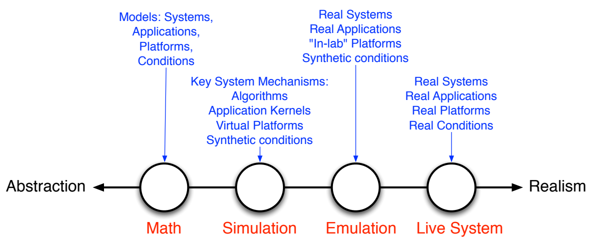
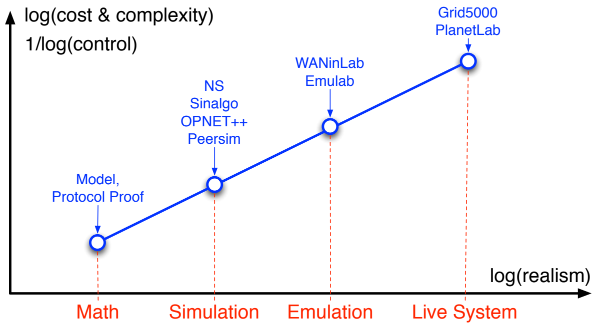
Making the Point

Sébastien Tixeuil
Univ. Pierre & Marie Curie - Paris 6
Sebastien.Tixeuil@lip6.fr

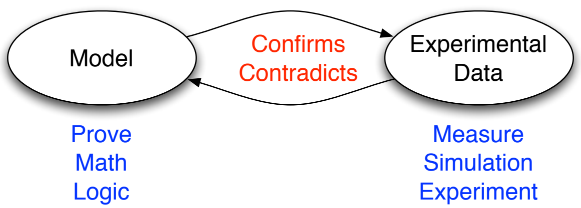
Studying Networks



Studying Networks



Agenda



Agenda

- **Writing Proofs**
- **Managing Experimental Data**
  - Classical *vs.* Exploratory
  - Practicalities

How to Write a Proof

How to write proofs: a quick guide. Eugenia Cheng.
http://www.math.uchcago.edu/~eugenia

Proof

- **Begining**: things we assume to be true, including the definitions of the things we talk about
- **Middle**: statements, each following logically from the things before it
- **End**: the thing we're trying to prove

Kinds of things to try and prove

$x = y$

$x \Rightarrow y$

$x \iff y$

x is purple

$\forall x, p(x)$ is true

$\exists x$ such that $p(x)$ is true

Example 1. *Using the field axioms, prove that $a(b-c) = ab - ac$ for any real numbers a, b, c. You may use the fact that $x.0 = 0$ for any real number x.*

BEGINNING
field axioms
definition $x - y = x + (-y)$
given $x.0 = 0$

MIDDLE

$$\begin{aligned} a(b-c) &= a(b+(-c)) && \text{definition} \\ &= ab + a(-c) && \text{distributive law} \\[1em] ac + a(-c) &= a(c+(-c)) && \text{distributive law} \\ &= a.0 && \text{additive inverse} \\ &= 0 && \text{given} \\ \therefore\ a(-c) &= -(ac) && \text{definition of additive inverse} \\[1em] \therefore\ ab + a(-c) &= ab - ac \end{aligned}$$

END
$\therefore$ by line 2, $a(b-c) = ab - ac$ as required □

Example 2. *Let f and g be functions $A \xrightarrow{f} B \xrightarrow{g} C$. Show that if f and g are injective then $g \circ f$ is injective*

BEGINNING: definition of injective
definition $(g \circ f)(a) = g(f(a))$
assumption that f and g are injective i.e.
$\forall a, a' \in A \quad f(a) = f(a') \Longrightarrow a = a'$
$\forall b, b' \in B \quad g(b) = g(b') \Longrightarrow b = b'$

MIDDLE:

$$
\begin{aligned}
(g \circ f)(a) = (g \circ f)(a') &\Longrightarrow g(f(a)) = g(f(a')) && \text{by definition}\\
&\Longrightarrow f(a) = f(a') && \text{since g is injective}\\
&\Longrightarrow a = a' && \text{since f is injective}\\
\therefore (g \circ f)(a) = (g \circ f)(a') &\Longrightarrow a = a'
\end{aligned}
$$

END: i.e. $g \circ f$ is injective, as required □

Example 3. *Prove by induction that $\forall n \in \mathbb{N}, 1 + \cdots + n = \dfrac{n(n+1)}{2}$*

BEGINNING: Principle of Induction

MIDDLE:

$$
\begin{aligned}
\text{for } n = 1, \text{ LHS} &= 1\\
\text{RHS} &= \frac{1(1+1)}{2}\\
&= 1
\end{aligned}
$$

$\therefore$ result is true for $n = 1$

If result is true for $n = k$ then

$$
\begin{aligned}
1 + \cdots + k + (k+1) &= \frac{k(k+1)}{2} + (k+1)\\
&= \frac{k(k+1) + 2(k+1)}{2}\\
&= \frac{(k+1)(k+2)}{2} \qquad \text{i.e. result true for } n = k+1
\end{aligned}
$$

$\therefore$ result true for k $\Longrightarrow$ result true for $k + 1$

END: $\therefore$ by the Principle of Induction, the result is true for all $n \in \mathbb{N}$ □

Traps and Pitfalls

What is Wrong ?

$$
\begin{aligned}
a(b-c) &= ab - ac\\
ab + a(-c) &= ab - ac\\
a(-c) &= -ac\\
ac + a(-c) &= 0\\
a(c + (-c)) &= 0\\
a.0 &= 0\\
0 &= 0 \qquad \square
\end{aligned}
$$

What is Wrong ?

$$
\begin{aligned}
a(b-c) &= ab + a(-c)\\
&= ab - ac \qquad \square
\end{aligned}
$$

What is Wrong ?

$$
\begin{aligned}
a(b-c) &= ab + a(-c)\\
&= ab + a(-c) + a.1\\
&= ab + a(1-c)\\
&= ab - ac \qquad \square
\end{aligned}
$$

What is Wrong ?

$$
\begin{aligned}
a(b-c) &= ab + a(-c)\\
a(-c) &= -ac
\end{aligned}
$$

because if you add ac to both sides then both sides vanish which means they're inverse

$$\therefore ab + a(-c) = ab - ac \qquad \square$$

Beware Incorrect Logic

- Negating a statement incorrectly
- proving the converse of something instead of the thing itself

$$\forall \varepsilon > 0\ \exists \delta > 0 \text{ s.t. } \forall x \text{ satisfying } 0 < |x - a| < \delta, \quad |f(x) - l| < \varepsilon$$

$$\exists \varepsilon > 0 \text{ s.t. } \forall \delta > 0 \quad \exists x \text{ satisfying } 0 < |x - a| < \delta \quad \text{s.t.} \quad |f(x) - l| \geq \varepsilon$$

