IDAHO TRANSPORTATION DEPARTMENT

RESEARCH REPORT

Bicycle and Pedestrian Safety Research Project

RP 301

By

Steve Jackson
Paula Raymond
Stephen Taylor

Prepared for

Idaho Transportation Department

ITD Research Program, Planning Services

Highways Development

October 2023

YOUR Safety • • • ▶ YOUR Mobility • • • ▶ YOUR Economic Opportunity

Disclaimer

This document is disseminated under the sponsorship of the Idaho Transportation Department and the United States Department of Transportation in the interest of information exchange. The State of Idaho and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the view of the authors, who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official policies of the Idaho Transportation Department or the United States Department of Transportation.

The State of Idaho and the United States Government do not endorse products or manufacturers. Trademarks or manufacturers' names appear herein only because they are considered essential to the object of this document.

This report does not constitute a standard, specification, or regulation.

Technical Report Documentation Page

1. Report No. FHWA-ID-23-301	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Bicycle and Pedestrian Safety Research Project		5. Report Date October 2023 6. Performing Organization Code
7. Author(s) Steve Jackson, https://orcid.org/0000-0002-3337-7846 Paula Raymond, https://orcid.org/0009-0001-8020-1501 Stephen Taylor		8. Performing Organization Report No.
9. Performing Organization Name and Address Toxcel 7140 Heritage Village Plaza Gainesville, VA 20155		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Idaho Transportation Department Highways Development, Planning Services, Research Program PO Box 7129 Boise, ID 83707-7129		13. Type of Report and Period Covered Final Report 11/14/2022 – 11/13/2023 14. Sponsoring Agency Code RP 301
15. Supplementary Notes Project performed in cooperation with the Idaho Transportation Department and Federal Highway Administration.		
16. Abstract Crash typing is the process of analyzing crash details to categorize crashes into similar groups, which can help identify hotspots, rising problems, and deploy the most effective countermeasures. This research analyzed crashes involving motor vehicles and bicycles or pedestrians in Idaho between 2012 and 2021. Using the PBCAT2 methodology, with supplementary machine learning tools, the study reviewed 2,739 bicycle and 2,209 pedestrian crash records. Findings indicate a decline in bicyclist crashes, but a steady rate in pedestrian crashes with a notable rise in fatalities, particularly in specific scenarios such as near parking lots. Key risk factors include motorist speeding, midblock crossings for bicyclists, and pedestrians walking on roads. The study suggests countermeasures such as expanded bicycle facilities, enhanced lighting, and enhanced law enforcement to improve safety for vulnerable road users in Idaho.		
17. Key Words pedestrian, bicyclist, crash typing, active transportation, safety	18. Distribution Statement Copies available from the [ITD Research Program]	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 63 / 22. Price None

Technical Advisory Committee

Each research project is overseen by a Technical Advisory Committee (TAC), which is led by an ITD project sponsor and project manager. The TAC is responsible for monitoring project progress, reviewing deliverables, ensuring that study objectives are met, and facilitating implementation of research recommendations, as appropriate. ITD's Research Program Manager appreciates the work of the following TAC members in guiding this research study.

- Project Sponsor: Scott Luekenga
- Project Manager: Margaret Havey
- TAC Members: Kelly Campbell, Margaret Pridmore, Denise Dinnauer
- FHWA-Idaho Advisor: Lance Johnson

Table of Contents

List of Tables

List of Figures

List of Maps

List of Abbreviations and Acronyms

CIEACT	Crash Information Extraction Analysis and Classification Tool
CSV	Comma-Separated Values
DUALIST	Utility for Active Learning with Instances and Semantic Terms
FARS	Fatality Analysis Reporting System
GES	General Estimates System
LMCM	Location-Movement Classification Method
M	Mean
ML	Machine Learning
MUTCD	Manual on Uniform Traffic Control Devices
N	Number (i.e., crash count, sample size, etc.)
NASS	National Automotive Sampling System
NHTSA	National Highway Traffic Safety Administration
PBCAT	Pedestrian and Bicycle Crash Analysis Tool
SD	Standard deviation

Executive Summary

Crashes involving motor vehicles and pedestrians or bicycles can vary in many ways. Each crash involves the specific movements and characteristics of the road users involved, the surrounding infrastructure, and a host of other factors including alcohol use and lighting. The ability to partition these crashes into similar groups can help to identify hotspots, identify rising problems, and deploy the most effective countermeasures. *Crash typing* describes this process of analyzing crash details to categorize crashes into a manageable number of groups.

The research team conducted a literature review to explore crash typing methodologies and their use in selecting countermeasures to improve safety. Pedestrian and Bicycle Crash Analysis Tool, Version 2 (PBCAT2) emerged as the preeminent methodology, though there is merit to clustering algorithms. PBCAT2 requires manual review of each crash narrative to extract the necessary data elements. Several tools powered by machine learning algorithms have been developed to facilitate this process.

This research effort collected 10 years (2012-2021) of records pertaining to crashes involving motor vehicles and bicycles (N = 2,739) or pedestrians (N = 2,209). Crash narratives were submitted to a large language model to extract additional information. Crashes were then aligned to PBCAT2 crash groups and typed using a hierarchical clustering algorithm.

The population in Idaho grew by 19.3% between 2012 and 2021, and vehicle miles traveled in Idaho increased by 23.3% over the same period. Results indicate that bicycle crashes of all types have declined over the ten-year period, while pedestrian crashes overall have held steady; however, fatalities have increased. Several crash types exhibit increasing trends: those occurring around parking lots, alleys, and driveways; and those involving motorists failing to signal at intersections, making improper left turns, speeding near turns and hills.

Risk factors for fatal or serious injury include motorist speeding and impairment, midblock crossings (for bicyclists), and walking along the roadway (pedestrians). With the help of crash type models, high occurrence corridors were identified and mapped. Countermeasures, including expanding bicycle facilities, restricting curbside parking and loading, minimizing visual clutter, installing additional lighting, reducing curb radii, and enforcing laws intended to prevent distracted driving are suggested to improve safety for bicyclists and pedestrians in Idaho. Also included as a strategy is providing advanced medical training to improve the chance of survival in the event of a crash.

1. Introduction

Crash typing is a method of categorizing crashes into a manageable number of groups. The term "crash type" is used liberally and carries different meanings in different contexts. For example, some sources use crash type and injury severity synonymously (Schnee et al., 2021), others use it as shorthand for the mode of the road users involved, and others use non-standardized typologies (Bhowmik et al., 2021; Intini et al., 2020; Li et al., 2020). More informative crash types describe events and maneuvers of the involved parties that led up to a crash (Thomas et al., 2018). These definitions aid in the selection of appropriate countermeasures and can provide valuable insight by producing homogeneous crash groups for researchers to drill down into when searching for factors that may influence injury severity.

The National Highway Traffic Safety Administration (NHTSA) funded the development of the first crash typing methodology in the 1970s. Snyder et al. (1971) collected highly detailed data from interviews with victims, witnesses, and police officers on 2,157 pedestrian crashes occurring in 13 U.S. cities. They developed a typology of 27 crash types with corresponding countermeasures by classifying crashes based on precipitating events, predisposing factors, and target groups. In 1977, researchers collected data from 919 bicyclist/motorist crashes in cities in California, Colorado, Florida, and Michigan (Cross & Fisher, 1977). They identified 36 "problem types," each one characterized by the traffic context in which the crash occurred, the operators' function failures, and the combination of factors causally related to the function failures. Smist (1982) introduced the first computer-assisted (pedestrian) crash typing procedure. These efforts culminated in the first official set of crash types endorsed by NHTSA, with 37 pedestrian-motor-vehicle crash types and 45 bicycle-motor-vehicle crash types (NHTSA, 1983b, 1983a).

This research report presents a review of crash typing methodologies and develops a methodology to assign crash types to crashes in Idaho involving bicycles and pedestrians. Trends and hot spots are identified, and appropriate countermeasures are recommended.

2. Literature Review

This review summarizes and evaluates the available methods and tools for pedestrian and bicycle ("pedbike") crash typing, and how crash typing has been used to select countermeasures to improve safety. One tool directly facilitates crash typing, while others seek to replace or supplement the manual review of crash narratives with machine learning techniques. The following sections describe these methods and their corresponding typologies.

The Pedestrian and Bicycle Crash Analysis Tool (PBCAT)

The Pedestrian and Bicycle Crash Analysis Tool (PBCAT) encompasses a methodology, typology, and tool; the tool guides users through the methodology to produce a particular typology by asking analysts to answer questions about the events leading up to the crash. Responses are recorded with user-provided crash IDs and resulting crash types, and a text file is produced for further analysis. Version 1 (PBCAT1), released in 1999 was developed to simplify the process and enable States to perform their own crash typing.

Ragland et al. (2003) used PBCAT1 to develop a pedestrian countermeasures plan for San Francisco. Authors also outline the steps to implement appropriate countermeasures: select the crash problem (crash type), map crashes and develop candidate zones, calculate injury densities, select final zones, then select countermeasures (based on cost, presumed safety effectiveness, ease of implementation, and other criteria) and evaluate.

Early users called for a reduction in the number of crash types and a better connection to countermeasures (Harkey & Blomberg, 2001). PBCAT2 was released in 2006 to address these criticisms. This version introduced crash type groups (16 groups to describe 56 pedestrian crash types, 21 groups to describe 79 bicyclist crash types) that closely correspond to popular countermeasure guides PEDSAFE (Harkey & Zegeer, 2004) and BIKESAFE (Hunter et al., 2006). Table 1 lists the PBCAT2 pedestrian and bicyclist crash type groups, and indicates which groups are included in these guides.

Table 1. PBCAT2 crash type groups, by mode.

Pedestrian	*Bicycle*
Backing Vehicle*	Backing Vehicle
Bus-Related*	Bicyclist Failed to Yield, Midblock*
Crossing Driveway or Alley	Bicyclist Failed to Yield, Signalized Intersection*
Crossing Expressway*	Bicyclist Failed to Yield, Sign-Controlled Intersection*
Crossing Roadway, Vehicle Not Turning*	Bicyclist Left Turn/Merge*
Crossing Roadway, Vehicle Turning*	Bicyclist Overtaking Motorist*
Dash/Dart-Out*	Bicyclist Right Turn/Merge*
Multiple Threat/Trapped*	Crossing Paths, Other Circumstances
Off Roadway*	Head-On
Other/Unknown, Insufficient Details	Loss of Control/Turning Error
Pedestrian in Roadway, Circumstances Unknown	Motorist Failed to Yield, Midblock*
Unique Midblock*	Motorist Failed to Yield, Signalized Intersection*
Unusual Circumstances	Motorist Failed to Yield, Sign-Controlled Intersection*
Waiting to Cross	Motorist Left Turn/Merge*
Walking Along Roadway*	Motorist Overtaking Bicyclist*
Working or Playing in Roadway*	Motorist Right Turn/Merge*
	Nonroadway
	Other/Unknown, Insufficient Details
	Other/Unusual Circumstances
	Parallel Paths, Other Circumstances
	Parking/Bus-Related

Note: crash type groups denoted with an asterisk (*) appear in PEDSAFE (D. L. Harkey & Zegeer, 2004) or BIKESAFE (Hunter et al., 2006)

PBCAT2 is the current standard in crash typing in the United States. This typology has appeared in the Fatality Analysis Reporting System (FARS), National Automotive Sampling System (NASS), and General Estimates System (GES) since 2010 (National Highway Traffic Safety Administration, 2022). Its use is recommended as part of systematic pedestrian safety analysis (Thomas et al., 2018) and in the selection of safety improvement projects (Natarajan et al., 2008).

Carter and Council (2006) applied PBCAT2 to the analysis of rural pedbike crashes in Florida and proposed countermeasures using PEDSAFE and BIKESAFE. Spainhour et al. (2006) also applied the PBCAT2 methodology to Florida crashes, focusing on fatal pedestrian crashes. Researchers found that pedestrian behaviors were the primary contributing factors in most cases. Zegeer et al. (2008) present the full lifecycle of a pedestrian safety improvement program in Miami, Florida, from crash typing using PBCAT2 to countermeasure selection, implementation, and evaluation. Researchers deployed problem-specific countermeasures and conducted a thorough before-after analysis to document significant improvements in key pedestrian groups including school-aged children.

Thomas et al. (2019) used PBCAT2 to compare data from FARS, the state of North Carolina, and the city of Boulder, Colorado. Vavrova et al. (2021) describe the process of applying the PBCAT2 methodology/tool to more than 10,000 pedbike crash reports in Texas and prescribe countermeasures specific to each crash type. Other applications in Arizona, Colorado, and Pennsylvania have also been documented (Chavis et al., 2018; Do & Harkey, 2006).

Version 3 was released in 2022 with several changes compared to prior versions (Thomas et al., 2022). It is a web application (available at https://www.pbcat3.org/), whereas prior versions required software installation and some technical expertise. More importantly, crash types are no longer determined separately for pedestrians and bicyclists, and “the precrash maneuvers of the motorist and non-motorist are the only two variables that influence the Crash Type" (Thomas et al., 2022, p. 22). Table 2 shows the resulting combinations that yield 79 *detailed* crash types involving a motor vehicle colliding with a pedestrian or bicyclist. Some non-motorist maneuvers are combined to produce 32 *basic* crash types.

The relation between PBCAT2 and PBCAT3 is unclear due to “substantial” changes to the logic and variable definitions (Thomas et al., 2022, p. 77). Because both versions require information typically only captured in crash narratives, this may impose a large burden for States conducting ongoing crash typing efforts.

PBCAT3 has been used to compare e-scooter and bicycle crashes in Tennessee (Shah et al., 2021), and to determine which types of vehicles were associated with various pedestrian crash types in North Carolina (Hu & Cicchino, 2022).

Table 2. PBCAT3 Detailed Crash Type Matrix. Source: (Thomas et al., 2022)

	CR: Crossing path from motorist’s right	CL: Crossing path from motorist’s left	CU: Crossing path, unknown direction	PS: Parallel path, same direction	PO: Parallel path, opposite direction	PU: Parallel path, unknown direction	MU: Moving in unknown path/direction	ST: Stationary	OU: Other/ unusual	UN: Unknown	FC: Non-motorist fall or crash
S: Going straight	S-CR	S-CL	S-CU	S-PS	S-PO	S-PU	S-MU	S-ST	S-OU	S-UN	S-FC
R: Turning right	R-CR	R-CL	R-CU	R-PS	R-PO	R-PU	R-MU	R-ST	R-OU	R-UN	R-FC
L: Turning left	L-CR	L-CL	L-CU	L-PS	L-PO	L-PU	L-MU	L-ST	L-OU	L-UN	L-FC
P: Parked	P-CR	P-CL	P-CU	P-PS	P-PO	P-PU	P-MU	P-ST	P-OU	P-UN	P-FC
E: Entering traffic lane	E-CR	E-CL	E-CU	E-PS	E-PO	E-PU	E-MU	E-ST	E-OU	E-UN	E-FC
B: Backing	B-CR	B-CL	B-CU	B-PS	B-PO	B-PU	B-MU	B-ST	B-OU	B-UN	B-FC
O: Other maneuver	O-CR	O-CL	O-CU	O-PS	O-PO	O-PU	O-MU	O-ST	O-OU	O-UN	O-FC
U: Unknown maneuver	U-CR	U-CL	U-CU	U-PS	U-PO	U-PU	U-MU	U-ST	U-OU	U-UN	U-FC
N: Non-collision	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	N-FC

The Location-Movement Classification Method (LMCM)

In 2016, a novel methodology and typology, the Location-Movement Classification Method (LMCM), expanded on PBCAT2 by explicitly considering, "(a) the location of the crash relative to an intersection or roadway segment and (b) the direction of pedestrian or bicyclist movement relative to the motor vehicle movement" (Schneider & Stefanich, 2016, p. 72). Note that PBCAT2 requests this information but does not use it to determine crash type. The LMCM typology includes 57 distinct crash types for both pedestrians and bicyclists (114 in total). **Error! Not a valid bookmark self-reference.** presents the LMCM coding scheme. For example, crash type N_RRD_X corresponds to a non-intersection crash on the right side of the roadway with no or unknown pedbike movement.

Table 3. Location-Movement Classification Method coding scheme (Schneider & Stefanich, 2016).

Main Crash Category	First Part of Code	Second Part of Code	Third Part of Code	Fourth Part of Code
Roadway intersection	General location I = roadway intersection	Side of intersection NS = nearside, or where motorist enters intersection FS = farside, or where the motorist exits intersection	Motorist movement ST = straight LT = left turn RT = right turn	Pedestrian or bicyclist movement relative to motorist's preturn direction R = approaching from motorist's right L = approaching from motorist's left S = same direction as motorist O = opposite direction as motorist X = no or unknown direction
Roadway nonintersection	General location N = roadway nonintersection	Location on the roadway RRD = right-side roadway lane LRD = left-side roadway lane RSH = right-side shoulder or bike lane LSH = left-side shoulder or bike lane RSW = right-side sidewalk LSW = left-side sidewalk	Pedestrian or bicyclist movement [a] R = approaching from motorist's right L = approaching from motorist's left S = same direction as motorist O = opposite direction as motorist X = no or unknown direction	None
Parking lot or private property	General location P = parking lot D = driveway	Motorist movement F = forward B = backward	None	None
Other [b]	All crashes have single code OTH	None	None	None

[a] R and L movements are not used for shoulder or sidewalk crashes.

[b] Other crashes include situations that do not fit into the categories, including driverless vehicle crashes or multiunit crashes where a pedestrian or bicyclist was struck by a vehicle that had already been struck by another vehicle.

Comparison of PBCAT2 and LMCM

The literature provides a limited comparison of the PBCAT2 and LMCM typologies. Table 4 compares the three most common crash types identified by two studies using both methodologies. Schneider and Stefanich (2016) applied both methodologies to 234 pedestrian and 155 bicycle fatal/severe injury crashes in Wisconsin. Authors cross-tabulate crashes to demonstrate the many-to-many relationships between the two typologies. For example, the most common pedestrian crash was 741 (Dash, PBCAT2) or N_RRD_X (non-intersection, right side of roadway, no or unknown pedbike direction, LMCM). This PBCAT2 crash type maps to five other LMCM types, and this LMCM type maps to five other PBCAT2 types.

Chavis et al. (2018) typed pedbike crashes in Washington, DC, using PBCAT2, the LMCM, and decision trees. The two methods provide very different types of information about the crash. Decision trees also identified traffic control type, crash time, alcohol, speeding, light condition, road type, city quadrant, and fault as contributing factors in more severe pedestrian crashes, and construction zones as factors in more severe bicycle crashes. PEDSAFE and BIKESAFE are used to recommend countermeasures.

Amsden and Huber (2006) also typed crashes in Wisconsin, though only bicyclist-motor vehicle crashes and only using the PBCAT2 methodology. Researchers typed over 1,000 crashes and found the following most common crash types: 141 (motorist drive-out, sign-controlled intersection), 144 (bicyclist ride-through, sign-controlled intersection), and 212 (motorist left turn, opposite direction). Several of these crash types were also identified in Schneider and Stefanich (2016) and Chavis et al. (2018).

Table 4. Comparison of most common crash types identified using PBCAT2 and LMCM methodologies (continued next page).

Location (Source)	*Road User*	*Most Common PBCAT2 Crash Types*	*Most Common LMCM Crash Types*
Wisconsin (Schneider & Stefanich, 2016)	Pedestrian	741 – Dash	N_RRD_X – non-intersection, right side of roadway, no or unknown pedbike direction
Wisconsin (Schneider & Stefanich, 2016	Pedestrian	770 – Motorist failed to yield	I_FS_ST_L – intersection, farside, motorist travelling straight, pedbike approaching from motorist's left
Wisconsin (Schneider & Stefanich, 2016	Pedestrian	742 – Dart out	N_RRD_R – non-intersection, right side of roadway pedbike approaching from motorist's right
Wisconsin (Schneider & Stefanich, 2016	Bicyclist	231 – Motorist overtaking, undetected bicyclist	N_RRD_X – non-intersection, right side of roadway, no or unknown pedbike direction
Wisconsin (Schneider & Stefanich, 2016	Bicyclist	212 – Motorist left turn, opposite direction	I_NS_ST_L – intersection, nearside, motorist travelling straight, pedbike approaching from motorist's left
Wisconsin (Schneider & Stefanich, 2016	Bicyclist	141 – Motorist drive-out, sign-controlled intersection	I_FS_ST_R – intersection, farside, motorist travelling straight, pedbike approaching from motorist's right

Location (Source)	*Road User*	*Most Common PBCAT2 Crash Types*	*Most Common LMCM Crash Types*
Washington, DC (Chavis et al., 2018)	Pedestrian	781 – Motorist left turn, parallel paths	I_NS_ST_X - Intersection, nearside, motorist travelling straight, no or unknown pedbike direction
Washington, DC (Chavis et al., 2018)	Pedestrian	770 – Motorist failed to yield	N_RRD_X – non-intersection, right side of roadway, no or unknown pedbike direction
Washington, DC (Chavis et al., 2018)	Pedestrian	760 – Pedestrian failed to yield	I_FS_LT_O – intersection, farside, motorist left turn, pedbike moving in opposite direction
Washington, DC (Chavis et al., 2018)	Bicyclist	244 – Bicyclist overtaking, extended door	N_RRD_S – non-intersection, right side of roadway, pedbike moving in opposite direction as motorist
Washington, DC (Chavis et al., 2018)	Bicyclist	212 – Motorist left turn, opposite direction	I_NS_ST_S – intersection, nearside, motorist travelling straight, pedbike moving in same direction
Washington, DC (Chavis et al., 2018)	Bicyclist	213 – Motorist right turn, same direction	I_FS_LT_O – intersection, farside, motorist left turn, pedbike moving in opposite direction

Clustering Algorithms

Both PBCAT2 and LMCM require crash narratives, which are often difficult to attain, require manual review, and are subject to judgment and error. These methods also ignore other potentially relevant factors such as lighting conditions, alcohol involvement, age, and dozens of other variables typically captured in State and nationwide crash databases. A variety of machine learning methods have been applied to deal with the lack of narratives and wealth of other data.

