PD Dr. rer. nat. Stefan Bosse

Universität Bremen - FB Mathematik und Informatik

Physiker & Informatiker

About the Lecturer

Bremen

> **Privatdozent im Fachbereich Mathematik und Informatik, Universität Bremen**

Bremen

> **Privatdozent im Fachbereich Mathematik und Informatik, Universität Bremen**
>
> 2016: **Habilitation** mit der Venia Legendi für die Informatik

Bremen

> **Privatdozent im Fachbereich Mathematik und Informatik, Universität Bremen**
>
> 2016: **Habilitation** mit der Venia Legendi für die Informatik
>
> 2018-2019: **Vertretungsprofessur Praktische Informatik** und Lehrbeauftragter in der Fakultät Informatik, Universität Koblenz-Landau

Bremen

> **Privatdozent im Fachbereich Mathematik und Informatik, Universität Bremen**
>
> 2016: **Habilitation** mit der Venia Legendi für die Informatik
>
> 2018-2019: **Vertretungsprofessur Praktische Informatik** und Lehrbeauftragter in der Fakultät Informatik, Universität Koblenz-Landau
>
> 2020-heute: Wissenschaftlicher Leiter und Projektleiter in der transregionalen **DFG Forschungsgruppe** 3022, *Datengetriebene Schadensdiagnostik mit ML*

Bremen

Privatdozent im Fachbereich Mathematik und Informatik, Universität Bremen

2016: **Habilitation** mit der Venia Legendi für die Informatik

2018-2019: **Vertretungsprofessur Praktische Informatik** und Lehrbeauftragter in der Fakultät Informatik, Universität Koblenz-Landau

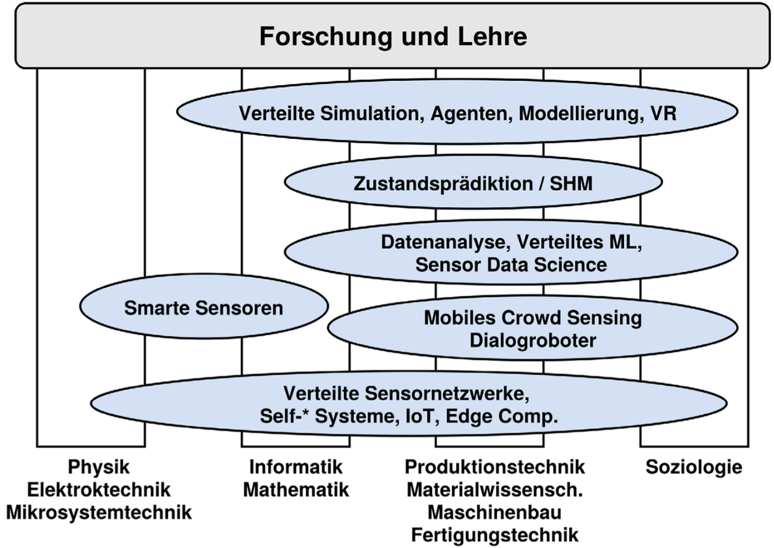
2020-heute: Wissenschaftlicher Leiter und Projektleiter in der transregionalen **DFG Forschungsgruppe** 3022, *Datengetriebene Schadensdiagnostik mit ML*

2022-heute: Wissenschaftlicher Leiter und Projektleiter in der **U Bremen Research Alliance AI Center for Health Care** (Add. Fertigung und Prozessoptimierung)

Siegen

> **Lehrbeauftragter im Fachbereich Msschinenbau, Lehrstuhl für Materialkunde und Werkstofftechnik, Universität Siegen**

Teaching and research topics



My Research Topics

> Sensor Data Sciences

- Material-integrated sensor systems (the machine or the material as a sensor)
- Sensor networks
- Image processing and sensor fusion
- Mobile data collection and mobile crowdsensing (the human being as a sensor)
- Process optimization

Research topics

> Distributed and Applied AI

- Agents and multi-agent systems (computation with agents, ABX)
- Machine Learning and Data Mining
- Ensemble and Distributed Machine Learning
- Structural Health Monitoring
- Chat Bots

Research topics

Simulation

- Agent-based Simulation (ABM/ABS)
- Coupling of real and virtual worlds (Augmented Virtuality, ABS+ABX)
- Parallel and distributed simulation methods and architectures (ABS/ABX)

