### 大数据可视化与可视分析-李国政-第9周 星期一 第3大节.mp4

00:00  
我们老师一会上来了，谢谢工作的休息。好，经常就是说我说咱们学院还没通过我的计划，我们放的应该是这个问题。好，然后谁知道这个讲过了。数据。数据、数据这个平台我们可以做好数据，当然就是说研究电源的发展，去年我们都不一样。

00:47  
谁知。

00:51  
谁知。

00:57  
还有的数据，所以我想死了。谁知。

01:11  
好的，我们开始上课了。非常荣幸我能跟大家来去共同去学习大数据，可视化个人分析这门课，然后在我看来所以可能大家都非常有趣，也是非常实用，也是非常激动人心的一个技术，然后我希望我通过这门课能让大家感受到它的意义这方面。

然后首先来简单的介绍一下我的印象，我觉得大家看到我现在可能觉得这个老师怎么造型那么别致，然后我为了避免让大家分心，其实这个不是开玩笑，我上周去看病的时候，然后去我们校医院看病，我总共花了10分钟时间看病，然后结果医生问我8分钟，你眼睛是怎么回事，你这怎么这么高级？

然后他直到我给他说了我这个眼镜，比如说他还试戴了一下，然后我还给他拍了，我还给他发了我一个怎么去网上购买眼镜的链接，然后他才给我开始看病。

然后如果大家需要的话，其实可以线下来找我要这个链接，然后线上可以不用说了。

是这样的，我跟大家解释一下，我现在我过年期间出了一个小事故，所以我没办法佩戴正常的眼镜，然后所以也很抱歉没有办法以最好的状态给大家上这个课，但是应该没什么关系。

所以我跟大家解释一下，然后我觉得我们我们上课的时候也可以大家避免困惑，然后我来介简单介绍一下我自己。我是2012年到16年的时候，在山东大学计算机学院获得本科学位，然后2016年到2021年在北京大学获得博士学位，然后2012 21年之后，我就来到北理工西南医学院继续从事科研工作。

我的研究方向是一直以来都是大数据可视化可视分析，然后以及跟人际交互结合的一些方向。

然后其实我可以简单的去介绍一下为什么我选择这个方向，当时我本科的时候可能也是你们大概你们这一年这个状态，因为大三下马上就要去选择方向，我当时参加了很多的数学建模比赛，我参加了美赛，也参加了国赛，我当时还拿过一等奖，然后但是我参加比赛之后，我发现我觉得用Matlab来去一直写一些仿真的程序，我都是觉得不过瘾，觉得好像我一直没有去解决这个问题，没有去解决实际的问题，反而是把它抽象成一个很很抽象的模型，然后这个模型到底有没有用，好像都是自说自话，然后后面我就想同时我还比较对于统计学比较感兴趣，然后当时我就觉得我以后一定要去做我能看得到的这样的一个工作，当时我觉得我现在还是很幸运的，我觉得我选择了一个我一直觉得非常有趣，非常激动人心的方向，然后我们来共同学习这个方向的时候，也希望大家能够感受到这个方向的有趣所在。

然后我发表了一些文章，然后其实可以看到发表文章不只是在位置，是在规划领域的领会，然后同时还有一些人际交互领域，还有数据挖掘领域的一些会议和期刊，其实可以从这个地方也能看到我们可是它这个方向它是一个比较交叉的方向，我们在讲的时候本节课讲的时候，大家也会感受得到它为什么是一个交叉的方向，它交叉在哪里。

然后同时我也说一些科研项目，然后我觉得比较我还是很自豪的，就是说其实我刚来北理工的时候，然后是带本科生去做毕设，然后还拿到了一个本科北理工的毕业设计的重点项目，同时去年的时候我我以为去年的时候也是同样的一个课程，然后有一个本科生，然后他当时做的特别好，然后后面暑假我们就一块去做工作，我当时后面也知道他他发表了一篇人交互可视化领域的主题会议，顶级期刊PPT的文章，然后所以我觉得北理工的学生真的特别牛，特别棒。

然后以上就是我对于我自己的简单介绍，然后本今天我觉得我主要来去回答4个问题，这4个问题可能会回答大家的疑虑，为什么第一个问题是为什么我要学习这个课程？第二个问题是我通过这个课程将学习到什么东西。第三个问题是这个课程8周的时间说短不短说长不长，但是我们是如何规划这个课程，大家在这个课程中是一步一步怎么去学校想学习的这部分内容。

第四个问题是本课程将会如何去评估？首先我们来去说一下第一个问题，第一个问题是首先我们都知道我们在大数据的时代，其实这个数据量有多大，大家有没有意识？其实首先第一个问题是2010年产生，全世界产生了多少数据？2010年的时候我还在我还没有读本科，我还在高二的时候，高二的时候我记得我不知道大家当时是什么阶段，高二的时候我记得智能手机还没有怎么去流行，我记得当时塞班我不知道你们还记不记得塞班系统，当时是诺基亚，现在都已经巨人已经倒下了，诺基亚，然后当时他们研制的塞班系统，然后当时还跟安卓系统来去安卓二位当时还在还占有很塞班还是很强大的，是现在来看其实安卓跟iOS已经是二分天下了。

所以当时的时候其实我们的移动互联网时代甚至还没有开始，对，然后当时所以那个时间点还是比较落后，还是信息还是没有那么信息化还是没有那么能够完善的，当时产生了多少数据是两个合并，然后隔壁国我们可以看数一下有几个0个十百千万、10万百万千万亿、10亿百亿千亿万元，11万亿个GB，大家可能对于GB是比较熟悉的，比如说我们现在人口现在有大概8080亿，然后如果我们以现在的人口来去平均的话，其实我们每个人那就有250GB的数据每年。

对，当然我觉得大家我对于我们所有的学习计算机的大学生，然后来说，其实我们产生的数据远比普通人会多得多，所以我觉得大概每天产生1GB的数据，我觉得肯定是可以达到的。

大家肯定会很好奇为什么我在学校里面遛一圈，或者是我周末出去玩一圈，我就产生了1GB的数据，一天1GB的数据，1GB的数据，可能大家的电脑可能500gnet然后500GB其实有这么多的数据都是在做什么呢？

我觉得这个可能是大家会比较可能会很好奇的问题，当然还是2010年的时候，但是其实现在这个数据量的上升速度几乎是每两年翻一番，我们现在在2024年的时候，现在当时是两个合并，现在可能是147个能力，这是一个预测。

这个数据量的话远比我们当时一天产生1GB的数据量要多得多得多了。这些数据中包含了什么？我觉得。

10:04  
这些数据非常的丰富。

10:07  
比如说这个是旧金山出租车GPS的数据，这个数据其实这个数据没有不是路网，但是其实你能看到这个数据把它画出来之后就是路网的数据，所以说这些汽车的所有的轨迹，我们把它收集起来之后，它就能够去描述我们这个社会，比如说很多的物理的信息。

当然其实大家回想起来，前几年的时候为什么滴滴在美国上市，然后被国家去调查呢？其实其实同理所有的滴滴行车的数据，我们把它画出来之后，没有不需要我们中国的路网所有的路网都已经拿到了，同时还有什么呢？比如说北京有很多的北京有很多北边有很多的军事跟军事相关的区域，对吧？比如说有什么研究所，那些所都是不能滴滴，出租车都是不能进。这时候如果我们把所有的都画出来之后，发现有一部分那是车是不能进入的，这时候其实这就是一些敏感的区域。

其实我在讲这个故事的时候，你会发现我们从数据就转到了知识，知识是什么知识，比如说知识路网，这就是知识是什么？知识是比如说某些地方是一些敏感的区域，是我们需要关注，是比如说敌对势力需要去对抗，我们需要去关注的，这就是一些知识，怎么从数据变成知识，这也是我们本系列的课程需要去回答的问题。

然后除此之外，除了我们行车的数据，还有医疗的数据，医疗的数据大家可能没有意识，但是如果比如说你们跟一些在科研的企业科研的时候，其实智慧医疗是一个很火的话题，然后我们如果跟医生来去合作的话，其实这些数据往往是非常难拿到，比如说人的心电图的数据，这些数据你在所有在医院看病的数据都是会被监控下来，但是这些监控下来的数据这是很宝贵的数据，然后这些数据就反映了人的真实，身体健康，然后对应的怎样的治疗方式，然后等等等。

