ACTIVITY14

Rishita Shukla

2025-11-14

Armed Forces Data Wrangling Redux (Activities #08 and #10)

Table 1: Group Absolute Freq

ServiceBranch	...10	...11	...12	...13	...14	...15	...16	...17	...18	
E1	7892	6564	1607	8171	177	30	207	30934	6299	3
E2	16292	6562	1698	8260	159	33	192	51141	10321	6
E3	38834	34981	9891	44872	797	210	1007	133057	29550	16
E4	31888	52399	15324	67723	598	134	732	202462	44874	24
E5	24204	42576	11038	53614	901	165	1066	177867	41291	21
E6	13037	32242	6746	38988	696	130	826	140384	25258	16
E7	8951	18367	4717	23084	517	110	627	75468	13007	8
E8	3834	3632	1133	4765	100	38	138	23078	3784	2
E9	1601	1956	506	2462	37	12	49	8937	1337	1
O1	2662	4867	1897	6764	387	137	524	19711	6307	2
O10	3	12	1	13	3	0	3	37	3	
O2	3951	4993	1889	6882	409	160	569	24432	7485	3
O3	5922	15532	5362	20894	1032	246	1278	57381	16796	7
O4	3943	10045	3395	13440	951	196	1147	35449	9143	4
O5	1947	7459	1804	9263	620	124	744	22335	4645	2
O6	709	2658	544	3202	203	33	236	9218	1635	1
O7	39	91	14	105	11	1	12	316	42	
O8	32	68	8	76	9	0	9	274	20	
O9	18	33	6	39	4	1	5	128	19	
Total	168032	245037	67580	312617	7611	1760	9371	1030687	223774	125
Total Enlisted	146533	199279	52660	251939	3982	862	4844	843328	175721	101
Total Officers	19226	45758	14920	60678	3629	898	4527	169281	46095	21

Total Warrant Officers	2273	0	0	0	0	0	0	18078	1958	2
W1	531	0	0	0	0	0	0	4330	450	
W2	785	0	0	0	0	0	0	6751	796	
W3	570	0	0	0	0	0	0	4098	450	
W4	276	0	0	0	0	0	0	2213	202	
W5	111	0	0	0	0	0	0	686	60	
Total	504096	735111	202740	937851	22833	5280	28113	3092061	671322	376

Table 2: Group Relative Frequency Table

ServiceBranch	...10	...11	...12	...13	...14	...15	...16	...17	...18	...19	...2	...3	..
E1	0.05	0.04	0.01	0.05	0.00	0.00	0.00	0.21	0.04	0.25	0.05	0.01	0.0
E2	0.11	0.04	0.01	0.05	0.00	0.00	0.00	0.34	0.07	0.41	0.12	0.02	0.1
E3	0.26	0.23	0.07	0.30	0.01	0.00	0.01	0.88	0.20	1.08	0.26	0.05	0.3
E4	0.21	0.35	0.10	0.45	0.00	0.00	0.00	1.34	0.30	1.64	0.54	0.10	0.6
E5	0.16	0.28	0.07	0.36	0.01	0.00	0.01	1.18	0.27	1.46	0.37	0.07	0.4
E6	0.09	0.21	0.04	0.26	0.00	0.00	0.01	0.93	0.17	1.10	0.33	0.05	0.3
E7	0.06	0.12	0.03	0.15	0.00	0.00	0.00	0.50	0.09	0.59	0.20	0.03	0.2
E8	0.03	0.02	0.01	0.03	0.00	0.00	0.00	0.15	0.03	0.18	0.06	0.01	0.0
E9	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.06	0.01	0.07	0.02	0.00	0.0
O1	0.02	0.03	0.01	0.04	0.00	0.00	0.00	0.13	0.04	0.17	0.05	0.02	0.0
O10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
O2	0.03	0.03	0.01	0.05	0.00	0.00	0.00	0.16	0.05	0.21	0.07	0.02	0.0
O3	0.04	0.10	0.04	0.14	0.01	0.00	0.01	0.38	0.11	0.49	0.14	0.04	0.1
O4	0.03	0.07	0.02	0.09	0.01	0.00	0.01	0.24	0.06	0.30	0.08	0.02	0.1
O5	0.01	0.05	0.01	0.06	0.00	0.00	0.00	0.15	0.03	0.18	0.05	0.01	0.0
O6	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.06	0.01	0.07	0.02	0.00	0.0
O7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
O8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
O9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
Total	1.12	1.63	0.45	2.08	0.05	0.01	0.06	6.85	1.49	8.33	2.46	0.47	2.9
Total Enlisted	0.97	1.32	0.35	1.67	0.03	0.01	0.03	5.60	1.17	6.77	1.96	0.35	2.3
Total Officers	0.13	0.30	0.10	0.40	0.02	0.01	0.03	1.12	0.31	1.43	0.40	0.11	0.5
Total Warrant Officers	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.01	0.13	0.09	0.01	0.1
W1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.03	0.00	0.0
W2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.01	0.05	0.04	0.00	0.0
W3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.02	0.00	0.0
W4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.01	0.00	0.0
W5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0

