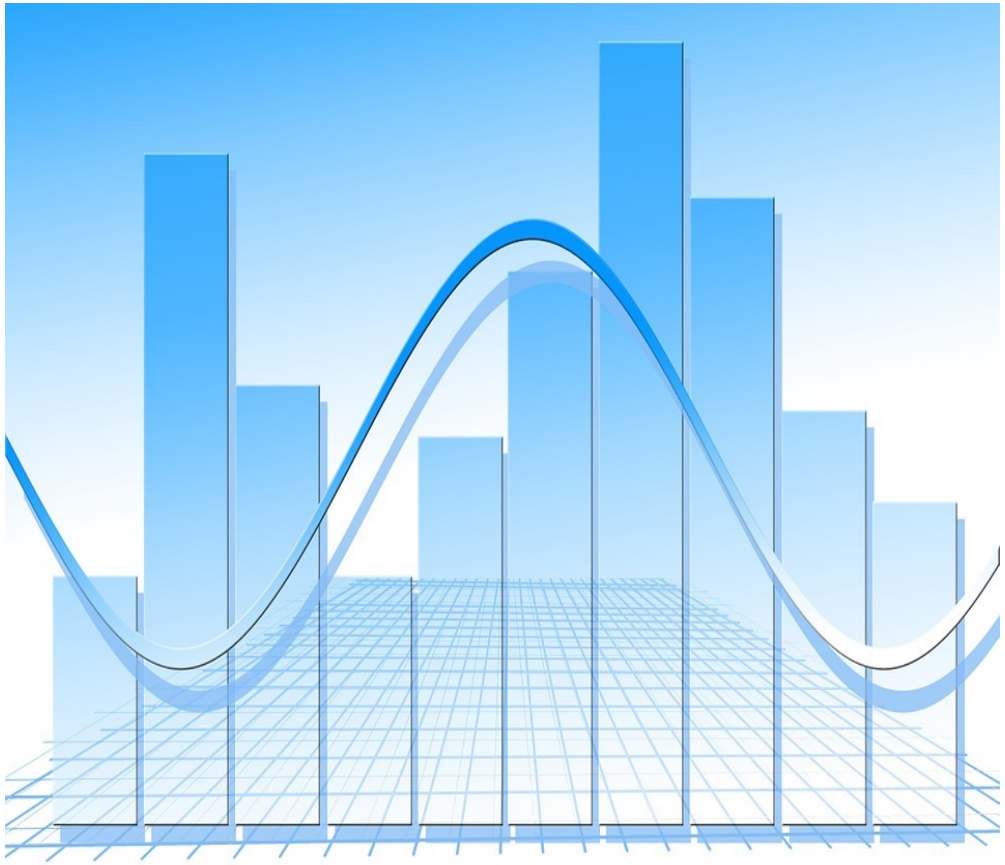
III. Analysis

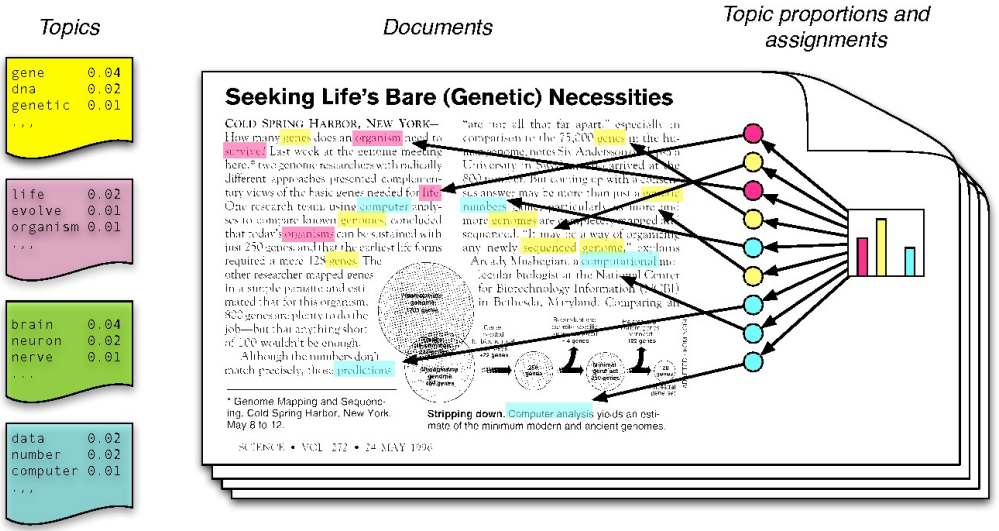
9. Quantification Methods

Prof. Dr. Martin Langner



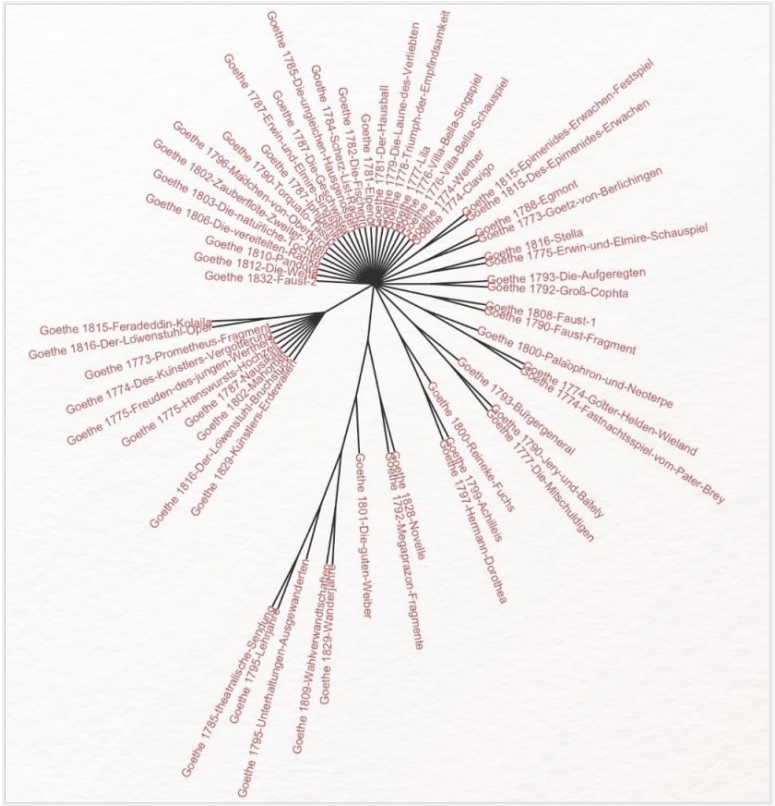
Schreibman / Siemens / Unsworth (2004) Kap. 28;
Jannidis / Kohle / Rehbein (2017) Kap. 10. 20

WORD STATISTICS



Topic Modelling:
http://www.cs.columbia.edu/~blei/papers/Blei2011.pdf

Jannidis / Kohle / Rehbein (2017) Kap. 20

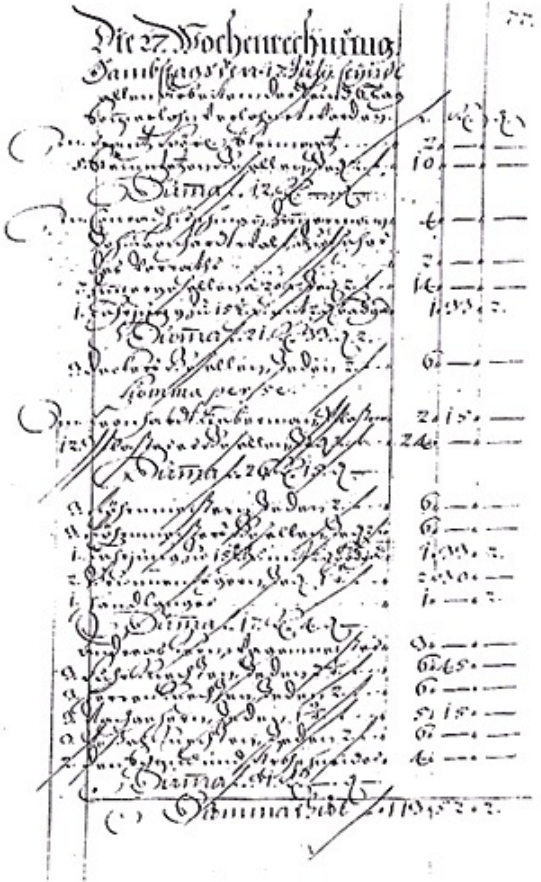


Bootstrap Consensus Tree of the Dramatic and Epic Oeuvre of Johann Wolfgang von Goethe
https://fortext.net/routinen/methoden/stilometrie

DATA COLLECTION AND ANALYSIS

https://www.flickr.com/photos/64379474@N00/3667281359

Accounting of the materials used for the Athena Parthenos by the overseers for the construction of the Parthenon, 440/439 BC



Extract of the weekly wages from the annual account of the master builder's office of 17 July 1700

Rainer Gömmel, Quellen zur Lohn- und Preisstatistik der Stadt Nürnberg vom 16. bis zum 18. Jahrhundert, in: Wolfram Fischer, Andreas Kunz (Hrsg.), Grundlagen der Historischen Statistik von Deutschland (Wiesbaden: Springer: 1991) 43–46

BEGINNINGS OF OFFICIAL AND SCIENTIFIC STATISTICS

- Königlich Preußisches Statistisches Bureau (1805)

- Kaiserliches Statistisches Amt (1872)
- Statistisches Reichsamt (1918)
- Statistische Amt des Vereinigten Wirtschaftsgebietes (1948)
- Statistisches Bundesamt (1950)

Sealing mark of the Königlich Preußisches StatistischensBureau Berlin

https://i.ebayimg.com/images/g/jMoAAOSwg2Rfetrd/s-l1600.jpg

"It will not occur to anyone with understanding to deny that the statistical approach to human affairs also has its great value; but one must not forget what it can and wants to achieve."

Johann Gustav Droysen 1863

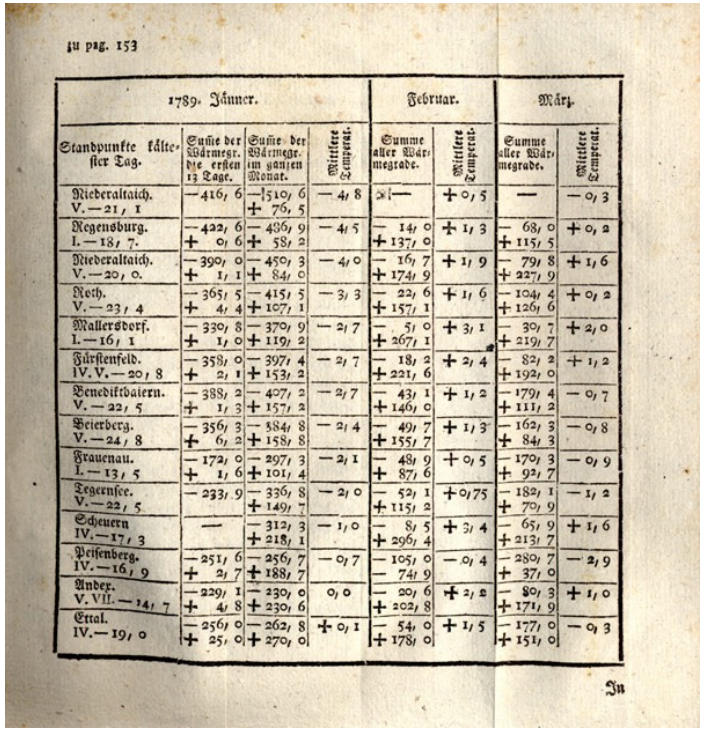
Johann Gustav Droysen, Die Erhebung der Geschichte zum Rang einer Wissenschaft, in: Historische Zeitschrift 9 (1863), 1–22, bes. 14.

Henry Thomas Buckle, History of Civilization in England (London: J. W. Parker & Son, 1857/61)

STATISTICS AS A HISTORICAL SOURCE

Statistics are "a specific knowledge practice [...] that orders and categorises phenomena in order to enable comparisons as well as inclusion and exclusion". They are thus "components of complex decision-making and communication processes“

Grundbegriffe und Konturen einer Kulturgeschichte der Statistik, in: Stefan Haas, Michael C. Schneider, Nicolas Bilo (Hrsg.), Die Zählung der Welt. Kulturgeschichte der Statistik vom 18. bis 20. Jahrhundert, Studien zur Alltags- und Kulturgeschichte 32 (Stuttgart: Steiner, 2019), 11



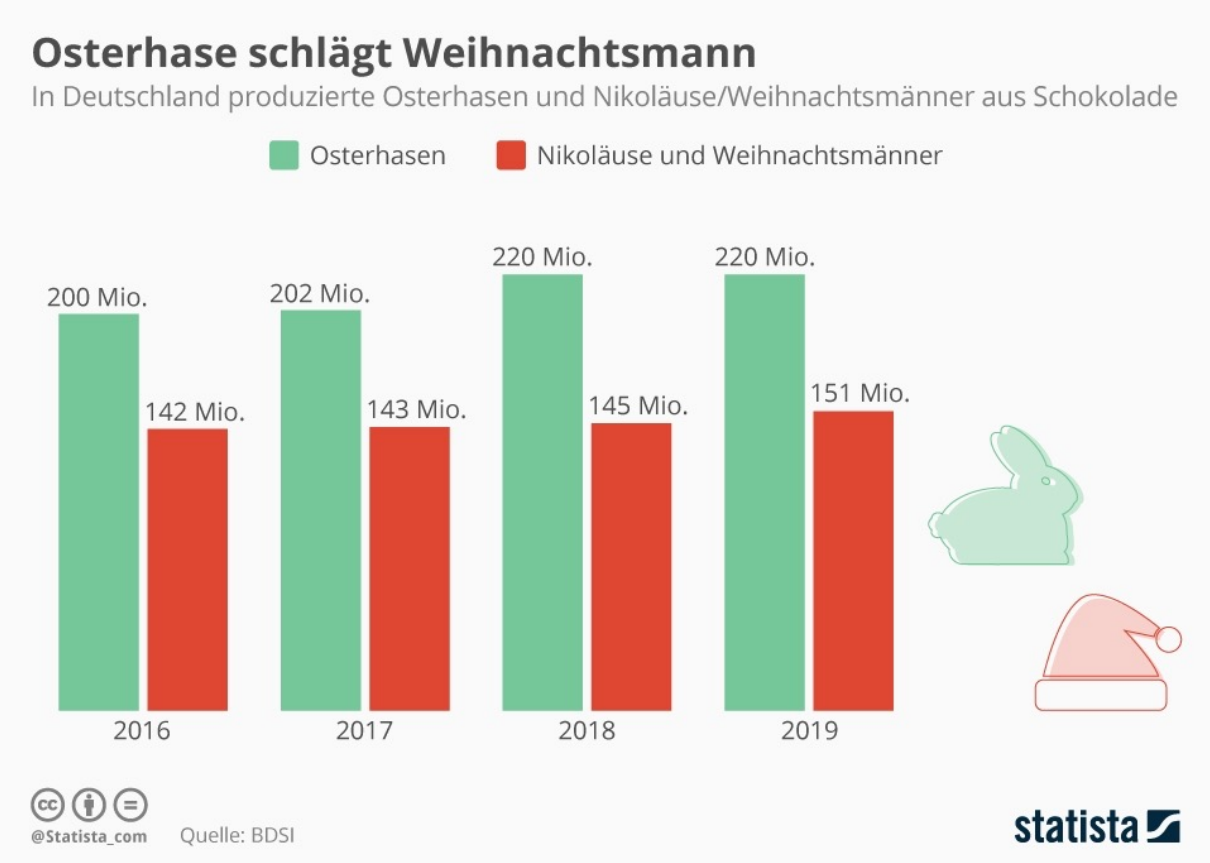
zu pag. 153

	1789. Jänner.			Februar.		März.	
Standpunkte kälteſter Tag.	Summe der Wärmegr. die erſten 15 Tage.	Summe der Wärmegr. im ganzen Monat.	Mittlere Temperat.	Summe aller Wärmegrade.	Mittlere Temperat.	Summe aller Wärmegrade.	Mittlere Temperat.
Niederaltaich. V. — 21, 1	— 416, 6	— 1510, 6 + 76, 5	— 4, 8	[illegible]	+ 0, 5	—	— 0, 3
Regensburg. I. — 18, 7.	— 422, 6 + 0, 6	— 436, 9 + 58, 2	— 4, 5	— 14, 0 + 137, 0	+ 1, 3	— 68, 0 + 115, 5	+ 0, 2
Niederaltaich. V. — 20, 0.	— 390, 0 + 1, 1	— 450, 3 + 84, 0	— 4, 0	— 16, 7 + 174, 9	+ 1, 9	— 79, 8 + 227, 9	+ 1, 6
Roth. V. — 23, 4	— 365, 5 + 4, 4	— 415, 5 + 107, 1	— 3, 3	— 22, 6 + 157, 1	+ 1, 6	— 104, 4 + 126, 6	+ 0, 2
Mallersdorf. I. — 16, 1	— 330, 8 + 1, 0	— 370, 9 + 119, 2	— 2, 7	— 5, 0 + 267, 1	+ 3, 1	— 30, 7 + 219, 7	+ 2, 0
Fürſtenfeld. IV. V. — 20, 8	— 358, 0 + 2, 1	— 397, 4 + 153, 2	— 2, 7	— 18, 2 + 221, 6	+ 2, 4	— 82, 2 + 192, 0	+ 1, 2
Benediktbaiern. V. — 22, 5	— 388, 2 + 1, 3	— 407, 2 + 157, 2	— 2, 7	— 43, 1 + 146, 0	+ 1, 2	— 179, 4 + 111, 2	— 0, 7
Beierberg. V. — 24, 8	— 356, 3 + 6, 2	— 384, 8 + 158, 8	— 2, 4	— 49, 7 + 155, 7	+ 1, 3	— 162, 3 + 84, 3	— 0, 8
Frauenau. I. — 13, 5	— 172, 0 + 1, 6	— 297, 3 + 101, 4	— 2, 1	— 48, 9 + 87, 6	+ 0, 5	— 170, 3 + 92, 7	— 0, 9
Tegernſee. V. — 22, 5	— 233, 9	— 336, 8 + 149, 7	— 2, 0	— 52, 1 + 115, 2	+ 0, 75	— 182, 1 + 70, 9	— 1, 2
Scheuern IV. — 17, 3	—	— 312, 3 + 218, 1	— 1, 0	— 8, 5 + 296, 4	+ 3, 4	— 65, 9 + 213, 7	+ 1, 6
Peiſenberg. IV. — 16, 9	— 251, 6 + 2, 7	— 256, 7 + 188, 7	— 0, 7	— 105, 0 — 74, 9	— 0, 4	— 280, 7 + 37, 0	— 2, 9
Andex. V. VII. — 14, 7	— 229, 1 + 4, 8	— 230, 0 + 230, 6	0, 0	— 20, 6 + 202, 8	+ 2, 2	— 80, 3 + 171, 9	+ 1, 0
Ettal. IV. — 19, 0	— 256, 0 + 25, 0	— 262, 8 + 270, 0	+ 0, 1	— 54, 0 + 178, 0	+ 1, 5	— 177, 0 + 151, 0	— 0, 3

In

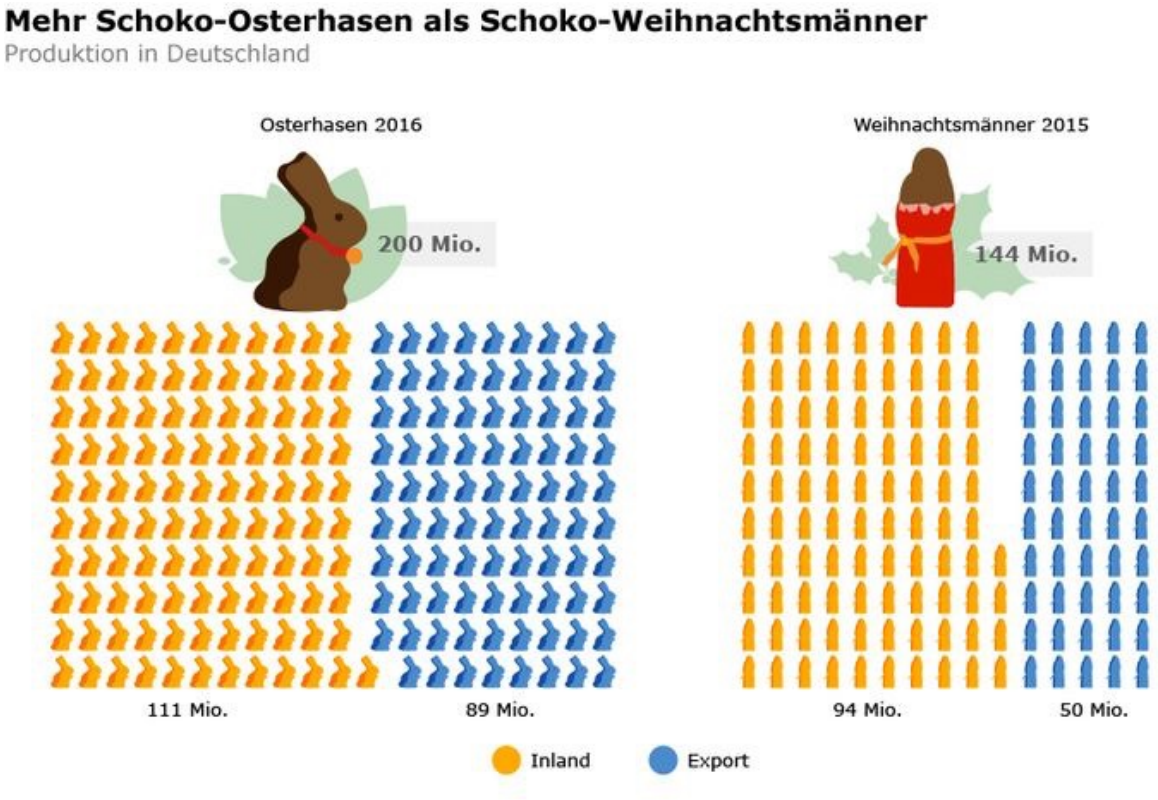
Klimadaten Bayerns 1789

https://www.historisches-lexikon-bayerns.de/Lexikon/Amtliche_Statistik



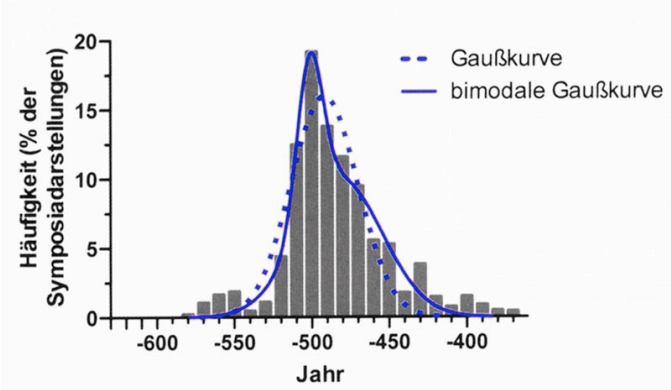
https://cdn.statcdn.com/Infographic/images/normal/7089.jpeg

VERKEHRUNG VON THESE UND BELEG

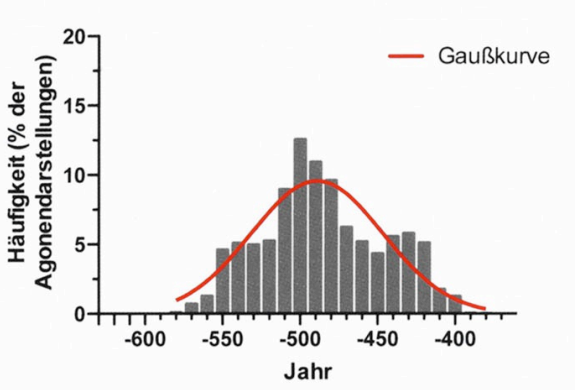


www.spiegel.de/wirtschaft/unternehmen/ostern-schokoladen-osterhase-schlaegt-weihnachtsmann-a-1083821.html

Wolfgang Filser, Die Elite Athens auf der attischen Luxuskeramik, IKON 16 (Berlin 2017) Taf. IX



1. Darstellungen der Symposia



2. Darstellungen der Athletik

What are the basics of evaluation?

How convincing are the results?

What procedures are there?

How do you create a good data basis?

What are typical mistakes?

