****

**交通数据分析与应用课程**

**第四次作业**

姓 名：任一玮

学 号：2310800

学 院：交通运输工程学院

研究方向：道路交通安全与环境

指导教师：王俊骅

二〇二三年十月

作业一：

使用IRI数据集： 获取STATE\_CODE为6且SHRP\_ID以050开头的子集。 获取每个部分的IRI摘要统计数据：最小值、最大值和平均值。 按平均IRI降序排列汇总数据（仅报告一个部分的结果）。 为选定部分生成平均IRI与时间的散点图，然后对图表进行解释：

提示1：平均IRI与日期

提示2：将STATE\_CODE和SHRP\_ID一起用作唯一确定部分的主键。 使用2017年CRSS数据集： 获取数据集accident和person的交集 按伤害严重程度（INJ\_SEV）统计观测总数。

提示：使用summarise()和group\_by() 合并事故数据集与车辆数据集，并报告结果的尺寸和右侧数据集中一个变量的缺失值数量。

提示：使用left\_join()。

获取STATE\_CODE为6且SHRP\_ID以050开头的子集，并得到IRI摘要统计数据最小值、最大值和平均值，按平均IRI降序排列汇总数据如下图所示。

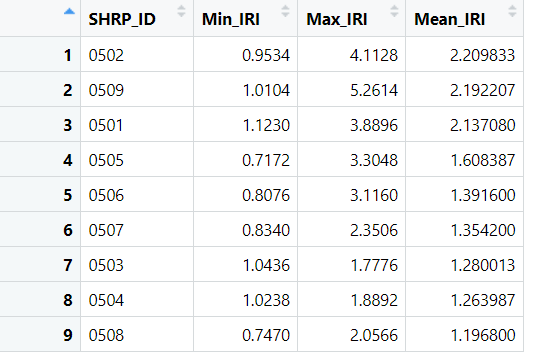


图 1 最小值、最大值和平均值

按平均IRI降序排列汇总数据（仅报告一个部分的结果），

选取平均IRI最高的部分的结果。

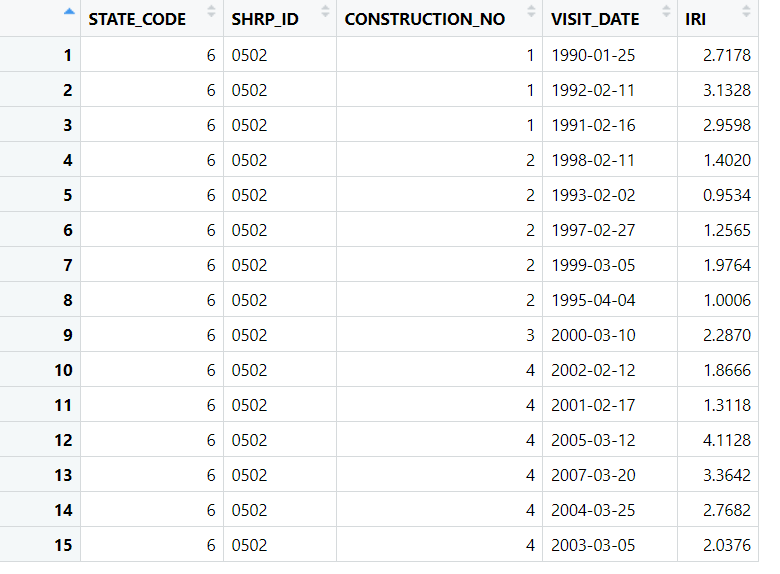


图 2 降序排列

为选定部分生成平均IRI与时间的散点图，然后对图表进行解释：

首先对于表格内数据的含义进行理解：

1. STATE\_CODE: 这通常指的是美国各州的编号，这个代码是按照一定的规则进行分配的，用来唯一标识美国的一个州。例如，STATE\_CODE 6 通常代表加利福尼亚州。
2. SHRP\_ID: SHRP代表Strategic Highway Research Program，一个由美国国家研究理事会发起的项目，用以改善高速公路系统。SHRP\_ID则是在这个项目下，某个公路部分的唯一标识符，它可以帮助研究人员准确找到某段公路的数据。
3. CONSTRUCTION\_NO: 这通常指的是施工项目的编号，用来标识公路施工或维护项目。这个编号在高速公路管理和维修记录中很重要，它帮助跟踪特定施工项目的细节。
4. VISIT\_DATE: 这表示对特定公路段进行观测或数据收集的日期。在IRI数据集中，这个日期与测量路面粗糙度时的实际日期相对应。
5. IRI (International Roughness Index): 国际粗糙度指数，是评价公路平整度的一个指标，用来量化道路表面的粗糙度。IRI的值越低，表明道路表面越平滑；反之，值越高则表明道路表面越粗糙。这个指标对车辆的行驶舒适度和安全性有直接的影响。通常用来评估道路状况，指导维修和施工决策。