Total	3.35	4.88	1.35	6.23	0.15	0.04	0.19	20.54	4.46	25.00	7.37	1.40	8.7

Table 3: Individual Absolute Fr

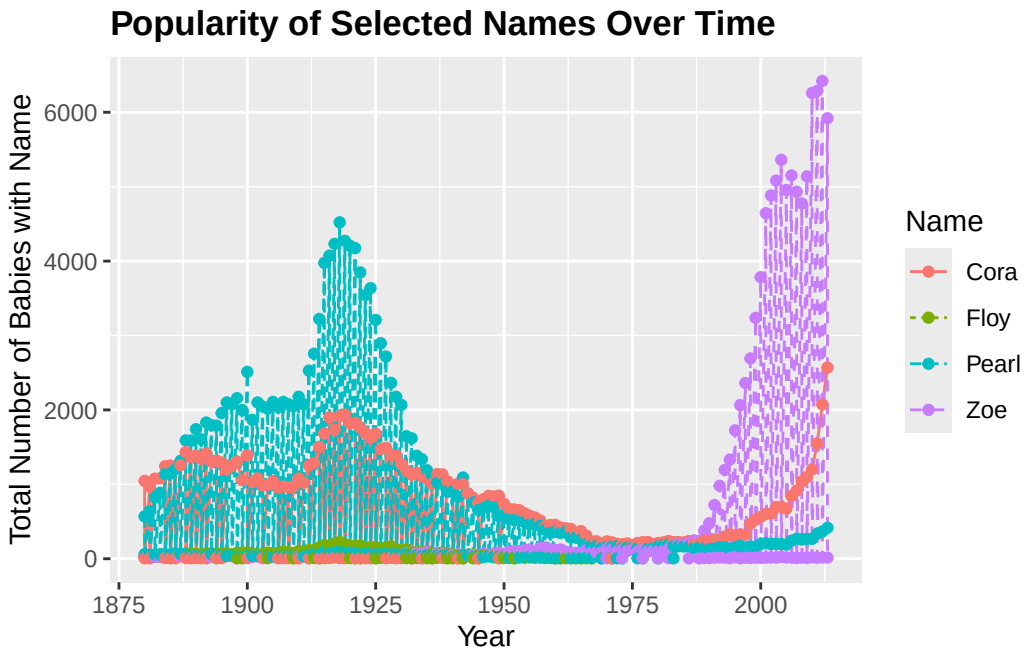
ServiceBranch	...10	...11	...12	...13	...14	...15	...16	...17	...18	
E1	7892	6564	1607	8171	177	30	207	30934	6299	3
E2	16292	6562	1698	8260	159	33	192	51141	10321	6
E3	38834	34981	9891	44872	797	210	1007	133057	29550	16
E4	31888	52399	15324	67723	598	134	732	202462	44874	24
E5	24204	42576	11038	53614	901	165	1066	177867	41291	21
E6	13037	32242	6746	38988	696	130	826	140384	25258	16
E7	8951	18367	4717	23084	517	110	627	75468	13007	8
E8	3834	3632	1133	4765	100	38	138	23078	3784	2
E9	1601	1956	506	2462	37	12	49	8937	1337	1
O1	2662	4867	1897	6764	387	137	524	19711	6307	2
O10	3	12	1	13	3	0	3	37	3	
O2	3951	4993	1889	6882	409	160	569	24432	7485	3
O3	5922	15532	5362	20894	1032	246	1278	57381	16796	7
O4	3943	10045	3395	13440	951	196	1147	35449	9143	4
O5	1947	7459	1804	9263	620	124	744	22335	4645	2
O6	709	2658	544	3202	203	33	236	9218	1635	1
O7	39	91	14	105	11	1	12	316	42	
O8	32	68	8	76	9	0	9	274	20	
O9	18	33	6	39	4	1	5	128	19	
Total	168032	245037	67580	312617	7611	1760	9371	1030687	223774	125
Total Enlisted	146533	199279	52660	251939	3982	862	4844	843328	175721	101
Total Officers	19226	45758	14920	60678	3629	898	4527	169281	46095	21
Total Warrant Officers	2273	0	0	0	0	0	0	18078	1958	2
W1	531	0	0	0	0	0	0	4330	450	
W2	785	0	0	0	0	0	0	6751	796	
W3	570	0	0	0	0	0	0	4098	450	
W4	276	0	0	0	0	0	0	2213	202	
W5	111	0	0	0	0	0	0	686	60	
Total	504096	735111	202740	937851	22833	5280	28113	3092061	671322	376

