Pollution Analysis

Isaiah Benny

2025-09-02

Which Year Had the Most Deaths?

```
library(tidyverse)
library(knitr)
```

```
pollution <- read_csv("ny_pollution.csv.gz")
```

```
## Rows: 3287 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (2): death, pollution
## date (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.
```

```
death_by_year <- pollution %>% group_by(Year = year(date)) %>%
  summarize(Deaths = sum(death)) %>% arrange(desc(Deaths))

kable(death_by_year[1,])
```

Year	Deaths
1999	63186

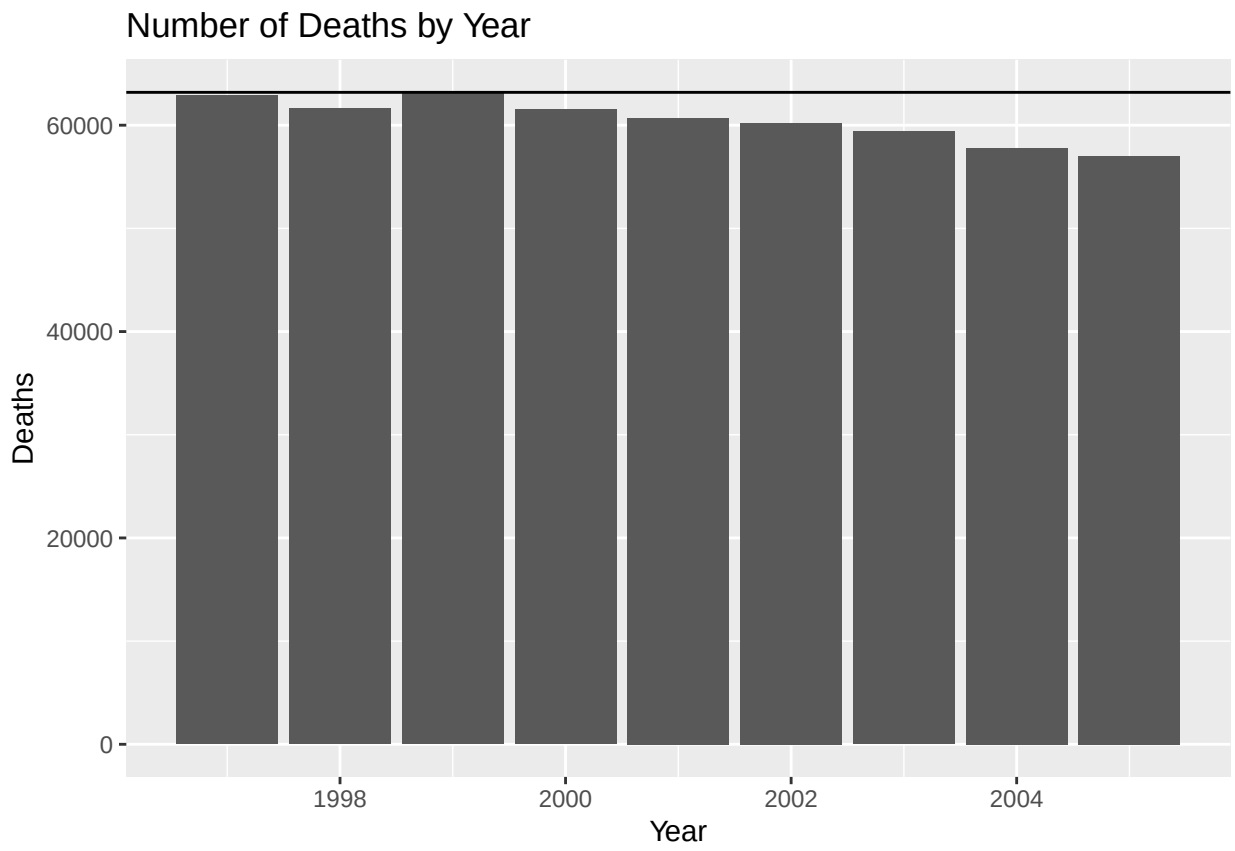
Statement

The year with the highest number of deaths is 1999

Supporting Premise

To support this conclusion, I will produce a graph showing the number of deaths for each year.

```
ggplot(death_by_year) + aes(x=Year,y=Deaths) + geom_col() +
  ggtitle("Number of Deaths by Year") + geom_hline(yintercept = max(death_by_year$Deaths))
```



Supporting Premise

This bar plot shows that the years other than 1999 have lower deaths than 1999.