Education and Technology Transfer: 2018 Summer Internship Report

Project Number: UTC Safe-D 03-049
Project Title: Data Fusion for Nonmotorized Safety Analysis

Ipek N. Sener, Ph.D.

Sirajum Munira

Atom Arce

Texas A&M Transportation Institute

October 2018

Disclaimer

Acknowledgments

This research was funded by a grant from the Safety through Disruption (Safe-D) National UTC supported by the U.S. Department of Transportation through the UTC Program. The authors would like to acknowledge Michelle Benoit from the Texas A&M Transportation Institute for her editorial review of an earlier version of this report.

Table of Contents

List of Tables

Introduction

Fueled by the inevitable changes in our transportation system, the Safety through Disruption (Safe-D) University Transportation Center (UTC) endeavors to maximize the potential safety benefits of disruptive technologies through targeted research that addresses the most pressing transportation safety questions. The Safe-D UTC focuses its efforts in three key areas:

- cutting-edge research by leading transportation safety experts and their students;
- education and workforce development with programs for all levels, from grade school through college and continuing education for professionals; and
- fully supported technology transfer including practitioner training partnerships, social networking, commercialization, and intellectual property management.

A Safe-D project titled Data Fusion for Nonmotorized Safety Analysis, headed by the principal investigator (PI), Dr. Ipek Sener from the Texas A&M Transportation Institute (TTI), began June 2018. One of the key tasks of this project is its education and technology transfer plan. As part of this task, a 5-week summer internship program was designed. The main goal of the internship program was to provide an undergraduate student with expanded opportunities for guided learning. The key objectives of the internship program were for the intern to gather knowledge about various aspects of the Safe-D program, and gain transferrable real-world research skills by assisting the researchers of the project. In addition, a secondary objective of this internship program was to provide a graduate student with opportunities for improved leadership skills by closely working with an undergraduate student. Both students were expected to develop work competencies and professional skills and gain experience that would facilitate their academic, career, and personal development.

This report describes the internship of Atom Arce, a recent high school graduate (High School for Math, Science and Engineering at CCNY - Class of 2018) and newly admitted first-year undergraduate student (Fall 2018) at the University of Toronto. During this internship, the undergraduate student intern worked with the TTI graduate student, Sirajum "Silvy" Munira, under the guidance of the project PI. The internship started July 5, 2018, and ended August 10, 2018.

Internship Structure

The internship started with the intern's standard TTI orientation, followed by a kick-off meeting between the PI, the graduate student and the undergraduate student intern. During this meeting, the team discussed what TTI does, what the Safe-D program is, and what the responsibilities during the internship period were. The key components of the discussion included:

- an overview of TTI as an organization,
- a brief description of the research areas of TTI,
- an overview of the Safe-D program and the current Safe-D project,
- the roles and responsibilities of the project team members,
- the goals of the internship, and
- the responsibilities during the 5-week internship (including tasks and deliverables).

At the beginning of each week, the intern received a detailed description of his responsibilities. The tasks of each week were designed according to his knowledge on the basic concepts of nonmotorized transport and its safety issues. The main task was that the intern read the assigned papers and then summarize and report his understanding of the topic. In addition to the assigned papers, he was encouraged to search and read other available materials for better understanding.

To maintain effective communication during internship, the intern attended scheduled meetings twice a week with the graduate student. As needed, occasional phone calls were also made in addition to the Skype call schedule. The Skype meeting schedule was:

- Monday, 12:00 p.m. — discuss the tasks, objectives, and deliverables (with schedule) for the week.
- Friday, 12.00 p.m. — discuss the progress of work during the week.

The intern received an Excel time sheet that identified the Skype call time, his work hours, and reporting time. The intern was allowed to deviate from the schedule with notice. Some of the articles were complicated, and required the intern to take more than one day to prepare the summary. After the intern submitted each summary report, the graduate student carefully examined each report and sent back her review the same day. The process was designed so that there is a continuous and effective communication between the students, the intern could promptly see and learn from the graduate student's review to prepare a better report the next time, and the graduate student could learn how to better manage the tasks and guide the intern.

During this process, the graduate student regularly updated the PI about the progress, and discussed the mentorship practice. The PI was also in continuous touch with the intern for any questions that he might have.

Work Plan

This section presents the details of each week's tasks and responsibilities.

Week 1

The first week's tasks were designed to enrich the intern's knowledge and understanding of different modes of transportation with a focus on nonmotorized activities, such as walking and biking. The papers outlined trends and patterns of walking and biking in the United States and other countries. The papers also provided information about transportation surveys including the American Community Survey (ACS) and the National Household Travel Survey (NHTS). Several socioeconomic and land use factors that affect bicycle and walking activities were discussed.

On Monday during the Skype call, the intern received detailed instruction about his tasks and the schedule of his work hours. The reports he was asked to read and summarize are:

- Report 1—McKenzie, B. (2014). *Modes Less Traveled: Bicycling and Walking to Work in the United States*. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau.
- Report 2—Pucher, J., Buehler, R., Merom, D., and Bauman, A. (2011). "Walking and Cycling in the United States, 2001–2009: Evidence from the National Household Travel Surveys." *American Journal of Public Health,* Vol. 101, No. S1, pp. S310–S317.
- Report 3—Kuzmyak, J. R., and Dill, J. (2012). "Walking and Bicycling in the United States: The Who, What Where, and Why." *TR News*, Vol. 280, pp. 4–7.

Table 1 provides the details of the work plan for week 1.

Table 1. Work Schedule for Week 1.