Additional Pitfalls

- Incorrect assumptions
- Incorrect use of definitions, or use of incorrect definitions

$$
\begin{aligned}
f(a) = f(a') &\Longrightarrow a = a'\\
g(a) = g(a') &\Longrightarrow a = a'
\end{aligned}
$$

$$
\begin{aligned}
(g \circ f)(a) = (g \circ f)(a') &\Longrightarrow g(a) \circ f(a) = g(a') \circ f(a')\\
&\Longrightarrow a = a'
\end{aligned}
$$

$\therefore$ $g \circ f$ is injective. □

Assumptions

- You need to *justify* everything *enough* for your *peers* to understand it
- If in doubt, *justify* things *more* rather than less

Practicalities

Practicalities

- Write the **begining** very carefully
- Write the **end** very carefully
- Try and manipulate both ends to meet in the middle, from *big* leaps to *smaller* ones
- Pretend to be more *stupid* (or *sceptical*, or *untrusting*) that you are

$x = y$ or "something equals something else"

$$\begin{aligned} x &= a \\ &= b \\ &= c \\ &= d \\ &= y \end{aligned}$$

$$\begin{aligned} x &= a \\ &= b \\ &= c \\ \\ y &= e \\ &= d \\ &= c \\ \\ \therefore x &= y \end{aligned}$$

$x \Longrightarrow y$

$$\begin{aligned} x &\Longrightarrow a \\ &\Longrightarrow b \\ &\Longrightarrow c \\ &\Longrightarrow d \\ &\Longrightarrow y \end{aligned}$$

We know that $a \Longrightarrow b$

Also $a \Longleftrightarrow x$
and $b \Longleftrightarrow y$

$\therefore x \Longrightarrow y$

$x \Longleftrightarrow y$

$$\begin{aligned} x &\Longrightarrow a \\ &\Longrightarrow b \\ &\Longrightarrow c \\ &\Longrightarrow d \\ &\Longrightarrow y \end{aligned}$$

Conversely
$$\begin{aligned} y &\Longrightarrow p \\ &\Longrightarrow q \\ &\Longrightarrow r \\ &\Longrightarrow x \end{aligned}$$

Hence $x \Longleftrightarrow y$

$$\begin{aligned} x &\Longleftrightarrow a \\ &\Longleftrightarrow b \\ &\Longleftrightarrow c \\ &\Longleftrightarrow d \\ &\Longleftrightarrow y \end{aligned}$$

x is purple

"x is purple" means y

We know a and

$$\begin{aligned} a &\Longrightarrow b \\ &\Longrightarrow c \\ &\Longrightarrow d \\ &\Longrightarrow y \end{aligned}$$

$\therefore$ x is purple as required

$\forall x, p(x)$ is true

Prove that any rational number can be expressed as $\frac{m}{n}$ where m and n are integers that are not both even.

Let x be a rational number. So x can be expressed as $\frac{p}{q}$ where p and q are integers and $q \neq 0$.

...

$\exists x$ s.t. $p(x)$ is true

$$\exists\ \delta > 0 \ \text{ s.t. } \ |x| < \delta \Longrightarrow |x^2| < \frac{1}{100}$$

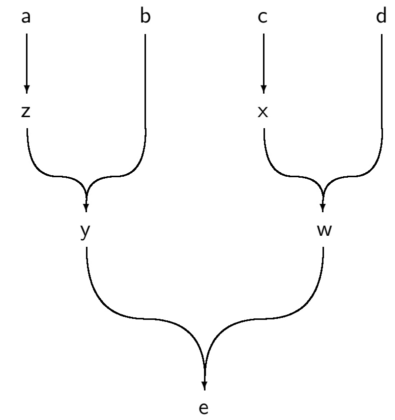
Put $\delta = \frac{1}{10}$. Now $|x^2| = |x|^2$ so we have

$$|x| < \frac{1}{10} \Longrightarrow |x^2| < \frac{1}{100}$$

If a, b, c, d are true then e is true

$$
\begin{array}{ccc}
\text{a} & \Longrightarrow & \text{z} \\
\text{b and z} & \Longrightarrow & \text{y} \\
\text{c} & \Longrightarrow & \text{x} \\
\text{x and d} & \Longrightarrow & \text{w} \\
\text{y and w} & \Longrightarrow & \text{e}
\end{array}
$$

How to write proofs: a quick guide. Eugenia Cheng.
http://www.math.uchcago.edu/~eugenia



How to write proofs: a quick guide. Eugenia Cheng.
http://www.math.uchcago.edu/~eugenia

Proof by Contradiction

- We are trying to prove that some statement *P* is true
- We say «suppose *P* were not true» and find a contradiction
- Since *P* being false gives a contradiction, we deduce that *P* must be true

How to write proofs: a quick guide. Eugenia Cheng.
http://www.math.uchcago.edu/~eugenia

Exploratory Data Analysis

NIST/SEMATECH e-Handbook of Statistical Methods,
http://www.itl.nist.gov/div898/handbook/

Approach

- **Exploratory Data Analysis** employs a variety of (*mostly graphical*) techniques to:
  - maximize *insight* into a data set
  - uncover underlying *structure*
  - extract *important* variables
  - detect *outliers* and *anomalies*
  - test underlying *assumptions*
  - develop parsimonious *models*
  - determine *optimal* factor settings

NIST/SEMATECH e-Handbook of Statistical Methods,
http://www.itl.nist.gov/div898/handbook/

Graphical techniques

- *Plotting the raw data* (data traces, histograms, bihistograms, probability plots, lag plots, block plots, and Youden plots)
- *Plotting simple statistics* such as mean plots, standard deviation plots, box plots, and main effect plots of the raw data
- *Positioning* such plots so as to maximize our natural *pattern recognition* abilities

NIST/SEMATECH e-Handbook of Statistical Methods,
http://www.itl.nist.gov/div898/handbook/

Classical *vs.* Exploratory

Classical Data Analysis

1. Problem
2. Data
3. Model
4. Analysis
5. Conclusion

NIST/SEMATECH e-Handbook of Statistical Methods,
http://www.itl.nist.gov/div898/handbook/

