## Data Description

NCHRP 22-49 The Effect of Vehicle Mix on Crash Frequency and Crash Severity

Data were collected during the NCHRP 22-49 study, “The Effect of Vehicle Mix on Crash Frequency and Crash Severity”. The study compiled facility-level data from 7 states (including California, Connecticut, Florida, Illinois, Minnesota, Texas, and Washington). The data includes crash information by severity level, traffic characteristics such as AADT, vehicle mix information such as percentage of truck traffic and percentage of single-unit truck traffic, and roadway characteristics such as lane width, shoulder type, shoulder width, median width, light condition, and speed limit information collected and compiled from different data sources such as department of transportation and highway safety information system. The study considered a total of 24 facilities (17 segments and 7 intersections) for developing crash prediction models (as shown below). For intersection facilities, crashes that occurred within a 250ft intersection buffer are assigned to intersections. A detailed description of the data sources and data preparation can be found in the project final report.

* Facility group 1 - Urban Limited Access:
  + Urban 4-lane divided (ULA4LD)
  + Urban 6-lane divided (ULA6LD)
  + Urban 8-lane divided (ULA8LD)
  + Urban 10-lane divided (ULA10LD)
* Facility group 2 - Rural Limited Access:
  + Rural 4-lane divided (RLA4LD)
  + Rural 6-lane divided (RLA6LD)
  + Rural 8-lane divided (RLA8LD)
* Facility group 3 - Urban Arterials:
  + Urban 2-lane undivided (UA2LUD)
  + Urban 3-lane (UA3L)
  + Urban 5-lane (UA5L)
  + Urban 4-lane undivided (UA4LUD)
  + Urban 4-lane divided (UA4LD)
* Facility group 4 - Rural Arterials:
  + Rural 2-lane undivided (RA2LUD)
  + Rural 3-lane (RA3L)
  + Rural 5-lane (RA5L)
  + Rural 4-lane undivided (RA4LUD)
  + Rural 4-lane divided (RA4LD)
* Facility group 5 - Urban Intersections:
  + Urban 3-leg STOP controlled (U3ST)
  + Urban 4-leg STOP controlled (U4ST)
  + Urban 3-leg signalized (U3SG)
  + Urban 4-leg signalized (U4SG)
* Facility group 6 - Rural Intersections:
  + Rural 3-leg STOP controlled (R3ST)
  + Rural 4-leg STOP controlled (R4ST)
  + Rural 4-leg signalized (R4SG)

For each facility, two sets of data were provided including a) estimation dataset, and b) validation dataset. The estimation dataset was used to develop the model while the validation dataset was used to check the model performance. All the datasets were provided in excel files. A Microsoft word document containing the data dictionaries was also provided.