

# Visualisierung

Matthias Hammerl

Fachhochschule Wiener Neustadt



# Plan

- 1 Modus
- 2 Einleitung und Historisches
- 3 Moderner Zugang und Grammatik der Visualisierung



# Warum Visualisierung?



*“Wir können uns dasjenige am deutlichsten vorstellen, was sich uns durch die Wahrnehmung unserer Sinne mitgeteilt und eingeprägt hat; der schärfste von all unseren Sinnen ist aber der Gesichtssinn. Deshalb kann man etwas am leichtesten behalten, wenn das, was man durch das Gehör oder durch Überlegung aufnimmt, auch noch durch die Vermittlung der Augen ins Bewusstsein dringt.”*

*Marcus Tullius Cicero: De Oratore (55 v. Chr.)*

# Gehirn ist mit sequentiellen Daten schnell überfordert



*Lewis Carroll: Alice in Wonderland*

**Queen:** Can you do Addition? What's one and one and one and one and one and one and one and one and one and one and one?

**Alice:** I don't know. I lost count.

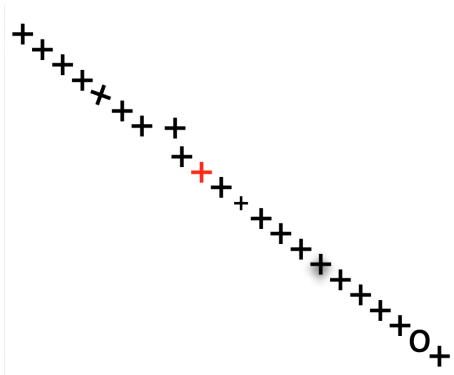
**Queen:** She can't do Addition!

# Evolutionär trainiert: Muster und Gestalterkennung



- Bemerkung: Visueller Cortex macht ca. 25 % des Cortex aus

# Visualisierung von Daten nutzt Mustererkennung des Gehirns



Das visuelle System des Menschen erkennt unter anderem...

- Lücken
- Ausreißer
- Trends und Muster

# Warum Visualisierung?

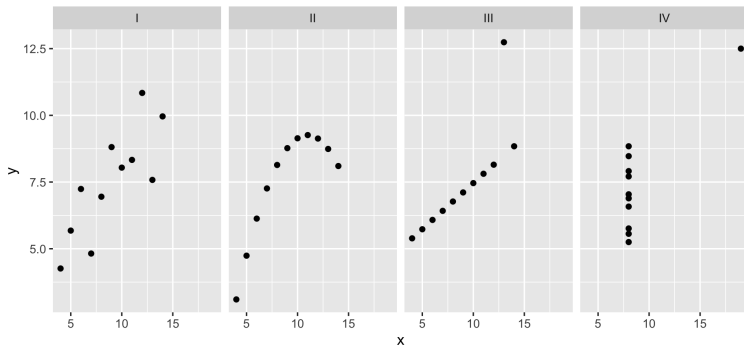
| I    |       | II   |      | III  |       | IV   |       |
|------|-------|------|------|------|-------|------|-------|
| x    | y     | x    | y    | x    | y     | x    | y     |
| 10.0 | 8.04  | 10.0 | 9.14 | 10.0 | 7.46  | 8.0  | 6.58  |
| 8.0  | 6.95  | 8.0  | 8.14 | 8.0  | 6.77  | 8.0  | 5.76  |
| 13.0 | 7.58  | 13.0 | 8.74 | 13.0 | 12.74 | 8.0  | 7.71  |
| 9.0  | 8.81  | 9.0  | 8.77 | 9.0  | 7.11  | 8.0  | 8.84  |
| 11.0 | 8.33  | 11.0 | 9.26 | 11.0 | 7.81  | 8.0  | 8.47  |
| 14.0 | 9.96  | 14.0 | 8.10 | 14.0 | 8.84  | 8.0  | 7.04  |
| 6.0  | 7.24  | 6.0  | 6.13 | 6.0  | 6.08  | 8.0  | 5.25  |
| 4.0  | 4.26  | 4.0  | 3.10 | 4.0  | 5.39  | 19.0 | 12.50 |
| 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15  | 8.0  | 5.56  |
| 7.0  | 4.82  | 7.0  | 7.26 | 7.0  | 6.42  | 8.0  | 7.91  |
| 5.0  | 5.68  | 5.0  | 4.74 | 5.0  | 5.73  | 8.0  | 6.89  |

*F. J. Anscombe: Graphs in Statistical Analysis (1973)*

- Deskriptive Statistik:

|                   | x    | y     |
|-------------------|------|-------|
| <b>Mittelwert</b> | 9,00 | 11,00 |
| <b>Varianz</b>    | 7,50 | 4,13  |

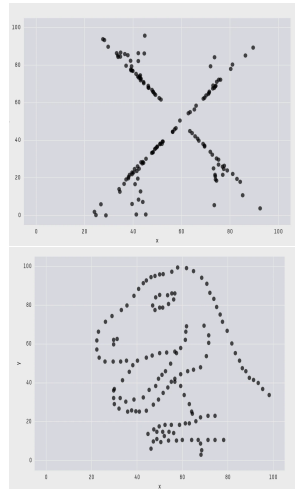
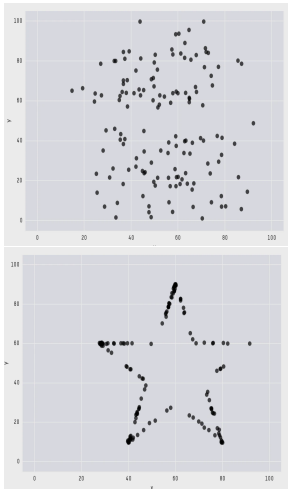
# Visualisierung macht Daten verständlich



*Plot der Anscombe-Datasets I-IV*

- Deskriptive Statistik:

|                   | x    | y     |
|-------------------|------|-------|
| <b>Mittelwert</b> | 9,00 | 11,00 |
| <b>Varianz</b>    | 7,50 | 4,13  |



*(Matejka and Fitzmaurice, Autodesk Research, 2017)*

→ 'Datasaurus Dozen': Tatsächlich 1-Parameter-Familie von 2-dimensionalen Daten mit gleichen statistischen Kenngrößen.

## Moderne Visualisierung: Teil der explorativen Datenanalyse



*"The greatest value of a picture is when it forces us to notice what we never expected to see."*

*John Tukey: Explorative Data Analysis (1977)*

- Gemäß Tukey sind die Ziele von explorativer Datenanalyse:
  - Unerwartete Entdeckungen und Einsichten gewinnen
  - Ideen für die Generierung von Hypothesen bilden
  - Die Auswahl von geeigneten statistischen Methoden unterstützen



# Warum Visualisierung?: Einsatzzwecke



- **Explorative Datenanalyse (EDA)**
  - Durch Visualisierung von Daten sollen Thesen entwickelt werden, welche anschließend durch statistische Modelle überprüft werden können
  - ↳ Generierung von Thesen
  - ↳ Validierung von Thesen
- **Visualisierung von Informationen (Informationsvisualisierung):**
  - Durch Visualisierung von Daten soll Information vermittelt werden

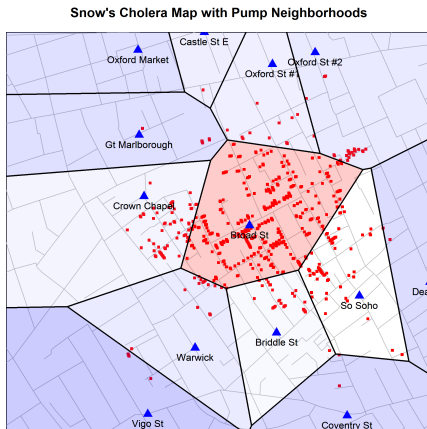
# John Snow: Ursache der Cholera



*John Snow: On the Mode of Communication of Cholera (1855)*

- Karte von Soho während des Ausbruchs von 1854:
- ◇ Balken vor den Adressen für gezählte Todesfälle

# John Snow: Ursache der Cholera (moderne Variante)



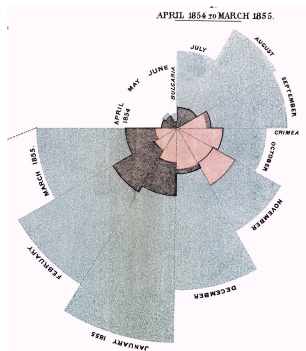
*Modern Version of John Snow's map / recreated in R (Friendly and Wainer, 2021)*

- Aus den 13 Koordinatenpunkten der Brunnen wird ein Voronoi-Diagramm definiert, welches die Ebene je nachdem welcher Brunnen als nächstes ist partitioniert.

## Florence Nightingale: Todesursachen im Krimkrieg

- Untersuchung zu den Todesursachen im Krimkrieg (1853-1856):

Vorwiegend schlechte sanitäre Verhältnisse verantwortlich



*Florence Nightingale: Causes of mortality in the army in the East (1858)*

- ◇ Rot: An Kriegsverletzungen gestorben
- ◇ Blau: An behandelbaren Krankheiten verstorben
- ◇ Grau: Alle sonstigen Todesursachen

# Historische Pioniere der Datenvisualisierung

- Joseph Priestley (1733-1804): *Visualisierung von historischen Zeitperioden / Chronographie*
- William Playfair (1759-1823): *Entwicklung von Liniendiagramm, Flächendiagramm, Balkendiagramm, Kreisdiagramm*
- Charles Joseph Minard (1781-1870): *Entwicklung von Flussdiagrammen*
- Florence Nightingale (1820-1910): *Weiterentwicklung und Einsatz des Polar-Area-Diagramms*
- Otto Neurath (1882-1945): *Visualisierung von sozialen und ökonomischen Größen mit Isotypen*

## Zeitalter vor Visualisierung



256

RENT OF LAND

[BK. I]

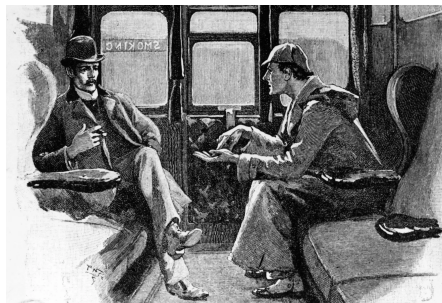
[CH. XI]

