

# **Activity 14**

Shane Papineau

2025-11-15

## **Table of contents**

<b>1 Exploring the Armed Forces</b>	<b>2</b>
<b>2 Popularity of Baby Names</b>	<b>4</b>
<b>3 The Box Problem</b>	<b>6</b>
3.1 Narative Text for Box Problem . . . . .	6
<b>4 Self Reflection</b>	<b>7</b>
<b>5 Code Appendix</b>	<b>8</b>

# 1 Exploring the Armed Forces

We will use the wrangled data to explore whether sex and rank are independent of one another through a two-way frequency table.

```
#| label: freq_table
#| echo: true

#Import web ranks

webRanks <- read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html") %>%
  html_elements(css = "table") %>%
  html_table()
rawRanks <- webRanks[[1]] # Extract the data frame of ranks

#Import Forces Google Doc
rawForces <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5qbwb_E/",
  col_names = TRUE, # Turn off Column Names
  skip = 1,
  na = c("N/A*") # Tells R to treat the N/A* as missing values
)

#Individually rename some of the column names to tidy up our data
rawForces <- rawForces %>%
  rename('Rank' = ...1)

rawForces <- rawForces %>%
  rename('NavyM' = Navy)

rawForces <- rawForces %>%
  rename('NavyF' = ...6)

#We want to use Navy as a sub Group as well as Navy E9's which are Master Chiefs

#First we want to make a cleaned up version of our data that shows only Navy Pay Grades of Ma

rawForces <- rawForces %>%
  select('Rank', "NavyM", "NavyF")
#From here we can now filter out our E9's which we will rename to Master Chief as that is the
```

Table 1: Frequency Table of Master Chiefs by Sex

Rank	Sex	Count
Master Chief	Female	368
Master Chief	Male	2574

```
#Rank for that Pay Grade in the Navy

cleanNavy <- rawForces %>%
  filter(Rank == "E9") %>%
  mutate(Rank = "Master Chief")

#After renaming E9 to its Rank name in the Navy we can now look at the total Males and Females

cleanNavy <- cleanNavy %>%
  pivot_longer(cols = c(NavyF, NavyM),
               names_to = "Sex",
               values_to = "Count") %>%
  mutate(Sex = recode(Sex,
                      NavyM = "Male",
                      NavyF = "Female"))

#We want to be able to view the frequency table in our pdf

cleanNavy %>%
  kbl(
    caption = "Frequency Table of Master Chiefs by Sex",
    booktabs = TRUE,
    col.names = c("Rank", "Sex", "Count")
  ) %>%
  kable_styling(full_width = FALSE)
```

In our Frequency Table (Table1) we can see the total amount of Master Chiefs that are Males as well as Females that are Master Chiefs. In the Data we can see that there are more than 2000 Male Master Chiefs compared to Female Master Chiefs. This helps us show how sex and rank are independent from each other.

## 2 Popularity of Baby Names

To explore the popularity of some names, we've decided to focus on the popularity of the first names of the instructional team: Veronica, Alyssa, Abby, and Neil. We will focus on the overall popularity of the name, without making a distinction based on the baby's sex.

```
#| label: Baby-Names-Popularity
#| echo: true

# Loads Baby Names Data

library(dplyr)
data(BabyNames)

# Filter Out Names Regardless of Sex

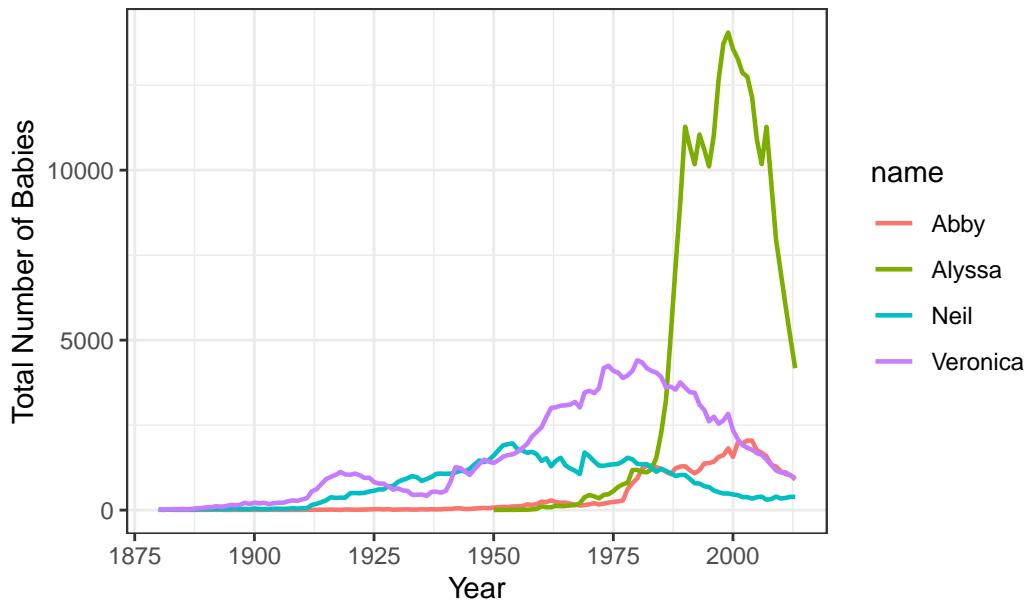
filteredNames <- BabyNames %>%
  filter(
    name %in% c("Alyssa", "Abby", "Neil", "Veronica")
  ) %>%
  group_by(name, year) %>%
  summarise(total = sum(count), .groups = "drop")

#We want to now plot our names to see how popular names change yearly

ggplot(
  data = filteredNames,
  mapping = aes(
    x = year,
    y = total,
    color = name,
    group = name
  )
) +
  geom_line(linewidth = 0.8) +
  labs(
    x = "Year",
    y = "Total Number of Babies",
    title = "Two Most Popular Names of Both Genders After 1875"
) +
```

```
theme_bw()
```

