



Analoge Informationsflut: Volkszählungen führten im 19. und Anfang des 20. Jahrhunderts zu Bergen an Papier, die meist händisch sortiert werden mussten.



# Daten zum Stapeln

Für Wissenschaftshistoriker ist Big Data kein neues Phänomen. Auch im 18. und 19. Jahrhundert sammelten Wissenschaft und staatliche Stellen enorme Datenmengen. Und schon damals war es eine Herausforderung, sie sinnvoll auszuwerten. Die Gruppe um **Elena Aronova**, **Christine von Oertzen** und **David Sepkoski** vom **Max-Planck-Institut für Wissenschaftsgeschichte** in Berlin untersucht, mit welchen – teils überraschenden – Methoden in der Vergangenheit gearbeitet wurde und wie der geänderte Umgang mit Daten auch Wissenschaft und Gesellschaft veränderte.

TEXT **TINA HEIDBORN**

**P**reußen, Mitte des 19. Jahrhunderts: In der Berliner Zensusbehörde liest ein Angestellter die Erhebungslisten der aktuellen Volkszählung laut vor. An einem großen Tisch hören die Auszähler aufmerksam zu, jeder ist für eine eigene Rubrik zuständig. Die Striche, die jeder dabei in seinem Abschnitt macht, werden am Ende zusammengezählt, so entsteht eine neue riesige Liste, die als Statistik veröffentlicht wird. Dieses sogenannte Ausstricheln ist ein langwieriges, teures und fehleranfälliges Verfahren.

Rund zwanzig Jahre später: In einer privaten Wohnung im Prenzlauer Berg sortiert und erfasst die Ehefrau eines Statistikangestellten die Zählkarten der

aktuellen Volkszählung. Die Karten hat das preußische Statistikbüro in großen Holzkisten mit 5000 oder 10000 Stück angeliefert, jetzt werden sie in der bürgerlichen guten Stube nach einem genau festgelegten Plan sorgfältig in Häufchen sortiert. Die Frau des Hauses hat sich eine Dienstbotin genommen, um diese Heimarbeit übernehmen zu können. Gemeinsam mit ihr verdienen sich ihre beiden Schwestern und ein Schwager, ein arbeitsloser Kaufmann, sowie zwei Witwen und zwei unverheiratete Fräulein aus der Nachbarschaft bei der Auswertung der Volkszählung zusätzliches Geld. Sie arbeiten mehr als zehn Stunden pro Tag, sieben Tage die Woche. Für die Wissenschaftshistorikerin Christine von Oertzen liegt zwischen



» Die Daten wurden aus den starren Listen gelöst, sie lernten sozusagen laufen. Das war der Beginn der modernen Datenverarbeitung.

den beiden Szenen ein entscheidender Sprung in der Geschichte des Umgangs mit Massendaten.

„Unter dem Begriff Daten kann man sehr Unterschiedliches fassen, damals wie heute. Auffallend ist, dass die preußische Verwaltung in den 1860er-Jahren ihr Volkszählungsverfahren umstellte. Und dabei benutzte sie erstmals einen speziellen Datenbegriff, er taucht

hier in den Quellen auf“, erklärt von Oertzen. „Die Verwaltung entwickelte eine Definition, was sie unter Daten versteht.“ Es war Ernst Engel, ab 1860 Direktor des Königlich Preußischen Statistischen Bureaus, der in dieser Zeit eine wichtige konzeptionelle Unterscheidung aufmachte: Er differenzierte zwischen dem „Urmaterial“, das bei den Volkszählungen in sogenannten Erfas-

sungslisten gesammelt wurde, und der Weiterverarbeitung in „Tabellen“. Eine Tabelle, so schrieb der Direktor, „enthält schon ein concentrirtes Resultat, eine Zusammenfassung und Gruppierungen der Angaben aus den Listen“. Engel war einer der führenden Köpfe der entstehenden wissenschaftlichen Statistik in Europa und führte 1867 in Preußen nach italienischem Beispiel das „Zählblättchen“ ein. Dieses erleichterte die Weiterverarbeitung des Urmaterials in Tabellen wesentlich: Von den Erhebungslisten wurden die gesammelten Informationen jetzt auf diese handlichen kleinen Karten übertragen, die entfernt an Spielkarten erinnerten.

