

Conditions of Car Accidents In Iowa City

Agenda

- Executive Summary
- Question Framing
- Problem Framing
- Data
- Methods
- Graphs and visualizations
- Conclusion

Executive Summary

Problem: Whether there are certain conditions that result in a driver being more likely to be involved in a harmful or fatal accident

Measure: Independence between variables

- Independent Variable: Conditions impacting accidents
- Dependent Variable: Frequency of crashes

Data: Data.Iowa.gov

Methods: Pivot tables, bar graphs, and Chi Square tests of independence

Conclusion: If an accident occurs in Iowa City, a resulting injury is most common during these conditions: in the afternoon, on a Friday, or in the summer.

Problem Framing

➤ **Context**

- Car crashes can happen to anyone who drives, and they can have very large financial impacts as well as the potential to cause serious harm or death

➤ **Motivation**

- We want to reduce injuries and deaths by discovering common factors in harmful accidents

Problem Statement

- We want to determine how different road conditions as well as timing factors relate to the frequency of crashes

Analytics Problem Framing

➤ **Through analyzing the data we will gain understanding regarding these questions:**

- Is whether the driver was under the influence of drugs/alcohol independent of the severity of injury reported in the crash?
- Is the likelihood of a crash causing injury independent of the day of the week the crash occurs?

➤ **Assumptions:**

- The data time frame is from 1/01/2017 to 12/31/2017
- All the accidents in the data set are a comprehensive list of every accident that occurred in Iowa City during this time frame
- The time the accident occurred is the same time as it was recorded
- All variables of the accidents have been categorized correctly
- Refusal to conduct a field sobriety test is inconclusive as to whether the driver was under the influence and will be treated as if they were sober

Hypothesis

Ho: The conditions of a crash are independent from the frequency of crashes occurring in Iowa City

Ha: The frequency of crashes in Iowa City is dependent to the conditions involved in the crash

Data

Taken from iowa.gov

Vehicle Accidents in Iowa by Location (Last Ten Years)

- 557,185 Rows
- 36 Columns

Filtered to only yield accidents during 2017

Refined to only show accidents occurring in Iowa City

2017 Iowa City Car Accident Data

- 1452 Rows
- 36 Columns

Data Cleaning for Accident Report

Eliminated:

- DOT case #
- Law enforcement #
- Work zone
- Property damage
- Rest update
- Accidents that occurred outside of Iowa City from January 1st through December 31st

Methodology

➤ **Pivot tables**

- To find the frequency of accidents in different scenarios
 - Ex. Crashes per month
- Visualized pivot tables results with the use of bar graphs

➤ **Chi Squared Test of Independence**

- To determine if there is a statistically significant relationship between driver or road conditions, and the frequency of crashes
- Dummy variables used

