Activity #15

Various data visualizations and what I have learned from Stat 184

Kyle Barber

2025-11-19

Table of contents

Frequency table of each sex in a specific rank when accounting for all enlisted soldiers in the Army

Table 1: Frequency table of each sex in a specific rank when accounting for all enlisted soldiers in the Army

Ranks	Female	Male	Total
Corporal OR Specialist	4.26%	22.30%	26.56%
First Sergeant OR Master Sergeant	0.41%	2.67%	3.08%
Private	1.59%	8.38%	9.97%
Private First Class	2.88%	12.32%	15.20%
Sergeant	3.08%	15.42%	18.51%
Sergeant First Class	1.24%	8.52%	9.76%
Sergeant Major OR Command Sergeant Major	0.11%	0.81%	0.92%
Staff Sergeant	2.07%	13.93%	16.00%
Total	15.66%	84.34%	100.00%

Figure 1: Frequency table of each sex in a specific rank when accounting for all enlisted soldiers in the Army

Figure 1 shows a frequency table showing what percent each sex represents in each rank of enlisted soldiers in the Army.

Popularity of Baby Names Plot

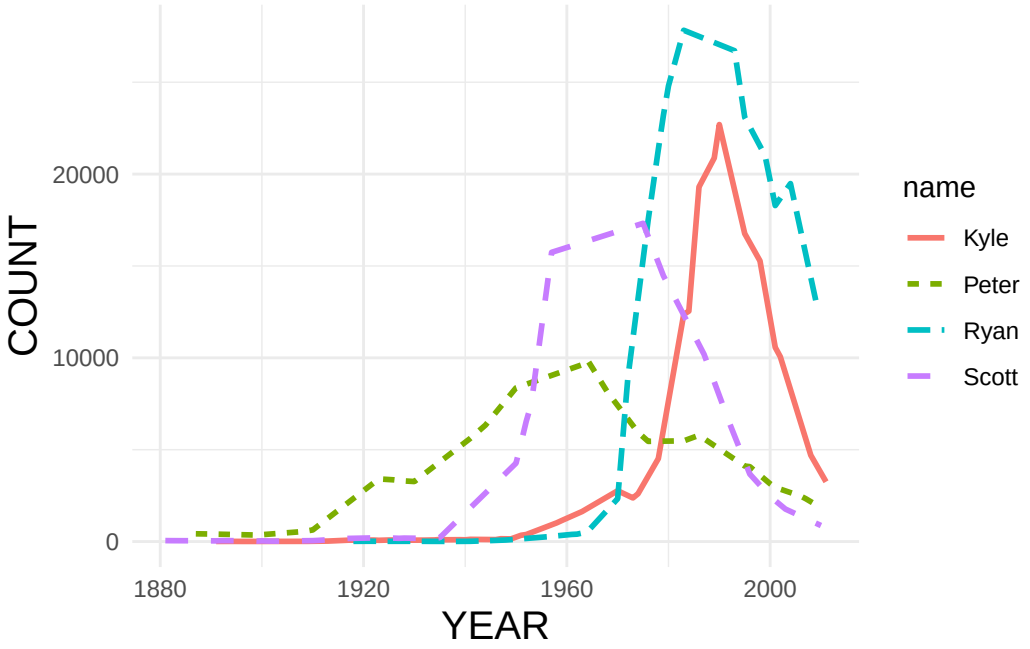

Figure 2: Plot of popularity of various baby names over time

Figure 2 demonstrates the popularity of the names: Kyle, Ryan, Scott and Peter. I chose these names because I was curious to see the popularity of the names of myself (Kyle) and my friends (Ryan, Scott and Peter) over time.

