数据可视化之美：经典案例与实践解析

随着DT时代的到来，传统的统计图表很难对复杂数据进行直观地展示。这几年数据可视化作为一个新研究领域也变得越来越火。成功的可视化，如果做得漂亮，虽表面简单却富含深意，可以让观测者一眼就能洞察事实并产生新的理解。可视化(visualization)和可视效果(visual)两个词是等价的，表示所有结构化的信息表现方式，包括图形、图表、示意图、地图、故事情节图以及不是很正式的结构化插图。

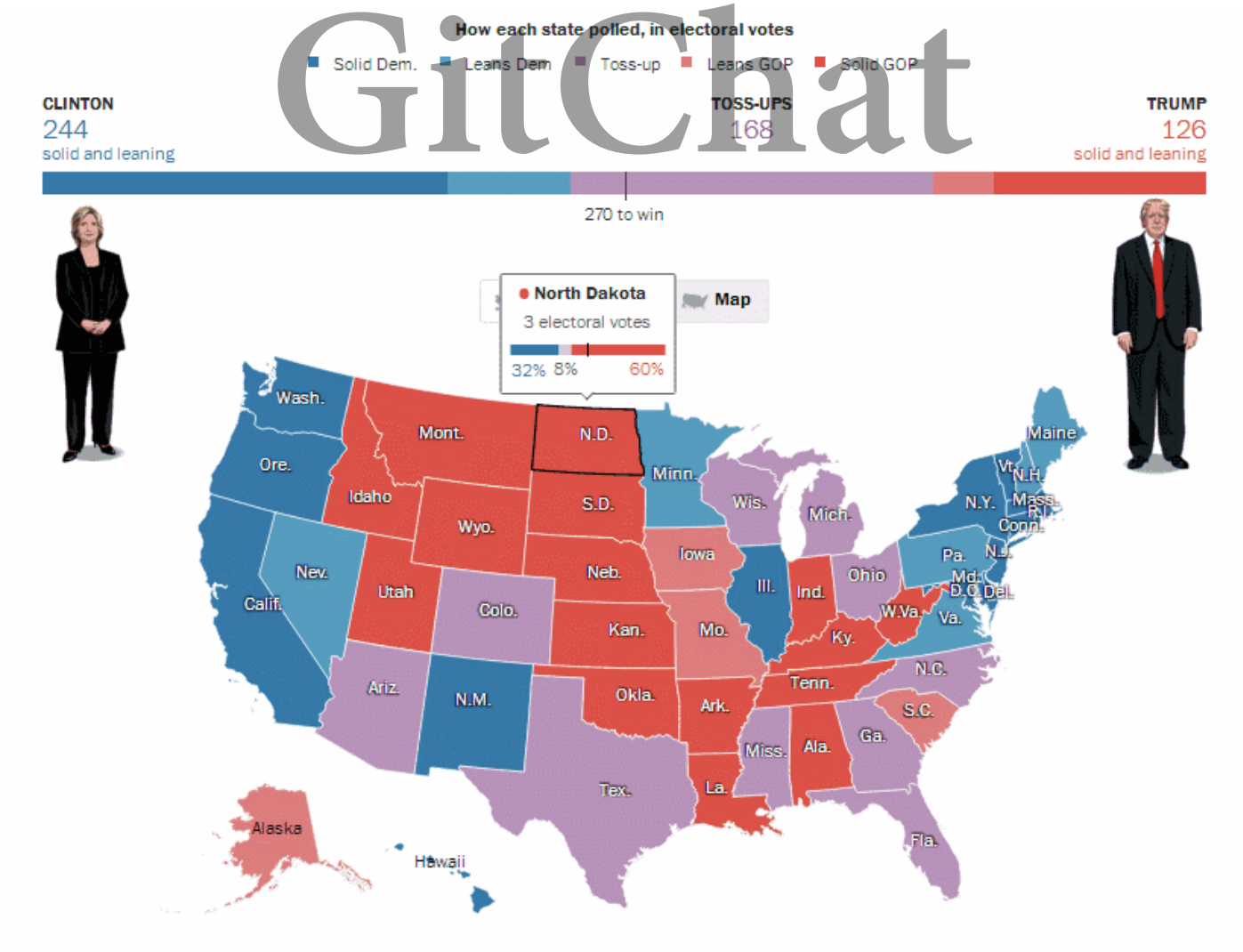
基本的可视化展现方式，如条形图、折线图、饼图、雷达图可以很容易通过各种软件（如Excel）容易生成，这些方法是常见可视化问题的良好且强大的解决方案。然而，使用这些方法的最佳方式局限于一些特定的数据类型，而且其标准型和普遍性意味着它们基本无法达到新颖性。如果对地理空间数据、社会网络关系、多维数据进行可视化，直观地传递数据期望表达的信息是需要特定的图表类型来展示。