PRICES OF WHEAT

257

| Wheat per quarter. | Year | Wheat per quarter. | Year |      |   |    |   |
|--------------------|------|--------------------|------|------|---|----|---|
| £                  | s.   | d.                 | Year |      |   |    |   |
| 1700               | 2    | 13                 | 0    | 1700 | 1 | 17 | 6 |
| 1701               | 2    | 17                 | 4    | 1701 | 1 | 9  | 0 |
| 1702               | 2    | 4                  | 10   | 1702 | 1 | 10 | 0 |
| 1703               | 2    | 4                  | 8    | 1703 | 2 | 0  | 0 |
| 1704               | 0    | 0                  | 0    | 1704 | 1 | 10 | 0 |
| 1705               | 0    | 0                  | 0    | 1705 | 1 | 6  | 0 |
| 1706               | 0    | 0                  | 0    | 1706 | 1 | 9  | 0 |
| 1707               | 0    | 0                  | 0    | 1707 | 2 | 1  | 0 |
| 1708               | 0    | 0                  | 0    | 1708 | 3 | 10 | 0 |
| 1709               | 0    | 0                  | 0    | 1709 | 3 | 10 | 0 |
| 1710               | 0    | 0                  | 0    | 1710 | 3 | 10 | 0 |
| 1711               | 0    | 0                  | 0    | 1711 | 3 | 10 | 0 |
| 1712               | 0    | 0                  | 0    | 1712 | 3 | 10 | 0 |
| 1713               | 0    | 0                  | 0    | 1713 | 3 | 10 | 0 |
| 1714               | 0    | 0                  | 0    | 1714 | 3 | 10 | 0 |
| 1715               | 0    | 0                  | 0    | 1715 | 3 | 10 | 0 |
| 1716               | 0    | 0                  | 0    | 1716 | 3 | 10 | 0 |
| 1717               | 0    | 0                  | 0    | 1717 | 3 | 10 | 0 |
| 1718               | 0    | 0                  | 0    | 1718 | 3 | 10 | 0 |
| 1719               | 0    | 0                  | 0    | 1719 | 3 | 10 | 0 |
| 1720               | 0    | 0                  | 0    | 1720 | 3 | 10 | 0 |
| 1721               | 0    | 0                  | 0    | 1721 | 3 | 10 | 0 |
| 1722               | 0    | 0                  | 0    | 1722 | 3 | 10 | 0 |
| 1723               | 0    | 0                  | 0    | 1723 | 3 | 10 | 0 |
| 1724               | 0    | 0                  | 0    | 1724 | 3 | 10 | 0 |
| 1725               | 0    | 0                  | 0    | 1725 | 3 | 10 | 0 |
| 1726               | 0    | 0                  | 0    | 1726 | 3 | 10 | 0 |
| 1727               | 0    | 0                  | 0    | 1727 | 3 | 10 | 0 |
| 1728               | 0    | 0                  | 0    | 1728 | 3 | 10 | 0 |
| 1729               | 0    | 0                  | 0    | 1729 | 3 | 10 | 0 |
| 1730               | 0    | 0                  | 0    | 1730 | 3 | 10 | 0 |
| 1731               | 0    | 0                  | 0    | 1731 | 3 | 10 | 0 |
| 1732               | 0    | 0                  | 0    | 1732 | 3 | 10 | 0 |
| 1733               | 0    | 0                  | 0    | 1733 | 3 | 10 | 0 |
| 1734               | 0    | 0                  | 0    | 1734 | 3 | 10 | 0 |
| 1735               | 0    | 0                  | 0    | 1735 | 3 | 10 | 0 |
| 1736               | 0    | 0                  | 0    | 1736 | 3 | 10 | 0 |
| 1737               | 0    | 0                  | 0    | 1737 | 3 | 10 | 0 |
| 1738               | 0    | 0                  | 0    | 1738 | 3 | 10 | 0 |
| 1739               | 0    | 0                  | 0    | 1739 | 3 | 10 | 0 |
| 1740               | 0    | 0                  | 0    | 1740 | 3 | 10 | 0 |
| 1741               | 0    | 0                  | 0    | 1741 | 3 | 10 | 0 |
| 1742               | 0    | 0                  | 0    | 1742 | 3 | 10 | 0 |
| 1743               | 0    | 0                  | 0    | 1743 | 3 | 10 | 0 |
| 1744               | 0    | 0                  | 0    | 1744 | 3 | 10 | 0 |
| 1745               | 0    | 0                  | 0    | 1745 | 3 | 10 | 0 |
| 1746               | 0    | 0                  | 0    | 1746 | 3 | 10 | 0 |
| 1747               | 0    | 0                  | 0    | 1747 | 3 | 10 | 0 |
| 1748               | 0    | 0                  | 0    | 1748 | 3 | 10 | 0 |
| 1749               | 0    | 0                  | 0    | 1749 | 3 | 10 | 0 |
| 1750               | 0    | 0                  | 0    | 1750 | 3 | 10 | 0 |
| 1751               | 0    | 0                  | 0    | 1751 | 3 | 10 | 0 |
| 1752               | 0    | 0                  | 0    | 1752 | 3 | 10 | 0 |
| 1753               | 0    | 0                  | 0    | 1753 | 3 | 10 | 0 |
| 1754               | 0    | 0                  | 0    | 1754 | 3 | 10 | 0 |
| 1755               | 0    | 0                  | 0    | 1755 | 3 | 10 | 0 |
| 1756               | 0    | 0                  | 0    | 1756 | 3 | 10 | 0 |
| 1757               | 0    | 0                  | 0    | 1757 | 3 | 10 | 0 |
| 1758               | 0    | 0                  | 0    | 1758 | 3 | 10 | 0 |
| 1759               | 0    | 0                  | 0    | 1759 | 3 | 10 | 0 |
| 1760               | 0    | 0                  | 0    | 1760 | 3 | 10 | 0 |
| 1761               | 0    | 0                  | 0    | 1761 | 3 | 10 | 0 |
| 1762               | 0    | 0                  | 0    | 1762 | 3 | 10 | 0 |
| 1763               | 0    | 0                  | 0    | 1763 | 3 | 10 | 0 |
| 1764               | 0    | 0                  | 0    | 1764 | 3 | 10 | 0 |
| 1765               | 0    | 0                  | 0    | 1765 | 3 | 10 | 0 |
| 1766               | 0    | 0                  | 0    | 1766 | 3 | 10 | 0 |
| 1767               | 0    | 0                  | 0    | 1767 | 3 | 10 | 0 |
| 1768               | 0    | 0                  | 0    | 1768 | 3 | 10 | 0 |
| 1769               | 0    | 0                  | 0    | 1769 | 3 | 10 | 0 |
| 1770               | 0    | 0                  | 0    | 1770 | 3 | 10 | 0 |
| 1771               | 0    | 0                  | 0    | 1771 | 3 | 10 | 0 |
| 1772               | 0    | 0                  | 0    | 1772 | 3 | 10 | 0 |
| 1773               | 0    | 0                  | 0    | 1773 | 3 | 10 | 0 |
| 1774               | 0    | 0                  | 0    | 1774 | 3 | 10 | 0 |
| 1775               | 0    | 0                  | 0    | 1775 | 3 | 10 | 0 |
| 1776               | 0    | 0                  | 0    | 1776 | 3 | 10 | 0 |
| 1777               | 0    | 0                  | 0    | 1777 | 3 | 10 | 0 |
| 1778               | 0    | 0                  | 0    | 1778 | 3 | 10 | 0 |
| 1779               | 0    | 0                  | 0    | 1779 | 3 | 10 | 0 |
| 1780               | 0    | 0                  | 0    | 1780 | 3 | 10 | 0 |
| 1781               | 0    | 0                  | 0    | 1781 | 3 | 10 | 0 |
| 1782               | 0    | 0                  | 0    | 1782 | 3 | 10 | 0 |
| 1783               | 0    | 0                  | 0    | 1783 | 3 | 10 | 0 |
| 1784               | 0    | 0                  | 0    | 1784 | 3 | 10 | 0 |
| 1785               | 0    | 0                  | 0    | 1785 | 3 | 10 | 0 |
| 1786               | 0    | 0                  | 0    | 1786 | 3 | 10 | 0 |
| 1787               | 0    | 0                  | 0    | 1787 | 3 | 10 | 0 |
| 1788               | 0    | 0                  | 0    | 1788 | 3 | 10 | 0 |
| 1789               | 0    | 0                  | 0    | 1789 | 3 | 10 | 0 |
| 1790               | 0    | 0                  | 0    | 1790 | 3 | 10 | 0 |
| 1791               | 0    | 0                  | 0    | 1791 | 3 | 10 | 0 |
| 1792               | 0    | 0                  | 0    | 1792 | 3 | 10 | 0 |
| 1793               | 0    | 0                  | 0    | 1793 | 3 | 10 | 0 |
| 1794               | 0    | 0                  | 0    | 1794 | 3 | 10 | 0 |
| 1795               | 0    | 0                  | 0    | 1795 | 3 | 10 | 0 |
| 1796               | 0    | 0                  | 0    | 1796 | 3 | 10 | 0 |
| 1797               | 0    | 0                  | 0    | 1797 | 3 | 10 | 0 |
| 1798               | 0    | 0                  | 0    | 1798 | 3 | 10 | 0 |
| 1799               | 0    | 0                  | 0    | 1799 | 3 | 10 | 0 |
| 1800               | 0    | 0                  | 0    | 1800 | 3 | 10 | 0 |
| 1801               | 0    | 0                  | 0    | 1801 | 3 | 10 | 0 |
| 1802               | 0    | 0                  | 0    | 1802 | 3 | 10 | 0 |
| 1803               | 0    | 0                  | 0    | 1803 | 3 | 10 | 0 |
| 1804               | 0    | 0                  | 0    | 1804 | 3 | 10 | 0 |
| 1805               | 0    | 0                  | 0    | 1805 | 3 | 10 | 0 |
| 1806               | 0    | 0                  | 0    | 1806 | 3 | 10 | 0 |
| 1807               | 0    | 0                  | 0    | 1807 | 3 | 10 | 0 |
| 1808               | 0    | 0                  | 0    | 1808 | 3 | 10 | 0 |
| 1809               | 0    | 0                  | 0    | 1809 | 3 | 10 | 0 |
| 1810               | 0    | 0                  | 0    | 1810 | 3 | 10 | 0 |
| 1811               | 0    | 0                  | 0    | 1811 | 3 | 10 | 0 |
| 1812               | 0    | 0                  | 0    | 1812 | 3 | 10 | 0 |
| 1813               | 0    | 0                  | 0    | 1813 | 3 | 10 | 0 |
| 1814               | 0    | 0                  | 0    | 1814 | 3 | 10 | 0 |
| 1815               | 0    | 0                  | 0    | 1815 | 3 | 10 | 0 |
| 1816               | 0    | 0                  | 0    | 1816 | 3 | 10 | 0 |
| 1817               | 0    | 0                  | 0    | 1817 | 3 | 10 | 0 |
| 1818               | 0    | 0                  | 0    | 1818 | 3 | 10 | 0 |
| 1819               | 0    | 0                  | 0    | 1819 | 3 | 10 | 0 |
| 1820               | 0    | 0                  | 0    | 1820 | 3 | 10 | 0 |
| 1821               | 0    | 0                  | 0    | 1821 | 3 | 10 | 0 |
| 1822               | 0    | 0                  | 0    | 1822 | 3 | 10 | 0 |
| 1823               | 0    | 0                  | 0    | 1823 | 3 | 10 | 0 |
| 1824               | 0    | 0                  | 0    | 1824 | 3 | 10 | 0 |
| 1825               | 0    | 0                  | 0    | 1825 | 3 | 10 | 0 |
| 1826               | 0    | 0                  | 0    | 1826 | 3 | 10 | 0 |
| 1827               | 0    | 0                  | 0    | 1827 | 3 | 10 | 0 |
| 1828               | 0    | 0                  | 0    | 1828 | 3 | 10 | 0 |
| 1829               | 0    | 0                  | 0    | 1829 | 3 | 10 | 0 |
| 1830               | 0    | 0                  | 0    | 1830 | 3 | 10 | 0 |
| 1831               | 0    | 0                  | 0    | 1831 | 3 | 10 | 0 |
| 1832               | 0    | 0                  | 0    | 1832 | 3 | 10 | 0 |
| 1833               | 0    | 0                  | 0    | 1833 | 3 | 10 | 0 |
| 1834               | 0    | 0                  | 0    | 1834 | 3 | 10 | 0 |
| 1835               | 0    | 0                  | 0    | 1835 | 3 | 10 | 0 |
| 1836               | 0    | 0                  | 0    | 1836 | 3 | 10 | 0 |
| 1837               | 0    | 0                  | 0    | 1837 | 3 | 10 | 0 |
| 1838               | 0    | 0                  | 0    | 1838 | 3 | 10 | 0 |
| 1839               | 0    | 0                  | 0    | 1839 | 3 | 10 | 0 |
| 1840               | 0    | 0                  | 0    | 1840 | 3 | 10 | 0 |
| 1841               | 0    | 0                  | 0    | 1841 | 3 | 10 | 0 |
| 1842               | 0    | 0                  | 0    | 1842 | 3 | 10 | 0 |
| 1843               | 0    | 0                  | 0    | 1843 | 3 | 10 | 0 |
| 1844               | 0    | 0                  | 0    | 1844 | 3 | 10 | 0 |
| 1845               | 0    | 0                  | 0    | 1845 | 3 | 10 | 0 |
| 1846               | 0    | 0                  | 0    | 1846 | 3 | 10 | 0 |
| 1847               | 0    | 0                  | 0    | 1847 | 3 | 10 | 0 |
| 1848               | 0    | 0                  | 0    | 1848 | 3 | 10 | 0 |
| 1849               | 0    | 0                  | 0    | 1849 | 3 | 10 | 0 |
| 1850               | 0    | 0                  | 0    | 1850 | 3 | 10 | 0 |
| 1851               | 0    | 0                  | 0    | 1851 | 3 | 10 | 0 |
| 1852               | 0    | 0                  | 0    | 1852 | 3 | 10 | 0 |
| 1853               | 0    | 0                  | 0    | 1853 | 3 | 10 | 0 |
| 1854               | 0    | 0                  | 0    | 1854 | 3 | 10 | 0 |
| 1855               | 0    | 0                  | 0    | 1855 | 3 | 10 | 0 |
| 1856               | 0    | 0                  | 0    | 1856 | 3 | 10 | 0 |
| 1857               | 0    | 0                  | 0    | 1857 | 3 | 10 | 0 |
| 1858               | 0    | 0                  | 0    | 1858 | 3 | 10 | 0 |
| 1859               | 0    | 0                  | 0    | 1859 | 3 | 10 | 0 |
| 1860               | 0    | 0                  | 0    | 1860 | 3 | 10 | 0 |
| 1861               | 0    | 0                  | 0    | 1861 | 3 | 10 | 0 |
| 1862               | 0    | 0                  | 0    | 1862 | 3 | 10 | 0 |
| 1863               | 0    | 0                  | 0    | 1863 | 3 | 10 | 0 |
| 1864               | 0    | 0                  | 0    | 1864 | 3 | 10 | 0 |
| 1865               | 0    | 0                  | 0    | 1865 | 3 | 10 | 0 |
| 1866               | 0    | 0                  | 0    | 1866 | 3 | 10 | 0 |
| 1867               | 0    | 0                  | 0    | 1867 | 3 | 10 | 0 |
| 1868               | 0    | 0                  | 0    | 1868 | 3 | 10 | 0 |
| 1869               | 0    | 0                  | 0    | 1869 | 3 | 10 | 0 |
| 1870               | 0    | 0                  | 0    | 1870 | 3 | 10 | 0 |
| 1871               | 0    | 0                  | 0    | 1871 | 3 | 10 | 0 |
| 1872               | 0    | 0                  | 0    | 1872 | 3 | 10 | 0 |
| 1873               | 0    | 0                  | 0    | 1873 | 3 | 10 | 0 |
| 1874               | 0    | 0                  | 0    | 1874 | 3 | 10 | 0 |
| 1875               | 0    | 0                  | 0    | 1875 | 3 | 10 | 0 |
| 1876               | 0    | 0                  | 0    | 1876 | 3 | 10 | 0 |
| 1877               | 0    | 0                  | 0    | 1877 | 3 | 10 | 0 |
| 1878               | 0    | 0                  | 0    | 1878 | 3 | 10 | 0 |
| 1879               | 0    | 0                  | 0    | 1879 | 3 | 10 | 0 |
| 1880               | 0    | 0                  | 0    | 1880 | 3 | 10 | 0 |
| 1881               | 0    | 0                  | 0    | 1881 | 3 | 10 | 0 |
| 1882               | 0    | 0                  | 0    | 1882 | 3 | 10 | 0 |
| 1883               | 0    | 0                  | 0    | 1883 | 3 | 10 | 0 |
| 1884               | 0    | 0                  | 0    | 1884 | 3 | 10 | 0 |
| 1885               | 0    | 0                  | 0    | 1885 | 3 | 10 | 0 |
| 1886               | 0    | 0                  | 0    | 1886 | 3 | 10 | 0 |
| 1887               | 0    | 0                  | 0    | 1887 | 3 | 10 | 0 |
| 1888               | 0    | 0                  | 0    | 1888 | 3 | 10 | 0 |
| 1889               | 0    | 0                  | 0    | 1889 | 3 | 10 | 0 |
| 1890               | 0    | 0                  | 0    | 1890 | 3 | 10 | 0 |
| 1891               | 0    | 0                  | 0    | 1891 | 3 | 10 | 0 |
| 1892               | 0    | 0                  | 0    | 1892 | 3 | 10 | 0 |

