Activity 2

5645242 Daniel Guo

2025-02-28

2.

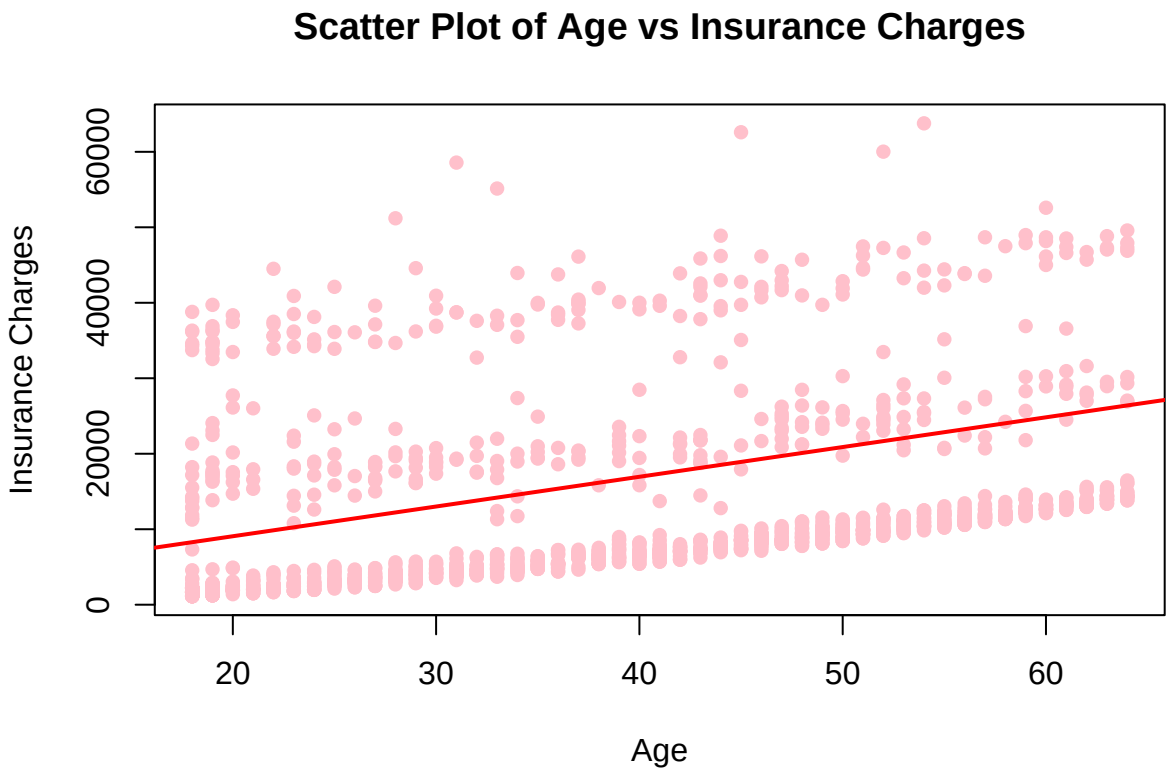
```
# Import the insurance dataset
Insurance <- read.csv(
  "/Users/danielguo/Desktop/University/Year 1/ST117/Activity 2/insurance.csv",
  sep = ",")
#print the first few rows
head(Insurance)
```

```
##   index age    sex    bmi children smoker    region   charges
## 1     0  19 female 27.900        0    yes southwest 16884.924
## 2     1  18   male 33.770        1     no southeast  1725.552
## 3     2  28   male 33.000        3     no southeast  4449.462
## 4     3  33   male 22.705        0     no northwest 21984.471
## 5     4  32   male 28.880        0     no northwest  3866.855
## 6     5  31 female 25.740        0     no southeast  3756.622
```

3. Make a scatterplot of two variables: BMI against charges

```
plot(Insurance$age, Insurance$charges,
     main = "Scatter Plot of Age vs Insurance Charges",
     xlab = "Age",
     ylab = "Insurance Charges",
     col = "pink",
     pch = 16)  # point shape

# Trend line
abline(lm(charges ~ bmi, data = Insurance), col = "red", lwd = 2)
```



4.

Trends and Observations:

Positive Relationship between age and insuance charges as the red trend line is upward sloping, indicating that as age increases, insurance charges also tend to increase, suggesting that older individuals generally have higher insurance costs.

There is a wide range of charges for each age group. Some individuals, at all ages have very high charges, possibly due to medical conditions, bmi, societal status, or other factors.

Possible Reasons for the Trend:

Older individuals tend to have higher medical risks, leading to increased insurance costs.

Factors like BMI, smoking status, and medical factors might explain the variations.

Different individuals may have different types of insurance plans, due to their wealth, conditions and status etc., which could influence costs.