

# 如何用数据讲一个好故事

Tableau可视化设计原则



优达学城  
UDACITY

# 1

## 选择数据集

GitHub数据集介绍

难度	数据集	概述	说明
初级	<a href="#">泰坦尼克号数据</a>	泰坦尼克号上2224名乘客及船员的数据子集，包括了人口统计资料和乘客信息。点击查看 <a href="#">更多信息</a> 。	创建数据可视化，展现幸存者和遇难者的人口统计资料或乘客信息。
初级	<a href="#">棒球数据</a>	该数据集包括1157名棒球球员的偏手性（惯用左手 / 右手）、身高（英寸）、体重（磅）、打击率和本垒打数量。	创建数据可视化，展示各个球员的不同表现。
中级	<a href="#">航班数据</a>	该数据集包括来自 <a href="#">RITA</a> 的美国航班延误和飞行情况。你可以直接从RITA下载，或通过左边的“航班数据”链接下载CSV压缩文件，该文件以年代为序，较原始文件更小。	研究随时间发展，不同航班飞行情况的变化，或研究某一特定年份的数据并创建图表，展示你的发现。

- 我对哪个主题更感兴趣（有兴趣探索的数据才能让你讲出好故事，仅仅为了更快通过项目而选择“简单”的数据显然不是个好主意。
- 对选择的主题我有什么想解答的疑问？
- 通过什么样的结论我可以回答我以上的疑问？



# 1

## 选择数据集

最受同学欢迎的3个数据集

- 泰坦尼克号数据（初级）


数据提供了泰坦尼克号上每位乘客的基本信息及生还情况，主要探索方向可以是所有以及生还乘客的社会组成。

- Prosper贷款数据（中级）

一个维度较广的数据源，来自美国P2P借贷平台Prosper。探索方向很广阔，结合贷款的利率（预期以及实际值），贷款人的职业，贷款目的，贷款状态，贷款人负债比，贷款人就业状况等可以得出很多有趣的结论。

- 摩拜上海成区用户使用（中级）

结构较简单，但需要简单清洗的数据。包含了一个月内所有行程的经纬度，难度在于经纬度并不能在Tableau中直接转化为地理信息，例如上海的行政辖区或地铁站。如果你对地图类图表（尤其是个性化地图，如[Mapbox](#)）感兴趣，这将是一个很好的练习。



## 1

## 选择数据集

工作中的数据集 - 各式各样的Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Violent Crimes In 2016 in the United States by City and State																
2																	
3																	
4	Location		Months														
5	city	state	Apr	Jun	Jul	Aug	Sep	Oct									
6	Albuquerque	New Mexico						46									
7	Anaheim	California			4												
8	Anchorage	Alaska	1						26								
9	Arlington	Texas					17										
10	Atlanta	Georgia						85									
11	Aurora	Colorado						16									
12	Austin	Texas					28										
13	Bakersfield	California			22												
14	Baltimore	Maryland							230								
15	Boston	Massachusetts						28									
16	Buffalo	New York						38									
17	Chandler	Arizona						3									
18	Charlotte-M	North Carolina			25												
19	Chicago	Illinois							536								
20	Chula Vista	California	2				1										
21	Cincinnati	Ohio						50									
22	Cleveland	Ohio						89									
23	Colorado Sp	Colorado					15										
24	Columbus	Ohio						70									
25	Corpus Chris	Texas			9												
26	Dallas	Texas					118										
27	Denver	Colorado					33										
28	Detroit	Michigan	5					221									
29	Durham	North Carolina							30								
30	El Paso	Texas						14									
31	Fort Wayne	Indiana						34									
32	Fort Worth	Texas	7					49									
33	Fresno	California				19											
34	Greensboro	North Carolina							20								

state	Total Crimes 2016
Alabama	12
Alaska	26
Arizona	132
California	515
Colorado	64
D.C.	105
Florida	210
Georgia	85
Hawaii	9
Illinois	536
Indiana	151
Kansas	10
Kentucky	95
Louisiana	127
Maryland	230
Massachuset	28
Michigan	221
Minnesota	26
Missouri	223
Nebraska	29
Nevada	128
New Jersey	86
New Mexico	46
New York	290
North Carolin	82
Ohio	217
Oklahoma	82
Oregon	14
Pennsylvania	255