Alternative Data and Methods Considered

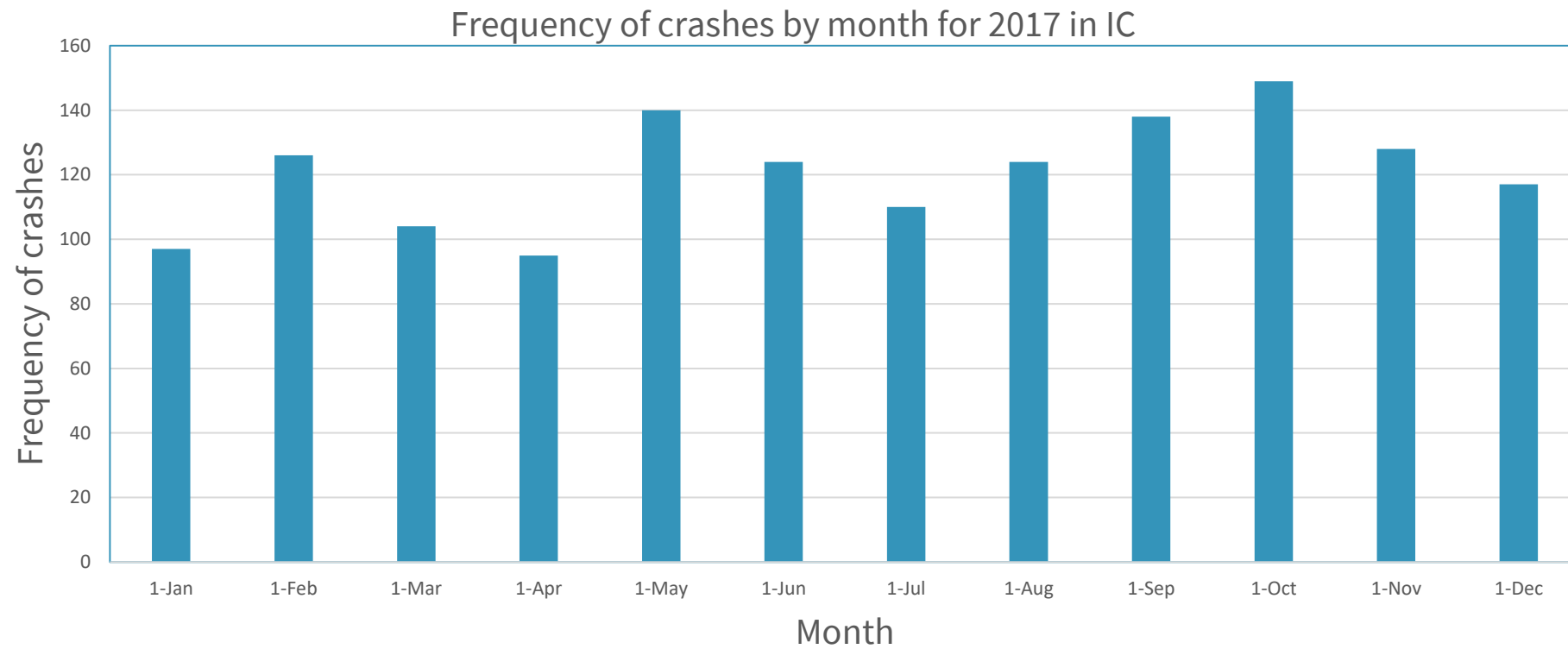
➤ **Alternative Data**

- Accident data for the entire state of Iowa
- Data covering multiple years

➤ **Alternative Methods**

- Linear Regression
- Low correlation coefficient values
 - Chi square test of independence already represents if there are statistical connections between variables.

Frequency of Crashes per Month



Is the quantity of people injured dependent on the month of the accident?

