STAT 184 – Activity 14: A First QMD File

Devin Charles

2025-11-10

Armed Forces Data Wrangling Redux

Question 3: Data Wrangling Code for Armed Forces Data

Here is where I tidied my code from activity #10, but as a slightly revised version.

```
library(dplyr)
library(tidyr)
library(knitr)

#load the data
armed_forces <- read.csv("active_duty_individual_8.csv",
                         stringsAsFactors = FALSE)

#rank mapping table
rank_map <- tribble(
  ~Pay_Grade, ~Army, ~Navy, ~MarineCorps, ~AirForce, ~SpaceForce, ~CoastGuard,
  "E1", "Private", "Seaman Recruit", "Private", "Airman Basic", "Specialist 1",
  "Seaman Recruit",
  "E2", "Private", "Seaman Apprentice", "Private First Class", "Airman",
  "Specialist 2", "Seaman Apprentice",
  "E3", "Private First Class", "Seaman", "Lance Corporal",
  "Airman First Class", "Specialist 3", "Seaman",
  "E4", "Corporal OR Specialist", "Petty Officer Third Class", "Corporal",
  "Senior Airman", "Specialist 4", "Petty Officer Third Class",
  "E5", "Sergeant", "Petty Officer Second Class", "Sergeant", "Staff Sergeant",
  "Sergeant", "Petty Officer Second Class",
  "E6", "Staff Sergeant", "Petty Officer First Class", "Staff Sergeant",
  "Technical Sergeant", "Technical Sergeant", "Petty Officer First Class",
  "E7", "Sergeant First Class", "Chief Petty Officer",
```

```
  "Gunnery Sergeant", "Master Sergeant OR First Sergeant", "Master Sergeant",
  "Chief Petty Officer",
  "E8", "First Sergeant OR Master Sergeant", "Senior Chief Petty Officer",
  "First Sergeant OR Master Sergeant", "Senior Master Sergeant OR First Sergeant",
  "Senior Master Sergeant", "Senior Chief Petty Officer",
  "E9", "Sergeant Major OR Command Sergeant Major",
  "Master Chief Petty Officer OR Fleet/Command Master Chief Petty Officer",
  "Sergeant Major OR Master Gunnery Sergeant",
  "Chief Master Sergeant OR First Sergeant",
  "Chief Master Sergeant",
  "Master Chief Petty Officer OR Fleet/Command Master Chief Petty Officer"
)

#join and clean up naming
armed_forces_ranked <- armed_forces %>%
  left_join(rank_map, by = "Pay_Grade") %>%
  mutate(
    Rank = case_when(
      Branch == "Army" ~ Army,
      Branch == "Navy" ~ Navy,
      Branch == "MarineCorps" ~ MarineCorps,
      Branch == "AirForce" ~ AirForce,
      Branch == "SpaceForce" ~ SpaceForce,
      Branch == "CoastGuard" ~ CoastGuard,
      TRUE ~ NA_character_
    )
  ) %>%
  select(Gender, Branch, Rank)
```

Question 4: Visualization for the Armed Forces

In this section I created a two-way frequency table to see the impact of sex and rank in the US Armed Forces.

```
freq_table <- armed_forces_ranked %>%
filter(Branch == "Army", !is.na(Rank)) %>% #filter out the N/A
group_by(Gender, Rank) %>% # group the gender and rank together
summarise(Count = n(), .groups = "drop") %>%
group_by(Gender) %>%
mutate(RelFreq = Count / sum(Count)) #calculate relative frequency
```

```
kable(freq_table, caption = "Army Enlisted Frequency Table: Gender by Rank")
```

