

Assignment-Final Project Checkpoint 3

John DeForest

Aug 5, 2025

```
g1 = read.csv('C:\\Users\\John DeForest\\Desktop\\qbs103 R DS\\QBS103_GSE157103_genes.csv')
sm1 = read.csv('C:\\Users\\John DeForest\\Desktop\\qbs103 R DS\\QBS103_GSE157103_series_matrix-1.csv')
#gene, metadata
#check df headers:
#print(head(g1))
#assign first column name as "Gene"
colnames(g1)[1] = 'Gene'
vars <- c("age", "ferritin.ng.ml.", "procalcitonin.ng.ml..", "lactate.mmol.l.")
```

```
#1 my sum fn
manual_sum <- function(x) {
  total <- 0
  for (i in x) {
    if (!is.na(i)) {
      total <- total + i
    }
  }
  return(total)
}
#helper fn to deal with other weird vals in data
clean_numeric <- function(x) {
  as.numeric(gsub("unknown", NA, x))
}

#lapply to vars
sm1[, vars] <- lapply(sm1[, vars], clean_numeric)
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
sm1$age <- as.numeric(sm1$age) #coerce x vals too
results_sum <- sapply(sm1[, vars], manual_sum) #my sum fn
results_base_sum <- sapply(sm1[, vars], function(x) sum(x, na.rm = TRUE)) #compare to base

#Display for comparison
results_sum
```

```
##          age          ferritin.ng.ml.  procalcitonin.ng.ml..
##      7532.00          91687.00          313.93
## lactate.mmol.l.
##      124.31
```

```
results_base_sum
```

```
##           age           ferritin.ng.ml. procalcitonin.ng.ml..  
##       7532.00           91687.00           313.93  
##    lactate.mmol.l..  
##       124.31
```

```
#2 my mean fn  
manual_mean <- function(x) {  
  valid_vals <- x[!is.na(x)] #check na  
  return(manual_sum(valid_vals) / length(valid_vals))  
}  
  
#apply mine, standard mean fn  
results_mean <- sapply(sm1[, vars], manual_mean)  
results_base_mean <- sapply(sm1[, vars], function(x) mean(x, na.rm = TRUE))  
  
#Display for comparison  
results_mean
```

```
##           age           ferritin.ng.ml. procalcitonin.ng.ml..  
##       61.235772           833.518182           3.077745  
##    lactate.mmol.l..  
##       1.462471
```

```
results_base_mean
```

```
##           age           ferritin.ng.ml. procalcitonin.ng.ml..  
##       61.235772           833.518182           3.077745  
##    lactate.mmol.l..  
##       1.462471
```

```
library(ggplot2)  
  
#3 scatter plot fn w mean/sd bar. dynamic input  
plot_var_vs_age <- function(df, yvar, ylab) {  
  mean_y <- mean(df[[yvar]], na.rm = TRUE)  
  sd_y <- sd(df[[yvar]], na.rm = TRUE)  
  
  p <- ggplot(df, aes(x = age, y = .data[[yvar]])) +  
    geom_point(alpha = 0.6) +  
    geom_hline(yintercept = mean_y, color = "blue", linetype = "solid", size = 1) +  
    geom_hline(yintercept = mean_y + sd_y, color = "red", linetype = "dashed") +  
    geom_hline(yintercept = mean_y - sd_y, color = "red", linetype = "dashed") +  
    labs(  
      title = paste(ylab, "vs Age"),  
      x = "Age (years)",  
      y = ylab  
    ) +  
    theme_minimal()  
  #save to png in cur wd for checkin
```

```

filename <- paste0(gsub("[^A-Za-z0-9]", "_", yvar), "_vs_Age.png")
ggsave(filename, plot = p, width = 7, height = 5, dpi = 300)
print(p)
}

#manual set axis labels
ylabs <- c(
  ferritin.ng.ml. = "Ferritin (ng/mL)",
  procaltitonin.ng.ml.. = "Procalcitonin (ng/mL)",
  lactate.mmol.l. = "Lactate (mmol/L)"
)

#Loop through y vars
for (v in names(ylabs)) {
  plot_var_vs_age(sm1, v, ylabs[[v]])
}