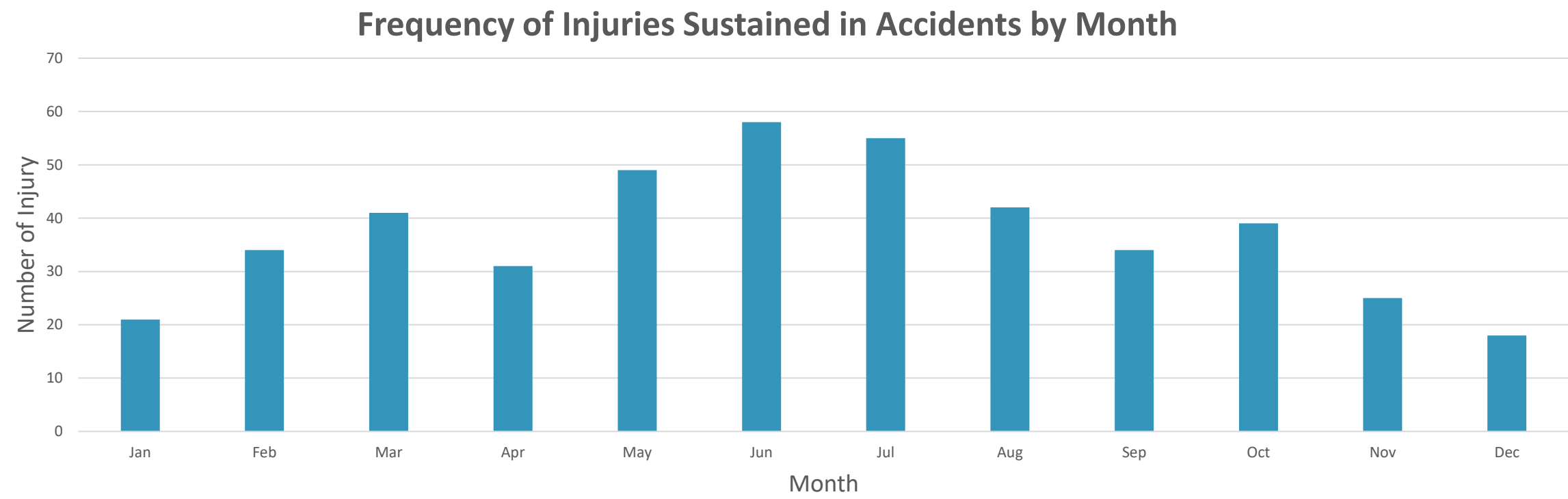
Test Statistic: 65.8051

Critical Value: 0.0182

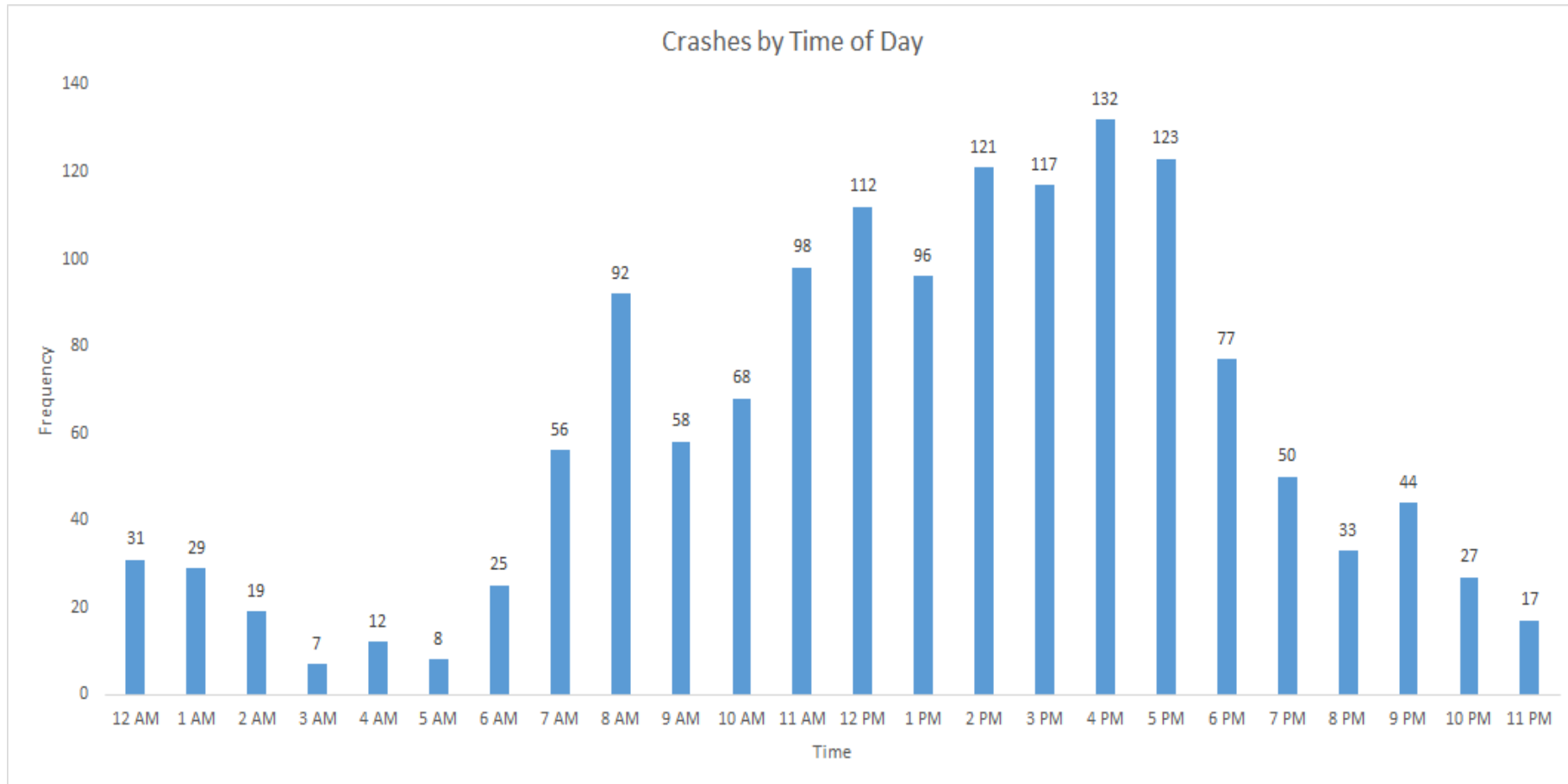
The months of June, July and December stand out

