# 0013 reading

## Week 3

### 50 Years of Date Science

这是一篇学术论文

它反映了数据科学的演变，追溯了他的历史，将其于传统统计进行比较，并为该领域的未来方向提出了想法。他没有提供新的证实数据或实验，而是提供了对该领域的批判性反思和思考其发展的框架。

**结构**

* **介绍：设定论文的动机，并将数据科学的概念作为一门独立的学科提出**
* **历史视角：探索数据科学的起源和演变，将其根源追溯到john tukey在探索性数据分析方面的工作**
* **统计于数据科学的比较：重点介绍传统统计学与新兴科学领域之间的主要区别**
* **未来挑战：该领域在未来几十年可能的发展方式**

**关键思想**

* **数据科学作为一个领域：donoho认为，数据科学不仅仅是统计学的重塑，而是一门独特的学科，包含由大数据和计算能力的兴起驱动的新工具、技术和范式**
* **对狭隘统计观点的批评：批评传统统计学过于关注形式的模型，没有接受机器学习和数据科学中更广泛的数据驱动方法**
* **算法和计算的作用：本文强调了计算机和算法作为数据科学核心不放呢的重要性，这与经典的统计方法不同**
* **探索性数据分析（EDA）：以Tukey的EDA理念为基础，Donoho强调了数据探索而不是拟合模型如何成为数据科学家工作的核心**

**主要贡献**

* **Donho的论文充当概念路线图，讲数据科学定义为一个独立于传统统计的领域，并强调关键的方法论和哲学的差异，有助于数据科学应该包含的持续辩论，为该领域未来发展制定了愿景**

**它响应了什么差距？**

* **解决传统统计学家于更广泛的数据科学社区人员之间的理解差距，特别是在算法重要性和探索性方法的作用方面**
* **涉及将数据科学视为一门独特学科的必要性**

**剩余问题**

* **上下文依赖性：虽然donoho认为数据科学是不同的，但数据学、机器学习和同级之间的界限仍然不稳定**

**未来工作**

* **Donoho建议数据科学家的教育需要持续发展，倡导纳入道德考虑和跨学科方法**
* **呼吁更加关注可重复性的数据来源、为该领域的关键挑战**

**结论**

**这篇论文为数据科学作为一门学科的构成辩论提供了反思和概念性的贡献。建立在历史基础之上，并回应对其与统计学和其他计算机领域关系的理解差距。虽然提供了有价值的见解，为未来关于方法论挑战和跨学科合作的工作留下了空间。**

### Quantitative methods III: Strength in number?

讨论地理学定量方法，没有提出原始的实证发现，而是批判性的评估了该领域并提出值得反思的领域。

**结构**

* **为讨论设定背景，并回顾在人文地理学中使用定量方法的最新趋势。**

**主要部分**

* **对大数据和数据分析的日益重视**
* **学科中定量和定性方法之间的平衡**
* **关于数值方法主导地位的批判性观点**
* **结论：反思过度依赖定量方法的潜在风险，并对未来提出建议**

**关键思想**

* **大数据和量化：本文讨论大数据日益增长的作用，以及大量信息可用性如何导致定量技术在地理学中的使用增加**
* **对定量优势的批评：弗拉克琳批评定量方法的主导地位，认为虽然数学可以提供力量，但他们也可以掩盖重要的背景和定性见解，质疑这种对数据的依赖是否会忽视人类经验和社会空间过程的细微差别**
* **平衡定量和定性方法：本文强调需要平衡定量方法与定性方法，以避免对地理现象的一维视图**

**剩余问题**

* **过度依赖数字：富兰克林强调定量方法本质上是客观或优越的假设，在人类经验和社会空间动态复杂且难以量化的情况下，这一假设可能并不成立**
* **实用上下文：阅读材料与使用定性和定量方法来理解人类行为和社会过程的领域尤其相关，例如城市研究、社会地理学或政策研究，在这些情况下，平衡的方法对于捕捉数字趋势和个人体验至关重要**

**对未来的工作**

* **作者建议未来的工作应该探索如何更好地将定量方法与定性见解相结合，呼吁方法论多元主义，其中这两种方法都受到重视并结合使用，以提供对地理问题的更全面理解。**

