NY Pollution

Juhi Malwade

2025-09-02

```r
library(tidyverse)
library(lubridate)

ny_pollution = read_csv('ny_pollution.csv.gz')

pollution_year = ny_pollution %>%
  group_by(year = lubridate::floor_date(date, "year")) %>%
  summarize(mean_pollution = mean(pollution, na.rm = TRUE))

pollution_year
```
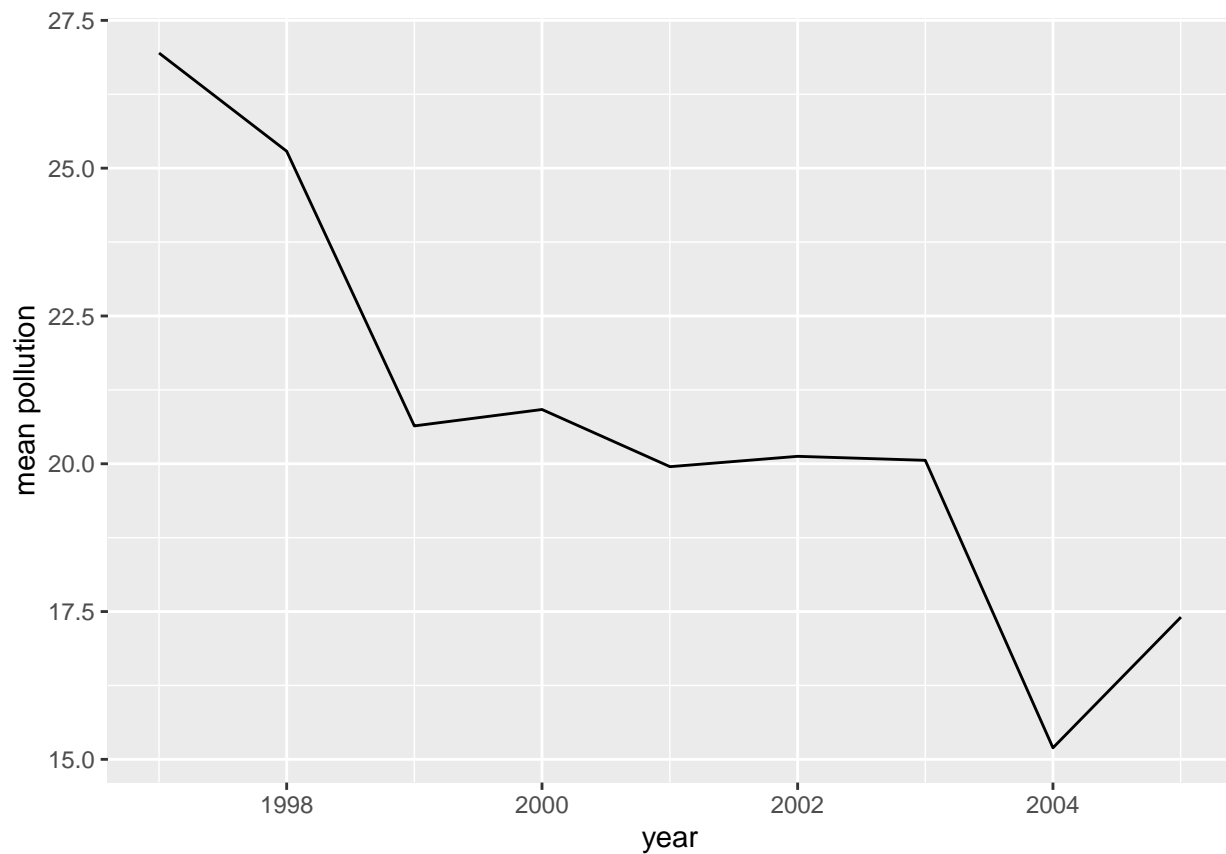
```
## # A tibble: 9 x 2
##   year       mean_pollution
##   <date>              <dbl>
## 1 1997-01-01           26.9
## 2 1998-01-01           25.3
## 3 1999-01-01           20.6
## 4 2000-01-01           20.9
## 5 2001-01-01           20.0
## 6 2002-01-01           20.1
## 7 2003-01-01           20.1
## 8 2004-01-01           15.2
## 9 2005-01-01           17.4
```

```r
ggplot(pollution_year) + aes(x = year, y = mean_pollution) + geom_line() + ylab("mean pollution")
```

The mean pollution per year in New York has decreased from 26.9 in 1997 to 17.4 in 2005. This decrease was more rapid from 1997-1999 and 2003-2004. However, there was a slight increase from 15.2 in 2004 to 17.4 in 2005. If we had more data from the years after 2005, we would be able to conclude if this slight increase continues or if the overall downward trend persists as New York takes more measures to decrease pollution.