Automated Feature Extraction with Machine Learning and Image Processing

Automatische Schadensdiagnostik in der Werkstofftechnik

Lecture and Winter School

PD Stefan Bosse

University of Siegen - Dept. Maschinenbau
University of Bremen - Dept. Mathematics and Computer Science

Introduction

Target Audience

- Materials Science and Engineering
- Mechanical Engineering
- Vehicle construction
- Industrial engineering
- International Production Engineering and Management

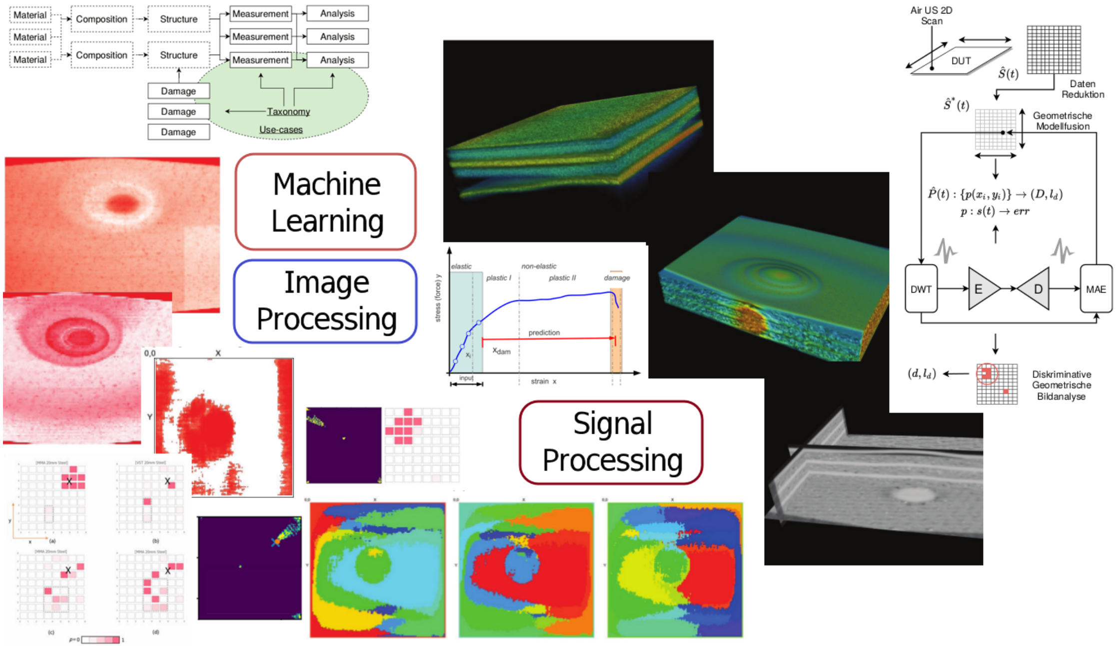
Student curriculum module area: Bachelor and Master Specialization

In addition to lecture students, scientific workers and thesis students are invited to participate to the winter school.

Main topics in this course

Metrics, taxonomies, methods, algorithms of ...

- Sensors in materials testing technology, digital sensor data, multidimensional and time-dependent sensor data
- Sensor data acquisition and processing (testing and measuring methods in materials testing technology and materials science)
- Features and feature analysis in image data
- Machine learning and automatic feature diagnostics (robustness, explainability, noise)
- Algorithms and models (especially for image data)
- Applications, demonstrations, examples, Digital laboratory exercises (integrated) with real and artificial measurement data from materials testing technology (X-ray CT, tensile tests, micrographs from the Add. Manufacturing, US time signals 2D/3D, and many more from YOU!)


Material
Material
Material
Composition
Composition
Structure
Structure
Measurement
Measurement
Measurement
Analysis
Analysis
Analysis
Damage
Damage
Damage
Taxonomy
Use-cases
Machine Learning
Image Processing
Signal Processing
elastic
non-elastic
plastic I
plastic II
damage
prediction
strain x
Air US 2D Scan
DUT
$\hat{S}(t)$
Daten Reduktion
$\hat{S}^*(t)$
Geometrische Modellfusion
$\hat{P}(t) : \{p(x_i, y_i)\} \rightarrow (D, l_d)$
$p : s(t) \rightarrow err$
DWT
E
D
MAE
(d, l_d)
Diskriminative Geometrische Bildanalyse

Special topics in this course

> ⚠ X-ray Radiography and X-ray Tomography

- From physical measurements to material structure analysis
- Image processing (i.e., for 2D and 3D data volumes) is discussed with practical lessons using real-world data
- Tomography fundamentals are introduced with practical applications and lessons using real-world data (including data from Siegen's Zeiss Xradia MicroCT device)
- Microscopy Imaging

Fields of Application

- Destructive and Non-destructive Testing in general
- Structural Health **Monitoring** (not addressed in this course)
- Structural Testing
- Material Design and Testing
- Processes and production (control and optimization)

? Your application?

