

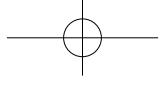
1.1 可视化的意义

在计算机科学领域,可视化将不可见或者难以直接显示的数据转换为可以感知的图形、符号、颜色、纹理等,增强数据识别效率,传递有效信息^[1]。可视化对应两个英文单词:visualize 和 visualization。前者 visualize 是动词,表示生成符合人类感知的图像,通过可视元素传递信息。后者 visualization 是名词,表示使某物或某事可见的动作或事实,使原本不可见的事物在人脑中形成可感知心理图像的过程或能力。数据可视化是一种关于数据属性和变量的视觉表现形式的科学技术。

可视化的意义在于帮助人们更好地分析数据。信息的质量很大程度上依赖于其表达方式,数据可视化对罗列数字组成的数据中所蕴含的意义进行分析,使分析结果可视化。其实数据可视化的本质就是视觉对话。数据可视化将技术与艺术完美结合,借助图形化的手段,清晰有效地传达信息。一方面,数据赋予可视化以价值;另一方面,可视化增加了数据的灵性,两者相辅相成,帮助人们从信息中提取知识、从知识中获取价值。精心设计的图形不仅可以提供信息,还可以通过强大的呈现方式增强信息的影响力,吸引人们的注意力并使其保持兴趣,这是纸质表格和电子表格无法做到的。可视化为读者更好地展现事物的全貌,很多讨论所涉及的主题都包括多个元素,其中一个元素会影响多个其他元素,如果不采取可视化,则无法看到全貌,也无法进行真正的讨论。可视化帮助读者增强理解,便于读者与信息对话和交流,简化信息的复杂性。可视化的作用体现在多个方面,如揭示想法或关系,形成论点或意见,观察事物演化的趋势,总结或积聚数据,资料存档和汇整,寻求真相和真理,传播知识和探索性分析数据等。

1.2 可视化简史

可视化的历史可谓悠久,通常,我们将可视化的历史分为“前计算机时代”和“后计算机时代”。如图 1.1 所示,在前计算机时代,绘制可视化图表和信息图主要采用手绘的



形式；而到了后计算机时代，更多的是利用计算机进行图表绘制。本节将基于 2004 年之前的可视化史进行简要举例，2004 年以后的可视分析将在本书 1.3.4 小节中阐述。

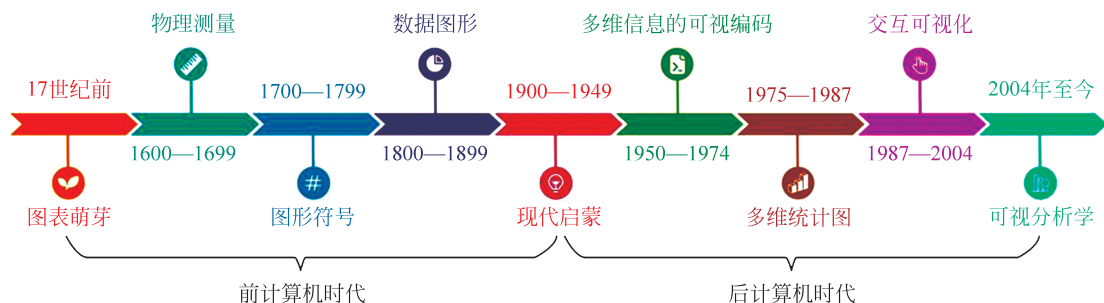


图 1.1 可视化简史

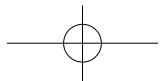
1.2.1 图表萌芽

17 世纪以前，图表表达还处于萌芽时期，主要形式是地图。大约公元前 366 年出现了人类历史上第一幅城市交通图，如图 1.2 所示。这张图显示了整个罗马世界，交通信息被绘在羊皮纸上，展示了从维也纳到意大利再到迦太基的地理信息。



