

## 4. Charakterisierung von Datensätzen und Darstellungsformen

### 4.1 Charakterisierung von Datensätzen

große Vielfalt von Daten für die Visualisierung möglich

#### Datenanforderungen:

- Zugriff (Indizierung, Cache-Ausnutzung)
- Platzeffizienz (Kompression, Auslagerung, ...)
- Verlustfrei oder verlustbehaftet
- Portabilität:
  - Binär: kompakt und schnell, aber nicht sehr portable (LSB, MSB)
  - Text: portable, lesbar, aber sehr groß

Für die Charakterisierung der Merkmale interessiert:

- **Wertebereich:** Ist eine Ordnungsrelation über dem Wertebereich definiert? In welchem Bereich liegen die Werte, die die Daten annehmen können und wie umfangreich ist dieser Bereich?
- **Datentyp:** skalare, vektorielle oder tensorielle Daten, wie wurden die Daten diskretisiert ?
- **Dimensionalität:** Anzahl der Datenwerte pro Beobachtungspunkt
- **Fehlergrenzen:** Angabe von Standardabweichung, Varianz
- **Strukturierung:** Beziehungen zwischen den Merkmalen

Entsprechend der Skalierung ihres Wertebereiches lassen sich die **Merkmale klassifizieren** als:

- **Qualitativ**: nicht-metrische Skalen; wenn Ordnungsrelation auf Skala definiert ist, sprechen wir von **ordinalen**, andererseits von **nominalen** Merkmalen.
- **Quantitativ**: metrische Skalen; werden auch als **metrische** Merkmale bezeichnet. Der Wertebereich quantitativer Merkmale kann **diskret** oder **kontinuierlich** skaliert sein.

Beim **Datentyp** unterscheiden wir:

**Skalare Daten**: sind durch eine Funktion  $f(x_1, \dots, x_n): \mathbb{R}^n \rightarrow \mathbb{R}$  in den  $n$  unabhängigen Variablen  $x_i$  gegeben.

**Vektorielle Daten**: repräsentieren Richtung und Betrag und ist als  $n$ -Tupel  $(f_1, \dots, f_n)$  definiert, mit  $f_k = f_k(x_1, \dots, x_n)$ ,  $n \geq 2$  und  $1 \leq k \leq n$ .

**Tensorielle Daten**: ein Tensor  $k$ -ter Stufe hat  $n^k$  Einträge, die über  $k$  Indizes indiziert sind,  $t_{j_1, j_2, \dots, j_k}(x_1, \dots, x_n)$  und ein spezielles Transformationsverhalten aufweisen. Skalare und Vektoren lassen sich als Tensoren 0-ter bzw. 1-ter Stufe auffassen.

Für die **Strukturierung der Merkmale** lassen sich folgende Varianten unterscheiden:

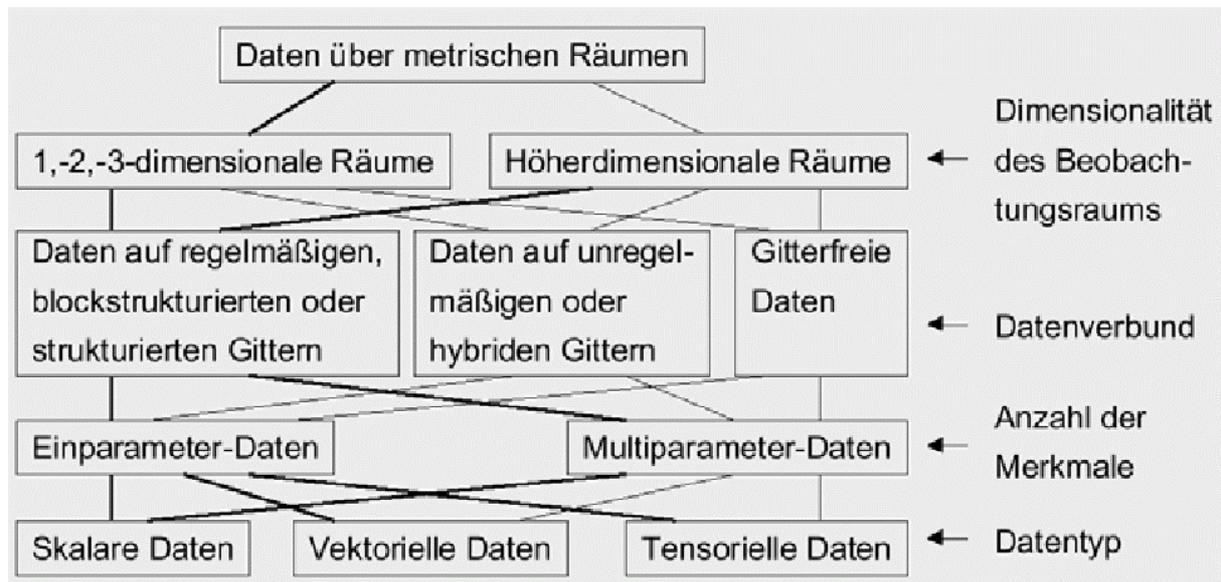
**Sequentiell**: Speicherung von Merkmalen und ihren Ausprägungen in einer Liste;

**Relational**: Speicherung von Merkmalen und ihren Ausprägungen in Tabellenform;

**Hierarchisch**: Speicherung von Merkmalen und ihren Ausprägungen in einer Baumstruktur;

**Netzwerkstrukturiert**: Speicherung von Merkmalen und ihren Ausprägungen in einer Netzwerkstruktur.

## Überblick – Typisierungen von Daten:



hierin nicht berücksichtigt: komplexe Datentypen (Texte, Webseiten, Multimedia-Daten) – Gegenstand der *Informations-Visualisierung* (siehe später)

### Weitere interessante Eigenschaften:

**Zusammensetzung der Datenmenge:** Werden unterschiedliche Datentypen erfasst?  
Wie unterscheiden sich die Wertebereiche?

**Umfang der Datenmenge:** Wie viele Merkmale werden wie oft und an welchen Beobachtungspunkten erfasst?

**Qualität der Datenmenge:** Sind alle Werte vollständig erfasst? Lassen sich fehlende Werte durch Interpolation ergänzen? Treten fehlerhafte Werte auf?

Die in der Datenmenge **enthaltene Information:** Gibt es bestimmte Zusammenhänge? Treten Redundanzen auf?

(Bartz 2005)

In der Literatur gibt es verschiedene Ansätze für eine formale Systematik der Datensatz-Typen.

### Spezifikation nach Bergeron & Grinstein, 1989:

Wissenschaftliche Daten sind  $m$ -dimensionale Daten auf einem  $n$ -dimensionalen Gitter, mit  $L_m^n$  bezeichnet, also  $m$ -dimensionale Daten

- auf beliebigen Positionen ( $L_m^0$ ),
- über einer Linie ( $L_m^1$ ),
- über einer Fläche ( $L_m^2$ ),
- auf einem (regelmäßigen) 3D-Gitter ( $L_m^3$ ) oder
- auf einem (regelmäßigen)  $n$ -dimensionalen Gitter ( $L_m^n$ ).

Diese Klassifikation ist einfach zu handhaben, es fehlen aber wichtige Aspekte wie die Spezifikation des Daten- oder Gittertyps

### Spezifikation nach Brodli 1992:

- Ausgangspunkt: "**Underlying Field**", ein Modellraum, in dem die Daten erhoben werden. Hieraus wird die zu "*visualisierende Entität*" ( $E$ ) extrahiert.
- $E$  ist eine Funktion, die durch ihren Definitions- und Wertebereich spezifiziert wird.
- Die unabhängigen Variablen werden in ihrer Dimensionalität und in ihrem Wirkungskreis beschrieben, die abhängigen Variablen durch Dimensionalität und Typ, z.B.:

$$E_n^{5S} \text{ oder } E_{[3]}^{V3}$$

### Spezifikation nach Wong 1997:

Hier wird lediglich zwischen der

- Dimensionalität des Beobachtungsraumes (Anzahl der unabhängigen Variablen  $\mathbf{d}$ ) und der
- Dimensionalität der Daten (Anzahl der abhängigen Variablen  $\mathbf{v}$ )

unterschieden.

Eine Datenmenge mit  $n$  unabhängigen und  $m$  abhängigen Variablen ist damit gegeben durch:

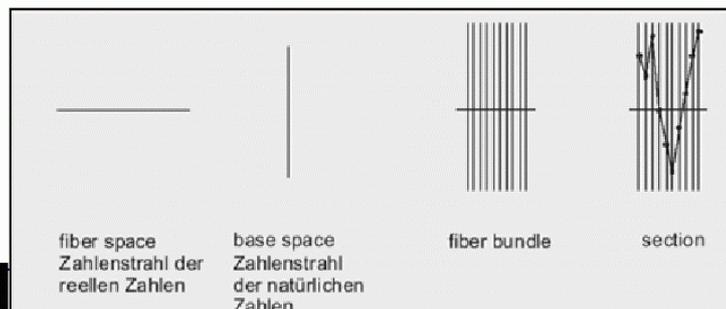
$$\mathbf{ndmv}$$

## Spezifikation von „Fiber-bundles“ nach Butler 1989:

Ein **Fiber-bundle** besteht aus 2 Räumen:

- **Base-space** - mehrdimensionaler Raum mit vorgegebener Topologie, enthält unabhängige Variable
- **Fiber-space** - enthält abhängige Variable (z.B. mehrdimensionaler Vektorraum)

Zur Spezifikation von Daten werden **Abschnitte (Sections)** im **Fiber-space** gebildet.



## Spezifikation nach Arndt (1993):

$$Zc^{T_1, \dots, T_k} n, g, w, t$$

- mit
- $Z$  : Umfang der Datenmenge
  - $c$  : Qualität der Datenmenge
  - $n$  : Dimensionalität des Beobachtungsraumes
  - $g$  : Datenverbund (Gitter)
  - $w$  : Wirkungskreis
  - $t$  : Zeitbezug mit  $t = \{t_a, t_{dis}, t_e\}$ , wobei
    - $t_a$  : Anfangszeitpunkt
    - $t_{dis}$  : Größe der Zeitintervalle
    - $t_e$  : Endzeitpunkt

$T_i$  steht für ein Tupel  $T_i = (ty_i, o_i, u_i, d_i)$  mit

- $ty_i$  : Datentyp
- $o_i$  : Existenz einer Ordnungsrelation über dem entsprechenden Wertebereich (ja/nein)
- $u_i$  : Umfang des Wertebereichs
- $d_i$  : Anzahl der Merkmale, die durch das Tupel beschrieben werden

Es gibt keine Visualisierungsmethoden, die die Vielfalt der wissenschaftlichen Daten behandelt.

**Beispiel:** Ungeordnete Menge von 137 Punkten mit skalaren Werten

- Bergeron & Grinstein  $L^0_1$
- Brodlie  $E^S_{\{0\}}$
- Butler **base = set, fiber = float:[ $-\infty, \infty$ ]**
- Wong **0d1v**
- Arndt  $137c^{(float,yes,[-\infty,\infty],1)}$  0, set, point

**Beispiel:** Geordnete Menge von 137 Punkten mit skalaren Werten

- Bergeron & Grinstein  $L^0_1$
- Brodlie  $E^S_{[0]}$
- Butler **base = ordered set, fiber = float:[ $-\infty, \infty$ ]**
- Wong **0d1v**
- Arndt  $137c^{(float,yes,[-\infty,\infty],1)}$  0, ord - set, point

**Beispiel:** Skalare Volumendatensatz von unsigned char auf regulärem 128x128x128 Gitter

- Bergeron & Grinstein  $L^3_1$
- Brodlie  $E^S_3$
- Butler **base = 3D-regular-grid, fiber = char:[0,255]**
- Wong **3d1v**
- Arndt  $2097152c^{(char,yes,[0,255],1)}$  3, regular - grid, local

**Beispiel:** Strömungsdaten auf curvilinearem Gitter mit 37x64x60 Gitterlinien

- Bergeron & Grinstein  $\mathbf{L}_3^3$
- Brodlie  $\mathbf{E}^{\mathbf{V}^3}_3$
- Butler **base = 3D-curvilinear-grid,  
fiber = float<sup>3</sup>:[-∞, ∞]<sup>3</sup>**
- Wong **3d3v**
- Arndt  $142080c^{(float^3, no, [-∞, ∞]^3, 1)} 3, curvilinear - grid, local$

**Beispiel:** Volumen (64x64x64) mit 3 skalaren, 2 vektoriellen und einem tensoriellen Merkmal

- Bergeron & Grinstein  $\mathbf{L}_3^{15}$
- Brodlie  $\mathbf{E}^{3\mathbf{S}2\mathbf{V}3\mathbf{T}^3}_3$
- Butler **base = 3D-regular-grid,  
fiber = float x float x float x float<sup>3</sup> x float<sup>3</sup> x float<sup>6</sup>**
- Wong **3d15v**
- Arndt  $262144c^{(float, yes, [-∞, ∞], 3), (float^3, no, [-∞, ∞]^3, 2), (float^6, no, [-∞, ∞]^6, 1)}$   
 $3, regular - grid, local$

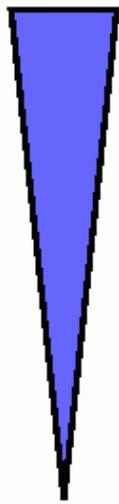
(Bartz 2005)

## 4.2 Darstellungstechniken

Theorie der Graphischen Semiotik (J. Bertin 1967):