Organisation

1. Introduction to basic methods and algorithms with simple "sandbox" tutorials and exercises (partially home/office work)

2. Discussion of scientific application cases and available data

3. Application of methods and algorithms to different complex application cases

4. Discussions and presentation of results

Data

What kind of data we have to process?

1. Input Data ***D*** ⇒ Measurements
2. Meta Data ***M*** ⇒ Context
3. Intermediate Data ***R*** ⇒ Preprocessed
4. Output Data ***F*** ⇒ Features

$$F(D, M) : \hat{D} \times \vec{M} \rightarrow \begin{cases} R \\ F \end{cases}$$

$$G(R) : \hat{R} \rightarrow \hat{F}$$

Data

Input Data

1. Dimension
    - 1D: Time-dependent vibration or guided ultrasonic wave signals measured at a specific geometrical position
    - 2D: Images (single projection, visible light, infrared, X-ray)
    - 3D: Time-dependent 2D images (videos) or 3D spatial data (multi-projection tomography, LIDAR, Coupled Ultrasonic, (Ultra)Sonography, Air-guided Ultrasonic)
    - 4D: Time-dependent 3D spatial data (tomography videos)

Data

2. Variable Class
    - Metric/Numerical, continuous (e.g., temperature)
    - Metric/Numerical, discrete (e.g., number of items)
    - Categorical/Symbolic (e.g., damage class)
    - Categorical/Interval (e.g., range intervals)
3. Size
    - One data value
    - 1000 data values
    - 1M data values
    - ..
4. Temporal Characteristics
    - Periodic
    - Sporadic
    - One-shot

Data

Your data?

Data

Meta Data

- Any experiment is situated in a context:
  - Date, Place
  - Set-up
  - Operator
  - Environmental conditions (temperature)
  - Measuring technology, device, parameters
- The device under test (DUT)
  - Dimensions
  - Material
  - ..
- Data formats and structure!

Data Sources and Storage

Storage

1. Files and File system
   - Linear Data Model
   - Tree Structure
2. Web Storage (HTTP)
   - Tree Structure
3. File Cloud Services (Seafile)
   - Tree Structure
4. Databases (SQL, NoSQL, ..)
   - Table Structure
5. Coded Data Structure in Files
   - File system in Files

Format and Coding

1. JavaScript Object Notation (JSON)
2. Comma Separated Values (CSV)
3. Yet Another Meta Language (YAML)
4. Numerical Python (numpy)
5. Hierarchical Data Format (HDF5)
6. Portable Network Graphics (PNG)
7. JPEG, BMP, TIFF, ..

Data Examples

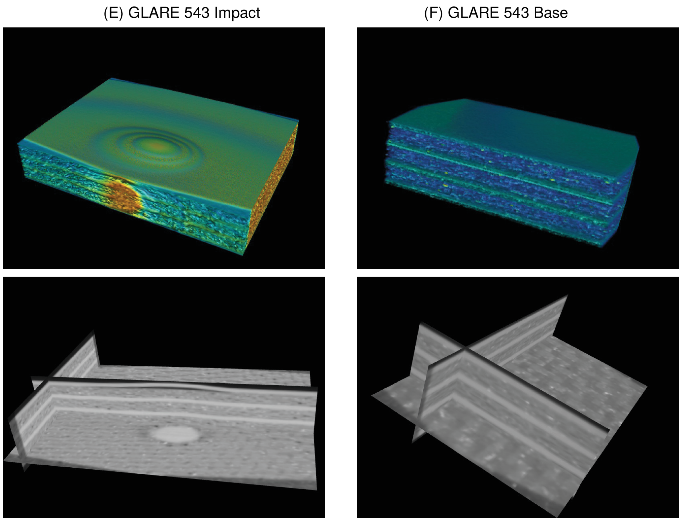


Fig. 1. Anomaly detection in 3D X-ray CT image volumes (composite materials with defects)

