

地理数据可视化

大数据可视分析导论

陈晴

<https://idvxlab.com>

同济大学

课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- 线数据可视化
- 区域数据可视化
- 地理数据集
- 地理数据可视化案例

课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- 线数据可视化
- 区域数据可视化
- 地理数据集
- 地理数据可视化案例

地图空间数据

- 描述对象在空间中的位置和属性
- 描述真实人类生活的空间
- 由移动设备和传感器产生

地图空间数据可视化

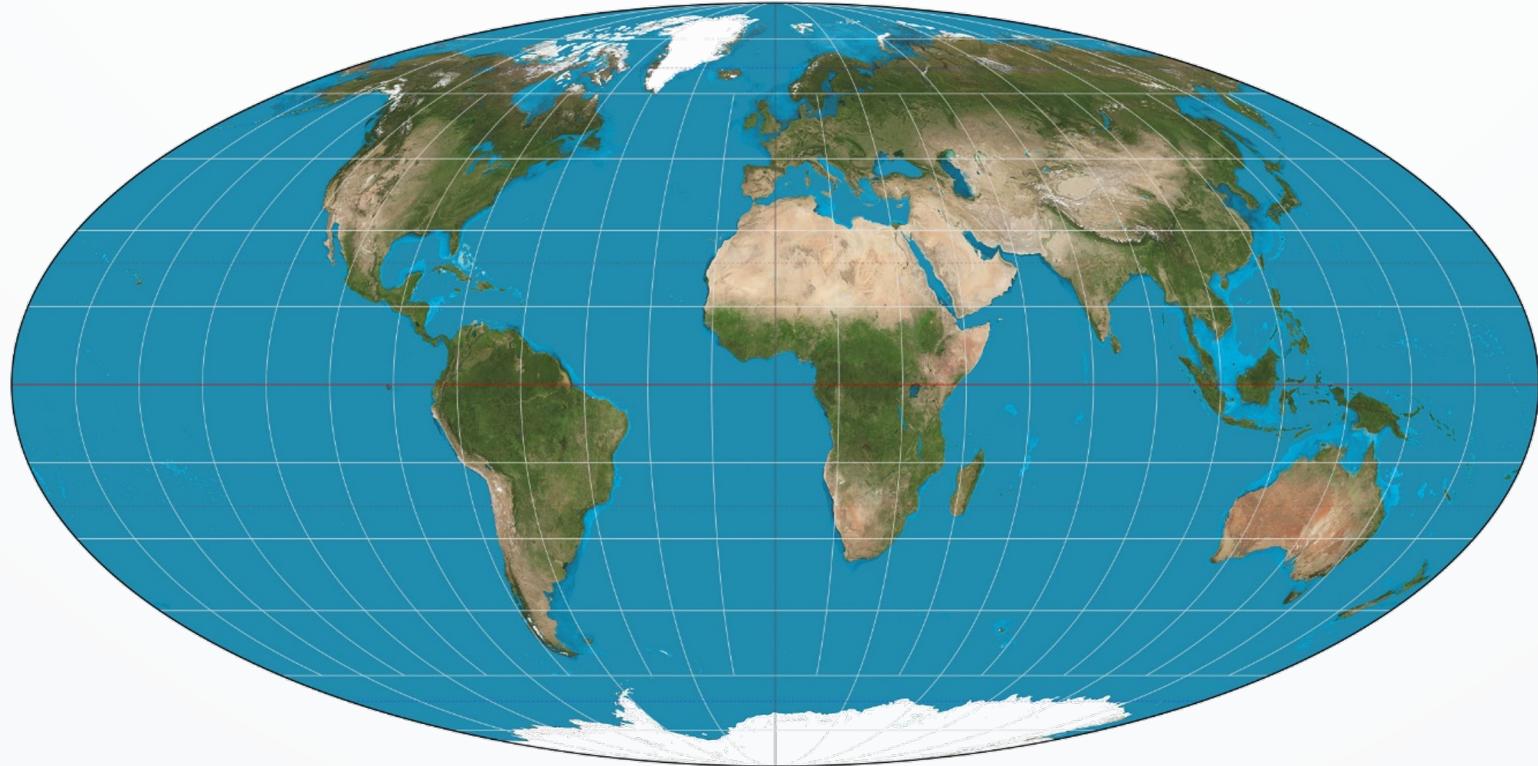
- 地图制图学 Cartography
- 地理信息系统科学 Geographic Information Systems (GIS)
- 地理空间数据可视化 Geographic Data Visualization

课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- 线数据可视化
- 区域数据可视化
- 地理数据集
- 地理数据可视化案例

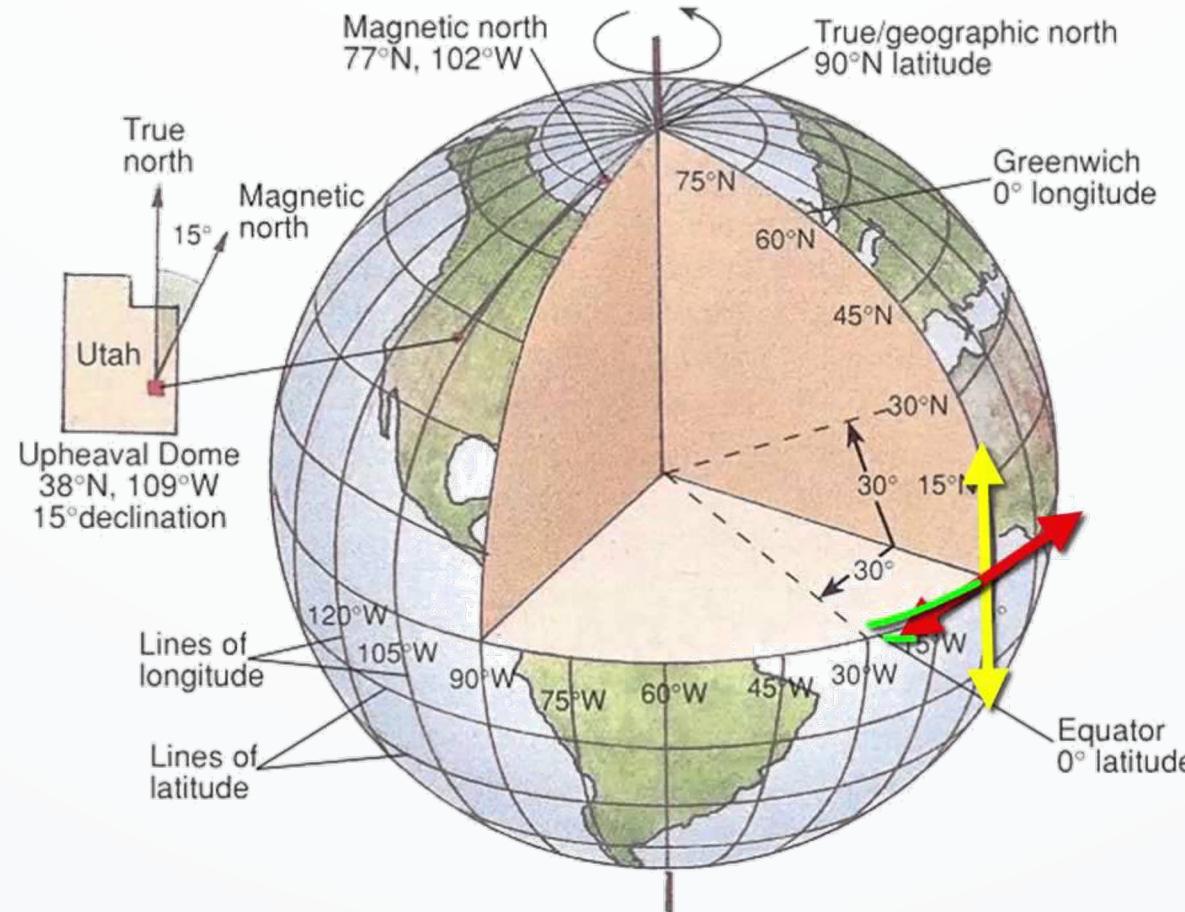
地图投影

- 将3D物体的表面投影到平面上



Wikipedia, Map projection.
https://en.wikipedia.org/wiki/Map_projection

经度与纬度Latitude-Longitude



Katryn Wiese, Latitude & Longitude.
<https://www.youtube.com/watch?v=2PlIX2YOAHo>

地图投影方式的分类

- 它们都要保持某种性质 – 但不可兼得
 - 等角度Conformal: 投影前后的任何位置的局部切向和法向方向组成的角度保持不变.
 - 等面积Equal-Area: 投影前后的相应图形面积大小保持不变.
 - 等距离Equidistant: 投影前后的各个位置在标准经纬线的长度保持不变

原则 – 在不同程度的变形中妥协 Compromise

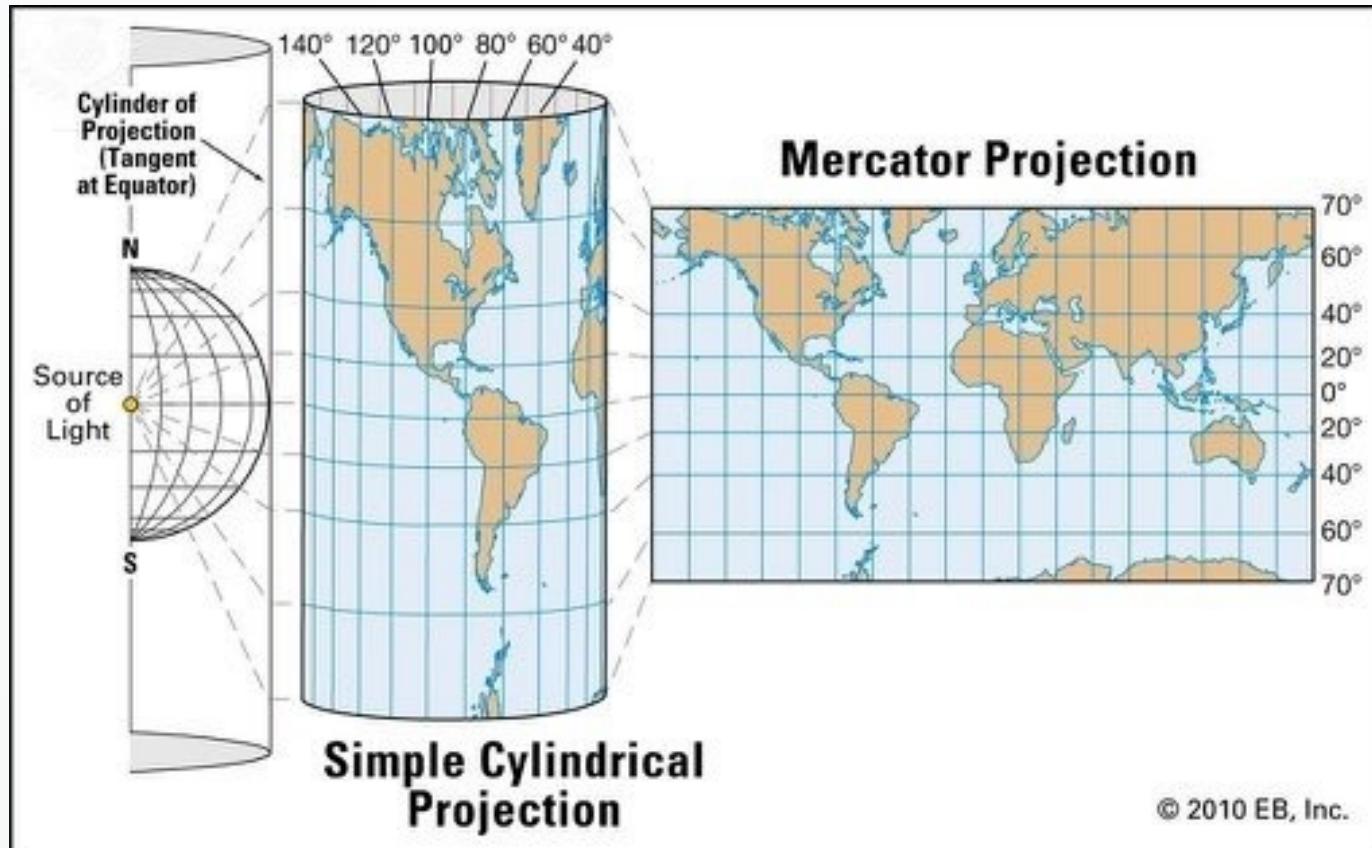
地图投影方式的分类

- 它们都要保持某种性质 – 但不可兼得
 - 等角度Conformal: 投影前后的任何位置的局部切向和法向方向组成的角度保持不变.
 - 等面积Equal-Area: 投影前后的相应图形面积大小保持不变.
 - 等距离Equidistant: 投影前后的各个位置在标准经纬线的长度保持不变

原则 – 在不同程度的变形中妥协 Compromise

等角度投影Conformal Projections

- 代表方式：墨卡托投影 Mercator Projection

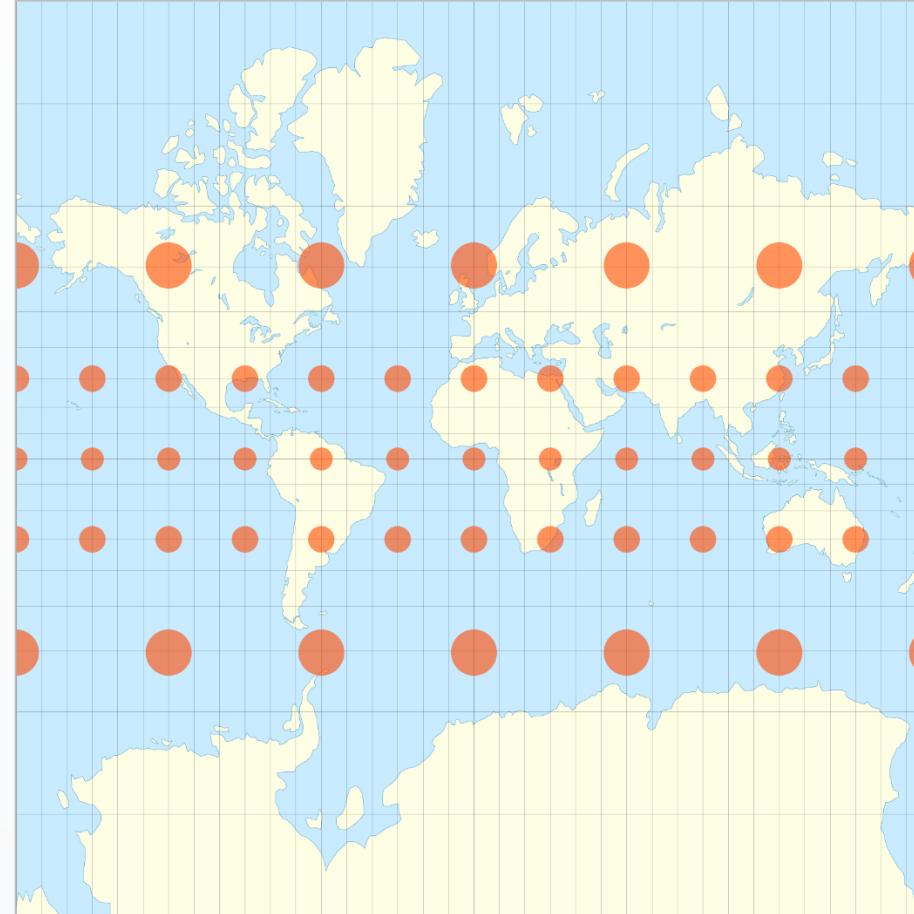


墨卡托投影中经纬度 (λ, φ) 转换成屏幕坐标 (x, y) 的数学公式：

$$\begin{aligned}x &= \lambda - \lambda_0 \\y &= \ln \left(\tan \left(\frac{\pi}{4} + \frac{\varphi}{2} \right) \right) = \frac{1}{2} \ln \left(\frac{1 + \sin \varphi}{1 - \sin \varphi} \right) \\&= \sinh^{-1}(\tan \varphi) = \ln(\tan \varphi + \sec \varphi)\end{aligned}$$

