Air Quality in Ukraine post Ukraine-Russia Dispute

Web address for GitHub repository

Curly Girlies: Shirley Fontanié, Rachel Gordon, and Julia Weinberg

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Research question: How does air quality in various Ukrainian cities differ before and after the Ukraine-Russia dispute?
******* Rationale: On February 24, 2022, Russia attacked Ukraine. The first city attacked was Lviv and Dnipro. Kyiv was hit February 24th.

2 Dataset Information

Describe sources of data here (input Julia paragraph)

```
# Initial wrangling

Ukraine_Processed <- UkraineData %>%
  drop_na(pm10) %>%
  drop_na(pm25)

# Setting Date
Ukraine_Processed$date <- as.Date(Ukraine_Processed$date, "%m/%d/%y")
write.csv(Ukraine_Processed, row.names = FALSE, file = "./Data/Processed/Ukraine_Process

# Creating subsets by city, month, and year

Dnipro_2021 <- Ukraine_Processed %>%
  filter(City %in% "Dnipro")%>%
  subset(date > "2021-2-28" & date < "2021-04-01") %>%
  mutate(Month = month(date),
         Day = day(date),
          Year = as.factor(year(date)))
write.csv(Dnipro_2021, row.names = FALSE, file = "./Data/Processed/Dnipro_2021.csv")


Dnipro_2022 <- Ukraine_Processed %>%
  filter(City %in% "Dnipro")%>%
  subset(date > "2022-2-28" & date < "2022-04-01") %>%
   mutate(Month = month(date),
         Day = day(date),
          Year = as.factor(year(date)))

write.csv(Dnipro_2022, row.names = FALSE, file = "./Data/Processed/Dnipro_2022.csv")

Lviv_2021 <- Ukraine_Processed %>%
  filter(City %in% "Lviv") %>%
  subset(date > "2021-2-28" & date < "2021-04-01") %>%
  mutate(Month = month(date),
         Day = day(date),
          Year = as.factor(year(date)))
write.csv(Lviv_2021, row.names = FALSE, file = "./Data/Processed/Lviv_2021.csv")

Lviv_2022 <- Ukraine_Processed %>%
  filter(City %in% "Lviv") %>%
  subset(date > "2022-2-28" & date < "2022-04-01") %>%
```

```
  mutate(Month = month(date),
         Day = day(date),
          Year = as.factor(year(date)))
write.csv(Lviv_2022, row.names = FALSE, file = "./Data/Processed/Lviv_2022.csv")

#creating subsets by city
FULL_DNIPRO <- bind_rows(Dnipro_2021,Dnipro_2022)
write.csv(FULL_DNIPRO, row.names = FALSE, file = "./Data/Processed/FULL_DNIPRO.csv")

FULL_LVIV <- bind_rows(Lviv_2021,Lviv_2022)
write.csv(FULL_LVIV, row.names = FALSE, file = "./Data/Processed/FULL_LVIV.csv")
```

Explain data wrangling process here (shirley do this)

Data File Name	Description
UkraineData	(Raw) Ukraine air quality data
Ukraine_Processed	(Processed) Ukraine air quality data, w/o na's
Dnipro_2021	Dnipro PM2.5 + PM10, Mar 2021
Dnipro_2022	Dnipro PM2.5+ PM10, Mar 2022
Lviv_2021	Lviv PM 2.5 + PM10, Mar 2021
Lviv_2022	Lviv PM 2.5 + PM10, Mar 2022
FULL_DNIPRO	Lviv_2021 + Lviv_2022 combined
FULL_LVIV	Dnipro_2021 + Dnipro_2022 combined

3 Exploratory Analysis

Table 2: PM2.5 Levels by City

City	Mean	Min	Max	Std Dev
Dnipro	49.41546	4	160	25.91608
Lviv	60.51086	8	518	34.79405

Table 3: PM10 Levels by City

City	Mean	Min	Max	Std Dev
Dnipro	24.73309	2	120	15.82186
Lviv	30.29246	4	606	26.78330

```
# DNIPRO PM2.5

#group=as.Date(year(Dnipro_2021$date)), color =as.Date(year(Dnipro_2021$date)))
#aes(month(date, label = TRUE, abbr = TRUE))

PM25_Dnipro_PLOT <-
  ggplot(FULL_DNIPRO) +
  (aes(x = Day, y = pm25, color = Year))+
  geom_line()+
  geom_point()+
  labs(x= "Date", y = "PM2.5",
       title = " Observing PM2.5 Values in Dnipro, Ukraine")
print(PM25_Dnipro_PLOT)
```

```
  #annotate(geom = "text", x = as.Date("2022-03-11"), y = 50, label ="March 11, 2022:"
  #annotate(geom = "text", x = as.Date("2022-03-11"), y = 45, label ="Missile Attack")




# LVIV PM2.5
PM25_Lviv_PLOT <-
  ggplot(FULL_LVIV) +
(aes(x = Day, y = pm25, color = Year)) +
              geom_line()+
  geom_point()+
  labs(x = "Date", y= "PM2.5",
       title = "Observing PM2.5 Values in Lviv Ukraine")
```