图 1.2 人类历史上第一幅城市交通图（图片来源于网络）

1569 年 8 月，墨卡托（Mercator）出版的世界地图是第一张真正意义上的世界地图，开创了地理学史上的新篇章。墨卡托发明了圆柱投影，用于在地图上描绘地球，以保持横线的直线性，地图上的直线在对照指南针时可以转换为恒定方位线，非常适合海上航行使用。这张图很受欢迎，并且方便打印，到目前为止仍是人类最常见的世界地图投影。墨卡



托投影的地图缺点在于和现实差别太大,变形严重。在墨卡托投影的地图上,变形最严重的就是非洲和格陵兰岛地区。

到了16世纪,用于精确观测和测量物理量的技术和仪器得到了很好的发展。人类产生了直接捕获图像并将数学函数记录在表格中的初步想法,这些便是可视化图表萌芽的开始。

1.2.2 物理测量

物理测量理论在17世纪有了巨大的发展。解析几何的兴起、测量误差的理论和概率论的诞生、人口统计学的形成与完善,以及政治版图的发展为数据可视化奠定了基础。17世纪末,数据可视化方法所需的基本要素已经具备,一些具有重大意义的真实数据、理论以及视觉表现方法的出现,使人类开启了可视化思考新模式,因此可以将17世纪视为可视化史的开端。

1626年,克里斯托弗·施纳(Christopher Scheiner)画出了表达太阳黑子随时间变化的图,如图1.3所示,这张图在一个视图上同时展示多个小图序列,是邮票图表法的雏形。

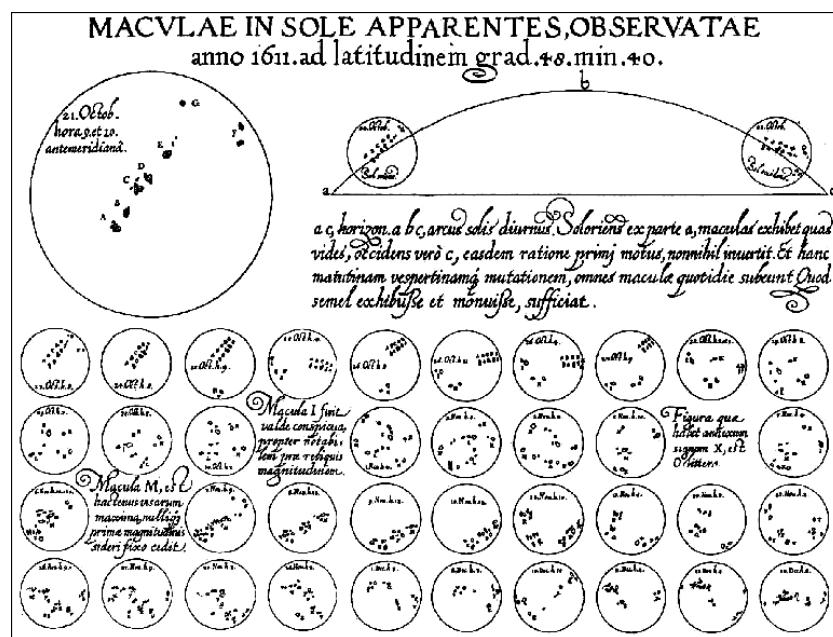
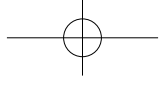


图 1.3 太阳黑子随时间变化的图(图片来源于网络)

1686年,埃德蒙·哈雷(Edmond Halley)绘制了迄今已知的第一个气象图显示了主流的风场分布,这也是向量场可视化的鼻祖。

1.2.3 图形符号

到了18世纪,社会和科技进步使数据价值开始被人们重视,人们不再满足于只在地



图上展示几何图形，抽象图形和函数图形的功能被大大扩展，因此许多崭新的数据可视化形式在这个世纪里诞生。

18 世纪是统计图形学的繁荣时期，其奠基人威廉·普莱费尔（William Playfair）发明的折线图、柱状图、饼状图，构成了当今数据可视化的核心要素。图 1.4 所示的柱状图展示了苏格兰与欧洲和新世界各个地区的贸易。通过柱状图的方式显示数据，可以一目了然地看到苏格兰与爱尔兰的紧密经济联系，以及与俄罗斯的贸易不平衡。图 1.5 所示的折线图展示了丹麦与挪威 1700—1780 年的贸易出口序列。图 1.6 所示的饼图用于展示局部与整体的关系，该图显示了土耳其各地区疆土所占比例。

