Modeling Colorado Car Accident Injuries

Brady Lamson - Fall 2024

Repository Link

Car Accidents in Colorado

Some background

- From 2021-2023 we saw around 100k car accidents per year in Colorado.
- Of those, about a quarter resulted in some degree of injury. ($\approx 25k$).
- Approximately 700 people die per year in car accidents in this state and around 35k are injured.
- Though our metrics don’t rank us too poorly in the county, any understanding we can get about car related injuries can help save lives.

Research Question

What are we doing here?

- On a county level, what influences injury rates in car accidents?
- Is it about the specific conditions of the crash itself, like the weather?
- Or is there a socioeconomic side as well? Do poorer counties see more injuries?
- What about the time we spend in our cars? Do longer commutes put us at risk?

Data Tables

Sources and Relevant Info

- This project leverages four separate tables.
- The primary table includes a plethora of information related to individual car accidents in Colorado.
- The rest serve as additional demographic information for Colorado counties to be included for modeling.

Table Description	Source
Colorado Car Accident Data	Colorado Department of Transportation
County Population	Colorado Department of Local Affairs
County Median Household Income	National Institute on Minority Health and Health Disparities
County Average Commute Time	Opendatasoft (data pulled from the Census Bureau)

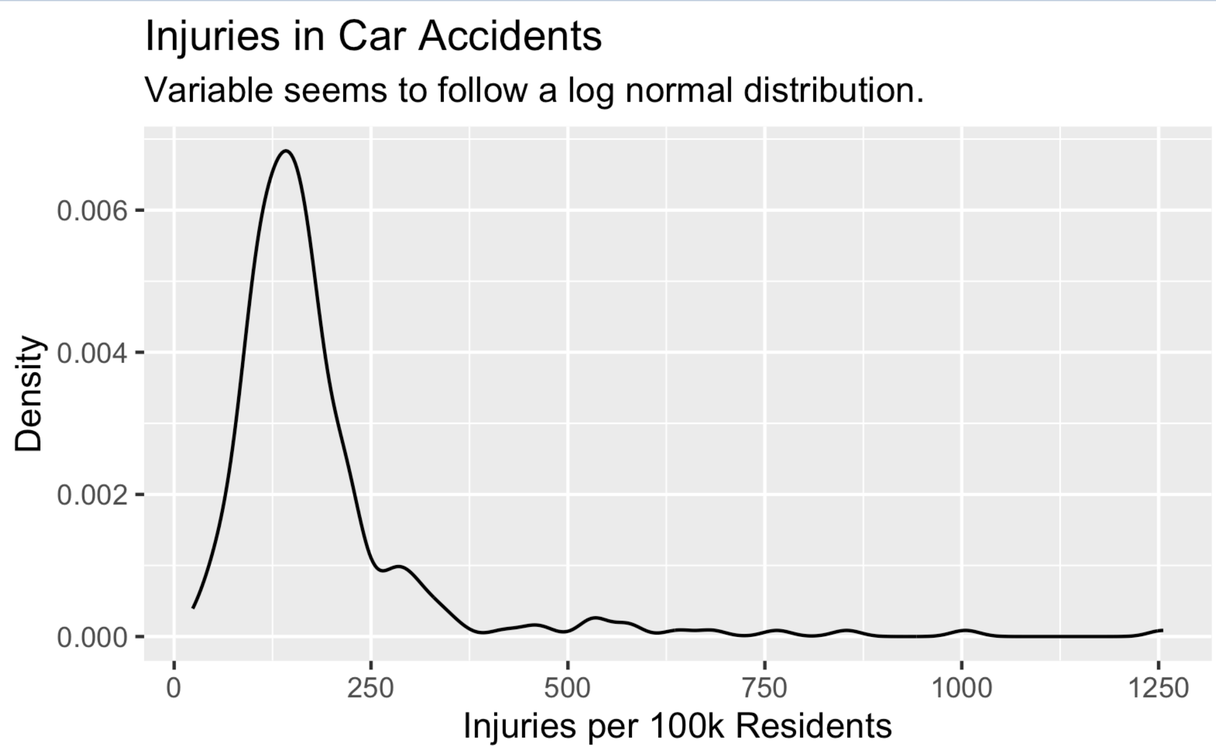
Data Aggregation and Overview

- Approximately 300k rows aggregated down to 264.
- Each row represents a Colorado county in a specific season.
- Values averaged over 3 years of data from 2021-2023.