Materials 2022, 15(13), 4645; https://doi.org/10.3390/ma15134645

Data Examples

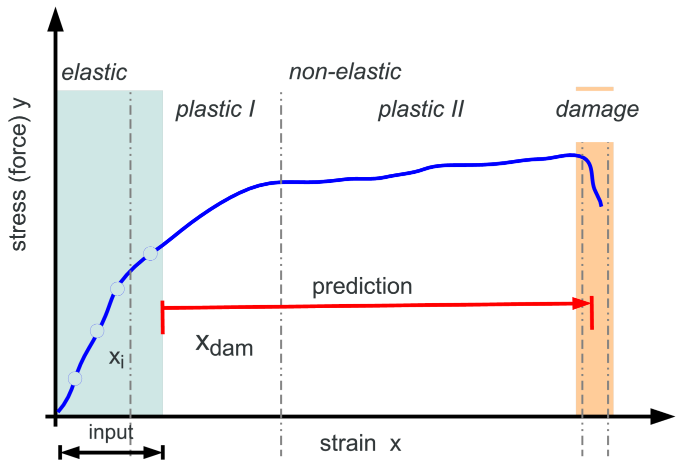


Fig. 2. Breakage prediction with tensile test data (metal probes)

SciForum ECSA 2020, https://doi.org/10.3390/ecsa-7-08279

Data Examples

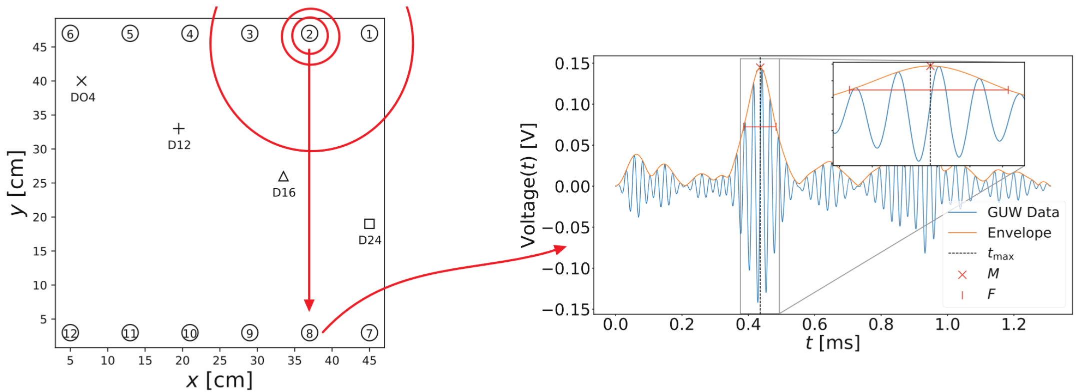


Fig. 3. Damage detection with Guided Ultrasonic Waves (composite material with pseudo defects)

SysInt 2022, DOI: 10.1007/978-3-031-16281-7_35

Data Examples

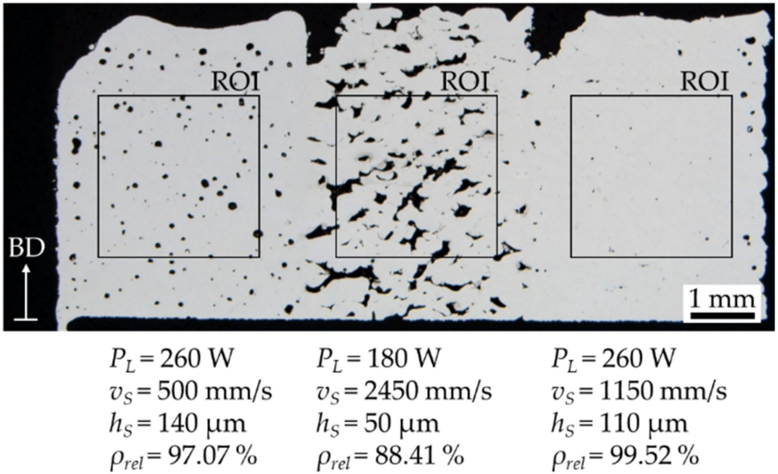
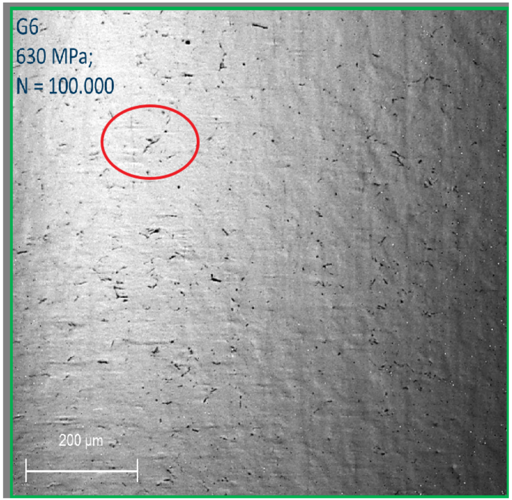


