

# ST301\_MP\_S18\_827

S18827

2024-01-22

“How does air quality vary across different monitoring sites in London throughout the year 2022-2023, and what are the major trends and patterns observed in the measured air pollutants?”

```
library(tidyverse)
```

```
## Warning: package 'lubridate' was built under R version 4.0.5
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.3      v readr      2.1.4
```

```
## v forcats   1.0.0      v stringr   1.5.0
```

```
## v ggplot2   3.4.4      v tibble    3.2.1
```

```
## v lubridate 1.8.0      v tidyr     1.3.0
```

```
## 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
```

```
library(skimr)
```

```
library(lubridate)
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.0.5
```

```
##
```

```
## Attaching package: 'reshape2'
```

```
##
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##      smiths
```

```
library(cowplot)
```

```
##
```

```
## Attaching package: 'cowplot'
```

```
##
```

```
## The following object is masked from 'package:lubridate':
```

```
##
```

```
##      stamp
```

```
library(sf)
```

```
## Warning: package 'sf' was built under R version 4.0.5
```

```
## Linking to GEOS 3.9.1, GDAL 3.2.1, PROJ 7.2.1; sf_use_s2() is TRUE
```

```
library(ggpubr)
```

```
##
## Attaching package: 'ggpubr'
##
## The following object is masked from 'package:cowplot':
##
##     get_legend
```

```
london_local_sites <- read_csv(file = "../Data/london_local_sites.csv", col_types = 'ccddc')
london_local_data_2022 <- read_csv(file = "../Data/london_local_data_2022.csv", col_types = 'ccddddd')
london_local_data_2022 <- london_local_data_2022 %>% mutate(date = ymd_hms(date))
```

```
london_local_data_2022 <- london_local_data_2022 %>%
  mutate(Month = month(date, label = TRUE),
         Day = weekdays(date))
```

```
london_local_sites %>%
  group_by(Parameter_name) %>%
  summarise(no_of_sites_measured = n_distinct(site)) %>%
  arrange(desc(no_of_sites_measured))
```

```
## # A tibble: 5 x 2
##   Parameter_name          no_of_sites_measured
##   <chr>                  <int>
## 1 Nitrogen dioxide             35
## 2 PM10 particulate matter (Hourly measured) 23
## 3 PM2.5 particulate matter (Hourly measured)  5
## 4 Ozone                       3
## 5 Sulphur dioxide             1
```

```
london_local_sites <- london_local_sites %>%
  group_by(code, site, latitude, longitude) %>%
  summarise(Parameter_name = str_c(Parameter_name, collapse = ", "))
```

```
## 'summarise()' has grouped output by 'code', 'site', 'latitude'. You can
## override using the '.groups' argument.
```

```
london_air_quality_2022 <- left_join(x = london_local_data_2022,
                                     y = london_local_sites,
                                     by = c("code", "site"))
```

```
summary(london_air_quality_2022[c("no", "no2", "nox", "pm10", "pm2_5", "o3", "so2")])
```

```
##           no           no2           nox           pm10
## Min.      : -3.30    Min.      : -7.20    Min.      : -4.50    Min.      : -6.50
## 1st Qu.:   2.60    1st Qu.:  16.70    1st Qu.:  22.50    1st Qu.:  11.10
## Median :   7.99    Median :  27.70    Median :  40.80    Median :  16.40
## Mean      : 19.34    Mean      : 33.15    Mean      : 62.81    Mean      : 19.47
## 3rd Qu.:  21.10    3rd Qu.:  44.00    3rd Qu.:  75.70    3rd Qu.:  23.90
## Max.      :631.50    Max.      :285.10    Max.      :1111.10    Max.      :300.80
## NA's      :46860    NA's      :46861    NA's      :46861    NA's      :121636
##      pm2_5           o3           so2
## Min.      : -3.0    Min.      : -2.00    Min.      : -4.90
## 1st Qu.:   6.0    1st Qu.:  32.20    1st Qu.:   1.80
## Median :   8.0    Median :  49.30    Median :   3.00
## Mean      : 10.6    Mean      : 49.61    Mean      :   3.08
## 3rd Qu.:  13.0    3rd Qu.:  65.70    3rd Qu.:   4.20
## Max.      :685.0    Max.      :189.60    Max.      :11.10
## NA's      :272308    NA's      :269738    NA's      :280607
```

```
london_shapefile <- st_read("../Data/ESRI/London_Borough_Excluding_MHW.shp")
```

```
## Reading layer 'London_Borough_Excluding_MHW' from data source
## 'E:\Projects\ST-306-Mini-Project\R Project\Data\ESRI\London_Borough_Excluding_MHW.shp'
## using driver 'ESRI Shapefile'
## Simple feature collection with 33 features and 7 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: 503568.2 ymin: 155850.8 xmax: 561957.5 ymax: 200933.9
## Projected CRS: OSGB 1936 / British National Grid
```

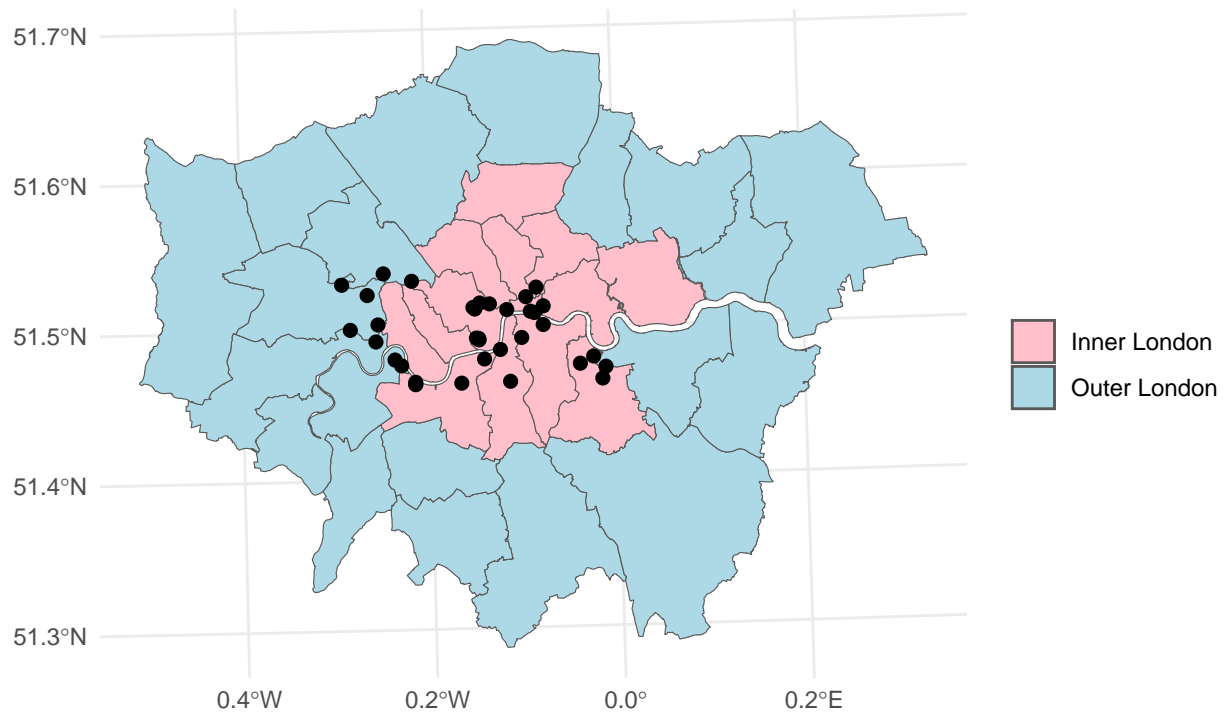
```
c <- st_read("../Data/ESRI/London_Borough_Excluding_MHW.dbf")
```

```
## Reading layer 'London_Borough_Excluding_MHW' from data source
## 'E:\Projects\ST-306-Mini-Project\R Project\Data\ESRI\London_Borough_Excluding_MHW.dbf'
## using driver 'ESRI Shapefile'
## Simple feature collection with 33 features and 7 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: 503568.2 ymin: 155850.8 xmax: 561957.5 ymax: 200933.9
## Projected CRS: OSGB 1936 / British National Grid
```

```
points_sp <- st_as_sf(london_local_sites, coords = c("longitude", "latitude"), crs = 4326)
```

```
ggplot() +
  geom_sf(data = london_shapefile, aes(fill = ONS_INNER)) +
  geom_sf(data = points_sp, size = 2) +
  scale_fill_manual(name = "",
                    values = c("T" = "pink", "F" = "lightblue"),
                    breaks=c("T", "F"),
                    labels=c("Inner London", "Outer London")) +
  theme_minimal() +
  labs(title = "Monitoring Sites in London")
```

## Monitoring Sites in London



```
ggsave("../Latex Report/london_map.eps")
```

```
## Saving 6.5 x 4.5 in image
```

```
no_data <- london_air_quality_2022 %>%
  filter(grepl("Nitrogen dioxide", Parameter_name)) %>%
  select(-c("pm10", "o3", "pm2_5", "so2")) %>%
  drop_na()

so2_data <- london_air_quality_2022 %>%
  filter(grepl("Sulphur dioxide", Parameter_name)) %>%
  select(-c("nox", "no2", "no", "pm10", "o3", "pm2_5")) %>%
  drop_na()

oz_data <- london_air_quality_2022 %>%
  filter(grepl("Ozone", Parameter_name)) %>%
  select(-c("nox", "no2", "no", "pm10", "pm2_5", "so2")) %>%
  drop_na()

pm2_5_data <- london_air_quality_2022 %>%
  filter(grepl("PM2.5", Parameter_name)) %>%
  select(-c("nox", "no2", "no", "o3", "pm10", "so2")) %>%
  drop_na()
```

```
pm10_data <- london_air_quality_2022 %>%
  filter(grepl("PM10", Parameter_name)) %>%
  select(-c("nox", "no2", "no", "o3", "pm2_5", "so2")) %>%
  drop_na()
```

```
summary(no_data[c("no", "no2", "nox")])
```

```
##           no           no2           nox
## Min.      : -3.300   Min.    : -7.20   Min.     : -4.50
## 1st Qu.:   2.600   1st Qu.:  16.70   1st Qu.:  22.50
## Median :   7.992   Median :  27.70   Median :  40.80
## Mean     :  19.335   Mean    :  33.15   Mean     :  62.81
## 3rd Qu.:  21.100   3rd Qu.:  44.00   3rd Qu.:  75.70
## Max.     : 631.500   Max.    : 285.10   Max.     :1111.10
```

```
summary(so2_data["so2"])
```

```
##           so2
## Min.      : -4.900
## 1st Qu.:   1.800
## Median :   3.000
## Mean     :   3.082
## 3rd Qu.:   4.200
## Max.     :  11.100
```

```
summary(oz_data["o3"])
```

```
##           o3
## Min.      : -2.00
## 1st Qu.:  32.20
## Median :  49.30
## Mean     :  49.61
## 3rd Qu.:  65.70
## Max.     : 189.60
```

```
summary(pm2_5_data["pm2_5"])
```

