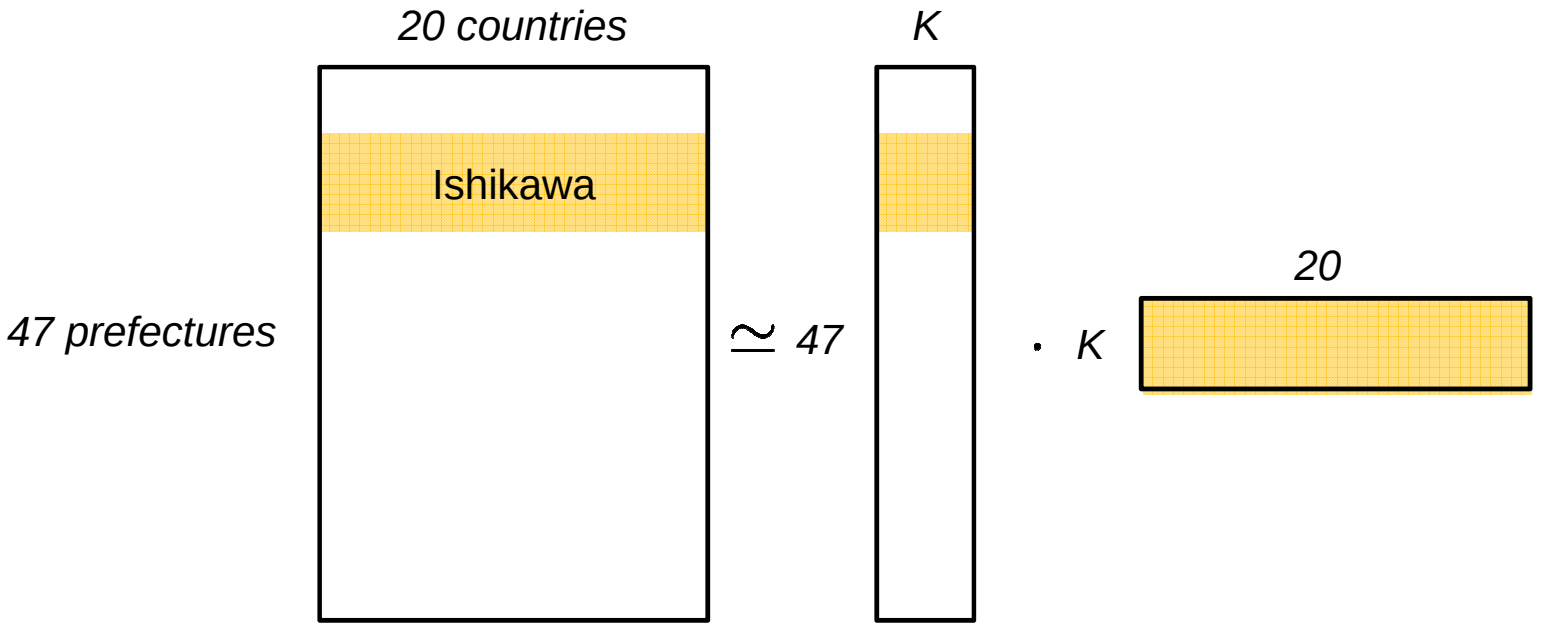
潜在パターンごとのカード20枚セット
セットごとに数字の出る確率は異なる



潜在パターン2のカード

「米国が出た」

【アイデア】
比較的少数の潜在パターンが隠れていて、データはその線形和になっているのでは？



$$X_{pc} \simeq \sum_{k=1}^{K} r_{pk} \cdot P_{kc}$$

$$\mathbf{X} \simeq \mathbf{R}\,\mathbf{P}$$

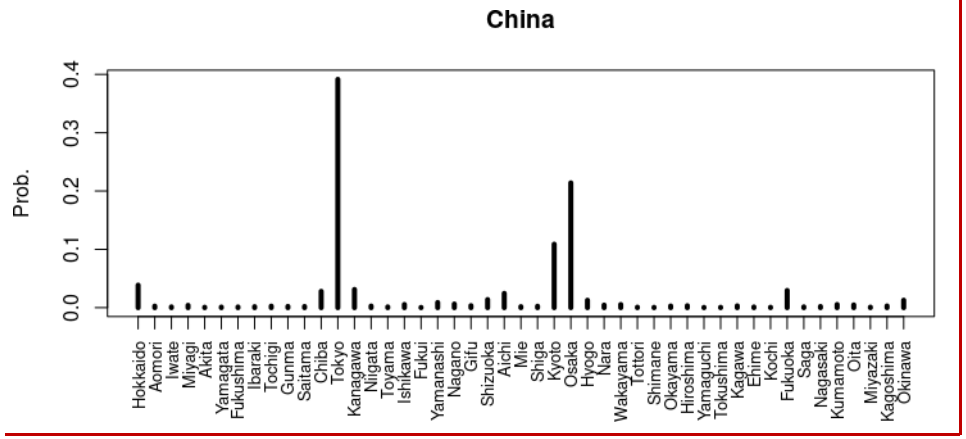
- ✓ 行列の要素は非負値
- ✓ 比較的疎な行列（ほぼゼロが多い）

非負値行列分解（non-negative matrix factorization）

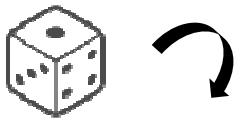
- ***R***, ***P*** とも未知、推定することに注意
- 確率モデルと見なせる
  トピックモデルやLDAなどと本質的にはほぼ同じ（ベイズ推定）

```
> library(NMF)

> dim(X)  # Data: prefectures by countries
[1] 47 20

> K <- 5
> set.seed(123)
> nmf(X, rank=K)
```

