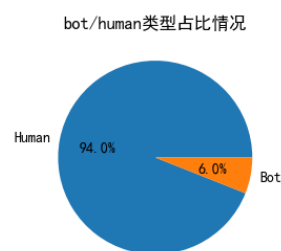


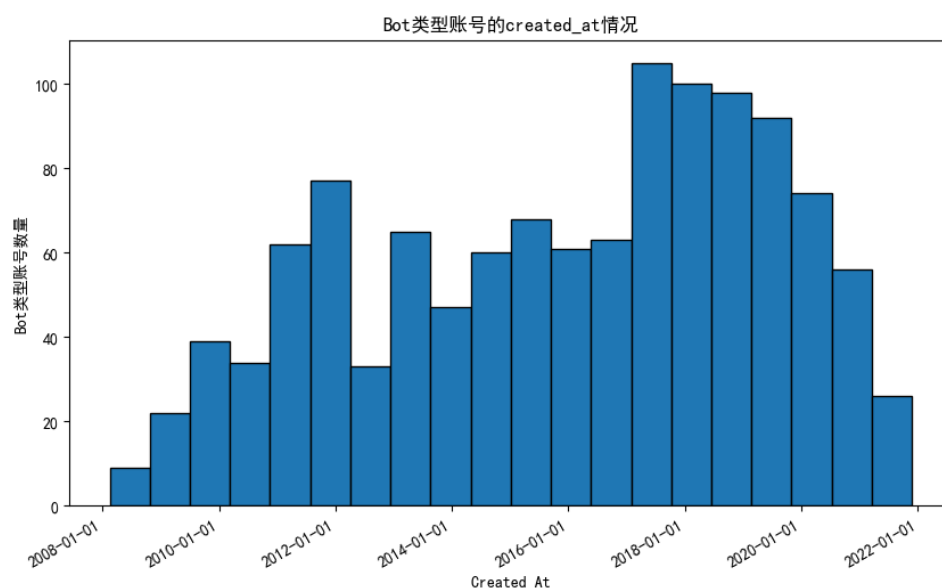
数据科学与工程导论 - Homework7报告

10235501469 邵乐怡

4.1 使用饼图可视化bot和human类型账号占比的情况，饼图可以较为直观地展示不同类型数据的占比情况。可以发现在所有记录的GitHub账号中，绝大部分为Human账号。

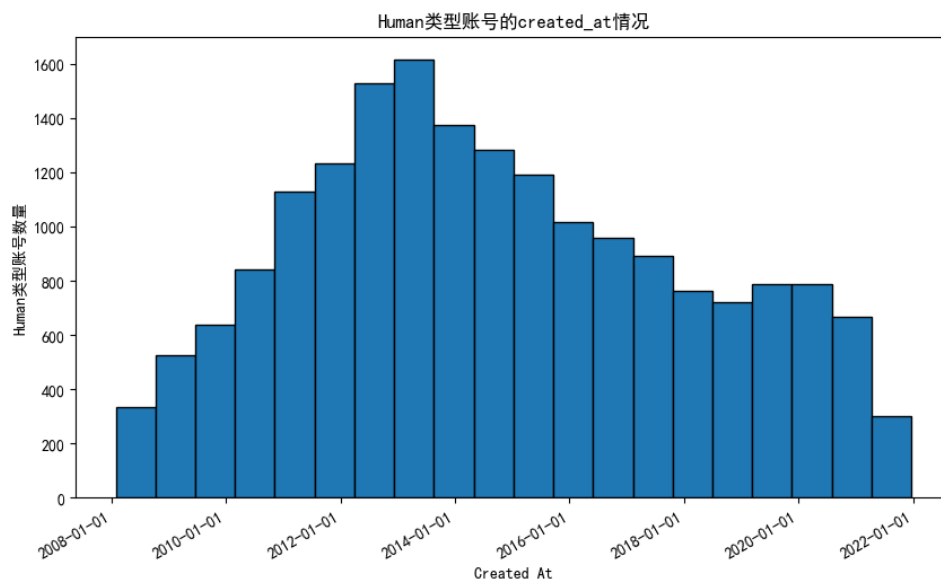


4.2 使用直方图可视化bot类型账号的created_at情况，直方图可以直观地展现bot类型账号的创建时间分布和趋势，每个条形的高度表示在该时间区间内创建的 bot 账号数量。可以发现2017年之前的bot账号创建数量波动程度较大但总体呈上升趋势，在2017年底达到一个高峰，此后开始减少，且减少的幅度随时间推移越来越大。

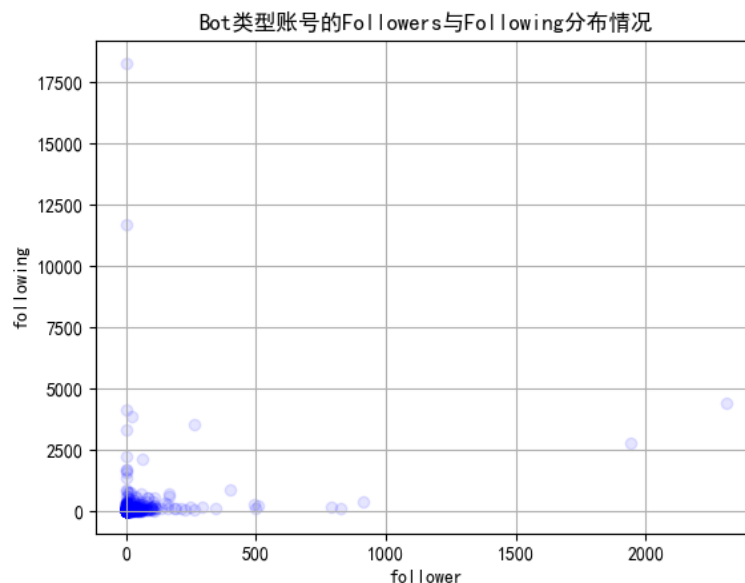


4.3 使用直方图可视化human类型账号的created_at情况，直方图可以直观地展现human类型账号的创建时间分布和趋势，每个条形的高度表示在该时间区间内创建的

human账号数量。可以发现human类型账号的创建数量自2008年开始逐年增加，在2013年底达到一个高峰，此后呈下降趋势。



4.4 使用散点图，直观地展示每个bot类型账号的followers与following之间的关系和这两个指标各自的分布情况。可以发现绝大部分bot账号followers和following的数量都集中在0附近，极少数账号有超过2500个following/超过500个follower。



4.5 使用散点图，直观地展示每个human类型账号的followers与following之间的关系和这两个指标各自的分布情况。可以发现绝大部分human账号following的数量都集中在0~20000，follower的数量都集中在0~2500。可以发现在社交影响力较大（following超过20000或followers超过2500）的账号中，follower比较多的账号following数量较少，反之亦然。

