School of Science, Computing and Engineering Technologies

COS30045

LAB 4.1 Design Studio

Overview

In this lab you will be given a sample data set and asked to identify the different data and attribute types. You will also think about some questions about this data set that might be answered by a visualisation.

ardd\_fatalities\_Jan2020\_0.xlsx (download from Canvas)

Download and review this data set before attempting this exercise.

1 Interpreting the data set

Complete the LAB 4.1 Quiz.

2 Visualisation Design

Think of three questions you would like to answer with that require a data visualistion.

For each data question you will need to consider the following:

Which data attributes (columns) do you need to answer this question?

Do you need to transform any of the data?

Does the data type change when you transform the data? If so how.

Make a sketch of how you think your visualisation might look and add to this document.

Your Question 1: Are there seasonal patterns in road fatalities that correlate with holiday periods and weather conditions?

**Data attributes needed:**

* Date (to extract month and day)
* State (different states have different seasonal patterns)
* Crash\_Type (weather-related vs other factors)
* Potentially Speed\_Limit and National\_Road\_Type (highway travel during holidays)

**Data transformations:**

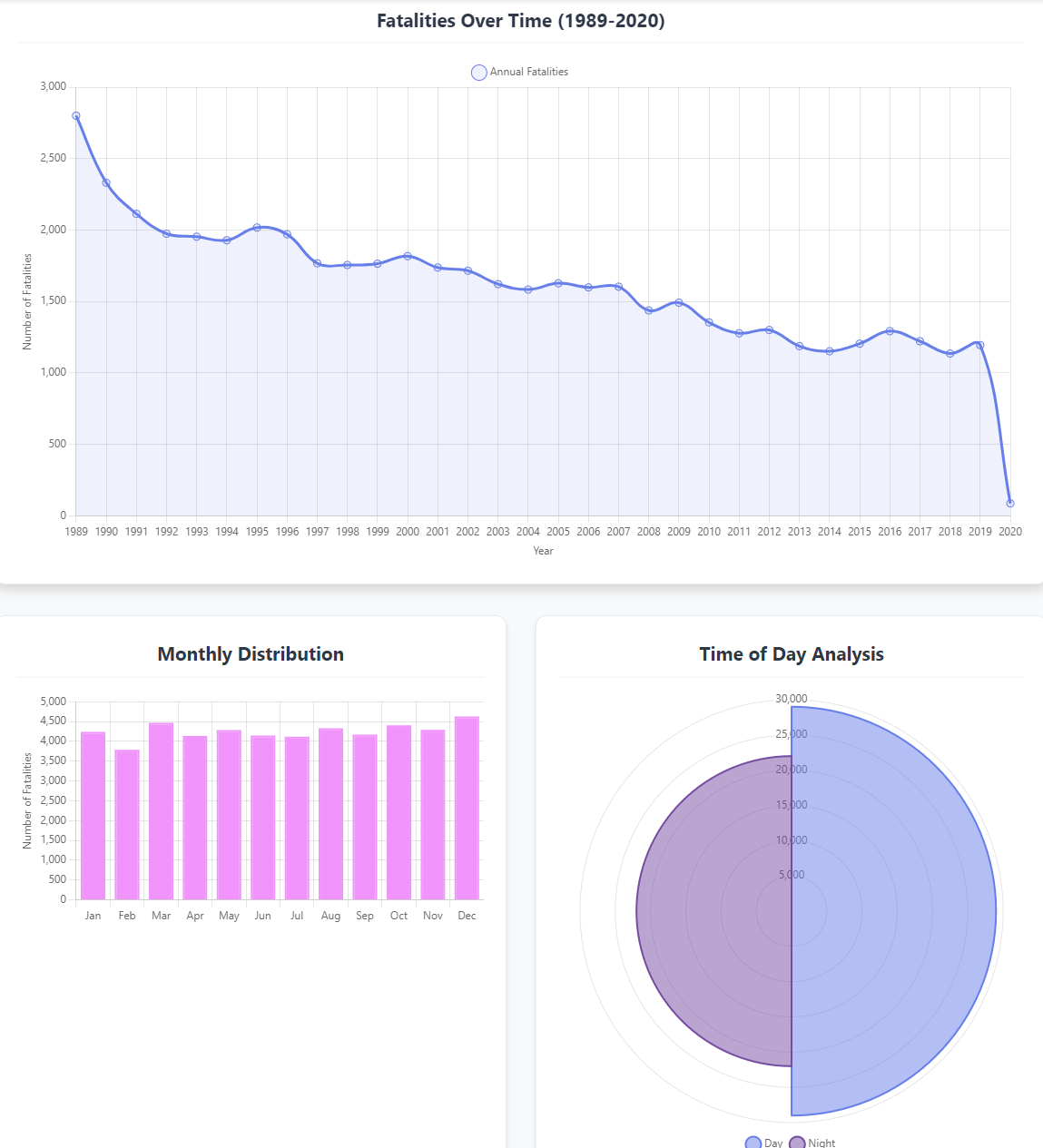
* Convert Date to month/season categories (Summer: Dec-Feb, Autumn: Mar-May, Winter: Jun-Aug, Spring: Sep-Nov)
* Create binary flags for major holiday periods (Christmas/New Year: Dec 20-Jan 10, Easter: varies, School holidays)
* Group crashes by seasonal weather patterns
* Calculate fatality rates per capita for each season to account for population changes