Table 1: Army Enlisted Frequency Table: Gender by Rank

Gender	Rank	Count	RelFreq
Female	Corporal OR Specialist	28519	0.2130541
Female	First Sergeant OR Master Sergeant	3495	0.0261098
Female	Private	22883	0.1709498
Female	Private First Class	35239	0.2632566
Female	Sergeant	22262	0.1663106
Female	Sergeant First Class	7720	0.0576731
Female	Sergeant Major OR Command Sergeant Major	1515	0.0113180
Female	Staff Sergeant	12225	0.0913281
Male	Corporal OR Specialist	79234	0.2643848
Male	First Sergeant OR Master Sergeant	9482	0.0316391
Male	Private	29767	0.0993253
Male	Private First Class	43775	0.1460666
Male	Sergeant	54803	0.1828644
Male	Sergeant First Class	30264	0.1009837
Male	Sergeant Major OR Command Sergeant Major	2865	0.0095598
Male	Staff Sergeant	49502	0.1651762
Total	Corporal OR Specialist	541	0.1250289
Total	First Sergeant OR Master Sergeant	112	0.0258840
Total	Private	365	0.0843541
Total	Private First Class	1015	0.2345736
Total	Sergeant	859	0.1985209
Total	Sergeant First Class	535	0.1236422
Total	Sergeant Major OR Command Sergeant Major	47	0.0108620
Total	Staff Sergeant	853	0.1971343

Question 5: Narrative Text for the Armed Forces Section

The visualization shows how gender and rank are distributed among enlisted people. Most men and women are revolving around the lower enlisted ranks, especially the Private First Class and Corporal/Specialist. Women make up a smaller share of higher ranks like Sergeant First Class and Sergeant Major. This pattern shows that gender and rank are not independent, men tend to lean towards senior ranks, while women tend to lean to the lower and middle class ranks.

Popularity of Baby

Question 6: Code for the Popular Baby Names Project

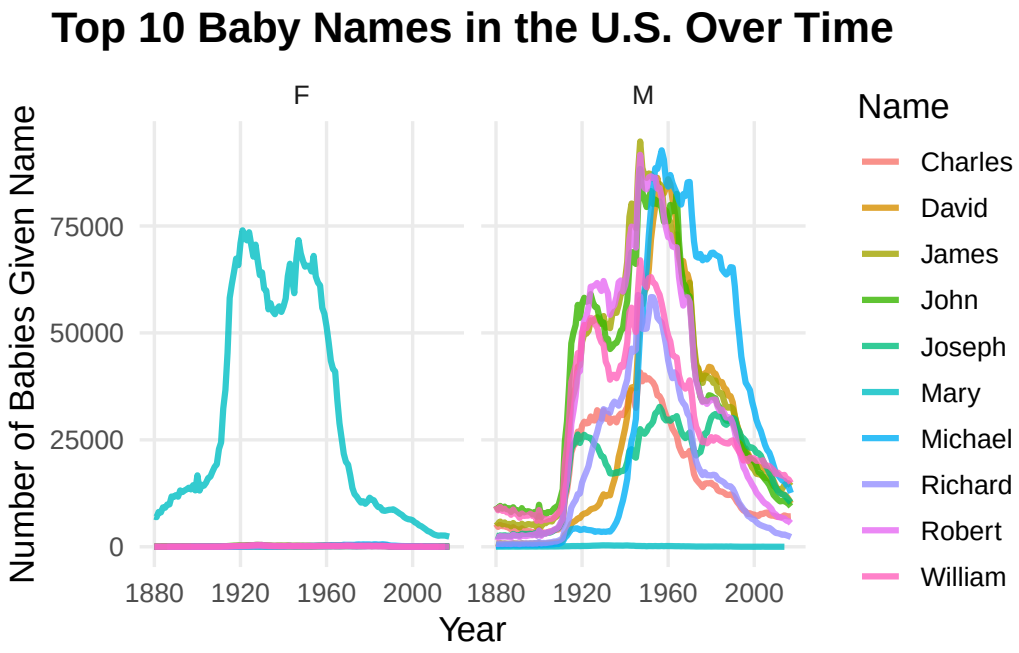
```
#load libraries
library(babynames)
library(dplyr)
library(ggplot2)

#load baby names data set
data("babynames")

#find 10 most popular baby names for all years and sexes
top10_names <- babynames %>%
group_by(name) %>%
summarise(total = sum(n)) %>%
arrange(desc(total)) %>%
slice_head(n = 10)

#filter the data set to include only those names
top10_data <- babynames %>%
filter(name %in% top10_names$name)

#time series plot of name popularity
ggplot(top10_data, aes(x = year, y = n, color = name)) +
geom_line(linewidth = 1.1, alpha = 0.8) +
facet_wrap(~sex) +
labs(
title = "Top 10 Baby Names in the U.S. Over Time",
x = "Year",
y = "Number of Babies Given Name",
color = "Name"
) +
theme_minimal(base_size = 13) +
theme(
plot.title = element_text(face = "bold", hjust = 0.5),
plot.subtitle = element_text(hjust = 0.5, color = "gray30"),
legend.position = "right",
panel.grid.minor = element_blank()
)
```



