



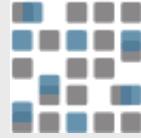
# R 入門

投野ゼミ合宿2019

謝辞:一部は以下の作者のスライドを翻訳:

Tyler K. Perrachione (MIT)

Eugene Tseytlin (Univ. of Pittsburgh)



# 統計ソフトの推移

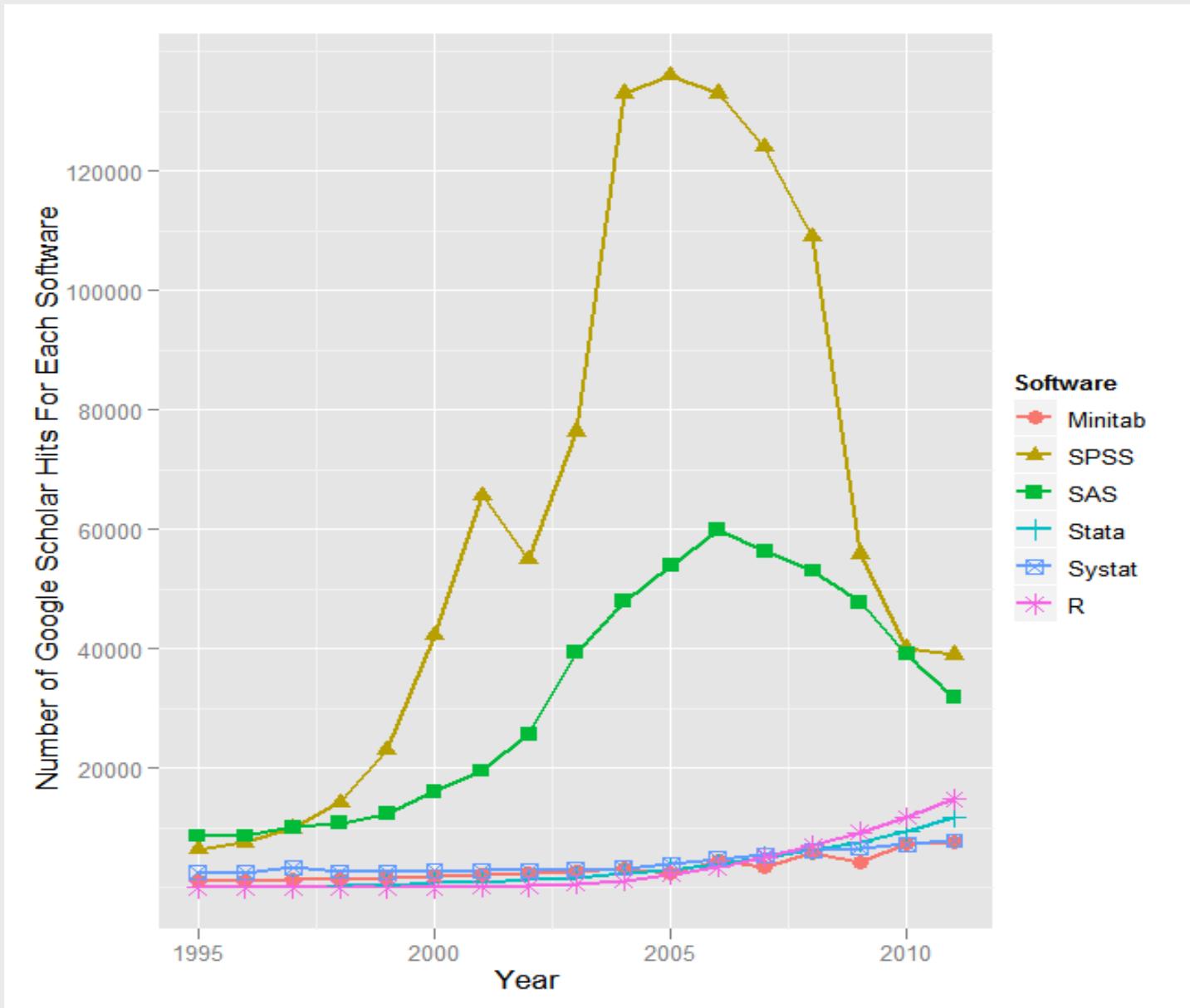
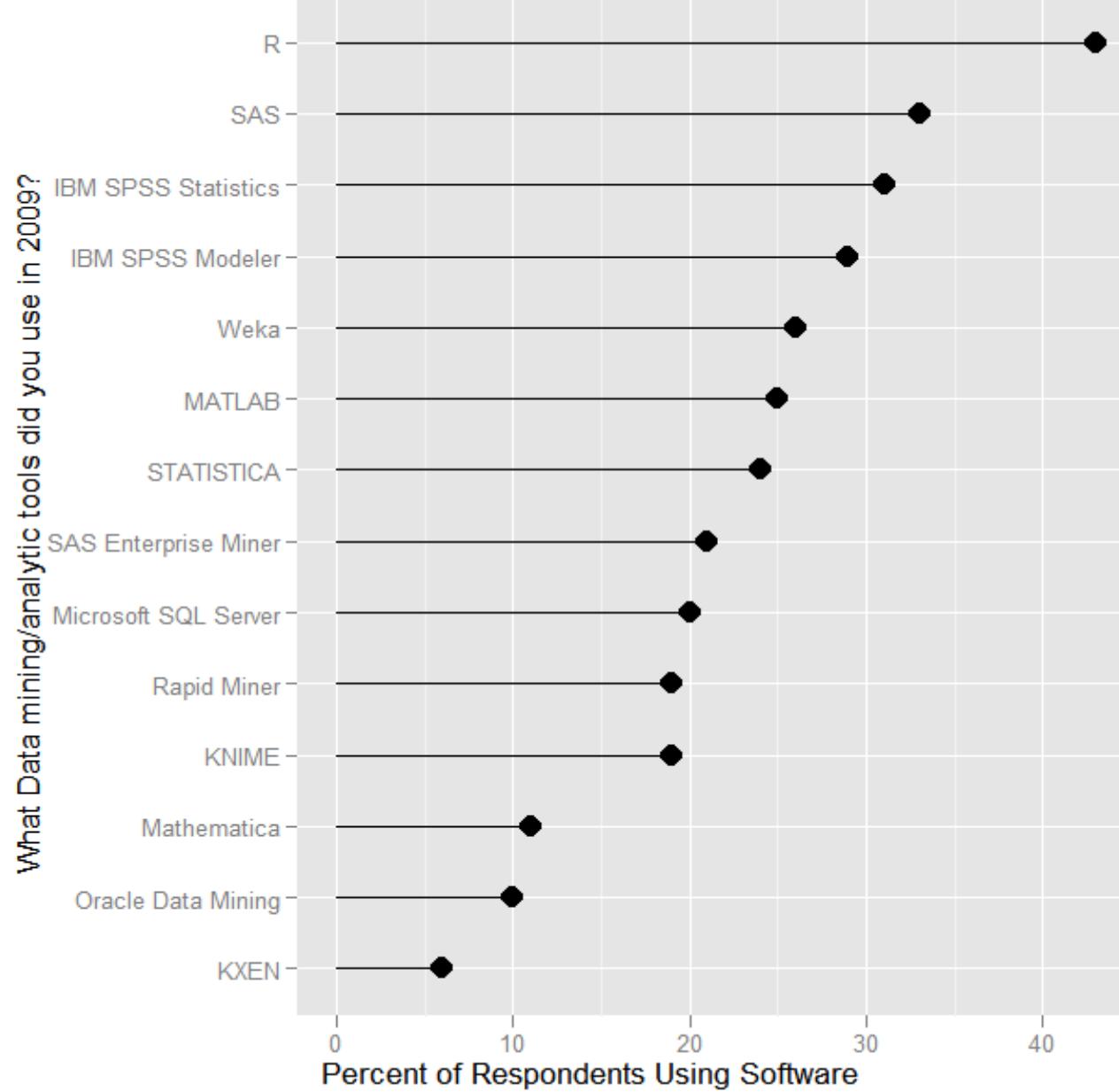


Figure 7a. Use of data analysis software in academic publications as measured by hits on Google Scholar.



