Australian road crash data analysis

2013 - 2023

project 3 - group 7

prepared by Praveen, Lee, Jesslyn, Athira Reji
for UWA Data Analytics Bootcamp 2023

team.

Praveen R
Lee Armstrong
Jesslyn Lengkong
Athira Reji

project overview

64% of surveyed Australians over 18 with driver's licenses have experienced at least one car crash. With a total of 39,755 people hospitalised due to car crashes in 2018-19, we look into some of the available data to review the demographics of road fatalities.

To further look at the statistics of road crashes and road fatalities in Australia, the Australian Road Deaths Database was analysed and visualisations were prepared for a more complete understanding of the demographics involved in road crashes and fatalities.

To understand the data better an AARD Dictionary was included to list all values that are included in the database

** Note dataset only provided data until September 2023 **

Aim: To create an interactive dashboard that will give the users an ability to perform various types of analysis to gain a better understanding of road crashes in Australia from year 2013 to 2023.

Trend analysis

Examining long-term trends in road fatalities to identify patterns and changes over time. This can help assess the effectiveness of road safety initiatives and interventions.

Geographic analysis

Analyzing crash data by location to identify high-risk regions. This information can be used to prioritize safety improvements and allocate resources effectively.

Demographic analysis

Investigating crash data by factors such as age, gender etc. This analysis can help identify vulnerable road user groups and tailor road safety campaigns accordingly.

Comparative analysis

Comparing road crash data across different states. This can help identify variations in road safety practices and outcomes, leading to the sharing of best practices.

Research Questions

How have overall state fatalities changed over the years from 2013 to 2023?

Are there any noticeable trends or patterns in the yearly fatalities data?

Are there significant variations in fatalities across different states?

What are the road user patterns, and how do they contribute to overall fatalities?

How are fatalities distributed among different age groups?

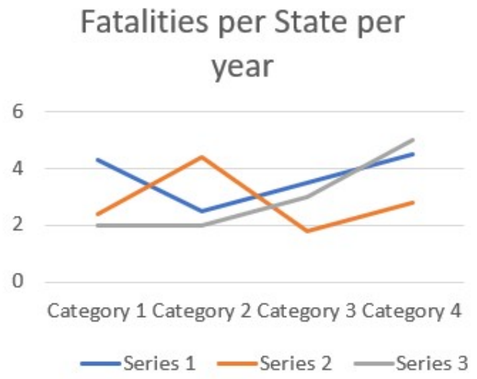
Is there a correlation between age group and the number of fatalities?

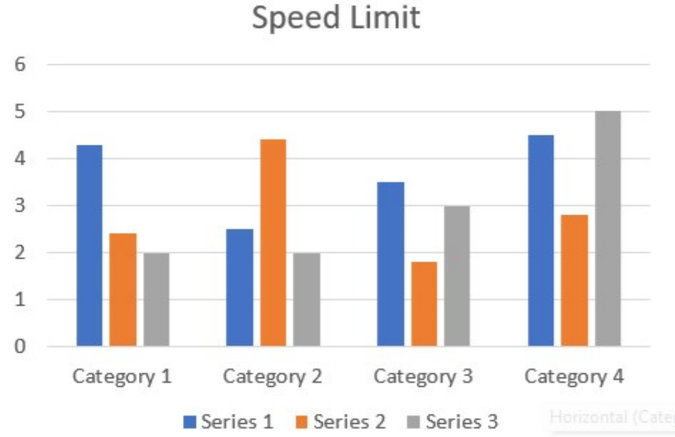
What is the correlation between fatalities and speed limits? Are higher speed limits associated with more severe accidents?

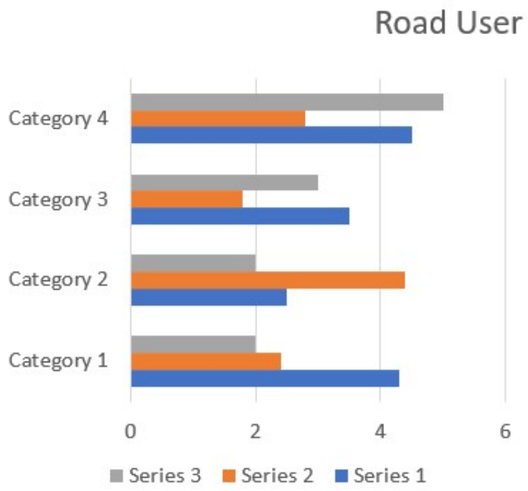
How do fatalities vary across different speed limit categories?

