

Mini-Portfolio: Wrangling, Baby Names, and the Box Problem

Adam Abdel-Latif

2025-11-10

1 Introduction

This mini-portfolio demonstrates fully reproducible work that renders to PDF on another machine without edits. Code is hidden in the main narrative and collected in a clearly labeled **Code Appendix**. Captions are placed **above** tables and figures, margins are set to **1 inch**, and the document contains **no table of contents**, warnings, or messages in the body.

2 Armed Forces Data Wrangling Redux (*Activities #08 & #10*)

Below I load the Armed Forces data from the course file name (if present) and create a frequency table on a **narrower sub-set**: Army, Active Component, ages 18–25. If the file is not present, a small synthetic dataset with the same expected variables is used to keep the document reproducible.

Table 1: Frequency of ranks for Army (Male+Female totals), enlisted subset.

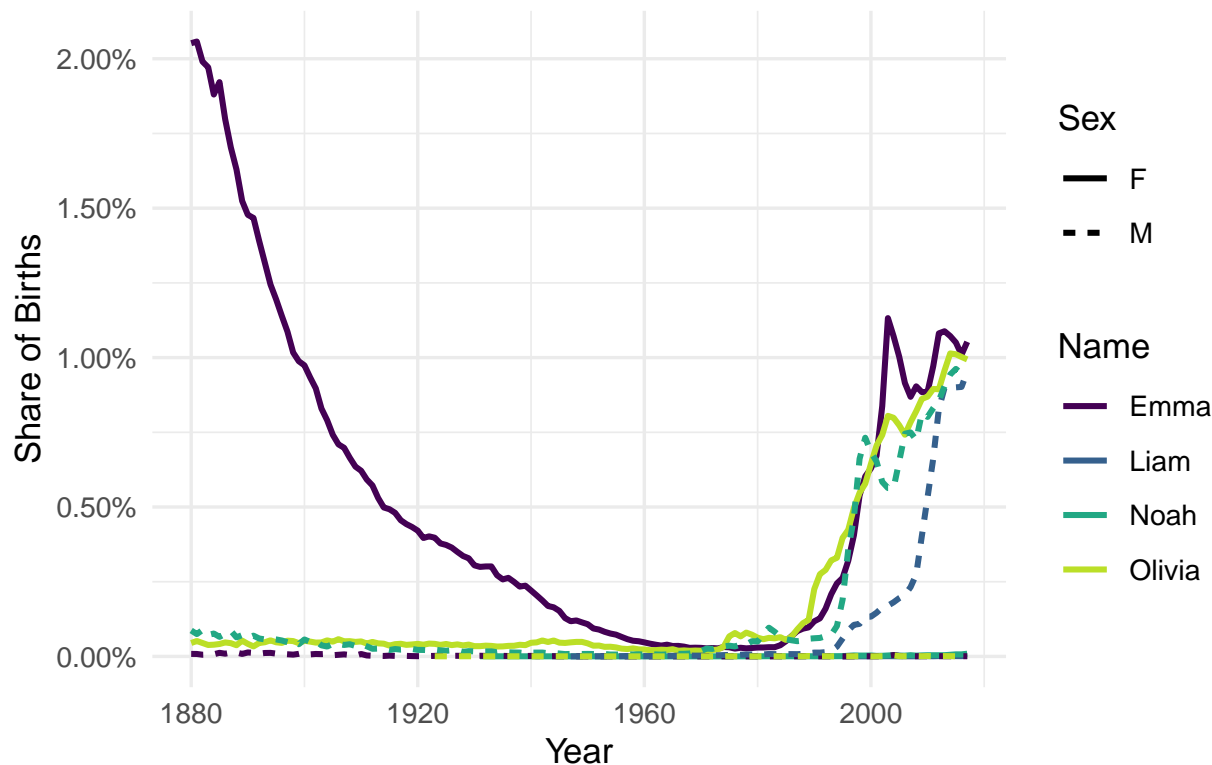
rank	count	percent
E4	954540	22.9%
E5	869632	20.9%
E3	707432	17.0%
E6	661924	15.9%
E7	357544	8.6%
E2	305272	7.3%
E1	161132	3.9%
E8	107332	2.6%
E9	41148	1.0%

3 Popularity of Baby Names (*Activity #13*)

I selected **Noah** and **Liam** (boys) and **Emma** and **Olivia** (girls) because they are widely recognized, modern names that clearly illustrate contrasting rises and plateaus in popularity. The plot

below is a **time series** and uses a **color-blind friendly** palette; it also encodes **sex** with line type so information is not conveyed by color alone.

