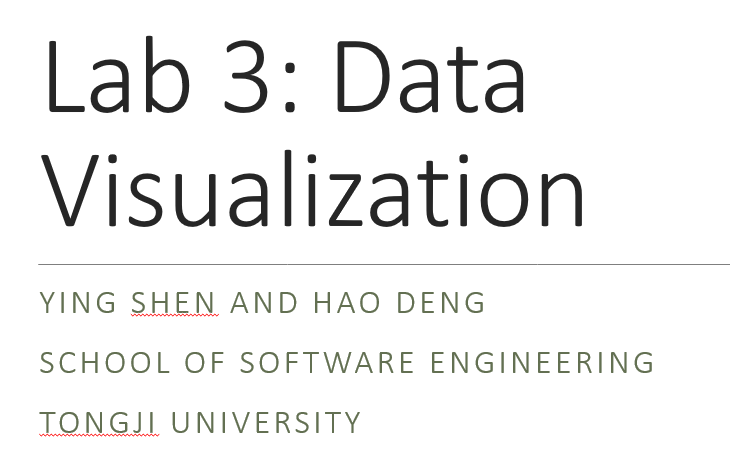
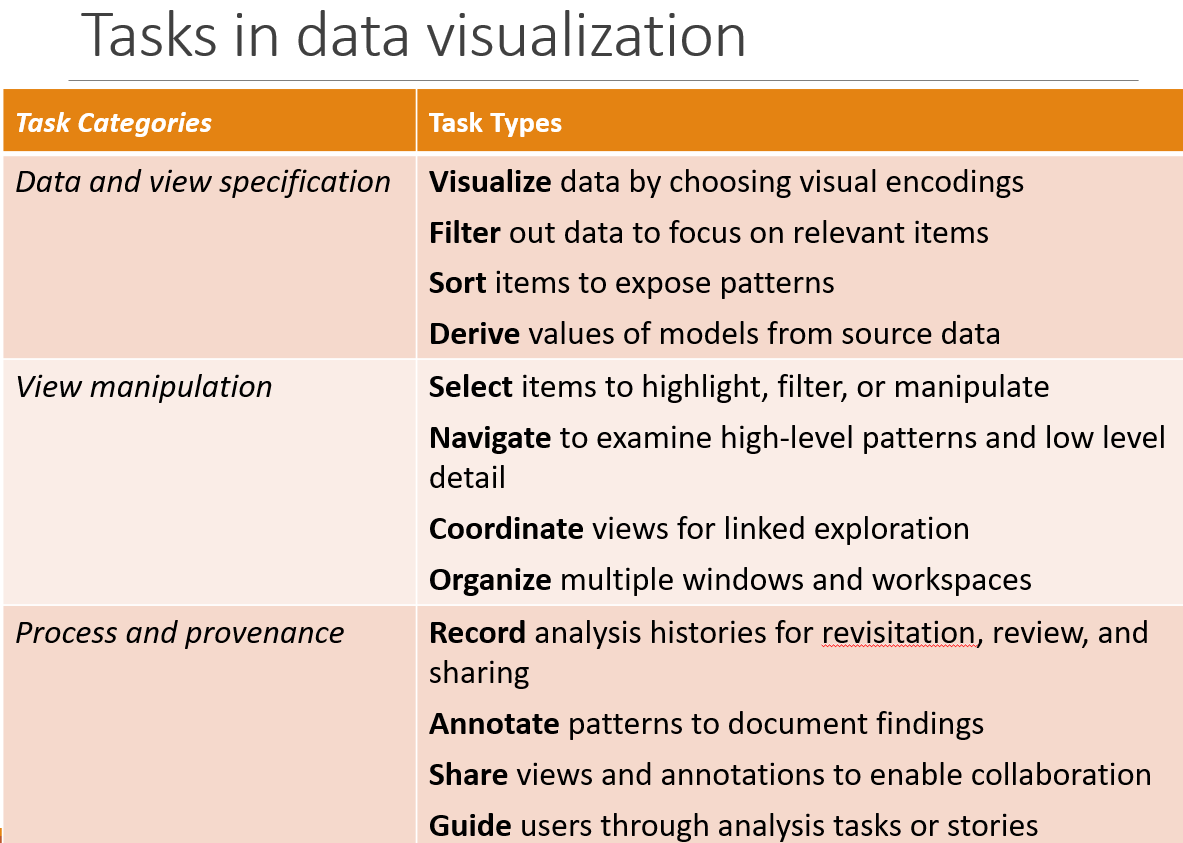
**P1：**



这是我们的第三次个人项目

**P2：**



我们回顾一下施耐德曼团队提出的数据可视化的任务分类，书上也称为面向数据可视化任务的交互式动力学分类。

这种分类方法结合了“用户想要执行的数据分析任务（用户意图）”，以及“用户在可视化工具中的实际操作（系统行为）”两个方面，进而形成了3个高级类别：

1. 数据和视图规范类任务
2. 视图操作类任务
3. 过程和出处类任务

其中，

数据和视图规范这类任务包括4个子任务：可视化、过滤、排序和派生

可视化任务，通过选择视觉编码来呈现数据

过滤任务，通过过滤数据，聚集相关数据条目；

排序任务，通过对数据条目排序，揭示数据中蕴含的模式；

派生任务，由源数据派生出其他信息（比如预算总额）

视图操作这类任务包括：选择、导航、协调和组织4个子任务

选择任务，包括选择需要突出显示、过滤和操作的数据条目

导航任务，通过导航来了解数据的高层次模式和低层次细节

协调任务，是为了协调数据的关联搜索

组织任务，包括组织多个数据窗口和工作区

过程和出处这类任务包括：记录、注释、共享和指导4个子任务

记录任务，包括记录“重览、审核、共享和分析数据”的历史

注释任务，主要提供注释文档的查找模式

共享任务，包括共享视图和注释，促进多用户协作

指导任务，通过分析任务或案例来指导用户

**P3：**



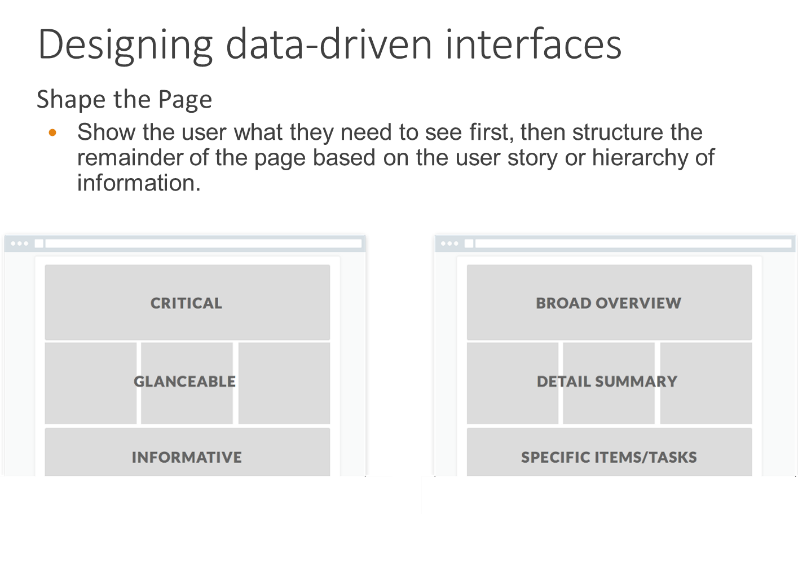
当我们设计复杂系统的时候，不可避免地会有多个用户或角色需要设计。

在设计时，首先需要确定用户和角色，然后围绕它们，组织我们的数据可视化任务。

图中是一个医疗报告应用程序。这个系统有不同的用户，每个人都有自己的数据可视化需求。对于这个应用程序来说：一旦确定了关键用户，它就可以把关键用户和要为用户提供的服务总结出来。