Wissenschaft der graphischen Zeichen

- Methode zur Erzeugung **effektiver** Visualisierung
- Grundlage: **Identifikation der Charakteristika** der Datenmenge
- Abbildung auf **visuelle Variablen**:
  - Position      • Helligkeit      • Farbe              • Form
  - Größe         • Textur         • Orientierung
- Klassifikation nach  
selektiv  $\leftrightarrow$  ordinal  $\leftrightarrow$  proportional

	quantitativ	ordinal	nominal	
<i>effektiv</i>	Position	Position	Position	
	Länge	Helligkeit	Farbton	
	Winkel	Sättigung	Textur	
	Orientierung	Farbton	Verbindung	
	Fläche	Textur	Umfassung	
	Volumen	Verbindung	Helligkeit	
	Helligkeit	Umfassung	Sättigung	
	Sättigung	Länge	Form	
	Farbton	Winkel	Länge	
	Textur	Orientierung	Winkel	
	Verbindung	Fläche	Orientierung	
	Umfassung	Volumen	Fläche	
	<i>weniger effektiv</i>	Form	Form	Volumen

[Mackinlay, TOG 1986] auf Basis von psychologischer Tests  
(Bartz 2005)

- Position kann in allen Bereichen optimal eingesetzt werden.
- Linienlänge kann für proportionalen Vergleich eingesetzt werden.
- **Vorsicht** bei der Wahrnehmung von Flächen: Bei der Flächenwahrnehmung ist  $n=0.7$  im Steven'schen Gesetz:

$$A_{empfundene} \approx A_{Reiz}^{0.7}$$

- Kreisradien sollten zum Beispiel wie folgt gewählt werden:

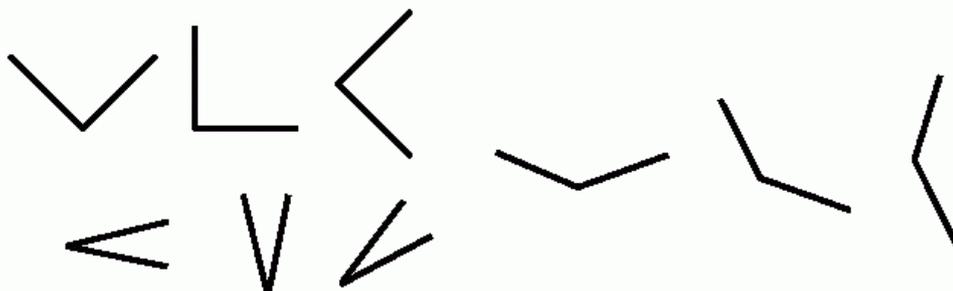
$$r_{Reiz} \approx (A_{empfundene}^{5/7}) / \sqrt{2\pi}$$



- Nominale und quantitative Daten können auf unterschiedliche Rotationen abgebildet werden. Bei nominalen Daten nicht mehr als 16 Stufen:

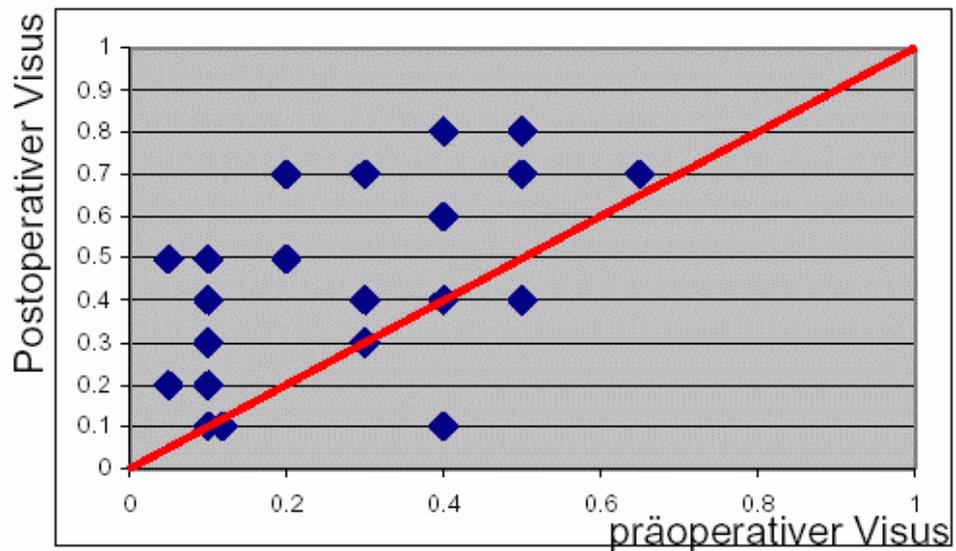


- **Vorsicht** bei der Darstellung mit Winkeln. Stumpfe Winkel oder Winkel mit horizontaler Winkelhalbierenden werden *überschätzt*.
- Spitze Winkel oder Winkel mit vertikaler Winkelhalbierenden werden *unterschätzt*.



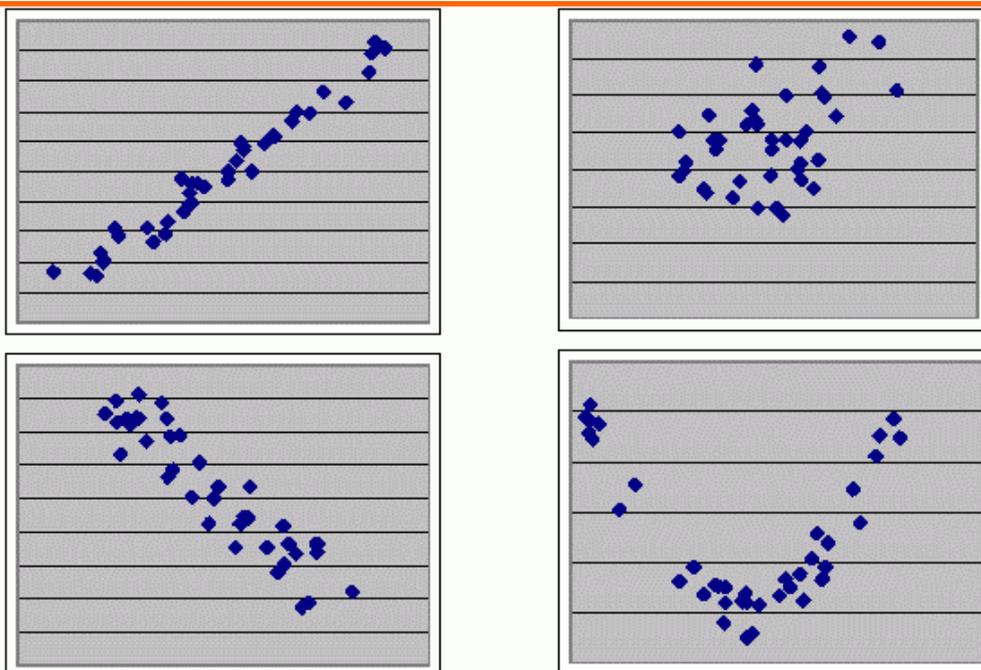
## Punktdiagramme / Scatter Plots:

Ausnutzung der guten Positionswahrnehmung.



## Hornhauttopographische Astigmatismusreduktion

(Beispiel; aus Bartz 2005.)

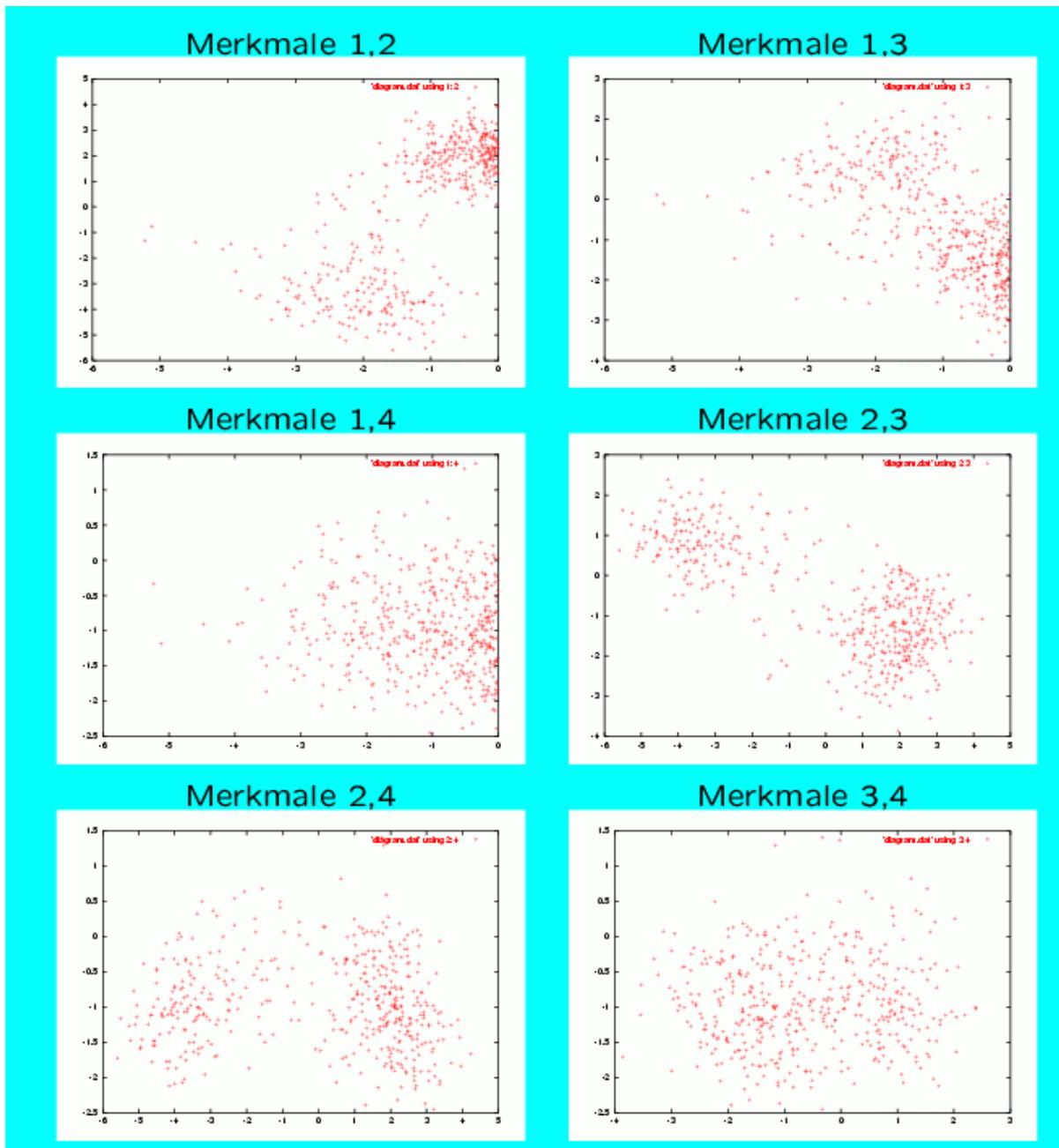


## Visuelle Korrelationsanalyse

(Bartz 2005)

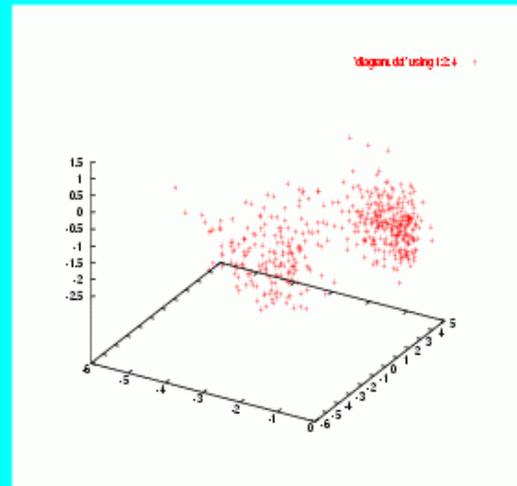
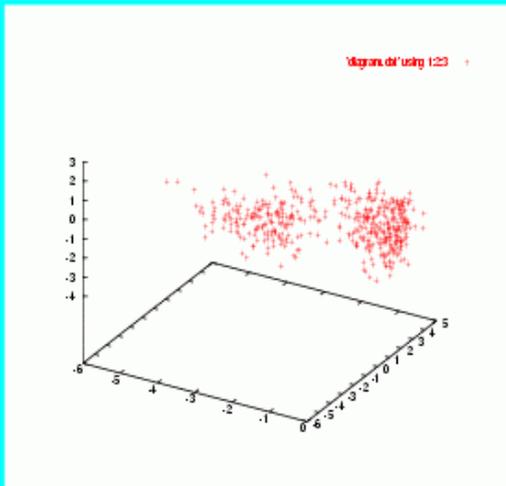
bei mehr als 2 Merkmalen: Alle möglichen Merkmalspaare in 2D-Scatterplots darstellen

Beispiel:



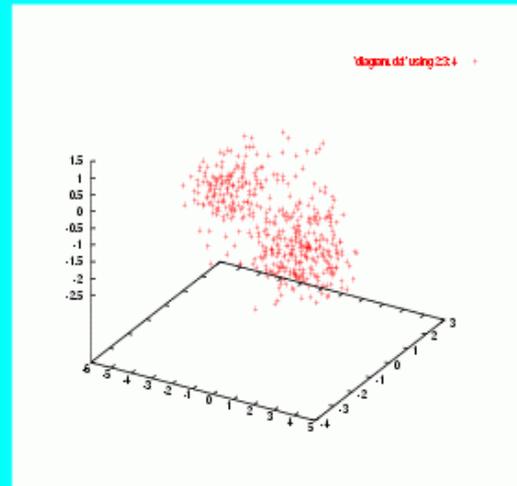
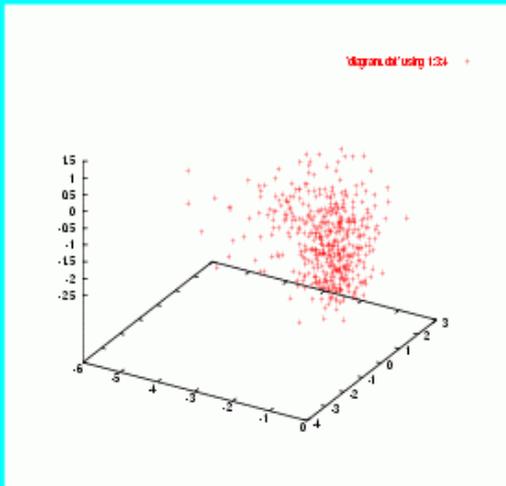
Es gibt  $\binom{N}{2} = (N-1) \cdot N / 2$  viele 2D-Streudiagramme!

