My First QMD File

Tiago Rundle

2025-11-12

1 Armed Forces Data Wrangling Redux (Activities #08 and #10)
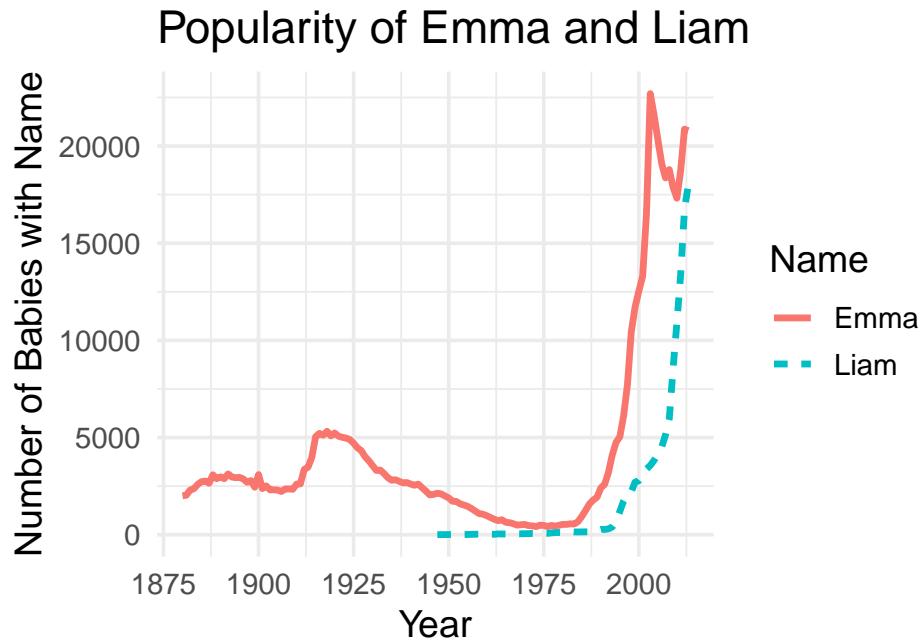
Table 1: Two-Way Frequency Table: Army Enlisted by Sex and Rank

Sex	Rank	Frequency
Female	Corporal OR Specialist	14619
Male	Corporal OR Specialist	81278
Female	First Sergeant OR Master Sergeant	1426
Male	First Sergeant OR Master Sergeant	9287
Female	Private	4705
Male	Private	26672
Female	Private First Class	8169
Male	Private First Class	38802
Female	Sergeant	11111
Male	Sergeant	55671
Female	Sergeant First Class	4322
Male	Sergeant First Class	30367
Female	Sergeant Major OR Command Sergeant Major	413
Male	Sergeant Major OR Command Sergeant Major	2908
Female	Staff Sergeant	7432
Male	Staff Sergeant	50030

The data show that sex and rank are not independent among enlisted Army personnel. While men outnumber women at every rank, the gap widens at higher ranks, with female representation decreasing steadily from lower enlisted positions like Private and Specialist to senior ranks such as Sergeant Major. This pattern suggests that women are less likely to occupy senior enlisted roles, indicating that rank advancement in the Army is associated with sex rather than being evenly distributed across genders.

2 Popularity of Baby Names (Activity #13)

Figure 1: Popularity of Emma and Liam vs Time

I chose the names that I did because I wanted a name that I knew has been around for centuries in the United States (Emma), and one that I wasn't sure about (Liam). The popularity of Emma remained relatively steady between 1875 and 1975. During this time interval, Liam was extremely uncommon and nearly nonexistent in the data. Then, towards the end of the 20th century and into the 2000's we see an explosion in the popularity of the two names. I believe that a factor contributing to this spike is the general spike in total population of the world.
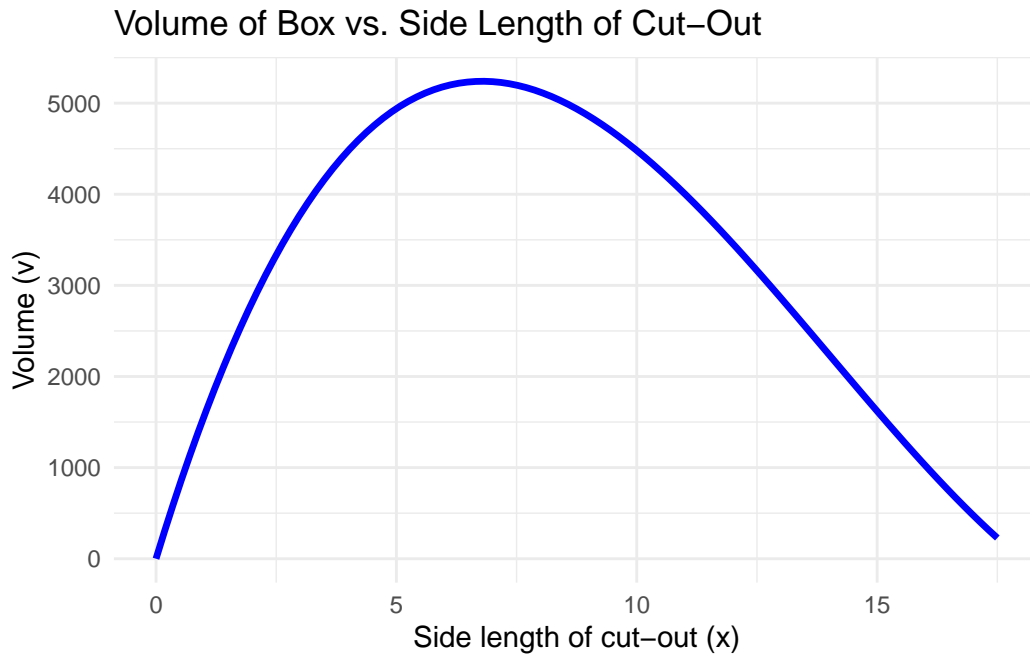
3 Plotting a Mathematical Function (Activity #04)

We want to visualize how the volume of an open-top box changes as we vary the side length of the square cutouts made from each corner of an 36" by 48" sheet of paper.

