



Basics of Data Visualization

数据可视化基础

中国地质大学（北京）信息工程学院

叶山

yes@cugb.edu.cn

Lab 1 第9-10题：大地测量学

可用 **Haversine** 公式计算两点间的球面距离。假设两点坐标分别为 $(\text{lat1}, \text{lon1})$ 和 $(\text{lat2}, \text{lon2})$,

则距离 d 的计算公式为：

```
 $\Delta\text{lat} = \text{lat2} - \text{lat1}$ 
 $\Delta\text{lon} = \text{lon2} - \text{lon1}$ 
 $a = \sin^2(\Delta\text{lat}/2) + \cos(\text{lat1}) * \cos(\text{lat2}) * \sin^2(\Delta\text{lon}/2)$ 
 $d = 2 * \arcsin(a^{1/2}) * R$ 
```

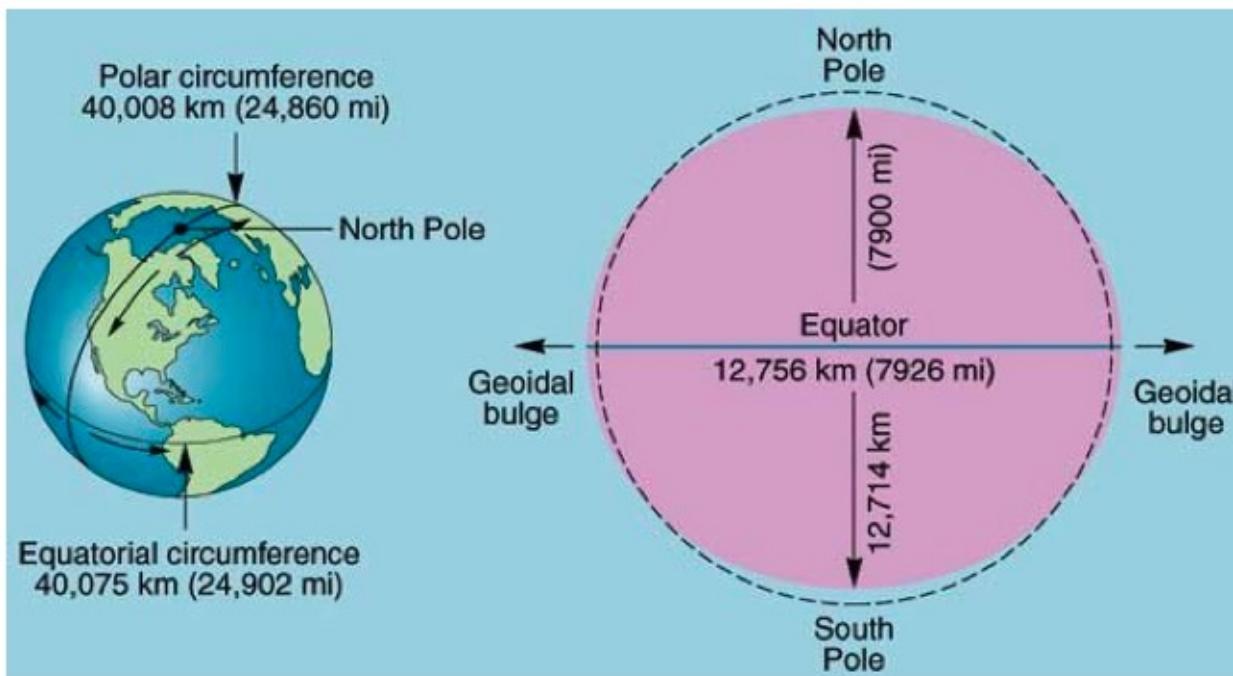
其中 R 为地球半径，取 6371 千米。注意：经纬度坐标本质上是角度，需要转换为弧度。

9. 在 R 中编写一个函数 `haversine_dist()`，依次输入两点的经纬度坐标，返回两点间的距离。该

功能可用 `geosphere` 包中的 `distVincentyEllipsoid` 函数轻松实现，但本次作业请勿使用。

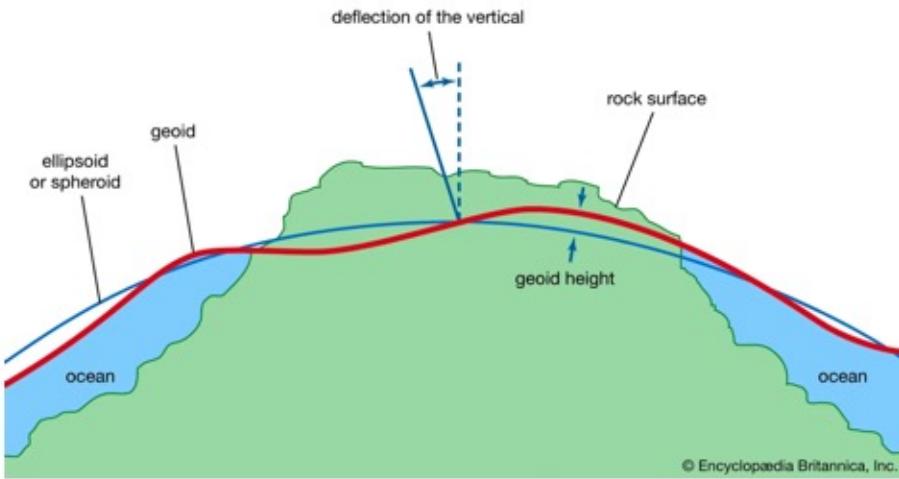
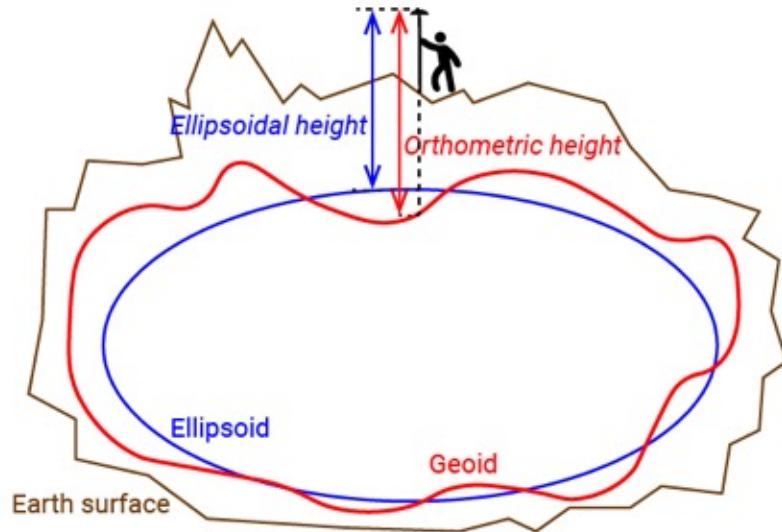
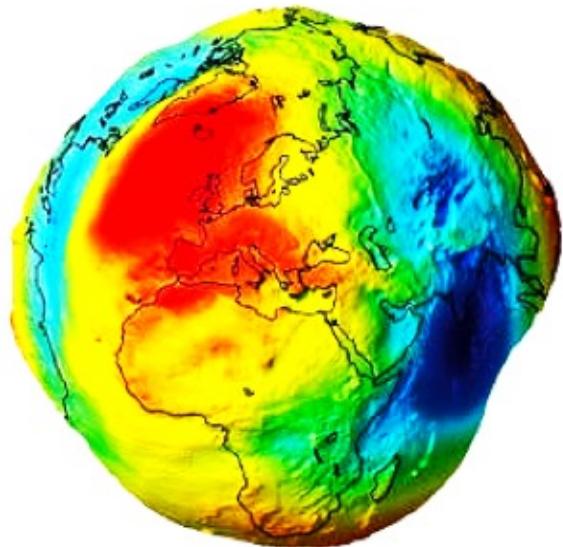
10. 在恐龙化石的数据集里，所有化石的经纬度都因板块活动而发生了改变，请问它们的平均移动距离
为多少千米？

Lab 1 第9-10题：大地测量学



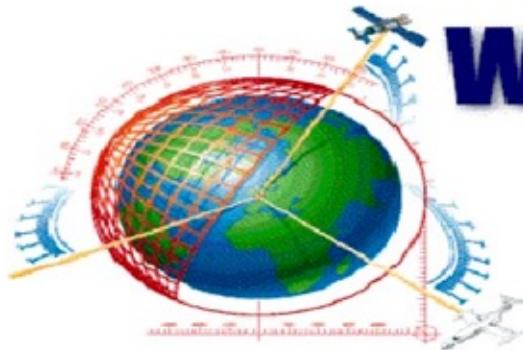
Lab 1 第9-10题：大地测量学

Geoid 大地水准面



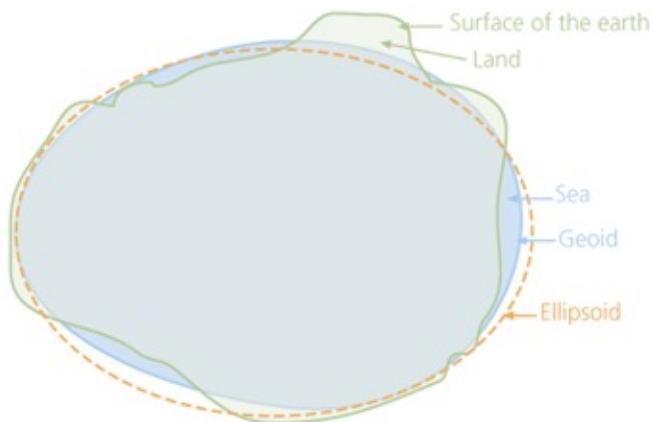
真实的地表高度，包括山脉、平原、海底等 → 去掉地表高度，模拟海水在重力和离心力的作用下自由流动，达到平衡后的地表（大地水准面 Geoid）→ 把大地水准面用数学方法近似为一个椭球体表面（参考椭球 Reference Ellipsoid）

Lab 1 第9-10题：大地测量学



WGS 84
World
Geodetic
System
1984

Model of the Earth



distVincentyEllipsoid说明文档

The WGS84 ellipsoid is used by default. It is the best available global ellipsoid, but for some areas other ellipsoids could be preferable, or even necessary if you work with a printed map that refers to that ellipsoid. Here are parameters for some commonly used ellipsoids:

ellipsoid	a	b	f
WGS84	6378137	6356752.3142	1/298.2572235
GRS80	6378137	6356752.3141	1/298.2572221
GRS67	6378160	6356774.719	1/298.25
Airy 1830	6377563.396	6356256.909	1/299.3249646
Bessel 1841	6377397.155	6356078.965	1/299.1528434
Clarke 1880	6378249.145	6356514.86955	1/293.465
Clarke 1866	6378206.4	6356583.8	1/294.9786982
International 1924	6378388	6356911.946	1/297
Krasovsky 1940	6378245	6356863	1/298.2997381

Pop Quiz

回答正确得2分
回答错误得1分
缺席不得分

1. 关系型数据库的结构化查询语言是?
A. Python | B. JavaScript | C. SQL | D. PostgreSQL

2. 以下哪一项是古气候学领域推出的新型数据格式?
A. Shapefile | B. LipD | C. XML | D. GeoJSON

3. 以下哪种方法被用于比较三个（及以上）组之间的均值差异?
A. 回归分析 | B. 学生T检验(T-test) | C. 卡方检验 (χ^2 test) | D. 方差分析 (ANOVA)

4. 分层型数据库中，数据记录被组织成什么结构?
A. 链表 | B. 栈 | C. 树 | D. 二维表

5. 以下哪种方法被用于分析总体中某一部分的重要程度?
A. 结构分析法 | B. 分组分析法 | C. 对比分析法 | D. 平均分析法



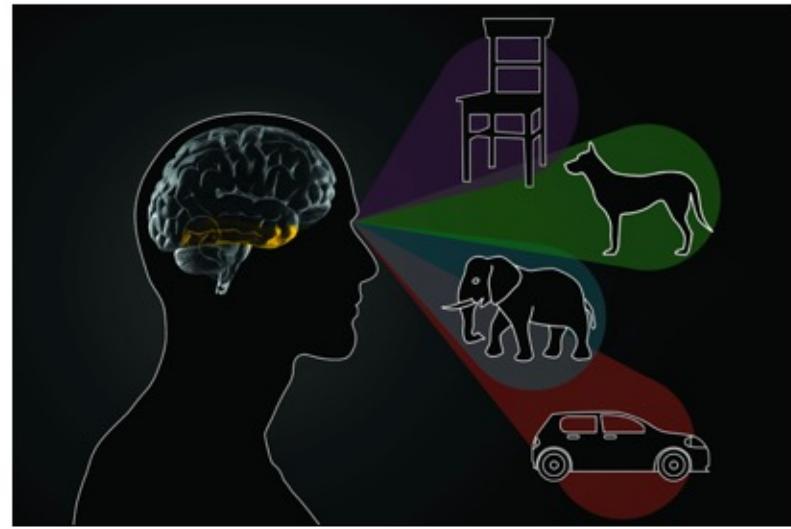
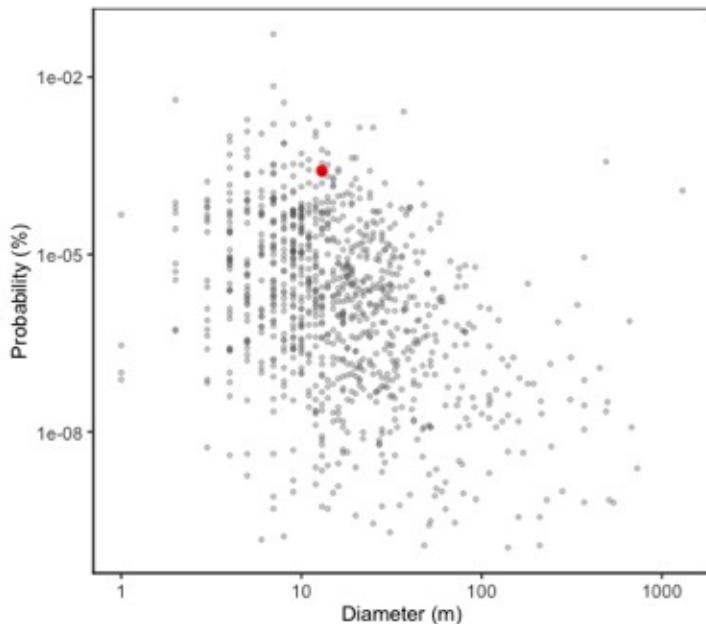
本节内容

