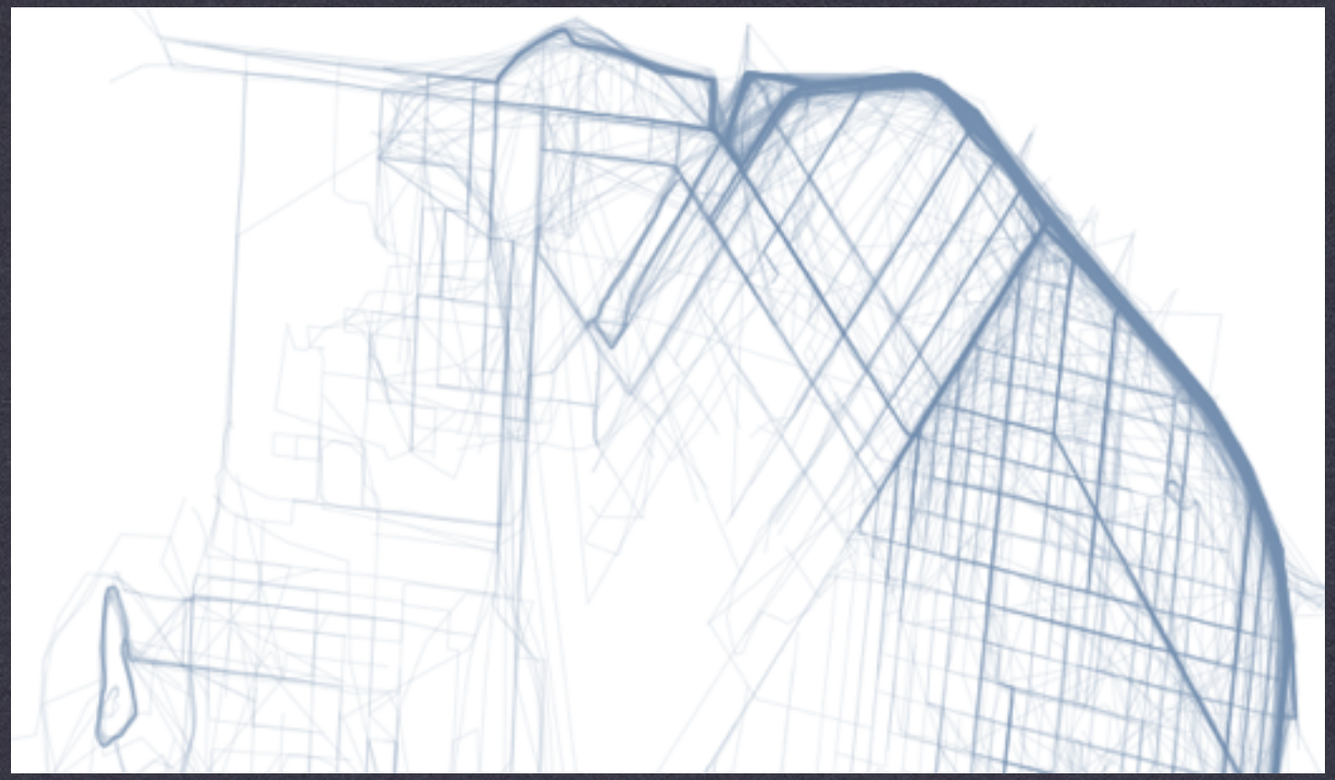
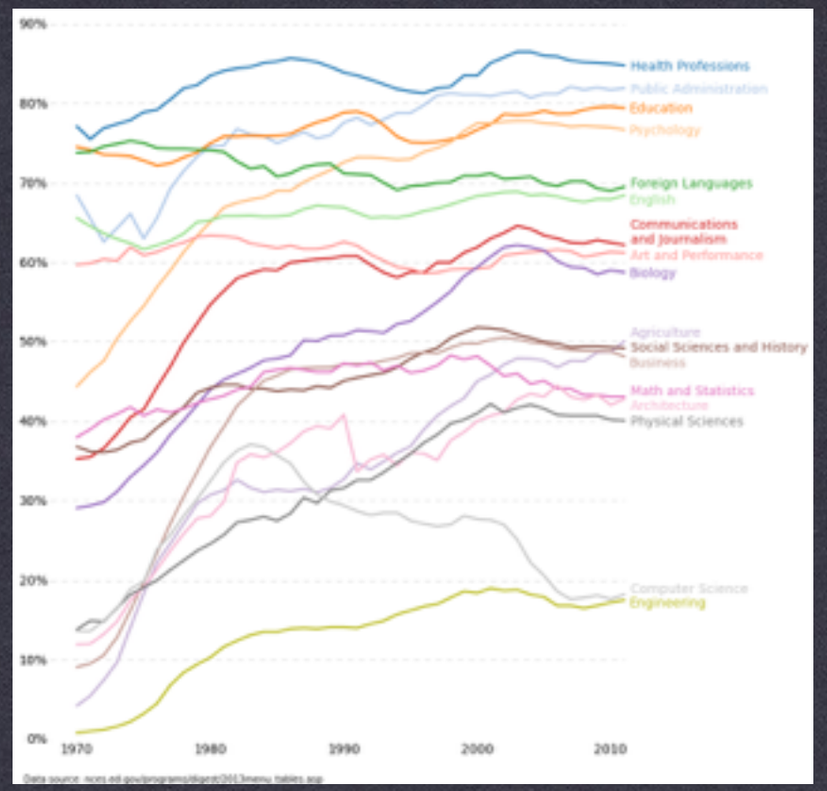
PYTHON与可视化实战