## Sine qua non der Visualisierung: Daten

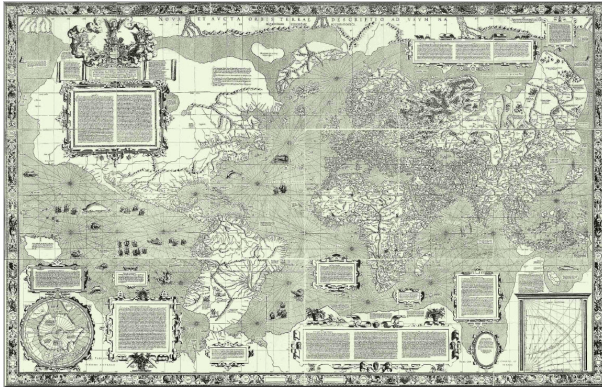


*'Data! Data! Data! I cannot  
make bricks without clay'*

Arthur Conan Doyle: *Sherlock Holmes, The Adventure of the Copper Beeches* (1892)

- Das 17. und 18. Jahrhundert sind das Zeitalter der Aufklärung und insbesondere in England die Zeit des Empirismus.
  - Es werden zunehmend von staatlicher Seite systematisch Daten gesammelt ('Statistik').
  - Anreize zur Entwicklung von Analyse und Darstellung / Visualisierung

## Historisch: Mercator-Weltkarte



*Gerardus Mercator (1596)*

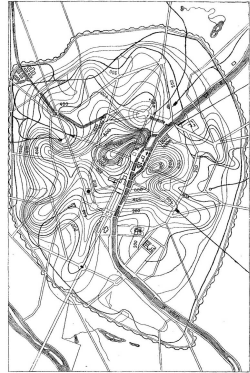
- Geraden auf der Karte = Routen entlang konstanter Kompassrichtung  
    ~ Gut geeignet für Navigation: Benötigt wird...
- ◇ Bestimmung Himmelsrichtung: Kompass
- ◇ Bestimmung Breitengrad: z.B. mit Sextant
- ◇ Bestimmung Längengrad: z.B. mit Chronometer



# Historische Beispiele Contourdiagramme



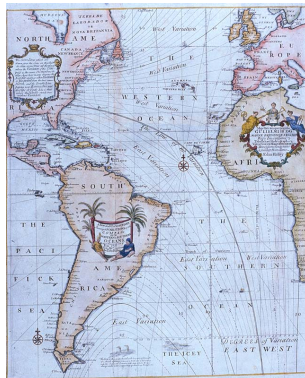
N. Cruquius: *Isobathen Merwede* (1730)



L.L. Vauthier: *Bevölkerungsdichte Paris* (1874)

- Beispiel links: Höhenlinien (Tiefenlinien) konkret  $\rightsquigarrow$  topographische Contourdiagramme, z.B. für Navigation / Atlanten
- Beispiel rechts: Höhenlinien abstrakt: entsprechen Linien gleicher Bevölkerungsdichte in Paris

## Beispiel Contourdiagramm: Magnetische Deklination

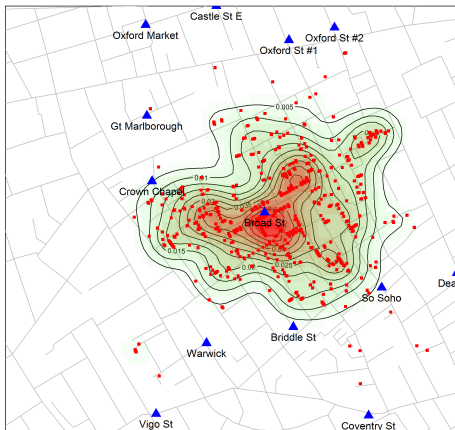


*Edmond Halley: Magnetic Variation (1702)*

- Edmond Halley sammelte seine Daten 1699-1700 auf seinen Reisen als Kapitän der HMS Paramour.
- Er benutzte vermutlich Newton-Interpolationspolynome (Methode der dividerten Differenzen), um zwischen seinen Messwerten zu interpolieren.

## John Snow: Ursache der Cholera (moderne Variante)

### Snow's Cholera Map, Death Intensity

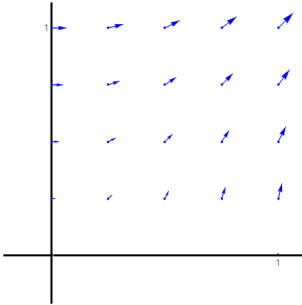


Modern Version of John Snow's map / recreated in R (Friendly and Wainer, 2021)

- Contourplots können auch für diskrete Datenpunkte geplottet werden, indem eine kontinuierliche Verteilung mittels Kernel-Density-Estimation berechnet wird.

# Vektorfelder

- **Vektorfeld:** Ordnet jedem Punkt in der Ebene einen Pfeil (Vektor) zu: Länge ('Stärke') und Richtung

