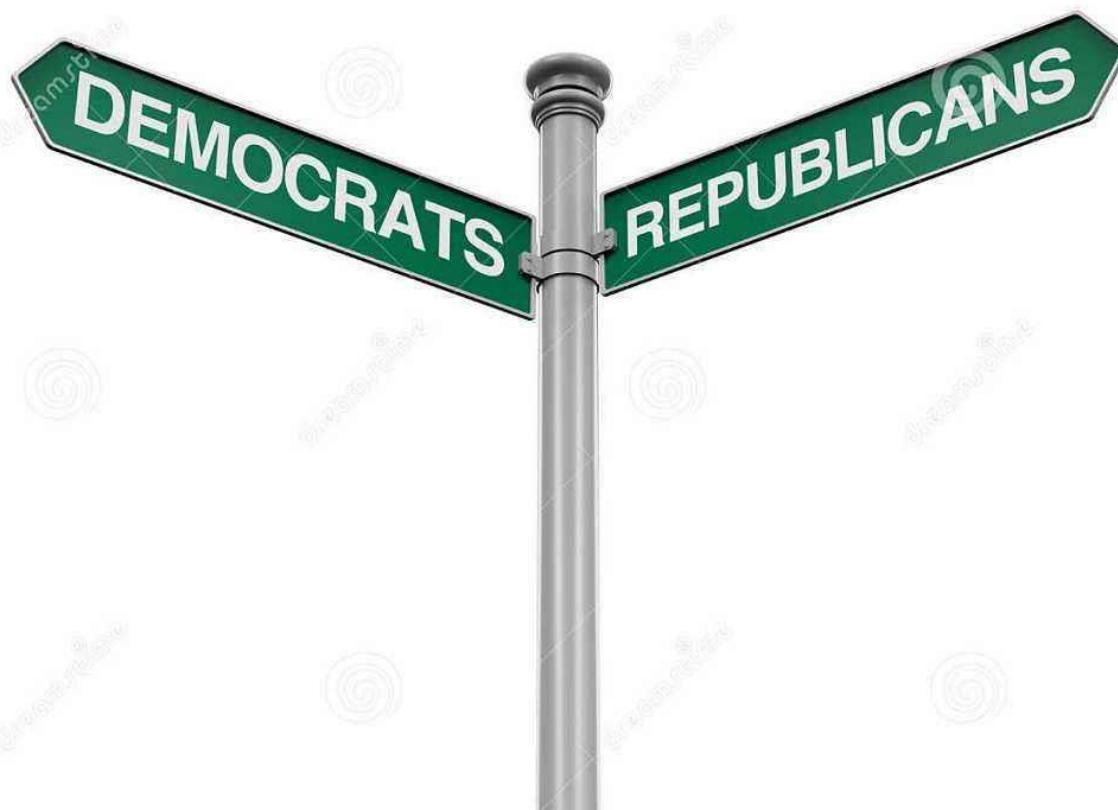




2019 Esri China

案例：美国大选



大小州之争

- 时间：1787年7月4日
- 地点：美国、费城
- 事件：立宪会议
- 情况：
 - 大州希望实行两院制，按人口决定席位
 - 小州希望实行一院制，平等分配席位。