其实这是一大部分的数据。

然后除此之外，还有我们的社交网络，提到社交网络现在对现在我们大家想到的就是微信，微信是一个对移动互联网时代其实是一个非常对是一个最重要的入口，然后他现在应该是月活，我觉得应该有13亿十几亿10多亿。

其实社交网络一直在不断的发生变化，比如说我感觉我们当时在本科的时候其实不是微信，可能是人肉，大家可能觉得这个名字都非常陌生，但当时微信没有了是吧？我们还是用然后所以当然现在还有抖音对吧？

然后美国其实为什么要去我们为什么一直要去收购配套？是因为这些社交网络中包含的数据，它不只是数据，反而是一些战略的资源，当然它去引导舆论，这是一方面，当然还有其他很多重要的方面，数据就是安全，数据就是一些重要的资本。

然后还有以上是比如说我们的行车，我们的地理空间的数据，然后还有一些我们生病时记录的一些规模化的数据，结构化的数据，然后还有一些我们的社交网络的数据，然后还有一些比如说在互联网上最多的数据就是文本数据，大家其实我们的媒体有一点是一个非常重要的数据来源，对不对？

其实我们看到一个不熟悉的名字的时候，我们都往往都会搜索 Ppi上它是怎么去定义的，它的一个最重要的方面就是说我要去这它的一个最重要的特征是所有人都可以去修改，都可以去编辑。

其实你现在看到的可能是在美国或者在欧洲那些人他们的理解，他们去编辑的结果，当然你也可以去编辑，然后编辑的结记录当然形成的就是一些知识，我有一个朋友他在这VP也是一个公益活动，公益性的一个网站，然后有一些比如说大公司去投资它，为什么投资它？为什么谷歌去投资它？

是因为这里面就含有一些重要的知识，然后他们去比如说谷歌他再去做推荐的时候，他们可能要用到知识图谱，知识图谱中的知识怎么去抽取，我就要从这些数据中来去抽取，所以数据中一定是蕴含着非常重要的知识，然后这时候我觉得这句话就能够很好的去概括数据它的一个重要意义，这是 Google的领导 the ability to take data，to be able to understand it，to process it，to extract value for me。 To realize this，to communicate，that is going to be a huge important deal in the next decade because now we really do have essential，Free and previously have so the complimentary separate factor is it the ability to understand that the data and extract values on this？我这边标红了一些关键的概念，就是我要去理解它，我要去处理，我要去提取关键的重要的价值，然后同时还要去把它可视化出来，然后跟他去交流。这个就是我们对于数据，而数据其实是一个很重要的战略资源，为什么数据这么重要？是因为数据中有一些重要的价值，我们要通过这些能力来去把这些价值去提炼出来，可以想象我们现在比如说每位同学我们都可以收集到大量的数据，我们其实都坐在金山上，其实往我们如果有这个能力来来去把这些数据中的价值去提取，获取的话，我们就可以比如说有了这个能力我们就掌握了，比如说后续我们能够去充分挖掘数据中价值的这个方面，所以我们这个课程后面这样希望你有这个能力，然后去掌握我们这个红字中所标识的一些重要的事情，你能从数据中去理解和挖掘，然后来去交流它的价值。

Ok说到了数据，现在大家都认同非常有价值的对吧？我们还要需要这样的能力来去从数据中来去提取价值。什么是数据可靠？我们可以从我们的比如说我们说一个概念的时候，往往就是说我来去Google上去搜一下，这发表的意思是什么，数据可视化是什么？比如说我来去搜一下石化，可以看到所有的清一色，我是有多少多大，清一色全是这样的东西，这个是什么大家知道吗？这个叫做数据大屏。

这个数据大屏是说我在一个公司里面往往有一个大屏幕，它可能叫做态势感知大屏，其实它的作用就是我来去监控我这个公司里面所有的比如说各个不同的指标，比如说人保集团他们有一些自己每天销售了多少数据，然后他们等等这些指标，我就有一个大平台去监控，为什么要大屏，是因为其实很大一部分程度我要去做交流和展示的原因，但是我可以告诉大家，这不是我们这堂课要讲的数据的套路知识，对。然后当我用英文去说，你会发现这个还是明显有区别的，它就不是清一色的结果的，可能会有一些什么宁波花费这样，当然也有数据大屏，然后也有一些新闻等等，你会发现其实国外的定义当然他也没有那么全面，比如说这些是一些可视化的库，然后这是这是我们收到的结果。

然后其实我们看到这个结果之后，我相信大家听到数据可视化这一个词语的时候，可能跟我们搜到的这些结果是类似的，可能我对于可视化的理解就是说，我要用一些炫酷的图表来去把这些数据展示出来，展示出来就可以了，我也不用管了，反正比如说Excel也可以展示，或者开展的一些工具也可以展示对吧？

但是这堂课我相信学完这堂课大家就会了解，其实可视化并不只是如此，说到这里我们就可以比如说举一个比较简单的例子，其实我们可视化是什么？其实是我将一些复杂的数据转换成人类可以理解看到并且理解的图形图像，然后为什么是这样的图形图像呢？其实可视化是这个步骤，对不对？但是从数据转化到可理解的图形，这个图形我们就可以把它转化成一个映射，这个映射怎样是最优的，比如说我要跟什么相关，其实有很多的方面取决于你的任务，取决于你的数据，取决于你的应用场景等等。

其实你的数据不一样，你的任务不一样，你的应用场景不一样，你从数据转换成可视化结果的时候，它不一样的，所以说这个事情并没有那么简单，然后我们可以回头来去看，当然其实转换得到之后，为什么它就有效？是因为我们人来去看一些枯燥的数字是非常效率低的。

比如说你看到这些数字的时候，你其实也不知道它有什么含义，我假设我告诉你说这是空气，这是导弹再去发射的时候它周围的一些空气流动的情况，所以你也不清楚这些数据它其中有什么样的模式，有什么样的不一样的东西呢？

但是如果我们把它转换成一些图形图像，其实你一眼就能看到它是这样的，但是由于一些专业的科学家，然后专业的人员，他们去理解这些数据，就能够进一步的去发现我要去怎么去调控，比如说导弹的发射，导弹的形状等等，然后我们为什么需要。

21:31  
可视化。

21:32  
可视化有什么作用。

21:33  
对不对？

21:36  
我这边列举了，其实可视化有几个比较典型的重要的作用，我们可以一会儿每一个都用一些详细的例子来去解释一下。

第一个例子作用是记录信息，说到记录信息，我们可能会想到其实这个名词就比较比较比较感觉比较远古一点，然后其实可视化在说说实话，其实可视化在出现数据出现信息之前就已经存在了，比如说我们人存在的时候，我们就喜欢用一些图形来去记录，比如说这是古巴古人的地图，就是中国在中国记录的考古时候发现的一些地图，这些地图都是一些复杂形式。

再往后推一点，其实在宋代的时候，比如说这是宋代的雨季图，也是考古发生的，能发现其实这时候比如说黄河长江，比如说黄河的井字系，长江的w系，然后我们的海岸线其实以及海南岛其实都已经跟我们现在相差5亿了，所以那个时候人类去理解这种地理的信息，同样是跟我们现在采用同样的方式，我们都会去选择我们更容易去理解的这种方式来去表达。

然后当然还有城市的地图，比如说这是苏州的地图，然后这就是一个我们只要去学习文化，一定避不开的一个最经典的案例。

这个案例我可以告诉大家，这个案例的背景是什么？是在拿破仑他去东征的地图，拿破仑在18世纪的时候他进攻俄国，然后整个的过程，这就是这个图就完整的去表达了拿破仑进攻俄国的整个过程。

然后我们都说一图胜千言，大家都听过这句话的，我用一个图就能够去概括很多的我想表达的话，而且这个会非常的形象，因为人去读文字往往都是线性的，但是读了一些图片我们都可以去理清楚。

我说到这个背景之后，大家能够理解这个图的含义是什么，比如说我可以进一步的去解释，然后这个条形它的宽度代表的是军队的人数，而这个颜色代表的是黄色代表的是进攻，黑色代表的是后盾。

