

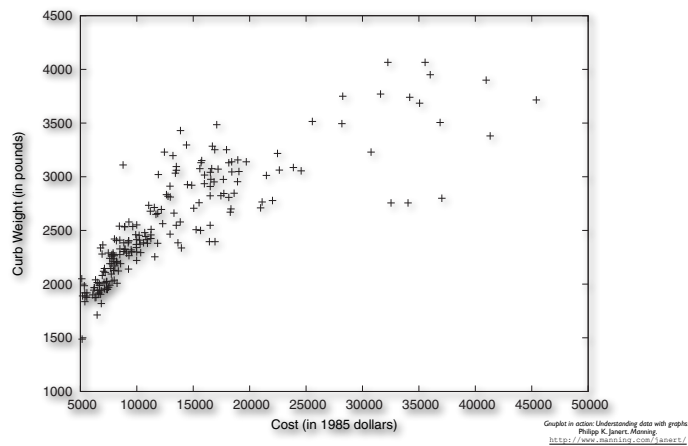
Investigating Relationships

Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. Manning.
<http://www.manning.com/janert/>

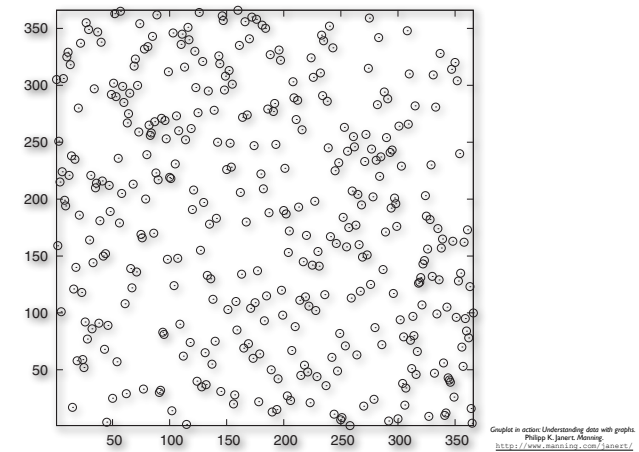
Scatter plots

- Assumes *Bivariate* data, i.e. lists of 2-tuples of responses
- The point is to check the nature of the *relationship* between the two responses
- Take care of *outliers*

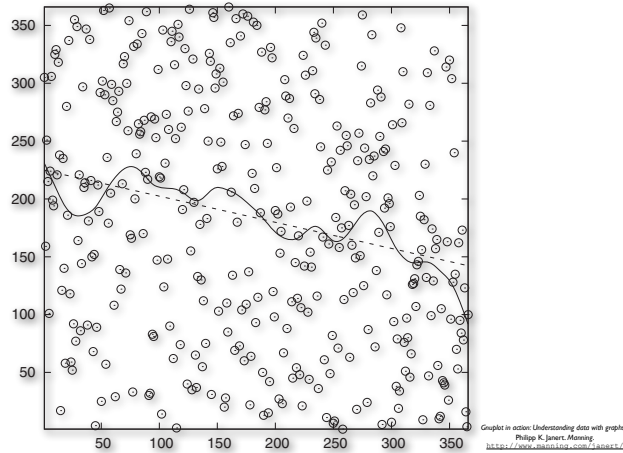
Example 1: Weight vs. Cost



Example 2: The 1970 Draft Lottery



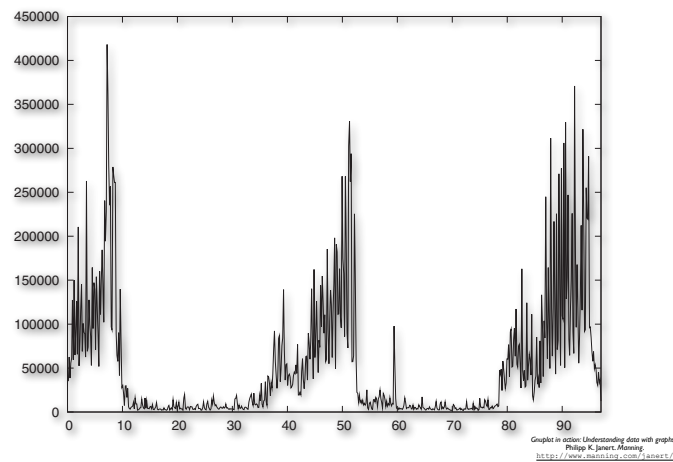
Example 2: The 1970 Draft Lottery



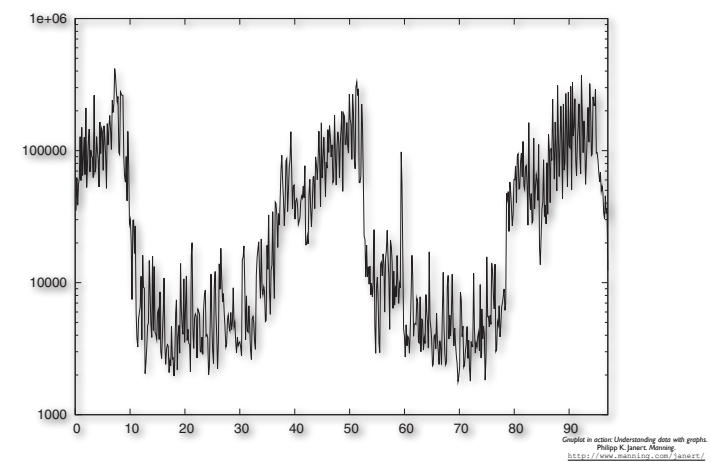
Logarithmic Scale

- **Serve three main purposes:**
 - Rein in *large variation* of the data
 - Turn *multiplicative* deviations into *additive* ones
 - Reveal *exponential* and *power law* behavior

