

# Activity 14 - Displaying Previous Codes

Kathryn Sujai Sundar

2025-11-12

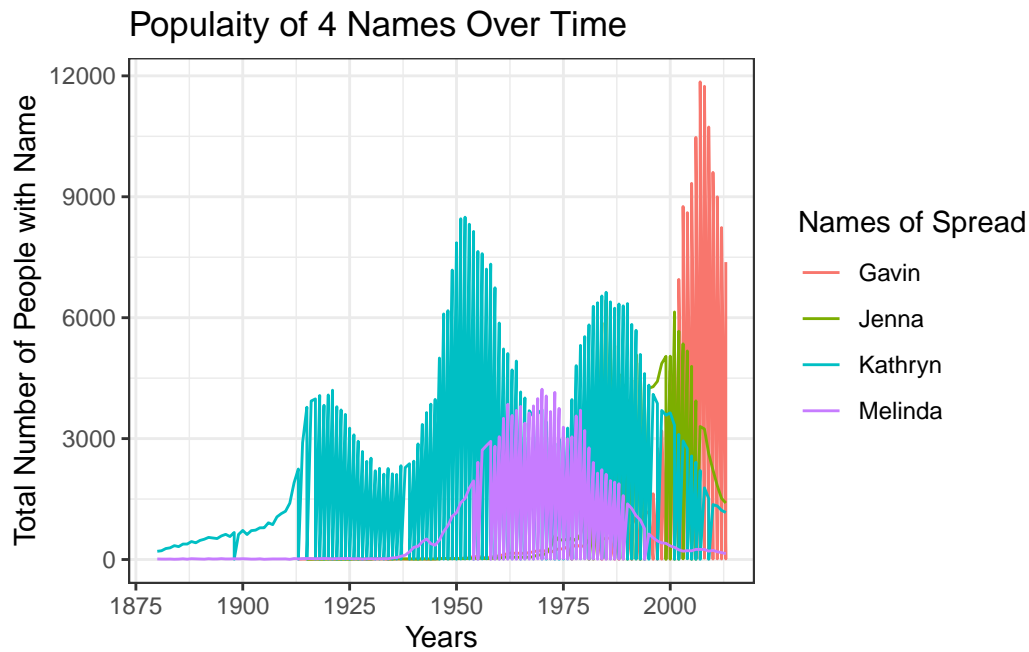
## Armed Forces Data Wrangling Redux

Table 1: Relationship Between Gender and Army Officers

rank_titles	female	male
Brigadier General	18	100
Captain	6053	20986
Colonel	452	3161
First Lieutenant	3006	9550
General	0	11
Lieutenant Colonel	1531	6939
Lieutenant General	5	46
Major	3044	12350
Major General	8	80
Second Lieutenant	2400	7122

