### 大数据可视化与可视分析-李国政-第9周 星期四 第4大节.mp4

00:00  
所以活动管理的管理成本，包括当然具体然后他的工作或者可能愿意去选择134的，我们这样的我们准备大家做好今天而且是什么样的问题，主要是时间过了，公告了解的太快，好了，谁知来一个市场，谁知企业的这个不怎么样？谁知对谁知我们这块能够这块怎么做好数据时间太好了，这个问题我银行如何开放的工作，谁知谁知道了证据的时代保护的工作需求？

01:27  
好的，各位同学，我们今天来开始我们的第二次课程，因为我的话筒就会发现不是不知道是没电了还是坏了，所以我明白我不知道后来的会不会金额来承担一些，所以目前对于前面国家整个方面也是然后我们上次课程主要去说了我说的4个问题，就是我们为什么学习本课程，然后我如果你学习本课程的话，大家学习到什么东西，然后如何去讲授以及15个方面。

我想因为上次我们学习的非常匆忙，而且下面还有一堂课程，所以我们并没有问过或者需要跟大家交流过有关的问题，大家有没有有什么相关的疑问，我们可以先在这次课程开始前先聊一下，有没有？有的话可以提出来，可以举手。

有的同学在上次课程结束之后，说我们的可视化编程是不是需要跟系统是有关系的，然后当时我觉得没关系，大家用什么系统都可以，无论是迈克尔迈克s还是 windows，这个都是没关系，然后有的同学说我们总本科课程，然后需要补充大家组队需要多少人，我说我们要看一下最终的人数，我希望的是然后我们每个团队每个人都能够去出一份力，因为通过上一次课程我发现有的可能做得非常好，但是我一问而且只有一个人在做，然后那就不是我们的目标了。

所以这是所以我要综合来去看一下，比如说几个人，比如说5个人左右，我觉得可能适合一点比较适合，但是也不局限或者没有1个明确的数字和限定。

还有同学问到我们课程点名，我说上次我不知道我是不是有没有说清楚，我说我们课程不点名，然后我希望我能够讲的有趣一点，然后希望大家能够觉得是因为有趣，然后来去选这门课来去听这门课。对，而不是因为只是因为学分，如果只是因为学分的话，我觉得你们还是太很浪费大家的时间了。

然后除此之外，大家有什么问题有问题吗？

04:07  
然后我们经过做作业的数据是否会比较好，这样的话应该是。

04:17  
能连上我们学校问了一个问题，说我们的数据是不是由大家自己来去找？我觉得我期望的是大家要凭借自己的兴趣来去做一些有意义的事情，比如说你感兴趣什么你就可以去找什么，当然有我们会提供一些可选的选择选项，但是这个选项不一定符合你的兴趣，所以你们要看一下你们会喜欢的可能会感兴趣的地方。

比如说有的非常的天非常冷的都可以，只要你感兴趣，比如说我看到过，有的你不能提餐具，原图这种思路也是可以的，然后有人去分析，比如说你喜欢非常喜欢刷短视频，我觉得你去分析一下短视频的一些相关的数据也非常有趣，或者我喜欢看小姐姐，你去分析一下历年的小品的数据也可以，只要你觉得非常感兴趣，能够运用到我们课程的知识，然后同时发现的信息也比较有趣。

我觉得这个事情能够主题，大家来举例，我们可以给大家一个选项，然后给大家一些可选项，但是不局限。

05:29  
好吧，还有什么问题，其实这种考试就是靠当时的数据。

05:38  
大数据进进，其实说实话说到这点我其实非常有感触，上学期的时候，我们给大家了两个题目，一个题目是都是非常有实际意义的题目，因为题目是我们中国的可视化大会的一个挑战赛，然后这个挑战赛应该有几百兆的数据。

他记得我上次给大家看了屏幕口的数据，然后我发现大家绝大多数的同学把60%的时间都在用于去处理数据，去清洗数据了，我觉得这个事情就就有点违背我自己原先的想法，所以我觉得数据的这部分大家可以看一看，也要看一下大家对于哪些数据比较容易收集的到，如果你说我花费了大量的时间来收集数据，可能可能你的工作量会比较大，然后你觉得你放在后面的时间会比较短，这个也会受一定影响。

对于你们收集的数据也没有什么限制，因为刚刚跟同学说，他说我们是大数据的分析，那是不是一定要公平的数据就非常大？

其实不是这样的，说实话我们要从大数据里面来去文化去支持，其实并不是说把所有所有的比如说500兆的数据都完成了，因为你是要从这个数据里面去做数据的变化，去做数据的清洗，然后你要去转化做非常的转化，为什么你做出来的数据一点问题，通过其实只是变换的话，你认为的一个有效的角度，其实这个数据量也不会特别大，但是它是从一个大量的数据中来去提取出来的，所以这也能够体现出我们所说的，因为大量的数据它并不是说你所有的数据都非常的有意义，然后从这个数据里面来去找到一些对有意义的数据，所以这个问题就是说简单来回答，就是说我们这个数据量也没有限制，但是这个问题就在于如果你想把这个事情做得非常有趣，你认为很有趣，你如果只是拿一个简单的表格来去说，我把它是一个非常简单的数据来去可视化的去做一些数据分析的工作的话，你会发发现你讲的故事就比较受限，对吧？

我可以给大家可以看一下这个例子，就是说其实咱们咱们这个课程在很多的网站上面，很国外都有很多的相关的课程的案例，比如说。

08:27  
我们看一下这种我们去把我们自己的谁知实际上的情况如何更好。

08:44  
我们打算后续会把大家的比较好的案例去建一个网站，然后供网联，你做来年的学生然后去看一看，然后其实有一些国外的课程，他们就有一些网站，后期我发给大家或者课间的时候发给大家，大家可以来看一下，可能算一个教程。

09:02  
其实我。

09:04  
不知道有没有回答你的问题。

09:08  
这个和大家想象的社会的领导。

09:14  
数据是不是比较困难的一点，其实其实这个要看他已经有一些现有的数据库，有一些地方是有一个比较好的开源的数据库，可能就拿去找一下，其实我们也有我们掌握一些现有数据的，然后你可以去用，但是我不想限定这个主题，所以所以比如说大家要是想去做大作业的时候，大家可以自己来去找一下。

09:48  
找一些比较。

09:51  
收集数据也比较简单一点，也可以全看大家。

09:56  
还有什么问题。

09:59  
大家还有其他人有问题吗？

10:02  
我们调整就是说可视化编程有没有这种形式提高发展成功？幸福，我们合作吴局长小吴。

10:12  
同学刚刚说的可视化编程有没有一些前期的学习的要求？前置的要求，比如说你必须要学会什么样的东西，你才能够学习文化的。其实并没有，我相信大家它主要是基于一个叫做议会的安卓来做，安卓是2011年的时候发表在发表在我们领域领会的一篇论文，这也是在张总在生产服务做的一些工作，然后它影响后续所有的转化。

然后这个编程我会再带大家去手把手去做一个例子，所以我们前期可以把我们学习，比如说我们学习基本知识的部分会稍微的压缩一点，然后给大家留在课上编程的时间多一点，这样大家可能更好的去上手。

然后有了上手之后，我们就可以我们大家想发挥的空间就会很大，所以简单来说没有什么线统计教学会学会用别人的语言才能去做这些事情，但我相信他这个账号找不到，因为找不到我。

比如说你们会开始会对大家估计应该是最基本的要求的。

11:26  
还有很多问题不知道，

11:30  
还有其他的问题吗？我们检测为什么。

11:36  
这个东西这是我们主要的技术达不到我们的政策，这个要量。

11:46  
我们同学说我们最后提交的。

11:49  
东西的形式是什么？形式就是我们会在最后一堂课的时候来去帮大家去汇报一下，相互汇报一下做的事情做的工作，然后去看一下，大家相互的去看一下，他做了这么一个词汇报，这是第一种形式第一个形式。

第二个形式就是我们再去发作业的时候会有一些标的编程的代码，我们就刚刚说我们希望把所有大家做的这些事情，大部分作业都汇总到一个大的网站上，大的网站就可以做很多的人来去。

