

# 数据可视化 漫谈

杨剑砺

Datawhale成员，外企数据分析师

2021/04/14

# 大纲

## 01 什么是可视化

从两张150年前的可视化图开始说起

## 02 可视化场景

个性化和通用性场景

## 03 可视化的艺术与技术

从艺术性和技术性两个角度研究可视化

## 04 优秀的案例资源

从模仿中精进

# 01

## 什么是可视化

Data visualization is the **graphical** representation of **information and data**. By using visual elements like charts, graphs, and maps, data visualization tools provide an **accessible way** to see and understand trends, outliers, and patterns in data.

—[tableau.com](https://tableau.com)

# 拿破仑远征图 (绘于1869年)

*Carte Figurative* des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.  
Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite, Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Thiers, de Fozendac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Moliow et ont rejoint vers Orscha et Wilkotsk, avaient toujours marché avec l'armée.

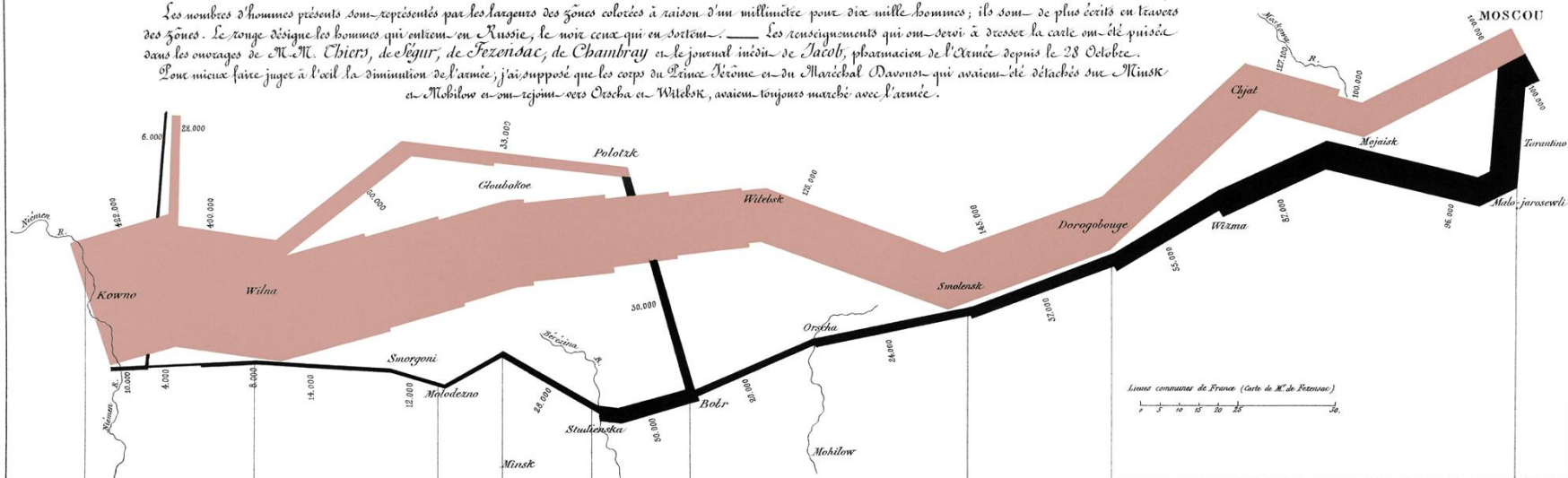
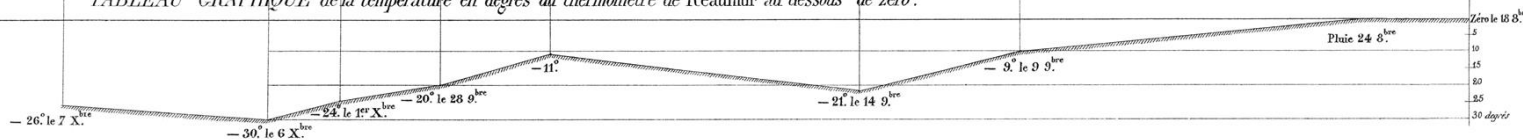
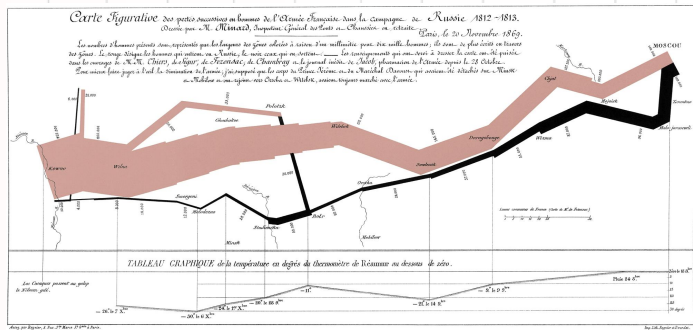


