

R Notebook

Code ▼

Hide

```
#1a Import data from text file
data <- read.delim("C:\\Users\\merve\\OneDrive\\Desktop\\Su_raw_matrix.txt")

#1b
mean(data$Liver_2.CEL) # Calculate the mean of Liver_2.CEL column
```

```
[1] 241.8246
```

Hide

```
sd(data$Liver_2.CEL) # Calculate the standard deviation of Liver_2.CEL column
```

```
[1] 1133.352
```

Hide

```
#1c
colMeans(data) # Compute the average of each column
```

Brain_1.CEL	Brain_2.CEL	Fetal_brain_1.CEL	Fetal_brain_2.CEL	Fetal_liver_1.CEL
204.9763	315.0924	198.3439	267.6551	209.8722
Fetal_liver_2.CEL	Liver_1.CEL	Liver_2.CEL		
399.1482	160.8558	241.8246		

Hide

```
colSums(data) # Compute the total of each column
```

Brain_1.CEL	Brain_2.CEL	Fetal_brain_1.CEL	Fetal_brain_2.CEL	Fetal_liver_1.CEL
2588031	3978357	2504290	3379413	2649846
Fetal_liver_2.CEL	Liver_1.CEL	Liver_2.CEL		
5039645	2030966	3053278		

Hide

#2

```
set.seed(123) # Set seed for reproducibility
```

#2a

```
a <- rnorm(10000, mean=0, sd= 0.2) # Generate 10,000 random numbers from N(0, 0.2)
# Save histogram as PNG
png("histogram_a.png", width=800, height=600)
hist(a, main = "Histogram: mean=0, sd=0.2", xlab = "Values",xlim=c(-5,5) )
dev.off()
```

```
null device
      1
```

Hide

#2b

```
b <- rnorm(10000, mean=0, sd= 0.5) # Generate 10,000 random numbers from N(0, 0.5)
# Save histogram as PNG
png("histogram_b.png", width=800, height=600)
```

Hide

```
hist(b, main = "Histogram: mean=0, sd=0.5", xlab = "Values", xlim=c(-5,5))
dev.off()
```

```
null device
      1
```

Hide

#In both graphs, the mean appears around 0, but in graph (a), the distribution is narrower and more concentrated because the standard deviation is small, whereas in graph (b), it is wider and the values are further from the mean