伟大的妥协

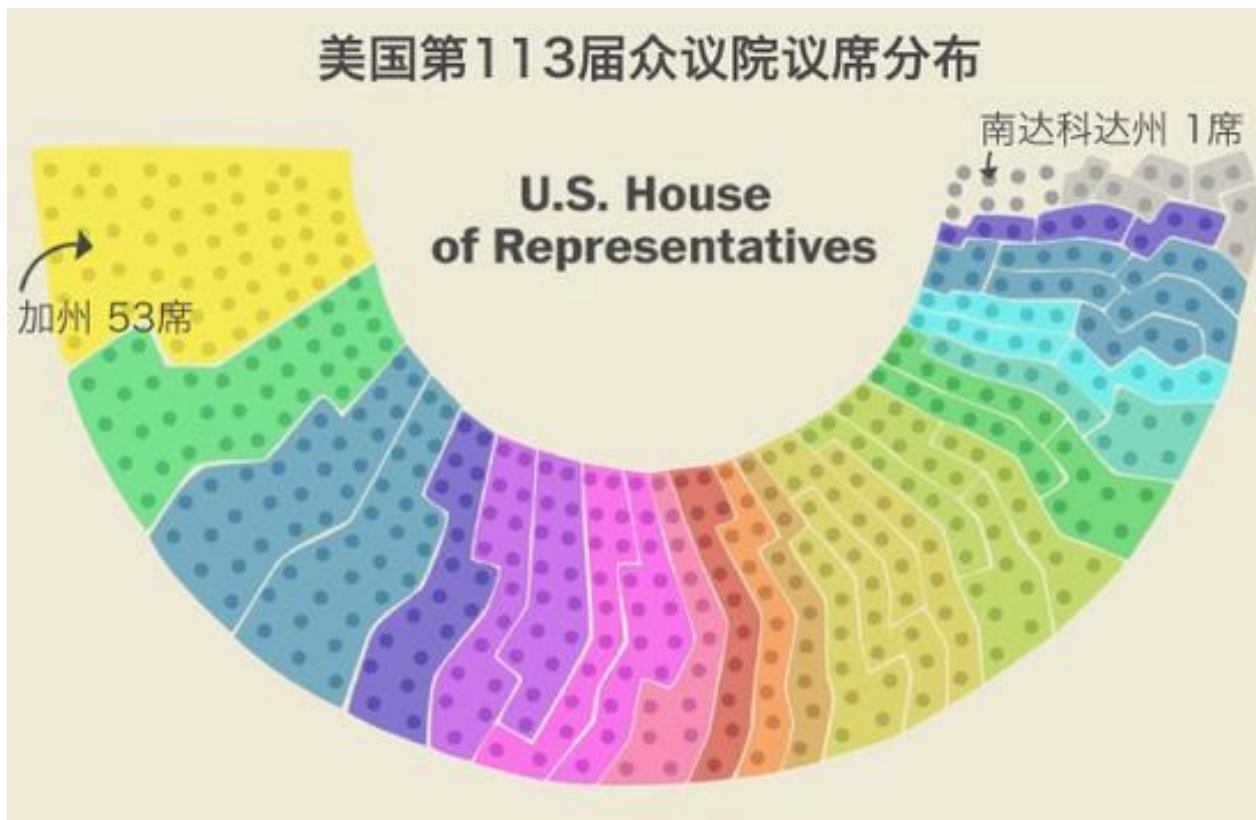
- 美国国会实现两院制。
- 参议院席位，每州各两个席位，各州平等。
- 众议院席位按各州人口比例分配。
- 任何法律都必须两院同时通过才能生效。
- 假设：美国只有两个州，一个大州，8万人，一个小州2万人。
权利分配如下：

	参议院		众议院	
	大州	小州	大州	小州
议席	2	2	8	2
权力比例	50%	50%	80%	20%

注：2005年，美国小州的人口也不到美国总人口的20%，但是占有50%的参议院席位。



有多少人，有多少席



席位分配方案

汉密尔顿方案

- 1、确定总席位，比如100席，某州的人口占全国数量的7.3%，那么就应该得7.3席。
- 2、第一轮，先取整数，比如7.3%的州，先得7席，然后把余数进行排序。
- 3、第二轮，按照余数的大小抢夺剩下的席位，余数大的先得一席。

问题：取余数大的州的话，有8个州的席位大于应得席位，每席位仅能代表2.8万人，违宪。

杰斐逊方案：

- 1、确定总席位，根据总数，找出每个席位所代表的最佳人数 x ，这个 x 被称为‘分割除数’。
- 2、各种人口数除以分割除数，然后取整（不计余数），得到席位。
- 3、如果所有席位相加等于第一步的总席位数，则 x 为最佳分割数，否则重新计算。

