Assignment_2

Joseph

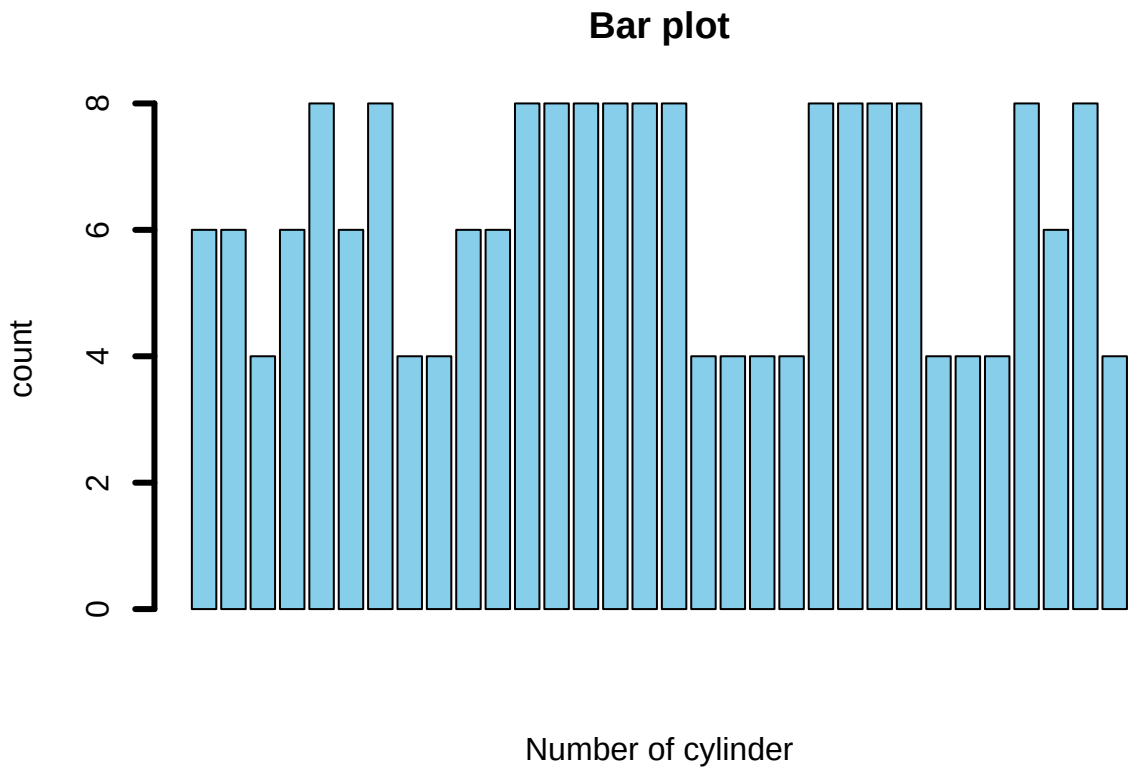
2024-07-11

Load the "mtcars" dataset from the built-in datasets in R. Write an R script to perform the following tasks:

```
### Calculate the average miles per gallon (mpg) for all cars in the dataset.
data <- mtcars
avg_miles_per_gallon <-  mean(data$mpg)

cat("Average miles per gallon(mpg) is ", avg_miles_per_gallon ,"\n")
```

```
## Average miles per gallon(mpg) is  20.09062
```

```
### Create a bar plot to visualize the number of cylinders (cyl) for each car.
barplot(data$cyl, main="Bar plot", xlab = "Number of cylinder", ylab = "count", col="skyblue", border =
```



Write an R script to read a CSV file named "sales_data.csv" containing sales data for different products. Perform the following tasks:

```
library(dplyr)
library(tidyverse)

# Calculate the total sales for each product category.
sales_data <-  read.csv("sales_data.csv")

total_sales_by_product <- sales_data  %>% group_by(Product) %>%  summarise(sum_total_sales = sum(Sales)
 )

print(total_sales_by_product)
```

```
## # A tibble: 3 x 2
##   Product   sum_total_sales
##   <chr>               <int>
## 1 Heater               1000
## 2 Kettle               1150
## 3 Microwave             550
```

```
# Generate a line plot to visualize the monthly sales trend for one of the product categories.
sales_data$Date <- as.Date(sales_data$Date)

print(sales_data)
```