12:30  
去学习的，如果是这样的问题。

12:32  
有一个人没有更大的好处，因为我们我觉得我们这个课程目前讲的是这样，希望在你以后的学习什么，你在你以后的工作中，如果觉得别的东西能够用得到，从这个层次走，让大家达到。

12:53  
好吧，我们没有清楚了。

12:56  
谁知好不要了，很正常的。

我们来去继续我们的课程带来的把。

13:09  
这个东西好好是很好的，谁知谁知。

13:30  
根据我们的计划，我们上次课主要是去介绍了一些以下这个课程，它的设计情况是怎样，这节课我们会来去跟大家去介绍一下可视化和可视分析它的概念是怎样的，然后我们怎么去解构这个概念，以及我们对这两个概念，通过一些实际的案例，然后来去让大家深入的去理解两这两个我们要去学习的东西是什么。

首先来去回顾一下我们的上述课程，那么首先我们说了数据量非常的大，现在以及他每隔两年的时候都会翻一翻，然后这些数据里面蕴含着丰富的知识，我们想要去其实每个人都做的这些数据都是非常具有非常重要的，有很大很多的资源，有很多的价值可以去挖掘。

我们希望大家有这样的挖掘数据中价值的能力，然后化可视分析的作用是什么？我们通过一些案例来给大家介绍，主要还有记录信息的能力，还有回答问题，交流发现，提出验证假设更详细，以发现模式制定决策分析原因，然后来去实现整个数据的上下文，然后来去组织一个数据故事，让你的发现更加的有趣以及体系用户的这样的作用。

然后比如说我们对于每个作用都用了一些实际的案例来去说文化的意义是什么，比如说一图胜千言，我们通过一张图片就能够去概括很多你想要去表达的这样的信息，然后还有说你可以将你所有的数据以上下文的形式来去呈现出来，比如说以地理空间的这种形式去呈现出来，呈现出来之后你就能够去发现数据不同的模式，然后以及我们通过一些抽象的，比如说节点链接的这种图像的数据，然后就能够看到比如说在学者合作网络中他们存在怎样的模式，然后以及我们也可以去用数据动画来去讲一个数据的故事，比如说在19八十、19 18 18 10年到2000年这200年的时间中，200个世界上200个国家，它在它的收入以及人口的寿命，他们的这样一个发展变化的情况，你会看到在前期的时候。他的贫富差距不大，但是各个情况都比较糟糕，但是后续发现贫富差距会越拉越拉越大，然后你会发现不同国家他们有不同的这种行为的模式，然后以及我们的通过方案，通过类似他们的论文之间的语言关系来去看到不同的方向的不同方向的这种不同的研究方向，他们的发展形成的情况。

以上我们都是通过一些案例来去介绍数据规划是什么，我们还没有去讲过可视化的定义是什么部分，可视化的定义是什么呢？因为毕竟是一个学科，最重要的事情就是我首先要定义清楚，我首先要建模好它对不对？这是一个更重要的事情。

其实说实话我们不是一个新的角色，我们不是一个非常老的角色，说其实你看很多的东西，很多的学科他们一开始提出了创立，它的人可能都已经失去很长时间了，但是说实话可视化这个学科，它的创立者现在还在一线的工作，还在发论文，你去开笔会的时候甚至还能看到，所以说其实它的定义其实也你可以看到没有，那么长时间距离我没有那么多。

然后比如说这些这是不同的人，很多大佬对于文化的定义，然后我们可以去看一下这个定义是啥。 Finding the future memory That for our natural meaningful protection and this is a birthday party。然后这个人的定义是传说微笑很高的一个阶段简单水平。然后这三个人他的定义是意识到在a点给我可以选宝贝一下就 amplify operation。

我觉得每个人你们看一下这三个定义，我相信其实会觉得有点懵，你们不知道，你们也不知道这几句话再去说什么，可能你们都不知道是比如说20倍上班或者是什么和单纯是什么，好内心是什么，对不对？

我们可以先来去介绍一下这几个关键的定义，所以这就涉及到一些心理学的方面的知识。因为上次课我们也说到，其实可视化是一个非常交叉的学科的任务，因为它是人来去用，所以你深入到人的这一个，你一定要涉及到一些心理学方面的知识。

所以我不知道大家有没有这么去看过这几个是吧？感觉知觉和认知这三个字。

其实这三个词可能你们觉得差不多，但是这是他们的差距还是还是非常大的，可能我觉得可能大家也都很好奇，我觉得你们会问我去学可视化的，我为什么要知道这些东西？对吧？就像我直接去学编程就好了，我直接去告诉我有哪些开发方法开发算法不就好了，对吧？

我问他好好学习，感觉是什么，知觉是什么，认知是什么？我问问他好好学习这些知识。其实我觉得这个道理就是说大家再去学习，如果我们是计算机科班出身的人，对不对？然后这时候其实很多的培训班他们也在改变，这时候他们教的编程的人跟我们学习编程的人有什么区别呢？这是我当时上学的时候老师告诉我的事情就是说我们学习是系统的来学习，我们知道大家怎么编译的，它的数据结构应该怎么设计比较好，他的办法本来是不是应该比较好，所以这就使得你编出来的程序成为更加的优点或者更快。

所以这就是要从原理上来去了解计算机这种运行，其实同理我们要先去了解人事的问题，而这样的话我们才能够更好的利用到更好的把文化这项技术能够用，感觉这件事情是啥呢？就是说它其实是一个最基本的概念，就是说人他有嗅觉、视觉、听觉、触觉和味觉，这个是与生俱来的这种最基本的能力。

然后知觉是什么呢？知觉是稍微上层一点，我通过这些人的感知人的心理过程，然后来去加工信息来去得到知识。然后认知是什么？认知就是要涉及到我的大脑，就是我要去思考一下这个事情是什么，你一定要涉及到认知。比如说我们举个例子，比如说晓峰在这边学习，然后门啪啪啪敲响，这时候你去开门是吧？这是一个正常的过程，但是这个过程中其实发生了什么？可能用到了我们上述说的感觉和知觉和认知。

这件事情就是说你听到它响了这件事情，你首先用到了你的听觉这个问题，对吧？但是你听到它响的是敲门声，而不是比如说拍击桌子的声音，这件事情就是用到了你的知觉，就是你要认识到你很自然的就想到这是一个门铃声，而进一步的你要去考虑到因为它是门铃声，所以有人在外面，所以你要去开门，所以这件事情你就要考虑一下，所以这件事情就是认识这个能力是在逐步的去提升，这是与生俱来的，可能是稍微能经过一点点训练就获得了，而这个是需要我们的经验和思考来去达到这个目的。

Ok你可能会去问我有这个能力，我为什么要去学习这些方面的知识？我们可以举一个例子，我们在做规划的时候，其实用好风险能力是非常关键的一个一个方面，比如说返点返点图是不是大家很自然的去看到这上面有4个方程，4个4类，对不对？但是这4类其实他们的边界不清楚，对不对？没有那么清楚，然后其实这是一篇论文，是我们领会的一篇论文，我觉得这篇论文特别的巧妙，他是怎么做的，他把这4类去把它建成1个等高线，然后我来去增加一些小的连线，小的段的连线。

这些段的连线其实后面我们会提到它是跟我们的格式法原则紧密相关的，格式法原则就是说其中有一个原则就是能看到这些虽然不闭合的原线，但是我还是要把它当做一个整体来去对待，这个就是一个格式化原则。

然后你会发现我们增加上那些小的连线之后，这4个类就会更加的清晰了，对吧？当然还可以这样，比如说这是4类，你增加上小的连线，你这个方向不一样，你会发现你看到的，比如说通过这边我很自然的去看到这是一类对不对？大家会应该持续这么来去理解这个数据对吧？但是假设我换一个方向来去变加小圆圈，你会发现这是一类。

