



蚂蚁金服数据可视化

新一代可视化引擎 G2

章伟星

蚂蚁金服



个人介绍

- 2013 年毕业于哈尔滨工业大学，软件工程硕士。
- 2013.07 - 2014.06 任职于百度，商务搜索质量部。
- 2014.07 - 至今，任职于蚂蚁金服，高级前端开发工程师，主要负责 G2 的开发以及相关数据可视化工作。

团队介绍

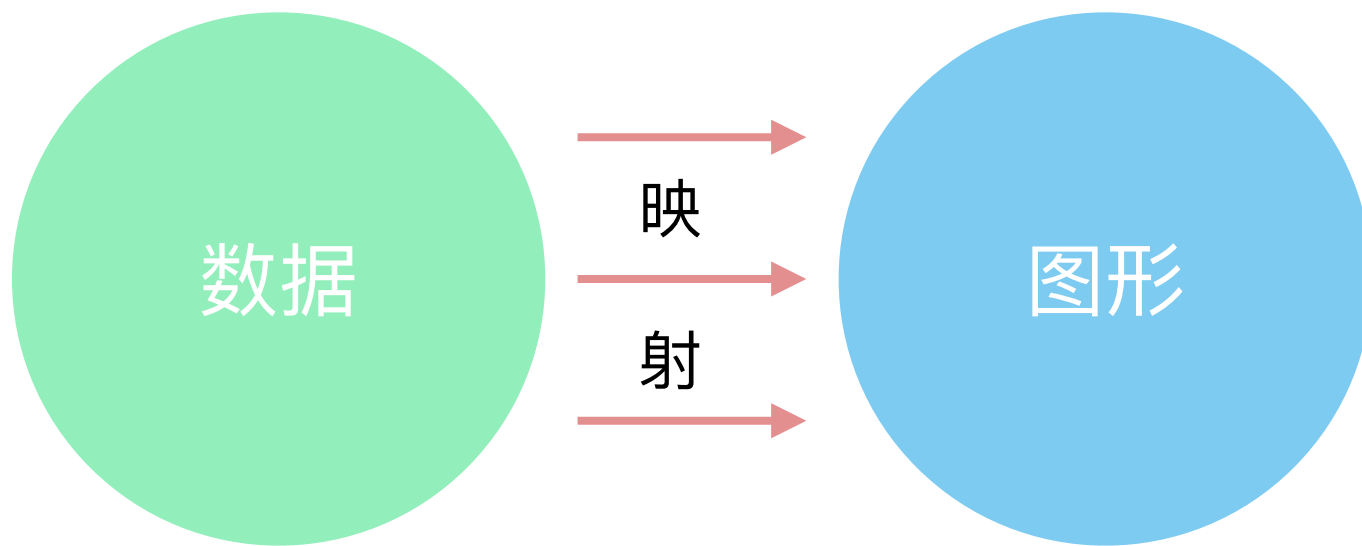
蚂蚁金服集团下的体验技术部，有很多热情似火的设计师，负责交互、视觉、用研、设计研究等工作，也有很多 Web 工程师，玩转 Java 与 Node 服务、多端 UI 架构、工程效能、数据可视化等种种领域。





数据可视化的本质 1

数据空间到图形空间的映射





什么是视觉编码 2

视觉编码

Points



Lines



Areas



几何标记

Position

→ Horizontal

→ Vertical

→ Both



Shape

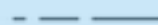


Size

→ Length

→ Area

→ Volume



Color



视觉编码

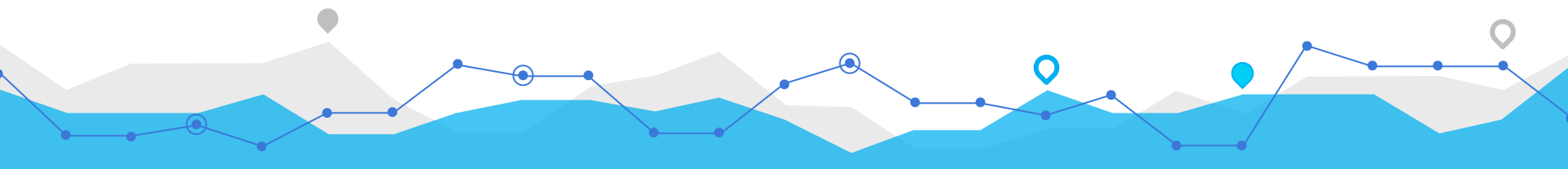
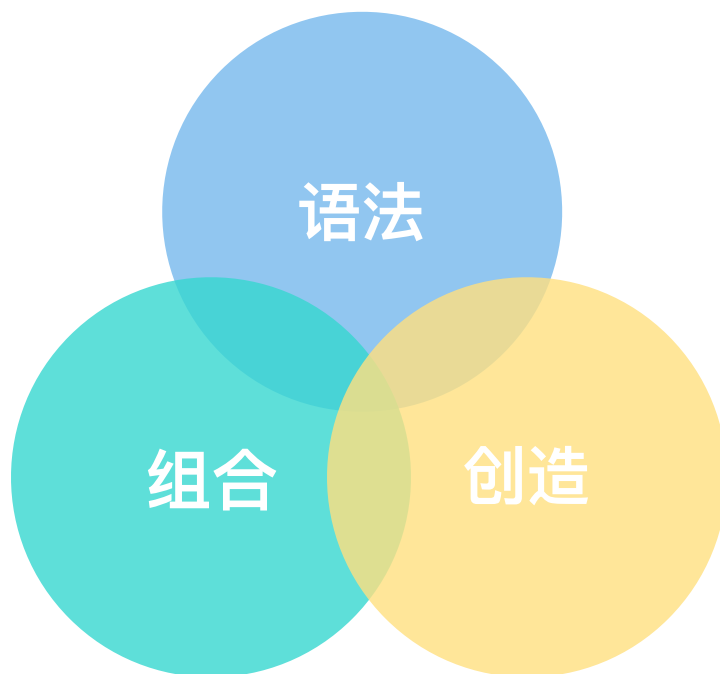