1. BASICS OF STATISTICS

Descriptive Statistics

Inferential Statistics

Normal distribution

Standard deviation

Mean of a distribution

Correlations

2. SAMPLING

Large Data Collections on the Internet

Canonisation of knowledge

Data Acquisition

Sampling methods

3. EXAMPLES

Sampling: Geo Data

Sampling: Artefacts

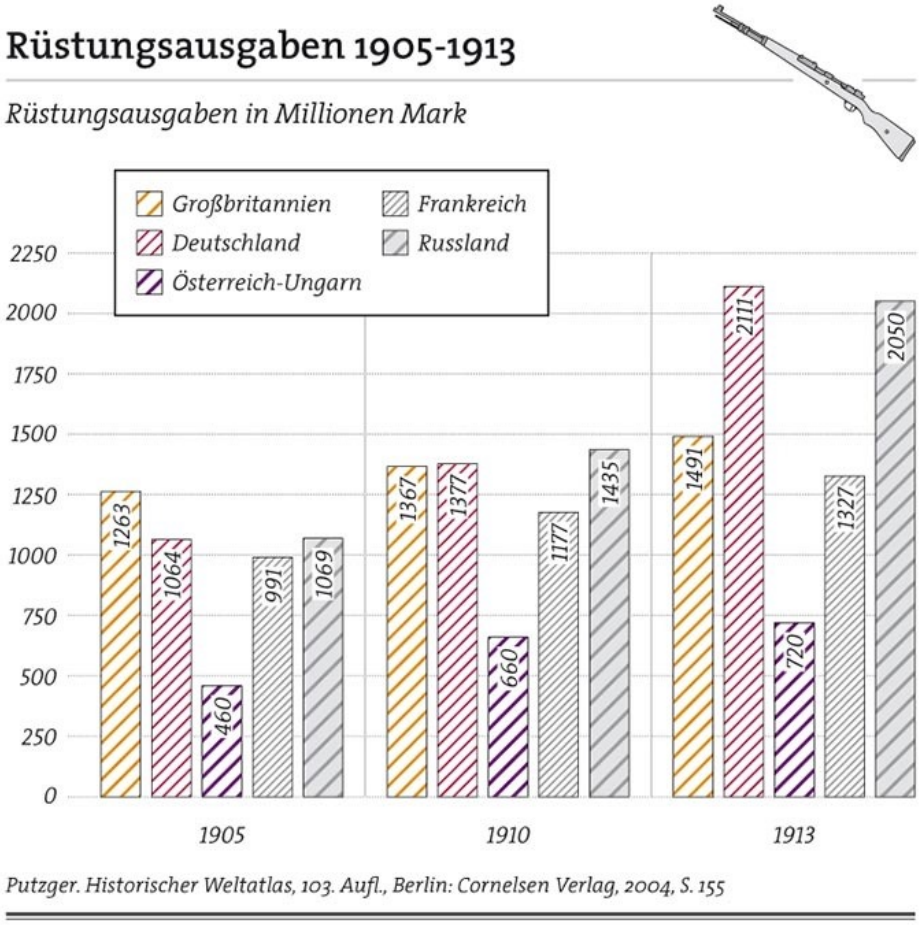
Tests of representativeness

Sampling: Posters

BASICS OF STATISTICS

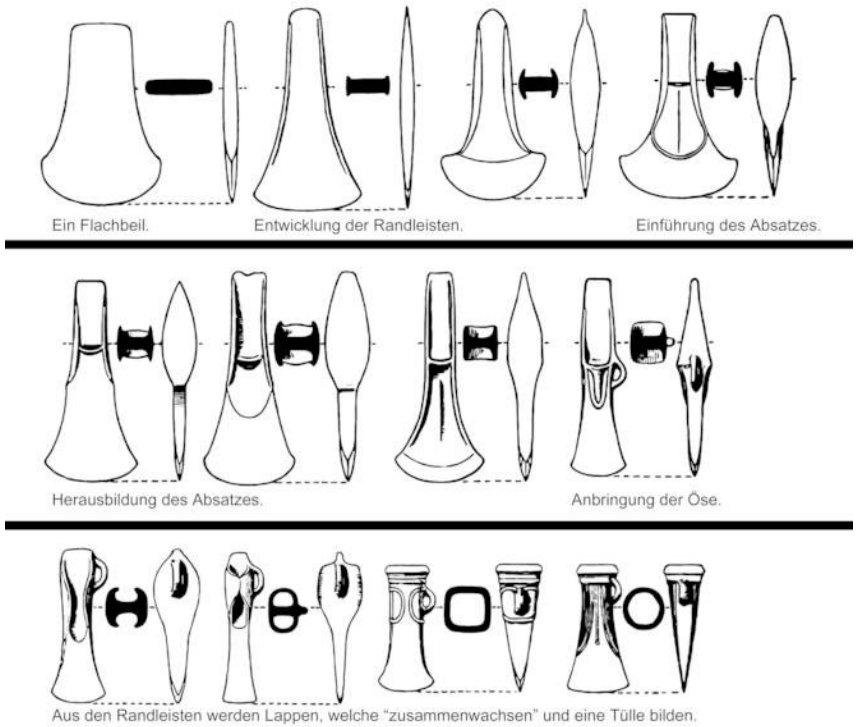
WIDTH INSTEAD OF DEPTH

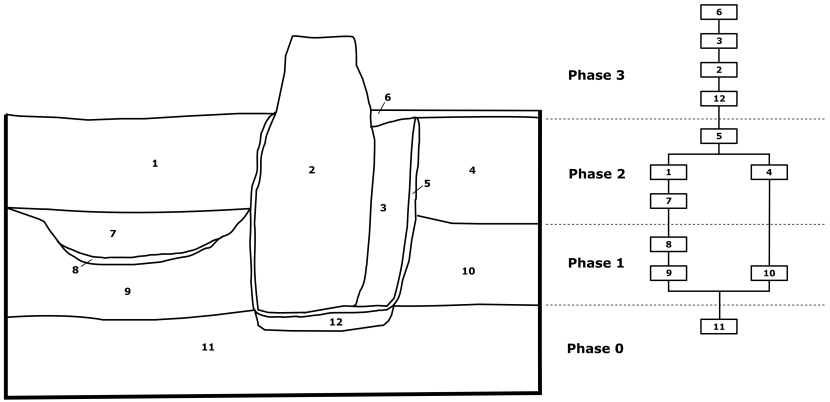
Quantitative research methods are complementary to qualitative approaches.
They can stimulate qualitative research or make its results more probable.



https://m.bpb.de/izpb/142137/1880-bis-1914

THE PROBABILITY PROOF



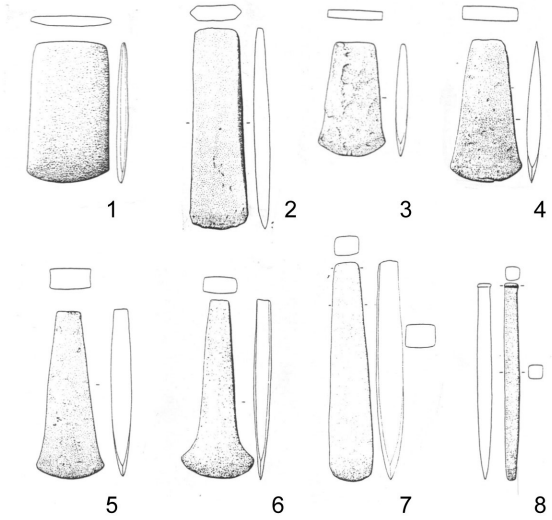


In the humanities, it is usually true that a thesis with the best arguments for its correctness is valid until other facts or better arguments against it are found.

www.praehistorische-archaeologie.de/wissen/datierung/typologie/typologie/

DEPENDENCY ON THE HYPOTHESIS

The data basis must be created depending on the hypothesis to be tested, i.e. it must be representative for this question.



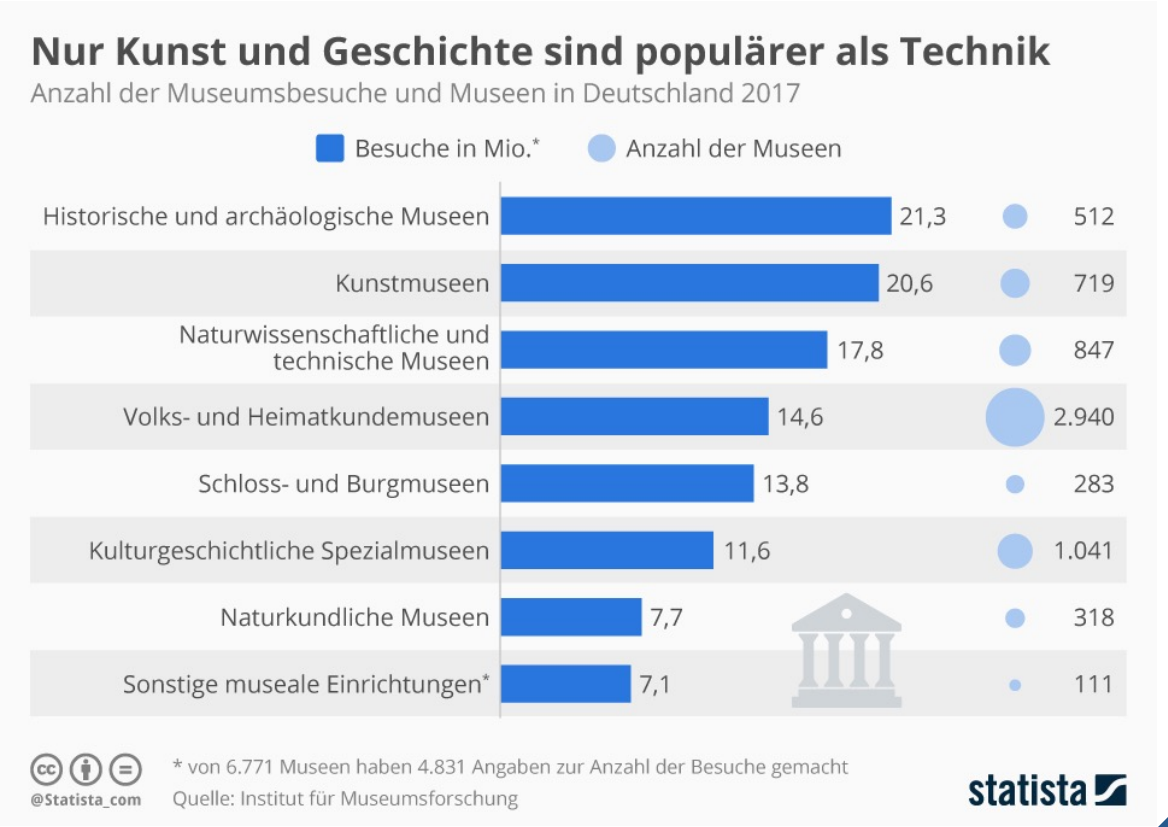
7.7.4 Schneidenformen

Schneidenformen	BF1	BF2	BF3	BF4	BF5	BF6	BF7	BF8
1	30	19	22	25	21	18	32	12
2	34	69	39	117	92	49	106	30
3	10	8	2	7	14	5	22	19
4	29	32	16	33	46	30	58	12
5	1	1	0	1	0	1	1	1
6	0	0	0	0	0	0	1	0
0	1	6	2	3	21	5	30	4
Gesamt	105	135	81	186	194	108	250	78
1	29 %	14 %	27 %	13 %	11 %	17 %	13 %	15 %
2	32 %	51 %	48 %	63 %	47 %	45 %	42 %	38 %
3	10 %	6 %	2 %	4 %	7,5 %	5 %	9 %	24 %
4	28 %	24 %	20 %	18 %	24 %	28 %	23 %	15 %
5	1 %	1 %		0,5 %		1 %	0,5 %	1 %
6							0,5 %	
0	1 %	4 %	2 %	2 %	11 %	5 %	12 %	5 %
Gesamt	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %

https://scidok.sulb.uni-saarland.de/bitstream/20.500.11880/23515/2/Albert Schmitz_ProfDrJanLichardus_Band1.pdf

DESCRIPTIVE STATISTICS

The ways of presentation in descriptive statistics are key figures, graphs and tables.



http://cdn.statcdn.com/Infographic/images/normal/17938.jpeg

DESCRIPTIVE STATISTICS

Categorical Data

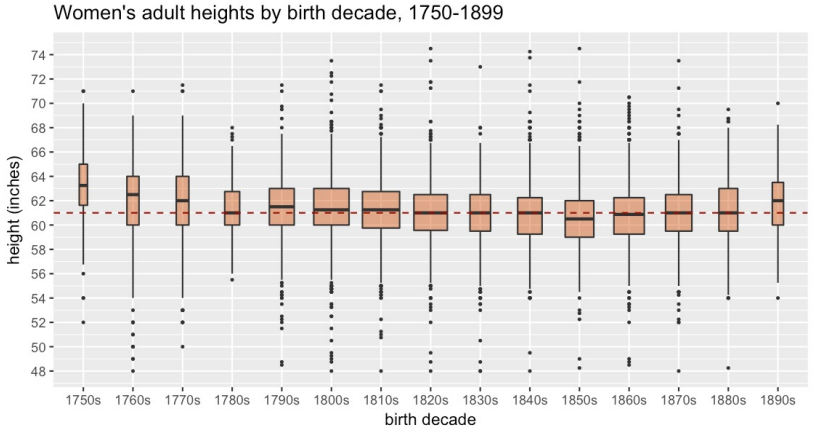
= Number of individuals in a group/category (frequency).

- Visualisation as bar chart (absolute) or pie chart (relative)

Numerical Data

= Measured values

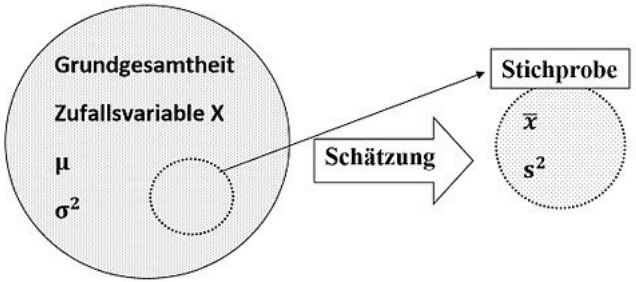
- Visualisation as histogram or box plot



https://www.dataquest.io/blog/box-plots-womens-heights/

INFERENTIAL STATISTICS

essentially deals with the question of the randomness of statistically measured phenomena. An attempt is made to classify the examined sample in a larger whole, whereby broad space is also given to the examination of the probability of correlations or differences.



Stichprobe Nr.	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	$\bar{x}$	z	$(x_{[1]} + x_{[9]})/2$
1	9,26	10,31	15,8	11,43	14,01	6,53	6,3	11,16	3,98	9,86	10,31	9,89
2	4,46	6,22	13,89	11,31	12,83	9,67	10,19	7,49	9,79	9,54	9,79	9,175
3	5,73	12,42	3,3	13,1	14,7	8,73	14,04	14,12	9,64	10,64	12,42	9,00
4	8,76	22,24	10,44	11,2	9,16	8,9	13,98	9,96	8,74	11,49	9,96	15,49
5	2,14	11,48	5,83	13,45	10,51	12,8	8,49	8,01	11,69	9,38	10,51	7,795
6	12,52	6,53	11,46	7,87	9,9	10,77	4,86	11,26	15,38	10,06	10,77	10,12
7	4,64	12,22	10,27	7,97	12,26	16,27	6,5	13,88	10,13	10,46	10,27	10,455
8	11,64	5,18	11,73	11,19	10,59	11,48	9,00	10,23	10,56	10,18	10,59	8,46
9	10,93	5,05	12,81	10,23	4,81	8,86	11,52	6,01	14,6	9,42	10,23	9,705
10	12,1	10,42	9,04	8,23	16,2	14,57	13,1	7,3	6,44	10,82	10,42	11,32
11	2,57	14,67	13,09	10,15	10,5	6,28	8,34	13,26	11,09	9,99	10,5	8,62
12	3,45	10,42	8,86	10,16	-1,17	8,71	10,25	-0,36	4,84	6,13	8,71	4,62
13	11,21	11,09	-2,77	16,24	11,59	9,08	5,38	12,57	9,14	9,28	11,09	6,735
14	8,62	6,78	9,62	15,45	12,9	7,19	7,61	16,49	15,04	11,08	9,62	11,63
15	13,23	7,92	10,17	15,38	7,6	7,8	13,85	13,58	13,41	11,44	13,23	11,49
16	9,35	12,09	11,76	9,05	11,89	12,76	11,42	9,07	11,81	11,02	11,76	10,905
17	6,6	4,16	7,8	17,3	10,22	10,74	6,66	13,61	5,47	9,17	7,8	10,73
18	4,01	15,34	8,28	11,49	7,83	7,37	8,51	9,98	14,21	9,67	8,51	9,675
19	6,21	1,72	0,55	4,85	7,14	12,3	13,33	0,39	12,96	6,61	6,21	6,86
20	9,66	10,17	13,75	8,3	11,32	12,09	11,79	5,23	16,5	10,98	11,32	10,865

https://de.wikibooks.org/wiki/Statistik:_Schätzen_und_Testen

BASIC CONCEPTS OF STATISTICS I

Characteristic (also variable) = the respective peculiarity of the object of investigation, the characteristics of which can vary (in contrast to a constant).

Expression = the totality of possible values of a characteristic.

Quantification = the numerical description of characteristic values of a variable on the basis of measurements or counts.

https://www.weltkunst.de/wp-content/uploads/2020/08/The-Olbricht-Collection-Wunderkammer-810x1077.jpg

Merkmalstyp	Skala	Variable oder Merkmal	Merkmalsausprägung oder Wert
qualitativ	Nominal	Parteizugehörigkeit	CDU, SPD, Grüne,...
qualitativ	Nominal	Wahrheitswert einer Aussage	Wahr, falsch
qualitativ	Nominal	Spielausgang beim Toto	0,1,2
Rang	Ordinal	Schulnoten	1,2,3,4,5,6
Rang	Ordinal	Hausnummer	...,12,14,16 ,...
Rang	Ordinal	Dienstgrade bei der Bundeswehr	Gefreiter, ..., General
quantitativ	Metrisch, Intervallskala, stetig bzw. quasi-diskret	Uhrzeit	2:00, 4:00
quantitativ	Metrisch, Intervallskala, stetig bzw. quasi-diskret	Temperatur in Grad Celsius	...,12,13,14,...
quantitativ	Metrisch, Verhältnisskala, stetig	Entfernung zwischen Wohn- und Arbeitsstätte	1 km, 1,5 km, ...
quantitativ	Metrisch, Verhältnisskala, stetig	Alkoholgehalt im Blut	0, 0,1, ...0,8,...

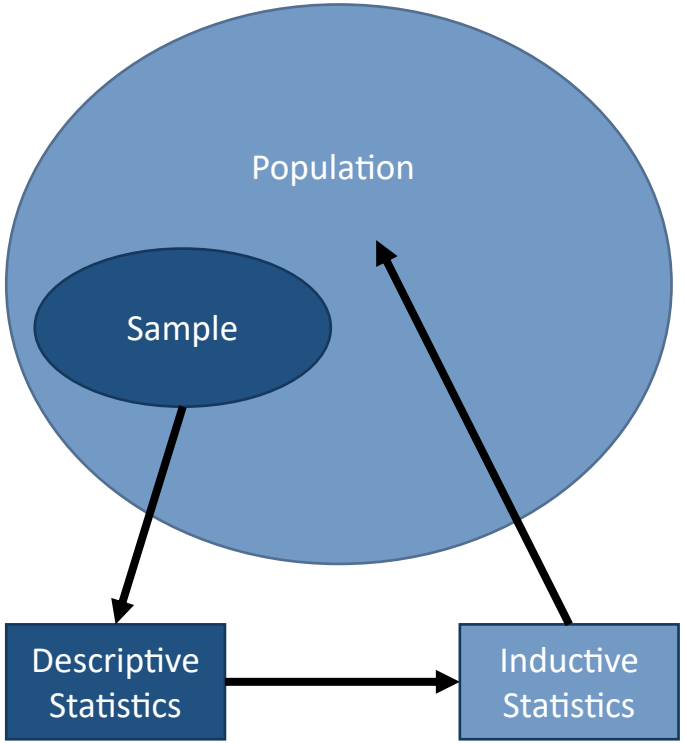
CHARACTERISTIC AND EXPRESSION

Claus Brell, Juliana Brell, Siegfried Kirsch, Statistik von Null auf Hundert (Springer, 2016)

BASIC CONCEPTS OF STATISTICS II

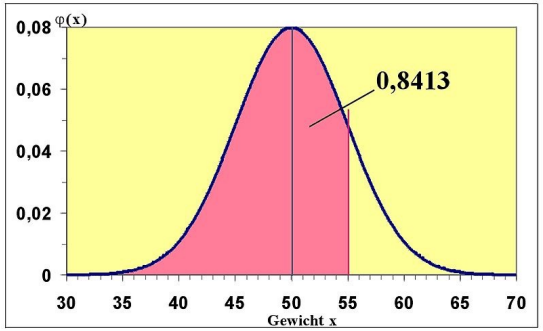
Population = the totality of all elements for which the statements of the study are to apply. Since a complete survey is rarely possible, a sample is usually selected.

Sample = the selection of observation units from a defined (basic) population. A sample should reflect this basic population without bias, e.g. through the model of representativeness.



THE GAUSSIAN NORMAL DISTRIBUTION

Normal distribution means a steady (continuous) distribution of randomly collected data (such as body size, the amount of ceramics in a deposit or the settlement density in a quadrant).

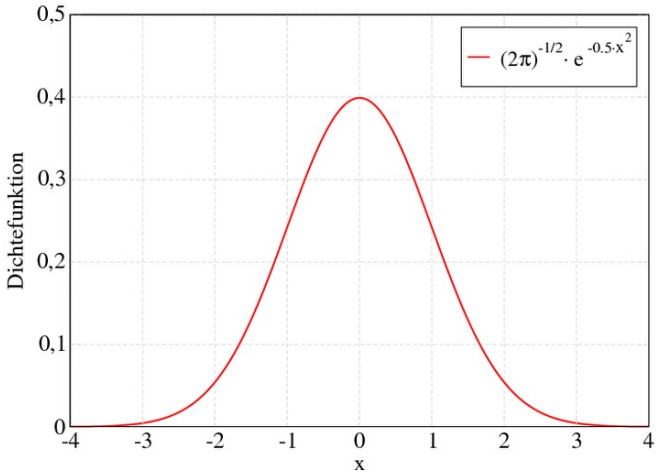


On a chicken farm with a lot of chickens, the individual eggs are weighed for one week. Let's define the random variable X as the weight of an egg in grams. It turns out that an egg weighs 50 g on average. The expected value is therefore 50. Let it also be known that the variance varX = 25 g2. One can approximate the distribution of weight as shown in the graph. We can see that most of the eggs are near the expected value of 50 and that the probability of getting very small or very large eggs becomes very small. We are looking at a normal distribution. It is typical for random variables that are made up of very many different influences that can no longer be separated, e.g. weight of the chicken, age, health, location, heredity, etc.

Thus, the probability that an egg weighs at most 55 g is 0.8413. This corresponds to the red area in the figure.

THE GAUSSIAN NORMAL DISTRIBUTION

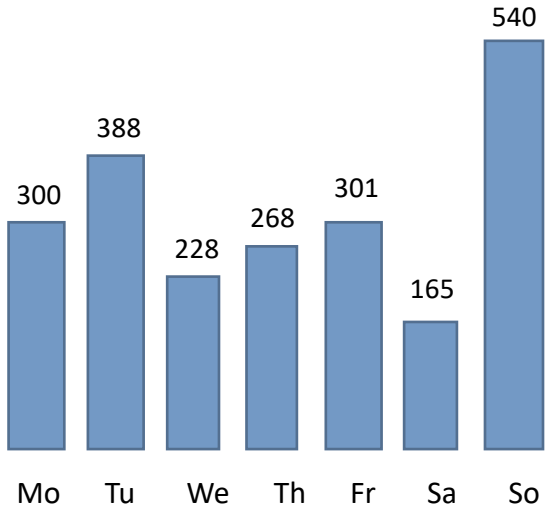
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$



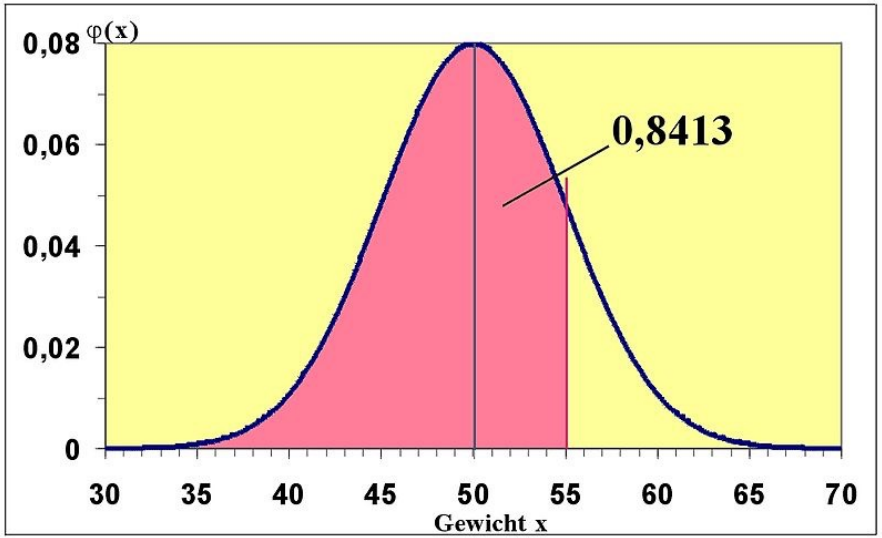
The special significance of normal distribution is based, among other things, on the central limit theorem, which states that a sum of n independent, identically distributed random variables with finite variance is normally distributed in the limit $n \to \infty$.