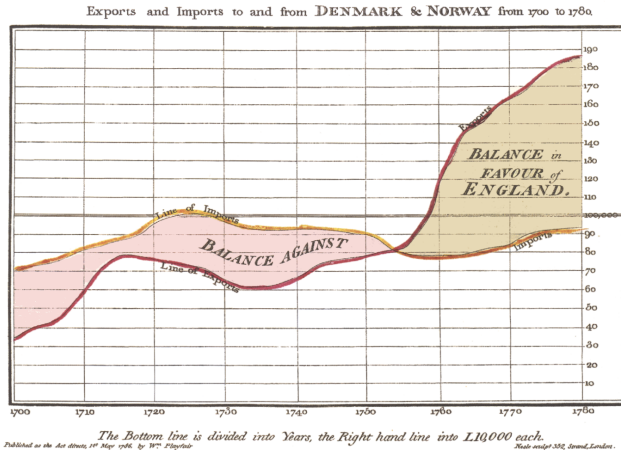


*Edmond Halley: Trade Winds, Philosophical Transactions of the Royal Society (1686)*

~~~~~ Darstellung von:

- Windrichtungen (und Stärken)
- Flussgeschwindigkeiten
- Feldstärken (Magnetisches Feld, Kraftfeld)

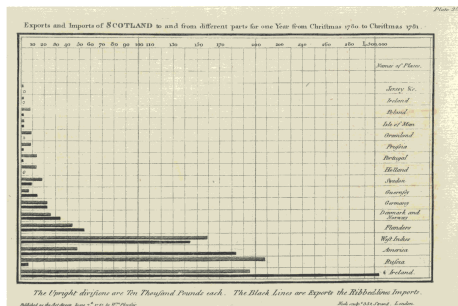
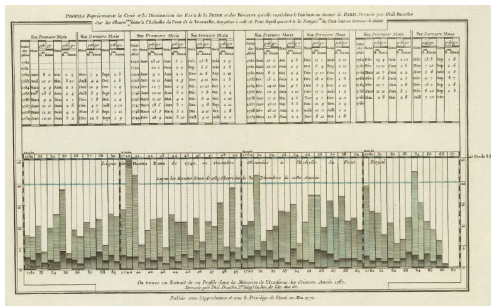
# Beginn Moderne Visualisierung: Lineplot



William Playfair: *The Commercial and Political Atlas*, London (1786)

- ◇ Orange Linie: Importe
- ◇ Rote Linie: Exporte

# Beginn Moderne Visualisierung: Barchart

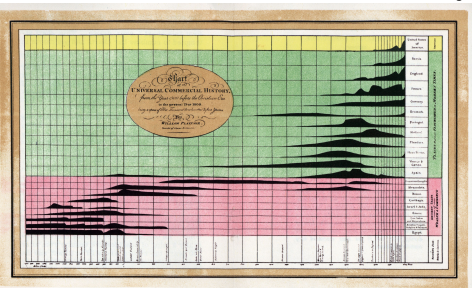


*Philippe Buache: Cartes et tables de la géographie physique ou naturelle (1754)*

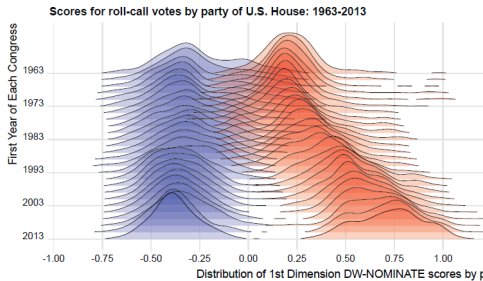
*William Playfair: The Commercial and Political Atlas, London (1786)*

- Links: Philippe Buache: Wasserstände der Seine von 1732-1766
- Rechts: William Playfair: Importe und Exporte von Schottland

# William Playfair: Ridgeplot



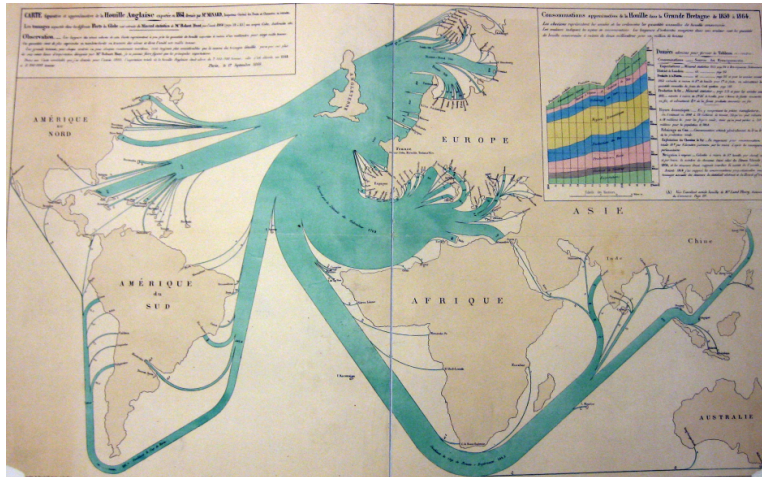
*Playfair: An Inquiry Into the Permanent Causes of the Decline and Fall of Powerful and Wealthy Nations. (1805)*



*Friendly and Wainer: Polarization of politics (2021)*

- Durch einen Ridgeplot können mehrere Linien-(bzw. Flächen-)Diagramme verglichen werden, indem sie bei gleichbleibender x-Achse und wiederholte y-Achse untereinander dargestellt werden.
- ◇ Beispiel links: Playfair's Grafik vom Aufstieg und Niedergang von Großmächten in der Geschichte
- ◇ Beispiel rechts (modern): Zunehmende Polarisierung im amerikanischen Kongress.

# Charles Joseph Minard: Flussdiagramm

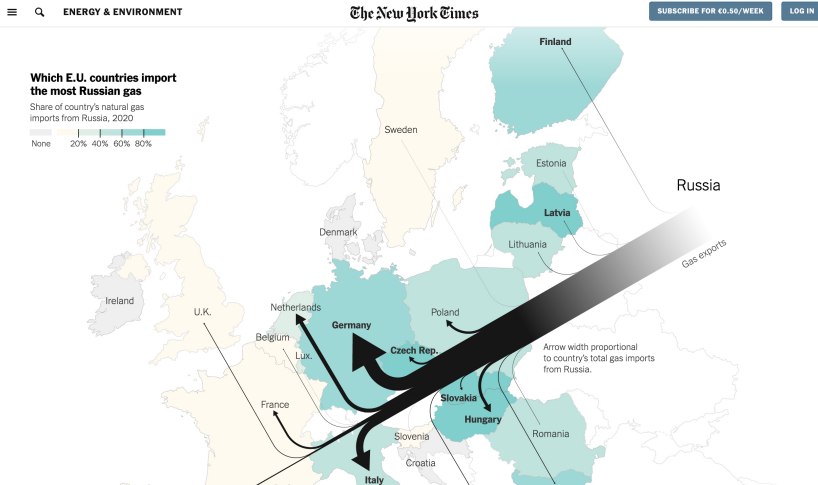


*Charles Joseph Minard: Tableaux Graphiques et des Cartes Figuratives (1864)*

- Kohleexporte aus England: 20.000 Tonnen Kohle pro mm
- Bemerkung: (Mengen-)Flussdiagramm



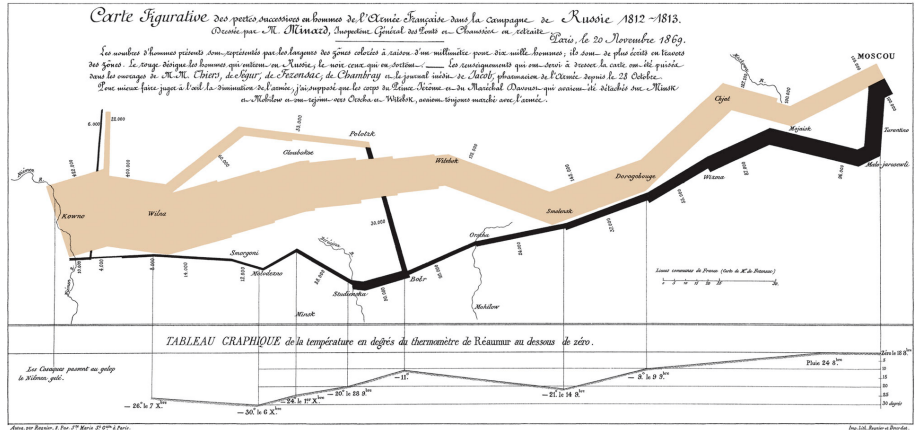
# Modernes Flussdiagramm: Russische Gas-Exporte 2020



*New York Times (2022)*

- Bemerkung: Heute Sankey-Diagramm genannt

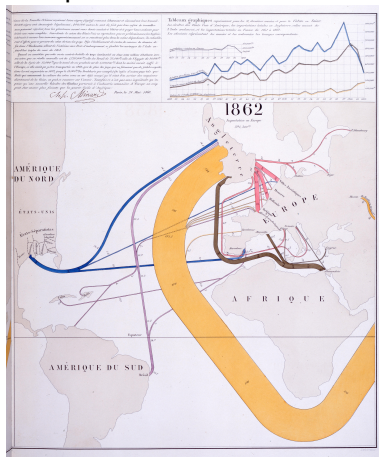
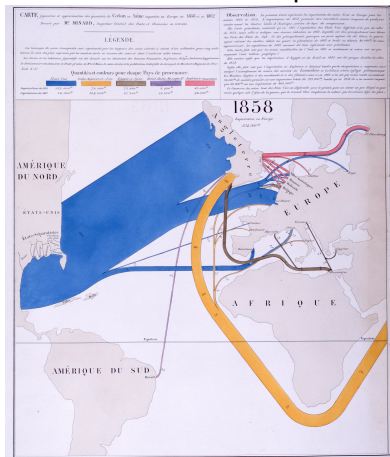
# Charles Joseph Minard: Napolenfeldzug 1812-1813



Charles Joseph Minard (1869)

- Jeder Millimeter entspricht 10.000 Soldaten:
- Beige Fläche: Einmarsch / Schwarze Fläche Rückzug
- Linie unten: Fallende Temperatur während Rückzug

# Charles Joseph Minard: Komparative Flüsse

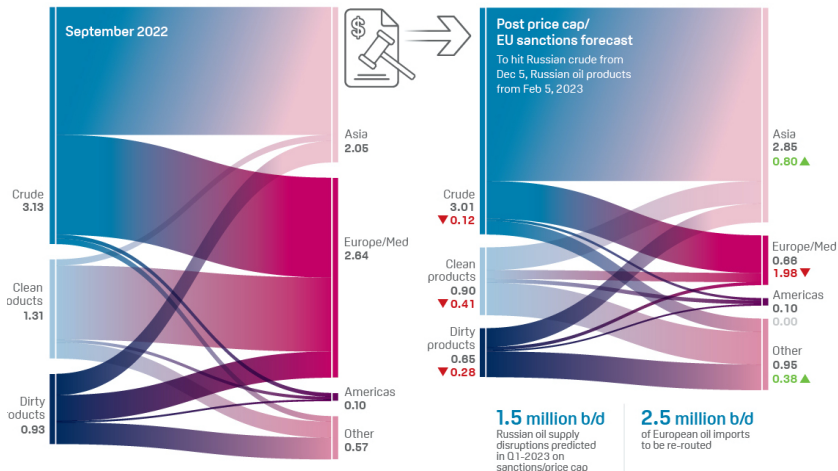


*Charles Joseph Minard: Carte figurative et approximative des quantités de coton en Europe en 1858 et 1862, Paris (1863)*

- Effekt des amerikanischen Bürgerkrieges auf europäische Baumwollimporte: Links: Bezugsquellen vor 1858 / Rechts: Bezugsquellen nach 1862

# Modernes komparatives Flussdiagramm: Russische Ölesporte vor und nach Sanktionen

Russia's seaborne oil exports before ..... and after new sanctions (million b/d)



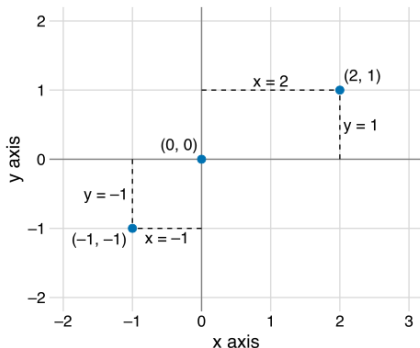
S&P Global 2022

# Kartesische Koordinaten



*Rene Descartes (1596-1650)*

Frans Hals (1648)

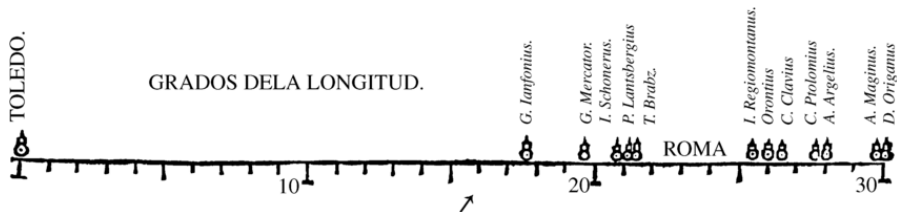


Kartesische Koordinaten

- Koordinatensystem  $\rightsquigarrow$  mathematische Beschreibung von (Daten-)Punkten
- Frans van Schooten (1615-1660): Idee von 2 orthogonalen Achsen
- Moderne Datenanalyse betrachtet einzelne Observations als Elemente in einem (a priori hoch-dimensionalen)  $\mathbb{R}^n$

# Van Langrens: 1-dimensionaler Plot von Datenpunkten

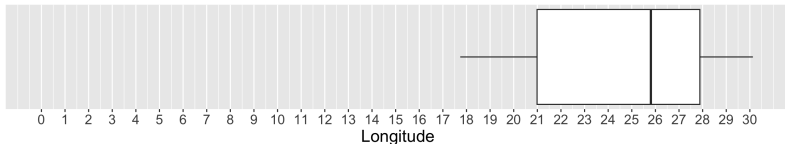
| Year | Name              | Longitude |
|------|-------------------|-----------|
| 150  | Ptolomeus, C.     | 27.7      |
| 1463 | Regiomontanus, I. | 25.4      |
| 1530 | Lantsbergius, P.  | 21.1      |
| 1536 | Schonerus, I.     | 20.8      |
| 1542 | Orontius          | 26.0      |
| 1567 | Mercator, G.      | 19.6      |
| 1567 | Clavius, C.       | 26.5      |
| 1578 | Brahe, T.         | 21.5      |
| 1582 | Maginus, A.       | 29.8      |
| 1601 | Organus, D.       | 30.1      |
| 1605 | Iansonius, G.     | 17.7      |
| 1610 | Argelius, A.      | 28.0      |



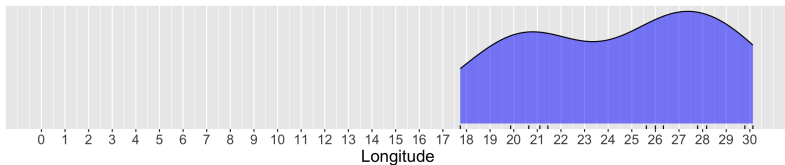
*Van Langrens: Longitude between Toledo and Rome (1644)*

# 1-Dimensionaler Plot von Datenpunkten: Statistik

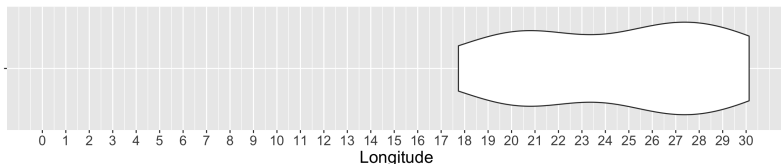
- John Tukey (1977): Boxplot der Quartile:



- Kernel-Density-Estimate  $\rightsquigarrow$  Dichteplot



- $\rightsquigarrow$  Moderne alternative zum Boxplot Violinplot



# Bemerkung: Bias (Systematischer Fehler)

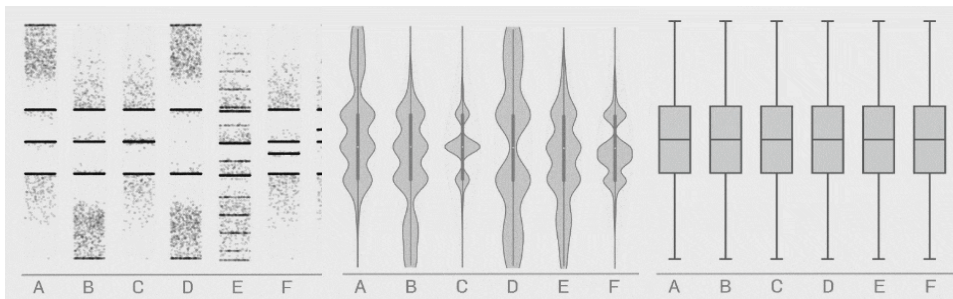
| Year | Name              | Longitude |
|------|-------------------|-----------|
| 150  | Ptolomeus, C.     | 27.7      |
| 1463 | Regiomontanus, I. | 25.4      |
