

Activity 14: QMD File

STAT 184

Harshini Mallipedhi

Monday, November 10, 2025

Table of contents

1 Armed Forces Data Wrangling Redux (Activities #08 and #10)	1
2 Popularity of Baby Names (Activity #13)	3
3 Plotting a Mathematical Function (Activity #04)	4
4 What You Feel You've Learned So Far	4
5 Code Appendix	5
5.1 Armed Forces Data Wrangling (Activities #08 and #10)	5
5.2 Popularity of Baby Names (Activity #13)	6
5.3 Plotting a Mathematical Function (Activity #04)	7

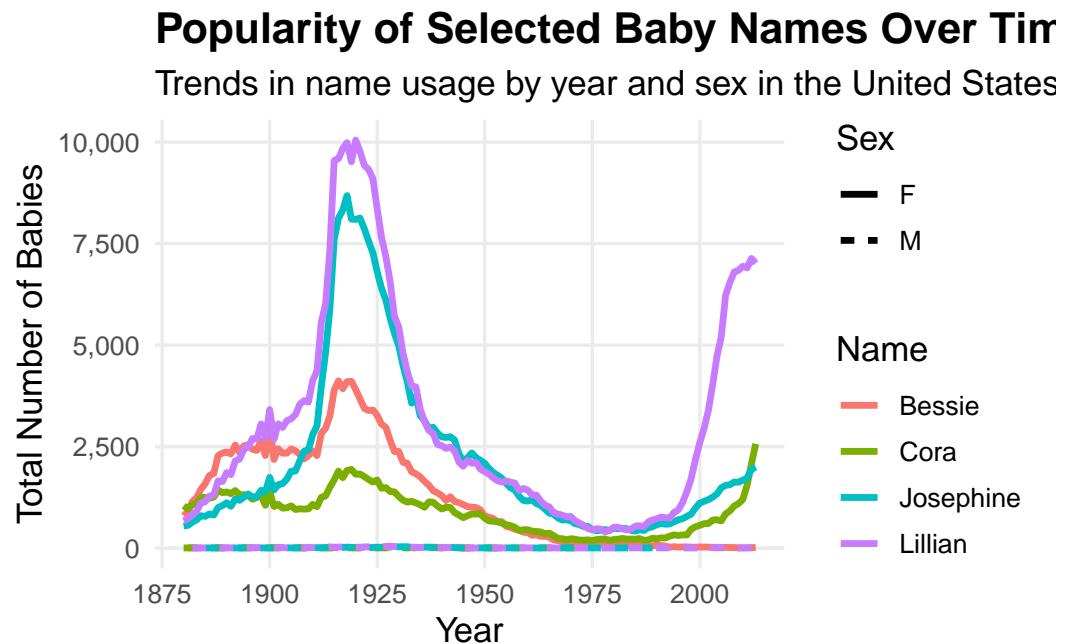
1 Armed Forces Data Wrangling Redux (Activities #08 and #10)