Date	Day	From	To	Action
July 9	Monday	12:00 p.m.	12:30 p.m.	Skype call
		12:30 p.m.	4:00 p.m.	Review Report 1
July 10	Tuesday	10:00 a.m.	4:00 p.m.	Review Report 1
		4:00 p.m.		Submit summary of Report 1
		6:00 p.m.		Receive feedback on Report 1 summary
July 11	Wednesday	12:00 p.m.	4:00 p.m.	Review Report 2
July 12	Thursday	9:00 a.m.	2:00 p.m.	Review Report 2
		2:00 p.m.		Submit summary of Report 2
		6:00 p.m.		Receive feedback on Report 2 summary
July 13	Friday	10:00 a.m.	12:00 p.m.	Review Report 3
		12:00 p.m.	12:30 p.m.	Skype call
		12:30 p.m.	2:00 p.m.	Review Report 3
		2:00 p.m.		Submit summary of Report 3
		6:00 p.m.		Receive feedback on Report 3 summary

The appendix contains the reports the intern prepared.

Week 2

The second week's task was designed to provide the intern with an overview of the safety issues of various travel modes. After gaining knowledge about overall crash statistics in the United States, the intern learned about the crash trends and statistics for both bicycle and pedestrian traffic. In addition to the sociodemographic distribution of fatal crashes of bicyclists and pedestrians, the reports provided important information about key factors contributing to nonmotorized crashes, including alcohol use, cell phone use, and lack of helmets. The reports also provided brief information about how different states are developing policies to make roads friendly for bicyclists and pedestrians.

After getting an overview of nonmotorized traffic activities and issues related to nonmotorized crashes and fatalities, the intern read and summarized three reports:

- Report 4—National Highway Traffic Safety Administration (2017). *2016 Motor Vehicle Crashes: Overview.* Traffic Safety Facts Research Note, 2017.
- Report 5—Retting, R. (2017). *Pedestrian Traffic Fatalities by State 2017 Preliminary Data.* Governors Highway Safety Association.
- Report 6**—**Williams, A. (2014). *Bicyclist Safety.* Governors Highway Safety Association.

The intern then created a summary report of the first six reports.

Table 2 provides the details of the work plan for week 2.

Table 2. Work Schedule for Week 2.

Date	Day	From	To	Action
July 16	Monday	12:00 p.m.	12:30 p.m.	Skype call
		12:30 p.m.	4:00 p.m.	Review Report 4
		4:00 p.m.		Submit summary of Report 4
		6:00 p.m.		Receive feedback on Report 4 summary
July 17	Tuesday	10:00 a.m.	4:00 p.m.	Review Report 5
		4:00 p.m.		Submit summary of Report 5
		6:00 p.m.		Receive feedback on Report 5 summary
July 18	Wednesday	12:00 p.m.	4:00 p.m.	Review Report 6
		4:00 p.m.		Submit summary of Report 6
		6:00 p.m.		Receive feedback on Report 6 summary
July 19	Thursday	9:00 a.m.	2:00 p.m.	Prepare Summary Report 1–6
July 20	Friday	10:00 a.m.	12:00 p.m.	Prepare Summary Report 1–6
		12:00 p.m.	12:30 p.m.	Skype call
		12:30 p.m.	2:00 p.m.	Prepare Summary Report 1–6
		2:00 p.m.		Submit Summary Report 1–6
		6:00 p.m.		Receive feedback on Summary Report 1–6

The appendix contains the reports the intern prepared.

Week 3

The objective of the first two weeks' tasks was to give the intern information about nonmotorized trends, crash issues, surveys to collect data, and factors affecting bicycle and pedestrian activities. To provide a brief overview of bicycle and pedestrian forecasting tools, the third week included reading a relatively complicated report about nonmotorized forecasting tools prepared for the Federal Highway Association.

The intern read and summarized:

- Report 7—Aoun, A., Bjornstad, J., DuBose, B., Mitman, M., Pelon, M., and Fehr and Peers (2015). *Bicycle and Pedestrian Forecasting Tools: State of the Practice.* DTFHGI-11-H-00024. Federal Highway Administration.

Table 3 provides the details of the work plan for week 3.

Table 3. Work Schedule for Week 3.

Date	Day	From	To	Action
July 23	Monday	12:00 p.m.	12:30 p.m.	Skype call
		12:30 p.m.	4:00 p.m.	Review Report 7
July 24	Tuesday	10:00 a.m.	4:00 p.m.	Review Report 7
July 25	Wednesday	12:00 p.m.	4:00 p.m.	Review Report 7
July 27	Friday	10:30 a.m.	12:00 p.m.	Review Report 7
		12:00 p.m.	12:30 p.m.	Skype call
		12:30 p.m.		Submit summary of Report 7
		6:00 p.m.		Receive feedback on Report 7 summary

The appendix contains the report the intern prepared.

Week 4

During the fourth week, the intern read a short literature review about emerging data sources for nonmotorized traffic.

The intern read and summarized:

- Report 8—A draft literature review report (work in progress - not yet published) about emerging data sources for nonmotorized traffic, prepared by the graduate student for the Safe-D project in 2018.

During this week, the intern started to work his internship presentation and overview report.

Table 4 provides the details of the work plan for week 4.

Table 4. Work Schedule for Week 4.

Date	Day	From	To	Action
July 30	Monday	12:00 p.m.	12:30 p.m.	Skype call
		12:30 p.m.	4:00 p.m.	Prepare internship overview report
July 31	Tuesday	10:00 a.m.	4:00 p.m.	Prepare internship overview report
August 1	Wednesday	12:00 p.m.	4:00 p.m.	Prepare internship presentation
August 2	Thursday	9:00 a.m.	2:00 p.m.	Review Report 8
		2:00 p.m.		Submit summary of Report 8
		6:00 p.m.		Receive feedback on Report 8
August 3	Friday	10:00 a.m.	12:00 p.m.	Prepare internship presentation
		12:00 p.m.	12:30 p.m.	Skype call
		12:30 p.m.	2:00 p.m.	Prepare internship presentation

The appendix contains the report the intern prepared.

