

全球幸福指数分析报告

胡延伸PB22050983

2025-06-07

World Happiness Report（世界幸福报告）是联合国可持续发展解决方案网络发布的全球性调查，基于经济、健康、自由、社会支持等多维度因素对全球国家居民的主观幸福感进行评估与排名。

本项目以 2015–2019 年的《World Happiness Report》数据为基础（数据来源：<https://www.kaggle.com/datasets/unsdsn/world-happiness?resource=download>（<https://www.kaggle.com/datasets/unsdsn/world-happiness?resource=download>）），围绕“全球幸福画像”主题展开系统的数据探索与分析。我们首先对各年度数据结构进行了统一清洗与合并，构建跨年完整数据集。随后，通过可视化与统计分析，深入探讨了哪些因素显著影响国家幸福指数，包括人均 GDP、社会支持、健康寿命、自由程度等。进一步地，我们构建回归模型进行验证，并利用聚类分析揭示不同国家的幸福特征分群差异，最终以地图形式呈现全球幸福分布的空间格局。

本报告的研究有助于从数据角度理解全球幸福感的结构性差异，并为相关公共政策制定提供量化支持。

数据探索与整理

本项目使用的《世界幸福报告》数据来自 Kaggle，涵盖 2015–2019 年度全球多个国家的幸福指数及其影响因素。由于数据按年份分别存储在多个 CSV 文件中，字段名称和结构略有差异，因此首先需对数据进行统一清洗和整合。主要工作包括：字段对齐、缺失值处理、变量初探和可视化探索等。

数据载入与结构统一

首先，使用 `tidyverse` 工具集批量读取各年度数据，并通过 `stringr` 从文件名中提取年份作为 `Year` 字段。在字段标准化过程中，考虑到不同年份中相同变量可能使用不同名称（如“GDP per capita”、“Economy..GDP.per.Capita.”），我们使用 `coalesce()` 函数统一重命名关键变量，包括：

- `HappinessScore`：幸福指数
- `GDP`：人均国内生产总值
- `SocialSupport`：社会支持程度
- `Health`：健康寿命预期
- `Freedom`：生活自由度
- `Generosity`：慷慨程度
- `Trust`：政府廉洁程度（或腐败感知）

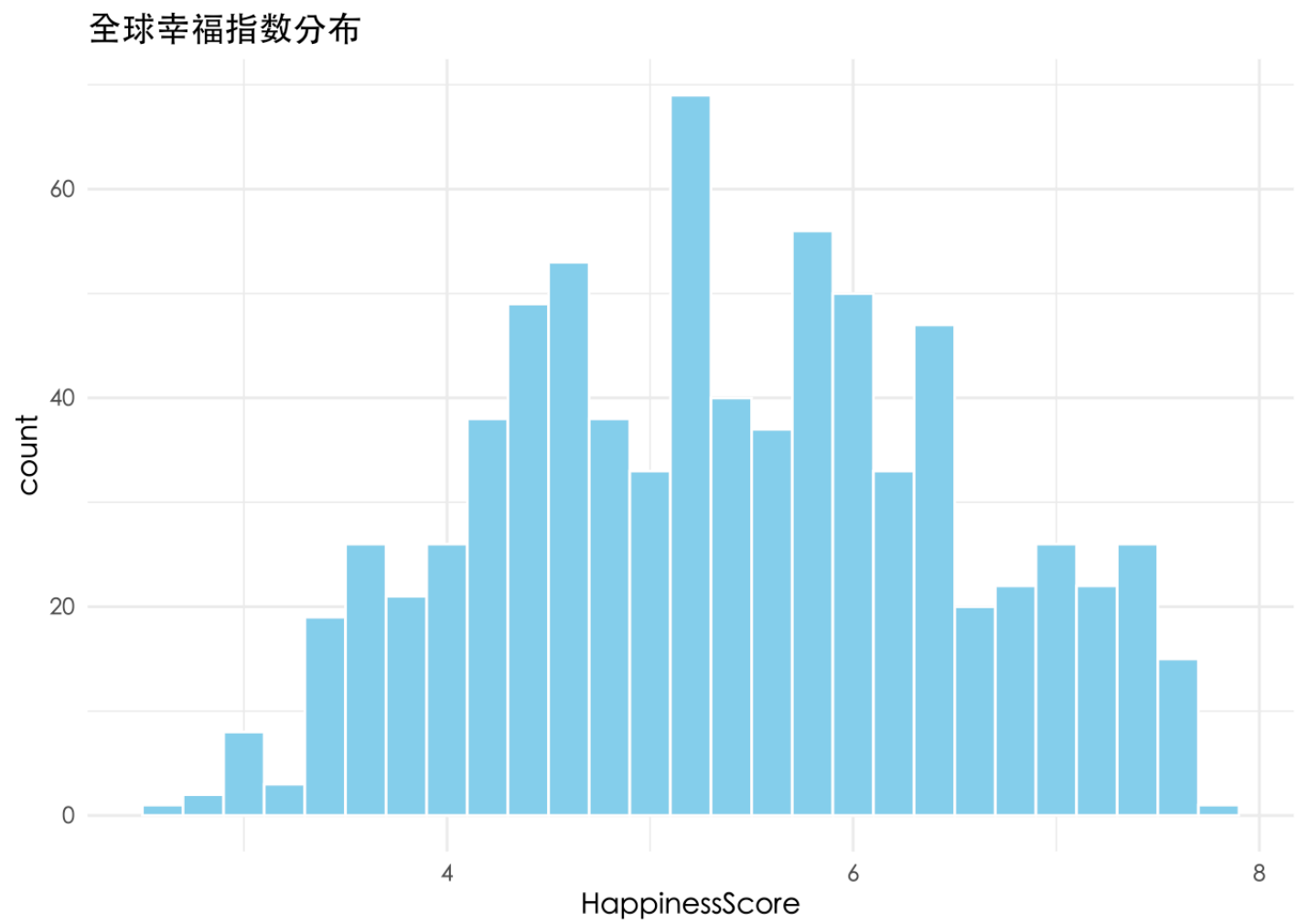
此外，所有数值型字段均通过 `mutate(across(...))` 转换为 `numeric` 类型，以便后续分析处理。

