The Birth of Statistical Thinking in the 18th and 19th Centuries

The rise of the moral or human sciences,

the scientific study of human beings

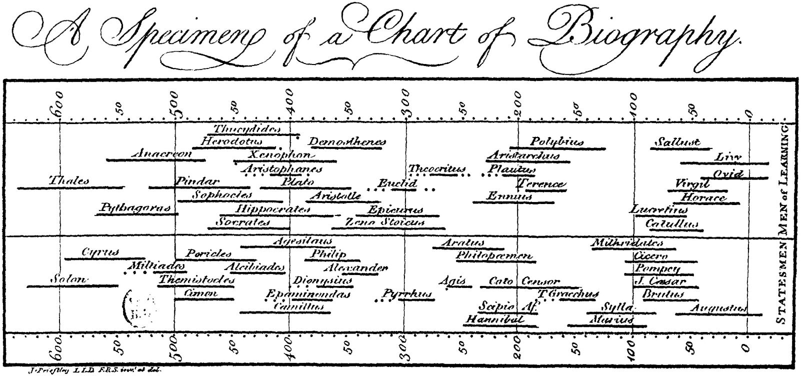
Waseda University, SILS,
History of Modern Earth and Life Sciences

Statistical thinking and understanding data

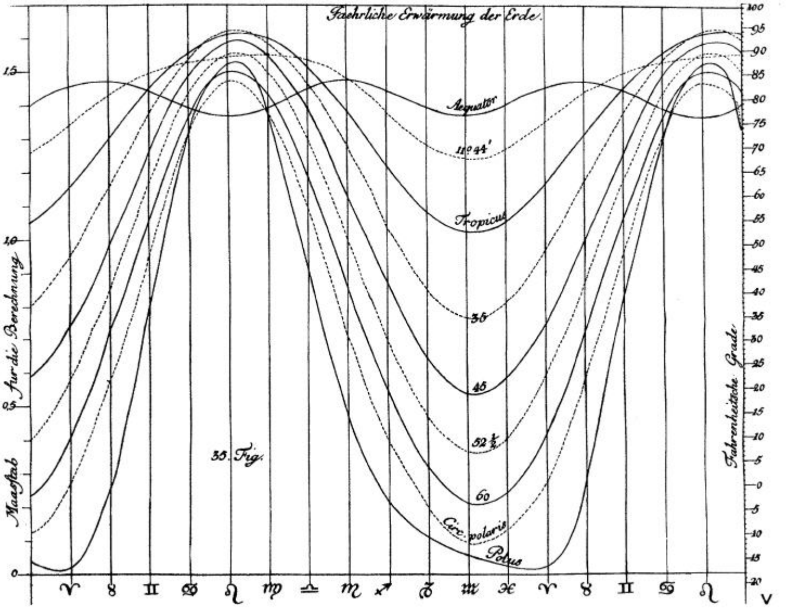
In the 18th and 19th centuries, there were social and political changes that involved collecting and organizing a constantly growing set of data – in the natural, life and moral sciences. This explosion of numerical information laid the groundwork for profound changes in the way that we thought of ourselves and our world – we moved away from concepts of *essential nature* and *ideal types*, towards concepts of *statistical normality*.

This period also saw the development of mathematical techniques for analyzing probability and statistics – mathematical equations dealing with frequencies and degrees of belief. Moreover, during this period, scholars of various types produced now-common visual representations of numerical, and factual, data, which allows the viewer to understand at a glance a great deal of information – and which had previously been presented in lists.