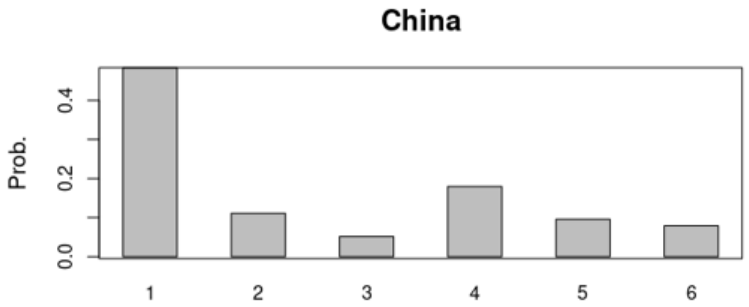
国籍別についても同様



China の「サイコロ」

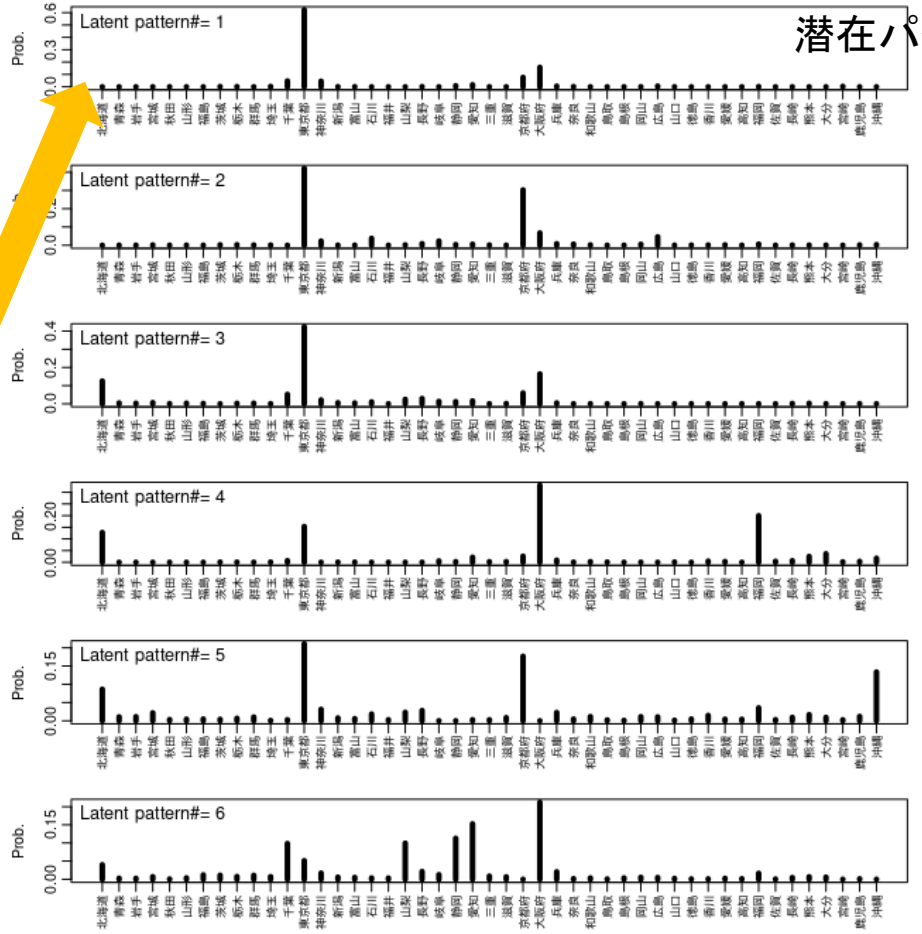


「面1が出た」



サイコロ6面＝潜在パターン6つ