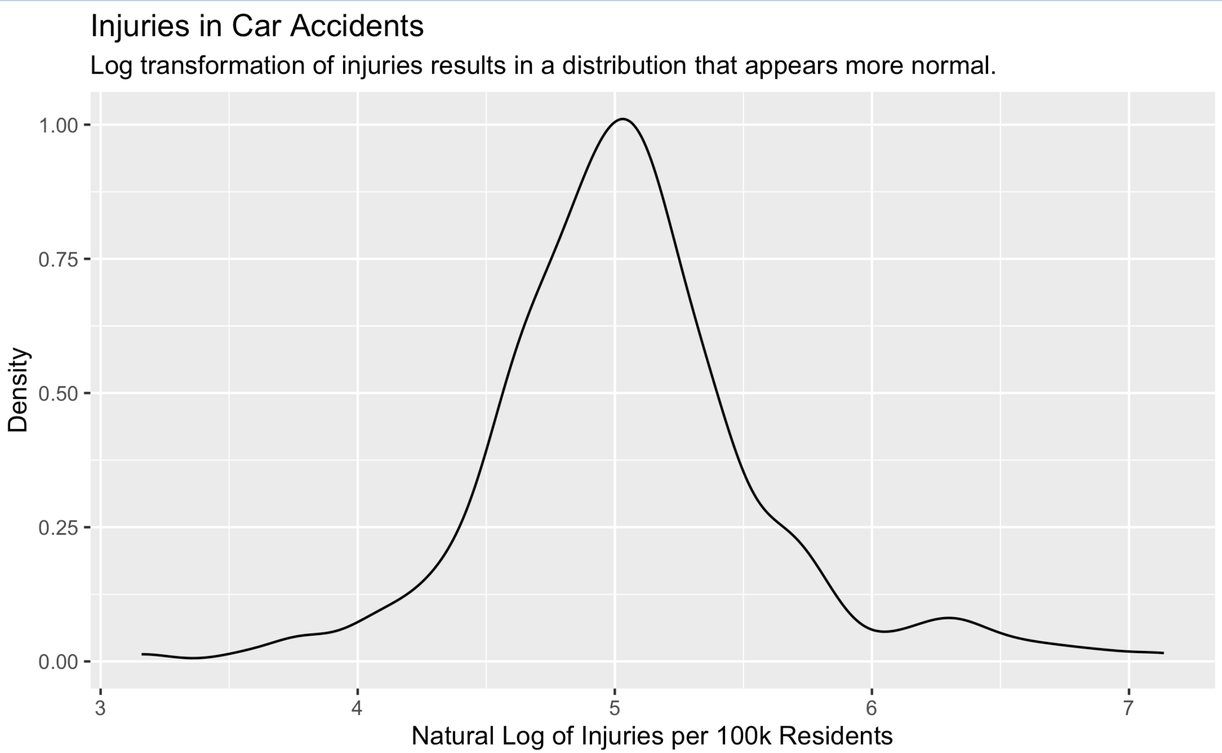
Variable	Data Type	Description
county	string	Name of Colorado county
season	factor	Spring, Summer, Fall, Winter
deaths	continuous numeric	# of deaths per 100k residents
injuries	continuous numeric	# of injuries per 100k residents
bad weather accidents	continuous numeric	# of accidents in poor weather per 100k residents
median income	continuous numeric	Median household income for county residents
mean commuting time	continuous numeric	Mean commuting time for county residents (minutes)

Response Distribution

A big component of this model is a log transformation on the injuries response variable. Notice the drastic shift in distribution.