**结论**

**本文对定量方法在人文地理学中的作用和讨论做出了反思和批判性贡献，回应大数据和计算方法的日益重视，并警告不要过度依赖这些方法。**

### HOUSING AND INEQUALITY IN LONDON

伦敦作为全球最繁华的城市之一，拥有高收入和多元文化，但同时也存在显著的社会经济不平等。报告旨在深入探讨伦敦内部的收入差距、贫困水平和社会流动性，为政策的制定者提供数据支持，促进社会公平。

**目的**

* 揭示收入不平等的现状和趋势：分析伦敦内部收入分配状况，比较不同地区和人群之间的差异
* 评估贫困水平：识别哪些人群和地区受到贫困影响，以及导致贫困的主要因素
* 探索社会流动性：研究个人在社会经济阶层中向上向下移动的机会和障碍
* 提供政策建议：为减少不平等和促进社会流动性提出具体的政策措施

**方法论**

* 数据收集与分析
* 地理空间分析
* 趋势比较
* 文献综述

**主要发现**

* 收入不平等持续加剧

1. 高收入群体增长
2. 低收入群体滞留
3. 顶端百分之1现象

* 贫困率居高不下

1. 相对贫困水平：约有28%伦敦居民生活在相对贫困线以下
2. 儿童贫困问题
3. 区域差异

* 住房成本导致经济压力

1. 高昂房价和租金
2. 住房短缺
3. 无家可归者增加

* 就业市场两极分化

1. 高技能与低技能工作：就业机会集中在高技能、高薪职位和低技能、低薪职位，中等技能岗位减少。
2. 失业与不稳定就业：某些地区的失业率较高，许多人从事零工或临时工作，缺乏就业保障。
3. 族裔和性别差异：少数族裔和女性在就业市场上面临更多障碍，薪酬和晋升机会受限。

* 教育和技能不平等

1. 教育资源分配不均：优质学校主要集中在富裕地区，贫困地区的教育质量较低。
2. 技能鸿沟：低收入群体获得高等教育和职业培训的机会有限，限制了就业选择。
3. 早期干预不足：幼儿教育投资不足，无法在早期阶段弥合差距。

* 健康和技能不平等

1. 寿命差异：富裕地区与贫困地区的居民预期寿命可相差数年。
2. 健康风险因素：低收入群体更容易患上慢性疾病，如糖尿病和心脏病。
3. 医疗服务可及性：贫困地区的医疗设施和服务质量相对较差。

**结论与贡献**

1. 不平等的系统性问题：伦敦的不平等并非孤立现象，而是由经济、社会、政策等多重因素共同作用的结果。
2. 对社会发展的影响：持续的不平等削弱了社会凝聚力，阻碍了经济的包容性增长，可能引发社会不稳定。
3. 需要综合性政策干预：单一政策无法解决复杂的不平等问题，需要从住房、教育、就业、健康等多个领域协同推进。

* 贡献

1. 提供全面分析：整合了大量数据和研究，全面描绘了伦敦不平等的现状。
2. 揭示关键问题：明确指出了影响不平等的主要因素，为政策干预指明了方向。
3. 倡导政策改革：提出了具体的政策建议，强调了政府和社会各界的共同责任。

**政策建议**

* **住房政策改革**

1. **增加经济适用房供应：政府应投资建设更多的经济适用房，满足低收入家庭的需求。**
2. **控制租金上涨：实施租金管制措施，防止租金过快上涨。**
3. **支持租房者权益：加强对租房者的法律保护，确保其居住稳定性**

* **提升教育机会**

1. **投资贫困地区教育：增加对落后地区学校的资金支持，改善教学设施和师资力量。**
2. **扩大早期教育：提供免费的高质量幼儿教育，促进儿童在关键发展阶段的成长。**
3. **推进终身学习：为成人提供职业培训和继续教育机会，提升技能水平。**

* **促进公平就业**

1. **创造就业机会：鼓励企业在高失业率地区投资，创造新的就业岗位。**
2. **消除就业歧视：制定反歧视政策，确保少数族裔和女性获得平等的就业机会。**
3. **支持小企业和创业：提供资金和技术支持，鼓励小型企业发展，带动就业。**

* **改善健康服务**

1. **加强初级保健：在贫困地区增加医疗设施，提供预防性保健服务。**
2. **推广健康教育：提高公众对健康生活方式的认识，减少疾病风险。**
3. **缩小健康差距：针对高风险群体实施专项健康计划。**