所以这就是我们用到了一些人的，但是人的一些感知的能力，然后来去我们一眼，首先我们去提升了他人来去发现聚类有几个doctor它的这种效果，同时我还可以去改变它，比如说这是12 3 4班，然后我们用颜色来去表达这4个班，然后我可以用增加这种小的小的这种棉线来去使得它搭配非常的明显，这是第一个例子。

第二个例子就是假设我们用葵花，我们可视化一定是有数据的构建的，可让形式我们可以把它分离来去看，这样的话你会发现我如果想要去看理解这个表达形式，那么首先我要去找到它对应的数据对不对？

我要找这个数据这件事情你一定要用到你的pop，就是我要去看这个数据，把这个数据是啥，然后我进一步的要去找到这个数据对应的它原始的数据是什么，这一部分这部分只要你经过思考了，你一定要用到你的宝贝水平，但是这件事情靠那些是人的非常高级的一个感知能力，所以要尽量避免，所以你要比如说怎么办？

比如说你直接把它放过来，放到那些表格单元里面，是不是？我就人来对我人不需要人的支持，来去建立文化跟数据之间的对应关系。

所以这时候你相当于你用的是你的process，就是你的第一种思路，就是我不需要思考，我就知道这个的话对应的数据是这样的，所以你的分析的效率就会更快。

所以我们我们就可以提出这种嵌入式的抓形式，它就是将你的二三十的能力，然后来去来去将你的需要用到你的感知好内生的地方，把它转换到或者在什么地方，这样你就会提升你的分析的效果。

当然这个工作是我们做的工作，比如他确实去整合到了一些人保集团的这种报表的分析平台里面去看，确实用的就是我们的这种方法，然后他用这种方法来去更有效的来去分析这个数据。

26:22  
然后。

26:23  
我们上面说到的一些关于数据可视化的概念，但是这些概念我们比较老了，然后最近其实在14年的时候，然后叫马尔旺夫人，他是在UK在他算是一个文化的塑造，就是他建立了可视化的这种理论的基础，然后他很有趣的是他的导师都是我不知道大家有没有关注过，我给你讲，所以他的导师就是图给你讲的过程中，就是大家可能看到了，他是斯坦福大学的一个当老师，然后他在应该是大前年的时候获得了图灵奖，然后它对于可视化的定义是啥？

基于计算机的可视化系统，通过提供数据的视觉表达形式，帮助人们更有效的去完成特定的分析任务。

我相信大家会觉得非常的绕特别绕，就不知道他再去说什么，但是其实这个定义。是我们对于可视化公认的一个比较系统比较完整的定义，我们可以一步一步的来去解构。

首先这个定义里面就提到了数据，因为我们是数据表达一定是要有数据，因为数据是你们原材料的建设，其实我们可以通过大数据的处理分析流程来去理解可视化是什么。比如说在数据管理里面需要数据的获取，需要数据的抽取和清洗，要数据的整合，数据的具体表示，在数据分析里面需要建模和分析，需要进一步的去提取他的知识，然后去传播给其他的用户。

然后在这些阶段中其实可怕都可以去发挥的作用，它的作用是什么呢？它可以去将人跟机器然后来去协同到一起，然后来去共同的去完成任务。所以第一个概念中包含了数据，是我们整个步骤的一个起点，就是它整个的一个原材料。

好，第二个步骤是人，你可以看到推广，你会想到为什么要把人来去引入到这个定义中，因为是这样的，我们整个的最终的目的是要去帮助人来去完成任务，这个过程服务的用户，服务的是人，所以我们刚才讲到为什么要把人包括在内，为什么我们的数据分析过程需要人来去参与？

这就涉及到我想问大家一个问题，首先第一个问题是你们觉得什么情况？我普通话是所有情况都会需要我们所有的数据分析，都会需要我们比如说大家有没有想回答一下的，你觉得什么情况下不需要可视化，大家可以考虑一下。

29:36  
谁知。

29:40  
对，其实我上面就写着，就是说如果这个数据分析的过程是完全自动化的，是不需要是非常明确的，比如说我们要去计算一个指标，计算这个指标是大是小，这样这个过程它就不需要规划和参与。因为这个过程很明确，但是有的问题分析问题是不明确的，他需要人的参与不清楚要探索的问题是什么，比如说给你了一个数据，你需要从这个数据中去分析出一些比较有趣的模式。

然后比如说假设我们现在在跟在跟一些保险金融公司来去紧密的去合作，然后他们需要去制定很多的经营的策略，这个策略他们也不知道怎么去制定，他们需要从这些现有的这些研究数据里面来去探索这些数据里面有怎样的模式。这时候所以这个问题就不是一个良性利益的问题，所以你就需要通过一些人机协同的方法来处理。当然还有一些可能的问题是说，比如说金融保险领域，它还有一个产品是。

30:58  
大家。

30:59  
我不知道有没有知道FBI就是我要去解释人工智能，这个地方是在于在欧洲来去评估的，为什么？

是因为欧洲他们去非常注重，比如说你这个模型是不是具有可解释性，是不是有偏的，比如说是不是它会有一些本来模型训练的数据，它会有一些偏差，然后导致模型它会有一些选择性，然后所以这个事情他们就会非常的注意，所以我就需要通过一些格化的方法来去辅助我这个模型的理解，看看这个模型是不是真正学到了我想学习的东西，或者这个模型是不是本身就有问题，我需要通过这个来去解释。

这是第二个场景。

第三个场景是说假设它涉及到一些非常重要的决策，比如说我们再去我们在跟一些运维管理部门去合作的时候，比如说铁路的运维，铁路运维是一个非常重大的决策，如果运维不好的话，可能涉及到整个铁路上的火车上的高铁上的人的生命安全，这时候你就不能完全的交给机器来应付，不能够把它交给一个自动化的办法来去解决，这时候就需要人的参与。

谁知所以这就是我们的一个你说这是高铁，他收集了很多的数据，然后他要去建模，然后他要去分析，但是分析的过程中不可能是说这是一个自动化，他肯定是要把人包括在内，当然还有说假设我们要去写新闻，新闻是一个通常我们通过大中心也好用，LP也好，它肯定可以生成大量的文本对不对？

但是我们去生成文本的时候会发现，而我们其实这个新闻的逻辑是非常重要的，比如说这个新闻，这个新闻讲的什么事情，比如说是在日本的时候把日本发生了一个危机的爆炸的事件，那么这个事件它就会导致如果跑没有人伤亡，但是但是有人伤亡，但是他很妥善的去解决了这个问题，本身这是一个负面的新闻。

但是通过这篇新闻的报道，他把它演变成主题就是说什么说其实发生了很多的报道，但是由于日本很妥善的处置，所以一点都没有出现事故。

然后这种新闻的介绍的逻辑就导致你本身想表达的一件事情本身就是一件负面的事情，但是通过这个新闻的写作的逻辑就变成了一个非常重要的事情，因为它妥善的去解决了，它一点都没有造成，比如说人员的死亡或者是经济的损失等等，所以逻辑是新闻的逻辑，这个逻辑肯定是要由新闻的报社记者，报社的编辑来去决定，所以我们用人工智能的方法自动去生成可能就不合适，所以就需要人机协同上来去做这件事情。

有比如说这就是我们需要人参与的一些案例，然后在这个概念中还有一个事情就是很多设备，就是我们要基于计算机的，当然我们上节课中也提到了，其实我们追溯来去看，我们在很早之前文化不需要计算机，我们使用手同样能够完成，但是那个时候可能可视化更多的是对于一种记录的作用，对于分析的意义来获取，所以有了计算机之后，我们就可以支持大规模的数据来构建文化。

然后这个的意义就在于我们可以让计算机来去做计算机更适合做的事情，比如说比如说我们人类去画这样一张图，你会觉得非常的费劲，非常工作量很大，这样重复繁琐耗时的事情就让计算机来去做，你只要写好可靠的程序做完就好了，那人来去做什么，人来去驱动跟这个形式来去交互，从这个形式中去发现进一步探索的方向，这个是人来去做的事情。