Alternative: 3D-Scatterplots



Merkmale 1,2,3 ⬆️ und 1,3,4 ⬇️

Merkmale 1,2,4 ⬆️ und 2,3,4 ⬇️



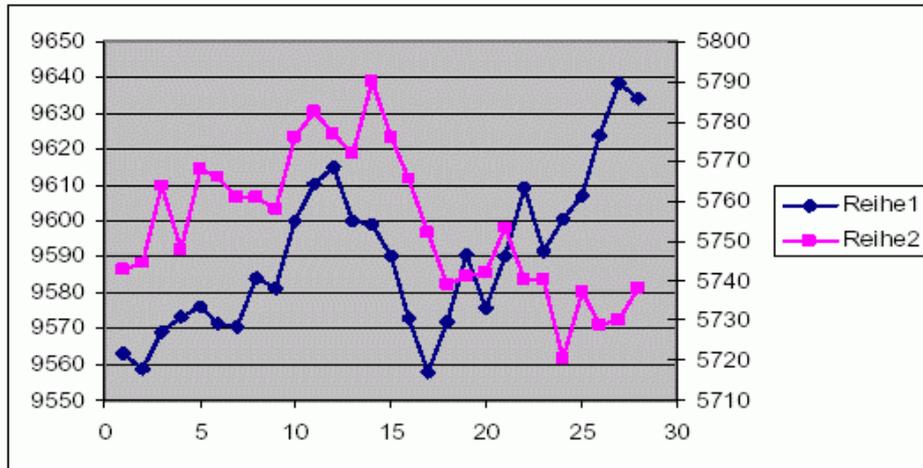
Es gibt  $\binom{N}{3} = (N-2)(N-1)N/6$  viele 3D-Streudiagramme!

Mehr als 3 Merkmale sind nicht gleichzeitig in einem Scatterplot darstellbar.

Vorsicht bei Projektionen von 3D-Punktwolken:  
Informationsverlust; Gefahr von Täuschungen!

## Kurvendiagramme:

Gruppierung durch Verbindung und zusätzliche Ausnutzung der guten Längenwahrnehmung.

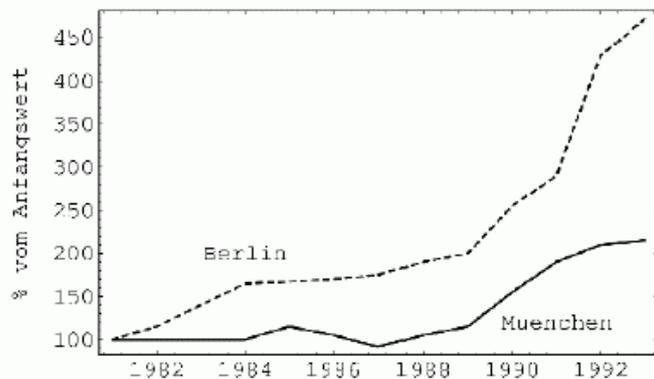
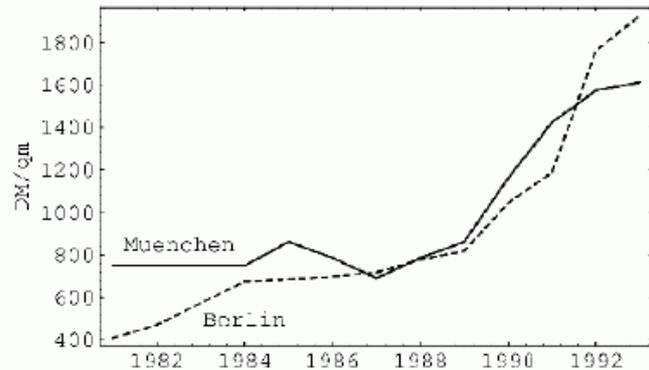


(Bartz 2005)

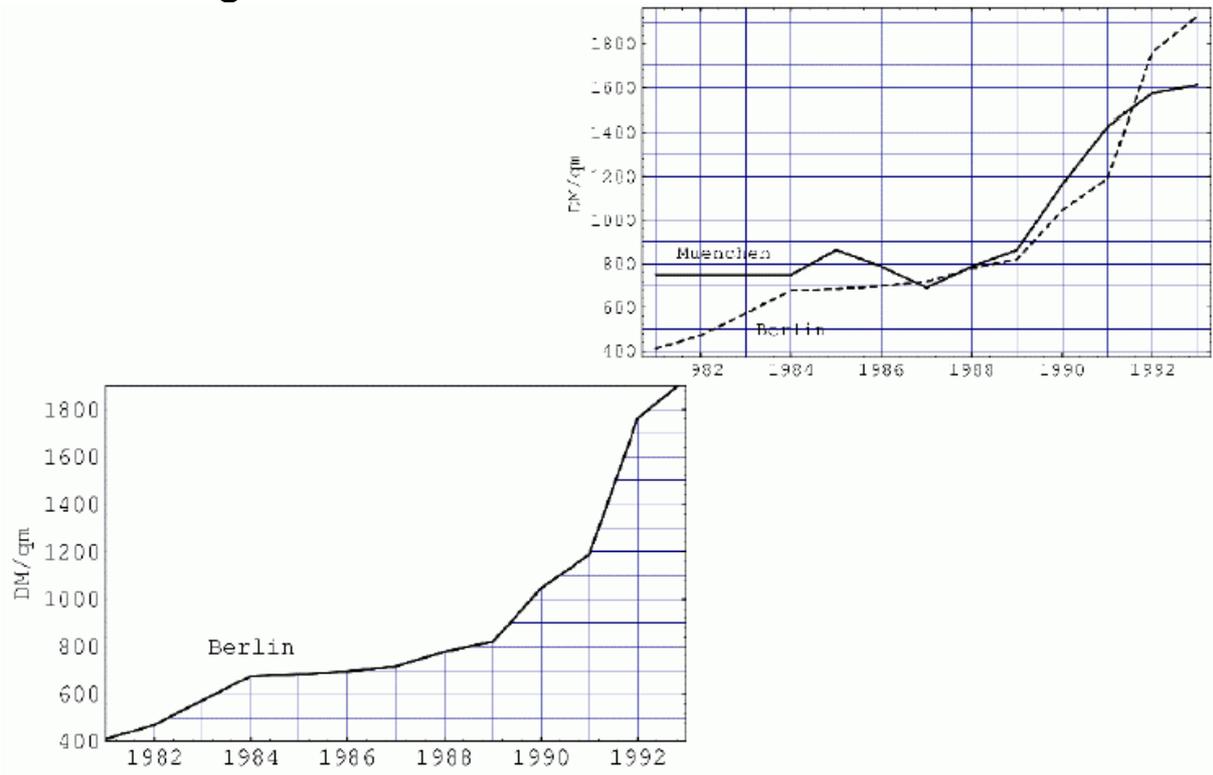
zu beachten:

Zwei Kurven in einem Diagramm:

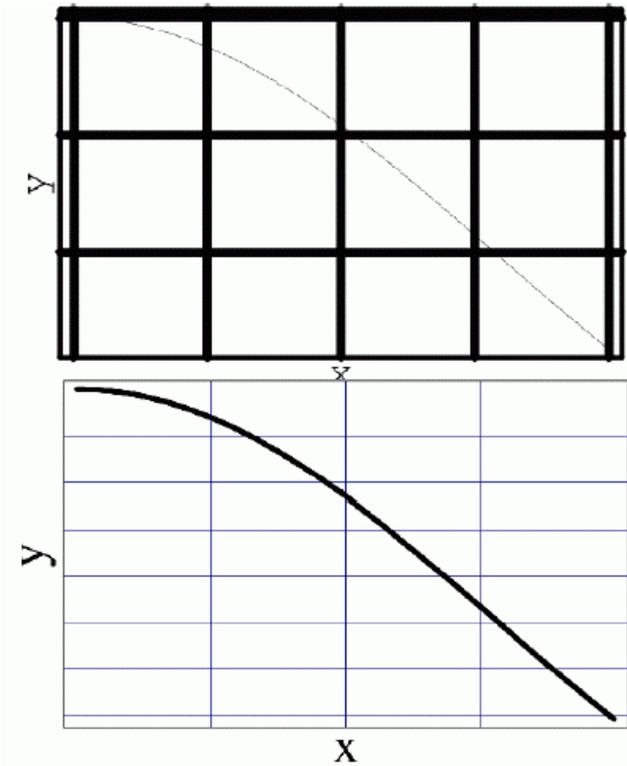
Beide mit denselben Maßeinheiten!



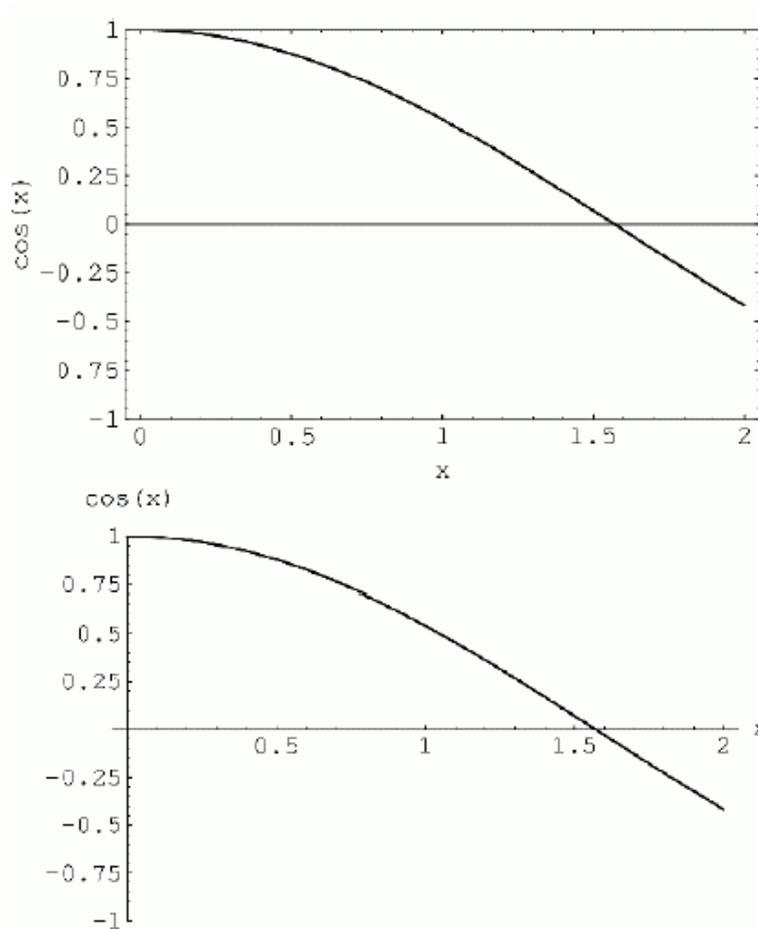
# Überflüssiges vermeiden:



Daten sind wichtiger als Gitter und Achsen!