Das Zählblättchen ermöglichte im Wortsinn einen neuartigen Zugriff auf die Informationen der Erhebungslisten: Die Blättchen waren beweglich. Sie konnten leicht aus- und nachgezählt oder nach verschiedenen Kriterien neu gestapelt und erfasst werden. Damit konnten unterschiedliche Informationen von den Erhebungslisten in Bezug zueinander gesetzt werden. Gerade das war ein Problem des Ausstrichelns gewesen: Für jedes Kriterium, das man aus den Erhebungslisten auswerten wollte, musste man eine neue Riesenliste erstellen. Mit dem Zählblättchen wurde es möglich, Daten zu korrelieren. Das sei, schrieb Engel 1868, „ja eben der Vorzug der Methode der Zählblättchen, dass sie unzählige Combinationen der einzelnen Daten des Inhalts der Blättchen gestattet“.

Und Ernst Engel optimierte weiter. Wenig später ersetzte er das Zählblättchen durch persönliche Zählkarten, die

**A. Volkszählung am 1. December 1871.** 202

Serzogthum Rauenburg.

Ort, Gemeinde \_\_\_\_\_  
 Straße oder Platz \_\_\_\_\_ Haus Nr. \_\_\_\_\_  
 Zählbezirk Nr. \_\_\_\_\_ Zählbrief Nr. \_\_\_\_\_ Zählkarte Nr. \_\_\_\_\_

*Man wolle vor Beantwortung der angelegten Fragen die Anleitung D. verlesen.*

1. Vor- und Familiennamen: \_\_\_\_\_
2. Geschlecht: \_\_\_\_\_
3. Geburtsort: \_\_\_\_\_  
 Kreis: \_\_\_\_\_ Staat: \_\_\_\_\_
4. Geburtstag und Geburtsjahr: \_\_\_\_\_
5. Familienstand: \_\_\_\_\_
6. Religionsbekenntniß: \_\_\_\_\_
7. Stand, Rang, Beruf, Erwerbszweig; Arbeits- oder Dienstverhältniß.  
 Hauptbeschäftigung: \_\_\_\_\_  
 Etwaige, mit Erwerb verbundene Nebenbeschäftigung: \_\_\_\_\_
8. Staatsangehörigkeit (Name des Staats): \_\_\_\_\_
9. Wohnort (der Personen, die für gewöhnlich nicht an der Haushaltung theilnehmen): \_\_\_\_\_  
 Kreis: \_\_\_\_\_ Staat: \_\_\_\_\_
10. Schulbildung, d. h. kann lesen und schreiben? \_\_\_\_\_
11. Besondere, die Bildungs- oder Erwerbsfähigkeit beeinträchtigende Mängel:  
 blind? \_\_\_\_\_ taubstumm? \_\_\_\_\_ blödsinnig? \_\_\_\_\_ irrsinnig? \_\_\_\_\_

Unscheinbare Revolution: Die preußische Zählkarte veränderte die Datenverarbeitung im 19. Jahrhundert grundlegend. Ihre Premiere hatte sie 1871 bei der ersten Volkszählung im Deutschen Reich.





Drangvolle Enge: Das Bevölkerungswachstum im 19. Jahrhundert führte vor allem in großen Städten zu elenden Verhältnissen. Die verbesserte Datenauswertung bei Volkszählungen brachte ans Licht, wo die Probleme besonders groß waren.

jeder Befragte selbst auszufüllen hatte: Sie hatten etwa DIN-A5-Format und waren damit knapp viermal größer als das Zählblättchen, aber ähnlich handlich. Auf ihnen mussten die Einwohner Preußens zahlreiche Angaben zu sich selbst (Alter, Geburtsort, Familienstand, Lesefähigkeit) machen. Damit sparte Engel die Erhebungslisten und den Zwischenschritt der manuellen Datenübertragung auf das Zählblättchen ein.