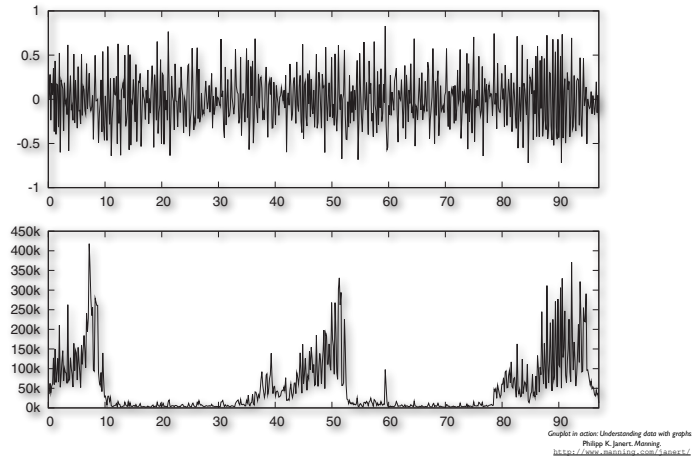
Example 1: Traffic Pattern at Website



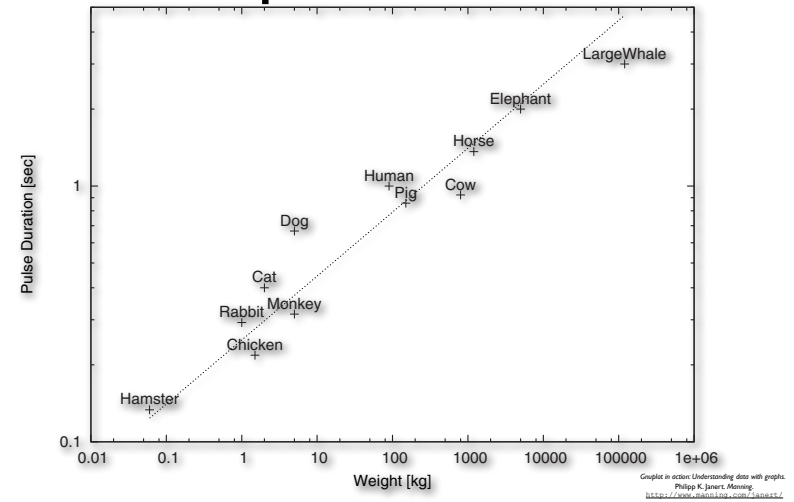
Example 1: Traffic Pattern at Website



Example 1: Traffic Pattern at Website



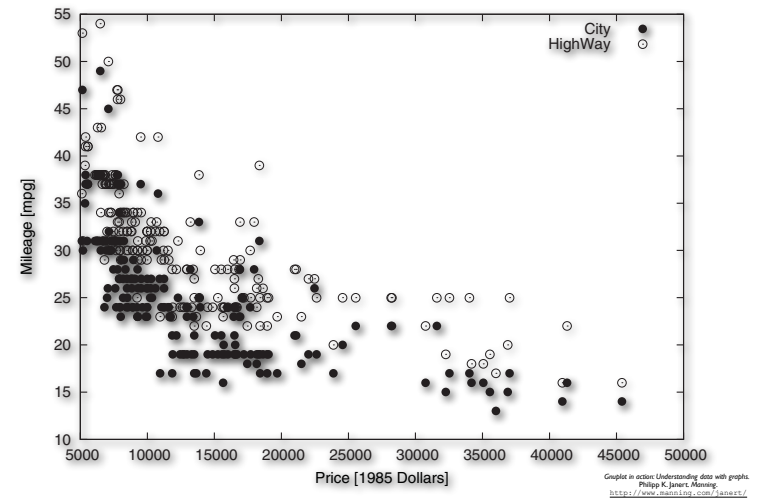
Example 2: Mammals



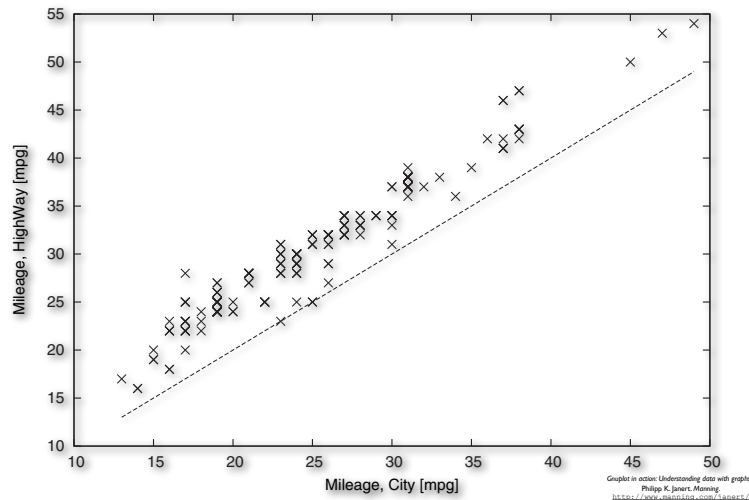
