

焦虑症、抑郁症的数据分析可视化报告

2023202128 沈逸尘

一、问题引入

21 世纪初期，全球范围内出现了大量精神健康相关疾病，且覆盖面大，流行范围广。据我国科研人员于 2016 年发表在 *The Lancet* 的调查结果和 WHO 发布的调查报告，2013~2015 年的精神疾病的发病率有显著上升的趋势，我国 15 岁以上人口中，各类精神疾病患者人数超过 1 亿人，而大型医院近年的抑郁症门诊就诊率每年 20% 的递增。

除了流行广泛，精神疾病也以其病因神秘、难治疗，症状复杂，可能导致的后果严重让人们恐惧。根据一篇发表在 *JAMA PSYCHIATRY* 的研究，精神疾病患者的死亡率明显高于对照人群，潜在寿命损失的中位数年数为 10 年。同时据估计，全球 14.3% 的死亡，即每年约 800 万例死亡，可归因于精神疾病。

基于这样的背景，对于精神疾病特别是最常见的焦虑症、抑郁症这类疾病发病的趋势、致病因素、具体表现和应对措施的调研有着极大的意义，本研究也试图回答在焦虑、抑郁背后到底发生了什么。

二、数据集介绍

本次数据分析的数据集主要有四个，数据来源均为 Kaggle 平台，分别存放在 ./data 路径下的四个文件夹。

Factors 文件夹下存放 95 位患有精神疾病的患者的患病原因，包括患者年龄、性别造成精神问题的诱发事件全貌、事件概述和对于该因素的抽象分类、心理学分类。获取链接：<https://www.kaggle.com/datasets/mdismielhossenabir/psychosocial-mental-health-analysis>

Message 文件夹下存放 6895 条患有抑郁症、焦虑症的患者的聊天纪录文件。获取链接：<https://www.kaggle.com/datasets/sahasourav17/students-anxiety-and-depression-dataset>。同时，该文件夹下还存放着用于文本分析的停用词数据集，包括哈工大停用词表、百度停用词表、四川大学机器智能实验室停用词库，获取链接：<https://github.com/endNone/stopwords>。

Prevalence 文件夹下存放 1990-2019 年间各国焦虑症的患病率，包括分不同年龄段的患病率、国家 gdp、标准年龄下不同性别患病率对比等数据。源数据采集自 Our world in data。获取链接：<https://www.kaggle.com/datasets/jaffidantonio/prevalence-of-anxiety-disorders-1990-to-2019?resource=download>。

Solutions 文件夹下存放 2016 年各国在精神健康方面的医疗资源配置情况，包括医院等专门机构情况，医疗、护理、科研人员等的配备情况两个分表，来自一个精神疾病与自杀的数据集，获取链接：<https://www.kaggle.com/datasets/twinkle0705/mental-health-and-suicide-rates?select=Facilities.csv>。

注：本项目所有代码文件、生成文件、数据文件以及报告文件均已在 Github 上传，链接：https://github.com/YichenShen0103/py_dv。

三、数据的分析处理以及可视化

模块一 焦虑症的患病率影响因素

本模块主要借助 Python 的 Pandas、Scipy 包进行数据处理工作，利用 Pyecharts 包（版本 1.8.1）进行可视化。

首先对于 age 分表，截取中国和美国的数据，并对各个年龄段的患病率利用 applymap 方法进行保留两位小数的格式化操作。利用 Pyecharts 的 Bar 类创建一个垂直柱状图对象，并调用 add_xaxis 方法将年龄数据导入，利用 add_yaxis 方法将两国患病率导入为两个 stack，并设置标签和标题。利用 render 方法将结果输出为 age.html 文件，并将结果保存在路径 ./result 下。

接下来处理 gender 分表，同样截取国家为中国和美国的数据，并对数据进行保留两位小数的格式化操作。利用 Pyecharts 的 Scatter 类创建一个散点图对象，以国家为横坐标，以男女性的患病率作为两组纵坐标数据，并设置横纵坐标轴标签、标题，并设置散点的大小来表示数据的绝对大小，利用 render 方法将结果输出为 gender.html 文件，同样保存在 ./result 路径下。

