

Activity 14

1. Armed Forces Data Wrangling Revised

Below, is the new data table for the Armed Forces

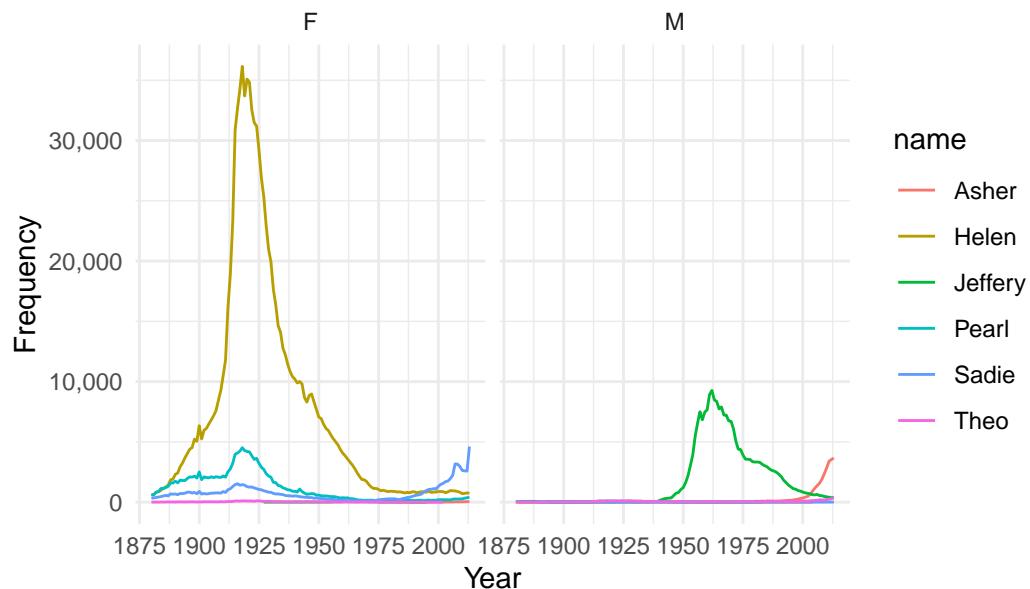
Table 1: Space Force Commissioned Officers

Rank	Female	Male
Brigadier General	1	11
Captain	246	1032
Colonel	33	203
First Lieutenant	160	409
General	0	3
Lieutenant Colonel	124	620
Lieutenant General	1	4
Major	196	951
Major General	0	9
Second Lieutenant	137	387

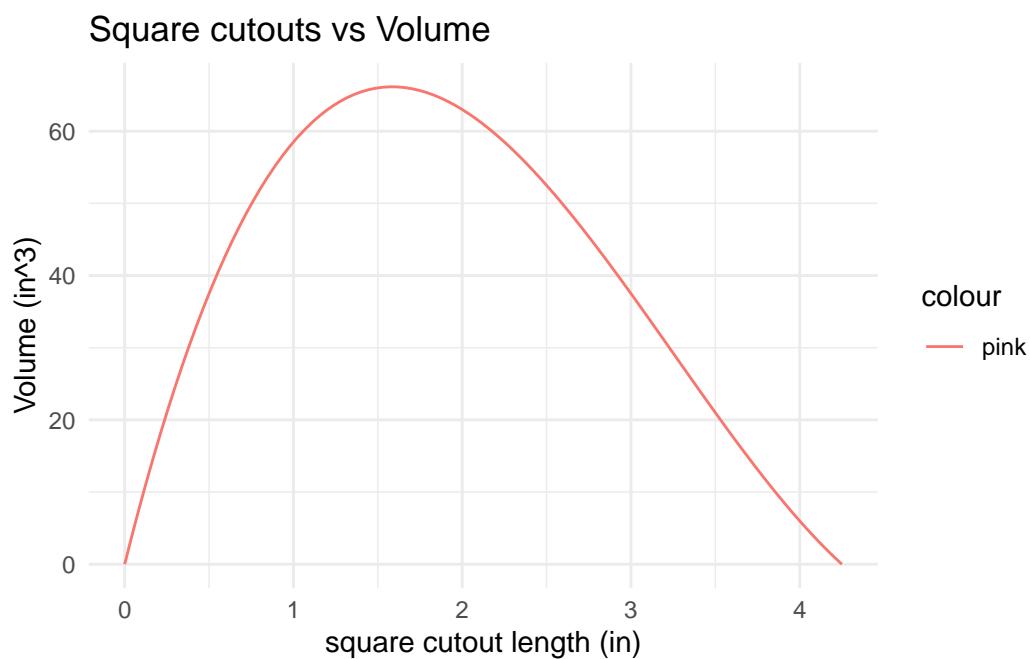
2. Popularity of Baby Names

I selected the names Helen, Sadie, Pearl, Jeffery, Theo, and Asher. I didn't want to select extremely popular names like Mary or John because those are popular and have always been quite popular. The names are selected are names that are rather uncommon. All the names have been heard of before; however, there is no "expected" trend for each name.

Popularity of Names over the Years



3. Plotting a Mathematical Function



4. Reflection

I have learned the basics of how to code in R. Since I started off the course not knowing how to code at all, I had to learn a lot of how coding works and how the syntax is very important to run the code. I learned how important comments are as well, especially after Activity 8, where the code started to become more advanced and layered. Plans (especially the steps section) also became very helpful when the activities had several parts to it. My favorite part was learning about the graph, coding them, and making them colorful. I feel like I've learned a lot about the basics of R and how to do simple things within it as well as how it's used in the real world.

