

# 基于Tableau的心脏病的预测与预防的数据可视化交互系统

朱琳, 赖泳孜, 刘欢, 刘昱汝, 郭艳芬, 李庆, 朱烨, 符颖

成都信息工程大学计算机学院, 四川 成都

收稿日期: 2022年7月10日; 录用日期: 2022年8月10日; 发布日期: 2022年8月16日

## 摘要

目前心脏病具有发病率高、致残率高、死亡率高、复发率高、并发症多等特点, 已经给人们生活带来诸多不便。本交互系统根据已有的心脏病相关数据, 基于Tableau的可视化分析软件, 利用数据可视化的表现方式, 对患者的相关病史、不良嗜好、体征数据等进行筛选处理, 分析与心脏病有关的因素及不同因素对于心脏病患病情况的影响, 为非专业人士和专业人士提供更加直观的数据分析以及疾病预防与治疗的相关建议。本实验结果表明生活习惯、饮食习惯、运动习惯等后天行为的好坏对于心脏病的影响较大, 建议人们在生活中健康饮食、规律作息、坚持锻炼。

## 关键词

心脏病, 可视化, 数据分析, 人机交互

# Data Visualization Interactive System for Prediction and Prevention of Heart Disease Based on Tableau

Lin Zhu, Yongzi Lai, Huan Liu, Yuru Liu, Yanfen Guo, Qing Li, Ye Zhu, Ying Fu

College of Computer Science & Technology, Chengdu University of Information Technology, Chengdu Sichuan

Received: Jul. 10<sup>th</sup>, 2022; accepted: Aug. 10<sup>th</sup>, 2022; published: Aug. 16<sup>th</sup>, 2022

## Abstract

At present, heart disease has the characteristics of high morbidity, high disability rate, high mortality rate, high recurrence rate and many complications, which has brought a lot of inconvenience to people. According to the existing data related to heart disease, this interactive system is based on Tableau's visual analysis software, and makes use of data visualization to screen and process

the related medical history, bad habits and physical signs of patients, analyzes the factors related to heart disease and the influence of different factors on heart disease, and provides more intuitive data analysis and relevant suggestions on disease prevention and treatment for non-professionals and professionals. The experimental results show that the acquired behaviors, such as living habits, eating habits and exercise habits, have a great influence on heart disease. It is suggested that people should have a healthy diet, work and rest regularly and keep exercising.

## Keywords

Heart Disease, Visualization, Data Analysis, Human-Computer Interaction

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着近年来我国医疗水平在逐步提升，人们对医疗疾病的关注度也在增高。2020年《国家卫生服务调查》数据显示，心脏病作为第三大慢性病，致死率仅低于恶性肿瘤及脑血管病。心脏病的发病率和患病率逐年上升，发病年龄也在提前[1]。目前国内心脏健康领域的专业化新闻报道才刚刚起步，可供参考的文献数据较少。高血压、高胆固醇和吸烟是导致心脏病的三大主要风险因素，据美国疾病控制与预防中心的数据统计，大约有一半的患者在患病前至少其中一种风险因素。其他引发心脏病的关键指标包括糖尿病状况、肥胖(BMI 值高)、缺乏体育活动或饮酒过多等。及时发现和提前预防对心脏病的影响因素对人们的医疗保健起到至关重要的作用。

数据可视化交互系统作为新兴数字媒体产业的一项衍生产物，能够提高用户间沟通效率、加快数据分析速度、更好地从结果追溯原因，从而更精准地呈现问题[2]。由于导致患心脏病的可能性因子较多，本文利用 Tableau 数据可视化工具对引起心脏病的各项因素进行可视化研究和统计分析，针对可能引发心脏病的各项指标数据生成直观的可视化图表，可以更好地帮助医护人员及民众了解到心脏病与影响因素的关系，对心脏病的预防和治疗起到帮助。

## 2. 数据可视化与技术

### 2.1. 可视化方法

数据可视化就是将复杂的问题简单化，将枯燥繁杂的数据转化为具有人性色彩的图表，便于更清晰直观地数据展示和业务分析。大数据体量巨大，来源众多格式复杂，使得汇总分析数据异常艰难，而将数据进行可视化的呈现，可以达到简洁直观，容易理解和记忆，传递更丰富的信息等作用。数据可视化旨在让用户在最短的时间内了解到数据呈现的信息，更好地进行数据的理解分析，方法主要包括明确目的、选择图表、视觉设计和突出信息四个步骤。设计要从需求和目标出发，不要盲目追求图表种类和美观，实用性要大于美观[3]。

数据可视化的工具众多，有 Echarts [4]、Tableau [5]、CiteSpace [6]、Highcharts [7]等软件都应用广泛，本文的可视化过程选择使用 Tableau 工具进行。

### 2.2. Tableau 工具简介

Tableau 是一个可视化的分析平台，提供了从数据连接到整体协作的完整功能，汇集数据，辅助人们

查看并快速分析数据。用户进行数据的连接后,通过简单的拖放式操作,无需编程,即可创建美观的交互图表、地图、仪表板和故事报告等,从而将繁杂的数据通过可视化呈现出来,方便进行深入分析,最后可以将可视化发布到社区,进行线上共享,可以获得帮助及有用的反馈。Tableau 具有快速分析、适用各种数据、自动更新、操作简单、智能仪表盘、实时共享等特点,功能强大,包括多个软件,都有各自的优点,本文进行的数据可视化分析主要使用了 Tableau Public 和 Tableau Desktop。

