**数据报告**

目录

[一、 数据处理 2](#_Toc522364174)

[有效数据： 2](#_Toc522364175)

[记录完整性： 2](#_Toc522364176)

[二、 单项分析： 3](#_Toc522364177)

[性别 3](#_Toc522364178)

[年龄 3](#_Toc522364179)

[村庄分布 4](#_Toc522364180)

[空腹血糖（FBG） 5](#_Toc522364181)

[体质指标（BMI） 5](#_Toc522364182)

[慢性病 6](#_Toc522364183)

[三、 联合分析 7](#_Toc522364184)

[BMI与FBG相关性 7](#_Toc522364185)

[多个指标联合分析 8](#_Toc522364186)

[四、 性别差异健康患病分析 9](#_Toc522364187)

[脉率 9](#_Toc522364188)

[空腹血糖： 10](#_Toc522364189)

[呼吸 10](#_Toc522364190)

# 数据处理

## 有效数据：

导入数据库的过程中，发现存在大量重复记录。故在之后的数据分析过程中，排除同一人的相同报告。处理后：

diabetes 表格有效记录108383条

health 表格有效记录640000条

## 记录完整性：

|  |  |
| --- | --- |
| diabetes 记录 | health 记录 |
|  |  |

由雷达图可知，多个类别记录完整性不足50%（例如锻炼频率、尿蛋白等）甚至没有记录（例如随机血糖、糖化血红蛋白等）。考虑到大部分数据的缺乏可能会影响数据分析的正确性，故在之后的数据分析过程中不再对其进行分析。

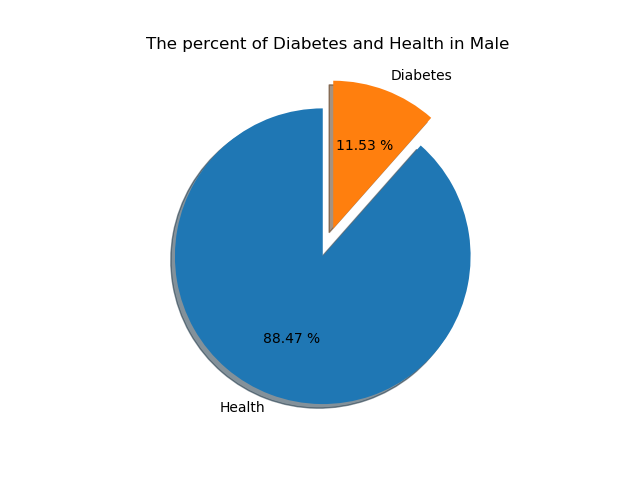
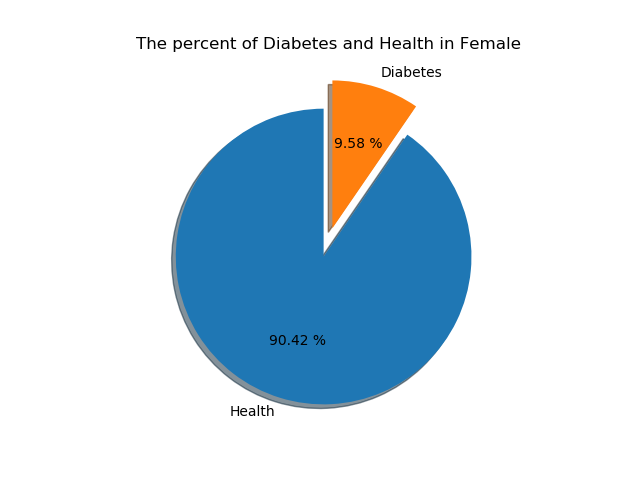
# 单项分析：

