

Activity #14 – A First QMD File

Joel Madavaram

2025-12-14

Table of contents

#Armed Forces

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    4.0.0      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.4
```

```
-- Conflicts ----- tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag()     masks stats::lag()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
# Read CSV and FORCE all columns except Pay Grade to numeric
```

```
personnel_counts <- read_csv(
  "US_Armed_Forces_(6_2025) - Sheet1.csv",
  skip = 2,
  col_types = cols(
    `Pay Grade` = col_character(),
    .default = col_double()
  )
)
```

New names:

```
* `Male` -> `Male...2`
* `Female` -> `Female...3`
* `Total` -> `Total...4`
* `Male` -> `Male...5`
* `Female` -> `Female...6`
* `Total` -> `Total...7`
* `Male` -> `Male...8`
* `Female` -> `Female...9`
* `Total` -> `Total...10`
* `Male` -> `Male...11`
* `Female` -> `Female...12`
* `Total` -> `Total...13`
* `Male` -> `Male...14`
* `Female` -> `Female...15`
* `Total` -> `Total...16`
* `Male` -> `Male...17`
* `Female` -> `Female...18`
* `Total` -> `Total...19`
```

Warning: One or more parsing issues, call ``problems()`` on your data frame for details, e.g.:

```
dat <- vroom(...)
problems(dat)
```

```
personnel_tidy <- personnel_counts %>%
  pivot_longer(
    cols = -`Pay Grade`,
    names_to = "Category",
    values_to = "Count"
  )

head(personnel_tidy)
```

```
# A tibble: 6 x 3
  `Pay Grade` Category Count
  <chr>         <chr>   <dbl>
1 E1           Male...2    NA
2 E1           Female...3  NA
3 E1           Total...4   NA
4 E1           Male...5    NA
5 E1           Female...6  NA
```

6 E1 Total...7 NA

The data visualization summarizes the distribution of enlisted Army personnel (pay grades E1–E4) by sex. Across all four entry-level enlisted ranks, male personnel substantially outnumber female personnel. While the total number of service members increases slightly at higher enlisted pay grades, the relative proportion of males to females remains fairly consistent across ranks.

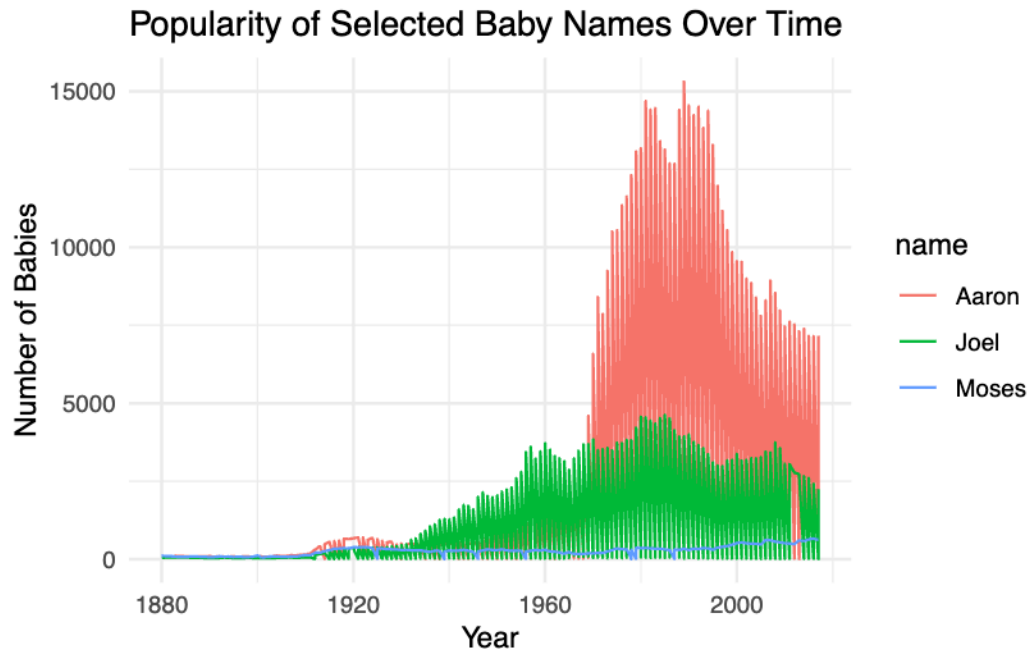
This suggests that sex and rank aren't independent within this subgroup of the U.S. Army. If sex and rank were independent, we would expect the proportion of male and female personnel to remain exactly the same across all pay grades.

#Popularity of Baby Names

```
#load
library(tidyverse)
library(babynames)

#filter data to selected baby names
selected_names <- babynames %>%
  filter(name %in% c("Moses", "Joel", "Aaron"))