The Core Principle

- Plot **exactly** what you want to **see**

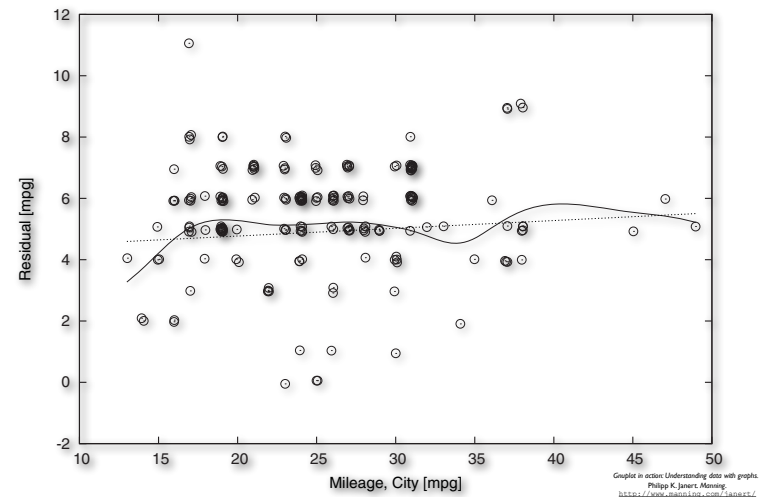
Iterate & Transform



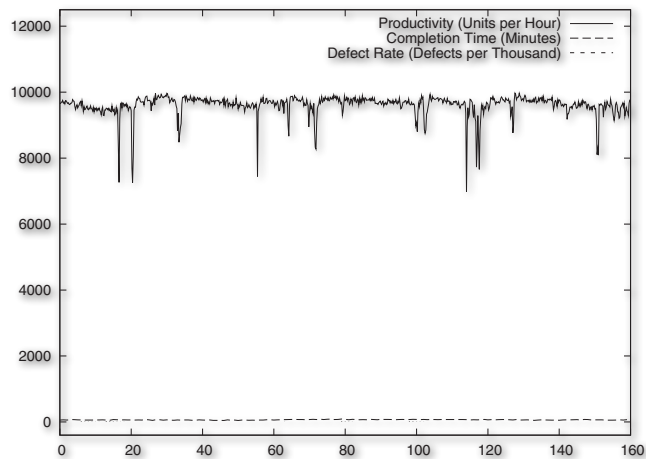
Iterate & Transform



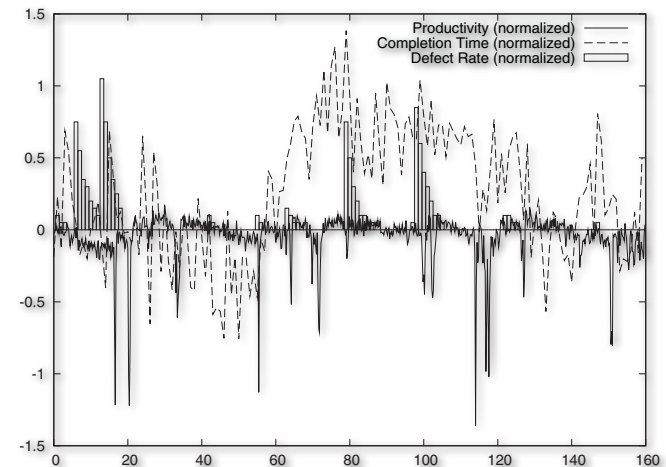
Iterate & Transform



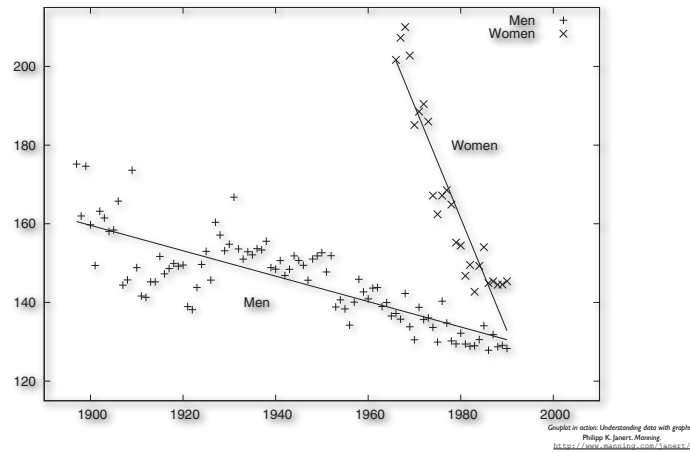
What's Wrong ?



