COMPREHENSIVE PORTFOLIO ASSESSMENT

BY

JEFFERY CHARLES SAUER

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy (Geographical Sciences)

at the

University of Maryland – College Park

2020

Contents

1 Curriculum Vitae

Jeffery C. Sauer

CONTACT INFORMATION

Center for Geospatial Information Science
University of Maryland
Room 1124 LeFrak Hall, 7251 Preinkert Dr.
College Park, MD 20740

Email: jcsauer@terpamil.umd.edu
Phone: 612 394 1115 (USA)
Web: https://jeffcsauer.github.io/

EDUCATION

University of Maryland, College Park, USA
PhD Student, Geospatial Information Science, Fall 2019 – Present

London School of Hygiene and Tropical Medicine, London, UK
MSc Epidemiology, granted November 2018

McGill University, Montreal, CAN
BA Geography (First-class honors, distinction), granted Spring 2017

EXPERIENCE

University of Maryland, Department of Geography, College Park, MD
Research Assistant and Teaching Assistant, Aug '19 - Present
Carry out dissertation-related research under the supervision of Dr. Kathleen Stewart, namely on geographic patterns of substance use. Instruct labs and manage coursework for GEOG373 (Fall 2019).

Geographic Information Center, Montreal, QC
Research Assistant Level I, Aug '18 - June '19
Helped manage big-data storage capabilities for spatial research projects at McGill University (PostGIS and PostgreSQL). Assisted faculty and students with R and Python. Conducted workshops on Python, SQL, and R.

Professors Leonardo Baccini & Franceso Amodio, Montreal, QC
Research Assistant, Sep '16 - May '19
Built and maintained custom dataset (100k+ observations) for research projects regarding United States legislative voting patterns and trade flows (STATA and R). Merged and validated datasets from 15+ secondary sources. Conducted text analysis of lobbying statements for descriptive and analytic purposes (R).

Institute for Health and Social Policy, Montreal, QC
Research Assistant Intern, Jan '17 - Aug '17
Worked under the Health and Well-Being Project to translate previous research findings into data visualizations. Curated specialized datasets for small populations from existing surveys (Aboriginal Peoples Survey, Canadian Long Form Survey) for research relating to household mobility and Indigenous health.

Indigenous Health Adaptation to Climate Change Lab, Montreal, QC
Research Assistant, Sep '15 - Oct '16
Worked under the supervision of Professor Lea Berrang-Ford to conduct statistical analyses (ArcMap, Stata) for the Uganda field site, ranging from exploratory descriptive statistics to inferential hierarchical multilevel models. Conducted bibliometric literature reviews regarding climate change (Sci2, Pajek, R).

PUBLICATIONS

Sauer J., Berrang-Ford L., Didamanya D., Donnelly B., Lwasa S., Zavaleta C. (2018). An analysis of the nutrition status of neighboring indigenous and non-indigenous populations in kanungu district, southwestern uganda: Close proximity, distant health realities. *Social Science & Medicine*. doi:10.1016/j.socscimed.2018.09.027

Breau S., Burkhart N., Shin M., Marchand Y., **Sauer J.** (2019). Is it time to start worrying about growing regional inequalities in Canada? *The Canadian Geographer*. Forthcoming 2020.

Manuscripts in preparation

Rich, S., Ba Tran, A., Williams, A., Holt, J., **Sauer, J.**, Oshan, T. (2020). arcos and arcospy: R and Python packages for accessing the DEA ARCOS database from 2006 - 2014. *Journal of Open Source Software (JOSS).*

PRESENTATIONS

Akande H.A., Cardille J., **Sauer J**. (September 2019). Autocorrelation and Spatial Non-Stationarity Effect on Bird Species Richness in Ontario, Canada. Presented at the 10th Student Conference on Conservation Science. NYC, USA.

Sauer J., Robinson B. (November 2016). The role of protected areas in forest restoration. Poster presented at: McGill University GIS Day. MTL, QC.

SOFTWARE PACKAGES

Sauer J. (November 2019). arcospy: a python translation of the R arcos module. Available via PyPI. https://github.com/jeffcsauer/arcospy.

RESEARCH TRAINING

Computer skills: R, Python, STATA, SPSS, SQL, ArcMAP, QGIS, PostGIS,

Accepted and completed as an undergraduate the Quebec Inter-University Center for Social Statistics graduate level summer short-course *Concept and application of multilevel statistical methods to health, social science and geographical research* (July 2016) taught by Professor S.V. Subramanian (Harvard).

Extended independent research project with Professor Brian Robinson (McGill) using matching methods to examine impact of protected areas on forest restoration (independently built longitudinal dataset of 50k+ observations from LANDSAT imagery, Global Forest Watch, and NASA NDVI).

EXTRA-CURRICULAR ACTIVITIES

McConnell Hall, Student Housing and Hospitality Services, Montreal, QC
Assistant Director and Resident Adviser, Aug '15 - April '17

Lived alongside first-year students as a support system for transition into university. Worked with a team to build community and provide safer spaces when sensitive issues.

McGill Undergraduate Geography Society (MUGS), Montreal, QC
Editor in Chief of Field Notes, the MUGS Academic Journal, Aug '15 - Apr '17

Recruited and led a team of editors to select undergraduate submissions for the 2015-16 and 2016-17 publication of *Field Notes.* Curated a special issue highlighting critical areas of geography, specifically feminist geography, queer geography, and colonial geography.

2 Goal Description

2.1 Original Statement of Purpose

While attending McGill University for an undergraduate Geography degree, I felt a voracity to immerse myself in the research process and develop the academic prowess required to conduct quality scientific investigation. I cultivated a data-oriented, statistical skill set that drew from the methods of health geography, comparative political economy, and GIS-based environmental modeling. However, underpinning the excited acquisition of these skills was an ever-present concern for how research informed the everyday decision-making strategies of policy makers.

This internal dialogue came to a head in my undergraduate thesis where I navigated the intersection of Indigenous land rights, the politics of recognition, and health inequalities. I analyzed and contextualized nutritional health inequalities between the Indigenous Batwa and non-Indigenous Bakiga populations of Southwest Uganda and found the prevalence of malnutrition to be markedly higher among the Batwa compared to their Bakiga neighbors. I recently published these findings in *Social Science & Medicine*. Through the thesis process I realized the complexity of the numerous pathways contributing to contemporary health distributions. Knowing that policy makers in Uganda played a key role in prioritizing certain social factors over others, I wanted a robust understanding of the approaches used in health sciences so I could quantify the impact of policy decisions on health.

I continued my education at the London School of Hygiene and Tropical Medicine in the MSc Epidemiology program. Epidemiology allowed me to pursue my affinity for health geography through the sub-discipline of spatial epidemiology while also receiving rigorous training in study design. Advancing my understanding of spatial techniques, my MSc thesis utilized semi-variograms and geographically weighted regression to investigate HIV prevalence in the Gem District of Western Kenya, surprisingly finding no evidence of spatial structure in the distribution of HIV. These results were shared with project collaborators at the regional health steward, the Kenya Medical Research Institute, to help inform policy questions of whether HIV cases were exhibiting spatial patterning.

In my PhD, I hope to join a research group that is asking interdisciplinary questions about pressing health issues impacting the United States. By applying quantitative skills, my goal is to generate original research that can reach a wide audience, contributing to high-impact journals and the public sector. As my career objective is to conduct and advocate for research that can inform national health policy, University of Maryland (UMD) is my top choice because of its established connection to the stakeholders and government agencies headquartered in Washington, D.C. The UMD Geography department's focus on student development – from geography-specific career fairs to strong presence at professional conferences – is especially

attractive as it allows for exploring diverse employment opportunities in both corporate and academic settings. A doctoral degree in Geography will provide me with an advanced, holistic approach to problem solving that is required to solve the complex policy decisions that affect human health.

2.2 Prospective Goals Since Entering Doctoral Program

The following section articulates my academic, research, and professional goals. I include some long-term professional goals that I intend to pursue after the doctoral program.

2.2.1 Academic Goals

My overarching academic goal is to become an expert in GIScience. Specifically, I want to study how GIS can be used to conceptualize and measure public health issues, namely substance use and overdose in the United States. Here I expand on four academic goals that can develop my expertise in GIScience.

I want to become fluent in the methods of spatial analysis.