可视化引擎 G2

3

一套属于 DT 时代的图形语法

the Grammar of Graphics





The Grammar of Graphics

一套属于 DT 时代的图形语法

丰富的图表类型

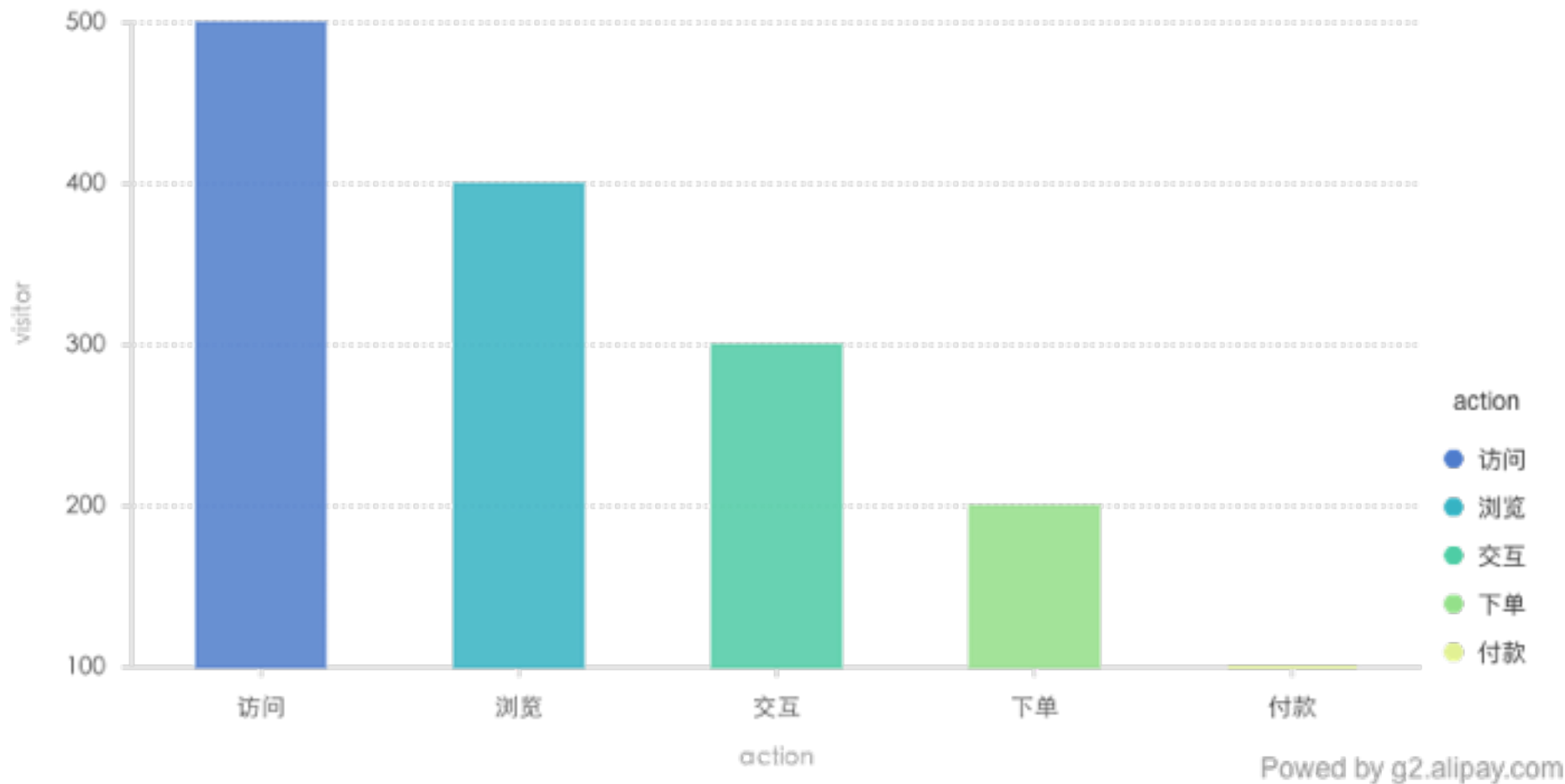
简单、易用
便于扩展

强大的数据分析
能力

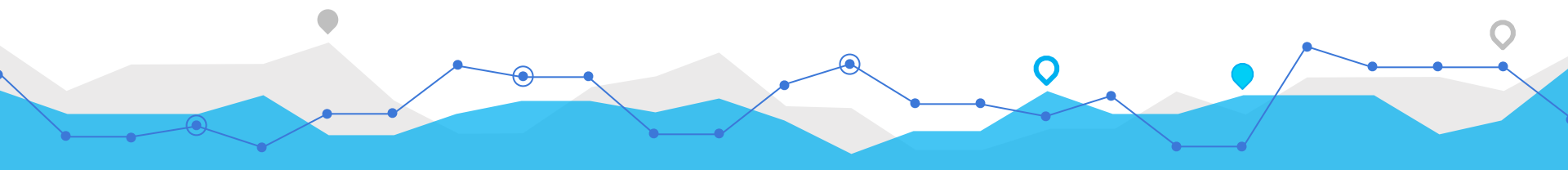


一个简单的例子

antvis.github.io/exec.html



那么，什么是图形语法？



图形语法

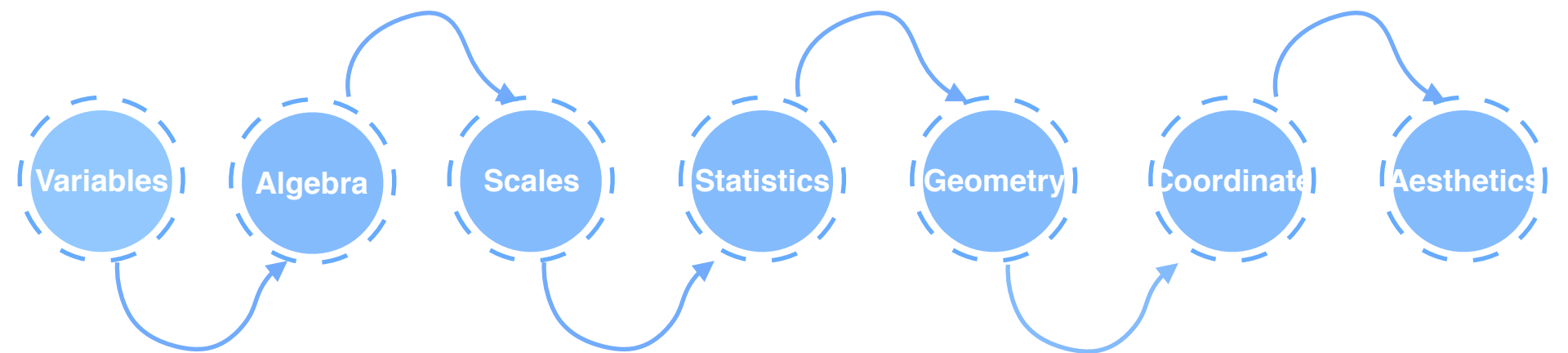
The Grammar of Graphics

Wilkinson(2005) 创建的一套用来描述所有统计图形深层特性的语法规则，该语法回答了『什么是统计图形』这一问题，以自底向上的方式组织最基本的元素形成更高级的元素。



