个人作业-2：数据可视化

- 作业必须以 .Rmd 自动报告形式提交（同时生成.html或.docx文件）
- 本次作业使用的数据集在nycflights13包中
  - 若尚未安装该包，请先安装install.packages("nycflights13")；注意：flights数据集中变量dep_time 中的缺失值表示航班取消了；
  - 在确保正确完成可视化要求的基础上，请尽可能美化图形；
- 1、箱线图和小提琴图
  - （1）为flights数据框中air_time变量绘制箱线图，并在同一张图中叠加air_time变量的小提琴图；
  - （2）4个季度air_time变量箱线图并列对比（一张图中）；
  - （3）12个月air_time变量箱线图并列对比（一张图中，且各箱线图水平放置）
- 2、直方图和核密度估计曲线
  - （1）画distance变量直方图，并在同一张图中叠加该变量的核密度估计曲线；
  - （2）在同一张图中绘制单航班平均distance排名前三位的航空公司（carrier）的distance变量的核密度估计曲线对比（即在一张图中绘制这三个carrier各自的distance的核密度估计曲线）；

作业

- 3、散点图与条形图
  - （1）绘制条形图，对比各个航空公司（carrier）单航班平均distance，要求按上述平均distance值从大到小排列条形（若航空公司数据量较多，可以考虑水平放置条形）；
  - （2）绘制堆叠条形图，对比各航空公司（carrier）的总distance，要求按总distance值从大到小排列，且每个条形分别由相应的4个季度的里程堆叠而成；
  - （3）绘制饼图，展示各公司航班数比例（要求将航班数小于10k的公司的航班数合为other展示，这些公司不在图中单独展示）；
  - （4）绘制散点图展示distance和air_time之间的关系，还要求在同一张图中叠加平滑线；（若点太密集，请考虑用高密度散点图重绘，或用hexbin处理）；
  - （5）使用分面变量season（季度，需新生成该变量）和carrier，按（4）中的要求绘制分面散点图；
  - （6）绘制散点图矩阵，展示distance、air_time、dep_delay、arr_delay这4个变量间的关系；

作业

- 4、其他
  - （1）绘制日历图，展示UA公司每天10点至12点（均为上午时间）之间到达的航班数量；
  - （2）以月为横轴绘制折线图，展示平均延误时间的变化；
- 5、请用适当的图形探索或展示下列问题：
  - 查看每个目的地，是否有些航班的速度快得可疑？（换言之，这些航班的数据可能有误）
  - 研究到每个目的地（dest）的平均距离（distance）与平均延迟到达时间（arr_delay）之间的关系；
  - 查看每天取消（若dep_time缺失则航班被取消）的航班数量，被取消航班存在模式吗？已取消航班的比例与平均延误时间有关系吗？
  - 哪个航空公司的延误情况最严重？你能否分清这是机场的问题，还是航空公司的问题吗？为什么能？为什么不能？提示：试一下flights %>% group_by(carrier, dest) %>% summarize(n())；
  - 若想尽量避免航班延误，你会在一天中的哪个时间搭乘飞机？