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Niémen gelé.



# 拿破仑远征图拆解



## 两个维度，六类数据

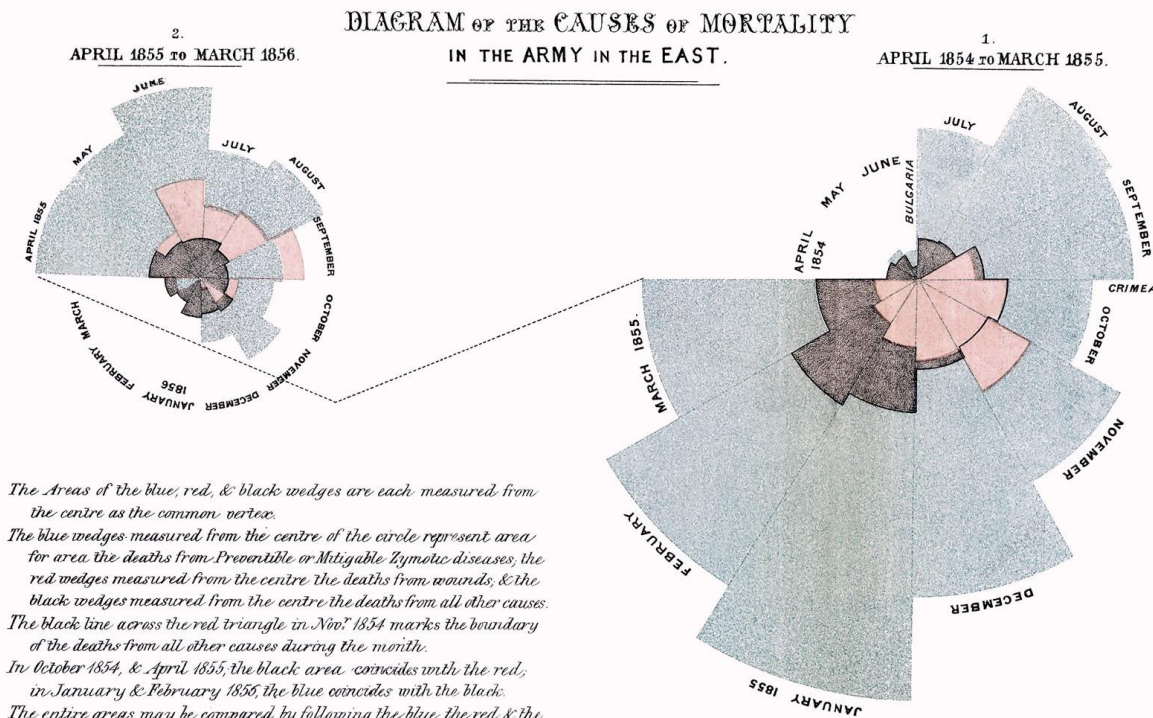
拿破仑军队的数量、行进的  
路程、温度、经纬度、行进  
方向、特定日期的位置

## 信息

军队的消减基本上都是在行  
军的过程中产生的，而不是  
由大的战役造成的损失

一图胜千言

# 另一个例子：南丁格尔玫瑰图的由来



死亡原因  
分布数据

信息

蓝色表示死于  
可预防的疾病、  
红色表示死于  
战争伤害、黑  
色表示死于其  
它原因

真正影响战争  
伤亡的并非战  
争本身，而是  
由于军队缺乏  
有效的医疗护  
理



1

好的可视化图要做到一图胜千言

2

可视化图表是为业务服务的

- ✓ Minard利用他的工程师的成就和绘制可视化图形的能力影响了1850年来法国的公用事业建设的计划编制
- ✓ 南丁格尔在得出战争伤亡结论后推动1857年英国皇家陆军卫生委员会和军医学校的成立