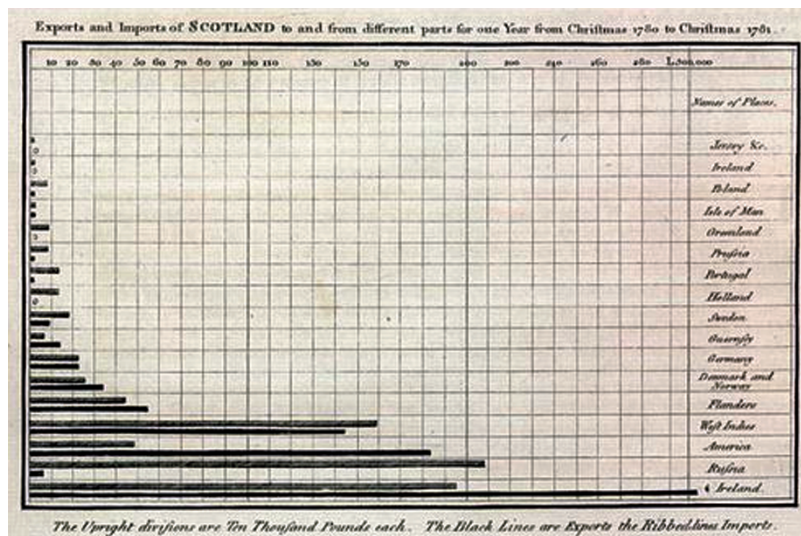


图 1.4 苏格兰与欧洲和新世界各个地区的贸易图（图片来源于网络）

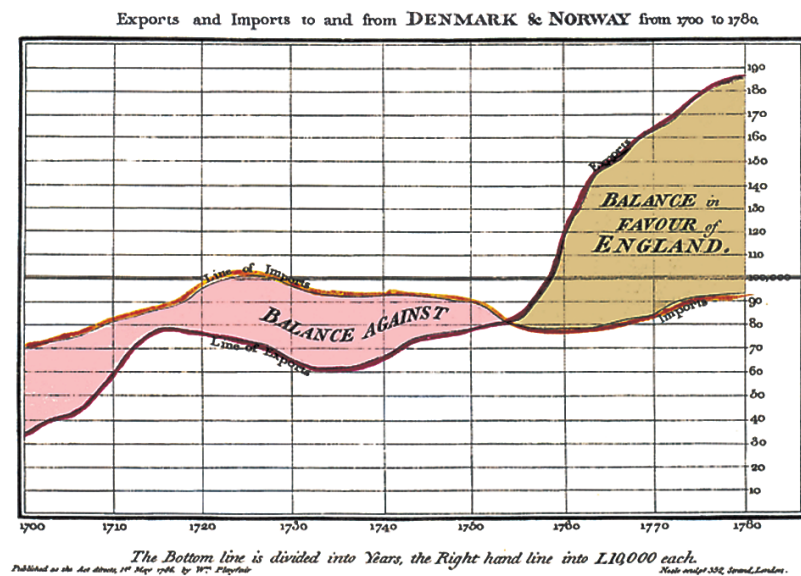
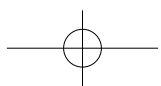