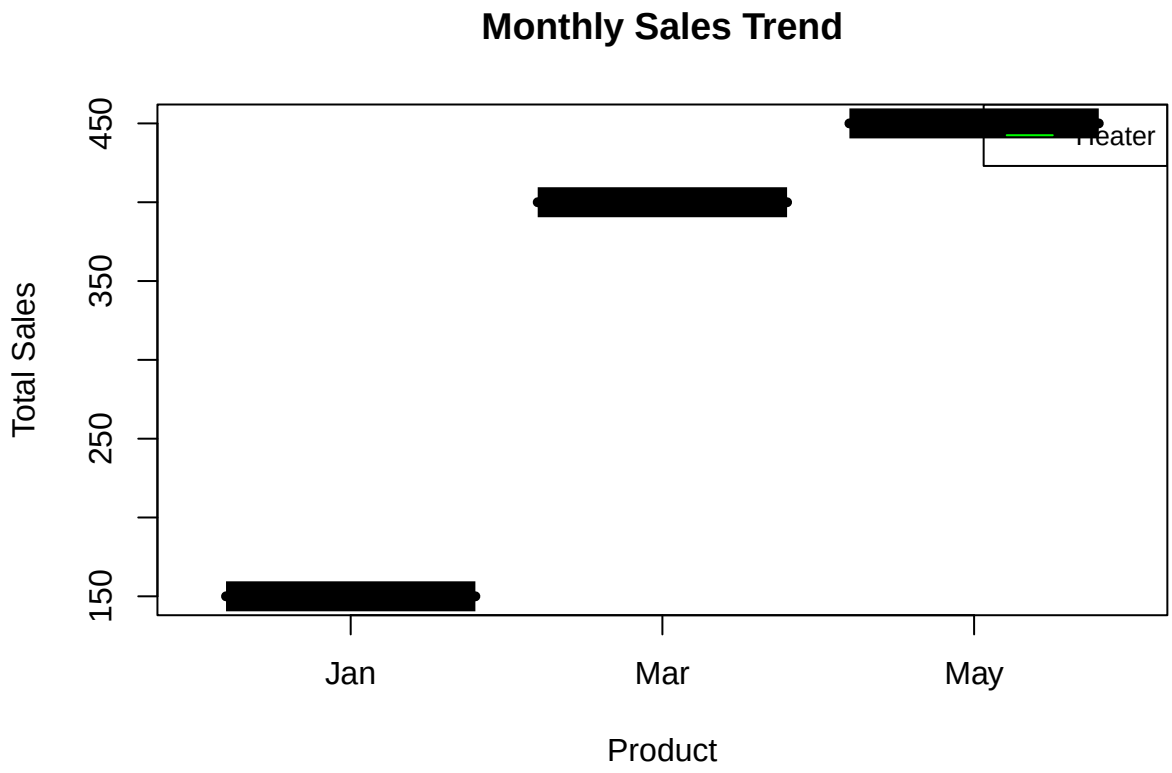
```
##     Product Sales Region       Date
## 1    Heater   150  North 2024-01-01
## 2 Microwave   200   East 2024-02-01
## 3    Kettle   300   West 2024-01-03
## 4    Heater   400  South 2024-03-04
## 5 Microwave   250  North 2024-01-05
## 6    Kettle   350   East 2024-04-06
## 7    Heater   450   West 2024-05-07
## 8 Microwave   100  South 2024-06-08
## 9    Kettle   500  North 2024-02-09
```

```
## Filter data for the specified product category and adding new month column
filtered_data <- sales_data %>%
mutate(Month = format(sales_data$Date, "%b")) %>%  filter(sales_data$Product == "Heater")

print(filtered_data)
```

```
##   Product Sales Region       Date Month
## 1  Heater   150  North 2024-01-01   Jan
## 2  Heater   400  South 2024-03-04   Mar
## 3  Heater   450   West 2024-05-07   May
```