Edmund Halley's map of magnetic lines, 1701

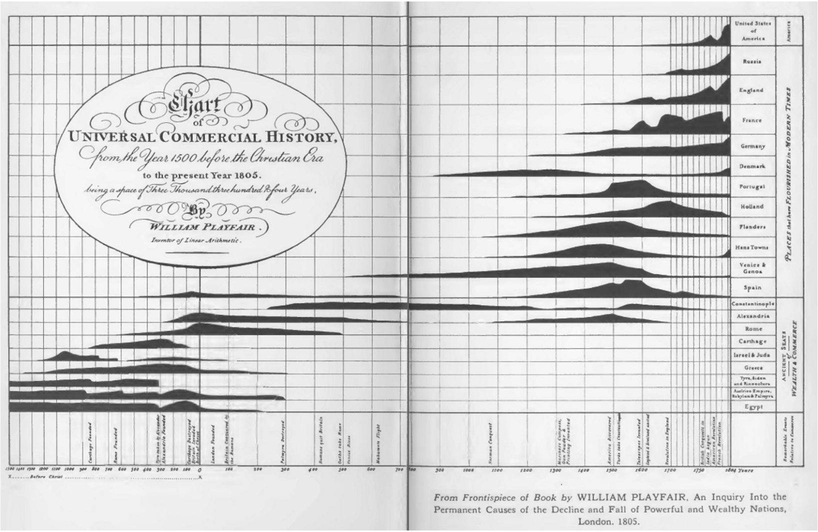


Joseph Priestley’s timeline of ancient Greek statesmen and thinkers, 1765

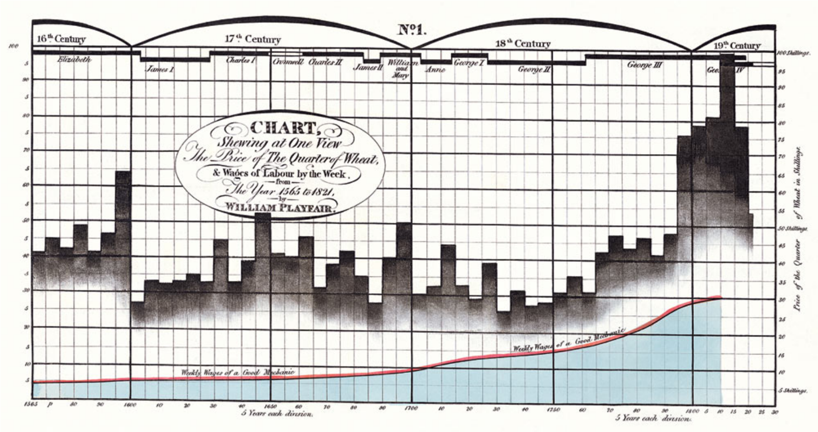


Lambert's graph of annual temperature variation at different latitudes, 1779

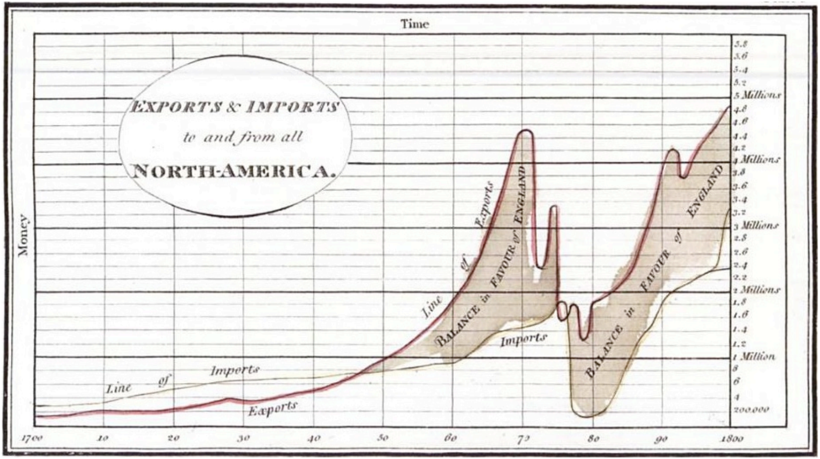
Marcellin du Carla-Boniface's topographical map, 1782



Graph of historical economies by William Playfair (brother of John), 1805

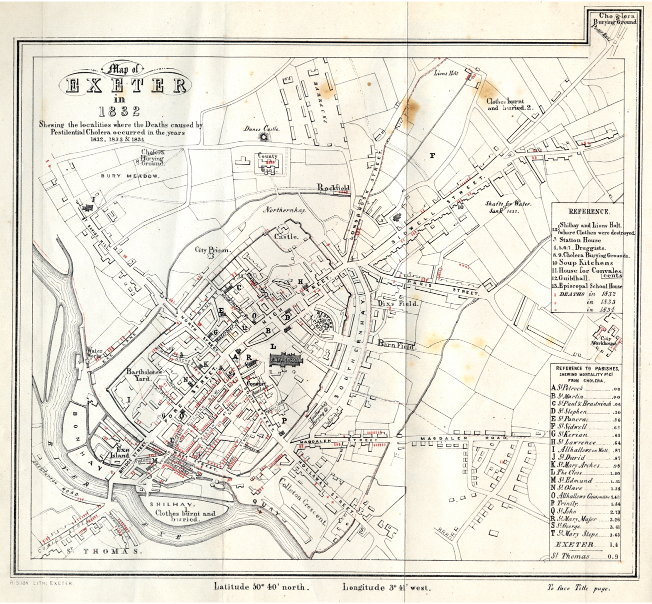


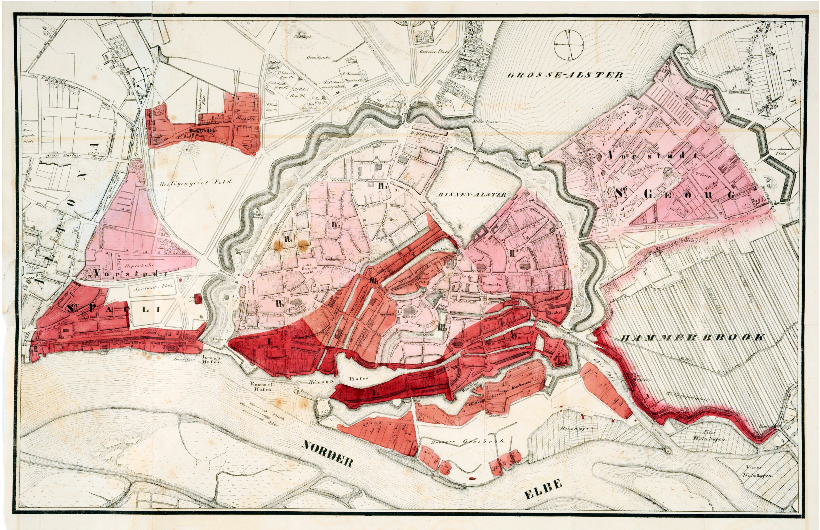
Playfair's graph of the price of wheat compared to wages, 1805



Playfair's graph England's trade balance with North America, 1805

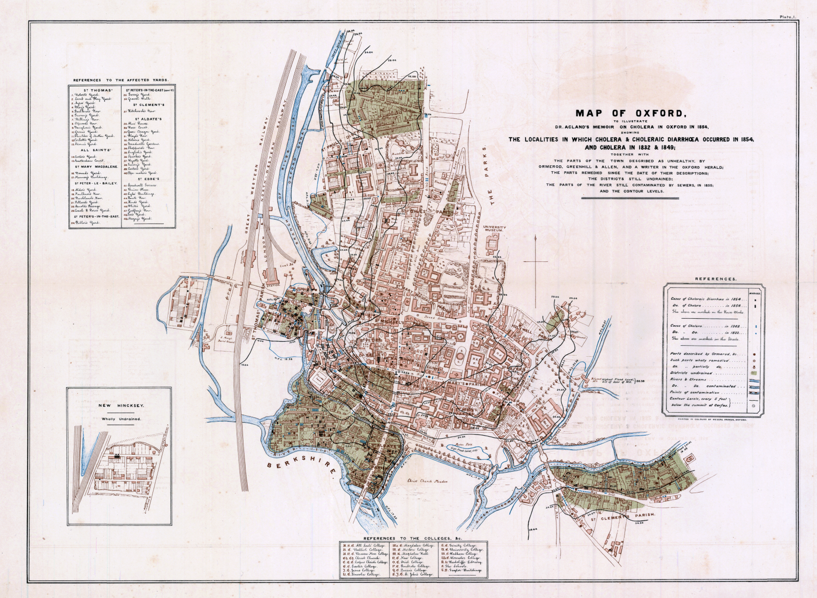
Shapter's map of the cholera epidemic in Exeter, England, 1832



