Analysis of Traffic Fatality Records

This manuscript (permalink) was automatically generated from uiceds/project-team-front-row@3d6205b on October 24, 2024.

Authors

- **Justin Rebholz** ✉
  · jrebholz12
  Department of Civil and Environmental Engineering, University of Illinois

- **Cameron Kimber** ✉
  · ckimber2
  Department of Civil and Environmental Engineering, University of Illinois

- **Hannah Daggett** ✉
  · hed2
  Department of Civil and Environmental Engineering, University of Illinois

- **Riley Kelch** ✉
  · rileykelch
  Department of Civil and Environmental Engineering, University of Illinois

✉ — Correspondence possible via GitHub Issues or email to Justin Rebholz <rebholz4@illinois.edu>, Cameron Kimber <ckimber2@illinois.edu>, Hannah Daggett <hed2@illinois.edu>, Riley Kelch <rjkelch2@illinois.edu>.

Abstract

Description

The dataset that will be used for this project is the Fatality Analysis Reporting System created by the National Highway Safety Administration. The data will be obtained from the NHTSA's FARS database, which is publicly accessible.The FARS dataset is available in the CSV format. The specific dataset that our project will be focused on is labeled "accidents" and includes 32K+ instances and 52 columns. The columns descriptions are described in the Fatality Analysis Reporting System (FARS) Analytical User's Manual 1975-2015. Columns are described in the below table:

Column	Description
ARR_HOUR	This data element records the hour when emergency services arrived at the scene.
ARR_MIN	This data element records the minute when emergency services arrived at the scene.
CF1, CF2, CF3	These data elements record contributing factors to the crash, such as driver behaviors or environmental conditions.
CITY	This data element identifies the city in which the crash occurred.
COUNTY	This data element identifies the county in which the crash occurred.
DAY	This data element records the day of the month on which the crash occurred.
DAY_WEEK	This data element identifies the day of the week on which the crash occurred.
DRUNK_DR	This data element records whether a driver involved in the crash was suspected of drinking alcohol.
FATALS	This data element records the number of fatalities resulting from the crash.
FUNC_SYS	This data element identifies the functional classification of the trafficway segment where the crash occurred.
HARM_EV	This data element records the first harmful event that occurred in the crash sequence.
HOSP_HR	This data element records the hour when the injured were admitted to the hospital.
HOSP_MN	This data element records the minute when the injured were admitted to the hospital.
HOUR	This data element records the hour when the crash occurred.
LATITUDE	This data element identifies the location of the crash using latitude coordinates.
LGT_COND	This data element identifies the light condition at the time of the crash, such as daylight, dark, or dusk.

Column	Description
LONGITUD	This data element identifies the location of the crash using longitude coordinates.
MAN_COLL	This data element identifies the manner of collision, such as rear-end, head-on, or angle.
MILEPT	This data element records the milepoint nearest to the crash location.
MINUTE	This data element records the minute when the crash occurred.
MONTH	This data element records the month in which the crash occurred.
NHS	This data element identifies whether the crash occurred on a National Highway System (NHS) route.
NOT_HOUR	This data element records the hour when the crash was reported to authorities.
NOT_MIN	This data element records the minute when the crash was reported to authorities.
PEDS	This data element records the number of pedestrians involved in the crash.
PERMVIT	This data element counts the number of persons in motor vehicles in transport (motorists) involved in the crash.
PERNOTMVIT	This data element counts the number of persons not in motor vehicles in transport (non-motorists) involved in the crash.
PERSONS	This data element is a count of the total number of persons involved in the crash.
PVH_INVL	This data element is the number of parked or working vehicles involved in the crash.
RAIL	This data element identifies if the crash involved a rail system or crossing.
RELJCT1	This data element identifies the relationship of the crash to a junction, such as intersection or non-intersection.
RELJCT2	This data element provides additional information about the crash's relationship to the junction.
REL_ROAD	This data element identifies the relationship of the crash to the road, such as on the roadway or off the roadway.
RD_OWNER	This data element identifies the entity responsible for the ownership of the road where the crash occurred.
ROUTE	This data element records the type of route where the crash occurred, such as Interstate, U.S. Highway, or State Highway.
RUR_URB	This data element identifies whether the crash occurred in a rural or urban area.