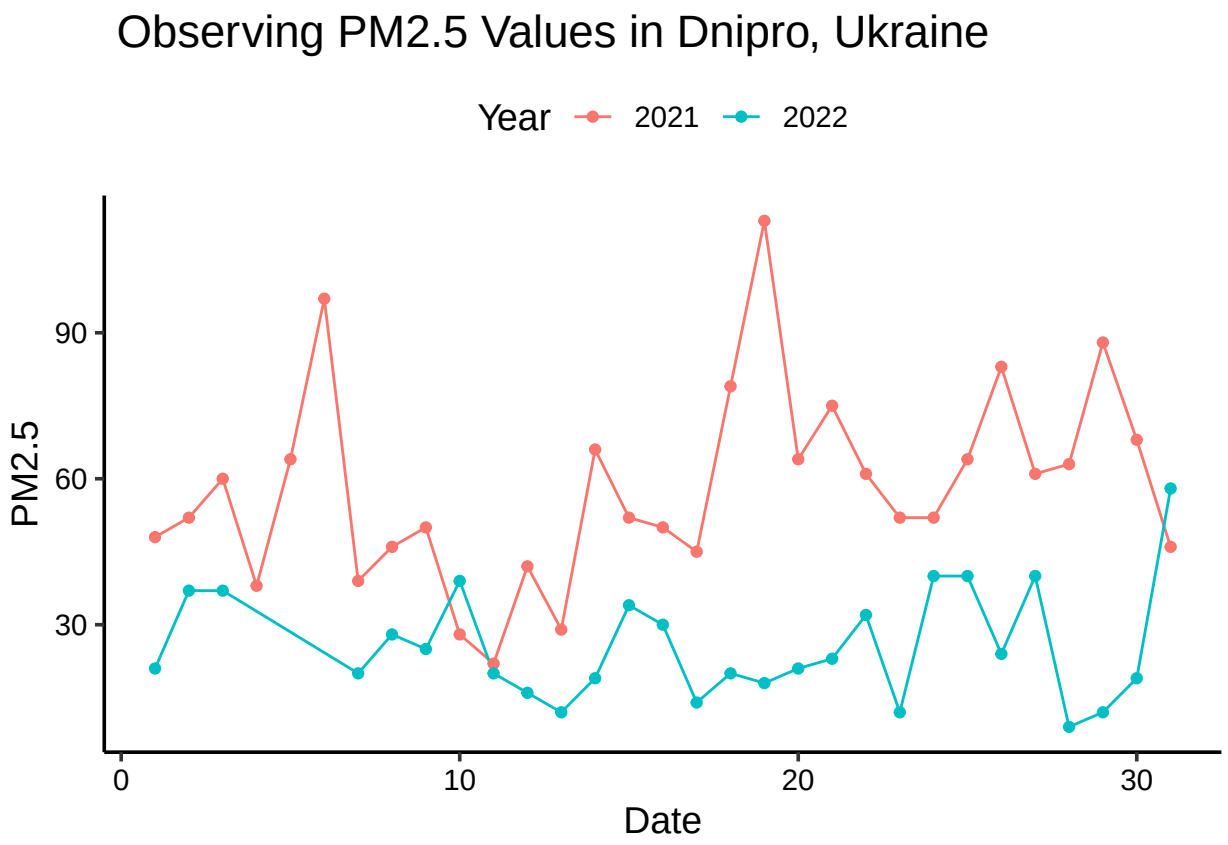

Figure 1: Comparing PM2.5 in Dnipro and Lviv, Ukraine

```
print(PM25_Lviv_PLOT)
```