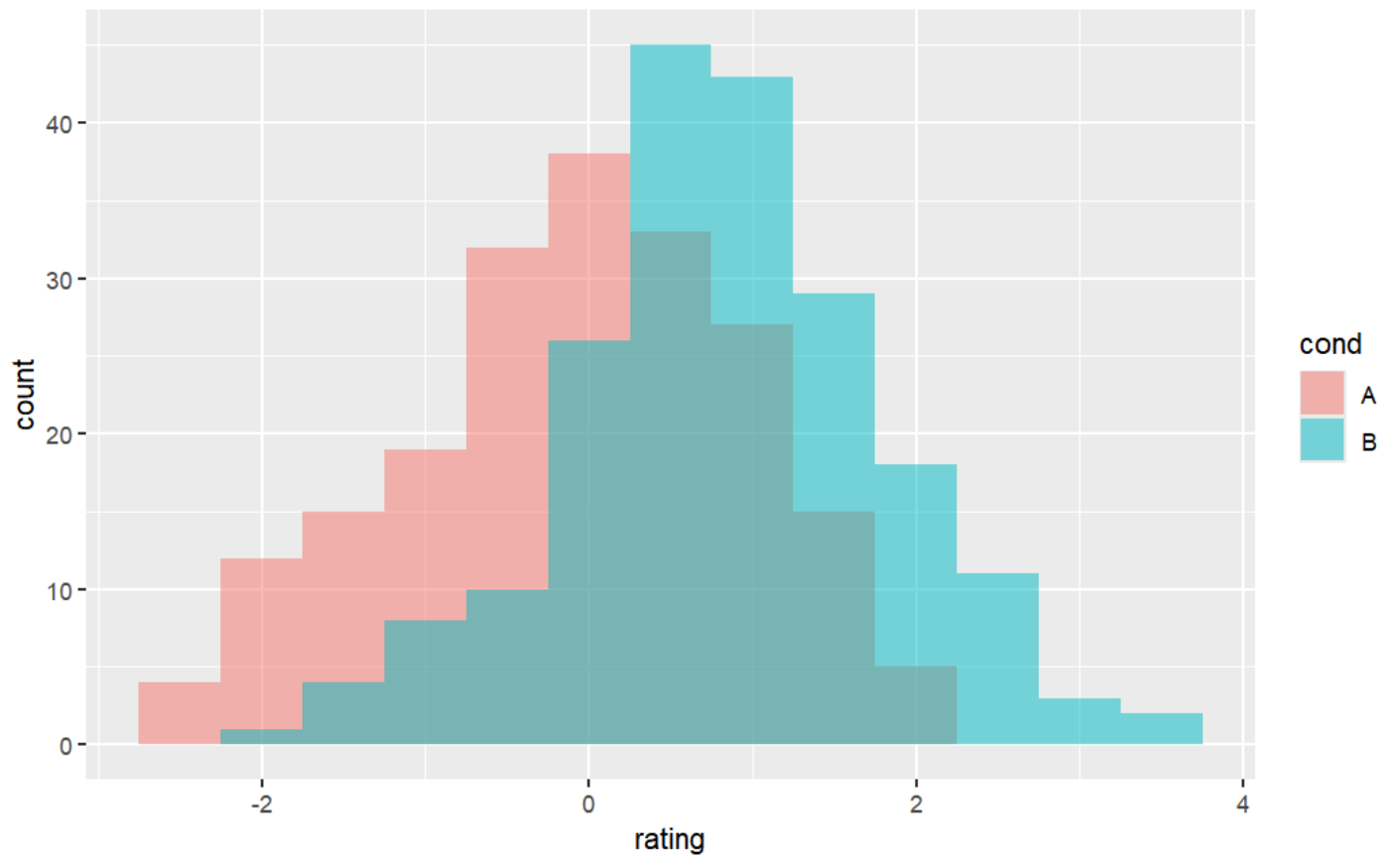
Hide

```
#install.packages("ggplot2")

library(ggplot2) # Import ggplot2 package

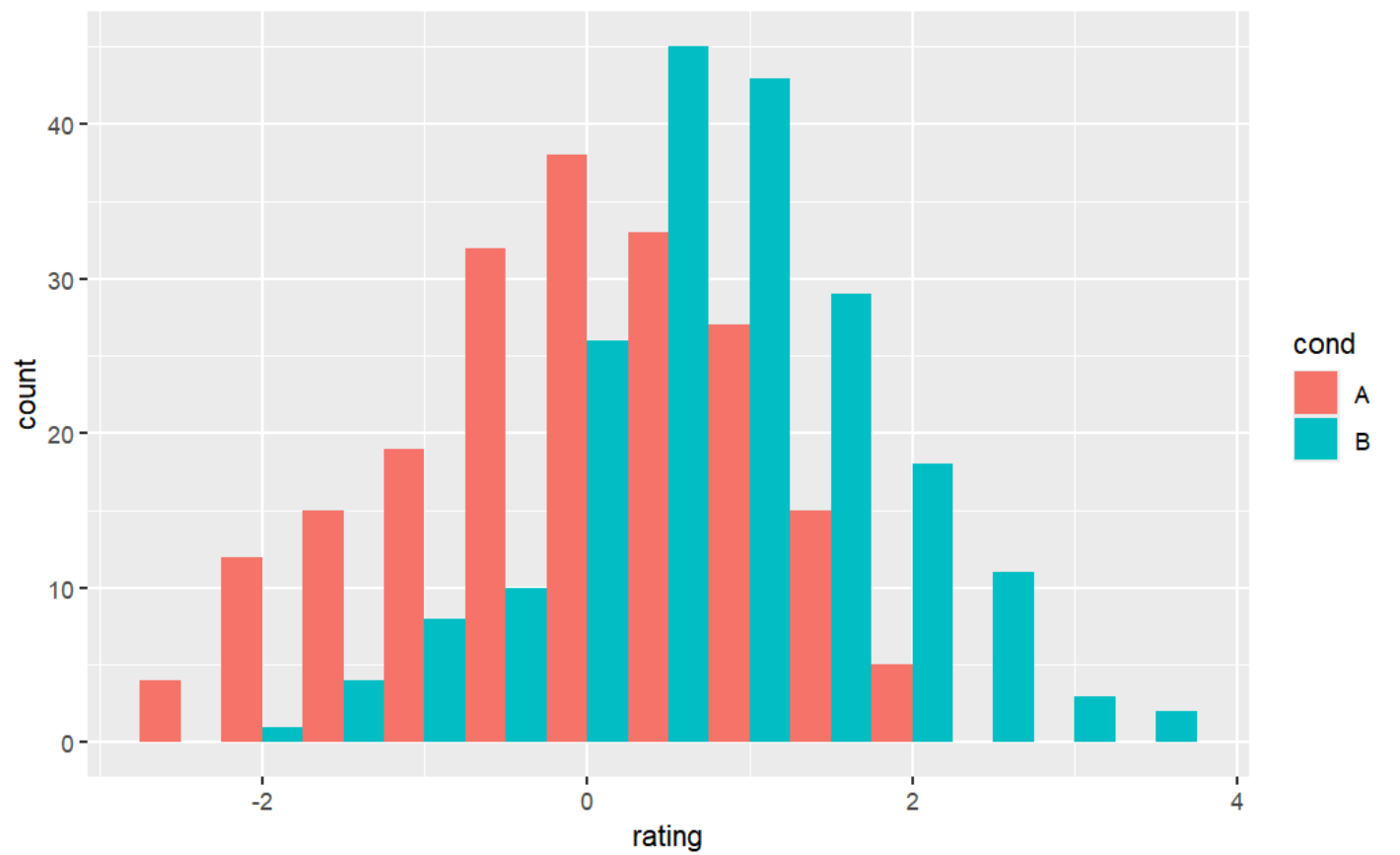
dat <- data.frame(cond = factor(rep(c("A","B"), each=200)),
  rating = c(rnorm(200),rnorm(200, mean=.8)))

#3b Overlaid histograms
ggplot(dat, aes(x=rating, fill=cond)) +
  geom_histogram(binwidth=.5, alpha=.5, position="identity")
```