潜在パターンごとのカード47枚セット
セットごとに数字の出る確率は異なる

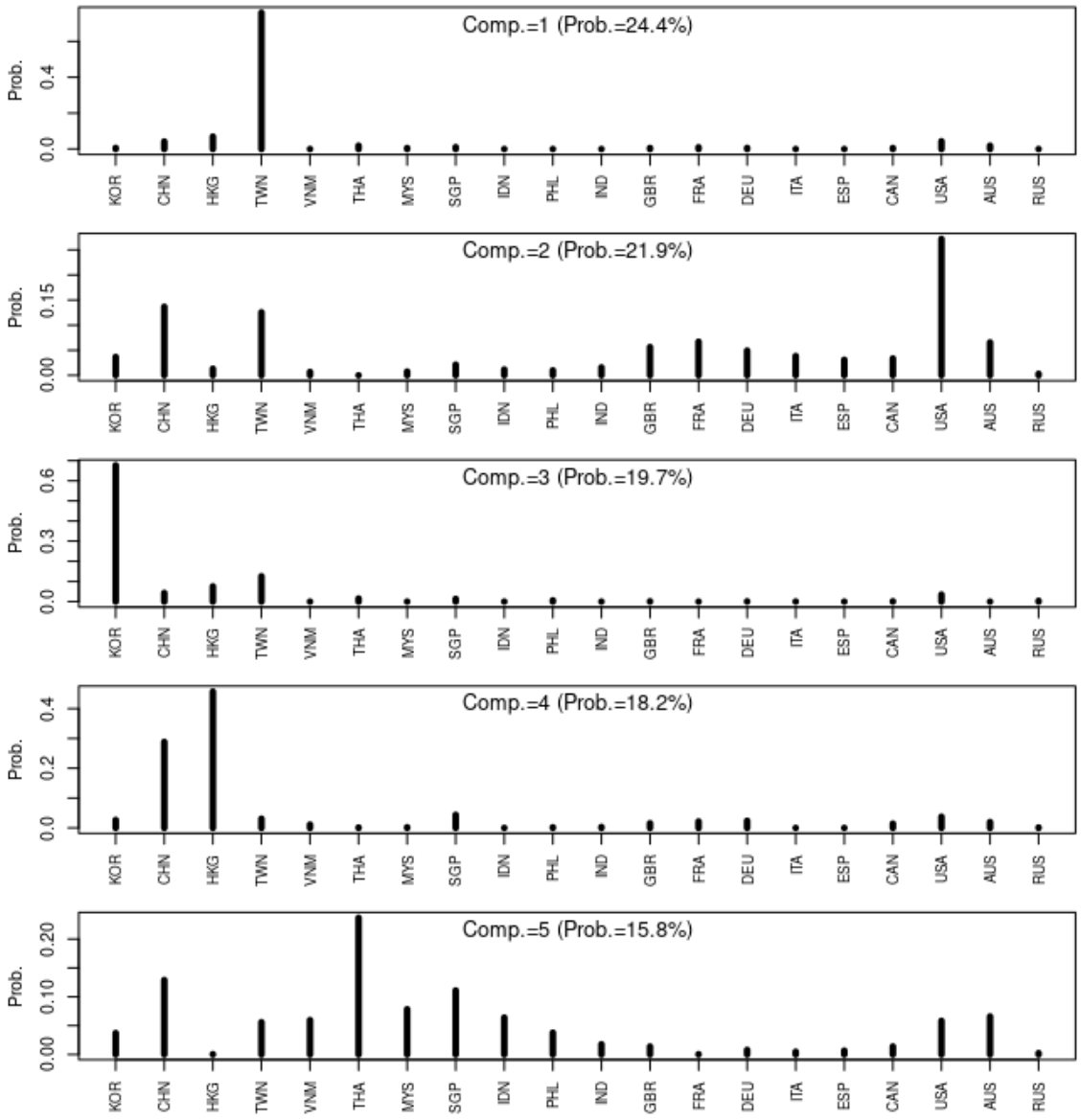


潜在パターン1のカード

「東京が出た」

主な結果

潜在パターン
国籍「空間」



「台湾」(24%)

