

Investigating Data - Covid Wider Impacts

Sarah Hughes

evidence of meeting PDA outcome 2.5. Exploratory data analysis

#Data Notes

First look at the wider covid impacts data.

About the data.

weekly variability - related to 2018/19 baseline.

revised version - now extracting all Scotland more easily.

```
library("tidyverse")
library("janitor")
library("lubridate")
library("skimr")
library("infer")
```

read in healthboard codes

```
hb_codes <- read_csv(here::here("lookup_tables/health_board_codes.csv"))
hb_codes <- clean_names(hb_codes)
```

load in datafiles

```
weekly_admissions_spec <- read_csv("Covid admissions by health board and speciality.csv")
weekly_admissions_spec <- clean_names(weekly_admissions_spec)

weekly_admissions_dep <- read_csv("Covid admissions by health board and deprivation.csv")
weekly_admissions_dep <- clean_names(weekly_admissions_dep)

weekly_admissions_demog <- read_csv("Covid admissions by health board, age and sex.csv")
weekly_admissions_demog <- clean_names(weekly_admissions_demog)
```

data cleaning note probably more cleaning needed before we finalise I changed the spelling but it turns out both spellings are correct!

```
weekly_admissions_spec <- weekly_admissions_spec %>%
  rename("speciality"= "specialty") %>%
  rename("speciality_qf"= "specialty_qf")
```

merge hbnames into datafiles

```

weekly_admissions_spec <- left_join(weekly_admissions_spec,hb_codes)
weekly_admissions_demog <- left_join(weekly_admissions_demog,hb_codes)
weekly_admissions_dep <- left_join(weekly_admissions_dep,hb_codes)

```

```

# for some reason this line is causing pdf error -?? as Neringa
#skim(weekly_admissions_spec)

```

do the mutates needed for identifying winter and 'crisis'

```

weekly_admissions_dep <- weekly_admissions_dep %>%
  mutate(year = as.integer(str_sub(week_ending,1,4))) %>%
  mutate(month = as.integer(str_sub(week_ending,5,6))) %>%
  mutate(day = as.integer(str_sub(week_ending,7,8))) %>%
  mutate(wdate = ymd(week_ending)) %>%
  # identify "All Scotland" data
  mutate(hb_name = ifelse(hb=="S92000003","All Scotland",hb_name)) %>%
  mutate(hb_name = ifelse(is.na(hb_name),"NHS Region Unknown",hb_name)) %>%
  mutate(iswinter = ifelse(month %in% c(4,5,6,7,8,9),FALSE,TRUE)) %>%
  mutate(above_thresh = ifelse(percent_variation>0,7,0))

```

```

weekly_admissions_demog <- weekly_admissions_demog %>%
  mutate(year = as.integer(str_sub(week_ending,1,4))) %>%
  mutate(month = as.integer(str_sub(week_ending,5,6))) %>%
  mutate(day = as.integer(str_sub(week_ending,7,8))) %>%
  mutate(wdate = ymd(week_ending)) %>%
  # identify "All Scotland" data
  mutate(hb_name = ifelse(hb=="S92000003","All Scotland",hb_name)) %>%
  mutate(hb_name = ifelse(is.na(hb_name),"NHS Region Unknown",hb_name)) %>%
  mutate(iswinter = ifelse(month %in% c(4,5,6,7,8,9),FALSE,TRUE)) %>%
  mutate(above_thresh = ifelse(percent_variation>0,7,0))

```

```

weekly_admissions_spec <- weekly_admissions_spec %>%
  mutate(year = as.integer(str_sub(week_ending,1,4))) %>%
  mutate(month = as.integer(str_sub(week_ending,5,6))) %>%
  mutate(day = as.integer(str_sub(week_ending,7,8))) %>%
  mutate(wdate = ymd(week_ending)) %>%
  # identify "All Scotland" data
  mutate(hb_name = ifelse(hb=="S92000003","All Scotland",hb_name)) %>%
  mutate(hb_name = ifelse(is.na(hb_name),"NHS Region Unknown",hb_name)) %>%
  mutate(iswinter = ifelse(month %in% c(4,5,6,7,8,9),FALSE,TRUE)) %>%
  mutate(above_thresh = ifelse(percent_variation>0,7,0))

```

How would we make monthly data?

```

monthly_admissions_spec <- weekly_admissions_spec

monthly_admissions_spec <- monthly_admissions_spec %>%
  group_by(hb, admission_type, speciality, year, month) %>%
  summarise(monthly_admissions = 4*mean(number_admissions, na.rm = TRUE)) %>%
  mutate(mdate = as.Date(make_datetime(year, month, 15))) %>%
  ungroup()

```

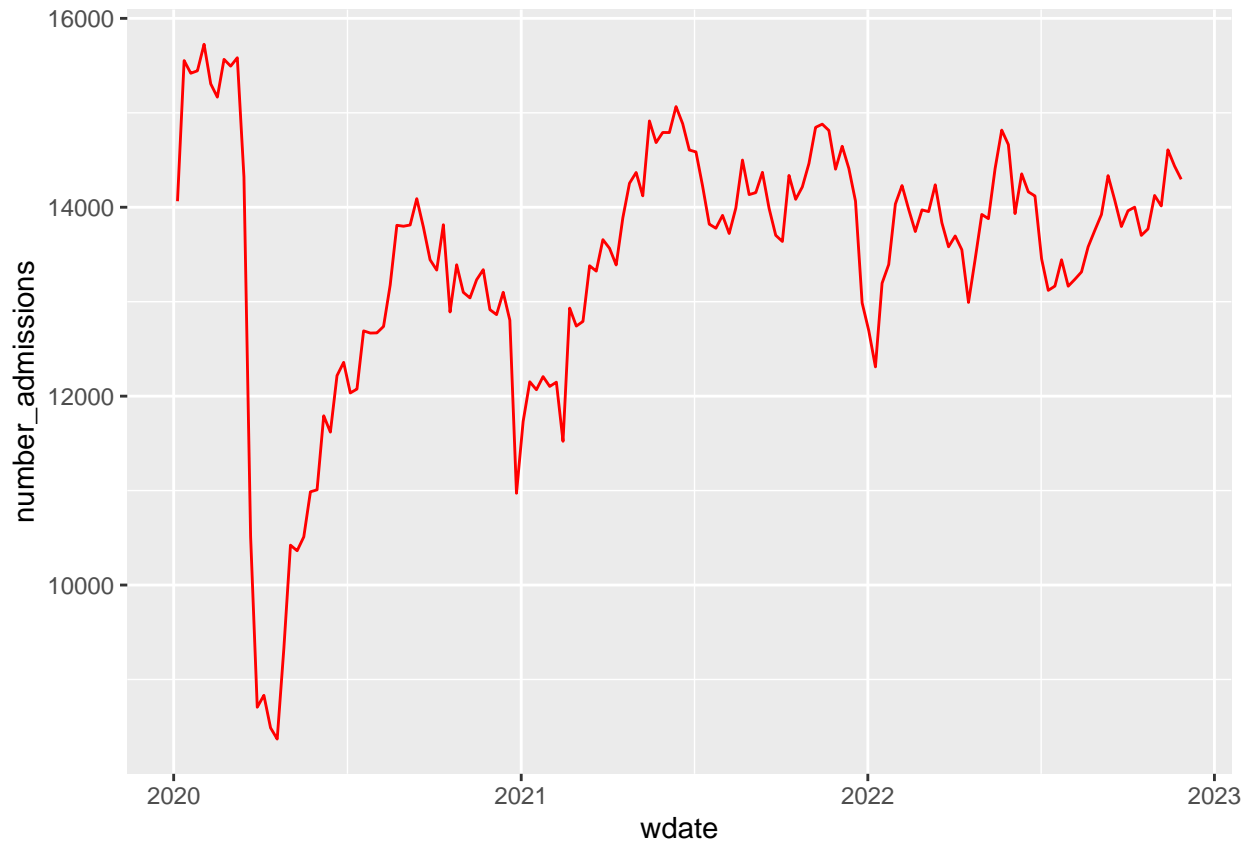
```
monthly_admissions_spec <- left_join(monthly_admissions_spec,hb_codes)
```

```
## Joining, by = "hb"
```

```
monthly_admissions_spec <- monthly_admissions_spec %>%
  mutate(hb_name = ifelse(is.na(hb_name), "NHS Region Unknown",hb_name)) %>%
  mutate(iswinter = ifelse(month %in% c(4,5,6,7,8,9),FALSE,TRUE))
```

plot weekly data

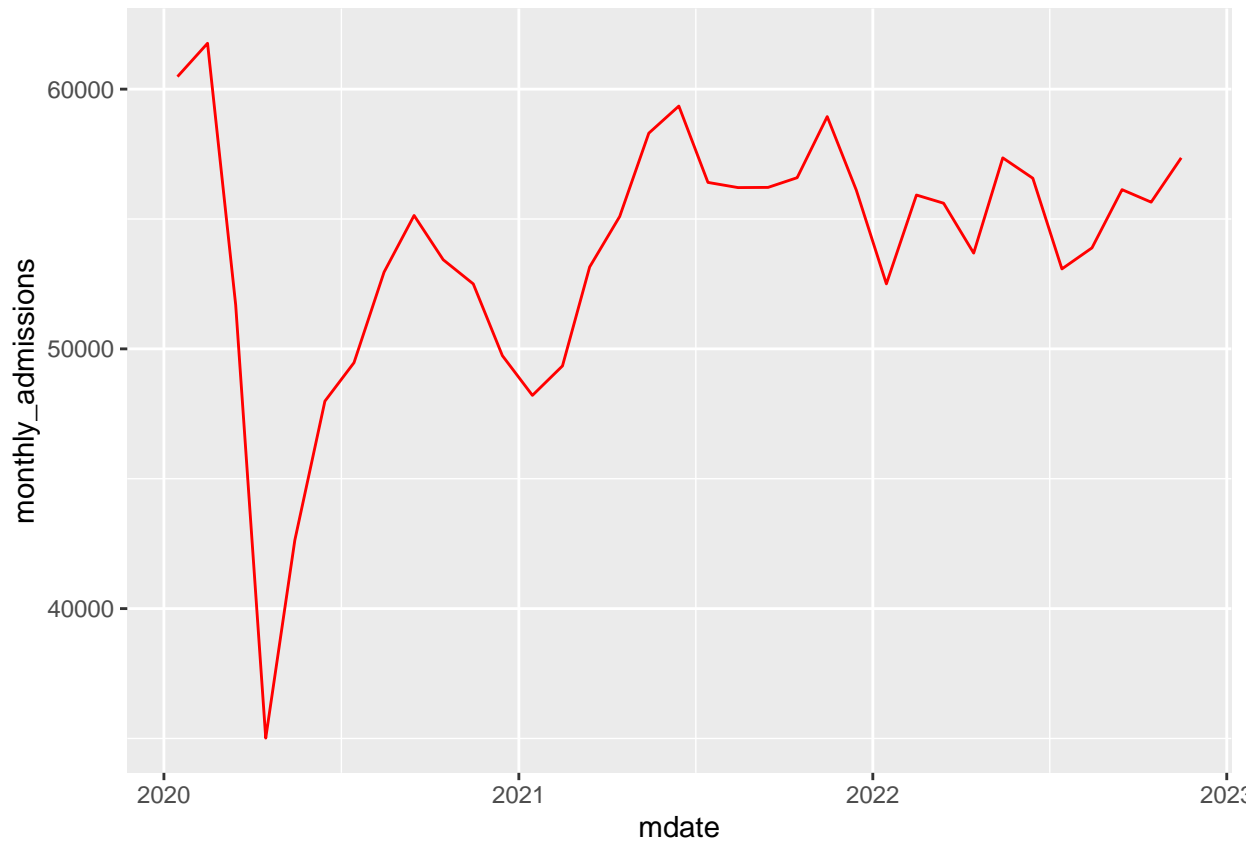
```
weekly_admissions_spec %>%
  filter(speciality == "All") %>%
  filter(admission_type == "All") %>%
  filter(hb_name == "All Scotland") %>%
  ggplot() +
  aes(x=wdate, y = number_admissions) +
  geom_line(colour='red')
```



check results - plot monthly data

```
monthly_admissions_spec%>%
  filter(speciality == "All") %>%
  filter(admission_type == "All") %>%
  filter(hb_name == "All Scotland") %>%
  ggplot() +
```

```
aes(x=mdate, y = monthly_admissions) +
geom_line(colour='red')
```



```
monthly_admissions_spec %>%
filter(year == 2021) %>%
filter(month==8) %>%
  filter(speciality == "All") %>%
  filter(admission_type == "All") %>%
  filter(hb_name == "All Scotland")
```

```
## # A tibble: 1 x 12
##   hb      admis~1 speci~2  year month month~3 mdate      hb_name hb_da~4 hb_da~5
##   <chr>   <chr>   <chr>   <int> <int>   <dbl> <date>   <chr>     <dbl>   <dbl>
## 1 S92000~ All     All     2021     8  56210. 2021-08-15 All Sc~      NA      NA
## # ... with 2 more variables: country <chr>, iswinter <lgl>, and abbreviated
## #   variable names 1: admission_type, 2: speciality, 3: monthly_admissions,
## #   4: hb_date_enacted, 5: hb_date_archived
```

try using smoother?

```
library(slider)

data_rolling <- weekly_admissions_spec %>%
  filter(wdate > as.Date("2020-05-01")) %>%
```

```

filter(speciality == "All") %>%
filter(admission_type == "All") %>%
filter(hb_name == "All Scotland") %>%
mutate(
  moving_avg = slide_dbl(
    .x = number_admissions,
    .f = ~ mean(., na.rm = TRUE),
    .before = 12,
    .after = 12,
    .complete = FALSE
  )
)
data_rolling