Week 5

The fifth week was reserved for finalization of the presentation and the internship overview report.

For the presentation, the intern prepared slides and sent them to the graduate student for feedback. The graduate student provided her feedback to improve and finalize the slides.

The intern delivered his presentation August 9 to the project team including the PI and the graduate student. After the presentation, the PI provided feedback on the overall presentation and asked the intern questions about his understanding and perception. The meeting continued with rich and engaging discussions on various topics related to the project and the internship. Achievements and potential future improvements were also discussed. The meeting was concluded with discussions on data science and being a good data scientist. The intern was suggested to incorporate what he learnt from this final discussion in his overview report. The following provides a representative snapshot from the intern's overview report on the topic:

> "The two main skills I practiced the most were analyzing the information and distilling the most important details from my readings. I believe my ability to write about academic research vastly improved, and the feedback I received on my reports were especially helpful in finding the weaknesses in my analysis. A key tip we discussed was always keeping in mind what the reader should take away while keeping the topic understandable.
>
> When looking at the internship as a whole, one area I wish I had more time to study is the data analysis process. More time to work with the data collection and manipulation process would have been beneficial. At the same time, if I had not had the context of the field already, I would not have

> understood the information, and balance is necessary to get the most out of the experience. This brought up another important topic, the role of a data scientist. Too often, this term is applied to individuals who have some data software experience. More importantly, the data scientist finds insight that software cannot."

The intern submitted his internship overview report as the final deliverable of his internship program. The internship overview report summarizes the intern's understanding of the concepts from the reports and in person discussions from weeks 1 through 5. It also provides insights about the knowledge he gathered during internship and how he might use this knowledge.

The appendix provides the presentation and the overview report the intern prepared.

Internship Accomplishments

The internship produced several significant accomplishments for both the undergraduate student intern and the graduate student of the project. The key benefits of the internship are as follows:

- **Career path**: The internship provided the intern with an opportunity to explore the transportation field with a focus on safety and big data use. Because he is just beginning college with a major and career path in engineering, the experience gave him valuable insight into what the transportation industry does and whether it is the right choice for him.
- **Technical proficiency**: The internship provided the intern with a great opportunity in developing technical proficiency. He improved in distilling the key takeaways from the reports he read and in converting technical information into easily understandable content.
- **Writing skills**: During this internship, the intern learned to adapt the tone and style of technical writing. Because he received a review of each of his reports from the graduate student, he had an opportunity to learn and improve continuously. The process helped improve his writing significantly.
- **Working relationships**: The internship provided the intern with an understanding of valuable workplace traits. He took time to understand his role and responsibilities and adjust to the workplace culture. During this process, he learned to complete the given task within the deadline. He also learned not to be afraid of asking questions.
- **Communication skills**: The intern got an opportunity to practice his presentation skills at the end of his internship. The presentation helped him feel more confident speaking in front of people, expressing thoughts in a concise way, and engaging in an interactive discussion with others.
- **Networking and establishing references**: Through the internship, the intern had the opportunity to acquire networking skills. He also established a network of professional contacts and references, which is expected to help him in his future career path.
- **Leadership/mentoring skills**: In addition to various benefits for the intern, the internship program also provided the graduate student with an opportunity to learn and practice the key qualities needed for successful leadership. This included various aspects of leadership, such as effective communication, problem solving, relationship building, time management, building empathy for team members, etc.

Conclusion

The intern successfully completed his five-week internship under the Safe-D project. The key activities in which the intern participated included:

- Article review—The intern studied the current research related to various aspects of the Safe-D project he was a part of.
- Report writing—The intern learned to prepare summary reports of research articles.
- Presentation—The intern prepared and delivered a presentation to the project team.

During this process, he acquired knowledge about nonmotorized traffic patterns and use, safety issues related to bicycle and pedestrian traffic, bicycle and pedestrian forecasting tools, and the use of emerging and big data sets in the transportation area.

At the end of the internship, the intern noted that he found the transportation engineering field very interesting, and he is looking forward to exploring the area more in the future. A specific quote from the intern's final report highlighted the success of the internship program and emphasized its importance in shaping young minds for a brighter future: "This fall, I will be studying engineering at the University of Toronto, and I will be sure to use the skills I have acquired during this internship. The mindset of an engineer overlaps with the skills of a data scientist, and I believe this internship will benefit me in the future".

The process not only offered several benefits to the intern but also provided the graduate student with an opportunity to practice the key qualities for becoming an effective mentor. In conclusion, the internship program successfully reached the goals of education and technology transfer.

References

Aoun, A., Bjornstad, J., DuBose, B., Mitman, M., Pelon, M., and Fehr and Peers (2015). *Bicycle and Pedestrian Forecasting Tools: State of the Practice.* DTFHGI-11-H-00024. Federal Highway Administration.

Kuzmyak, J. R., and Dill, J. (2012). “Walking and Bicycling in the United States: The Who, What Where, and Why.” *TR News*, Vol. 280, pp. 4–7.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. *META Group Research Note*, Vol 6, pp. 70.

McKenzie, B. (2014). *Modes Less Traveled: Bicycling and Walking to Work in the United States*. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau.

National Highway Traffic Safety Administration (NHTSA) (2017). *2016 Motor Vehicle Crashes: Overview.* Traffic Safety Facts Research Note, 2017.

Pucher, J., Buehler, R., Merom, D., and Bauman, A. (2011). “Walking and Cycling in the United States, 2001–2009: Evidence from the National Household Travel Surveys.” *American Journal of Public Health,* Vol. 101, No. S1, pp. S310–S317.

Retting, R. (2017). *Pedestrian Traffic Fatalities by State 2017 Preliminary Data.* Governors Highway Safety Association.

