

DEATH RATES IN VICTORIA

Rashmitha Karkera

30438101 | FIT5147

Table of Contents

INTRODUCTION	2
1.1 MOTIVATION	2
1.2 PROBLEM DESCRIPTION.....	2
1.3 QUESTIONS	2
DATA WRANGLING AND DATA CHECKING	2
DATA EXPLORATION.....	4
CONCLUSION.....	8
REFLECTION.	8
REFERENCES	8

Introduction

1.1 Motivation

The rising rates of death amidst the recent COVID-19 pandemic spiked a curiosity in me to find the death rates that may have been happening before the pandemic. On finding the desired dataset, the visualization gave a broader view on these perspectives. The death rates are found and plotted across various parameters and are displayed in graphs: Pie Chart, Graph, etc for a better understanding.

1.2 Problem Description

The dataset of mortality rates helps create a vivid picture of how death arrives in daily lives. The dataset is used to find causes that may have resulted in death of people in Australia based on age groups, year, States and other constraints. Keeping these factors in mind, tests have been conducted across the dataset to find facts/ rates in Australia and limitations to Victoria.

1.3 Questions

From the dataset obtained the following questions can be answered

- Q1) Which type of death has high death rate in Victoria?
- Q2) Which age group has highest % of death rate by intentional self-harm in Victoria?
- Q3) Which has higher possibility of death between intentional self-harm and accidents among the age group from 15-25?

Data Wrangling and Data Checking

Given below are the datasets that are expended in providing parameters of various death rates in Australia. The data is being obtained from the site <https://www.abs.gov.au/ausstats/abs@.nsf/mf/3302.0>

For this project , I have made use of 2 .xls files which contains data about the death rates, types of Australia

- A) Cube set:- Underlying causes of death (Victoria) (Australian Bureau Of Statistics, 2019)
- B) Cube set -Intentional self-harm (Suicide) (Australia) (Australian Bureau Of Statistics, 2019)

Making use of R :

Initially we need to check the format of the dataset. Next, check for missing values and null values.

Here we make use of library “dplyr” .

The next step is to call the is.null function to check the null values.

```
1 library(tidyverse)
2 library(dplyr)
3 library(data.table)
4 rdata<- read.csv("vicdata.csv")
5 view(rdata)
6 is.null(rdata)
```

Snapshot 1

It shows FALSE. The dataset obtained didn't contain null values.

```
> is.null(rdata)
[1] FALSE
> |
```

Snapshot 2

Since, there are no null values present, we move on to the next step.

DEATH RATES IN VICTORIA

For answering the questions, the entire dataset isn't needed. Hence the data is reformatted and restructured into a new sheet. Here we make use of Excel as the tool for Data Wrangling and checking before feeding the dataset for visualization .

Q1. We have taken **Table 3.3 Underlying cause of death, Selected causes by age at death, numbers and rates, Victoria, 2018**

From the Excel sheet it is observed that the dataset has 156 rows , and there are multiple duplicate columns. The table is categorised age wise and for each age group the type of death is assigned. We need a subset of 3 columns, type of death males, female and person for reading the data. Here we make use of Consolidate option from data tab in Excel Sheet.

This will total all the duplicate since the type of death is split via age. And the data looked as follow

Abc Sheet1 Cause of death an...	# Sheet1 Males	# Sheet1 Females	# Sheet1.Females Persons
Fetus and newborn a...	45	36	81
Congenital malforma...	13	6	19
Congenital malforma...	11	7	18
Ill-defined and unkno...	11	7	18
Other congenital mal...	12	3	16
Haemorrhagic and ha...	8	5	13
Chromosomal abnor...	6	7	13
Disorders related to l...	8	4	12

Snapshot 3

As we can see that the data seems very inaccurate. So we make use of Data Interpreter tool in Tableau to check the authenticity of the dataset and organise the table. Next we make use of pivot to format the dataset in a refined way. The new version of data is shown as below.

Abc Pivot gender	# Pivot no of death	Abc Sheet5 Cause of death an...	# Sheet5 total death/ cause	Abc Sheet5 Age
Females	36	Fetus and newborn a...	81	Under 1 year
Males	45	Fetus and newborn a...	81	Under 1 year
Females	6	Congenital malforma...	19	Under 1 year
Males	13	Congenital malforma...	19	Under 1 year
Females	7	Congenital malforma...	18	Under 1 year
Males	11	Congenital malforma...	18	Under 1 year
Females	7	Ill-defined and unkno...	18	Under 1 year
Males	11	Ill-defined and unkno...	18	Under 1 year

Snapshot 4

Q2 Here we utilise the **Table 11.2 Intentional self-harm, Age-specific death rates, 5 year age groups by sex, 2009-2018**

Here again, a certain section of the dataset is utilised to check the death rate as per the age group for the whole of Australia.