Normalized Metrics



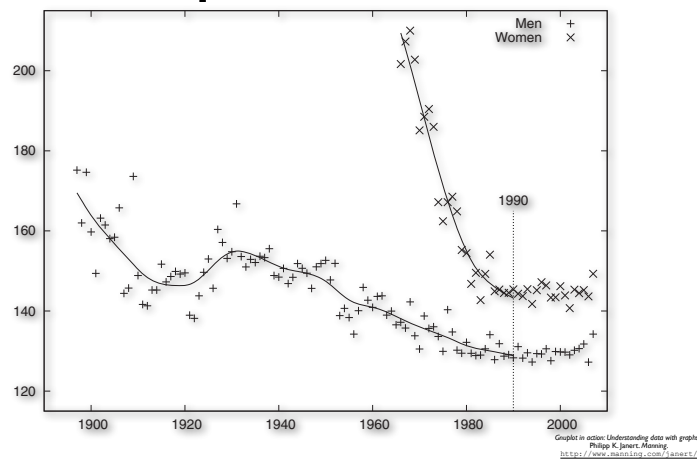
Truncation & Responsiveness



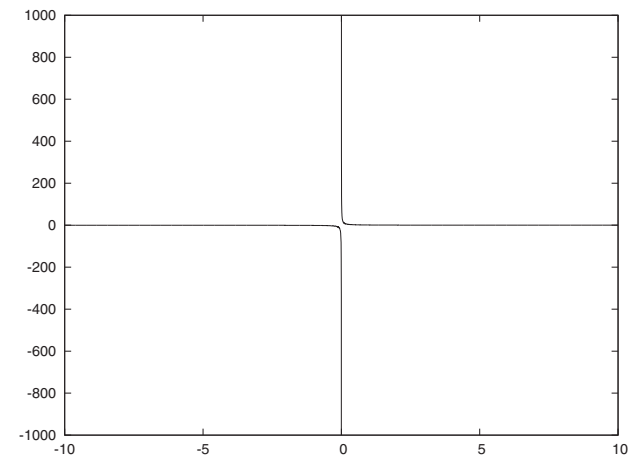
Truncation & Responsiveness

- Outlier removal
- Sampling bias
- Edge effects

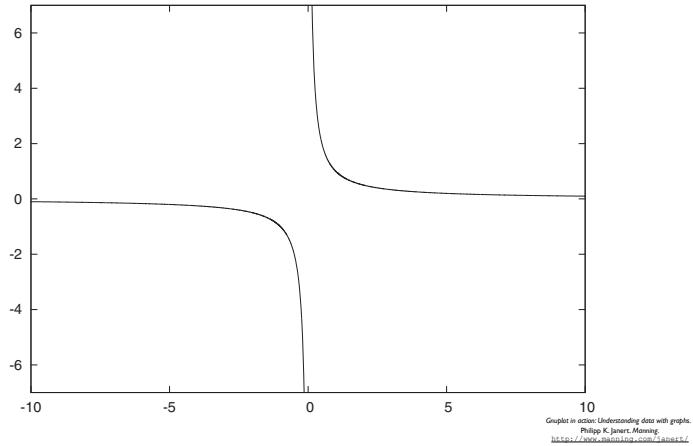
Truncating & Responsiveness



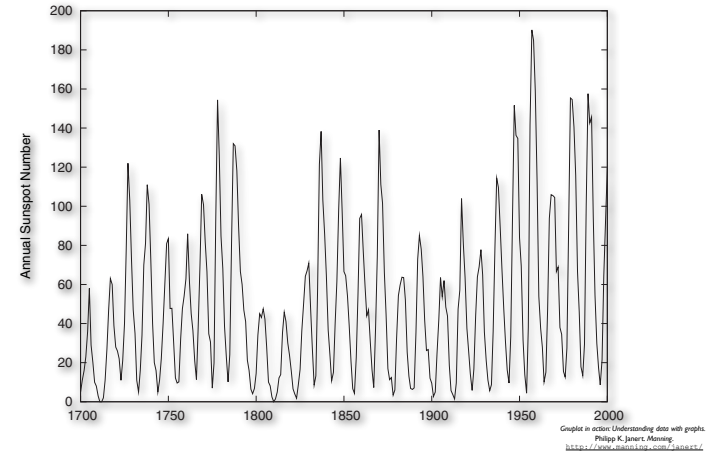
Improving Perception



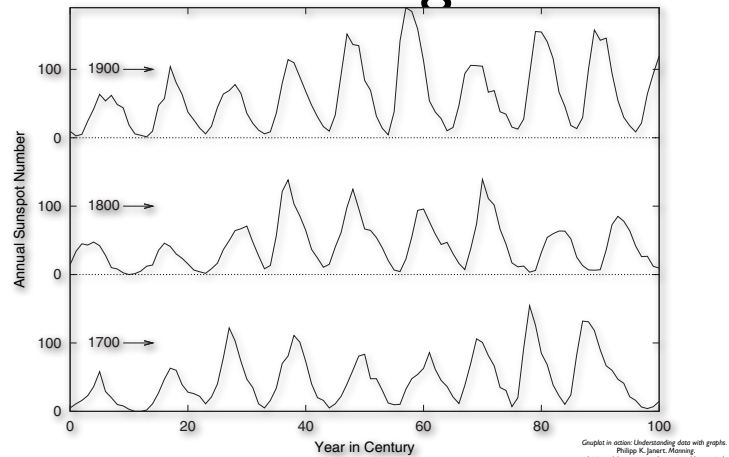
Improving Perception: Banking



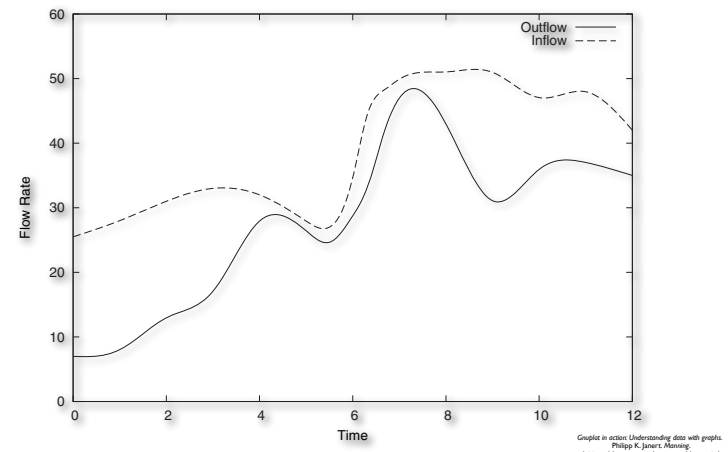
Improving Perception: Banking



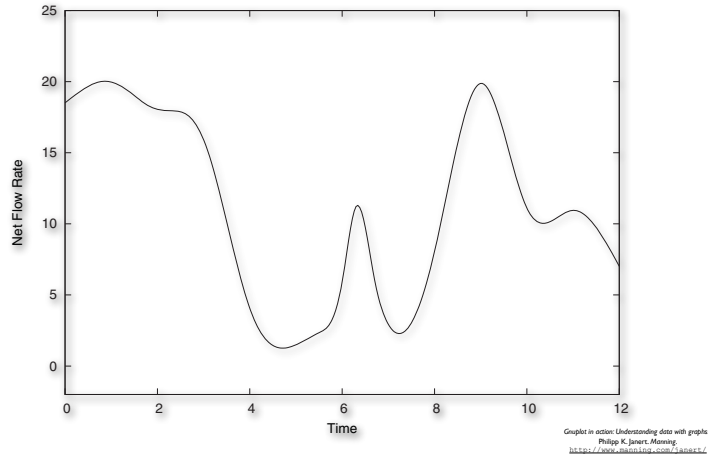
Improving Perception: Banking



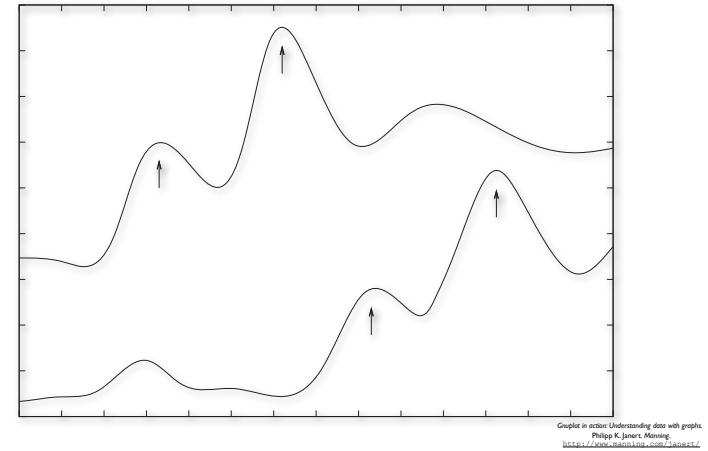
Judging Lengths and Distances



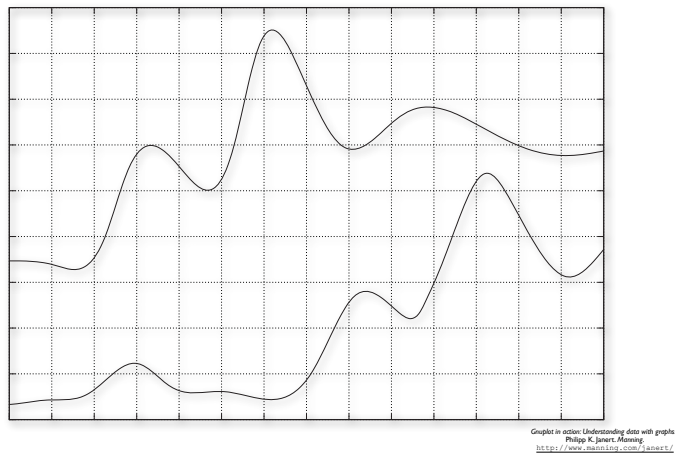
Judging Lengths and Distances



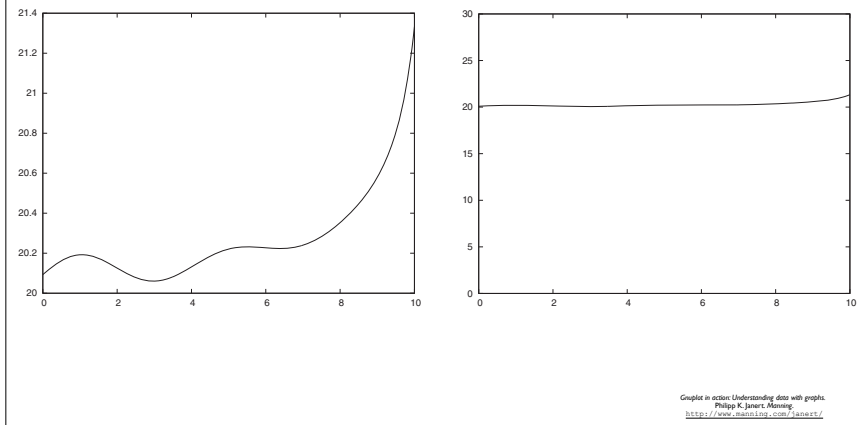
Enhancing Quantitative Perception



Enhancing Quantitative Perception



Plot Ranges ?



The Core Principle

- Plot **exactly** what you want to **see**

GNU PLOT 101