Williams, A. (2014). *Bicyclist Safety.* Governors Highway Safety Association.

Appendix: Intern Reports[1]

Intern Report 1 (Week 1)

Intern's report based on: McKenzie (2014).

- Nonmotorized transportation—bicycling and walking.
- Many state and local agencies are promoting nonmotorized travel to complement major transportation systems.
- The American Community Survey (ACS) is a survey by the U.S. Census Bureau conducted every year and sent to around 3.5 million households.
- The data compiled have been used to highlight differences in biking/walking populations.
- The main population groups are social, economic, and geographic.
- This report uses ACS data from a 5-year span: 2008–2012.
- More U.S. workers are biking to work (greatest increase in commuting mode).
- Biking is more common in big cities.
- Younger workers tend to walk more.
- Means of transportation—principal mode of travel a worker takes while commuting.
- Principal city—largest city in each metropolitan area.
- Bicycling and walking are the least common modes of commuting at 0.6 percent and 2.8 percent, respectively.
- Automotive vehicles and public transportation are the leading modes at 86.2 percent and 5.0 percent, respectively.
- Walking has dropped steadily since 1980 from 5.6 percent to 2.8 percent.
- Bicycling has increased from 0.5 percent to 0.6 percent.
- Walking rates are highest in the Northeast and lowest in the South.
  - Walking commutes are more common in large cities.
- Bicycling is most popular in the West and least used in the South.
  - Biking is more common in large cities.
- Population density and infrastructure keep nonmotorized travel concentrated.
  - Smaller areas can also have higher rates of biking compared to their corresponding principal city, such as Davis and Sacramento, California.
- Biking and walking to work are both most common with workers age 16 to 24 and steadily decline across older age brackets.
- As household income increases, biking and walking percentages drop.
- Higher education leads to lower biking/walking rates until the graduate/professional degree level.
- Differentiating by race demonstrates how walking and biking percentages are consistent at 4 percent and 2 percent, respectively.
- Walking and biking both have lower average commute times than motorized travel.

[1] The intern's reports have gone through editorial review as part of the publication process of this report; however, no changes have been made to the content.

Intern Report 2 (Week 1)

Intern's report based on: Pucher et al. (2011).

This report by Pucher et al. uses data from the National Household Travel Survey (NHTS) for 2001 to 2009. The U.S. Department of Transportation conducts the NHTS to improve transportation policy and provide an in-depth view of how the American public travels. The report explains how active travel has numerous health benefits and reduces pollution attributed to automobiles. There has been a big push to promote nonmotorized transportation because the rates of workers walking and biking to work have declined in the past few decades.

The NHTS used landline telephone surveys and in-person interviews. The response rates were low and can be attributed to more Americans using cell phones. What the study revealed is an increase from 8.6 percent to 10.5 percent of walking trips and a slight increase of biking trips, 0.9 percent to 1.0 percent. The reasons Americans use active travel vary. Recreation/exercise accounted for 27 percent of walking trips and 49 percent of biking trips. Ninety percent of all public transportation trips are combined with walk trips as well. The prevalence of 30 minutes of physical activity due to walking/biking increased by 0.7 percent yet dropped in the categories of children, seniors, and women. When differentiating by race, White Americans do significantly less walking than minority Americans.

I thought this report added a vital analysis of the results, explaining the shortcomings of the techniques used and the potential factors that affected the rates of walking and biking. One area that could support the overall analysis is looking at city funding for nonmotorized travel spaces. This would allow the study of how successful certain rollouts have been.

Intern Report 3 (Week 1)

Intern's report based on: Kuzmyak and Dill (2012).

The authors draw a link between health and physical exercise and how decreasing access to nonmotorized travel is a health concern. The authors use data and information from the National Household Travel Survey to juxtapose America's low walking and biking rate to the rest of the world. Even compared to similarly wealthy countries, the United States' 11 percent walking and 1 percent biking rates are much lower than those of Switzerland, Spain, Germany, Sweden, and the United Kingdom.

As public interest has grown for walking and bicycling, data tools have had to evolve. Standard travel forecast models fail to include nonmotorized travel, and the traffic analysis zone is not refined enough to monitor walking and biking. Improved models that incorporate the required level of detail will not be available for the next few years. The most comprehensive biking and walking data come from regional surveys.

The majority of the data come from the count method, which gives researchers minimal information besides activity levels in a particular area. The most active demographic is children followed by young adults. Walking is densely connected to income because higher earners have better access to motorized travel. Biking and salary are consistent across income brackets, ranging from 0.9 percent to 1.3 percent. Education has a similar effect, with college graduates walking/biking less until the graduate level.

Some new data collection techniques include new sensors, pressure-sensitive devices, and image-processing programs. Although each method has its shortcomings, connecting these datasets will yield a more complete understanding of walking/cycling behavior. The average walking distance is 0.7 miles (15 minutes), and the biking average is 2.3 miles (19 minutes). The main reasons for both modes of travel are social and recreational (35.4 percent and 47.3 percent, respectively), while work-related travel is much lower at 4.5 percent and 10.9 percent, respectively.

The environment affects the price and accessibility of different modes of travel. Hills have a negative impact on the number of people willing to walk and bike. More variable weather leads commuters to gravitate to cars and public transportation. Infrastructure is critical because there has to be a proper setting to support walking and biking commuters. Trips have to be shorter and communities more compact for there to be more cyclists and pedestrians. Safety ties into a setting. Where there are designated spaces for cyclists, people feel safer traveling near normal traffic routes.

This paper does a great job of analyzing the main factors that contribute to low nonmotorized travel. The explanation of the different data collection processes gave me a better understanding of why transportation policies for biking and walking are so behind those for motorized vehicles. Many new techniques exist. However, it will take time for them to be added to current models and be used for future policies.

Intern Report 4 (Week 2)

Intern's report based on: NHTSA (2017).