可视化的定义、作用、发展

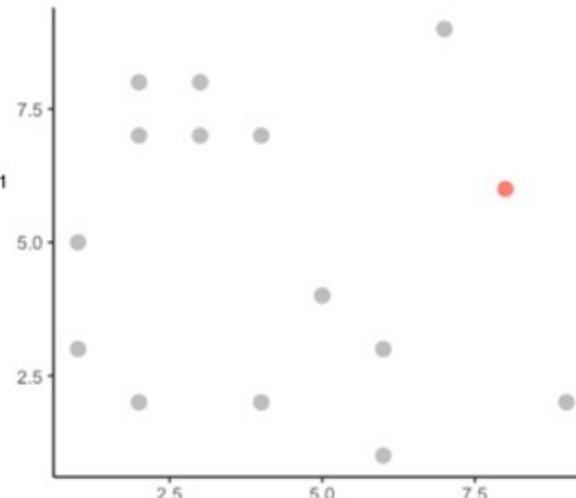
人类视觉



最高带宽：100MB/s



超过50%的人脑功能被用于视觉的感知：解码可视信息、高层次可视信息处理、思考可视符号等



视觉处理是并行的
(视觉突变现象)

可视化的定义

Visualize (v.)：生成符合人类感知的图像。

Visualization (n.)：通过视觉导向的方式把某个事物呈现出来的过程、能力或结果。

可视化的定义

可视化是一种用来帮助某人完成某项任务的图像。

可视化定义的两个要素

- 必须是一种图像：通过视觉感官传递信息。
- 具有现实的目的性。

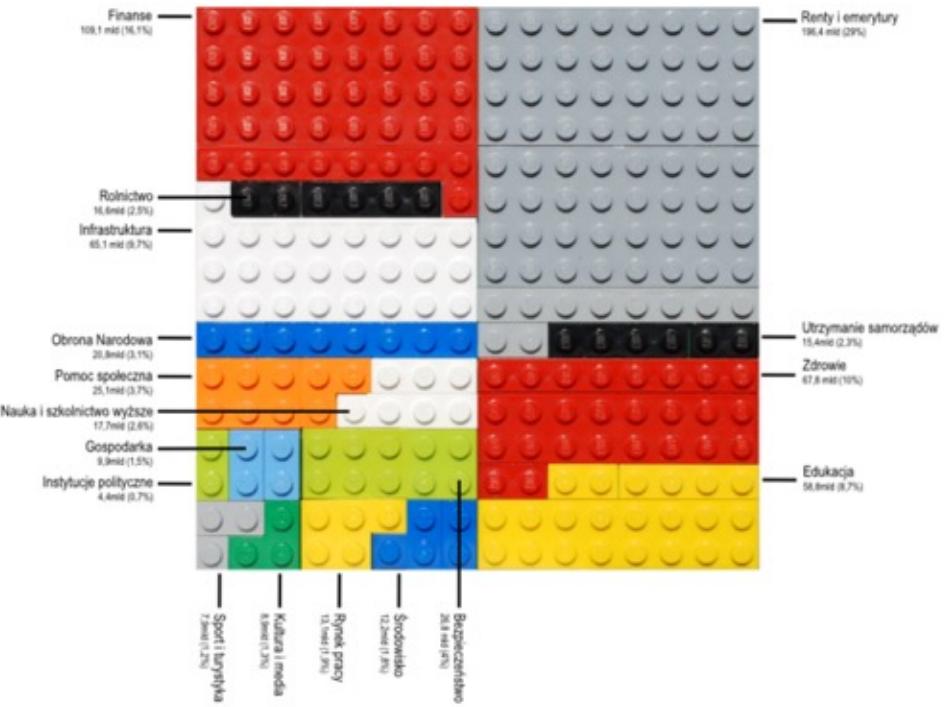


可视化的定义

- 可视化是“视觉”的。
- 更加广义的“图像”：动图、动画、交互性图像、非传统意义的数据图片

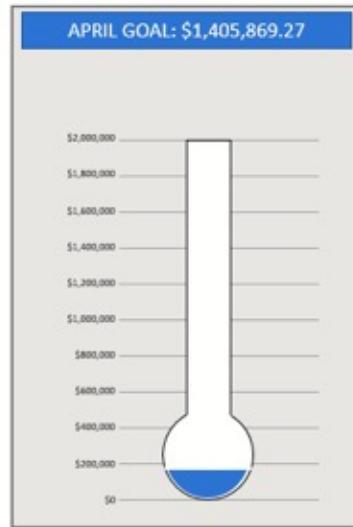


- 80%
- 60%
- 40%
- 20%
- 0%



可视化的定义

- 可视化存在的目的：被用于某些特定的场景或任务。
- 在进行数据可视化的操作时，应该思考可视化作品要用来做什么。

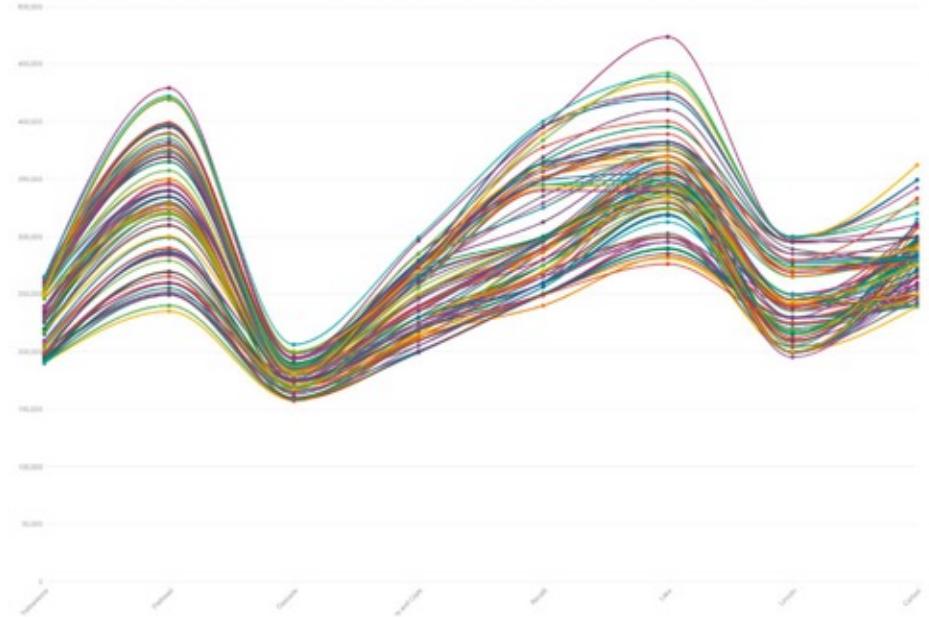
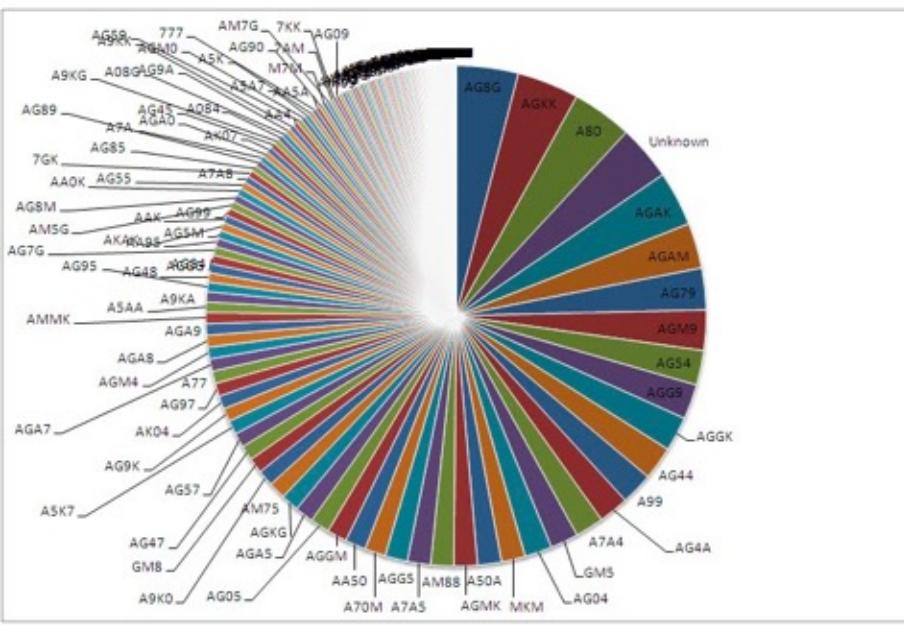


这是数据可视化作品吗？
它和雪堆、乐高图的区别在哪里？

没有明确的目的性，因
此不属于可视化

可视化的定义

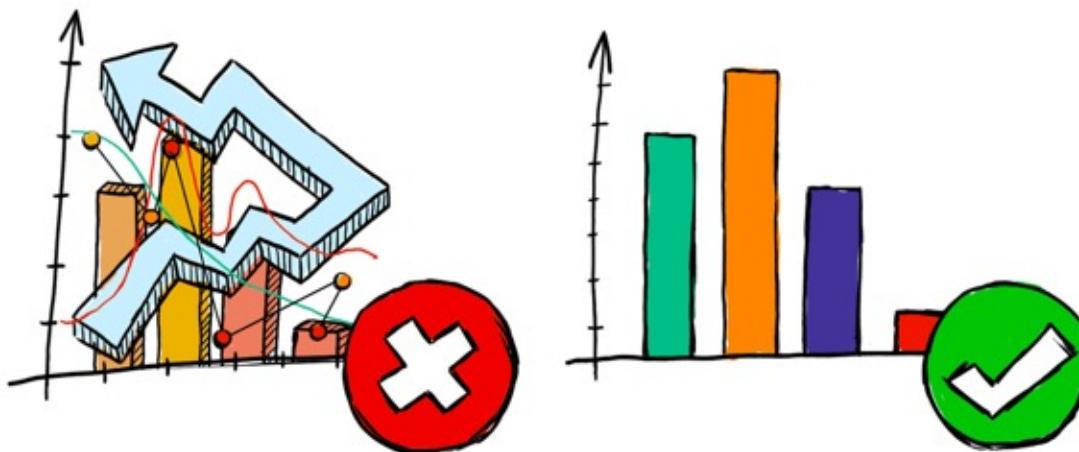
- 目的 ≠ 结果



无效的可视化作品属于数据可视化吗？

可视化的定义

- 制作可视化作品并不难，但是制作优秀的、有效的可视化作品不容易。



可视化的定义

从计算机科学的角度定义可视化：利用人眼的感知能力对数据进行交互的可视表达以增强认知的技术称为可视化。它将不可见或难以直接显示的数据转化为可感知的图形、符号、颜色、纹理等，以增强数据识别效率并传递有效信息。

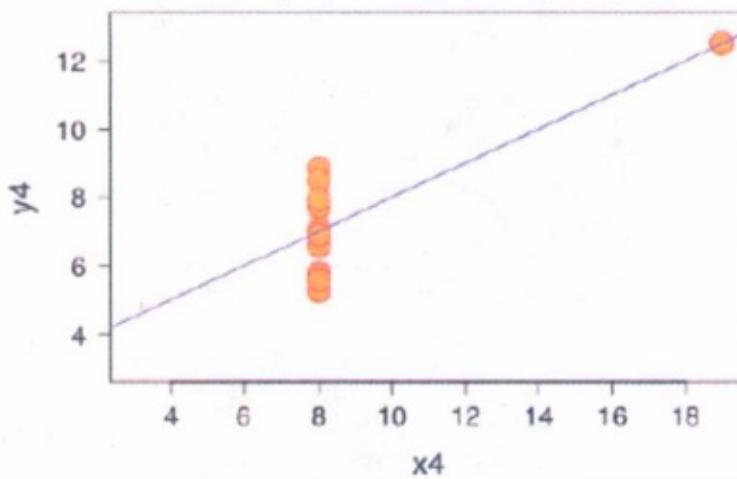
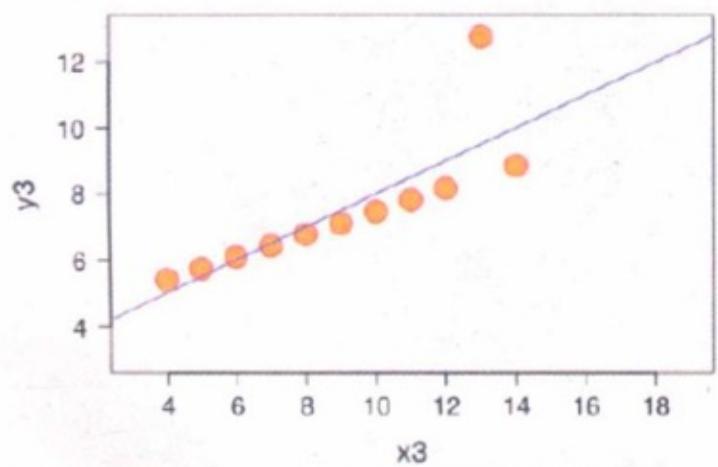
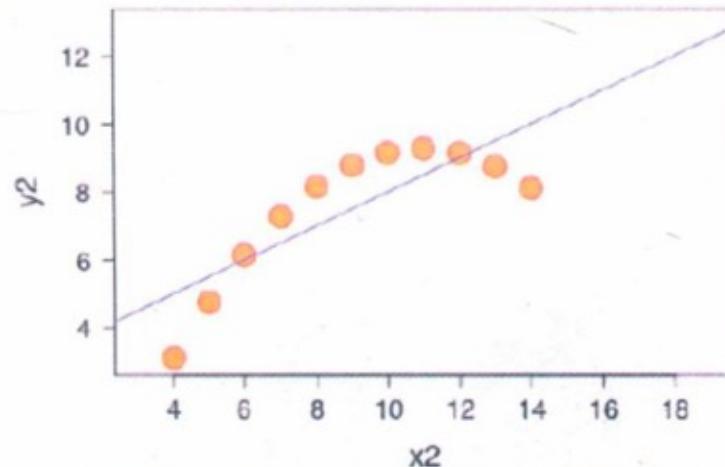
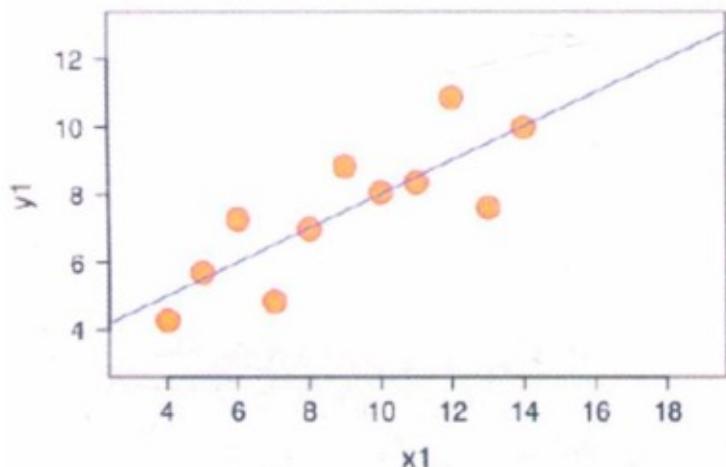
可视化的定义