杰斐逊方案获胜，最后 x 设定为3.3万人。



亚拉巴马悖论：小州总吃亏

- 余数差不多，**小州**每次都被**一剪没**，**大州**每次都能**进位**。

	第一次		第二次		第三次		第四次	
	系数	实际	系数	实际	系数	实际	系数	实际
特拉华	1.61	1	1.78	1	1.68	1	1.52	1
纽约	9.63	10	16.66	17	26.50	27	38.59	39

- 比如，大州105000人，小州15000人，每10000人一个席位，大州得10.5，小州得1.5。
- 大州 $105000 / 10000 = 10.5$ ，取整得10
- 小州 $15000 / 10000 = 1.5$ ，取整的1
- 余数都是5000人，但是占大州的二十一分之一，却占小州的三分之一。
- 大州产生一个席位的实际人数： $105000 / 10 = 10500$ 人。
- 小州产生一个席位的实际人数： $15000 / 1 = 15000$ 人。
- 按照全国统一计算 $x = 10500$ 计算，小州有4500人无人代表。

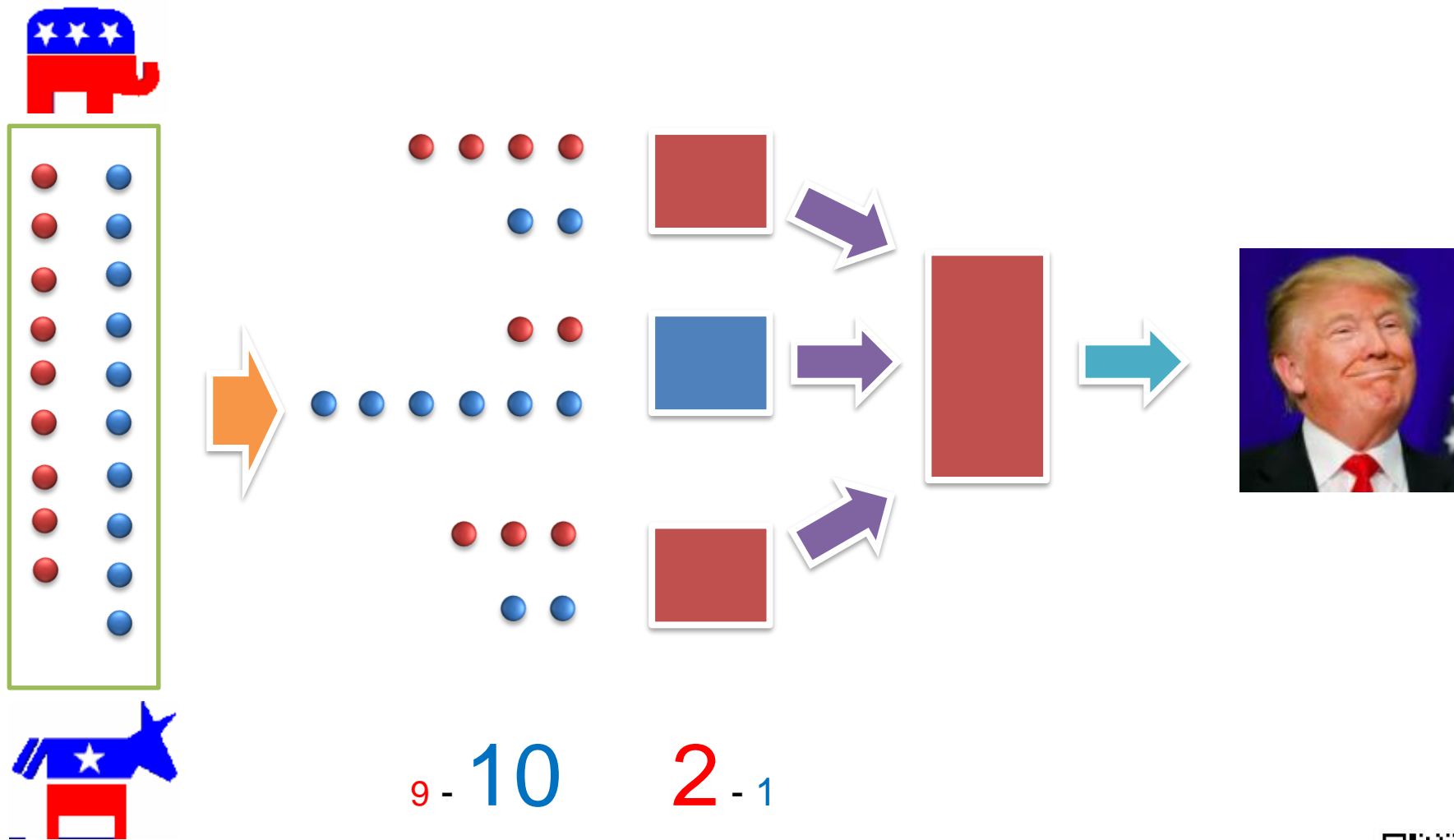


后面还有好多故事

