Mini Project

S/18/841

2024-01-30

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✔dplyr     1.1.3     ✔readr     2.1.4
## ✔forcats   1.0.0     ✔stringr   1.5.0
## ✔ggplot2   3.4.3     ✔tibble    3.2.1
## ✔lubridate 1.9.3     ✔tidyr     1.3.0
## ✔purrr     1.0.2
## ── Conflicts ────────────────────────────────────────
tidyverse_conflicts() ──
## ✖dplyr::filter() masks stats::filter()
## ✖dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(tinytex)
library(skimr)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(latex2exp)
```

```
London_Local_data_2022 <- read_csv("../data/london_local_data_2022.csv")
```

```
## Rows: 289069 Columns: 10
## ── Column specification
──────────────────────────────────────────────────────
## Delimiter: ","
## chr  (2): site, code
## dbl  (7): nox, no2, no, pm10, o3, pm2_5, so2
## dttm (1): date
##
## 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.
```

```r
head(London_Local_data_2022)
```

```
## # A tibble: 6 × 10
##   site       code  date                 nox   no2    no  pm10    o3 pm2_5
so2
##   <chr>      <chr> <dttm>             <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
<dbl>
## 1 Brent - J… BT6   2022-01-01 00:00:00  13.4  10.1   2.2  29.9    NA    NA
NA
## 2 Brent - J… BT6   2022-01-01 01:00:00  16    11.3   3    17.5    NA    NA
NA
## 3 Brent - J… BT6   2022-01-01 02:00:00  11.1   7     2.6  16      NA    NA
NA
## 4 Brent - J… BT6   2022-01-01 03:00:00   7.8   5.3   1.7  16.5    NA    NA
NA
## 5 Brent - J… BT6   2022-01-01 04:00:00   8.6   5.7   1.9  14.8    NA    NA
NA
## 6 Brent - J… BT6   2022-01-01 05:00:00  10.1   6.9   2.1  11.3    NA    NA
NA
```

```r
london_local_sites <- read_csv("../Data/london_local_sites.csv")
```

```
## Rows: 67 Columns: 5
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr (3): code, site, Parameter_name
## dbl (2): latitude, longitude
##
## 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.
```

```r
head(london_local_sites)
```

```
## # A tibble: 6 × 5
##   code  site                                latitude longitude
Parameter_name
##   <chr> <chr>                                  <dbl>     <dbl> <chr>
## 1 BT8   Brent - ARK Franklin Primary Academy    51.5   -0.218  Nitrogen
dioxide
## 2 BT8   Brent - ARK Franklin Primary Academy    51.5   -0.218  PM10
particulat…
## 3 BT6   Brent - John Keble Primary School       51.5   -0.248  Nitrogen
dioxide
## 4 BT6   Brent - John Keble Primary School       51.5   -0.248  PM10
particulat…
## 5 CT4   City of London - Beech Street           51.5   -0.0961 Nitrogen
dioxide
## 6 CT4   City of London - Beech Street           51.5   -0.0961 PM10
particulat…
```

cleaning data

```
London_Local_data_2022 <- London_Local_data_2022 %>%
  remove_empty(c("cols","rows"))

London_Local_data_2022[is.na(London_Local_data_2022)] <- 0

str(London_Local_data_2022)

## tibble [289,069 × 10] (S3: tbl_df/tbl/data.frame)
##  $ site : chr [1:289069] "Brent - John Keble Primary School" "Brent - John
Keble Primary School" "Brent - John Keble Primary School" "Brent - John Keble
Primary School" ...
##  $ code : chr [1:289069] "BT6" "BT6" "BT6" "BT6" ...
##  $ date : POSIXct[1:289069], format: "2022-01-01 00:00:00" "2022-01-01
01:00:00" ...
##  $ nox  : num [1:289069] 13.4 16 11.1 7.8 8.6 10.1 16.5 12.8 14.6 16.7 ...
##  $ no2  : num [1:289069] 10.1 11.3 7 5.3 5.7 6.9 12.2 9.2 10.2 11.1 ...
##  $ no   : num [1:289069] 2.2 3 2.6 1.7 1.9 2.1 2.8 2.4 2.9 3.7 ...
##  $ pm10 : num [1:289069] 29.9 17.5 16 16.5 14.8 11.3 12.9 11.9 9.6 11.6
...
##  $ o3   : num [1:289069] 0 0 0 0 0 0 0 0 0 0 ...
##  $ pm2_5: num [1:289069] 0 0 0 0 0 0 0 0 0 0 ...
##  $ so2  : num [1:289069] 0 0 0 0 0 0 0 0 0 0 ...

