

「ことわざ」を創る！

永田 倭大

今回の目標

- 連関規則の復習
- データの並び替えをする

連関規則の復習

- 連関規則とは
- 前提確率
- 条件付き確率(confidence)
- 同時確率(support)
- 事前確率
- 改善率(lift)

連関規則

- ・属性Aを持つオブザベーションは属性Bを持つ傾向にある
例)「枝豆を買う」→「ビールを買う」

前提確率

- Aが生じる確率が高いと規則を適用できる可能性が高い
- $P(A)$ とかく

条件付き確率

- Aという属性を持っているオブザベーションの中でBという属性を持っているオブザベーションの確率
- Confidenceなどと呼ばれる
- 条件付き確率が高い連関規則は良い規則
- $P(B | A)$ とかく

$$P(B | A) = P(A \cap B) / P(A)$$

同時確率

- 前提確率と条件付き確率の積
- Supportなどと呼ばれる
- $P(A,B)$ とかく

$$P(A,B) = P(A) \times P(B|A)$$

事前確率

- 条件付き確率と比較して事前確率の低い連関規則は良い規則
- $P(B)$ とかく

改善率

- Liftと呼ばれる
- 連関規則の良し悪しを判断するもの

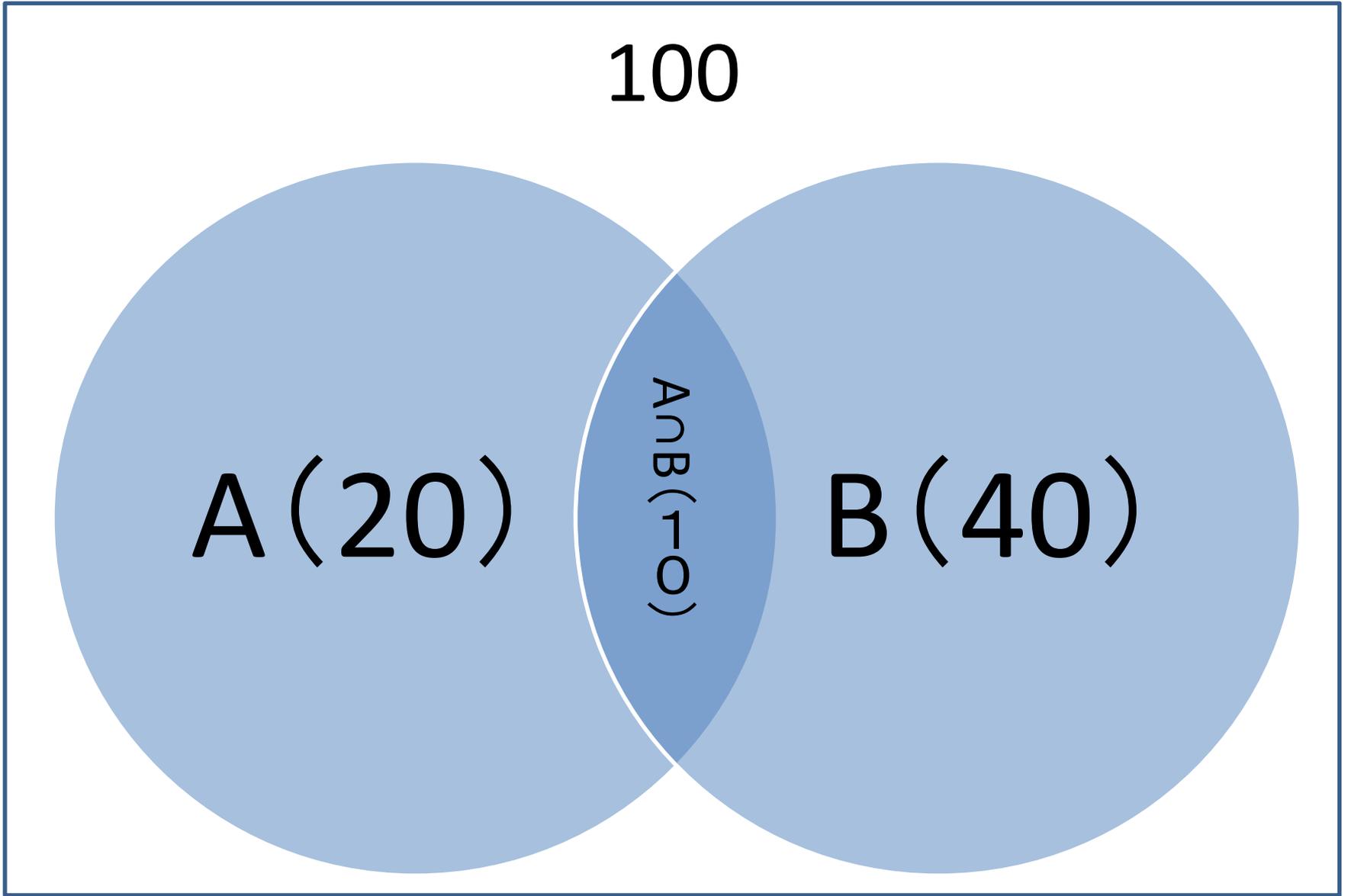
$$\text{Lift} = P(B|A) / P(B) = P(B,A) / (P(A) \times P(B))$$

