

2024春季学期金融学本科选修课程

国际投资- 方法论系列2

数据可视化：原则与技巧

陈方豪 助理教授

经济学院 特区港澳经济研究所

2024年4月11日



暨南大学
JINAN UNIVERSITY

议程

◆ 数据图的重要性

为什么需要数据图？

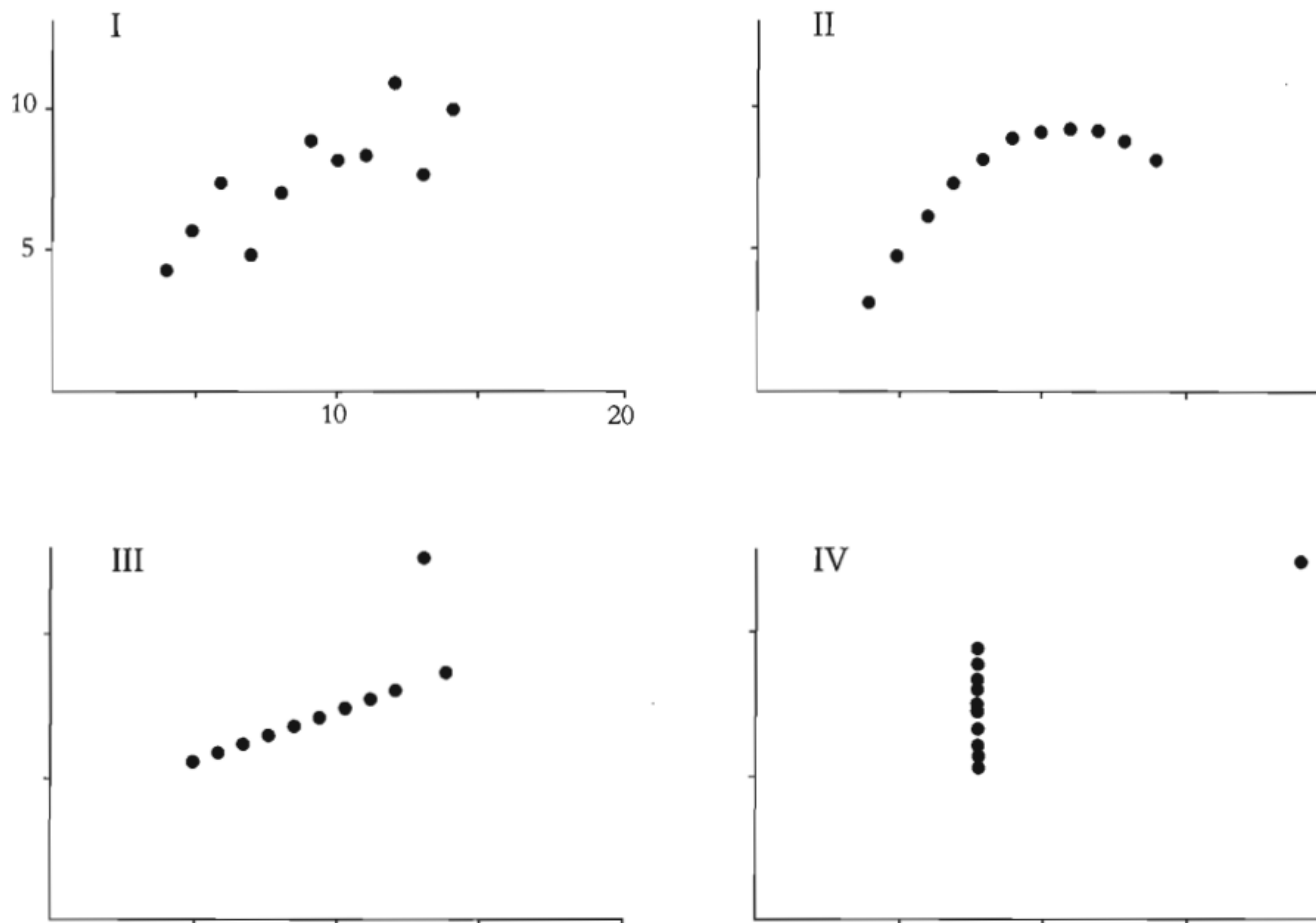
- ◆ 图可以用来发掘一些从数据中难以发掘的规律
- ◆ 呈现关键的信息，降低认知的成本

图 vs. 表: 安斯库姆四重奏 (Anscombe's Quartet)

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

$N = 11$
 mean of X's = 9.0
 mean of Y's = 7.5
 equation of regression line: $Y = 3 + 0.5X$
 standard error of estimate of slope = 0.118
 $t = 4.24$
 sum of squares $X - \bar{X} = 110.0$
 regression sum of squares = 27.50
 residual sum of squares of Y = 13.75
 correlation coefficient = .82
 $r^2 = .67$

图 vs. 表： 安斯库姆四重奏（Anscombe's Quartet）



这张图标志着卫生统计学的诞生



John Snow (1854)