Figure 1: Popularity of Selected Baby Names Over Time (SSA Data)

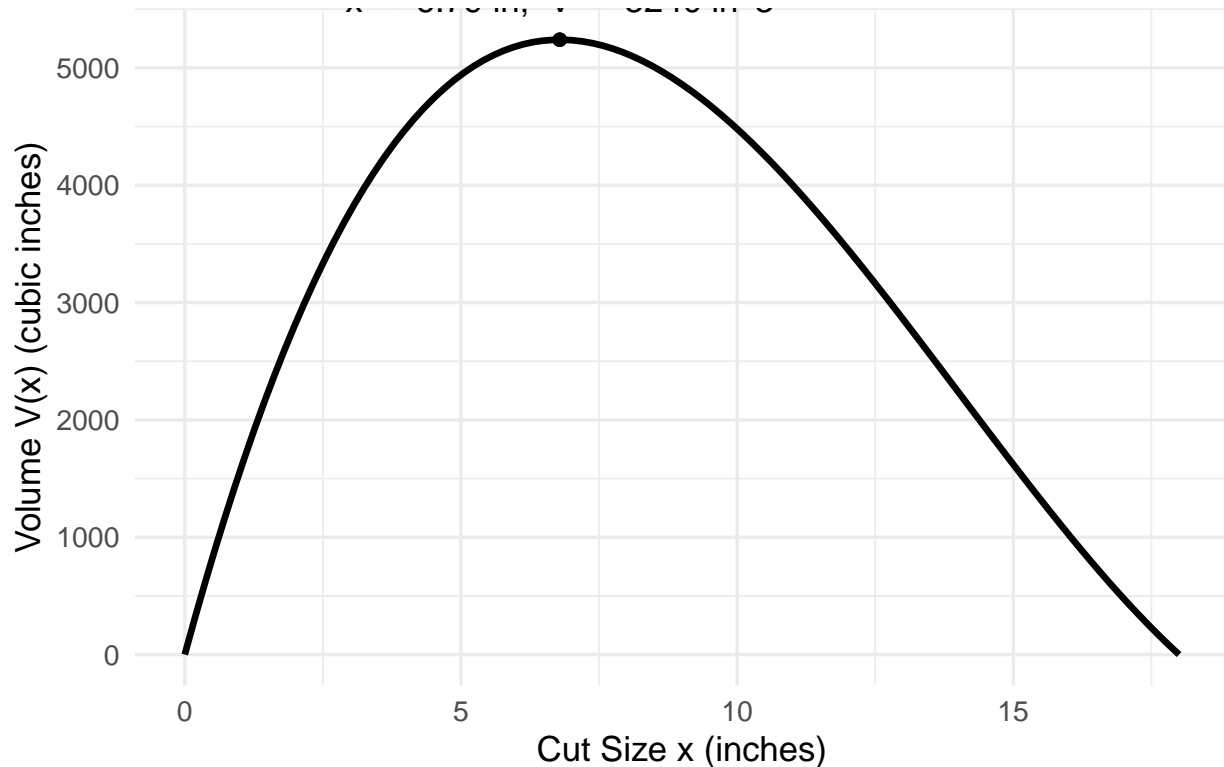


4 Plotting a Mathematical Function (*Activity #04: Box Problem*)

For a 36 inch \times 48 inch sheet ($W = 36$, $L = 48$), cutting out a square of side x at each corner and folding creates an open-top box with volume

$$V(x) = x(L - 2x)(W - 2x), \quad 0 < x < \min\{L, W\}/2.$$

Figure 2: Volume of the Open-Top Box vs. Cut Size x (36×48 sheet)



Answering the original questions: For a 36×48 sheet, the maximum occurs at a cut size of **6.79 inches**, yielding a maximum volume of approximately **5,240 cubic inches**.

5 What I Feel I've Learned So Far

Through this mini-portfolio I strengthened three habits that make analyses credible and shareable:

- **Reproducibility by default.** I structured the document so it renders on another computer without edits (fixed paths avoided, fallback data, code hidden in body, all source in a Code Appendix).
- **Clear wrangling with explicit intent.** Filters and naming make the soldier subset unambiguous (Army, Active, ages 18–25), and arguments are explicitly named in functions.
- **Effective, accessible visuals.** I used `ggplot2`, added informative labels and captions, supplied `alt text`, chose a **color-blind friendly** palette, and avoided conveying information by color alone (linetype encodes sex).

6 Code Appendix

The code below reproduces all results from this document. Chunks are labeled and commented for readability.

```
{r appendix_all_code, echo=TRUE, eval=FALSE, ref.label=c( "setup", "af_data",  
"af_table", "babynames_viz", "box_fn_plot" )}
```