100

A (20)

$A \cap B$ (10)

B (40)



- 前提確率: $P(A) = 20/100 = 1/5$
- 条件付き確率: $P(B | A) = (10/100) / (20/100) = 1/2$
- 同時確率: $P(A, B) = 1/10$
- 事前確率: $P(B) = 40/100 = 2/5$
- 改善率: $\text{Lift} = (1/2) / (2/5) = 5$

演習

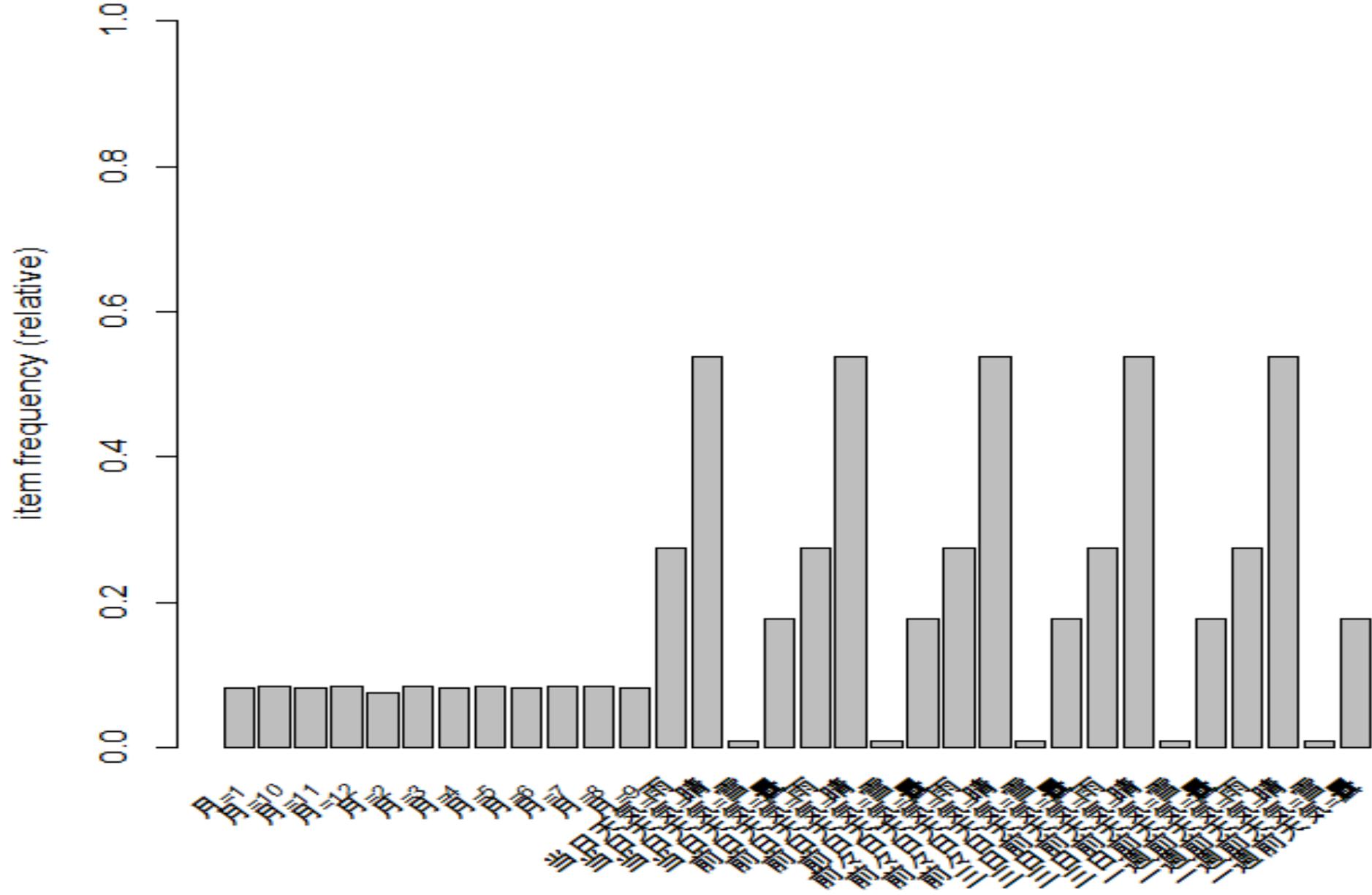
tenki.csvから規則を見つけ並び替えてみよう

```
> setwd("../Dropbox/Share/DL02045/Program/chap5")
> library(arules)
> data<-read.csv("tenki.csv",header=TRUE,row.names=1,colClasses="factor")
> data.trans<-as(data,"transactions")
> itemFrequency(data.trans)
> itemFrequencyPlot(data.trans,ylim=c(0,1),cex.names=0.7)
```