对于国家 GDP 分表，首先去除空项，然后截取 2019 年的数据。利用 Pyecharts 的 Map 类创建一个地图对象，将焦虑患病率作为数据填入地图制作热力图，并设置最大值和最小值、标签、标题等，利用 render 方法将结果输出为 map.html 文件，保存在路径 ./result

下。利用 `scipy.stats` 的 `ttest_ind` 方法对焦虑症患病率和国家 `gdp` 进行相关性检验。为了可视化两者之间的关系，利用 `Pyecharts` 的 `Scatter` 类创建一个散点图对象，将焦虑患病率作为 `y` 轴，`GDP` 作为 `x` 轴创建一幅散点图，利用 `render` 方法将结果输出为 `gdp.html` 文件，保存在 `./result` 路径下。

模块二 焦虑抑郁患者的聊天记录文件的文本分析

本模块的目的是找出所有聊天记录中的包含消极情绪的文本数据制作词云图。主要借助 `Python` 的 `Pandas`、`Nltk`、`TextBlob` 包进行数据处理工作，利用 `Wordcloud` 包（版本 1.8.1）进行可视化。

首先利用代码从 `Nltk` 包的官网下载 `vader` 库的情感词典。对数据进行初步清洗，即转化为小写，去除非英文字符和空项后，利用 `SentimentIntensityAnalyzer` 类创建 `Vader` 情感分析器。对于数据集中的每一条聊天记录，分别用 `vader` 分类器和 `TextBlob` 包的情感分析工具分别进行情感分析，得到 `vader` 综合得分，`TextBlob` 极性、主观性三组数据，并筛选出 `vader` 综合得分和 `TextBlob` 极性和 `vader` 综合得分低于 0 的数据，并存入在 `./data/messages` 目录下新创建的 `messages.txt` 文件中。

读入 `messages.txt` 以及停用词文件，利用 `Wordcloud` 包的 `WordCloud` 类创建一个词云图对象并设置停用词为读入的文件，利用 `messages.txt` 生成一幅词云图，存入 `./result` 路径下的 `wordcloud.png` 中。

模块三 导致精神疾病的因素分析

本模块主要借助 `Python` 的 `Pandas` 包进行数据处理工作，利用 `Pyecharts` 包（版本 1.8.1）进行可视化。

首先读入数据，进行数据清洗，利用 `pd.str.strip` 方法去除所有数据首尾的空格，并小写，去除空项。

根据不同的造成心理疾病的问题分类，统计各个问题分类的占比。排序后以元组方式存入列表中。利用 `Pyecharts` 的 `Pie` 类创建一个饼图对象，导入列表中的数据，生成反映各个因素导致精神疾病情况的占比图，将结果存放在 `./result` 路径下的 `factors.html` 文件中。

模块四 各国对于精神疾病的防治资源数据分析

本模块主要借助 Python 的 Pandas 包进行数据处理工作，利用 Pyecharts 包（版本 1.8.1）进行可视化。

读入两张数据表后，分别对两张表进行数据清洗，去除空项。对于包含各国精神医学医疗设施的表格，分别关于专门精神医学医疗中心和综合医院精神医学中心进行排序；对于包含各国精神医学人力资源的表格，分别关于精神科医生和专门护理人员进行排序。对于四个结果表截取前 10 名，利用 Pyecharts 的 Bar 类创建四个水平条形图对象，以国家为纵坐标，对应数据为横坐标绘制水平条形图，并利用 Tab 类创建一个用于切换不同图表的按钮组件，将四张表组合进一个图表中，将结果存入 ./result 目录下的 solution.html 文件中。

四、数据分析的故事化

“昨天我感觉我担心的要死了，我的生活已经僵硬了，一遍又一遍”

“我像是中毒了一样，每天晚上都害怕忘记了什么。”

“这让我感到头疼，就像是受伤了一样。”

“我感觉好难过，就像心碎了一样，但我甚至找不到原因！”

“哦，我的天哪！”