等角度投影Conformal Projections

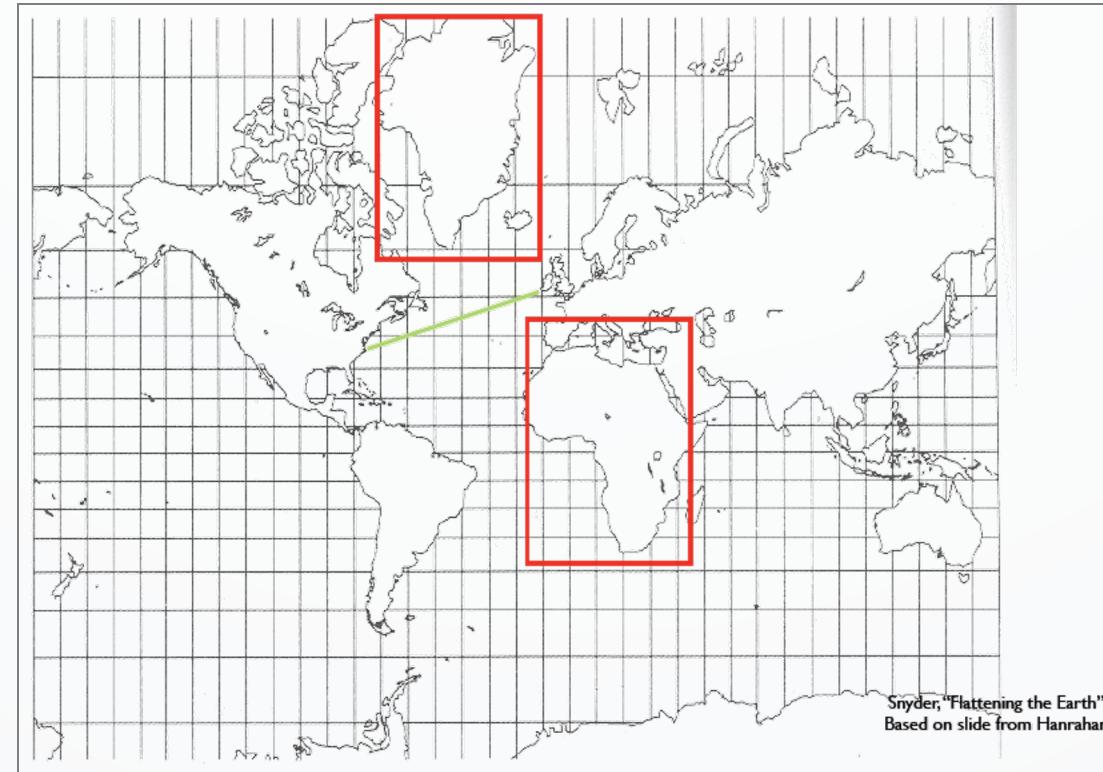
- 二维投影变形



Wikipedia, Mercator projection.
https://en.wikipedia.org/wiki/Mercator_projection

等角度投影Conformal Projections

- 区域大小变形



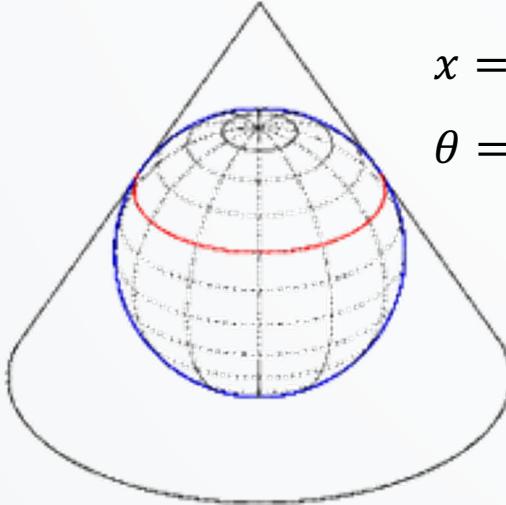
地图投影方式的分类

- 它们都要保持某种性质 – 但不可兼得
 - 等角度Conformal: 投影前后的任何位置的局部切向和法向方向组成的角度保持不变.
 - 等面积Equal-Area: 投影前后的相应图形面积大小保持不变.
 - 等距离Equidistant: 投影前后的各个位置在标准经纬线的长度保持不变

原则 – 在不同程度的变形中妥协 Compromise

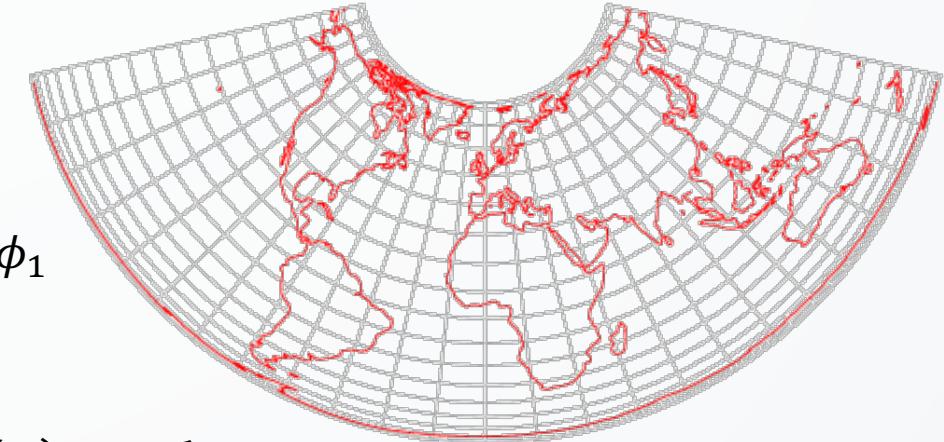
等面积投影 Equal-Area Projections

- 代表方式：亚尔勃斯投影 Albers Conic Projection



$$\begin{aligned}x &= \rho \sin \theta, y = \rho_0 - \rho \cos \theta, \text{ 其中} \\ \theta &= n(\lambda - \lambda_0), \rho = \frac{\sqrt{C-2n \sin \phi}}{n}, \rho_0 = \frac{\sqrt{C-2n \sin \phi_0}}{n}, \\ n &= \frac{1}{2}(\sin \phi_1 + \sin \phi_2), C = \cos^2 \phi_1 + 2n \sin \phi_1\end{aligned}$$

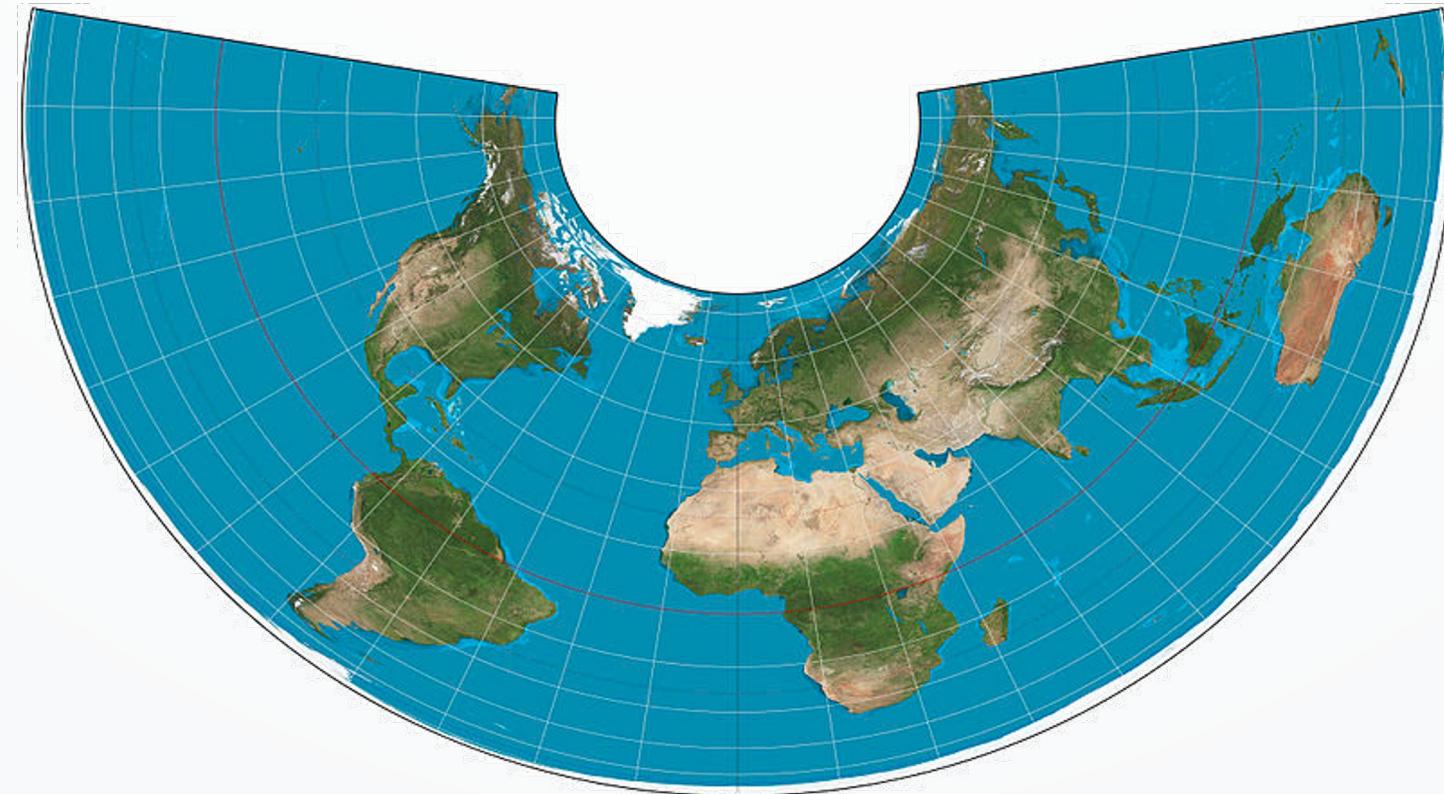
公式中 λ_0 为基准的中央经线， ϕ_0 为坐标起始纬度。 ϕ_1 和 ϕ_2 分别为第一和第二标准纬线。



MathWorks, Manage distortions with map displays.
<https://www.mathworks.com/discovery/map-projection.html>

等面积投影 Equal-Area Projections

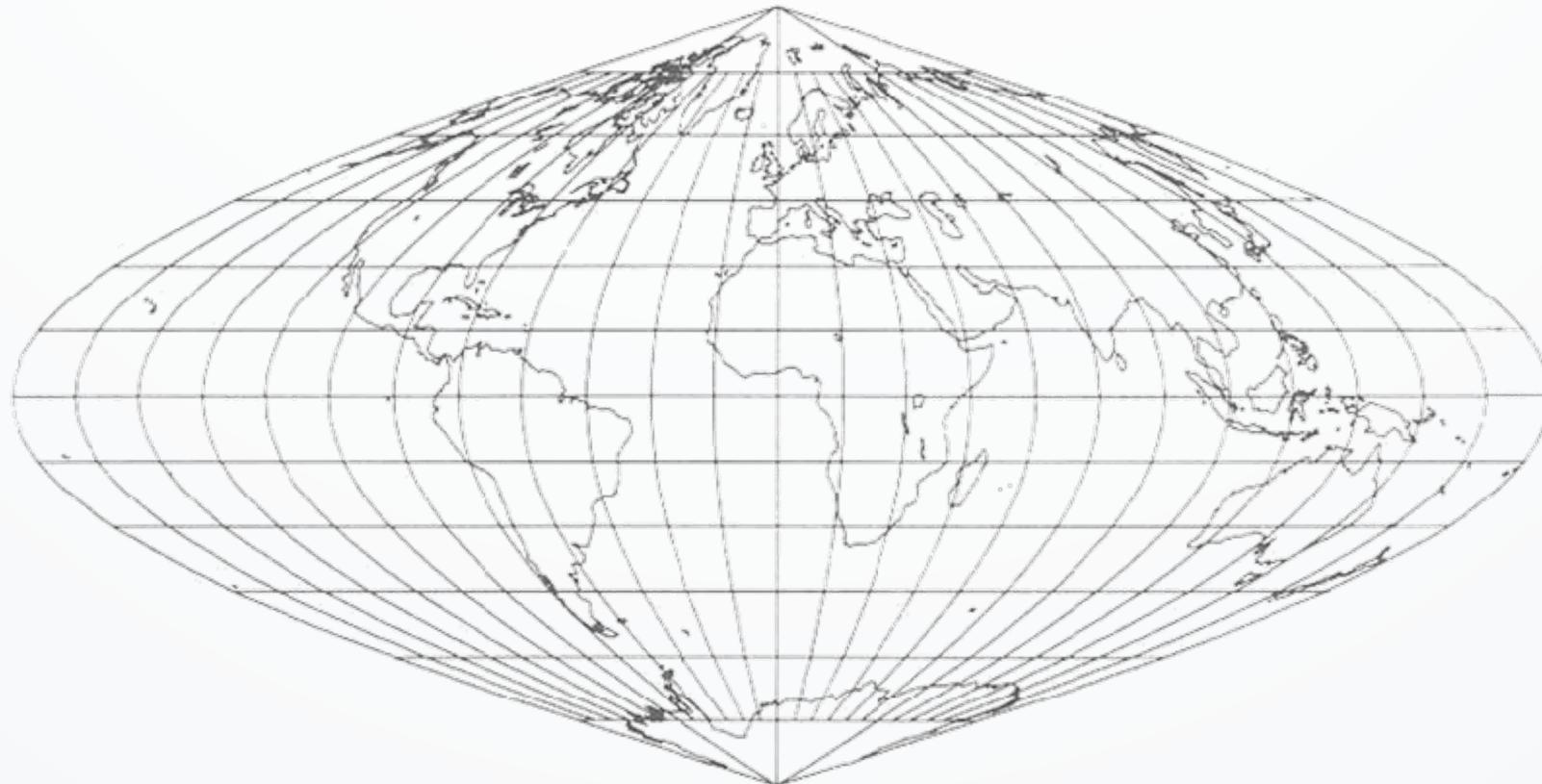
- 代表方式：亚尔勃斯投影 Albers Conic Projection



Wikipedia, Albers projection.
https://en.wikipedia.org/wiki/Albers_projection

等面积投影 Equal-Area Projections

- 正弦曲线等面积伪圆柱投影 Sinusoidal Equal-Area Projection



Snyder, Flattening the Earth.
Based on slide from Hanrahan.