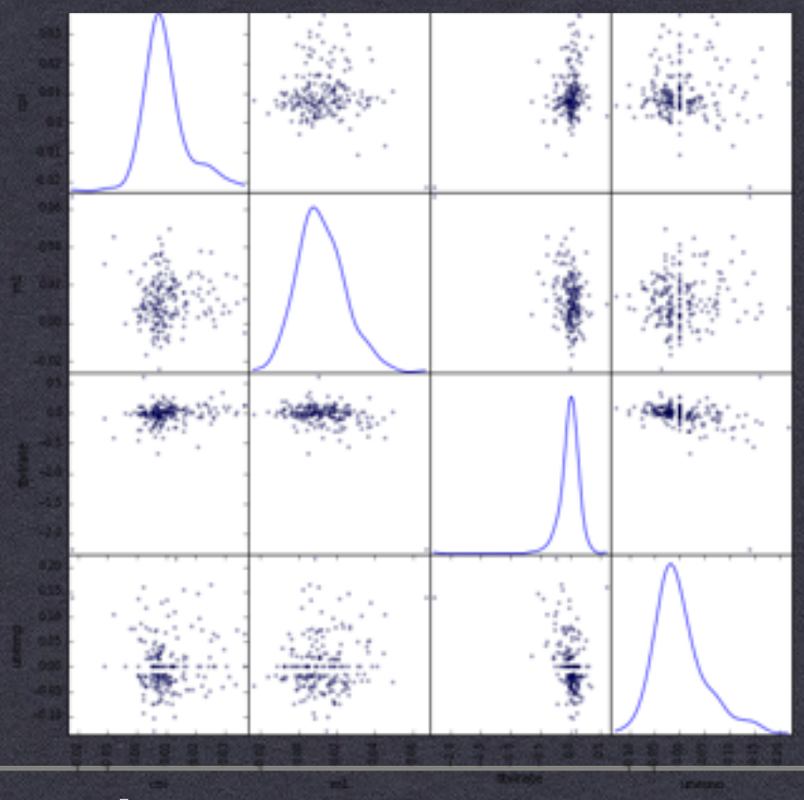
JULYEDU @老冯

内容提要

- Beyond柱状图：可视化能够为我们做些什么
    - 可视化的理论介绍
- 可视化项目入门实战实战
    - 如何使用python进行初步的可视化工作
    - Coding实战
- 知道画什么，比知道怎么画更重要！！！