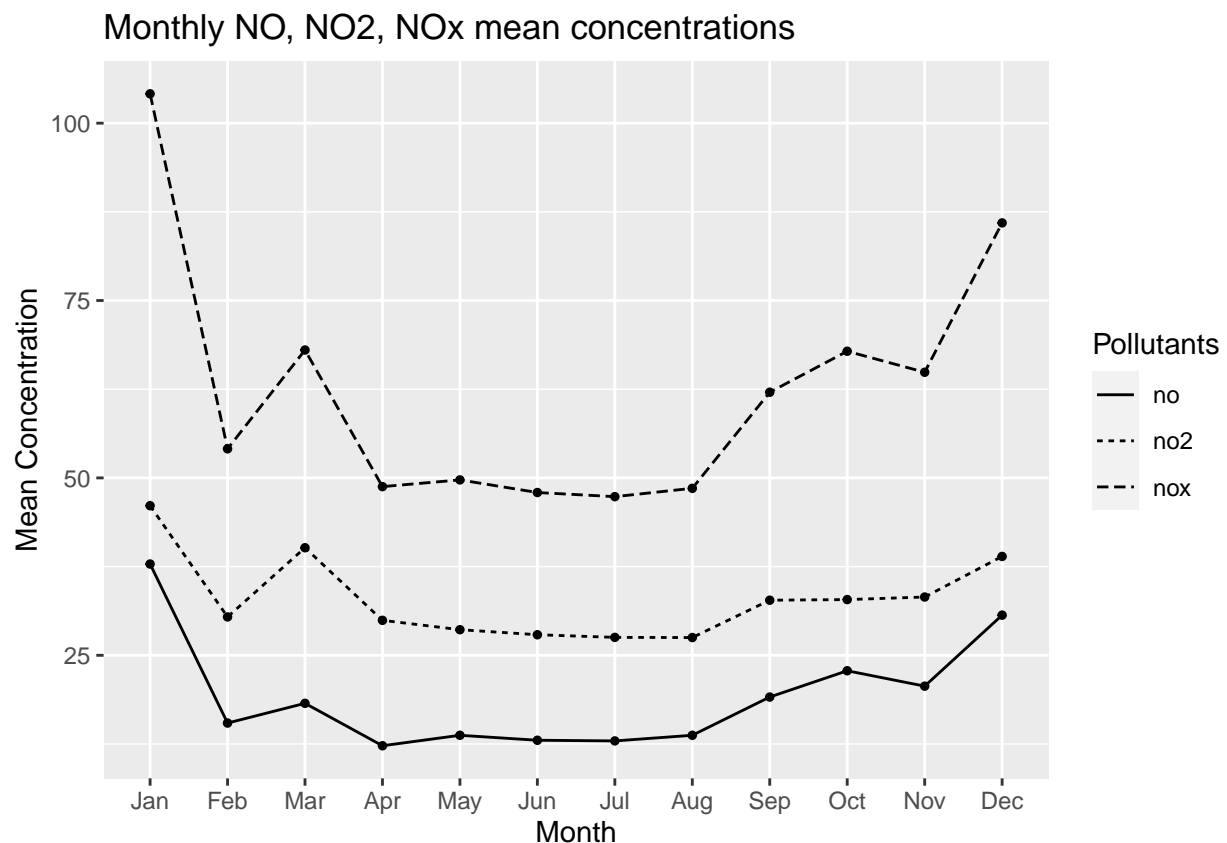
```
##           pm2_5
## Min.      : -3.0
## 1st Qu.:   6.0
## Median :   8.0
## Mean     :  10.6
## 3rd Qu.:  13.0
## Max.     : 685.0
```

```
summary(pm10_data["pm10"])
```

```
##           pm10
## Min.      : -6.50
```

```
## 1st Qu.: 11.10
## Median : 16.40
## Mean   : 19.47
## 3rd Qu.: 23.90
## Max.   :300.80
```

```
no_month <- no_data %>%
  group_by(Month) %>%
  summarise(no = mean(no),
            no2 = mean(no2),
            nox = mean(nox)) %>%
  melt(id.vars = "Month") %>%
  ggplot(aes(x = Month, y = value)) +
  geom_point(size = 1) +
  geom_line(aes(group = variable, linetype = variable)) +
  labs(title = "Monthly NO, NO2, NOx mean concentrations",
       x = "Month",
       y = "Mean Concentration",
       linetype = "Pollutants")
no_month
```



```
#ggsave(file = "../.. /Latex Report/NO_Months.eps")
```

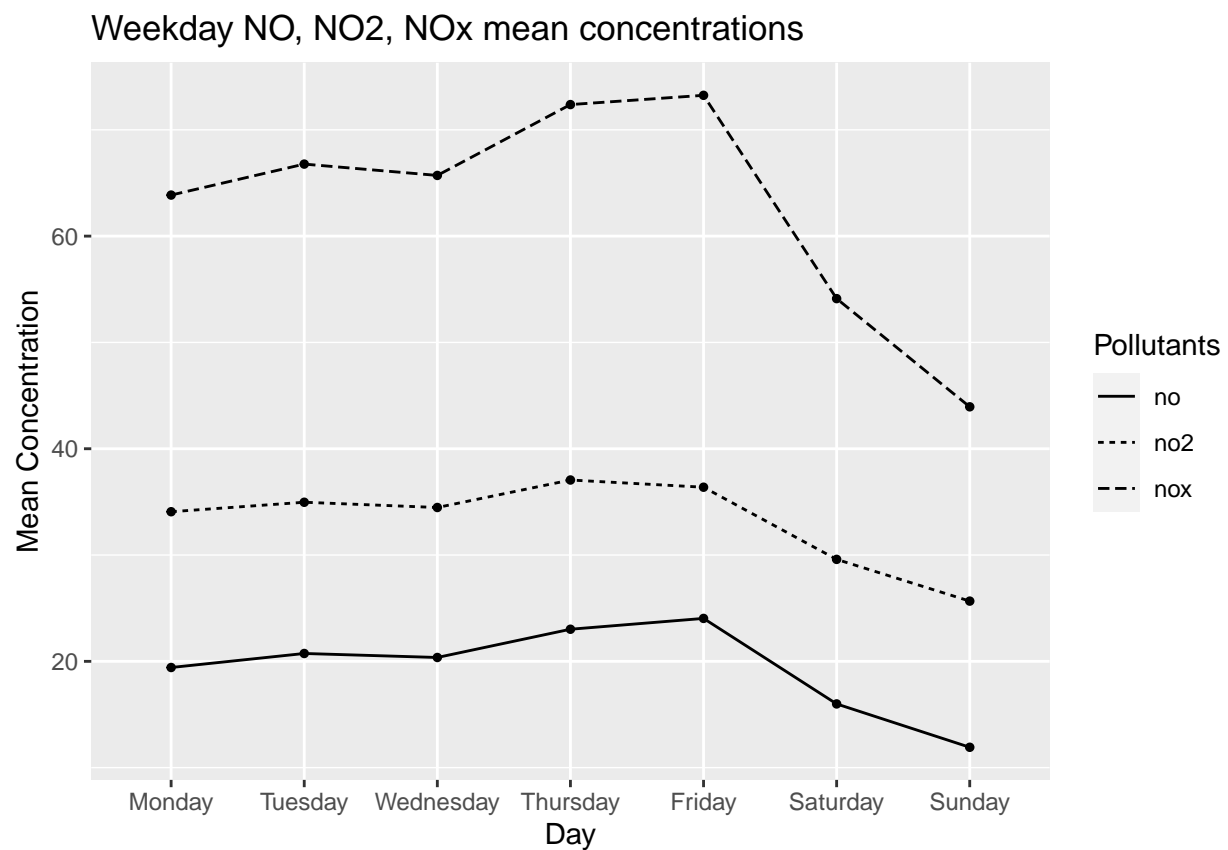
```
no_day_data <- no_data %>%
  group_by(Day) %>%
```

```

summarise(no = mean(no),
          no2 = mean(no2),
          nox = mean(nox)) %>%
melt(id.vars = "Day")

no_day <- no_day_data %>%
  ggplot(aes(x=factor(Day, level=c("Monday", "Tuesday", "Wednesday", "Thursday","Friday","Saturday", "Sunday"),
    y = value)) +
  geom_point(size = 1) +
  geom_line(aes(group = variable, linetype = variable)) +
  labs(title = "Weekday NO, NO2, NOx mean concentrations",
       x = "Day",
       y = "Mean Concentration",
       linetype = "Pollutants")
no_day

```

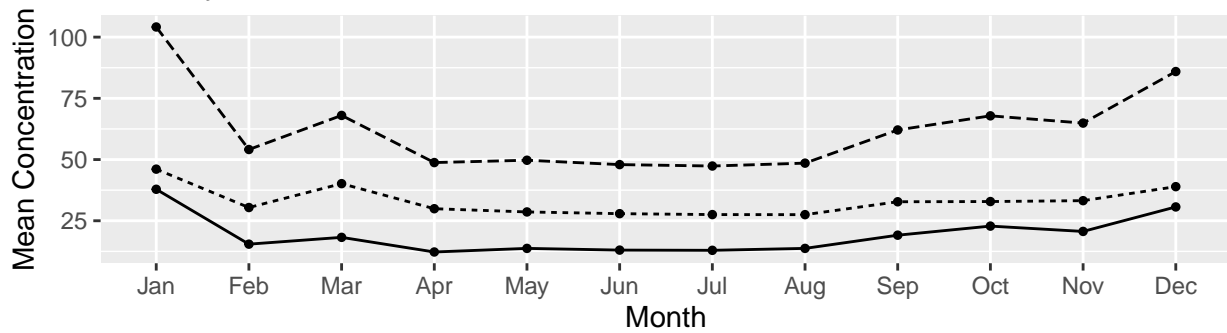


```

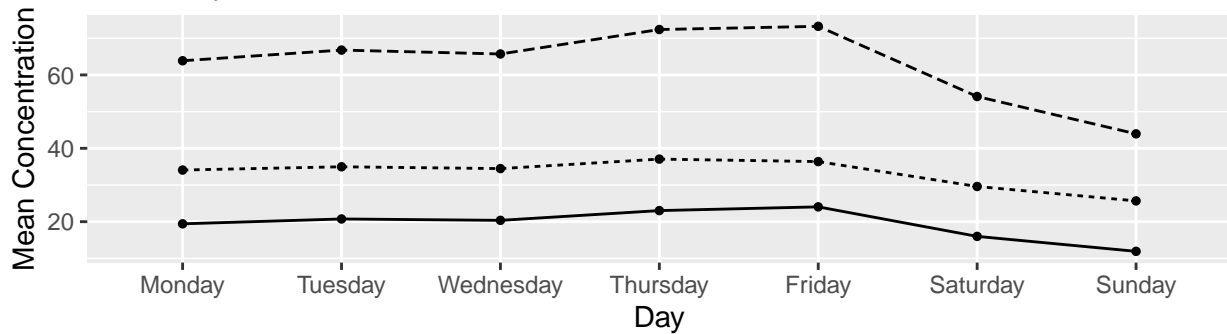
ggarrange(no_month, no_day, ncol=1, common.legend = TRUE, legend="bottom")