Table 4: Individual Relative Frequency Table

ServiceBranch	...10	...11	...12	...13	...14	...15	...16	...17	...18	...19	...2	...3	..

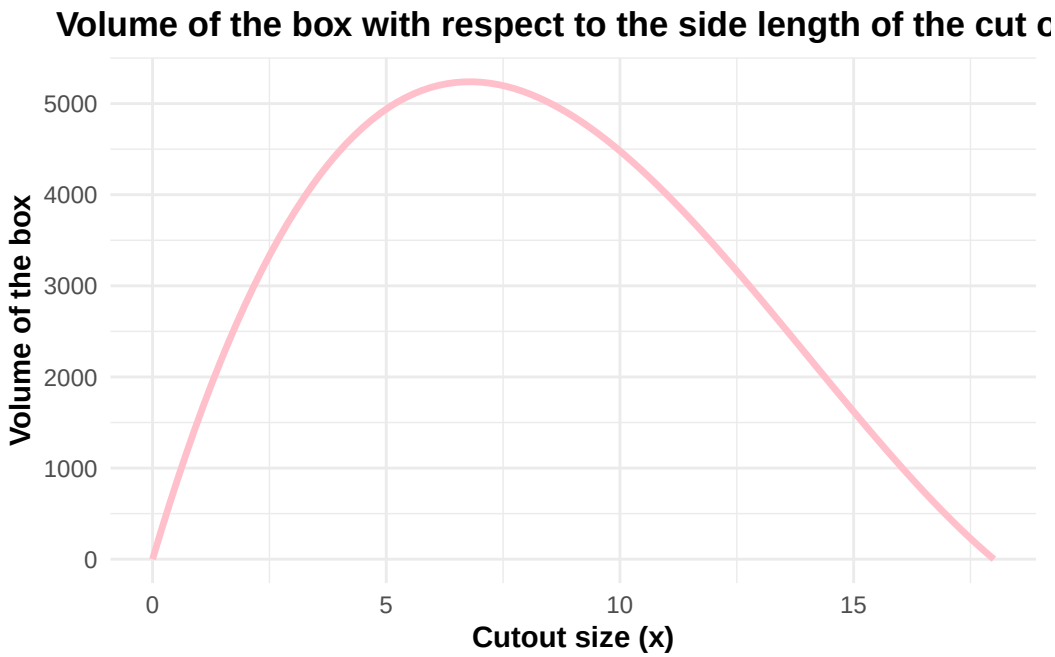
E1	0.05	0.04	0.01	0.05	0.00	0.00	0.00	0.21	0.04	0.25	0.05	0.01	0.0
E2	0.11	0.04	0.01	0.05	0.00	0.00	0.00	0.34	0.07	0.41	0.12	0.02	0.1
E3	0.26	0.23	0.07	0.30	0.01	0.00	0.01	0.88	0.20	1.08	0.26	0.05	0.3
E4	0.21	0.35	0.10	0.45	0.00	0.00	0.00	1.34	0.30	1.64	0.54	0.10	0.6
E5	0.16	0.28	0.07	0.36	0.01	0.00	0.01	1.18	0.27	1.46	0.37	0.07	0.4
E6	0.09	0.21	0.04	0.26	0.00	0.00	0.01	0.93	0.17	1.10	0.33	0.05	0.3
E7	0.06	0.12	0.03	0.15	0.00	0.00	0.00	0.50	0.09	0.59	0.20	0.03	0.2
E8	0.03	0.02	0.01	0.03	0.00	0.00	0.00	0.15	0.03	0.18	0.06	0.01	0.0
E9	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.06	0.01	0.07	0.02	0.00	0.0
O1	0.02	0.03	0.01	0.04	0.00	0.00	0.00	0.13	0.04	0.17	0.05	0.02	0.0
O10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
O2	0.03	0.03	0.01	0.05	0.00	0.00	0.00	0.16	0.05	0.21	0.07	0.02	0.0
O3	0.04	0.10	0.04	0.14	0.01	0.00	0.01	0.38	0.11	0.49	0.14	0.04	0.1
O4	0.03	0.07	0.02	0.09	0.01	0.00	0.01	0.24	0.06	0.30	0.08	0.02	0.1
O5	0.01	0.05	0.01	0.06	0.00	0.00	0.00	0.15	0.03	0.18	0.05	0.01	0.0
O6	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.06	0.01	0.07	0.02	0.00	0.0
O7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
O8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
O9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
Total	1.12	1.63	0.45	2.08	0.05	0.01	0.06	6.85	1.49	8.33	2.46	0.47	2.9
Total Enlisted	0.97	1.32	0.35	1.67	0.03	0.01	0.03	5.60	1.17	6.77	1.96	0.35	2.3
Total Officers	0.13	0.30	0.10	0.40	0.02	0.01	0.03	1.12	0.31	1.43	0.40	0.11	0.5
Total Warrant Officers	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.01	0.13	0.09	0.01	0.1
W1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.03	0.00	0.0
W2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.01	0.05	0.04	0.00	0.0
W3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.02	0.00	0.0
W4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.01	0.00	0.0
W5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.0
Total	3.35	4.88	1.35	6.23	0.15	0.04	0.19	20.54	4.46	25.00	7.37	1.40	8.7