```

```

## # A tibble: 135 x 22
##       id week_ending hb      hbqf  admis-1 admis-2 speci-3 speci-4 numbe-5 avera-6
##   <dbl>      <dbl> <chr> <chr> <chr>    <chr>    <chr>    <chr>    <dbl>    <dbl>
## 1  2416  20200503 S920~ d      All     d      All     d      10422  15608.
## 2  2548  20200510 S920~ d      All     d      All     d      10363  15114.
## 3  2682  20200517 S920~ d      All     d      All     d      10510  15783
## 4  2815  20200524 S920~ d      All     d      All     d      10987  15634.
## 5  2953  20200531 S920~ d      All     d      All     d      11008  15458.
## 6  3088  20200607 S920~ d      All     d      All     d      11792  15490.
## 7  3221  20200614 S920~ d      All     d      All     d      11619  15508.
## 8  3357  20200621 S920~ d      All     d      All     d      12217  15556
## 9  3493  20200628 S920~ d      All     d      All     d      12358  15455
## 10 3633  20200705 S920~ d      All     d      All     d      12034  15082.
## # ... with 125 more rows, 12 more variables: percent_variation <dbl>,
## #   hb_name <chr>, hb_date_enacted <dbl>, hb_date_archived <dbl>,
## #   country <chr>, year <int>, month <int>, day <int>, wdate <date>,
## #   iswinter <lgl>, above_thresh <dbl>, moving_avg <dbl>, and abbreviated
## #   variable names 1: admission_type, 2: admission_type_qf, 3: speciality,
## #   4: speciality_qf, 5: number_admissions, 6: average20182019

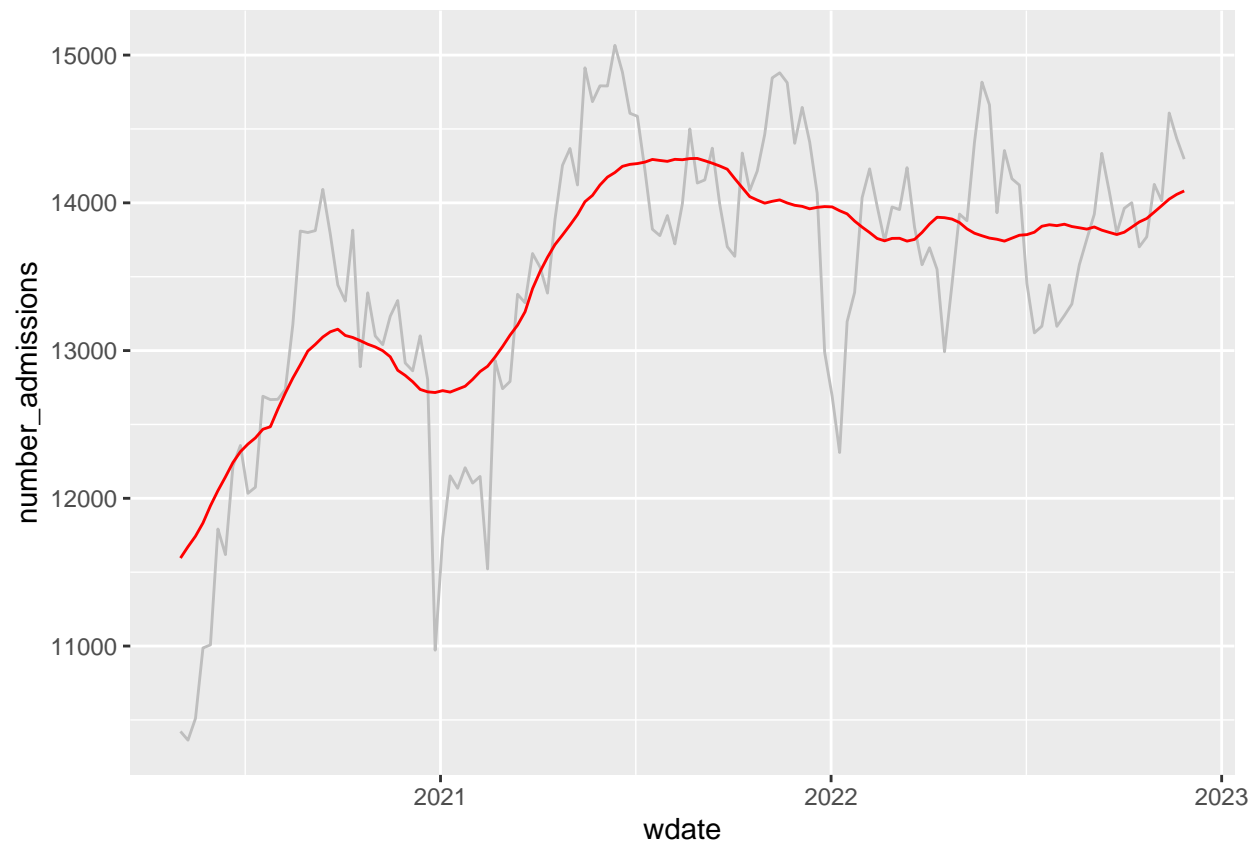
```

plot - check the smoother makes sense and tweak values if they dont

```

ggplot(data_rolling)+
  geom_line(aes(x = wdate, y = number_admissions), colour = "gray") +
  geom_line(aes(x = wdate, y = moving_avg), colour = "red")

```



this is a smoothed version of weekly data - its not the right way to create monthly values

take a look at what is in each dataset

```
weekly_admissions_spec %>%
  distinct(hb_name)
```

```
## # A tibble: 16 x 1
##   hb_name
##   <chr>
## 1 NHS Ayrshire and Arran
## 2 NHS Borders
## 3 NHS Dumfries and Galloway
## 4 NHS Forth Valley
## 5 NHS Grampian
## 6 NHS Highland
## 7 NHS Lothian
## 8 NHS Orkney
## 9 NHS Shetland
## 10 NHS Western Isles
## 11 NHS Fife
## 12 NHS Tayside
## 13 NHS Greater Glasgow and Clyde
## 14 NHS Lanarkshire
## 15 All Scotland
## 16 NHS Region Unknown
```

14 health boards plus 'all scotland' and an NA - need to work out what to do with this

```
weekly_admissions_spec %>%  
  distinct(admission_type)
```

```
## # A tibble: 3 x 1  
##   admission_type  
##   <chr>  
## 1 All  
## 2 Emergency  
## 3 Planned
```

Data neatly divided into emergency and planned

```
weekly_admissions_spec %>%  
  distinct(speciality)
```

```
## # A tibble: 13 x 1  
##   speciality  
##   <chr>  
## 1 Accident & Emergency  
## 2 All  
## 3 Cardiology  
## 4 Medical (excl. Cardiology & Cancer)  
## 5 Medical (incl. Cardiology & Cancer)  
## 6 Paediatrics (medical & surgical)  
## 7 Paediatrics (medical)  
## 8 Surgery  
## 9 Community  
## 10 Gynaecology  
## 11 Cancer  
## 12 Paediatrics (surgical)  
## 13 <NA>
```

note that some groupings are also combined. need to take a closer look at what these mean.

```
weekly_admissions_demog %>%  
  distinct(age_group)
```

```
## # A tibble: 8 x 1  
##   age_group  
##   <chr>  
## 1 Under 5  
## 2 5 - 14  
## 3 15 - 44  
## 4 45 - 64  
## 5 65 - 74  
## 6 75 - 84  
## 7 85 and over  
## 8 All ages
```

7 age groups and "all ages"

```
weekly_admissions_demog %>%  
  distinct(sex)
```

```
## # A tibble: 3 x 1  
##   sex  
##   <chr>  
## 1 All  
## 2 Female  
## 3 Male
```

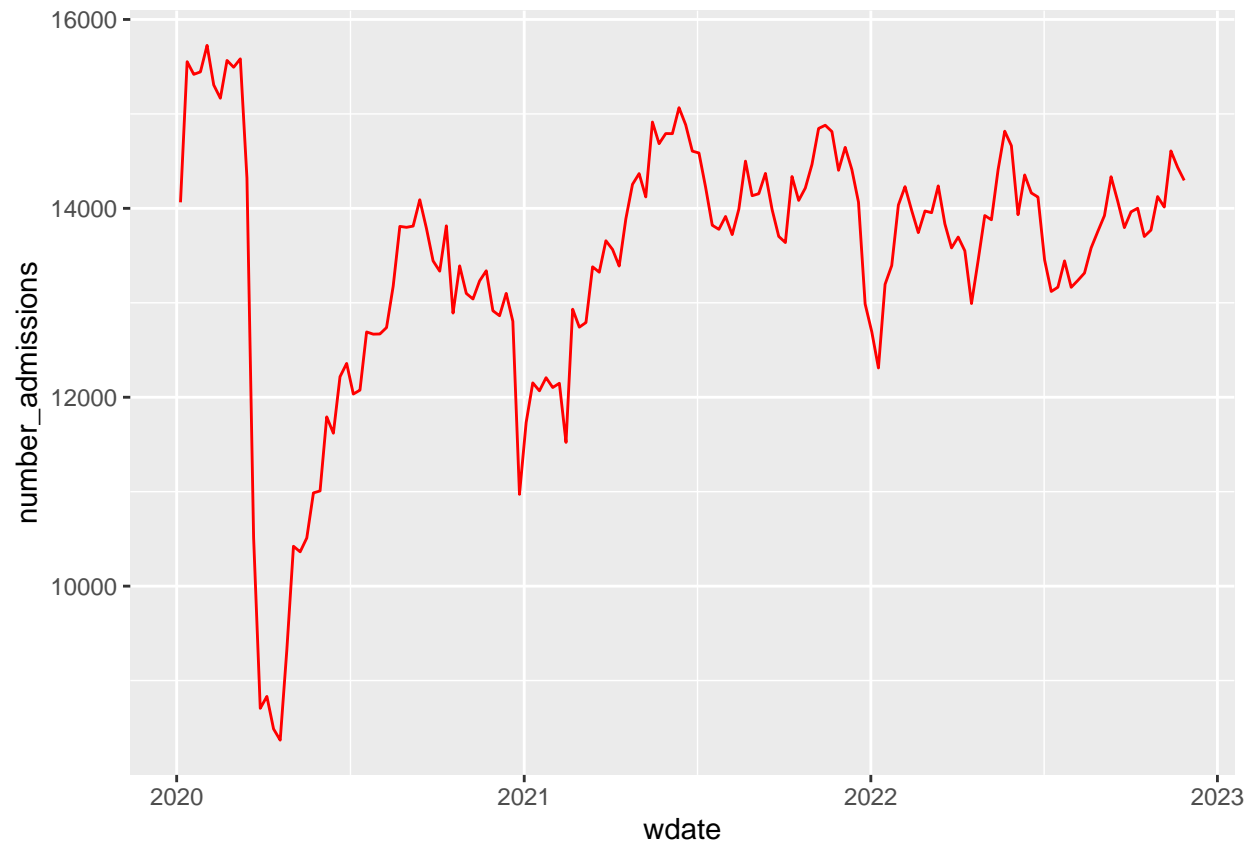
just male - female and all

```
weekly_admissions_dep %>%  
  distinct(simd_quintile)
```

```
## # A tibble: 5 x 1  
##   simd_quintile  
##           <dbl>  
## 1             1  
## 2             2  
## 3             3  
## 4             4  
## 5             5
```

note there is no 'all' category in this dataset

```
weekly_admissions_spec %>%  
  filter(speciality == "All") %>%  
  filter(admission_type == "All") %>%  
  filter(hb_name == "All Scotland") %>%  
  ggplot() +  
  aes(x=wdate, y = number_admissions) +  
  geom_line(colour='red')
```

weekly hospital admissions hover around 14000 per week expect 168,000 per quarter?

pre=pandemic hospital admissions hovered around 15500 per week expect 186,000 per quarter

Overall(all specialities, all admissions, all healthboards) admissions are only at 90% of pre-pandemic levels

quick look at demographics

```
weekly_admissions_demog %>%
  filter(sex == "All") %>%
  filter(age_group != "All ages") %>%
  filter(hb_name == "All Scotland") %>%
  filter(admission_type == "All") %>%
ggplot() +
  aes(x=wdate, y = number_admissions, color = age_group) +
  geom_line() #+
```



```
#scale_colour_brewer(palette = "Set2")
```

age plot shows some variation between age groups.