Tableau Desktop 是一款强大的可视化软件,提供了对数据源访问,可视化到分析所需要的全部功能,可以实时探索各种数据。利用简便的拖放式界面,自定义视图、页面、布局、标记、筛选器等,数分钟内生成美观的可视化交互图表。Tableau Public 是一个免费的软件,用户可以在线探索创建,共享数据可视化,其界面操作与 Desktop 大致相同,只是在一些细致的功能上有所缺失。Public 拥有规模极大的可视化数据库,世界各地数以万计的用户通过 Public 分享交流可视化数据,创建免费的个人资料,可以将可视化保存到个人资料中,实现 Web 上任何位置的在线共享,方便团队协作。

### 3. 心脏病预测与预防的数据可视化交互

#### 3.1. 需求分析

据美国心脏协会调查显示,2020 年全球约有 1900 万人死于心血管疾病[8]。根据《中国心血管健康与疾病报告 2019》显示[9],我国目前患心血管病人达 3.3 亿,心血管患病率持续上升,心血管病死亡率居于全国首位,其中每 5 例因病死亡病例中有 2 例是死于心血管疾病。心血管病作为世界三大疾病之一,其相应的发现和预防措施应该得到人们的关注和重视。

心血管疾病相关文本数据体量巨大、内容冗杂,数据可视化可将文本数据转化为图表的形式进行直观地表达,使得重点信息更加突出、主次分明。同时,数据可视化可以和用户进行更深层次的互动,让用户选择自己更感兴趣的信息,从而更自由、更灵活地输出信息。数据可视化鲜明、形象的表达还可以让用户记忆深刻,更便于用户理解和记忆相关信息。

从医生的角度来看,将心血管疾病数据进行数据可视化,可将冗杂的数据进行归类、分析,帮助医生迅速地、直观地获得相关信息,完成其相应的工作和研究。从患者的角度来看,将心血管疾病数据进行数据可视化,可以帮助患者充分认识自己患病成因,并帮助患者改正不良生活习惯,对患者病情的恢复有很大帮助。从其他未患病的普通用户的角度来看,将心血管疾病数据进行数据可视化,可以帮助用户认识到自身的不良生活习惯,并且通过自我观察和医院检查等手段,了解自身是否存在相关病症,从而及时发现和预防相关心血管疾病。

#### 3.2. 心脏病数据可视化研究思路及方法

心脏病数据可视化设计重点在于对导致患心脏病的客观因素进行分析,找到患心脏病的关键因素,帮助医生、患者或普通用户对数据进行理解,从而达到对心血管疾病的发现和预防。基于对患心脏病相关因素的分析,本文从以下几个方面探讨患心脏病因素的影响,以预测患者病情:1) 日常习惯(酗酒、抽烟、缺乏锻炼、缺乏睡眠)对患心脏病的影响;2) 过往病史(糖尿病、肾病、皮肤病、中风)和心脏病的关系;3) 年龄和心脏病的关系;4) 性别和心脏病的关系。

通过分析以上几种因素与心脏病成因的关系,本文拟采用五种可视化图表对心脏病数据进行呈现:桑基图、花瓣图、散点图、跑道图和玫瑰图[10]。

1) 桑基图。桑基图是一种特定类型的流程图,其最明显的特征是始末端的分支宽度总和相等,即所有主支宽度的总和应与所有分出去的分支宽度的总和相等,保持能量的平衡。

本文将各年龄段人群有病史的患者进行区分统计,采用桑基图来呈现,制图过程为:首先设置数据

桶, 将各类数据分组以便计算频数。其次分别设置三个图表, 其中两个图表分别按照各病症频数和各年龄段的占比对应列表, 并通过设置字段保持对称性再创建一个连接用曲线表。最后调整图表大小形成左右对应。

2) 花瓣图。花瓣图又称韦恩图, 是一种关系型图表, 通过图形之间的重叠来反映数据集之间的相交关系。对于数据集之间交叠关系的可视化, 花瓣图是表达数据的最好方式。

本文将数据集中是否由不良习惯诱发的心脏病数据制成花瓣图。制图过程为: 首先创建单个花朵的圆心和花瓣的坐标轴并设置象限位置。其次将通过筛选器筛选的数据导入, 定义花瓣和圆心的大小。再次, 按照年龄段分类创建多个花朵, 再设置以屏幕左上角为原点的二维坐标轴将花朵按年龄段递增的顺序从左至右、从上到下摆放。

3) 散点图。散点图能呈现数据点的分布, 表现两个元素的相关性; 同时也能像折线图一样表示时间推移下的数据发展趋势。其适用于展现两个或三个变量之间的关系; 数据量越大, 图像呈现的效果越好, 且能快速关注离群点(异常点)对整体的影响。

睡眠不足也是心脏病的一个诱因, 为了探索患病和睡眠时间的关系, 且男女性别因睡眠不足诱发的心脏病患者差异性, 本文采用散点图进行该部分数据可视化。制图过程为: 首先设置聚合, 计算睡眠不正常的人在各个年龄段的患病比例。其次, 将睡眠时间拖入筛选器, 便于表格的动态交互。最终将用户图标素材载入 Tableau 图库, 在标记中更改散点图形。

