

8
AKADEMISCHE
MITTAGSPAUSE 2016

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386



SPRECHEN SIE

8
MATH & MATIK?

Ein Bild sagt mehr als tausend Worte

Graphische Darstellungen komplexer Daten

Bastian Rieck

Warum graphische Darstellungen?

Die Ausgangssituation

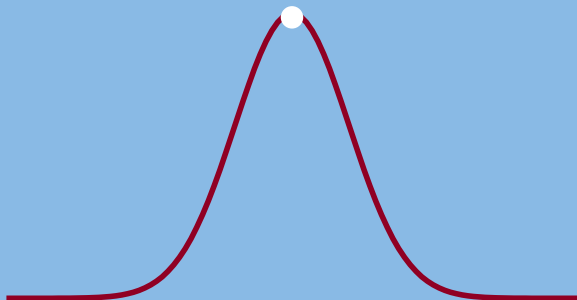
x	y
10.0	8.04
8.0	6.95
13.0	7.58
9.0	8.81
11.0	8.33
14.0	9.96
6.0	7.24
4.0	4.26
12.0	10.84
7.0	4.82
5.0	5.68

I

Was tun?



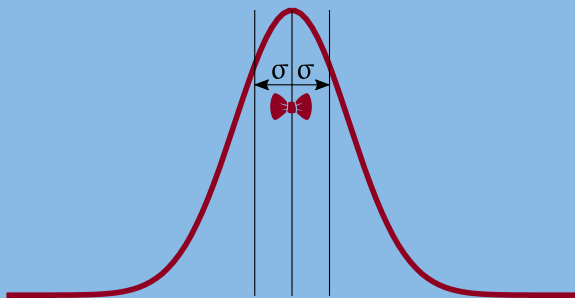
Arithmetisches Mittel



$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Hier: $\bar{x} = 9, \bar{y} = 7.50$

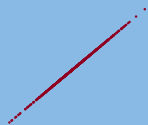
Standardabweichung



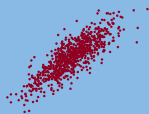
$$\sigma_x^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2)$$

Hier: $\sigma_x^2 = 11$, $\sigma_y^2 = 4.12$

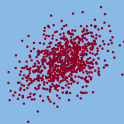
Korrelationskoeffizient



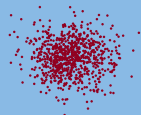
$$\text{cor}(x, y) = 1$$



$$\text{cor}(x, y) = 0.8$$



$$\text{cor}(x, y) = 0.4$$

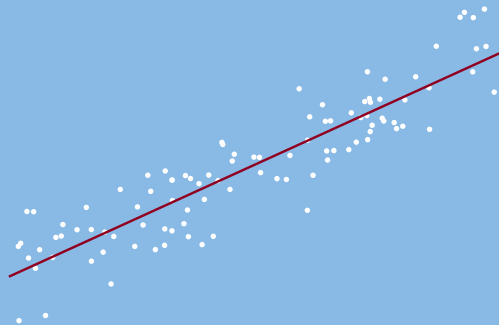


$$\text{cor}(x, y) = 0$$

$$\text{cor}(x, y) := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

Hier: $\text{cor}(x, y) = 0.816$

Regressionsgerade



$y = ax + b$, mit...

$$a = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (4)$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (5)$$

Hier: $y = 3.00 + 0.500x$

Zusammengefasst

	x	y
Arithmetisches Mittel	9	7.50
Stichprobenvarianz	11	4.12
Korrelationskoeffizient		0.816
Regressionsgerade		$y = 3.00 + 0.500x$

Noch mehr Daten!

x	y
10.0	9.14
8.0	8.14
13.0	8.74
9.0	8.77
11.0	9.26
14.0	8.10
6.0	6.13
4.0	3.10
12.0	9.13
7.0	7.26
5.0	4.74

II

x	y
10.0	7.46
8.0	6.77
13.0	12.74
9.0	7.11
11.0	7.81
14.0	8.84
6.0	6.08
4.0	5.39
12.0	8.15
7.0	6.42
5.0	5.73

III

x	y
8.0	6.58
8.0	5.76
8.0	7.71
8.0	8.84
8.0	8.47
8.0	7.04
8.0	5.25
19.0	12.50
8.0	5.56
8.0	7.91
8.0	6.89

IV

Noch mehr Ergebnisse!