Sun et al. (2019) applied a k-means clustering algorithm to 14,236 pedestrian crashes in Louisiana, then examined each cluster individually for factors associated with increased injury severity. Researchers noted differences between generally predictive factors and cluster-specific factors. For example, pedestrians crossing or entering the roadway was not predictive of fatal/severe injury overall, but it was in the cluster of nighttime crashes involving alcohol or drugs. Depaire et al. (2008) conducted a similar analysis on 4,028 pedbike crashes in Belgium, producing seven clusters with similar conclusions. Another study identified seven clusters in New York City and five in Montreal, concluding that clustering should be used “not only for descriptive analysis, but also as a preliminary segmentation tool for a more detailed, standard statistical analysis” (Mohamed et al., 2013, p. 35).

Song et al. (2021) presents an exemplary analysis of crash types with countermeasure recommendations. Researchers used PBCAT2 crash type groups in part of a larger clustering effort aimed at identifying upward trending hotspots of fatal pedestrian and bicyclist crashes and provided results to planners and the public in an interactive web application. The application incorporates secondary data – including exposure metrics – maps crashes, assesses corridors, and recommends appropriate countermeasures.

One approach applied a clustering algorithm to questionnaire responses to develop a unique typology of bicyclist crash types. Billot-Grasset et al. (2014) collected information from 1,078 French bicyclists involved in crashes with motor vehicles, including: crash date and place, injury severity, age, gender, trip purpose, traffic environment, infrastructure, road surface conditions, objects avoided or hit, road users' trajectories, weather and light conditions, and bicyclist equipment. The analysis produced 17 clusters. Authors note important differences between their findings and the early work of Cross et al. (1977) such as the absence of a nighttime factor despite its emergence as a key variable in the clustering process. Chavis et al. (2018) typed pedbike crashes in Washington, DC, using PBCAT2, the LMCM, and decision trees. PBCAT2 and LMCM results are discussed in the previous section of this review. The decision trees identified traffic control type, crash time, alcohol, speeding, light condition, road type, city quadrant, and fault as contributing factors in more severe pedestrian crashes, and construction zones as factors in more severe bicycle crashes. These factors are not considered in PBCAT2 or LMCM methodologies.

These techniques can leverage a wealth of information but can also produce unique typologies. This may be advantageous within a single State's analysis, as States vary in terms of crash environments, infrastructure, and other factors. However, a non-standardized typology makes inter-State comparisons difficult or impossible.

Other Machine Learning Applications

More recently, machine learning (ML) techniques have been applied to the crash narratives themselves. The review identified two tools that seek to facilitate this process.

The Utility for Active Learning with Instances and Semantic Terms (DUALIST) was released in 2012 as a resource to help researchers train models to classify texts into custom groups (Settles, 2012). Users read through texts (crash reports) and select which terms and phrases are indicative of each group (crash type). The model can then be applied to new texts to categorize them accordingly. This tool has been used to significantly reduce the time to classify the narrative text of traumatic brain injuries (Chen et al., 2016).

DUALIST was released prior to recent advances in natural language processing, (Radford et al., 2019), pre-training and fine-tuning (Howard & Ruder, 2018), and transfer learning (Pan & Yang, 2010). Sayed et al. (2022) developed a publicly available tool specifically to read and classify crash narratives with less manual labor. The Crash Information Extraction Analysis and Classification Tool (CIEACT) accepts CSV files of crash narratives and allows users to apply pre-existing models or train new ones. Current models are limited to identifying crashes involving work zones and driver distraction. This tool represents a remarkable step forward in the typing of crashes from crash narratives.

Such techniques have been applied to motorcycle crashes (Das et al., 2021) and heavy vehicle crashes in Australia (Arteaga et al., 2020) and workers' compensation claims (Bertke et al., 2016; Marucci-Wellman et al., 2017). Das et al. (2020) applied three ML algorithms to read crash narratives and determine fault; the best algorithm accurately determined fault with 77% accuracy in a sample of roughly 150 pedestrian crashes (trained on 265 crashes). Montella et al. (2011) used classification trees and association rules to explore the severity of pedestrian crashes in Italy. Other machine learning and language processing algorithms can be implemented in popular programming languages such as Python and R (Mesevage, 2020; Wolff, 2020).

3. Methodology

This research effort sought to improve safety for bicyclists and pedestrians by aggregating relevant data into a georeferenced database, implementing crash typing methodologies, and conducting systematic analysis to identify safety hotspots and trends. Findings were used to make targeted countermeasure recommendations. This section provides further detail on the methodologies used.

The Idaho Transportation Department (ITD) stores and analyzes crashes in a non-public database known as WebCars. The research team worked with ITD staff to gain access to WebCars and extract 10 years of records pertaining to crashes involving motor vehicles and bicycles or pedestrians. This data includes 123 unique data elements covering 2,739 bicycle crashes and 2,209 pedestrian crashes between 2012 and 2021. These elements exclude identifying information such as names and officer badge numbers but include the full text of the officer-generated crash narrative. All data manipulation and analysis were conducted using the R language for statistical computing (R Core Team, 2023).

Crash data stored in WebCars is organized in three different data tables pertaining to the crash, the vehicles involved, and the people involved. The research team conducted standard data validation and cleaning steps, then aggregated the three tables into one "flat" table. The "flat" table uses indicator variables to organize all information pertaining to one crash onto one row. For example, a crash involving a 45-year-old driver and 20-year-old bicyclist would appear as `driver.age_41_60=1` and `pedbike.age_25_under=1`, with other variables capturing additional details. This manipulation was necessary to cluster the crashes without losing information about the people and vehicles involved. Only data describing vehicle drivers, bicyclists, and pedestrians was considered; data for non-driving passengers was discarded. A georeferenced database including all collected data and crash type labels, as well as documentation, was delivered to ITD.

Two crash typing methodologies were implemented: PBCAT2 and clustering.

Table 5 provides the PBCAT2 crash group definitions. Note that this report uses the term "crash type" to refer to what the PBCAT2 technique calls "crash groups." Most crash group definitions were successfully implemented by using a combination of data from crash records and narratives. Some crash groups (bicyclist 220, 225, 240; pedestrian 800) were omitted due to a lack of data, while others (bicyclist 190, 290, 850; pedestrian 600, 990) were omitted for lack of specificity.

Table 5. PBCAT2 crash group definitions (continued next page). Source: (D. Harkey et al., 2006)

Mode	*Crash Group Number*	*Crash Group Label*	*Definition*	*Source*
Bicycle	110	Loss of Control/Turning Error	Either the motorist or the bicyclist lost control of their vehicle or made a turning error and inadvertently moved into the path of the other operator. Note: Includes loss of control due to mechanical problems or operator error or turning errors such as traveling into the opposing lane.	Narratives
Bicycle	140	Motorist Failed to Yield, Sign-Controlled Intersection	The motorist drove into the crosswalk area or intersection and collided with the bicyclist. The motorist either violated the sign or did not properly yield right-of-way to the bicyclist. Note: Crashes at traffic circles or roundabouts with yield control are included here.	Crash records, narratives
Bicycle	145	Bicyclist Failed to Yield, Sign-Controlled Intersection	The bicyclist rode into the intersection and collided with the motorist. The bicyclist either violated the sign or did not properly yield right-of-way to the motorist. Note: Crashes at traffic circles or roundabouts with yield control are included here.	Crash records, narratives
Bicycle	150	Motorist Failed to Yield, Signalized Intersection	The motorist drove into the crosswalk area or intersection and collided with the bicyclist. The motorist either violated the signal or did not properly yield right-of-way to the bicyclist.	Crash records, narratives
Bicycle	158	Bicyclist Failed to Yield, Signalized Intersection	The bicyclist rode into the intersection and collided with the motorist. The bicyclist either violated the signal or did not properly yield right-of-way to the motorist.	Crash records, narratives
Bicycle	190	Crossing Paths, Other Circumstances	The bicyclist and motorist were on initial crossing paths, but the crash cannot be further classified.	Omitted
Bicycle	210	Motorist Left Turn/Merge	The motorist made a left turn or merge into the path of a bicyclist traveling in the same or opposite direction.	Crash records, narratives
Bicycle	215	Motorist Right Turn/Merge	The motorist made a right turn or merge into the path of a bicyclist traveling in the same or opposite direction.	Crash records, narratives
Bicycle	219	Parking/Bus-Related	The bicyclist was struck by a motorist entering or exiting a parking space or by a bus or delivery vehicle pulling into or away from the curb.	Crash records
Bicycle	220	Bicyclist Left Turn/Merge	The bicyclist made a left turn or merge into the path of a motor vehicle traveling in the same or opposite direction.	Omitted
Bicycle	225	Bicyclist Right Turn/Merge	The bicyclist made a right turn or merge into the path of a motor vehicle traveling in the same or opposite direction.	Omitted
Bicycle	230	Motorist Overtaking Bicyclist	The motorist was overtaking the bicyclist at the time of the crash.	Crash records
Bicycle	240	Bicyclist Overtaking Motorist	The bicyclist was overtaking the motorist at the time of the crash. Note: This group includes crashes involving bicyclists striking parked cars or extended doors.	Omitted
Bicycle	258	Head-On	Either operator was going the wrong way and the two parties collided head-on.	Crash records, narratives
Bicycle	290	Parallel Paths, Other Circumstances	The bicyclist and motorist were on initial parallel paths, but the crash cannot be further classified.	Omitted
Bicycle	310	Bicyclist Failed to Yield, Midblock	The bicyclist rode into the street from a nonintersection location (including residential or commercial driveway or other midblock location) without yielding to the motorist.	Crash records, narratives
Bicycle	320	Motorist Failed to Yield, Midblock	The motorist drove across the sidewalk or into the street from a nonintersection location (including residential or commercial	Crash records, narratives

Mode	***Crash Group Number***	***Crash Group Label***	***Definition***	***Source***
			driveway or other midblock location) without yielding to the bicyclist.	
Bicycle	600	Backing Vehicle	The motorist was backing up at the time the crash occurred.	Crash records, narratives
Bicycle	850	Other/Unusual Circumstances	There were unusual circumstances surrounding the crash, but the crash cannot be further classified.	Omitted
Bicycle	910	Nonroadway	The crash occurred off the road network such as in a parking lot, driveway, on a multi-use path separated from the road right-of-way, in an open grassy area or yard, etc.	Crash records, narratives
Pedestrian	100	Unusual Circumstances	The crash involved a disabled vehicle, emergency vehicle or vehicle in pursuit, play vehicle, driverless vehicle, or the pedestrian was struck intentionally, was clinging to a vehicle, or was struck as a result of other unusual circumstances.	Crash records, narratives
Pedestrian	200	Backing Vehicle	The pedestrian was struck by a vehicle that was backing at the time.	Crash records
Pedestrian	310	Working or Playing in Roadway	The pedestrian was working or playing in the roadway.	Crash records
Pedestrian	340	Bus-Related	The pedestrian was struck while crossing/walking to a bus or bus stop or while waiting at a bus stop.	Crash records, narratives
Pedestrian	350	Unique Midblock	The crash was associated with a vendor truck, mailbox, or other roadside 'destination' that was not a bus, or the pedestrian was struck while entering or exiting a parked vehicle.	Crash records
Pedestrian	400	Walking Along Roadway	The pedestrian was standing or walking along the roadway on the edge of a travel lane, or on a shoulder or sidewalk.	Crash records, narratives
Pedestrian	460	Crossing Driveway or Alley	The pedestrian was crossing a driveway on a sidewalk crossing, shared-use path, shoulder, or edge of the travel lane.	Crash records, narratives
Pedestrian	500	Waiting to Cross	The pedestrian was standing on the curb or near the roadway edge waiting to cross the roadway when struck.	Narratives
Pedestrian	600	Pedestrian in Roadway, Circumstances Unknown	The pedestrian was standing, walking, or lying in the road right-of-way at an intersection or midblock location but the circumstances do not otherwise fit any previously described or are unknown.	Omitted
Pedestrian	720	Multiple Threat/Trapped	The pedestrian entered the roadway on a green signal or in front of standing or slowing traffic and was trapped when the signal changed, and traffic started moving or was struck by a vehicle traveling in the same direction as the stopped traffic. Note: Multiple threat may occur at nonsignalized locations.	Narratives
Pedestrian	740	Dash/Dart-Out	The pedestrian either ran into the roadway in front of a motorist whose view of the pedestrian was not obstructed or walked or ran into the road and was struck by a motorist whose view of the pedestrian was blocked until an instant before impact.	Narratives
Pedestrian	750	Crossing Roadway, Vehicle Not Turning	The pedestrian was struck while crossing the roadway (not an expressway) by a vehicle that was traveling straight through.	Crash records
Pedestrian	790	Crossing Roadway, Vehicle Turning	The pedestrian was struck while crossing a non-expressway road by a vehicle that was turning or about to turn.	Crash records
Pedestrian	800	Off Roadway	The pedestrian was struck in a parking lot, driveway, open area or other or unknown, nonroadway area (vehicle not backing).	Omitted
Pedestrian	910	Crossing Expressway	The pedestrian was on an expressway or expressway ramp when struck by a motor vehicle.	Crash records
Pedestrian	990	Other/Unknown, Insufficient Details	The circumstances do not clearly fit any of the situations described or are unknown.	Omitted

PBCAT2 has specific information requirements (e.g., did the bicyclist lose control or make a turning error) that are not consistently captured in standard crash reports. The research team used a large language model (GPT-3) to extract additional information from the crash narratives. Narratives were first sanitized by randomly changing all detected names. For example, “John was travelling east on Marshall Avenue…John did not see the bicyclist” might be changed to “Tabatha was travelling east on Arnold Avenue…Tabatha did not see the bicyclist.” Note that names were changed consistently within each narrative so that the large language model could persistently identify all individuals involved. As PBCAT2 generates mode-specific crash types, different questions were posed depending on the mode involved (bicycle or pedestrian). To further facilitate the dialogue, possible responses were enumerated. Questions and responses by mode are presented in Table 6.

Table 6. Questions and possible responses submitted to large language model.

Mode	*Question*	*Possible responses*
Bicycle	Did the bicyclist lose control or make a turning error?	Yes, no
Bicycle	Did the driver of the vehicle lose control or make a turning error?	Yes, no
Bicycle	Did the bicyclist fail to yield?	Yes, no
Bicycle	Did the driver fail to yield	Yes, no
Bicycle	Were the bicyclist and driver travelling in the same, opposite, or intersecting directions?	Same, opposite, intersecting
Bicycle	Did the driver contribute to the crash by opening their door?	Yes, no
Pedestrian	Did this crash involve any of the following: a disabled vehicle, an emergency vehicle, a police vehicle in pursuit, a driverless vehicle? If so, which one?	Disabled vehicle, emergency vehicle, police vehicle in pursuit, driverless vehicle
Pedestrian	Did the driver hit the pedestrian intentionally?	Yes, no
Pedestrian	Was the pedestrian hit while walking to a bus stop?	Yes, no
Pedestrian	Was the pedestrian hit while waiting at a bus stop?	Yes, no
Pedestrian	Was the pedestrian hit while exiting a parked vehicle?	Yes, no
Pedestrian	Was the pedestrian hit while walking along the roadway on the edge of a travel lane, or on a shoulder or sidewalk? If so, which one?	Pedestrian was not hit while walking along the roadway, travel lane, shoulder, sidewalk
Pedestrian	Was the pedestrian hit while crossing a driveway?	Yes, no
Pedestrian	Was the pedestrian hit while waiting to cross the road?	Yes, no
Pedestrian	Was the pedestrian trapped at a median?	Yes, no
Pedestrian	Was the pedestrian hit by a vehicle traveling in the same direction as the stopped traffic?	Yes, no
Pedestrian	Did the pedestrian run into the roadway in front of a motorist whose view of the pedestrian was not obstructed?	Yes, no
Pedestrian	Did the pedestrian run into the roadway in front of a motorist whose view of the pedestrian *was* obstructed?	Yes, no
Pedestrian	Did the driver see the pedestrian?	Yes, no

The research team assessed the performance of the large language model by manually reviewing 100 crashes involving each mode; overall, it was considered accurate in 83% of bicycle questions and 93% of pedestrian questions. Table 7 provides further details on the accuracy of the model on each set of questions tied to an individual crash: approximately 80% of crashes included 1 error or less. Although this technique does not produce perfect responses, it is comparable to a human effort while requiring a fraction of the time and resources.

Table 7. Large language model performance details

Number of inaccurate responses per crash	*Bicycles*	*Pedestrians*
0	45%	38%
1	34%	47%
2	13%	9%
3 or more	8%	6%

Whereas PBCAT2 uses a predetermined set of crash types, clustering methods allow crash types to emerge from the available data. The hierarchical clustering algorithm (with a binary distance metric and Ward's agglomeration method) used hundreds of indicator variables to identify ten unique crash types for each mode. Most indicator variables were created from categorical variables. For example, the roadway functional class variable was converted to four separate indicators: `functional_class_Local`, `functional_class_Major_Collector`, `functional_class_Minor_Arterial`, and `functional_class_Principal_Arterial`. Others binned numeric variables into relevant groups (e.g., `driver.age_41_60`, `pedbike.age_25_under`). The resulting clusters were labelled based on the relative prevalence of variables. For example, if 90% of crashes in Hypothetical Cluster A occurred on major collector roads, compared to less than 5% in other clusters, Cluster A may be labelled as "major collector roads."

The research team documented crash trends overall and by crash type. Those with worsening trends were further examined to recommend appropriate countermeasures.

4. Results

This section visualizes and describes trends in bicycle and pedestrian crashes overall, then presents the crash typing results by mode and method.

Overall Trends

Figure 1 shows annual bicycle and pedestrian crashes, fatalities, and injuries. Bicycle crashes and injuries show a strong downward trend, while fatalities remain very low and relatively flat. On the contrary, pedestrian crashes and injuries are flat, while fatalities show a strong increasing trend.

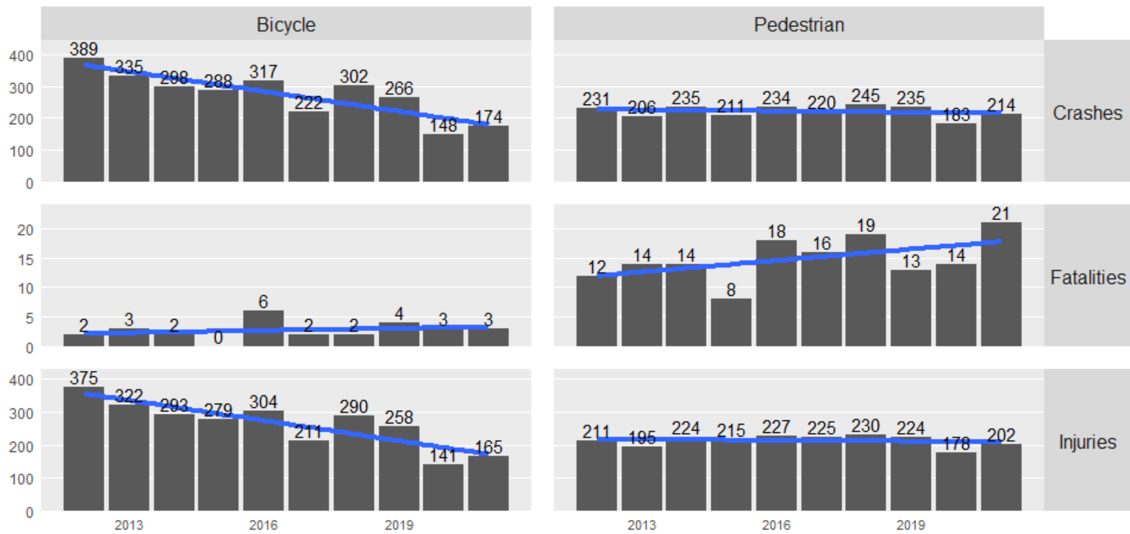


Figure 1. Annual bicycle and pedestrian crashes, fatalities, and injuries (2012-2021).

Figure 2 shows the temporal aspects of crashes. Both bicycle and pedestrian crashes spike during the morning and evening commute times, with the former outnumbering the latter between rush hours. Crashes involving both modes are higher during the week, likely coinciding with traditional commute patterns. There is significant variation throughout the year: bicycle crashes increase during the warmer months while pedestrian crashes increase slightly during the colder months.

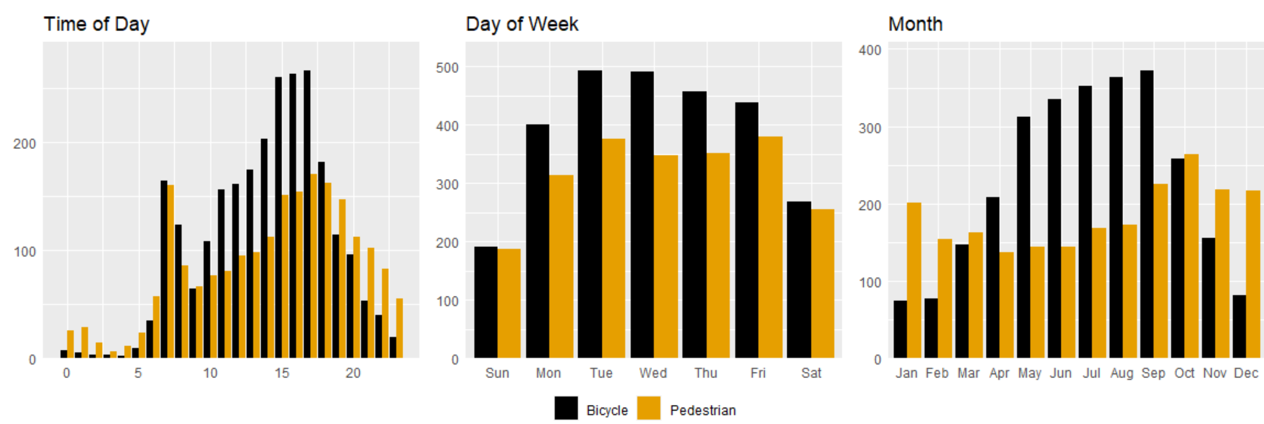


Figure 2. Bicycle and pedestrian crashes (2012-2021) by time of day, day of week, and month

Figure 3. Annual bicyclist and pedestrian injuries by severity.

shows annual bicycle and pedestrian injuries by severity. Fatalities were consistently higher among pedestrians than bicyclists. Suspected serious injuries show a notable difference between the two road users: serious injuries among bicyclists have decreased since 2012, but similar injuries among pedestrians show a large increase around 2016. Both minor and possible injuries appear to be significantly lower in 2020 and 2021, likely due to changes brought on by the COVID-19 pandemic. Non-injury crashes were consistently low over the ten-year period.