4) 跑道图。跑道图是柱状图基础上的一种变形, 二者区别在于, 跑道图不仅能看到不同指标的占比情况, 还可以直观看到指标在当前维度下的排名情况。

本文将有过往病史(糖尿病、肾病、皮肤病、中风)的患者做统计分析, 判断出病因与心脏病的关联, 制成跑道图。制作过程为: 首先, 建立链接表, 添加半径、角度等变量, 即由于跑道图的圆周属性, 每一个变量都应对应 360 行(创建角度字段: 每一个角度对应一行, 即 0 度~360 度, 跑道图共由 361 个点组成), 其次利用极坐标表示每个点对应的坐标, 最终调节图表形状属性。

5) 玫瑰图。玫瑰图, 又叫南丁格尔玫瑰图, 是将柱图转化为更美观饼图形式, 是极坐标化的柱图。不同于饼图用角度表现数值或占比, 南丁格尔玫瑰图使用扇形的半径表示数据的大小, 各扇形的角度则保持一致。对照饼图, 由于半径和面积的关系是平方的关系, 玫瑰图会将数据的比例大小夸大, 尤其适合对比大小相近的数值。

为了分析年龄对患病的可能性的影响, 本文使用玫瑰图来呈现。制图过程为: 通过创建多个计算字段计算角度、片数、X 和 Y 轴等, 将度量 Age Category 放到标记的颜色栏中以区分不同年龄, 将度量 Heart Disease 放在标记的颜色中方便观察不同年龄的心脏病患病占比, 将做好的表格转换为数据桶, 将表计算的的计算依据对应好, 即可得出不同年龄阶段患心脏病的占比分布。

## 4. 实验及结果分析

### 4.1. 数据来源

本文使用的数据集来自美国疾病控制与预防中心, 是行为风险因素监测系统(BRFSS)的主要组成部分, 该系统每年进行电话调查, 收集美国居民健康状况的数据。数据集包含患者年龄、性别、种族、日常习惯(是否酗酒、是否抽烟、睡眠时间等)、身体健康状况、既往病史(哮喘、糖尿病、心脏病、肾病、皮肤病等)等信息。本文将该数据集导入 Tableau 中, 通过对造成心脏病的可能因素的组合并设置计算字段得出分析图表, 数据来源可靠。

出于统计分析的考量, 本文在保留筛选后原变量的基础上加入了一些整合后的变量, 其中所有涉及到的变量定义如表 1 所示。

**Table 1.** Definition of variables in data analysis**表 1.** 数据分析中的变量定义

编号	变量	定义	编号	变量	定义
1	Age Category	具体年龄	11	Physical Activity	是否定期锻炼
2	Alcohol Drinking	是否酗酒	12	Race	种族
3	Asthma	是否患哮喘病	13	Sex	性别
4	Diabetic	是否患糖尿病	14	Skin Cancer	是否有皮肤病
5	DiffWalking	是否残疾	15	Smoking	是否抽烟
6	Gen Health	平时健康状况(以一个月天数计)	16	Stroke	是否中风
7	Heart Disease	是否患心脏病	17	BMI	身体质量指数
8	Kidney Disease	是否患肾病	18	Mental Health	心理健康情况
9	Physical Health	身体健康情况	19	Sleep Time	睡眠时间
10	Bad Habits	筛选整合值, 取编号 2、11、15 中为否者, 编号 19 小于 6 者	20	Age Category*	以年龄相差 10 为一组分组的年龄段

## 4.2. 数据清洗

数据清洗环节分为两个环节: 1) 第一遍先将所有数据都导入 Tableau Public, 将大量存在的残缺的、信息不全即存在缺省值(NA)的数据通过筛选器进行逐个相关变量筛选剔除。2) 第二遍对已经得出的数据图表进行集中数据分析, 将明显异常和无意义的的数据剔除[11]。

