Exploring Data with Visualizations

Ting Huang

2025-11-12

1 Exploring the Armed Forces

For this section, we want you to create a two-way frequency table to explore the impact of sex and rank in the US Armed Forces. In this section I focus on the Air Force airman series, that is, service members in the ranks of Airman Basic, Airman, and Airman First Class. Tables 1 and 2 show the distribution of these ranks separately for men and women. Among male airmen there are 8,537 Airman Basic (16.0%), 7,343 Airman (13.8%), and 37,324 Airman First Class (70.2%). Among female airmen there are 1,933 Airman Basic (13.5%), 2,019 Airman (14.1%), and 10,369 Airman First Class (72.4%). The conditional distributions are therefore very similar: for both sexes, about 70% of individuals are Airman First Class and only a small minority are in the two lower grades. Because the percentage breakdown across ranks is almost the same for males and females, sex and rank appear to be approximately independent in this Air Force airman subgroup—knowing a person's sex does not substantially change the probability that they hold any one of these three ranks.

Table 1: Male Activity Duty Soldiers by Rank and Branch

Rank	Air Force	Army	Marine Corps	Navy	Space Force	Total
Admiral	0 (0.00%)	0 (0.00%)	0 (0.00%)	8 (0.00%)	0 (0.00%)	8 (0.00%)
Airman	7,343 (0.70%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	7,343 (0.70%)
Airman Basic	8,537 (0.82%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	8,537 (0.82%)
Airman First Class	37,324 (3.57%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	37,324 (3.57%)
Brigadier General	99 (0.01%)	100 (0.01%)	36 (0.00%)	0 (0.00%)	11 (0.00%)	246 (0.02%)
Captain	15,715 (1.50%)	20,986 (2.01%)	5,385 (0.52%)	2,644 (0.25%)	997 (0.10%)	45,727 (4.38%)
Chief Master Sergeant	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	47 (0.00%)	47 (0.00%)
Chief Master Sergeant OR First Sergeant	1,903 (0.18%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1,903 (0.18%)
Chief Petty Officer	0 (0.00%)	0 (0.00%)	0 (0.00%)	19,046 (1.82%)	0 (0.00%)	19,046 (1.82%)
Chief Warrant Officer	0 (0.00%)	10,690 (1.02%)	1,612 (0.15%)	1,886 (0.18%)	0 (0.00%)	14,188 (1.36%)
Chief Warrent Officer	33 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	33 (0.00%)
Colonel	2,663 (0.25%)	3,161 (0.30%)	656 (0.06%)	0 (0.00%)	206 (0.02%)	6,686 (0.64%)
Commandar	0 (0.00%)	0 (0.00%)	0 (0.00%)	5,525 (0.53%)	0 (0.00%)	5,525 (0.53%)
Corporal	0 (0.00%)	0 (0.00%)	28,519 (2.73%)	0 (0.00%)	0 (0.00%)	28,519 (2.73%)
Corporal OR Specialist	0 (0.00%)	79,234 (7.58%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	79,234 (7.58%)
Ensign	0 (0.00%)	0 (0.00%)	0 (0.00%)	5,497 (0.53%)	0 (0.00%)	5,497 (0.53%)
First Lieutenant	5,045 (0.48%)	9,550 (0.91%)	3,162 (0.30%)	0 (0.00%)	437 (0.04%)	18,194 (1.74%)
First Sergeant OR Master Sergeant	0 (0.00%)	9,482 (0.91%)	3,495 (0.33%)	0 (0.00%)	0 (0.00%)	12,977 (1.24%)
General	11 (0.00%)	11 (0.00%)	3 (0.00%)	0 (0.00%)	3 (0.00%)	28 (0.00%)
Gunnery Sergeant	0 (0.00%)	0 (0.00%)	7,720 (0.74%)	0 (0.00%)	0 (0.00%)	7,720 (0.74%)
Lance Corporal	0 (0.00%)	0 (0.00%)	35,239 (3.37%)	0 (0.00%)	0 (0.00%)	35,239 (3.37%)
Lieutenant	0 (0.00%)	0 (0.00%)	0 (0.00%)	14,480 (1.39%)	0 (0.00%)	14,480 (1.39%)
Lieutenant Colonel	7,373 (0.71%)	6,939 (0.66%)	1,830 (0.18%)	0 (0.00%)	657 (0.06%)	16,799 (1.61%)
Lieutenant Commander	0 (0.00%)	0 (0.00%)	0 (0.00%)	7,983 (0.76%)	0 (0.00%)	7,983 (0.76%)