* **税收和福利制度调整**

1. **提升福利支持：提高社会福利的水平和覆盖面，保障基本生活需求。**
2. **税收公平化：调整税收结构，增加高收入者的税负，减轻低收入者的压力。**
3. **鼓励慈善和社会投资：通过税收优惠，激励企业和个人参与社会公益事业。**

**局限性与未来方向**

* **数据局限性：报告的数据可能无法反映最新的社会经济变化，需要持续更新和监测。**
* **缺乏长时间序列分析：对不平等趋势的长期变化缺乏深入研究，未来应加强历史数据的分析。**
* **政策实施评估不足：需要对已采取的政策措施进行效果评估，以调整和优化策略。**

**未来研究和行动方向**

* **加强实时数据收集：利用大数据和技术手段，建立实时监测系统，及时发现和应对不平等问题。**
* **深化社区参与：鼓励社区和居民参与研究和政策制定，确保措施符合实际需求。**
* **跨部门合作：促进政府各部门、私营企业和非营利组织的合作，共同解决不平等问题。**

**总结**

**这份报告深入剖析了伦敦复杂而严峻的不平等问题，既有宏观的统计分析，又有微观的现状描述。通过详实的数据和多维度的研究，报告为理解不平等的根源和影响提供了重要的洞见。其提出的政策建议具有可操作性，强调了综合治理和多方合作的必要性。对于政策制定者、研究者和社会各界而言，这都是一份具有重要参考价值的报告。**

## **Week 4**

### Politicizing Space And Place

梅西试图重塑我们对空间和地方的理解，反对这些概念的静态、固定观念，而是提出它们是动态的、关系性的和政治性的。她的工作挑战了既定的地理和社会思想，并促进了一种更进步、更多元化的空间性观点。

**结构**

* **文字通过几个关于空间和地方的理论讨论，批判了对这些术语的传统、固定的理解，而是提出讲空间视为关系和政治的观点**
* **Massey还引入空间是通过社会关系和权力动态产生的概念**
* **强化了他关于空间和地方应该政治化的论点，呼吁以新的方式将这些术语视为流动的、开放的和受全球力量影响的。**

**关键思想**

* **空间关系：Massey挑战观点，太空是一个中立、空的容器，事情在这里发生，相反，他提出空间永远是关系型的，通过人、地方和事件之间的互动产生。**
* **空间与地方的政治性质：梅西认为，空间和地方本质上是政治性的，他们受到权力关系、不平等和斗争的影响，并不受到不同社会群体的争斗**
* **挑战固定主义和本质主义的地方观：批评了讲地方视为固定的、有界限的或本质的事物概念。相反，梅西主张将地方视为动态的，通过与更广泛的全球进程的互动而不断变化**
* **全球化和空间政治：他讨论了全球化如何改变我们对地方的理解，使本地和全球复杂化，地方不是孤立的或独立的，而是通过全球人员、货物和思想的流动相互联系**

### **Date Feminism**

Data Feminism. Ch.3, On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints

这是一本对数据科学和女权主义理论的交叉点做出理论和概念的贡献，阐述通过女权主义看待数据的重要性，讨论数据收集、分析和解释中嵌入的权力结构

**主要部分**

**每个部分都涉及数据科学的不同女权主义原则或方法，例如权力动态、交叉性以及数据工作中对参与方法的需求**

**案例研究或示例：作者通过真实世界的例子来说明如何将女权主义原则应用于数据实践，使抽象的概念更加具体**

**关键思想**

* **女权主义的数据方法：数据科学受到现有权力结构的影响，并且通常反映了系统性的不平等，通过应用女权主义原则，我们可以发现和挑战这些偏见**
* **交叉性:核心女权主义原则之一是交叉性，它承认个人以多种交叉的方式（种族、性别、阶级）经历压迫。**
* **挑战数据中的权力结构：权力在数据收集、分析和解释中如何运作的重要性，并指出谁控制数据往往决定了谁的声音被听到，谁的经历被边缘化。**
* **数据是证据的工具：女权主义原则处理数据时，可以通过强调不平等和倡导边缘化群体来用作实现社会正义的工具**

# **week 5**

1. 阅读 D'Ignazio 和 Klein （[2020a）（第](https://jreades.github.io/fsds/sessions/week5.html#ref-dignazio:2020ch5) 5 章）反思：