## 4.3. 可视化实现

### 4.3.1. 可视化图表

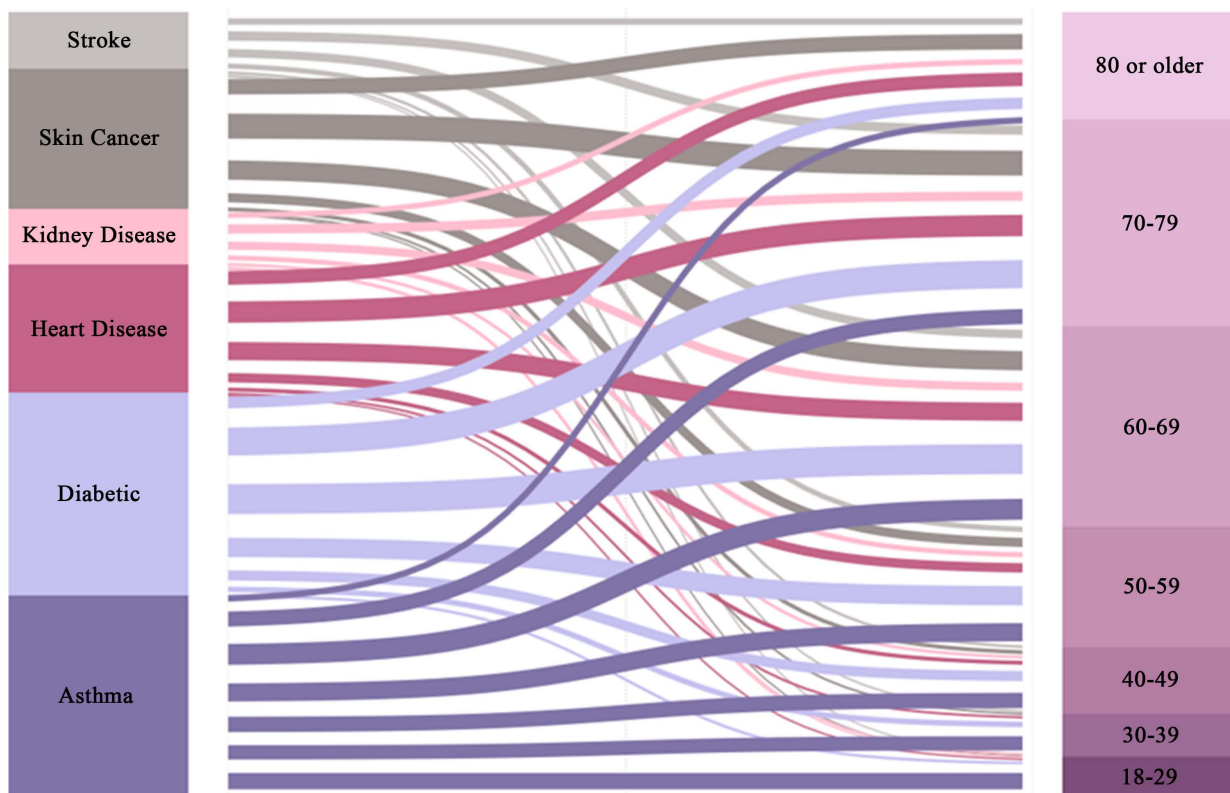
桑基图用来表示数据的流向, 本文将各年龄段人群有病史的患者采用桑基图进行区分统计, 可以更清晰的展现出各年龄段患病人数与各个病症的对应关系。对于分析各年龄阶段不同疾病患病人数流量分析如图 1 所示。

图 1 左侧为疾病类型, 右侧为各年龄段, 方便观察两侧数据类别的对应情况和数量关系。从图 1 可以看出, 患病人数排名前三的是糖尿病、哮喘病和心脏病, 心脏病的患病年龄主要集中在 50 岁以上的人群中, 其中 70~79 岁的人数占比最高, 其次是 60~69 岁人群, 在 80 岁以上的患病人数依然较多。采用桑基图, 更好地展示各年龄段人群患病情况的流量趋势。

花瓣图可以对数据集之间交叠关系进行可视化, 直观且美观。在数据类别上多种因素导致结果因素的分析时通过花瓣图可以更清晰地看到各阶段内的不同因素对圆心大小的影响程度。将患病人数和酗酒、抽烟等不良习惯从原始数据中抽取出来, 给年龄段进行分段处理得到一张新的数据表。图 2 所示为不良习惯对心脏病患病情况的影响。

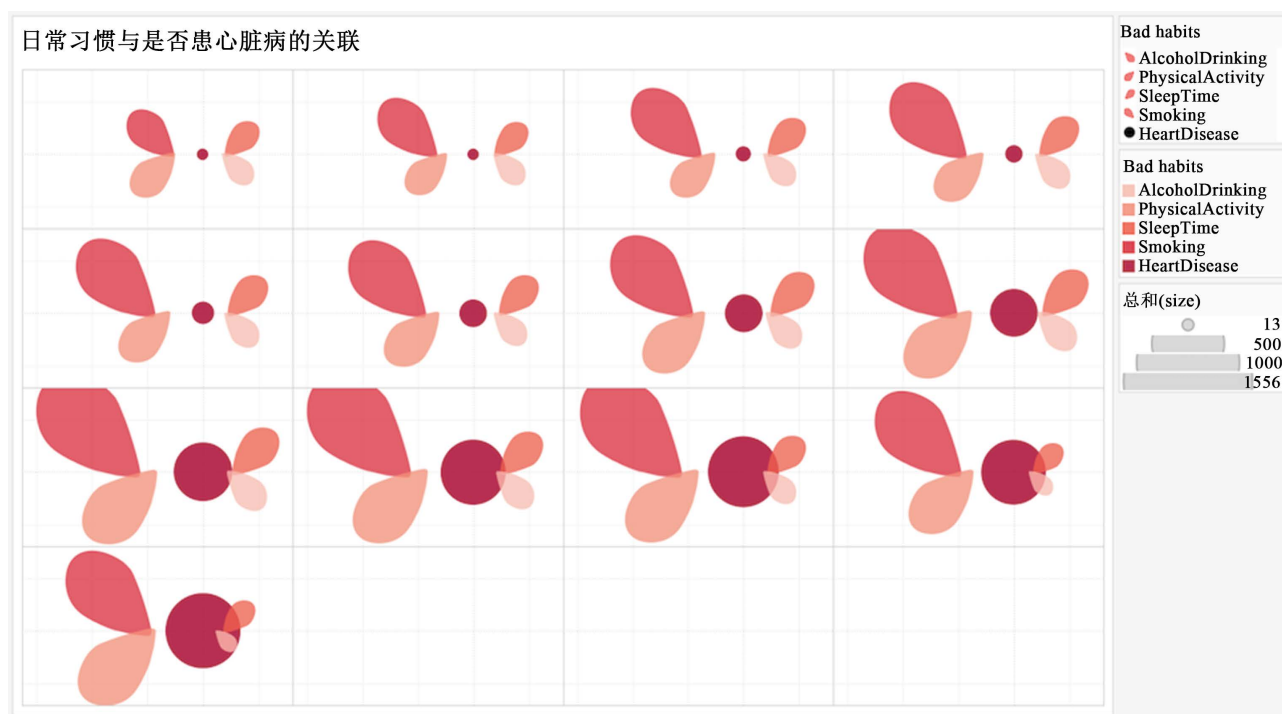
图 2 按各年龄段区分朵数, 圆心大小对应患心脏病人数, 四个花瓣则分别对应“是否抽烟”、“是否酗酒”、“是否睡眠不足”、“是否定期锻炼”的统计人数, 从左至右从上至下年龄段逐渐提升。从结果上可以看到心脏病患病人数随年龄增大逐渐增多, 各因素对心脏病的患病概率都有影响, 其中抽烟和缺乏锻炼是主要的致病因素。随着时间的变化, 酗酒和睡眠不足对心脏病的影响也在持续增加。

