

R入門

第二回 グラフィクス・データセット

横浜国立大学
酒井信介

数値計算の妥当性把握のためにグラフィックス機能は重要

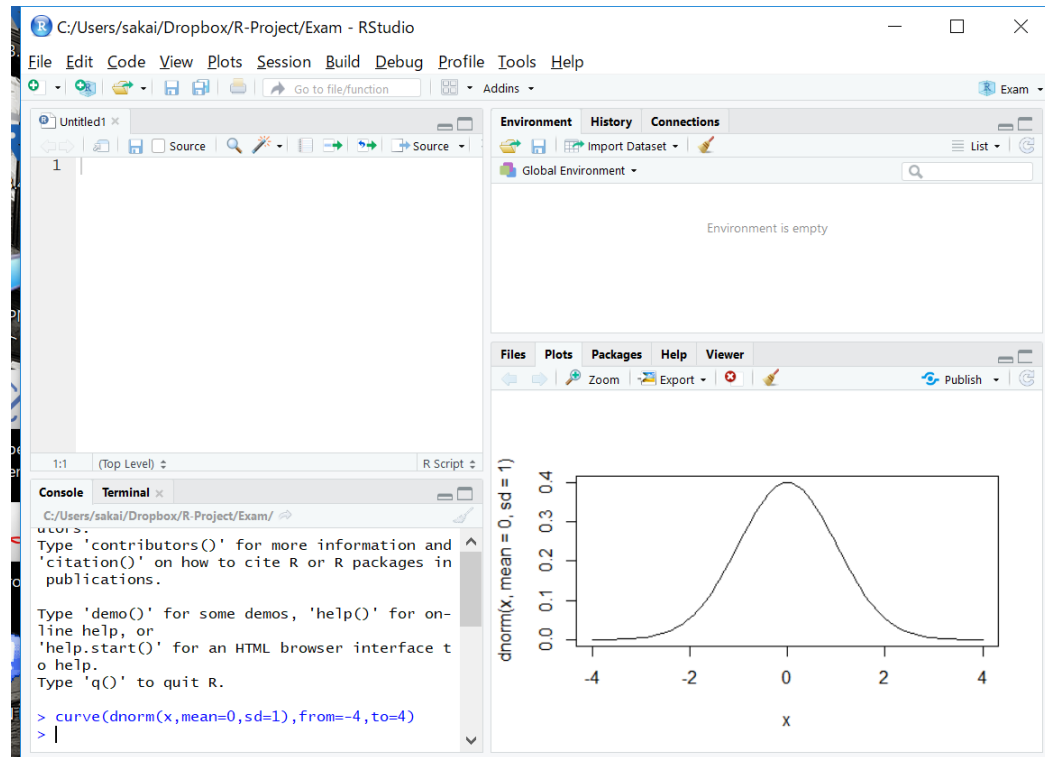
- Rには極めて強力なグラフィックス機能がある
- 論文使用にも十分耐えられる(これだけの目的に使ってもよい)
- 現象の理解にも極めて有用
- プログラムのバグ取りにも利用できる
- 新たなアルゴリズム適用には、まず概要を理解する必要がある
- パッケージのexampleを実行してみればよい

使用例

正規分布の確率密度関数はどうのような形状か？

```
curve(dnorm(x,mean=0,sd=1),from=-4,to=4)
```

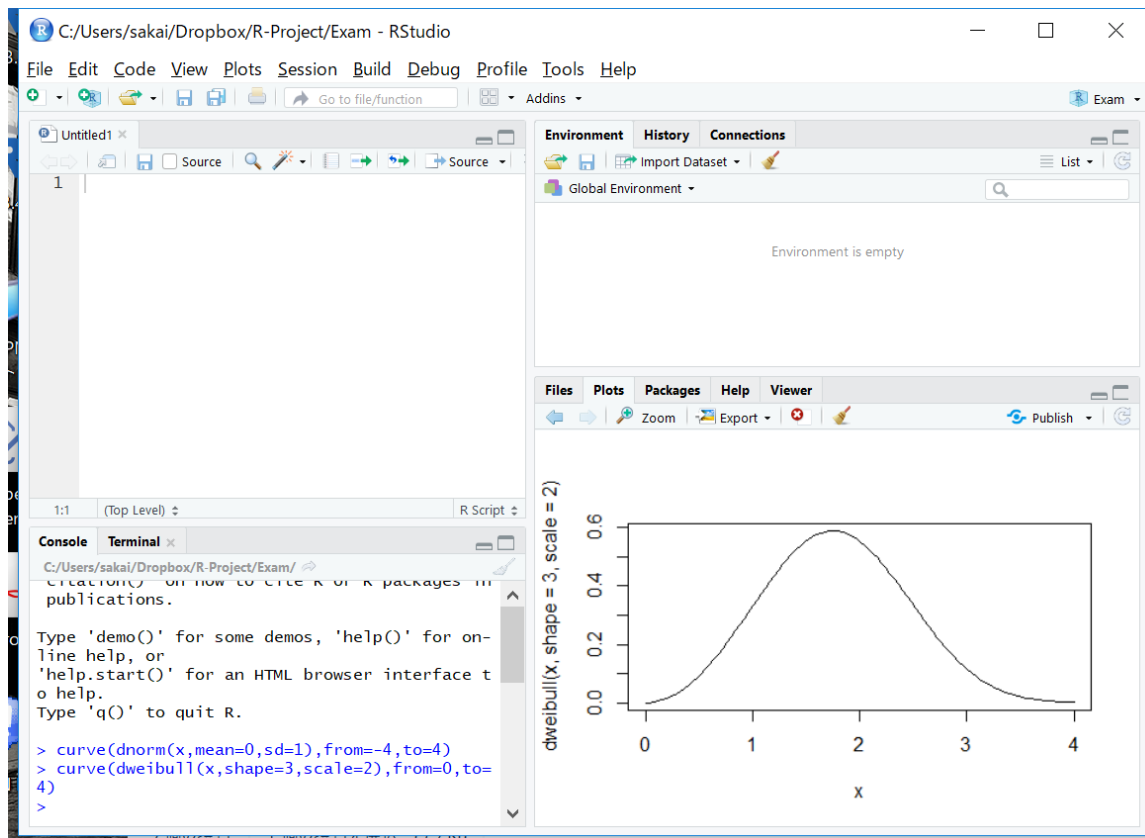
curve()も, dnorm()も標準でもっている関数なので, いきなり実行できる.



Weibull分布の例

それでは、形状母数3, 尺度母数2のワイブル分布は?