This means that random variables can also be regarded as normally distributed if they result from the superposition of a large number of independent influences, whereby each individual influencing variable makes an insignificant contribution in relation to the total sum.

Average (arithmetic mean) = sum of the values divided by their numberl

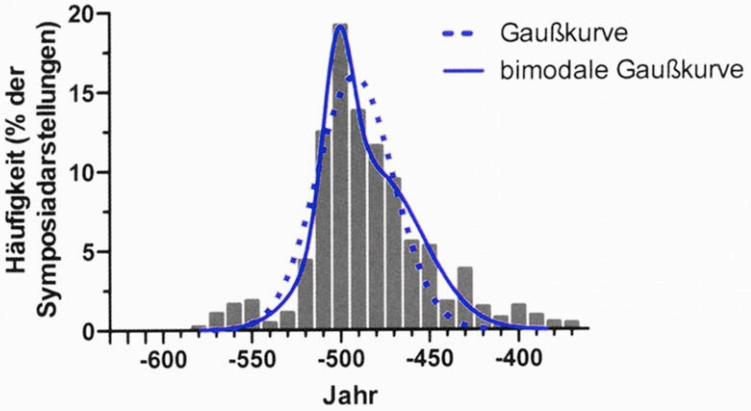


Quantity distribution of the units

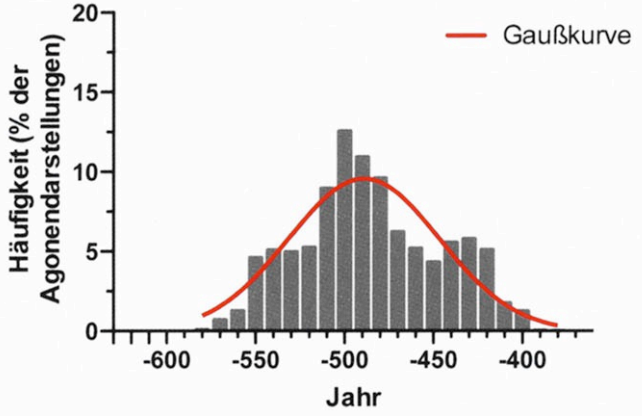


Frequency distribution of the measured values

THE GAUSSIAN NORMAL DISTRIBUTION



1. Darstellungen der Symposia

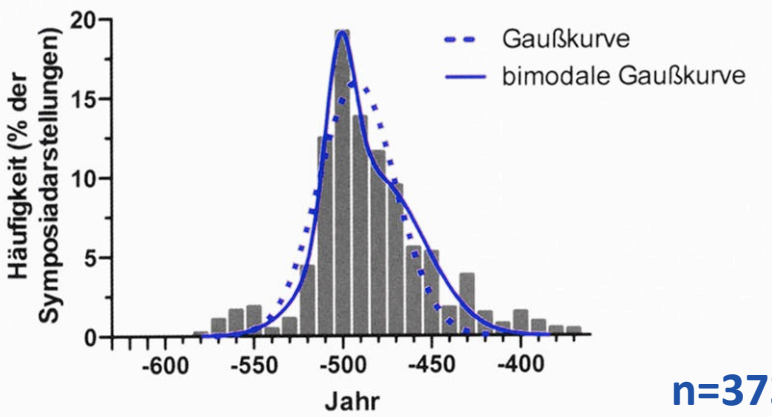


2. Darstellungen der Athletik

Annual figures are not measured values and have nothing to do with normal distribution!

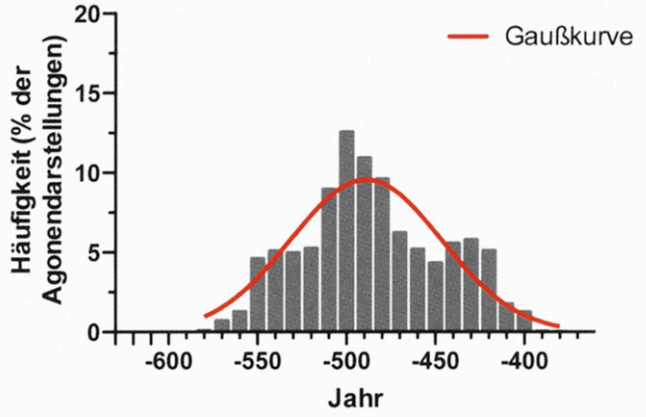
Wolfgang Filser, Die Elite Athens auf der attischen Luxuskeramik, IKON 16 (Berlin 2017) 100. 591–594 Taf. IX
zu 373 Symposiendarstellungen und 1216 Vasenbildern athletischer Thematik

HOW REPRESENTATIVE ARE SUCH STATISTICS?



n=373

1. Darstellungen der Symposia



n=1216

2. Darstellungen der Athletik

Legende
1. Prozentuale Häufigkeitsverteilungen der Darstellungen von Symposia in Abhängigkeit vom Jahrzehnt ihrer Produktion und hieraus berechnete Verteilungskurven (Gaußkurve: blau, gestrichelt; bimodale Normalverteilung: blau, durchgezogen).
2. Prozentuale Häufigkeitsverteilungen der Darstellungen von athletischen Szenen in Abhängigkeit vom Jahrzehnt ihrer Produktion und hieraus berechnete Gaußsche Verteilungskurve (rot).

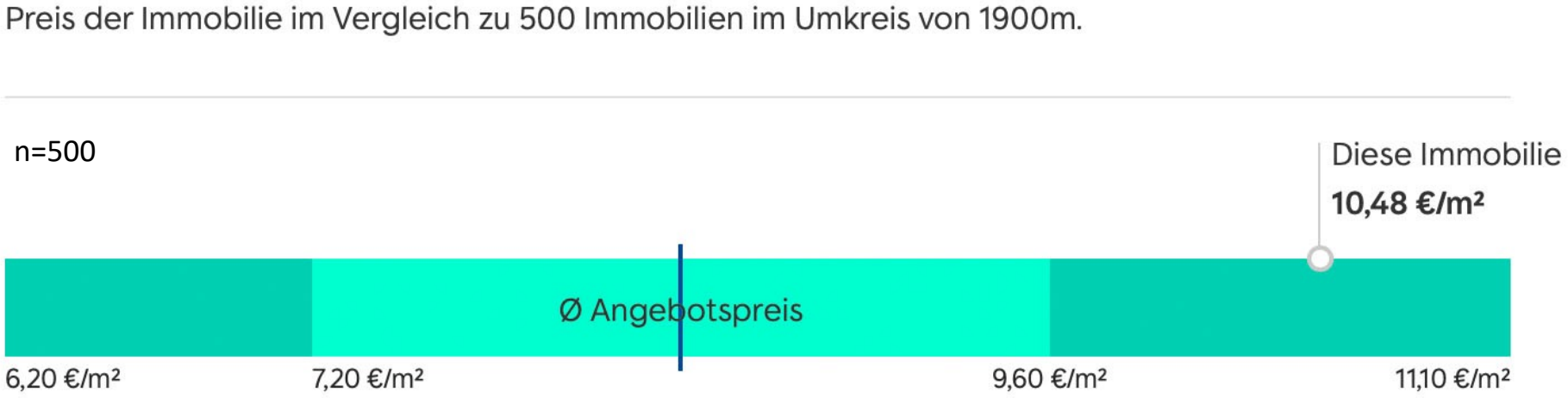
Die Kurvenberechnungen wurden mit dem Programm „Prism5 for Mac OS X" von GraphPad Software durchgeführt.

Wolfgang Filser, Die Elite Athens auf der attischen Luxuskeramik, IKON 16 (Berlin 2017) 100. 591–594 Taf. IX

on 373 symposium representations and 1216 vase paintings with athletic themes

STANDARD DEVIATION

An estimated value (s) is usually given as the standard deviation depending on the sample size (n).



STANDARD DEVIATION

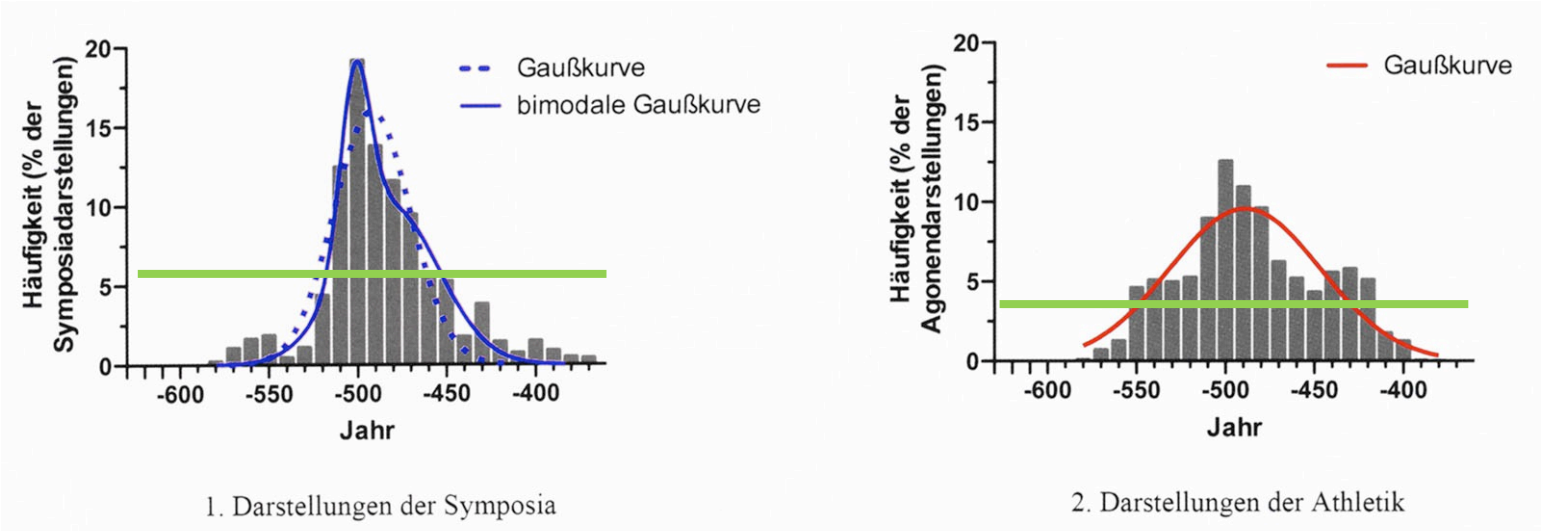
The standard deviation as a function of the sample size (n) is usually given as an estimated value (s) of ± s / √n, which means in percent ± 100 / √n.

For orientation, a few confidence intervals (e) are given to indicate the error limits:

n = 30: e = ± 18,3%
n = 50: e = ± 14,1%
n = 75: e = ± 11,5%
n = 100: e = ± 10,0%
n = 200: e = ± 7,1%

n = 400: e = ± 5,0%
n = 800: e = ± 3,5%
n = 1000: e = ± 3,2%
n = 2000: e = ± 2,2%
n = 5000: e = ± 1,4%

HOW REPRESENTATIVE ARE SUCH STATISTICS?



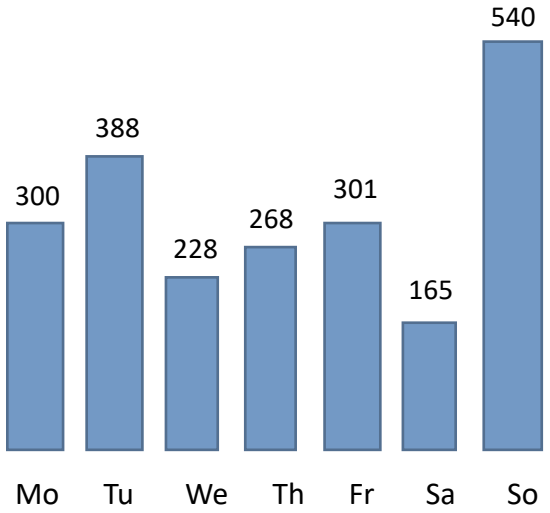
n = 373: e = ± 5,8%

n = 1216: e = ± 3,0%

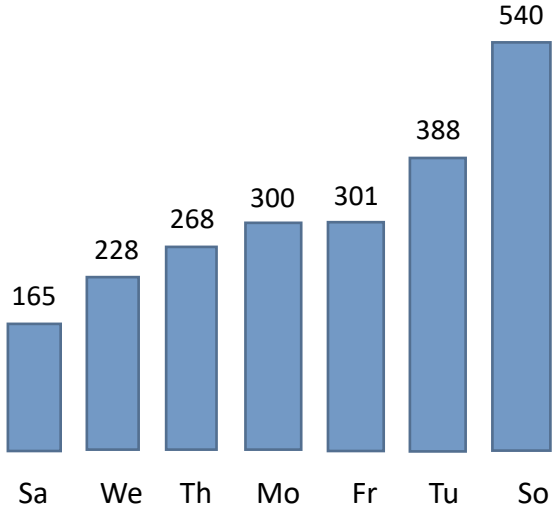
The specification of the surveyed sample size (***n***) is mandatory!

Average (arithmetic mean) = sum of the values divided by their number

Mean (median) = the middle value in a sorted series of values



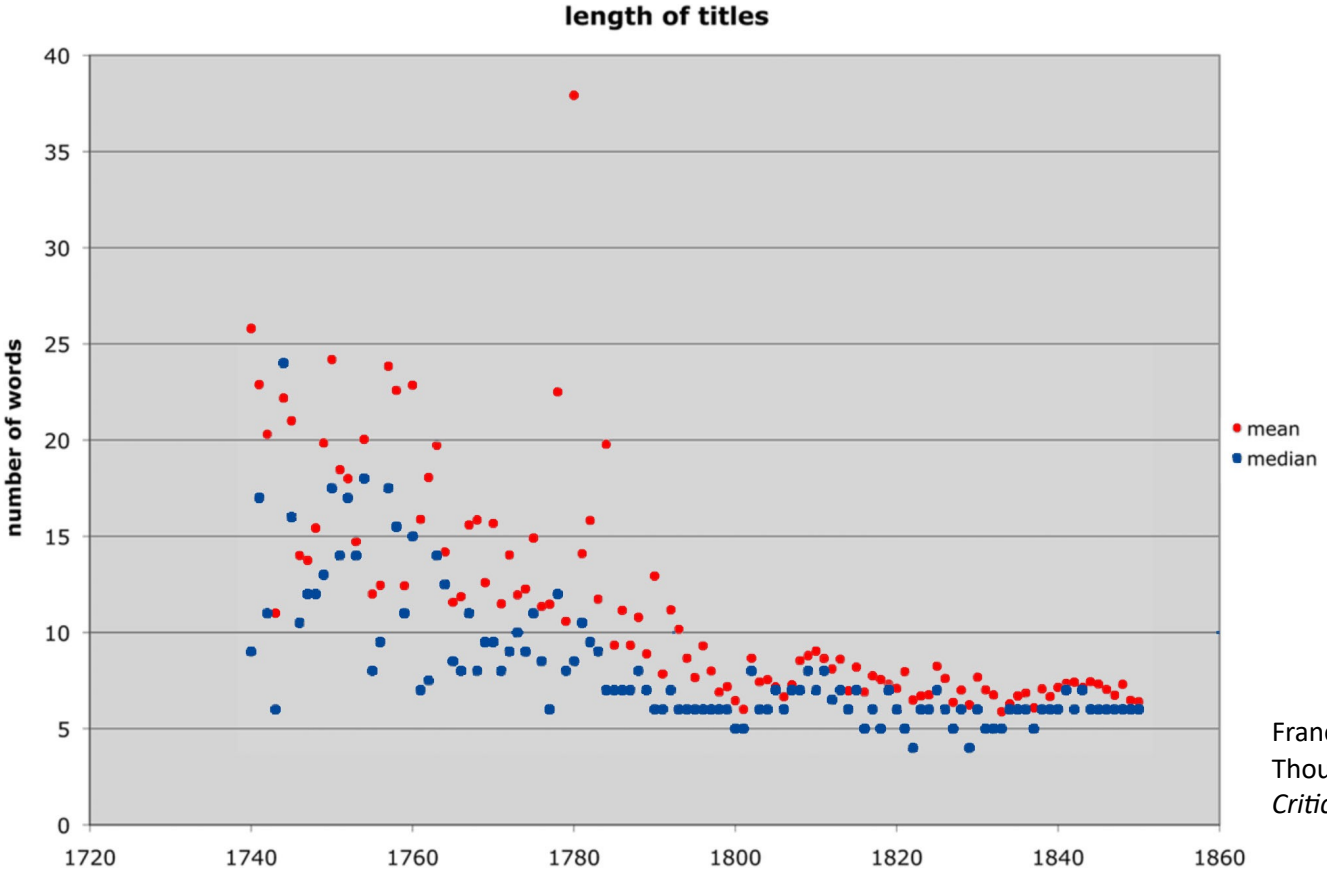
Quantity distribution of the units



Quantity distribution of the units in a sorted row

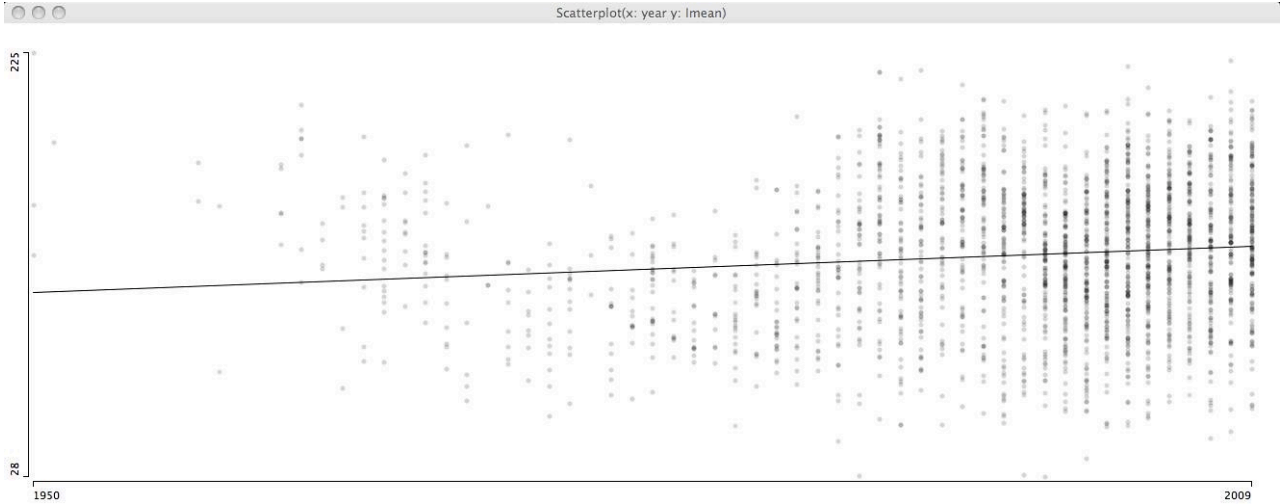
Ø=312,86

m=300



Franco Moretti, “Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740– 1850),” *Critical Inquiry* 36, no. 1 (2009): 135 Abb. 1

MEAN OF A DISTRIBUTION



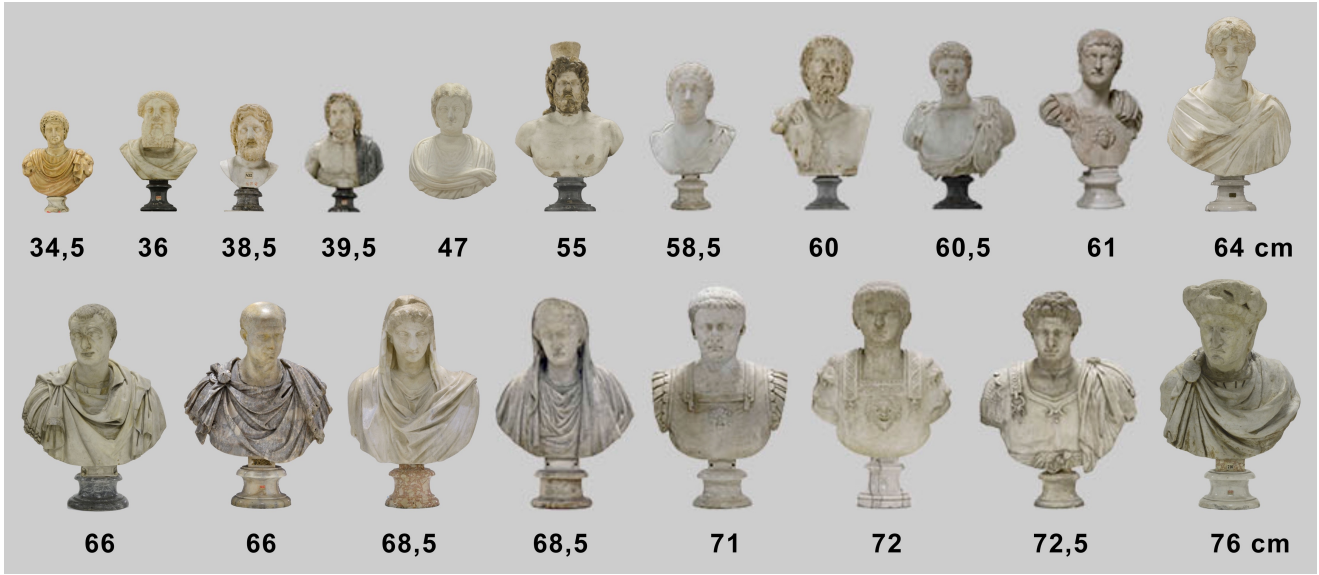
https://live.staticflickr.com/4072/5137599115_b19fc75568_h.jpg

Peter Leonard, Vogue Cover Averages (2013):
https://thegogglesdonothing.com/archives/2013/10/vogue_cover_averages.shtml

Average (arithmetic mean) = sum of the values divided by their number [=58,68]

Mean (median) = the middle value in a sorted series of values [=61]

Mode (modal value) = most frequent value of a distribution, value with the greatest probability. The value can be unimodal, bimodal or multimodal.



Height of the 19 marble busts (without pedestals) in the Obizzi collection, the value of which was estimated at two Venetian Zecchini each in the 1806 inventory of the estate.

Central tendency:

- Arithmetic mean (= sum of the values divided by their number)
- Median (= the middle value in a sorted series of values)
- Modal value (= most frequent value of a distribution, value with the greatest probability)

Statistical dispersion:

- Variance (= mean squared deviation from the arithmetic mean).
- Range (min/max, = difference between largest and smallest observation)
- Standard deviation (= root of the variance)

https://en.wikipedia.org/wiki/Summary_statistics

BASIC CONCEPTS OF STATISTICS III

Representativeness is given if the composition of the basic population is replicated or approximated by tests when selecting the elements of the sample.

Probability = the classification of phenomena according to their degree of certainty. The probability p is represented by values between 0 (impossibility) and 1 (certainty of occurrence).

Relationship = systematic correspondence between the expressions of two variables.

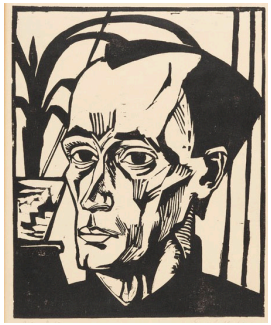
Correlation = relationship between two quantitative characteristics. The strength of the correlation is expressed by the correlation coefficient. It lies between the extremes -1 and +1. If it is positive, it means that a high value of variable A is accompanied by a high value of variable B. The same applies to low values.