# 統計ソフト利用調査





# Microsoft Excel

Sales Data - Microsoft Excel

Sheet Insert Page Layout Formulas Data Review View

PivotTable Table Column Line Pie Bar Area XY (Scatter) Other Charts Picture IGX Shapes Hyperlink Text Box Header WordArt Signature Object Symbol Links Illustrations

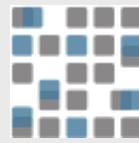
H18 f<sub>x</sub>

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3	Total Sales by Store			Total Sales by Region									
4	Store	Sales		Region	Sales								
5	Store 1	\$ 313,765		West	✓ \$ 1,718,258								
6	Store 2	\$ 107,160		South	✗ \$ 534,389								
7	Store 3	\$ 351,751		Midwest	⌚ \$ 1,009,268								
8	Store 4	\$ 131,047		East	✗ \$ 900,431								
9	Store 5	\$ 252,136		Total	\$ 4,162,346								
10	Store 6	\$ 167,462											
11	Store 7	\$ 210,073											
12	Store 8	\$ 308,092		Total Sales by Category									
13	Store 9	\$ 97,492		Category	Sales								
14	Store 10	\$ 393,484		Automotive	\$ 86,285								
15	Store 11	\$ 396,891		Gardening	\$ 52,048								
16	Store 12	\$ 151,168		Electronics	\$ 83,026								
17	Store 13	\$ 251,390		Jewelry	\$ 93,035								
18	Store 14	\$ 392,776		Sporting	\$ 50,016								
19	Store 15	\$ 259,654		Houseware	\$ 19,149								
20	Store 16	\$ 225,184		Books	\$ 42,247								
21	Store 17	\$ 335,785		Games	\$ 18,420								
22													
23													
24													
25													
26													

TOTAL SALES BY CATEGORY

Legend:

- Automotive
- Gardening
- Electronics
- Jewelry
- Sporting
- Housewares
- Books
- Games



# SPSS

\*stroke\_survival.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

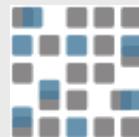
Reports Descriptive Statistics Tables Compare Means General Linear Model Generalized Linear Models Mixed Models Correlate Regression Loglinear Neural Networks Classify Dimension Reduction Scale Nonparametric Tests Forecasting Survival Multiple Response Missing Value Analysis... Multiple Imputation Complex Samples Quality Control ROC Curve...

Visible: 42 of 42 Variables

	patid		gender	active	obesity	diabetes	bp
1	9735702127	29	-54	Female	Yes	No	Hypotension
2	4852351830	79	-74	Male	Yes	Yes	Hypotension
3	3434994256	79	-74	Female	Yes	Yes	Hypotension
4	6053971728	82	-74	Male	Yes	No	Normal
5	9370757269	29				No	No Hypotension
6	3537185320	29				Yes	No Normal
7	0275365329	82				No	Yes Normal
8	3906583332	79				No	No Normal
9	4785366661	82				No	No Normal
10	9589919145	82				No	No Hypotension
11	4598012219	79				Yes	No Normal
12	3629441662	79				No	No Normal
13	5307816588	79				No	No Hypotension
14	5357069859	82				Yes	No Normal
15	5132742071	29				Yes	Yes Normal
16	2660586207	29				Yes	No Hypotension
17	5408312498	79				No	No Hypotension
18	9069087682	29				No	No Hypotension
19	8173197592	799998	58	55-64	Female	No	No Normal
20	8808732689	822229	83	75+	Male	Yes	No Hypotension
21	5666440246	822229	67	65-74	Female	Yes	Yes No Normal

Data View Variable View

Linear... IBM SPSS Statistics Processor is ready



# SAS

SAS

File Edit View Go Tools Solutions Window Help

Log - (Untitled)

```
2308 goptions reset=all;
2309 goptions hsize=5 in vsize=4 in ;
2310 ods html file="fig4_short.html" nogtitle nogfootnote opt
2310! ;
2311 options noimageprint;
2312 title "2008 Year to Date Weekly Report";
2313 proc tabulate data=yr2008 noseps ;
2314 var volnew high low close;
2315 table date ='', (high='Weekly High' low='Weekly Low'
2315! volnew='Volume(100,000)')
2316 * mean='' * f=comma15. / rts=15;
2317 class date;
2318 run;
2319 title;
2320 proc gchart data=work.sectors;
2321 pie Sector / sumvar=Percentage descending detail=Issuer
2322 value=none other=5 otherlabel='Combined'
2323 noheading legend html=htmlvar name='figure_
2324 run;
2324! quit;
2325 ods html close;
```

Editor - Untitled1 \*

```
goptions reset=all;
goptions hsize=5 in vsize=4 in ;
ods html file="fig4_short.html" nogtitle nogfootnote optio
goptions noimageprint;
title "2008 Year to Date Weekly Report";
proc tabulate data=yr2008 noseps ;
var volnew high low close;
table date ='', (high='Weekly High' low='Weekly Low' cl
* mean='' * f=comma15. / rts=15;
class date;
run;
title;
proc gchart data=work.sectors;
pie Sector / sumvar=Percentage descending detail=Issuer de
value=none other=5 otherlabel='Combined'
```

Results Viewer - file:///C:/SAStemp/fig4\_short.html

### 2008 Year to Date Weekly Report

	Weekly High	Weekly Low	Weekly Close	Volume(100,000)
04JAN08	13,365	12,789	12,800	10,789
11JAN08	12,931	12,502	12,606	15,895
18JAN08	12,795	12,022	12,099	20,082
25JAN08	12,487	11,635	12,207	18,246

A pie chart titled '2008 Year to Date Weekly Report' showing the distribution of sectors. The sectors are Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Information Technology, Materials, and Combined. The 'Combined' sector is the largest, followed by Consumer Staples, Financials, and Consumer Discretionary.

Sector	Percentage
Consumer Discretionary	~15%
Consumer Staples	~10%
Energy	~5%
Financials	~8%
Health Care	~3%
Information Technology	~7%
Materials	~2%
Combined	~30%

Output - (Untitled) Log - (Untitled) Editor - Untitled... Explorer Results Viewe...



# R

```
leisch@galadriel:~/work/tmp
R> n <- 5
R> g <- gl(n, 100, n*100)
R> x <- rnorm(n*100) + sqrt(codes(g))
R> boxplot(split(x,g), col="lavender", notch=TRUE)
R> title(main="Notched Boxplots", xlab="Group", font.main=4, font.lab=1)
R>
R> ctl <- c(4.17,5.58,5.18,6.11,4.50,4.61,5.17,4.53,5.33,5.14)
R> trt <- c(4.81,4.17,4.41,3.59,5.87,3.83,6.03,4.89,4.32,4.69)
R> group <- gl(2,10,20,labels=c("Ctl","Trt"))
R> weight <- c(ctl,trt)
R> anova(lm,D9 <- lm(weight~group))
```

Analysis of Variance Table  
Response: weight

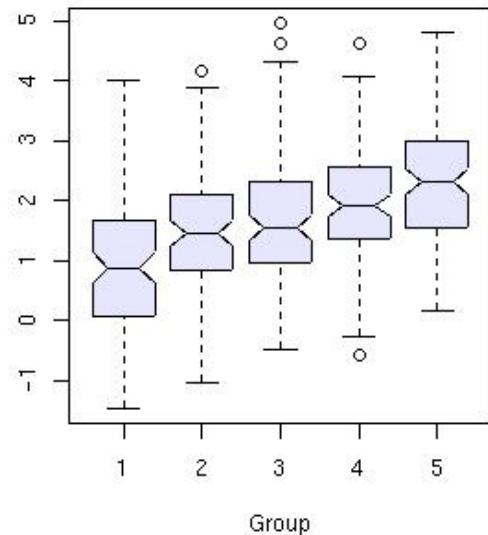
	Df	Sum Sq	Mean Sq	F	Pr(>F)
group	1	0.6882	0.6882	1.419	0.249
Residual	18	8.7293	0.4850		

```
R>
```

```
R> [
```

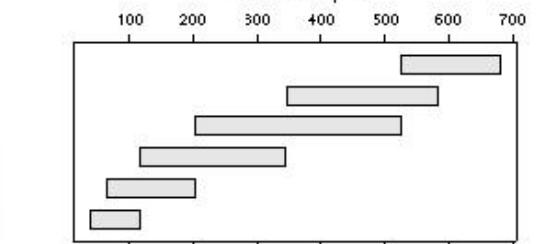
R Graphics: Device 4 (ACTIVE)