skim(London_Local_data_2022)
```

Data summary

Name	London_Local_data_2022
Number of rows	289069
Number of columns	10

_____

Column type frequency:	
character	2
numeric	7
POSIXct	1

_____

Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
site	0	1	17	40	0	34	0
code	0	1	3	3	0	34	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
nox	0	1	52.63	67.70	-4.5	13.3	32.8	66.2	1111.1	▇▁▁▁▁
no2	0	1	27.78	23.94	-7.2	10.5	23.1	40.0	285.1	▇▂▁▁▁
no	0	1	16.20	31.48	-3.3	0.9	5.4	17.3	631.5	▇▁▁▁▁
pm10	0	1	11.28	14.01	-6.5	0.0	8.4	18.0	300.8	▇▁▁▁▁
o3	0	1	3.32	14.14	-2.0	0.0	0.0	0.0	189.6	▇▁▁▁▁
pm2_5	0	1	0.61	3.46	-3.0	0.0	0.0	0.0	685.0	▇▁▁▁▁
so2	0	1	0.09	0.60	-4.9	0.0	0.0	0.0	11.1	▁▇▁▁▁

Variable type: POSIXct

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2022-01-01	2022-12-31 23:00:00	2022-07-02 11:00:00	8760

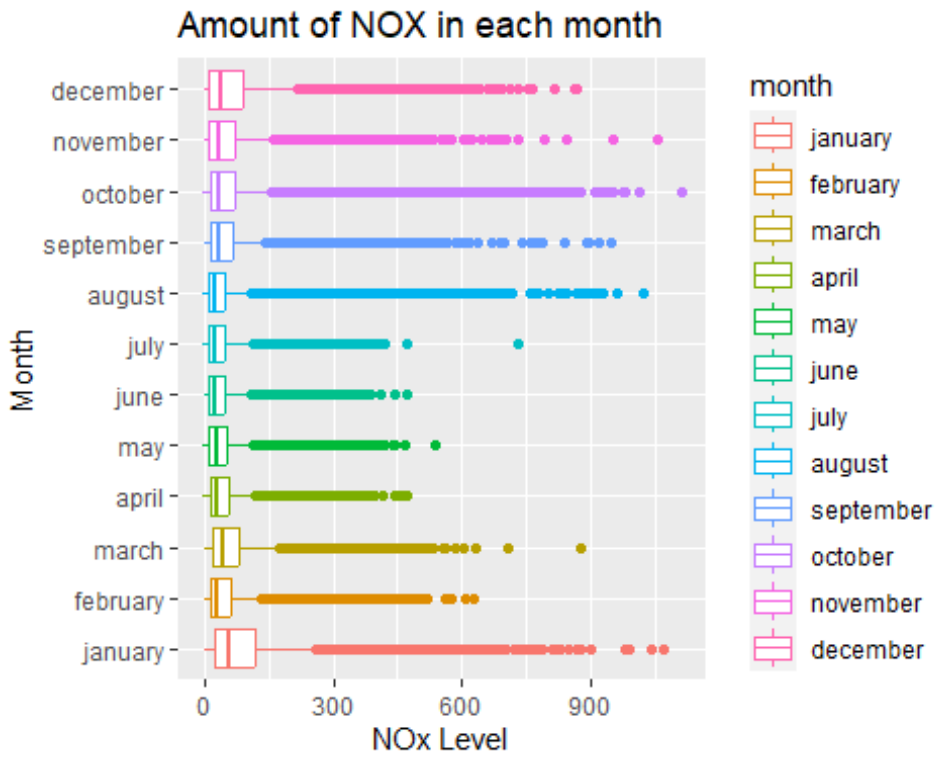
```
London_Local_data_2022 <- London_Local_data_2022 %>%
  mutate(month = factor(case_when(
    month(date)==1 ~ "january",
    month(date)==2 ~ "february",
    month(date)==3 ~ "march",
    month(date)==4 ~ "april",
    month(date)==5 ~ "may",
    month(date)==6 ~ "june",
    month(date)==7 ~ "july",
    month(date)==8 ~ "august",
    month(date)==9 ~ "september",
    month(date)==10 ~ "october",
    month(date)==11 ~ "november",
    month(date)==12 ~ "december",
  )), year = year(date))
```

Plots of data sets with different air types