<http://www.gnuplot.info>

Gnuplot in action: Understanding data with graphs.

Philipp K. Janert. Manning.

<http://www.manning.com/janert/>

GNU PLOT

- **Free** software for plotting data
- **NOT** «push-button-limited-capacities» type of software
- **Multiplatform**
- Integrates well with **LaTeX**

GNU PLOT Invocation

```
Mac-Pro:metho tixeuil$ gnuplot
```

```
GNU PLOT  
Version 4.3 patchlevel 0  
last modified March 2009  
System: Darwin 9.8.0
```

```
Copyright (C) 1986-1993, 1998, 2004, 2007-2009  
Thomas Williams, Colin Kelley and many others
```

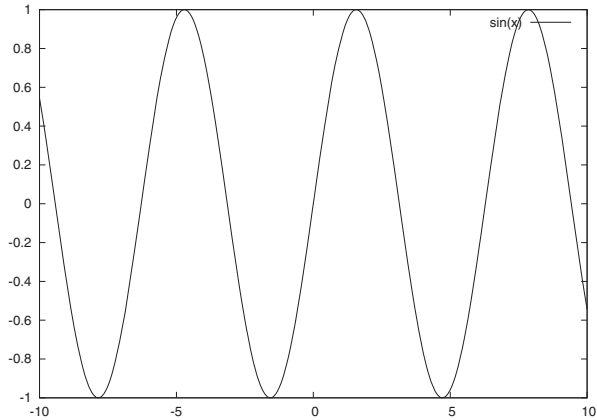
```
Type 'help' to access the on-line reference manual.  
The gnuplot FAQ is available from  
http://www.gnuplot.info/faq/
```

```
Send comments and help requests to <gnuplot-beta@lists.sourceforge.net>  
Send bug reports and suggestions to <gnuplot-beta@lists.sourceforge.net>
```

```
Terminal type set to 'x11'  
gnuplot>
```

