

Practical

2025-11-18

LOAD PACKAGES

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## vforcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyverse  1.3.1
## v purrr    1.2.0
## -- 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
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(here)
```

```
## here() starts at /home/myke/Desktop/Intro2HDS_R_WEEK5_MAIN
```

READ DATA

```
ara_data <- read_csv(here("INPUTS/WEEK7/alcohol_related_admissions.csv"))
```

```
## Rows: 12790 Columns: 5
## -- Column specification -----
## Delimiter: ","
```

```

## chr (3): area_name, area_code, health_board
## dbl (2): year, alcohol_admissions
##
## 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.

aad_data <- read_csv(here("INPUTS/WEEK7/alcohol-admissions_deprivation-data.csv"))

```

```

## Rows: 216 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr (5): indicator, geography_code, quintile, period, definition
## dbl (1): indicator_measure
##
## 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.


```

```
scotpho_data <- read_csv(here("INPUTS/WEEK7/scotpho_data_extract.csv"))
```

```

## Rows: 15160 Columns: 12
## -- 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...
##
## 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.


```

```
izc_data <- read_csv(here("INPUTS/WEEK7/iz2011_codes_and_labels_21042020.csv"))
```

```

## Rows: 1279 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (9): IntZone, IntZoneName, CA, CAName, HSCP, HSCPName, HB, HBName, Country
##
## 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.


```

EXERCISE 1

Exercise 1: For the alcohol-related hospital admission data per intermediate data zone (the dataset we have produced in the Data Analysis in R section), plot the data distributions for each health board in 2019.

```
head(ara_data)
```

```

## # A tibble: 6 x 5
##   area_name           area_code  year alcohol_admissions health_board
##   <chr>                <chr>    <dbl>            <dbl> <chr>
## 1 Culter              S02001236  2010             475. Grampian
## 2 Cults, Bieldside & Milltimber~ S02001237  2010             732. Grampian

```

```

## 3 Cults, Bieldside & Milltimber~ S02001238 2010          372. Grampian
## 4 Garthdee                  S02001239 2010          1306. Grampian
## 5 Braeside, Mannofield, Broomhi~ S02001240 2010          359. Grampian
## 6 Braeside, Mannofield, Broomhi~ S02001241 2010          417. Grampian

```

```

hb19_ara_data <- ara_data %>%
  filter(year == 2019) %>%
  select(alcohol_admissions, health_board) %>%
  group_by(health_board) %>%
  summarise(total = sum(alcohol_admissions))

head(hb19_ara_data)

```

```

## # A tibble: 6 x 2
##   health_board      total
##   <chr>              <dbl>
## 1 Ayrshire and Arran    70033.
## 2 Borders                15597.
## 3 Dumfries and Galloway  17639.
## 4 Fife                   74894.
## 5 Forth Valley            29890.
## 6 Grampian                70716.