```
London_Local_data_2022$month <- factor(London_Local_data_2022$month,
                                       levels =
c("january","february","march","april","may",

"june","july","august","september","october","november","december"))
```

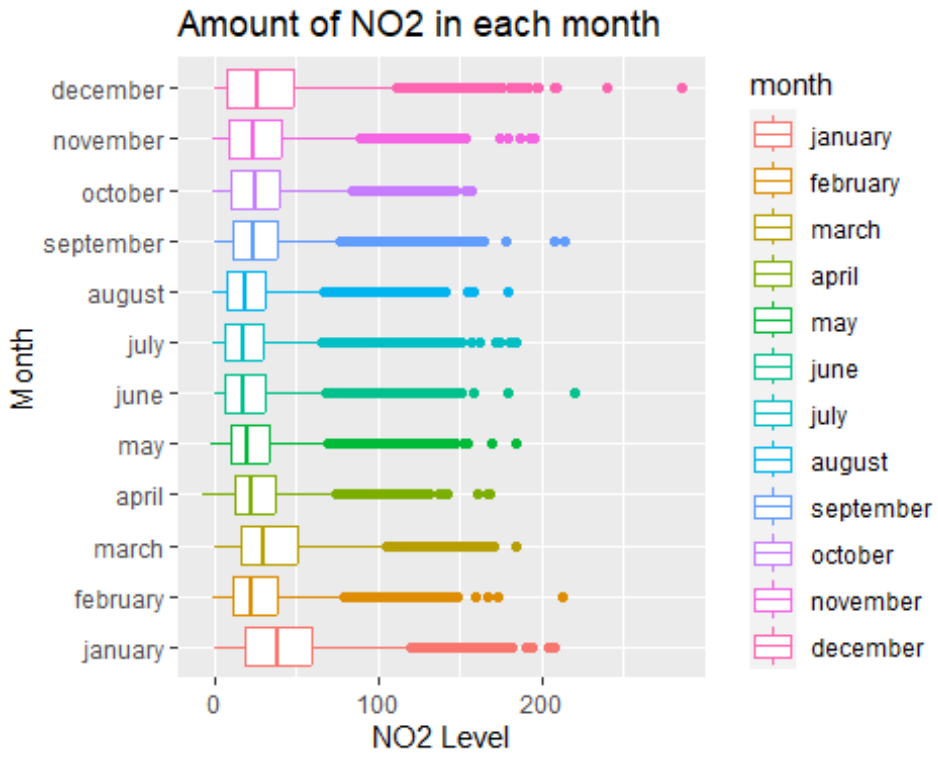
NOX

