gnuplot; Easy, Automated, Chart plotting

Gary Briggs <gbriggs@rand.org>

November 6, 2010

Cautionary Tale: Anscombe's Quartet

Mentioned on first page, first chapter, Tufte's "*The Visual Display of Quantitative Information*"

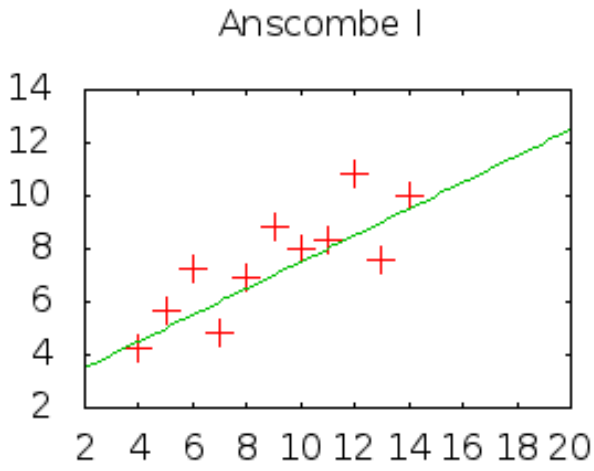
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.1	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.1	4.0	5.39	19.0	12.5
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

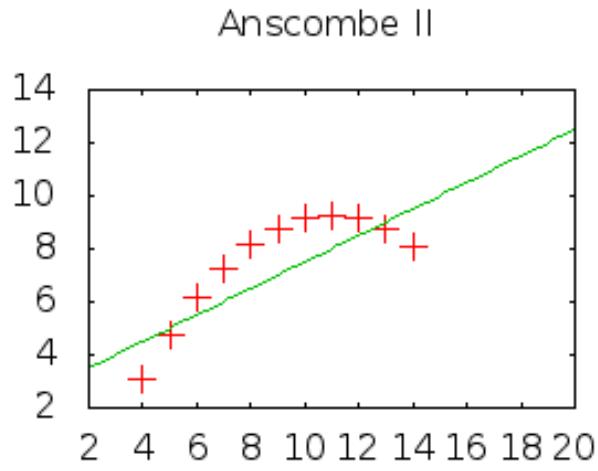
Anscombe's Quartet

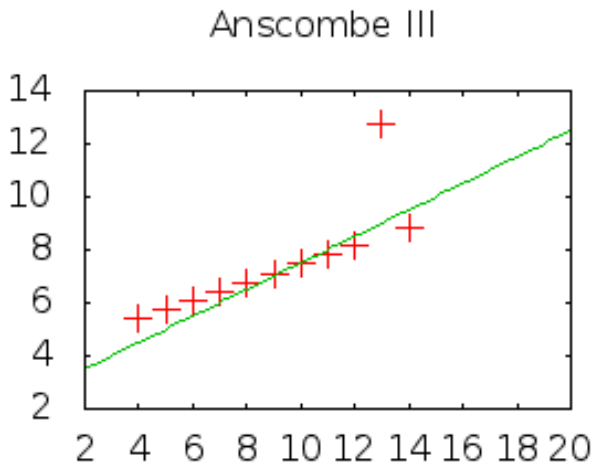
- Some of the usual suspects...
  - all data sets look the same

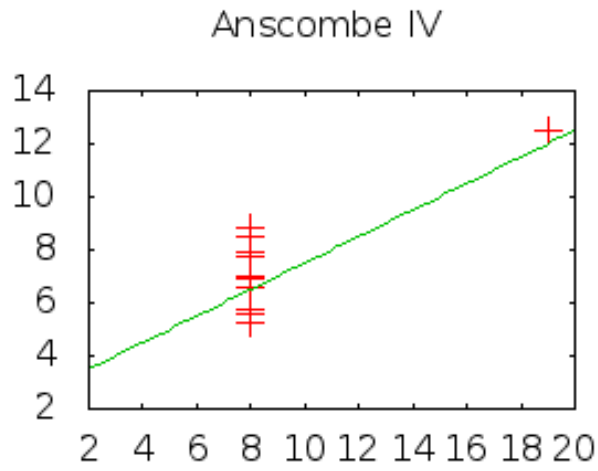
Property	Value
mean(x)	9
variance(x)	10
mean(y)	7.5
variance(y)	3.75
Correlation(x,y)	0.898
linear regression	y=2.5+0.5x