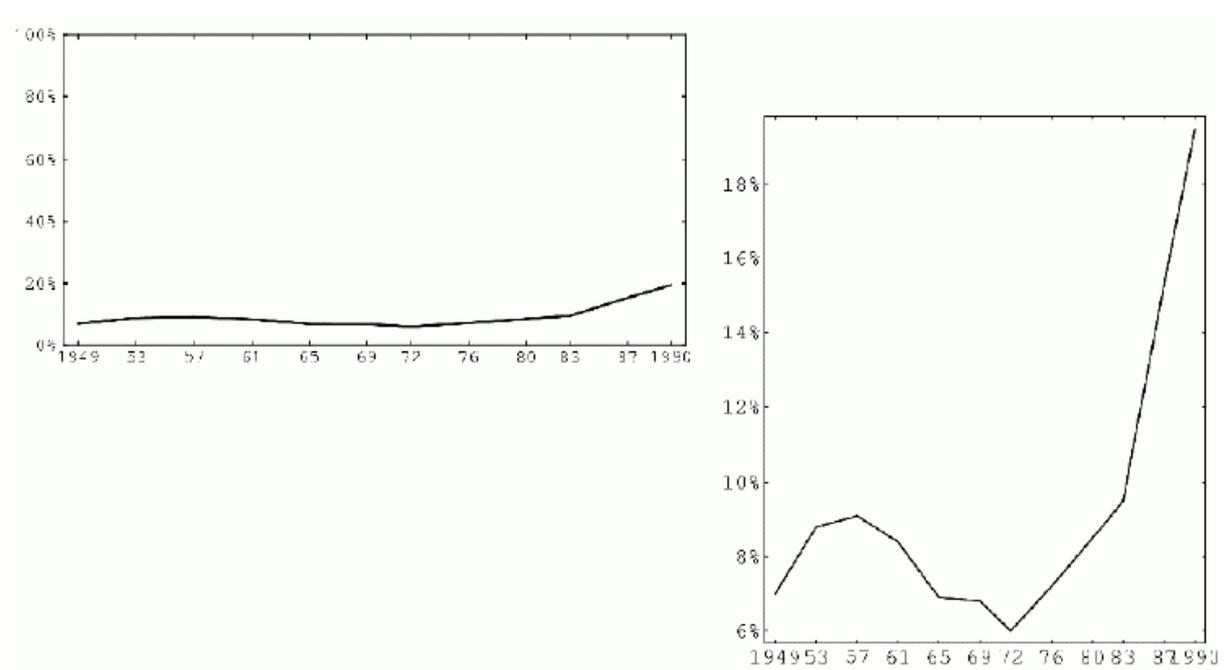


## Achsen oder Rahmen?



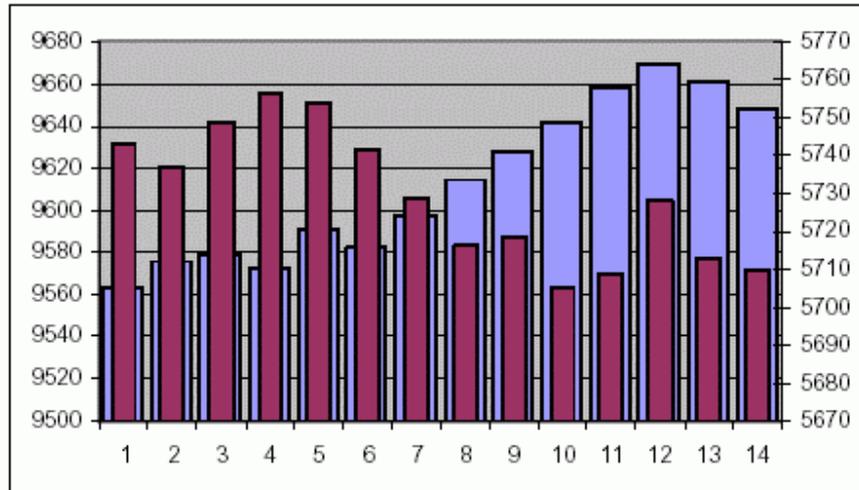
(Weimar 2005)

## Vorsicht vor Manipulationen!



## Säulen- / Balkendiagramme:

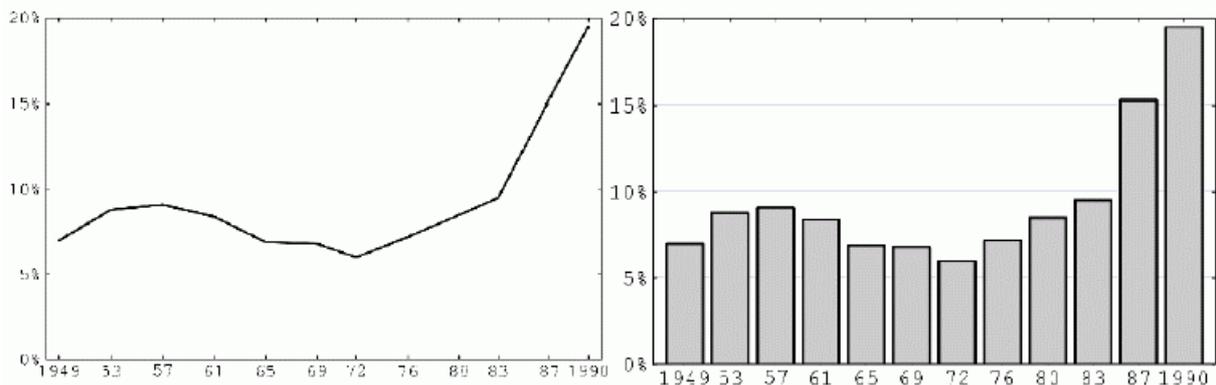
Bei diskreter unabhängiger Variable und quantitativen abhängigen Variablen.



(Bartz 2005)

## Kurven- oder Balkendiagramm?

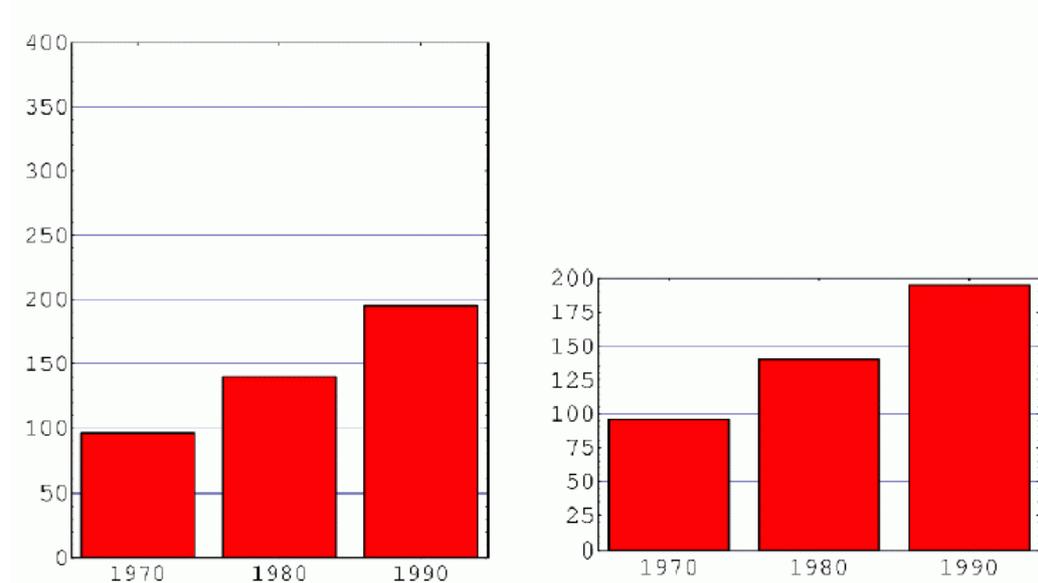
### Anteil der Frauen in deutschen Bundestag



(Weimar 2005)

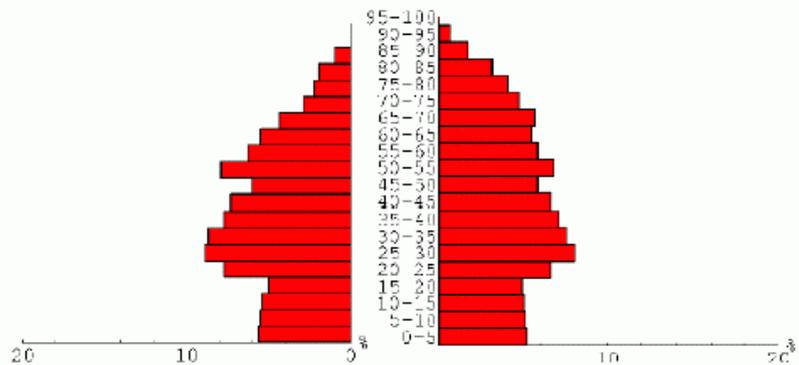
zu beachten:

- Keinen Platz verschwenden

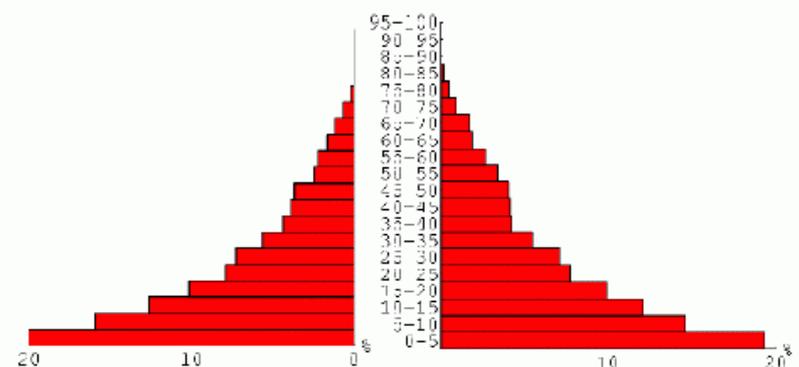


auf gleiche Maßstäbe bei Vergleichen achten

Deutschland (1991)

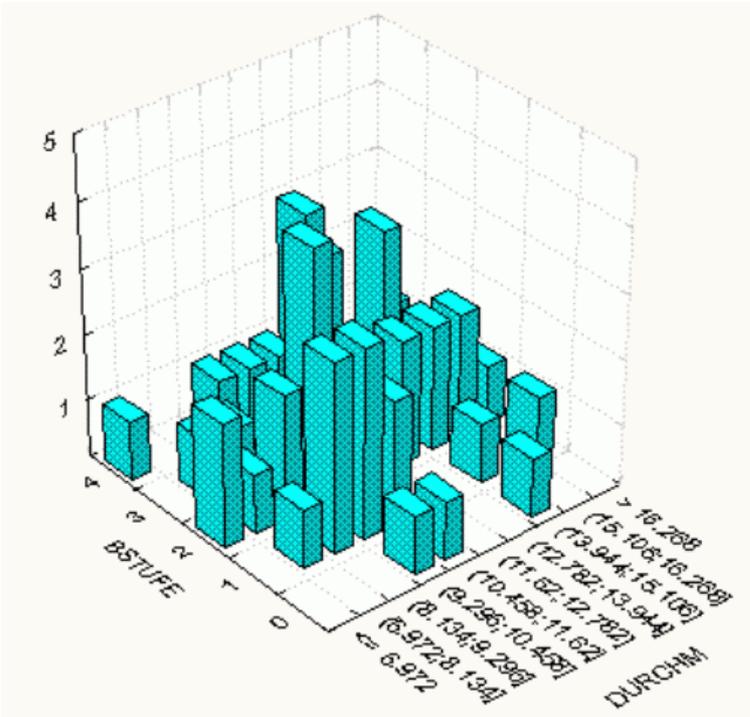


Tansania (1990)

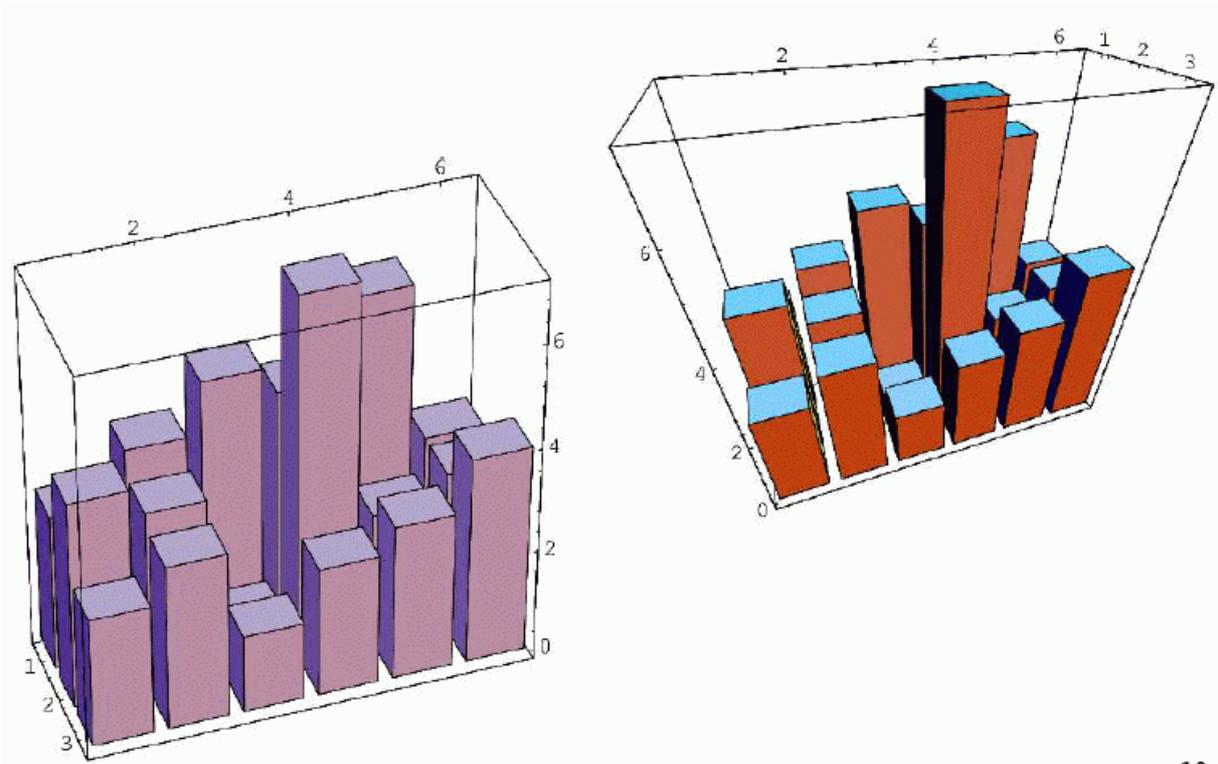


zwei Bevölkerungspyramiden (aus Weimar 2005)