Code for boxvolume function and plot of the maximum box volume vs height

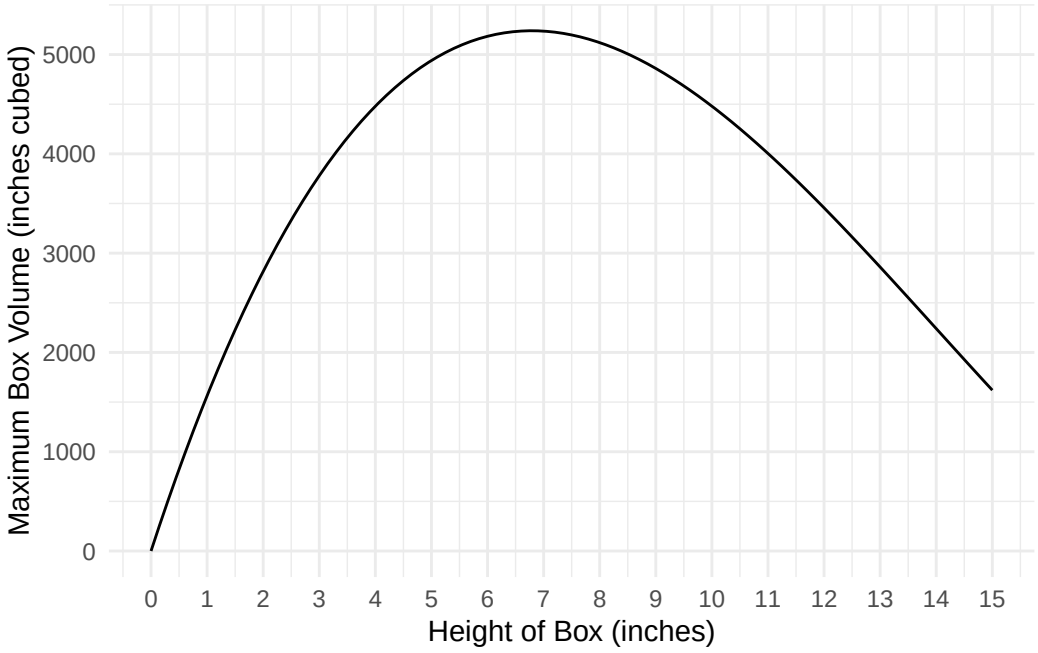

Figure 3: Plot of maximum box volume vs height using my boxvolume function

Figure 3 displays the maximum volume of a box over various height values from 0 to 15. This graph calls a box volume function that calculates the volume of a box when given the length width and height.

What I have learned from Stat 184

Stat 184 has been a very beneficial course that has made me a better statistician. It has helped me improve my skills in working with data. It has also changed the way I think about data and approaching problems. I have learned a lot about the importance and power of data and data visualizations, RStudio and the vast and inclusive world of open science.

This class has helped show me the wonderful world of data. I learned that there is data on almost everything in our world. Also, I learned that any data set, no matter how big or small, is able to be worked with. Before taking this course I was not prepared to work with extremely large or intimidating data sets. Now I understand, that was an irrational fear and I am comfortable working with any data set despite the size or contents. Also, I have learned that data can be used to solve almost any problem. Data is not just used to find statistics and trends in numbers but can pinpoint origins of disease and help the world's best engineers reach space. I learned that data is not just numbers and words in a chart. Data can include maps, diagrams and even art. This is inspiring and further shows the variety that data comes in and how all data can be used to solve issues.

Another major thing I have learned in this class is RStudio. Stat 184 is centered around RStudio and before taking this course I had little to no experience with this language. I have learned the basics of the language and how to upload and work with data in a console and environment. Also, I learned the organization of R and many of its packages and functions. Learning RStudio has helped me become comfortable with data sets and has shown me that all types of data can be worked with. Also, RStudio has allowed me to improve as a statistician. This class has taught me how to create various data visualizations and their importance. Creating data visualization is a very important skill and is used every day to solve world issues. Data visualizations serve as a way for statisticians to display their ideas in a way that is easily accessible and understood by all.

Lastly, I have learned about open science. This has helped me improve in the way I approach and tackle problems. I have learned to carefully map out plans and ideas before executing. Also, this has helped me understand what to do when faced with any problem, not just problems including data. I have also learned about the community of open science. Open science is a family with its sole purpose to make everyone a better scientist and problem solver. Although Stat 184 has taught me a lot it has not made me a perfect statistician. Open science serves as a way for me to learn from others and enhance my skills even when I feel I do not know where to go. I have learned of many beneficial sites like Github and R Markdown Gallery that display the work of other data scientists.

Stat 184 has taught me a lot about working with data, RStudio and others. It has helped me improve my data science skills in many ways. It has laid a foundation that I am excited to build off of and better myself in the world of open science and data.