```

Monthly NO, NO2, NOx mean concentrations



Weekday NO, NO2, NOx mean concentrations



Pollutants — no ..... no2 - - - - - nox

```
suppressMessages({
  ggsave("../Latex Report/no_plot.eps")
})
```

```
#ggsave(file = "../Latex Report/NO_Day.eps")
```

```
no_day_data %>%
  filter(Day %in% c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")) %>%
  group_by(variable) %>%
  summarise(mean_value = mean(value))
```

```
## # A tibble: 3 x 2
##   variable mean_value
##   <fct>         <dbl>
## 1 no           21.5
## 2 no2          35.4
## 3 nox          68.4
```

```
no_day_data %>%
  filter(Day %in% c("Saturday", "Sunday")) %>%
  group_by(variable) %>%
  summarise(mean_value = mean(value))
```

```
## # A tibble: 3 x 2
```



```
##   variable mean_value
##   <fct>         <dbl>
## 1 no            14.0
## 2 no2           27.6
## 3 nox           49.0
```

```
no_site <- no_data %>%
  group_by(site) %>%
  summarise(no = mean(no),
            no2 = mean(no2),
            nox = mean(nox))
```

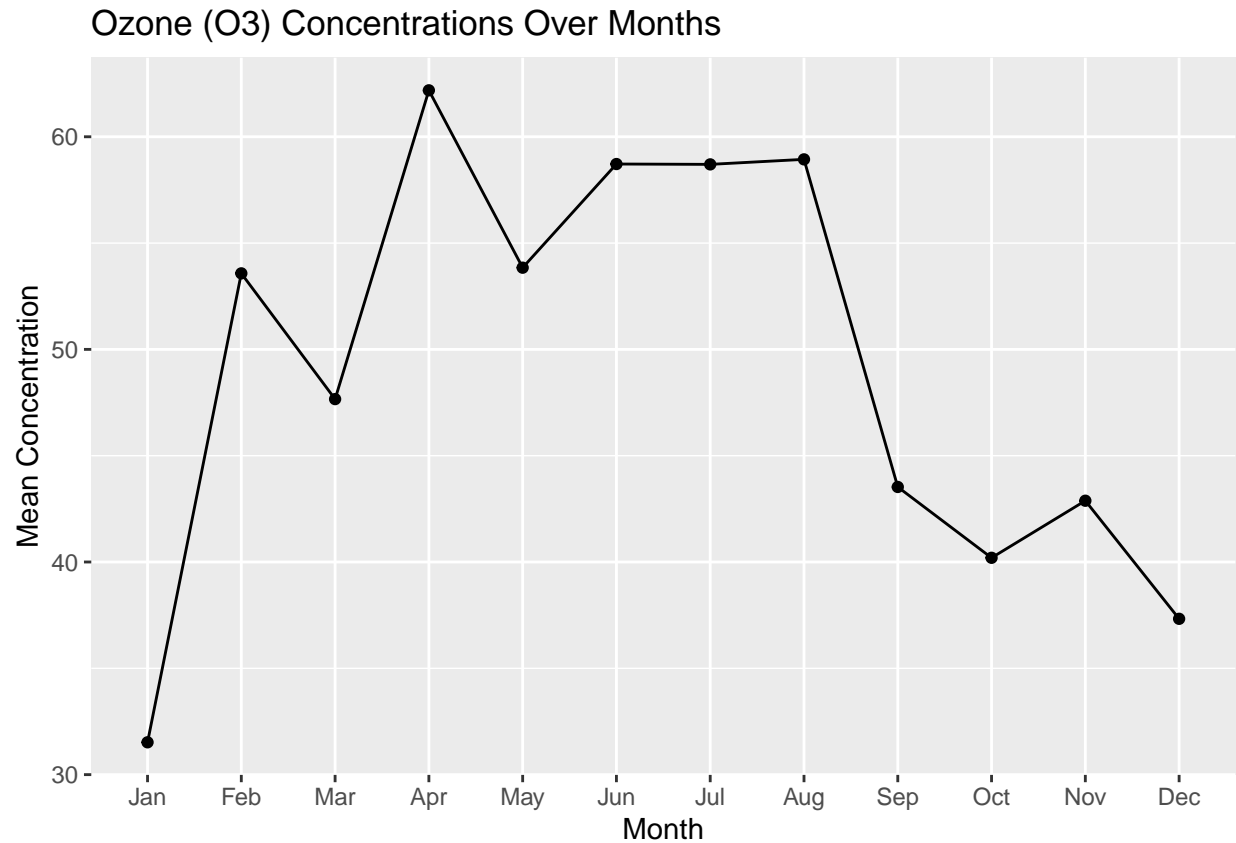
```
no_site %>%
  arrange(nox)
```

```
## # A tibble: 33 x 4
##   site                                no   no2   nox
##   <chr>                             <dbl> <dbl> <dbl>
## 1 Richmond Upon Thames - Barnes Wetlands 2.06 14.3 17.4
## 2 Lewisham - Deptford                    4.98 19.1 26.7
## 3 Southwark - Elephant and Castle        4.89 21.8 29.3
## 4 City of London - Sir John Cass School  5.05 22.8 30.6
## 5 Ealing - Acton Vale                    7.13 21.7 32.6
## 6 Westminster - Ebury Street (Grosvenor) 7.00 24.8 35.7
## 7 Richmond Upon Thames - Castelnau      8.44 23.3 36.3
## 8 Wandsworth - Putney                   10.2 26.9 42.6
## 9 Wandsworth - Battersea                10.8 27.1 43.6
## 10 Westminster - Duke Street (Grosvenor) 8.73 31.6 45.0
## # i 23 more rows
```

```
oz_data_month <- oz_month <- oz_data %>%
  group_by(Month) %>%
  summarise(o3 = mean(o3))
```

```
oz_month <- oz_data_month %>%
  ggplot(aes(x = Month, y = o3)) +
  geom_point() +
  geom_line(aes(group = 1)) +
  labs(title = "Ozone (O3) Concentrations Over Months",
       x = "Month",
       y = "Mean Concentration")
```

```
oz_month
```

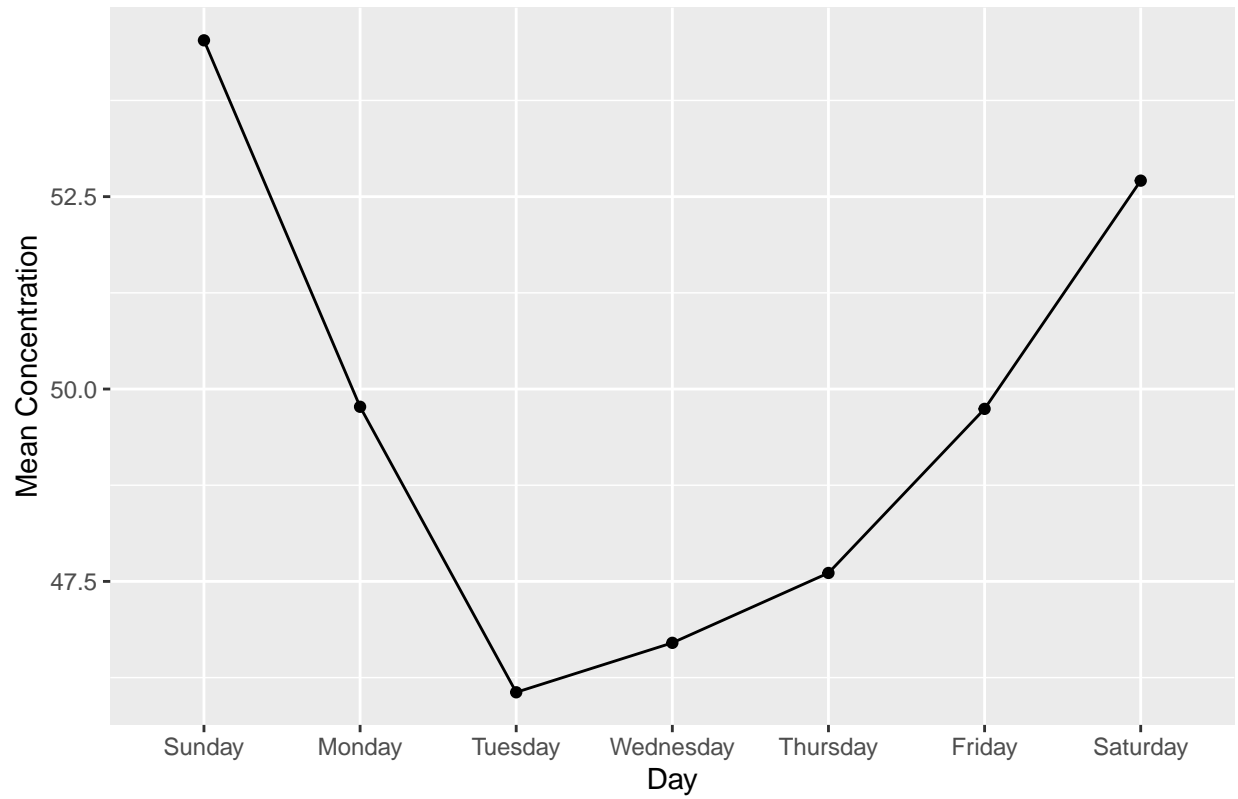


```
#ggsave(file = "../..\\Latex Report/Ozone_Months.eps")
```

```
oz_day <- oz_data %>%
  group_by(Day) %>%
  summarise(o3 = mean(o3)) %>%
  ggplot(aes(x=factor(Day, level=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday")),
    geom_point() +
    geom_line(aes(group = 1)) +
    labs(title = "Ozone (O3) Concentrations Over Weekdays",
      x = "Day",
      y = "Mean Concentration")
```

```
oz_day
```

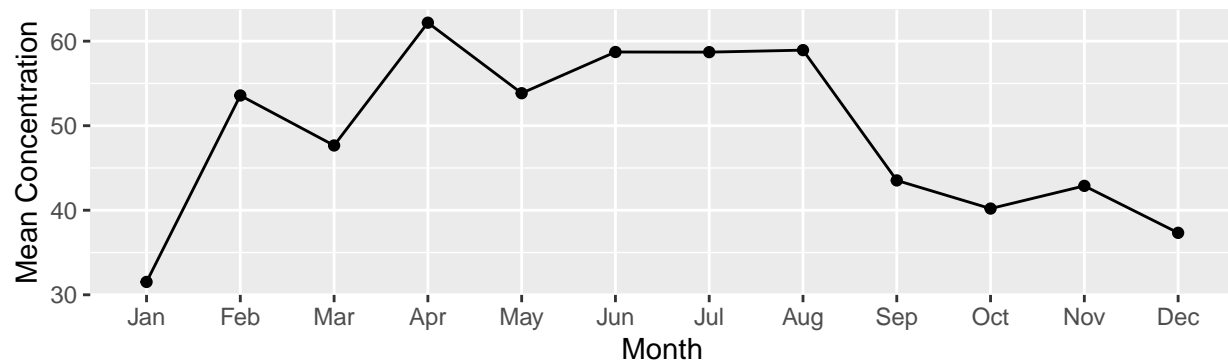
Ozone (O<sub>3</sub>) Concentrations Over Weekdays