图 1.5 丹麦与挪威于 1700—1780 年的贸易出口序列图（图片来源于网络）



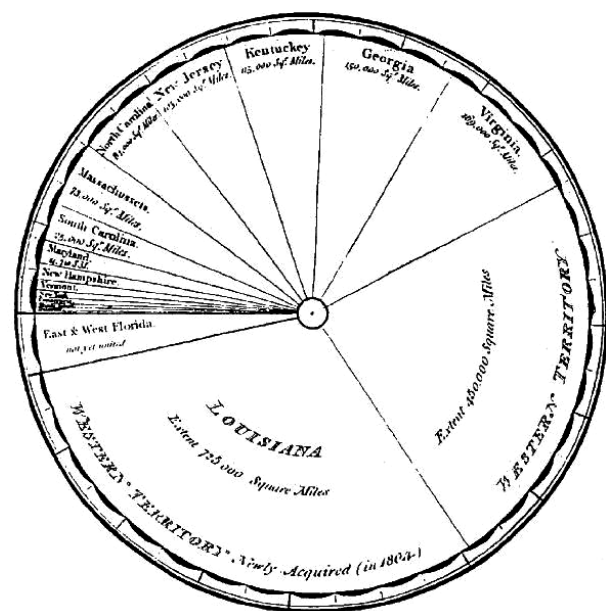


图 1.6 使用饼图显示土耳其各地区疆土所占比例（图片来源于网络）

随着工艺设计的完善，统计图形和主题制图的种类日益丰富，包括柱状图、饼图、直方图、折线图、时间线、轮廓线等。在专题制图学中，制图从单一地图发展为全面的地图集，描绘了涉及各种主题（经济、社会、道德、医学、身体等）的数据，同时演化出了可视化思考的新方式。

1.2.4 数据图形

19 世纪中期，可视化快速发展的所有条件已经具备。人们认识到数字信息对社会计划、工业化、商业和运输的重要性在日益提高，欧洲各地开始建立官方的国家统计局。

1854 年约翰·斯诺（John Snow）在《伦敦暴发的霍乱病例群》图中使用点图映射了当年的宽街霍乱疫情，如图 1.7 所示。他还使用了统计数据来说明水源质量与霍乱病例之间的联系，结果表明该疾病是通过受污染的水传播的，而不是以前认为的通过空气传播。斯诺的研究是公共卫生和地理历史上的重大事件，它被认为是流行病学的创始事件。

弗洛伦斯·南丁格尔（Florence Nightingale）不仅是受人尊敬的现代护理学创始人，也是一位才华横溢的数学家，是统计学图形表示的先驱。1857 年，弗洛伦斯·南丁格尔主动申请，自愿担任战地护士。她率领 38 名护士抵达前线，在战地医院服务。她竭尽全力排除各种困难，仅用半年左右的时间就将伤病员的死亡率下降到 2.2%，战争结束后，南丁格尔回到英国，被人们推崇为民族英雄。她以 Playfair 的思想为基础，绘制了极坐标面积图 Coxcomb，并将图表插入自己的许多出版物中。图 1.8 所示为东方军队士兵死亡原因图，该图表按月描绘了克里米亚战争期间士兵死伤的相关原因，每个扇形的面积代表了

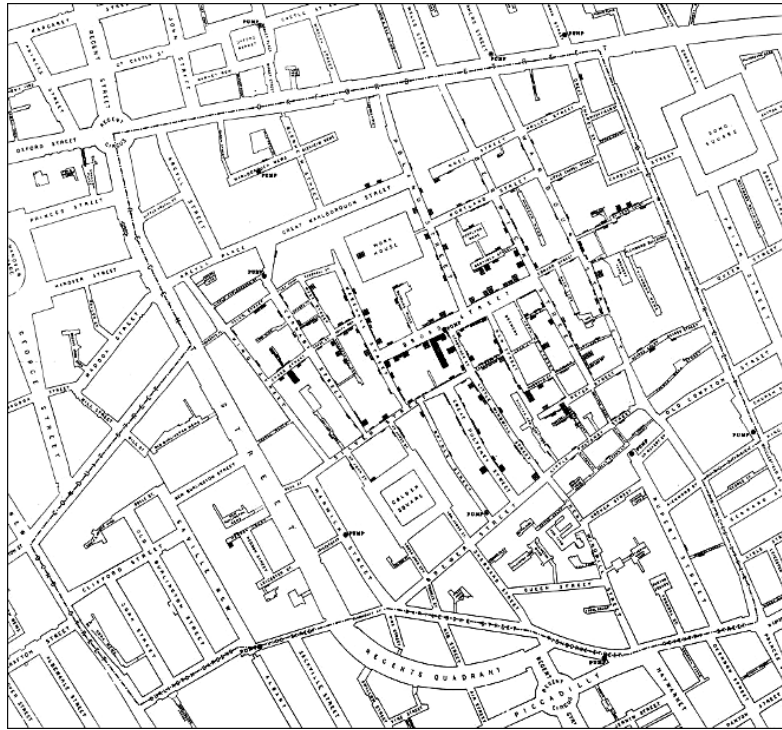
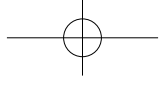