```
graph_NOX <- London_Local_data_2022 %>%
  ggplot(aes(x=month, y=nox))+geom_boxplot(aes(col=month))+
  coord_flip()+xlab('Month')+ylab('NOx Level')+
  ggtitle("Amount of NOX in each month")

graph_NOX
```



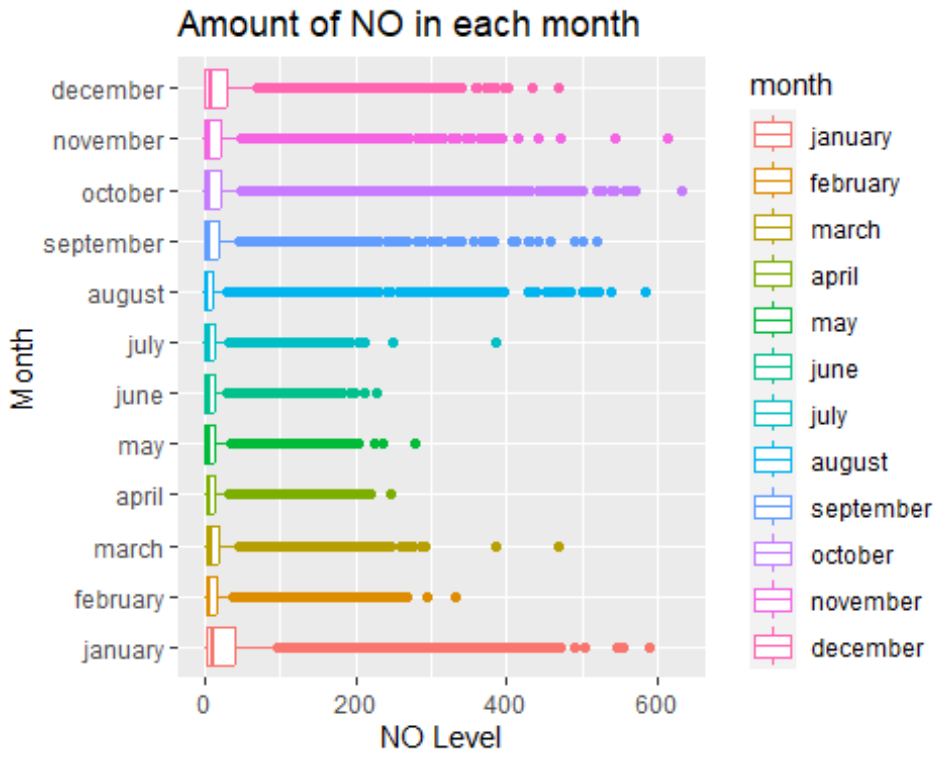
NO2

```
graph_NO2 <- London_Local_data_2022 %>%
  ggplot(aes(x=month, y=no2))+geom_boxplot(aes(col=month))+
  coord_flip()+xlab('Month')+ylab('NO2 Level')+
  ggtitle("Amount of NO2 in each month")

