

企业数据可视化从0到1

EXPLORATORY DATA ANALYSIS

演讲人：程凯征 - DataHunter

跨界互联
数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT

北京 中国大饭店 2017.07

The chalkboard is densely packed with hand-drawn content in white chalk. At the top left, there is a large equation:
$$€^2 \cdot \left(\frac{1271}{31} \right) + 44 \left(\frac{01231}{1888} \right) \div \$2\sqrt{H^0}$$
 Below this, there are several smaller diagrams and equations. On the left, a bar chart with three bars of increasing height is drawn. Next to it is a network diagram with nodes and connecting lines. In the center, a coordinate system shows a circle and a line intersecting. To the right of the center, there are more equations:
$$319211$$
$$72 \begin{matrix} 1132 \\ 123012 \end{matrix} \begin{matrix} (1) \\ //0 \end{matrix} p_0 \div \mathcal{L}_0$$
 Below these, another equation is written:
$$€^2 (2/3 + 10) H \quad | \quad 11$$
$$72 (00 | 0 \frac{1}{2}) 342$$
 Further down, there are several dollar signs (\$\$\$\$) and a series of upward-pointing arrows. At the bottom left, there is a diagram with a house icon and the text "HP". In the bottom right, there are more equations and a small diagram with a circle and a line. The woman stands in the foreground, her hands clasped behind her back, looking at the board.

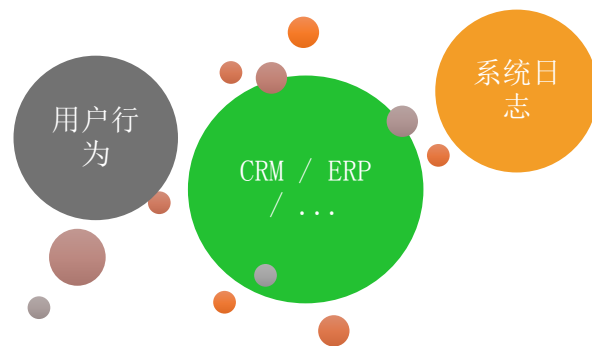
“表哥” / “表妹” 的一天

- 数据采集难
 - 人工填报，录入，系统分散，到处要数据
- 数据处理难
 - 数据质量低，滞后，要做大量的ETL工作
- 重复劳动多
 - 每天都在进行重复劳动，没时间真正的分析
- 分析结果分享麻烦
 - 除了报表，还要做PPT，各种版本分享难