4组二维数据，通过表格呈现，难以直观识别它们的区别。

x1		x2		x3		x4	
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

平均值、方差、标准差都相同
简单的统计学方法也无法区分

可视化的定义



把肉眼难以观察出规律的数据用可视化作品来呈现后，能迅速地发现它们的模式和规律。

可视化的定义

- 从信息加工的角度定义可视化：通过图形化符号来提高信息处理效率的工具。



 **1/10**
OF A SECOND

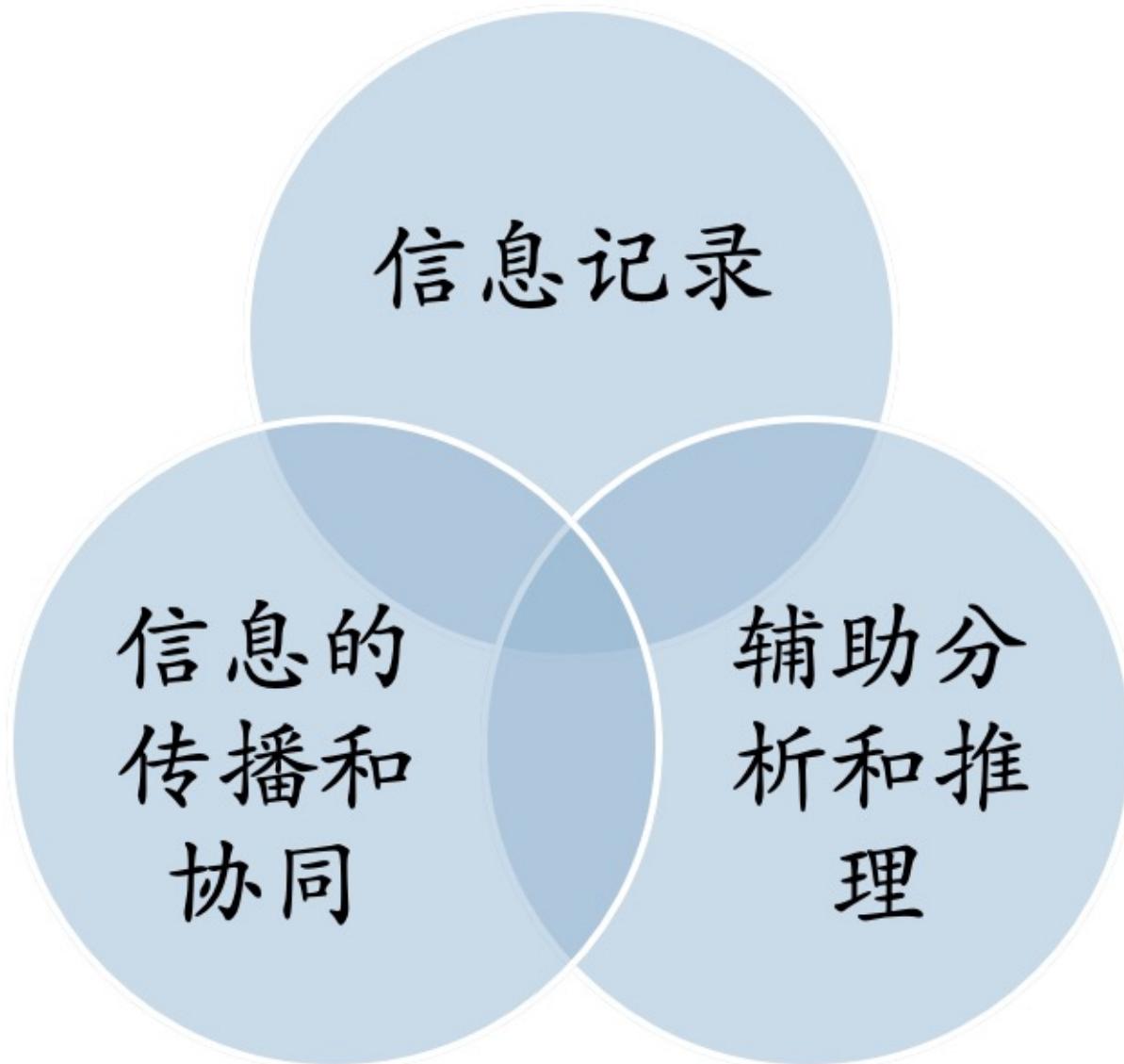
is all it takes you to
understand a **visual scene**



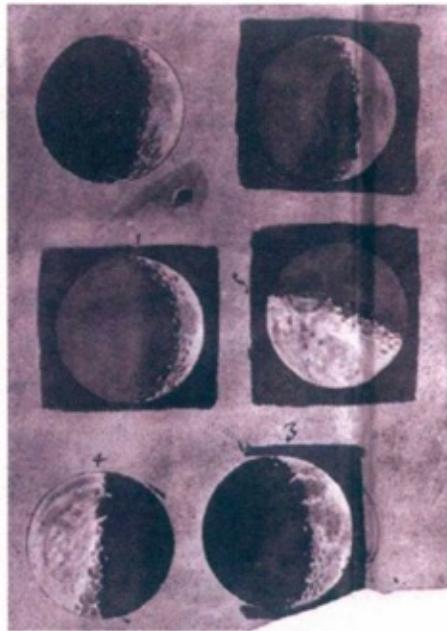
 **60**
SECONDS

is the average time it takes
to read **200-250 words**

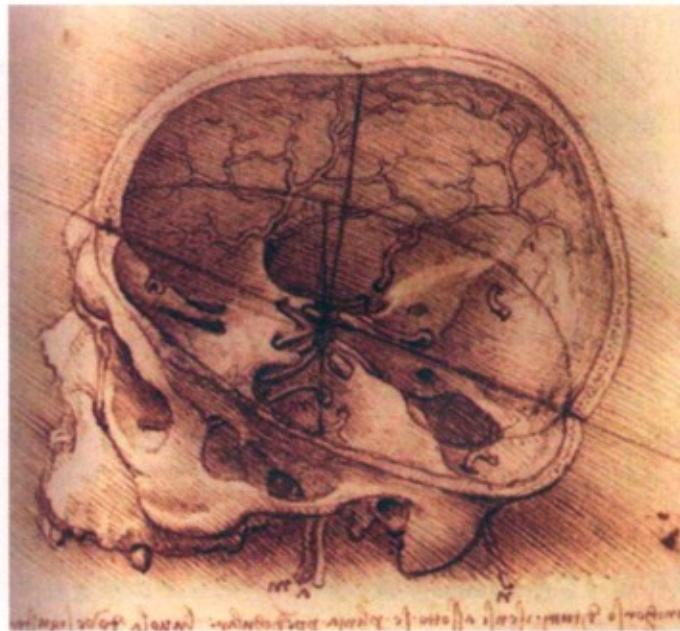
可视化的 作用



信息记录



伽利略于1616年手绘的月亮周期示意图



达芬奇绘制的人类头盖骨示意图

图像化的方式能让数据和信息更加长久地留存。

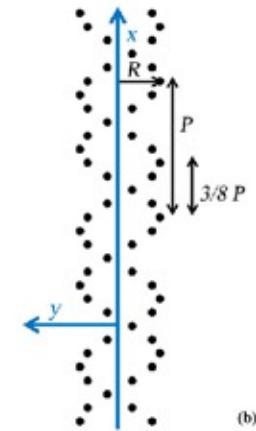
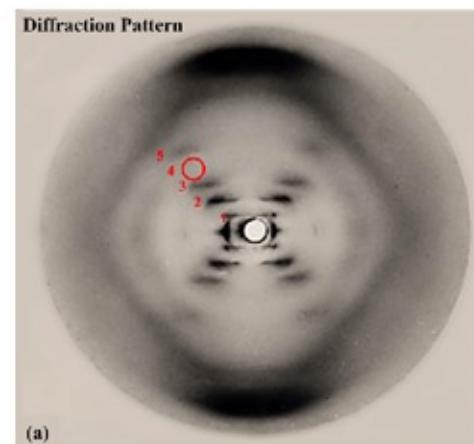
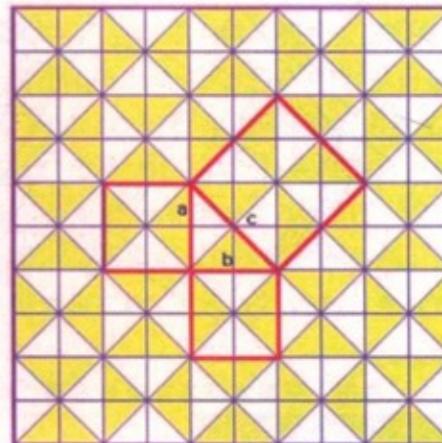
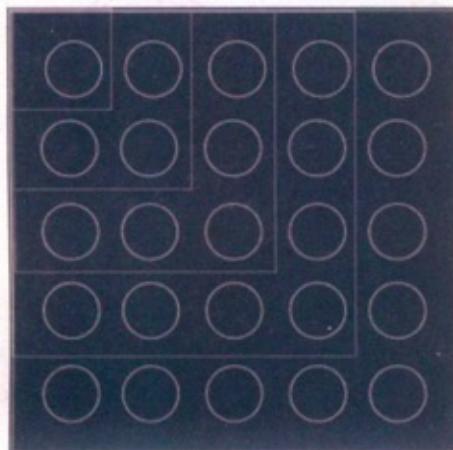


Photo 51: 1952年由伦敦大学学院的罗莎琳·富兰克林所拍摄的一张DNA的X光衍射图片，阐明了DNA的双螺旋结构。被誉为史上最重要的
一张照片（BBC）。

辅助分析和推理

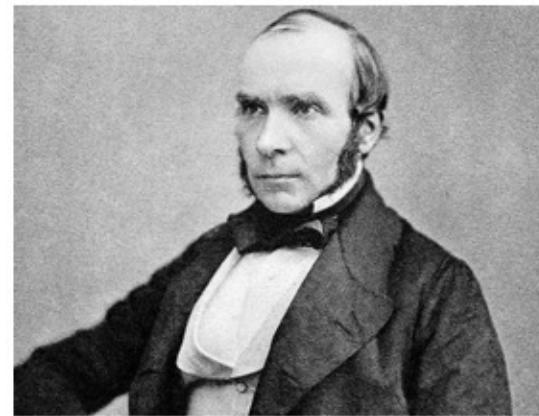
- 通过可视化呈现能提升对信息认知的效率，并引导别人从可视化结果里分析和推理出有效信息。这种直观的信息感知机制极大地降低了数据理解的复杂度，突破了常规统计分析方法的局限性。



左：对奇数的和的可视化， $1 + 3 + 5 + 7 + 9 = 25$ ；

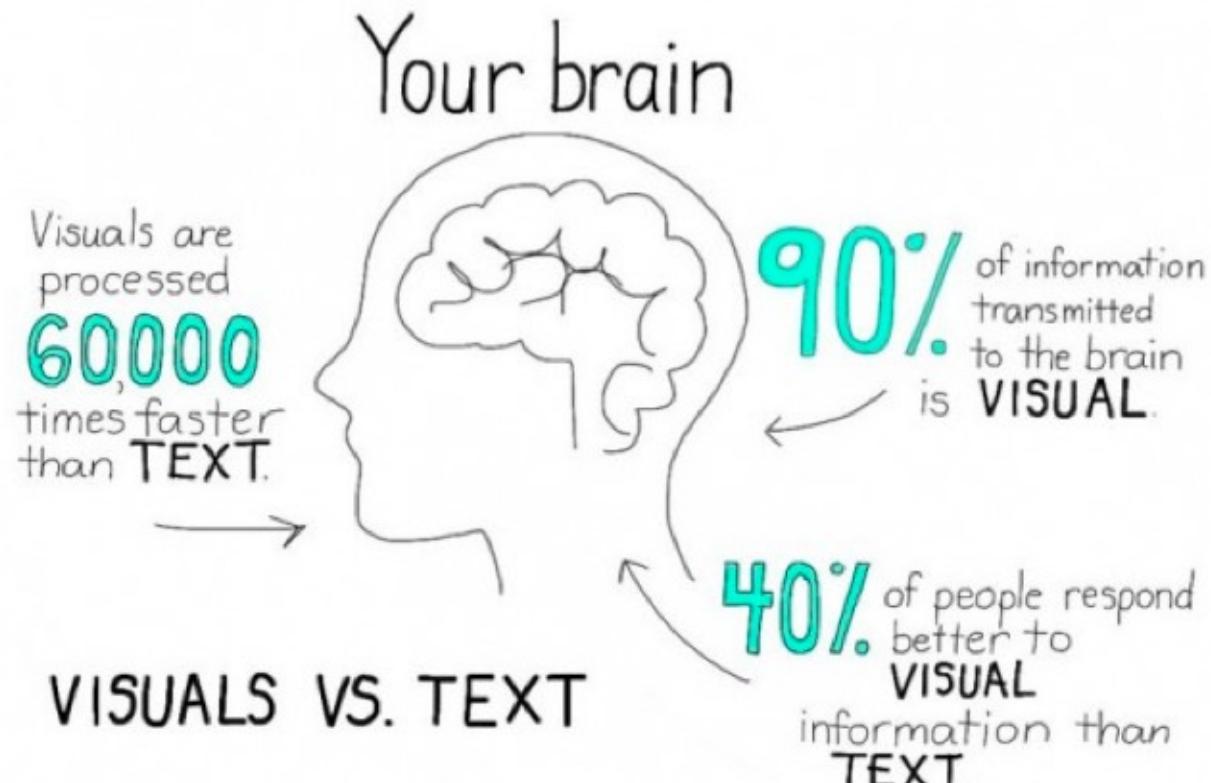
右：中国古代用于证明勾股定理的图形化证明方法， $c^2 = a^2 + b^2$