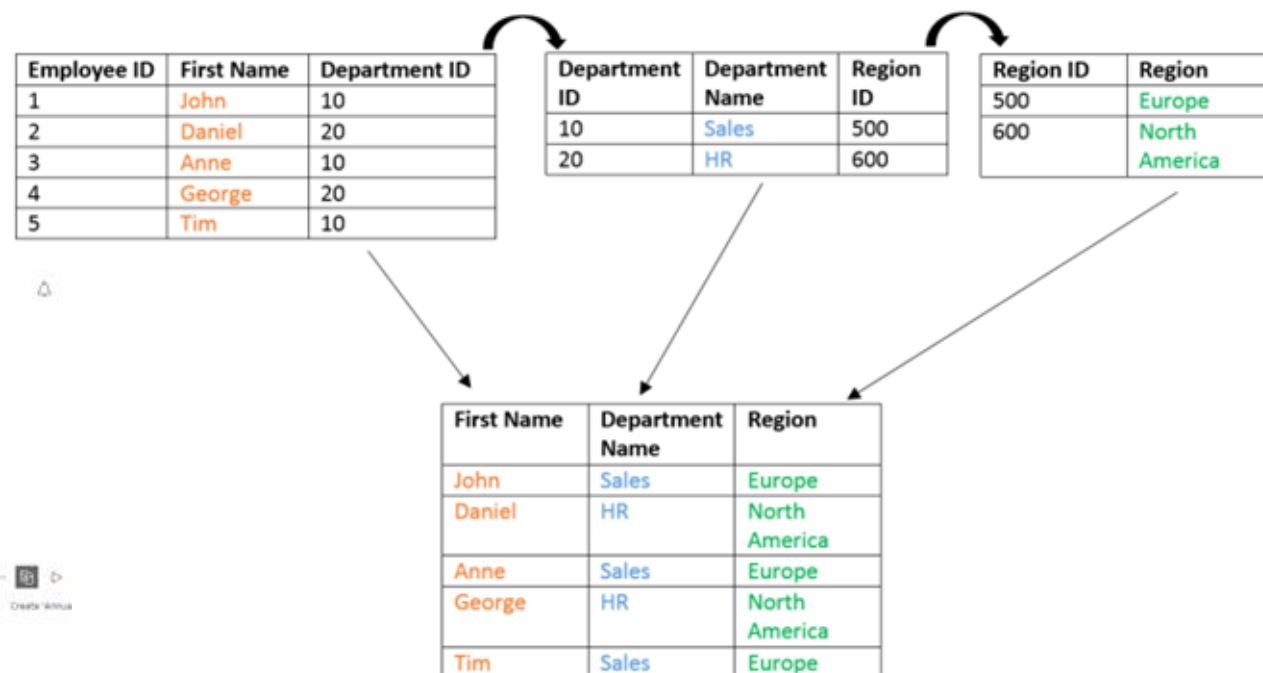
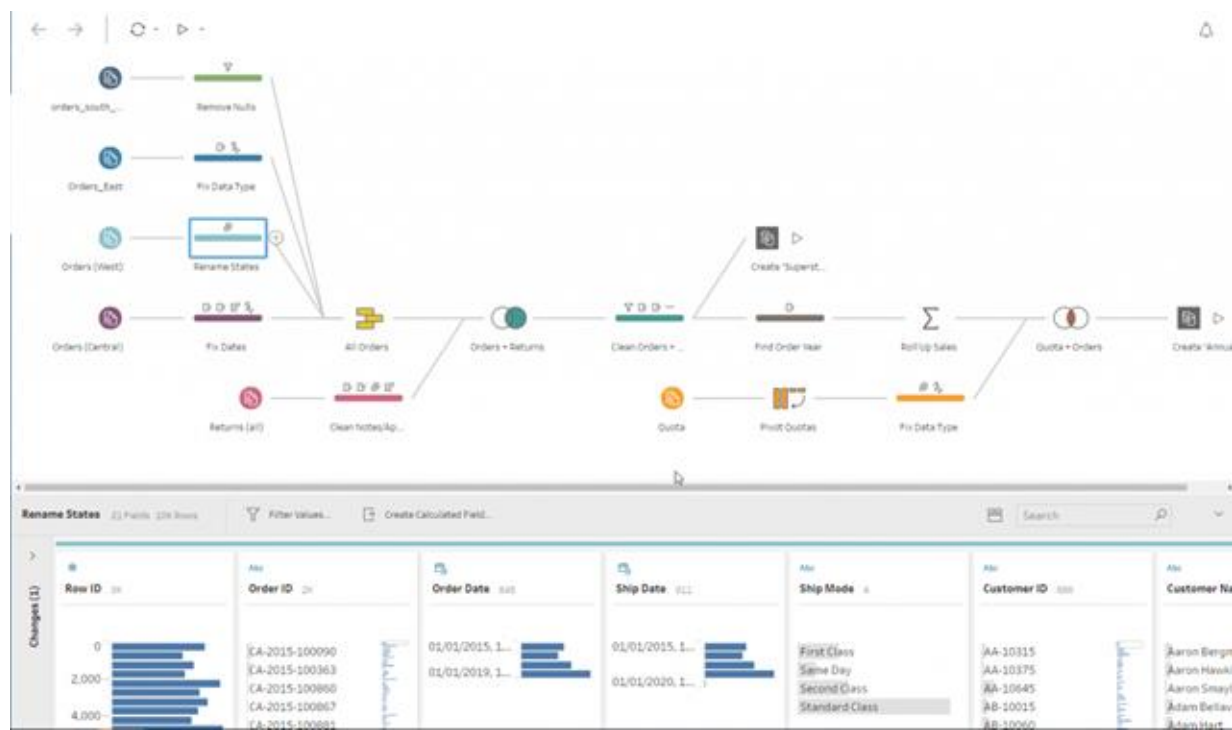
State	Population 2016
Alabama	4860545
Alaska	741522
Arizona	6908642
Arkansas	2988231
California	39296476
Colorado	5530105
Connecticut	3587685
Delaware	952698
District of Co	684336
Florida	20666589
Georgia	10313620
Hawaii	1428683
Idaho	1680036
Illinois	12835726
Indiana	6634007
Iowa	3130869
Kansas	2907731
Kentucky	4436113
Louisiana	4686157
Maine	1330232
Maryland	6024752
Massachusetts	6823731
Michigan	9933445
Minnesota	5525050
Mississippi	2985415
Missouri	6091176
Montana	1038656
Nebraska	1907603
Nevada	2939254