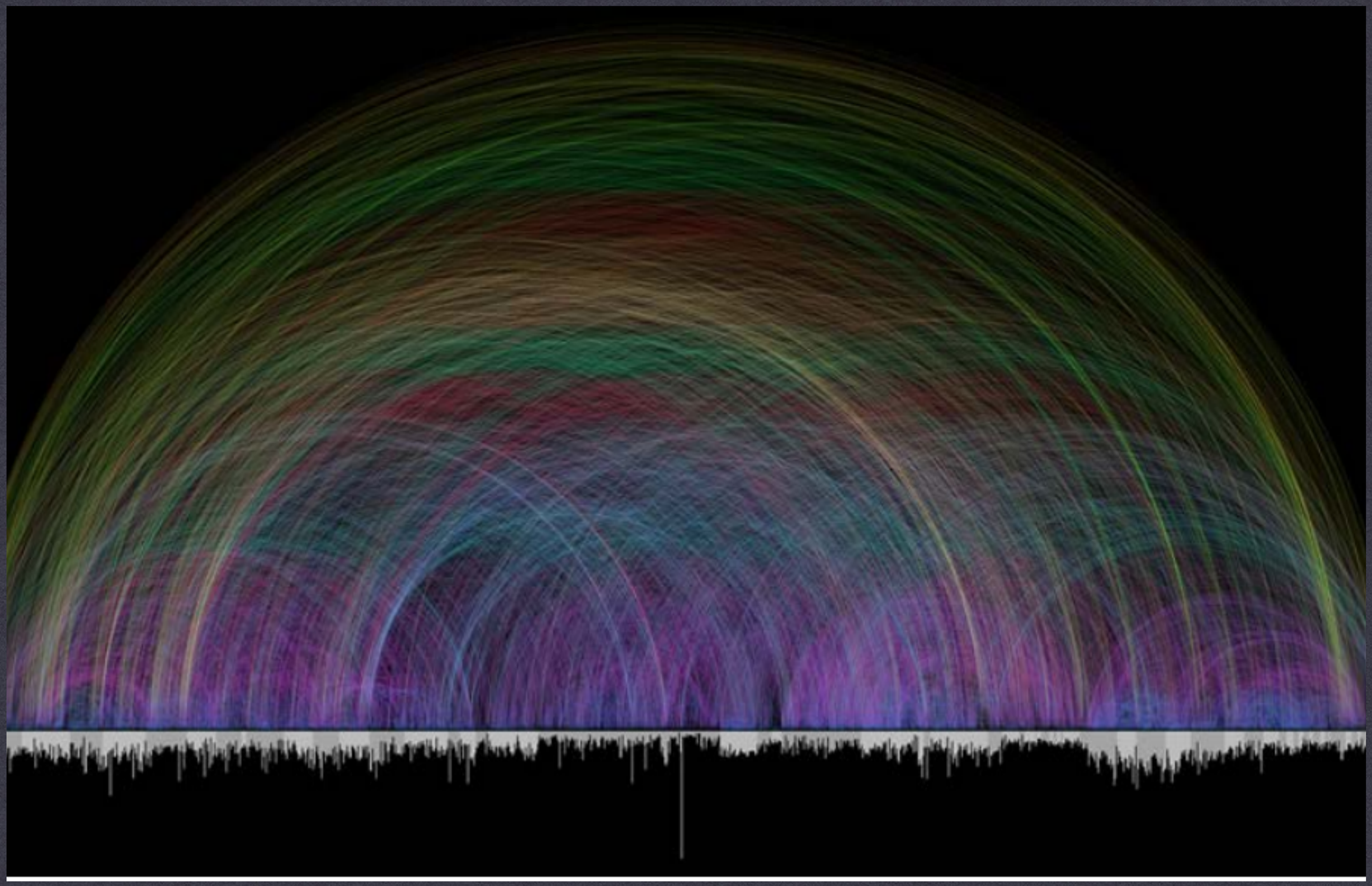




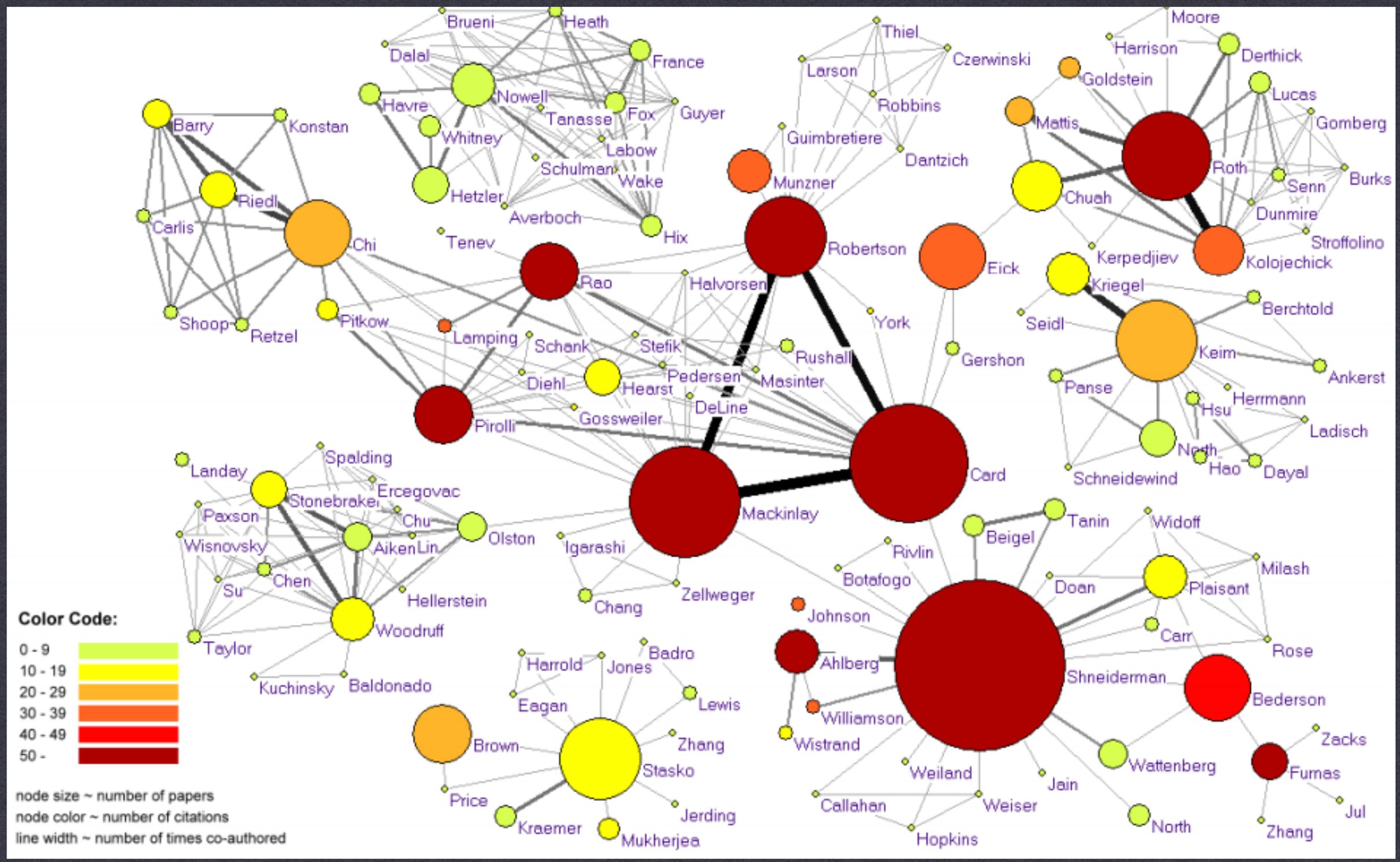


今天课程结束后，你将会利用PYTHON绘制这些图表

更重要的是，学习到可视化的一些理论基石。



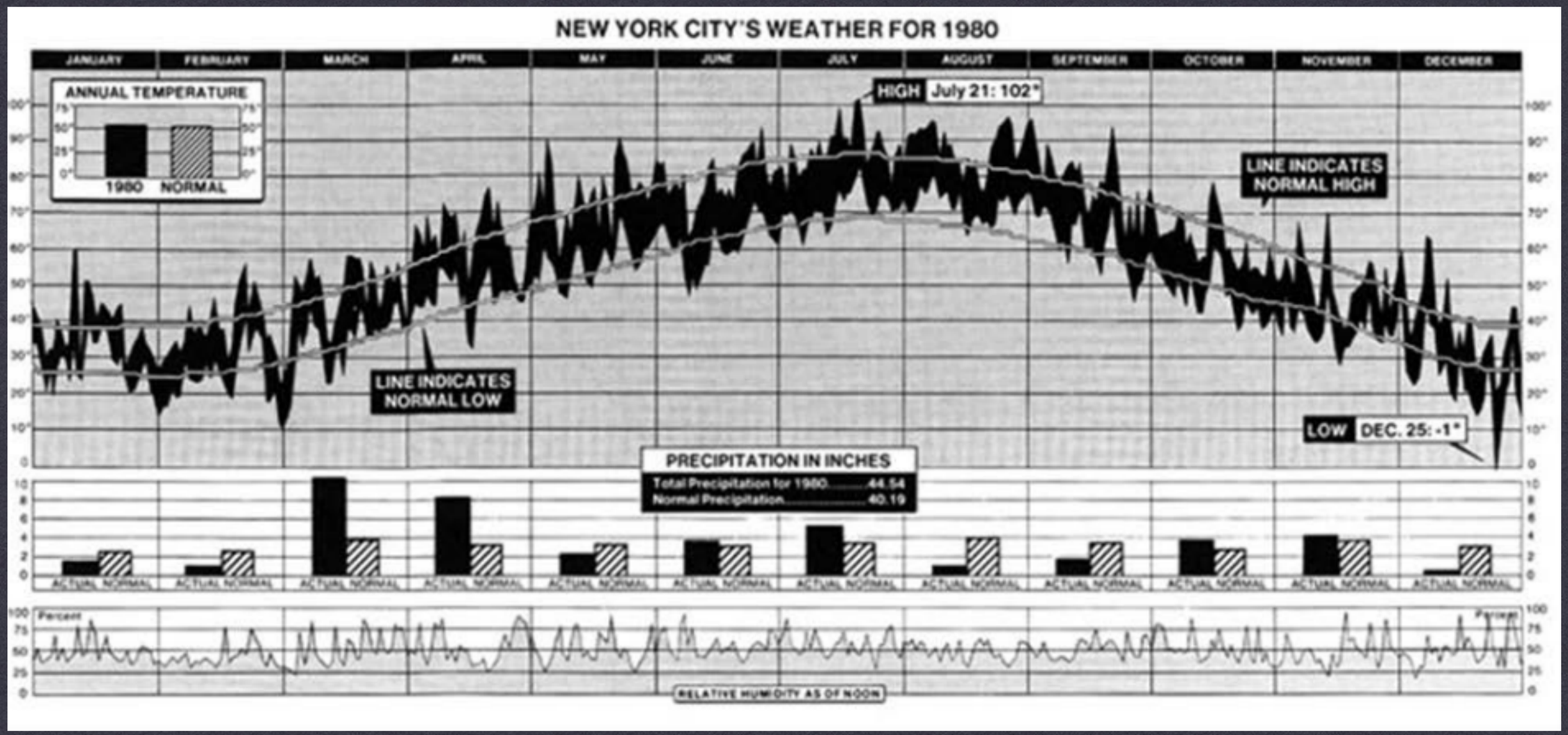
圣经引用可视化



GRAPH VISUALIZATION

洞察数据内涵

历史记载的第一个数据可视化：CHOLERA OUTBREAK



寻找潜在模式

NEW YORK WEATHER

糟糕的可视化：

一些具体案例

