University of Edinburgh
Health Data Science Course
R programming assignment - an analysis of NHS Scotland hospital admissions and deaths data

Exam number: B010574

17 November 2021

Contents

1 Overview **1**

2 Data Processing **2**

2.1 Libraries 2

2.2 Import data 2

2.3 Cleaning, wrangling, analysis 2

3 Results and Discussion **5**

3.1 Table 1 - total numbers of each injury type for each age group, with rates of death for admission, by age group 5

3.2 Table 2 - total numbers of each injury type for each sex, with rates of death for admission, by sex 7

3.3 Plot 1 - rates of death for admissions of each injury type, by age group 8

3.4 Plot 2 - rates of death for admissions of each injury type, by sex 9

3.5 Discussion 10

1 Overview

This assignment uses data from the Scottish Health and Social Care Open Data platform (https://www.opendata.nhs.scot/dataset/unintentional-injuries), downloaded in CSV format.

The data shows all hospital admissions and deaths (associated with unintentional injuries) recorded annually in Scotland between 2010 and 2019. The data also contains parameters such as age group, sex and injury type, for each admission and death.

This report outlines an analysis of this data, with a focus on injury types, and how these are related to age and sex demographic groups in the data.

2 Data Processing

2.1 Libraries

```
library(tidyverse)
library(janitor)
library(ggplot2)
library(knitr)
library(kableExtra)
library(formatR)
```

2.2 Import data

```
# import data
orig_ui_admissions <- read_csv("ui_admissions_2021.csv")
orig_ui_deaths <- read_csv("ui_deaths_2021.csv")
```

2.3 Cleaning, wrangling, analysis

The data is 'tidy' and in 'long' format, and does not require further arrangement in this respect.

The data has been aggregated in several ways (by geographic location, injury type, age group and sex) and special consideration has been taken to filter out these aggregated rows, as a first step in the manipulation, to focus on the aims of the analysis.

The admissions data is manipulated to show the total number of each injury type, by age group.

```
# cleaning, wrangling - admissions, age
ui_admissions_age <- orig_ui_admissions %>%
    clean_names() %>%
    # filter out 'Scotland' data
filter(hbr != "S92000003") %>%
    # filter to use only injuries with labels, excluding 'All'
    # aggregated category
filter((injury_type != "All Diagnoses")) %>%
    # filter to use only disaggregated categories for age and sex,
    # and aggregated categories for location = 'All'
filter((age_group != "All") & (sex == "All") & (injury_location == "All")) %>%
    # turn age_group into factor type, set order of levels
mutate(age_group = factor(age_group, levels = c("0-4 years", "5-9 years",
    "10-14 years", "15-24 years", "25-44 years", "45-64 years", "65-74 years",
    "75plus years"))) %>%
    # calculate total no. of admissions by age and injury type
group_by(age_group, injury_type) %>%
    summarise(tot_by_age_injury = sum(number_of_admissions)) %>%
    arrange(desc(tot_by_age_injury), .by_group = TRUE)
```

The deaths data is manipulated to show the total number of each injury type, by age group.

```
# cleaning, wrangling - deaths, age
ui_deaths_age <- orig_ui_deaths %>%
    clean_names() %>%
    # modify injury names to match those in admissions, to enable
    # join - assumptions: 'Land transport accidents' -> 'RTA',
    # 'Accidental exposure' -> 'Accidental Exposure', 'Struck
    # by,against' -> 'Struck by, against'
mutate(injury_type = str_replace(injury_type, "Land transport accidents",
    "RTA")) %>%
    mutate(injury_type = str_replace(injury_type, "Accidental exposure",
        "Accidental Exposure")) %>%
    mutate(injury_type = str_replace(injury_type, "Struck by,against",
        "Struck by, against")) %>%
    # filter out 'Scotland' data
filter(hbr != "S92000003") %>%
    # filter to use only injuries with labels, excluding 'All'
    # aggregated category
filter((injury_type != "All")) %>%
    # filter to use only disaggregated categories for age and sex,
    # and aggregated categories for location = 'All'
filter((age_group != "All") & (sex == "All") & (injury_location == "All")) %>%
    # turn age_group into factor type, set order of levels
mutate(age_group = factor(age_group, levels = c("0-4 years", "5-9 years",
    "10-14 years", "15-24 years", "25-44 years", "45-64 years", "65-74 years",
    "75plus years"))) %>%
    # calculate total no. of deaths by age and injury type
group_by(age_group, injury_type) %>%
    summarise(tot_by_age_injury = sum(numberof_deaths)) %>%
    arrange(desc(tot_by_age_injury), .by_group = TRUE)
```