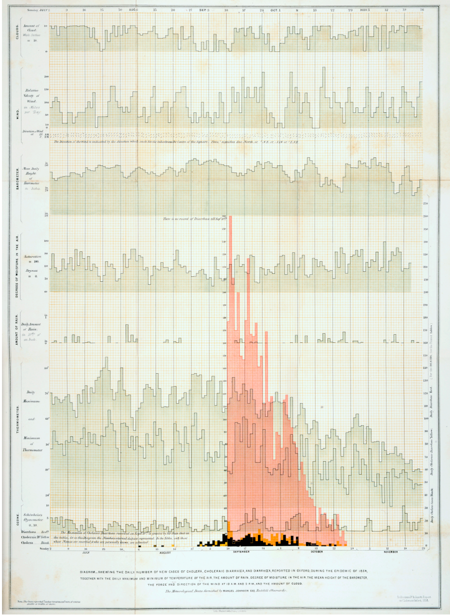


Rothenburg's map of the cholera epidemic in Hamburg, Germany, 1832

Chadwick's map of the cholera epidemic in Leeds, England, 1843



Acland's map of three cholera epidemics in Oxford, England, 1854

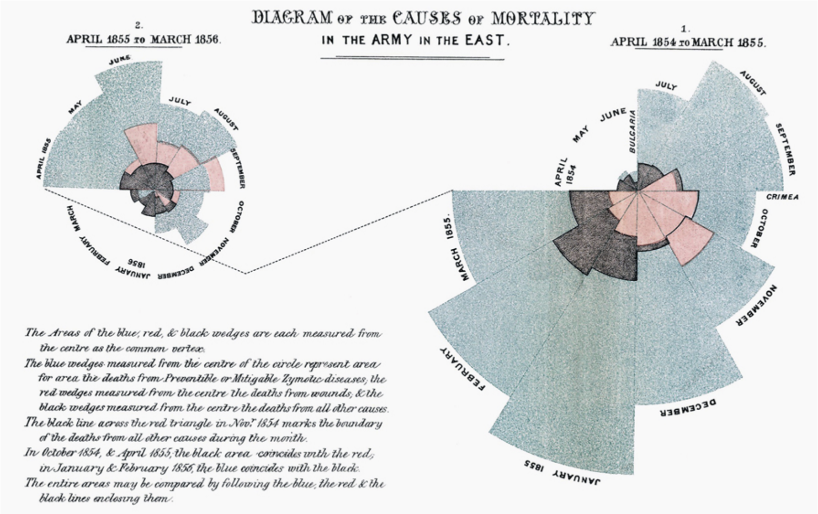


Acland's graph of three cholera epidemics in Oxford, England, 1854

red = diarrhea,

yellow = choleraic dysentery,

black = cholera



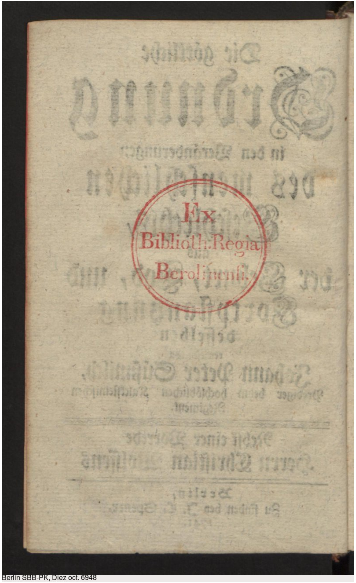
Nightingale's graphic of the causes of death in the Crimean War, 1858

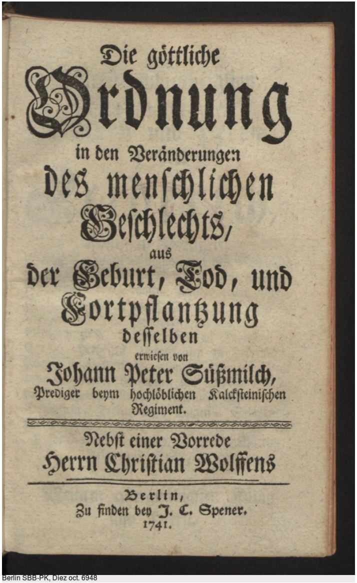
blue = infections, red = wounds, black = other causes

The divine order

Statistical work during the 18th century was done by practically minded philosophers and liberal, enlightenment reformers. There was little access to actual numbers except through church records and local records scattered around in unrelated departments (marriages, deaths, births, etc.). By the end of the century, however, Prussia began to establish more regular protocols for collecting data and putting it to use by the state – the word *statistics* comes from the German term for statecraft.

One of the first important works of statistics was Johann Süssmilch's (1707-1766), *Die göttliche Ordnung* (1741), in which he argued that the conditions of the world, in and of themselves, provided proof of God's benevolence. He collected data from church registers and complied it into a sort of *demographic theology*. By looking at the numbers, he claimed he could understand the wisdom of God's plan.





Die göttliche
Ordnung
in den Veränderungen
des menſchlichen
Geſchlechts,
aus
der Geburt, Tod, und
Fortpflantzung
deſſelben
erwieſen von
Johann Peter Süßmilch,
Prediger beym hochlöblichen Kalckſteiniſchen
Regiment.

Nebſt einer Vorrede
Herrn Chriſtian Wolffens

Berlin,
Zu finden bey J. C. Spener.
1741.

Süssmilch's *Die göttliche Ordnung* (1741)

Besondres Verzeichniß

Der in Wien Gestorbenen nach dem Geschlecht und Alter.

Jahre.	Gestorbenen.				
	Mans-Person.	Frauens-L.	Knäbelein.	Mägdlein.	Summa
1722	1038	942	1551	1430	4961
1723	1079	974	1758	1632	5443
1726	1082	944	1951	1733	5710
1728	1716	1331	2085	1869	7001
1738	1534	1298	2339	2192	7363
1739	1446	1283	1776	1653	6158
Sum. v. 6 Jahre	7895	6772	11460	10509	36636
Mittel-Zahl im Durchschnitt von 6 Jahr					6106

LISTE

LISTE

Aller der in Breßlau Verstorbenen und Getauften so unter der Jurisdiction des Stadt-Magistrats befindlich.

