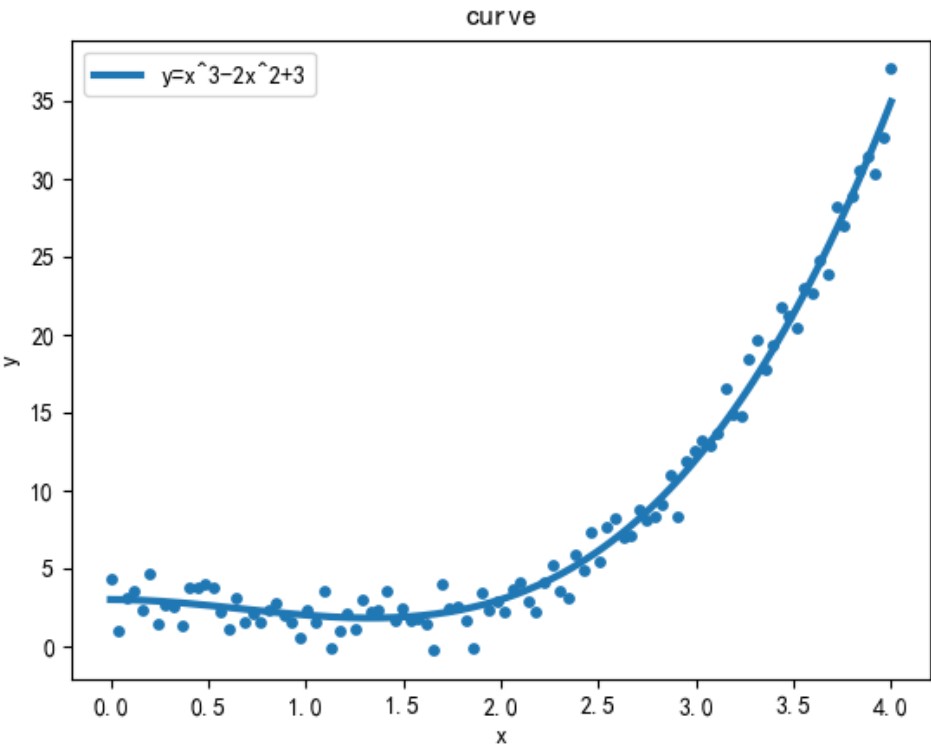


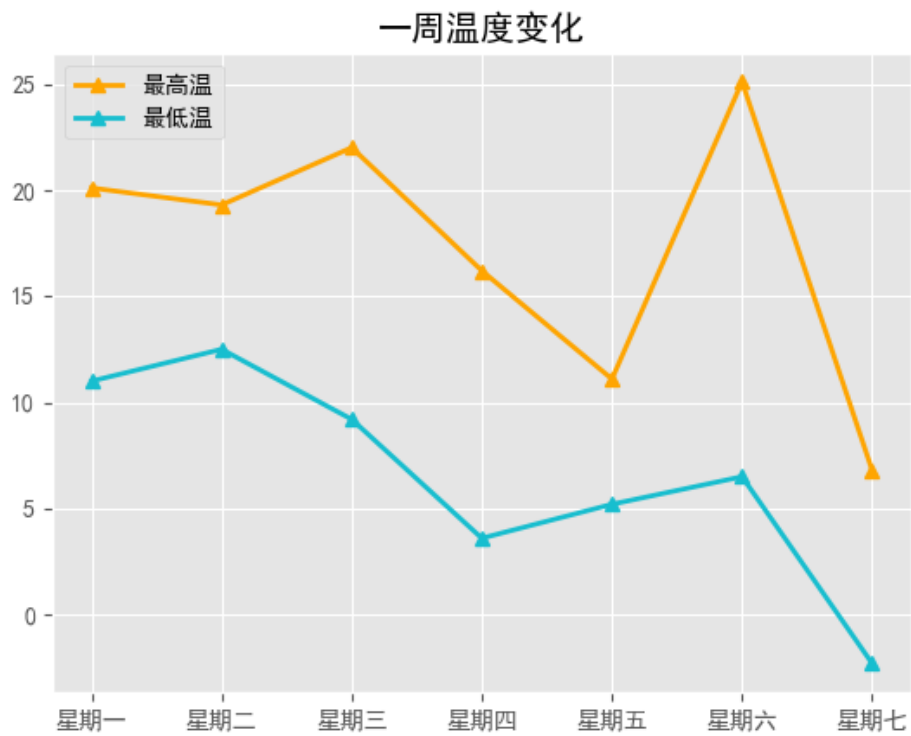
每个人均自行完成所有题。其中第五题需要自学一下新画图函数。

1. (1) 在[0,4]区间等距取 50 个数存到列表 `x` 中，然后生成数组 `d = np.random.randn(50)`,
(2) $f(x) = x^3 - 2x^2 + 3$ ，求出 `x` 数组的 $f(x)$ 的函数值，再与 `d` 相加，结果为 `Y`。用 `x` 为横坐标，`Y` 为纵坐标画点。点的形状为实心圆，大小为 15。