Exploratory Data Analysis

1. Problem
2. Data
3. **Analysis**
4. **Model**
5. Conclusion

NIST/SEMATECH e-Handbook of Statistical Methods,
http://www.itl.nist.gov/div898/handbook/

Classical vs. Exploratory

- **Models**
- **Focus**
- **Techniques**
- **Rigor**
- **Data Treatment**
- **Assumptions**

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Model

- **Classical**
  - *imposes models* (both deterministic and probabilistic). *e.g.* regression models, analysis of variance. The most common probabilistic model assumes that the errors are normally distributed.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Model

- **Exploratory**
  - does not impose deterministic or probabilistic models on the data. In fact, EDA allows the data to suggest admissible models that best fit the data.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Focus

- **Classical**
  - *On the Model*. Estimate model parameters, generate predicted values from the model.
- **Exploratory**
  - *On the Data*. Structure, outliers, and models suggested by the data.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Techniques

- **Classical**
  - *Quantitative*. Mean, Variance, ANOVA, T-test, chi^2 tests, F-Test.
- **Exploratory**
  - *Graphical*. Scatter plots, Character plots, box plots, histograms, bihistograms, probability plots, residual plots, mean plots.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Rigor

- **Classical**
  - Probabilistic *foundation* of Science. Rigorous, formal, objective.
- **Exploratory**
  - Suggestive, indicative, insightful. Subjective, depend on interpretation.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Data Treatment

- **Classical**
  - Maps all data into *few* numbers. Loss of information.
- **Exploratory**
  - Shows *all* data. No loss of information.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Assumptions

- **Classical**
  - Tests based on classical techniques are very sensitive. Yet they depend on underlying assumptions. that could be *unkown* or *untested*.
- **Exploratory**
  - Makes no assumptions.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Quantitative Techniques

- Hypothesis testing
- Analysis of variance
- Point estimate and confidence intervals
- Least squares regression

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Graphical Techniques

- Testing assumptions
- Model Validation
- Estimator Selection
- Relationship identification
- Factor Effect determination
- Outlier Detection

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

EDA Example

EDA Example

X	Y
10.00	8.04
8.00	6.95
13.00	7.58
9.00	8.81
11.00	8.33
14.00	9.96
6.00	7.24
4.00	4.26
12.00	10.84
7.00	4.82
5.00	5.68

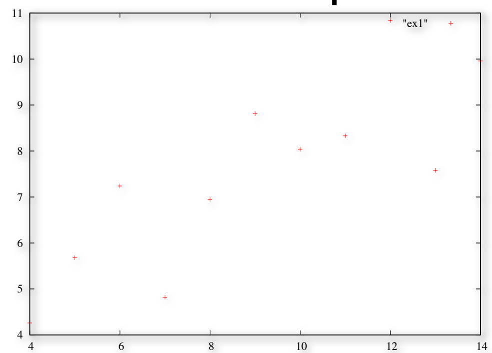
NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

EDA Example (DS1)

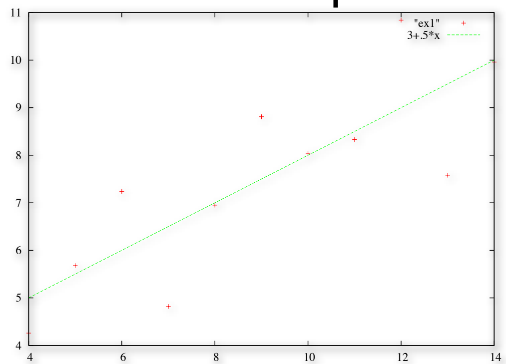
- N = 11
- Mean of X = 9.0
- Mean of Y = 7.5
- Intercept = 3
- Slope = 0.5
- Residual Standard Deviation = 1.237
- Correlation = 0.816

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

EDA Example



EDA Example



EDA Example

X2	Y2	X3	Y3	X4	Y4
10.00	9.14	10.00	7.46	8.00	6.58
8.00	8.14	8.00	6.77	8.00	5.76
13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.77	9.00	7.11	8.00	8.84
11.00	9.26	11.00	7.81	8.00	8.47
14.00	8.10	14.00	8.84	8.00	7.04
6.00	6.13	6.00	6.08	8.00	5.25
4.00	3.10	4.00	5.39	19.00	12.50
12.00	9.13	12.00	8.15	8.00	5.56
7.00	7.26	7.00	6.42	8.00	7.91
5.00	4.74	5.00	5.73	8.00	6.89

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

EDA Example (DS2)

- N = 11
- Mean of X = 9.0
- Mean of Y = 7.5
- Intercept = 3
- Slope = 0.5
- Residual Standard Deviation = 1.237
- Correlation = 0.816

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

EDA Example (DS3)

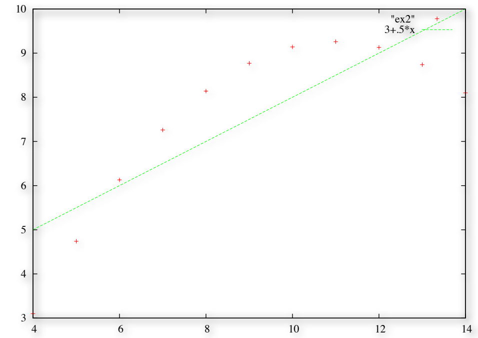
- N = 11
- Mean of X = 9.0
- Mean of Y = 7.5
- Intercept = 3
- Slope = 0.5
- Residual Standard Deviation = 1.236
- Correlation = 0.816

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

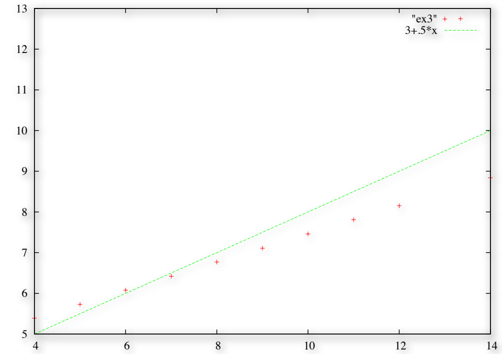
EDA Example (DS4)

- N = 11
- Mean of X = 9.0
- Mean of Y = 7.5
- Intercept = 3
- Slope = 0.5
- Residual Standard Deviation = 1.236
- Correlation = 0.817

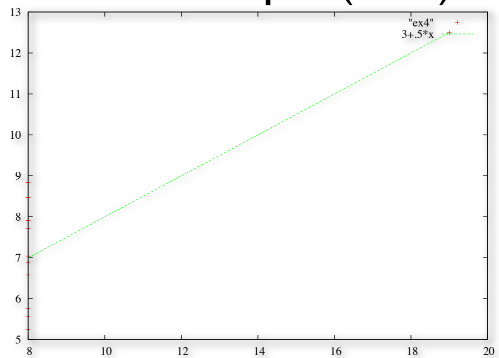
EDA Example (DS2)



EDA Example (DS3)



EDA Example (DS4)



Four Basic Tools

Univariate Data

- Most basic tools operate on *univariate* data, *i.e.* a list of *single* responses

Data Sets

- **Flow DS**: This data set was collected by Bob Zarr of NIST in January 1990 from a heat flow meter calibration and stability analysis. The response variable is a calibration factor.