Table 1: Distribution of Males by Pay Grade and Branch

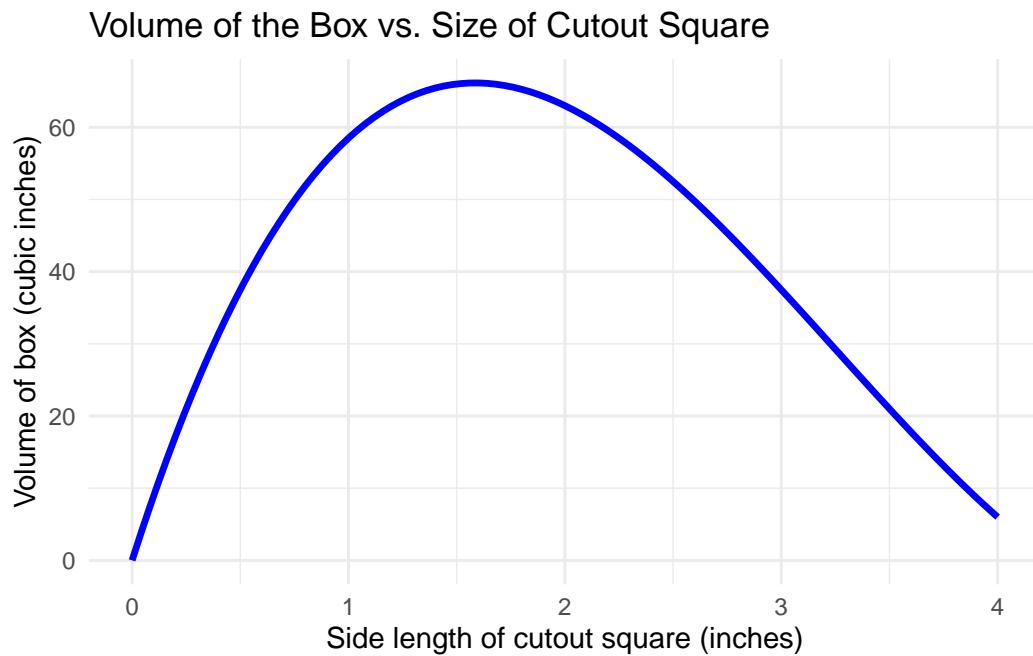
Pay_Grade	AirForce	Army	MarineCorps	Navy	Space-Force	Total
E1	8,537 (0.8%)	7,429 (0.7%)	7,849 (0.8%)	8,903 (0.9%)	179 (0.0%)	32,897 (3.1%)
E2	7,343 (0.7%)	22,338 (2.1%)	15,034 (1.4%)	17,504 (1.7%)	186 (0.0%)	62,405 (6.0%)
E3	37,324 (3.6%)	43,775 (4.2%)	35,239 (3.4%)	25,436 (2.4%)	1,015 (0.1%)	142,789 (13.7%)

Pay_Grade	AirForce	Army	MarineCorps	Navy	Space-Force	Total
E4	53,185 (5.1%)	79,234 (7.6%)	28,519 (2.7%)	33,859 (3.2%)	541 (0.1%)	195,338 (18.7%)
E5	40,614 (3.9%)	54,803 (5.2%)	22,262 (2.1%)	58,142 (5.6%)	859 (0.1%)	176,680 (16.9%)
E6	31,400 (3.0%)	49,502 (4.7%)	12,225 (1.2%)	45,833 (4.4%)	853 (0.1%)	139,813 (13.4%)
E7	18,309 (1.8%)	30,264 (2.9%)	7,720 (0.7%)	19,046 (1.8%)	535 (0.1%)	75,874 (7.3%)
E8	3,876 (0.4%)	9,482 (0.9%)	3,495 (0.3%)	6,007 (0.6%)	112 (0.0%)	22,972 (2.2%)
E9	1,903 (0.2%)	2,865 (0.3%)	1,515 (0.1%)	2,574 (0.2%)	47 (0.0%)	8,904 (0.9%)
O1	5,048 (0.5%)	7,122 (0.7%)	2,412 (0.2%)	5,497 (0.5%)	412 (0.0%)	20,491 (2.0%)
O10	11 (0.0%)	11 (0.0%)	3 (0.0%)	8 (0.0%)	3 (0.0%)	36 (0.0%)
O2	5,045 (0.5%)	9,550 (0.9%)	3,162 (0.3%)	5,544 (0.5%)	437 (0.0%)	23,738 (2.3%)
O3	15,715 (1.5%)	20,986 (2.0%)	5,385 (0.5%)	14,480 (1.4%)	997 (0.1%)	57,563 (5.5%)
O4	9,682 (0.9%)	12,350 (1.2%)	3,637 (0.3%)	7,983 (0.8%)	941 (0.1%)	34,593 (3.3%)
O5	7,373 (0.7%)	6,939 (0.7%)	1,830 (0.2%)	5,525 (0.5%)	657 (0.1%)	22,324 (2.1%)
O6	2,663 (0.3%)	3,161 (0.3%)	656 (0.1%)	2,644 (0.3%)	206 (0.0%)	9,330 (0.9%)
O7	99 (0.0%)	100 (0.0%)	36 (0.0%)	101 (0.0%)	11 (0.0%)	347 (0.0%)
O8	63 (0.0%)	80 (0.0%)	28 (0.0%)	62 (0.0%)	10 (0.0%)	243 (0.0%)
O9	30 (0.0%)	46 (0.0%)	17 (0.0%)	32 (0.0%)	4 (0.0%)	129 (0.0%)
W1	27 (0.0%)	3,727 (0.4%)	494 (0.0%)	44 (0.0%)	0 (0.0%)	4,292 (0.4%)
W2	33 (0.0%)	6,024 (0.6%)	725 (0.1%)	641 (0.1%)	0 (0.0%)	7,423 (0.7%)
W3	0 (0.0%)	2,794 (0.3%)	518 (0.0%)	744 (0.1%)	0 (0.0%)	4,056 (0.4%)
W4	0 (0.0%)	1,378 (0.1%)	265 (0.0%)	432 (0.0%)	0 (0.0%)	2,075 (0.2%)
W5	0 (0.0%)	494 (0.0%)	104 (0.0%)	69 (0.0%)	0 (0.0%)	667 (0.1%)
Total	248,280 (23.8%)	374,454 (35.8%)	153,130 (14.7%)	261,110 (25.0%)	8,005 (0.8%)	1,044,979 (100.0%)

