

Homework 9

due Apr 6, 2021

(For questions 1, 2, 3) Consider the following paired data sets of length 20:

```
x <- c(6.82, 1.44, 9.39, 8.51, 10.38, 4.59, 14.96, 9.68, 13.54, 6.42, 11.03,
        3.53, 16.91, 9.52, 8.16, 8.97, 8.32, 3.58, 13.57, 9.99)
y <- c(36.69, 6.39, 49.59, 45.65, 52.18, 27.66, 79.35, 54.10, 71.01, 34.60, 61.17,
        22.79, 91.20, 50.57, 44.11, 53.51, 45.96, 22.20, 73.01, 55.70)
```

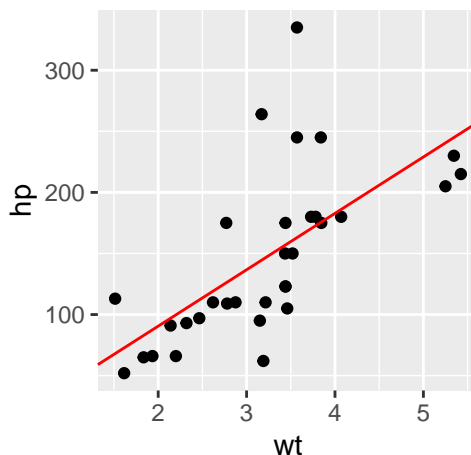
1. Create a scatter plot to visualize the data (*Hint*: you may want to start with making a data frame and then use `geom_point()`, `x` as x -axis and `y` as y -axis).
2. Do you think there is a strong linear association between `x` and `y`? Compute the sample correlation coefficient between `x` and `y` to justify your claim.
3. Assume that `y` is an outcome in a certain experiment, and `x` is a predictor. Find the best fitting line describing the association between `x` and `y` by specifying its y -intercept (β_0) and slope (β_1). Overlay the best fitting line to the plot you obtained in 1. above.

(For questions 4, 5) Consider the `mtcars` data set we used in class:

```
head(mtcars, n = 3)
```

```
##           mpg cyl  disp  hp  drat    wt  qsec vs am gear carb
## Mazda RX4    21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

4. Obtain the following scatter plot with the best fitting line:



5. Compute r , the correlation coefficient.