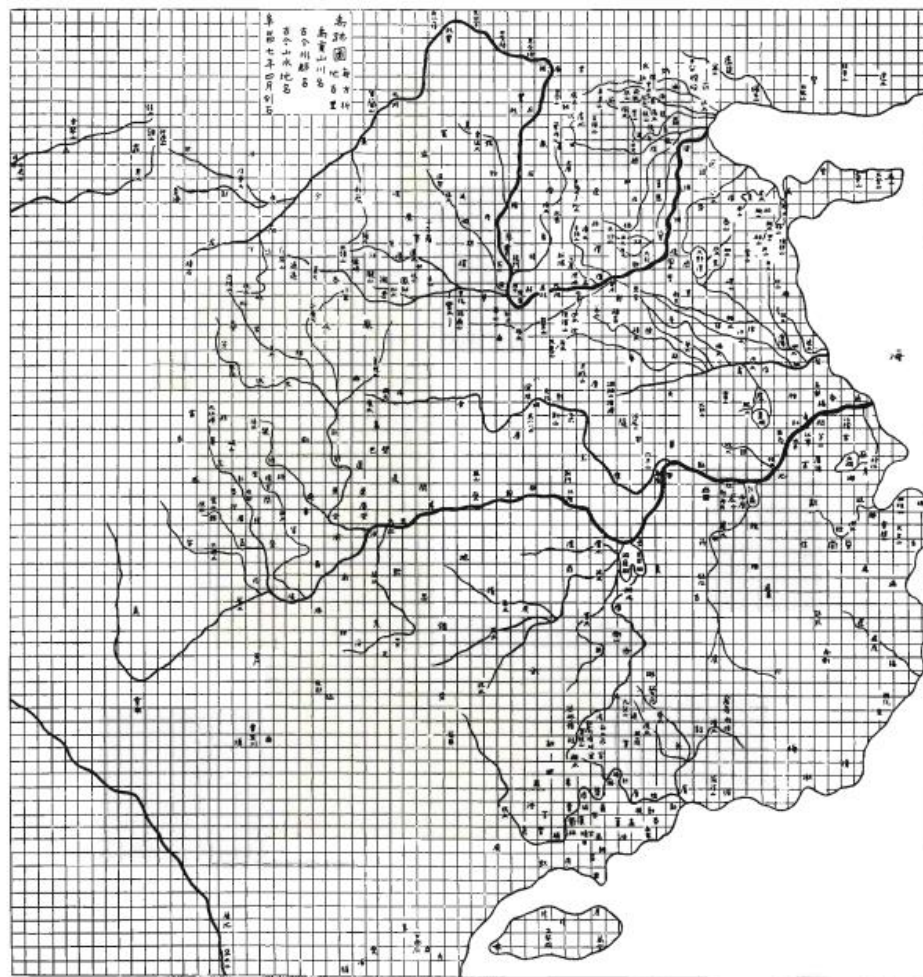
暨南大学
JINAN UNIVERSITY

议程

◆ 数据图的重要性

◆ 数据图的本质

最早的数据图：带坐标的地图



禹迹图（~1000）



最早的数据图：带坐标的地图



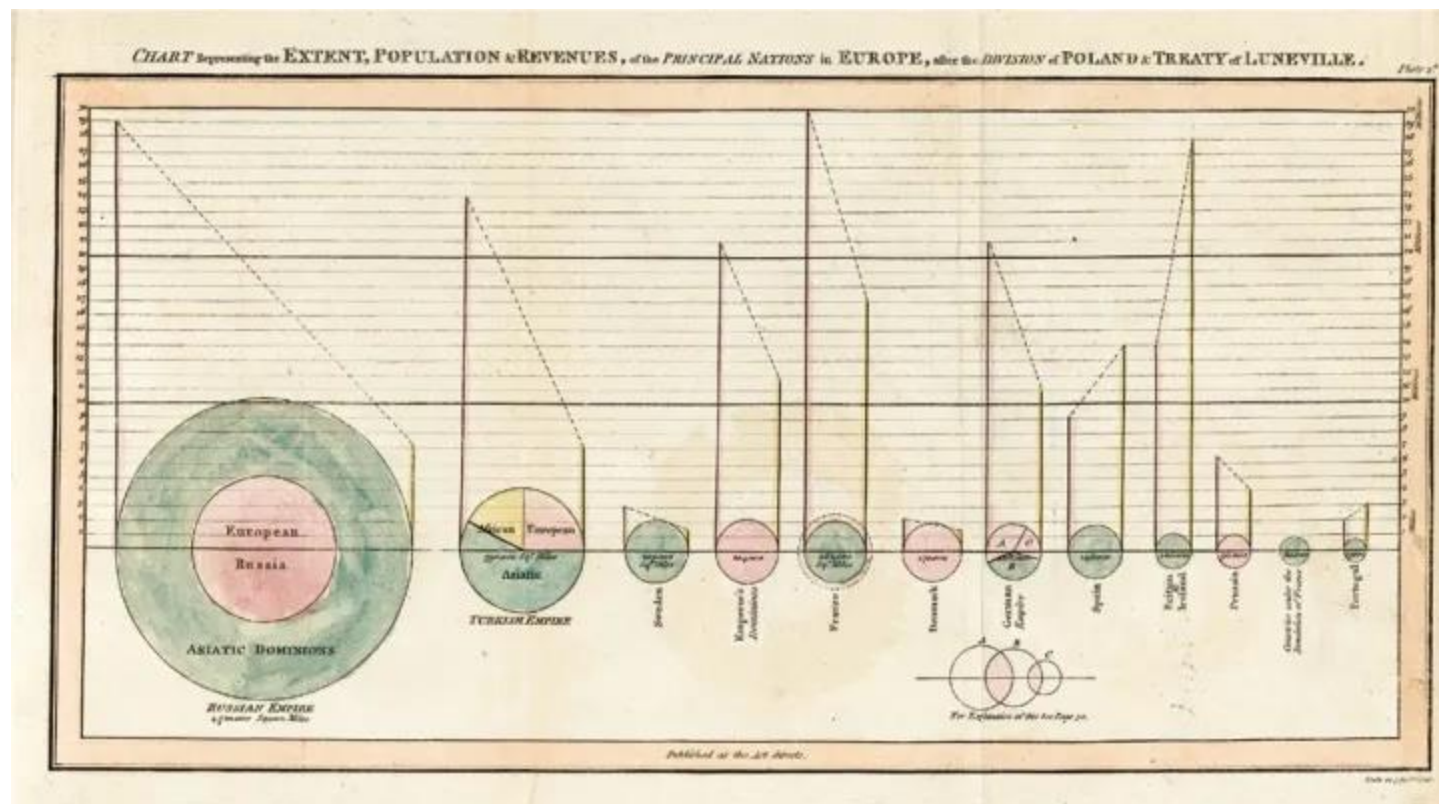
Cosmographia (1546)

重要的一步：抽象

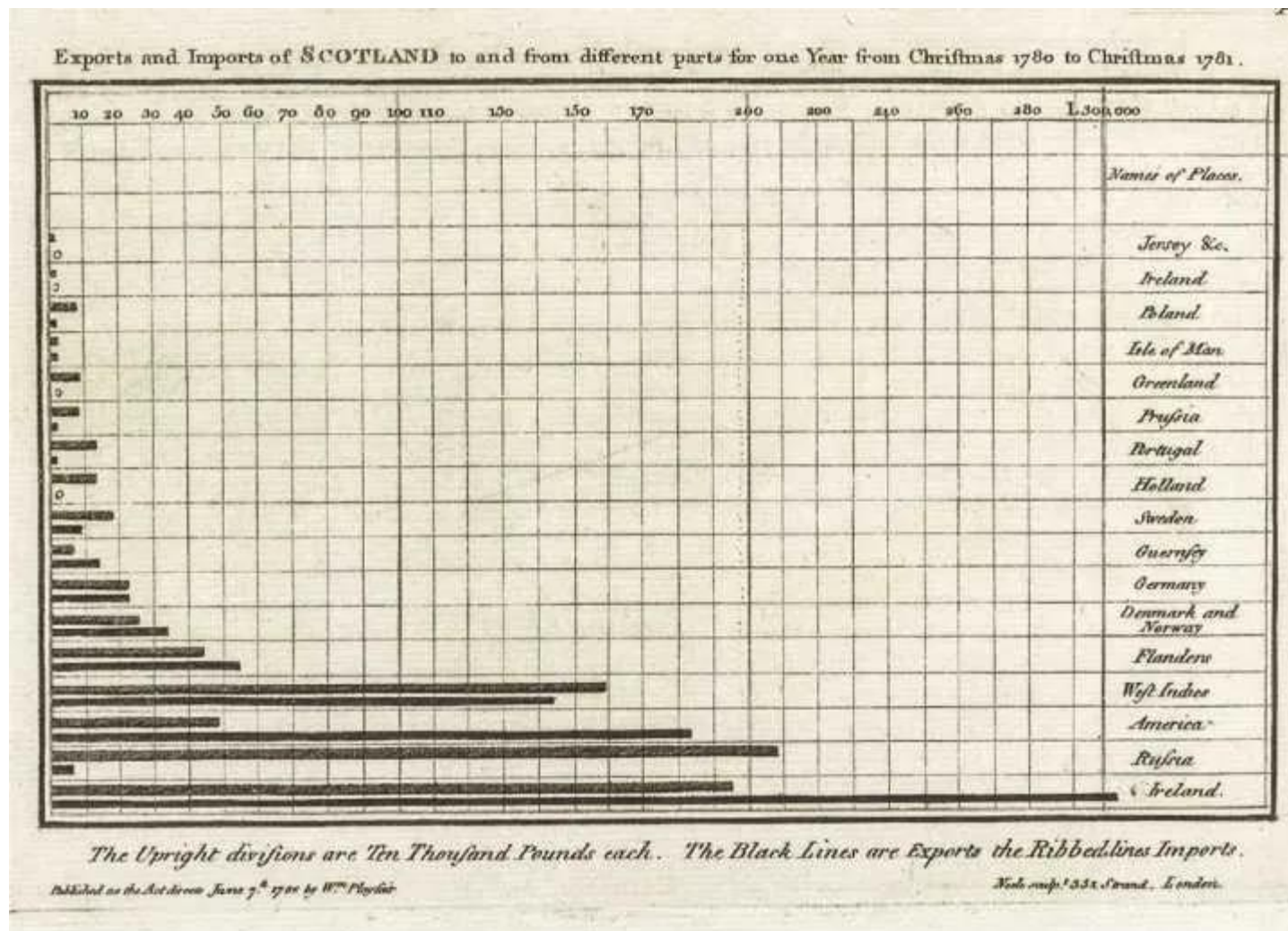
- ◆ 任意一个地点=经度和纬度的相交点
- ◆ 抽象：坐标系可以用来表达地理坐标之外的任何数量关系
 - 地点→观测
 - 经度→变量1，x轴
 - 纬度→变量2，y轴
- ◆ 笛卡尔（1637）：数据可视化的一大步
- ◆ 拓展：抽象的方向可以是反过来的
 - cosine similarity