◉ IT 人员

- 各个系统独立，数据孤岛严重
- 业务需求变化快，IT 资源瓶颈较大



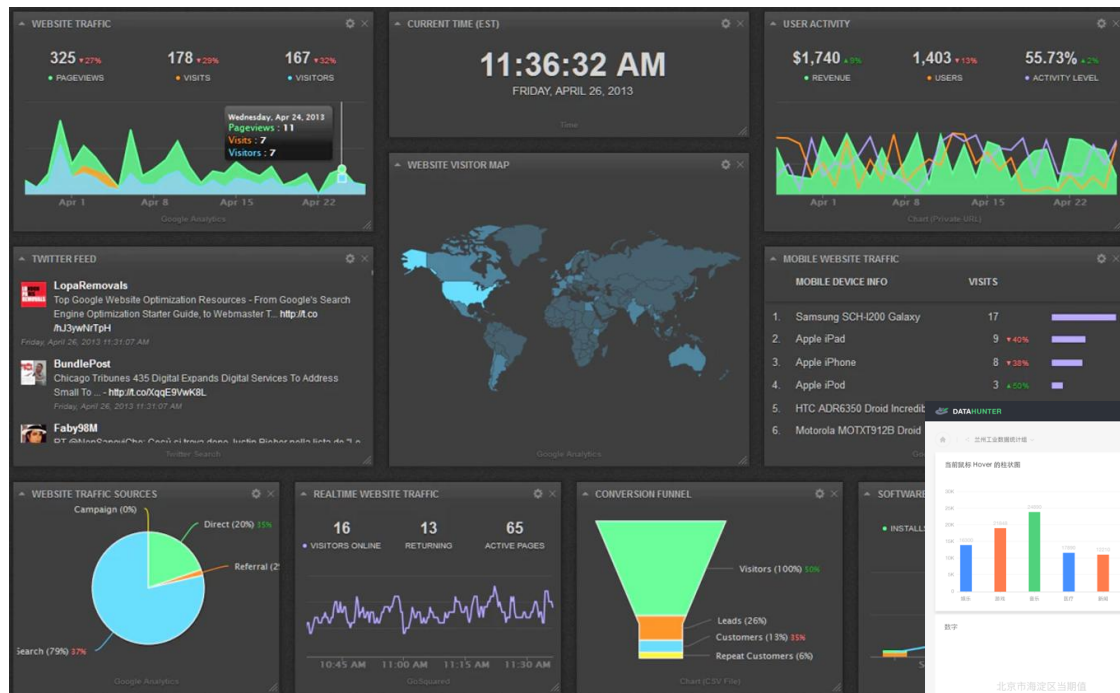
◉ 数据人员

- 重复性劳动大于分析工作
- 手工统计准确性差，时效性差

◉ 业务人员

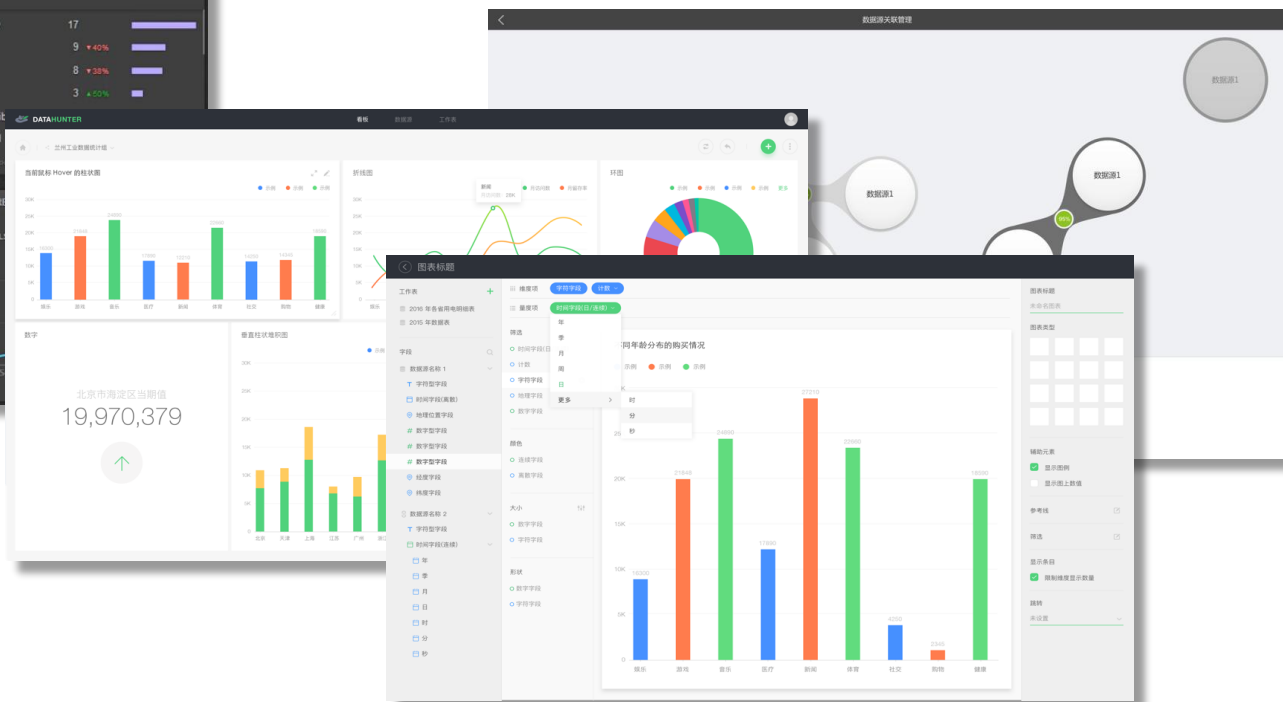
- 业务报表获取困难，决策缺乏依据
- 无法进行业务预测 / 预警

国外的企业数据可视化



实时业务看板

Real Time Dashboard



探索式商业智能

3rd Generation BI

什么是数据探索式分析

◉ 验证式分析

- 先有想法，再通过数据进行验证
- 目前大部分企业内数据需求属于此类

◉ 探索式分析

- 没有目标，通过数据和图形发现问题
- 走在业务之前，变被动为主动

固定报表

验证式分析

动态报表

探索式分析

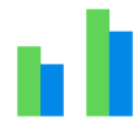


企业数据可视化阶段

探索式分析 – 理论基础

数据类型 + 数据之间的关系 + 正确的图表选择 + 合理的视觉呈现
= 数据可视化分析（探索式分析）

0100101010
1010001001
0110011101
1101010110

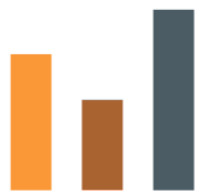


探索式分析 - 数据的类型

- 分类数据
 - 可以进行分组或排序，通常都是文字类型（可以分为有序和无序，均为离散数据）
- 量化数据
 - 可以测量，所有的值都是数字（可以是连续数据或离散数据）
- 时间数据
 - 以时间作为数据内容（既可以是连续数据，又可以是离散数据）
- 地理数据
 - 用作地理位置的标示（地名 / 经纬度信息，属于离散数据）