Fig. 4. Porosity analysis with metallurgical microphotographs of sliced surfaces (additive manufacturing)

Materials 2022, 15(20), 7090; https://doi.org/10.3390/ma15207090

Data Examples



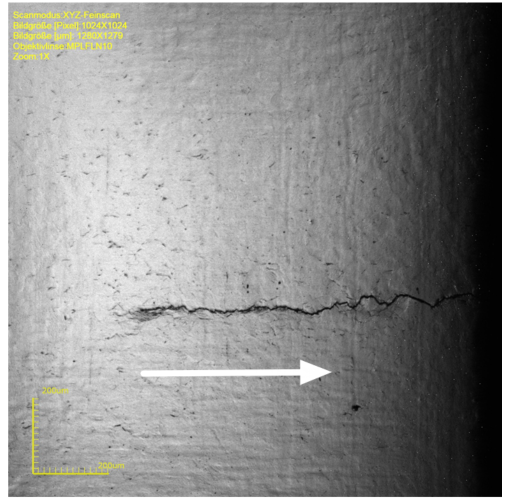


Fig. 5. Feature marking of micro cracks and crack propagation of metal probes after applying stress

AG Brandt, U Siegen

Features

A feature is any kind of aggregated or condensed information from data.

- Categorical features, e.g., damage classes, machine states, etc.
- Metric/numerical features, e.g., length of a crack, spatial position, time, temperature
- Regions of interest (labelled polygons)
- Anomaly

The output of a ML model is typically a categorical or metric feature variable.

Features

There are input and output features.

Input signal features

Selected features from the raw signal data that pose a strong correlation with the output target features $\Rightarrow$ Transformation of signal data to **Intermediate variables**

Examples: Statistical features of time data series (average, deviation, skewness, ..), frequency spectrum, ROI, wavelet transformation coefficients.

Output target features

The final relevant information that is obtained from the data, i.e., the answer to a question.

Features

Your features?

Methods and Algorithms

A typical "Data Mining" work flow consists of different stages:

1. Data pre-processing (e.g., filtering)
2. Signal feature selection (what are the best/strongest intermediate variables) and their computation
3. Selecting a predictor and analysis model (e.g., an artificial neural network or a decision tree)
4. Fit the model to training data
5. Apply the model to test data
6. validation and improvement.

> ⚠ Different data processing methods and algorithms are used in the different stages.

Methods and Algorithms

Statistical Methods

Statistical analysis of signal data and computation of statistical aggregate variables. Mostly applied to data vectors (data series) or matrix data (images) $\Rightarrow$ **Horizontal Analysis**

Algebraic and geometrical Methods

Analysis of data variables with respect to correlation strength and information content, e.g., Principle Component Analysis, with optional following data transformation $\Rightarrow$ **Vertical Analysis**

Transformation Methods

Transformation of data spaces, e.g., time-to-frequency (fourier) or time-to-time/frequency (wavelet) spaces. Image processing applies often kernel-based transformations, e.g., contrast amplification, edge detection.

Methods and Algorithms

Regression Methods

A numerical parametrized model (e.g., polynomial function) is fitted to labelled training data ⇒ **Supervised Training**

Selection Methods

A tree structure is created from labelled training data ⇒ **Supervised Training**

Clustering Methods

A tree or functional graph structure is created from unlabelled training data ⇒ **Unsupervised Training** and grouping of experiments.

Methods and Algorithms

In general, data processing and machine learning is a functional composition.

