Activity 14

Brian Schiele

2025-11-18

First we will look at the US Armed Forces data frame from Activity 8, the code used to wrangle the data can be found in the code appendix.

We will look at a frequency table that looks at enlisted soldiers in the Army and attempts to establish a connection with gender.

Table 1: Table 1: Frequency table of male Army soldiers by Rank

Rank	Army
Corporal OR Specialist	79,234 (26.44%)
First Sergeant OR Master Sergeant	9,482 (3.16%)
Private	29,767 (9.93%)
Private First Class	43,775 (14.61%)
Sergeant	54,803 (18.29%)
Sergeant First Class	30,264 (10.10%)
Sergeant Major OR Command Sergeant Major	2,865 (0.96%)
Staff Sergeant	49,502 (16.52%)
Total	299,692 (100.00%)

Fig 1. Frequency tables of Men(top) and Women(bottom) in the Army as Enlisteds

Table 2: Table 2: Frequency table of female Army soldiers by Rank

Rank	Army
Corporal OR Specialist	15,143 (27.22%)
First Sergeant OR Master Sergeant	1,472 (2.65%)
Private	5,662 (10.18%)
Private First Class	10,229 (18.39%)
Sergeant	10,954 (19.69%)
Sergeant First Class	4,410 (7.93%)

Rank	Army
Sergeant Major OR Command Sergeant Major	394 (0.71%)
Staff Sergeant	7,363 (13.24%)
Total	55,627 (100.00%)

Fig 1. Frequency tables of Men(top) and Women(bottom) in the Army as Enlisteds

The first table pictured is for men, the second is for women. The rank columns shows the rank of a selected group of soldiers, the next column shows the count of those soldiers, and the Army column shows the percentage of that rank by gender, ex. there are 29,767 male privates which is 9.93% of all male army enlisted soldiers. From this we can see that overall proportions are very similar between genders, but at higher levels such as staff sergeant or sergeant first class there is a slightly higher proportion of men. Conversely there is a higher proportion of women who are private first class.

Next we will look at a data visualization for baby names across a wide set of years. The code for the wrangling and production of this visualization can be found in the code appendix.
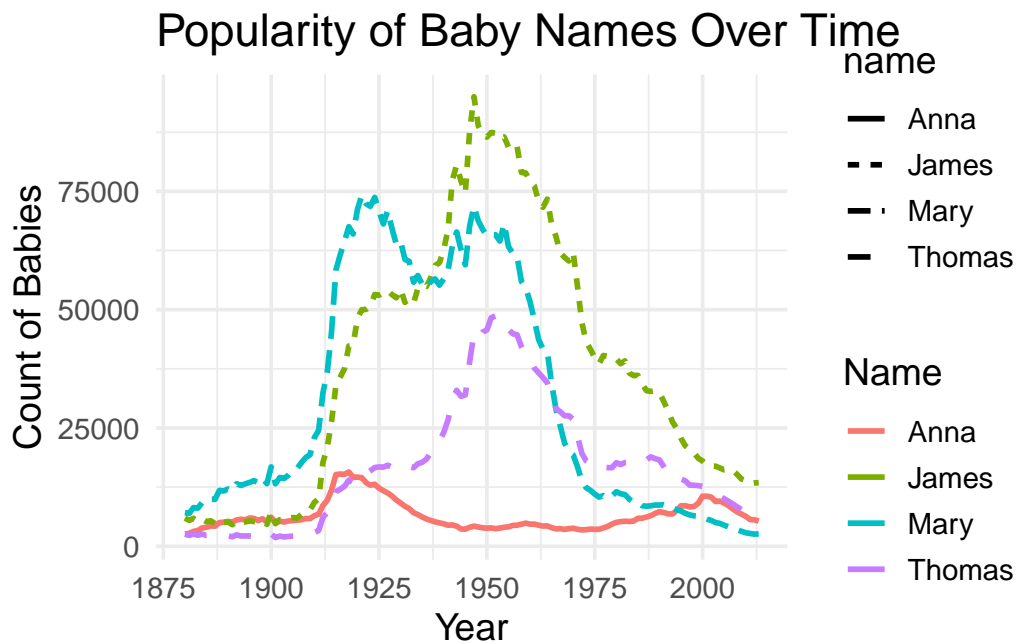

Figure 1: Fig 2. The Baby Names Mary Thomas James and Anna plotted by year and count

With this visualization, we can see the popularity of baby names from about 1875-2010. The Y-axis denotes the number of babies born that year named a certain name, and the X-axis

denotes the year. The legend on the right shows both the color and line type used for each name. With this graph we can see that the name Anna was overall the least popular, and that James and Mary competed for the top spot, makes sense, they are both biblical names. All of the names besides Anna peak from 1940-1960, during the Baby Boomer generation.

Next, we will look at a visualization for a math problem called the **Box Problem**. The code for this problem and the visualization can be found in the appendix.