地图投影方式的分类

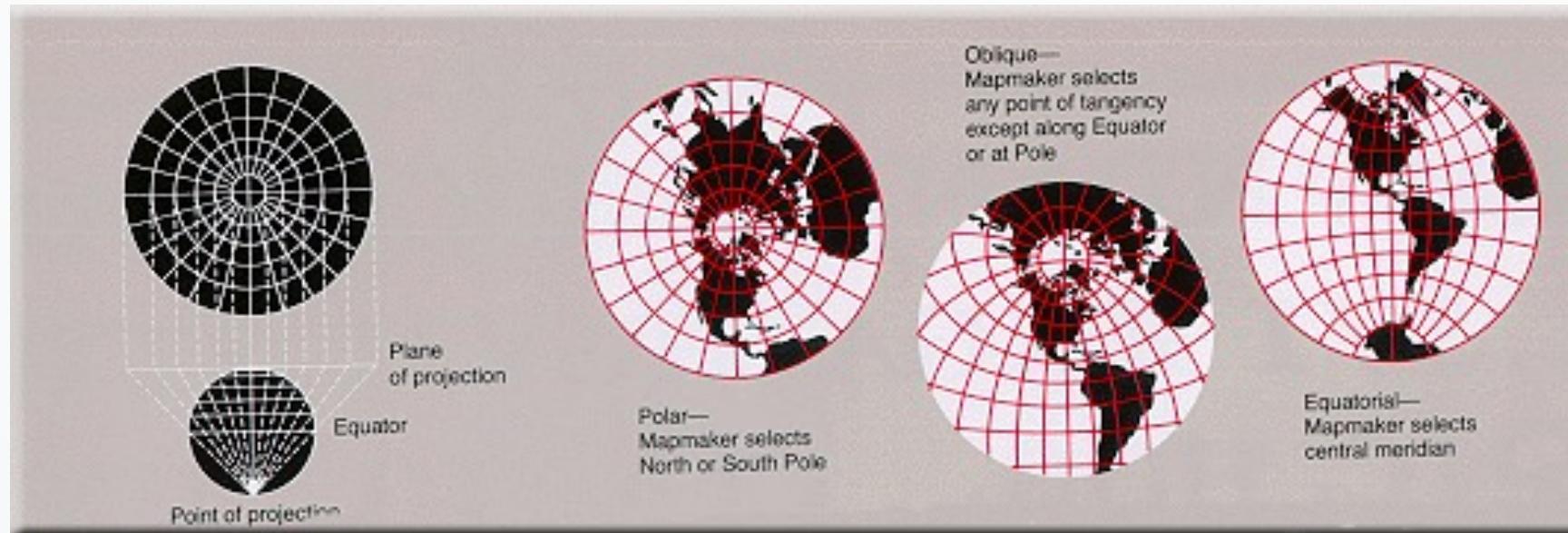
- 它们都要保持某种性质 – 但不可兼得
 - 等角度Conformal: 投影前后的任何位置的局部切向和法向方向组成的角度保持不变.
 - 等面积Equal-Area: 投影前后的相应图形面积大小保持不变.
 - 等距离Equidistant: 投影前后的各个位置在标准经纬线的长度保持不变

原则 – 在不同程度的变形中妥协 Compromise

等距离投影Equidistant Projections

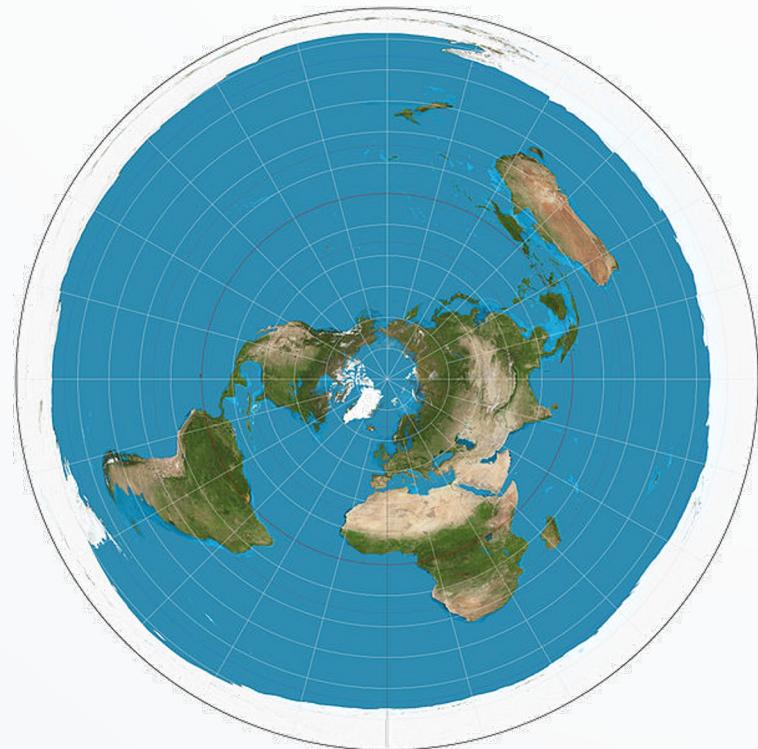
- 代表方式：方位角投影 Azimuthal Projections

- 从中心点开始的方向将被保留。
- 通过中心点的大圆在地图上用直线表示。



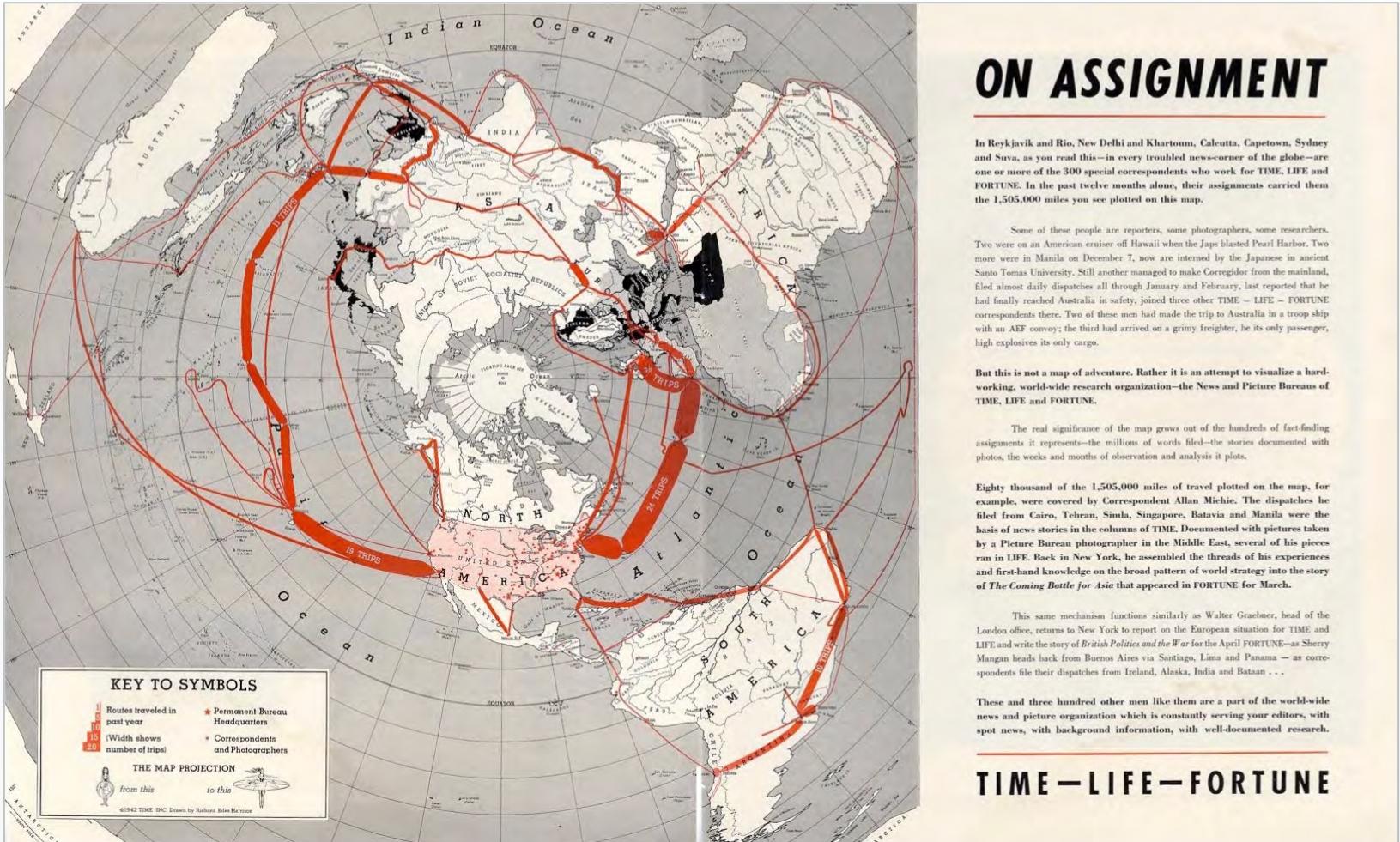
等距离投影 Equidistant Projections

- 代表方式：方位角投影 Azimuthal Projections



Wikipedia, Azimuthal equidistant projection.
https://en.wikipedia.org/wiki/Azimuthal_equidistant_projection

等距离投影Equidistant Projections



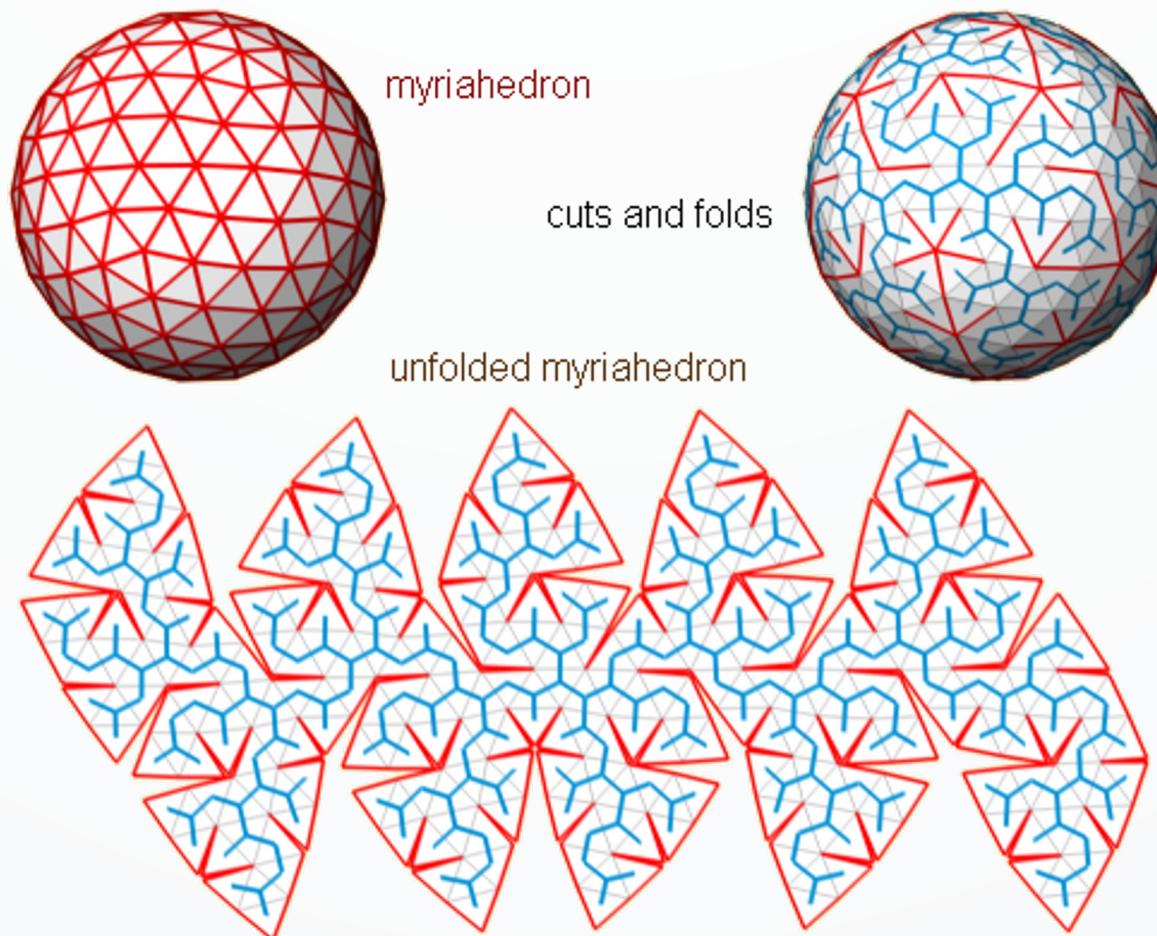
Richard Edes Harrison, Fortune magazine, 1942.
<https://www.fulltable.com/vts/f/fortune/reh/mn.htm>

地图投影方式的分类

- 它们都要保持某种性质 – 但不可兼得
 - 等角度Conformal: 投影前后的任何位置的局部切向和法向方向组成的角度保持不变.
 - 等面积Equal-Area: 投影前后的相应图形面积大小保持不变.
 - 等距离Equidistant: 投影前后的各个位置在标准经纬线的长度保持不变

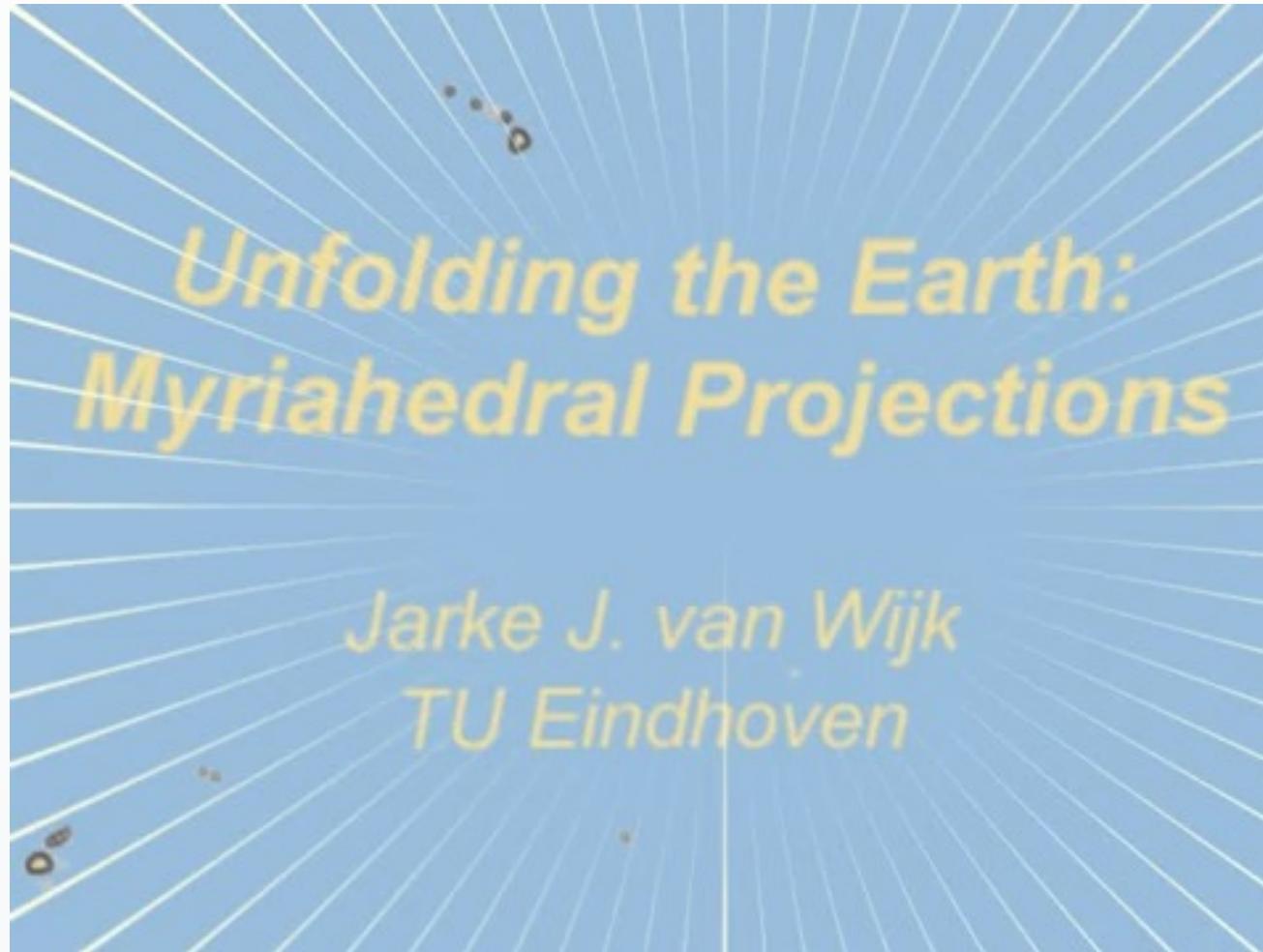
原则 – 在不同程度的变形中妥协 Compromise

多面体投影 Myriahedral Projection



Jack van Wijk, Unfolding the Earth: Myriahedral Projections.
<https://www.win.tue.nl/~vanwijk/myriahedral/>

多面体投影 Myriahedral Projection



Jack van Wijk, Unfolding the Earth: Myriahedral Projections.
<https://www.win.tue.nl/~vanwijk/myriahedral/>

