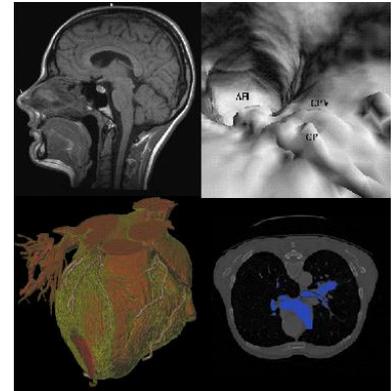
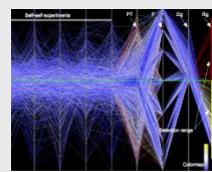


Informations- visualisierung

Thema:	8. Spezifische Verfahren - Histogramme und Scatterplots
Dozent:	Prof. Dr. Geric Scheuermann scheuermann@informatik.uni-leipzig.de
Sprechstunde:	nach Vorlesung
Umfang:	2
Prüfungsfach:	Modul Fortgeschrittene Computergraphik Medizininformatik, Angewandte Informatik





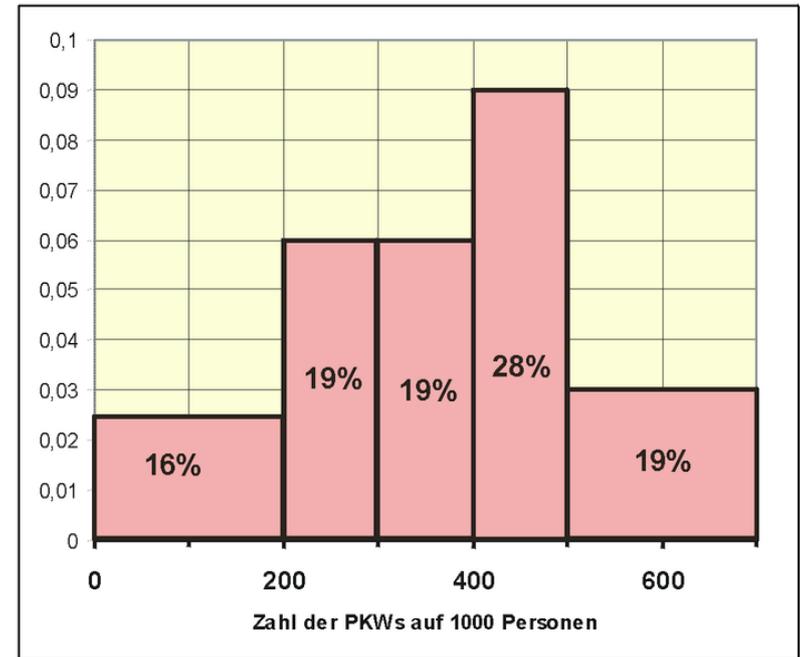
8. Spezifische Verfahren

Verfahren für die Darstellung von Tabellen

- Histogramme
- Scatterplots
- Treemap
- Stacked Display / Dimensions
- Scatterplot Matrix
- Parallele Koordinaten

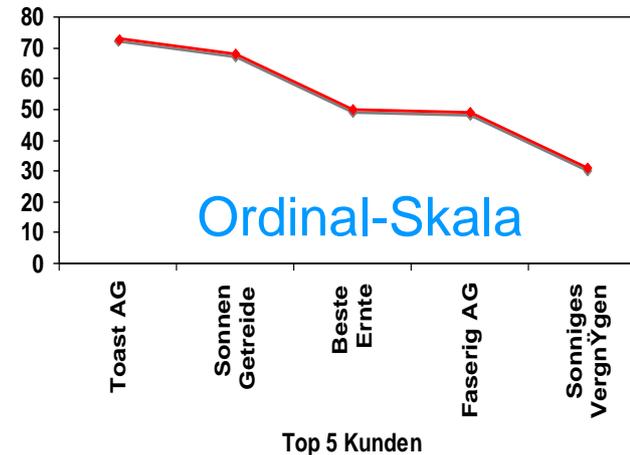
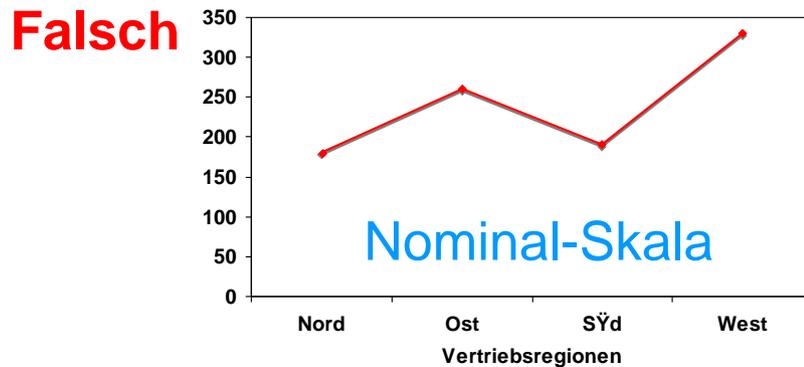
8.1 Histogramme

- Balkendiagramm oder BarChart
- Sind **Häufigkeitsverteilungen**
- Abszisse: **Werteverlauf** / Kategorien
- IdR. **geordnete** Kategorien
- Ordinate: **Häufigkeit** der Kategorie
- Je mehr Kategorien, desto mehr **pseudokontinuierlicher** Verlauf
- Kontinuierlicher Werteverlauf
 - **Diskretisierung**
 - Erfordert **Klassifikation** in „Eimer“ bzw. „Buckets“
 - **Gleichmäßige** oder **willkürliche** Klassifikation
- Klassenbreite kann / soll Größe des Wertebereichs entsprechen: **Häufigkeitsdichte**

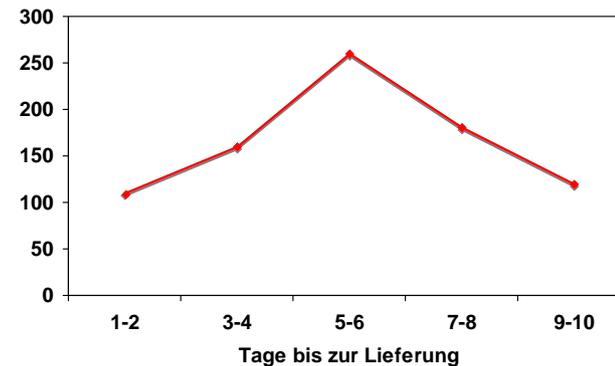
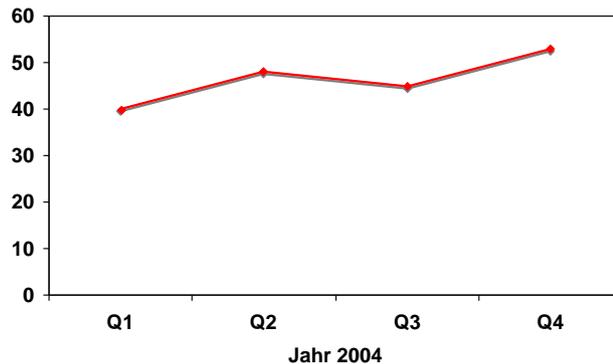


8.1 Histogramme

- Variante: Liniengraph
- Wann sind entsprechende Skalen **geeignet**?

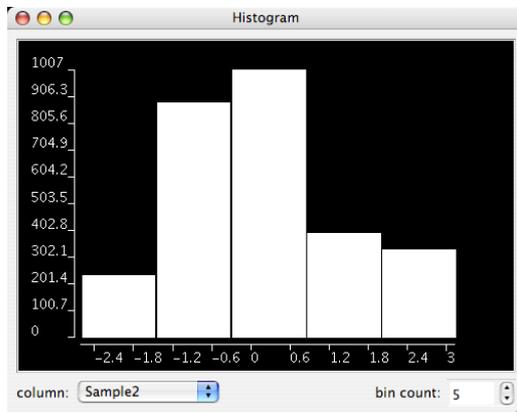


Falsch

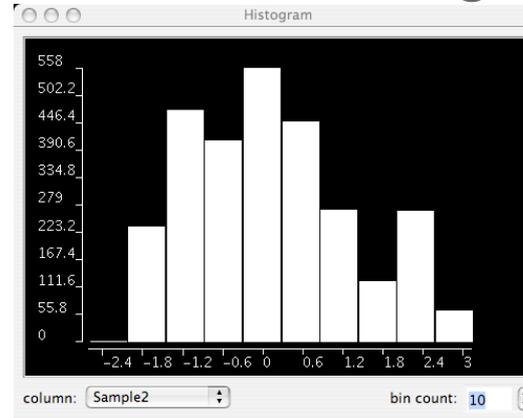


8.1 Histogramm

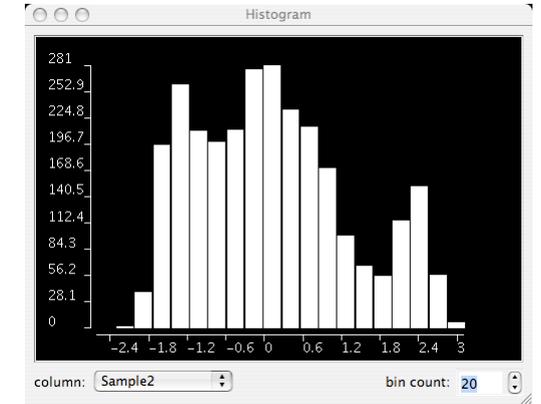
- Anzahl der Klassen verändert Histogrammform



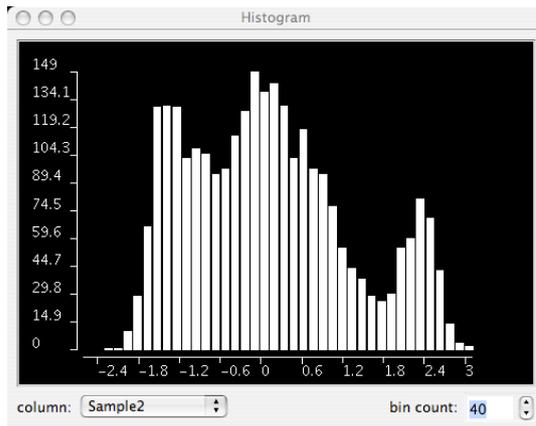
5 Klassen



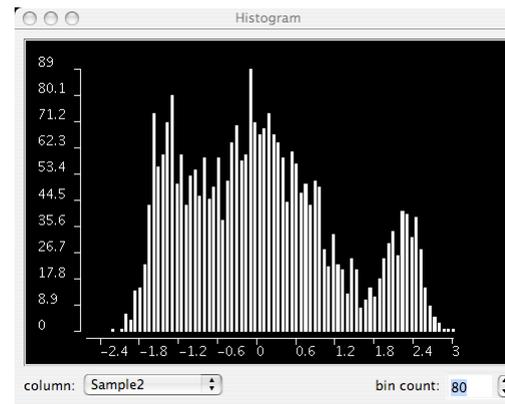
10 Klassen



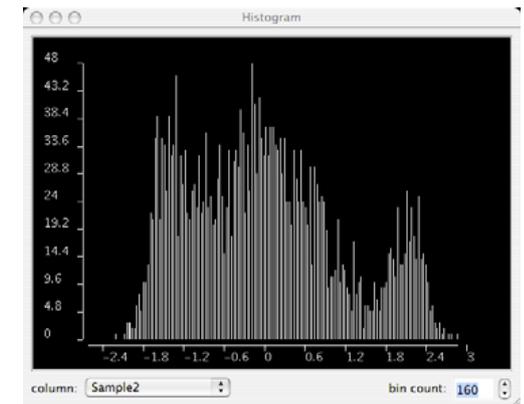
20 Klassen



40 Klassen



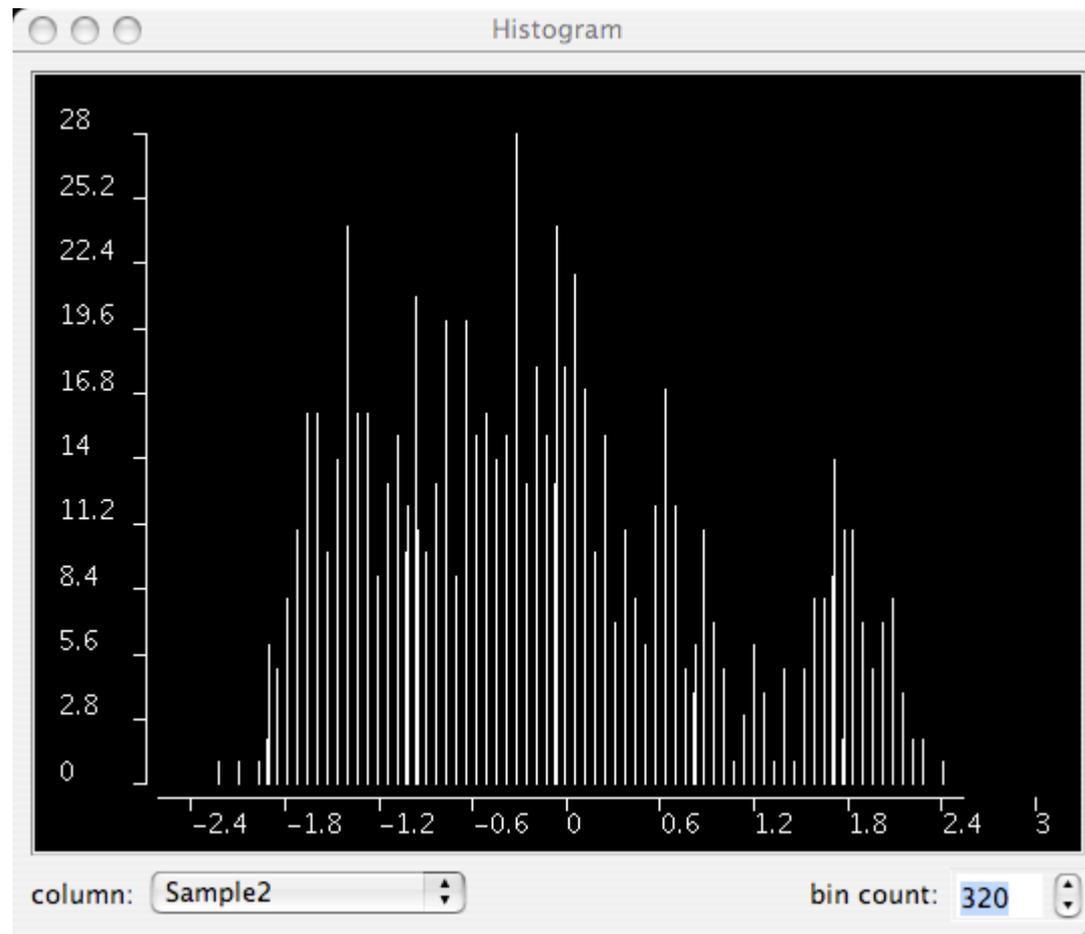
80 Klassen



160 Klassen

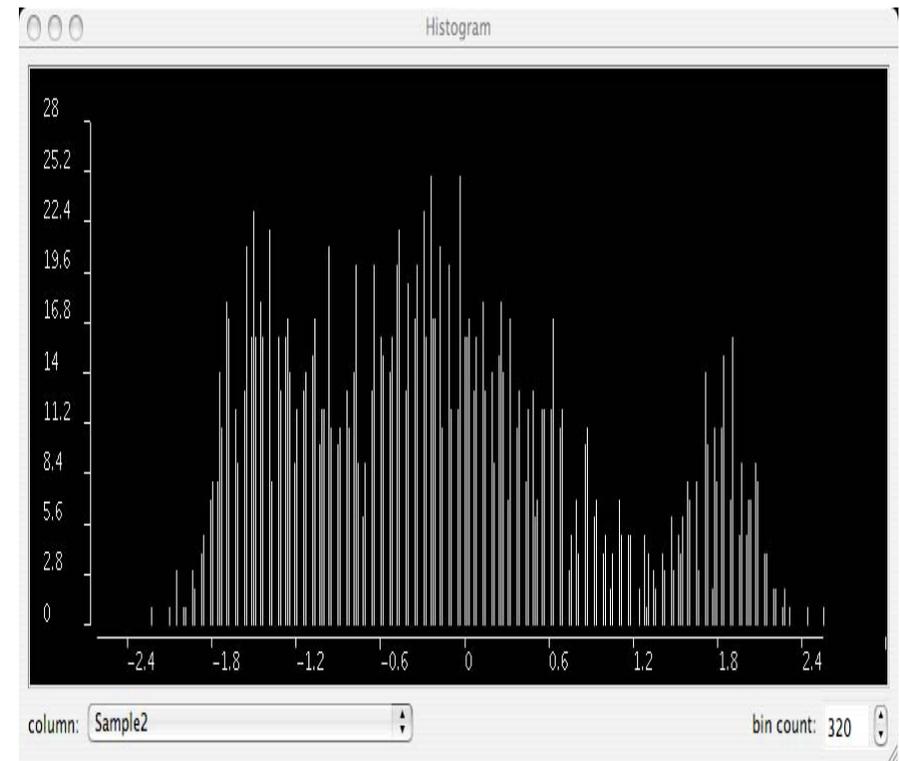
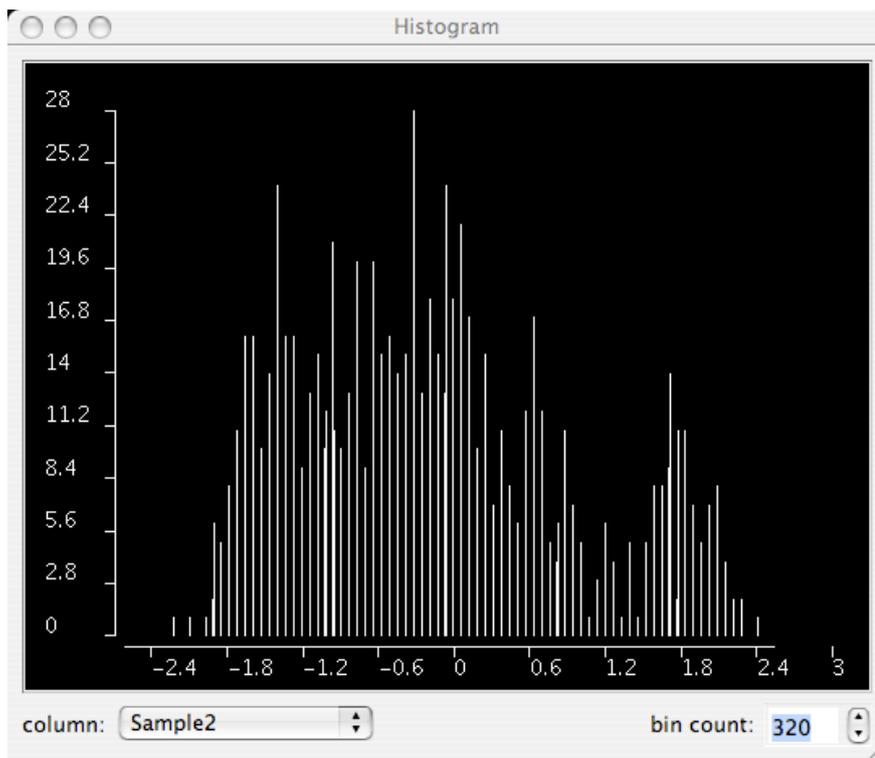
8.1 Histogramm