辅助分析和推理



John Snow伦敦霍乱地图（1854）：空间分析、流行病学的起源

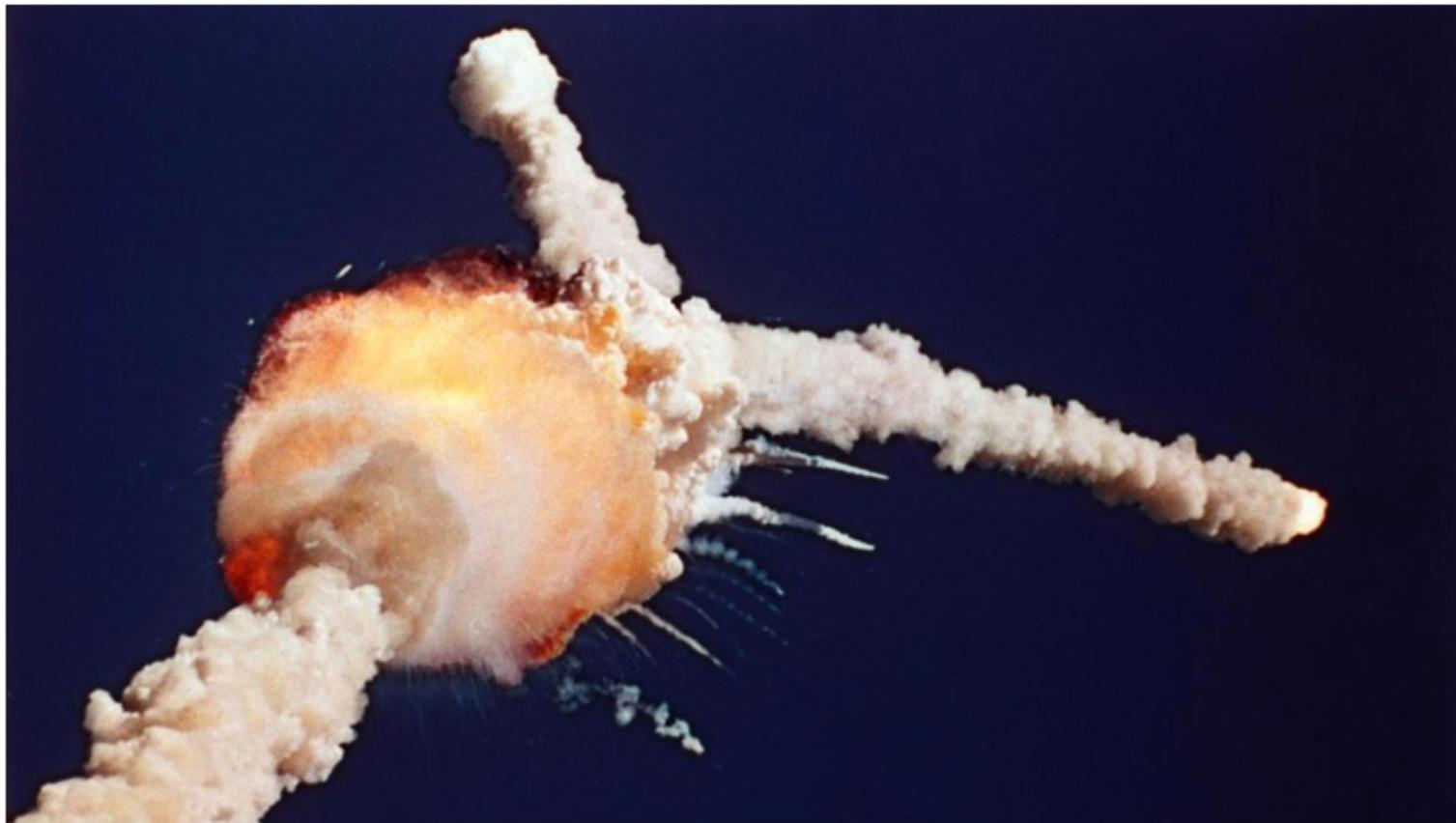
信息的传播和协同



annemacoll.com

人类从外界获取的信息中，90%从**视觉**输入。

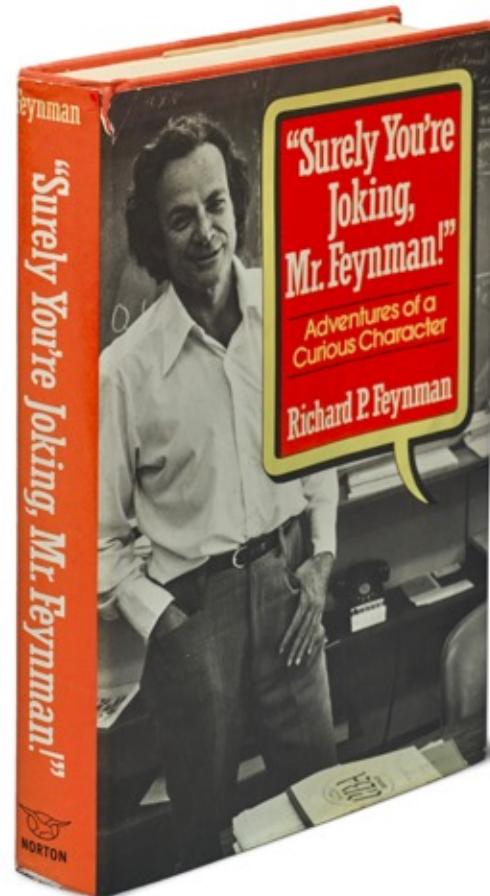
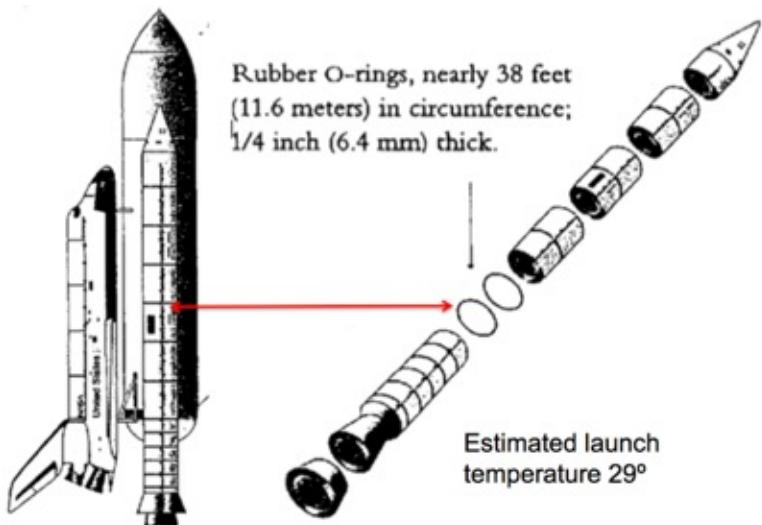
信息的传播和协同