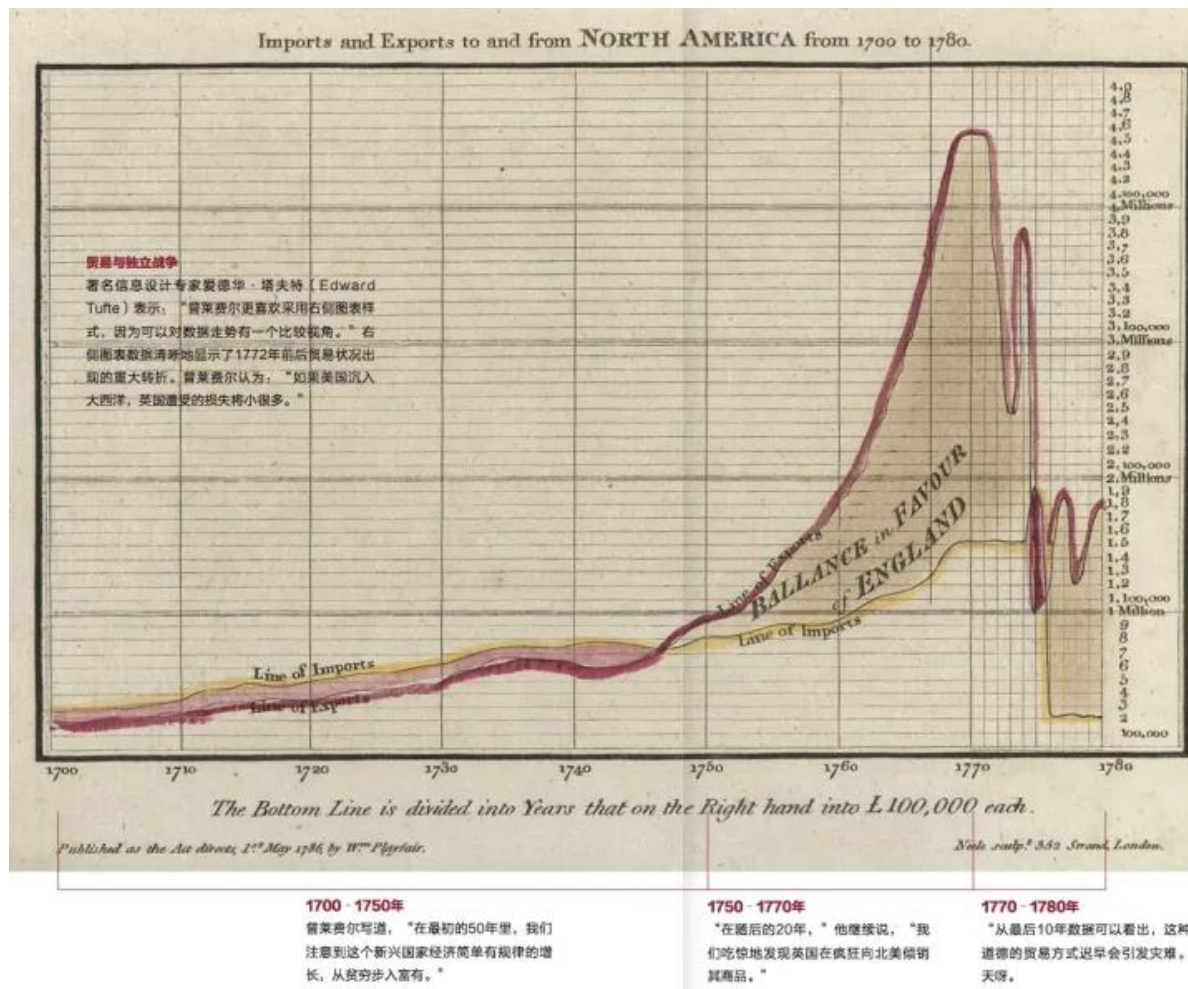
数据可视化之父：Playfair（1786）



数据可视化之父：Playfair（1786）



数据可视化之父：Playfair（1786）



数据图的本质：展示不同维度变量之间的关系

基本图元/ 几何对象	第 1 维度	第 2 维度	第 3 维度	第 4 维度	第 5 维度
树形图	面积大小	颜色			
饼状图/ 环状图	扇面大小				
折线图	X 轴	Y 轴	次坐标轴	颜色	虚实
面积图	X 轴	Y 轴	次坐标轴	颜色	
柱形图	X 轴	Y 轴	次坐标轴	颜色	宽窄
散点图	X 轴	Y 轴	大小	颜色	(符号)形状
填充地图	坐标位置	颜色			
符号地图	坐标位置	大小	颜色	(符号)形状	
雷达图	高低	颜色			
股价图	X 轴	Y 轴	颜色		

数据图的本质：展示不同维度变量之间的关系

- ◆ 像“GDP在增长”这样一个只包含GDP一个对象的论述不是也可以用图形描述吗？
- ◆ 这是因为：其实当你在使用“增长”这个词的时候已经隐含了另一个维度的信息：时间。
 - 事实上，这一类图表有一个专门的名字：时间序列（Time Series）
- ◆ 所以，很多看似在描述一个变量的语句其实也隐蔽地描述了变量之间的关系。