* 《数据女权主义》的作者提出了从“数据向善”到“数据共同解放”的转变。使用源中的示例解释这两个框架之间的核心区别。

1. 数据向善：这种框架的核心是好意的数据分析，即数据科学家或组织通过数据来解决社会问题，eg，通过数据分析来识别贫困地区，以期提供帮助。典型的数据向善的项目可能由技术专家领导，主要侧重于数据的科学分析，而非社区的直接参与。
2. 数据共同解放：这种框架更进一步，主张赋予数据参与者自主权，确保数据收集、处理和决策过程都充分反映社区声音和需求。它提供“共创”，即社区成员不仅是受益者，更是参与者和所有者。这意味着项目目标是通过数据来促进社区成员和数据科学家共同的解放，强调协作和自主性。

* 比较和对比反驱逐映射项目 （AEMP） 和 Eviction Lab 采用的数据科学方法。他们的数据收集、清理和分析方法如何反映不同的价值观和优先事项？这些选择有哪些道德含义，它们与“认知暴力”的概念有何关系？

1. AEMP的数据收集和分析更为社区化，合作性强。它于社区组织密切合作，关注数据背景和文化，保留数据的混乱性和原始特征，以呈现出更真实的社区声音。这种方法符合反驱逐和反种族主义的立场，也意在突出社区对抗驱逐的复杂性和抗争的过程。
2. Ecivtion Lab：该项目采用了更标准化的清理数据的方法，倾向于获取更大规模、更清洁的数据，这使数据更适合全国性比较，但可能忽视了地方独特的文化和社会背景，eviction lab这种清理方法代表了规范化和标准化的数据科学，可能导致“认知暴力”，既在数据中抹去多样化的社区声音和复杂性，使数据与真实背景分离。

* 用于描述数据科学家的主要隐喻，例如“独角兽”、“巫师”和“摇滚明星”，可能会加剧数据工作中的权力不平衡。分析这些隐喻的含义，并讨论数据科学劳动的替代框架如何促进更具协作性和公平性的实践。

1. 隐喻分析：形象强调了数据科学家的稀缺性和特殊技能，隐含着技术领域的单一权威形象。这种比喻使得数据工作变得孤立、忽视团队协作，并强化个人主义倾向。此外，这些隐喻多址向男性角色，带有潜在的性别和种族偏见。
2. 替代框架：若采用“协作者”、“伙伴”或社区导向型设计者等替代框架，将有助于在数据科学实践中建立协作和公平性。这些框架鼓励数据科学家将自己视为社区的一部分，以共同目标为驱动。这不仅有助于知识转移，还能为社区赋权，促进更具有包容性的数据实践。

## Data Feminism. Ch.5, Unicorns Janitors, Ninjas, Wizards, and Rock Stars

Here’s a detailed analysis of Chapter 5 of *“Data Feminism”* by Catherine D’Ignazio and Lauren F. Klein based on your questions:

第五章是概念、理论和证实贡献的混合体，因为它提出女权主义概念，也使用现实世界的案例研究和示例来应用和说明这些概念。

第五章主要是概念和理论，提供了关于如何将女权主义理论应用于数据收集、分析和表示的见解，也通过包括在实践中应用女权主义原则的数据项目示例来做出实证贡献。

### 结构

首先对关键主题进行了一般性阐述，确立了将应用于数据的女权主义原则

本章分为不同部分，分解了解女权主义原则，并将其应用于数据收集、分析和表示问题

使用case study研究来说明女权主义对数据实践的干预如何解决偏见、排斥或不公平的问题。

### 关键思想

* 挑战数据科学的客观性：中心思想之一是对数据科学的客观性的主张的批评。作者认为，数据从来都不是中立的；它总是反映了收集、分析和解释他的人的偏见。
* 数据表示功能：数据的表示方式有可能复制或挑战现有的权力结构。作者讨论了女权主义的数据表示方法如何阐明在传统数据实践中经常被排除或歪曲的边缘化声音和经历。
* 数据公正和公平：一个关键主题是数据实践必须积极致力于实现社会正义和公平的想法。女权主义数据实践不仅涉及识别偏见，还涉及重新思考如何收集、分析和表示数据以确保公平性和包容性的整个过程。
* Collective good的数据：作者促进使用数据来增强集体赋权，特别是对于边缘化社区，并倡导数据科学更具参与性，确保数据收集影响的人参与到这一过程。