24:21  
然后背后。

24:23  
是它的地理位置，当然还有一些河流可以看到这是河流。

24:29  
过去了。

24:32  
然后还有底部，它有一个温度的曲线，就表示它当前的温度是怎样。我们看到这里之后大家会有什么发现吗？比如说我可以概括整个拿破仑东征的过程中，或者是他最后概括我的他东征的一个事件，他发生了什么？

我们可以看到其实他一开始的时候人数会非常多，然后非常多，然后在走的过程中就在锐减，然后到了莫斯科的时候会发现其实人数都已经只有他出发时候的1/3了，然后当然它中间也会发生了一个小的变故，就是说一一堆人可能走散了，然后走到了这里，然后到了莫斯科之后，这时候兵败如山倒，然后就回来了，这就灰溜溜的回来之后发现人数也是越来越少。

然后每次经过一个河流的时候，它的人数就会减少一些，每次经过一个河流人数就会减少一些。然后直到到法国的时候，其实人数1/20都不到，而且一个很有意思的发现是你会发现其实大多数回来活着回来的人其实往往都是是之前走散的那批人，然后当然他跟他的时间有关，温度有关系，比如说它温度下降的时候，它的人数就会下降，就会减少的越来越剧烈，越来越剧烈。

26:05  
对。

26:06  
这就是用一张图来去概括这样一个整个的拿破仑中间的过程，这就是我们记录信息，我用一个图就可以记录所有发生的信息。

而第二个可视化作用是回答问题，比如说这个是一个英国的摄影师，然后他当时想要去探索的一个问题是马在奔跑的时候4条腿是否是同时离地，然后基于这个问题，他就做了一个这样的装置，这个装置是我这边排的这样子，然后我再去让马跑过，然后我一直拍照，然后把拍的照片给我记录下来，然后这样就得到这样的一个结果，这个结果我们就可以看到有些有些地方麻雀都会出表比例的，这样就可以去解答他的问题。

然后第三个方面，第三个作用是分析原因，我大家如果爱好航天的话可能会清楚，比如说美国它有一个挑战者号是在1019~80年代的时候，时候，它发生了一个载人航天事故的重大的致命的事故。然后这个事故是在它飞机起飞了之后，当时应该是探发上去是为了去探索发的卫星，然后它发上去之后过了几十秒飞机就爆炸就解体，7名航空员就全部名单，这时候就肯定我们需要去探索它，为什么导致它的原因对不对？

这样的话我们才能总结经验，要不然我们我们找不到原因，我们后续发射还可能会出现这种问题，这就是一个重要的一个任务。

这时候一个其实。

28:02  
你。

28:04  
就能发现他把发射的记录要把它画出来，画出来之后它跟温度有关系，比如说横轴是温度，纵轴是。

28:16  
出现的事故的次数。

28:19  
你会发现温度高的时候，几乎都是成功的出现事故的内容，然后温度越低，其实大概是这样一个曲线，温度越低，出现了事故的次数会越来越高。

然后其实当时挑战者号当时发射的时候是在。

28:39  
这个温度。

28:40  
所以它自然的它就会联想到是不是因为温度的问题，所以导致了这个事情也导致了这个事故。对。

后面他们就去分析，基于假设我就去分析一个启发，我就分析这个原因，结果发现果然是因为助推器它的o型环失效，然后导致它出现气温气在气温极低的时候，它的密封性失效之后，它就会。

29:10  
出现了这个事故。

29:12  
这就是通过比如说我们把一些数据把它给展示出来，展示出来之后你就能够发现给你一些启发，就发现数据中存在的一些模式，然后有了这个模式之后，你就可以去验证你的假设，进一步的去探索你的发现。

29:30  
然后。

29:34  
这是第三个作用。

第四个作用这是一个很典型的例子，典型的例子是可怕在疫情传播领域中最早的应用，现在在在英国的时候还立着案例的一个，然后他的一个事迹，他的纪念碑，他这个事情说的是19 19世纪中期的时候，张之洞他的一个伦敦的地图死亡地图当时出现了一个事件，就是我当时出伦敦出现了大型的霍乱，所有人找不到霍乱的原因，可能当时大部分人都认为它是由于空气传播所导致的，然后这时候这个人他就很创造性的然后来去做一件什么事情，他把所有的死亡的病例他就挂在伦敦的地图上，画在伦敦地图上之后，就会有一些很有意思的发现，这发现是啥呢？

我们把这个地方放大了去看，其实这个地方一眼就能看到这个地方是最多的，对不对？死亡的人数最多最密集，然后以这个地方为中心，慢慢的向周围来去扩散。然后他同时发现这个地方有一个水泵，然后他就来去验证为什么这个地方死亡的人会特别多，因为如果是空气传播的话，肯定是呈现出这种模式，可能分布的会非常均匀，所有的地方都是很均匀的一个对死亡的特征，但是这个特征不能够去验证它是空气传播的，然后他就会去解释为什么这个地方是因为这个地方水泵它有一个它当时这个地方是有问题的，水泵水是有问题的，而这一个楼里面的人就会从水泵来去取水，然后就导致这个楼里面的人他的犯病率会特别的高，然后所以就一个猜测就是饮用水被污染，然后进一步他就把水泵把它给封住了，封住之后霍乱之后所有的都给解除了。

对，所以当时如果大家有机会去英国的话，还可以看到当时现在水泵它的一个记录。

31:59  
以上是。

32:03  
我们用可视化来去解决问题的，其实这边还有一个案例是南丁格尔玫瑰谷，这个玫瑰图我们落地大家都说它是集才华与美貌一身的数据图表，而这个图表它其实可以看到非常漂亮，像一个玫瑰一样，所以我们就叫它玫瑰，当时的来源是南丁格尔，这是一个故事。

他在战地医院去做工作的时候，他会发现其实大多数的士兵他死亡的原因不是由于战伤战战受作战时所受的伤，而是由于他的医护条件不够，然后慢慢感染，做到比如说蓝色区域就是原本被避免的死亡的视频数就是蓝色区域，然后红色区域是因为受伤过重所导致的是黑色区域是属于其他原因的事情。

而这个图怎么去理解，这样是一个环形的环形就导致它的时间，然后它当然是从这边开始，然后到这儿它这个时间然后又从这边开始。

好到这儿，然后你会发现它一开始的时候在这个地方它的你因为死于原本可以避免感染的士兵数会越来越多，其实可以看到这个图非常的有感染力。当你想要去表达你想表达的信息的时候，假设我们以后同学工作对吧？

然后你想我发现了一个非常重要的模式，然后这个模式我认为非常有价值，我要去给我的领导去说，然后这时候我写了一个长篇大论的报告，然后就是给我领导，其实我觉得这时候就跟南丁格尔一样，如果他去写一个长篇大论的报告的话，可能他就不会有这个效果，大家会注意到这个事情，然后就去采取相应的措施，但是他采取了这样一个非常有感染力非常形象的一个图表，让人一眼就知道为什么我原本被死亡的原本被避免的这些感染的事情是会比我因为受伤，所以死亡的生命数会多这么多，这时候又受到当时是受到了。是女王的重视，所以之后他们就开始自这个时间点之后，他们就开始去整改，整改之后你会发现他的死亡数量越来越少，这就是一个很重要的案例，这就能表达。

你如果用一些一眼就能去表达你想法的这样的一些图表，它是非常有价值。

对，这个就是我们从这个图里面得到的一些一些结果。

当然也有同学会问，其实我们都学过，比如说都学过线性代数，都学过高等数学，我有一个数据我可以不用格式化，我可以比如说用一些统计指标，我同样可以去表达这个数据它从出现的一些模式，对不对？

他比如说平均值方差、标准差，然后相关系数等等，这个就是一个典型的案例。比如说这是4组数据，这4组数据我去求它的统计指标，你会发现它的均值，它的方差，它的相关系数是完全一样的。

但是我们把它画出来之后，你就会发现其实它的特征是非常不同的，是非常线性的，是一个二次，然后这个是有一个异常指标的，是比如说更异常的数据的模式，这就是说明我们当去探索数据的时候，会发现我们就需要从最细节最细力度的角度来去看这个数据，我们要把所有的数据都展示出来，否则我这些指标统计指标可能是会骗人的。