Lieutenant General	30 (0.00%)	46 (0.00%)	17 (0.00%)	0 (0.00%)	4 (0.00%)	97 (0.01%)
Lieutenant Junior Grade	0 (0.00%)	0 (0.00%)	0 (0.00%)	5,544 (0.53%)	0 (0.00%)	5,544 (0.53%)
Major	9,682 (0.93%)	12,350 (1.18%)	3,637 (0.35%)	0 (0.00%)	941 (0.09%)	26,610 (2.55%)
Major General	63 (0.01%)	80 (0.01%)	28 (0.00%)	0 (0.00%)	10 (0.00%)	181 (0.02%)
Master Chief Petty Officer OR Fleet/Command Master Chief Petty Officer	0 (0.00%)	0 (0.00%)	0 (0.00%)	2,574 (0.25%)	0 (0.00%)	2,574 (0.25%)
Master Sergeant	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	535 (0.05%)	535 (0.05%)
Master Sergeant OR First Sergeant	18,309 (1.75%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	18,309 (1.75%)
Petty Officer First Class	0 (0.00%)	0 (0.00%)	0 (0.00%)	45,833 (4.39%)	0 (0.00%)	45,833 (4.39%)
Petty Officer Second Class	0 (0.00%)	0 (0.00%)	0 (0.00%)	58,142 (5.56%)	0 (0.00%)	58,142 (5.56%)
Petty Officer Third Class	0 (0.00%)	0 (0.00%)	0 (0.00%)	33,859 (3.24%)	0 (0.00%)	33,859 (3.24%)
Private	0 (0.00%)	29,767 (2.85%)	7,849 (0.75%)	0 (0.00%)	0 (0.00%)	37,616 (3.60%)
Private First Class	0 (0.00%)	43,775 (4.19%)	15,034 (1.44%)	0 (0.00%)	0 (0.00%)	58,809 (5.63%)
Rear Admiral (Lower)	0 (0.00%)	0 (0.00%)	0 (0.00%)	101 (0.01%)	0 (0.00%)	101 (0.01%)
Rear Admiral (Upper)	0 (0.00%)	0 (0.00%)	0 (0.00%)	62 (0.01%)	0 (0.00%)	62 (0.01%)
Seaman	0 (0.00%)	0 (0.00%)	0 (0.00%)	25,436 (2.43%)	0 (0.00%)	25,436 (2.43%)
Seaman Apprentice	0 (0.00%)	0 (0.00%)	0 (0.00%)	17,504 (1.68%)	0 (0.00%)	17,504 (1.68%)
Seaman Recruit	0 (0.00%)	0 (0.00%)	0 (0.00%)	8,903 (0.85%)	0 (0.00%)	8,903 (0.85%)
Second Lieutenant	5,048 (0.48%)	7,122 (0.68%)	2,412 (0.23%)	0 (0.00%)	412 (0.04%)	14,994 (1.43%)
Senior Airman	53,185 (5.09%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	53,185 (5.09%)
Senior Chief Petty Officer	0 (0.00%)	0 (0.00%)	0 (0.00%)	6,007 (0.57%)	0 (0.00%)	6,007 (0.57%)
Senior Master Sergeant	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	112 (0.01%)	112 (0.01%)
Senior Master Sergeant OR First Sergeant	3,876 (0.37%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3,876 (0.37%)
Sergeant	0 (0.00%)	54,803 (5.24%)	22,262 (2.13%)	0 (0.00%)	859 (0.08%)	77,924 (7.46%)
Sergeant First Class	0 (0.00%)	30,264 (2.90%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	30,264 (2.90%)
Sergeant Major OR Command Sergeant Major	0 (0.00%)	2,865 (0.27%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	2,865 (0.27%)
Sergent Major OR Master Gunnery Sergeant	0 (0.00%)	0 (0.00%)	1,515 (0.14%)	0 (0.00%)	0 (0.00%)	1,515 (0.14%)
Specalist 2	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	186 (0.02%)	186 (0.02%)
Specialist 1	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	179 (0.02%)	179 (0.02%)
Specialist 3	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1,015 (0.10%)	1,015 (0.10%)
Specialist 4	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	541 (0.05%)	541 (0.05%)
Staff Sergeant	40,614 (3.89%)	49,502 (4.74%)	12,225 (1.17%)	0 (0.00%)	0 (0.00%)	102,341 (9.79%)
Technical Sergeant	31,400 (3.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	853 (0.08%)	32,253 (3.09%)
Vice Admiral	0 (0.00%)	0 (0.00%)	0 (0.00%)	32 (0.00%)	0 (0.00%)	32 (0.00%)
Warrant Officer	0 (0.00%)	3,727 (0.36%)	494 (0.05%)	44 (0.00%)	0 (0.00%)	4,265 (0.41%)
Warrent Officer	27 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	27 (0.00%)
Total	248,280 (23.76%)	374,454 (35.83%)	153,130 (14.65%)	261,110 (24.99%)	8,005 (0.77%)	1,044,979 (100.00%)