Jahre.	Verstorben.	Getaufte.	Verhältniß der Geb. zu den Gestorbenen.
Von 1555	2298	1636	
Von 1556 b. 1567 also in 12 Jahr.	13960	16323	1,1:1
Die Mittel-Zahl	1163	1360	
Von 1568	9251	1053	Pest-Jahr.
Von 1569 bis 1571 in 3 Jahren	2660	3041	1,1:1
Die Mittel-Zahl	886	1013	
Von 1572	2441	1129	Pest-Jahr.
Von 1573 bis 1584 in 12 Jahren	12867	15307	1,1:1
Die Mittel-Zahl	1072	1275	
Von 1585	8931	1018	Pest-Jahr.
Von 1586 bis 1598 in 13 Jahren	15115	15650	1,03:1
Die Mittel-Zahl	1162	1203	
Von 1599	3942	1100	Pest-Jahr.
Von 1600 bis 1612 in 13 Jahren	15477	14017	1:1,1
Die Mittel-Zahl	1190	1078	
Von 1613	2357	1093	Pest-Jahr.
Von 1614 bis 1624 in 11 Jahren	14927	12410	1:1,2
Die Mittel-Zahl	1357	1128	
Von 1625	3000	920	Pest-Jahr.
Von 1626 bis 1632 in 7 Jahren	9583	7066	1:1,3
Die Mittel-Zahl	1369	1009	

)()()()(Von

Those who died in Vienna; those who died and were baptized in Wrocław

The social good

Birth rates were said to be a function of available farmland. As the population grows the available arable land decreases and the birth rate drops. In due time, there is a shortage of labor and land becomes more available. These patterns were mostly born out in Prussia at that time. (There was also a fair amount of moralizing. The higher mortality rates in cities were blamed on sin. Süssmilch warned against alcohol, gambling, prostitution, city life, priestly celibacy, and so on.)

Süssmilch believed that the true wealth of a nation lay in its *population*. He argued against war and claimed that it was based on a basic misunderstanding on the part of princes. They tried to use it to increase the number of their subjects. Instead they should introduce medical care and systematic inoculations, reduce taxes, and redistribute arable land. Finally, he advocated the creation of government offices to collect population statistics and to act on them.

The moral sciences

The moral sciences were developed by Enlightenment philosophers for the study of human society and its rational improvement – such as the Marquis de Condorcet, Pierre-Simon de Laplace, Thomas Bayes, and others, who developed mathematical tools for dealing with probability and statistics.

Condorcet, *Discours prononcés* … (1782)

"Those sciences … the object of which is man himself, the direct goal of which is the happiness of man, will enjoy a progress no less sure than that of the physical sciences; and this idea so sweet, that our nephews will surpass us in wisdom as in enlightenment, is no longer an illusion."

Condorcet's moral sciences had two approaches: historical (progressive) and statistical (analytical). These began to diverge into separate traditions in the 19th century.

The Malthus doctrine

Thomas Malthus (1766-1834) was a Cambridge educated clergyman who wrote *An Essay on the Principle of Population as it affects the Future Improvement on Society* (1798), which argued that "population increases in a geometric ratio, while the means of subsistence increases in an arithmetic ratio." So, there will quickly be too little food and people will have to compete for the little that there is. Hence, vice and misery – war, famine and disease – are necessary checks on population growth, not simply the effects of a mismanaged society.

In this view, society itself was its own worst enemy and the state required numerical studies of political economy to soften these dangers. The people should be educated on the real sources of their misery. The fundamental claim was that "society" had its own set of essentially natural laws, independent of either state or government.

Malthus, A restrictive law

Malthus, *An Essay on the Principle of Population …* (1798)

"Through the animal and vegetable kingdoms, nature has scattered the seeds of life abroad with the most profuse and liberal hand. She has been comparatively sparing in the room, and the nourishment necessary to rear them… The race of plants, and race of animals shrink under this great restrictive law. And the race of man cannot, by any efforts of reason, escape from it. Among plants and animals its effects are waste of seed, sickness, and premature death; among mankind, misery and vice … For let every corruption and abuse of power be entirely got rid of; let virtue, knowledge and civilization be advanced to the greatest height … the sooner will they be overthrown again, and the more inevitable and fatal will be the catastrophe."

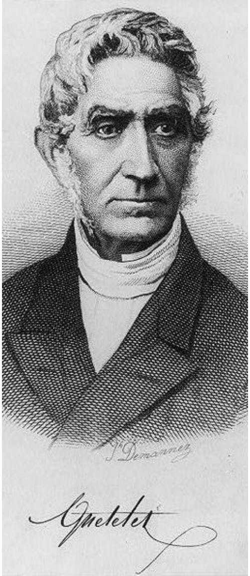
Malthus, The principle of population

Malthus, *An Essay on the Principle of Population …* (1798)

"For the principle of population will still prevail, and from the comfort, ease and plenty that will abound, will receive an increasing force and impetus. The number of mouths to be fed will have no limit; but the food that is to supply them cannot keep pace with the demand for it… There will be no remedy; the wholesome checks of vice and misery (which have hitherto kept this principle within bounds) will have been done away; the voice of reason will be unheard; the passions only will bear sway; famine, distress, havoc and dismay will spread around; hatred, violence, war and bloodshed will be the infallible consequence; and from the pinnacle of happiness, peace, refinement and social advantage shall be hurled once more into a profounder abyss of misery, want, and barbarism than ever by the sole operation of the principle of population!"

Adolphe Jacques Quetelet (1796–1874)

- Educated in mathematics in Gent and Brussels.
- He studied for a while in Paris, under Laplace and Fourier.
- Founded and directed the astronomical observatory in Brussels.
- He was elected to a number of scientific academies – Royal Academy of London, Swedish Academy, and others.
- He founded a number of statistical journals and societies.
- Some famous works are *Sur l'homme et le developpement de ses facultés, essai d'une physique sociale* (1835), and *Recherches statistiques* (1844).



The mathematics of society

Quetelet was an astronomer and he applied the mathematics of astronomical observations to statistical problems. In particular, he was interested in the statistical methods used to compute "true" position from many observations of same object – known as the *problem of least squares*.