单项分析结果可以了解患病与健康在记录指标上的不同，以得到与患病有关的指标。

## 性别

男性与女性中的患病与健康比例分别如下

男性人群中，患病者比例明显高于女性人群

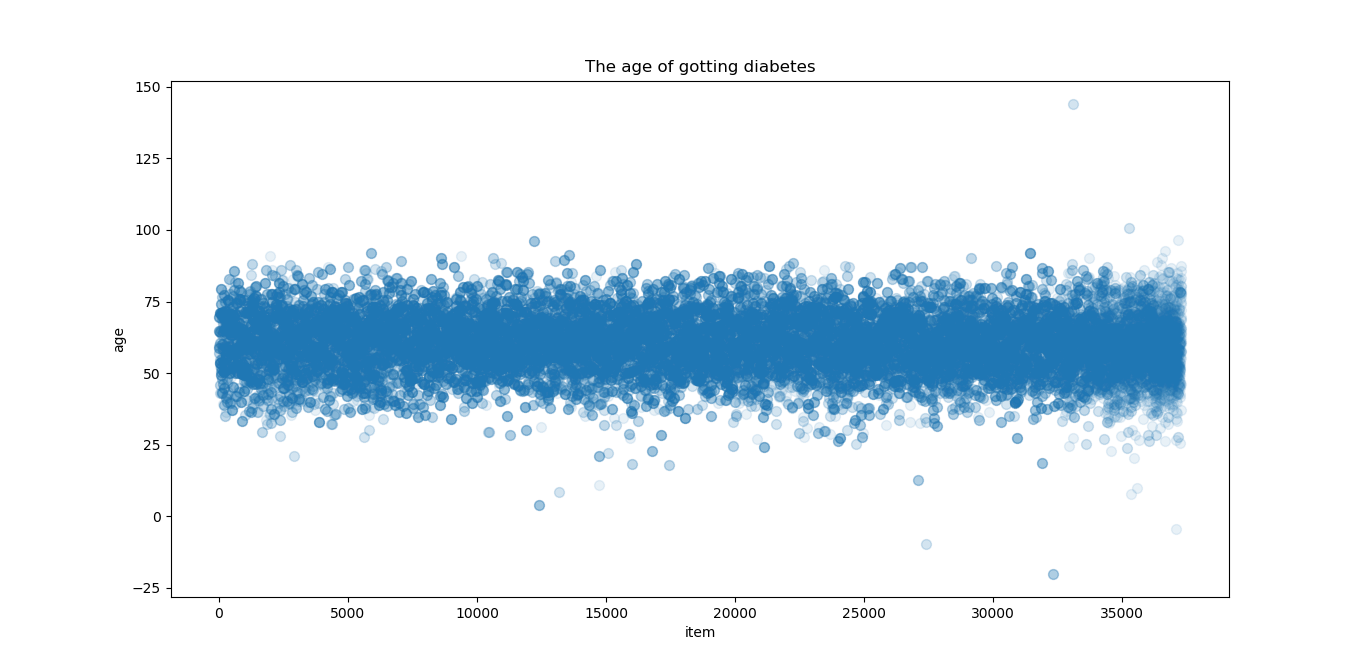
 

## 年龄

用发现时间 – 出生日期可以得出患病时年龄

可以看出患病年龄多在35-80年龄段，即步入中年要时刻关注身体指标

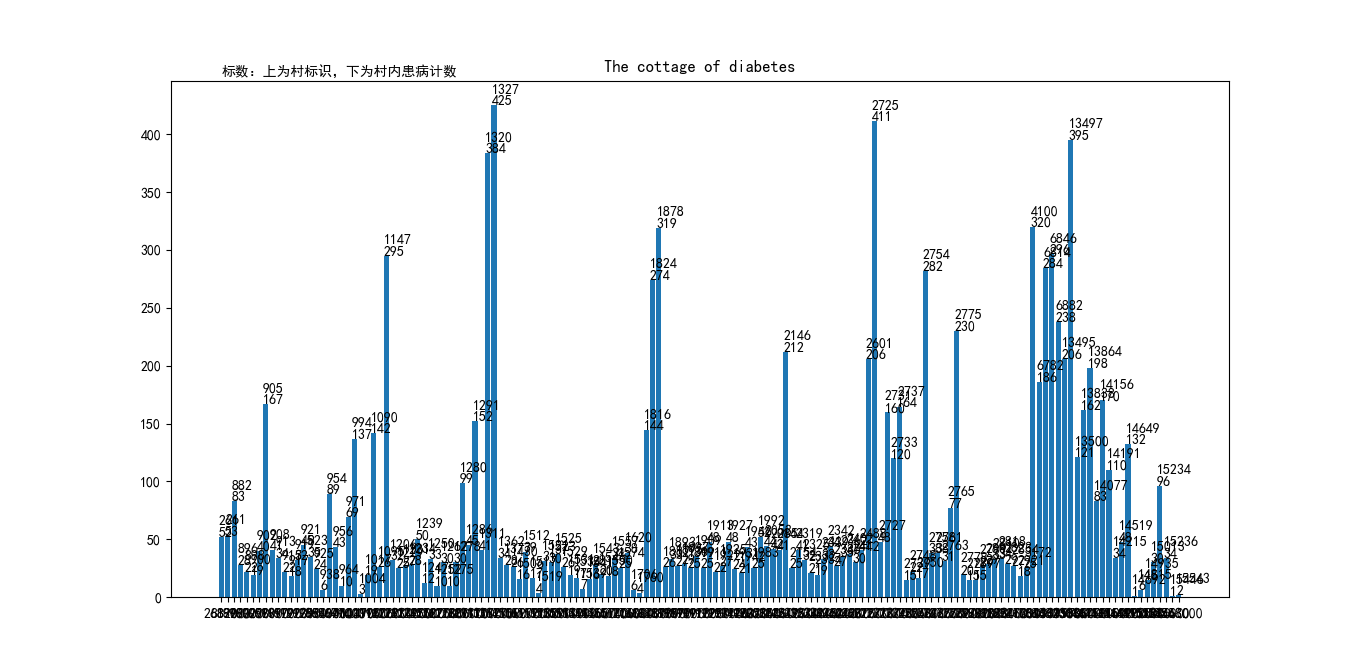
（患病时年龄存在负数和大于100的，可见数据存在误差）



## 村庄分布

以柱状图显示村庄中患病人数

可见存在几个村庄患病人数颇高（例如村标识为1174，1327，1878，1824，2725，13497等），考虑到糖尿病与平时饮食有较大关系，可能这些村庄附近存在水源或食物中糖类物质含量高等其他诱发糖尿病的因素