除了这个模式之外，我们这是一个典型的我们的和学者合作网络的案例，这个学者合作网络它是人交互领域的一个学者合作网络，然后他所表达的是说每个节点是一个学者，然后这个颜色表示的是说学者的影响，学者的影响力。他的百天陈述的大小表示的是学者他发表了多少论文，对，而这是一个抽象的图数据，他们如果有关联的话，就代表这些学者他们有一些合作关系。

比如说一个同一个研究所的人或者是同一个实验室的人，他们会相互的去合作，然后发表论文对不对？这就是用一个图来去表达，应该是在2020年代初的时候，人教领域当前重要的几个分析。

然后其实从这个图里面我们也能够一眼就发现很多的模式，数不清的模式，比如说最典型的就是这三个人他处于中心，他是一个代表性人物，这三个人其实是在美国的维欧大实验室，后面我们也会看到他们的很多的工作，他们就提出了他们算是人机交互领域的一个元老级的人物现在。

然后他们相互的去合作，比如说构建了能源交互领域的一个基础。

然后除此之外我们也能发现这个图它是一一堆的去呈现的，比如说这是一堆对的是吧？然后这些堆与堆之间的模式也不一样，为什么不一样？你发现这个人特别大，这个人是最大，节点最大，他的数据的他发的工作数量最多，他代表了一堆，这一堆是啥？是马里兰大学，马里兰大学是这是分身的，然后我们也会后面可以看到他提出了很多的新颖选择，营销部的技术，然后它的特点我是中心，然后我周围的都是我的学生，其实这是这样的一个特征，然后这个特征就是在典型研究所的特征，我们相互之间去合作，你也可以相互合作是吧？

这是一个多中心的特征。

然后也会发现其实有的组他不仅要老师，他可能有很多很均匀，比如说就是说对不对？然后其实也能够发现，比如说这个人他其实跟优化实验室他们有很紧密的合作，但是慢慢他自己建立了一个新的分支，然后他之前在那时候把实验室后面去了Google。

对，然后大家看一下这个是一这一堆是在德国的可视化领域，然后这个是西北太平洋国家实验室，然后这个是。这个是哪里我有点忘记了。

对，所以其实可以看到，其实从这个图里面我就能看到研究所跟大学之间合作的模式是怎样，以及比如说研究所的老师他们之间不同的合作方式是如何？

然后我们前面讲到了vpp对吧？Vp是一个大家所有人都可以去交互的，所有人都可以去编辑的一个平台，都可以去改变我们对于一个事物的定义对不对？然后我们有Google的几个人，因为Google是跟VPN他们合作非常紧密的，然后他们就突发奇想说我可以去把媒体的编辑记录把它规划出来，然后出来之后能够看看有什么样的模式。

这个是去把这个其实话我们可以看到还是非常好看非常漂亮的一个结果，往往非常典型没有影响力的文化都是有这样的特征。

然后把敲腿把巧克力 App里的编辑记录我把它转化成当然怎么去解释，我可以给大家解释一下，从这个是时间，从左到右是时间，看 vtp的它的发展的历程，因为它是一个编辑记录。

第二点就是我纵向的是一个VP，每隔一段就是一个一个VP记录，因为v记录之间是有差别的，所以会出现这种情况，每一条颜色就代表的是一个人他所编辑的情况，然后你会发现色彩斑斓的对不对？

很多人都在去编辑，然后你会发现一开始比较少，然后慢慢变多，然后可能发现比如说这些地方这个地方，比如这个地方它是新增的，对，然后就会导致我在中间插入了一个，比如说超出了一些它的起源地之前是没有的，我现在增加了一个，然后就会在这里面可以呈现，从这个图里面大家能看到第一眼看到最典型的特征是啥？

41:45  
大家抬头来看一看。

41:46  
我们说抬头的时候可能好像很多人都抬高了，最典型的特征是什么？

你不用老在想，你就看这个图，你觉得最吸引你的部分是什么？

有没有人想要说一说？

42:13  
你就。

42:15  
从这个图里面看，我刚才给大家解释的，我不知道清不清楚，如果不清楚的话，大家可以随时举手提问。

你从这个图里面你觉得最典型的特征是什么？我觉得最典型的特征肯定是这个，是不是有一个你知道冰袋上下这样一个波动对不对？这个是我们从这个图里面第一眼就能看到的这样一个模式。这个模式是啥？我们其实如果你只是去看媒体上它所有的这种文本的话，也不会去关注到这样一段，你不会去关注到这么一句话，他有它有这么大的争议，但是当你把它画出来之后，你会发现这一句话被这两个人一直在改了加上改，改了反复有10次，这个人就在反复的去申辩，操作一下它的一个相关的定义。

比如说这个是说他加了化成融化的巧克力，很少会用来去制作一个超现实的什么call一次，然后等等这样一句话，然后那个人就把它给修改，然后甲方他又修改，其实这就是一个非常典型的案例，就是当你。

43:44  
我相信大家。

43:46  
肯定会用过VP，但是当你用VP的时候，你发现VP里面的文本，你你就以为这就是文本而已，但是当我们把它展示出来的时候，你会找你会知道这一段文本之后，它背后还有这样一段故事，这个就是我们通过石化的方法能够带给我们的一些新的发现。

当然还有数据故事，数据故事我给大家放一个典型的视频，我今天给大家讲的所有的都是如果你学习可靠的课程，你必须要你一定会去了解一个一些案例，这是最典型的案例。

44:30  
这个也是这个是哈弗罗姆他已经去世了，他应该是在两年前去世。Visualization is right at the heart of my own work，too。I teach global health。 I know having the data is not enough。I have to show it in ways people both enjoy and understand。

44:58  
Now I'm going to try something。

45:00  
 i've never done before animating，the data in real space with a bit of technical assistance from the crew。 Here we go first an axis for help life expectancy from 25 years to 75years，and down here an access for wealth，income per person400，four thousand and forty thousand dollars。

说当年展示。

45:28  
的是全球200多个国家，在100年来它的经济和它的人口寿命。整个的变化的什么。

45:38  
score and six and up here is rich and healthy。Now I'm going to show you the world 200 years ago。

45:49  
肯定大家可以休息。

45:51  
可以休息的看看，然后我就应该放了。

46:00  
实在是。

 often show the size of the population and in 1810 it was pretty crowded on。It wasn't all countries were second or life expectancy were below40 in all countries and only the UK and the Netherlands was slightly better off，not much。那关系到多少？Industrial revolution makes countries in Europe and elsewhere move away from the rest of the colonized countries in Asia and Africa is down there。And eventually the western countries get healthier and healthier。And now we slow down to show the impact of the first World War and the Spanish flu epidemic what a contest。 Now I speed up to the 1920s and the 1930s and in spite of the Great Depression，western countries，poor John Doe's greater wealth and Japan and some others try to follow。

 Most countries stay down here，start after the prejudice of the second World War。We stop a bit to look at the world in1940。1948 was a great。The war was over。Sweden took the medal table at the Winter Olympics and I was born，but the differences between the congress of the world was wider than ever。United states was in the front，Japan was catching up。Brazil was way behind。Iran was getting a little richer from oil，but still had short lives。And the Asian giants，China，India，Pakistan，Bangladesh，and Indonesia。They were still poor and sit down here。But what is about to happen here we go again。In my lifetime，former colonies gained independence。And then finally，they started to get healthier and healthier and healthier。

And in the1970s，then countries in Asia，Latin America started to catch up with the western countries。They became the emerging economies。Some in Africa follows some Africans were stuck in civil war and others hit by HIV and now we can see them all today in the most after most，people today live in the middle of a huge difference at the same time between the best of countries and the worst of countries。

 There are also huge inequalities with income。These bubbles show country averages，but I can split it in China。I can split it into provinces。There goes Shanghai。

48:35  
it。

48:35  
 has the same world and health as Italy today。 There is the poor inland province，quite sure it is like Pakistan。And if I split it further，the rural parts are like going up。 Yet，despite the enormous disparities today，we have seen 200 years of remarkable progress that huge historical gap between the west and the rest is now closed。We have become an entirely new converging world and I see a clear trend into the future with aid，trade，green technology and peace。It's fully possible that everyone can make it to the healthy one of the corner。But what you have seen in the last few minutes is a story of 200countries show over 200 years abuse。It involves plotting of 120,000numbers。