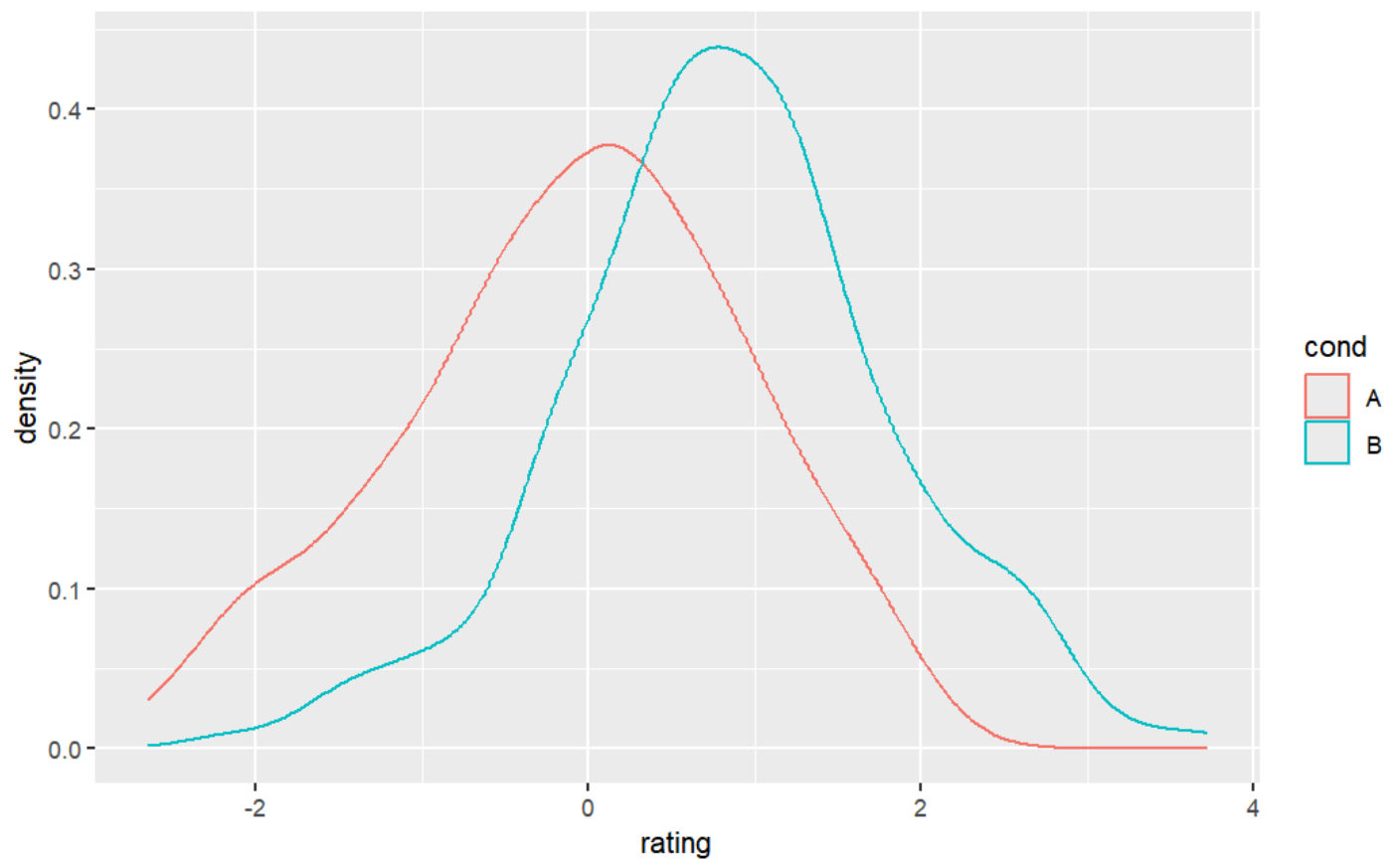
[Hide](#)