让我们一起来看几个经典的可视化，观测它们是如何充分利用其源数据结构的。

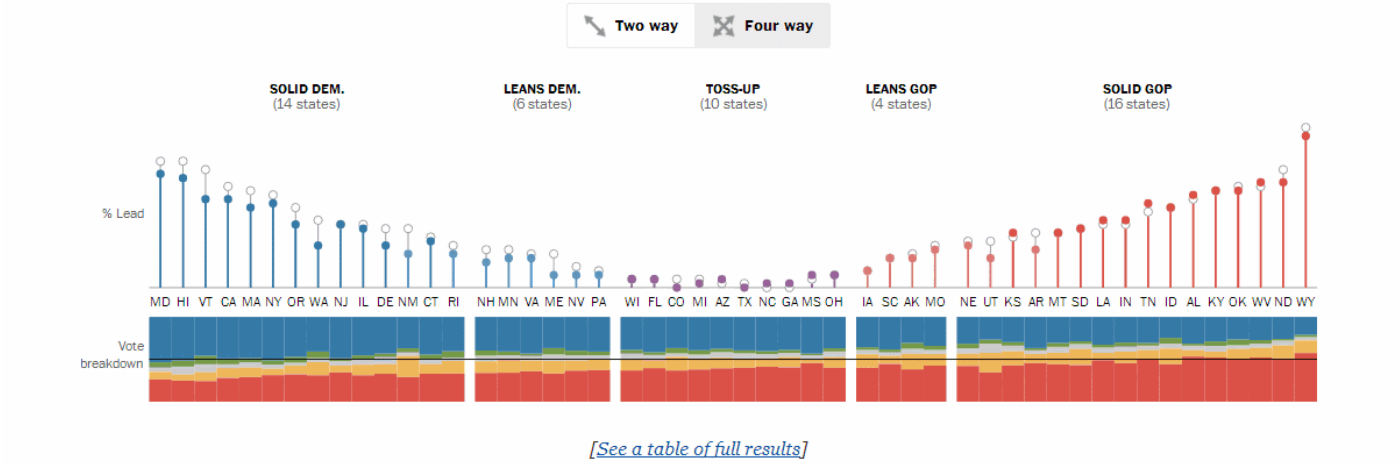
1. “美国大选”数据可视化

在美国大选期间，美国媒体做了不少与之相关的数据报道，让我们来回顾一下，他们是如何将美国大选的数据可视化的吧!

下图为各洲“选举人票”的占比情况。作者设计了两种表现方法，一是以“选举人票”的分布做为底图，一是直接以美国地图作为底图。除此图上方双方选举人票总体数量对比外，鼠标移至各洲上方还能显示各洲“选举人票”数量及对希拉里与特朗普的支持比例。



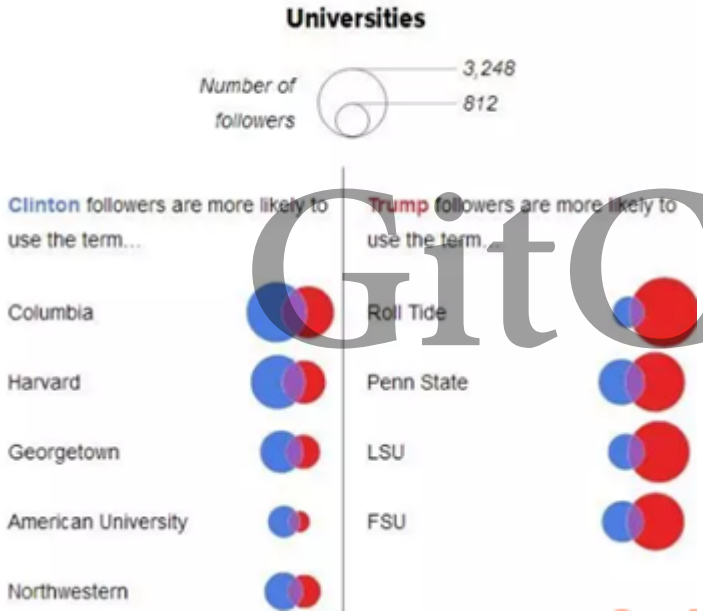
关于第三方选举人影响的情况(在只有两种选择和有四个选举人的情况下，选民的态度又是怎样呢?)



[See a table of full results]

Special Questions

特朗普的粉丝更多的是公立学校出身，而希拉里的则大都为精英阶层。

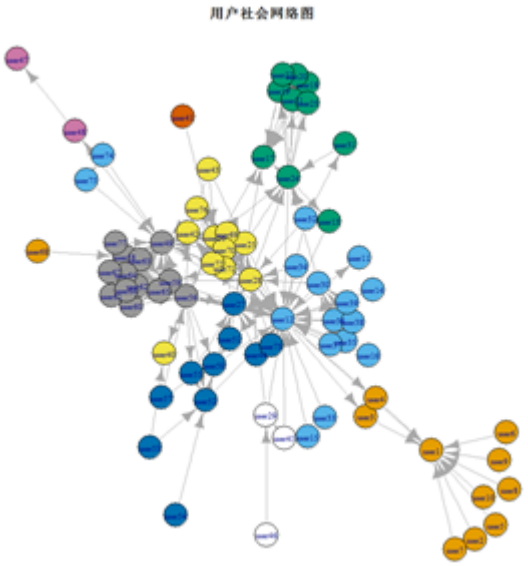


希拉里的粉丝大都较为“书生气”，使用与书籍相关的词汇，其中有很多被认证为教授或博士;而特朗普的粉丝更加喜欢流行文化，他们可能同时是流行歌手的粉丝，也更加关注球类运动等。

