alcohol_admissions_analysis_practive_JW

JW

2024-11-07

Exercise 1: For the alcohol-related hospital admission data per intermediate data zone (the data we used in the previous section), plot the data distributions for each health board in 2019.

Load packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

read in data

```
raw_data <- read_csv("scotpho_data_extract.csv")
```

```
## New names:
## * `` -> `...13`
## * `` -> `...14`
## * `` -> `...15`
## * `` -> `...16`
## * `` -> `...17`
## * `` -> `...18`
## * `` -> `...19`
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 15160 Columns: 19
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (7): indicator, area_name, area_code, area_type, period, definition, dat...
## dbl (5): year, numerator, measure, lower_confidence_interval, upper_confiden...
## lgl (7): ...13, ...14, ...15, ...16, ...17, ...18, ...19
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
inter_zone_codes <- read_csv("iz2011_codes_and_labels_21042020.csv")
```

```
## New names:
## Rows: 1279 Columns: 18
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (10): IntZone, IntZoneName, CA, CAName, HSCP, HSCPName, HB, HBName, Coun... lgl
## (8): ...10, ...11, ...12, ...13, ...14, ...15, ...16, ...17
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
## * `` -> `...15`
## * `` -> `...16`
## * `` -> `...17`
## * `` -> `...18`
```

joint data

```
iz_raw_data <- raw_data %>%
  filter(area_type == "Intermediate zone") %>%
  select(area_name,area_code,year,measure)

iz_codes_HBname <- inter_zone_codes %>%
  select(IntZone,HBName)

admission_data <- left_join(iz_raw_data,iz_codes_HBname, by = c("area_code" = "IntZone"))

admission_data <- admission_data %>%
  mutate(HBName = gsub("NHS ", "", HBName)) %>%
  rename(health_board = HBName,
         alcohol_admissions = measure) %>%
  select('year','alcohol_admissions','health_board')

glimpse(admission_data)
```

```
## Rows: 12,790
## Columns: 3
```

```
## $ year                <dbl> 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 201~
## $ alcohol_admissions <dbl> 475.18, 732.32, 372.46, 1306.45, 359.21, 416.62, 23~
## $ health_board        <chr> "Grampian", "Grampian", "Grampian", "Grampian", "Gr~
```

##select data for each health board in 2019

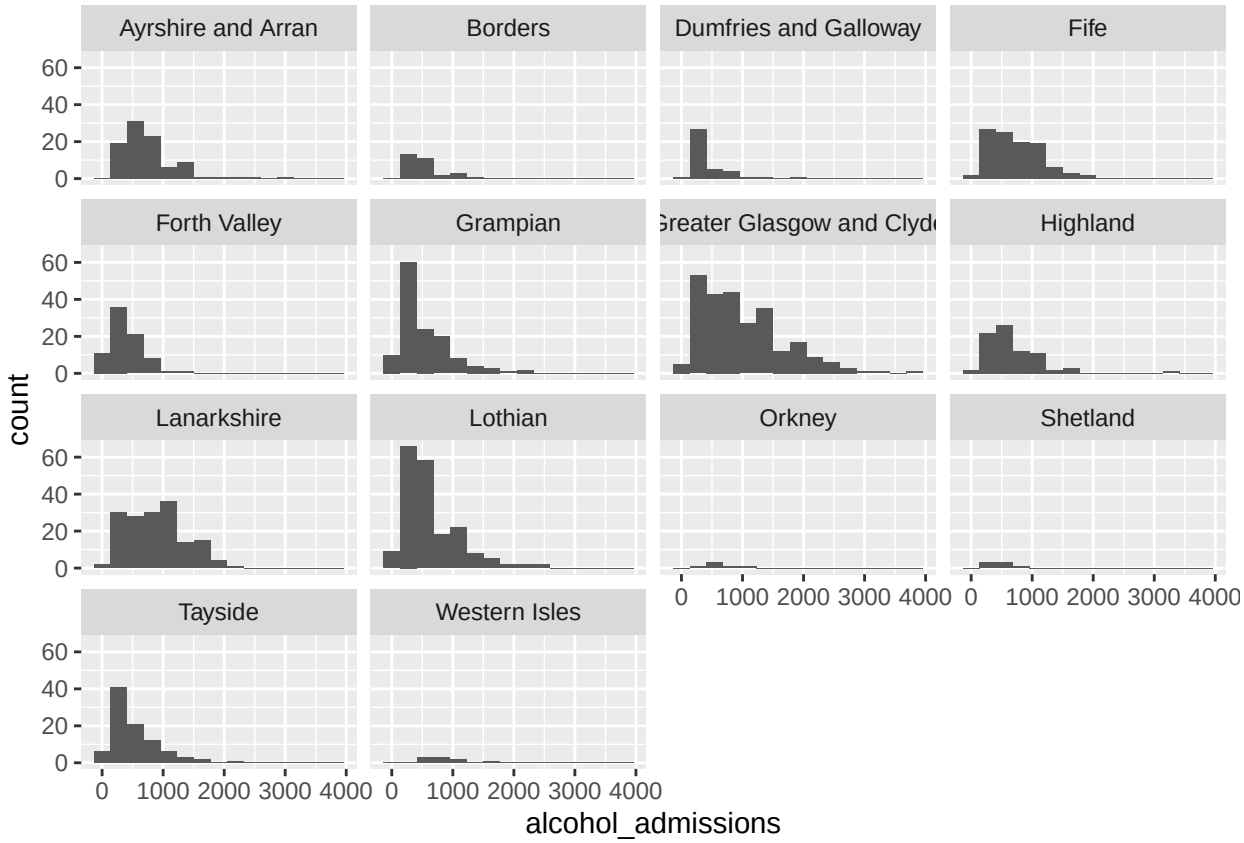
```
all_health_board_2019 <- admission_data %>%
  filter(year == 2019)

glimpse(all_health_board_2019)
```