`curve(dweibull(x,shape=3,scale=2),from=0,to=4)`

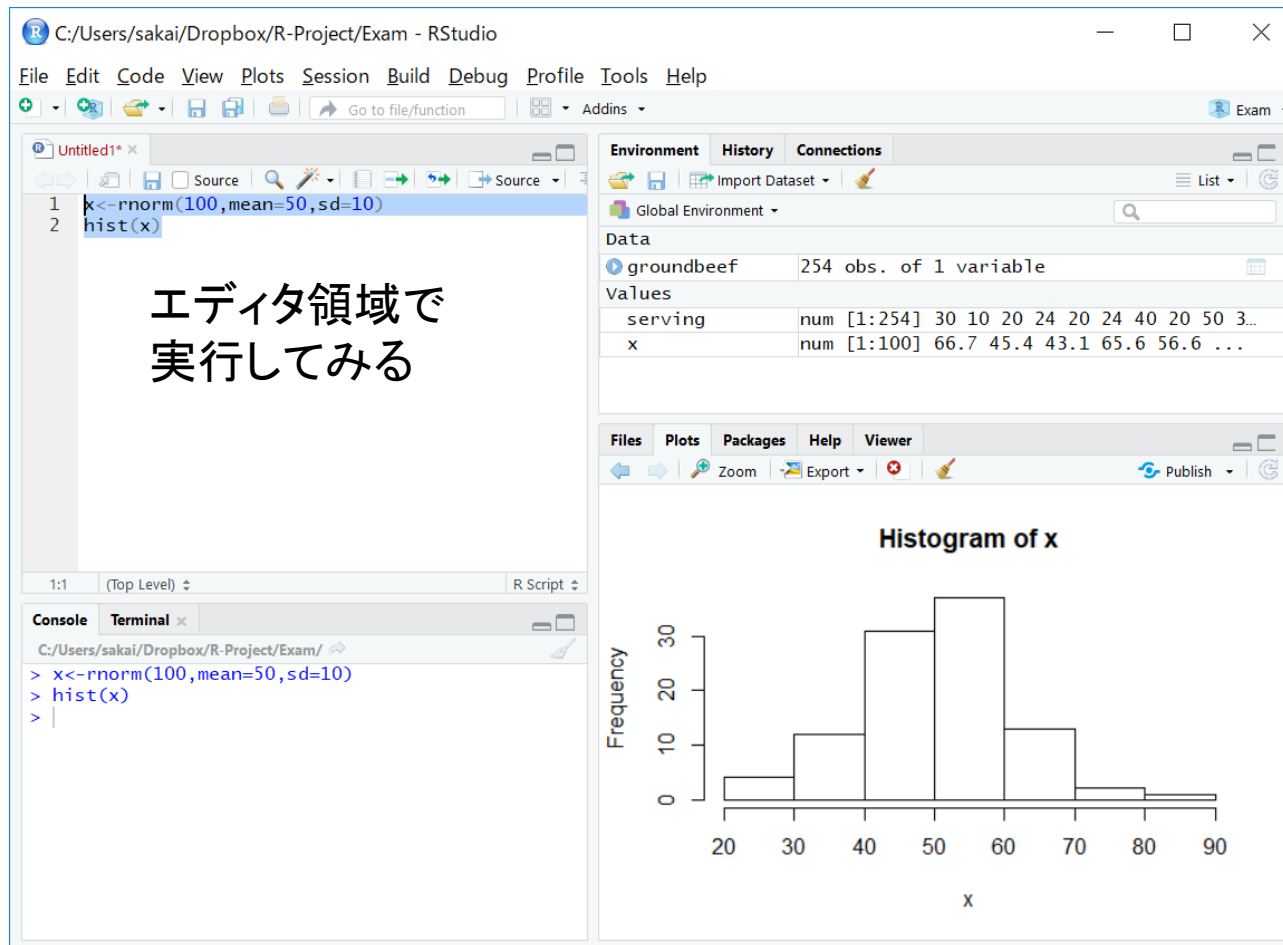


サンプルデータのヒストグラム表示

```
x<- rnorm(100,mean=50,sd=10)
```

```
hist(x)
```

エディタ領域で
実行してみる

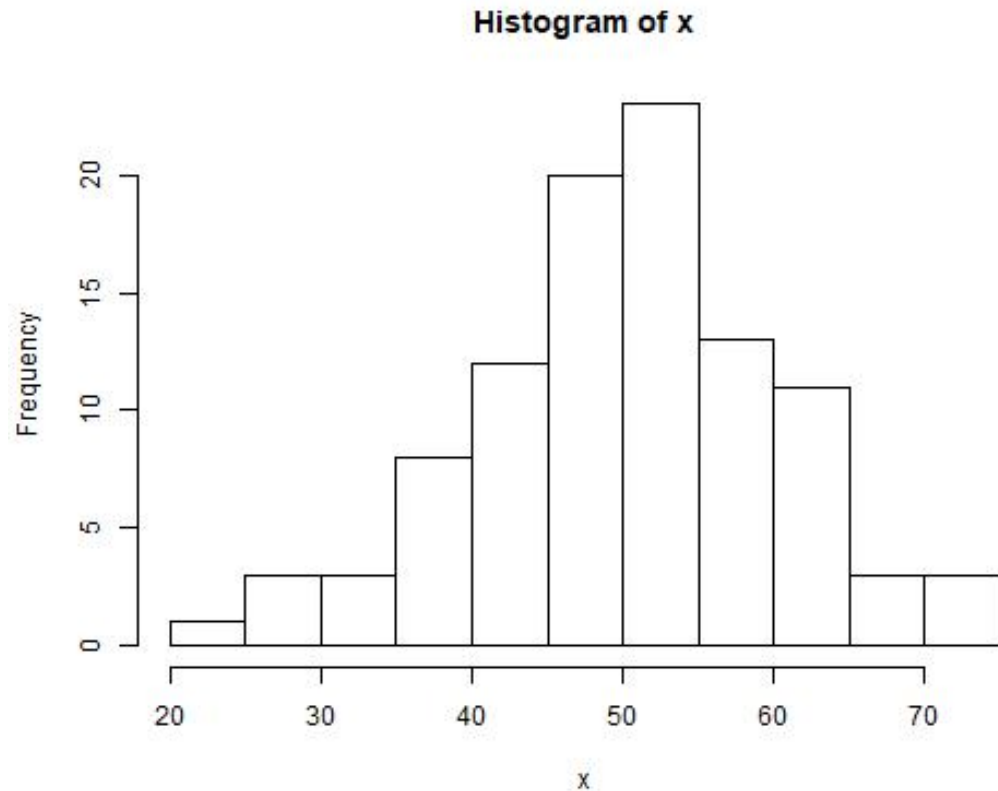


画像として保存するためには

```
x<-rnorm(100,mean=50,sd=10)  
jpeg("hist.jpg",width=500,height=400)  
hist(x)  
dev.off()
```

jpeg画像の出力

pdf,png,postscript
などの形式も可能



想定される問題

今サンプルデータが得られたとしたとき, この分布がどのような分布に従っているのか調べたい. このような問題には, 研究上たびたび遭遇する. Rでこの問題を扱うためのノーハウをまとめる. まず, Googleで探す.