**Data type changes:**

* Date (datetime) → Categorical (season names)
* Date (datetime) → Boolean (is\_holiday\_period)
* Numeric counts → Rates per 100,000 population

**Visualization sketch:**A multi-panel dashboard with:

* Line chart showing monthly fatalities over multiple years with holiday periods highlighted
* Heatmap calendar showing daily fatality patterns
* Grouped bar chart comparing seasonal fatality rates by state
* Small multiples showing crash type distribution by season



Your Question 2: How do demographic risk factors (age and gender) interact with road user types to identify the most vulnerable populations?

**Data attributes needed:**

* Age
* Gender
* Road\_User (Driver, Passenger, Pedestrian, Motorcyclist, Cyclist, etc.)
* Crash\_Type
* Time (to see if risk patterns vary by time of day)

**Data transformations:**

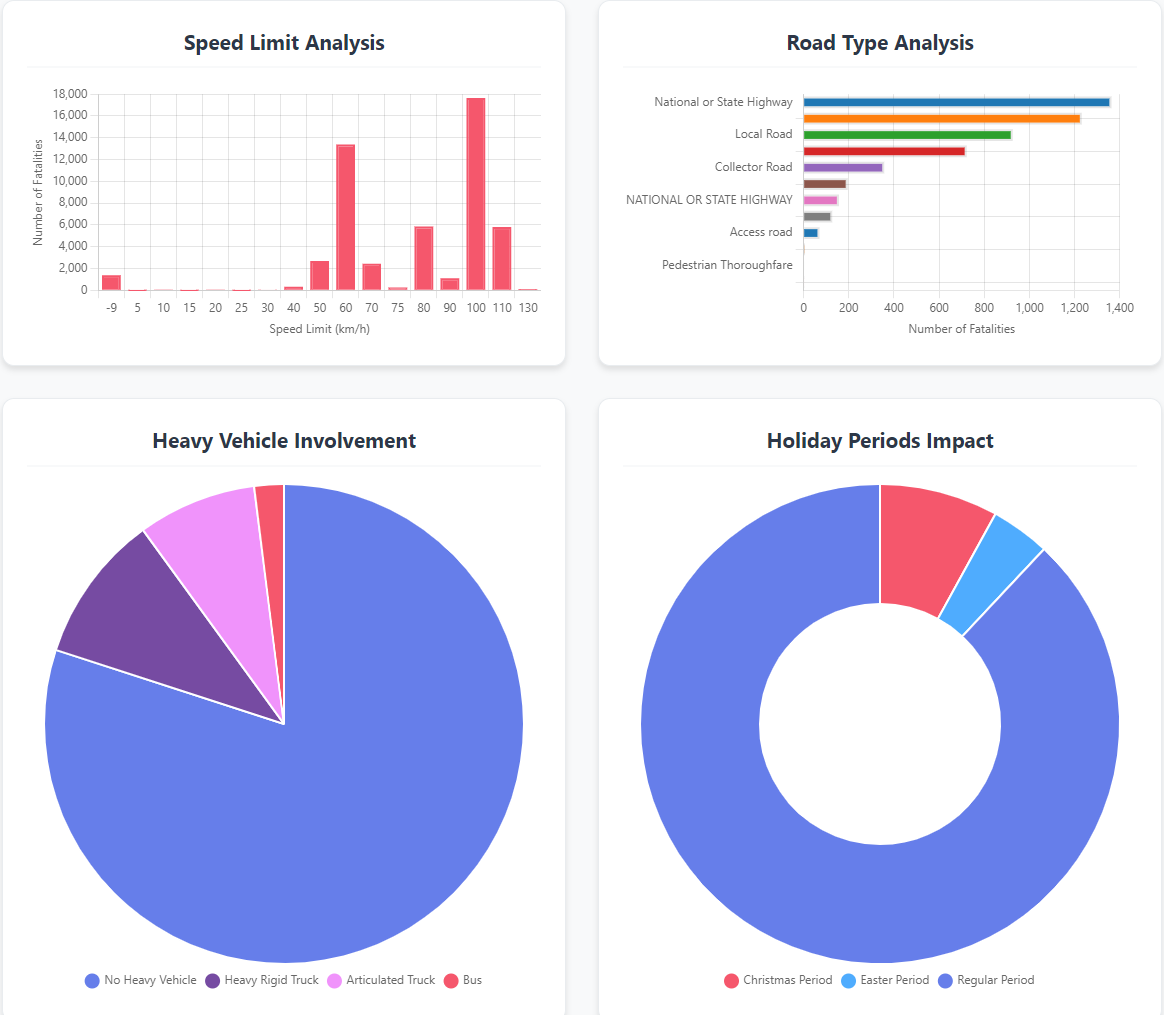
* Group Age into meaningful cohorts (17-25, 26-39, 40-59, 60+)
* Create interaction variables combining age groups, gender, and road user type
* Calculate risk ratios compared to baseline population demographics
* Transform time into peak/off-peak categories