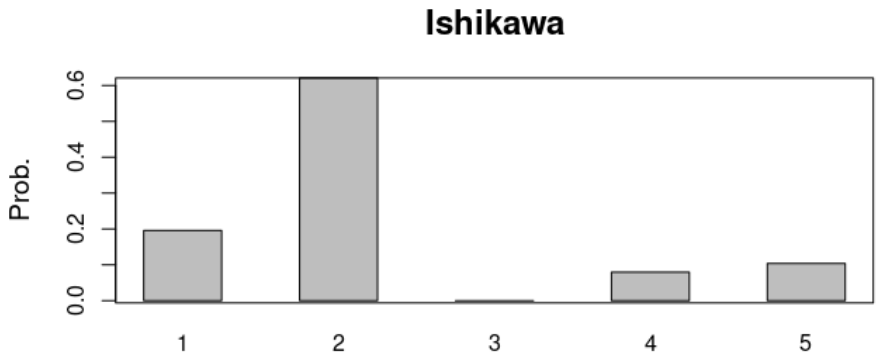
「欧米+中台」(22%)

「韓国」(20%)

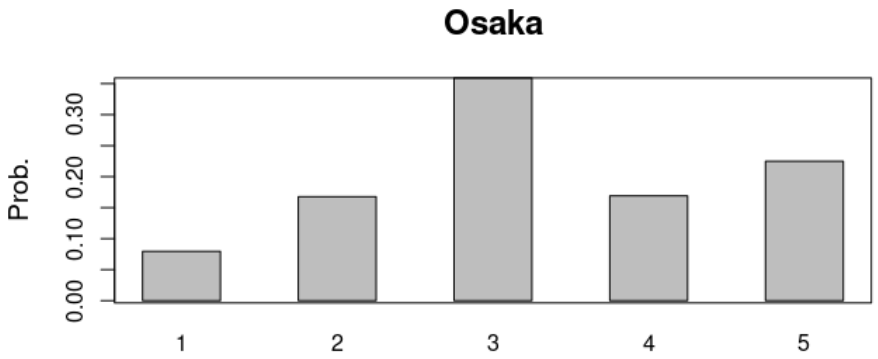
「中国・香港」(18%)

「東南アジア」(16%)