然后给不同的用户提供不同的数据

**P4：**



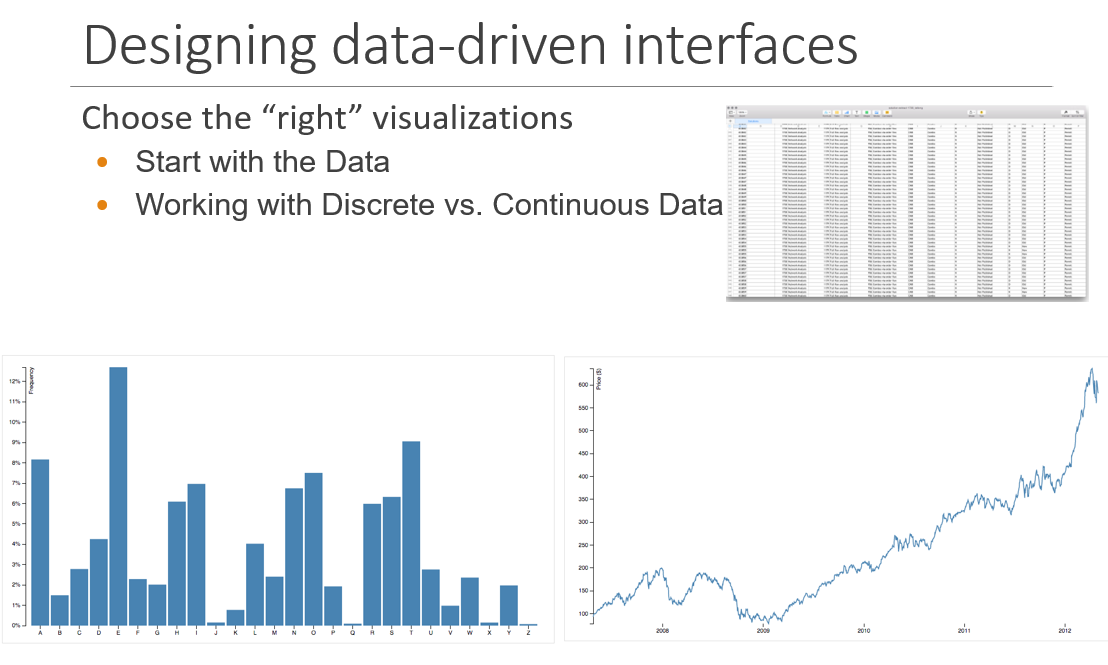
其次，设计时，要通过塑造页面，向用户展示他们需要看到的东西。然后根据用户故事或信息的层次结构来构建页面的其余部分。

图中是塑造页面的两种常用方法。

一种方法是：顶部呈现关键信息，中部给出一目了然的内容，底部提供大量信息

一种方法是：顶部给出概览，中部给出详细概述，底部给出特定的条目或者任务

**P5：**



最后，设计时，要选择正确的可视化方式。

避免滥用图表，以及偏向于美学的设计倾向。

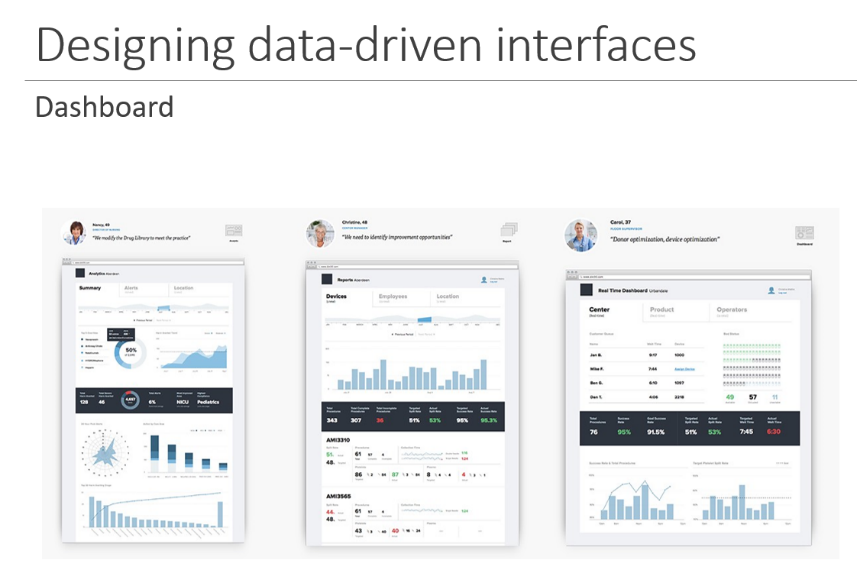
当前的数据可视化设计里，这些坏习惯正在广泛传播。设计出正确的可视化方式，要从数据本身开始（从原始数据思考），通过利用一些不错的工具来给你启发，比如Tableau。Tableau是目前最好的工具之一，不过比较贵。

也要处理好离线数据和连续数据

有些图表比另一些图表更能呈现你的数据。如果你处理的是离散数据，比如左图，用柱状图就是不错选择。如果你处理的是连续数据，比如右图，用线形图效果会比较好。

这只是简单的一个说明。实际选择可视化方式的时候，要比这个问题复杂得多。**我们看看我这边项目组正在做的原型系统：….**

**P6：**



现在，针对我们本次个人项目，我们来看看如何设计数据驱动的界面，尤其是如何设计用于展示数据的仪表盘。

无论哪一种数据产品，仪表盘（Dashboard）都是最核心的功能。它作为用户接触数据的第一个页面，相当于数据产品的门户，担负着提纲挈领，引导分析的重要职能，帮助用户能够快速判断业务情况，支持他们做出决策并行动。

