

第2回基礎ゼミ(プログラミングゼミ) 分析結果の集約・視覚化の方法(R)

2017/5/2

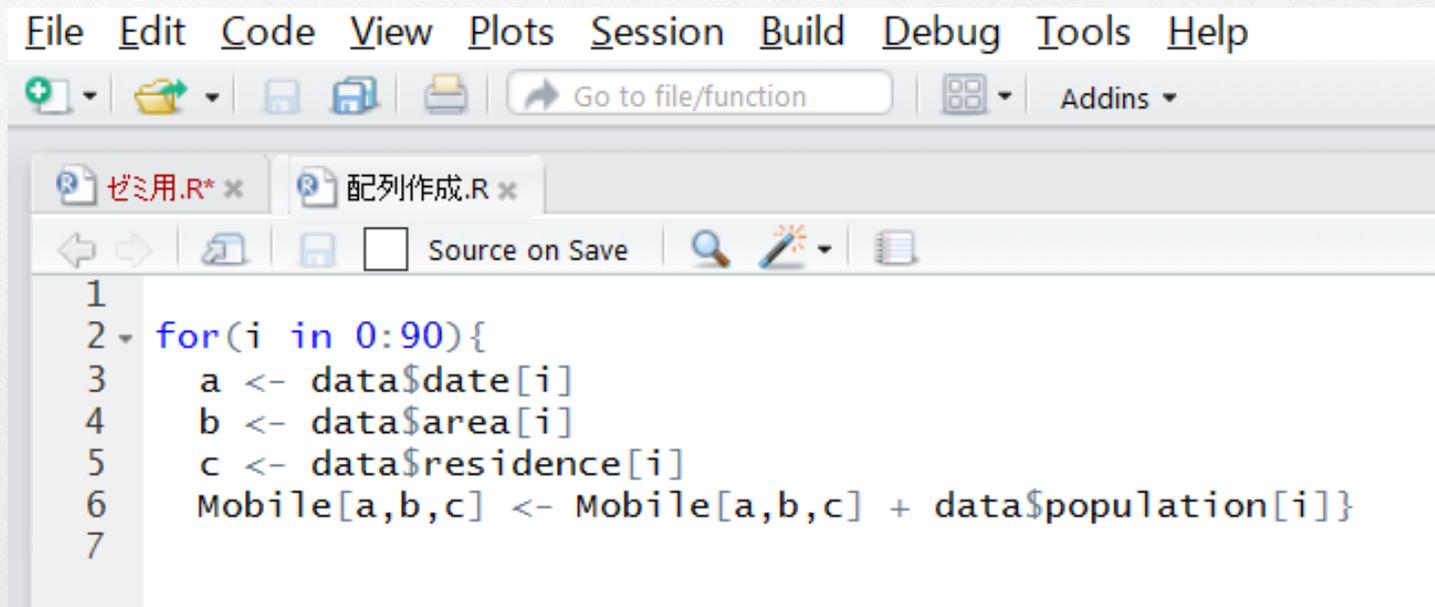
鈴木 新
福田研究室

目次

- 保存したプログラムの読み込み
- 分析・調査結果の集約
 - データ説明
 - データの読み込み
 - データセットの構築
- 結果の可視化(グラフ作成)
 - 折れ線グラフ
 - 棒グラフ
 - 円グラフ
- グラフの応用

保存したプログラムの読み込み

File→Open File→保存したRファイルを選択



The screenshot shows the RStudio interface. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. The toolbar contains icons for file operations and a search box labeled 'Go to file/function'. The file explorer shows two open files: 'ゼミ用.R*' and '配列作成.R'. The script editor displays the following R code:

```
1
2 - for(i in 0:90){
3   a <- data$date[i]
4   b <- data$area[i]
5   c <- data$residence[i]
6   Mobile[a,b,c] <- Mobile[a,b,c] + data$population[i]}
7
```