5. Code Appendix

All code used in this document is included below.

```
# Load libraries
library(tidyverse)
library(googlesheets4)
library(gsheets)
library(rvest)
library(janitor)
library(knitr)

# Uploading the table to R
google_group_raw <- gsheet2tbl("https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJi")

# gs4_deauth()
# google_group_raw <- read_sheet(
#   ss = "https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bw"
#   na = c("", "N/A")
# )

# Removing columns and rows that aren't necessary
google_group_clean <- google_group_raw %>%
  slice(-c(1, 2, 12, 18, 29, 30, 31)) %>%
  select(-c(4, 7, 10, 13, 16, 17, 18, 19))

# Renaming each column
colnames(google_group_clean) <- c(
  "Pay Grade",
```

```

"Army:Male", "Army:Female",
"Navy:Male", "Navy:Female",
"Marine Corps:Male", "Marine Corps:Female",
"Air Force:Male", "Air Force:Female",
"Space Force:Male", "Space Force:Female"
)

# Changing table orientation to long table - Not one of the final tables
google_group_long <- google_group_clean %>%
  pivot_longer(
    cols = -`Pay Grade`,
    names_to = c("Service Branch", "Gender"),
    names_sep = ":" ,
    values_to = "Soldiers"
  )

# Uploading the second link which contains the ranks associated with its respective pay grade
soldier_raw <- read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html") %>%
  html_elements(css = "table") %>%
  html_table()

# Cleaning the data to remove categories (like Coast Guard) and renaming columns
soldier_indiv_raw <- soldier_raw[[1]] %>% clean_names()
soldier_indiv_clean <- soldier_indiv_raw %>%
  slice(-c(1, 26)) %>%
  select(-c(1, 8)) %>%
  rename(
    "Pay Grade" = `pay_grade`,
    "Army" = `ranks_by_branch_of_service`,
    "Navy" = `ranks_by_branch_of_service_2`,
    "Marine Corps" = `ranks_by_branch_of_service_3`,
    "Air Force" = `ranks_by_branch_of_service_4`,
    "Space Force" = `ranks_by_branch_of_service_5`
  )

# Changing the table to pivot longer - Table 2 contains Rank attribute
soldier_indiv_long <- soldier_indiv_clean %>%
  pivot_longer(
    cols = -`Pay Grade`,
    names_to = c("Service Branch"),
    values_to = "Rank"
  )

```

```

# Combining the Table 1 and 2 into one data set so the attributes will be Pay grade, service
soldier_group_combo <- google_group_long %>%
  left_join(soldier_indiv_long, by = c("Pay Grade", "Service Branch")) %>%
  select(`Pay Grade`, `Service Branch`, Gender, Soldiers, Rank)

# Removing missing data from dataset
soldier_group_combo_no_NA <- subset(soldier_group_combo, Soldiers != "N/A")

# 2nd Dataset where each case is a person, using uncount function
soldier_group_uncount <- soldier_group_combo_no_NA %>%
  mutate(Soldiers = Soldiers %>% gsub(", ", "", .) %>% as.numeric()) %>%
  tidyverse::uncount(weights = Soldiers)

# Table
space_officers <- soldier_group_uncount %>%
  filter(`Service Branch` == "Space Force" & grepl(pattern = "0", x = `Pay Grade`))
tabyl(dat = space_officers, Rank, Gender) %>%
  kable()

# load libraries
library(dcData)
library(ggplot2)
library(tidyverse)

# Load data
data("BabyNames")

# Assigning the table to a variable
BabyNames_raw <- BabyNames

# New variable has the data altered and rearranged so that only the selected names for all y
BabyNames_sample <- BabyNames_raw %>%
  filter(name %in% c("Helen", "Sadie", "Pearl", "Jeffery", "Theo", "Asher")) %>%
  arrange(name, year)

# Table for the data
sample_table <- BabyNames_sample %>%
  semi_join(BabyNames_sample, by = c("name", "sex"))

# Make a line graph for the table, where color is the names and there are two graphs that are
ggplot(BabyNames_sample, aes(x = year, y = count, group = name,
                             color = name,)) +

```

```

facet_wrap(~ sex) +
geom_line() +
scale_y_continuous(labels = scales::comma) +
labs(
  title = "Popularity of Names over the Years",
  x = "Year",
  y = "Frequency"
) +
theme_minimal()
library(ggplot2)

# Calculate Volume
calculate_V <- function(r) {
  length <- 11 - 2 * r
  width <- 8.5 - 2 * r
  height <- r
  Volume <- length * width * height
  return(Volume)
}

#Assign values to axises
x_values <- seq(from = 0, to = 4.25, by = 0.01)
y_values <- calculate_V(x_values)

# Make a table
table_volume <- data.frame(x_values, y_values)

# Create graph
ggplot(table_volume, aes(x = x_values, y = y_values, col = "pink")) +
  labs(title = "Square cutouts vs Volume", x = "square cutout length (in)",
       y = "Volume (in^3)") +
  stat_function(fun = calculate_V) +
  theme_minimal()

```