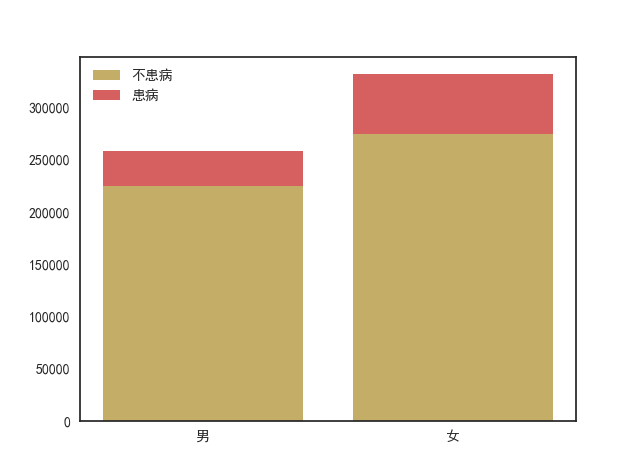
数据分析说明文档

—————基于性别的diabete分析

前期工作：根据用户标识的唯一性对数据进行了去重，一个用户保留一组数据，不考虑多次体检，且对男女数据进行分割。

1. 总体上的男女患病比例



横坐标为性别 纵坐标为数目

男性患病频率=0.1293465349599347

女性患病频率=0.1720627096019935

P(女性患病)>P(男性患病)