### ZUM ERSTEN MAL WURDEN MISSTÄNDE SICHTBAR

„Die preußischen Statistiker waren ganz begeistert von der neuen Möglichkeit, unterschiedliche Kriterien zu kombinieren“, sagt die Historikerin Christine von Oertzen. Sie gingen dazu über, die Kärtchen in drei Auszahlungsdurchgängen auszuwerten, jeweils nach mehreren Kriterien. Jetzt konnte man etwa gezielt verwitwete katholische Frauen auf

dem Land erfassen oder alleinstehende evangelische Arbeiter in Kleinstädten. Genau das war das Ziel des ambitionierten Statistikers Engel gewesen: Er wollte ein Verfahren, das nicht nur das Auszählen verbesserte, sondern das Material für eine weiterführende Lesbarkeit aufbereitete. „Heutzutage ist es schwierig, sich bewusst zu machen, was für ein Entwicklungsschritt das war“, sagt von Oertzen. Ein Durchbruch, der eine bis dahin unerreichbare Differenzierung der Datenauswertung ermöglichte. „Die preußische Volkszählung sollte eine Momentaufnahme sein und als Gegenwartsbeschreibung dienen.“ Jetzt wurden zum ersten Mal Missstände wie die hohe Kindersterblichkeit in der gesammelten Materialmasse sichtbar. Oder man schlüsselte die Angaben danach auf, wo besonders viele Menschen, die nicht miteinander verwandt waren, unter einem Dach lebten – auch dies ein Armutsindikator.

Als Wissenschaftshistorikerin interessiert sich Christine von Oertzen besonders für die Entwicklung von Technologien und für deren konkrete Anwendungen. In der Umstellung von Listen auf bewegliche Papierdatenträger wie Zählblättchen und Zählkarte in Preußen nach 1860 sieht sie eine bisher wenig beachtete Revolution im Umgang mit Daten: „Die Daten wurden aus den starren Listen gelöst, sie lernten sozusagen laufen. Das ist der Beginn der modernen Datenverarbeitung, nicht die Einführung von Hollerith-Maschinen und die Mechanisierung.“ Die vermeintlich bahnbrechende Bedeutung des Hollerith-Verfahrens hält sie für überschätzt.

Im Jahr 1889 hatte der Ingenieur Herman Hollerith seine Erfindung auf der Pariser Weltausstellung vorgestellt: ein maschinelles Auszählverfahren mit Lochkarten, Sortier- und Tabulationsmaschinen. 1890 wurde es erstmals bei



# SCIENTIFIC AMERICAN

[Relieved at the Post Office of New York, N. Y., as Second Class Matter. Copyrighted, 1890, by Munn & Co.]

A WEEKLY JOURNAL OF PRACTICAL INFORMATION, ART, SCIENCE, MECHANICS, CHEMISTRY, AND MANUFACTURES.

Vol. LXIII, No. 35.  
ESTABLISHED 1845.

NEW YORK, AUGUST 30, 1890.

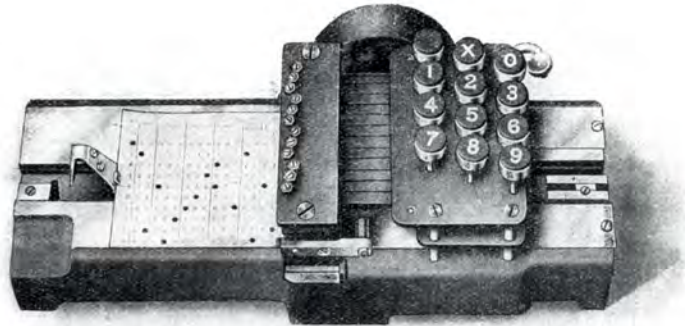
\$3.00 A YEAR.  
Weekly.



THE NEW CENSUS OF THE UNITED STATES—THE ELECTRICAL ENUMERATING MECHANISM.—[See page 132.]

**Links** Bei der US-amerikanischen Volkszählung 1890 kamen erstmals die sogenannten Hollerith-Maschinen, Tabelliermaschinen auf der Basis von Lochkarten, zum Einsatz. Das bedurfte damals noch mühsamer Vorarbeiten, denn die handgeschriebenen Informationen mussten manuell in Lochkarten gestanzt werden.

**Rechts** Erst in der Folgezeit wurden die Kartenlocher so weiterentwickelt, dass sie mit Tasten schnell bedient werden konnten – wie dieses Modell aus den 1920er-Jahren.



einer amerikanischen Volkszählung eingesetzt. Hollerith war auf die Idee gekommen, als er sah, dass Eisenbahnschaffner in Amerika beim Lochen der Fahrkarten Informationen mittransportierten: Je nachdem, wer die Fahrkarte vorzeigte, lochten sie diese an unterschiedlichen Stellen (etwa für Einstiegsort, Fahrtziel, Reiseklasse und Preis). Der Verfahrensvorteil der Hollerith-Karten bestand zu Anfang nur darin, dass sie maschinell – und damit schnell – auszulesen waren. Aber die Informationen mussten bei der Volkszählung 1890 noch manuell in die 63 Millionen Lochkarten eingestanzt werden.