这些文字看上去有些荒诞，但这却是美国一个真实的焦虑症患者和他最亲密的朋友的聊天记录，这里记录着他的心声。

焦虑，或者扩大了说，心理健康问题，已经离我们每一个人越来越近。文明助力人类建起了规模空前的大型城市，恢弘的摩天大楼，然而——

高塔下的阴影里，光明未曾到达

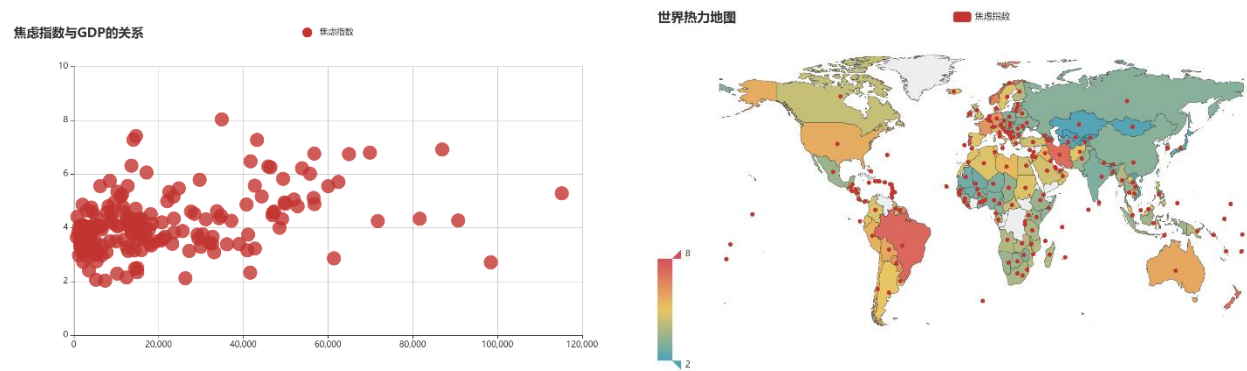
一篇发表在医学界最具影响力的期刊——《柳叶刀》上的调查显示，随着经济的发展，步入二十一世纪后全球精神疾病的发病率有显著上升的趋势，其中仅看中国的精神

疾病终身发病率就达到了为 **13.2%**,按这个比例来推算,我国 15 岁以上人口中,各类精神疾病患者人数已经超过了 **1 亿人**。同时,美国医学会的研究表明,精神疾病患者的死亡率明显高于普通人,潜在的寿命损失达到了 **10 年**。同时据估计,全球 **14.3%**的死亡,即每年约 **800 万例**死亡,可归因于精神疾病。

精神障碍就像来自我们无法揣测的世界的魔鬼,人类越是企图摆脱来自物质的拘束,他就越是贪婪的侵蚀,随时准备吞没每一个越来越渺小的人,犹如**等价交换**般,注定了你我作为祭品的命运,他说:

(一) 你只可到这里, 不能越过¹

那么,精神疾病的发病真的有那么神秘吗?或许也只有从数据中,我们能窥得一些奥秘。

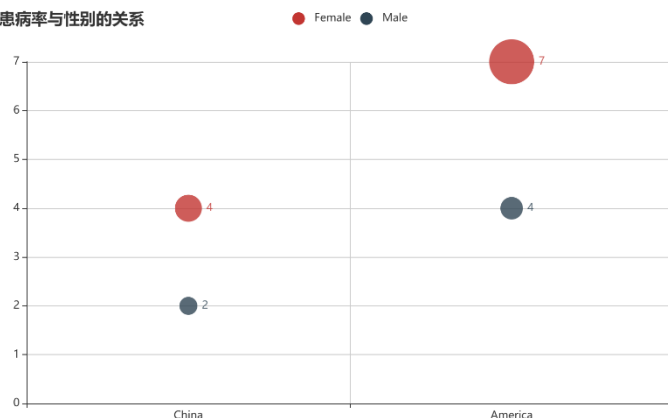


右边的这一幅是全球焦虑症患病率的热力地图,不难发现,焦虑症的患病率似乎和经济发展有着千丝万缕的关系,越发达的地区,象征着焦虑和抑郁的红色就越是频繁的出现。欧洲、美国、巴西、阿拉伯国家、澳洲,这些往往被我们称为**经济发达地区的国家却成了精神疾病的肆虐地**。而透过左侧的这幅散点图,这种关联被更直观的展现出来。

