Boston Area Research Initiative

RADCLIFFE INSTITUTE FOR ADVANCED STUDY HARVARD UNIVERSITY

Dan O'Brien • Research Director • Boston Area Research Initiative
Radcliffe Institute for Advanced Study • Harvard University • 10 Garden Street, Cambridge, MA 02138
Phone: 617-496-9975 • E-Mail: daniel_obrien@radcliffe.harvard.edu • Web: www.bostonarearesearchinitiative.net

DOCUMENTATION FOR THE BICYCLE COLLISIONS IN BOSTON DATABASE

Overview

This document describes the structure and organization of the Bicycle Collisions in Boston database, which draws off of Boston Police Department records to document those bike collisions that occurred in Boston from 2009 – 2012. The database was constructed as part of a research-policy collaboration between Dahianna Lopez (Harvard School of Public Health) and the Boston Police Department.

The database contains a single spreadsheet (*Bicycle Collisions.xlsx*) and an accompanying shape file that enables mapping and spatial analysis (*Bicycle Collisions.shp* and associated files.

Table of Contents

1. SUMMARY OF BICYCLE COLLISIONS DATABASE (*Bicycle Collisions.xlsx*)

1.1 Description of Contents

The Bicycle Collision database contains all bicycle collisions in the city of Boston during 2009 – 2012 that were recorded in Boston Police Department records. The database itself is derived predominantly from these records. The database contains 1,815 collisions.

1.2 Description of Variables

The database contains 51 variables grouped into seven categories:

1) Initial Report Content, or basic information associated with the initial emergency report;
2) Geographic Location describes where the collision occurred;
3) Context of Collision describes various man-made factors that might have contributed to the collision;
4) Weather describes the environmental conditions at the time of the collision
5) Liability includes information referencing potential fault and legal liability in the collision.
6) Demographic characteristics of the cyclist.
7) The official Boston Police Department narrative describing the incident.

Some of the variables are dichotomous, indicating whether something was or was not true. These all take the numerical values: 1 = "Yes"; 0 = "No"; 99, 98 or [Blank] = "Unknown" except where noted.

Also, many variables are derived from the content of the narratives, coded by a team of readers. This is true for most of the measures describing the Context of Collision and Liability.

1.2.1. Initial Report Content

- *ID* is the unique code used to track bicycle collisions in the document.
- *YEAR* is the year of the collision.
- *DATE* is the date of the collision
- *DAY_WEEK* is the day of the week on which the collision occurred.
- *TIME* is the time of the collision

- *TYPE* indicates the vehicles involved in the collision.
    - *Note:* Possible values include: CO = Cyclist Only; BVA = Bike vs. Auto; BVP = Bike vs. Pedestrian; BVB = Bike vs. Bike; BR = Bicycle-Related (during narrative reading, it was clear that a bicycle was involved, but the exact details were unknown)
- *SOURCE* indicates how the report was received.
    - *Note:* Possible values include: CALL SCREENING = ; E911 = 911 Call; N/A = Non-applicable; ON-SITE = Officer in area at time of collision; PHONE-IN = Direct call to police station; WALK-IN = Walk-in report at police station

1.2.2. Geography

- *XFINAL* is the x-coordinate generated by the BPD (in feet).
- *Xkm* is the x-coordinate in meters.
- *YFINAL* is the y-coordinate generated by the BPD (in feet).
- *Ykm* is the y-coordinate in meters.
- *Address* is the nearest address to the collision.
- *Main* indicates whether the street was a main street/major thoroughfare or not (based on BARI's Geographical Infrastructure for the City of Boston).
- *RoadType* indicates the predominant zoning on the road where the collision occurred.
    - *Note:* Possible Values: Com = Commercial, Emp = No Zoning, Exe = Exempt (i.e., government, non-profit), Ind = Industrial, Res = Residential.
- *ISINTERSECTION* indicates whether the crash occurred at an intersection.
- *TRACT* is the FIPS code for the census tract where the collision occurred.
- *CouncilDISTRICT* indicates the city council districts in which the collision occurred.
- *Councillor* is the name of the councilor representing that district as of May 2013.
- *PlanningDistrict* indicates the Boston Redevelopment Authority planning district in which the collision occurred.

1.2.3. Context of Collision

- *OIF1* are factors that potentially contributed to the crash. These factors were specified through consultation with various stakeholders, including city agencies and bicycle advocates. Their presence in a given collision was identified through the

narrative (see section *1.2.6*). This is true for each of the *OIF* variables. See Appendix A for a complete list of OIFs.

