Stat 184 Activity 14

Lance He

2025-11-12

Table of contents

Introduction

This is the 'First qmd file' for Activity 14 of Stat 184 class. In the document, I will discuss the following topics:

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

2. Popularity of Baby Names (Activity #13)

3. Plotting a Mathematical Function (Activity #04)

4. What I Feel I've Learned So Far

Armed Forces Data Wrangling Redux

In this section, I am building two tables that organize the cases by rank and branch, one table for male, one for female.

From the table, some obvious conclusions that we can get are: 1. There are no female generals or admirals in any of the branches. 2. There are most people in ranks that have the paygrade from E3, E4, and E5.

Table 1: Frequency of Male ENLISTED Personnel (E1-E9) by Rank and Branch

Rank	Army	Navy	Marine_Corps	Air_Force	Space_Force	Total
E1	7429	8903	7849	8537	179	32897
E2	22338	17504	15034	7343	186	62405
E3	43775	25436	35239	37324	1015	142789
E4	79234	33859	28519	53185	541	195338
E5	54803	58142	22262	40614	859	176680
E6	49502	45833	12225	31400	853	139813
E7	30264	19046	7720	18309	535	75874
E8	9482	6007	3495	3876	112	22972
E9	2865	2574	1515	1903	47	8904

Table 2: Frequency of Female ENLISTED Personnel (E1-E9) by Rank and Branch

Rank	Army	Navy	Marine_Corps	Air_Force	Space_Force	Total
E1	1326	3434	655	1933	38	7386
E2	4336	5833	1684	2019	41	13913
E3	10229	9103	4174	10369	194	34069
E4	15143	9959	2961	15055	179	43297
E5	10954	16169	2670	10762	173	40728
E6	7363	9950	1529	6679	147	25668
E7	4410	3434	747	4807	114	13512
E8	1472	850	293	1221	25	3861
E9	394	368	82	523	16	1383

Popularity of Baby Names

Summary and Observations:

This line chart tracks the popularity of four different names (two male, two female) from 1880 to the present. The x-axis shows the year, the y-axis shows the total number of births, and

each name gets its own colored line. It's really interesting to see how "James" and "Michael" peaked in the mid-1900s, while "Jennifer" and "Jessica" had their massive spike in popularity much later, around the 1980s.
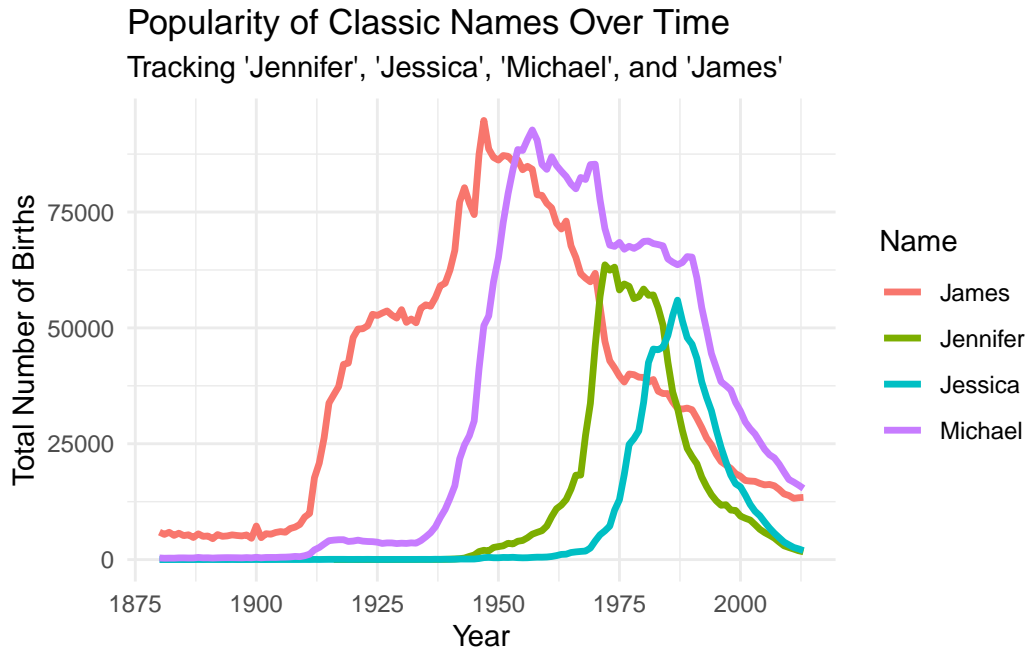

Figure 1: Popularity of Classic Names

Plotting a Mathematical Function

Summary and Observations:

1. The estimated rounded maximum volume of the box (to one decimal place) is 66.15 cubic inches.
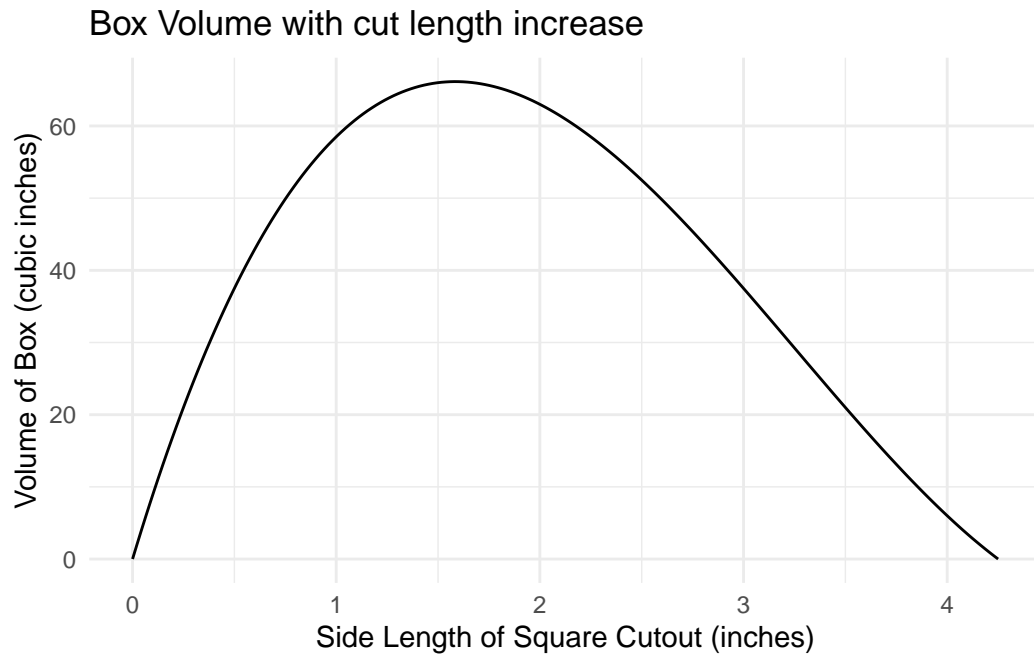2. The side length of the square that results in this volume (to two decimal places) is 1.59 inches.

Figure 2: Plotting a Mathematical Function

What You Feel You've Learned So Far

I learned how to use the quarto files and use the pdf format of it. Also, using descriptive language to give the readers contexts about the plots and tables. Last but not least, I improved some of my older code and makes it being able to run on other computers without alternations.