根据行业人员研究过的一些 Dashboard 设计案例和文章，仪表盘的设计可以总结如下：**一个原则，两个时期，三个细则（板书）。**

**一个原则即「提纲挈领，引导分析」**，具体来说，就是在一个页面里要明确告诉用户当前业务状况好坏，并支持能够针对某个问题进行分析，从而串联起整个数据平台。由此原则从而可推出以下**三个细则（板书）**：

1. 因人而异，细分场景：根据用户和场景设计产品
2. 少胜于多，分清主次：展示最关键的指标，区分优先级
3. 深入分析，落地行动：支持深入分析，并落地为具体行动

这就像讲一个用户故事。经过起承转合慢慢铺垫，最终进入正题，告诉用户应该怎么去执行，怎么去优化产品和业务。

**各类数据产品在实现前面提到的不同原则时，又可以分为两个时期：Report（定制化）和 Customize（个性化）**。

定制化是早期数据可视化的实现形态，如 Google Analytics 和 Mixpanel 等数据分析产品，特点是根据指定的分析思路，严谨地从汇总到细分，层层下钻。它们为实现「因人而异」，往往会设定多个 Dashboard 页面，或者本身就只针对某一类用户群体。

**个性化更常见于一些比较新型的分析产品**，如 Domo 和 Looker等数据可视化工具，特点是报表自定义程度较高，通过自由选择单图的方式来组建 Dashboard。随着时间的发展，这两者的界限在慢慢模糊，但主要的特点依旧鲜明。

**P7：**



【1】

数据面板设计原则的第1个细则——因人而异，细分场景，主要说的是：

产品的核心就是解决问题，解决某些用户在某些场景下的某个问题。对于不同行业，不同业务，不同职位的人们来讲，关注的内容自然不一样。在设计一个数据产品或页面时，**我们需要围绕着用户和场景来做设计**。我们看到，图中不同用户都有自己的目标场景。这也就是刚才我们所说的：给不同的用户提供不同的数据

http://www.woshipm.com/data-analysis/691262.html

【1.1】

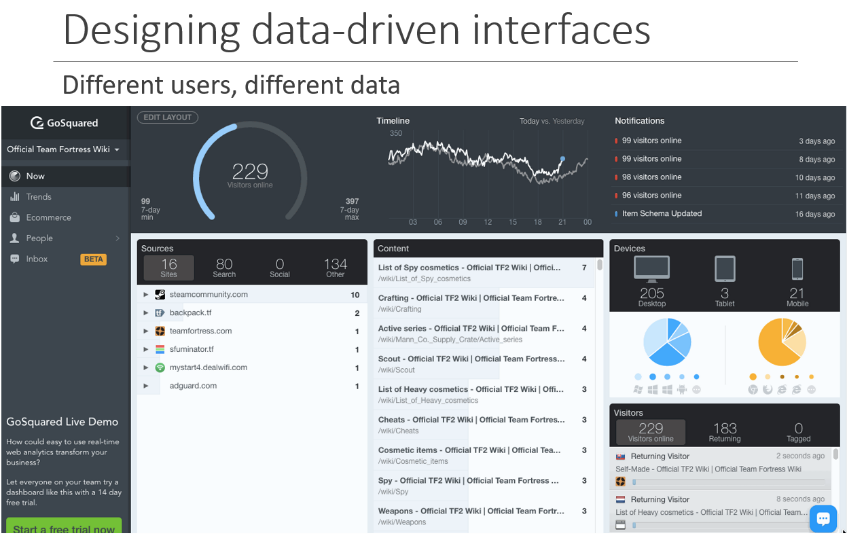
因此，在第1个细则中，划分用户是1个关键点。划分用户，一般会从业务线或岗位入手。不同业务线间，关注的核心指标自然不同，比如各种业务部门，即使大家都关注大盘数据，但每天更关心的，还是自己的业务细节指标。

岗位也是同理，管理层重在把握全局，而执行层重在每个细节的执行效果，关注的数据层级和指标也会有所差异。

以图中这个 Domo 提供的例子为例，它为每个职位的人单独设一个 tab 来显示，每一个岗位都有自己预设的 Dashboard 页面。

所有设计仪表盘的第一件事情，就是明确自己的用户，然后根据不同的用户群体设计仪表盘，以及确定开发的优先级。

**P8：**



【1.2】

划分场景是第二个关键点。要划分的主要场景包括但不局限以下情况：实时监控场景，指定主题分析场景，移动查询场景，周日会汇报场景和大屏显示场景等。

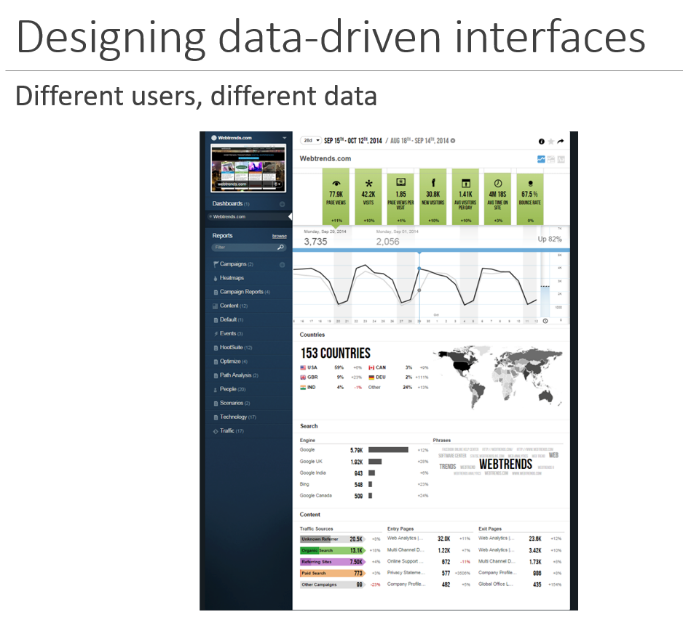
从实时监控场景来讲，Dashboard 会被分为实时和历史两种，两种略有差异。实时侧重于监控，历史侧重于了解和分析。这两种没有明显的分界，实时需要历史的信息作为对比，来判断当前的数据是否正常。而历史也需要准实时的信息来更快地了解当前的情况。这两种形态的Dashboard 对于数据产品来讲都必不可少。

这个图是1个较好的实时 Dashboard 设计方案，**核心在于细分维度的多维监控，并确定合适的阈值点**。

移动场景考虑到屏幕等硬件条件的限制，则侧重于通知和展示，不深入到分析部分。到具体产品设计上，则是通过APP 等方式实现，提供最核心的数据查看和智能挖掘，不做过多的查询功能和复杂的交互。同时，基于现有大部分人通过手机进行沟通，那么页面或数据的分析也变得重要。