Column	Description
SCH_BUS	This data element identifies if a school bus was involved in the crash.
SP_JUR	This data element identifies if the crash occurred in a special jurisdiction, such as military or Indian reservations.
STATE	This data element identifies the state in which the crash occurred. The codes are from the General Services Administration's (GSA) publication of worldwide Geographic Location Codes (GLC).
ST_CASE	This data element is the unique case number assigned to each crash. It appears on each data file and is used to merge information from the data files together.
TWAY_ID	This data element identifies the primary trafficway on which the crash occurred.
TWAY_ID2	This data element identifies the secondary trafficway associated with the crash, if applicable.
TYP_INT	This data element identifies the type of intersection involved in the crash, if applicable.
VE_FORMS	This data element is a count of all vehicle forms applicable to this crash.
VE_TOTAL	This data element is the number of contact motor vehicles that the officer reported on the PAR as a unit involved in the crash.
WEATHER	This data element identifies additional weather factors at the time of the crash.
WEATHER1	This data element records the primary weather condition at the time of the crash.
WEATHER2	This data element records the secondary weather condition at the time of the crash.
WRK_ZONE	This data element identifies if the crash occurred in a work zone.
YEAR	This data element records the year in which the crash occurred.

Link: https://www.kaggle.com/datasets/nhtsa/2015-traffic-fatalities

Plan and Proposal

Using the FARS dataset we aim to understand the trends in traffic fatalities in a given year and what factors are affecting those trends. We will also look at how the different variables play a role in the severity of the accident and identify geographic regions that are more prone to accidents. The trends in traffic fatalities found through this project can be used to inform policy makers and ultimately decrease the number of traffic fatalities.

This manuscript is a template (aka "rootstock") for Manubot, a tool for writing scholarly manuscripts. Use this template as a starting point for your manuscript.

The rest of this document is a full list of formatting elements/features supported by Manubot. Compare the input (`.md` files in the `/content` directory) to the output you see below.

Basic formatting

Bold text

Semi-bold text

Italic text

Combined *italics and **bold***

~~Strikethrough~~

1. Ordered list item
2. Ordered list item
    a. Sub-item
    b. Sub-item
        i. Sub-sub-item
3. Ordered list item
    a. Sub-item

- List item
- List item
- List item

subscript: H_2O is a liquid

superscript: 2^{10} is 1024.

unicode superscripts^{0123456789}

unicode subscripts$_{0123456789}$

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to editing and version control.

Line break without starting a new paragraph by putting
two spaces at end of line.

Document organization

Document section headings:

Heading 1

Heading 2

Heading 3

Heading 4

Heading 5

Heading 6

A heading centered on its own printed page

Horizontal rule:

`Heading 1`'s are recommended to be reserved for the title of the manuscript.

`Heading 2`'s are recommended for broad sections such as *Abstract*, *Methods*, *Conclusion*, etc.

`Heading 3`'s and `Heading 4`'s are recommended for sub-sections.

Links

Bare URL link: https://manubot.org

Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah

Link with text

Link with hover text

Link by reference

Citations

Citation by DOI [1].

Citation by PubMed Central ID [2].

Citation by PubMed ID [3].

Citation by Wikidata ID [4].

Citation by ISBN [5].

Citation by URL [6].

Citation by alias [7].

Multiple citations can be put inside the same set of brackets [1,5,7]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [2,3,7,8].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

Referencing figures, tables, equations

Figure 1

Figure 2

Figure 3

Figure 4

Table 1

Equation 1

Equation 2

Quotes and code

> Quoted text

> Quoted block of text
>
> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.

Code `in the middle` of normal text, aka `inline code`.

Code block with Python syntax highlighting:

```
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskih-
        konstruktsiyah"
```