#3c Interleaved histograms

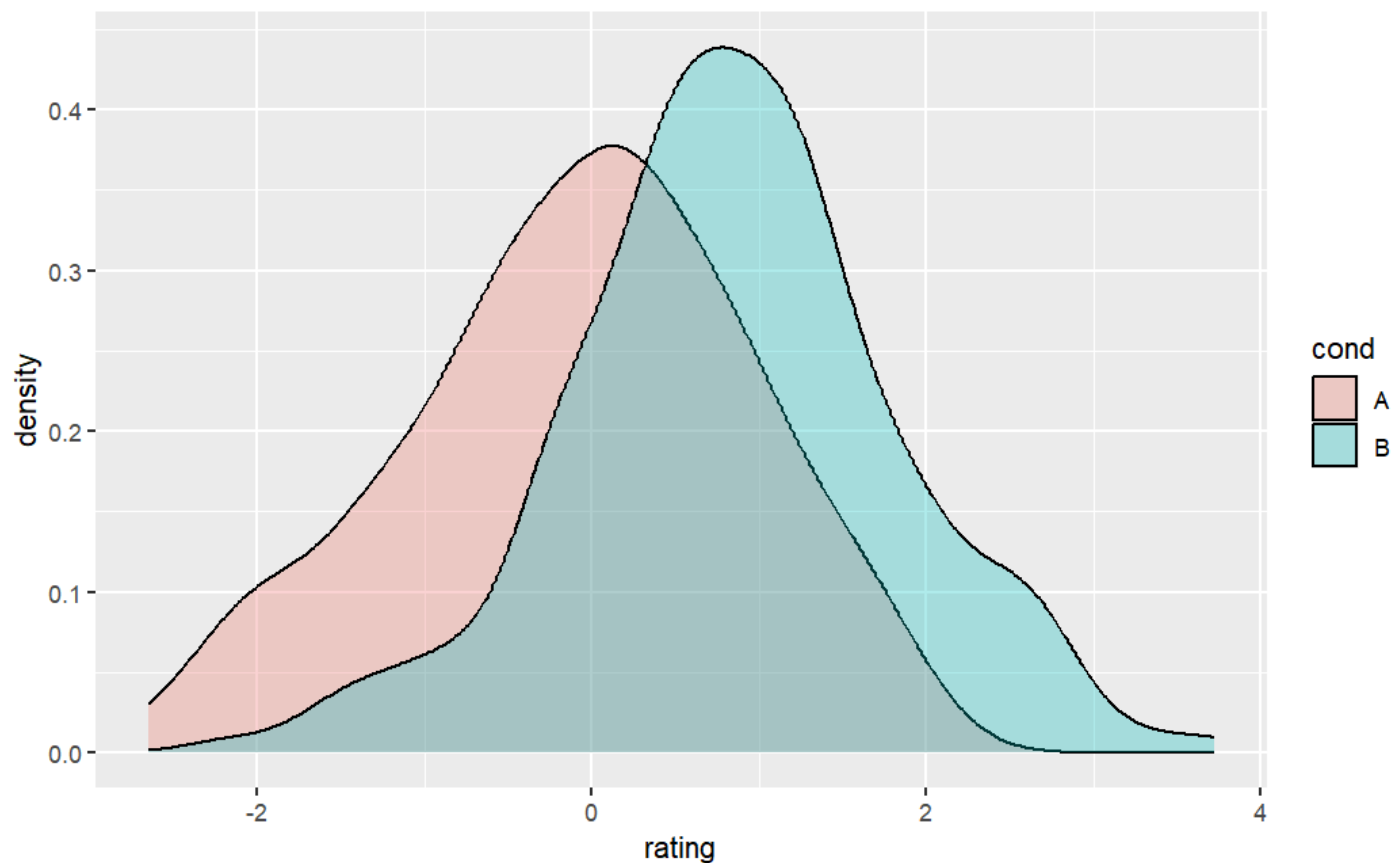
```
ggplot(dat, aes(x=rating, fill=cond)) + geom_histogram(binwidth=.5, position="dodge")
```

[Hide](#)

```
#3d Density plots  
ggplot(dat, aes(x=rating, colour=cond)) + geom_density()
```

[Hide](#)

```
#3e Density plots with semitransparent fill  
ggplot(dat, aes(x=rating, fill=cond)) + geom_density(alpha=.3)
```


[Hide](#)

#3f

```
diabetes_data <- read.csv("C:\\Users\\merve\\OneDrive\\Desktop\\diabetes_train.csv")
```

```
#Overlaid histograms
```

```
p1 <- ggplot(diabetes_data, aes(x=mass, fill=class)) +  
  geom_histogram(binwidth=.5, alpha=.5, position="identity")  
ggsave("overlaid_hist.png", plot = p1, width = 8, height = 6)
```

```
#Interleaved histograms
```

```
p2 <- ggplot(diabetes_data, aes(x=mass, fill=class)) + geom_histogram(binwidth=.5, position="dodge")  
ggsave("interleaved_hist.png", plot = p2, width = 8, height = 6)
```

```
#Density plots
```

```
p3 <- ggplot(diabetes_data, aes(x=mass, colour=class)) + geom_density()  
ggsave("density_plots.png", plot = p3, width = 8, height = 6)
```

```
#Density plots with semitransparent fill
```

```
p4 <- ggplot(diabetes_data, aes(x=mass, fill=class)) + geom_density(alpha=.3)  
ggsave("dp_withfill.png", plot = p4, width = 8, height = 6)
```

[Hide](#)

#4

```
remove.packages("tidyverse")
install.packages("tidyverse")
```

Error in install.packages : Updating loaded packages

Hide

```
library(tidyverse)
library(tidyr)

passengers <- read.csv("C:\\Users\\merve\\OneDrive\\Desktop\\titanic.csv")
```

#4a

```
passengers %>%
  drop_na() %>% # Remove rows with NA values
  summary()     # Show summary statistics of each column
```