# 02

## 可视化场景

# 常见的可视化场景

A

自助式探索分析

B

专题数据分析报告

C

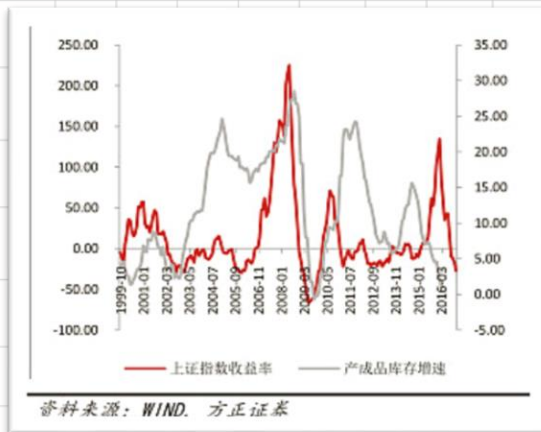
BI可视化平台

# A 自助式探索分析

- 数据的探索与发现
- 模型的尝试与构建

## 宏观经济研究

产成品库存增速的阶段  
库存周期和股指的滞后关系



## 机器学习建模

在机器学习建模的多个过程  
都依赖可视化

数据获取

特征工程

模型评估

数据清洗

算法训练

模型展示

# B 专题数据分析报告

背景介绍

分析目的

分析思路

分析内容

结论与建议



# C BI可视化平台

需求讨论

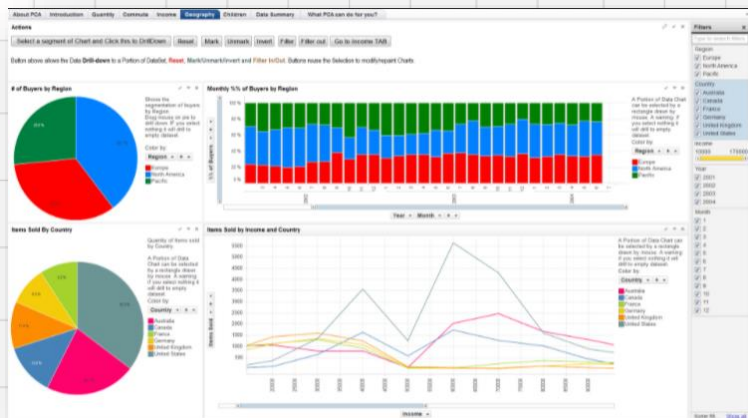
数据建模

ETL开发

BI开发

项目交付

常规界面



三个趋势

- 敏捷开发
- BI内化到业务部门
- 向智能化迈进

规范性分析  
(prescriptive)

预测性分析  
(predictive)

诊断性分析  
(diagnostic)

描述性分析  
(descriptive)