1. Title
2. Merged header cells
3. Extra white space
4. Sub-tables

# 1

## 选择数据集

工作中的数据集 – 高度标准化的数据库






# 1

## 选择数据集

数据集的探索

我有一个年级期末考试的表格，共有三列，分别是学号，科目和成绩。我可能会想知道：

- 一共有多少学生
  - 一共有几门课
  - 每个学生都参加了所有课程吗
  - 如果不是，每个学生平均参加了几门课
  - 哪几门课参加的人数最多，是因为这些是必修课吗
  - 每门课的分数分布是如何的，有空值的分数吗
  - 满分是100吗，及格率是多少
  - 哪门课均分比较高
  - 课程的受欢迎程度和分数分布或得高分的难易程度有关吗
- ( 一个相对有趣的问题被提出 )
- 

# 2

## 项目要求

可视化聚焦于一个具体、清晰的发现



向受众展示一切（探索性分析）是非常诱人的，这可以证明你工作及分析的可靠性，但你需要抑制住这个冲动而把注意力集中在解释性分析（将数据抽象为受众能消化的信息）上。试着在组织结论时回答以下问题：

- 你需要观众了解什么或者做什么
- 你在跟谁沟通  
（你的受众越具体，你就越能成功地沟通）
- 沟通时使用什么样的语气  
（是轻松还是严肃，是庆祝成功还是鼓励行动）
- 什么样的数据可以用来支持你的观点

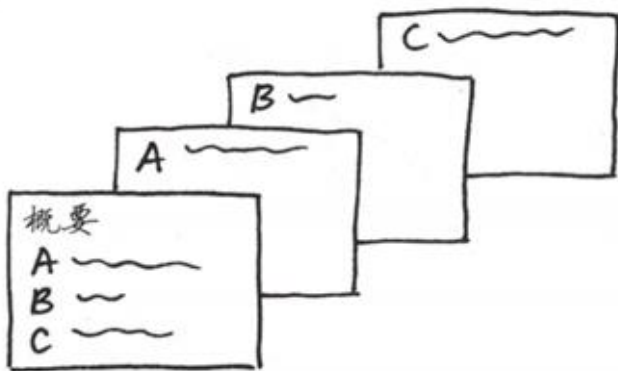
# 2

## 项目要求

可视化聚焦于一个具体、清晰的发现

从这个故事的结构来看，仅仅平铺式地叙述数据事实肯定不够精彩，常用的故事结构有：

- 总分总  
最简单的作文结构，适用于3-5页的故事和较简单的数据
- 前言- A1 - A2 - B1- B2 - 结论  
常用的论文结构，举个例子，A1和A2是A的宏观和微观两方面，B1和B2是B的正反两面
- WHO-WHERE-WHEN-WHAT-WHY-HOW  
5W1H回答式地解析你的中心思想



示例：

待解决的问题：公司Q4利润率过低

提出问题：如何在来年提高利润率（不考虑增加新业务）

- A. 内部分析：是否利润率低于历史记录
  - A1 宏观：Q4盈亏表分析
  - A2 微观：项目点分析，业绩最好和最差的10个项目点
- B. 外部分析：是否低于市场利润率
  - B1 市场标杆分析
  - B2 竞争对手分析
- C. 可实施方案分析
  - C1 价格分析：是否涨价或折扣
  - C2 成本分析：是否削减成本或增加投入

总结，给出决策建议



# 2

## 项目要求

特定发现得以清楚表达

91%

Simple text



Scatterplot



Vertical bar



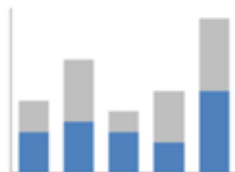
Horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