内容太多

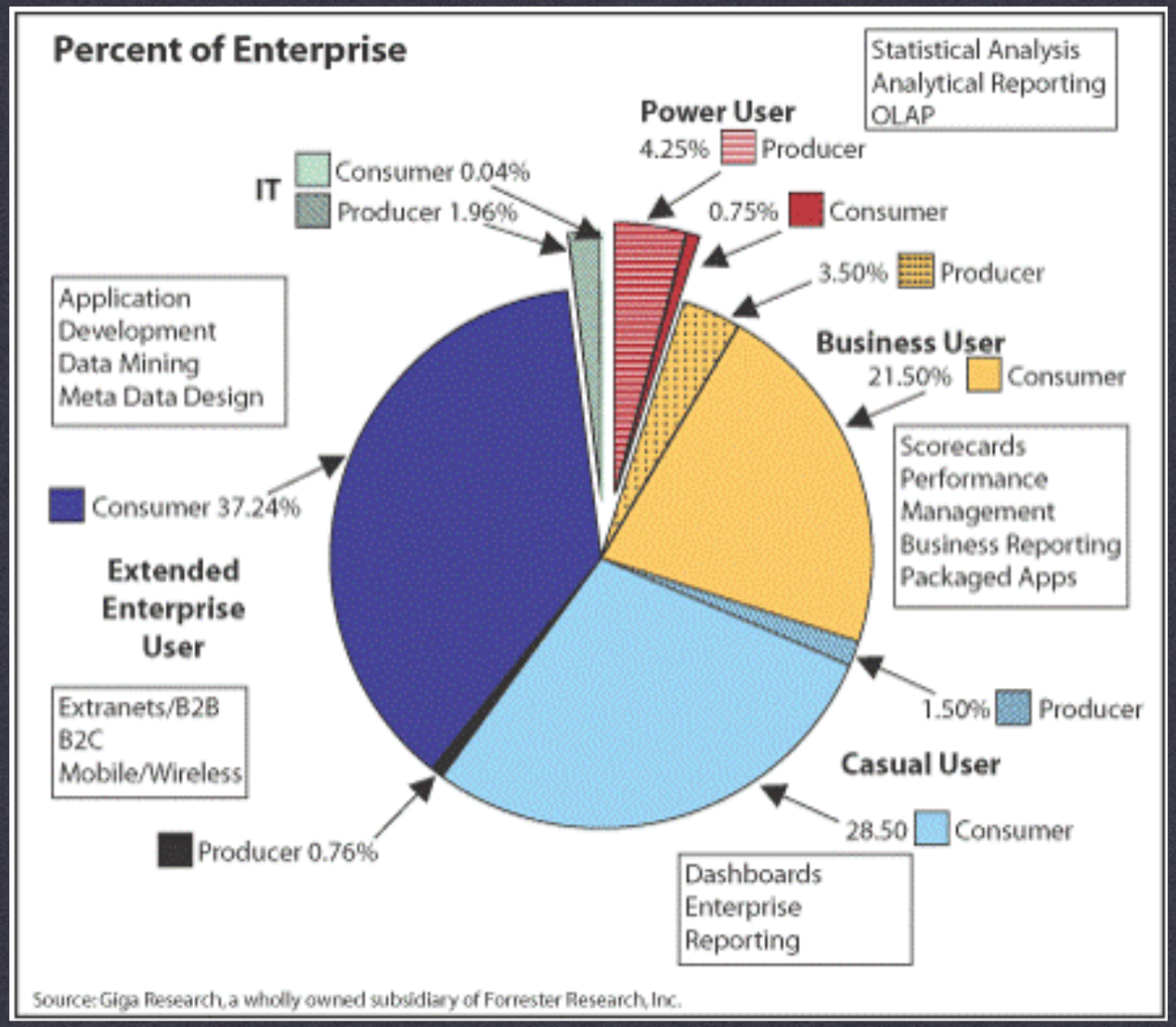
KISS :

KEEP

IT

SIMPLE

STUPID



WRONG SCALE


AVERAGE COST PER QUALIFIED LEAD
$5.28 Print and Traditional Media
$3.09 Adwords
$2.68 Online Advertising
$1.41 SEO
$0.67 LOCAL SEO
According to research done by Statistic Land and CMO

此图问题在哪?