同样以 Domo 为例，它在移动端也有相应的场景划分，支持消息的移动推送和展示：

**P9：少胜于多，分清主次**



数据面板设计原则的第2个细则——少胜于多，分清主次，主要说的是：

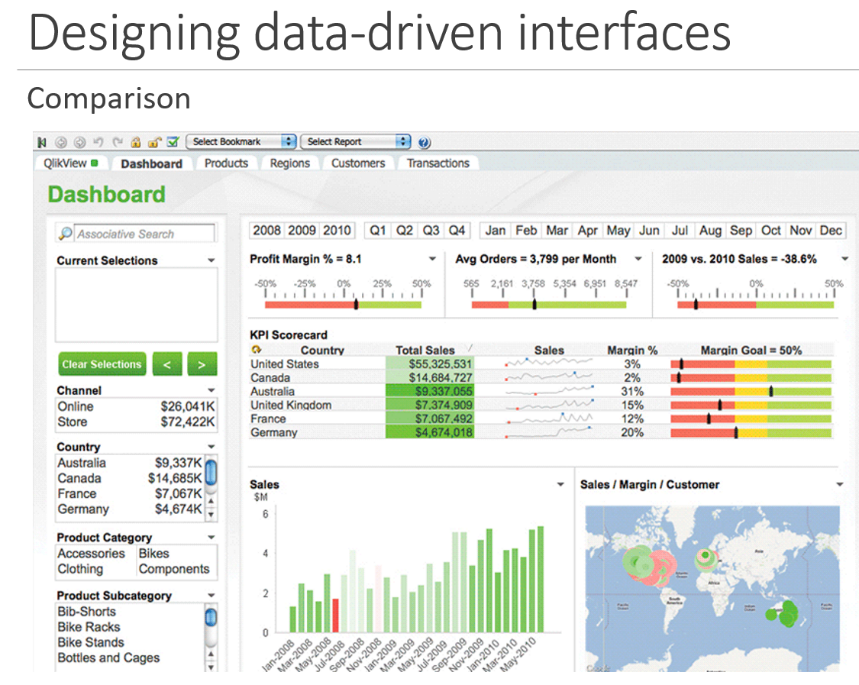
根据「提纲挈领，引导分析」的原则，仪表盘的界面上优先展示的指标应该是用户最关心的，且因为每个用户时间精力有限，在数量上要极度精简。在挑选指标时，要遵循以下原则：

1. 从用户的需求出发
2. 精准，精确反映当前业务情况
3. 核心指标不超过7个
4. 确定核心指标间的联系及优先级

图中Webtrend网站创立于 1993 年，是目前公认市场占有率第一的商业网站分析方案，为接近 2000 个公司提供服务。作为一款通用的商业分析产品，它在 Dashboard 页面提供了网站用户最关心的 7 个指标，包括最近 30 天的汇总访问量，访问深度，新访问用户，日均访问用户数，平均访问时间，跳出率。同时通过与上个30天的环比的增减，跌涨与红绿对应，能够使用户一眼就了解到当前业务情况。

同时，这几个业务从几个侧面整体衡量了网站的用户数量及质量，既能监测网站访问用户数的情况，同时通过跳出率等指标来分析访问用户的质量，可监测爬虫程序或及时识别伪造用户。

**P10：深入分析，落地行动**



数据面板设计原则的第3个细则——深入分析，落地行动，主要说的是：

由于分析无非「对比，细分，溯源」，从这个角度讲，仪表盘的设计需要让用户能够在「提纲挈领」之余，也要能提供给用户分析思路。当发现数据异常时，能够沿着思路自主得到答案，或者分析方向

【3.1】

对于分析来说，首先：没有对比就没有分析，从各种指标的对比中才能看出指标的偏离。对比可分为三大维度

1. 时间维度：同比，环比
2. 空间维度：地区对比，团队对比，商品／服务/渠道对比
3. 设计维度：目标值，业界值，极限值等与实际值的对比

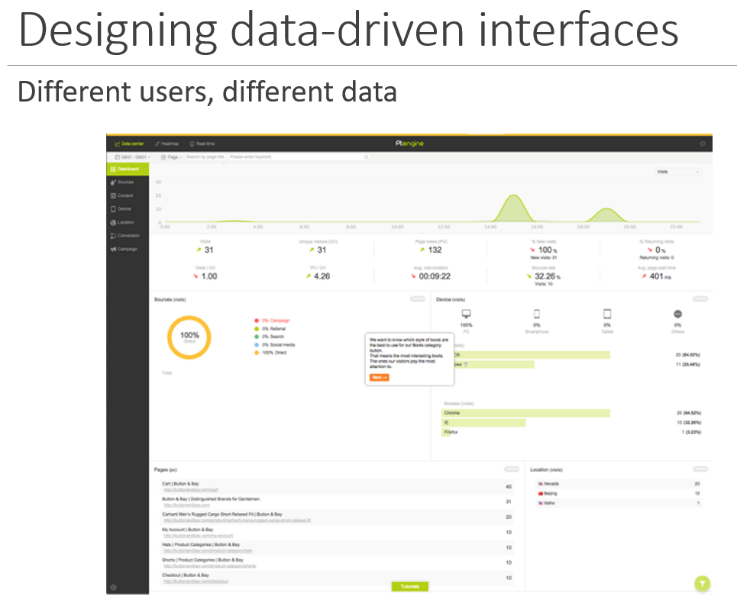
图中是老牌的可视化厂商的产品 QlikView，它将对比的概念发挥的淋漓尽致，基本涉及了以上各种对比类别。

在它的 Sales 图表中，提供了各个月份间的时间维度对比。在 Sales/Margin/Customer 上和左侧的列表中，则提供了地区，渠道，品类等各空间维度的对比。而在 KPI Scorecard（计分卡）图表中，是设计维度的对比：「当前销售额」和「某个时间段内的最高值和最低值」，「利润率」和「目标利润率」&&「利润率区间」，从而使所有地区的销售情况一目了然。

从上面的例子可以看出，对比对于显示关键信息来讲非常重要。在设计相关的对比模块时，一要在合适的维度上进行对比，这个维度应该对当前指标的变动起主导作用。二要突出对比的结果，方式有颜色变化，内容闪烁或数据报警等。

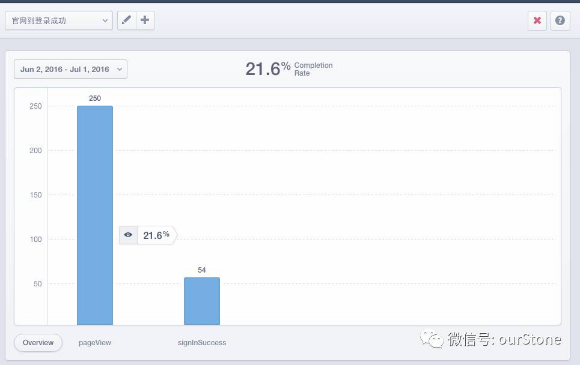
这里顺便提下对于红绿色含义的认知差别。有些国内的数据产品会以为在中国股市中，红涨绿跌，并且在中国传统文化中红色代表喜庆，所以在数据产品中红色应该也代表数据变好。其实数据可视化的原则是，高效地向用户传递数据信息。而用户天然对红色的内容会比较敏感，红色应该用来传递更重要的信息。数据下跌比数据上涨要重要得多，因此数据产品在指标的显示上，应该仍遵循红坏绿好的原则。