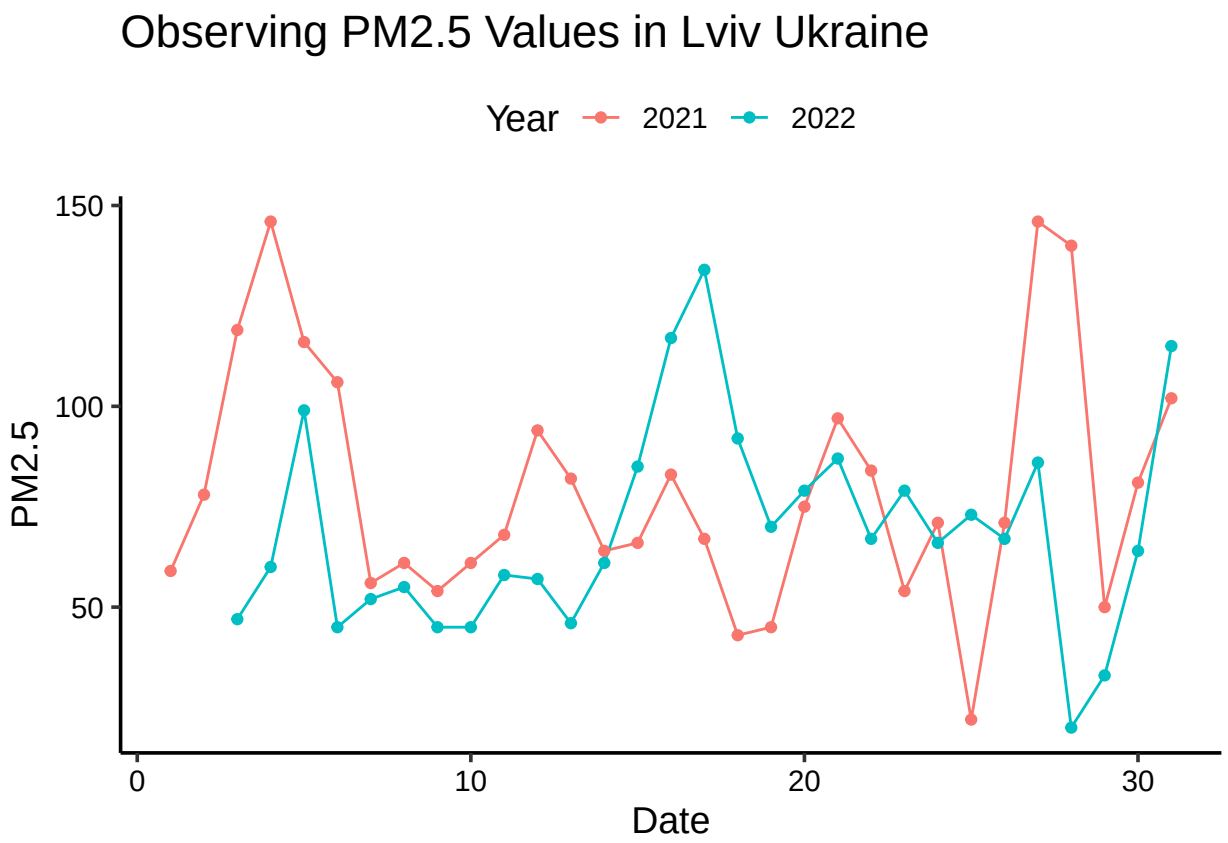

Figure 2: Comparing PM2.5 in Dnipro and Lviv, Ukraine

```
  #scale_x_date(limits = as.Date(c("2022-03-01", "2022-03-31")))+
  #xlab("Date")+
  #ylab("PM2.5")+
 # labs(title = "        Observing PM2.5 Values in Lviv, Ukraine")+
  #ylim(0,180)+
 # annotate(geom = "text", x = as.Date("2022-03-18"), y = 165, label ="March 18, 2022:
  #annotate(geom = "text", x = as.Date("2022-03-18"), y = 155, label ="1st Missile Att
  #annotate(geom = "text", x = as.Date("2022-03-26"), y = 110, label ="March 26, 2022:
  #annotate(geom = "text", x = as.Date("2022-03-26"), y = 100, label ="2nd Missile Att
```

```
#DNIPRO PM2.5
PM10_Dnipro_PLOT <-ggplot(FULL_DNIPRO) +
  (aes(x = Day, y = pm10, color = Year))+
  geom_line()+
  geom_point()+
  labs(x= "Date", y = "PM10",
       title = " Observing PM10 Values in Dnipro, Ukraine")
print(PM10_Dnipro_PLOT)
```