散点图可以清晰的描述分布, 为了探索患病和睡眠时间的关系, 本文选用散点图(如图 3 所示)呈现因睡眠不足对不同性别在各年龄段的患病情况, 且用不同颜色和形状区分性别, 使单一纵列对比清晰。从



**Figure 1.** Sankey chart is used to analyze the flow of patients with different diseases at different ages

**图 1.** 桑基图对于分析各年龄段不同疾病患病人数流量分析



**Figure 2.** Petal plots were used to analyze the impact of bad habits on heart disease

**图 2.** 花瓣图对于分析不良习惯对于心脏病患病情况的影响

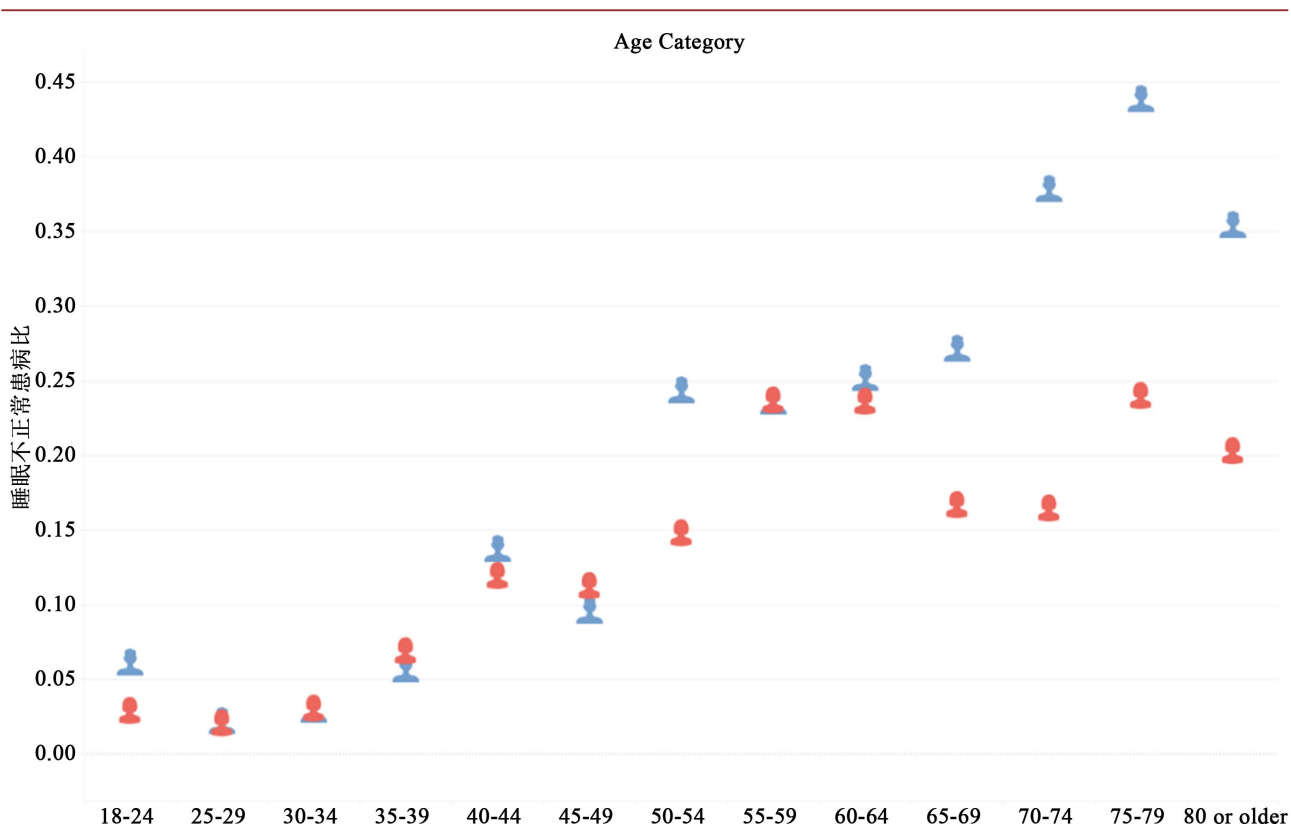


Figure 3. Analysis of the incidence of sleep deficiency in different genders at different ages