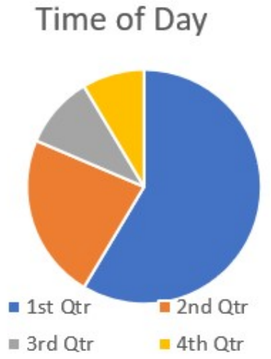
concept stage

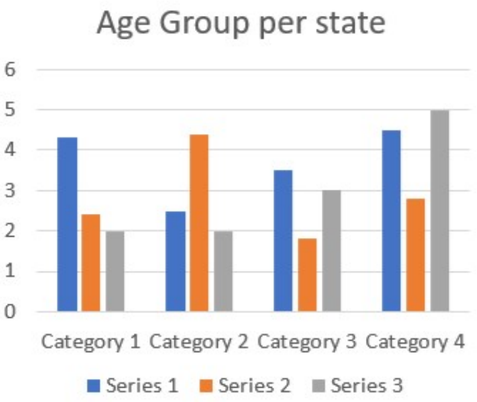
State checkboxes only











EXTRACT, TRANSFORM and LOAD

01  DOWNLOADING DATA

02 DATA CLEANING

03  DATABASE SETUP

04 SQLAlchemy PROJECT DATABASE

05  CREATING ROUTES

EXTRACT, TRANSFORM AND LOAD

01

DOWNLOADING DATA

Data Download from "DATA.GOV.AU" through API

ardd_fatalities.csv, ardd_fatal_crashes.csv

```
data_df = requests.get(
    "https://data.gov.au/data/api/3/action/package_show?id=5b530fb8-526e-4fbf-b0f6-aa24e84e4277").json()

results = data_df['result']['resources']

csv = []

for i in range(0, len(results)):
    csv.append(results[i]['url'])

open('Master_Folder/ardd_fatalities.csv', 'wb').write(requests.get(csv[1]).content)
open('Master_Folder/ardd_fatal_crashes.csv', 'wb').write(requests.get(csv[2]).content)
open('Master_Folder/calendar.csv', 'wb').write(requests.get(csv[4]).content)
wget.download(csv[3], out='Master_Folder/')
wget.download('https://github.com/Elkfox/Australian-Postcode-Data/blob/master/au_postcodes.csv', out = 'Master_Folder/')
```

EXTRACT, TRANSFORM AND LOAD

02

CLEANING DATA

Filling empty columns and -9 values with “Undefined” - Renaming columns - removing unwanted columns - merging data frames

Fatal_Crash_df, Modified_Fatalities_df

```
Fatal_Crash = pd.read_csv('Master_Folder/ardd_fatal_crashes.csv')
Fatalities = pd.read_csv('Master_Folder/ardd_fatalities.csv')

Fatal_Crash_df = Fatal_Crash.fillna({'Time': '00:00', 'Crash Type': 'Undetermined', 'Bus Involvement': 'Undetermined',
                                     'Heavy Rigid Truck Involvement': 'Undetermined',
                                     'Articulated Truck Involvement': 'Undetermined',
                                     'Speed Limit': '-9', 'National Remoteness Areas': 'Undetermined',
                                     'SA4 Name 2021': 'Undetermined',
                                     'National LGA NAme 2021': 'Undetermined', 'National Road Type': 'Undetermined'
Fatalities_df['SA4 Name 2021'] = Fatalities_df['SA4 Name 2021'].replace(['Unknown'], 'Undetermined')
Fatalities_df['National LGA Name 2021'] = Fatalities_df['National LGA Name 2021'].replace(['Unknown'], 'Undetermined')
Fatalities_df['National Road Type'] = Fatalities_df['National Road Type'].replace(['Unknown'], 'Undetermined')

Fatalities_df.to_csv("output/Cleaned_Fatalities_Data.csv", index=False)
Fatalities_df.to_json("output/Cleaned_Fatalities_Data.json", orient="index")

Modified_Fatalities_df = Fatalities_df[['Crash ID', 'Road User', 'Gender', 'Age', 'Age Group']]
```