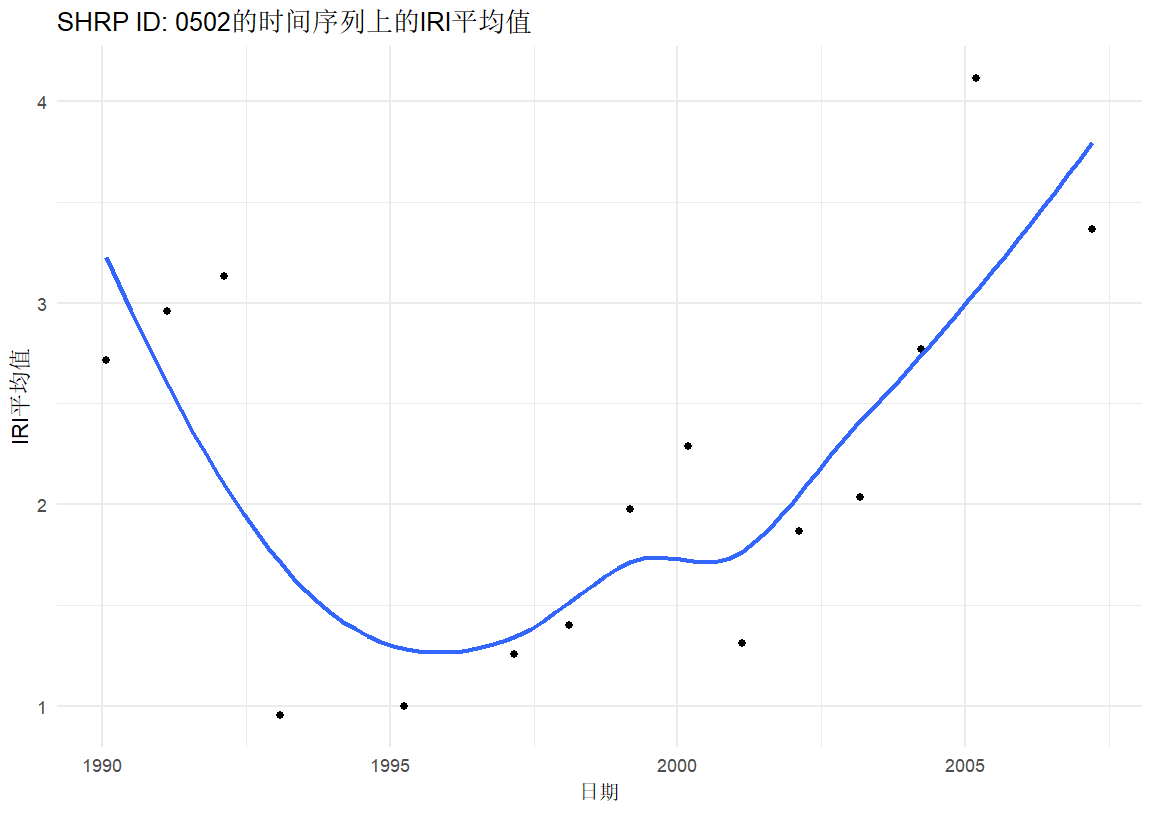


图 3 0502的时间序列上的IRI平均值

首先，对于SHRP\_ID为0502的道路段，从1990年至1995年的数据表现出IRI指数有一个显著的下降趋势，从大约3.3下降至1左右。IRI指数的下降通常与道路维护措施的实施，比如重新铺设路面、填补坑洼和裂缝处理等有关。在这种情况下，较低的IRI值表明在这5年的时间里，道路管理部门可能进行了有效的道路改善工作，从而提高了车辆的行驶舒适度和减少了行驶时的振动与噪声。