如果说 GDP 和焦虑症发病率的关系没有那么明显,那么性别与焦虑症发病率的关系显得更加神秘而又值得深思。在下面这幅图中,代表女性焦虑症发病率的红点比代表男性发病率的灰黑色点高出了不止一个档次。作为东西方世界的代表国家,无论在中国

¹ 引自《圣经·旧约·约伯记》

焦虑症患病率与性别的关系

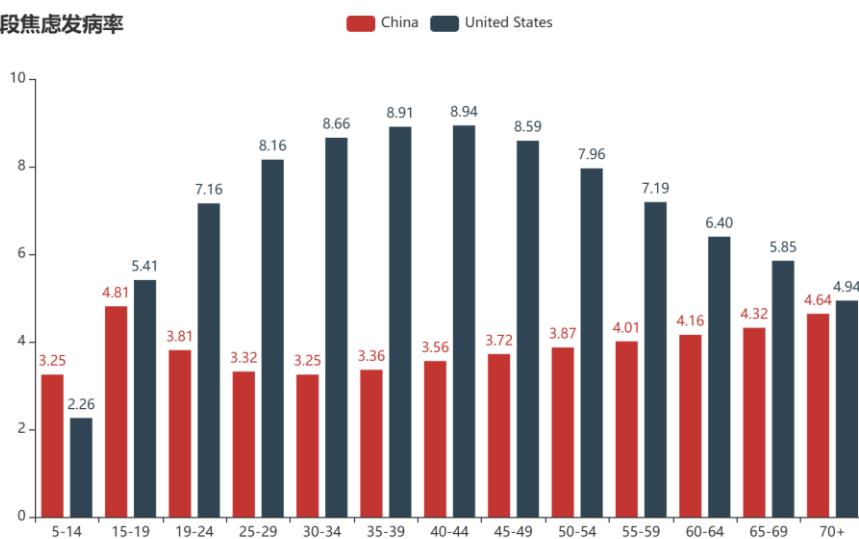


还是在美国,女性的焦虑症发病率似乎都比男性高出接近一倍。一项由北京大学第六医院等 44 个单位,历经 3 年多的时间完成的调研项目也显示,女性的抑郁症发病率显著高于男性。对此,发表在

《神经医学前沿》期刊上的一项科研指出,从一般生物学角度,男女性之间的基因表达模式、神经解剖结构和神经可塑性以及免疫特征之间的差别都导致了女性更容易患精神疾病,特别是严重精神疾病;另外女性围产期的独特心理周期也是重要的因素。

除开 GDP 和性别之外,年龄也是重要的影响精神疾病发病率的因素。在下图中展示的便是中美两国各个年龄段焦虑症发病率。

各年龄段焦虑发病率



可以看到,不同的国家集中发病的年龄段也有所不同,中国焦虑发病率在青少年时期多发,随着 30 岁后年龄的增长再次呈现上升趋势;美国焦虑症发病率则在年龄分布上呈现随年龄先上升后下降的趋势,在 40 岁左右达到最大值。这种国家间的差异,年龄间的差异往往难以用生物学或者医学解释,学界认为这可能与各文化之间差异有关。

(二) 那些溺死的星辰所告诉我的

还记开篇提到的那位焦虑症患者的聊天记录吗,如果我说你所看到的只是 **3000 多条**聊天记录中的一部分呢?如果我说你所看到的 **3000 多条**聊天记录也只是**为数不多**的几位精神障碍患者的一部分聊天记录呢

