Does Driving a Car Without a Seatbelt Increase the Likelihood of Additional Infractions?

ATE-252 Capstone Project
by Zach Trozenski (SP 2020)

Central Research Question:
Are unbelted drivers more likely to incur secondary traffic violations?

Data Preparation:
Data was downloaded from the Commonwealth of Pennsylvania's OpenData website. The dataset contains crash incident details from Pennsylvania logged from 1997 to the current year reported to the Pennsylvania Department of Transportation (PennDOT).

Initial attempts to load and process the data into OpenRefine and Excel failed due to size of the data set. As a classmate pointed out, the data could be pre-filtered on the website and redownloaded. Using this filtering functionality, I created two subsets: one containing belted crashes and the other containing unbelted crashes.

However, the subsets proved yet again to be too large. Using the coreutils gshuf tool, I randomly sampled 500 rows of each subset. 500 samples were chosen as the tools were able to load the data sets properly. The next section ("About the Data") provides an analysis of the sampled datasets used in this investigation.

Methodology:
In order to aggregate tabulations, all "Yes/No" values in the data sets for violations, driver age, injury, etc. were converted to binary (i.e 1 or 0) with OpenRefine. Subsequently, columns which were not investigated were removed. A full process log of what was done to the data sets can be found on my GitHub repository. Infractions which drivers incur in addition to being unbelted are considered secondary violations for the purposes of this investigation. It should not be confused with secondary enforcement which refers to the statute governing seat belt citations.

Once the data had been cleaned and normalized, datasets were loaded into Jupyter Notebooks. Using the pandas library, the two datasets were manipulated using the same code (in order to maintain consistency) and plotted using matplotlib.pyplot. The code used can be found in my GitHub repository for further investigation, if desired.

Additionally, I created QGIS maps showing the crashes for each road type. I pulled shapefiles and data from PASDA, TIGER, and the PA Turnpike website in order to construct these maps. I created 5 maps for the following roads: state, local, unpaved, interstate, and Turnpike. While the first three aforementioned road types had proper shapefiles, I could not find interstate or Turnpike shapefiles. As a replacement for an interstate shapefile, I plotted each mile marker and exit which creates pseudo lines that can be traced. However, for the Turnpike I was only able to plot the interchanges along the road, relying on the underlying PA Roads shapefile to assist in tracing the road's path.

Larger PDF versions of maps can be found on my GitHub repository.

About the Data:
An analysis reveals the sampled data sets (unbelted vs. belted crashes) are similarly distributed across age cohorts (Data section,

Fig A), crash year (Fig. B), road type (Fig. C), and geographic location across the state (Appendix, Fig's 1-5).
The distribution for the cohorts of drivers aged 16, 18, and 20 sees more unbelted drivers, while cohorts for ages 17, 19, 65-74, and 75+ have fewer unbelted drivers. At face value, this would suggest younger drivers being belted less frequently, the cohorts for age 17 and 19 having more belted than unbelted drivers suggest there is not necessarily a bias towards younger drivers neglecting seat restraints.

The Commonwealth of Pennsylvania's Traffic Code contains legislation that designates the commonwealth as a Secondary Enforcement state. Drivers may not be stopped solely for failing to wear a seatbelt but can be cited for failing to use a seatbelt if stopped for a Primary Enforcement offense (such as speeding). For drivers under the age of 18, as well as minors ages 8-17 in all seats, failure to wear a seatbelt falls under Primary Enforcement and minors may be stopped for this alone.

Pennsylvania's law mandating use of seatbelts went into effect in 1983 and became a secondary enforcement state in 1987. The enactment of the law may be correlated to the data represented in belted versus unbelted crashes by year (Fig B.), as it may be driving the decrease of unbelted crashes over time. The inverse (drivers steadily wearing seatbelts more in crashes) does not hold true as the data indicates a fluctuating number of belted drivers across the sampled year range. Further research into different data would be needed to explain trends of seatbelt use over time.