graph_NO2
```



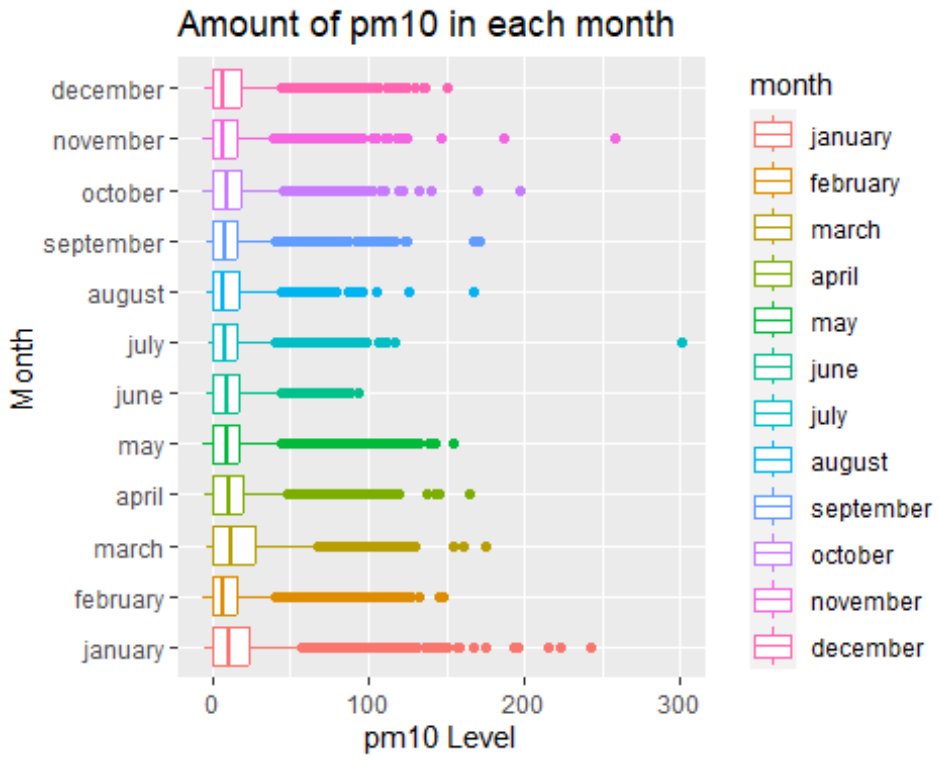
NO

```
graph_NO <- London_Local_data_2022 %>%
  ggplot(aes(x=month, y=no))+geom_boxplot(aes(col=month))+
  coord_flip()+xlab('Month')+ylab('NO Level')+
  ggtitle("Amount of NO in each month")

graph_NO
```



pm10

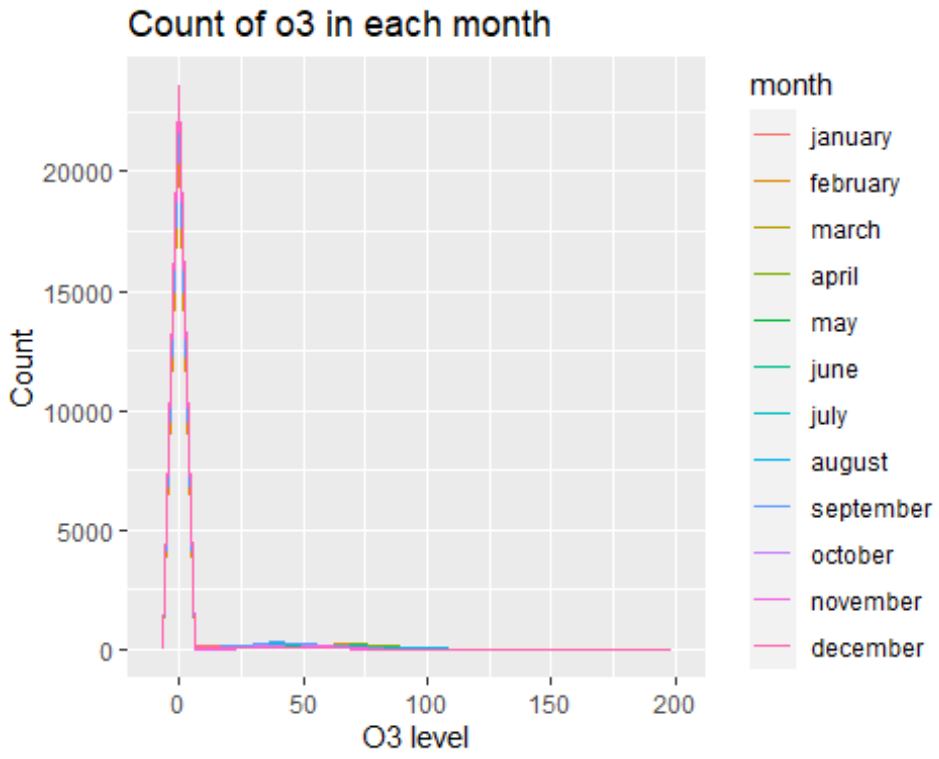
```
graph_pm10 <- London_Local_data_2022 %>%
  ggplot(aes(x=month, y=pm10))+geom_boxplot(aes(col=month))+
  coord_flip()+xlab('Month')+ylab('pm10 Level')+
  ggtitle("Amount of pm10 in each month")

graph_pm10
```



o3 count

```
graph_o3 <- London_Local_data_2022 %>%
  ggplot(aes(x=o3))+geom_freqpoly(aes(col=month))+
  xlab("O3 level")+ylab("Count")+
  ggtitle("Count of o3 in each month")