Mac用のデータは"tenkimac.csv"

colClasses="factor": データの読み込み速度を上げる

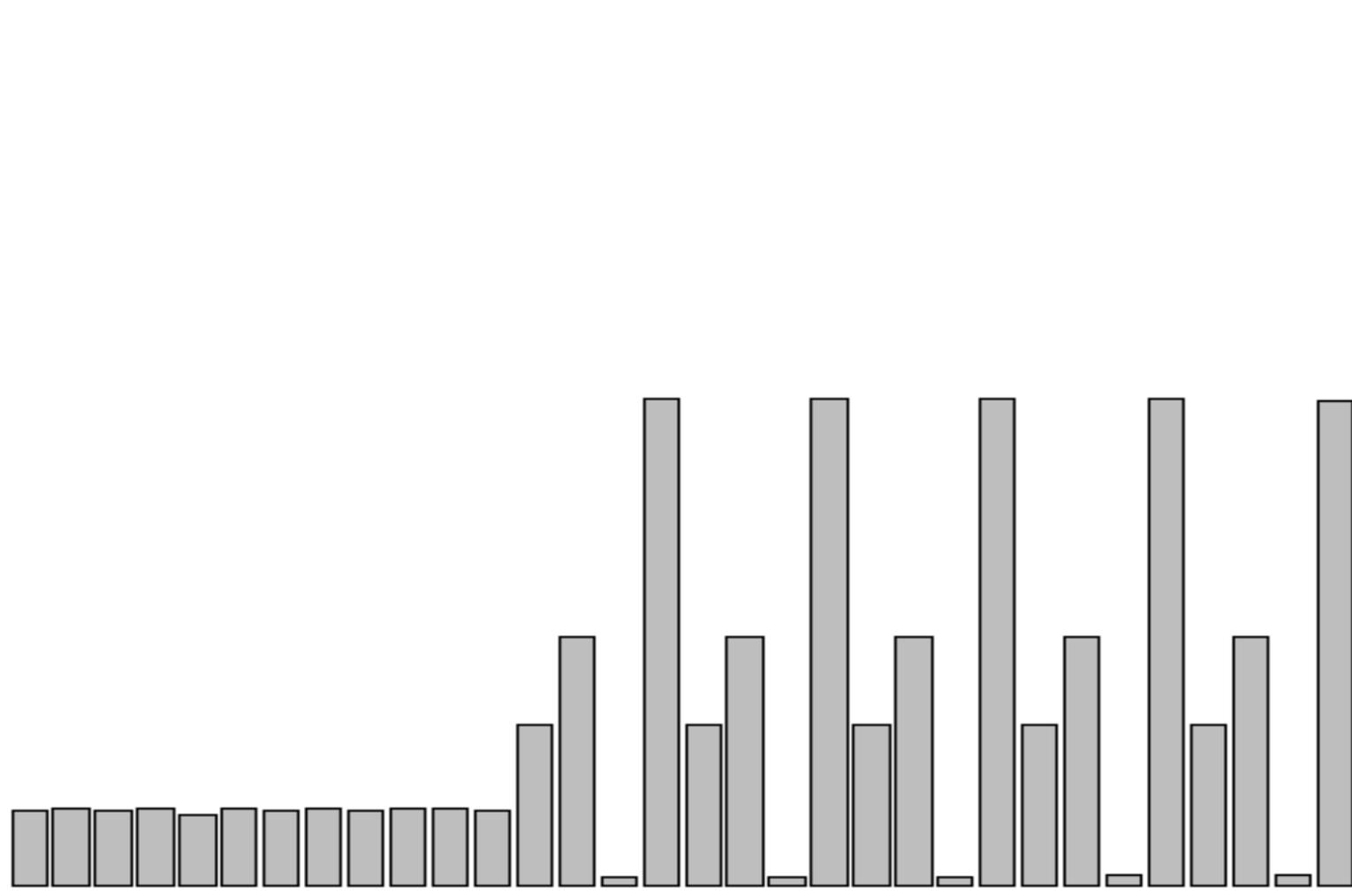
cex.names: 軸ラベルの文字の倍率



item frequency (relative)