图形语法

The Grammar of Graphics



G2 的可视化框架设计



G2 的数据处理流程





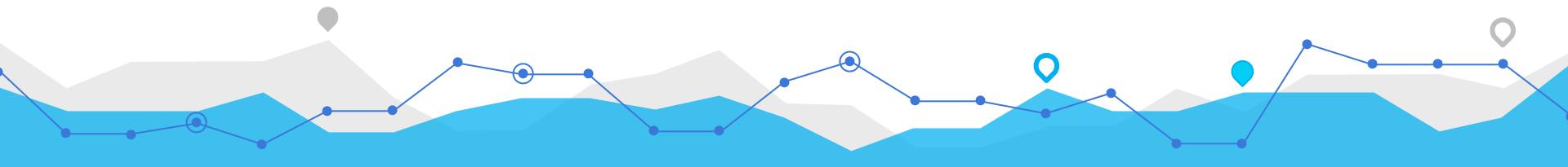
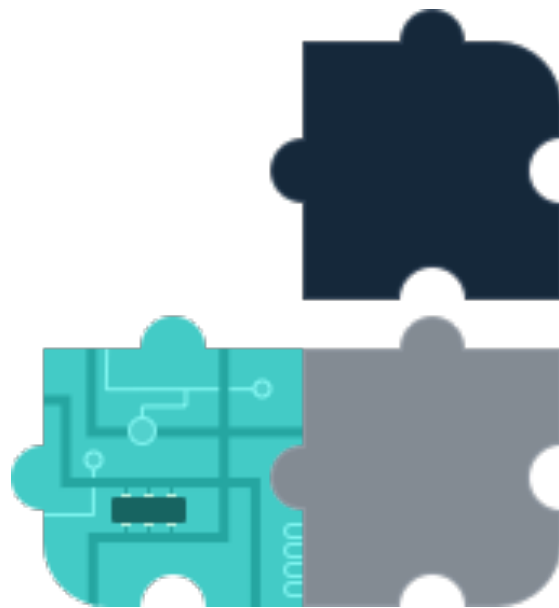
可视化引擎 G2

3

G2 图表模块简介

G2 的图形语法

一张图表就是从数据到几何标记对象的图形属性的一个映射，此外图形中还可能包含数据的统计变换，最后绘制在某个特定的坐标系中。



Data, 数据

 <https://g2.alipay.com/tutorial/data.html>

```
var data = [  
  {action: '访问', visitor: 500},  
  {action: '浏览', visitor: 400},  
  {action: '交互', visitor: 300},  
  {action: '下单', visitor: 200},  
  {action: '付款', visitor: 100}  
];
```

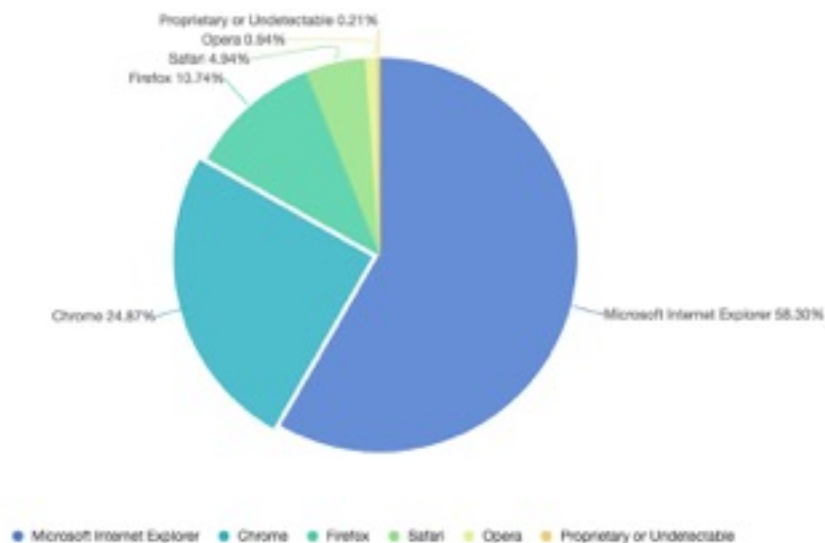
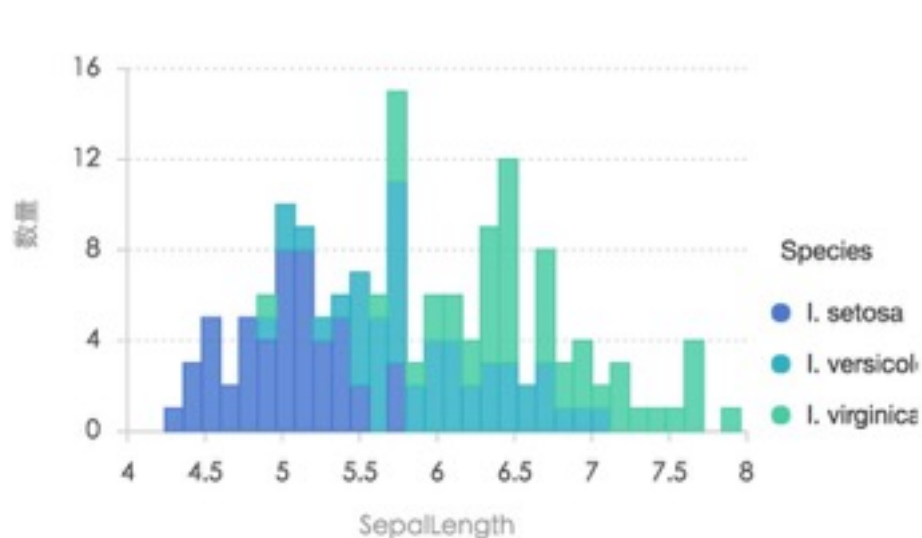
JSON 数组



Stat, 统计

<https://g2.alipay.com/tutorial/stat.html>

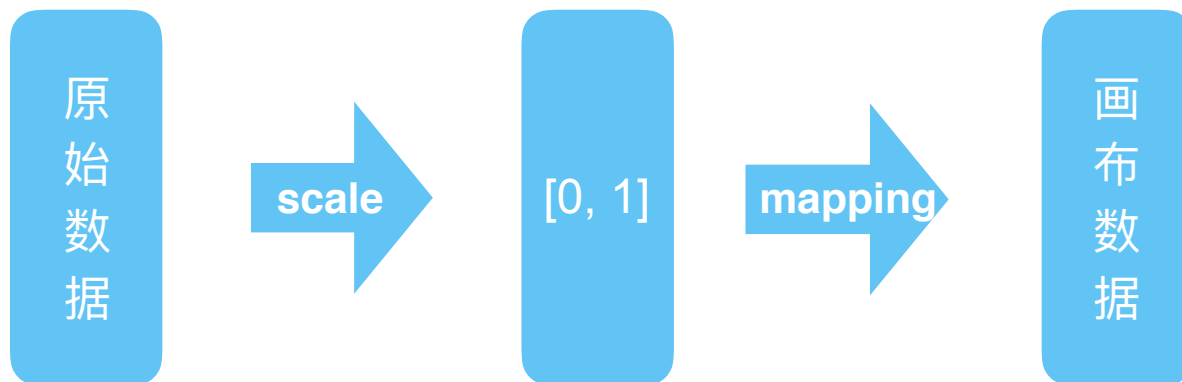
统计变换是对数据进行的某种汇总，是可选的，但通常非常有用。



```
chart.intervalStack().position(Stat.summary.proportion()).color('cut');
```