然而，从1995年至2007年，IRI指数的整体上升趋势表明了道路表面逐渐变得粗糙。道路表面的粗糙化通常与多种因素相关，如重载车辆的频繁通行、恶劣的气候条件以及时间的推移导致的材料老化等。增加的表面粗糙度可能导致车辆悬挂系统过度磨损，增加轮胎与路面的摩擦，从而可能增加燃油消耗和减少行车安全性。此外，高IRI值与增加的事故率相关联，特别是在湿滑路况下，因为车辆的操控性和制动效能可能受到影响。

从科研的角度出发，对IRI的长期跟踪分析不仅可以帮助理解特定道路段的状况，还可以为道路养护策略的制定提供量化的依据。对于该道路段，在1995年之后的上升趋势应当引起管理者的关注，可能表明需要更频繁的养护干预或甚至是彻底的重建工程。此外，IRI的变化趋势也可能反映了地下基础设施的老化程度和道路材料的长期性能。

从政策制定和资源配置的角度来看，IRI数据提供了一种评价道路资金投资回报的手段。如果投资在道路改善上的资金能够显著减少IRI，那么相关的成本就能够通过降低车辆运营成本、减少事故发生和提高通行效率等途径得到回收。

使用2017年CRSS数据集： 获取数据集accident和person的交集。

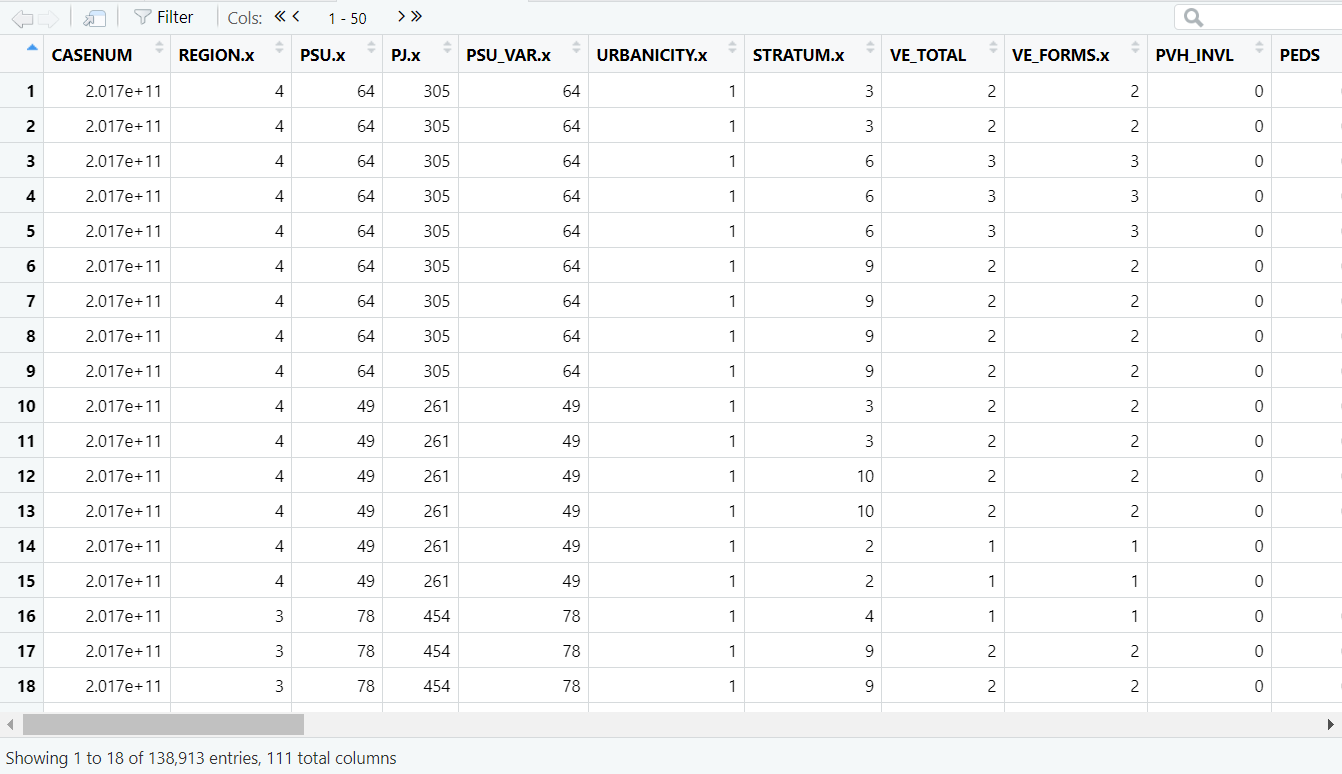


图 4 交集

按伤害严重程度（INJ\_SEV）统计观测总数。

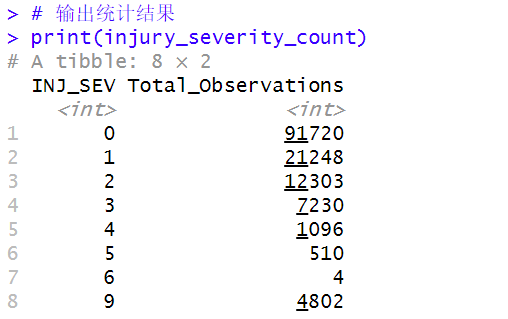


图 5 统计观测总数

并报告结果的尺寸和右侧数据集中一个变量的缺失值数量。

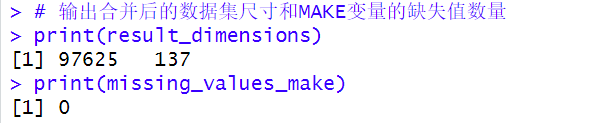


图 6 尺寸以及缺失数量

使用file.choose()函数让用户选择文件，不给定固定路径，解决固定路径的问题，以下是R语言完整代码。



作业二：

将你的作业推送到一个git仓库中（额外问题，可得5分） 提供你的仓库链接以及你的仓库截图。你可以使用集成开发环境或git的命令行界面 Rstudio旁边的“连接”和“教程”标签页中有一个Git标签。