图 1.7 伦敦爆发的霍乱病例群（图片来源于网络）

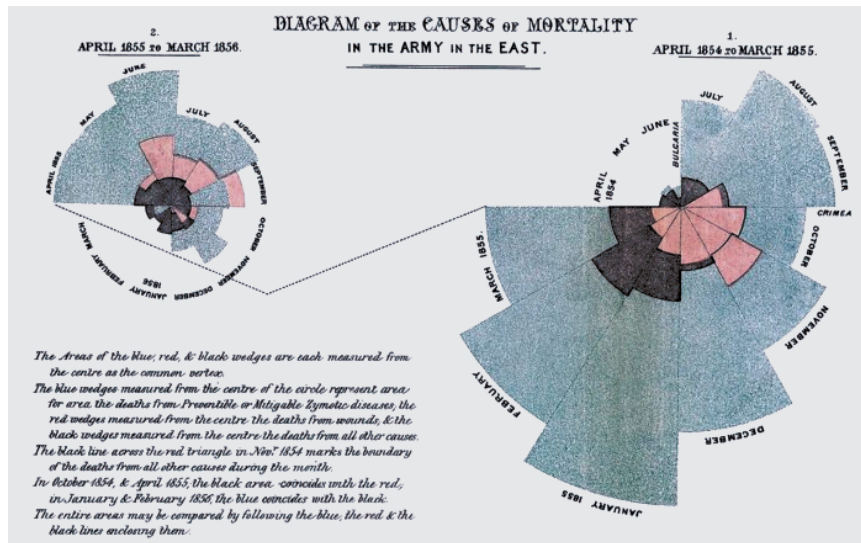
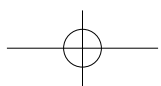


图 1.8 东方军队士兵死亡原因图（图片来源于网络）

统计数据的大小。

1869 年查尔斯·约瑟夫·米纳德（Charles Joseph Minard）发布的拿破仑 1812 年东征俄罗斯事件的流图，被誉为有史以来最好的数据可视化，如图 1.9 所示。他的流图呈现了拿破仑军队的位置、行军方向、军队分散和重聚的时间地点以及减员等信息。