He paid particular attention to the patterns of statistical data, and argued that these exhibited strong regularities, despite the freedom of the individuals.

Quetelet, *Nouveaux mémoires …* (1829)

"We know in advance how many individuals will dirty their hands with the blood of others, how many will be forgers, how many poisoners, nearly as well as one can enumerate in advance the births and deaths that must take place."

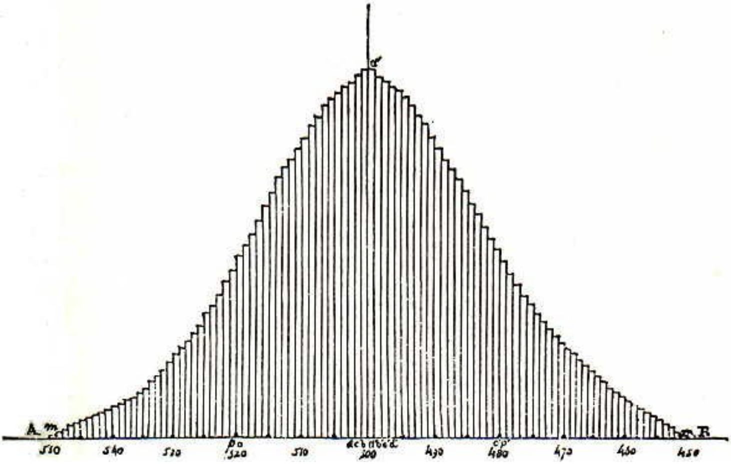
He argued that these statistics provided evidence for a law of large numbers.

The error curve, or normal curve

The error curve, or normal curve, had been derived either as an observational error curve or as the limit of a coin-tossing game. In either case, we are talking about something we take to be real (the celestial body really is somewhere, the coin has some actually tendency), and so on.

Quetelet imported this *claim about reality* into biological and social phenomena by subjecting them to the normal curve. He argued, or rather assumed, that where we find an error curve in the observations, we can infer that there is some *real trait* or characteristic being observed. That is, he made the mean into a *real object* – what we call reification.

Quetelet's error curve



The "average man" – as an ideal

In his *Treatise on Man*, Quetelet introduced the idea that each "people" or "nation" had a racial, or ethnic, type – which was understood to be a real type. A race would be defined by finding the distribution of the measurements of physical and moral qualities. He believed that this average can be thought of as a real feature of a population. In this way, the fundamental object of his study shifted away from individuals.

He asserted that each individual in a society has within them a tendency towards every activity or disposition in the degree to which it is found in the society (or in some subgroup) at large. For Quetelet, this mean was both *healthy* and *proper*.

Quetelet , "Researches sur le penchant au crime ...," (1832)

"Deviations more or less great from the mean have constituted ugliness in body as well as vice in morals and a state of sickness with regard to the constitution."

Properties of a population

In 1844, he added a new twist. He argued as follows: (1) if we measure an individual many times – say, height – we get a normal curve, (2) this curve is produced by the fact that there is *one real value*, (3) measuring a number of individuals of different groups many times each would give a different curve, hence (4) we can use the normal curve as a test for the *homogeneity of a population*.

In this way, he transformed the normal curve into a test for an underlying cause – such as *a type* and its natural variations. The object of study was now the abstraction that this curve described.

Scottish solders

At the time, there was not enough data to find this curve in vary many circumstances.

Quetelet used a paper from the *The Edinburgh Medical Journal* which had the measurements of 5,000 Scottish solders. He fit the chest figures to the curve and claimed it was the same phenomena that would result from someone measuring a single solder 5,000 times.

In the 1870s, he attempted to prove his doctrine using data from French conscripts and American volunteers for the Union army.

MESURES de la POITRINE.	NOMBRE d'hommes.	NOMBRE PROPORTIONNEL.	PROBABILITÉ d'après L'OBSERVATION.	RANG dans LA TABLE.	RANG d'après le CALCUL.	PROBABILITÉ d'après LA TABLE.	NOMBRE D'OBSERVATIONS calculé.
Pouces.							
33	3	5	0,5000			0,5000	7
34	18	31	0,4995	52	50	0,4993	29
35	81	141	0,4964	42,5	42,5	0,4964	110
36	185	322	0,4823	33,5	34,5	0,4854	323
37	420	732	0,4501	26,0	26,5	0,4531	732
38	749	1305	0,3769	18,0	18,5	0,3799	1333
39	1073	1867	0,2464	10,5	10,5	0,2466	1838
			0,0597	2,5	2,5	0,0628	
40	1079	1882	0,1285	5,5	5,5	0,1359	1987
41	934	1628	0,2913	13	13,5	0,3034	1675
42	658	1148	0,4061	21	21,5	0,4130	1096
43	370	645	0,4706	30	29,5	0,4690	560
44	92	160	0,4866	35	37,5	0,4911	221
45	50	87	0,4953	41	45,5	0,4980	69
46	21	38	0,4991	49,5	53,5	0,4996	16
47	4	7	0,4998	56	61,8	0,4999	3
48	1	2	0,5000			0,5000	1
	5738	1,0000					1,0000

Quetelet's table for the chest measurement of Scottish solders

Cause and variation

For Quetelet, variation itself was inherently uninteresting. He took it to be the result of *minute unknown causes* acting on the individual and driving it away from the *true type*. The mean, on the other hand, was the proper object of study.

Quetelet, *Sur l'homme …* (1835)

"All things will occur in conformity with the mean results obtained for a society. If one seeks to establish… a social physics, it is he [that is, the *average man*] whom one should consider, without disturbing oneself with particular cases or anomalies, and without studying whether some given individual can undergo a greater or lesser development in one of his faculties."