Regular Distribution



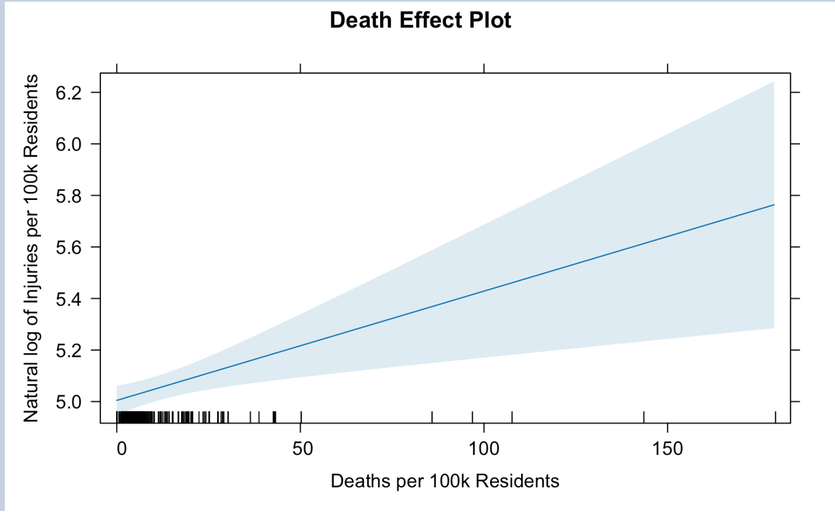
Transformed Distribution

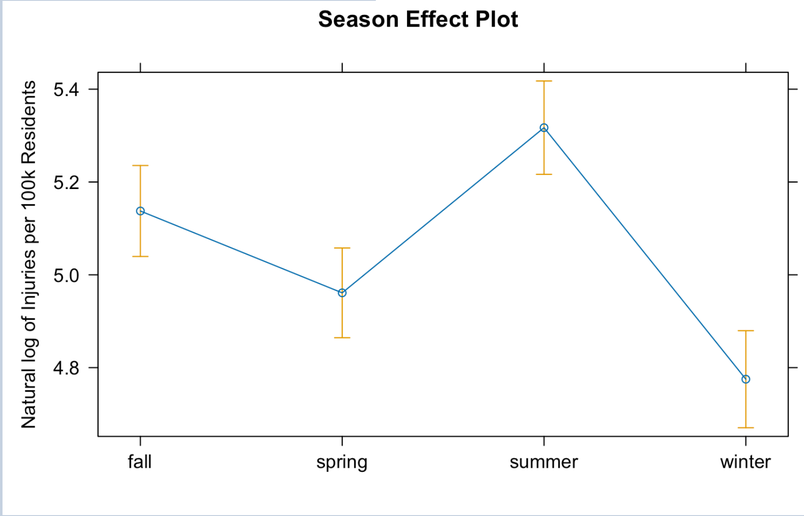
Bivariate Relationships

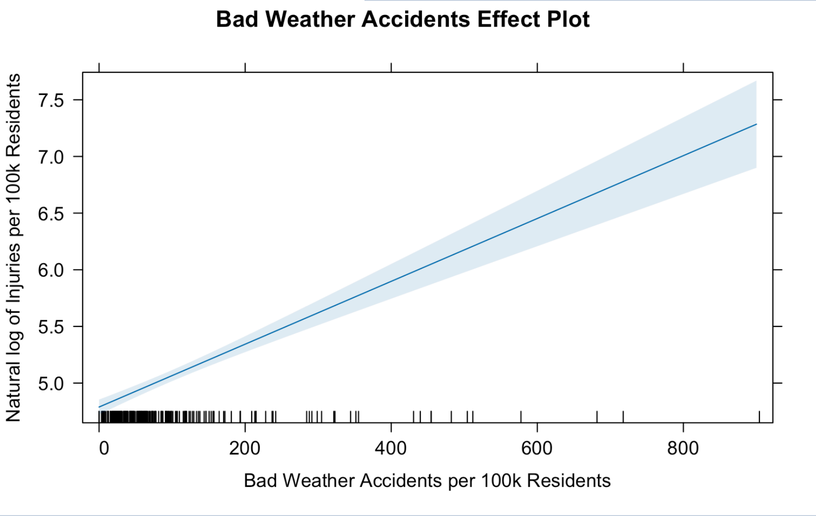
- Seemingly no linear relationship for commute time and income vs injuries.
- Deaths and bad weather accidents appear to have a linear relationship with injuries.