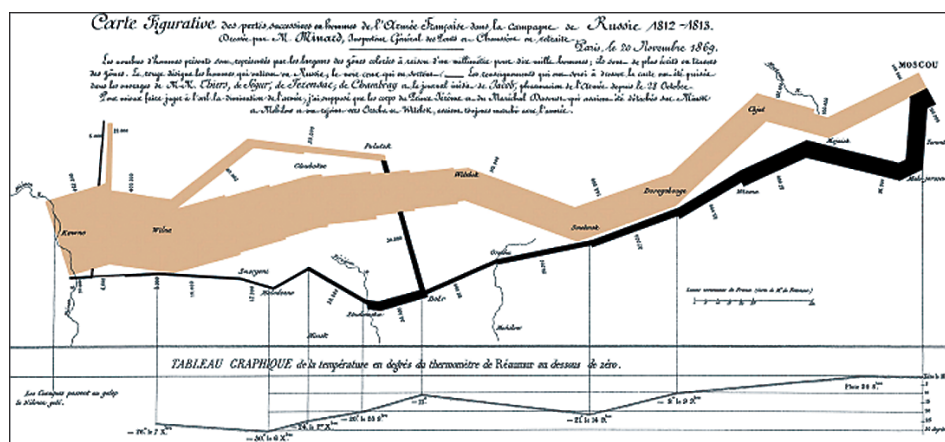


图 1.9 拿破仑 1812 年东征俄罗斯事件的流图（图片来源于网络）

1879 年路易吉·佩罗佐（Luigi Perozzo）绘制了三维人口金字塔立体图，如图 1.10 所示，该图以实际数据为依据（瑞典人口普查，1750—1875 年）。此图与之前出现的可视化形式有一个明显的区别，即开始使用三维形式，并使用彩色表示数据值之间的区别，提高了视觉感知。