Data Sets

- **Walk DS**: A random walk can be generated from a set of uniform random numbers by the formula :

$$R_i = \sum_{j=1}^{i} (U_j - 0.5)$$

- where U is a set of uniform random numbers

Data Sets

- **Beam DS**: This data set was collected by H.S. Lew of NIST in 1969 to measure steel-concrete deflections. The response variable is the deflection of a beam from center point.

Run-sequence Plot

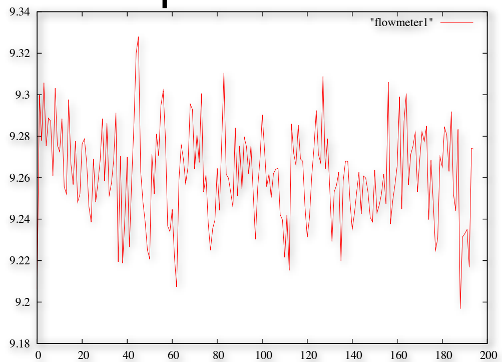
- Considers *Univariate* Data
- **Vertical axis:** response variable Y(i)
- **Horizontal Axis:** Index i (i=1,2,3,...)

Run-sequence Plot

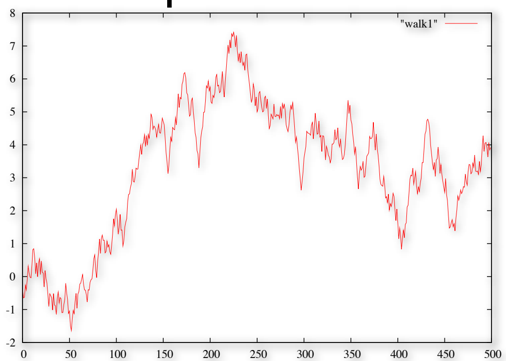
- Used to answer the questions
  - Are there any *shifts* in *location* ?
  - Are there any *shifts* in *variation* ?
  - Are there any *outliers* ?

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

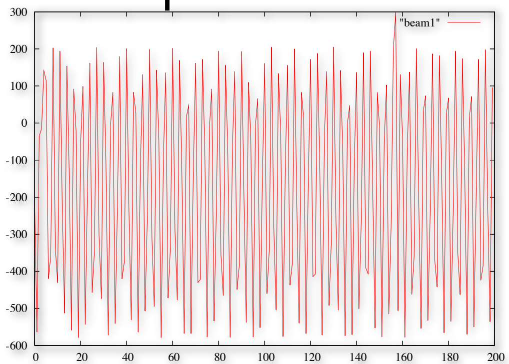
Run-sequence Flow DS



Run-sequence Walk DS



Run-sequence Beam DS



Lag Plot

- Considers *univariate* data
- **Vertical Axis**: *Y(i)* for all *i*
- **Horizontal Axis**: *Y(i-1)* for all *i*

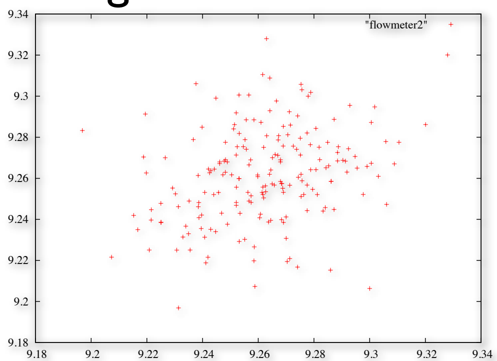
NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Lag Plot

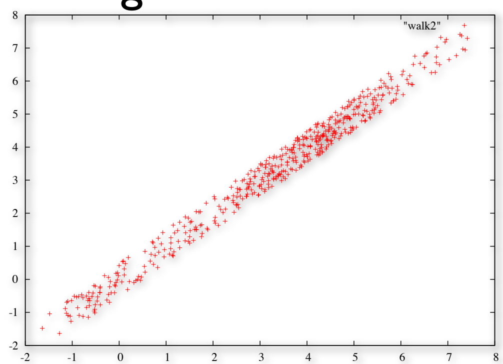
- Are the data *random* ?
- Is there *serial correlation* in the data ?
- What is a suitable *model* for the data ?
- Are there *outliers* in the data ?

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

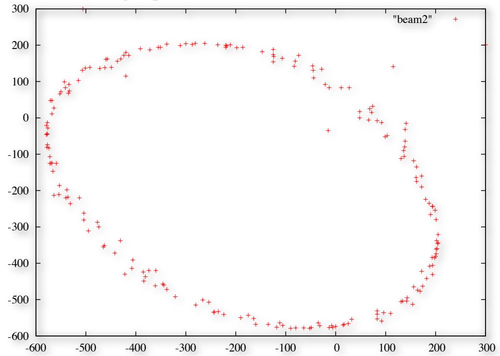
Lag Plot Flow DS



Lag Plot Walk DS



Lag plot Beam DS



Histogram

- Considers univariate data
- Split the range of the data into equal-sized bins, then for each bin the number of points from the data for each bin are counted
- **Vertical axis:** Frequency
- **Horizontal axis**: Response variable

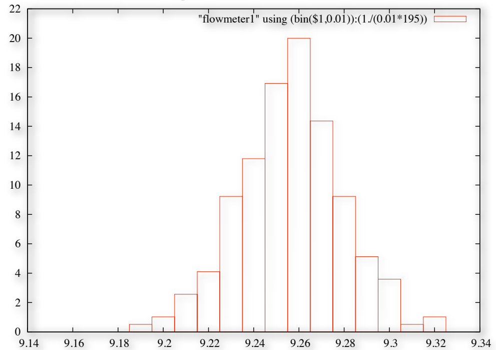
NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

Histogram

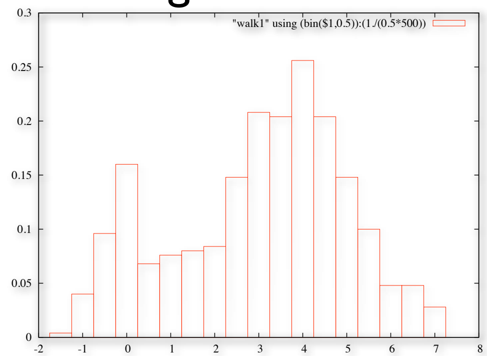
- Used to answer the following questions
  - What kind of population do the data come from ?
  - Where are the data located ?
  - How spread out are the data ?
  - Are the data symmetric or skewed ?
  - Are there outliers in the data ?