Notched Boxplots

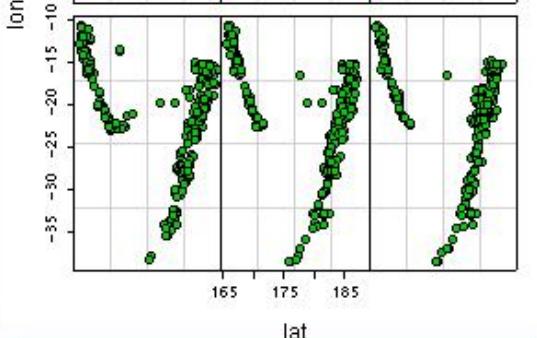
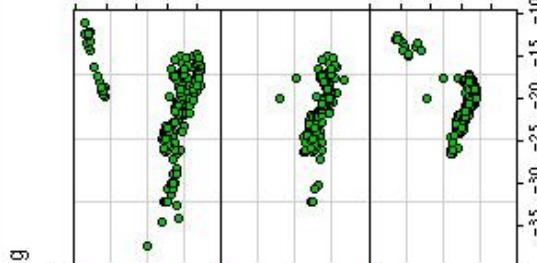


R Graphics: Device 3 (inactive)

Given : depth

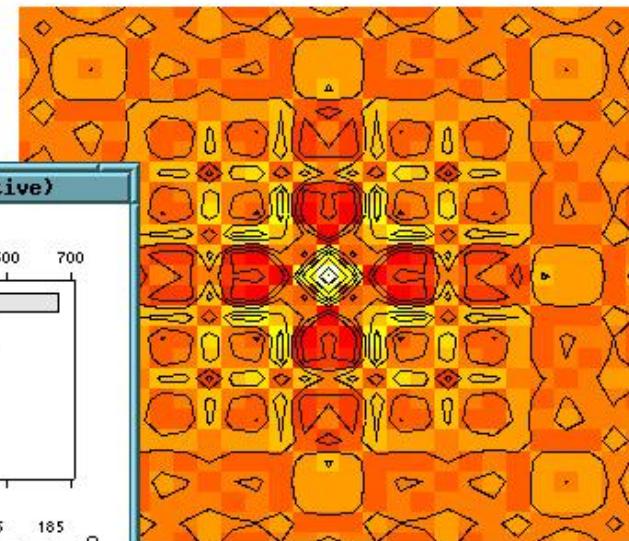


165 175 185 165 175 185



R Graphics: Device 2 (inactive)

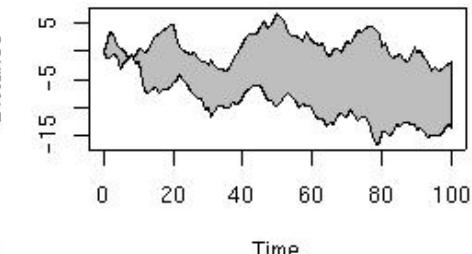
Math can be beautiful ...



$$\cos(r^2)e^{-r/6}$$

R Graphics: Device 5 (inactive)

Distance Between Brownian Motions





R

---

## PRO

- 企業や研究所・大学で広範囲に利用されている
- パワフルで柔軟
- ユーザー層が膨大
- 本やマニュアルが多い
- User Interface も改善が進み使いやすくなっている  
(RStudio)

## COST

- オープンソース
- CON
- あまりユーザーフレンドリーではない
- 学習にはそれなりの努力が必要



# Rとは?

---

- 行列データの演算操作が強力
- データ分析のツール群を大量に実装
- データ分析結果の視覚化ツールが豊富
- プログラム言語としても基本的な機能を完備
- 無料ソフト



# インストール

---

- How to get R:
  - <http://www.r-project.org/>
  - Google: “R”
  - Windows, Linux, Mac OS X, source
  
- Files for this tutorial:
  - 投野研究室ページからダウンロード
  - このスライドのファイル
  - Rコードとサンプル・データ
    - R\_commands.R / T\_Tutorial\_Data.txt



# Installing, Running, and Interacting with R

RStudio interface showing the Environment, History, and Plots panes.

**Environment:**

- Global Environment: myData (63 obs. of 8 variables)
- Values: x=1, y=2, z=3
- Functions: arc (function (x))

**History:**

```
> y  
[1] 2  
> z  
[1] 3  
> (x + y) * z  
[1] 9  
> # Functions:  
> arc <- function(x) 2*asin(sqrt(x))  
> arc(0.5)  
[1] 1.570796  
> ## Reading data from files:  
> myData <- read.table("R_Tutorial_Data.txt", header=TRUE, sep="\t")  
> View(myData)  
> plot(myData)  
>
```

**Plots:**



# Installing, Running, and Interacting with R

---

Rコードを打ちながら解説します：“R\_commands.R”

- データ入力
  - Math(数式)
  - Variables(変数)
  - Arrays(配列)
  - Math on arrays(配列の演算)
  - Functions(関数)
- Getting help(HELP の表示)
- Reading data from files(ファイルの読み込み)
- Selecting subsets of data(部分セットの抽出)



# Installing, Running, and Interacting with R

## Math:

```
> 1 + 1
```

```
[1] 2
```

```
> 1 + 1 * 7
```

```
[1] 8
```

```
> (1 + 1) * 7
```

```
[1] 14
```

## Variables:

```
> x <- 1
```

```
> x
```

```
[1] 1
```

```
> y = 2
```

```
> y
```

```
[1] 2
```

```
> z -> z
```

```
> z
```

```
[1] 3
```

```
> (x + y) * z
```

```
[1] 9
```



# Installing, Running, and Interacting with R

## Arrays:

```
> x <- c(0,1,2,3,4)
> x
[1] 0 1 2 3 4

> y <- 1:5
> y
[1] 1 2 3 4 5

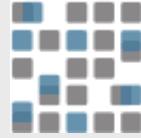
> z <- 1:50
> z
[1]  1   2   3   4   5   6   7   8   9   10  11  12  13  14  15
[16] 16  17  18  19  20  21  22  23  24  25  26  27  28  29  30
[31] 31  32  33  34  35  36  37  38  39  40  41  42  43  44  45
[46] 46  47  48  49  50
```



# Installing, Running, and Interacting with R

## Math on arrays:

```
> x <- c(0,1,2,3,4)
> y <- 1:5
> z <- 1:50
> x + y
[1] 1 3 5 7 9
> x * y
[1] 0 2 6 12 20
> x * z
 [1] 0 2 6 12 20 0 7 16 27 40 0
[12] 12 26 42 60 0 17 36 57 80 0 22
[23] 46 72 100 0 27 56 87 120 0 32 66
[34] 102 140 0 37 76 117 160 0 42 86 132
[45] 180 0 47 96 147 200
```



# Installing, Running, and Interacting with R

## Getting help: ヘルプの表示

```
> help(t.test)
> help.search("standard deviation")
```

The screenshot shows the RStudio interface with the following details:

- Code Editor:** Displays the script `R_commands.R` containing the commands `help(t.test)` and `help.search("standard deviation")`.
- Console:** Shows the output of running the script, including messages about locale settings and warnings from `read.xls`.
- Environment View:** Shows the global environment with two datasets: `myData` and `myData2`.
- Help View:** The right panel displays the documentation for `t.test`, titled "Student's t-Test". It includes sections for Description, Usage, and Arguments.



# Installing, Running, and Interacting with R

- サンプル実験データ:
  - 被験者が新しい task に取り組む:
  - 被験者は2グループ(適性:high vs. low)
    - (“A” and “B”; high and low aptitude learners)
  - トレーニング手法が2通り:
    - (“High variability” and “Low variability”)
  - 4回のトレーニング前の評価テスト
- ファイル: “R\_Tutorial\_Data.txt”





# Installing, Running, and Interacting with R

## ファイルからのデータ読み込み:

```
> myData <- read.table("R_Tutorial_Data.txt",
+ header=TRUE, sep="\t")
> myData
   Condition Group Pre1 Pre2 Pre3 Pre4 Learning
1       Low      A  0.77  0.91  0.24  0.72      0.90
2       Low      A  0.82  0.91  0.62  0.90      0.87
3       Low      A  0.81  0.70  0.43  0.46      0.90
...
61      High     B  0.44  0.41  0.84  0.82      0.29
62      High     B  0.48  0.56  0.83  0.85      0.48
63      High     B  0.61  0.82  0.88  0.95      0.28
```