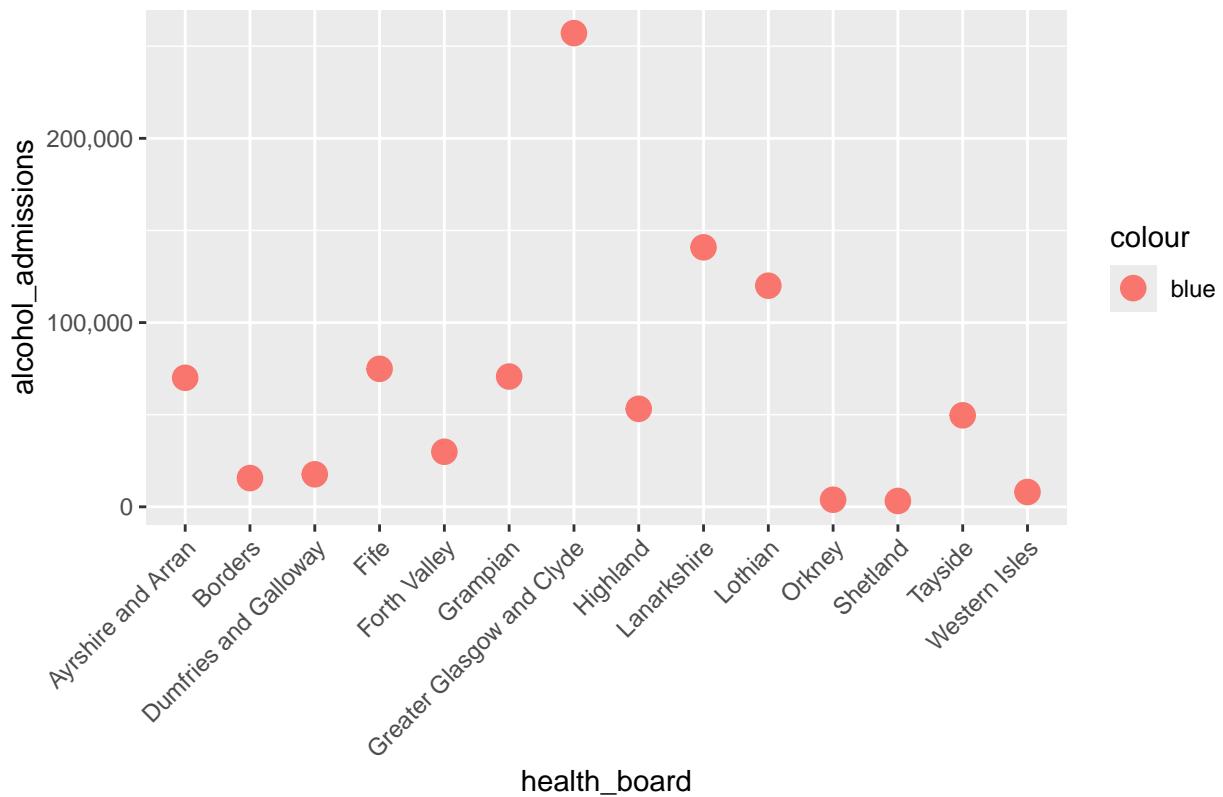
```

```

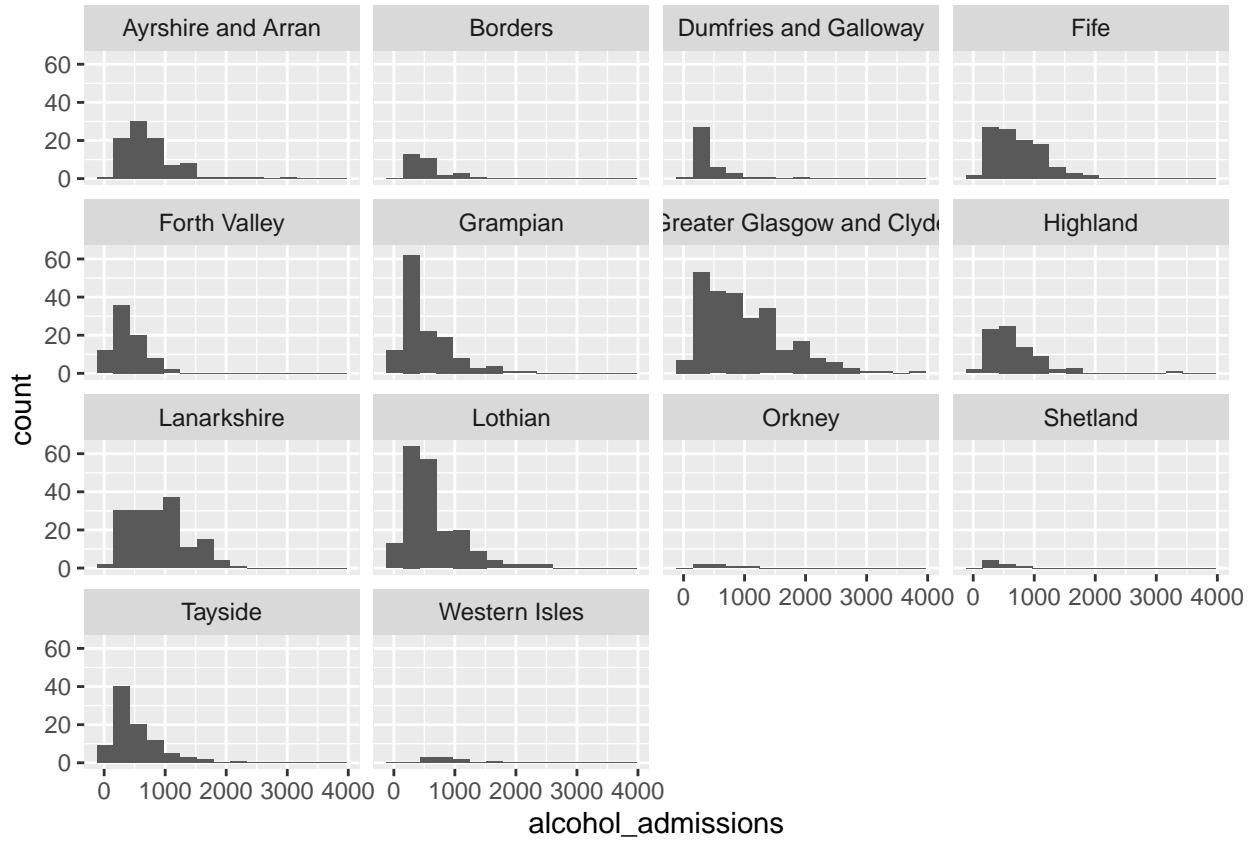
hb19_ara_data %>%
  ggplot(aes(x = health_board, y = total, color = "blue")) +
  geom_point(size = 4) +
  labs(title = "hb19 admissions",
       x = "health_board",
       y = "alcohol_admissions") +
  scale_y_continuous(labels = comma) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