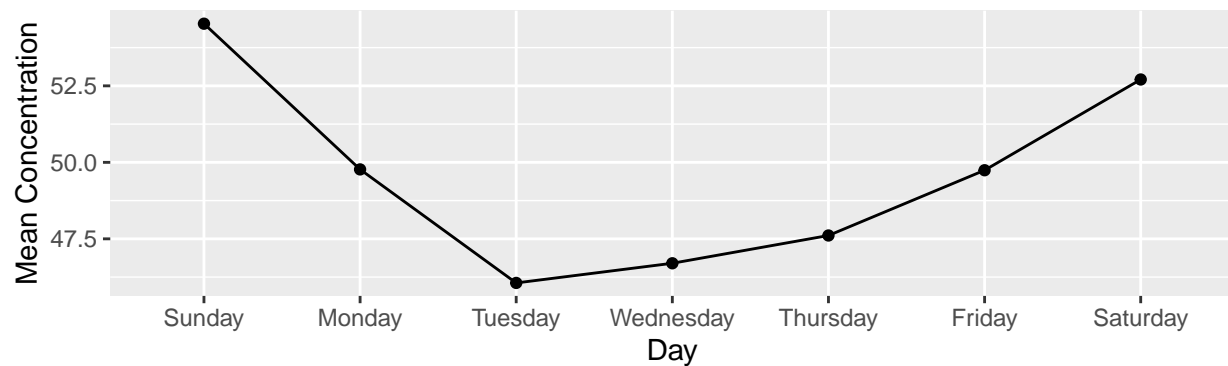
```
#ggsave(file = "../Latex Report/Ozone_Days.eps")
```

```
ggarrange(oz_month, oz_day, ncol=1)
```

### Ozone (O3) Concentrations Over Months



### Ozone (O3) Concentrations Over Weekdays

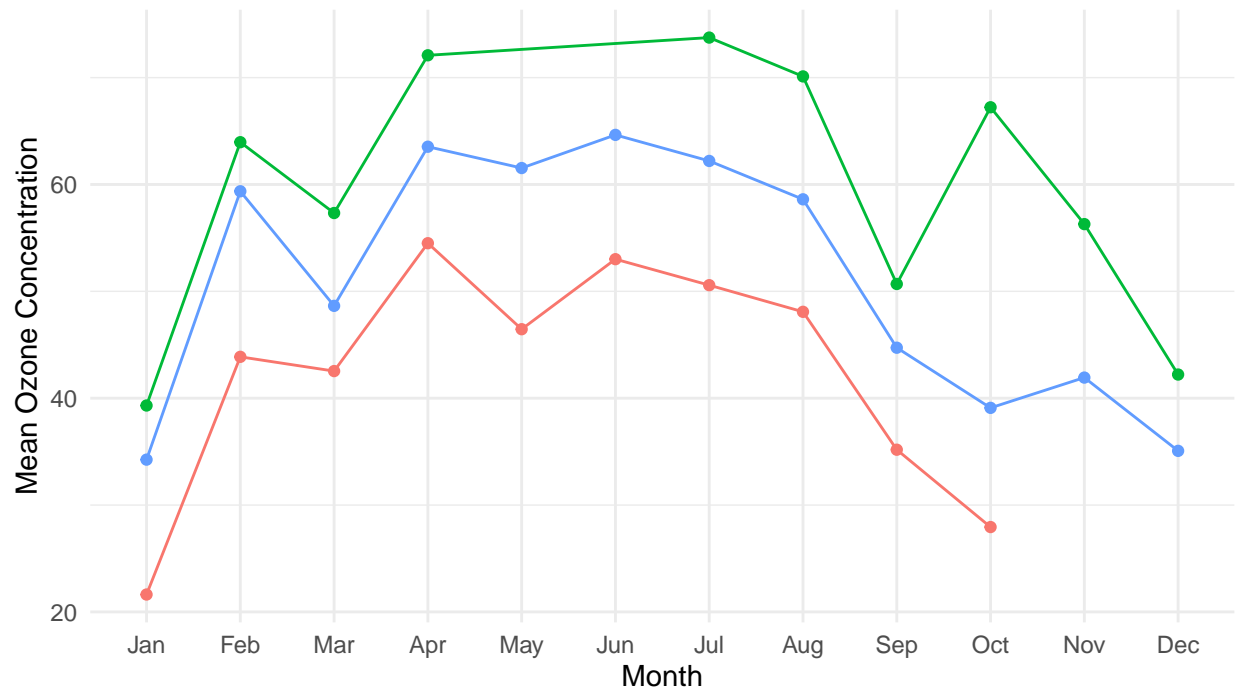


```
suppressMessages({
  ggsave("../Latex Report/oz_plot.eps")
})

oz_data %>%
  group_by(Month, code, site) %>%
  summarise(o3 = mean(o3)) %>%
  ggplot(aes(x = Month, y = o3, color = site)) +
  geom_point() +
  geom_line(aes(group = site)) +
  labs(
    title = "Mean Ozone Concentrations Over Months for Each Monitoring Site",
    x = "Month",
    y = "Mean Ozone Concentration",
    color = "Monitoring Site") +
  theme_minimal() +
  theme(legend.position = "bottom")
```

## 'summarise()' has grouped output by 'Month', 'code'. You can override using the  
## '.groups' argument.

Mean Ozone Concentrations Over Months for Each Monitoring Site

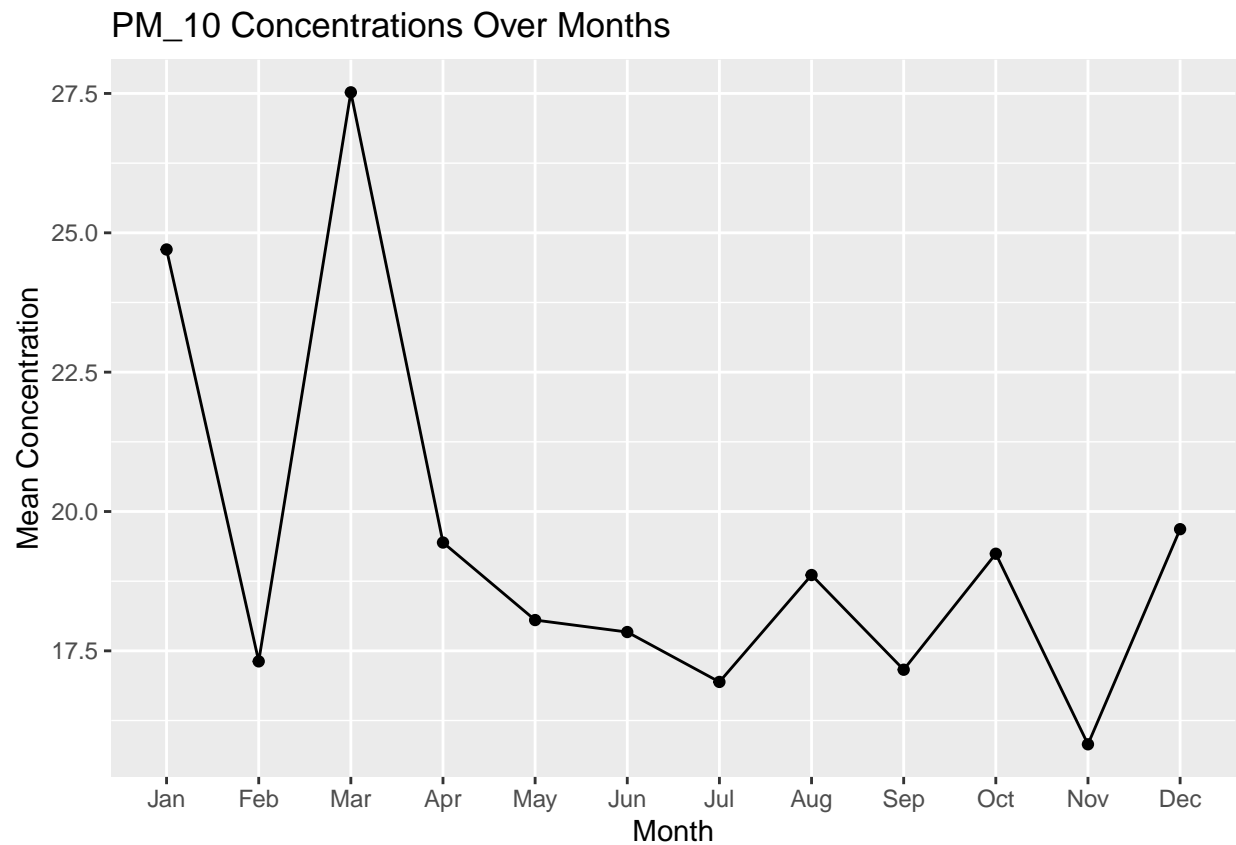


Monitoring Site — Hackney – Old Street — Richmond Upon Thames – Barnes Wetlands — Southwark – Elephant

```
ggsave("../Latex Report/o3_month_sites.eps")
```