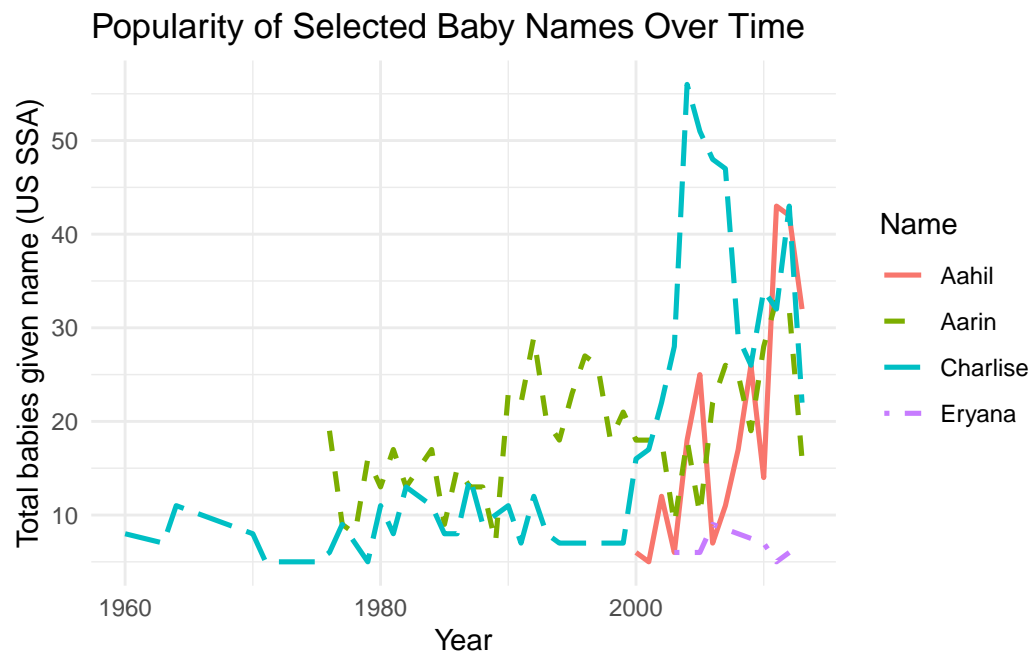Table 2: Female Activity Duty Soldiers by Rank and Branch