	Number of Injuries Per Accident				
	0	1	2	3	4
Jan	0.369	0.306	1.006	0.534	0.200
Feb	0.435	1.548	0.280	0.135	2.102
Mar	0.212	0.223	0.021	3.554	2.869
Apr	0.075	0.202	0.294	0.523	0.196
May	0.160	0.134	0.848	0.068	0.289
Jun	2.131	4.145	1.614	0.147	2.159
Jul	2.232	2.870	2.625	9.456	0.227
Aug	0.398	2.056	0.003	0.683	0.256
Sep	0.396	1.157	0.016	0.760	0.285
Oct	0.089	0.025	0.217	0.821	0.308
Nov	0.827	0.844	2.046	0.705	0.264
Dec	2.286	5.590	0.696	0.645	0.242

Crashes Resulting in Injuries Are Most Common in the Months of May Through July



From 12:00 PM to 5:00 PM Crashes Are Most Common



Time of Day that Alcohol Related Injuries Occur	Column1	Column2	Column3
12:00 AM	0.707819757	6.327908628	
1 AM	5.955951416	53.24620565	
2 AM	4.577007909	40.9184507	
3 AM	0.099263365	0.887414487	
4 AM	0.180664899	1.615144199	
5 AM	0.022506403	0.201207243	
6 AM	0.101278814	0.905432596	
7 AM	0.005490503	0.049085099	
8 AM	0.270076837	2.414486922	
9 AM	0.131085169	1.171901408	
10 AM	0.131085169	1.171901408	
11 AM	0.258823635	2.3138833	
12 PM	0.337596046	3.018108652	
1 PM	0.254210835	2.272644869	
2 PM	0.122427806	1.094504589	
3 PM	0.472634465	4.225352113	
4 PM	0.450128061	4.024144869	
5 PM	0.191596877	1.71287608	
6 PM	0.092690558	0.828653588	
7 PM	0.558212101	4.990416181	
8 PM	0.005294069	0.047328974	
9 PM	0.255256726	2.281995129	
10 PM	0.266691874	2.384225352	
11 PM	0.054923726	0.491018109	
		Test Statistic	154.0970072
		Critical Value	2.16965E-21

How Does Alcohol Consumption Relate to What Time Injuries are Sustained During a Car Crash?

Are Injuries Dependent on the Day of the Week?

Weekday	Number of Injuries Per Accident					
	0	1	2	3	4	
Friday	0.239649	1.630656	0.210902	0.147109	0.319606	
Saturday	0.180729	0.185655	0.489196	0.003775	0.39876	
Sunday	0.102268	0.398962	0.149773	0.683196	0.256198	
Monday	0.1053	1.047825	0.00447	2.842568	0.712383	
Tuesday	0.776425	4.427195	0.1289	1.068871	0.400826	
Wednesday	0.383066	0.698183	0.460951	0.541214	0.448347	
Thursday	0.162263	0.178922	0.288225	1.289256	0.551847	
					Test Statistic	21.91347
					Critical Value	0.584431

Are Injuries Dependent on Road Conditions?

Road Condition	Number of Injuries Per Accident					
	0	1	2	3	4	
Dry	0.126223826	0.214526272	0.391503823	0.339988672	0.132752761	
Ice/frost	0.081432128	0.013312601	0.661157025	0.08815427	0.033057851	
Mud, dirt	0.082798797	0.200413223	0.041322314	0.005509642	0.002066116	
Not Reported	0.993585564	2.404958678	0.495867769	0.066115702	0.024793388	
Oil	0.750688705	3.190103945	0.041322314	0.005509642	0.002066116	
Sand	0.750688705	3.190103945	0.041322314	0.005509642	0.002066116	
Slush	0.167432456	0.400826446	10.18264463	0.011019284	0.004132231	
Snow	1.029268369	3.402654983	0.005149396	0.143250689	0.053719008	
Unknown	0.002887356	0.195485218	1.033057851	0.137741047	0.051652893	
Wet	0.091317425	0.135675229	1.06010125	3.604697619	0.394628099	
					Test Statistic	36.48626334
					Critical Value	0.803866332

Methodology Weakness

- Lack of frequency of driving fatalities makes it hard to model what conditions most likely result in death
- Possible unreported accidents that not apart of the data
- Human error is a large factor that can not be measured
- We can not accurately forecast future injuries caused by accidents given our methods

Conclusion

- We reject the null hypothesis and accept the alternative hypothesis
- The high chi square test statistics received indicates that the condition of summer season, result in a higher likelihood of an injury if a crash occurs
- The highest frequency of injuries is on Friday, however, the percent of crashes resulting in injury is highest on Tuesday
- Accidents on snowy roads are much less likely to result in injury, but accidents on wet roads are more likely to result in injury
- Drinking and driving is most likely to result in a car crash injury during the early morning hours of the day