图 3. 睡眠不足对不同性别在各年龄段的患病情况分析

图 3 可以看出，随着年龄的增长，睡眠不足对心脏病的影响也越来越大；男女睡眠不足且患有心脏病的情况在 50 岁之前基本一致，50~54 岁以及 65 岁以上开始体现出较大的差异性；睡眠不足对于患心脏病的影响中，普遍男性患病几率大于女性，且在 75~79 岁阶段男女睡眠时间对心脏病的影响较大。

由于跑道图侧重于将数据按一定比例进行缩放处理，使得各个类别间的数量差别得以更清晰地呈现，本文主要分析四种患病(哮喘、皮肤病、糖尿病、肾病)对心脏病的影响程度，如图 4 所示。从图 4 中可以更清晰的比较不同数据间的差别，看到各因素对于心脏病的影响。其中，在多种并发症中肾病对患心脏病的影响最大，占患有心脏病群体的 29%，其次是糖尿病，占患有心脏病群体的 21%，皮肤病、哮喘分别占心脏病患者的 17%和 12%。

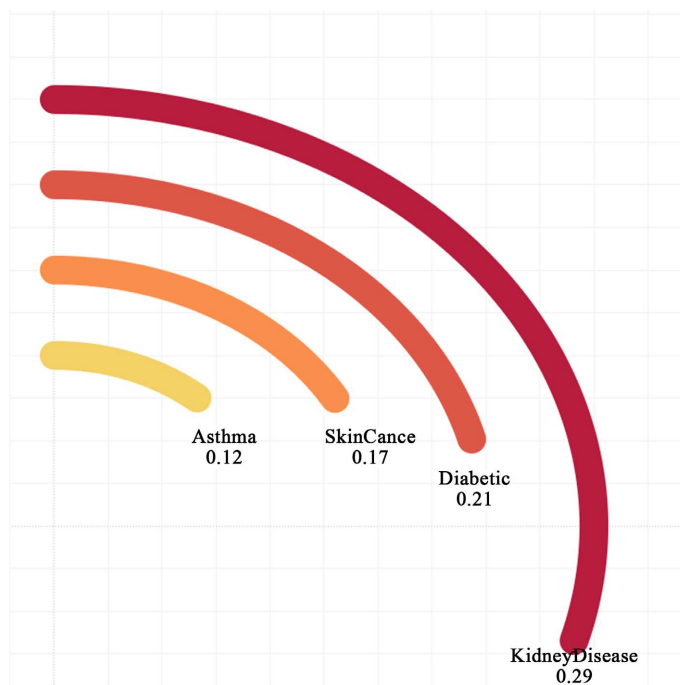
玫瑰图中每个扇形的弧长大小表示该种类占总体的比例，由于数据集中的指标较多，且各年龄占比之间差距较小，本文使用玫瑰图呈现不同年龄段患心脏病的占比情况，如图 5 所示。

从图 5 可以看出，随着年龄的增长，玫瑰图单片的弧长和面积有增大的趋势，说明年龄越大，患心脏病的人数占比也逐渐增大。因此，可以得出随着年龄的增长，人们患心脏病的风险增大。