1.0
0.8
0.6
0.4
0.2
0.0

month=1
month=10
month=11
month=12
month=2
month=3
month=4
month=5
month=6
month=7
month=8
month=9
day=cloudy
day=rainy
day=snow
day=sunny
yesterday=cloudy
yesterday=rainy
yesterday=snow
yesterday=sunny
two days=cloudy
two days=rainy
two days=snow
two days=sunny
three days=cloudy
three days=rainy
three days=snow
three days=sunny
weekdays=cloudy
weekdays=rainy
weekdays=snow
weekdays=sunny



```
options(digits=3)
```

```
sunnyRule=apriori(data.trans,list(maxlen=5,supp=0.05,conf=0.65,ext=TRUE),list(rhs="当日天気=晴",default="lhs")) ※day=sunny
```

```
inspect(SORT(sunnyRule,"support")[1:10,])
```

```
inspect(SORT(sunnyRule,"confidence")[1:10,])
```

```
sunnyRuleOf1=subset(sunnyRule,lhs %in% "月=1")
```

↑sunnyRuleにおいてlhsが月=1のものを抽出

```
Inspect(sunnyRuleOf1)
```

rhs="当日天気=晴:ルールボディを設定する

Default="lhs":rhsで設定した項目以外をルールヘッドに設定

	lhs	rhs	support	confidence	lhs.support	lift
1	{前日天气=晴}	=> {当日天气=晴}	0.3678	0.683	0.5386	1.27
2	{前日天气=晴, 前々日天气=晴}	=> {当日天气=晴}	0.2613	0.711	0.3676	1.32
3	{前日天气=晴, 三日前天气=晴}	=> {当日天气=晴}	0.2355	0.716	0.3288	1.33
4	{前日天气=晴, 一週前天气=晴}	=> {当日天气=晴}	0.2226	0.706	0.3154	1.31
5	{前日天气=晴, 前々日天气=晴, 三日前天气=晴}	=> {当日天气=晴}	0.1930	0.739	0.2610	1.37
6	{前日天气=晴, 前々日天气=晴, 一週前天气=晴}	=> {当日天气=晴}	0.1628	0.733	0.2221	1.36
7	{前日天气=晴, 三日前天气=晴, 一週前天气=晴}	=> {当日天气=晴}	0.1537	0.742	0.2072	1.38
8	{前々日天气=晴, 三日前天气=晴, 一週前天气=晴}	=> {当日天气=晴}	0.1529	0.691	0.2212	1.28
9	{前日天气=晴, 前々日天气=晴, 三日前天气=晴, 一週前天气=晴}	=> {当日天气=晴}	0.1271	0.762	0.1669	1.41
10	{月=12}	=> {当日天气=晴}	0.0681	0.800	0.0851	1.49

lhs rhs support confidence lhs.support lift

1 {月=12,
前日天气=晴,
前々日天气=晴} => {当日天气=晴} 0.0505 0.885 0.0571 1.64

2 {月=12,
前日天气=晴} => {当日天气=晴} 0.0582 0.862 0.0675 1.60

3 {月=12,
前々日天气=晴} => {当日天气=晴} 0.0560 0.840 0.0667 1.56

4 {月=12,
三日前天气=晴} => {当日天气=晴} 0.0544 0.828 0.0656 1.54

5 {月=12,
一週前天气=晴} => {当日天气=晴} 0.0535 0.812 0.0659 1.51

6 {月=12} => {当日天气=晴} 0.0681 0.800 0.0851 1.49

7 {前日天气=晴,
前々日天气=晴,
三日前天气=晴,
一週前天气=晴} => {当日天气=晴} 0.1271 0.762 0.1669 1.41

8 {前日天气=晴,
三日前天气=晴,
一週前天气=晴} => {当日天气=晴} 0.1537 0.742 0.2072 1.38

9 {前日天气=晴,
前々日天气=晴,
三日前天气=晴} => {当日天气=晴} 0.1930 0.739 0.2610 1.37

10 {前日天气=晴,
前々日天气=晴,

感想

- 並べ替えができて楽しかった