**P11：**



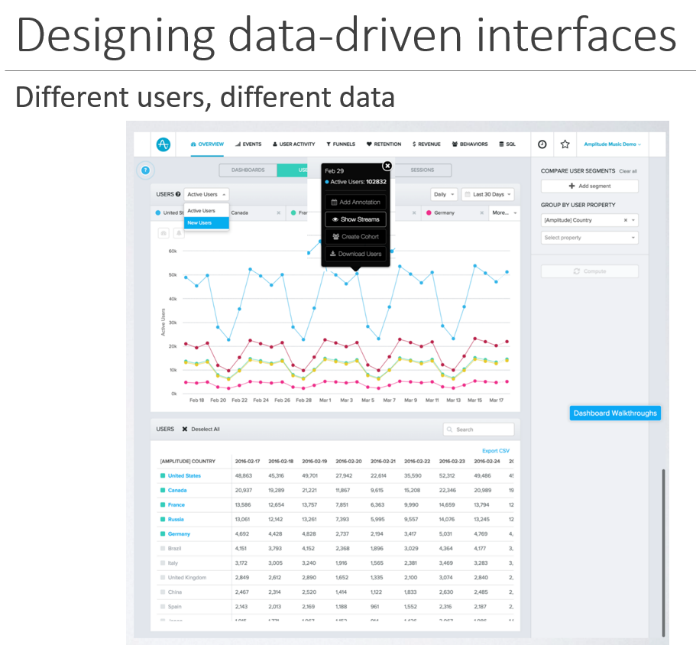
【3.2】

细分是对核心指标进行多维度的划分，分为单维度细分，多维度细分，流程细分及TopN细分（TopN 细分则是注重看中某个维度下占比前列的维度值的变化，来直接反映某些指标值的变化）等等。



以图中例子为例，它提供了对不同核心指标的细分，左侧包括访问来源，访问地区，访问设备及页面。同时，用户可以在页面对不同指标进行切换。当你对某些数据有疑问时，你可以通过点击相应的维度跳转进入各个子页面，起到了「引领分析」的作用。

**P12：**



【3.3】

溯源作为一个对于数据异变根本原因的追查过程，很难融合在以简洁为原则的 Dashboard 中，不过 图中例子 通过隐藏选项并且通过和内在其他功能的融合，很好地解决了这个问题。

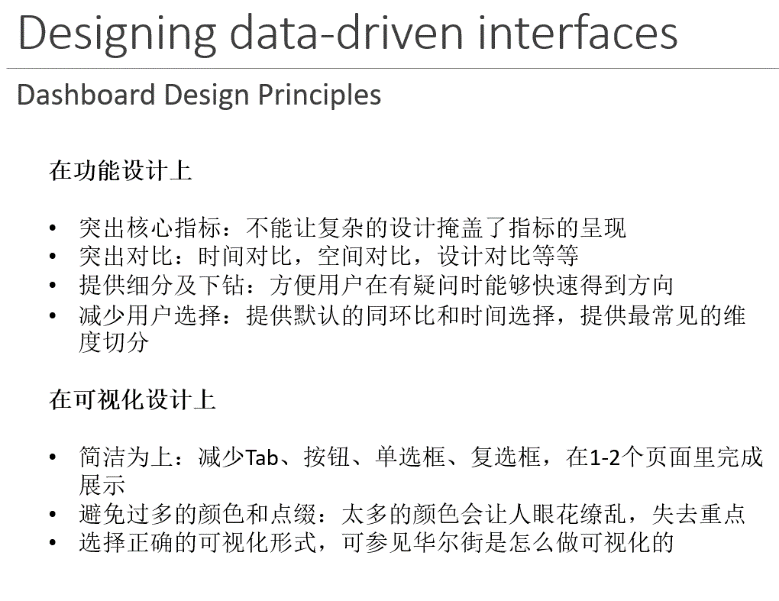
图中例子中首先提供了不同主题，不同形式的数据的对比细分展示。而在某类具体业务和数据点上，我们可以选取某个时间点的用户群体作为 Cohort（队列）对象或者针对这批用户的流向结合产品内其他功能进行分析。它甚至能够将整个用户id列表下载下来，做线下分析。

分析后，需要落地行动，在这个过程中，要反复地问自己“So What”。从设定用户场景，到确定指标和优先级，再经历对比、细分、溯源三个分析流程，最后要做的，就提供给用户决策和行动的建议和方向。

一个典型技巧是：先假定几个异常的场景，然后通过设计出来的仪表盘，演练拆解场景中出现的问题。如果能够在若干个场景中都顺利走通，那就证明你整个设计能够支持用户做出决策和行动。

再往上一层，就是能够直接给出业务建议的层次了。这个一方面需要对业务的极度熟悉，另一方面可能还需要熟悉数据挖掘和机器学习的内容。

**P13：Dashboard 设计原则**



刚开始做 Dashboard 设计时，不要一上来就画原型图，而是先划分用户和场景，然后用思维脑图确认核心指标和衍生维度。再根据思维脑图画出草图，演练几遍，在整个流程都走顺之后，再开始仪表盘的设计。

在设计细节上，不要刻意追求炫酷的效果，而是注重设计的内在逻辑和传递消息的有效性。这里有一些设计原则可以参考：

在功能设计上

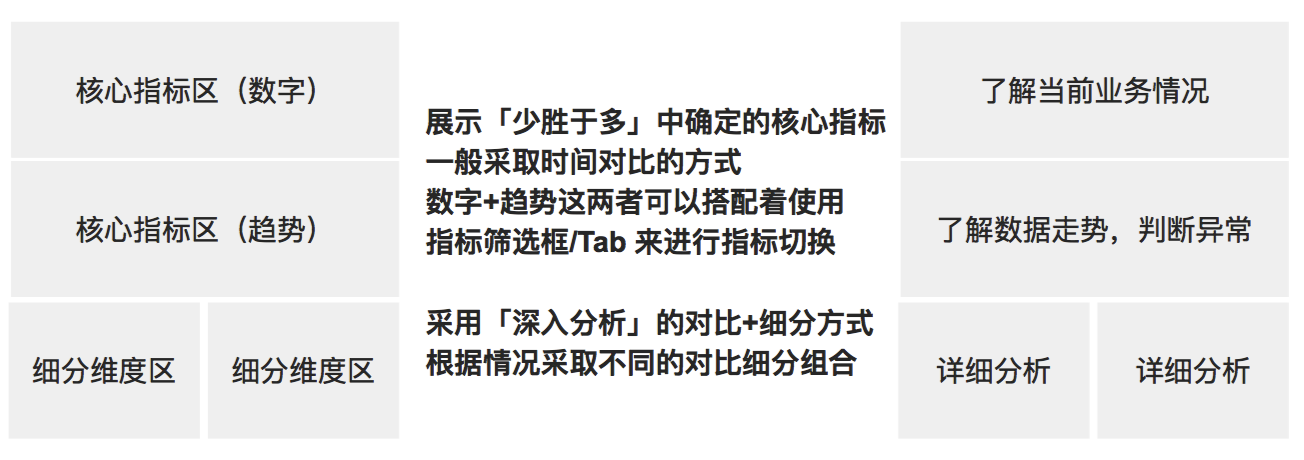
* 突出核心指标：不能让复杂的设计掩盖了指标的呈现
* 突出对比：时间对比，空间对比，设计对比等等
* 提供细分及下钻：方便用户在有疑问时能够快速得到方向
* 减少用户选择：提供默认的同环比和时间选择，提供最常见的维度切分