With regards to road type, there are fewer unbelted crashes on larger roadways (i.e. Turnpike, interstates, and state roads). Most crashes (between either data set) occurred on state roads. The only road type where unbelted crashes occurred more than belted crashes was on local roads. The data suggests, and is supported by the visualization (Appendix, Fig. 4), that the bulk of the roads in Pennsylvania are state roads which may explain the glut of crashes having occurred on this road type.

Data Analysis:
The analysis of the data is focused on the likelihood of unbelted drivers incurring a second violation (beyond being unbelted) in crash situations (Fig. I). As a control group, belted crashes were used to give a baseline (Fig H). In addition to exploring the statistical significance between the two data sets (Fig J) there will also be an assessment of the severity of unbelted crashes compared to belted crashes measured in injuries and fatalities (Fig's. C – G).

There are twenty categories that have been selected: Alcohol Related, Drinking Driver, Underage Drinking Driver, Unlicensed, Distracted, Cell Phones, No Clearance, Running Red Light, Tailgating, Speeding, Speeding Related, Aggressive Driving, Fatigue/Asleep, National Highway Transportation Safety Administration (NHTSA) Aggressive Driving, Running Stop Sign, Hit Parked Vehicle, Drug Related, Illegal Drug Related, Drugged Driver, and Impaired Driver. While some of these categories appear to overlap, the data dictionary for the PA crash data draws separations between related categories, as such they have been kept separate entities with the caveat there may be some double counting.

An examination of raw counts of violations between the two data sets (Fig. J) reveals four categories out of the total twenty where belted drivers incurred more violations than unbelted drivers in the sampled data. Belted drivers hit parked vehicles, tailgated, drove distracted, and drove aggressively according the NHTSA more than unbelted drivers by an average difference of 5 occurrences.

In the remaining sixteen categories (alcohol related, drinking driver, underage drinking driver, unlicensed, using a cell phone, no clearance, running red light, speeding, speeding related, aggressive driving, fatigue or asleep, running stop sign, drug related, illegal drug related, drugged driver, and impaired driver) unbelted drivers incur more secondary violations by an average difference of 23 occurrences.

In order to determine whether or not the difference between the two dataset is statistically significant a *t* test was conducted using the Analysis Tookpak in Excel. The result (compiled in Fig. K) indicates *p* values of 0.0049 for one tailed distribution and 0.0099 for two tailed distribution. When compared to their belted counterparts, unbelted drivers are more likely to incur a secondary violation.

When it comes to injuries the data confirms what is already widely known and accepted regarding the severity of injuries associated with unrestrained persons in crashes. According to the sampled data, unbelted crashes have resulted in more injuries and fatalities than belted crashes (Fig. D and Fig. E).

Finally, as Fig's. F and G illustrate, unbelted crashes result in higher counts of persons injured, sometimes increasing the total persons injured from 4 to 8.

The data suggests drivers who are belted are in less severe crashes and are less likely to incur secondary violations. Using the data borne from this investigation, a case can be made to PennDOT to advocate for the Commonwealth to adopt Primary Enforcement legislation into the traffic code. By strengthening enforcement of seat belt usage among Pennsylvanians, PennDOT can reduce the severity of crashes and promote safer driving practices.

Further Questions:

Further investigation can be conducted into the statistical significance of the injuries between the data sets, rather than simply tabulated raw counts and comparing totals.

The perception of wearing a seatbelt in the context of the urban rural divide would be another interesting investigation. Determining what the perception and importance is of wearing a seat belt between two demographics may uncover interesting conclusions about Pennsylvania's populace.

I was also struck but the number of unpaved roads in Pennsylvania. I am curious to the extent of which those roads belong to farms. Since farm vehicles occupy a special status in the PA vehicle code, I am wondering if they follow similar trends to what is already logged in the data set.