# 03

## 可视化的艺术与技术

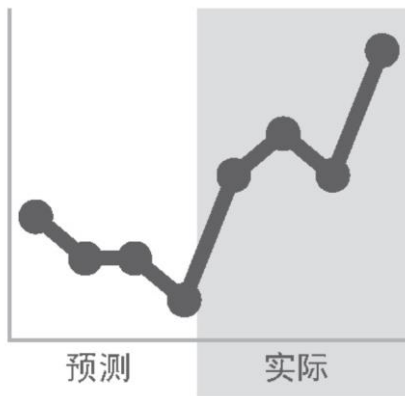
# 可视化的艺术性



视觉认知的格式塔原则

相似/临近/**包围**  
闭合/连续/连接

包围原则的使用场景  
之一是为数据添加一个视觉上的区分，  
我们会认为物理上包围在一起的物体从属于同一个群体

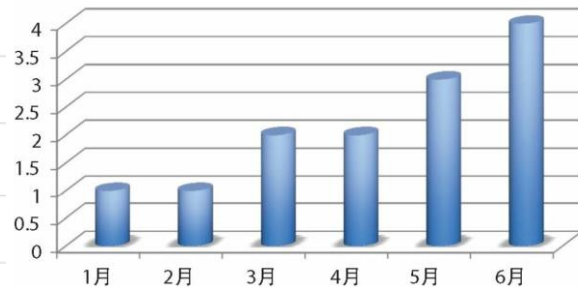


设计

选择

数据可视化黄金定律

永远不要用3D图形





# 可视化的艺术性



引导受众视觉沟通  
前注意属性

视线

颜色

少用颜色  
一致性  
考虑色盲患者  
考虑跨国文化



方向



形状



线条长度



线条宽度



大小



曲率



添加标记



包围



色调



强度



空间位置



动作

国家级医药销量前5位

图中用颜色分布表示销量在指定国家的排名，范围是1（红色）到10及以上（紫色）。

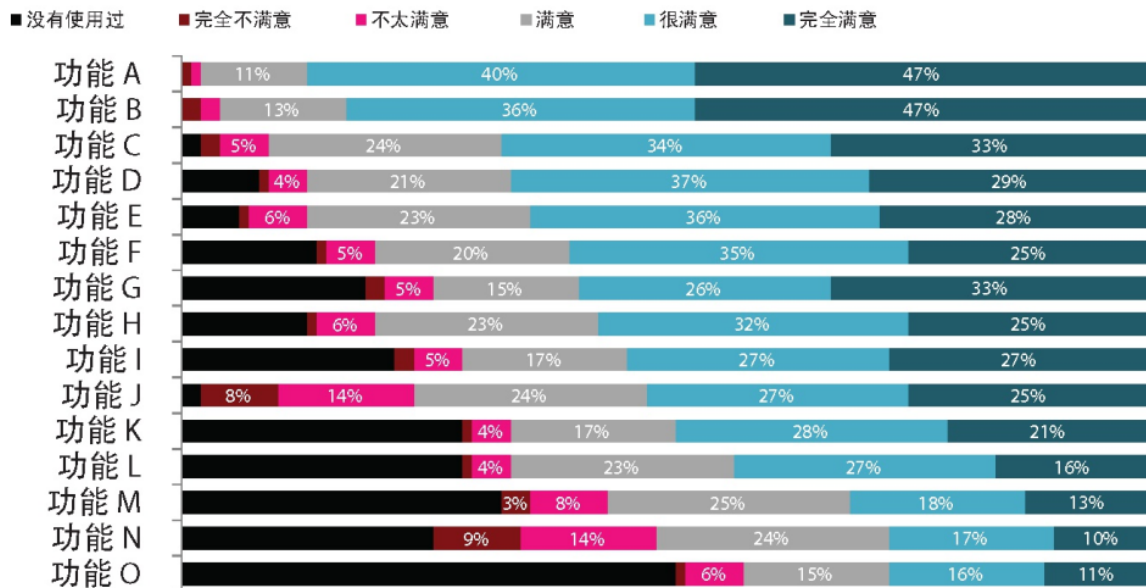
Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5

药品前5名：国家级销量排名

排名	1	2	3	4	5+
国家   药品	A	B	C	D	E
Australia	1	2	3	6	7
Brazil	1	3	4	5	6
Canada	2	3	6	12	8
China	1	2	8	4	7
France	3	2	4	8	10
Germany	3	1	6	5	4
India	4	1	8	10	5
Italy	2	4	10	9	8
Mexico	1	5	4	6	3
Russia	4	3	7	9	12
Spain	2	3	4	5	11
Turkey	7	2	3	4	8
United Kingdom	1	2	3	6	7
United States	1	2	4	3	5