分析結果の集約

分析結果の集約

- 調査でサンプルデータを収集
- マスターデータの状態だとよく分からない
- データセットの構築し、それから分析を行う

	date	area	residence	population	
1					
2		1	1	8838799	
3		1	1	138629	
4		1	1	215917	
5		1	2	1186455	
6		1	2	1042802	
7		1	2	203692	
8		1	3	1	2974746
9		1	3	2	218569
10		1	3	3	730391
11		1	3	1	8567867
12		1	3	2	185797
13		1	3	3	222783
14		2	1	1	57017
15		2	1	1	728
16		2	1	1	0083
17		2	1	1	75116
18		2	3	1	107475
19		2	3	1	737134
20		3	1	1	8016227
21		3	1	2	151582
22		3	1	3	203713
23		3	2	1	1304611
24		3	2	2	1158083
25		3	2	3	191063
26		3	3	1	3679162
27		3	3	2	90335
28		3	3	3	755224
29		4	1	1	8418944
30		4	1	2	123099
31		4	1	3	234537
32		4	1	1	1751732

見にくい
分析しづらい

データ説明

- 基本データはcsvファイル
(Rもcsvファイルじゃないと読み込めない)

xlsxファイル

セルに打ち込んだ式やグラフ、値が保存される。

	A	B	C	D
1	1			
2	2		=A2*A4-A6	
3	3			
4	4			
5	5			
6	6			
7	7			
8				

csvファイル

1枚目のsheetのセルに打ち込んだ値のみ保存される。

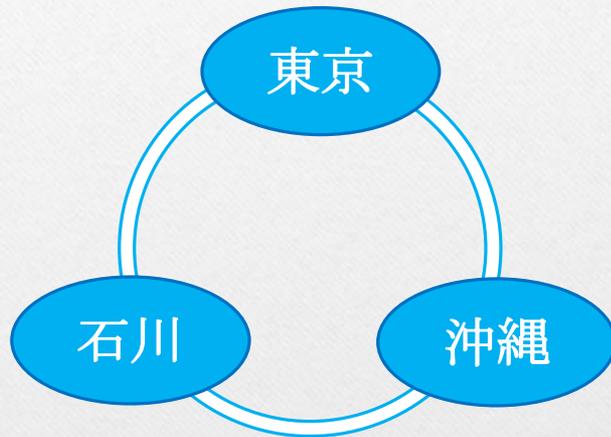
	A	B	C
1	1		
2	2		2
3	3		
4	4		
5	5		
6	6		
7	7		

csvファイルで保存する時は式とかグラフとか消えるから注意！

データ説明

～今回使用するマスターデータの紹介～

- ある10日間の東京都・沖縄県・石川県のOD(3×3)表
(モバイル空間統計仮想データ)



	date	area	residence	population	
1					
2		1	1	1	8838799
3		1	1	2	138629
4		1	1	3	215917
5		1	2	1	1186455
6		1	2	2	1042802
7		1	2	3	203692
8		1	3	1	2974746

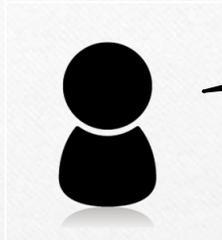
1日目に石川県出身で
東京都にいる人の数

～マスターデータあるある～

東京=1, 沖縄=2, 石川=3

といったように文字ではなく数字で表される

元データの読み込み



データ読み込んでほしい！

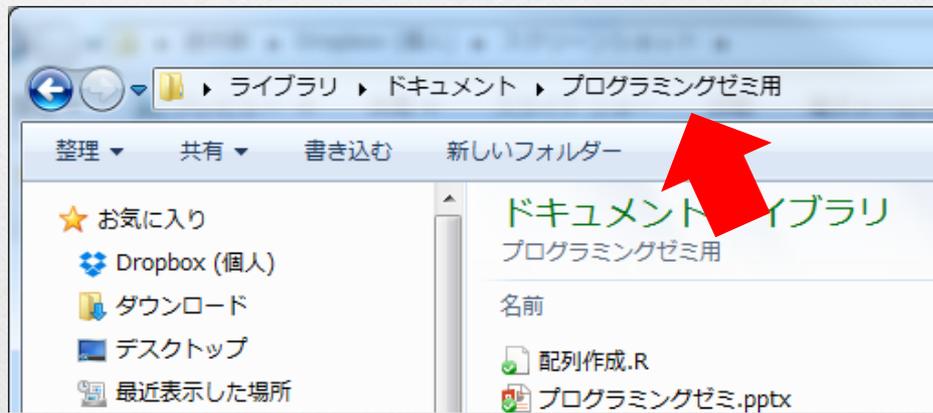
どこにあるの？



このパソコン上のどこどこにデータがある

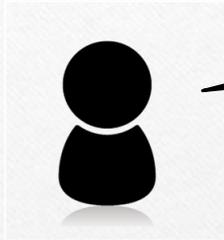
→setwd(“**C:¥Users¥a.suzuki**”)