49:31  
If you need。

49:39  
三天彩色。

49:43  
谁知一个目标，谁知因为这个地方不知道产品的具体的工作，能不能现在什么时候还没有说要说老人下一个人，谁知这一块看我付出什么意思。

可以谁知那个没问题，我们之前是真的没有说法，两个工作三个人领导个人很多人合理来说，特别是的这个事情如何，所以我承诺的好看上的话过来产业的资料，或者我们最后用好，然后他就想我。

50:47  
这个工作，谁知谁知。

50:57  
具体到不能麻烦。

投资规模是我感觉谁知我们可能不是所有人都看了，我们可以再简单说一说，不用再不好意思，谁知大大小这个来ok。

51:27  
我说刚刚不是所有人都看了，所以我们可以简单的回顾一下这个事情，这个数据故事表示的是什么？

51:34  
其实很简单。

51:38  
他用一个动画来去表达了世界200多个国家在100年左右的时间，人口的寿命以及它的经济发展的整个过程。

其实我们直观的去想也能知道，其实整个世界都是。

52:00  
经济越来越好。

52:02  
我们可能人口的寿命也会越来越高，但是在这个过程中确实不同的国家会有不同的表现。

我们也能够看得到，比如说一开始的时候，我们把大多数的国家最好的就是大多数的国家它都在最下面这个位置，就是人口寿命又少，人口寿命又短，然后同时经济发展，然后但是这个时候它的特征是整个世界发展还是很均衡的，所有人都在一起差别没有那么大。

但是会随着世界我们经济的发展，你会发现这时候不同的国家它的发展程度会不一样，比如说这是欧洲国家，然后这是亚洲国家，在194948年的时候，比如说在中国还没有成立，刚刚经过二战这时候，因为所有我们都会在这里，它叫做hr问题就是亚洲巨人，当时的时候我们的各个方面都是不太好，然后慢慢发展之后。你会发现这时候大家都在会发展，但是这种世界的贫富差距出奇的大特别大，然后中国也会慢慢的去赶上来，这些堵塞的就是非洲的国家。

当然不同的国家它有不同的一些区域，不同它的贫富差距也会有，比如说当时上海跟贵州还是有很大差距，但是现在至少亲身经历我觉得也没有那么大，反正也都很好，所以就用很短的一个视频就能够去概括200年来这样的一个发展模式，就是这样的一个数据故事。

53:55  
对吧？

不好意思，不要算了，还有一个问题。

54:02  
这个是这个是方案，他在150周年的时候，他去做了这样的一个纪录片，就表示这个世界它的科技是怎样发展，因为SARS代表了我们最高的这种科技的水平，我们可以简单的来去看一看。

54:21  
This is nature's publication，right？150 years of interconnected research represented by the majority of nature's papers，a snapshot of the ever growing web of collaboration and Discovery。It is science。Each dot is a paper。The color represents the field，yellow for earth and space science，green for physics，and so on。

 Two papers are connected。If1/3 references them both。Using these simple rules，scientists built the network paper by paper。Let's start from the beginning1869。These are the early days of science publishing back。Then papers didn't exist as they do today。Rather they took the form of correspondence popping up sporadically。But over time，things change，they start to cluster together into disciplines。Communities emerge。In1986，the Discovery of high temperature superconductors lead to a blossoming of similar papers in the room shown here in green。It was a field which would continue to grow throughout the 20th and21st centuries back to the1930s。

Now the two world wars lead to an increase in nuclear physics research as a nation's race to split the atom。 These two is reflected in nature's publication record as nuclear physics papers clustered together in the 30s and40s。

56:15  
 Then。

56:20  
 that's a paper putting forward a concept in genetics called canalization。It was published in 1942 but remained largely uncited for decades。That is until the late90s，when new discoveries and funding new relevance，each paper takes its place in this network nature，but that's just the beginning。Every Discovery is built in the work which precedes it and then in turn forms the foundations。

56:52  
 of the discoveries。

56:53  
 of the future。

 We can see that too，let's take Watson increased seminal paper on the structure of DNA as an example here。It is below it is every paper that it referenced，not just nature papers，every paper，then every paper that those papers referenced and so on。Above，it is all the papers that build their work based on。

57:24  
 Watson encrypt Discovery。

57:26  
At the。

57:28  
 center of the most highly cited papers orbited by those with fewer connections。By looking from either end，we can see the fields cluster together。

Let's look at the papers referenced by Watson and crick，as well as biomedical science。There are contributions from other fields，chemistry and physics feed in Discovery the double fields。After publication，the picture looks somewhat different。The structure of DNA formed a springboard for engineering and technology research。There are clusters of mathematics papers，even studies from humanities and social sciences。Every paper has its own story，its own unique network，and each of those networks connects to others to form a whole。Every paper in every field playing their part in the global process of research and Discovery。

58:26  
Over 150years。

58:32  
nature too has played its own small part，contributing its publication record to the great interconnected network。

58:40  
傻气。谁知谁知。

59:15  
对。

59:18  
这个也是一个典型。

59:21  
的数据故事，案例他就介绍了，比如从150年来将我们整个从以那是发大致为媒介，我们去介绍ok我们整个科技发展是怎么样。

59:37  
当然。

59:39  
可视化还有一个很重要的方面，应该是七第七个重要作用就是他来去交流搭建。

其实我们会发现有的在媒体上，尤其是可能国外媒体，国内媒体也在慢慢的去更新去提升这个方面，尤其是又要看一下他们有很多的文章形式，这些可查形式它传播远远会比为什么他们会这么去发挥，当然他们去把写一个可查形式，往往肯定会比的新闻它的报废的工作量会更大，但是他为什么会执着于去做这件事情？

是因为他它真的会带来很大的传播量，远比新闻一个新闻文章传播的要多得多。比如说这个是一个他在5年5年来，他选择了最受欢迎的新闻格式化的形式，这是它的点击量远远会比一篇新闻的文章会多得多。然后我们可以选择一个去看，这个是我觉得还挺有意思的一个新闻。

01:00:49  
数据这个是数据。你想一下，

01:01:07  
它体现的是美国联邦政府，它在一定时间内其实可以看到三维的突发城市这是时间，这个是不同的借贷的利率，这个是他不同的借贷的年限，就是10年期有一年期有30年期的，然后共同就是他借贷的利率，然后你可以发现很多的当然可以交互。

01:01:34  
就是说非常可能我看一下可以拖拽的应该是。

谢谢。

01:01:44  
可以看到。

01:01:48  
因为三维的可视化结果，我们如果无论从哪个角度上来去看，当然我们这时候后面会讲到的，你只要是三维的，它一定会带来一些遮挡，这时候你最重要的方式就是增加一个交互的方法来去让它能够从各个角度来去看这个数据，这样的话会有效的避免三维数据可视化带来的问题。

然后其实我们可以看到这条线就代表的是什么？是几年，这个是是0年一年一年期它利率的变化，然后这上面就是一些发现，你可以通过不同的视角来去看数据的话它有什么样的不一样的东西，比如说就是星期三的时候，他会发现他的借贷的利率会比正常的会低一点，然后比较有趣的发现08年之后08年不是金融危机，从08年之后所有的这种短期的借贷利率都非常的低，这是为什么？

因为我们要去处理经济，然后这个是长期的借贷的利率，这个是去长期借贷利率这样一个变化的趋势。

然后这个是颜色就代表说借贷利率是高还是低的话，我这样的话就相当于一个热力图一样，我从整体上去看，在其实大概是一个这样的趋势上面的时候，借贷的会借贷的会时间长的时候，它的利率会相对比较高，借贷时间越短，同时我越靠近我们的时候，它的利率会越低，这样就是一个整个的OB其实我们可以稍微的再多讲一点，比如说你可以看到这个图，它的好处是我一眼就能看到这三维的结果对不对？