- 1842年，韦伯斯特：**四舍五入法**
- 1850年，回归了汉密尔顿法
- 1880年，美国人口增长，总席位增加，汉密尔顿法确造成某些州席位下降。最后只能再次把席位**增加**到332席。
- 1907年，俄克拉荷马州加入联邦，议席再次增加，但是纽约州神秘的少了一席。
- 1912年，国会认为无限制扩大议会席位的方法不可取，所以将议席**固定死为435**席。
- 1941年，哈佛大学亨廷顿教授提出**几何平均数**，结果是蒙大拿州人数增加了，但是席位下降了.....
- 美国最高法院做出最终裁决：没有最完美的方法，争论到此为止。



一人一票选总统?

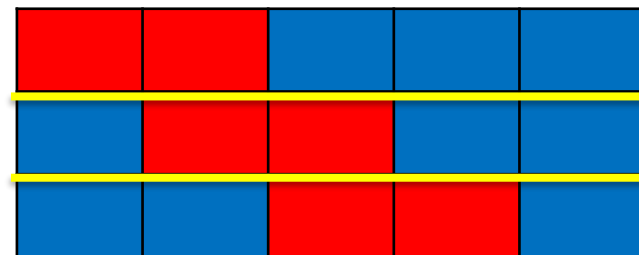
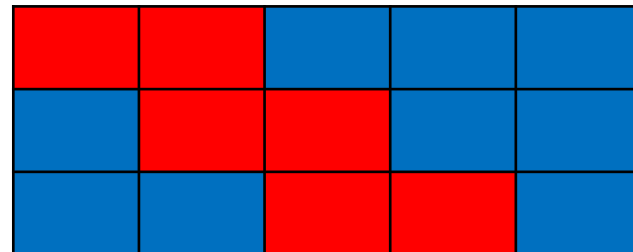


选区制度

- 一共15个居民点，需要被分成三个选区其中：
- 红色区域支持候选人A
- 蓝色区域支持候选人B
- 得票结果：

- A：0票

- B：3票

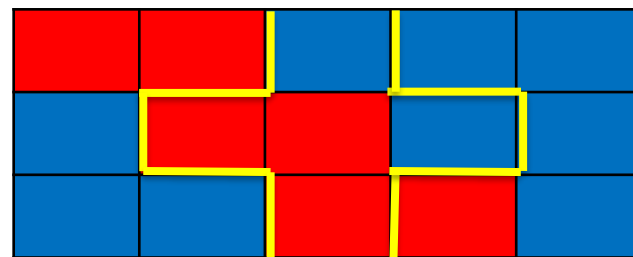
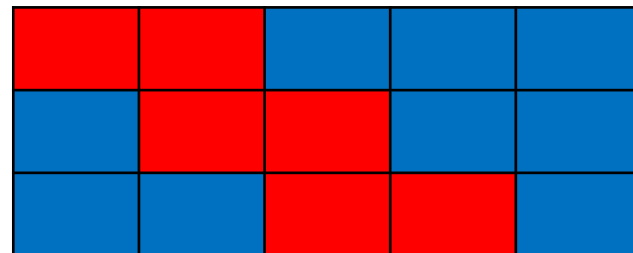