graph_o3
```

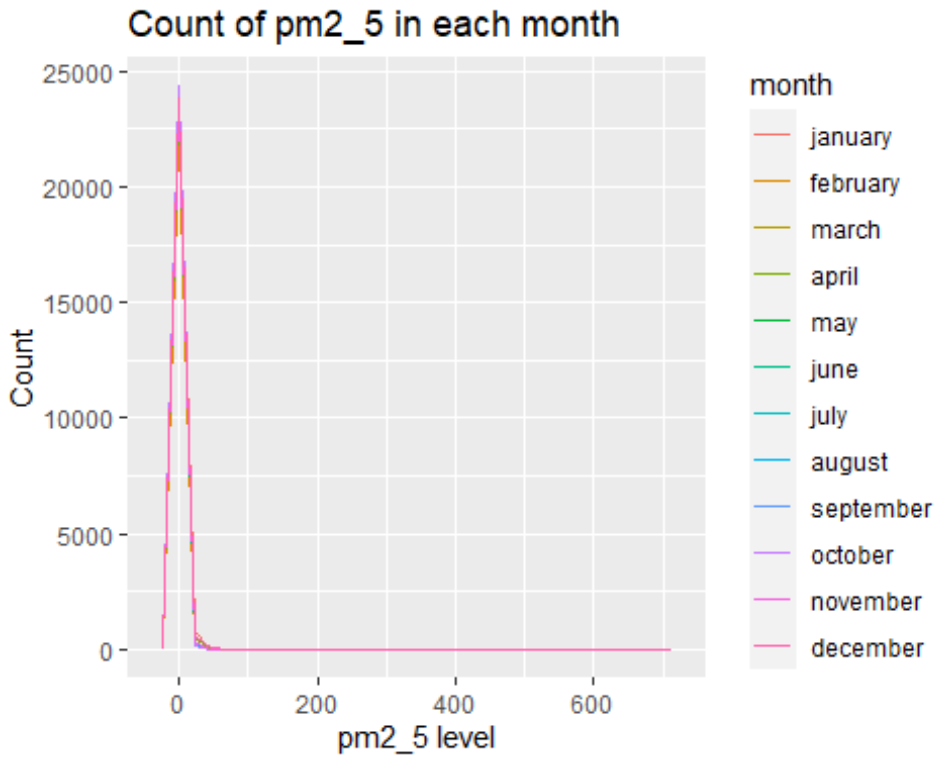
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



pm2_5 count

```
graph_pm2_5 <- London_Local_data_2022 %>%
  ggplot(aes(x=pm2_5))+geom_freqpoly(aes(col=month))+
  xlab("pm2_5 level")+ylab("Count")+
  ggtitle("Count of pm2_5 in each month")