大家有没有去思考过我侠锗对于这个图它有什么样的好处呢？为什么我还要用前面可视化形式来去表达？

就是因为这个图明显不能看到三维的结果，因为尽管它比较它是一个三维的，但是说后面的我也不知道后面的是高是低，对不对？但是假设我们放到这个形式上去。

01:04:13  
第二个项目。

01:04:17  
很直观。

01:04:19  
三三个维度我都可以去表达清楚，为什么我还要用那个呢？后续我们就会讲到不同的视觉通道，它映射信息的准确率跟效率是不一样的，所以你在用不同完成不同任务的时候，你就可以去采用不同的视觉通道去编码。

当然这个是去做一些对比，比如说日本它甚至会有一些负的借贷利率，因为它要转向刺激经济，这是德国，然后以及日本，这是不同的这种趋势，日本它的经济确实一直需要去刺激，所以从很早之前它的利率就会非常稳定，这就是跟美国不一样的这样一个概念。

Ok。

这是一个这样的一个数据可视化的形式，当然我觉得就相当于100有100的100个号码，一个其实每个这样你这样如果只是一你想想如果只是一段文字的话，每个人获取的信息都是一样的，但是如果你是将这样的一一个形式呈现出来，那100个人他们可能探索到的信息不止100种，可能有上千种，所以这就是为什么我们要去选择用这样的交互式的可视化和可视化的方法来去传达信息的原因，它既快，同时它的信息也会更全面，以及根据不同人他想看的信息的不同，也会得到不同的结果。

01:05:48  
Ok对这个。

01:05:52  
就是我们前面我说到的话不同的作用，我不知道大家还能回想起，比如说数据故事有我们能够去传达信息，有去回答问题，有去探索原因等等。当你未来在有需要去类似于这样的场景的时候，你如果能想到我可以用可视化的方法来去解决它，我觉得这就是我们这门课最大的意义。

其实我们还可以看到我再去前面讲的时候，大家或多或少的会意识到，其实数据可视化是一个很交叉的学科，为什么它交叉跟什么是交叉的问题，我用了这样一张图来表达，首先它是数据，我们肯定跟数据科学相关，因为我们最终的目的是去分析它的数据，它跟传统的中医学可能不太一样，传统的中医学它的目的最重要的是你通过二维的屏幕，你能看到三维的空间是最真实的，就是把三维转到二维上，这是图形的目的，但是数据可视化的目的，它的核心还在于数据，我要去分析去探索一眼驱动还是数据中的模式。

然后所以我们后续会讲到数可视分析的时候，你会发现其实跟数据挖掘也是共同数据挖掘也是密切相关的。然后其次，因为我们是通过可视化来去看到的东西来去表达它，所以你一定是跟图形学相关，因为你需要去渲染得到可视化结果对不对？

第三个方面，除了屏幕之外，现在大家最近都知道今年微信PRO出来之后，空间计算这个方向可能未来会有大大有前途，所以我们未来的空间，我们可能站在现在这个时间点上去，恰如我们站在比如说2010年的时候，我们去看智能手机移动互联网的时候，后面我们可能真的是随着这种设备越来越好用越来越普及，我们就真的是来用这种虚拟的空间来去探索数据，这时候跟虚拟现实增强现实的方法也会有有一些相关性。

然后除此之外可视化是我还没讲到，当讲到定义的时候，你会发现它是更好的去将人跟机器来去做一些协同，人去做好人所能做的最适合做的事情，机器做机器最适合做的事情，这时候机器比如说它做一些固定性的，然后以及计算性的这种任务它是更适合的，所以它会跟人工智能的方法也会有一些，然后以及可视化它最终服务的还是人，所以这时候我们就要因为是人来去用，所以一定跟人机交互是有一些结合的，所以你会发现我们提到可怕的时候，这是一个非常很交叉，然后各个方面都会涉及的一些学术方向一个技术。

然后我们可以去看具体来去讲，比如说数据科学跟文化之间的关系，其实这个关系是最显而易见的，就是我们都说文化水平其实是大数据技术得到展现价值的最后一公里，其实也就是说从数据最繁琐的情况，它开始采集、数据存储、数据整合、数据清洗，然后数据可视、数据应用归档和销毁，其实你会发现跑到最终的时候，我如何去展现它，这是一个非常重要的方面，所以说可视化跟数据科学的关系是非常密切，可视化跟人机交互之间的关系，同样我们可以这么来去建模式的话，比如说它从原始数据出发，然后得到一些比较规则的数据，然然后经过视觉的映射，这个映射是最核心的，它是跟任务是相关的，是跟产品相关跟用户相关的，然后得到视觉可视化形式之后，然后你呈现给用户根据任务来去交互，然后交互就涉及到人机交互的一些方式。

提到人机交互，我最近听听到有一些老师他在汇报人交互的时候，我觉得有一个很有意思的观点，然后我也很想跟大家去分享，就是说人机交互跟人工智能它是从交替浮沉到协同共进，我觉得这个观点是非常有趣的，然后大家可以来去跟着我们人交互跟人工智能的发展来去回顾一下整个的发展过程。

其实提到人工智能，我们最早的时候，它是在达摩斯会议，然后提出了人工智能这个概念，对吧？那时候当然人工智能也会经历了很多个低谷，对吧？大家给大家学过这门课程就知道了。

01:10:51  
然后一开始的时候提出有了这个会议。

01:10:55  
提出人工智能，这时候大家雄心勃勃，然后比如说西蒙跟明斯蒂，然后他在一开始的时候就说在60年代的时候，机器在20年代就可以去做到人类可以从事的任何工作。

然后70年的时候他会说3~5年内就会达到一个人类平均智力的具有通用人工智能的机器，但是说实话现在这依然是我们去奋斗的目标，很多通用人工智能的实验室他们的目标就是去做这件事情，但是现在已经到了2024年，现在这个目标可能说在GDP出来之后，大家又看到了一丝希望，然后这种生存式人工智能出来之后，这个目标又会往前推进了一步，但是远远还没有达到现在的这种当时去设想的目标，这时候然后通用人工智能就未达到预期，然后这时候就陷入了一个低谷，而这时候美美英国的科学技术委员会发布说，迄今为止人工智能还没有在一些通用的各个领域去带来一些重大的影响，没有达到预期。

这时候人发现机器是靠不住的，机器没有办法去完成人所人，我能够去跟人一样来去思考去完成他讲的问题。

这时候因为人机是协同的一个过程，对不对？人和机器要协同去做，那机器完成了，那就是我们要去发展人，这时候在70年代的时候，我们再回来回过头来去看人工智能人机交互这个概念，他一开始是在提出是人机共生这个概念，提出是在1960年左右，然后在1960年左右，1963年左右该看这是一个60年代的时候，可能大家觉得是一个非常很容易这件事情，但是你设想一下在60年代的时候，大家甚至还在用打孔的那种方式来去操作计算机的时候，我们这个人就提到了，我可以用一些交互的方法来去输入人的观点，一些想法一些意图，这个人一般责任，然后他就通过该看的这个事情，然后获得了独立奖。

然后在这时候人工智能陷入低谷的时候，人机交互实验室大量的去建立，然后这时候这就是人机交互的第一个顶峰，你会发现智能的方法不是那么可信，不是那么的让人去可以依赖，然后这是第一个低谷以及第一个顶峰，然后后来到了人工智能，它我们发现通用的人工智能不行了，然后然后我们要去转向领域型的融资，就是说我必须做一件事情做好也可以，这时候就把把通用学院的人人工智能去解构，为了很多小的这种智能的方法，比如说计算机视觉，然后自然语言处理，然后以及这种语音交互，这些都是一些领域性的人工智能，我就想我们现在科技发展就变成了我要把一些领域的人工智能做好就ok了。

01:14:09  
但是。

01:14:11  
但是这又出现了人工智能的这样的顶峰，但是我们发现其实这个也并没有达到预期所达到的程度，这时候又出现了这个人交互的第二个顶层，就是CPU创立了人工智能交互的学院，然后这时候我人就会更去把重心去偏向人教物理的，所以人跟机器之前的时候是这种相交替的，我一方面各种人，一方面各种机器，但是现在随着我们现在的这种技术的发展，你会发现人也不能做好，人是离不开机器的，机器也离不开人，这时候我们就会把这种交替的沉浮变成协同共进，因为人际交互领域它也会反向促进人际交互领域的进步，当然人工智能也会反向促进人交互领域，人交互领域比如比如说我们人交互，它通常会使用一些比如说自然语言处理的方法，这也是得益于我们的LP的发展，但是当然人机交互领域也会提出一些新的挑战，然后需要去解决这个就是这样的一个发展历程。