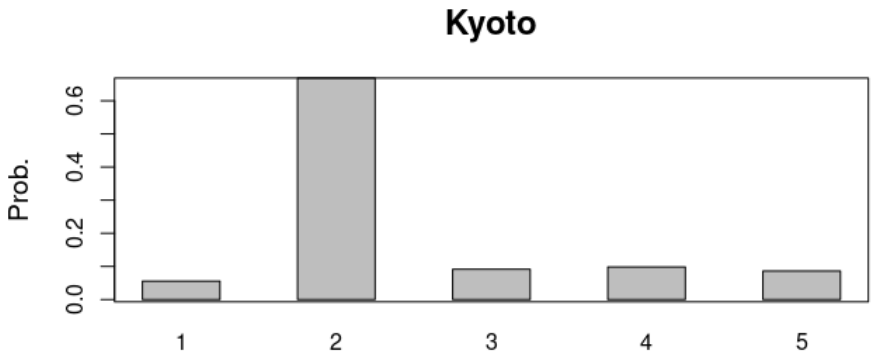
都道府県ごとの潜在パターンへの分解（例）



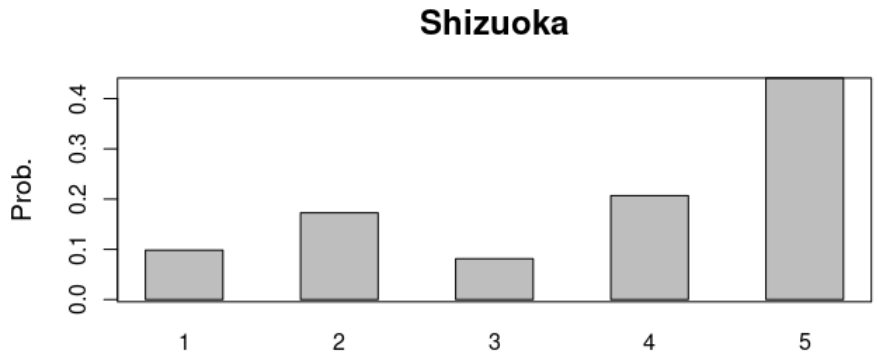
1「台湾」2「欧米+中台」3「韓国」4「中国」5「東南アジア」



1「台湾」2「欧米+中台」3「韓国」4「中国」5「東南アジア」

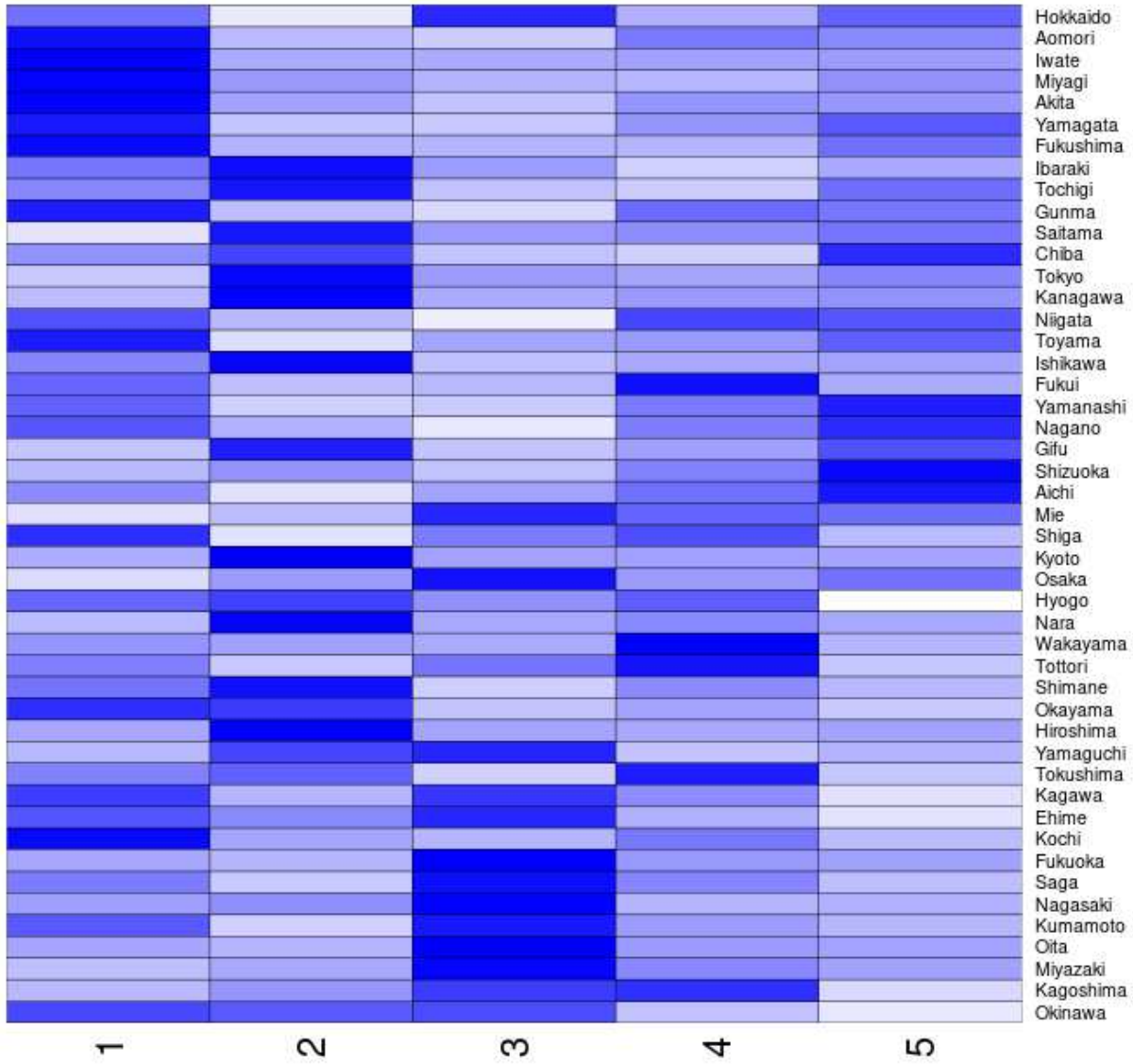