This report is an overview of fatalities on U.S. roadways in 2016. The data come from the National Highway Traffic Safety Administration. In 2016, 37,361 people were killed in crashes, which was a 5.6 percent increase from 2015. In 2015, the 8.4 percent increase in fatalities was the greatest since 1964. Vehicle miles traveled (VMT) grew 2.2 percent in 2016, and the fatality rate per 100 million VMT increased by 2.6 percent.

The 2016 fatality count (37,461) was the highest since 2007, and the rate (1.18) was the highest since 2008. The fatality composition has changed significantly since 2007. Passenger car occupant fatalities have dropped from 40 percent to 36 percent. The primary increase has been pedestrian fatalities, which in 10 years have gone from 13 percent to 18 percent. Within the 5.6 percent overall fatality increase, the groups that had the highest rates of change were pedestrians (9.0 percent) and large-truck occupants (8.6 percent). The proportion of fatalities inside/outside the vehicle was 67 percent inside and 33 percent outside in 2015 and 2016, which shows a shift from the 80 percent inside and 20 percent outside from 1996 to 2000. When looking at fatalities from a crash-type perspective, rollover multivehicle and nighttime had the greatest increase at 9.1 percent and 6.3 percent, respectively. In human choice situations, unrestrained passenger fatalities grew the most at 4.6 percent, while alcohol-impaired driving only increased by 1.7 percent. Within the alcohol-impaired-involved accidents, passenger cars are the largest group with 3.1 percent in 2015. Fatalities for people who used restraints increased by 4.8 percent (+520), and fatalities for those who did not increased by 4.6 percent (+460). By age, the highest increase in fatalities from 2007 to 2016 was 65+, and the age bracket 16–24 had the most significant decrease at −26.8 percent.

This report had detailed data that resulted in a very concise analysis of the different group fatalities. I would like to see more data on traffic safety policies and how funding has changed awareness and the implementation of better traffic systems.

Intern Report 5 (Week 2)

Intern's report based on: Retting (2017).

This report comes from preliminary data reported by State Highway Safety Offices to study the recent increase in pedestrian deaths during the last few years. Pedestrian fatalities have grown 27 percent from 2007 to 2016, while all other traffic deaths decreased by 14 percent.

The U.S. Census Bureau's American Community Survey reported 4 million Americans walk to work, which is a 4 percent increase since 2007. At the same time, pedestrian fatalities went from 11 percent to 16 percent as a percentage of total motor vehicle deaths. The fatality rate per 100,000 people is over 2 for 15 states, demonstrating how widespread the problem is. The rates range from New Mexico's 3.45 to Nebraska's 0.68. Local municipal streets account for 33 percent of pedestrian fatalities, with state and U.S. highways following at 25 percent and 16 percent, respectively. In 2016, 77 percent of all pedestrian deaths happened after dark, with five states having more than 80 percent.

A majority of on-foot accidents happen in non-intersection areas, with only 18 percent of fatalities happening at intersections. Alcohol involvement is reported at 46 percent of pedestrian fatalities, with pedestrians making up 33 percent and drivers 13 percent. During the first six months of 2017, five states, which make up 30 percent of the U.S. population (California, Florida, Texas, New York, and Arizona), accounted for 43 percent of pedestrian deaths. In the 10 major cities, the number of fatalities from 2015 to 2016 increased by 28 percent.

To reduce the number of fatalities, there are a couple different approaches. The first is to increase the separation of pedestrians from motorists, which can be achieved by adding more sidewalks or giving the pedestrian more control over intersections. The second is to increase the visibility of the walker to traffic, and the most common way to solve this is by adding more lighting. The most important counter is to improve public knowledge so people have a better awareness of the problem and how they can contribute to safer travel habits.

Intern Report 6 (Week 2)

Intern's report based on: Williams (2014).

This report is written by Dr. Allan Williams and was sponsored by the Governors Highway Safety Association. Biking is one of the least common methods of transportation, but its popularity has grown and so have fatalities. Even though cycling accidents only represent 2 percent of all motor vehicle deaths, between 2010 and 2012, American bicyclist deaths increased by 16 percent. However, 54 percent of these deaths are concentrated in six states—California, Florida, Illinois, New York, Michigan, and Texas, while 23 states average five or fewer accidents per year.

Biking fatalities once mainly involved children, but now more adults (age 20+) are involved, going from 21 percent in 1975 to 84 percent in 2012, being primarily male at 74 percent. The setting has shifted to urban areas from 50 percent in 1975 to 69 percent in 2012. Nonmotorists are known vulnerable road users. This is because during bicycle and motor vehicle incidents, the risk is asymmetric; cyclists are prone to serious injuries. At the same time, cyclists tend not to wear helmets, with 46 percent saying they never use them. In fatal accidents, 65 percent of those killed are not wearing any headgear. Twenty-one states have laws about youth helmet use, but there are no laws for all-age bikers. Alcohol use is also very prevalent in fatal cycling accidents, with 25 percent found with blood alcohol levels over 0.08 percent.

The proposed solutions revolve around separation of motorists from cyclists and providing better cycling infrastructure. Visibility is a huge problem, and more cities are building bike lanes with better lighting. Education is also key since most biking accidents happen due to bad biking behavior, whether it be related to alcohol or not wearing a helmet for protection.

Intern Summary Report 1–6 (Week 2)

Intern's report based on: McKenzie (2014), Pucher et al. (2011), Kuzmyak and Dill (2012), NHTSA (2017), Retting (2017), and Williams (2014).