缺失值处理

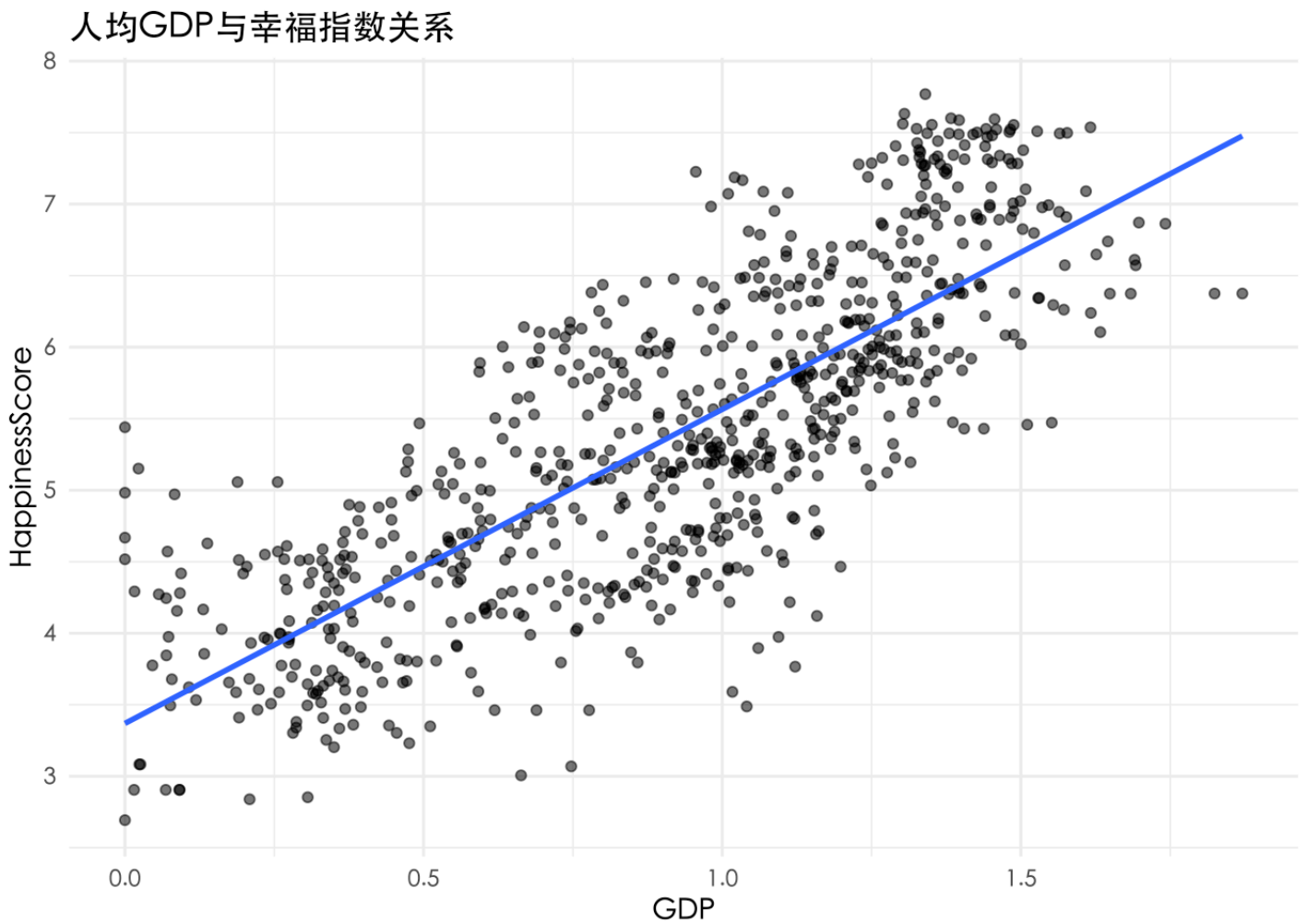
通过 `colSums(is.na())` 检查缺失值情况后，决定使用 `na.omit()` 删除缺失观测。这是因为本项目聚焦于幸福因素之间的相对关系分析，适量删除缺失数据不会对总体趋势造成实质影响，反而能简化后续建模与聚类分析。

初步可视化分析

为了分析全球幸福指数的总体分布情况，我们绘制了幸福指数的直方图（如下图）。从图中可以看出，全球大多数国家的幸福指数集中在 4.5–6.5 分之间，呈现出偏正态分布特征。极少数国家的幸福指数高于 7 分，反映出少数高福利国家的优势地位。而幸福得分低于 4 的国家数量相对较少，可能受到经济、健康或社会环境等多重因素影响。



此外，我们还绘制了人均 GDP 与幸福指数的散点图及线性拟合趋势线，可以直观地观察到二者存在较强的正相关性，这为后续的相关性分析和回归建模提供了初步依据。

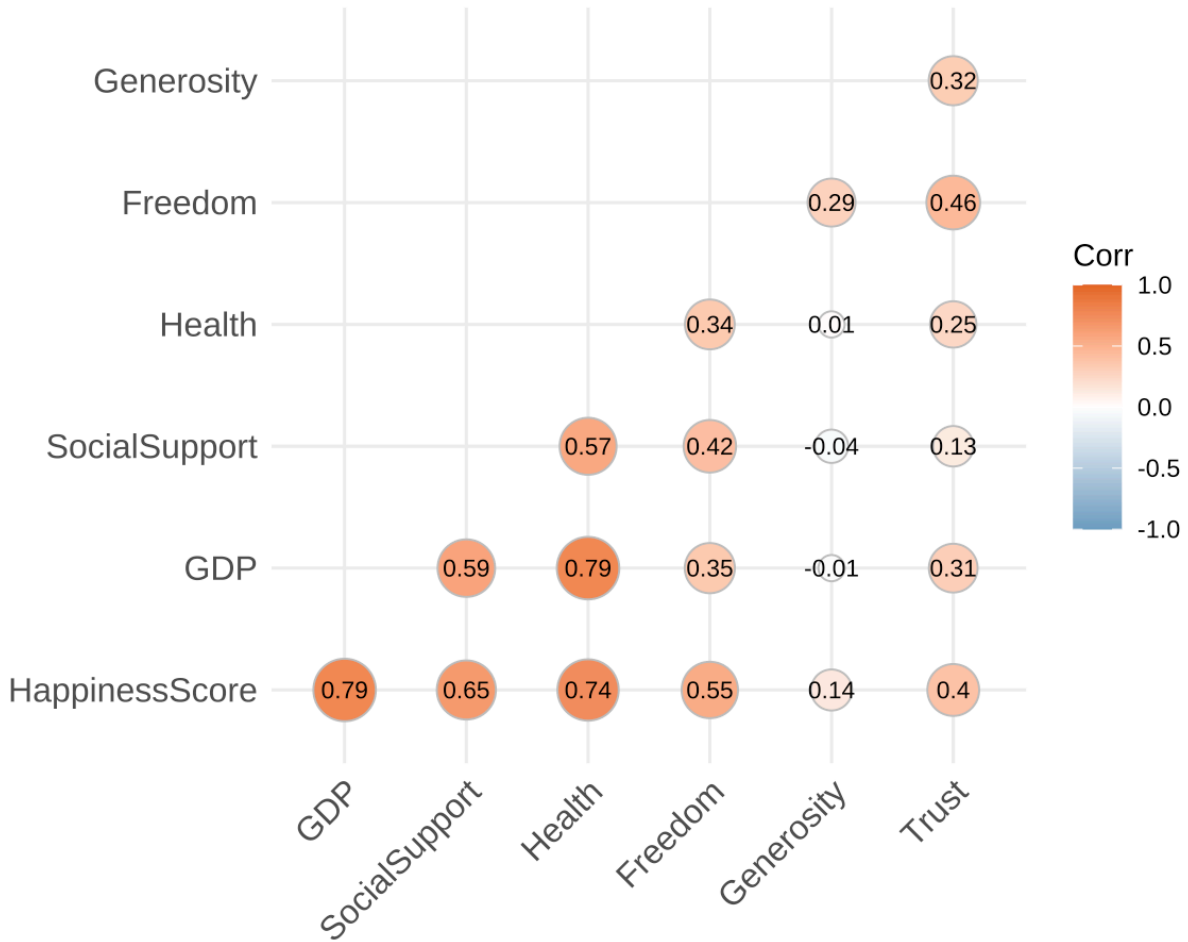


任务一：相关性分析

为了探索影响幸福指数（Happiness Score）的关键因素，我们选取了六个与幸福密切相关的变量，分别为：人均 GDP（GDP）、社会支持（SocialSupport）、健康寿命预期（Health）、自由度（Freedom）、慷慨程度（Generosity）与对政府信任（Trust），构建了皮尔逊相关性分析矩阵：

并用 `ggcorrplot` 可视化呈现（如下图所示）。

幸福指数相关性矩阵



从图中可以看出，幸福指数与各因素的相关性强弱排序如下：

- 人均 GDP：0.79，与幸福指数高度正相关；
- 健康寿命：0.74，正相关性也非常显著；
- 社会支持：0.65；
- 自由度：0.55；
- 政府信任：0.40；
- 慷慨程度：0.14，相关性较弱。