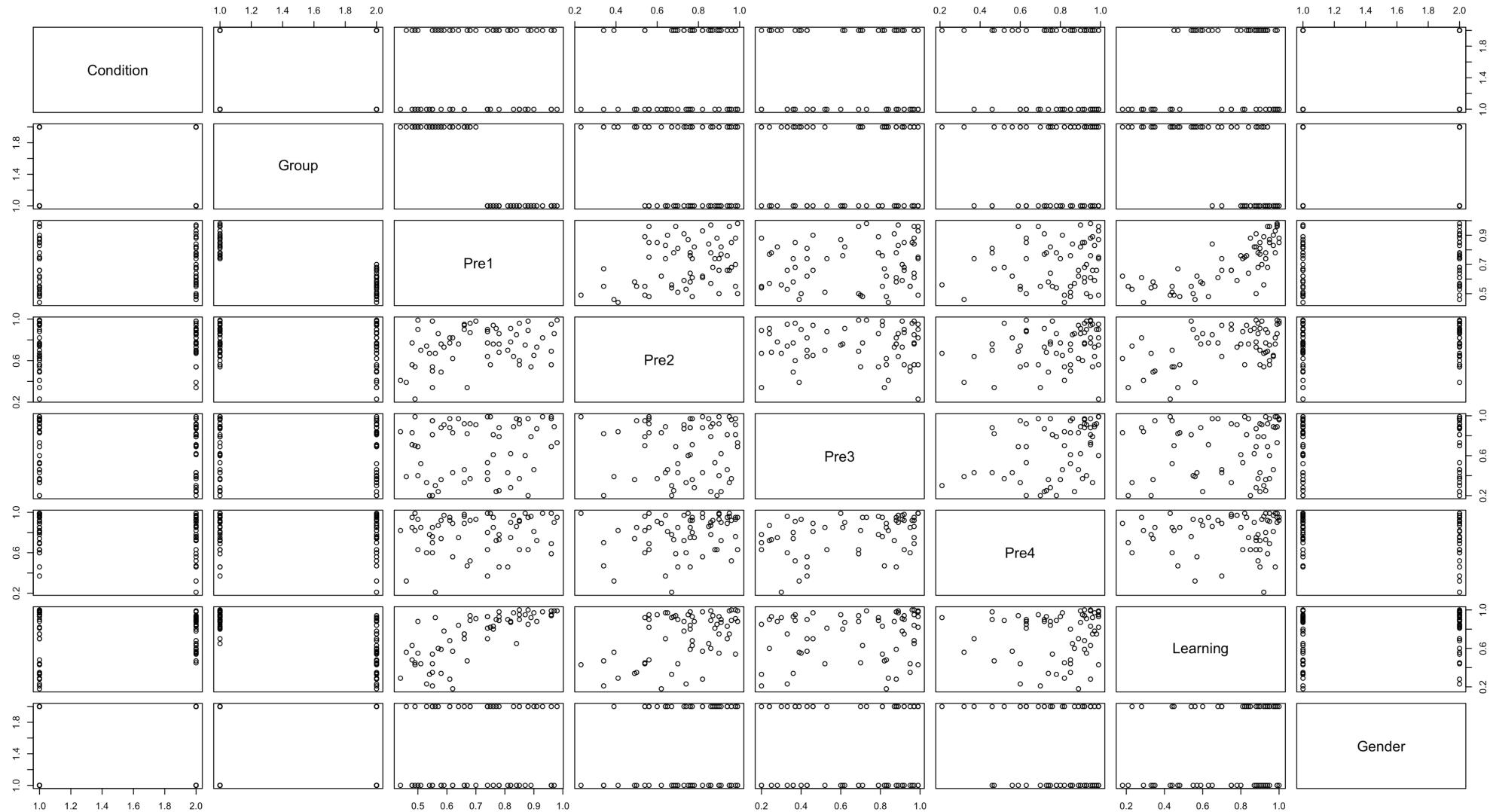
	Condition	Group	Pre1	Pre2	Pre3	Pre4	Learning	Gender
1	Low	A	0.77	0.91	0.24	0.72	0.9	
2	Low	A	0.82	0.91	0.62	0.90	0.87	
3	Low	A	0.81	0.70	0.43	0.46	0.90	
4	Low	A	0.88	0.89	0.2	0.63	0.85	
5	Low	A	0.78	0.68	0.25	0.73	0.93	F
6	Low	A	0.74	0.9	0.99	0.99	0.93	M
7	Low	A	0.78	0.86	0.79	0.78	0.89	F
8	Low	A	0.76	0.76	0.61	0.85	0.8	F
9	Low	A	0.93	0.82	0.99	0.99	0.98	M
10	Low	A	0.82	0.78	0.28	0.75	0.88	F
11	Low	A	0.91	0.73	0.87	0.72	0.88	M
12	Low	B	0.96	0.69	0.69	0.59	0.94	F
13	Low	A	0.97	0.86	0.89	0.9	0.99	F
14	Low	A	0.89	0.54	0.79	0.96	0.92	F
15	Low	A	0.76	0.94	0.81	0.95	0.83	M
16	Low	B	0.84	0.85	0.97	0.86	0.65	F
17	Low	B	0.62	0.82	0.43	0.56	0.57	F
18	Low	B	0.5	0.9	0.4	0.93	0.55	F
19	Low	B						

Condition	Group	Pre1	Pre2	Pre3	Pre4
Low	A	0.77	0.91	0.24	0.72
Low	A	0.82	0.91	0.62	0.9
Low	A	0.81	0.7	0.43	0.46
Low	A	0.88	0.89	0.2	0.63
Low	A	0.78	0.68	0.25	0.73
Low	A	0.74	0.9	0.99	0.99
Low	A	0.78	0.86	0.79	0.78
Low	A	0.76	0.76	0.61	0.85
Low	A	0.93	0.82	0.99	0.99
Low	A	0.82	0.78	0.28	0.75
Low	A	0.91	0.73	0.87	0.72
Low	A	0.96	0.69	0.69	0.59
Low	A	0.97	0.86	0.89	0.9
Low	A	0.89	0.54	0.79	0.96
Low	A	0.76	0.94	0.81	0.95
Low	A	0.84	0.85	0.97	0.86
Low	B	0.62	0.82	0.43	0.56



# Installing, Running, and Interacting with R

```
> plot(myData)
```



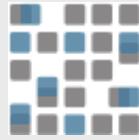


# Installing, Running, and Interacting with R

## Selecting subsets of data:

```
> myData$Learning
[1] 0.90 0.87 0.90 0.85 0.93 0.93 0.89 0.80 0.98
[10] 0.88 0.88 0.94 0.99 0.92 0.83 0.65 0.57 0.55
[19] 0.94 0.68 0.89 0.60 0.63 0.84 0.92 0.56 0.78
[28] 0.54 0.47 0.45 0.59 0.91 0.98 0.82 0.93 0.81
[37] 0.97 0.95 0.70 1.00 0.90 0.99 0.95 0.95 0.97
[46] 1.00 0.99 0.18 0.33 0.88 0.23 0.75 0.21 0.35
[55] 0.70 0.34 0.43 0.75 0.44 0.44 0.29 0.48 0.28

> myData$Learning [myData$Group=="A"]
[1] 0.90 0.87 0.90 0.85 0.93 0.93 0.89 0.80 0.98
[10] 0.88 0.88 0.94 0.99 0.92 0.83 0.65 0.98 0.82
[19] 0.93 0.81 0.97 0.95 0.70 1.00 0.90 0.99 0.95
[28] 0.95 0.97 1.00 0.99
```



# Installing, Running, and Interacting with R

## Selecting subsets of data:

```
> myData$Learning
[1] 0.90 0.87 0.90 0.85 0.93 0.93 0.89 0.80 0.98
[10] 0.88 0.88 0.94 0.99 0.92 0.83 0.65 0.57 0.55
[19] 0.94 0.68 0.89 0.60 0.63 0.84 0.92 0.56 0.78
[28] 0.54 0.47 0.45 0.59 0.91 0.98 0.82 0.93 0.81
[37] 0.97 0.95 0.70 1.00 0.90 0.99 0.95 0.95 0.97
[46] 1.00 0.99 0.18 0.33 0.88 0.23 0.75 0.21 0.35
[55] 0.70 0.34 0.43 0.75 0.44 0.44 0.29 0.48 0.28
> attach(myData)
> Learning
[1] 0.90 0.87 0.90 0.85 0.93 0.93 0.89 0.80 0.98
[10] 0.88 0.88 0.94 0.99 0.92 0.83 0.65 0.57 0.55
[19] 0.94 0.68 0.89 0.60 0.63 0.84 0.92 0.56 0.78
[28] 0.54 0.47 0.45 0.59 0.91 0.98 0.82 0.93 0.81
[37] 0.97 0.95 0.70 1.00 0.90 0.99 0.95 0.95 0.97
[46] 1.00 0.99 0.18 0.33 0.88 0.23 0.75 0.21 0.35
[55] 0.70 0.34 0.43 0.75 0.44 0.44 0.29 0.48 0.28
```



