

Debian での数学ことはじめ。

gnuplot, Octave, R 入門

まえだこうへい mkouhei@debian.or.jp

IRC nick: mkouhei

2009年11月14日

スプレッドシートって落ちませんか？

- データ量増えてくるとレスポンス最悪。
- 循環参照するとOOoなんて落ちる。
- OSまで固まるなんて最悪。

その他の不満。

- 組み込み関数の精度がそもそも怪しい。
- テキストデータじゃないから、diffできないし。
git diff --binary できるけど、読めん！

魂の叫びが聞こえそうです。



もうこんな
生活嫌だ！

今日の献立。

- gnuplot
- GNU Octave
- GNU R

以上、三本をお送りします。



特徴。

- gnuplot
 - グラフ描画が得意なコ。
 - 外部プログラムから呼び出せます。
 - LaTeX とも連携できちゃいます。
- GNU Octave
 - 数値計算が得意なコ。
 - グラフ描画は gnuplot 使います。
 - MATLAB の互換を目指してます。¹
- GNU R
 - 統計処理が得意なコ。
 - 拡張機能が豊富です。
 - グラフ描画機能もあります。

¹別の見方では MATLAB のクローン (clone) なんだとか。

インストール方法。

gnuplot

```
$ sudo apt-get install gnuplot
```

GNU Octave

```
$ sudo apt-get install octave3.2
```

GNU R

```
$ sudo apt-get install r-base
```

今月の具材。

我が家の光熱費

```
"date", "days", "kWh", "kWh/d", "yen/d", "yen", "days", "m^3", "m^3/d", "yen/d", "yen",  
2001/04, 11, 34, 3.09, 83.5, 918, 36, 4.9, 0.14, 108.3, 3900, 58, 9, 0.16, 22.8, 1320, ,  
2001/05, 33, 95, 2.88, 73.8, 2434, 29, 3.5, 0.12, 114.1, 3310, , , , , , , ,  
2001/06, 30, 129, 4.3, 102.7, 3082, 32, 3, 0.09, 96.9, 3100, 61, 3, 0.05, 15.7, 960, ,  
2001/07, 28, 155, 5.54, 131.3, 3676, 28, 1.7, 0.06, 91.1, 2550, , , , , , , ,  
(snip)
```

image200911/konetsu.csv にあるので、手元に勉強会のリポ
ジトリがある人は試してみると良いよ。

今月の具材。

我が家の光熱費

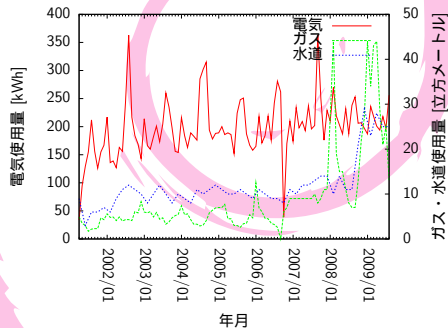
- 2001年4月から2009年9月までのデータ。
- 2回大きく変わる。大雑把に把握しているのは、以下。
- 2006年9月に市川から多摩に引越し。
 - 電気契約が30Aから40Aに。
 - ガスがプロパンから都市ガスに。
 - 水道代が上がる&下水道代が追加に。
- 2008年10月に結婚。
 - ガス、水道使用量が倍増。
 - 電気使用量はあまり変わらず。

gnuplot で処理した場合。

基本は、データのプロット。

set で行っているのはグラフの見栄えが主。

```
reset
set terminal epslatex input color
set output "gnuplot.tex"
set datafile separator ","
set xdata time
set timefmt "%Y/%m"
set format x "%Y/%m"
set xtics rotate by -90
set mxtics 12
set xrange ["2001/04":"2009/08"]
set yrange [0:400]
set xlabel "年月"
set ylabel "電気使用量 [kWh]"
set y2range [0:50]
set y2label "ガス・水道使用量 [立方メートル]"
set ytics nomirror
set y2tics
plot "kohnetsuhi.csv" using 1:3 axes x1y1 \
    title "電気" with line,\
    "" using 1:8 axes x1y2 title "ガス" with line,\
    "" using 1:13 axes x1y2 title "水道" with line
```