The formula for the volume is:

$V(x) = (36 - 2x)(48 - 2x)x$

Figure 2: Plot of Volume vs Cut-Out Length

The graph shows how the box's volume changes as the cut-out size increases. Volume rises quickly at first because larger cut-outs make the box deeper, reaching a clear maximum at about x = 7, where the volume is just about 5250. After that point the volume decreases because cutting out too much reduces the base dimensions. From the plot, the cut-out length that maximizes the box's volume is therefore about 7 units.

4 What I Feel I've Learned So Far

So far in STAT 184, I've learned how to work within the R and RStudio environment and have become more comfortable using R for data analysis. I've gained experience writing basic R syntax, using functions, and installing and loading packages. I've also learned how to read and write CSV files and perform basic web scraping to collect data. Using the Tidyverse package, I can now clean, organize, and summarize datasets efficiently. Additionally, I've learned how to generate descriptive statistics to better understand data, and I've started creating clear and effective visualizations using the ggplot2 package. Overall, I feel that I've developed a strong foundation in using R for statistical computing and data visualization.

5 Code Appendix

```
### Armed Forces Data Wrangling Redux

# Load Packages
```

```r
library(tidyverse)
library(rvest)
library(googlesheets4)
library(knitr)

# Scrape Rank Data
webRanks <- read_html(
  "https://neilhatfield.github.io/Stat184_PayGradeRanks.html"
  ) %>%
  html_elements(css = "table") %>%
  html_table()

rawRanks <- webRanks[[1]]

# Wrangle Rank Data
rawRanks[1, 1] <- "Type"
rankHeaders <- rawRanks[1, ]
names(rawRanks) <- rankHeaders[1,]
rawRanks <- rawRanks[-c(1, 26), ]

cleanRanks <- rawRanks %>%
  dplyr::select(!Type) %>%
  pivot_longer(
    cols = !`Pay Grade`,
    names_to = "Branch",
    values_to = "Rank"
  ) %>%
  mutate(
    Rank = na_if(x = Rank, y = "--")
  )

# Load Armed Forces Data
gs4_deauth()
forcesHeaders <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bwo/ed:
  col_names = FALSE,
  n_max = 3
)

rawForces <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bwo/ed:
  col_names = FALSE,
  skip = 3,
  n_max = 28,
  col_types = "c"
)
```

4

```r
# Wrangle Armed Forces Data
branchNames <- rep(
  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 %>%
  rename(Pay.Grade = `.Pay Grade`) %>%
  dplyr::select(!contains("Total")) %>%
  filter(Pay.Grade != "Total Enlisted" &
           Pay.Grade != "Total Warrant Officers" &
           Pay.Grade != "Total Officers" &
           Pay.Grade != "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)
  )

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

# Transform Group into Individual
key_individualRanks <- key_forcesRanks %>%
  filter(!is.na(Frequency)) %>%
  uncount(
    weights = Frequency
```

```r
  )

# Focus on Army Enlisted personnel
army_enlisted <- key_individualRanks %>%
  filter(Branch == "Army", str_detect(Pay.Grade, "^E"))

# Create two-way frequency table of Sex vs Rank
table_df <- as.data.frame(table(army_enlisted$Sex, army_enlisted$Rank)) %>%
  rename(
    Sex = Var1,
    Rank = Var2,
    Frequency = Freq
  )

kable(table_df, caption="Two-Way Frequency Table: Army Enlisted by Sex and Rank")

### Popularity of Baby Names

# Load Packages
library(tidyverse)
library(dcData)
library(ggplot2)

# Data Wrangling
selected_names <- BabyNames %>%
  filter(name %in% c("Emma", "Liam")) %>%
  filter((name == "Emma" & sex == "F") | (name == "Liam" & sex == "M"))

# Data Visualization
ggplot(selected_names, aes(x=year,
                           y=count,
                           color=name,
                           linetype=name,
                           group=interaction(name, sex))) +
  geom_line(size = 1.2) +
  labs(
    title = "Popularity of Emma and Liam",
    x = "Year",
    y = "Number of Babies with Name",
    color = "Name",
    linetype = "Name"
  ) +
  theme_minimal(base_size = 14)

### Plotting a Mathematical Function

library(ggplot2)
```

```r
# Function Definition
volume = function(x) {
    v = (36 - 2*x)*(48 - 2*x)*x
    return(v)
}

x_range <- data.frame(x = c(0, 17.5))

# Function Visualization Plot
ggplot(x_range, aes(x = x)) +
  stat_function(fun = volume, color = "blue", size = 1.2) +
  labs(
    title = "Volume of Box vs. Side Length of Cut-Out",
    x = "Side length of cut-out (x)",
    y = "Volume (v)"
  ) +
  theme_minimal()
```