```

```

## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

```

```

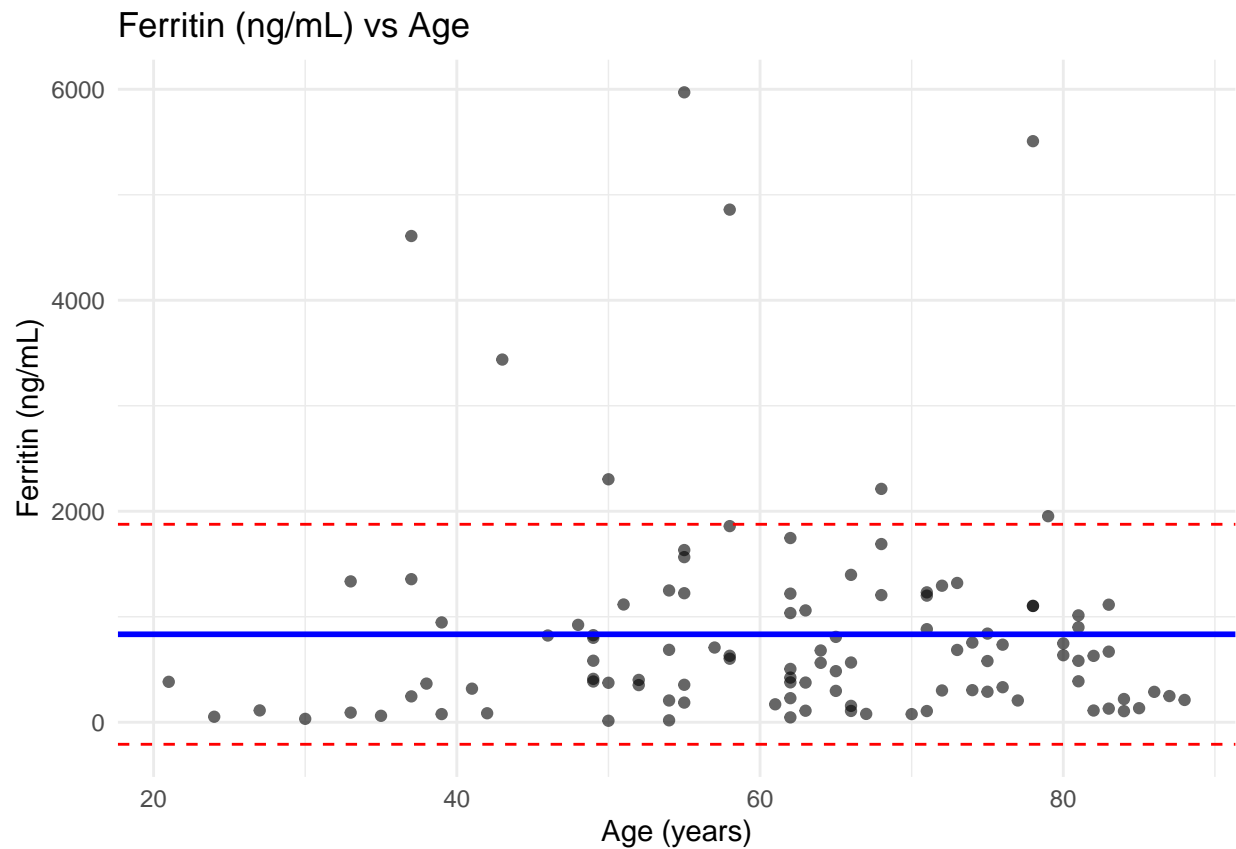
## Warning: Removed 19 rows containing missing values or values outside the scale range
## ('geom_point()').
## Removed 19 rows containing missing values or values outside the scale range
## ('geom_point()').