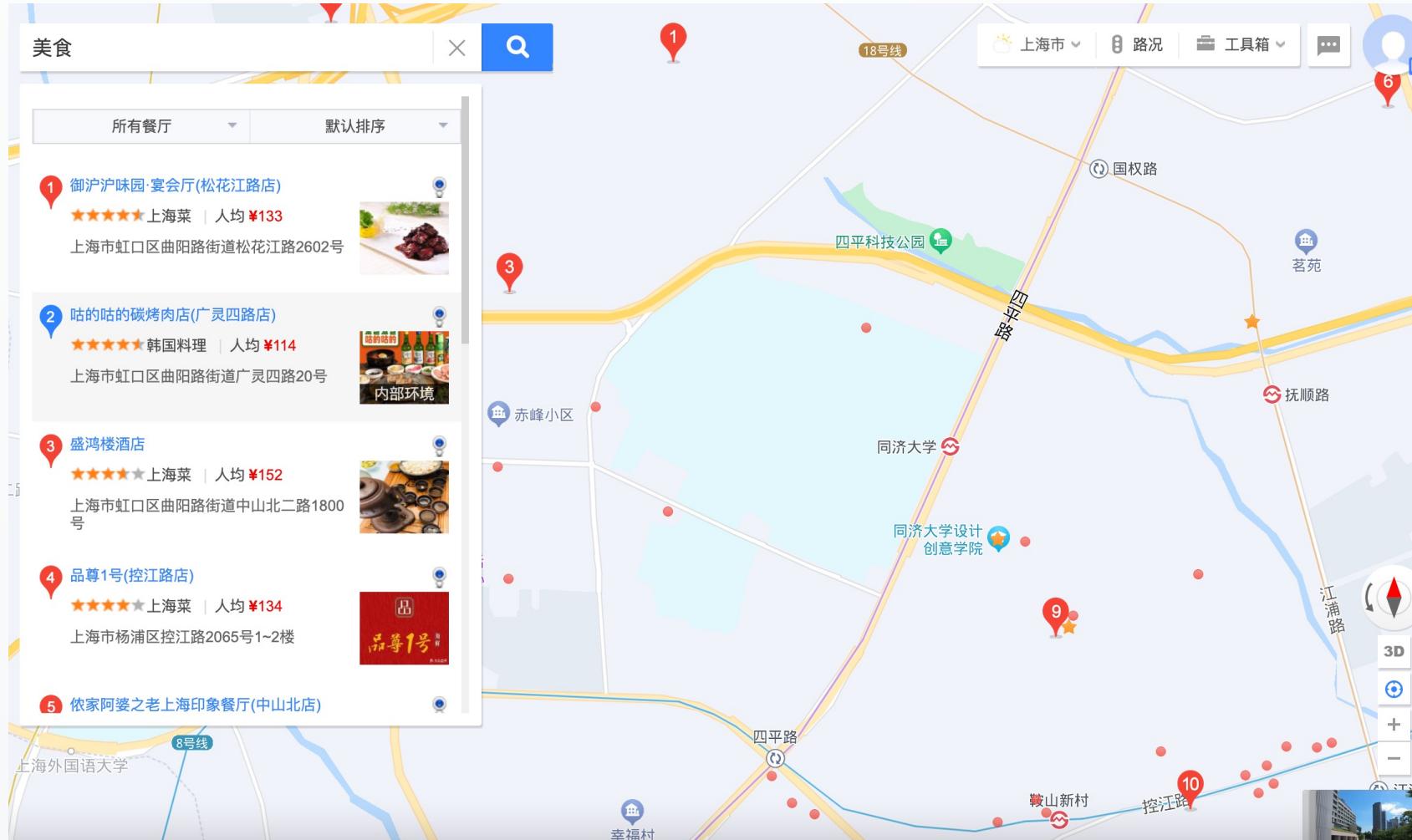
课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- 线数据可视化
- 区域数据可视化
- 地理数据集
- 地理数据可视化案例

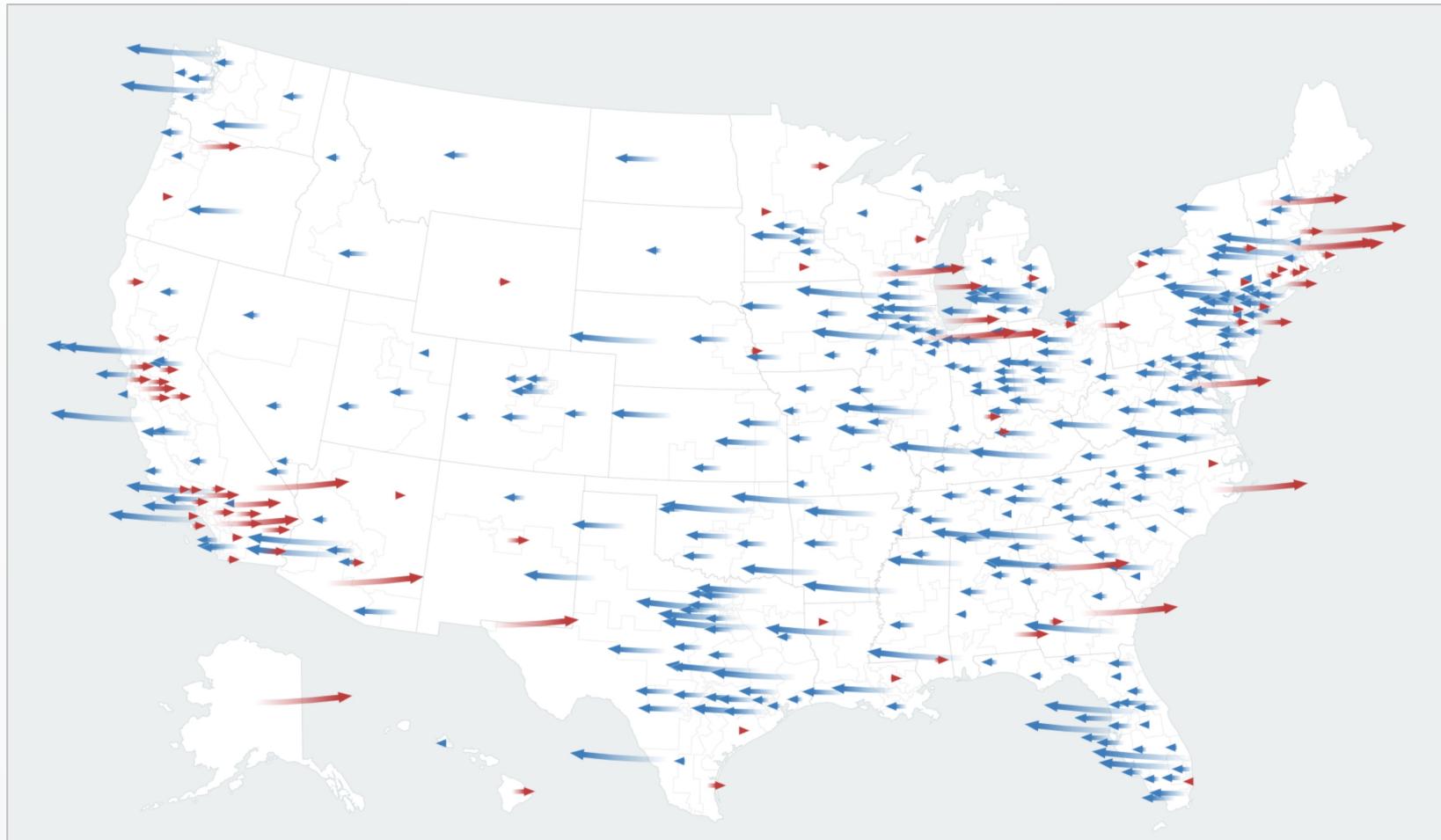
点数据

- 散布在地理空间中的点，带有经度和纬度信息，也可以有权重。
 - 地图上的地标
 - 您所在社区的餐馆
- 在地图上标记点数据将是点数据可视化的最直接、最简单的方式。
- 基于点的可视化被广泛使用。

点数据可视化



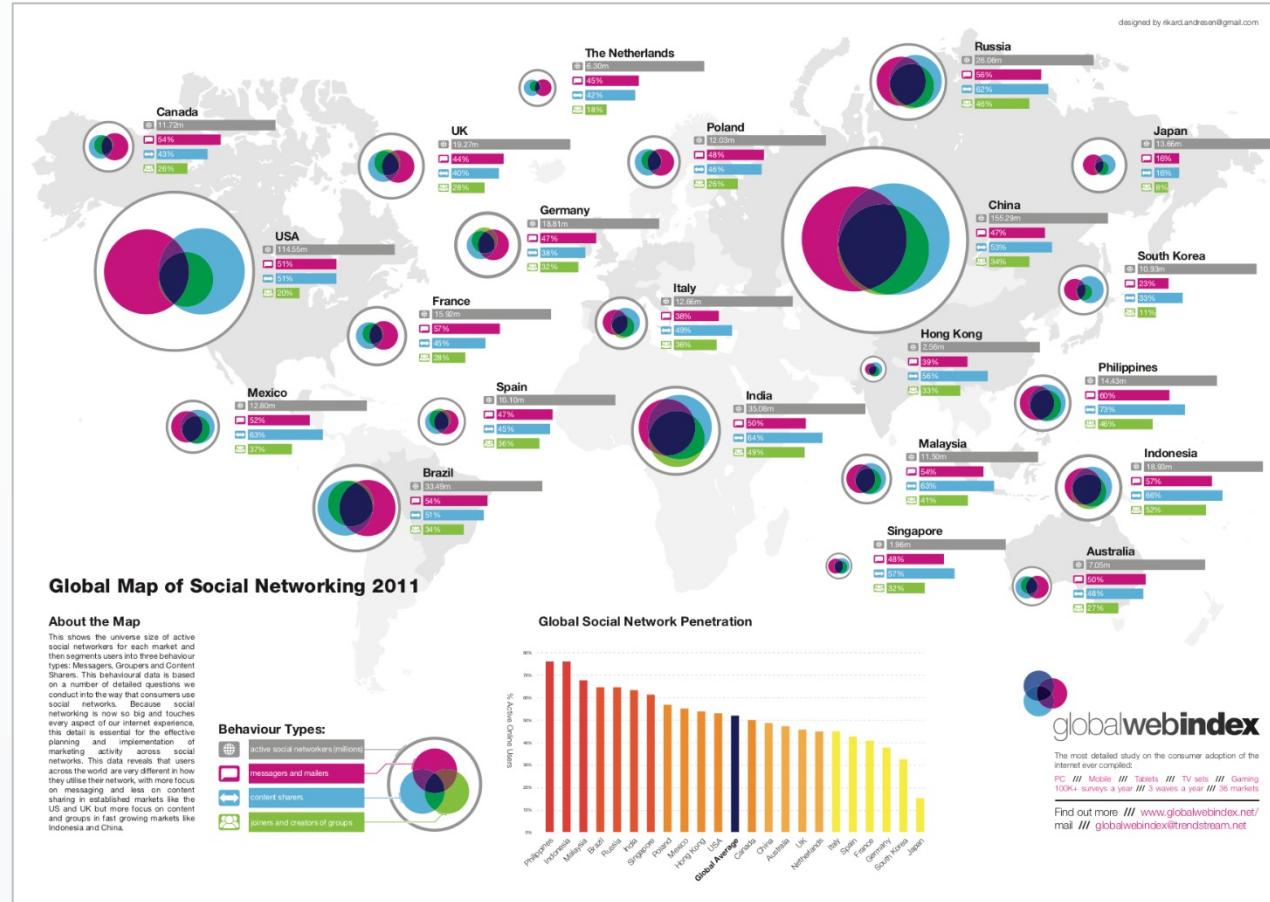
矢量型点数据



The New York Times, How Democrats Took the House.

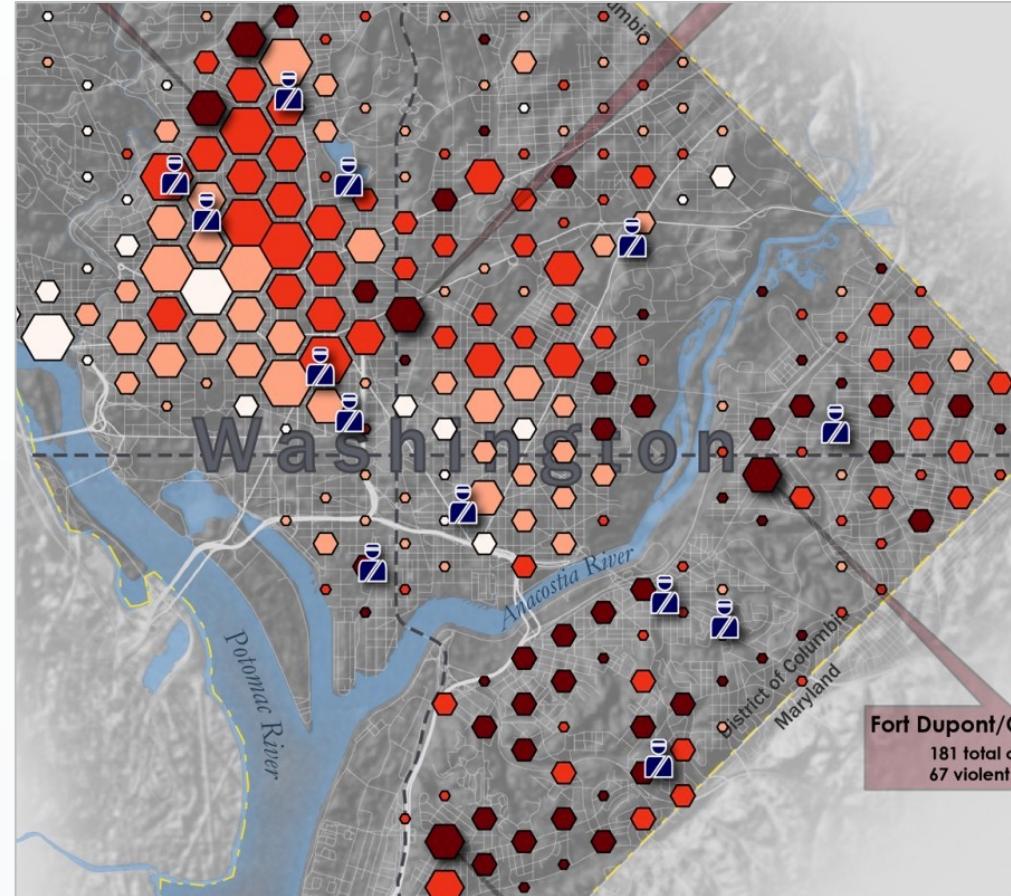
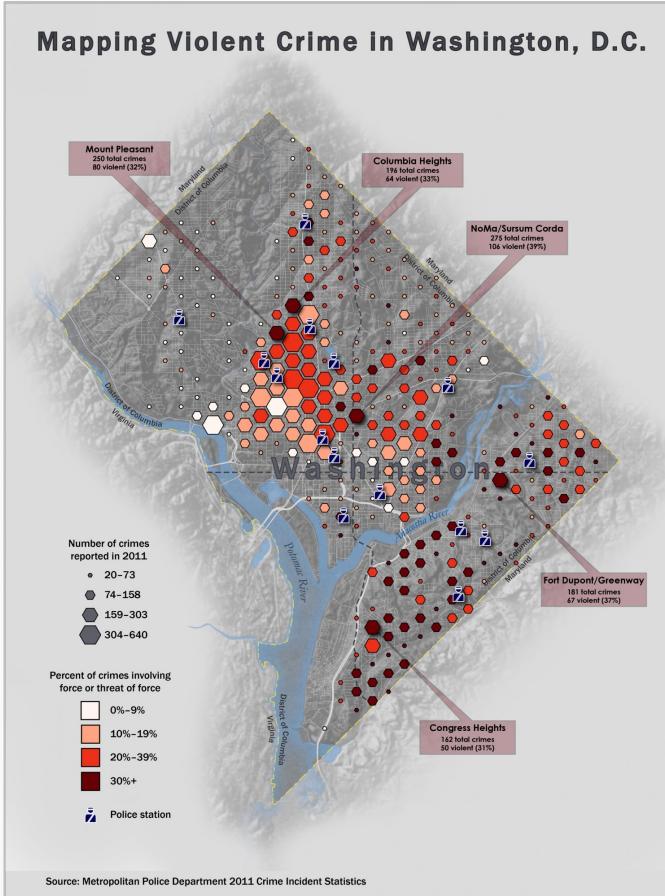
<https://www.nytimes.com/interactive/2018/11/07/us/politics/how-democrats-took-the-house.html>