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

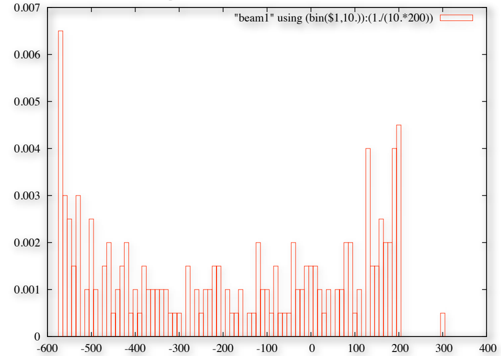
Histogram Flow DS



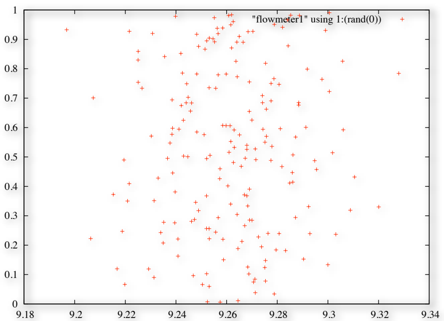
Histogram Walk DS



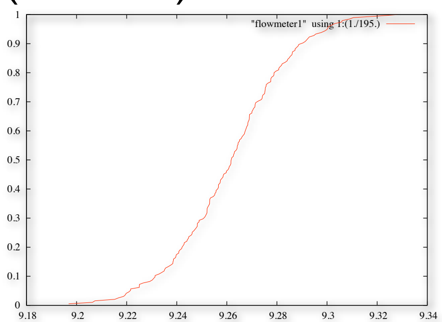
Histogram Beam DS



Beyond Histograms : Jitter Plots



Beyond Histograms : (Normal) Cumulative



(Normal) Probability Plot

- Considers univariate data
- **Vertical axis:** Ordered Response values
- **Horizontal axis:** Normal order statistics median

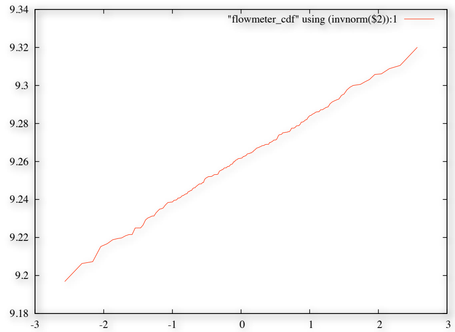
NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

(Normal) Probability Plot

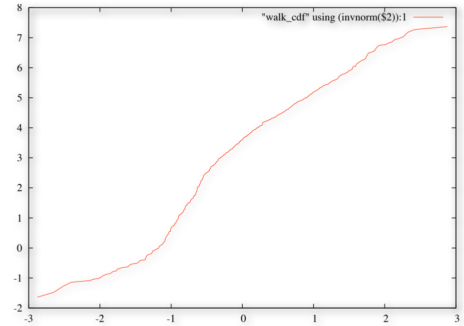
- Used to answer the following questions:
  - Are the data normally distributed ?
  - What is the nature of the departure from normality (data skewed, shorted than expected tail, longer than expected tails, etc.) ?

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/

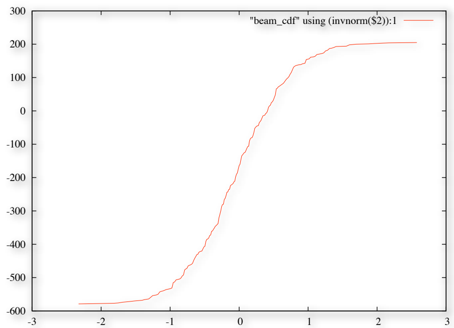
(Normal) Probability Plot Flow DS



(Normal) Probability Plot Walk DS



(Normal) Probability Plot Beam DS



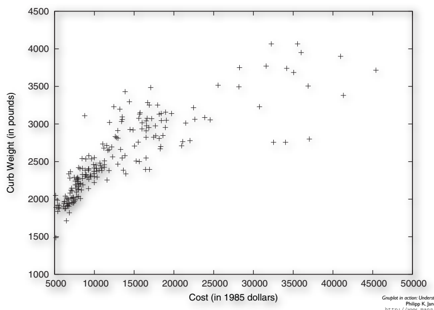
Investigating Relationships

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

Scatter plots

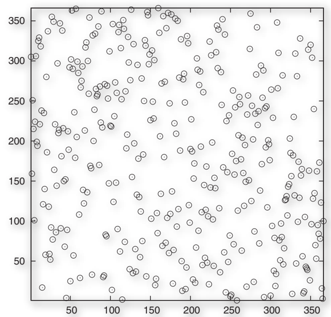
- Assumes *Bivariate* data, i.e. lists of 2-tuples of responses
- The point is to check the nature of the *relationship* between the two responses
- Take care of *outliers*

Example 1: Weight vs. Cost



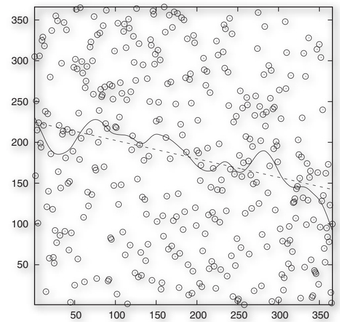
Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