↑読み込むcsvファイルがあるフォルダ



- フォルダのディレクトリをクリックしてコピー
- スラッシュの向きを逆に (＼を／に)

元データの読み込み



データ読み込んでほしい！

ファイル名は？



～って名前のファイルを読み込んでほしい

→`read.csv("モバ空サンプル.csv")`
↑ファイル名

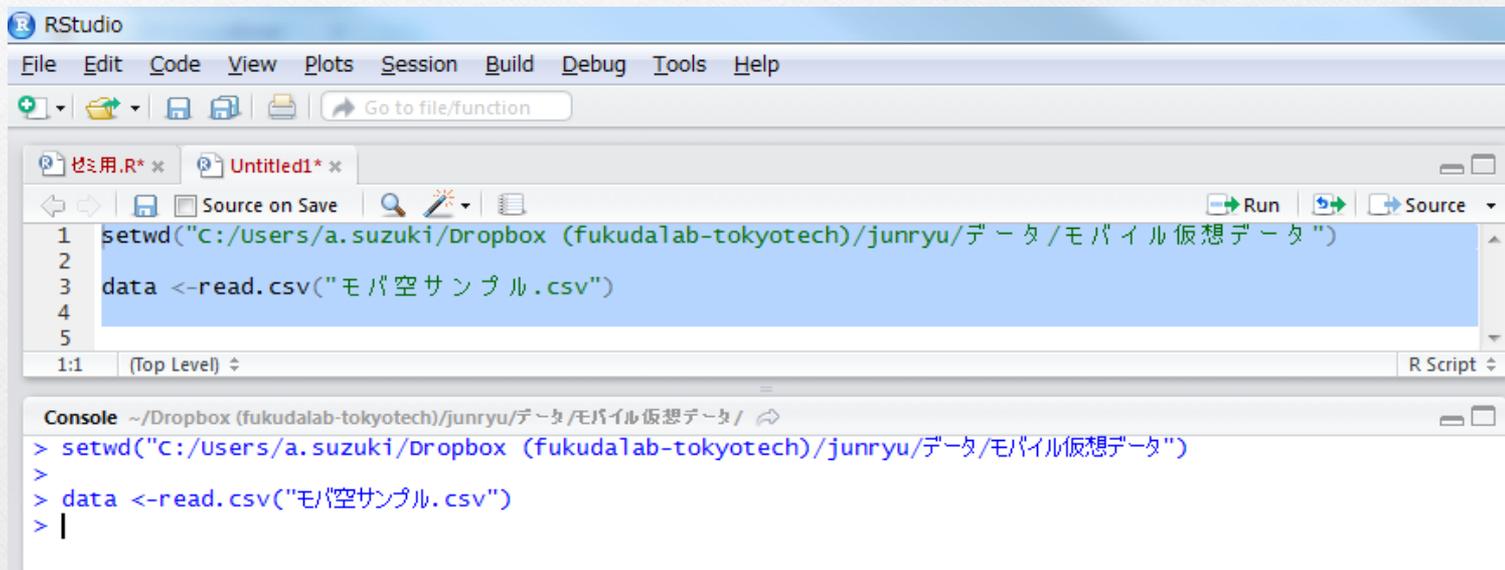
読み込むだけじゃなくてそのデータを
すぐ呼び出せるようにする

```
5 data <-read.csv("モバ空サンプル.csv")
6   #dataと打てばすぐ読み込んだデータが出てくる
7
8 |
```

“<-”は代入って意味

元データの読み込み

- 読み込んでみよう



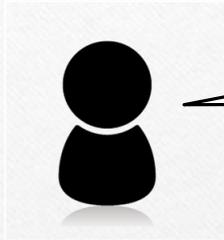
The screenshot shows the RStudio interface. The source editor contains the following R code:

```
1 setwd("C:/Users/a.suzuki/Dropbox (fukudalab-tokyotech)/junryu/データ/モバイル仮想データ")
2
3 data <-read.csv("モバ空サンプル.csv")
4
5
```

The console shows the execution of the code:

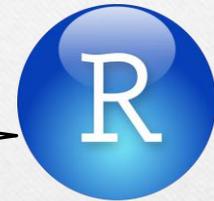
```
> setwd("C:/Users/a.suzuki/Dropbox (fukudalab-tokyotech)/junryu/データ/モバイル仮想データ")
>
> data <-read.csv("モバ空サンプル.csv")
> |
```

データセットの構築



1日目の東京→沖縄の
旅客数が知りたい！

それはdataの何行何列目
にあるの？



データを読み込んだだけじゃまだやりにくい
→データセットの作成

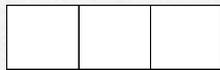
～配列(array)～

- 項目(次元)ごとに整理できるため、データを取り出す時や計算する際に便利
- 今回だと何日目の東京出身の人が何県にいる人数は
何人かすぐに分かる

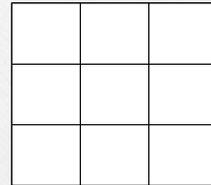
データセットの構築

- 配列のイメージ

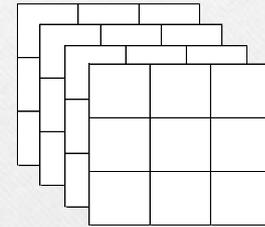
... □ マス一個にひとつの値が入る



1項目 (1次元)



2項目 (2次元)



3項目 (3次元)

- 書き方

最初空の配列 (0行列) を作り, 値を代入していく

```
array(0:0, dim=c(10,3,3))
```

入れる数値の範囲

次元: 10(日数) × 3(滞在地) × 3(居住地)の場合

データセットの構築

- 配列を作成してみよう

```
1 setwd("C:/Users/a.suzuki/Dropbox (fukudalab-tokyotech)/junryu/データ/モバイル仮想デー  
2  
3 data <- read.csv("モバ空サンプル.csv")  
4 #dataと打てばすぐ読み込んだデータが出てくる  
5  
6 Mobile <- array(0:0, dim=c(10,3,3))#空の配列, Mobileは配列の名前(何でもいい)  
7  
8 for(i in 0:90)  
9 {  
10 a <- data$date[i]  
11 b <- data$area[i]  
12 c <- data$residence[i]  
13 Mobile[a,b,c] <- Mobile[a,b,c] + data$population[i]}  
14  
15 Mobile[3,,]#(日付*旅行先*居住地)
```

次回やるよ!!!

```
> Mobile <- array(0:0, dim=c(10,3,3))#空の配列, Mobileは配列の名前(何  
>  
> for(i in 0:90)  
+ {  
+ a <- data$date[i]  
+ b <- data$area[i]  
+ c <- data$residence[i]  
+ Mobile[a,b,c] <- Mobile[a,b,c] + data$population[i]}  
>  
> Mobile[3,,]#(日付*旅行先*居住地)  
      [,1] [,2] [,3]  
[1,] 10009370 120354 242897  
[2,] 1699304 1181866 165863  
[3,] 1291326 97780 741240
```

Mobile[,,]

に日数, 滞在地, 居住地
を打ち込むと値が出てくる

東京:1

沖縄:2

石川:3

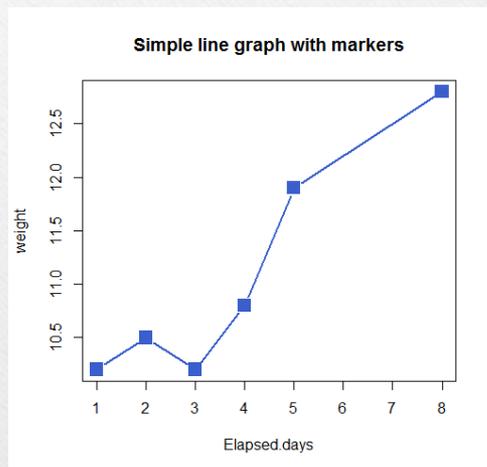
データの可視化

代表的な図

集約したデータを見やすくしたい

→グラフの作成

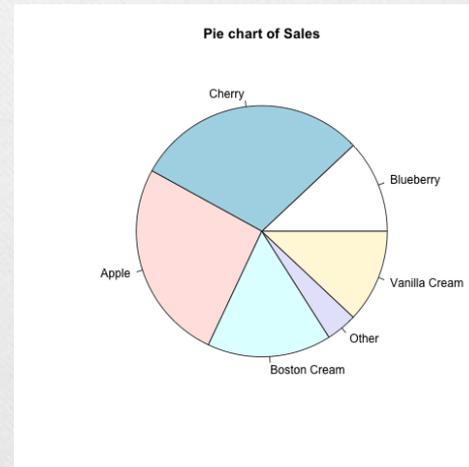
●代表的な図



折れ線グラフ(plot)