The dataset has columns including the age group and reference year from 2009 to 2018 with death rates. The age group ranges from 0-14 to 85 and over. For reorganising the data we make use of Tableau. When the excel sheet is read onto the Tableau, it is noticed that there are multiple unwanted values present. The unwanted values needs to be remove. Here we make use of Data Interpreter to clear the values. The table is organised by making use of the pivot function. We pivot the columns from 2009-2019 .The new pivoted column will have values as year. Rename the table column as Year, Death Rate and Age group.

# Sheet4 Year	# Sheet4 Death Rate	Abc Sheet4 age group
2009	7.2000	15-19
2010	8.0000	15-19
2011	8.0000	15-19
2012	9.0000	15-19
2013	10.5000	15-19

Snapshot 5

Q3. We have taken **Table 3.3 Underlying cause of death, Selected causes by age at death, numbers and rates, Victoria, 2018**

The same dataset has been utilised used for Question 1.

Data Exploration

We use Tableau for the exploration of the data.

Q1) Which type of death has high death rate?

The dataset is for the whole of Victoria, and it is found that Ischaemic Heart Disease is the cause for higher death rates in Victoria. The pie chart representation of all causes is denoted and plotted with the total death rates.

Which cause has the highest death date



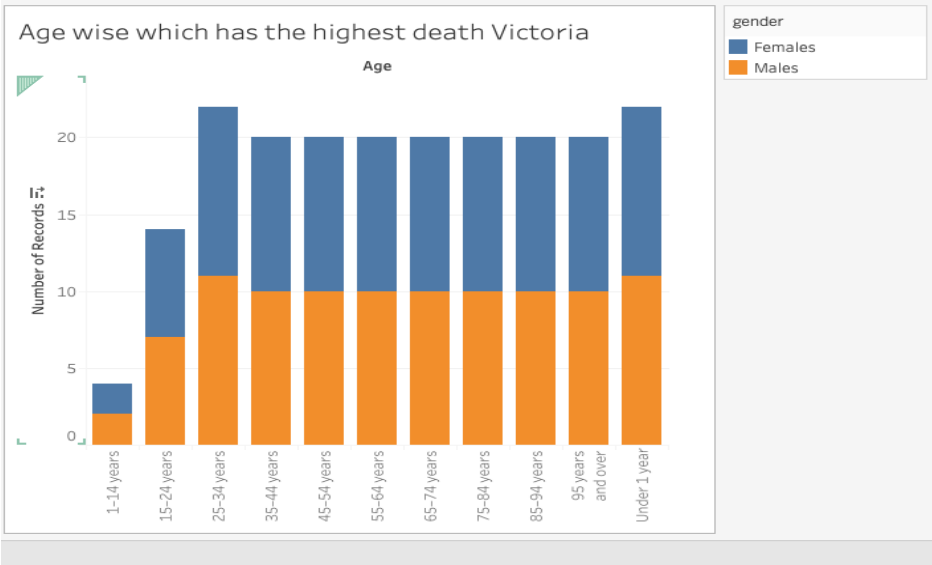
Cause of death and ICD-10 code: **Ischaemic heart diseases (I20-I25)**
Avg. total death/ cause: **535.4**

Cause of death and ICD...
Accidental poison..
Car occupant injur..
Cerebrovascular d..
Chromosomal abn..
Chronic lower res..
Congenital malfor..
Congenital malfor..
Diabetes mellitus ..
Diseases of liver (..
Disorders related ..
Episodic and paro..
Event of undeter..
Falls (W00-W19)
Fetus and newbor..
Haemorrhagic an..
Hypertensive dise..
Ill-defined and un..
Infections specific..
Influenza and pne..
Intentional self-h..
Ischaemic heart d..
AVG(total death/ cause)