Code Appendix

```
# CODE FOR MILITARY FREQUENCY TABLE
# 1. Load necessary packages: tidyverse, rvest, googlesheets4.
library(tidyverse)
library(rvest)
library(googlesheets4)
library(dplyr)
library(janitor)
library(kableExtra)

# 2. Scrape the Pay Grade and Rank dataset.
PayGrade_rank <- read_html("https://neilhatfield.github.io/Stat184_PayGradeRanks.html")%>%
  html_elements(css = "table")%>%
  html_table()
PayGrade_rank_raw <- PayGrade_rank[[1]]

# 3. Rename the headers of the Pay Grade and Rank dataset.
names(PayGrade_rank_raw) <- c("membership", "Pay_Grade", "Army", "Navy",
                              "Marine.Corps", "Air.Force", "Space.Force",
                              "Coast.Guard")

# 4. Wrangle the Pay Grade and Rank data set.
clean_PayGrade_rank <- PayGrade_rank_raw%>% ## get rid of the first column and row
  select(-membership)%>%
  slice(-1)%>% ## combine all columns besides Pay Grade to represent rank and branch.
  pivot_longer(
    cols = Army:Coast.Guard,
    names_to = "Branch",
    values_to = "Ranks"
  )%>% ## get rid of unnecessary rows
  slice(-(145:150))

# 4. Load in the Armed Forced data
gs4_deauth()
military_Raw <- read_sheet(
  ss = "https://docs.google.com/spreadsheets/d/19xQnI1cBh6Jkw7eP8YQuuicMlVDF7Gr-nXCb5qbwb_E/e
  col_names = TRUE,
  skip = 2,
  na = c("", "N/A*")
)
```

```r
# 5. Wrangle the Armed Forces data
military_clean <- military_Raw%>% ## get rid of all total columns and rows
  select(!contains("Total"))%>%
  slice(c(-10, -16, -27, -28, -29))%>%
  select(-"Male...17", -"Female...18")%>% ## Rename all male and female columns and provide
  rename(
    Army_Male = Male...2,
    Army_Female =  Female...3,
    Navy_Male = Male...5,
    Navy_Female = Female...6,
    Marine.Corps_Male = Male...8,
    Marine.Corps_Female = Female...9,
    Air.Force_Male = Male...11,
    Air.Force_Female = Female...12,
    Space.Force_Male = Male...14,
    Space.Force_Female = Female...15
  )%>% ## create new column combining all columns besides pay grade.
  pivot_longer(
    cols = Army_Male:Space.Force_Female,
    names_to = "Branch_Sex",
    values_to = "frequency"
  )%>% ## separate new column by name and branch.
  separate_wider_delim(
    cols = Branch_Sex,
    names = c("Branch", "Sex"),
    delim = "_"
  )%>% ## Add the ranks
  mutate(
    RankClass = case_when(
      grepl(pattern = "^E", x = `Pay Grade`) ~ "Enlisted",
      grepl(pattern = "^W", x = `Pay Grade`) ~ "Warrant Officer",
      grepl(pattern = "^O", x = `Pay Grade`) ~ "Officer"
    )
  )


# 6. Join tables
final_frequency_table <- left_join(
  x = military_clean,
  y = clean_PayGrade_rank,
  by = join_by(`Pay Grade` == Pay_Grade, Branch == Branch)
)
```

```r
# 7. Make case an individual soldier
final_table <- final_frequency_table %>%
  filter(!is.na(frequency)) %>% ## Remove all cases with missing counts
  uncount(
    weights = frequency
  )

# 8. Make the frequency data set
data <- final_table
Frequency_table <- data%>%
  filter(Branch == "Army" & RankClass == "Enlisted")%>% #Choose which branch and rank class t
  tabyl(Ranks, Sex)%>% #Choose what frequencies to use
  adorn_totals(where = c("row", "col"))%>% #set totals
  adorn_percentages(denominator = "all")%>% #set percentages
  adorn_pct_formatting(digits = 2) #set format

# 9. Make a frequency table
Frequency_table %>%
  kable( #set caption
    caption = "Frequency table of each sex in a specific rank when accounting for all enliste
    booktabs = TRUE,
    align = c("l", rep("c", 6))
  ) %>%
  kable_styling( #set style
    bootstrap_options = c("striped", "condensed"),
    font_size = 16
  )
# CODE FOR POPULARITY OF BABY NAMES
# 1. Load necessary packages
library(tidyverse)
library(ggplot2)
library(dplyr)
library(dcData)
data(BabyNames)
# 2. Filtering our specific names
Baby_Name <- BabyNames%>%
  filter(name == c("Kyle", "Peter", "Scott", "Ryan"), sex == "M")
# 2. Creating plot
ggplot(Baby_Name) +
  ## Setting the aesthetic
  aes(x = year, y = count, colour = name, linetype = name) +
  geom_line(size = 1.0) +
```

```
  scale_color_hue(direction = 1) +
  ## Labeling plot
  labs(
    x = "YEAR",
    y = "COUNT",
  ) +
  ## Setting theme of plot and size of lines
  theme_minimal() +
  theme(
    plot.title = element_text(size = 20L,
                              hjust = 0.5),
    axis.title.y = element_text(size = 15L),
    axis.title.x = element_text(size = 15L)
  )
# CODE FOR BOX PROBLEM
# 1. Load necessary packages
library(ggplot2)
# 2. Function fo find the volume of a box with given length width and height.
boxvolume <- function(length, width, height){
  newLength = length - 2 * height
  newWidth = width  - 2*height
  volume = newLength*newWidth*height
  return(volume)
}

# 3. Plot of maximum box volume vs height of box using my boxvolume function
ggplot()+
  stat_function( ## calling the function
    fun = boxvolume,
    args = list(length = 36, width = 48),
    xlim = c(0, 15)
  ) +
  scale_x_continuous( ## set parameters of the x-axis
    breaks = seq(from = 0, to = 15, by = 1)
  ) +
  labs( ## labeling the plot
  x = "Height of Box (inches)",
  y = "Maximum Box Volume (inches cubed)"
) +
  theme_minimal() ## set the theme of the plot
```