CONTINGENCY TABLE


K. Schmidt-Rottluff


H. M. Pechstein

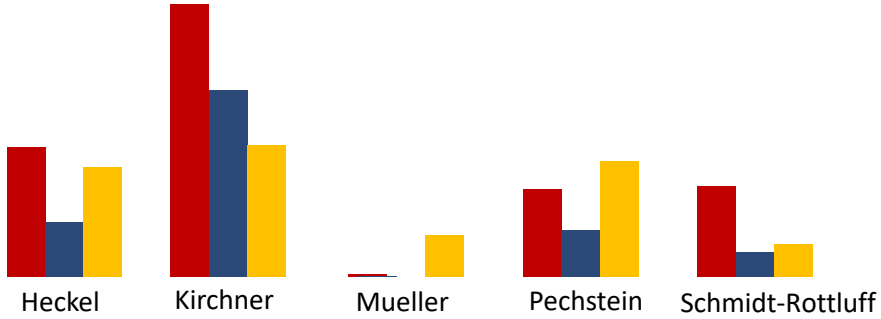

E. L. Kirchner


E. Heckel


O. Mueller

Absolute frequency

Künstler \ Technik	Holz-schnitt	Radie-rung	Litho-graphie	Gesamt
Heckel	465	197	401	**1063**
Kirchner	985	675	479	**2139**
Mueller	13	1	153	**167**
Pechstein	316	165	423	**904**
Schmidt-Rottluff	334	92	118	**544**
Gesamt	**2113**	**1130**	**1574**	**4817**



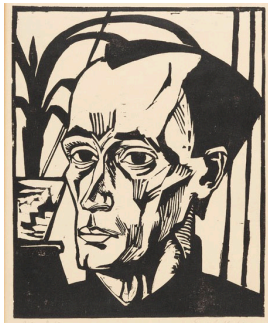
Prints of the Brücke members by artist and printing technique (basis: current catalogues raisonnés)

CONTINGENCY TABLE


K. Schmidt-Rottluff


H. M. Pechstein

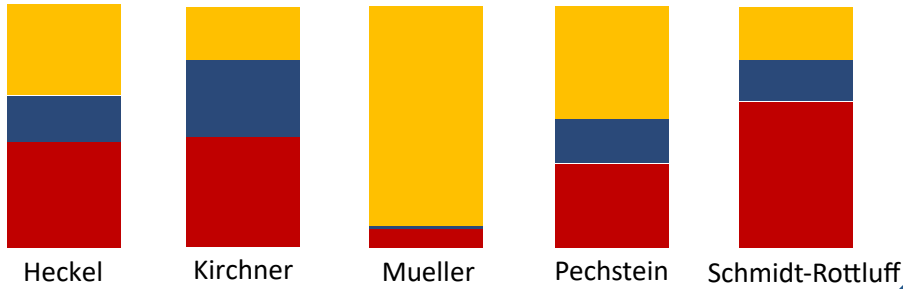

E. L. Kirchner


E. Heckel


O. Mueller

Relative frequency in %

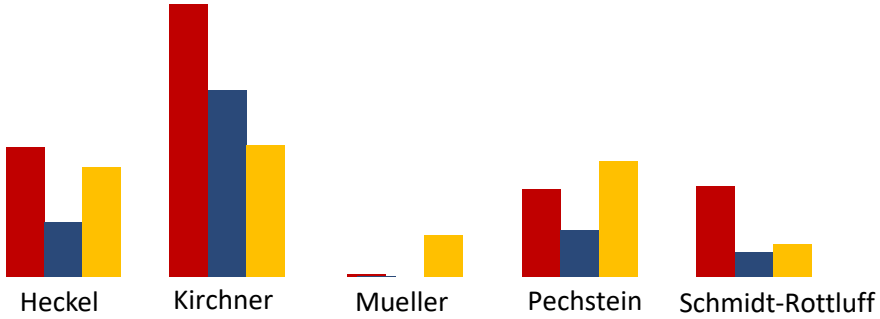
Künstler \ Technik	Holz-schnitt	Radie-rung	Litho-graphie	Gesamt
Heckel	43,7	18,5	37,7	**100**
Kirchner	46,0	31,6	22,4	**100**
Mueller	7,8	0,6	91,6	**100**
Pechstein	35,0	18,3	46,8	**100**
Schmidt-Rottluff	61,4	16,9	21,7	**100**

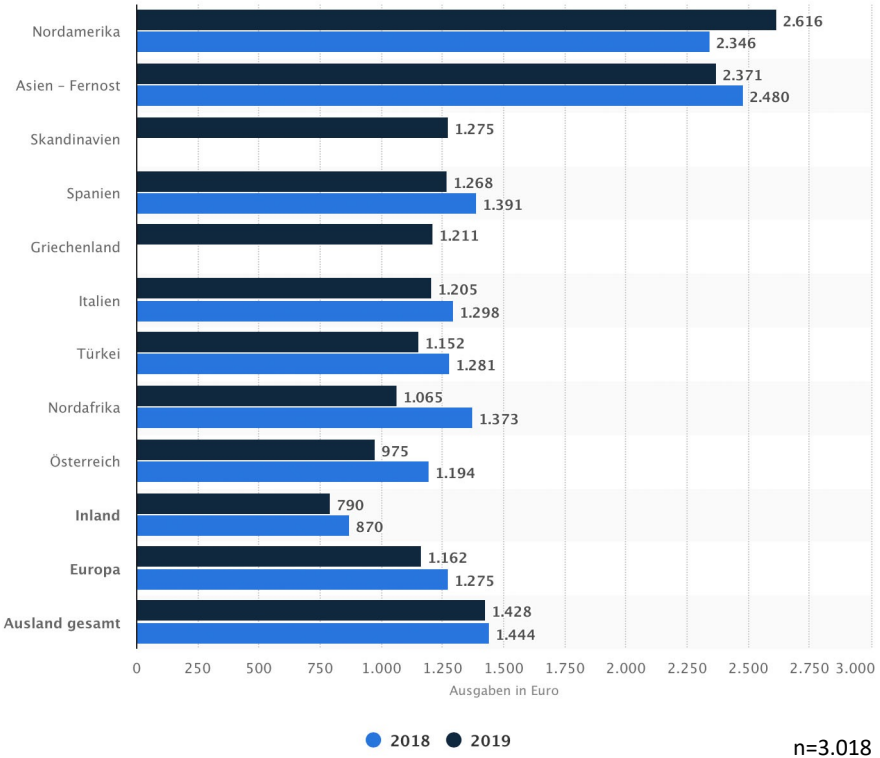


Prints of the Brücke members by artist and printing technique (basis: current catalogues raisonnés)

DESCRIPTIVE STATISTICS

	Holz-schnitt	Radie-rung	Litho-graphie	Gesamt
Heckel	465	197	401	**1063**
Kirchner	985	675	479	**2139**
Mueller	13	1	153	**167**
Pechstein	316	165	423	**904**
Schmidt-Rottluff	334	92	118	**544**
Gesamt	**2113**	**1130**	**1574**	**4817**

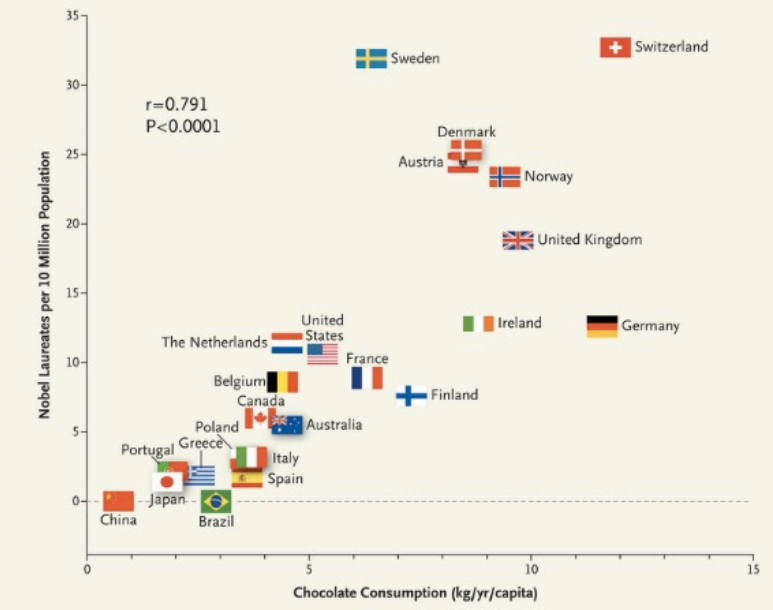




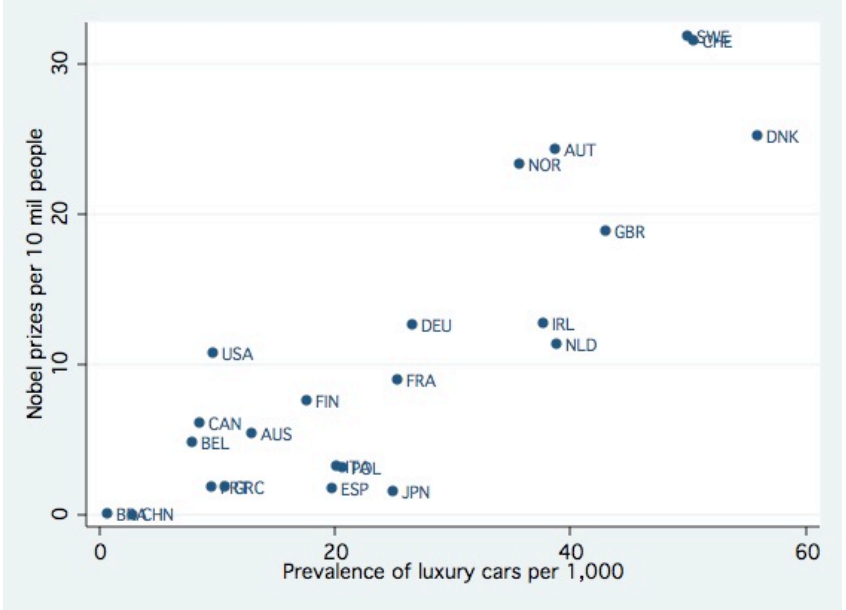
n=3.018

https://de.statista.com/statistik/daten/studie/217740/umfrage/ausgaben-der-deutschen-fuer-urlaub/

CORRELATIONS AND CAUSAL RELATIONSHIPS



Franz H. Messerli, M.D., Chocolate Consumption, Cognitive Function, and Nobel Laureates, New England Journal of Medicine 367:16, 2012, 1562–1564: https://www.nejm.org/doi/full/10.1056/NEJMon1211064

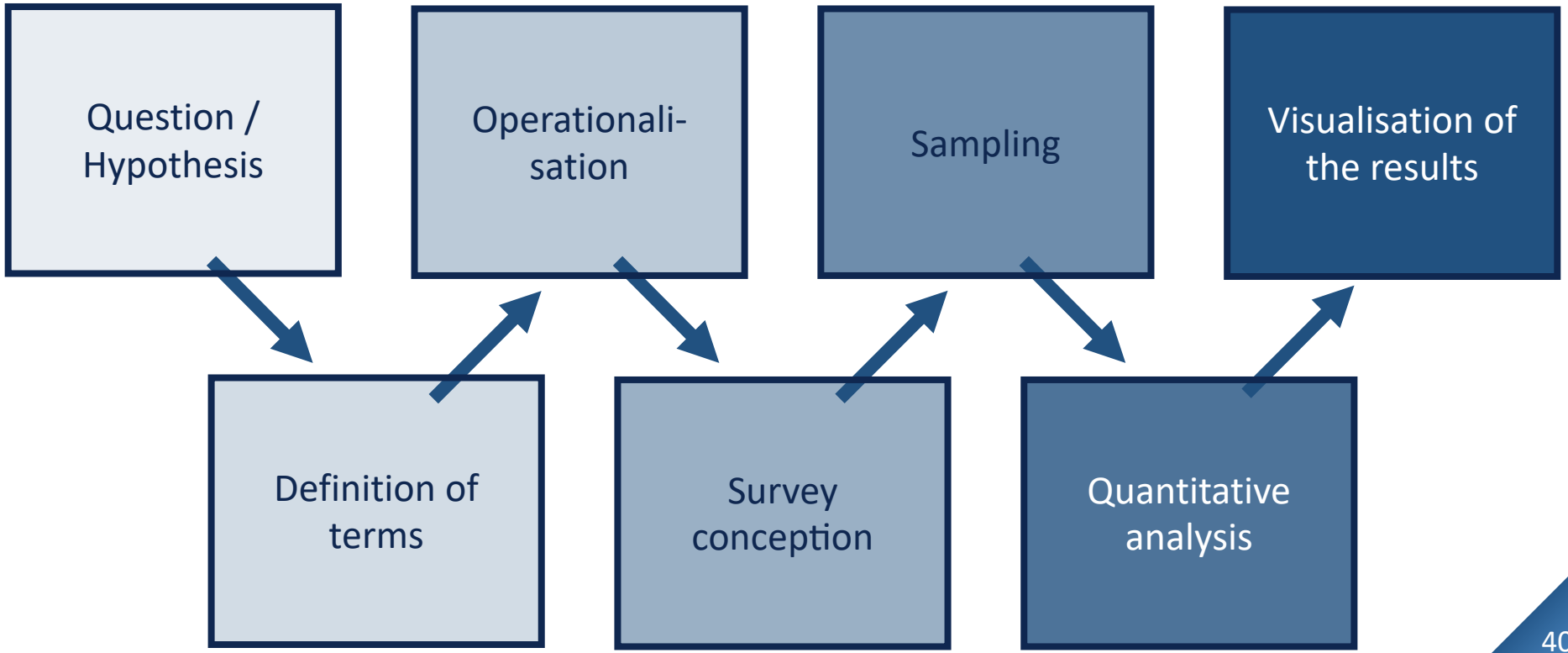


https://epianalysis.wordpress.com/2012/11/19/chocolate/

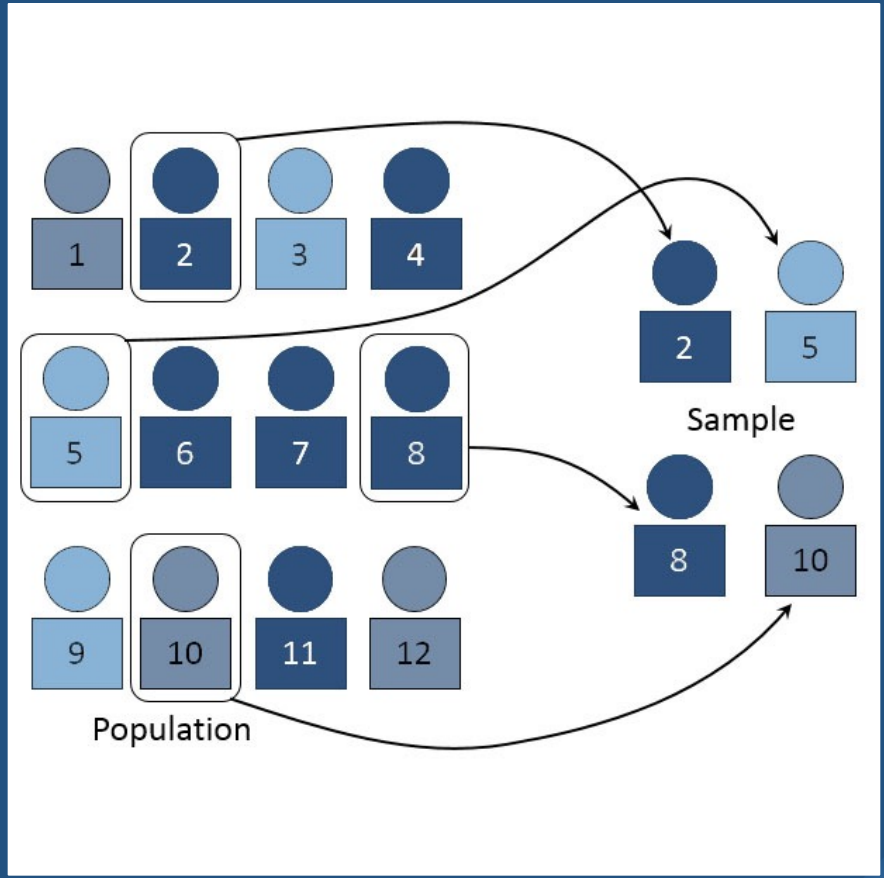
ESTIMATES

https://digitalhumanities.umass.edu/pbmp/?p=1565

PROCEDURE OF A QUANTIFICATION STUDY



SAMPLING: BASICS



PROJECT GUTENBERG
(www.gutenberg.org; www.projekt-gutenberg.org; über 60.000 e-books)

PROJEKT GUTENBERG-DE

Startseite ∞ Alle Werke ∞ Neu ∞ Information ∞ Shop ∞ Lesetips ∞ Themen ∞ Buchverlag ∞ Impressum ∞ Datenschutz ∞ Quellenangabe

Autoren: A · B · C · D · E · F · G · H · I · J · K · L · M · N · O · P · Q · R · S · T · U · V · W · X · Y · Z · Alle Autoren

10.000 Werke lokal lesen: Gutenberg-DE Edition 15 auf USB. Information und Bestellung in unserem Shop

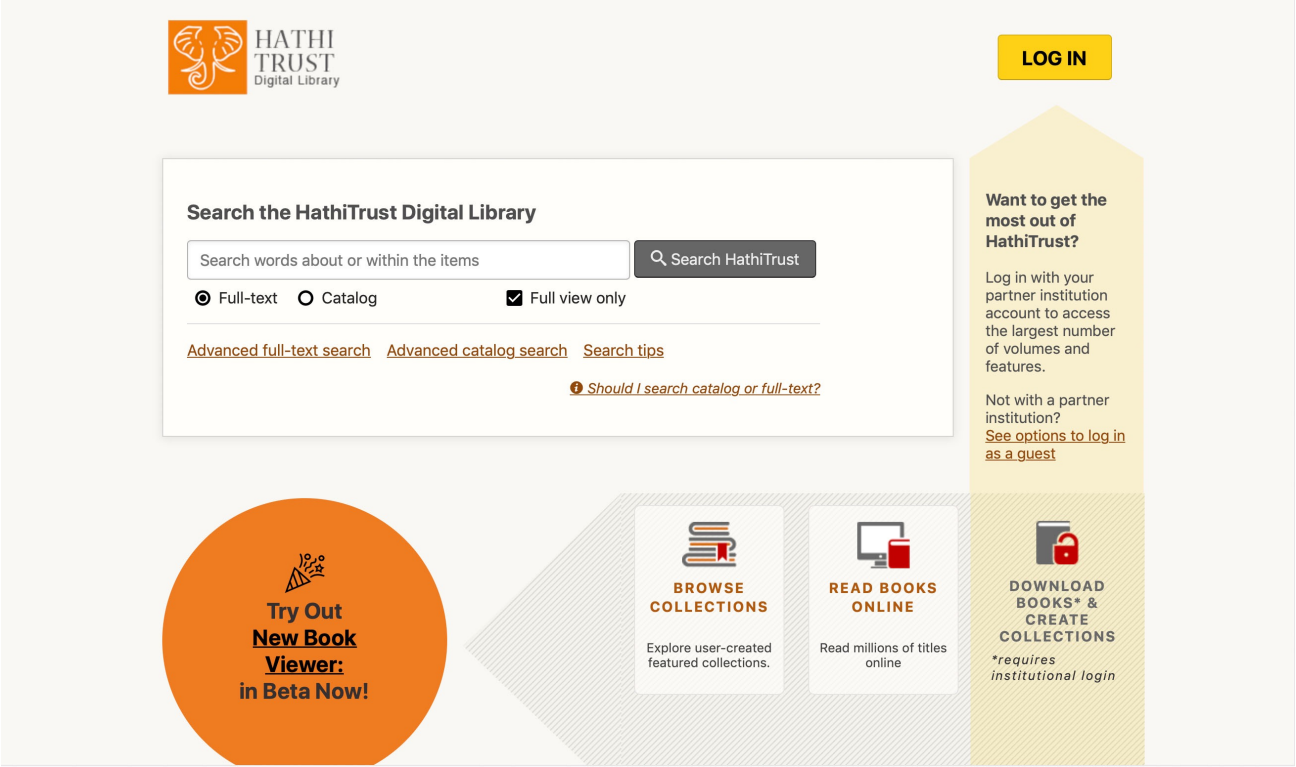
STARTSEITE

Wer vieles bringt, wird manchem etwas bringen;
Und jeder geht zufrieden aus dem Haus.

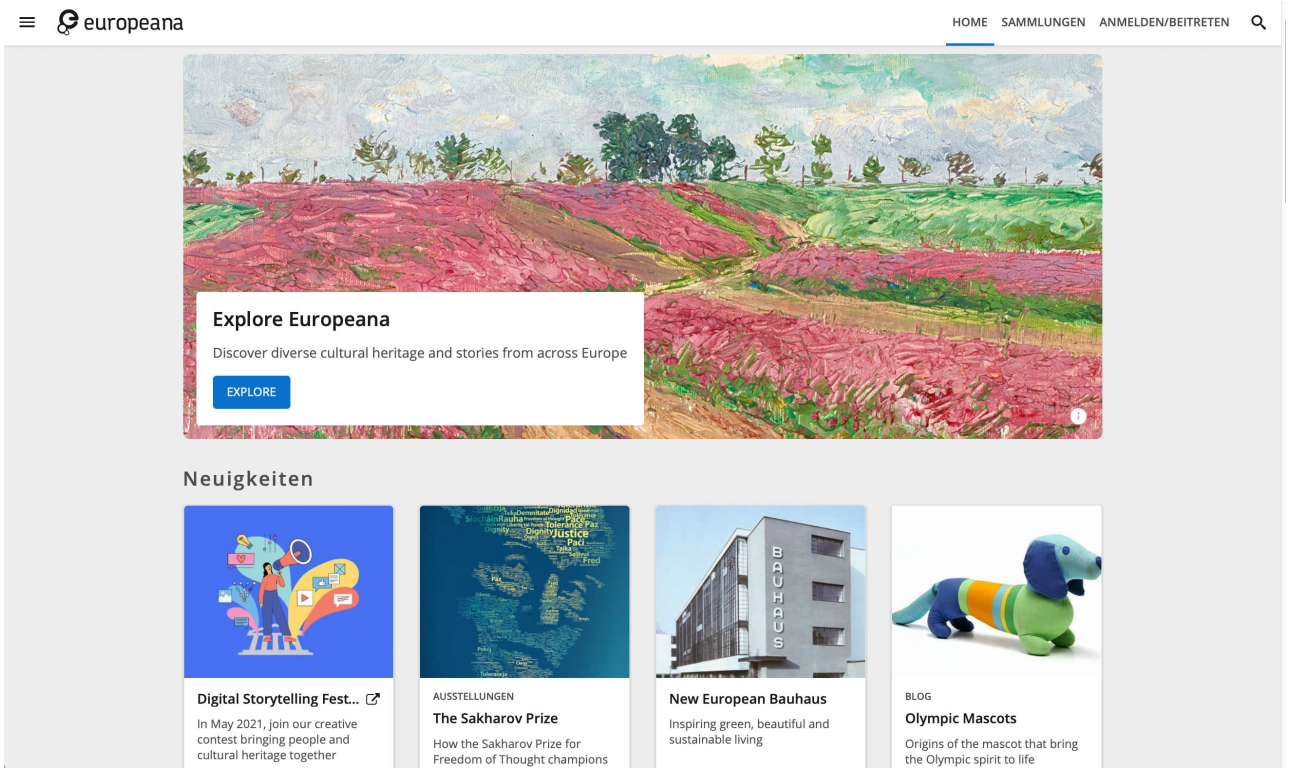
Das möchten wir erreichen und machen deshalb auf einige Themen abseits der großen Klassiker im Projekt Gutenberg aufmerksam.