在可视化设计上

* 简洁为上：减少Tab、按钮、单选框、复选框，在1-2个页面里完成展示
* 避免过多的颜色和点缀：太多的颜色会让人眼花缭乱，失去重点
* 选择正确的可视化形式，可参见华尔街是怎么做可视化的

可供参考的设计框架有以下两种，

**（1）总分式，先展示核心指标，再对核心指标进行拆分**



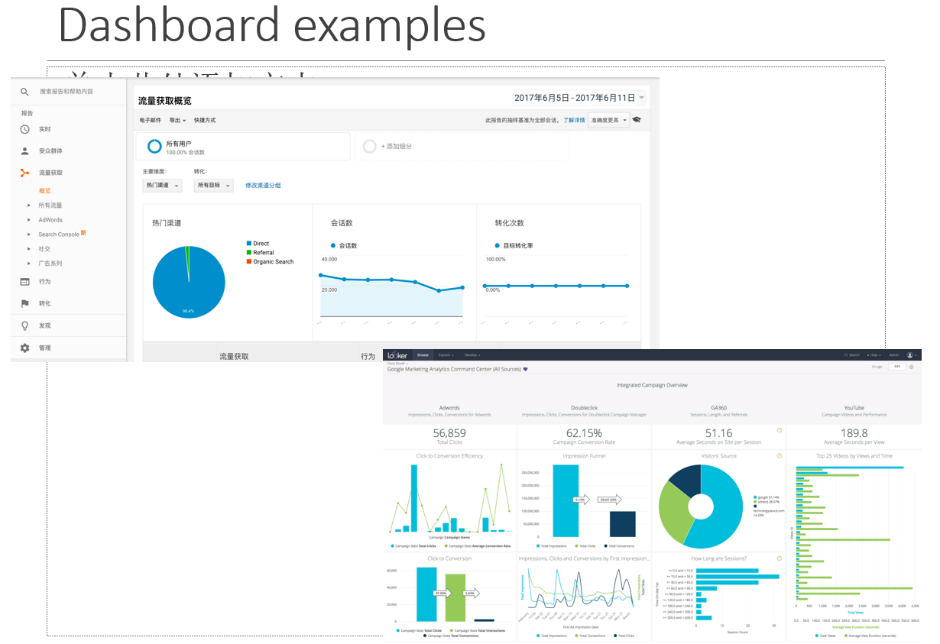
**（2）分散式，常见于 Customize （个性化）类型的数据产品，形式是若干个报表集中在一个页面展示**

此类产品一般没有等级明确的金字塔结构来承接分析思路。此类形式胜在自由，但缺陷在于信息量太大，让用户一下子不知道该关心哪个指标。所以建议此类产品必须可以定制每个单图的大小，从而起到「少胜于多，分清主次」的作用。

为了追求视觉效果，在大屏展示场景下，也会经常采取这种布局方式。



**P14：**



我们来看一下这个数据看板例子：

第一个图是Report Dashboard，在Google Analytics 中根据分析主题划分出受众群体、流量获取、行为\转化四个子菜单，每个菜单会再按照菜单层次依次往下拆分，就像一层层金字塔一样。

第二个图是Customize Dashboard，在Looker 中，会采取这种分散的图表的组织方式，每个图表都可以再次编辑。整个分析页面，不再是简单的图和表的堆砌，而是变成了一个画布。用户可以在这个画布上放置他们任意想要的内容。

**这两种方式各有利弊。**

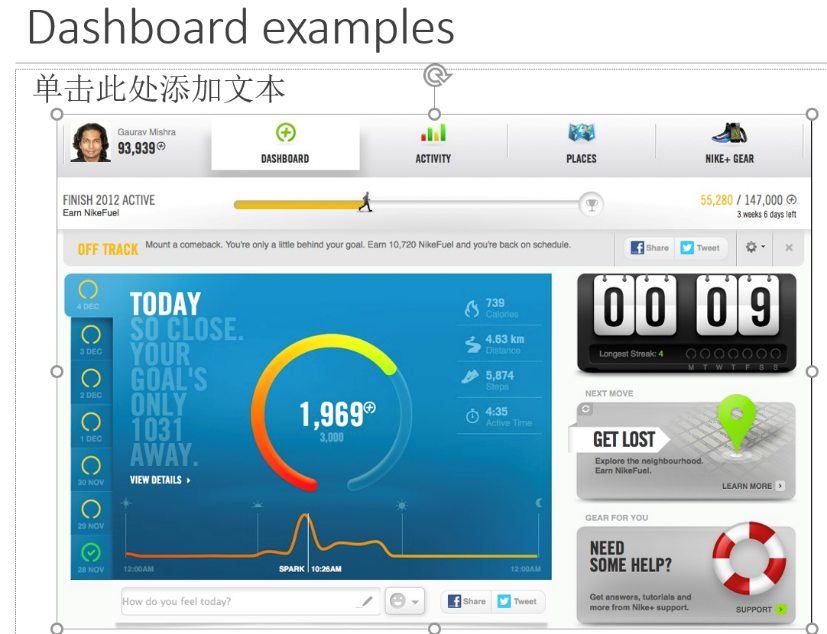
Report Dashboard 适用于整个市场的数据利用意识还处于早期阶段的时候。此时大家的分析思路匮乏，对指标的需求比较单一，只能由设计者提供一个抽象普适的分析思路，供用户日常使用。

而随着数据运营的方法论越来成熟，整个数据采集和分析工具越来越先进时，原先的方案已经不能满足人们的需求。不同的人，同一个人不同时期关注的内容都不一样。因此对个性化的要求便能越来越高，Customize Dashboard越来越重要。

Customize Dashboard 会面临几个问题：1，对底层数据的规范要求较高；2，对使用者要求较高，这需要用户有十分明确的分析思路及定义指标的概念。

为了解决这个问题，很多产品便提出了两者相融合的方式：在 Customize Dashboard 的基础上，提供一些默认的报表，方便用户使用。允许用户修改，从而具备更大的自由性，适用于跨业务跨部门的情况。