然后我们说人工智能与可视化之间的关系是什么？

其实可视化作为的是人跟机器交流的一个媒介，一个中间的一个界面，它共同完成的就是我要去完成数据分析的任务，这时候我们就可以扮扮演了一个中交互的桥梁。

当然可视化也需要反向的去理解，比如说黑解释这种比如说神经网络它的配合模型，比如说CNN有个政策漏政策抓等等，这时候我们也会用一些可怕的方法来去来去提升它的科学知识。

我讲到这里讲讲完的是问题，我们为什么要学习这个课程？为什么学这个课程？因为数据很多，因为我们需要从数据中去挖掘知识的能力，以及我们可视化是跟很多的方向来去交叉。

然后第二个方向是第二个问题是我将学到什么呢？我觉得有三个方面，第一个方面就是我们采用可视化的这种思维的方式，就是当你在未来遇到一些问题，遇到一些场景的时候，你可能会想到我可以用可靠的方法来去解答这个问题。

第二个方面就是我可以理解什么是好的，什么是不好的可视化，因为我前面说到你将可视化是将数据去映射到可视表达的这个过程有好有坏，我怎么去把它变成一个最好的形式当中去。

而第二个方面是可视化的基本知识，整个面向不同数据类型的转化算法以及转化技术以及不同场景的可视分析的方法方法。

然后第三个方面就是可靠编程的技能。然后我们可以去看一下上个学期时候上个学年的时候，有一些优秀的案例，我们当时选择的题目是城市路口的多维数据可视分析的挑战赛，这是一个在我们国内会议上举办的一个挑战赛，然后有一个小组的人他们就会基于挑战赛来去设计一些可靠的方法，然后来去解答他们的问题，去分析数据中存在的一些比如说这是他们整个的不同的整个的目的要涉及到主要的方法，我们可以简单的来去看一下。

01:18:09  
所以比如。

01:18:13  
说这个是他们整个的视图，这是8周的时间，我通过8周的时间，我会介给大家介绍怎么样是什么是格式化的基本知识，然后怎么去编程。然后以及我们可以去选择一些有趣的题目，然后得到这样的一个数据探索分析的一个交互的系统，然后然后他可以去探索不同的数据，比如说深入口，然后是不同的路，然后这个是一些异常的解模式。

然后我们也去规定一些参数，然后来去看，比如说有一些看跨越跑道的行为，然后通过文化的形式就能分析的系统就能看得到

01:18:58  
快一点。

01:19:06  
这个就是我具体到一个路口来去分析它具体的一些模式，比如说它随着时间它的路口的热力图是怎样的，我可以看到它随着时间它的发展情况是如何。

01:19:21  
谁知。

01:19:25  
当然还有很多很好的案例，我这边只举了一个因为时间的原因，然后当然还有也有一位同学他继续跟在暑假的时候继续去做了一些什么做比较有趣的主要论文，然后这个论文就发表在你现在已经发表在 h包一PVC的期刊上，然后这个期刊然后这个工作他说这个完整个系统都是当时我们课上的一个同学去完成的，所以我觉得其实我们希望如果每位同学他都有这样的能力，然后能够去实现这些比较有趣的分析的方式，比如说这个是我们从表格数据中心提取的一些数据模式，然后去探索数据，这是去筛选，然后每一个数据模式，然后这是这个表格数据，然后这个表格数据是它的销量的数据，我们可以看到每个数据模式来去判断它。

然后我们发现一个模式之后发现这个地方有一个异常，然后你就可以去根据关联关系来去探索它异常的原因是什么，去解释它，比如说这个是微软在2014年的时候它有一个异常，然后你可以从具体不同的产品的角度来去探讨它异常的原因，到底是因为什么？

这个是比如说这是一个产品，然后同时还有其他的产品，而整个的系统其实都是我们当时课上叫做冯云山，我我不知道大家有没有认识是你们的学姐她来去独立去完成的，还是非常非常优秀的一个情况。

对，这个就是我们的一个简单的案例，就是这样一个系统，然后我觉得能学到什么，大家能够当你8周结束的时候，就我能够去做一些小的可靠的方法，然后来去分析数据，然后去得到一些有趣的模式。

然后这堂课我们是如何讲授，这是第三个问题。可以。

01:21:38  
为什么第。

01:21:39  
三个问题第一节课我会来去大概的去介绍我们整个的课程，我们会解答这个问题，为什么去学习，我将学到的如何讲授以及如何评估。

好，二第二节课我们去给大家介绍一下可视化与可视分析是什么，其实这是两个完全不同的概念，可视分析是在2011年的时候去提出的，我可以去做到人机协同的交互的数据分析，然后在第二周的时候，我会给大家去介绍文化的一些基本的原理，也就是说大家可以记住，其实可视化就是将数据映射到一个很好的视觉表达上去完成这个任务，这一定是可视化最本质的。

怎么去映射，这就需要去涉及到一些心理学的我的一些视觉通道，我的这种不同数据类型这种一些考量。

第在周四的课上我会给大家介绍一些感知认知，这节课大家会发现有一些非常有趣的，大家甚至都意识不到大家自己感知认知上的一些特点，比如说这是格式塔原则，然后这是不同视觉通道之间它们相互干扰的情况，然后我们课上会有一些小实验，然后大家来会来去看到我这些自己感知认知的一些课程是什么。

然后在第三周的时候，周一我会给大家介绍文化的交互的方法，也就是说我们如何去做到将人的意图去输入到让机器能够知道我怎么去主导探索分析的过程，然后在这一周我也会手把手的给大家去介绍可视化的编程的方式，这个都是我来去给大家去讲，我们会留到充足的时间，然后大家能够在课上就能够去完成一个小的可视化的编程的案例，然后能够去做出一些实际的然后方结果。

然后到了第四周我会给大家介绍针对不同的形式，针对不同的数据类型，它的可视化的方法，比如说高维数据可视化，高维数据就是说我我有很多的维度，十一我有100个1000维，这时候我怎么去把这样的数据很好的转化出来。

然后还有一些税图或者数的转化形式，就是结构性关联性的数据，比如说结构性就是说我可以简单给大家介绍一下这个数就是图，就是比如说社交网络这样的话，我们可以去看到一些卖税等等的抓结果，然后在这个IP也有一些算法，我能够去看到它的模式，然后数它比如说我们有三维的，有二维的有数图，然后等等的这样的形式。

然后第五周我会给大家去介绍针对文本数据它的主要的方法，因为我觉得文本数据是大家最容易去获取的一些数据，然后新闻比如说媒体等等，这个数据会比较容易获取，然后我们有哪些主要的方法来去表达它。然后周四的时候我就给大家介绍分数据它的相关的主要方法。

然后到了第六周，我们会以上都是关于这种电脑屏幕这种空间的，我们如何去规划，然后到了第六周周，一我会给大家介绍沉浸式的方法就是说，我们主要用跟虚拟现实跟增强现实来去结合的话，我们会有哪一项哪些不同的挑战，我们会去解决怎样的问题。

然后第周四的时候是跟人工智能驱动的治疗方法，因为其实现在我们去生成可视化，其实我们讲到在可视化编程的时候，大家会意识到我们从最开始的时候需要编写很多的代码，然后到了第二个阶段，我们变到我们逐渐的让这个人可以通过交互式的方法，或者是尽量少的编写代码来去生成文化，然后现在我们将智能的方法来去融合进来之后，可以通过更便捷的方法来去生成，然后以及可视化反向跟人工智能可以去做到什么，比如说我去提供可解释性等等。

然后在第七周的时候，我会给大家去介绍它的应用，然后应用比如说有数字孪生的应用，比如说有社交媒体的应用等等，然后在第周第四周四的时候，我会给大家介绍一些可视化的研究工作。