X	PassengerId	Survived	Pclass	Name
Min. : 0.0	Min. : 1.0	Min. :0.0000	Length:714	Length:714
1st Qu.:221.2	1st Qu.:222.2	1st Qu.:0.0000	Class :character	Class :character
Median :444.0	Median :445.0	Median :0.0000	Mode :character	Mode :character
Mean :447.6	Mean :448.6	Mean :0.4062		
3rd Qu.:676.8	3rd Qu.:677.8	3rd Qu.:1.0000		
Max. :890.0	Max. :891.0	Max. :1.0000		

Sex	Age	SibSp	Parch	Ticket
Length:714	Min. : 0.42	Min. :0.0000	Min. :0.0000	Length:714
Class :character	1st Qu.:20.12	1st Qu.:0.0000	1st Qu.:0.0000	Class :character
Mode :character	Median :28.00	Median :0.0000	Median :0.0000	Mode :character
	Mean :29.70	Mean :0.5126	Mean :0.4314	
	3rd Qu.:38.00	3rd Qu.:1.0000	3rd Qu.:1.0000	
	Max. :80.00	Max. :5.0000	Max. :6.0000	

Fare	Cabin	Embarked
Min. : 0.00	Length:714	Length:714
1st Qu.: 8.05	Class :character	Class :character
Median :15.74	Mode :character	Mode :character
Mean :34.69		
3rd Qu.:33.38		
Max. :512.33		

Hide

#4b

```
passengers %>%
  filter(Sex == "male") # Keep only male passengers
```

X	PassengerId	Survived	Pclass	Name	Sex
<int>	<int>	<int>	<chr>	<chr>	<chr>
0	1	0	3	Braund, Mr. Owen Harris	male
4	5	0	3	Allen, Mr. William Henry	male
5	6	0	3	Moran, Mr. James	male
6	7	0	1	McCarthy, Mr. Timothy J	male
7	8	0	3	Palsson, Master. Gosta Leonard	male
12	13	0	3	Saunderscock, Mr. William Henry	male
13	14	0	3	Andersson, Mr. Anders Johan	male
16	17	0	3	Rice, Master. Eugene	male
17	18	1	2	Williams, Mr. Charles Eugene	male
20	21	0	2	Fynney, Mr. Joseph J	male

1-10 of 577 rows | 1-8 of 13 columns

Previous123456...58Next

Hide

```
#4c

passengers %>%
  arrange(desc(Fare)) # Sort by Fare in descending order
```

X	PassengerId	Survived	Pclass
<int>	<int>	<int>	<chr>
258	259	1	1
679	680	1	1
737	738	1	1
27	28	0	1
88	89	1	1
341	342	1	1
438	439	0	1
311	312	1	1
742	743	1	1
118	119	0	1

1-10 of 891 rows | 1-4 of 13 columns

Previous123456...90Next


```
#4d

passengers %>%
  mutate(FamSize = Parch + SibSp) # Create new column 'FamSize' as the sum of 'Parch' and 'SibSp'
```

X	PassengerId	Survived	Pclass	Name
<int>	<int>	<int>	<chr>	<chr>
0	1	0	3	Braund, Mr. Owen Harris
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)
2	3	1	3	Heikkinen, Miss. Laina
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)
4	5	0	3	Allen, Mr. William Henry
5	6	0	3	Moran, Mr. James
6	7	0	1	McCarthy, Mr. Timothy J
7	8	0	3	Palsson, Master. Gosta Leonard
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)

1-10 of 891 rows | 1-6 of 14 columns

Previous123456...90Next

```
#4e

passengers %>%
  group_by(Sex) %>%
  summarise(
    meanFare = mean(Fare), # Average Fare per Sex
    numSurv = sum(Survived) # Total number of survivors per Sex
  )
```

Sex	meanFare	numSurv
<chr>	<dbl>	<int>
female	44.47982	233
male	25.52389	109

2 rows

#5

```
quantile(diabetes_data$skin, probs= c(0.1, 0.3, 0.5, 0.6))
```

```
10% 30% 50% 60%  
 0   10   23   27
```