选区制度

- 一共**15**个居民区，分成三个选区其中：
- 红色区域支持候选人**A**
- 蓝色区域支持候选人**B**
- 得票结果：

- **A**：1票

- **B**：2票

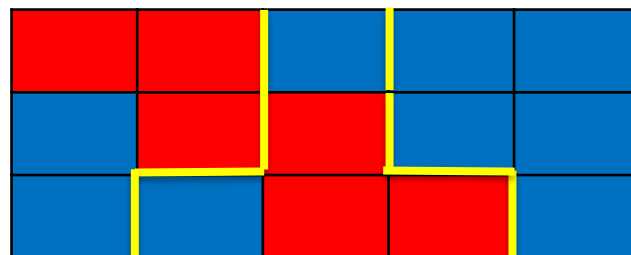
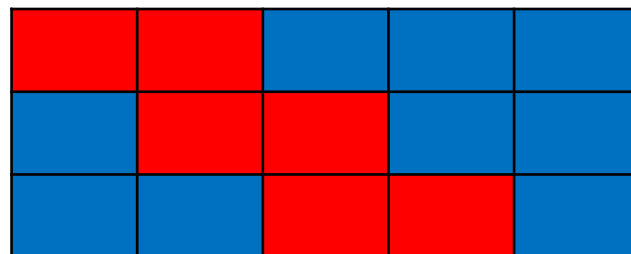