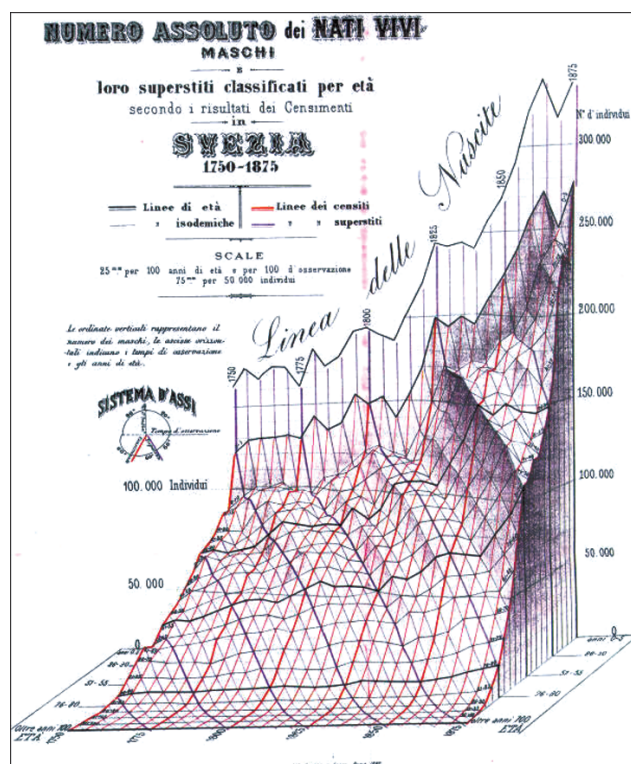


图 1.10 三维人口金字塔立体图（图片来源于网络）

1885 年法国工程师伊布里（Ibry）绘制的火车时刻表（见图 1.11），显示了从巴黎到里昂这一路线上火车的行驶速度，并且此绘制方法一直沿用至今。

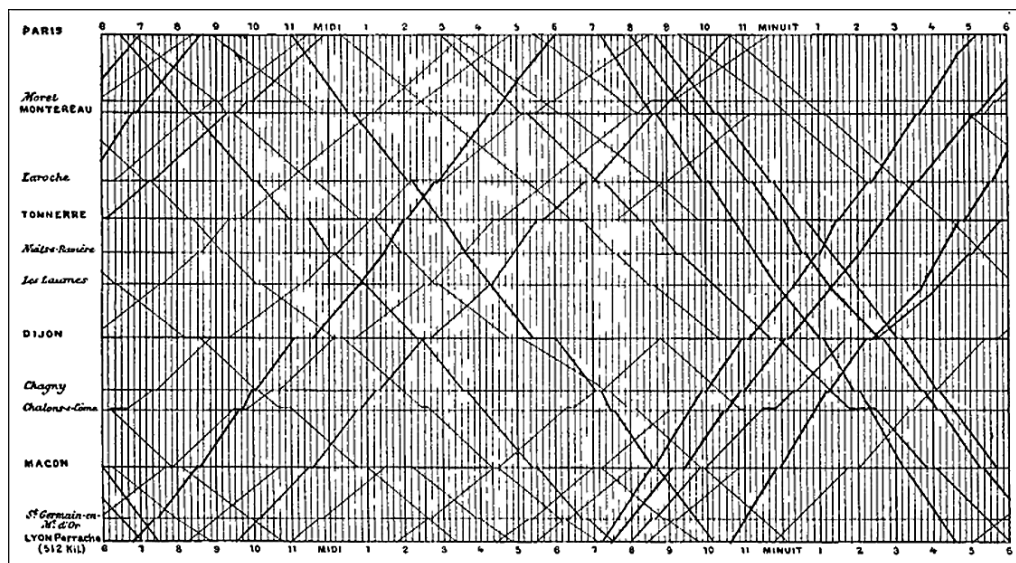


图 1.11 显示行驶速度的火车时刻表（图片来源于网络）

1.2.5 现代启蒙

如果说 19 世纪初是统计图形和专题制图的“黄金时代”，那么 20 世纪初则可称为可视化的“现代黑暗时代”。这一阶段少有图形创新，直到 20 世纪 30 年代中期，社会科学中用于量化的统计模型的兴起，这一局面才有所扭转。

1904 年曼努德（Manuder）绘制了蝴蝶图，如图 1.12 所示，该图研究了黑子随时间的变化，验证了太阳黑子的周期性并对未来黑子的变化进行了预测。

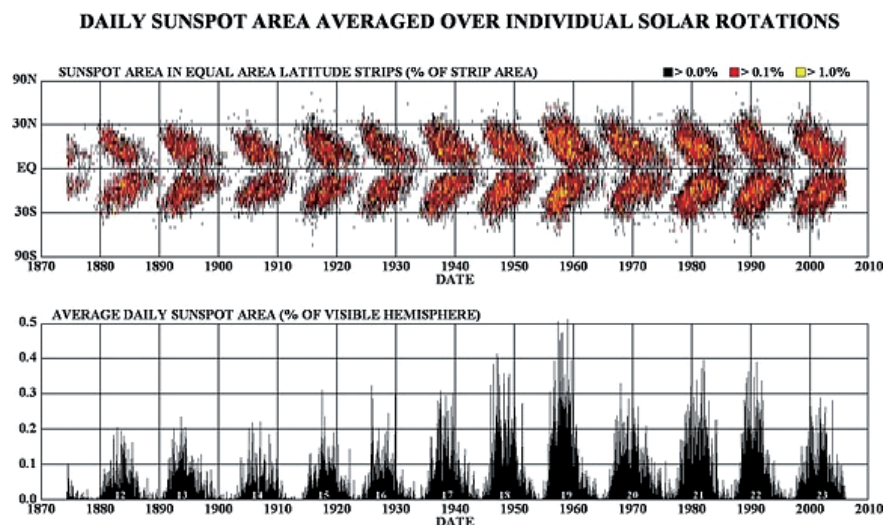


图 1.12 蝴蝶图——展示了黑子随时间的变化（图片来源于网络）

图 1.13 所示为 1933 年绘制的伦敦地铁线路图，该地铁线路图出版后迅速为乘客接受，并成为今日交通线路图的一种主流表现形式。



图 1.13 1933 年绘制的伦敦地铁线路图（图片来源于网络）

此外，统计应用的发展唤醒了可视化，数理统计把数据可视化变成了科学，世界大战和随后的工业及科学发展衍生的对数据处理的迫切需求把这门科学运用到了各行各业。

1.2.6 多维信息的可视编码

现代电子计算机的诞生是一个划时代的事件。计算机的出现彻底改变了数据分析工作^[2]。到 20 世纪 60 年代晚期，大型计算机已广泛应用于西方的大学和研究机构中，使用计算机程序绘制数据可视化图形逐渐取代了手绘图形。高分辨率的图形和交互式的图形分析，提供了手绘时代无法实现的表现能力。

1971 年出现了不规则多边形“星图”形态的表达。“星图”可直观地展示高维多元数据，查看哪些变量具有相似的值、哪些变量在数据集内得分较高或较低、变量之间是否有异常值。图 1.14 所示为美国城市犯罪率星图。