Final Model: Injuries as Response

Variable	Relationship
Deaths	Positive
Bad Weather Accidents	Positive
Season: Spring	Negative compared to Fall
Season: Summer	Positive compared to Fall
Season: Winter	Negative compared to Fall

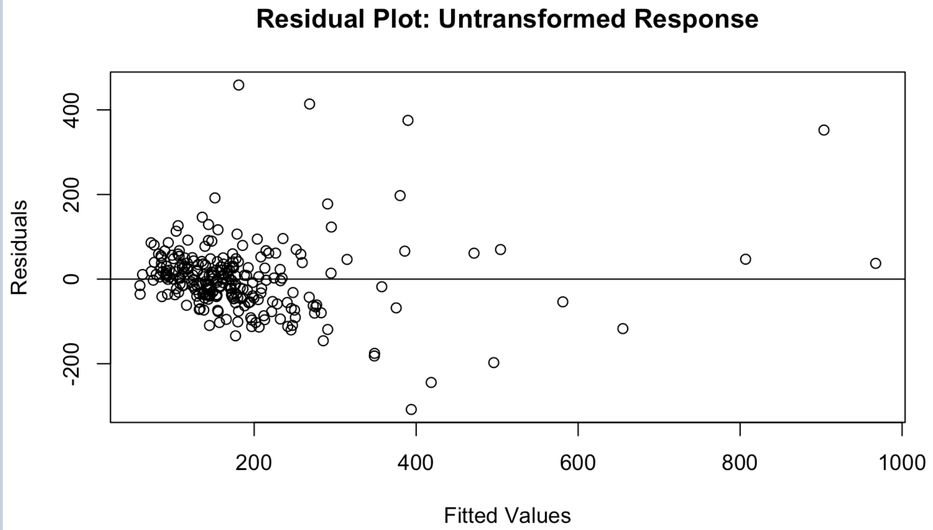


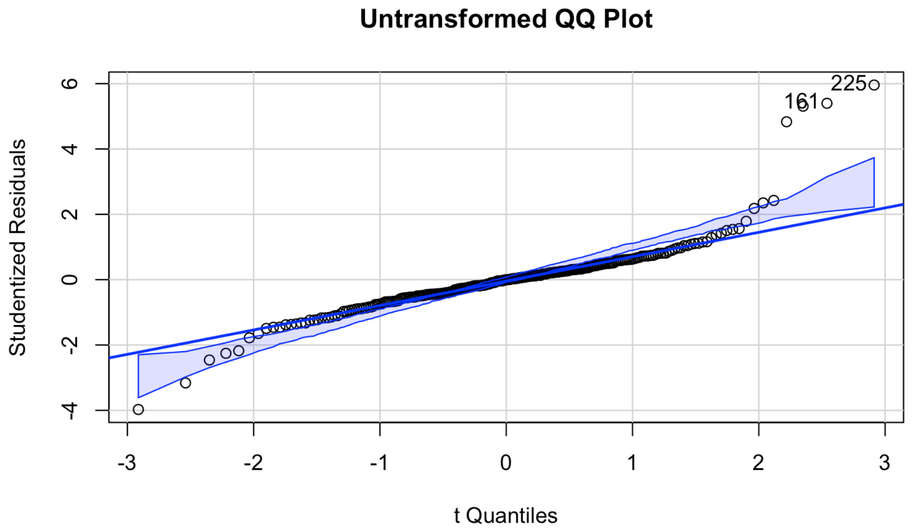




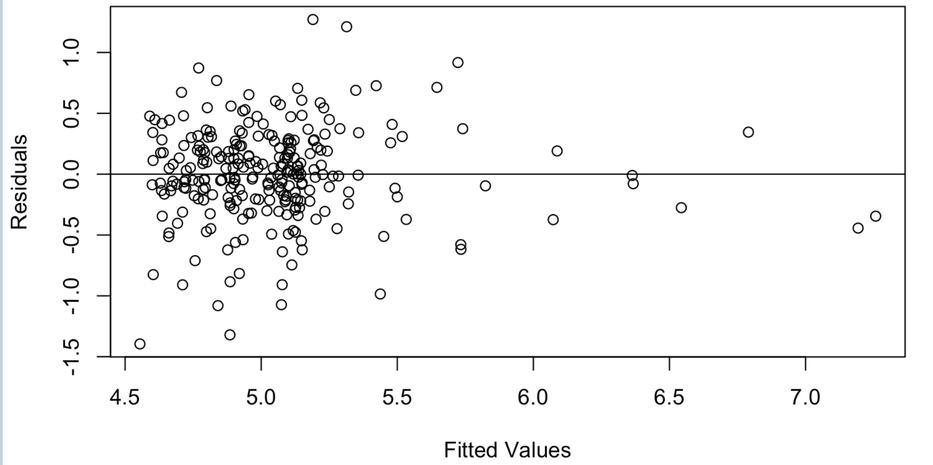
Structural Checks

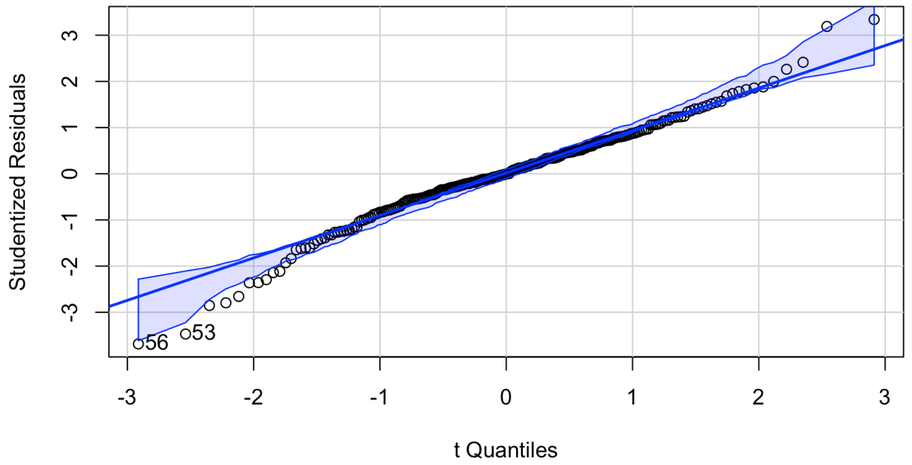
Untransformed





Log Transformed





Limitations

And influential points

- As values like injuries and deaths are per-capita, extremely small counties (< 2000 residents) have vastly inflated values that throw off the model.
- These values don't impact things to a significant degree, however it's likely the smaller counties would benefit from a separate model or approach entirely.
- Only about 6 of these counties overall.

Example Small Counties

County	Season	Population	Injuries per 100k per year	Injuries per year
Mineral	Fall	929	1255.8	11.7
Hinsdale	Spring	776	43.0	0.67

Injuries Summary

Min	1st Quartile	Median	3rd Quartile	Max
0	116.09	150.63	197.19	1255.83

Conclusion:

Interpreting Results

- Summer appears to have the largest influence on the number of accidents resulting in injuries.
- Though deaths and bad weather accidents have a positive relationship with injuries, it’s a lot smaller than I had anticipated.
- Median household income appears to have no significant impact on the number of injuries a county sees in a given season.
- Mean commute time as well doesn’t appear to have a linear relationship with injuries which I found surprising.

Conclusion:

Real World Implications and future research

- The main car accident table contains very specific information down to the street an accident occurred on, future examination of injury prone streets and intersections is recommended.

- Do the locations of injuries differ depending on season? If so, research into variable speed limits or other such traffic controlling measures based on season may be worth investigating.