2. 社会关系可视化

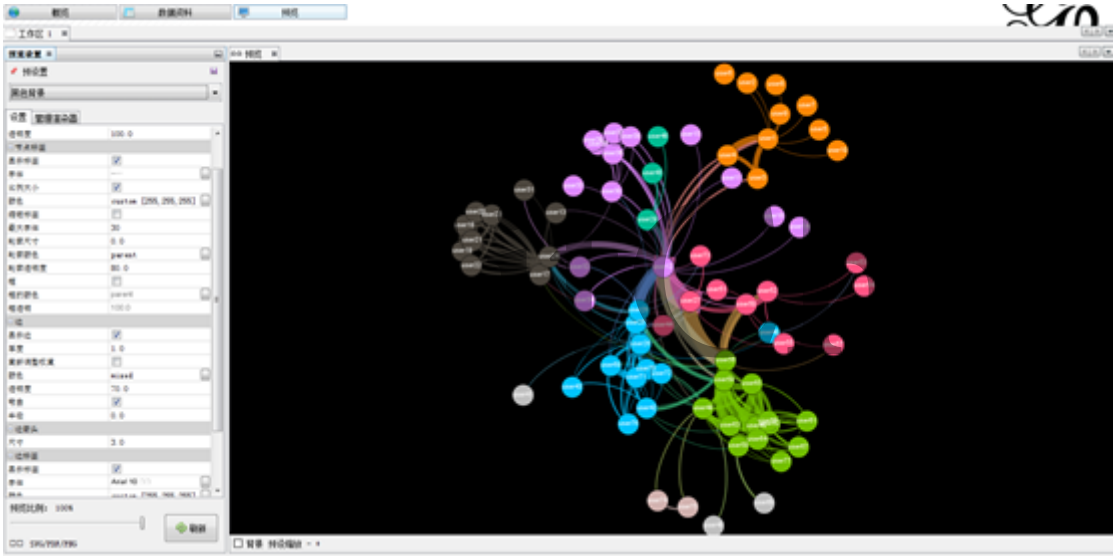
社会网络分析(Social Netwrok Analysis，SNA)是在传统的图与网络的理论之上对社会网络数据进行分析的方法。随着人类进入了移动互联网时代，社会网络数据成了重要的数据资源。SNA的本质是利用各样本间的关系来分析整体样本的群落现象，并分析样本点在群落形成中的作用以及群落间的关系。

近几年手机端网游越来越重视游戏用户社交性设计。这款游戏的玩法设计特别强调强社交性：用户可以在游戏内组建家族，家族成员有不同的职务等级，用户也可以在游戏内给好友赠送道具。我们从数据库中收集抽取了部分用户的家族数据(Nodes)和好友沟通数据(Links)。其中Nodes数据集包括Id(用户ID)、Label(用户名称)、Group(所属家族)、Level(等级)的信息；Links数据集包括Source(发起方)、Target(接收方)和Weight(斗气数量)信息。



从网络图可以看出，不同家族的成员基本紧密联系在一起，并通过一些关键成员与其他家族成员联系。例如我们发现右下角的那个社群的成员先通过user1用户、再通过user12用户跟其他社团成员联系在一个大网络图中。

我们也可以用Gephi软件快速绘制社会网络图，并对其进行美化。



3. 地理信息可视化

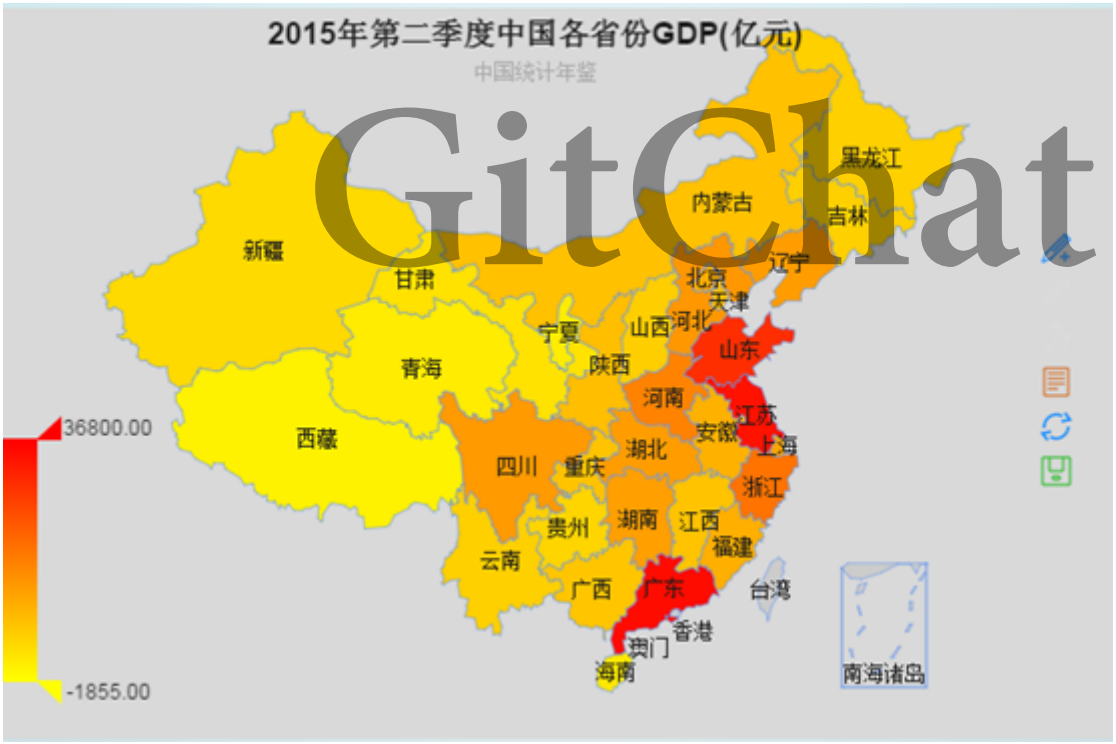
在第一个例子中，我们已经见识到了地理信息可视化的魅力。接下来我们简单了解下如何利用Remap包快速绘制可交互的地图数据可视化。目前托管在github，https://github.com/lchiffon/REmap。