hb19 admissions



```
ara_data %>%
  filter(year == 2019) %>%
  ggplot(aes(x = alcohol_admissions)) +
  geom_histogram(bins = 15) +
  facet_wrap(~health_board)
```



EXERCISE 2

Exercise 2: For which health boards do we have data on alcohol-related hospital admissions for each deprivation quintile?

```
head(aad_data)
```

```
## # A tibble: 6 x 6
##   indicator      geography_code quintile period indicator_measure definition
##   <chr>          <chr>        <chr>     <chr>           <dbl> <chr>
## 1 Alcohol-related h~ S08000020    1 - mos~ 2002/~       1289. Age-sex s~
## 2 Alcohol-related h~ S08000020    2          2002/~       673.  Age-sex s~
## 3 Alcohol-related h~ S08000020    3          2002/~       380.  Age-sex s~
## 4 Alcohol-related h~ S08000020    4          2002/~       286.  Age-sex s~
## 5 Alcohol-related h~ S08000020    5 - lea~ 2002/~       225.  Age-sex s~
## 6 Alcohol-related h~ S08000020    Total     2002/~       556.  Age-sex s~
```

```
hb_code_quintile_data <- aad_data %>%
  select(hb_code = geography_code, quintile)
```

```
head(hb_code_quintile_data)
```

```
## # A tibble: 6 x 2
##   hb_code  quintile
```

```

##   <chr>      <chr>
## 1 S08000020 1 - most deprived
## 2 S08000020 2
## 3 S08000020 3
## 4 S08000020 4
## 5 S08000020 5 - least deprived
## 6 S08000020 Total

head(izc_data)

## # A tibble: 6 x 9
##   IntZone  IntZoneName      CA    CAName HSCP  HSCPName HB     HBName Country
##   <chr>    <chr>          <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 S02001236 Culter        S120~ Aberd~ S370~ Aberdee~ S080~ NHS G~ S92000~
## 2 S02001237 Cults, Bieldside a~ S120~ Aberd~ S370~ Aberdee~ S080~ NHS G~ S92000~
## 3 S02001238 Cults, Bieldside a~ S120~ Aberd~ S370~ Aberdee~ S080~ NHS G~ S92000~
## 4 S02001239 Garthdee       S120~ Aberd~ S370~ Aberdee~ S080~ NHS G~ S92000~
## 5 S02001240 Braeside, Mannofie~ S120~ Aberd~ S370~ Aberdee~ S080~ NHS G~ S92000~
## 6 S02001241 Braeside, Mannofie~ S120~ Aberd~ S370~ Aberdee~ S080~ NHS G~ S92000~

hb_code_hb_name_data <- izc_data %>%
  select(HB, HBName) %>%
  distinct()

head(hb_code_hb_name_data)

## # A tibble: 6 x 2
##   HB      HBName
##   <chr>  <chr>
## 1 S08000020 NHS Grampian
## 2 S08000030 NHS Tayside
## 3 S08000022 NHS Highland
## 4 S08000019 NHS Forth Valley
## 5 S08000017 NHS Dumfries and Galloway
## 6 S08000015 NHS Ayrshire and Arran

hb_pairs_data <- hb_code_quintile_data %>%
  group_by(quintile) %>%
  summarise(
    count_hb = n_distinct(hb_code),
    member = paste(unique(hb_code), collapse = ", ")
  )

hb_code_quintile_data %>%
  left_join(hb_code_hb_name_data, by = c("hb_code" = "HB")) %>%
  group_by(quintile) %>%
  summarise(
    count_hb = n_distinct(hb_code),
    hb_codes = paste(unique(hb_code), collapse = ", "),
    hb_names = paste(unique(HBName), collapse = ", ")
  )