R 分布 推定

こんなサイトが見つかった

[2017-01-16](#)

[単変量の分布母数の推定からあれこれ: fitdistrplus パッケージが便利](#)

どうやらfitdistrplusというCRANから提供されているらしい. 早速インストールしてみる. インストールは1回だけやればよい.

```
>install.packages("fitdistrplus")
```

以後は使う度ごとに, 以下のlibraryコマンドを実行する.

```
>library(fitdistrplus)
```

このパッケージにはどのようなコマンドがあるかを知りたいければ

```
>help(package=fitdistrplus)
```

パッケージの魅力

- 統計解析に必要なとなる分析機能を有するパッケージは大半が提供されている
- 自分でプログラムすると多くの労力が必要
- Rの場合, ネット上で必要とするパッケージの場所に到達できれば, たった一行で解決できることも多い!

Untitled1 x

Source

1

1:1

(Top Level) ↕

R Script ↕

Console Terminal x

C:/Users/sakai/Dropbox/R-Project/Exam/ ↗

> help(package=fitdistrplus)

>

これが使えそうだ、内容を調べる
ためにクリック

Environment

History

Connections

Import Dataset

Global Environment

Environment is empty

Files

Plots

Packages

Help

Viewer

R: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data

Find in Topic

[genscomp](#)

(for non-censored data)

[descdist](#)

Description of an empirical distribution for non-censored data

[detectbound](#)

Detect bounds for density function

[endosulfan](#)

Species Sensitivity Distribution (SSD) for endosulfan

[fitdist](#)

Fit of univariate distributions to non-censored data

[fitdistcens](#)

Fitting of univariate distributions to censored data

[fitdistrplus](#)

Overview of the 'fitdistrplus' package

[fluazinam](#)

Species-Sensitivity Distribution (SSD) for Fluazinam

[gofstat](#)

Goodness-of-fit statistics

[graphcomp](#)

Graphical comparison of multiple fitted distributions (for non-censored data)

[groundbeef](#)

Ground beef serving size data set

[llcurve](#)

(Log)likelihood surfaces or (log)likelihood curves

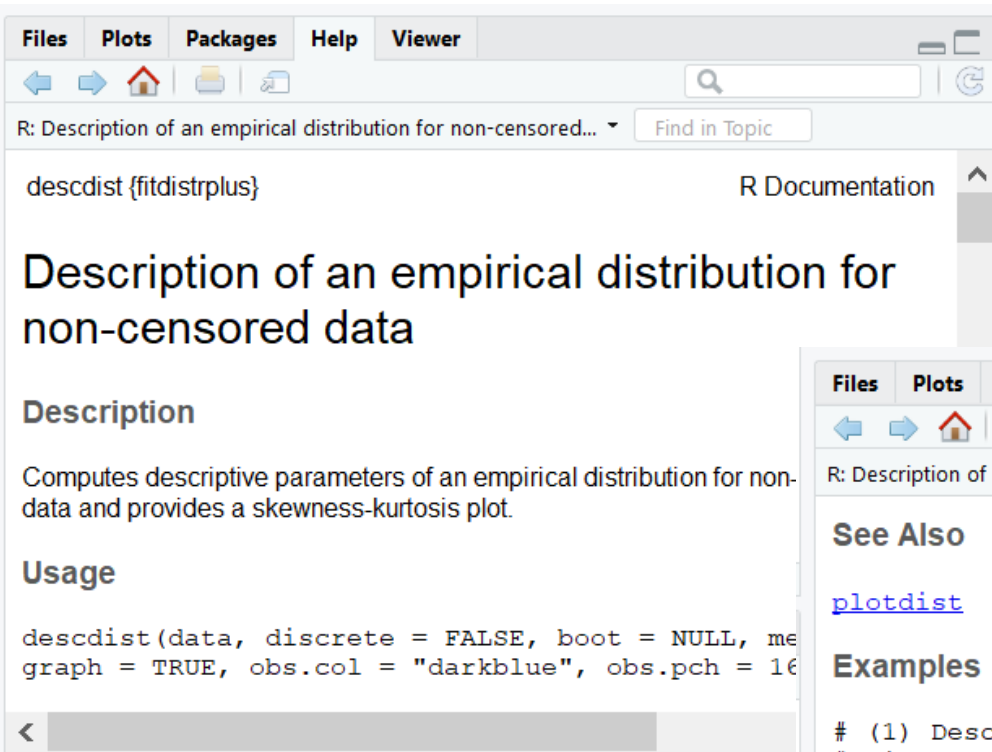
[llplot](#)

(Log)likelihood plot for a fit using maximum likelihood

[llsurface](#)

