

HW1

2025-02-07

Question 1

(a)

```
su <- read.delim("Su_raw_matrix.txt")
```

(b)

```
mean_liver2 <- mean(suLiver2.CEL, na.rm = TRUE) sd_liver2 <- sd(suLiver_2.CEL, na.rm = TRUE)
```

(c)

```
col_means <- colMeans(su, na.rm = TRUE) col_sums <- colSums(su, na.rm = TRUE)
```

Question 2

```
set.seed(123) par(mfrow=c(1,2)) # Set layout for two plots
```

```
data1 <- rnorm(10000, mean = 0, sd = 0.2) hist(data1, breaks=50, main="Mean=0, SD=0.2", xlim=c(-5,5),  
col="blue")
```

```
data2 <- rnorm(10000, mean = 0, sd = 0.5) hist(data2, breaks=50, main="Mean=0, SD=0.5", xlim=c(-5,5),  
col="red")
```

Question 3

```
library(ggplot2)
```

```
dat <- data.frame(cond = factor(rep(c("A","B"), each=200)), rating = c(rnorm(200), rnorm(200,  
mean=.8)))
```

(b) Overlaid histograms

```
ggplot(dat, aes(x=rating, fill=cond)) + geom_histogram(binwidth=.5, alpha=.5, position="identity")
```

(c) Interleaved histograms

```
ggplot(dat, aes(x=rating, fill=cond)) + geom_histogram(binwidth=.5, position="dodge")
```

(d) Density plots

```
ggplot(dat, aes(x=rating, colour=cond)) + geom_density()
```

(e) Density plots with semitransparent fill

```
ggplot(dat, aes(x=rating, fill=cond)) + geom_density(alpha=.3)
```

(f) Applying same plots to diabetes dataset

```
diabetes <- read.csv("diabetes_train.csv")  
ggplot(diabetes, aes(x=mass, fill=class)) + geom_histogram(binwidth=.5, alpha=.5, position="identity")  
ggplot(diabetes, aes(x=mass, fill=class)) + geom_histogram(binwidth=.5, position="dodge")  
ggplot(diabetes, aes(x=mass, colour=class)) + geom_density()  
ggplot(diabetes, aes(x=mass, fill=class)) + geom_density(alpha=.3)
```

Question 4

```
library(tidyverse)  
passengers <- read.csv("titanic.csv")
```

(a)

```
passengers %>% drop_na() %>% summary()
```

(b)

```
passengers %>% filter(Sex == "male")
```

(c)

```
passengers %>% arrange(desc(Fare))
```

(d)

```
passengers %>% mutate(FamSize = Parch + SibSp)
```

(e)

```
passengers %>% group_by(Sex) %>% summarise(meanFare = mean(Fare, na.rm = TRUE), numSurv =  
sum(Survived, na.rm = TRUE))
```

Question 5

```
quantile(diabetes$skin, probs = c(0.1, 0.3, 0.5, 0.6), na.rm = TRUE)
```