```

```

## # A tibble: 6 x 4
##   quintile      count_hb hb_codes      hb_names
##   <chr>        <int> <chr>        <chr>
## 1 1 - most deprived    2 S08000020, S08000029 NHS Grampian, NHS Fife
## 2 2                2 S08000020, S08000029 NHS Grampian, NHS Fife
## 3 3                2 S08000020, S08000029 NHS Grampian, NHS Fife
## 4 4                2 S08000020, S08000029 NHS Grampian, NHS Fife
## 5 5 - least deprived 2 S08000020, S08000029 NHS Grampian, NHS Fife
## 6 Total            2 S08000020, S08000029 NHS Grampian, NHS Fife

```

EXERCISE 3

Exercise 3: Create boxplots showing alcohol admissions for each deprivation quintile. Create a different plot for each health board in the data set.

```
head(aad_data)
```

```

## # A tibble: 6 x 6
##   indicator      geography_code quintile period indicator_measure definition
##   <chr>        <chr>        <chr>     <chr>        <dbl> <chr>
## 1 Alcohol-related h~ S08000020    1 - mos~ 2002/~          1289. Age-sex s~
## 2 Alcohol-related h~ S08000020    2           2002/~          673.  Age-sex s~
## 3 Alcohol-related h~ S08000020    3           2002/~          380.  Age-sex s~
## 4 Alcohol-related h~ S08000020    4           2002/~          286.  Age-sex s~
## 5 Alcohol-related h~ S08000020    5 - lea~ 2002/~          225.  Age-sex s~
## 6 Alcohol-related h~ S08000020    Total     2002/~          556.  Age-sex s~

```

```

grampian_aad_data <- aad_data %>%
  filter(geography_code == "S08000020", quintile != "Total") %>%
  select(geography_code, quintile, indicator_measure)