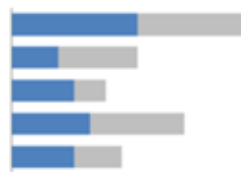
Table



Line



Stacked vertical bar



Stacked horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

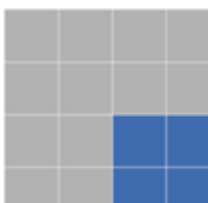
Heatmap



Slopegraph



Waterfall



Square area

### 简单文本

只有一两个数据进行分享时，完全不需要任何图表，只是用数字（尽可能突出）配合辅助性的文字是最清晰有效的

### 表格（包括热力图）

展示多维度的数据，让受众根据需要自行选择关注的点

热力图是表格的变种，通过单色（注意不要使用多种颜色）的饱和度突出数值的变化

表格注意淡化边框（窄边框或无边框）

### 散点图

展示两件事情的关系，观察是否存在及何种关系  
往往配合平均线或趋势线进行使用

### 折线图

暗示点之间存在离散数据（一系列数据分割成不同类别）间没有的联系，通常，连续性数据都以时间为单位

### 柱状图

柱状图的应用广泛，可以代替饼图表示百分比，也可以直观地利用长短比较数据大小

### 斜率图

适用于两个时间段或两组对比数据点，无需解释线的意义和具体变化是多少而直接展示数据的提升或下降以及变化速度（斜率），唯一的缺点是缺少绝对值

### 瀑布图

抽离堆叠条形图的一部分进行重点关注，展示最终数据的组成或其中上升下降等变化

### 面积图

比较极大的数值变化时更实用

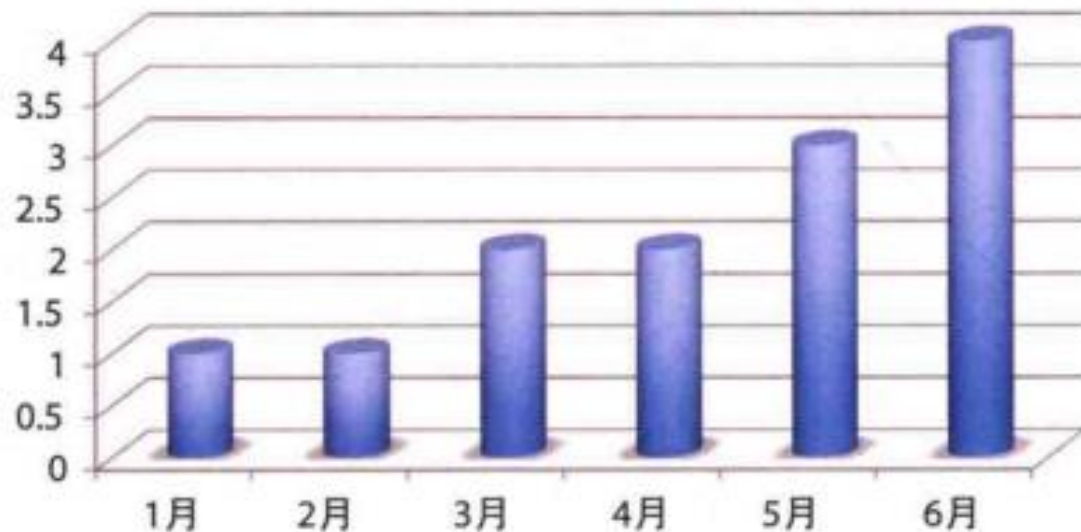
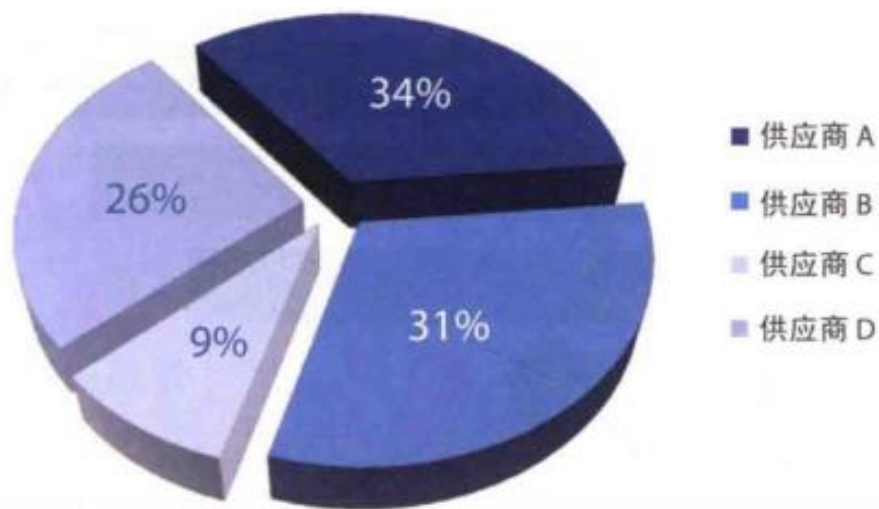
常见的图表有以上几种，无需多“花哨高端”即可满足绝大部分需要

