My Report of UN Data on Deaths

Mohammed Faizan

# Introduction to the research question and context

This report is based on Open Demographic data published by the United Nations Statistics Division. "The United Nations Statistics Division collects, compiles and disseminates official demographic and social statistics on a wide range of topics.[1]" This analysis is made on deaths recorded in Australia in the years, 2015, 2016 and 2017. The records were further filtered and the records with deaths below 311 were eliminated due to data size constraints. Detailed overview of the data is present in Section @ref(overview). The census data is the collection population data longitudinally every year to draw remarkable insights and general trends to allow adoption and implementation of strategies for the betterment of the studied aspect. This data from the United Nations could drive better implementation of vaccination programs, prepare hospital facilities and child care programmes among all others. Summary statistics for distribution of deaths across different variables is performed to gain insights about deaths in each category.

* Deaths by Year @ref(year)
* Deaths by Gender@ref(sex)
* Deaths by Age@ref(age)

# Exploratory Data Analysis

## Overview of the data

The UN data on death has **5** variables and **996** records. The variables in the data are: **Year, Sex, Age, Cause, Value**

checking missingness:**FALSE**

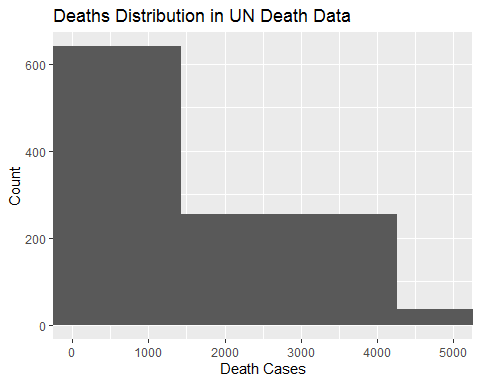
**There is no missingness in the data.**

### Unique Values of each varibale in the data set are displayed below:

* **Year**: 2017, 2016, 2015
* **Sex**: Male, Female
* **Age**: Total, 0, 20 - 24, 25 - 29, 30 - 34, 35 - 39, 40 - 44, 45 - 49, 50 - 54, 55 - 59, 60 - 64, 65 - 69, 70 - 74, 75 - 79, 80 - 84, 85 - 89, 90 - 94, 95 +
* **Cause of Death**: 44
* **Summary of Cases**:
  + Minimum Cases: 312
  + Median Cases: 850
  + Maximum Cases: 82858

### Cases Distribution

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



The deaths distribution is skewed and has most records with near to 1500 deaths and has around 200 records for with to 2000, 3000 and 4000 deaths each.

Top ten causes of death are shown in Table @ref(tab:deaths-cause).

Deaths by cause over all years, ages, sex

|  |  |
| --- | --- |
| Cause | n\_deaths |
| All causes, ICD10 | 952917 |
| Neoplasms, ICD10 | 274931 |
| Malignant neoplasms, ICD10 | 267857 |
| Diseases of the circulatory system, ICD10 | 262050 |
| Ischaemic heart diseases, ICD10 | 111656 |
| Diseases of the respiratory system, ICD10 | 87129 |
| Cerebrovascular diseases, ICD10 | 58061 |
| Mental and behavioural disorders, ICD10 | 56511 |
| External causes, ICD10 | 53682 |
| Diseases of the nervous system, ICD10 | 47502 |
| Malignant neoplasm of trachea, bronchus and lung, ICD10 | 46104 |

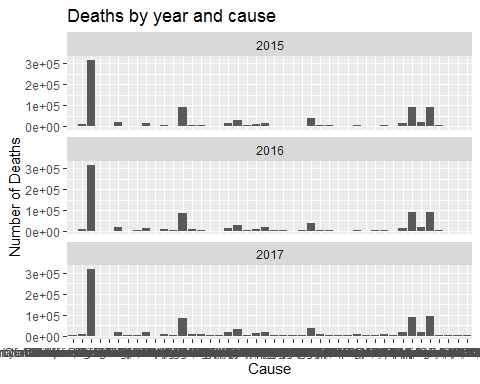
## Deaths by Year

Analysis on death by year envisions the pattern deaths of each country in terms of number of deaths. Death trend becomes charactristic of a country if the deaths are similar each year.

It can be seen from Table @ref(tab:deaths-year) that each year there are over 850,000 deaths in Australia. Of the 3 years, 2017 has seen most deaths summing to 865709. And from Figure @ref(fig:deaths-cause-year), we can see that the distribution of the cause of death is similar each year.