Question 7: Visualization for Popular Baby Names Project

This visualization shows the most common baby names in the United States over time. It compares how certain names rise and fall in popularity for boys and girls, showing cultural and social influences over different decades.

```
#time series plot of name popularity
ggplot(top10_data, aes(x = year, y = n, color = name, linetype = name)) +
  geom_line(linewidth = 1.1, alpha = 0.8) +
  facet_wrap(~sex) +
  labs(
    title = "Top 10 Baby Names in the U.S. Over Time",
    x = "Year",
    y = "Number of Babies Given Name",
    color = "Name",
    linetype = "Name"
  ) +
  scale_color_brewer(palette = "Dark2") +
  theme_minimal(base_size = 13) +
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5, color = "gray30"),
```

```
    legend.position = "right",
    panel.grid.minor = element_blank()
  )
```

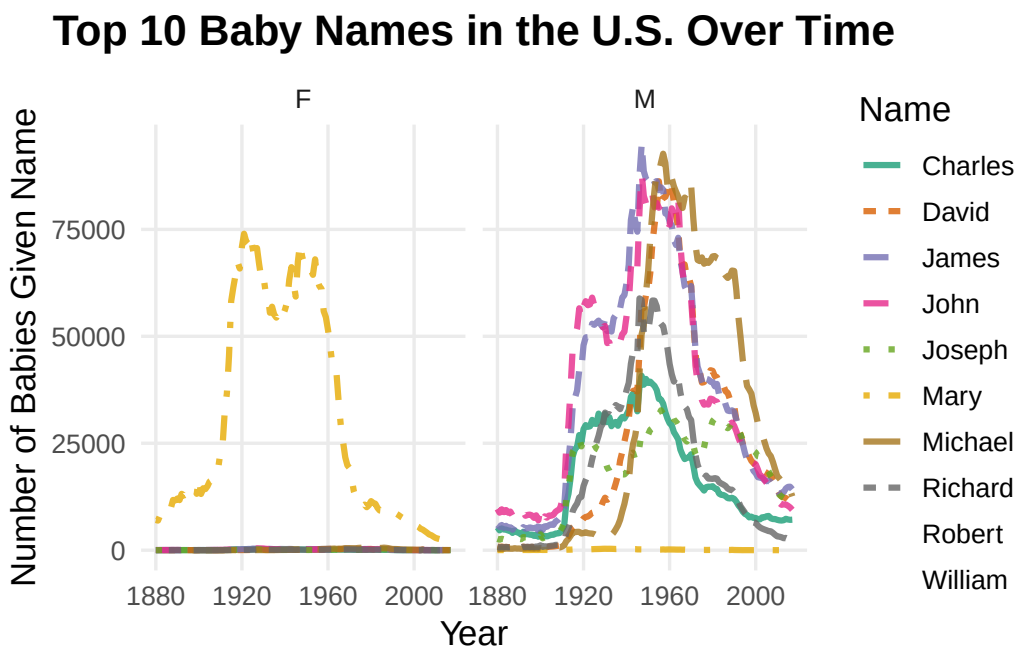


Figure 1: Trends in the Top 10 U.S. Baby Names Over Time

Question 8: Narrative Text for the Popular Baby Names Project

The time plot shows the changing popularity of the 10 most common names across 120 years. Some names were common, then dropped off, while other stayed high for most of it. These different shifts show how culture and society can affect people's baby naming for new parents.