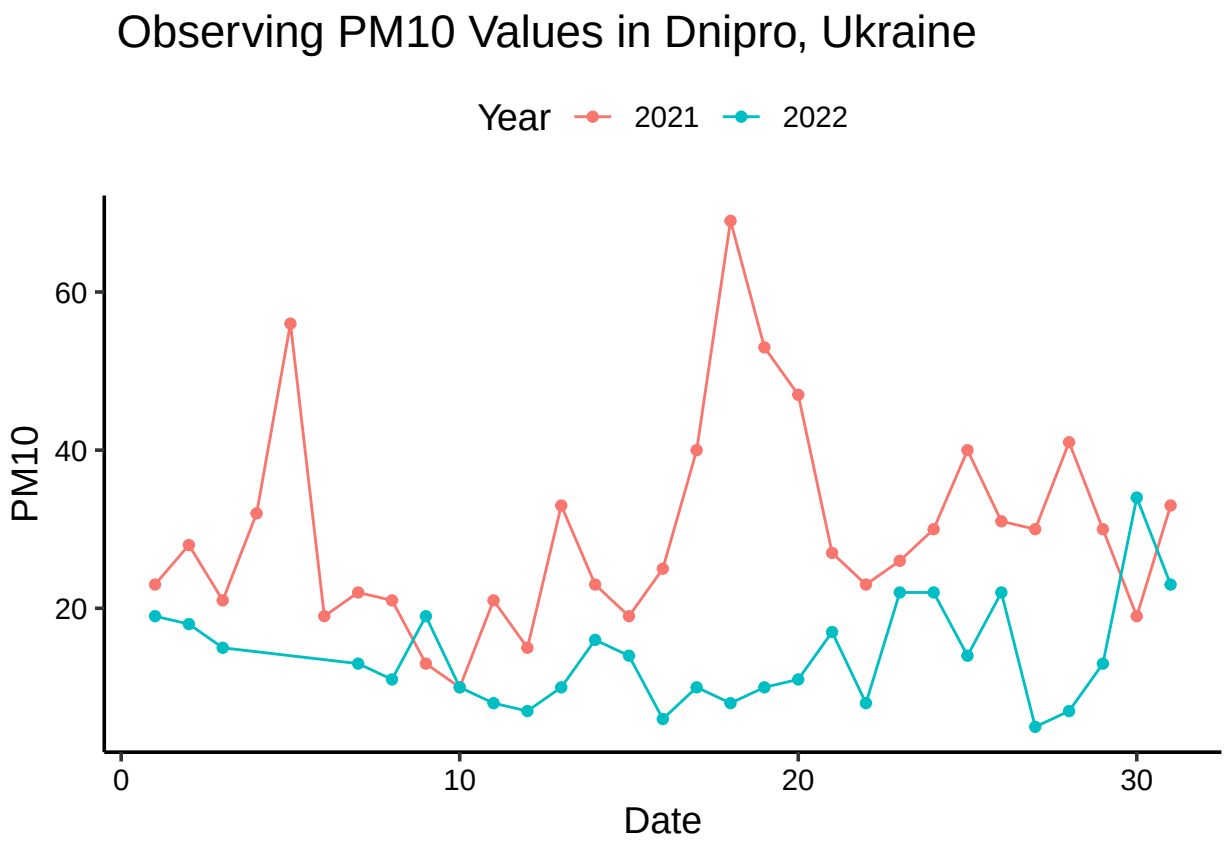

Figure 3: Comparing PM10 in Dnipro and Lviv, Ukraine

```
            #  geom_line()+
 # geom_point()+
  #scale_x_date(limits = as.Date(c("2022-03-01", "2022-03-31")))+
  #xlab("Date")+
  #ylab("PM10")+
  #labs(title = "        Observing PM10 Values in Dnipro, Ukraine")+
  #ylim(0,40)+
  #annotate(geom = "text", x = as.Date("2022-03-11"), y = 25, label ="March 11, 2022:"
  #annotate(geom = "text", x = as.Date("2022-03-11"), y = 23, label ="Missile Attack")



#LVIV PM10
PM10_Lviv_PLOT <-
  ggplot(FULL_LVIV) +
(aes(x = Day, y = pm10, color = Year)) +
              geom_line()+
  geom_point()+
  labs(x = "Date", y= "PM10",
       title = "Observing PM10 Values in Lviv Ukraine")
print(PM10_Lviv_PLOT)
```

```
  #geom_point()+
  #scale_x_date(limits = as.Date(c("2022-03-01", "2022-03-31")))+
  #xlab("Date")+
  #ylab("PM10")+
  #labs(title = "        Observing PM10 Values in Lviv, Ukraine")+
  #ylim(0,75)+
  #annotate(geom = "text", x = as.Date("2022-03-18"), y = 70, label ="March 18, 2022:"
  #annotate(geom = "text", x = as.Date("2022-03-18"), y = 66, label ="1st Missile Atta
  #annotate(geom = "text", x = as.Date("2022-03-26"), y = 60, label ="March 26, 2022:"
  #annotate(geom = "text", x = as.Date("2022-03-26"), y = 56, label ="2nd Missile Atta
```

```
FULL_Air_quality <- bind_rows(Dnipro_2022,Lviv_2022)

PM25_Lviv_and_Dnipro_PLOT <-
  ggplot(FULL_Air_quality) +
(aes(x = Day, y = pm25, color = City)) +
              geom_line()+
  geom_point()+
  labs(x = "Date", y= "PM25",
       title = "Observing PM25 Values in 2022 Lviv and Dnipro Ukraine")
print(PM25_Lviv_and_Dnipro_PLOT)
```