Snapshot 6

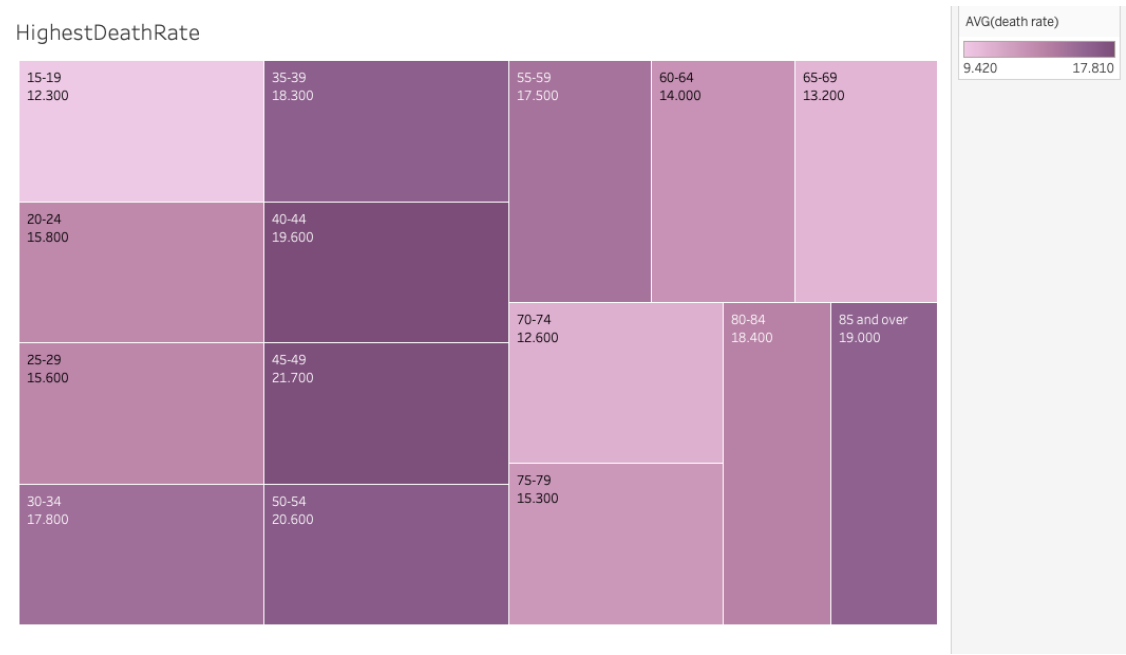
DEATH RATES IN VICTORIA

A further exploration was done to find the age group which has high death rates in the state of Victoria and it was found that age 25-34 seems to have higher rates of deaths. The bar graph below gives the depiction of age wise highest death rate in Victoria. It is obtained by plotting age across X-axis and the number of records across the Y-axis. Colour Filter is used to differentiate between the gender.



Snapshot 7

Q2) Which age group has highest % of death rate by intentional self-harm in Victoria?
Initially, using the subset 2 we , have found that the age group with high death rates in term of Intentional Self harm was around the age group 45-49.The following graph is the death rate for whole of Australia.