Francis Galton (1822–1911)

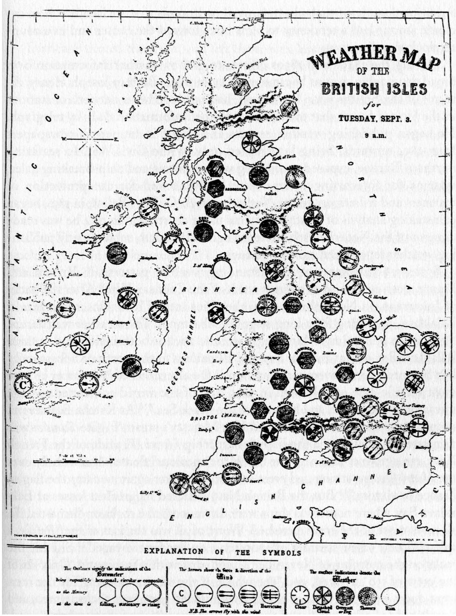
- From a wealthy Quaker family, which had gone Anglican. Grandson of Erasmus Darwin.
- Educated at Cambridge (Passman).
- Explored Africa.
- Worked on meteorology.
- On reading Darwin's *Origins of Species* (1859), he turned his attention to heredity.
- Became one of Britain's most prominent statistical scientists.
- Wrote prolifically for both the scientific and lay public.

Galton's work

Galton made significant contributions to African exploration, meteorology, statistics, psychology, personal identification, and human heredity. He developed the statistical ideas of *regression* and *correlation*. More disturbingly, he founded the scientific discipline of eugenics.

In 1884, he established the Anthropometric Laboratory for collecting detailed statistical data on human subjects.

Some of his most influential books were *Hereditary Genius* (1869), *English Men of Science: their Nature and Nurture* (1874), *Inquiries into Human Faculty and Its Development* (1883), *Natural Inheritance* (1889), and *Finger Prints* (1892).



Galton's weather map of the British Isles

Heredity

Darwin's *Origins of Species* turned Galton toward heredity – which he dealt with in a text called *Hereditary Genius* (1869).

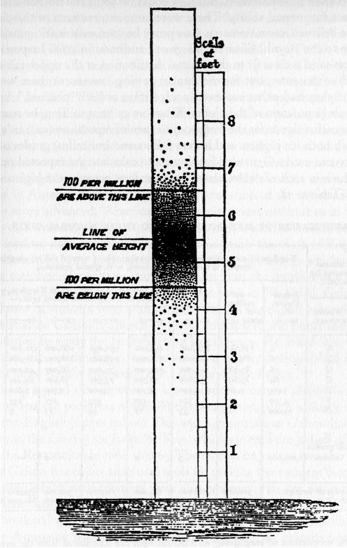
He was particularly interested in human heredity and used *pedigree analysis* to try to determine what traits are passed down. He assumed that social rank was an unproblematic indicator of talent and capacity. (Of course, it was difficult to rule out the possibility that environmental factors played a role, but he simply disregarded this.)

He introduced the idea that certain people should be encouraged to have children, while others should be discouraged.

Galton argued that characteristics like talent and genius are *inheritable traits* and, hence, must be distributed in the population like height.

He produced a normal distribution by imagining the whole population measuring their height on a single board.

By this, and other equally dubious arguments, Galton calculated that 1 *man* in 4,000 would be "eminent."



Nature and nurture

Although there was a strong public response to Galton's book, many were fairly cautious in accepting his main position, and a few wrote strongly against his position – for example, in Switzerland, Augustin Pyramus de Candolle argued that it was social positioning, means and education that led scientific "talent" to run in families.

Galton countered with *English Men of Science: Their Nature and Nurture* (1874), which was a superficial work based on a questionnaire sent to his friends.

This debate also led Galton to develop the technique of twin studies. He compared sets of twins in order to try to determine which characteristics depend on hereditary factors as opposed to environmental.

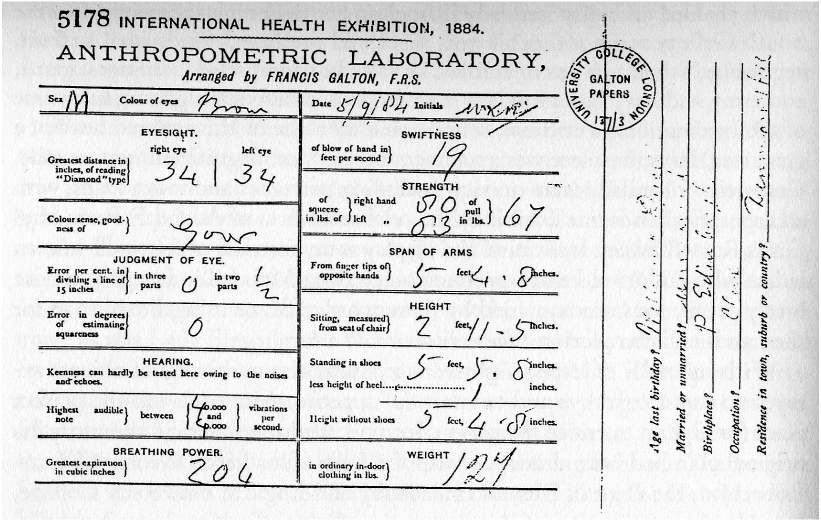
Gathering data

In 1884, Galton used the International Health Exhibition to set up an Anthropometric Laboratory "for testing and measuring the efficiency of the various mental and bodily powers."

The examination took about 1/2 hour and was designed to test for

> "Keenness of Sight and of Hearing; Colour Sense, Judgment of Eye; Breathing Power, Reaction Time; Strength of Pull and Squeeze; Force of Blow; Span of Arms; Height, both standing and sitting; and Weight."

After the Exhibition, he set up the Laboratory on a permanent basis with a more detailed test.



5178 INTERNATIONAL HEALTH EXHIBITION, 1884.

ANTHROPOMETRIC LABORATORY,

Arranged by FRANCIS GALTON, F.R.S.