- Klassendarstellung ist für Aliasing anfällig

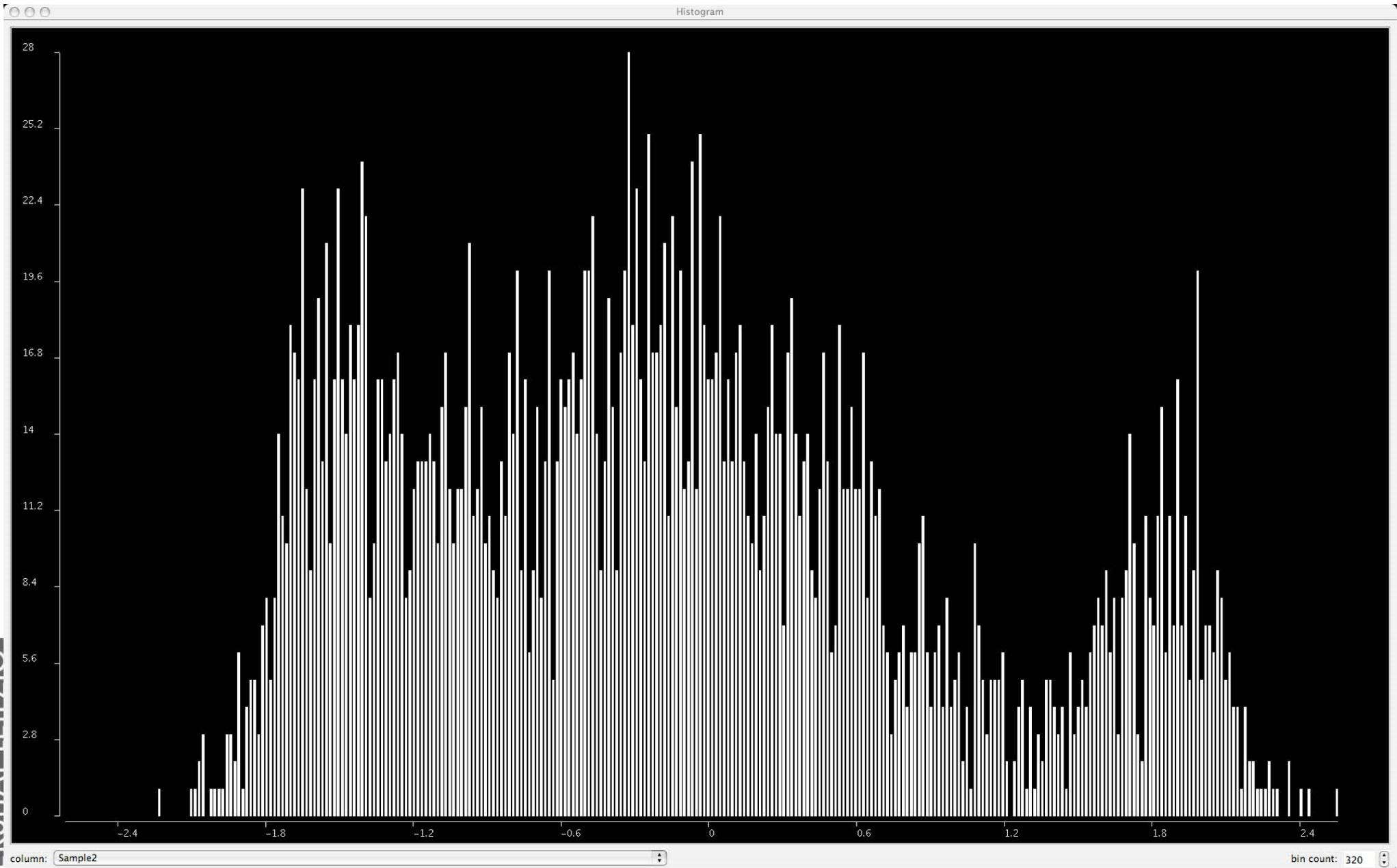
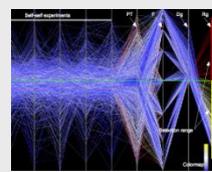


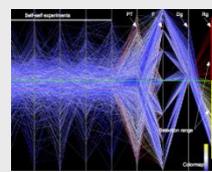
8.1 Histogramm

- Klassendarstellung ist für **Aliasing-Artefakt** anfällig
 - Unterschiedliche Darstellungsgrößen
 - **Einzelklassen fehlen**, schlecht kontrollierbar
 - Für **große Anzahlen** schlecht geeignet



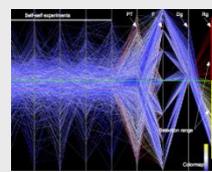
8.1 Histogramm





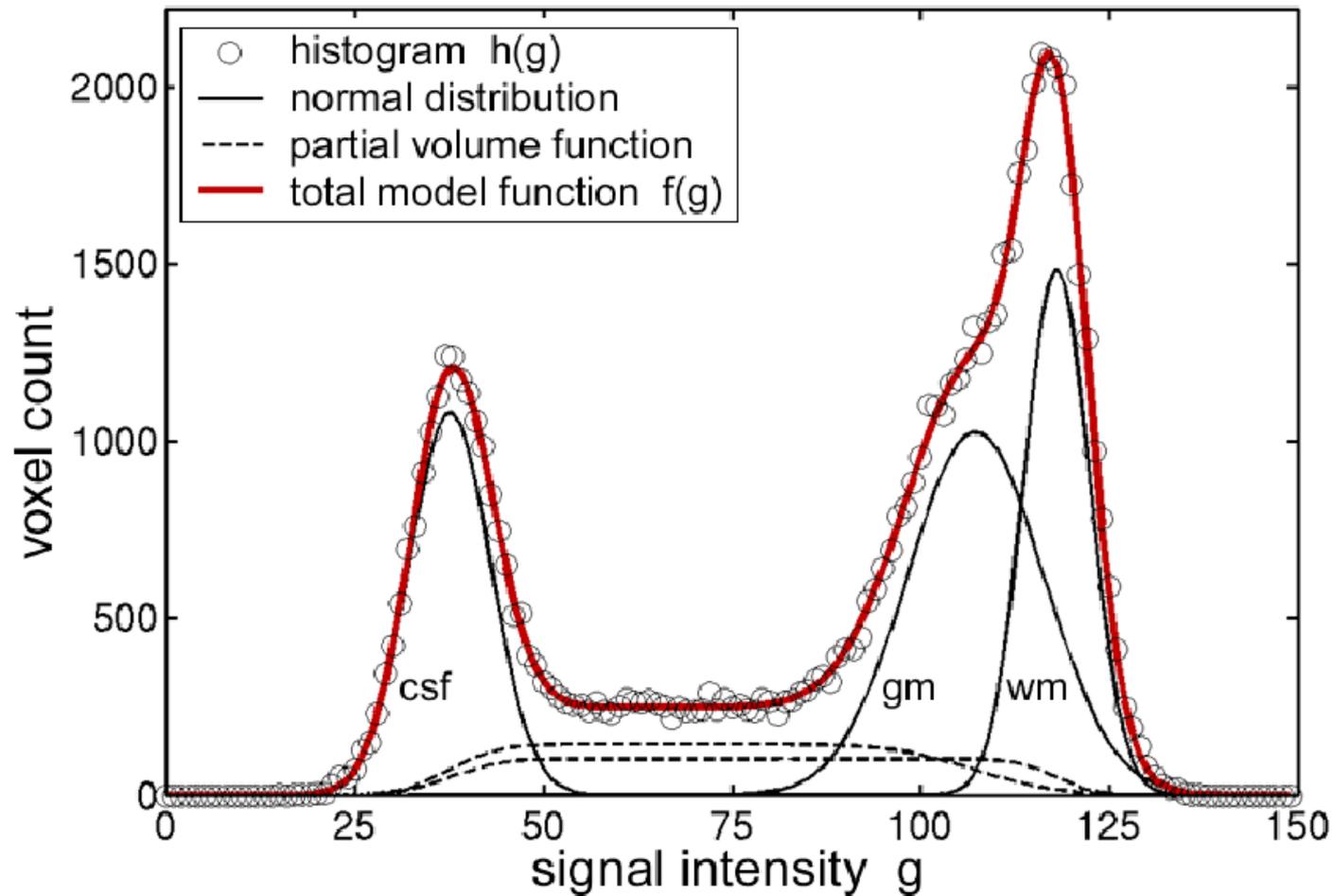
8.1 Histogramm

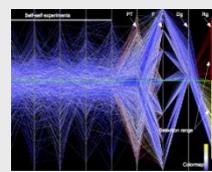
- **Anzahl der Klassen** muss sich also richten nach
 - Bildschirmauflösung
 - Anzahl der Datenwerte
- Histogramm stellt nur Abhängigkeit von einem Parameter dar
- Wenn der richtige Parameter dargestellt wird, so kann man oft eine **Gauss-Kurve** erkennen (Normalverteilung)
- Bei mehreren überlagernden Materialien (Bilddaten) spricht man auch vom **Mixed Gaussian Model**



8.1 Histogramm

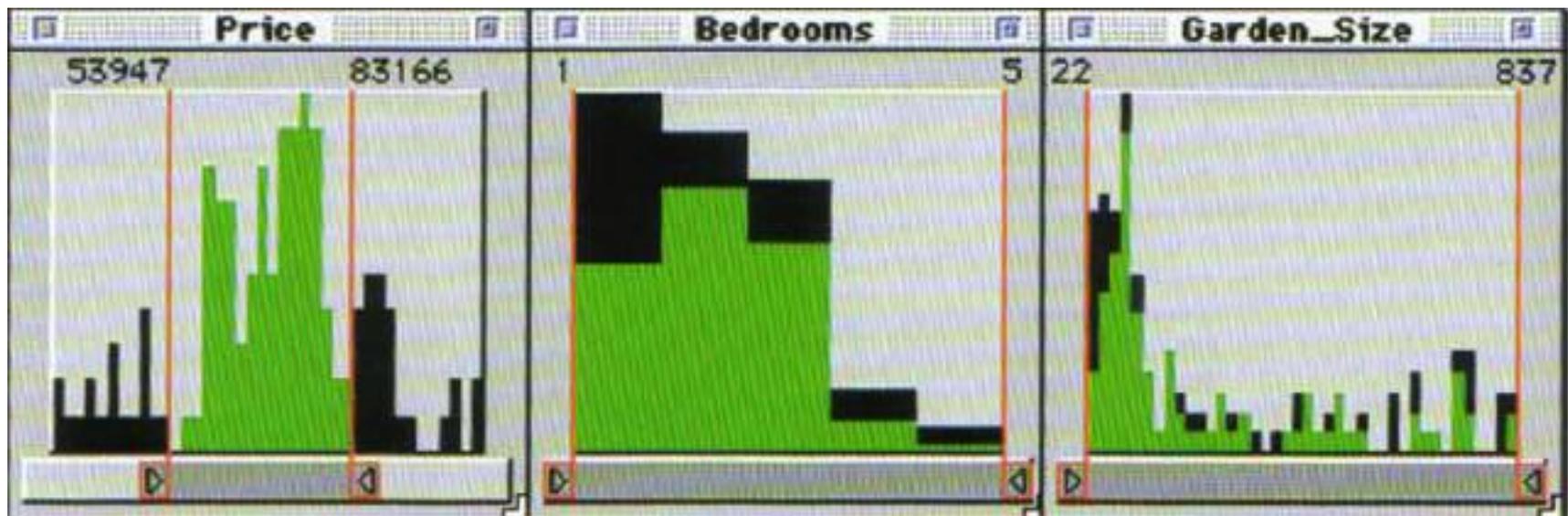
- Model-Fitting für drei Materialien (Normalverteilung)