(2) 在上图中继续画出函数 $f(x) = x^3 - 2x^2 + 3$ 在区间[0,4]之间的曲线，线宽为 3。设置 `x` 坐标轴标签为 '`x`'，设置 `y` 坐标轴标签为 '`y`', 曲线标记为 $y = x^3 - 2x^2 + 3$ 。图名为 '`curve`'。
结果保存为 1.png



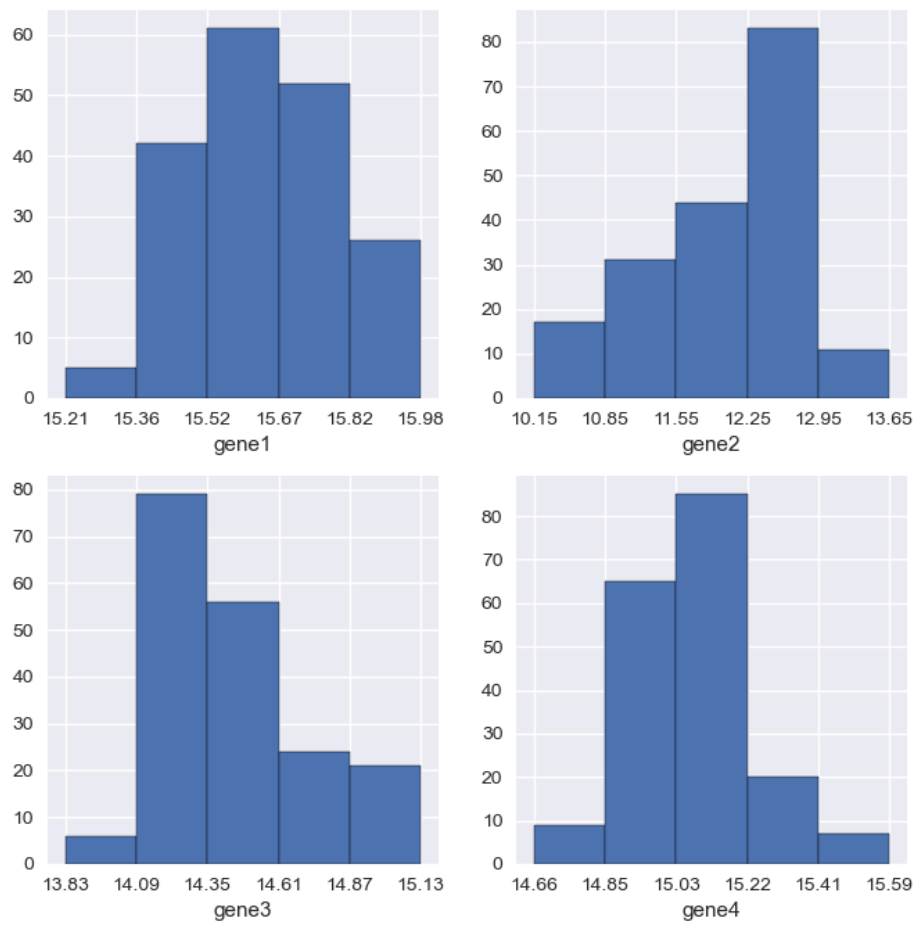
- 2.下表是某地一周内最高温和最低温变化，在一张图中画出最低温和最高温曲线。风格采用 '`ggplot`'（其他自选风格也可以）。
参数如下：
最高温曲线：标签：最高温，`marker`：三角形，颜色：橙色，线宽：2
最低温曲线：标签：最低温，`marker`：三角形，颜色：#17becf，线宽：2
`x` 轴刻度如图所示
图名为 '`一周温度变化`'
结果保存为 2.png