我们难以想象精神疾病患者在经历了怎样窒息般的煎熬,他们的人生本该如天上的星星一般璀璨。

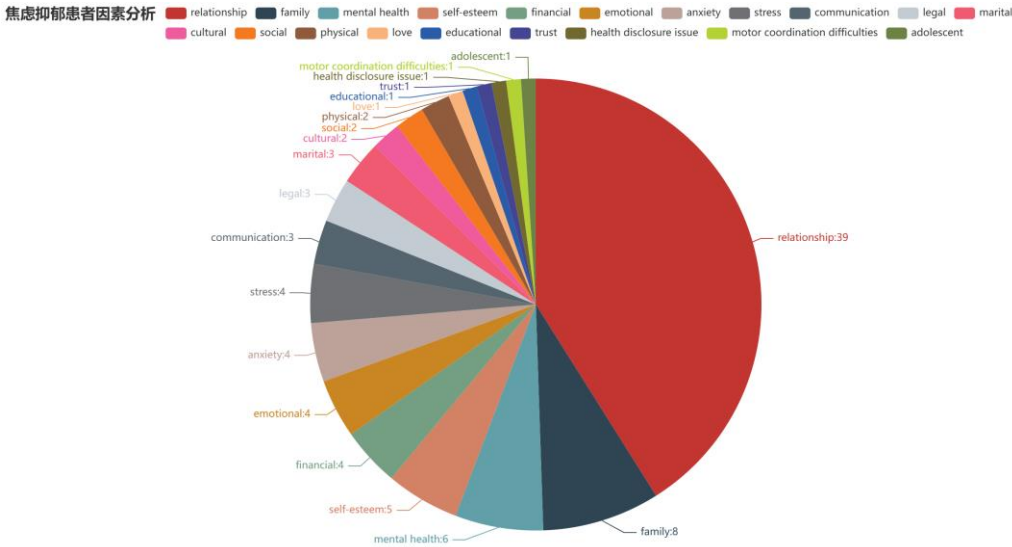


这是用他们的聊天记录所制成的一幅词云图,映入眼帘的便是“失眠”、“累”、“焦虑”、“害怕”、“时间”、“人们”等字样,而这些零散的词汇背后,却是每一个患者真真切切经历过的**恐惧与挣扎**。

或许有人发现了,这些词语背后都隐藏着一个共同的主旋律——**压力**,而这似乎就成了解释一切的关键。为什么经济越发达的地方的人们越焦虑?为什么女性比男性有着更高的患病几率?为什么青少年和中年的精神状态比其他年龄区间更加脆弱?归根结底,是这些群体特有的压力。**工作上的压力、生活中的压力、家庭里的压力、学业压力、上有老下有小的压力,甚至是来自文化和整个社会的压力,他们并不是不够坚强,只是承受的太多。**

我们找到了一份记录了 95 位专业医疗机构确诊精神障碍和精神疾病的患者自述的自己的病因的数据集。或许从这份数据中,我们能更直观的看到,这些特别的人们到底在承受这什么样的痛苦。

从下面这幅饼图中不难看出,大多数心理问题和精神问题产生的原因来自于**各种社会关系、家庭和经济**,当然我们也看到了一些少数,比如自尊心、青春期、健康状况等等。



(三) 所以，不要为明天忧虑²

2013 年 5 月,第六十六届世界卫生大会通过了世界卫生组织的《**2013-2020 年精神卫生综合行动计划**》,强调了要在社区环境中提供全面、综合的精神卫生和社会照护服务。这被认为是全球范围内减少精神疾病的重要计划,并提出了一系列举措。

庆幸的是,截止目前,许多国家都在积极响应这一号召。

2022 年 4 月 27 日,中国国务院办公厅印发了《“十四五”国民健康规划》明确提出要全方位干预健康问题和影响因素,完善心理健康和精神卫生服务,并提出了精神健康发展目标——到 2025 年,心理相关疾病发生的上升趋势减缓,严重精神障碍得到有效控制。