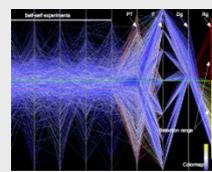




8.1 Histogramm

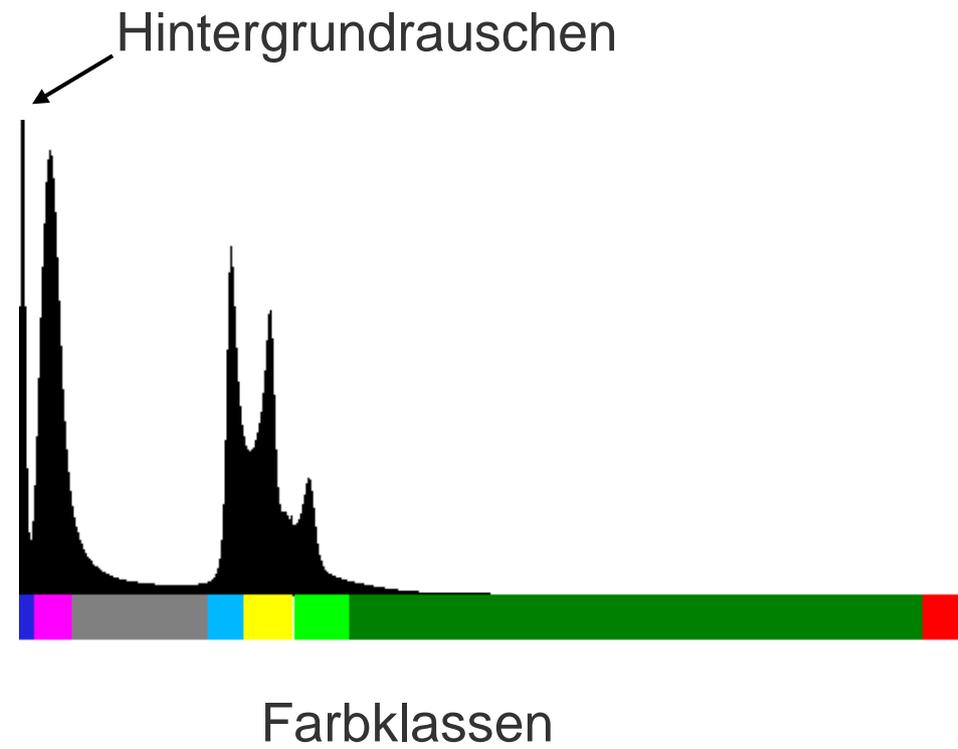
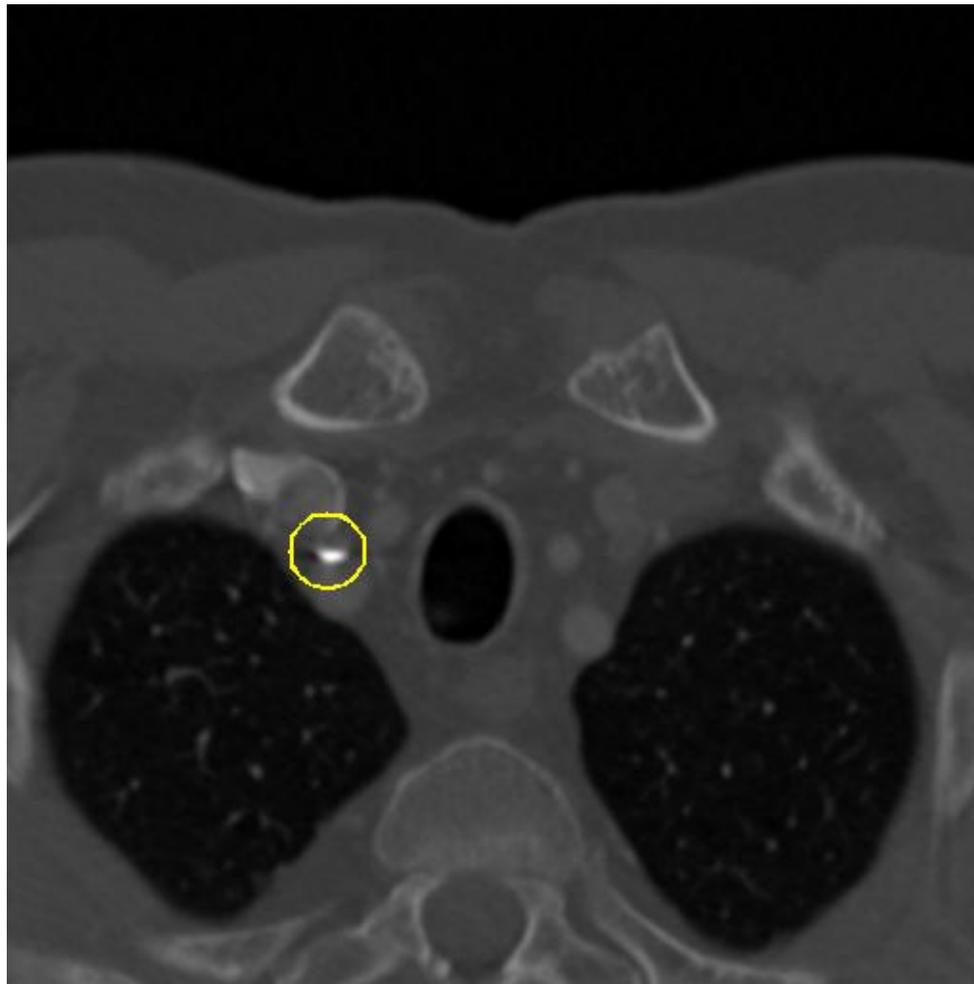
- **Brushing** in einer Variable induziert **Selektion in anderen Variablen** (von unten)

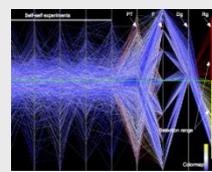




8.1 Histogramm

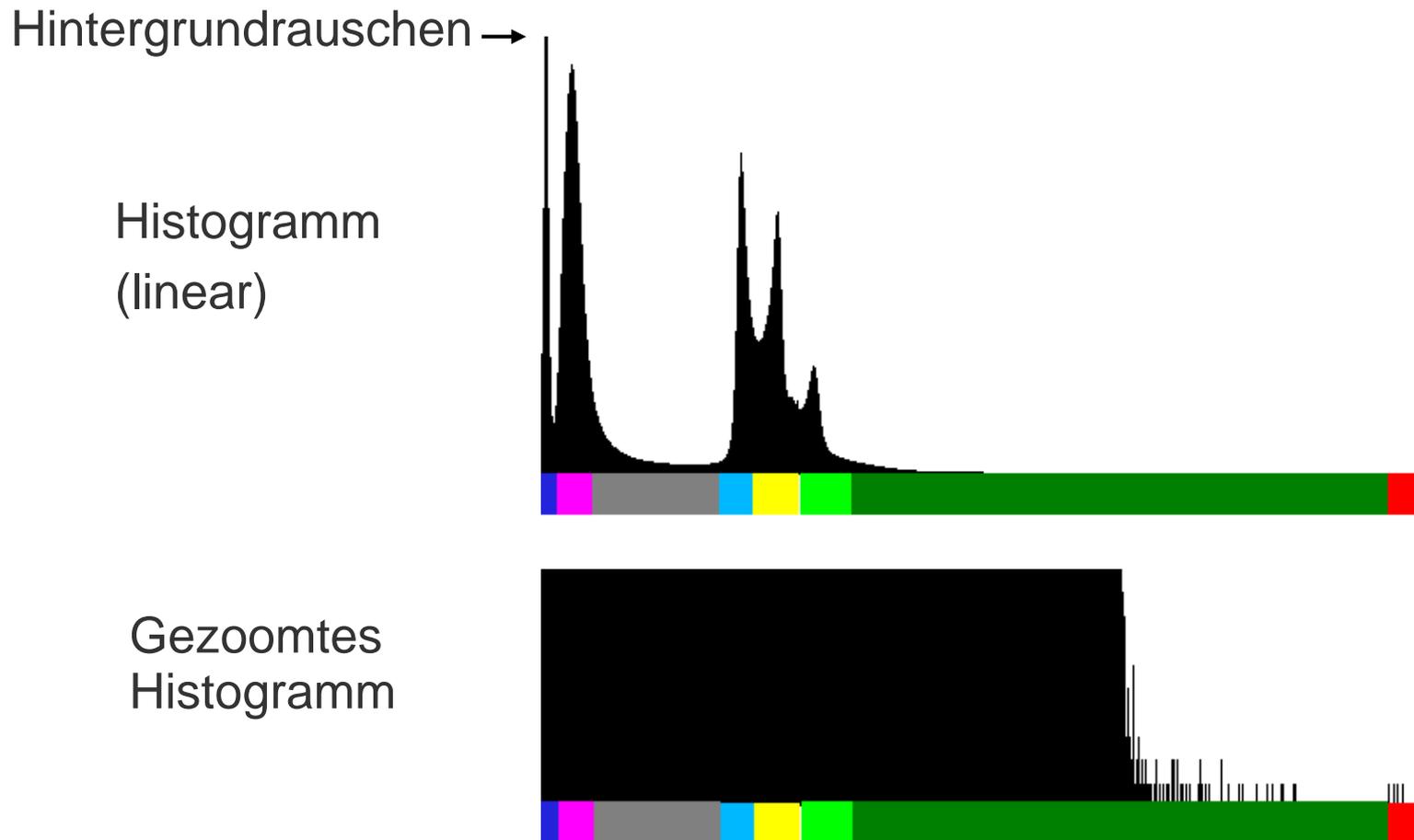
- CT Lunge

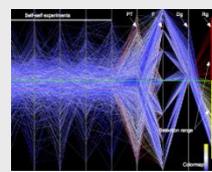




8.1 Histogramm

- Bildhistogramme werden **häufig von Hintergrundrauschen** dominiert werden

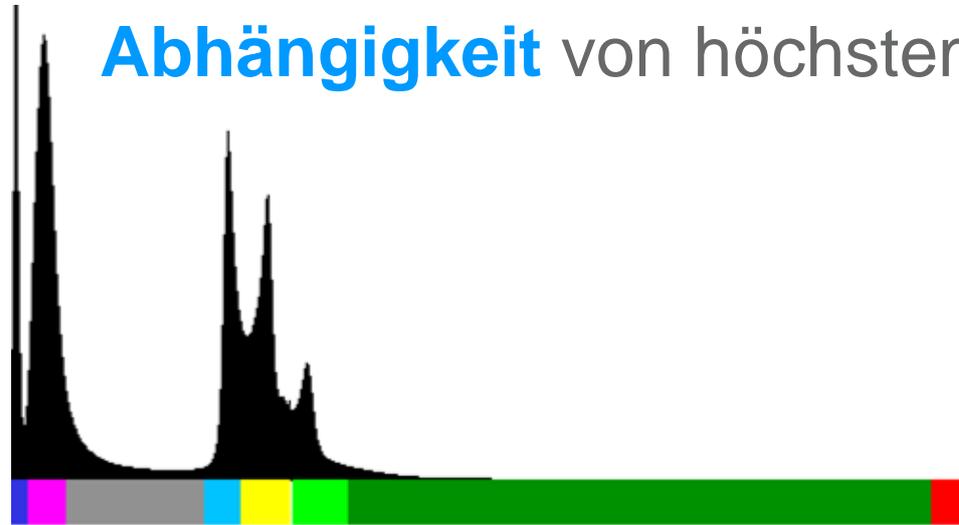




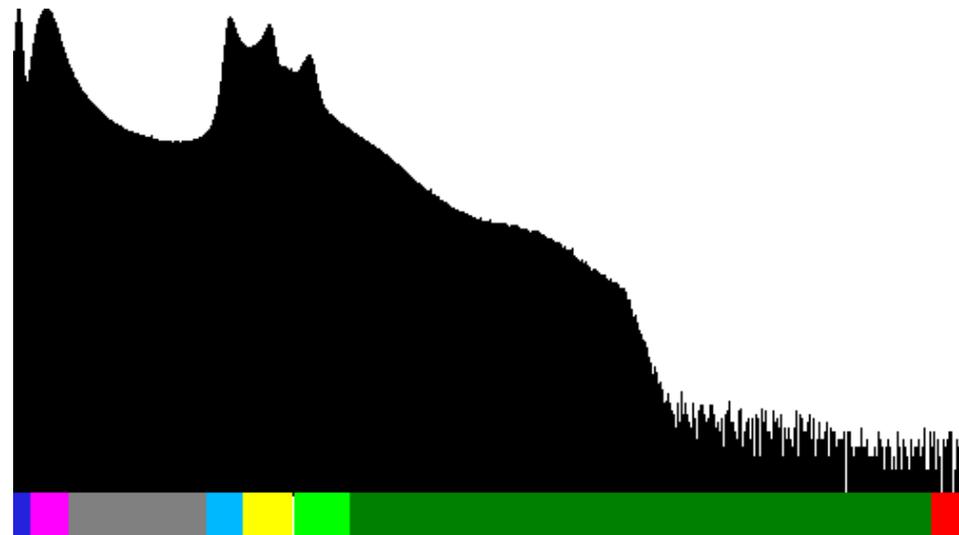
8.1 Histogramm

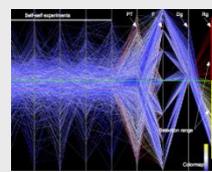
- Y-Achse wird **logarithmische skaliert** (unten)
- **Reduziert** (Peak) **Abhängigkeit** von höchstem Scheitel

Histogramm
(linear)



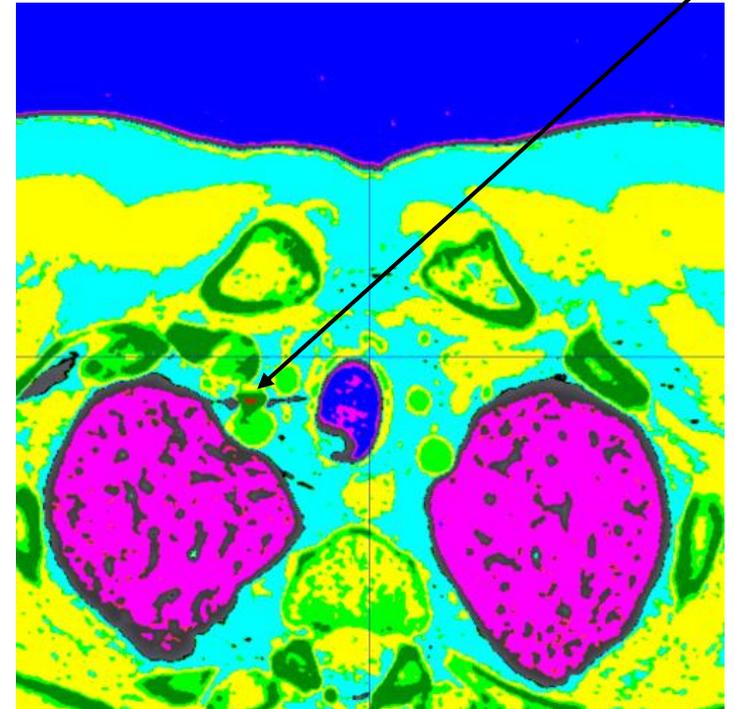
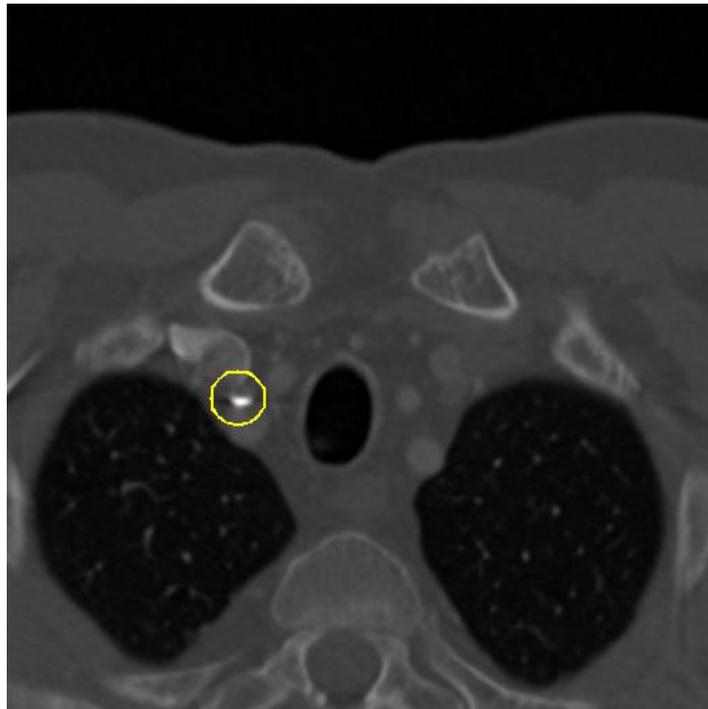
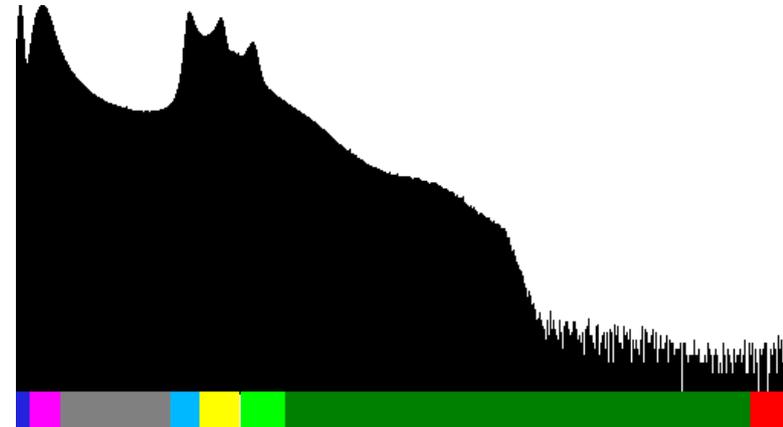
Logscale
Histogramm