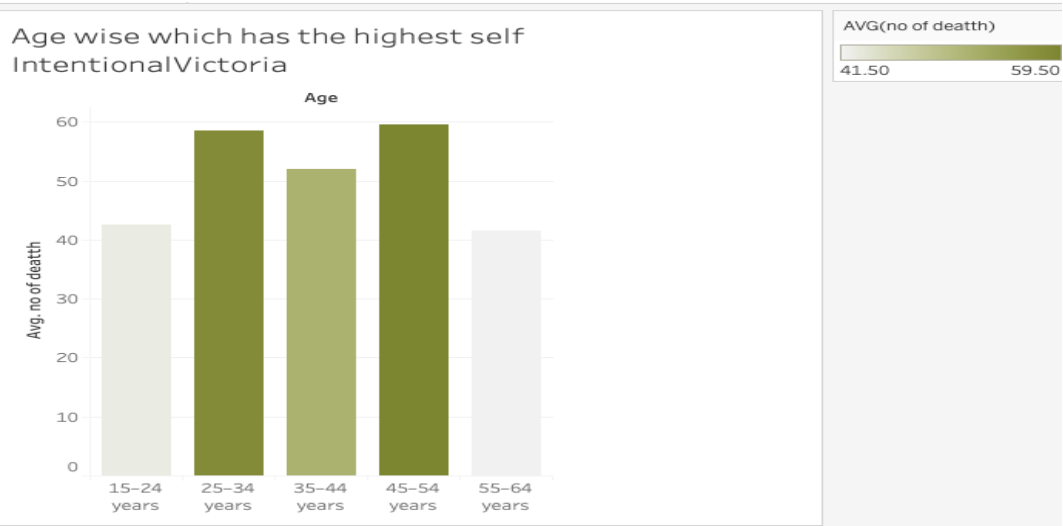


Snapshot 8

DEATH RATES IN VICTORIA

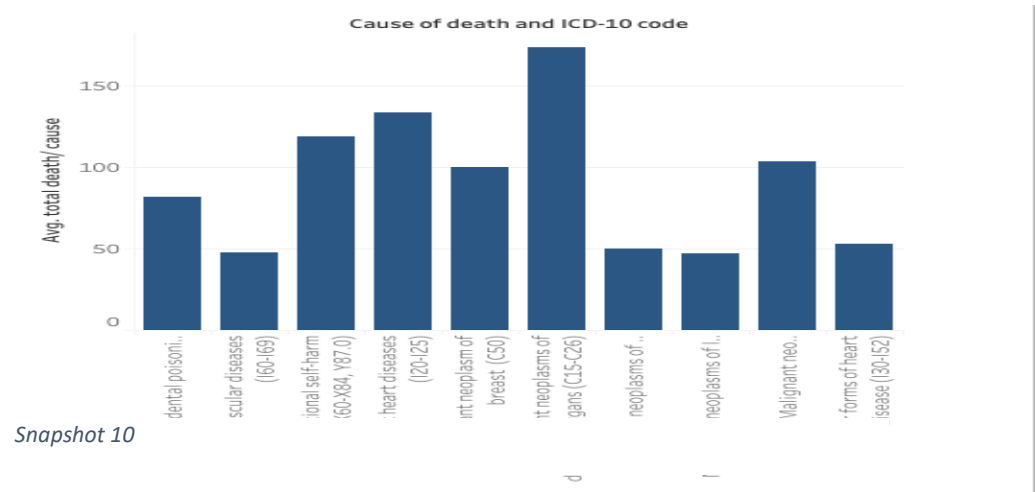
Using the subset 1 , the bar graph is plotted for the highest death rate for Victoria. Here we filter the cause of death to “Intentional Self Harm” and plot my considering the age and AVG(no of death) across the X-axis and Y-axis respectively.

Comparing the 2 graphs, it’s evident that nationwide as well as in the state of Victoria , the age group with maximum death rate in terms of Intentional Self-Harm is 45-49.



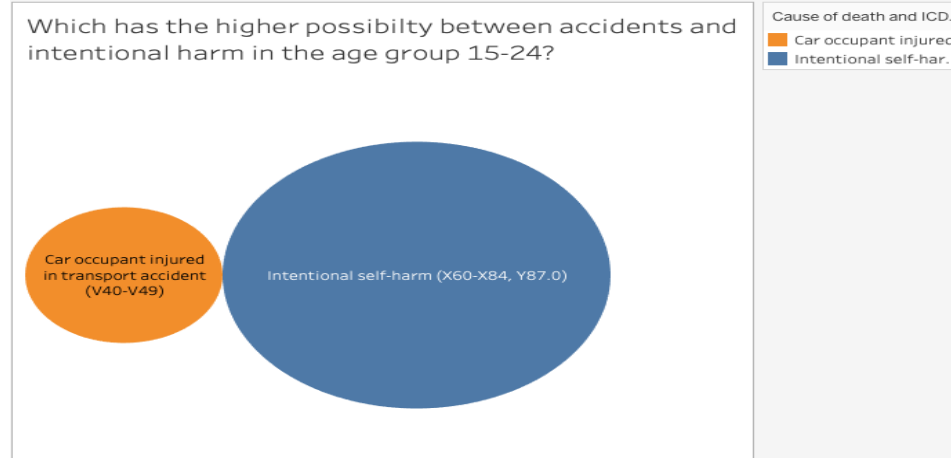
Snapshot 9

One might one wonder, what maybe the cause of death between the age 45-54 years? A further analysis was done and the cause for that is ‘ Malignant neoplasm of Digestive Organs’.



Snapshot 10

Q3) Which has higher possibility of death between intentional self-harm and accidents among the age group from 15-25? Age group between 15-25 are supposed to be rebels and take many outrageous spontaneous. There has been a rise in the global death rates among youth because of suicide. (Suicide facts and stats, n.d.) Also, the youth are said to be reckless drivers. These tests are done to check which cause of death has higher rates.