Deaths by year

|  |  |
| --- | --- |
| Year | n\_cases |
| 2015 | 856605 |
| 2016 | 853311 |
| 2017 | 865709 |



Deaths by year and cause

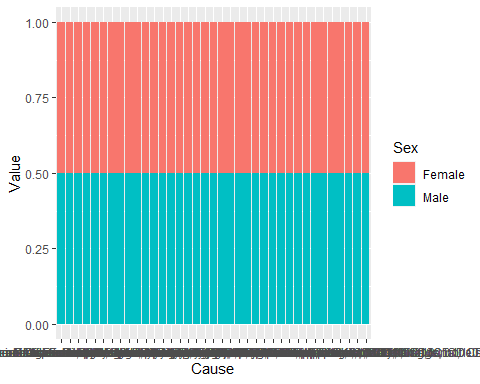
## Death by Gender

Analysis of deaths by gender is of utmost importance to recognise trends of the death causes that affect each gender individually and death causes that are general. It can be seen from Table @ref(tab:deaths-sex) that male deaths exceed female deaths by over 100,000 across 2017, 2016, 2015. And from Figure @ref(fig:deaths-cause-sex), it is noted that equal number of men and women die due to general causes of death.

Deaths by Gender

|  |  |
| --- | --- |
| Sex | n\_cases |
| Male | 1355419 |
| Female | 1220206 |

Furthermore, there are 37 causes by which both men and women die. These causes are listed in @ref(fig:deaths-cause-sex) and the deaths are even by each gender for every cause @ref(fig:deaths-cause-sex). The causes of death that affect only females are listed in @ref(fig:deaths-cause-sex) and for males it is listed in @ref(fig:deaths-cause-sex). Malignant neoplasm of prostate and Malignant neoplasm of female breast is the cause of most deaths in men and women respectively. Thus, with this information steps can be taken to educate men and women separately with respect to the precautions and treatments related to these causes.



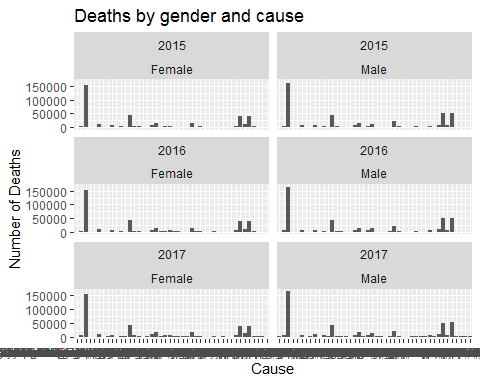
Deaths by gender and cause

Female Only deaths

|  |  |
| --- | --- |
| Cause | n\_deaths |
| All other diseases, ICD10 | 968 |
| Malignant neoplasm of female breast, ICD10 | 11477 |

Male Only Deaths

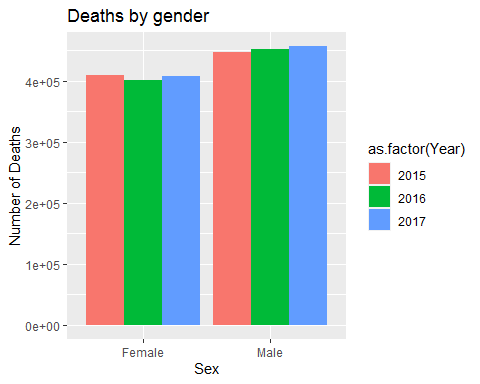
|  |  |
| --- | --- |
| Cause | n\_deaths |
| All other external causes, ICD10 | 380 |
| Certain conditions originating in the perinatal period, ICD10 | 990 |
| Congenital malformations, deformations and chromosomal abnormalities , ICD10 | 976 |
| Malignant neoplasm of lip, oral cavity and pharynx, ICD10 | 1838 |
| Malignant neoplasm of prostate, ICD10 | 17734 |



Deaths by gender and cause

It is seen from Figure @ref(fig:death-sex-year) that male deaths increase over the years and the death trend is similar for all three years.

## `summarise()` has grouped output by 'Sex'. You can override using the `.groups` argument.



death-sex-year

## Death by Age

for the analysis, the new age groups have been formed, accordingly:

* Age > 60 ~ "senior citizens",
* between(Age, 35, 60) ~ "middle aged",
* between(Age, 25, 34) ~ "youth",
* between(Age, 15, 24) ~ "teenage and youth",
* Age < 15 ~ "children"

It can be seen from Table @ref(tab:deaths-age) that senior citizen deaths far exceed other age groups across 2017, 2016, 2015.