Scale, 数据度量

<https://g2.alipay.com/tutorial/scale.html>



根据不同的数据类型定义对应的度量：

- identity
- linear pow log time
- cat timeCat



Attr, 图形属性

 <https://g2.alipay.com/tutorial/attr.html>

图形属性对应视觉编码中的视觉通道。

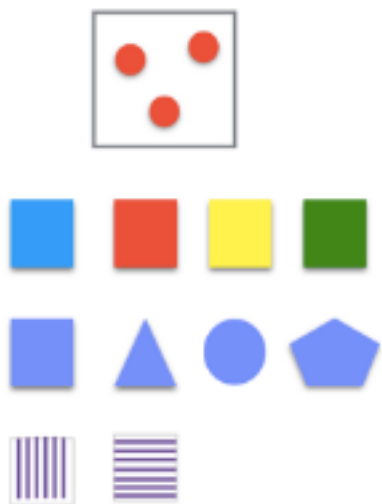
那么，如何将变量映射到图形属性上？



视觉通道的表现力

强
到
弱

分类类型



位置

颜色

形状

纹理

连续/有序类型



坐标轴位置



长度



角度

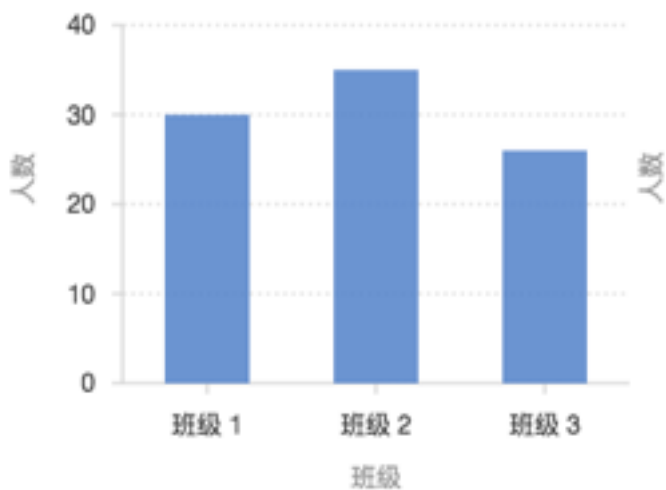


面积



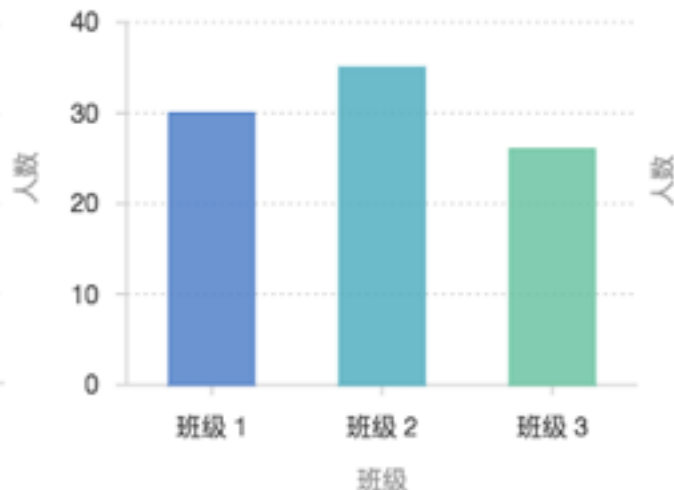
亮度/饱和度

数据和视觉通道的映射



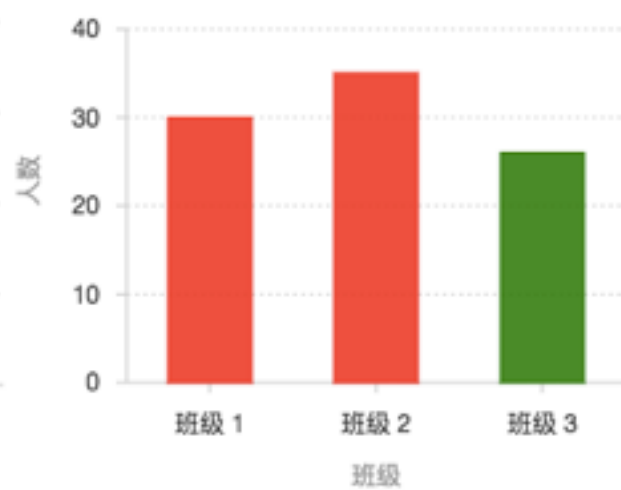
1: 1

一个数据字段对应一个视觉通道



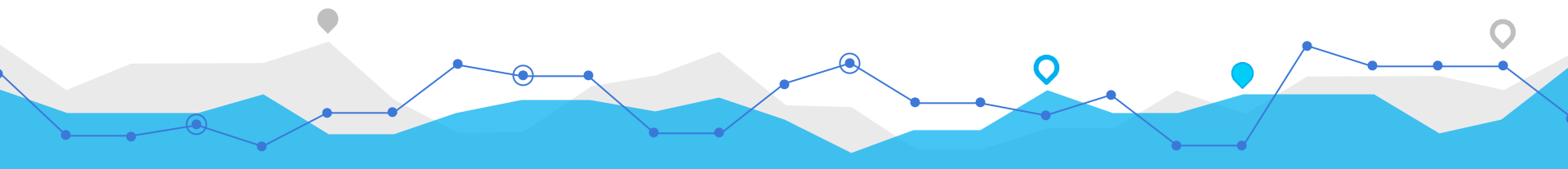
1: n

一个数据字段对应多个视觉通道



n: 1

多个数据字段对应一个视觉通道



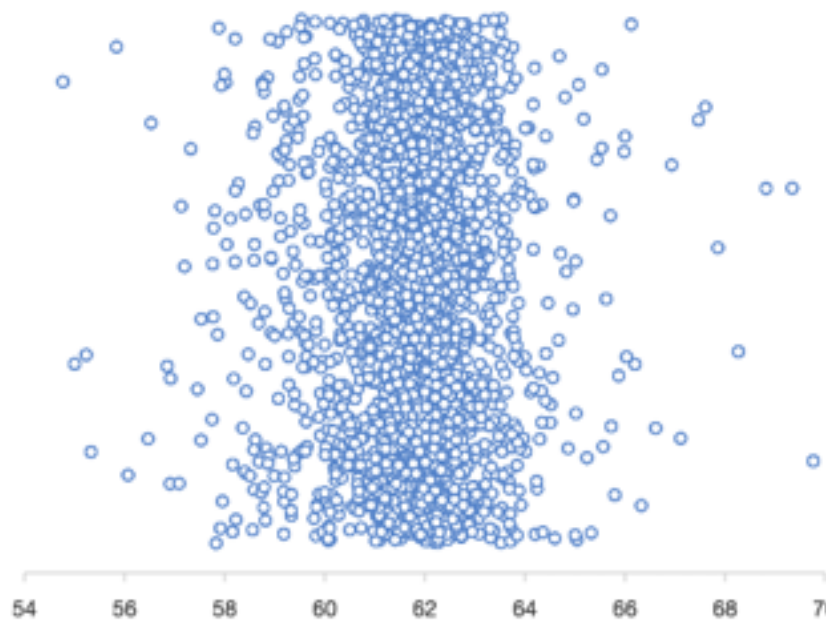
G2 使用的视觉通道

position	位置, <code>chart.point().position('x*y')</code>
color	颜色, <code>chart.interval().position('x*y').color('x');</code>
size	大小, <code>chart.point().position('x*y').size('z');</code>
shape	形状, <code>chart.point().position('x*y').shape('square');</code>
opacity	透明度, <code>chart.interval().position('x*y').opacity(0.8);</code>

