

# Activity 14 — A First QMD File

## STAT 184

Mason Spikula

Tuesday, November 4, 2025

### Table of contents

0.1	Armed Forces Data Wrangling Redux (Activities #08 and #10) . . . . .	1
0.2	Popularity of Baby Names (Activity #13) . . . . .	2
0.3	Plotting a Mathematical Function (Activity #04) . . . . .	2
0.5	What You Feel You've Learned So Far . . . . .	3
0.6	Code Appendix . . . . .	4

### 0.1 Armed Forces Data Wrangling Redux (Activities #08 and #10)

This dataset shows the distribution of members of the armed forces based on their branch, sex, and pay grade. I created a graph showing the distribution of females officers, their branches, and pay grade.

Table 1: Distribution of Female Officers by Pay Grade and Branch

Pay_Grade	AirForce	Army	MarineCorps	Navy	SpaceForce	Total
O1	1,985 (4.2%)	2,400 (5.1%)	366 (0.8%)	1,766 (3.7%)	152 (0.3%)	6,669 (14.1%)
O2	2,037 (4.3%)	3,006 (6.4%)	525 (1.1%)	1,716 (3.6%)	155 (0.3%)	7,439 (15.7%)
O3	5,485 (11.6%)	6,053 (12.8%)	707 (1.5%)	4,830 (10.2%)	280 (0.6%)	17,355 (36.7%)
O4	3,440 (7.3%)	3,044 (6.4%)	338 (0.7%)	2,306 (4.9%)	209 (0.4%)	9,337 (19.7%)
O5	1,890 (4.0%)	1,531 (3.2%)	137 (0.3%)	1,151 (2.4%)	124 (0.3%)	4,833 (10.2%)

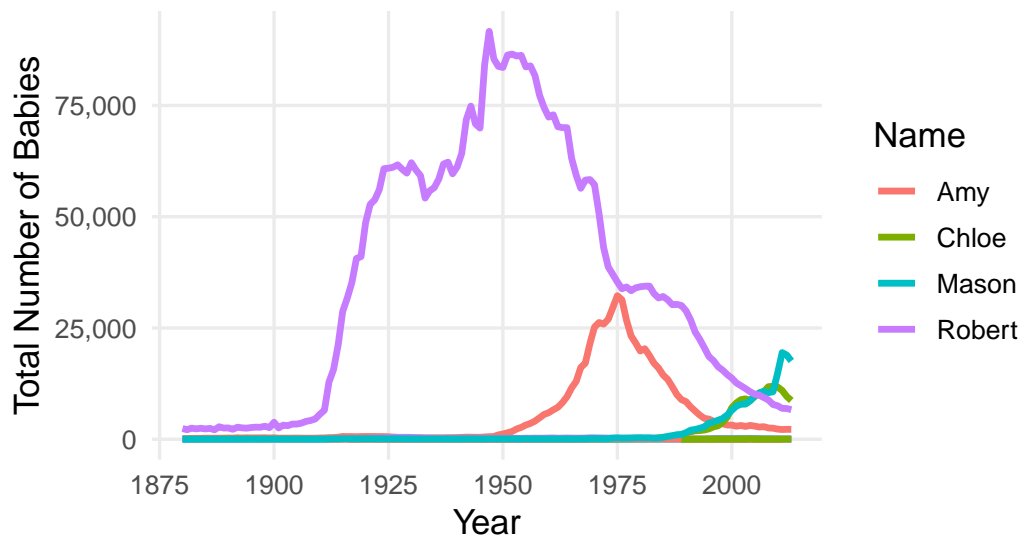
Pay_Grade	AirForce	Army	MarineCorps	Navy	SpaceForce	Total
O6	569 (1.2%)	452 (1.0%)	54 (0.1%)	452 (1.0%)	42 (0.1%)	1,569 (3.3%)
O7	18 (0.0%)	18 (0.0%)	2 (0.0%)	5 (0.0%)	2 (0.0%)	45 (0.1%)
O8	6 (0.0%)	8 (0.0%)	2 (0.0%)	6 (0.0%)	0 (0.0%)	22 (0.0%)
O9	7 (0.0%)	5 (0.0%)	1 (0.0%)	2 (0.0%)	1 (0.0%)	16 (0.0%)
Total	15,437 (32.6%)	16,517 (34.9%)	2,132 (4.5%)	12,234 (25.9%)	965 (2.0%)	47,285 (100.0%)

## 0.2 Popularity of Baby Names (Activity #13)

I wanted to see the popularity of the names of my family throughout the years. I did my dad's name, Robert, my mom's name, Amy, my sister's name, Chloe, and my name, Mason. Because I noticed that some of these names were used in both genders throughout the data, I filtered it down to just be for the respective gender.