8.1 Histogramm

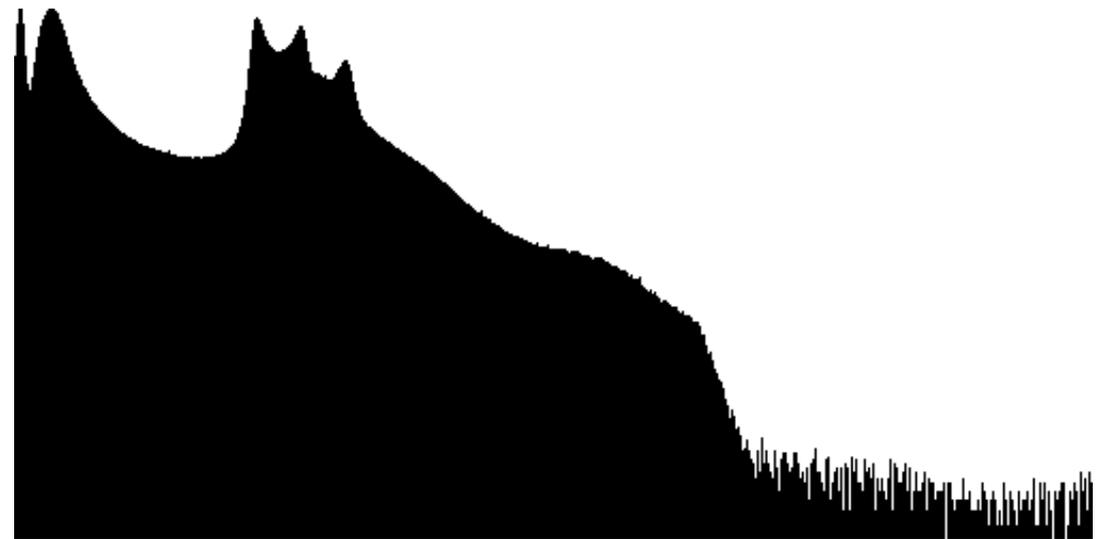
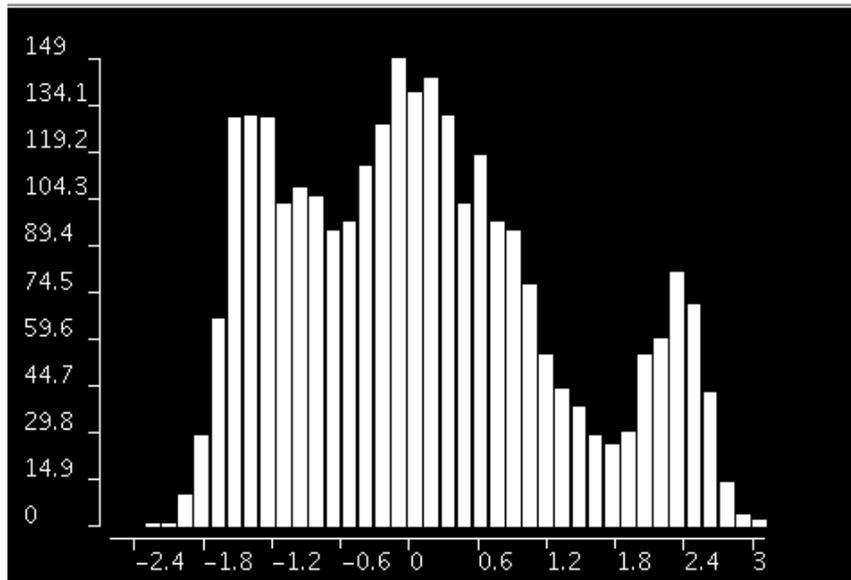
- CT Lunge **klassifiziert**



8.1 Histogramm

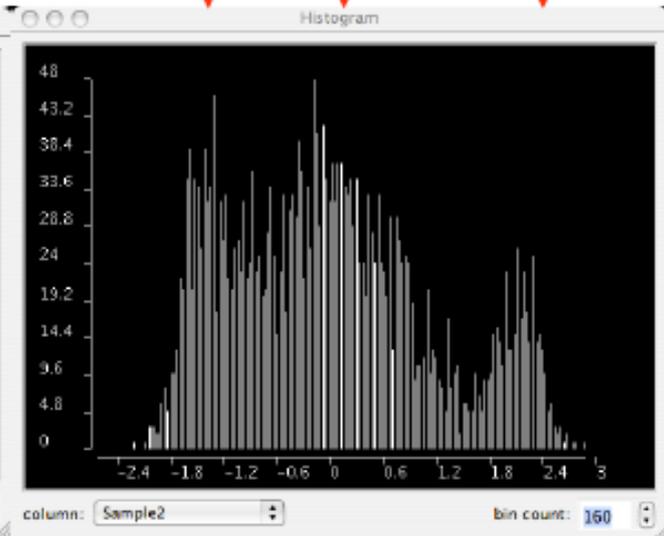
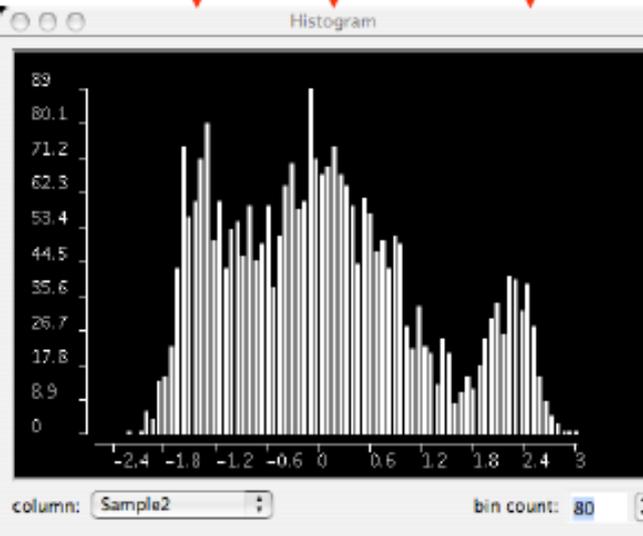
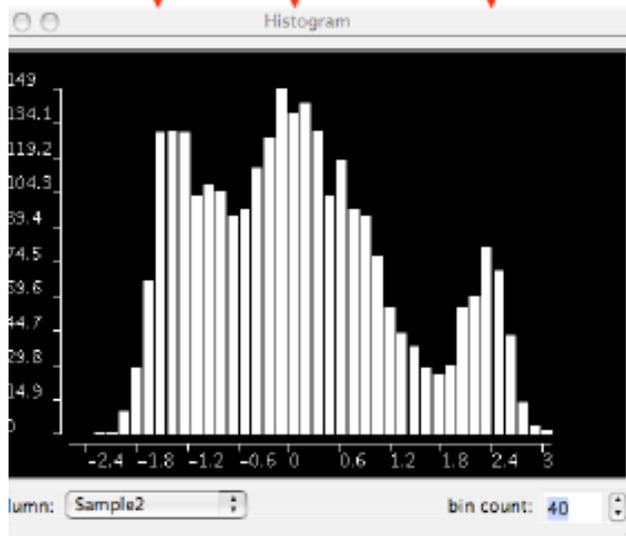
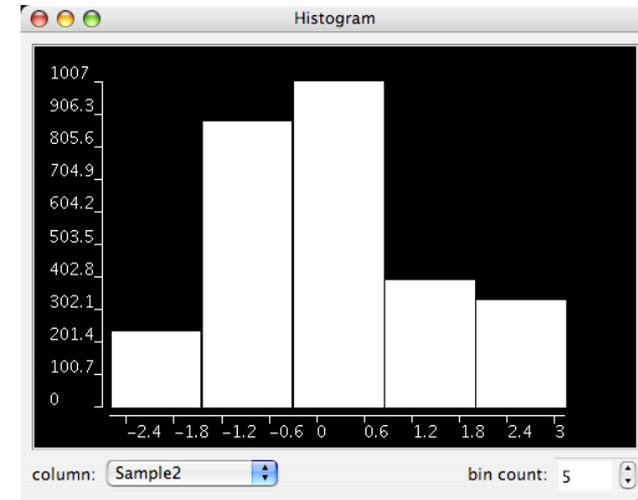
Balkendiagramm vs. Histogramm

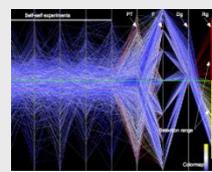
- Balkendiagramme haben **erkennbare** Breite
- ... haben **Zwischenräume**
- Histogramme haben **glatte Übergänge**



8.1 Histogramm

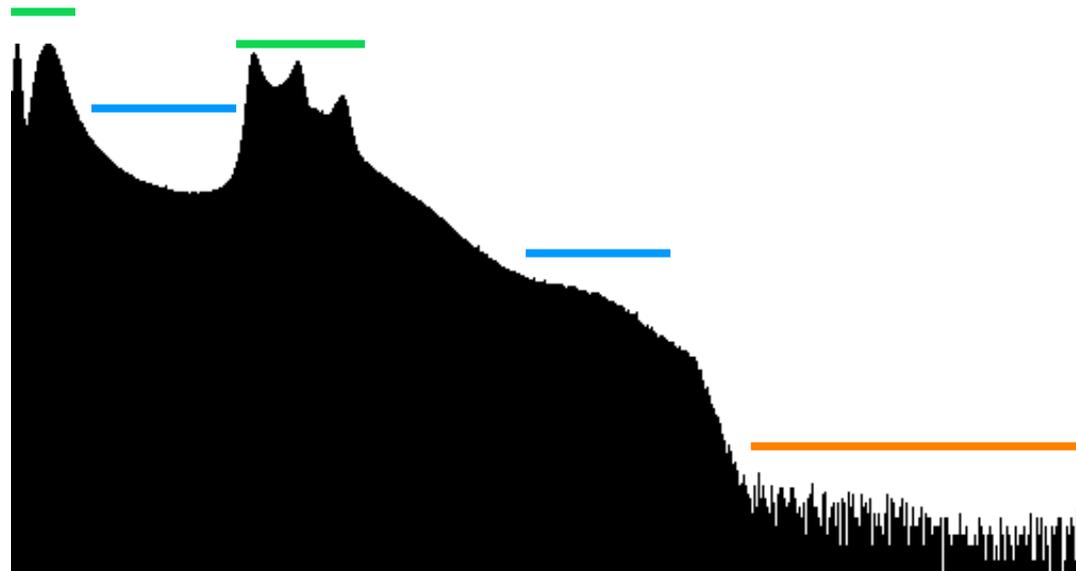
- Der Scheitel erkennbar (wenn genügend Klassen vorhanden)

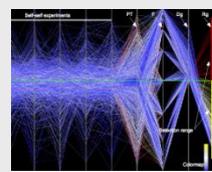




8.1 Histogramm

- **Ebenen** deuten auf ähnliches Verhalten
- **Flanken** auf Änderungen
- **Rauschen** kann Information überdecken
- Oft aber nur bei Bilddaten deutlich

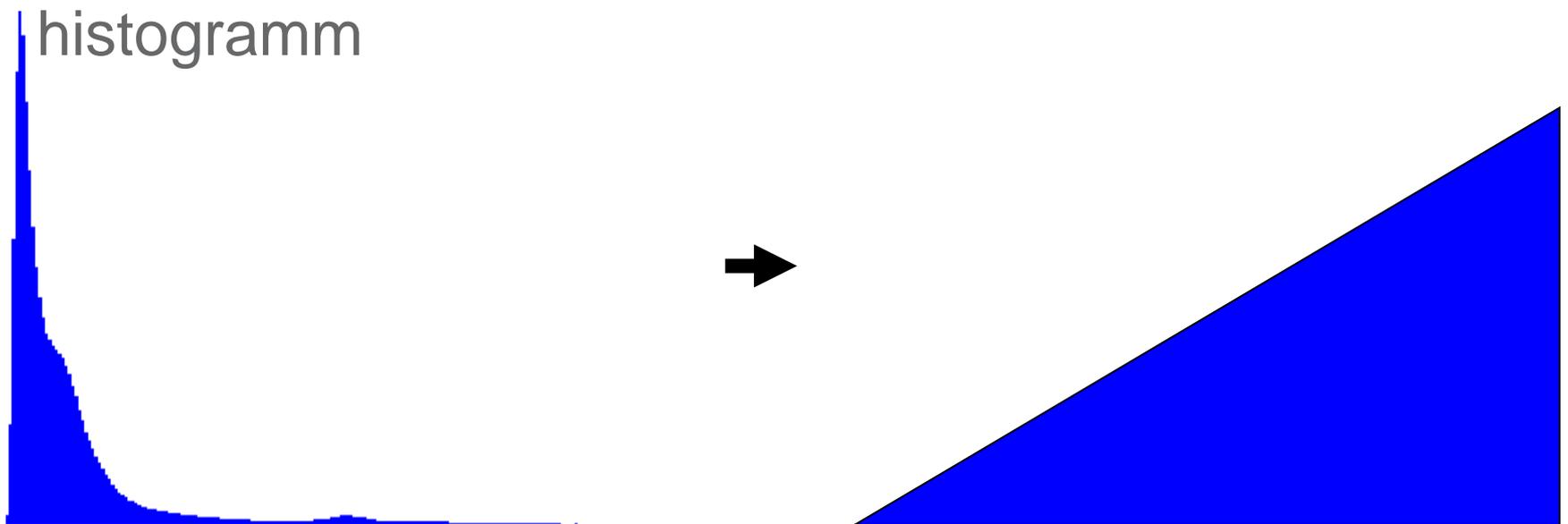


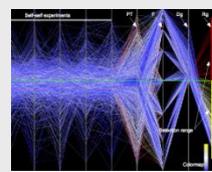


8.1 Histogramm

Summenhistogramm / Kumulatives Histogramm

- Auch Summenhäufigkeitsdiagramm
- Nächste Klasse ist **Summe** aus vorhergehenden Klassen und aktueller Klasse
- Verwendung in der **Bildverarbeitung**
- **Histogrammäqualisierung** basiert auf Summen-

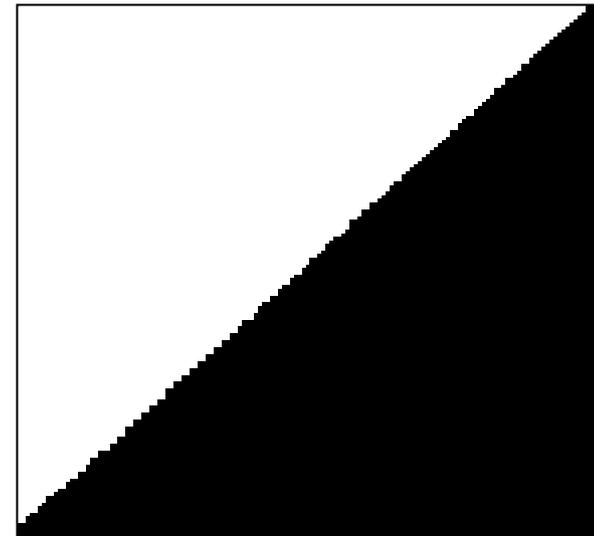
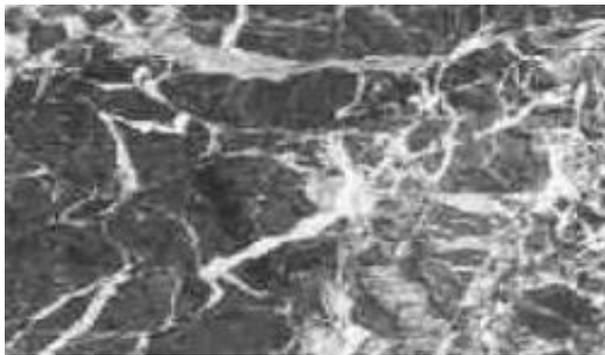




8.1 Histogramm

Summenhistogramm/Histogrammäquivalisierung

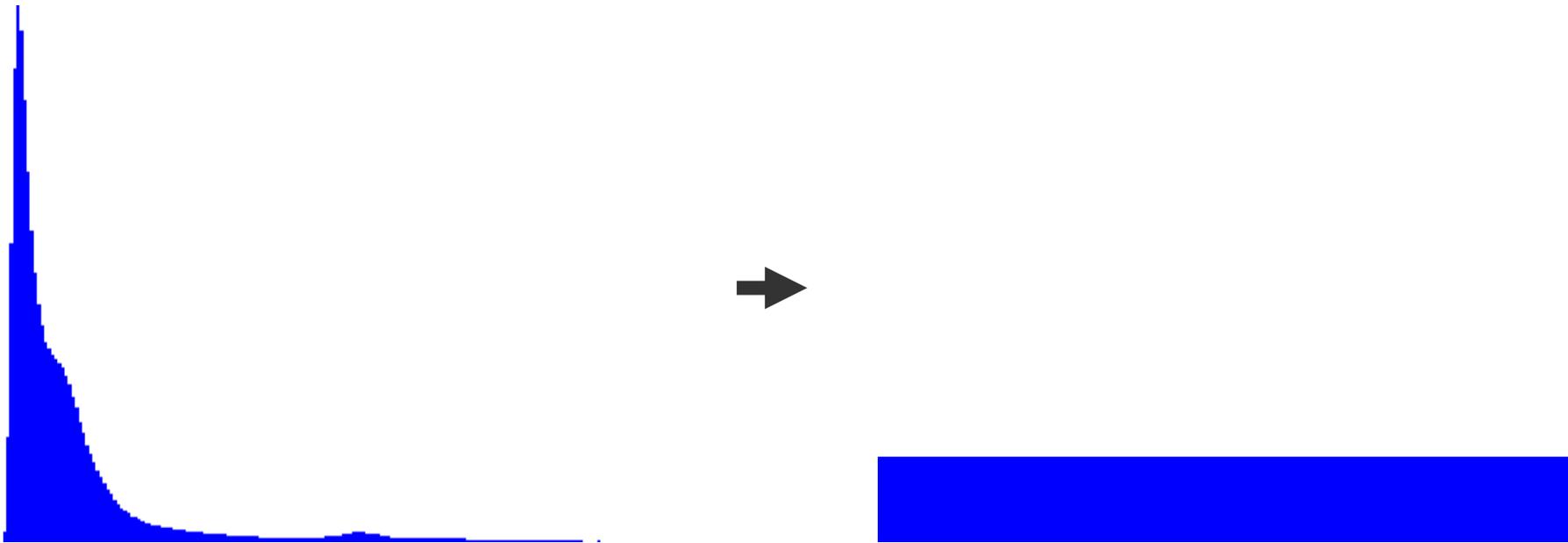
- Diskret, daher nicht perfekt

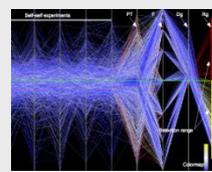


8.1 Histogramm

Histogramm nach Äqualisierung

- **Gleichverteilung** der Häufigkeit
- **Neuaufteilung** in Klassen
- **Stärkere Repräsentation** der großen Häufigkeiten
- Nicht gut für alle Datenarten