**P15：**



在数据产品之外，在管理后台产品，用户端产品时，也有这种集中展示数据，帮助用户快速了解情况的场景。大多数的这类场景集中在运动类产品，理财类产品和工具型产品上。大家可以借鉴一下

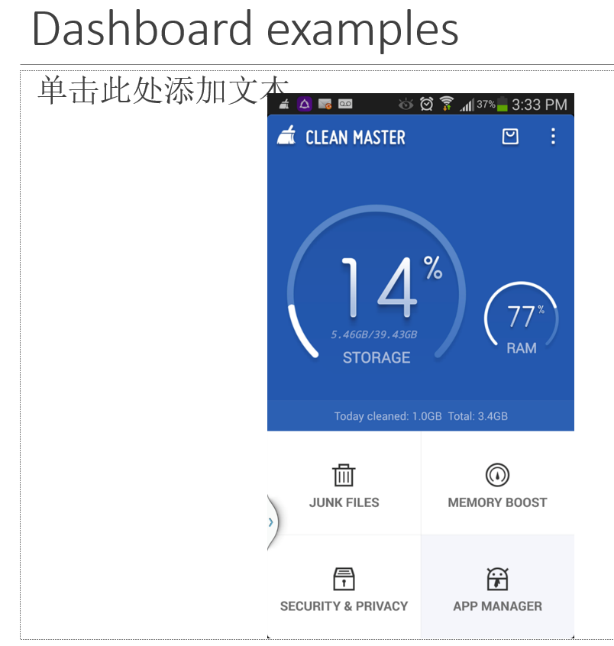
图中是运动类产品的数据看板设计案例

**P16：**



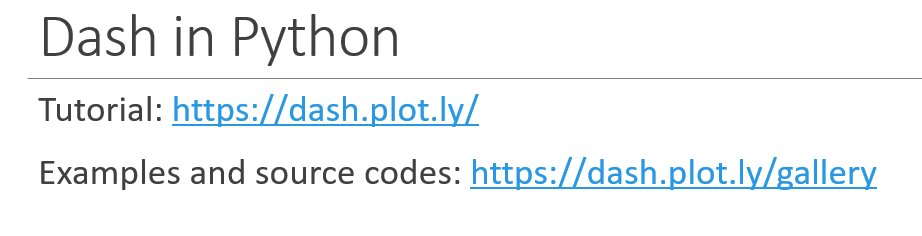
图中是理财类产品的数据看板设计案例

**P17：**



图中是工具类产品的数据看板设计案例

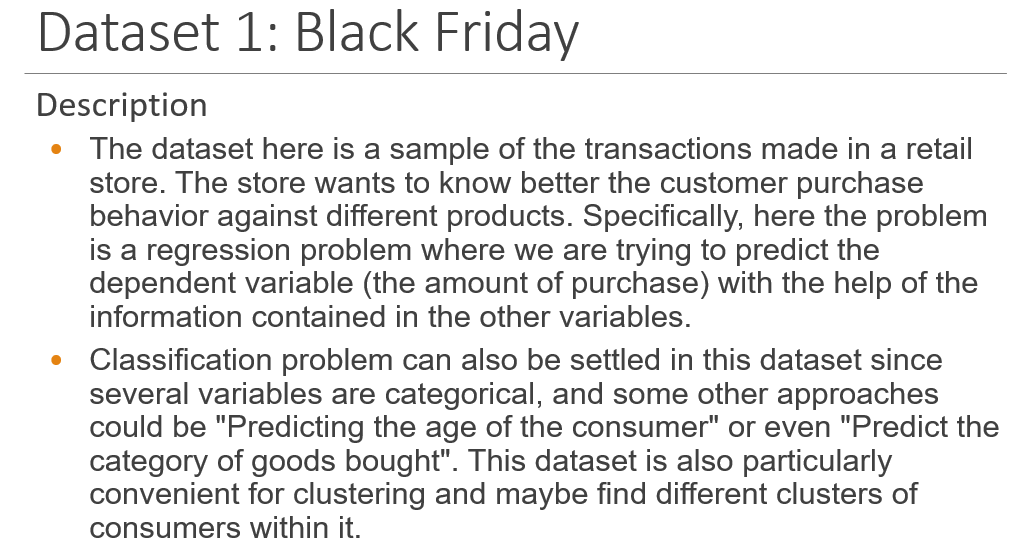
**P18：如果用Python实现数据看板或者仪表盘**



大家可以从这个教程入手。

它里边有例子和源代码。

**P19：**



最后，我们为本次个人项目准备了3个数据集。第一个数据集是一个零售商店的黑色星期五交易数据。

该商店想通过已获得的数据，更好地了解顾客对不同产品的购买行为。

具体来说，这里的问题是一个回归问题，我们试图借助其他变量中包含的信息，来预测因变量（也就是购买量）。

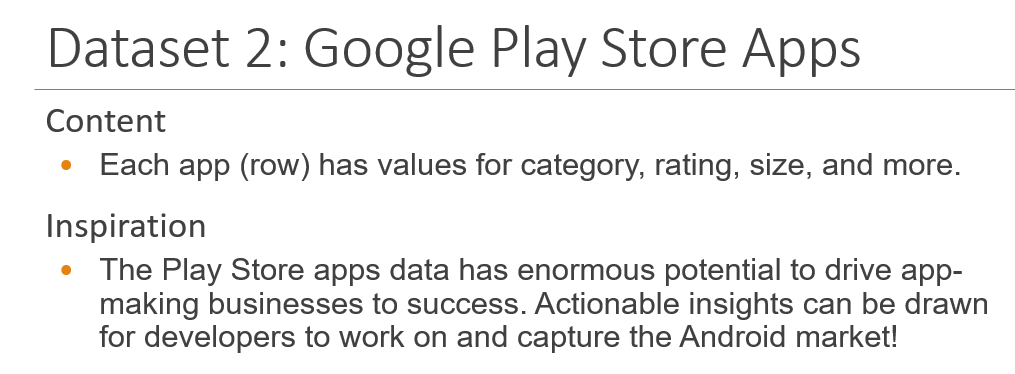
---

分类问题也可以在这个数据集中解决，因为有几个变量是可用于分类的。

这个数据集还可以用来 "预测消费者的年龄"，甚至是 "预测购买的商品类别"。

这个数据集对于聚类也特别方便，可以在其中找到不同的消费者聚类。

**P20：**



第二个数据集是谷歌商店的数据。内容包括了：每个应用程序的类别、评级、大小等方面的数值。

谷歌商店的数据是很有应用潜力的。它可以推动应用制作企业走向成功。可以得出可操作的见解，供开发者研究并占领安卓市场

大家可以想想如何针对应用，对这个数据进行可视化

**P21：**



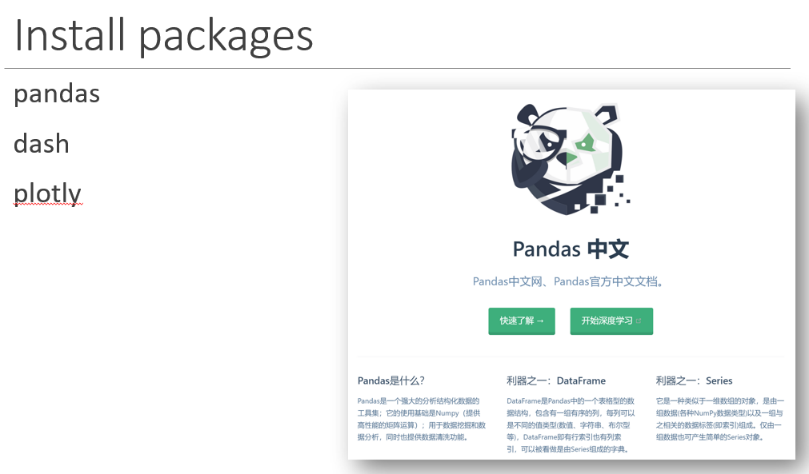
第三个数据集是高校薪资

里边包括了：

* 按学院划分的薪资增长情况