这表明经济、健康、社会和政治环境是幸福感的主要来源，尤其是人均 GDP 和健康寿命，对整体幸福得分的提升起到了决定性作用。

回归模型验证

为了进一步验证上述结论，我们使用多元线性回归模型拟合幸福指数与六个因素之间的关系。模型结果如下：

$$\text{HappinessScore} = 2.18 + 1.14 \cdot \text{GDP} + 0.65 \cdot \text{SocialSupport} + 1.01 \cdot \text{Health} + 1.48 \cdot \text{Freedom} + 0.59 \cdot \text{Generosity} + 0.86 \cdot \text{Trust}$$

- 所有变量的回归系数均为正，且 p 值远小于 0.001，统计上显著；
- 模型 $R^2 = 0.76$ ，说明该模型能解释 76% 的幸福指数变化；
- 回归系数最大的是自由（Freedom = 1.48），其次是 GDP、健康与信任。

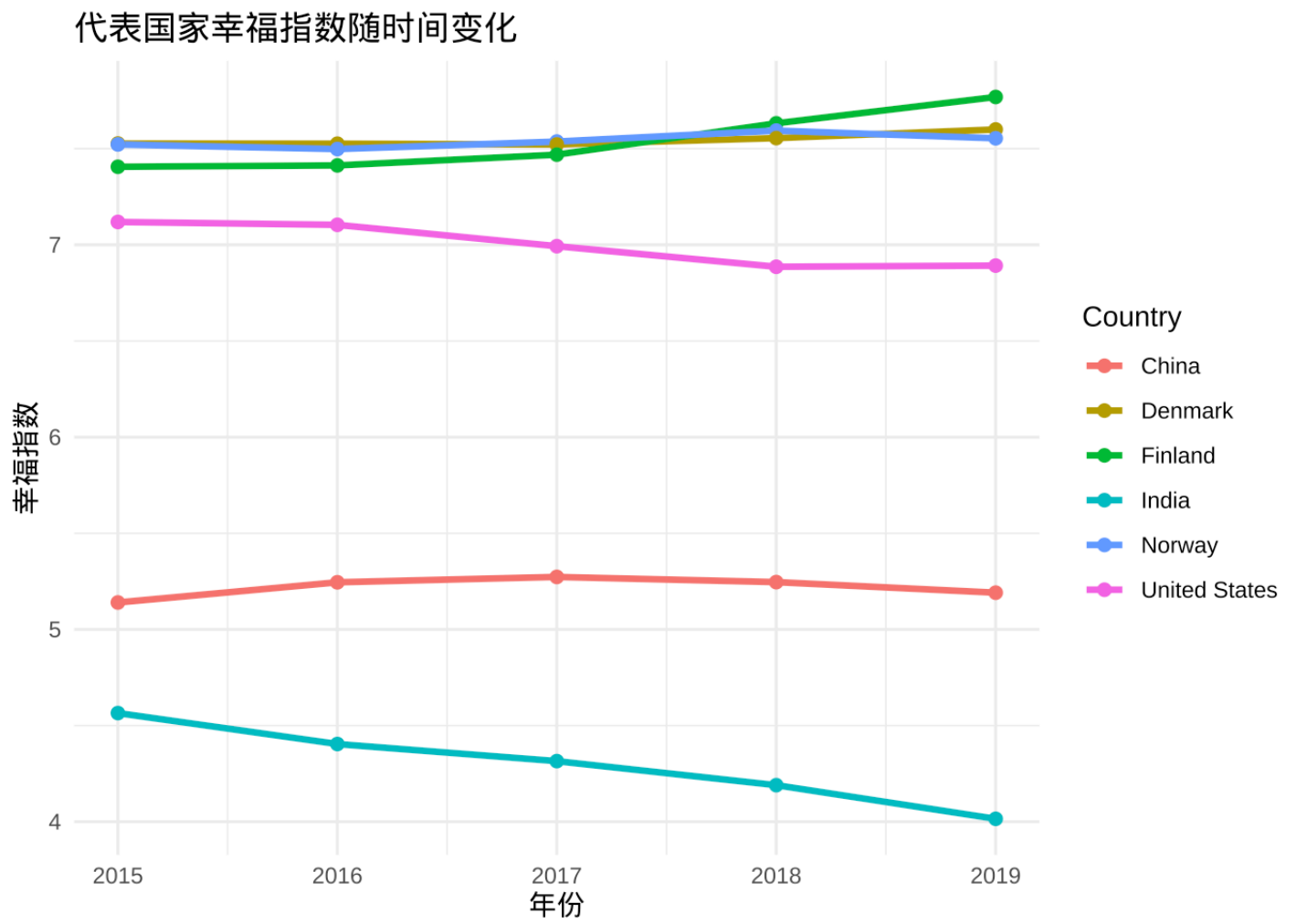
这说明在控制其他变量的情况下，自由度对幸福指数的边际提升效果最大，其次是经济收入与健康水平。

任务二：幸福指数随时间变化趋势分析

为了探究全球及特定国家的幸福指数如何随时间演变，我们从国家层面与全球平均层面分别开展了趋势分析。同时，我们还统计了在 2015–2019 年间幸福指数提升或下降幅度最大的国家，挖掘其中的结构性变动特征。