## 空腹血糖（FBG）

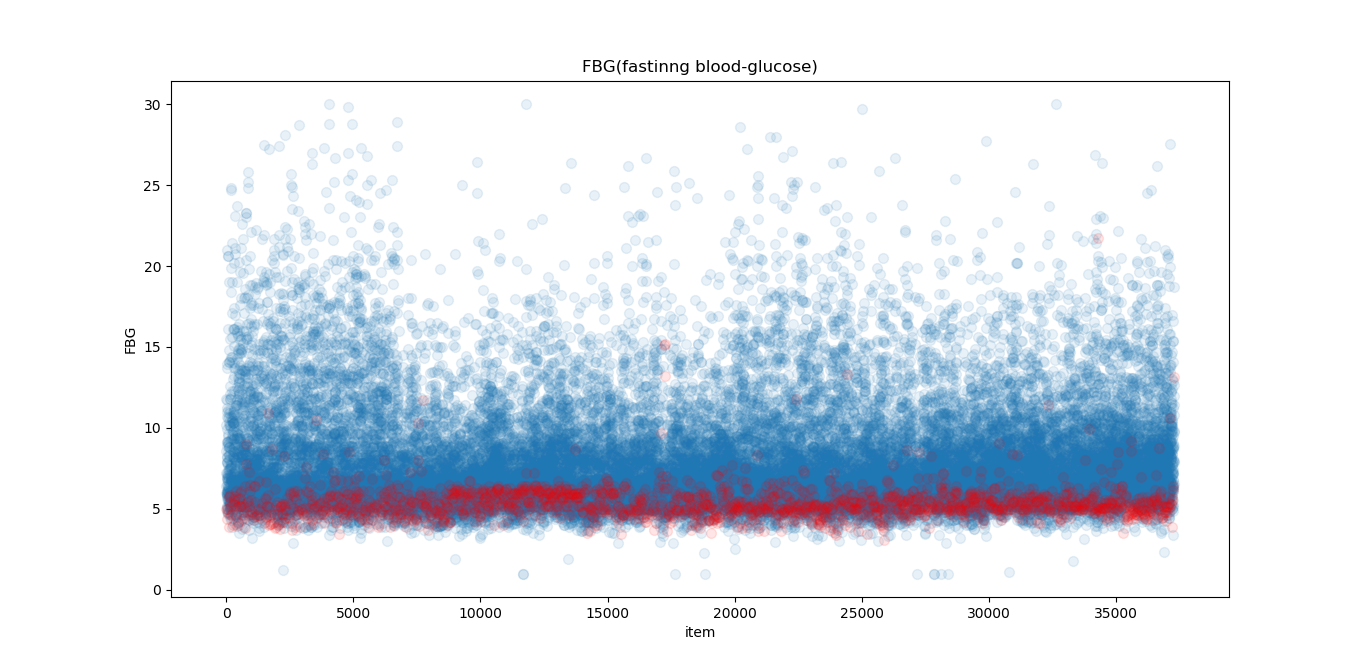
蓝色为患病者（FBG） 红色为健康者FBG

颜色深浅则表示数据出现频率高低

可以看出健康者的FBG在5左右，未有明显高值

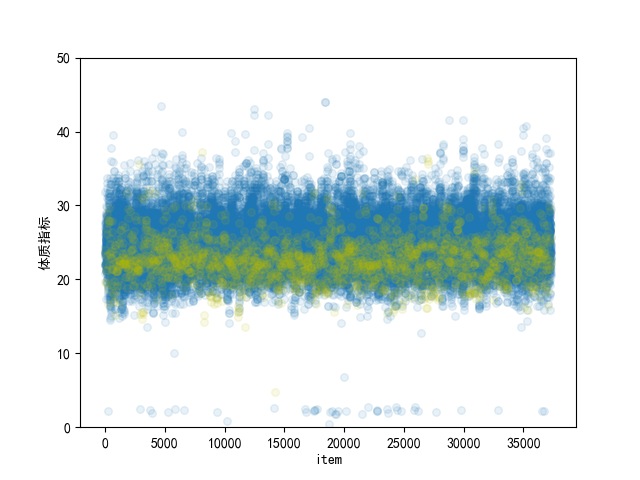
而多数患病者的FBG远远高于5，高者能达到健康者的五倍左右