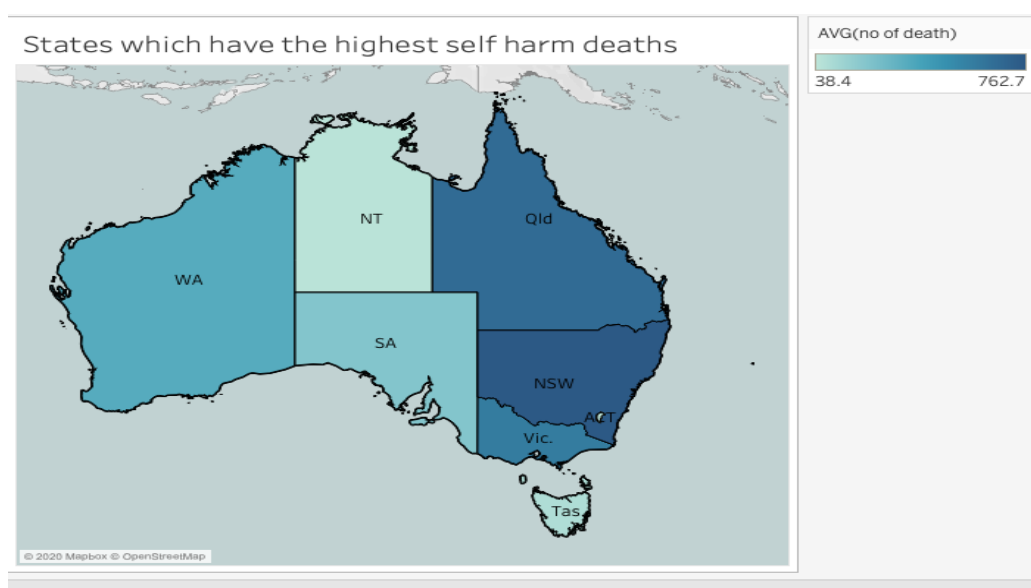


Snapshot 11

It is found that the rates are higher in terms of Intentional Harm than the accidents among the youngsters in Victoria.

Extra Findings:

The dataset which was found on the Australian Bureau of Statistics where restricted to certain limitations. Even though the dataset answered all the questions there are quite a few drawbacks. The ability to not narrow it down to the city of Melbourne is one of them. However, all the answers were found using only the dataset 1, certain other section of dataset captured my interest. One such is that, the ability to plot the death rates as per Intention self-harm from the year 2009-2018 across the map of Australia. New South Wales has a higher rate in these terms and Tasmania has the least. The geographic map was plotted by setting the datatype of location to the a certain geographic and Tableau auto generates the latitude and longitude for the graph respectively.



Snapshot 12

Conclusion

A certain level of in-depth exploration was conducted by me for the death rate datasets. While doing so, I was able to find and learn multiple ways and approaches to wrangle the data, and also clean them along the process. Initially, the questions were to find the death rates in Melbourne, however, the lack of available dataset made it hard to conduct the analysis. Yet, it provided me an insight on plotting and conducting analysis for a broader perspective. The exploration has given me an insight how each group affects the death rate, or how the causes affects each group. In addition to the exploration, the analysis has given me a chance to find purpose of some in-built functions in Tableau, R and Excel.

Reflection.

The assignment helped me brush up my skills in utilizing my skills on using the Tableau Public. The Tableau Public has made arranging and grouping the dataset in a better format. It has given me an opportunity to view the dataset in a broader prospect. Exploration of data and representing and finding answers to ever question arise is fascinating. The R tool also minimize the time for reading through the entire dataset to find errors and missing values. While exploring the data, I did learn a great deal about the statistics of the death rates and its causes and much more intriguing facts about the software tools and things they can do.

References

- A) Tabular data: -Includes both textual and number data 1962 Rows & 11 Columns. Australian Bureau Of Statistics. (2019, 9 25). Australian Bureau Of Statistics. Retrieved from 3303.0 - Causes of Death, Australia, 2018: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3303.02018?OpenDocument>
- B) Tabular data: -Includes textual and number data / Has 13 tables Australian Bureau Of Statistics. (2019, 9 25). Australian Bureau Of Statistics. Retrieved from 3303.0 - Causes of Death, Australia, 2018: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3303.02018?OpenDocument>
- C) *Suicide facts and stats*. (n.d.). Retrieved from Life in a MIND: <https://lifeinmindaustralia.com.au/about-suicide/suicide-data/suicide-facts-and-stats>