# 2

## 项目要求

特定发现得以清楚地表达

注意，这里我们不推荐使用饼图及3D图形，理由是有更简单的可视化可以达到目的。人眼不擅长比较饼图的弧度（相对于长度）而3D图形使图表变得复杂和扭曲。

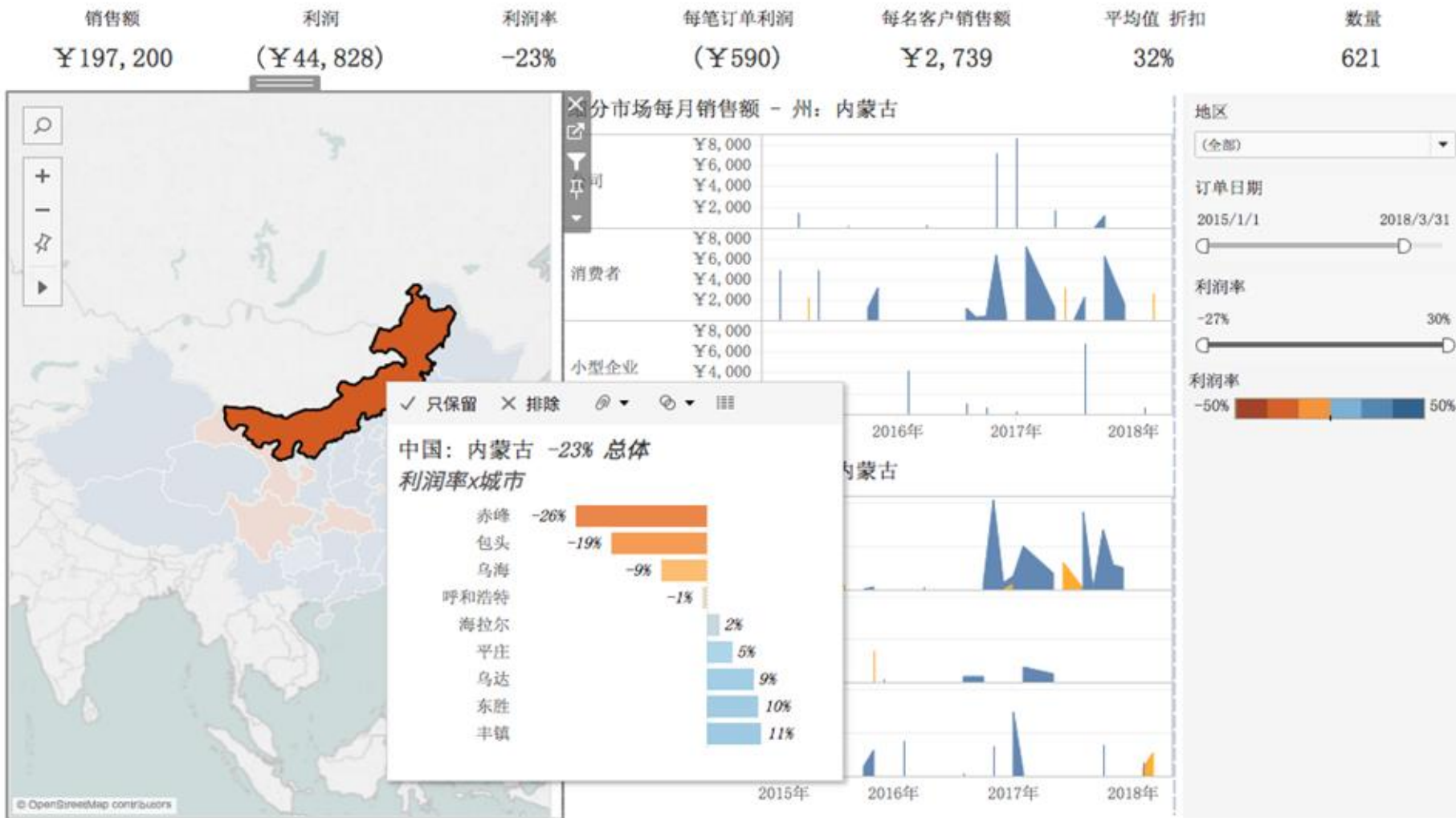


# 2

## 项目要求

可视化交互，仪表盘操作

为高管提供的概述 - 盈利性 (内蒙古)



# 2

## 项目要求

可视化交互，仪表盘操作

编辑筛选器操作

名称:

源工作表

田 概述

☐ 产品销售额  
☐ 细分市场销售额  
☐ 销售总额  
☒ 销售额地图

运行操作方式:

悬停

选择

菜单

☐ 仅在单选时运行

目标工作表

田 概述

☒ 产品销售额  
☒ 细分市场销售额  
☒ 销售总额  
☒ 销售额地图

清除选定内容将会:

☐ 保留筛选器  
☒ 显示所有值  
☐ 排除所有值

bleau - 示例超市 [只读]

操作

外部 Web 资源，或使用筛选操作和突出显示操作将工作表连接到

继续运行	源	字段
选择	概述 (销售额地图)	省/自治区

编辑...