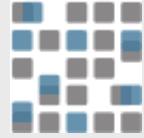
# Installing, Running, and Interacting with R

## Selecting subsets of data:

```
> Learning[Group=="A"]
 [1] 0.90 0.87 0.90 0.85 0.93 0.93 0.89 0.80 0.98
[10] 0.88 0.88 0.94 0.99 0.92 0.83 0.65 0.98 0.82
[19] 0.93 0.81 0.97 0.95 0.70 1.00 0.90 0.99 0.95
[28] 0.95 0.97 1.00 0.99

> Learning[Group!="A"]
 [1] 0.57 0.55 0.94 0.68 0.89 0.60 0.63 0.84 0.92
[10] 0.56 0.78 0.54 0.47 0.45 0.59 0.91 0.18 0.33
[19] 0.88 0.23 0.75 0.21 0.35 0.70 0.34 0.43 0.75
[28] 0.44 0.44 0.29 0.48 0.28

> Condition[Group=="B"&Learning<0.5]
 [1] Low Low High High High High High High High
[10] High High High High High
Levels: High Low
```



# Useful functions for text processing

---

```
>substr("internationalization", 6, 13)
```

```
[1] "national "
```

```
>example <- c("I", "do", , "not" "know")
```

```
>nchar(example)
```

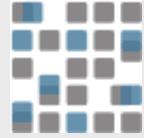
```
[1] 1 2 3 4
```

```
>tolower(example)
```

```
[1] "i "... "do"... "not"..."know"
```

```
>toupper(example)
```

```
[1] "I"..."DO"... "NOT"..."KNOW"
```



# Useful functions for text processing

```
>chartr("o", "x", example)  
[1] "l"   "dx"  "nxt" "knxw"
```

```
>example.2 <- "I do not know"  
>strsplit(example.2, " ") # " " はスペース  
[[1]]  
[1] "l"   "do"  "not" "know"  
# ""と何もいれないと文字単位で分割  
# data 形式的には vector ではなく, list の状態
```



```
>unlist(strsplit(example.2, " "))  
[1] "I"  "do" "not" "know"
```

```
>text <- c("This is a first example sentence.",  
"And this is a second example sentence.")
```

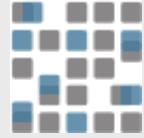
```
>grep("second", text)
```

```
[1] 2
```

# grep はマッチした位置を返す

```
>grep("second", text, value=T)
```

```
[1] "And this is ..."
```



# Package gsubfn (1)

---

```
>library(gsubfn)
```

```
>text <- c("This is a first  
example sentence.", "And  
this is a second example  
sentence.")
```

```
>strapply(text, "first")
```

```
[[1]]
```

```
[1] "first"
```

```
[[2]]
```

```
NULL
```

```
>strapply(text, "is")
```

```
[[1]]
```

```
[1] "is" "is"
```

```
[[2]]
```

```
[1] "is" "is"
```



# Package gsubfn (2)

```
gsub(pattern, replacement, x, ignore.case=F)
```

```
>gsub("a", "the", text)
```

```
[1] "This is the first exthemple sentence."
```

```
[2] "And this is the second exthemple sentence."
```

```
>gsub(" a ", " the ", text)
```

```
[1] "This is the first example sentence."
```

```
[2] "And this is the second example sentence."
```

```
>text.2 <- gsub(" a ", " the ", text) #テキストに格納
```



# Statistics and Data Analysis

---

- Parametric Tests パラメトリック・テスト
  - Independent sample t-tests
  - Paired sample t-tests
  - One sample t-tests
  - Correlation
- Nonparametric tests ノンパラメトリック・テスト
  - Shapiro-Wilks test for normality
  - Wilcoxon signed-rank test (Mann-Whitney U)
  - Chi square test
- Linear Models and ANOVA 線形モデル



# Basic parametric inferential statistics

## Independent sample t-tests:

```
> t.test(Pre2[Group=="A"],  
+ Pre2[Group=="B"],  
+ paired=FALSE)
```

} Pre2テストの AとBグループの比較

Welch Two Sample t-test

```
data: Learning[Group == "A"] and Learning[Group == "B"]  
t = 1.6117, df = 53.275, p-value = 0.1129  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-0.0179193 0.1645725  
sample estimates:  
mean of x mean of y  
0.7764516 0.7031250
```



# Basic parametric inferential statistics

## Independent sample t-tests:独立2標本 t検定

```
> t.test(Pre2[Group=="A"],  
+ Pre2[Group=="B"],  
+ paired=FALSE,  
+ var.equal=TRUE)
```

Welch Two Sample t-test

```
data: Learning[Group == "A"] and Learning[Group == "B"]  
t = 1.601, df = 61, p-value = 0.1145  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-0.0179193 0.1645725  
sample estimates:  
mean of x mean of y  
0.7764516 0.7031250
```



# Basic parametric inferential statistics

## Independent sample t-tests:独立2標本 t検定

```
> t.test(Pre2[Group=="A"],  
+ Pre2[Group=="B"],  
+ paired=FALSE,  
+ var.equal=TRUE,  
+ alternative="greater")
```

Welch Two Sample t-test

```
data: Learning[Group == "A"] and Learning[Group == "B"]  
t = 1.601, df = 61, p-value = 0.5727  
alternative hypothesis: true difference in means is greater than 0  
95 percent confidence interval:  
-0.003169388 Inf  
sample estimates:  
mean of x mean of y  
0.7764516 0.7031250
```



# Basic parametric inferential statistics

## Paired sample t-test:対応のある2標本 t検定

```
> t.test(Pre4[Group=="A"],  
+ Pre3[Group=="A"],  
+ paired=TRUE)
```

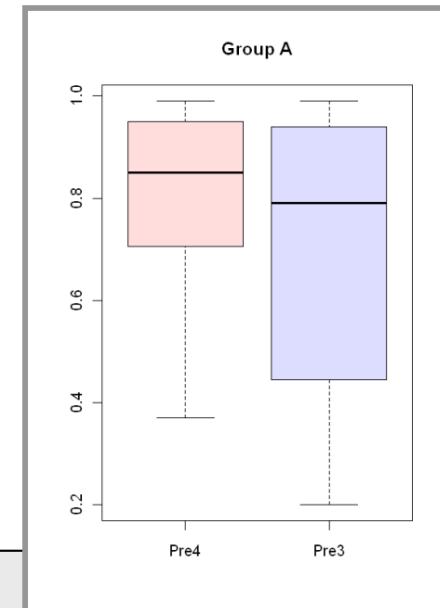
}

グループAの3回目と4回目の比較

Paired t-test

```
data: Pre4[Group == "A"] and Pre3[Group == "A"]  
t = 2.4054, df = 30, p-value = 0.02253  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 0.01641059 0.20100876  
sample estimates:  
mean of the differences  
 0.1087097
```

```
> boxplot(Pre4[Group=="A"],  
+ Pre3[Group=="A"],  
+ col=c("#ffdddd", "#ddffff"),  
+ names=c("Pre4", "Pre3"), main="Group A")
```





# Basic parametric inferential statistics

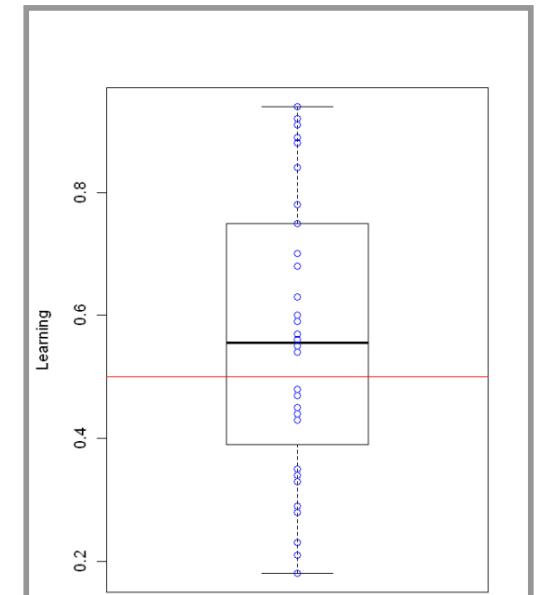
## One sample t-test: 母平均との差の検定

```
> t.test(Learning[Group=="B"], mu=0.5, alternative="greater")
```

```
One Sample t-test
```

```
data: Learning[Group == "B"]
t = 1.5595, df = 31, p-value = 0.06452
alternative hypothesis: true mean is greater than 0.5
95 percent confidence interval:
 0.4945469      Inf
sample estimates:
mean of x
 0.5625
```

```
> boxplot(Learning[Group=="B"],
+ names="Group B", ylab="Learning")
> lines(c(0,2), c(0.5, 0.5), col="red")
> points(c(rep(1,length(Learning[Group=="B"]))),
+ Learning[Group=="B"], pch=21, col="blue")
```