Data:

Fig A. Distribution of Driver's Age

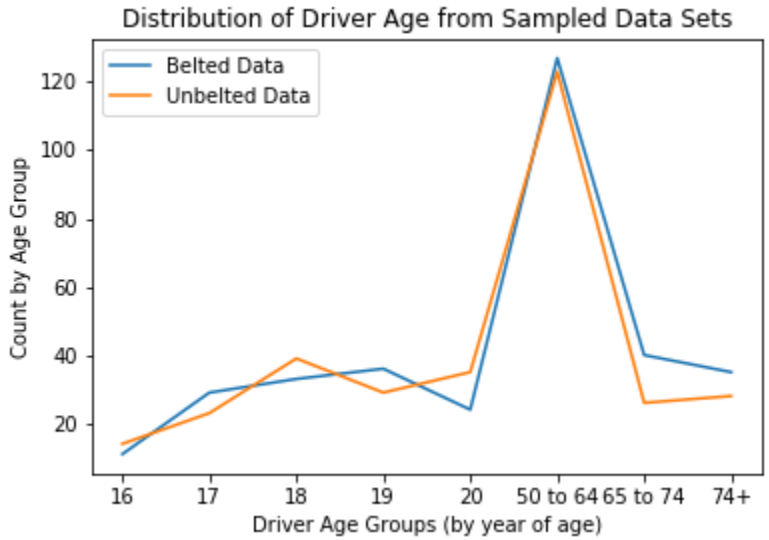


Fig B. Distribution of Crash Year

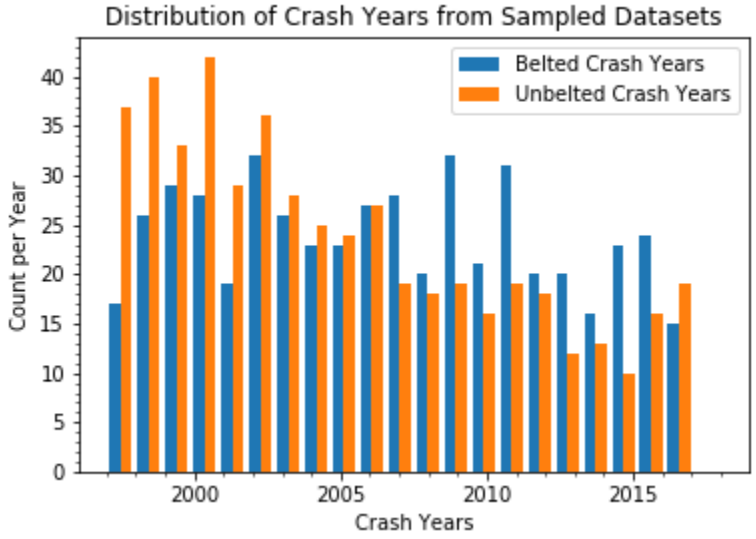


Fig C. Road type of crash location

Road Type	Belted Crashes	Unbelted Crashes
Interstate	34	20
State Road	361	348
Local Road Only	131	149
Turnpike	8	3

Fig D. Unbelted Injury/Fatality

Injury or Fatality in Crash?	Injury	Fatal
No	137	482
Yes	363	18

Fig E. Belted Injury/Fatality

Injury or Fatality in Crash?	Injury	Fatal
No	256	497
Yes	244	3

Fig F. Belted Injury Count

Person Count	Injury Count	Suspected Serious Injury Count	Suspected Minor Injury Count	Possible Injury Count	Total Injury Count
0	244	490	459	326	256
1	191	8	37	140	182
2	45	2	3	26	43
3	15	0	1	7	15
4	5	0	0	1	4

Fig G. Unbelted Injury Count

Person Count	Injury Count	Suspected Serious Injury Count	Suspected Minor Injury Count	Possible Injury Count	Total Injury Count
0	125	461	389	271	137
1	232	36	84	156	225
2	88	2	21	43	87
3	28	1	4	19	27
4	13	0	2	5	11
5	5	0	0	4	5
6	7	0	0	2	6
7	1	0	0	0	1
8	1	0	0	0	1