Rank	Air Force	Army	Marine Corps	Navy	Space Force	Total
Airman	2,019 (0.87%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	2,019 (0.87%)
Airman Basic	1,933 (0.83%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1,933 (0.83%)
Airman First Class	10,369 (4.45%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	10,369 (4.45%)
Brigadier General	18 (0.01%)	18 (0.01%)	2 (0.00%)	0 (0.00%)	0 (0.00%)	40 (0.02%)
Captain	5,485 (2.35%)	6,053 (2.60%)	707 (0.30%)	452 (0.19%)	280 (0.12%)	12,977 (5.57%)
Chief Master Sergeant	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	16 (0.01%)	16 (0.01%)
Chief Master Sergeant OR First Sergeant	523 (0.22%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	523 (0.22%)
Chief Petty Officer	0 (0.00%)	0 (0.00%)	0 (0.00%)	3,434 (1.47%)	0 (0.00%)	3,434 (1.47%)
Chief Warrant Officer	0 (0.00%)	1,218 (0.52%)	100 (0.04%)	253 (0.11%)	0 (0.00%)	1,571 (0.67%)
Chief Warrent Officer	1 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.00%)
Colonel	569 (0.24%)	452 (0.19%)	54 (0.02%)	0 (0.00%)	42 (0.02%)	1,117 (0.48%)
Commandar	0 (0.00%)	0 (0.00%)	0 (0.00%)	1,151 (0.49%)	0 (0.00%)	1,151 (0.49%)
Corporal	0 (0.00%)	0 (0.00%)	2,961 (1.27%)	0 (0.00%)	0 (0.00%)	2,961 (1.27%)
Corporal OR Specialist	0 (0.00%)	15,143 (6.49%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	15,143 (6.49%)
Ensign	0 (0.00%)	0 (0.00%)	0 (0.00%)	1,766 (0.76%)	0 (0.00%)	1,766 (0.76%)
First Lieutenant	2,037 (0.87%)	3,006 (1.29%)	525 (0.23%)	0 (0.00%)	155 (0.07%)	5,723 (2.45%)
First Sergeant OR Master Sergeant	0 (0.00%)	1,472 (0.63%)	293 (0.13%)	0 (0.00%)	0 (0.00%)	1,765 (0.76%)
Gunnery Sergeant	0 (0.00%)	0 (0.00%)	747 (0.32%)	0 (0.00%)	0 (0.00%)	747 (0.32%)
Lance Corporal	0 (0.00%)	0 (0.00%)	4,174 (1.79%)	0 (0.00%)	0 (0.00%)	4,174 (1.79%)
Lieutenant	0 (0.00%)	0 (0.00%)	0 (0.00%)	4,830 (2.07%)	0 (0.00%)	4,830 (2.07%)
Lieutenant Colonel	1,890 (0.81%)	1,531 (0.66%)	137 (0.06%)	0 (0.00%)	124 (0.05%)	3,682 (1.58%)
Lieutenant Commander	0 (0.00%)	0 (0.00%)	0 (0.00%)	2,306 (0.99%)	0 (0.00%)	2,306 (0.99%)
Lieutenant General	7 (0.00%)	5 (0.00%)	1 (0.00%)	0 (0.00%)	1 (0.00%)	14 (0.01%)
Lieutenant Junior Grade	0 (0.00%)	0 (0.00%)	0 (0.00%)	1,716 (0.74%)	0 (0.00%)	1,716 (0.74%)
Major	3,440 (1.48%)	3,044 (1.31%)	338 (0.14%)	0 (0.00%)	209 (0.09%)	7,031 (3.02%)
Major General	6 (0.00%)	8 (0.00%)	2 (0.00%)	0 (0.00%)	0 (0.00%)	16 (0.00%)
Master Chief Petty Officer OR Fleet/Command Master Chief Petty Officer	0 (0.00%)	0 (0.00%)	0 (0.00%)	368 (0.16%)	0 (0.00%)	368 (0.16%)
Master Sergeant	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	114 (0.05%)	114 (0.05%)
Master Sergeant OR First Sergeant	4,807 (2.06%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	4,807 (2.06%)
