

2017~2018 年第 2 学期期末考试试题

1. 数据: nycflights13, 如果没有请先安装包(nycflights13)
 - (1) 将 flights 和 planes 进行连接, 其中, 最终结果应当保存 flights 中的所有观测值并且保存 planes 中对应的观测值, 并将结果命名为 flights_planes
 - (2) 选取 flights_planes 中 tailnum 中含有字母"W"的观测值, 并将选取出来的数据框保存为新的 flights_planes
 - (3) 剔除 flights_planes 中 engines 和 seats 缺失的观测值, 然后将数据按照 engines 和 seats 分组, 计算各组飞机在 2013 年的 distance 和 arrive_delay 的总和, 计算结果不能为 NA, 将计算结果保存为新的数据 flights_sum
 - (4) 利用上一问所得的 flights_sum, 计算不同 engines 的总到达迟到时间和总飞行距离
2. 本题使用数据 flights
 - (1) 请首先剔除数据中实际出发时间(dep_time)缺失的观测值, 然后生成一个新的变量 dep_interval, 用来将数据中的实际出发时间分为上午(6:01-12:00)、下午(12:01-18:00)、晚上(18:01-24:00)和凌晨(0:01-6:00)四组, 将每一年每一月每一天内每个 dep_interval 分组 (year, month, day, dep_interval), 计算每一组的平均到达延误时间和平均到达的机场数量 (注意: 相同机场不能重复计算) 以及到达延误的方差
 - (2) 选取平均到达延误在 30 分钟以内同时平均到达机场数据不低于 50 个的小组, 并按照平均到达延误由大到小排列
3. 利用数据集合 nycflights13 中 flights 和 planes
 - (1) 利用 flights 和 planes, 找出平均服役时间最长的前 10 种飞机型号(model)
 - (2) 将数据集 flights 根据目的地(dest)进行分组, 统计出各小组的行数, 将行数最多的前 10 个观测值保存为 top10_dest。用同样的操作, 将数据集 flights 根据目的地(dest)进行分组, 统计出每个目的地的平均起飞延误时间(avg_dep_delay), 将结果命名为 avg_dep_delay
 - (3) 将 top10_dest 与 avg_dep_delay 按照 dest 进行合并, 得到数据集 dest_delay, 并画图表示出以 dest 为 x 轴、avg_dep_delay 为 y 轴的折线
4. 编写函数分别实现以下功能:
 - (1) 对于任意给定的一个数 n, 找出 1 到 n 之间的所有质数
 - (2) 找出 1000 以内所有能够被最多两组不同质数相加得到的偶数 (例如: 12=5+7 为一组不同的质数, 14=7+7 为一组相同的质数, 18=5+13 和 18=7+11 为两组不同的质数, 56=3+53、56=13+43 和 56=19+37 为三组不同的质数), 注意最后输出的结果要包括偶数本身, 以及对应相加的质数