As shown in *Modes Less Traveled* by Brian McKenzie, bicycling and walking are the least common modes of commuting at 0.6 percent and 2.8 percent, respectively, according to 2008–2012 American Community Survey data. In J. Richard Kuzmyak's article, he juxtaposes America's low walking and biking with rates in the rest of the world. Even compared to similarly wealthy countries, the United States' 11 percent walking and 1 percent biking rates are much lower than those of Switzerland, Spain, Germany, Sweden, and the United

Kingdom. However, nonmotorized travel is becoming more popular with its links to health benefits and low environmental impact.

At the moment, walking and biking are concentrated in large cities and for short commutes. Compared to other modes of transportation, walking trips average 0.7 miles (15 minutes), and the biking average is 2.3 miles (19 minutes). Young workers (age 16–24) have the highest rates of nonmotorized commuting at 6.8 percent for walking and 1.0 percent for biking, and rates drop off the higher the age bracket. Education has a similar effect, with college graduates walking/biking less until the graduate degree level. Biking and salary are consistent across income brackets, ranging from 0.9 percent to 1.3 percent. According to Kuzmyak, the main reason for people to walk or bike is social and recreational reasons (35.4 percent and 47.3 percent, respectively), while work-related travel is much lower at 4.5 percent and 10.9 percent, respectively. Yet, Pucher et al. explain that 90 percent of all public transportation trips are combined with walking trips, demonstrating the demand for walking in conjunction with other forms of travel.

Nonmotorized transportation is not as well documented as automotive travel. The majority of walking and biking data come from the count method, which gives researchers minimal information besides activity levels in a particular area. As public interest has grown for walking and bicycling, the data tools used have had to evolve. Standard travel forecast models fail to include nonmotorized travel, and the traffic analysis zone is not refined enough to monitor walking and biking. Some new data collection techniques include new sensors, pressure-sensitive devices, and image-processing programs. Although each method has its shortcomings, connecting these datasets will yield a more complete understanding of walking/cycling behavior. Improved models that incorporate the required level of detail will not be available for the next few years. The most comprehensive biking and walking data still rely on regional surveys.

Even though nonmotorized travel is uncommon, fatalities are still a problem. The National Highway Traffic Safety Administration reported 37,361 people were killed in crashes in 2016, which was a 5.6 percent increase from 2015. This increase in deaths is connected to the vehicle miles traveled (VMT), which grew 2.2 percent in 2016, and the fatality rate per 100 million VMT increased by 2.6 percent. As road accidents increase, one group is more at risk. The primary fatality increase has been pedestrians; in 10 years, pedestrian deaths have gone from 13 percent to 18 percent. The proportion of fatalities inside/outside the vehicle was 67 percent inside and 33 percent outside in 2015–2016, which shows a shift from the 80 percent inside and 20 percent outside from 1996 to 2000. During the first six months of 2017, five states, which make up 30 percent of the U.S. population (California, Florida, Texas, New York, and Arizona), accounted for 43 percent of pedestrian deaths. A majority of on-foot accidents happen in non-intersection areas, with only 18 percent of fatalities happening at intersections. Some proposed reasons for the sudden increase in fatalities are the legalization of marijuana, alcohol consumption, smartphone use, lack of biking headgear, and weak transportation infrastructure.

To reduce the number of fatalities, there are a couple different approaches. The first is to increase the separation of pedestrians from motorists, which can be achieved by adding more sidewalks or giving the pedestrian more control over intersections. The second is to increase the visibility of the walker to traffic, and the most common way to do this is adding more lighting. The third is to enforce the law more, whether it be to mandate helmets for cyclists or punish infractions such as drunk driving. The most important counter is to improve public knowledge so people have better awareness of the problem and how they can contribute to safer travel habits.

Intern Report 7 (Week 3)

Intern's report based on: Aoun et al. (2015).

This report is sponsored by the Federal Highway Administration and focuses on transportation forecasting models and how they are changing. These models are used to assess activity and inform future transportation policy and new development. Currently, nonmotorized travel does not have the same level of analysis. To improve these models, data are pulled from the U.S. Census and even cell phones. Also, though many communities are pushing for active transportation, biking and walking make up only 11 percent of national travel and 16 percent of travel in big cities.

Key definitions:

- Big data: Large datasets that need unique processing and integration with other datasets to be understood.
- Civic technologies: Apps created to increase communication between the government and its constituents.
- Crowdsourcing: Using information gathered by users to answer a specific problem.
- Forecasting: Using models to predict an outcome.
- Regression: A statistical process for determining the relationship between variables.

The different types of forecasting tools can be classified by structure, purpose, and geographic scope. In structure, aggregate tools view a collection of datasets to make predictions, whereas disaggregate tools analyze decisions at the individual level and make assumptions about the different types of individuals within a given population. In purpose, demand tools describe activity by time, length, and reason, while project prioritization looks at usage levels to determine new decisions. The geographic scope is broken up into regional, subarea, and local/community planning levels.

At the local level, which is most appropriate for biking and pedestrian forecasting because these trips are usually under 3 miles, there are three methods: factor/sketch planning tools, aggregate demand models, and biking share forecasting. Factor/sketch planning tools use existing count data and are useful for the ease of extensive data collection but lack contextual factors and produce weaker forecasts. Aggregate demand models create an equation based on the existing dataset and can be formed using simple statistical software yet fail to include individual choice and factors. Biking share forecasting is a combination of geographical information systems (GIS) and spatial tools to make a regression equation of ridership within a given area, which benefits from the availability of data but struggles to account for riders from outside the system.

With the quick forecasting model tools, there are different ways of estimating future demand. Conclusions can be drawn from the American Community Survey and census data, traffic volume analysis, traffic stress, and viewing trips from travel models. A majority of these techniques rely on count and survey data, which are easily attainable; they differ by the level of knowledge needed to understand the trends. These techniques are limited by location-based attributes that are difficult to account for, and also the number of resources needed to conduct these studies.