| 1530 | Lantsbergius, P.  | 21.1      |
| 1536 | Schonerus, I.     | 20.8      |
| 1542 | Orontius          | 26.0      |
| 1567 | Mercator, G.      | 19.6      |
| 1567 | Clavius, C.       | 26.5      |
| 1578 | Brahe, T.         | 21.5      |
| 1582 | Maginus, A.       | 29.8      |
| 1601 | Organus, D.       | 30.1      |
| 1605 | Iansonius, G.     | 17.7      |
| 1610 | Argelius, A.      | 28.0      |



*Friendly and Wainer: History of Data Visualization (2021)*



# 1-Dim. Plot von Datenpunkten: Modernes Beispiel



*(Matejka and Fitzmaurice, Autodesk Research, 2017)*

**Plot Rohdaten  
(Scatterplot)**

**Plot Dichtefunktionen  
(Violinplot)**

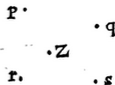
**Plot der Quartile und  
Extrema (Boxplot)**

- Die Daten A-G haben gleichen Median, 25%-Quantil, 75%-Quantil, Minimum und Maximum  $\rightsquigarrow$  Gleiche Boxplots
- Violinplot interpoliert zwischen Rohdaten (links, unübersichtlich) und Boxplots (rechts, ungenau)

# Roger Cotes: Mittlung von Datenpunkten in Ebene

## 22 ÆSTIMATIO ERRORUM

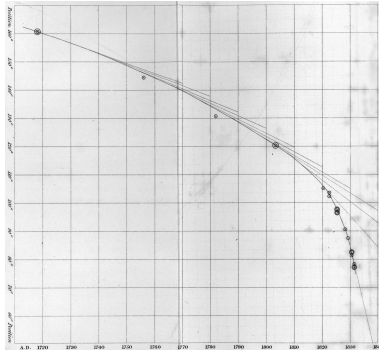
Ad eundem fere modum in aliis casibus Limites inveniuntur Errorum qui ex minus accuratis observationibus ortum ducunt, quin & Positiones ad Observandum commodissimæ deprehenduntur: ut mihi vix quidquam ulterius desiderari videatur postquam ostensum fuerit quæ ratione Probabilitas maxima in his rebus haberi possit, ubi diversæ Observationes, in eundem finem institutæ, paululum diversas ab invicem conclusiones exhibent. Id autem fiet ad modum sequentis Exempli. Sit  $p$  locus Objecti alicujus ex Observatione prima definitus,  $q, r, s$  ejusdem Objecti loca ex Observationibus subsequenter; sint insuper  $P, Q, R, S$  pondera reciproce proportionalia spatiis Evagationum, per quæ se diffundere possint Errores ex Observationibus singulis procedentes, quæque dantur ex datis Errorum Limitibus; & ad puncta  $p, q, r, s$  posita intelligantur pondera  $P, Q, R, S$ , & inveniatur eorum gravitatis centrum  $Z$ : dico punctum  $Z$  fore Locum Objecti maxime probabilem, qui pro vero ejus loco tutissime haberi potest.



*Roger Cotes: Aestimatio Errorum in Mixta Mathesis, per Variationes Planitum Trianguli Plani et Spherici (1722)*

- Roger Cotes: Versuch, um aus mehreren fehlerbehafteten Messungen  $p, q, r, s$  in der Ebene durch gewichtete Mittlung eine Schätzung  $Z$  des tatsächlichen Koordinatenpunktes zu erhalten.

# John Herschel: Virginis (Erster Scatterplot)



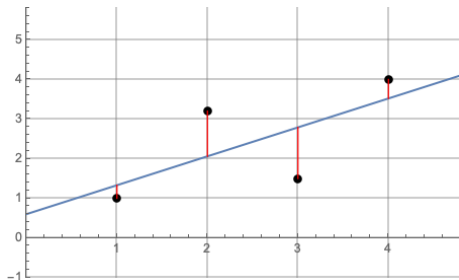
*John F. W. Herschel, "On the investigation of the orbits of revolving double stars: Being a supplement to a paper entitled 'micrometrical measures of 364 double stars,'" Memoirs of the Royal Astronomical Society, 5, (1833)*

- Der erste überlieferte Scatterplot stammt vom Astronomen John Herschel: Messungen der Konfiguration des Virginis-Doppelsterns.
- Außerdem legte Herschel eine (manuell generierte) glatte Interpolationskurve durch die Daten.

# Gauss: Methode der kleinsten Quadrate



Carl Friedrich Gauss (1777-1855)

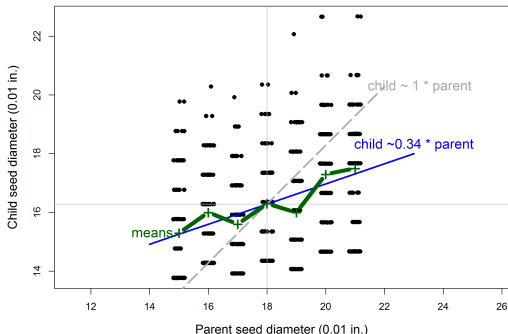


Methode der kleinsten Quadrate

- Die systematische Methode, um aus mehreren fehlerbehafteten Messungen (in beliebiger Dimension) ein optimales (wahrscheinlichstes) Model zu erhalten stammt von Gauss:
- Unter der Annahme, dass die Fehler normalverteilt sind, liefert die Methode der kleinsten Quadrate jene Parameter, für welche die gegebenen Datenbeobachtungen am wahrscheinlichsten sind (Gauss-Markow-Theorem).

# Francis Galton: Regressionsgerade

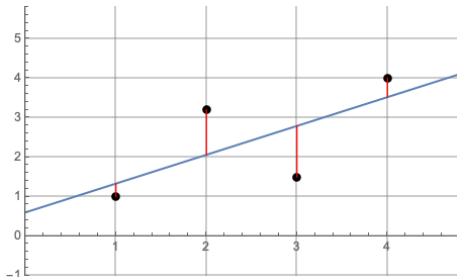
- Francis Galton verteilte an 7 Personen je 7 Packungen von 10 Dufterbsensaat von ungefähr gleicher Größe. Die Nachwuchsgeneration ließ er sich zuschicken



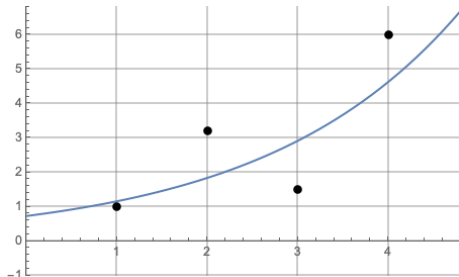
*Francis Galton: Natural Inheritance (1894) (modern recreation)*

- Blau: Regressionslinie durch Gruppenmittel
- Grau: Linie mit Steigung 1
- 'Regression to mediocrity'

# Modern: (Non-)Linear Least Squares - Fit



Lineares Model:  $y = a \cdot x + b$

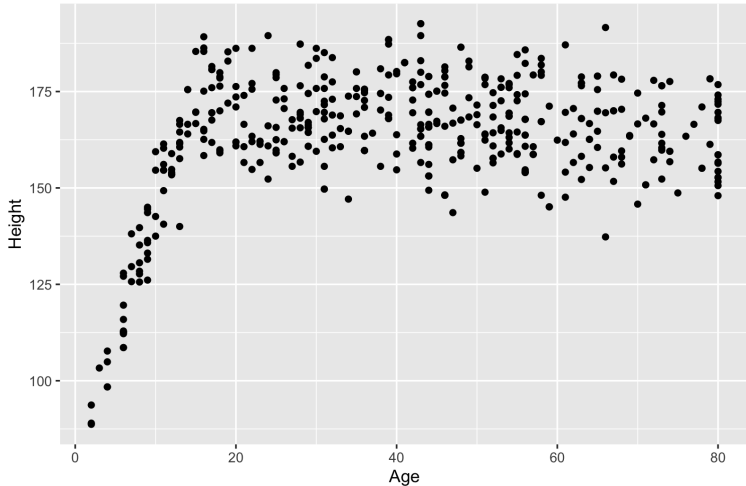


Exponentielles Model:  $y = a \cdot e^{b \cdot x}$

- Kommen alle Parameter linear vor: Linear Least Squares  $\rightsquigarrow$  Matrixmethoden zur Berechnung der Parameter (eindeutig)
- Kommen Parameter nicht-linear vor: Non-linear Least Squares  $\rightsquigarrow$  Iterativer Optimierungsprozess

## Modern: Non-Parametric Least Squares

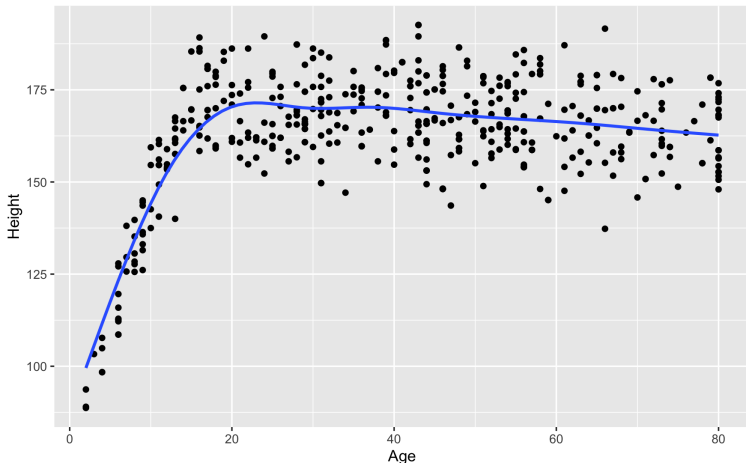
- Beispiel: Scatterplot von Körpergröße (Height) vs. Alter:



- Was ist eine gute approximative Kurve?