Popularity of Baby Names (Activity #13)



I chose the names Cora, Floy, Pearl, and Zoe since their popularity patterns differ and its easy to observe trends in the data visualization.

Plotting a Mathematical Function (Activity #04)



Reflection

Throughout the semester, I've gained an excellent skill-set including:

1. Creating functions in R -I learned how to use R to define variables and create functions that perform certain actions using those variables.
2. Data Wrangling - I learned how to use R to tidy any form of data into a clean and easy to understand form, which I can then use in various projects.
3. Data visualizations - I learned how to use packages in R to create visualizations such as tables, graphs, plots, etc. and then make various statistical inferences.
4. Quarto - I learned how to use R and Quarto to create QMD files.

Code Appendix

```
#DATA WRANGLING - US ARMED FORCES DATA
# STEP 1: Load tidyverse
library(tidyverse)

# STEP 2: Load the data
activity_08_data <- read_csv("https://docs.google.com/spreadsheets/d/1cn4i0-ymB1ZytWXCwsJiq6i

# STEP 3: Use pivot_longer() to remove extra columns
activity_08_tidy_data <- activity_08_data %>%
  pivot_longer(
    cols = -1,
    names_to = "PayGrade",
    values_to = "Count"
  )

# STEP 4: Rename columns
first_col_name <- names(activity_08_data)[1]

group_df <- activity_08_tidy_data %>%
  rename(
    ServiceBranch = all_of(first_col_name),
    Soldiers = Count
  ) %>%

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

# STEP 5: Create Pay Gradetable from the given data
paygrade_rank <- tibble(
  PayGrade = c("E1", "E2", "E3", "E4", "E5", "E6", "E7", "E8", "E9",
               "W1", "W2", "W3", "W4", "W5",
               "O1", "O2", "O3", "O4", "O5", "O6", "O7", "O8", "O9", "O10"),
  RankCategory = c(rep("Enlisted", 9), rep("Warrant Officer", 5), rep("Commissioned Officer"
)

# Join Armed Forces data with Pay Grade data
group_df <- group_df %>%
```

```r
  left_join(paygrade_rank, by = "PayGrade")

# STEP 6: Use uncount to create individual soldier records
individual_df <- group_df %>%
  uncount(weights = Soldiers)
#FREQUENCY TABLES
#STEP 1: Load necessary packages
library(tidyverse)
library(janitor)
library(knitr)
library(kableExtra)


#STEP 2: Create group absolute frequency table
group_abs <- group_df %>%
  group_by(ServiceBranch, PayGrade) %>%
  summarise(Soldiers = sum(Soldiers), .groups = "drop") %>%
  pivot_wider(names_from = PayGrade, values_from = Soldiers, values_fill = 0) %>%
  adorn_totals(where = c("row", "col"))


#STEP 3: Make wellformatted table
kable(group_abs, caption = "Group Absolute Frequency Table") %>%
  kable_styling(full_width = FALSE, position = "center", bootstrap_options = c("striped", "ho


#STEP 4: Create group relative frequency table
total_soldiers <- sum(group_df$Soldiers)
group_rel <- group_df %>%
  group_by(ServiceBranch, PayGrade) %>%
  summarise(Soldiers = sum(Soldiers), .groups = "drop") %>%
  mutate(RelFreq = Soldiers / total_soldiers * 100) %>%   # convert to %