The benefits of aggregate demand models are the minimum software requirements and the use of existing public data. The disadvantages are that aggregate demand models undercut the importance of the individual's choice and are costly for the necessary count data. Bike share forecasting uses elements of aggregate demand models but also uses GIS tools to get information about population, land use, density, and transportation frequency. The data are heavily reliable because they come directly from these bike share stations; however, it is difficult to link the data to other datasets.

Network simulation tools use various links and nodes with datasets to determine activity levels, travel time, and collisions. These tools are highly detailed and sophisticated to the degree of pedestrians and cyclists. However, developing one of these systems is highly time consuming, and collecting the data is very costly.

Regional travel demand forecasting tools use the following steps: trip generation, trip distribution, mode choice, and traffic assignment. With these analyses, governments can plan and maintain transportation networks. The tools are readily accessible since all major urban areas require these models for forecasting and have plenty of financial support. However, the tools are not effective for nonmotorized travel because the scale of the model is better suited to automotive transportation. The tools also require specialized software and need to be geographically calibrated for the quality of the data.

Activity and tour-based models take into account the complexity of different households, focus on the individual's trip decisions and how they interact with the transportation system around them, and assess how demand will change. This approach is highly nuanced, analyzing multi-stop trips, which creates highly detail conclusions. On the other hand, this technique is extremely resource intensive, requires much more information than a standard household survey, and is challenging to scale up to a large population.

Overall, the field of cycling and pedestrian forecasting is changing rapidly, and there is no one best way to analyze nonmotorized travel. Forecasting uses elements of spatial analysis, simulation models, and general household surveys, and more research is needed about what techniques can link different datasets. Funding is key to any of these tools becoming useful and producing significant findings.

Intern Report 8 (Week 4)

> Intern's report based on a draft report (work in progress - not yet published) about emerging data sources for nonmotorized traffic, prepared by the graduate student for the project (2018).

According to Laney, big data have three components: volume (size), velocity (speed of generation or collection), and variety (synthesizing a range of sources). Big data have immediate connections to nonmotorized travel because analysis of these modes of transportation requires much more detail. These large datasets come from three sources: global positioning systems (GPS), live point, and journey data.

GPS data come from cell phone apps, wearable tech, and bike sharing systems. The process researchers use to generate and collect all this information from stakeholders is called *crowdsourcing*. It incorporates amateur input in scientific projects, which has been coined *citizen science.* The emergence of crowdsourcing allows researchers to fill in gaps in studies that normal funding cannot address. The difficulty with using crowdsourced data is reliability since the conditions may not have the same standards as in a normal scientific study from a big agency. Certain apps and websites are geared toward specific demographics, and their datasets do not properly represent a population.

Currently, demand forecasting models have plenty of variance in quantity, specificity, and scale of data used. For future predictions about nonmotorized travel, these models generally struggle to capture the nuances of walking and cycling. While household surveys and GPS tracking are well established in studying motorized travel, nonmotorized travel primarily relies on count data done by volunteers in limited areas. Not only are the data limited to certain areas, but biking and pedestrian traffic also varies more depending on time and location, making short-term counts inefficient. To improve count data, devices such as pneumatic tubes, inductive loops, thermal cameras, infrared sensors, magnetometers, piezoelectric devices, radar sensors, and video imaging are being implemented in nonmotorized systems. These automated counters give an accurate continuous summary of activity and traffic but are expensive to install and may not be suitable for a large areas.

GPS data can get to the level of tracking cyclists' travel patterns in conjunction with roadway environments. With these data, profiles of riders' preferences can be built to create a model of where routes are most needed. GPS apps also allow cyclists to document their trips. This documentation gives researchers critical characteristics of what types of facilities work best for bicyclists, and allows researchers to find sources of stress in major routes. Fitness app data can also be used to determine preferences among riders. These crowdsourced data help identify the different purposes of trips: commuting, exercise, and travel—and how they affect the areas riders gravitate toward. Factors such as slope, speed, and traffic density affect biking negatively. The only problem with these findings is that the data come from a small sample size with variance from app to app. Biking sharing system data have been used to find links between biking rates and weather condition, timing, and location. Data from existing biking sharing systems have been used to develop new bike share programs in new cities and countries.

Overall, big data can help analyze the nuances of nonmotorized travel better than standard forecasting models. With the many different datasets and crowdsourcing efforts, big data can be more cost efficient than expensive count surveys and more widely used.

Internship Presentation (Week 5)

TTI Internship Overview

Texas A&M Transportation Institute

Atom Arce

Table of Contents

1. Internship structure
2. Week 1- Nonmotorized Transportation
3. Week 2- Fatalities
4. Week 3- Cyclist Safety
5. Week 4- Forecasting Models and Big Data
6. Takeaways

Internship Structure

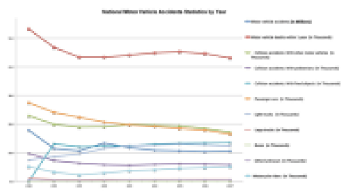
1. Read and write summary of reports
2. Learned from the review of the reports
3. Develop a weekly subject report

Nonmotorized Transportation

- ❑ Read three papers on Nonmotorized activity
- ❑ Understood various surveys such as the ACS and NHTS
- ❑ Learned about trend, concentration, influencing factors etc.

Fatalities

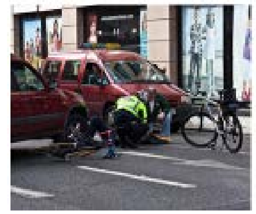
I then transitioned to fatalities. The papers covered were 2016 Fatal Motor Vehicle Crashes- National Highway Traffic Safety Administration and Pedestrian Traffic Fatalities by State by the Governors Highway Safety Association. 37,361 people were killed in crashes in 2016 which was a 5.6% increase from 2015. This increases in deaths is connected to the Vehicle Miles Traveled (VMT) growing 2.2% in 2016, and the fatality rate per 100 million VMT increasing by 2.6%. Even though biking rates are extremely low across the U.S, bikers have a higher fatality rate growing at a rate of 9.0% in a year.