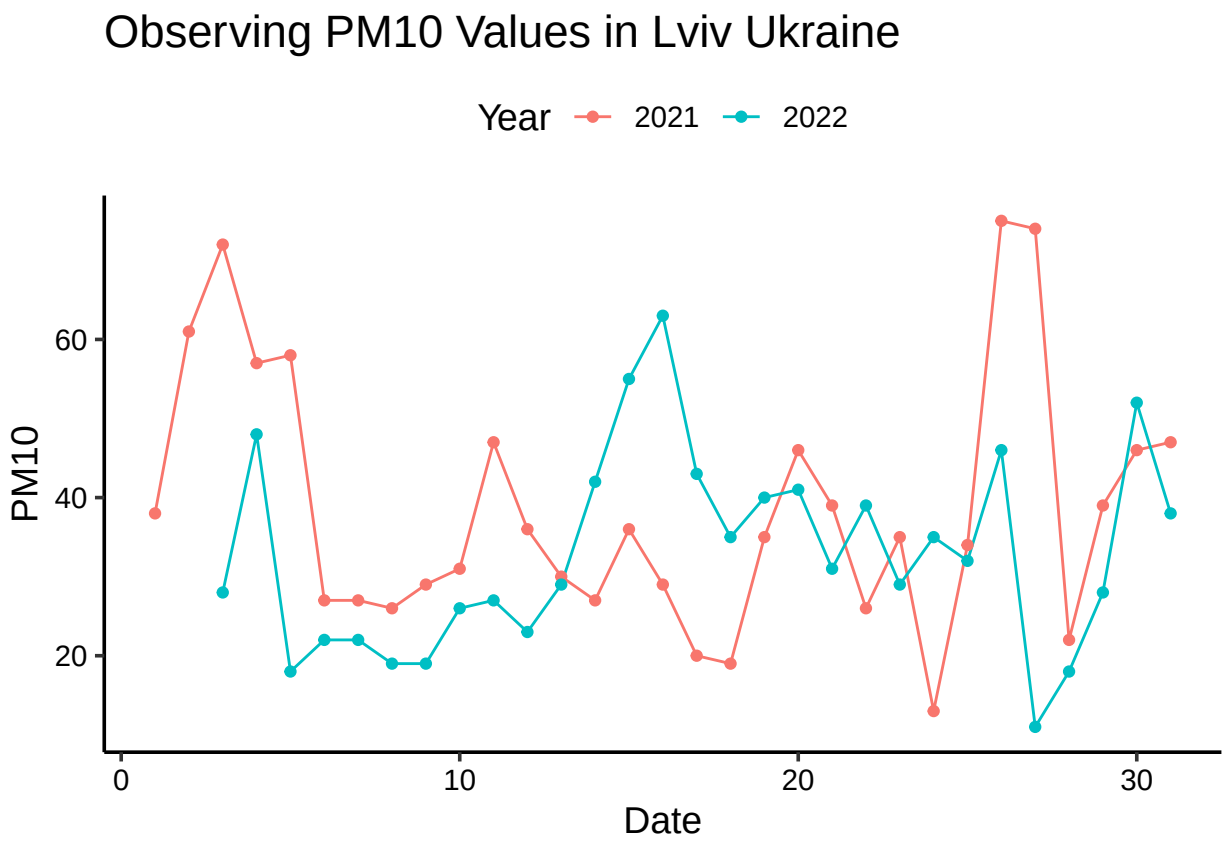

Figure 4: Comparing PM10 in Dnipro and Lviv, Ukraine

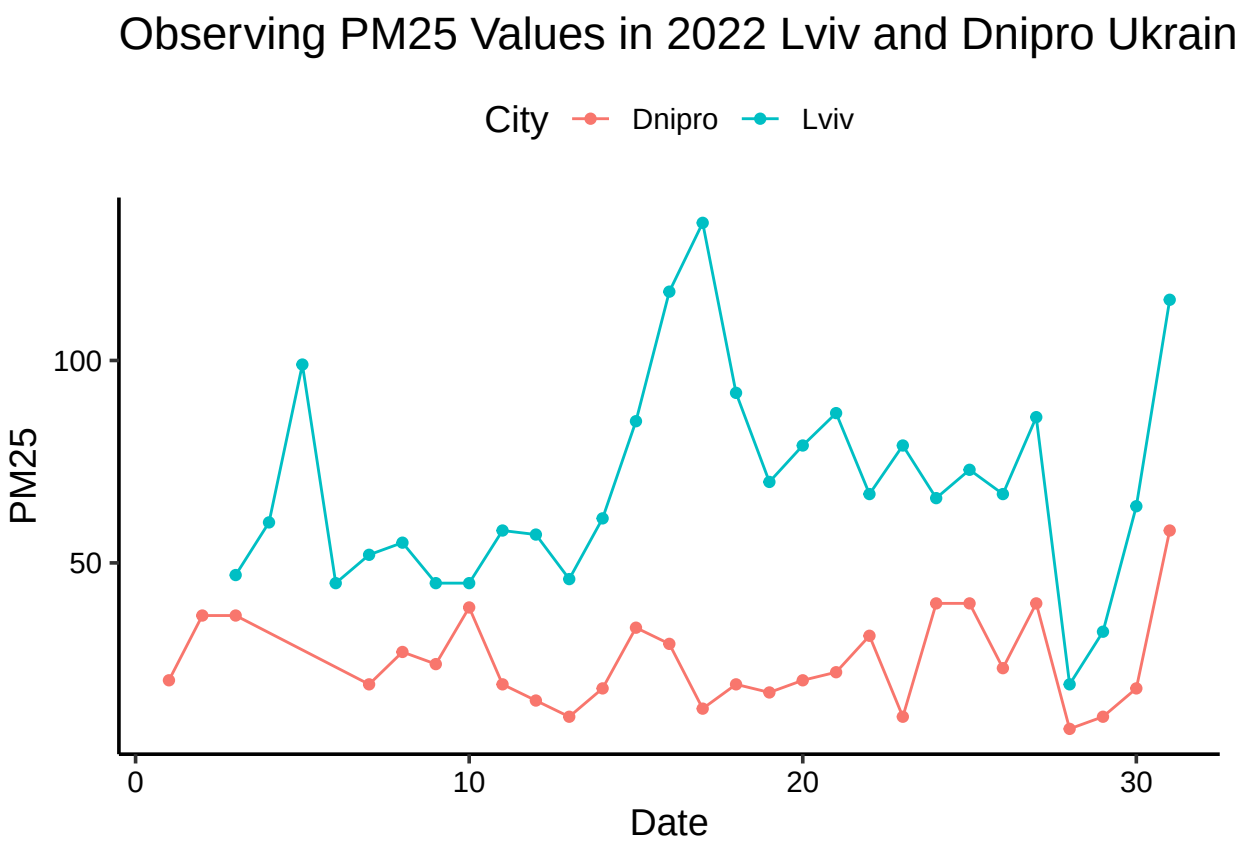

Figure 5: Comparing Air Pollution in Lviv vs Dnipro

```
PM10_Lviv_and_Dnipro_PLOT <-
  ggplot(FULL_Air_quality) +
(aes(x = Day, y = pm10, color = City)) +
              geom_line()+
  geom_point()+
  labs(x = "Date", y= "PM10",
       title = "Observing PM10 Values in 2022 Lviv and Dnipro Ukraine")
print(PM10_Lviv_and_Dnipro_PLOT)
```