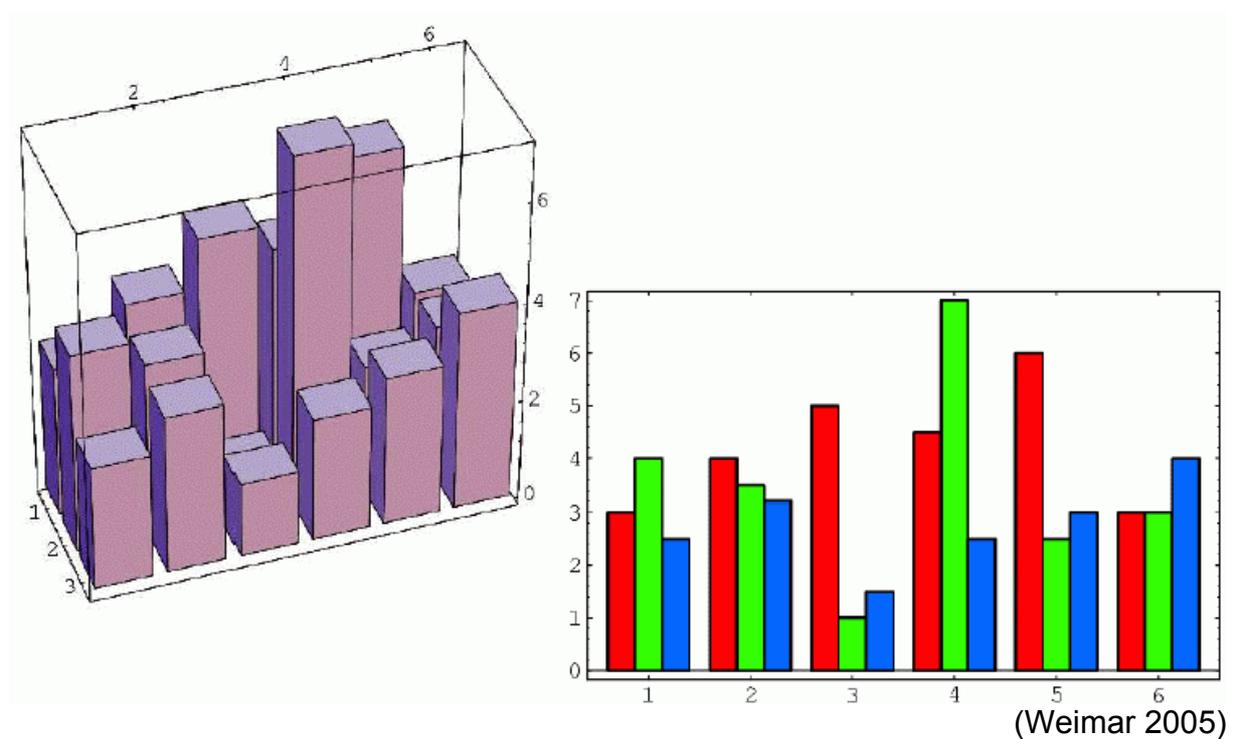
# 3D-Balkendiagramm:



# Vorsicht mit Perspektive!

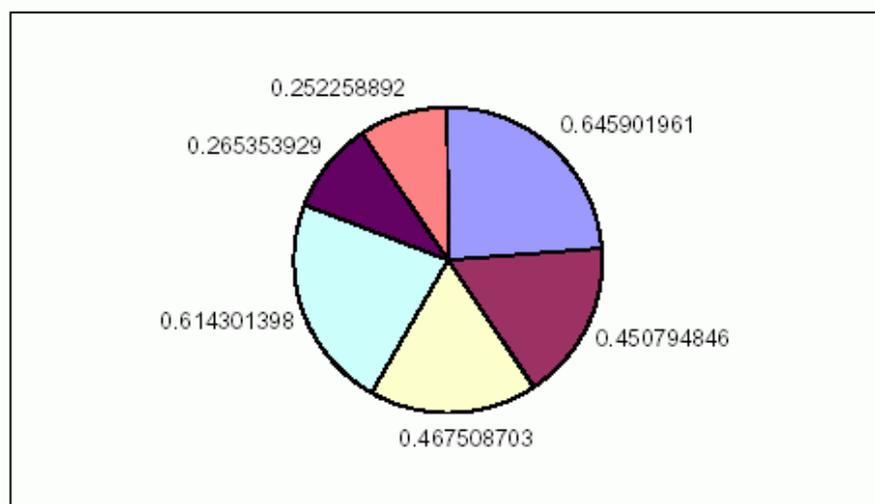


bei wenigen Ausprägungen des zweiten Merkmals:  
evtl. Balkendiagramm lieber in 2D



### Kreisdiagramme:

Für quantitative Größen, die sich zu einer Gesamtheit addieren. Wird im Allgemeinen von abgeraten.



(Bartz 2005)

- Torte als Teil einer Torte: schlecht.

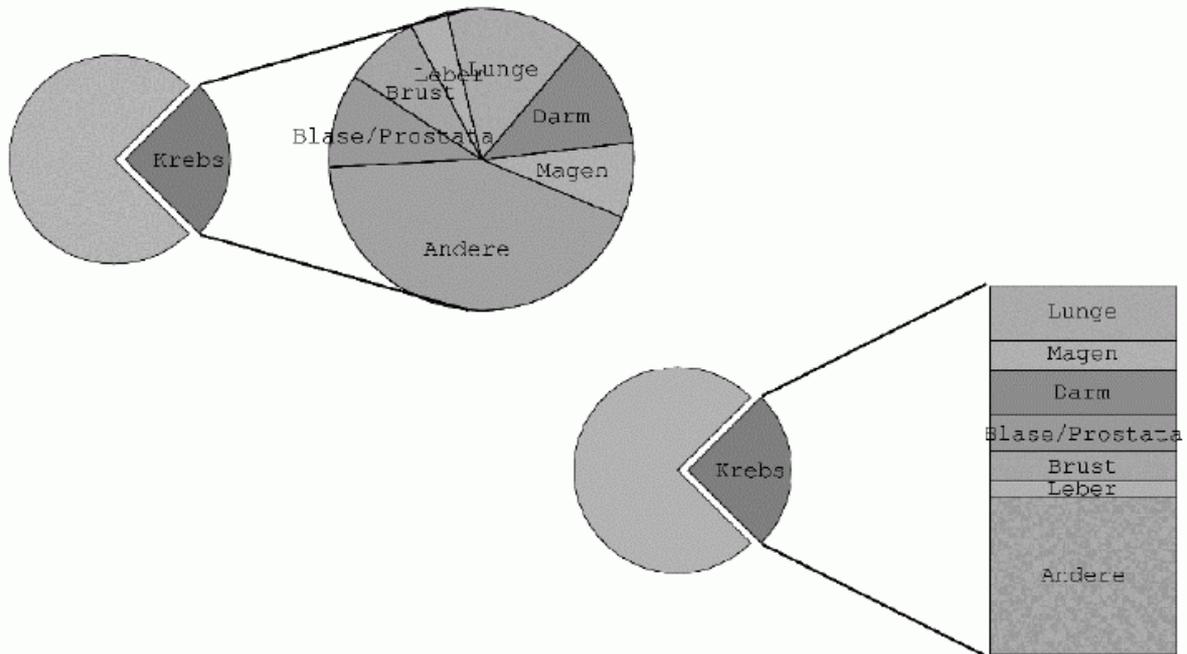
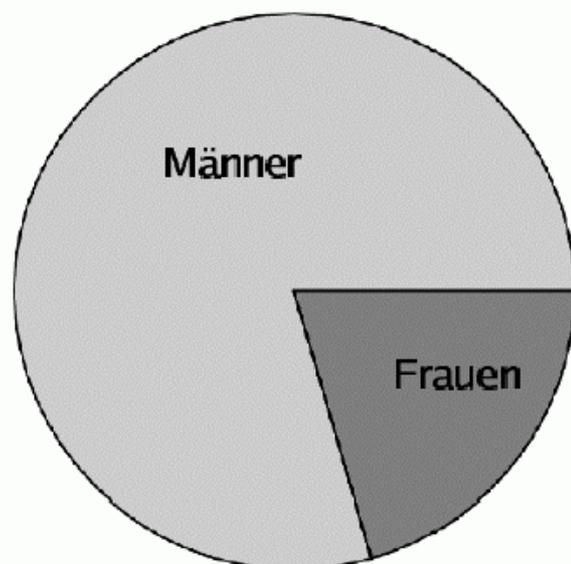


Tabelle oder Grafik?

### Zweitstimmen

CDU/CSU	43.8 %
SPD	33.5 %
FDP	11.0 %
Grüne	3.8 %
PDS	2.4 %
Sonstige	5.4 %

### Frauenanteil



- **Tabelle**

- ♦ **Exakte Zahlen wichtig**
  - Bilanz
  - Kontoauszug
  - Fahrplan
  - Gehaltsabrechnung
- ♦ **Langzeitgedächtnis**
- ♦ **Nachkommastellen**
  - **Wieviele signifikant?**

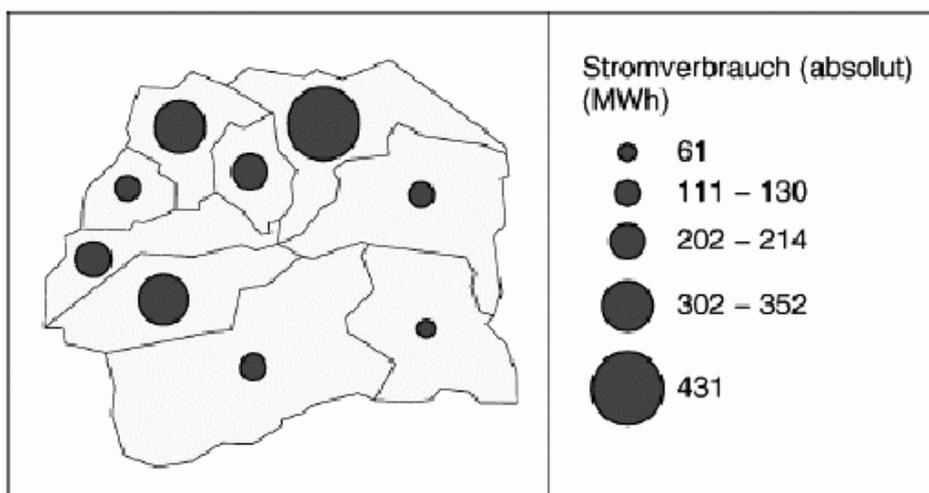
- **Graphik**

- ♦ **Schnelle Übersicht**
- ♦ **Für Leute, die % nicht verstehen.**
- ♦ **Kurzzeitgedächtnis**
- **Welche Graphik**
  - ♦ **Welche Daten?**
  - ♦ **Wer ist Empfänger?**
  - ♦ **Was will ich hervorheben?**

(Weimar 2005)

### **Kreissignaturen:**

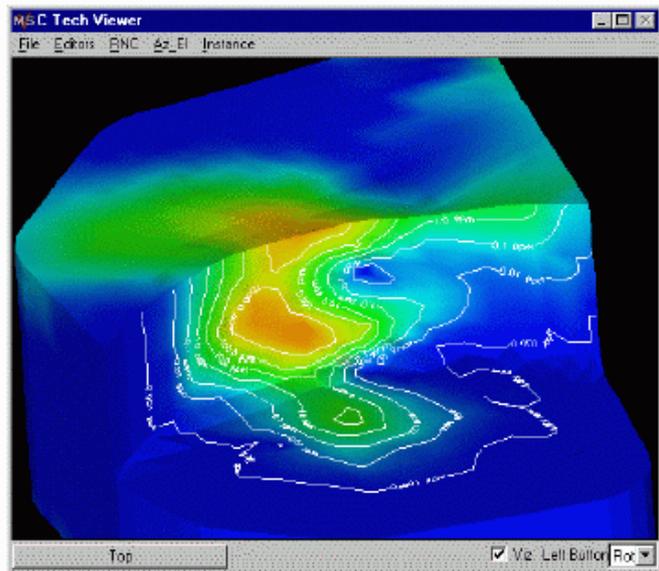
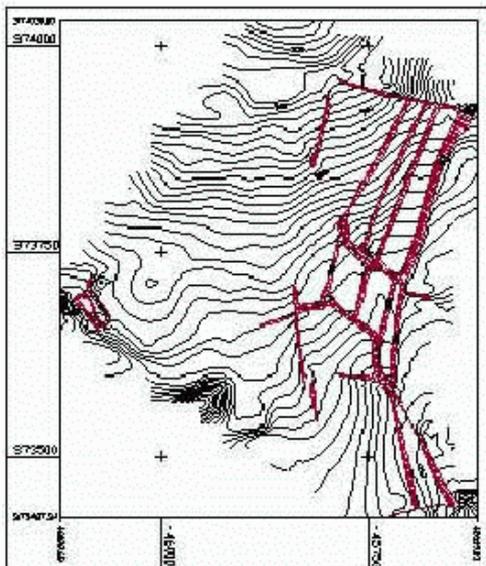
Quantitative Merkmale werden auf Kreisgröße abgebildet, meist bei Karten.



(Bartz 2005)

## Isolinien und Höhenfelder

- Gegeben sei ein qualitatives Merkmal über der Ebene:  $f(x,y) \rightarrow [a,b]$  (z.B. Höhenfeld)
- **Isolinie** bezeichnet Punktmengen mit  $f(x,y) = \text{const.}$
- **Isostreifen** bezeichnet Punktmengen mit  $f(x,y) \in [v, v+\Delta v]$ .
- Falls  $f(x,y)$  auf keinem ausgedehnten Bereich konstant ist, sind Isolinien Linien.



Anwendungen: Höhenlinien, Wetterkarten,...

(Bartz 2005)

wie lassen sich Isolinien erzeugen?