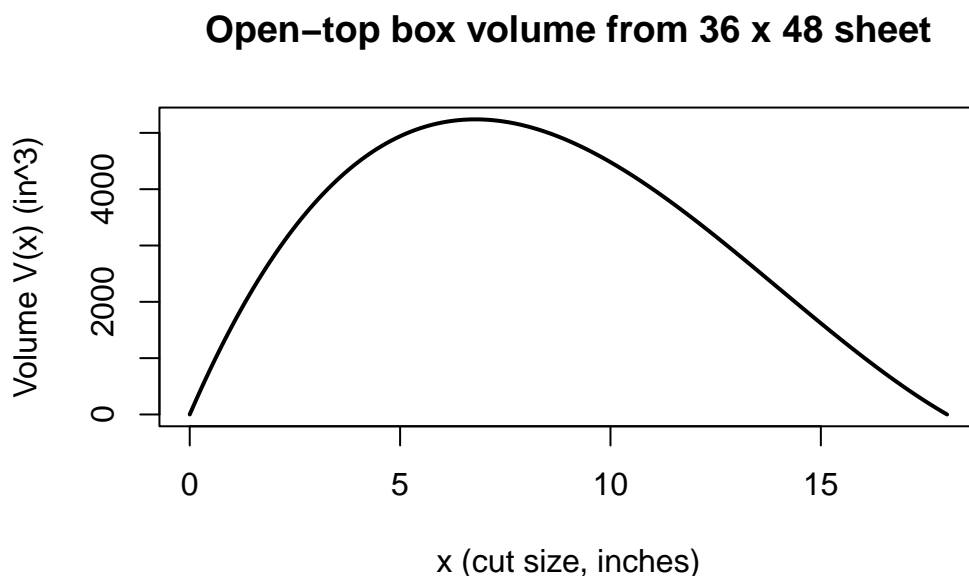**Open–top box volume from 36 x 48 sheet**

Figure 2: Fig 3. The Box Problem When the Sheet of Paper is 36 x 48

From this visualization, we see the maximum volume we can get is 5239.8 cubic inches, from a cut of length 6.8 inches. Interpreting this, we can maximize the volume of our box with a cut of about 7 inches, and we can see that that it is a much slower decline of volume with cuts greater than 6.8 as opposed to less than 6.8.

Finally, lets talk about what I have learned in this class. I love sports statistics, and I like playing around with them in my spare time. This class has opened up that world to me so much more, I had some experience with data wrangling, but topics like the join functions never made sense until this class, and now I use them in personal projects. A really cool unit to me was Chartjunk, I knew when I saw a data visualization I liked, but I didn't know WHY I liked it. That unit helped put into words why some visualizations worked for me and some didn't. Now when I make my own I try and incorporate those techniques the best I can.

Appendix A: Code

This code is for the **Data Wrangling** Section of the Armed Forces data.

```r
library(googlesheets4)
library(tidyverse)
gs4_deauth()
forcesHeaders <- read_sheet( ##set the headers we will use
  ss = 'https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5qbwb_E/e
  col_names = FALSE,
  n_max = 3
)

rawForces <- read_sheet( ##get the unwrangled data
  ss = 'https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5qbwb_E/e
  col_names = FALSE,
  skip = 3,
  n_max = 28,
  col_types = 'c'
)

branchNames <- rep( ##set the branch names we will use
  x = c("Army", "Navy", "Marine Corps", "Air Force", "Space Force", "Total"),
  each = 3
)

tempHeaders <-paste(
  c("", branchNames),
  forcesHeaders[3,],
  sep = "."
)

names(rawForces)<- tempHeaders

cleanForces <- rawForces %>% ##clean the data by renaming and getting the frequencys by pars
  rename(Pay.Grade = '.Pay Grade') %>%
  dplyr::select(!contains("Total")) %>%
  pivot_longer(cols = !Pay.Grade,
               names_to = "Branch.Sex",
               values_to = "Frequency") %>%
  separate_wider_delim(
    cols = Branch.Sex,
```

```
    delim = ".",
    names = c("Branch", "Sex")
) %>%
mutate(
  Frequency = na_if(Frequency, y = "N/A*"),
  Frequency = parse_number(Frequency)
)
```

Now here is the code for the frequency table of enlisted Army Soldiers.