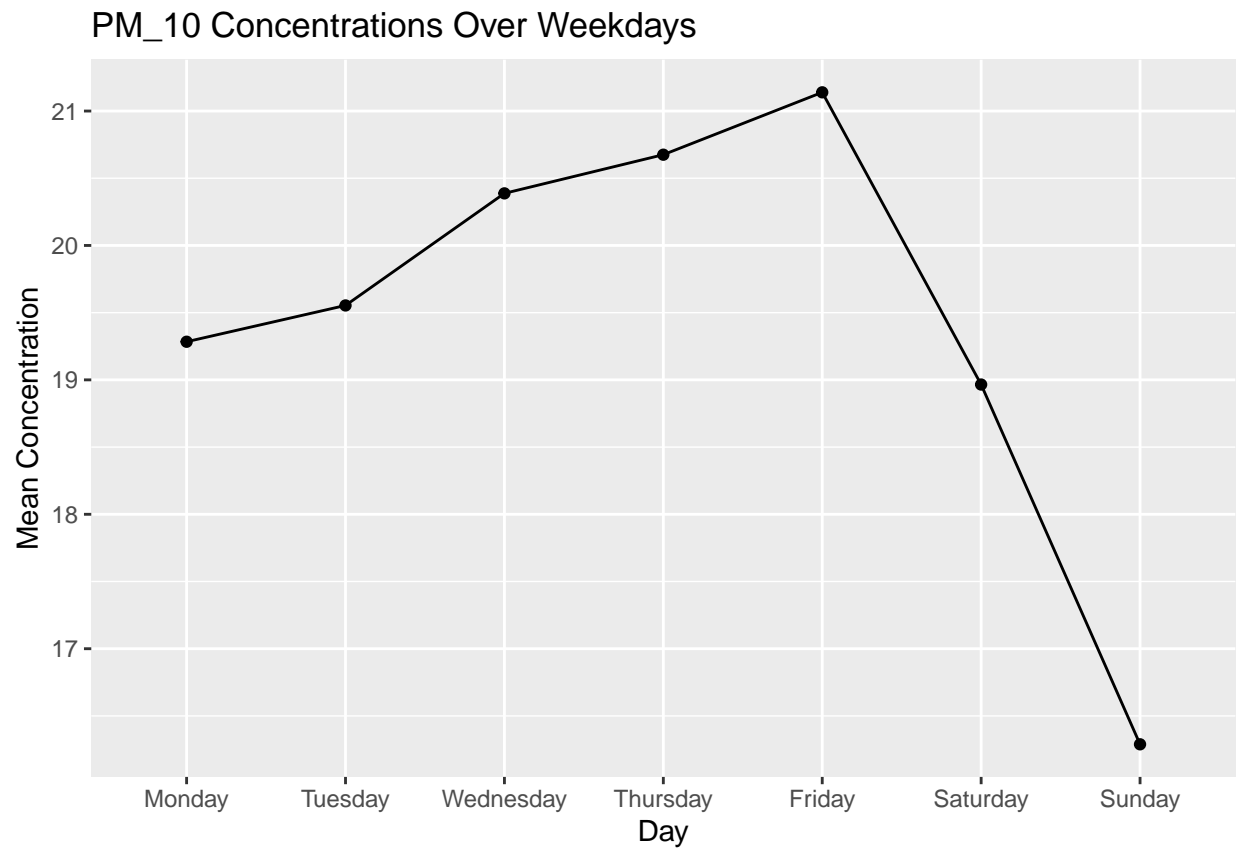
```
## Saving 6.5 x 4.5 in image
```

```
pm10_month <- pm10_data %>%
  group_by(Month) %>%
  summarise(pm10 = mean(pm10)) %>%
  ggplot(aes(x = Month, y = pm10)) +
  geom_point() +
  geom_line(aes(group = 1)) +
  labs(title = "PM10 Concentrations Over Months",
       x = "Month",
       y = "Mean Concentration")
pm10_month
```

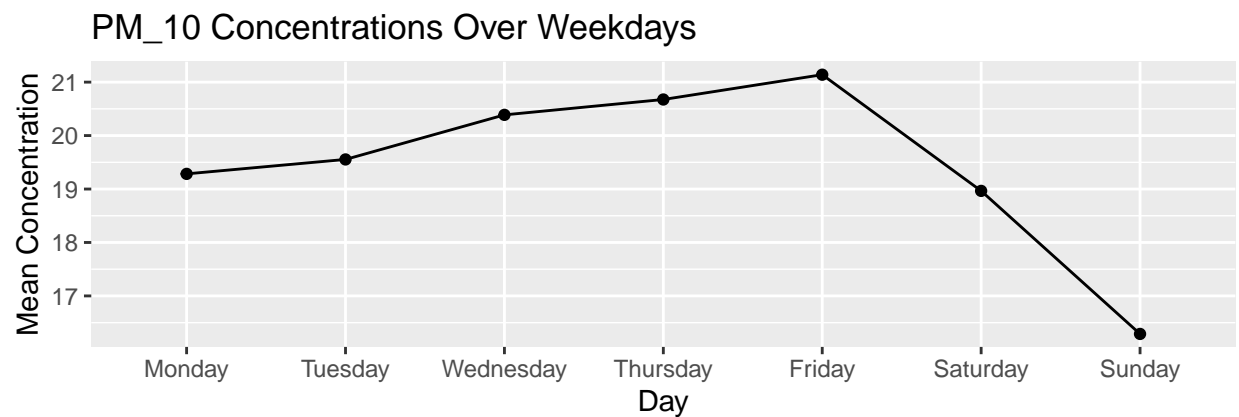
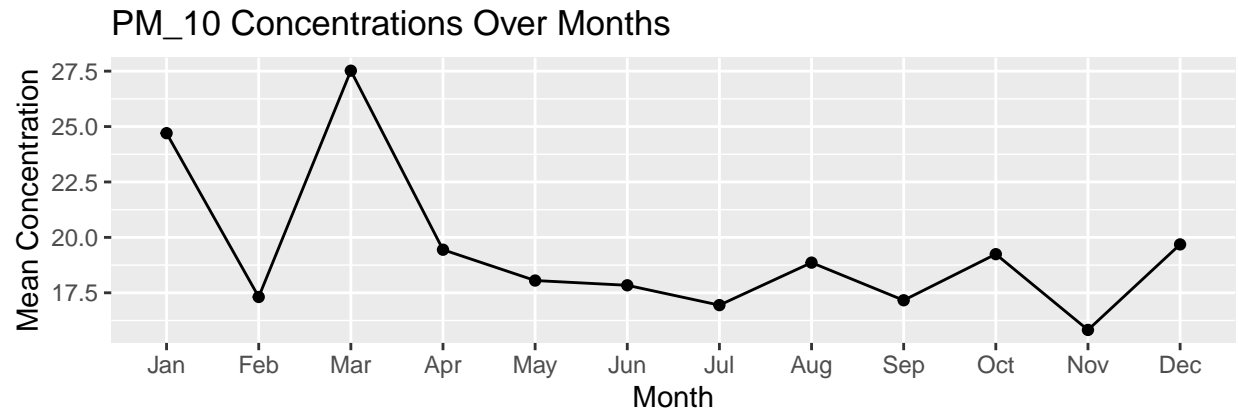


```
#ggsave(file = "../Latex Report/PM_10_Months.eps")
```

```
pm10_day <- pm10_data %>%
  group_by(Day) %>%
  summarise(pm10 = mean(pm10)) %>%
  ggplot(aes(x=factor(Day, level=c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")),
    y = pm10)) +
  geom_point() +
  geom_line(aes(group = 1)) +
  labs(title = "PM10 Concentrations Over Weekdays",
    x = "Day",
    y = "Mean Concentration")
pm10_day
```



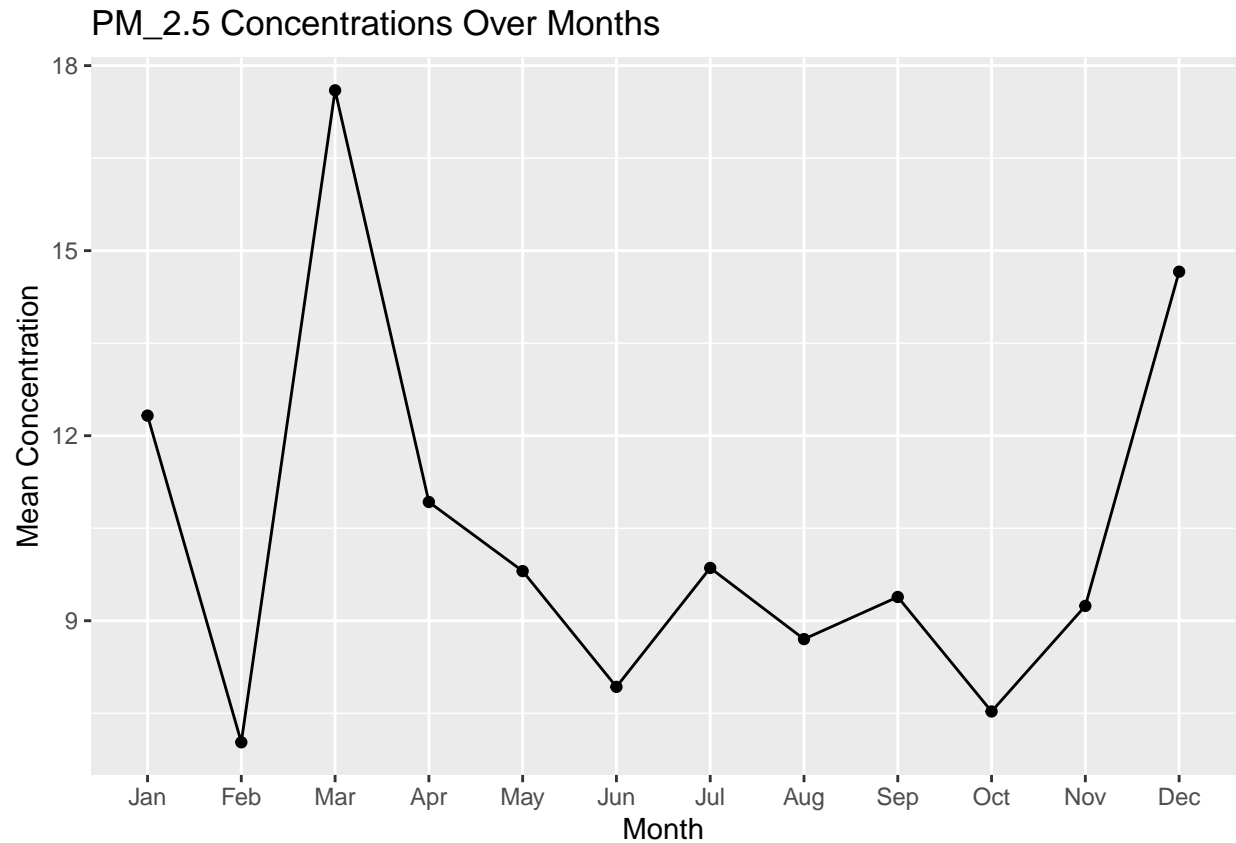
```
#ggsave(file = "../.. /Latex Report/PM_10_Days.eps")  
ggarrange(pm10_month, pm10_day, ncol=1)
```



```
suppressMessages({
  ggsave("../Latex Report/pm10_plot.eps")
})
```