2011年全球社交网络地图



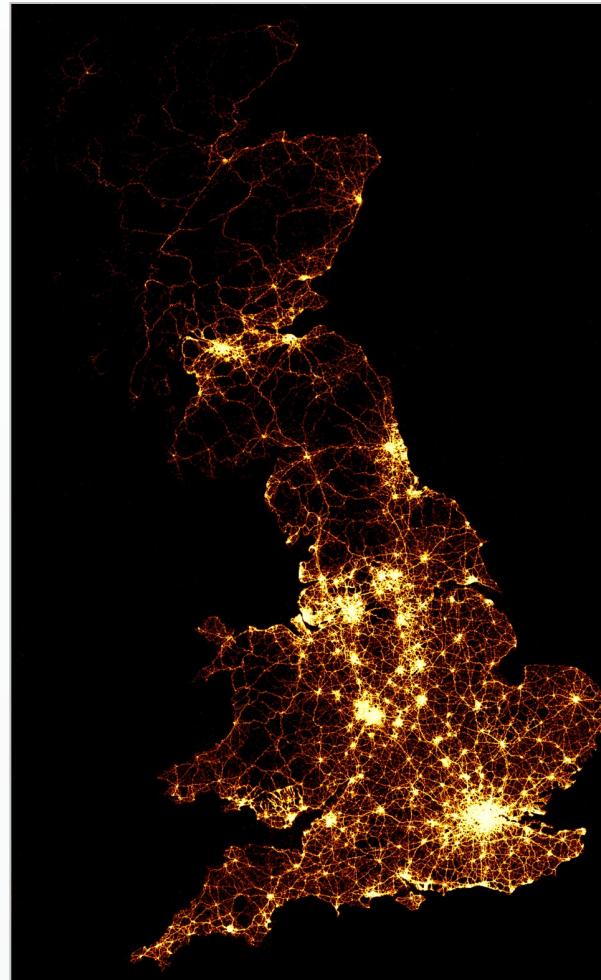
Global Web Index, Global Map of Social Networking.
<https://blog.globalwebindex.com/chart-of-the-day/new-globalwebindex-infographic/>

华盛顿特区的暴力犯罪



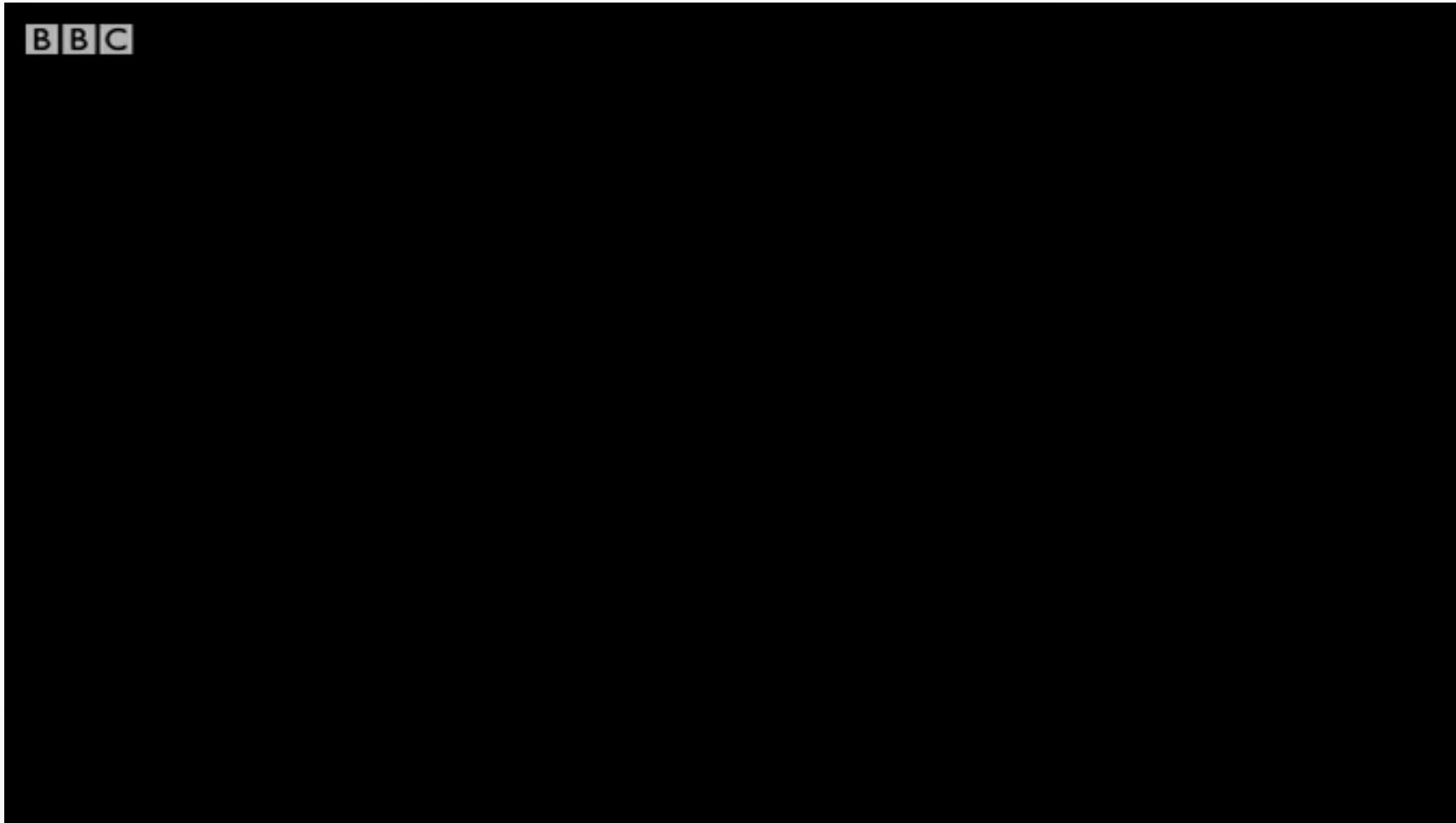
Cooper Thomas, Violent Crime in Washington, D.C.
<https://visual.ly/community/infographic/geography/violent-crime-washington-dc>

英国交通事故地理可视化



BBC, Every death on every road in Great Britain 1999-2010.
<https://www.bbc.co.uk/news/uk-15975724>

英国交通事故地理可视化



BBC, Every death on every road in Great Britain 1999-2010.
<https://www.bbc.co.uk/news/uk-15975724>

课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- **线数据可视化**
- 区域数据可视化
- 地理数据集
- 地理数据可视化案例

线数据

- 线数据指的是位置和位置之间的路径。
- 每条线都有一个长度属性，即两个位置之间的距离。
- 最常见的可视化方法也是直接绘制，也有很多可以编码的地方（颜色、宽度等）。



社交网络关系



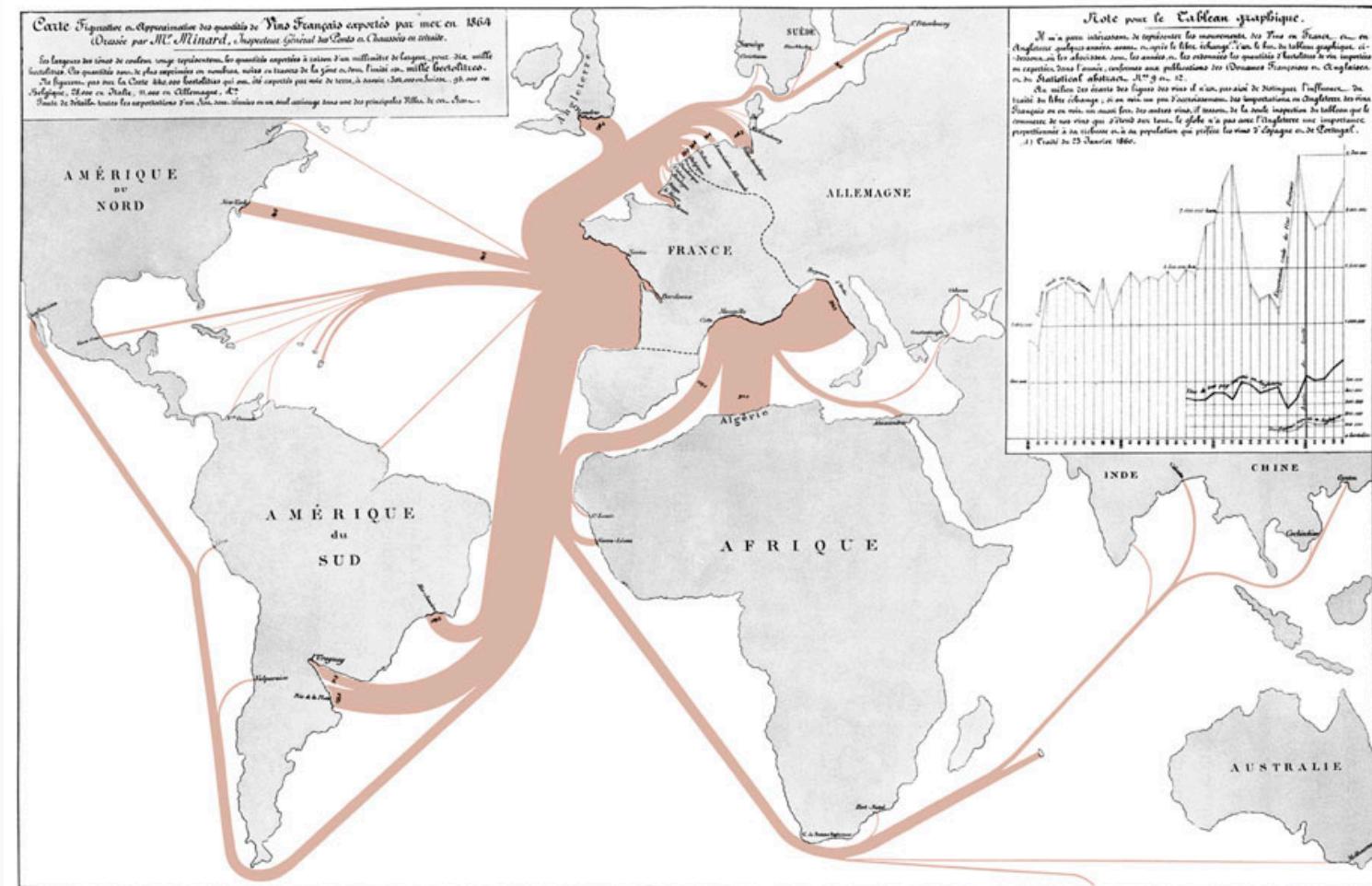
Facebook, 2013.
<https://bit.ly/2P7WZep>

空中交通网络



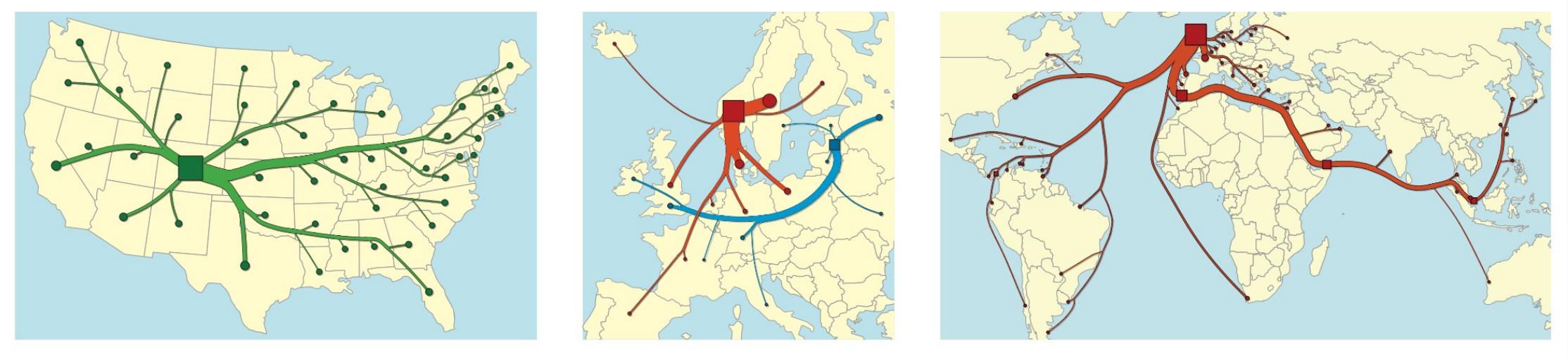
Martin Grandjean, Connected World: Untangling the Air Traffic Network.
<http://www.martingrandjean.ch/connected-world-air-traffic-network/>

线绑定 Line Binding



Charles Joseph Minard, French wine exports, 1864.
https://en.wikipedia.org/wiki/Flow_map

流量图 Flow Map

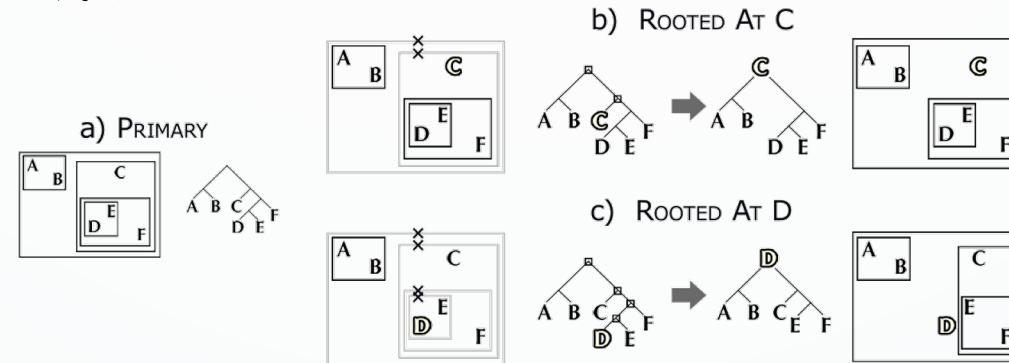


Flow maps: Migration from Colorado, migration from Norway and Latvia, whisky exports from Scotland.

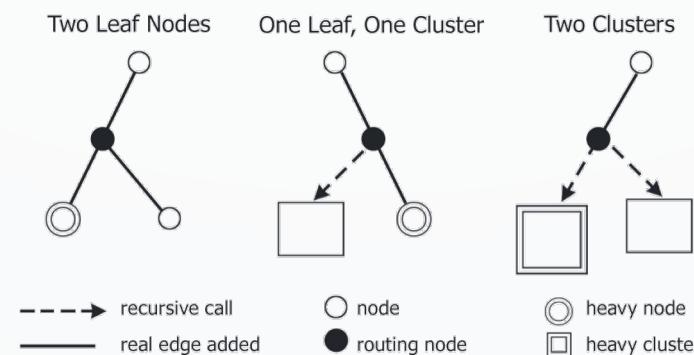
Buchin, Kevin, Bettina Speckmann, and Kevin Verbeek. "Flow map layout via spiral trees." IEEE transactions on visualization and computer graphics 17.12 (2011): 2536-2544.

流量图 Flow Map

- 通过层次聚类计算连线绑定



- 连线的布局算法



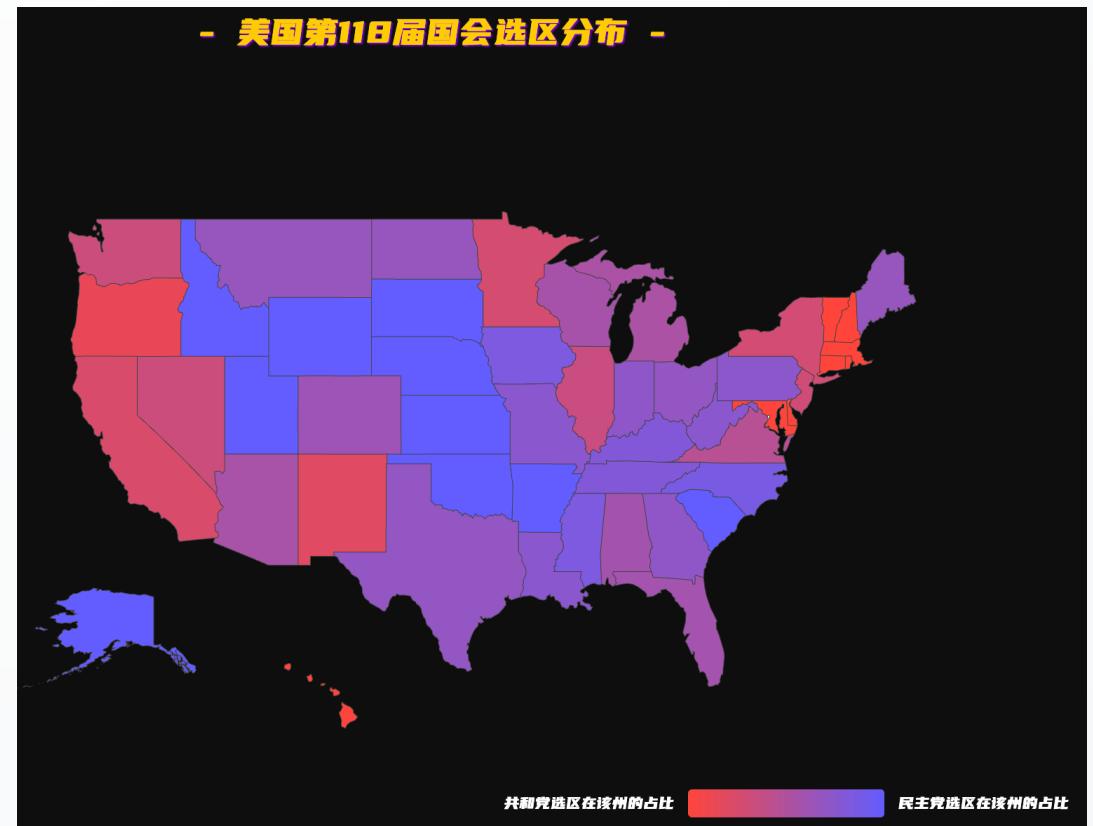
Buchin, Kevin, Bettina Speckmann, and Kevin Verbeek. "Flow map layout via spiral trees." IEEE transactions on visualization and computer graphics 17.12 (2011): 2536-2544.