## Two Most Popular Names of Both Genders After 1875



In our visualization, we can see how the four names that we selected either increase or decrease over time starting from 1875. One of the points on our visualization is that some names take longer than others to see growth. One example is the name Alyssa which shows major popularity after 1975, we can also see that it peaks around the year 2000. We have our Years on the X-axis with total number of babies on the Y-axis. I decided make a time graph that shows 4 names with their respected colors to help demonstrate popularity.

### 3 The Box Problem

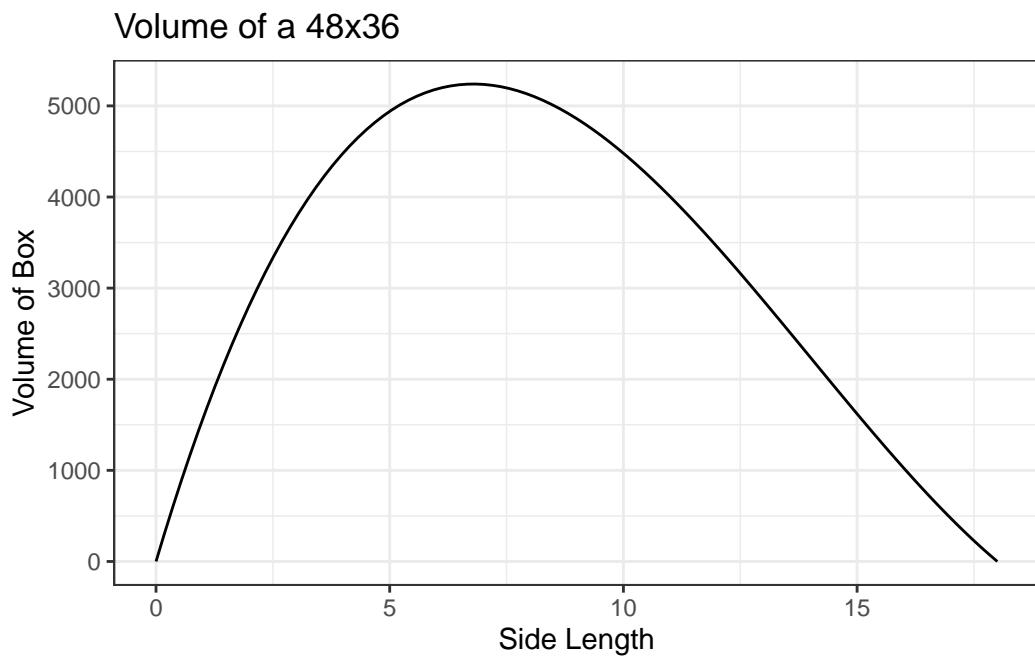


Figure 1: Figure2: Volume of Box Function

#### 3.1 Narrative Text for Box Problem

In the plot of the Volume vs Side Length, we can see that the volume peaks around 6-7 inches. As the side length increases(X-axis) we can see it start at 0 and peak around 6-7. We can also see that once it hits this peak, the line begins to decrease. This will help us find the maximum volume.

## **4 Self Reflection**

I believe that I have a strong concept on how to visualize data that I am given. I also believe that I am understanding more and more about R each day. At first it did not come easy to me and I had to put in a lot of work to feel like I understand what I am doing now. I believe that even within this assignment I can still improve many aspects of my code and understanding of R.