# create time series plot
ggplot(selected_names, aes(x = year, y = n, color = name)) +
  geom_line() +
  labs(
    title = "Popularity of Selected Baby Names Over Time",
    x = "Year",
    y = "Number of Babies"
  ) +
  theme_minimal()
```



I picked these names because they consist of my name and my siblings. I was just curious the popularity of our names changed over time.

#Plotting a Mathematical Function activity 4

```
library(ggplot2)

# Volume function for a 36x48 inch sheet
box_volume <- function(x) {
  x * (36 - 2 * x) * (48 - 2 * x)
}

# Find maximum volume numerically
opt_result <- optimize(box_volume, interval = c(0, 18), maximum = TRUE)

max_cut <- opt_result$maximum
max_volume <- opt_result$objective

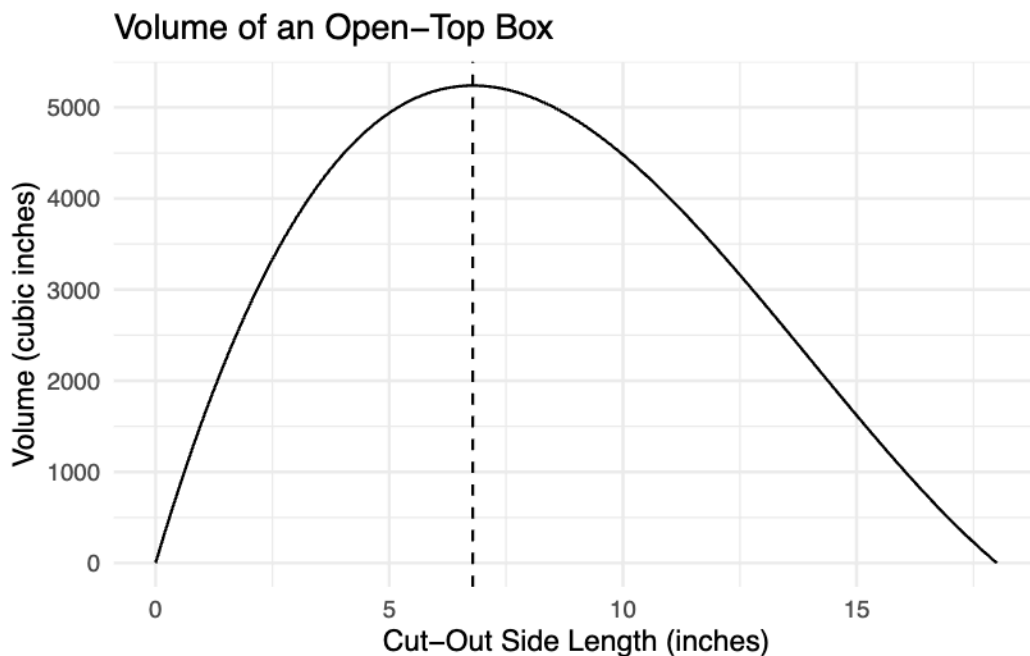
# Output results
cat("Optimal cut length:", round(max_cut, 2), "inches\n")
```

Optimal cut length: 6.79 inches

```
cat("Maximum volume:", round(max_volume, 2), "cubic inches\n")
```

Maximum volume: 5239.82 cubic inches

```
# Plot using stat_function
ggplot(data.frame(x = c(0, 18)), aes(x = x)) +
  stat_function(fun = box_volume) +
  geom_vline(xintercept = max_cut, linetype = "dashed") +
  labs(
    title = "Volume of an Open-Top Box",
    x = "Cut-Out Side Length (inches)",
    y = "Volume (cubic inches)"
  ) +
  theme_minimal()
```



The graph shows that there is an optimal cut out length for a 36in X 48in sheet of paper. And if you make the cutout smaller or larger than that optimal cutout length, you will get box with a volume that will be less than maximum. If you go to the highest cutout length possible which in this case is 18 inches since the smallest side of the paper is 36 inches, you will get a volume of 0. The opposite will give the same result as well, that is you don't make a cutout at all. The maximum volume is 5239.82 in³. To achieve the max volume, you would need to make a cutout of 6.79 inches.

#What You Feel You've Learned So Far

Throughout the course, I have learned how to work with data in a more thoughtful and reproducible way. The importance of reproducibility was one of the things that was repeated often throughout the semester.

I have also developed stronger skills in data wrangling and reshaping. The Armed Forces data analysis required me to diagnose issues in raw data, reshape it into a usable format, and verify that my results made sense. One of the problems I ran into was When R reads a .csv file it takes the first row and assumes them to be the column headers. But in the .csv file the column headers didn't appear until the third row. So I solved it by using skip=2, which made it so R started reading the file from the third row as the previous rows were the title of the data set. Its things like that that helped me see how much careful preparation is needed before any meaningful analysis can take place.

And then of course creating visualizations with the thought of ease of readability in the back of my mind for the reader. Making charts and graphs as simple to read, getting rid of unnecessary visual obstructions.