Natur:

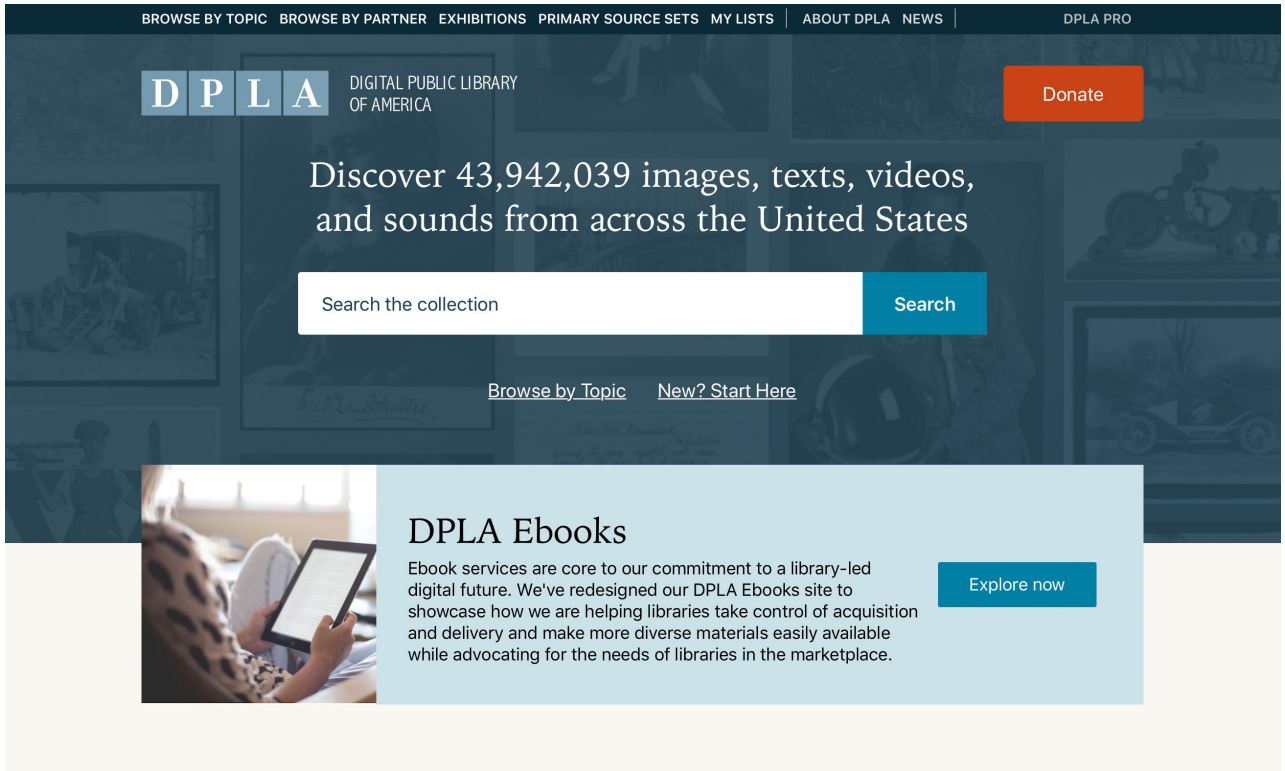
HAITHI TRUST DIGITAL LIBRARY
(www.hathitrust.org; 9.055.899 digitised books)



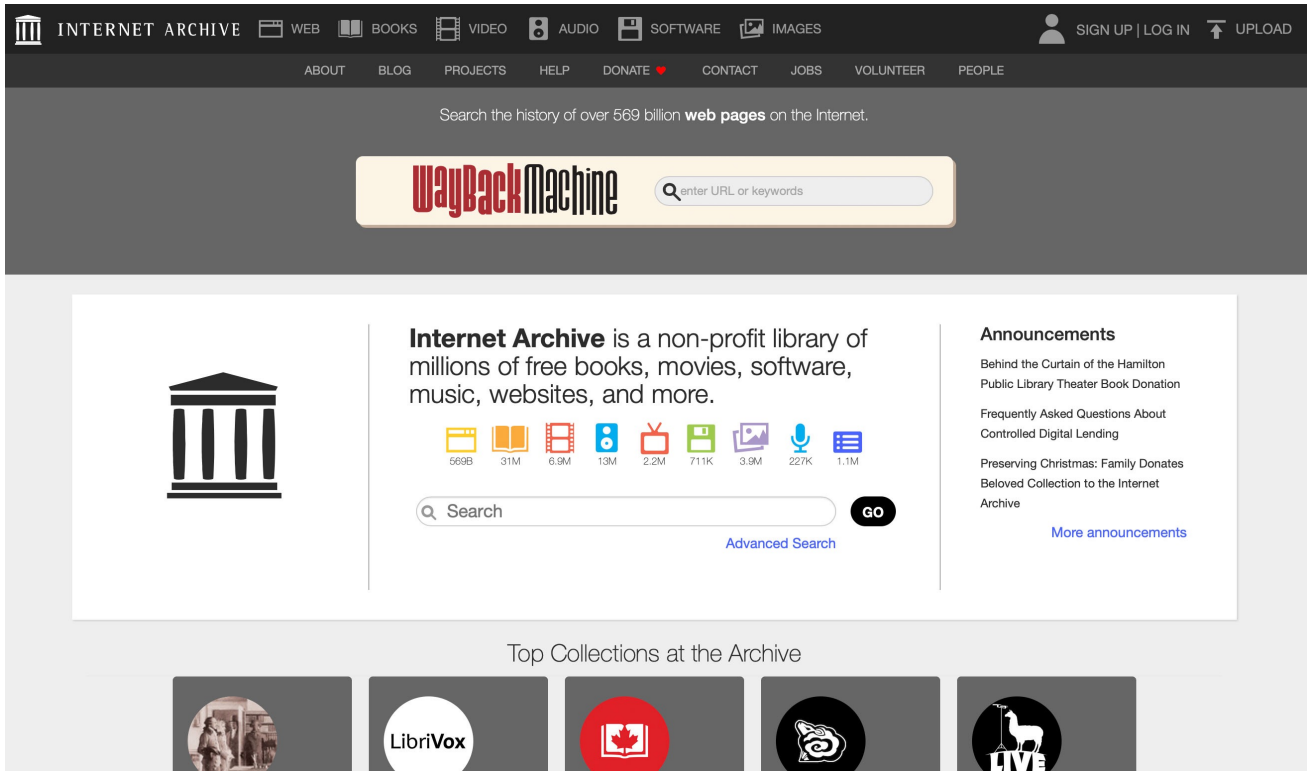
EUROPEANA (www.europeana.eu; 52,254,880 Works of art, collectors' items, books, films and pieces of music from European museums, archives and libraries)



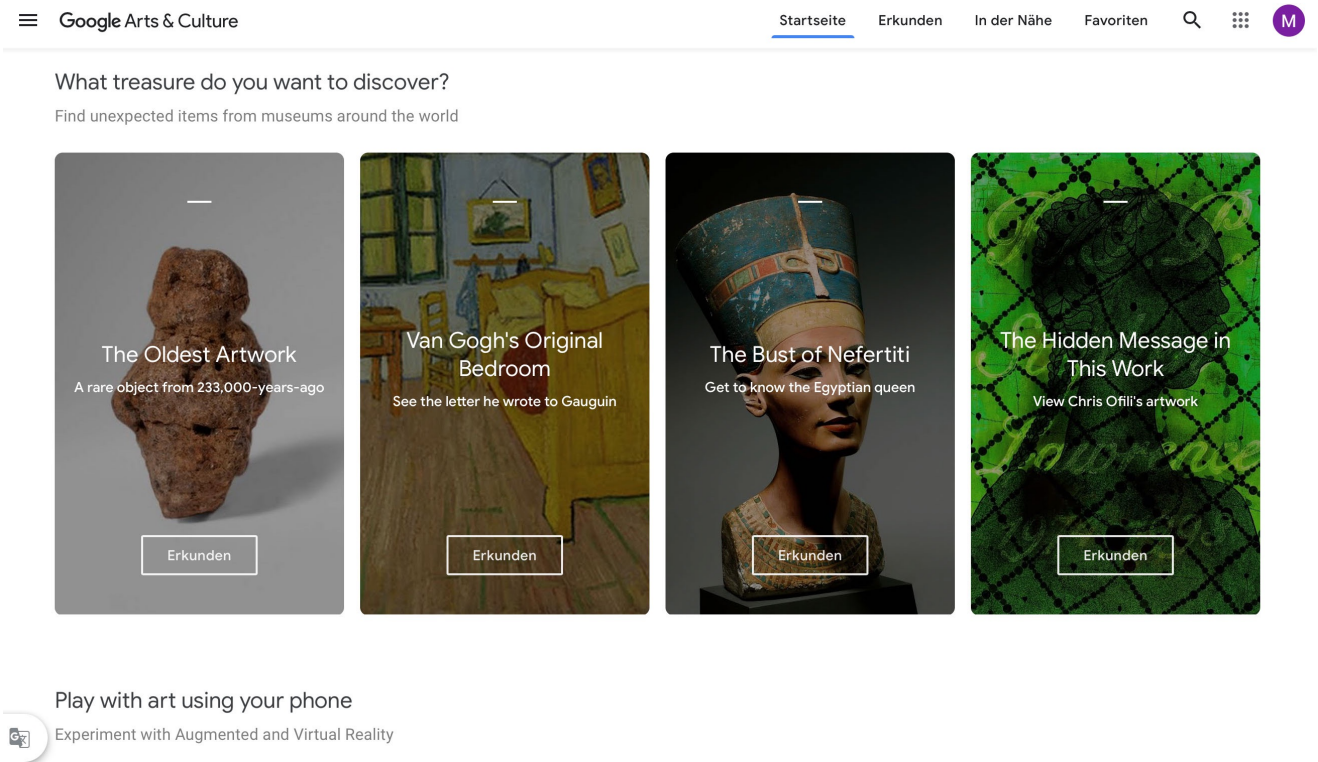
DIGITAL PUBLIC LIBRARY OF AMERICA
(https://dp.la; 43.942.039 images, texts, videos and sound files)



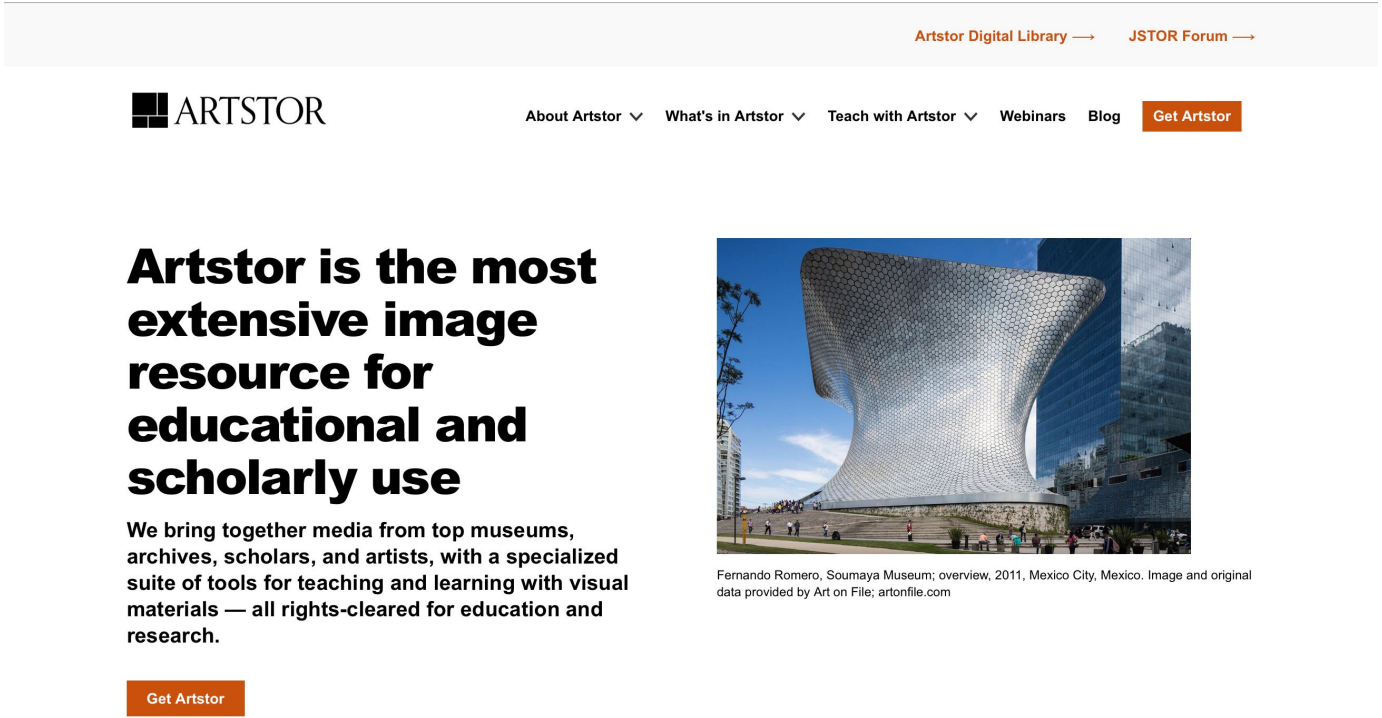
THE INTERNET ARCHIVE
(https://archive.org; with 569 Mio. digitised web pages since 1996)



GOOGLE ARTS & CULTURE (https://artsandculture.google.com)



ARTSTOR.ORG
(over 1.3 million works from public collections, 2.5 million for subscribers)



PROMETHEUS
(www.prometheus-bildarchiv.de)

FOTO MARBURG
(www.bildindex.de)

LARGE DATA COLLECTIONS ON THE INTERNET

s.a. https://en.wikipedia.org/wiki/Category:Digital_library_projects

https://en.wikipedia.org/wiki/Category:Discipline-oriented_digital_libraries

https://en.wikipedia.org/wiki/Category:Full_text_scholarly_online_databases

https://en.wikipedia.org/wiki/Category:Virtual_art_museums_and_galleries

https://en.wikipedia.org/wiki/Category:Web_archiving_initiatives

DATA COLLECTIONS ON MODERN HISTORY

Our World in Data (https://ourworldindata.org)

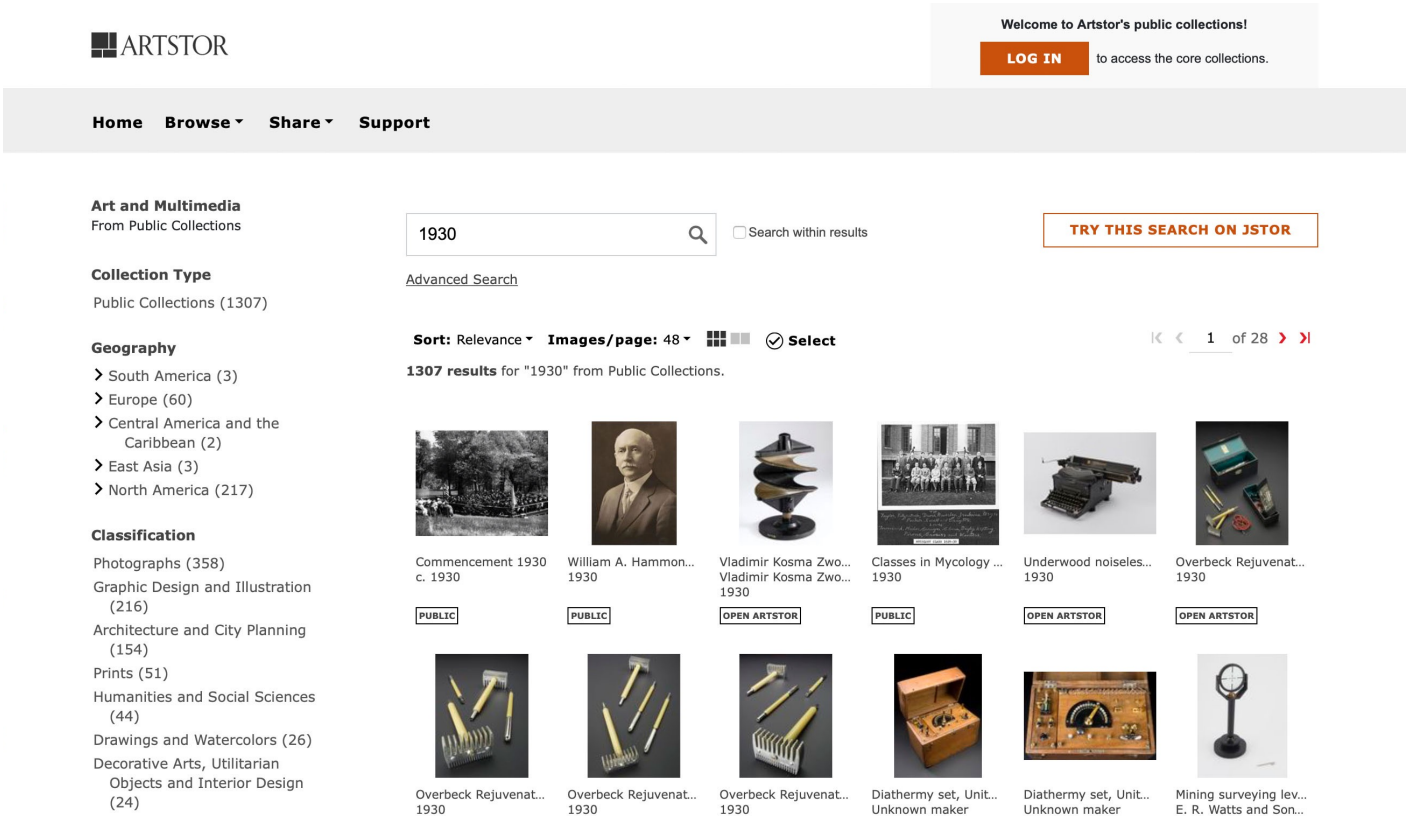
Histat: Zeitreihen zur Historischen Statistik (www.gesis.org/histat/de/index)

Deutschland in Daten: Zeitreihen zur historischen Statistik (www.deutschland-in-daten.de/datensatz)

Centre for Global Economic History Databases (www.cgeh.nl/data; mit Liste weiterer Quellen)

The ifo Prussian Economic History Database (https://www.ifo.de/iPEHD)

CANONISATION OF KNOWLEDGE



Lev Manovich, How to Follow Global Digital Cultures, or Cultural Analytics for Beginners, in: Konrad Becker, Felix Stalder (Hrsg.), Deep Search. The Politics of Search beyond Google (Innsbruck: Studien Verlag 2009) 198–211: http://manovich.net/index.php/projects/how-to-follow-global-digital-cultures

CANONISATION OF KNOWLEDGE

Astronomy
Galileo
Johannes Kepler
William Herschel
Pierre-Simon de Laplace
Nicolas Copernicus

Biology
Charles Darwin
Aristotle
Jean-Baptiste Lamarck
Georges Cuvier
Thomas Hunt Morgan

Chemistry
Antoine Lavoisier
Jöns Berzelius
Carl Scheele
Joseph Priestley
Humphrey Davy

Earth Sciences
Charles Lyell
James Hutton
William Smith
Agricola
Abraham Werner

Physics
Isaac Newton
Albert Einstein
Ernest Rutherford
Michael Faraday
Galileo

Mathematics
Leonhard Euler
Isaac Newton
Carl Gauss
Euclid
Pierre-Simon de Laplace

Medicine
Louis Pasteur
Robert Koch
Hippocrates
Galen
Paracelsus

Technology
Thomas Edison
James Watt
Leonardo da Vinci
Christiaan Huygens
Archimedes

Chinese Art
Gu Kaizhi
Zhao Mengfu
Wu Daozi
Mu Yuan
Dong Qichan

Chinese Literature
Du Fu
Li Bo
Bo Juyi
Su Dongpo
Han Yu

Chinese Philosophy
Confucius
Laozi
Zhu Xi
Mencius
Zhuangzi

Arabic Literature
al-Mutanabbi
Abu Nuwas
al-Ma'arri
Imru' al-Qays
Abu Tammam

Japanese Art
Toyo Sesshu
Tawaraya Sotatsu
Ogata Korin
Hasegawa Tohaku
Kano Eitoku

Japanese Literature
Matsuo Basho
Chikamatsu Monzaemon
Murasaki Shikibu
Ihara Saikaku
Mori Ogai

Indian Literature
Kalidasa
Vyasa
Valmiki
Asvaghosa
Bhartrhari

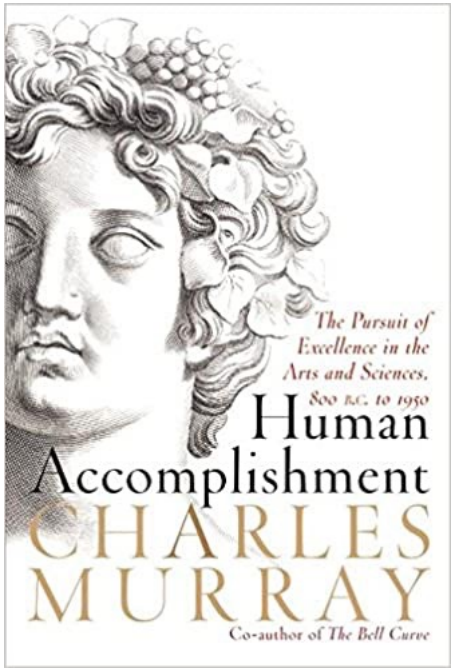
Indian Philosophy
Sankara
Nagarjuna
Ramanuja
Buddha
Madhva

Western Art
Michelangelo
Pablo Picasso
Raphael
Leonardo da Vinci
Titian

Western Literature
William Shakespeare
Johann von Goethe
Dante Alighieri
Virgil
Homer

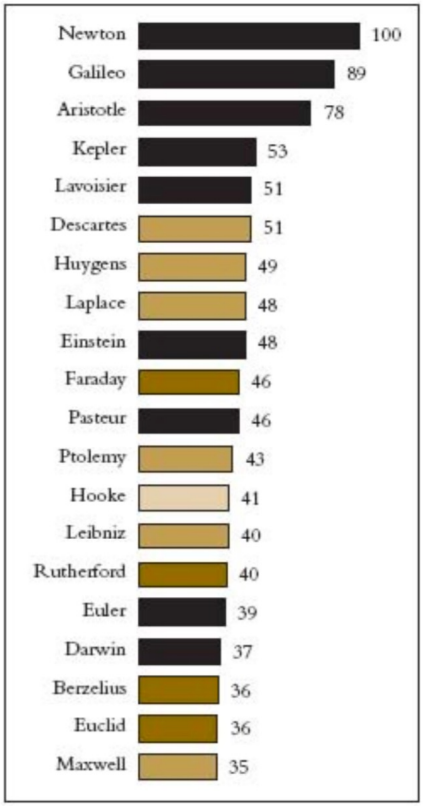
Western Music
Ludwig van Beethoven
Wolfgang Amadeus Mozart
Johann Sebastian Bach
Richard Wagner
Franz Joseph Haydn

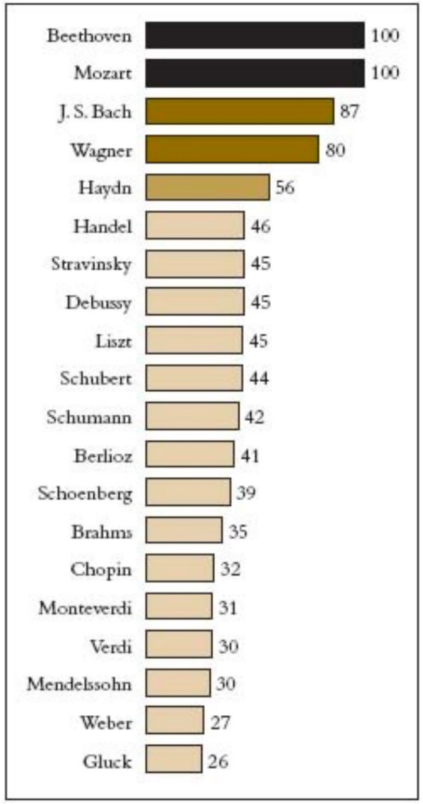
Western Philosophy
Aristotle
Plato
Immanuel Kant
Rene Descartes
Georg Hegel

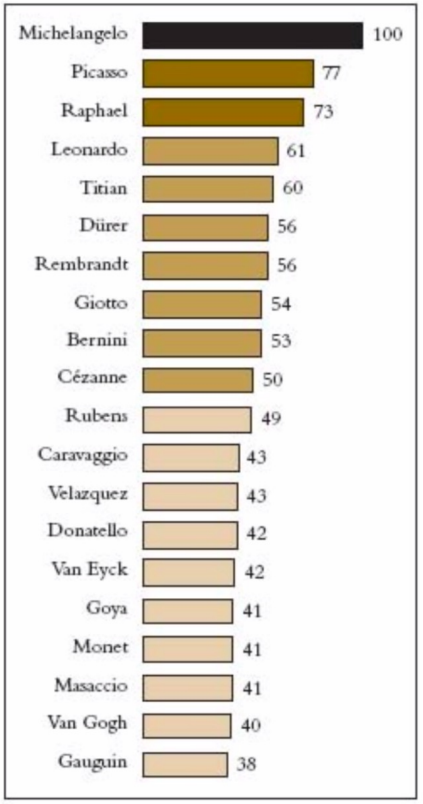


Charles Murray, Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 B.C. to 1950 (HarperCollins, 2003)

CANONISATION OF KNOWLEDGE



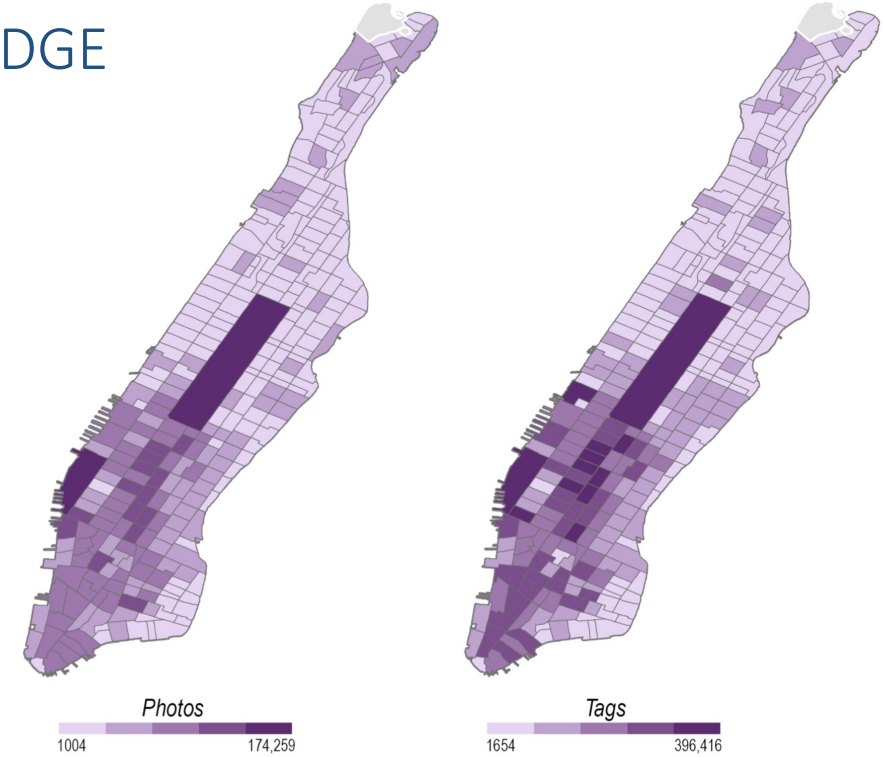




Charles Murray, Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 B.C. to 1950 (HarperCollins, 2003)