```

```

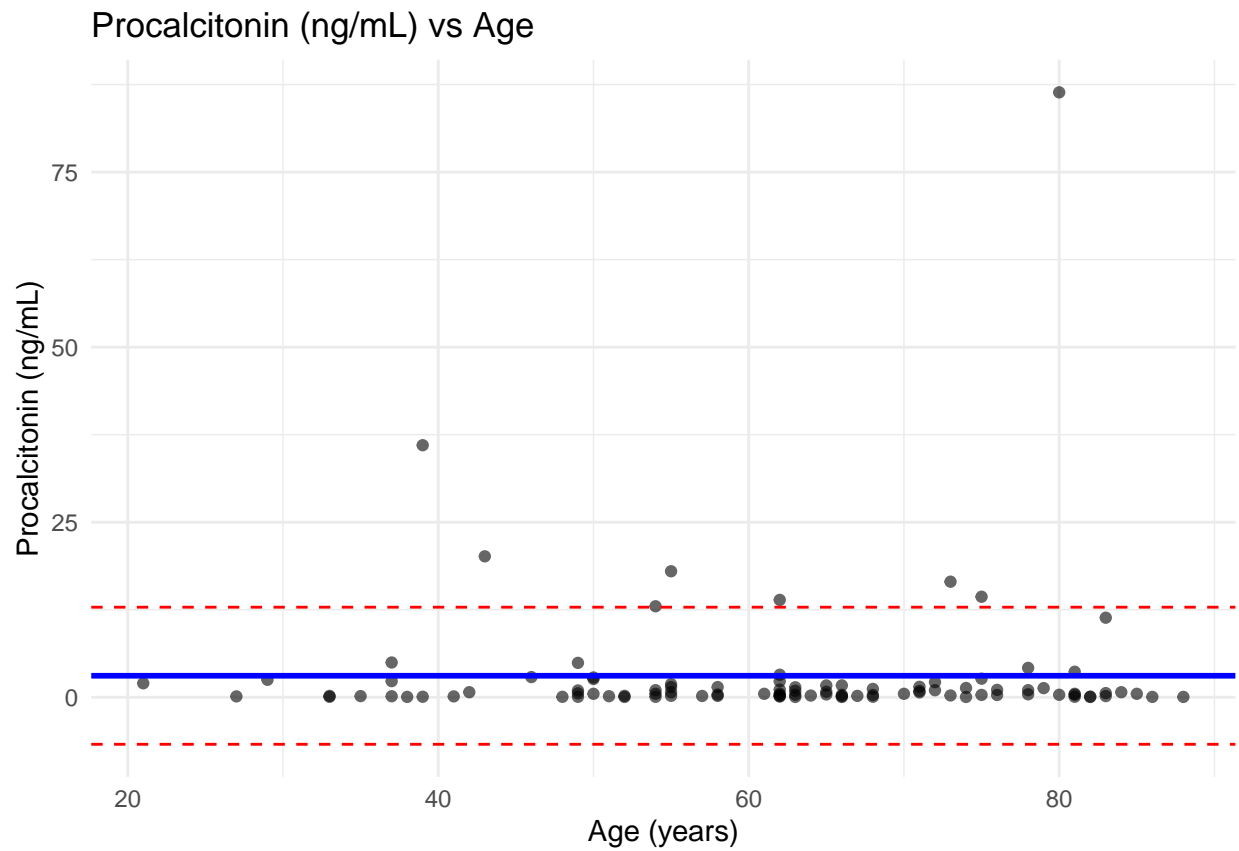
## Warning: Removed 27 rows containing missing values or values outside the scale range
## ('geom_point()').

```



```
## Warning: Removed 27 rows containing missing values or values outside the scale range  
## ('geom_point()').
```

```
## Warning: Removed 44 rows containing missing values or values outside the scale range  
## ('geom_point()').
```



```
## Warning: Removed 44 rows containing missing values or values outside the scale range
## ('geom_point()').
```