- für alle Pixel überprüfen, ob Isolinie hindurchgeht
- oder
- gröberes Gitter verwenden, interpolieren

erster Ansatz:

### ***Pixel by Pixel Contouring:***

geg.:  $f : (1, \dots, x_{max}) \times (1, \dots, y_{max}) \rightarrow \mathbf{R}$

Isowerte  $l_1, \dots, l_n$  und Isofarben  $c_1, \dots, c_n$

### **Algorithmus:**

```
for all  $(x,y) \in (1, \dots, x_{max}) \times (1, \dots, y_{max})$  do
  for all  $k \in \{1, \dots, n\}$  do
    if  $|f(x,y) - l_k| < \varepsilon$  then
      draw( $x, y, c_k$ )
```

**Problem:** bei großen Steigungen kann Isolinie trotz  $\varepsilon$  verfehlt werden.

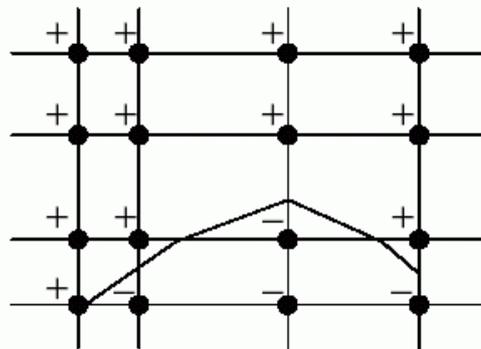
zweiter Ansatz:

### **Marching Squares:**

- Die skalare Funktion wird evtl. auf einem gröberem Gitter abgetastet, als dargestellt wird.
- Für gegebenen Isowert werden Schnitte der Isolinien mit den Gitterkanten berechnet.
- In jeder Gitterzelle werden die Schnittpunkte mit Isoliniensegmenten verbunden.
- Bei mehrdeutigen Fällen werden weitere Punkte im Innern der Zellen ausgewertet.

(Bartz 2005)

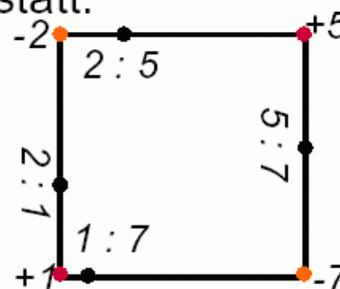
## Marching Squares Beispiel:



## Marching Squares Schnittberechnung:

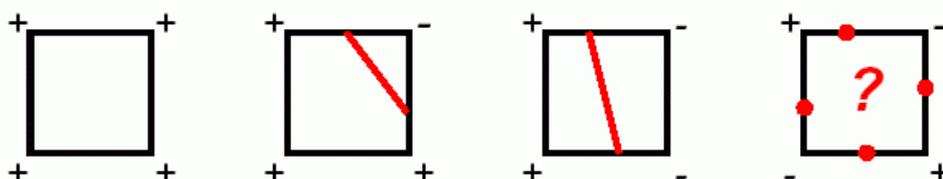
- Für gegebenen Isowert  $I$ , werden die Ecken der Zellen mit  $+$  oder  $-$  markiert, je nachdem, ob der Funktionswert größer oder kleiner  $I$  ist.
- Bei Kanten mit Ecken unterschiedlichen Vorzeichens, findet ein Schnitt statt.
- Die Schnittposition  $x$  auf der Kante  $[0, 1]$  geht aus linearer Interpolation hervor:

$$x = (I - f_0) / (f_1 - f_0)$$



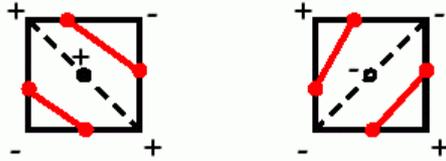
## Marching Squares Fallunterscheidung:

Unter Berücksichtigung von Symmetrien (Rotationen und Komplement) bleiben noch vier verschiedenen Vorzeichenzuordnungen:



## Auflösung der Mehrdeutigkeit:

- Im Falle von 4 Schnittpunkten gibt es zwei Möglichkeiten, die gültige Isolinien erzeugen:



- Wenn die Isolinien keine Feinstruktur in der Zelle haben, kann über den Funktionswerte im Zellenmittelpunkt zwischen den Fällen unterschieden werden.

## Marching Squares Algorithmen:

- **In Zellenreihenfolge:**

Durchlaufe alle Zellen und generiere Segmente → jeder Punkt 2 mal, höhere Interpolation schwierig

- **Linienverfolgung:**

Verfolge Isolinien entlang von Saatzellen bis Rand erreicht wird oder die Linie geschlossen wird

→ Hermite-Interpolation mit Hilfe von Gradienten aus zentralen Differenzen

**Problem:** Wie können Saatzellen schnell gefunden werden ?

(zur Interpolationstechnik und zu Gradienten später mehr!)

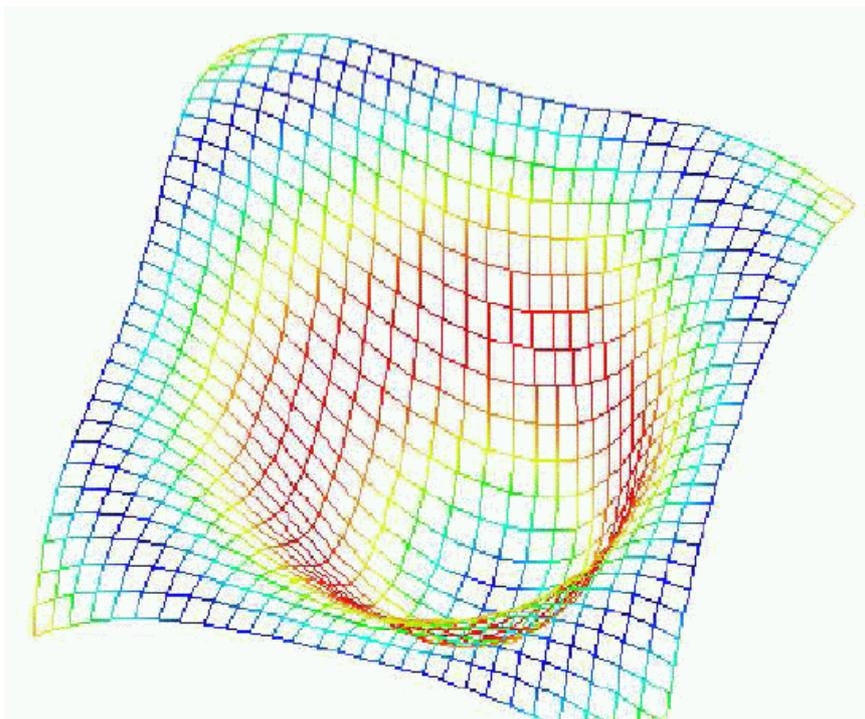
## Marching Squares vieler Isolinien:

- Komplexität des einfachen Marching Squares Algorithmus bei  $z$  Zellen und  $n$  Isolinien:  
 $O(z \cdot n)$
- Sei  $s \leq z \cdot n$  Zahl der Zellen-Isolinien Schnitte.
- Segmentbaum über Zellen:  
 $O(z \log z + s + \log^2 z)$ .
- Bereichsbaum über Isowerten:  
 $O(z \log n + n \log n + s)$

(Bartz 2005)

## Höhenfelder:

Das qualitative Merkmal wird über der Ebene nach oben aufgetragen, als beleuchtete Fläche oder mit Gitternetzlinien dargestellt.



(Weimar 2005)

## Darstellungstechniken / Farbe:

### Grundsatzfragen

- Welche Art von Daten sind farbkodierbar?
- Welche Art der Zusammenhänge lassen sich wie effizient identifizieren?
- Wie steht es mit Farbenblinden?
- Welche Farbbelegungen sind intuitiv oder zu vermeiden (Metaphern)?

### Einsatzmöglichkeiten von Farbe:

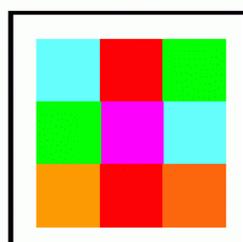
- Unterscheidung verschiedener Elemente
- Betonung von Elementen oder Regionen
- Nachbildung natürlicher Erscheinungsweisen
- Hervorrufen von symbolischen Assoziationen
- Ausdruck von Stimmung

### Grundsatzproblem:

- Um eine parallele bzw. wenig anstrengende Wahrnehmung zu ermöglichen sollten **nie mehr als sieben bzw. 5-8 Farben** gleichzeitig benutzt werden.
- Das ist die Grenze der **Wahrnehmungskapazität** vieler menschlicher Wahrnehmungskanäle.

### Aufgaben:

- Unterscheidbarkeit der Farben
- Lokalisierung einer Merkmalsausprägung



welche Farben soll man verwenden?

## Nach Wellenlängen:

Die folgenden Farben lassen sich unter fast allen Beleuchtungsverhältnissen unterscheiden:

	Farbe	Wellenlänge	Farbe	Wellenlänge	
	violett	430 nm	gelbgrün	556 nm	
	blau	476 nm	gelb	582 nm	
	grünblau	494 nm	orange	596 nm	
	blaugrün	504 nm	orangerot	610 nm	
	grün	515 nm	rot	642 nm	

## Basisfarben:

Wähle 5-8 Basisfarben aus. Ein Vorteil ist, dass diese mit gängigen Namen belegt sind.

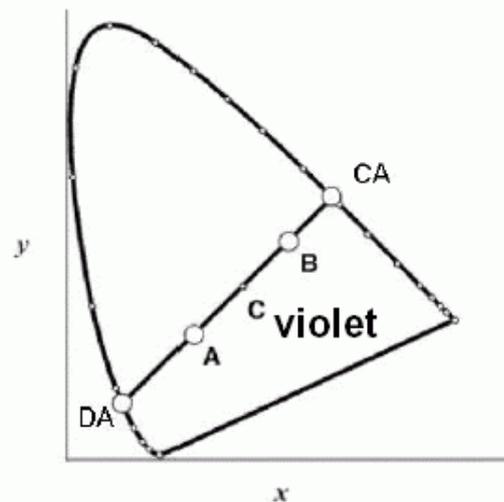
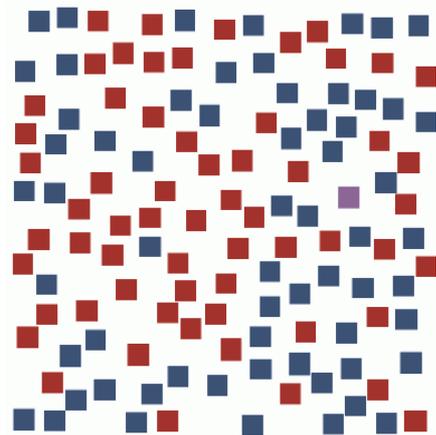
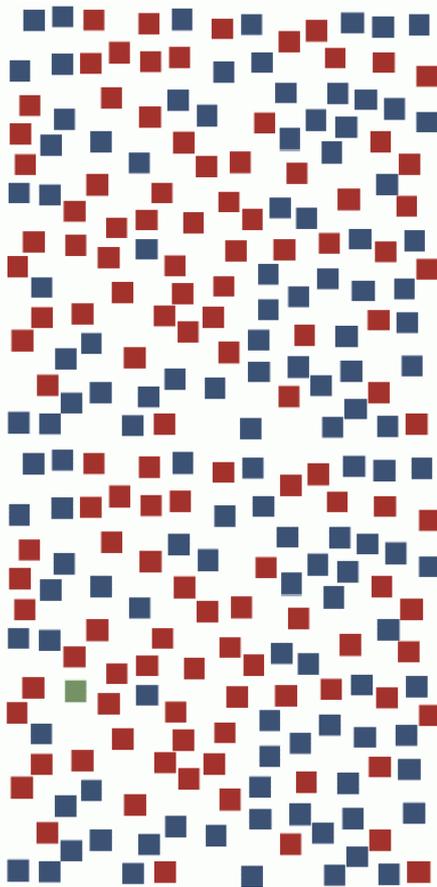
Farbe	CIE-x-Wert	CIE-y-Wert
rot	0,51	0,32
pink	0,35	0,30
lila	0,27	0,20
blau	0,19	0,25
grün	0,31	0,52
gelb	0,46	0,48
orange	0,54	0,40
braun	0,46	0,38

## Farbdifferenzen in CIE L\*u\*v\*:

- Grundsätzlich kann natürlich eine beliebige Menge von Farben gewählt werden, die den wahrgenommenen Farbabstand im L\*u\*v\*-Farbraum maximieren.
- Für eine schnelle Erkennbarkeit einzelner Farben sind jedoch zusätzlich wichtig:
  - Lineare Separation
  - Farbkategorien

## Lineare Separation:

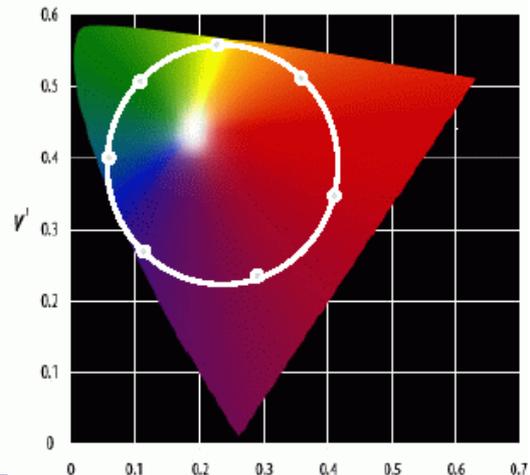
- **Experiment:** Testperson sucht unter  $n$  Klecksen, der Farbe  $f_1$  oder  $f_2$ , einen Zielklecks der Farbe  $f$ .
- **Ergebnis:** sind  $f$ ,  $f_1$  und  $f_2$  affin abhängig, so wird  $O(n)$  Zeit benötigt, andernfalls  $O(1)$ .
- **Interpretation:** wenn sich die Zielfarbe durch eine Gerade im Farbraum von den nicht-Zielfarben trennen lässt, kann ein Zielfleck in den Vorverarbeitungsstufe des Auges erkannt werden  
→ wähle Farben möglichst auf Kreis





## Kombination von Farbabstand, linearer Separation und Kategorie:

- Wähle Farben auf Kreis, so dass sie in möglichst unterschiedliche Kategorien oder auf Kategoriengrenzen fallen.



außerdem ähnliche / leicht abgestufte Helligkeiten der verwendeten Farben beachten (Abgleich mit Referenzgrau).