	星期一	星期二	星期三	星期四	星期五	星期六	星期天
最高温	20.1	19.3	22.0	16.2	11.1	25.1	6.8
最低温	11.0	12.5	9.2	3.6	5.2	6.5	-2.3

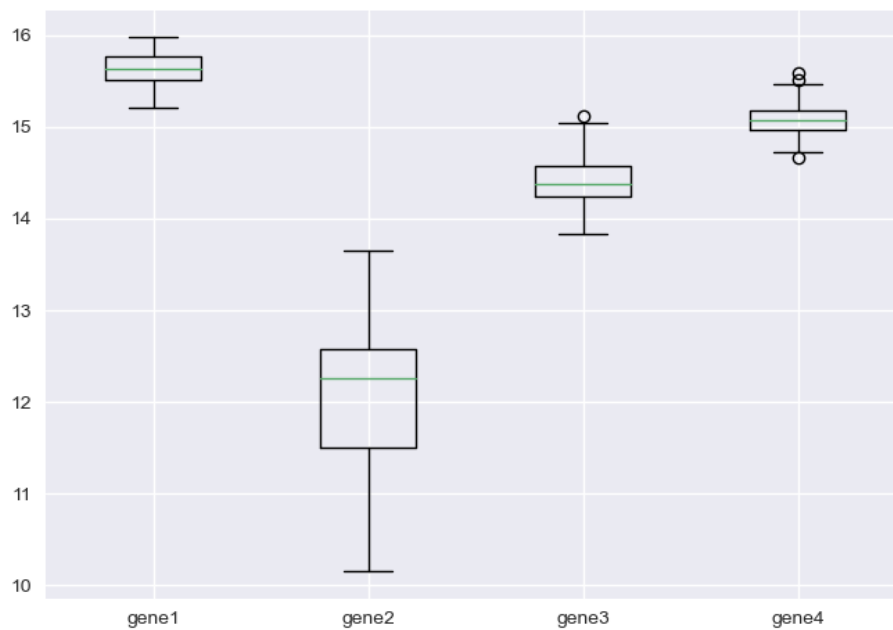


3. 文件：exp.txt 中有 4 个基因的表达数据，每一行是一个基因。

问题：(1)在一张图中分别画 4 个子图，每个子图画一个基因表达情况的分布图。即把基因的表达量范围等分为 5 分，统计每一个表达范围内的数量。x 轴刻度显示为每个区域范围，如下图所示。结果保存为 3.1.png。