Spatial analysis has become an essential toolbox for geography. Like measures of central tendency to a statistician, so too are techniques like Moran's I and Spatial Weights to the quantitative geographer. Increasingly, spatial analysis requires working across programming languages, geographic information systems, and data repositories. This family of methods is important as it is the leading approach to understanding the relationship between space and its individual, environmental, or constructed occupants. Academically, spatial analysis can be situated within the Modeling section of the Geographic Data Science (GDS) research agenda (Singleton & Arribas-Bel, 2019). From the computational implementation to the equation, I want to master spatial analysis.

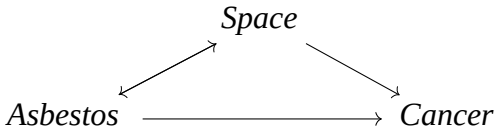
I want to champion the geospatial analysis of health outcomes.

Perhaps the most fundamental objective in epidemiology is to show how an exposure relates to an outcome (e.g. asbestos and cancer, conceptualized below).

$$Asbestos \longrightarrow Cancer$$

Yet from the moment of birth to the end of life, individuals occupy space. Throughout the life course, place and health are always connected, even if the intensity of the connection varies. Although space has a constant confounding presence (conceptualized on following page), different considerations of space in the health sciences has resulted in several fractured

subdisciplines like Health/Medical Geography, Spatial Epidemiology, Sociology of Health, and others (Rosenberg, 1998).



Each of these sub-disciplines offer techniques that fit complement the others, and I want to use my experience navigating health geography and spatial epidemiology to help unify these closely related fields. I also want to advance the way in which spatial analysts conceptualize and measure space. With the ever-increasing amount of data products, there is more opportunity than ever to question and re-conceptualize how space plays a role in the exposure-outcome relationship.

I want to create academic workflows that are reproducible and have real-world application.

Following the early work of Claerbout and Karrenbach, I want my academic work, from coursework to publications to the penultimate thesis, to be reproducible (Claerbout & Karrenbach, 1992). I believe the extra effort of making research as reproducible as possible is important for several reasons. First, reproducible analyses allow for students and other researchers to engage with learning beyond the article format. Opaque models or enticing figures can be deconstructed with the data and code. Second, fully reproducible work acts as an invitation to other interested parties to get involved with, correct, and improve the research. Third, I see the act of making research reproducible as a symbolic acknowledgement of imperfect research. Whether the imperfection is as slight as inefficient code or as large as a fundamental error in the analysis, there is ample opportunity for the academic community to improve the way in which errors are handled in the research process. Cultivating appropriate modes of response to inevitable errors in the research process is especially important when communicating science to an increasingly skeptic public.

I want to understand the philosophical system in which I carry out my research.

While I have a great affinity for critical theory, I have only limited classroom exposure to the study of critical theory specifically in the veins of Feminism, Neoliberalism, and Postcolonialism. I would like to broaden my understanding of theory and be able to articulate the philosophical system or critical viewpoint of my research. This focus on research-from-theory follows in the footsteps of previous geographers like Curry and Tuan, although the focus on the relationship between health and space draws on the work of Koch and other medical ethicists (Curry, 1996; Koch, 2017; Tuan, 1991). Additionally, as the distinct questions motivating the field of geography have been called into question before, there is a responsibility for upcoming researchers to be able to articulate the place of geographical research as it relates

to other domains like sociology, political science, and environmental science (Cutter et al., 2002). I believe that developing and practicing this articulation advances my academic maturity and helps communicate the value of a geographical perspectives to collaborators.

2.2.2 Research Goals

My research goals seek to understand what aspects of the environment – built, natural, and social – influence substance use in the United States. In the following section I describe three research goals and how I intend to realize each in my dissertation.

I want to research how spatial modeling techniques perform under different data constraints, especially in the analysis of health outcomes.

As spatial modeling continues to grow in breadth and popularity there is a need to understand the general behavior of spatial models when using different epidemiological data and study designs. For example, there are no established best practice guidelines as to what types of spatial models should be used in cohort studies versus convenience samples, or the relative benefits and drawbacks of spatial models when studying different health outcomes. This is an area of study where interdisciplinary GIScientists can lead the research agenda by systematically examining the sensitivity of spatial models to different data constraints. Within my own research I intend to deploy spatial modeling techniques to different sources of routine health data, namely administrative hospital and prescription record datasets. This research will lead to a better understanding of the behavior of spatial models for routine health datasets which are routinely used in cohort, retrospective, and cross-sectional study designs.

I want to research the integration of GIScience principles into machine learning workflows.

I am keenly interested in exploring how machine learning algorithms can be made spatially explicit. Machine learning algorithms are increasingly being used to identify optimal features for regression and classification problems. There is ample opportunity to examine how these machine learning approaches might be improved by incorporating aspects of spatial analysis like weight matrices, distance measures, neighborhood window delineation, diffusion networks and more. Classical examples of machine learning within Geography were recently reviewed by Singleton and Arribas-Bel, who note the success of unsupervised spatial algorithms like regionalization compared to a relative lack of lack of spatially sexplicit supervised learning techniques (Singleton & Arribas-Bel, 2019). In the context of GIScience for the analysis of health outcomes, I intend to explore how the explicit inclusion of space in machine learning algorithms can improve the performance of modelling tasks for public health. Examples applications of this research include forecasting areas that may have high disease burden or predicting the diffusion of a disease over space and time.

I want to use data integration to overcome methodological issues in the study of substance use.

Substance use is an incredibly complicated act with different motivations depending on the individual, substance, and societal context. Substances are usually illegal or highly restricted, and once ingested they can directly change the behavior of users from the non-using population. These factors and more make the use of a single data source unappealing when trying to study substance use (Johnson, 2014). Thus, I want to use data integration to create original data products that borrow from each other's strength to analyze geographic patterns in substance use. Specifically, I want to focus on opportunities to spatialize administrative data – such as hospital and prescription record data – and combine it with government census products to carry out novel spatial analyses. As my career develops, I see this goal maturing into the exploration and integration of unconventional datasets with ambient location ambient location – such as street sensors, communication technologies, and transportation sources – that may be able to shed light on the geographic patterns of substance use.

2.2.3 Professional Goals

I have several professional goals that I intend to initiate during the doctoral program. These goals relate to my capabilities as a researcher as well as where I would like to seek employment after the doctoral program.

I want to become a researcher that can design studies, attract funding, and execute complex research agendas.

I want to carry out studies that leverage my training in geography and epidemiology to carry out complex study designs of space and health. This is an important skill for both formulating independent and collaborative research projects. I would like to engage in the grant writing process, especially towards grants that apply to US funding institutions like the National Institutes of Health (NIH) and the National Science Foundation (NSF). Through the course of my PhD I want to improve my abilities to carry out each stage of the research process from idea formulation to publishing results. Ultimately, I want to produce academic papers, technical reports to participating non-academic institutions, and engage the next generation of researchers through the dissemination of science in public outlets.

I want to use the GIScience as a common ground for bridging different areas of research on substance use in the United States.

GIScience offers the potential to combine space, time, and human phenomenon into unified analytical and data processing frameworks (M. P. Kwan, 2013). Although GIS is used extensively by GIScientists, health professionals, criminologists, social scientists and more,

there remains several relatively siloed areas that produce knowledge on substance use. Within the United States, these areas are personified with separate government agencies addressing substance use (e.g. the Drug Enforcement Agency and the National Institute for Drug Abuse). I want to leverage the common use of GIS across disciplines to promote interdisciplinary research on the geographic context of substance use. Such research might manifest as the integration of routinely collected datasets on substance use, collaborations between researchers at different agencies to develop spatial models, or new visualizations and data-driven tools for use by local actors. During the course of my doctoral degree I hope to develop my GIScience expertise and professional skills so that I can realize these grand goals.

