

# **STATS/CSE 780 - Homework assignment 1**

Pao Zhu Vivian Hsu (Student number: 400547994)

2023-09-25

## Introduction

Asbestos was a common construction material prior to the 1990s that was later found to be linked to diseases such as lung cancer and asbestosis (Government of Canada, 2023). Although it was banned in 2018, asbestos is still prevalent in old buildings and actively used in the military, nuclear, and chlor-alkali industries in Canada (Government of Canada, 2018). This report examines asbestos waste and identifies key sectors that provinces can target to further reduce the toxin from the environment.

## Methods

To begin the study, disposal data was downloaded from the Open Government Portal (Environment and Climate Change Canada, 2022) and filtered down to asbestos waste only. While the original data consisted of 17 variables (see Supplementary Materials for details), only the year, province, North American Industry Classification System (NAICS) code, and quantity of waste were important for the aim of this study. The first two digits of the NAICS code were mapped to sector names using data scraped from the Statistics Canada website (2023). This reduces the granularity of the original industry variables and allows for a high-level analysis later in the study. All quantities were converted to tonnes to standardize differences in measurement methods and the data was then aggregated by year, province, and sector to eliminate unused variables. Finally, three line graphs were created to examine country-level, province-level, and industry trends. All analyses were done using R and the last plot was created using Shiny (R Core Team, 2023a, 2023b).

## Results

*Shiny app*

## Discussion

Based on the visuals, what do you recommend ppl to do?

Where were there limitations in the study? - Certain provinces/territories do not collect data on these substances. This could mean a couple of things - there is no waste disposal places there,

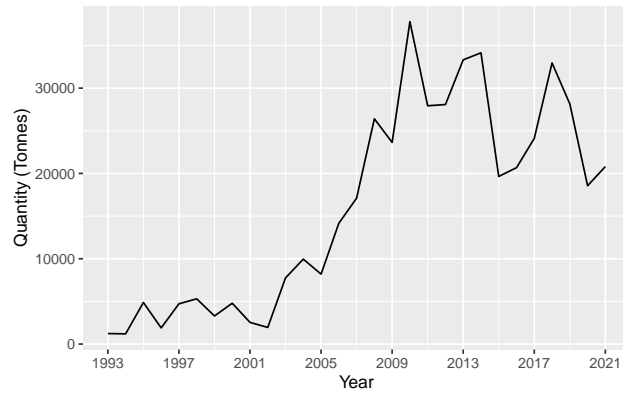


Figure 1: Asbestos waste across Canada by year

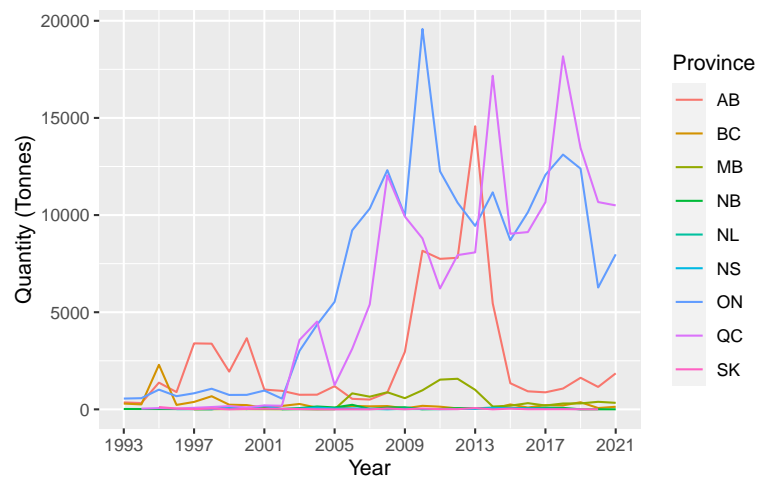


Figure 2: Asbestos waste by year and province

they do not report on waste quantities, or they do not track that particular substance. - The data is measured and estimated differently by institution. While different methods could produce different amounts of variation from the true quantity, it is not a concern this is the best that can be provided. If in doubt, can look to standardize measurement techniques.

Results can be used to guide the development of stricter laws to reduce asbestos in the environment.

- As a whole, Canada's asbestos waste has been trending

## Supplementary material

### Report Code

```
# ----- PACKAGES ----- #

library(tidyverse)
library(ggplot2)
library(stringi)

# ----- LOAD DATA ----- #

disposalDataRaw <- read_csv(file="NPRI-INRP_DisposalsEliminations_1993-present.csv",
                             locale=locale(encoding="latin1"))
naicsCodesRaw <- read_lines("https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=136")

# ----- DATA CLEANSING ----- #

# --- Step 1: Create 2-digit NAICS code lookup table --- #
# Pull 2-digit NAICS codes / code ranges and their descriptions from the source website.
naicsCodes <- data.frame(x=naicsCodesRaw) %>%
  filter(grepl("<th id=", x)) %>%
  mutate("Sector Code (2-digit NAICS Code)" = str_match(x, "CPV=\\s*(.*?)\\s*&")[,2],
         "Sector Name" = str_match(x, '"wb-inv">\\s*(.*?)\\s*</span>')[,2]) %>%
  select(`Sector Code (2-digit NAICS Code)`, `Sector Name`)
naicsCodes$`Sector Name` <- stri_replace_all_regex(naicsCodes$`Sector Name`,
                                                    pattern = c("&#40;", "&#41;", "&#44;"),
                                                    replacement = c("(", ")", ",",),
                                                    vectorize = F)
```

```

# Break code ranges down to their own rows
codeRangesOnly <- naicsCodes %>%
  filter(str_length(`Sector Code (2-digit NAICS Code)`>2) %>%
  mutate(repStart = as.integer(str_match(`Sector Code (2-digit NAICS Code)`,
                                         "([0-9]{2})[-]([0-9]{2})")[,2]),
         repEnd = as.integer(str_match(`Sector Code (2-digit NAICS Code)`,
                                         "([0-9]{2})[-]([0-9]{2})")[,3])
  ) %>%
  group_by(`Sector Name`) %>%
  group_modify(~ tibble("Sector Code (2-digit NAICS Code)" =
                        seq(.$repStart, .$repEnd))) %>%
  ungroup()

# Replace rows with code ranges with the broken down rows
naicsCodes <- rbind(naicsCodes, codeRangesOnly) %>%
  filter(str_length(`Sector Code (2-digit NAICS Code)`==2)

# --- Step 2: Filter disposal data for asbestos and join 2-digit naics codes --- #
disposalData <- disposalDataRaw %>%
  filter(grepl("asbestos",
              `Substance Name (English) / Nom de substance (Anglais)`,
              ignore.case = TRUE)) %>%
  mutate("Quantity (Tonnes)" = if_else(`Units / Unités` == "kg",
                                         `Quantity / Quantité`/1000,
                                         `Quantity / Quantité`),
         "Sector Code (2-digit NAICS Code)" = substr(`NAICS / Code_SCIAN`, 1, 2)) %>%
  left_join(naicsCodes,
            by = c("Sector Code (2-digit NAICS Code)" =
                  "Sector Code (2-digit NAICS Code)")) %>%
  group_by(`Reporting_Year / Année`,
          `PROVINCE`,