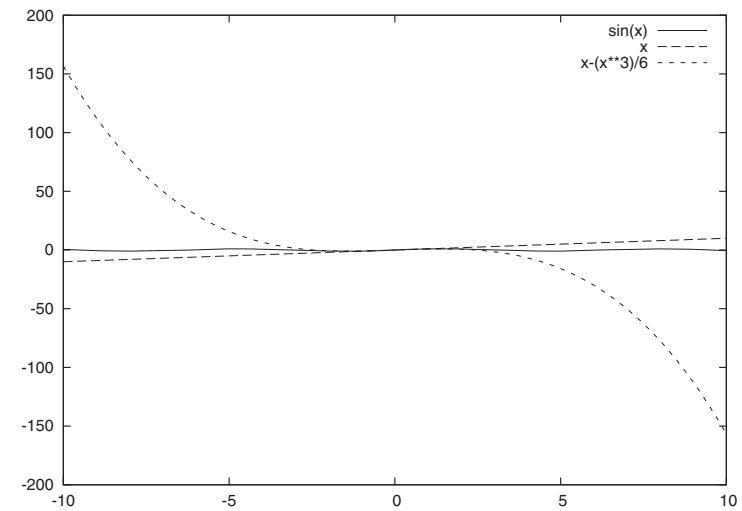

First plots

```
gnuplot>  
plot sin(x)
```



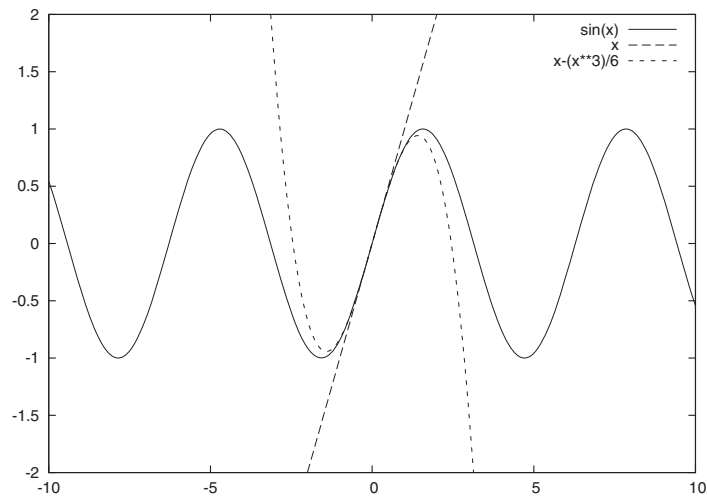
Gnuplot in action: Understanding data with graphs.
 Philipp K. Janert, Manning.
 <http://www.manning.com/Janert/>

```
plot sin(x), x, x-(x**3)/6
```



Gnuplot in action: Understanding data with graphs.
 Philipp K. Janert, Manning.
 <http://www.manning.com/Janert/>

```
plot [-2:2] sin(x), x, x-(x**3)/6
```



Gnuplot in action: Understanding data with graphs.
 Philipp K. Janert, Manning.
 <http://www.manning.com/Janert/>

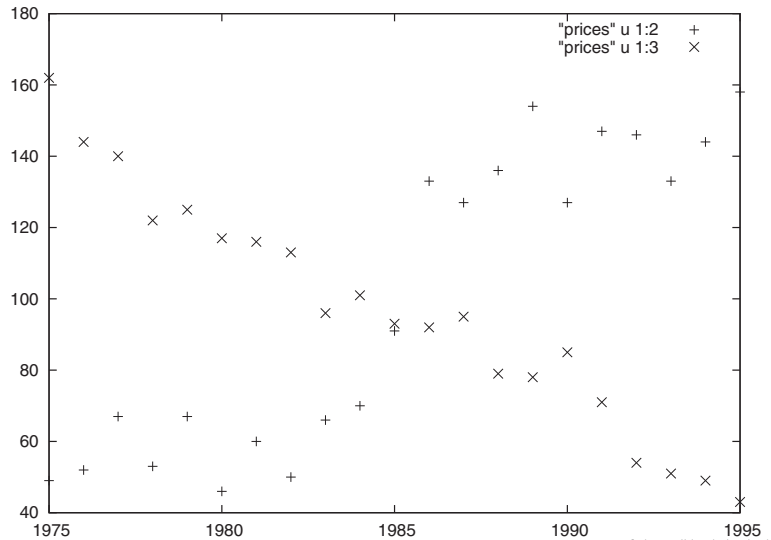
Plotting From Data

Average PQR and XYZ stock price (in dollars per share) per calendar year

1975	49	162
1976	52	144
1977	67	140
1978	53	122
1979	67	125
1980	46	117
1981	60	116
1982	50	113
1983	66	96
1984	70	101
1985	91	93
1986	133	92
1987	127	95
1988	136	79
1989	154	78
1990	127	85
1991	147	71
1992	146	54
1993	133	51
1994	144	49
1995	158	43

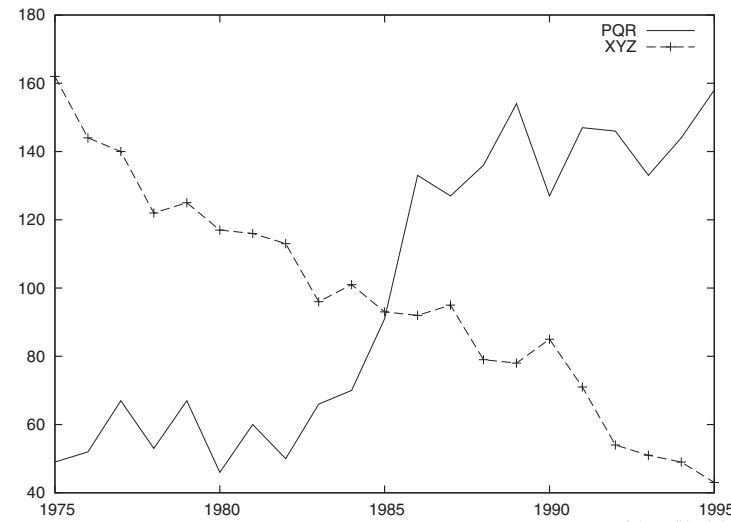
Gnuplot in action: Understanding data with graphs.
 Philipp K. Janert, Manning.
 <http://www.manning.com/Janert/>

```
plot "prices" using 1:2, "prices" using 1:3
```



Gnuplot in action: Understanding data with graphs.
Philip K Janert, Manning.
<http://www.manning.com/Janert/>

```
plot "prices" using 1:2 title "PQR" with lines,  
"prices" using 1:3 title "XYZ" with linespoints
```



Gnuplot in action: Understanding data with graphs.
Philip K Janert, Manning.
<http://www.manning.com/Janert/>