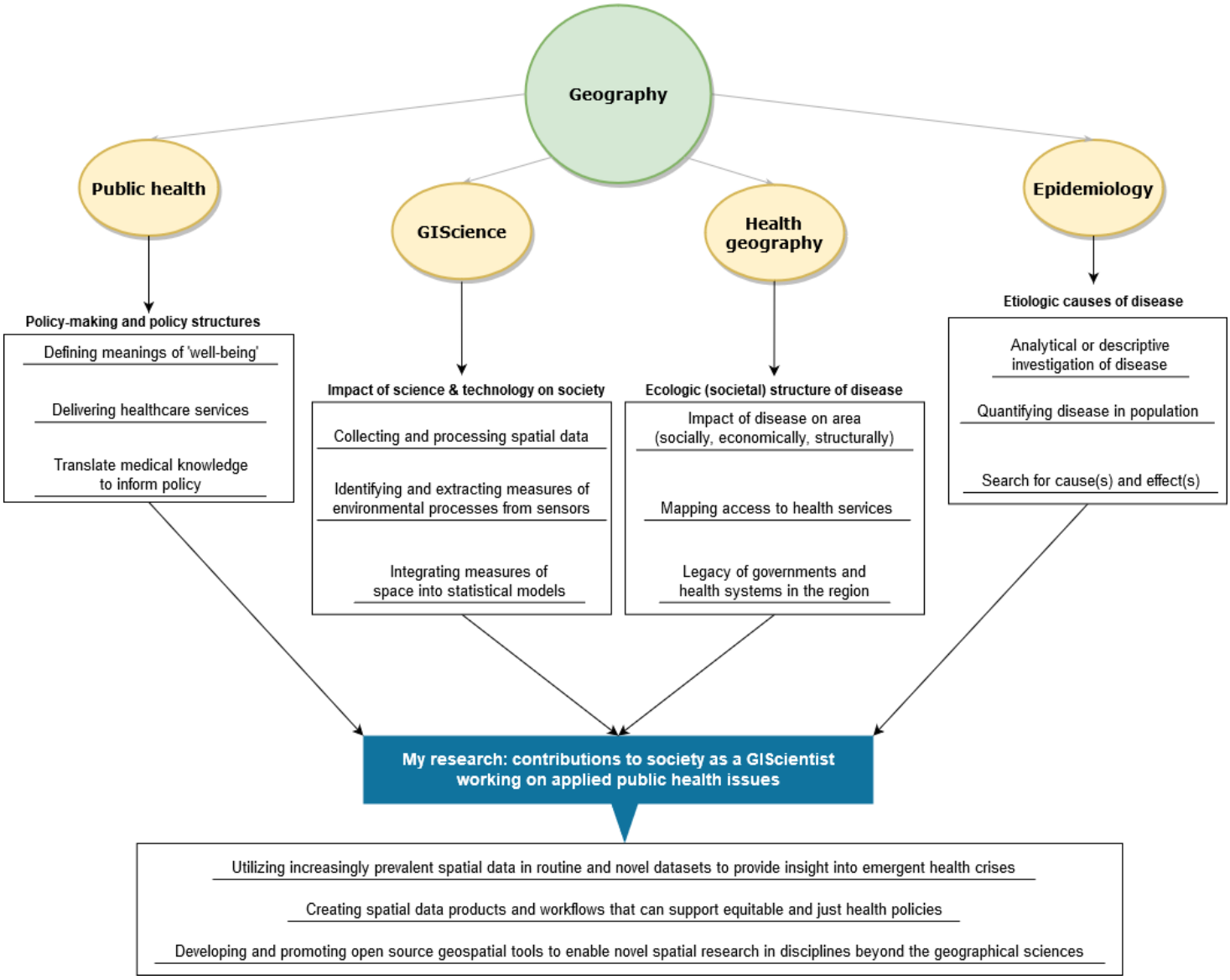
I want to use mixed methods research to capture the complexities of substance use.

Quantitative and qualitative studies offer distinct benefits that can help researchers approach the different components of substance use. For example, quantitative studies can use GIS to describe and visualize the time and place of substance use, as well as quantitatively assess individual and built factors that associate with substance use. Conversely, qualitative studies can use techniques like interviews to understand factors that are not readily quantifiable, such as motivations for using substances or understanding how substance users perceive concepts like rehabilitation, treatment options, risk of arrest, and more. Combining these approaches has some precedent by prominent geographers, although there is ample room for experimental combination of quantitative and qualitative methods (M.-P. Kwan et al., 2019). Mixed methods research is important for the study of substance use because the fundamental approaches of both quantitative and qualitative studies risk omitting something that is communicated by the other. While a quantitative study could carry out a sentiment analysis on words transcribed from an interview, such an approach would miss the intent of qualitative research to render the experience of participants. Similarly, qualitative studies are generally ill-equipped to provide important estimates for phenomenon like drug overdose or mortality. I want to use the strong tradition of both disciplines in geography - as well as the potential for these methods to emerge through GIS - to describe the complex reality of substance use.

I want to teach spatial analysis and geography to the next generation of researchers.

During my time as a residence advisor at McGill and the London School of Hygiene and Tropical Medicine (LSHTM) I gained a deep appreciation for enhancing the undergraduate student experience. Through teaching assistantships, I have been able to continue developing this appreciation and contribute directly to the educational experience of undergraduates at the University of Maryland (UMD). I hope to become an instructor of record for the Department of Geographical Sciences and help excite the next generation of geographers. In a broader sense, I view teaching in higher education as a way to provide students with tools to deal with the unexpected, challenging, and open-ended situations encountered throughout life.

2.2.4 Conceptual representation of scholarly identity



3 Coursework

In this section I provide a summarized narrative of my journey through higher education. I reflect on my academic performance while highlighting key parts of my scholarly training.

3.1 Programs of Study

Like many students, the first year of university was a time of great uncertainty. My course selections were determined largely by how interesting I found the title as well as the time the class started (perhaps the best example being a class called Jesus of Nazareth that began at 3pm). By the end of the first year I had started taking my studies more seriously, and one course in particular would set me on the path leading to UMD. The course was GEOG221, Environment and Health, co-instructed by Dr. Nancy Ross and Dr. Ian Strachan. It was a mixed-discipline course that focused on the health impacts of atmospheric and environmental effects. One unit focused on the ramifications of extreme weather events for those who were isolated with chronic diseases, and this topic reminded me of my own arthritic sensitivity to certain types of weather. Reflecting as a PhD student in Geographical Sciences, that lecture and class sold me on the importance of interdisciplinary approaches in health research.

Table 1. (following page) provides information on courses that were highly influential to my undergraduate studies in Geography. Courses at McGill University are on a semester system and are comparable to other institutions of higher education in North America (i.e. Fall, Spring, and Summer semesters).

In my second year at McGill there were two major developments in my academic curiosity. The first was a newfound curiosity in statistics and GIS. I took back-to-back semesters of geographical statistics (GEOG202 and GEOG351) and GIS (GEOG201 and GEOG306). GEOG306 was a particularly important course in my studies and pivoted me towards computational geography. Focused on raster analysis, I found the course extremely challenging as my general academic abilities were still maturing and I did not have a grasp of the basic concepts that united statistics, GIS, and computer science. The second important development was my progression into the African Studies minor. Through the minor program I was exposed to a suite of literature, knowledge, and discussions that radically changed my perspective on history and the systems that support society today. The coursework of African Studies provided me with an understanding as to how the recent history of colonialism and slavery are tied to ongoing issues of racism and inequality. Growing up in a rural, homogenous, and white community played a part in my severe gaps of knowledge, and pursuing African Studies was one step towards remedying my ignorance. This ignorance cannot be completely erased, and it should not be. This ignorance is a part of my history and I try to reckon with it every day through personal betterment and recognizing injustice when it occurs.

Table 1: Influential courses undertaken while pursuing Bachelor of Arts (Major: Geography) at McGill University. Asterisk denotes graduate level courses.

Course Code	Course Title	Grade & Credits
GEOG202	Statistics & Spatial Analysis	A, 3cr
GEOG221	Environment and Health	B+, 3cr
GEOG272	Geomorphology: Earth's Changing Surface	A-, 3cr
GEOG303	Health Geography	A, 3cr
GEOG351	Quantitative Methods	A-, 3cr
GEOG381	Geographic Thought & Practice	A, 3cr
COMP364	Computational tools for the Life Sciences	A, 3cr
SOCI550*	Developing Societies	A, 3cr
GEOG512*	Adv. Quant. Methods in Social Field Research	A, 3cr
PPHS525*	Healthcare Systems in a Comparative Perspective	A, 3cr
AFRI598*	Research Seminar in African Studies	A-, 3cr
POLI618*	Quantitative Analysis	A, 3cr

Through these personal transformations I wanted to contribute to projects that were analyzing health inequalities. Wanting to have the technical expertise to be of use in a research lab, I sought out upper-level coursework that could advance my skills in modeling, study design, and general knowledge of conducting health research. I also began the honors program with Dr. Lea Berrang-Ford (described in more detail in Section 4). During this time coursework went smoothly and I was able to expand my expertise by completing activities like the summer workshop entitled *Concept and application of multilevel statistical methods to health, social science and geographical research* taught by Professor Subu Subramanian of Harvard. While I appreciated exploring multiple disciplines with a Geographical perspective, I often felt like I lacked certain fundamentals of health research when reading papers in epidemiology and other journals. Thus, as my time at McGill ended, I targeted my MSc applications at Epidemiology programs.