EXTRACT, TRANSFORM AND LOAD

03

DATABASE SETUP

Database Setup using SQLITE3

Fatalities_Crash

```
conn = sqlite3.connect('output/Master_Database.sqlite')
c = conn.cursor()

Master_Database_df = pd.merge(Fatal_Crash_df, Modified_Fatalities_df, on='Crash ID')
Master_Database_df = Master_Database_df[
    ['Crash ID', 'State', 'Dayweek', 'Day of week', 'Month', 'Year', 'Time', 'Time of Day', 'Speed Limit', 'Crash Type',
     'Number of Fatalities', 'Road User', 'Gender', 'Age', 'Age Group', 'Christmas Period', 'Easter Period',
     'National Road Type', 'National Remoteness Areas', 'SA4 Name 2021', 'National LGA Name 2021',
     'Bus Involvement', 'Heavy Rigid Truck Involvement', 'Articulated Truck Involvement']]

Master_Database_df['Speed Limit'] = Master_Database_df['Speed Limit'].astype('int64')
Master_Database_df['Time'] = pd.to_datetime(Master_Database_df['Time'], format='%H:%M')

Modified_Master_DB = Master_Database_df.set_axis(['Crash_ID', 'State', 'Dayweek', 'Day_of_week', 'Month', 'Year', 'Time', 'Time_of_Day', 'Speed_Limit', 'Crash_Type',
     'Number_of_Fatalities', 'Road_User', 'Gender', 'Age', 'Age_Group', 'Christmas_Period', 'Easter_Period',
     'National_Road_Type', 'National_Remoteness_Areas', 'SA4_Name_2021', 'National_LGA_Name_2021',
     'Bus_Involvement', 'Heavy_Rigid_Truck_Involvement', 'Articulated_Truck_Involvement'], axis = 'columns')

Modified_Master_DB.to_json("output/Master_Database_df.json", orient="index")
Modified_Master_DB.to_sql('Fatalities_Crash', conn, if_exists='replace', index=False)

for row in c.execute('SELECT * FROM Fatalities_Crash'):
    print(row)

conn.close()
```

EXTRACT, TRANSFORM AND LOAD

04 SQLA

SETTING PROJECT DATABASE & JSON FOR VISUALISATION

Creates a web API using Flask, connects to a SQLite database, retrieves specific data, and serves it in JSON format

Visualisation_data_v02.json

```
from sqlalchemy.ext.automap import automap_base
from sqlalchemy.orm import Session
from flask import Flask, jsonify
from sqlalchemy import Column, Integer, String, create_engine, text, DateTime
import pandas as pd
from gevent.pywsgi import WSGIServer
from gevent import monkey

engine = create_engine("sqlite:///output/Master_Database.sqlite", echo=False)
Base = automap_base()

class Fatalities_Crash(Base):
    __tablename__ = "Fatalities_Crash"
```

```
@app.route("/api/v1.0/Data_for_Dashboard")
def data_for_dashboard():
    session = Session(engine)

    dashboard_data = session.execute(text('SELECT * FROM Fatalities_Crash WHERE Year >= 2013')).fetchall()

    session.close()
    Data = (pd.DataFrame(dashboard_data,
                         columns=['Crash_ID', 'State', 'Dayweek', 'Day_of_week', 'Month', 'Year', 'Time',
                                  'Time_of_Day', 'Speed_Limit', 'Crash_Type',
```

EXTRACT, TRANSFORM AND LOAD

05

DATABASE SETUP

Database Setup using SQLITE3

Fatalities_Crash

```python
from flask import Flask, render_template, jsonify, request
import json
import os

app = Flask(__name__)

@app.route('/')
def dashboard():
    return render_template('index.html')

@app.route('/crash_map2')
def leaflet_map_page():
    return render_template('crash_map2.html')
    # return render_template('crash_map4.html')

@app.route('/years', methods=['GET'])
def get_years():
    try:
        file_path = os.path.join(app.static_folder, 'data', 'Visualisation_data_v02.json')
        with open(file_path, 'r') as f:
            data = json.load(f)

        # Extract years and sort them
        years = sorted({str(record['Year']) for record in data.values()})
```

