课程总结报告

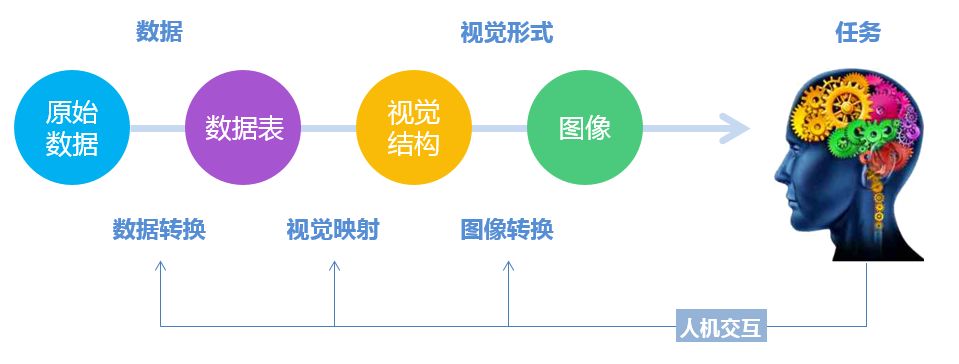
计科2002 赵梓洋 2007010230

github账号zzy1956142723

我与小组成员的小组课题《大数据可视化》演讲之后受到了老师的鼓励与批改，现总结报告如下

可视化的概念：可视化（Visualization）是利用计算机图形学和图形处理技术，将数据转换成图形或图像在屏幕上显示出来，并进行交互处理的理论、方法和技术。它涉及到计算机图形学、图像处理、[计算机视觉](https://baike.baidu.com/item/%E8%AE%A1%E7%AE%97%E6%9C%BA%E8%A7%86%E8%A7%89" \t "https://baike.baidu.com/item/%E5%8F%AF%E8%A7%86%E5%8C%96/_blank)、  [计算机辅助设计](https://baike.baidu.com/item/%E8%AE%A1%E7%AE%97%E6%9C%BA%E8%BE%85%E5%8A%A9%E8%AE%BE%E8%AE%A1" \t "https://baike.baidu.com/item/%E5%8F%AF%E8%A7%86%E5%8C%96/_blank)等多个领域，成为研究[数据表示](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E8%A1%A8%E7%A4%BA" \t "https://baike.baidu.com/item/%E5%8F%AF%E8%A7%86%E5%8C%96/_blank)、数据处理、决策分析等一系列问题的综合技术。目前正在飞速发展的虚拟现实技术也是以图形图像的可视化技术为依托的。

上一步的主要工作是为数据可视化配置好了数据，接下来便可根据业务的需要来绘制不同的图，如折线图/柱状图/散点图等等。具体的实现方法在后面的章节中会细致讲解，这里重点提一下绘图函数里的stat参数。这个参数是对冲突样本点做统计，该参数默认为identity，表示保留样本点原(y)值，还可以是sum，表示对出现在这点的(y)值进行求和等等。





数据采集要大、全、细、时

大 充分考虑用户规模与数据规模的增长，做好数据资产积累的准备。

全 多种数据源，多种方法全量采集，贯穿用户使用产品的整个生命周期。

细 采集足够全面的属性、维度、指标，让积累的数据资产更加优质。

时 提高数据采集的时效性，从而提高后续数据应用的时效性。

# 根据孙主任的反馈，**如何将枯燥的大数据呈现为可视化的图和动画？**

数据的可视化也是一样，组成一幅内容清晰、表达力强、美观的可视化信息图的也仅仅是一些基本的元素，这些元素的不同组合却可以产生出令人着迷的力量。

要列出“可视化元素之根”很容易：位置、长度、角度、形状、纹理、面积（体积）、色相、饱和度等几种有限的元素，邱南森在他的[《数据之美》](https://link.zhihu.com/?target=https://book.douban.com/subject/25833225/" \t "_blank)中提供了一张视觉元素的图，其中包含了大部分常用的元素。