quick look at sex differences

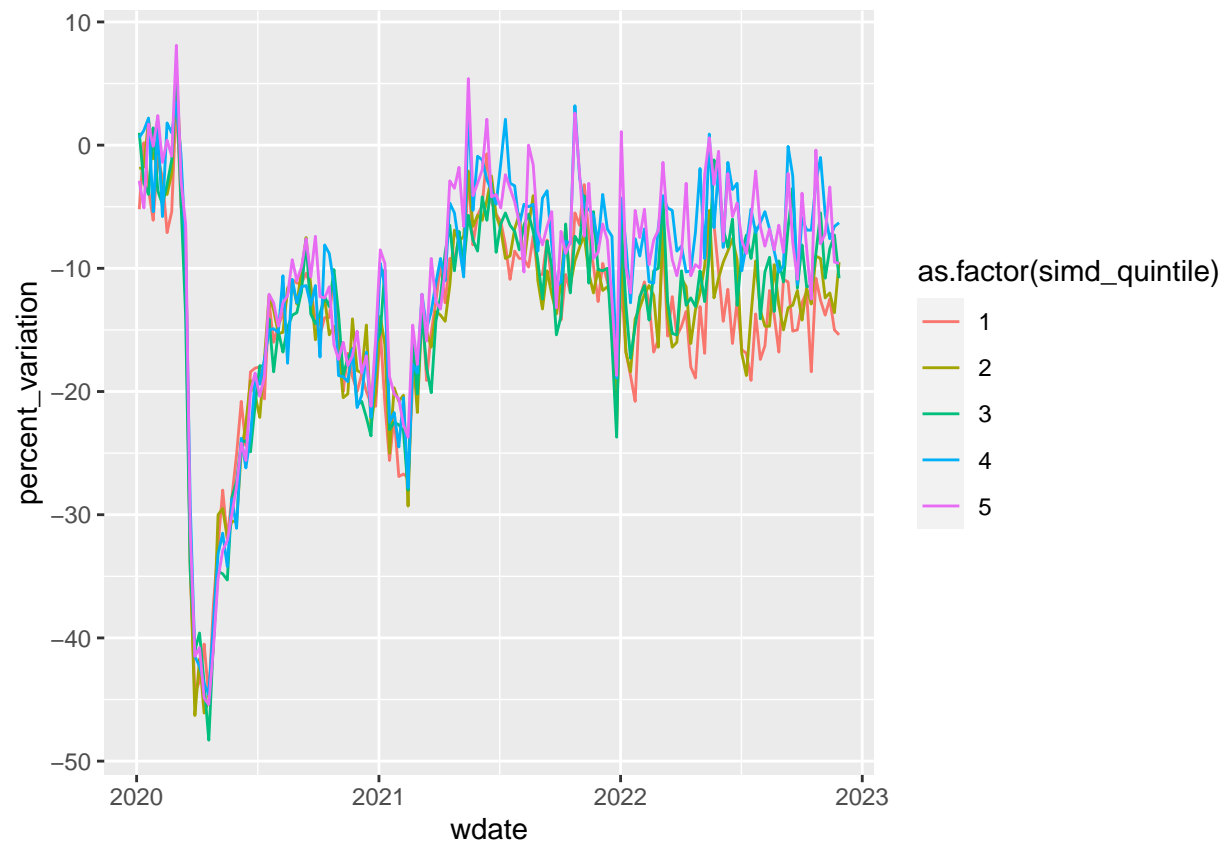
```
weekly_admissions_demog %>%
  filter(sex != "All") %>%
  filter(age_group == "All ages") %>%
  filter(hb_name == "All Scotland") %>%
  filter(admission_type == "All") %>%
  ggplot() +
  aes(x=wdate, y = percent_variation, color = sex) +
  geom_line() #+
```



```
#scale_colour_brewer(palette = "Set2")
```

quick look at index of deprivation. 1 is most deprived and 5 is least deprived

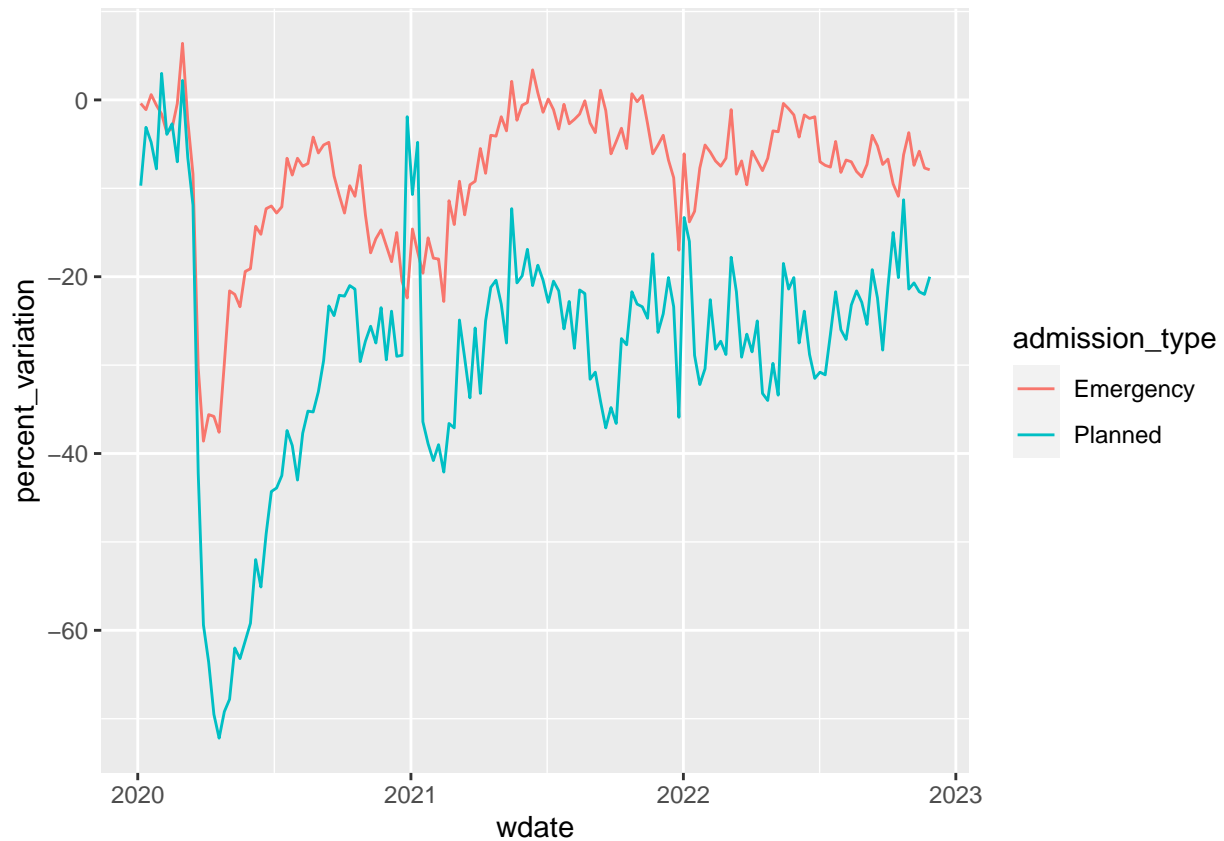
```
weekly_admissions_dep %>%
  filter(admission_type == "All") %>%
  filter(hb_name == "All Scotland") %>%
  ggplot() +
  aes(x=wdate, y = percent_variation, color = as.factor(simd_quintile)) +
  geom_line()
```



something happening in 2022 - why is there divergence between 1 and 5?

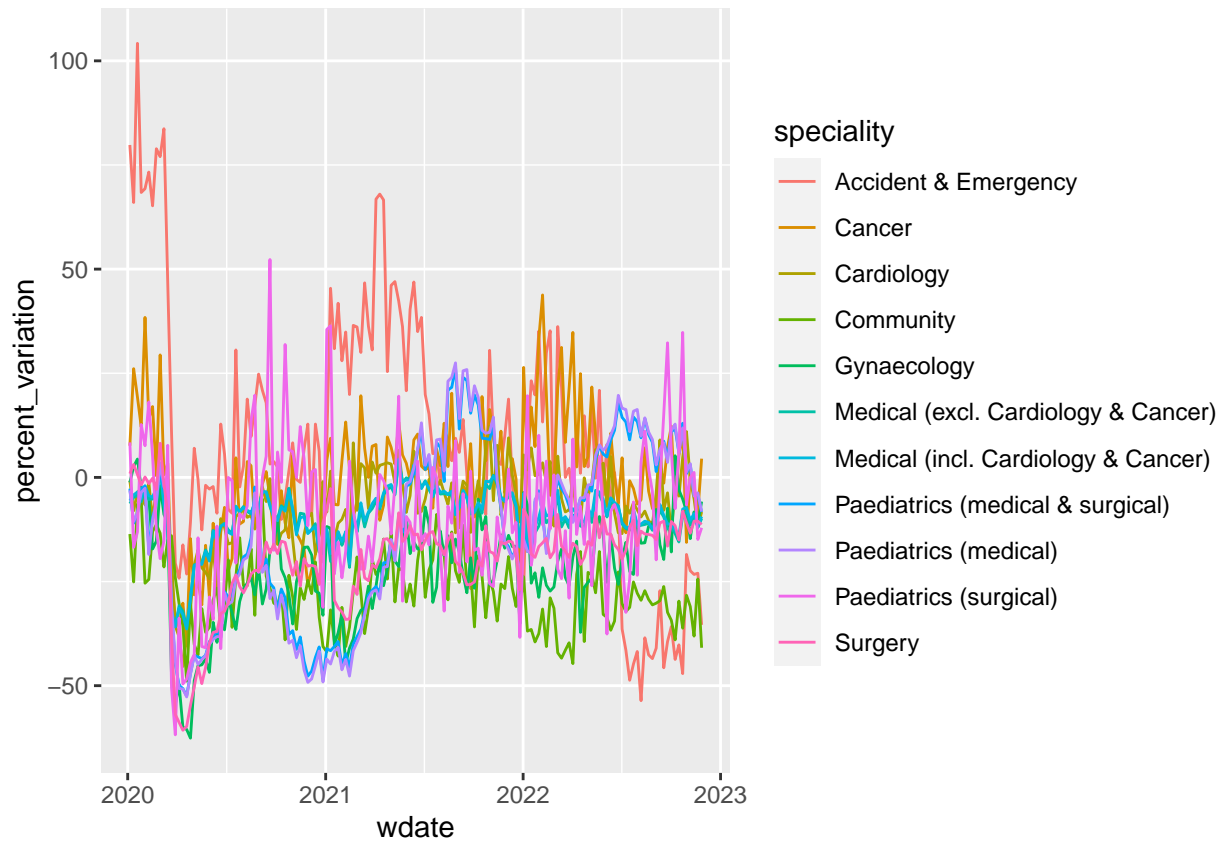
look at admissions type

```
weekly_admissions_spec %>%
  filter(admission_type != "All") %>%
  filter(speciality=="All") %>%
  filter(hb_name == "All Scotland") %>%
  ggplot() +
  aes(x=wdate, y = percent_variation, colour = admission_type) +
  geom_line()
```



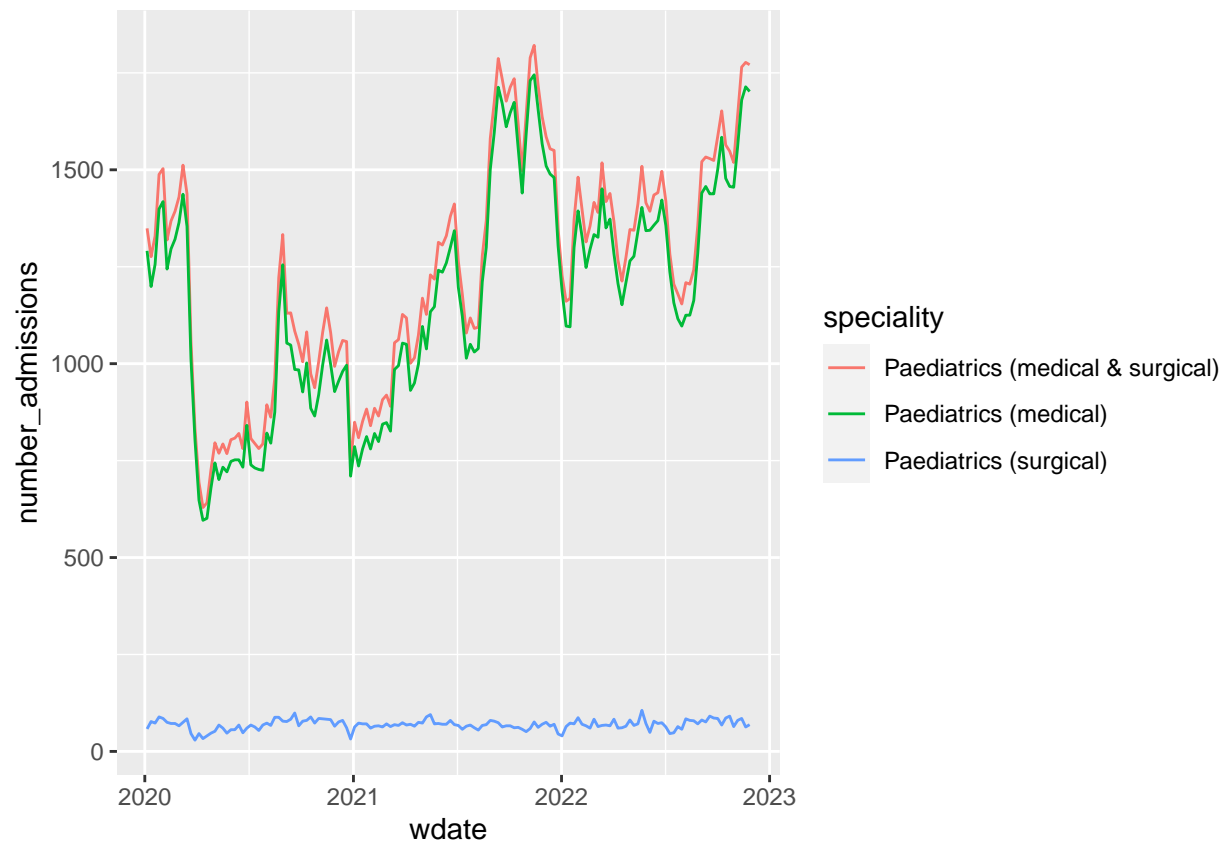
need to check what the 'spike' in planned at end of 2021 is related to
 quick look at specialities

```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality != "All") %>%
  filter(hb_name == "All Scotland") %>%
  ggplot() +
  aes(x=wdate, y = percent_variation, colour=speciality) +
  geom_line()
```