议程

◆ 数据图的重要性

◆ 数据图的本质

◆ 作图原则



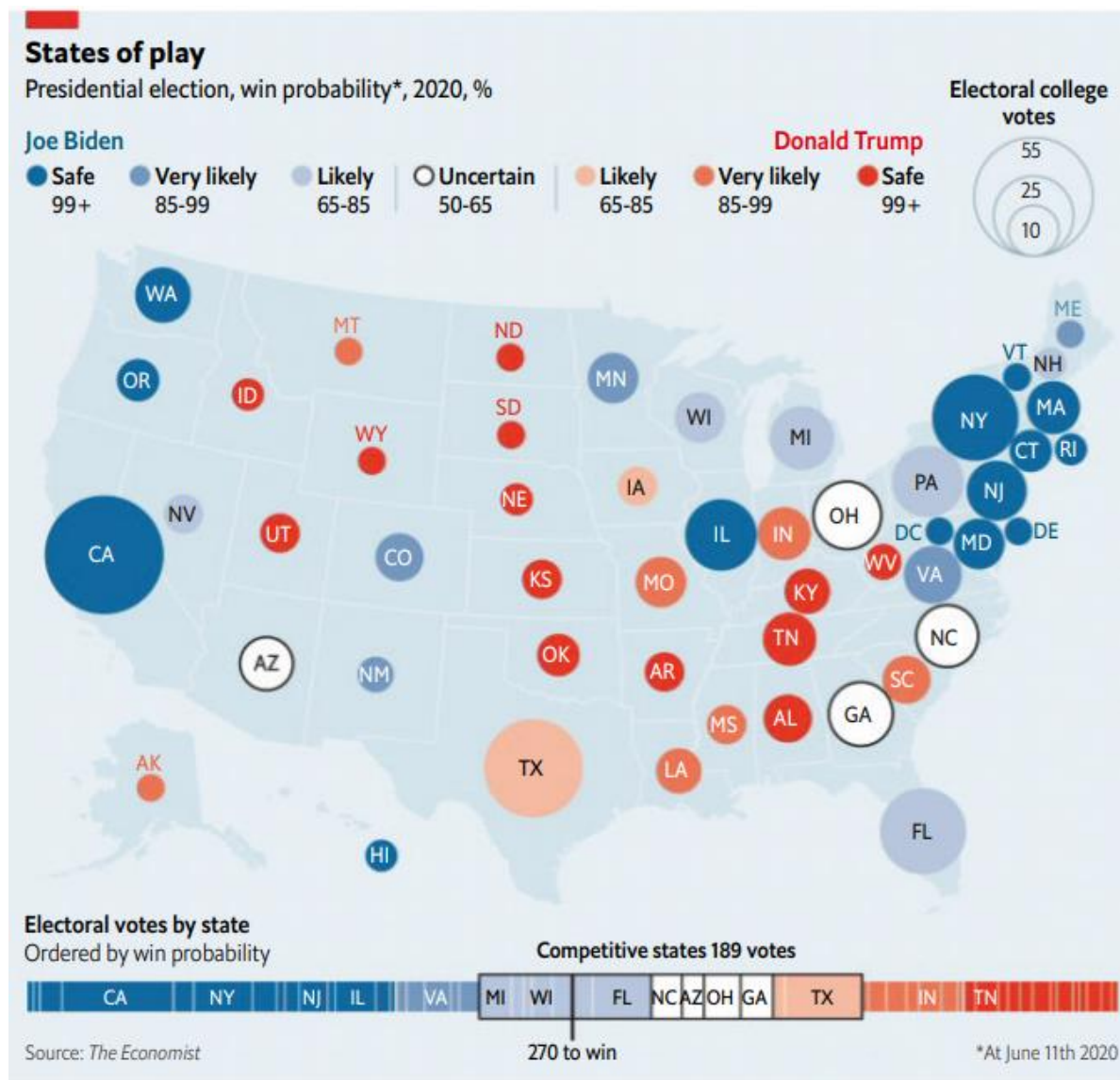
信息太少→信息太多

- ◆ 当一件事物的成本降低的时候，人们倾向于去多做或多用它。
- ◆ 现代统计软件将制作一张复杂图表的成本几乎降至为零。这反而催生出了很多副作用：我们看到很多现代人制作的图表，包含了太多无关的庞杂信息，形式华丽但并不直观，反而是增加了人的认知难度。
- ◆ 为了克服这个问题，我们必须秉持“从一而终”的态度来进行数据作图，从自己想要传达的信息出发，“如无必要，勿增实体”。

《经济学人》 Economist



《经济学人》 Economist

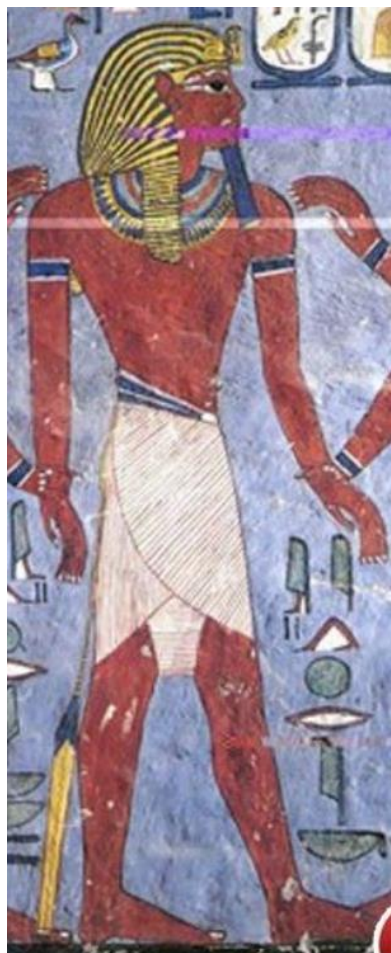


作图原则

- ◆ 原则1：保持风格的一致和优美
- ◆ 原则2：少既是多，排除无关信息
- ◆ 原则3：减少多余的展示元素和维度
- ◆ 原则4：注释靠近其所解释的元素
- ◆ 原则5*：让图自己讲述故事

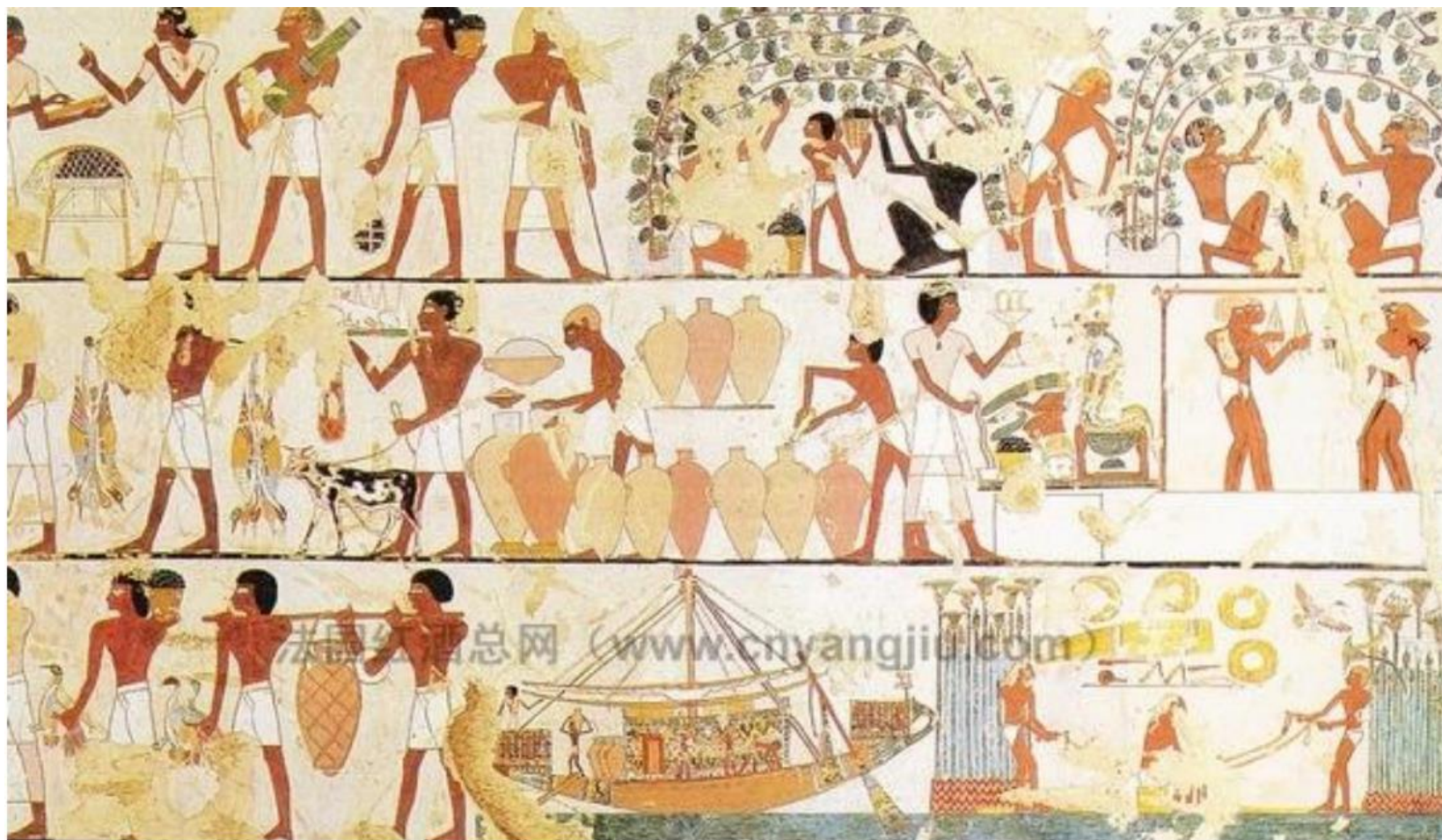


原则1：保持风格的一致与优美： 古埃及的人像难道不奇怪嘛？



- ◆ 侧面的头
- ◆ 正面的躯干
- ◆ 侧面的手臂
- ◆ 侧面的脚

原则1：保持风格的一致与优美： 统一的风格带来美感



原则1：保持风格的一致与优美： 视觉上的和谐，跟图像的视觉可压缩性直接相关

- ◆ 布局规律（元素少，位置规律）
- ◆ 色彩简单（主题颜色较少）
- ◆ 可预测性（图片中元素的规律性）
- ◆ 图形/物体常见（易于找到匹配的模式）

原则1：保持风格的一致与优美： 视觉上的和谐，跟图像的视觉可压缩性直接相关

易于压缩

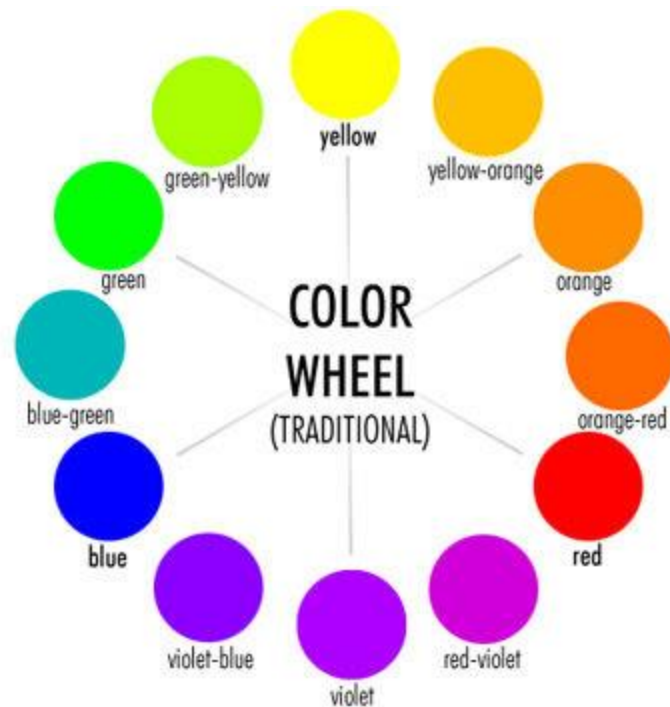


不易压缩



原则1：保持风格的一致与优美： 配色原理

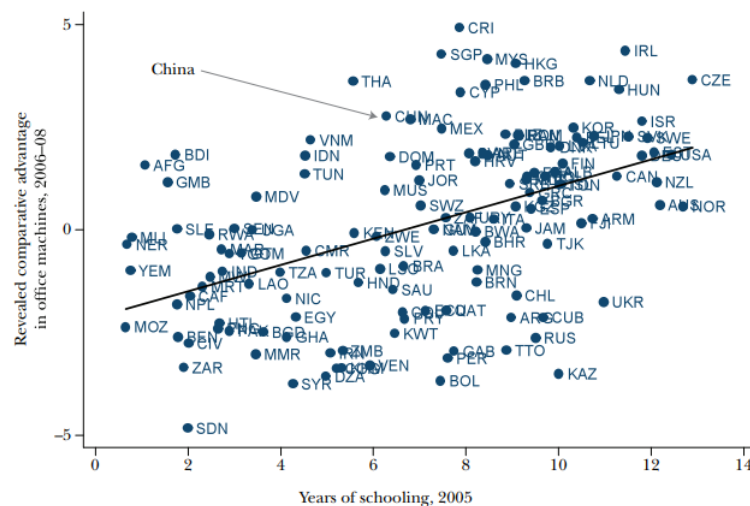
- ◆ 相似色，互补色，拟自然色
- ◆ 有时出版物要求提交不对颜色敏感的图片；对彩色图片额外收费
 - 黑白配色
 - 公式：{深，浅} × {实，虚}
 - 超过4个要素尽量不要画在一幅图



原则2：少既是多，排除无关信息

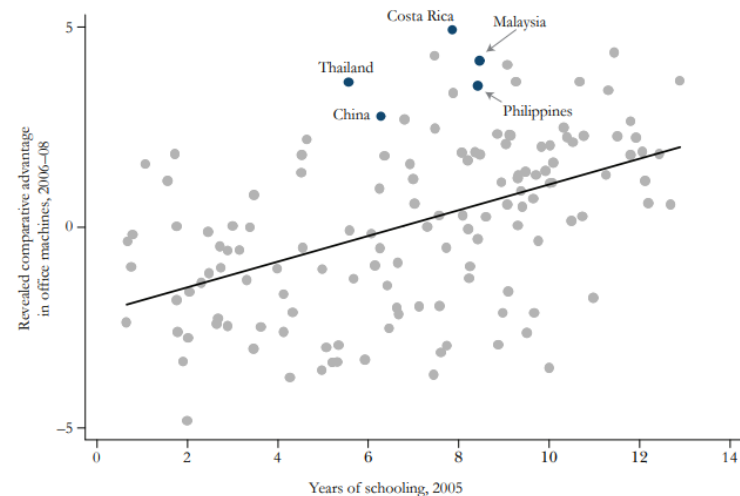
修改前

Education and Exports of Office Machines



修改后

Education and Exports of Office Machines



Source: Hanson (2012).

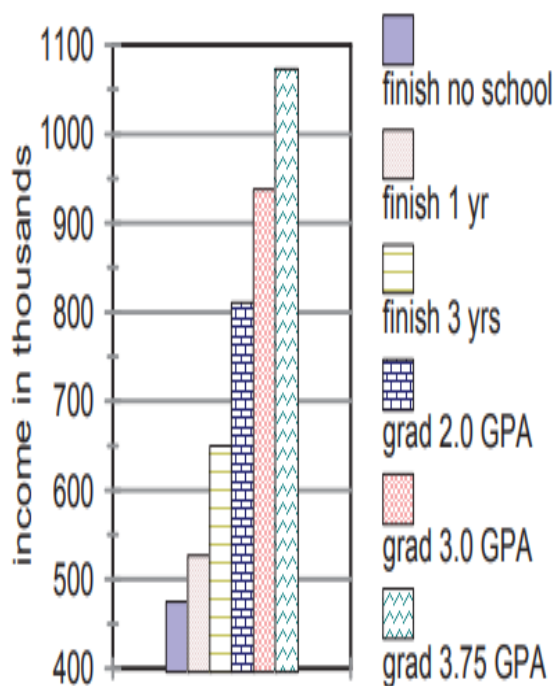


暨南大学
JINAN UNIVERSITY

原则3：减少多余的展示元素和维度

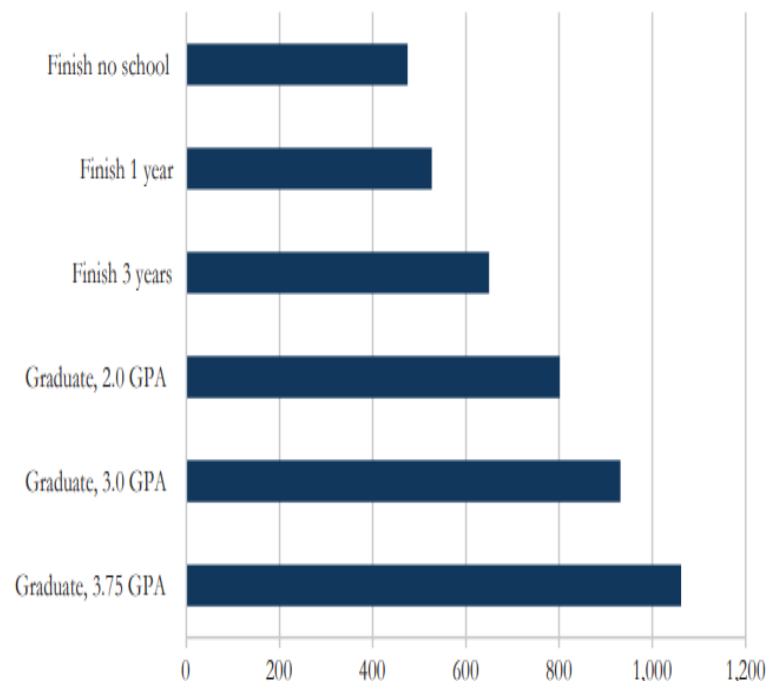
修改前

Figure 2 Discounted Expected Lifetime Earnings, VN(t')



修改后

Discounted Expected Lifetime Earnings, VN(t')
(Income in thousands)

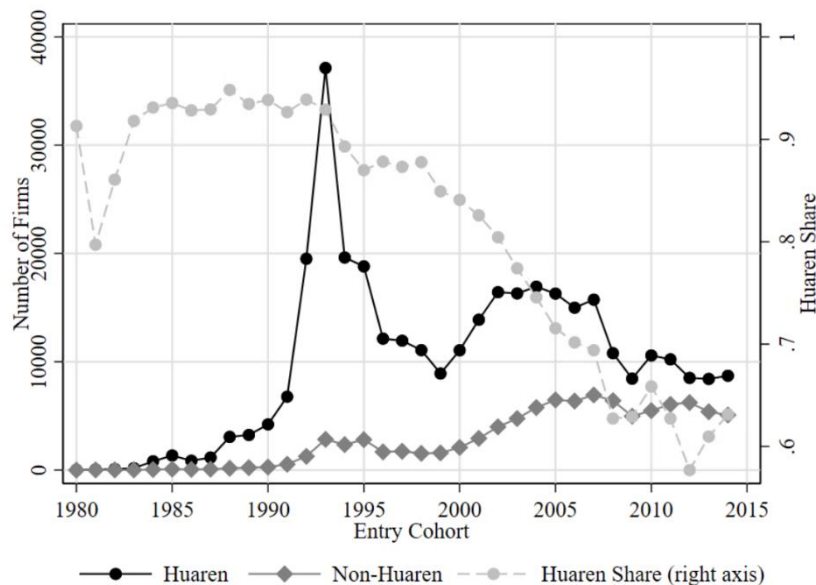


Source: Stinebrickner and Stinebrickner (2013).

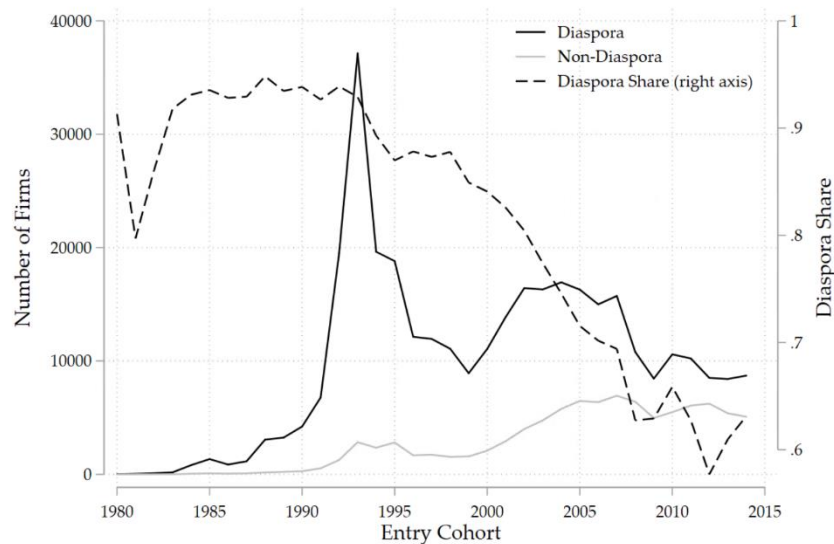


原则3：减少多余的展示元素和维度

修改前



修改后

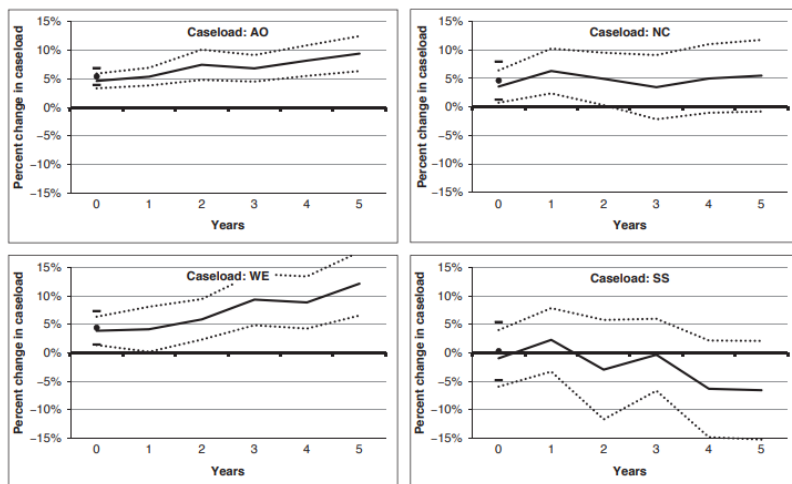


Chen, F., Xiong, R. & Zhang, X. “Familiar Strangers: The Role of Diaspora Networks in Foreign Investment and Long-run Development.”