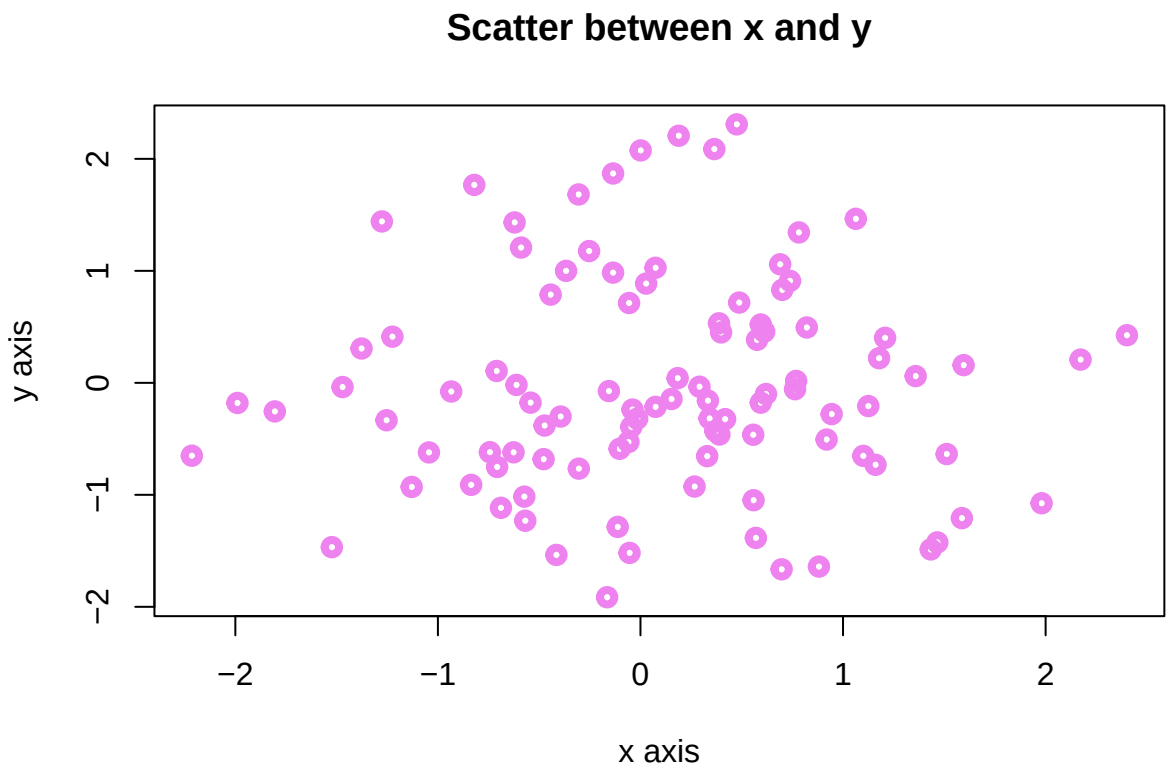
```
## Line Plot
plot(factor(filtered_data$Month),filtered_data$Sales,
```

```
     type="l", # type "o" for points and lines
     col = "blue",
     xlab = "Product",
     ylab = "Total Sales",
     lwd=5,
     main = paste("Monthly Sales Trend"))

## Add a legend
legend("topright", legend = "Heater", col = "green", lty = 1, cex = 0.8)
```



Write an R script to generate a scatter plot using random data. The scatter plot should have 100 points with x and y coordinates ranging from 1 to 100. Add labels and a title to the plot.

```
set.seed(1)

x <- rnorm(100)
y <- rnorm(100)

plot(x,y , main="Scatter between x and y", xlab = "x axis", ylab = "y axis", col="violet", lwd=4)
```



Load the "iris" dataset from the built-in datasets in R.

Create a box plot to visualize the distribution of petal widths (Petal.Width) for each species.

```
## Calculate the average sepal length (Sepal.Length) for each species.

grouped_data <- iris %>% group_by(Species)%>%
 summarise(avg_sep_length <- mean(Sepal.Length))

print(grouped_data)
```

```
## # A tibble: 3 x 2
##   Species    `avg_sep_length <- mean(Sepal.Length)`
##   <fct>                                       <dbl>
## 1 setosa                                       5.01
## 2 versicolor                                   5.94
## 3 virginica                                    6.59
```

```
### Create a box plot to visualize the distribution of petal widths (Petal.Width) for each species
library(ggplot2)
ggplot(iris, aes(x = Species, y=Petal.Width)) +
```

```
  geom_boxplot(fill="skyblue") +
  ggtitle("Boxplot : Petal Width by species")
```