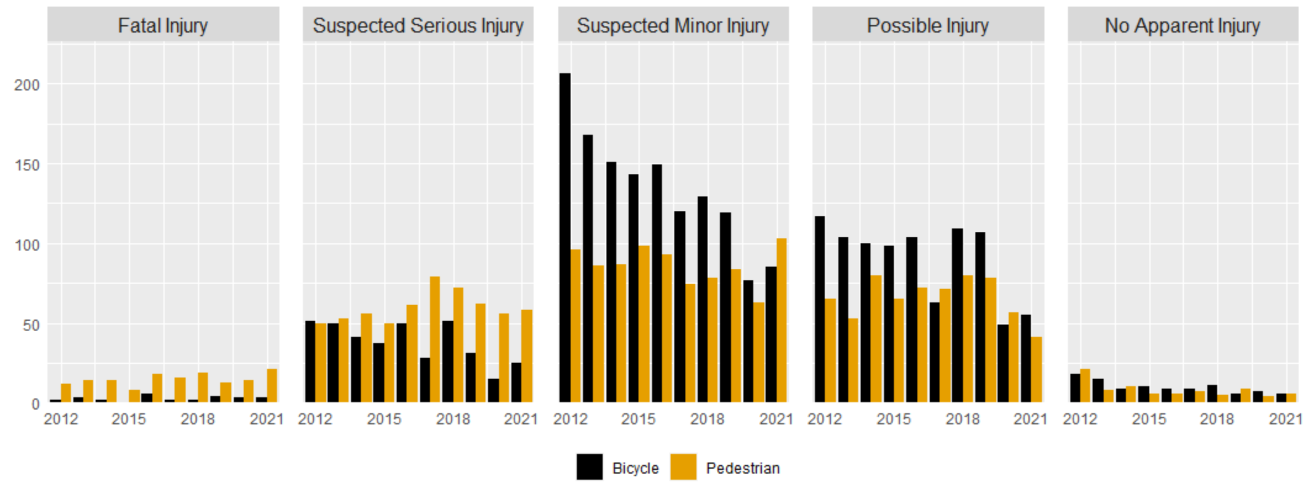


Figure 3. Annual bicyclist and pedestrian injuries by severity.

Figure 4 shows the ages of individuals involved in these crashes, separately for crashes involving bicycles and those involving pedestrians. The median age for drivers was 40 years in bicycle crashes and 39 years in pedestrian crashes. Bicyclists tended to be slightly younger than pedestrians, with a median age of 24 years versus 28 years.

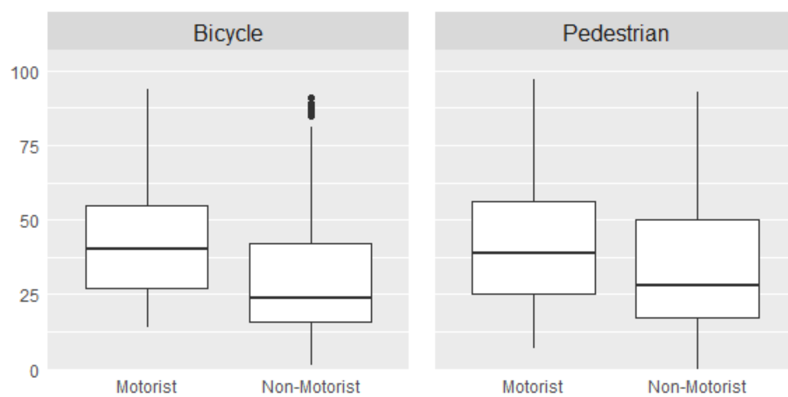


Figure 4. Ages of drivers, bicyclists, and pedestrians involved in crashes (2012-2021).

Figure 5 further explores the ages of individuals involved in crashes involving bicycles and pedestrians. Despite some short-term increases, all age groups are generally trending downward. By 2021, the most common age group was 21-30 for drivers, and 11-20 for both bicyclists and pedestrians.

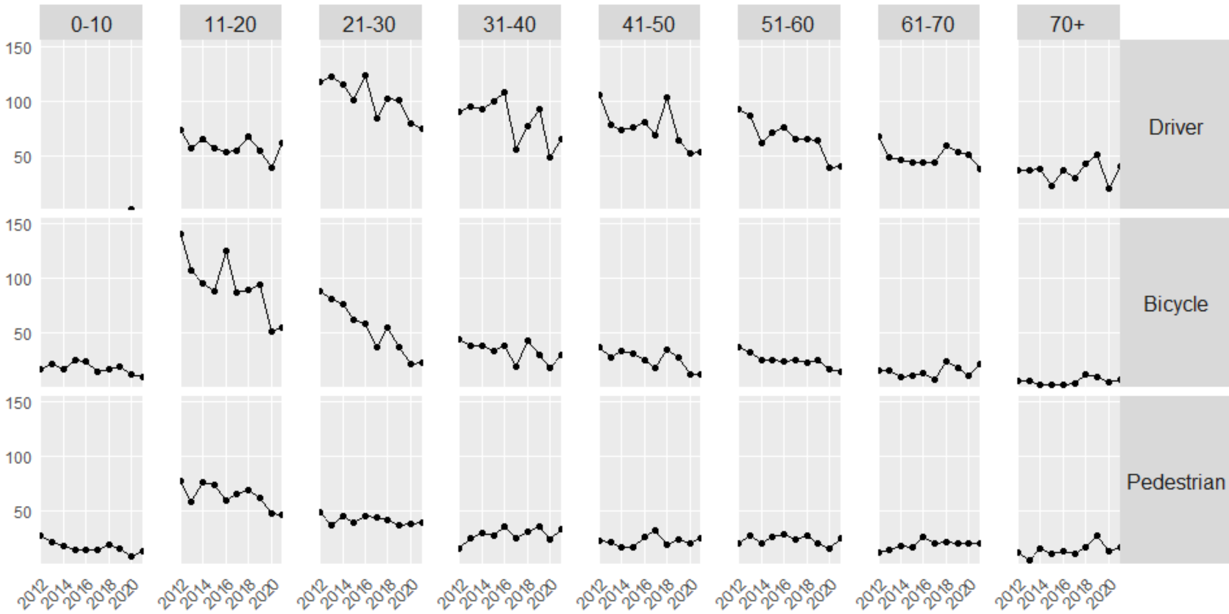


Figure 5. Annual number of crashes involving road users of various age ranges.

The research team further investigated the concept of age by identifying crashes that may be related to traveling to school. School addresses and grades served were collected from the Idaho State Department of Education. Each grade corresponded to an age range, with Kindergarten at 5 $\pm$ 1 years, first grade at 6 $\pm$ 1 years, etc. Excluding weekend crashes, each crash was then matched to its closest age-appropriate school. Crashes occurring within a one-mile radius of an age-appropriate school were considered related to school travel. For example, a crash involving a 9-year-old child 0.5 miles from an elementary school would be considered school-travel-related, but a child of the same age and distance from a high school would not. Figure 6 shows the annual number of bicycle and pedestrian crashes related to school travel. Over the 10-year period, these crashes trend downward. However, an upward trend (with a COVID-induced downward shift in 2020) can be seen starting in 2017.

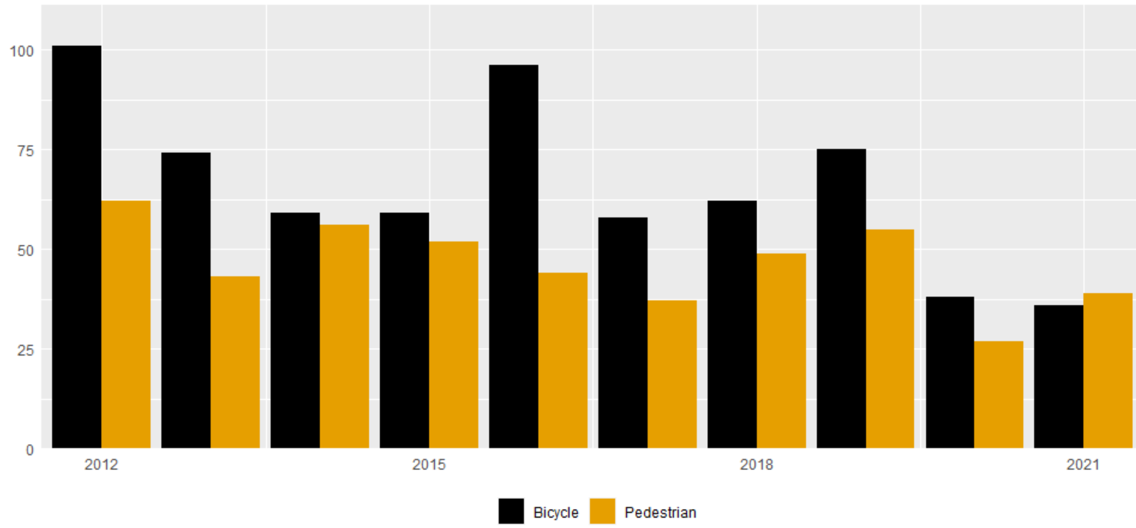


Figure 6. Annual bicycle and pedestrian crashes (2012-2021) related to school travel.

Figure 7 shows crash urbanicity. Note that the "rural/urban" data element included in crash records was often missing, prompting the use of a related variable indicating whether or not the crash occurred within city limits; those that did were considered urban while those that did not were considered rural. The vast majority (93%) of bicycle and pedestrian crashes occurred in urban areas. Annual bicycle crashes exhibit a decline while pedestrian crashes hold relatively flat. Crashes involving both road user types fell significantly in 2020 and began to rise again the following year.

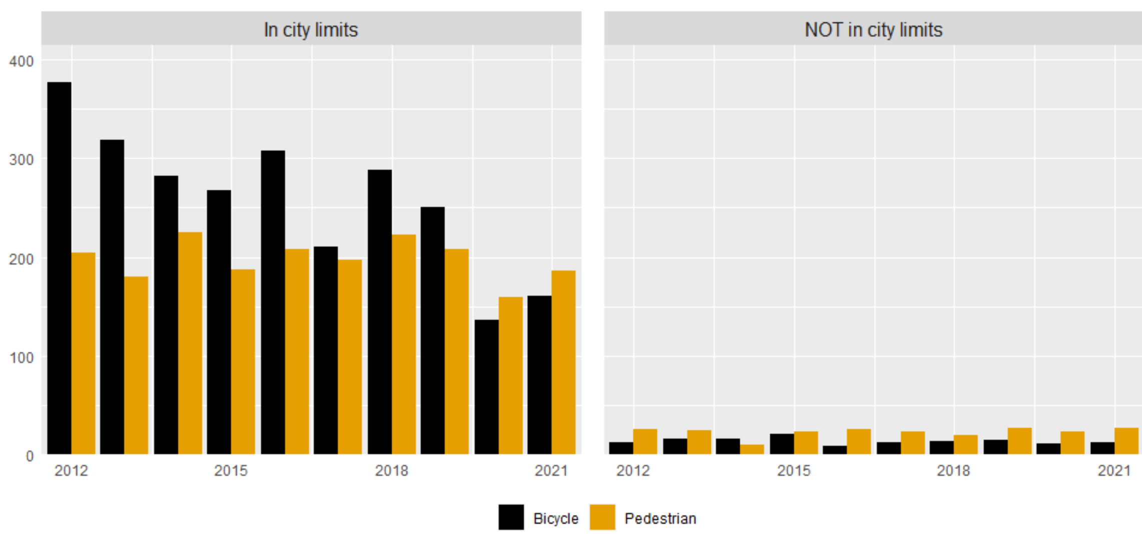


Figure 7. Annual bicycle and pedestrian crashes (2012-2021) in urban and rural areas.

Figure 8 shows how post-crash care differs in urban and rural areas. Over the 10-year period, average response time (the time between notification and arrival) in rural areas (M = 10.7 minutes, SD = 5.7, N =

281 crashes) exceeded that of urban areas (M = 6.3 minutes, SD =4.0, N = 3,346 crashes). Response time exhibits a slight upward trend in both urban and rural areas.

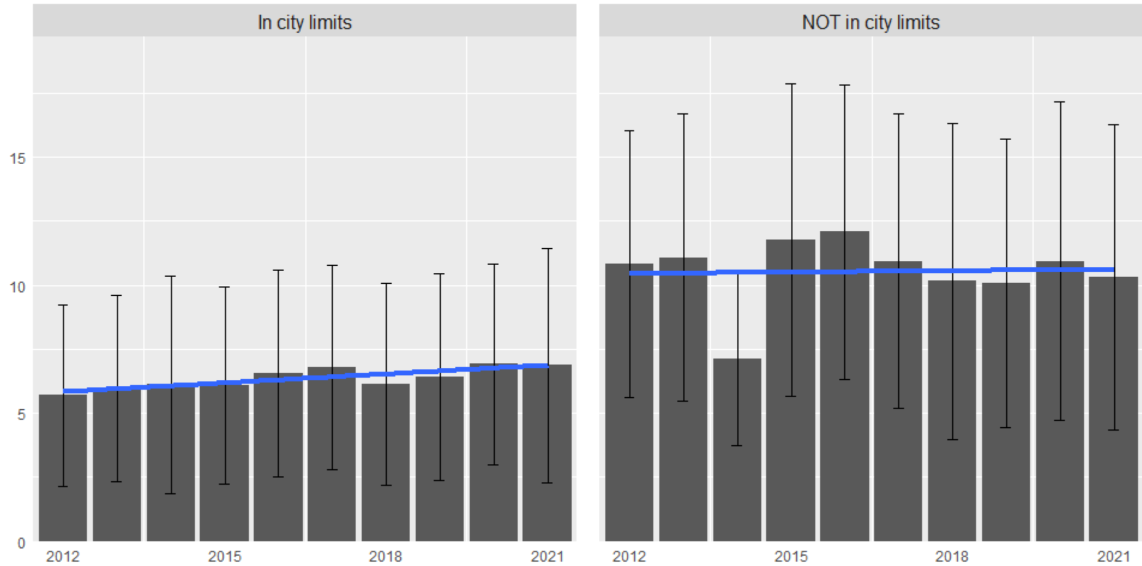


Figure 8. Annual mean (± 1 SD) ambulance response time (minutes) in urban and rural areas.

Helmet use also differs in urban and rural areas. Figure 9 shows annual helmet use among bicyclists involved in crashes in urban and rural areas. Helmet use is lower within city limits (approximately 25%) but increasing, whereas helmet use in rural areas shows a sharp decline.

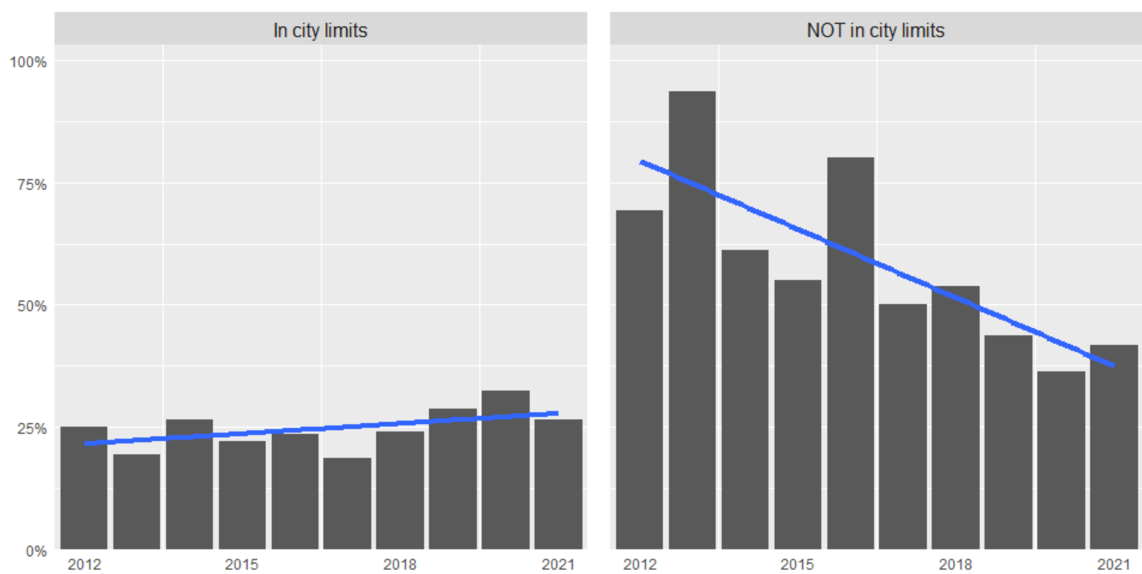


Figure 9. Annual helmet use (%) among bicyclists involved in crashes in urban and rural areas.

Figure 10 describes the lighting conditions. Daytime is the most common setting for both bicycle and pedestrian crashes. Both are similarly represented in dawn/dusk conditions as well, but pedestrian crashes largely outnumber bicycle crashes in the dark, both with and without street lights activated.

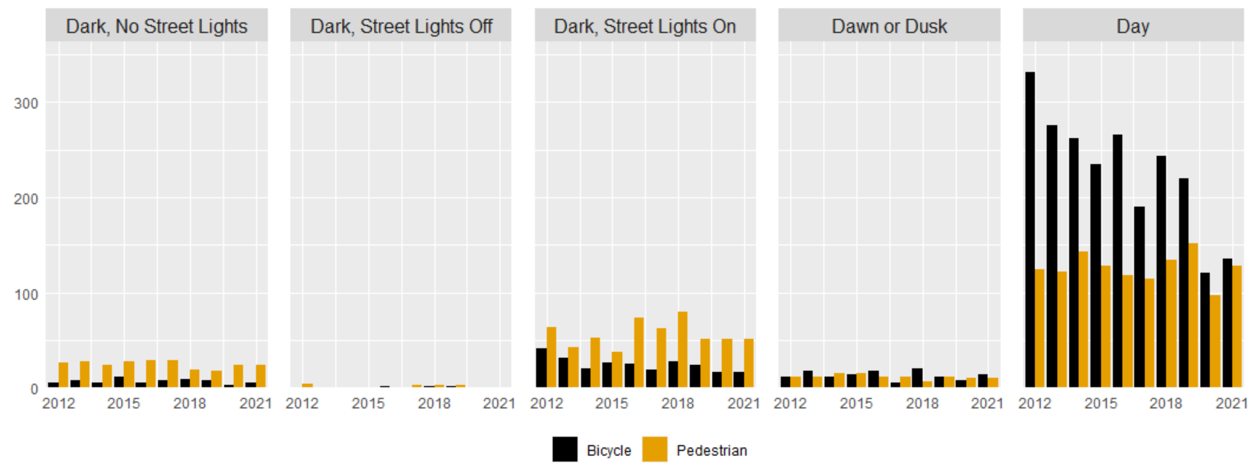


Figure 10. Annual bicycle and pedestrian crashes by lighting condition.

Figure 11 shows annual bicycle and pedestrian crashes by the traffic control device(s) present at the time of the crash. Notably, intersections with stop signs on the cross streets only represent a safety challenge for bicyclists. Throughout the 10-year period, bicycle crashes at these intersections far outnumber similar pedestrian crashes.

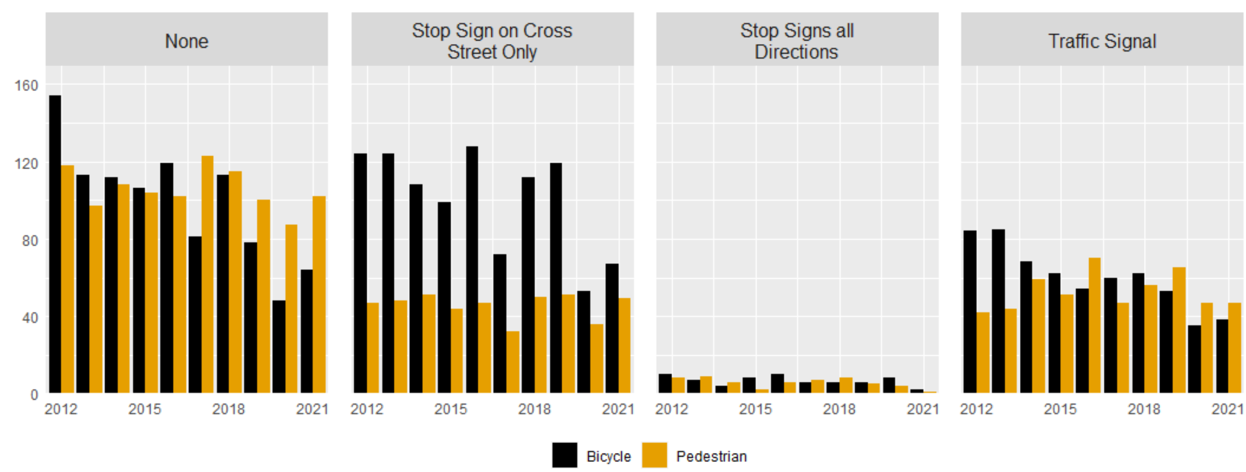


Figure 11. Annual bicycle and pedestrian crashes by traffic control device(s) present.