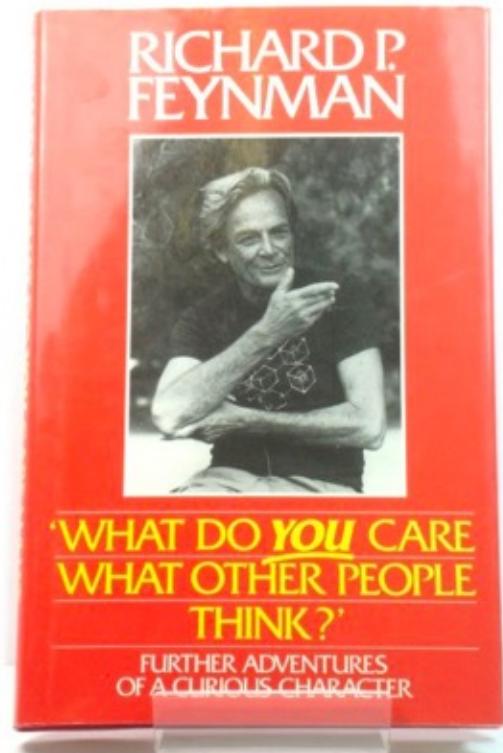
1986年1月28日美国“挑战者”号航天飞机爆炸事故

信息的传播和协同

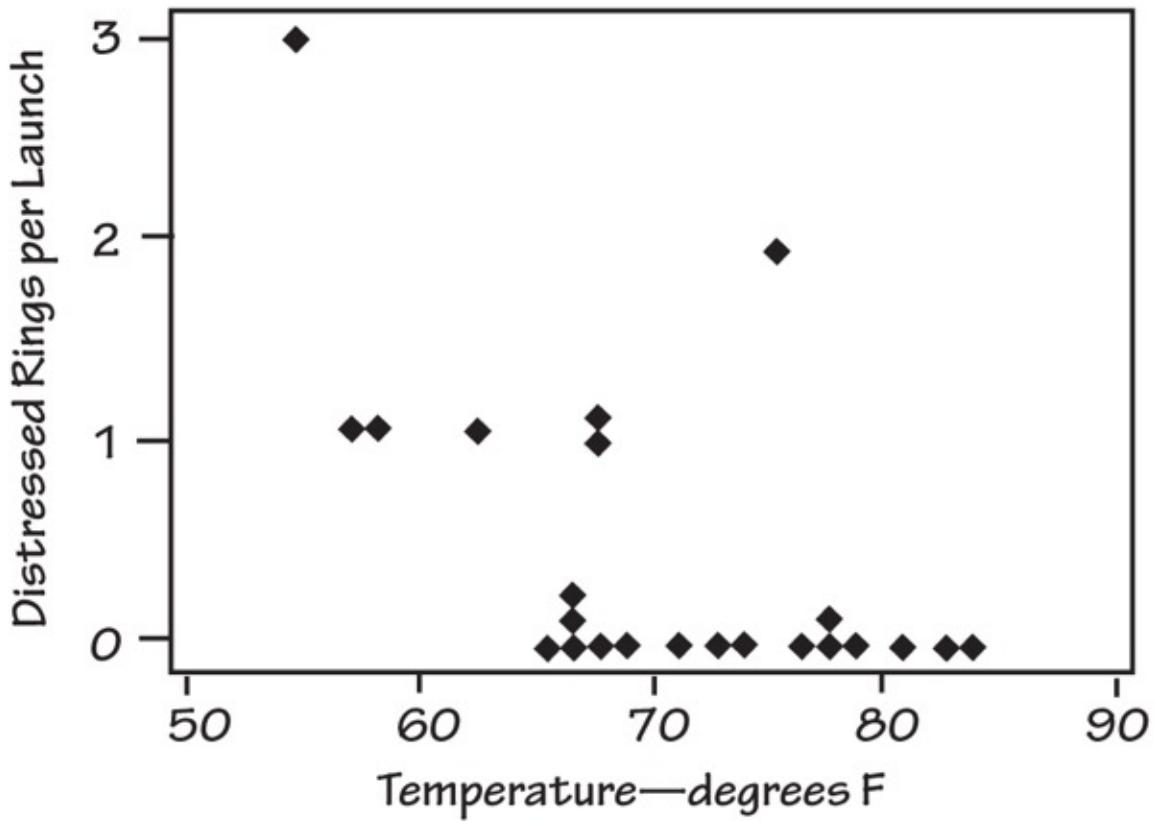
O形环



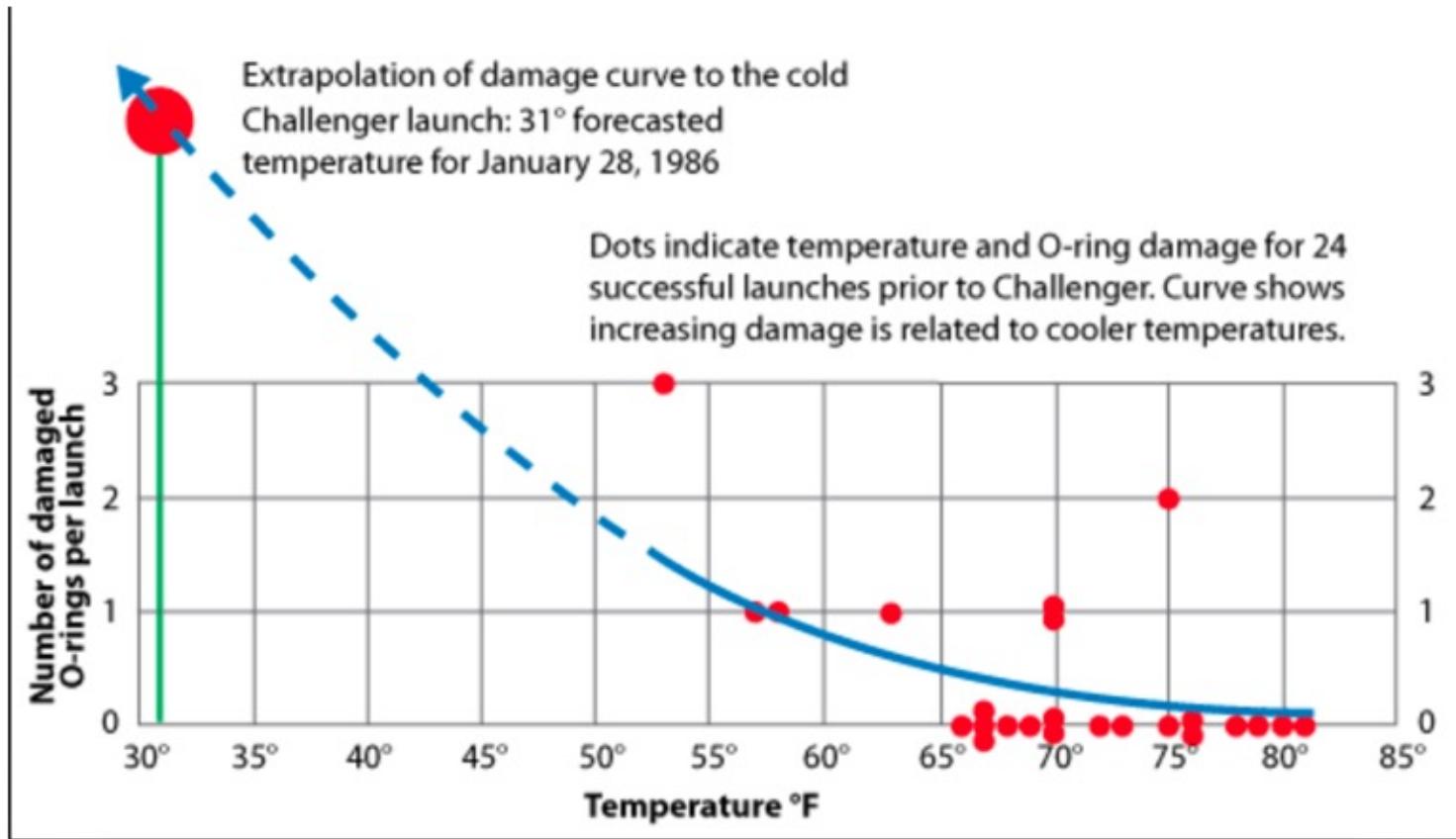
理查德·费曼



信息的传播和协同



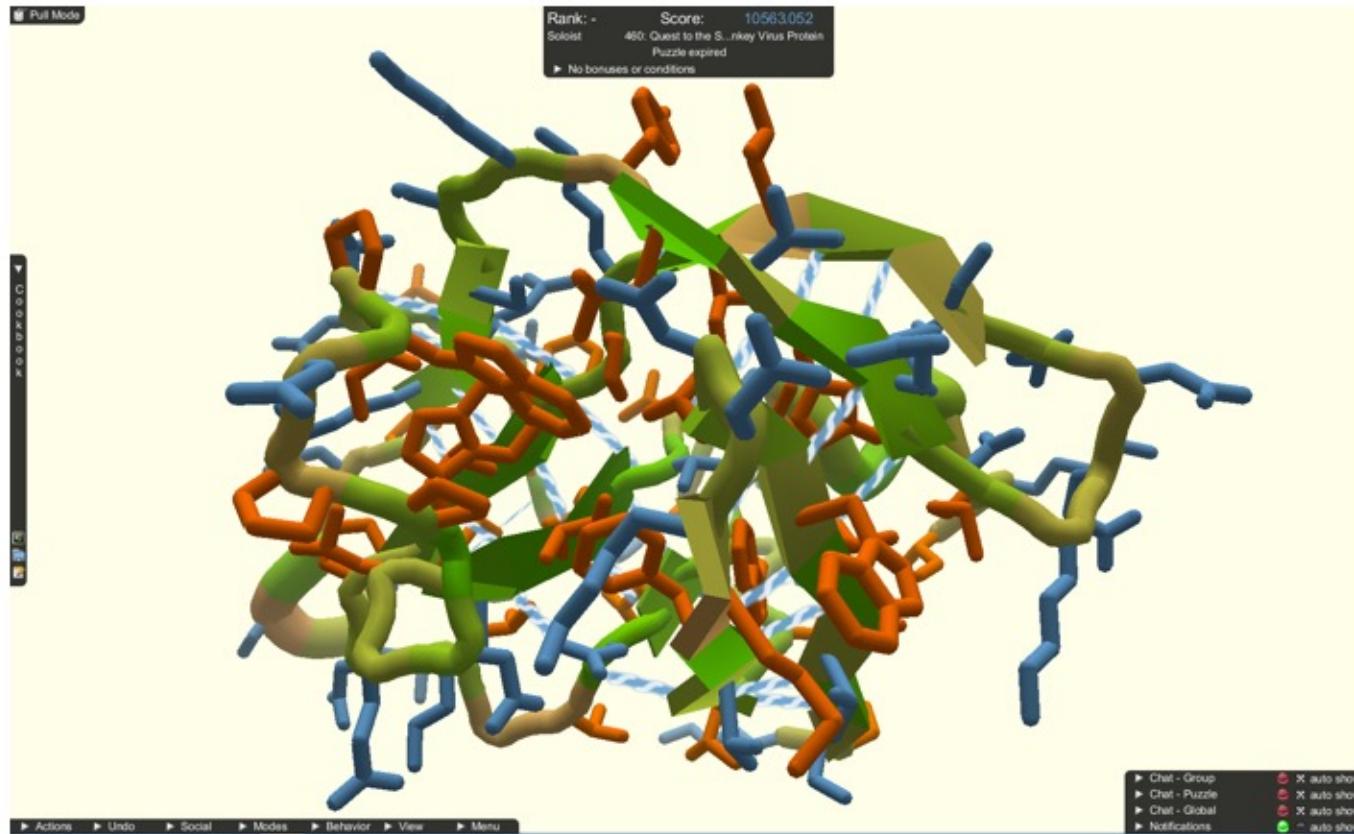
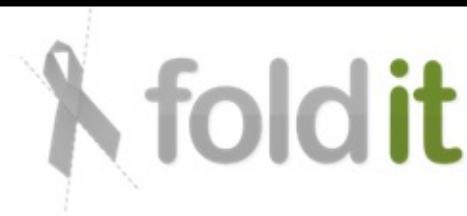
信息的传播和协同