### 主要贡献

通过将女权主义原则作为重新思考数据科学实践的框架做出重大贡献，批判了数据客观性的神话，并倡导采用更具包容性、参与性的方法，优先参与公平和正义。

### 它响应了什么差距？

本章回应了数据科学中的空白，其中权力、代表性和偏见问题经常被忽视或理论化不足。通过引入女权主义观点，作者挑战了传统方法，并呼吁重新思考如何收集、分析、共享数据。

### 结论

数据不是中立的，应用女权主义框架来挑战现有的权力结构并确保数据实践具有包容性和公正性至关重要，女权主义数据时间对于创造一个更公平的世界至关重要，在这个世界中，数据可以满足所有人的需求，尤其是边缘化群体。

### 剩余问题

* 在不同上下文中的适用性：在大型机构或企业环境中应用女权主义数据原则可能存在困难，因为在这些机构或企业环境中，效率和利润往往优先于道德问题，尽管作者提出了令人信服的论点，但将这些原则转化为企业或政府环境可能会遇到阻力。
* 需要实践指导：虽然本章提供了一些理论见解和一些示例，但在为数据科学家或组织提供有关如何在日常工作中实施女权主义数据实践的详细应用策略方面可能存在差距。在实际数据项目中应用这些原则的更具体的工具或框架可能会有所帮助。

### 其他背景

这些想法可以应用于广泛的环境，特别是涉及边缘化社区，社会正义问题的环境，例如公共卫生数据、警务和刑事司法，或社交媒体平台中的算法偏见。例如，在医疗保健的背景下，女权主义数据实践可以确保在研究和政策制定中不会忽视代表性不足的群体（例如女性、少数族裔、LGBTQ）的需求。

### 结论

“数据女权主义”第五章为通过女权主义视角重新思考数据实践提供了一个关键框架，挑战了数据客观性的神话，强调了数据科学中的包容性、公平性和社会正义的重要性。本章的贡献是概念性的，也是实证性的，它批判了传统方法，并提供女权主义原则作为替代方案。虽然提出的想法很有价值，但在某些机构环境中应用这些原则存在挑战，未来的工作可能会探索更应用的策略，以便在各种环境中实施女权主义数据实践。

## Teaching introductory GIS programming to geographers using an open source Python approach

阅读 Etherington （2016） 反思：

为什么 Etherington 提倡使用开源 Python 方法向地理学家教授 GIS 编程？

Etherington 提倡开源 Python 主要是因为 Python 具有跨平台、免费和可扩展性优势。与商业软件相比，Python 没有财务或硬件障碍，更适合在任何设备上学习和使用。此外，Python 是目前 GIS 编程的主流语言，特别是对于 ArcGIS。这种开源方法使学生能够掌握普遍适用的编程技能，提供更灵活的学习和应用机会，即使他们没有编程背景也能逐步上手​

描述 Etherington 用于构建其 GIS 编程课程的六个核心空间概念，并思考它们是否提高了我们对这两个概念的理解。

Location：让学生了解地理坐标系统和位置再投影，使用pyproj python包进行坐标转换。

Neighbouthood：涉及空间距离的计算，要求学生编写之间的函数来计算欧几里得距离，增强他们对临近性概念的理解。

Event：关注随实践变化的空间事件，结合detetime模块和matplotlib包来帮助学生理解事件在空间中的变化和可视化。

Field：引入连续现象的栅格数据结构，使用GDAL包读取和写入栅格数据，使学生能够处理大规模空间数据，

Object：解释点、线和多边形等离散空间对象的矢量数据结构，使用OGR包处理矢量数据。

Network：网络数据结构用于表示位置之间的连通性，networkX包被用于分析和绘制地理网络。

Etheringtong 是一篇学术论文，旨在以地理为中心的受众分享有关使用开源python方法教授gis编程的见解

贡献主要是实证和教育学，借鉴作者的教学经验，并提出向没有强大编程背景的学生有效教授gis编程的方法。

### 结构

* Intro：本文首先解释了教授gis编程的基本原理，特别是使用python作为开源工具，强调了现代地理和gis职业对编程技能的需求。
* 方法：解释了etherington课程中gis编程的入门结构。讨论教学方法，包括动手编码练习、特定python库的使用以及学生面临的挑战。
* 挑战和解决方法：提供一个部分来解决学生遇到的常见挑战，尤其是当他们几乎没有编程背景时，并提供了如何克服这些障碍的解决方案。
* Case study：使用一些实际案例演示python在解决实际gis问题中的应用。
* 结论：对该方法的成功进行了反思，并提供进一步改进gis编程教学的建议。