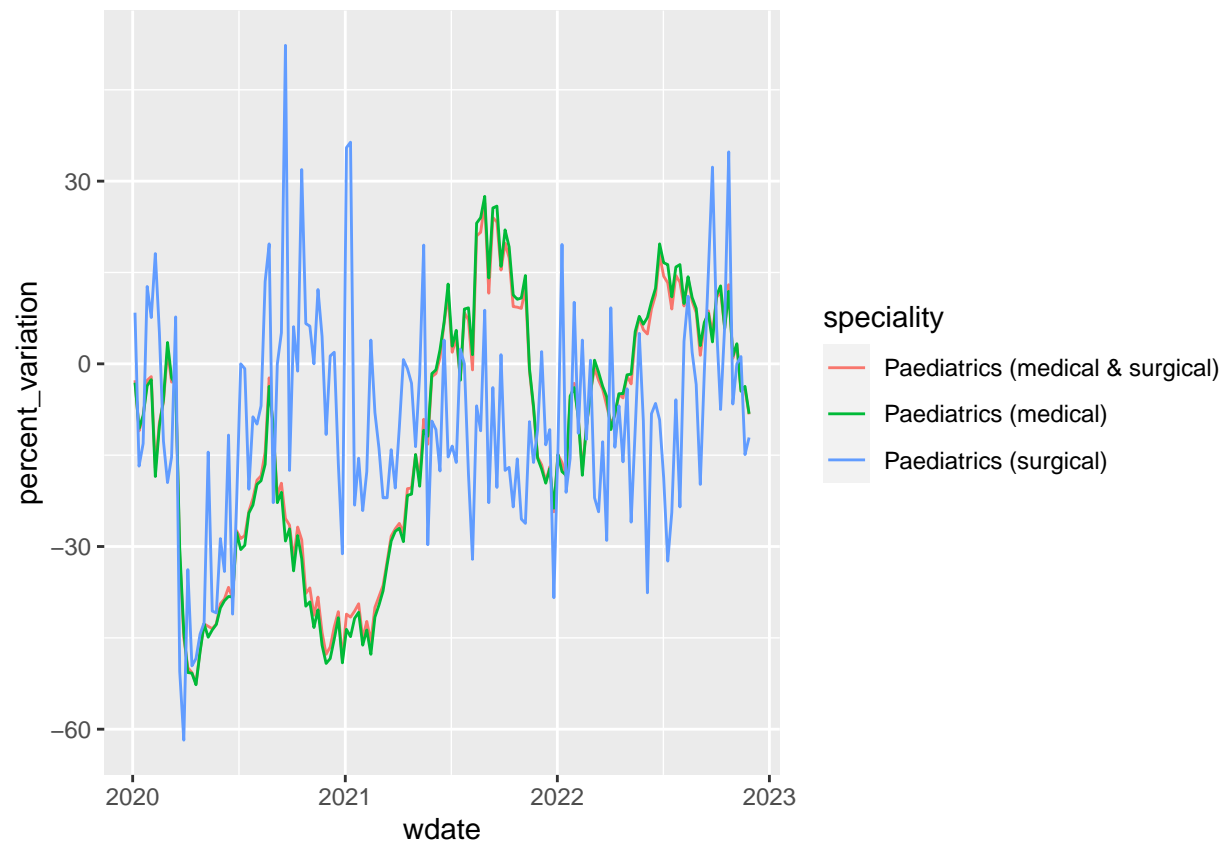


take a closer look at paediatrics

```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality != "All") %>%
  filter(str_detect(speciality, "Paed")) %>%
  filter(hb_name == "All Scotland") %>%
ggplot() +
  aes(x=wdate, y = number_admissions, colour=speciality) +
  geom_line()
```

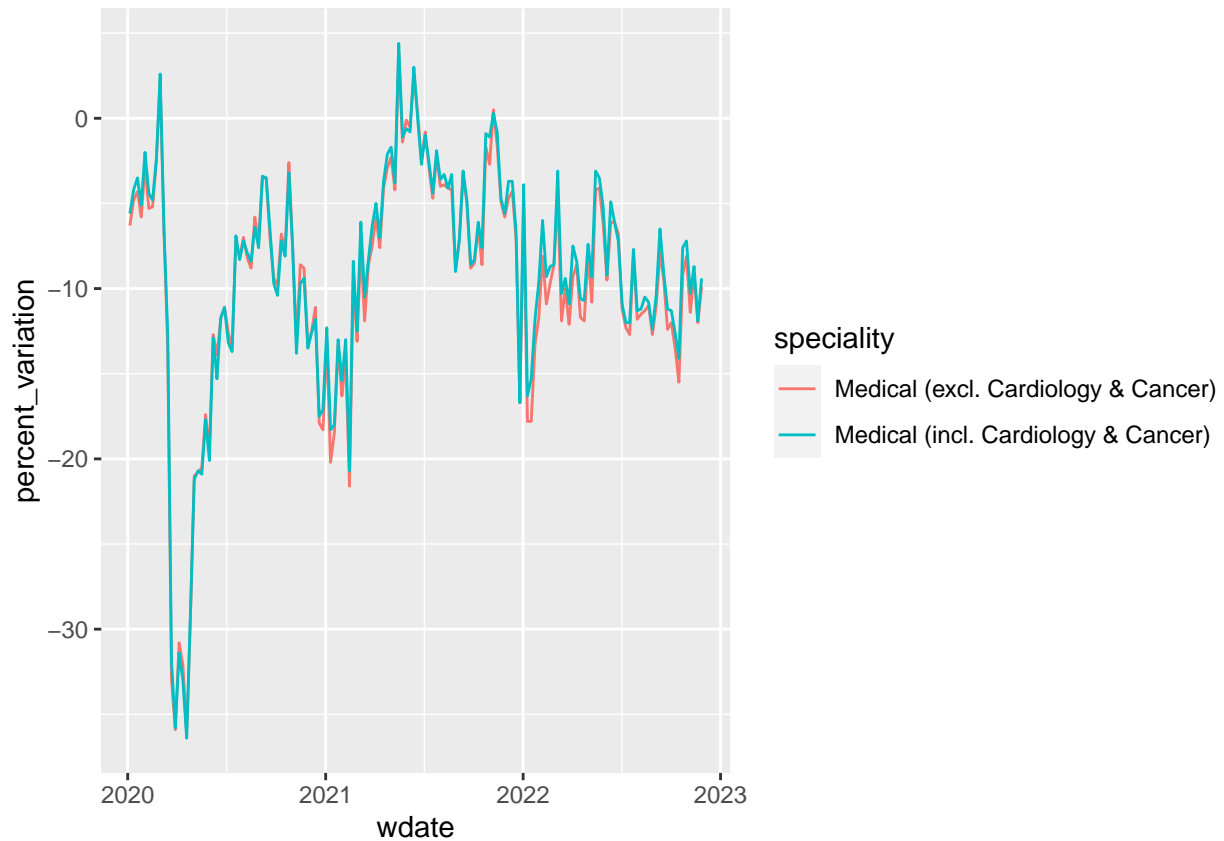


```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality != "All") %>%
  filter(str_detect(speciality, "Paed")) %>%
  filter(hb_name == "All Scotland") %>%
ggplot() +
  aes(x=wdate, y = percent_variation, colour=speciality) +
  geom_line()
```



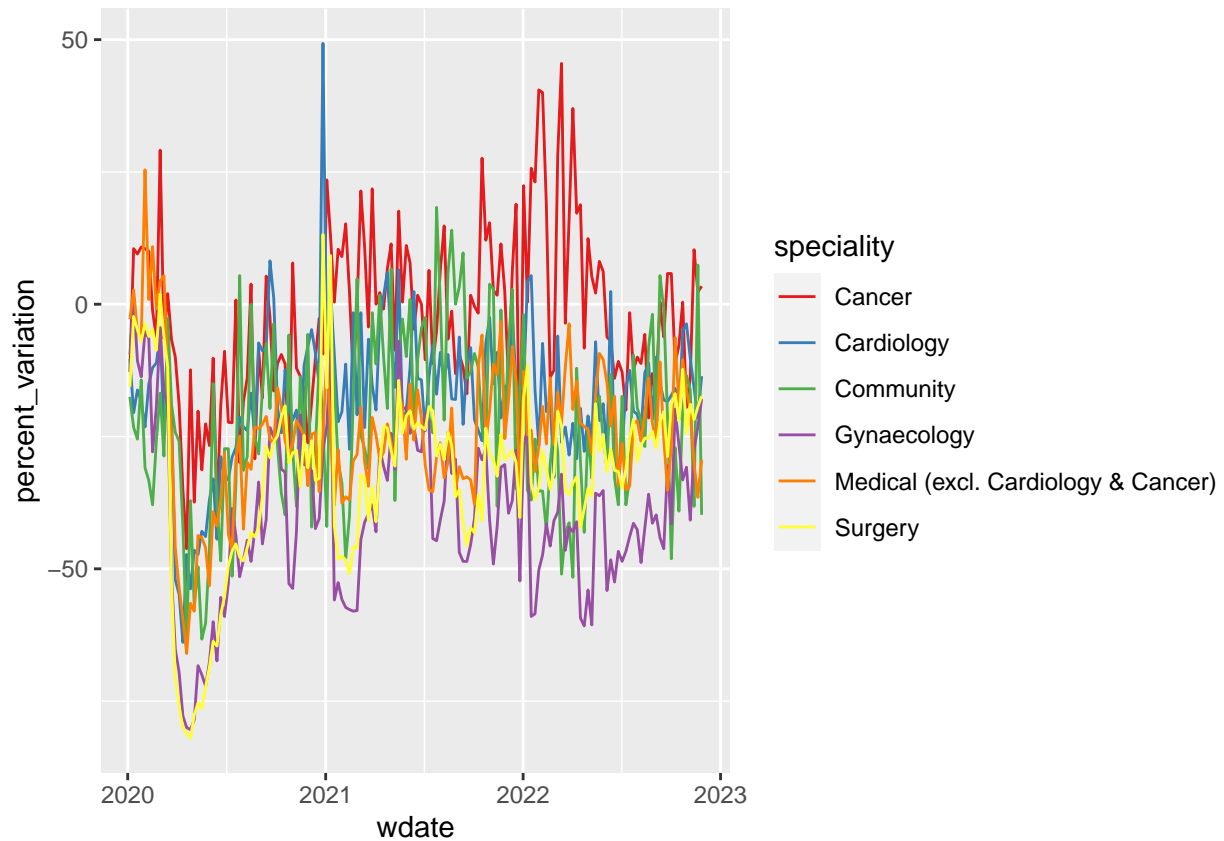
take a closer look at medical

```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality != "All") %>%
  filter(str_detect(speciality, "Medical")) %>%
  filter(hb_name == "All Scotland") %>%
ggplot() +
  aes(x=wdate, y = percent_variation, colour=speciality) +
  geom_line()
```

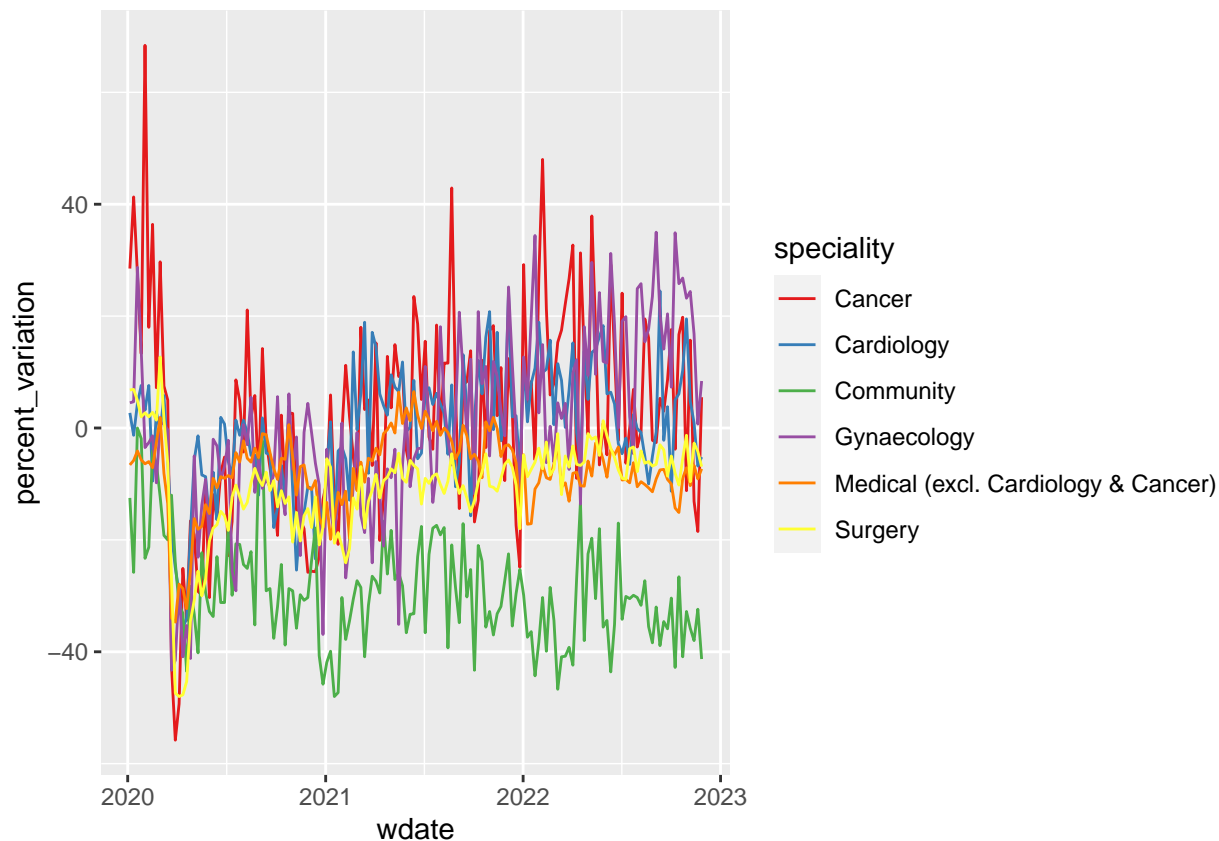
simpler plot of specialities

```
weekly_admissions_spec %>%
  filter(admission_type == "Planned") %>%
  filter(speciality != "Medical (incl. Cardiology & Cancer)") %>%
  filter(!str_detect(speciality, "Paed")) %>%
  filter(speciality != "All") %>%
  filter(speciality != "Accident & Emergency") %>%
  filter(hb_name == "All Scotland") %>%
  ggplot() +
  aes(x=wdate, y = percent_variation, colour=speciality) +
  geom_line() +
  scale_colour_brewer(palette = "Set1")
```



we cant present data like this - its just too busy. how would we illustrate these differences - better with a barplot

```
weekly_admissions_spec %>%
  filter(admission_type == "Emergency") %>%
  filter(speciality != "Medical (incl. Cardiology & Cancer)") %>%
  filter(!str_detect(speciality, "Paed")) %>%
  filter(speciality != "All") %>%
  filter(speciality != "Accident & Emergency") %>%
  filter(hb_name == "All Scotland") %>%
ggplot() +
  aes(x=wdate, y = percent_variation, colour=speciality) +
  geom_line() +
  scale_colour_brewer(palette = "Set1")
```