这个现象由人类的大脑构造而固有，因此在设计可视化作品时也应该充分考虑，比如：

* 避免使用面积图作为对比
* 在做对比类图形时，当差异不明显时需要考虑采用非线性的视觉元素
* 选用多种颜色作为视觉编码时，差异应该足够大

孙主任的计算科学导论课程让我了解整个计算机的主体框架，同时了解到了时代大背景，目前，计算机技术的飞速发展，使得计算机技术的应用及渗透的领域越来越广，这迫切需求高技能高素质的人才。如何将计算机专业的人才培养成合格的、适应市场需求的人才，成为每一位计算机专业教育工作者的首要任务。%20计算机科学导论作为一门计算机专业大学一年级学生的专业必修课，重点旨在勾画计算机科学体系的框架，通过计算机科学基础理论与应用操作相结合课程内容安排，使学生对计算机学科的知识体系结构有一个较为全面和系统的了解，继而激发并培养自己对本专业的兴趣，为以后大学四年的知识学习、能力素质和职业道德的塑造奠定坚实基础基于课程至关重要的作用，主要从计算机科学导论的教学方法和考核方法等方面进行了讨论，以期在今后的课程教学过程中达到最佳教学效果。

在ggplot2中，图形的绘制是一个个图层添加上去的。举个例子来说，我们首先决定探索一下身高与体重之间的关系；然后画了一个简单的散点图；然后决定最好区分性别，图中点的色彩对应于不同的性别；然后决定最好区分地区，拆成东中西三幅小图；最后决定加入回归直线，直观地看出趋势。这是一个层层推进的结构过程，在每一个推进中，都有额外的信息被加入进来。在使用ggplot2的过程中，上述的每一步都是一个图层，并能够叠加到上一步并可视化展示出来

计算机科学与技术这一门科学深深的吸引着我们这些同学们，原先不管是国内还是国外都喜欢把这个系分为计算机软件理论、计算机系统、计算机技术与应用。后来又合到一起，变成了现在的计算机科学与技术。我一直认为计算机科学与技术这门专业，在本科阶段是不可能切分成计算机科学和计算机技术的，因为计算机科学需要相当多的实践，而实践需要技术；每一个人(包括非计算机专业)，掌握简单的计算机技术都很容易（包括原先Major们自以为得意的程序设计），但计算机专业的优势是：我们掌握许多其他专业并不"深究"的东西，例如，算法，体系结构，等等。非计算机专业的人可以很容易地做一个芯片，写一段程序，但他们做不出计算机专业能够做出来的大型系统。今天我想专门谈一谈计算机科学，并将重点放在计算理论上。  
  
  
1)计算机语言  
 随着20世纪40年代第一台存储程序式通用电子计算机的研制成功，进入20世纪50年代后，计算机的发展步入了实用化的阶段。然而，在最初的应用中，人们普遍感到使用机器指令编制程序不仅效率低下，而且十分别扭，也不利于交流和软件维护，复杂程序查找错误尤其困难，因此，软件开发急需一种高级的类似于自然语言那样的程序设计语言。1952年，第一个程序设计语言Short Code出现。两年后，Fortran问世。作为一种面向科学计算的高级程序设计语言，Fortran的最大功绩在于牢固地树立了高级语言的地位，并使之成为世界通用的程序设计语言。Algol60的诞生是计算机语言的研究成为一门科学的标志。  
  