1973 年出现了神奇的卡通脸谱图——切尔诺夫脸谱（Chernoff Faces），如图 1.15 所示。该图用脸谱来分析多维度数据，即将 P 个维度的数据用人脸部位的形状或大小来表征。脸谱图分析法的基本思想是由 15~18 个指标决定脸部特征。若实际资料变量更多，则多出的变量将被忽略；若实际资料变量较少，则脸部有些特征将被自动固定。统计学曾给出了几种不同的脸谱图的画法，而对于同一种脸谱图的画法，将变量次序重新排列，得到的脸谱的形状也会有很大不同。

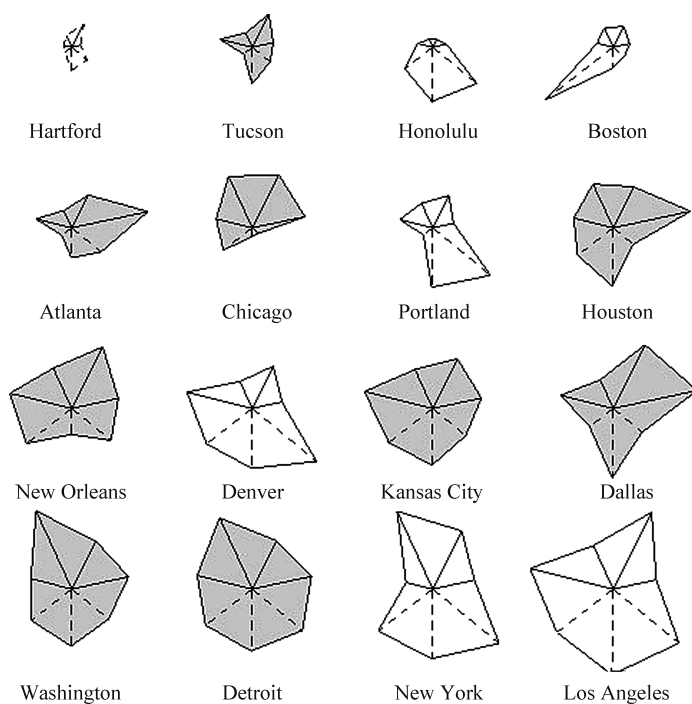


图 1.14 美国城市犯罪率星图（图片来源于网络）

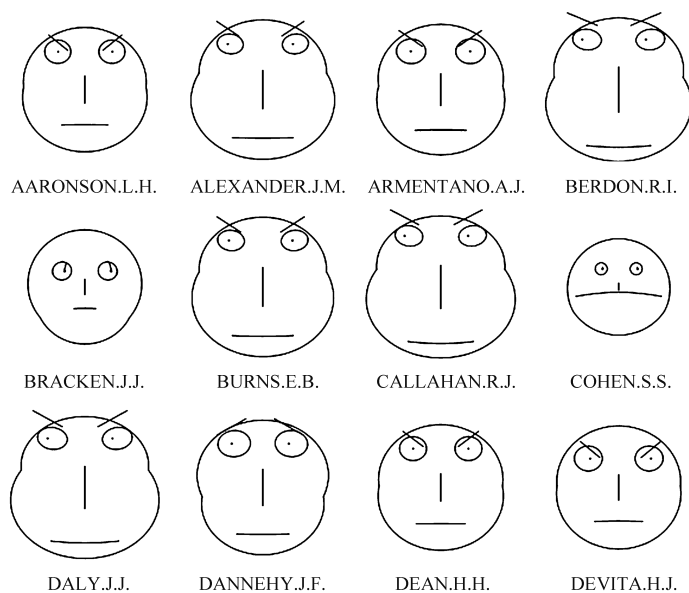


图 1.15 卡通脸谱图（图片来源于网络）

1.2.7 多维统计图

各种计算机系统、计算机图形学、图形显示设备、人机交互技术的发展激发了人们对