Data Transformation

```
plot "data" using 1:( sqrt($2) ) with lines  
plot "data" using 1:( ($2+$3)/2 ) with lines  
plot "data" using 1:2 with lines, "" using 1:( $3/100 ) with lines  
plot "data" using ( log($1) ):( log($2) ) with lines
```

```
set logscale  
set logscale x  
set logscale y  
  
unset logscale  
unset logscale x  
unset logscale y
```

Gnuplot in action: Understanding data with graphs.
Philip K Janert, Manning.
<http://www.manning.com/Janert/>

Plotting Unix /etc/passwd

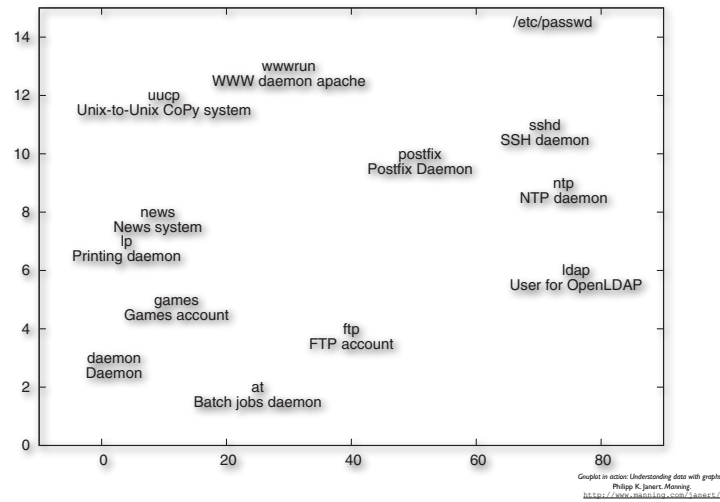
```
at:x:25:25:Batch jobs daemon:/var/spool/atjobs:/bin/bash  
daemon:x:2:2:Daemon:/sbin:/bin/bash  
ftp:x:40:49:FTP account:/srv/ftp:/bin/bash  
games:x:12:100:Games account:/var/games:/bin/bash  
ldap:x:76:70>User for OpenLDAP:/var/lib/ldap:/bin/bash  
lp:x:4:7:Printing daemon:/var/spool/lpd:/bin/bash  
mail:x:8:12:Mailer daemon:/var/spool/clientmqueue:/bin/false  
man:x:13:62:Manual pages viewer:/var/cache/man:/bin/bash  
mysql:x:60:108:MySQL database admin:/var/lib/mysql:/bin/false  
news:x:9:13:News system:/etc/news:/bin/bash  
ntp:x:74:103:NTP daemon:/var/lib/ntp:/bin/false  
postfix:x:51:51:Postfix Daemon:/var/spool/postfix:/bin/false  
sshd:x:71:65:SSH daemon:/var/lib/ssh:/bin/false  
uucp:x:10:14:Unix-to-Unix CoPy system:/etc/uucp:/bin/bash  
wwwrun:x:30:8:WWW daemon apache:/var/lib/wwwrun:/bin/false
```

Gnuplot in action: Understanding data with graphs.
Philip K Janert, Manning.
<http://www.manning.com/Janert/>

```

set datafile separator ':'
set datafile commentschar "m"
plot [-20:150][:27] "/etc/passwd"
↳ u 3:($0+2):( stringcolumn(1) . "\n" . stringcolumn(5) ) w labels

```



Exporting Graphics

- «Web» graphics
 - *JPG, SVG, PNG, GIF*
- «Print» graphics
 - *EPS, EPSLaTeX, PDF*

Exporting EPS

```

... # plot commands
set terminal postscript eps enhanced
set output 'enhanced.eps'
replot

```

Including EPS in LaTeX

```

\documentclass{article}
\usepackage{graphicx}

\begin{document}

\section{The First Section}

\begin{figure}[h]
\begin{center}
\includegraphics[width=10cm]{enhanced}
\end{center}
\caption{A Postscript file, included in \LaTeX}
\end{figure}

\end{document}

```

Including EPS in LaTeX

1 The First Section

Here is a very short paragraph. The plot will be included after this paragraph.

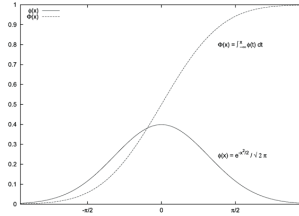


Figure 1: A Postscript file, included in LaTeX

And here is a second paragraph. The graph should have been included before.

2 The Second Section

The second section really contains only a very short text.

Graphic in action: Understanding data with graphs.
Philipp K. Janert, Manning
<https://www.it-ebooks.info/item/2484837/>

Implementing EDA 4BT

- Run-sequence Plot
- Lag Plot
- Histogram
- (Normal) Probability Plot

Run-sequence Plot

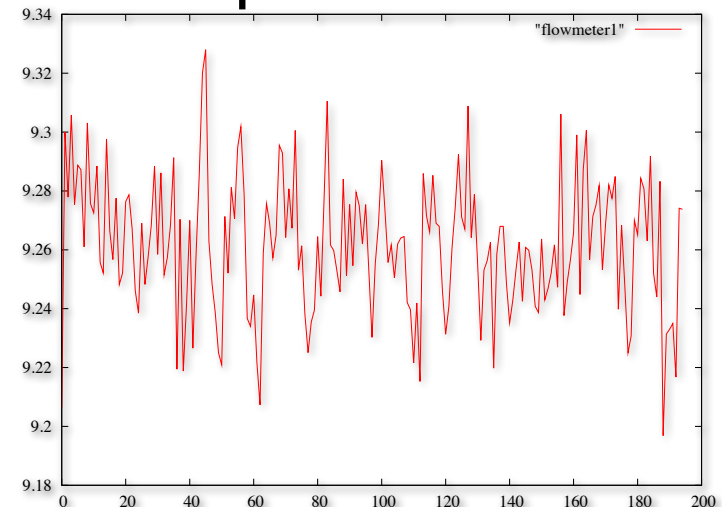
```
set terminal postscript eps color  
"Times-Roman" 16
```

```
set output "flowmeter_runseq.eps"  
plot "flowmeter1" with lines
```

flowmeter I

```
9.206343  
9.299992  
9.277895  
9.305795  
9.275351  
9.288729  
9.287239  
9.260973  
...
```