# Basic parametric inferential statistics

## Correlation: 相關分析

```
> cor.test(Pre1, Learning, method="pearson")
```

Pearson's product-moment correlation

data: Pre1 and Learning

t = 9.2461, df = 61, p-value = 3.275e-13

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

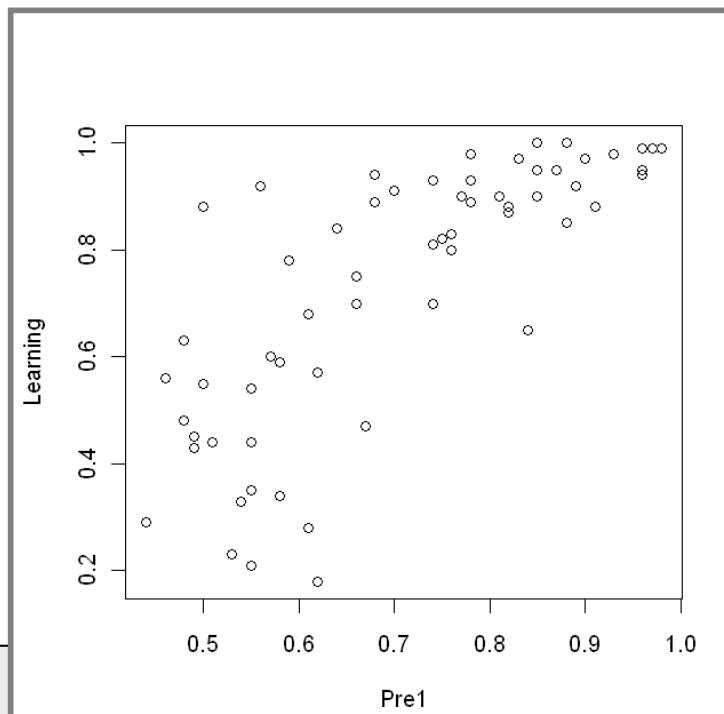
0.6366698 0.8506815

sample estimates:

cor

0.7639292

```
> plot(Pre1, Learning)
```





# Basic parametric inferential statistics

## Correlation (fancier plot example):

```
> cor.test(Pre1, Learning, method="pearson")
```

Pearson's product-moment correlation

data: Pre1 and Learning

t = 9.2461, df = 61, p-value = 3.275e-13

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

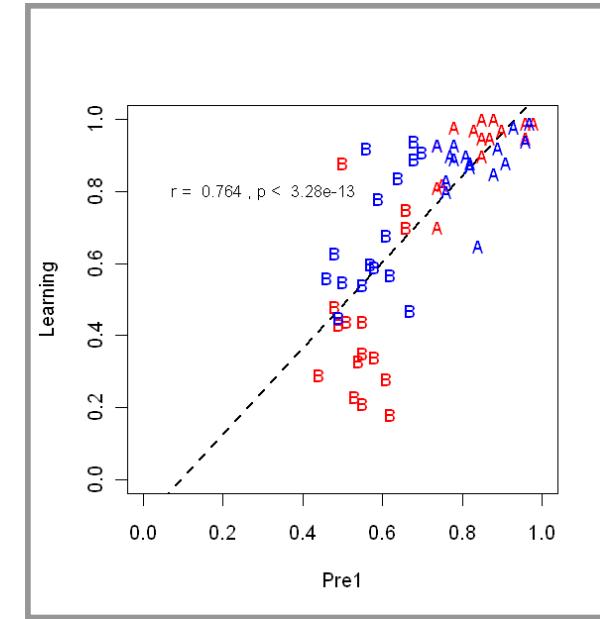
0.6366698 0.8506815

sample estimates:

COR

0.7639292

```
> plot(Learning~Pre1, ylim=c(0,1), xlim=c(0,1), ylab="Learning", xlab="Pre1", type="n")
> abline(lm(Learning~Pre1), col="black", lty=2, lwd=2)
> points(Learning[Group=="A"&Condition=="High"]~Pre1[Group=="A"&Condition=="High"],
+ pch=65, col="red", cex=0.9)
> points(Learning[Group=="A"&Condition=="Low"]~Pre1[Group=="A"&Condition=="Low"],
+ pch=65, col="blue", cex=0.9)
> points(Learning[Group=="B"&Condition=="High"]~Pre1[Group=="B"&Condition=="High"],
+ pch=66, col="red", cex=0.9)
> points(Learning[Group=="B"&Condition=="Low"]~Pre1[Group=="B"&Condition=="Low"],
+ pch=66, col="blue", cex=0.9)
> legend(2.5,1.0, c("LV Training", "HV Training"), pch=c(19), col=c("blue","red"), bty="y")
> yCor <- cor.test(Pre1, Learning, method="pearson")
> text(0.3,0.8, paste("r = ", format(myCor$estimate,digits=3), ", p < ", format(myCor$p.value,digits=3)), cex=0.8)
```





# Statistics and Data Analysis

## Are my data normally distributed? 正規性の検定

```
> t.test(Learning[Condition=="High"&Group=="A"],  
+ Learning[Condition=="Low"&Group=="A"])
```

Welch Two Sample t-test

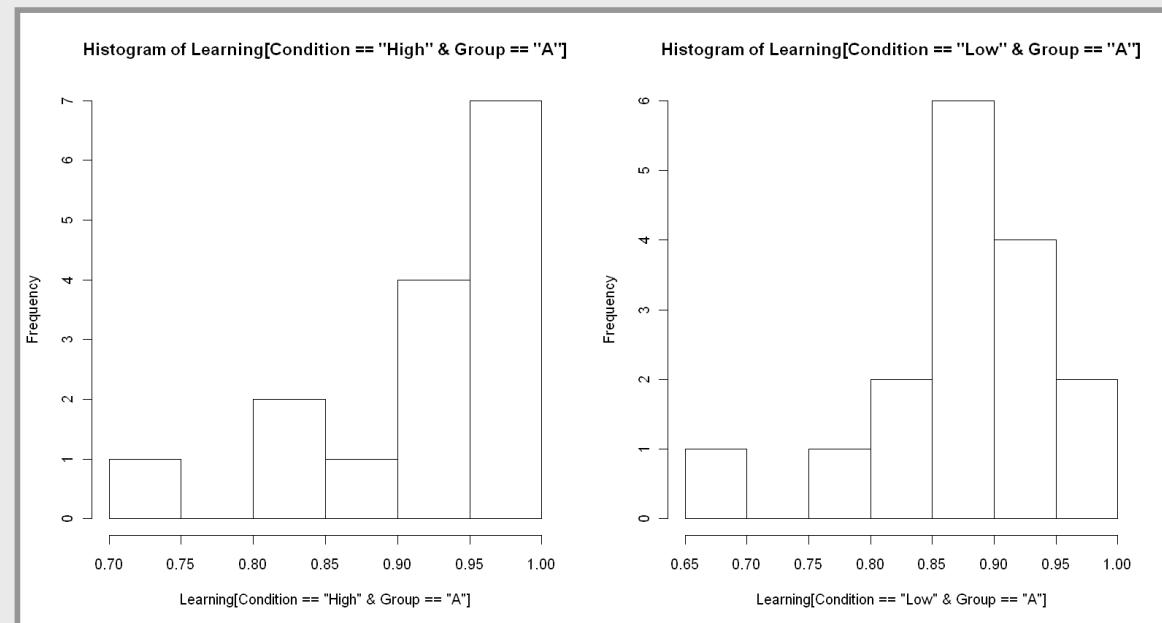
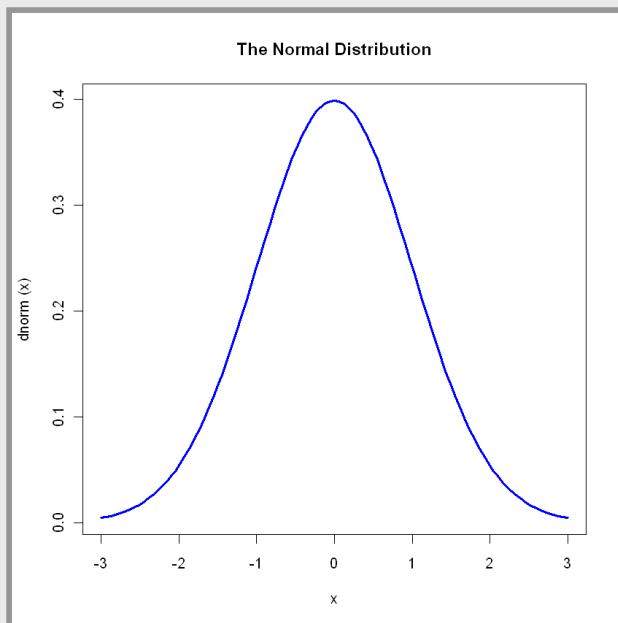
```
data: Learning[Condition == "High" & Group == "A"] and  
Learning[Condition == "Low" & Group == "A"]  
t = 1.457, df = 28.422, p-value = 0.1561  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-0.01764821 0.10481488  
sample estimates:  
mean of x mean of y  
0.9273333 0.8837500
```