UNIVERSITY COLLEGE LONDON GALTON PAPERS 137/3

Sex *M* Colour of eyes [illegible]	Date *5/9/84* Initials [illegible]
EYESIGHT.	**SWIFTNESS**
Greatest distance in inches, of reading "Diamond" type — right eye *34* — left eye *34*	of blow of hand in feet per second *19*
Colour sense, goodness of [illegible]	**STRENGTH** — of squeeze in lbs. of right hand *80*, left " *83*; of pull in lbs. *67*
JUDGMENT OF EYE. — Error per cent. in dividing a line of 15 inches — in three parts *0* — in two parts *1/2*	**SPAN OF ARMS** — From finger tips of opposite hands *5* feet, *4.8* inches.
Error in degrees of estimating squareness *0*	**HEIGHT** — Sitting, measured from seat of chair *2* feet, *11.5* inches.
HEARING. Keenness can hardly be tested here owing to the noises and echoes.	Standing in shoes *5* feet, *5.8* inches. — less height of heel.......... *1* inches.
Highest audible note — between [illegible] and [illegible] vibrations per second.	Height without shoes *5* feet, *4.8* inches.
BREATHING POWER. Greatest expiration in cubic inches *204*	**WEIGHT** in ordinary in-door clothing in lbs. *124*

Age last birthday? [illegible]

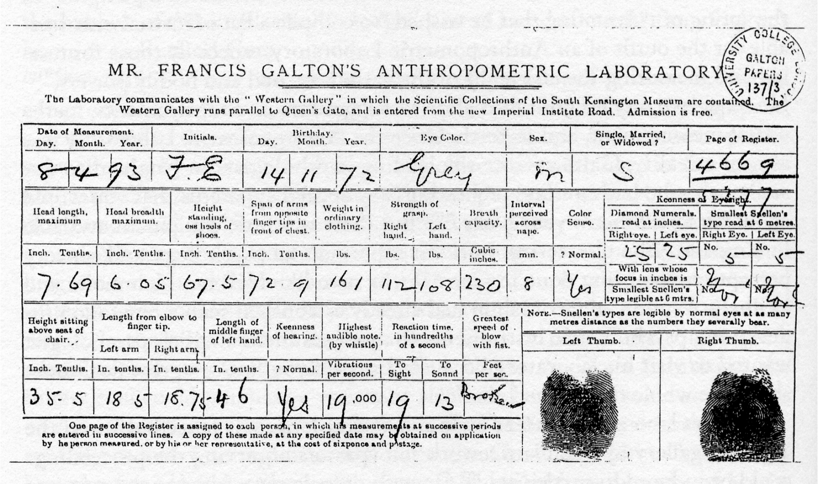
Married or unmarried? [illegible]

Birthplace? [illegible]

Occupation? [illegible]

Residence in town, suburb or country? [illegible]

Galton's data collection card, International Health Exhibition



MR. FRANCIS GALTON'S ANTHROPOMETRIC LABORATORY

The Laboratory communicates with the "Western Gallery" in which the Scientific Collections of the South Kensington Museum are contained. The Western Gallery runs parallel to Queen's Gate, and is entered from the new Imperial Institute Road. Admission is free.

Date of Measurement. Day. Month. Year.	Initials.	Birthday. Day. Month. Year.	Eye Color.	Sex.	Single, Married, or Widowed?	Page of Register.
8 4 93	F.E	14 11 72	Grey	m	S	4669

Head length, maximum	Head breadth maximum.	Height standing, less heels of shoes.	Span of arms from opposite finger tips in front of chest.	Weight in ordinary clothing.	Strength of grasp. Right hand.	Strength of grasp. Left hand.	Breath capacity.	Interval perceived across nape.	Color Sense.	Keenness of Eyesight. Diamond Numerals read at inches. Right eye.	Left eye.	Smallest Snellen's type read at 6 metres. Right Eye.	Left Eye.
Inch. Tenths.	Inch. Tenths.	Inch. Tenths.	Inch. Tenths.	lbs.	lbs.	lbs.	Cubic inches.	mm.	? Normal.	25	25	No. 5	No. 5
7.69	6.05	67.5	72.9	161	112	108	230	8	Yes	With lens whose focus in inches is Smallest Snellen's type legible at 6 mtrs.		No.	No.

Height sitting above seat of chair.	Length from elbow to finger tip. Left arm	Right arm	Length of middle finger of left hand.	Keenness of hearing.	Highest audible note. (by whistle)	Reaction time, in hundredths of a second. To Sight	To Sound	Greatest speed of blow with fist.
Inch. Tenths.	In. tenths.	In. tenths.	In. tenths.	? Normal.	Vibrations per second.			Feet per sec.
35.5	18.5	18.75	4.6	Yes	19,000	19	13	Broken

NOTE.—Snellen's types are legible by normal eyes at as many metres distance as the numbers they severally bear.

Left Thumb. Right Thumb.

One page of the Register is assigned to each person, in which his measurements at successive periods are entered in successive lines. A copy of these made at any specified date may be obtained on application by the person measured, or by his or her representative, at the cost of sixpence and postage.

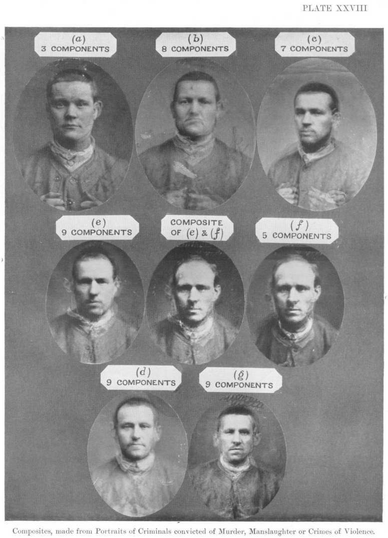
Galton's data collection card, Anthropometric Laboratory

Interest in variation

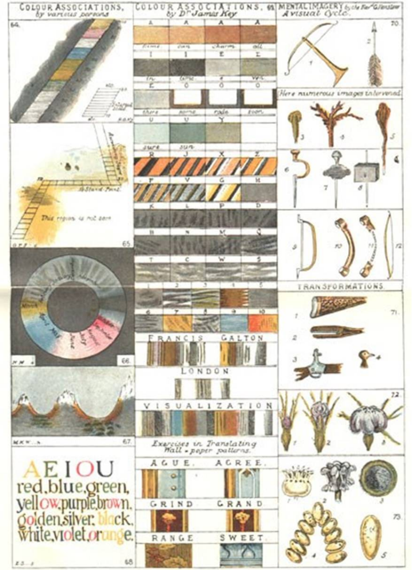
Galton was fascinated by variation as a *real phenomena*, not just as a "perturbation."

