

3.2 中美对比

案情回放

2011 年 1 月 19 日，英国《卫报》发表了一篇关于中美国力对比的文章<sup>1</sup>。此时正是中国国家主席胡锦涛访美期间，《卫报》称此次访问是“硬实力与软实力的会晤”。文中分析了中美一些经济社会指标的对比，配图 3.3。

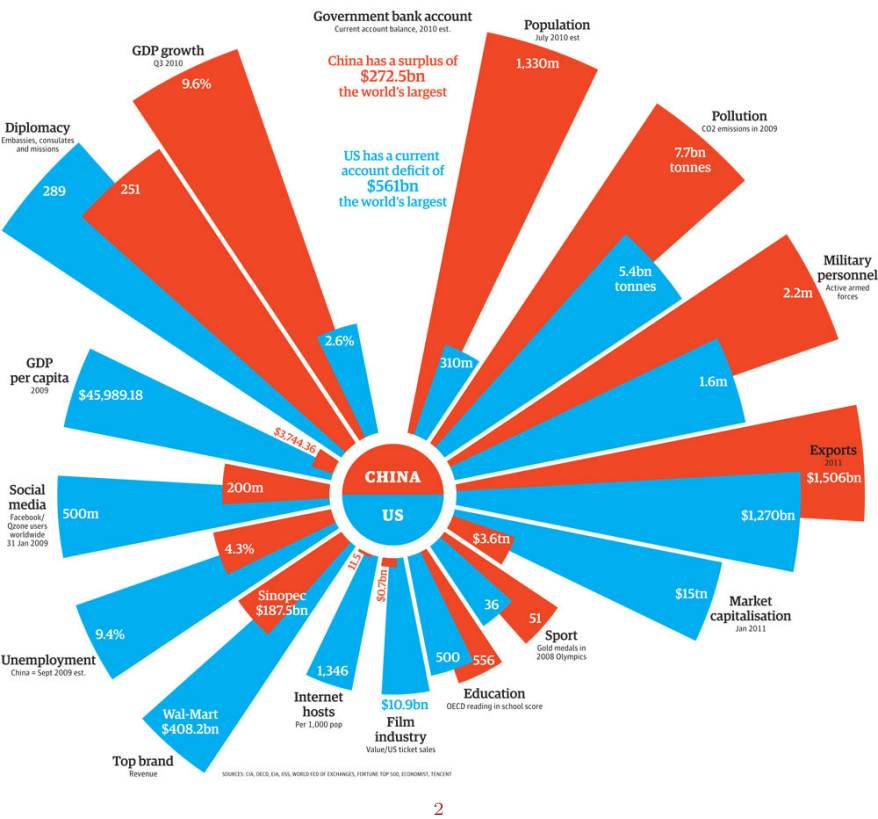


图 3.3: 2010 年中美国力对比：中国人口多、污染重、GDP 增长快；美国人均 GDP 高、市值高、失业率高

总体来说，这幅图形比较有冲击力，能突出两国各自的优势劣势，尤其值得称道的是作者将各个指标进行了排序，这样便于读者阅读。然而，如果我们再细看一下，就能发现一个问题：由于指标数值都是用扇环表

<sup>1</sup><https://www.theguardian.com/news/datablog/2011/jan/19/china-social-media>

表 3.2: 《卫报》提供的 2010 年中美国力数据。未经订正

country	metric	value	units
China	Current account balance	272.5	\$ billion
America	Current account balance	-561.0	\$ billion
China	Population	1330.0	millions of people
America	Population	310.0	millions of people
China	Pollution	7.7	billion tonnes
America	Pollution	5.4	billion tonnes
China	Active armed forces	2.2	million military personnel
America	Active armed forces	1.6	million military personnel
China	Exports	1506.0	\$ billion
America	Exports	1270.0	\$ billion
China	Market capitalisation	15.0	\$ trillion
America	Market capitalisation	3.5	\$ trillion
China	2008 Olympics	51.0	Gold medals in 2008 Olympics
America	2008 Olympics	36.0	Gold medals in 2008 Olympics
China	Reading ability	556.0	OECD reading score in schools
America	Reading ability	500.0	OECD reading score in schools
China	Diplomacy	251.0	number of embasies, consulates and missions
America	Diplomacy	289.0	number of embasies, consulates and missions
China	Unemployment	4.3	% of working population
America	Unemployment	9.4	% of working population
China	GDP growth	9.6	GDP growth for % on previous year
America	GDP growth	2.6	GDP growth for % on previous year
China	\$ GDP per capita	374436.0	\$ GDP per capita
America	\$ GDP per capita	459891.2	\$ GDP per capita

示，因此容易造成一个误会，即数值大小究竟体现在扇环的半径上还是面积上？不同读者的视角自有不同，但这幅图如果按照面积大小去解读数据就错了。该文提供了部分原始数据<sup>3</sup>，见表 3.2，从中可以看到，中国人口是美国人口的 4.3 倍，但图中中国扇环（红色）的面积显然不止是美国的 4.3 倍，这样可能给读者造成一个印象，似乎中国人口是美国的七八倍。