所以我们有了这种被子之后，可视化之后，我们就可以就可以更高效的来去获取普通话程序，同时我们也可以获取更大规模数据的普通话，然后比如说构建这种复杂的普通话跟普通话来去探索交互这种能力，这是我们数据中的概念中的第三个，当然还有比如说这个概念中还有一个叫做视觉表达形式，为什么我们要通过一些视觉表达形式来去高效的完成特定的任务呢？

这个就涉及到我们可以举个例子，可能刚刚大家还没有清楚，比如说靠内生跟profession之间有什么区别，大家可以看这样的例子。

36:09  
或者产品操作。

36:10  
这些数字大家来看一下，我如果说这些数字中呈现了一个增长的趋势，这样的话对吧？

大家去验证的话是不是一个来进行。

36:21  
我要去看。

36:25  
我要想一想，其实还是要花很长时间，我要去验证。

36:29  
遇到可怕的问题。

36:32  
你这个验证的过程你要去统一的过程，这就涉及到你的跑分数，就是你要用到的，你要思考一下，你要动一下脑筋，但是只是看这张图，你马上就能看到这是一个增长的趋势，我不需要动你们，对吧？

这个速度会非常大，这就是假设我们要去看一个数据中呈现怎样的趋势，你这样的话就能解释文化它的意义是在什么，他用你的ps来去辅助你的考核意识，就是将你需要用到你考核意识的地方，需要用到你认识的地方去把它减负，转移到用你那块单选题库，因为本身是非常快，所以你就能够很高效的去完成这样的任务。

然后我们在整个流程的靠推手这两种不同的方式，我觉得大家记住这个例子，就可以能够很清楚的去理解我们这两个方法，两个不同的人的这种感觉通道，它的特点不同，所以说文化就是要借助视觉感知来去增强你的视觉认知能力。

37:43  
这就是我们想做的事情。

37:45  
为什么要通过视觉呢？为什么我们通过听觉？我们刚刚说了人有很多的感觉器官，比如说我们有听觉，我们有嗅觉，我们有速度对吧？这些为什么我们都能够去选我们选，我是觉得那是因为我们大脑再去接触去收集信息的时候，我们的视觉占据的比例是绝大多数占据了80%以上。

其实第一个点就是说我们的信息大多数都是通过视觉的这种通道去收取这块。

好，第二个特征是说大量的视觉信息他都可以去并行的去处理，比如说你要去阅读一篇新闻，我给你举一个例子，假设你去阅读一个新闻阅读一定是线性的，因为你要一条一条的读完了之后你要加还是第一点，但是一张图片不是这样的，所以说一图胜千言，其实你看图片的效率远比你看文字要高得多，你第一眼就能看到这个图片中的哪些，比如说你想去发现的信息，但是听觉系统就是说听觉也是一个陷阱，因为所有的我们的基本信号都是这样渐进而去。

39:03  
传递到我们一个整体，我们要考核你。

39:08  
但是它的嗅觉味觉注意这些视觉的通道就会非常的受限。

比如说我们举一个例子就是一个这个就是一个典型例子，就是说我们的视觉来去处理信息的时候，其实跟我们信息量多少是没有太明显的关系的。比如说我通过这个图里面，这个图里面可能有15个点，而15个点它其中有一个红色的点，你要去找到这个红色的点，但是从50个点中去找到红色的点，它的效率是完全不能说完全一样，就是差距几乎没有什么差别。

但是假设你是一个线性的文本，你要去找到一个一行中的文本它有错误，那你就要线性的来假设它是10行跟50行，你的差距没有太大就太大了，所以这就是为什么视觉它是一个命题处理，我很快就能查到。

40:10  
这里面有什么哪些异常的信息。

40:13  
这就是为保障了事故的原因，这是我们解构的概念，就是布局的。

然后后来我就说这是数据的事情咋办？

这个数据跟我们说的不一样，这个数据前面说的数据是这个数据的起文化的起点，而这边是说我们为什么要把这个数据变成一个视觉表达，这个例子我前面说过了，有大家上节课应该听的话应该也是有印象的就是说这2个4个数据集它是1个二维的xy然后我就说我们可以去计算它的一些统计指标，如果去算它的平均值，把它的方差标准差，把它的相关性，我就可以去代表这些数据了，我为什么还要把这些数据把它展现成一些视觉的表达？

但这是因为有什么？这当然是我们一个连续构造的例子，你会发现它的平均值方差、标准差，相关系数是完全一样的，但是它的模式就会非常差异会非常明显。

这就是为什么我们要去注重原始的数据。刚刚小东同学其实刚刚提到过说我们是不是要把比如这就是一个权衡，其实你会发现我们在讲文化视觉表达构建的时候会发现越来越不够权衡，你要去权衡很多方面，而且权衡人的能力，而且权衡机器的能力，要去权衡你的视觉的把这个屏幕的空间等等。

刚刚同学们说了，说我大数据我要把所有的东西都把它展现出来，这个就是一个权衡，你要把所有的数据都展示出来，你会涉及到你会发现屏幕是没有那么强的渲染能力的人，也是没有那么强的感知能力的，所以你要去对数据去做一些处理，但是你去做数据处理，这不就涉及到这样的问题，我去做了数据处理之后就发现所有数据指标都一样的，这时候你去做这个事，你这个处理你去发现的模式就会做了一个很大的限制。

所以这就是一个我们要再去平衡的一个事情，就是我们不能把所有的原始数据都把它转化成合适的表达，但是我们也不能说完全不对数据落实，然后来去直接去转化，这个就是一个大家会发现这是一个非常有意思的事情，然后比如说非常传统媒体也是文化中一个非常重要的部分，我怎样对数据做变化是最有效的，这个就涉及到我们对于数据的理解，对于后来我们的任务，我们的用户等多方面的考虑，很清楚。

43:08  
ok。

43:09  
当然还有一些可视化的这个定义中，第七个重要的概念是什么？

就是我们要把数据转化成一些实话而去完成，那文化它有一些很重要的特点。

43:26  
比如说谁知。

43:29  
其实我们刚刚说到做文化将数据转化为文化形式，它最重要的事情就是要去做选择。

我们其实会发现它的一个重要的特点就是它的设计空间会非常大，比如说你要将数据转化成一些识别表达，你可以转化为各种各样的每个人应该都有自己的我们再去后面去讲的时候，你会发现不同的数据，一个数据可能有上百种不同的视点可视化的表达形式，你要怎样去选择，或者你要怎样去创新，这个是一个最重要的问题。

当然它的设计空间很大，它本身会因为我本身要把数据去映射的一些识别形式，所以这个数据的处理变换视觉映射都是会影响到自己这种结果，当然规划它一定会涉及到我的交互，这就是为什么可怕，一定涉及到交互，我们后面会去讲到这个事情。文化是作为人跟机器来去共同去协同探索数据的一个中间的桥梁，这时候要我们的交互就是将人的知识来去跟文化来去协同，让知识输入到这个。

44:43  
机器里面去。

44:46  
然后还有一个最重要的方面就是我们要去完成特定的分析任务，而这个分析任务并不是说这个任务是一个对于我们去完成数据可视化构建的一个核心，有了这个任务的限制条件，比如说我们整个可视化的设计当中，那么这个图就是陈拉拉他做的一个成绩，就是说这个矩形的空间是说我一个数据我能映射成多种多样的可视化，各种各样都有好的有坏的、有适合的不适合的，这些你都可以都在这个空间里面。

那么我们要去做的事情是啥？我们要就从这个空间里面去找到一个适合的适合的适合你的任务，适合你的产品，适合你的用户，如果是不好的这样一个案例是这样的，我在这个空间里面我只了解一小部分，比如说我相信这是现在大家的状态，我可能只了解一小部分数据存放，然后我就在这个空间里面去找，然后这些都是坏的，然后你就去找了一个坏的，然后这样就完成了你的任务，但是这个完成任务完成不好，然后我们希望做的是扩展你的空间，然后让你知道更多的这样的不同的文化的方法，然后同时在这个方法里面，你能够有效的去衡量哪些是好的，哪些是坏的，ok我就可以去从这个空间里面去找到一个好的这样。