Figure 12 shows the types of motorized vehicles often involved in crashes with bicycles and pedestrians. Trends are similar for cars, pickup trucks, SUVs and vans, with a notable increase in crashes involving pickups in 2021.

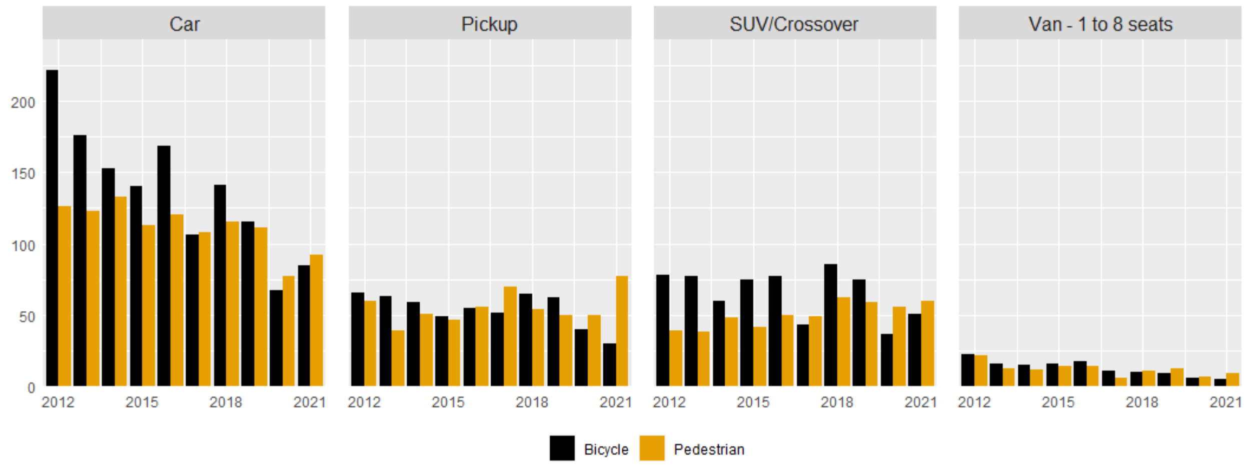


Figure 12. Annual bicycle and pedestrian crashes by type of motor vehicle involved.

Bicycle Crash Types

Figure 13 shows annual bicycle crashes by PBCAT2 crash type. All 14 crash types exhibit a downward trend. Crashes involving a loss of control or turning error are the most common over the ten-year period and continue to be so in 2021 (PBCAT2 110: N = 1,046, 24%). Nonroadway crashes (those occurring in parking lots, driveways, etc.) are the second most common (910: 631, 15%). Crashes involving a failure to yield at stop signs are the next most common type, with motorists (140: 497, 12%) and bicyclists (145: 482, 11%) roughly equally at fault. Midblock crashes are next, with motorists (320: 482, 11%) failing to yield more often than bicyclists (310: 337, 8%); followed by bicyclists (158: 298, 7%) and motorists (150: 278, 6%) failing to yield at signalized intersections. The remaining crash types account for less than 5% of bicycle crashes.

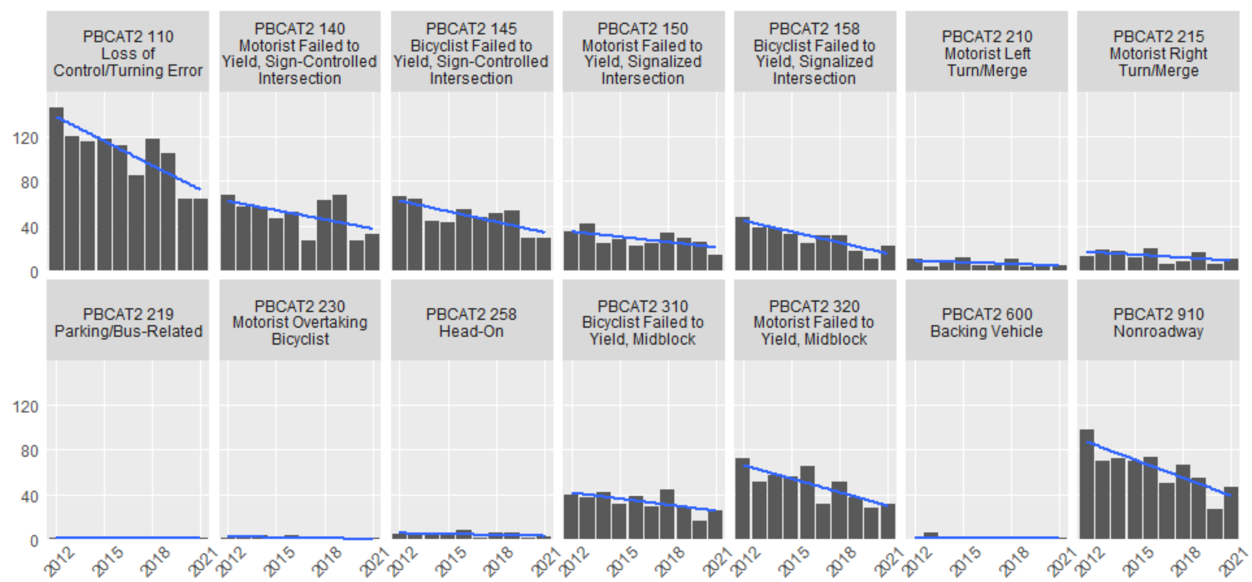


Figure 13. Annual bicycle crashes by PBCAT2 crash type.

The PBCAT2 methodology can apply multiple crash types to a single crash. As implemented here, 1,025 (37%) of bicycle crashes were characterized by a single PBCAT2 crash type, 886 (32%) were characterized by two, and 470 (17%) were characterized by three or more. The remaining 358 (13%) did not align to any PBCAT2 crash type using the available data.

Figure 14 shows annual bicycle crash types identified via clustering. As with PBCAT2 bicycle crash types, all of these crash types exhibit a visible downward trend. There is significant overlap between the both methods' crash types, but the clustering approach identified three novel types. Clusters 3 – 9 are similar to PBCAT2 bicycle crash types 158, 600, 140, 150, 145, 258, and 310, respectively. Three cluster crash types do not closely correspond to any PBCAT2 types: Cluster 1 includes sideswipes at intersections, Cluster 2 includes crashes involving drivers speeding and fleeing the scene, and Cluster 10 includes motorist errors and impairments such as improperly changing lanes or overtaking other road users, and driving while drowsy or under the influence of alcohol.

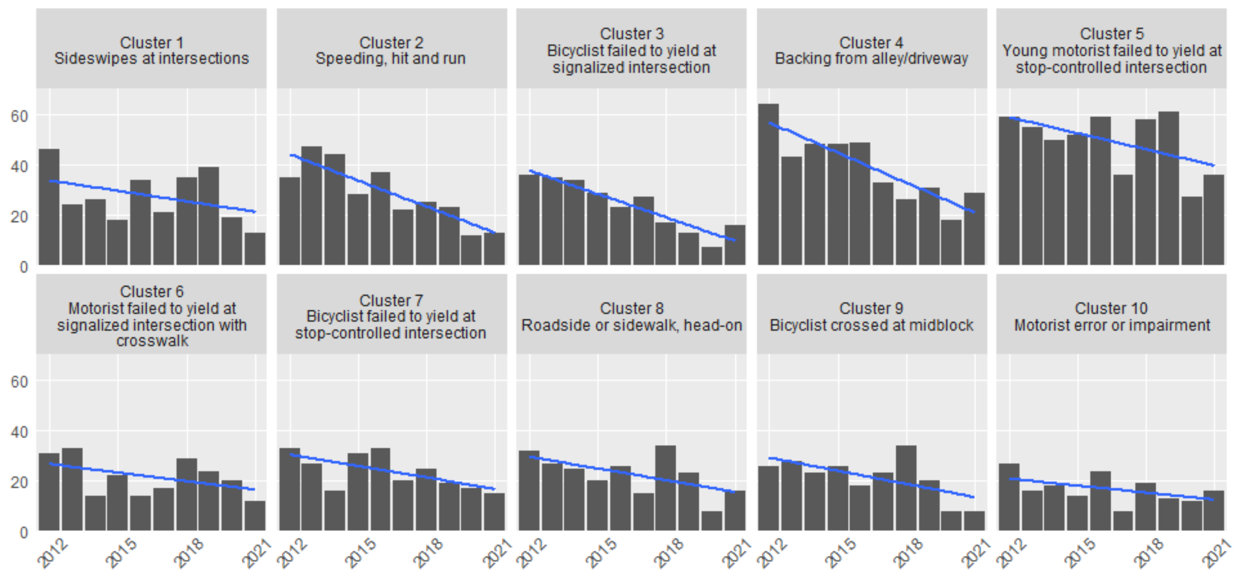


Figure 14. Annual bicycle crashes by clustering crash type.

Unlike PBCAT2, the clustering technique assigns one crash type to each crash. Cluster sizes are also more uniform in size, ranging from 161 crashes (Cluster 10) to 493 (Cluster 5) over the 10-year period.

Figure 15 shows the prevalence of injury severity among bicycle crash types (determined via clustering). Clusters 2, 9, and 10 have the highest portion of fatal injuries; clusters 9 and 10 also have the highest portion of suspected serious injuries. These crashes involve speeding, midblock crossings, and motorist impairment – key risk factors for fatal and severe bicycle crashes.

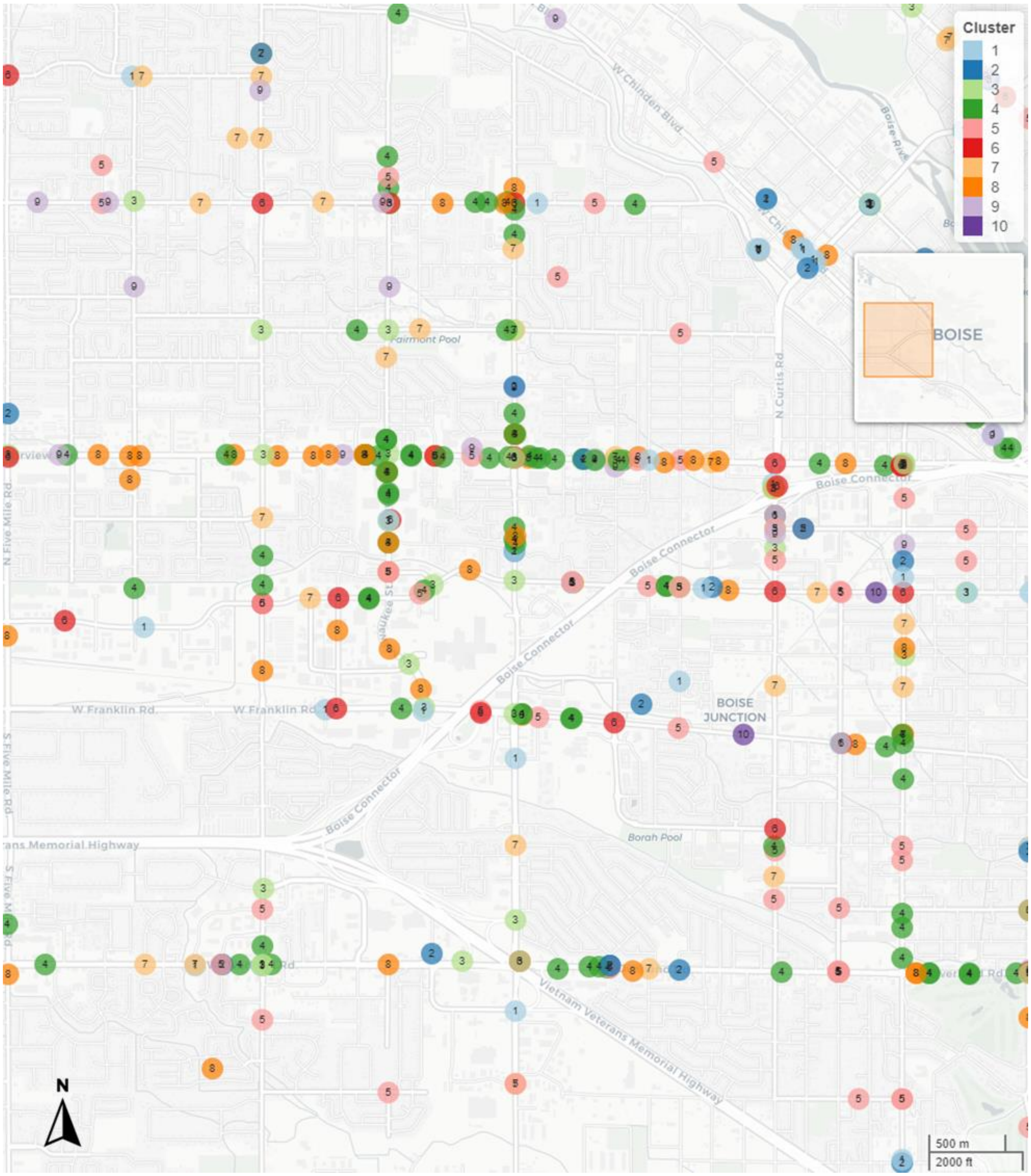
Figure 15. Prevalence of injury severity levels in bicycle crash types.

The maps on the following pages represent static snapshots of the georeferenced database provided by the research team to ITD. Each of these cities (Boise, Coeur d'Alene, Nampa, Meridian, Idaho Falls, and Pocatello) reported at least 100 bicycle crashes during the 10-year period, representing 68% of all bicycle crashes during the same period. The maps can be used to identify locations and corridors with high incidences of particular crash types. Table 8 lists High occurrence corridors for bicycle crashes in each location. Crashes of various types tend to occur along these corridors. Implementing countermeasures along these corridors could eliminate a substantial portion of bicycle crashes in Idaho.

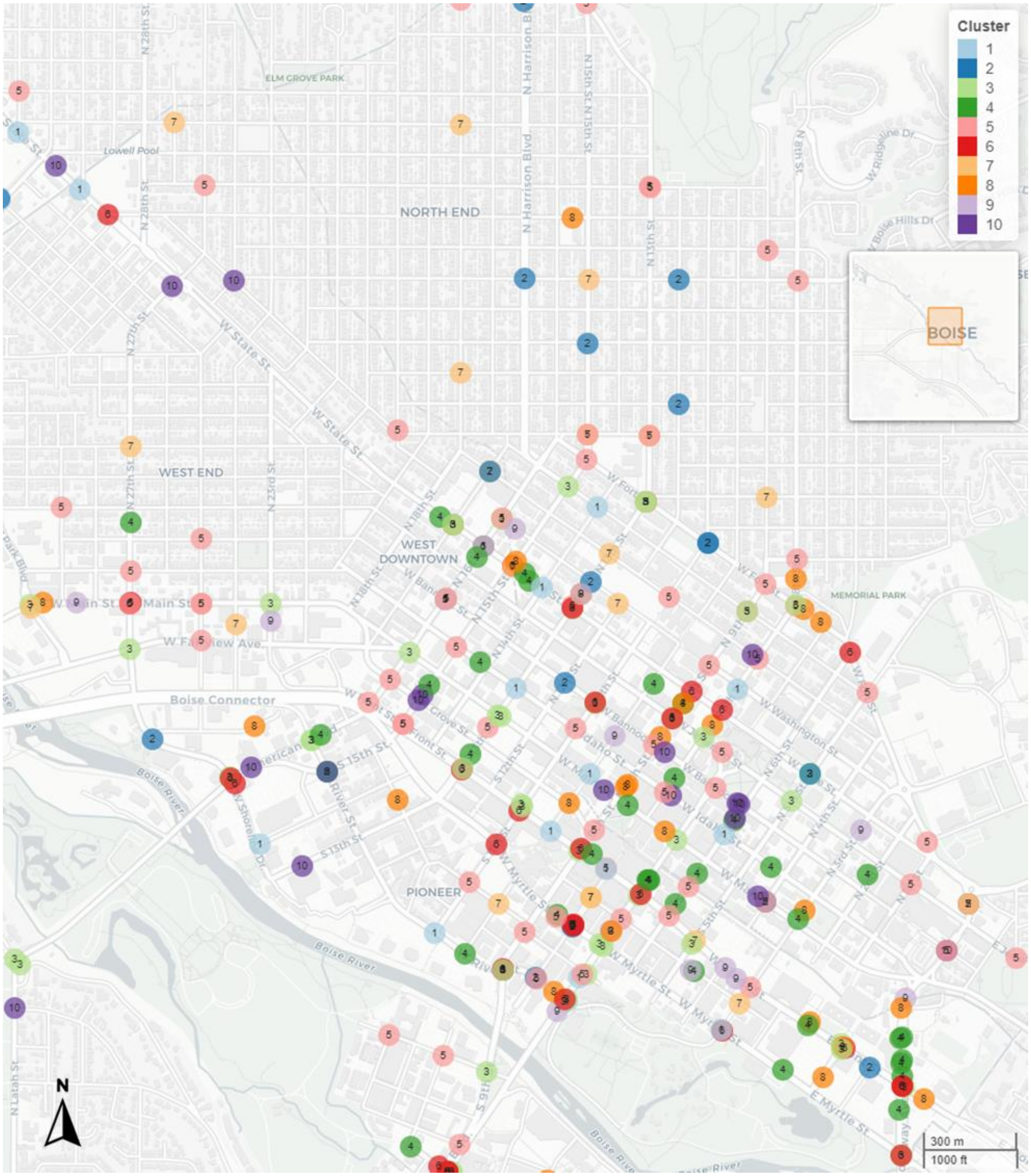
Table 8. Highest occurrence corridors for bicycle crashes.

Location	***Corridors***
Boise	West Fairview Avenue North Orchard Street South Vista Avenue West Overland Road South Broadway Road North 9th Street
Coeur d'Alene	West Appleway Avenue North Government Way
Nampa	12th Avenue Road* 12th Avenue South* Caldwell Boulevard*
Meridian	West Cherry Lane/Fairview Avenue North Eagle Road
Idaho Falls	West Broadway Street East Sunnyside Road East 17th Street*
Pocatello	Pole Line Road Yellowstone Avenue* 4th Avenue*

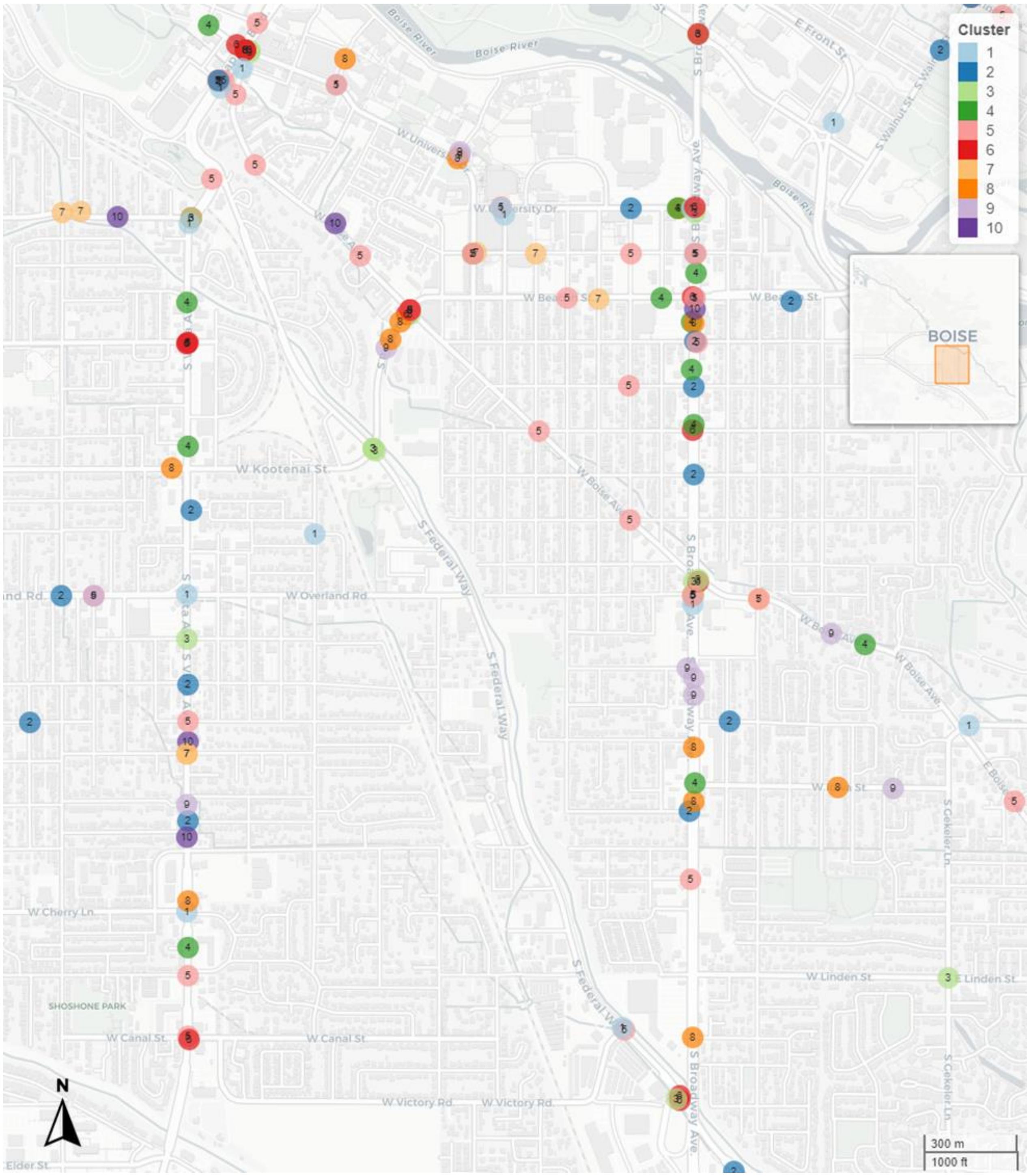
*Also a corridor with highest occurrence for pedestrian crashes