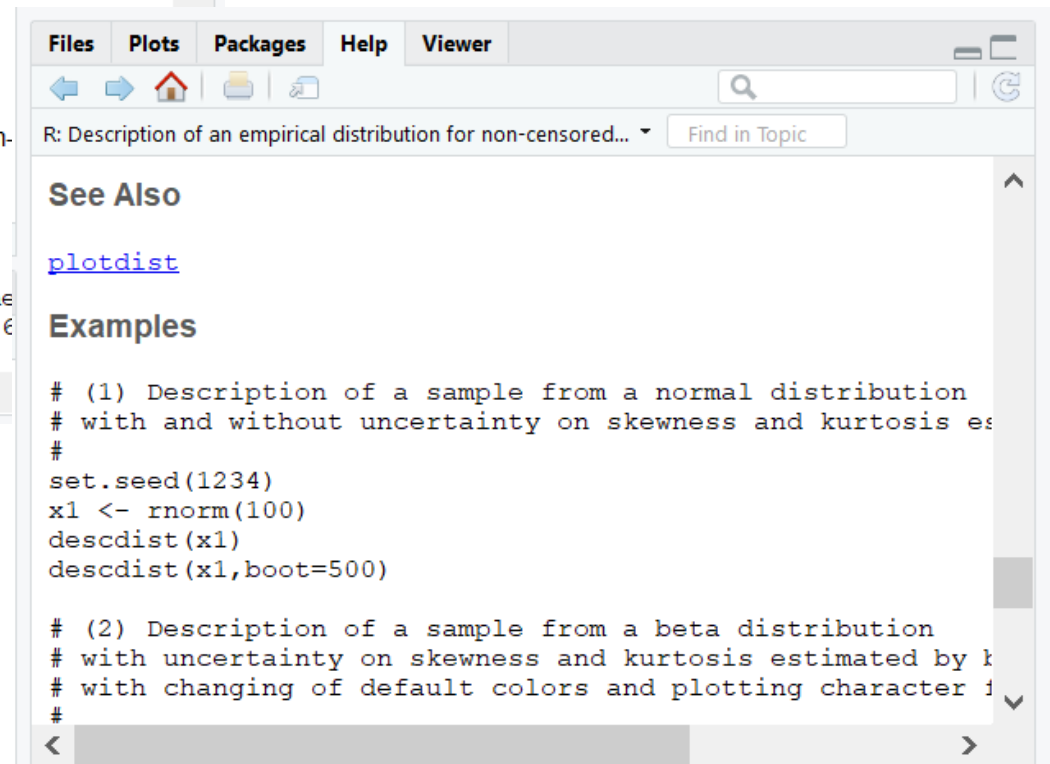
(Log)likelihood surfaces or (log)likelihood curves

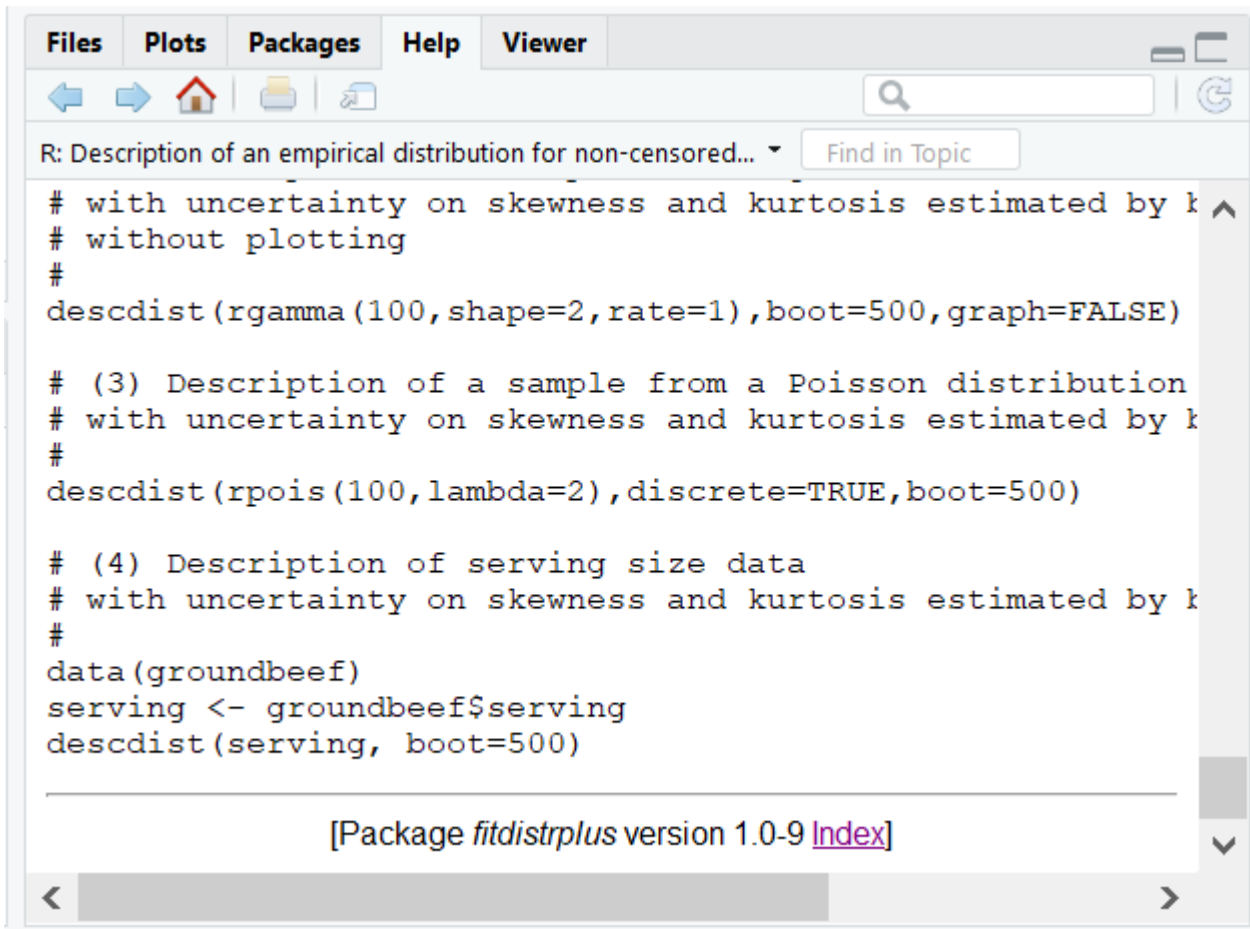
コマンドの詳細



引数, 使い方の説明が出てくる

スクロールすると下の方にExampleが出てくる. これが便利!とりあえず, 実行してから理解できる.