Code Appendix

Armed Forces Data Wrangling (Activities #08 & #10)

```r
# Coding Style: Tidyverse
library(tidyverse)


file_path <- "https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5q

col_names <- c(
  "Rank",
  "Army_Male", "Army_Female", "Army_Total",
  "Navy_Male", "Navy_Female", "Navy_Total",
  "Marine_Corps_Male", "Marine_Corps_Female", "Marine_Corps_Total",
  "Air_Force_Male", "Air_Force_Female", "Air_Force_Total",
  "Space_Force_Male", "Space_Force_Female", "Space_Force_Total",
  "Total_Male", "Total_Female", "Total_Total"
)

# Load the raw data from the CSV file with custom column names.
armed_forces_raw <- read_csv(
  file_path,
  skip = 3,
  col_names = col_names,
  na = "N/A*",
  n_max = 26
)

rank_titles <- tibble::tribble(
  ~Rank, ~Title,
  "E1", "E1: Private / Seaman Recruit / Private / Airman Basic / Specialist 1",
  "E2", "E2: Private / Seaman Apprentice / Private First Class / Airman / Specialist 2",
  "E3", "E3: Private First Class / Seaman / Lance Corporal / Airman First Class / Specialist
  "E4", "E4: Corporal OR Specialist / Petty Officer Third Class / Corporal / Senior Airman /
  "E5", "E5: Sergeant / Petty Officer Second Class / Sergeant / Staff Sergeant / Sergeant",
  "E6", "E6: Staff Sergeant / Petty Officer First Class / Staff Sergeant / Technical Sergean
  "E7", "E7: Sergeant First Class / Chief Petty Officer / Gunnery Sergeant / Master Sergeant
  "E8", "E8: First Sergeant OR Master Sergeant / Senior Chief Petty Officer / First Sergeant
  "E9", "E9: Sergeant Major OR Command Sergeant Major / Master Chief Petty Officer OR Fleet/C
  "W1", "W1: Warrant Officer / Warrant Officer / Warrant Officer / Warrant Officer",
```

```
  "W2", "W2: Chief Warrant Officer / Chief Warrant Officer / Chief Warrant Officer / Chief Wa
  "W3", "W3: Chief Warrant Officer / Chief Warrant Officer / Chief Warrant Officer / Chief Wa
  "W4", "W4: Chief Warrant Officer / Chief Warrant Officer / Chief Warrant Officer / Chief Wa
  "W5", "W5: Chief Warrant Officer / Chief Warrant Officer / Chief Warrant Officer / Chief Wa
  "O1", "O1: Second Lieutenant / Ensign / Second Lieutenant / Second Lieutenant / Second Lieu
  "O2", "O2: First Lieutenant / Lieutenant Junior Grade / First Lieutenant / First Lieutenant
  "O3", "O3: Captain / Lieutenant / Captain / Captain / Captain",
  "O4", "O4: Major / Lieutenant Commander / Major / Major / Major",
  "O5", "O5: Lieutenant Colonel / Commander / Lieutenant Colonel / Lieutenant Colonel / Lieut
  "O6", "O6: Colonel / Captain / Colonel / Colonel / Colonel",
  "O7", "O7: Brigadier General / Rear Admiral (Lower) / Brigadier General / Brigadier General
  "O8", "O8: Major General / Rear Admiral (Upper) / Major General / Major General / Major Gen
  "O9", "O9: Lieutenant General / Vice Admiral / Lieutenant General / Lieutenant General / Li
  "O10", "O10: General / Admiral / General / General / General",
)

# Clean and reshape the data into a tidy, long format.
armed_forces_tidy <- armed_forces_raw %>%
  left_join(rank_titles, by = "Rank") %>%
  filter(!is.na(Title)) %>%
  # The description of the rank makes the table messy, so we only keeps the 'Rank' column
  # This keeps the original 'Rank' column (E1, E2, etc.).
  select(-Title) %>%
  relocate(Rank) %>%
  select(-ends_with("_Total")) %>%
  pivot_longer(
    cols = -Rank,
    names_to = "Branch_Sex",
    values_to = "Count"
  ) %>%
  extract(Branch_Sex, into = c("Branch", "Sex"), regex = "(.*)_(Male|Female)$") %>%
  mutate(
    Count = as.numeric(gsub(",", "", Count)),
    Count = replace_na(Count, 0)
  )


enlisted_tidy_data <- armed_forces_tidy %>%
  filter(str_starts(Rank, "E"))


# Create a final frequency table for male ENLISTED personnel.
```

```
male_frequency_table <- enlisted_tidy_data %>%
  filter(Sex == "Male") %>%
  select(-Sex) %>%
  pivot_wider(
    names_from = Branch,
    values_from = Count
  )

# Create a final frequency table for female ENLISTED personnel.
female_frequency_table <- enlisted_tidy_data %>%
  filter(Sex == "Female") %>%
  select(-Sex) %>%
  pivot_wider(
    names_from = Branch,
    values_from = Count
  )

# Print the tables using knitr::kable() so they appear in the PDF.
knitr::kable(male_frequency_table, caption = "Frequency of Male ENLISTED Personnel (E1-E9) by
knitr::kable(female_frequency_table, caption = "Frequency of Female ENLISTED Personnel (E1-E9
```

Popularity of Baby Names (Activity #13)

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

# 1. Filter the data and create the 3-attribute plot
BabyNames %>%

  filter(
    (name == "Jennifer" & sex == "F") |
      (name == "Jessica" & sex == "F")  |
      (name == "Michael" & sex == "M")   |
      (name == "James" & sex == "M")
  ) %>%

  # 2. Create the plot
  ggplot(mapping = aes(x = year, y = count, color = name)) +
```

```
  # 3. Add the line geometry
  geom_line(linewidth = 1.2) +

  # 4. Add a professional theme and polished labels
  theme_minimal() +
  labs(
    title = "Popularity of Classic Names Over Time",
    subtitle = "Tracking 'Jennifer', 'Jessica', 'Michael', and 'James'",
    x = "Year",
    y = "Total Number of Births",
    color = "Name"
  )
```

Plotting a Mathematical Function (Activity #04)

```
find_volume <- function(cut_length){
  length = 8.5
  width = 11
  box_length = length - 2 * cut_length
  box_width = width - 2 * cut_length
  height = cut_length

  Volume <- box_length*box_width*height


  ifelse(cut_length >= 0 & cut_length <= 8.5 / 2, Volume, NA)
}

# 2. Create the ggplot using stat_function as required by the assignment
#    The domain is from 0 to the maximum possible cut (4.25)
ggplot(data.frame(x = c(0, 4.25)), aes(x = x)) +
  stat_function(fun = find_volume, n = 200) + # n=200 makes a smoother curve
  labs(
    title = "Box Volume with cut length increase",
    x = "Side Length of Square Cutout (inches)",
    y = "Volume of Box (cubic inches)"
  ) +
  theme_minimal()
```