Example 2: The 1970 Draft Lotery



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

Example 2: The 1970 Draft Lotery

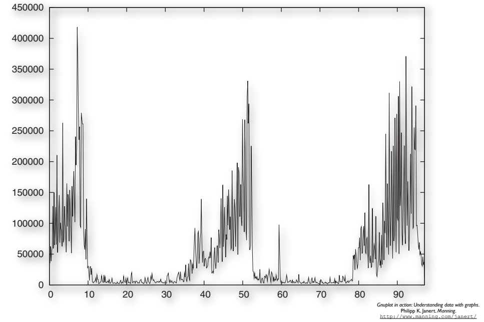


Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

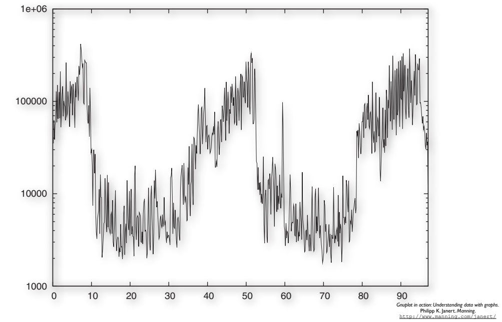
Logarithmic Scale

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

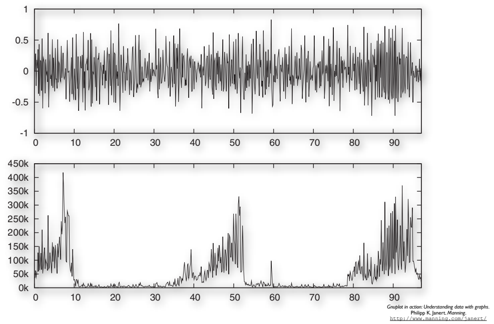
Example 1: Traffic Pattern at Website



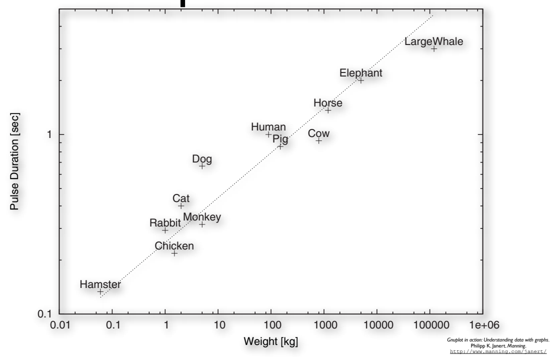
Example 1: Traffic Pattern at Website



Example 1: Traffic Pattern at Website



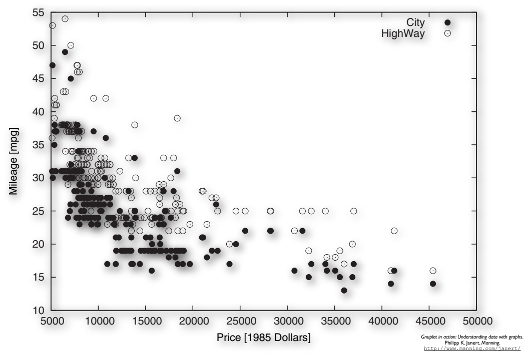
Example 2: Mammals



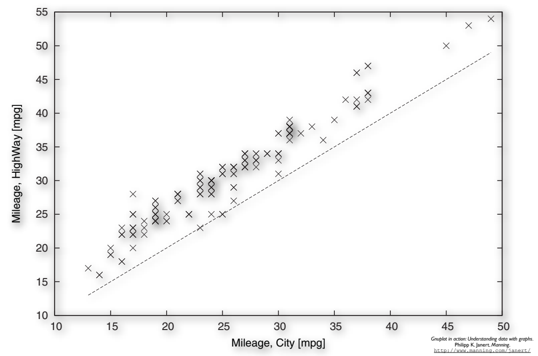
The Core Principle

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

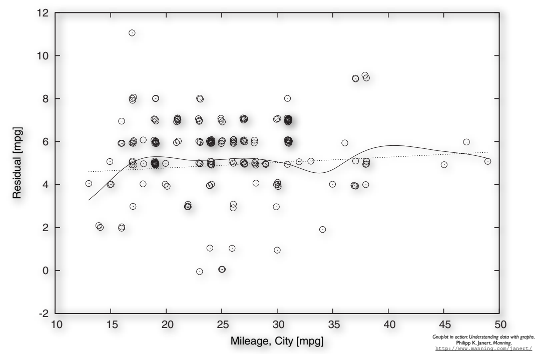
Iterate & Transform



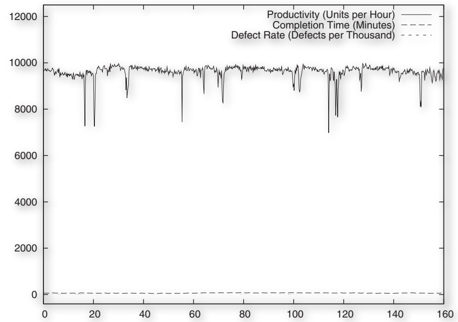
Iterate & Transform



Iterate & Transform

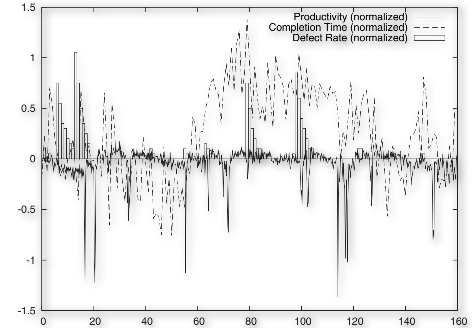


What's Wrong ?



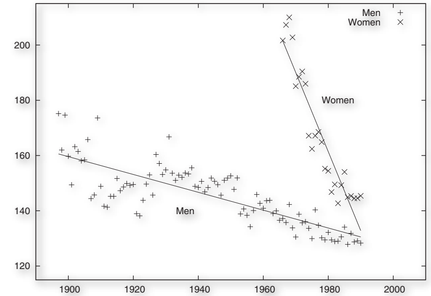
Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

Normalized Metrics



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

Truncation & Responsiveness

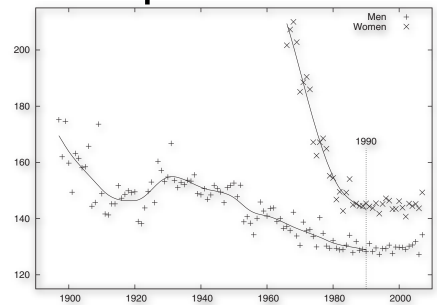


Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

Truncation & Responsiveness

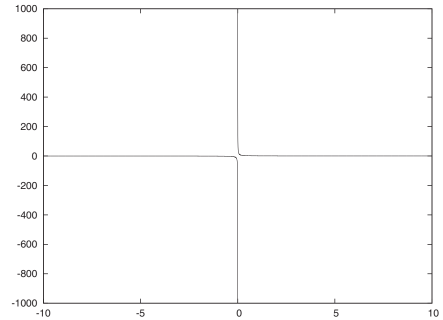
- Outlier removal
- Sampling bias
- Edge effects