课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- 线数据可视化
- **区域数据可视化**
- 地理数据集
- 地理数据可视化案例

区域数据

- 区域数据可以理解为地图上的一个区域。
- 有长度、宽度。
- 是由一系列的点围成的一个封闭的二维空间。

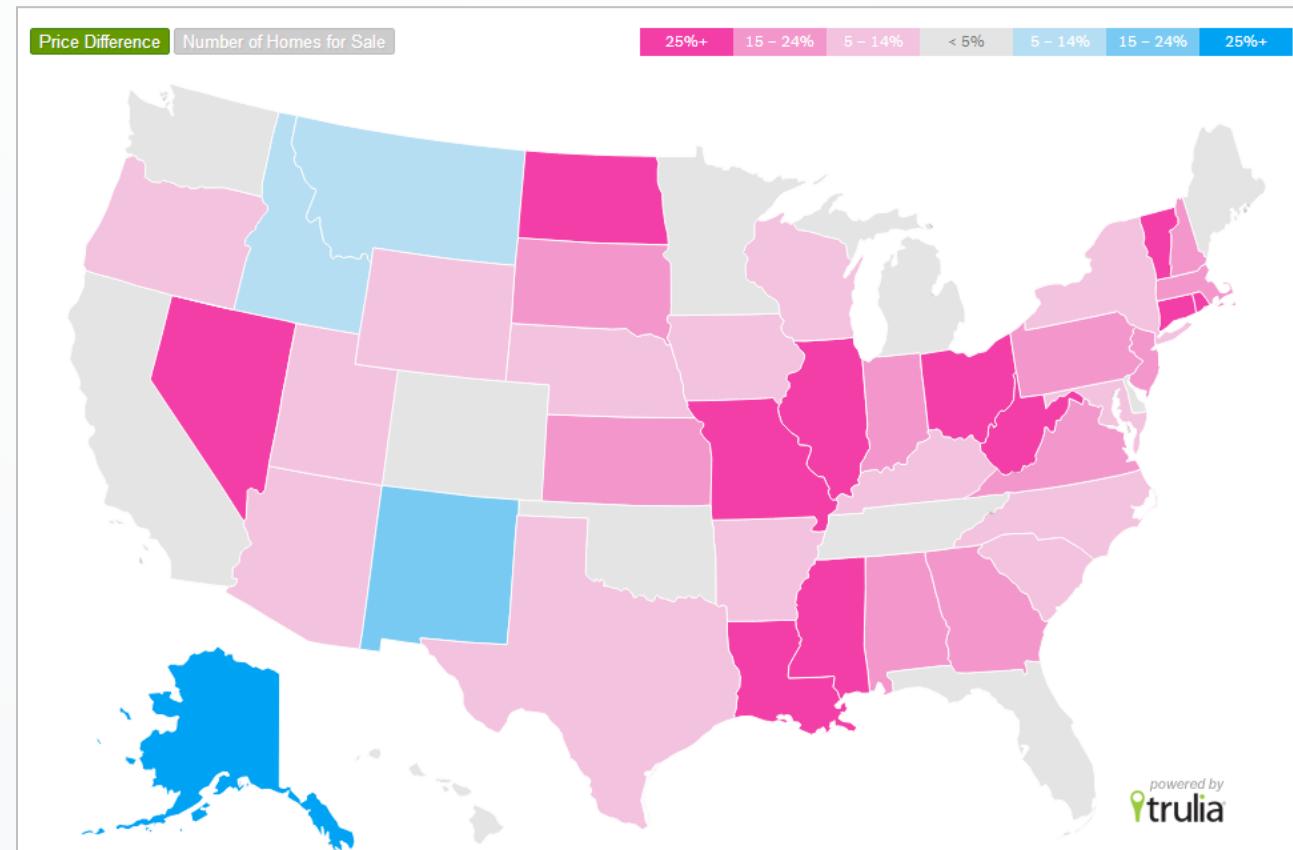


等值线图 Choropleth Map

- 通过等值线显示区域内的分布特征，两个等值线之间的区域具有相同或者相似的数据值。
- 假设数据在该区域均匀分布。
- 用颜色表示数据的比例。

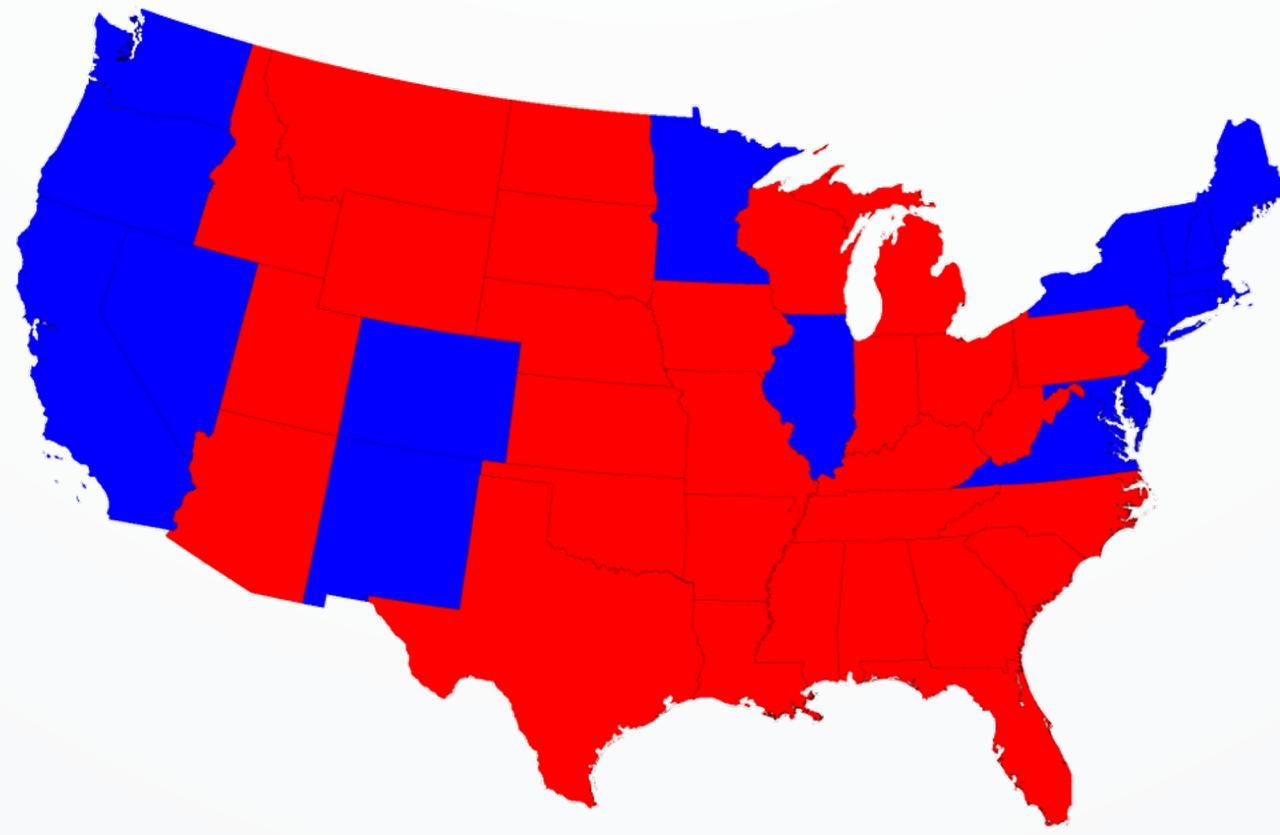


房地产销售代理分布图 Selling Real Estate: Men vs. Women



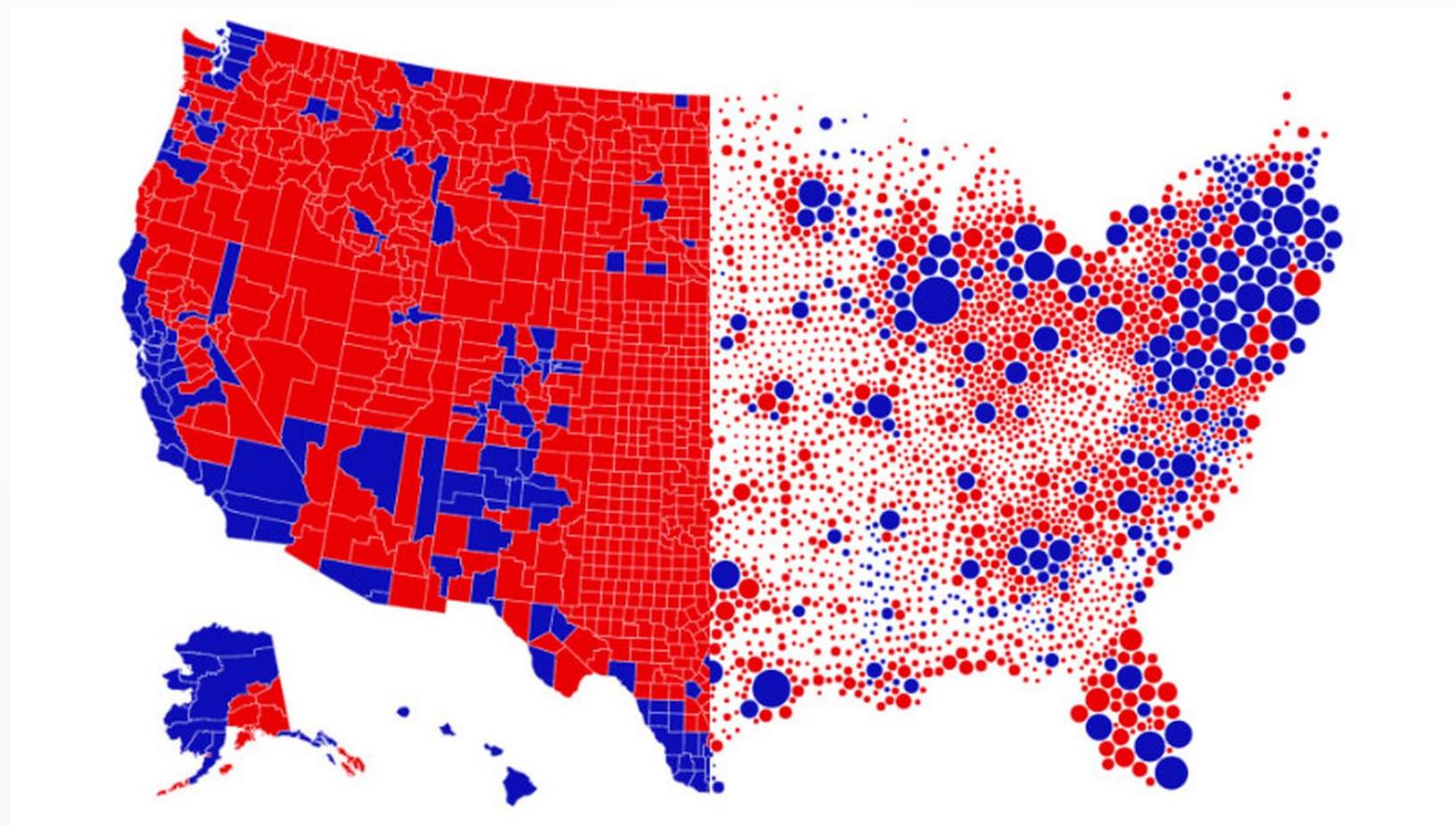
Trulia, Is real estate a man's or woman's world.
<https://bit.ly/2Klh9LE>

Choropleth Map 有什么缺点？



Mark Newman, 2016 Maps of the 2016 US presidential election results.
<http://www-personal.umich.edu/~mejn/election/2016/>

Choropleth Map 有什么缺点？

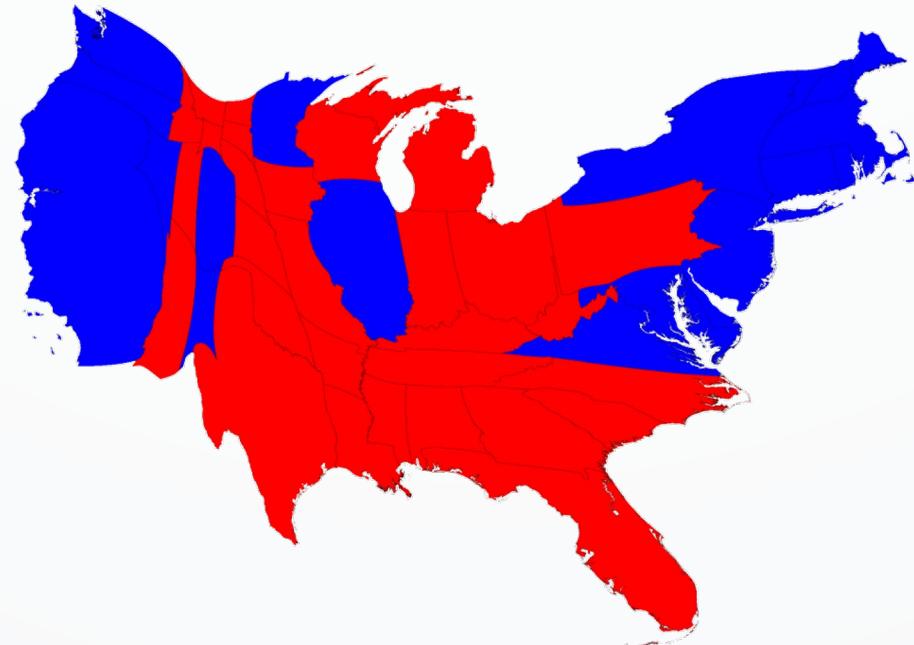


Choropleth Map 有什么缺点？

- 假设数据在该区域均匀分布。
- 这张地图没有考虑到人口分布。
- 它没有考虑到这样一个事实，即红色州的平均人口大大低于蓝色州的平均人口。蓝色的面积可能很小，但它们代表了大量的选民，这是选举中最重要的。
- 当数据积累在显示空间较小的地方时，就会出现不匹配，反之亦然。
- 这种可视化设计很可能会误导观众。

