

《七天数据可视化之旅》第一天 数据可视化过程

原创：Destiny 木东居士 8月5日

《七天数据可视化之旅》第一天 数据可视化过程

Destiny, 某物流公司数据产品经理, 目前从事数据平台搭建和可视化相关的工作。持续学习中, 期望与大家多多交流数据相关的技术和实际应用, 共同成长。

0x00 前言

之前写过一篇文章, 叫做数据可视化的基本流程, 是站在可视化过程中的上下游关系, 来梳理和介绍数据可视化的全过程, 及其各个环节的主要工作内容。今天这篇文章, 也是介绍数据可视化的过程, 但是更多的是站在产品经理的视角, 去呈现一个数据需求, 从需求提炼、分析, 到最终可视化呈现的一个过程。

可视化的终极目标是洞悉蕴含在数据中的现象和规律, 从而帮助用户高效而准确的进行决策。

0x01 可视化过程

一个完整的数据可视化过程, 主要包括以下4个步骤:

- 确定数据可视化的主题;
- 提炼可视化主题的数据;
- 根据数据关系确定图表;
- 进行可视化布局及设计;



0x02 确定数据主题

可视化过程的第一步：确定数据主题，即确定需要可视化的数据是围绕什么主题或者目的来组织的。

在可视化过程中的4个步骤之中，第一步是相对来说较容易的一步。

业务运营中的具体场景和遇到的实际问题，公司层面的某个战略意图，都是确定数据可视化主题的来源和依据。简而言之，一个具体问题或某项业务、战略目标的提出，其实就可以对应一个数据可视化的主题。

比如，银行分析不同城市用户的储蓄率、储蓄金额，电商平台进行双十一的实时交易情况的大屏直播，物流公司分析包裹的流向、承运量和运输时效，向政府机构或投资人展示公司的经营现状等，都可以确定相应的数据主题。

0x03 提炼数据

确定数据围绕什么主题进行组织之后，我们接下来要了解我们拥有哪些数据，如何来组织数据，这里面又衍生出另外三个问题。

1. 确定数据指标

分析和评估一项业务的经营现状通常有不同的角度，这也就意味着会存在不同的衡量指标。同样一个业务问题或数据，因为思考视角和组织方式的不同，会得出截然不同的数

据分析结果。

例如，要评估寄件这项业务，有人想了解寄件量，有人想知道不同快递公司的运输时效，有人想知道寄件用户的下单渠道，还有人想了解寄件收入。

拿起数据，就开始画图，会让整个数据可视化作品没有重点、杂乱无章，是一种用战术上的勤劳掩盖战略上的懒惰，最终的呈现效果一般不理想。

2. 明确数据间的相互关系

基于不同的分析目的，所关注的数据之间的相互关系也截然不同，这一步实质上是在进行数据指标的维度选择。

例如，都要统计寄件量，有人希望知道各个快递公司的寄件量是多少，有人想了解一天内的寄件量高峰位于哪个时段，还有人想知道寄件量TOP10的城市排名。这里的快递公司、时段、城市，都是观察寄件量这个指标的不同维度。

通常，数据之间的相互关系包含如下几类：

- 趋势型：
通常研究的是某一变量随另一变量的变化趋势，常见的有时间序列数据的可视化。
- 对比型：
对比两组或者两组以上的数据，通常用于分类数据的对比。
- 比例型：
数据总体和各个构成部分之间的比例关系。
- 分布型：
展现一组数据的分布情况，如描述性统计中的集中趋势、离散程度、偏态与峰度等。
- 区间型：
显示同一维度上值的不同分区差异，常用来表示进度情况。
- 关联型：
用于直观表示不同数据之间的相互关系，如包含关系、层级关系、分流关系、联结关系等。
- 地理型：
通过数据在地图上的地理位置，来展示数据在不同地理区域上的分布情况，根据空间维度不同，通常分为二维地图和三维地图。

3. 确定用户关注的重点指标

确定了要展示的数据指标和维度之后，就要对这些指标的重要性进行一个重要性排序。

因为对于一个可视化展示的终端设备而言，其屏幕大小有限，且用户的时间有限、注意力也极其容易分散。如何让用户在短时间内，更有效率的获取到重要的信息，这是评估一个可视化产品好坏的重要因素。

在可视化设计之前，不妨问用户两个问题：

（1）如果整个版面只能展示一个最重要的信息，你希望是什么？

（2）你希望展现这些信息的理由是什么？通过用户对这些问题的回答，你能了解到，在已确定的指标和维度中，用户最关注的是哪个或哪些。

通过确定用户关注的重点指标，才能为数据的可视化设计提供依据，从而通过合理的布局和设计，将用户的注意力集中到可视化结果中最重要的区域，提高用户获取重要信息的效率。

0x04 确定图表

数据之间的相互关系，决定了可采用的图表类型。常见的数据关系和图表类型的对应关系如下图所示：



通常情况下，同一种数据关系，对应的图表类型是有多种方式可供选择，是不是随机选择一种方式就可以了呢？

当然不是，图表的目的是为了更好的去呈现数据中的现象和规律，那么必然，可视化图表的效果也极大的受到实际数据的影响，这个后续再分享。

0x05 可视化设计

在做好了以上的需求收集和整理之后，接下来就要开始进入可视化的设计和呈现的阶段。这一步主要包括两个方面：一是进行可视化布局的设计，二是数据图形化的呈现。

1. 页面布局

可视化设计的页面布局，要遵循以下三个原则：

(1) 聚焦

设计者应该通过适当的排版布局，将用户的注意力集中到可视化结果中最重要的区域，从而将重要的数据信息凸显出来，抓住用户的注意力，提升用户信息解读的效率。

(2) 平衡

要合理的利用可视化的设计空间，在确保重要信息位于可视化空间视觉中心的情况下，保证整个页面的不同元素在空间位置上处于平衡，提升设计美感。

(3) 简洁

在可视化整体布局中，要突出重点，避免过于复杂或影响数据呈现效果的冗余元素。

2. 图表制作

影响图表呈现效果的，主要有两个影响因素，一个是数据层面的，一个是非数据层面的。

(1) 数据层面

若数据中存在极端值或过多分类项等，会极大影响可视化的效果呈现，如柱形图中柱形条的高度、气泡图中气泡的大小、饼图中的分类项太多等。

对于数据本身造成的可视化效果不佳的情况，我们是不是就束手无策了呢？当然不是，在以往的可视化过程中，本人虽然也踩了很多坑，但是对于如何解决这类问题也积累了一些经验，下次专题分享。

(2) 非数据层面

非数据层面，但是影响图表呈现效果的因素，通常在设计过程中就可以解决。

比如图表的背景颜色、网格线的深浅有无、外边框等等，这类元素是辅助用户理解图表的次要元素，但如果不加处理全部放出，视觉上就不够聚焦，干扰到你真正想展示的数据信息。

因此，对于此类非数据层面，但是影响图表视觉呈现的元素，应该尽量隐藏和弱化。

0xFF 总结

最后一句个人经验：作为数据可视化的设计者，你应该在可视化设计之前，全面了解此次数据的分布情况、量级，通常几行sql就可以搞定，这样在进行可视化设计的时候，可以少踩很多坑。

作者更多文章：

[数据可视化的基本流程](#)

[如何设计数据字典](#)