1. 代表性国家幸福指数变化趋势

我们选取了六个具有代表性的国家：芬兰、丹麦、挪威、美国、中国和印度，绘制其 2015–2019 年的幸福指数变化曲线（见下图）。



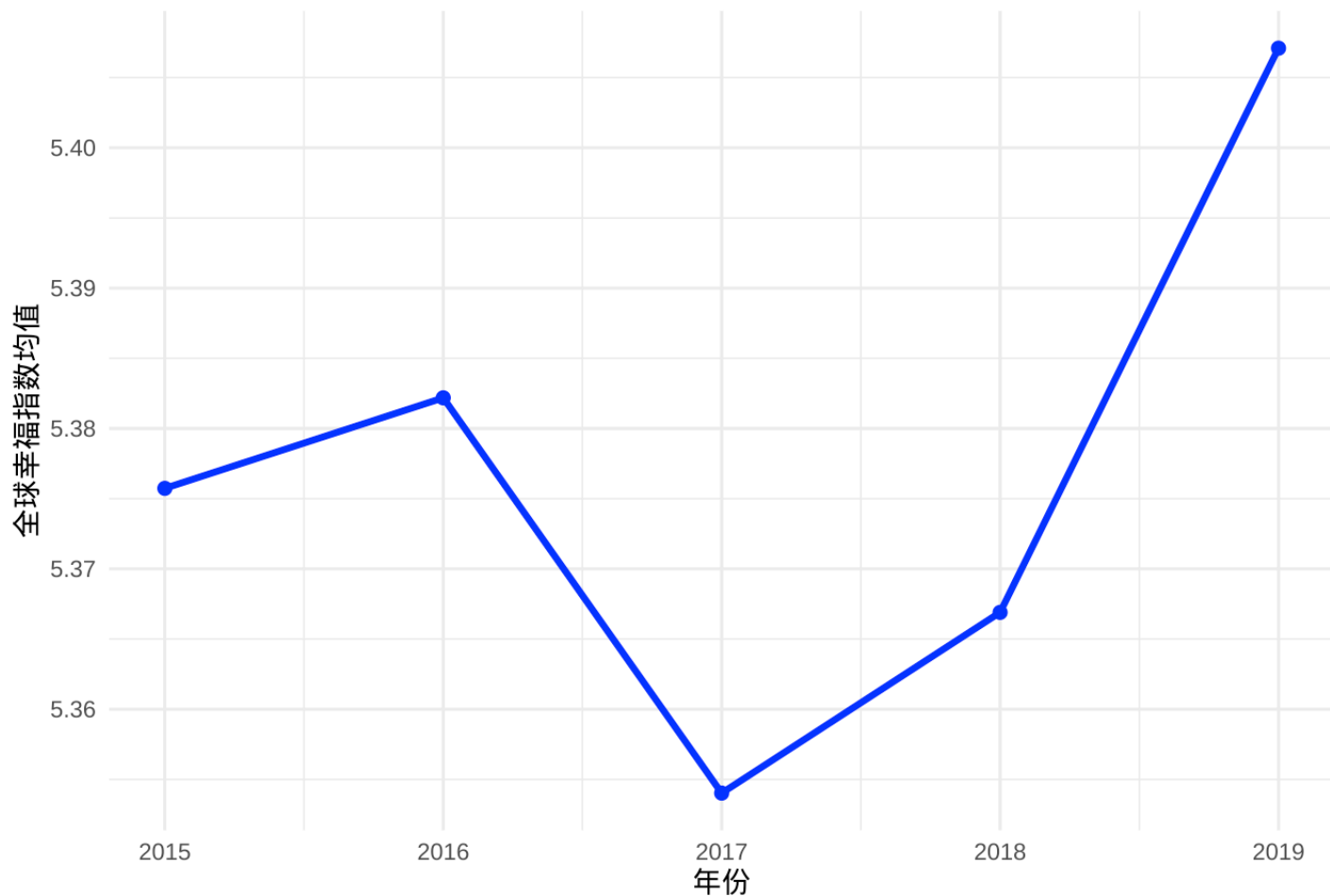
从图中可以看出：

- 北欧三国（芬兰、丹麦、挪威）长期维持在 7.5 分以上，整体趋势**稳定或缓慢上升**，幸福指数处于全球领先地位；
- 美国在该时段略有下降，可能受社会撕裂或心理健康等因素影响；
- 中国的幸福指数在这五年内**稳步上升**，显示出经济增长和生活改善的正向反馈；
- 印度的幸福指数呈现**持续下降趋势**，从约 4.5 降至 4.0，下降幅度显著，值得关注。

2. 全球幸福指数趋势

为了观察整体水平的变动，我们计算了所有国家在各年度的幸福指数平均值并绘图展示。

全球平均幸福指数随时间变化



可以看出：

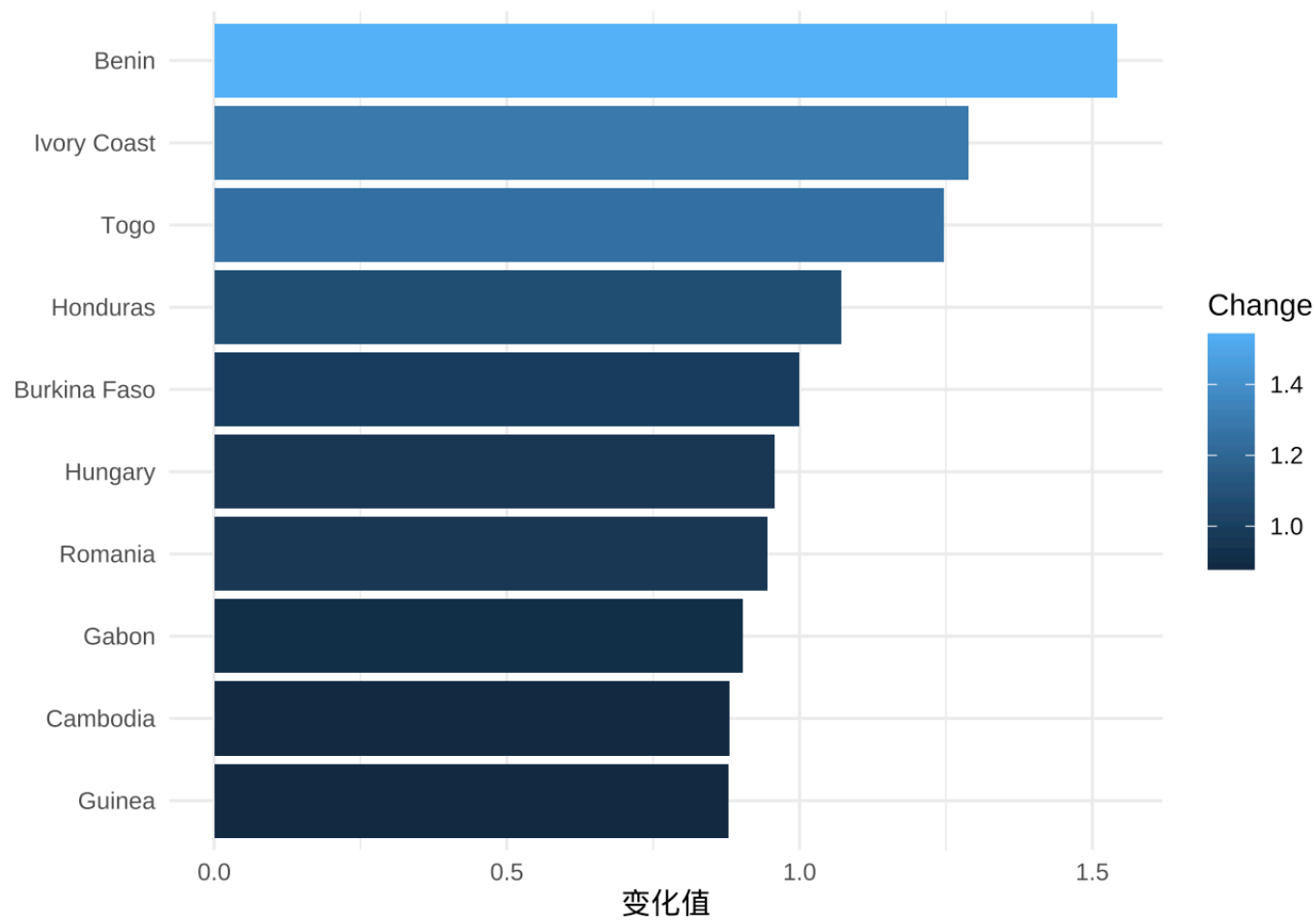
- 2015–2019 年间全球幸福指数平均值总体波动较小，在 5.36–5.41 之间；
- 2017 年出现小幅下降，但在 2018–2019 年逐渐回升，显示出全球范围内幸福感总体保持稳定。