信息的传播和协同

玩家通过交互式可视化游戏预测出正确的蛋白质结构的速度比任何算法都快。

人类能借助可视化，凭直觉解决计算机没办法解决的问题。



David Baker



Zoran Popović

<https://fold.it/>

可视化简史

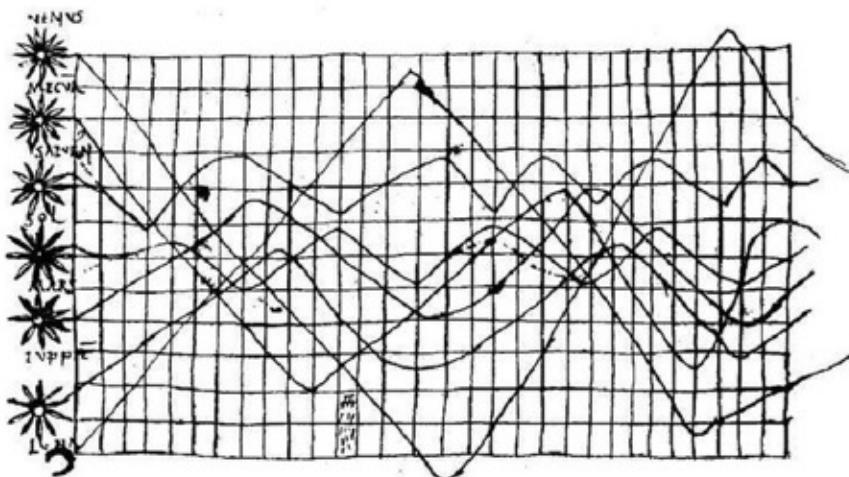
可视化的
发展

地图

科学制图

工程制图

统计制图

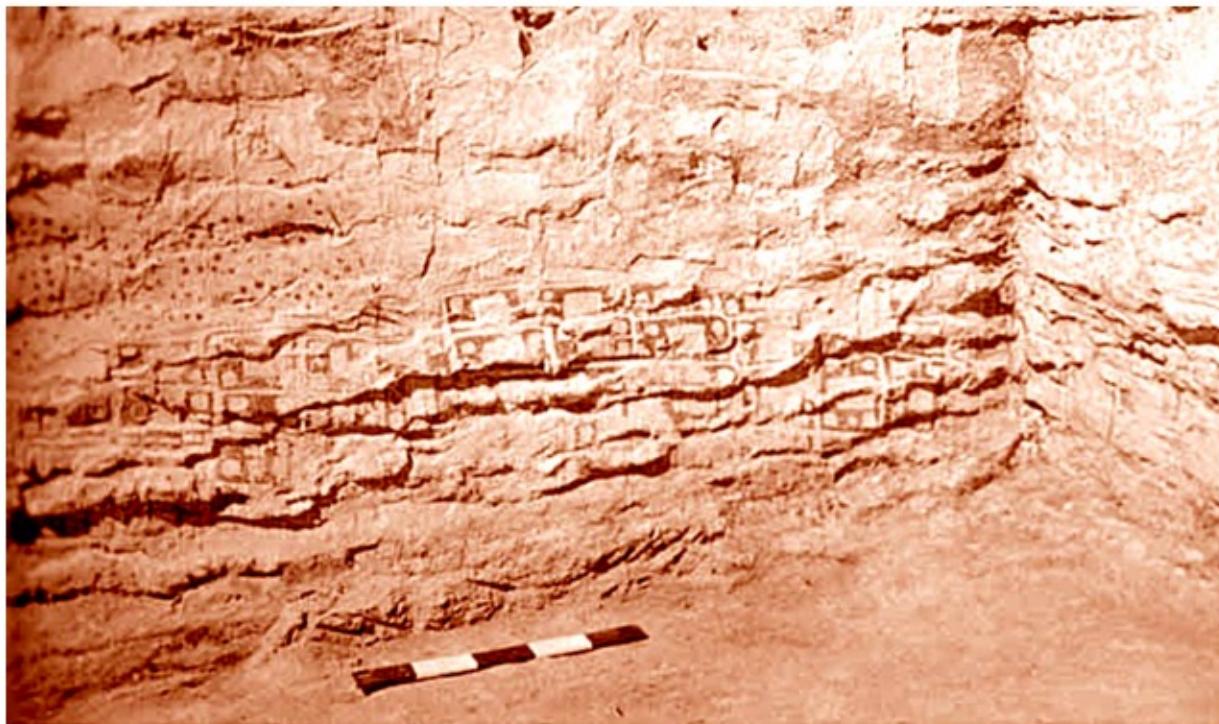


可视化简史



可视化简史

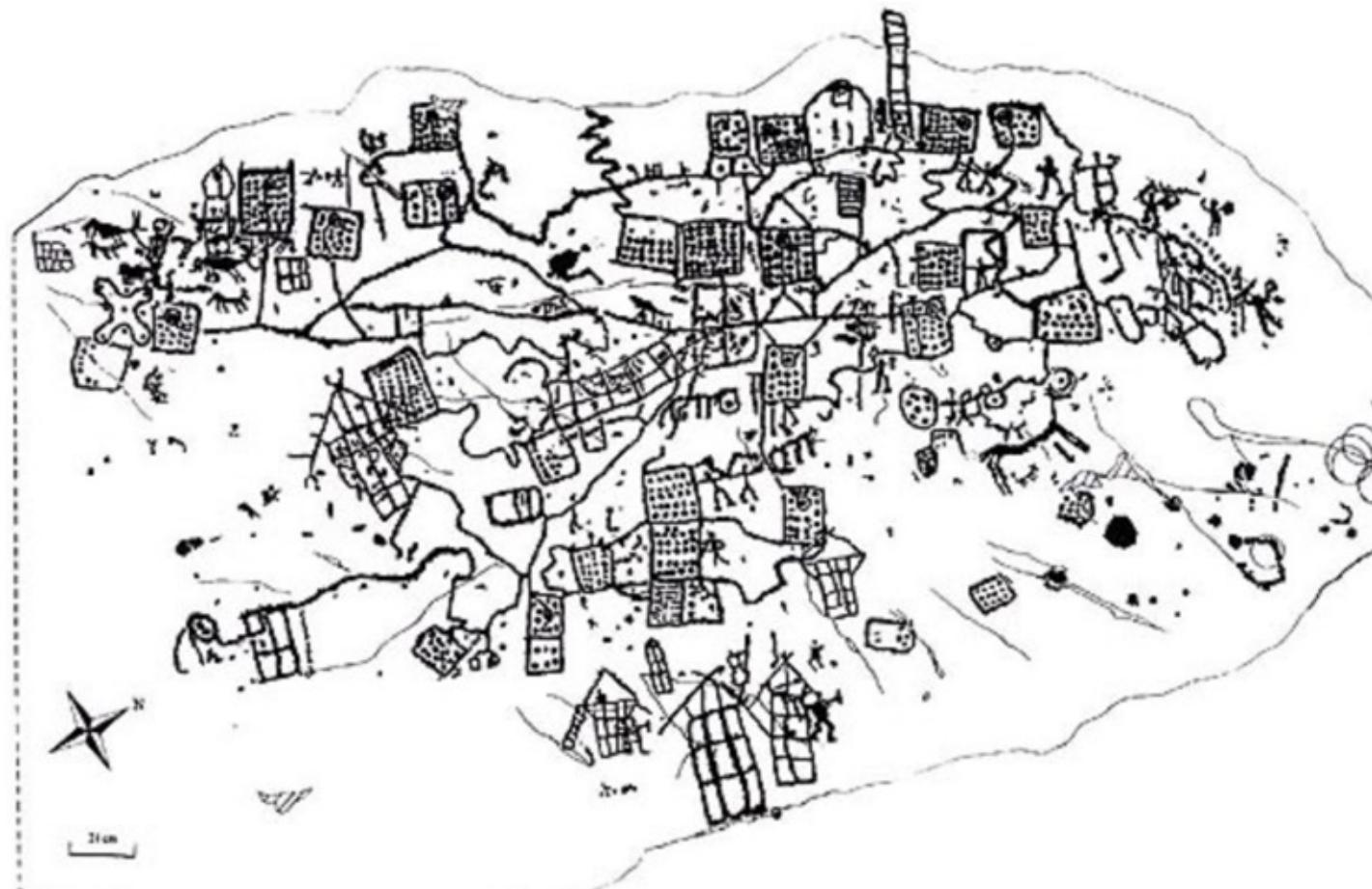
早期发展



土耳其
加泰土丘



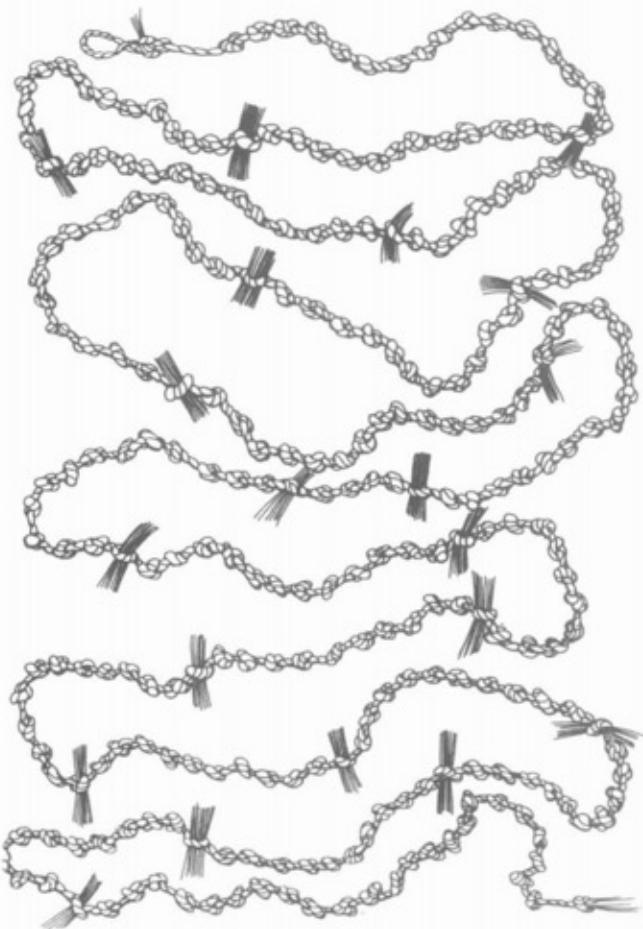
可视化简史



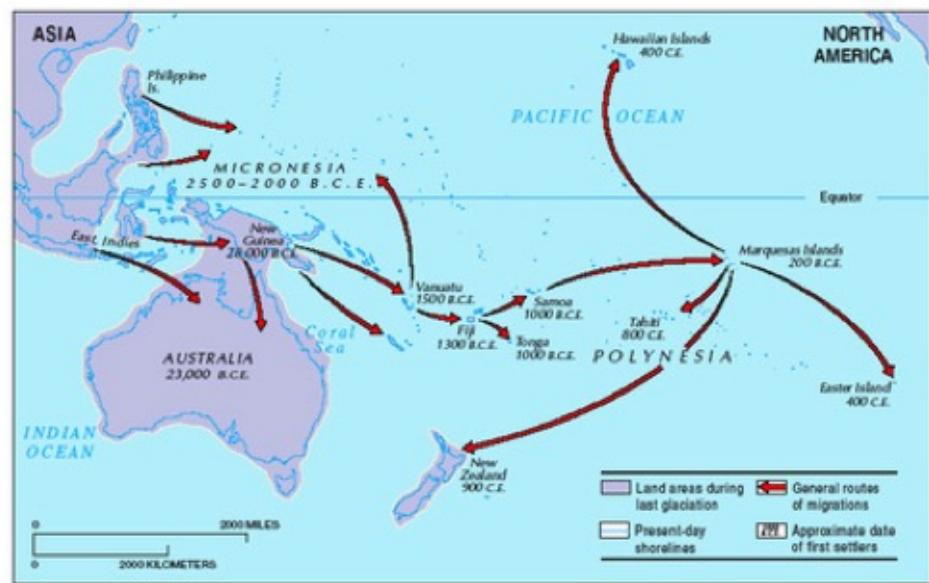
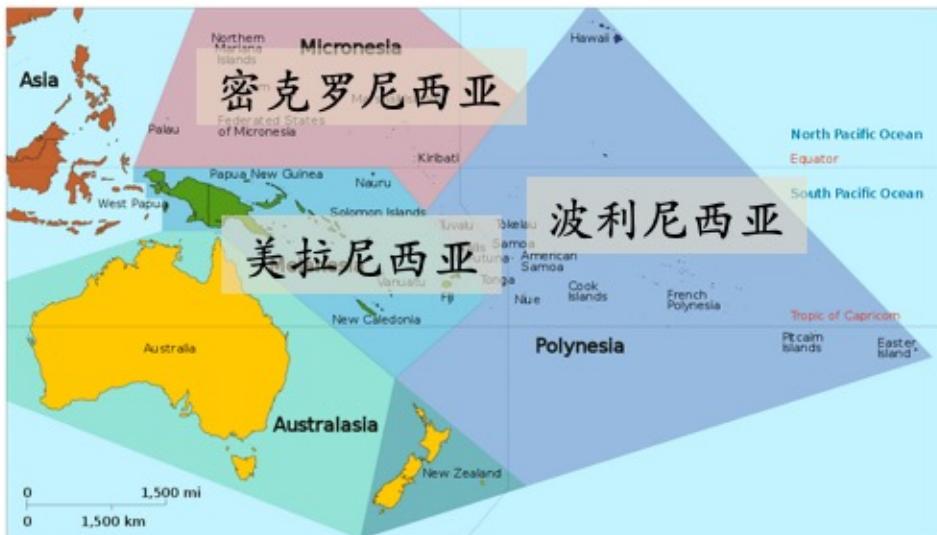
意大利伦巴第地区发现的贝多莱纳地图（Bedolina Map）：宽4.3米、高2.4米的岩画地图。该图被认为是当地早期居民用来宣称土地所有权的“地契”。

可视化简史

早期发展

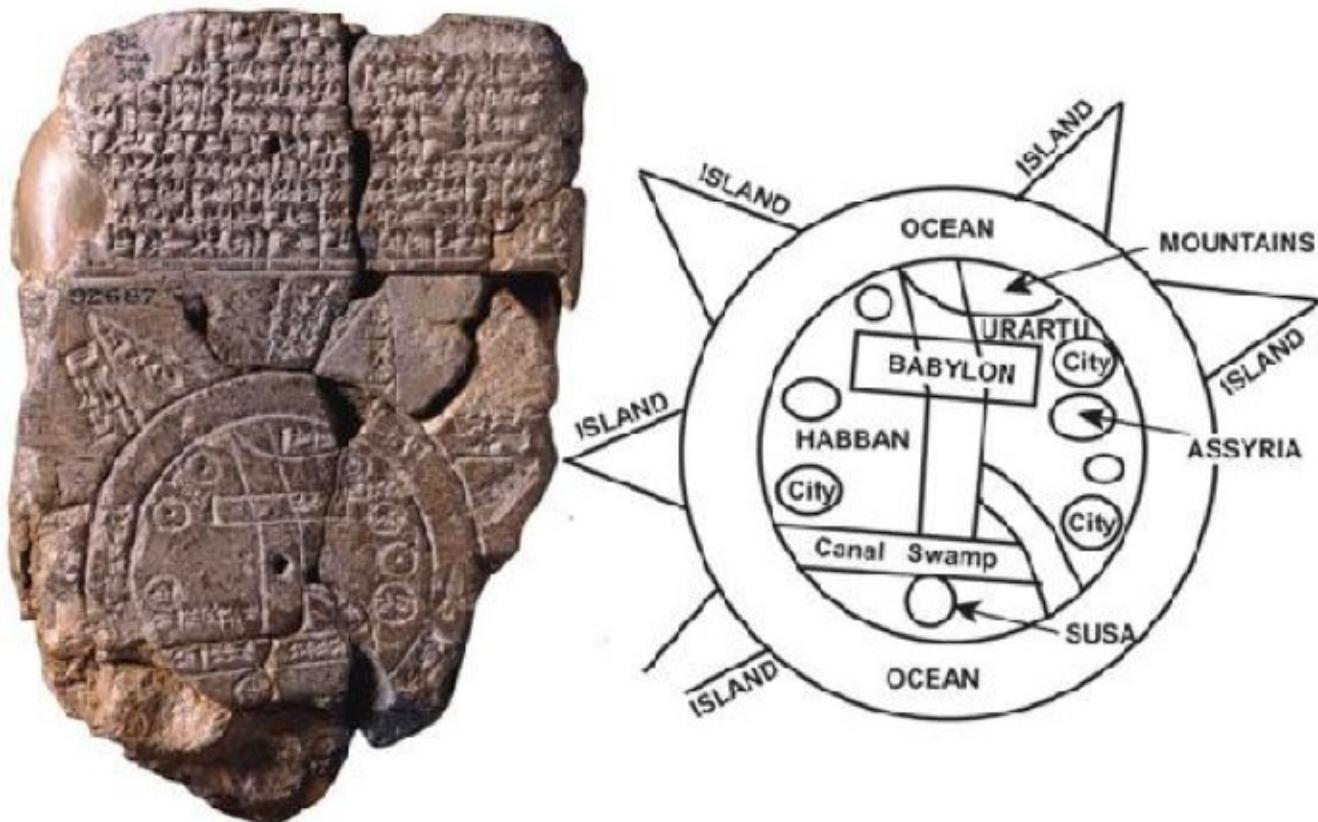


绳结地图



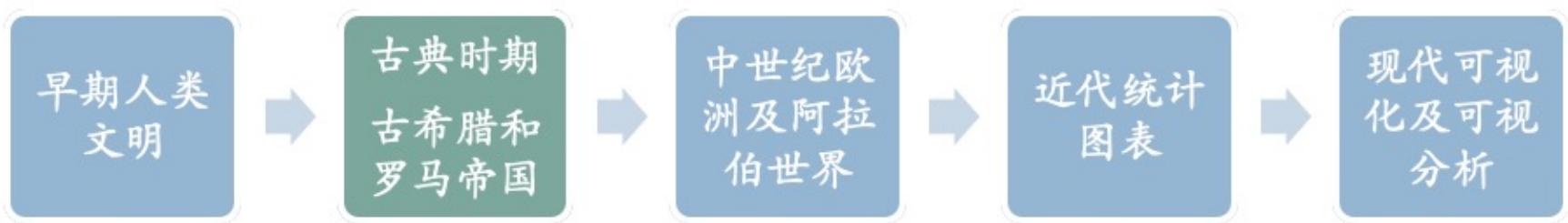
可视化简史

早期发展



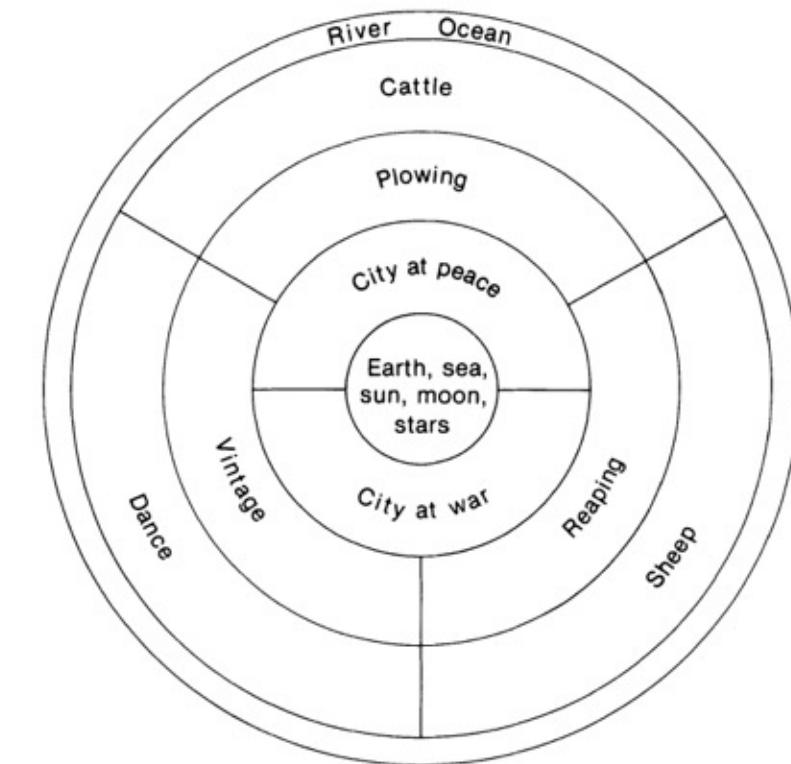
巴比伦地图

可视化简史



可视化简史

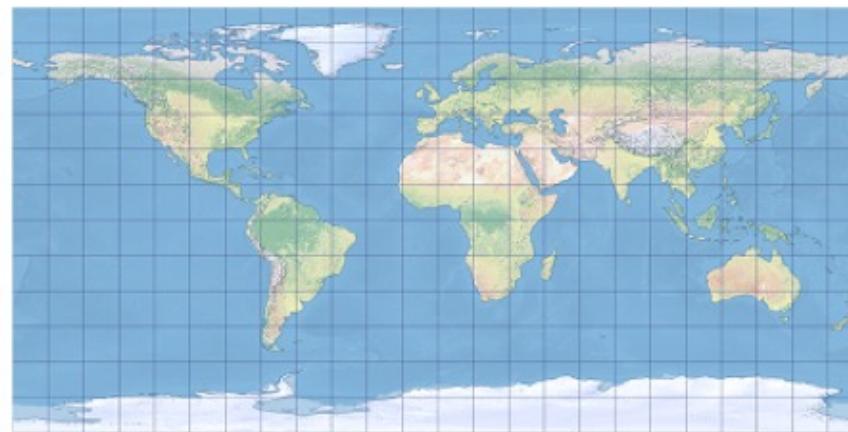
古典时期



荷马史诗：阿喀琉斯的盾牌（复原图）

可视化简史

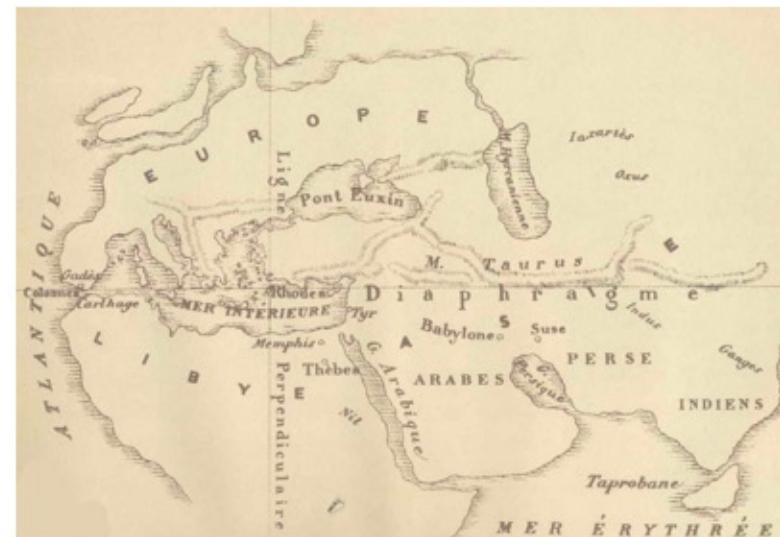
古典时期



米利都的泰勒斯：简易投影地图（plate carree）



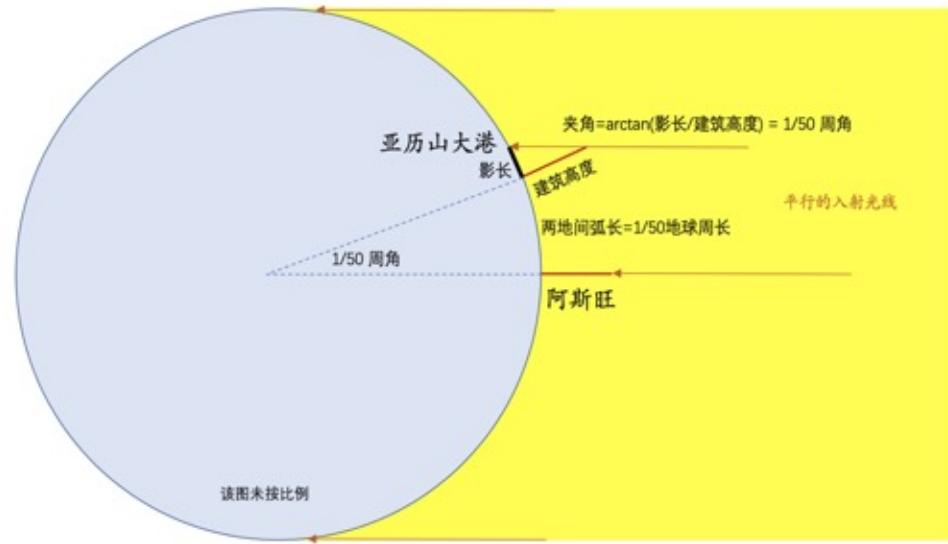
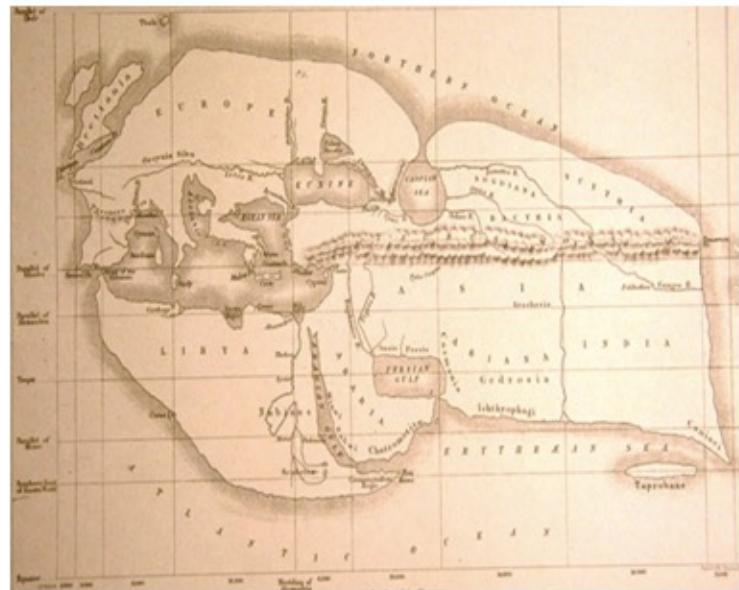
阿那克西曼德的三洲地图（复原图）