  - *Note:* Possible values: 1 = Car ran red light; 2 = Car ran a stop sign; 3 = Car was speeding; 4= Bike ran red light; 5 = Bike ran a stop sign; 6 = Bike was speeding; 7 = Cyclist riding to oncoming traffic; 8 = Other (__write in___); 10 = Poor visibility of bike; 11 = Poor visibility of car; 14 = Personal item caught; 15 = Car door opened and struck biker; 16 = Cyclist hit parked care; 17 = Car made right turn and crashed into bicyclist in bike lane.

- *OIF2* indicates a second influencing factor identified in the narrative (see *OIF1*).
- *OIF3* indicates a second influencing factor identified in the narrative (see *OIF1*).
- *OIF4* indicates a second influencing factor identified in the narrative (see *OIF1*).
- *BLFinal* indicates whether the narrative mentioned the presence of a bike lane.
  - *Note:* This variable should not be used to identify where bike lanes are in the city, only whether they were mentioned in the narrative. The locations of bike lanes can be requested from the Boston Transportation Department.
- *CS* answers whether there was a conflict of stories between parties? (i.e., did the cyclist state they were stopped at the light and did the motorist state that the cyclist was speeding through the light?) ("1" = Yes, "2" = No)
- *LIGHTING* is the lighting situation at the time of the collision.
- *Indoor* indicates whether the crash occurred inside of a covered parking lot or other covered space.
- *Light* refers to whether the police officer stated a problem with street lighting based on narrative content.
- *LightEng* indicates whether the police officer noted a problem specifically with lighting fixtures.

1.2.4. Weather

Variables describing the ambient conditions were accessed from the National Oceanic and Atmospheric Administration (NOAA). They are derived from the sensor at Logan Airport.

- *WEATHER* is a qualitative description of the weather at the time of the collision.
- *PrecipCond* describes the amount of precipitation on the day of the collision.
  - *Note:* Possible values: 1 = No/Trace Precipitation (<.05 in.); 2 = Light Precipitation (.05 - .29 in.); 3 = Heavy (>=0.30 in.).

- *AtmosCond* refers to the temperature at the time of the collision.
    - *Note:* Possible values: 1 = >= 60 °F; 2 = 32-59 °F; 3 = <=31.
- *DayNight* refers to the period of day or night at the time of the collision.
    - *Note:* 1 = Day; 2 = Night; 3 = Sunrise; 4 = Sunset.
- *Tmax* is the maximum temperature on the day of the collision.
- *Tmin* is the minimum value of temperature on the day of the collision.
- *Tavg* is the average value of temperature on the day of the collision.
- *Temprange* is the range of temperature on the day of the collision.
- *SunriseTime* is the time of sunrise on the day of the collision.
- *SunsetTime* is the time of sunset on the day of the collision.
- *Snowfall* is the amount of snowfall on the day of the collision.
- *PrecipTotal* is the total amount of precipitation on the day of the collision.

1.2.5. Liability

- *Fault* indicates whether the police officer assigned fault to a party.
- *Doored* indicates whether the bicyclist came into contact with an opening vehicle door.
- *HelmetDocumented* indicates whether the use of helmet by a cyclist was documented.
    - *Note:* In most cases, there was no documentation of the use of a helmet by a cyclist. “0” indicates that it was noted that a helmet was not worn. “Unknown” indicates that neither the presence nor absence of a helmet was documented.
- *TaxiFinal* indicates whether the vehicle involved was a taxi cab or other livery vehicle.
- *Hitrunfinal* indicates whether the collision was a hit-and-run.
- *AlcoholFinal* indicates whether alcohol use was noted for any party involved in the collision.
- *INJURED* is whether one or more people were injured due to the collision.
- *TRANSPORTRED* is whether one or more people were transported to a hospital via ambulance due to the collision.

- *TREATED* is whether one or more people were treated medically after the collision.

1.2.6. Demographics of Cyclist

- *GENDER* indicates the gender of the cyclist.
- *ETHNICITY* indicates the ethnicity of the cyclist.
- *AGE* indicates the age of the cyclist.

1.2.7. Narrative

- *Narrative* presents the complete account of the collision as recorded by police officers.
    - *In order to protect client confidentiality, researchers at Harvard University used computer science methods to strip personal identifiers from the narratives. Any potential personal identifiers were replaced by "xxx" sequences.*
    - *Note:* Not available in mappable shape file.