x	y
9	7.50
11	4.12
	0.816
$y = 3.00 + 0.500x$	

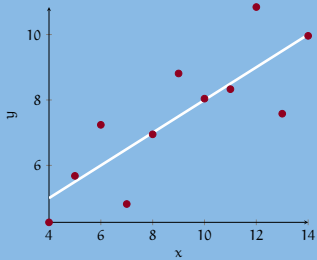
x	y
9	7.50
11	4.12
	0.816
$y = 3.00 + 0.500x$	

x	y
9	7.50
11	4.12
	0.816
$y = 3.00 + 0.500x$	

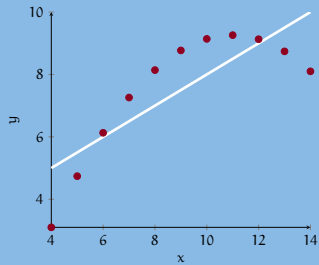
Was nun?



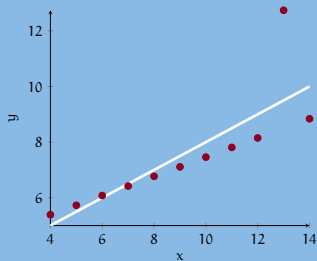
René Descartes



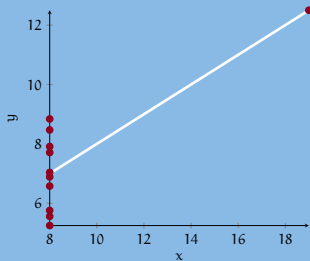
I



II



III



IV

Was ist mit echten Daten?

Echte Daten

„USDA National Nutrient Database for Standard Reference“

Datenbank von ca. 8800 unterschiedlichen Nahrungsmitteln mit knapp 700000 Nährstoffangaben.

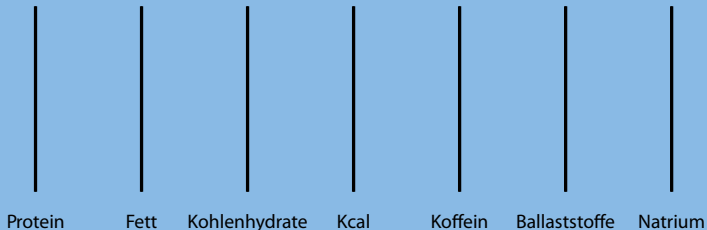
- 1 Protein (g)
- 2 Fett (g)
- 3 Kohlenhydrate (g)
- 4 Kcal
- 5 Koffein (mg)
- 6 Ballaststoffe (g)
- 7 Natrium (mg)

Name	1	2	3	4	5	6	7
Butter, gesalzen	0.85	81.11	0.06	717	0.0	0.0	643
Käse, Feta	14.21	21.28	4.09	264	0.0	0.0	917
Froschschenkel, roh	16.40	0.30	0.00	73	0.0	0.0	58

Wie darstellen?

Parallele Koordinaten

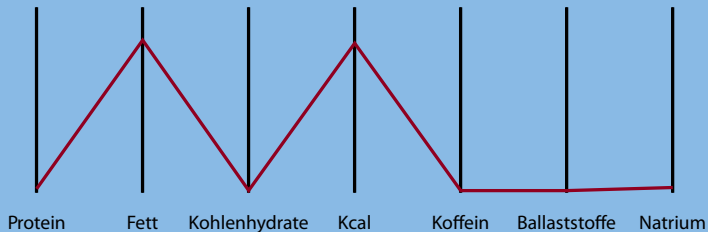
Name	1	2	3	4	5	6	7
Butter, gesalzen	0.85	81.11	0.06	717	0.0	0.0	643
Käse, Feta	14.21	21.28	4.09	264	0.0	0.0	917
Froschschenkel, roh	16.40	0.30	0.00	73	0.0	0.0	58



Wie darstellen?

Parallele Koordinaten

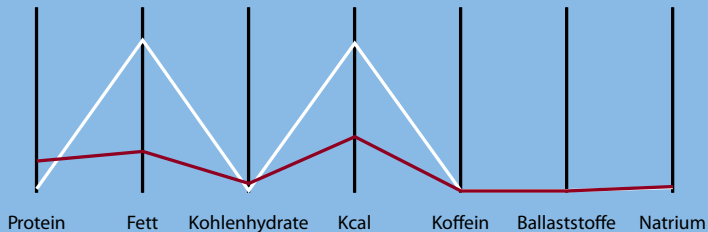
Name	1	2	3	4	5	6	7
Butter, gesalzen	0.85	81.11	0.06	717	0.0	0.0	643
Käse, Feta	14.21	21.28	4.09	264	0.0	0.0	917
Froschschenkel, roh	16.40	0.30	0.00	73	0.0	0.0	58



Wie darstellen?