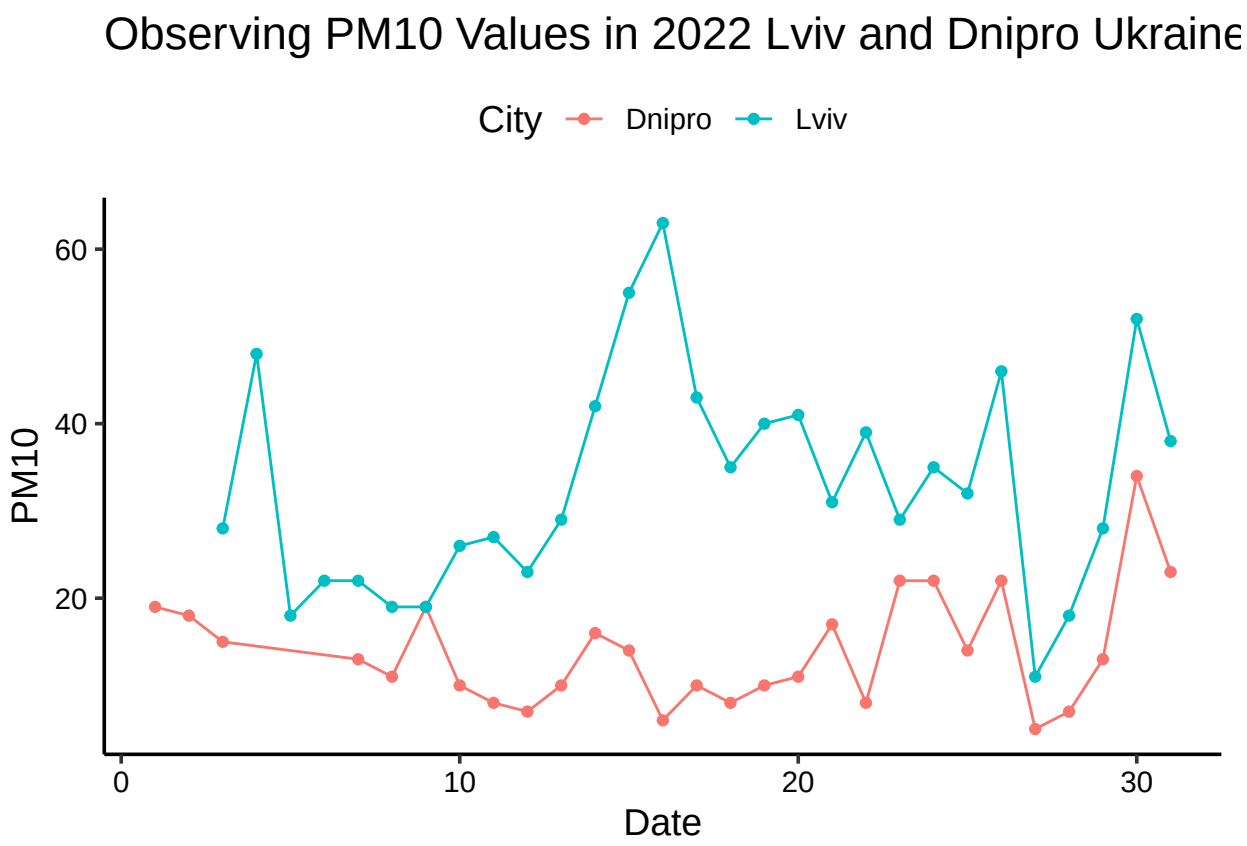

Figure 6: Comparing Air Pollution in Lviv vs Dnipro

4 Analysis

```
#after doing all these anovas i realized im not sure what this is really telling us an

lviv.pm25.anova <- aov(data = FULL_LVIV , pm25 ~ Year)
summary(lviv.pm25.anova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Year          1   1585  1584.5   1.988  0.164
## Residuals    58  46232   797.1
```

```
lviv.pm10.anova <- aov(data = FULL_LVIV , pm10 ~ Year)
summary(lviv.pm10.anova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Year          1    482   482.2   2.281  0.136
## Residuals    58  12262   211.4
```

```
dnipro.pm25.anova <- aov(data = FULL_DNIPRO , pm25 ~ Year)
summary(dnipro.pm25.anova)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Year          1  15305   15305   55.92 5.07e-10 ***
## Residuals    57  15599     274
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
dnipro.pm10.anova <- aov(data = FULL_DNIPRO , pm10 ~ Year)
summary(dnipro.pm10.anova)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Year          1   3616    3616   32.57 4.33e-07 ***
## Residuals    57   6329     111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
dnipro.lviv.pm25.anova <- aov(data = FULL_Air_quality , pm25 ~ City)
summary(dnipro.lviv.pm25.anova)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## City          1  26819   26819   66.65 4.72e-11 ***
## Residuals    55  22130     402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
dnipro.lviv.pm10.anova <- aov(data = FULL_Air_quality , pm10 ~ City)
summary(dnipro.lviv.pm10.anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## City         1   5180    5180   51.52 1.93e-09 ***
## Residuals   55   5530     101
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lviv.25.lm <- lm(data = FULL_LVIV, pm25 ~ Year)
summary(lviv.25.lm)
```

```
##
## Call:
## lm(formula = pm25 ~ Year, data = FULL_LVIV)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.387 -18.887  -4.745  16.147  66.613
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   79.387      5.071   15.66   <2e-16 ***
## Year2022     -10.284      7.294   -1.41    0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.23 on 58 degrees of freedom
## Multiple R-squared:  0.03314,    Adjusted R-squared:  0.01647
## F-statistic: 1.988 on 1 and 58 DF,  p-value: 0.1639
```

```
lviv.10.lm <- lm(data = FULL_LVIV, pm10 ~ Year)
summary(lviv.10.lm)
```

```
##
## Call:
## lm(formula = pm10 ~ Year, data = FULL_LVIV)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.742 -11.069  -3.242   8.013  36.258
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   38.742      2.611   14.84   <2e-16 ***
## Year2022      -5.673      3.756   -1.51    0.136
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 14.54 on 58 degrees of freedom
## Multiple R-squared:  0.03784,    Adjusted R-squared:  0.02125
## F-statistic: 2.281 on 1 and 58 DF,  p-value: 0.1364
```

```
dnipro.25.lm <- lm(data = FULL_DNIPRO, pm25 ~ Year)
summary(dnipro.25.lm)