Table 2. (following page) describes the modules completed at LSHTM. Unlike universities in North America, courses at LSHTM are referred to as 'modules' which are usually 4 to 6 weeks in duration. Although shorter in duration, modules contain a similar amount of educational content and a student takes between 8 and 10 modules in a given academic year. Courses are assessed on a different grading scheme where: 0 = Very poor (fail), 1 = Poor (fail), 2 = Satisfactory Pass (only allowed once, not allowed on exams), 3 = Good Pass, 4 = Very Good Pass, 5 = Excellent Pass (reserved for top 5-10% of students depending on the module).

At LSHTM my grades and research productivity were not as high as I would have liked (overall module average of 3.68). I moved to the UK with my partner and found myself balancing cultural adjustment, new living spaces, and financial pressures. The focus of the year was as much ensuring decent living quarters as it was learning about epidemiology. I was also managing

ongoing research responsibilities that had carried over from McGill. Nevertheless, I was able to learn the leading methods and trends on how to analyze epidemiological data, as well as achieve top marks in courses that focused on environmental and spatial approaches in health science. I completed the degree and was able to convert the capstone project into a manuscript.

Table 2: Influential courses undertaken while pursuing Master of Science (Epidemiology) at the London School of Hygiene and Tropical Medicine (LSHTM).

Module Title	Module Mark
Study Design: Writing a Study Proposal	3.00
Statistical Methods in Epidemiology	4.00
Spatial Epidemiology in Public Health	5.00
Environmental Epidemiology	5.00
Adv. Statistical Methods in Epidemiology (focus: Clinical Trials)	3.00

Table 3. describes completed and planned courses for the first year of study in the PhD program at UMD. Since enrolling at the University of Maryland I have made significant progress towards completing the course requirements of the PhD degree (Table 3). One area identified for development by both the PAC committee and the program director, Dr. Laixiang Sun, was ensuring competency in the earth sciences. In addition to courses taken at McGill (GEOG221, 272) and LSHTM (Environmental Epidemiology), I have enrolled in AOSC647: Machine Learning in Earth Science. This course combines fundamentals of earth and atmospheric science with a survey of machine learning methods. This course allows me to both accomplish the PhD course requirement as well as advance my interests in statistical modeling.

Table 3: Courses completed and proposed in pursuit of Doctor of Philosophy (Geographical Sciences) at the University of Maryland.

Course Code	Course Title	Grade & Credits
	Fall 2019 - Completed	
GEOG601	The Nature and Practice of Science	A, 3cr
GEOG606	Quantitative Spatial Analysis	A+, 3cr
GEOG788P	Models and Methods for Spatial Data Science	A+, 3cr
GEOG798	Seminar Series	A+, 1cr
	Spring 2020 - Ongoing	
SURV799B	Topics in Survey Methodology; Small Area Estimation	TBD, 3cr
AOSC647	Machine Learning in Earth Science	TBD, 3cr
TLTC798	University Teaching and Learning	TBD, 2cr
GEOG608	Comprehensive Portfolio Assessment Summary	TBD, 1cr
GEOG798	Seminar Series	TBD, 1cr

4 Research Experience

I have been actively employed as a research assistant in varying capacity since September of 2015. Beyond the development of a wide variety of research skills, I have spent considerable time reflecting on how my personal experiences have affected my trajectory of research interests. In the following section I chronologically describe the major research activities I have carried out since 2015 across my undergraduate and graduate studies.

4.1 Reflection on Previous Research Activities

My introduction to research was in the area of Indigenous health inequalities. My principal task was to analyze nutritional health characteristics sampled from Indigenous Batwa and non-Indigenous Bakiga of Southwestern Uganda. The project supervisor was Dr. Lea Berrang-Ford and the overarching Indigenous Health Adaptation to Climate Change (IHACC) multidisciplinary, multi-center research organization. As it was my first experience in research, I rapidly developed skills in inputting, cleaning, and validating survey data, as well as carrying out statistical analysis in STATA. The primary objective was to classify the malnutrition status of individuals in the sample and examine if there were malnutrition differences between the Indigenous and non-Indigenous populations. Considerable time was spent understanding how to measure malnutrition, especially because the Batwa are a short-statured people. While I had taken undergraduate-level statistics courses, the focus of the models in this research project were learned from the previously-mentioned summer course.

Wanting to stay in the area of Indigenous health inequalities, I successfully applied to the internship program at the Institute for Health and Social Policy (IHSP) of McGill University under Dr. Mylene Riva. At the time, Dr. Riva's work focused on the health status of Indigenous populations in the high north of Canada. Because my work with Dr. Berrang-Ford had been primarily quantitative I wanted to gain more skills in research translation. With Dr. Riva I sketched and designed posters that translated research findings into visual, easy-to-understand figures that could be disseminated to community partners. This process taught me how such research translation is ultimately iterative as drafts must be produced, reviewed, and improved before information can be disseminated.

Throughout these research experiences I was engaging in a dialogue with myself about researcher positionality and societal impact. I felt that there were complicated, unanswered questions about the power dynamics between research institutions and Indigenous participants and partners. I engaged with critical literatures discussing these power dynamics and spoke at length with my advisors and research partners. While I offer no widespread insight on the issue, by the end of my time with IHACC and IHSP I had realized for myself that I would seek opportunities in other areas of health inequalities research.

At LSHTM I learned the leading-edge methods in the spatial analysis of health outcomes. This included collecting and organizing spatial health data, applying spatial statistics like point pattern analysis and spatial regression, and, most importantly, translating spatial outputs into rational elements of disease control programs. I worked as a research assistant carrying out exploratory spatial analysis on a subset of data from the DREAMS project - a large, multinational project between the United States PEPFAR program and several countries of sub-Saharan Africa to reduce new HIV infections among young women. The spatial analysis was principally concerned with identify potential hotspots within a known area of high HIV burden. I carried out a series of point pattern analysis and kriging techniques to assess the degree of spatial dependency in the data. All analyses were coordinated with Dr. Daniel Kwaro, the LSHTM-Kenyan Ministry of Health (KMOH) partner. Intermittent updates and reports were provided to Dr. Kwaro, and both an MSc capstone project and journal-ready manuscript was produced.

Concurrent to these experiences in Indigenous health inequalities and throughout my MSc I worked for Dr. Leonardo Baccini in the Department of Political Science in at McGill University. My time under Dr. Baccini proved to be immensely beneficial to the development of my data processing skills. Working with STATA and R, I regularly collected, scraped, and processed government datasets relating to economic, political, and social indicators in the United States.

I worked with Dr. Baccini from September of 2016 to May of 2019. During that time, I was the sole code maintainer for two large projects relating to United States' legislative voting patterns, trade flows, and agricultural productivity. In fact, there was a time during my MSc where I considered abandoning health sciences in favoring of switching to political science. While I ultimately decided against this change, for the majority of 2019 Dr. Baccini and I acted as collaborators on a project examining comparative advantage and legislative voting patterns. A large portion of the internal project proposal is included as an example of critical thinking. While I greatly appreciated my time with Dr. Baccini, the projects were increasingly away from topics I had intrinsic interest in (namely health sciences). Thus, I chose to exit the project and focus on my upcoming move to Maryland to begin my PhD in the Department of Geographical Sciences. The technical skills I gained with Dr. Baccini help me to this day and through our work I gained an above average familiarity with the multitude of United States statistical products.

4.2 Additional Research Competencies

I have the research maturity to acknowledge that I do not yet possess all the skills necessary to complete the dissertation. While I am proficient in data preparation, technical writing, health geography, statistics, and geographic information science, there are several specific domains that I intend to develop throughout the course of the dissertation.