### 关键思想

* 编程在gis教育中的重要性：地理学家和gis专业人员越来越需要编程技能。Python作为一种开源语言，被强调为用于这些目的的可访问和广泛使用的工具。
* 教育中的开源工具：etherington 倡导使用python等开源编程语言，使gis教育更易于广泛使用的工具。
* Gis教育更易于访问且更实惠
* Gis编程教学策略：讨论了有效的教学方法，包括动手学习，协作解决问题和增量学习方法，以帮助学生建立对编程的信心。
* 向地理学家教授程序涉及的挑战：学生对学习编程的抵制或他们最初在语法和逻辑方面的挣扎，以及教育工作者如何通过结构化教学方法解决这些问题。

主要贡献：通过提供使用python教授gis编程的使用策略来做出贡献，这些策略为没有编程经验的地理学家量身定制，强调循序渐进的支持学习过程，并演示使用python开源库进行gis工作的好处。

响应差距：解决gis教育领域的差距，其中许多学生，尤其是地理课程的学生，缺乏编程技能，但越来越多地需要他们从事与gis相关的工作，本文回应对平易近人教学方法的需求，这些方法可以弥合地理和编程之间的差距。

### 结论：python是在gis教育中引入编程的高效工具，并且通过适当的教学支持，学生可以克服最初的编程困难，python开源的性质使其成为软件预算有限的机构的宝贵资源

剩余问题：

* 所介绍的教学方法可能不容易转化为所有机构环境，尤其是学生先前编程知识水平不同的机构环境，例如某些课程需要更深入的基础编程课程，才能账务用于gis的python。
* 学习曲线和资源：尽管python是一种相对容易使用的编程语言，但学生可能仍然会发现学习曲线陡峭，etherington通过推荐增量学习和动手实践来解决这个问题，但并没有完全解决如何在更大、更多样化的课堂中扩展这种方法的问题。

## Cars and socio-economics: understanding neighbourhood variations in car characteristics from administrative data

阅读兰斯利 （2016） 反思：

为什么 Lansley 要介绍一篇对营销和广告文献进行广泛回顾的地理研究文章？

Lansley引入营销和广告文献的回顾，是为了说明不同车型作为消费商品如何反映消费者的身份，生活方式，社会经济地位。这种消费视角帮助解释了为何某些车辆在某些社区更受欢迎。这些车辆的特性和定位通常营销来强化，以满足特地消费者群体的需求，由此车辆选择反映了居民的社会经济属性。通过这种回顾，lansley能够将汽车消费与更广泛的社会和经济特征联系起来，使得研究具有更深的社会学视角。

Lansley 如何将汽车划分为不同的类别，这种细分对理解英国的社会和经济格局有什么影响？

Lansley将汽车分为是个类别，如城市车、超小型车、家庭轿车、suv、豪华车等，此划分基于车辆的结构、用途和市场定位，例如家庭车通常在经济型社区更受欢迎，而高端豪华车则集中中富裕地区。通过这种细分，研究揭示了英国各地不同社会经济群体的分布，例如在伦敦等大城市，较小经济型车辆更受欢迎，而在富裕地区，豪华车，suv更常见，这种细分提供了一个有效框架，用于分析车辆类型与社区经济状况之间的联系，揭示出区域经济差异空间格局。

使用行政数据（例如 DVLA 汽车登记数据）来研究社会经济趋势有哪些限制？

* 数据隐私和聚合限制：为了保护隐私，数据只能以聚合形式提供，可能会掩盖个体或更细微的社区差异。
* 数据质量和一致性：行政数据通常为管理目的而非研究设计，因此在数据完整性和一致性上可能存在问题，eg某些车辆可能注册在公司名下，且不反映真实的家庭车辆所有权。
* 数据缺乏社会经济背景：虽然车辆登记数据提供了车辆类型和分布信息，但缺乏关于车主的具体经济和人口信息，需要与其他数据（人口普查）结合分析才能准确解释社会经济趋势。