FINAL DATABASE

CREATING GEOJSON

FINAL DATABASE

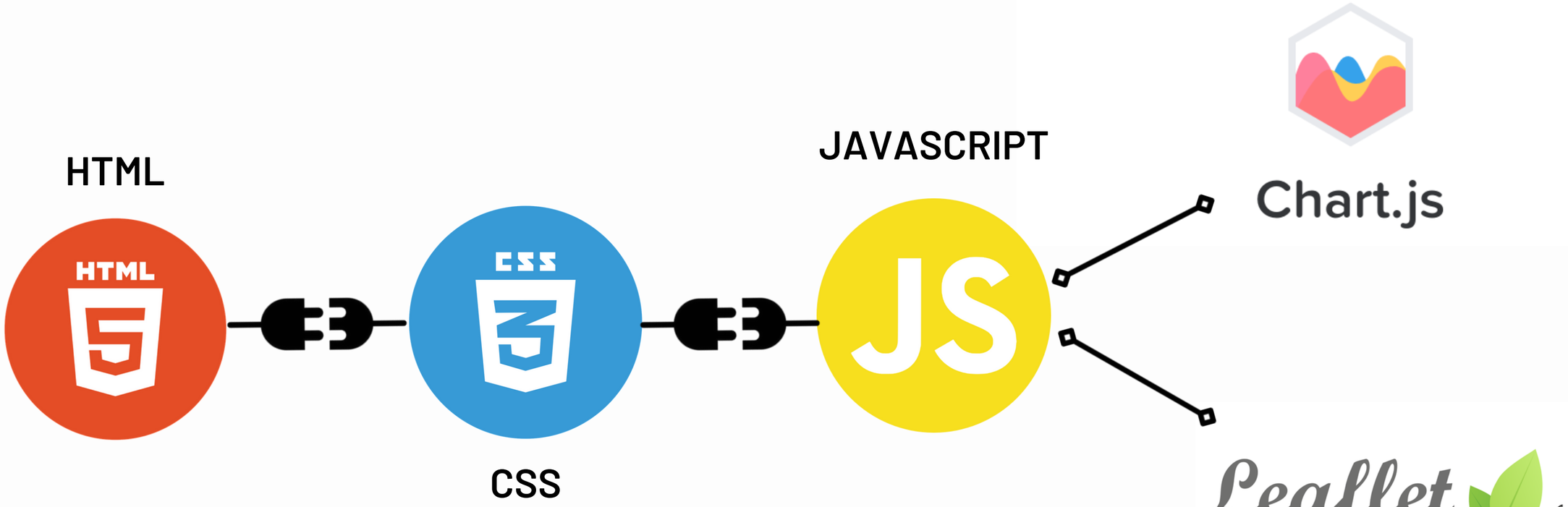
+

au_postcodes

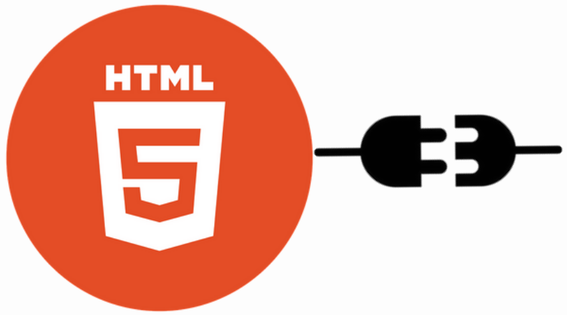
=

output_geojson.geojson

DASHBOARD DEVELOPMENT



DASHBOARD DEVELOPMENT

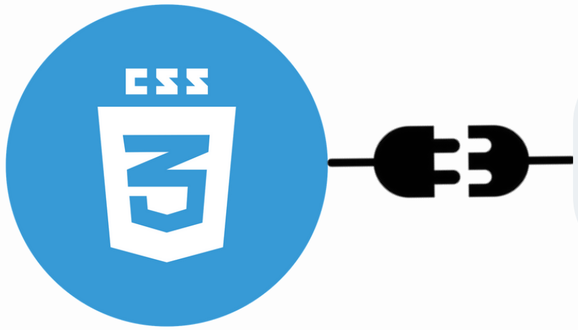


HTML

For creating the HTML, we found a template on W3-School that we worked with. Here we created divs and blocks to create the containers that we required to place the graphs in.
Responsive Web Design Templates (w3schools.com)

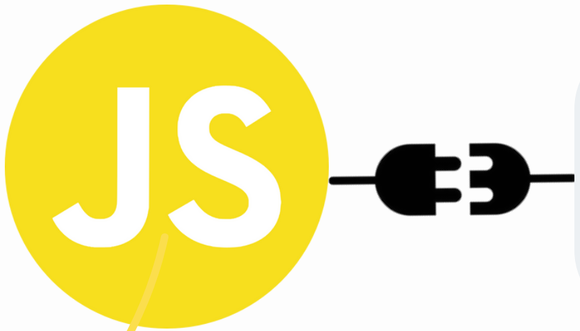
W3 also provided us with a sidebar to work with to include some information about the data source
- About the Project and reasoning for choosing this dashboard
- Data Source- AARD CSVs
- Leaftlet Index for map presentation of fatalities in Australia
- GitHub repository link

DASHBOARD DEVELOPMENT



CSS

styles.css
Styling of tables, fonts, and events for scrolling



JAVASCRIPT

In our logic. js
- Scrolling down with the Div continuing to show as you scroll down the page
- Dropdown box for year selection:
- Functions for creating each chart/table with chart.js Chart.js | Chart.js (chartjs.org)

dashboard snapshot

Smashing Data

Welcome to Smashing Data

This dashboard explores the data of fatalities and crashes around Australia between the years 2013 to 2023

Below are some further information of the dataset and information analysed

Menu

- About the Project
- AARD Data Source
- JSON Data File
- Leaflet Visualisation
- GitHub Repository

Dashboard

About the Project

64% of surveyed Australians over 18 with driver's licenses have experienced at least one car crash. With a total of 39,755 people hospitalised due to car crashes in 2018-19, we look into some of the available data to review the demographics of road fatalities.