# Statistics and Data Analysis

## Are my data normally distributed?

```
> plot(dnorm, -3, 3, col="blue", lwd=3, main="The Normal Distribution")
> par(mfrow=c(1, 2))
> hist(Learning[Condition=="High"&Group=="A"])
> hist(Learning[Condition=="Low"&Group=="A"])
```





# Statistics and Data Analysis

## Are my data normally distributed?

```
> shapiro.test(Learning[Condition=="High"&Group=="A"])
```

Shapiro-Wilk normality test

```
data: Learning[Condition == "High" & Group == "A"]
W = 0.7858, p-value = 0.002431
```

```
> shapiro.test(Learning[Condition=="Low"&Group=="A"])
```

Shapiro-Wilk normality test

```
data: Learning[Condition == "Low" & Group == "A"]
W = 0.8689, p-value = 0.02614
```



# Basic nonparametric inferential statistics

## Wilcoxon signed-rank / Mann-Whitney U tests:

```
> wilcox.test(Learning[Condition=="High"&Group=="A"],  
+ Learning[Condition=="Low"&Group=="A"],  
+ exact=FALSE,  
+ paired=FALSE)  
  
Wilcoxon rank sum test with continuity correction  
  
data: Learning[Condition == "High" & Group == "A"] and  
Learning[Condition == "Low" & Group == "A"]  
W = 173.5, p-value = 0.03580  
alternative hypothesis: true location shift is not equal to 0
```



# Basic nonparametric inferential statistics

## Chi square test: カイ<sup>2</sup>乗検定

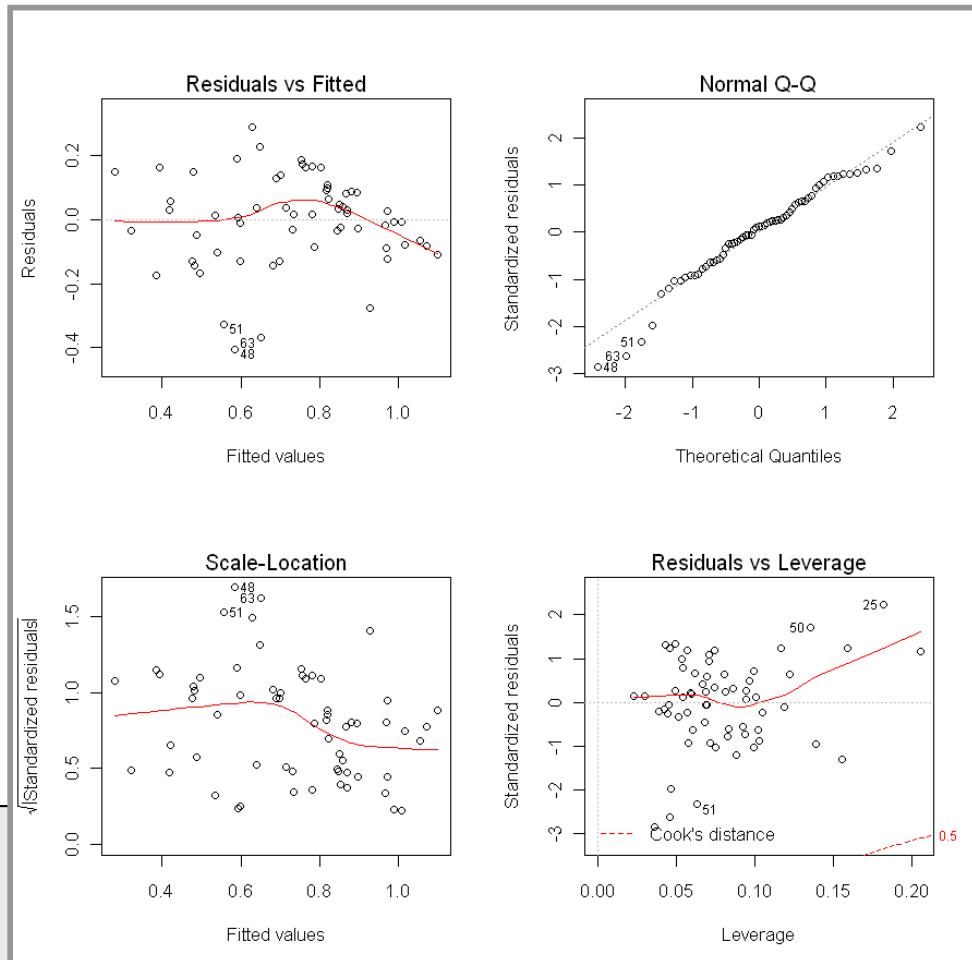
```
> x <- matrix(c(  
+ length(Learning[Group=="A"&Condition=="High"&Gender=="F"] ),  
+ length(Learning[Group=="A"&Condition=="Low"&Gender=="F"] ),  
+ length(Learning[Group=="B"&Condition=="High"&Gender=="F"] ),  
+ length(Learning[Group=="B"&Condition=="Low"&Gender=="F"] )),  
+ ncol=2)  
> x  
      [,1] [,2]  
[1,]    4   12  
[2,]   10    7  
> chisq.test(x)  
  
Pearson's Chi-squared test with Yates' continuity correction  
  
data: x  
X-squared = 2.5999, df = 1, p-value = 0.1069
```



# Linear models and ANOVA

## Linear models: 線形モデル

```
> myModel <- lm(Learning ~ Pre1 + Pre2 + Pre3 + Pre4)  
> par(mfrow=c(2,2))  
> plot(myModel)
```





# Linear models and ANOVA

## Linear models: 線形モデル

```
> summary(myModel)
```

Call:

```
lm(formula = Learning ~ Pre1 + Pre2 + Pre3 + Pre4)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.40518	-0.08460	0.01707	0.09170	0.29074

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.22037	0.11536	-1.910	0.061055 .
Pre1	1.05299	0.12636	8.333	1.70e-11 ***
Pre2	0.41298	0.10926	3.780	0.000373 ***
Pre3	0.07339	0.07653	0.959	0.341541
Pre4	-0.18457	0.11318	-1.631	0.108369

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1447 on 58 degrees of freedom

Multiple R-squared: 0.6677, Adjusted R-squared: 0.6448

F-statistic: 29.14 on 4 and 58 DF, p-value: 2.710e-13



# Linear models and ANOVA

## Linear models: 線形モデル

```
> step(myModel, direction="backward")
```

Start: AIC=-238.8

Learning ~ Pre1 + Pre2 + Pre3 + Pre4

	Df	Sum of Sq	RSS	AIC
- Pre3	1	0.01925	1.2332	-239.81
<none>			1.2140	-238.80
- Pre4	1	0.05566	1.2696	-237.98
- Pre2	1	0.29902	1.5130	-226.93
- Pre1	1	1.45347	2.6675	-191.21

Step: AIC=-239.81

Learning ~ Pre1 + Pre2 + Pre4

	Df	Sum of Sq	RSS	AIC
- Pre4	1	0.03810	1.2713	-239.89
<none>			1.2332	-239.81
- Pre2	1	0.28225	1.5155	-228.83
- Pre1	1	1.54780	2.7810	-190.58

...

...

Step: AIC=-239.89

Learning ~ Pre1 + Pre2

	Df	Sum of Sq	RSS	AIC
<none>			1.2713	-239.89
- Pre2	1	0.24997	1.5213	-230.59
- Pre1	1	1.52516	2.7965	-192.23

Call:

```
lm(formula = Learning ~ Pre1 + Pre2)
```

Coefficients:

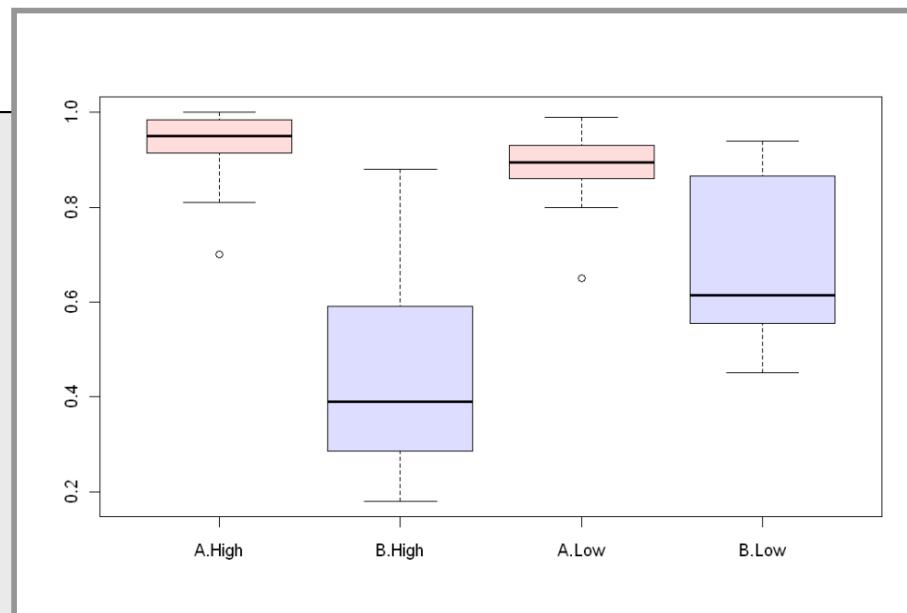
(Intercept)	Pre1	Pre2
-0.2864	1.0629	0.3627



# Linear models and ANOVA

## ANOVA: 分散分析

```
> myANOVA <- aov(Learning~Group*Condition)
> summary(myANOVA)
      Df Sum Sq Mean Sq F value    Pr(>F)
Group          1 1.8454 1.84537 81.7106 9.822e-13 ***
Condition       1 0.1591 0.15910  7.0448 0.0102017 *
Group:Condition 1 0.3164 0.31640 14.0100 0.0004144 ***
Residuals      59 1.3325 0.02258
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> boxplot(Learning~Group*Condition, col=c("#ffdddd", "#ddffff"))
```

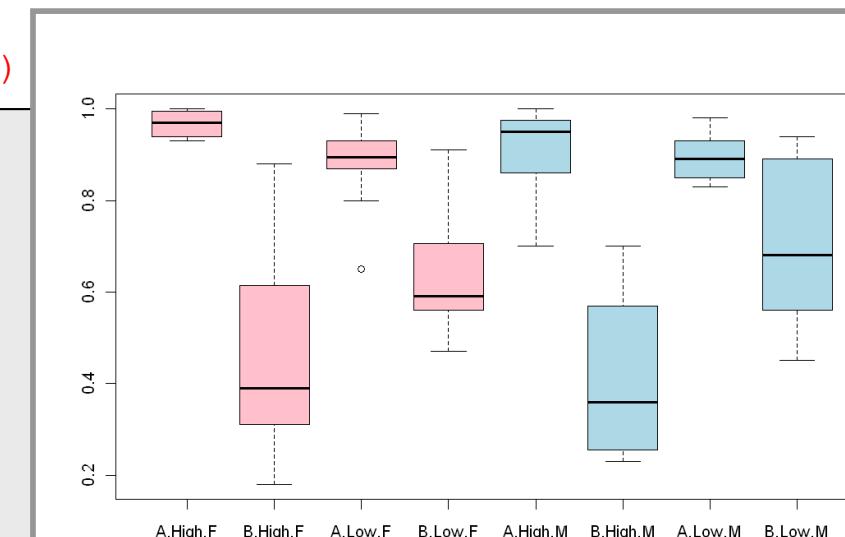


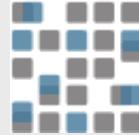


# Linear models and ANOVA

## ANOVA: 分散分析

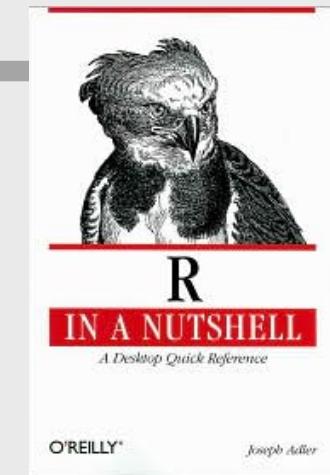
```
> myANOVA2 <- aov(Learning~Group*Condition+Gender)
> summary(myANOVA2)
      Df  Sum Sq Mean Sq F value    Pr(>F)
Group          1 1.84537 1.84537 80.3440 1.523e-12 ***
Condition      1 0.15910 0.15910  6.9270  0.010861 *
Gender         1 0.04292 0.04292  1.8688  0.176886
Group:Condition 1 0.27378 0.27378 11.9201  0.001043 **
Residuals     58 1.33216 0.02297
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> boxplot(Learning~Group*Condition+Gender,
+ col=c(rep("pink",4),rep("light blue",4)))
```





# その他の R 関連サイト

- R wiki:  
<http://rwiki.sciviews.org/doku.php>
- R graph gallery:  
<http://addictedito.free.fr/graphiques/thumbs.php>
- Kickstarting R:  
<http://cran.r-project.org/doc/contrib/Lemon-kickstart/>



ISBN: 9780596801700

R plot residuals - Google Search - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.google.com/search?hl=en&safe=off&client=firefox-a&hs=...

R plot of the day

Web Images Videos Maps News Shopping Gmail more ▾

Google

R plot residuals

About 2,740,000 results (0.24 seconds) Advanced search

**R\_Plot Residuals**  
plot.residuals (multi), R Documentation. **Plot Residuals**. Description. **plot residuals** is used for plotting residuals from models obtained from dynamic ...  
rss.acs.unt.edu/Rdoc/library/multiln/html/plot.residuals.html - Cached

**R\_Residual plot for lm or nls objects.**  
Constructs a residual plot for lm or nls objects. Different symbols for different groups can be added to the plot if an indicator variable regression is ...  
www.Rforge.net/doc/packages/NCStats/residual.plot.html - Cached - Similar

**R\_Residuals from a Logistic Regression Model Fit**  
The plot.lm.partial function computes partial residuals for a series of binary ... Under R, pl="loess" uses lowess and does not provide confidence bands: ...  
lib.stat.cmu.edu/S/Harrel/Help/Design.../residuals.lm.html - Cached - Similar

**R help archive [R] plotting residuals**  
Nov 4, 2006 ... [R] plotting residuals. This message : [ Message body ] | More options ... Does anyone know how to obtain a plot of residuals by predicted ...  
tolstoy.newcastle.edu.au/R/e2/help/06/11/4466.html - Cached - Similar

**poc Plots of Residuals**  
File Format: Microsoft Word - View as HTML  
Notice that the R has gone up a lot and is now significant, and the residuals plot looks fine.  
Let us have a look at the regression line ...  
core.ecu.edu/psyc/wuenschl/spss/Residual-Plots-SPSS.doc

**PDF R FUNCTIONS FOR REGRESSION ANALYSIS**  
File Format: PDF/Adobe Acrobat - Quick View  
Oct 14, 2005 ... qqnorm.gls: Normal Plot of Residuals from a gls Object (nlme) .... rsq.rpart: Plots the Approximate R-Square for the Different Splits (rpart ...  
cran.r-project.org/doc/contrib/Ricci-regression.pdf - Similar

**Normal Probability Plot of Residuals | R Tutorial**  
Jan 4, 2010 ... An R tutorial on the normal probability plot for the residual of a simple linear

Done

R Graph Gallery :: thumbnails gallery - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://addictedito.free.fr/graphiques/thumbs.php

R start [R.Wiki]

read my blog RELATED SITES: R-project | CRAN | Bioconductor | R Movies Gallery | R Wiki | SGDV | Graphics CTV

Graph Gallery : Enhance your data visualization with R

Home Browse Related Source code Graphics List Thumbnails

Image scatter plot matrix

The ipairs function produces an image scatter plot matrix of large datasets where the colors encode[...]

http://addictedito.free.fr/graphiques/RGraphGallery.php?graph=159