狄凯尔卡斯的世界地图（经纬线雏形）

可视化简史

罗马帝国



埃拉托斯特尼

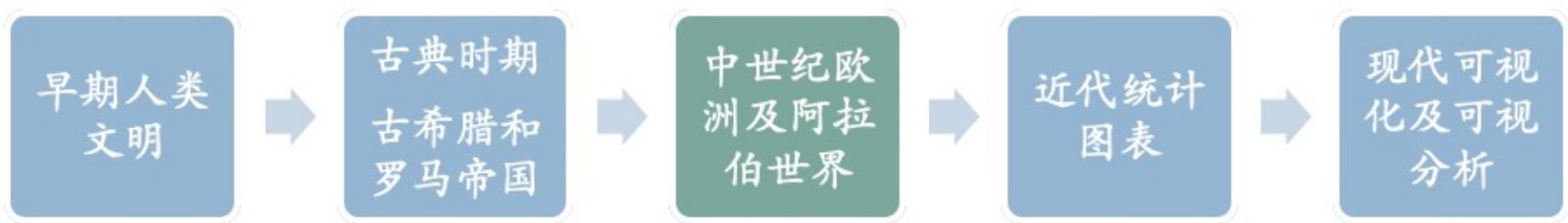
可视化简史

罗马帝国



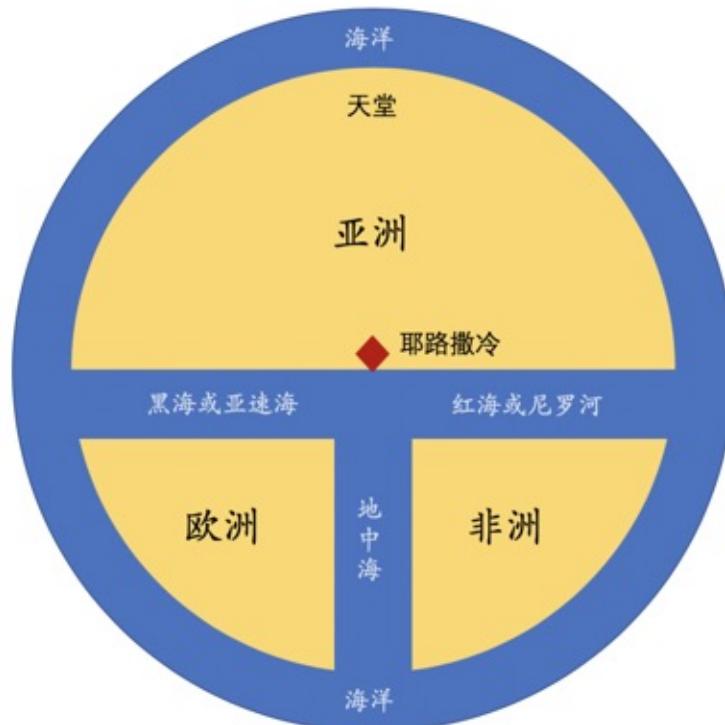
克劳狄乌斯·托勒密

可视化简史



可视化简史

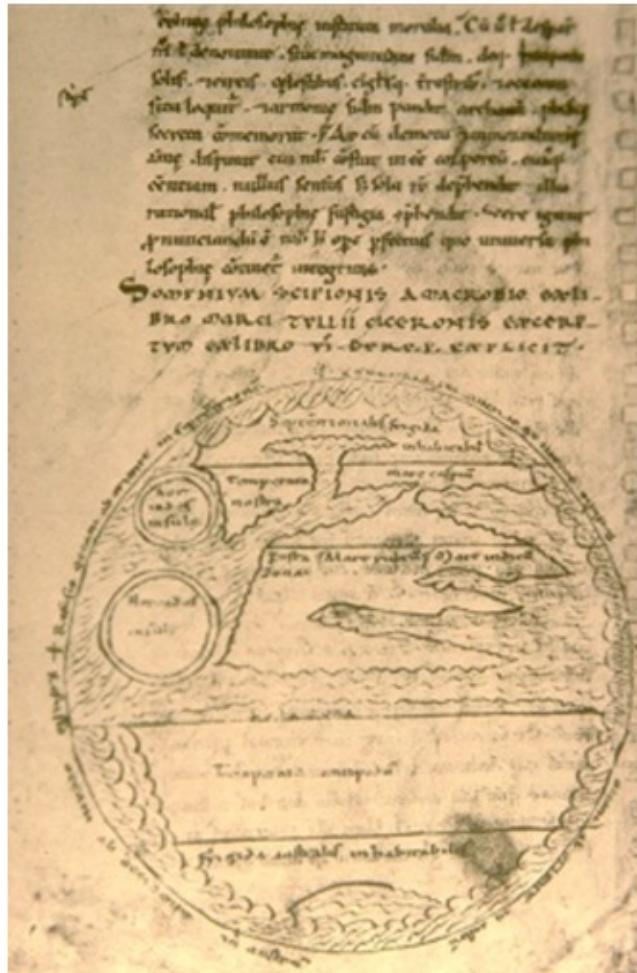
中世纪欧洲



T-O型地图

可视化简史

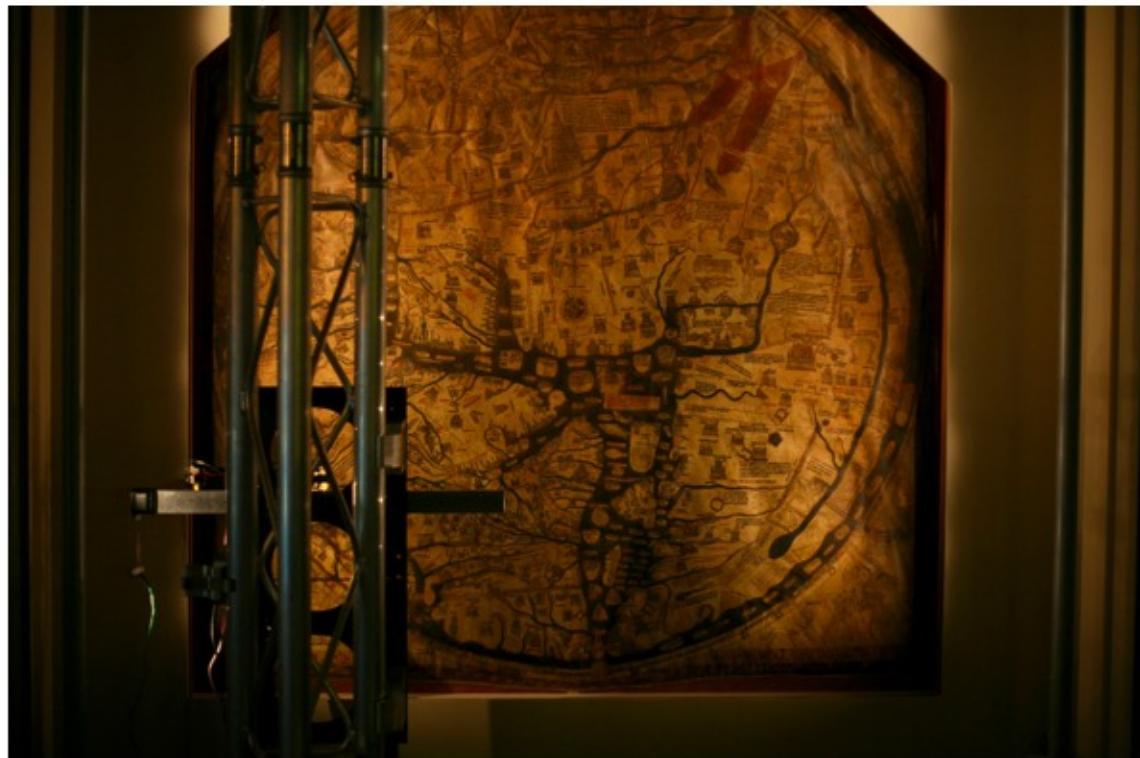
中世纪欧洲



带状地图

可视化简史

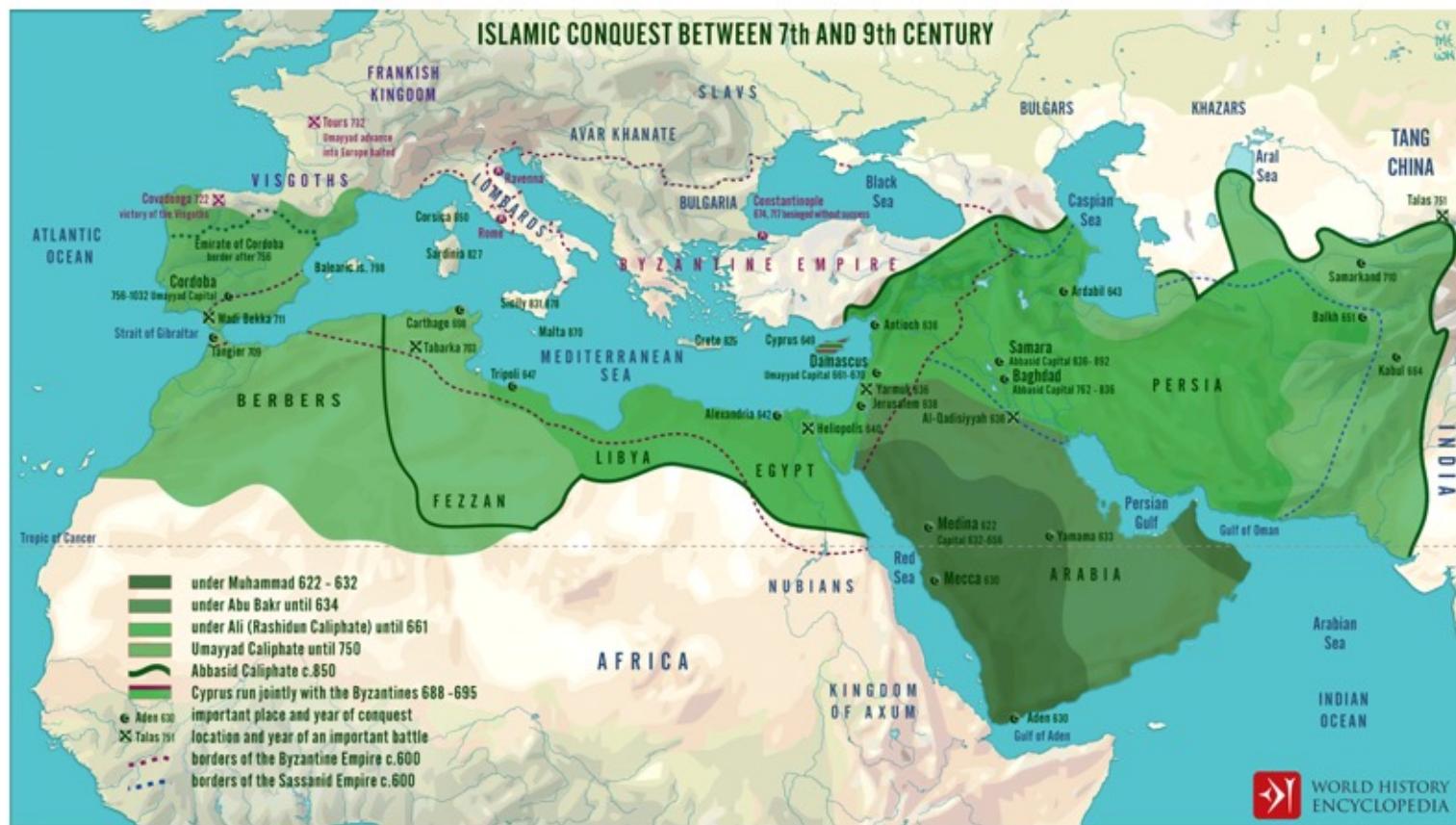
中世纪欧洲



赫里福德地图（上南下北）

可视化简史

中世纪伊斯兰世界



可视化简史

中世纪伊斯兰世界



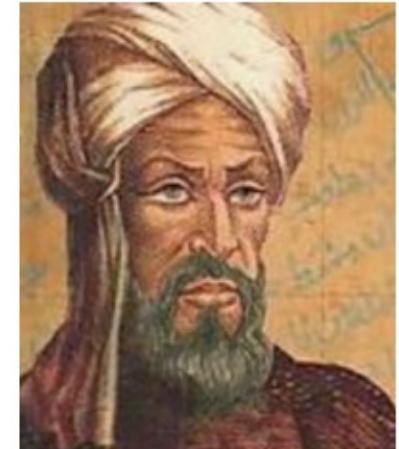
尼罗河地图

المسافة إلى المطر	نحو نوبه
ناموس وحمل	نحو نوبه
مدينون والبلط	نحو نوبه
خرون وآس	نحو نوبه
مسن	نحو نوبه