Run-sequence Flow DS



Lag Plot

```
#!/usr/bin/perl

$previous = <>;
chomp($previous);
while ( $current = <> ) {
    chomp($current);
    print $current . "\t" .
$previous . "\n";
    $previous = $current;
}
```

Lag Plot

```
$> perl lag.pl < flowmeter1 > flowmeter2
```

flowmeter1

```
9.206343
9.299992
9.277895
9.305795
9.275351
9.288729
9.287239
9.260973
...
```

flowmeter2

```
9.299992 9.206343
9.277895 9.299992
9.305795 9.277895
9.275351 9.305795
9.288729 9.275351
9.287239 9.288729
9.260973 9.287239
...
```

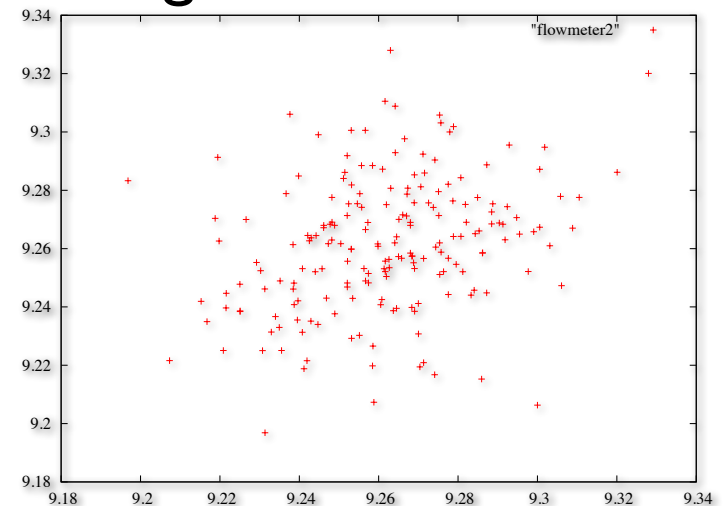
Lag Plot

```
set output "flowmeter_lag.eps"
plot "flowmeter2"
```

flowmeter2

```
9.299992 9.206343
9.277895 9.299992
9.305795 9.277895
9.275351 9.305795
9.288729 9.275351
9.287239 9.288729
9.260973 9.287239
...
```

Lag Plot Flow DS



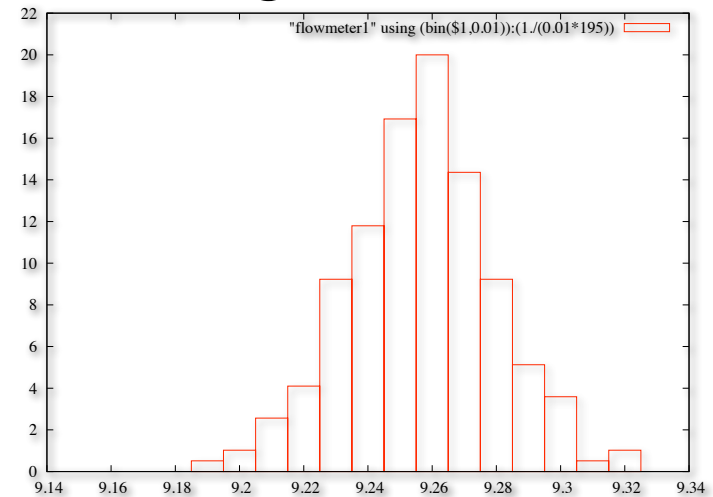
Histogram

```
set output  
"flowmeter_histogram.eps"  
bin(x,s) = s*int(x/s)  
set boxwidth 0.01  
plot "flowmeter1" using  
(bin($1,0.01)):(1./(0.01*195))  
smooth frequency with boxes
```

flowmeter1

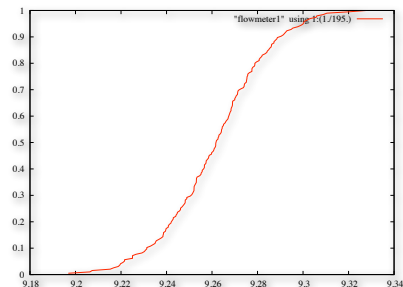
```
9.206343  
9.299992  
9.277895  
9.305795  
9.275351  
9.288729  
9.287239  
9.260973  
...
```

Histogram Flow DS



(Normal) Probability Plot

```
set output "flowmeter_cumulative.eps"  
plot "flowmeter1" using 1:(1./195.)  
smooth cumulative
```



(Normal) Probability Plot

```
set table "flowmeter_cdf"  
replot  
unset table
```

flowmeter1

```
9.206343  
9.299992  
9.277895  
9.305795  
9.275351  
9.288729  
9.287239  
9.260973  
...
```

flowmeter_cdf

```
9.19685 0.00512821 i  
9.20634 0.0102564 i  
9.20733 0.0153846 i  
9.21527 0.0205128 i  
9.21675 0.025641 i  
9.21881 0.0307692 i  
...
```

(Normal) Probability Plot

```
Set output  
"flowmeter_isnormal.eps"  
plot "flowmeter_cdf" using  
(invnorm($2)):1 with lines
```

flowmeter_cdf

```
9.19685 0.00512821 i  
9.20634 0.0102564 i  
9.20733 0.0153846 i  
9.21527 0.0205128 i  
9.21675 0.025641 i  
9.21881 0.0307692 i  
...
```

(Normal) Probability Plot Flow DS