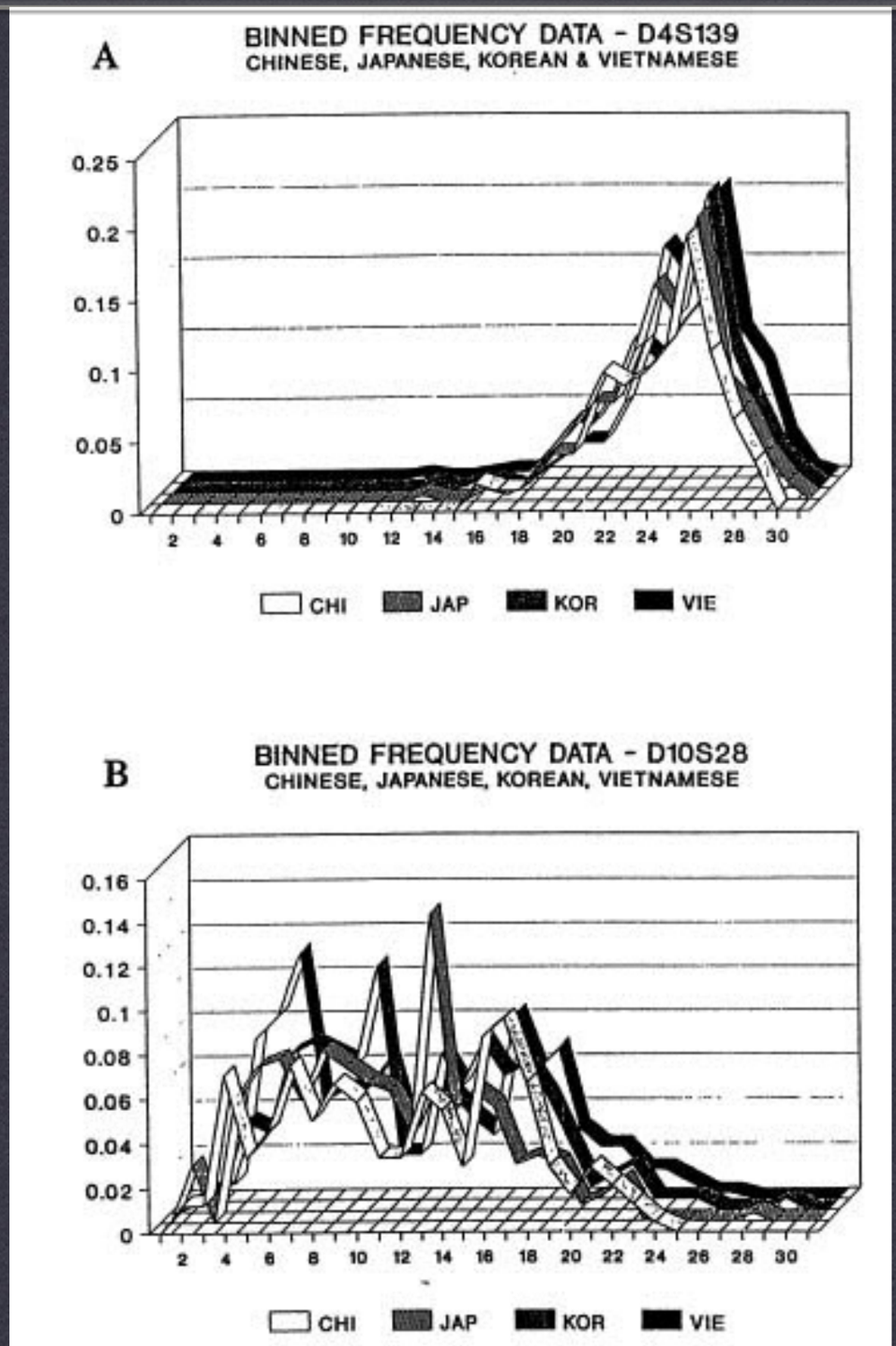


FIG. 4. *Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the* 30 *bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean and Vietnamese are* 103, 125, 93 *and* 215 *for* D4S139 *and* 120, 137, 100 *and* 193 *for* D10S28. *The horizontal axis is the bin number; bins are not of equal length.*

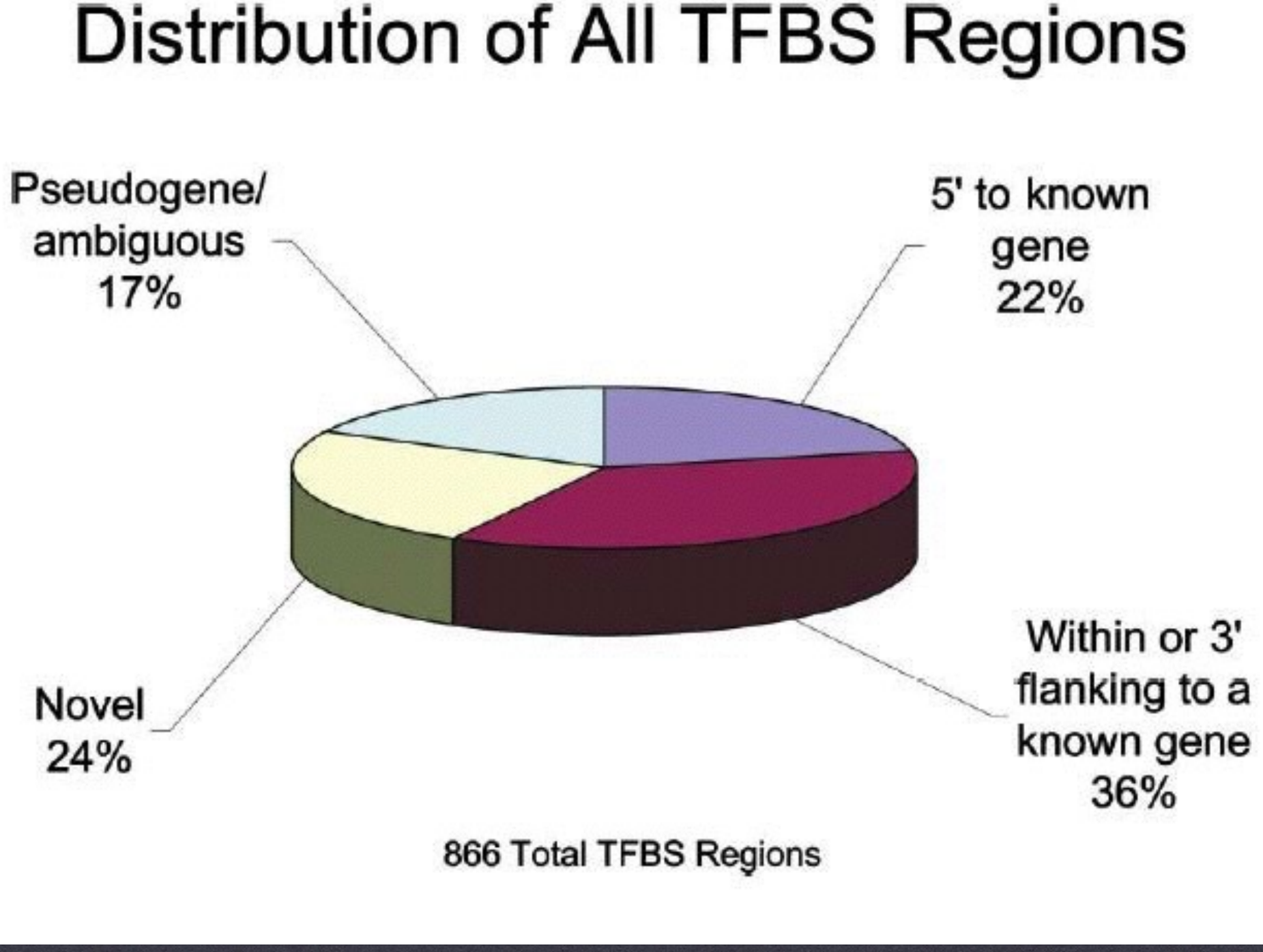
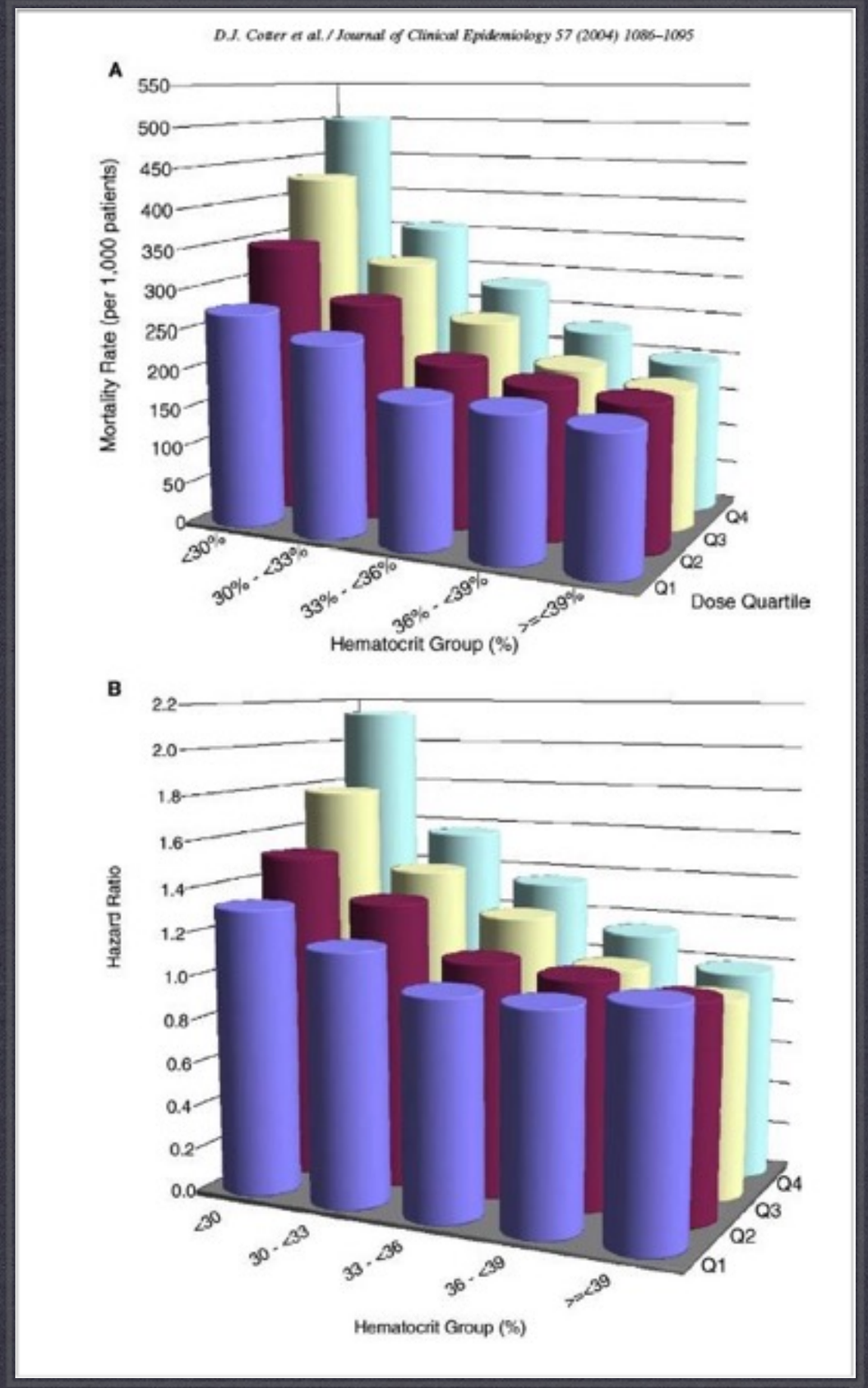