Petty Officer First Class	0 (0.00%)	0 (0.00%)	0 (0.00%)	9,950 (4.27%)	0 (0.00%)	9,950 (4.27%)
Petty Officer Second Class	0 (0.00%)	0 (0.00%)	0 (0.00%)	16,169 (6.93%)	0 (0.00%)	16,169 (6.93%)
Petty Officer Third Class	0 (0.00%)	0 (0.00%)	0 (0.00%)	9,959 (4.27%)	0 (0.00%)	9,959 (4.27%)
Private	0 (0.00%)	5,662 (2.43%)	655 (0.28%)	0 (0.00%)	0 (0.00%)	6,317 (2.71%)
Private First Class	0 (0.00%)	10,229 (4.39%)	1,684 (0.72%)	0 (0.00%)	0 (0.00%)	11,913 (5.11%)
Rear Admiral (Lower)	0 (0.00%)	0 (0.00%)	0 (0.00%)	5 (0.00%)	0 (0.00%)	5 (0.00%)
Rear Admiral (Upper)	0 (0.00%)	0 (0.00%)	0 (0.00%)	6 (0.00%)	0 (0.00%)	6 (0.00%)
Seaman	0 (0.00%)	0 (0.00%)	0 (0.00%)	9,103 (3.90%)	0 (0.00%)	9,103 (3.90%)
Seaman Apprentice	0 (0.00%)	0 (0.00%)	0 (0.00%)	5,833 (2.50%)	0 (0.00%)	5,833 (2.50%)
Seaman Recruit	0 (0.00%)	0 (0.00%)	0 (0.00%)	3,434 (1.47%)	0 (0.00%)	3,434 (1.47%)
Second Lieutenant	1,985 (0.85%)	2,400 (1.03%)	366 (0.16%)	0 (0.00%)	152 (0.07%)	4,903 (2.10%)
Senior Airman	15,055 (6.46%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	15,055 (6.46%)
Senior Chief Petty Officer	0 (0.00%)	0 (0.00%)	0 (0.00%)	850 (0.36%)	0 (0.00%)	850 (0.36%)
Senior Master Sergeant	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	25 (0.01%)	25 (0.01%)
Senior Master Sergeant OR First Sergeant	1,221 (0.52%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1,221 (0.52%)
Sergeant	0 (0.00%)	10,954 (4.70%)	2,670 (1.15%)	0 (0.00%)	173 (0.07%)	13,797 (5.92%)
Sergeant First Class	0 (0.00%)	4,410 (1.89%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	4,410 (1.89%)
Sergeant Major OR Command Sergeant Major	0 (0.00%)	394 (0.17%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	394 (0.17%)
Sergent Major OR Master Gunnery Sergeant	0 (0.00%)	0 (0.00%)	82 (0.04%)	0 (0.00%)	0 (0.00%)	82 (0.04%)
Specalist 2	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	41 (0.02%)	41 (0.02%)
Specialist 1	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	38 (0.02%)	38 (0.02%)
Specialist 3	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	194 (0.08%)	194 (0.08%)
Specialist 4	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	179 (0.08%)	179 (0.08%)
Staff Sergeant	10,762 (4.62%)	7,363 (3.16%)	1,529 (0.66%)	0 (0.00%)	0 (0.00%)	19,654 (8.43%)
Technical Sergeant	6,679 (2.86%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	147 (0.06%)	6,826 (2.93%)
Vice Admiral	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (0.00%)	0 (0.00%)	2 (0.00%)
Warrant Officer	0 (0.00%)	460 (0.20%)	44 (0.02%)	4 (0.00%)	0 (0.00%)	508 (0.22%)
Warrent Officer	1 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.00%)
Total	68,807 (29.51%)	73,822 (31.66%)	17,071 (7.32%)	71,591 (30.70%)	1,892 (0.81%)	233,183 (100.00%)