### Popularity of Selected Baby Names

Trends in name usage by year in the United States (1880–2010)

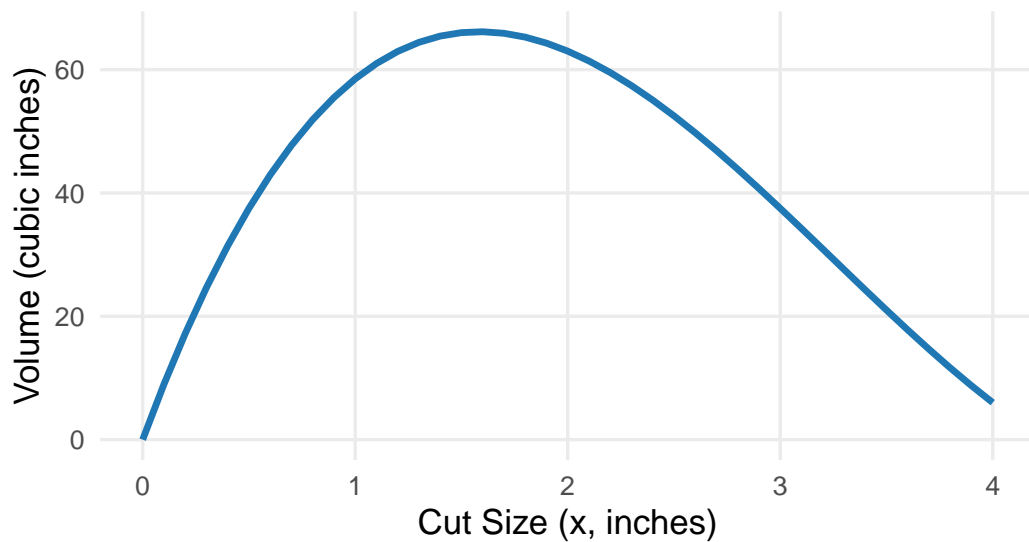


## 0.3 Plotting a Mathematical Function (Activity #04)

Create a plot of the volume of the box with respect to the side length of the cut out

## Box Volume as the Size of the Cut (x) Changes

Length = 11 in, Width = 8.5 in



### 0.4

### 0.5 What You Feel You've Learned So Far

Throughout this course so far, I feel like I've learned a lot of both technical skills and analytical skills. I think the most important technical things I've learned are data wrangling, tidying, and visualization. I've learned how to use tools like dplyr, tidyr, and ggplot2 to transform raw datasets into tidied datasets that are more suitable for analysis. My favorite part of this class was probably creating visualizations. This is because to do so, you had to tidy the data, but this part was coding but also somewhat of an art because it is important to make the graphs visually appealing and make your point clear. A good visualization can let anyone understand the data. Something cool that helped further this was using Quarto to code, show the visualizations, and explain them all in the same document. This lets users fully understand work and have it all in the same place which is very helpful. Quarto is way simpler than I thought it was in how it kind of acts like jupyter notebooks but for R.

In addition to the technical skills I've learned, I've also picked up more analytical skills. This class has taught me how to analyze visualizations better and what to look for datawise. It also taught me how viewers analyze visuals so that I could make sure that I made powerful ones that are easily understood, even by those unfamiliar with the topic.

An important lesson this class has also taught me is how to manage my time with some weeks having no assignments and some having multiple big assignments.