Anscombe's Quartet, Graphically









In which Gary kvetches about Excel

- If you want something unusual, excel may be a pain
- Difficult to automate
  - Even at only 10 seconds a chart, plotting a bunch of charts takes a while
  - And then your researcher asks for something *slightly* different...
  - ... or you realise you screwed something up and have to redo the plots anyway

Simple gnuplot stuff, input data

- Data must be space-separated
- Data must be in columns
- Separate groups with a blank line
- Built in help is great. Type “help”, or “help topic”
- Simple command:

```
plot “filename.dat” using 1:2 with points # 2d plot
splot “filename.dat” using 1:2:3 with lines # 3d plot
```

Simple gnuplot stuff, concepts

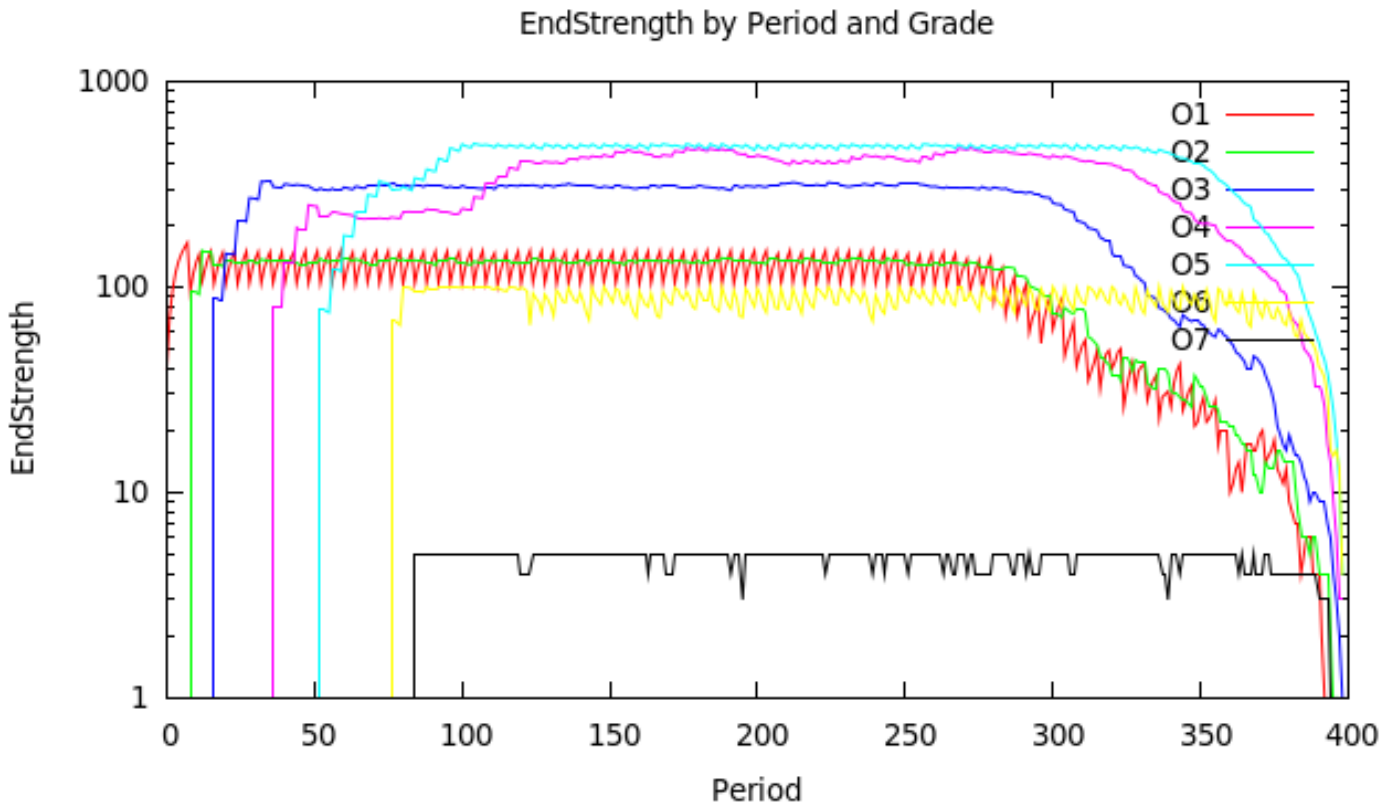
- Terminals
  - Default is currently “wxt”, which does an on-screen plot
    - “wxt” includes the ability to scale, zoom in, rotate 3d plots...
    - hit “h” in the wxt window to spam the keyboard shortcuts to the terminal
  - Set to “png” to output a png, “jpeg” for jpeg, etc
    - set output “out.png”
  - “help terminal” to get a list, and what each terminal’s parameters are
  - eg: set terminal png size 320,240
- set ...
  - set logscale y # or other axes
  - set xlabel “Test” # or other axes
  - set title “Cookies” # Note the quotes
  - unset key # To hide the legend
  - set xrange [0:100] reverse # Force it to go 100..0
- replot
  - Re-runs the last plot command after you’ve changed other stuff