棒グラフ(barplot)

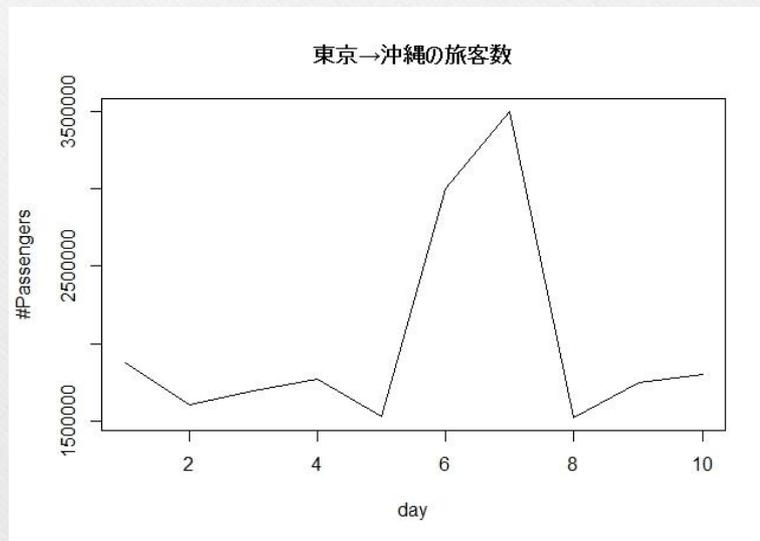


円グラフ(pie)

折れ線グラフ (Plot関数)

東京→沖縄に行った人の10日間の変動を見る

```
19 plot(Mobile[,2,1], #グラフ化するデータ
20     main = "東京→沖縄の旅客数", #タイトル
21     xlab="day", #横軸名称
22     ylab="#Passengers", #縦軸名称
23     type = "l", #グラフのタイプ (折れ線グラフ)
24     col="black", #グラフの色
25 )
26
```

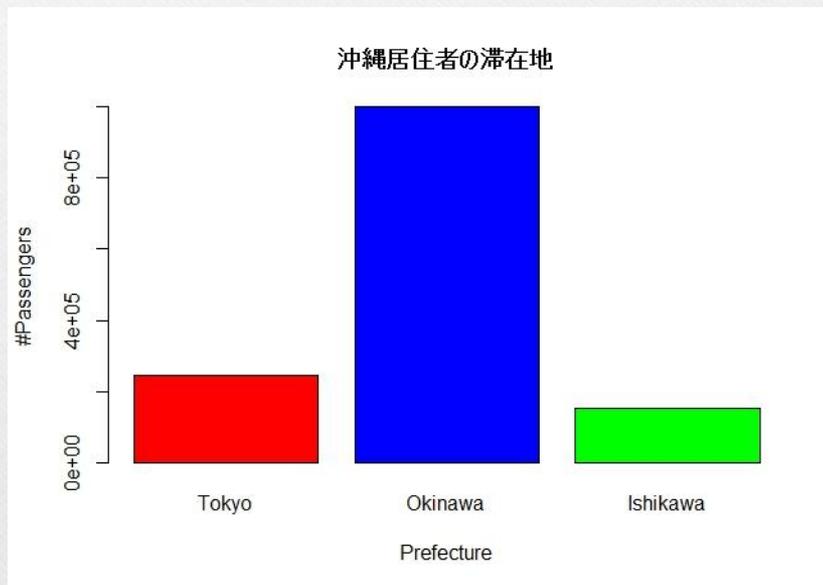


6日目, 7日目はかなり多くの人が沖縄に...
→土日?