The admissions data is manipulated to show the total number of each injury type, by sex.

```
# cleaning, wrangling - admissions, sex
ui_admissions_sex <- orig_ui_admissions %>%
    clean_names() %>%
    # filter out 'Scotland' data
filter(hbr != "S92000003") %>%
    # filter to use only injuries with labels, excluding 'All'
    # aggregated category
filter((injury_type != "All Diagnoses")) %>%
    # filter to use only disaggregated categories for age and sex,
    # and aggregated categories for location = 'All'
filter((age_group == "All") & (sex != "All") & (injury_location == "All")) %>%
    # calculate total no. of admissions by sex and injury type
group_by(sex, injury_type) %>%
    summarise(tot_by_sex_injury = sum(number_of_admissions)) %>%
    arrange(desc(tot_by_sex_injury), .by_group = TRUE)
```

The deaths data is manipulated to show the total number of each injury type, by sex.

```
# cleaning, wrangling - deaths, sex
ui_deaths_sex <- orig_ui_deaths %>%
    clean_names() %>%
```

```
    # modify injury names to match those in admissions, to enable
    # join - assumptions: 'Land transport accidents' -> 'RTA',
    # 'Accidental exposure' -> 'Accidental Exposure', 'Struck
    # by,against' -> 'Struck by, against'
mutate(injury_type = str_replace(injury_type, "Land transport accidents",
    "RTA")) %>%
    mutate(injury_type = str_replace(injury_type, "Accidental exposure",
        "Accidental Exposure")) %>%
    mutate(injury_type = str_replace(injury_type, "Struck by,against",
        "Struck by, against")) %>%
    # filter out 'Scotland' data
filter(hbr != "S92000003") %>%
    # filter to use only injuries with labels, excluding 'All'
    # aggregated category
filter((injury_type != "All")) %>%
    # filter to use only disaggregated categories for age and sex,
    # and aggregated categories for location = 'All'
filter((age_group == "All") & (sex != "All") & (injury_location == "All")) %>%
    # calculate total no. of deaths by sex and injury type
group_by(sex, injury_type) %>%
    summarise(tot_by_sex_injury = sum(numberof_deaths)) %>%
    arrange(desc(tot_by_sex_injury), .by_group = TRUE)
```

The manipulated admissions and deaths data frames are joined by age group and injury type, to show total numbers of each injury types for each age group, in both admissions and deaths, to enable a calculation of the rate of deaths for each admission, by age group.

```
# join and calculate rate of deaths per admission, as a percentage,
# by age group, for each injury type - an assumption is that this
# calculation makes sense in terms of the supplied two NHS datasets
joined_df1 <- inner_join(ui_admissions_age, ui_deaths_age, by = c("age_group",
    "injury_type")) %>%
    mutate(rate = round(((tot_by_age_injury.y/tot_by_age_injury.x) * 100),
        digits = 2))
```

The manipulated admissions and deaths data frames are joined by sex and injury type, to show total numbers of each injury types for each sex, in both admissions and deaths, to enable a calculation of the rate of deaths for each admission, by sex.

```
# join and calculate rate of deaths per admission, as a percentage,
# by sex, for each injury type - an assumption is that this
# calculation makes sense in terms of the supplied two NHS datasets
joined_df2 <- inner_join(ui_admissions_sex, ui_deaths_sex, by = c("sex",
    "injury_type")) %>%
    mutate(rate = round(((tot_by_sex_injury.y/tot_by_sex_injury.x) * 100),
        digits = 2))
```