3. 幸福指数变化最大的国家

我们进一步统计了每个国家从 2015 到 2019 年的幸福指数“净变化值”，并绘制了变化幅度最大的前 10 个“上升国家”与“下降国家”。

幸福提升最多的国家：

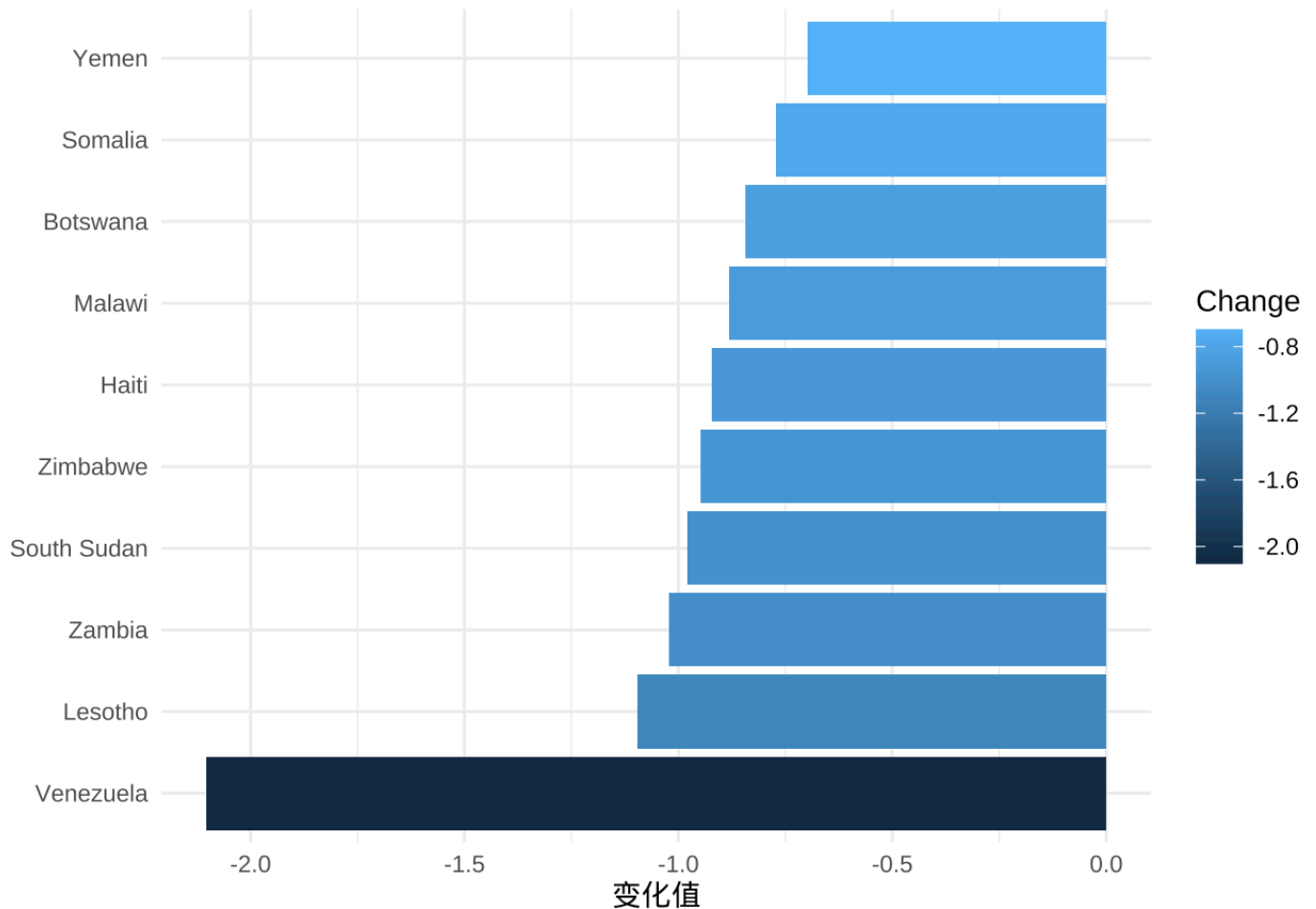
幸福指数提升最多的国家



- 排名第一的 贝宁 (Benin) 幸福指数提高了 1.543 分；
- 其它如 科特迪瓦 (Ivory Coast)、多哥 (Togo)、洪都拉斯 (Honduras)、匈牙利 (Hungary) 等国也有显著上升；
- 多数来自非洲和东欧发展中地区，反映了其近年发展成果和民众生活改善。

幸福下降最多的国家：

幸福指数下降最多的国家



- 委内瑞拉 (Venezuela) 以 **-2.103** 分的跌幅位居首位，受政治与经济危机影响显著；
- 其次为 莱索托、赞比亚、南苏丹、津巴布韦 等非洲国家；
- 整体来看，这些国家的共同特征是**社会动荡、经济衰退或治理不善**，反映出幸福感下降与国家发展密切相关。

任务三：幸福特征聚类分析

为了进一步了解不同国家幸福指数背后的构成模式，我们选取了 2019 年的数据，对全球各国的幸福影响因素进行了**K-means** 聚类分析。聚类所使用的变量包括：

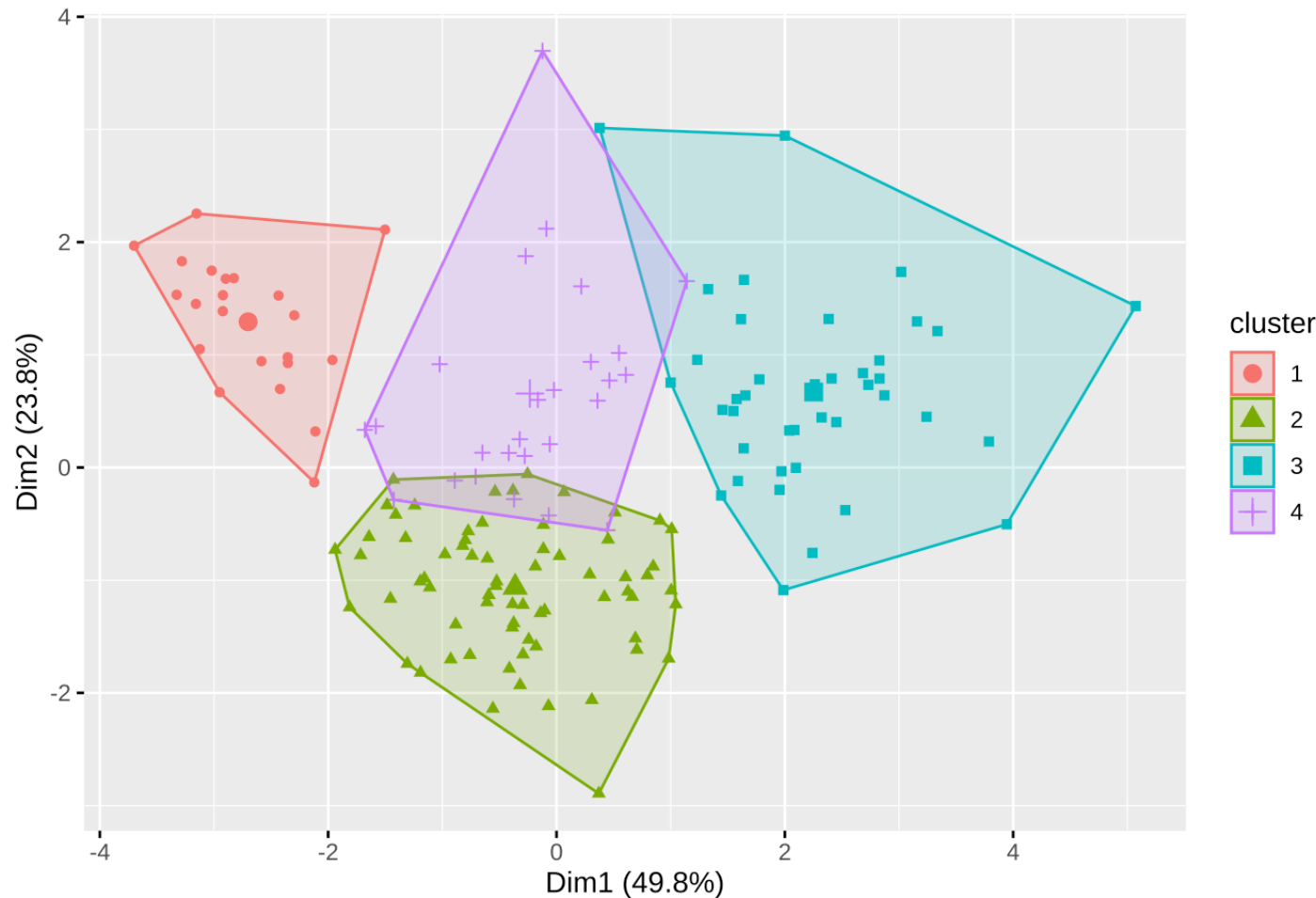
- 人均 GDP (GDP)
- 社会支持 (SocialSupport)
- 健康寿命预期 (Health)
- 生活自由度 (Freedom)
- 慷慨程度 (Generosity)
- 对政府信任度 (Trust)

所有变量均进行了 Z-score 标准化处理，以消除量纲影响。最终设定聚类数为 4，应用 K-means 算法对数据进行划分。

聚类结果可视化

通过 `fviz_cluster()` 可视化聚类结果（如下图所示）：

2019年国家幸福特征聚类



图中每个点代表一个国家，不同颜色和图形表示隶属于不同的幸福特征类别。二维空间通过主成分分析（PCA）降维，可视化国家间的特征相似性。

各聚类类别分布情况如下：

聚类编号	国家数量	特征大致描述
Cluster 2	68 个	中等幸福水平，多数发展中国家，特征较均衡但无突出项
Cluster 3	40 个	低幸福群体，通常是政治动荡或经济落后地区，信任度较低
Cluster 4	26 个	自由高但经济较低，如拉美国家，表现出“非经济型幸福”特征
Cluster 1	22 个	高幸福群体，如北欧国家，GDP、健康、自由等指标优越