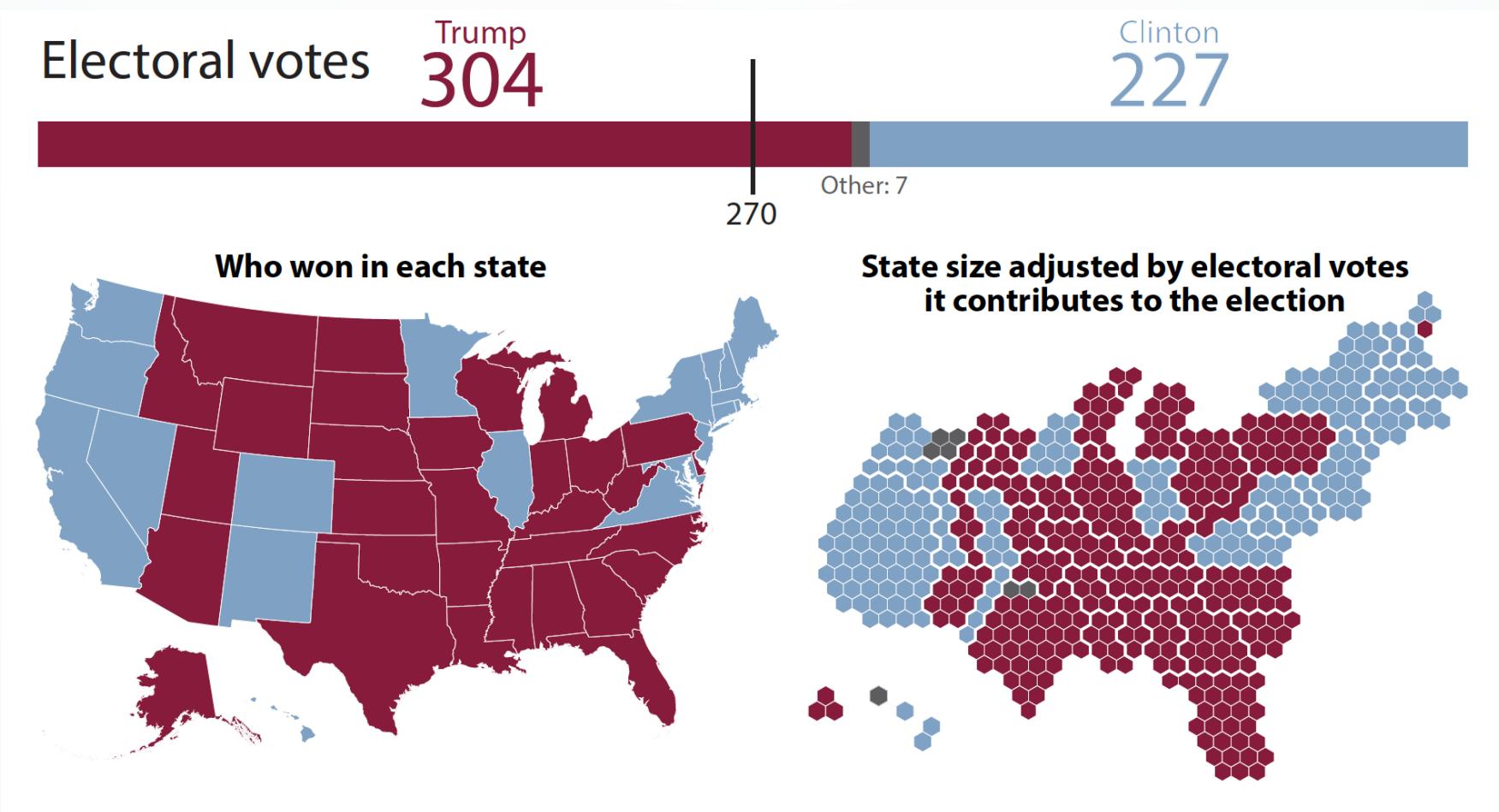
比较统计图 Cartogram Map

- 利用某些数据变量（如旅行时间、人口或国民生产总值）来代替被土地面积或距离，对区域进行适当变形。

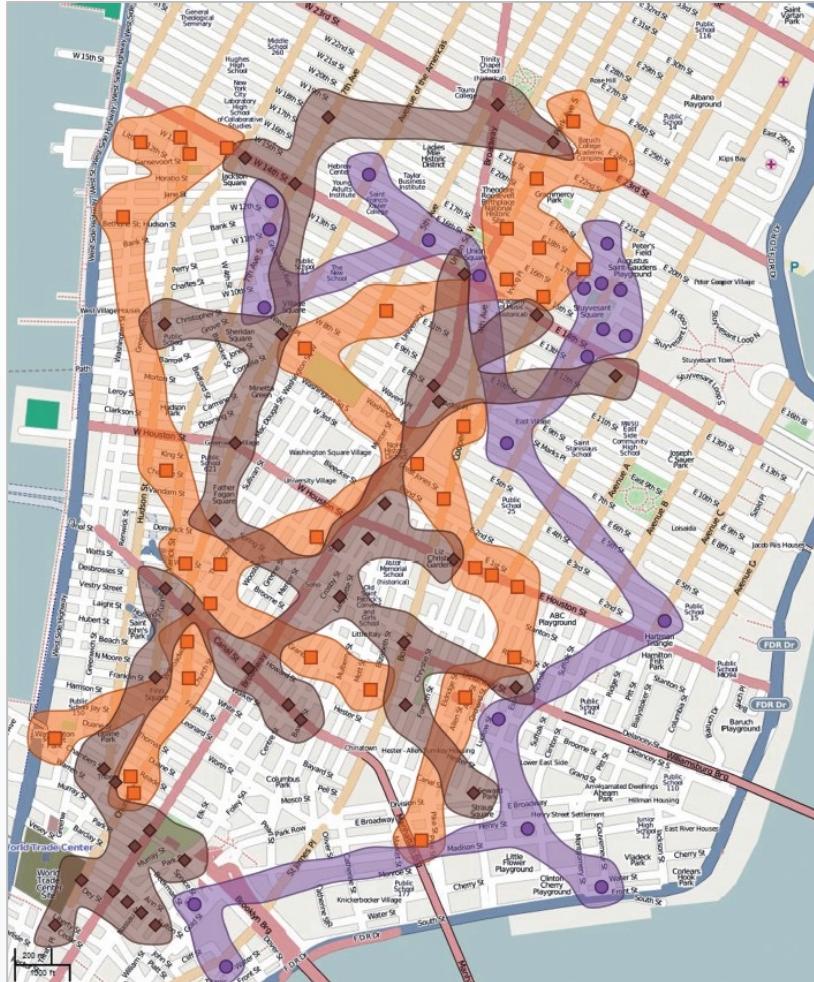


Mark Newman, 2016 Maps of the 2016 US presidential election results.
<http://www-personal.umich.edu/~mejn/election/2016/>

比较统计图 Cartogram Map

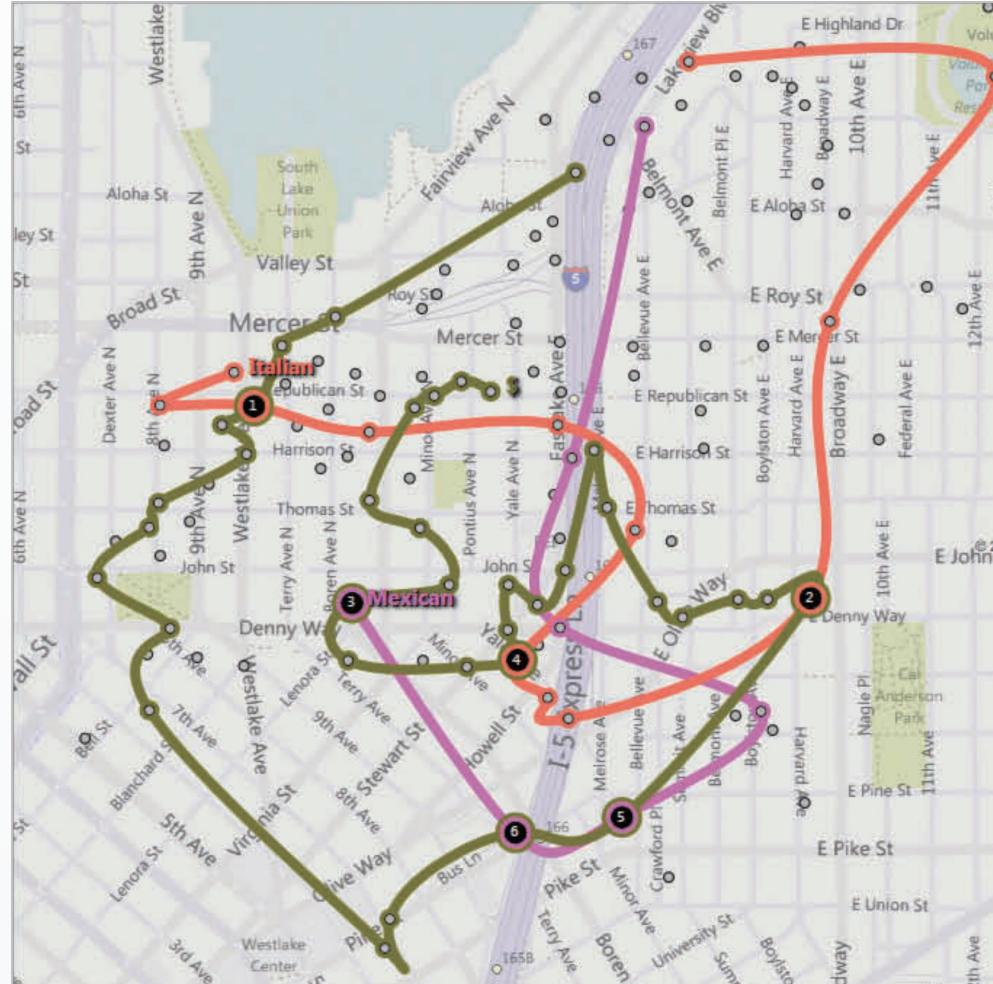


多集合地图：气泡集合Bubble Sets



Collins, Christopher; et al. Bubble sets: Revealing set relations with isocontours over existing visualizations.
IEEE Transactions on Visualization & Computer Graphics 6 (2009): 1009-1016.

多集合地图：线集合LineSets

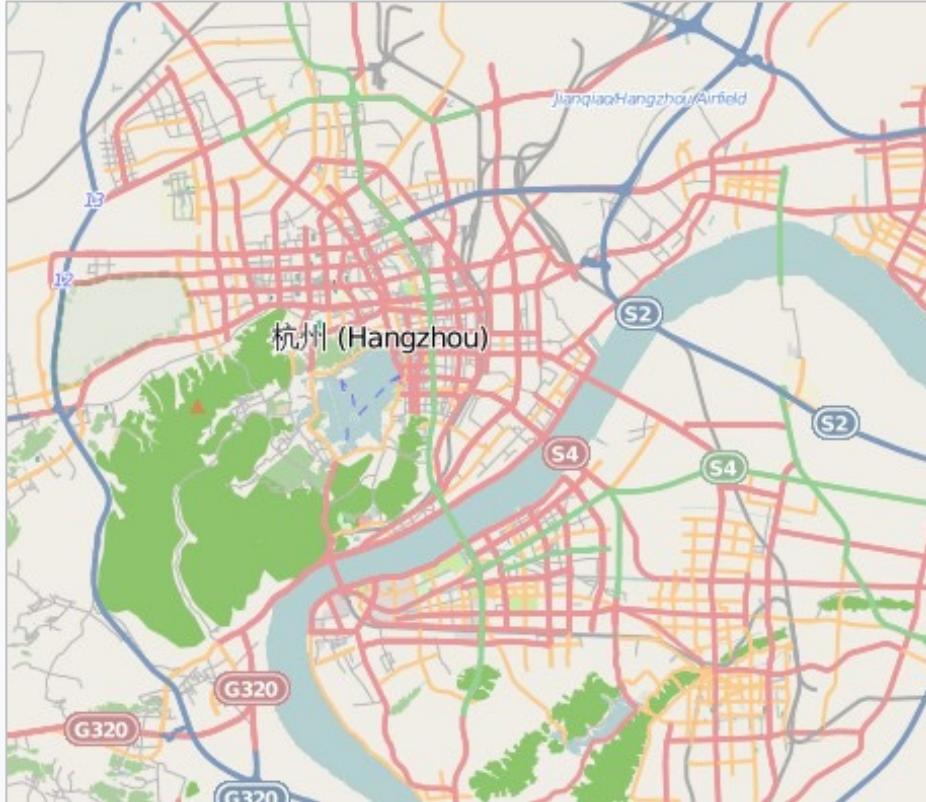


Alper, Basak, et al. Design study of linesets, a novel set visualization technique.
IEEE Transactions on Visualization & Computer Graphics 12 (2011): 2259-2267.

课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- 线数据可视化
- 区域数据可视化
- **地理数据集**
- 地理数据可视化案例

地理数据集



OpenStreetMap
<https://www.openstreetmap.org/>



HK Government Datasets
<https://data.gov.hk>

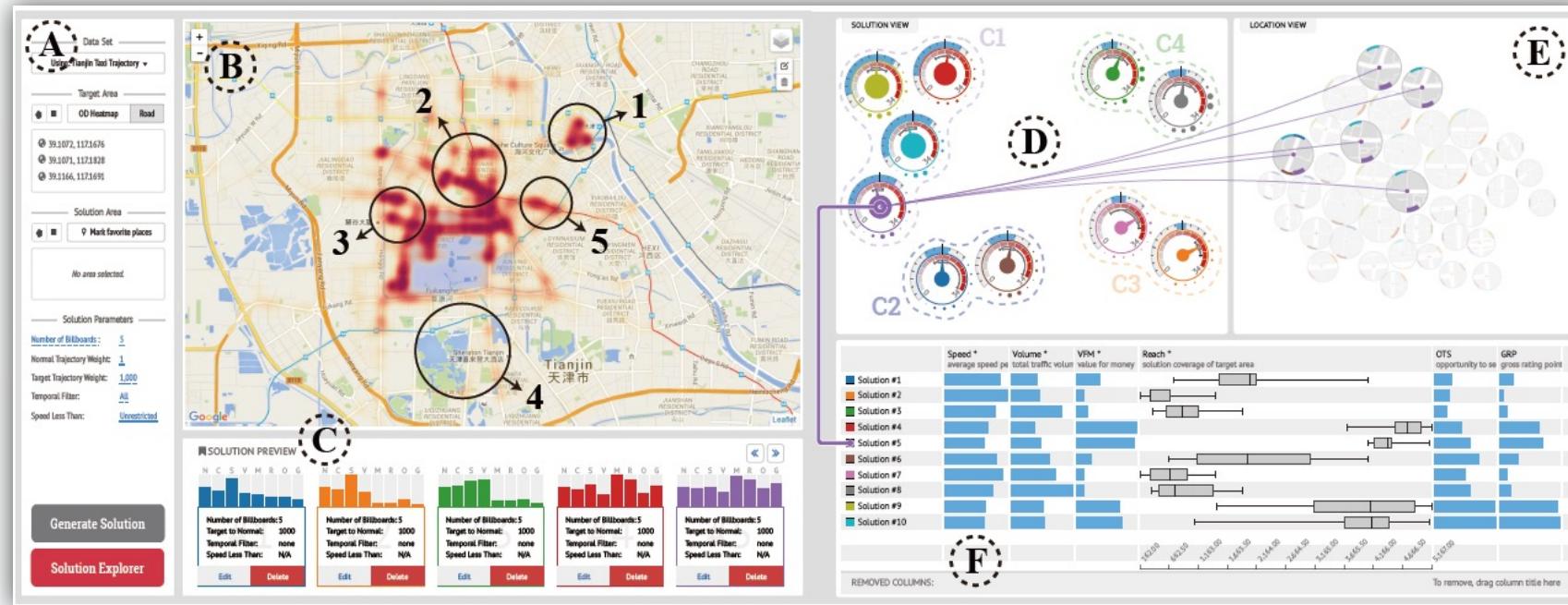
地理数据集

- Google/Baidu API
- OpenStreetMap (<http://www.openstreetmap.org>)
- HK government dataset (<http://data.gov.hk>)
- NYC Open Data (<https://opendata.cityofnewyork.us>)
- Data.gov (<https://www.data.gov>)

课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- 线数据可视化
- 区域数据可视化
- 地理数据集
- 地理数据可视化案例

地理数据可视化案例



SmartAdP: Visual Analytics of Large-scale Taxi Trajectories
for Selecting Billboard Locations

IEEE VAST 2016 (IEEE TVCG)