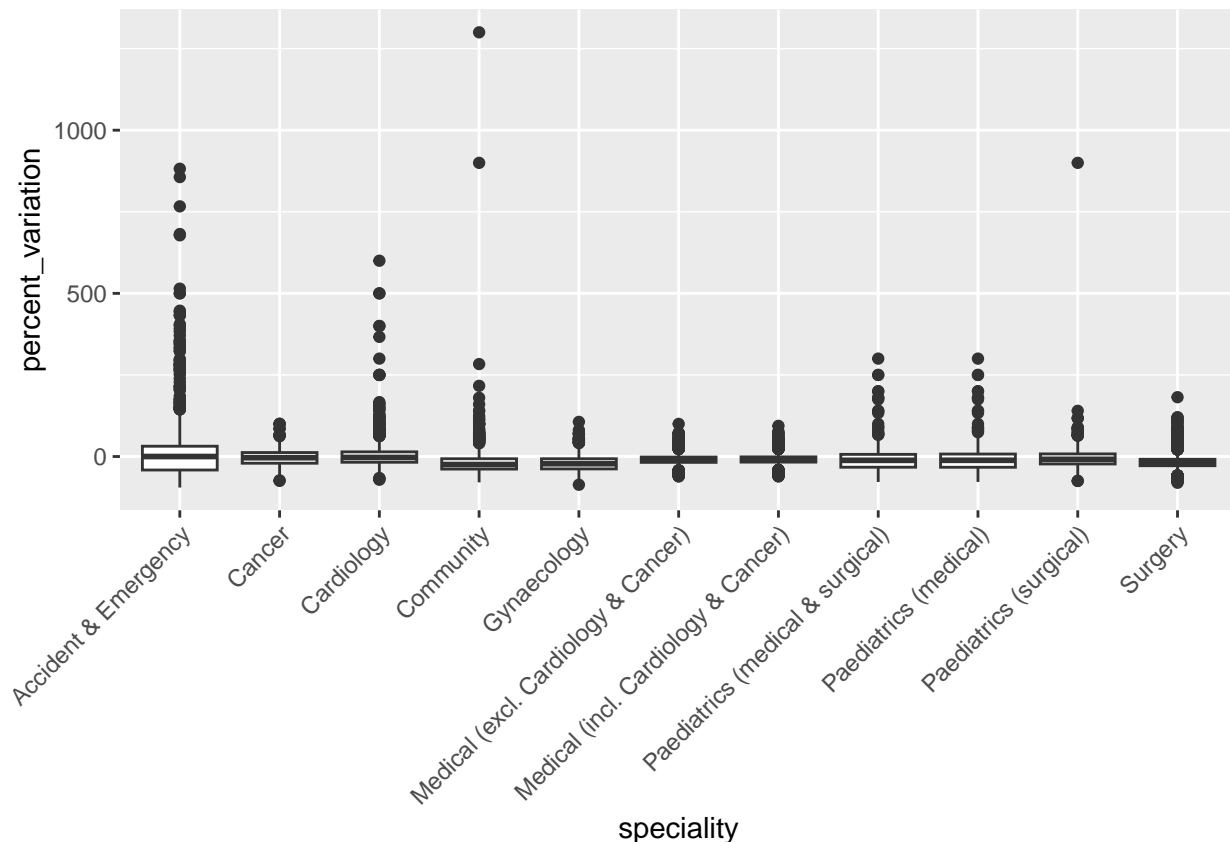
boxplot for different specialities? some values greater than 8000% (i have filtered them out) indicates we need some more work on data cleaning!

```
weekly_admissions_spec %>%
  #filter(admission_type == "All") %>%
  #filter(speciality!="All") %>%
  #filter(hb_name == "NHS Region Unknown") %>%
  #group_by(speciality) %>%
  #need to take a good look at what these values are!
  filter(percent_variation>500)
```

```
## # A tibble: 230 x 21
##       id week_ending hb      hbqf admis~1 admis~2 speci~3 speci~4 numbe~5 avera~6
##   <dbl>      <dbl> <chr> <chr> <chr>    <chr> <chr>    <chr>    <dbl>    <dbl>
## 1 325    20200119 S080~ <NA> All     d      Accide~ <NA>      263     27.5
## 2 3429   20200628 S080~ <NA> All     d      Commun~ <NA>       5       0
## 3 3507   20200628 <NA> <NA> All     d      Gynaec~ <NA>       5       0
## 4 3570   20200705 S080~ <NA> All     d      Commun~ <NA>       5       0
## 5 3977   20200726 S080~ <NA> All     d      Commun~ <NA>       8       0
## 6 4118   20200802 S080~ <NA> All     d      Commun~ <NA>       8       0
## 7 4255   20200809 S080~ <NA> All     d      Commun~ <NA>       9       0
## 8 4333   20200809 <NA> <NA> All     d      Gynaec~ <NA>       5       0
## 9 4396   20200816 S080~ <NA> All     d      Commun~ <NA>       7       0
## 10 4535  20200823 S080~ <NA> All     d      Commun~ <NA>       6       0
## # ... with 220 more rows, 11 more variables: percent_variation <dbl>,
## #   hb_name <chr>, hb_date_enacted <dbl>, hb_date_archived <dbl>,
```

```
## # country <chr>, year <int>, month <int>, day <int>, wdate <date>,
## # iswinter <lgl>, above_thresh <dbl>, and abbreviated variable names
## # 1: admission_type, 2: admission_type_qf, 3: speciality, 4: speciality_qf,
## # 5: number_admissions, 6: average20182019
```

```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality!="All") %>%
  group_by(speciality) %>%
  #need to take a good look at what these values are!
  filter(percent_variation<8000) %>%
  ggplot()+
  aes(x=speciality, y=percent_variation)+
  geom_boxplot()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

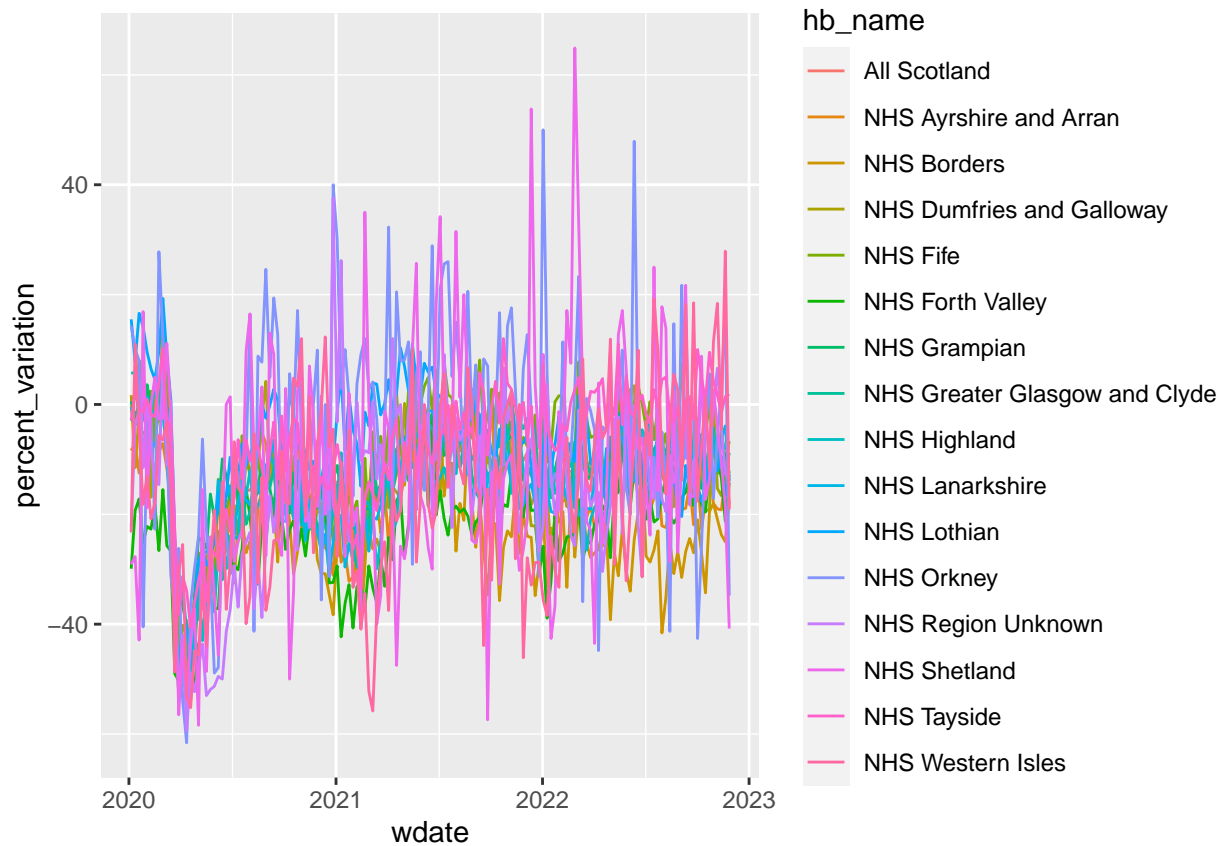


some specialities have faced higher admission levels than others.

Now look by health boards. high variability in some health boards. is this true variability - e.g indicator of crisis period or just data artefact? values are small so more affected by transformation to %percentage var. especially when related to weekly values

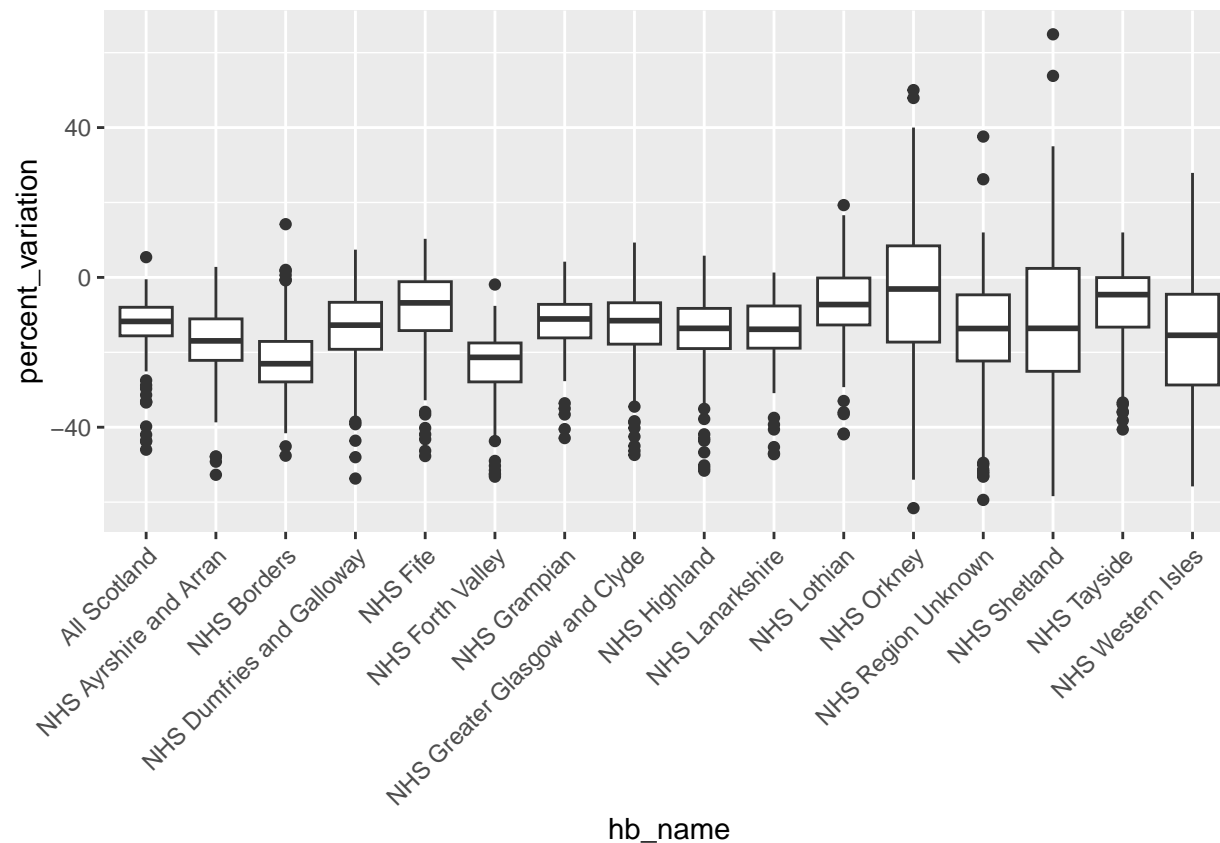
```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality=="All") %>%
  group_by(hb_name) %>%
  ggplot()+
```