2 Popularity of Baby Names (Activity #13)



3 Plotting a Mathematical Function (Activity #04)



4 What You Feel You've Learned So Far

I have learned how to use R for data analysis and visualization. I can now read data files, clean them, organize them, and use different functions to explore patterns in the data. I have learned how to use different packages (like dplyr, tidyr, and ggplot2) to make my work easier and more efficient. Writing code in R has helped me understand how data can be transformed step by step. I realized how important data wrangling is because it makes sure that the data is accurate, consistent, and ready for analysis. Without proper wrangling, it would be difficult to draw reliable conclusions. I also learned how to interpret statistics and understand what the numbers in my analysis really mean. I have come to appreciate how much clear and thoughtful visualizations can impact how results are understood. The example of the water pump map from England showed me how data visualization can reveal problems that might not be seen otherwise. And finally, I learned the importance of having good code comments. In the beginning, R seemed confusing, but now I feel more confident using it. I understand how coding and visualization work together to solve real problems by making patterns clear and easy to understand.

5 Code Appendix

5.1 Armed Forces Data Wrangling (Activities #08 and #10)

```
#load all packages
library(dplyr)
library(tidyr)
library(readr)
library(janitor)
library(knitr)

#read in data
ArmedForcesRaw <- read.table(
  file   = "/Users/harshinimallipedhi/Desktop/US_Armed_Forces_(6_2025) - Sheet1.csv",
  header = FALSE, sep = ",", skip = 1, stringsAsFactors = FALSE
)

ArmedForcesRaw <- ArmedForcesRaw[
  -c(1, 2, 12, 18, 29, 30, 31),
  -c(4, 7, 10, 13, 16, 17, 18, 19)
]

ArmedForcesRaw <- ArmedForcesRaw %>%
  rename(
    Pay_Grade = V1, Army_Male = V2, Army_Female = V3,
    Navy_Male = V5, Navy_Female = V6,
    MarineCorps_Male = V8, MarineCorps_Female = V9,
    AirForce_Male = V11, AirForce_Female = V12,
    SpaceForce_Male = V14, SpaceForce_Female = V15
  )

AF_tidy_individualCase <- ArmedForcesRaw %>%
  pivot_longer(cols = -Pay_Grade,
               names_to = c("Branch", "Sex"), names_sep = "_",
               values_to = "Count") %>%
  mutate(Count = parse_number(Count)) %>%
  filter(!is.na(Count) & Count > 0) %>%
  uncount(Count)

#narrower subset of soldiers is male soldiers
male_table <- AF_tidy_individualCase %>%
```

```

filter(Sex == "Male") %>%
tabyl(Pay_Grade, Branch) %>%
adorn_totals(c("row","col")) %>%
adorn_percentages("all") %>%
adorn_pct_formatting(digits = 1) %>%
adorn_ns(position = "front")

kable(male_table, caption = "Distribution of Males by Pay Grade and Branch")