# 可视化的艺术性

你对这每一项功能的满意度如何？



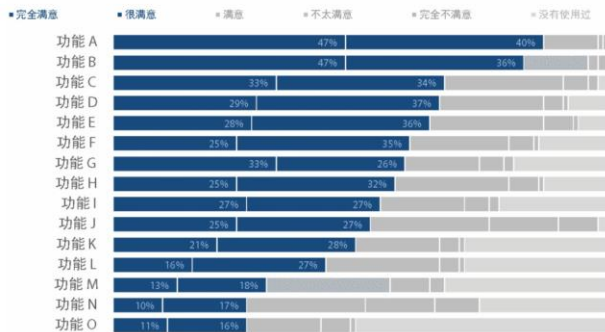
原图

# 可视化的艺术性

## 只突出某一个重点

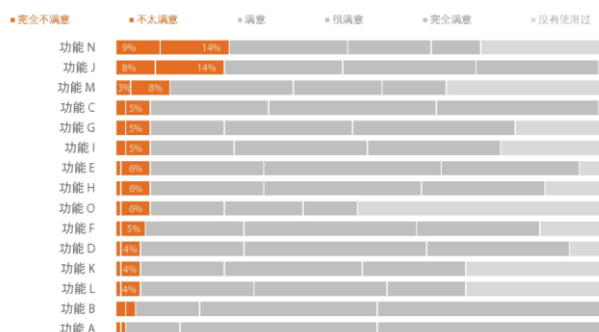
功能A和功能B获得了最高的用户满意度

产品X的用户满意度：按功能分



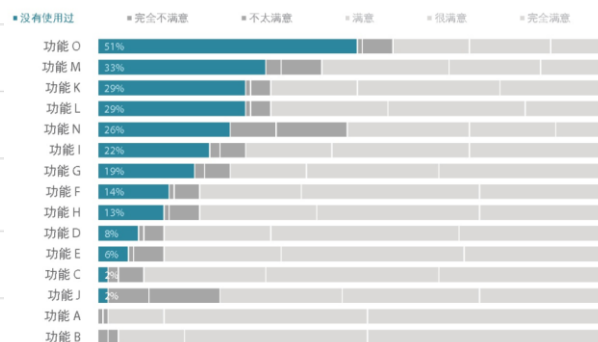
用户对功能N和功能J最不满意

产品X的用户满意度：按功能分



功能O被使用得最少

产品X的用户满意度：按功能分

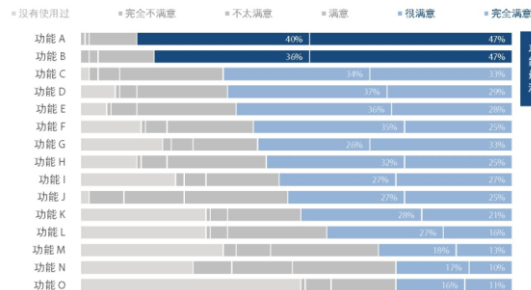


# 可视化的艺术性

## 突出三个重点

用户满意度在不同功能之间差异很大

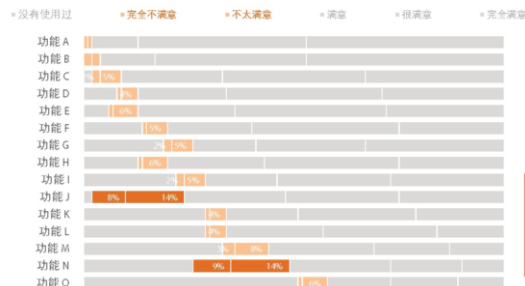
产品X的用户满意度：按功能分



功能A和功能B获得了最高的用户满意度

用户满意度在不同功能之间差异很大

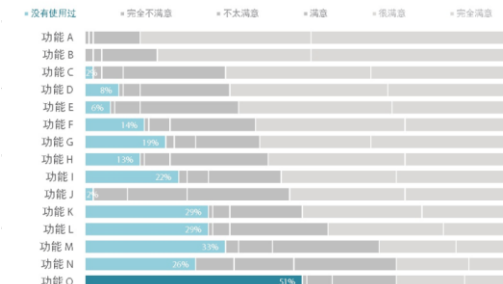
产品X的用户满意度：按功能分



用户对功能N和功能J最不满意，我们能作出何种改进以提升用户体验呢？

用户满意度在不同功能之间差异很大