gnuplot で処理した場合。

基本は、データのプロット。

set で行っているのはグラフの見栄えが主。

```
reset # 設定の初期化
set terminal epslatex input color # EPS LaTeX でカラー出力
set output "gnuplot.tex" # ファイル名 gnuplot.tex で保存
set datafile separator "," # 読み込むデータファイルはカンマ区切り
set xdata time # x 軸のパラメータは時間データ
set timefmt "%Y/%m" # 読み込む時間データの書式
set format x "%Y/%m" # 出力する x 軸の書式を指定
set xtics rotate by -90 # x 軸のメモリラベルを-90 度回転
set mxtics 12 # x 軸の補助メモリの間隔
set xrange ["2001/04":"2009/08"] # x 軸のデータ範囲
set yrange [0:400] # y 軸のデータ範囲
set xlabel "年月" # x 軸のラベル
set ylabel "電気使用量 [kWh]" # y 軸の名前
set y2range [0:50] # 第 2y 軸のデータ範囲
set y2label "ガス・水道使用量 [立方メートル]" # 第 2y 軸のラベル
set ytics nomirror # 第 2y 軸は右側のみに表示
set y2tics # 第 2y 軸を表示
plot "kohnetsuhi.csv" using 1:3 axes x1y1 \ # 第 1,3 列のデータで x 軸、y 軸で表示
    title "電気" with line, \ # 凡例のタイトルと線プロット
    "" using 1:8 axes x1y2 title "ガス" with line, \
    "" using 1:13 axes x1y2 title "水道" with line
```

gnuplot での処理。

gnuplot を起動してバッチをロード。

```
$ gnuplot
> load 'gnuplot.gp'
> exit
$ ls
gnuplot.eps gnuplot.tex
```

これを LaTeX で取り込むと、先ほどのような図に。

- デフォルトでは、別ウィンドウで表示される。
- 3次元グラフならグルグル回転もできる。
- 他の画像フォーマットにも出力できる。

GNU Rでの処理。

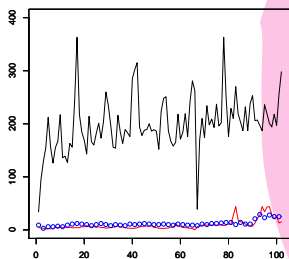
データの取り込み。

```
$ R  
> konetsu <- read.csv("konetsu.csv")
```

データのプロット。

```
> plot(konetsu$kWh, type="l", ylim=c(0,400), ann=F)  
> par(new=T)  
> plot(konetsu$m.3, type="l", ylim=c(0,400), ann=F, col="red")  
> par(new=T)  
> plot(konetsu$m.3.1, ylim=c(0,400), ann=F, col="blue")
```

GNU Rでの処理。



gnuplot のときに比べ、

- 第 2y 軸を表示していない。
- ラベル表示していない。
- 水道使用量が線じゃない。

といったのは、GNU Rのせいではありません。
私のカルマが足りないだけです。

GNU Rでの処理。

電気とガスの使用量の相関関係を見てみる。

```
> cor.test(konetsu$kWh, konetsu$m.3)

Pearson's product-moment correlation

data: konetsu$kWh and konetsu$m.3
t = 1.3154, df = 100, p-value = 0.1914
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.06572595  0.31685457
sample estimates:
      cor
0.1304159
```

まあ、分かっていたことですがかなり低いです。平均気温となら、いずれかは相関関係があるかもしれません。

力及ばず。





Debian パッケージ化に関して。

拡張パッケージあります。

- GNU Octave は、Octave-Forge。
- GNU R は、CRAN(the Comprehensive R Archive Network)
- Perl の CPAN のようなもん。

拡張パッケージを deb パッケージにする。

既に結構 deb パッケージになっているけど全てじゃない。

```
$ apt-cache search "GNU R" | wc -l  
213  
$ apt-cache search octave | wc -l  
133
```

拡張パッケージを deb パッケージにする。

deb パッケージにするためのパッケージが用意されている。

- Octave は、octave-pkg-dev。
 - debian/control の Build-Depends に octave-pkg-dev, Depends に `${octave Depends}` を追記。
 - debian/rules に “include /usr/share/cdbbs/1/class/octave-pkg.mk” を追記。
 - Octave-Forge 由来でない場合は、rules に “SOURCEFORGE=NO” を追記。
- R は、r-base-dev。
 - Debian パッケージポリシーを考慮して、バイナリ形式の実行ファイルは、`/usr/lib/R/bin/R.binary` に分離する必要がある。
 - とあったのだけど、よく分からず。Java のようにバイナリコードで配布されている、ということだろうか？
 - 教えて、エロいヒト。

まとめ

- gnuplot, GNU R, GNU Octave の導入方法。
- gnuplot, GNU R の簡単な使い方。
- 拡張パッケージの Debian パッケージ化の際の考慮。
- 統計処理については、月の平均気温との相関関係を出すつもりだった。
- が、気象庁のデータから引っ張ってくるのは結構面倒なので、今後やってみる。