```

5.2 Popularity of Baby Names (Activity #13)

```

#| echo: true
#| eval: false

#load the packages
library(tidyverse)
library(dcData)
library(scales)

#load the dataset
data("BabyNames")

#select the names
names_selected <- c("Josephine", "Bessie", "Cora", "Lillian")

#filter the dataset
selected_names <- BabyNames %>%
filter(name %in% names_selected) %>%
select(name, sex, count, year)

#summarize data
plot_df <- selected_names %>%
group_by(name, sex, year) %>%
summarise(count = sum(count), .groups = "drop")

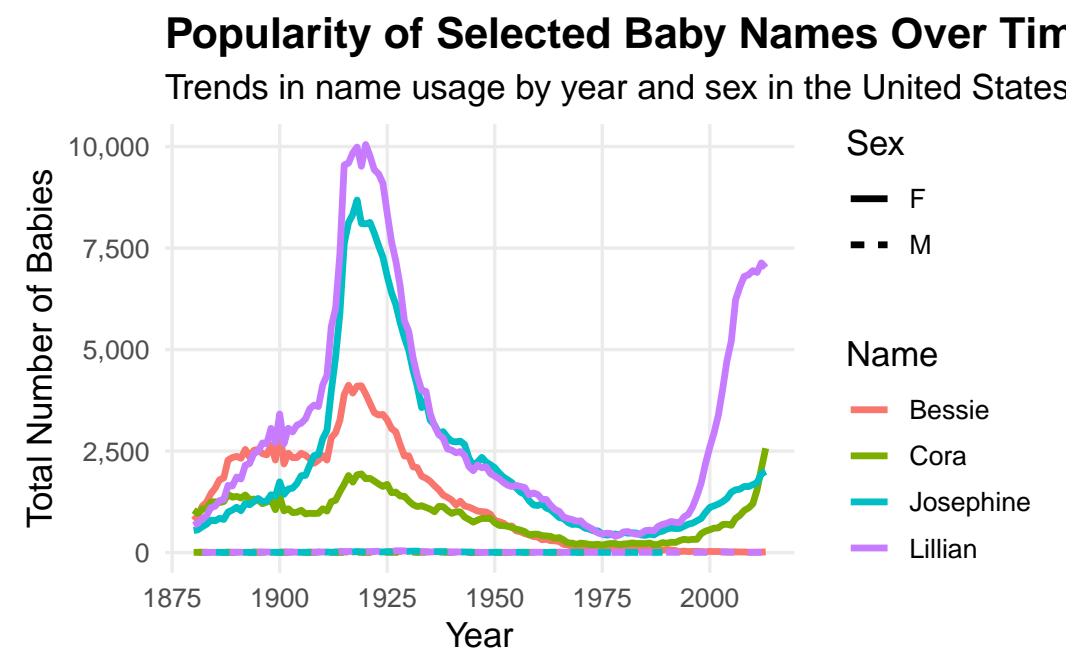
#line graph
ggplot(plot_df, aes(x = year, y = count, color = name, linetype = sex, group = interaction(name, sex))) +
geom_line(linewidth = 1.2, na.rm = TRUE) +
scale_y_continuous(labels = comma) +
labs(
  title = "Popularity of Baby Names Over Time",
  subtitle = "Count vs Year for Selected Names"
)

```

```

title = "Popularity of Selected Baby Names Over Time",
subtitle = "Trends in name usage by year and sex in the United States (1880–2022)",
x = "Year",
y = "Total Number of Babies",
color = "Name",
linetype = "Sex"
) +
theme_minimal(base_size = 13) +
theme(
plot.title = element_text(face = "bold"),
panel.grid.minor = element_blank()
)

```



5.3 Plotting a Mathematical Function (Activity #04)

```

#load packages
library(ggplot2)

#createing the volume function
volume_box <- function(x) {

```

```

volume <- x * (11 - 2 * x) * (8.5 - 2 * x)
return(volume)
}

#making the plot

ggplot(data.frame(x = c(0, 4)), aes(x = x)) +
  stat_function(fun = volume_box, color = "blue", size = 1.2) +
  labs(
    title = "Volume of the Box vs. Size of Cutout Square",
    x = "Side length of cutout square (inches)",
    y = "Volume of box (cubic inches)"
  ) +
  theme_minimal()

```