Domain knowledge on substance use research

My previous training in health science research has been on the topics of malnutrition and HIV. Since entering the doctoral program under Dr. Stewart, I have begun to familiarize myself with the literature – especially geographic literature – on substance use. I have set up journal alerts for the leading academic publications in substance use and I have read a number of books on the opioid epidemic in the United States. As I move forward in the dissertation, I expect to define and execute a systematic literature review on the spatial analysis of substance use. There are also relevant professional meetings – such as The Prescription Rx and Heroin Summit and Addiction Conference – that would enhance my knowledge on current topics in substance use research.

Domain knowledge on advanced spatial modeling techniques

Although I have taken several courses on GIS and statistics, I have had few courses that were tailored specifically to spatial modeling. Since entering the doctoral program, I have completed both GEOG606 and GEOG788P which placed significant emphasis on spatial modeling techniques like spatially lagged regression and geographically weighted regression. I am developing my skills in this area through independent study and contributing to the Python Spatial Analysis Library (PySAL). I have also begun reviewing leading textbooks in the statistical analysis of spatial data (Banerjee, n.d.). Developing this competency is essential as I plan to apply spatial regression methods throughout my dissertation.

Domain knowledge on machine learning

The burgeoning field of machine learning is an area of applied statistics and big data analysis. While I have completed introductory exercises on the subject, I lack formal training that can enable me to apply machine learning methods on my own datasets. AOSC647 is my first step towards gaining this training and the class has already provided an immense amount of information on supervised learning techniques including k-means, lasso, ridge, and elastic net regression. After completing the course, I intend to focus on applications of machine learning that integrate geographic data.

4.3 Planned Future Research

In addition to the dissertation research (see Section 7), I plan to carry out the following research activities.

Research assistantship under Dr. Kathleen Stewart for the Center for Substance Abuse Research (CESAR).

Under a BSOS Dean's Research Initiative (DRI) collaboration between Dr. Kathleen Stewart and Dr. Eric Wish, I am undertaking a geospatial analysis of patients presenting to University of Maryland Medical System (UMMS) hospitals in downtown Baltimore for drug related health problems. Details of this research are explained in Section 5 and Section 7.

Expand arcos/arcospy with additional data scraped from Drug Enforcement Agency (DEA) Automated of Reports and Consolidated Orders System (ARCOS) annual reports.

ARCOS represents a nationwide, longitudinal dataset of information pertaining to controlled substances. While *The Washington Post* has made available a sizeable portion of the ARCOS dataset, their data was at the county-level and focused on the total number of prescriptions for oxycodone and hydrocodone. A significantly larger dataset was recently made available via Retail Drug Summary Reports. However, this data has yet to see much scientific use due to the unwieldy format of the reports (most are more than 500 pages in .pdf format). I have begun writing scripts to process the reports and convert them into datasets for statistical and spatial analysis. I intend to use this data in my own dissertation work (see Section 7) as well as make the data easily accessible to other researchers.

4.4 Significant Research Outputs

Rich, S., Ba Tran, A., Williams, A., Holt, J., Sauer, J., Oshan, T. (2020). arcos and arcospy: R and Python packages for accessing the DEA ARCOS database from 2006 2014. Journal of Open Source Software (JOSS). Ready for submission. *Sole translator of Python module. Primary author of JOSS submission and information pages on Github*.

Sauer, J. (2019). An exploratory spatial data analysis of the 2006-2012 ARCOS dataset: the geographic distribution of prescription opioid doses and their association with arrests for the possession of synthetic narcotics. *University of Maryland, term paper for GEOG788P: Models and Methods for Spatial Data Science.*

Sauer, J., Baccini, L. (2019). Comparative Advantage Internal Build Guide. *Primary author of reproducibility guide for an extended research project with Dr. Leonardo Baccini and Dr. Francesco Amodio.*

Sauer, J., Berrang-Ford, L., Patterson, K., Donnelly, B., Lwasa, S., Namanya, D., . . . Harper, S. (2018). An analysis of the nutrition status of neighboring indigenous and non-indigenous populations in Kanungu district, southwestern Uganda: Close proximity, distant health realities. *Social Science & Medicine*, 217, 55-64. doi:10.1016/j.socscimed.2018.09.027

Sauer, J. (2018) An exploratory analysis of HIV Prevalence and its Spatial Characteristics in Gem District, Western Kenya. *London School of Hygiene and Tropical Medicine, Master's Thesis.*

Sauer, J., Riva, M. (2017) Visualizing research for dissemination. *McGill University Institute for Health and Social Policy (IHSP) Internship Cumulative Presentation.*

Sauer, J. (2016) Malnutrition among Indigenous Batwa in Southwestern Uganda. *McGill University, Bachelor's Thesis.*

5 Professional Experiences

Developing arcospy with data journalists at The Washington Post

While in communication with Dr. Stewart about my plans at UMD I began to follow media coverage of the opioid epidemic more closely. In this day-to-day coverage I happened across a breaking story by *The Washington Post* about a large amount of DEA data they were making publicly available as part of ongoing journalism and litigation against Purdue Pharmaceuticals. After reading the story and exploring the data I realized its immense research potential. In the late summer of 2019, I began to familiarize myself with the data, and in the Fall 2019 semester I incorporated the data into two substantial class projects. A component of one class project - specifically GEOG788P - was the opportunity to carry out a mini software development exercise. At the time I did not have any experience in this process, but I wanted to push myself and so I decided to translate *The Washington Post*'s data API from R to Python. This project would eventually expand into a partnership with *The Washington Post* to fuse the software together, jointly author paper ready for submission to *The Journal of Open Source Software (JOSS)*, and release python module called arcospy that has since had several hundred downloads. This process provided invaluable experience producing a simple software package and building a relationship with external partners.

Research assistantship (as PhD student) with Dr. Kathleen Stewart, the University of Maryland

As a research assistant I have spent the Fall of 2019 and Spring of 2020 working as a research assistant on the collaborative partnership between Dr. Stewart and the Center for Substance Abuse Research (CESAR). I have carried out substantial mapping and exploratory spatial data analysis of an ongoing pilot study that examines urinalysis results from patients reporting to the University of Maryland Medical System in downtown Baltimore. Tens of maps have been produced, refined, and presented to both Dr. Stewart and partners at CESAR to help provide insight into relevant geospatial analyses that might be applied to the study. To promote both

reproducibility and rapid adaptability, all analyses and maps were written in R and are shared with CESAR.

Research assistant (as professional) at Geographic Information Centre, McGill University

After completing my master's degree, I began as a fulltime research assistant to the Director of the Geographic Information Centre (GIC), Dr. Tim Elrick. At the GIC I was responsible for assisting Dr. Elrick with daily information technology (IT) operations and ensuring that university researchers had access to high-powered workstations. I was also tasked with a sizeable amount of workshop organization and teaching from topics ranging from introductory GIS to developing Git/Github workflows. I also acted as an at-large research assistant, contributing GIS, statistical analysis, data management scripts, or general workflow consulting to researchers across several career levels.

Sauer et al. (2018). An analysis of the nutrition status of neighboring indigenous and non-indigenous populations in Kanungu district, southwestern Uganda: Close proximity, distant health realities. Social Science & Medicine, doi:10.1016/j.socscimed.2018.09.027

As previously elaborated, I spent several years working for Dr. Lea Berrang-Ford as one of the last McGill-based undergraduate researchers on the IHACC project. I carried out the analysis over an extended period, first as an undergraduate thesis and then later as s publication to *Social Science & Medicine*. This was my first substantial academic endeavor and it is difficult to overstate my growth over the project. Most importantly, this publication was a practice in project completion. Due to Dr. Berrang-Ford's transition to the UK, I had sole responsibility of submitting the article, responding to reviewers, and making final corrections that would appear online. I also had to deliver the article and produce summarized versions of the research to partners at the Ugandan Ministry of Health as part of IHACC project's integrated knowledge translation.

Research assistant (as student and professional) with Dr. Leonardo Baccini

I worked for Dr. Leonardo Baccini in the Department of Political Science in at McGill University for an extended period, occupying a significant role in the data gathering, preparation, and analysis process for several of his projects. I would routinely gather statistical products produced by the United States government (for example, the Quarterly Workforce Indicators). I was responsible for maintaining a large, shared drive between Dr. Baccini and external collaborators, as well as providing frequent updates on the progression of the data collection and processing. Although I wanted to take on more even more responsibility in the projects, my commitments to other research endeavors and training outside the discipline prohibited this. Nevertheless, I gained significant exposure to methods used in quantitative political science and policy analysis.