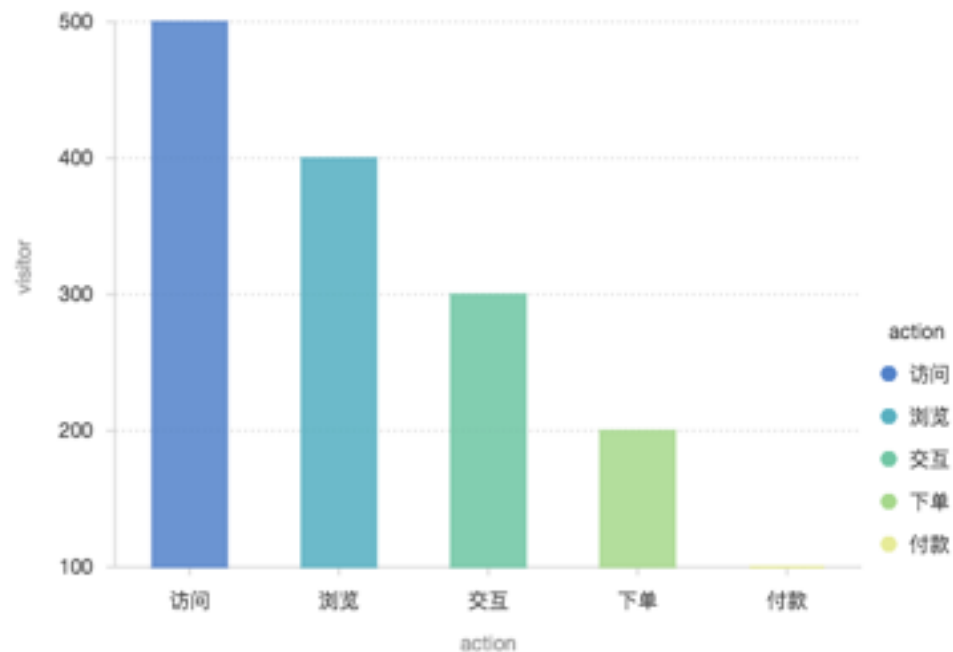


Position 位置

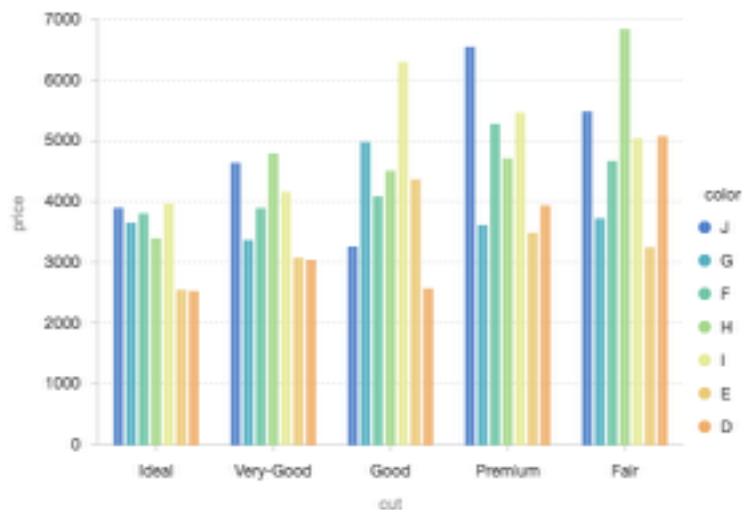
1. position('dim')



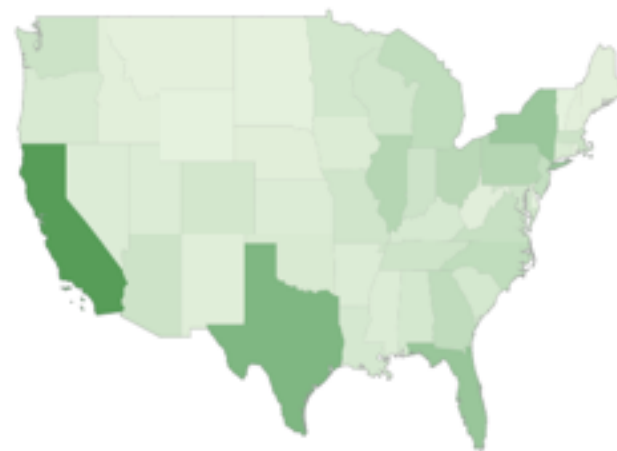
2. position('dim1*dim2')



