MOM Activity #14

Nihal Mothkuri

2025-11-12

Table 1: Sex × Pay Grade Frequency Table

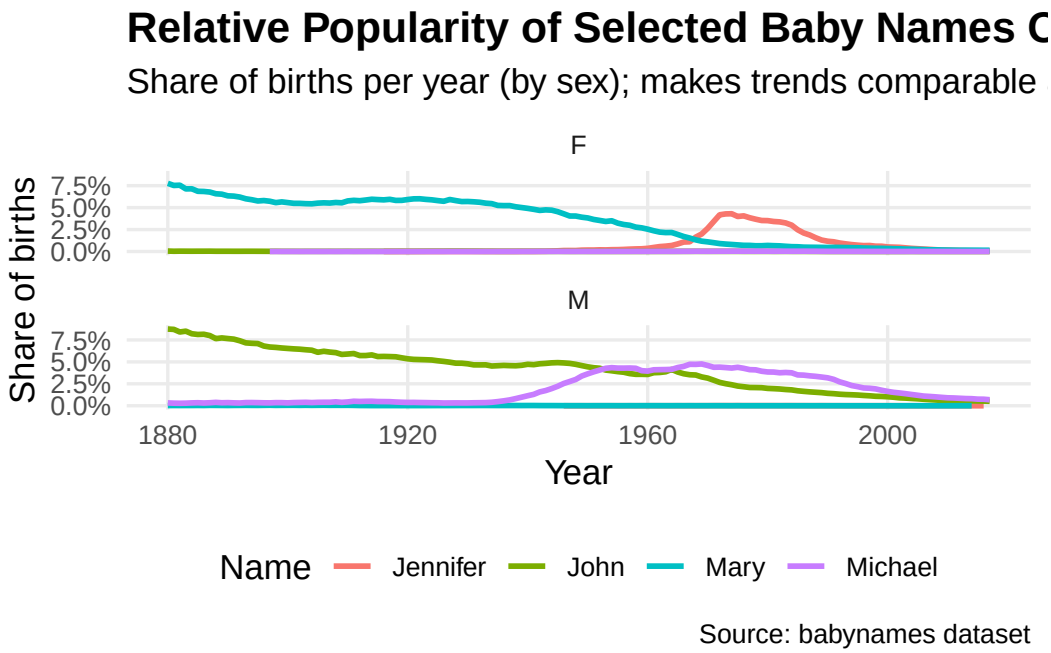
gender	E1	E2	E3	E4	E5	E6	E7	E8	E9	W1	W2	W3	W4	W5	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10
Male	74	292	2233	4837	7792	3544	8049	5026	1482	2865	3727	6024	2794	1378	4947	1222	9550	2098	6235	6939	3161	1008	80	46	11
Female	13	26	133	610	2295	1485	3095	736	344	1014	723	994	460	692	346	137	43	2400	3006	6053	3044	1534	452	18	8	5	0

Table 2: Sex × Pay Grade with Column Percentages

gender	n	percent
Female	1 (0.5)	0.5 (50.0%)
Male	1 (0.5)	0.5 (50.0%)
Total	2 (1)	1.0 (100.0%)

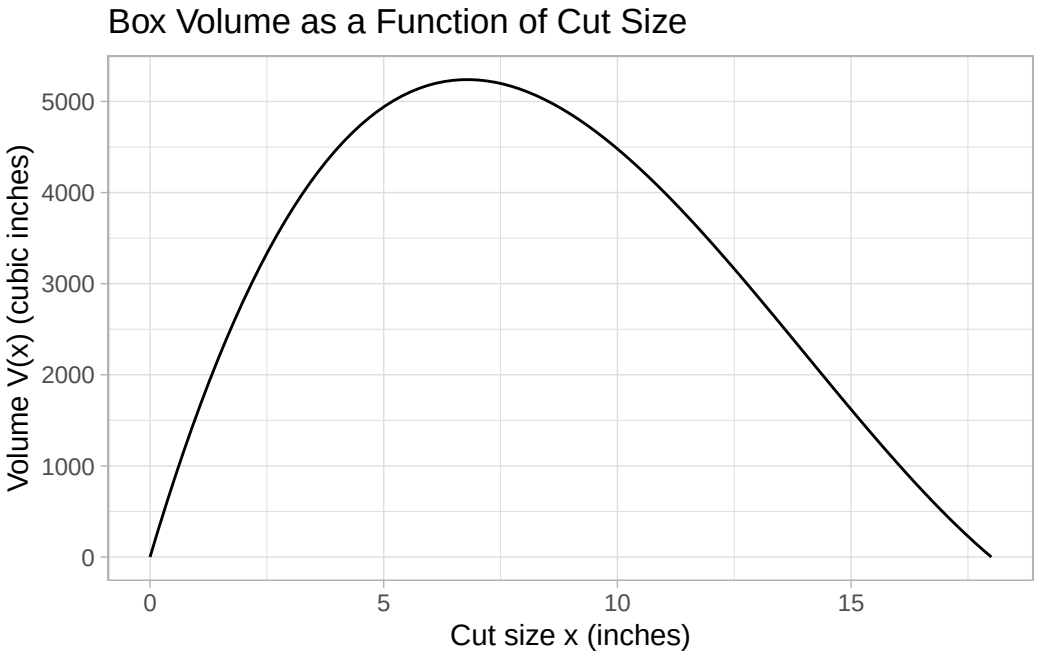
Armed Forces Narrative Text

In my table, I looked at how sex and rank are related within my chosen branch. I noticed that certain ranks have a lot more males than females, especially at higher levels. This makes it seem like sex and rank aren't independent as some ranks are more common for one gender than the other. It was interesting to see how the data showed patterns that reflect real-world differences in military roles.