Code block with no syntax highlighting:

```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

Figures

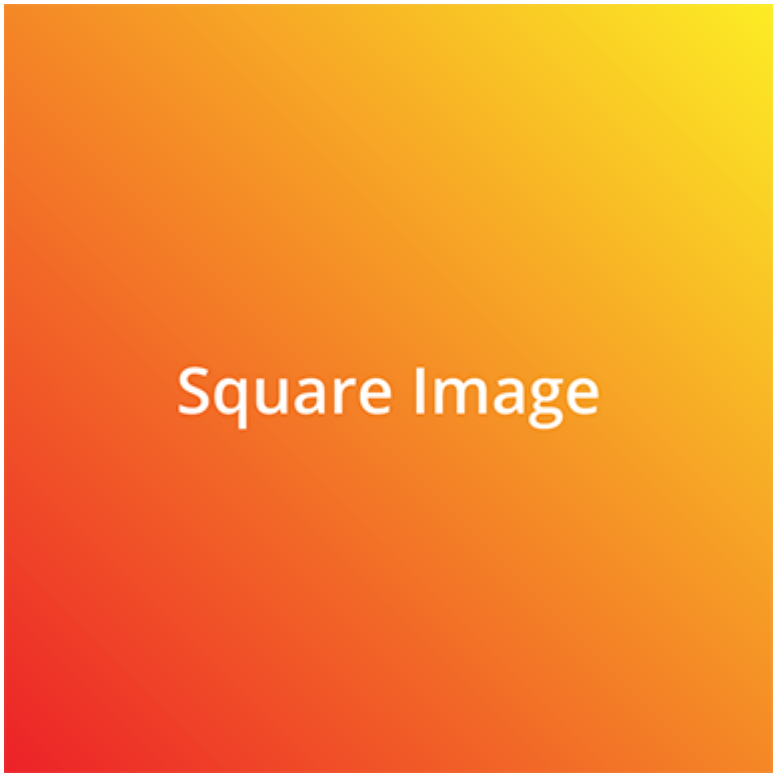

Figure 1: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.

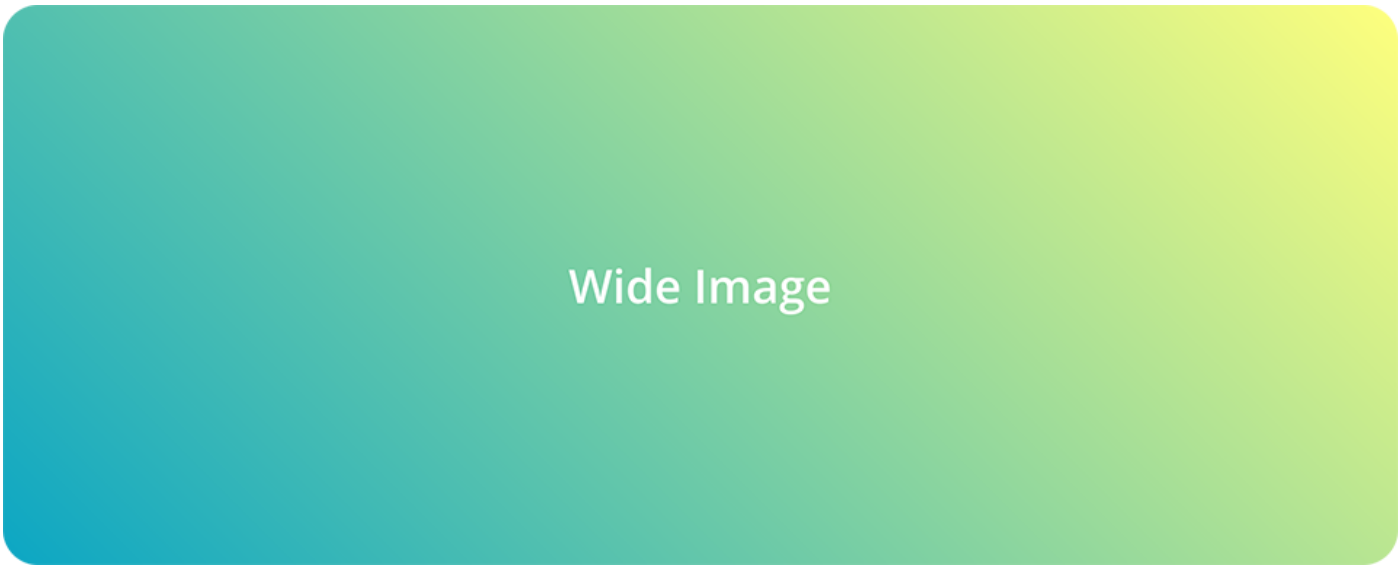

Figure 2: An image too wide to fit within page at full size. Loaded from a specific (hashed) version of the image on GitHub.

Figure 3: A tall image with a specified height. Loaded from a specific (hashed) version of the image on GitHub.

Figure 4: A vector `.svg` image loaded from GitHub. The parameter `sanitize=true` is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image. Note that if you want to export to Word (`.docx`), you need to download the image and reference it locally (e.g. `content/images/vector.svg`) instead of using a URL.

Tables

Table 1: A table with a top caption and specified relative column widths.

Bowling Scores	**Jane**	**John**	**Alice**	**Bob**
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

Table 2: A table too wide to fit within page.

	Digits 1-33	**Digits 34-66**	**Digits 67-99**	**Ref.**
pi	3.14159265358979323 846264338327950	28841971693993751 0582097494459230	78164062862089986 2803482534211706	`piday.org`
e	2.71828182845904523 536028747135266	24977572470936999 5957496696762772	40766303535475945 7138217852516642	`nasa.gov`

Table 3: A table with merged cells using the `attributes` plugin.

	Colors	
Size	**Text Color**	**Background Color**
big	blue	orange
small	black	white

Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{1}$$

An equation too long to fit within page:

$$\begin{aligned} x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t \\ + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \end{aligned} \tag{2}$$

Special

⚠ **WARNING** *The following features are only supported and intended for* `.html` *and* `.pdf` *exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as* `.docx`*.*

LINK STYLED AS A BUTTON

Adding arbitrary HTML attributes to an element using Pandoc's attribute syntax:

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot `attributes` plugin (more flexible than Pandoc's method in terms of which elements you can add attributes to):

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Available background colors for text, images, code, banners, etc:

`white` `lightgrey` `grey` `darkgrey` `black` `lightred` `lightyellow` `lightgreen` `lightblue` `lightpurple` `red` `orange` `yellow` `green` `blue` `purple`

Using the Font Awesome icon set:

✔ ? ★ 🔔 ⊗ •••

Light Grey Banner
useful for *general information* - manubot.org

Blue Banner
useful for *important information* - manubot.org

Light Red Banner
useful for *warnings* - manubot.org

Analysis

Exploratory Data Analysis