所以体检结果中的FBG可以作为衡量患糖尿病与否的标准，但FBG为5并不一定以为正常，部分患者FBG也会为5

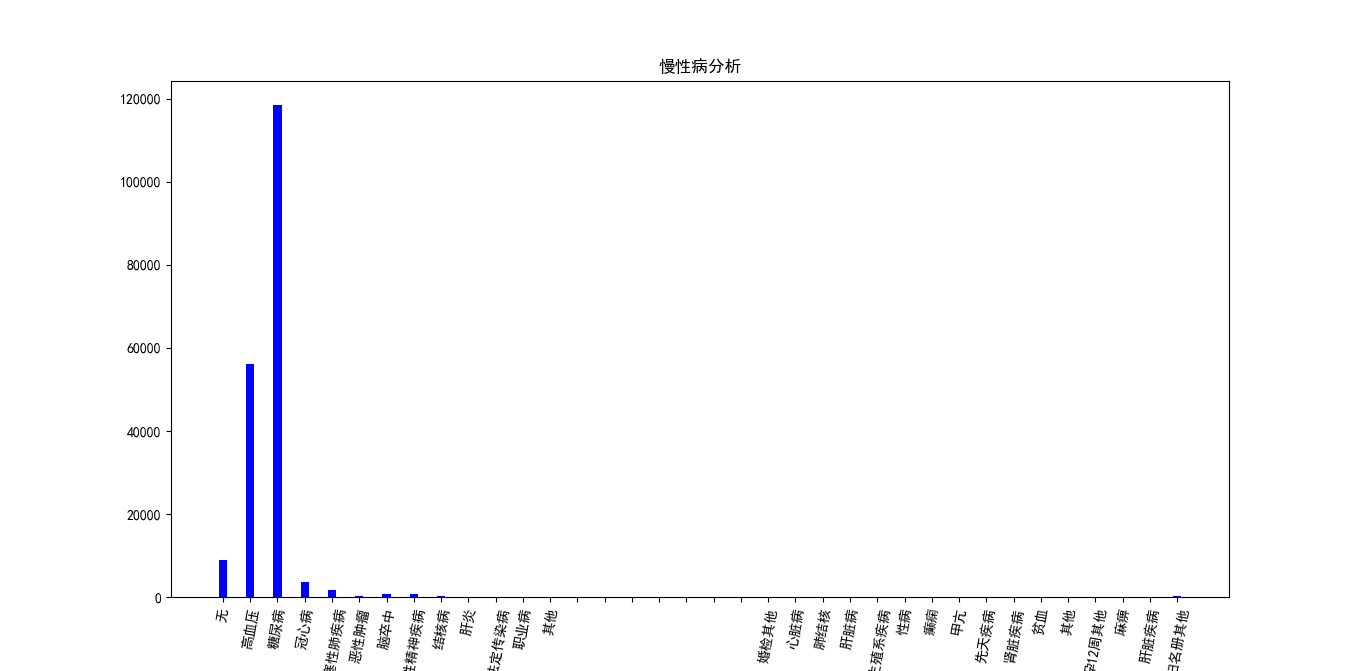


## 体质指标（BMI）

可见健康者BMI多在正常范围（18.5-23.9），而糖尿病患者多数BMI偏高，符合常人对于糖尿病人体质的理解



## 慢性病



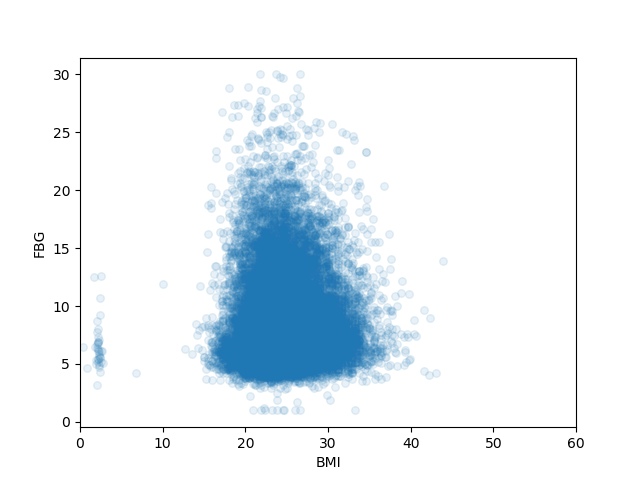
由图可知，糖尿病患者常伴随着高血压的出现，冠心病也偶有发生。