百度迁徙图是近年来非常流行的一种地理信息可视化，可以通过连线动态查看人口流向。此处给大家绘制一幅动态航班图的地理信息可视化图，大家点击链接可查看动态效果。

也可以利用Remap快速实现未来天气预报。

还可以把一些各地举行的会议事件在地图上进行可视化展示，下图是2015年中国R语言会议在各个城市举行的可视化展示（动图链接）。

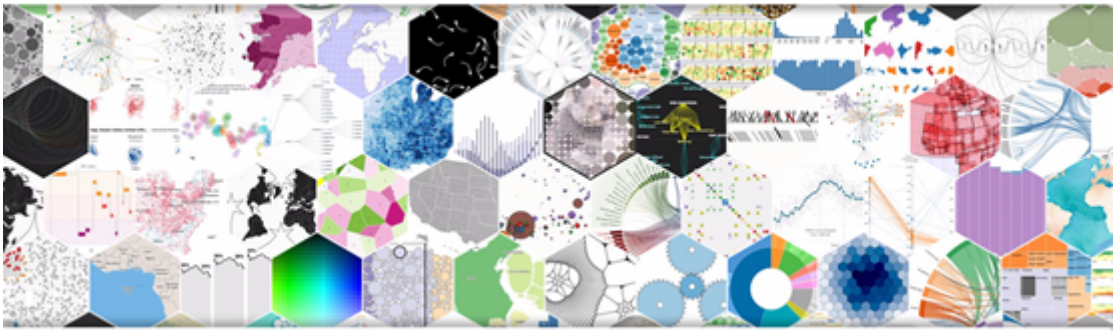
城市热力图也是近年来非常流行的一种地理信息可视化方式，通过颜色的深浅表示不同地区的实际数值大小（动图链接）。



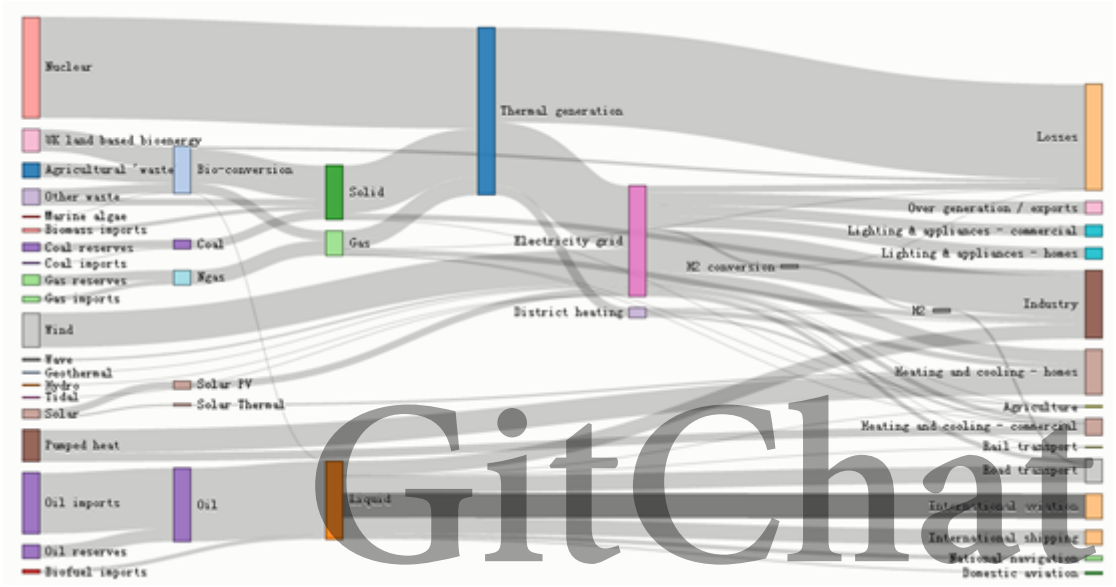
通过以上的几个小例子，相信大家已经惊叹于上面的可视化效果，给人眼前一亮、耳目一新的感觉。以上可视化并未运用到很高深的技术，如果你也掌握以下一些可视化知识，也能绘制出以上图表的效果。接下来，就给大家介绍几个常用的交互数据可视化手段：D3、Echarts和R（R是一款数据分析挖掘软件，但是其拥有强大的可视化功能，并能集成D3、Echarts图库，实现交互绘图）。