## Minimale Größe des Farbzeiges:

Art der Information	vergleichende Farbunterscheidung (2-7 Farben)	absolute Farbidentifizierung (2-4 Farben)	Helligkeit des Symbols
kritische Information, variable Position	> 20'	> 20'	>60'
kritische Information, feste Position	> 16'	> 20'	>60'
unkritische Information	> 12'	> 16'	>60'

## Abbildung der Ordnung:

Die Merkmalsausprägungen  $x_i$  werden auf Farben  $c_i$  abgebildet, so dass die empfundenen Farbeindrücke  $E(c_i)$  die Ordnung repräsentiert:

$$x_1 < x_2 < \dots < x_n \rightarrow E(c_1) < E(c_2) < \dots < E(c_n)$$

Bei quantitativen Merkmalen sollte auch gelten:

$$\Delta E(c_i, c_j) \approx k \cdot \Delta(x_i, x_j)$$

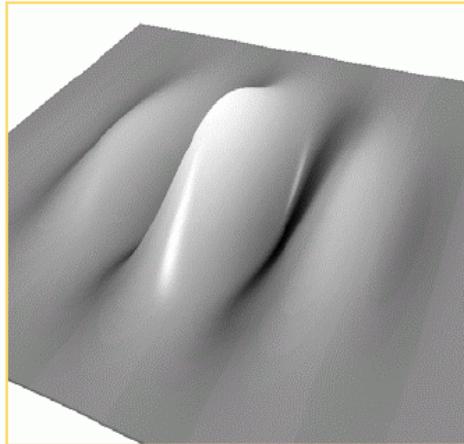
gebräuchliche Farbskalen (vgl. Kapitel 3):

### Grauwertskala:



- Benutze empfindungsbezogenen Helligkeit  $L^*$
- Ca. 150 wahrnehmbare Helligkeitsstufen
- Bei Schaukasten für Röntgenbilder mehrere Adaptionstufen (Lampen) → bessere Abstufung
- Durchgängige intuitive Ordnung
- Oft bildgebende Verfahren schwarz weiß
- Keine Störfaktoren wie Täuschungseffekte oder Fehlsichtigkeiten

**Grauwertskala:**

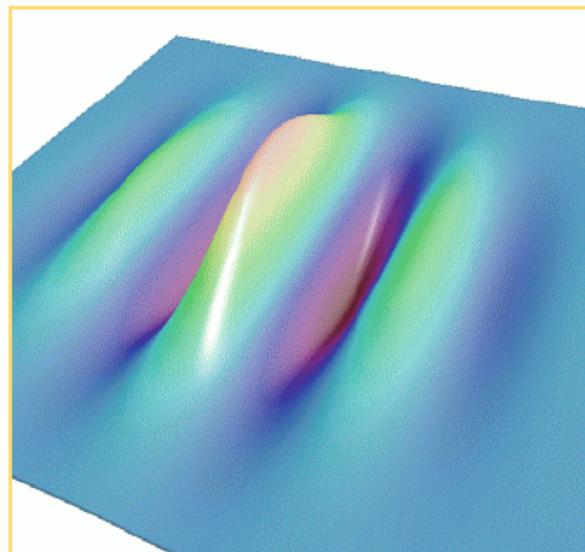
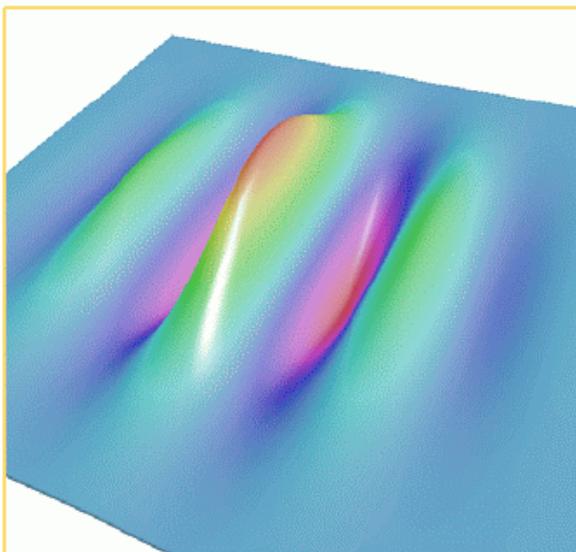


**Regenbogenskala:**



- Benutze empfindungsbezogene Skalierung und konstante Sättigung
- Nur für geübte Benutzer intuitiv, da manche Farben einen geringeren Abstand suggerieren
- Kreisförmige Ordnung → benutze Ausschnitt
- Kombiniere mit Helligkeitsvariation
- Angabe einer Legende

**Regenbogenskala:**

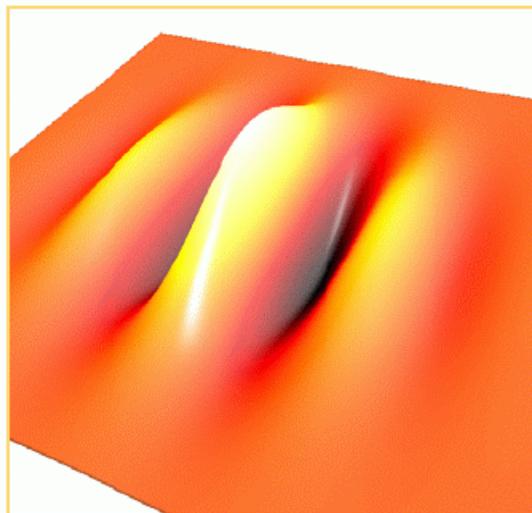


### Temperaturskala:



- Drittle Farbskala und interpoliere von schwarz über rot, gelb nach weiß → Variation von Farbton und Helligkeit
- Intuition: entspricht schwarzem Objekt mit wachsender Temperatur
- Bessere Ordnungsempfindung

### Temperaturskala:



### Magentaskala:



- Drittle Farbskala und interpoliere von Schwarz über Rot, Magenta nach Weiß → Variation von Farbton und Helligkeit
- Nutzt bessere empfindungsbezogene Farbunterschiede im Magenta-Farbbereich aus
- Gute Ordnungsempfindung

## Optimale Farbskalen [Levkowitz, CG&A 1992]:

Maximiere Farbabstände bei intuitiver Ordnung

- Darstellung im RGB-Raum  $c_i = (r_i, g_i, b_i)$
- $c_1 = (0,0,0)$  und  $c_n = (1,1,1)$
- $r_i \leq r_{i+1}$  und  $g_i \leq g_{i+1}$  und  $b_i \leq b_{i+1}$
- vollständig chromatisch oder achromatisch
- Sättigung ist monoton steigend oder fallend

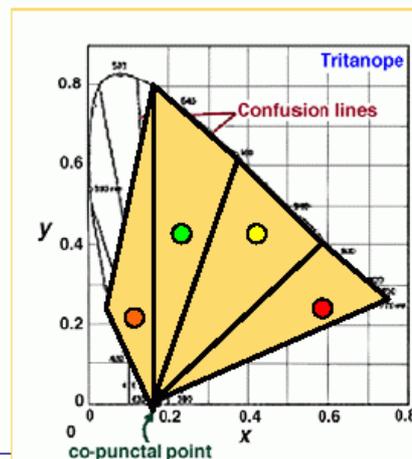
Experimente zeigen leider **keine Überlegenheit** gegenüber Temperaturskala, aber **Unterlegenheit** gegenüber Grauwertskala!

## Bivariate und Trivariate Farbskalen:

- Sind nur sehr eingeschränkt zu empfehlen.
- Meist gibt es keine intuitive Ordnung,
- oder man muss sich auf wenige Ausprägungen beschränken,
- oder ist die Zahl der Ausprägungen in einer Dimension von der Ausprägung in der (den) Dimension(en) abhängig (vgl. Munsell Farbraum).

## Berücksichtigung von Farbfehlsichtigkeit:

- Es können zwar Farbfehlsichtigkeiten simuliert werden. Empfindungsqualitative Aussagen sind jedoch kritisch.
- Zu empfehlen sind grundsätzlich Farbskalen, die orthogonal zu Verwechslungsgeraden liegen oder die zusätzliche Kodierung in der Helligkeit.



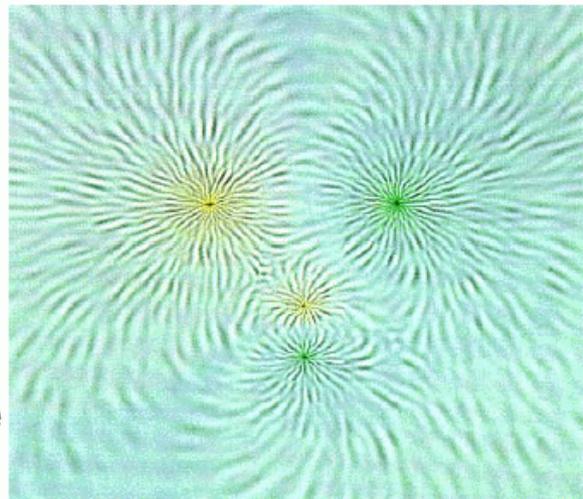
(Bartz 2005)

## *Darstellung von Merkmalen mittels Texturen:*

- nach Möglichkeit nur die 3 "primären Texturkanäle" verwenden:
  - Orientierung,
  - Größe,
  - Kontrast.
- Abbildungen von Merkmalen auf andere Kanäle, wie: Regularität, Grobheit, Rauigkeit usw., sind nicht gut untersucht bzw. ergeben keine guten Erkennungsergebnisse.

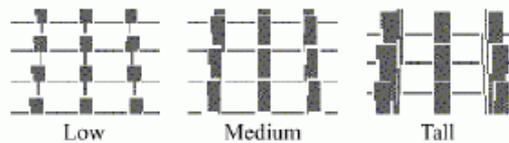
### Beispiel:

- Shows the magnetic field generated by two dipoles.
- Field orientation is mapped to texture orientation, and field strength is mapped to texture contrast and inversely to texture size.
- Texture is displayed using the luminance dimension, while field potential is displayed by a color sequence.

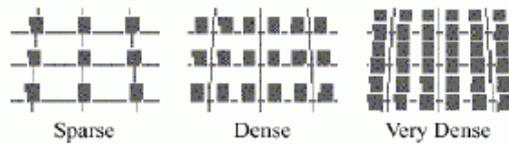


## Healey: Drei Texturattribute + Farbe

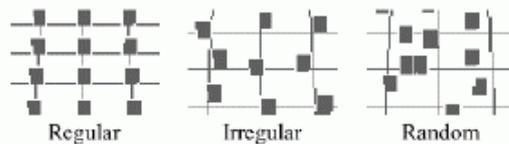
Größe/Höhe



Dichte



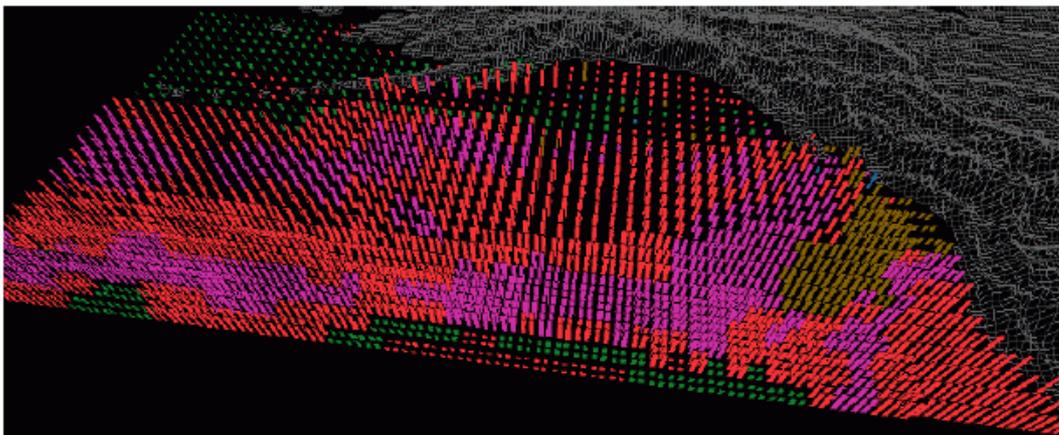
Regularität



Basierend auf Untersuchungen der  
Wahrnehmungsgeschwindigkeit

---

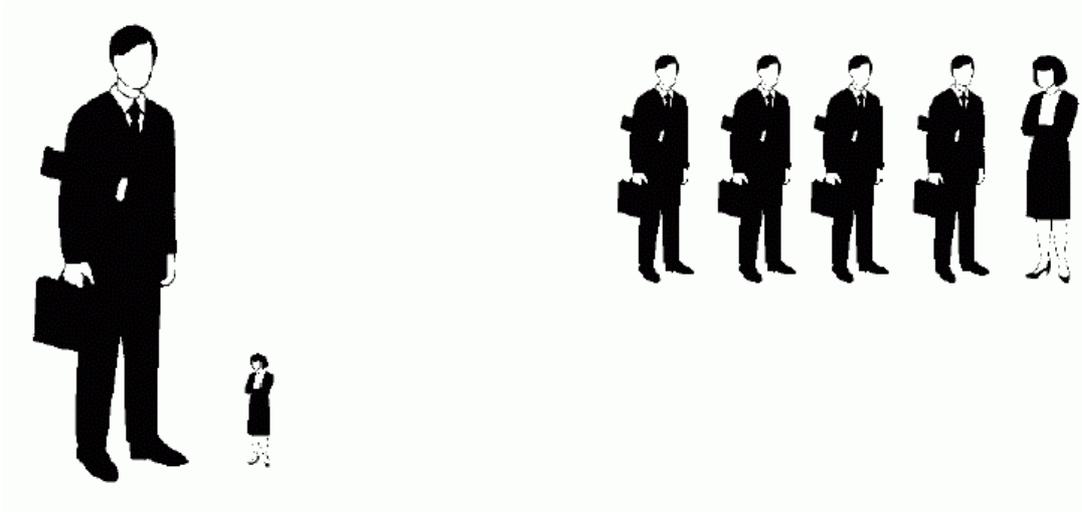
Beispiel, wo Textur mit eingesetzt wurde:



Visualization of the oceanography datasets, **color** used to represent **plankton density** (blue, green, brown, red, and purple represent lowest to highest densities), **height** used to represent **current strength**, **texture density** used to represent **Sea Surface Temperature**: June, 1956

(Bartz 2005)

## Piktogramme



(aus Weimar 2005)

### Glyphen = Piktogramme (Icons)

- Exakt positionierbare Primitive
- Werte von Merkmalen werden auf **Länge, Winkel, Form, Farbe oder Transparenz** abbilden.