As my move to UMD approached I created exit plans with Dr. Baccini such that my exit from the projects would not cause too much disturbance. This required extensive reproducibility guides, training of new research assistants, and extended discussions with collaborators.

6 Evidence of Analytical and Integrative Thinking

Here I expand on five examples from my list of significant research outputs that demonstrate analytical and integrative thinking. The selection highlights both my ability to use different analytical frameworks as well as integrate my Geographical expertise into the research process.

Example 1. Rich, S., Ba Tran, A., Williams, A., Holt, J., Sauer, J., Oshan, T. (2020). arcos and arcospy: R and Python packages for accessing the DEA ARCOS database from 2006 2014. *Journal of Open Source Software (JOSS)*. Ready to submit.

As previously explained, arcospy is a result of a class project started in GEOG788P. However, it has since expanded to become an ongoing example of analytical and integrative thinking shaping my research career. Carrying out this project required rapidly learning the key elements of a software development framework, including computational efficiency, testing, documentation, and deployment. I championed this project because of the immense research potential residing within the data and the intuition that expanding access to the data would allow a wider user base to explore the data. During my literature review on the opioid crisis I have come to appreciate how a broader access to data sources like ARCOS might have been able to reveal the role of prescription opioids in the crisis. Since building a relationship with partners at *The Washington Post*, I have offered novel ideas to expand the dataset, primarily through the parsing of massive pdf reports made available online. This has spurred a new round of investigative interest and I intend to continue working on the data such that it can be made easily accessible for other interested citizens and researchers. I hope that these twin software packages can become a go-to resource for analyzing the contextualizing the US opioid crisis as one chapter in the extended history on the medical use of opioids.

Example 2. Sauer, J. (2019). An exploratory spatial data analysis of the 2006-2012 ARCOS dataset: the geographic distribution of prescription opioid doses and their association with arrests for the possession of synthetic narcotics. *University of Maryland, term paper for GEOG788P: Models and Methods for Spatial Data Science.*

This paper was my first analysis of the ARCOS data and the complementary component to the arcospy software development project. GEOG788P placed immense emphasis on the implementation and examination of spatial analysis, especially exploratory spatial data analysis and spatial regression models. I set about employing both techniques on the relationship between ARCOS data and arrests for the sale of synthetic opioids. I applied global and local

autocorrelation statistics as well as spatial lag and spatial error multivariate regression models to understand the role of space in the relationship between exposure and outcome. I integrated feedback from both Dr. Stewart and Dr. Taylor Oshan as to how the models might be improved - primarily through the addition of state-level fixed effects - and sought out further information on issues with the arrest data. While I have since shifted my focus towards other health measures as the outcome of interest, the experience from this paper advanced my understanding of spatial analysis. The place-based component between being exposed to prescription pain medicine and seeking drug-related health services provides a strong theoretical backing for spatial models in my future analyses.

Example 3. Sauer, J., Baccini, L. (2019). Comparative Advantage Internal Build Guide. *Replication materials for ongoing research projects led by Dr. Leonardo Baccini and Francesco Amodio.*

While exiting my research responsibilities with Dr. Baccini I was responsible for reviewing a large analysis and data repository shared between multiple collaborators. Analytically, this project required the proposition of several novel measures that could both capture agricultural advantage as well as voting measures. The goal of the project was to use these measures to assess the degree to which US legislators did or did not vote in line with their empirical agricultural productivity in international trade agreements. This required the integration of global satellite products and US legislative voting records. I had the responsibility of managing unifying the nonspatial and spatial data, contributing to the novel measures, carrying out preliminary descriptive analyses, and ensuring reproducible workflows. While the 'reproducibility crisis' has thoroughly changed the way, quantitative analysis is carried out in fields like political science, economics, and epidemiology, the discussion has appeared in only select circles of geography (Nüst et al., 2018). As I advance in my research career I intend to advocate for greater reproducibility in the geographical sciences, especially in the area of GIScience.

Example 4. Sauer, J., Berrang-Ford, L., Patterson, K., Donnelly, B., Lwasa, S., Namanya, D., . . . Harper, S. (2018). An analysis of the nutrition status of neighboring indigenous and non-indigenous populations in Kanungu district, southwestern Uganda: Close proximity, distant health realities. *Social Science & Medicine*, 217, 55-64. doi:10.1016/j.socscimed.2018.09.027

As discussed elsewhere in the document, this paper was a critical step in my development as a researcher. The greatest analytical challenge in this paper was identifying a way to meaningful quantify the impact of indigeneity in two closely related communities. While several traditional statistical operations were available - such as difference in means tests and logistic regression - we opted for multilevel models. While these models are more complex, they also allowed us to explicitly capture variation across individual, household, and community-ethnicity levels. Resoundingly, models that incorporated community-ethnicity explained the greatest amount of

variance in the measure of malnutrition. A more conceptual challenge was the navigation of the Discussion section of the paper. The topics covered in the Discussion section include short-statured population evolutionary theory, inequality brought on by 18th century colonization, and recent identity-based discrimination. Integrating the core components of these topics and balancing them alongside others was challenging, even more so when considering how these topics intersect with the differing notions of Indigeneity across Africa. After discussions with both Dr. Berrang-Ford and partners from the Ugandan Ministry of Health, we ultimately realized that it was our responsibility present the complex intersection of these topics to the best of our ability. While our study may not provide the skeleton key to unlocking a coherent grand narrative that seamlessly links the numerous topics, it provided me with a perspective that could appreciate the scale and gravity required by interdisciplinary research.

Example 5. Sauer, J. (2018) An exploratory analysis of HIV Prevalence and its Spatial Characteristics in Gem District, Western Kenya. *London School of Hygiene and Tropical Medicine, Master's Thesis.*

This project was the culmination of my master's degree in Epidemiology. Although areas with a high burden of disease can be identified using traditional epidemiological approaches, the Modifiable Area Unit Problem (MAUP) complicates this profess and has yet to be thoroughly understood. In this project I examined a study area that was known to have a higher burden of HIV compared to the surrounding area and nation at large. However, there was a notion that researchers might be able to pinpoint specific places within the study area that had an especially high burden. I was tasked with investigating this claim and choosing appropriate methods to analyze the data across different spatial levels (individual, village, or community). While we eventually concluded that there lacked evidence that any such places existed within the study area - and that all indicated a higher burden of disease - it was an excellent exercise in integrating both geographic and epidemiologic thinking into a single study.

7 Dissertation Plans

7.1 Tentative title

Geospatial analysis of overdose and treatment admission in the United States opioid epidemic using administrative data sources

7.2 Real-world problem to be addressed

Since the late 1990s there has been increasing attention to rises in opioid-related morbidity, mortality, and criminality in the United States. The National Center for Health Statistics (NCHS), a program run by the Centers for Disease Control and Prevention (CDC), estimates that in 1999

there were approximately 8,000 deaths involving opioids whereas 2017 saw over 47,000 deaths (Hedegaard et al., 2018). Opioids are a chief concern of the Drug Enforcement Agency (DEA), who consider both prescription and non-prescription substances to be 'highly' available in cities throughout the United States (DEA, 2018).

The combination of public outcry and alarming statistics led the US government to declare the Opioid Crisis a public health emergency in October of 2017 (U.S. Department of Health & Human Services, 2017). This declaration led to the Substance Use-Disorder Prevention that Promotes Opioid Recovery and Treatment (SUPPORT) for Patients and Communities Act, which promises $6 billion USD in funding towards addiction services (Walden, 2018).

While the 'Opioid Crisis' is often understood as a national issue, its local manifestations vary immensely. Opioids are not a novel substance within the United States - cities like Baltimore, Maryland, have dealt with forms of opioids for close to 100 years, with multigenerational families dependent on the substance in its various forms (Jonnes, 1999; Mars et al., 2018). The goal of my dissertation is to explore this variation through the spatial analyses of datasets that capture different realities of the ongoing crisis.

7.3 Research Literature and Methodological Approaches

Research Literature

In the proposal preparation process, I am reviewing three primary bodies of literature: (1) the medical history of opioids; (2) opioids in the United States, and (3) spatial analysis of health outcomes. I am also actively monitoring criminology-related journals for information on opioids, although I do not intend to rely on these publications due to the different set of objectives.