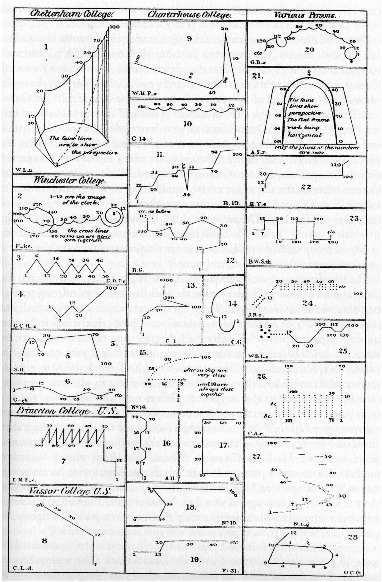
This led him to study the ways in which variations

- could be grouped together to form types (composite photography),
- could be used to distinguish the unique features of individuals (psychology and personal identification),
- could be correlated in individuals and across generations (correlation and regression).



Composite photographs of criminals





Galton, *Inquiries into Human Faculty* (1883)

Regression and correlation

Regression

Regression is the tendency of the progeny of extreme parents to regress toward the mean of the population, or the parents of extreme progeny to regress towards the mean. Each group is normally distributed, but the mean is closer to that of the whole population.

Correlation

Correlation is the tendency of certain traits of an individual to be statistically linked. The traits are still normally distributed, but the means are close together.

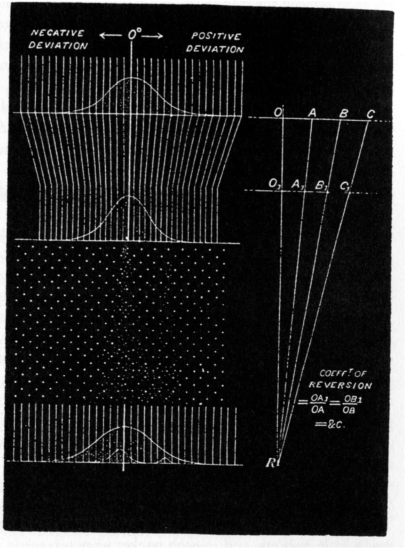
Galton discovered these as phenomena and helped developed mathematical tests for them.

Galton's work on regression

Galton introduced the concept of regression to a broad audience at a Friday Lecture at the Royal Institution.

He used a wooden model, called a *quincunx*, to drop tiny led balls through a scattering device. Each column produced a normal distribution that was closer to the mean. All of them produced a single normal distribution.

Plotting the means for the parents and progeny gave two lines. The inclination of the lines could be used to determine the rate of regression to the mean.



Galton's slide showing the quincunx

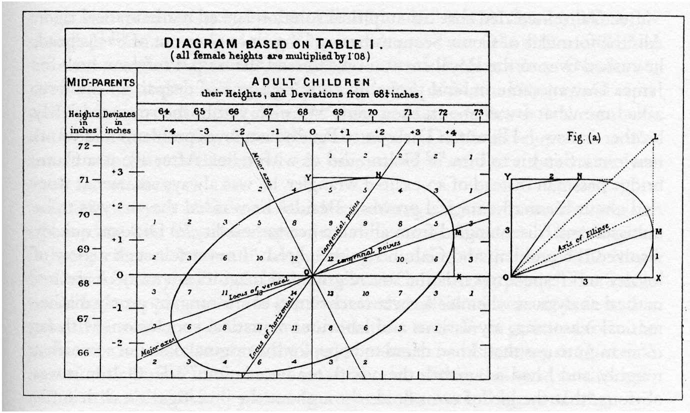
Regression to the mean

Galton performed a series of experiments in which he tracked the heights of plants over a number of generations, comparing those of parents with their offspring.

Galton, "Regression towards Mediocracy ...," 1886

"It appeared from these experiments that the offspring did *not* tend to resemble their parent seeds in size, but to be always more mediocre than they – to be smaller than the parents, if the parents were large; to be larger than the parents if the parents were very small... The experiments showed further that the mean filial regression towards mediocracy was directly proportional to the parental deviation from it."

Moreover, the phenomena of regression to the mean is time-independent, so that it also goes from offspring to parents. This shows us that this a *purely statistical phenomena*, and does not call for a causal explanation.



The traits were normally distributed for each trait. When all the data was plotted, lines joining equal deviations formed ellipses. These mathematical properties could then be used to look for correlation.

Eugenics

Eugenics was both a science and a major social movement of the end of the 19th and first half of the 20th centuries. There were international conferences, journals, academic positions, and so on – and its practitioners believed that it was *supported by the biological sciences*.

It advocated for the breeding of certain human populations by using "negative" reinforcement to discourage, or prevent, certain people from having children, and "positive" reinforcement to encourage other people to have children. Whereas most of Galton's suggestions were positive, many nations developed state-run programs to institutionalize negative measures.

Sterilizations

There were programs of state enforced sterilization in many countries. Just as some examples:

- Over 60,000 court-ordered sterilizations in the US.
- Involuntary sterilizations in Sweden, Norway, Switzerland and Alberta.
- There was a series of eugenics laws in Japan that governed the sterilization of criminals, and those with genetic and mental diseases.
- In Nazi Germany, some 400,000 people were sterilized for alcoholism, feeblemindedness, and schizophrenia.
  - The 1935, the Nuremburg Laws were designed to “cleanse” the German population of unwanted elements (Jews, Roma, homosexuals, etc).

Resistance to eugenics

There was almost no resistance to eugenics before the stock market crash of 1930, and still little before the atrocities of Nazi Germany. After the market crash, some biologists (for example, L.C. Dunn) began to question the assumed link between social position, wealth and biological viability or hereditary worth.

From the very beginning there were a few outspoken opponents, such as G.K. Chesterton.

Chesterton, *Eugenics and Other Evils* (1922)

"Eugenics, as discussed, evidently means the control of some men over the marriage and unmarriage of others; and very probably means the control of the few over the marriage and unmarriage of the many."

Overview

- We have looked at the rise of statistical information and numerical data, and the use of visual representation to codify and explicate this information.
- We discussed the development and use of statistics in Quetelet's articulation of the normal as the ideal type.
- We looked at Galton's investigation of variation and his conception of the normal as mediocre.
- Finally, we looked briefly at the rise and decline of eugenics.