## Modern: Non-Parametric Least Squares

- Beispiel: Scatterplot von Körpergröße (Height) vs. Alter:

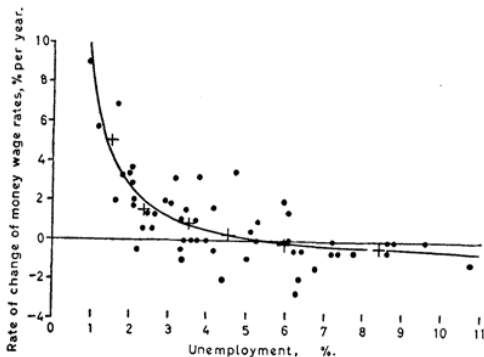


- ~ General Additive Model (GAM): Spline-Interpolationskurve
- ◇ Alternative: loess (locally estimated scatterplot smoothing)



## Beispiel nicht-linearer Zusammenhang: Phillip's Curve

- Scatterplot + Model ist eine der wichtigsten Standardroutinen
  - ↪ 2 Variablen, bei welchen man einen Zusammenhang vermutet werden "gegeneinander" geplottet
  - ↪ detailliertere Beobachtungen als mit einzelnen Statistiken (z.B. Korrelation)



A. W. Phillips: "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957," (1958)

# Gefahren und Nachteile von Visualisierung

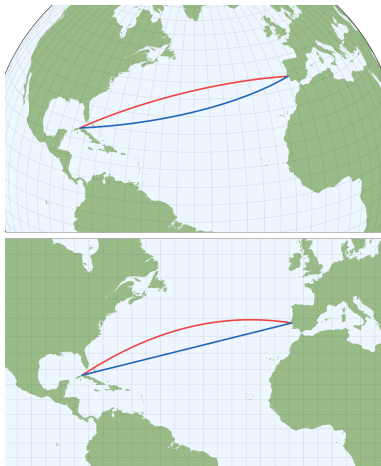
- Neigung des Gehirns zur Gestaltwahrnehmung kann Muster erkennen wo keine sind
- Scheinkorrelationen  $\rightsquigarrow$  visuell überzeugende Grafik ohne reale Fundierung
- Überparametrisierung / Overfitting  $\rightsquigarrow$  Model für bestehende Daten scheinbar sehr gut, jedoch nicht auf echte (neue) Daten übertragbar
- Fehlende Daten  $\rightsquigarrow$  Potentielle Fehlinterpretationen / Bias

# Grundproblem für Weltkarte: Krümmung der Erde



- Sphäre ist nicht isometrisch zur euklidischen Ebene
- keine 'perfekte' Karte möglich

# Mercator-Weltkarte: Geraden sind nicht die kürzeste Wege



*Oben: Sphäre, Unten: Mercator-Karte*

- Rot: Kürzester Weg zwischen Lissabon und Havanna
- Gerade auf Mercator-Karte zwischen Lissabon und Havanna

## Alternative: Lambert-Weltkarte

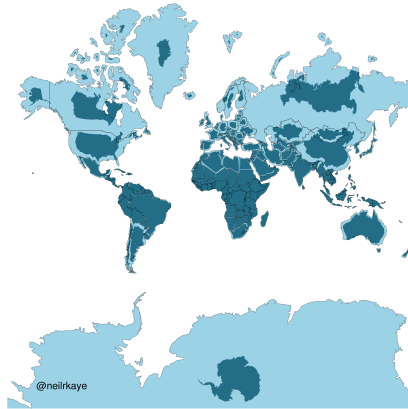


*Johann Heinrich Lambert (1772)*

- Geraden auf der Karte  $\approx$  kürzeste Routen
- Einsatz in Luftfahrt

# Mercator-Weltkarte: Illustration Flächenverzerrung

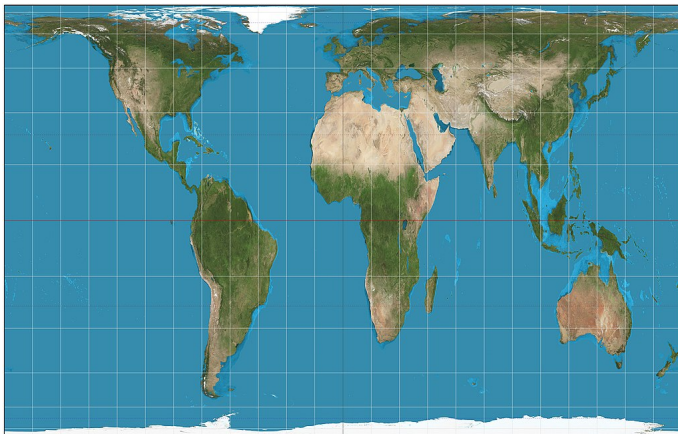
- Weiteres Problem der Mercator-Karte: Flächen werden verzerrt



*Mercator Weltkarte (moderne Darstellung): Flächenverzerrung*

- ◇ Am genauesten sind die Flächen in der Nähe des Äquators
- ◇ Zu den Polen hin werden Flächen immer stärker vergrößert

## Alternative: Gall-Peters-Karte



*Gall-Peters-Karte (1974))*

- Vorteil: Flächen werden korrekt dargestellt
- Nachteil: Starke Verzerrung der Formen

# Gestalt und Muster



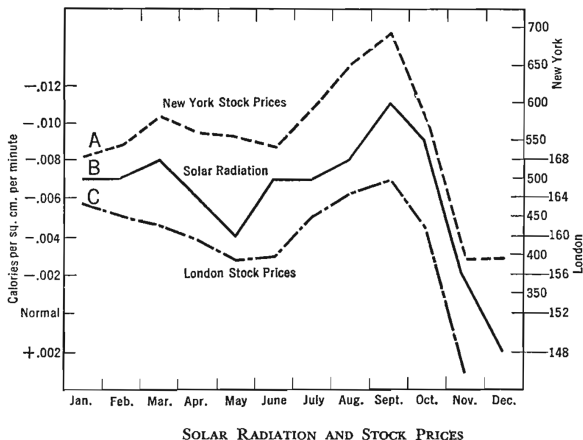
*Rosetta Nebula*



*Galle Krater*

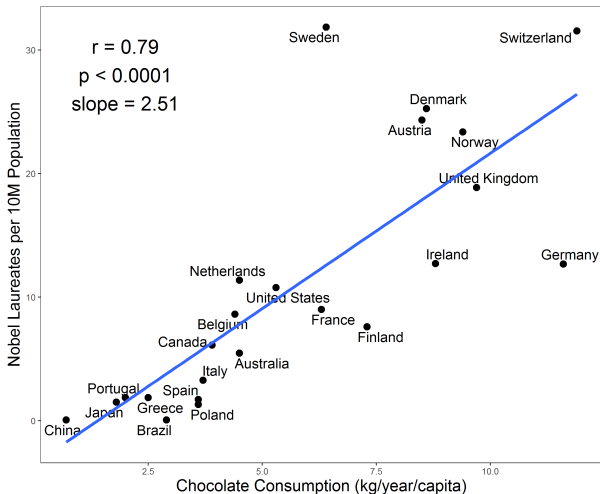


# Beispiel: Scheinkorrelation

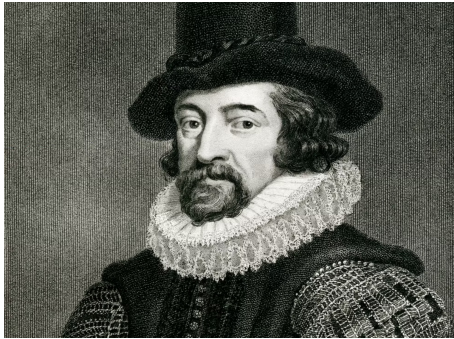


*Garcia Mata and Schaffner: Solar and Economic Relationships: A Preliminary Report (1934)*

# Beispiel: Spurious Correlations



*Franz H. Messerli: Chocolate Consumption, Cognitive Function, and Nobel Laureates, The New England Journal of Medicine, 367:16 (2012)*



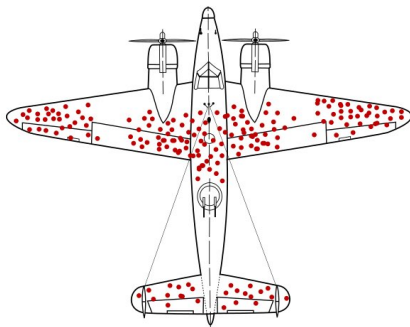
*Francis Bacon (1561-1626)*

Der Philosoph Francis Bacon berichtet folgende Anekdote aus der Antike:

*Als dem ungläubigen Diagoras im Tempel des Neptun die vielen Dankesinschriften derjenigen gezeigt werden, die einem Schiffbruch entkamen und gefragt wurde, ob er dank solcher Rettungszeichen nicht doch an die Wirksamkeit von Gebeten glaube erwiderte dieser:  
'Und wo sind die Dankestafeln der Ertrunkenen?'*

## Beispiel: Gefahr der fehlenden Daten

- Während dem 2. Weltkrieg wurde die Statistical Research Group beauftragt, anhand von Datenanalysen festzustellen, wie Flugzeuge am besten durch Panzerung geschützt werden sollen.



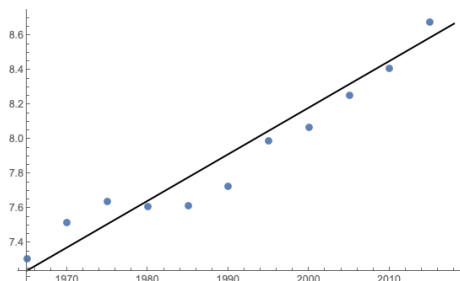
*Ellenberg: How Not to Be Wrong: The Power of Mathematical Thinking (2014)*

→ Wo sollte am besten in mehr Panzerung investiert werden?

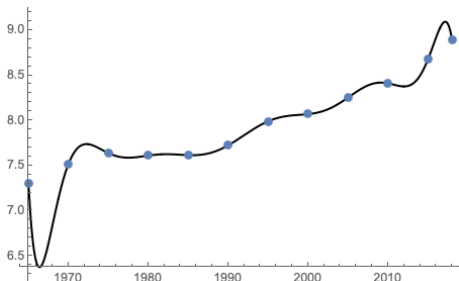
# Overfitting: Beispiel

- Wir betrachten die demographische Entwicklung Österreichs (in Mio.):

| 1965 | 1970 | 1975 | 1980 | 1985 | 1990 | 1995 | 2000 | 2005 | 2010 | 2015 | 2018 |
|------|------|------|------|------|------|------|------|------|------|------|------|
| 7,31 | 7,52 | 7,64 | 7,61 | 7,62 | 7,72 | 7,99 | 8,07 | 8,25 | 8,41 | 8,68 | 8,90 |



Plot Datenpunkte +  
Ausgleichsgerade



Plot Datenpunkte + Model mit 12  
Parametern

## Overfitting



*Johann von Neumann*

*"With four parameters I can fit an elephant,  
and with five I can make him wiggle his  
trunk."*

# Neumann's parametrisierte Elefanten

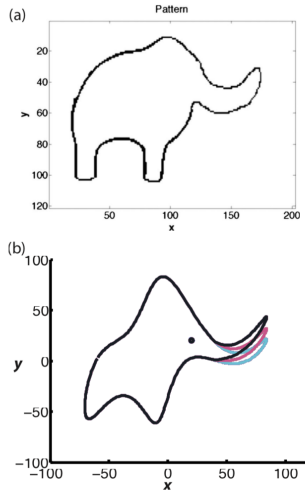


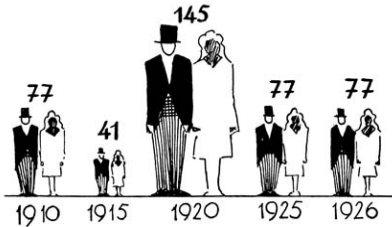
Fig. 1. (a) Outline of an elephant. (b) Three snapshots of the wiggling trunk.

*Mayer-Khairi-Howard: Drawing an elephant with four complex parameters,*  
American Journal of Physics (2010)

## Bemerkung: Otto Neurath: Isotypen

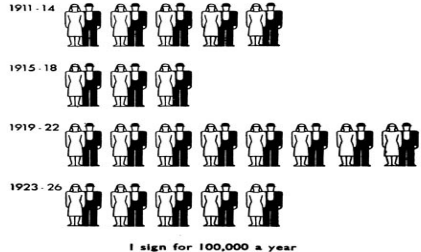
Was ist hier falsch:?