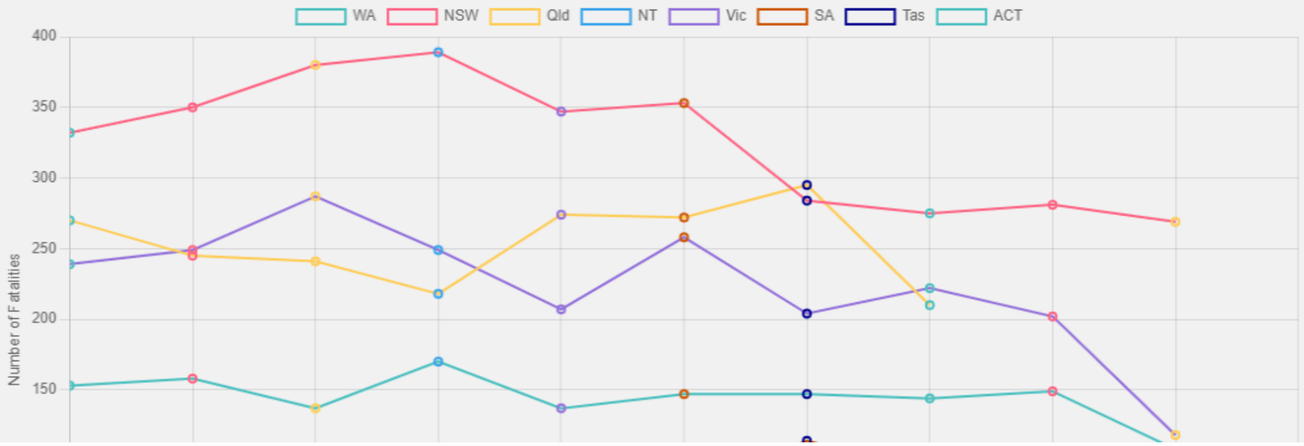
To further look at the statistics of road crashes and road fatalities in Australia, the **Australian Road Deaths Database** was analysed and visualisations were prepared for a more complete understanding of the demographics involved in road crashes and fatalities.

To understand the data better an **AARD Dictionary** was included to list all values that are included in the database

*** Note dataset only provided data until September 2023 ***

Overall State Fatalities from 2013 to 2023

Below represents the fatalities per year per state.
Click the States to remove them from the list for relative comparisons per year.



Summary Table

This is the current state record for road fatalities from 2013 to 2023 as of September 2023

State	Fatalities
NSW	3874
Vic	2579
Qld	2439
WA	1727
SA	1211
NT	469
Tas	425
ACT	87

Conclusions

How have overall state fatalities changed over the years from 2013 to 2023?

Are there any noticeable trends or patterns in the yearly fatalities data?

The fatalities shows a downward trend over the years and the numbers have significantly reduced over the years.

Are there significant variations in fatalities across different states?
NSW has the highest number of fatalities over the last decade and also when studying yearly data, followed by VIC and QLD.

What are the road user patterns, and how do they contribute to overall fatalities?
Drivers are the most impacted during the crash.

How are fatalities distributed among different age groups?

Is there a correlation between age group and the number of fatalities?

40-46yrs age group has the most number of fatalities followed by 25-39yrs, followed by 16-24yrs.

What is the correlation between fatalities and speed limits? Are higher speed limits associated with more severe accidents?

How do fatalities vary across different speed limit categories?

The number of fatalities change with different speed limits. The 60km/hr-80km/hr range has the most number of fatalities.

references

Green Color Codes (html-color.codes)

How to Highlight Table Row on hover mouse Using CSS (tutorialdeep.com)

CSS Table Style (w3schools.com)

Canvas is already in use. Chart with ID '3' must be destroyed before the canvas with ID 'eachStateChart' can be reused - Search (bing.com)

How To Create an On Scroll Fixed Header (w3schools.com)