探索式分析 - 数据之间的关系



简单对比

对分类的量化数据进行简单的对比，从而更直观的了解两者的量化对比情况，通常用来发现问题。



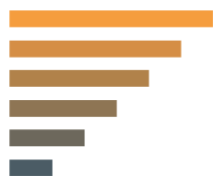
时间序列

显示同一维度下数值随时间的变化，通常可以帮助人们发现趋势，进行预测。



相关性

同一维度下两个数值的变化关系对比，从而发现正相关性或负相关性，通常用于因果关系的发现。



分级排序

两个以上的数值互相之间的关系，通常用于排序分级，从而查看顺序和数量。



偏差性

通过观察数据点之间的关系，发现一些特殊的，与普通数据有明显不同的情况，用于观察数据偏离度。



分布情况

描述数据围绕核心数值的一个分布情况，用于观察数据的分布方式和情况。



局部与整体

用于对比部分数据和整体的关系，比如：小学生在“王者农药”玩家群里面的分布比例。

** 还有更多，七种武器仅供参考 **

探索式分析 - 视觉处理依据

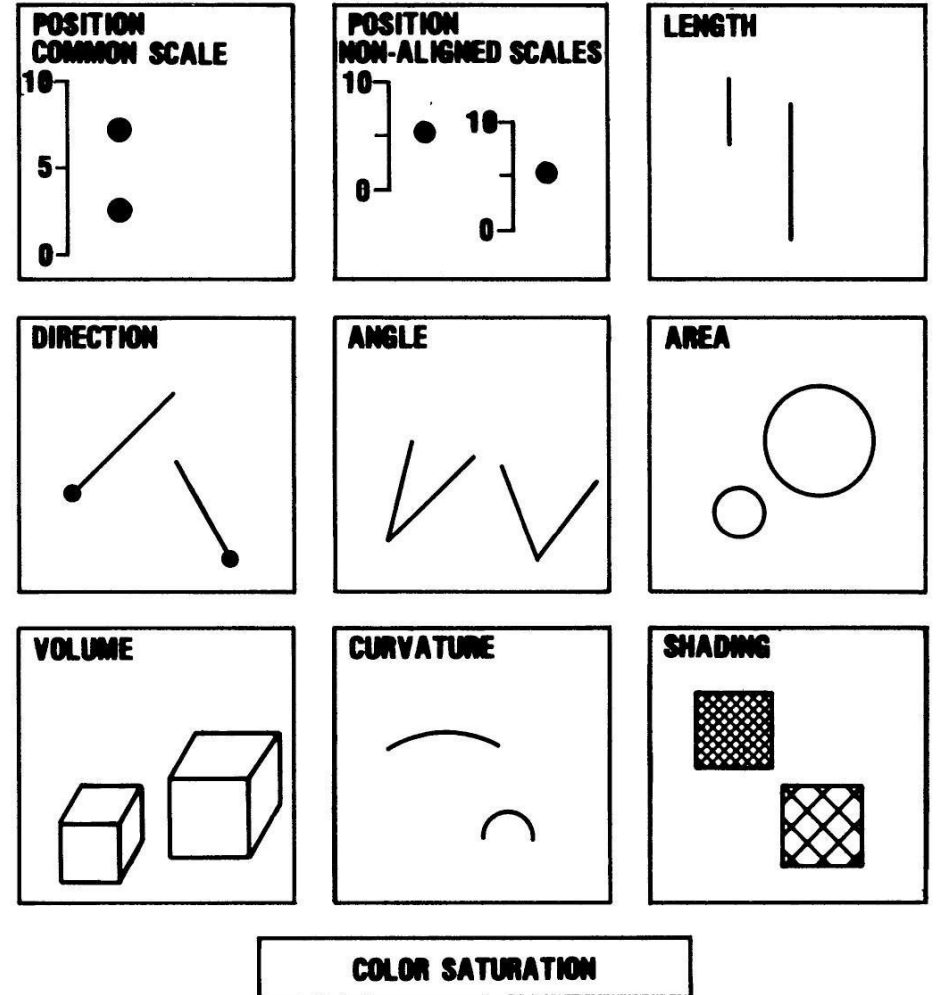
刻度定位 > 长度 > 方向 > 角度

> 面积 > 体积 > 曲率

> 阴影 > 颜色 > 饱和度

**** 不要迷信3D图形****

**** 小心使用颜色****



探索式分析 - 选择图表

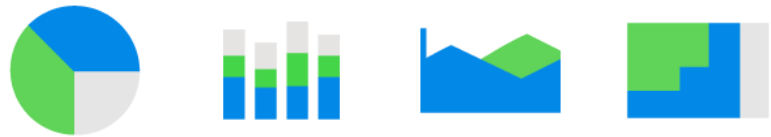
- 是否需要进行比较 / 对比？



- 是否需要了解数据的趋势？



- 是否需要查看数据构成（组成）？



- 是否需要分析数据之间的关系？



- 是否需要理解数据分布情况？



** 五个问题，确保图表使用正确**

新一代企业数据可视化工具

- Gartner 2017 BI (商业智能) 魔力象限报告：传统的BI厂商已从领导者象限出局，自助探索式分析将成为趋势



自助式数据处理能力

探索式分析能力

数据展示分享能力

团队协作沟通能力

异构数据源整合

整合各业务系统数据