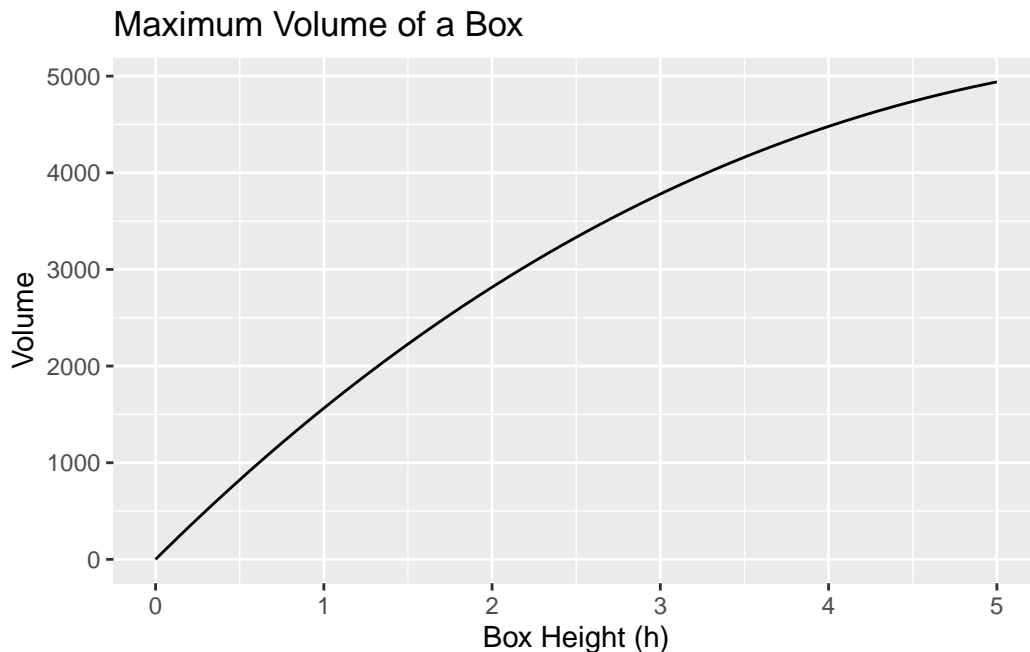
In this frequency table comparing the rank titles of female and male army officers, there is a significantly larger population of male army officers than female. For the title Captain, there are 20,986 men but only 6,053. Similarly, the position Lieutenant Colonel has 6,939 men but 1531 female in that position. This table proves that gender and rank are not independent of each other due to the strong correlation of being a male and having a title compared to being a woman and having a title.

## Popularity of Baby Names



This is a line graph that compares the popularity of four names through time. I choose the names Kathryn, Gavin, Melinda and Jenna because they are all names in my family. Gavin is my brothers name and Melinda and Jenna are my cousins. The graph shows that the names Gavin and Jenna had an increase around the 2000s while the names Kathryn and Melinda were popular in the 1900's. Kathryn had around three majors maximums at 1920's, 1950's and the 1990's. Melinda has one increase around 1975.

## Plotting a Mathematical Function



This graph portrays the relationship between the height of the box and its maximum volume. Through this graph, the box height hits a maximum between the interval (1,2). So, to maximize the volume, the best height in inches would be around 1-2 inches. The maximum volume for the side length of 1.5 would be 66 inches cubed.

## What I've Learned So Far

Through this class, I have the learned the importance of planning and the effect of lack of planning on projects. Planning out how to execute code helps organize and visualize what exactly needs to happen. I have also learned how to create tables that are beneficial for both the creator and consumer. I better understand what the components of a good graph/visual representation include. In general, I also learned how to use and operate R and the various packages and their functions. In turn, I learned the importance of looking into different functions, and knowing what they do and to experiment with different concepts.

## Code Appendix

```

# | echo: false
## Transform the Armed Forces Active Duty Data from a frequency (pivot) table into two data.

# Step 1: Load Packages ----
## Needed Packages:({tidyverse}, {rvest}, {googlesheets4})

library(googlesheets4)
library(tidyverse)
library(rvest)

# Step 2: Load Data ----
## Needed data: (Armed Forces Active Duty Data, Pay Grade and Rank)
armyRaw <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicM1VDF7Gr-nXCb5qbwb_E/",
  skip = 2,
  col_names = TRUE,
  na = c("", "N/A*")) #turn google sheet into a dataframe in R

armyRank_raw <- read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html")%>%
  html_elements(css = "table") %>%
  html_table() #turn Army Ranks data website into R dataframe

# Step 3: Make group case data frame ----
usArmy_clean <- armyRaw %>%
  #rename column names for better readability, filter out unnesessary data, and combine data
  rename(
    pay_grade = `Pay Grade`,
    army_male = Male...2,
    army_female = Female...3,
    navy_male = Male...5,
    navy_female = Female...6,
    marine_male = Male...8,
    marine_female = Female...9,
    airforce_male = Male...11,
    airforce_female = Female...12,
    spaceforce_male = Male...14,
    spaceforce_female = Female...15
  ) %>%
  select(-Total...4, -Total...7, -Total...10, -Total...13, -Total...16, -Male...17, -Female...18)
  filter(!pay_grade %in% c(