广告牌选址问题

- 广告牌之所以有效，是因为路上的每个人都能看到它。

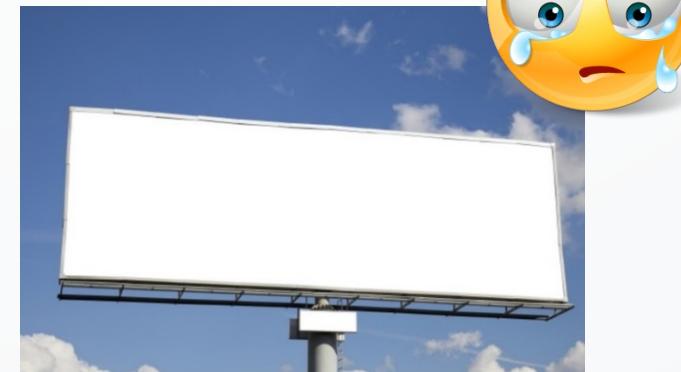


Billboards along roads



Billboards along roads

广告牌选址问题



选址因素

- 需要考虑什么样的选址因素？



广告牌选址因素

Traffic volume

Traffic speed

Traffic OD

Environment



广告牌选址因素

Traffic volume

Traffic speed

Traffic OD

Environment



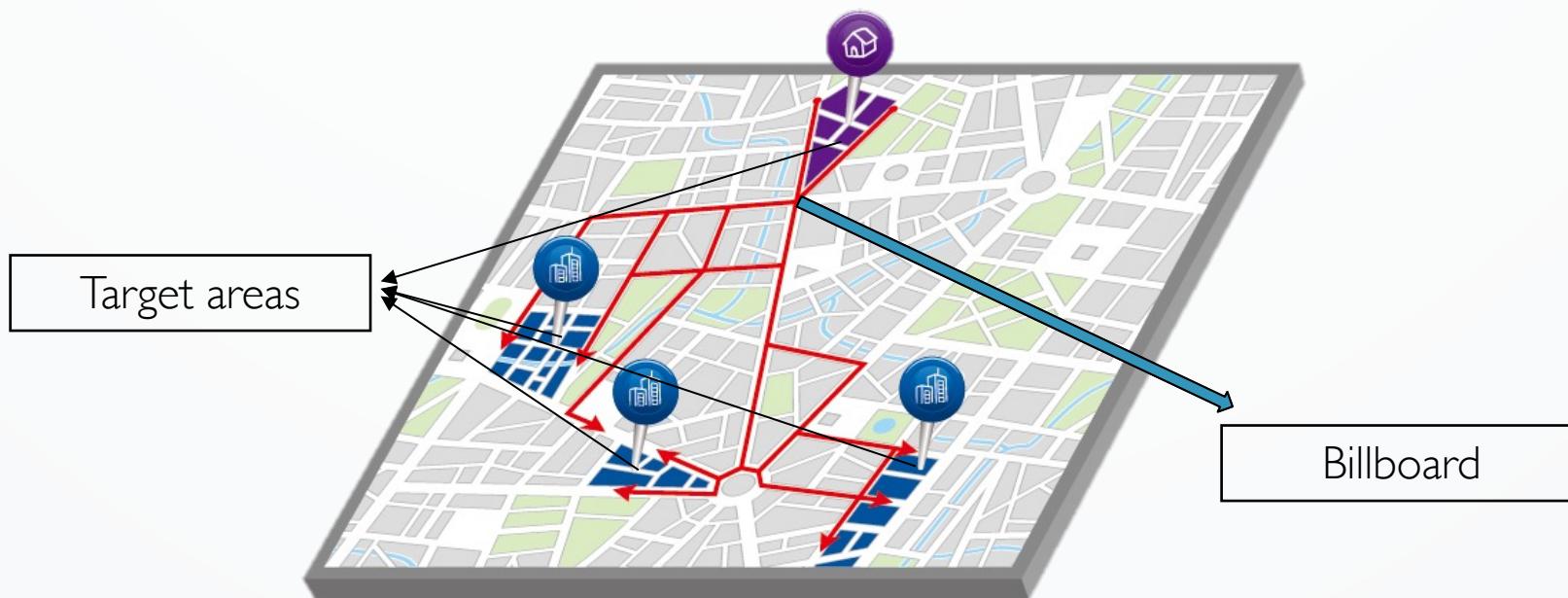
广告牌选址因素

Traffic volume

Traffic speed

Traffic OD

Environment



广告牌选址因素

Traffic volume

Traffic speed

Traffic OD

Environment



Home



Work



Shopping
& Errands



Dining &
Entertainment

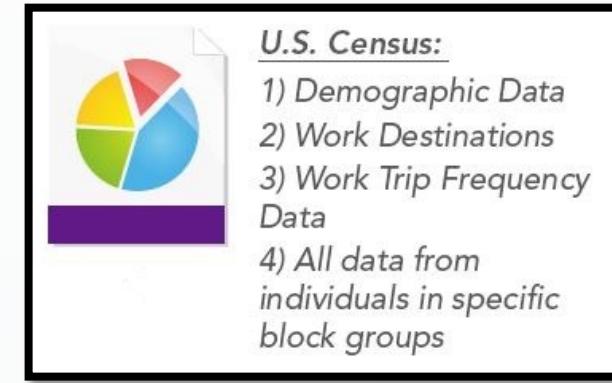
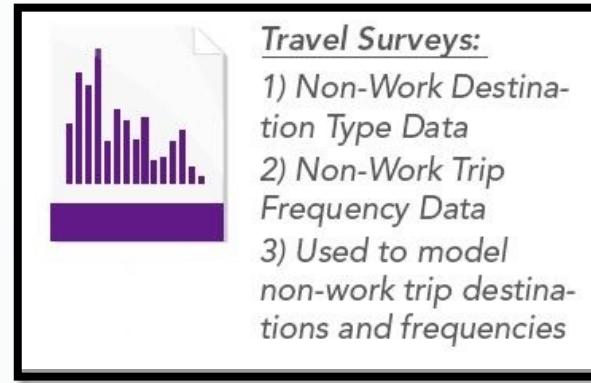


Education
Health &
Professional
Services



关键挑战 (1/3)

- 如何以低廉的价格获得如此全面的数据。
传统的方法既昂贵又耗时。

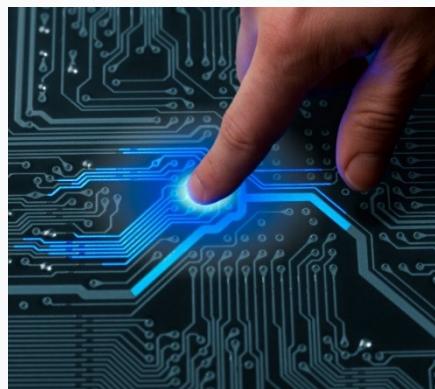


Easy to collect

Reveal underlying traffic patterns

关键挑战 (2/3)

- 如何快速制定出好的解决方案?
 - 巨大的方案空间
 - 部分因素很难建模
 - 关于最优选择的不同看法



Computational Power

+



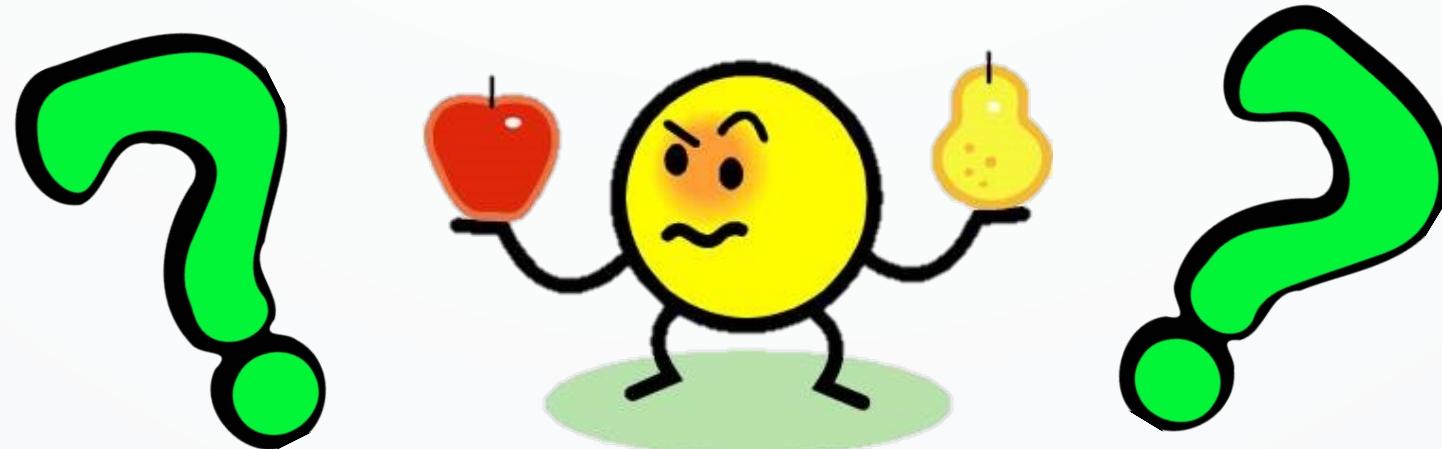
Human Knowledge

=

Cost-effective
Advertising Plan

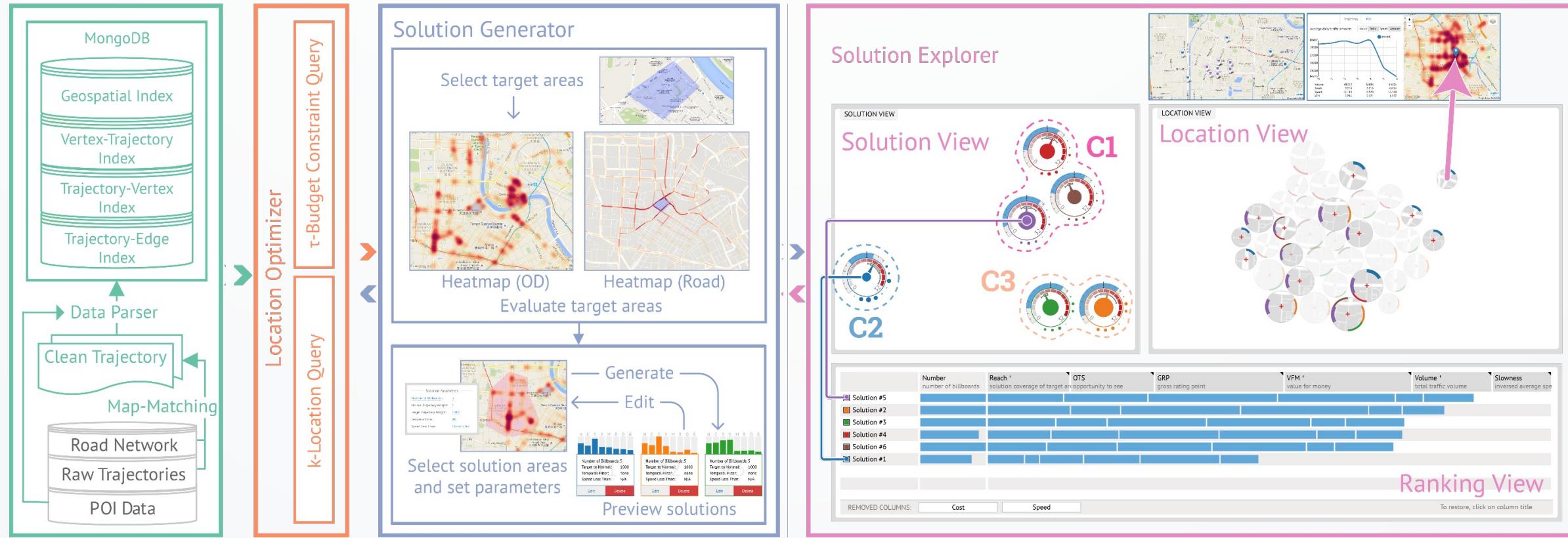
关键挑战 (3/3)

- 如何直观地比较几个候选方案以找到最优方案。
 - 哪种广告策略更好?
 - 如果我多花50,000美元可以怎么选址?
 - ...



解决方案

- 结合先进的数据挖掘方法和交互式可视化技术开发了可视化分析系统。



分析任务

Location
Optimizer

Solution
Generator

Solution
Explorer

Which areas befitting for placing billboards?
Where are the optimal locations?
How good is a billboard location/solution?

What are the differences and similarities?
How many groups exists?
What is the rank? Which ones are optimal?

Solution Generator Demo

Solution Generator

分析任务

Location
Optimizer

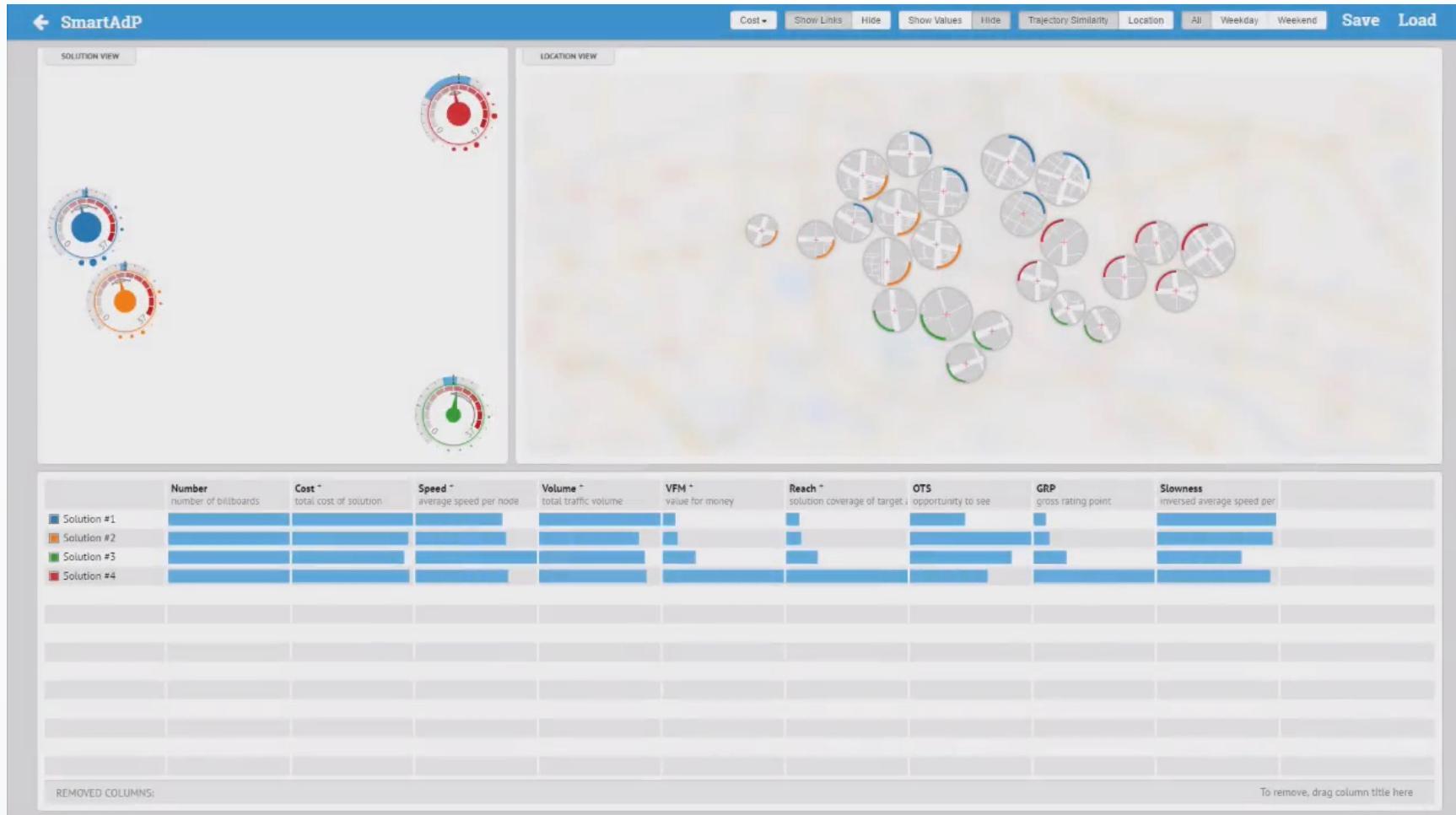
Solution
Generator

Solution
Explorer

Which areas befitting for placing billboards?
Where are the optimal locations?
How good is a billboard location/solution?

What are the differences and similarities?
How many groups exists?
What is the rank? Which ones are optimal?

Solution Explorer Demo



实例研究

- Case study of the Tianjin dataset
 - Road network data, 133,726 road segments and 99,007 vertices
 - GPS trajectory data, 4 million trajectories
 - POI data, 154,633 points

实例研究视频

Case Studies

课程大纲

- 什么是地理数据？
- 地图投影
- 点数据可视化
- 线数据可视化
- 区域数据可视化
- 地理数据集
- 地理数据可视化案例