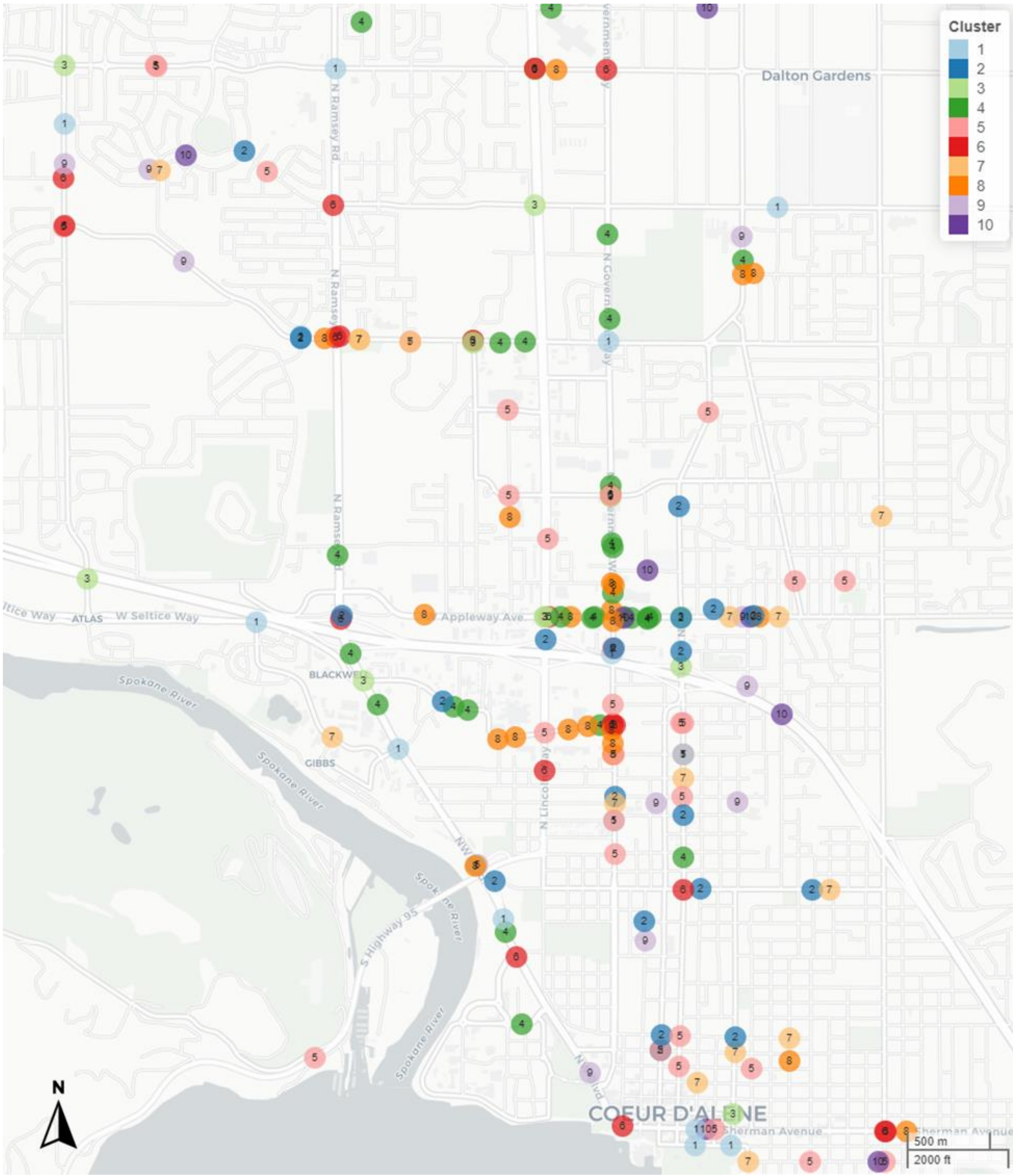
Map 1. Bicycle crashes in Boise (1 of 3).



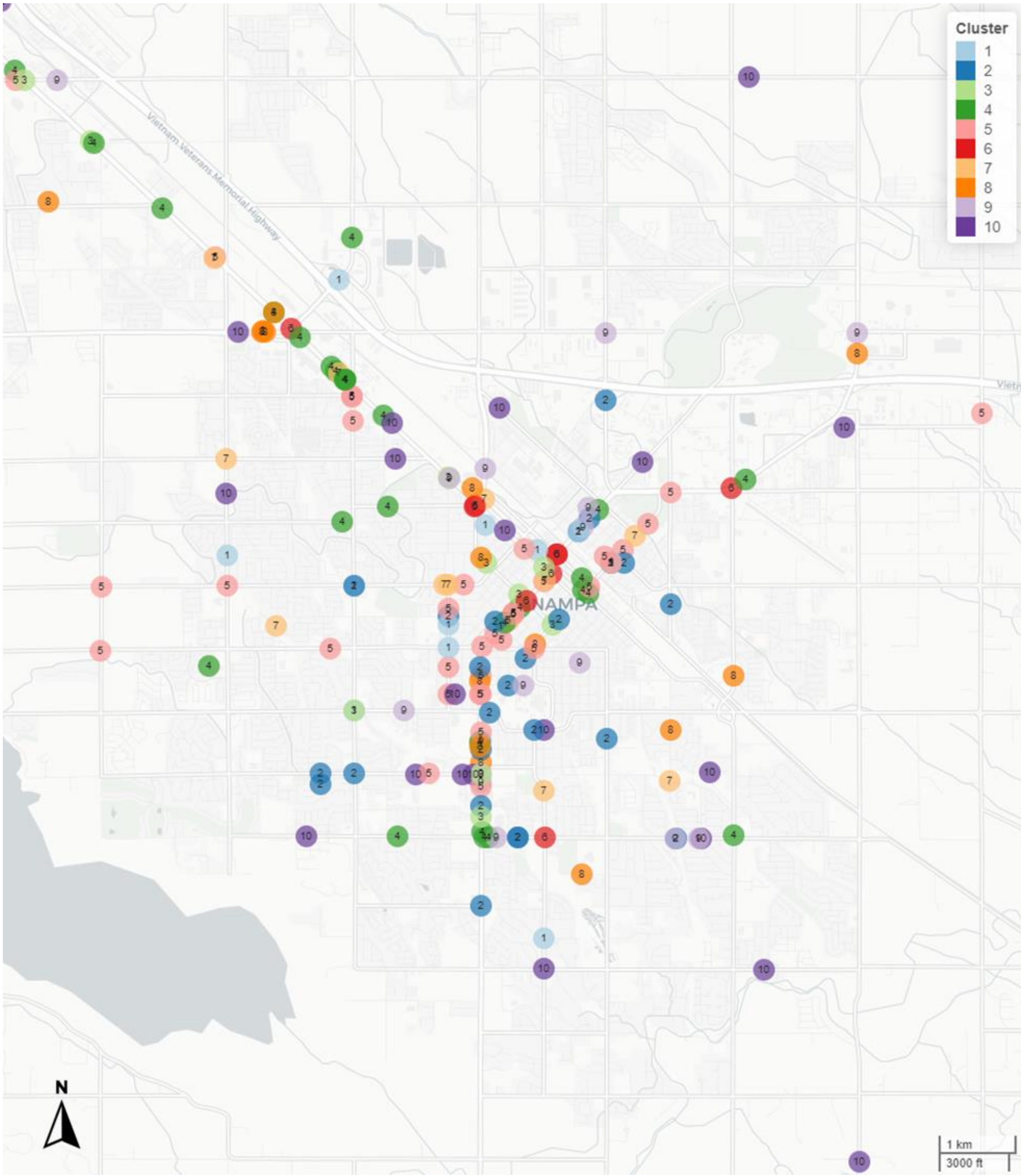
Map 2. Bicycle crashes in Boise (2 of 3).



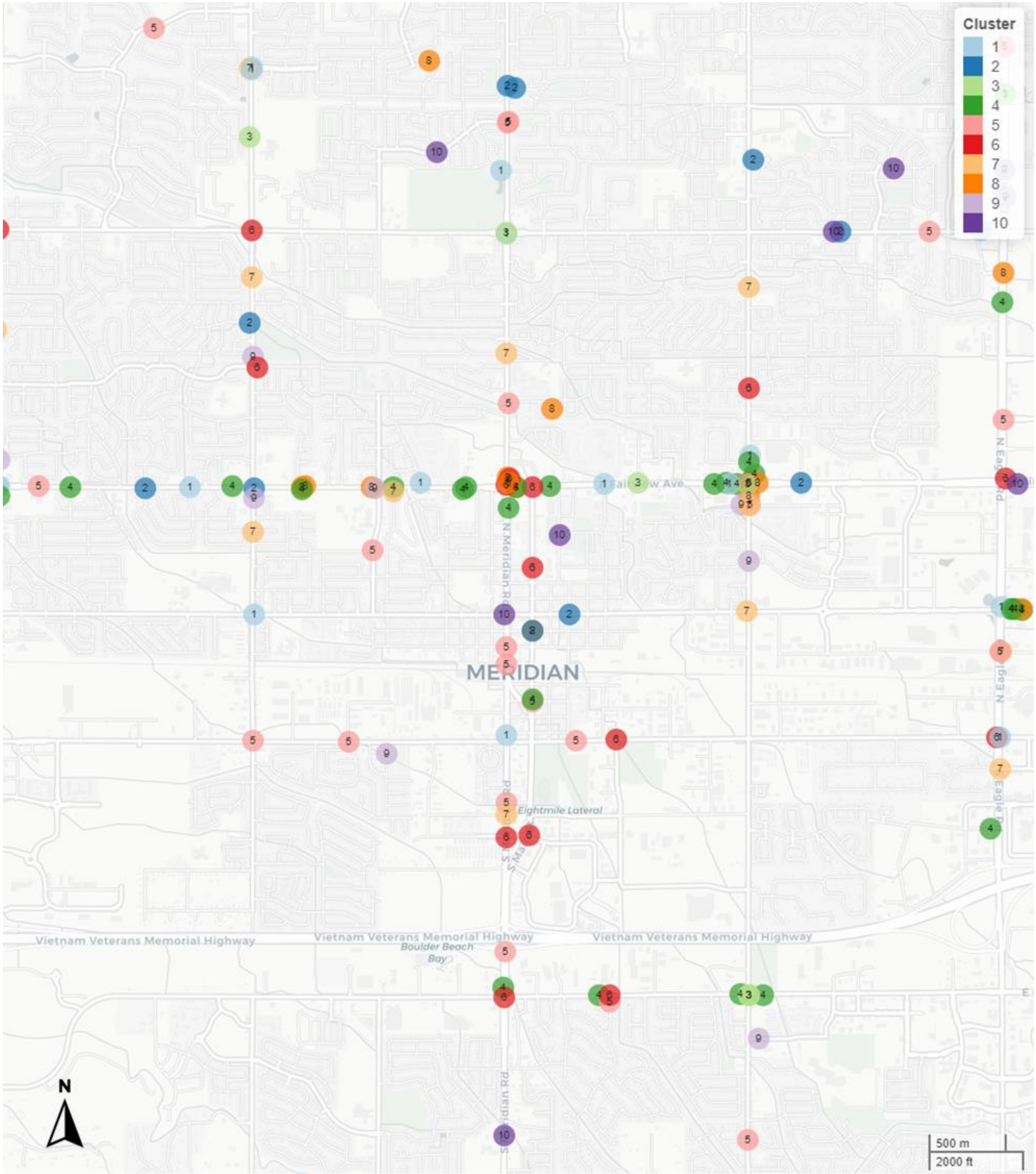
Map 3. Bicycle crashes in Boise (3 of 3).



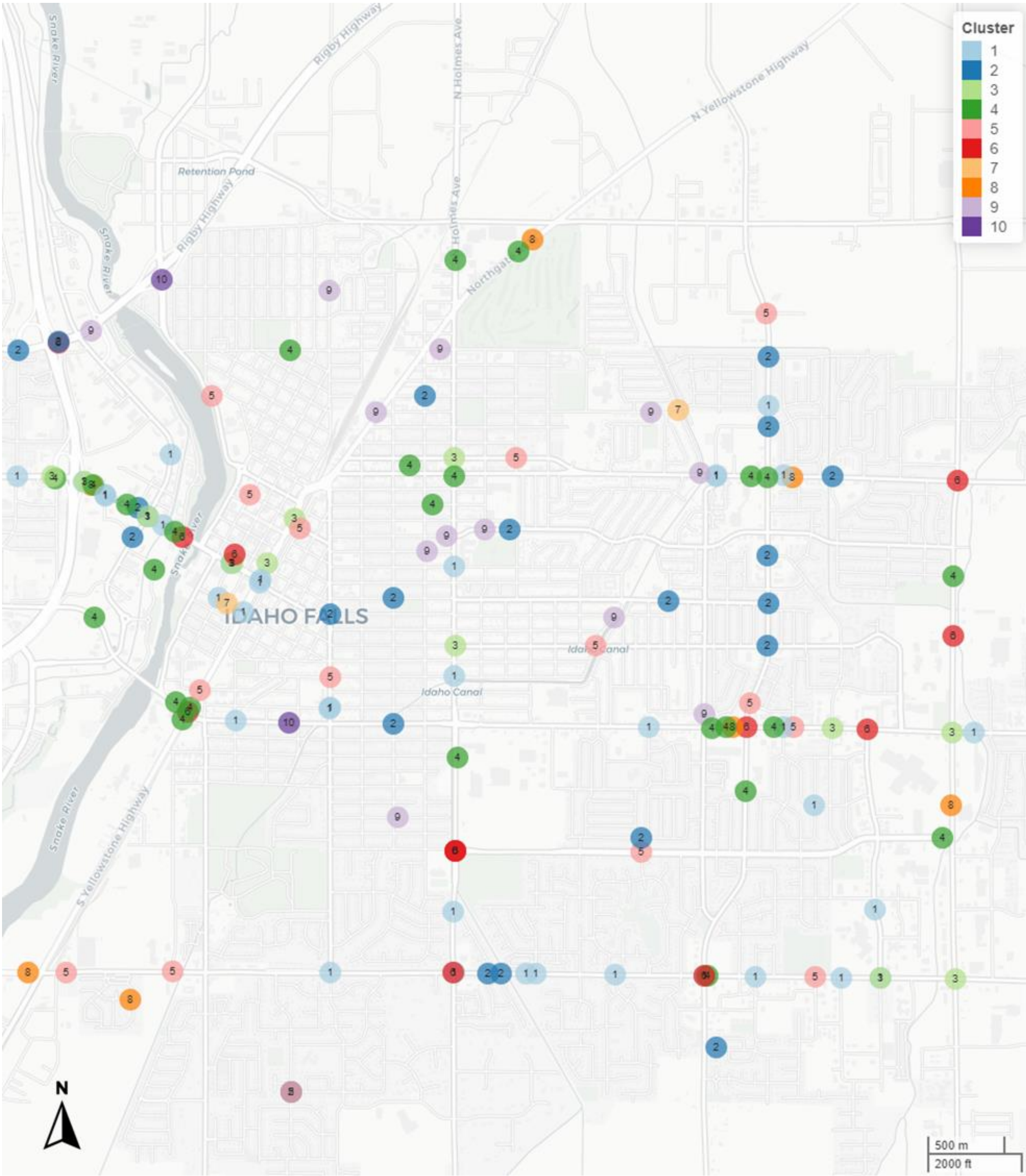
Map 4. Bicycle crashes in Coeur d'Alene.



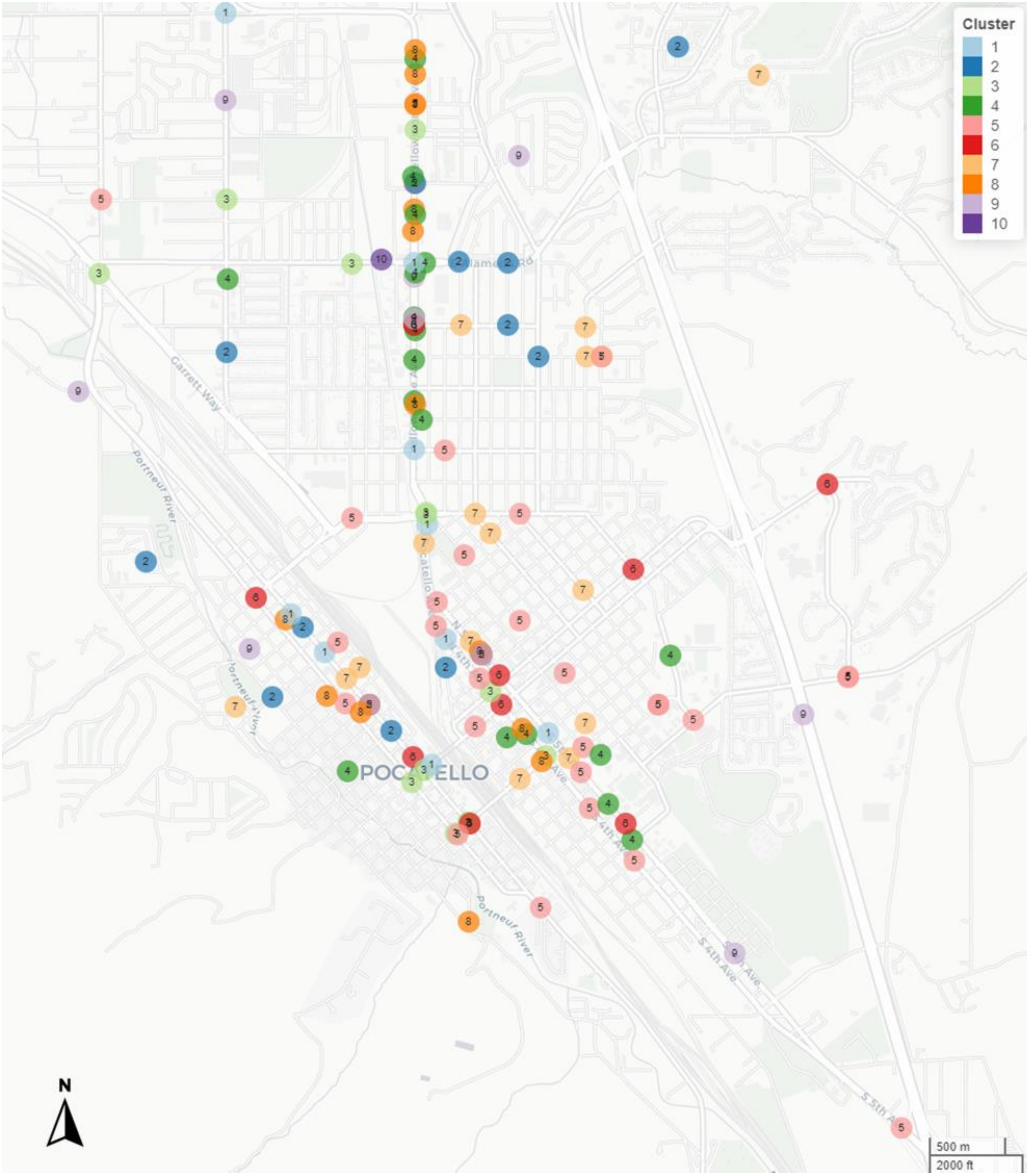
Map 5. Bicycle crashes in Nampa.



Map 6. Bicycle crashes in Meridian.



Map 7. Bicycle crashes in Idaho Falls.



Map 8. Bicycle crashes in Pocatello.

Pedestrian Crash Types

Figure 16 shows annual pedestrian crashes by PBCAT2 crash type. The two most common crash types involve pedestrians crossing a roadway, with those involving a vehicle going straight (PBCAT2 750: N = 766, 22%) decreasing over time and those involving a turning vehicle increasing (790: 637, 18%). The next three most common crash types involve midblock crossings (350: 424, 12%) and pedestrians walking (400: 411, 12%) or working/playing (310: 298, 8%) along the roadway. Crashes involving pedestrians crossing driveways or alleys (460: 260, 7%) have increased over time, while those involving dash/dart-outs (740: 251, 7%) or waiting to cross (500: 224, 6%) have decreased. Multiple threat crashes (involving a pedestrian getting trapped on a median while crossing) have also increased (720: 93, 3%). The remaining crash types account for less than 5% of pedestrian crashes.

Five crash types show increasing trends to various degrees. Crash types 100 (unusual circumstances), 340 (bus-related) and 720 (multiple threat/trapped) are increasing but represent a small fraction of all pedestrian crashes. On the contrary, crash types 460 (crossing driveway or alley) and 790 (crossing roadway, vehicle turning) are more common and increasing. Notably, both of these crash types involve a pedestrian crossing the path of a motor vehicle when the driver's vision may be blocked or focused elsewhere.

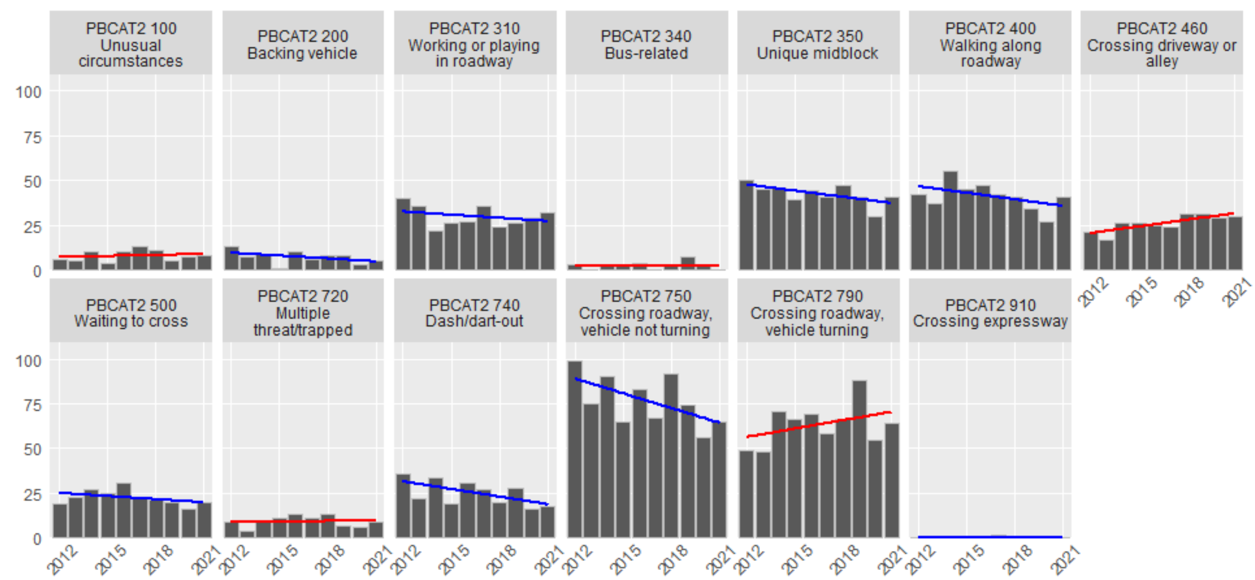


Figure 16. Annual pedestrian crashes by PBCAT2 crash type.