Color 颜色



分类数据的颜色映射



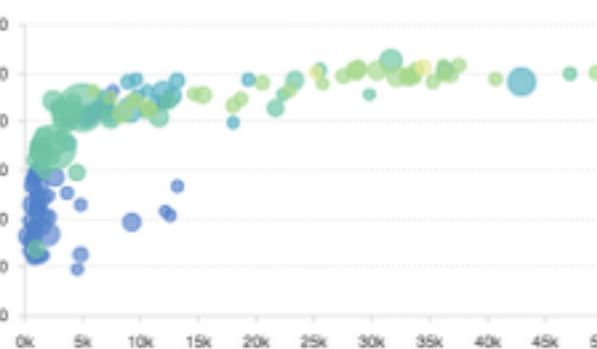
定量（连续）数据的颜色映射：

PS: opacity 透明度在视觉编码过程中，只能进行定量（连续）数据的映射，作为颜色的一个补充使用。

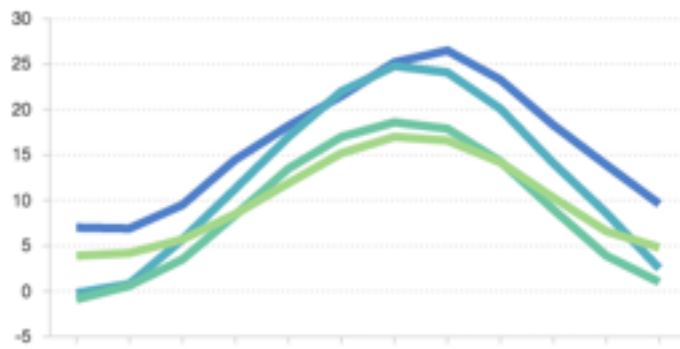


Size 大小

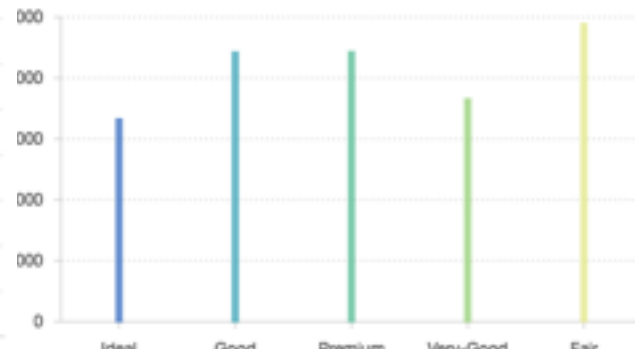
在这里，不同的视觉标记对应不同的视觉表现。



点（半径）



线（粗细）

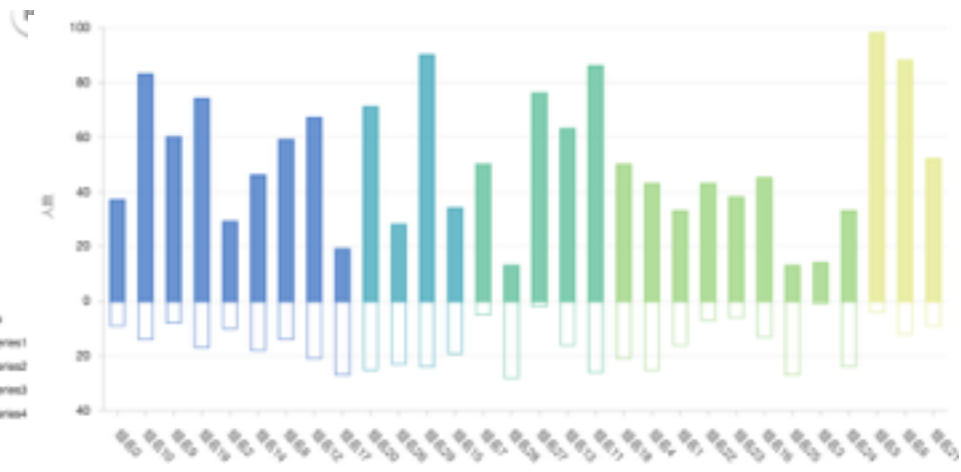
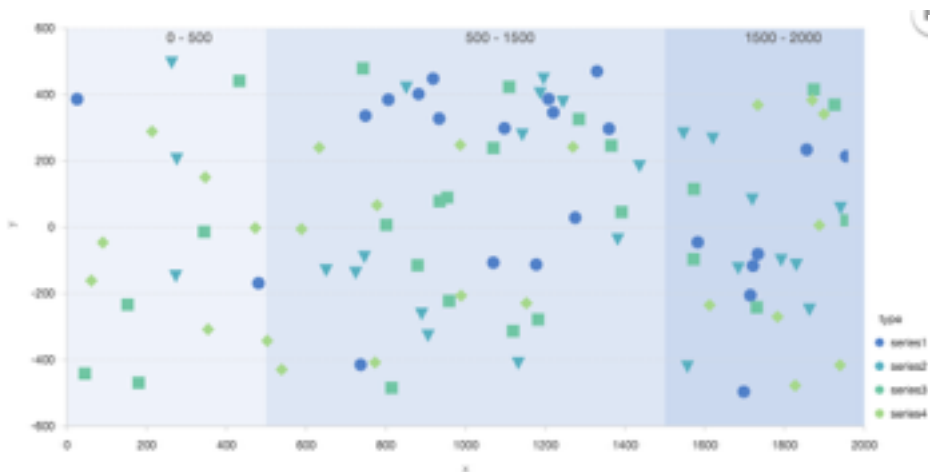


柱状图（宽度）



Shape 形状







不同的对应不同的 shapes。



Geom, 几何标记

chart.point()