1. D3（https://d3js.org/）

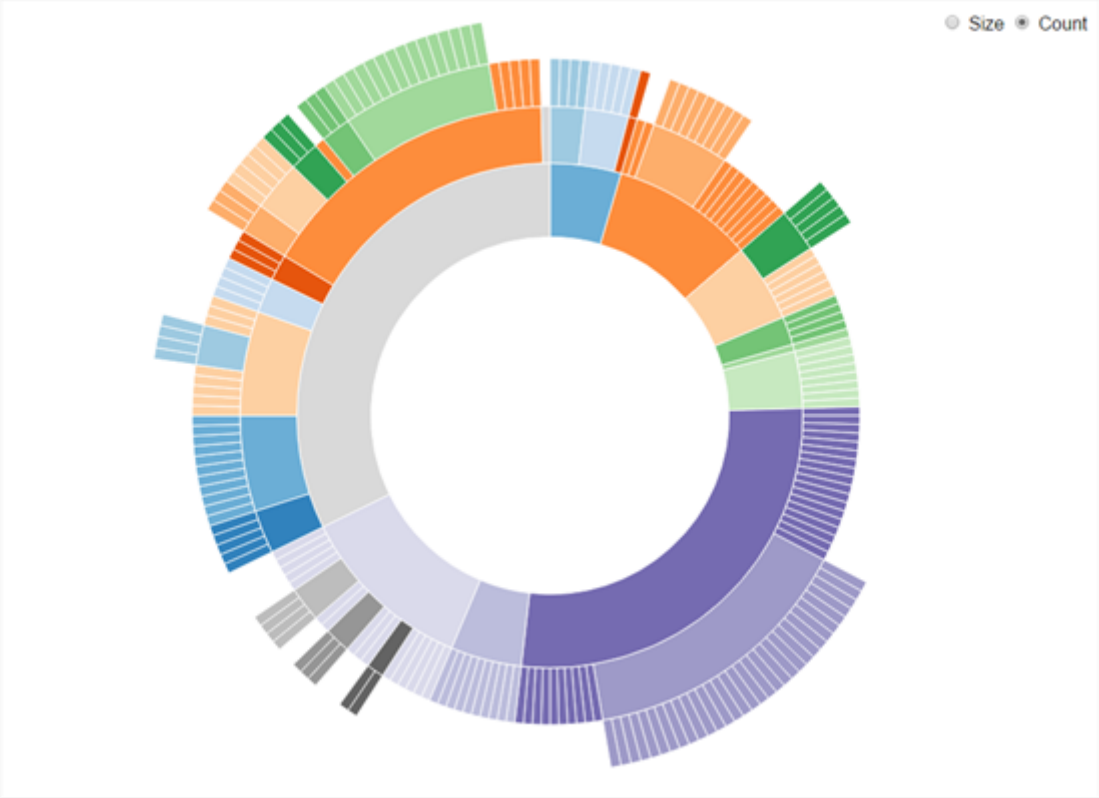
D3 是最流行的可视化库之一，它被很多其他的表格插件所使用。它允许绑定任意数据到DOM，然后将数据驱动转换应用到Document中。你可以使用它用一个数组创建基本的HTML表格，或是利用它的流体过度和交互，用相似的数据创建惊人的SVG条形图。



比如D3可以非常容易地绘制交互桑基图。桑基图（Sankey diagram），即桑基能量分流图，也叫桑基能量平衡图。它是一种特定类型的流程图，图中延伸的分支的宽度对应数据流量的大小，通常应用于能源、材料成分、金融等数据的可视化分析。桑基图最明显的特征就是，始末端的分支宽度总和相等，即所有主支宽度的总和应与所有分出去的分支宽度的总和相等，保持能量的平衡（动图链接）。



可以通过D3对Sunburst Partition可视化探索。通过解析布点获得的用户行为路径数据，我们可以用最简单与直接的方式将每个用户的事件路径点击流数据进行统计，并用数据可视化方法将其直观地呈现出来。 D3.js是当前最流行的数据可视化库之一，我们可以利用其中的Sunburst Partition来刻画用户群体的事件路径点击状况。从该图的圆心出发，层层向外推进，代表了用户从开始使用产品到离开的整个行为统计；Sunburst事件路径图可以快速定位用户的主流使用路径。通过提取特定人群或特定模块之间的路径数据，并使用Sunburst事件路径图进行分析，可以定位到更深层次的问题。灵活使用Sunburst路径统计图，是我们在路径分析中的一大法宝（动图链接）。