CANONISATION OF KNOWLEDGE

https://fastenurseatbelts.com/de/13-instagram-spots-in-new-york/



Agustin Indaco, Lev Manovich, Urban Social Media Inequality. Definition, Measurements, and Application, arXiv.org, July 7, 2016: https://arxiv.org/abs/1607.01845

europeana

20th century
Explore collections and stories from 20th century history and culture

Animals
Boost your knowledge of the animal world, history and culture

Art Nouveau
Explore Europeana's extensive archive of Art Nouveau stories

Architecture
Discover the wonders of European architecture

Asian art & heritage
Explore stories and objects from Asia in European collections

Black history
Explore the histories of Black people in Europe

Chinese heritage
Explore stories and objects from China's rich history in European collections

Colouring Books
Bring culture to life with Europeana's range of colouring books

Colours
Bring a splash of colour into your life with these curated colours

Composers
Meet the people who shaped Europe's musical history and culture

Discovering Europe
Take a virtual journey through Europe from the comfort of your home

Diversity & Inclusion
Find resources to learn and teach about inclusion and diversity

Dragons, myths and legends
Know more about legendary creatures in Europe

Environment
Stories from Europeana about the environment, the climate, and nature

European identities
Discover curated content about European identities, cultures and history

Food and drink
Explore European cultural heritage through food and drink

Colourin

Bring cultu
European
colouring

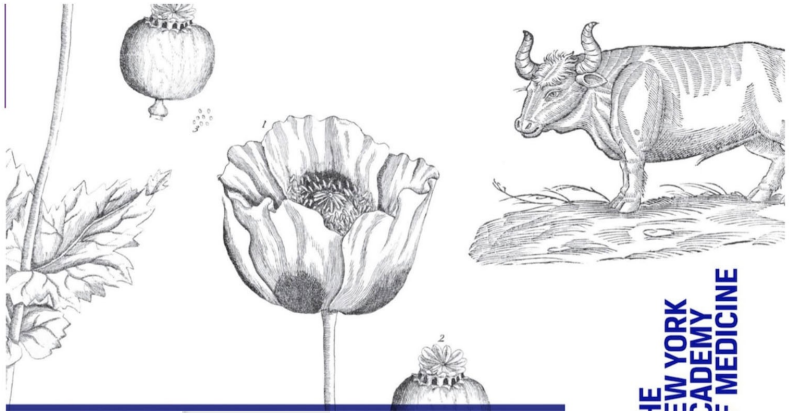
MY MODERN MET
SHARE | INSPIRE | CONNECT

POPULAR MORE TOPICS ENGLISH SEARCH

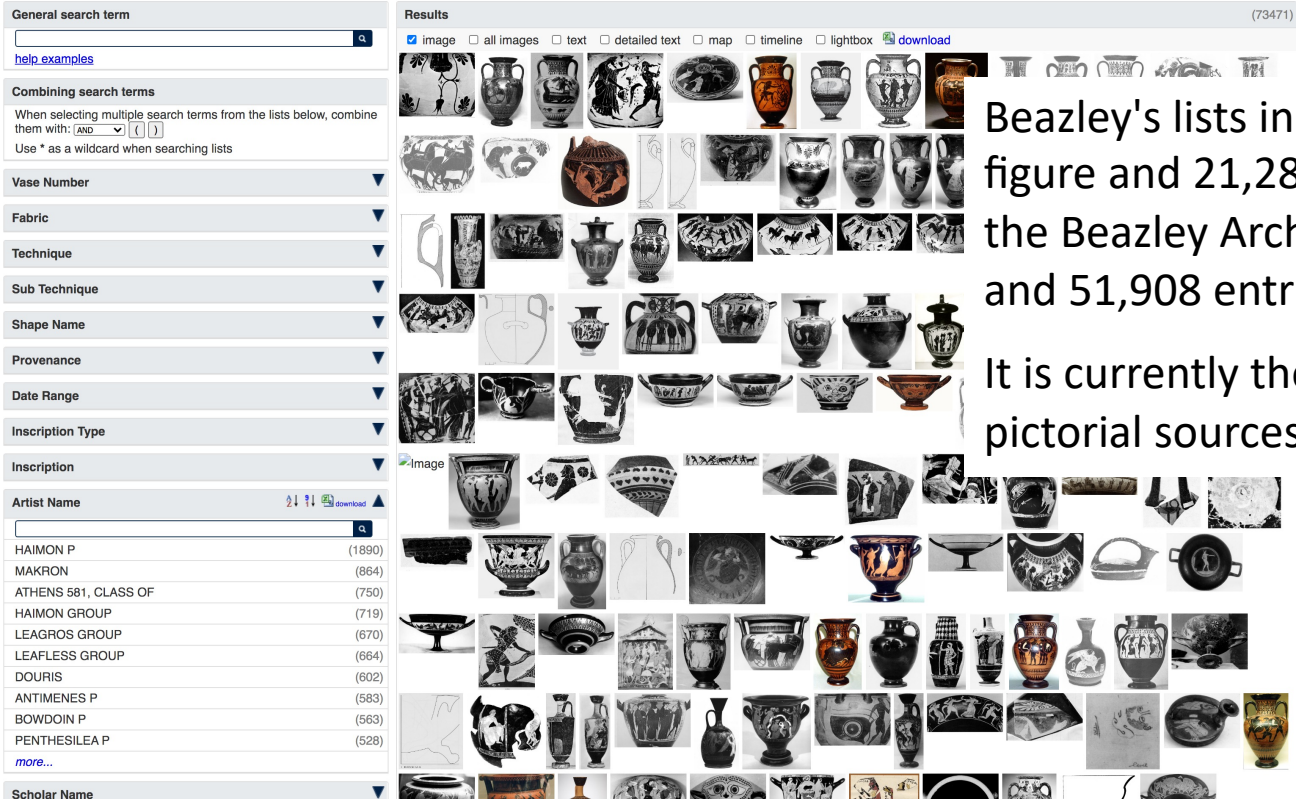
HOME / ART

You Can Download Thousands of Coloring Book Pages From Museum Collections

By Megan Cooper on April 9, 2020



BEAZLEY ARCHIVE POTTERY DATABASE



Beazley's lists include 12,786 Attic black-figure and 21,286 Attic red-figure vessels, the Beazley Archives database has 42,265 and 51,908 entries (as of May 2021).

It is currently the largest database on pictorial sources from Greek antiquity.

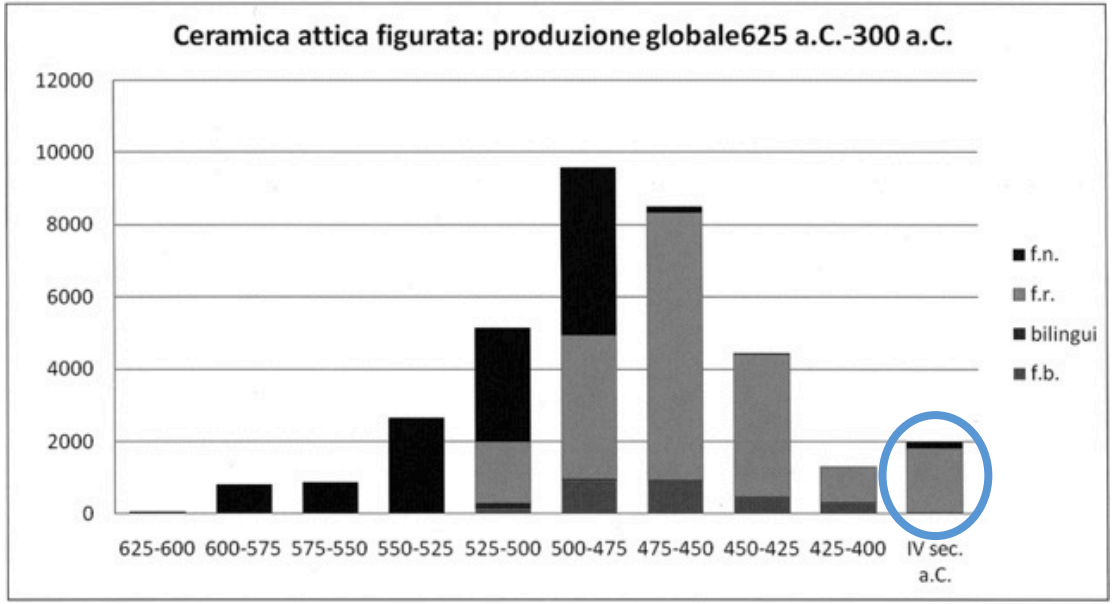
www.beazley.ox.ac.uk/XDB

DATA BASIS

Beazley (and the researchers after him) was not interested in the 4th century.

Again, no Gaussian distribution curve, but the art-historical model of rise, bloom and decline.

Thus, at www.vasenrepertorium.de you'll find 12,131 att. vases of the 4th century BC.

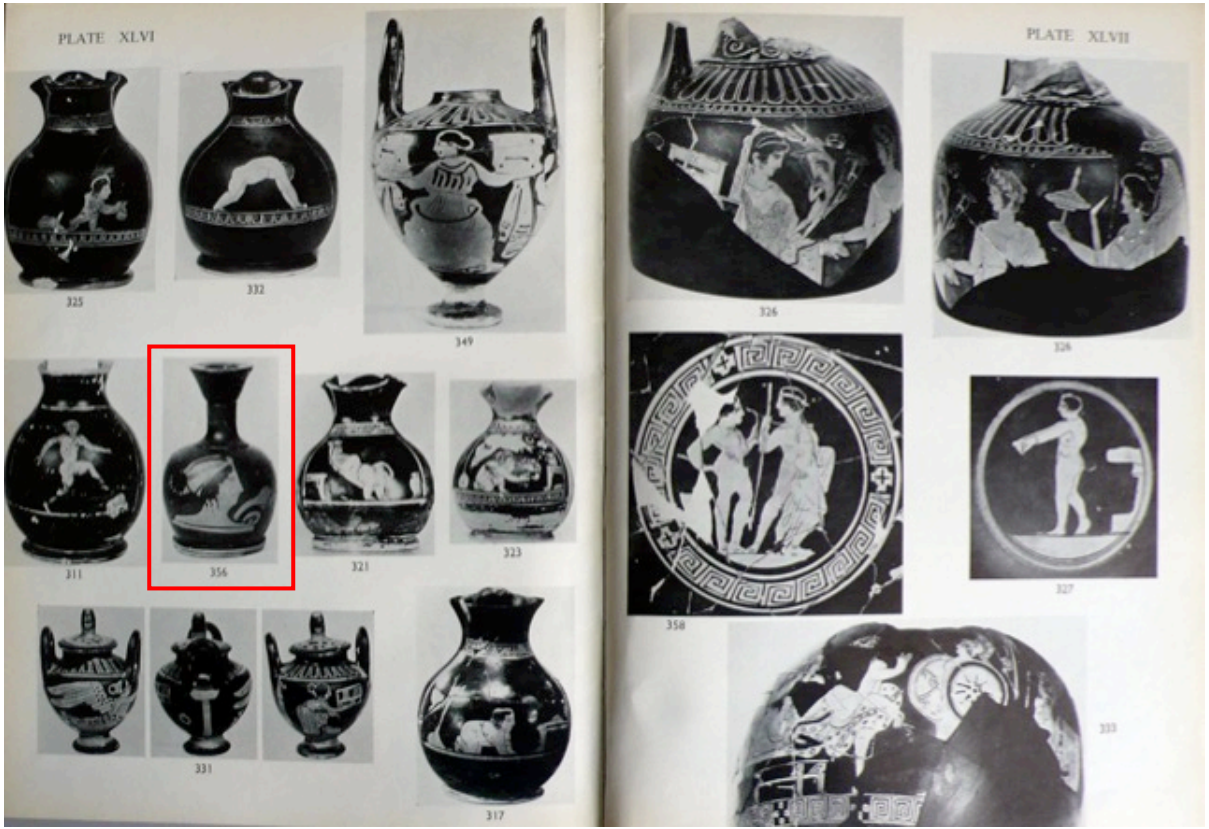


F. Giudice – R. Scicolone – S.L. Tata, Vedere il vaso attico. Costruzione del quadro di riferimento delle forme dal 635 al 300 a.C., in: S. Schmidt – A. Stähli (Hrsg.), Vasenbilder im Kulturtransfer. Zirkulation und Rezeption griechischer Keramik im Mittelmeerraum (München 2012) 27–34 Taf 1,1

DATA BASIS

Beazley only recorded what he could assign to a painter or a workshop.

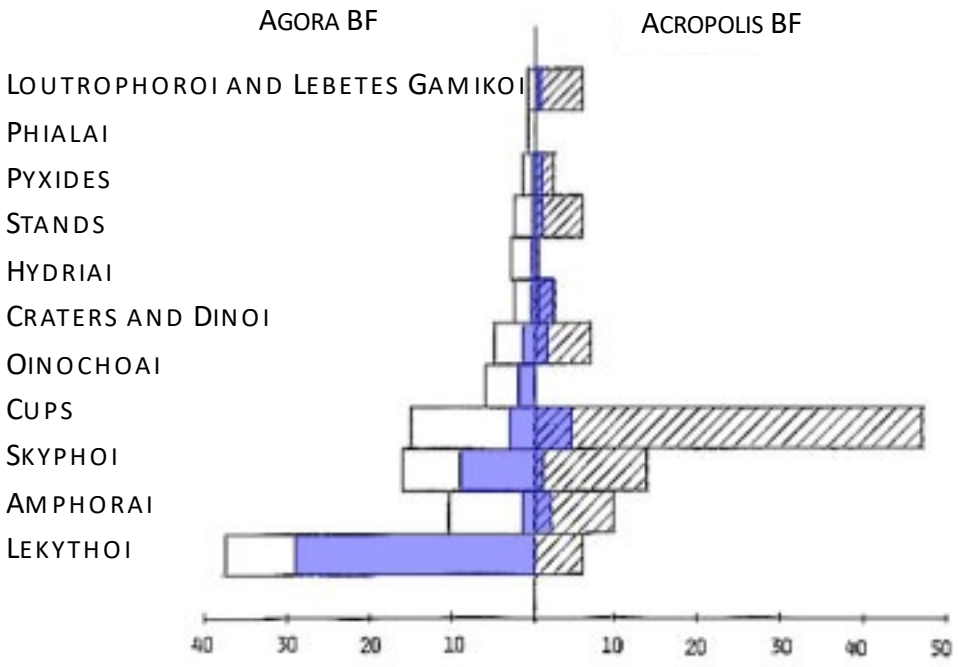
The recording criteria of the data basis are essential for the evaluation!



Ashmolean Museum, Select Exhibition of Sir John and Lady Beazley's Gifts 1912-1966 (Oxford 1967)

DATA BASIS

Beazley did not record a proportionally constant share compared to the known amount of found pottery from the Agora and Acropolis of Athens. His sample is therefore not suitable for making statements about the distribution of Attic pottery.



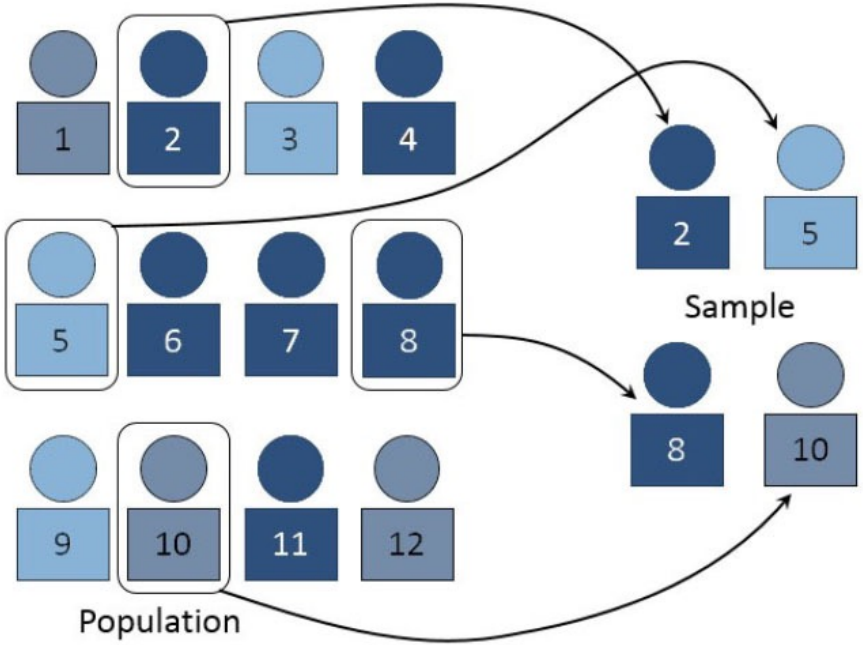
L. Hannestad, in: Ancient Greek and Related Pottery (Kopenhagen 1988) 224–227

SAMPLING

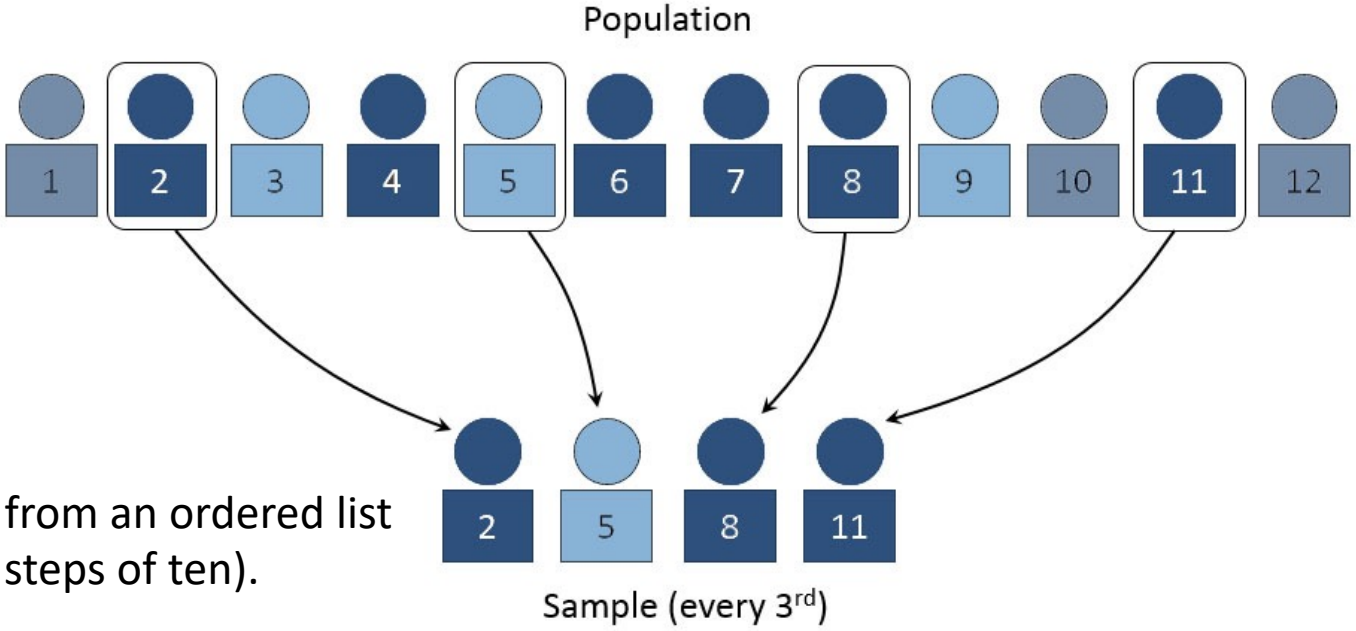
is a defined method for selecting data for a statistical investigation in such a way that analyses on these data allow conclusions to be drawn about the population without systematic error.

SIMPLE RANDOM SAMPLING

- Every entity and every subset in the population has an equal chance of being selected.
- Small samples may result in an unrepresentative selection.



SYSTEMATIC SAMPLING



- Elements are selected from an ordered list at regular intervals (e.g. steps of ten).
- Small samples may result in an unrepresentative selection.

SYSTEMATIC SAMPLING

- particularly susceptible to periodicities in the list that systematically produce errors

Outline of the Iconclass system

The 10 main categories - click one to start browsing and searching

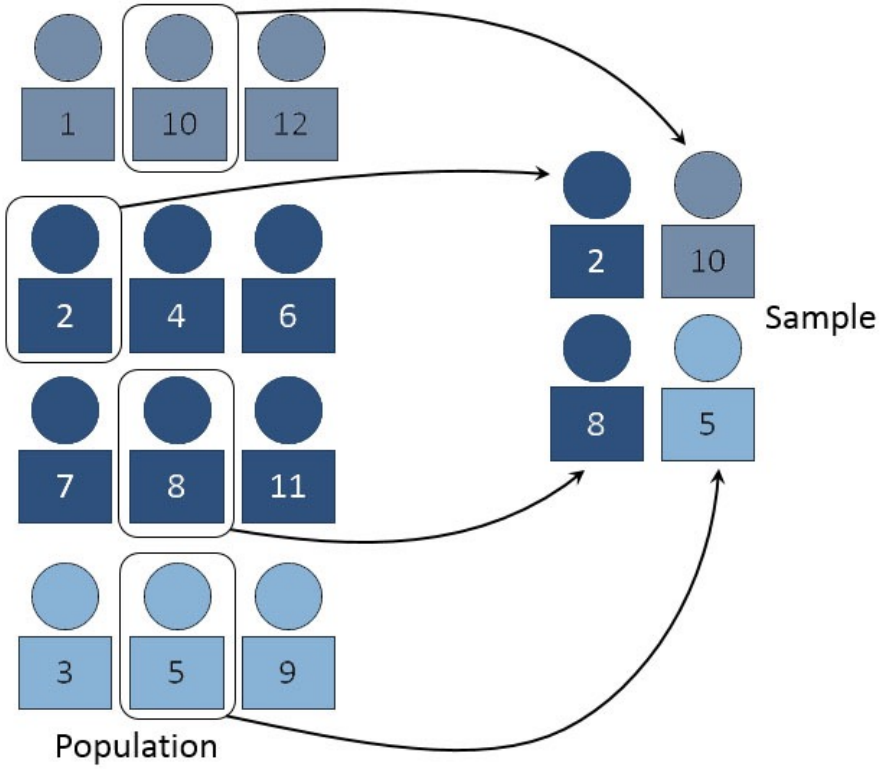
0 · Abstract, Non-representational Art
1 · Religion and Magic
2 · Nature
3 · Human Being, Man in General
4 · Society, Civilization, Culture
5 · Abstract Ideas and Concepts
6 · History
7 · Bible
8 · Literature
9 · Classical Mythology and Ancient History

http://www.iconclass.org/help/outline

STRATIFIED SAMPLING

- Equal numbers of individuals are randomly drawn from previously determined subpopulations.

www.akpool.de



STRATIFIED SAMPLING

- Statements about subgroups possible
- Particularly suitable if elements are better documented in one sub-area than in another





STRATIFIED SAMPLING

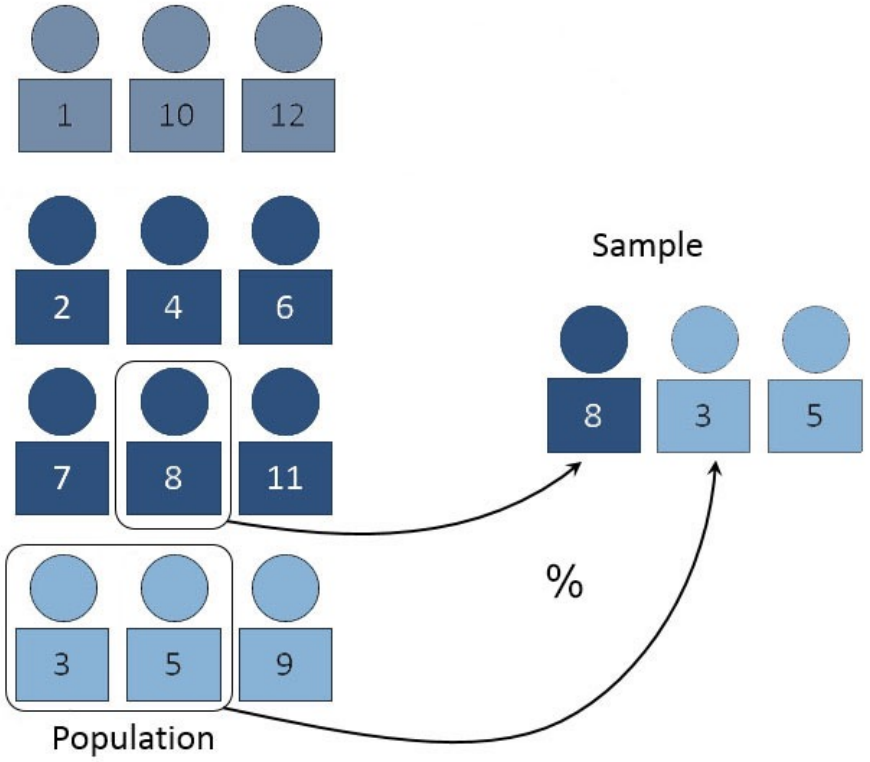
Advantages:

- Variability within layers is minimised.
- Variability between strata is maximised.
- The variables by which the population is stratified are strongly correlated with the desired dependent variable.

