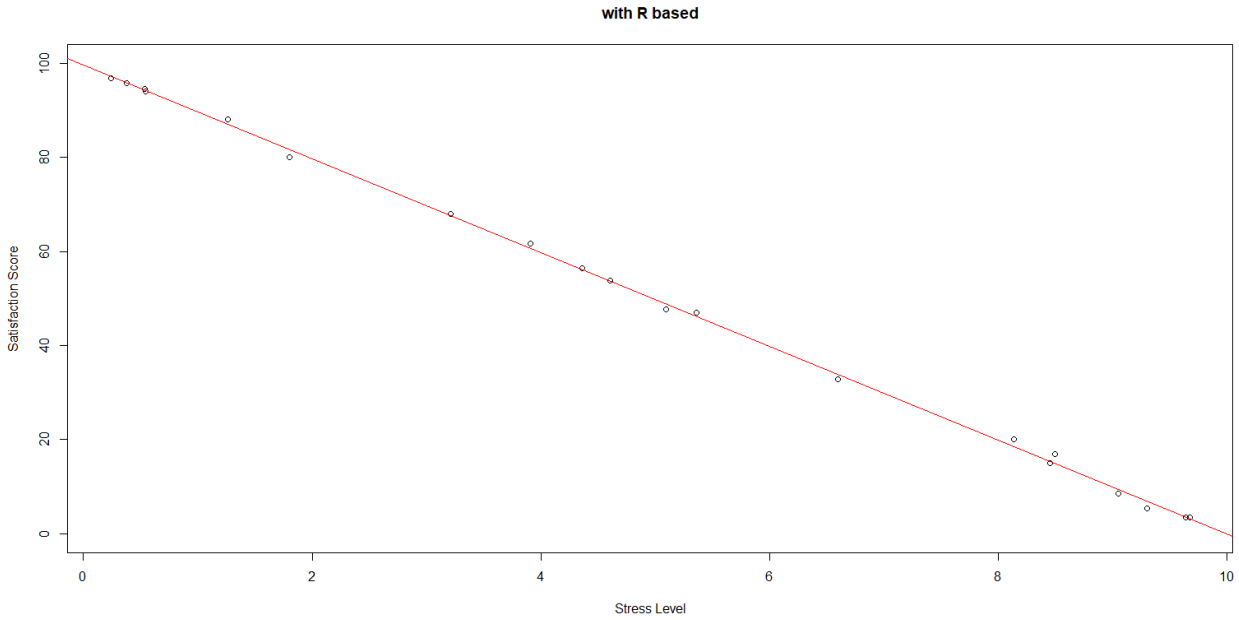
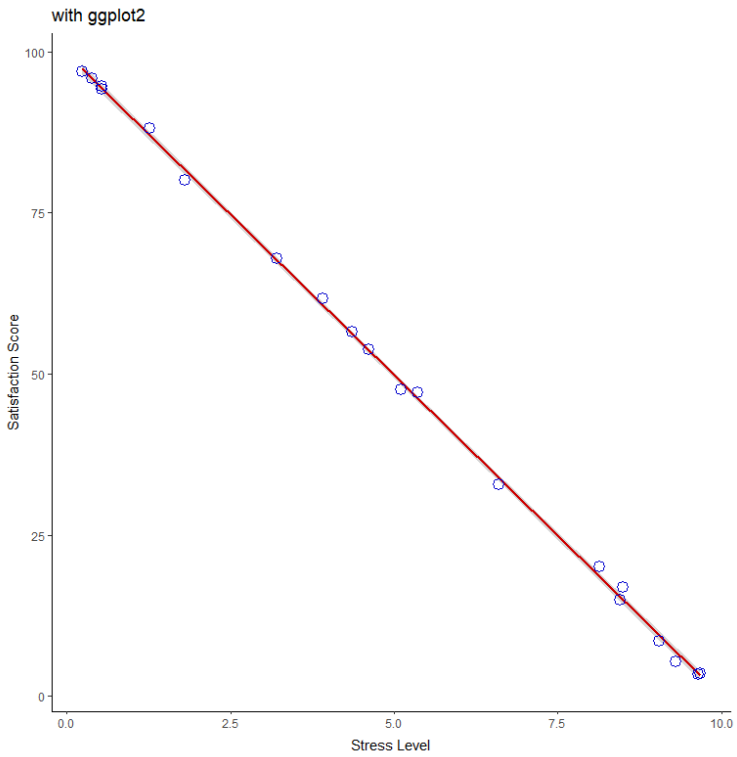
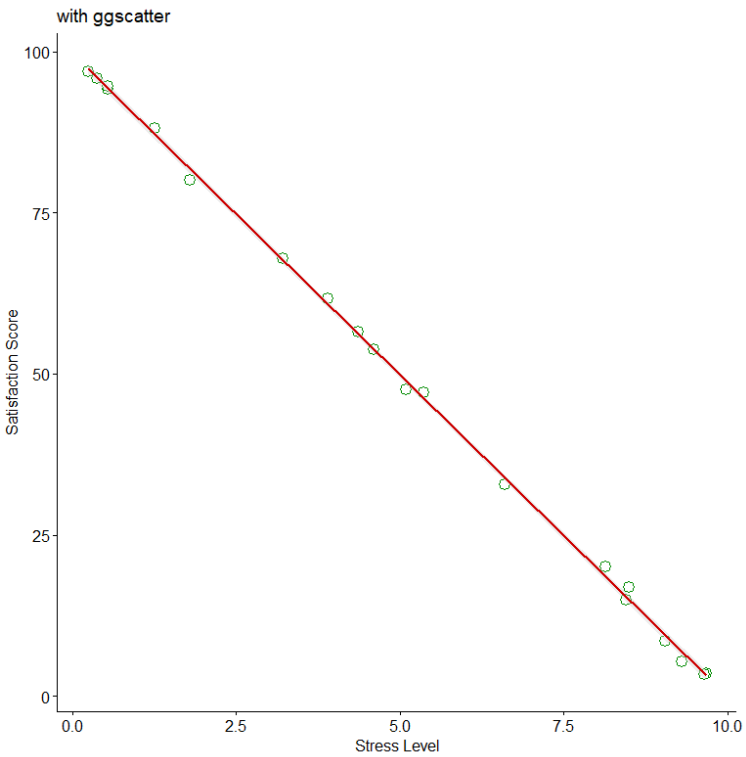
```r
# R course for beginners
# Week 6 part 1
# assignment by Amir Mano, id 205779788

#### prepare workspace ----
rm(list = ls()) #or Ctrl + Shift + F10 & Ctrl + L

# import packages
library(tidyverse)

#### creating and saving variables ----
N = 20
df <- data.frame(
  sub_id = seq(1,N),
  age = runif(N, 18, 60),
  gender = factor(sample(c('M','F'), N, replace=T)),
  stress = runif(N, 0, 10)
)

# adding satisfaction score
b0 = 100
b1 = -10
df <- df |>
  mutate(satisfaction = rnorm(N,b0 + stress*b1),30)
df <- df |>
  mutate(satisfaction = if_else(satisfaction > 100, 100, satisfaction),
         satisfaction = if_else(satisfaction < 0, 0, satisfaction))

# save
save(df, file='HW_6_1.RData')
```

```r
# R course for beginners
# Week 6 part 2
# assignment by Amir Mano, id 205779788

#### prepare workspace ----
rm(list = ls()) #or Ctrl + Shift + F10 & Ctrl + L
load('HW_6_1.RData')

# import packages
library(tidyverse)
library(ggplot2)
library(patchwork)
library(ggpubr)

#### descriptive statistics ----
cat('Age statistics:\n', summary(df)[c(1,6,4,3) ,2],'\n')
cat('Female numbers:\n', df$sub_id[df$gender=='F'],'\n')
cat('Male numbers:\n', df$sub_id[df$gender=='M'],'\n')

#### plotting ----
# with R based
plot(df$stress,df$satisfaction, ylim= c(0,100),
     xlab = 'Stress Level', ylab = 'Satisfaction Score')
title('with R based')
abline(lm(df$satisfaction ~ df$stress), col = "red")
plot1 <- recordPlot()

# with ggplot
plot2 <- ggplot(data = df, aes(x = stress, y = satisfaction)) +
  geom_smooth(method = "lm", color = "red3") +
  geom_point(color = "blue3", shape = 21, size = 4) +
  labs(title = "with ggplot2",
       x = "Stress Level",
       y = "Satisfaction Score") +
  theme_classic()


plot3<- ggscatter(data = df, x = "stress", y = "satisfaction",
          color = "green4", shape = 21, size = 4,
          add = "reg.line",
          add.params = list(color = "red3", fill = "lightgray"),
          conf.int = TRUE) +
  labs(title = "with ggscatter",
       x = "Stress Level",
       y = "Satisfaction Score")

plot1
plot2 + plot3
```