1. 不同性别患病不患病的年玲玲分布

纵坐标-年龄

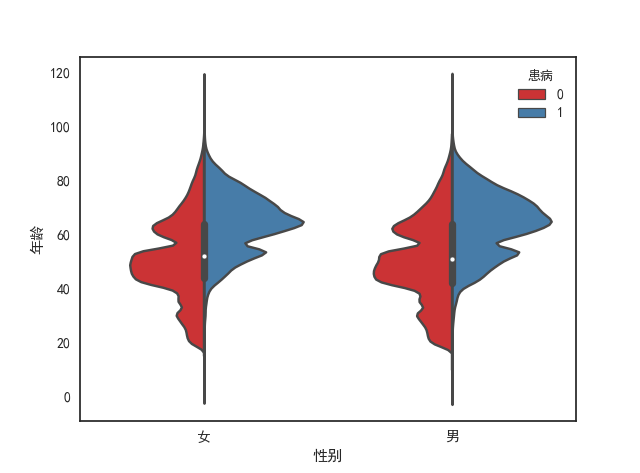
横坐标-性别

red不患病

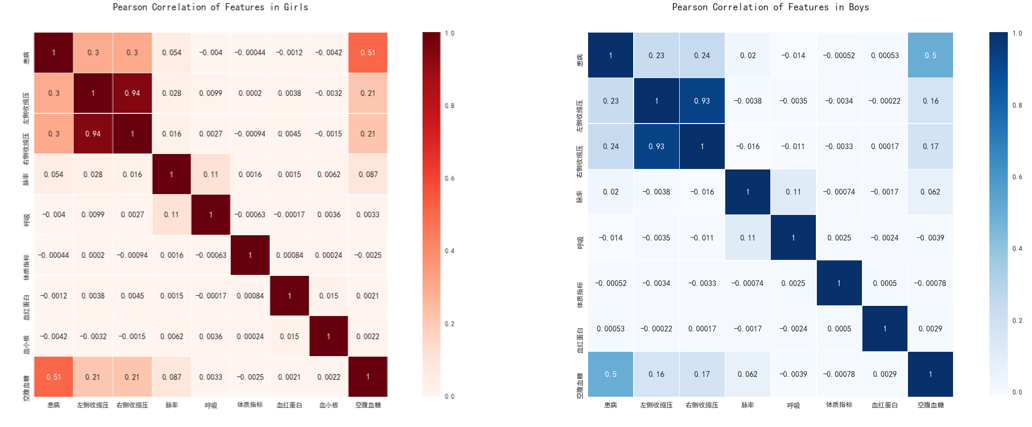
blue患病

男女均为60-80区间患病人数多

60岁左右均出现人数骤减

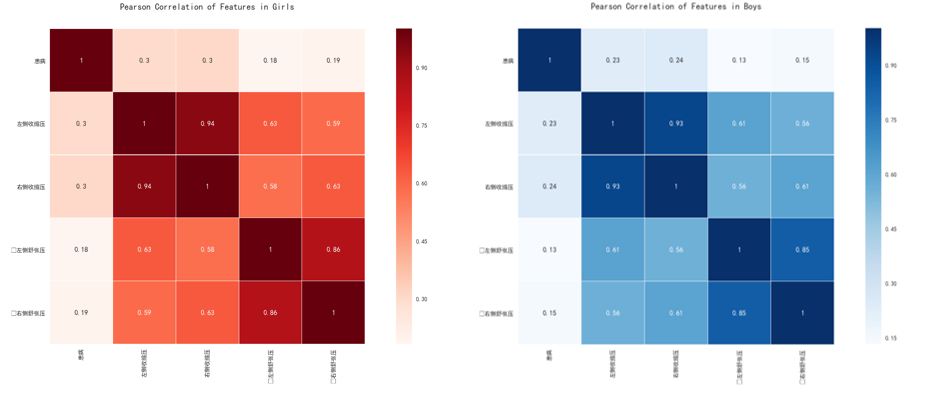


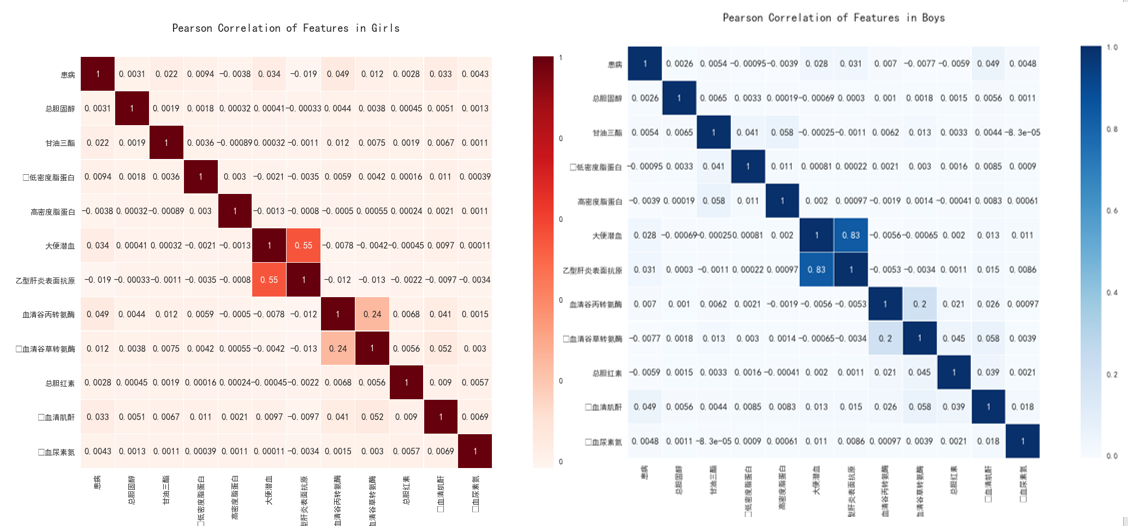
1. 各项指标相关系数的呈现（red女blue男）



患病与空腹血糖以及收缩压相关性很大

下面是与收缩压舒张压相关系数





这两张表上并没有特别影响患病的指标，但是可以明显看出大便潜血与乙型肝炎表面抗原有很大相关性，尤其是在男性中

1. 男性女性中收缩压与患病的关系

red患病 blue不患病

正常收缩压范围60-140，可以看出150以上的患病人数非常多 图片包含 屏幕截图

已生成极高可信度的说明 图片包含 屏幕截图

已生成极高可信度的说明

左侧收缩压

图片包含 天空, 时钟, 物体, 天线

已生成极高可信度的说明

右侧收缩压

图片包含 物体, 天线

已生成高可信度的说明

