

```

install.packages("tidyverse")

library(tidyverse)


# reading in the data
df <- read.csv("StudentsPerformance.csv")

# taking a quick look
glimpse(df)


is.na(df$math.score)
is.na(df$reading.score)
is.na(df$writing.score)
is.na(df$test.preparation.course)

#non standard na
df <- df %>%

  mutate(test.preparation.course = replace(test.preparation.course, test.preparation.course == "none",
NA)) %>%

  mutate(test.preparation.course = replace(test.preparation.course, test.preparation.course == "N/A",
NA))

is.na(df$test.preparation.course)


# replacing "--" with NA
df <- df %>%

  mutate(test.preparation.course= replace(test.preparation.course, is.na(test.preparation.course),
"unavailable"))

df$test.preparation.course


#data visual
scatter.smooth(x=df$math.score,y=df$reading.score,xlab="maths score",

```

```

        ylab="reading score",

        col="red",pch=19)

scatter.smooth(x=df$math.score,y=df$writing.score,xlab="maths score",
        ylab="writing score",

        col="blue",pch=19)
#box

install.packages("vioplot")
library("vioplot")

vioplot(df$math.score~df$gender ,col=2:length(levels(df$gender )))
#bar
couts <- table(df$gender , df$race.ethnicity )
barplot(couts,main = "simple com",xlab = "gears", legend=rownames(couts),col = c("red","blue"))
#hist
getwd()
x = read.csv("StudentsPerformance.csv")
x
hist(x$math.score,
        main = "Marks in maths",
        xlab = "Students in nos.",
        ylab = "Marks",
        las = 1,
        col = c("skyblue", "chocolate2")
)

```

```
model <- lm(math_score ~ reading_score,data=data) print(model) summary(model) a<-  
data.frame(reading_score=90) predict(model,a) sqrt(mean((df$math.score-  
predicted_value)^2))
```

Predicting math scores for reading and writing score

```
model3 = lm(math.score~reading.score + writing.score, data =df)
```

```
print(model3)
```

```
#Review the results
```

```
summary(model3)
```

```
#reading <-df$reading.score
```

```
#writing <-df$writing.score
```

```
b <-data.frame(reading.score=90,writing.score=88)
```

```
predicted_value <-predict(model3,b)
```

```
predicted_value
```

```
##Redidual Standard Error
```

```
RSE=sigma(model3)/mean(df$math.score)
```

```
#Accuracy
```

```
sqrt(mean((predicted_value)^2))
```