The screenshot shows the RStudio interface with the 'Viewer' pane active. The pane displays R code for the `fitdistrplus` package. The code includes comments and function calls for describing empirical distributions. The code is as follows:

```
# with uncertainty on skewness and kurtosis estimated by k
# without plotting
#
descdist(rgamma(100, shape=2, rate=1), boot=500, graph=FALSE)

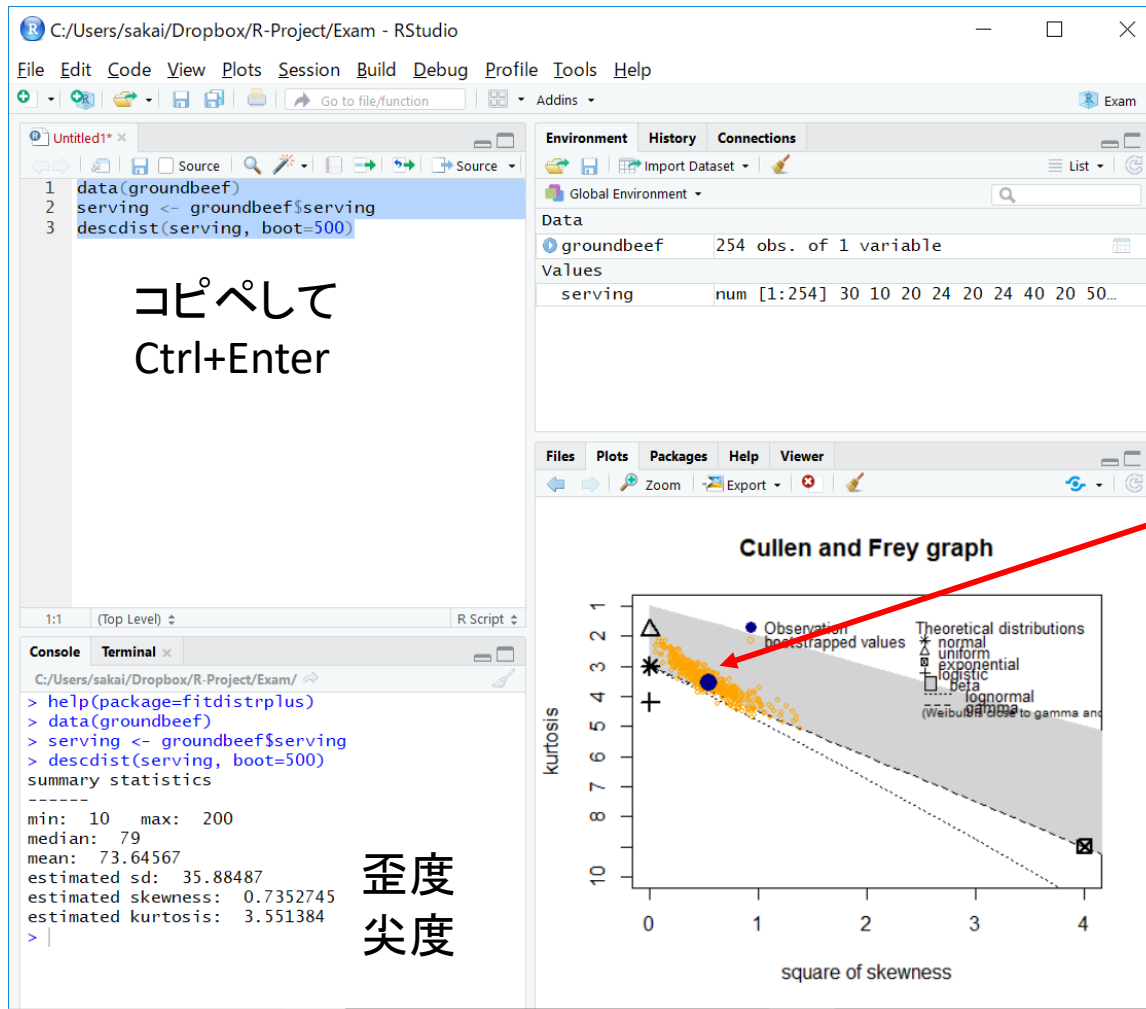
# (3) Description of a sample from a Poisson distribution
# with uncertainty on skewness and kurtosis estimated by k
#
descdist(rpois(100, lambda=2), discrete=TRUE, boot=500)

# (4) Description of serving size data
# with uncertainty on skewness and kurtosis estimated by k
#
data(groundbeef)
serving <- groundbeef$serving
descdist(serving, boot=500)
```

At the bottom of the pane, it says: [Package *fitdistrplus* version 1.0-9 [Index](#)]

(4)の例題は、全く不明のデータから統計分布の推定する一般的手順を示す例題. たったの3行!

Groundbeefというデータはパッケージに付随して提供されるもの。このように、試行するためのデータも一緒に提供されることは素晴らしい。これがどのような結果を示すのか、エディタ領域にコピーしてRstudioで実施してみる。
kurtosis(尖度)とskewness(歪度)の関係から、ガンマ分布か対数正規分布の可能性を知ること。



コピーして
Ctrl+Enter

対数正規分布
の可能性

次に、対数正規確率分布
に適合するかどうかをもう
少し調べたい。どうすれば
よいか？確率紙を使うこと
が候補になる。確率紙とは
何か？

確率紙機能を提供するパッケージ

By 酒井信介

[確率紙の考え方](#) 参照のこと

パッケージを正式に提供するサイトはCRAN.

開発途上のパッケージを提供するサイトGitHub. --→当面これを利用
しかし、インターネット上から誰もが利用できる.

利用法

```
>install.packages("devtools")
```

```
>library(devtools)
```

```
>install_github("ShinsukeSakai0321/ProbPaper")
```