الخط الرياحي خط الاستواء

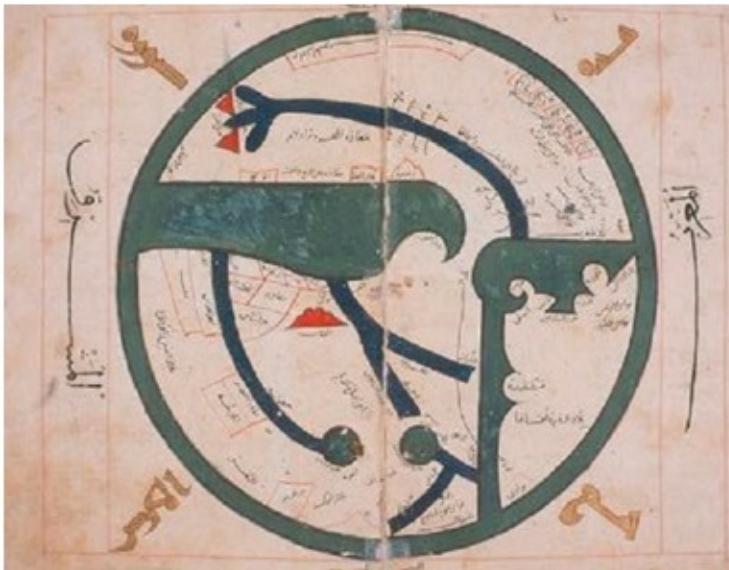
الاسم	الطول	العرض
الوز	نحو نوبه	نحو نوبه
حصار	نحو نوبه	نحو نوبه
بـ جـ	نـ	نـ
جـ	نـ	نـ

花拉子米
地理坐标列表的手稿

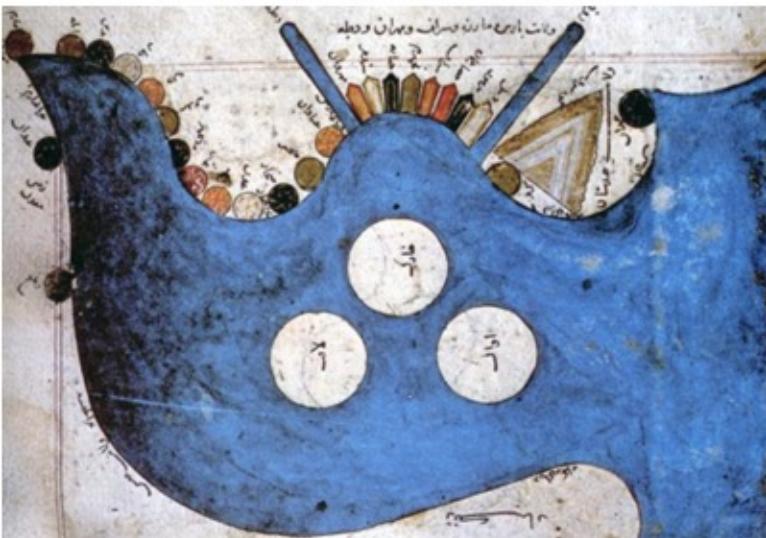


花拉子米

可视化简史



由巴尔希亲自绘制的一幅地图



伊斯塔赫里绘制的波斯湾地图

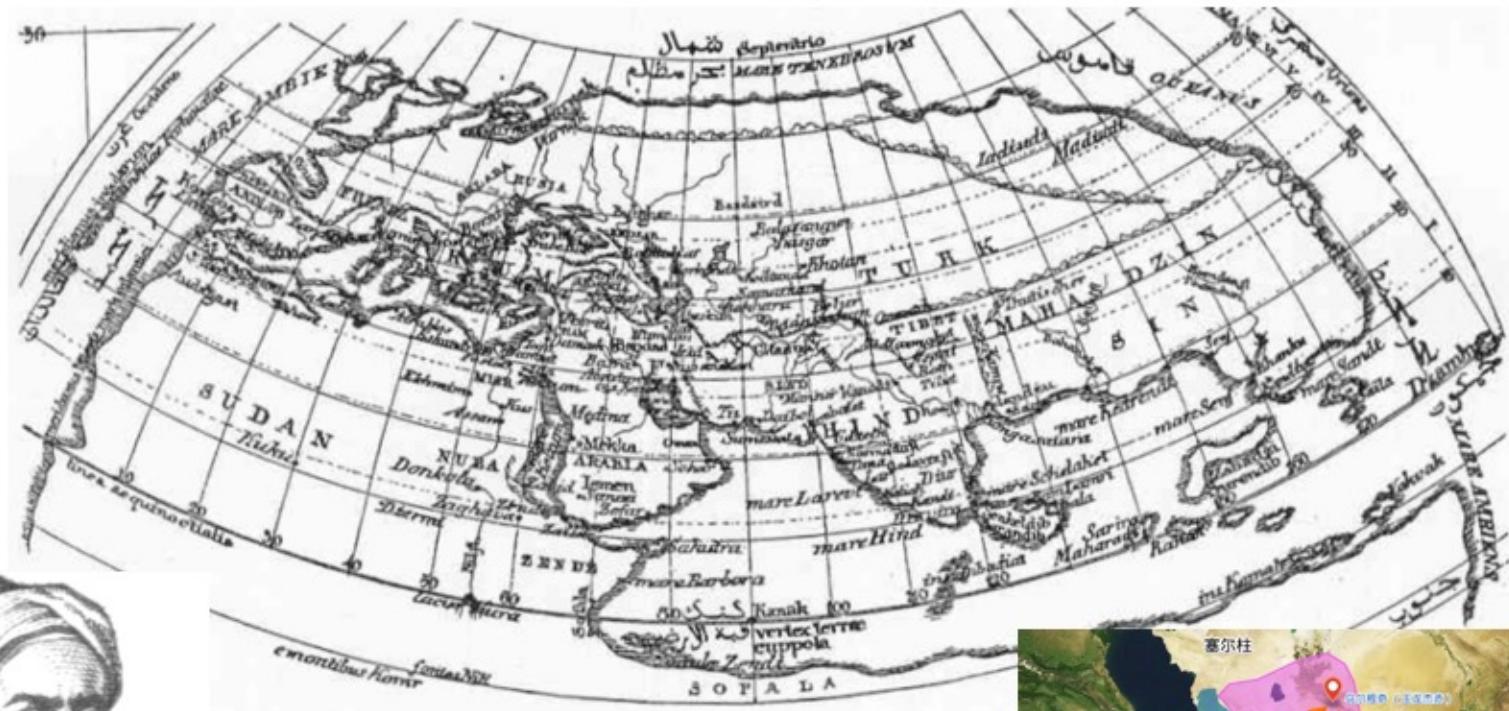


巴尔希学派的著名人物伊斯塔赫里绘制的直布罗陀海峡地图，左南右北

巴尔希地理学派
(Balkhi School of Geographers)

可视化简史

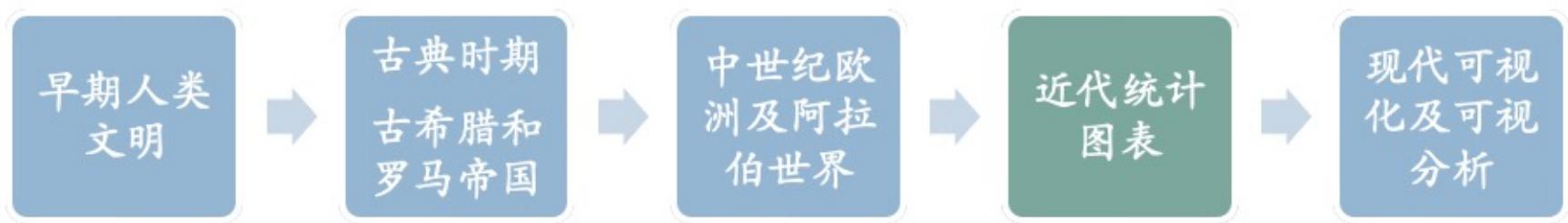
中世纪伊斯兰世界



比鲁尼的世界地图



可视化简史



可视化简史

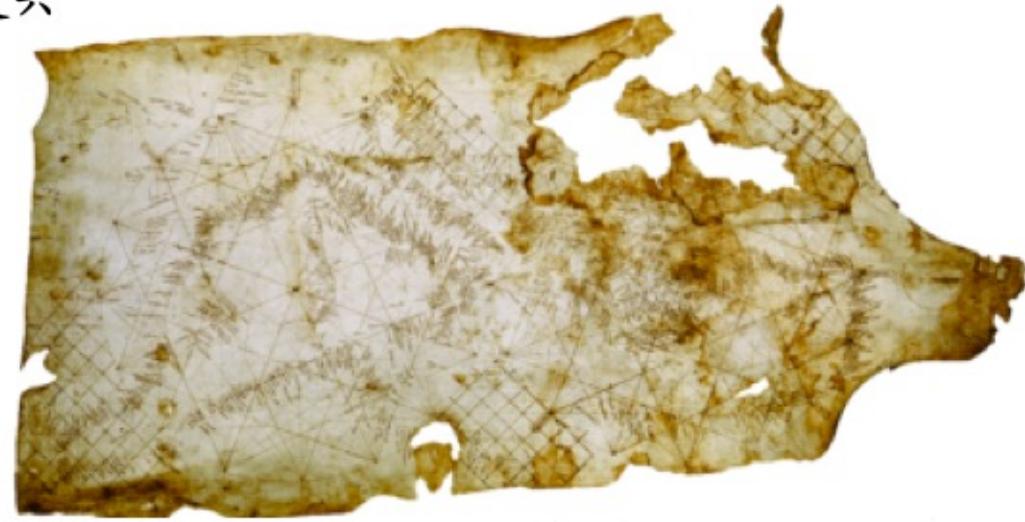
文艺复兴



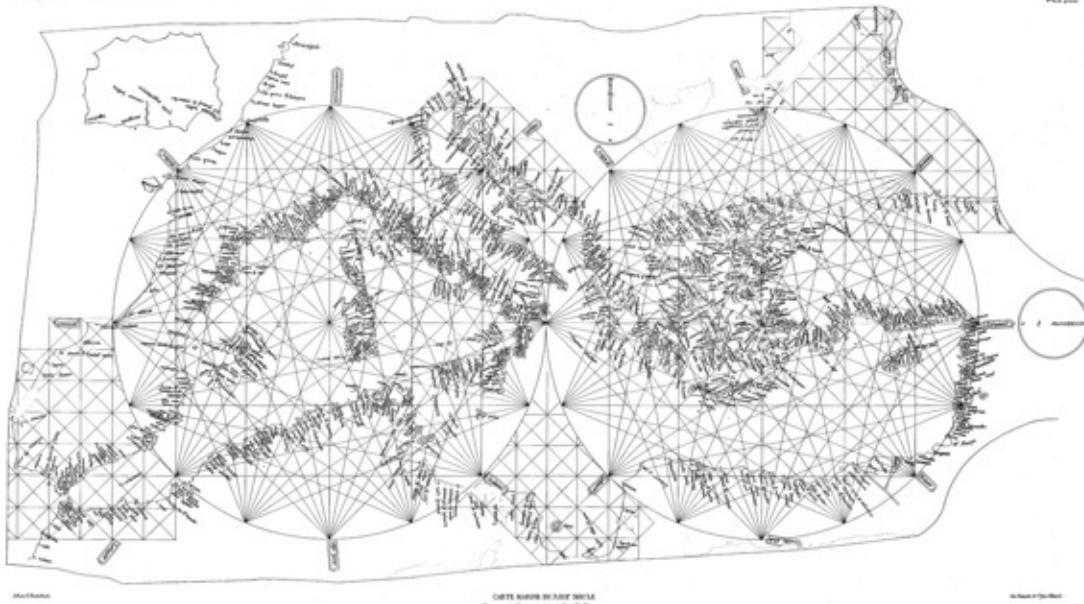
波尔图兰航海图 (portolan charts)

可视化简史

文艺复兴

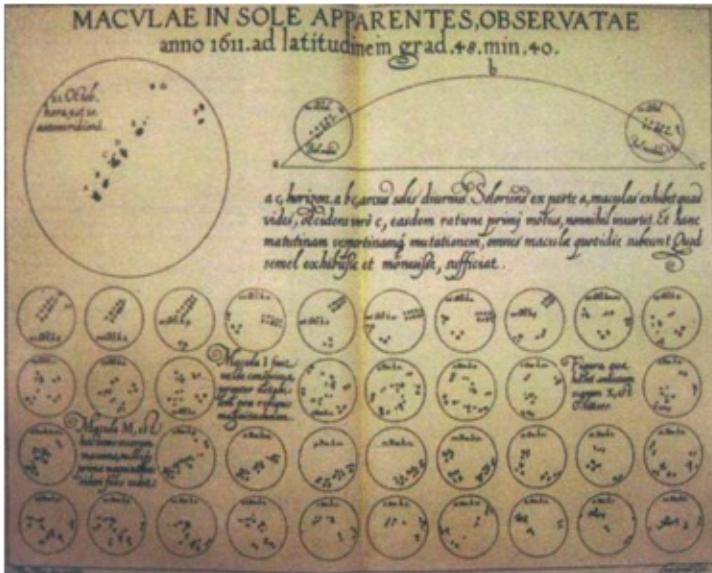


1275年的比萨航海图
(Carta Pisana)

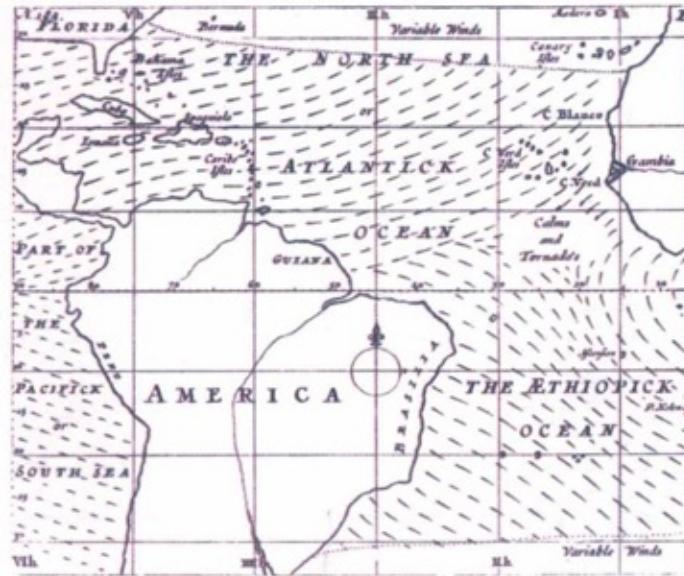


可视化简史

十七世纪



太阳黑子变化示意图 (1626)



最早的天气预报图 (1686)