我们希望比如说这些现在的状态，这就是我们希望大家能够达到的功能，好的大家可以先休息。谁知我们先下课，然后一会来上课。

46:27  
谁知了是腐败的这个问题会有问题。谁知检测的包括文化质量有关叫我的历史，还有谁知我们吃饭的事情，谁知这可能有点矛盾点不行。

企业的合作关系，那么现在根本的这种然后我们的思想，社会的过程，企业改革的方面的问题，数据我们也是希望不同的很多的产品的一些空间。

你说谁知他们有可能包括我们自己一次过去世界的目标公司信息，谁知能够再次的考虑，这个时候我们想考虑的长一点，行业部门的支持是一个不少规模的安排不好，我孩子的重要角色不是重要的这个东西，我想我们至少还有很多工作的事情，谁知所以现在孩子长大了之后也不能启动的时候，这个就非常的问题。

三有可能这块我们所有的绩效增长，目前前面咱们发展的思路，谁知企业的这种方案的信息去支持这样去考虑，但是也很好，你也不应该使用是一种谁知他拿了好多了。

49:18  
你做到花了好几个人，

49:23  
我们的情况以及社会的合作的我们市场的安全大家进行一下。

我想问这个工作，谁知没有了，对人家的作为他公司我们是服务好的，绿色的，应该说都是很明显，所以资料这两块客户资料发展多少，谁知市场要进行过程。

50:02  
中的了。

50:05  
这个项目的市场问题，谁知大家很高兴，数据并不是快，我们20 25，谁知市场的发展我们这个工作很好，支持我们党员的这个级别我们非常清楚。

谁知谁知咱们的客户管理，谁知数据。

51:01  
我们继续来上课。我不知道大家是不是觉得大家觉得我说我讲的是不是有什么大家听不懂的，然后或者是或者是太浅显了，或者等一下大家都可以回去举手，马上都可以。

然后我们也可以增强一下我们的互动，大家也可以我希望让大家觉得这个事情还是比较有趣的，关键是我想表达的这一点，你会发现对如果你觉得一件事情有趣的话，比如说你会发现你更有动力去做那些事情。

51:40  
我觉得做科研。

51:41  
同样是这样的，因为大家比如说面临的现在是大三，所以你们后面可能会去我觉得应该有大部分都会去，比如说继续去做一些研究型的工作，对不对？

其实比如说我对于科研工作，比如说我可以跟大家来去交流一下，我觉得你一定要觉得你这件事情是非常有意义的，非常有趣，然后我觉得两个点就是有趣和有用，你会觉得这个事情非常你感兴趣，同时它有用应用的地方，然后我觉得如果有了这两点，你会觉得你的动力会十足，然后觉得也不会觉得很枯燥，然后有遇到困难你也能够更好的去坚持下去。

我觉得这是我做了很长时间以来的一些经验和感触，很有可能想跟大家去分享，也希望通过这堂课，然后来去让大家能够感受到我们对一个事情感兴趣，然后是什么状态。

我觉得刚刚提到这一点。

52:46  
就是说我。

52:48  
我们可以这样建模可视化的空间，它有很多的可视化形式，然后对于同一个数据都有很多，然后你要从这个形式中去找到一个它最适合任务的，用一个它最适合用户去完成任务的一样一个一个是可视化，而这件事情是不容易的，一开始不容易的原因在几点，比如说第一点，大多数的核查方法可能是没有那么高效的，所以你要去了解更多。

其次你要去从这个里面去找到，你要去判断一下哪一个是更好的，然后这样你才能够更好的去确定它适合的格拉是什么。支持。比如说我跟大家来去让大家来去思考一个问题，我们可以课堂来去做一个小的实验。

然后比如说我想我跟大家说一下，假设我们有一个红色结构处理，然后我假如说我们的电脑的硬盘，电脑硬盘是一个典型的层次结构数据对吧？你有一个很多的文件夹，然后最下层是文件对吧？大家认同这一点对吧？假设我想可视化我这个硬盘，然后我的目的是啥？我的目的是我想去看一下我哪些地方是是需要去整理的，哪些地方结构是不好的，这就是我的目的。

你们想一下我们用什么样的表达形式来去完成这样的任务，你们手里都有电脑，你们都想有这个任务，我觉得都很自然。你们想一下你们自己，我觉得花5分钟的时间我们到13，然后大家可以想一下，你们觉得你们可以相互讨论，或者自己来去搜都可以。

我觉得我想来验证一下，其实大家现在的状态可能然后我们需要让大家去达到这个状态，你要能够去找到适合的可能是什么意思，等你上完课或者三分钟我们到11好吧，你给大家来去想一想，可以随便讨论，你们怎么讨论都可以，可以大声不用带入这个不对，不用可以大声讨论，都可以都没问题。

55:11  
你出来就可以了。

55:12  
我们要去可视化一个电脑的硬盘，你的任务是我要去管理一下硬盘，我要去看一看哪些地方是不好的，我要去我要去重新归置一下我的电脑，你觉得用什么样的形式比较好？

当然我给大家解释了，这个是你的电脑的磁盘是一个层次结构数据，就是从最根结点，然后一直到你的文件。

55:38  
下面一个层次结构，谁知谁知。

55:53  
我会找大家来上来画一下，你们先自己考虑，然后手画一画。

56:02  
企业那么可能隔离出来，谁知还有谁知谁知谁知谁知其他的我们可以更好的问题，是不是？是东西。

56:52  
其实文件它肯定有大小，肯定有类型。

其实这是很基本的非常简单的一个数据结构了。

57:00  
这是最简单的数据结构，谁知自己都出来了，这两个人谁知我们创业是问题，谁知很好，谁知1/2的是是吧？谁知谁知谁知好的，谢谢。

工作质量谁知是这个产品

58:45  
我们来画一下，原来想画思想。不讲话的话。

58:52  
我们可以是国内的，怎么办？不给你了坏了，但是我们的事业这个话题的理解，数据的发展的支持来之前的，谁知外部投资的服务的方式说不好的话，换句话一点，你觉得有什么样。

59:21  
的选择出来了？

59:22  
不知道此行系统非常好，这个也是既然是c位同学。

59:33  
我要再见。

59:35  
他是个公司公司的目标，没有这个东西，一半的时光的时间你就可以咱们可以如果说这个是重点电话消化情况这句话来做，然后给你看长期作用。

另外一个。

59:56  
没事不要紧，你还有什么不能做的对不对？

59:59  
就是边。

01:00:00  
玩边学。

01:00:01  
谁知没有对错，你就画一个什么？谁知

01:00:23  
大家可以天马行空一点，我觉得这个事情你们可以怎么做都可以，只要能够完成你们的任务就可以了。

01:00:37  
自己做得很的很好。是的。是吧？

01:00:51  
还是不要弄这一点，我觉得以后其实大家比如说你们以后工作。

01:01:00  
还是以后做研究。

任何情况，这个背景的告诉我怎么样，好的挺好的。谢谢。

好的行业很渴望的中国客户的企业，我感觉非常的谁知的。不能不可能。然后怎么样？谁知谁知谁知数据，谁知谁知没有什么的不错，谁知你怎么看？觉得不一样。谁知好，要是是这个问题出来了，而且成本的投入到的谁知，谁知谁知。谁知这次的事情很多了，好吧，这个事情的问题。谁知谁知。

01:03:57  
来，大家可以抬头看一下，来。

这位同学的作品已经出炉了，来我来。

01:04:10  
企业先看。

01:04:14  
一下机会服务，我们要不然每个人来去大概讲一讲，我们简单说几句我们自己的想法来。我们从这边来开始来。

应该这个是你的是吧？对你叫你姓什么？你好好小亮。

小娜你来讲一讲你的工作，你的作品是怎么想的。

01:04:44  
因为文件系统它是以书的形式组织的，所以我们把我们把它没有能追踪到最终文件的看成一个树的节点，然后最终能落实的文件称为树的一个最后的叶子节点，因为在文件系统它很关键问题，也是索引和查找的效率，所以其实我觉得文件的大小这个东西倒反而是次要的，我们需要关注的更多应该是层次结构的失衡，所以我们更多时候我觉得应该关注是树的深度，比如说在数据结构里面是否我们把如果更要关注的是索引的重整，我建议就是说做这种东西，我再想一下如何整理你的文件系统，你就可以把它可视化。