1「台湾」2「欧米+中台」3「韓国」4「中国」5「東南アジア」



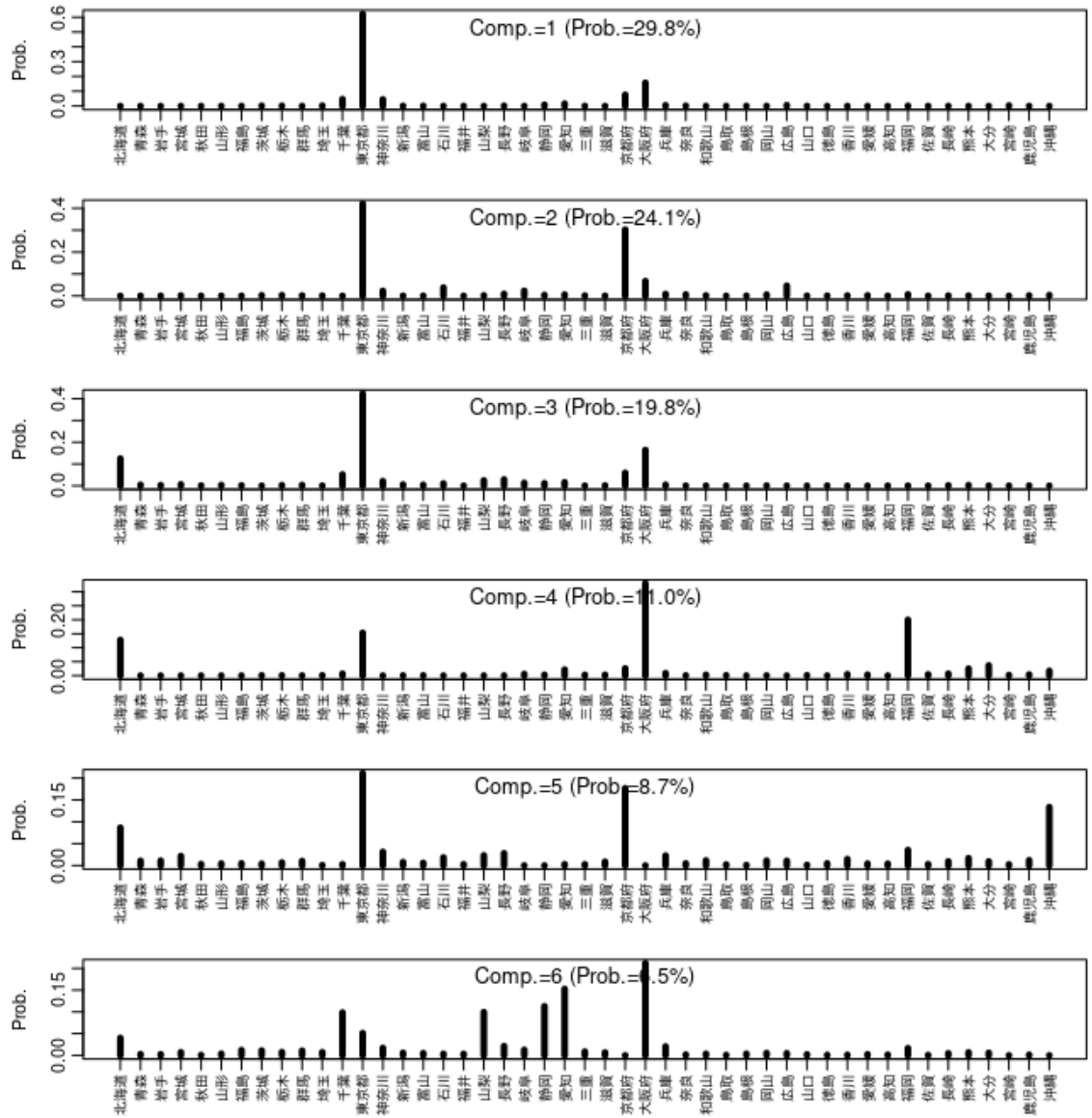
1「台湾」2「欧米+中台」3「韓国」4「中国」5「東南アジア」

都道府県ごとの
潜在パターンへの分解



1「台湾」2「欧米+中台」3「韓国」4「中国」5「東南アジア」

潜在パターン
都道府県「空間」



「東京」(30%)

「東京+京都」(24%)