  select(-Soldiers) %>%
  pivot_wider(names_from = PayGrade, values_from = RelFreq, values_fill = 0) %>%
  adorn_totals(where = c("row", "col"))


#STEP 5: Make well formatted table
kable(group_rel, digits = 2, caption = "Group Relative Frequency Table") %>%
  kable_styling(full_width = FALSE, position = "center", bootstrap_options = c("striped", "ho
```

```r
#STEP 6: Create individual absolute frequency table
individual_abs <- individual_df %>%
  count(ServiceBranch, PayGrade) %>%
  pivot_wider(names_from = PayGrade, values_from = n, values_fill = 0) %>%
  adorn_totals(where = c("row", "col"))


#STEP 7: Make well formatted table
kable(individual_abs, caption = "Individual Absolute Frequency Table") %>%
  kable_styling(full_width = FALSE, position = "center", bootstrap_options = c("striped", "ho


#STEP 8: Create individual relative frequency table
total_individuals <- nrow(individual_df)
individual_rel <- individual_df %>%
  count(ServiceBranch, PayGrade) %>%
  mutate(RelFreq = n / total_individuals * 100) %>%
  select(-n) %>%
  pivot_wider(names_from = PayGrade, values_from = RelFreq, values_fill = 0) %>%
  adorn_totals(where = c("row", "col"))

#STEP 9: Make well formatted table
kable(individual_rel, digits = 2, caption = "Individual Relative Frequency Table") %>%
  kable_styling(full_width = FALSE, position = "center", bootstrap_options = c("striped", "ho
#POPULARITY OF BABY NAMES
#STEP 1: Load necessary packages
library(dcData)
library(tidyverse)
library(ggplot2)

#STEP 2: Load dataset
data(BabyNames)

#STEP 3: Select names
selected_names <- c("Cora", "Pearl", "Zoe", "Floy") #Replace names

#STEP 4: Wrangle data
tidy_data <- BabyNames %>%
  filter(name %in% selected_names) %>%

#STEP 5: Group by year, name, sex
  group_by(year, name, sex)
```

```r
#STEP 6: Create line plot
ggplot(tidy_data, aes(x = year, y = count, color = name, linetype = name)) + #Use color and l
  geom_line() +
  geom_point() +
  labs(
    title = "Popularity of Selected Names Over Time",
    x = "Year",
    y = "Total Number of Babies with Name",
    color = "Name",
    linetype = "Name"
  ) +
  theme(
    plot.title = element_text(face = "bold")
  )
#PLOTTING OF MATHEMATICAL FUNCTION
# Box Problem
#Finds volume of an open top box made from a rectangular sheet and then uses it to find the l
library(ggplot2)

#STEP 1: Define dimensions of the rectangular sheet.
L <- 36 # numeric length of sheet
W <- 48 #numeric width of sheet

#STEP 2: Define volume function using the size of the cutout x
volume_box <- function(x){x * (L-2*x) * (W-2*x)}

#STEP 3: Define range
x_min <- 0
x_max <- min(L/2, W/2)

# STEP 4: Create plot
ggplot(data = data.frame(x = c(x_min, x_max)), aes(x = x)) +
  stat_function(fun = volume_box, color = "pink", size = 1.2) +
  labs(
    title = "Volume of the box with respect to the side length of the cut out",
    x = "Cutout size (x)",
    y = "Volume of the box"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    axis.title = element_text(face = "bold")
```

```
)
```