**Data type changes:**

* Age (numeric) → Categorical age groups
* Time (datetime/numeric) → Categorical (Morning Peak, Day, Evening Peak, Night)
* Raw counts → Risk ratios and rates per demographic group

**Visualization sketch:**

* Bubble chart with age groups on x-axis, road user types on y-axis, bubble size = fatalities, color = gender
* Stacked bar chart showing road user type distribution within each age-gender combination
* Risk matrix heatmap showing fatality rates per 100,000 for each demographic-road user combination
* Small multiples radar charts showing risk profiles by time of day



Your Question 3: What is the relationship between road infrastructure (speed limits, road types) and crash severity across different geographic regions?

**Data attributes needed:**

* State
* Speed\_Limit
* National\_Road\_Type (Local Road, State Highway, National Highway, etc.)
* Remoteness (Major Cities, Inner Regional, Outer Regional, Remote, Very Remote)
* Population density data (external data for per capita calculations)
* Crash\_Type (to understand severity patterns)

**Data transformations:**

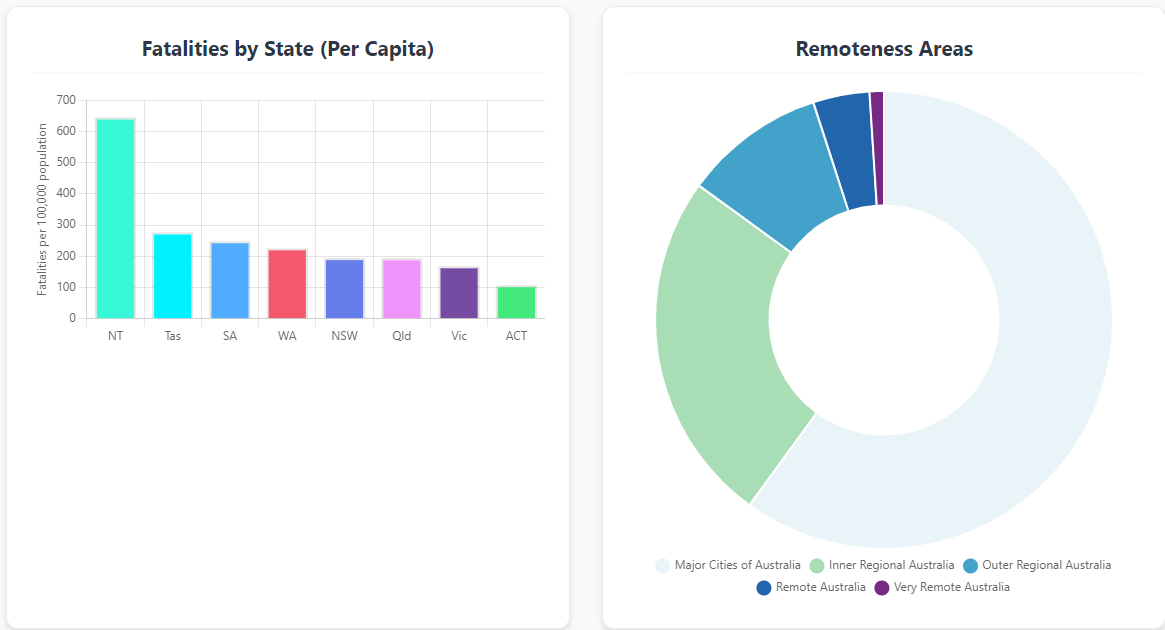
* Create speed limit categories (≤60km/h urban, 61-80km/h suburban, 81-100km/h rural, 100+km/h highway)
* Combine road type and speed limit into infrastructure risk categories
* Calculate fatality rates per kilometer of each road type (requires external road length data)
* Create remoteness-adjusted risk scores

**Data type changes:**

* Speed\_Limit (numeric) → Categorical speed zones
* Geographic coordinates → Remoteness categories
* Raw fatality counts → Fatalities per km of road infrastructure
* State names → Geographic regions (Metro vs Regional vs Remote)

**Visualization sketch:**

* Choropleth map of Australia showing fatality rates by remoteness areas
* Scatter plot with speed limit on x-axis, fatalities per km on y-axis, colored by road type, sized by traffic volume
* Parallel coordinates plot showing the relationship between state, remoteness, road type, speed limit, and fatality rate
* Box plots comparing fatality distributions across infrastructure categories



Include this file as evidence for your Demonstration 2