3 Results and Discussion

3.1 Table 1 - total numbers of each injury type for each age group, with rates of death for admission, by age group

```
joined_df1 %>%
    rename(`Age Group` = age_group, `Injury Type` = injury_type, Admissions = tot_by_age_injury.x,
        Deaths = tot_by_age_injury.y, `Rate (%)` = rate) %>%
    kable(caption = "Total numbers of each injury type for each age group, with rates of death for
        admission, by age group") %>%
    row_spec(0, bold = TRUE) %>%
    kable_styling(font_size = 7, latex_options = c("striped", "HOLD_position"),
        full_width = T)
```

Table 1: Total numbers of each injury type for each age group, with rates of death for admission, by age group

Age Group	Injury Type	Admissions	Deaths	Rate (%)
0-4 years	Falls	13240	6	0.05
0-4 years	Poisoning	4809	0	0.00
0-4 years	Other	3942	58	1.47
0-4 years	Struck by, against	2510	0	0.00
0-4 years	Crushing	2468	0	0.00
0-4 years	Accidental Exposure	2042	1	0.05
0-4 years	Scalds	1896	1	0.05
0-4 years	RTA	487	11	2.26
5-9 years	Falls	12788	0	0.00
5-9 years	Other	3805	9	0.24
5-9 years	Struck by, against	1878	3	0.16
5-9 years	Crushing	1423	0	0.00
5-9 years	Accidental Exposure	1307	1	0.08
5-9 years	RTA	1130	23	2.04
5-9 years	Poisoning	475	0	0.00
5-9 years	Scalds	235	0	0.00
10-14 years	Falls	9745	1	0.01
10-14 years	Other	4290	16	0.37
10-14 years	Struck by, against	2730	2	0.07
10-14 years	RTA	1453	14	0.96
10-14 years	Accidental Exposure	1369	0	0.00
10-14 years	Crushing	593	0	0.00
10-14 years	Poisoning	455	2	0.44
10-14 years	Scalds	137	0	0.00
15-24 years	Falls	12807	40	0.31
15-24 years	Other	11044	81	0.73
15-24 years	RTA	5573	325	5.83
15-24 years	Struck by, against	5483	1	0.02
15-24 years	Accidental Exposure	4795	5	0.10
15-24 years	Poisoning	3757	382	10.17
15-24 years	Crushing	1178	0	0.00
15-24 years	Scalds	316	0	0.00
25-44 years	Falls	30098	134	0.45
25-44 years	Other	22324	274	1.23
25-44 years	Accidental Exposure	9732	83	0.85
25-44 years	Poisoning	9528	3887	40.80
25-44 years	RTA	7897	516	6.53
25-44 years	Struck by, against	5939	5	0.08
25-44 years	Crushing	2341	0	0.00
25-44 years	Scalds	754	1	0.13
45-64 years	Falls	65723	507	0.77
45-64 years	Other	19641	467	2.38
45-64 years	Accidental Exposure	10113	239	2.36
45-64 years	RTA	7841	547	6.98
45-64 years	Poisoning	6013	2380	39.58
45-64 years	Struck by, against	3267	14	0.43
45-64 years	Crushing	2238	5	0.22
45-64 years	Scalds	694	5	0.72
65-74 years	Falls	50994	645	1.26
65-74 years	Other	5829	294	5.04
65-74 years	Accidental Exposure	4266	222	5.20
65-74 years	RTA	2583	207	8.01
65-74 years	Poisoning	1650	225	13.64
65-74 years	Struck by, against	1101	9	0.82
65-74 years	Crushing	493	2	0.41
65-74 years	Scalds	243	2	0.82
75plus years	Falls	154280	6872	4.45
75plus years	Accidental Exposure	7421	261	3.52
75plus years	Other	6093	449	7.37
75plus years	RTA	3514	303	8.62
75plus years	Poisoning	2599	139	5.35
75plus years	Struck by, against	2234	3	0.13
75plus years	Crushing	447	1	0.22
75plus years	Scalds	398	9	2.26