```

```

    "Total Enlisted",
    "Total Warrant Officers",
    "Total Officers", "Total",
    "Source: DMDC Active-Duty Military Personnel Master File (June 2025)")
  ) %>%
pivot_longer(
  cols = -pay_grade,
  names_to = c("branch", "gender"),
  names_sep = "_",
  values_to = "total_soldiers"
) %>%
filter(!is.na(total_soldiers), total_soldiers > 0) %>%
filter(!is.na(pay_grade), pay_grade != "")

#select table with needed information from Army Rank data website
armyRank_Raw2 <- (armyRank_raw[[1]])

names(armyRank_Raw2) <- make.unique(names(armyRank_Raw2)) #Clean names in columns
names(armyRank_Raw2)[names(armyRank_Raw2) == ""] <- "junk" #remove unnecessary information
names(armyRank_Raw2) <- make.names(names(armyRank_Raw2))
#rename column names for better readability, filter out unnesessary data, and combine data.
armyRank_clean <- armyRank_Raw2 %>%
  rename(
    pay_grade = Pay.Grade,
    army = Ranks.by.Branch.of.Service,
    navy = Ranks.by.Branch.of.Service.1,
    marine = Ranks.by.Branch.of.Service.2,
    airforce = Ranks.by.Branch.of.Service.3,
    spaceforce = Ranks.by.Branch.of.Service.4,
    junk1 = Ranks.by.Branch.of.Service.5
  ) %>%
  select(-junk) %>%
  select(-junk1) %>%
  slice(2:25) %>%
  pivot_longer(
    cols = -pay_grade,
    names_to = "branch",
    values_to = "rank_titles"
  )
#Join the two cleaned data sets together
usarmyRanks_clean <- usArmy_clean %>%
  left_join(armyRank_clean, by = c("pay_grade", "branch")) %>%

```

```

mutate(
  total_soldiers = as.numeric(gsub(",", "", total_soldiers))
) %>%
filter(!is.na(total_soldiers), total_soldiers > 0)

# Step 4: Make single case data frame ----
## Duplicate group data to fit each count of soldiers
usArmy_individual <- usArmyRanks_clean %>%
  uncount(weights = total_soldiers)

##Turn Individual Data into Frequency Table
library(tidyr)
library(knitr)
library(janitor)
library(dplyr)

#filter army officers from the individual dataframe
army_officers <- usArmy_individual %>%
  filter(branch == "army") %>%
  filter(grepl("^0", pay_grade)) %>%
  count(rank_titles, gender) %>%
  pivot_wider(
    names_from = gender,
    values_from = n,
    values_fill = 0
  )
#create a table
kable(
  army_officers,
  caption = "Relationship Between Gender and Army Officers"
)
# | echo: false

#Popular Baby Names Into Trend Graph
library(tidyverse)
library(dcData)
library(ggplot2)
data("BabyNames")

#Select four popular names and create a new dataframe
names_clean <- BabyNames %>%
  filter(name %in% c("Kathryn", "Gavin", "Jenna", "Melinda"))

```

```

#use new dataframe to create a plot of the trend of names
names_clean %>%
  ggplot(
    mapping = aes(
      x = year,
      y = count,
      color = name
    )
  ) +
  geom_line () +
  theme_bw() +
  labs(
    title = "Populaity of 4 Names Over Time",
    x = "Years",
    y = "Total Number of People with Name",
    color = "Names of Spread"
  )
#Box PLOT Function Into a Graph

#create a function that finds the max volume of the paper box
get_max_vol <- function(x, side_length = 36, side_width = 48){
  volume <- (side_length - 2 * x) * (side_width - 2 * x) * (x)
  return (volume)
}

library(ggplot2)
#use function to create a plot of the function
ggplot(
  data.frame(h = c(0,5)), #create a dataframe to store the value h
  mapping = aes(x = h)
) +
  #get the graph of the function
  stat_function(fun = get_max_vol) +
  labs(
    title = "Maximum Volume of a Box",
    x = "Box Height (h)",
    y = "Volume"
  )

```