graph_pm2_5
```

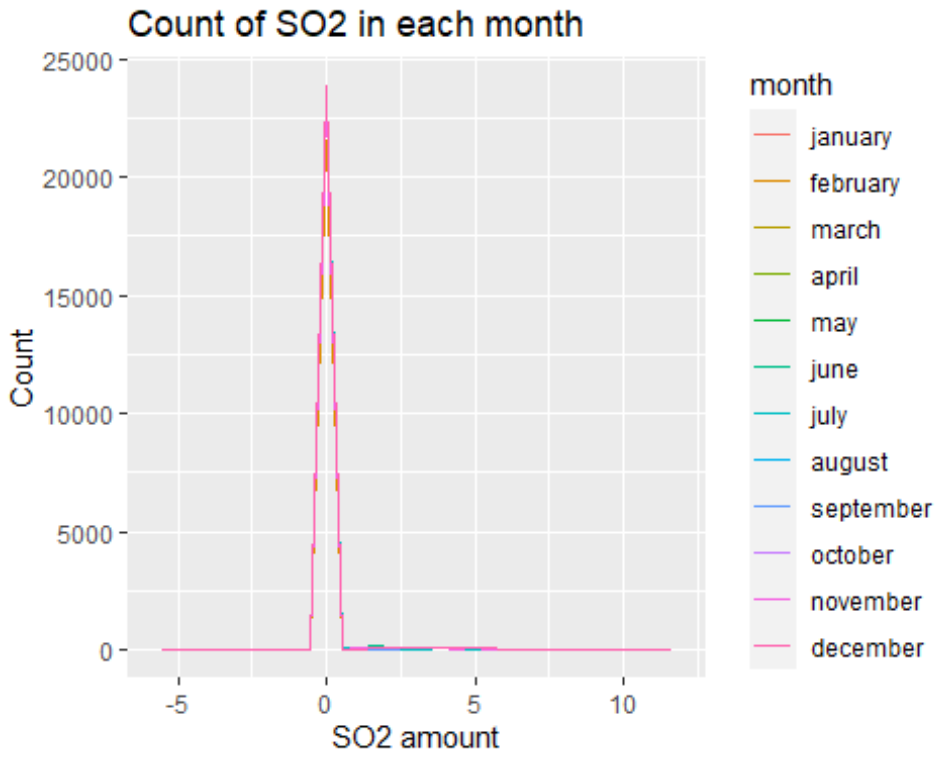
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



SO2 Count

```
graph_SO2 <- London_Local_data_2022 %>%
  ggplot(aes(x=so2))+geom_freqpoly(aes(col=month))+
  xlab("SO2 amount")+ylab("Count")+
  ggtitle("Count of SO2 in each month")