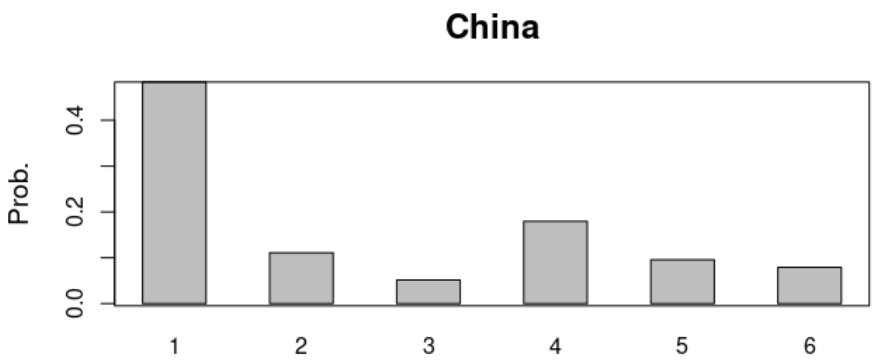
「東京+北海道」(20%)

「大阪+福岡」(11%)

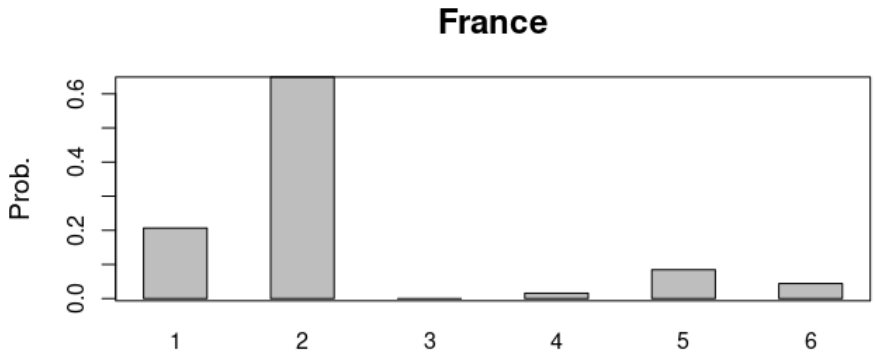
「東京+沖縄」(9%)

「愛知・静岡・
山梨・千葉」(7%)

