

# 刘高彤

gaotongLIU@163.com | 6462044317 | 13614007168

领英: [www.linkedin.com/in/gaotong-liu-gloria](https://www.linkedin.com/in/gaotong-liu-gloria) | 个人网站: <https://gaotongliu.github.io/>



## 教育经历

哥伦比亚大学, 梅尔曼公共卫生学院

2019年9月-2021年5月

生物统计学 硕士, GPA: 4.1 (5%)

纽约

- 相关课程: 数据科学(A), 概率论(A), 生物统计方法学(A+), 纵向数据分析, 生存分析, SQL

香港大学, 交换项目

2018年9月 - 2018年12月

- 相关课程: 市场分析, Python, 统计推断

香港

香港理工大学

2014年9月 - 2018年5月

应用生物与生物科技 理学学士, GPA: 3.94/4.0

香港

- 授予一级荣誉毕业生 (5%)

- 相关课程: 统计学, 生物科技产品商业化

昆士兰大学, 交换项目

2017年2月 - 2017年6月

- 相关课程: 生物统计学

布里斯班, 澳大利亚

## 工作经历

哥伦比亚大学, 梅尔曼公共卫生学院, [COVID-19 DSCovR Dashboard](#)

2020年5月 - 2020年9月

研究助理, Dr. Shing Lee 研究小组, 美国新冠肺炎可视化与统计分析

纽约

- 合作开发了美国首个 Rshiny 新冠肺炎可视化网站, 为大众提供美国及国际疫情趋势, 各州病例特征统计以及各州政策的比较, 网站浏览人数超过 700 人, 日均浏览人数 50 人
- 负责从 52 个各州卫生部官网收集, 抓取, 清理, 解析数据以及对收集到的 25000 多条数据进行质量监控
- 对各州新冠患病, 住院与死亡人群的族裔, 年龄, 性别数据进行可视化分析, 同时研发新的可视化图表展示各州近 1000 条疫情政策的阶段性调整
- 使用 R 爬虫收集 52 个州的新冠肺炎死亡情况与 8 种常见慢性病发病率数据, 利用随机效果模型对其关系进行 meta 分析, 研究发现患有慢性病的人群更易因新冠肺炎而死亡, 目前已完成英文论文初稿

## 项目经历

香港大学轻餐市场分析

2018年9月 - 2018年12月

市场研究助理, Dr. Kwan CW 研究小组, 香港大学

香港

- 设计并收集 60 份左右轻餐相关网络问卷调查, 使用 R 清理相关数据并进行描述性数据分析
- 应用多维标度模型, 市场反应模型和对应分析方法对收集的问卷调查进行商业分析, 最终根据潜在顾客数量, 喜好等指标预测轻餐大学市场规模和可获利润, 完成市场分析报告

机器学习预测模型

2020年4月

- 使用 R caret, 根据近房屋面积, 建筑材料, 住宅地址等近 80 项指标, 通过交叉检验比较岭(ridge)回归, lasso 回归, 广义加性模型(GAM) 和 MARSpline 等模型结果, 优化回归模型, 预测房屋价格
- 根据酸度, 酒精含量等 11 项红酒物理化学检验结果, 通过交叉检验比较决策树, 随机森林, AdaBoost 算法, 支持向量机(SVM) 等分类模型的性能, 最终选定随机森林预测红酒质量等级, 并根据变量重要性 (VI), 部分依赖图(PDP), 个体条件期望图 (ICE), 解释预测结果

建立急诊记录数据库

2020年3月

- 使用 MySQL, 根据来访患病人群特征, 急诊到访记录和医生诊疗记录, 建立医院急诊系统的关系数据库

医学院性别薪资差异分析

2019年12月

梅尔曼公共卫生学院, 哥伦比亚大学

纽约

- 根据工作经验, 论文发表, 职称, 性别, 薪资等 10 项数据鉴别出性别与薪资关系中的混杂偏倚因子, 交互作用因子, 最终确立多元线性回归模型, 完成统计分析报告

## 综合技能

技能: 掌握 Microsoft Office 和 R 编程语言, 熟悉 Tableau, SQL, SAS 和机器学习 (逻辑回归, 随机森林, 决策树, Adaboost), 可使用 Python (Pandas, NumPy, Matplotlib, Scikit-learn) 进行可视化与建模

语言: 英语 (流利), TOEFL (104), 日语 (基础)

兴趣爱好: 琵琶 (十级)