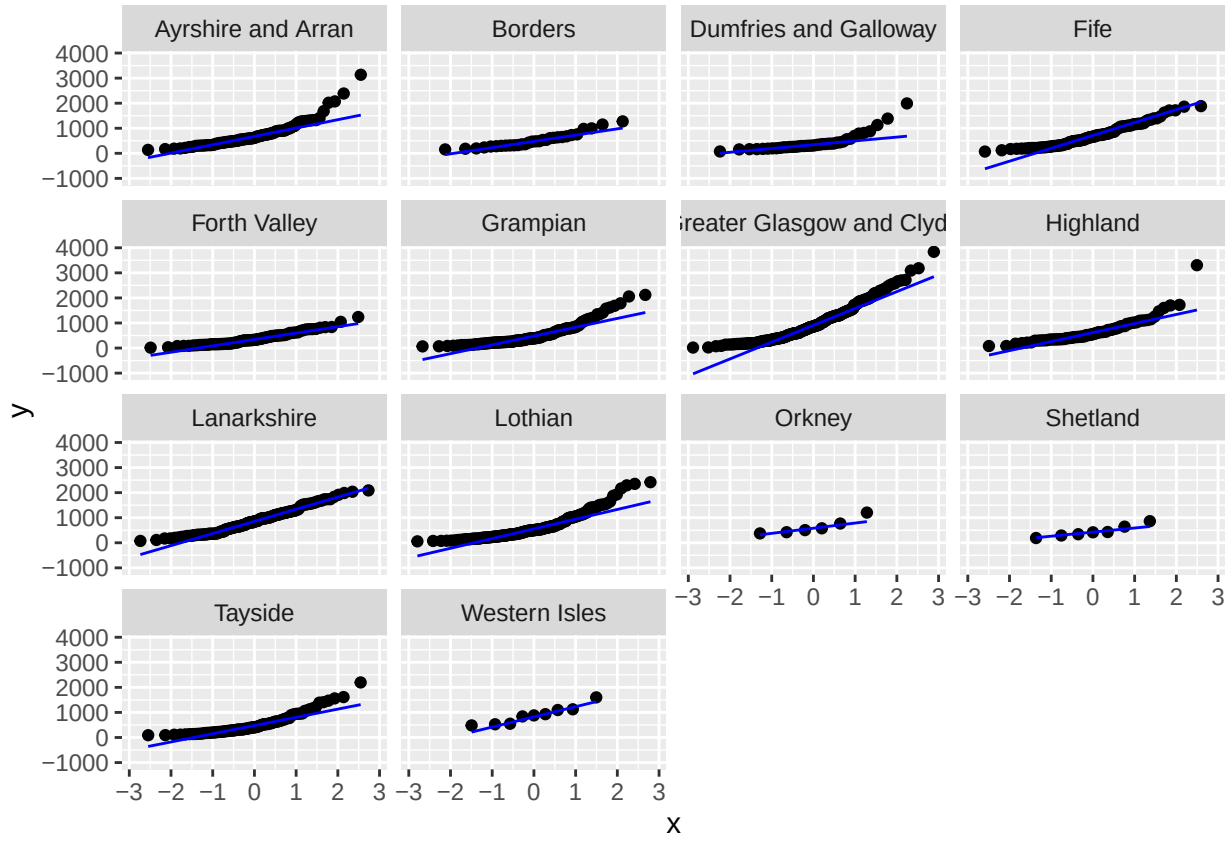
```
## Rows: 1,279
## Columns: 3
## $ year                <dbl> 2019, 2019, 2019, 2019, 2019, 2019, 2019, 2019, 201~
## $ alcohol_admissions <dbl> 520.74, 353.82, 197.59, 909.61, 252.83, 719.23, 220~
## $ health_board        <chr> "Grampian", "Grampian", "Grampian", "Grampian", "Gr~
```