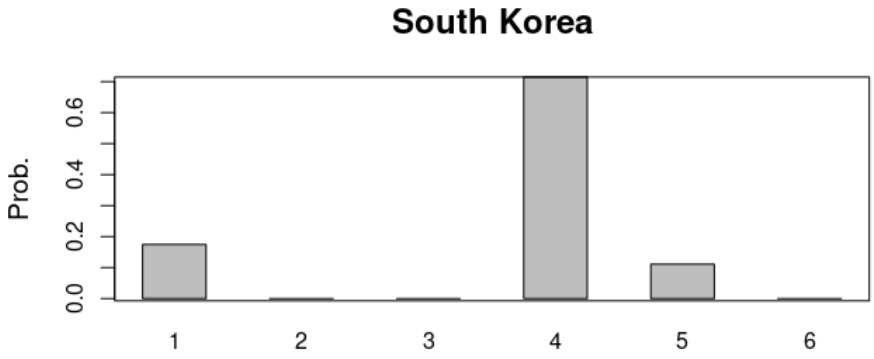
国籍ごとの潜在パターンへの分解（例）



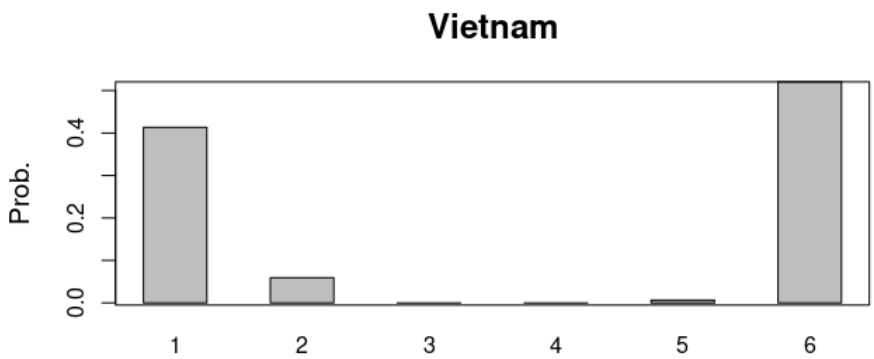
1「東京」2「+京都」3「+北海道」4「大阪+福岡」5「+沖縄」6「愛知他」



1「東京」2「+京都」3「+北海道」4「大阪+福岡」5「+沖縄」6「愛知他」

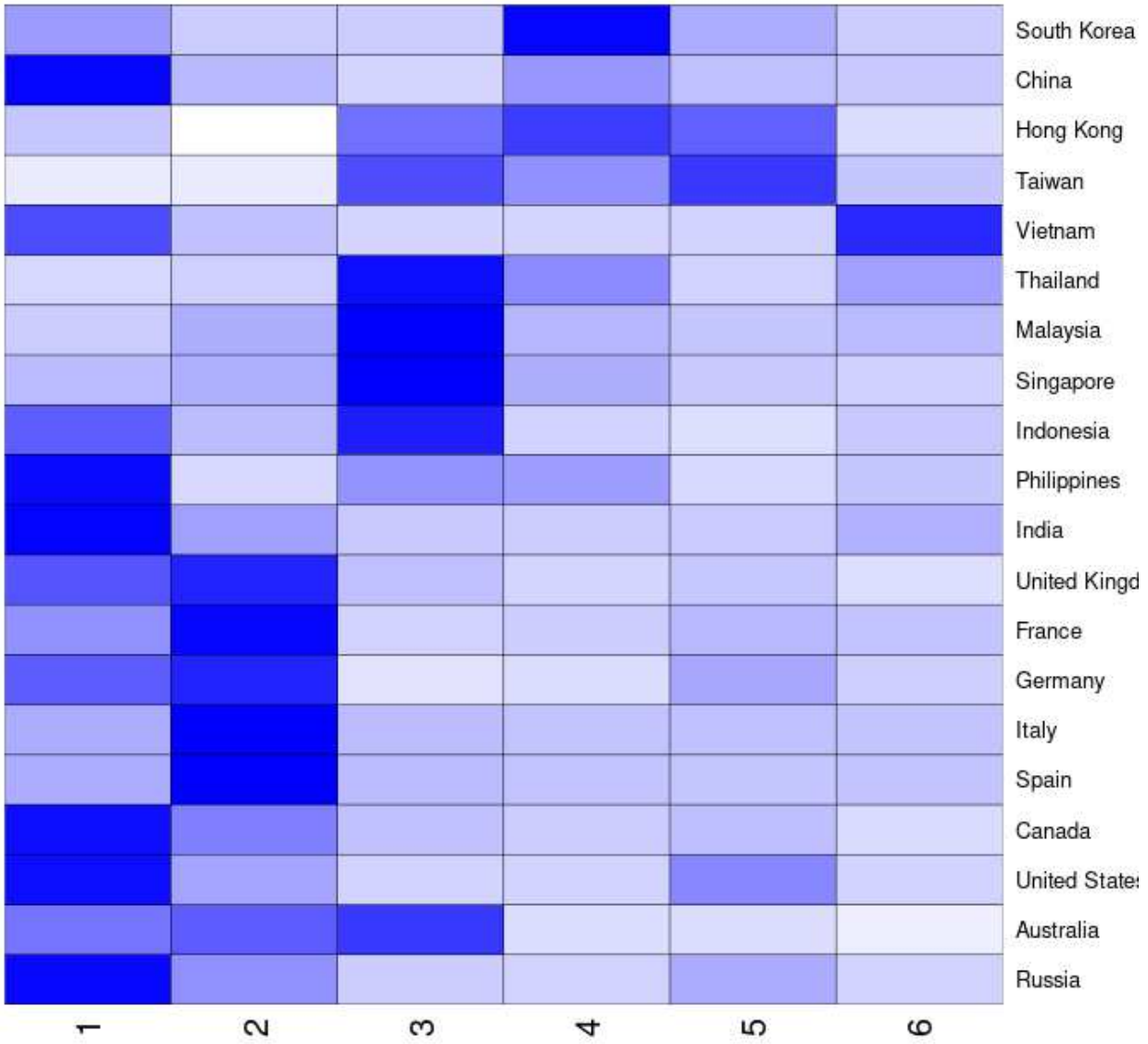


1「東京」2「+京都」3「+北海道」4「大阪+福岡」5「+沖縄」6「愛知他」



1「東京」2「+京都」3「+北海道」4「大阪+福岡」5「+沖縄」6「愛知他」

国籍ごとの
潜在パターンへの分解



1「東京」2「+京都」3「+北海道」4「大阪+福岡」5「+沖縄」6「愛知他」

FAQ

1. 潜在パターンの数はどうやって決めるのか？
   → 確率モデルとして、IC, CV などのアプローチ

2. 確率モデルは正しいのか？
   → 観光、特に宿泊の背景知識との整合性など
   （例えば、空港など交通機関など）

3. 時間的な変化は？
   → 暦年ごとに解析して、安定性の発見や変化の検出