* 按地区划分的薪资情况
* 按专业划分的薪资涨幅情况

所有数据均来自《华尔街日报》，基于Payscale公司的数据。

**P22：**



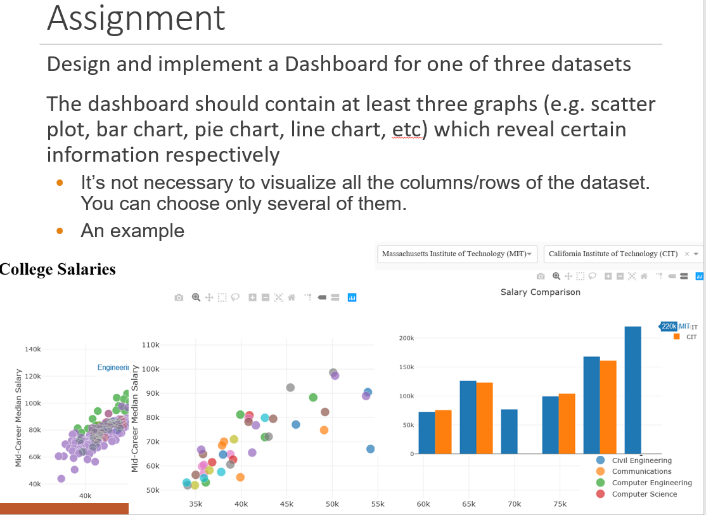
大家在实现的时候，需要装pandas。它是Python 语言的一个扩展程序库，应用于数据分析。它的使用基础是Numpy（Numpy提供高性能的矩阵运算）；pandas除了用于数据挖掘和数据分析，同时也提供数据清洗功能。

也要安装dash库。Dash是一个用于构建Web应用程序的Python数据可视化库。它基于Python Web框架Flask，以及Javascript绘图库Plotly.js，和用于构建用户界面的Javascript库React.js，所以它非常适合用于构建后端基于Flask，前端数据可视化的Web网页。因为Dash应用程序可以在Web浏览器中显示，所以可以将这些Dash应用程序部署到服务器，通过URL共享。

https://blog.csdn.net/LesuperV/article/details/100533632

绘图库用plotly就可以。plotly绘图种类丰富，效果美观。ploty默认的绘图结果，是一个HTML网页文件，通过浏览器可以直接查看，可以与Web无缝集成

**P23：项目任务**



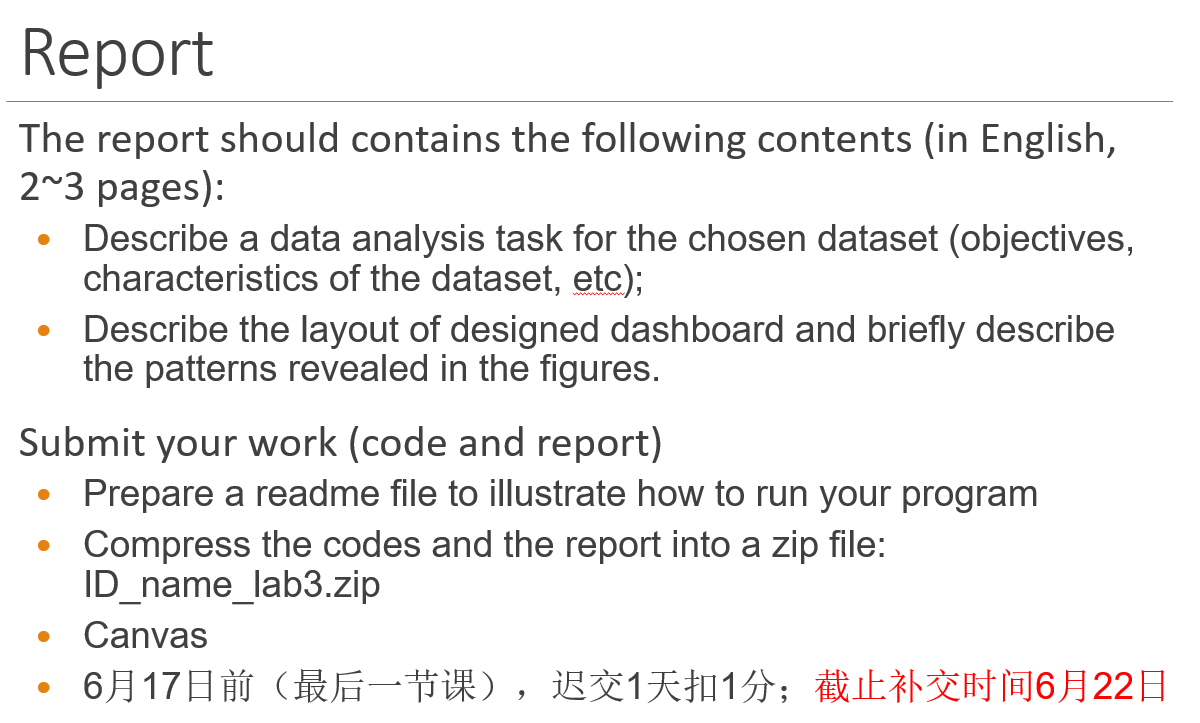
为三个数据集中的一个设计，并实现一个仪表盘

仪表盘应该至少包含三个图形（如散点图、柱状图、饼状图、折线图等），分别揭示某些信息

不一定要把数据集的所有列/行都可视化。你可以只选择其中的几个。

这里图是一个任务例子，供大家参考

**P24：任务要求**



报告应包含以下内容（英文，2~3页）。

描述所选数据集的数据分析任务（目标、数据集的特点等）。

描述所设计的仪表盘的布局，并简要描述数字中所揭示的模式。

提交代码和报告。报告我这里有个模板可供参考，需要的话，请大家联系助教