3.2 Table 2 - total numbers of each injury type for each sex, with rates of death for admission, by sex

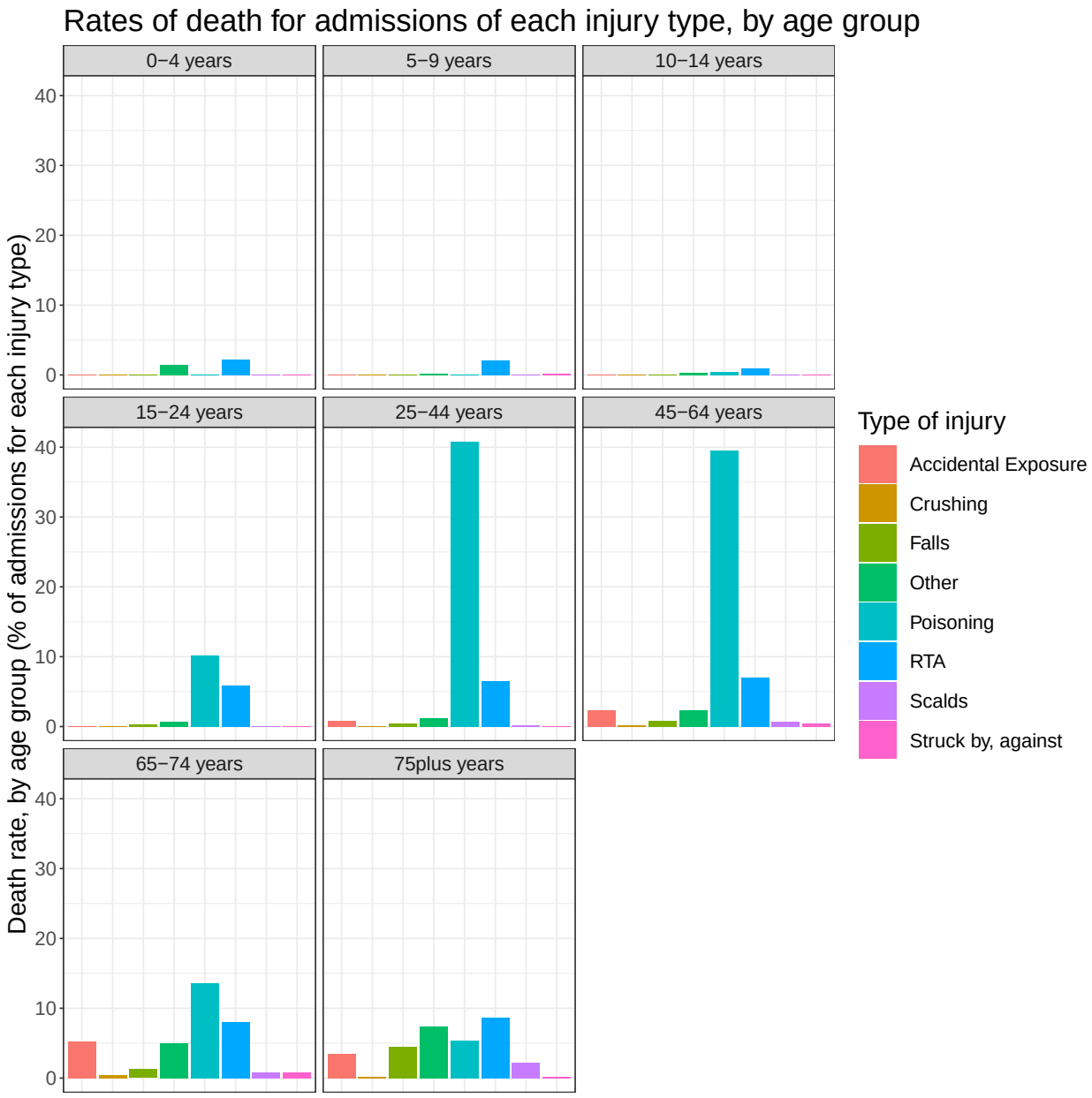
```
joined_df2 %>%
    rename(Sex = sex, `Injury Type` = injury_type, Admissions = tot_by_sex_injury.x,
        Deaths = tot_by_sex_injury.y, `Rate (%)` = rate) %>%
    kable(caption = "Total numbers of each injury type for each sex, with rates of death for
    admission, by sex") %>%
    row_spec(0, bold = TRUE) %>%
    kable_styling(font_size = 7, latex_options = c("striped", "HOLD_position"),
        full_width = T)
```