几何标记 Geom 决定了图表类型，而对于每一种图表类型来说，又可以使用不同的形状绘制图形。

Geom 类型	Shape
point	  
line	   
area	 
interval	  
polygon	  
schema	 



Coord, 坐标系

描述了数据是如何映射到图形所在的平面，我们通常使用的是笛卡尔坐标系，但也可以将其变换为其他类型，如极坐标和地图投影。



Geom

+

Shape

+

Coord

共同决定

图表类型





实例演示 4

DEMO 1

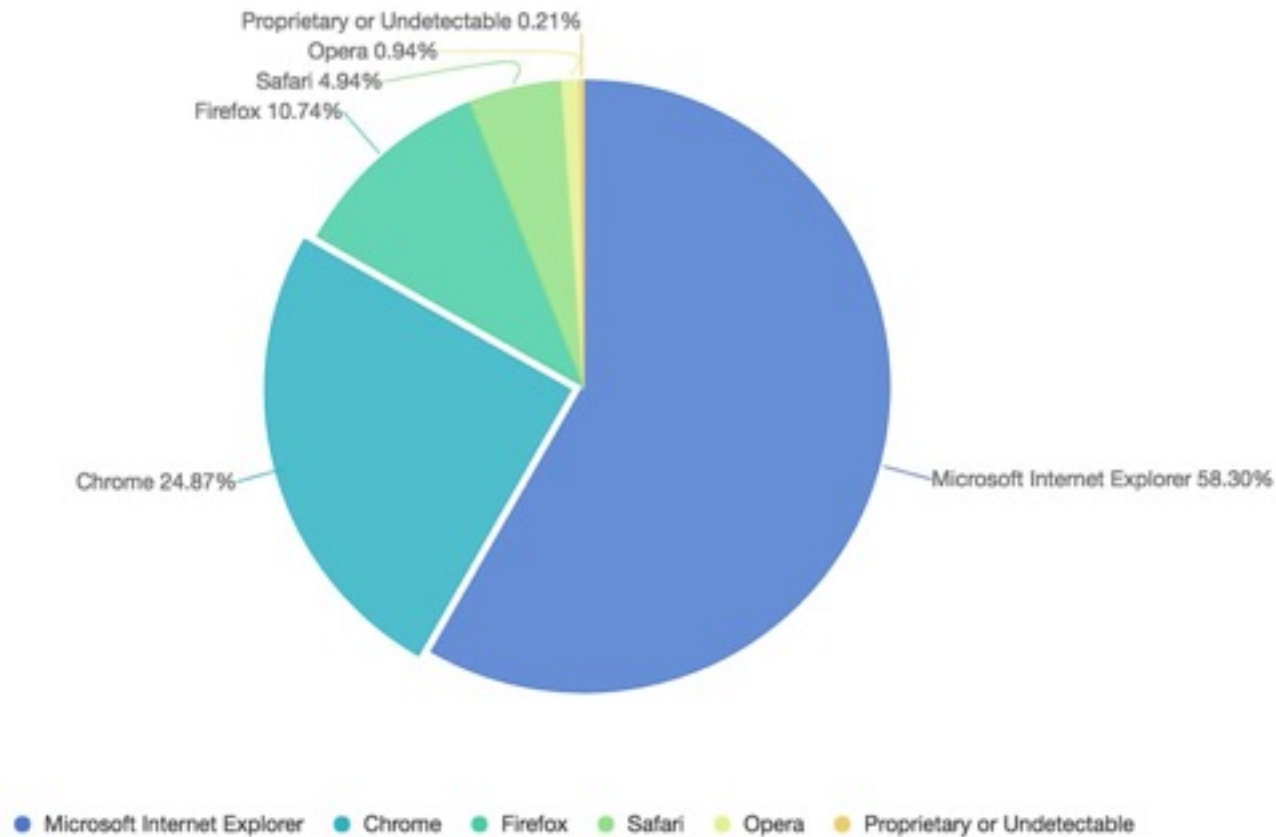
R 语言中的钻石数据

表 2.1 diamonds 数据集。变量 depth、table、x、y 和 z 代表了钻石的各种测量尺度，如图 2.1 所示。

carat	cut	color	clarity	depth	table	price	x	y	z
0.2	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
0.2	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
0.2	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
0.2	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
0.2	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75
0.2	Very Good	J	VVS2	62.8	57.0	336	3.94	3.96	2.48