2.ECharts（http://echarts.baidu.com/）

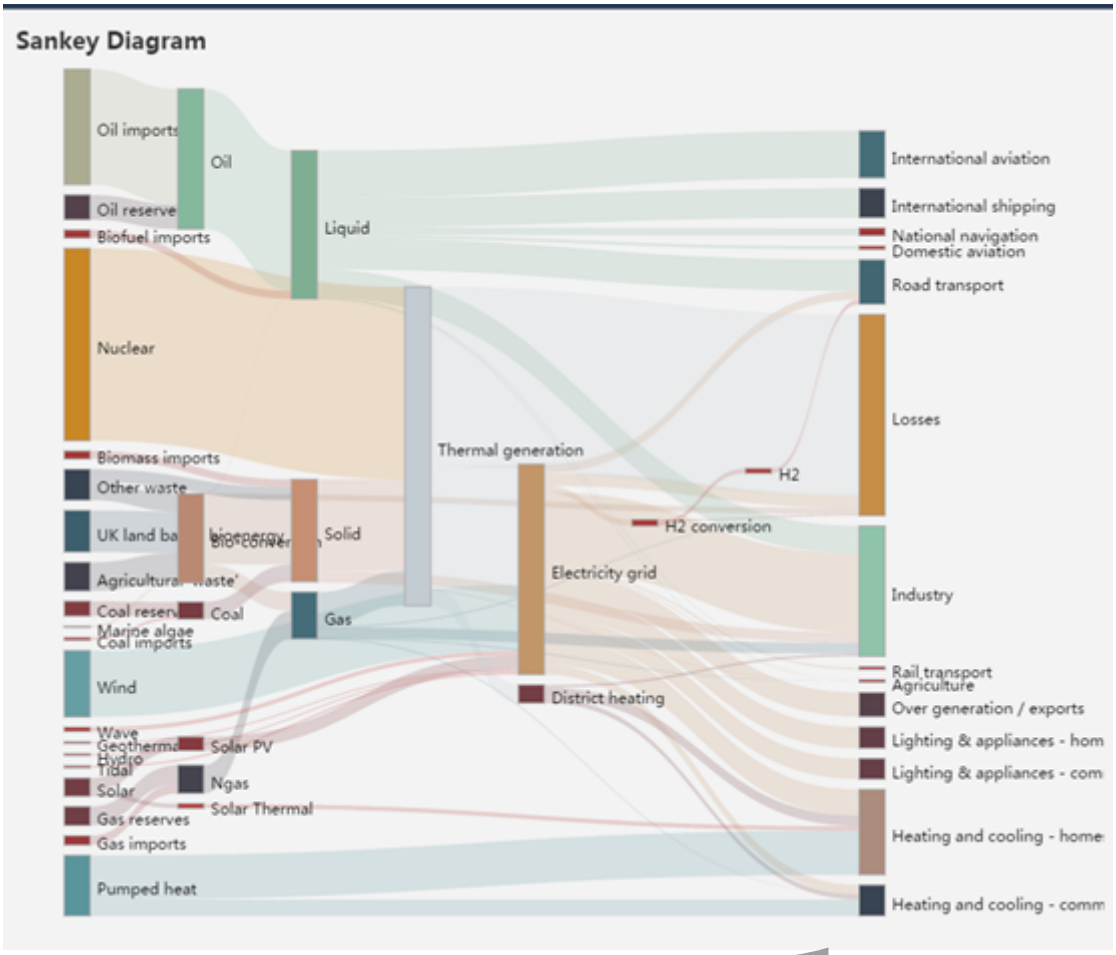
ECharts，缩写来自Enterprise Charts，商业级数据图表，一个纯Javascript的图表库，可以流畅的运行在PC和移动设备上，兼容当前绝大部分浏览器（IE6/7/8/9/10/11，chrome，firefox，Safari等），底层依赖轻量级的Canvas类库ZRender，提供直观，生动，可交互，可高度个性化定制的数据可视化图表。创新的拖拽重计算、数据视图、值域漫游等特性大大增强了用户体验，赋予了用户对数据进行挖掘、整合的能力。