```
aes(x=wdate, y=percent_variation, colour = hb_name)+
geom_line()
```



try different visualisations

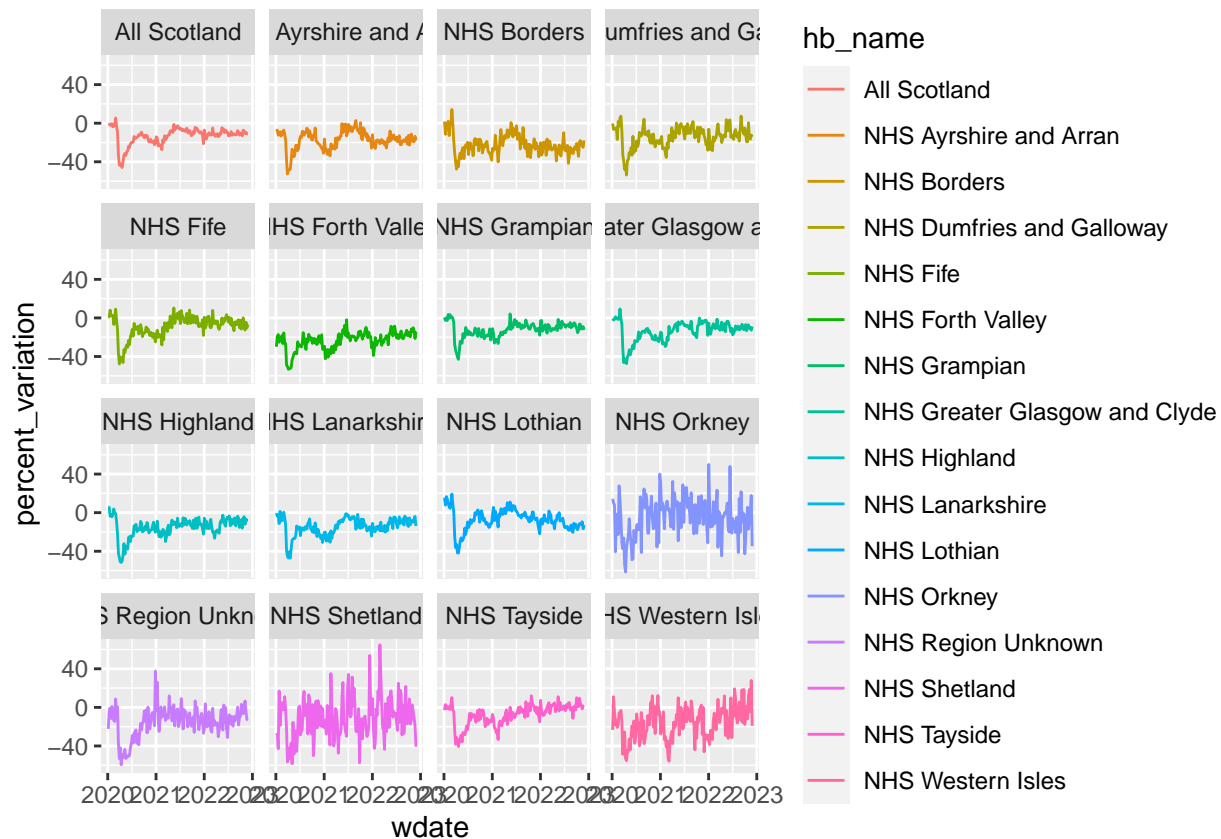
```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality=="All") %>%
  group_by(hb_name) %>%
  ggplot()+
  aes(x=hb_name, y=percent_variation)+
  geom_boxplot()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



for some health boards a significant number of admissions levels are above the levels seen in 2018-19. could we perhaps try and use this as an indicator?

Need to think more carefully about what the data in 'region unknown' means

```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality=="All") %>%
  ggplot() +
  aes(x=wdate, y=percent_variation, colour=hb_name) +
  geom_line()+
  facet_wrap(~hb_name)
```



How is NA calculated - still need to check??

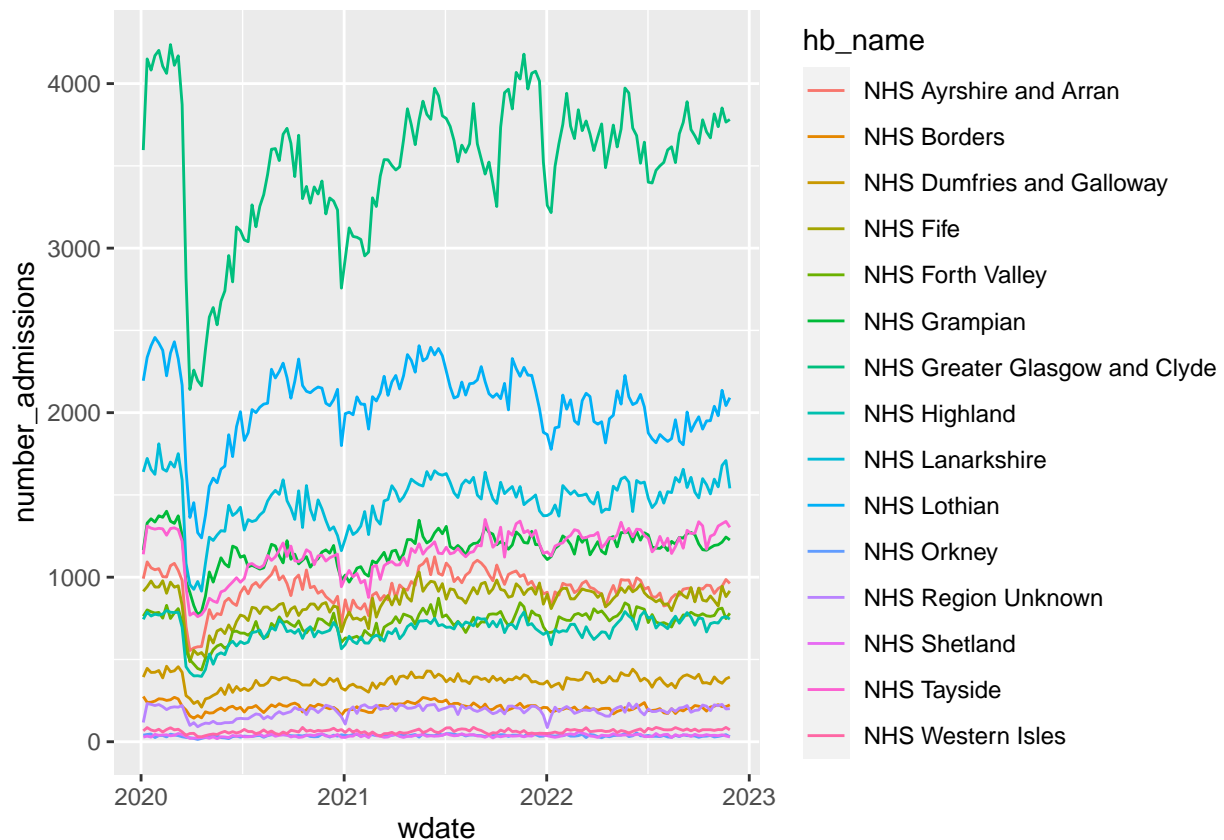
summarise admissions by health board - need to show how they differ.

```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality == "All") %>%
  filter(hb_name != "All Scotland") %>%
  group_by(hb_name) %>%
  summarise(mean = mean(number_admissions)) %>%
  arrange(desc(mean))
```

```
## # A tibble: 15 x 2
##   hb_name                mean
##   <chr>                  <dbl>
## 1 NHS Greater Glasgow and Clyde 3522.
## 2 NHS Lothian                2052.
## 3 NHS Lanarkshire            1468.
## 4 NHS Grampian               1170.
## 5 NHS Tayside                1154.
## 6 NHS Ayrshire and Arran      932.
## 7 NHS Fife                   849.
## 8 NHS Forth Valley           721.
## 9 NHS Highland               679.
## 10 NHS Dumfries and Galloway  367.
## 11 NHS Borders                210.
## 12 NHS Region Unknown         186.
```

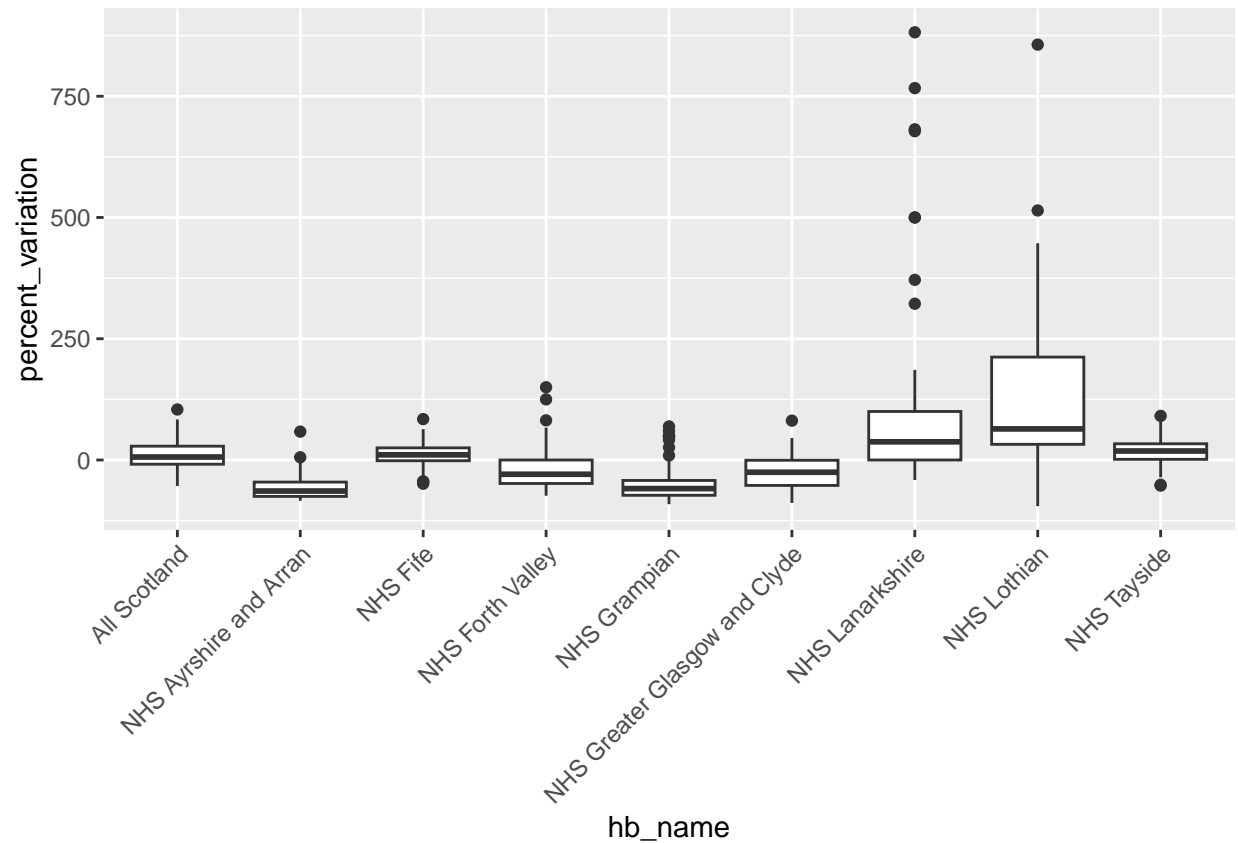
```
## 13 NHS Western Isles          63.8
## 14 NHS Shetland              36.6
## 15 NHS Orkney                36.5
```

```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality == "All") %>%
  filter(hb_name != "All Scotland") %>%
ggplot() +
  aes(x=wdate, y = number_admissions, colour = hb_name) +
  geom_line()
```