## 0.6 Code Appendix

```
# =====  
# Armed Forces Data Wrangling Redux (Activities #08 and #10)  
# =====  
  
# packages  
library(dplyr)  
library(tidyr)  
library(knitr)  
library(janitor)  
library(readr)  
  
ArmedForcesRaw <- read.table(  
  file = "~/Desktop/US_Armed_Forces_(6_2025) - Sheet1.csv",  
  header = FALSE,  
  sep = ",",  
  skip = 1,  
  stringsAsFactors = FALSE  
)  
  
# drop extra rows/cols (header remnants, totals, etc.)  
ArmedForcesRaw <- ArmedForcesRaw[-c(1, 2, 12, 18, 29, 30, 31), -c(4, 7, 10, 13, 16, 17, 18,  
  
# rename columns to readable names (no trailing comma)  
ArmedForcesRaw <- ArmedForcesRaw %>%  
  rename(  
    Pay_Grade = V1,  
    Army_Male = V2,  
    Army_Female = V3,  
    Navy_Male = V5,  
    Navy_Female = V6,  
    MarineCorps_Male = V8,  
    MarineCorps_Female = V9,  
    AirForce_Male = V11,  
    AirForce_Female = V12,  
    SpaceForce_Male = V14,  
    SpaceForce_Female = V15  
  )  
  
# Case = individual soldier (one row per person)  
AF_tidy_individualCase <- ArmedForcesRaw %>%
```

```

pivot_longer(
  cols      = -Pay_Grade,
  names_to  = c("Branch", "Sex"),
  names_sep = "_",
  values_to = "Count"
) %>%
mutate(Count = parse_number(Count)) %>% # converts to numeric
filter(!is.na(Count) & Count > 0) %>%   # gets rid of na's and 0;s
uncount(Count)                          # expand to one row per person

female_counts <- AF_tidy_individualCase %>%
  filter(Sex == "Female" & grepl("^0", Pay_Grade)) %>% #gets just females that are officers
  tabyl(Pay_Grade, Branch) %>%
  adorn_totals(c("row", "col"))

female_table <- female_counts %>%
  adorn_percentages("all") %>%
  adorn_pct_formatting(digits = 1) %>%
  adorn_ns(position = "front")

# Display nicely formatted female table
kable(female_table, caption = "Distribution of Female Officers by Pay Grade and Branch")

# =====
# Popularity of Baby Names (Activity #13)
# =====

#load the packages
library(tidyverse)
library(dcData)
library(scales)

#load the dataset
data("BabyNames")

#select the names
names_selected <- c("Robert", "Amy", "Chloe", "Mason")

#filter the dataset
selected_names <- BabyNames %>%
  filter(name %in% names_selected) %>%
  select(name, sex, count, year)

```

```

#summarize data
plot_df <- selected_names %>%
  group_by(name, sex, year) %>%
  summarise(count = sum(count), .groups = "drop")

#line graph
#x-axis: Year
#y-axis: Total number of babies
#color: Name
ggplot(plot_df, aes(x = year, y = count, color = name, group = interaction(name, sex))) +
  geom_line(linewidth = 1.2, na.rm = TRUE) +
  scale_y_continuous(labels = comma) +
  labs(
    title = "Popularity of Selected Baby Names",
    subtitle = "Trends in name usage by year in the United States (1880-2022)",
    x = "Year",
    y = "Total Number of Babies",
    color = "Name",
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold"),
    panel.grid.minor = element_blank()
  )

# =====
# Box Problem (Activity #04)
# =====

# Box volume function
getBoxVolume <- function(length = 11, width = 8.5, x = 1) {
  (length - 2 * x) * (width - 2 * x) * x
}

# range of cut sizes
x_vals <- seq(0, 4, by = 0.1)

# given function to find volume
volumes <- sapply(x_vals, function(x) getBoxVolume(11, 8.5, x))

# crates data frame for ggplot
box_df <- data.frame(x = x_vals, volume = volumes)

```

```

# creates visual
ggplot(box_df, aes(x = x, y = volume)) +
  geom_line(linewidth = 1.2, color = "#1f78b4") +
  labs(
    title = "Box Volume as the Size of the Cut Changes",
    subtitle = "Length = 11 in, Width = 8.5 in",
    x = "Cut Size (x, inches)",
    y = "Volume (cubic inches)"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold"),
    panel.grid.minor = element_blank()
  )

```