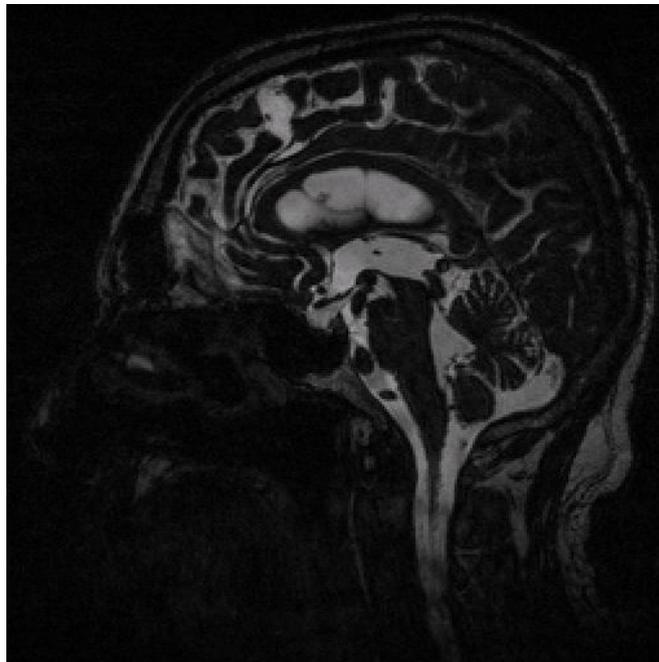


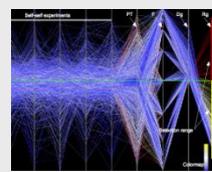


8.1 Histogramm

Histogramm nach Äqualisierung

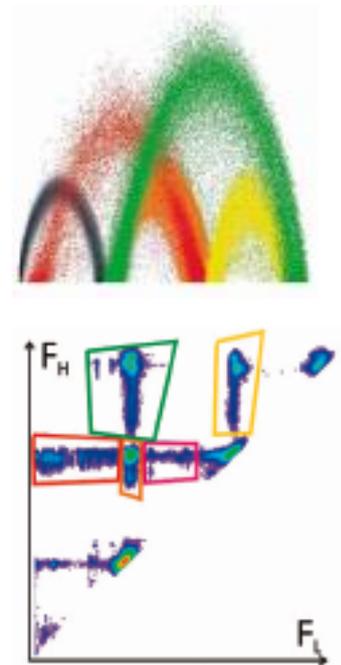
- Nicht gut für alle Datenarten:
Betonung des Hintergrundrauschens



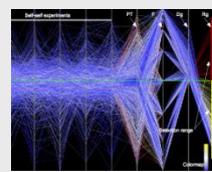


8.1 Histogramm

- Viele Ansätze versuchen **mehr Variablen** in Histogramm zu packen:
 - Gradientenhistogramm (2D-Histogramm) in der Volumenvisualisierung
 - LH-Histogramme bauen darauf auf (Low-High)
- Zwei Parameter:
 - **Scatterplot** (2D-Histogramm)
 - 3D-Histogramm
 - Stacked Histogramm

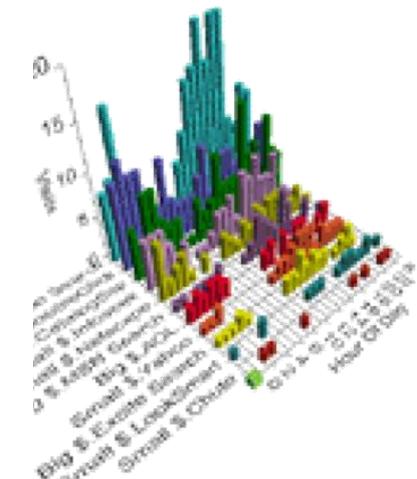
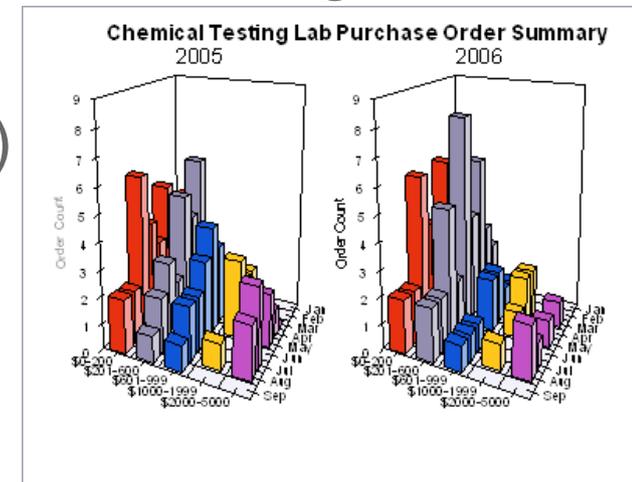


Gradientenhistogramm / LH-Histogramme

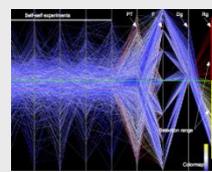


8.1 Histogramm

- Viele Ansätze versuchen **mehr Variablen** in Histogramm zu packen:
 - Gradientenhistogramm(2D-Histogramm) in der Volumenvisualisierung
 - LH-Histogramme bauen darauf auf (Low-High)
- Zwei Parameter:
 - Scatterplot (2D-Histogramm)
 - **3D-Histogramm**
 - Stacked Histogramm

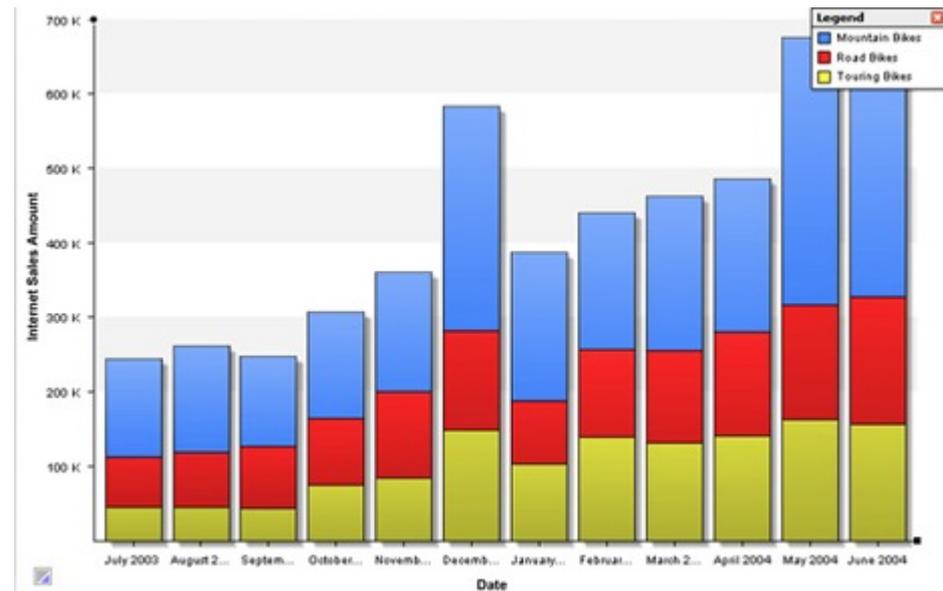


3D-Balkendiagramme



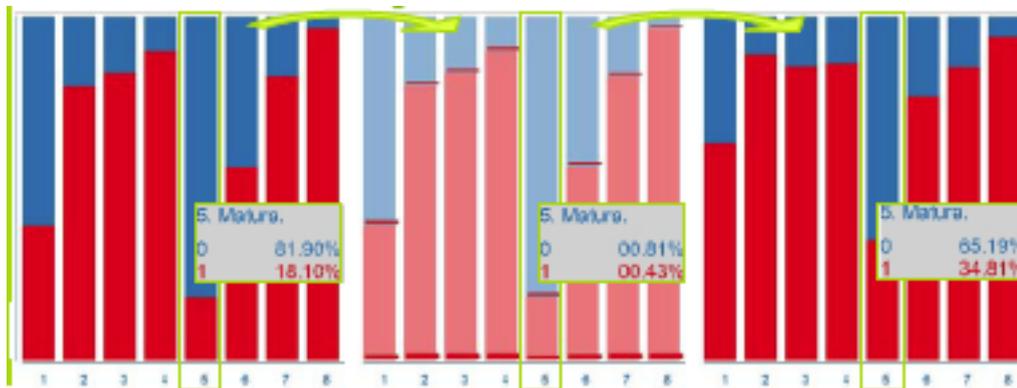
8.1 Histogramm

- Viele Ansätze versuchen **mehr Variablen** in Histogramm zu packen
- Zwei Variablen:
 - Scatterplot (2D-Histogramm)
 - 3D-Histogramm
 - **Stacked Histogram**

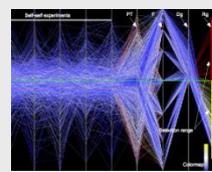


Absolutes Histogramm

[Wikipedia.de]



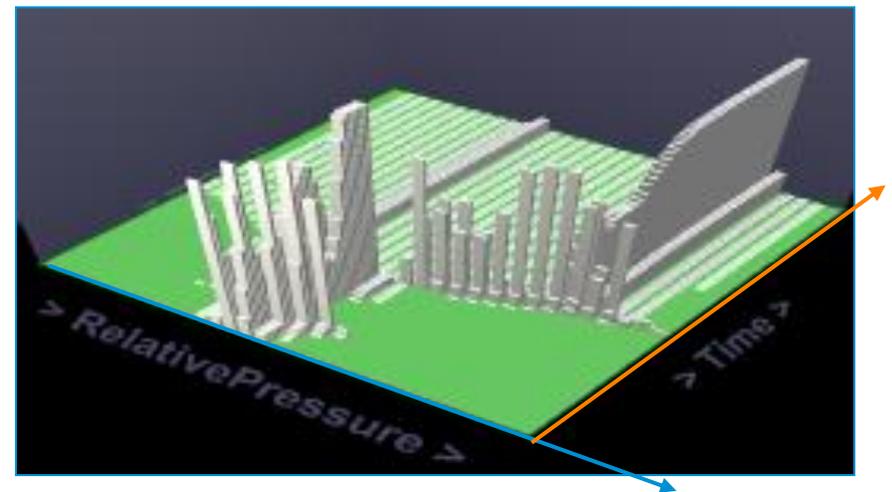
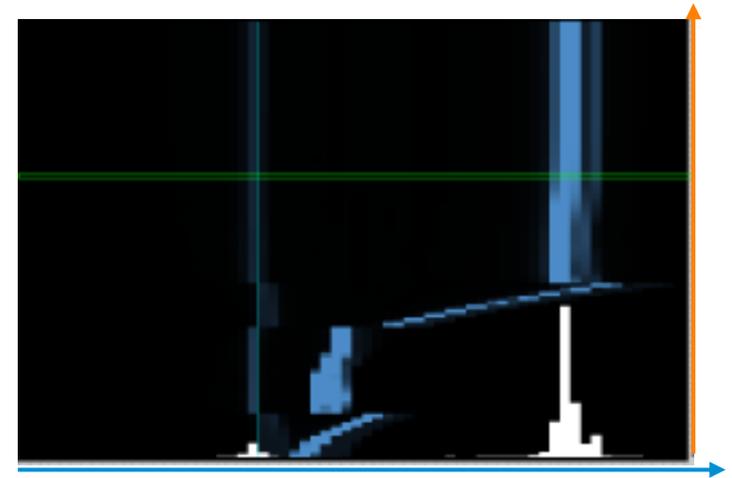
Relatives Histogramm [Hauser, 2006]

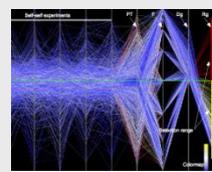


8.1 Histogramm

TimeHistogram

- Verbessert Histogramme für **Zeitreihen** (CFD)
- Eigentlich ein Scatterplot/2D Histogramm
- Beschränkt in **Klassenanzahl**
- Kontinuierliches Histogramm **nicht so gut lesbar**

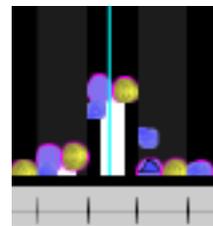
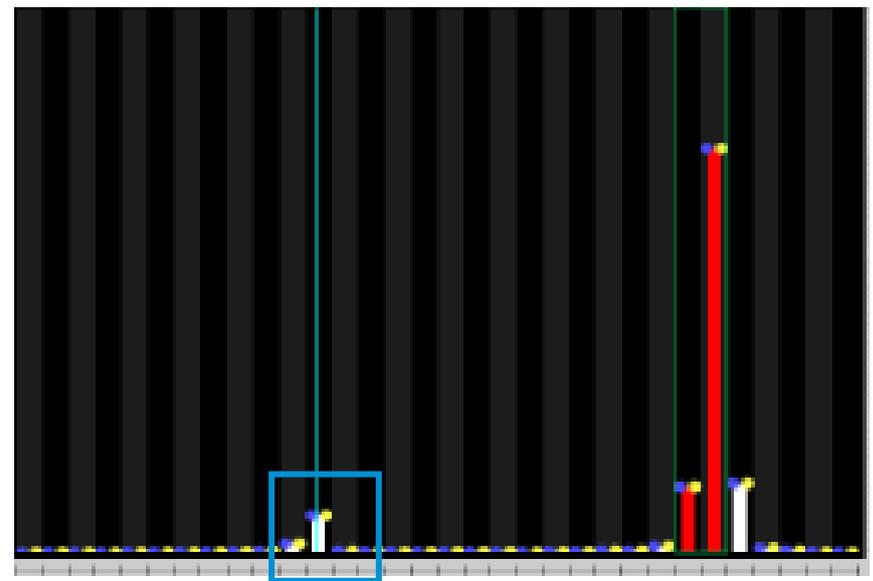
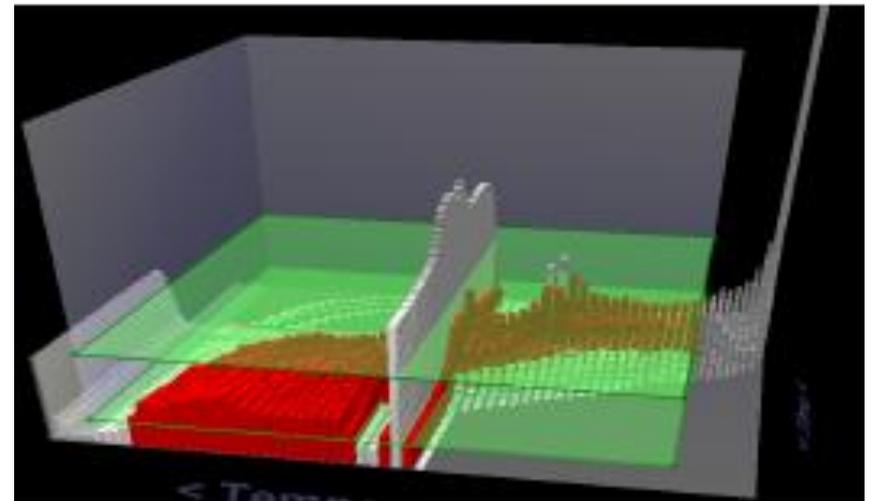


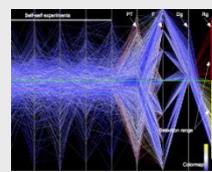


8.1 Histogramm

TimeHistogram

- Interaktion / Brushing **in 3D nicht so einfach**
- Deshalb meistens **in 2D**
- **Unmittelbare** zeitliche Nachbarschaft als Punkte
 - Links/Lila: Vergangenheit
 - Rechts/gelb: Zukunft