移除

作表的操作

取消

确定

- 添加仪表盘操作的方法有两种
- 直接在仪表盘上点击灰色区域内的漏斗形图标“用作筛选器”
  - 依次点击，仪表盘-操作-添加操作

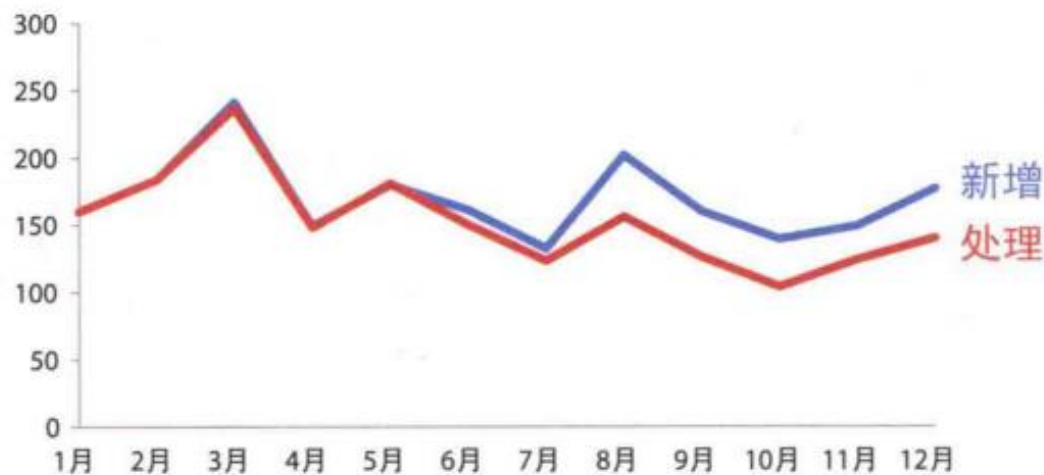
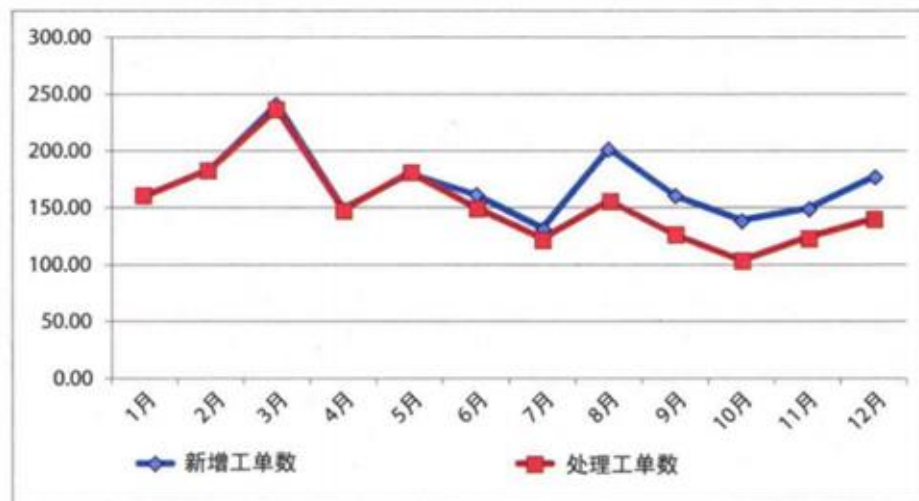
# 2

## 项目要求

图表细节

比较初稿和终稿的优化过程：

- 取消边框及不必要的参考线或网格线
- 去除数据标记
- 清理坐标轴标签，改变缩进并调整文字的倾斜
- 直接标记数据，利用文字及颜色代替图标





## 2

## 项目要求

图表细节

比较修改前后的优化：弱化边框，利用单色饱和度消除过多颜色，强调前5，使用完整文字

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	9	4	7
FRA	3	2	4	9	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



	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



# 2

## 项目要求

### 聚合数据与计算字段

可视化项目的此项要求在于考察你进行汇总或更改数据的粒度的能力。聚合函数不一定通过创建新的计算字段来完成，使用快速计算同样可以达到效果。

聚合计算的规则：

- 任何聚合计算中不得同时包括聚合值和解聚值  
例如， $\text{SUM}(\text{Price}) * [\text{Items}]$  不是有效的表达式，因为  $\text{SUM}(\text{Price})$  已聚合，而  $\text{Items}$  则没有  
不过， $\text{SUM}(\text{Price} * \text{Items})$  和  $\text{SUM}(\text{Price}) * \text{SUM}(\text{Items})$  均有效
- 表达式中的常量可根据情况充当聚合值或解聚值  
例如： $\text{SUM}(\text{Price} * 7)$  和  $\text{SUM}(\text{Price}) * 7$  均为有效的表达式。
- 所有函数都可用聚合值进行计算。但是，任何给定函数的参数必须或者全部聚合，或者全部解聚  
例如： $\text{MAX}(\text{SUM}(\text{Sales}), \text{Profit})$  不是有效的表达式，因为  $\text{Sales}$  已聚合，而  $\text{Profit}$  则没有  
不过， $\text{MAX}(\text{SUM}(\text{Sales}), \text{SUM}(\text{Profit}))$  为有效的表达式
- 聚合计算的结果始终为度量

# 2

## 项目要求

聚合数据与计算字段

常用的命令包括但不限于：

- ATTR(expression): 如果它的所有行都有一个值，则返回该表达式的值。否则返回星号。会忽略 Null 值。
- AVG(expression): 返回表达式中所有值的平均值。AVG 只能用于数字字段。会忽略 Null 值。
- CORR(expression 1, expression2): 返回两个表达式的皮尔森相关系数。皮尔森相关系数衡量两个变量之间的线性关系。结果范围为 -1 至 +1（包括 -1 和 +1），其中 1 表示精确的正向线性关系，比如一个变量中的正向更改即表示另一个变量中对应量级的正向更改，0 表示方差之间没有线性关系，而 -1 表示精确的反向关系。
- COUNT(expression): 返回组中的项目数。不对 Null 值计数。
- COUNTD(expression): 返回组中不同项目的数量。不对 Null 值计数。
- MAX(expression): 返回表达式在所有记录中的最大值。如果表达式为字符串值，则此函数返回按字母顺序定义的最后一个值。
- MIN(expression): 返回表达式在所有记录中的最小值。如果表达式为字符串值，则此函数返回按字母顺序定义的第一个值。
- SUM(expression): 返回表达式中所有值的总计。SUM 只能用于数字字段。会忽略 Null 值。
- 详细级别表达式, 又成为LOD，包括FIXED, INCLUDE和EXCLUDE: 简单来说类似与SQL里的WHERE和HAVING，我们在这里不做进一步的展开，更多关于LOD的教程可参见[这里](#)。