Figure 1. Cla
TFBS regions
classified ba
tions (RefSeq
tions, GenBa
sembl predic
calculated fr
gion. TFBS re
within 5 kb
within 5 kb o
a gene, nove
and pseudog
ping or flan
limited to ch
falling into m
gories).

此图问题在哪儿？？？

此图问题在哪儿


D.J. Cotter et al. / Journal of Clinical Epidemiology 57 (2004) 1086–1095
A
Mortality Rate (per 1,000 patients)
550
500
450
400
350
300
250
200
150
100
50
0
<30%
30% - <33%
33% - <36%
36% - <39%
>=<39%
Hematocrit Group (%)
Q4
Q3
Q2
Q1
Dose Quartile
B
Hazard Ratio
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
<30
30 - <33
33 - <36
36 - <39
>=<39
Hematocrit Group (%)
Q4
Q3
Q2
Q1

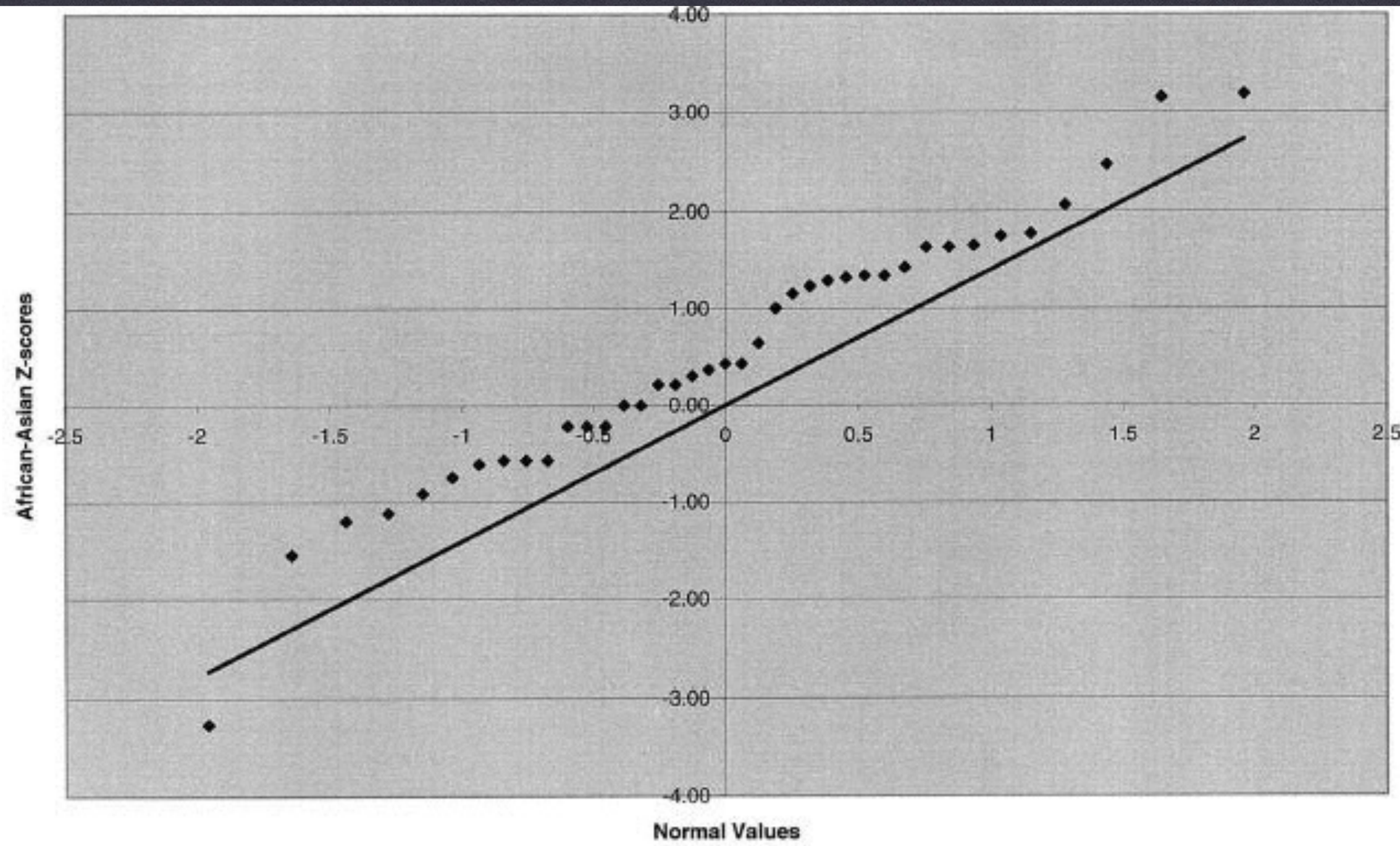
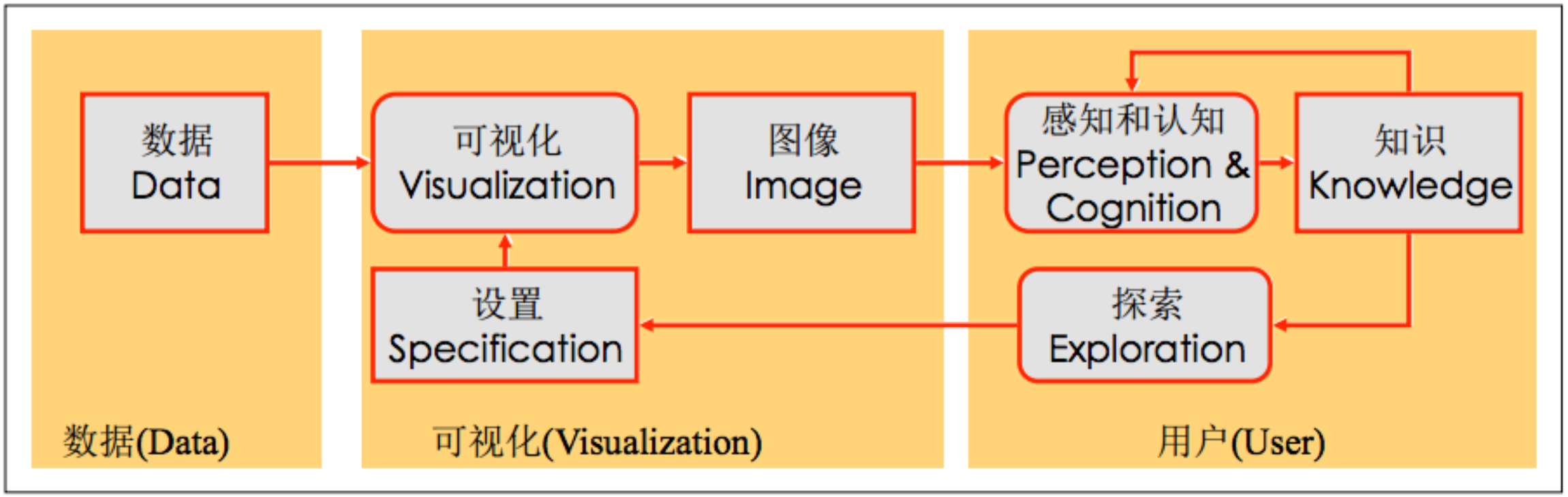


Figure 2 Q-Q plots of Z scores for telomeric interval-length differences. *a*, African Americans versus Asians. *b*, Whites versus Asians.

此图问题在哪儿？？？？？

"The Purpose of Data Visualization is to Convey Information to People."