plot the data distributions for each health board in 2019

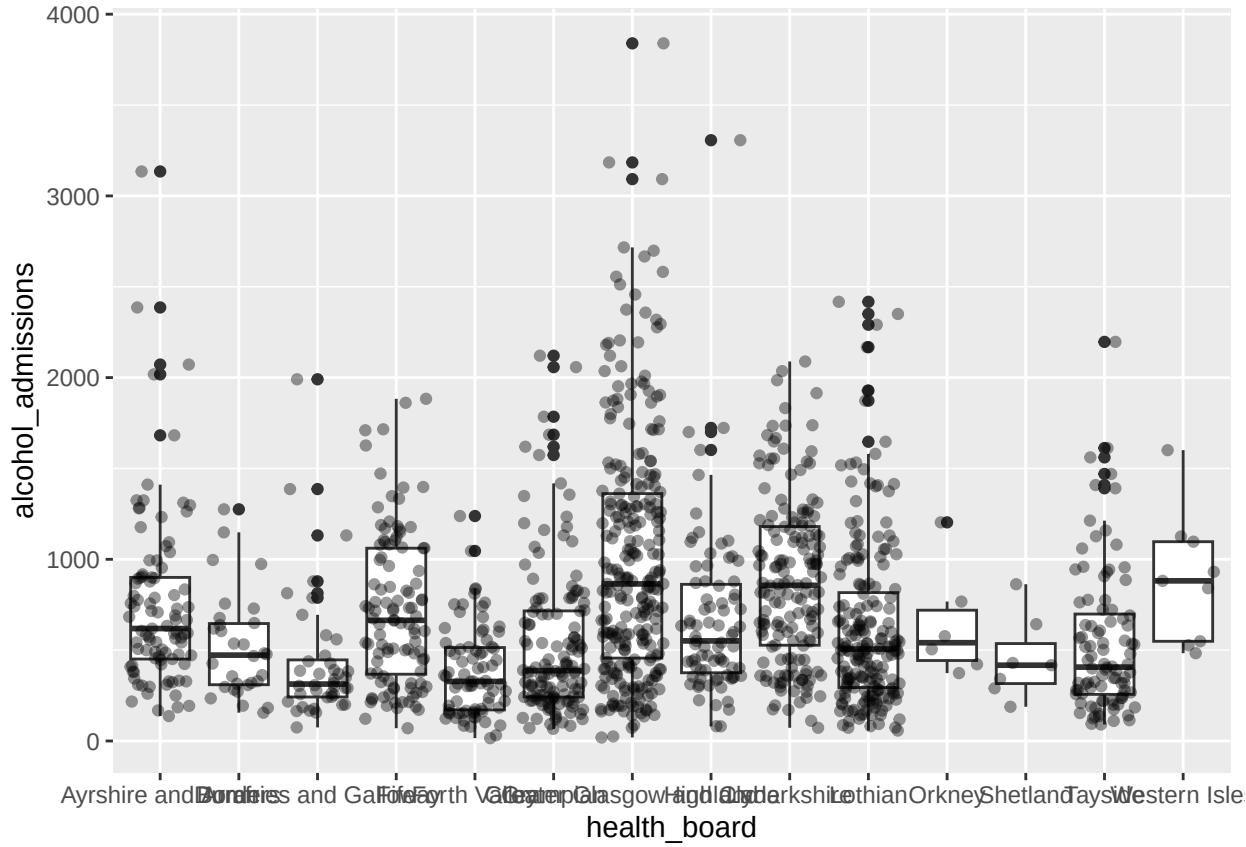
```
#histogram
all_health_board_2019 %>%
  ggplot(aes(x = alcohol_admissions)) +
  geom_histogram(bins = 15) +
  facet_wrap(~health_board)
```



```
# Q-Q plot
all_health_board_2019 %>%
  ggplot(aes(sample = alcohol_admissions)) +
  geom_qq() +
  geom_qq_line(colour = "blue") +
  facet_wrap(~health_board)
```