那么，中美国力的实际对比到底是怎样的？我们应该如何用图形恰当地表达其差异？

<sup>3</sup><https://docs.google.com/spreadsheets/d/1Fs0B5uCUwt6p7Csji5CwiZViHspS83pCJ6i-130iL7I>

探案过程

我们利用这些数据重画了《卫报》的图形，如图 3.4 所示。

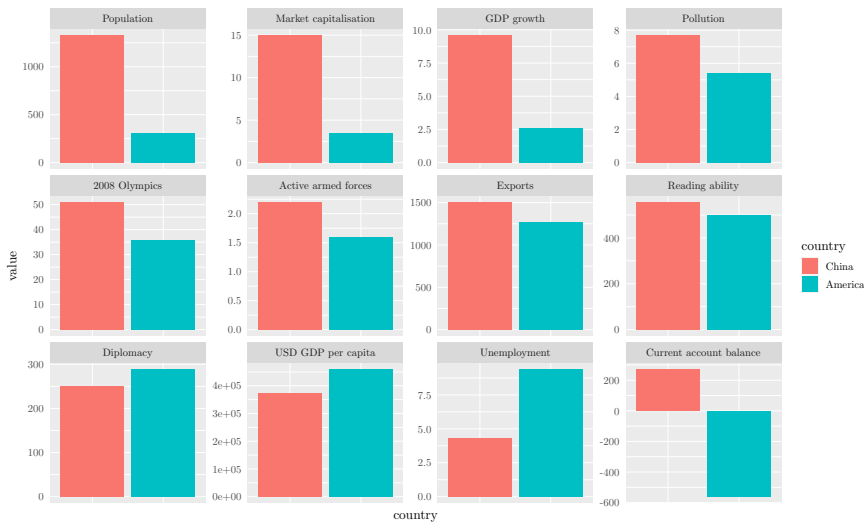


图 3.4: 重新调整后的 2010 年中美国力对比图：条形图的高度表达数值大小更清楚（图中红色代表中国）。各子图显示的指标详见表 3.2。注意：本图的“市值”（Market Capitalisation）数据与图 3.3 所示的数据中美两国颠倒了，这源于《卫报》给出的数据表中的错误。数据表中还有两处错误，你能找到吗？

由于我们的主要目的是比较两国的指标，因此简单的条形图足以完成这个任务。我们大致上按中国指标与美国指标的比率大小排列这些条形图。虽然这幅图看上去比《卫报》的图形逊色，但不会造成任何误解，也能更客观反映两国差异。当然，更好的解决方案可能是保持原有的“放射性”设计，但把扇形替换为矩形条。

我们在检查《卫报》提供的数据表时，发现表中“市值”（Market Capitalisation）一栏中美数据颠倒了，对比原图和我们重画的图让我们很快注意到了这个差异。数据的录入错误无处不在，而“看数字不如看图形更直观”这一点也让我们相信图形在检查错误方面有其特殊价值。数据表中还有两处错误，你能通过对比两图找到它们吗？

## 探案工具

本案例使用了条形图（详见 4.1 小节）和玫瑰图（详见 1.3 和 5.5 小节），所使用的数据收集在 **MSG** 包中，前文所述的《卫报》数据表中的错误经过了订正。加载数据和绘制图形的代码如下：

```
# ggplot2 绘制 2010 年中美国力对比图
library(MSG)
library(ggplot2)
data(cn_vs_us, package = "MSG")
print(
  ggplot(cn_vs_us, aes(x = country, y = value, fill = country)) +
  geom_col() + facet_wrap(~metric, scales = "free_y", ncol = 4)
)
```

## 3.3 新冠疫情

### 3.3.1 案情回放

2020 年春节前后，新冠疫情爆发，亿万人的生活和工作都或多或少受到疫情的影响。笔者在关注疫情统计数据的过程中，遇到了一些有意思的问题。这里试举两例。

第一个例子是，笔者看到朋友分享的一幅疫情地图（如图 3.5），显示了 1 月 22 日全国各省确诊和疑似病例，只见沿海和中部地区一片火红（红色表示新冠肺炎确诊病例），唯有江苏省是白色。图上搭配文字：“在吗，救我”，在地图上格外抢眼。笔者即将赴江苏就职，对此格外关注的同时，不免产生疑问：江苏和周边省份差别真的有这么大吗？为什么？

第二个例子是，有一段时间，一种酷似鹦鹉螺的玫瑰图非常流行，如图 3.6 所示，显示了 2020 年 3 月 27 日除我国以外的世界各国新冠疫情。图中用不同颜色、相同圆心角的扇形来代表确诊 707 例以上的不同国家，各国新冠肺炎的确诊和死亡人数以数字标注。显然，图形经过了精心设计，令人过目不忘。然而，仔细观察却不免心生困惑：与 3.2 小节所举的例子类似，如果以半径代表数值，那么单从图形看，会给人一种印象，即美国确诊数是德国的 2 倍，而德国大约是英国的 2 倍；但一看数字，德国却大约是英国的 4 倍。那么，确诊数量究竟体现在扇形的哪个属性呢？从图上如何看出各国的实际差异呢？