with R based
Satisfaction Score
Stress Level
0
20
40
60
80
100
0
2
4
6
8
10


with ggplot2
Satisfaction Score
Stress Level
0
25
50
75
100
0.0
2.5
5.0
7.5
10.0


with ggscatter
Satisfaction Score
Stress Level
0
25
50
75
100
0.0
2.5
5.0
7.5
10.0

```r
# R course for beginners
# Week 6 part 3
# assignment by Amir Mano, id 205779788

#### prepare workspace ----
rm(list = ls()) #or Ctrl + Shift + F10 & Ctrl + L
load('HW_6_1.RData')

# import packages
library(tidyverse)

#### regression models ----
# without scaling
model <- lm(data = df, satisfaction ~ stress)
model$coefficients[1:2]
# N = 20
# (Intercept)      stress
#   99.956724   -9.962657
# N = 200
# (Intercept)      stress
#   99.888501   -9.983768
# Original 100, -10

# Pearson
correlation <- cor(df$satisfaction, df$stress)
correlation
# -0.9993229

# scaling
df<- df|> mutate(s_satisfaction = scale(satisfaction))
df<- df|> mutate(s_stress = scale(stress))
model_scaled <- lm(data = df, s_satisfaction ~ s_stress)
model_scaled$coefficients[2]
# -0.9993229 - same as Pearson's r
```