# 联合分析

从单项分析中的结果，对一些指标进行联合分析，判断是否存在相互关联关系。

## BMI与FBG相关性

在发现患者的FBG和 BMI都偏高（或有此趋势）后，考虑是否二者存在某种相关性，故作下图。但由图显示，二者并没有显著联系



## 多个指标联合分析

|  |  |
| --- | --- |
| 患病 | 健康 |
|  |  |

上述图表已相关系数绝对值为准绘出

可知，无论患病与否，左右侧舒张压、左右侧收缩压两两存在强相关性

心脏心率与脉率几乎一致

白蛋白与血钠浓度存在较强相关性

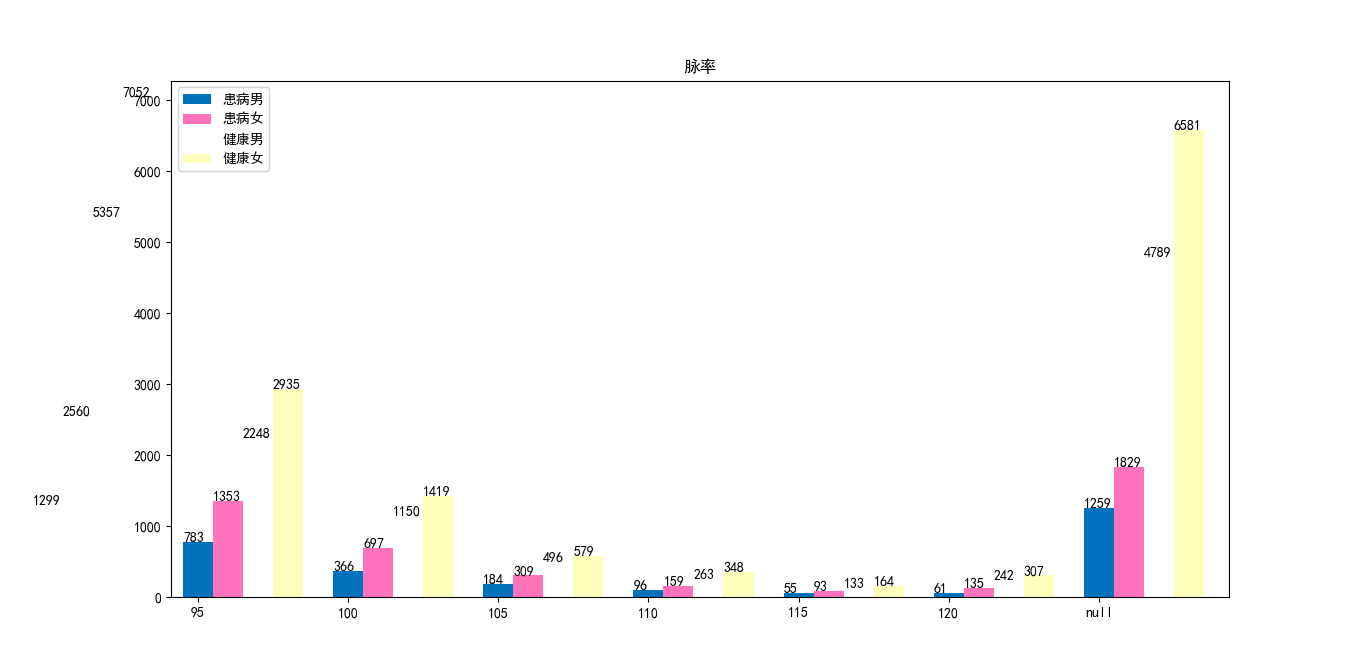
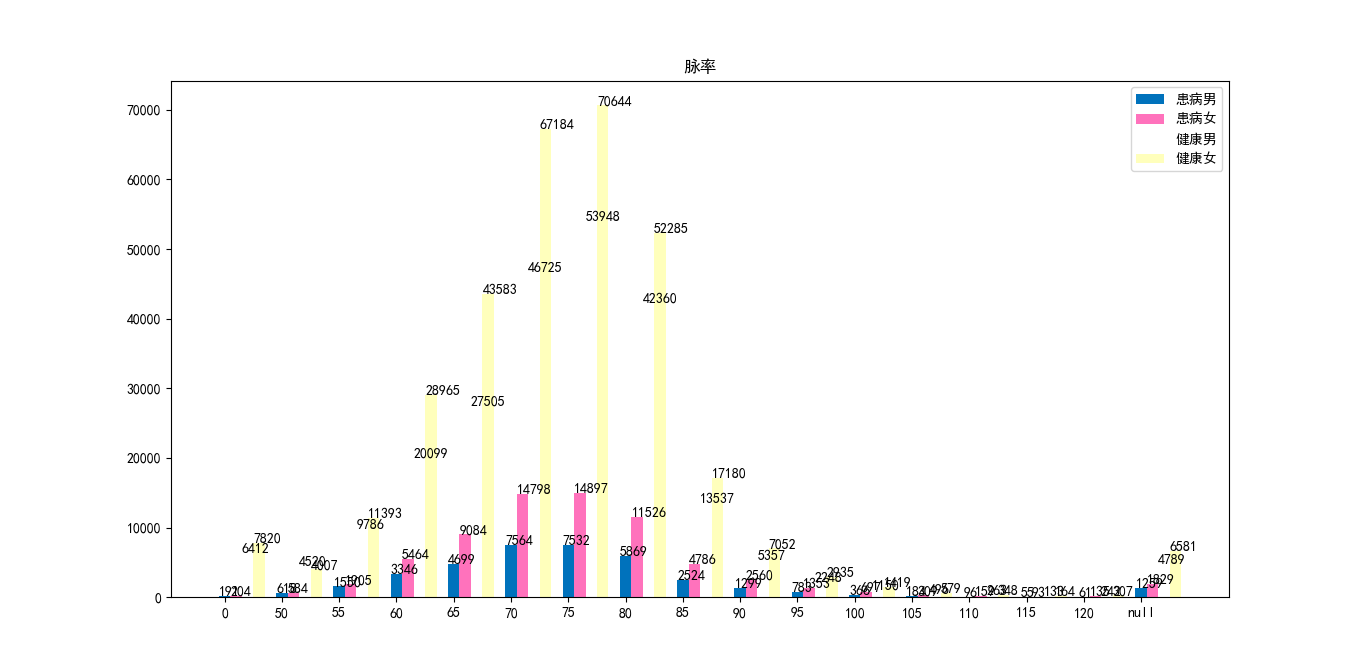
健康人与患病者差异体现在健康人的餐后血糖与空腹血糖不像患病者存在较高的相关性，可以反映出健康人能更好地消化吸收，降低空腹血糖含量，是与患病者的主要区别。