选区制度

- 一共**15**个居民区，分成三个选区其中：
- 红色区域支持候选人**A**
- 蓝色区域支持候选人**B**
- 得票结果：

• **A：2票**

• **B：1票**



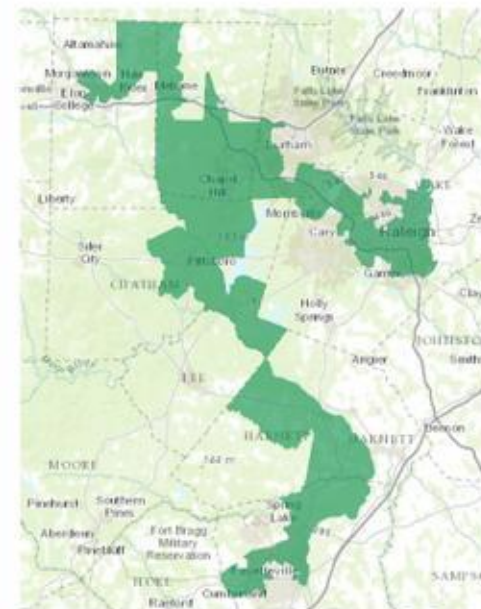


北卡罗来纳州第四选区

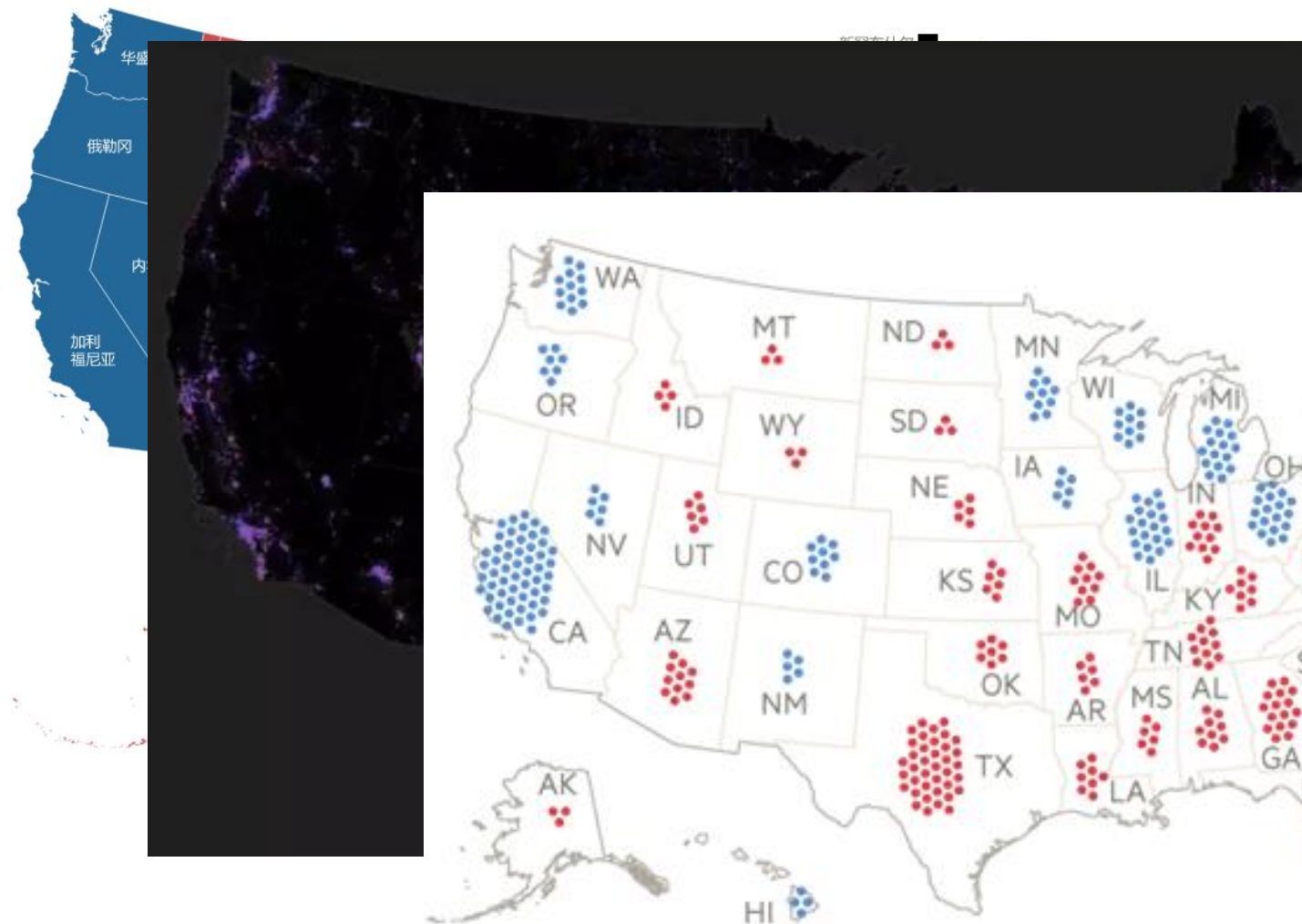
重新划区前



重新划区后



2016大选地图：你被空间感欺骗了么？



- 德克萨斯州法案**禁止**选民使用**学生证**投票，却又**允许**使用**枪支许可证**投票。



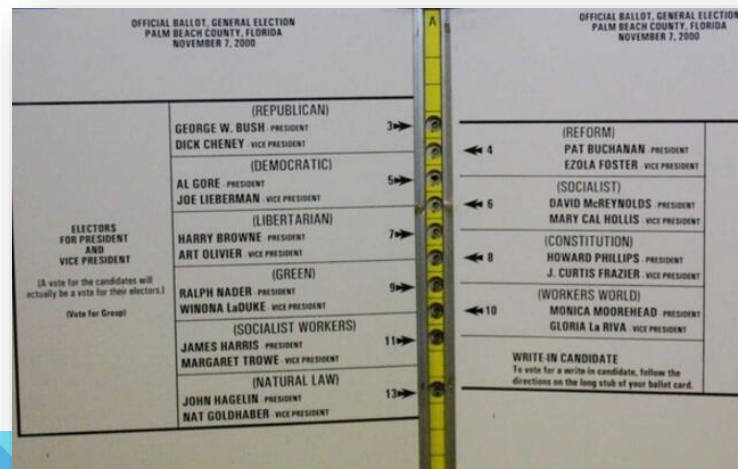
数据分析表明：**学生**通常支持**民主党**的居多，而**枪支爱好者**一般会选择支持**共和党**。