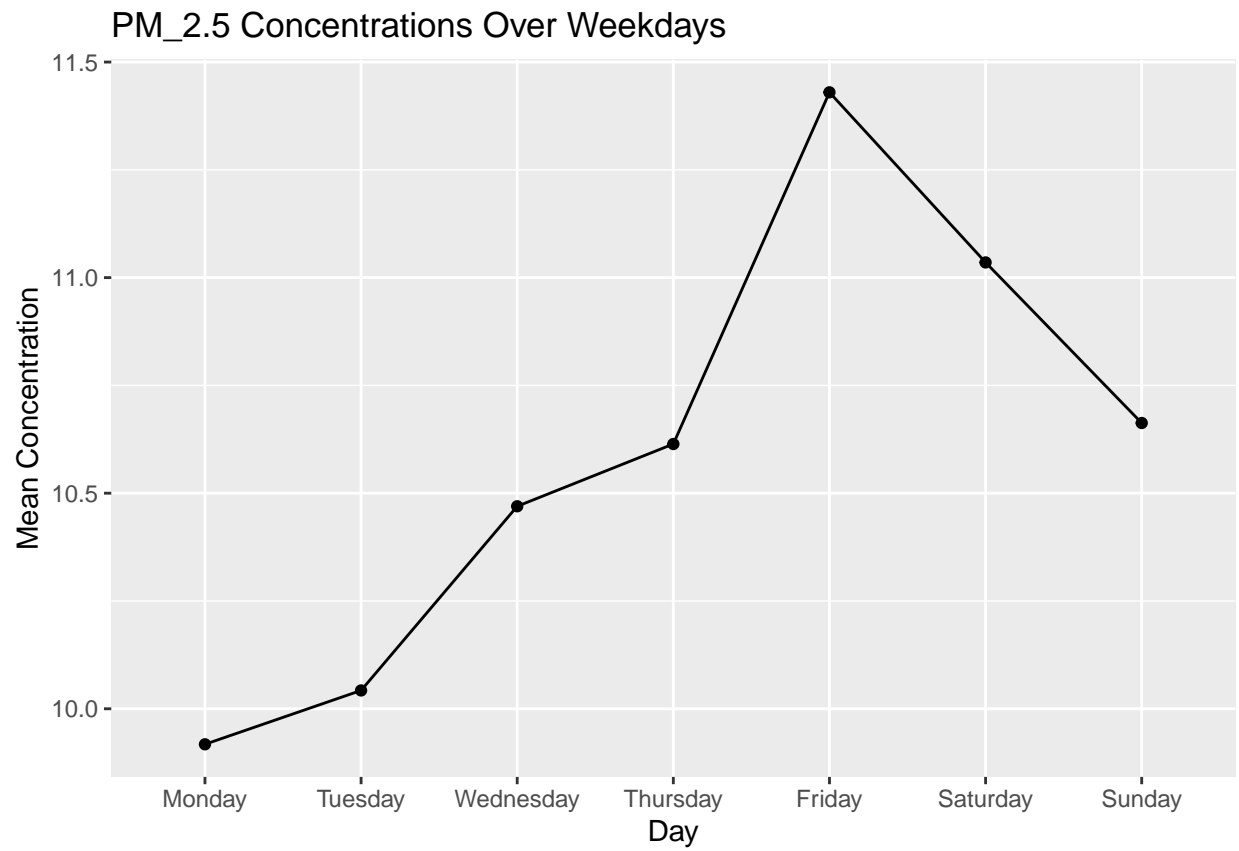
```
pm2_5_month <- pm2_5_data %>%
  group_by(Month) %>%
  summarise(pm2_5 = mean(pm2_5)) %>%
  ggplot(aes(x = Month, y = pm2_5)) +
  geom_point() +
  geom_line(aes(group = 1)) +
  labs(title = "PM2.5 Concentrations Over Months",
       x = "Month",
       y = "Mean Concentration")
pm2_5_month
```



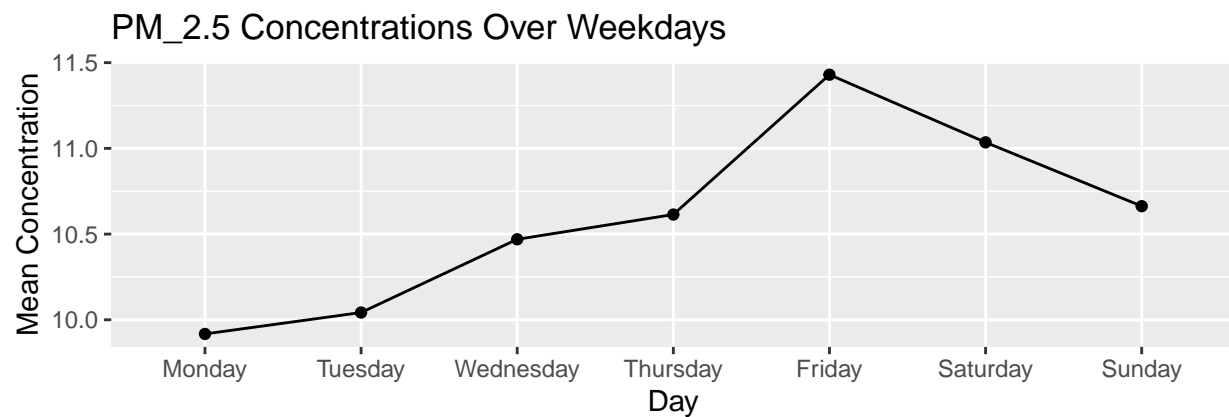
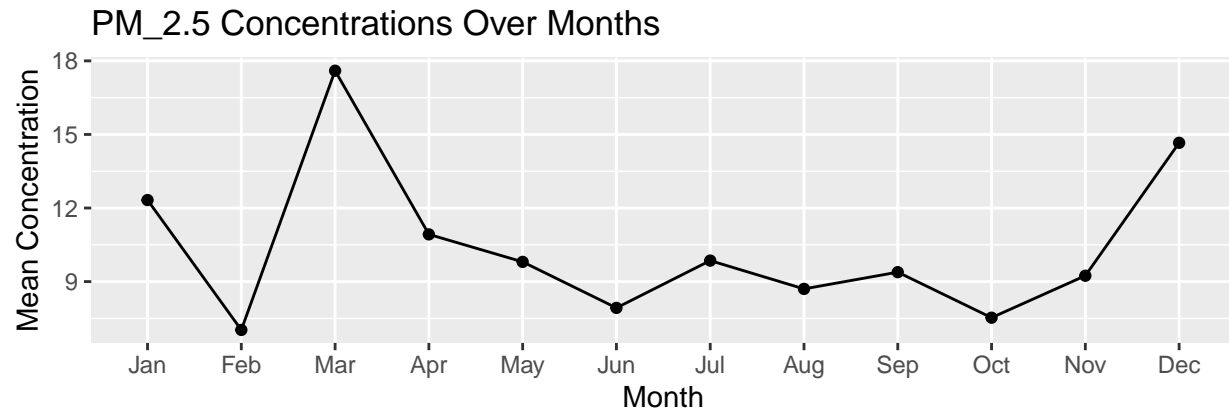


```
#ggsave(file = "../Latex Report/PM_2.5_Months.eps")

pm2_5_day <- pm2_5_data %>%
  group_by(Day) %>%
  summarise(pm2_5 = mean(pm2_5)) %>%
  ggplot(aes(x=factor(Day, level=c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"),
    y = pm2_5)) +
  geom_point() +
  geom_line(aes(group = 1)) +
  labs(title = "PM2.5 Concentrations Over Weekdays",
    x = "Day",
    y = "Mean Concentration")
pm2_5_day
```



```
#ggsave(file = "../Latex Report/PM_2_5_Days.eps")  
ggarrange(pm2_5_month, pm2_5_day, ncol=1)
```

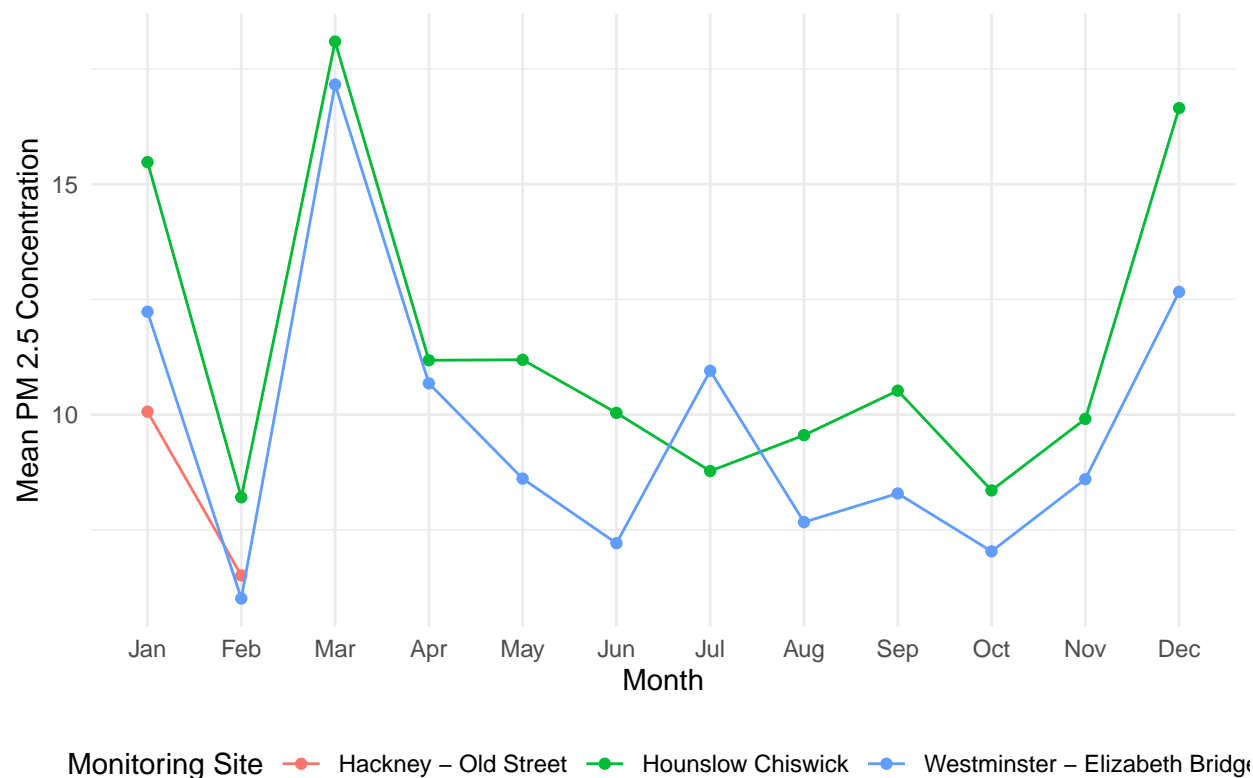


```
suppressMessages({
  ggsave("../Latex Report/pm2_5_plot.eps")
})

pm2_5_data %>%
  group_by(Month, site) %>%
  summarise(mean_pm2_5 = mean(pm2_5)) %>%
  ggplot(aes(x = Month, y = mean_pm2_5, color = factor(site))) +
  geom_point() +
  geom_line(aes(group = site)) +
  labs(
    title = "Mean PM 2.5 Concentrations Over Months for Each Monitoring Site",
    x = "Month",
    y = "Mean PM 2.5 Concentration",
    color = "Monitoring Site") +
  theme_minimal() +
  theme(legend.position = "bottom")
```