如果一个树的两边太重了，你就可以把它重整成一个平衡树，我这主要是从搜索的效率角度来分析一下这个文件系统。

01:05:37  
好的，非常感谢小廖，我觉得这个形式可能大家提到数，我觉得80%可能都会想到我要用节点链接数的形式来去做，对不对？这是最直观的形式。然后刚刚想到说我想去看的不是它的文件大小，我想去看的它的平衡的程度，我觉得这件事情非常对，因为比如说你要去我不想让我的文件系统特别的失衡，对不对？

比如说有的文件我藏得特别深，我藏了10层，我这时候我就需要去重新整理一下，因为这样的话我很多文件都可能找不到，这有了你的前提，我觉得这个形式确实是最高效，为什么？因为节点链接数它最适合去表达的是这个层次结构，它的拓扑结构的信息，就是说这个结构它哪些地方比较长，哪些地方比较深，哪些地方比较宽，因为这个形式是最容易。

01:06:35  
去直观的来去表达出来的。

01:06:37  
我觉得他能够去做的是我来去找到我这个文件系统里面哪些结构不太有些失衡，对不对？

来第二位同学我们来说一下。

01:06:54  
暂时。

01:06:57  
这个姓张好怎么着，我就想到了普通windows文件管理系统里边的管理方式是基本上一个磁盘下面有好多文件夹，然后我想让用户直观的看到每个文件夹在这个文件磁盘中它所的占用比例。

01:07:18  
ok只能。

01:07:19  
让用户更好的管理人家系统，因为我觉得现在大多数用户都已经很习惯那种传统文件管理器的管理方式，所以用这种是最容易用版本不畅， You must always have it，okay？

你知道这个形式叫什么，不知道这个形式叫冰柱图。

对，这是一个也是一个非常典型的层次结构，数据的开发形式，就是说它我可以解释一下，小张应该就是说宽度表示的是整个的能量对吧？然后这个可能是说你的比如说CC1比如说director，director就是文件夹一占了50%，然后进一步的占30%这样的一个深度，然后这是你的深度对吧？你说是这样，然后技术还可以，二下面还有一个字库，他也占了第二二50%，包括他这个的好处是啥？

大家你们觉得对于这个的好处是啥？

01:08:29  
谁知。

01:08:31  
比如说当然完成的任务不一样，因为刚刚小张同学们说的是我想我更看我也看中他的文件夹的大小，所以这个就相当于我们综合了一下我的透明结构的信息以及我的节点书里面的信息，而其实我同样能够看到它均衡是均衡，比如它的深度，它的大小等等我都可以看到。

但是比如说我在这个图里面要去体现它的节点的属性信息的时候，就没有它会更清楚，但是它的结构更清楚，因为这个是我更用这符合小廖当时的设想，我就是想看它的透骨结构它是是不是均衡，而我加上它的这种属性信息我就可以用，我觉得做的也很好，所以选择也是最容易的。

Ok然后上面

01:09:25  
特别是来谢什么？

谢我，好。

01:09:35  
上面我想对于如果说是一整块磁盘的话，可以把它划分成可以把它在如果说最底层的最小块的它实际的物理存储空间，然后在此基础上，比如说在整个磁盘内，它也可能会有第一节所有的文件夹，对于这一次第一层的文件夹可以首先对于这第一层的文件夹就可以先直观的在这个图上看得出来，然后当然也可以选择，比如说选定了某个区域某个文件夹了以后，再可以点击进去了以后再来观察在这个文件夹它下辖的第二级文件夹以及更深层的文件夹，每一个它实际占用内存的空间， Ok。

好的，请你知道这个形式叫什么吗？不知道这个形式叫吹卖叫树图的形式，就是。

01:10:43  
这个。

01:10:44  
形式是它最适合用来去做什么做，我不关心它的拓扑结构的信息，就是翠阳就刚刚小廖他强调的说我拓扑结构它均衡不均衡，这个图很难看出来，因为它的深度。

01:10:58  
它都是通过嵌套的方式来表达的，对吧？

01:11:02  
所以它更强调的是啥？

我的整个的文件夹的大小的信息，这个是最容易能够表达出来的，然后同时小张来说我可以通过一些交互的方法来去探索它，然后也能够去增强他的这种表达能力，然后你可以看到这三种不同的形式，每一个都有自己的侧重点，每一个都有自己的适用的场景。然后好请回。

然后还有第四个来。

01:11:31  
是因为我可以做比较多资产管理，我感觉他太满了，然后想知道到底是哪个地方，它占地不是大小，然后如果说饼状图的话，我觉得这种比较直观的比较哪个部分它在磁盘里面占的比较多的，然后它为什么量化？

是因为我的想法是一开始你只能看到它最顶层的文件夹，然后你要看到再详细的信息的话，你就给小宇把鼠标放上去，然后让他展示这个文件夹内部的大小的分布，所以我们叫不叫画了。

01:12:06  
好的。

01:12:07  
我觉得这个形式可能是我们姓什么路，小路同学觉得这个形式我最能够去看到他同意的信息，这个形式可能是我们觉得最美观的信息，假设我们就这样想让它变得更美观，我们可以用这种形式来去表达它。

Ok你会发现同样一个任务，每个人做的事情，每个人强调的方面都大相径庭，可能差异非常的大。我们来看一下这个问题。Ok这个是啥？我只是举了一个例子，因为大家刚刚画的都是层次结构的数据，然后我就举了一个这个例子，就是层次结构数据可视化的公诉的网站，而这个网站上你可以看到列举了多少个系统，339个不同的数字来显示。

我们来看一下前面大家画的在哪儿。

01:13:10  
比如说我们的。

01:13:16  
催脉。

01:13:17  
福建。

01:13:18  
目前情况还是就是1999年的时候，亚克兰会给他提出的这种对面的主要方法，能够去很好的去看集团它占用的空间的情况。

然后我们还看到刚刚说的冰柱图的形式。

01:13:36  
我讲一讲说了。

01:13:50  
我们的系统里面。不知道，不好意思。

01:13:58  
我来找一找我们的冰。

01:14:00  
幅图去哪了？

01:14:03  
Iphone boss。这个对吧？我自己叫做IP客户。

01:14:11  
为什么不叫IP不多？就是你可以想象成那种因数就是兵兵、兵六兵六对不对？从上到下，然后他就是反映的是刚刚我们从通过它的宽度表示的是它的整个文件的主题，而一层一层下来之后就能看到某一个小的属性，小的文件。

01:14:31  
它的属性的这个图。

01:14:33  
当然还有很多。

01:14:34  
比如说我们的。

01:14:39  
比如说你分配的那边催判断这个就是说什么，其实我们也经常会用到层次数据可查形式，假设你是用的windows，然后你的左侧是不是有一个资源管理器，你的资源管理器其实更重要的不是在于它的文件占多少，而是在于它的文件的目录名怎么去展示，所以这时候他们就用这种横向的排布的来去处理复杂显示，因为这样的话我能够更好的去展示文件的目录的名称。

当然还有很多，比如说我要去跟强调按照时间的呈现的信息，我就可以用这种方式来去表达。

01:15:19  
然后还有比如说我的。

01:15:25  
比如说这个是当时我做的工作，我做的工作是说我把它变成一个leader的什么结构数据，它能够去更好的去做数的大规模数的比较，比如说它之前的数我们只能去比较2个或者10个，用我们的可查形式就能去比较成百个或者上千个这种核查形式。

而这个工作也是发表在pvcp的期刊上，然后当然还有比如说树它还有一些应用场景，比如说家谱树，那家谱树就是它的特点可能是有一个父亲有一个母亲有很多的孩子对吧？他不是说一个路子节点，一个父亲节点下面有很多孩子，而这时候你就可以用加速器的方法来寻找它。