これにて自分のPCにインストールされる以後は使う度に

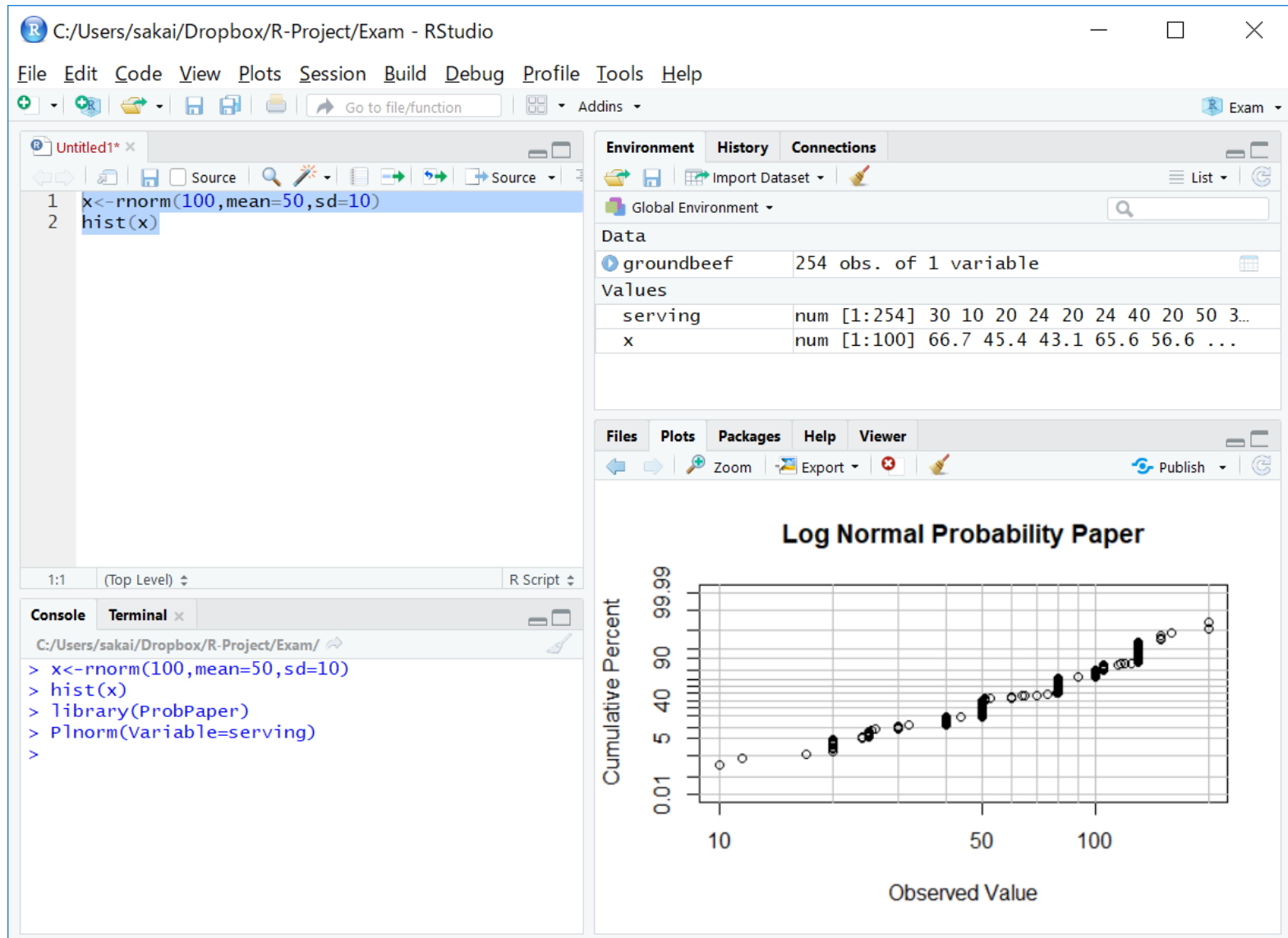
```
>library(ProbPaper)
```

対数正規確率紙上へのプロット

```
>Plnorm(Variable=serving)
```

直線性が確認され確かに対数正規確率分布に適合することが確認できる

確率紙へのプロット

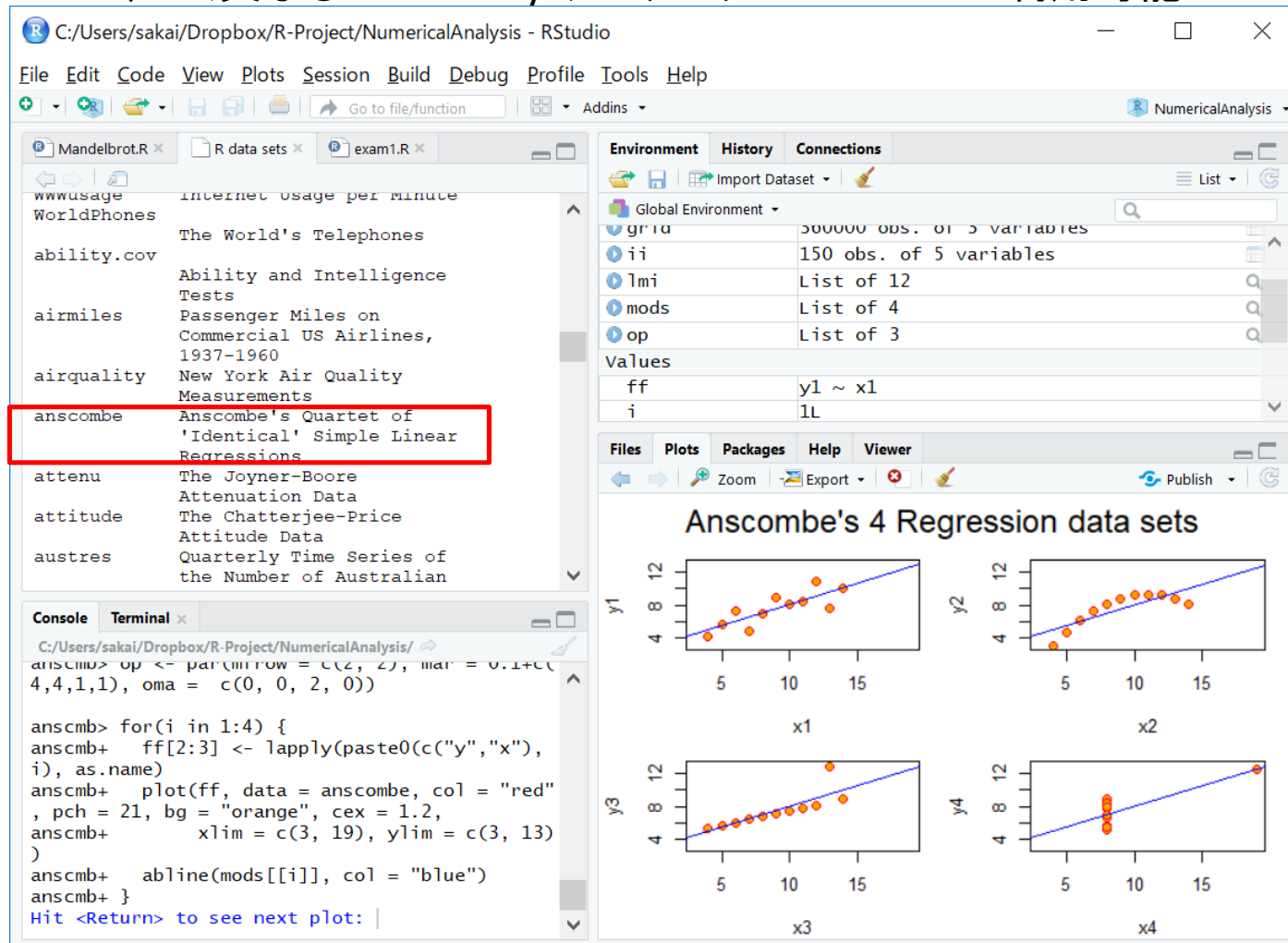


Rに登録されているデータセット

- デフォルトで100以上のものが利用できる
- 現在利用可能なデータセットを表示するには
 - `data()` と入力
- 有名なものの一つがiris
- その使い方を知らなければ
 - `example("dataset name")`