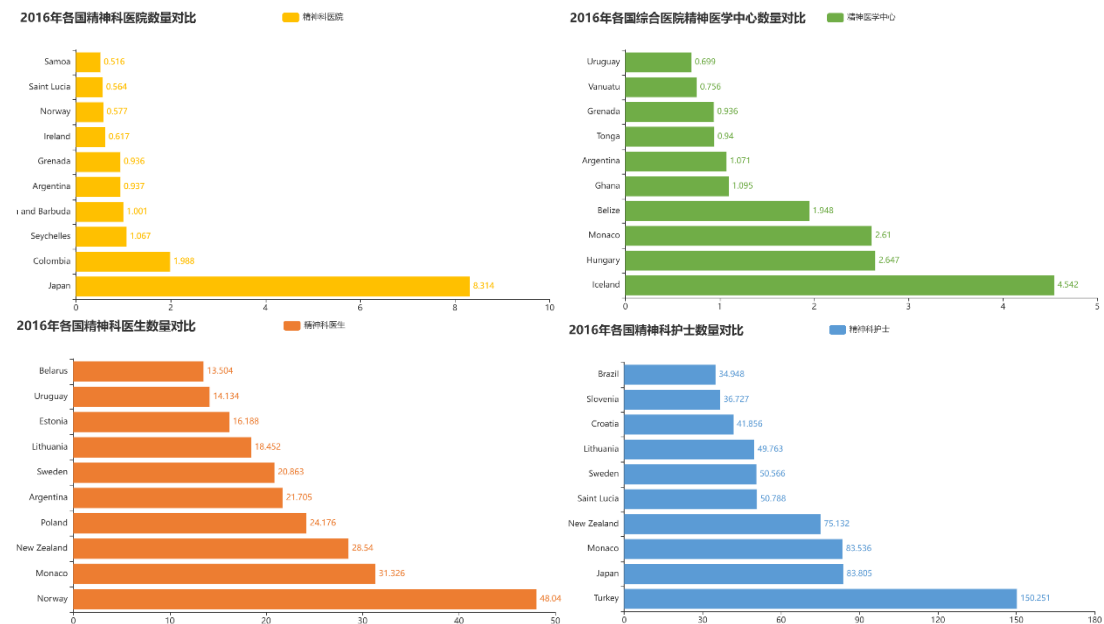
英国国家健康服务平台(NHS)发布了《长期规划》心理健康五年前景计划》等文件,并专门建立了心理健康服务平台和热线电话。

美国卫生与公众服务部发布了《2017-2022 年度国家精神卫生计划》,该文件概述了美国在心理健康服务、预防、治疗和研究方面的优先事项和目标。同时国家精神卫生研究所(NIMH)也发布了一些关于心理健康研究的战略计划。

² 引自《圣经·新约·马太福音》

.....

从以下几幅图中,我们也可以对目前各国的精神卫生和心理健康保障措施的实施现状窥见一斑。



随着全世界范围内的不断发展,对精神疾病的**重视**只会越来越强;随着心理问题越来越称为被**正视**的问题,对于这些神秘的疾病的恐惧和羞耻也会日益**消弭**;随着我们的**预防和保护措施**越来越完善,越来越多正在坠落的星星会被我们携手**托举**起来。

我们也有理由相信,在开篇所提到的那一位患有焦虑症和抑郁症的患者,能够在不远的未来,获取到真正便利而又普及的医疗服务,重新拥抱一个美好的人生。

所以...不必为明天忧虑!

五、参考文献

- E.R. Walker, R.E. McGee, B.G. Druss. (2015). Mortality in Mental Disorders and Global Disease Burden Implications A Systematic Review and Meta-analysis. *JAMA PSYCHIATRY*, (72), 334-341.
- R. C. Kessler et al. (2009). The global burden of mental disorders: An update from the WHO World Mental Health (WMH) Surveys. *EPIDEMIOLOGIA E PSICHIATRIA SOCIALE-AN INTERNATIONAL JOURNAL FOR EPIDEMIOLOGY AND PSYCHIATRIC SCIENCES*, (18), 23-33.

- Yueqin Huang et al. (2019). Prevalence of mental disorders in China: a cross-sectional epidemiological study. *LANCET PSYCHIATRY*, (6), 211-224.
- R. S. Eid, A. R. Gobinath, L. A. M. Galea, (2019). Sex differences in depression: Insights from clinical and preclinical studies. *PROGRESS IN NEUROBIOLOGY*, (176), 86-102.

注：本项目所有源代码文件、原数据集、中间文件以及最终报告均以在 Github 上传，链接：https://github.com/YichenShen0103/py_dv