Truncating & Responsiveness



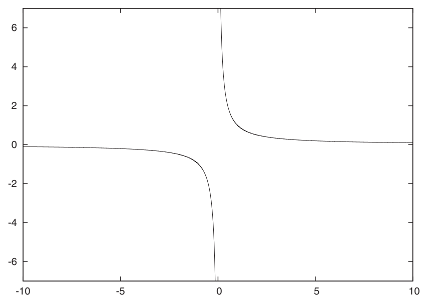
Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

Improving Perception



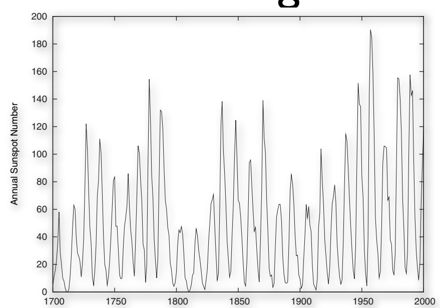
Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

Improving Perception: Banking



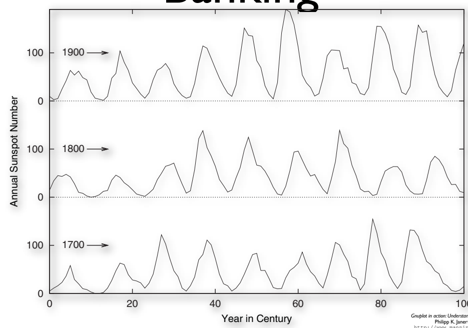
Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

Improving Perception: Banking



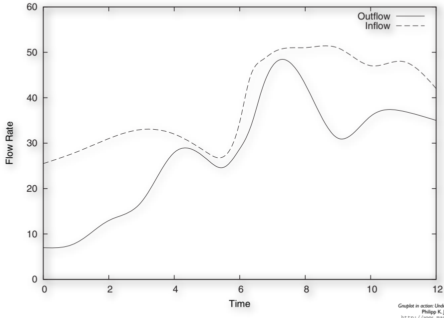
Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

Improving Perception: Banking



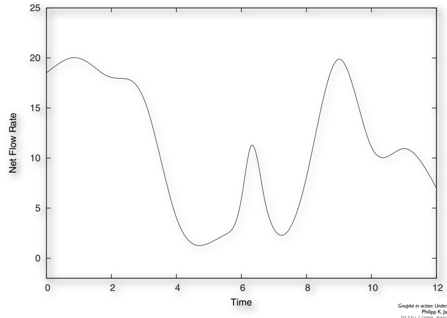
Gnuplot in action: Understanding data with graphs.
Philipp K. Janert. *Manning.*
http://www.manning.com/janert/

Judging Lengths and Distances



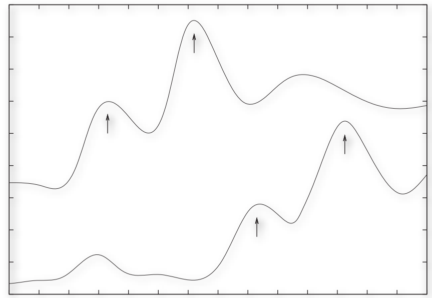
Gnuplot in action: Understanding data with graphs. Philipp K. Janert. *Manning.* http://www.manning.com/janert/

Judging Lengths and Distances



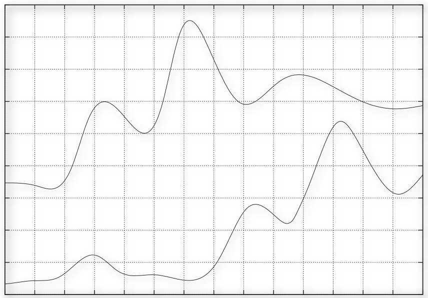
Gnuplot in action: Understanding data with graphs. Philipp K. Janert. *Manning.* http://www.manning.com/janert/

Enhancing Quantitative Perception



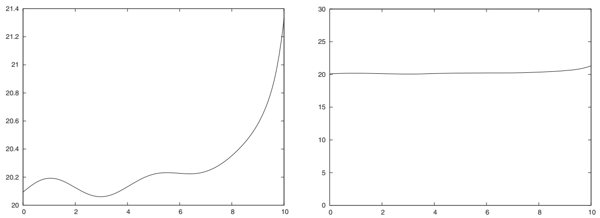
Gnuplot in action: Understanding data with graphs. Philipp K. Janert. *Manning.* http://www.manning.com/janert/

Enhancing Quantitative Perception



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

Plot Ranges ?



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

The Core Principle

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

GNUPLOT 101

http://www.gnuplot.info

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

GNUPLOT

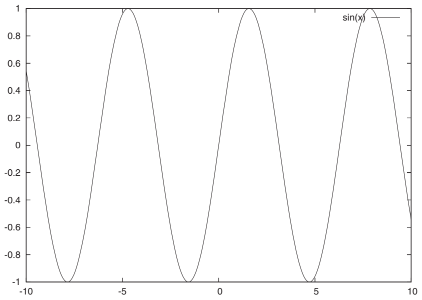
- **Free** software for plotting data
- **NOT** «push-button-limited-capacities» type of software
- **Multiplatform**
- Integrates well with ***LaTeX***

GNUPLOT Invocation

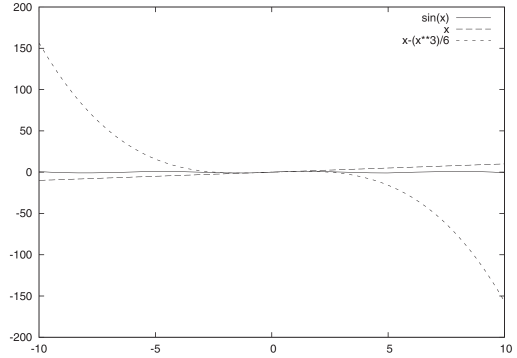
```
Mac-Pro:metho tixeuil$ gnuplot

    G N U P L O T
    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)
```