# 性别差异健康患病分析

按照患病男、患病女、健康男、健康女分别对脉率、空腹血糖、呼吸进行分析。每个分类配图两份，第一张为总览，第二张为数据值小的放大图。

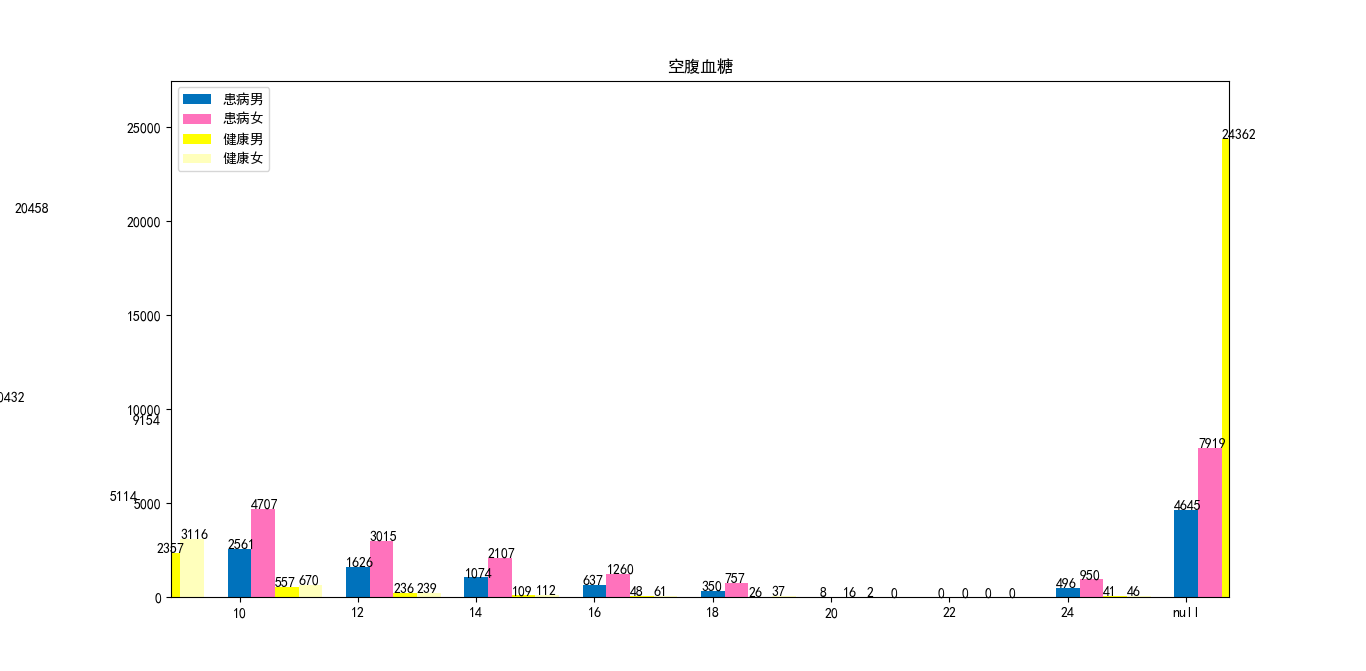
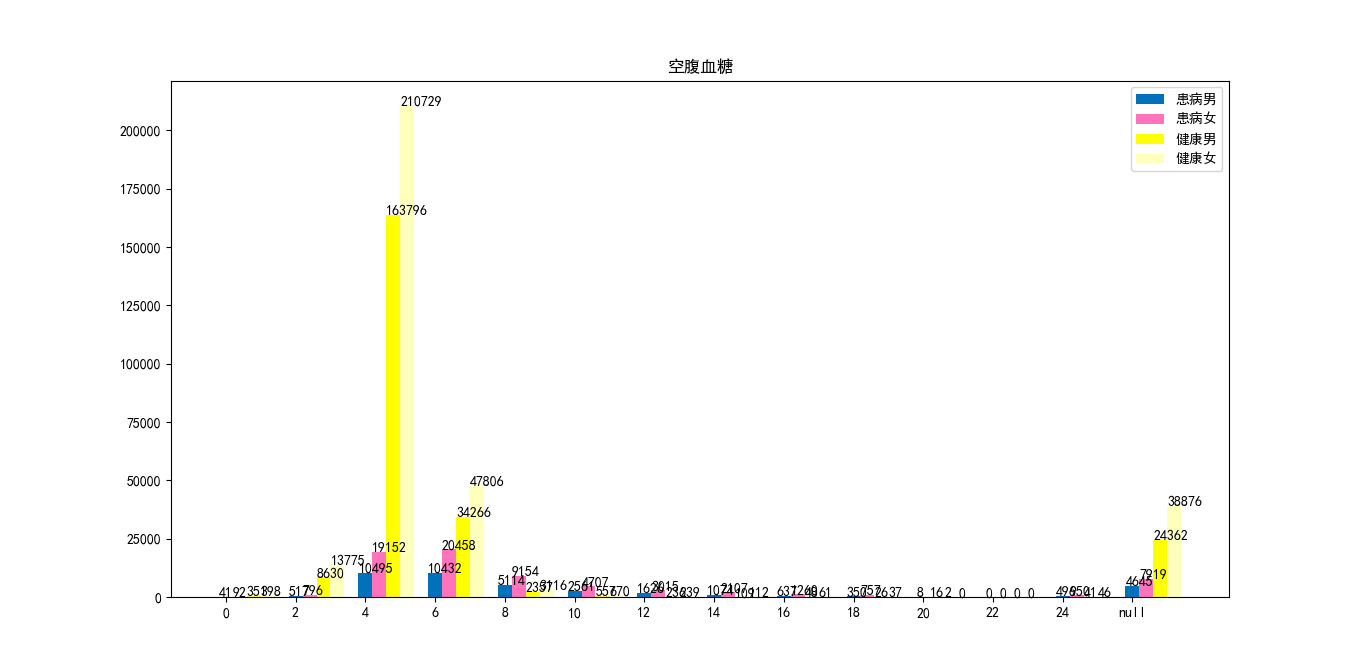
## 脉率

由图片可以看出，四类人群的脉率值分布相似，75~80的最多，其次是70~75、80~85。可见脉率对患病与否及男女差异并无影响。



## 空腹血糖：

患病者空腹血糖值偏高已在第一次分析中进行绘图分析，就男女差异而言，并无直观体现。



## 呼吸

图片显示：健康男、患病男峰值出现在15~20，其次是20~25，两者无差。而健康女峰值出现在20~25，其次是25~30；患病女峰值则出现在15~20，其次是20~25。可见女性患病或许会导致呼吸频率下降。

特作折线图以凸显峰值位置