结论：德克萨斯在总统大选中历来是共和党的根据地，民主党对其**38**张选举人票绝无奢望，但新法案对**60**万亲民主党选民投票权的剥夺，却能保证共和党在同时进行的（以及未来几届的）参众两院选举中高枕无忧。



• 蝴蝶选票

- 当年佛州用的是如今臭名昭著的“蝴蝶选票”。该选票分左右两列，各党候选人按数字顺序，以“左右左右”的方式分列两侧。按规定，选民要在被选人的名字旁打孔。但两列候选人呈交错分布，这样的设计很容易让人分不清自己究竟把票投给了谁。
- “布什”的名字放在了左侧，可它的右侧却不是“戈尔”，而是美国小党派改革党候选人“布坎南”。结果就是，很多想选戈尔的选民误把选票投给了布坎南。
- 很多选民先是在“布坎南”名字旁打孔，发现错误后又在“戈尔”名字旁打一个更大的孔。这种选票就会因为“打两个孔”被视为废票。



facebook



Cambridge Analytica



OCEAN测试

- 姓名：
 - 迈克尔科辛斯基 (Michal Kosinski)
- 职业：
 - 斯坦福大学助理教授，心理学家，心理测验学（一个由数据驱动的心理学分支）领域领军人物。



开放性Openness

- 对新经验的开放程度

尽责性conscientiousness

- 完美主义程度

外向性extraversion,

- 社交活跃程度

随和性agreeableness

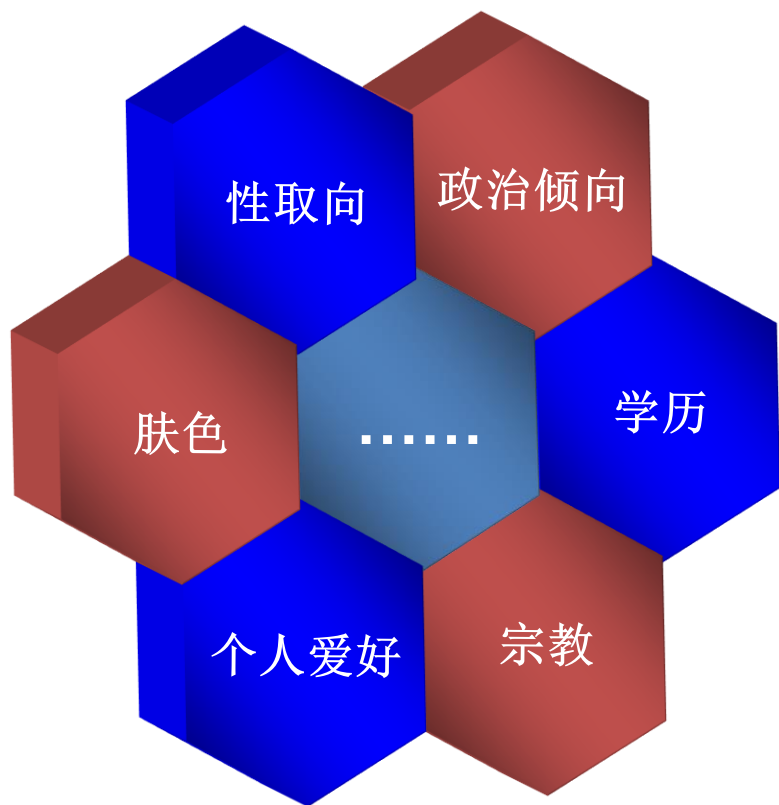
- 体贴和合作程度

情绪稳定性neuroticism

- 焦躁不安程度的简称



Data drives all we do



只给你看你想看到的，只给你看我们想给你看到的

- 我们利用Facebook收集了数百万人的个人资料。并建立模型来利用我们对他们的了解，并瞄准他们内心的恶魔。这是整个公司的基础。



克里斯托弗·怀利
数据泄密事件的举报者，
剑桥分析公司高级数据分
析师



插播广告

- 所有的PPT、数据、文章、代码.....均可以通过此公众号获取。

公众号：
虾神daxialu