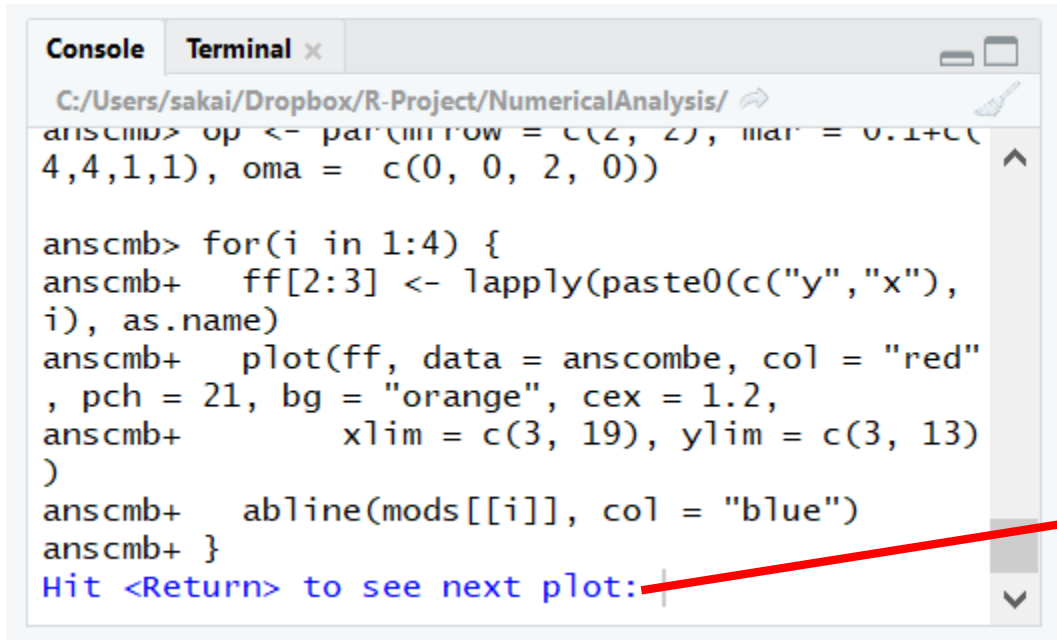
例えば線形単回帰について勉強 したければ

同じ通常の統計的性質(平均、分散、相関、回帰直線)を持つ
が、全く異なる 四つの x-y データセットanscombeが利用可能



exampleの実行

example("anscombe")とコンソール領域に入力すると



```
Console Terminal x
C:/Users/sakai/Dropbox/R-Project/NumericalAnalysis/
anscomb> op <- par(mfrow = c(2, 2), mar = 0.1+c(
4,4,1,1), oma = c(0, 0, 2, 0))

anscomb> for(i in 1:4) {
anscomb+   ff[2:3] <- lapply(paste0(c("y","x"),
i), as.name)
anscomb+   plot(ff, data = anscombe, col = "red"
, pch = 21, bg = "orange", cex = 1.2,
anscomb+     xlim = c(3, 19), ylim = c(3, 13)
)
anscomb+   abline(mods[[i]], col = "blue")
anscomb+ }
Hit <Return> to see next plot: |
```

自動的にコマンド入力,
出力結果を例示する

これだけで, データセット
の使い方を理解できる

この状態で停止するので,
Returnすると

plot領域に描画結果を表示する



Rはこのように、豊富な自習機能を有している. 自分でどんどん学習できる!

anscombeの中身を見してみる

```
Console Terminal x
C:/Users/sakai/Dropbox/R-Project/NumericalAnalysis/

anscmb> mtext("Anscombe's 4 Regression data set
s", outer = TRUE, cex = 1.5)

anscmb> par(op)
> anscombe
      x1 x2 x3 x4      y1      y2      y3      y4
1  10 10 10  8  8.04 9.14  7.46  6.58
2   8  8  8  8  6.95 8.14  6.77  5.76
3  13 13 13  8  7.58 8.74 12.74  7.71
4   9  9  9  8  8.81 8.77  7.11  8.84
5  11 11 11  8  8.33 9.26  7.81  8.47
6  14 14 14  8  9.96 8.10  8.84  7.04
7   6  6  6  8  7.24 6.13  6.08  5.25
8   4  4  4 19  4.26 3.10  5.39 12.50
9  12 12 12  8 10.84 9.13  8.15  5.56
10  7  7  7  8  4.82 7.26  6.42  7.91
11  5  5  5  8  5.68 4.74  5.73  6.89
> |
```