5

Cyclist Safety

Within the subject of fatalities, I focused on cyclist safety and read the Bicyclist Safety report created by Governors Highway Safety Association. In fatal bike accidents, 65% are not wearing any headgear and alcohol use is also very prevalent with 25% found with BAC levels over 0.08%. Now more adults(20+) are involved going from 21% in 1975 to 84% in 2012. 54% of these deaths are concentrated in six states: California, Florida, Illinois, New York, Michigan, and Texas, demonstrating the concentration of accidents to bigger cities.



Forecasting Models and Big Data

A common conclusion was bicycle and pedestrian models are much less accurate compared to motorized forecasting models. I wanted to learn more about the different techniques in building datasets, so I read Bicycle and Forecasting Tools: State of the Practice created by the Federal Highway Administration and and Sirajum's Emerging Big Data Sources for NMT. I learned changing variables such as scope, source, trip details, and household data, and including influence the use of a model. The prevalence of Crowdsourcing (collecting amateur data) means there are more opportunities to amass direct data from cyclists and pedestrians therefore better models.

7

Takeaways

- Writing skills
- Professionalism
- Greater appreciation for the transportation research community
- Data scientist mindset

8

Internship Overview Report (Week 5)

Over the last several weeks, I have learned about nonmotorized transportation. My work schedule revolved around analyzing academic papers and creating reports summarizing my understanding of the topic. Apart from the set papers, I did seek out other sources to better understand the intricacies of transportation forecasting.

I started the internship by exploring the different modes of transportation and how Americans choose to commute to work. A majority of the U.S. transportation data come from surveys like the American Community Survey and the National Household Travel Survey. Even though these studies have different purposes and scopes, what is clear is nonmotorized travel is much less popular than all other modes of transportation. Bicycling and walking are the least common modes of commuting at 0.6 percent and 2.8 percent, respectively, according to McKenzie. These rates are significantly lower than European countries like Switzerland, Spain, and Germany. Because of the layout of American transportation systems, biking and walking are concentrated in large cities and for short commutes. Walking trips average 0.7 miles (15 minutes), and the biking average is 2.3 miles (19 minutes). Young workers (age 16–24) have the highest rates of nonmotorized commuting at 6.8 percent for walking and 1.0 percent for biking, and rates drop off the more senior the age bracket.

I studied vehicle fatalities in the United States and saw that roadway deaths are up. The National Highway Traffic Safety Administration reported 37,361 people were killed in crashes in 2016, which was a 5.6 percent increase from 2015. This increase in deaths is connected to the vehicle miles traveled (VMT) growing 2.2 percent in 2016, and the fatality rate per 100 million VMT increasing by 2.6 percent. Within the 5.6 percent overall fatality increase, the groups that had the highest rates of change were pedestrians (9.0 percent) and large-truck occupants (8.6 percent). What is interesting is how more likely nonmotorists are to be killed in these accidents, which demonstrates a need for change in transportation safety policies.

Within the focus on fatalities, I moved on to bicyclist safety and why this demographic is in danger. In fatal accidents, 65 percent of those killed are not wearing any headgear. Twenty-one states have laws about youth helmet use, but there are no laws for all-age bikers. Alcohol use is also very prevalent in fatal cycling accidents, with 25 percent found with blood alcohol levels over 0.08 percent. My takeaway from viewing biking statistics was that it is more challenging to regulate cyclists, and individual decisions such as alcohol use have a much higher potential for accidents. Factors such as cell phone use, intoxication, and lack of helmets are all connected to bad cycling behavior, and education has a considerable role in correcting these careless accidents.

A common theme in the first two weeks was the emergence of cities and communities pushing for more people using nonmotorized modes of transportation. At the same time, bicycling and pedestrian data are not as fleshed out as forecast models for motorized transportation systems. The proposed solutions included improving count data sources using new technology, but big data seem to solve many of the problems standard studies miss. I then moved on to studying forecasting tools and how big data can improve these techniques.

What I learned was that there are many ways to develop a model of transportation, by changing variables such as scope, source, trip details, and household data, and including multiple datasets. Big data have the advantage of linking different datasets from user data. Crowdsourcing, which is the process of researchers using amateur input to fill gaps regular surveys do not have the resources to cover, is becoming more of a trend in transportation studies. The problems with using big data are that the sources can be incompatible because data points can come from bike sharing services, smartphone apps, and a range of other devices and services. However, in the future, gathering data with this technique will allow for greater analysis as software improves.

I entered this internship with minimal knowledge of the transportation industry and will leave with a greater appreciation for how researchers conduct studies and how the field is changing. During my presentation, we discussed topics such as privacy and the importance of explaining how data benefit both the researchers and the users who provide their information. The two main skills I practiced the most were analyzing the information and distilling the most important details from my readings. I believe my ability to write about academic research vastly improved, and the feedback I received on my reports were especially helpful in finding the weaknesses in my analysis. A key tip we discussed was always keeping in mind what the reader should take away while keeping the topic understandable.

When looking at the internship as a whole, one area I wish I had more time to study is the data analysis process. More time to work with the data collection and manipulation process would have been beneficial. At the same time, if I had not had the context of the field already, I would not have understood the information, and balance is necessary to get the most out of the experience. This brought up another important topic, the role of a data scientist. Too often, this term is applied to individuals who have some data software experience. More importantly, the data scientist finds insight that software cannot.

This fall, I will be studying engineering at the University of Toronto, and I will be sure to use the skills I have acquired during this internship. The mindset of an engineer overlaps with the skills of a data scientist, and I believe this internship will benefit me in the future.