### DATENAUSWERTUNG WAR WIE DAS EINBRINGEN DER ERNTE

Auch einige europäische Staaten wie Österreich-Ungarn und das Russische Reich führten Holleriths System umgehend ein. Es gilt gemeinhin als entscheidender Entwicklungsschritt moderner Datenverarbeitung.

Doch im Preußen der Jahrhundertwende hielt man das eigene, manuelle Verfahren für mindestens gleichwertig. In der Tat stand eine preußische Zählkarte der ursprünglichen Hollerith-Karte als Technologie nicht wesentlich nach, lautet die Analyse Christine von Oertzens. Immerhin hatten die preußi-

schen Statistiker mit dem Prinzip der Kartenauszahlung schon 20 Jahre zuvor eine wesentliche Grundlage des Informationszeitalters geschaffen.

Die Einführung von Zählblättchen und Zählkarte eröffnete der preußischen Verwaltung außerdem die Möglichkeit, die Datenauswertung im Wortsinne auszulagern: Sie wurde zu einer typisch weiblichen Heimarbeit. Der Staat übertrug sie den Ehefrauen seiner Zensusbediensteten, die er dafür in Haftung nahm. Bei Schlampigkeit musste nachgearbeitet werden, Lohnkürzungen drohten. Bei ihrer Archivrecherche ist die Max-Planck-Forscherin auf das Beispiel des oben erwähnten, ungewöhnlich großen Zählteams im Prenzlauer Berg gestoßen. „Die Daten mussten schnell ausgewertet werden, es war wie Saisonarbeit, wie das Einbringen der Ernte“, erklärt sie. „Daten kommen uns ja oft unphysisch vor.“ Doch als sich von Oertzen in die preußische Datenverarbeitungsgeschichte um 1900 vergrub, bekamen die Daten „Hände und Füße“, wie sie es nennt, und wurden fassbar: als Millionen von Kärtchen, die in Berlin zwischen dem Zensusbüro und ausgewählten Privatwohnungen hin- und hergeschickt wurden.

Sozialpolitisch führten die preußischen Statistiker um 1900 übrigens gern an, dass die neuartigen Hollerith-

Maschinen Menschen die Arbeit wegzunehmen drohten. Emil Blenck, der Amtsnachfolger von Ernst Engel, betonte stets eine Fürsorgepflicht gegenüber Kriegsveteranen, die traditionell mit Hilfsarbeiten für das Zensusbüro versorgt worden waren – und ließ dabei unter den Tisch fallen, dass mittlerweile nicht mehr verarmte Veteranen, sondern mittelständische Ehefrauen die Hauptarbeit erledigten.

### ES GALT, DIE MEHRDEUTIGE WIRKLICHKEIT ZU ORDNEN

Beim Sortieren, Stapeln und Auszählen in ihren guten Stuben standen die Frauen vor einem Grunddilemma jeglicher Datenverarbeitung: die vielgestaltige, oft mehrdeutige Wirklichkeit in vermeintlich trennscharfe Statistik-kategorien zu pressen. So sollte zum Beispiel bei der preußischen Volkszählung am 1. Dezember 1890 angegeben werden: „Verwandtschaft oder sonstige Stellung zum Haushalts-Vorstand“. Die Antworten kamen nicht nur in Millionen unterschiedlichen, teils schwer lesbaren Handschriften an, sondern auch in einer riesigen Bandbreite. Schließlich musste hier jeder Befragte in eigener Einschätzung und Formulierung etwas eintragen. Die Frauen sollten die eingehenden





Frauen erledigten oft die Datenerfassung, wie hier im US-amerikanischen Zensusbüro 1908. Klavierspielerinnen wurden bevorzugt, weil sie die Stanztastatur schnell und fehlerfrei bedienen konnten.