As with bicycles, the PBCAT2 methodology applied multiple crash types to many pedestrian crashes: 991 (45%) of pedestrian crashes were characterized by a single PBCAT2 crash type, 739 (33%) were characterized by two, and 332 (15%) were characterized by three or more. The remaining 151 (7%) did not align to any PBCAT2 crash type using the available data.

Figure 17 shows annual pedestrian crashes by clustering crash type. Some of these crash types are similar or related to PBCAT2 crash types. Clusters 1 and 9 involve different causes (different road users failing to obey a traffic control device) but may correspond to PBCAT2 750 (crossing roadway, vehicle not turning). Clusters 3 and 4 involve a motorist making an improper turn and may correspond to PBCAT2 790 (crossing roadway, vehicle turning). Cluster 5 involves midblock crossings and may correspond to PBCAT2 350 (unique midblock) or 740 (dash/dart out). Clusters 2 (crossing parking lot, alley, driveway), 6 (backing), and 7 (pedestrian walking along roadway) are directly correlated with PBCAT2 types 460, 200, and 400, respectively. Two novel crash types emerged from the clustering process: cluster 8 involves motorists speeding near horizontal or vertical curves, and cluster 10 includes hit-and-run pedestrian crashes.

Five crash types exhibit increasing trends. Cluster 4 crashes (improper left turn) are the most numerous and exhibit a slight annual increase. Crashes in clusters 2 (parking lot, alley, driveway), 3 (failed to signal at intersection), and 8 (speeding near turns and hills) occur at similar rates, with those in cluster 2 increasing the fastest. Hit-and-run crashes (cluster 10) are the rarest but are also increasing. All other crash types show an overall decreasing trend over the ten-year period.

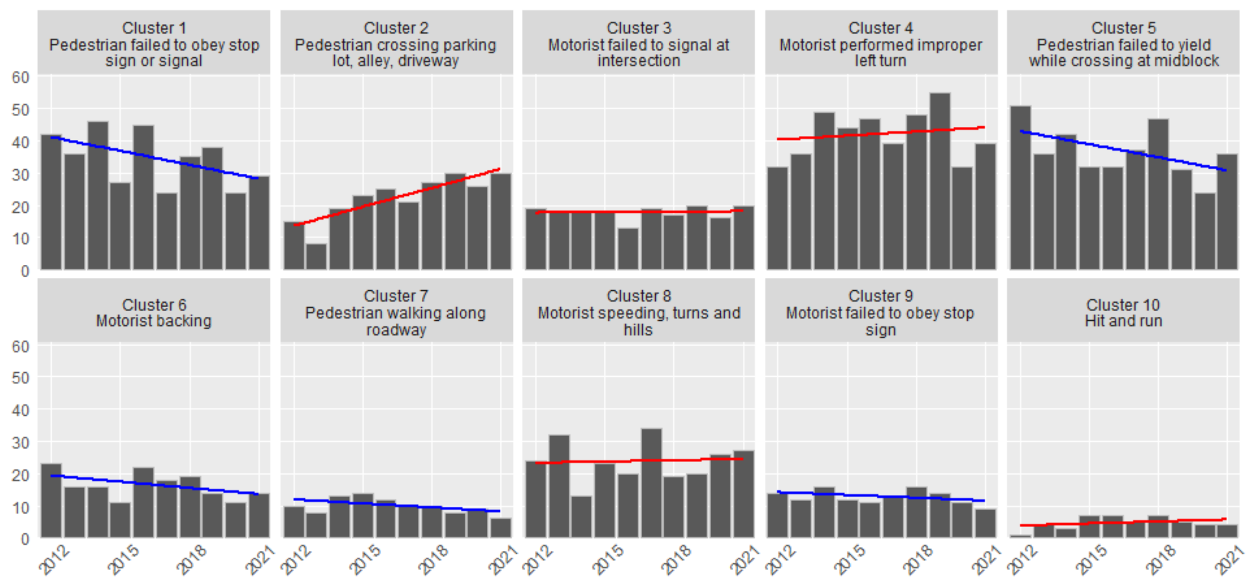


Figure 17. Annual pedestrian crashes by clustering crash type.

Figure shows the prevalence of injury severity among pedestrian crash types (determined via clustering). Clusters 7 and 8 have the highest proportion of fatal and suspected serious injuries. These crashes involve walking along the roadway and speeding around turns and hills. Clusters 1 and 5 also have high rates of suspected serious injuries; these crashes are also likely to involve higher speeds as they involve failing to yield to traffic control devices and midblock crossings.

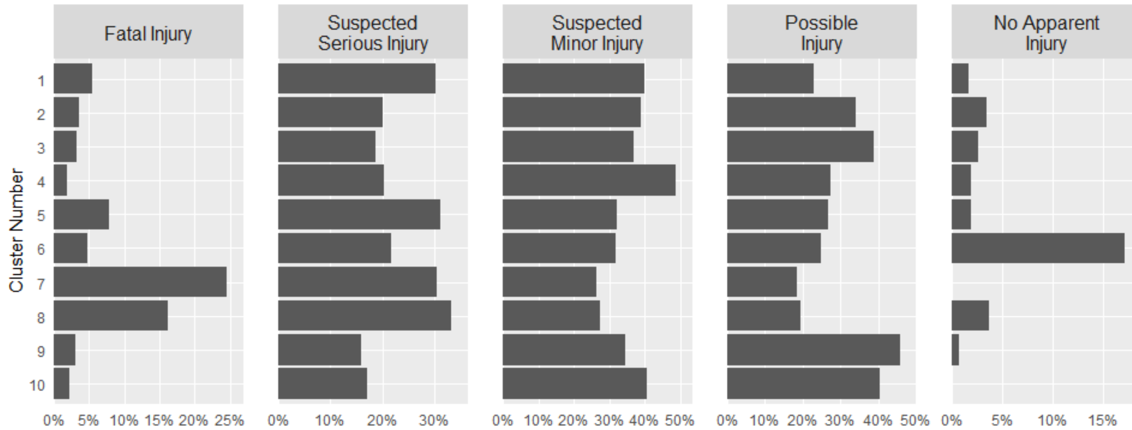


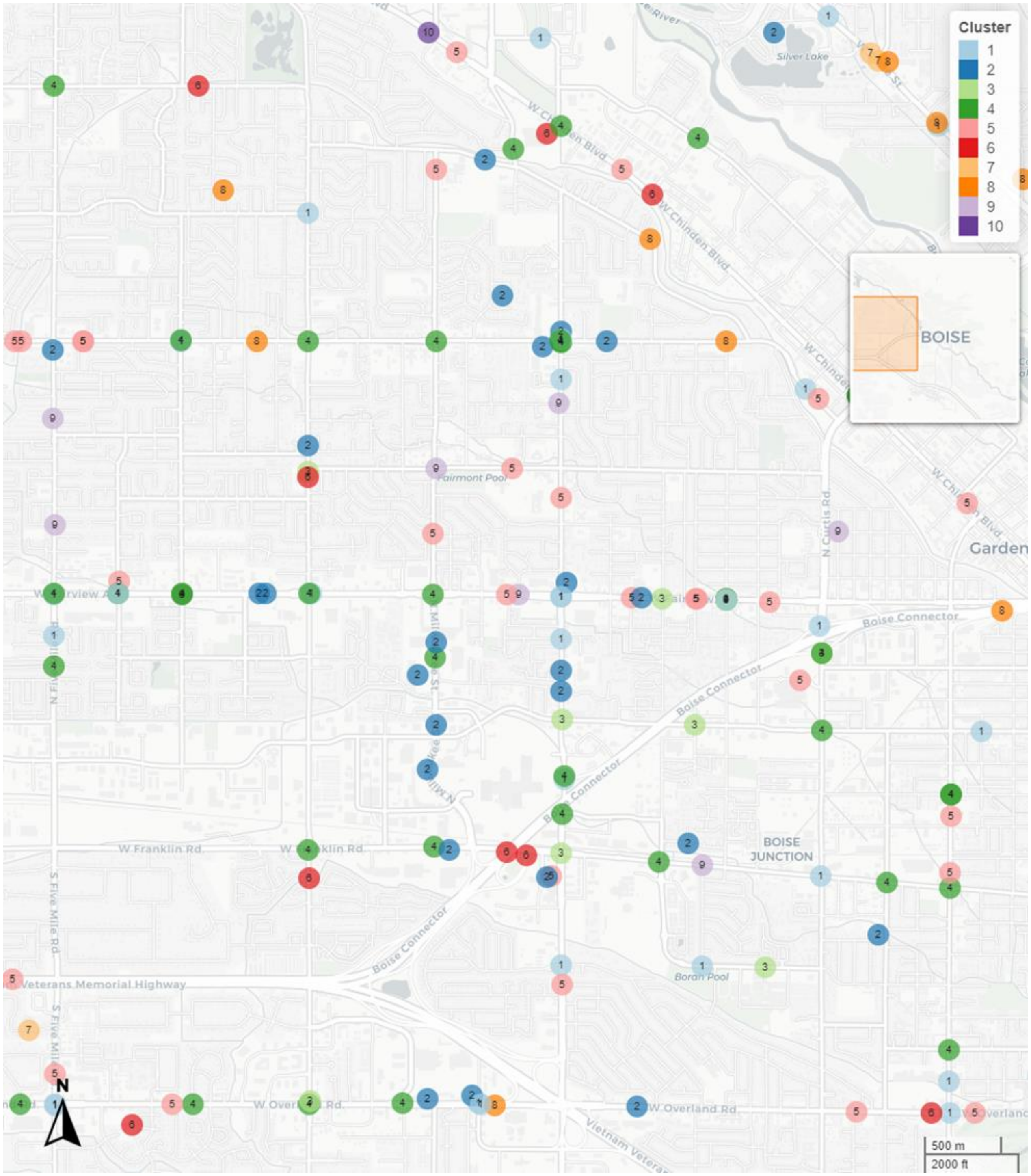
Figure 18. Prevalence of injury severity levels in pedestrian crash types.

The maps on the following pages represent static snapshots of the georeferenced database provided by the research team to ITD. Each of these cities (Boise, Nampa, Pocatello, Idaho Falls, Coeur d'Alene, Twin Falls, and Meridian) reported at least 100 pedestrian crashes during the 10-year period, representing 62% of all pedestrian crashes during the same period. The maps can be used to identify locations and corridors with high incidences of particular crash types. Table 9 lists high occurrence corridors for pedestrian crashes in each location. Implementing countermeasures along these corridors could eliminate a substantial proportion of pedestrian crashes in Idaho.

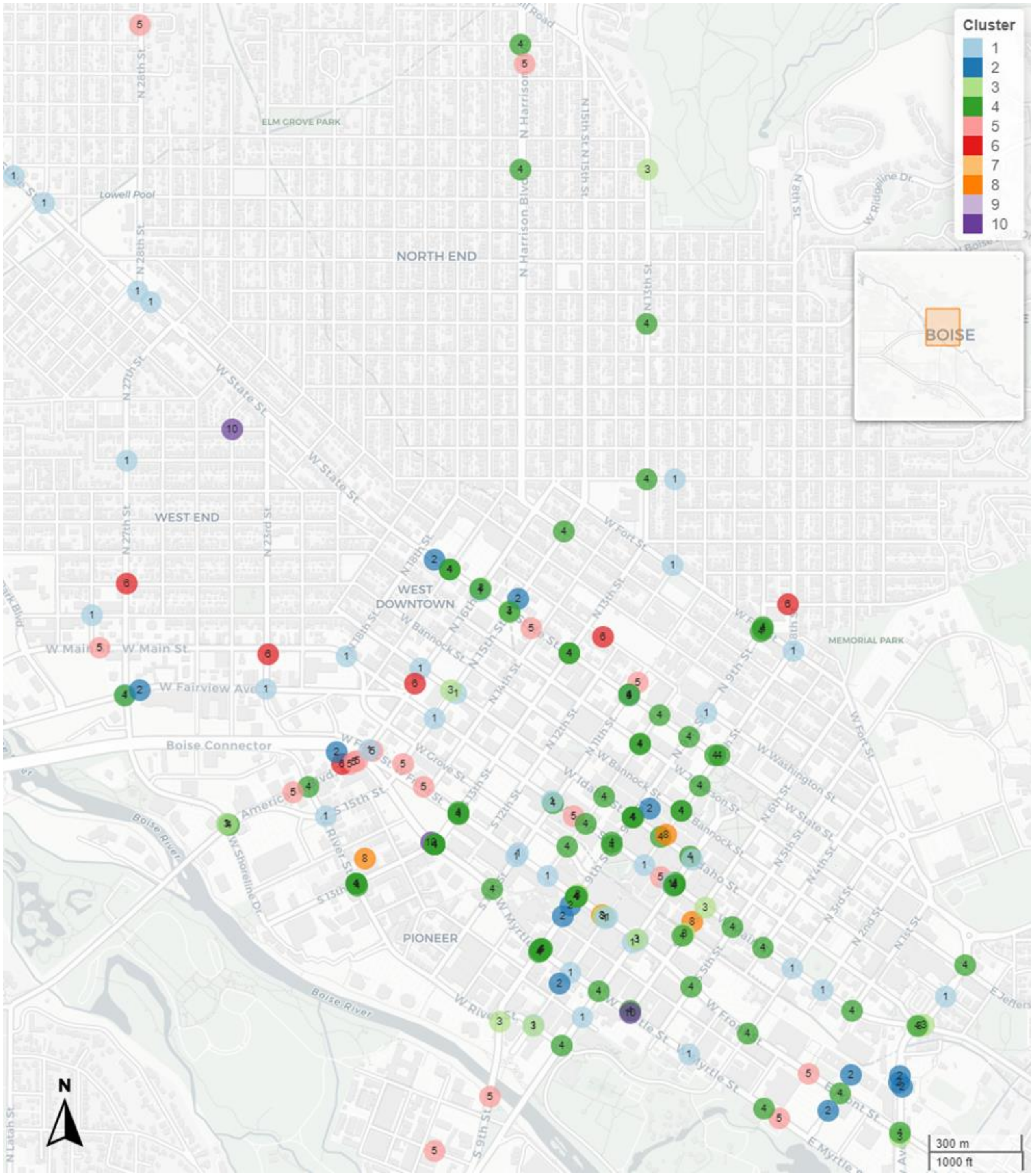
Table 9. Highest occurrence corridors for pedestrian crashes.

Location	***Corridors***
Boise	West Fairview Avenue North Cole Road West State Street West Main Street 9th Street
Nampa	12th Avenue Road* 12th Avenue South* Caldwell Boulevard*
Pocatello	North Arthur Avenue Yellowstone Avenue* 4th Avenue* 5th Avenue
Idaho Falls	East 17th Street* South Woodruff Avenue
Coeur d'Alene	Sherman Avenue
Twin Falls	Blue Lakes Boulevard North Addison Avenue Washington Street
Meridian	Meridian Road

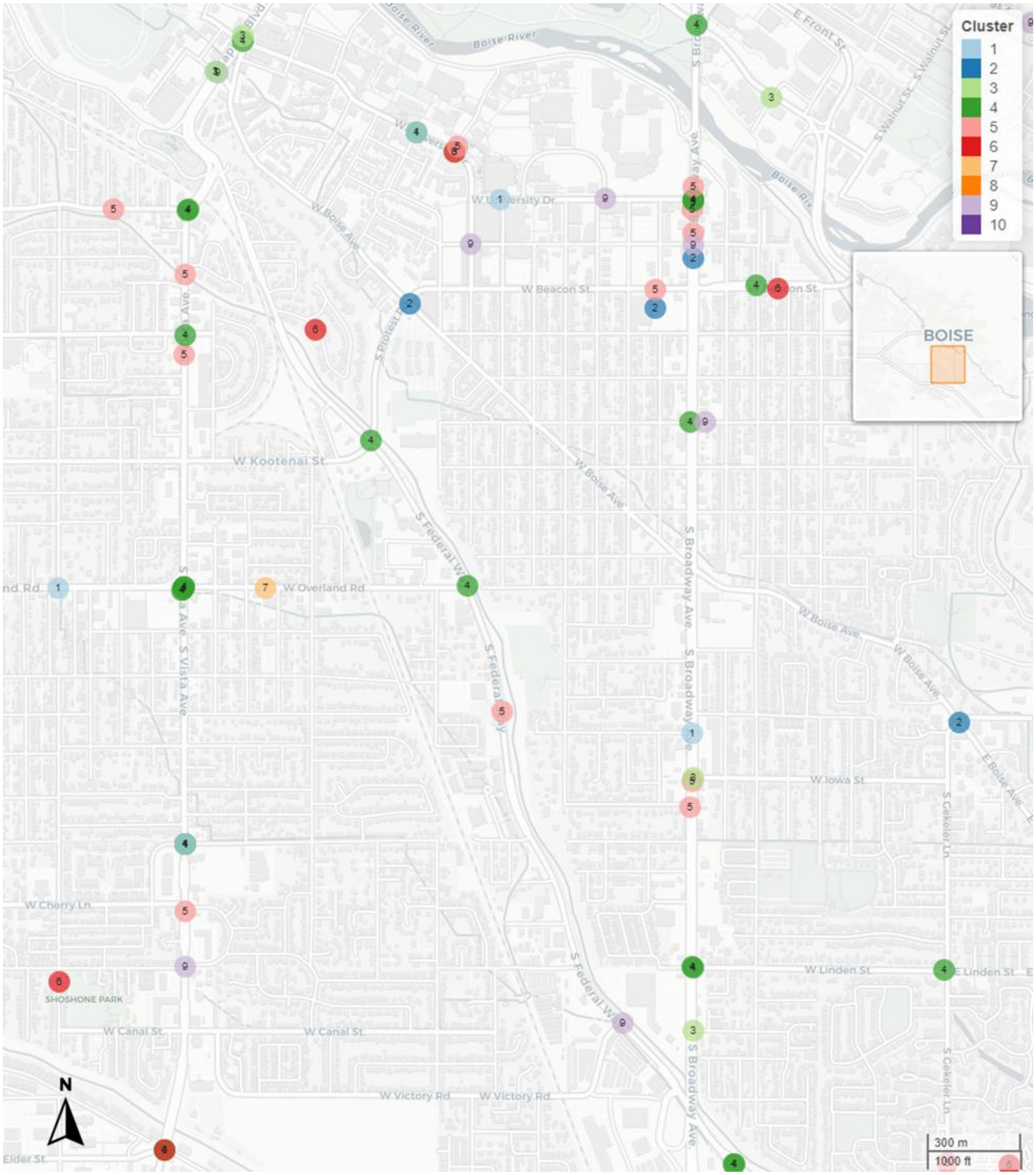
*Also a corridor with highest occurrence for bicycle crashes



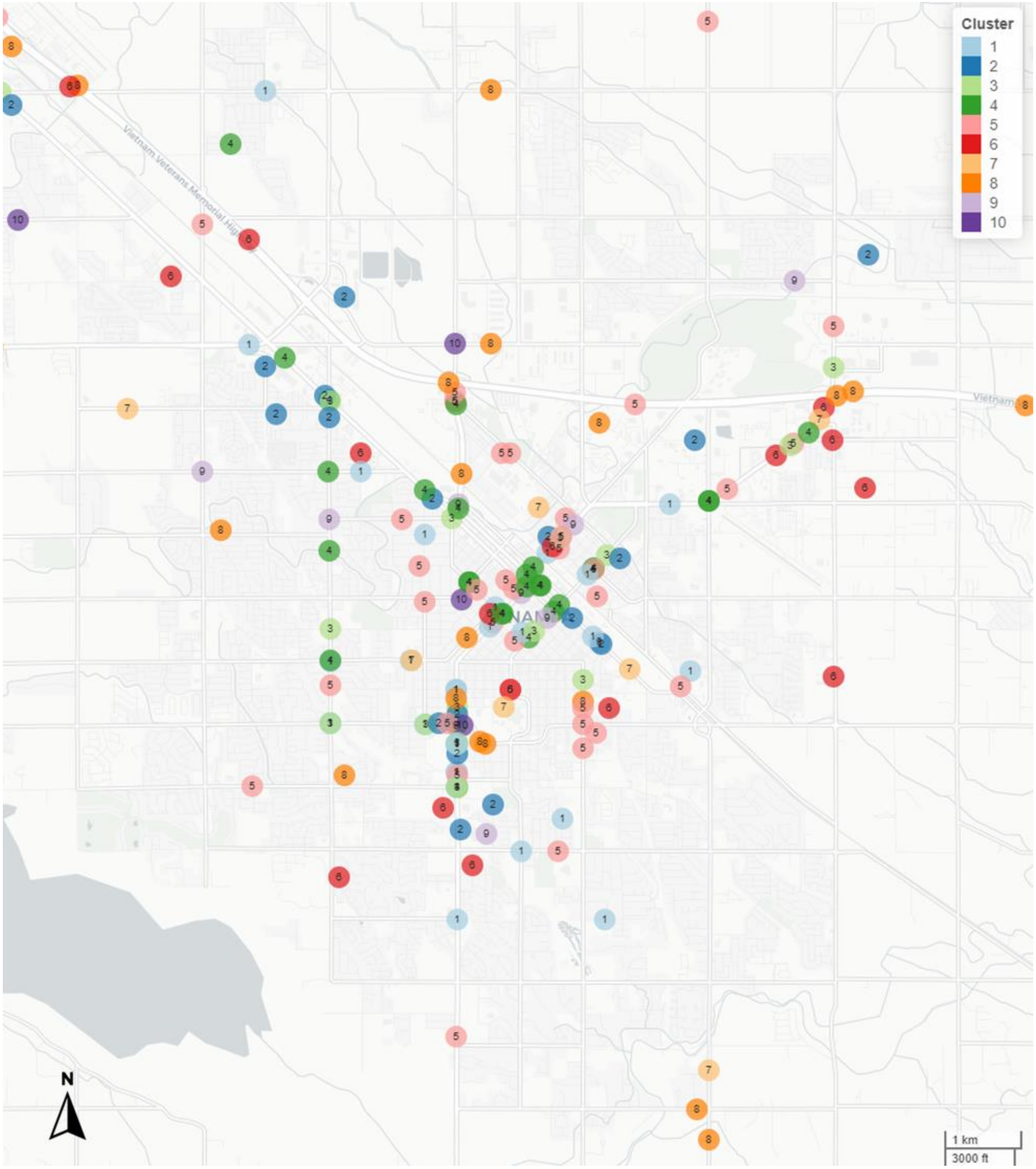
Map 9. Pedestrian crashes in Boise (1 of 3).



