

Activity14

JennyT

2025-11-14

Table of contents

Armed Forces Data	2
Popular Baby Names Project	3
Plotting a Mathematical Function	4
What I've Learned	4
Code Appendix	5

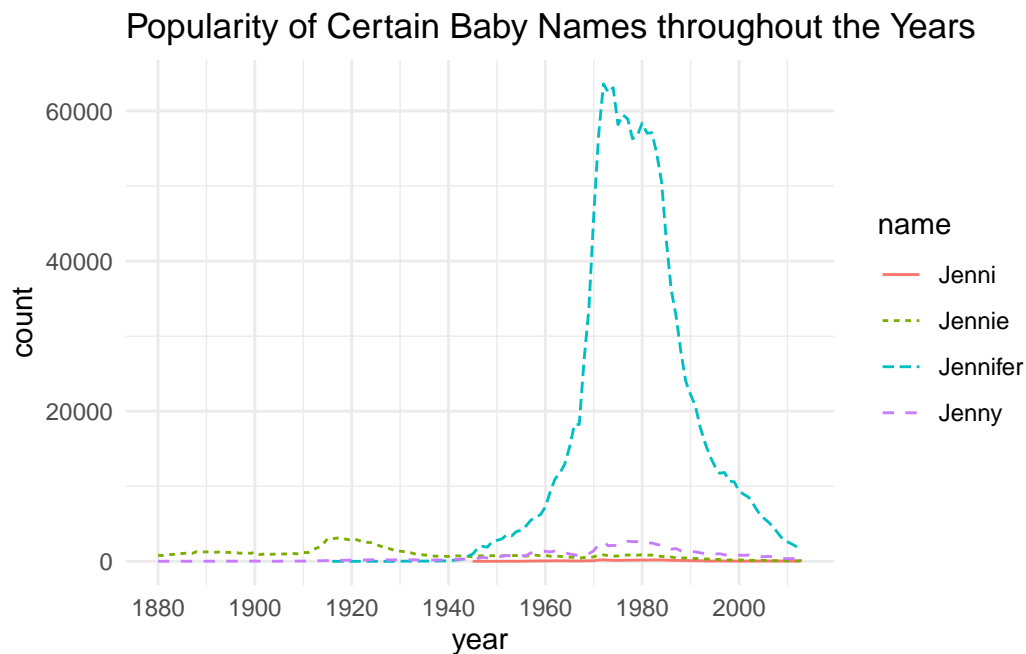
Table 1: Two Way Frequency Data for Armed Forces

Pay Grade	Male	Female	Total
E1	7429	1326	8755
E2	22338	4336	26674
E3	43775	10229	54004
E4	79234	15143	94377
E5	54803	10954	65757
E6	49502	7363	56865
E7	30264	4410	34674
E8	9482	1472	10954
E9	2865	394	3259
Total	299692	55627	355319

Armed Forces Data

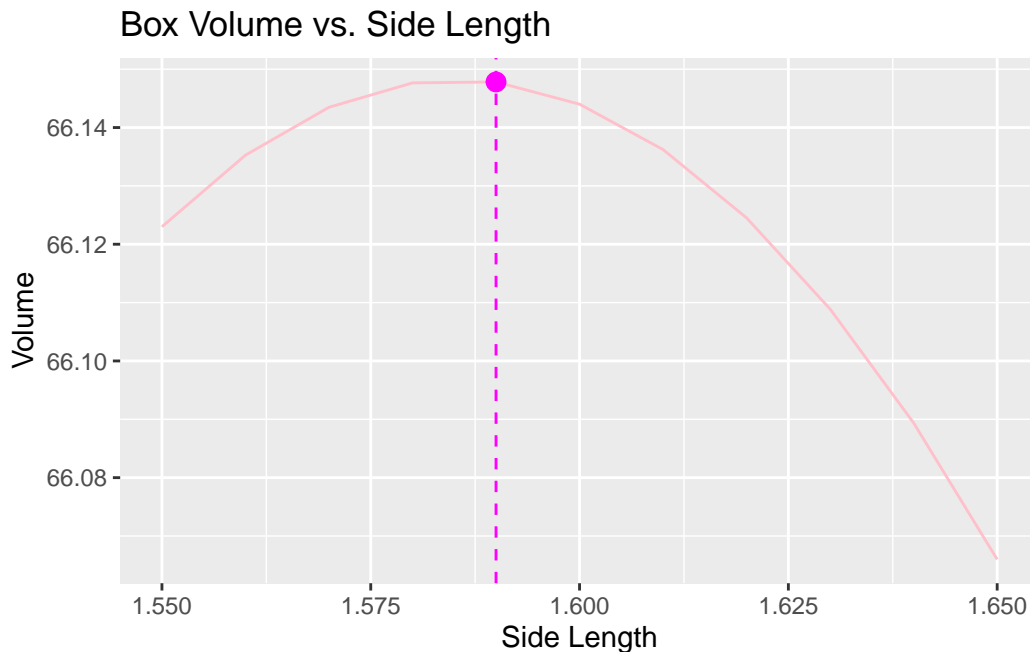
The table above showcases a two-frequency table of Enlisted Ranks and Sex for the Army branch of the US Armed forces. One can compare both horizontally and vertically. For example, there's much more male officers than female officers, by more than 5 times; as for the distribution of ranks, for both sexes, it seems like there's a normal distribution, with the peak being around E4. The same case would be true for the sexes individually as well, where most people falls in E4. Out of all ranks, E9 (the highest) has the least personnel, which is not surprising since it's the highest rank.

