Activity 14

Ellie Wallace

2025-11-07

0.1 Armed Forces Data Wrangling Redux

Table 1: Table of Enlisted Army Members by Gender

rank	Female	Male	Total
Corporal OR Specialist	15,143 (4.26%)	79,234 (22.30%)	94,377 (26.56%)
First Sergeant OR Master Sergeant	1,472 (0.41%)	9,482 (2.67%)	10,954 (3.08%)
Private	5,662 (1.59%)	29,767 (8.38%)	35,429 (9.97%)
Private First Class	10,229 (2.88%)	43,775 (12.32%)	54,004 (15.20%)
Sergeant	10,954 (3.08%)	54,803 (15.42%)	65,757 (18.51%)
Sergeant First Class	4,410 (1.24%)	30,264 (8.52%)	34,674 (9.76%)
Sergeant Major OR Command Sergeant Major	394 (0.11%)	2,865 (0.81%)	3,259 (0.92%)
Staff Sergeant	7,363 (2.07%)	49,502 (13.93%)	56,865 (16.00%)
Total	55,627 (15.66%)	299,692 (84.34%)	355,319 (100.00%)

Table of Enlisted Army Members

Table 1 shows the comparison of sex and rank for all enlisted positions in the army. The rank columns shows different ranks for all enlisted members, such as Private, Sargent, and Staff Sargent. The following rows show the proportion of males and females that hold these positions, followed by the total. Both the number and percentages are included in this table. In every single category, males hold a higher percentage of the roles, outnumbering females every time. Overall, males hold about 84% of all roles. while females hold the other 16%.

0.2 Popularity of Baby Names

Figure 1: Baby Names over Time
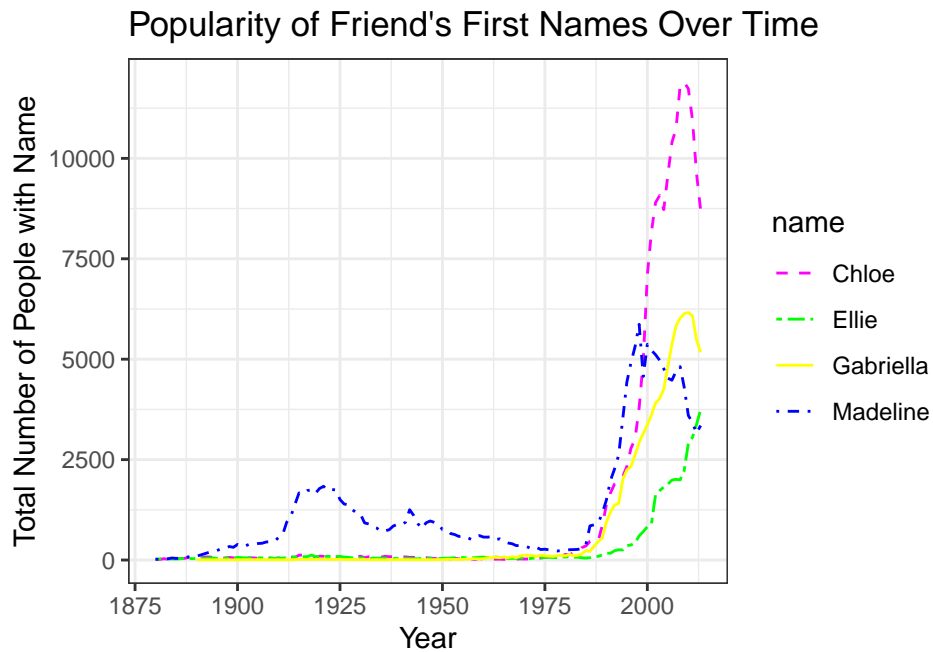
Popularity of Friend's First Names Over Time

Figure 1 displays a line graph titled "Popularity of Friend's First Names Over Time". The y-axis displays the total number of people with a name, ranging from 0 to 10000 and the x-axis shows the year, ranging from 1875 to 2020. There are four names shown, Chloe with a magenta dashed line, Ellie with a green dot-dash line, Gabriella with a solid yellow line , and Madeline with a blue dash-dot line. This information is located in a legend on the right. All of the lines are relatively low until about 1975 when the numbers spike, specifically for Chloe and Madeline. The names were chosen because they are the names of my friends who live in my building.

0.3 Plotting a Mathematical Function

Figure 2: Graph of Box Volume vs Cut Out Size in Inches
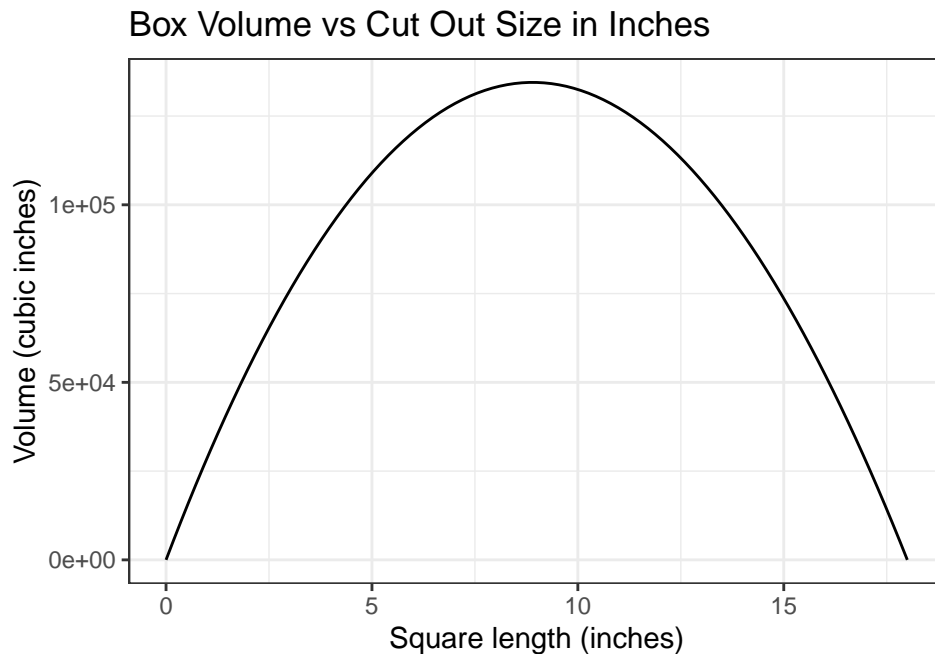
Box Volume vs Cut Out Size in Inches

Figure 2 shows a line graph that plots the volume of a box, in cubic inches, by the cut out size, in inches. The Y axis, shows the volume, ranging from 0 to 120,000 cubic inches, the x axis shows the square length, ranging from 0 to 20 inches. The curve begins at the origin, and increases, hitting a peak at about 10 inches and 10,000 cubic inches. It then symmetrically decreases back down.

0.4 What I've Learned So Far

So far in this course, I have learned a wealth of knowledge on how to operate R. I've learning how to access data sets, and tidy them to make them easier to understand. I have also learned how to create functions to do a specific task. On top of this, I have learned how to calculate important information from data sets, such as the 5 number summary. With all if this data that I have worked with, I have been able to create data visualizations that enhance understandings of data, such as line and bar graphs, as well as frequency tables. Besides the work in R, this course has taught me the value of planning when embarking on an assignment and the proper way carry out those steps. I have also learned the importance of revision, so that the final submission of any assignment is adequate. These skills are applicable not just to the Stat 184 course, but to the rest of my college courses and future career.

#Code Appendix

```
#Armed Forces Data where group is a case
library(tidyverse)
library(dplyr)
```

```r
# install.packages("rvest")
library(rvest)
# install.packages('googlesheets4')
library(googlesheets4)
#download the nessecary packages

gs4_deauth() #deauthorizes google sheet
#create new data frame and imports data from sheet
Armed_ForcesRaw1 <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5qbwb_E/ed
  na = c("", "N/A*"), #eliminates NAs
  col_names = TRUE,
  skip = 2
)

af_clean_data <- Armed_ForcesRaw1 %>% #creates new data frame for cleaned data
  select(!contains("Total")) %>% #select all columns that don't contain "Total"
  rename( #rename all of the columns
    pay_grade = 'Pay Grade',
    Army_Male = 'Male...2',
    Army_Female = 'Female...3',
    Navy_Male = 'Male...5',
    Navy_Female = 'Female...6',
    Marines_Male ='Male...8',
    Marines_Female = 'Female...9',
    Air.Force_Male = 'Male...11',
    Air.Force_Female = 'Female...12',
    Space.Force_Male = 'Male...14',
    Space.Force_Female = 'Female...15',
    junk1 = 'Male...17',
    junk2 = 'Female...18'
  ) %>%
  dplyr::select( #remove junk columns
    -junk1,
    -junk2
  ) %>%
  filter(!pay_grade %in% c("Pay Grade", "", "Total", "Total Enlisted",
"Total Warrant Officers", "Total Officers",
"Source: DMDC Active-Duty Military Personnel Master File (June 2025)")
#filter by pay grade
  ) %>%
  pivot_longer( #split up the branch and gender columns into just a count
    cols = 2:11,
    names_to = "branch_gender",
    values_to = "count"
  ) %>%
  separate_wider_delim(
```

```r
#separate the branch and gender column by _ and place values into thier own seperate columns
    cols = branch_gender,
    delim = "_",
    names = c("branch", "gender")
  ) %>%
  mutate(
    branch = gsub("\\.", " ", branch) #elimates any . in data and replaces it with a space
  )

#Tidying rank data
#creating a new data fram for ranks data read_html
afRanks_raw <-read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html") %>%
  html_elements(css = "table") %>%
  html_table()
#View(afRanks_raw) #choosing what parts are being kept after scraping
afRanks_raw <- afRanks_raw[[1]]

names(afRanks_raw) <- make.unique(names(afRanks_raw))
names(afRanks_raw)[names(afRanks_raw) == ""] <- "junk"

afRanks_clean_data <- afRanks_raw %>% #making tidy data frame
  rename( #renaming columns to make proper
    pay_grade = 'Pay Grade',
    Army = 'Ranks by Branch of Service',
    Navy = 'Ranks by Branch of Service.1',
    Marines = 'Ranks by Branch of Service.2',
    Air.Force = 'Ranks by Branch of Service.3',
    Space.Force = 'Ranks by Branch of Service.4',
    junk9 = 'Ranks by Branch of Service.5'
  ) %>%
  dplyr::select( #removing junk
    -junk,
    -junk9
  ) %>%
  filter(pay_grade != "Pay Grade") %>%
  filter(pay_grade !=
  "Note: -- indicates that a pay grade is not currently used by a service branch") %>%
  #filter unnecessary columns
  pivot_longer( #expanding columns longer
    cols = 2:6,
    names_to = "branch",
    values_to = "rank"
  ) %>%
  mutate( #eliminating . and making it a space
    branch = gsub("\\.", " ", branch)
  ) %>%
  mutate( #treating all of the types of NAs as character
```

```r
    rank = as.character(gsub("--","NA", rank))
  ) %>%
  left_join( #joining the two frames together
    x = af_clean_data,
    y = .,
    by = join_by(pay_grade == pay_grade, branch == branch)
  ) %>%
  mutate(rank = na_if(rank, "NA"))

#Person is a case
afRanks_clean_person <- afRanks_clean_data %>% #making new data frame for a person as a case
  filter(!is.na(count)) %>% #eliminating NAs
  uncount(
    weights = count
  )

#install.packages("janitor")
library(janitor)
#install.packages("knitr")
library(knitr)
#install.packages("kableExtra")
library(kableExtra)

#Filter for navy and enlisted
army_enlisted <- afRanks_clean_person %>% #making a new data frame for the specific group
  filter(branch == "Army") %>% #filter by the branch
  filter(str_detect(pay_grade, "E")) #filtering by rank

#Creating the 2-way table
army_table <- army_enlisted %>% #making the new data frame for the table
  tabyl(rank, gender) %>%
  adorn_totals(where = c("row", "col")) %>%
  adorn_percentages(denominator = "all") %>%
  adorn_pct_formatting(digits = 2)
#adding table features

 army_freqtable <- army_table %>% #creating the frequency table
  adorn_ns(
    position = "front",
    format_func = function(x) {
      format(x, big.mark = ",")
    }
  )
army_freqtable %>% #added polished features to the table
  kable(
    caption = "Table of Enlisted Army Members by Gender",
    align = c("l", rep("c", 6))
```

```r
) %>%
kable_classic(
  font_size = 10,
  lightable_options = "striped"
)




#Baby Names Data
#Step 1: Load packages needed for making plots, tidying and data
library(dcData)
library(ggplot2)
library(tidyverse)
library(dplyr)
data("BabyNames")

#Step 2: Tidy the data
BabyNames_tidy <- BabyNames %>% #create a new data frame
#filter out every other name besides the ones chosen
  filter(name %in% c("Chloe", "Gabriella", "Madeline", "Ellie")) %>%
  filter(sex %in% "F") #filter out all of the male names so it's just females

#Step 3: Create the visual
ggplot( #set the framework
  data = BabyNames_tidy, #choose the tidied data frame
  mapping = aes( #set the variables
    x = year,
    y = count,
    color = name,
    linetype = name
  )
) +
  geom_line() + #create the line plot
  labs(
    title = "Popularity of Friend's First Names Over Time", #Add the title
    x = "Year", #label the variables
    y = "Total Number of People with Name"
  ) +
  #edit the colors of the lines for each name
  scale_color_manual(values = c("Chloe" = "magenta", "Ellie" = "green",
  "Gabriella" = "yellow", "Madeline" = "blue")) +
  #change the type of line for each name
  scale_linetype_manual(values = c("Chloe" = "dashed", "Ellie" = "twodash",
  "Gabriella" = "solid", "Madeline" = "dotdash")) +
  theme_bw() #set the theme
```

```r
#Box Volume function
#creating the function
box_Problem <- function(
    square_length,
    paper_width = 36,
    paper_length = 848)
  { #adding dimensions
volume <- (paper_length - 2 * square_length) *
  (paper_width - 2 *square_length) *
  square_length #creating the formula
return(volume) #adding what is to be returned in the function
}

f <- function(x) {
  box_Problem(x, paper_width = 36, paper_length = 848)
} #creating a function with the box problem so it works bettwe with the geom functions

#creating the graph
ggplot() +
  geom_function(fun = f) + #adding in the function
  xlim(0,18) + #setting bounds
  labs(
    title = "Box Volume vs Cut Out Size in Inches",
    x = "Square length (inches)",
    y = "Volume (cubic inches)",
    alt = "Graph showing box volume vs cut out size in inches"
    ) + #adding labels
  theme_bw() #adding a theme
```