Opioid is a catch-all term used to refer to substances that activate and bind to opiate receptors (Rosenblum et al., 2008). Like many other substances with abuse potential, opium has a distinct historical legacy in several countries around the world. Booth, Meldrum, and Brill et al. have authored numerous articles and books that describe the various histories of opium across the globe (Booth, 1999; Brill et al., 2005; Meldrum, 2003).

Opioids are usually acquired in either legal (e.g. prescription pain medication) or illegal (e.g. heroin) form, although this distinction is heavily influenced by legal and societal contexts. Someone in extreme pain may have a need for pharmaceutically produced opium derivatives, such as oxycodone. However, if a dependence emerges, the perceived use of the prescription pain medicine might transform to be conceived as illicit use. Making things more complicated, an addicted user may in fact use opioids to avoid withdrawal symptoms, complicating traditional understandings as to why an individual uses a substance.

Opioids in the US have a complicated medical and social history. Medical professionals have long tried to balance potent analgesic effects with issues of dependence, while agencies tasked with law enforcement have focused on opioid diversion (Drug Enforcement Administration, 2002). Presently there are a numerous opioid products in both pure or combination forms available via prescription in the United States, as well as substantial recovery programs that distribute methadone to individuals with addiction (Rosenblum et al., 2008).

A common historical reference point for drug legislation in the United States is the 1914 Harrison Act. Originally a tax act, a judicial interpretation expanded the use of the act to control the distribution of substances that carried a risk of addiction (Meier, 2003). I intend to contextualize my research within contemporary legislation, identifying what legislation since the Harrison Act may have been important in the opioid epidemic.

Spatial analysis has increasingly been used by geographers, epidemiologists, and criminologists in an attempt to capture dynamic relationships between built, natural, and social systems, many of which have an explicit spatial component (Elliott & Wartenberg, 2004; Symanzik, 2014; Ye & Wu, 2011). Particularly relevant to me are analyses of administrative datasets wherein the exposure is a prescription opioid or illicit opioids. Exemplary examples include Brownstein et al., Fulton et al, and Stopka et al. (Brownstein et al., 2010; Fulton et al., 2019; Stopka et al., 2019).

While systematic comparison between these studies can be difficult due to differences in study design, available data or samples, and specific geospatial techniques employed, spatial analysis and geographic information science (GIS) remains a useful way that researchers can approach questions of "demand, supply, harms, and harm reduction" at geographic units that are logistically and politically meaningful in the US landscape (Mazumdar, Mcrae, & Islam, 2015).

Methodological Approaches

My methodological approaches draw from GIS and Geographic Data Science (GDS) workflows (Singleton & Arribas-Bel, 2019). After gathering and cleaning the spatial data I intend to carry out exploratory spatial data analysis (ESDA) to determine the extent of spatial autocorrelation present (L Anselin, 1999). I will produce descriptive maps for the attribute values as well as spatial autocorrelation measures. Spatial autocorrelation statistics that are relevant for my research includes local and global indicators of spatial autocorrelation, Lee's statistic, and local join count statistics (Luc Anselin & Li, 2019; Lee, 2001).

As the above exploratory steps are reaching the conclusion, I will determine an appropriate spatial modeling technique to apply to the data. Example models include spatial lag models, spatial error models, geographically weighted regression, and conditional autoregressive models.

Given the longitudinal nature of the datasets I intend to work with there may be opportunity for fitting spatio-temporal models (Ransome et al., 2020). The covariates included in the model are be informed by background literature as well as automated feature selection (e.g. ridge or random forest methods).

7.4 Research Questions

I am planning for a three-paper structure for the dissertation. Each of the three research questions is related to the geospatial modeling of exposure and outcome data relevant to the opioid crisis.

Paper 1: Developing risk-surface models for drug overdose using a subset of electronic health records from emergency departments in Baltimore, Maryland.
Research objectives(s) to be addressed: What are the residency patterns for patients who present with a drug-related health problem in Baltimore, Maryland? Can a spatio-temporal risk model account for these residency patterns, or are there additional social and built factors that explain variation in risk between areas that present patients with a drug-related health problem?
Brief description of paper: For the first paper I am analyzing residency patterns for an exhaustive cross-section of adult patients presenting to University of Maryland Medical System (UMMS) emergency departments with a drug or overdose-related health problem from 2016 onward. I have carried out extensive ESDA to describe patient residency patterns and I am deploying hierarchical Bayesian Poisson regression models to develop spatio-temporal risk surfaces. These risk surfaces identify specific areas that have above- or below-average risk of patients presenting to emergency departments with a drug-related health problem. Moreover, the models can control for known confounding factors that mitigate overdose risk, such as neighborhood socioeconomic characteristics. This approach has the added benefit of being able to consider joint outcomes, for example, both heroin and methamphetamine risk in a given area. The dataset includes more than 8000 electronic health records including physician diagnoses and corresponding ICD-10 codes, patient characteristics, and residency information at the zip code level.

Paper 2: Examining the geographic relationship between prescription opioids and admission to treatment facilities.
Research question(s) to be addressed: How does the geographic distribution of prescription opioids to an areal unit and its neighbors impact admission to treatment facilities for opioid use disorder?
Brief description of paper: For the second paper I intend to quantify the distribution of prescription opioids and explore their geographic relationship with a tangible public health outcome, specifically the admission to publicly funded treatment facilities for opioid use disorder. Data will be combined from the DEA ARCOS database and the Substance Abuse & Mental Health Services Administration (SAMHSA) Treatment Episode Data Set (TEDS-A). I

intend to analyze concurrent relationship between the amount of prescription opioids in a given area and the number of admissions to publicly funded treatment facilities for opioid use disorder through spatial and non-spatial modeling techniques. Given that opioid use disorder can develop over time, I am hypothesizing ways to fit lagged models at different time intervals.

Paper 3: In development.
Brief description: I am currently developing several potential research questions for **Paper 3.** One potential plan is to examine the impact of COVID-19 on patients seeking medical services for drug and overdose-related health problems using the previously mentioned EPIC data. For example, it may be interesting to examine the relative 'reach' of the UMMS system before, during, and after COVID-19. The reach of the UMMS systems might be conceptualized via a convex hull using the centroid of the patient zip codes. Alternatively, I could use the zip code level data from scraped ARCOS annual reports in relation to observed patient counts from the EPIC data.

8 References

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. *Geographic Information Systems: Principles, Techniques, Management and Applications*, 253–266.

Anselin, Luc, & Li, X. (2019). Operational local join count statistics for cluster detection. *Journal of Geographical Systems*, *21*, 189–210. https://doi.org/10.1007/s10109-019-00299-x

Banerjee, S. (n.d.). *Hierarchical Modeling and Analysis for Spatial Data*.

Booth, M. (1999). *Opium: a history* (1st ed.). Griffin.

Brill, S., Gurman, G. M., & Fisher, A. (2005). A history of neuraxial administration of local analgesics and opioids. *European Journal of Anaesthesiology*, *20*(9), 682–689. https://doi.org/10.1017/S026502150300111X

Brownstein, J. S., Green, T. C., Cassidy, T. A., & Butler, S. F. (2010). Geographic information systems and pharmacoepidemiology: using spatial cluster detection to monitor local patterns of prescription opioid abuse. *Pharmacoepidemiology and Drug Safety*, *19*(6), 627–637. https://doi.org/10.1002/pds.1939

Claerbout, J., & Karrenbach, M. (1992). Electronic documents give reproducible research a new meaning. *Proceedings of the 62nd Annual International Meeting of the Society of Exploration Geophysics*, 601–604.

Curry, M. (1996). On space and spatial practice in contemporary geography. In E. Carville, M. Kent, & K. S. Martin (Eds.), *Concepts in human geography* (pp. 3–32). Rowman and Littlefield.