Popular Baby Names Project



The graph displays the frequency of four feminine baby names: Jenni, Jennie, Jennifer, and Jenny, from 1875 to 2020. The names were chosen for that they are variants to my name (Jenny). The y-axis represents the count, ranging from 0 to 60,000, and the x-axis represents the years from 1875 to 2020. The graph includes four lines: a thin red line for Jenni, a thin green dotted line for Jennie, a bold cyan dashed line for Jennifer, and a purple dashed line for Jenny. The line for Jennifer shows a significant spike between 1950 and 1980, peaking around 1970, with counts reaching over 60,000. The other names remain relatively low and stable over the years.

Plotting a Mathematical Function



The line graph above, titled “Box Volume vs. Side Length”, which compares different side lengths and the volume they correspond to. The x-axis is labeled “Side Length” and ranges from 1.550 to 1.650, while the y-axis is labeled “Volume” and ranges from 66.08 to 66.14. The graph depicts a curve that starts at a lower volume, increases to a peak, and then gradually decreases. A magenta vertical dashed line intersects the peak of the curve at a side length around 1.590, with a corresponding max volume of 66.15.

What I’ve Learned

I think this activity is great for learning about the practical usages of Quarto Markdown documents, I had experiences with HTML so the syntax wasn’t unfamiliar, but I did learn about stuff like setting the appropriate YAML headers and how to code chunk output options. One of my graphs was originally in HTML format, so I also had to learn how to convert that to be compatible with PDF.

Code Appendix

```
library(tidyr)
library(dplyr)
library(tidyverse)
library(kableExtra)
library(janitor)

#inmort dataset
df <- read.csv("C:/school/stat184/US_Armed_Forces_(6_2025) - Sheet1.csv", header = TRUE)

#organize the dataset so the columns and rows with "total" are removed and distribute the br
df <- df[!grepl("Total", df$Active.Duty.Personnel.by.Service.Branch..Sex..and.Pay.Grade), ] %
  select(-c(4,7,10,13,16,17,18,19))
df[1,3] <- "Army"
df[1,5] <- "Navy"
df[1,7] <- "Marine Corps"
df[1,9] <- "Air Force"
df[1,11] <- "Space Force"

#merge the first two rows
merged_row <- sapply(df[1:2, ], function(x) paste(x, collapse = "_"))
df_new <- rbind(merged_row, df[3:nrow(df), ])

#set the first row as column headings
colnames(df_new) <- df_new[1, ]
df_new <- df_new[-1, ]

#transform the dataset so that it's in the order of pay grade, branch, sex, and population
#filter the data so it only shows those in the Army branch and Enlisted rank
#transpose the dataframe with pivot_wider and accumulate the values for the totals
df_cleaned <- df_new %>%
pivot_longer(
  cols = ~`_Pay Grade`,
  names_to = c("Branch", "Sex"),
  names_sep = "_"
) %>%
filter(startsWith(`_Pay Grade`, "E") & Branch == "Army") %>%
select(-Branch) %>%
pivot_wider(names_from = "Sex", values_from = "value") %>%
mutate(across(c("Male", "Female"), ~ as.numeric(gsub(",", "", .)))) %>%
adorn_totals(c("row", "col"))
```

```

table <- df_cleaned %>%
  kable(format = "latex",
        caption = "Two Way Frequency Data for Armed Forces",
        booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped"), full_width = FALSE)

table

```

```

library(dcData)
library(dplyr)
library(ggplot2)

#filter out the names and sex that we are looking for, as well as unwanted columns
df <- BabyNames %>%
  filter(name %in% c("Jenny", "Jenni", "Jennie", "Jennifer")) %>%
  filter(sex %in% "F") %>%
  select(name, count, year)

# create a graph with the appropriate stats
ggplot(df, aes(x = year, y = count, color = name)) +
  geom_line(aes(linetype = name)) +
  theme_minimal() +
  labs(title = "Popularity of Certain Baby Names throughout the Years") +
  scale_x_continuous(breaks = seq(1880, 2013, by = 20))

```

```

library(ggplot2)

#define the function
volume <- function(sideLength) {
  (8.5 - sideLength * 2) * (11 - sideLength * 2) * sideLength
}

#generate the data for the function
squareSideLength <- seq(from = 1.55, to = 1.65, by = 0.01)
df <- data.frame(
  sideLength = squareSideLength,
  volume = volume(squareSideLength)
)

#find the sideLength that would generate the largest volume
vertex_index <- which.max(df$volume)

```

```

vertex <- df[vertex_index, ]

#create the graph
ggplot(df, aes(x = sideLength, y = volume)) +
  geom_line(color = "pink") +
  geom_vline(xintercept = vertex$sideLength, linetype = "dashed", color = "magenta") +
  geom_point(data = vertex, aes(x = sideLength, y = volume), color = "magenta", size = 3) +
  labs(
    title = "Box Volume vs. Side Length",
    x = "Side Length",
    y = "Volume"
  )

```