解读与发现：

- Cluster 1 国家普遍具备高收入、高社会支持与高健康预期，如芬兰、丹麦、挪威等；
- Cluster 2 为多数国家所处的中间群体，具有一定发展基础，但幸福构成多样且波动大。
- Cluster 3 为低幸福群体，应重点关注其公共治理、健康保障和基础发展问题；
- Cluster 4 国家自由度较高，但 GDP 和信任度较低，代表一类文化乐观但发展受限的国家；

任务四：地图可视化展示

为了更直观地展示不同国家的幸福特征及其地理分布，本任务使用地理信息数据（GeoData）与聚类结果结合，绘制了两张关键地图：

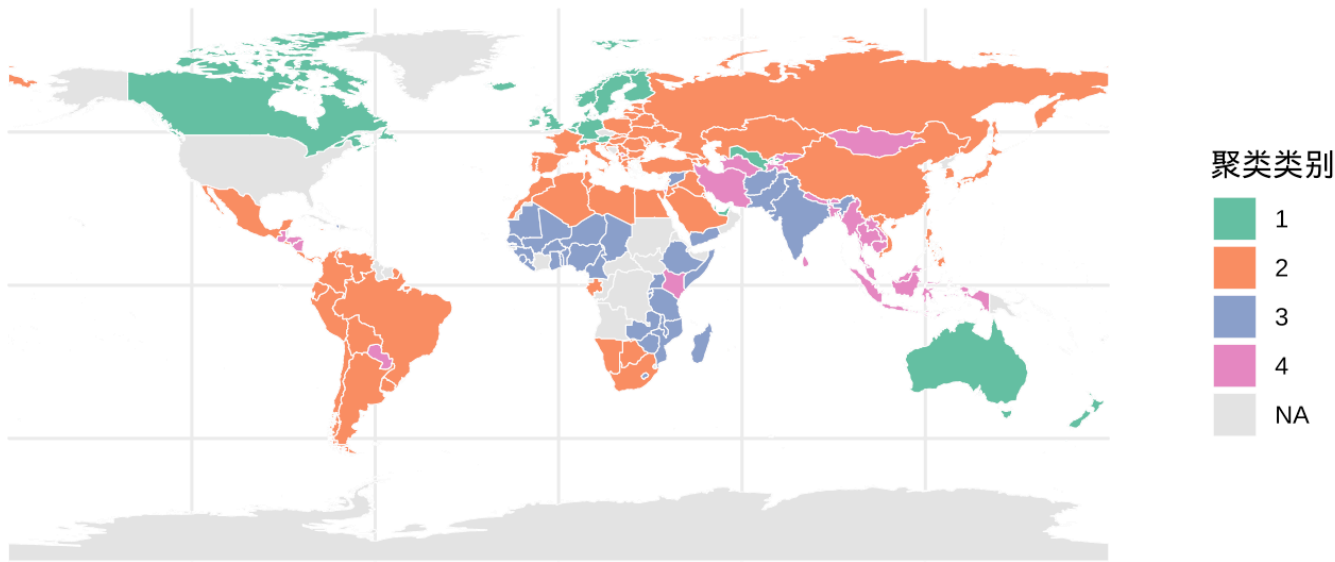
- 1. 2019 年国家幸福特征聚类地图
- 2. 2019 年全球幸福指数热力图

地图基于 `rnaturalearth` 提供的世界国界数据，通过 `sf` 和 `ggplot2` 进行可视化绘制。

1. 幸福特征聚类地图

我们将上一任务中得到的 4 类幸福特征聚类结果与地图数据合并，并绘制如下：

2019年全球幸福特征聚类地图



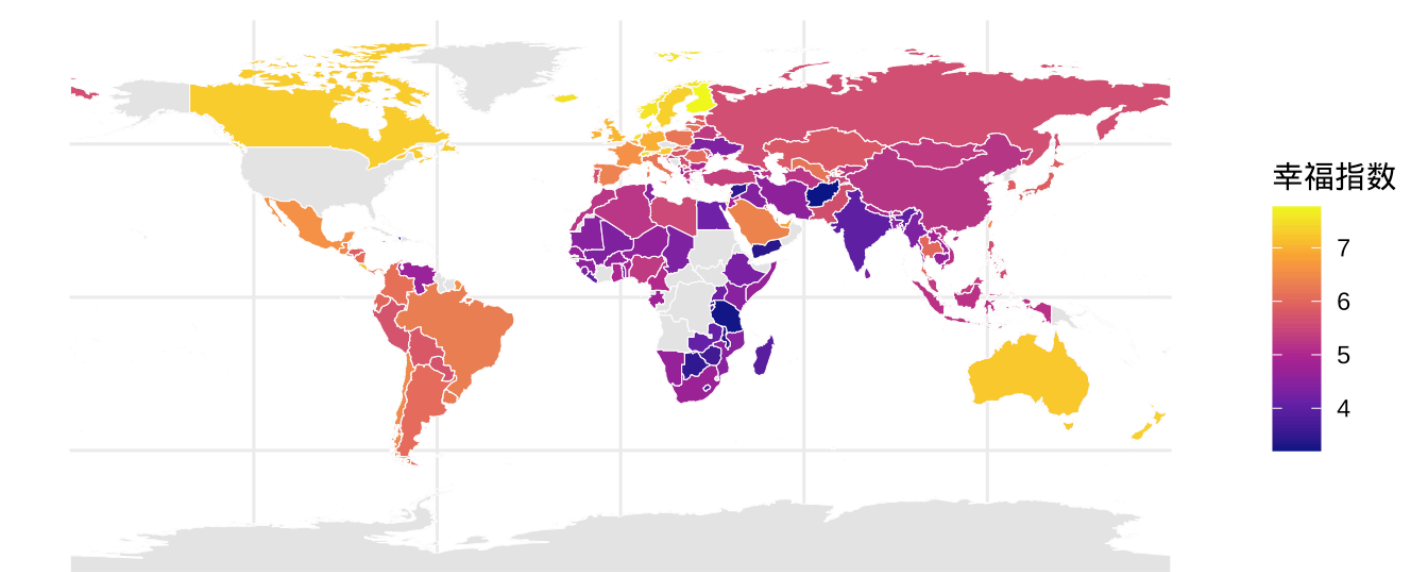
从图中可以看出：

- **绿色区域（Cluster 1）**：主要分布于北欧、澳大利亚和北美等高幸福国家，代表经济发展、社会支持与健康水平俱优的“高幸福群体”；
- **紫色区域（Cluster 4）**：多为拉美、南亚等自由度较高但经济较低的国家，代表“自由驱动型幸福”；
- **蓝色区域（Cluster 3）**：主要集中在撒哈拉以南非洲和战乱国家，显示为“低幸福群体”，值得国际援助关注；
- **橙色区域（Cluster 2）**：是最大的一类，覆盖中国、印度、俄罗斯等，处于“中等发展阶段”，幸福结构平衡但提升空间大。