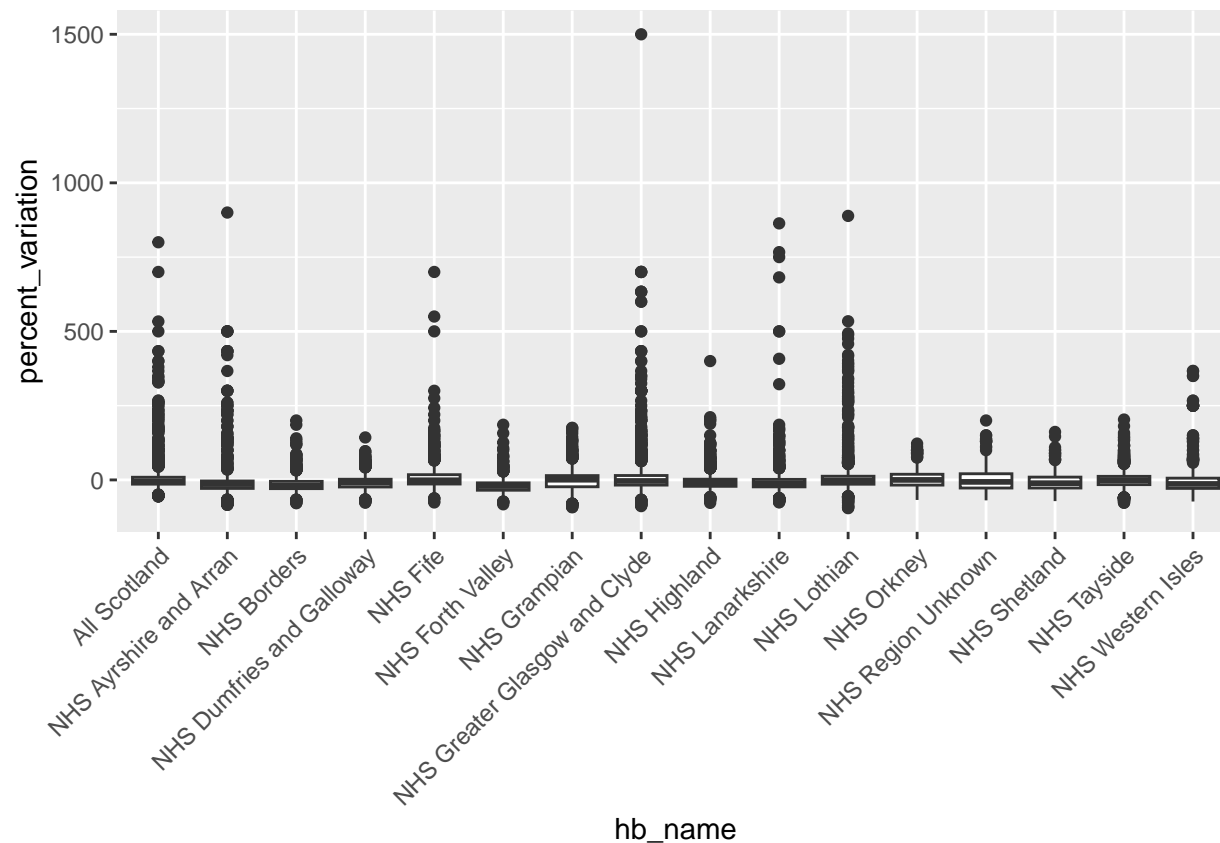
admission values shows all health boards - health boards vary massively in size and admission numbers but mostly same overall pattern.

```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality=="Accident & Emergency") %>%
  #take a good look at what these values are!
  filter(percent_variation<8000) %>%
  group_by(hb_name) %>%
  ggplot()+
  aes(x=hb_name, y=percent_variation)+
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

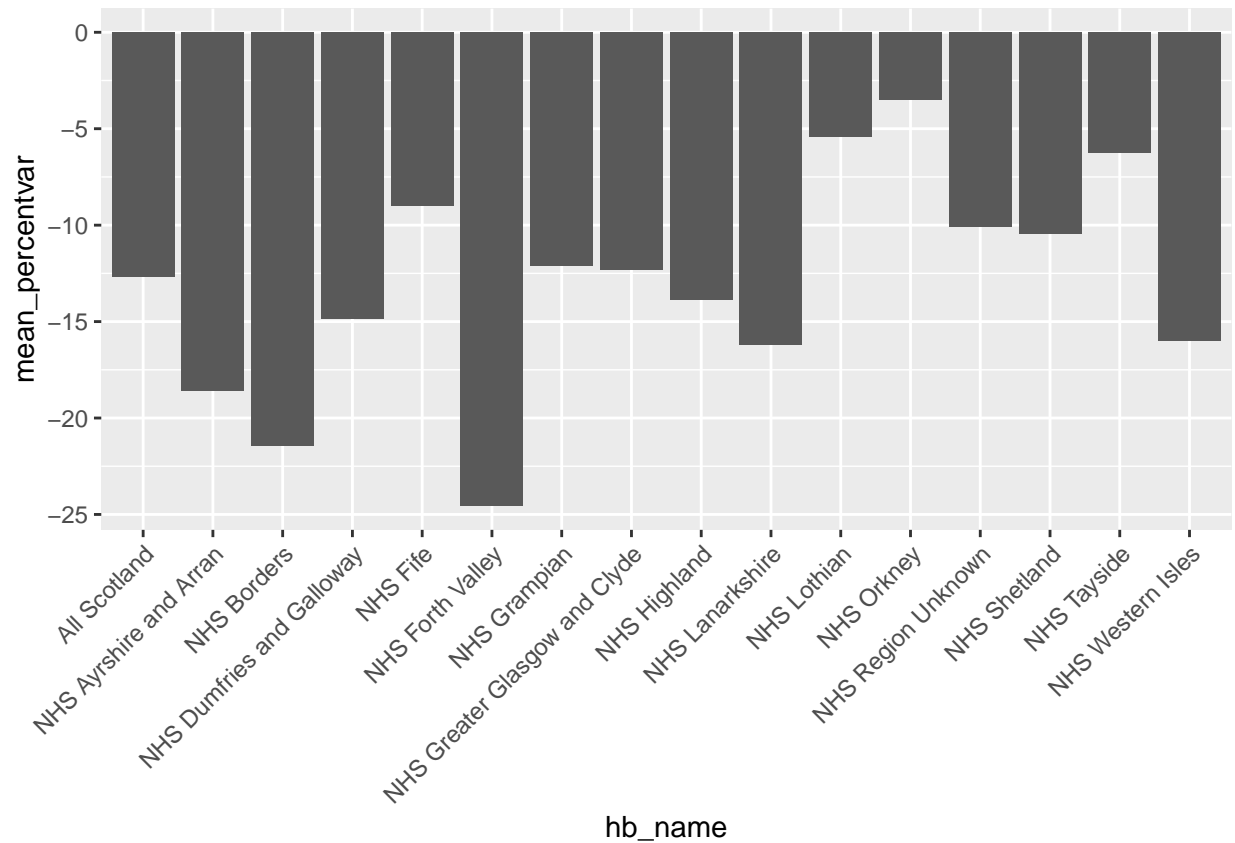
do only certain health boards have an A+E dept?

```
weekly_admissions_spec %>%
  filter(admission_type == "Emergency") %>%
  #filter(speciality=="Accident & Emergency") %>%
  #take a good look at what these values are!
  filter(percent_variation<8000) %>%
  group_by(hb_name) %>%
  ggplot()+
  aes(x=hb_name, y=percent_variation)+
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



first attempt at a 'crisis' calculation. this will be affected by variability - need a smoothed or monthly version for summary

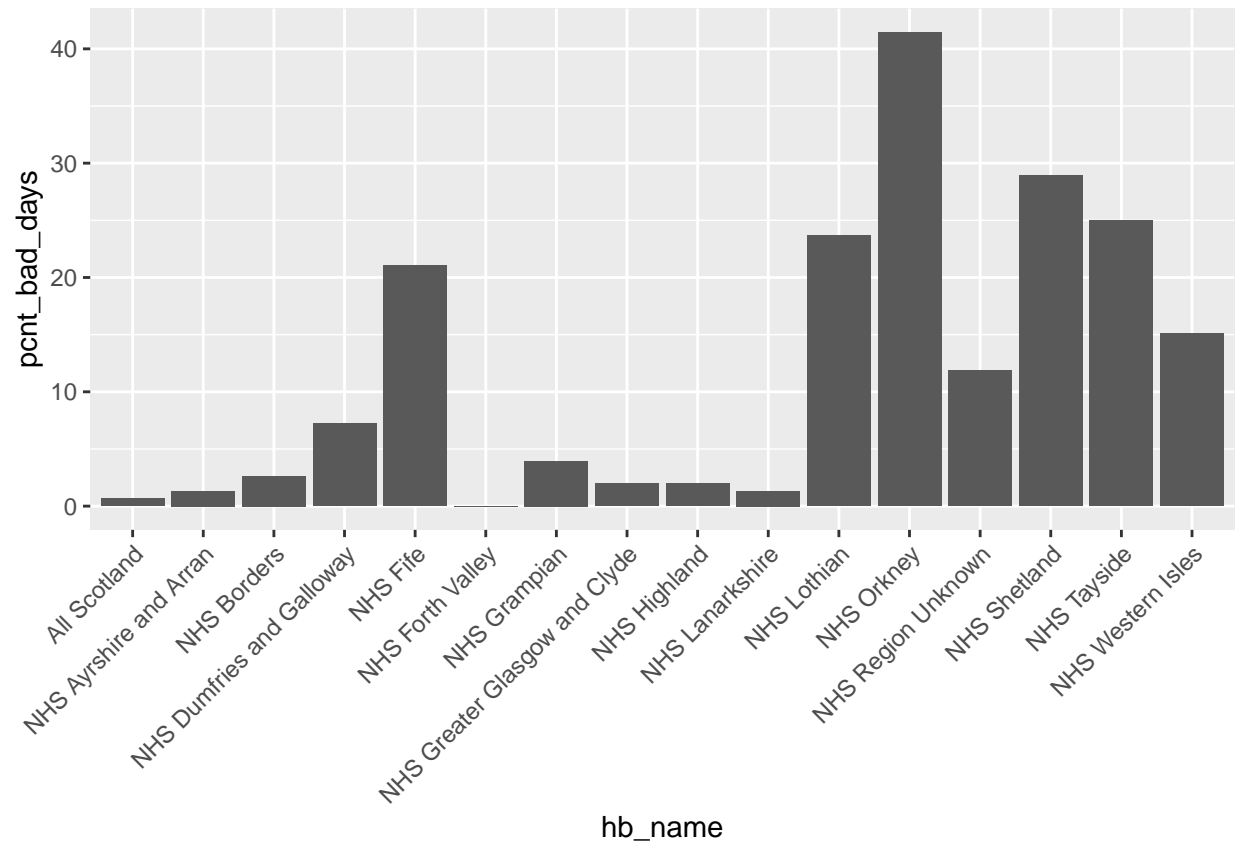
```
weekly_admissions_spec %>%
  filter(admission_type == "All") %>%
  filter(speciality=="All") %>%
  #filter(admission_type == "Emergency") %>%
  #filter(speciality=="Accident & Emergency") %>%
  #filter(year == 2020) %>%
  filter(iswinter) %>%
  group_by(hb_name) %>%
  summarise(count_bad_days = sum(above_thresh),
            count_days = n()*7,
            pcnt_bad_days = 100*(sum(above_thresh)/(n()*7)),
            mean_percentvar = mean(percent_variation)) %>%
  ggplot() +
  aes(x=hb_name, y=mean_percentvar) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```

weekly_admissions_spec %>%
  filter(speciality=="All") %>%
  filter(admission_type=="All") %>%
  #filter(year == 2021) %>%
  #filter(iswinter) %>%
  group_by(hb_name) %>%
  summarise(count_bad_days = sum(above_thresh),
            count_days = n()*7,
            pcnt_bad_days = 100*(sum(above_thresh)/(n()*7)),
            mean_percentvar = mean(percent_variation)) %>%
  ggplot() +
  aes(x=hb_name, y=pcnt_bad_days) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

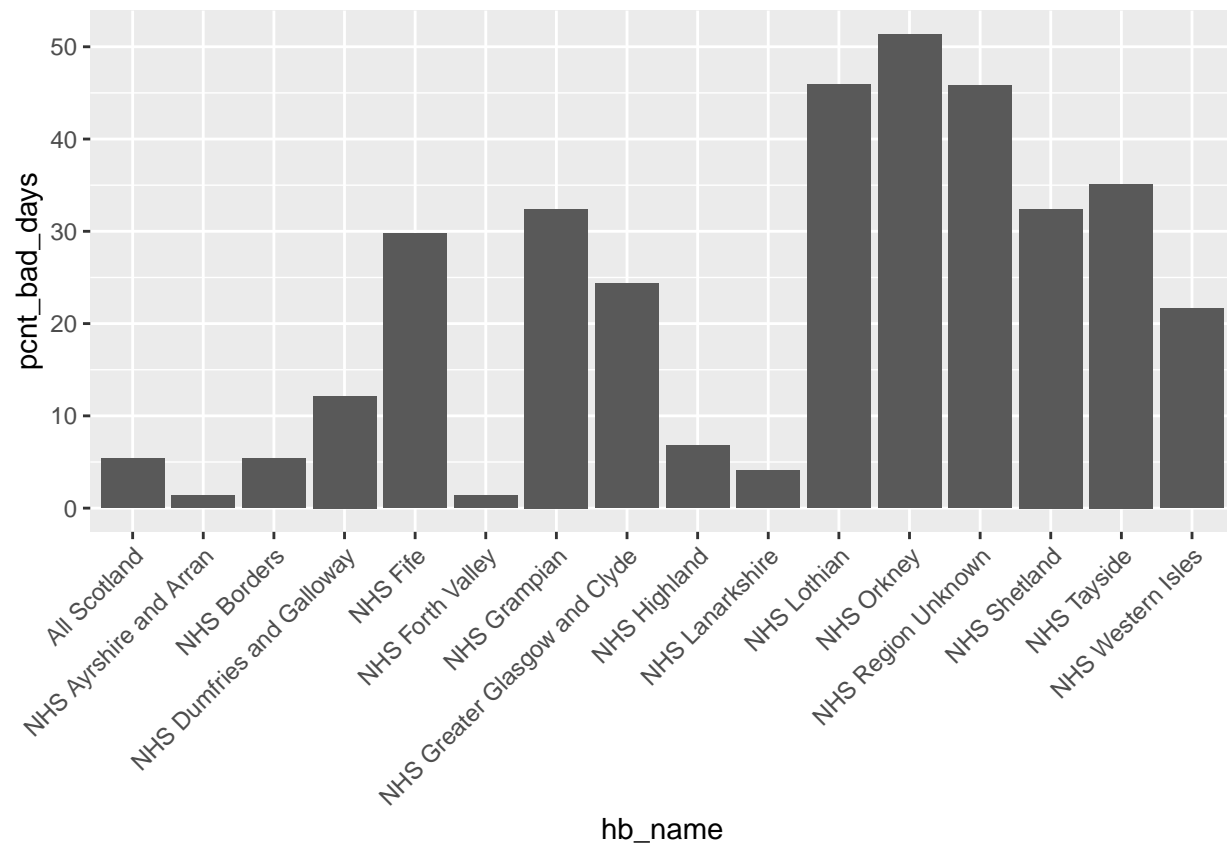
```



```

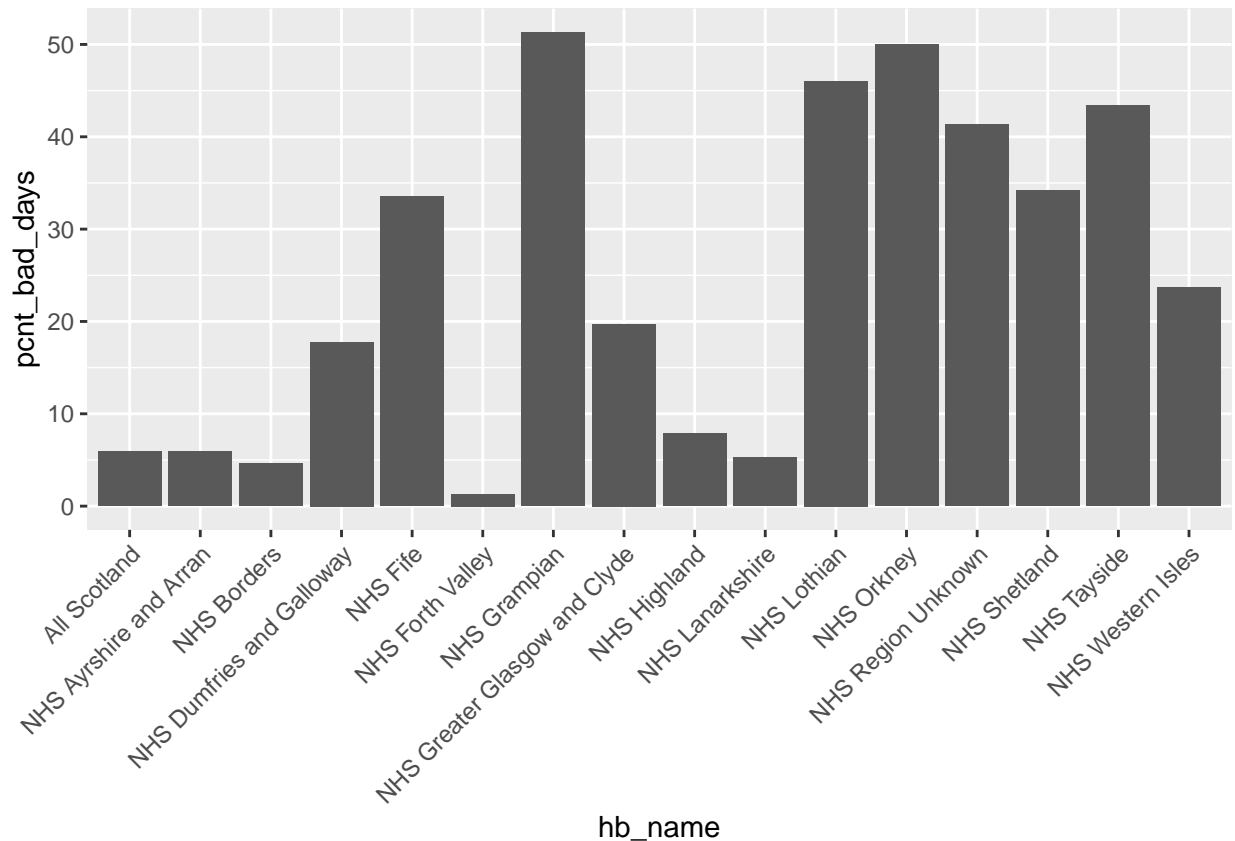
weekly_admissions_spec %>%
  filter(speciality=="All") %>%
  filter(admission_type=="Emergency") %>%
  #filter(speciality=="Accident & Emergency") %>%
  #filter(year == 2020) %>%
  filter(iswinter) %>%
  group_by(hb_name) %>%
  summarise(count_bad_days = sum(above_thresh),
            count_days = n()*7,
            pcnt_bad_days = 100*(sum(above_thresh)/(n()*7)),
            mean_percentvar =mean(percent_variation)) %>%
  ggplot() +
  aes(x=hb_name, y=pcnt_bad_days) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



take a look by health board