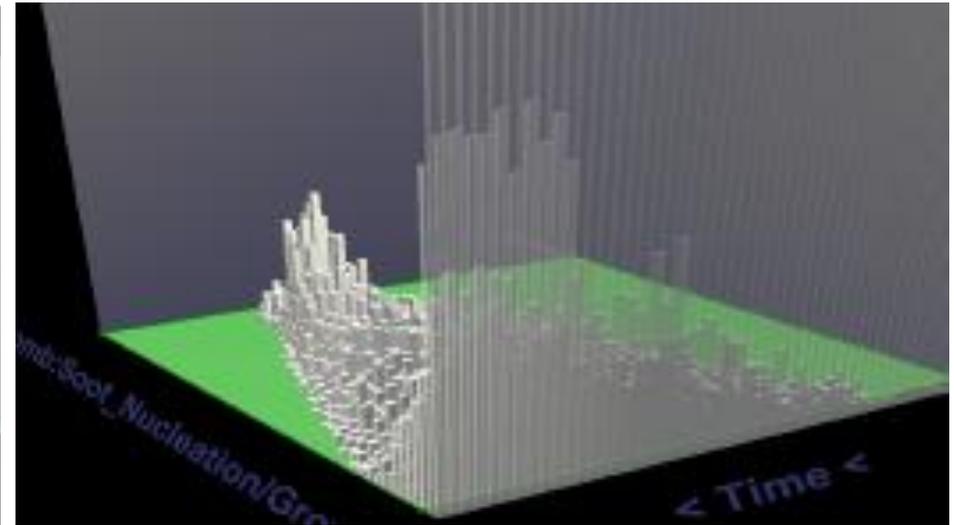
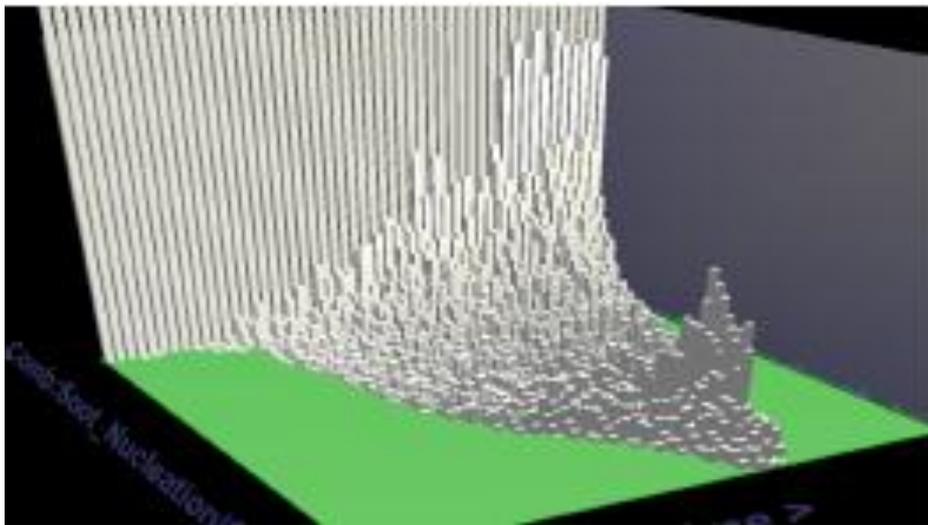


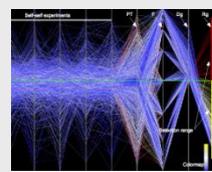


8.1 Histogramm

TimeHistogram

- **Verdeckungsreduktion** durch
 - Projektion auf den Rand
 - Transparenz des Rands

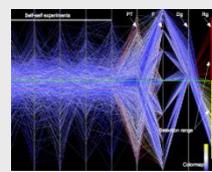




8.1 Histogramm

TimeHistogram

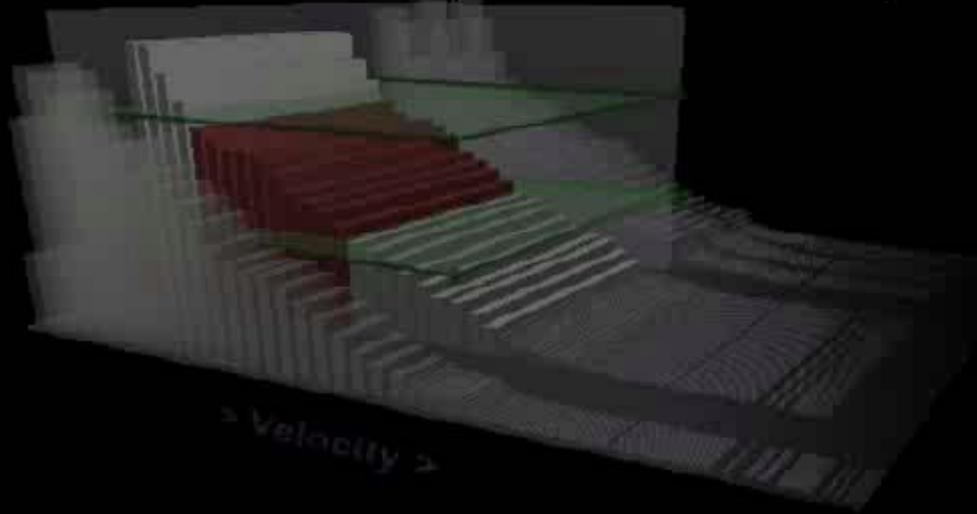
- Enge Verzahnung mit Zeitachse
- Behauptet deutliche Verbesserungen
- 3D Interaktion schwierig(er)
- Verdeckung nicht wirklich gelöst

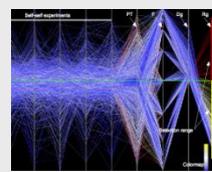


8.1 Histogramm

TimeHistograms for Large, Time-Dependent Data

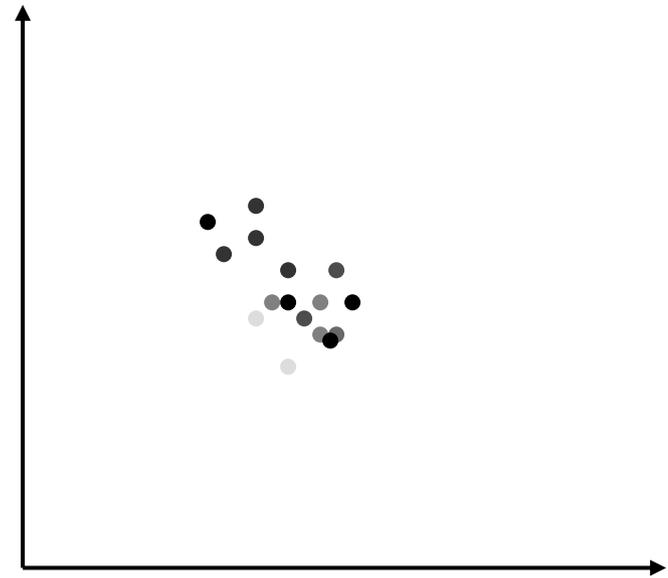
Robert Kosara, Fabian Bendix, Helwig Hauser

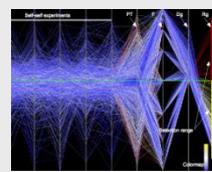




8.2 Scatterplot

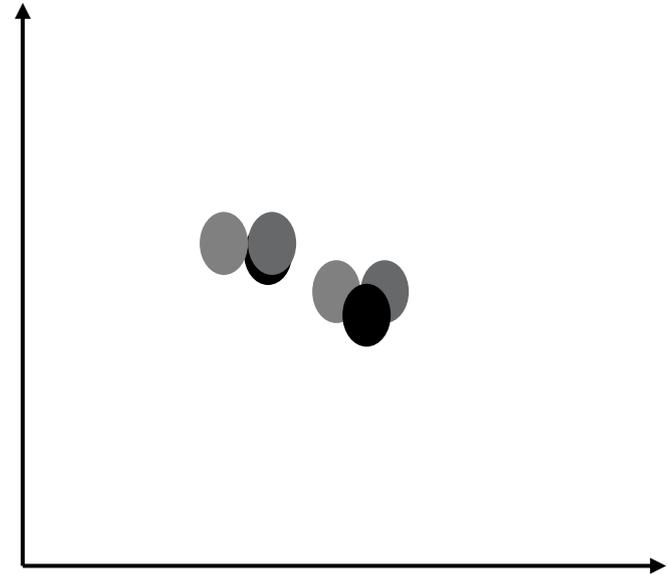
- Auch **Streudiagramm** oder **PointGraph**
- Bildet **zwei Variablen / Merkmale** auf Abszisse und Ordinate
- Ist eine **Projektion**
- Häufigkeit wird als **Dichte, Opazität, oder Luminanz** abgebildet
- Pseudokontinuierlich
- Potentielle **Aliasing**-Probleme





8.2 Scatterplot

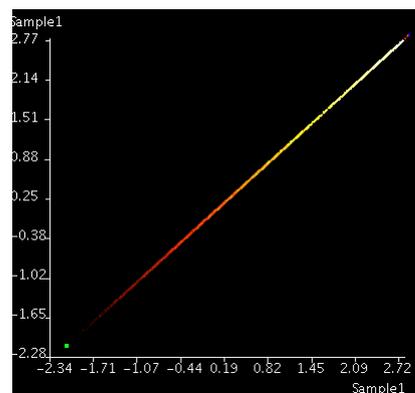
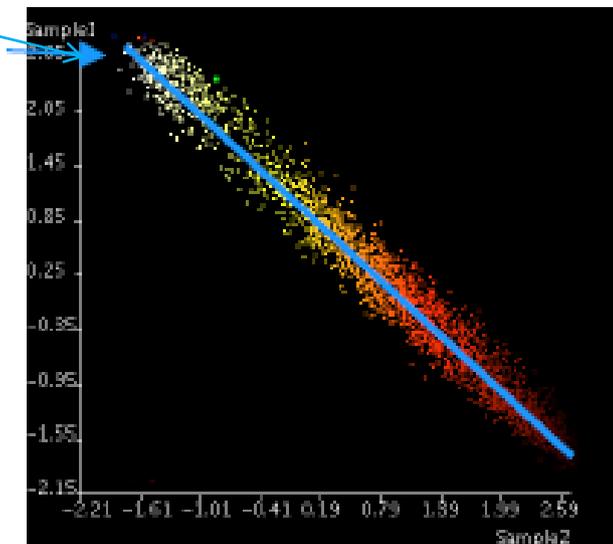
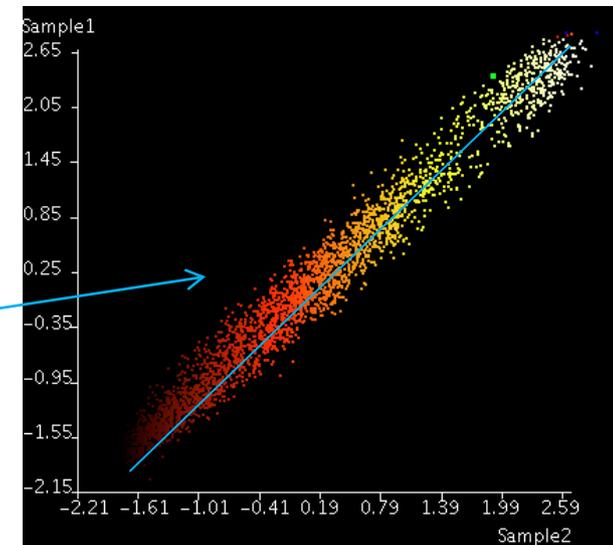
- **Keine feste Reihenfolge**
(Compositing)
- Letzter Punkt in der Liste wird **als letztes** dargestellt
- Wird **über bisherige** Punkte gezeichnet:
Verdeckung (Occlusion) oder **Overdraw**

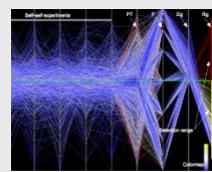


8.2 Scatterplot

Analyse von Abhängigkeiten

- Lineare Korrelation:
Sample 1/Sample2
 - Positive Korrelation
 - Negative Korrelation
 - Trendlinie
- Selbstvergleich liefert perfekte Korrelation

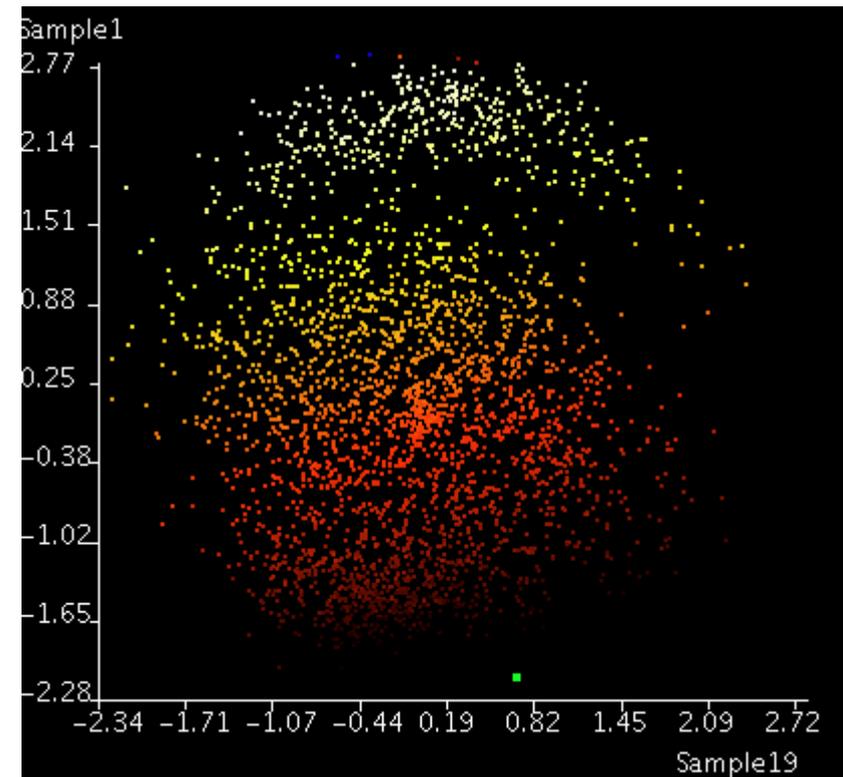




8.2 Scatterplot

Analyse von Abhängigkeiten

- Lineare Korrelation: Sample1/Sample2
- Nicht alle Variablen sind direkt korreliert



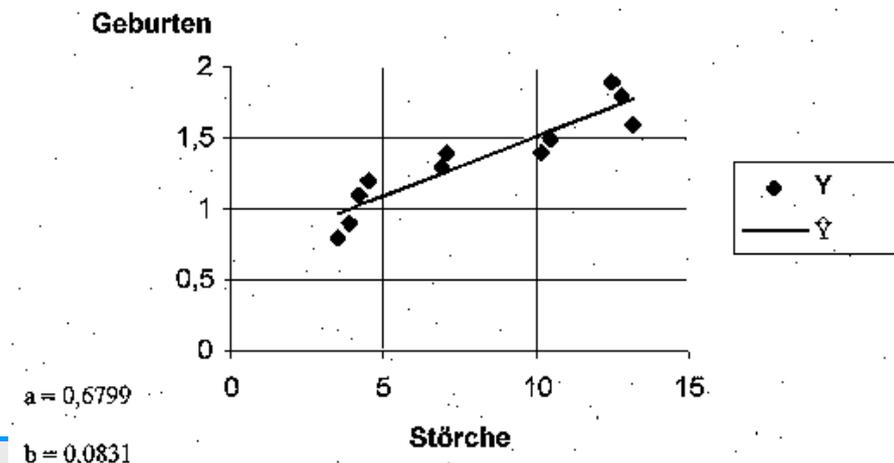
8.2 Scatterplot

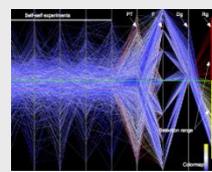
Korrelationen

- Bedeuten **nicht eine echte Abhängigkeit**
- Deuten auf **gleiche** (positive Korrelation) **Datenentwicklung**
- Beispiel: Störche und Geburtenrate

Störche und Geburten in ausgewählten Ländern, Ausgangs- und Prognosewerte und graphische Darstellung

Land	Störche pro 100 Quadratkilometer X	Geburten pro 1000 Einwohner Y	\hat{Y}
Deutschland	3,5	0,8	0,971
Frankreich	3,9	0,9	1,004
USA	4,2	1,1	1,029
Brasilien	7,1	1,4	1,270
Argentinien	6,9	1,3	1,253
Rußland	4,5	1,2	1,054
Ägypten	12,8	1,8	1,744
Namibia	10,5	1,5	1,552
China	13,2	1,6	1,777
Indonesien	10,2	1,4	1,528
Indien	12,5	1,9	1,719