## 5 Code Appendix

```
#| label: freq_table
#| echo: true

#Import web ranks

webRanks <- read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html") %>%
  html_elements(css = "table") %>%
  html_table()
rawRanks <- webRanks[[1]] # Extract the data frame of ranks

#Import Forces Google Doc
rawForces <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5qbwb_E/",
  col_names = TRUE, # Turn off Column Names
  skip = 1,
  na = c("N/A*") # Tells R to treat the N/A* as missing values
)

#Individually rename some of the column names to tidy up our data
rawForces <- rawForces %>%
  rename('Rank' = ...1)

rawForces <- rawForces %>%
  rename('NavyM' = Navy)

rawForces <- rawForces %>%
  rename('NavyF' = ...6)

#We want to use Navy as a sub Group as well as Navy E9's which are Master Chiefs

#First we want to make a cleaned up version of our data that shows only Navy Pay Grades of Ma

rawForces <- rawForces %>%
  select('Rank', "NavyM", "NavyF")
#From here we can now filter out our E9's which we will rename to Master Chief as that is the
#Rank for that Pay Grade in the Navy
```

Table 2: Frequency Table of Master Chiefs by Sex

Rank	Sex	Count
Master Chief	Female	368
Master Chief	Male	2574

```

cleanNavy <- rawForces %>%
  filter(Rank == "E9") %>%
  mutate(Rank = "Master Chief")

#After renaming E9 to its Rank name in the Navy we can now look at the total Males and Females

cleanNavy <- cleanNavy %>%
  pivot_longer(cols = c(NavyF, NavyM),
               names_to = "Sex",
               values_to = "Count") %>%
  mutate(Sex = recode(Sex,
                      NavyM = "Male",
                      NavyF = "Female"))

#We want to be able to view the frequency table in our pdf

cleanNavy %>%
  kbl(
    caption = "Frequency Table of Master Chiefs by Sex",
    booktabs = TRUE,
    col.names = c("Rank", "Sex", "Count")
  ) %>%
  kable_styling(full_width = FALSE)

#| label: Baby-Names-Popularity
#| echo: true

# Loads Baby Names Data

library(dcData)
data(BabyNames)

# Filter Out Names Regardless of Sex

```

```

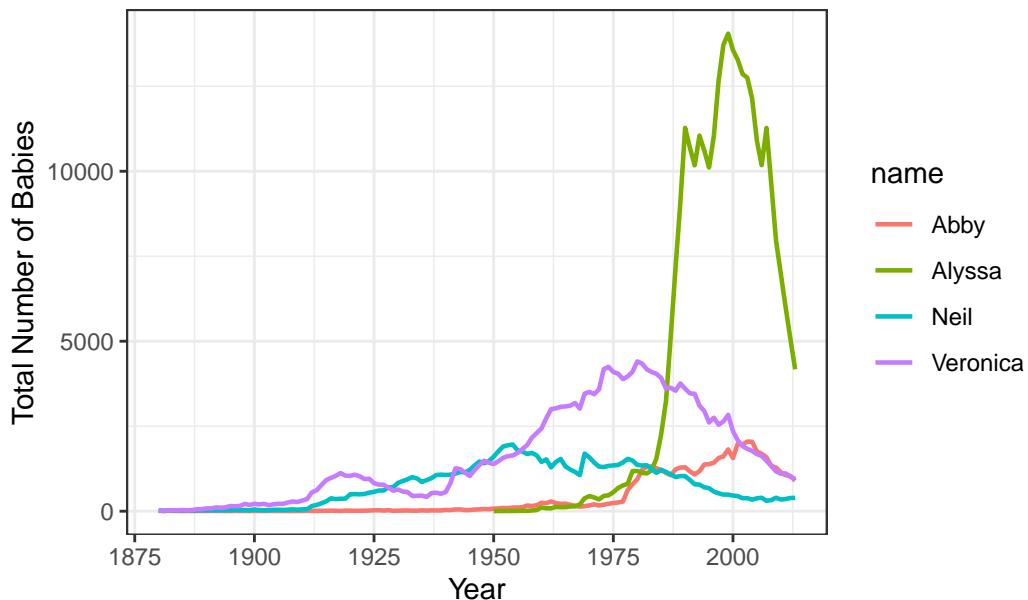
filteredNames <- BabyNames %>%
  filter(
    name %in% c("Alyssa", "Abby", "Neil", "Veronica")
  ) %>%
  group_by(name, year) %>%
  summarise(total = sum(count), .groups = "drop")

#We want to now plot our names to see how popular names change yearly

ggplot(
  data = filteredNames,
  mapping = aes(
    x = year,
    y = total,
    color = name,
    group = name
  )
) +
  geom_line(linewidth = 0.8) +
  labs(
    x = "Year",
    y = "Total Number of Babies",
    title = "Two Most Popular Names of Both Genders After 1875"
) +
  theme_bw()

```

## Two Most Popular Names of Both Genders After 1875



```
#| label: Box Problem
#| echo: false
#| fig-cap: "Figure2: Volume of Box Function"
#| fig-alt: "A curve will show us the volume when a box is cut. We wil be able to see its pe

#Here we Define our volume function
getVolume <- function(sideLength, paperLength = 48, paperWidth = 36){
  vol <- sideLength * (paperWidth - 2*sideLength) * (paperLength - 2 * sideLength)
  return(vol)
}

# we will plot now
ggplot(
  data = tibble(sideLength = c(0,18)),
  mapping = aes(x = sideLength)
) +
  stat_function(
    fun = getVolume
  ) +
  labs(
    title = "Volume of a 48x36",
    x = "Side Length" ,
    y = "Volume of Box"
```

```
) +  
theme_bw()
```