```
#Exploritory data analysis

#plot number of accidents per day over the course of the year

#time of day of drunk driving accidents

#drunk driving fatalities per state

#plot the locations of accidents

```

```
#Predictive Modeling

#predict future accident locations: clustering algorithm to determine hotspots and
machine learning to predict future locations of accidents
```

Figure 5: Test Image

Predictive Modeling

References

National Highway Traffic Safety Administration. “2015 Traffic Fatalities.” Kaggle, https://www.kaggle.com/datasets/nhtsa/2015-traffic-fatalities. Accessed 24 Oct. 2024.

National Highway Traffic Safety Administration. Fatality Analysis Reporting System (FARS) Analytical User’s Manual 1975-2015. U.S. Department of Transportation, Aug. 2016.

1. **Sci-Hub provides access to nearly all scholarly literature**
   Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene
   eLife (2018-03-01) https://doi.org/ckcj
   DOI: 10.7554/elife.32822 · PMID: 29424689 · PMCID: PMC5832410

2. **Reproducibility of computational workflows is automated using continuous analysis**
   Brett K Beaulieu-Jones, Casey S Greene
   Nature biotechnology (2017-04) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103790/
   DOI: 10.1038/nbt.3780 · PMID: 28288103 · PMCID: PMC6103790

3. **Bitcoin for the biological literature.**
   Douglas Heaven
   Nature (2019-02) https://www.ncbi.nlm.nih.gov/pubmed/30718888
   DOI: 10.1038/d41586-019-00447-9 · PMID: 30718888

4. **Plan S: Accelerating the transition to full and immediate Open Access to scientific publications**
   cOAlition S
   (2018-09-04) https://www.wikidata.org/wiki/Q56458321

5. **Open access**
   Peter Suber
   MIT Press (2012)
   ISBN: 9780262517638

6. **Open collaborative writing with Manubot**
   Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter
   Manubot (2020-05-25) https://greenelab.github.io/meta-review/

7. **Opportunities and obstacles for deep learning in biology and medicine**
   Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, … Casey S Greene
   Journal of The Royal Society Interface (2018-04) https://doi.org/gddkhn
   DOI: 10.1098/rsif.2017.0387 · PMID: 29618526 · PMCID: PMC5938574

8. **Open collaborative writing with Manubot**
   Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter
   PLOS Computational Biology (2019-06-24) https://doi.org/c7np
   DOI: 10.1371/journal.pcbi.1007128 · PMID: 31233491 · PMCID: PMC6611653