Antworten für die preußische Statistik in sieben Rubriken klassifizieren. So wollte das Zensusbüro etwa Pflinglinge und Pensionäre in einer Kategorie ausgezählt haben ebenso wie „einquartierte Soldaten“ oder die damals nicht seltenen „Schlafgänger“ – Nachtarbeiter, die sich tagsüber ein dann ungenutztes privates Bett mieteten. „Die Frauen mussten zum Zählen sortieren, und das war alles andere als stumpfes Abarbeiten“, schildert Christine von Oertzen diesen elementaren Arbeitsgang. „Da steckte jede Menge Interpretation und Analyse drin. Fleiß und Zuverlässigkeit reichten nicht, die Frauen mussten auch relativ gut gebildet sein, um die Informationen richtig zuzuordnen zu können.“

### **BIG DATA BRAUCHT EBENFALLS MENSCHLICHE ARBEIT**

Die Zensusbehörde legte ein Hilfsblatt mit Antwortbeispielen bei. Das macht deutlich, wie schwierig es war, die Angaben in Rubriken einzupassen. Bei der Stellung zum Haushalts-Vorstand sollte die Statistik „Rubrik 2: Dienstboten des Haushalts-Vorstands“ und „Rubrik 3: Gewerbs- und Arbeits-helfen des Haushalts-Vorstands“ voneinander getrennt ausweisen. Die

Beispiele in der Anleitung informierten die Auswerterinnen, dass unter Rubrik 2 etwa das Arbeitsmädchen auf dem Land, die Erzieherin, Gesellschafterin, „Gehülfin des Haushalts“, Haushälterin, Haushaltsstütze und Magd, außerdem Knecht und Kutscher erfasst werden sollten. Während unter Rubrik 3 etwa aufzuführen waren: „Arbeiter, Hofmeister, Lehmmädchen, Mamsell“, aber auch eine so allgemeine Angabe wie „in Arbeit“. Wieso kam, wer sich als „Haushälterin“ bezeichnete, in Rubrik 2; wer angab, „Mamsell“ zu sein, in Rubrik 3?

„Es gibt ja diese Vorstellung, dass sich das mit den Daten von allein macht, weil in ihnen schon alles drinsteckt. Dass man sie nur noch zählen muss und dass das leicht ist. Das halte ich für eine Illusion“, sagt Christine von Oertzen. Ihre Untersuchung zeigt plastisch, wie viel Analyse das Auswerten von gesammelten Daten vor mehr als 100 Jahren brauchte. Und heute im viel beschworenen Zeitalter von Big Data? „Natürlich interessiert uns die Frage nach Kontinuitäten beziehungsweise Brüchen“, sagt die Historikerin. Der Digitalisierung zum Trotz – auch bei Big Data heute, zu Beginn des 21. Jahrhunderts, sei viel menschliche Arbeit im Spiel: Um die Datenmassen

kompatibel zu machen, zu pflegen und nutzbar zu halten. „Das sind Dinge, die wir heute so gern vergessen“, meint von Oertzen.

### **AUCH FRÜHER GLAUBTE MAN, DIE MASSE MACHT'S**

Und die Annahme, dass Daten zum neuen Rohstoff werden? Dass in der digitalisierten Welt zukünftig Daten den Weg wissenschaftlicher Innovation weisen werden? „Manchmal ist ja die Behauptung zu hören, dass wissenschaftliche Forschung sehr viel stärker datengetrieben sein wird“, sagt die Forscherin. Statt Hypothesen aufzustellen und zu überprüfen, müsse man sich nur an die auflaufenden Daten halten. Da ist sie durch ihre Beschäftigung mit historischen Massendaten-Beispielen eher skeptisch geworden.

Auch die vermeintlichen Vollständigkeitsträume, die in Zeiten von Big Data wieder verstärkt geträumt werden, kommen Christine von Oertzen bekannt vor. „Im 19. Jahrhundert gab es eine große Begeisterung und den Glauben an die vollständige Erfassbarkeit der Wirklichkeit durch Daten“, sagt sie. Auch damals versuchten Wissenschaftler in den unterschiedlichsten Disziplinen, massenhaft Einzeler-

» Es gibt die Vorstellung, dass sich das mit den Daten von allein macht, weil in ihnen schon alles drinsteckt. Das ist eine Illusion.

eignisse zu erfassen, auf der Suche nach einem Gesamtbild in Astronomie, Sprachwissenschaft, Evolutionsbiologie oder auch in der Taxonomie. Für viele Forschungsprojekte galt schon damals: Die Masse macht's.