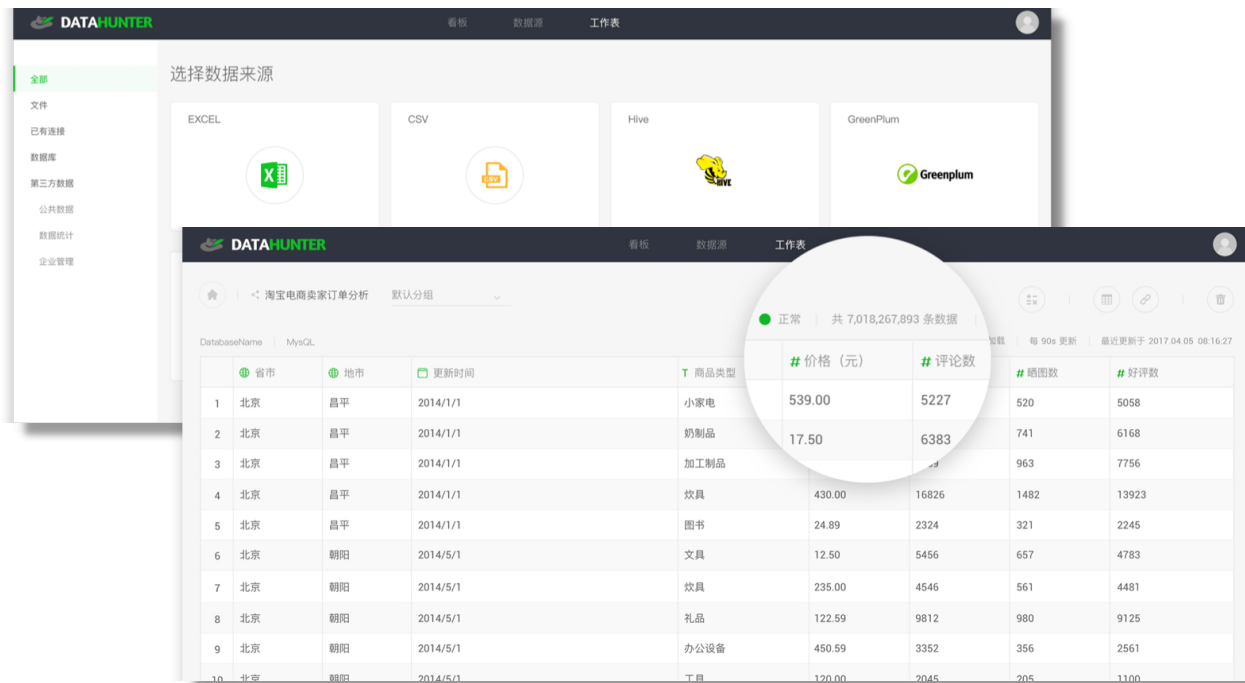
- CRM / ERP / 其他系统
- 支持Excel、CSV等数据文件

支持导入互联网数据

- 公共数据市场（天气、汇率等）
- 第三方互联网服务（微博等）

支持实时数据更新

- 实时业务看板成为可能



探索式分析能力

- 根据数据智能推荐图形
 - 根据数据特点，推荐可视化分析方式
 - 可以在可视化图形间任意切换

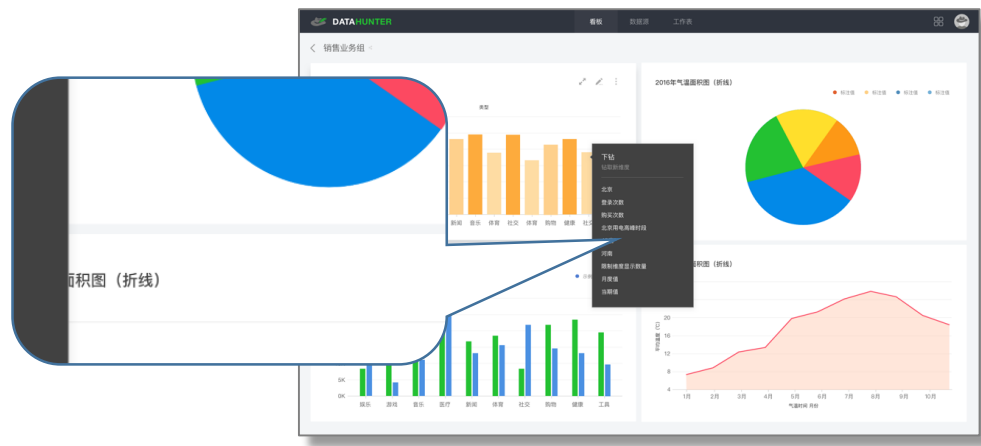
- 动态数据建模

- 无需预先建模
- 无需预计算



交互式数据展示 (Dashboard)

- 数据联动，协同过滤
 - 看板上的图表可以进行协同过滤
 - 方便数据查看分析



- 数据钻取、排序、筛选
 - 将一定的分析能力交给用户
 - 让分析随心而动



分享、讨论、沟通

跨屏幕数据分享

- 兼容PC、平板和手机
- 支持PPT形式内容分享

讨论 + 沟通

- 针对数据看板进行讨论
- 随时随地进行沟通



大屏展示投放

- 支持TV电视无缝投影
 - 直接将Dashboard内容投影至TV
 - 支持触摸屏互动



- 支持超大屏幕实时数据展示
 - 专业设计
 - 实时数据更新展示



关于我们

- ◉ DH 数据分析系统 (DH Data Analytics)

- 多数据源汇聚, 轻松数据建模
- 探索式, 交互式的数据分析模式

- ◉ DH 数据大屏幕 (DH Data Screen)

- 用于业务指挥中心, 会议室, 展厅, 展会等场景
- 整合多种业务数据, 用最合适的方式进行实时动态的可视化

- ◉ 大数据技术服务 (Data Service)

- 大数据系统咨询规划以及建设
- 数据清洗抽取 (ETL)





CDA 数据分析师
www.cda.cn

THANKS

跨界互联 数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT