Activity #14 - A First QMD

Ryan Sulaiman

2025-11-10

Data Wrangling Code for Armed Forces Data

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
v dplyr     1.1.4      v readr     2.1.5
v forcats   1.0.0      v stringr   1.5.1
v ggplot2   4.0.0      v tibble    3.3.0
v lubridate 1.9.4      v tidyr     1.3.1
v purrr     1.1.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(rvest)
```

```
Attaching package: 'rvest'

The following object is masked from 'package:readr':

    guess_encoding
```

```r
library(googlesheets4)

# Scrape Rank Data
webRanks <- read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html") %>%
  html_elements(css = "table") %>%
  html_table()

rawRanks <- webRanks[[1]] # Extract the data frame of ranks

# Wrangle Rank Data
## Enter a value in the first cell (1, 1)
rawRanks[1, 1] <- "Type"
## Extract actual column headers
rankHeaders <- rawRanks[1, ]
## Apply headers as column names
names(rawRanks) <- rankHeaders[1,]
```

Warning: The `value` argument of `names<-()` must be a character vector as of tibble
3.0.0.

```r
## Remove redundant first row and last row
rawRanks <- rawRanks[-c(1, 26), ]

cleanRanks <- rawRanks %>%
  dplyr::select(!Type) %>% # Remove extra column
  pivot_longer(
    cols = !`Pay Grade`, # The improper name requires backticks
    names_to = "Branch",
    values_to = "Rank"
  ) %>%
  mutate(
    Rank = na_if(x = Rank, y = "--")
  )

# Load Armed Forces Data
gs4_deauth()
forcesHeaders <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bwo/e
  col_names = FALSE, # Turn off Column Names
  n_max = 3 # read only the first three rows
)
```

```
v Reading from "US Armed Forces (6/2024)".
v Range 'Sheet1'.
New names:
```

```r
rawForces <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bwo/e
  col_names = FALSE, # Turn off Column Names
  skip = 3, # Skip the first three rows
  n_max = 28, # Read only the next 28 rows; drops footer
  col_types = "c" # Tells R to read everything as character data
)
```

```
v Reading from "US Armed Forces (6/2024)".
v Range '4:10000000'.
New names:
```

```r
# Wrangle Armed Forces Data
## Create good column names
### Pattern is Pay Grade followed by 3 columns for each branch in the order
### Army, Navy, Marine Corp, Air Force, Space Force, and Total
branchNames <- rep( # Create three copies of each branch
  x = c("Army", "Navy", "Marine Corps", "Air Force", "Space Force", "Total"),
  each = 3
)
tempHeaders <- paste( # Combine branch with other headers
  c("", branchNames),
  forcesHeaders[3,],
  sep = "."
)

names(rawForces) <- tempHeaders

cleanForces <- rawForces %>%
  rename(Pay.Grade = `.Pay Grade`) %>%
  dplyr::select(!contains("Total")) %>% # Remove total columns
  filter(Pay.Grade != "Total Enlisted" &
           Pay.Grade != "Total Warrant Officers" &
           Pay.Grade != "Total Officers" &
           Pay.Grade != "Total") %>% # Remove total rows; see note below
  pivot_longer( # Reshape data
    cols = !Pay.Grade,
    names_to = "Branch.Sex",
```

```r
    values_to = "Frequency"
  ) %>%
  separate_wider_delim( # Separate branches and sex
    cols = Branch.Sex,
    delim = ".",
    names = c("Branch", "Sex")
  ) %>%
  mutate(
    Frequency = na_if(Frequency, y = "N/A*"), # Convert N/A* to missing
    Frequency = parse_number(Frequency) # Parse values as numbers
  )

# Merge Data Frames
key_forcesRanks <- left_join(
  x = cleanForces,
  y = cleanRanks,
  by = join_by(Pay.Grade == `Pay Grade`, Branch == Branch)
)

# Transform Group into Individual
key_individualRanks <- key_forcesRanks %>%
  filter(!is.na(Frequency)) %>% # Remove all cases with missing counts
  uncount(
    weights = Frequency
  )



# When removing the total rows, a more concise approach would be to use a logical
# test for global regular expression printing (grepl) to search for the pattern
# of "Total" within each cell of the Pay Grade column
altExample <- rawForces %>%
  rename(Pay.Grade = `.Pay Grade`) %>%
  filter(!grepl(pattern = "Total", x = Pay.Grade))

# Alternative reading of the Google Sheet that comes from reading through
# the documentation of read_sheet
rawForces <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bwo/
  col_names = FALSE,
  skip = 3,
  n_max = 28,
  na = c("N/A*") # Tells R to treat the N/A* as missing values
```

4

```
)
```

```
v Reading from "US Armed Forces (6/2024)".
v Range '4:10000000'.
New names:
```

Visualization for the Armed Forces

```
echo=FALSE
library(tidyverse)
library(rvest)
library(googlesheets4)
library(kableExtra)
```

```
Attaching package: 'kableExtra'
```

```
The following object is masked from 'package:dplyr':

    group_rows
```

```r
# Scrape Rank Data
webRanks <- read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html") %>%
  html_elements(css = "table") %>%
  html_table()

rawRanks <- webRanks[[1]] # Extract the data frame of ranks

# Wrangle Rank Data
## Enter a value in the first cell (1, 1)
rawRanks[1, 1] <- "Type"
## Extract actual column headers
rankHeaders <- rawRanks[1, ]
## Apply headers as column names
names(rawRanks) <- rankHeaders[1,]
## Remove redundant first row and last row
rawRanks <- rawRanks[-c(1, 26), ]

cleanRanks <- rawRanks %>%
```

```
  dplyr::select(!Type) %>% # Remove extra column
  pivot_longer(
    cols = !`Pay Grade`, # The improper name requires backticks
    names_to = "Branch",
    values_to = "Rank"
  ) %>%
  mutate(
    Rank = na_if(x = Rank, y = "--")
  )

# Load Armed Forces Data
gs4_deauth()
forcesHeaders <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bwo/e
  col_names = FALSE, # Turn off Column Names
  n_max = 3 # read only the first three rows
)
```

v Reading from "US Armed Forces (6/2024)".

v Range 'Sheet1'.

New names:
* `` -> `...1`
* `` -> `...2`
* `` -> `...3`
* `` -> `...4`
* `` -> `...5`
* `` -> `...6`
* `` -> `...7`
* `` -> `...8`
* `` -> `...9`
* `` -> `...10`
* `` -> `...11`
* `` -> `...12`
* `` -> `...13`
* `` -> `...14`
* `` -> `...15`
* `` -> `...16`
* `` -> `...17`
* `` -> `...18`
* `` -> `...19`

```
rawForces <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/1cn4iO-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bwo/e
  col_names = FALSE, # Turn off Column Names
  skip = 3, # Skip the first three rows
  n_max = 28, # Read only the next 28 rows; drops footer
  col_types = "c" # Tells R to read everything as character data
)
```

```
v Reading from "US Armed Forces (6/2024)".
v Range '4:10000000'.
New names:
```

```
# Wrangle Armed Forces Data
## Create good column names
### Pattern is Pay Grade followed by 3 columns for each branch in the order
### Army, Navy, Marine Corp, Air Force, Space Force, and Total
branchNames <- rep( # Create three copies of each branch
  x = c("Army", "Navy", "Marine Corps", "Air Force", "Space Force", "Total"),
  each = 3
)
tempHeaders <- paste( # Combine branch with other headers
  c("", branchNames),
  forcesHeaders[3,],
  sep = "."
)

names(rawForces) <- tempHeaders

cleanForces <- rawForces %>%
  rename(Pay.Grade = `.Pay Grade`) %>%
  dplyr::select(!contains("Total")) %>% # Remove total columns
  filter(Pay.Grade != "Total Enlisted" &
           Pay.Grade != "Total Warrant Officers" &
           Pay.Grade != "Total Officers" &
           Pay.Grade != "Total") %>% # Remove total rows; see note below
  pivot_longer( # Reshape data
    cols = !Pay.Grade,
    names_to = "Branch.Sex",
    values_to = "Frequency"
  ) %>%
  separate_wider_delim( # Separate branches and sex
    cols = Branch.Sex,
```

```r
    delim = ".",
    names = c("Branch", "Sex")
  ) %>%
  mutate(
    Frequency = na_if(Frequency, y = "N/A*"), # Convert N/A* to missing
    Frequency = parse_number(Frequency) # Parse values as numbers
  )


# Merge Data Frames
key_forcesRanks <- left_join(
  x = cleanForces,
  y = cleanRanks,
  by = join_by(Pay.Grade == `Pay Grade`, Branch == Branch)
)


# Transform Group into Individual
key_individualRanks <- key_forcesRanks %>%
  filter(!is.na(Frequency)) %>% # Remove all cases with missing counts
  uncount(
    weights = Frequency
  )



# When removing the total rows, a more concise approach would be to use a logical
# test for global regular expression printing (grepl) to search for the pattern
# of "Total" within each cell of the Pay Grade column
altExample <- rawForces %>%
  rename(Pay.Grade = `.Pay Grade`) %>%
  filter(!grepl(pattern = "Total", x = Pay.Grade))

# Alternative reading of the Google Sheet that comes from reading through
# the documentation of read_sheet
rawForces <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6fZ9PhGLUvbMBHlzqG4bwo/
  col_names = FALSE,
  skip = 3,
  n_max = 28,
  na = c("N/A*") # Tells R to treat the N/A* as missing values
)


v Reading from "US Armed Forces (6/2024)".
v Range '4:10000000'.
```

New names:

```r
kable(key_forcesRanks, caption = "Armed Forces Visualization")
```

Table 1: Armed Forces Visualization

Pay.Grade	Branch	Sex	Frequency	Rank
E1	Army	Male	7909	Private
E1	Army	Female	1322	Private
E1	Navy	Male	9051	Seaman Recruit
E1	Navy	Female	2681	Seaman Recruit
E1	Marine Corps	Male	7233	Private
E1	Marine Corps	Female	659	Private
E1	Air Force	Male	6564	Airman Basic
E1	Air Force	Female	1607	Airman Basic
E1	Space Force	Male	177	Specialist 1
E1	Space Force	Female	30	Specialist 1
E2	Army	Male	18763	Private
E2	Army	Female	3383	Private
E2	Navy	Male	10969	Seaman Apprentice
E2	Navy	Female	3603	Seaman Apprentice
E2	Marine Corps	Male	14688	Private First Class
E2	Marine Corps	Female	1604	Private First Class
E2	Air Force	Male	6562	Airman
E2	Air Force	Female	1698	Airman
E2	Space Force	Male	159	Specalist 2
E2	Space Force	Female	33	Specalist 2
E3	Army	Male	38802	Private First Class
E3	Army	Female	8169	Private First Class
E3	Navy	Male	23430	Seaman
E3	Navy	Female	7493	Seaman
E3	Marine Corps	Male	35047	Lance Corporal

Pay.Grade	Branch	Sex	Frequency	Rank
E3	Marine Corps	Female	3787	Lance Corporal
E3	Air Force	Male	34981	Airman First Class
E3	Air Force	Female	9891	Airman First Class
E3	Space Force	Male	797	Specialist 3
E3	Space Force	Female	210	Specialist 3
E4	Army	Male	81278	Corporal OR Specialist
E4	Army	Female	14619	Corporal OR Specialist
E4	Navy	Male	39241	Petty Officer Third Class
E4	Navy	Female	11855	Petty Officer Third Class
E4	Marine Corps	Male	28946	Corporal
E4	Marine Corps	Female	2942	Corporal
E4	Air Force	Male	52399	Senior Airman
E4	Air Force	Female	15324	Senior Airman
E4	Space Force	Male	598	Specialist 4
E4	Space Force	Female	134	Specialist 4
E5	Army	Male	55671	Sergeant
E5	Army	Female	11111	Sergeant
E5	Navy	Male	57238	Petty Officer Second Class
E5	Navy	Female	16254	Petty Officer Second Class
E5	Marine Corps	Male	21481	Sergeant
E5	Marine Corps	Female	2723	Sergeant
E5	Air Force	Male	42576	Staff Sergeant
E5	Air Force	Female	11038	Staff Sergeant
E5	Space Force	Male	901	Sergeant
E5	Space Force	Female	165	Sergeant
E6	Army	Male	50030	Staff Sergeant
E6	Army	Female	7432	Staff Sergeant
E6	Navy	Male	45749	Petty Officer First Class
E6	Navy	Female	9580	Petty Officer First Class

Pay.Grade	Branch	Sex	Frequency	Rank
E6	Marine Corps	Male	11667	Staff Sergeant
E6	Marine Corps	Female	1370	Staff Sergeant
E6	Air Force	Male	32242	Technical Sergeant
E6	Air Force	Female	6746	Technical Sergeant
E6	Space Force	Male	696	Technical Sergeant
E6	Space Force	Female	130	Technical Sergeant
E7	Army	Male	30367	Sergeant First Class
E7	Army	Female	4322	Sergeant First Class
E7	Navy	Male	18026	Chief Petty Officer
E7	Navy	Female	3098	Chief Petty Officer
E7	Marine Corps	Male	8191	Gunnery Sergeant
E7	Marine Corps	Female	760	Gunnery Sergeant
E7	Air Force	Male	18367	Master Sergeant OR First Sergeant
E7	Air Force	Female	4717	Master Sergeant OR First Sergeant
E7	Space Force	Male	517	Master Sergeant
E7	Space Force	Female	110	Master Sergeant
E8	Army	Male	9287	First Sergeant OR Master Sergeant
E8	Army	Female	1426	First Sergeant OR Master Sergeant
E8	Navy	Male	6500	Senior Chief Petty Officer
E8	Navy	Female	912	Senior Chief Petty Officer
E8	Marine Corps	Male	3559	First Sergeant OR Master Sergeant
E8	Marine Corps	Female	275	First Sergeant OR Master Sergeant
E8	Air Force	Male	3632	Senior Master Sergeant OR First Sergeant
E8	Air Force	Female	1133	Senior Master Sergeant OR First Sergeant
E8	Space Force	Male	100	Senior Master Sergeant
E8	Space Force	Female	38	Senior Master Sergeant
E9	Army	Male	2908	Sergeant Major OR Command Sergeant Major
E9	Army	Female	413	Sergeant Major OR Command Sergeant Major

Pay.Grade	Branch	Sex	Frequency	Rank
E9	Navy	Male	2518	Master Chief Petty Officer OR Fleet/Command Master Chief Petty Officer
E9	Navy	Female	323	Master Chief Petty Officer OR Fleet/Command Master Chief Petty Officer
E9	Marine Corps	Male	1518	Sergent Major OR Master Gunnery Sergeant
E9	Marine Corps	Female	83	Sergent Major OR Master Gunnery Sergeant
E9	Air Force	Male	1956	Chief Master Sergeant OR First Sergeant
E9	Air Force	Female	506	Chief Master Sergeant OR First Sergeant
E9	Space Force	Male	37	Chief Master Sergeant
E9	Space Force	Female	12	Chief Master Sergeant
W1	Army	Male	3817	Warrant Officer
W1	Army	Female	399	Warrant Officer
W1	Navy	Male	30	Warrant Officer
W1	Navy	Female	3	Warrant Officer
W1	Marine Corps	Male	483	Warrant Officer
W1	Marine Corps	Female	48	Warrant Officer
W1	Air Force	Male	NA	Warrent Officer
W1	Air Force	Female	NA	Warrent Officer
W1	Space Force	Male	NA	NA
W1	Space Force	Female	NA	NA
W2	Army	Male	5398	Chief Warrant Officer
W2	Army	Female	653	Chief Warrant Officer
W2	Navy	Male	615	Chief Warrant Officer
W2	Navy	Female	96	Chief Warrant Officer
W2	Marine Corps	Male	738	Chief Warrant Officer
W2	Marine Corps	Female	47	Chief Warrant Officer
W2	Air Force	Male	NA	Chief Warrent Officer
W2	Air Force	Female	NA	Chief Warrent Officer
W2	Space Force	Male	NA	NA

Pay.Grade	Branch	Sex	Frequency	Rank
W2	Space Force	Female	NA	NA
W3	Army	Male	2826	Chief Warrant Officer
W3	Army	Female	316	Chief Warrant Officer
W3	Navy	Male	734	Chief Warrant Officer
W3	Navy	Female	102	Chief Warrant Officer
W3	Marine Corps	Male	538	Chief Warrant Officer
W3	Marine Corps	Female	32	Chief Warrant Officer
W3	Air Force	Male	NA	Chief Warrent Officer
W3	Air Force	Female	NA	Chief Warrent Officer
W3	Space Force	Male	NA	NA
W3	Space Force	Female	NA	NA
W4	Army	Male	1532	Chief Warrant Officer
W4	Army	Female	149	Chief Warrant Officer
W4	Navy	Male	416	Chief Warrant Officer
W4	Navy	Female	42	Chief Warrant Officer
W4	Marine Corps	Male	265	Chief Warrant Officer
W4	Marine Corps	Female	11	Chief Warrant Officer
W4	Air Force	Male	NA	Chief Warrent Officer
W4	Air Force	Female	NA	Chief Warrent Officer
W4	Space Force	Male	NA	NA
W4	Space Force	Female	NA	NA
W5	Army	Male	514	Chief Warrant Officer
W5	Army	Female	47	Chief Warrant Officer
W5	Navy	Male	69	Chief Warrant Officer
W5	Navy	Female	5	Chief Warrant Officer
W5	Marine Corps	Male	103	Chief Warrant Officer
W5	Marine Corps	Female	8	Chief Warrant Officer
W5	Air Force	Male	NA	Chief Warrent Officer
W5	Air Force	Female	NA	Chief Warrent Officer

Pay.Grade	Branch	Sex	Frequency	Rank
W5	Space Force	Male	NA	NA
W5	Space Force	Female	NA	NA
O1	Army	Male	7153	Second Lieutenant
O1	Army	Female	2358	Second Lieutenant
O1	Navy	Male	5012	Ensign
O1	Navy	Female	1545	Ensign
O1	Marine Corps	Male	2292	Second Lieutenant
O1	Marine Corps	Female	370	Second Lieutenant
O1	Air Force	Male	4867	Second Lieutenant
O1	Air Force	Female	1897	Second Lieutenant
O1	Space Force	Male	387	Second Lieutenant
O1	Space Force	Female	137	Second Lieutenant
O2	Army	Male	10013	First Lieutenant
O2	Army	Female	3093	First Lieutenant
O2	Navy	Male	5640	Lieutenant Junior Grade
O2	Navy	Female	1769	Lieutenant Junior Grade
O2	Marine Corps	Male	3377	First Lieutenant
O2	Marine Corps	Female	574	First Lieutenant
O2	Air Force	Male	4993	First Lieutenant
O2	Air Force	Female	1889	First Lieutenant
O2	Space Force	Male	409	First Lieutenant
O2	Space Force	Female	160	First Lieutenant
O3	Army	Male	20694	Captain
O3	Army	Female	5739	Captain
O3	Navy	Male	14825	Lieutenant
O3	Navy	Female	4825	Lieutenant
O3	Marine Corps	Male	5298	Captain
O3	Marine Corps	Female	624	Captain
O3	Air Force	Male	15532	Captain

Pay.Grade	Branch	Sex	Frequency	Rank
O3	Air Force	Female	5362	Captain
O3	Space Force	Male	1032	Captain
O3	Space Force	Female	246	Captain
O4	Army	Male	12758	Major
O4	Army	Female	3002	Major
O4	Navy	Male	8085	Lieutenant Commander
O4	Navy	Female	2217	Lieutenant Commander
O4	Marine Corps	Male	3610	Major
O4	Marine Corps	Female	333	Major
O4	Air Force	Male	10045	Major
O4	Air Force	Female	3395	Major
O4	Space Force	Male	951	Major
O4	Space Force	Female	196	Major
O5	Army	Male	6969	Lieutenant Colonel
O5	Army	Female	1539	Lieutenant Colonel
O5	Navy	Male	5478	Commandar
O5	Navy	Female	1040	Commandar
O5	Marine Corps	Male	1809	Lieutenant Colonel
O5	Marine Corps	Female	138	Lieutenant Colonel
O5	Air Force	Male	7459	Lieutenant Colonel
O5	Air Force	Female	1804	Lieutenant Colonel
O5	Space Force	Male	620	Lieutenant Colonel
O5	Space Force	Female	124	Lieutenant Colonel
O6	Army	Male	3084	Colonel
O6	Army	Female	588	Colonel
O6	Navy	Male	2613	Captain
O6	Navy	Female	421	Captain
O6	Marine Corps	Male	660	Colonel
O6	Marine Corps	Female	49	Colonel

Pay.Grade	Branch	Sex	Frequency	Rank
O6	Air Force	Male	2658	Colonel
O6	Air Force	Female	544	Colonel
O6	Space Force	Male	203	Colonel
O6	Space Force	Female	33	Colonel
O7	Army	Male	87	Brigadier General
O7	Army	Female	18	Brigadier General
O7	Navy	Male	90	Rear Admiral (Lower)
O7	Navy	Female	7	Rear Admiral (Lower)
O7	Marine Corps	Male	37	Brigadier General
O7	Marine Corps	Female	2	Brigadier General
O7	Air Force	Male	91	Brigadier General
O7	Air Force	Female	14	Brigadier General
O7	Space Force	Male	11	Brigadier General
O7	Space Force	Female	1	Brigadier General
O8	Army	Male	100	Major General
O8	Army	Female	7	Major General
O8	Navy	Male	67	Rear Admiral (Upper)
O8	Navy	Female	3	Rear Admiral (Upper)
O8	Marine Corps	Male	30	Major General
O8	Marine Corps	Female	2	Major General
O8	Air Force	Male	68	Major General
O8	Air Force	Female	8	Major General
O8	Space Force	Male	9	Major General
O8	Space Force	Female	0	Major General
O9	Army	Male	41	Lieutenant General
O9	Army	Female	8	Lieutenant General
O9	Navy	Male	32	Vice Admiral
O9	Navy	Female	4	Vice Admiral
O9	Marine Corps	Male	18	Lieutenant General

Pay.Grade	Branch	Sex	Frequency	Rank
O9	Marine Corps	Female	0	Lieutenant General
O9	Air Force	Male	33	Lieutenant General
O9	Air Force	Female	6	Lieutenant General
O9	Space Force	Male	4	Lieutenant General
O9	Space Force	Female	1	Lieutenant General
O10	Army	Male	12	General
O10	Army	Female	1	General
O10	Navy	Male	7	Admiral
O10	Navy	Female	1	Admiral
O10	Marine Corps	Male	3	General
O10	Marine Corps	Female	0	General
O10	Air Force	Male	12	General
O10	Air Force	Female	1	General
O10	Space Force	Male	3	General
O10	Space Force	Female	0	General

Narrative Text for the Armed Forces Section

The visualization displays a table that displays the data of branches of military, sexes of the personnel, pay-grade, and frequency. There are ratios between men and women, however the sex and ranks are independent of each other. There is no causation between the two for their differences. With this table, their are insights that are highlighted within. Such as, the Army consistently holds the largest branch, while the Air Force and Space Force yield the last. Showing the differences across the services, the data provides context for military composition changes.

Popularity of Baby Names

Code for the Popular Baby Names Project

```
library(ggplot2)
library(dplyr)
library(tidyverse)
BabyNames <- read.csv("BabyNames.csv")
instructors <- c("Mabel", "Bessie", "Catherine", "Elizabeth")

subsetNames <- BabyNames %>%
            filter(name %in% instructors) %>%
            group_by(name, year) %>%
            summarize(
              total = sum(count),
              .groups = "drop"
            )
```

Visualization for Popular Baby Names Project

```
echo=FALSE

library(ggplot2)
library(dplyr)
library(tidyverse)
data("BabyNames", packages="dcData")
```

```
Warning in data("BabyNames", packages = "dcData"): data set 'BabyNames' not
found
```

```
Warning in data("BabyNames", packages = "dcData"): data set 'dcData' not found
```
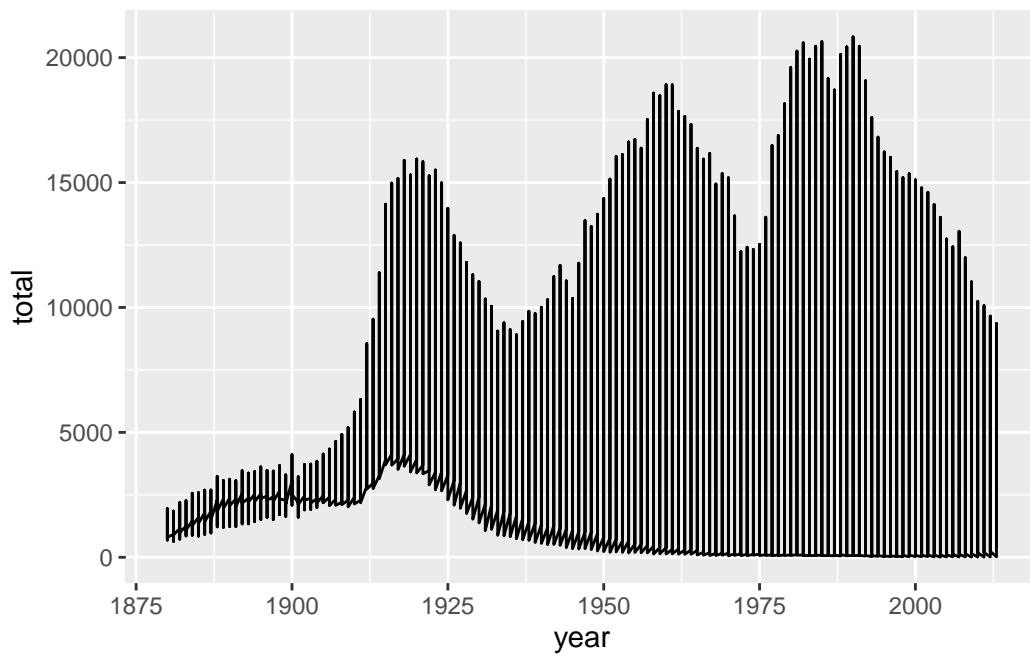
```
instructors <- c("Mabel", "Bessie", "Catherine", "Elizabeth")

subsetNames <- BabyNames %>%
            filter(name %in% instructors) %>%
            group_by(name, year) %>%
            summarize(
```

```
                total = sum(count),
                .groups = "drop"
            )
ggplot(subsetNames, mapping = aes(x = year, y = total)
        ) + geom_line()
```



```
psuPalette <- c("#1E407C", "#BC204B", "#3EA39E", "#E98300",
                "#999999", "#AC8DCE", "#F2665E", "#99CC00")

ggplot(subsetNames, mapping = aes(
  x = year,
  y = total,
  color = name,
  linetype = name
  )
       ) + geom_line(linewidth = 0.75) +
  labs(
    title = "Popularity of the Stat 184 Instructor's First Names Over Time",
    x = "Year",
    y = "Total Number of People with Name",
    color = "Name",
    linetype = "Name"
```
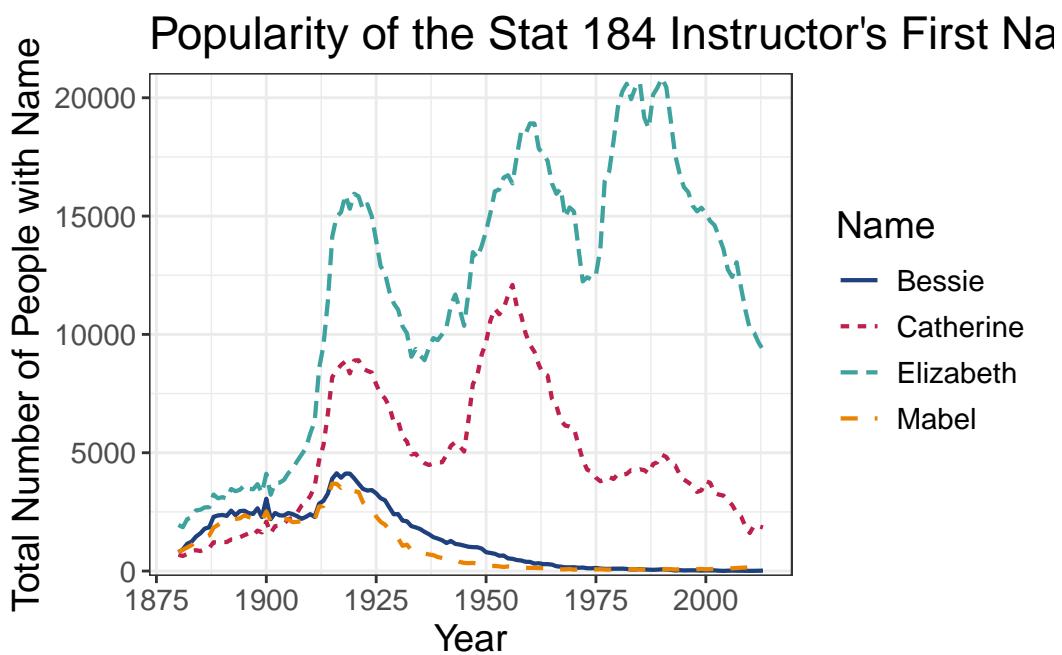
```
) + scale_y_continuous(
  expand = expansion(mult = 0.01)
) + scale_color_manual(
  values = psuPalette
) +
theme_bw() +
theme(
  text = element_text(size = 14),
  legend.kermy.size = unit(1, "cm")
)
```

Warning in plot_theme(plot): The `legend.kermy.size` theme element is not
defined in the element hierarchy.

Narrative Text for the Popular Baby Names Project

Alt Text: Displaying several line graphs on one graph to show the total number of people with
a name, respective to the year.

Long Description: This plot has values for a given names and years. There are 4 names that
are being used and the years that name has been given. The graph demonstrates the amount

of times the name has been used over time. The different lines indicate the different names based on the legend. There are different growth rates of the lines through the years, with names bouncing back in popularity, steady decreases, and many other rates.

Brief Paragraph: This is a data visualization that has the names of how many people had that specific name for a given year. Creating a line graph for each name, showing the growth of the name throughout the years.

Plotting a Mathematical Function

Code

```
volume.box <- function(x){
    vol <- (8.5-2*x)*(11-2*x)*(x)
    return(vol)
}
```
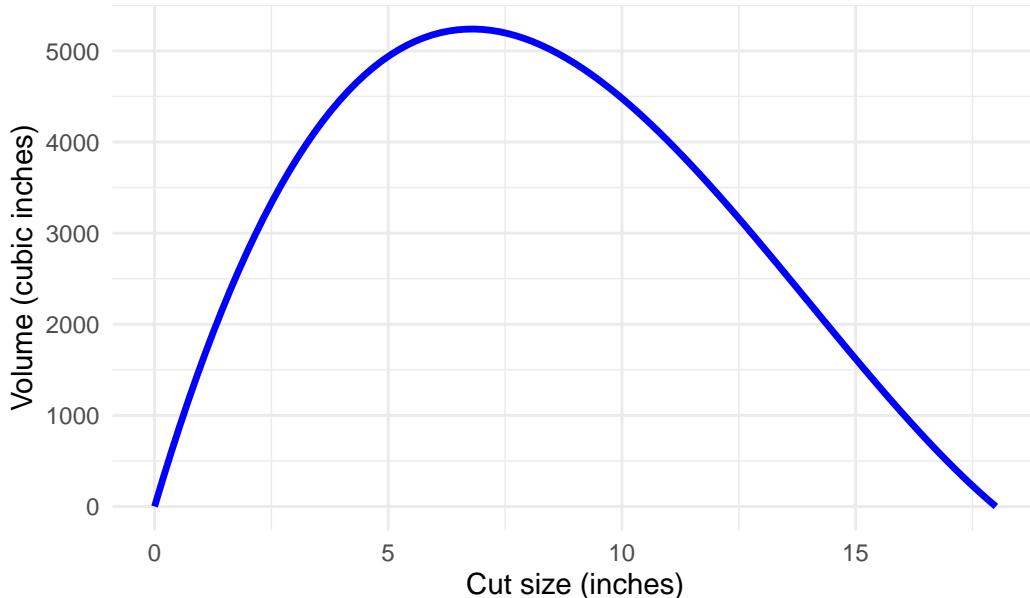
Visualization

```
echo=FALSE
# Define the volume function
box_volume <- function(x, length = 48, width = 36) {
  volume <- x * (length - 2 * x) * (width - 2 * x)
  return(volume)
}

# Load ggplot2
library(ggplot2)

# Create the plot
box_plot <- ggplot(data.frame(x = c(0, 18)), aes(x = x)) +
  stat_function(fun = box_volume, args = list(length = 48, width = 36),
                color = "blue", size = 1.2) +
  labs(
    title = "Box Problem: Volume vs. Cut Size",
    x = "Cut size (inches)",
    y = "Volume (cubic inches)"
  ) +
  theme_minimal()
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

```
print(box_plot)
```

Box Problem: Volume vs. Cut Size


```
# Graph of box problem displays volume function for a 36 by 48 inch sheet of paper. Showing
```

Narrative Text for the Box Problem

From this visualization, it displays the volume in cubic inches, to the cut size in inches. The max volume is the peak of the parabolic function, the y axis value. Finding the cut size at the maximum as well for the x axis. Looking for the size of the cuts can find the size of the other volumes if looking for specifics, due to limits.

What You've Learned So Far

There are several meaningful concepts that I have learned throughout the Stat 184 course. For example, how to use R functions and its installable packages. To begin, I learned how to tidy the data. During activity #05, learning how to get specific data values that are useful and taking out data that I do not need. Also, data wrangling, collecting raw data and cleaning up several factors. Handling missing values, removing duplicates, and structuring data into tables.

Learning how to make data plots. Using geometric plots to make visualizations. Throughout all the activities, I have utilized these functions and concepts to answer the questions.