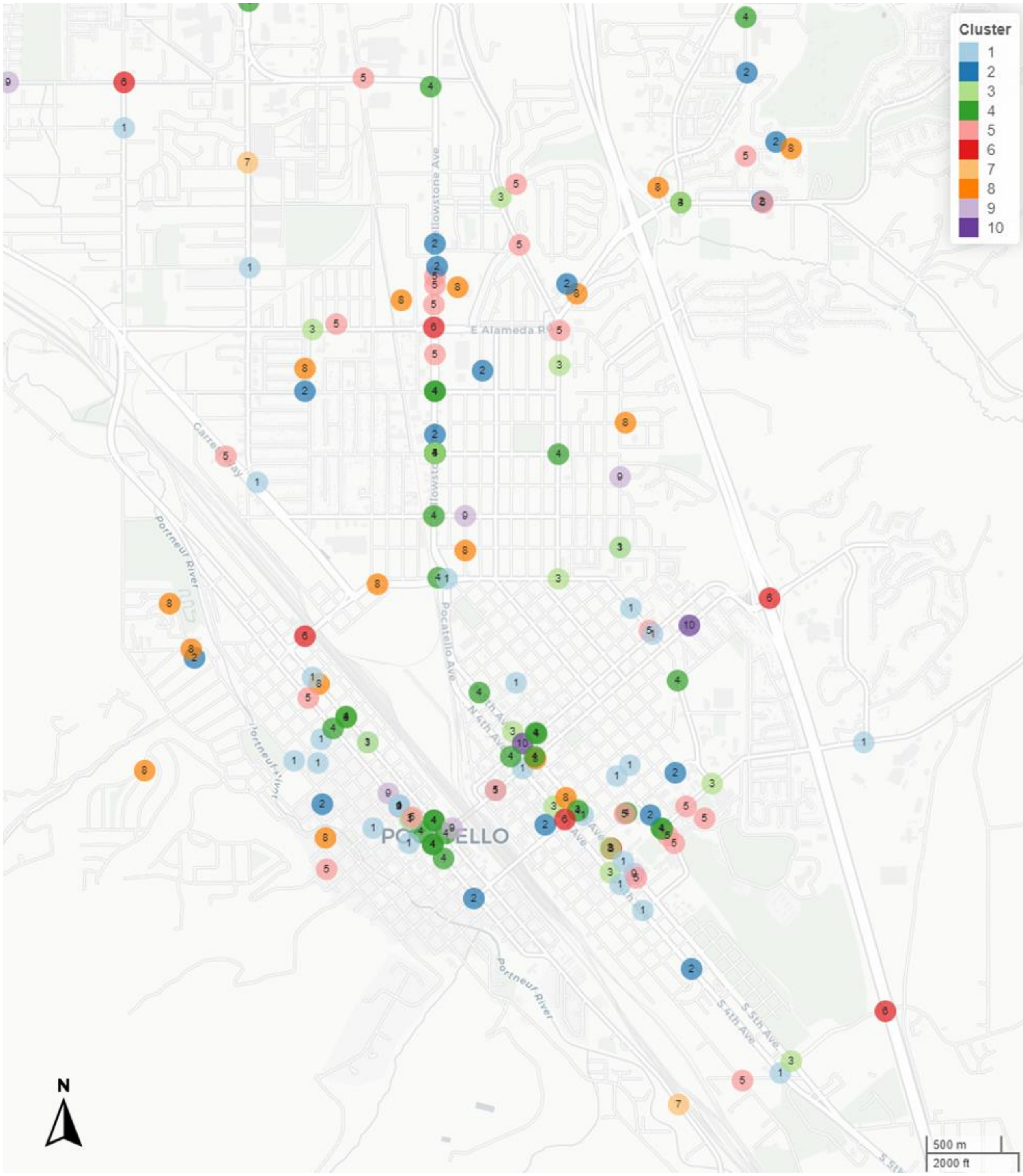
Map 10. Pedestrian crashes in Boise (2 of 3).



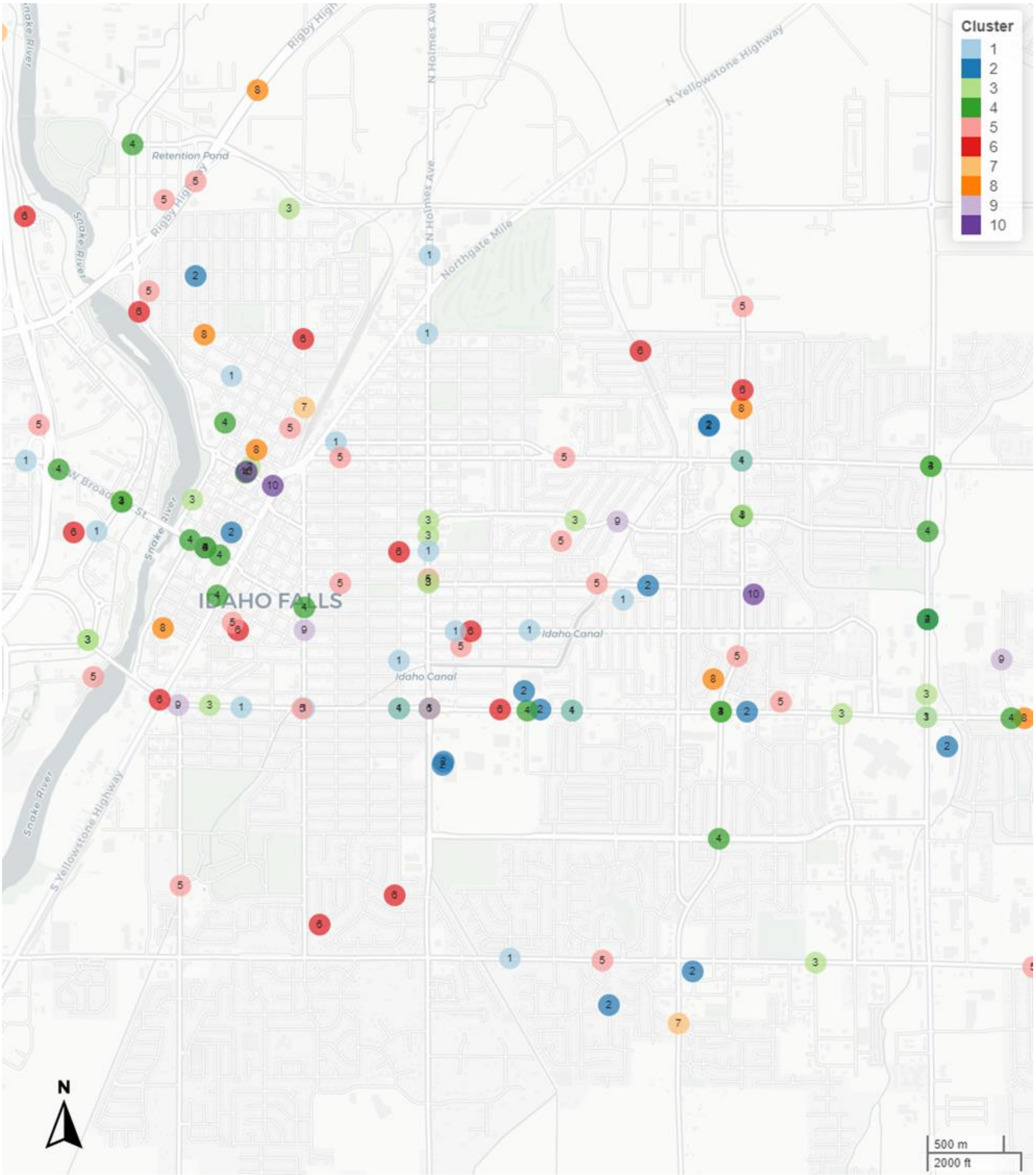
Map 11. Pedestrian crashes in Boise (3 of 3).



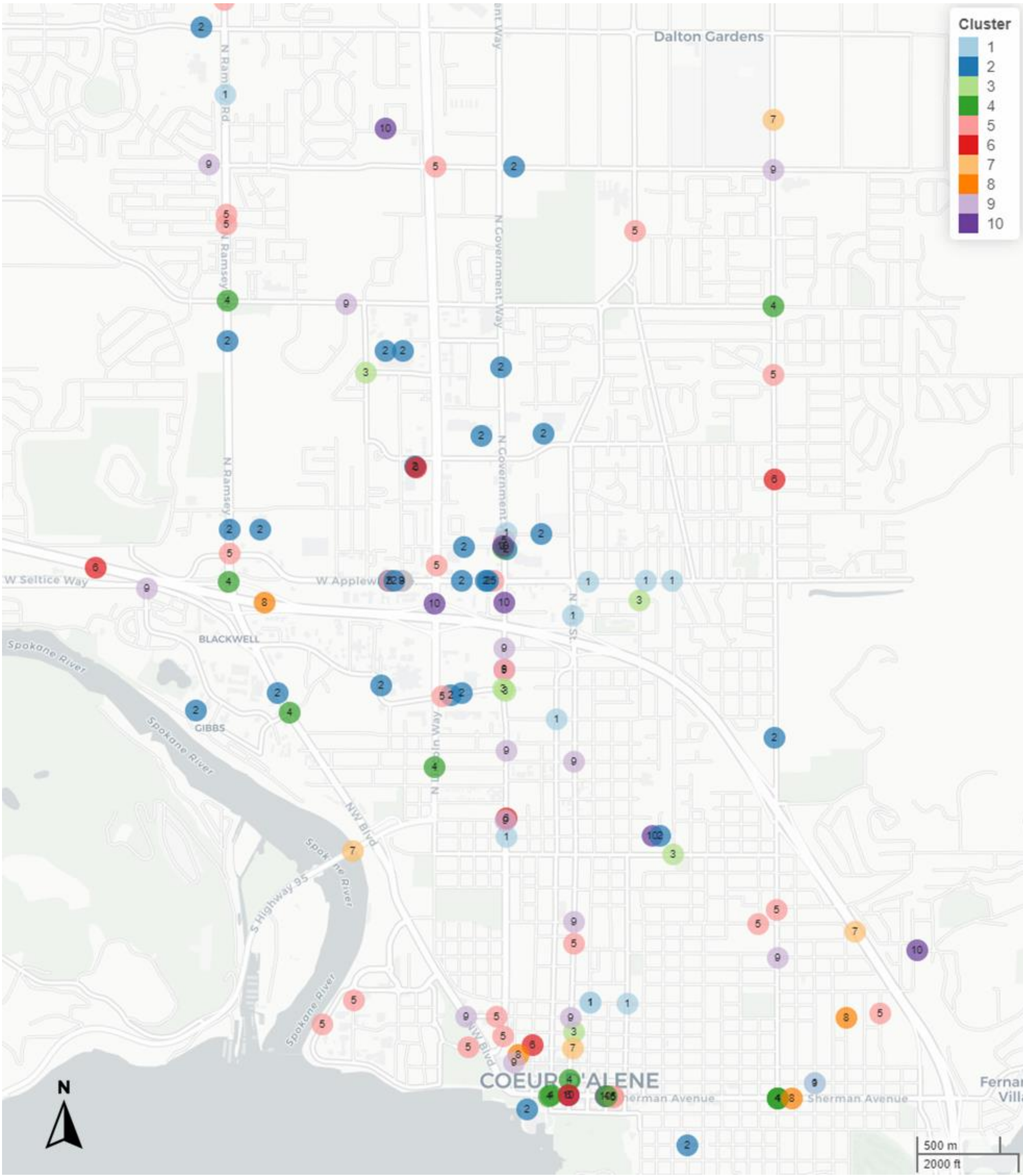
Map 12. Pedestrian crashes in Nampa.



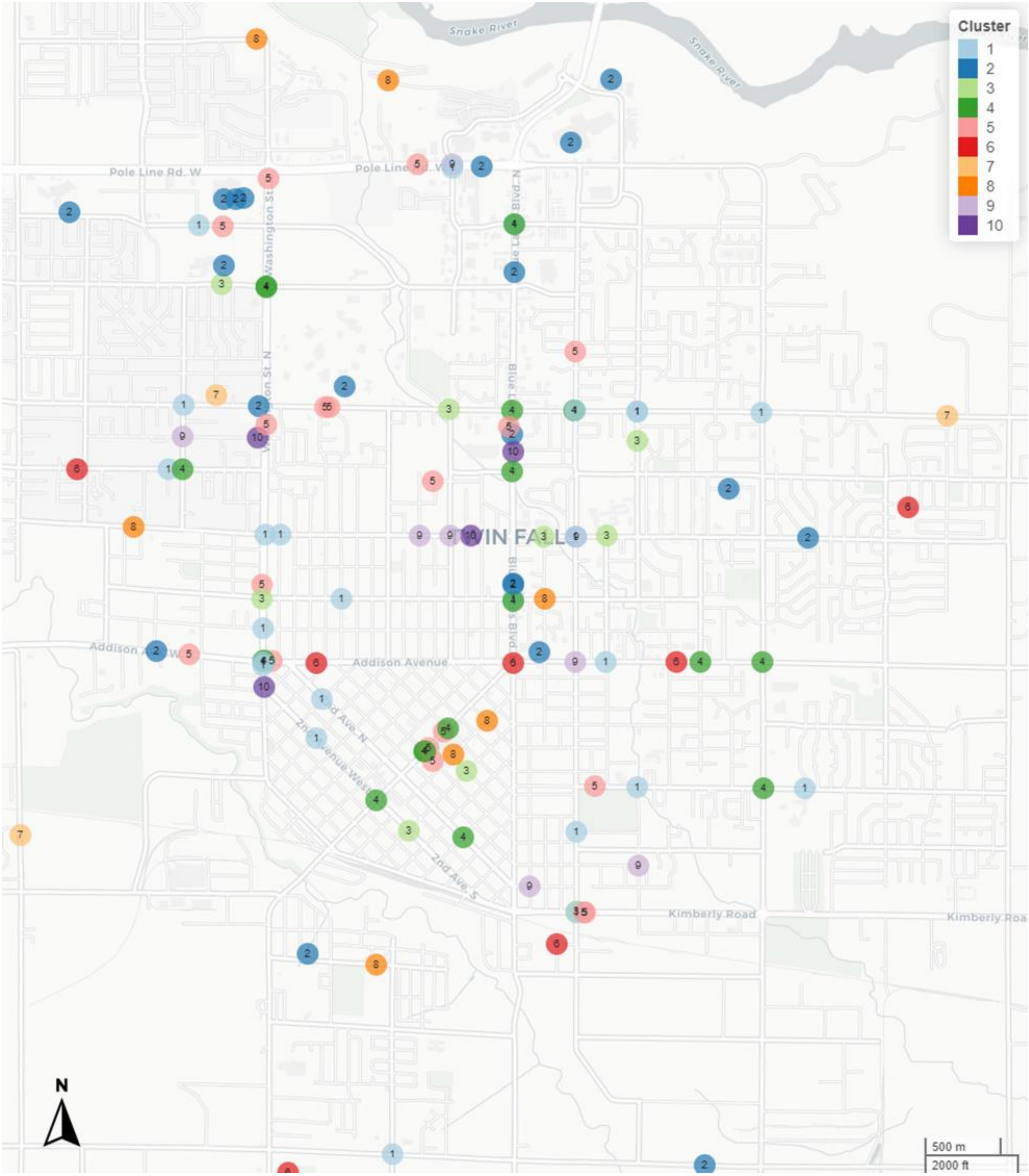
Map 13. Pedestrian crashes in Pocatello.



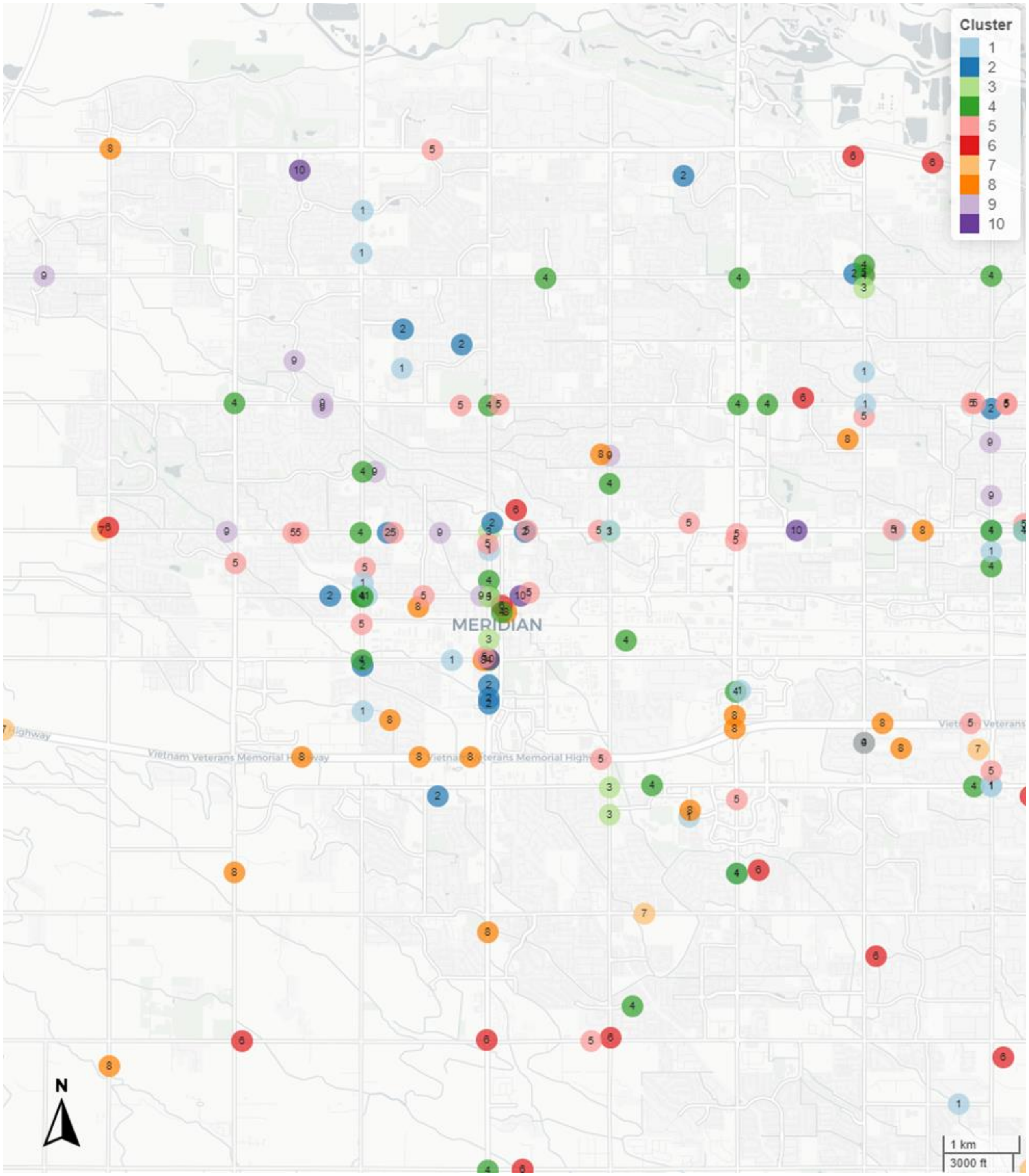
Map 14. Pedestrian crashes in Idaho Falls.



Map 15. Pedestrian crashes in Coeur d'Alene.



Map 16. Pedestrian crashes in Twin Falls.



Map 17. Pedestrian crashes in Meridian.

5. Countermeasure Recommendations

This section provides countermeasure recommendations to mitigate worsening trends, target high occurrence corridors, and address the ten crash clusters for each mode described in the previous section.

All bicycle crash types have decreased between 2012 and 2021. As active transportation increases in popularity, expanding bicycle facilities can encourage further safety improvements. Spikes in crashes occur during rush hour and tend to be on arterials, major collectors, and at intersections. Focusing efforts on the corridors identified in Table 8 can yield the largest improvements.