所以说我们通过一个简典型的例子，我觉得大家画的都非常好，其实可能再回到我们那个图，再回到我们这个图，可能大家的范围其实并没有那么宽，并没有那么窄，可能还是比较宽，但是通过我们讲，我们可以让大家知道，其实你所想的这些都是有迹可循的，比如说为什么这个形式它就以这个形，比如说我们这个形式为什么它更适合来去表达拓扑结构信息，这个就可以通过我们在课上后面去讲的一些实际的案例来去做一些解答。

 Ok这是我们的第八个概念。当然其实讲到这儿，我们通过大家来去画了之后，你会发现可视化设计一定考虑了很多方面，包括它的计算能力，包括人类的认知能力，包括我们的展示空间，我们要在这几个方面去做一个权衡，然后来去将数据去更好的去映射到一个视觉表达上来，去辅助我们更好的去完成任务。

01:17:20  
好的。

01:17:21  
以上就是我们对于可视化的定义，第一节课的时候我们讲了通过一些案例来去说可视化视频，这节课我们通过一些实际的定义来去告诉大家可视化是什么。我们可以再看一下可视化的学科发展，刚刚我们也说了，其实可视化它的发展时间很短，在短短的几十年中，可能新有一大批人他们在这个方向上去持续的去做，然后持续的去拓展探索。

01:17:49  
然后得到了很多新颖的很多有效的方。

01:17:52  
方式，当然现在学科也是已经比较比较大的，数据主要是什么？

其实我们这里面可以通过数据格式数据类型来去划分，比如说有信息可视化，有科学可视化可视分析，然后信息可视化，它更强调的是我的一些抽象的数据，比如说我这种层次结构数据，那就属于信息可视化的范畴。科学可视化指的是说我有一些及空间结构化的数据，这个是最早兴起的，然后这个是在后续，然后这个是在2012年左右来去提出的。

然后我来去看一下不同的发展阶段，在计算机发明之前，其实就有很多人他通过一些手绘的方式来去记录信息，来去分析分析数据，他们就是口号的第一个发展阶段，最典型的代表人物就是威廉瑞贝尔。

01:18:50  
他是英国的，然后超同样的他也是欧洲的。

01:18:54  
然后等等这类，因为你可以看到他们之前画的这些数据是啥，这是英国的进出口贸易的数据，就说我们进口多还是出口多，通过这个形式就能很容易的判断。

当然还有一些最典型的，比如说我们8号可能就是来源于我发一下，还有一些比如说这个是一个很有趣的例子，我们的地铁图地铁图普查形式是很有代表性的，你其实这个地方也是一个权衡所在，就是说当我们再去乘坐地铁的时候，你在地下其实你不关心你在比如说我们在北京乘坐地铁，你不关心你在北京的具体哪一个地理位置，而在你关心的是哪一个站点，从这个站点到哪一个站点去，这个站点名字叫什么？我怎么去换乘对不对？

所以这时候其实北京市大家可以想到环形的地铁图。环形的地铁图其实并不代表了那个点站点就在北京市的位置上去，它只是代表了我大概的一个位置，但是我更多权衡的是我要让北京市的地解读更加的清晰的去表达出他们之间的关系是怎样的，我怎么去转这样。然后这就是还是看我们可以看到它是一个伦敦的地点，它最早提出的是他是一个英国人，然后他就最早画出了伦敦的地铁图，然后一直沿用到现在这个规则都没有改变过，这个规则是啥？

就是说我的线路斜的一定是45度，一定是水平柱子的，你会发现北京市的地铁图也是这样的，一定是45度，只要是斜着的一定是45度，它没有说30度60度，为什么？是因为它更好的去放置这些标签的信息。

对。站点的信息，因为我在乘坐地铁的时候，我要知道这个站点的名字是什么，这是最关键的，而不是说站点的地理位置。

所以这个就是我们说它是一个阶段，一它在计算机发明和应用之前就已经出现了，然后阶段二这时候出现了一个很重要的节点，就是说在计算机的时代，我们把它变成一个学科标志性的节点，在1986年时候，1986年87年的时候，而美国的NSF我们叫做nss就是美国的自然科学基金委，其实如果大家比如说以后从事教职，你会发现基金委它是一个风向标，它指导了我们说未来我们要去研究什么方向，未来的基金要投到哪个方向上去，这个是一个很重要的风向标。

然后他们在86年的时候就召开了一个叫做可视化的会议，然后这就认为这是可视化当时的在计算机时代的可视化的一个起点，其实1986年时间也不算长，据我们现在。它的起点是什么？当时在医学界里面在科学界产生了大量的数据，比如说CT数据、mi数据，而这些数据人需要去理解这些数据，我们科学家需要去理解，医生需要去理解，然后怎么去理解，我们就需要通过一些可视化的方法去表达出来，跟传统的图形学有一些不一样，传统图形学它可能表达的是没有那些抽象数据的一些没有这种抽象的数据，我想比如说把一个大家学过图形学的话，可能会记得，比如说我们要做微信，要把一个房间的光线把它给计算出来渲染出来，这个就是没有数据的这样一个渲染的情况，但我们面向的是背后有数据，我们要把数据开发出来，这个就有了一个新的挑战。

在这样的一个背景下，就出现了我们的第二个阶段。

比如说89年的时候，有一个美国有一个可视人体计划，可是人体计划当时说的是在其实有它的起源是有一个罪犯，好像他把他死亡了之后，他把他的遗体给捐献出来，然后做了一些切片，做每个切片是一毫米，然后把人体切了之后切了出来之后我们这时候我们想要去研究生命科学的时候，就想去看这个人它的结构是怎样的，我要去把它呈现出来，其实这件事情不容易，那时候人对于中国科学家对于人体还没有那么了解，我们就需要通过一些格式的方法来去表达出来，然后这时候就有了这种科学数据的回放，然后还有一些流体数据，因为这时候我们要去科学家去做一些科学的实验，我们要把这种流体数据来去把。

01:23:45  
它表达出来。

01:23:47  
然后这时候1990年的时候就建立了PPT阅读的期刊，然后他的代表性人物就是l Os。

01:23:55  
然后这一次的机会。

01:23:57  
被双子对这个人是我的导师，也就是可能隔了我导师是三倍这个人，而这个人现在还在做着科学研究，他是在纽约的纽约的纽约大学州立大学十七分校，他是近代文化的创始人，就是最早去创立了 PVC的期刊，创立了类似的会议。

01:24:26  
这是他来去做。

01:24:28  
然后到了第三个阶段，就是我们在90年代的时候，其实大家会发现除了这种空间的数据，还有一些结构化的数据，比如说文字数据，比如说网络数据，比如说文本数据，比如说表格数据、高位数据等等，这些数据跟空间数据不一样，我们这时候就需要去发展，针对这种抽象数据我们要怎么去开发，这时候就有了一些代表性的人物，然后比如说这时候就跟人机交互紧密的去结合，然后他们代表性的人物就是说pass，然后他提出了这种空税，就是锥形瓶在单位空间中，那时候在单位空间中来去做，然后跟实际上他提出了催卖，这个人是这个人是在迈马里兰大学毕业。

当然如果我们的同学如果在90年代的时候你提出了催麦的话，当然你那时候可能慢慢你也发展成了美国科学院是这样的级别的人物，然后这个是图灵奖的获得者，他提出了他构建了拼，他没有 BI的工具，然后还有比特力，它主要是针对图数据来去做这块。

比如说我们后面学习到最典型的图布局的方法，力导向的图布局就是 PPT他提出的他其实对于中国还是很有感情的，他经常来中国去讲学，对吧？

然后到了第四个阶段，其实会发现除了抽象数据之外，我们有大量的其他的数据也在去产生的，不只是估价数据，可能是说因为数据非常的复杂，假设假设我们在我们的比如说社交媒体数据中既有图数据对吧？