产品X的用户满意度：按功能分



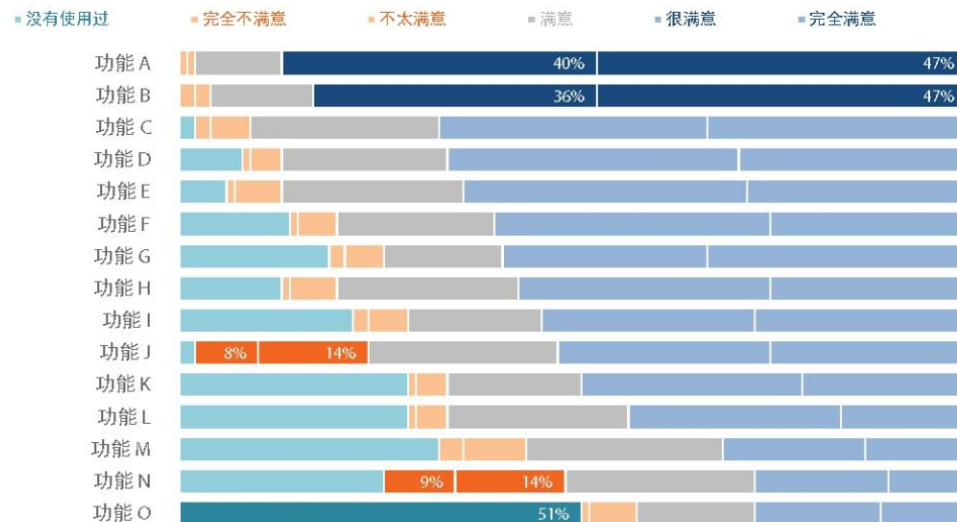
功能O被使用得最少，我们可以在现有用户群中采用何种措施以提升使用率呢？

# 可视化的艺术性

## 浓缩成一张图的报告

### 用户满意度在不同功能之间差异很大

产品X的用户满意度：按功能分

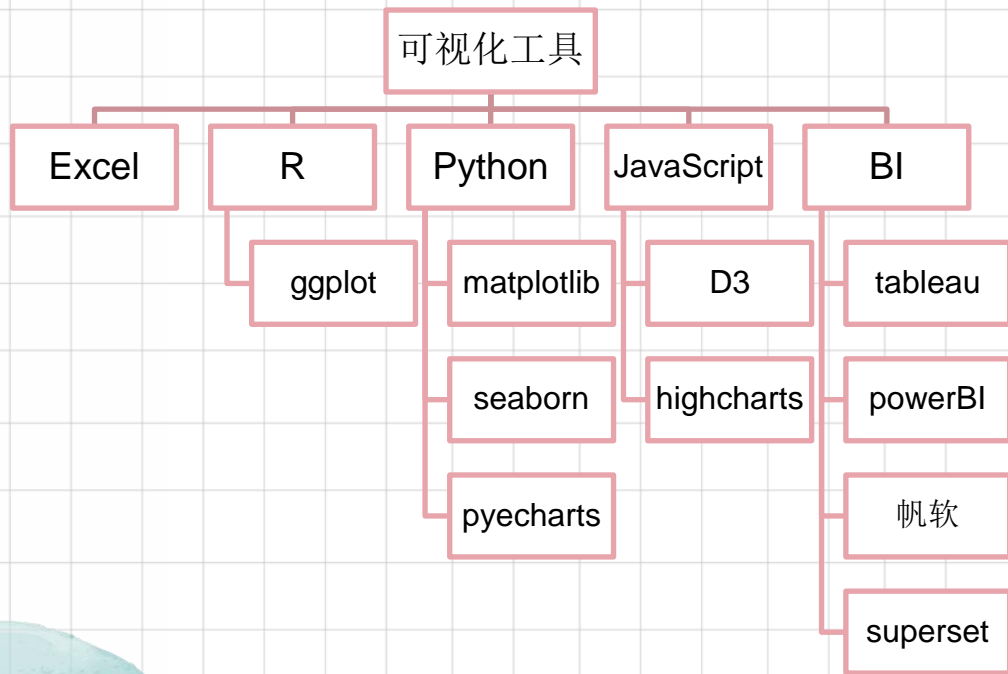


功能A和功能B获得了最高的用户满意度

用户对功能N和功能J最不满意，我们能作出何种改进以提升用户体验呢？

功能O被使用得最少。我们可以在现有用户群中采取何种措施以提升使用率呢？

# 可视化的技术性



# 04

## 优秀的案例资源



# 可视化资源

可视化教学

<https://www.data-to-viz.com>

优秀可视化案例

<https://flowingdata.com>

在线制图

<https://flourish.studio>

<http://tuzhidian.com>

<https://www.visualcapitalist.com>

<https://rawgraphs.io>

<https://ourworldindata.org>



**Q&A**