2 Popularity of Baby Names

The visualization below revisits Activity #13. Figure X shows how the popularity of four relatively uncommon baby names—Aahil, Aarin, Charlise, and Eryana—has changed over time in the U.S. Social Security Administration data. The x-axis is the year and the y-axis is the total number of babies given each name. Charlise appears the earliest and has several small waves, peaking in the early 2000s at over 50 babies in a year before declining. Aarin first appears in the mid-1970s and is never very common, but it fluctuates between about 10 and 30 births per year. Aahil is a much more recent name: it starts to appear around 2000 and rises quickly, reaching more than 40 babies per year by the late 2010s. Eryana is the rarest of the four, only showing up for a short period in the 2000s with fewer than 10 births per year.

I chose these names because they are unusual but not completely absent, which makes their trends
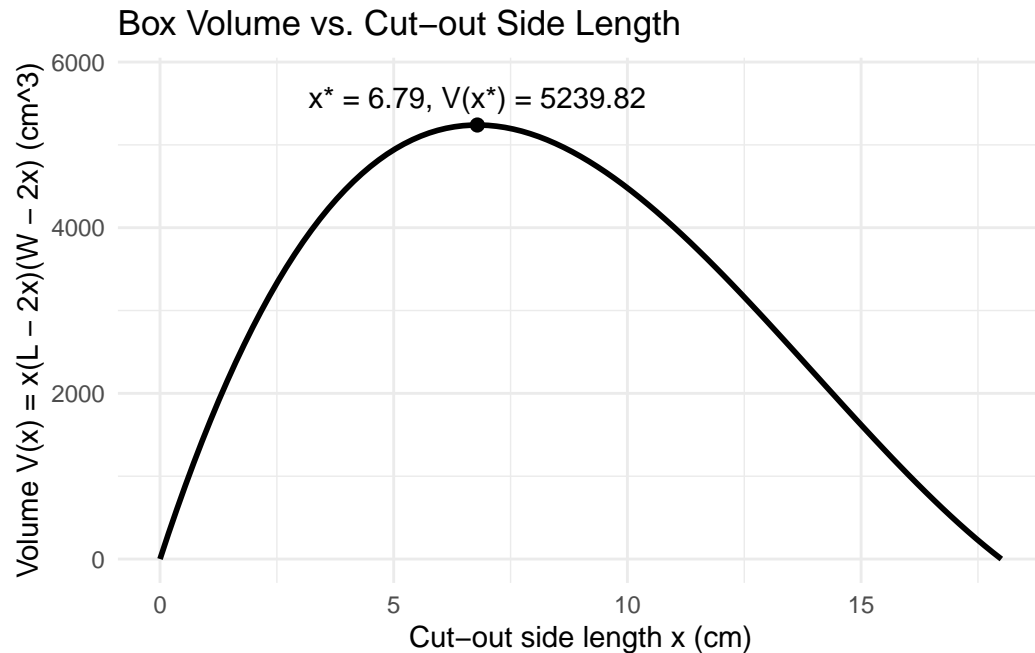
easier to see without the lines overlapping too much. They also represent different patterns of popularity: an older name with several waves (Charlise), a steady but low-level name (Aarin), a newer name that is currently growing (Aahil), and a very rare name that appears only briefly (Eryana). Together, they illustrate how baby name popularity can rise, fall, or stay small over time.

3 The Box Problem

This graph shows how the volume of an open-top box changes as we vary the side length x of the squares cut from each corner of a 36 inch by 48 inch sheet of cardboard. The horizontal axis is the cut-out length x (in inches), and the vertical axis is the resulting box volume V(x) in cubic inches. The curve starts at 0, increases as x gets larger, reaches a single peak, and then decreases back toward 0 when x becomes too large and the cardboard can no longer form a usable box.

The labeled point on the curve marks the maximum volume. From the graph, the box volume is largest when x* is about 6.79 inches, giving a maximum volume of about 5,239.82 cubic inches. So the visualization tells us that to get the biggest possible box from a 36" by 48" sheet, we should cut out squares with side length roughly 6.8 inches.

Box Volume vs. Cut–out Side Length

x* = 6.79, V(x*) = 5239.82

Volume V(x) = x(L − 2x)(W − 2x) (cm^3)

Cut–out side length x (cm)

4 Self Reflection

So far in this course I feel that I have moved beyond just "doing calculations" and have started to think like someone who uses data to communicate. In the baby names project, I learned how to choose a small set of names, make a clear time–series plot with ggplot, and then explain what each line means in words. I realized that a good visualization must tell a story – for example, showing which names are rising, which are declining, and which are always rare – not just show numbers.

The box problem helped me connect algebra, calculus, and graphs. By writing the volume as a function and plotting it, I could see why there is a single maximum and how the optimal cut size around 6.8 inches gives the largest volume. In the Armed Forces activity, I practiced building two–way tables and thinking about independence by comparing the distributions of ranks for males and females in a specific subgroup. These examples made abstract ideas like "optimization" and "independence" much more concrete.

I also learned many technical skills that are important for real data work: using R, ggplot2, and janitor, organizing a Quarto document, fixing problems with chunk labels, wide tables, and figure placement. Altogether, this course has taught me that statistics is not only about formulas; it is about combining code, graphics, and clear writing to support my conclusions.

5 Code Appendix

5.1 A. Setup & Packages

```
# install.packages("tinytex")
# below is the command for Pyhton if anyone wants to use instead of R. Run the command in term
# python -m pip install jupyter pandas numpy matplotlib

# Load Packages ----
library(googlesheets4) # For loading Armed Forces Data
library(rvest) # For scraping Armed Forces Ranks
#library(dcData) # For loading Baby Names Data
library(babynames)  # For loading Baby Names Data
library(tidyverse) # For data wrangling and visualizing
library(janitor) # For making two-way table
library(knitr) # For creating professional tables
library(kableExtra) # For styling professional tables
```

