Q1. Features: The median line is closer to the top(Q3), therefore Q2 - Q1 > Q3 - Q2, the distribution is skewed to the left. The IQR (height of the box) is approximately 10 and there is one outlier at approximately 135.

Q2. a) C

b) A

c) B

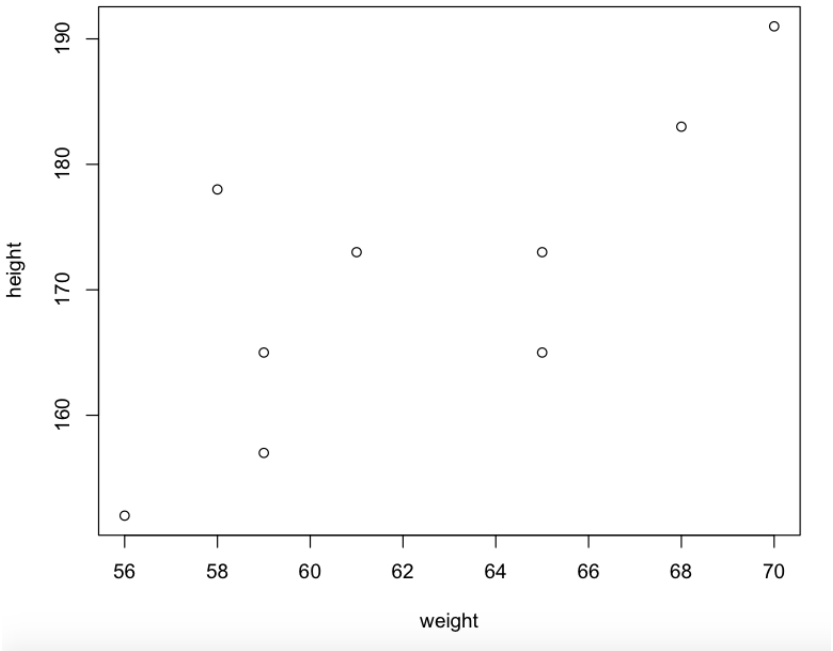
d) B

e) C

f) E

Q3. a)
```
> weight <- c(59,70 ,68 ,61, 65, 65, 56, 59, 58)
> height <- c(165 ,191, 183, 173, 173, 165, 152, 157, 178)
> plot(weight,height)
```

The plot looks linear, positive, strong

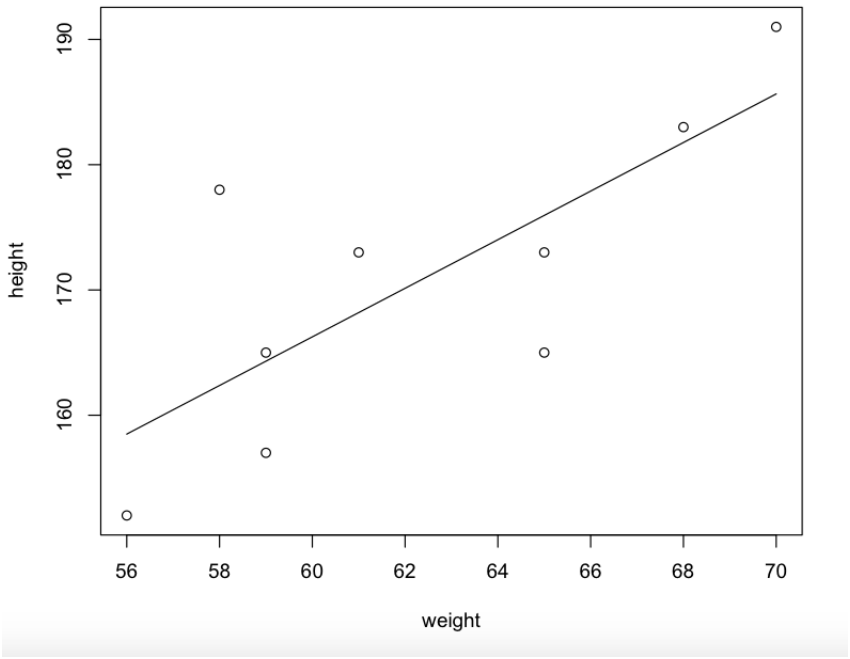


b)
```
> xbar <- mean(weight)
> ybar <- mean(height)
> rho <- cor(weight, height)
> sd_x <- sd(weight)
> sd_y <- sd(height)
> b_1 <- rho*(sd_y/sd_x)
> b_0 <- ybar - b_1*xbar
```

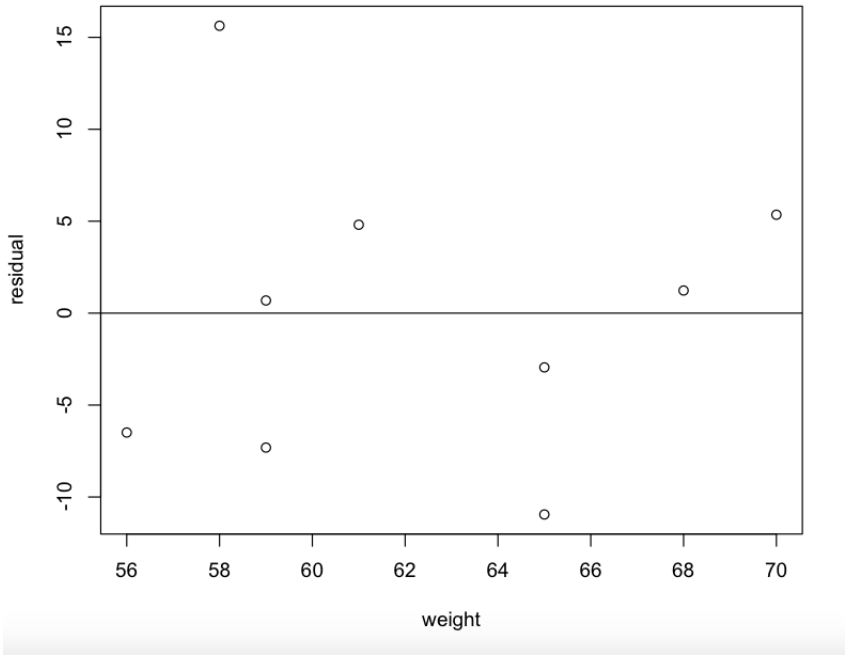
$y = 49.87 + 1.94x$

c) $49.87 + 1.94(60) = 166.27$

d)
```
> plot(weight, height)
> lines(weight, coef(lm(height~weight))[1] +
coef(lm(height~weight))[2]*weight)
```



e)
```
> plot(weight, residual)
> abline(0,0)
```



The plot looks like there is a random dispersion of points suggesting that a linear model is appropriate. Non-random patterns suggest non-linear relationships. Clearly, this is not the case.