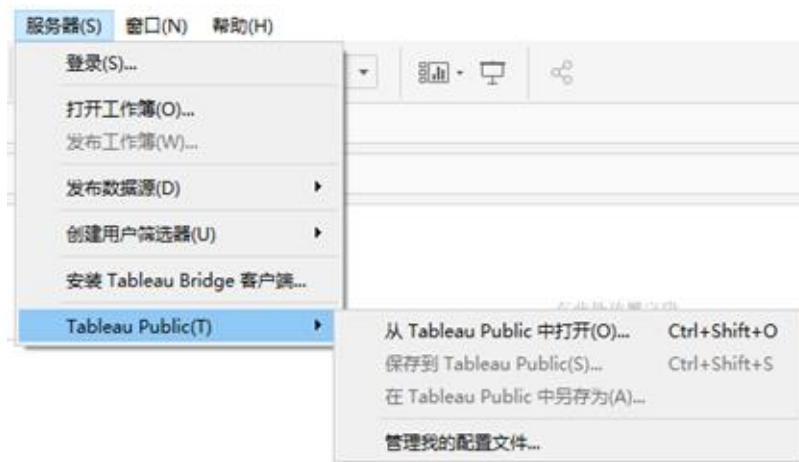
# 2

## 项目要求

文字报告的撰写

文字报告的提交必须以PDF或MD格式，报告（开始处）必须包含最初稿及最终稿的Tableau Public链接。

- 提交的ZIP打包文件中，如果包含了作品原文件，则注意保存为TBLX格式而非TBL格式
- 数据提取（HYPER或TDE）文件不是必须的
- 注意上传至Public的文件和ZIP中版本的一致性
- 隐藏/删除最终故事以外的所有标签页