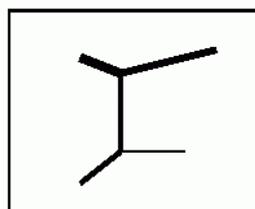
#### Regeln für die Erzeugung:

- einzelne Merkmale sollten gut kombinier- und unterscheidbar sein
- Ikonen sollten separat erkennbar sein
- Differenzen in der Merkmalsausprägung sollten sich in den Ikonen widerspiegeln

#### Stick-Figure-Piktogramme

[Picket & Grinstein, IEEE SMC 1988]:

- 2d vierarmige Figur,
- Kodierung der Merkmale in Armlänge,
- Armbreite und Armwinkel zur Vertikalen  
→ 12 Merkmale

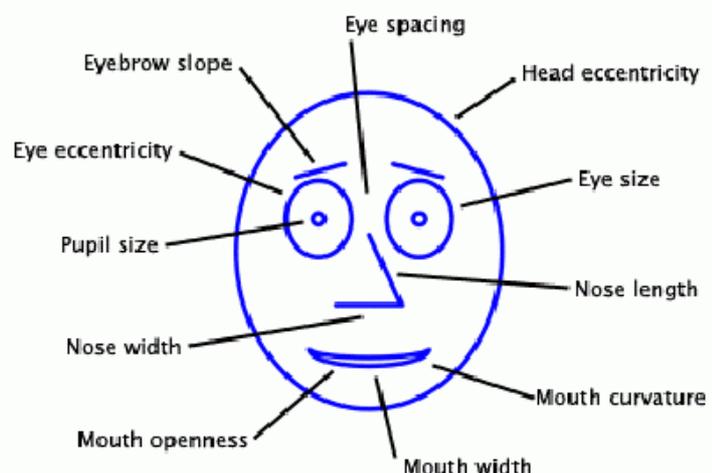


## Farbpiktogramme [Levkowitz, Vis 1991]:

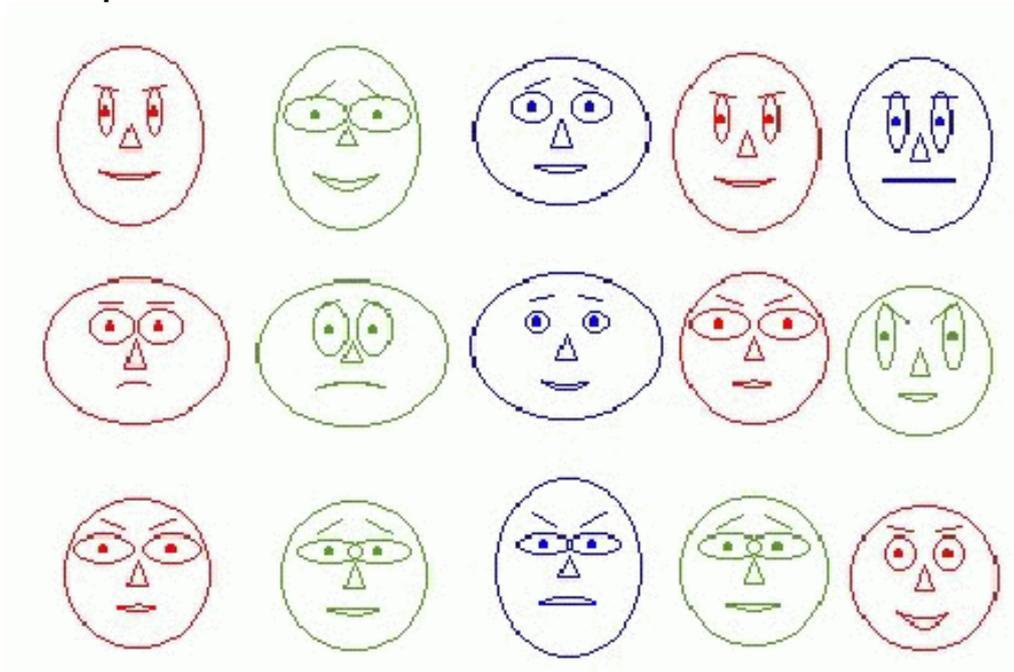
- Unterteilung einer Grundfigur (Dreieck, Quadrat, Sechseck,...) in ein Kanten- und Flächennetz.
- Merkmale werden Kanten oder Flächen zugeordnet und auf Farbskalen abgebildet
- Bei Abbildung auf Kanten baryzentrische oder bilineare Interpolation auf den Flächen
- Möglichkeit zur Gruppierung durch Hervorhebung von Kanten und oder Flächen

## Chernoff Piktogramme/Gesichter [Chernoff 73]:

- Kodierung von 12 Merkmalen in über Form und Größe von Kopf, Nase, Mund und Augen.
- Ausnutzung der menschlichen Fähigkeit Nuancen in Gesichtern zu unterscheiden.
- Häufige Anwendung der sogenannten Chernoff-Gesichter in der Statistik.
- Kodierung von 12 Merkmalen in über Form und Größe von Kopf, Nase, Mund und Augen.



Beispiel:



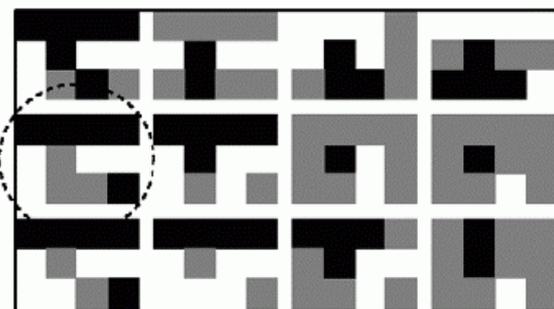
## Shape Coding [Beddow, Vis 1990], cont'd :

Spezialfall der Farbpiktogramme

- Basisform ist Rechteck
- Flächennetz ist regelmäßiges Gitter
- Abbildung auf Flächen

Beispiel:

1	2	3	4
5	6	7	8
9	10	11	12

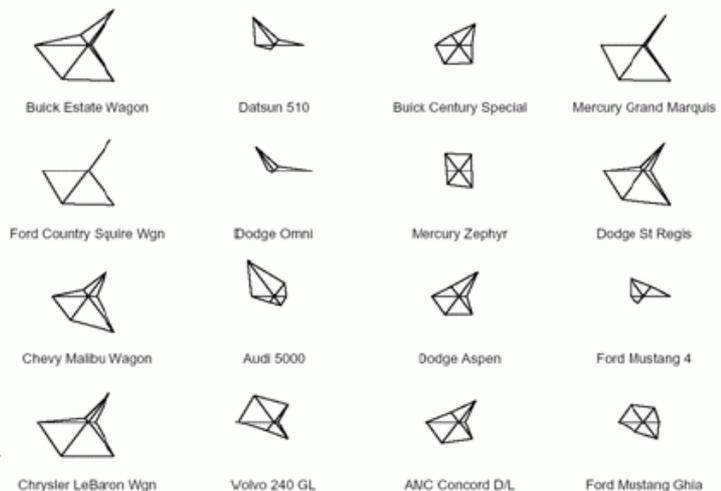


## weitere Piktogramm-Typen

- Kreisförmige Glyphen/Piktogramme:
  - Star Glyphs
  - Sun Ray Plots/Glyphs
  - ....
- Ähnlich wie Radspeichen
- Werte werden als Distanzen vom Drehpunkt repräsentiert

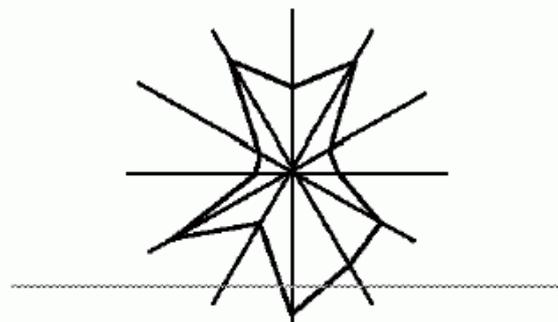
### Star Glyphs [Fienberg, Am. Stat., 1979]:

- Stern setzt sich aus gleichmäßigen verteilten Zacken vom Zentrum zusammen
- Erstes Attribut startet rechts
- Dann im Gegen-  
uhrzeigersinn
- Enden der Zacken werden verbunden



### Sun Plot:

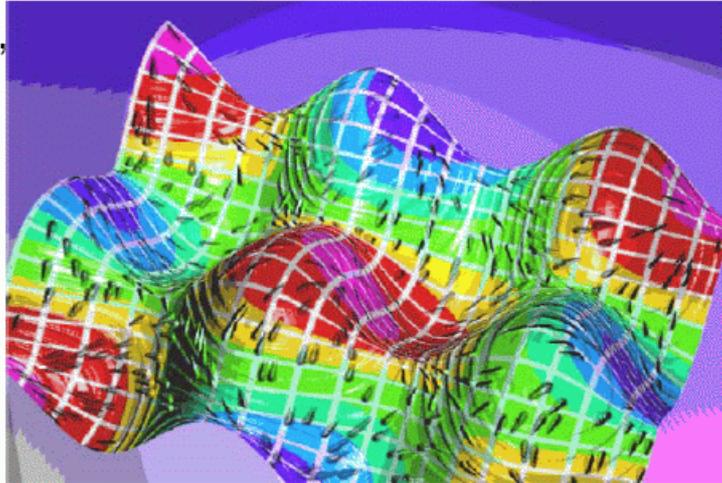
- Variation des Star glyph
- Feste Stern-Struktur als Basis



## Kombination verschiedener Darstellungsmöglichkeiten

### Multiple Attribute:

- Geometrie, Höhenfeld
- Farbe: Sättigung, Ton, Intensität
- Texture/Glyphen



(Bartz 2005)

### 4.3 Animation

Hinzunahme von Bewegung als weiteres Wahrnehmungsprimitiv

#### **Animation:**

(1) die Gesamtheit der Methoden zur Erzeugung synthetischer *Bewegtbilder*;

(2) ein vermittels (1) erzeugtes Produkt auf Bewegtbildmedien. "Zeitvariante" der Computergrafik.

Ursprünge: "animus" = Lufthauch, Atem; "animos" = beleben;  $\alpha\nu\epsilon\mu\omicron\sigma$  = Wind, Atem.

2 grundsätzliche Erzeugungstypen:

- Keyframe-Animation (Erzeugung einer Serie von Standbildern (*Frames*), Interpolation zwischen diesen)
- parametrische Animation (1 Parameter / Merkmal wird auf die Zeit abgebildet, alle Frames werden automatisch berechnet)

Variante: *interaktive Animation*  
(Erzeugung der Animation direkt im  
Visualisierungssystem);

bei nicht-interaktiver Animation kann die Bildsequenz vor  
der Präsentation berechnet werden (darf dann auch  
länger dauern)

was soll animiert werden?

- Zeit
- andere Parameter
- Objektposition, Kameraposition (z.B. Rundumfahrt um  
ein Objekt), Skalierung (z.B. Zoom-in)

je nach Anwendung: Drehbuch nötig?

Datenformate: mpeg, avi, wmv, mng, gif, ...

Tradeoff zwischen Größe und Geschwindigkeit  
(Einschränkungen insbes. bei Übertragung übers  
Internet, mobilen Geräten, Speicherplatz-Knappheit)

Beschränkung durch das langsamste Glied der Kette:  
Grafik, Bus, Festplatte, Netz, DVD/CD-ROM

MPEG-Standards: Framegrößen 160\*120, 320\*240,  
640\*480 Pixel

Geschwindigkeit min. 5 Bilder/s, max. 30 Bilder/s.