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

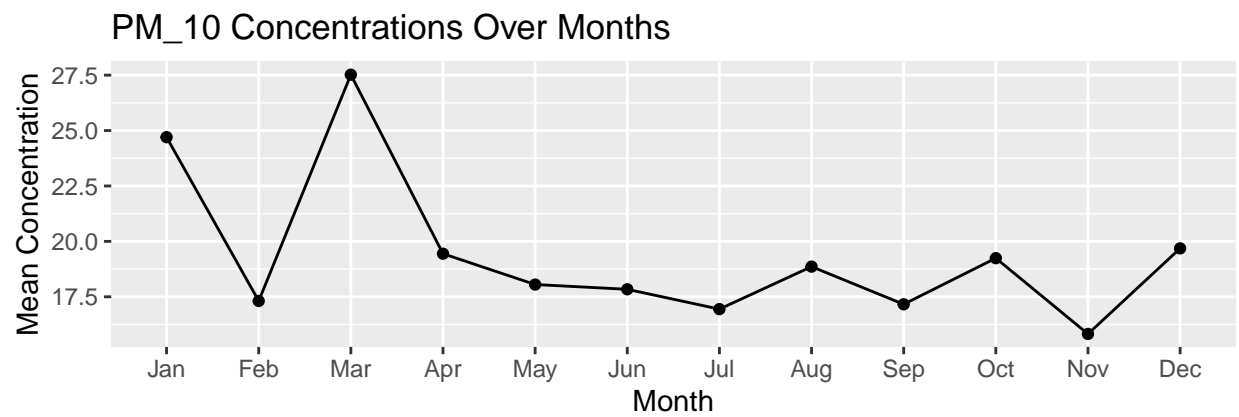
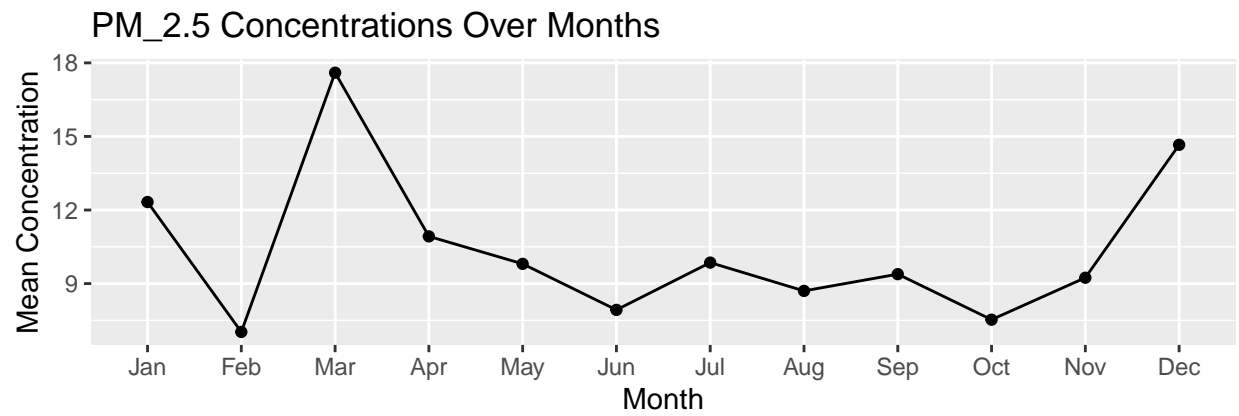
Mean PM 2.5 Concentrations Over Months for Each Monitoring Site



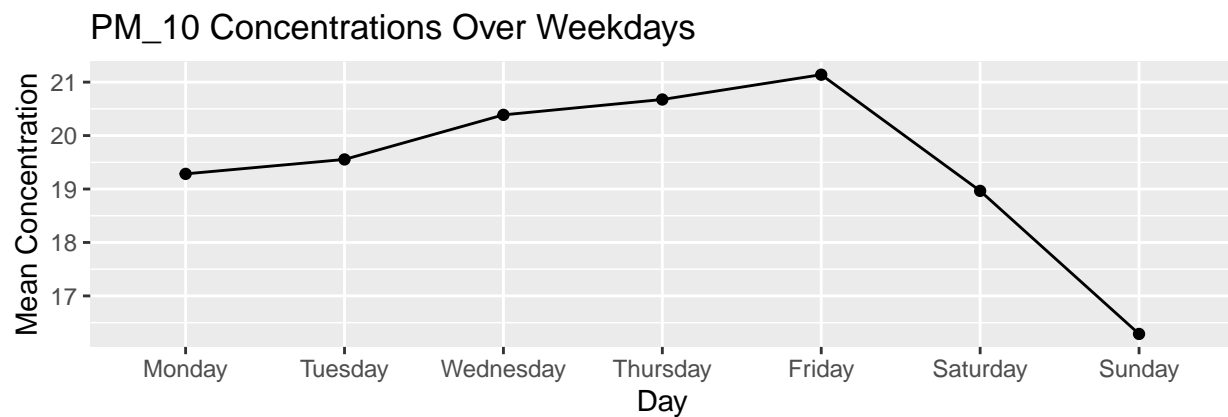
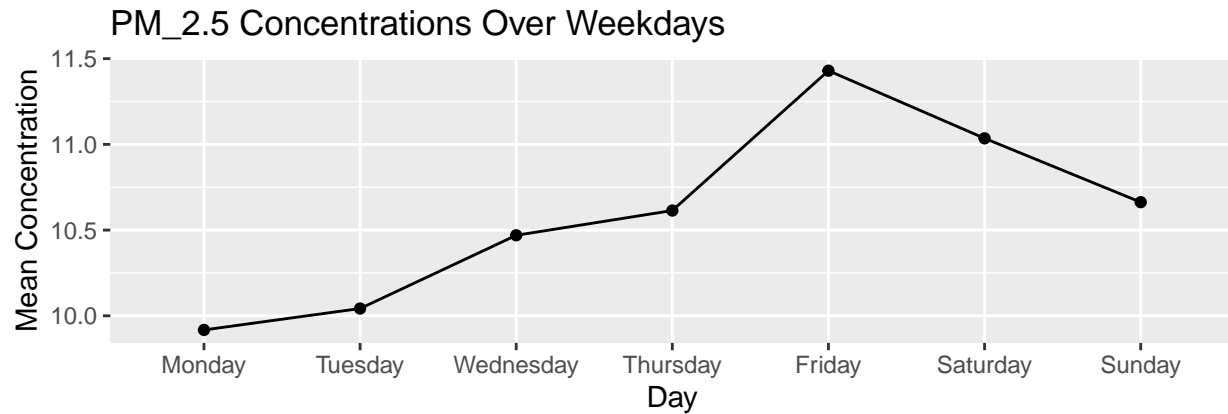
```
ggsave("../Latex Report/pm2_5_month_sites.eps")
```

```
## Saving 6.5 x 4.5 in image
```

```
ggarrange(pm2_5_month, pm10_month, ncol=1)
```



```
suppressMessages({  
  ggsave("../Latex Report/pm_month_plot.eps")  
})  
  
ggarrange(pm2_5_day, pm10_day, ncol=1)
```



```
suppressMessages({
  ggsave("../Latex Report/pm_day_plot.eps")
})
```

```
NO_month <- no_data %>%
  group_by(Month, code, site) %>%
  summarise(no = mean(no),
            no2 = mean(no2),
            nox = mean(nox))
```

## 'summarise()' has grouped output by 'Month', 'code'. You can override using the  
## '.groups' argument.

```
NO_month %>%
  arrange(desc(no)) %>%
  head() %>%
  select(c("Month", "site", "no"))
```

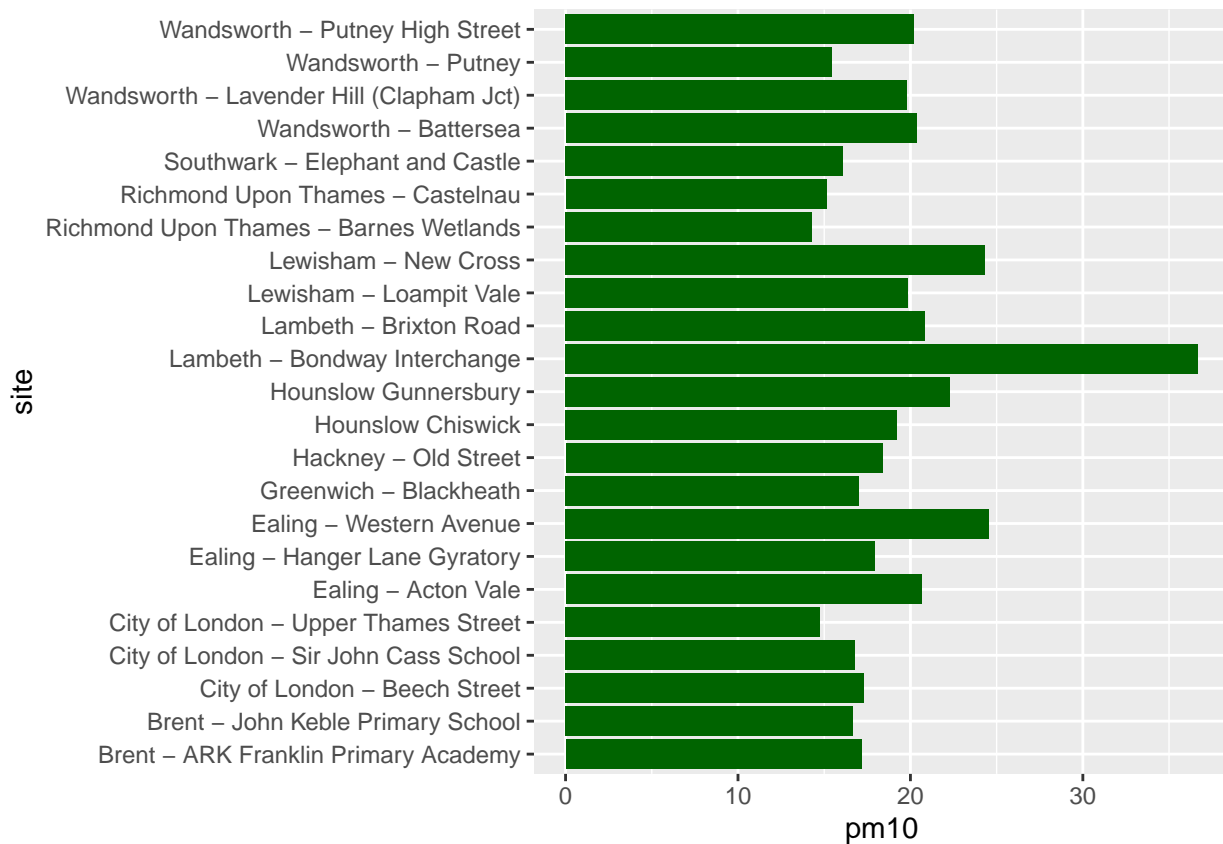
## Adding missing grouping variables: 'code'

```
## # A tibble: 6 x 4
## # Groups:   Month, code [6]
##   code Month site          no
##   <chr> <ord> <chr>      <dbl>
```

```
## 1 EA6 Jan Ealing - Hanger Lane Gyratory 110.
## 2 LW4 Oct Lewisham - Loampit Vale 105.
## 3 LB4 Jan Lambeth - Brixton Road 88.9
## 4 WA7 Jan Wandsworth - Putney High Street 86.5
## 5 LW4 Aug Lewisham - Loampit Vale 78.2
## 6 LB4 Dec Lambeth - Brixton Road 73.5
```

```
pm10_data %>%
  group_by(code,site) %>%
  summarise(pm10 = mean(pm10)) %>%
  mutate(site_x = str_extract(code, "[A-Z]+")) %>%
  ggplot(aes(x = site, y = pm10)) +
  geom_bar(stat = "identity", fill = "darkgreen") +
  coord_flip()
```