### 结构

* Intro:介绍英国车辆所有权的普遍性和其作为社会经济指标的可能性。强调了基于政府部门的车辆登记数据可能提供的社会经济洞察，特别实在人口普查数据发布之前的空白期内。
* 背景：车主选择车型的因素，包括日常出行需求，代替交通方式以及经济能力。还介绍了过往研究对汽车所有权作为社会经济地位指标的关注，指出了人口密度和城市环境对汽车类型的影响。
* 数据与方法：数据来自英国驾照和车辆管理局，并将数据按地理层次进行分组。文章对车辆进行了分类，以分析每种类型的地理分布。
* 分析与讨论：通过对比不同地区的汽车类型、车龄等，展示了各社区在汽车类型上的明显分布差异，特别是在伦敦等大城市，车辆类型表现出明显的社会经济差异。
* 多变量分析：运用典型相关分析（CCA）来检验不同变量间的复杂关系。结果表明，高收入社区倾向于拥有更昂贵的车辆，低收入社区倾向于经济型车辆。

### 核心概念

汽车类型分类：将车辆按尺寸和市场细分为10个类别，例如城市车（city car）、家庭轿车（family car）、豪华车（luxury car）等。这种分类法帮助分析了不同类型车辆在各社区的分布情况。

社会经济地位与汽车类型的关系：研究重点在于探讨各社区的车辆类型是否能够反映居民的社会经济地位。例如，高档车通常出现在收入较高的社区，而较旧或便宜的车则分布在经济水平较低的区域。

### 结论

Lansley 发现，社区汽车的地理分布与社会经济特征有着显著的关联。不同车型在各区域的分布反映了居民的收入、生活方式和社会地位，例如豪华车在富裕社区的集中分布。而通过DVLA数据，研究展示了行政记录数据在小区域内研究社会经济差异的潜力。

本研究支持了使用车龄、车辆类型等细分特征作为衡量社会经济差异的新方法，提供了一种替代传统人口普查数据的补充方式，尤其适用于两次人口普查之间的时间段内。

阅读三个文献回答：

这些来源凸显了分析中对“干净”数据的需求与数据清理可能掩盖多样性之间的紧张关系。考虑来自说明这种紧张关系的来源的具体示例，并分析不同的数据清理方法如何影响数据中“异常值”的表示（无论是边缘化群体还是极端值）。

#### 数据清理和多样性之间的紧张关系：

* AEMP vs. Eviction Lab 数据清理方法：在 D'Ignazio 和 Klein 的讨论中，AEMP 采用了更“混乱”的数据方法，与社区紧密合作，以保留数据中的多样性和背景。与之相对，Eviction Lab 采用更“干净”的方法，去除不必要的细节，以便于大规模分析。然而，这种方法有可能将边缘化群体的声音淹没在标准化数据中。AEMP 的方法允许保留独特的社区故事和复杂性，而 Eviction Lab 的清理方法可能使异常值（如无房家庭或遭遇多次驱逐的家庭）变得不显眼。
* Lansley 的车辆数据清理：在 Lansley 的研究中，为了符合数据隐私和质量的需求，车辆数据在公开前被高度聚合，例如，某些公司车辆被从家庭数据中剔除。这种清理限制了对不同社区车辆所有权的全面分析，因为企业车辆在某些区域也可能反映当地的经济活动或居民生活方式。清理后，研究数据中代表性下降，可能影响到边缘化群体的汽车选择和使用模式分析。

#### 不同数据清理方法对“异常值”的影响：

* 极端值和边缘化群体的影响：在 Etherington 提供的教学示例中，Python 开源 GIS 编程中也面临类似挑战。例如，通过标准化和清理地理数据，可能导致一些少数群体的地理活动（如短时移动或特定地区的高度集聚现象）被视作“异常”而被排除。这些极端值和异常现象在未清理的数据中可能更能代表特定群体的行为模式或经济状况。
* 数据清理的伦理含义：无论是为简化分析还是确保数据标准，数据清理往往带有主观性选择。清理过程中，数据分析者会面临是否保留或剔除某些看似“无用”或“异常”的数据，而这些数据可能恰恰是描述边缘化群体或揭示社区差异的关键。对于 AEMP 的项目来说，保留这种“混乱性”有助于放大弱势群体的声音，而对于需要清晰模式的项目，标准化数据或许是有效分析的必须。