注意: データセットはpackageとともに提供される場合もある(groundbeefなど)

```
library(fitdistrplus)
```

Packageを呼び出した時点でデータセットgroundbeefが登録されている

data()に対して下方に

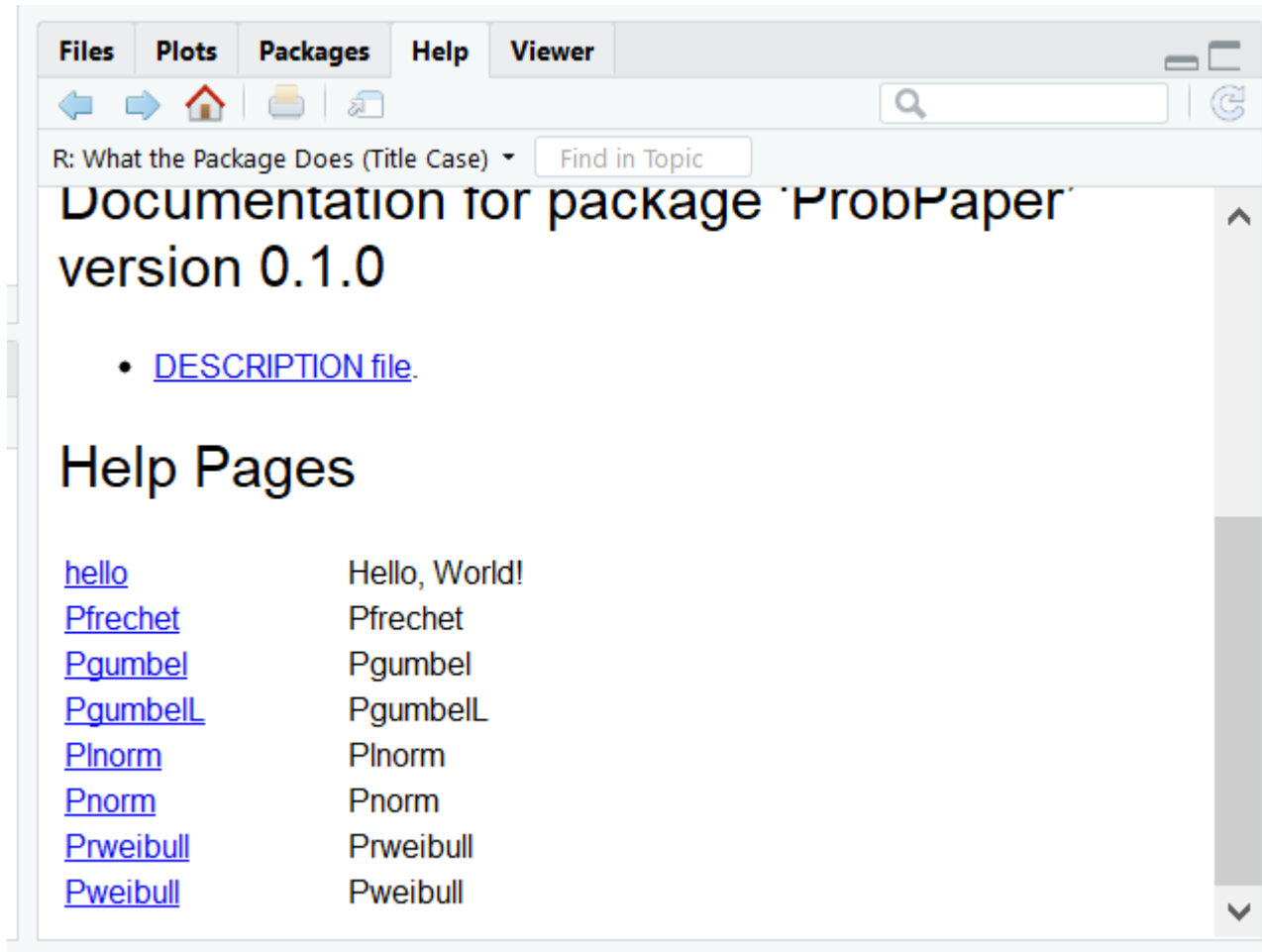
Data sets in package ‘fitdistrplus’:

と表示され、その中に

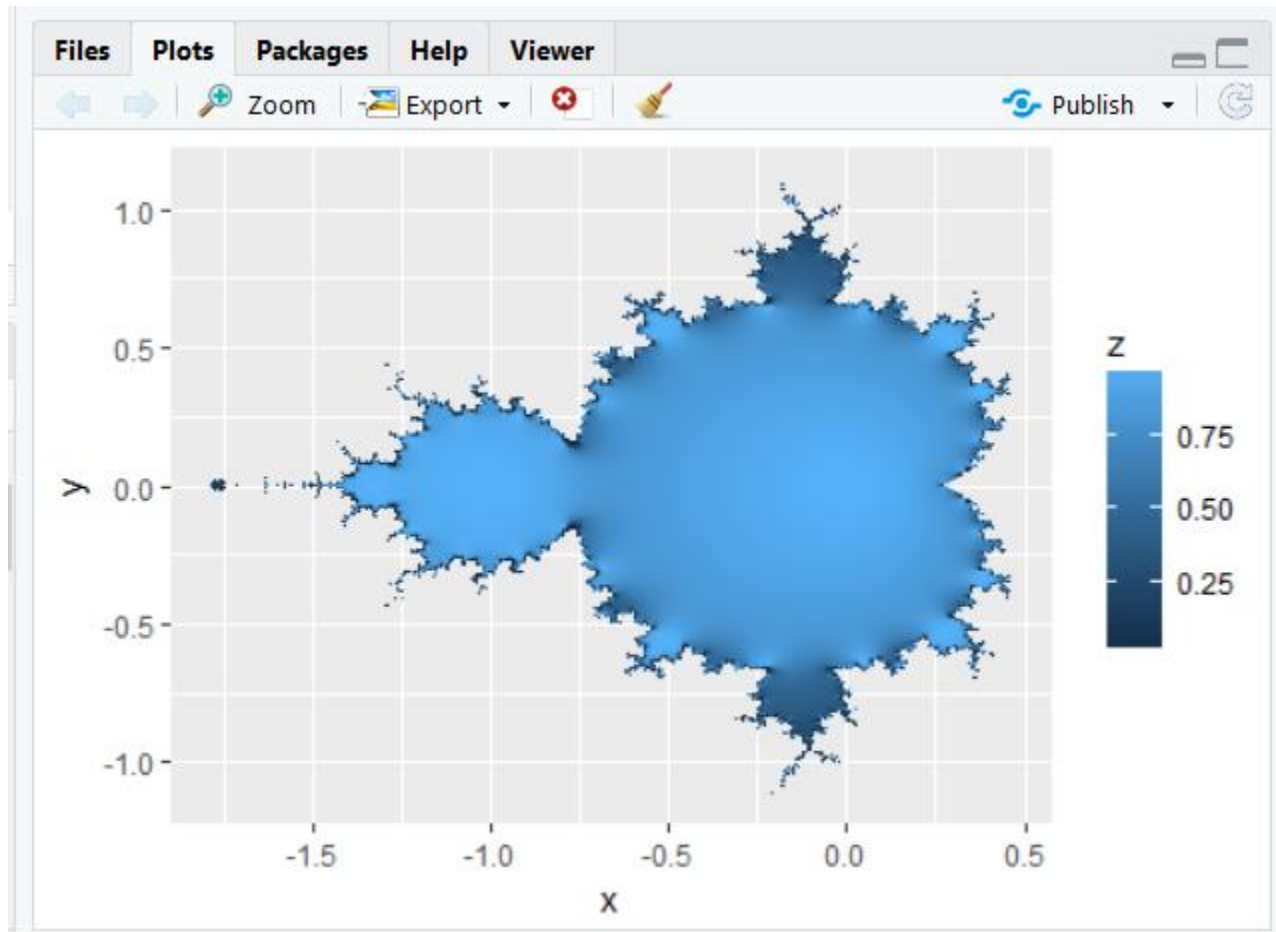
groundbeef Ground beef serving size data set

と

用意されている確率紙



Rでいろいろ遊べる Mandelbortのフラクタル図形



UKgas(1960-1986)ガス使用量 スペクトラム

data(UKgas) → たったの二行!
spec.pgram(Ukgas)