Table 3: Table 1: Frequency table of male Army soldiers by Rank

Rank	Army
Corporal OR Specialist	79,234 (26.44%)
First Sergeant OR Master Sergeant	9,482 (3.16%)
Private	29,767 (9.93%)
Private First Class	43,775 (14.61%)
Sergeant	54,803 (18.29%)
Sergeant First Class	30,264 (10.10%)
Sergeant Major OR Command Sergeant Major	2,865 (0.96%)
Staff Sergeant	49,502 (16.52%)
Total	299,692 (100.00%)

Fig 1. Frequency tables of Men(top) and Women(bottom) in the Army as Enlisteds

Table 4: Table 2: Frequency table of female Army soldiers by Rank

Rank	Army
Corporal OR Specialist	15,143 (27.22%)
First Sergeant OR Master Sergeant	1,472 (2.65%)
Private	5,662 (10.18%)
Private First Class	10,229 (18.39%)
Sergeant	10,954 (19.69%)
Sergeant First Class	4,410 (7.93%)
Sergeant Major OR Command Sergeant Major	394 (0.71%)
Staff Sergeant	7,363 (13.24%)
Total	55,627 (100.00%)

Fig 1. Frequency tables of Men(top) and Women(bottom) in the Army as Enlisteds

Here is the code for wrangling and creating the visualization for BabyNames data.
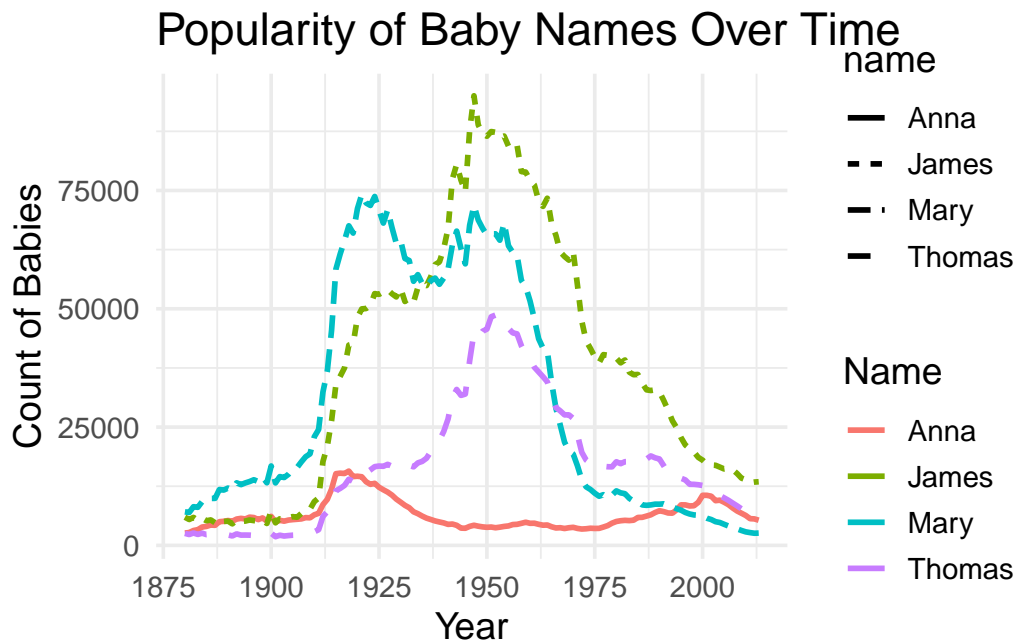
Figure 3: Fig 2. The Baby Names Mary Thomas James and Anna plotted by year and count

Here is the code used for the function and plot of the box problem.

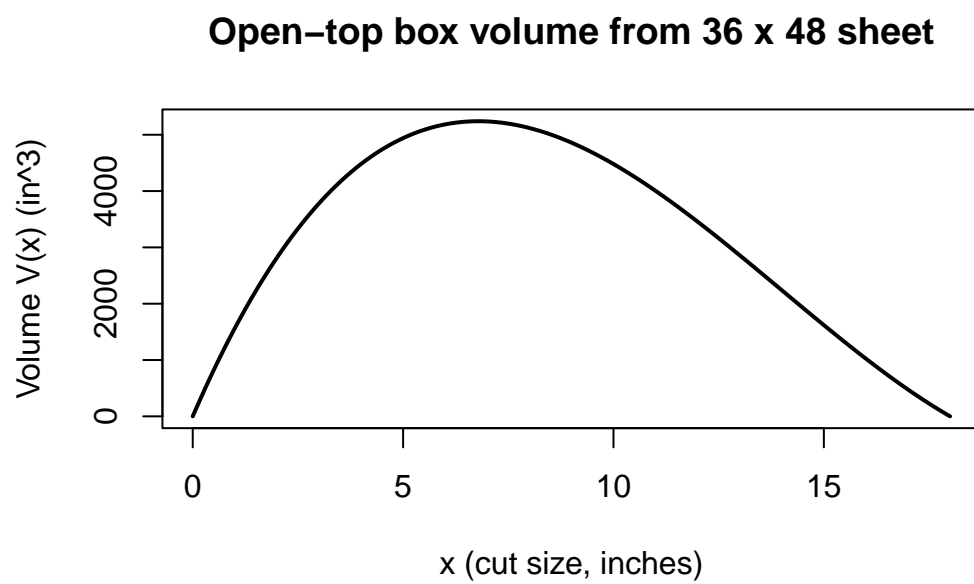**Open–top box volume from 36 x 48 sheet**

Volume V(x) (in^3)

x (cut size, inches)

Figure 4: Fig 3. The Box Problem When the Sheet of Paper is 36 x 48