(2)计算机模型与软件开发方法  
 20世纪80年代是计算机网络、分布式处理和多媒体大发展的时期。在各种高级程序设计语言中增加并发机构以支持分布式程序设计，在语言中通过扩展绘图子程序以支持计算机图形学程序设计成为当时程序设计语言的一种时尚。之后，在模数/数模转换等接口技术和数据库技术的支持下，通过扩展高级语言的程序库又实现了多媒体程序设计的构想。  
(3)计算机应用  
 用计算机来代替人进行计算，就得首先研究计算方法和相应的计算机算法，进而编制计算机程序。由于早期计算机的应用主要集中在科学计算领域，因此，数值计算方法就成为最早的应用数学分支与计算机应用建立了联系。最初的时候，由于计算机的存储器容量很小，速度也不快，为了计算一些稍稍大一点的题目，人们常常要挖空心思研究怎样节省存储单元，怎样减少不需要的操作。  
  
 在计算机应用领域，科学计算是一个长久不衰的方向。该方向主要依赖于应用数学中的数值计算的发展，而数值计算的发展也受到来自计算机系统结构的影响。早期，科学计算主要在单机上进行，经历了从小规模数值分析到中大规模数值分析的阶段。随着并行计算机和分布式并行计算机的出现，并行数值计算开始成为科学计算的热点，处理的问题也从中大规模数值分析进入到中大规模复杂问题的计算。 几何是数学的一个分支，它实现了人类思维方式中的数形结合。在计算机发明之后，人们自然很容易联想到了用计算机来处理图形的问题，由此产生了计算机图形学。计算机图形学是使用计算机辅助产生图形并对图形进行处理的科学。并由此推动了计算机辅助设计（CAD）、计算机科学（CAI）、计算机辅助信息处理、计算机辅助测试（CAT）等方向的发展。  
 在各种实际应用系统的开发中，有一个重要的方向值得注意，即实时系统的开发。  
利用计算机证明数学定理被认为是人工智能的一个方向。人工智能的另一个方向是研究一种不依赖于任何领域的通用解题程序或通用解题系统，称为GPS。特别值得一提的是在专家系统的开发中发展了一批新的技术，如知识表示方法、不精确性推理技术等，积累了经验，加深了对人工智能的认识。20世纪70年代末期，一部分学者认识到了人工智能过去研究工作基础的薄弱，开始转而重视人工智能的逻辑基础研究，试图从总结和研究人类推理思维的一般规律出发去研究机器思维，并于1980年在《Artificial Intelligence》发表了一组非单调逻辑的研究论文。他们的工作立即得到一大批计算机科学家的响应，非单调逻辑的研究很快热火朝天地开展起来，人工智能的逻辑基础成为人工智能方向发展的主流。数据库技术、多媒体技术、图形学技术等的发展产生了两个新方向，即计算可视化技术与虚拟现实技术。  
随着计算机网络的发展，分布在全世界的各种计算机正在以惊人的速度相互连接起来。网络上每天都在进行着大量政治、经济、军事、外交、商贸、科学研究与艺术信息的交换与交流。网络上大量信息的频繁交换，虽然缩短了地域之间的距离，然而同时也使各种上网的信息资源处在一种很难设防的状态之中。于是，计算机信息安全受到各国政府的高度重视。除了下大力气研究对付计算机病毒的软硬件技术外，由于各种工作中保密的需要，计算机密码学的研究更多地受到各国政府的重视。  
实际上，在计算机科学中计算机模型和计算机理论与实现技术同样重要。但现在许多学生往往只注重某些计算机操作技术，而忽略了基础理论的学习，并因为自己是“操作高手”而沾沾自喜，这不仅限制了自己将研究工作不断推向深入，而且有可能使自己在学科发展中处于被动地位。他们的工作立即得到一大批计算机科学家的响应，非单调逻辑的研究很快热火朝天地开展起来，人工智能的逻辑基础成为人工智能方向发展的主流。数据库技术、多媒体技术、图形学技术等的发展产生了两个新方向，即计算可视化技术与虚拟现实技术。  
例如，在20世纪50年代和20世纪60年代，我国随着计算机研制工作和软件开发工作的发展，陆续培养了在计算机制造和维护中对计算机某一方面设备十分精通的专家，他们能准确地弄清楚磁芯存储器、磁鼓、运算器、控制器，以及整机线路中哪一部分有问题并进行修理和故障排除，能够编制出使用最少存储单元而运算速度很快的程序，计算机类：计算机系统结构、微型计算机系统、并行分布%2F处理与智能计算机系统、计算机软件、人工智能与智能控制、计算机图形学及计算机辅助设计、计算机信息处理与应用、计算机设计自动化与计算机科学理论对机器代码相当熟悉