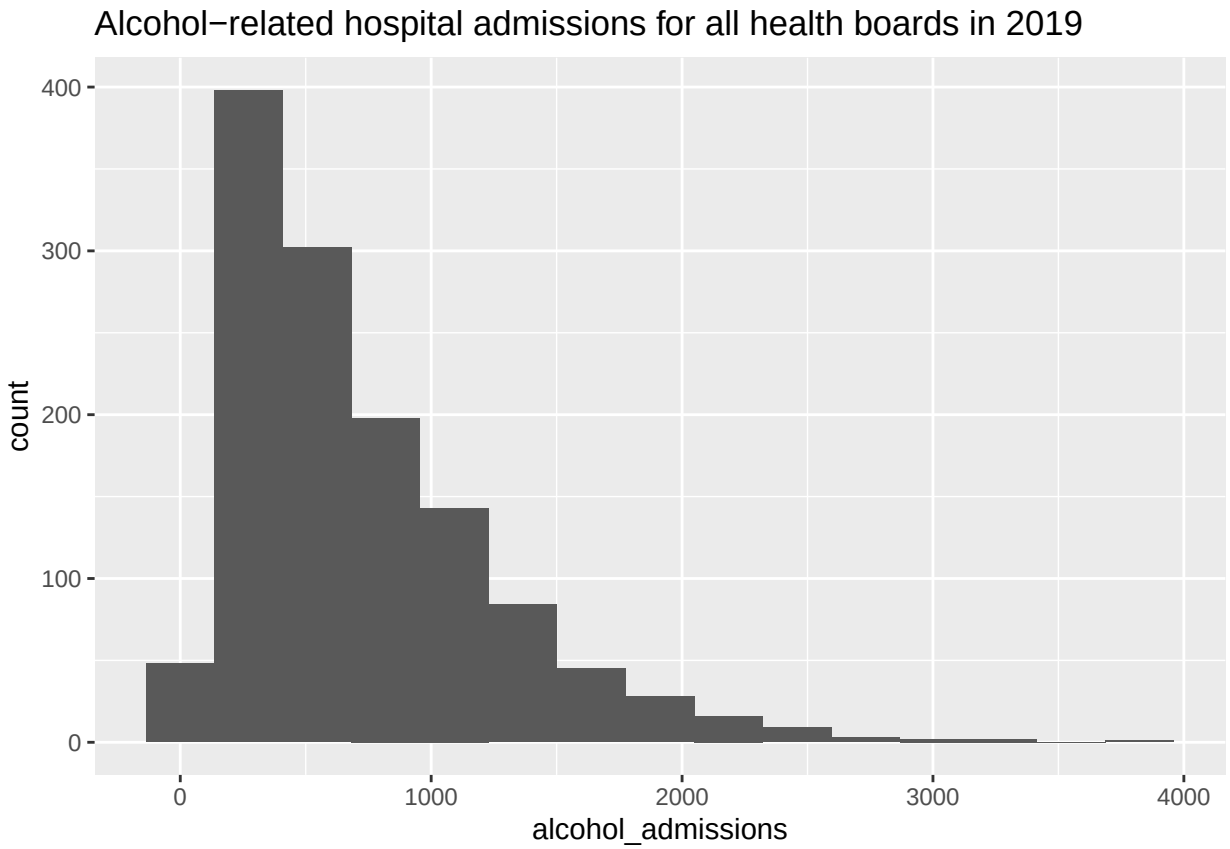

```
# boxplot
all_health_board_2019 %>%
  ggplot(aes(x = health_board,
             y = alcohol_admissions)) +
  geom_boxplot() +
  geom_jitter(alpha = 0.4) +                # add data points
  theme(legend.position = "none")
```