ECharts 3还新增更多图表类型，更好的满足不同数据的处理需求 更多的搭配方案让你的数据呈现方式更个性和完美。

比如地图信息可视化：

动图链接

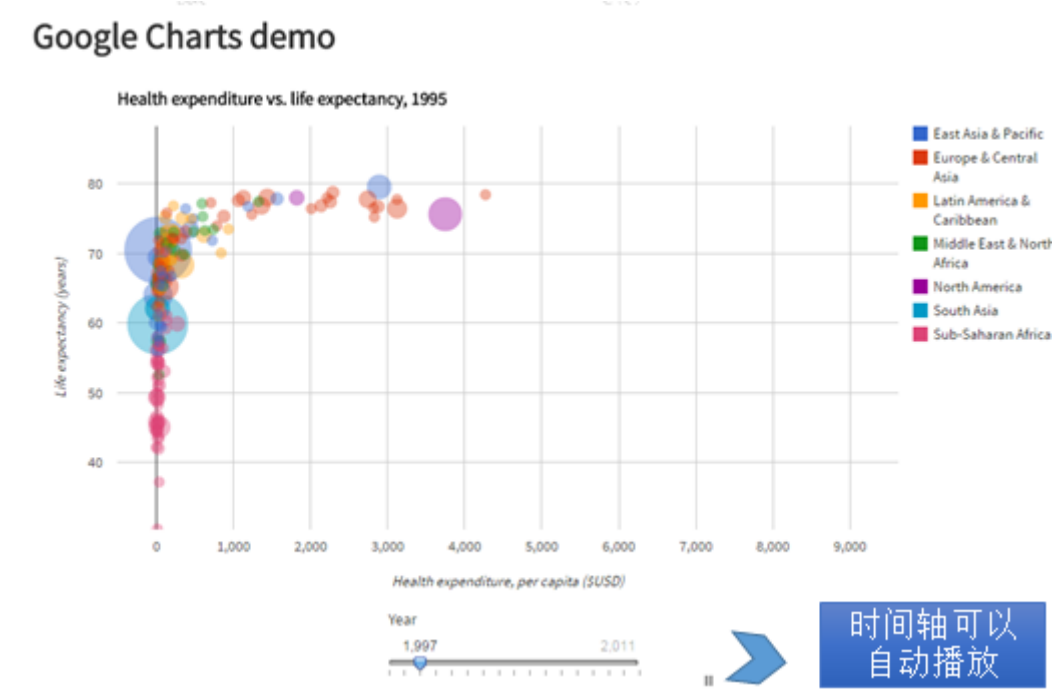
利用ECharts绘制桑基图。



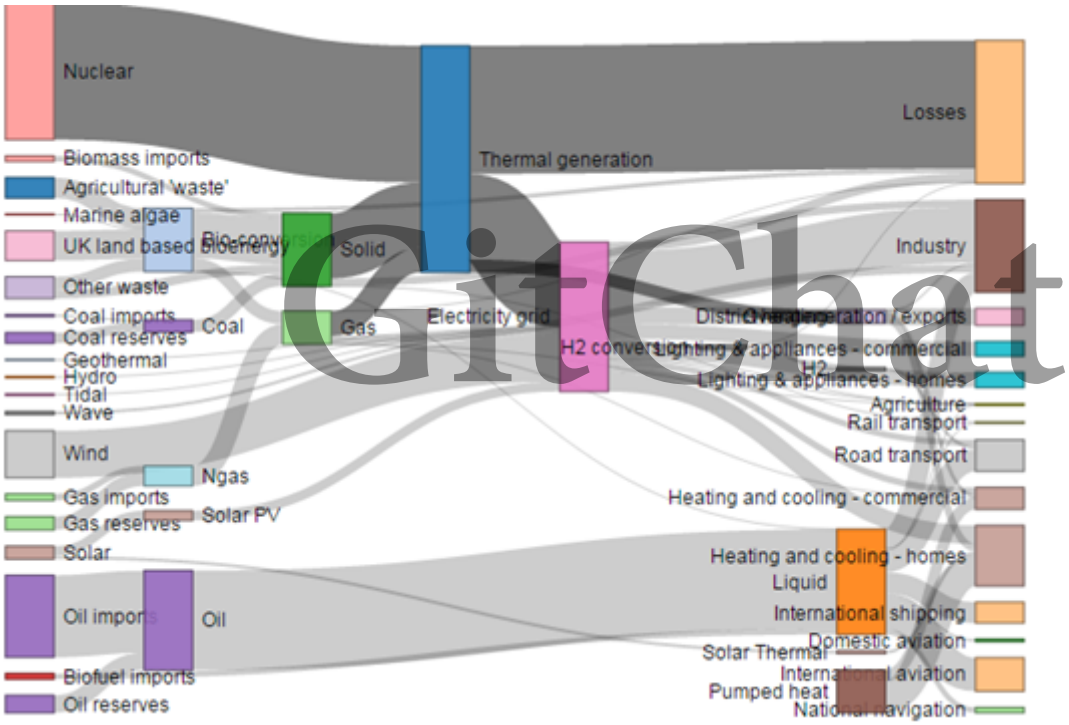
3. R（https://www.r-project.org/）

R语言是一套开源的数据分析解决方案，几乎可以独立完成数据处理、数据可视化、数据建模及模型评估等工作，而且可以完美配合其他工具进行数据交互。5) R语言拥有顶尖的制图功能。不仅有 lattcie包、ggplot2包对复杂数据进行可视化，更有rCharts包、recharts包、plotly包实现数据交互可视化，甚至可以利用功能强大的shiny包实现R与web整合部署，构建网页应用，帮助不懂CSS，HTML的用户能利用R快速搭建自己的数据分析APP应用。