Table 2: Total numbers of each injury type for each sex, with rates of death for admission, by sex

Sex	Injury Type	Admissions	Deaths	Rate (%)
Female	Falls	201816	4732	2.34
Female	Other	26140	553	2.12
Female	Accidental Exposure	15962	315	1.97
Female	Poisoning	13694	1970	14.39
Female	RTA	10682	539	5.05
Female	Struck by, against	6866	8	0.12
Female	Crushing	3751	1	0.03
Female	Scalds	2026	9	0.44
Male	Falls	147859	3473	2.35
Male	Other	50828	1095	2.15
Male	Accidental Exposure	25083	497	1.98
Male	RTA	19796	1407	7.11
Male	Struck by, against	18276	29	0.16
Male	Poisoning	15592	5045	32.36
Male	Crushing	7430	7	0.09
Male	Scalds	2647	9	0.34

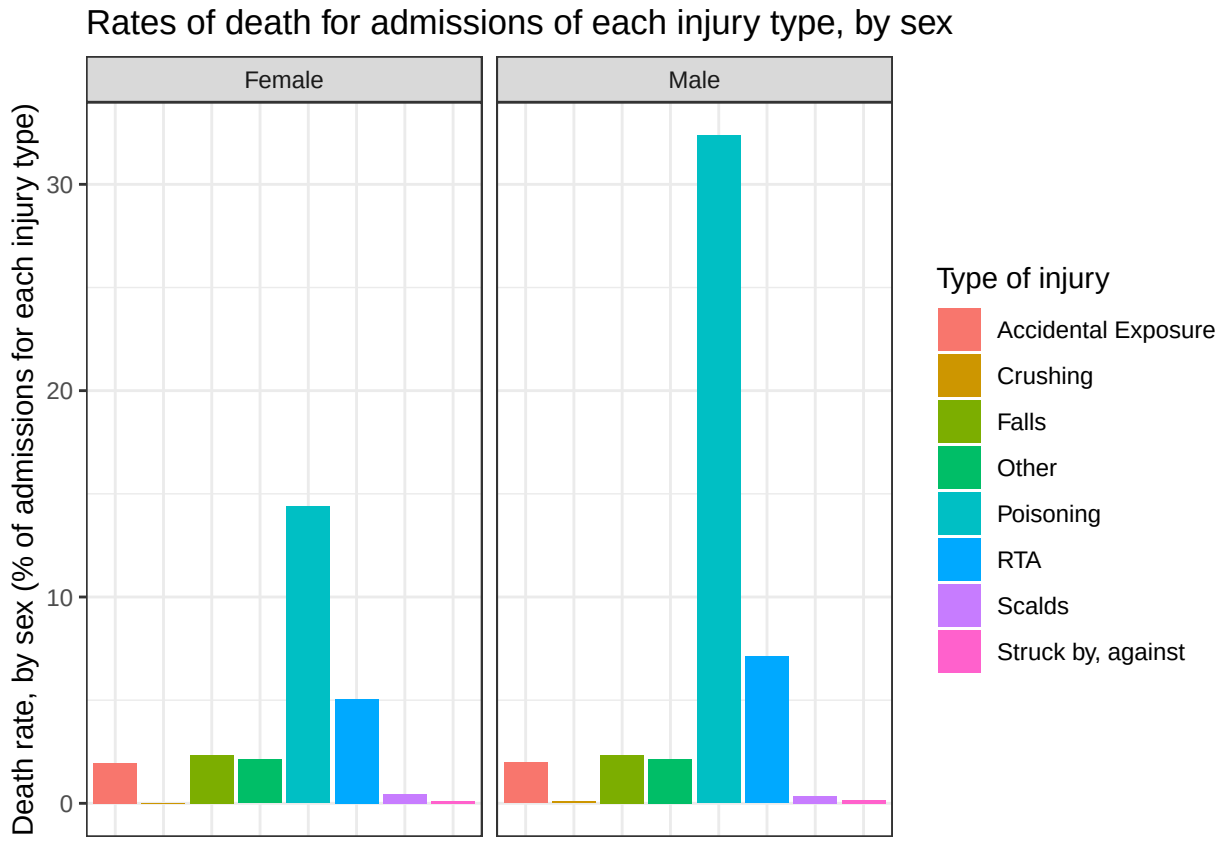
3.3 Plot 1 - rates of death for admissions of each injury type, by age group

```
# plot
plot1 <- ggplot(data = joined_df1, aes(x = injury_type, y = rate, fill = injury_type)) +
    ggtitle(paste("Rates of death for admissions", "of each injury type, by age group")) +
    geom_bar(stat = "identity") + theme_bw() + facet_wrap(~age_group) +
    labs(x = "", y = paste("Death rate, by age group", "(% of admissions for each injury type)")) +
    scale_fill_discrete(name = "Type of injury") + theme(text = element_text(size = 18),
    legend.key.height = unit(1, "cm"), legend.key.width = unit(1, "cm"),
    axis.title.x = element_blank(), axis.text.x = element_blank(), axis.ticks.x = element_blank())
plot1
```