```

```

      `Sector Code (2-digit NAICS Code)`,
      `Sector Name`) %>%

summarize("Quantity (Tonnes)" = sum(`Quantity (Tonnes)`) %>%
ungroup() %>%

rename("Year" = `Reporting_Year / Année`,
       "Province" = `PROVINCE`)

# ----- SAVE DATA FOR SHINY ----- #

save(disposalData, file="shiny/disposalData.RData")

# ----- COUNTRY-LEVEL LINE GRAPH ----- #

# Preliminary data transformation
disposalData_fig2 <- disposalData %>%
  group_by(`Year`) %>%
  summarize("Quantity (Tonnes)" = sum(`Quantity (Tonnes)`)

# Plot line graph
disposalData_fig2 %>%
  ggplot(aes(x=`Year`, y=`Quantity (Tonnes)`) +
  geom_line() +
  scale_x_continuous(breaks = round(seq(min(disposalData_fig2$`Year`),
                                       max(disposalData_fig2$`Year`), by = 4),1))

# ----- PROVINCE-LEVEL LINE GRAPH ----- #

# Preliminary data transformation
disposalData_fig1 <- disposalData %>%
  group_by(`Year`, `Province`) %>%

```

```

    summarize("Quantity (Tonnes)" = sum(`Quantity (Tonnes)`)

# Plot line graph
disposalData_fig1 %>%
  ggplot(aes(x=`Year`, y=`Quantity (Tonnes)`, color=`Province`)) +
  geom_line() +
  scale_x_continuous(breaks = round(seq(min(disposalData_fig1$`Year`),
                                         max(disposalData_fig1$`Year`), by = 4),1))

```

## Shiny App Code

```

library(shiny)
library(tidyverse)
library(ggplot2)
library(stringi)

# ----- DATA PRE-PROCESSING ----- #

# Load cleaned disposal data
load("disposalData.RData")

# Drop down options
provinceOptions <- disposalData %>%
  select(`Province`) %>%
  distinct(`Province`) %>%
  pull()

# ----- APP UI ----- #
ui <- fluidPage(

```



```

# Application title
titlePanel("Yearly Asbestos Waste by Province and Sector"),

# Sidebar with a slider input for number of bins
sidebarLayout(
  sidebarPanel(
    selectInput(inputId = "province",
               label = "Province",
               choices = provinceOptions
    )
  ),

  # Show a plot of the generated distribution
  mainPanel(
    plotOutput("lineGraph")
  )
)

# ----- SERVER LOGIC ----- #
server <- function(input, output) {

  output$lineGraph <- renderPlot({

    # Filter waste data by user's province selection
    disposalData_line <- disposalData %>%
      filter(`Province` == input$province) %>%
      group_by(`Year`, `Sector Name`) %>%
      summarize("Quantity of Asbestos (Tonnes)" = sum(`Quantity (Tonnes)`)
  })
}

```

```

# Plot line graph showing waste quantity by year and sector for the selected province
disposalData_line %>%
  ggplot(aes(x=`Year`, y=`Quantity of Asbestos (Tonnes)`, color=`Sector Name`)) +
  geom_line() +
  scale_x_continuous(breaks = round(seq(min(disposalData_line$`Year`),
                                         max(disposalData_line$`Year`), by = 4),1))
  })

}

# ----- RUN APP ----- #
shinyApp(ui = ui, server = server)

```

## References

- Environment and Climate Change Canada. (2022). *Bulk data files for all years – releases, disposals, transfers and facility locations*. Government of Canada. <https://doi.org/10.18164/774eeb0c-a069-4674-a9f7-82f4adf54369>
- Government of Canada. (2018). *Prohibition of asbestos and products containing asbestos regulations*. <https://laws-lois.justice.gc.ca/eng/regulations/SOR-2018-196/page-1.html#docCont>
- Government of Canada. (2023). *Asbestos and your health*. <https://www.canada.ca/en/health-canada/services/air-quality/indoor-air-contaminants/health-risks-asbestos.html>
- R Core Team. (2023a). *Easy web apps for data science without the compromises*. R Foundation for Statistical Computing. <https://shiny.posit.co/>
- R Core Team. (2023b). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Statistics Canada. (2023). *North american industry classification system (NAICS) canada 2022 version 1.0*. <https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=1369825>