Parallele Koordinaten

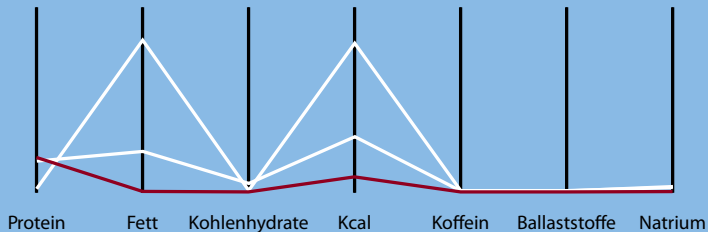
Name	1	2	3	4	5	6	7
Butter, gesalzen	0.85	81.11	0.06	717	0.0	0.0	643
Käse, Feta	14.21	21.28	4.09	264	0.0	0.0	917
Froschschenkel, roh	16.40	0.30	0.00	73	0.0	0.0	58



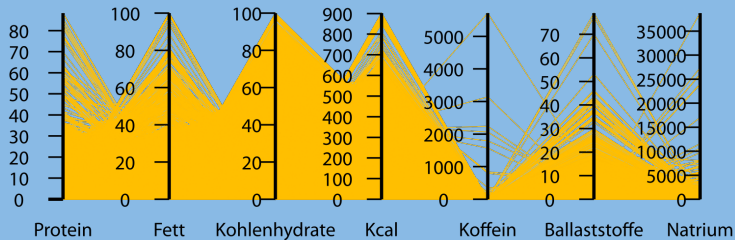
Wie darstellen?

Parallele Koordinaten

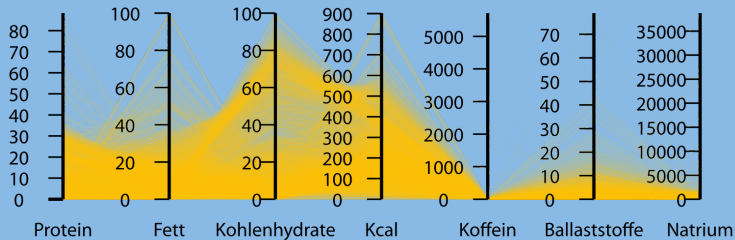
Name	1	2	3	4	5	6	7
Butter, gesalzen	0.85	81.11	0.06	717	0.0	0.0	643
Käse, Feta	14.21	21.28	4.09	264	0.0	0.0	917
Froschschenkel, roh	16.40	0.30	0.00	73	0.0	0.0	58



Parallele Koordinaten

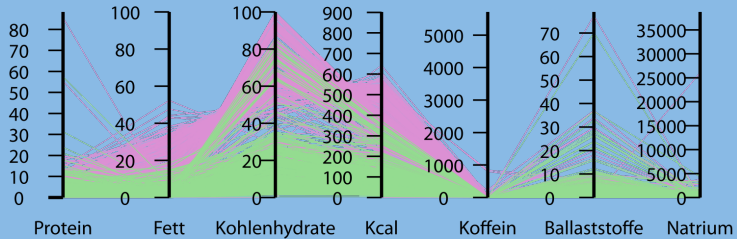


Parallele Koordinaten

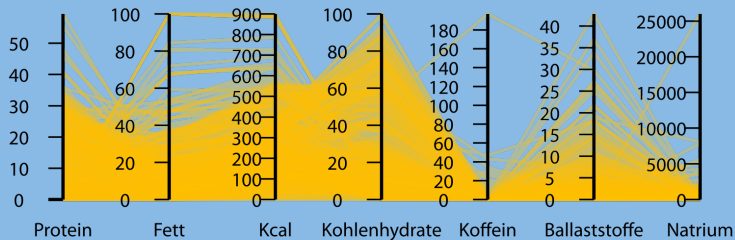


Welche Fragen können wir damit beantworten?

Wie unterscheiden sich Süßigkeiten und Gemüse?



Wie hängen Fett, Kohlenhydrate und Kcal zusammen?



Visualisierung hilft...

- ...komplexe Daten besser zu erfassen
- ...Zusammenhänge zu finden
- ...Fehler schneller zu erkennen

Mehr unter:

