

# 经济数据解读02 (丽湖校区)

## 第二次线下课

黄嘉平  
中国经济特区研究中心  
[huangjp.com](http://huangjp.com)  
2024.3.28

# 关于小组汇报的分组

- 最终选课人数为51人，总可用时间为240分钟  
⇒ 3人一组，共17组，每组约12-13分钟  
        第四次课（6组）、第五次课（6组）、第六次课（5组）
- 分组第一阶段：自由组队，在4月14日前报给助教姜宏卓（可以在QQ里私聊）。
- 分组第二阶段：第一阶段没有组队的同学，我们会随机分组，并在4月18日第三次课上公布结果。
- 汇报顺序：按随机原则分配，在4月18日第三次课上公布结果。确实需要调整的小组可先在课下和其他组自行协调，双方同意后将调整结果报给助教

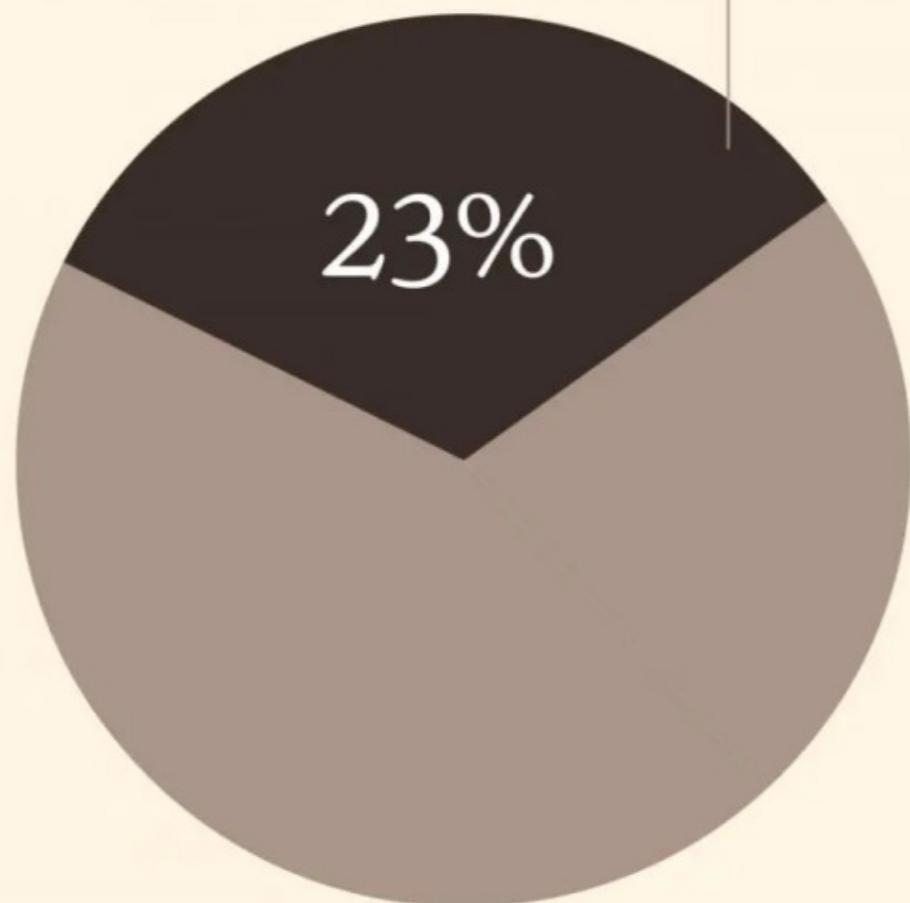
# 数据的可视化

UCDavis DataLab, *Principles of Data Visualization*,  
[https://ucdavisdatalab.github.io/workshop\\_data\\_viz\\_principles/](https://ucdavisdatalab.github.io/workshop_data_viz_principles/)

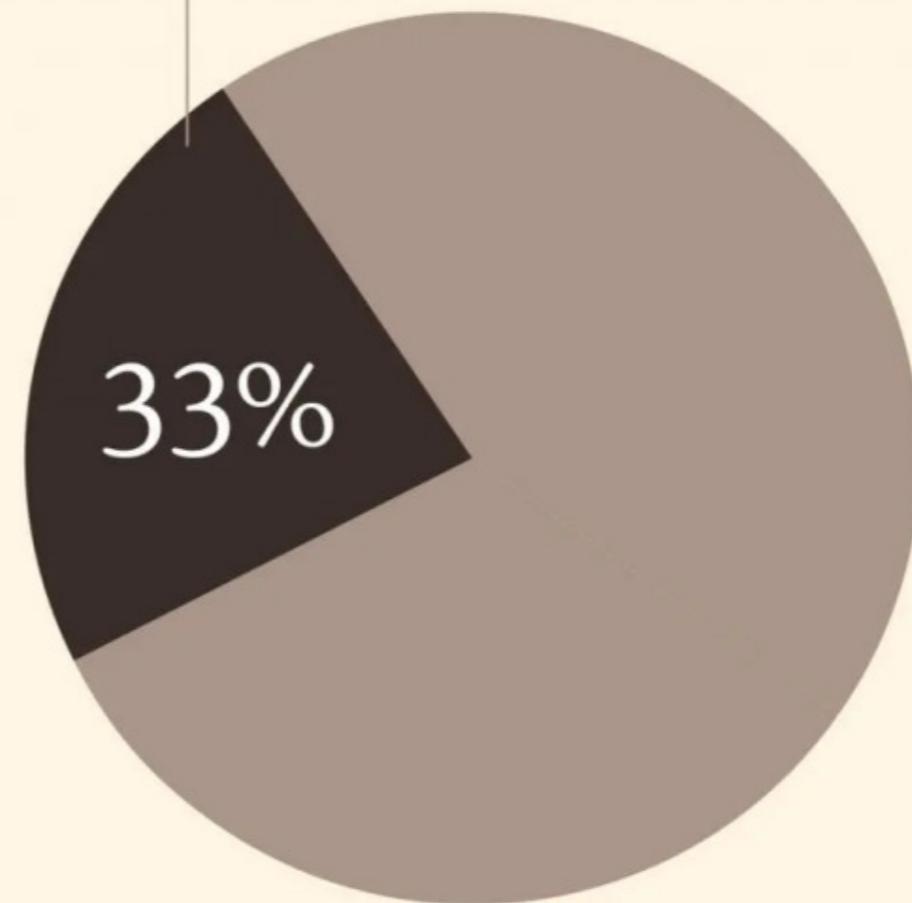
*A picture is worth  
a thousand words.*

一张图片胜过千言万语

Did not see a doctor when they needed to because of fear of being mistreated as a transgender person.



Reported having at least one negative experience related to being transgender with a health care provider.



National Center for Transgender Equality. US Transgender Survey, 2015. Available at: <http://www.ustranssurvey.org>

# VOTELINE

## SATURDAY'S RESULT

Can Julia Gillard win  
the next federal election?

## TODAY'S QUESTION

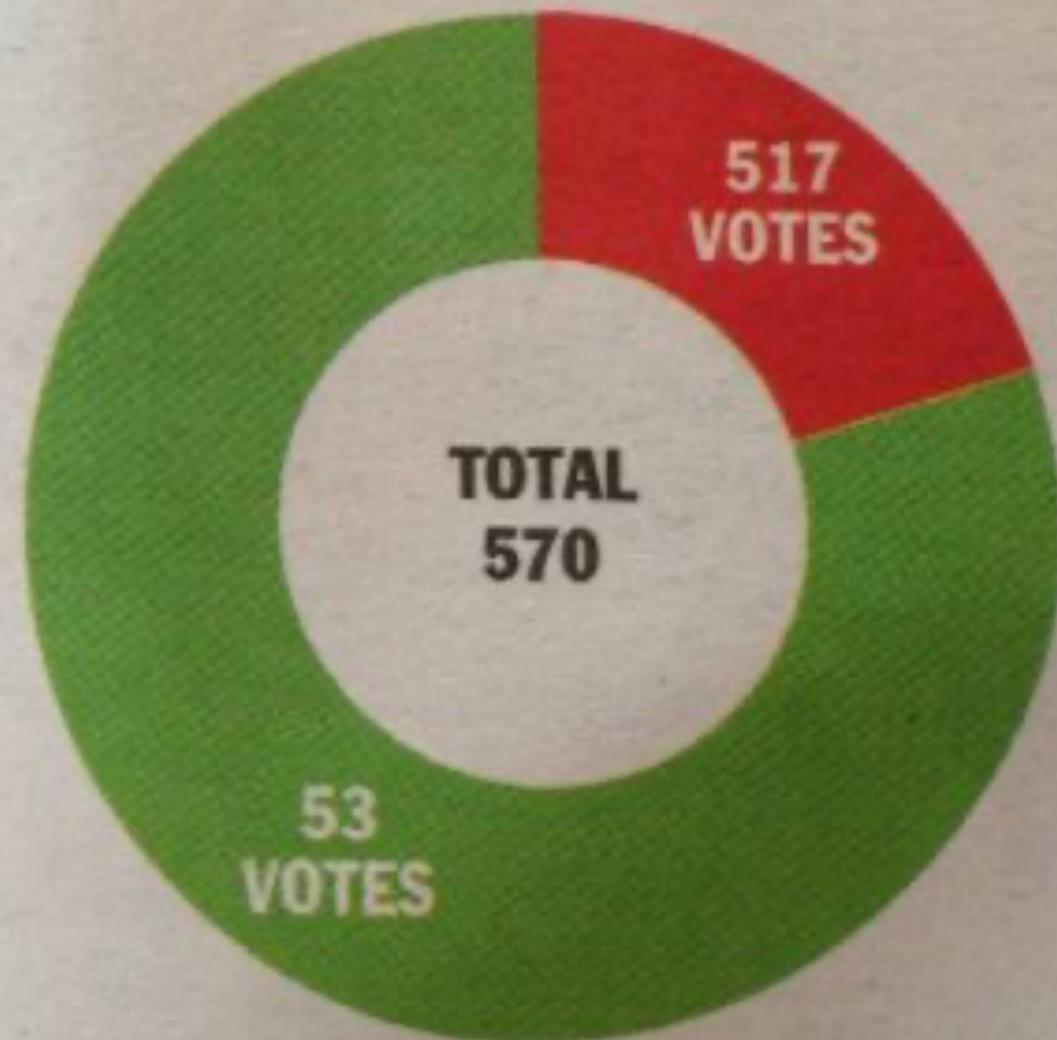
Do you like the plan to remove  
a car lane of the Princes Bridge  
for sole use by cyclists?

<b>YES</b>	<b>1900 956 434</b>
<b>NO</b>	<b>1900 956 435</b>

Calls cost 38.5c including GST. You  
can also have your say at  
[heraldsun.com.au](http://heraldsun.com.au) or [facebook.com/heraldsun](https://facebook.com/heraldsun)

**10%**  
**YES**

**90%**  
**NO**



in private residences. Men accounted for 79 per cent of these deaths.

Kat Wahamaa, regional

wouldn't even know drug users."

She said the problem is not that people are taking too

people working long hours.

As of May 31, there have been 851 deaths in B.C.,

Waha  
prob  
will."

"If  
outb  
publ  
respo  
publ  
craft  
exact  
now.

She  
have  
deca  
tion  
serv  
prof  
safe  
drug  
nali  
into

W  
ina  
200  
exa  
dic  
Wa  
do  
the

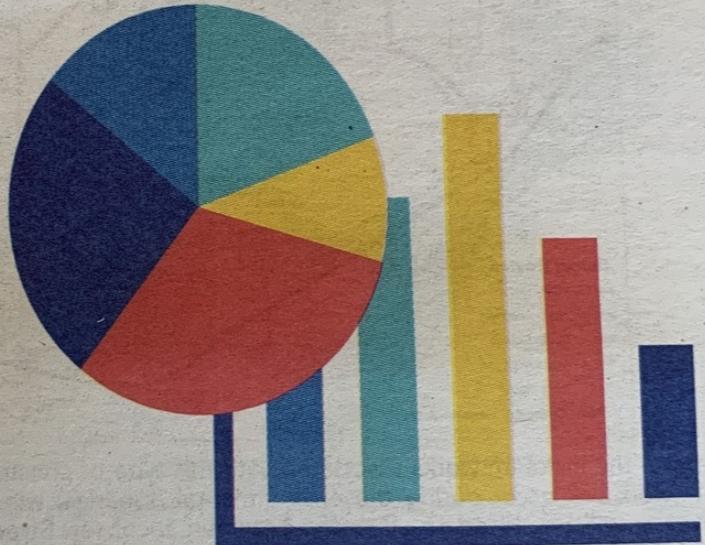
A  
sio  
ing  
ve  
su  
sp  
gr  
w  
th  
as  
C  
ac  
sa  
ad  
ti  
w  
p  
c

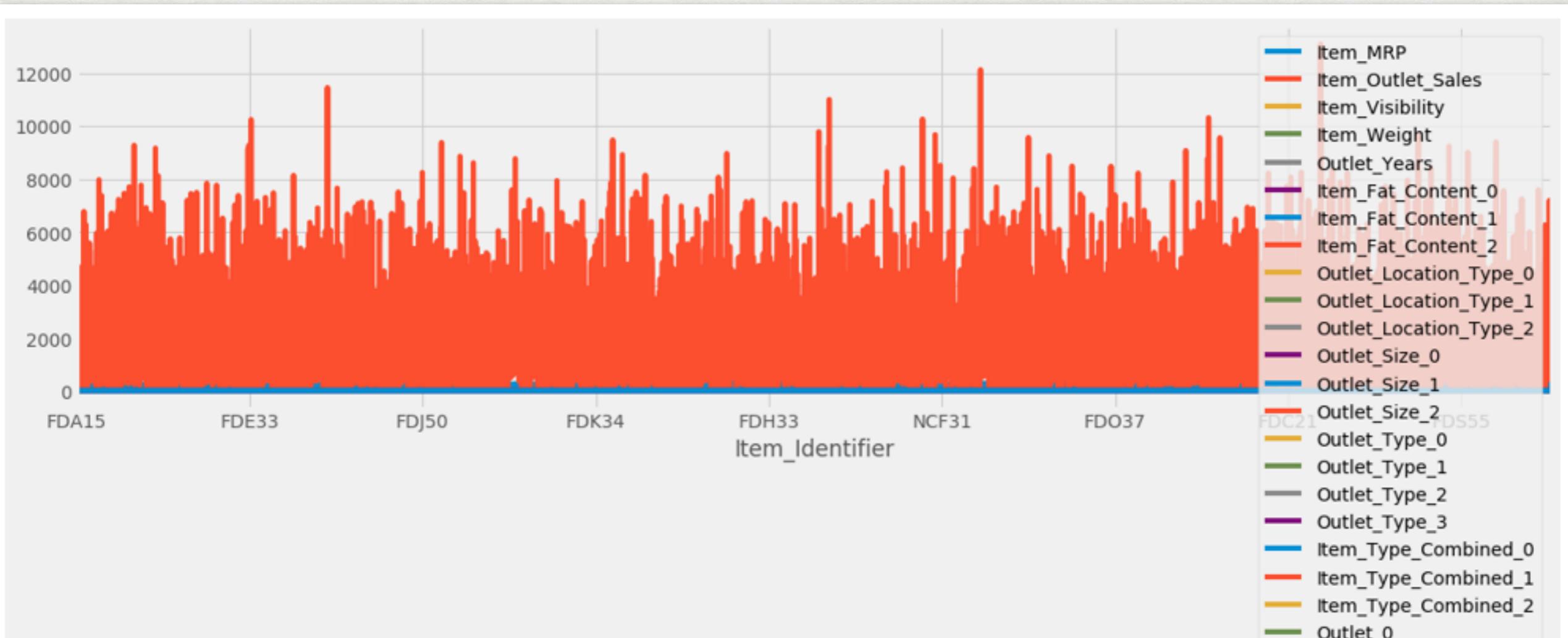
## OVERDOSE EPIDEMIC BY THE NUMBERS:

- Overdose deaths in B.C. in 2020: 1,726  
Deaths in Fraser Health in 2020: 574
- Overdose deaths in B.C. in first 5 months of 2021: 851  
Deaths in Fraser Health in first 5 months: 291

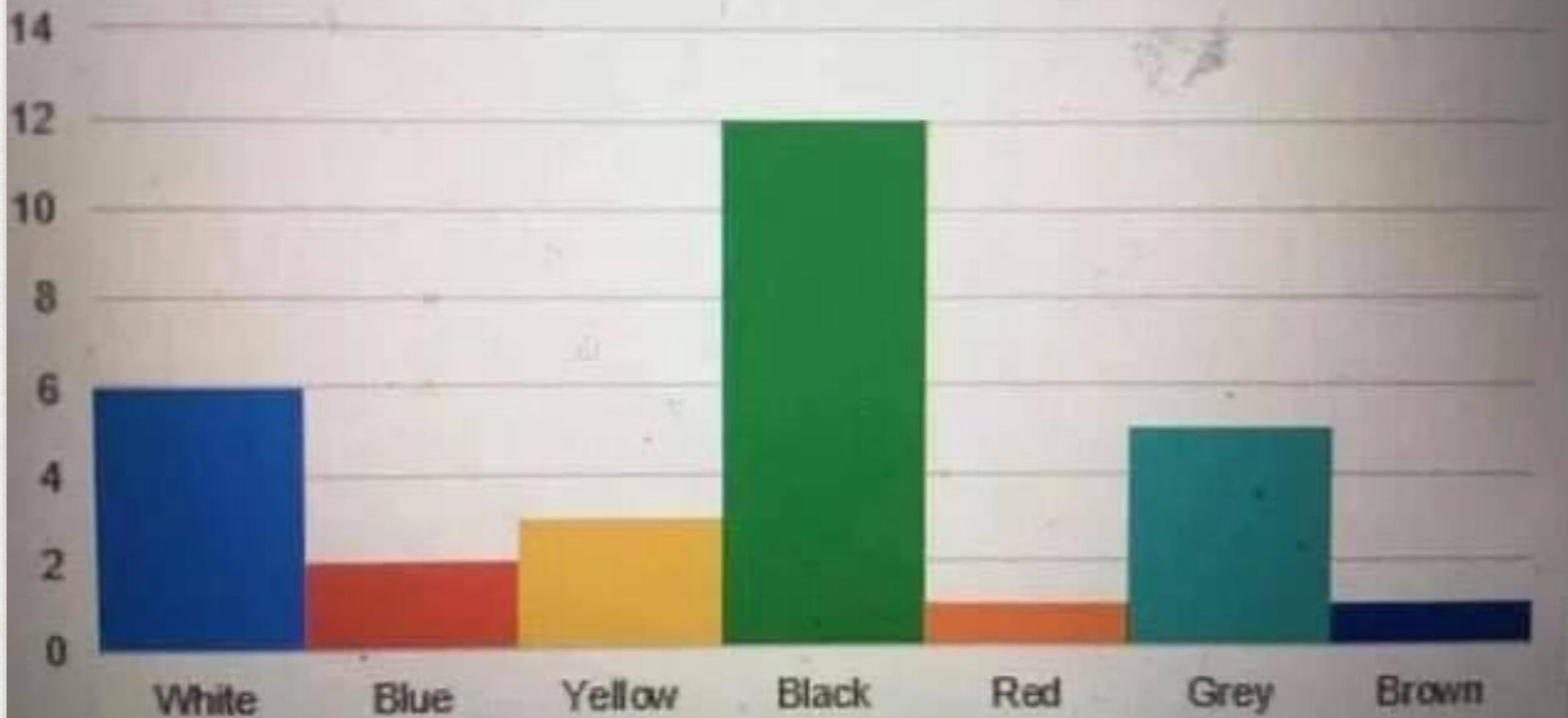
### Provincial data from Jan. 1 to May 31, 2021:

- In May 2021, there were 160 suspected illicit drug toxicity deaths – equates to about 5.2 deaths per day. This is the second largest number of suspected deaths ever recorded in the month of May (May 2020 had 177 deaths).
- In 2021, 70 per cent of people dying have been aged 39 to 59. Males have accounted for 80 per cent of the deaths.
- In 2021, 85 per cent of deaths occurred inside, 56 per cent in private residences and 29 per cent in social and supportive housing, SROs, shelters, hotels. Only 14 per cent occurred outside.
- No deaths have been reported at supervised consumption or drug overdose prevention sites.





## Shoe color Frequency



**\$599**

**Lawn-Level**

**\$795**

**Bronze**

**\$900**

**Bevel**

**\$1,250**

**Slant**

**\$1,900**

**Upright**

**\$500**

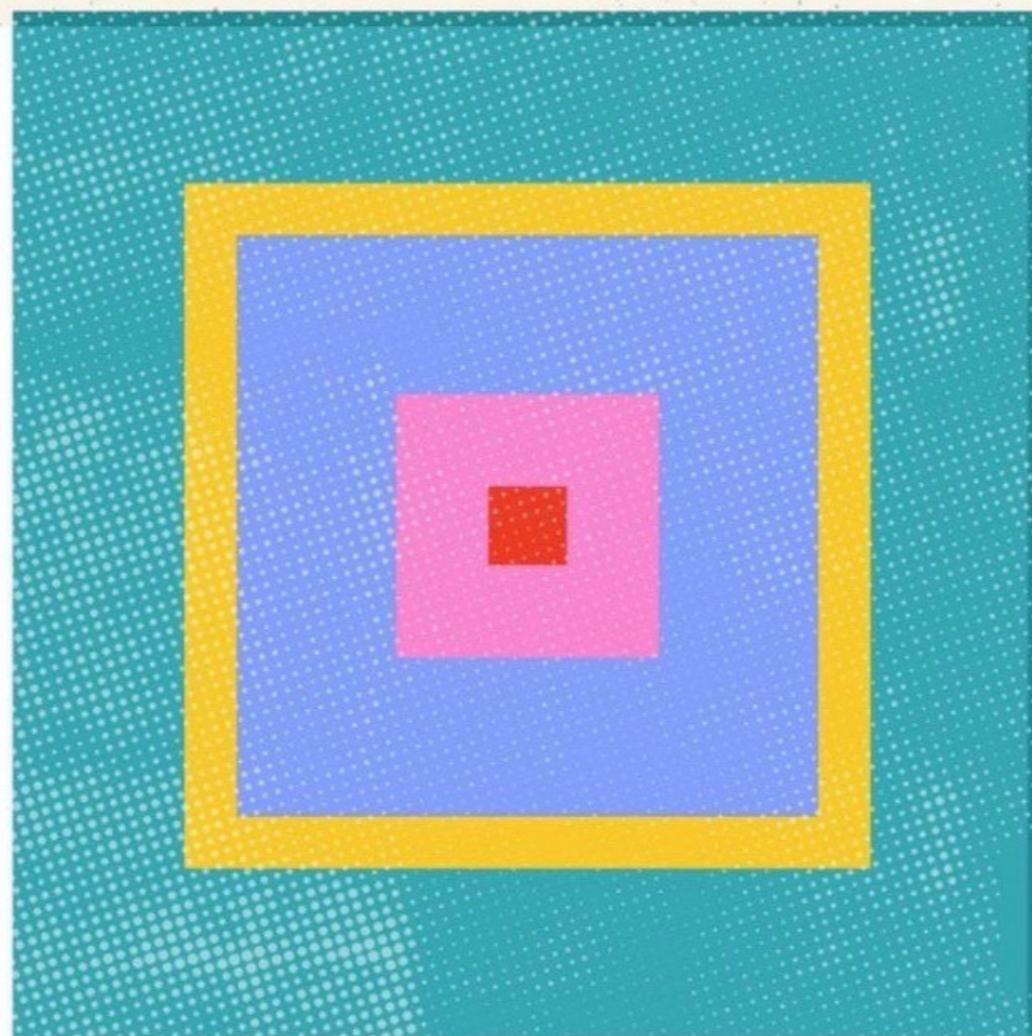
**\$1,000**

**\$1,500**

**\$2,000**

**\$2,500**

# HOW MUCH DO YOU SPEND ON GROCERIES EVERY WEEK?



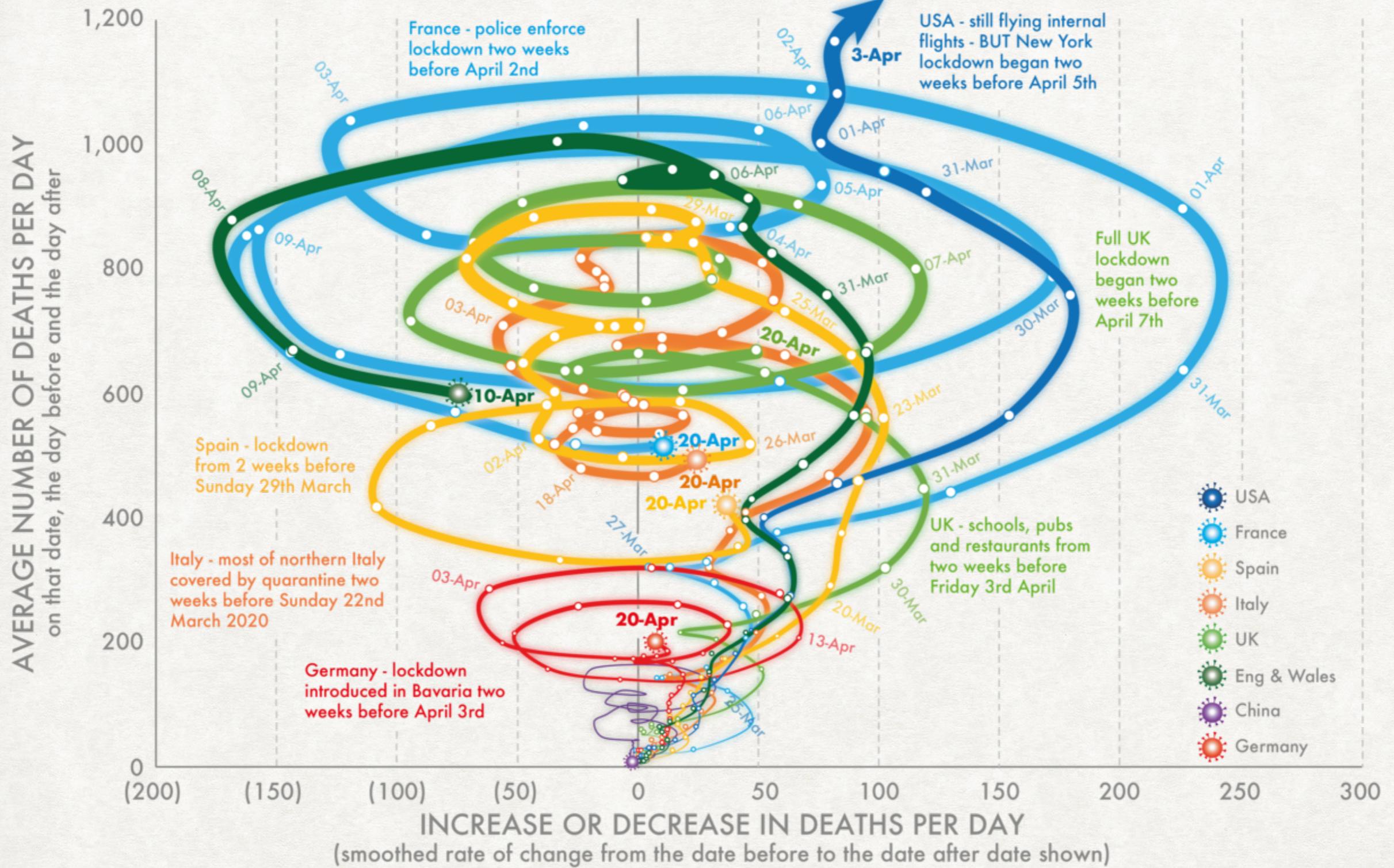
**22%** UNDER \$100

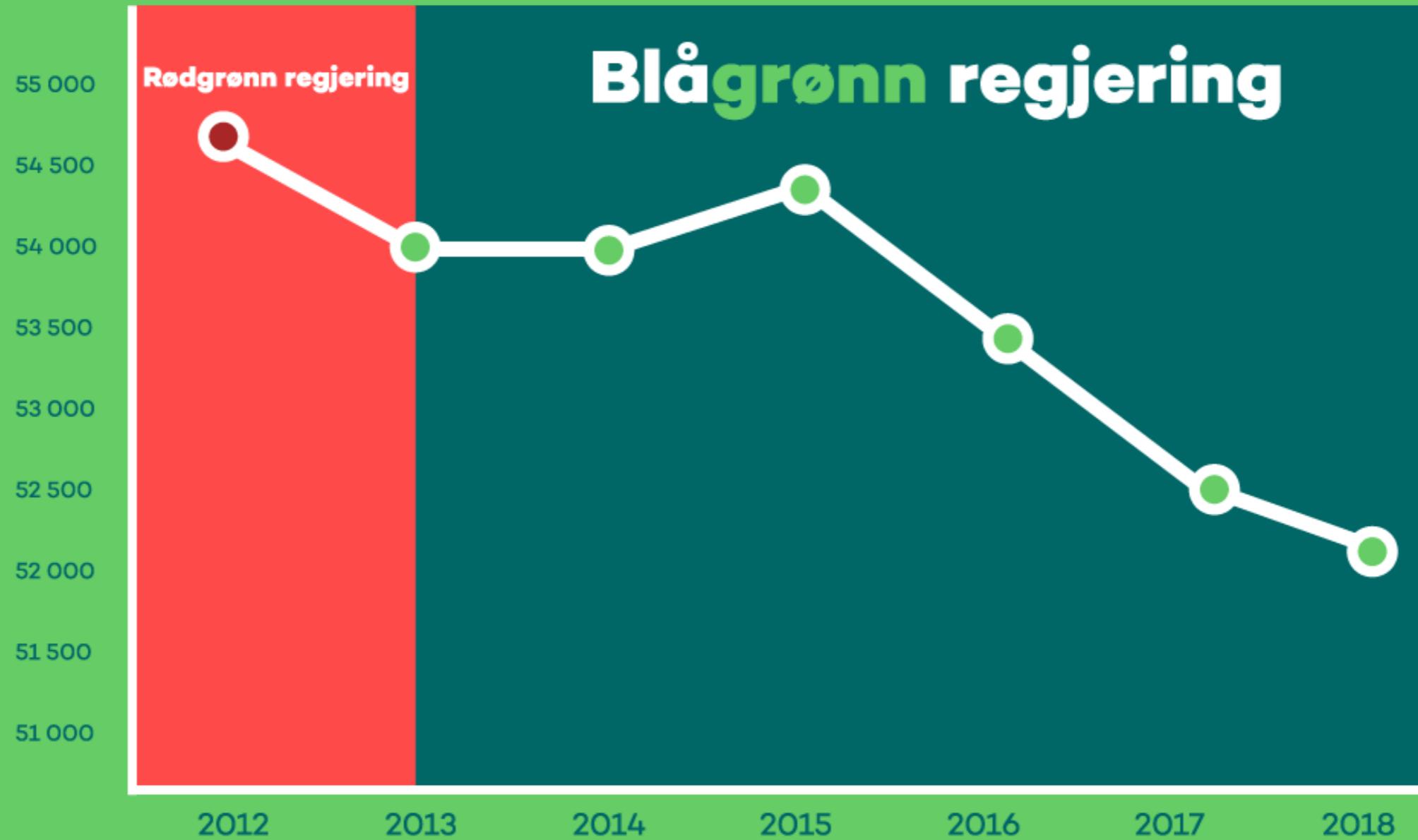
**26%** ABOUT \$100

**39%** \$100 TO \$200

**10%** \$200 TO \$300

**3%** MORE THAN  
\$300





$$(52200 - 54200) \div 54200 = -3.7\%$$

# Best in Show: The Ultimate Data Dog



Inexplicably Overrated



Bulldog



Hot Dogs!



our data score



intelligence



costs



longevity



grooming



ailments



appetite

The Rightly Ignored

Overlooked Treasures

# 应该怎样将数据可视化

- 好的可视化
  - 能够轻松读取图中的数据
  - 忠实于数据，并且能体现数据背后的故事
  - 富有表现力，能给人留下深刻印象
- 差的可视化
  - 含有过多或过少的信息
  - 存在自相矛盾的地方
  - 超出人类的认知水平
  - 用错误的方式展示了数据
  - 使用了错误的数据

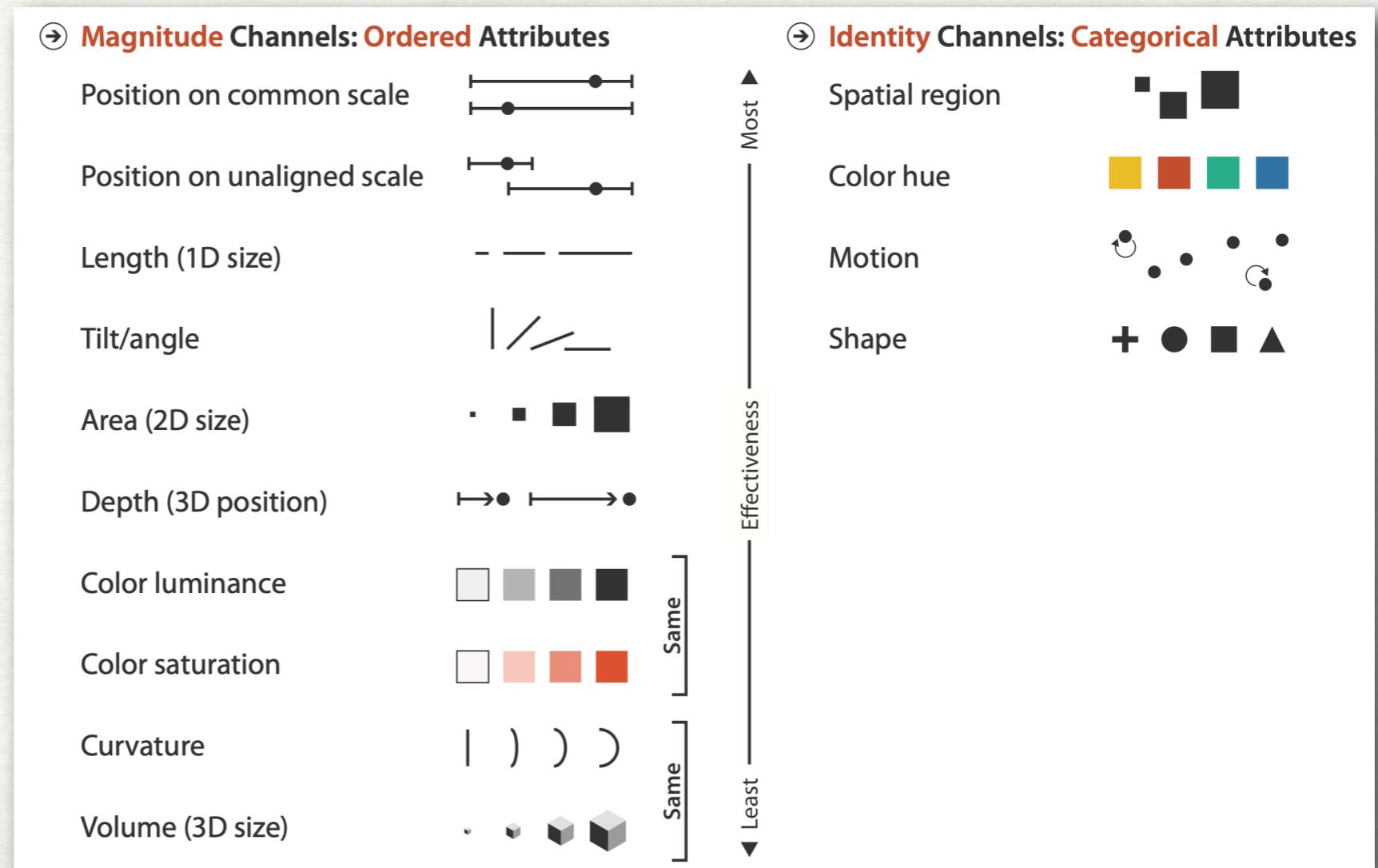
# 可视化之前应该思考的问题

- 目的：我为什么要做这幅图？
  - 受众：我的图是给谁看的？
  - 媒介：我将怎样利用或分享这幅图？
  - 工具：我怎么制作这幅图？
  - 寓意：我要通过这幅图讲什么故事？
  - 影响：这幅图对那些人有用，可能忽视哪些人？
- 
- 例1：“我为了完成一篇课程论文，需要展示一组数据的分析结果。论文的读者是老师（具有较高学术素养和学术储备的专业人士）。我并不擅长数据可视化，因此选择 Excel 进行绘图。”
  - 例2：“我在一家咨询公司实习，正在负责协助撰写行业分析报告。报告的读者主要是同行业的从业人员。公司有专业可视化软件可以进行制图。”

# 图的基本组成部分

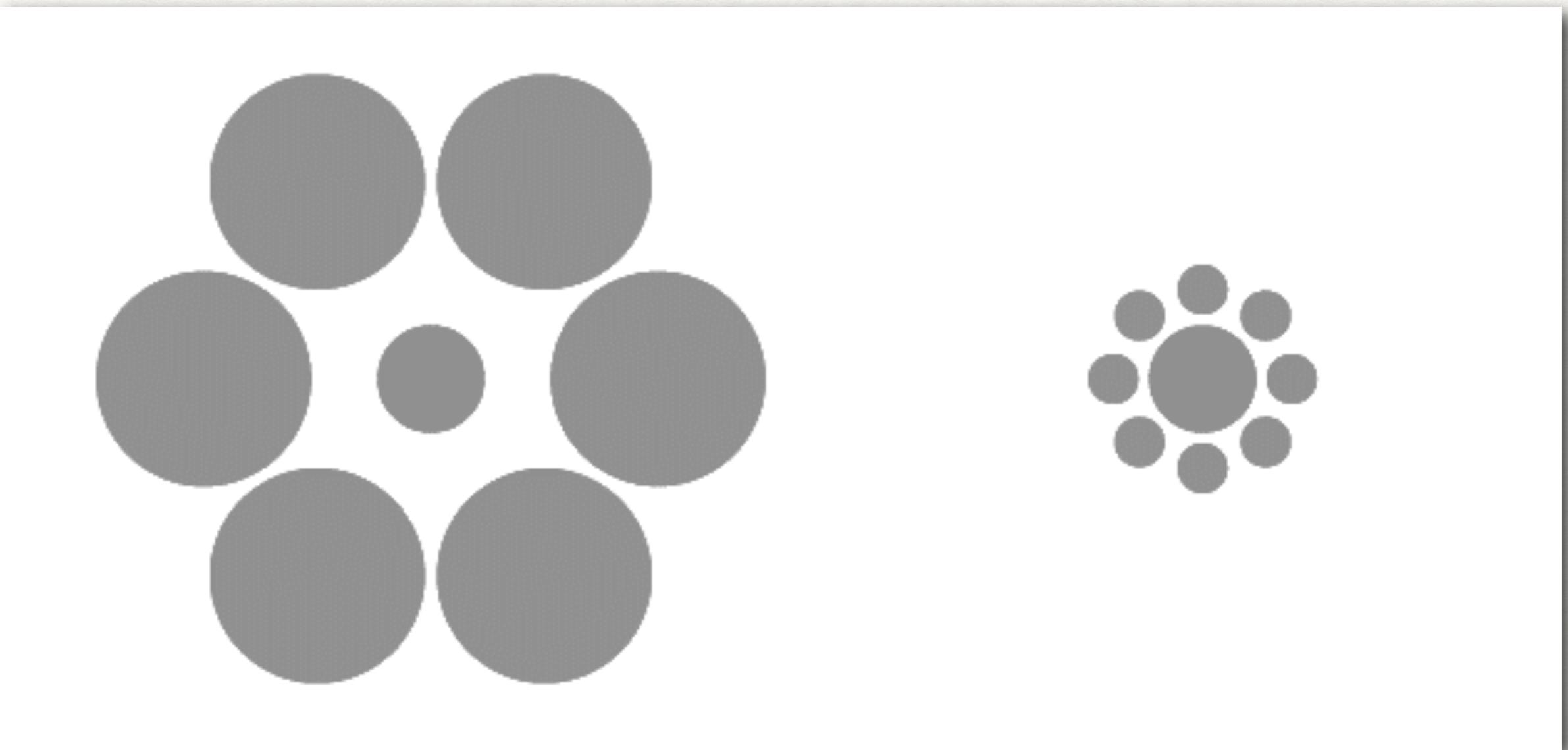
绘图的目的是让人的眼睛和大脑能够认知并理解图中的信息

- 符号 (mark) 是基本的图形表达：点 (0D)、线 (1D)、区域 (2D) 等
- 通过选择不同的表达方式 (channel, 或称视觉通道) 决定如何描绘符号，使图能够展示数据的含义。



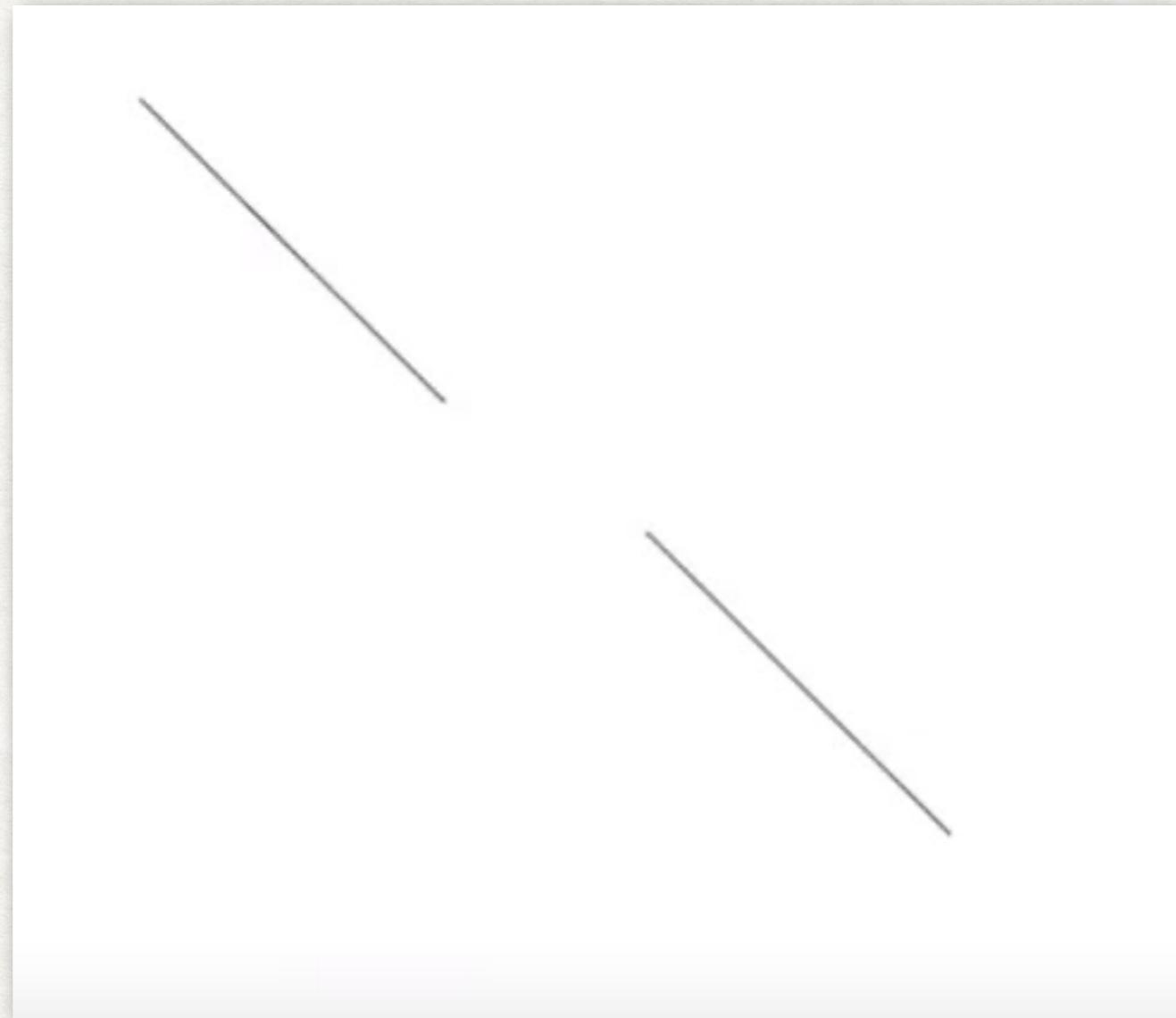
# 人类的视觉认知偏差

- 位于中间的圆哪个大？



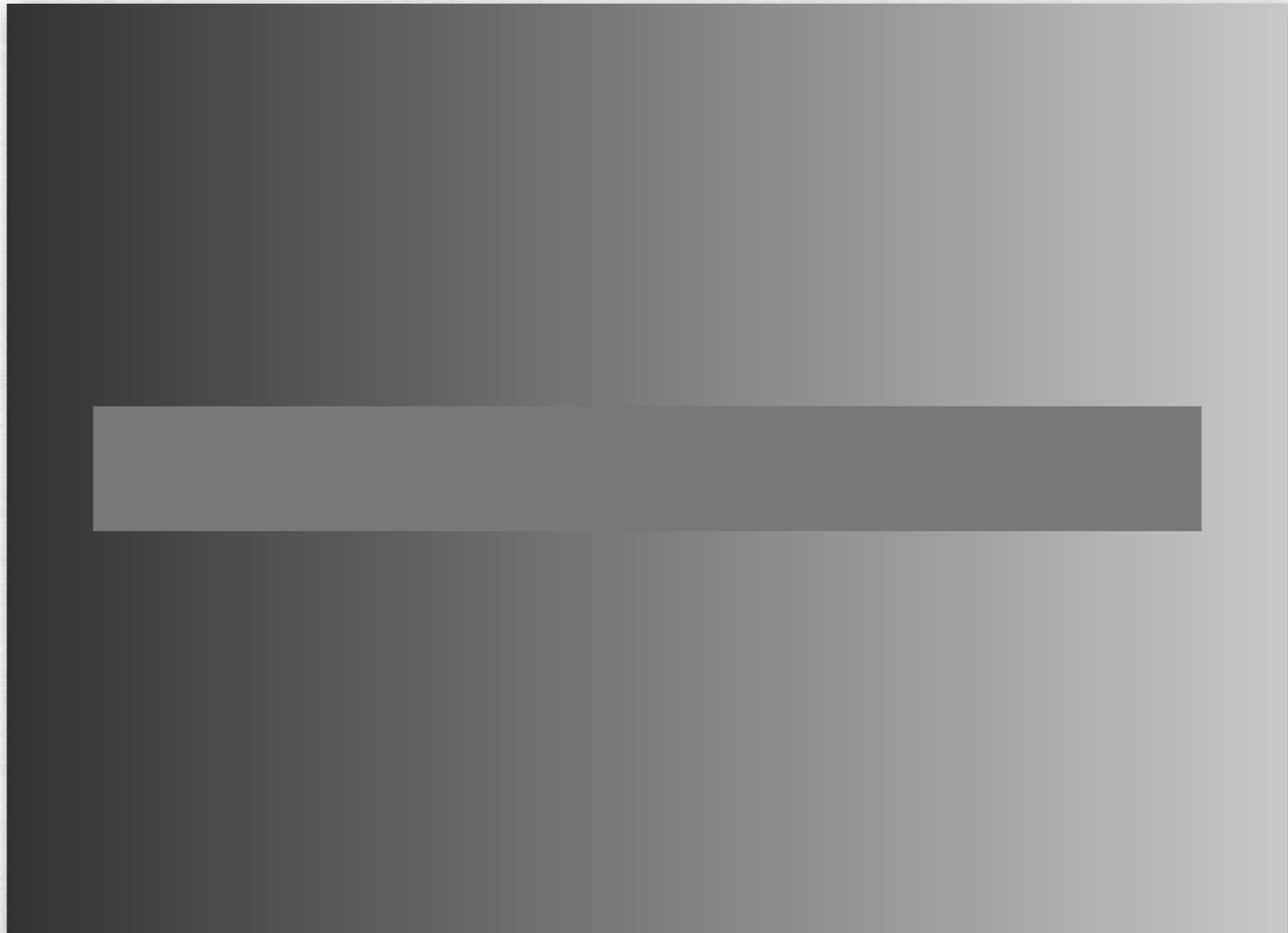
# 人类的视觉认知偏差

- 两条线段在同一条直线上吗？



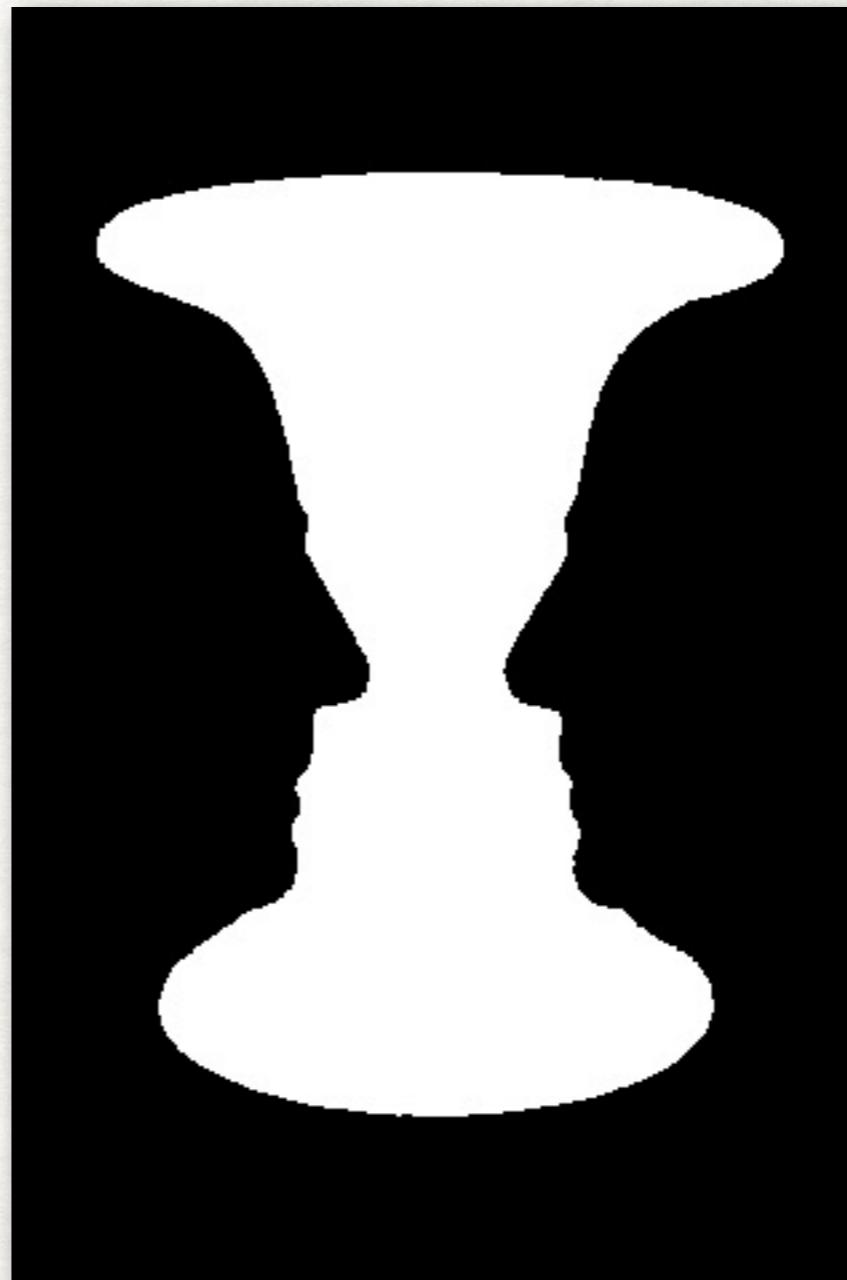
# 人类的视觉认知偏差

- 中间的条形部分的颜色是渐变的吗？



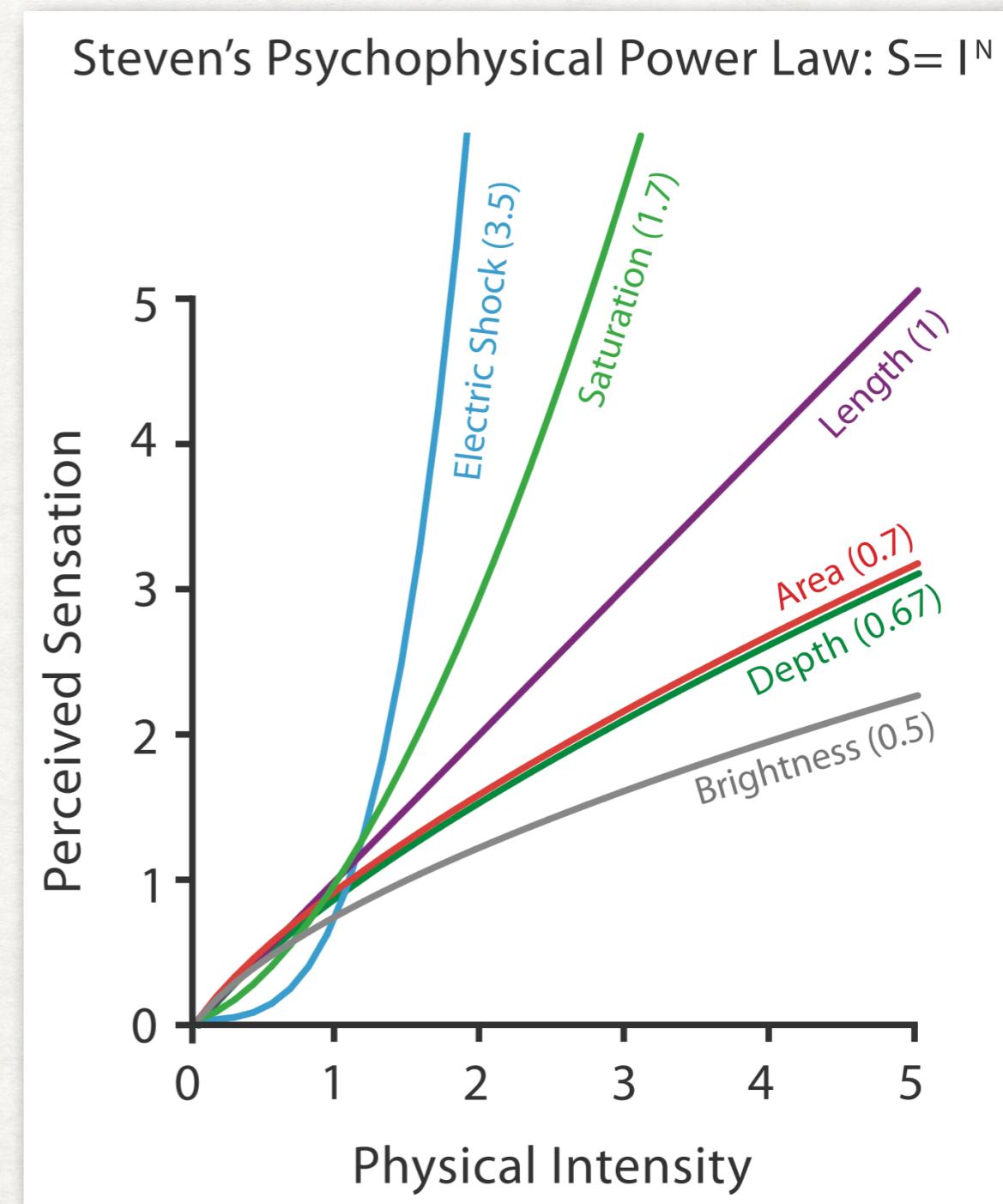
# 人类的视觉认知偏差

- 你看到的是什么？



# 人类的视觉认知偏差

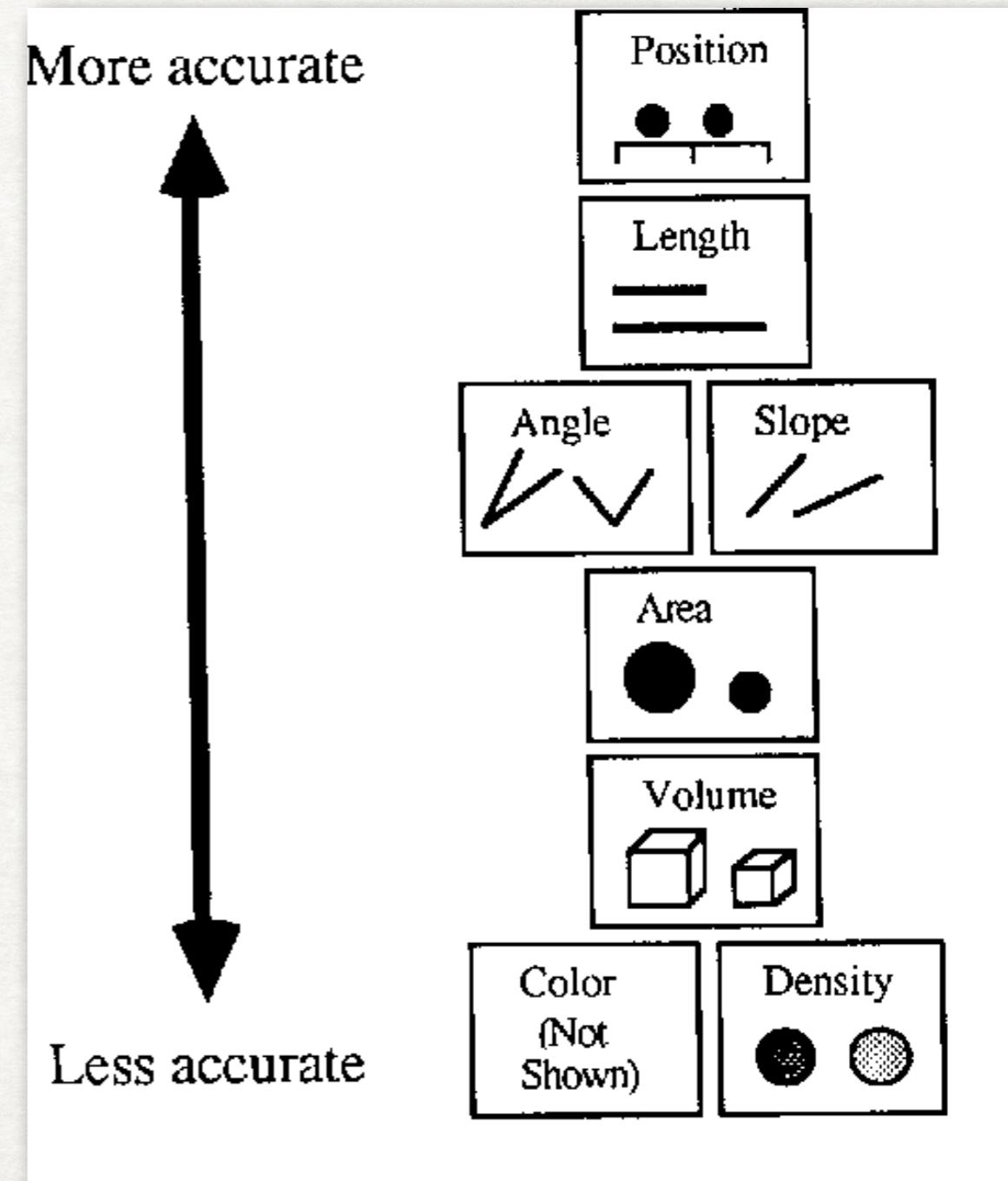
- 对于不同类型的刺激，人类的感知程度是不一样的
- 对电流刺激，我们通常感觉到的强弱变化会大于真实的强弱变化
- 对光的刺激，我们感觉到的强弱变化则会小于真实的强弱变化
- 对于长度点变化，我们的感知和真实变化基本一致



# 选择合适的表达方式

按照能否正确认知的程度排序

1. 具有相同刻度的位置
2. 具有不同刻度的位置
3. 长度
4. 方向
5. 角度、倾斜度
6. 面积
7. 体积
8. 曲度
9. 密度、色彩饱和度、阴影
- 10.色调



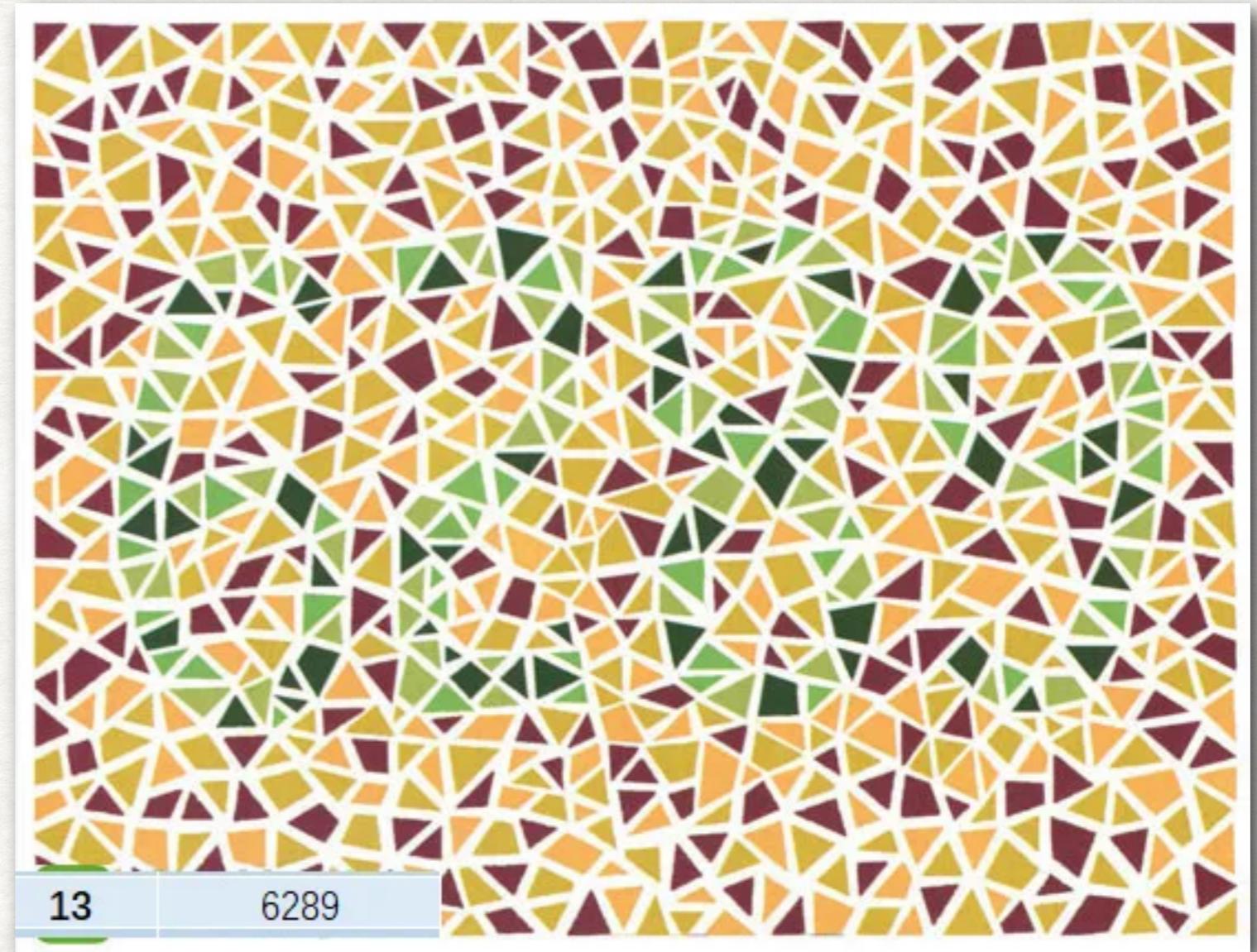
# 对可视化效果的评价

- 表现力原则：图应该完整地表达数据的所有属性
- 有效性原则：图的阅读者能够从图中读取正确的信息

因此，针对数值类变量（身高、体重、家庭成员数、好评度等），我们应当选取更容易区分变化程度的表达方式（如位置、长度等）；而针对分类变量（性别、民族、出生地等）则应用形状、颜色或纹理图案区分不同类别

# 关于颜色的选择

- 不同的人对颜色的感知是不一样的
- 你能看清右图中的数字码? (6289)

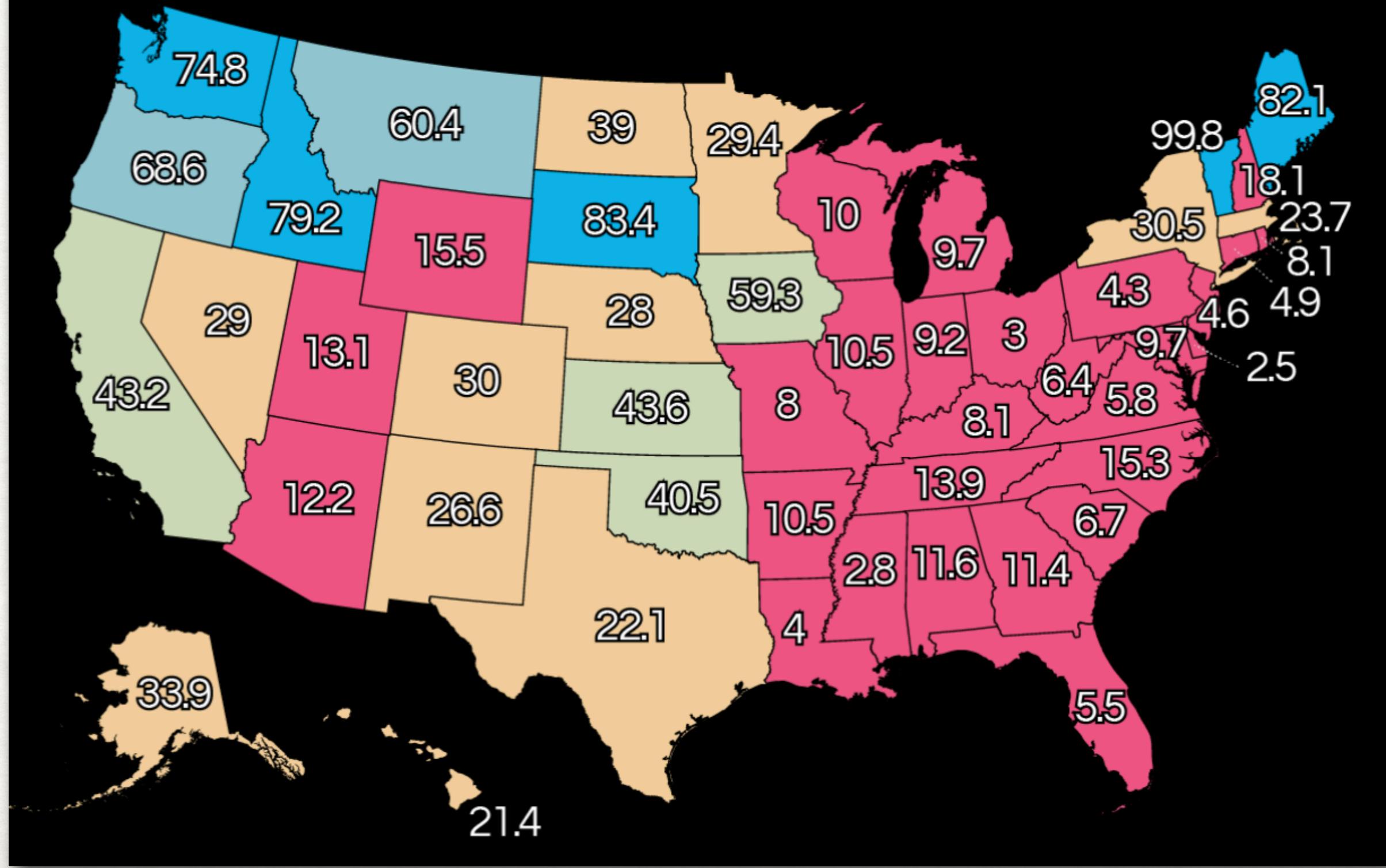


# 视锥细胞对比测试 (Cone Contrast Test)

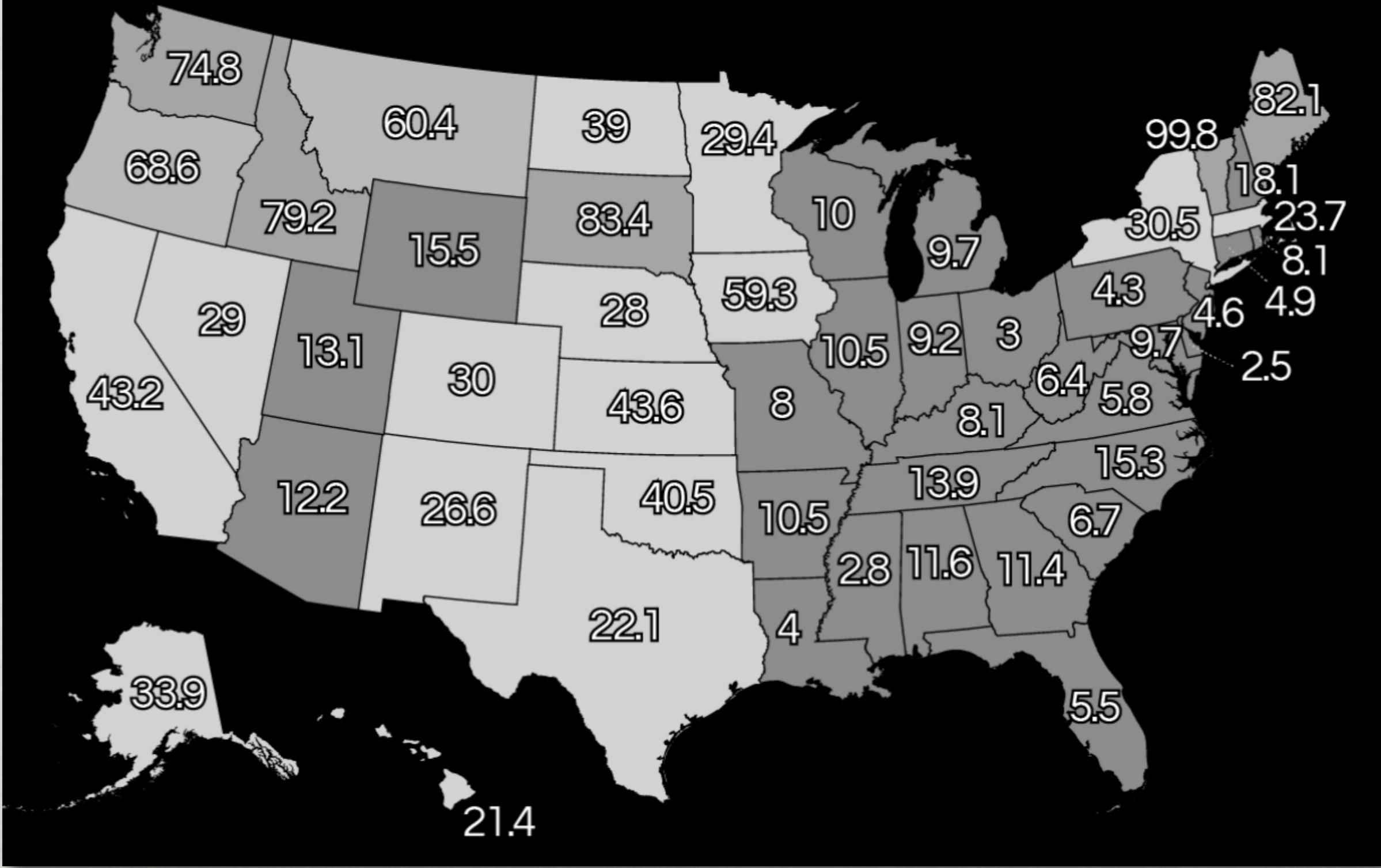
Rabin, Gooch, & Ivan. Rapid Quantification of Color Vision: The Cone Contrast Test.  
*Investigative Ophthalmology & Visual Science*. February 2011, Vol.52, 816-820.

Score	Cone Contrast Test			Cone Contrast (%)		
	L Cone	M Cone	S Cone	L, M	S	
10	V Z	N F	E Z	27.5	173	
20	F V	Z U	N R	19.1	120	
30	R P	E P	F D	13.2	83	
40	Z E	N F	Z V	9.1	57	
50	H R	E D	R P	6.3	39	
60	D R	H P	Z N	4.4	27	
70	N Z	D U	E D	3.0	19	
80	H F	F H	V R	2.1	13	
90	G T	H T	F D	1.4	10	
100			H P	1.0	7	

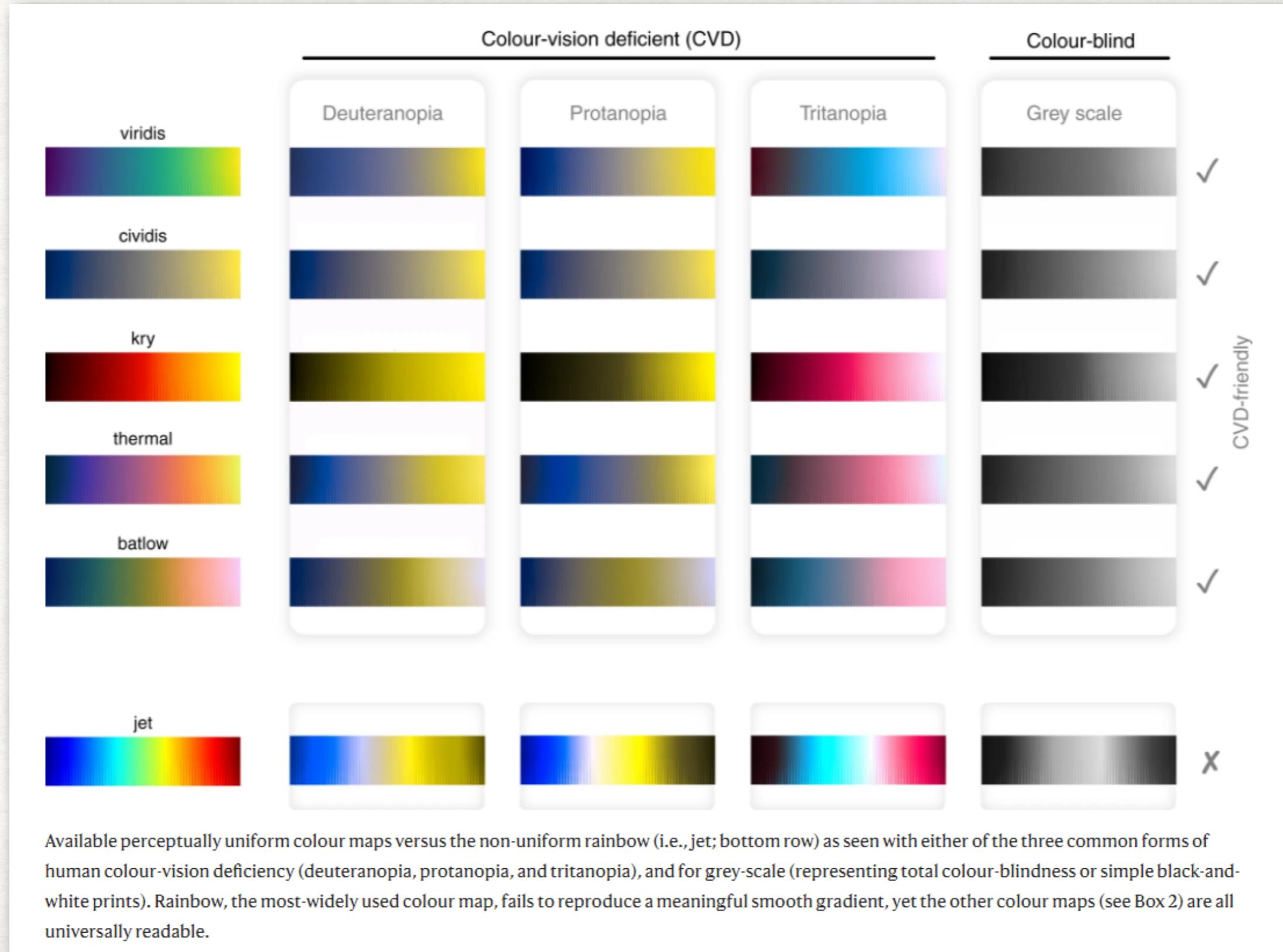
# % Renewable Electricity



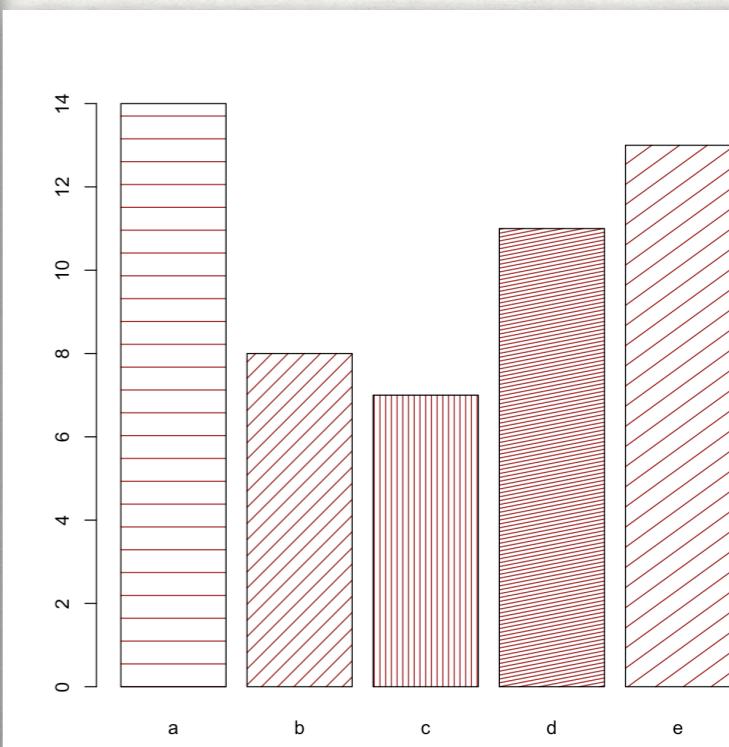
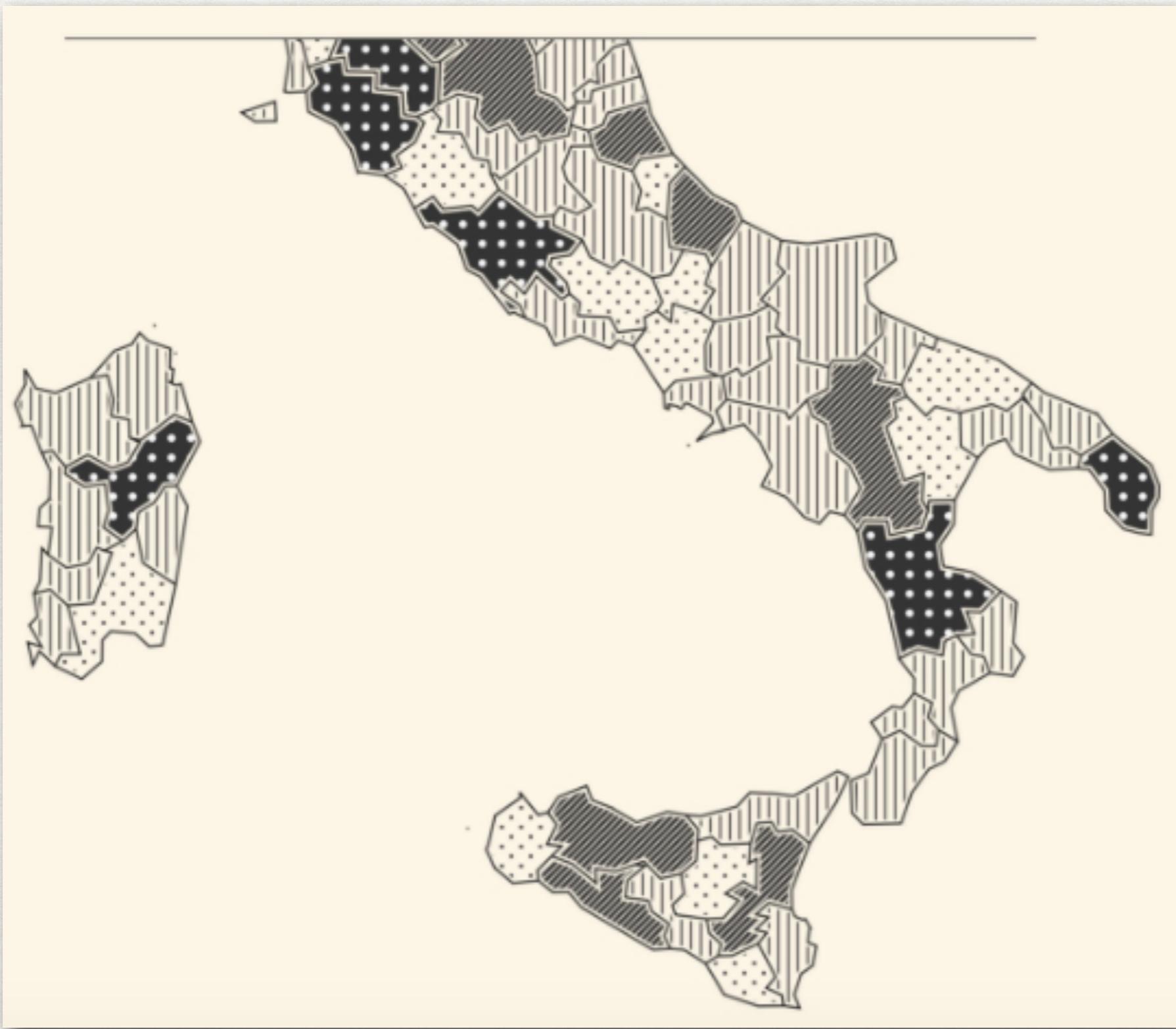
# % Renewable Electricity



# 避免使用容易产生问题的颜色组合



# 辅助使用纹理图案、符号等

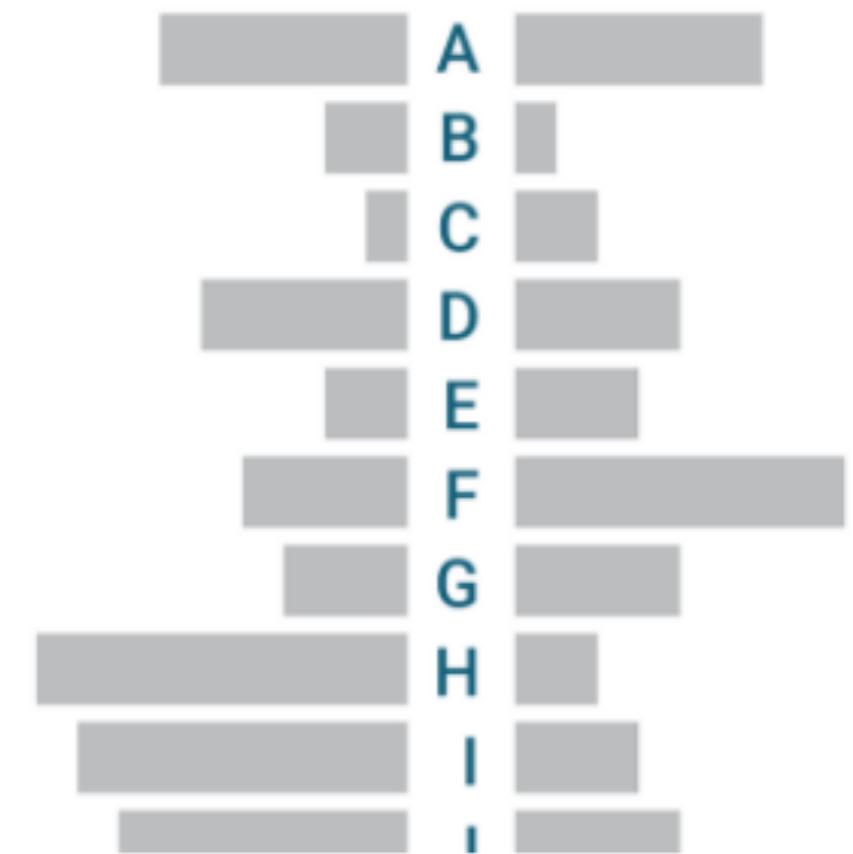


# 重新考慮圖的表达方式

NOT IDEAL



BETTER



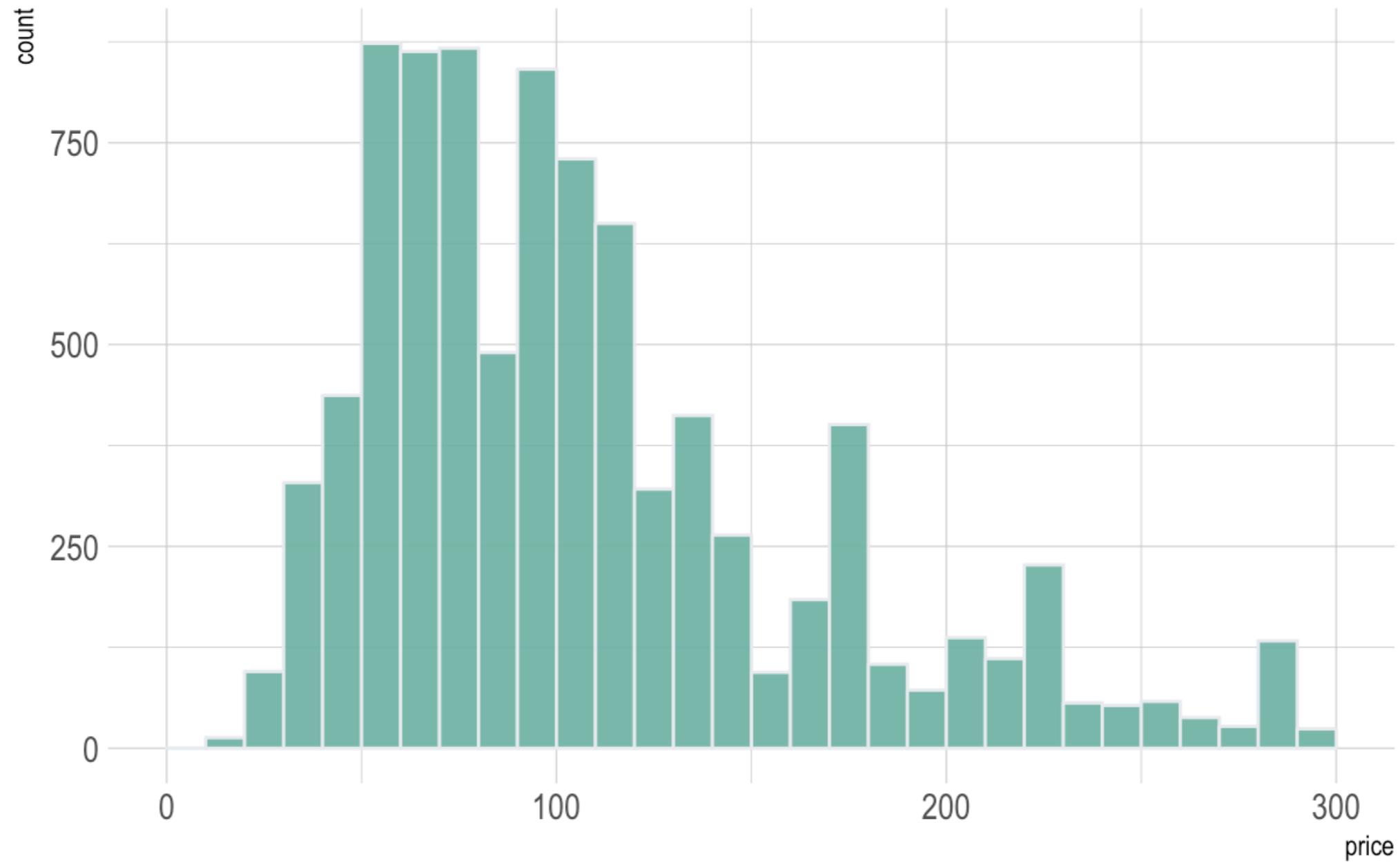
# 统计数据的可视化

- 区分数据的种类：定量（连续或离散数值）、定性（有序分类、无序分类）
- 决定你想展示的数据特征：数值间的比较？数值的变化趋势？数值的分布特征？
- 选择合适的图：
  - 一个定量变量：直方图、箱线图
  - 两个定量变量：散点图、折线图
  - 一个定量变量 + 一个定性变量：柱形图、箱线图、饼图（不建议使用）
  - 包含更复杂的信息：气泡图、网络图等

<https://www.data-to-viz.com/>

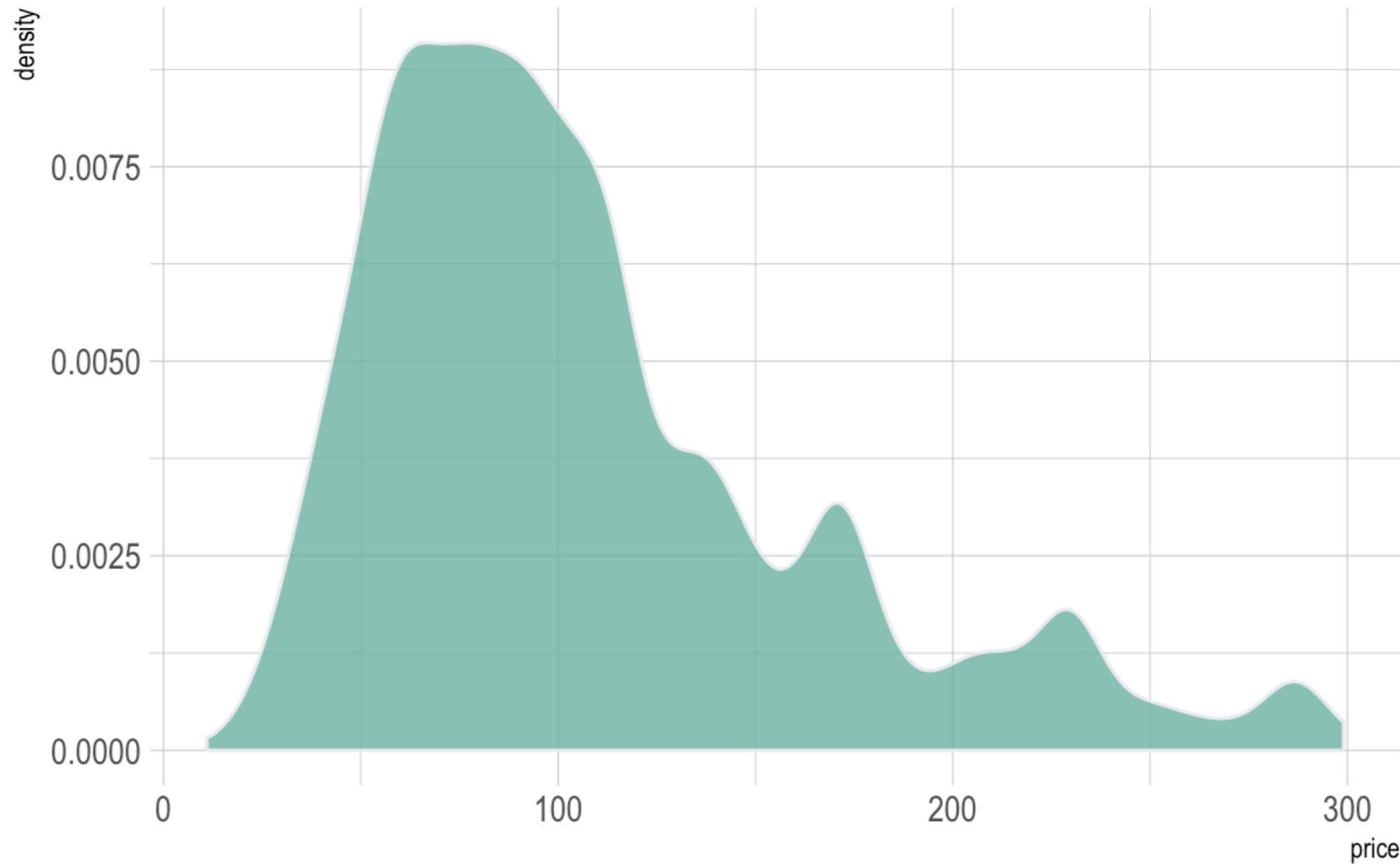
# 直方图 histogram

Night price distribution of Airbnb appartements

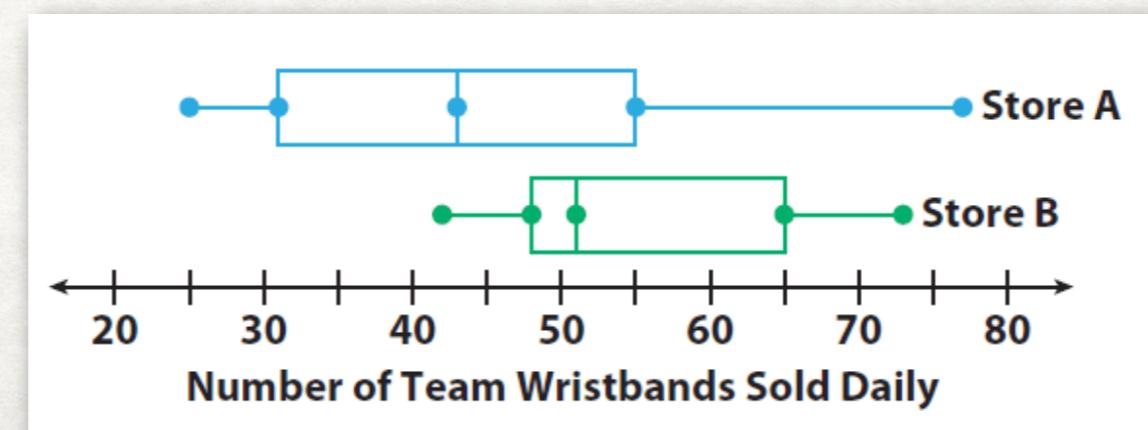
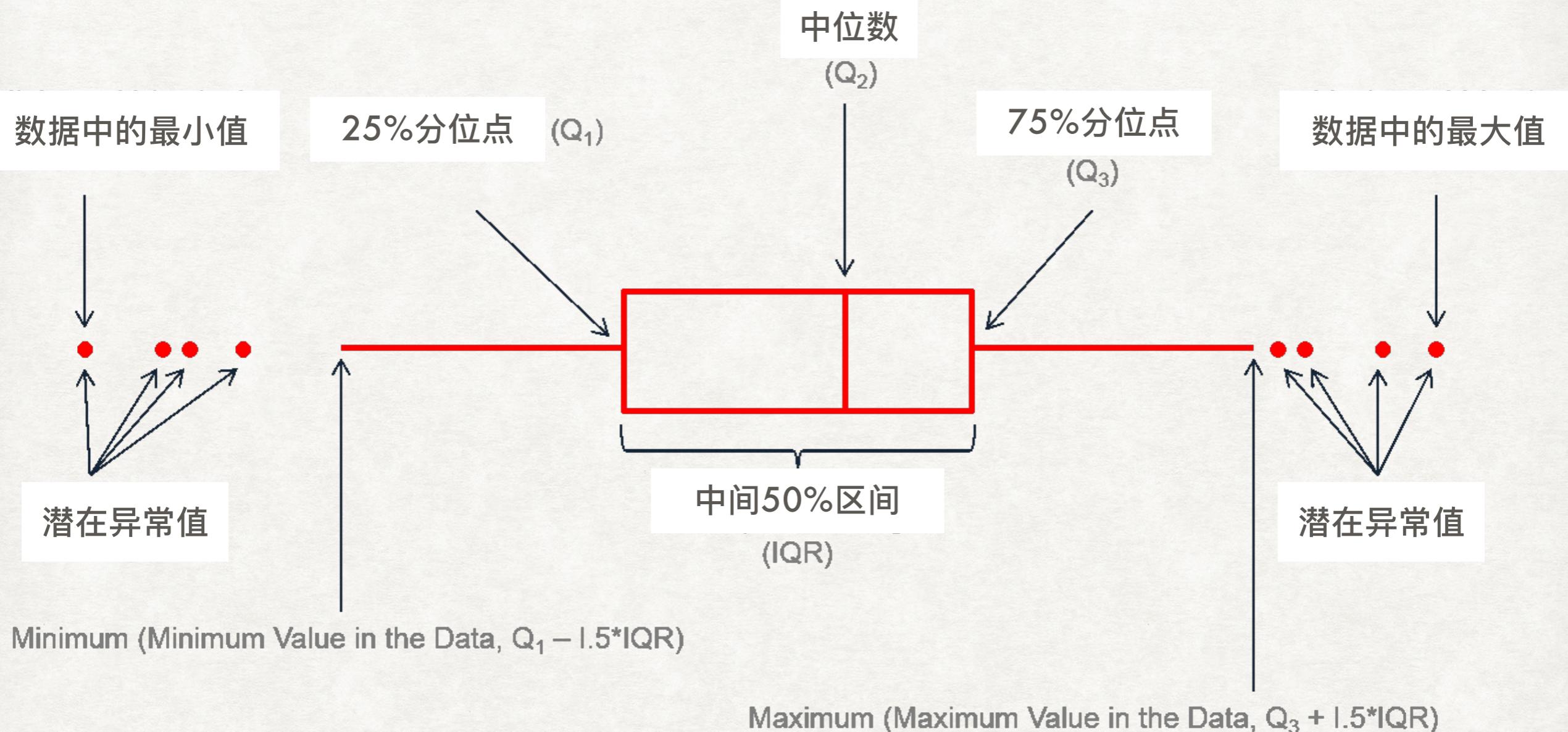


# 密度图 density plot

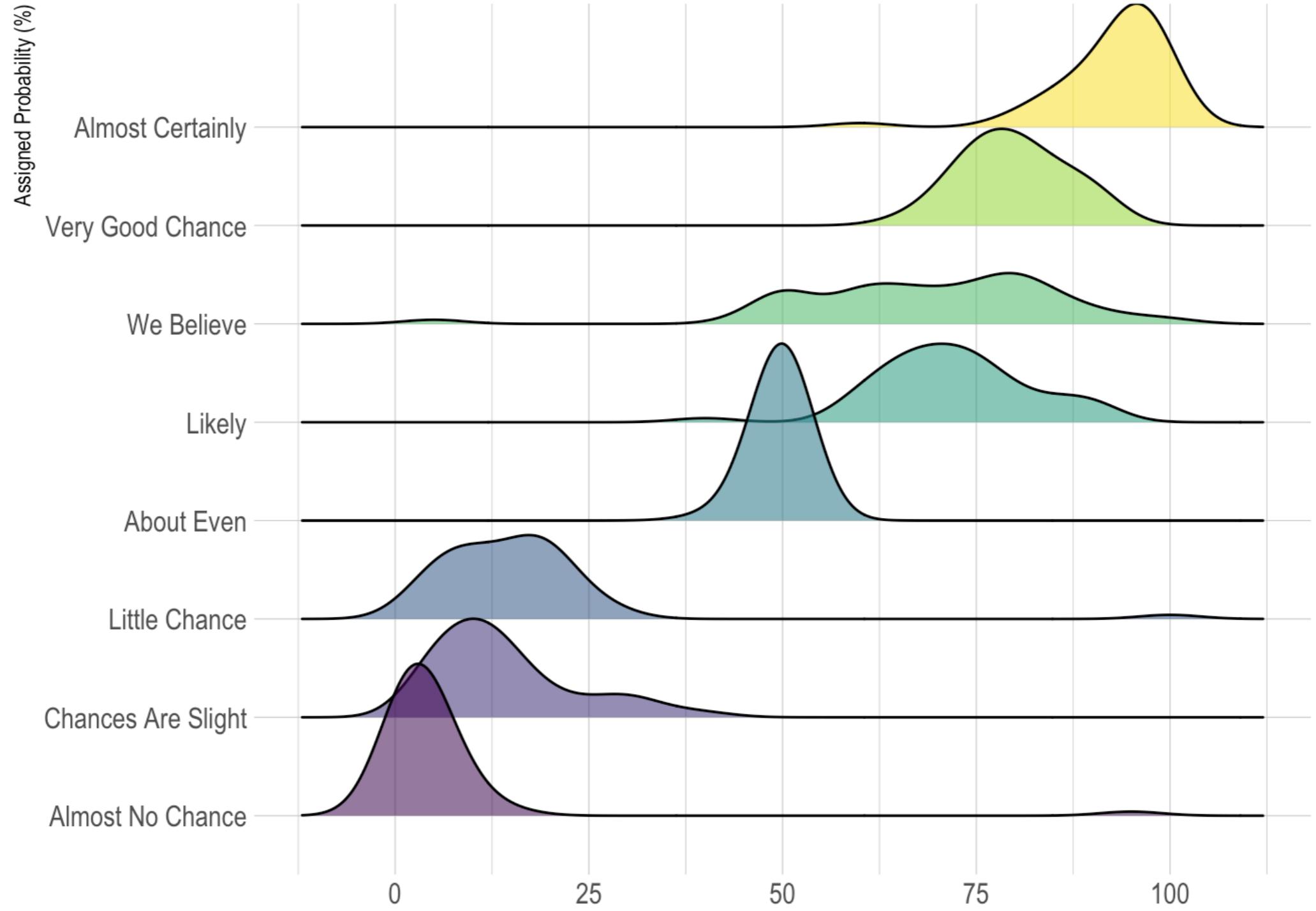
**Night price distribution of Airbnb appartements**



# 箱线图 boxplot



# 比较多个变量的分布

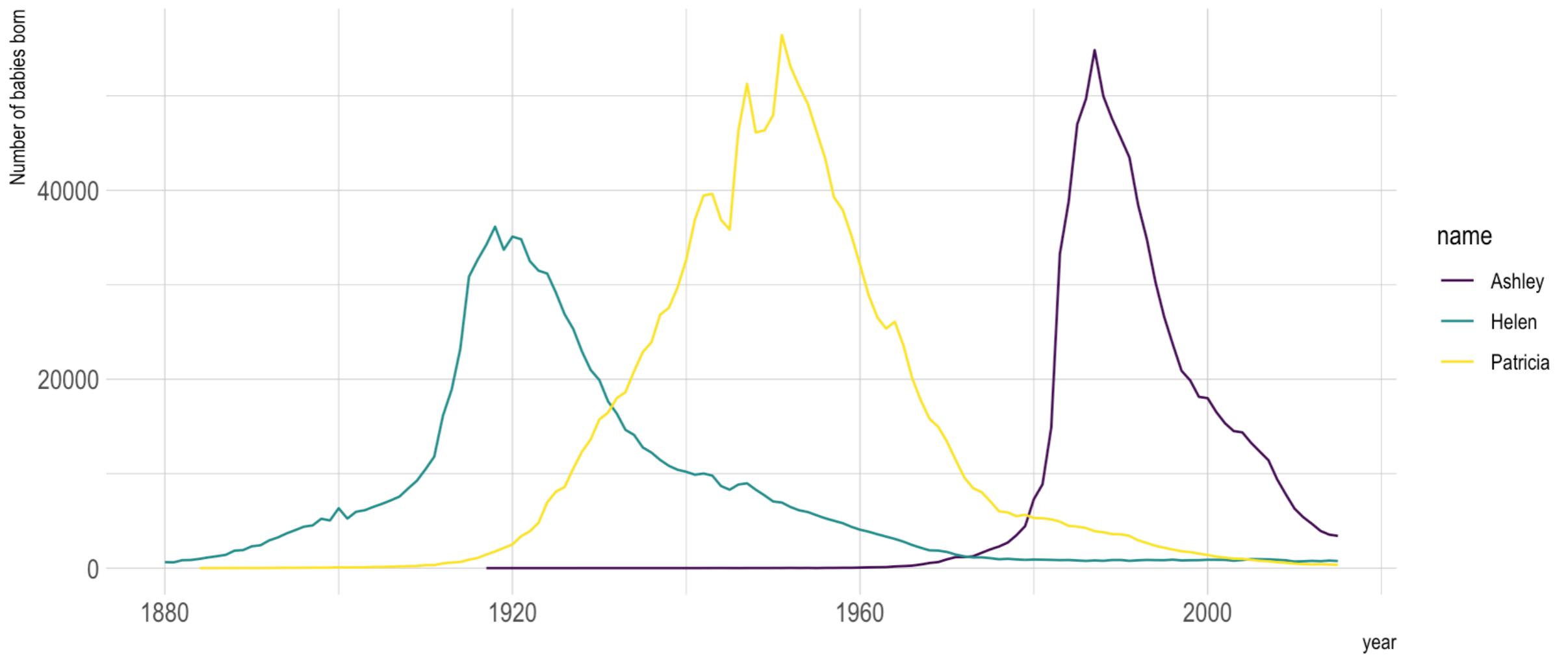


# 散点图 scatter plot

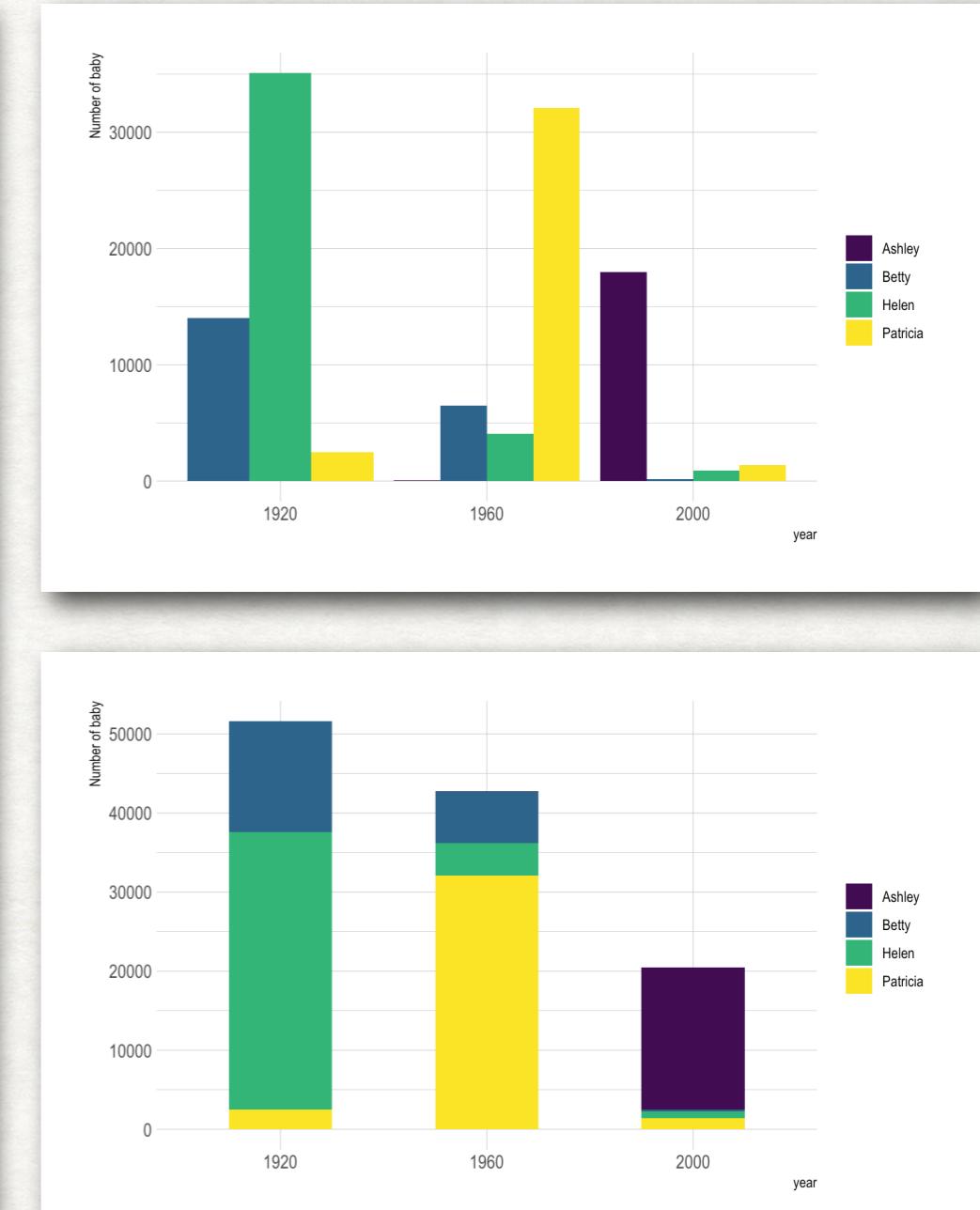
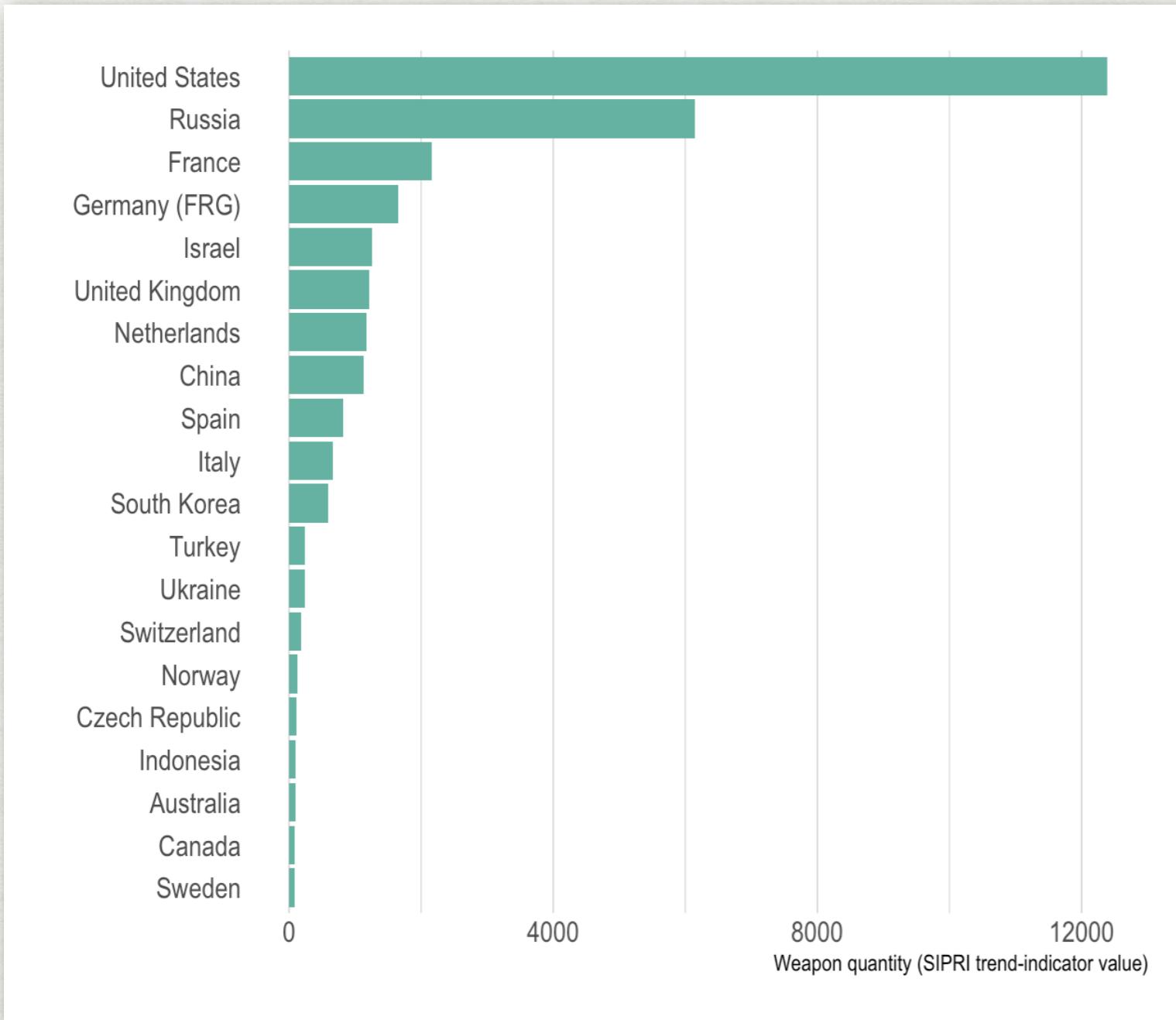


# 折线图 line chart

Popularity of American names in the previous 30 years

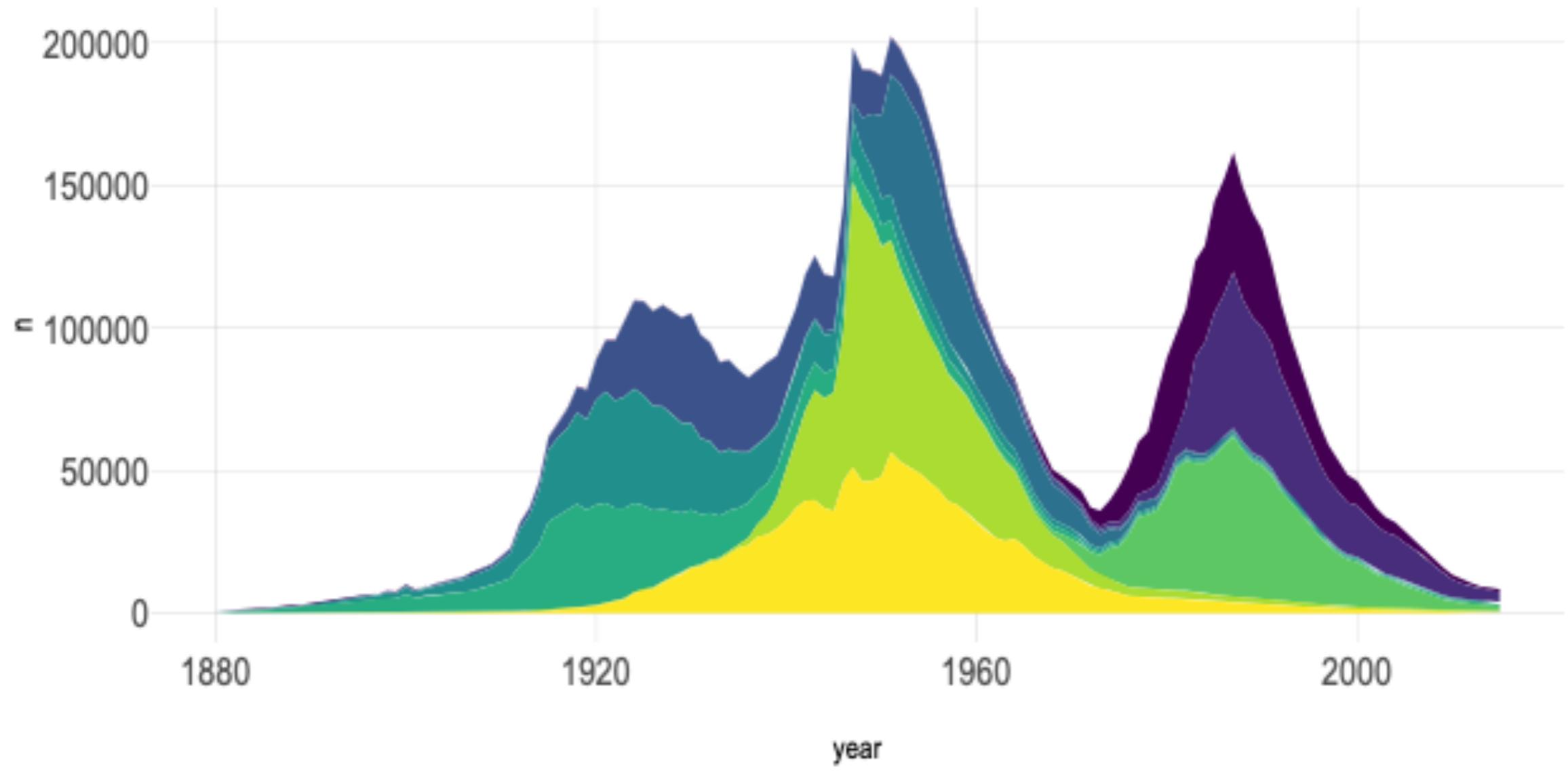


# 柱形图 barplot

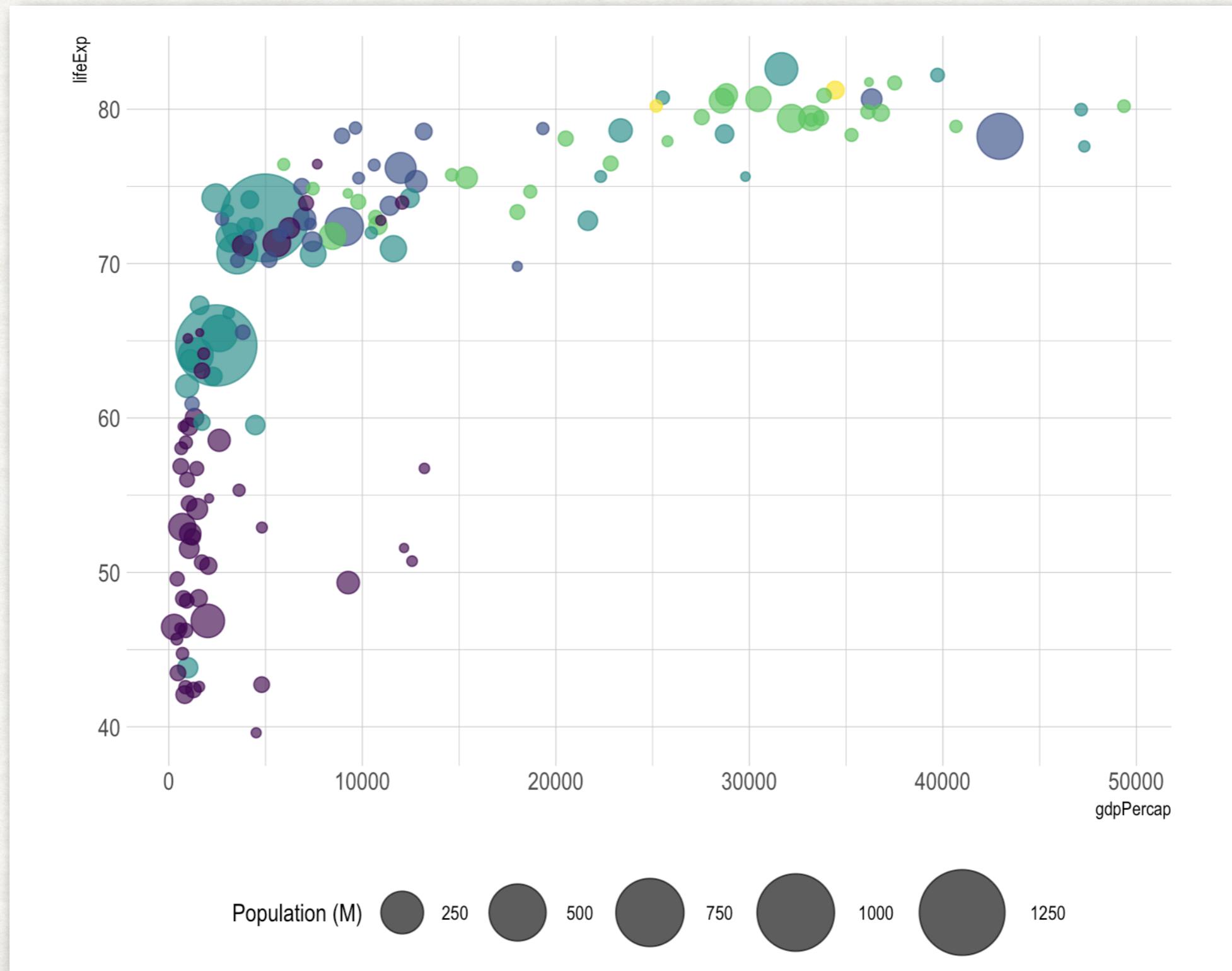


# 重叠区域图 stacked area graph

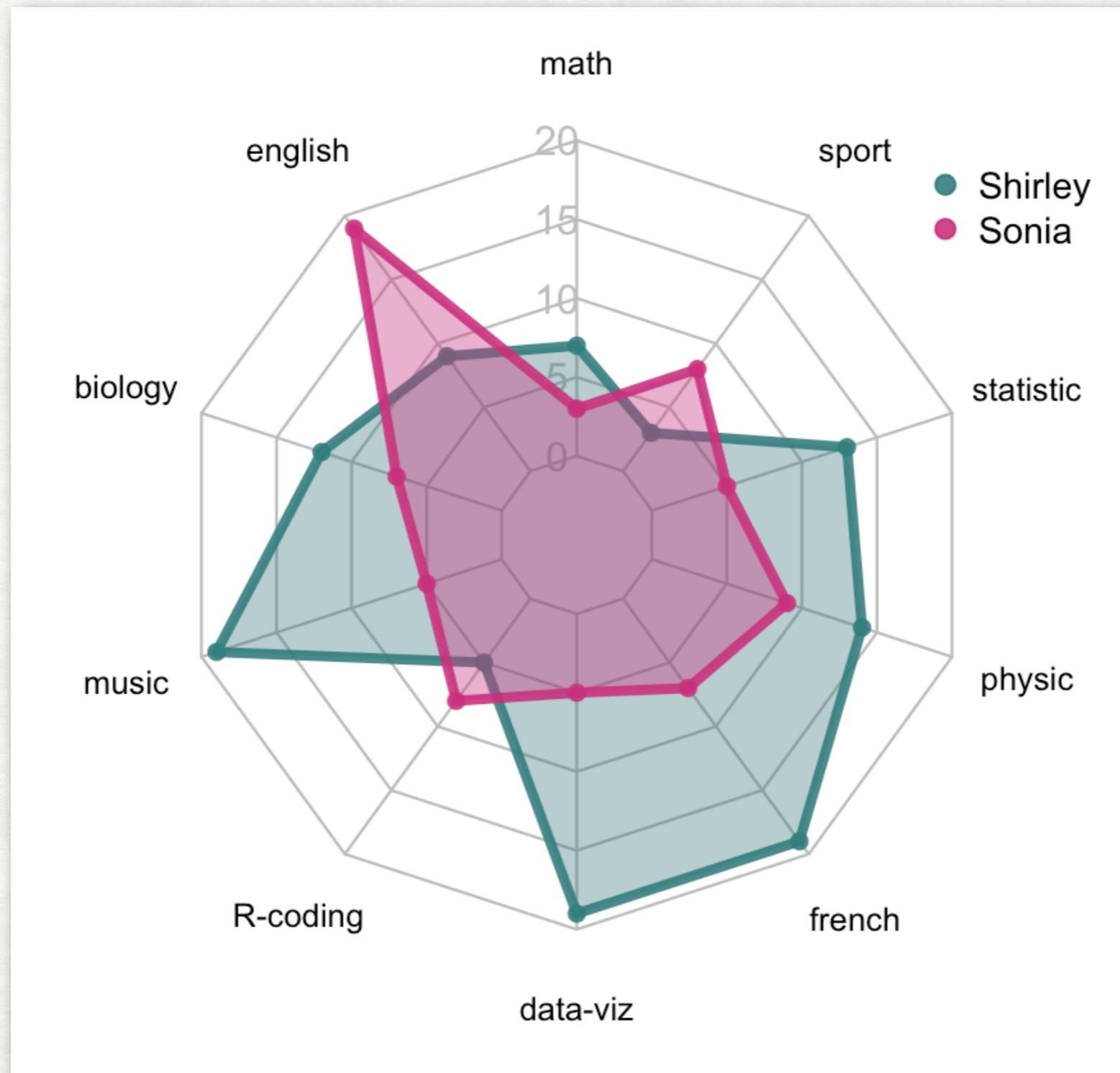
Popularity of American names in the previous 30 years



# 气泡图 bubble plot



# 雷达图（蛛网图） radar (spyder) chart

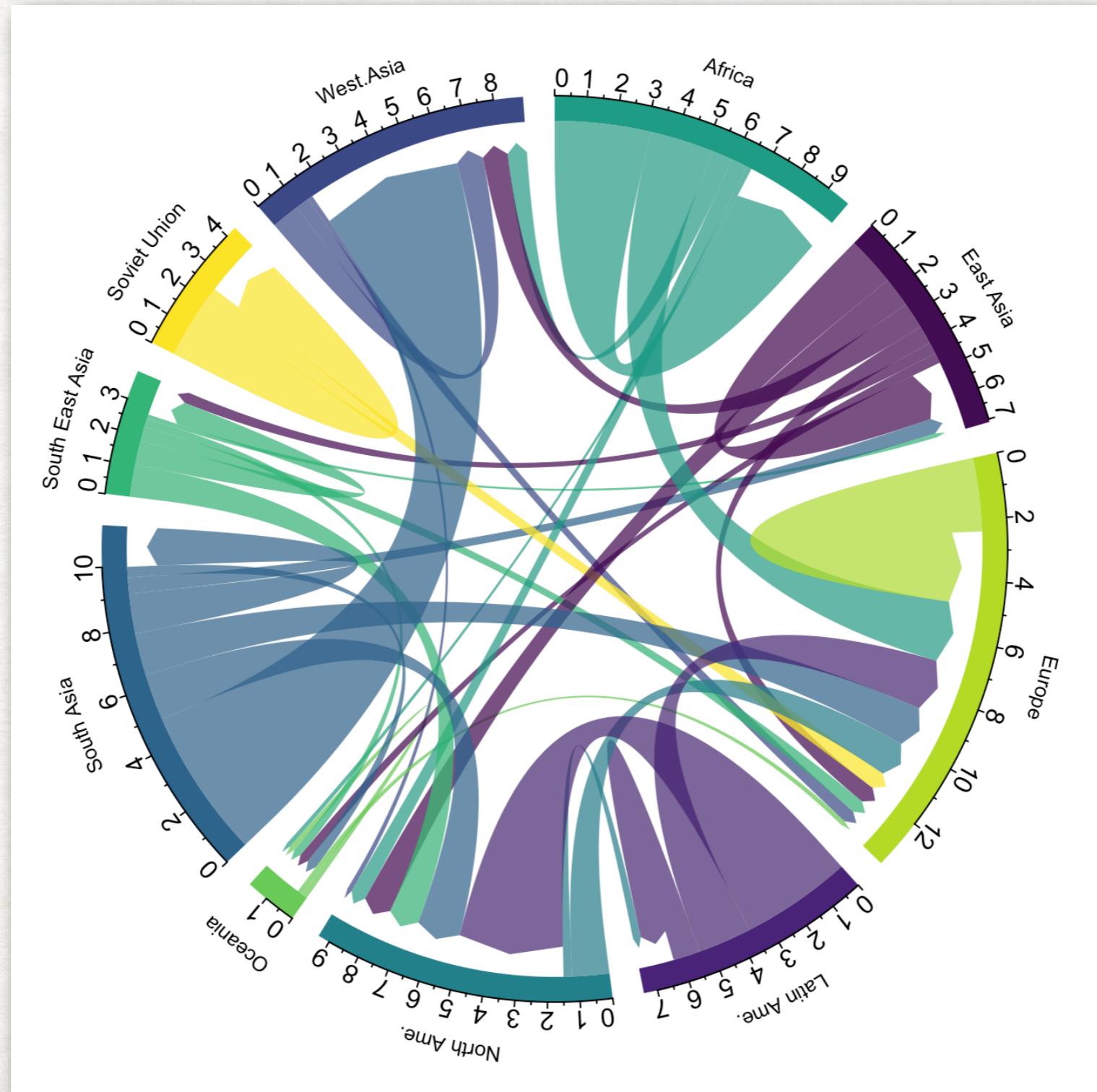


# 词云图 word cloud

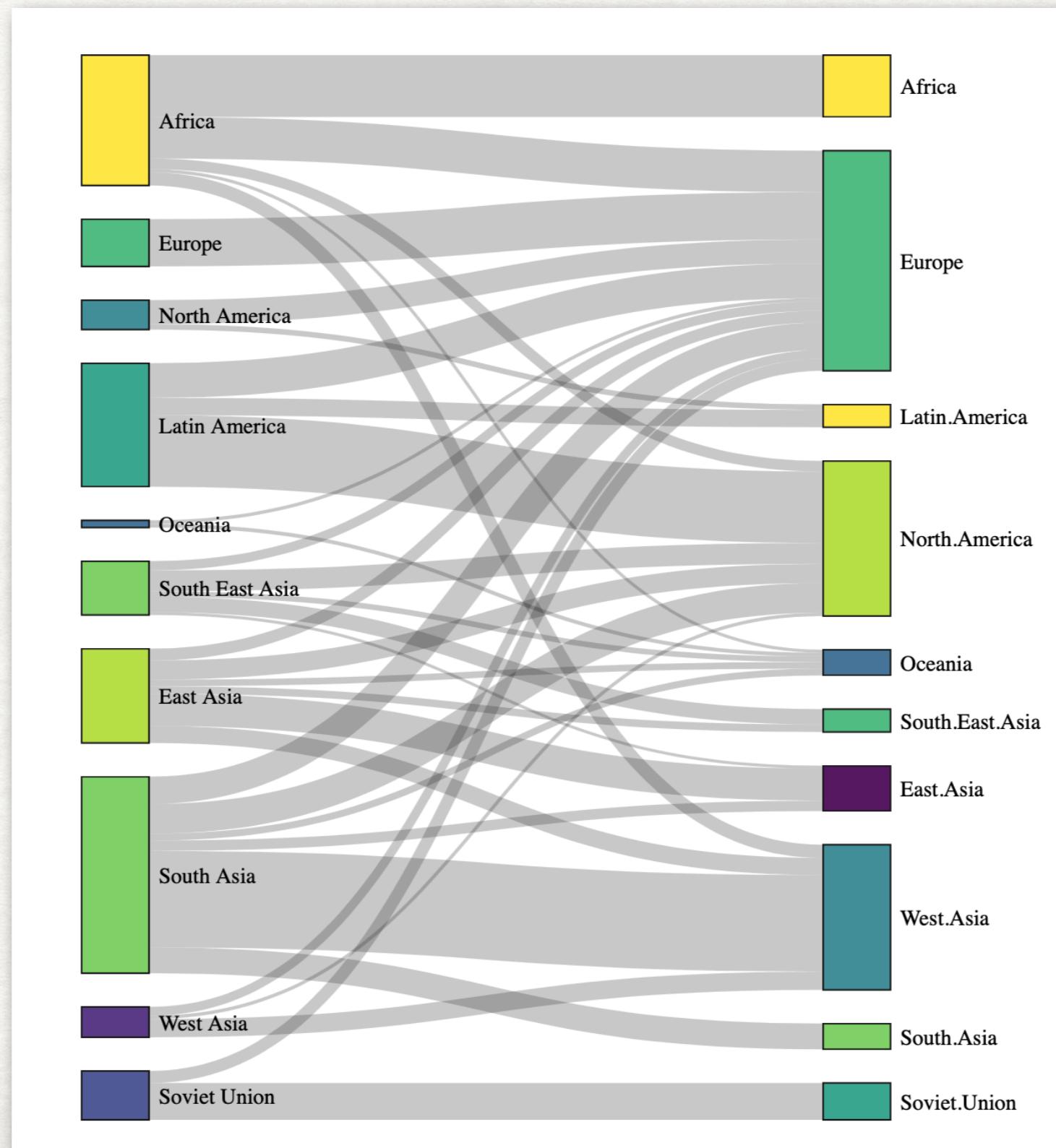


<https://wmx.cnu.edu.cn/tgp/Mihl/e56e52b31bc44adda3b62defba2762d4.htm>

# 和弦图 chord diagram



# 桑基图 Sankey diagram



地区间的移民

# 优化你的图

- 提供必要信息（图例等），删除冗余（多余的网格线、过多的留白等）
- 选择合适的字体和字号、调整符号的大小和粗细
- 给坐标轴标注名称和单位
- 必要时恰当的添加注释
- 可以参考的书籍和网站：
  1. Chen, Härdle, & Unwin. (2008). *Handbook of Data Visualization*. Springer.
  2. <https://informationisbeautiful.net/>
  3. <https://flowingdata.com/>
  4. <https://www.data-to-viz.com/>