Example of bad system: signs of different sizes  
Number of men getting married in Germany out of every 10,000 persons



Mit 'Isotypen':

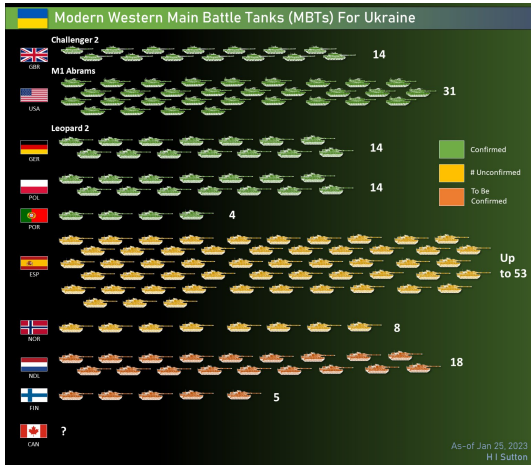
Men Getting Married in Germany in a Year



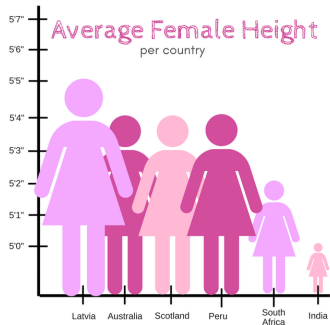
*Otto Neurath: International Picture Language, London (1936)*



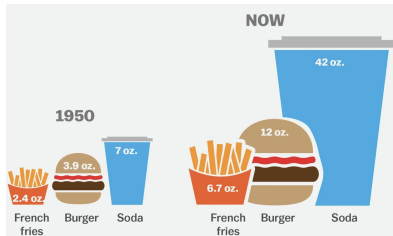
- Suboptimale Verwendung von 'Isotypen':



- Falsche Darstellung (Flächen proportional zu Werten<sup>2</sup>):



- Korrekte Darstellung (Flächen proportional zu Werten):

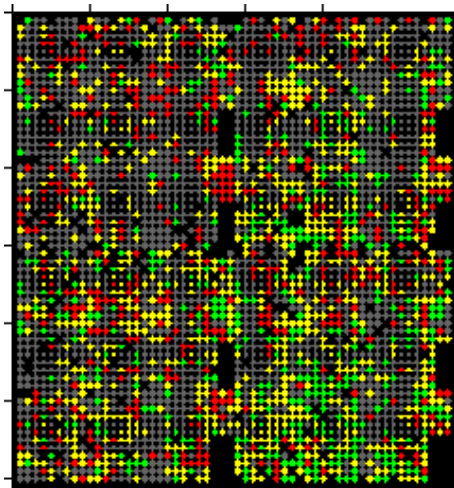


# Grenzen der Visualisierung

- Nur numerische Daten direkt visualisierbar: Sonst Umwandlung von kategorischen oder ordinalen Daten in numerische Skalen notwendig
- Nur 1-2-dimensionale Daten direkt visualisierbar: Sonst Projektion der Daten notwendig
- Visualisierung im allgemeinen auf flachen Grund (Papier / Bildschirm)  
    ~> Probleme mit gekrümmten Daten: Kompromiss zwischen Größe des gezeigten Bereiches und der Genauigkeit

# Wie kann man hochdimensionale Daten visualisieren?

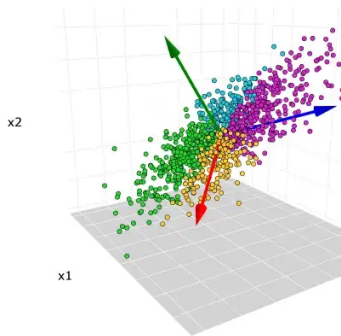
- Beispiel: Gene-Expression-Heatmap



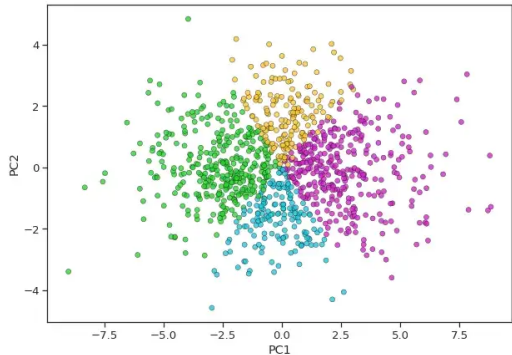
⇒ Reduktion der Dimension notwendig

⇒ Correlation Matrix oder Principal-Component-Analysis (PCA)

# Illustration Principal-Component-Analysis(PCA)



Originaler Datenraum mit  
PC-Richtungen

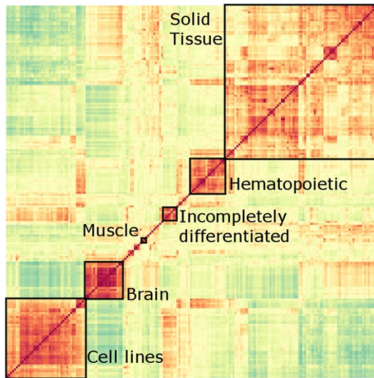


Projektion der Daten auf PC1 und PC2

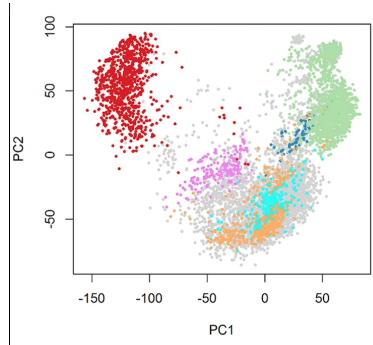
- Es werden konsekutiv Richtungen bestimmt, in welchen die Daten maximale Varianz aufweisen: zunächst **PC1**, anschließend (orthogonal zu PC1) **PC2**, anschließend (orthogonal zu PC1 und PC2) **PC3**:

## Beispiel: Gene-Expression

- Beispiel mit 369 Genen:



Correlation Matrix



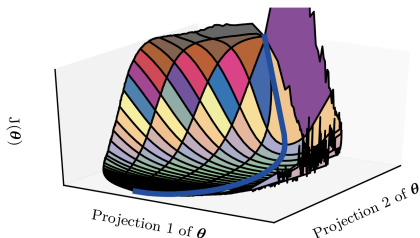
PCA

*Lenz et al.: Principal components analysis and the reported low intrinsic dimensionality of gene expression microarray data, Scientific Reports (2016)*

## Beispiel: Illustration Konvergenz neuronales Netz

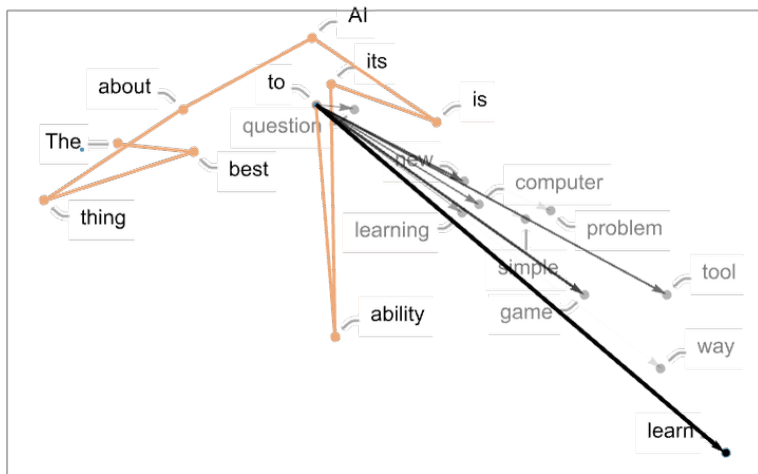
- Moderne neuronale Netze können weisen üblicherweise Millionen von Parametern (Gewichte und Biase der Nodes) auf, welche mittels Gradientenabstiegsverfahren optimiert werden (Minimierung der Kostenfunktion)

→ Konkrete Visualisierung der Optimierung / Konvergenz schwierig



*Ian Goodfellow, Yoshua Bengio and Aaron Courville: Deep Learning (2015)*

# Beispiel: Illustration probabilistische Vorhersage eines neuronalen Netzes



*Stephen Wolfram: What Is ChatGPT Doing ... and Why Does It Work? (2023)*



## Beispiel: Umfrage (kategorische Daten)

- Für gegebene kategorielle Daten ist umwandlung in numerische Größen oder Zuweisung von diskreten Symbole notwendig.
- Beispiel: 'Harvey Balls': Ordinale Daten (0-4) werden visuell dargestellt:



- Beispiel:

|               | Price | Accessibility | Quality | Branding |
|---------------|-------|---------------|---------|----------|
| Product One   |       |               |         |          |
| Product Two   |       |               |         |          |
| Product Three |       |               |         |          |
| Product Four  |       |               |         |          |

## Beispiel: Umfrage (lexikographische Daten)

- Darstellung von Wortdaten: tag cloud ('word cloud'):
- Beispiel:

2016 Remain voters



### 2016 Leave voters

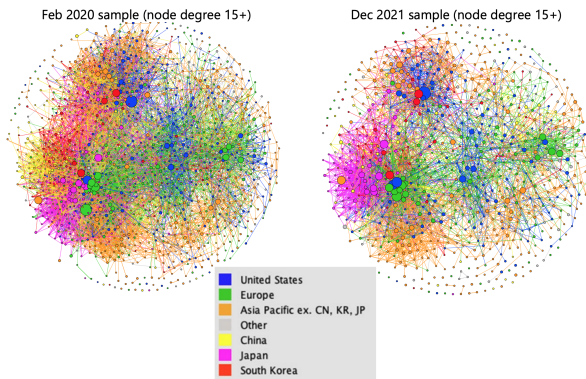


*Tony Blair Institut (2022)*

- ◆ Bei einer Umfrage zu einem Thema werden 2 Personengruppen nach einem 'Stichwort' gefragt, welches Sie mit diesem Thema assoziieren.
- ◆ Es werden aus allen gegebenen Antworten Frequenztabellen erstellt und die häufigsten Antworten gemäß proportionaler Frequenz angezeigt.

## Beispiel: Handelsnetzwerk

- Beispiel Visualisierung Netzwerk:
- ◇ Internationale Wertschöpfungsketten vor und nach COVID-Pandemie:



*Bank of International Settlements (BIS) (2023)*

- ◇ Knoten: Firmen
- ◇ Kanten: Handelstransaktionen

# Modern: Programmiersprachen für Statistik und Visualisierung

- **Beispiele:**

- R (primär für Statistik und Datenanalyse)
- Python (Programmierung, Statistik, Datenanalyse, Machine Learning)
- ...

- **Vorteile von "programmierten Grafiken":**

- ◇ Geschwindigkeit (Einsatz moderner Systeme, effiziente Algorithmen)
- ◇ Reproduzierbarkeit / Pipelines (Skript / Code)
- ◇ Deployment in Anwendungen / Dynamische Reports
- ◇ Skalierbarkeit (Übertragung auf leistungsfähigeres System)
- ◇ Libraries/Packages für konkrete Probleme
- ◇ ...

# Charts für spezielle Einsatzzwecke

- Abweichungen (Deviations)