2. 全球幸福指数热力图

下图展示的是各国在 2019 年的幸福指数（Happiness Score）分布情况：

2019年全球幸福指数地图



图例以颜色深浅反映幸福指数数值：

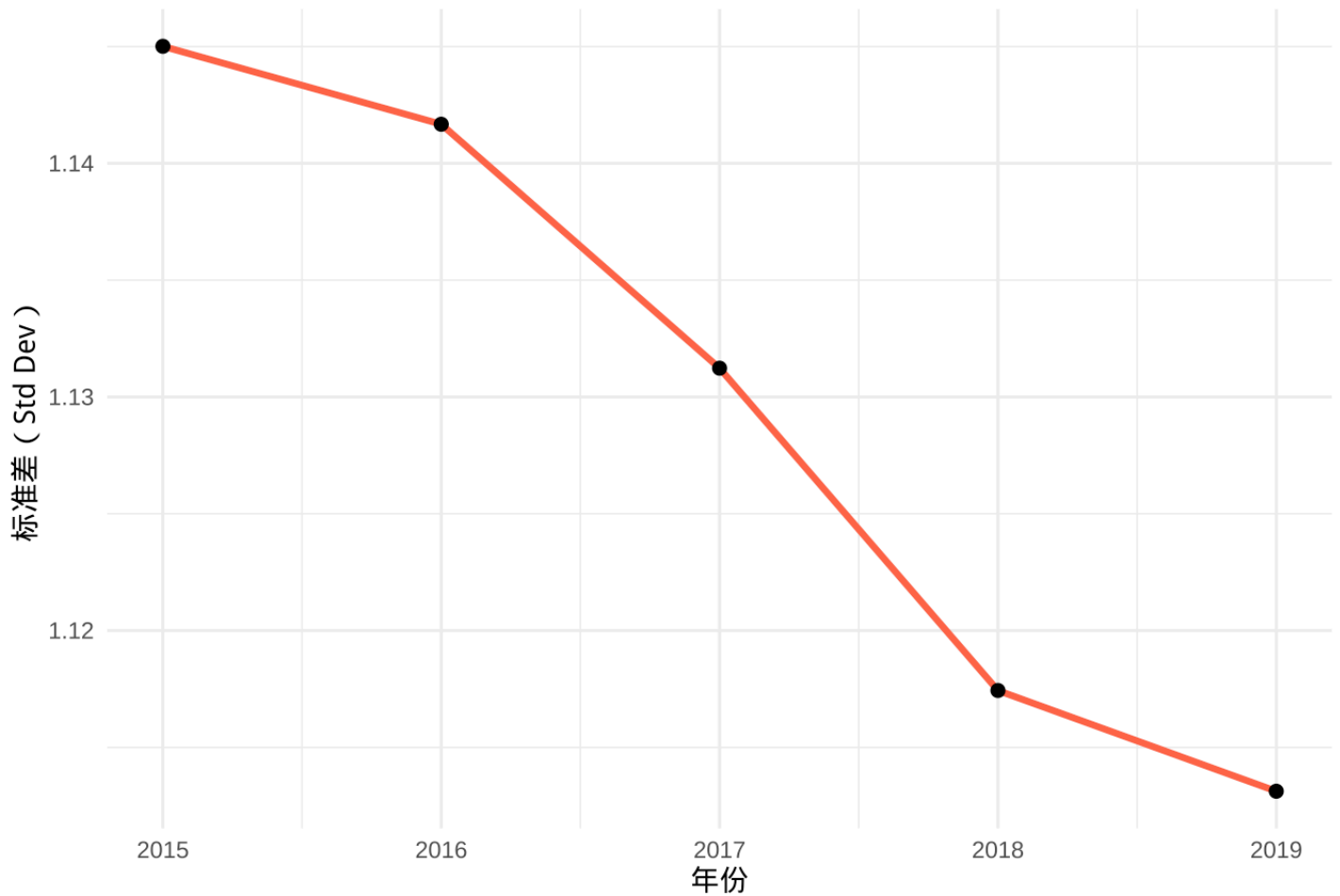
- **黄色区域 (>7)**：代表全球幸福水平最高的国家，如芬兰、丹麦、瑞士等；
- **紫色区域 (<5)**：主要分布于非洲、西亚与部分南亚国家，显示出明显的区域性幸福劣势；
- **中国** 处于 5.1–5.5 区间，幸福感仍有提升空间；
- **拉美国家** 整体表现不低，尽管经济发展中，但可能受文化与社会联系影响。

任务五：幸福不平等分析 —— 幸福分布的差异与变化趋势

在前面的分析中，我们聚焦于各国幸福指数的整体水平与影响因素。然而，幸福的公平性同样值得关注：**全球幸福是否正在向少数国家集中？幸福差距是否在扩大或缩小？**为此，我们从标准差、四分位差（IQR）以及年度分布箱型图三个角度，考察 2015–2019 年全球幸福指数的不平等变化趋势。

1. 年度标准差变化趋势

每年全球幸福指数标准差变化

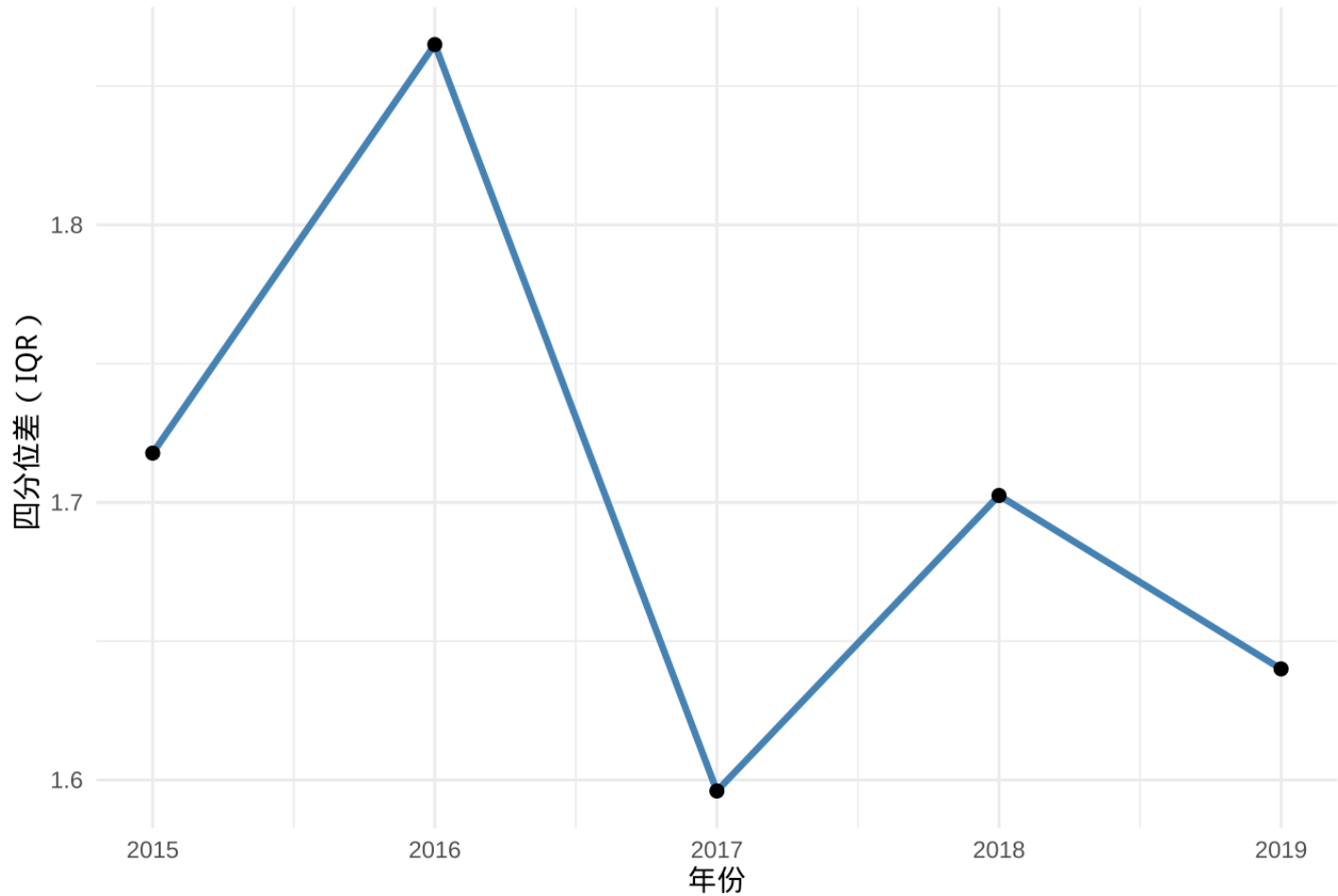


如上图所示，从 2015 到 2019 年，全球各国幸福指数的标准差逐年下降，从约 **1.146** 降至 **1.116**。这说明：

- 各国之间的幸福得分逐步趋于收敛；
- **全球幸福差距在缓慢缩小；**
- 在幸福构成主要受经济、健康等结构性因素决定的前提下，全球发展存在一定的“追赶效应”。