- only make sense if there are homogeneous subgroups.

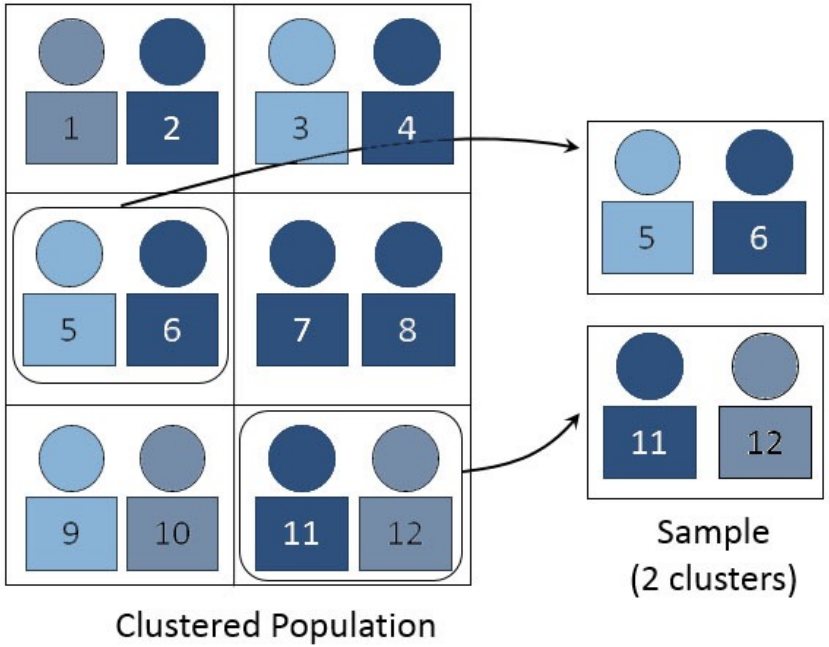
QUOTA SAMPLING

- The population is first divided into mutually exclusive subgroups in order to then select a certain proportion ("quota") of units from each segment.
- is not a probability sample, because the selection of the sample is not random

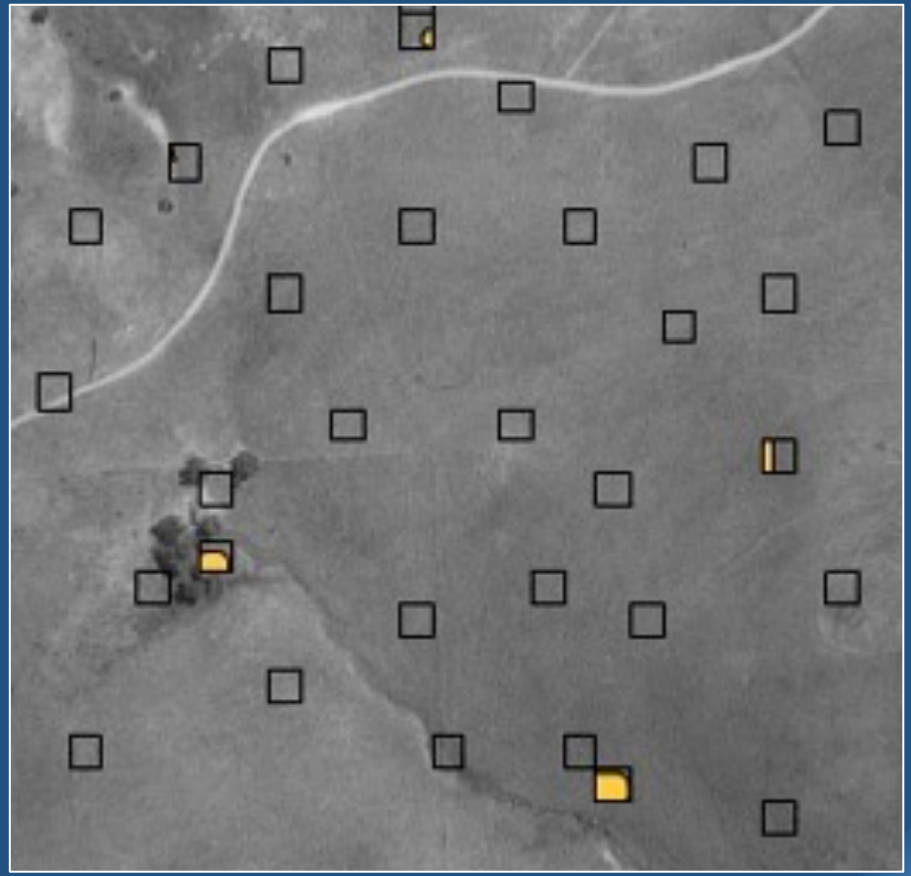


CLUSTER SAMPLING

- Selection by groups ("clusters") clustered by geography or by time period.
- requires a larger sample than simple random sampling because of the higher variance.



SAMPLING: EXAMPLES



GEODATA

Information about the settlement structure of past times can be gained from

- above-ground remains of settlement structures

- Picking up surface finds (survey)

- cleaning and smaller excavations

- geodetic or geomagnetic methods

- large-scale excavations

Luftbild von West-Montana (USA)

www.utexas.edu/courses/denbow/labs/survey.htm

RESEARCH ON REPRESENTATIVE SAMPLES

Hypothetical example: Landscape in Western Montana (USA)

Division of the area into 27 x 37 (= 999) quadrants

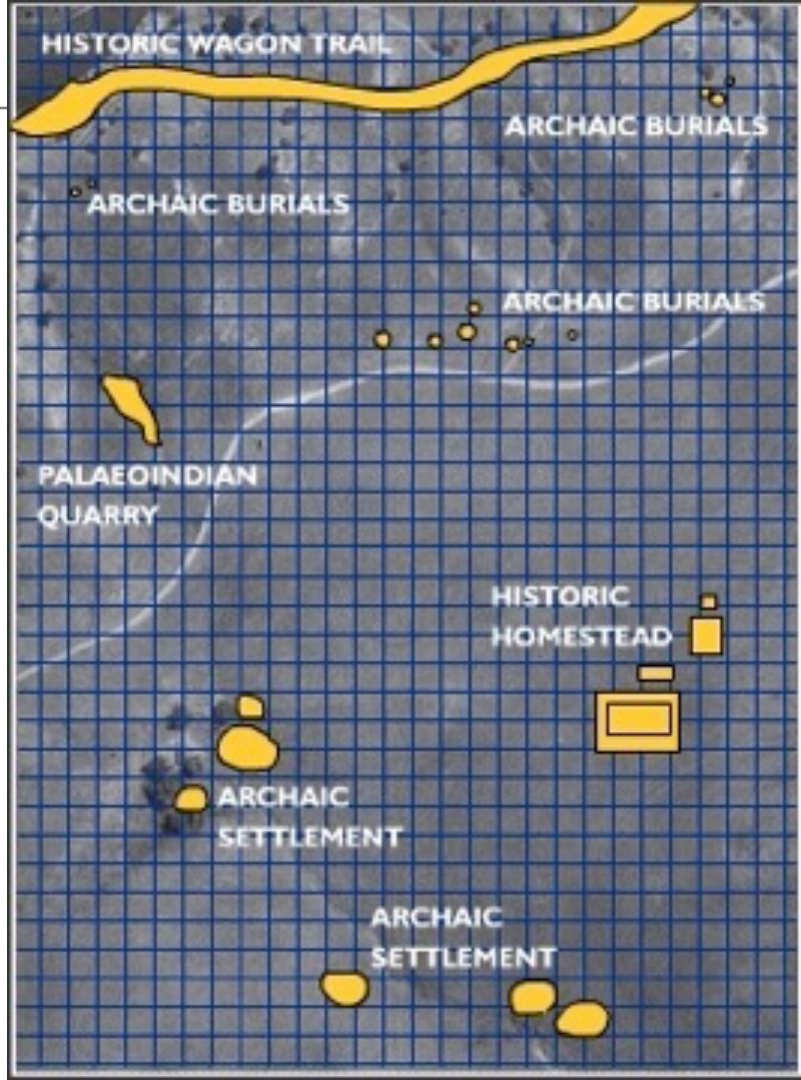


Illustration of all structures

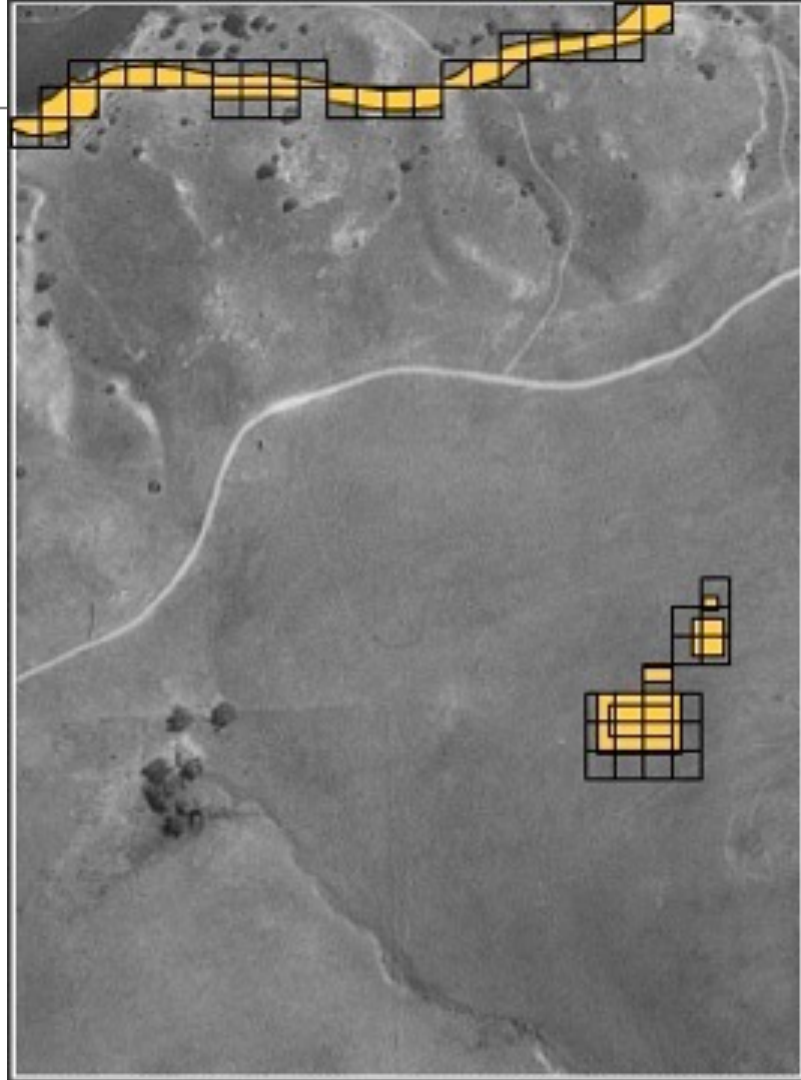
www.utexas.edu/courses/denbow/labs/survey.htm

TARGETED SELECTION OF THE STUDY AREA (NON-PROBABILISTIC SAMPLING)

is used when one is only interested in certain structures that are already known.

i.e. the six prehistoric structures remain unknown in this way.

- Not a representative sample



Non-probabilistic sampling

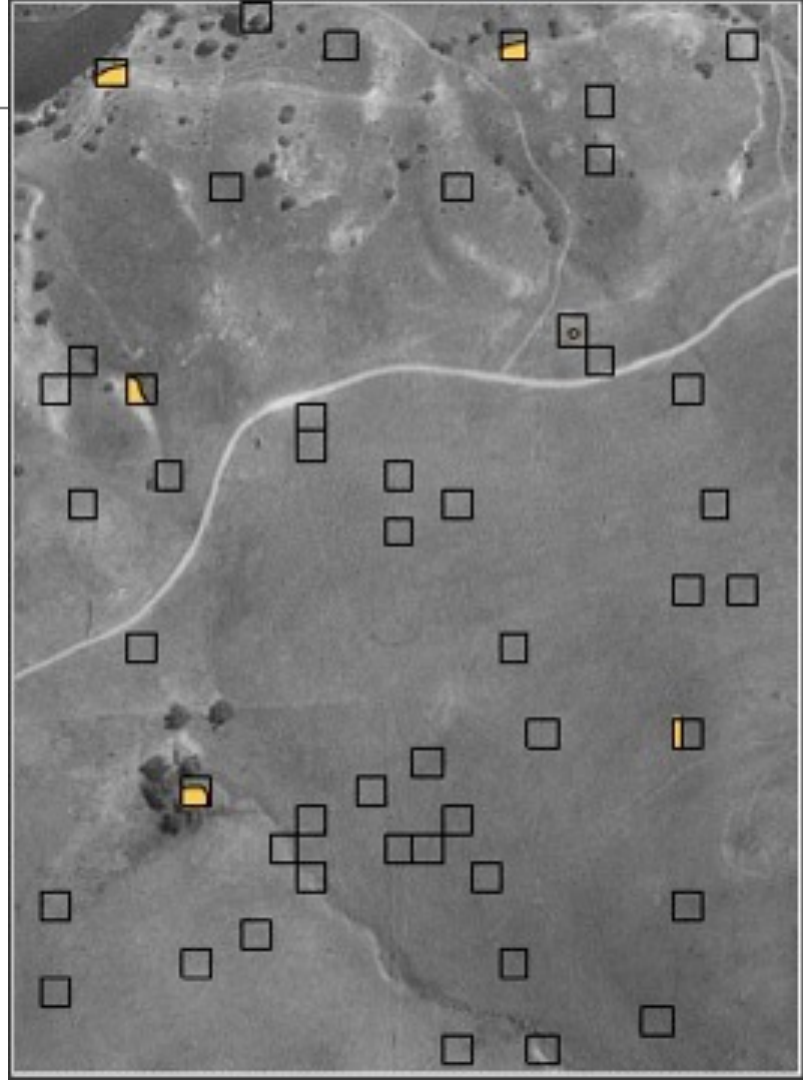
www.utexas.edu/courses/denbow/labs/survey.htm

SIMPLE RANDOM SAMPLE

Probabilistic sampling uses statistical methods to study only representative areas of an area.

Here, only 5% of randomly selected quadrants were investigated, which means that six of the eight sites are covered.

However, larger areas remain unexplored.

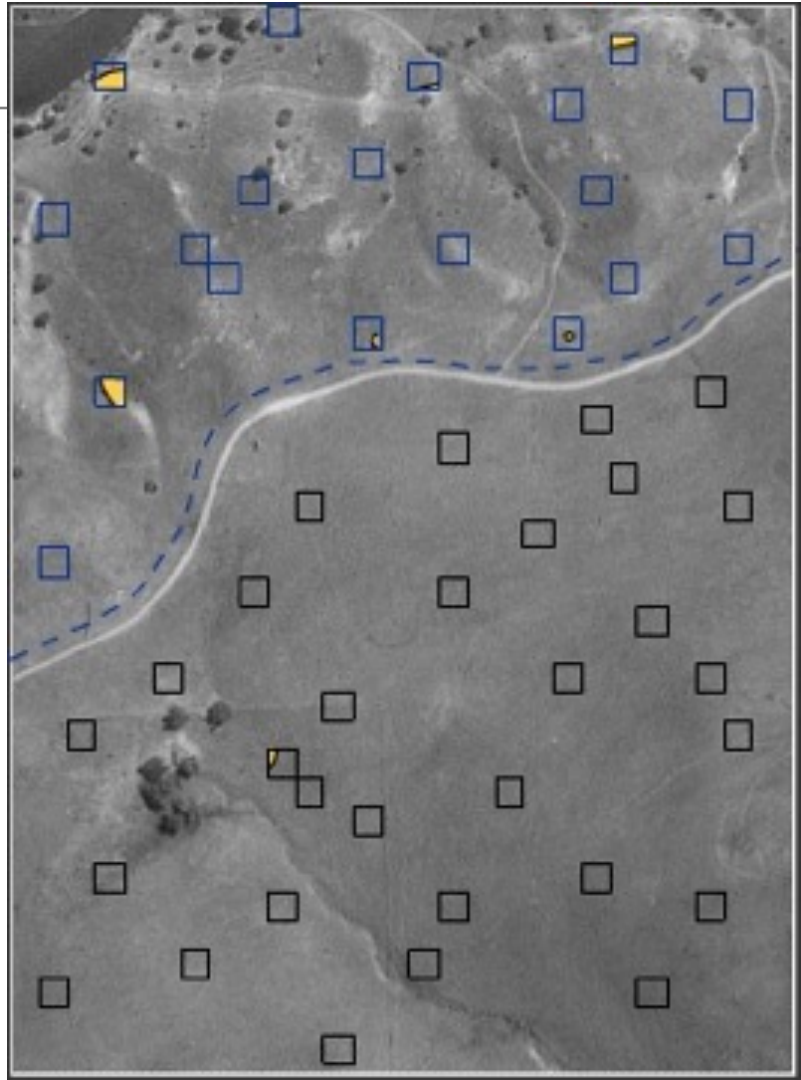


Simple random sampling

www.utexas.edu/courses/denbow/labs/survey.htm

STRATIFIED RANDOM SAMPLING

The study area is first roughly divided into different, topographically distinguishable areas, within which quadrants are again randomly selected.



Stratified random sampling

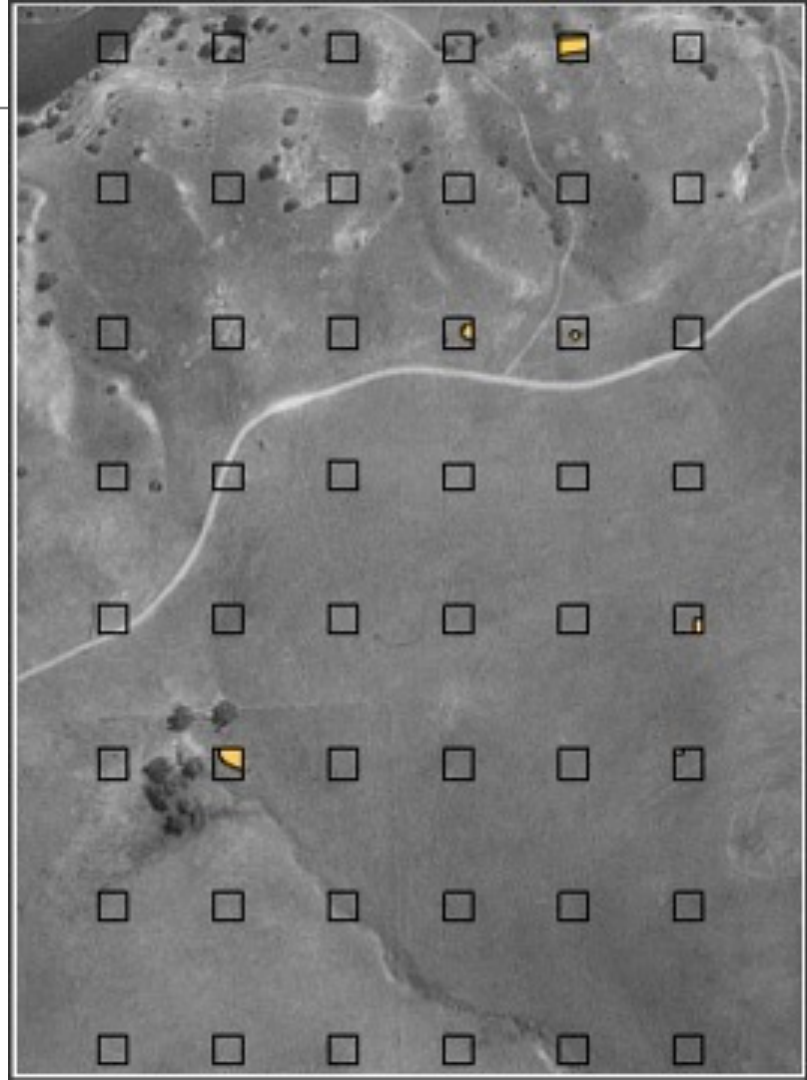
www.utexas.edu/courses/denbow/labs/survey.htm

SYSTEMATIC SAMPLING

The units are evenly distributed over the study area.

- No large areas that are not acquired

- Already known structures or areas of special interest (such as valleys or hills) may indicate settlement forms that could be investigated on a smaller scale.



Systematic Sampling

www.utexas.edu/courses/denbow/labs/survey.htm

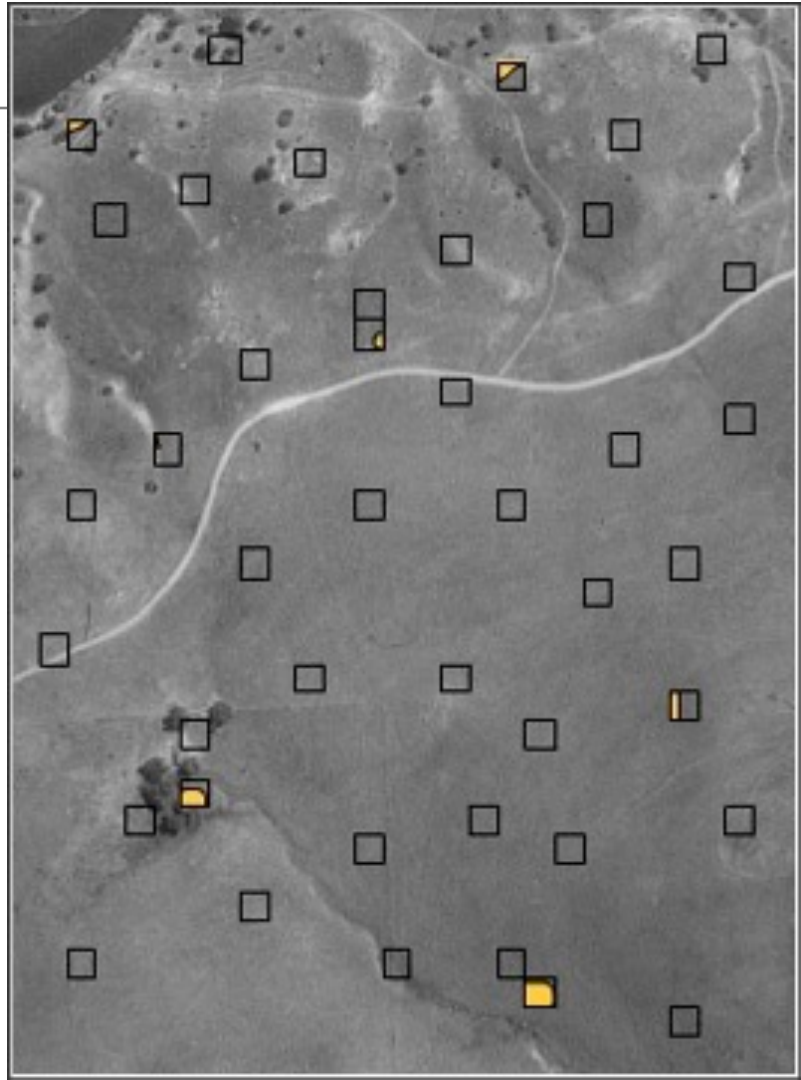
COMBINATION OF STRATIFIED AND SYSTEMATIC SAMPLING

The study area is divided into smaller, topographically distinguishable areas, within which quadrants are randomly selected.

- Each of the samples mentioned can be useful.

- None has any prospect of complete acquisition of the existing structures.

- Large-scale structures are more likely to be acquired.



Systematic Unaligned Sampling

EXCAVATIONS AS A RANDOM SAMPLE

In archaeology, the amount of what is preserved is usually considered a random sample.