Allerdings machte die Masse häufig auch Probleme. Bibliotheken und Gelehrte benutzten Zettelkästen, um die massiv wachsenden Informationen handhabbar zu machen. David Sepkoski, Mitorganisator der Arbeitsgruppe, geht dieser historischen Entwicklung am Beispiel von Zoologie und Paläontologie nach. Er verfolgt darin zurück, wie die im 19. Jahrhundert entstehende Paläontologie über einen langen Zeitraum Klassifikationssysteme für Fossilien entwickelte und wie Wissenschaftler – lange vor dem Aufkommen von Computern – die Informationen über ausgestorbene Lebewesen vergangener Zeiten auf Papier ordneten und speicherten. Der Paläontologe Heinrich Georg Bronn (1800 bis 1862) etwa nutzte vorhandene Kataloge und Kompendien und bereitete deren Datenmasse neu auf: Er wertete sie quantitativ aus und strukturierte sie entsprechend den eigenen wissenschaftlichen Hypothesen neu. Für seine Bücher erstellte er Schaubilder und Diagramme, welche die Entstehung, Verbreitung und die Diversifizierung sowie das Aussterben von Gattungen auf einen Blick vermittelten. Das System seiner Aufbereitung auf Papier diente später als Modell für elektronische und digitale paläontologische Datenbanken.

Auch in Disziplinen wie der beobachtenden Astronomie, deren Kernbestandteil schon immer das Sammeln von Daten gewesen war, schwollen die

Datenmengen immer weiter an, etwa durch neue technische Möglichkeiten wie das Abfotografieren des Sternhimmels oder elektronische und schließlich digitale Supertelekope. Die eigentliche wissenschaftliche Arbeit verlagerte sich dabei immer mehr auf das schwierige Zusammenbringen unterschiedlicher Datenformate und die Analyse wie auch sinnvolle Korrelation der gesammelten Informationen. Daten zu teilen und zu zirkulieren, entwickelte sich so zum Kerngeschäft der Astronomie und veränderte die Kultur der gesamten Disziplin.

### HEUTE LASSEN SICH DATEN AUS DEM KONTEXT LÖSEN

Regelrecht zur politischen Tauschewährung wurden geophysische Großdaten zu Zeiten des Kalten Kriegs, wie Elena Aronova in der Berliner Arbeitsgruppe herausgefunden hat. Amerikanische und sowjetische Datenzentren sammelten und archivierten Massen an Material in analoger Form. Die Vision, dieses Material Wissenschaftlern in Ost und West tatsächlich frei zur Verfügung zu stellen, ließ sich allerdings nur zum Teil verwirklichen. Sie scheiterte aber nicht nur an politischen Zwän-

gen, sondern auch an den technischen Grenzen analoger Speichermedien.

Historisch neu im Digitalzeitalter ist in den Augen der Wissenschaftshistoriker die Möglichkeit, Daten ganz aus ihrem ursprünglichen Kontext zu lösen. Einmal gesammelte und digitalisierte Informationen sind nicht mehr – wie noch in den Datenzentren des Kalten Krieges – lokal verortet und können heute, völlig aus ihrem ursprünglichen Zusammenhang gelöst, weiter genutzt werden. So wie beim Pima Data Set: Die medizinischen Daten von Angehörigen eines amerikanischen Indianerstamms waren ursprünglich mit Einverständnis der Betroffenen erfasst worden, um Übergewicht und Diabetesneigung in dieser Gruppe zu untersuchen. Mittlerweile ist die Datensammlung im Internet frei zugänglich und wird vorwiegend als Lerndatensatz zur Optimierung von computerbasiertem maschinellem Lernen genutzt.

Die Erfassung der Welt durch Daten wirft neue Probleme auf und hat durch die Digitalisierung heutzutage auch neue Dimensionen erreicht. Doch beim Rückblick auf die Datenpraktiken der Vergangenheit wird klar, wie alt die Fundamente sind, die unsere heutige Verdatung prägen. ◀

### AUF DEN PUNKT GEBRACHT

- Auch im 18. und 19. Jahrhundert sammelten Wissenschaftler große Datenmengen in der Erwartung, damit die Wirklichkeit zu erfassen. Die wissenschaftliche Arbeit verlagerte sich zunehmend auf die Analyse der Daten.
- Das preußische Statistikkamt revolutionierte Mitte des 19. Jahrhunderts mit Zählkarten die Datenauswertung. Damit konnten Daten nach unterschiedlichen Kriterien korreliert und unbekanntes Zusammenhänge aufgedeckt werden.