原则4：注释靠近其所解释的元素

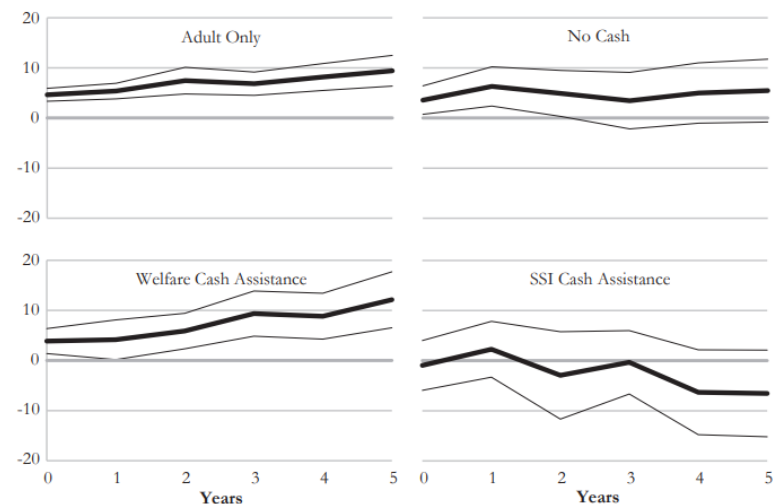
修改前



Source: Klerman and Danielson (2011).

修改后

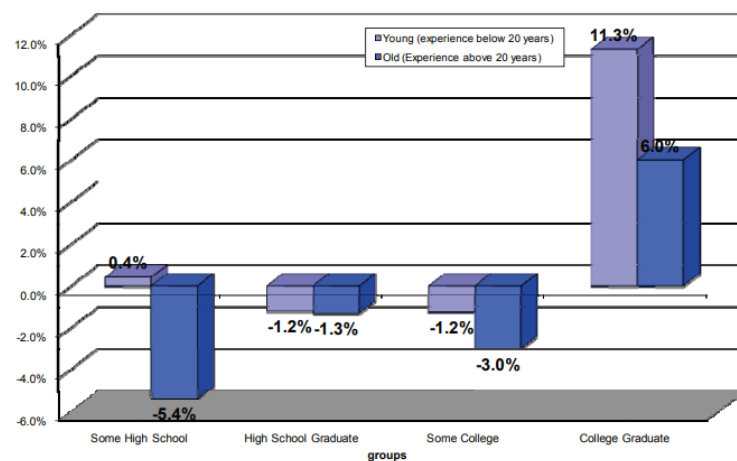
Implied Impulse Response Functions for Different Caseloads
(Percent change)



综合修改

修改前

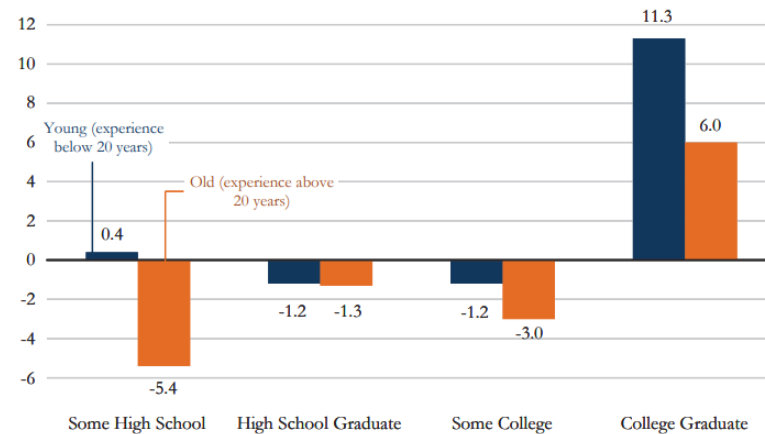
Change in real weekly wages of US-born workers by group, 1990-2006



Source: Ottaviano and Peri (2008).

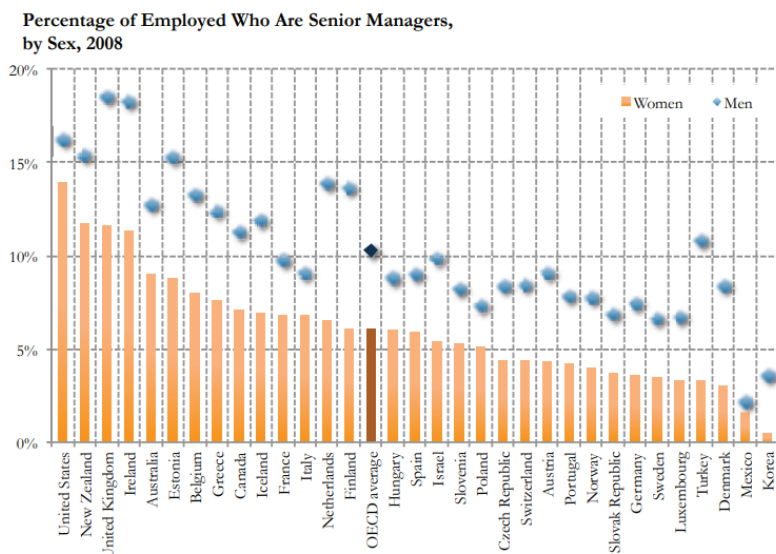
修改后

Change in real weekly wages of US-born workers by group, 1990-2006
(Percent)



综合修改

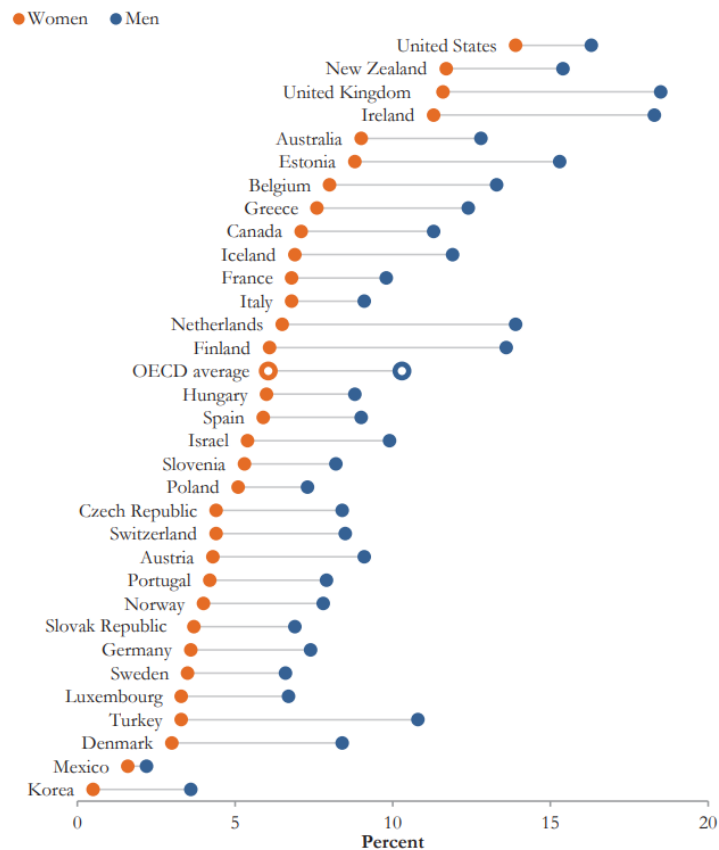
修改前



Source: Author, based on OECD (no date) and Rampell (2013).