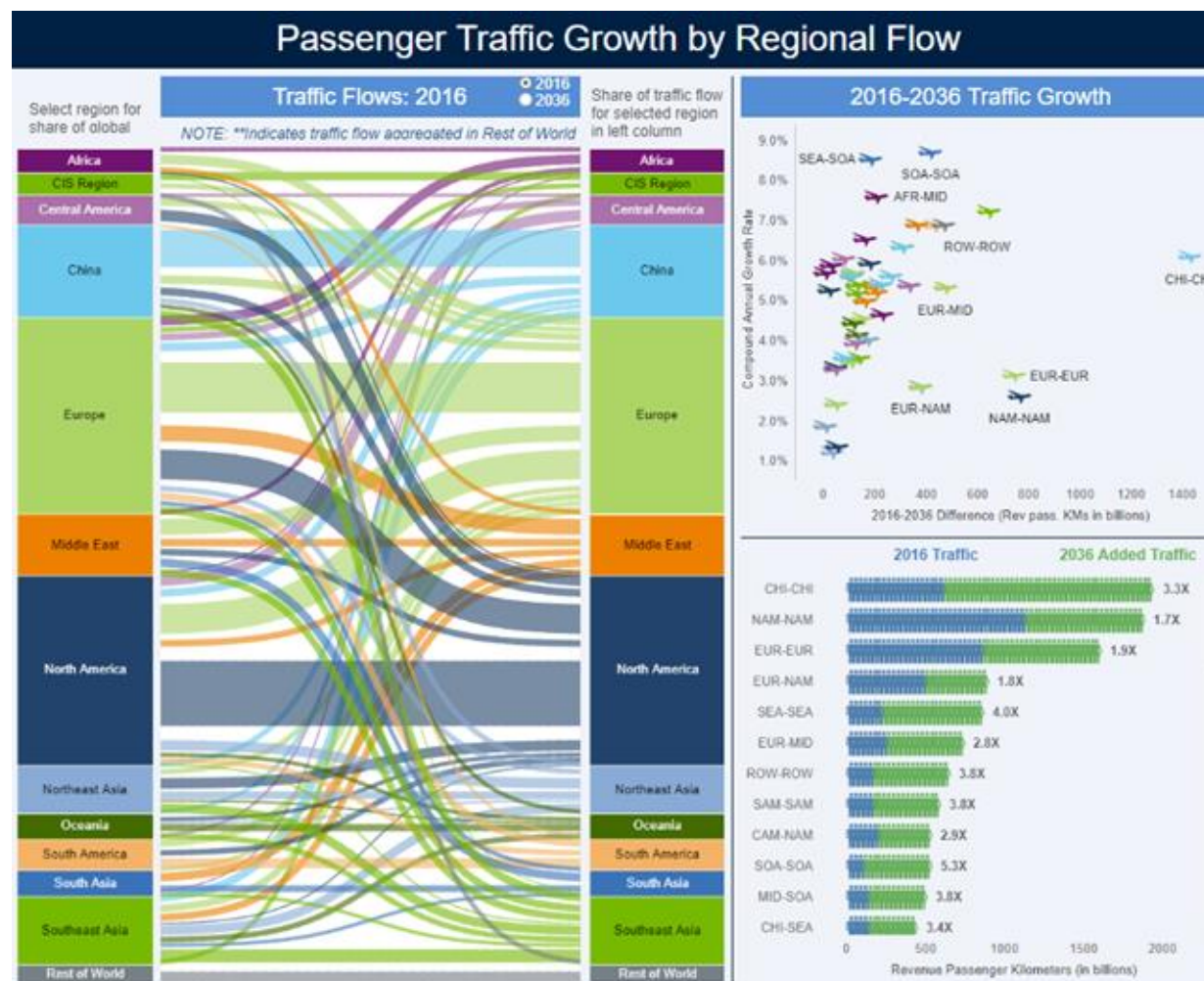
# 3

## 资源及使用

Tableau Public

[Tableau Public Gallery](#), 上有大量持续更新的Tableau官方选出的特色作品。一部分作品的原文件可以被下载在本地打开，你可以学习他们仪表盘元素的构成、计算字段的使用、颜色的组成等。

需要注意的是，这些特色作品作为了解Tableau作品的多样性及可实现性很好，但是这些作品和实际工作中的需求相差非常远。以Tableau官方使用的Tableau报表为例，所有商业中使用视图都以简单、明了以及实际为目的，非常的“冷淡”但要求数据“精准”。



Boeing Current Market Outlook

# 3

## 资源及使用

### Forums

[Udacity Forum](#)（中文版），优达中国官方论坛，你可以按项目板块提出任何问题。当然提问前先搜索是否有已经回答的相似历史问题，并注意提问的问题描述，一般48-72小时内会有论坛导师为你解答

请问如何自定义中国大陆的地理编码以创建填充地图? <a href="#">提问</a>		8	8	4天前
请问如何在动画中进行排序时，每一帧只显示当前排序的内容? <a href="#">提问</a>		4	4	4天前
创建故事求助 <a href="#">提问</a>		5	12	8天前
Tableau联网保存出错~~做了很久，就这样没了??		3	31	11天前
日期由高散改为连续问题		7	10	16天前
请问如果已经保存了所需的版本的story，那么跟这些版本关联的worksheet，dashboard是否都需要保存?		3	10	22天前
LinkedIn 最热技能 <a href="#">提问</a>		4	78	10月7日
Tableau做数据合并		26	22	9月30日
创建一个tableau故事，求项目反馈		2	22	9月20日
Tableau public保存不了。如何处理?		10	24	8月30日
数据进阶用Tableau讲故事项目——寻求项目反馈，谢谢!		1	29	8月27日

[Tableau Forum](#)（英文），这里是Tableau的官方论坛，绝大部分你遇到的技术性问题都可以在这里找到答案。软件本身的bug（如连接数据源报错）可以提交工单联系技术支持人员，一般5-7个工作日内Tableau官方支持会通过WebEx方式远程协助你解决。

Recent Discussions			
<a href="#">How To Filter Using a Secondary Data Source?</a> (20.2)	2 MINUTES AGO IN FORUMS		BY MICHAEL YANOFSKY
<a href="#">Columns change dynamically based on filter</a>	9 MINUTES AGO IN FORUMS		BY LUSIANA DUE
<a href="#">Running Python API through Tableau Desktop</a>	17 MINUTES AGO IN FORUMS		BY LIONEL MESSI
<a href="#">Login to Tableau Server Admin Page - Tableau 2018.3</a>	51 MINUTES AGO IN SERVER ADMINISTRATION		BY SINDHUJA GANGADURAI
<a href="#">Unable to connect to PostgreSQL db</a>	1 HOUR AGO IN SERVER ADMINISTRATION		BY KB0030
Popular Content			
Browse a collection of the most active, insightful threads across our forums.			
<a href="#">BROWSE ALL DISCUSSIONS</a>			
<a href="#">Missing a month in view</a>	11 HOURS AGO		BY A J
<a href="#">Tableau Data Extract Command Line Utility on Tableau Server</a>	20 HOURS AGO		BY BRADY ATOMOS
<a href="#">Need help with time series data - Excel spreadsheet</a>	4 HOURS AGO		BY HUNTER BRANSOME
<a href="#">Map Visualisation: Different color when equal values</a>	8 HOURS AGO		BY RALF BATTAY
<a href="#">Hide nulls without affecting date filter range</a>	4 HOURS AGO		BY PATRICK MYLES
<a href="#">START A DISCUSSION</a>			

**Q&A**  
***Thanks***



优达学城  
UDACITY



## Acknowledgement

感谢以下资深助教对本辅导资料的贡献

虞振远 (Bill Yu)



优达学城  
UDACITY