2. 年度四分位差（IQR）变化趋势

每年全球幸福指数四分位差（IQR）变化

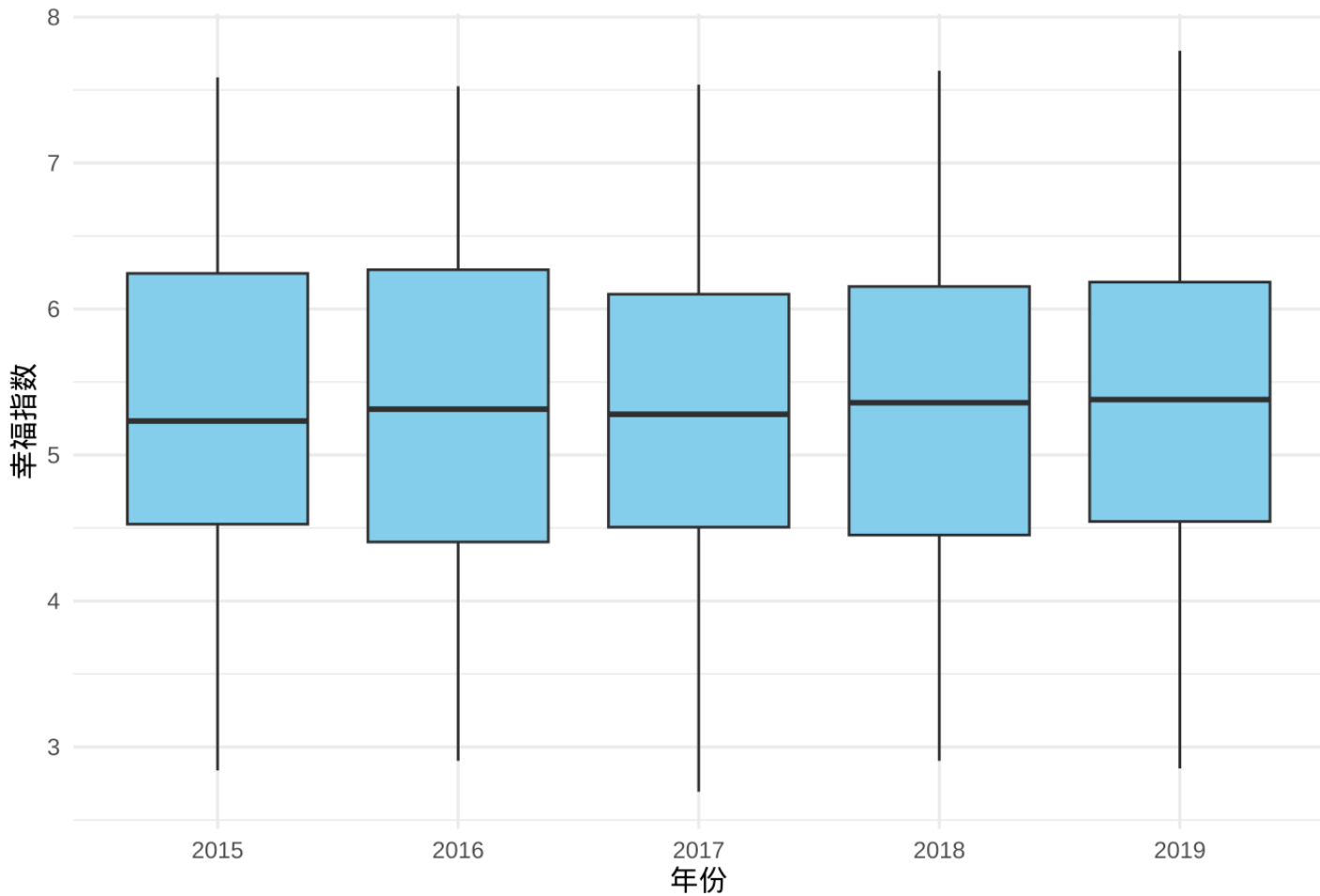


四分位差（IQR）更能反映“中间 50% 国家”之间的幸福波动范围。可以观察到：

- IQR 在 2016 年达到峰值后有所回落；
- 整体在 **1.6–1.9** 之间波动；
- 说明中等幸福国家的内部差距略有起伏，但不存持续扩大趋势。

3. 年度幸福指数分布箱型图

幸福指数年度分布箱型图



从年度箱型图中可以进一步确认：

- 幸福得分的中位数保持稳定，说明多数国家的幸福水平未发生剧烈变化；
- 但箱体上下边界（Q1-Q3）略有收窄，与前两项趋势一致；
- 极端值（如低幸福国家）依然存在，但未主导整体结构。

本部分分析表明，**2015–2019 年全球幸福的不平等程度总体呈下降趋势**。无论是从标准差、IQR，还是箱型分布角度，全球国家之间的幸福差距并未持续扩大，反而表现出轻微的收敛特征。

这可能与以下因素相关：

- 发展中国家（如中国、东南亚、非洲部分国家）经济与健康状况改善；
- 国际援助与社会支持体系初显成效；
- 高幸福国家维持稳定状态，而低幸福国家略有上升。

未来仍需关注：**极端低幸福国家的底线保障、中等国家群体的幸福波动性管理**，以及避免“表面均衡、结构不稳”的风险。

政策建议与分析总结

本项目通过系统分析 2015–2019 年全球幸福指数数据，从因素识别、变化趋势、聚类分类与地理可视化等角度，描绘了一个立体的“全球幸福画像”。基于上述研究结果，我们提出以下几点政策性建议与思考：

1. 提升基础民生仍是关键路径

人均 GDP、健康寿命和社会支持被证明是最显著的幸福决定因素。对于中低收入国家，持续加强公共医疗体系建设、教育普及、社会保障覆盖与经济机会平等，是提升国民幸福感的核心战略。

2. 关注自由与信任的制度环境

自由度和对政府信任虽然在回归模型中系数略低于经济指标，但其独立效应显著，说明幸福不仅来自“生活质量”，还来自“治理质量”。政策制定者应重视民主参与、反腐败制度、公民自由与法治环境的建设，以增强社会信任与凝聚力。

3. 因地制宜设计幸福政策

聚类结果显示，不同国家或地区的幸福特征存在明显异质性。例如，北欧国家属于“全维优势型”，而部分南美国家尽管经济较弱，却在自由度或社会连结上具有优势。因此，提升幸福感不能简单“照搬”他国经验，而应根据本国文化、制度与发展阶段制定差异化策略。

4. 加强对“低幸福群体”的国际支持

聚类与地图可视化表明，非洲、中东等部分地区长期处于幸福指数低位，且缺乏基本经济与健康支撑。国际社会应在公共卫生、粮食安全、基础教育等方面提供更多援助，同时推动和平稳定，以创造提升幸福感的基本条件。

5. 建立动态监测与反馈机制

幸福感不是静态的指标，而是随时间与社会结构演变的主观评价。因此，建议各国建立长期幸福感监测系统，结合民调、大数据与心理指标，动态调整社会发展优先级，增强公共政策对人民福祉的敏感性与响应力。