软件是与程序密切相关的一个概念。在计算机发展的初期，硬件设计和生产是主要问题，那时的软件就是程序。后来，随着计算技术的发展，传统软件的生产方式已不适应发展的需要，于是人们将工程学的基本原理和方法引入软件设计和生产中。现在计算机软件一般是指计算机系统中的程序及其文档，也可以指在研究、开发、维护以及使用上述含义下的软件所涉及的理论、方  
计算机理论的一个核心问题 我国计算机科学系里的传统是培养做学术研究，尤其是理论研究的人（方向不见得有多大的问题，但是做得不是那么尽如人意）。而计算机的理论研究，说到底了，如网络安全学，图形图像学，视频音频处理，哪个方向都与数学有着很大的关系，虽然也许是正统数学家眼里非主流的数学。这里我还想阐明我的一个观点：我们都知道，数学是从实际生活当中抽象出来的理论，人们之所以要将实际抽象成理论，目的就在于想用抽象出来的理论去更好的指导实践，有些数学研究工作者喜欢用一些现存的理论知识去推导若干条推论，殊不知其一：问题考虑不全很可能是个错误的推论，其二：他的推论在现实生活中找不到原型，不能指导实践。严格的说，我并不是一个理想主义者，政治课上学的理论联系实际一直是指导我学习科学文化知识的航标（至少我认为搞计算机科学与技术的应当本着这个方向）。  
　我个人的浅见是：计算机系的学生，对数学的要求固然跟数学系不同，跟物理类差别则更大。通常非数学专业的所?高等数学"，无非是把数学分析中较困难的理论部分删去，强调套用公式计算而已。而对计算机系来说，数学分析里用处最大的恰恰是被删去的理论部分。记上一堆曲面积分的公式，难道就能算懂了数学？那倒不如现用现查，何必费事记呢？再不然直接用Mathematica或是Matlab好了。退一万步。华罗庚在数学上的造诣不用我去多说，但是他这光辉的一生做得我认为对我们来说，最重要的几件事情：首先是它筹建了中国科学院计算技术研究所，这是我们国家计算机科学的摇篮。在有就是他把很多的高等数学理论都交给了做工业生产的技术人员，推动了中国工业的进步。第三件就是他一生写过很多书，但是对高校师生价值更大的就是他在病期间在病床上和他的爱徒王元写了《高等数学引论》（王元与其说是他的爱徒不如说是他的同事，是中科院数学所的老一辈研究员，对歌德巴赫猜想的贡献全世界仅次于陈景润）这书在我们的图书馆里居然找得到，说实话，当时那个书上已经长了虫子，别人走到那里都会闪开，但我却格外感兴趣，上下两册看了个遍，我的最大收获并不在于理论的阐述，而是在于他的理论完全的实例化，在生活中去找模型。这也是我为什么比较喜欢具体数学的原因，正如我在上文中提到的，理论脱离了实践就失去了它存在的意义。正因为理论是从实践当中抽象出来的，所以理论的研究才能够更好的指导实践，不用于指导实践的理论可以说是毫无价值的。  
　   正如上面所论述的，计算机系的学生学习高等数学：知其然更要知其所以然。你学习的目的应该是：将抽象的理论再应用于实践，不但要掌握题目的解题方法，更要掌握解题思想，对于定理的学习：不是简单的应用，而是掌握证明过程即掌握定理的由来，训练自己的推理能力。只有这样才达到了学习这门科学的目的，同时也缩小了我们与数学系的同学之间思维上的差距。  
 