- Korrelationen (Correlations)
- Verteilungen (Distributions)
- Veränderung (Change over Time)
- Größe (Magnitude)
- Anteile (Part-to-whole)
- Räumlich (Spatial)
- Fluss (Flow)

## Deviation

Emphasize variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

**Example FT uses**  
Trade surplus/deficit, climate change

### Diverging bar



A simple standard bar chart that can handle both negative and positive magnitude values.

### Diverging stacked bar



Perfect for presenting survey results which involve sentiment (eg disagree/neutral/agree).

### Spine



Splits a single value into two contrasting components (eg male/female).

### Surplus/deficit filled line



The shaded area of these charts allows a balance to be shown – either against a baseline or between two series.

## Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

**Example FT uses**  
Inflation and unemployment, income and life expectancy

### Scatterplot



The standard way to show the relationship between two continuous variables, each of which has its own axis.

### Column + line timeline



A good way of showing the relationship between an amount (columns) and a rate (line).

### Connected scatterplot



Usually used to show how the relationship between 2 variables has changed over time.

### Bubble



Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

### XY heatmap

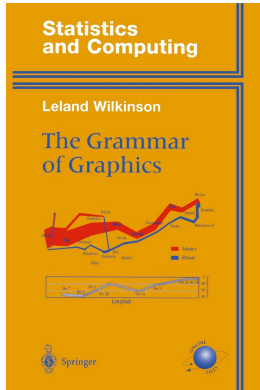


A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

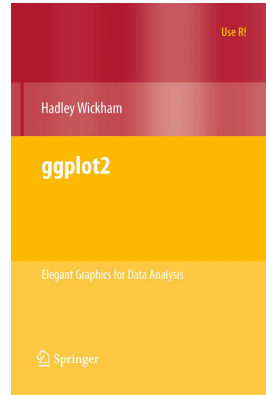
*Financial Times: Visual Vocabulary (2023):  
(Ausschnitt)*

# Moderner Zugang: Grammatik der Visualisierung

- **Idee:** Visualisierungen sind nicht ein 'Katalog' aus vorgefertigten Darstellungstypen sondern entstehen gemäß einer logischen Struktur.



Wilkinson



Wickham

# Grammatik: Grundkomponenten (Layer)



- **Elementare Layer:**

- **Daten** (*Data*): Was soll gezeigt werden?
- **Variablenzuweisungen** (*Aesthetics*): Welche Variablen der Daten entsprechen  $x$  und  $y$ -Koordinaten im Plot? Welche Variablen werden zu Farbskalen? Welche Variablen werden zu Formen?
- **Darstellungsform** (*Geometries*): Welche Darstellungsobjekte? (e.g. Balken, Punkte, Linien)

# Grammatik: Grundkomponenten (Layer)

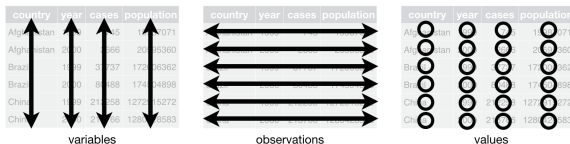


- Weitere Layer:
  - **Aufteilung** (*Facets*): Wie soll die Darstellung auf einzelne Teildarstellungen aufgeteilt werden?
  - **Statistik** (*Statistics*): Welche statistischen Operationen sollen auf die Variablen angewandt werden?
  - **Koordinatensysteme** (*Coordinates*): Welches Koordinatensystem soll der Darstellung zugrunde liegen?
  - **Themen** (*Themes*): Hintergrundfarbe, Schrift, Design



# Vorbereitung der Daten: Tidy Data

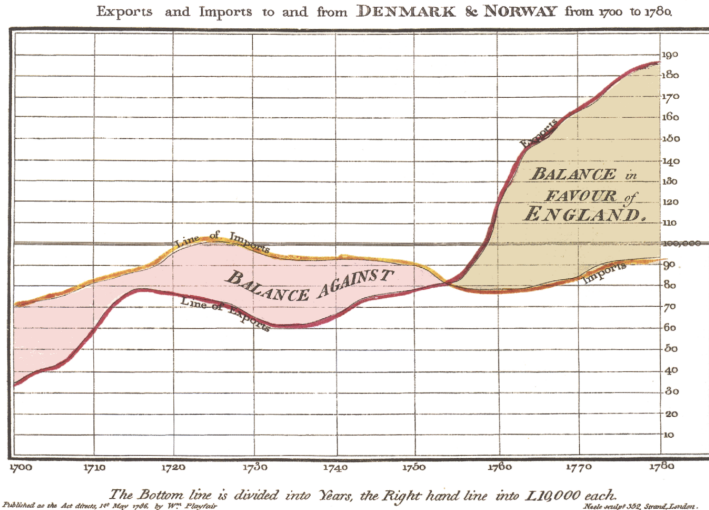
*“Tidy datasets are all alike, but every messy dataset is messy in its own way.” — Hadley Wickham*



*Hadley Wickham and Garrett Grolemund: R for Data Science*

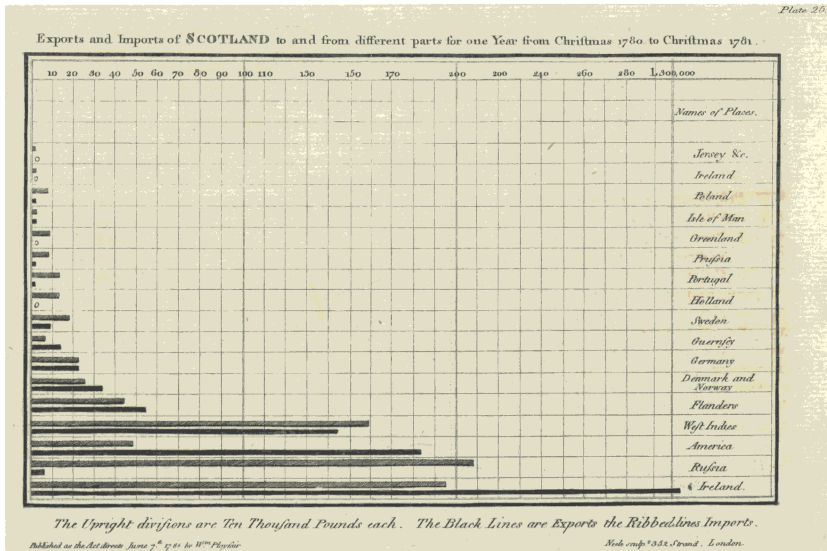
- Ein Dataset ist tidy (deutsch: aufgeräumt, geordnet), wenn...
  - Jede Variable einer Spalte entspricht
  - Jede Beobachtung einer Zeile entspricht
  - Jede Zelle genau einem Wert entspricht

## Beispiel: Playfair's Lineplot



William Playfair: *The Commercial and Political Atlas*, London (1786)

# Beispiel: Playfair's Barchart



William Playfair: The Commercial and Political Atlas, London (1786)

# Beispiel: Anscombe's Dataset

| I    |       | II   |      | III  |       | IV   |       |
|------|-------|------|------|------|-------|------|-------|
| x    | y     | x    | y    | x    | y     | x    | y     |
| 10.0 | 8.04  | 10.0 | 9.14 | 10.0 | 7.46  | 8.0  | 6.58  |
| 8.0  | 6.95  | 8.0  | 8.14 | 8.0  | 6.77  | 8.0  | 5.76  |
| 13.0 | 7.58  | 13.0 | 8.74 | 13.0 | 12.74 | 8.0  | 7.71  |
| 9.0  | 8.81  | 9.0  | 8.77 | 9.0  | 7.11  | 8.0  | 8.84  |
| 11.0 | 8.33  | 11.0 | 9.26 | 11.0 | 7.81  | 8.0  | 8.47  |
| 14.0 | 9.96  | 14.0 | 8.10 | 14.0 | 8.84  | 8.0  | 7.04  |
| 6.0  | 7.24  | 6.0  | 6.13 | 6.0  | 6.08  | 8.0  | 5.25  |
| 4.0  | 4.26  | 4.0  | 3.10 | 4.0  | 5.39  | 19.0 | 12.50 |
| 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15  | 8.0  | 5.56  |
| 7.0  | 4.82  | 7.0  | 7.26 | 7.0  | 6.42  | 8.0  | 7.91  |
| 5.0  | 5.68  | 5.0  | 4.74 | 5.0  | 5.73  | 8.0  | 6.89  |

*F. J. Anscombe: Graphs in Statistical Analysis (1973)*



| x  | y    | set |
|----|------|-----|
| 10 | 8.04 | I   |
| 10 | 9.14 | II  |
| 10 | 7.46 | III |
| 8  | 6.58 | IV  |
| :  |      | :   |

# Anscombe's Dataset

- Variablen: x, y (numerisch)
- Darstellungsform: Punkte
- Aufteilung: Gitter