Two PBCAT2 crash types increased over the 10-year period and account for 25% of pedestrian crashes: crossing driveway or alley (PBCAT2 460) and crossing roadway while vehicle turning (PBCAT2 790). These PBCAT2 crash types correspond to clusters 2, 3, and 4, which account for 37% of all pedestrian crashes. To prevent pedestrian crashes and reduce the severity of the outcomes when a crash does occur, drivers must see the pedestrian sooner. Countermeasures for driveways, alleys, and intersections include:

- Implement parking restrictions around driveways and alleys (see: Section 2B.39 Parking, Standing, and Stopping Signs of the MUTCD), and ensure stringent enforcement of these regulations;
- Restrict loading and unloading in those same areas, and install clear signage (see: Section 2B.39 Parking, Standing, and Stopping Signs of the MUTCD);
- Reduce visual clutter (e.g., roadside advertising, unnecessary regulatory signs, etc.);
- Add lighting to further enhance pedestrian conspicuity.

Crashes involving pedestrians crossing the roadway and turning vehicles can benefit most from reducing turning speeds. Doing so can increase the time that drivers have to spot pedestrians, as well as reduce the severity of crashes that are not prevented. Turning speeds can be reduced without affecting through speed in adjacent lanes by reducing curb corner radii, however consideration should be given to vehicle size and off-tracking. Larger curb radii allow drivers to take turns at higher speeds; smaller radii require motorists to reduce speed to make a relatively sharp turn. They can also shorten crossing distances while providing larger pedestrian waiting areas at corners and improving sight distance.

These countermeasures can increase pedestrian conspicuity and sight distance, but they cannot mitigate the effects of distractions. According to the crash records, driver distraction was involved in just 1% (N=25) of bicycle crashes and 3% (63) of pedestrian crashes. These rates are likely to underestimate the role of distraction as drivers are reluctant to admit to being distracted. In 2021, Idaho implemented the Hands-Free Device law, requiring all electronic devices to be in hands-free mode while driving, including when stopped at a red light or stop sign[1] . Enforcement may be difficult as it requires officers to see

[1] Idaho State Legislature House Bill No. 5, I.C. § 49-1401A, 2021.

drivers in the act, but the law's penalties may serve as effective deterrents. Educating the public about this law, why it is important to drive without distractions, and the consequences for not doing so may reduce pedestrian crashes of various types.

Table 10 and Table 11 present crash counts and countermeasures for bicycle and pedestrian crash type clusters. Some countermeasures can mitigate a large variety of crash types, while others have been developed to address the unique challenges posed by certain scenarios. The recurring emphasis on education and enforcement as countermeasures across multiple crash clusters signifies their foundational role in traffic safety. Education, which is about imparting knowledge and cultivating awareness, is a proactive approach. It seeks to equip road users with the information they need to navigate traffic situations safely, thereby preventing crashes from occurring in the first place. On the other hand, enforcement, which involves the application of laws and regulations, acts as a reactive measure. It serves as a deterrent, ensuring that road users adhere to safe practices, with the knowledge that deviations might lead to penalties. Infrastructure-related countermeasures, such as traffic calming roadway configurations, raised medians, and protected intersections emphasize the role of urban planning and infrastructure design in influencing road user behavior. Well-designed infrastructure can naturally guide users towards safer practices, reducing the reliance on active enforcement or continuous education. Countermeasures such as back-up warning devices in vehicles and legislation to prohibit sidewalk riding underscore the complexity of traffic safety. Each crash cluster presents its own set of challenges, and while broad-based measures can address overarching issues, there's a clear need for specialized solutions to tackle unique problems. Together, these broad and specific countermeasures form a comprehensive toolkit that, when applied judiciously, can significantly improve road safety for all users.

Table 10. Bicycle crash counts and countermeasures by cluster.

Cluster	*Statewide Crashes (2012-2021)*	*Countermeasures*
1. Sideswipes at intersections	275	Wider and more retroreflective pavement markings, bike lanes, bike boxes, education, safe passing laws
2. Speeding, hit and run	286	Reduced corner radii, protected bike lanes, traffic calming roadway configuration
3. Bicyclist failed to yield at signalized intersection	237	Education, enforcement, signage (directed at bicyclists)
4. Backing from alley/driveway	389	Back-up warning devices in vehicles, education, parking restrictions to improve visibility
5. Young motorist failed to yield at stop-controlled intersection	493	Education, enforcement, signage (larger, double, or retroreflective stop signs)
6. Motorist failed to yield at signalized intersection with crosswalk	216	Facilities intended to increase conspicuity (high visibility crosswalks, rectangular rapid-flashing beacons, etc.)
7. Bicyclist failed to yield at stop-controlled intersection	236	Education, enforcement, signage (directed at bicyclists)
8. Roadside or sidewalk, head-on	226	Education, enforcement, signage (directed at bicyclists), legislation to prohibit sidewalk riding
9. Bicyclist crossed at midblock	214	Education, raised medians, signage (directed at bicyclists)
10. Motorist error or impairment	167	Education, enforcement

Table 11. Pedestrian crash counts and countermeasures by cluster (continued next page).

Cluster	*Statewide Crashes (2012-2021)*	*Countermeasures*
1. Pedestrian failed to obey stop sign or signal	346	Education, enforcement, signage (directed at pedestrians), facilities intended to accommodate pedestrians (leading pedestrian intervals, protected intersections, etc.)
2. Pedestrian crossing parking lot, alley, driveway	224	Signage to increase awareness of potential conflict points, clearly marked pedestrian walkways, speed humps, raised crossings, lighting
3. Motorist failed to signal at intersection	178	Education, enforcement, facilities intended to accommodate pedestrians (leading pedestrian intervals, protected intersections, etc.)

Cluster	*Statewide Crashes (2012-2021)*	*Countermeasures*
4. Motorist performed improper left turn	421	Education, enforcement, facilities intended to accommodate pedestrians (leading pedestrian intervals, protected intersections, etc.)
5. Pedestrian failed to yield while crossing at midblock	368	Facilities that encourage proper crossing (offset crossings, raised crosswalks, refuge islands, etc.)
6. Motorist backing	164	Back-up warning devices in vehicles, education, parking restrictions to improve visibility
7. Pedestrian walking along roadway	100	Grade-separated paths, enforcement
8. Motorist speeding, turns and hills	238	Reduced corner radii, traffic calming roadway configuration, advance warning signs at crosswalks
9. Motorist failed to obey stop sign	128	Education, enforcement, signage (larger, double, or retroreflective stop signs)
10. Hit and run	47	Traffic calming roadway configuration, advance warning signs at crosswalks, signage to increase awareness of potential conflict points

6. Discussion

Crashes entail many distinct circumstances, causes, and nuances that can make safety improvement efforts difficult. Crash typing attempts to reduce these details to a manageable number of groups to inform countermeasure selection and implementation. This process, however, can be time-consuming, resource-intensive, and error-prone. The methods described in this report seek to streamline the crash typing process but is not without its limitations.

PBCAT2 crash types require information that is not readily available from crash records derived from police crash reports. A large language model was used to extract the additional required details. Large language models are a relatively new innovation in the field of artificial intelligence, capable of interpreting substantial amounts of text, but they are (currently) unable to decipher images such as crash diagrams. Proper question phrasing and response enumeration can help these tools produce the desired output, but they still require human validation. As this technology matures, researchers can fine tune transportation-specific models to potentially improve the data extraction process.

Clustering is a powerful technique that has been applied in many domains to group similar observations when the volume of detail exceeds human capabilities. It is, however, a stochastic process that can produce non-unique results. This may be beneficial when analyzing geographically diverse areas but may also pose challenges to planning organizations. The process of labelling individual clusters is also challenging as it involves a degree of subjectivity and domain expertise. The number of clusters to generate is another decision that must be made. Clustering algorithms can determine an “optimal” number based on minimum size requirements and similarity statistics. The research team chose to generate ten clusters for practicality purposes; fewer may have hidden the increasing trend among some crash types, while more may yield very small groups.

This analysis relies on information recorded by law enforcement officers and is therefore subject to incompleteness and subjectivity. Several data elements were often missing, including posted speed limits, roadway function class and geometry, and point of impact. Speed of travel would also be helpful in determining crash types and modelling injury severity, but it was not present in the data. In addition, the KABCO injury severity scale is useful for crash analysis, but researchers have identified significant inaccuracies (Burdett et al., 2015; Farmer, 2003; Tsui et al., 2009).

This analysis was also limited by a lack of exposure data (level of travel among motorists, bicyclists and pedestrians). While crash counts are important, highly populated areas often experience more crashes simply because there are more people driving, biking and walking. Incorporating exposure metrics could help to identify areas with elevated crash rates. Many agencies collect data on vehicle travel, but few collect long-term data for active modes of transportation. Commercial entities can provide estimates, though often at a high cost. Future research may produce further insight into crash rates by acquiring and integrating such data.

7. Conclusions

This research collected data on bicycle and pedestrian crashes in Idaho and applied crash typing methodologies to identify safety hotspots and trends, and support data-driven planning decisions that prioritize safety and minimize risk to vulnerable road users. A large language model extracted details from crash narratives to supplement readily available crash records. The research team used this collection of information to apply the PBCAT2 crash typing methodology and hierarchical clustering to further analyze crash groups and recommend appropriate countermeasures. The clustering technique produced crash types similar to those established in PBCAT2, as well as several novel types.

Bicycle crashes decreased over the ten-year period (2012-2021), both overall and within each crash type. Speeding, midblock crossings, and motorist impairment were determined to be the key risk factors for fatal and severe bicycle crashes. Continuing to support active transportation by expanding bicycle facilities can prolong the positive trends into the future.

Overall, pedestrian crashes did not appear to be increasing, despite an upward trend in fatalities. Crash typing, however, identified several groups of crashes that did increase over the ten-year period: those occurring around parking lots, alleys, and driveways; and those involving motorists failing to signal at intersections, making improper left turns, and speeding near turns and hills. There was also an increase in the number of incidents where someone fled the scene. The crash types with the highest proportion of fatal and serious injuries involve walking along the roadway and motorists travelling at high speeds. Countermeasures that reduce motorist speeds, increase pedestrian visibility, increase driver education and awareness, and reduce distractions could help to mitigate these crashes and reverse the increasing trends among some crash types.

Review of bicycle and pedestrian crashes in Idaho from 2012 through 2021 showed that the vast majority of these crashes (93%) occurred in urban areas (defined as within the city limits of an Idaho community). Many of these crashes occurred in a small number of Idaho's largest cities. From 2012-2021, 68% of all bicycle crashes occurred in six cities (Boise, Coeur d'Alene, Nampa, Meridian, Idaho Falls, and Pocatello). Similarly, 62% of all pedestrian crashes during the same period occurred in Boise, Nampa, Pocatello, Idaho Falls, Coeur d'Alene, Twin Falls, and Meridian. The research team developed maps to identify high occurrence corridors in highly populated areas throughout Idaho. Most corridors exhibited a high propensity of crashes for either bicyclists or pedestrians, but several appeared to be problematic for both. All underlying data was submitted to ITD for further analysis and collaboration with local agencies to develop safety action plans.

8. Cited Works

Amsden, M., & Huber, T. (2006). *Bicycle crash analysis for Wisconsin using a crash typing tool (PBCAT) and geographic information system (GIS)* (Final Report). http://wisdotresearch.wi.gov/wp-content/uploads/05-18bicycle-f.pdf

Arteaga, C., Paz, A., & Park, J. W. (2020). Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. *Safety Science*, *132*, 104988. https://doi.org/10.1016/J.SSCI.2020.104988

Bertke, S. J., Meyers, A. R., Wurzelbacher, S. J., Measure, A., Lampl, M. P., & Robins, D. (2016). Comparison of methods for auto-coding causation of injury narratives. *Accident Analysis & Prevention*, *88*, 117–123. https://doi.org/10.1016/J.AAP.2015.12.006

Bhowmik, T., Yasmin, S., & Eluru, N. (2021). A new econometric approach for modeling several count variables: a case study of crash frequency analysis by crash type and severity. *Elsevier*. https://www.sciencedirect.com/science/article/pii/S0191261521001818

Billot-Grasset, A., Viallon, V., Amoros, E., & Hours, M. (2014). Typology of bicycle collisions and falls from a survey of 1078 injured people: some hints for safety advices. *Transport Research Arena*. https://doi.org/10.13140/RG.2.1.2282.7682

Burdett, B., Li, Z., Bill, A. R., & Noyce, D. A. (2015). Accuracy of Injury Severity Ratings on Police Crash Reports. *Https://Doi.Org/10.3141/2516-09*, *2516*, 58–67. https://doi.org/10.3141/2516-09

Carter, D. L., & Council, F. M. (2006). *Factors contributing to pedestrian and bicycle crashes on rural highways*.

Chavis, C., Lee, Y.-J., Dadvar, S., & Studies, M. S. University. Dept. of T. and U. I. (2018). *Analysis of bicycle and pedestrian crash causes and interventions*. https://doi.org/10.21949/1503647

Chen, W., Wheeler, K. K., Lin, S., Huang, Y., & Xiang, H. (2016). Computerized "learn-as-you-go" classification of traumatic brain injuries using NEISS narrative data. *Accident; Analysis and Prevention*, *89*, 111. https://doi.org/10.1016/J.AAP.2016.01.012

Cross, K. D., & Fisher, G. (1977). *A study of bicycle/motor-vehicle accidents: Identification of problem types and countermeasure approaches*.

Das, S., Dutta, A., & Tsapakis, I. (2021). *Topic models from crash narrative reports of motorcycle crash causation study*. *2675*(9), 449–462. https://doi.org/10.1177/03611981211002523

Das, S., Le, M., & Dai, B. (2020). Application of machine learning tools in classifying pedestrian crash types: A case study. *Transportation Safety and Environment*, *2*(2). https://doi.org/10.1093/tse/tdaa010

Depaire, B., Wets, G., & Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. *Accident Analysis & Prevention*, *40*(4), 1257–1266. https://doi.org/10.1016/J.AAP.2008.01.007

Do, A., & Harkey, D. (2006). *Living to tell the tale*. FHWA-HRT-06-005. https://highways.dot.gov/public-roads/julyaugust-2006/living-tell-tale

Farmer, C. M. (2003). Reliability of police-reported information for determining crash and injury severity. *Traffic Injury Prevention*, *4*(1), 38–44. https://doi.org/10.1080/15389580309855

Harkey, D. L., & Blomberg, R. D. (2001). *Beta testing of the pedestrian and bicycle crash analysis tool*. http://www.nhtsa.dot.gov/people/injury/pedbimot/ped/pbcatjan01/technicalreport.html

Harkey, D. L., & Zegeer, C. V. (2004). *PEDSAFE: Pedestrian safety guidelines and countermeasure selection system*. United States. Department of Transportation. Federal Highway Administration. https://doi.org/10.21949/1503647

Harkey, D., Tsai, S., Thomas, L., & Hunter, W. (2006). *Pedestrian and Bicycle Crash Analysis Tool (PBCAT): Version 2.0 Application Manual*.

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, *1*, 328–339. https://doi.org/10.48550/arxiv.1801.06146

Hu, W., & Cicchino, J. B. (2022). Relationship of pedestrian crash types and passenger vehicle types. *Journal of Safety Research*, *82*, 392–401. https://doi.org/10.1016/J.JSR.2022.07.006

Hunter, W. W., Thomas, L., & Stutts, J. C. (2006). *BIKESAFE: Bicycle countermeasure selection system*. United States. Federal Highway Administration. https://doi.org/10.21949/1503647

Intini, P., Berloco, N., Fonzone, A., Fountas, G., & Ranieri, V. (2020). The influence of traffic, geometric and context variables on urban crash types: A grouped random parameter multinomial logit approach. *Analytic Methods in Accident Research*, *28*. https://doi.org/10.1016/j.amar.2020.100141

Li, G., Lai, W., & Qu, X. (2020). Association between crash attributes and drivers' crash involvement: a Study based on police-reported crash data. *International Journal of Environmental Research and Public Health 2020, Vol. 17, Page 9020*, *17*(23), 9020. https://doi.org/10.3390/IJERPH17239020

Marucci-Wellman, H. R., Corns, H. L., & Lehto, M. R. (2017). Classifying injury narratives of large administrative databases for surveillance-A practical approach combining machine learning ensembles and human review. *Accident; Analysis and Prevention*, *98*, 359–371. https://doi.org/10.1016/J.AAP.2016.10.014

Mell, H. K., Mumma, S. N., Hiestand, B., Carr, B. G., Holland, T., & Stopyra, J. (2017). Emergency Medical Services Response Times in Rural, Suburban, and Urban Areas. *JAMA Surgery*, *152*(10), 983. https://doi.org/10.1001/JAMASURG.2017.2230

Mesevage, T. (2020, May 28). *Best text classification APIs – Automatically organize data*. MonkeyLearn. https://monkeylearn.com/blog/text-classification-apis/

Mohamed, M., Saunier, N., Miranda-Moreno, L. F., & Ukkusuri, S. V. (2013). A clustering regression approach: A comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US and Montreal, Canada. *Safety Science*, *54*, 23–37. https://doi.org/10.1016/j.ssci.2012.11.001

Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2011). Data-mining techniques for exploratory analysis of pedestrian crashes. *Transportation Research Record*, *2237*, 107–116. https://doi.org/10.3141/2237-12

Natarajan, S., Demetsky, M., & Lantz, K. (2008). *Framework for selection and evaluation of bicycle and pedestrian safety projects in Virginia*. https://rosap.ntl.bts.gov/view/dot/19892

National Highway Traffic Safety Administration. (2022). *2020 FARS/CRSS pedestrian bicyclist crash typing manual: a Guide for coders using the FARS/CRSS ped/bike typing tool*.

NHTSA. (1983a). *Manual Accident Typing for Bicyclist Accidents: Coder's Handbook*.

NHTSA. (1983b). *Manual Accident Typing for Pedestrian Accidents: Coder's Handbook*.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. https://github.com/codelucas/newspaper

Ragland, D. R., Markowitz, F., & Macleod, K. E. (2003). *An intensive pedestrian safety engineering study using computerized crash analysis*.

Sayed, M. A. (2022). *Use of machine learning and natural language processing to enhance traffic safety analysis*. University of Wisconsin-Milwaukee.

Schnee, J., Stegmaier, J., & Li, P. (2021). A probabilistic approach to online classification of bicycle crashes. *Accident Analysis & Prevention*, *160*, 106311. https://doi.org/10.1016/J.AAP.2021.106311

Schneider, R., & Stefanich, J. (2016). Application of the location-movement classification method for pedestrian and bicycle crash typing. *Transportation Research Record*, *2601*, 72–83. https://doi.org/10.3141/2601-09

Settles, B. (2012). *dualist: Interactive machine learning for text analysis* (0.3). https://github.com/burrsettles/dualist#readme

Shah, N. R., Aryal, S., Wen, Y., & Cherry, C. R. (2021). Comparison of motor vehicle-involved e-scooter and bicycle crashes using standardized crash typology. *Journal of Safety Research*, *77*, 217–228. https://doi.org/10.1016/J.JSR.2021.03.005

Smist, T. E. (1982). *Pedestrian injury causation study (pedestrian accident typing)*. United States. National Highway Traffic Safety Administration. https://doi.org/10.21949/1525143

Snyder, M. B., & Knoblauch, R. L. (1971). *Pedestrian safety: the Identification of precipitating factors and possible countermeasures*. United States. National Highway Traffic Safety Administration. https://doi.org/10.21949/1524846

Song, L., Fan, W. (David), Li, Y., & Wu, P. (2021). Exploring pedestrian injury severities at pedestrian-vehicle crash hotspots with an annual upward trend: A spatiotemporal analysis with latent class random parameter approach. *Journal of Safety Research*, *76*, pp 184-196. https://doi.org/https://doi.org/10.1016/j.jsr.2020.12.008

Spainhour, L. K., Wootton, I. A., Sobanjo, J. O., & Brady, P. A. (2006). Causative factors and trends in Florida pedestrian crashes. *Transportation Research Record*, *1982*(1), 90–98. https://doi.org/10.1177/0361198106198200112

Sun, M., Sun, X., & Shan, D. (2019). Pedestrian crash analysis with latent class clustering method. *Accident Analysis & Prevention*, *124*, 50–57. https://doi.org/10.1016/J.AAP.2018.12.016

Thomas, L., Levitt, D., Vann, M., Blank, K., Nordback, K., & West, A. (2022). *PBCAT pedestrian and bicycle crash analysis tool version 3.0 user guide*. United States. Federal Highway Administration. Office of Research, Development, and Technology. https://doi.org/10.21949/1503647

Thomas, L., Nordback, K., & Sanders, R. (2019). Bicyclist crash types on national, state, and local levels: a New look. *Transportation Research Record*, *2673*(6), pp 664-676. https://doi.org/10.1177/0361198119849056

Thomas, L., Sandt, L., Zegeer, C., Kumfer, W., Lang, K., Lan, B., Horowitz, Z., Butsick, A., Toole, J., & Schneider, R. J. (2018). Systemic pedestrian safety analysis. In *Systemic Pedestrian Safety Analysis*. Transportation Research Board. https://doi.org/10.17226/25255

Tsui, K. L., So, F. L., Sze, N. N., Wong, S. C., & Leung, T. F. (2009). Misclassification of injury severity among road casualties in police reports. *Accident Analysis and Prevention*, *41*(1), 84–89. https://doi.org/10.1016/J.AAP.2008.09.005

Vavrova, M., Chang, C., Kumar, S., Ruiz, D., Gonzalez, S., & Benitez, M. (2021). *North Texas bicycle and pedestrian crash analysis*. University of Texas at El Paso. Center for Transportation Infrastructure Systems. https://doi.org/10.21949/1503647

Wolff, R. (2020, August 26). *5 types of classification algorithms in machine learning*. MonkeyLearn. https://monkeylearn.com/blog/classification-algorithms/

Zegeer, C. V, Blomberg, R., Henderson, D., Masten, S., Marchetti, L., Levy, M., Fan, Y., Sandt, L., Brown, A., Stutts, J., & Thomas, L. (2008). Evaluation of the Miami-Dade pedestrian safety demonstration project. *Transportation Research Record*, *2073*, 1–10.