因为人跟人之间构建的这种连接关系，它构建了实际上就是一个让客户然后还有比如说人发了一条微博，人发了一条博客，这个博客就是一个文本数据对吧？

还有比如说图片人可以发一些图片，然后当然它也是一个实时数据，因为人可能会他持续不断的发微博，然后还有他可能跟地理空间又再去结合，因为比如说人再去发微博的时候，他可能打了一些地理位置的标签，这样的话你这个数据不止不只能用一种类型来去报，你往往一个数据它涉及到各种各样的纷纷复杂的类型，你要去分析这个数据，它就带来了新的挑战，这时候是在2006年的时候，它的一些新的研究方向叫做可视分析，就是我要通过可视化的方法来去分析数据，然后同时也面向铺垫了一些面向大众的文化，比如说因为在那个时间段建立起来的就是公司改变了所有的规划规程，然后还有一些和分析的框架，然后他们的典型代表人物就是杰克里杰克毕业的，他是在之前在工厂读大学，现在在在华盛顿大学，然后招CS，他在佐加理工，他主要做的是一些跟文本相关的主要形式，然后带着一本跟两条线，他主要是做分析相关的工作。

然后这就是回顾一下我们比如说1986年到现在，可能得有三四十年的历史，然后整个文化阶段它在慢慢的演进，慢慢的发展，当然这个也已经是06年的事情了。

现在其实我们去看可视化领域大家的一个工作，更多的是我们要跟其实算是智能驱动的智能可视化的一个时代，就是我们要跟人工智能的方法去做结合，然后将可视化构建的难度进一步的去降低，让那人他更多的精力更多的是集中在比如说分析的这种意图。

01:28:39  
然后再去。

01:28:41  
这个是我们最近的一些。

01:28:43  
研究的工作。

01:28:45  
然后当然我们可以重新来去看一下，可视化和可视分析流程它的这两个概念，我们要怎么去建模方案，就是在整个的发展历程中也有很多的工作来去提升，在这个方面也做了处理。

当然最直接的就是说起我们可视化就去做一个映射，它的映射是啥？将数据、任务、场景通过视觉映射、视觉编码听以及处理的办法和方式文化的结果。

01:29:17  
这就是一个。

01:29:19  
最简单的编码，最简单的建筑，然后在1999年的时候。提出了信息可视化的流水线，这个流水线就是从原始数据出发，然后经过数据变换，数据变换就是前面前面同学们提出的，我要一个大量的数据，我要怎样的去变换到我我找到一个关系的分析的角度，这个是数据变换，然后做了数据变换之后，我要去做视觉的映射，然后得到人的表达，然后通过视觉的变化，比如说人的一些交互，然后得到你看到的模式，然后来去完成这个任务。

当然人要根据这个任务反向的来去调节不同的方面，比如说你调节数据变化的方面，调节视觉映射的方面，调节视觉视图变化的角度，这些都是交互人的交互来去做的事情。这个是在大家还记得是1987年的时候，算是可视化正式建立，到了2005年的时候，正好是可视化学科建立18岁18岁大家都知道是一个成年的年纪对吧？

在国外也是这样的，18岁就代表了这个学科已经比较成熟，这时候顶会上的亚团会他发表在文化领域的顶会上获得了最佳论文奖，他这个工作就是在去说什么，他已经发展了18年了，我要去回想一下，可视化什么样的可视化是有价值的，什么样的可视化是没有价值的，我们要指导我们未来的发展方向，这就是他的工作的名字叫做the value或者没什么我的可怕的价值在哪里，然后他就提出了这样一个模型，这个模型是说什么也很简单，就是我从数据出发可以然后道德形式，然后我将通过一些图形图像动画的方式传达给用户，传达给经过人的发展性感知，然后得到knowledge得到支持。

对得到知识之后，反向我可以原来去探索它之后去修改文化的构建过程，然后修改得到并不是怕形式又发生了变化，然后这样的一个迭代，这样就是一个他认为的这样的一个迭代过程，然后进一步的他又去做了一些他有一个经济学模型来去来去解构可视化的价值，比如说他经济学模型是啥，价值是什么？

价值就是我付出了多少努力，我得到了多少的回报，如果我得到的回报越多，比我的努力很多，我这个价值就是最大的。

如果付出了很多结果只得到了一点点，对影响非常小，我这个价值就很小，这个就是他最核心的想法。

然后他这样的话，我基于这个框架，我怎么来去看我的努力，我的开销是什么？我的回报是什么呢？我的开销就是人付出的一些劳动，比如说人要去实现可刷系统，人要实现可刷形式，比如说大家要是这个形式我要去实现出来，这个是一个开销。

而其次开销是说单个用户我要去比如说开销是整个垃圾为它实现的开销，然后这个是用户来去基于垃圾来去实现合法的开销。

而进一步的人类探索的过程中也有一些判断，因为人要去交互，交互就要花你的时间精力，这也算开销。然后还有感知探索的开销，对，这个就是你的开销不同的开销，它的次数是不一样的，比如说我们构建一个壳的话，它的开萧是啥？

我首先构建了IOA开销和CII，然后n次比如说用户的数量是 n个，然后乘以n就是用户的开销，然后n乘m就是说用户构建了多少个这样的复杂程序，n乘m乘CM三个阶段探索了多少次，然后以其用户的感知认知就是这样的一个建模的结果。然后这个回报就是说我每次探索我反馈增加了多少分到位，然后你就说我有价值的信息，我的价值体现在哪里？这个技术，然后回过头来去看，这个时候它的收益就是纪检，所以非常容易很简单，但是有了这个表达之后，你会发现什么？

他会发现你可视化，你让用户用的人越多，他可能价值就越大，因为你 n摆在这， n是前面这个技术，所以他要去干做更朴实的一些表达形式，就是让用的人越来越多，然后你会发现它构建的开销越来越小，你这个收益越来越大，所以我就要尽可能的降低人构建转化的开销，把 ci其实他没有技术，比如说我构建一个很好用的拉格尔，他花了很多的精力，但是对于单个的转化的开销收益来说，其实它是微乎其微的。

然后所以说有两个方向，第一个方向就是我要去让这个形式更普世，让更多的人能用得到，这是我们的一个发展方向。然后第二个方向就是我要让可视化去在更大的场景里面，更大的应用场景里面去扮演的价值，比如说我的一个分析系统，我能够帮助比如说人保，比如说金融这种保险，能够一个系统，我能够帮助他省几十万几百万，我觉得你这个价值就会很大，所以说所以尽管它可能场景比较有限，没有这么朴实，但是你这个场景里面很有用，那也够了。

所以当然你会觉得可能想想也是一个顺理成章的事情，但是他就通过这种建模的方法表达了出来，给我们提供了一个方向。因为后续就是要么用股市，要么这个场景都不一样，我们要去往这个方向上去发展。在2010年的时候，在这是欧洲他提出了一个文化建模的框架，然后这是可以分析的框架。就是说我们可以去看这个本质分析框架是啥，数据来了之后我要去做数据的变换。

01:35:44  
为啥？

01:35:46  
好，现在已经下课了，好吧，我们下节课再继续讲，谢谢大家。

01:35:52  
数据谁知。所以谁知谁知的我们还有谁知我的给我们谁知。

01:36:33  
谁知。

01:36:38  
谁知我们行业的，但是特别是我们的谁知。目前也是有一模一样的业务目标。

01:36:54  
然后说我一看就是抓进去了，怀疑起来了。

01:37:00  
数据我们在上面的也是不是说法律，我们要落实。

01:37:11  
市场运行的复杂的讨论过上面的内容都这么的热血，然后我们这样的生活概念也是让公司的的人能够吃下去的，不错。

谁知我去那里找你去举办这个例子。

可以的，最后的话最后我们的资源扩大的话，可以调整我们有限资产。

01:37:38  
我们也可以去利用所有的事情不要的一个成本还没有明确，你还要做工作当中，你要根本就看不到要不要钱，他们都不是，我真的感觉目标到了为什么？多少人。

01:37:58  
因为上周水平他们自己的电话的数据，目前跟我们说有没有什么时间，他们应该不会。

01:38:07  
学校和服务关系我们不知道。

01:38:09  
然后我们中央这一块的发展的情况，

01:38:13  
谁知我说后面的我这个方法搞出来，谁知其实有什么意思？谁知的策略，谁知谁知不开心了。