棒グラフbarplot関数

4日目に沖縄県出身の人がどこに何人行ったか

```
27 barplot(Mobile[4,,2], #グラフ化するデータ
28     main = "沖縄居住者の滞在地", #タイトル
29     xlab="Prefecture", #横軸名称
30     ylab="#Passengers", #縦軸名称
31     names.arg = c("Tokyo", "Okinawa", "Ishikawa"), #各データ名称, c(,)はベクトル
32     col = c("red", "blue", "green"), #グラフの色, 左から順に
33 )
```

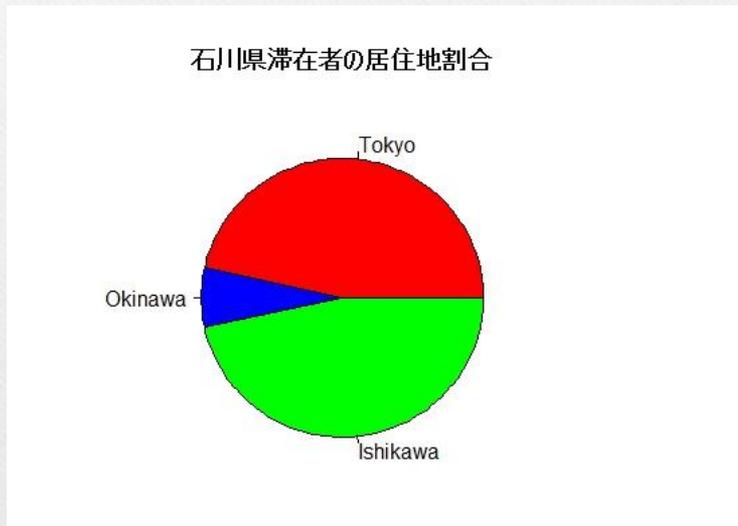


- 多くの方は沖縄にいる。(居住地だから?)
- 石川より東京に行く人のほうが多い(飛行機の直行便が多い分?)

円グラフpie関数

5日目の石川県滞在者の居住地別割合

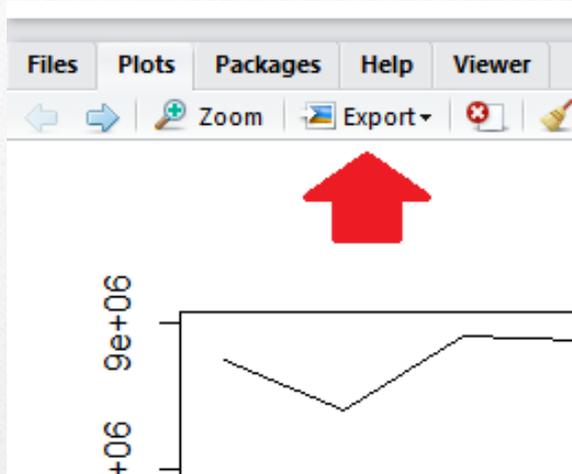
```
34 pie(Mobile[5,3,], #グラフ化するデータ
35     main = "石川県滞在者の居住地割合", #タイトル
36     labels <- c("Tokyo", "Okinawa", "Ishikawa"), #各データ名称, c(,,)はベクトル
37     col = c("red", "blue", "green"), #グラフの色,
38 )
```



- 東京からの旅客が多い
- 東京多すぎない？
(東京の人口は1300万人に対し石川は115万人, 人口の差)