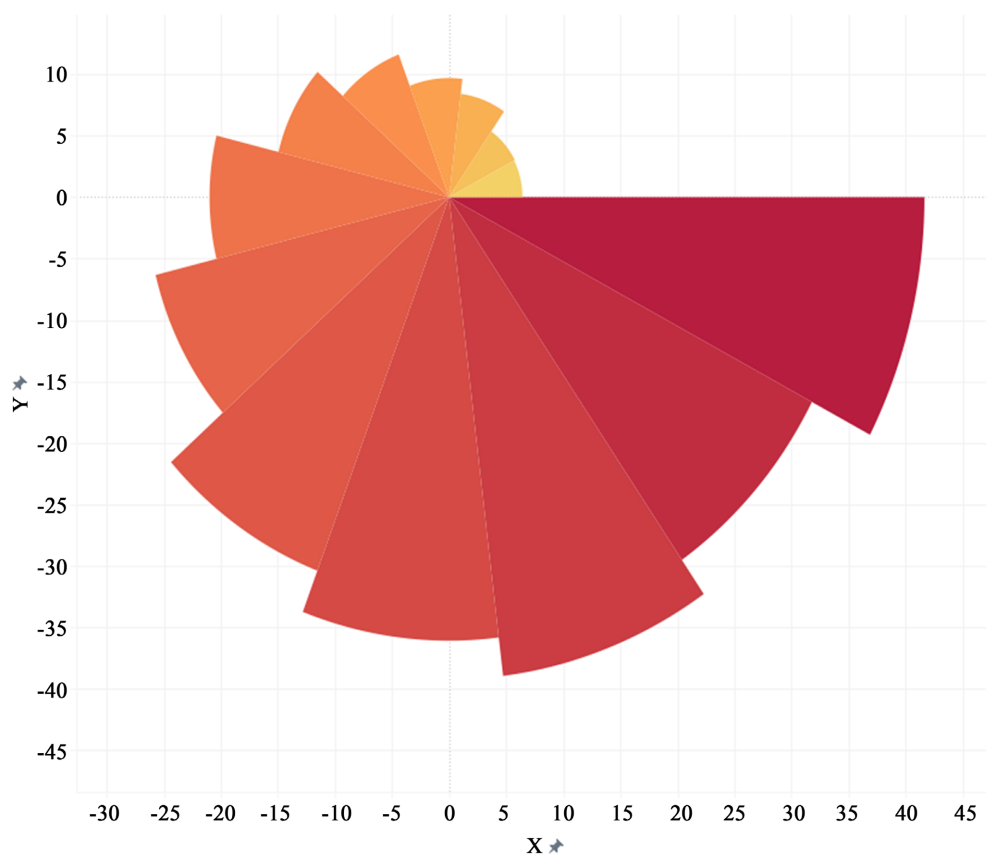
#### 4.3.2. 仪表盘的设计与制作

仪表板是若干图表视图的集合，方便同时比较各种数据，本文制作分析型仪表板，生成一个完整的数据可视化交互系统。仪表板的设计主要包括布局、字体色彩等几个方面[12]。仪表板的局部图如图 6 所示。

1) 确定框架，梳理层级结构，整体布局遵循突出重点的原则，大致以均等布局为主，平均分配每一块展示区域，用于展示平级分析数据，图标之间无明显主次时，这样均等布局在视觉上不会有过多地干



**Figure 4.** Analysis of the influence of other diseases (asthma, skin disease, diabetes, kidney disease) on heart disease  
**图 4.** 其他病症(哮喘、皮肤病、糖尿病、肾病)对心脏病的影响分析



**Figure 5.** Rose chart of analysis of the influence of age on the likelihood of disease  
**图 5.** 年龄对患病的可能性的影响分析玫瑰图



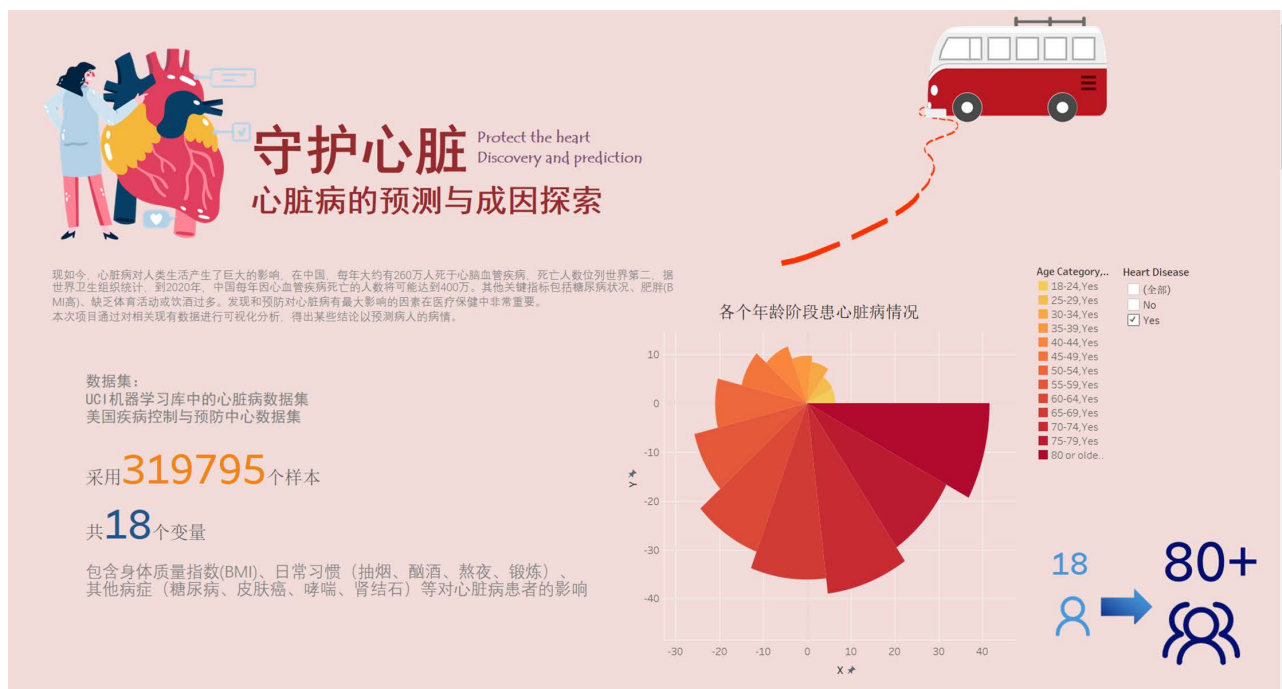


Figure 6. Local view of dashboard

图 6. 仪表板局部图

扰，所有指标都可找到，一目了然。

2) 字体设计主要从可识别性和风格融合度来考虑，主要使用默认字体，局部提炼关键数据作为先导，用加粗且大字号的字体来展示数据，配以不同的颜色来突出标识，让用户能更清晰明确数据和结论。