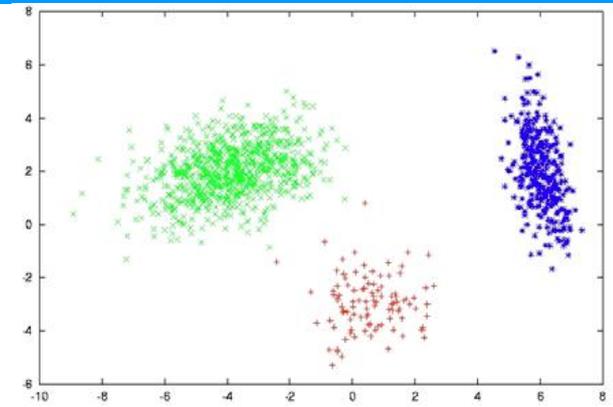




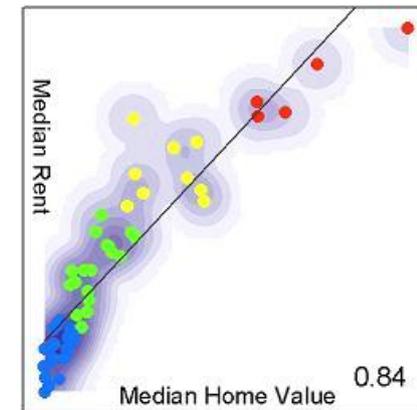
8.2 Scatterplot

Analyse von Abhängigkeitsstrukturen

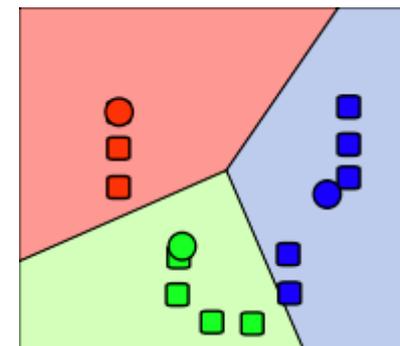
- **Cluster: Anhäufung** von Punkten
- **Clusteranalyse:**
 - Multivariate Analyseverfahren
 - Strukturentdeckend
- Erfordert Maß für **Zusammengehörigkeit**
 - Distanzmetrik (Single/complete/average Linkage)
 - Schwerpunkt/Mittelpunkt
 - Ähnlichkeitsfunktion



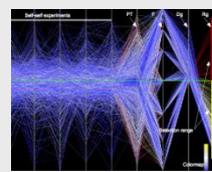
[GFal.de]



[ESRI.com]



[Wikipedia.de]



8.2 Scatterplot

Visuelle Clustererkennung

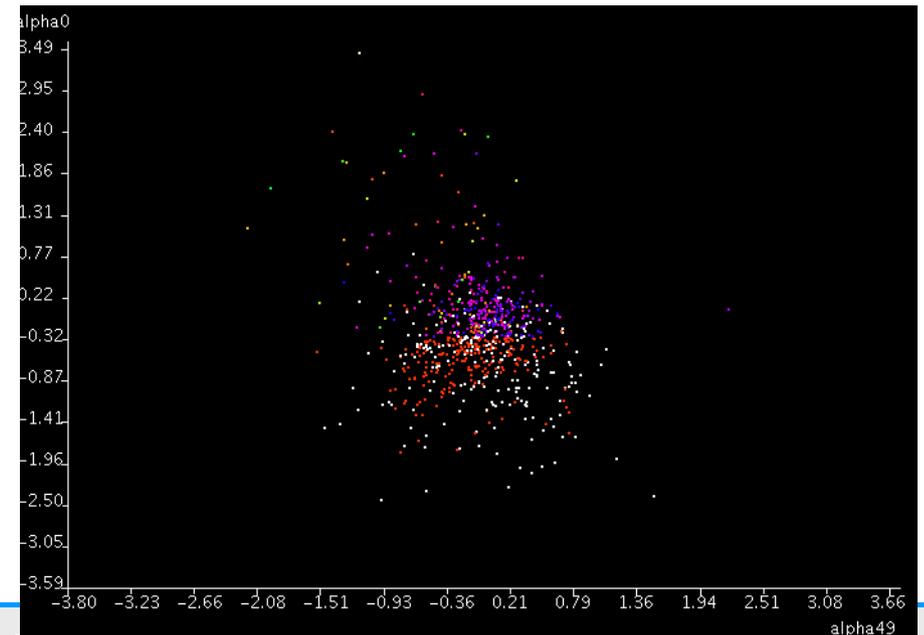
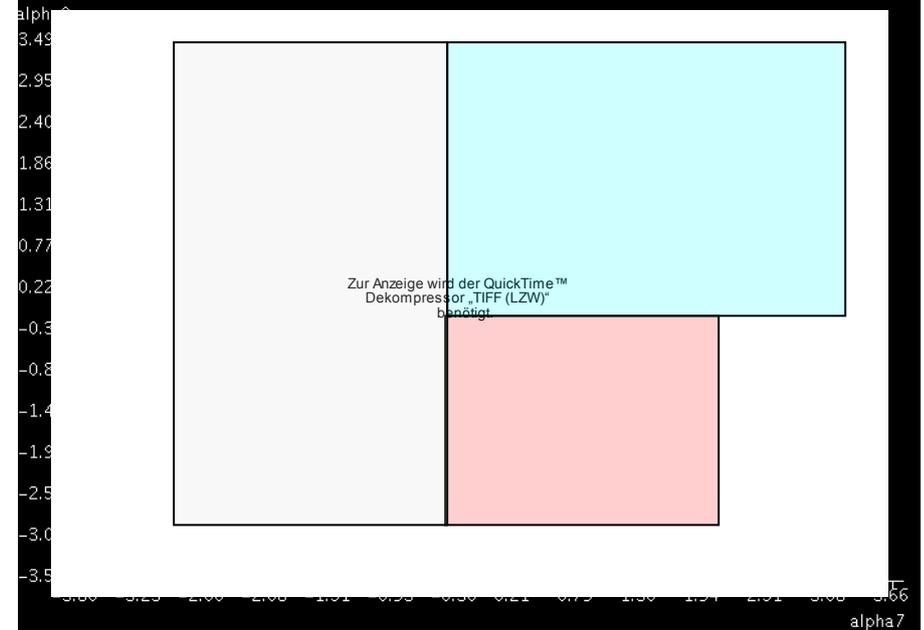
- **Opazität**modellierung: Je mehr Punkte, **desto opaker**
- Verschiedene **Gewichtung**modelle möglich
 - Global
 - Klassenspezifisch

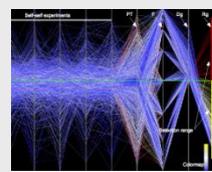


8.2 Scatterplot

Brushing und Linking

- Brushing in einer Projektion ergibt **andere Selektion** in anderer Projektion

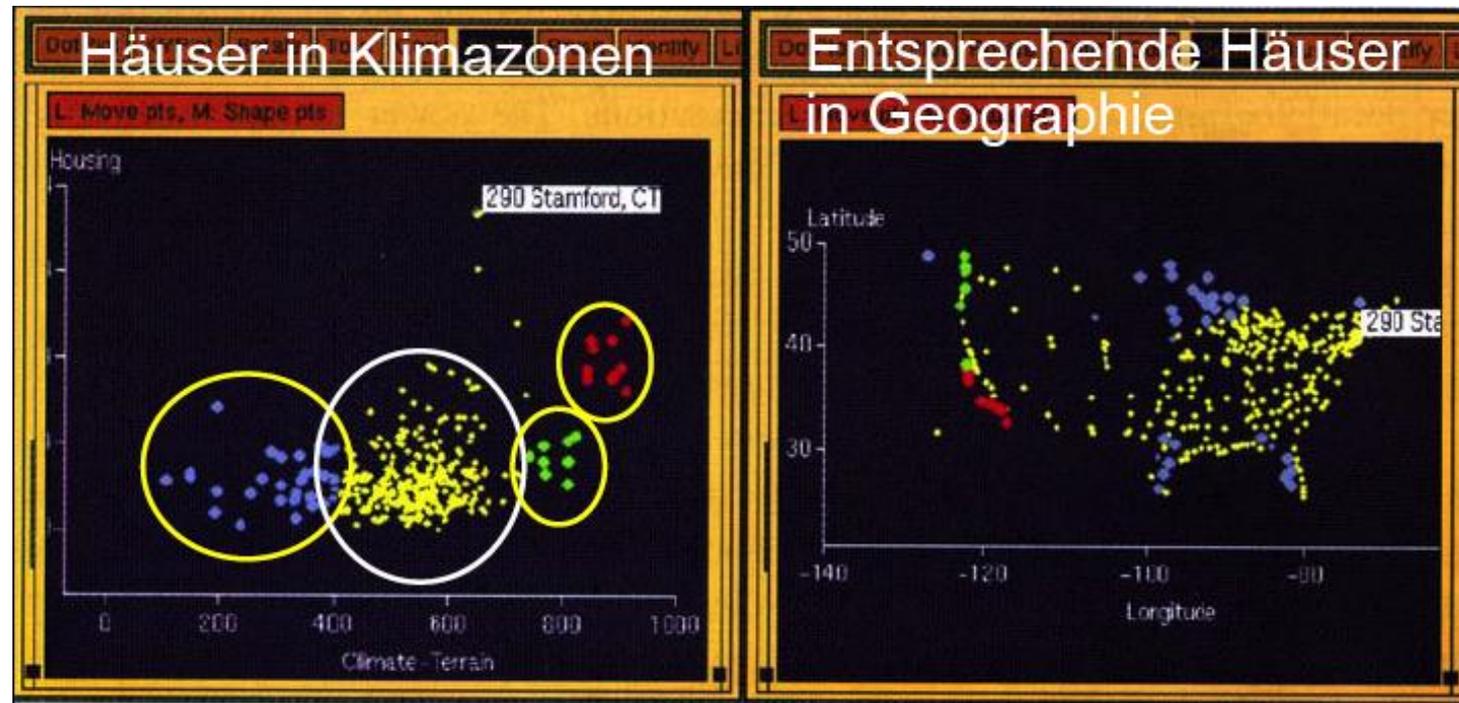




8.2 Scatterplot

Brushing und Linking

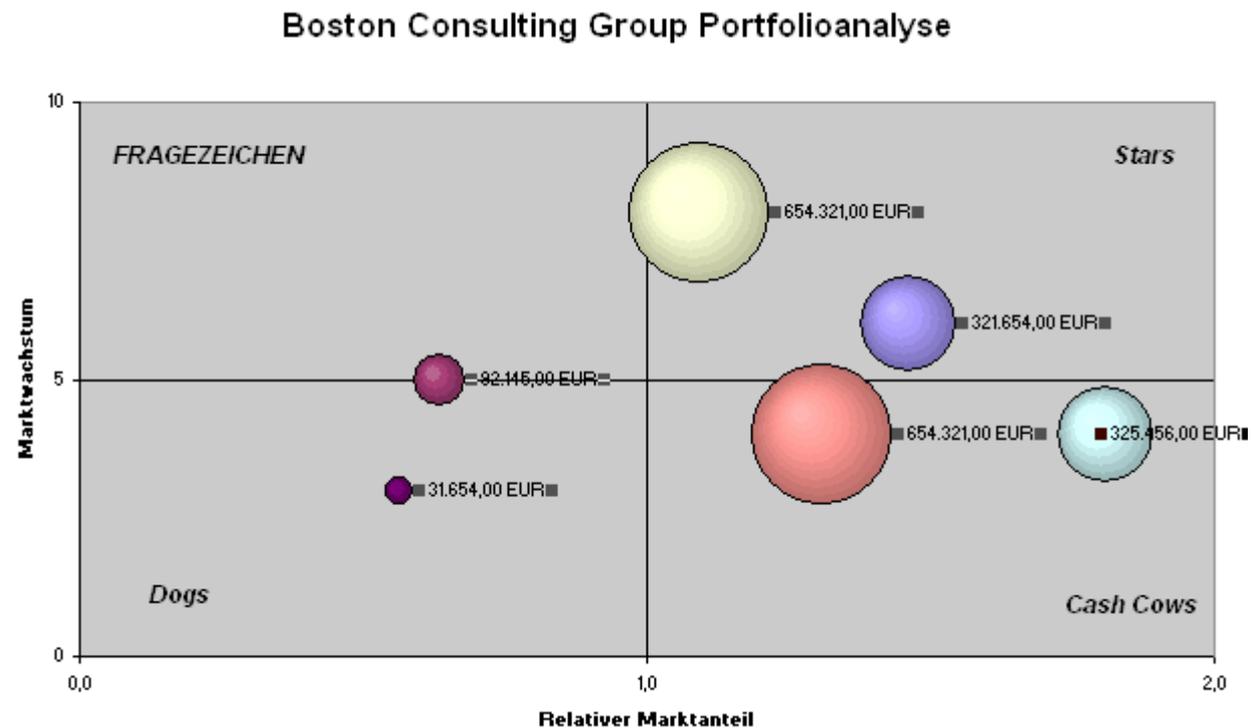
- Brushing in einer Projektion ergibt **andere Selektion** in anderer Projektion



8.2 Scatterplot

Integration von mehr Variablen

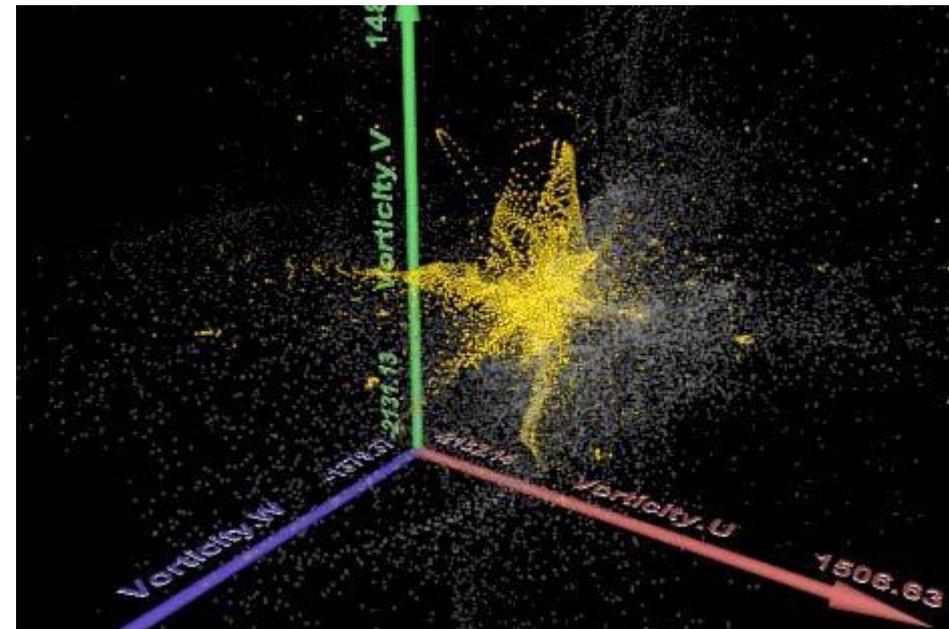
- BubbleGraphs
- Weitere Variable wird mit **Kugelgröße** repräsentiert
- Verstärkt **Verdeckungsproblem**

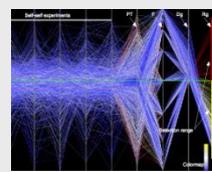


8.2 Scatterplot

Integration von mehr Variablen

- 3D-Scatterplots
- Integration weiterer Variable in **3. Dimension**
- Erhebliches **Verdeckungsproblem**
- **Räumliches** Verständnis schwierig
- **Brushing schwierig**
- **Erfordert Tiefensortierung** für 3D und ggf. Alphablending

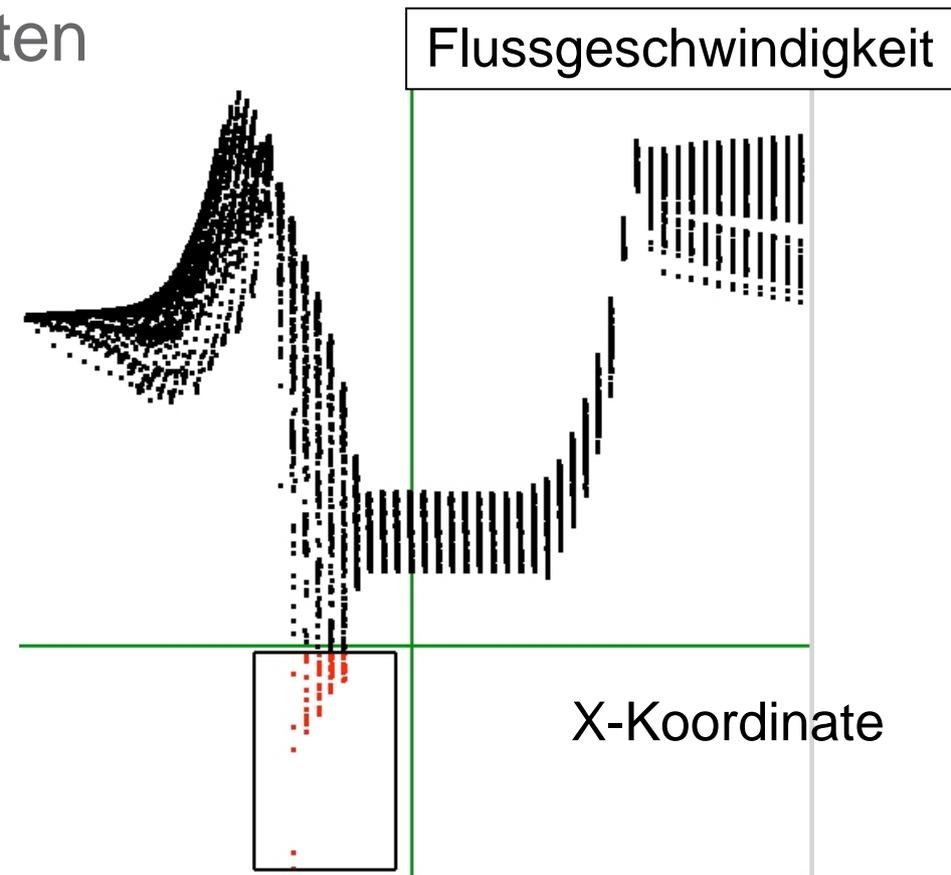


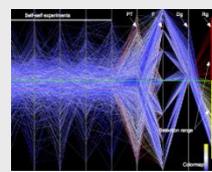


8.2 Scatterplot

Integration von mehr Variablen

- 3D Scatterplots
- Integration von räumlichen Daten
- Scatterplot wird dünnem Volume Rendering ähnlich.