- ML searches mapping functions $F(X)$: $X \rightarrow Y$.
- ML is commonly an approximation of a hypothesis model $H \approx F$ of an unknown real (world) model M
- There are infinite approximations (hypothesis models), the so called **model space**

Methods and Algorithms

- A model function that maps *X* on *Y* is commonly parametrized by *P* (**parameter space**):

$$F(X,P) : X \times P \rightarrow Y$$

$$arg \min_P (Err)$$

$$Err = |H - M|, H = F$$

- The structure of the model function itself can be parametrized by *S*, e.g., the degree of a polynomial function:

$$F(X,P)_S : X \times P \rightarrow Y$$

$$P = \{p_0, p_1, .., p_S\}$$

$$F_i = p_0 + \sum_i p_i x^i$$

Methods and Algorithms

Toolbox

- Signal filtering and transformation (FFT, DWT, Hilbert and analytical signal,..)
- Image Processing (Edge detection, clustering, filtering, binarization ,..)
- Statistical analysis (basics, PCA, ..)
- Machine Learning, supervised, unsupervised (Decision trees, regression, art. neural networks, LSTM, Self-organizing Maps, ..)
- Agent-based approaches (pattern search, cellular automata)

Methods and Algorithms

Your methods?

Software

The Web Browser is the Laboratory!

1. WorkBook (fully self-contained HTML file) for the Web browser or node-webkit)
2. WorkShell (node.js worker, terminal console, optional)
3. SQLite3 Database (node.js)
4. wex (local access to file system)
5. Additional plug-ins (for WorkBook and WorkShell)
6. Notebook (fully self-contained digital lessons for the Web browser)

Software

Your software?

Software Architecture

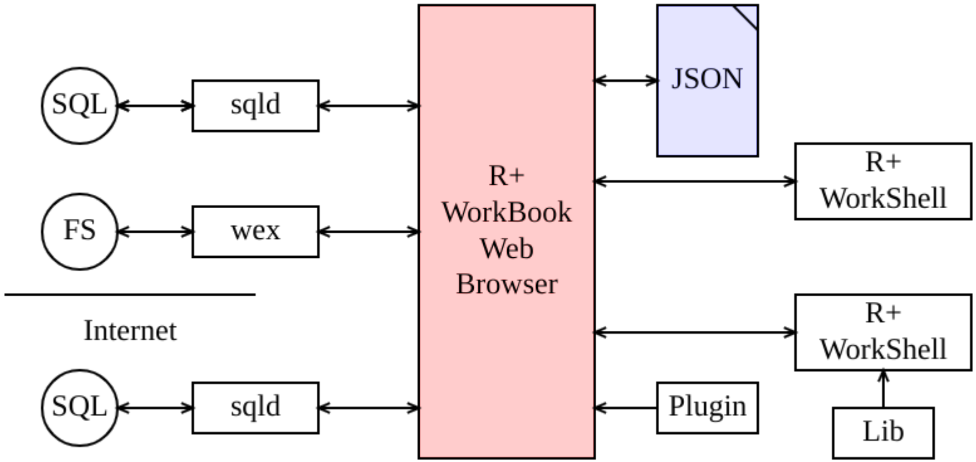


Fig. 6. Work flow architecture

Toolbox

- The WorkBook and WorkShell software provide a broad range of methods and algorithms, partially using well-known software packages:
  - Image processing with OpenCV.js (https://docs.opencv.org/3.4/d5/d10/tutorial_js_root.html)
  - Image processing with image.js (https://github.com/image-js/image-js)
  - Artificial Neural networks with neataptic (https://wagenaartje.github.io/neataptic/)
  - Convolutional Neural Networks with convnet.js (https://cs.stanford.edu/people/karpathy/convnetjs/)

... andy many more ...

- But most of all: The numerical **R** programming language, widely used in science and industry (not only for statistical analysis of table data)

R+

> R+ is an R dialect with a run-time environment entirely programmed in JavaScript and hence usable in any Web browser.

- Introduction to R programming
  - Basic data structures
  - Basic computational statements
  - Advanced computational statements
  - Control statements
  - Input and Output operations
  - Image processing
- R+ extends the core R syntax (e.g., simplified list, vector, and matrix constructors)

R+

> ⚠ KISS: Keep it simple and safe and focus on algorithms and methods, not programming!

```
use math,plot
m = matrix(runif(100),10,10)
plot(m,auto.scale=TRUE)
```