比如我们可以绘制动态交互的气泡图，通过下面的时间轴播放动态查看不同年份的气泡情况。



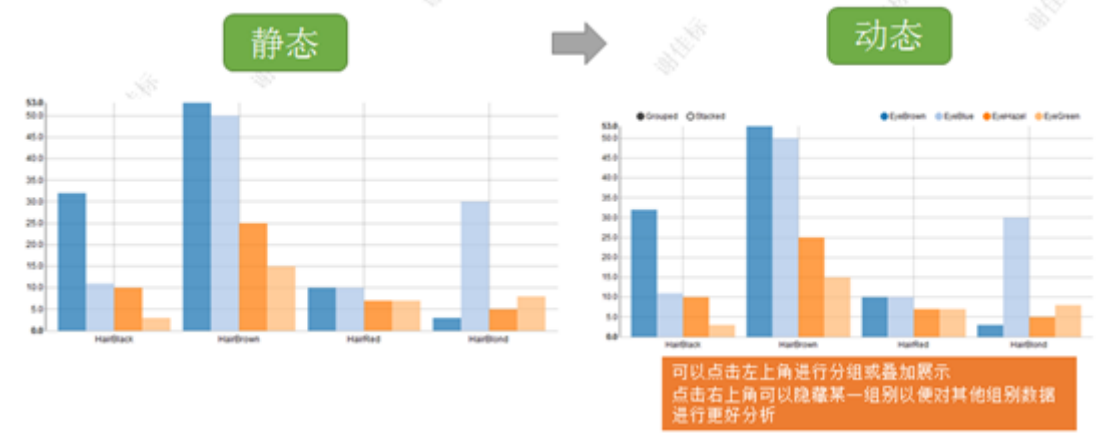
也可以利用networkD3包调用D3.js库，绘制社会网络图和桑基图。



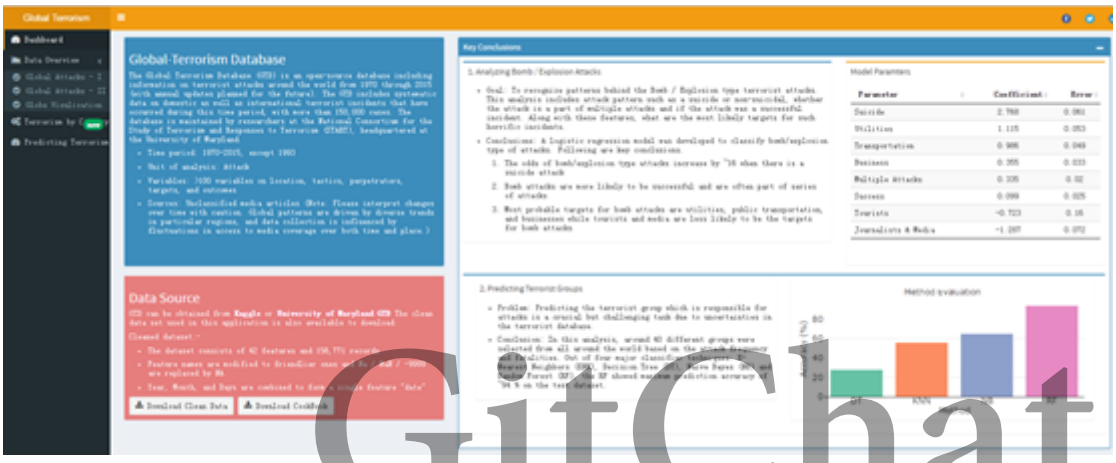
前文我们已经了解了几种常用的数据可视化技术。接下来，让我们一起来学习下创建有效的可视化的步骤。我们通常会按照下述的几个关键步骤进行：

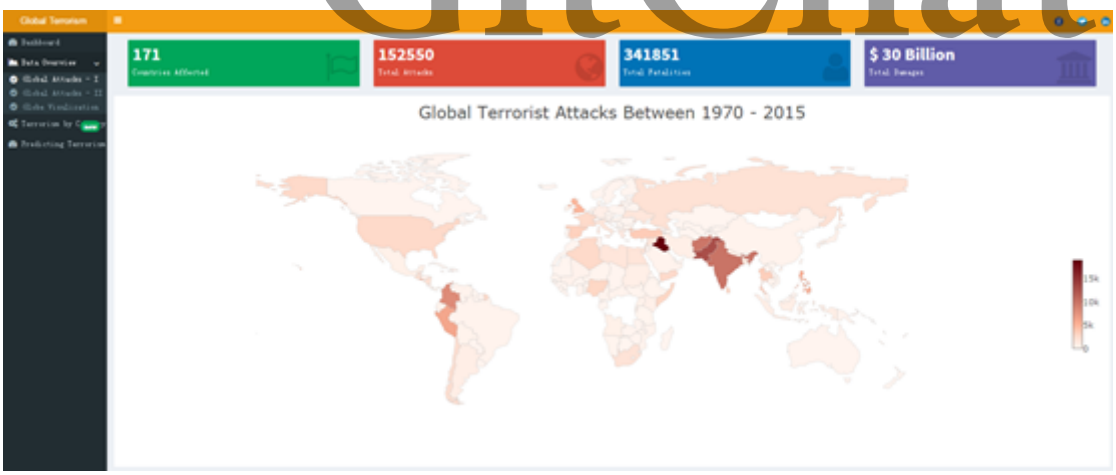
1. 你有什么数据？数据有哪些分类？
2. 关于数据你想了解什么？
3. 应该使用哪种可视化方式？
4. 能够进行可视化的工具有哪些？
5. 透过可视化你看见了什么，有什么意义？

最后，复杂高维数据无法用单一的静态图表进行直观地展示，因此需要借助可视化手段让数据动起来，更好地发现数据价值。比如说有不同组别的数据，我们想查看各组别间的数据和总计时，此时就可以通过交互式探索的形式进行展示。



还可以结合自己掌握的数据分析和可视化技术，搭建数据可视化平台，从而实现智能BI的可视化功能。比如说，我们不需要具备开发能力，利用R工具的shiny包可以快速搭建数据可视化原型。下面这个例子（https://www.showmeshiny.com/global-terrorism/）就是一个通过shiny包结合可视化技术实现的一个可视化平台。






Global Terrorism
Query Builder
North America
873
3268
4674
$12 Billion
Attacks
Targets
Weapons
Explosion
Facility Attack
Armed Assault
Assassination
Kidnapping
Private property
Government
Public services
Hate crimes
Bombs
Incendiary
Firearms
Unknown
Melee
Major Attacks (Click on points for more information)
Major Terrorist Groups
Group
Fatalities