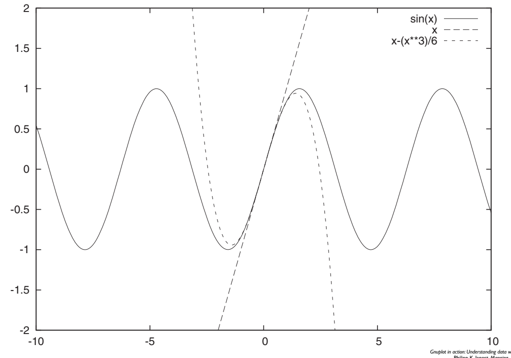



```
plot sin(x), x, x-(x**3)/6
```



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
```



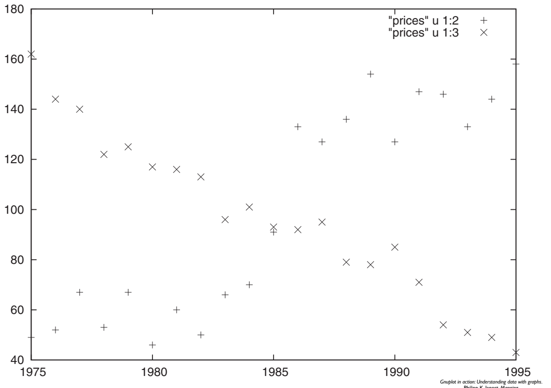
Gnuplot in action: Understanding data with graphs. Philipp K. Janert. Manning. http://www.manning.com/janert/

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
```

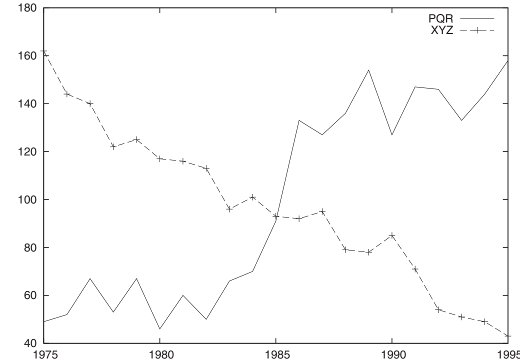
Gnuplot in action: Understanding data with graphs. Philipp K. Janert. Manning. http://www.manning.com/janert/

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



Gnuplot in action: Understanding data with graphs. Philipp 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. Philipp 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
```

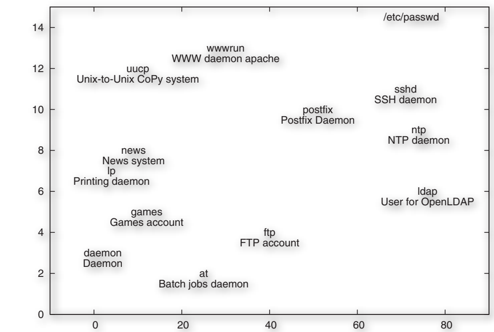
Gnuplot in action: Understanding data with graphs. Philipp K. Janert. Manning. http://www.manning.com/janert/

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/sshd:/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. Philipp 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
```



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

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.

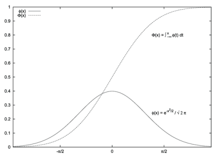


Figure 1: A Postscript file, included in LaTeX

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.

1

Implementing EDA 4BT

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

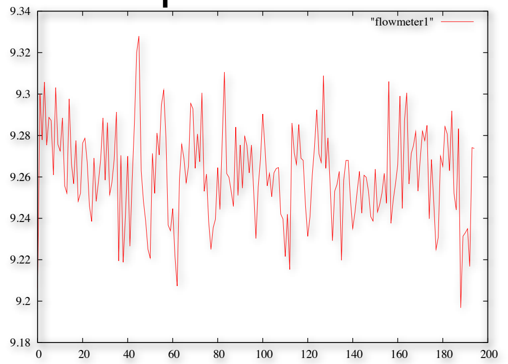
Run-sequence Plot

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

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

flowmeter1

```
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
...
```

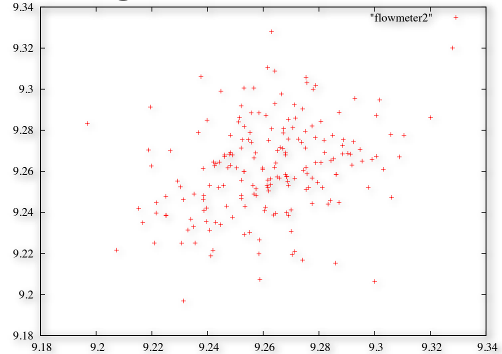
Lag Plot

```
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



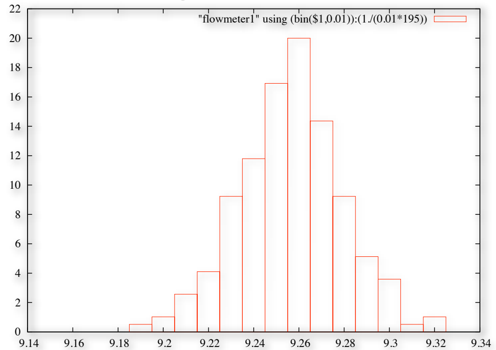
Histogram

```
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

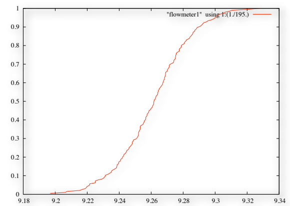
```
9.206343
9.299992
9.277895
9.305795
9.275351
9.288729
9.287239
9.260973
...
```

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
...
```

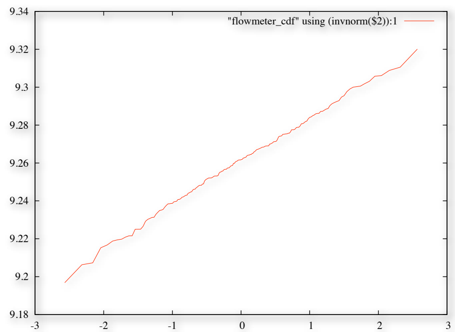
(Normal) Probability Plot

```
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



Homework 6

Deadline:
Tuesday 6 Nov., 17:00

- Step 1: Say something *meaningful* about the data located at the METHO class web site:
  - **Format**:
    - node x, node y, begin contact, end contact
- One page limit, One graph limit

Deadline:
Tuesday 6 Nov., 17:00

- **Step 2**: download *some* experimental data:
  - http://sensorscope.epfl.ch/index.php/Environmental_Data
  - http://fta.inria.fr/apache2-default/pmwiki/index.php?n=Main.DataSets

Can you *confirm/invalidate* associated publications ?
Can you extract *new insight* on the data ?

No page limit, no graph limit, include source code

Step 1 + Step 2 = One PDF