```
weekly_admissions_spec %>%
  filter(speciality=="All") %>%
  filter(admission_type=="Emergency") %>%
  #filter(speciality=="Accident & Emergency") %>%
  #filter(year == 2020) %>%
  #filter(iswinter) %>%
  group_by(hb_name) %>%
  summarise(count_bad_days = sum(above_thresh),
            count_days = n()*7,
            pcnt_bad_days = 100*(sum(above_thresh)/(n()*7)),
            mean_percentvar = mean(percent_variation)) %>%
  ggplot() +
  aes(x=hb_name, y=pcnt_bad_days) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



quick look at winter v summer

```
weekly_admissions_spec %>%
  filter(speciality=="Accident & Emergency") %>%
  filter(admission_type=="All") %>%
  group_by(iswinter, year) %>%
  summarise(pcnt_bad_days = sum(above_thresh)/n()*7, mean_percentvar =mean(percent_variation))
```

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

```
## # A tibble: 6 x 4
## # Groups:   iswinter [2]
##   iswinter year pcnt_bad_days mean_percentvar
##   <lg1>    <int>      <dbl>      <dbl>
## 1 FALSE   2020      21.4        1.35
## 2 FALSE   2021      28.1       17.6
## 3 FALSE   2022      19.1       -13.1
## 4 TRUE    2020      28.3       43.7
## 5 TRUE    2021      25.0       14.4
## 6 TRUE    2022      21.1        2.10
```

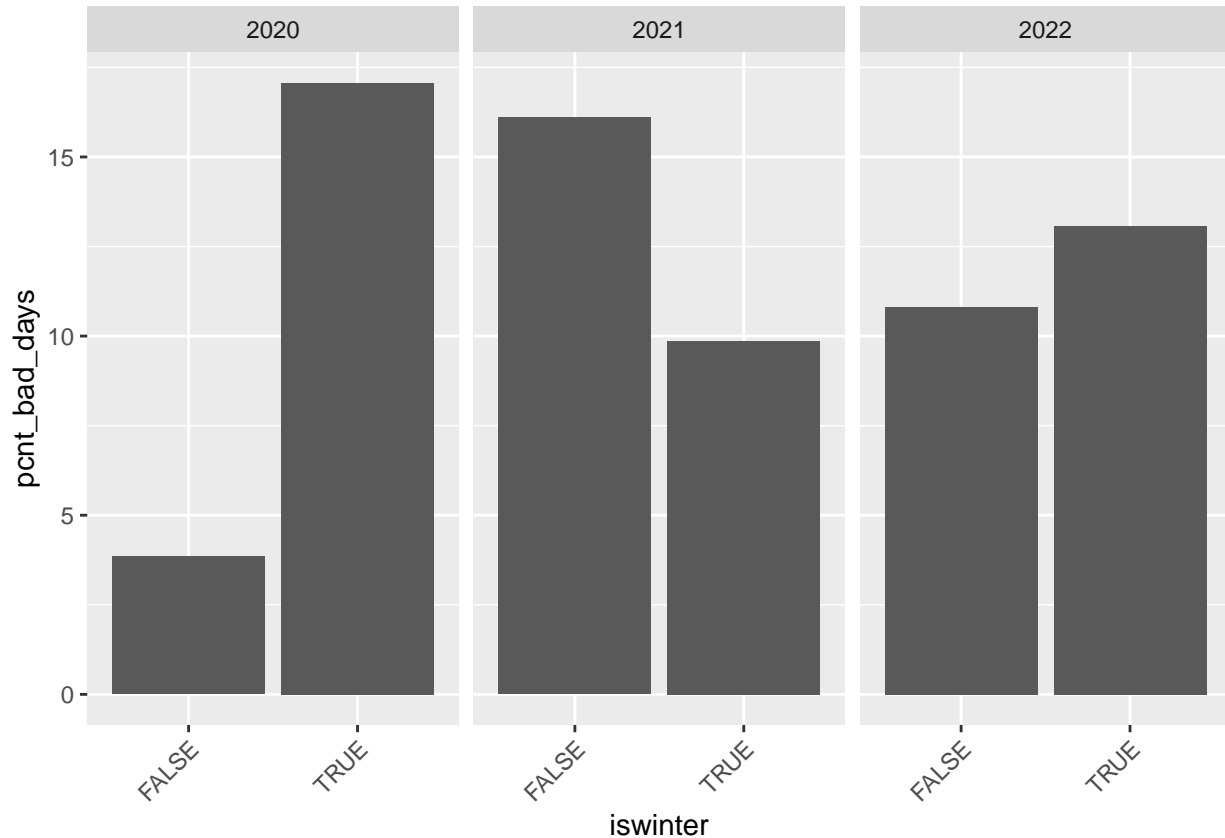
no obvious big discrepancy in winter admissions for this parameter winters of 2020 and 2022 are 'worse' than summers, percentages are small and maybe not significant?

```

weekly_admissions_spec %>%
  filter(speciality=="All") %>%
  filter(admission_type=="All") %>%
  group_by(iswinter, year) %>%
  summarise(count_bad_days = sum(above_thresh),
            count_days = n()*7,
            pcnt_bad_days = 100*(sum(above_thresh)/(n()*7)),
            mean_percentvar = mean(percent_variation)) %>%
  ggplot() +
  aes(x=iswinter, y=pcnt_bad_days) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  facet_grid(~year)

```

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



```

weekly_admissions_spec %>%
  filter(speciality=="All") %>%
  filter(admission_type=="Emergency") %>%
  group_by(iswinter, year) %>%
  summarise(count_bad_days = sum(above_thresh),
            count_days = n()*7,
            pcnt_bad_days = 100*(sum(above_thresh)/(n()*7)),

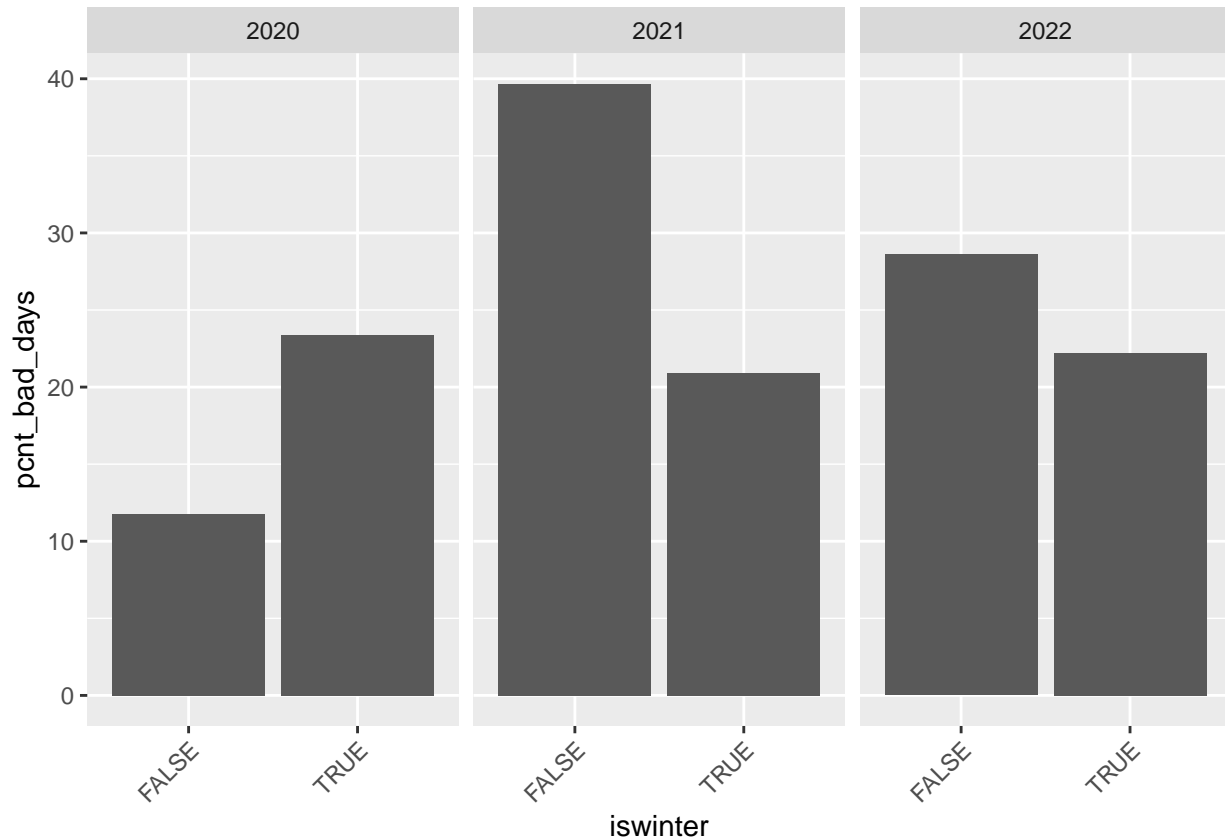
```

```

    mean_percentvar = mean(percent_variation)) %>%
  ggplot() +
    aes(x=iswinter, y=pcnt_bad_days) +
    geom_col() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    facet_grid(~year)

```

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



```

weekly_admissions_spec %>%
  filter(speciality=="Accident & Emergency") %>%
  filter(admission_type=="Emergency") %>%
  group_by(iswinter, year) %>%
  summarise(count_bad_days = sum(above_thresh),
            count_days = n()*7,
            pcnt_bad_days = 100*(sum(above_thresh)/(n()*7)),
            mean_percentvar = mean(percent_variation)) %>%
  ggplot() +
    aes(x=year, y=pcnt_bad_days) +
    geom_col() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    facet_grid(~iswinter)

```

'summarise()' has grouped output by 'iswinter'. You can override using the


```
## '.groups' argument.
```