3) 色彩设计主要使用了自带的红色面板区中不同程度的红色，该色板契合“心脏病”的研究主题；蓝紫色为辅色进行局部搭配，起到突出数据的作用，使整体具有美观性又不突兀杂乱。单个元素的颜色和整体背景色系保持一致，协调的配色加上多样式的图表，使整个页面具有设计感，数据呈现更清晰丰富。

4) 点缀设计方面，适当地给标题、图表等添加诸如边框、按钮图像等点缀效果，在局部合理地插入图画用来配合描述数据信息或结论，使整体美观度上升不显单调，同时帮助更好区分布局，明确数据内容。

仪表板的制作过程添加了交互功能，可以让用户主动探索感兴趣的数据信息。使用筛选器，设置约束条件以实现数据的动态查询，在工作表的右上角，启用“用作筛选器”的选项，可以用一个工作表中的标记作为仪表板中其他有相同标记的工作表的筛选器。点击仪表板菜单中的操作，添加“转到工作表”动作，选择启动操作的源工作表，实现仪表板到另一个仪表板的导航，通过两个文字按钮来指定该操作。

## 5. 小结

现如今处于信息化时代，在享受科技便利的同时也给我们的生活带来了许多困扰，如过快的生活节奏使身体超负荷运行，冗长的信息列表中难以快速找到自己需要了解的数据结果等，而数据可视化的方式可以清晰快速地将数据动态直观地呈现在人们面前。本文以心脏病为例，针对心脏病患病的可能影响因素进行了分析，并讨论了可视化仪表盘的制作排版方式。分析发现，心脏病患病人数随年龄增大逐渐增多，酗酒、抽烟、睡眠不足、缺乏锻炼对心脏病的患病概率均有影响，其中睡眠不足对男性的心脏病

患病几率影响较大, 抽烟和缺乏锻炼是主要的致病因素, 心脏病患者同时患有肾病的占比最大等。

本文选用的数据集在年龄段变量中, 数据分布不均匀, 可能会影响数据分析结果, 在筛选数据的过程中, 既要保持数据的客观性又同时需要满足各个变量的数量统一性是我们今后要努力的地方[13]。为此, 本文在日后的研究中也更加重视数据清理过程, 为数据研究做出更大贡献。希望在数字化的未来, 繁杂的数据可以更多地图像化, 更多的用于医疗健康领域, 即使是非专业的用户也可以清晰快速地了解自己的身体状况从而调整自己的生活状态, 让更多的人拥有健壮的体质, 共创一个幸福的社会。

## 基金项目

四川省科技计划项目(重点研发项目) (2020YFG0442)。

四川省科技计划项目(重点研发项目) (2020YFG0453)。

四川省自然科学基金(青年) (2022NSFSC0907)。

## 参考文献

- [1] 白玲. 基于 Tableau 工具的医疗数据可视化分析[J]. 中国医院统计, 2018, 25(5): 399-401.
- [2] 刘佳. 网约车应用频率及其影响因素分析——基于武汉市数据可视化分析[J]. 艺术与设计(理论), 2022, 2(2): 104-106.
- [3] 丽姿·贝内特, 肖俊洪. 从学习的角度看学生对仪表盘数据的反应[J]. 中国远程教育, 2019(1): 67-78+93.
- [4] 彭曙光, 王梦梅, 赵麒博, 申刘宝, 彭玉杰, 陆雪艳. 面向 ECharts 的疫情信息可视化系统[J]. 福建电脑, 2022, 38(4): 80-83.
- [5] 于晓明, 史胜楠, 甘克勤. 基于 Tableau 的标准起草单位大数据可视化分析[J]. 标准科学, 2019(9): 80-84.
- [6] 沈慧金. 中国 STEM 教育的研究现状、热点与前沿的可视化研究——基于 CiteSpace 的文献计量可视化分析[J]. 电脑与信息技术, 2022, 30(3): 88-91.
- [7] 葛丽, 汪祎, 任思蓉, 孙悦红. 基于 ECharts 和 Highcharts 的进口风险工业品数据可视化[J]. 信息与电脑(理论版), 2021, 33(10): 27-29.
- [8] Tsao, C.W., Aday, A.W., Almarzooq, Z.I., *et al.* (2021) Heart Disease and Stroke Statistics—2022 Update: A Report from the American Heart Association. *Circulation*, **145**, e153-e639. <https://doi.org/10.1161/CIR.0000000000001052>
- [9] 国家心血管病中心. 中国心血管健康与疾病报告 2019 [M]. 北京: 科学出版社, 2020: 142-144.
- [10] 陈为, 赵焯, 张嵩, 鲁爱东. 可视化导论[M]. 北京: 高等教育出版社, 2020.
- [11] 周开宇, 李一飞, 沈建通, 等. 基于多元统计和社会网络分析的中国先天性心脏病介入诊断治疗学科知识可视化分析[J]. 中国循证儿科杂志, 2012, 7(1): 44-49.
- [12] 姚忆南. 数据可视化大屏 UI 设计的研究和分析[J]. 自动化仪表, 2021, 42(5): 107-110.
- [13] 慕欣. 心脏病的统计研究[J]. 中国新通信, 2018, 20(23): 227-229.