Demo

- Help system
  - help plot style
  - help terminal
- Simple plots, mostly of Anscombe's Quartet, endstrength, retentionrates
  - plot parameters: using, with <style>, title "<title>"
- 3d plots

Demo Redux, Endstrength

```
set title "EndStrength by Period and Grade"
set xlabel "Period"
set ylabel "EndStrength"
set logscale y
plot \
        "endstrengthbygrade.txt" using 1:3 with lines t "O1", \
        "endstrengthbygrade.txt" using 1:4 with lines t "O2", \
        "" using 1:5 with lines t "O3", \
        "" using 1:6 with lines t "O4", \
        "" using 1:7 with lines t "O5", \
        "" using 1:8 with lines t "O6", \
        "" using 1:9 with lines t "O7"
```



Demo Redux, Retention Models

```
set title "Retention Rate by Grade, TOS for two retention models"
set xlabel "Grade"
set xrange [0:7] reverse
set ylabel "YOS"
set zlabel "Retention%"
set ticslevel 0
set view 53,122
splot "simplebygraderetentionmodel.txt" using 1:2:4 with lines title "SimpleByGrade Retention", \
       "nullretentionmodel.txt" using 1:2:4 with lines title "Null Retention"
```

Automating

1. Output data to a file, space separated, in columns
2. Use gnuplot manually to figure out how you want to plot stuff
3. Write all those commands in a file, and pipe it into gnuplot
4. You can do multiple plots from one script, by multiple calls to "set output" then "plot"
5. ... I tend to write stuff so that my program outputting the data also output the commands to plot
   - ./program input.file output.file | gnuplot

Anscombe's Quartet, gnuplot

```
set terminal png size 320,240 # Output PNGs this size
set pointsize 2 # Make the points bigger
set nokey # No need for a legend

set xrange [2:20] # gnuplot auto scales each plot by default
set yrange [2:14] # Make them all the same range

set title "Anscombe I"
set output "anscombe1.png"
plot "anscombe.txt" using 1:2 with points, 2.5+0.5*x

set title "Anscombe II"
set output "anscombe2.png"
plot "anscombe.txt" using 3:4 with points, 2.5+0.5*x

set title "Anscombe III"
set output "anscombe3.png"
plot "anscombe.txt" using 5:6 with points, 2.5+0.5*x

set title "Anscombe IV"
set output "anscombe4.png"
plot "anscombe.txt" using 7:8 with points, 2.5+0.5*x
```

Obtaining gnuplot

- Linux
  - It's already on most unix boxes here
  - If not, "sudo yum install gnuplot"
  - Or ask your local friendly sysadmin
- Windows
  - When you go looking, you want the "MSYS" one, not the "cygwin one"
  - `http://www.tatsuromatsuoka.com/gnuplot/Eng/winbin/`
- OSX
  - It's in fink or darwinports