Cutter, S. L., Golledge, R., & Graf, W. L. (2002). The Big Questions in Geography. *The Professional Geographer*, *54*(3), 305–317. https://doi.org/10.1111/0033-0124.00332

DEA. (2018). *DEA National Drug Threat Assessment 2018 Report*. https://www.dea.gov/sites/default/files/2018-11/DIR-032-18 2018 NDTA final low resolution.pdf

Drug Enforcement Administration. (2002). A joint statement from 21 health organizations and the Drug Enforcement Administration. Promoting pain relief and preventing abuse of pain medications: a critical balancing act. *Journal of Pain and Symptom Management*, *24*(2), 147. http://www.ncbi.nlm.nih.gov/pubmed/12269265

Elliott, P., & Wartenberg, D. (2004). Spatial Epidemiology: Current Approaches and Future Challenges. *Environmental Health Perspectives*, *112*(9), 998–1006. https://doi.org/10.1289/ehp.6735

Fulton, L., Dong, Z., Zhan, F. B., Kruse, C. S., & Stigler Granados, P. (2019). Geospatial-Temporal and Demand Models for Opioid Admissions, Implications for Policy. *Journal of Clinical Medicine*, *8*(7), 993. https://doi.org/10.3390/jcm8070993

Hedegaard, H., Minino, A., & Warner, M. (2018). *Drug Overdose Deaths in the United States, 1999-2017*. National Center for Health Statistics - Data Brief. https://www.cdc.gov/nchs/data/databriefs/db329-h.pdf

Johnson, T. P. (2014). Sources of Error in Substance Use Prevalence Surveys. *International Scholarly Research Notices*, *2014*, 1–21. https://doi.org/10.1155/2014/923290

Jonnes, J. (1999). Everybody Must Get Stoned. In *Hep-cats, narcs, and pipe dreams : a history of America's romance with illegal drugs*. Johns Hopkins University Press.

Koch, T. (2017). *Ethics in Everyday Places: Mapping Moral Stress, Distress, and Injury*. ESRI Press.

Kwan, M.-P., Wang, J., Tyburski, M., Epstein, D. H., Kowalczyk, W. J., & Preston, K. L. (2019). Uncertainties in the geographic context of health behaviors: a study of substance users' exposure to psychosocial stress using GPS data. *International Journal of Geographical Information Science*, *33*(6), 1176–1195. https://doi.org/10.1080/13658816.2018.1503276

Kwan, M. P. (2013). Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility: Space-Time Integration in Geography and GIScience. *Annals of the Association of American Geographers*, *103*(5), 1078–1086. https://doi.org/10.1080/00045608.2013.792177

Lee, S. Il. (2001). Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *Journal of Geographical Systems*, *3*(4), 369–385. https://doi.org/10.1007/s101090100064

Mars, S. G., Ondocsin, J., & Ciccarone, D. (2018). Sold as Heroin: Perceptions and Use of an Evolving Drug in Baltimore, MD. *Journal of Psychoactive Drugs*, *50*(2), 167–176. https://doi.org/10.1080/02791072.2017.1394508

Mazumdar, S., Mcrae, I., & Islam, M. (2015). How Can Geographical Information Systems and Spatial Analysis Inform a Response to Prescription Opioid Misuse? A Discussion in the Context of Existing Literature. *Current Drug Abuse Reviews*, *8*(2), 104–110. https://www.ingentaconnect.com/content/ben/cdar/2015/00000008/00000002/art00004

Meier, B. (2003). *Pain Killer: An Empire of Deceit and the Origin of America's Opioid Epidemic*. Penguin Random House. https://www.penguinrandomhouse.com/books/575714/pain-killer-by-barry-meier/

Meldrum, M. L. (2003). A Capsule History of Pain Management. *JAMA*, *290*(18), 2470. https://doi.org/10.1001/jama.290.18.2470

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). Reproducible research and GIScience: An evaluation using AGILE

conference papers. *PeerJ*, *2018*(7), e5072. https://doi.org/10.7717/peerj.5072
Ransome, Y., Subramanian, S. V., Duncan, D. T., Vlahov, D., & Warren, J. (2020). Multivariate spatiotemporal modeling of drug- and alcohol-poisoning deaths in New York City, 2009–2014. *Spatial and Spatio-Temporal Epidemiology*, *32*, 100306. https://doi.org/10.1016/J.SSTE.2019.100306
Rosenberg, M. W. (1998). Medical or health geography? Populations, peoples and places. *International Journal of Population Geography*, *4*(3), 211–226. https://doi.org/10.1002/(SICI)1099-1220(199809)4:3<211::AID-IJPG83>3.0.CO;2-O
Rosenblum, A., Marsch, L. A., Joseph, H., & Portenoy, R. K. (2008). Opioids and the treatment of chronic pain: controversies, current status, and future directions. *Experimental and Clinical Psychopharmacology*, *16*(5), 405–416. https://doi.org/10.1037/a0013628
Singleton, A., & Arribas-Bel, D. (2019). Geographic Data Science. *Geographical Analysis*, *March*, 1–15. https://doi.org/10.1111/gean.12194
Stopka, T. J., Amaravadi, H., Kaplan, A. R., Hoh, R., Bernson, D., Chui, K. K. H., Land, T., Walley, A. Y., LaRochelle, M. R., & Rose, A. J. (2019). Opioid overdose deaths and potentially inappropriate opioid prescribing practices (PIP): A spatial epidemiological study. *International Journal of Drug Policy*, *68*, 37–45. https://www.sciencedirect.com/science/article/pii/S0955395919300908
Symanzik, J. (2014). Exploratory Spatial Data Analysis. In *Handbook of Regional Science* (pp. 1–1732). https://doi.org/10.1007/978-3-642-23430-9
Tuan, Y.-F. (1991). A View of Geography. *Geographical Review*, *81*(1), 99. https://doi.org/10.2307/215179
U.S. Department of Health & Human Services. (2017). *Determination that a Public Health Emergency Exists Nationwide as the Result of the Opioid Crisis*. Public Health Emergency. https://www.phe.gov/emergency/news/healthactions/phe/Pages/opioids.aspx
Volkow, N. D., & Collins, F. S. (2017). The Role of Science in Addressing the Opioid Crisis. *New England Journal of Medicine*, *377*(4), 391–394. https://doi.org/10.1056/NEJMsr1706626
Walden, G. (2018). *H.R.6 - 115th Congress (2017-2018): SUPPORT for Patients and Communities Act*. https://www.congress.gov/bill/115th-congress/house-bill/6
Ye, X., & Wu, L. (2011). Analyzing the dynamics of homicide patterns in Chicago: ESDA and spatial panel approaches. *Applied Geography*, *31*(2), 800–807. https://doi.org/10.1016/J.APGEOG.2010.08.006

9 Documentation

Accompanying the PAC document is a folder containing copies of the following documents, including all documents discussed in the PAC document as well as additional documents demonstrating integrative and analytical thinking. Documents are ordered chronologically, with the most recent documents appearing first.

Document 1. Rich, S., Ba Tran, A., Williams, A., Holt, J., Sauer, J., Oshan, T. (2020). arcos and arcospy: R and Python packages for accessing the DEA ARCOS database from 2006 2014. Journal of Open Source Software (JOSS). Ready for submission. *Sole translator of Python module. Primary author of JOSS submission and information pages on Github*.

Document 2. Sauer, J. (2019). An exploratory spatial data analysis of the 2006-2012 ARCOS dataset: the geographic distribution of prescription opioid doses and their association with arrests for the possession of synthetic narcotics. *University of Maryland GEOG788P, term paper.*.

Document 3. Sauer, J., Baccini, L. (2019). Comparative Advantage Internal Build Guide. *Primary author of reproducibility guide for an extended research project with Dr. Leonardo Baccini and Dr. Francesco Amodio.*

Document 4. Sauer, J., Berrang-Ford, L., Patterson, K., Donnelly, B., Lwasa, S., Namanya, D., . . . Harper, S. (2018). An analysis of the nutrition status of neighboring indigenous and non-indigenous populations in Kanungu district, southwestern Uganda: Close proximity, distant health realities. *Social Science & Medicine*, 217, 55-64. doi:10.1016/j.socscimed.2018.09.027

Document 5. Sauer, J. (2018). An exploratory analysis of HIV Prevalence and its Spatial Characteristics in Gem District, Western Kenya. *London School of Hygiene and Tropical Medicine, Master's Thesis.*

Document 6. Sauer, J., Riva, M. (2017). Visualizing research for dissemination. *McGill University Institute for Health and Social Policy (IHSP) Internship Cumulative Presentation.*

Document 7. Sauer, J. (2017). Unpacking Deworming Externalities: a reproduction of Miguel and Kremer (2004). *McGill University, POLI618, Term paper*.

Document 8. Sauer, J. (2016) Malnutrition among Indigenous Batwa in Southwestern Uganda. *McGill University, Bachelor's Thesis.*

Document 9. Sauer, J. (2015). Coarsened Exact Matching (CEM) to Analyze Reforestation via Landsat Imagery in Celaque National Park. *McGill University, GEOG512, Term paper*.