graph_SO2
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Summarization

```
table_site <- London_Local_data_2022 %>%
  group_by(site) %>%
  summarise(NO=sum(no),
            NO2=sum(no2),
            NOx=sum(nox),
            pm10=sum(pm10),
            O3=sum(o3),
            pm2_5=sum(pm2_5),
            SO2=sum(so2))
table_site
```

```
## # A tibble: 34 × 8
##    site                                  NO    NO2    NOx   pm10    O3 pm2_5
SO2
##    <chr>                              <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl>
<dbl>
##  1 Brent - ARK Franklin Primary A… 1.24e5 2.49e5 4.39e5 1.34e5     0     0
0
##  2 Brent - John Keble Primary Sch… 1.01e5 2.40e5 3.94e5 1.42e5     0     0
0
##  3 City of London - Beech Street   1.42e5 3.49e5 5.67e5 1.39e5     0     0
0
##  4 City of London - Sir John Cass… 4.37e4 1.98e5 2.64e5 1.19e5     0     0
0
##  5 City of London - Upper Thames … 0      0      0      2.51e2     0     0
```

```
0
##  6 City of London - Walbrook Wharf 3.44e5 4.42e5 9.69e5 0          0     0
0
##  7 Ealing - Acton Vale             2.82e4 8.58e4 1.29e5 4.73e4     0     0
0
##  8 Ealing - Hanger Lane Gyratory   5.01e5 4.49e5 1.22e6 1.47e5     0     0
0
##  9 Ealing - Western Avenue         1.96e5 3.06e5 6.07e5 2.14e5     0     0
0
## 10 Greenwich - Blackheath          1.06e5 2.25e5 3.87e5 1.42e5     0     0
0
## # ℹ 24 more rows
```

```
table_monthly <- London_Local_data_2022%>%
  group_by(month)%>%
  summarise(NO=sum(no),
            NO2=sum(no2),
            NOx=sum(nox),
            pm10=sum(pm10),
            O3=sum(o3),
            pm2_5=sum(pm2_5),
            SO2=sum(so2))
table_monthly
```

```
## # A tibble: 12 × 8
##    month          NO      NO2      NOx    pm10      O3  pm2_5   SO2
##    <fct>       <dbl>    <dbl>    <dbl>   <dbl>   <dbl>  <dbl> <dbl>
##  1 january   833924. 1014988. 2293633. 359324.  67611. 25071. 2431.
##  2 february  311137.  612346. 1089415. 228271.  84328. 10755  2342.
##  3 march     394539.  869004. 1472031. 413958.  84647. 24516  2412.
##  4 april     256342.  626266. 1020618. 273625  116097. 15558  2061.
##  5 may       279166.  581898. 1011325. 260697.  78288. 13570  1968.
##  6 june      245197.  525196.  902578. 247301.  82322.  7603  1727
##  7 july      252477.  537435.  924698. 231689.  99910. 14055  1732.
##  8 august    273364.  547810.  966785. 244677. 131431. 10208  2305.
##  9 september 377110.  646310. 1224609. 234723.  93936. 13055  1925.
## 10 october   461804.  664810. 1373086. 273280   46951.  8922  2464.
## 11 november  401433.  645585. 1261436. 223239.  32976. 12900  1921.
## 12 december  596718.  757815. 1672765. 268571.  40538. 21490  2790.
```

```
table_monthly_total <- London_Local_data_2022%>%
  group_by(month)%>%
  summarise(Total_pollutant=sum(no)+sum(no2)+sum(nox)+
              sum(pm10)+sum(o3)+sum(pm2_5)+sum(so2))
table_monthly_total
```

```
## # A tibble: 12 × 2
##    month     Total_pollutant
##    <fct>               <dbl>
##  1 january           4596983.
##  2 february          2338595.
```

```
##  3 march            3261107.
##  4 april            2310568.
##  5 may              2226913.
##  6 june             2011924.
##  7 july             2061997.
##  8 august           2176580.
##  9 september        2591667.
## 10 october          2831317.
## 11 november         2579491.
## 12 december         3360687.
```

```
table_site_total <- London_Local_data_2022%>%
  group_by(site)%>%
  summarise(Total_pollutant=sum(no)+sum(no2)+sum(nox)+
              sum(pm10)+sum(o3)+sum(pm2_5)+sum(so2))
table_site_total
```

```
## # A tibble: 34 × 2
##    site                                  Total_pollutant
##    <chr>                                           <dbl>
##  1 Brent - ARK Franklin Primary Academy          946047.
##  2 Brent - John Keble Primary School             876809.
##  3 City of London - Beech Street                1196247.
##  4 City of London - Sir John Cass School         624413.
##  5 City of London - Upper Thames Street             251 
##  6 City of London - Walbrook Wharf              1755354.
##  7 Ealing - Acton Vale                           290427.
##  8 Ealing - Hanger Lane Gyratory                2313841.
##  9 Ealing - Western Avenue                      1322689.
## 10 Greenwich - Blackheath                        859664.
## # ℹ 24 more rows
```

```
max(table_monthly_total$Total_pollutant)
```

```
## [1] 4596983
```

```
min(table_monthly_total$Total_pollutant)
```

```
## [1] 2011924
```

```
max(table_site_total$Total_pollutant)
```

```
## [1] 2355163
```

```
min(table_site_total$Total_pollutant)
```

```
## [1] 251
```

eps files

```r
postscript(file ="graph_NOX.eps",width =5 ,height = 6,horizontal = FALSE )
graph_NOX
dev.off()

## png
##   2

postscript(file ="graph_NO2.eps",width =5 ,height = 6,horizontal = FALSE )
graph_NO2
dev.off()

## png
##   2

postscript(file ="graph_NO.eps",width =5 ,height = 6,horizontal = FALSE )
graph_NO
dev.off()

## png
##   2

postscript(file ="graph_o3.eps",width =5 ,height = 6,horizontal = FALSE )
graph_o3

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

dev.off()

## png
##   2

postscript(file ="graph_pm10.eps",width =5 ,height = 6,horizontal = FALSE )
graph_pm10
dev.off()

## png
##   2

postscript(file ="graph_pm2_5.eps",width =5 ,height = 6,horizontal = FALSE )
graph_pm2_5

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

dev.off()

## png
##   2

postscript(file ="graph_SO2.eps",width =5 ,height = 6,horizontal = FALSE )
graph_SO2

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

dev.off()
```

```
## png
##   2
```