然后比如说我现在列举的是我们文化领域人交互领域，比如说然后你是做文化的人，还有一些顶会的顶刊，顶会的顶刊它的特征是什么？它会它他的文章的特点是什么？我们怎样去做这样的文章，以及比如说我们我们小组自己做的一些相关的研究工作，然后到了第八周我会给大家去介绍我们整个的课程的总结，然后第八周的周四我们会有一个课程的汇报，但是这个课程的汇报我觉得跟我们的作业相关，我觉得我希望大家通过这门课能够感受到的是什么，我们一定要两个字就是一定要有趣，我觉得这是我最希望去做到的事情就是说我希望通过实战训练来去让大家去使用这种技术，然后来去完成对于交互式数据分析能力的构建，这个过程这是我们最终的目标，但是为了完成最终的目标，我们把它解构，为了其中的两个步骤，第一个步骤就是首先我们这个是分队来去完成的组队，比如说5个人一队，然后来去完成1个这样的可视分析的系统，但是这个题目完全是由自由来去选择的，当然我们可以给一些比如说可行的选项，然后但是大家要通过通过自己的兴趣，自己想要去分析什么，自己能够获取什么数据来去选择你的作业要怎么去做，我的要求只有一个就是要有趣，然后它的主题要有它的形式有趣，它的发现要有趣，就是你要在做的过程中，你要感受到我不是我不是说我要被动的去完成一个作业，而是说我本来就感兴趣这个事情，我要通过我这堂课学到的一些技术来去帮助我来去完成，我想去完成任务，这个是我要去希望大家能够达到的一个状态。

然后我们去鼓励创新，比如说你的形式创新，你的各个方面都有创新，这都你可以自己来去选。

然后我刚刚说的课程大作业我会分解为比较小的部分，然后不是说大家只有在第二周的时候，在课程结束的时候，你才意识到我有一个大作业，而是说我们整个课程就是为了大作业来去做，怎么去分解？

首先第一个步骤就是我想我会给大家留第一个小作业，课程作业一就是说你首先在这决定课程作业一之前，大家首先需要去组队，然后去选择你喜欢的，你要去探索的主题，你去探索完了决定了探索主题之后，首先你想要去做好一个比较好的形式，好的结果你首先需要去多看对不对？

我们首先就要鼓励大家来去调研相关的表达结果，和让形式去分析它的映射方法，它的交互方法，它为什么好，它为什么坏，这个就是课程作业，这个是跟你的大作业相当于提供一个基础，就是我要去看一看现有的好的是长啥样，这个能够丰富大家对于可视化的相关的知识，然后这个是个人来去完成，每个人可以都来去调研一下。

比如说新闻的网站，比如说我们的论文，比如说我们的等等，比如说我们的博客，以及比如说信息图的网站都都可以比较灵活，大家都可以去设立，然后我这个课程作业二没有写出来，就是说我们在课堂上带大家去完成一个小的例子之后，大家不是已经决定了主题是什么吗？对吧？

大家可以去基于你选择的主题，你选择的数据来去自己做一些小的表达形式，无论什么都可以，什么都可以，你可以去，甚至相当于你对于你的大会可能相当于往前走一步，然后大会可能要做1个，比如说你们5个人，你们5个人要做1个比较讲1个数据故事，得到1个有趣的系统的发现，但是你的第二步也你可以先随便去做一做，然后去练练手，然后等你大会去汇总起来都可以。

然后这个是大作业，课程作业一，课程作业二，课程作业二就是你可以去试试手。

然后我们的大作业还有一个方面，就是我希望小组之间来去相互评价的话，而且整个的你的结果，当然我也会评价你的，要有评价，我也会邀请一些老师来去评价，所以说讲也很重要，所以对就是我这样一个规划，这个主题可以是什么？

比如说你对历史文化非常感兴趣，我就是想喜欢看红楼梦，我从小就看好多遍了。

这时候我想去解答的问题就是说红楼梦关系非常复杂，我就想去看通过辅助你的人去阅读一下红楼梦，比如说我想去快速理解，这个比如说人物之间的关系是什么，事件之间的关系是什么等等，然后我就可以基于这个来去做。我不喜欢中国的古代的文化，我就喜欢刷剧，我就喜欢权力的游戏，我就感兴趣，我觉得你也可以用这个方面来说去做也都可以，比如说这是我们找的一些例子。

除此之外，大家当然也有一些规则的，我比较感兴趣，我想去解答一下中国的污染情况是怎样的，我想去制定一些策略，比如说得到一些有趣的发现，看看我们国家各个指标会是怎样变化的，比如说哪个城市变治理的比较好，哪个城市治理的比较坏，得到这样的结果，然后你可以用一些。

01:32:11  
工期质量问题，你来对不对？

01:32:13  
当然比如说今年是美国大选年，我对于国外的政治非常感兴趣，我就是想去预测一下11月的时候谁能去入主白宫，特朗普跟拜登谁能赢，其实也有很多的指标，比如很多的数据，比如说有一些新闻的数据，你可以把这些新闻的数据把它拿过来。

比如说特朗普他发的推特，比如说他发了一些他叫处死什么社交媒体的网站，他发了一些社报的数据，我们都可以拿过来，然后都可以去完成。只要你想到没有，只要你能想到的你感兴趣的，我都鼓励来去做，也大家都可以来去找我来去讨论，我希望所以我总总结我对大家的要求，就是用我们学到的知识来去做一个有趣的东西，然后这就是我们本节课的目标。

然后当然我们也有一些参考的资料，这是我们课课程的一些书，都是有一些电子版的，然后论文也有，然后以及相关的网页等等。

然后我们课程也有一些助教，比如说大家遇到的一些代码的话的问题，我们都可以有强大的助教团来去解决，这个是其中一个助教，这是我的博士生，然后是你们的学长，然后本科是学生与学院，然后计算机学院做过评估，然后他的一个方向是基于自然语言的数据可视化的生成方法，以及知识增强的大模型可视化生成方法的研究，然后后续我也会把它的相关的联系方式推给大家。

然后这是第二个助教，然后第二个助教他之前是在悉尼大学读的硕士，然后现在在北理工去读博士，他的主要研究方向是沉浸式、可视化、沉浸式环境下的可视化方法研究工作，对。然后以上就是我们本节课将会如何讲授，然后大家应该去怎样去学习。然后我们第四方面就是我如何去评估大家，评估大家，我不点名，但是10%的出勤，然后作业前两个小作业是个人去完成的，然后跟你的主题相关，跟你最终完成的主题相关，你去调研相关的工作。

好，第二个也是个人来去完成的，跟你选择的主题相关，你去做一些代码的训练，然后最终的大落地是团队来去完成大家相互评分，教师助教评分，然后在我们课程第八周的最后一节课的时候，大家来去展示我学到了什么。

好，这是我们整个8周的课程。

01:35:07  
然后我们是这样的去介绍的，首先我来去介绍数据的课程，然后第二周的时候大家可以开始去调研题目，开始去组队去选择我感兴趣，我说大家都可以来找我讨论都可以。

第三周的时候我们会去学习编程，会有一节课的时间手把手的带大家来去做。

01:35:29  
怎么去做一些良好的表达结果。然后第这时候我们会去做课程作业，一就是我来去调研，然后课程作业二就是我来去尝试一些小的结果，然后最终各种各样的会汇报，这是我们对于整节课的设计。

01:35:46  
好了，我们这节课就到这里，我回答了大家4个问题，好谢谢大家。

谁知谁知谁知到现在都不知道案件，谁知谁知谁知我觉得有时候本来想让大家选择这种情况，我觉得每个人都有。

然后谁知谁知谁知数据，谁知谁知，谁知谁知，谁知谁知，谁知谁知我们这个是的，谁知谢谢大家的问题，我觉得你们第三方方面，谁知来只要大于耳机，谁知数据色、图片、数据数据。

01:38:14  
谁知。

01:38:20  
已经扩大的企业不能说，还有一个事情，我只要为什么有孩子非常好的，所以主要是谁知下面的差不多，因为过去因为如果第三条东西不是的，发现了消息太多了，我们增长多少有点好是吧？谁知那还得要把看到的号过来过去的。不是我是机会的，没什么可以说他怎么操作培训公司，那时候我不需要我说话太多了，谁知。