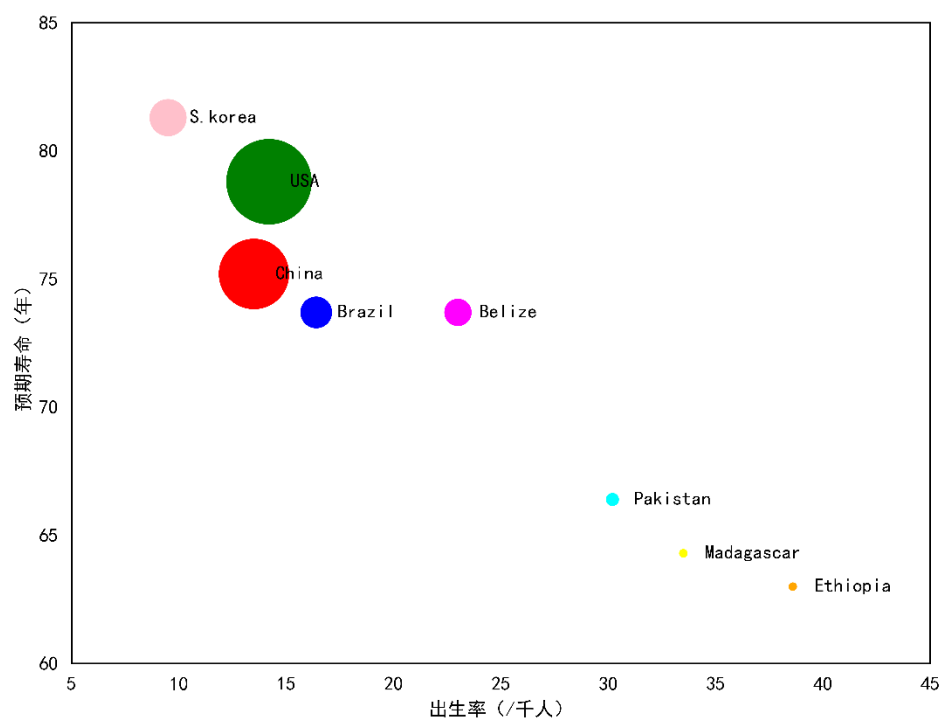


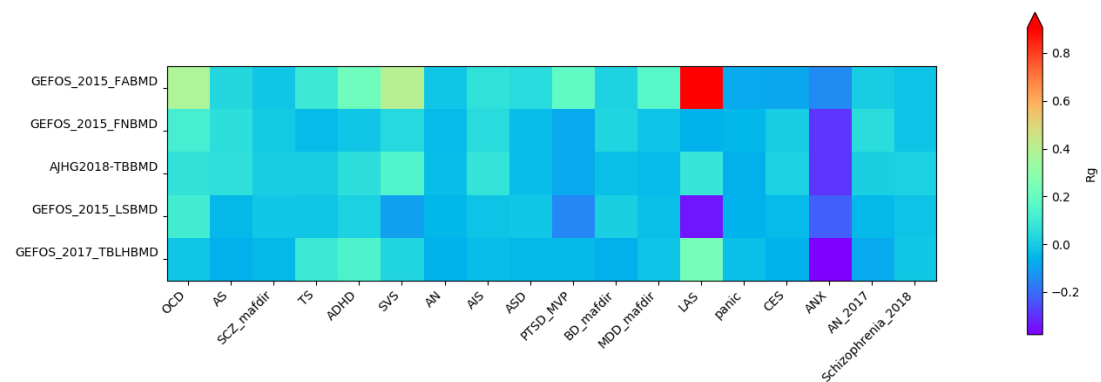
(2)在另一个图中画出基因表达的箱线图，结果保存为 3.2.png。



4. 文件 countries.txt

问题：中包含 8 个国家的数据，数据分别为出生率（每千人）、预计寿命（岁）。GDP。以出生率为横坐标，预计寿命为纵坐标，每个国家画一个点（圆形），点的大小为 GDP/20。并在对应点圆心右边 1 个单位处 (x+1) 标出国家名称。X 轴范围[5,45],Y 轴范围[60,85],坐标轴名称如下图，颜色自己设置。结果保存到 4.png 中





6.文件: BMJ2018-fracture.ma 为骨折 GWAS summary 数据

问题: ma 文件提供了四列, 其中 CHR 为 SNP 所在染色体序号, BP 为 SNP 在染色体上的物理位置, P 代表了目标 SNP 与表型是否显著关联。曼哈顿图一般用于 GWAS 结果展示。其中 X 轴代表了 SNP 的物理位置, Y 轴用 $-\log_{10}(P)$ 值代表了 SNP 显著性。

利用 matplotlib 完成曼哈顿图的绘制。

首先根据 CHR, BP 两列对 SNP 进行 X 轴上的定位, 然后根据 P 值计算 $-\log_{10}(P)$ 进行 Y 轴上的定位, 再根据 CHR 对 SNP 点赋予颜色属性。利用 `.scatter(x, y, color)` 进行绘制。

利用 `.xticks()` 对 X 轴进行重命名。

根据 SNP 总数进行 bonferroni 校正, 即阈值 $p=0.05/n$, n 为 SNP 总数。在图上添加代表阈值的横线。

最终保存为 6.png。