3.4 Plot 2 - rates of death for admissions of each injury type, by sex

```
# plot
plot2 <- ggplot(data = joined_df2, aes(x = injury_type, y = rate, fill = injury_type)) +
    ggtitle("Rates of death for admissions of each injury type, by sex") +
    geom_bar(stat = "identity") + theme_bw() + facet_wrap(~sex) + labs(x = "",
    y = "Death rate, by sex (% of admissions for each injury type)") +
    scale_fill_discrete(name = "Type of injury") + theme(axis.title.x = element_blank(),
    axis.text.x = element_blank(), axis.ticks.x = element_blank())
plot2
```



3.5 Discussion

From the tables and plots, several features of the data can be seen:

1. The most common type of injury for all of the demographic groups used in this analysis (age group and sex), is Falls. This is particularly acute for the youngest and oldest age groups, with 42% of all admissions in the 0-4 years age group being for Falls, and 87% in the 75plus years age group.
2. Rates of deaths by admissions for the first 3 age groups (0-4 years, 5-9 years and 10-14 years) are relatively low (<2.5%), for all injury types ('relatively' in this context refers to other age groups).
3. For the next four age groups (15-24 years, 25-44 years, 45-64 years and 65-74 years), rates of death for all injury types remain relatively low (<5.5%), except Poisoning and RTA. In particular, the rate of death by Poisoning in the 25-44 and 45-64 age groups is significantly higher (40.80% and 39.58% respectively) than for any other type of injury.
4. Rates of deaths for females and males are roughly the same for all injury types, except Poisoning, for which the rates for males (32.36%) are significantly higher than for females (14.39%).

The first conclusion which can be drawn from this analysis is that Falls contribute to a significant amount of hospital admissions, more than any other type of injury.

The second conclusion which can be drawn is that death rates by Poisoning of males in the age groups of 25 to 64 years old in Scotland could indicate an issue which requires attention by public and policy makers, for potential prevention strategies. 'Poisoning' includes deaths from drug abuse, specifically acute intoxication.

These conclusions match similar conclusions in a 2021 publication by Public Health Scotland (PHS) (https://publichealthscotland.scot/media/5838/2021-03-23-ui-2021-report.pdf).

Further analysis could include:

1. Variations of these rates over geographical regions covered by the data (Heath Board Regions or Council Areas in Scotland)
2. Variations of these rates over the time period covered by the data (2010-2019)
3. Correlations with Scottish Index of Multiple Deprivation (SIMD) data