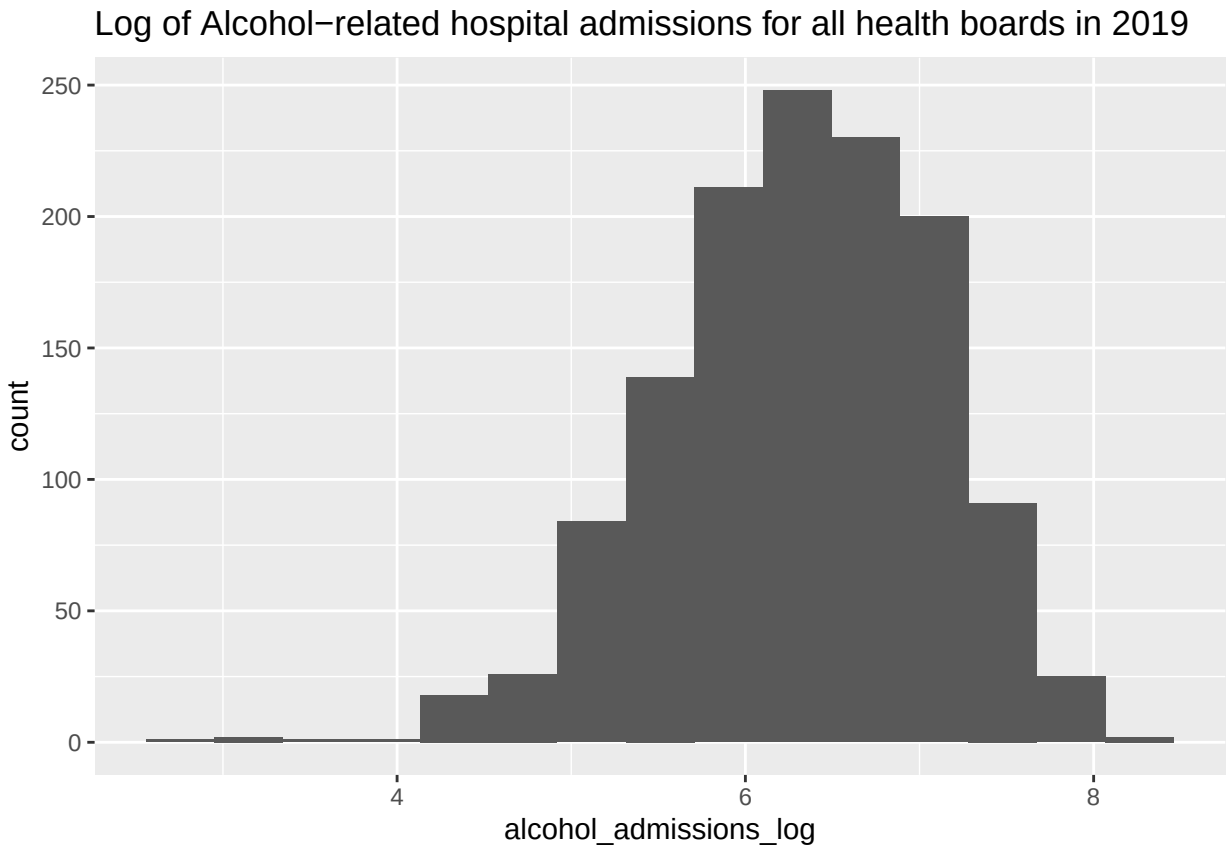
Transformation to normal

```
all_health_board_2019_log <- all_health_board_2019 %>%
  mutate(alcohol_admissions_log = log(alcohol_admissions))

all_health_board_2019 %>%
  ggplot(aes(x = alcohol_admissions)) +
  geom_histogram(bins = 15) +
  ggtitle("Alcohol-related hospital admissions for all health boards in 2019")
```




```
all_health_board_2019_log %>%
  ggplot(aes(x = alcohol_admissions_log)) +
  geom_histogram(bins = 15) +
  ggtitle("Log of Alcohol-related hospital admissions for all health boards in 2019")
```


Log of Alcohol−related hospital admissions for all health boards in 2019
count
250
200
150
100
50
0
4
6
8
alcohol_admissions_log