Example: The American excavations at the Agora of Athens

Deposits of the 5th and 4th century BC with red-figure pottery

POTTERY FINDS

Example: Pelikai from the settlement excavation in Olbia

O. E. Buravčuk, Red-figured pottery, in: N. A. Lejunskaja u.a., The Lower City of Olbia (Sector NGS) in the 6th century BC to the 4th century AD (Aarhus 2010) 171–184 Taf. 88–89

POTTERY FINDS

Only fragments that can be clearly assigned to a motif are suitable for iconographic questions.







POTTERY FINDS

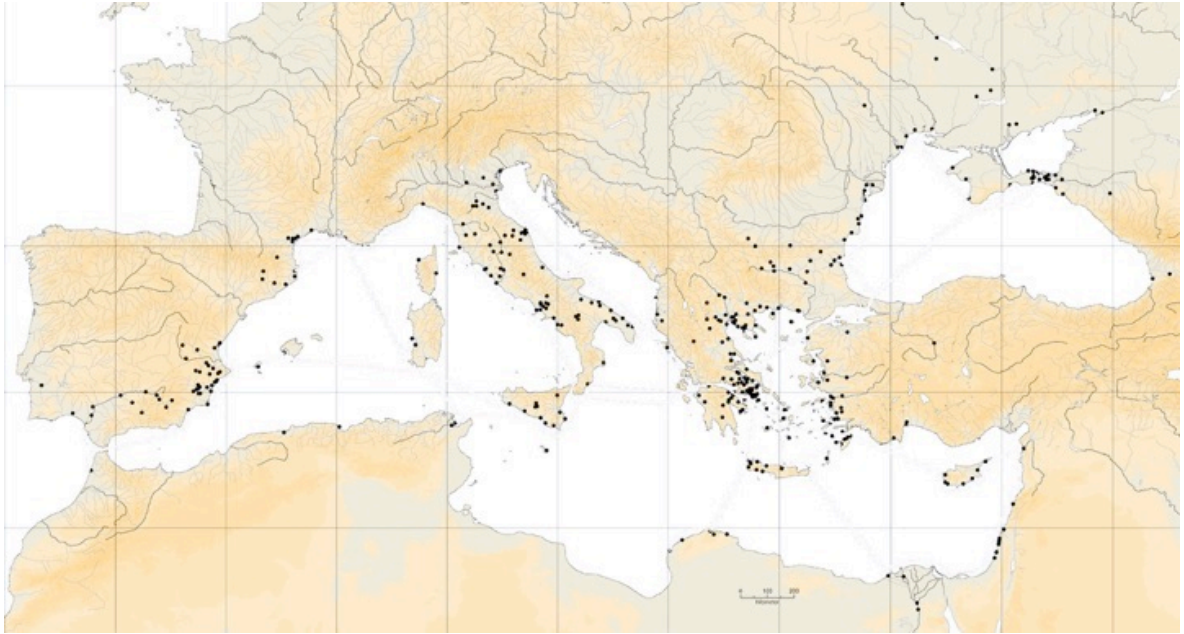
Example: Pelikai from the settlement excavation in Olbia

O. E. Buravčuk, Red-figured pottery, in: N. A. Lejunskaja u.a., The Lower City of Olbia (Sector NGS) in the 6th century BC to the 4th century AD (Aarhus 2010) 171–184 Taf. 88–89

M. B. Moore, Attic Red-Figured and White-Ground Pottery. Agora 30 (Princeton 1997) Taf. 12–13

CONCLUSION

A stratified sample can minimise the imbalance caused by the selective publication of the pieces if the same inclusion criteria apply to all sites and contexts. Then, as with random sampling, each element of the population has the same chance of being included in the sample.



Significant sites of late red-figure vases from Athens

Test for representativeness with regard to the publication situation:

Growth over the last 25 years

Our sample contains 3363 entries that were submitted for the first time in the last 25 years, i.e. between 1986 and 2010. This corresponds to 35.9% of the total material. With regard to the distribution among vessel forms, this increase leads to minor deviations, which are less than 2%.

Gesamt		bis 1985	1986–2010
1053 (11,4 %)	Pelike	659 (11 %)	398 (11,8 %)
52 (,6 %)	Lutrophoros	44 (,7 %)	14 (,4 %)
38 (,4 %)	Lebes Gamikos	19 (,3 %)	21 (,6 %)
507 (5,5 %)	Kelchkrater	353 (5,9 %)	159 (4,7 %)
1683 (18,2 %)	Glockenkrater	1101 (18,4 %)	608 (18,1 %)
331 (3,6 %)	Hydria	257 (4,3 %)	77 (2,3 %)
436 (4,7 %)	Oinochoe Typ 2	350 (5,8 %)	87 (2,6 %)
383 (4,2 %)	Oinochoe Typ 3 (Chous)	286 (4,8 %)	97 (2,9 %)
67 (,7 %)	sonst. Kannen	64 (1,1 %)	5 (,1 %)
983 (10,7 %)	Bauchlekythos	606 (10,1 %)	378 (11,2 %)
460 (5 %)	Askos	253 (4,2 %)	207 (6,2 %)
72 (,8 %)	Pyxis Typ A	54 (,9 %)	19 (,6 %)
23 (,2 %)	Pyxis Typ B	11 (,2 %)	12 (,4 %)
107 (1,2 %)	Pyxis Typ C	73 (1,2 %)	34 (1 %)
111 (1,2 %)	Pyxis Typ D	66 (1,1 %)	45 (1,3 %)
641 (6,9 %)	Lekanis	371 (6,2 %)	277 (8,2 %)
74 (,8 %)	Teller	61 (1 %)	13 (,4 %)
228 (2,5 %)	Fischteller	143 (2,4 %)	85 (2,5 %)
607 (6,6 %)	Skyphos	347 (5,8 %)	261 (7,8 %)
281 (3 %)	Schalenskyphos	112 (1,9 %)	169 (5 %)
305 (3,3 %)	Kylix (fußlos)	186 (3,1 %)	143 (4,3 %)
442 (4,8 %)	Kylix	376 (6,3 %)	111 (3,3 %)
62 (,7 %)	Sonstige	50 (,8 %)	12 (,4 %)
9228 (100 %)	Gesamt	6000 (100 %)	3363 (100 %)

RESULT ACCURANCY

denotes the accuracy of a data basis with regard to a change in the data basis by a certain number of data.

Example: 400 E(3)=0.75

The result accuracy E(3) below the total number indicates how much the percentage distribution of the data would change if one value was unilaterally increased by three.

The result accuracy refers to a fundamental statistical inaccuracy, which is particularly noticeable with small amounts of data, while the confidence interval expresses the estimated representativeness and maximum deviation of the data, which cannot be determined exactly in archaeological evaluations anyway.

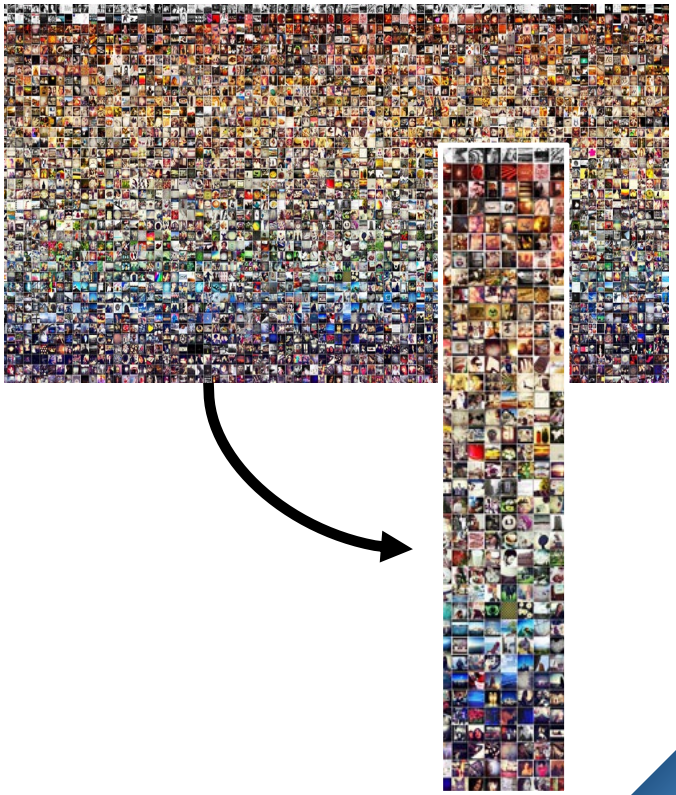
ANALYSES ON THE POPULATION

- e.g. on the basis of complete surveys (censuses, inventories, corpora and catalogues raisonnés) or born-digital content (Instagram, Twitter, youtube, TV etc.)
- apart from checking completeness, no preliminary investigations necessary
- Calculations of location and dispersion measures (especially for big data) are useful.

https://1.bp.blogspot.com/-6N6jB-NaPg4/VwR67VjiL7I/AAAAAAAAAnyk/WQ-VFpJKrI8z0Q_k4kK_uFV3Om8V8O4BA/s1600/01_734x615.png

ANALYSES ON REPRESENTATIVE SAMPLES

- e.g. surveys, image sets, geo-surveys etc., i.e. a proportion selected according to uniform rules of all works available in a certain medium, time or place.
- Preliminary research on the composition of the population necessary.
- Sampling (composition, size) only makes sense depending on the characteristics to be investigated, because every reduction of the data basis bears the risk of excluding something important
- Calculations of location and dispersion measures possible if necessary

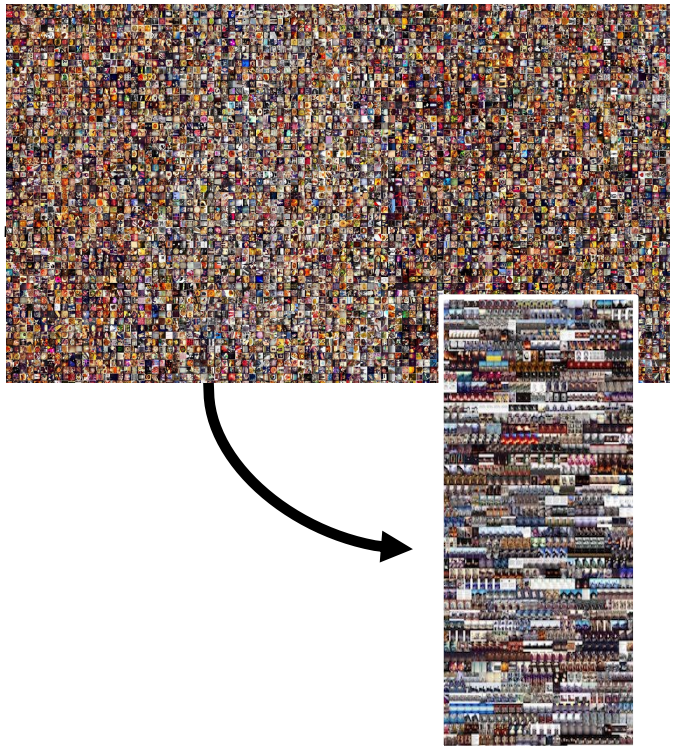


https://phototrails.net/wp-content/uploads/2013/06/homepage-img_1.jpg

ANALYSES ON INDETERMINATE SUBSETS OF THE POPULATION

(depending on the preservation for historical periods)

- e.g. internet portals and databases, collection catalogues, excavation finds etc.
- Knowledge about the origin of the data collection necessary
- Re-sampling (composition, size) necessary depending on the characteristics to be investigated
- Calculations of location and scattering measures will possibly mask individual phenomena



https://i.pinimg.com/originals/7b/8b/72/7b8b7296c18dabd98c7bae49cb736e17.jpg

HOW REPRESENTATIVE ARE SURVEY EXHIBITIONS?

Archive

28 February until 20 September 2020

THE POSTER

200 YEARS OF ART AND HISTORY

With nearly 400 exhibits by around 200 artists and designers, the exhibition The Poster at the Museum für Kunst und Gewerbe Hamburg (MKG) offers a large-scale, representative overview of the history of the poster from its beginnings in the early nineteenth century to today. Art and history, design and advertising meet in this medium.

https://www.mkg-hamburg.de/en/exhibitions/archive/2020/the-poster.html

COLLECTING INTERESTS AND FOCI:

- 1880s: "Examples of lithographic colour printing".
- 1915: Printed matter on the occasion of the war
- 1964: Exhibition of the Alliance Graphique International (graphic design)
- Exhibitions of posters by individual designers
- Private collection of artists' posters
- 1990s: thematic exhibitions

MK&G MUSEUM FÜR KUNST UND GEWERBE HAMBURG

PRESSE | DAS MK&G | NEWSLETTER

NEUES | AUSSTELLUNGEN | SAMMLUNG | BESUCH | KALENDER | VER

Sammlung › Sammlungen › Plakat

Plakat

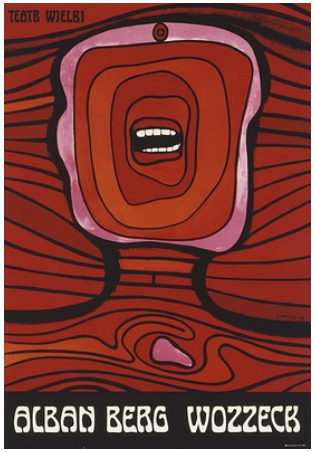
Seit dem 19. Jahrhundert gehört das Plakat zu den führenden Bild-Medien unserer Welt. Von den Werken eines Toulouse-Lautrec bis zu den Plakaten der italienischen Modemarke Benetton aus den 1990er Jahren entstanden hier viele der markantesten Bilder ihrer Epoche. Die meisten Plakate sind viel seltener als man denken mag – denn hängen sie einmal an der Wand, so werden sie nach zehn Tagen überklebt und sind für die Nachwelt verloren. Eine Plakatsammlung hat also vor allem die Aufgabe zu bewahren. Darüber hinaus geht es aber auch um die Auswahl, um das Hervorheben besonders gelungener Beispiele. Nur so lässt sich eine Geschichte des Plakates darstellen. Die Plakatsammlung des MKG ist – auch im internationalen Vergleich – eine der ältesten

https://www.mkg-hamburg.de/de/sammlung/sammlungen/plakat.html

HOW REPRESENTATIVE ARE SURVEY EXHIBITIONS?

Sample (400 posters)
- too small for 250 years
- weighted with regard to certain regions (France, Russia, Germany) and phases (around 1890, 1930, 1970)
- weighted with regard to important artists

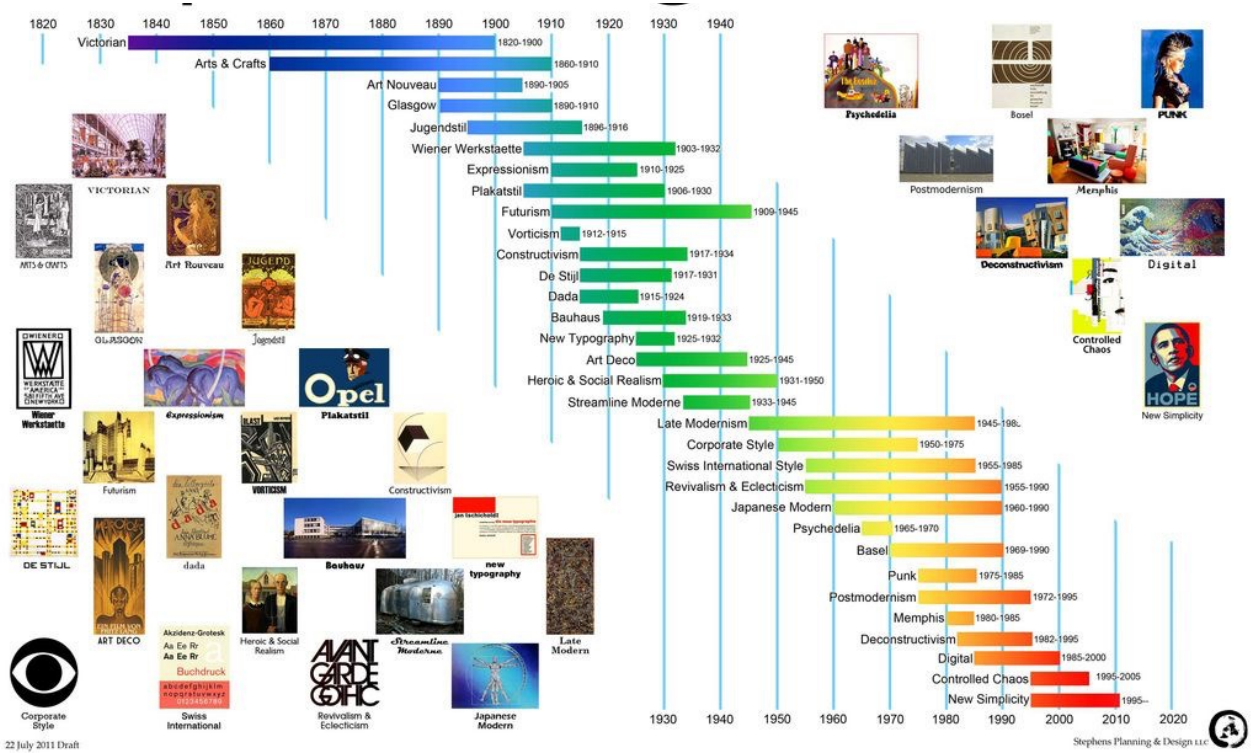
= canonical instead of representative



https://www.mkg-hamburg.de/de/sammlung/sammlungen/plakat.html

STRATEGIES

- Dependence on the research question
- Size of the sample depends on the number of characteristic expressions



https://i.pinimg.com/564x/7b/c6/f5/7bc6f523596df3e58c34fc38775f85eb.jpg

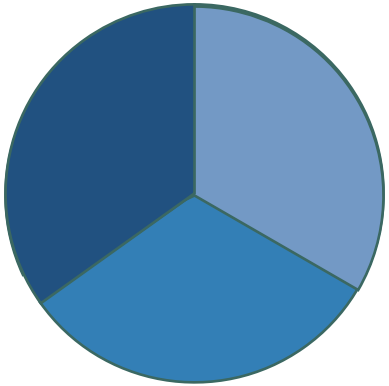
STRATEGIES

• No one-sided expansion of the data basis, but equal distribution of all relevant characteristics

https://static.dw.com/image/55730573_101.jpg

STRATEGIES

• If necessary, split the sampling frame into several subpopulations!
However, you can then no longer examine the relationship of the strata to each other.



https://www.mkg-hamburg.de/de/sammlung/sammlungen/plakat.html

HOW REPRESENTATIVE ARE SURVEY EXHIBITIONS?

Archive

28 February until 20 September 2020

THE POSTER

200 YEARS OF ART AND HISTORY

With nearly 400 exhibits by around 200 artists and designers, the exhibition The Poster at the Museum für Kunst und Gewerbe Hamburg (MKG) offers a large-scale, representative overview of the history of the poster from its beginnings in the early nineteenth century to today. Art and history, design and advertising meet in this medium.

https://www.mkg-hamburg.de/en/exhibitions/archive/2020/the-poster.html

CHALLENGES

- Reflective handling of data in the humanities (source criticism)

  i.e. also discussion of the right data basis, the appropriate statistical method and adequate visualisation.

- Modelling of fuzziness(ies) with regard to statistical evaluability

- different quantitative methods
- basics of statistics (Bedeutung der Normal-verteilung, Arten der Stichprobe, Konfidenz-intervalle, Ergebnis-genauigkeit etc.)
- Analysis methods for common measures and structures
- Sampling concepts and theories
- Basics of georeferenced analyses

- quantitative and qualitative description and analysis of specialised data and appropriate selection (sampling) of data
- calculate different central tendencies; understand and verify results using statistical methods
- model correlations between two characteristics; distinguish correlations from causal relationships.

What do you mean by result accuracy? What is the difference to standard deviation?

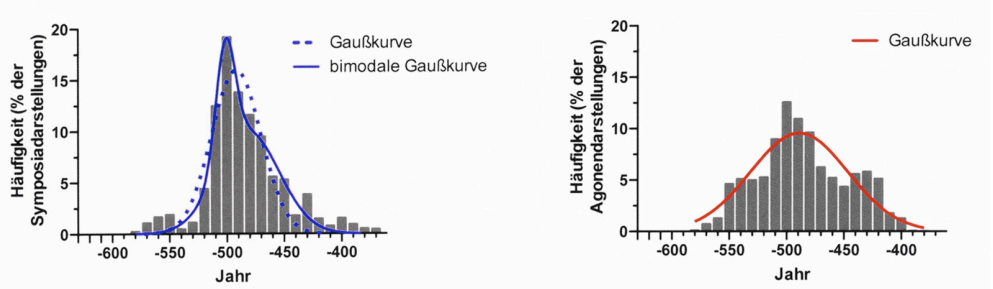
Folie 17. 73

Which sampling methods do you know? Describe a method using an example from image science and name its advantages

Folie 26–38

When should statistical methods be used in the humanities? What are the advantages and difficulties?

Folie 39–54



1. Darstellungen der Symposia

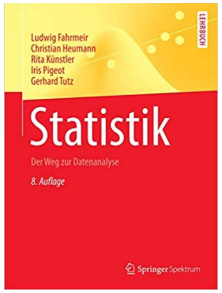
2. Darstellungen der Athletik

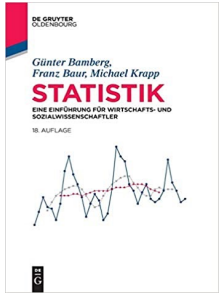
Legende
1. Prozentuale Häufigkeitsverteilungen der Darstellungen von Symposia in Abhängigkeit vom Jahrzehnt ihrer Produktion und hieraus berechnete Verteilungskurven (Gaußkurve: blau, gestrichelt; bimodale Normalverteilung: blau, durchgezogen).
2. Prozentuale Häufigkeitsverteilungen der Darstellungen von athletischen Szenen in Abhängigkeit vom Jahrzehnt ihrer Produktion und hieraus berechnete Gaußsche Verteilungskurve (rot).

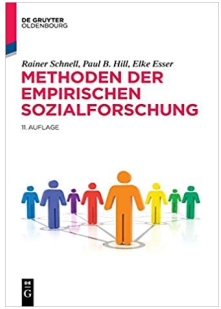
Die Kurvenberechnungen wurden mit dem Programm „Prism5 for Mac OS X" von GraphPad Software durchgeführt.

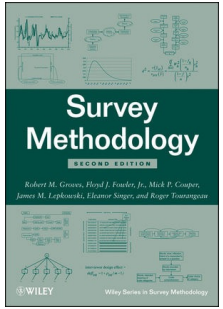
Why should one take a critical view of the visualisation of the data collected by Wolfgang Filser shown opposite?

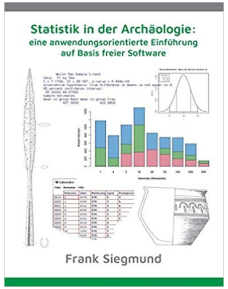
Folie 9. 16–18. 70

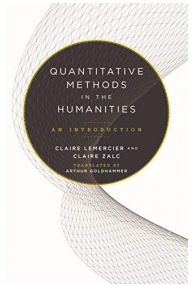












Ludwig Fahrmeir, Christian Heumann, Rita Künstler, Iris Pigeot, Gerhard Tutz, Statistik: Der Weg zur Datenanalyse. 8. Aufl. (Springer, 2016)

Günter Bamberg, Franz Baur, Michael Krapp, Statistik: Eine Einführung für Wirtschafts- und Sozialwissenschaftler, 18. Aufl. (Berlin: De Gruyter 2017)

Rainer Schnell, Paul B. Hill, Elke Esser, Methoden der empirischen Sozialforschung, 11. Aufl. (De Gruyter Oldenbourg, 2018)

Robert Groves, et al. Survey Methodology, 2nd ed. (Wiley, 2010)

Frank Siegmund, Statistik in der Archäologie: eine anwendungsorientierte Einführung (Selbstverlag 2020)

Claire Lemercier, Claire Zalc, Quantitative Methods in the Humanities: An Introduction (Univ. of Virginia Press, 2019)

IMPRESSUM

Inhaltlich verantwortlich:

Prof. Dr. Martin Langner
Institut für Digital Humanities
Georg-August-Universität Göttingen
37073 Göttingen
Martin.Langner@uni-goettingen.de
Telefon 0551 / 3926790
https://www.uni-goettingen.de/digitalhumanities

Nutzungsbedingungen:

Bitte informieren Sie uns, wenn Sie der Ansicht sind, dass ein hier veröffentlichter Scan eine Urheberrechtsverletzung darstellt; wir werden diesen Fall umgehend lösen.

Haftungshinweis:

Trotz sorgfältiger inhaltlicher Kontrolle übernehmen wir keine Haftung für die Inhalte externer Links. Für den Inhalt der verlinkten Seiten sind ausschließlich deren Betreiber verantwortlich.