The Box Problem

Question 9: Code for the Box Problem

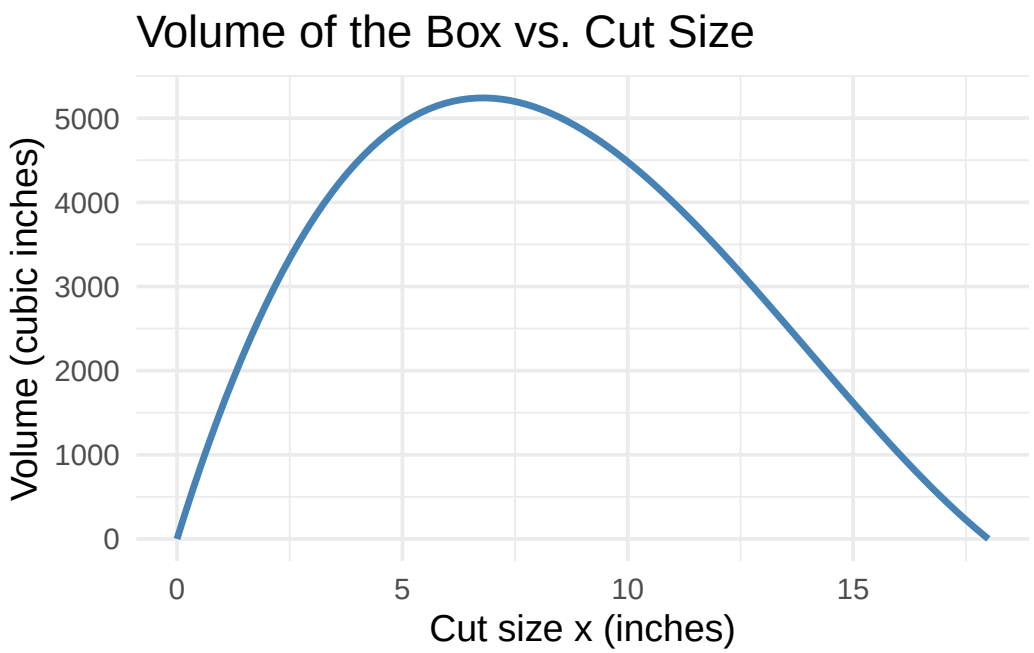
```
#volume box
volume_box <- function(x) {
  V <- x * (48 - 2*x) * (36 - 2*x)
```

```r
  return(V)
}

#load libraries
library(ggplot2)

#create the ggplot
box_plot <- ggplot(data.frame(x = c(0, 18)), aes(x = x)) +
  stat_function(fun = volume_box, color = "steelblue", linewidth = 1.2) +
  labs(
    title = "Volume of the Box vs. Cut Size",
    x = "Cut size x (inches)",
    y = "Volume (cubic inches)"
  ) +
  theme_minimal(base_size = 14)

#add alt text
box_plot + labs(
  alt = "Line graph showing how the volume of a box changes as cut size
  increases. The curve rises, peaks near x = 6 in, then falls back to zero
  as cuts become too large."
)
```



Question 10: Narrative Text for the Box Problem

The plot shows a volume increase until the cut size reaches around 6 inches with a maximum volume around 5200 cubic inches. After the maximum the volume decreases because the dimensions shrink too much.

Self-Reflection

Throughout this course, I have learned how to use data visualizations and stat tools in RStudio to explore real-world data. I have become very familiar with R and a lot of features within it, even though I am taking other 400 level R classes. Having good visuals that follow a good design from Tufte and Kosslyn is a necessity. I have also learned the importance of commenting sections of code in order to show what each block does. Overall, I developed good skills so far and want to keep advancing that.