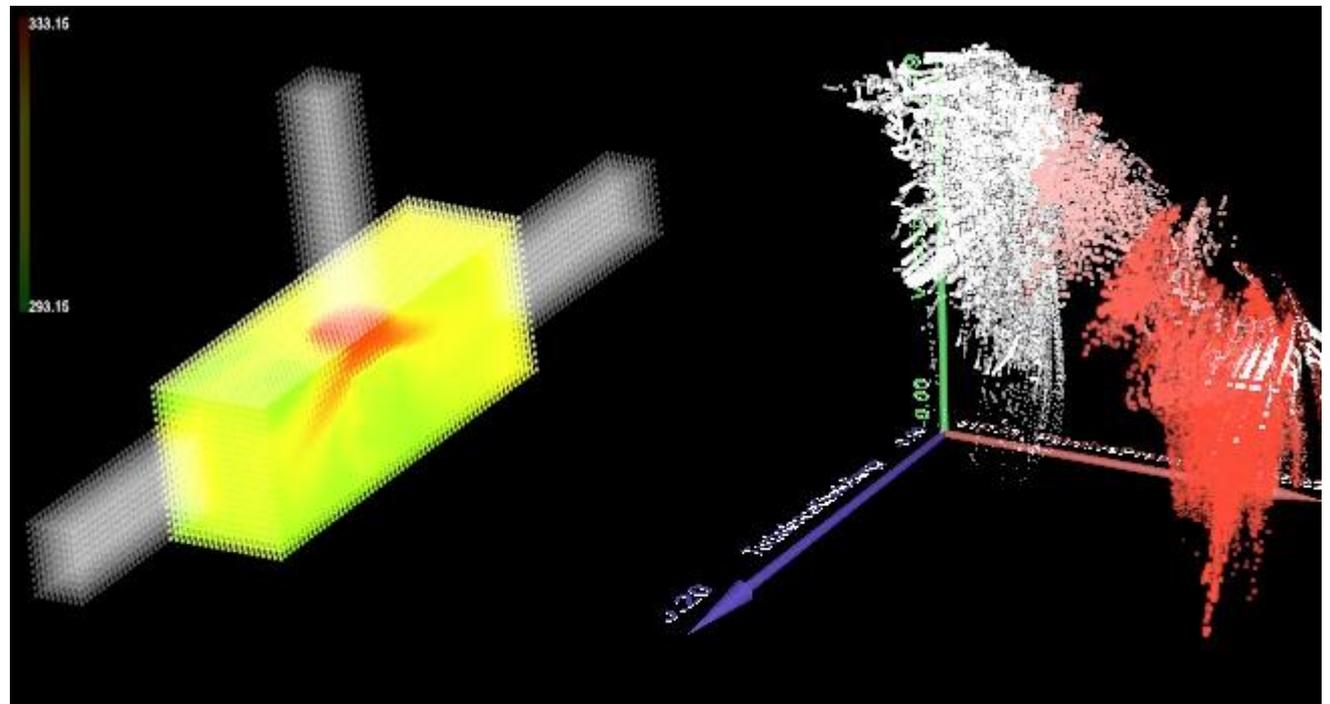


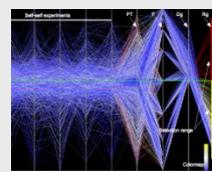


8.2 Scatterplot

Integration von mehr Variablen

- 3D-Scatterplots
- Integration von räumlichen Daten bietet sich an.
- Scatterplot wird dünnem Volumerendering ähnlich

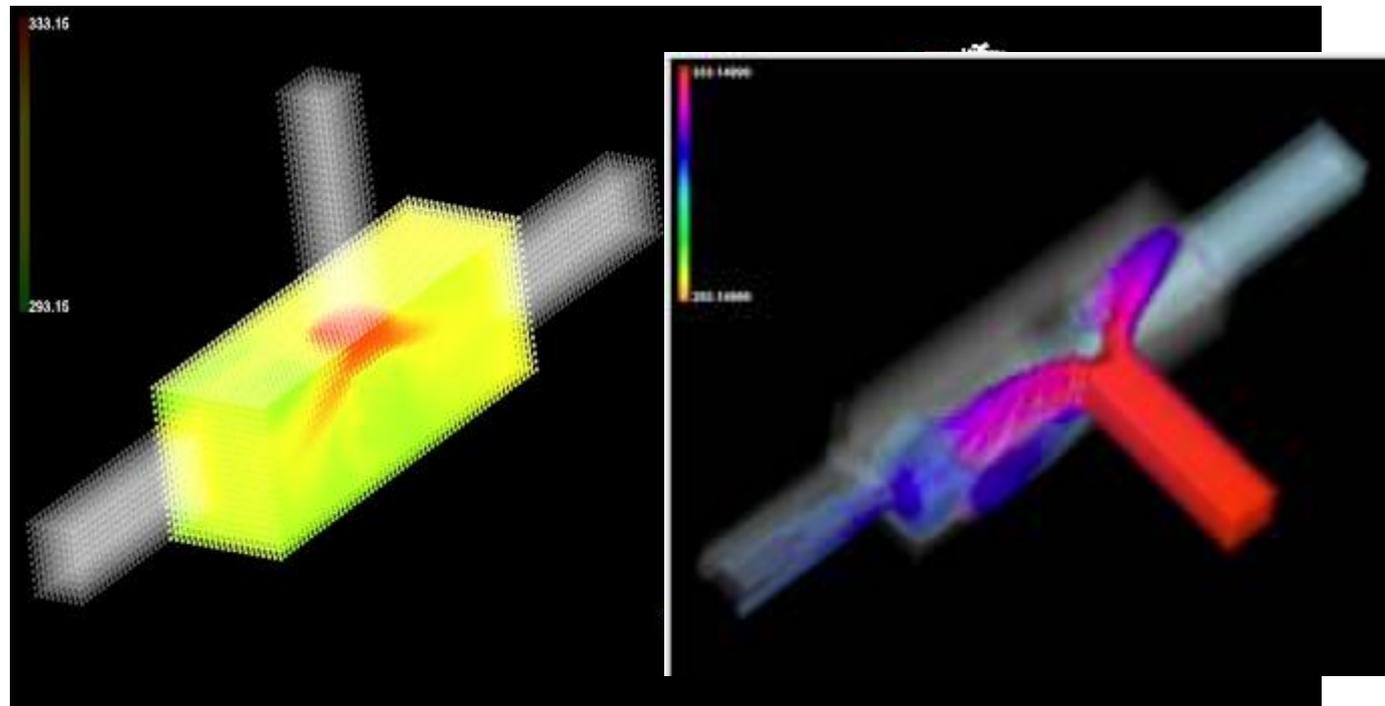




8.2 Scatterplot

Integration von mehr Variablen

- 3D-Scatterplots
- Integration von räumlichen Daten bietet sich an.
- Scatterplot wird dünnem Volumerendering ähnlich

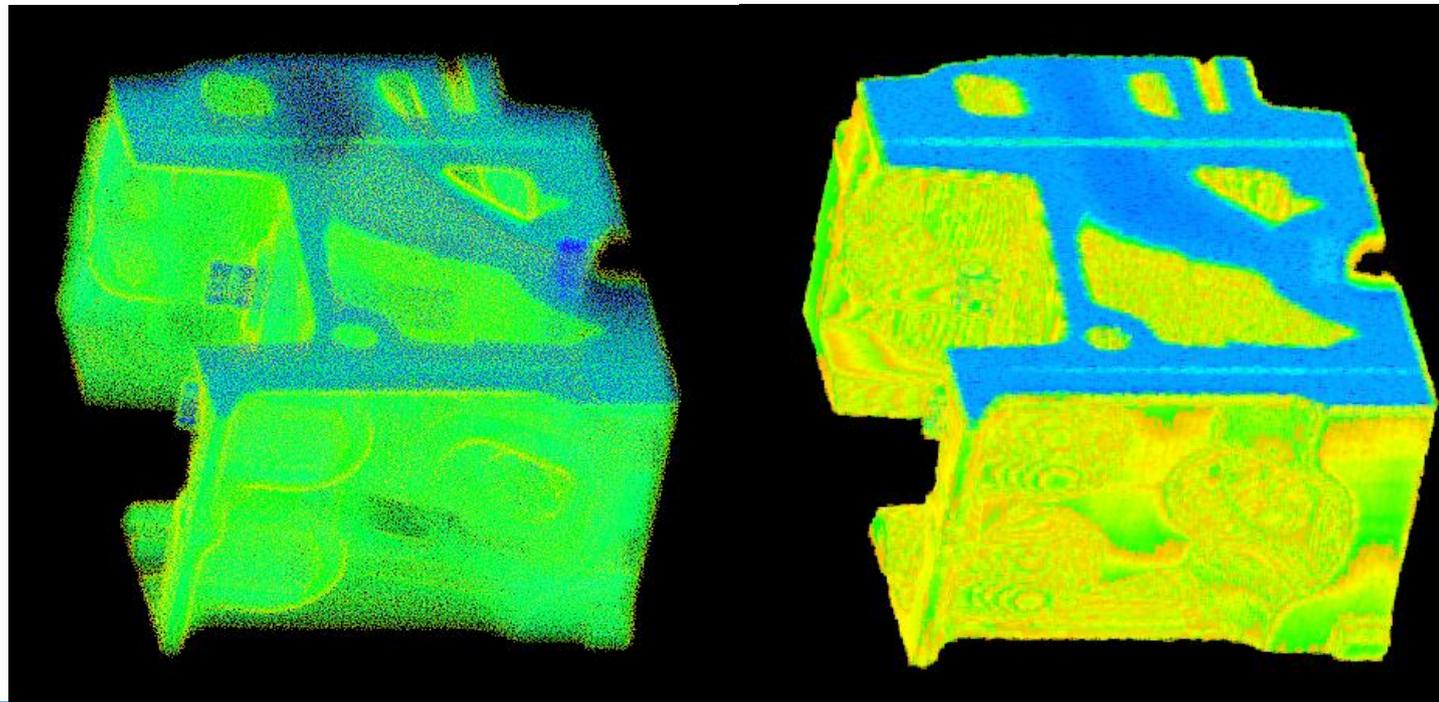


8.2 Scatterplot

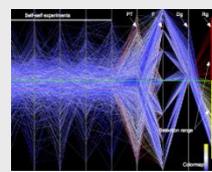
Integration von mehr Variablen

- 3D-Scatterplots
- Integration von **räumlichen Daten** bietet sich an
- Scatterplot wird **dünnem Volumerendering** ähnlich (Punktbasiert/Splatting)

Unterschiedliche
Splatgrößen



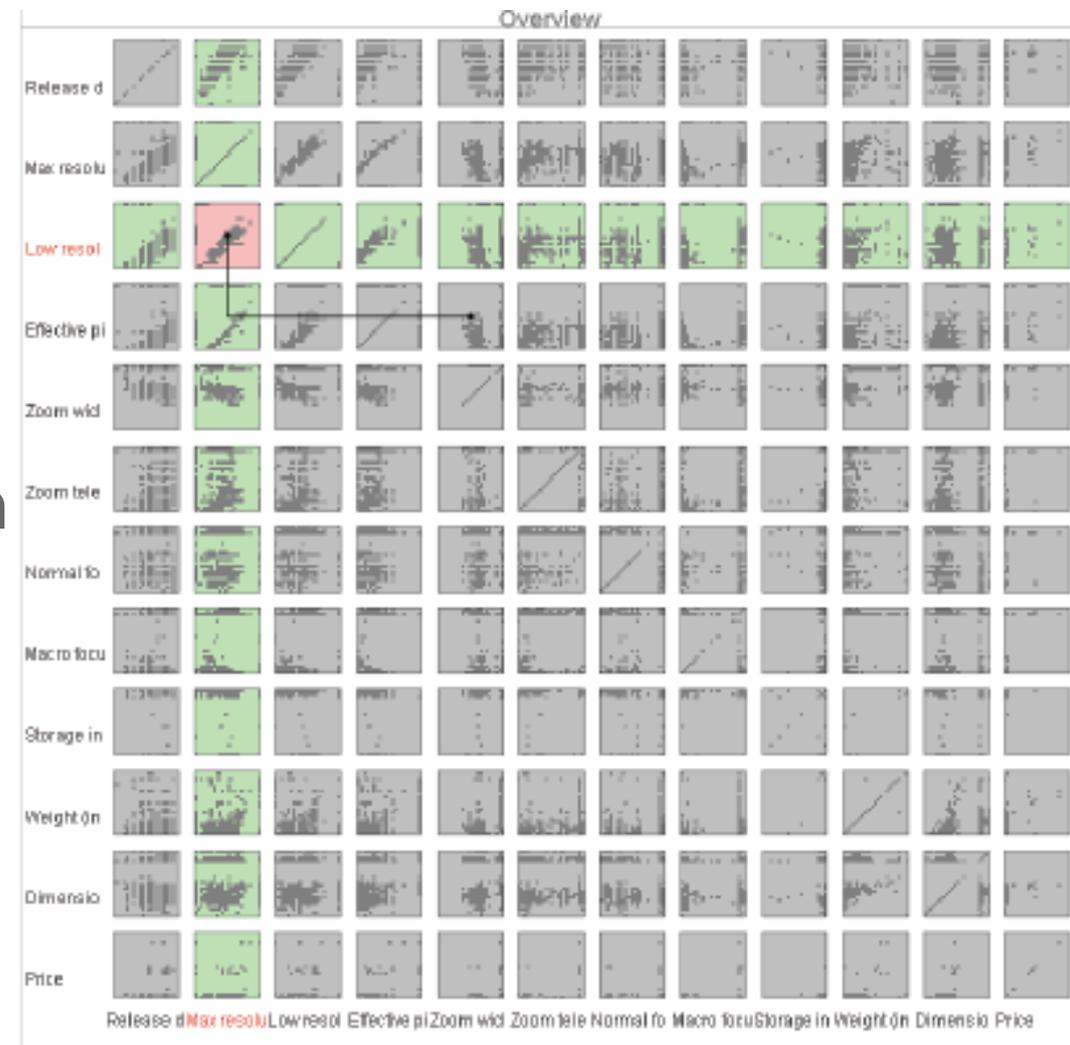
Informationsvisualisierung

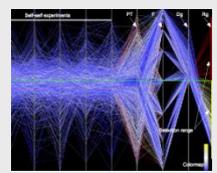


8.2 Scatterplot

Integration von mehr Variablen

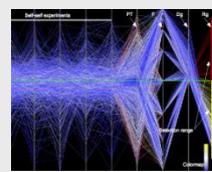
- Scatterplots-**Matrizen**
- **Kombination** aller Parameter
- **Histogramm** auf der Diagonalen
- **Schwierige** Interpretation
- Problematische bei **hoher Dimensionszahl**





8.2 Scatterplot

- Continuous scatterplots
- lit scatterplots
-



Literatur

- Elmqvist, Dragicevic, Fekete: Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation, Proc. InfoVis 2008.
- Few: Information Dashboard Design, O'Reilly, 2006.
- Hahn, Peitgen: IWT- Interactive Watershed Transformation, Proc. SPIE Medical Imaging, pp. 643-653, 2003.
- Hauser, Inselberg, Ward, Yang: Modern Parallel Coordinates: from Relational Information to Clear Patterns, Tutorial at IEEE Visualization, 2006.
- Kosara, Bendix, Hauser: TimeHistograms for Large, Time-Dependent Data. Proc. EG/IEEE Symposium on Visualization, pp. 45-54, 2004.
- Pereda, Vilanova, Serlie, Gerritsen: Visualization of Boundaries in Volumetric Data Sets Using LH Histograms, IEEE TVCG 12(2):208-218, 2006.
- Piringer, Kosara, Hauser: Interactive Focus+Context Visualization with Linked 2D/3D Scatterplots, Proc. Coordinated & Multiple Views in Exploratory Visualization, 2004.
- Spence: Information Visualization: Design for Interaction, Prentice Hall, 2006.