<http://antvis.github.io/demo.html>

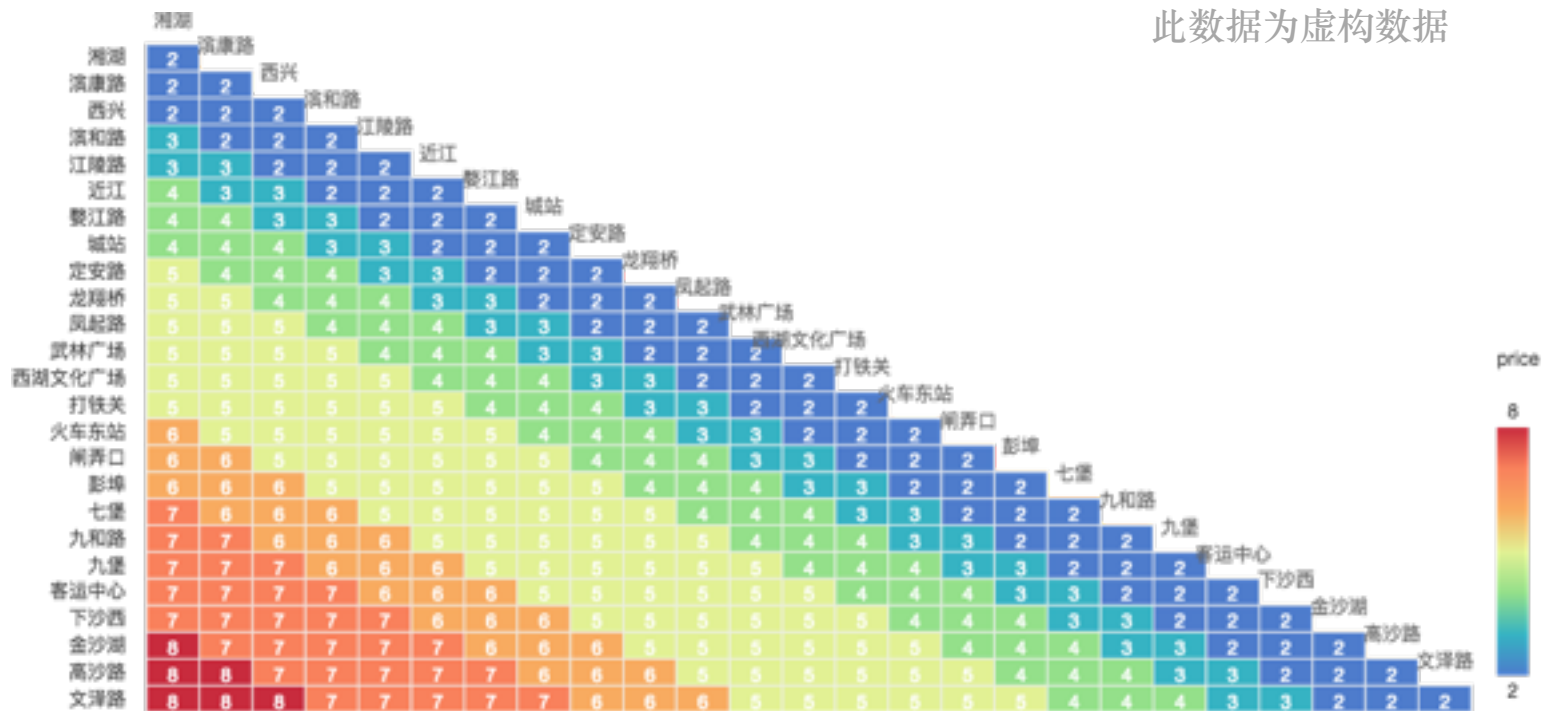
DEMO 2



<https://g2.alipay.com/demo/05-pie/pie.html>

DEMO 3

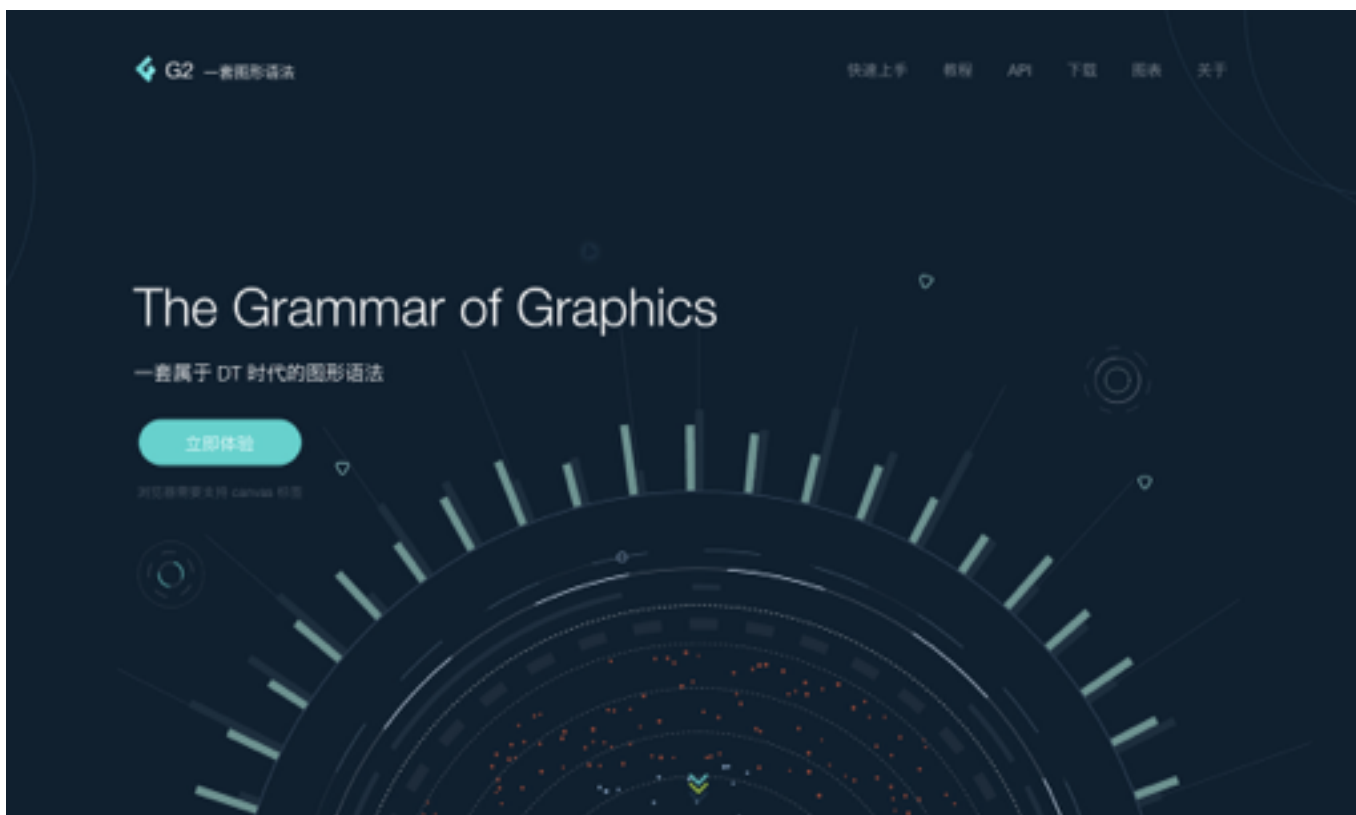
此数据为虚构数据



杭州地铁票价可视化

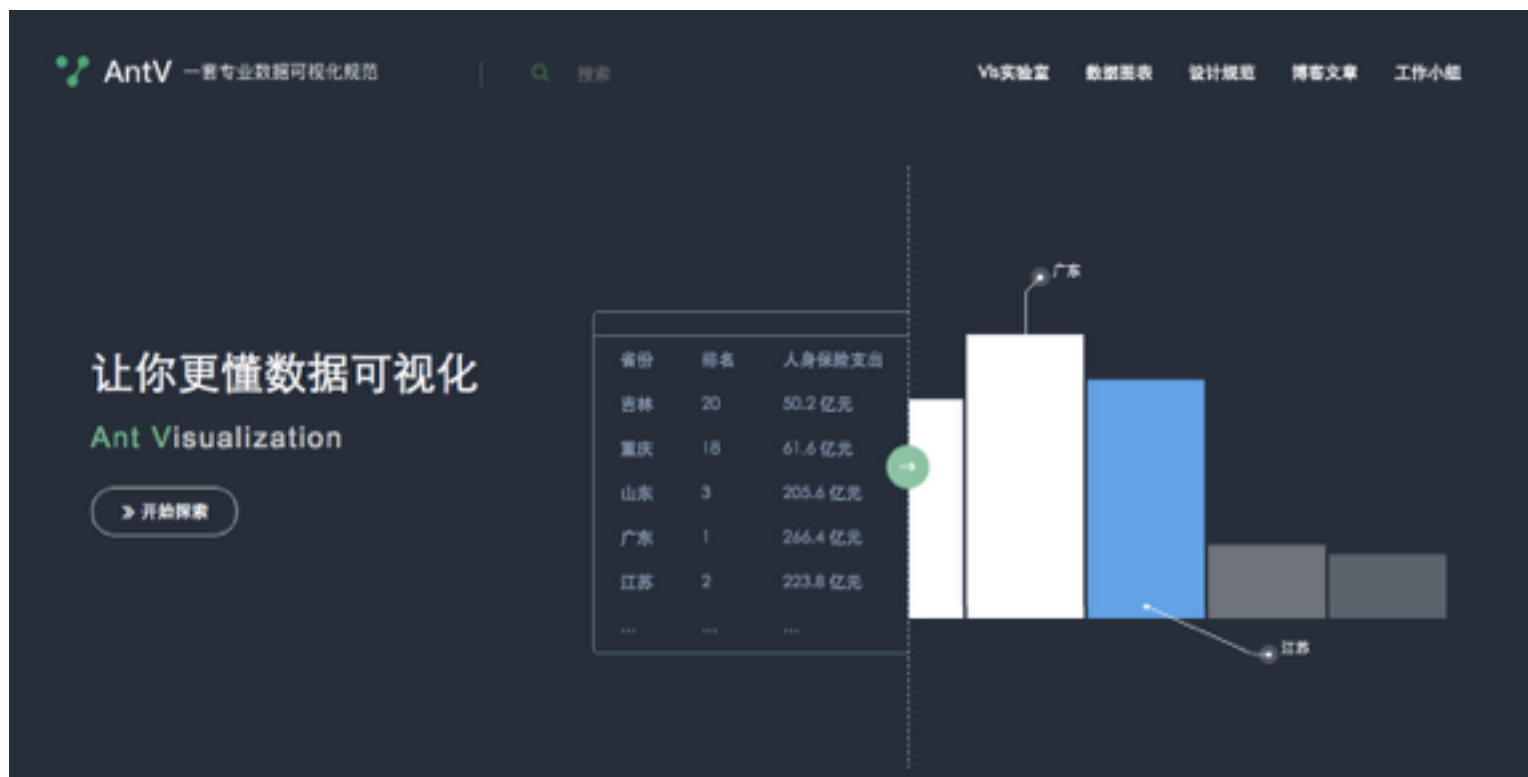
<http://g2.alipay.com/demo/08-heatmap/subway.html>

更多内容请访问 **G2 官网**
<https://g2.alipay.com>



AntV 一套专业数据可视化规范

<https://antv.alipay.net/>



联系我们



<https://github.com/antvis>



<https://github.com/antvis/feedback/issues/new>





总结、展望 4

G2 的目标

以数据为中心

融入统计学知识

专业的可视化引擎



目前的 G2

1.0版

常见图表
数据映射
统计函数
...

1.1版

多坐标系图形
的统一展示
数据流程改造

1.2版

性能优化
统一精度
加法运算符



Roadmap

数据统计结果图表化

使图表可视化更简单，
通过简洁易用的语法服
务于更多更广的用户群
体。

G2 1.x

图形语法、数据映射、多坐标
系统一、简单统计函数、统一
显示精度

数据结果展示化

以 D3 为目标，成为一套
完整的以数据驱动的可
视化工具，并将 PC 与
mobile 端技术统一。

G2 2.x

Shape 扩展、异构数据多
view 支持、交互语法、多图
联动、多端统等

数据分析过程可视化

将技术产品化，打造一
款 web 端的 ggplot2，贯
通数据分析过程的可视
化产品，今年起步。

G2 3.x

数据分析、数据统计、可视
步骤、产品化等



谢谢

联系方式

团队：@王保平 / yubo@alipay.com

个人：@章伟星 / zaifei.zwx@alibaba-inc.com