Fig H. Belted Violations

	Alcohol Related	Drinking Driver	Underage Drinking Driver	Unlicensed	Distracted
No	474	474	496	499	458
Yes	26	26	4	1	42
	Cell Phone	**No Clearance**	**Running Red Light**	**Tailgating**	**Speeding**
No	496	468	482	470	481
Yes	4	32	18	30	19
	Speeding Related	**Aggressive Driving**	**Fatigue / Asleep**	**NHTSA Aggressive Driving**	**Running Stop Sign**
No	390	222	490	476	483
Yes	110	278	10	24	17
	Hit Parked Vehicle	**Drug Related**	**Illegal Drug Related**	**Drugged Driver**	**Impaired Driver**
No	464	495	497	496	470
Yes	36	5	3	4	30

Fig I. Unbelted Violations

	Alcohol Related	Drinking Driver	Underage Drinking Driver	Unlicensed	Distracted
No	400	400	487	494	459
Yes	100	100	13	6	41
	Cell Phone	**No Clearance**	**Running Red Light**	**Tailgating**	**Speeding**
No	495	464	472	480	460
Yes	5	36	28	20	40
	Speeding Related	**Aggressive Driving**	**Fatigue / Asleep**	**NHTSA Aggressive Driving**	**Running Stop Sign**
No	364	218	487	477	480
Yes	136	282	13	23	20
	Hit Parked Vehicle	**Drug Related**	**Illegal Drug Related**	**Drugged Driver**	**Impaired Driver**
No	472	472	489	472	387
Yes	28	28	11	28	113

Fig J. Belted vs. Unbelted Driver Violations

	Alcohol Related	Drinking Driver	Underage Drinking Driver	Unlicensed	Distracted
Belted	26	26	4	1	42
Unbelted	100	100	13	6	41
	Cell Phone	**No Clearance**	**Running Red Light**	**Tailgating**	**Speeding**
Belted	4	32	18	30	19
Unbelted	5	36	28	20	40
	Speeding Related	**Aggressive Driving**	**Fatigue / Asleep**	**NHTSA Aggressive Driving**	**Running Stop Sign**
Belted	110	278	10	24	17
Unbelted	136	282	13	23	20
	Hit Parked Vehicle	**Drug Related**	**Illegal Drug Related**	**Drugged Driver**	**Impaired Driver**
Belted	36	5	3	4	30
Unbelted	28	28	11	28	113

Fig. K t-Test: Paired Two Sample for Means (Belted vs. Unbelted)

	Belted	***Unbelted***
Mean	35.95	53.55
Variance	3816.05	4368.365789
Observations	20	20
Pearson Correlation	0.909762413	
Hypothesized Mean Difference	0	
df	19	
t Stat	-2.863557401	
P(T<=t) one-tail	0.004971357	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	0.009942714	
t Critical two-tail	2.093024054	

Appendix:

Fig. 1

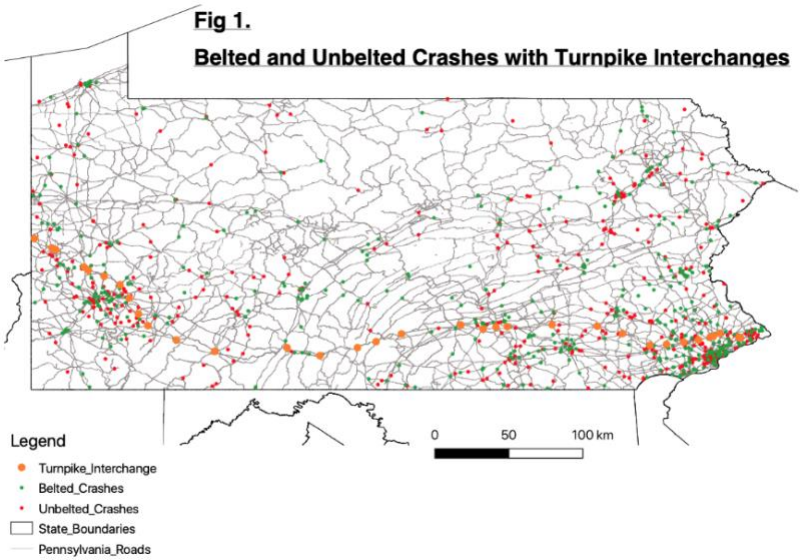


Fig. 2

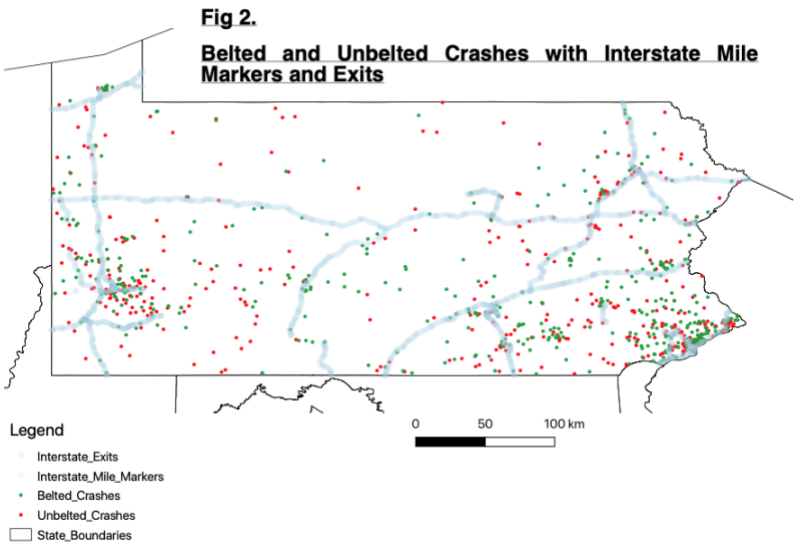


Fig. 3

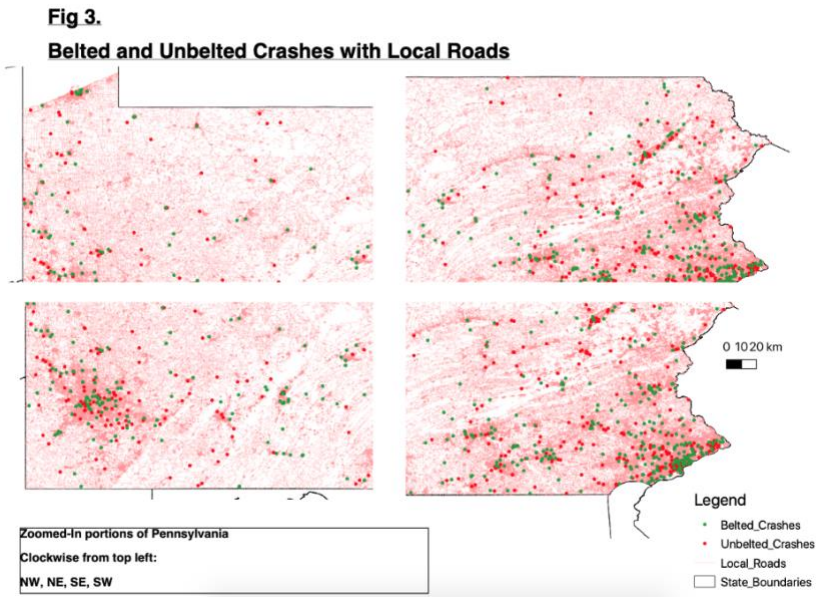


Fig. 4

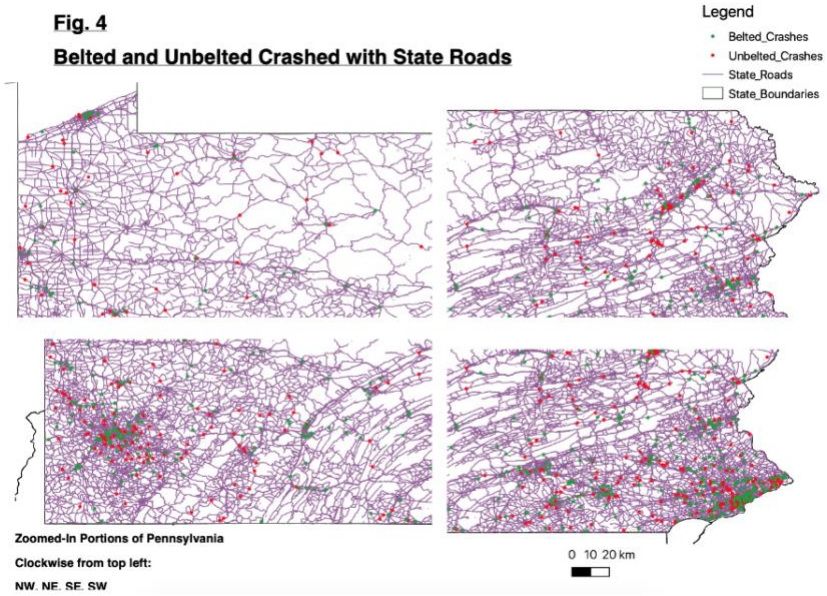
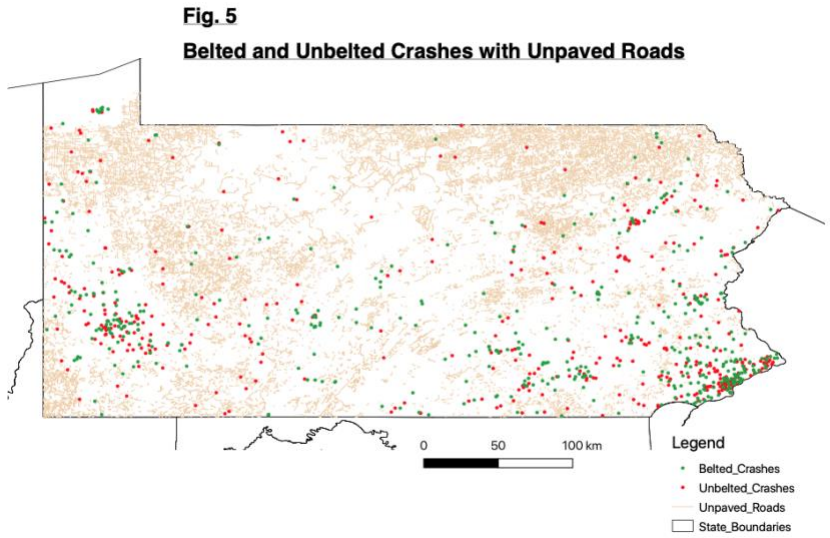


Fig. 5



Sources:

PA State Crash Data: https://data.pa.gov/Public-Safety/Crash-Incident-Details-CY-1997-Current-Annual-Coun/dc5b-gebx

TIGER/Line Shapefiles: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2018.html

Statute Enforcement source: https://en.wikipedia.org/wiki/Seat_belt_laws_in_the_United_States

PA Title 75: https://www.legis.state.pa.us/WU01/LI/LI/CT/HTM/75/75.HTM

Shapefiles for PA Roads: https://www.pasda.psu.edu/

Turnpike Interchange Coordinates: https://www.paturnpike.com/travel/interchanges.aspx

Coordinate conversion tool: http://www.zonums.com/online/coords/cotrans.php?module=11