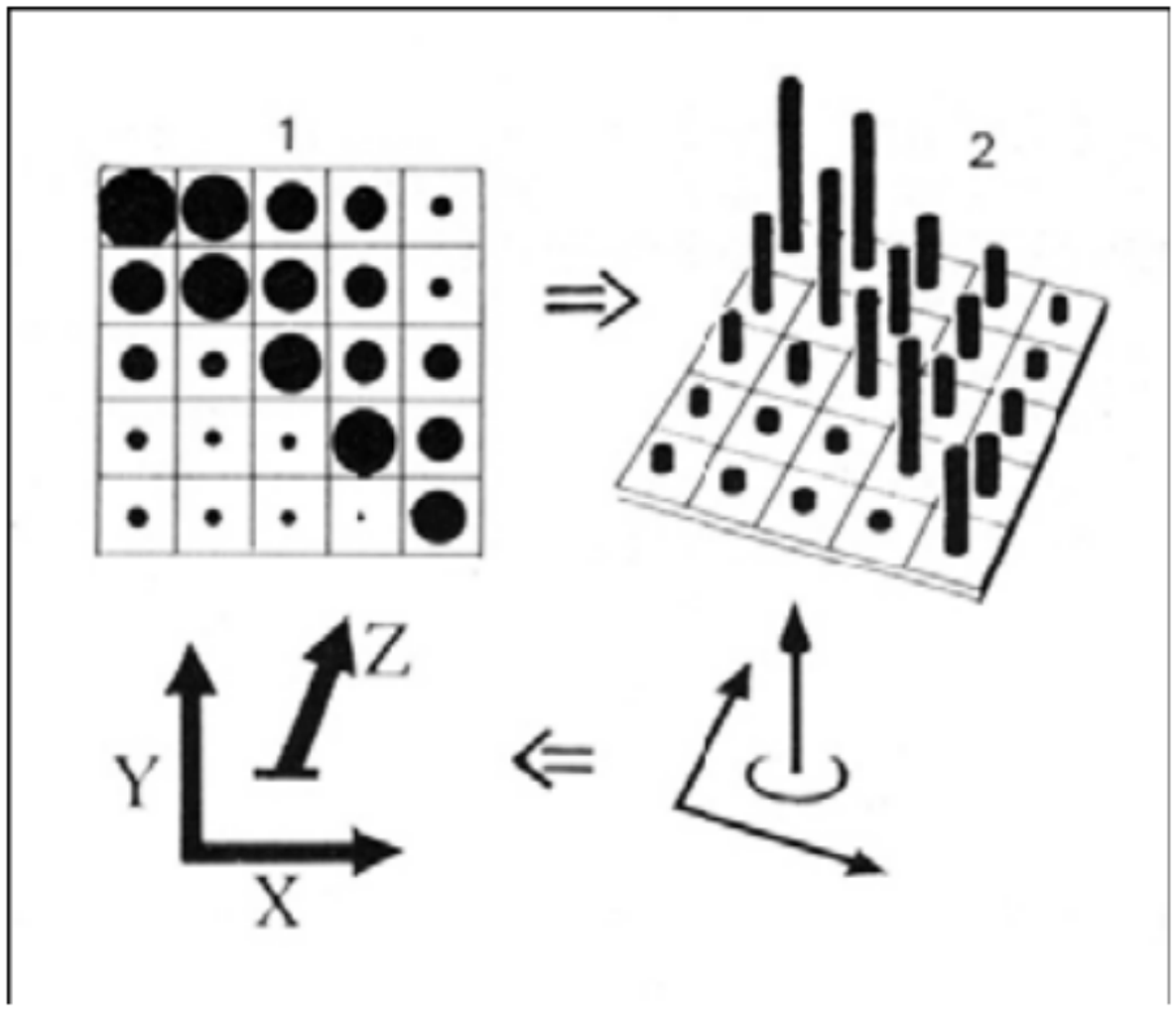
Pat Haranhan, Stanford



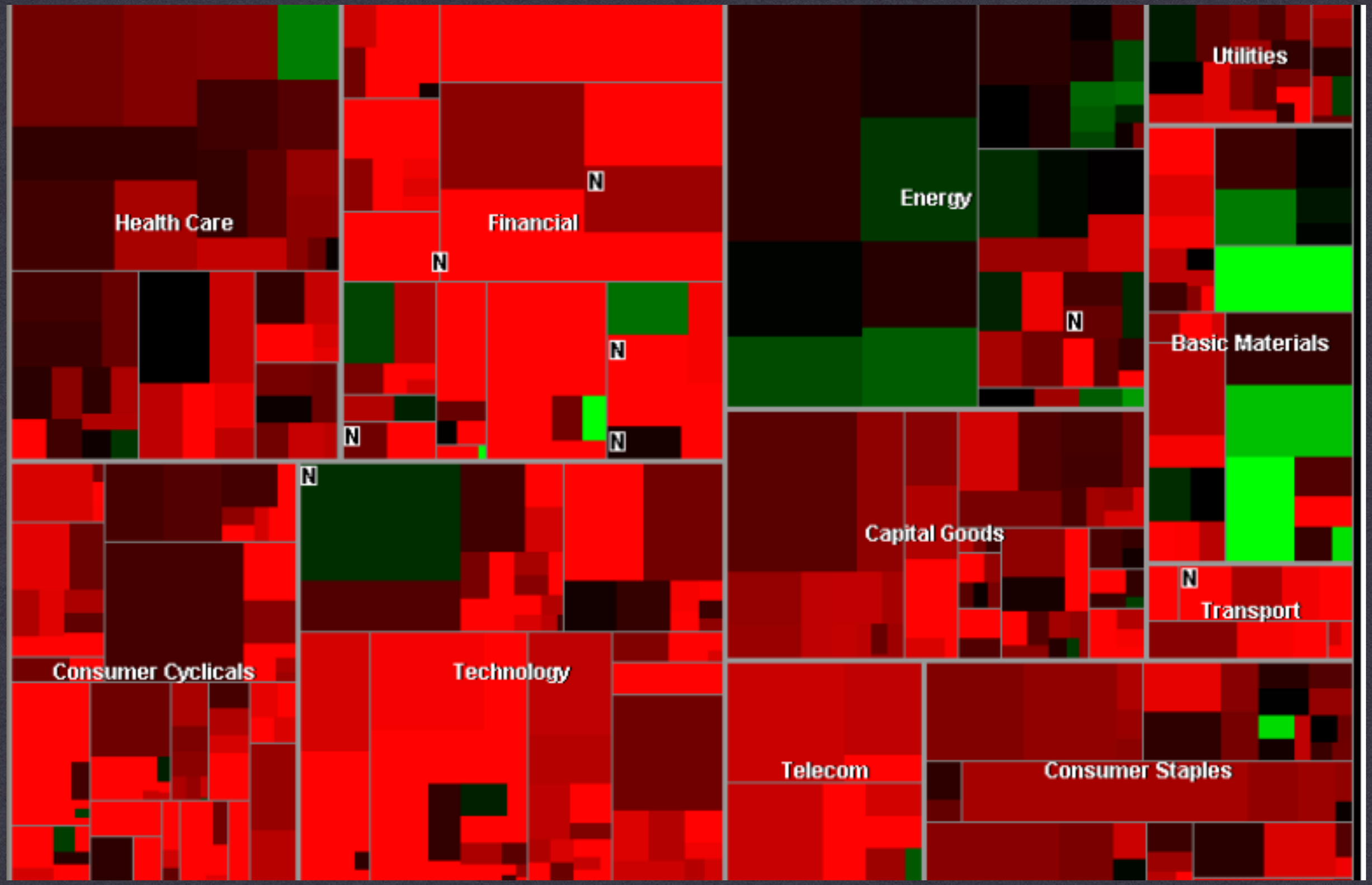
一些可视化设计原则

- 位置
- 大小
- 纹理
- 颜色
- 方向
- 形状

- Two variables [x,y] can map to points
  - Scatterplots, maps, …
- Third variable [z] must use …
  - Color, size, shape, …



MORE THAN 2 DIMENSION..


Health Care
Financial
N
N
N
N
N
Energy
N
Utilities
Basic Materials
N
Transport
Capital Goods
Consumer Cyclicals
Technology
Telecom
Consumer Staples

TREE MAP

代码时间！