```

```
head(grampian_aad_data)
```

```

## # A tibble: 6 x 3
##   geography_code quintile      indicator_measure
##   <chr>        <chr>        <dbl>
## 1 S08000020    1 - most deprived 1289.
## 2 S08000020    2             673.
## 3 S08000020    3             380.
## 4 S08000020    4             286.
## 5 S08000020    5 - least deprived 225.
## 6 S08000020    1 - most deprived 1473.

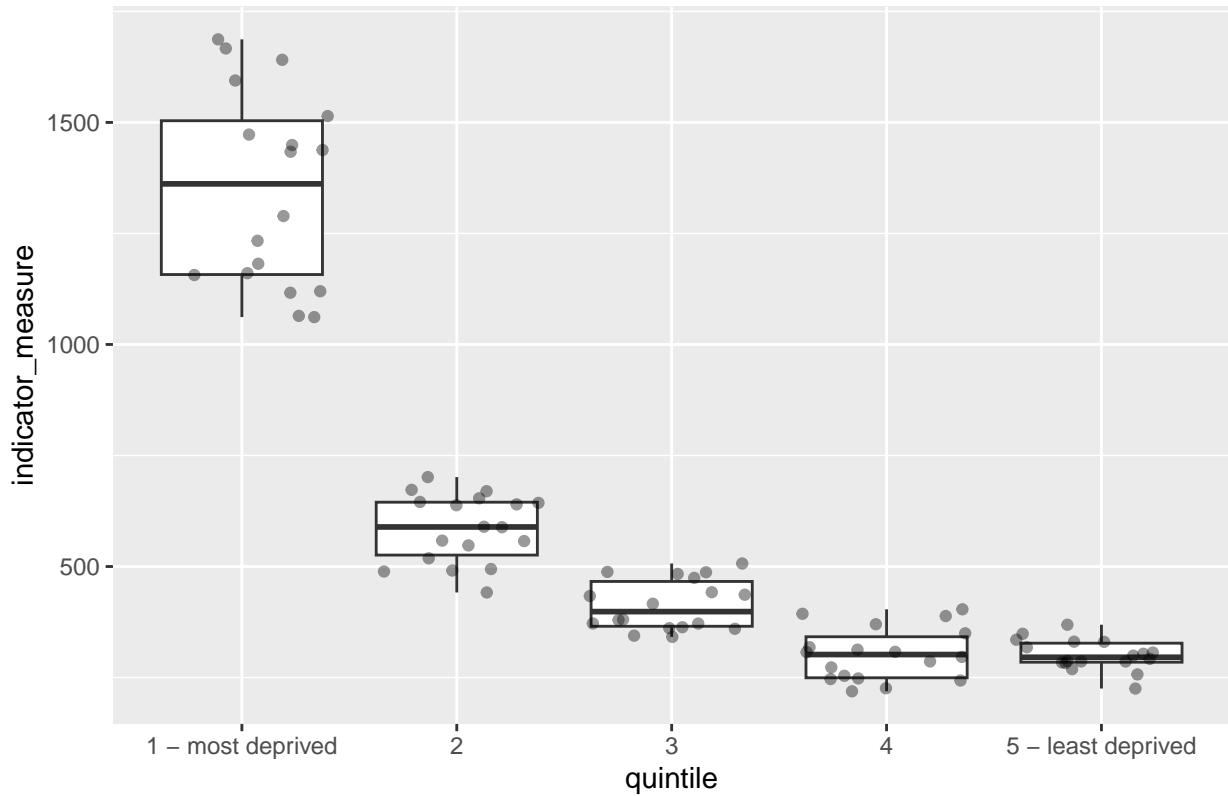
```

```

grampian_aad_data %>%
  ggplot(aes(x = quintile,
             y = indicator_measure)) +
  geom_boxplot() +
  geom_jitter(alpha = 0.4) +
  labs(title = "grampian_data") +
  theme(legend.position = "none")

```

grampian_data



```
fife_aad_data <- aad_data %>%
  filter(geography_code == "S08000029", quintile != "Total") %>%
  select(geography_code, quintile, indicator_measure)
```

```
head(fife_aad_data)
```

```
## # A tibble: 6 x 3
##   geography_code quintile      indicator_measure
##   <chr>          <chr>                  <dbl>
## 1 S08000029     1 - most deprived    901.
## 2 S08000029     2                      653.
## 3 S08000029     3                      457.
## 4 S08000029     4                      344.
## 5 S08000029     5 - least deprived   205.
## 6 S08000029     1 - most deprived    893.
```

```
fife_aad_data %>%
  ggplot(aes(x = quintile,
             y = indicator_measure)) +
  geom_boxplot() +
  geom_jitter(alpha = 0.4) +
  labs(title = "fife_data") +
  theme(legend.position = "none")
```

fife_data