Deaths by Age

|  |  |
| --- | --- |
| Age | n\_cases |
| senior citizens | 2406887 |
| middle aged | 154732 |
| youth | 7946 |
| children | 3315 |
| teenage and youth | 2745 |

# Conclusion

This open data can be used under the condition that "all data and metadata provided on UNdata’s website are available free of charge and may be copied freely, duplicated and further distributed provided that UNdata is cited as the reference"[1]. Open data empowers statistians and data experts with enormous analysis effectivness for their projects especially when such data is available from authenticated sources such as UNdata. This report is made to showcase the use of open data in a legitimate context.

Analysis of United Nations death data for Australia has provided for the insights as seen in @ref(overview).Over these 3 years Australia has recorded deaths with similar trends.This open data suggests that there exists a characteristic death trend for Australia.

# Resources

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

## tidyverse

## @Article{,  
## title = {Welcome to the {tidyverse}},  
## author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani},  
## year = {2019},  
## journal = {Journal of Open Source Software},  
## volume = {4},  
## number = {43},  
## pages = {1686},  
## doi = {10.21105/joss.01686},  
## }

##   
##   
## plotly

## @Book{,  
## author = {Carson Sievert},  
## title = {Interactive Web-Based Data Visualization with R, plotly, and shiny},  
## publisher = {Chapman and Hall/CRC},  
## year = {2020},  
## isbn = {9781138331457},  
## url = {https://plotly-r.com},  
## }

##   
##   
## visdat

## @Article{,  
## title = {visdat: Visualising Whole Data Frames},  
## author = {Nicholas Tierney},  
## doi = {10.21105/joss.00355},  
## url = {http://dx.doi.org/10.21105/joss.00355},  
## year = {2017},  
## publisher = {Journal of Open Source Software},  
## volume = {2},  
## number = {16},  
## pages = {355},  
## journal = {JOSS},  
## }

##   
##   
## naniar

## @Manual{,  
## title = {naniar: Data Structures, Summaries, and Visualisations for Missing Data},  
## author = {Nicholas Tierney and Di Cook and Miles McBain and Colin Fay},  
## year = {2020},  
## note = {R package version 0.6.0},  
## url = {https://CRAN.R-project.org/package=naniar},  
## }

##   
##   
## bookdown

## @Manual{,  
## title = {bookdown: Authoring Books and Technical Documents with R Markdown},  
## author = {Yihui Xie},  
## year = {2020},  
## note = {R package version 0.21},  
## url = {https://github.com/rstudio/bookdown},  
## }  
##   
## @Book{,  
## title = {bookdown: Authoring Books and Technical Documents with {R} Markdown},  
## author = {Yihui Xie},  
## publisher = {Chapman and Hall/CRC},  
## address = {Boca Raton, Florida},  
## year = {2016},  
## note = {ISBN 978-1138700109},  
## url = {https://github.com/rstudio/bookdown},  
## }

##   
##   
## knitr

## @Manual{,  
## title = {knitr: A General-Purpose Package for Dynamic Report Generation in R},  
## author = {Yihui Xie},  
## year = {2021},  
## note = {R package version 1.31},  
## url = {https://yihui.org/knitr/},  
## }  
##   
## @Book{,  
## title = {Dynamic Documents with {R} and knitr},  
## author = {Yihui Xie},  
## publisher = {Chapman and Hall/CRC},  
## address = {Boca Raton, Florida},  
## year = {2015},  
## edition = {2nd},  
## note = {ISBN 978-1498716963},  
## url = {https://yihui.org/knitr/},  
## }  
##   
## @InCollection{,  
## booktitle = {Implementing Reproducible Computational Research},  
## editor = {Victoria Stodden and Friedrich Leisch and Roger D. Peng},  
## title = {knitr: A Comprehensive Tool for Reproducible Research in {R}},  
## author = {Yihui Xie},  
## publisher = {Chapman and Hall/CRC},  
## year = {2014},  
## note = {ISBN 978-1466561595},  
## url = {http://www.crcpress.com/product/isbn/9781466561595},  
## }

Data Source:

* UNdata | Demographic Statistics Database | United Nations Statistics Division <http://data.un.org/Data.aspx?d=POP&f=tableCode%3a105%3bcountryCode%3a36%3brefYear%3a2015%2c2> 016%2c2017&c=2,3,6,8,10,12,14,16,17,18&s=\_countryEnglishNameOrderBy:asc,refYear:desc,areaCod e:asc&v=1 Accessed 10 April 2020