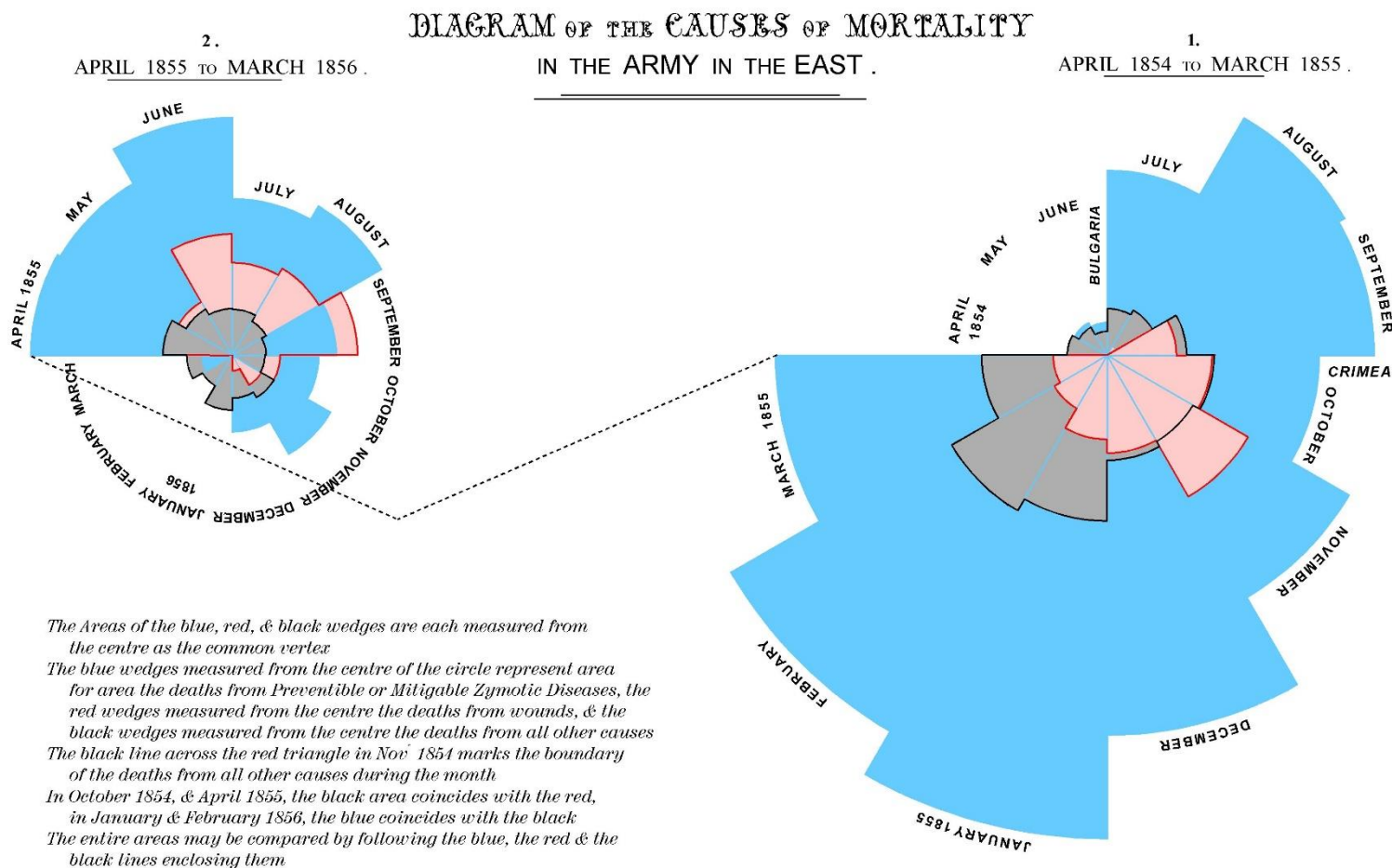
修改后

Percentage of Employed Who Are Senior Managers, by Gender, 2008 (percent)



原则5*：让图自己讲述故事

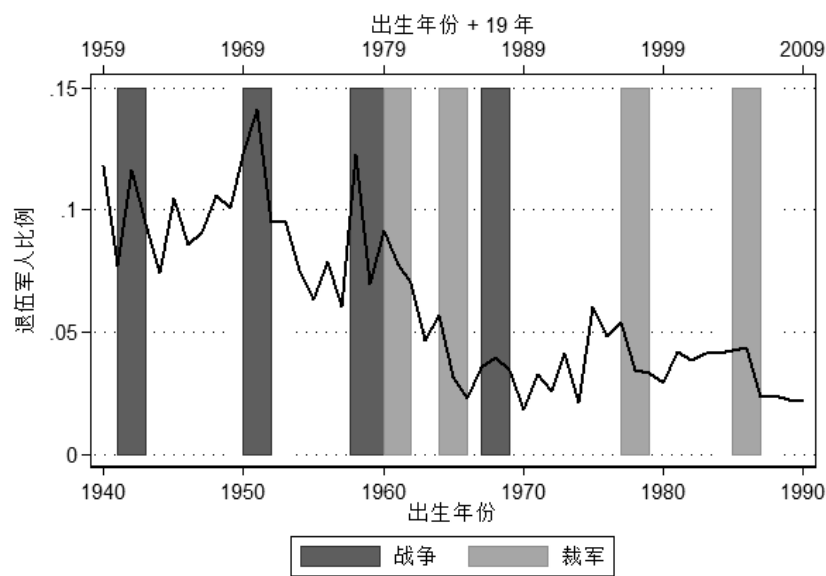
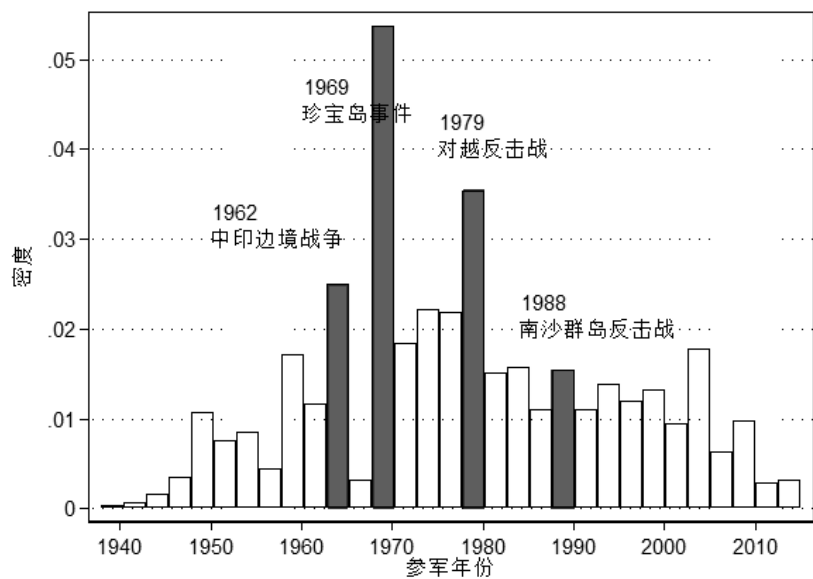
<http://www.Florence-Nightingale-Avengeing-Angel.co.uk/Corcomb.htm>



南丁格尔玫瑰图



原则5*：让图自己讲述故事



陈方豪，张晓波，《军人是天生的创业者吗？——参军经历与创业行为》

谢 谢

- ◆ 陈方豪 助理教授
- ◆ 经济学院 特区港澳经济研究所
- ◆ 2022年毕业于北京大学国家发展研究院，获经济学博士学位，多伦多大学访问学者；主要研究领域为发展经济学、城市经济学与国际经济学；研究主线是中国的区域产业发展与全球市场的关系，中国在全球价值链中的位置，以及中国企业的国际化进程。当前的研究重心为华侨华人与中国产业发展之间的联系。
- ◆ 邮箱：chenfanghao@jnu.edu.cn
- ◆ 个人网页：<https://fanghaochen.github.io/homepage/>