5.2 B. Armed Forces Data Wrangling & Frequency Table

```
# ---- B. Armed Forces Wrangling & Table -----------------------------------
# Tidy Armed Forces Data ---- ## Make two tied data frames for the June 2025 armed forces data
# Step 1: Load Packages ----
library(tidyverse)
library(rvest)
library(googlesheets4)
# Step 2: Scrape Rank Data ----
webRanks <- read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html") %>% html_ele
html_table()
rawRanks <- webRanks[[1]] # Extract the data frame of ranks
# Step 3: Wrangle Rank Data ----
## Enter a value in the first cell (1, 1)
rawRanks[1, 1] <- "Type"
## Extract actual column headers
rankHeaders <- rawRanks[1, ]
## Apply headers as column names
names(rawRanks) <- rankHeaders[1,]
## Remove redundant first row and last row
rawRanks <- rawRanks[-c(1, 26), ]

cleanRanks <- rawRanks %>%
  dplyr::select(!Type) %>% # Remove extra column
  pivot_longer(
    cols = !`Pay Grade`, # The improper name requires backticks
    names_to = "Branch",
    values_to = "Rank"
  ) %>%
  mutate(
    Rank = na_if(x = Rank, y = "--")
```

```r
  )
# Step 4: Load Armed Forces Data ----
gs4_deauth() # Prevents needing to sign into a Google account
forcesHeaders <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5qbwb_E/ed:
  col_names = FALSE, # Turn off Column Names
  n_max = 3 # read only the first three rows
)

rawForces <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5qbwb_E/ed:
  col_names = FALSE, # Turn off Column Names
  skip = 3, # Skip the first three rows
  n_max = 28, # Read only the next 28 rows; drops footer
  na = c("N/A*") # Tells R to treat the N/A* as missing values
)


# Step 5: Wrangle Armed Forces Data ----
## Step 5a: Create good column names ----
### Pattern is Pay Grade followed by 3 columns for each branch in the order
### Army, Navy, Marine Corp, Air Force, Space Force, and Total
branchNames <- rep( # Create three copies of each branch
  x = c("Army", "Navy", "Marine Corps", "Air Force", "Space Force", "Total"),
  each = 3
)
tempHeaders <- paste( # Combine branch with other headers
  c("", branchNames),
  forcesHeaders[3,],
  sep = "."
)

names(rawForces) <- tempHeaders

## Step 5b: Wrangle Armed Forces Data ----
cleanForces <- rawForces %>%
  rename(Pay.Grade = `.Pay Grade`) %>%
  dplyr::select(!contains("Total")) %>% # Remove total columns
  filter( # Remove total rows; see note below
    Pay.Grade != "Total Enlisted" &
   Pay.Grade != "Total Warrant Officers" &
   Pay.Grade != "Total Officers" &
   Pay.Grade != "Total"
    ) %>%
  pivot_longer( # Reshape data
    cols = !Pay.Grade,
    names_to = "Branch.Sex",
```

```r
    values_to = "Frequency"
  ) %>%
  separate_wider_delim( # Separate branches and sex
    cols = Branch.Sex,
    delim = ".",
    names = c("Branch", "Sex")
  )

# Step 6: Merge Data Frames ----
key_forcesRanks <- left_join(
  x = cleanForces,
  y = cleanRanks,
  by = join_by(Pay.Grade == `Pay Grade`, Branch == Branch)
)

# Step 7: Transform Group into Individual ----
key_individualRanks <- key_forcesRanks %>%
  filter(!is.na(Frequency)) %>% # Remove all cases with missing counts
  uncount(
    weights = Frequency
  )
```

```r
# Armed Forces Frequency Tables ----
#Goal: Make two frequency tables showing rank and branch, by sex.

# I am going to use the key_IndividualRanks data frame
# Load Packages ----
library(tidyverse)
library(janitor)
library(knitr)
library(kableExtra)

# Filter Data on Sex ----
maleSoldiers <- key_individualRanks %>%
  filter(Sex == "Male")
femaleSoldiers <- key_individualRanks %>%
  filter(Sex == "Female")

# Create Male Table ----
## Base Table ----
maleTable <- maleSoldiers %>%
  tabyl(Rank, Branch) %>% # Original Branch, Rank; flipped for improvement
  adorn_totals(where = c("row", "col")) %>%
  adorn_percentages(denominator = "all") %>%
  adorn_pct_formatting(digits = 2)

## Add Absolute Frequencies ----
```

```r
formatNs <- attr(maleTable, "core") %>%
  adorn_totals(where = c("row", "col")) %>%
  mutate(
    across(where(is.numeric), .fns = ~format(.x, big.mark = ","))
  )

## Finalize Male Table ----
maleTable <- maleTable %>%
  adorn_ns(position = "front", ns = formatNs)

# Create Female Table ----
## Base Table ----
femaleTable <- femaleSoldiers %>%
  tabyl(Rank, Branch) %>%
  adorn_totals(where = c("row", "col")) %>%
  adorn_percentages(denominator = "all") %>%
  adorn_pct_formatting(digits = 2)

## Add Absolute Frequencies ----
formatNs <- attr(femaleTable, "core") %>%
  adorn_totals(where = c("row", "col")) %>%
  mutate(
    across(where(is.numeric), .fns = ~format(.x, big.mark = ","))
  )

## Finalize Female Table ----
femaleTable <- femaleTable %>%
  adorn_ns(position = "front", ns = formatNs)

# Polish Tables ----
## Male Table ----
maleTable %>%
  kable(
    caption = "Male Activity Duty Soldiers by Rank and Branch",
    format  = "latex"
  ) %>%
  kable_styling(
    latex_options = c("hold_position", "scale_down")
  )

## Female Table ----
femaleTable %>%
  kable(
    caption = "Female Activity Duty Soldiers by Rank and Branch",
    format  = "latex"
  ) %>%
  kable_styling(
```

```
    latex_options = c("hold_position", "scale_down")
  )
```