关于本课程的建议，本课程的内容（数据处理技术、数据应用算法等） 编程语言与编程工具。

仓库链接：https://github.com/321980r/jiaotongdashuju.git

上传文件：

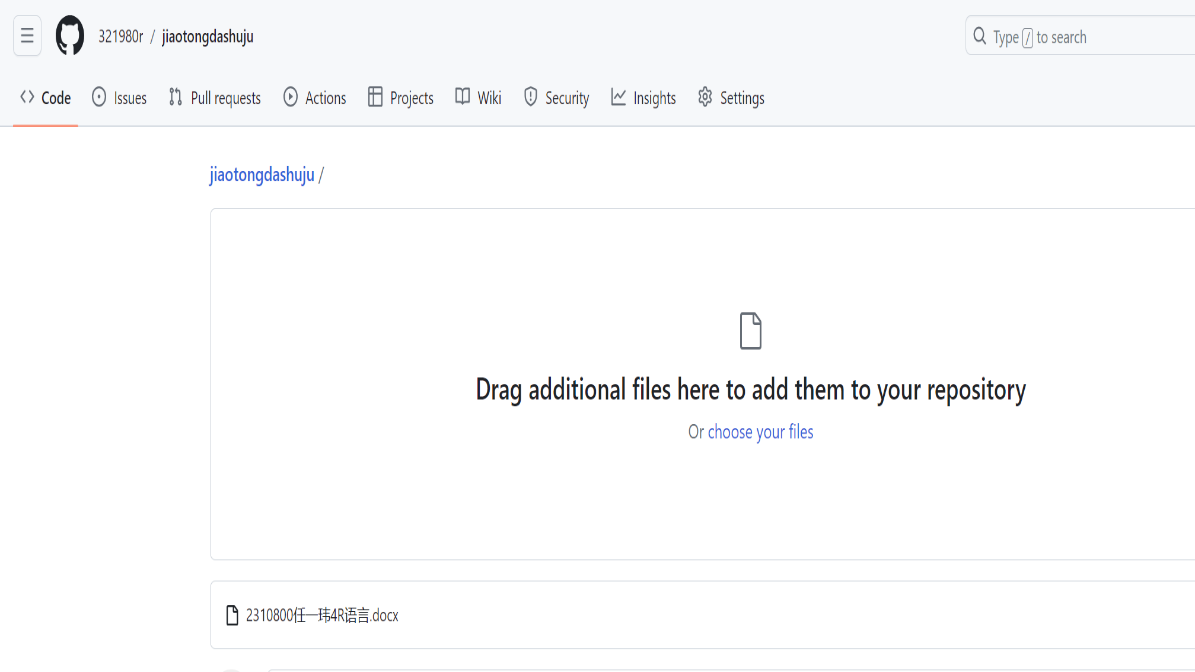


图 7 上传文件

仓库截图：

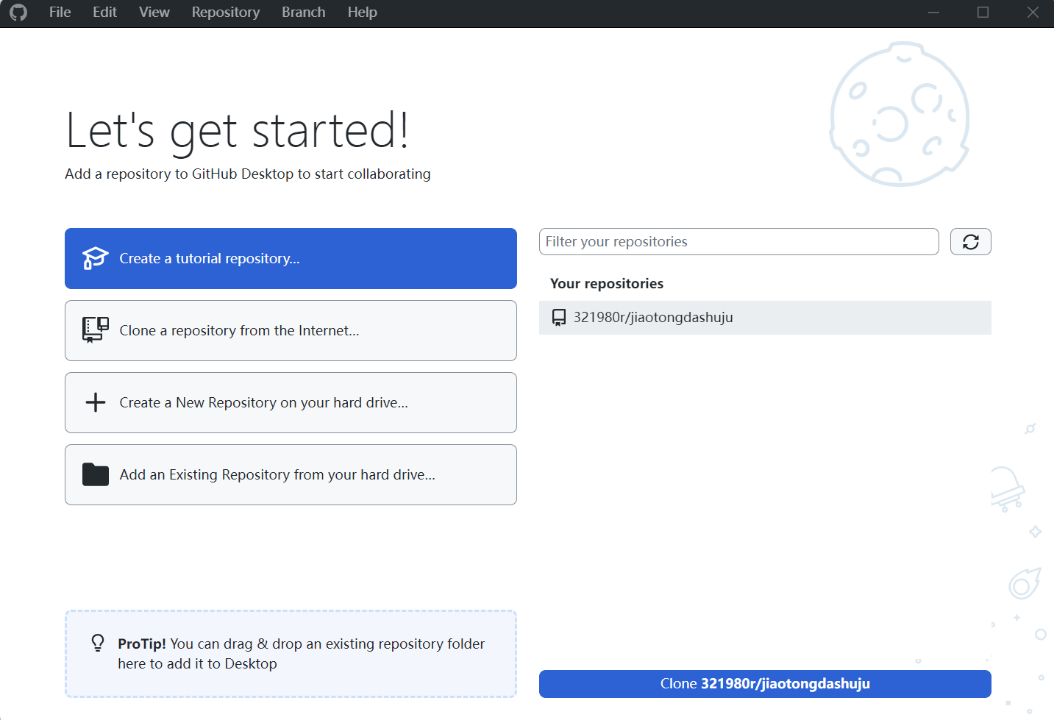


图 8 仓库截图

关于本课程的建议：

首先，我想表达对这几节R语言课程的喜爱。在课堂上，我对数据处理的技术和数据应用的算法有了更深的认识，这对我日后的科研工作将产生深远的影响，可能在数据分析的时候会尝试用R语言替代python。

一个建议是，可以增加一些关于R语言在现实世界数据分析中的应用案例，尤其是在交通运输工程领域研究中的实际应用。例如，交通系统分析、交通流量预测、道路安全研究等方面，从处理真实数据集入手，学习如何导入、清洗、变换数据，进而使用R语言的多种统计模型来进行假设检验、回归分析、时间序列分析等。通过结合交通运输领域的专业知识，不仅能学习到如何编写代码，更重要的是理解背后的数据分析逻辑和方法论。

此外，如果可能的话，引入一些关于R编程语言最佳实践的讨论，比如代码优化技巧、编写高效R代码的策略，以及如何使用RStudio和其他工具来提高编程效率，这些都将是非常有价值的。了解如何利用版本控制系统，比如Git进行协作编程，也是一个非常重要的技能，我相信同学们也会从中受益。

最后，课程中如果能提供更多关于RMarkdown的使用讲解，包括如何整合代码和分析结果到可交付的报告中，将使我们能够更加专业地展示工作成果。

作业三：

欢迎使用R Markdown文件！（去帮助 -> Markdown快速参考）或 • R Markdown：权威指南（点击链接到页面） • 对于R markdown的解决方案将额外给予五分奖励。