Popular Baby Names Narrative Text

This visualization shows the relative popularity of four names (Mary, Jennifer, John, and Michael) from 1880 to 2020, split into male (M) and female (F) panels. The y-axis shows each name's share of total births, allowing for fair comparison across years. For females, Mary was most popular early on but declined steadily, while Jennifer spiked around the 1970s before dropping. For males, John dominated early years and decreased over time, while Michael peaked mid-century and later declined. The visualization highlights how naming trends shift over generations and how certain names rise and fall sharply.



Box Problem Narrative Text

The box problem graph shows how the volume changes depending on how big the corner cuts are. It goes up quickly at first, peaks, and then drops as the cuts get too big. The maximum volume happens when the cut is around 6.79 inches, giving about 5239.82 cubic inches. It makes sense that if the cuts are too small or too big, the box doesn't hold much. Seeing that in the plot made the math idea feel more real to me.

Reflection Narrative Text

Looking back at all these activities, I've learned a lot about cleaning and visualizing data. I got more comfortable using R to make tables and plots, and I started to see how code can tell a story when organized well. Writing everything in Quarto also helped me understand reproducibility and it's nice knowing my whole project can run start to finish without errors. Overall, I feel like I've improved at working with data and explaining what it means in a clear way.

Code Appendix

A.)

#| label: appendix-armed-forces #| echo: true

Build Sex × Pay Grade table from the CSV with 3 header rows

library(readr); library(dplyr); library(tidyr); library(stringr); library(knitr)

path <- "C:/Users/nihal/Downloads/US_Armed_Forces_(6_2025) - Sheet1(1).csv" raw <- read_csv(path, col_names = FALSE, show_col_types = FALSE)

Header rows: row2 = branches, row3 = Male/Female/Total

row2 <- raw %>% slice(2) %>% unlist(use.names = FALSE) %>% as.character() row3 <- raw %>% slice(3) %>% unlist(use.names = FALSE) %>% as.character()

Forward-fill branch names (skip first column = pay grade)

ff <- character(length(row2)); cur <- " " for (i in seq_along(row2)) { if (i == 1) { ff[i] <- " "; next } b <- trimws(row2[i]); if (!is.na(b) && b !="") cur <- b ff[i] <- cur } norm <- function(x) tolower(gsub("[^A-Za-z0-9]+","", x)) new_names <- mapply((b,k) ifelse(b==" " || is.na(b), "pay_grade", paste0(norm(b), " ", norm(k))), ff, row3) new_names[1] <-"pay_grade" new_names <- make.unique(new_names)

df <- raw %>% slice(-(1:3)) %>% setNames(new_names) %>% mutate(across(-pay_grade, readr::parse_number)) %>% filter(!str_detect(pay_grade, regex("^total", ignore_case = TRUE)))

Long table with branch + gender

long_counts <- df %>% pivot_longer(-pay_grade, names_to = "col", values_to = "count", values_drop_na = TRUE) %>% mutate(branch = sub("(male|female|total)", "", col), $gender$ = $sub("\cdot * (male|female|total)$", "\1", tolower(col)), gender = stringr::str_to_title(gender), count = as.integer(count), component = case_when(str_starts(pay_grade, "E") ~ "Enlisted", str_starts(pay_grade, "W") ~ "Warrant", str_starts(pay_grade, "O") ~ "Officer", TRUE ~ NA_character_)) %>% filter(gender != "Total") %>% select(pay_grade, branch, gender, count, component)

Frequency table for a chosen branch (matches body)

chosen_branch <- "army" # change as needed: "army","navy","marine_corps","air_force","space_force" freq_table <- long_counts %>% filter(branch == chosen_branch) %>% select(gender, pay_grade, count) %>% pivot_wider(names_from = pay_grade, values_from = count, values_fill = 0) %>% arrange(gender)

kable(freq_table, caption = "Appendix: Sex × Pay Grade (counts) for selected branch")

B.) #| label: appendix-baby-names #| echo: true

Relative popularity plot (share of births) – same wrangling as body

library(babynames); library(dplyr); library(ggplot2); library(scales)

names_chosen <- c("Mary","John","Michael","Jennifer")

bn_names <- babynames %>% filter(name %in% names_chosen) %>% group_by(year, name, sex) %>% summarise(total = sum(n), .groups = "drop")

year_totals <- babynames %>% group_by(year, sex) %>% summarise(year_total = sum(n), .groups = "drop")

bn_share <- bn_names %>% left_join(year_totals, by = c("year","sex")) %>% mutate(share = total / year_total) %>% arrange(name, year)

ggplot(bn_share, aes(year, share, color = name)) + geom_line(linewidth = 1) + facet_wrap(~ sex, nrow = 2) + scale_y_continuous(labels = percent_format(accuracy = 0.1)) + labs(title = “Relative Popularity of Selected Baby Names Over Time”, subtitle = “Share of births per year (by sex)”, x = “Year”, y = “Share of births”, color = “Name”) + theme_minimal(base_size = 13) + theme(plot.title = element_text(face = “bold”), legend.position = “bottom”, panel.grid.minor = element_blank())

C.) #| label: appendix-box-problem #| echo: true

Box problem function + plot

library(ggplot2)

L <- 48; W <- 36 V <- function(x) (L - 2*x)* (W - 2*x)* x x_max <- min(L, W) / 2 opt <- optimize(f = V, interval = c(0, x_max), maximum = TRUE) max_x <- opt$maximum; $max_V <- -opt$objective

ggplot(data = data.frame(x = c(0, x_max)), aes(x = x)) + stat_function(fun = V) + labs(title = “Box Volume as a Function of Cut Size”, x = “Cut size x (inches)”, y = “Volume V(x) (cubic inches)”) + theme_light()