って言うより仮想データだからです, すみません

グラフの出力



- 図はpdfやjpegに保存できる
- 綺麗なpdfのほうが論文とかには使われやすい

Rで他にどんな図が描けるか

→ `demo(graphics)`

をCtrl+R押し続ける

グラフの応用

グラフの応用

●図の重ね合わせ

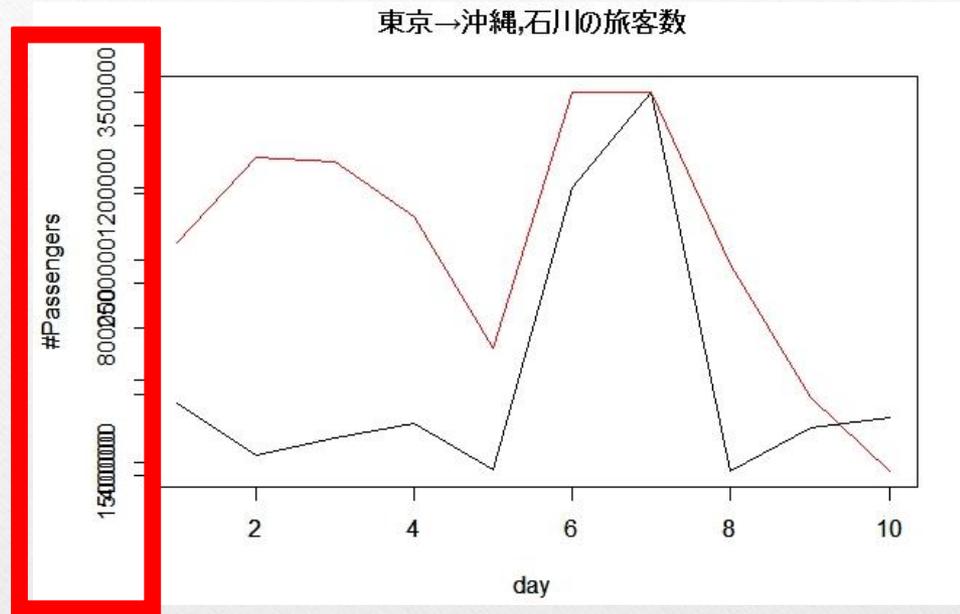
複数の折れ線グラフを1枚の図に描きたい場合

→plot()とplot()の間に**par(new=T)**関数を書く

東京→沖縄, 東京→石川に行った人の10日間の変動を見る

```
43 plot(Mobile[,2,1], #グラフ化するデータ
44     main = "東京→沖縄,石川の旅客数", #タイトル
45     xlab="day", #横軸名称
46     ylab="#Passengers", #縦軸名称
47     type = "l", #グラフのタイプ(折れ線グラフ)
48     col="black", #グラフの色
49 )
50 par(new=T)
51 plot(Mobile[,3,1], #グラフ化するデータ
52     xlab="day", #横軸名称
53     ylab="#Passengers", #縦軸名称
54     type = "l", #グラフのタイプ(折れ線グラフ)
55     col="red", #グラフの色
56 )
```

グラフの応用



縦軸が変なことに...

`par(new=T)`はあくまでグラフの重ね合わせ

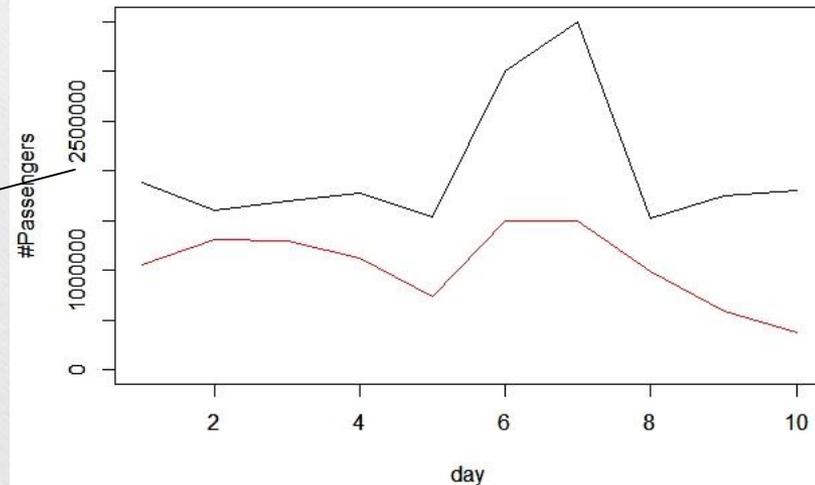
→軸の範囲やタイトルは自分であわせなきゃいけない

グラフの応用

```
43 plot(Mobile[,2,1], #グラフ化するデータ
44       main = "東京→沖縄,石川の旅客数", #タイトル
45       xlab="day", #横軸名称
46       ylab="#Passengers", #縦軸名称
47       ylim=c(0,3500000), #縦軸範囲
48       type = "l", #グラフのタイプ(折れ線グラフ)
49       col="black", #グラフの色
50 )
51 par(new=T)
52 plot(Mobile[,3,1], #グラフ化するデータ
53       xlab="day", #横軸名称
54       ylab="#Passengers", #縦軸名称
55       ylim=c(0,3500000), #縦軸範囲
56       type = "l", #グラフのタイプ(折れ線グラフ)
57       col="red", #グラフの色
58 )
```

縦軸の範囲を追加

東京→沖縄,石川の旅客数



重ね合わせのダブリ
が無くなった