可以明显看出患病的收缩压要明显高于不患病的（中位数来代表水平）

1. 舒张压

图片包含 屏幕截图

已生成极高可信度的说明

图片包含 屏幕截图

已生成极高可信度的说明

1. 男女中饮酒情况与患病关系

无明显差别

1. 男女中图片包含 屏幕截图

   已生成极高可信度的说明图片包含 屏幕截图

   已生成极高可信度的说明饮酒种类与患病关系

无明显差别

图片包含 屏幕截图

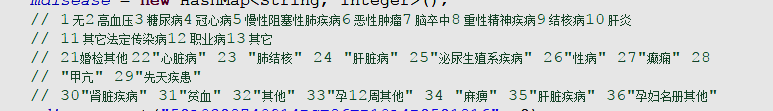
已生成极高可信度的说明

1. 男女中禽畜栏与患病关系

无明显关系图片包含 屏幕截图

已生成极高可信度的说明

1. 男女中慢性病与患病关系



#2病

#0.4504636553993419

#0.47380218548613057

#4病0.020101705055339515

#0.031013589240683665

#5病0.016691594376308704

#0.009351358924068366

#6病0.0022135806162129823

#0.0025042028579434017

#7病0.007717618905174992

#0.005726393947884561

#8病0.004756206999700868

#0.005743905855982068

·· #9病0.006880047861202512

#0.0017161669935556179

#10病0.003709243194735268

#0.0022765480526758196

#11病0.00017947950942267426

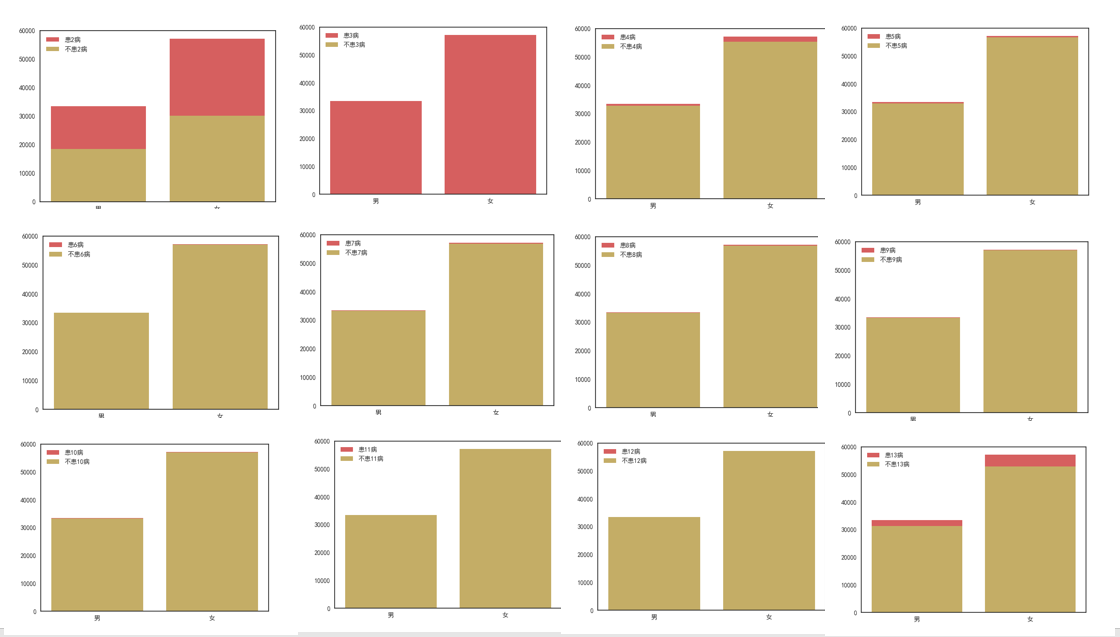
#8.755954048753153e-05

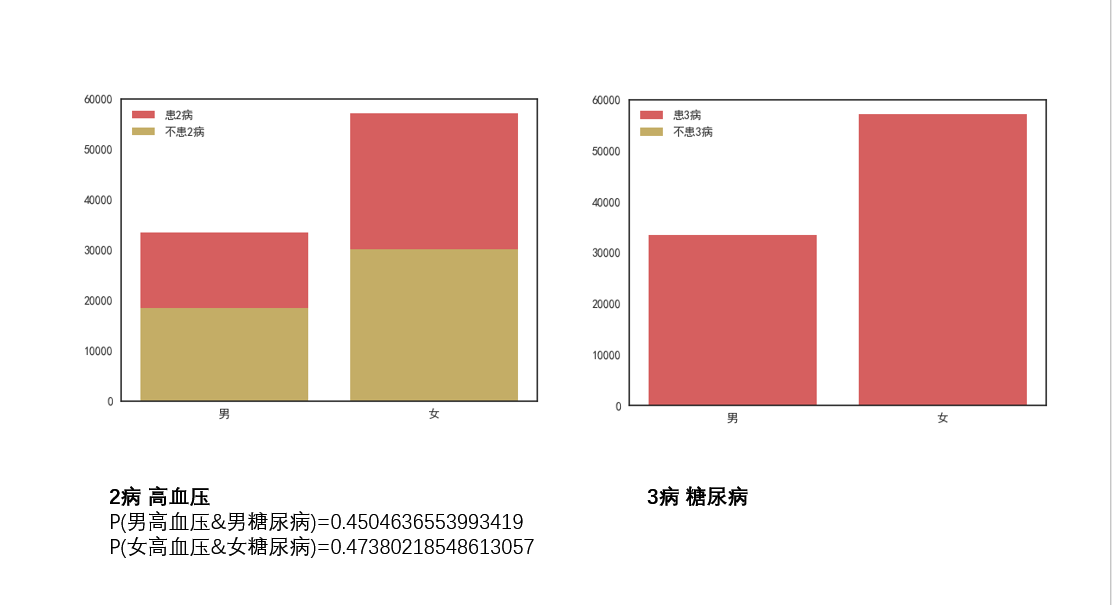
#12病0.0005982650314089142

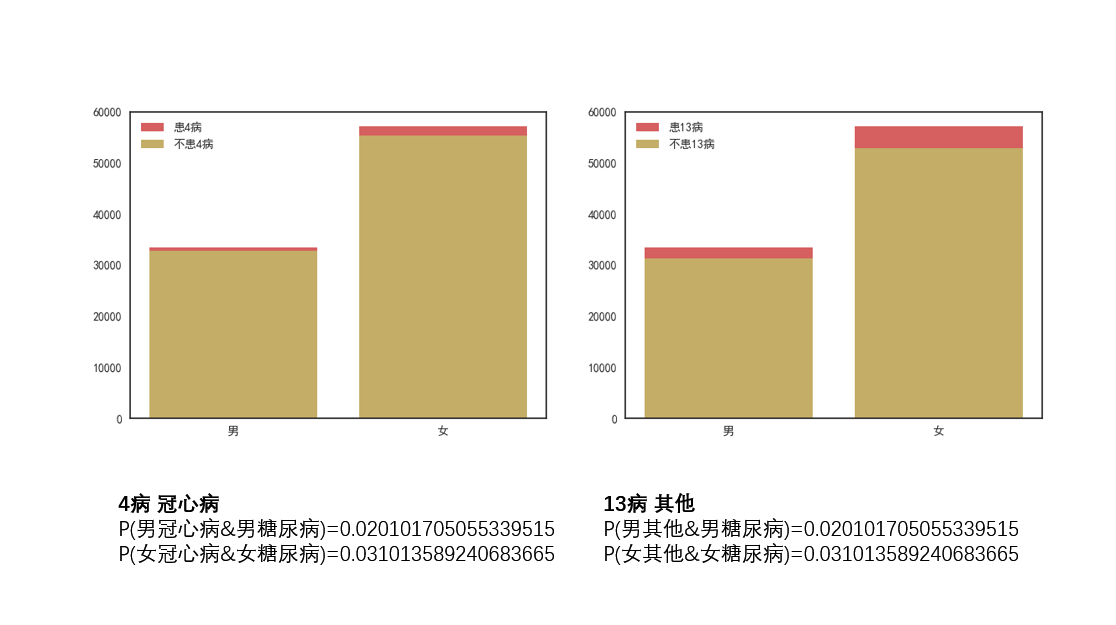
#0.0003327262538526198

#13病0.06215973676338618

#0.07419795460913421







1. 结论：

基于以上针对性别的实验研究，并没有发现明显的不同导致男女患糖尿病病的因素，但是可以得出以下结论：

1. 女性患病率略高于男性，且60-80年龄段患病人数最多
2. 空腹血糖与糖尿病患病相关性最大
3. 糖尿病患病与血压，尤其是收缩压，具有很强相关性