5.3 C. Popularity of Baby Names (Activity #13)

```r
# ---- C. Baby Names Visualization --------------------------------------
# Load baby names
bn <- NULL
if (requireNamespace("dcData", quietly = TRUE)) {
  data("BabyNames", package = "dcData", envir = environment())
  bn <- get("BabyNames")
} else {
  stop("dcData package is required for BabyNames.")
}

# Standardize and verify columns
names(bn) <- gsub("\\.+", "_", tolower(names(bn)))
required <- c("name","year","count")
if (!all(required %in% names(bn))) stop("dcData::BabyNames must contain columns: name, year, c

# Select names and aggregate
instructors <- c("Aahil", "Aarin", "Charlise", "Eryana")
subsetNames <- bn %>%
  dplyr::filter(name %in% instructors) %>%
  dplyr::group_by(name, year) %>%
  dplyr::summarise(total = sum(count), .groups = "drop")

# Build linetype mapping that honors special choices when present
lt_all <- setNames(rep("solid", length(unique(subsetNames$name))), sort(unique(subsetNames$name
lt_special <- c(
  "Aahil"    = "solid",
  "Aarin"    = "dashed",
  "Charlise" = "longdash",
  "Eryana"   = "dotdash"
)
overlap <- intersect(names(lt_special), names(lt_all))
lt_all[overlap] <- lt_special[overlap]

# ggplot with color + linetype legends merged
ggplot(subsetNames, aes(x = year, y = total, color = name, linetype = name)) +
  geom_line(linewidth = 0.9) +
  scale_linetype_manual(values = lt_all) +
  labs(
    title = "Popularity of Selected Baby Names Over Time",
    x = "Year",
```

```
    y = "Total babies given name (US SSA)",
    color = "Name",
    linetype = "Name"
  ) +
  theme_minimal()
```

5.4 D. Plotting the Box Volume Function (Activity #04)

```
# ---- D. Box Problem: V(x) with stat_function ------------------------------
L <- 48; W <- 36
V <- function(x) x * (L - 2*x) * (W - 2*x)
x_max <- min(L, W) / 2

# Numeric maximization
opt <- optimize(V, interval = c(0, x_max), maximum = TRUE)
x_star <- opt$maximum
V_star <- opt$objective

# Plot
ggplot(data.frame(x = c(0, x_max)), aes(x = x)) +
  stat_function(fun = V, linewidth = 1) +
  geom_point(data = data.frame(x = x_star, y = V_star), aes(x = x, y = y), inherit.aes = FALSE
  annotate("text", x = x_star, y = V_star, vjust = -0.8,
           label = paste0("x* = ", round(x_star, 2), ", V(x*) = ", round(V_star, 2))) +
  expand_limits(y = V_star * 1.1) +
  labs(
    title = "Box Volume vs. Cut-out Side Length",
    x = "Cut-out side length x (cm)",
    y = "Volume V(x) = x(L - 2x)(W - 2x) (cm^3)"
  ) +
  theme_minimal()
```