```

```
##
## Call:
## lm(formula = pm25 ~ Year, data = FULL_DNIPRO)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.968  -9.841  -4.714   8.159  55.032
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   57.968      2.971  19.510  < 2e-16 ***
## Year2022     -32.253      4.313  -7.478 5.07e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.54 on 57 degrees of freedom
## Multiple R-squared:  0.4952, Adjusted R-squared:  0.4864
## F-statistic: 55.93 on 1 and 57 DF,  p-value: 5.074e-10
```

```
dnipro.10.lm <- lm(data = FULL_DNIPRO, pm10 ~ Year)
summary(dnipro.10.lm)
```

```
##
## Call:
## lm(formula = pm10 ~ Year, data = FULL_DNIPRO)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.677  -6.677  -1.677   3.661  39.323
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   29.677      1.893  15.681  < 2e-16 ***
## Year2022     -15.677      2.747  -5.707 4.33e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.54 on 57 degrees of freedom
## Multiple R-squared:  0.3636, Adjusted R-squared:  0.3524
```

```
## F-statistic: 32.57 on 1 and 57 DF,  p-value: 4.325e-07
```

```
dnipro.lviv.pm25.lm <- lm(data = FULL_Air_quality, pm25 ~ City)
summary(dnipro.lviv.pm25.lm)
```

```
##
## Call:
## lm(formula = pm25 ~ City, data = FULL_Air_quality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.103 -11.714  -3.103  11.286  64.897
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   25.714      3.791   6.783 8.54e-09 ***
## CityLviv      43.389      5.315   8.164 4.72e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.06 on 55 degrees of freedom
## Multiple R-squared:  0.5479, Adjusted R-squared:  0.5397
## F-statistic: 66.65 on 1 and 55 DF,  p-value: 4.715e-11
```

```
dnipro.lviv.pm10.lm <- lm(data = FULL_Air_quality, pm10 ~ City)
summary(dnipro.lviv.pm10.lm)
```

```
##
## Call:
## lm(formula = pm10 ~ City, data = FULL_Air_quality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.069  -6.000  -1.069   5.931  29.931
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   14.000      1.895   7.388 8.72e-10 ***
## CityLviv      19.069      2.657   7.178 1.93e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.03 on 55 degrees of freedom
## Multiple R-squared:  0.4837, Adjusted R-squared:  0.4743
## F-statistic: 51.52 on 1 and 55 DF,  p-value: 1.929e-09
```

```
#I'm confused if we need to upload a shapefile of ukraine so that we can make a map bu

#https://simplemaps.com/data/ua-cities
#we can download a csv of all the cities with lat and long from this link and we could
```

4.1 Question 1: Are there significant differences in air quality levels between affected Ukrainian cities during the Russian invasion?

4.2 Question 2: Do air quality levels worsen in affected cities around missile attack events?

4.3 Question 3: Are there significant differences in air quality levels in affected Ukrainian cities before and during the Russian attacks?

5 Summary and Conclusions

6 References

<add references here if relevant, otherwise delete this section>