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



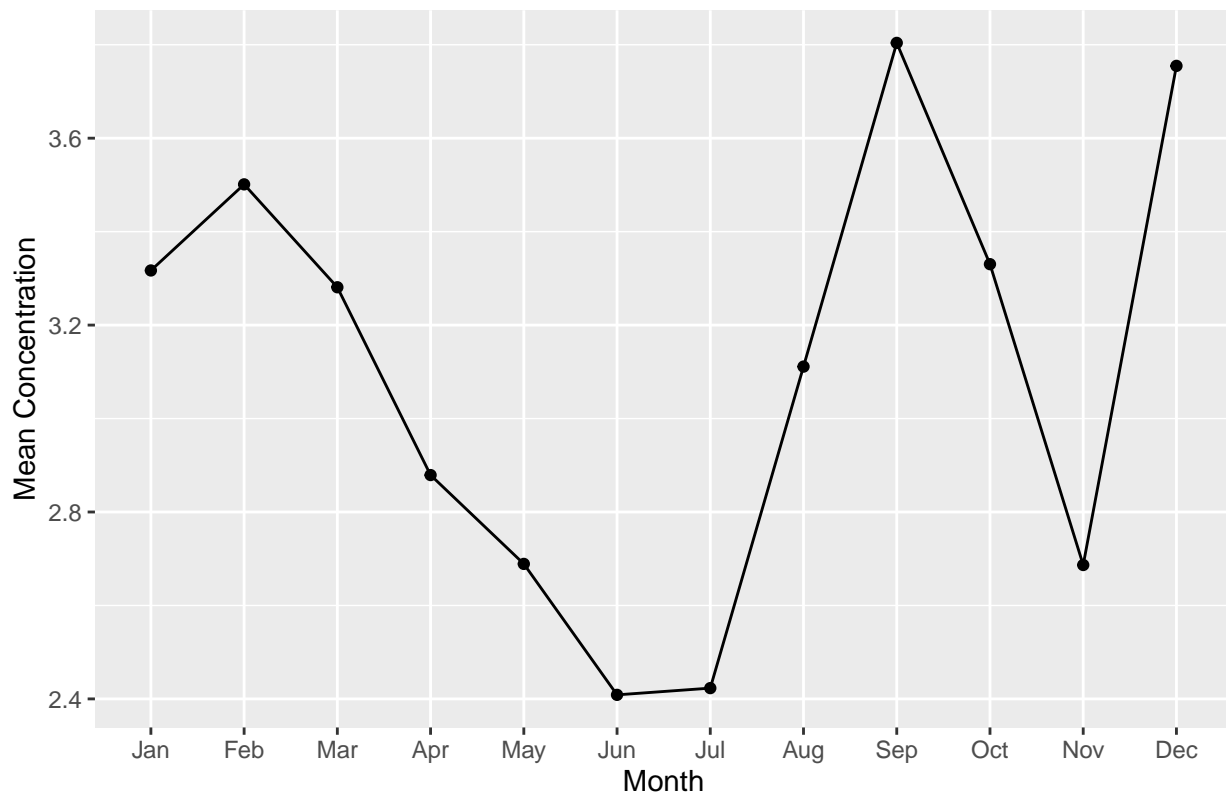
```
ggsave("../..../Latex Report/pm10_month_site.eps")
```

```
## Saving 6.5 x 4.5 in image
```

```
so2_month <- so2_data %>%
  group_by(Month) %>%
  summarise(so2 = mean(so2)) %>%
  ggplot(aes(x = Month, y = so2)) +
  geom_point() +
  geom_line(aes(group = 1)) +
  labs(title = "SO2 Concentrations over Months",
       x = "Month",
       y = "Mean Concentration",
       color = "Pollutants")

so2_month
```

SO2 Concentrations over Months



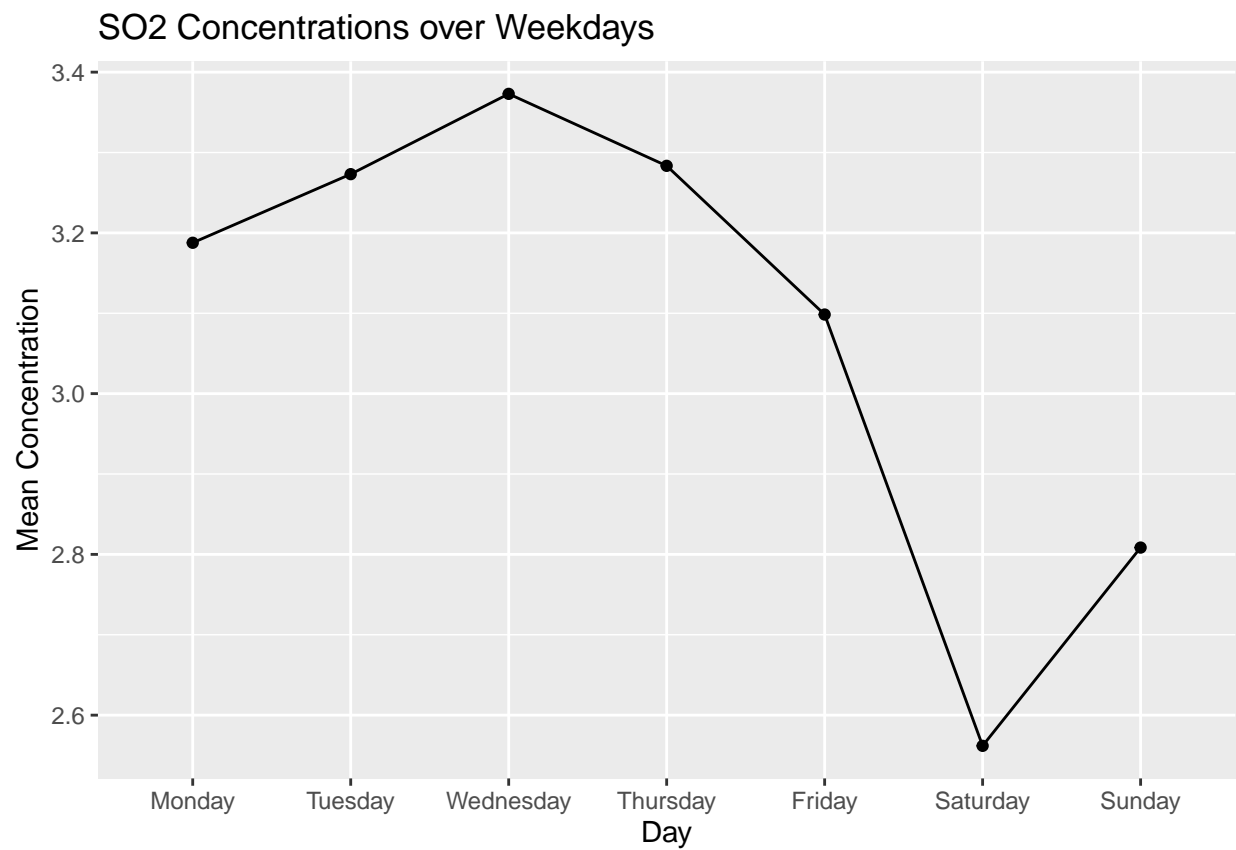
```
#ggsave(file = "../.. /Latex Report/SO2_Months.eps")

so2_day <- so2_data %>%
  group_by(Day) %>%
  summarise(so2 = mean(so2)) %>%
  ggplot(aes(x=factor(Day, level=c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"),
                             y = so2)) +
  geom_point() +
  geom_line(aes(group = 1)) +
  labs(title = "SO2 Concentrations over Weekdays",
       x = "Day",
       y = "Mean Concentration",
```



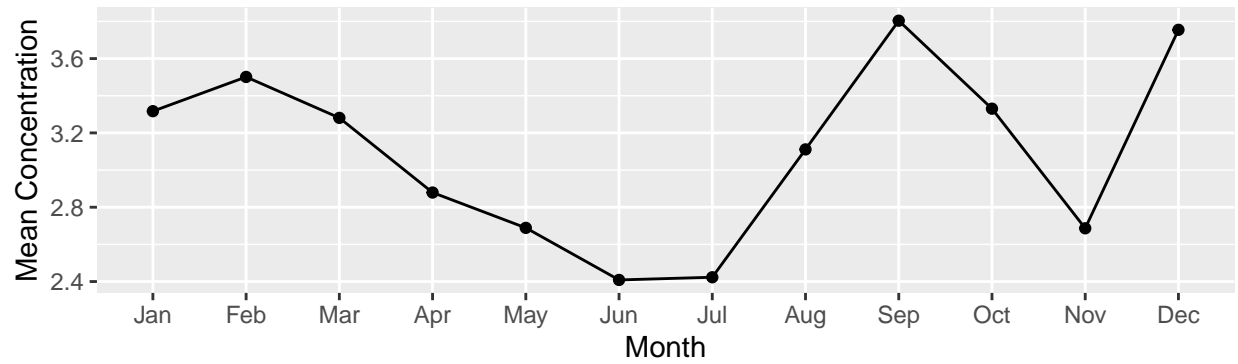
```
color = "Pollutants")
```

```
so2_day
```

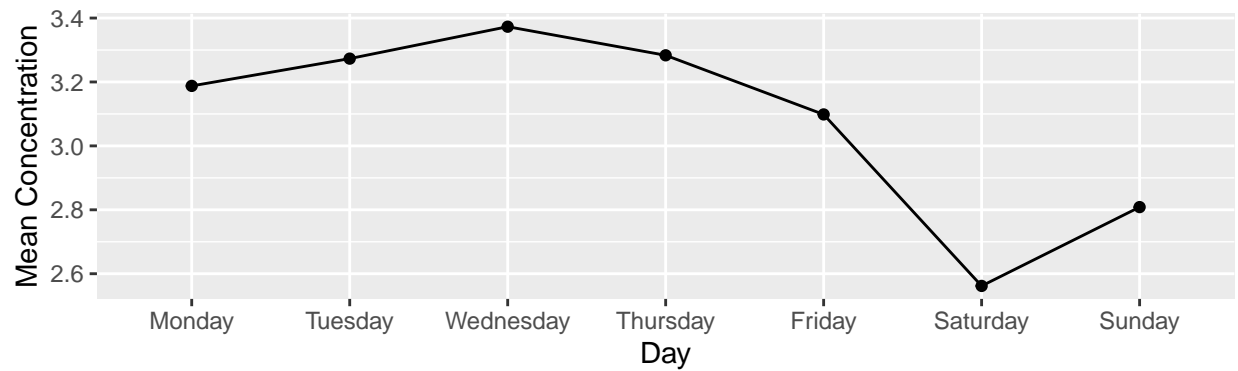


```
ggarrange(so2_month, so2_day, ncol=1)
```

SO2 Concentrations over Months



SO2 Concentrations over Weekdays



```
suppressMessages({  
  ggsave("../Latex Report/so2_plot.eps")  
})
```