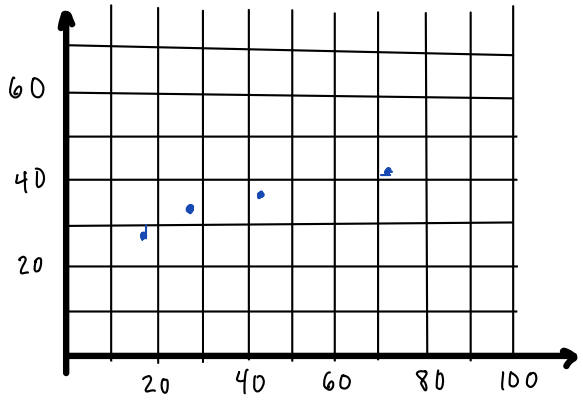
Yellow

1. (a) There appears to be a positive linear association between variables x and y from this data set.



(b) We have

$$\bar{x} = \frac{1}{4}(28+43+17+71) = \frac{1}{4}(159) = 39.75$$

$$\bar{y} = \frac{1}{4}(33+37+28+41) = \frac{1}{4}(139) = 34.75$$

$$s_x^2 = \frac{1}{3}\left[(28-39.75)^2+(43-39.75)^2+(17-39.75)^2+(71-39.75)^2\right] = 547.583$$

$$\Rightarrow s_x \doteq 23.40$$

So, $b = (0.9652)\frac{5.56}{23.4} \doteq 0.229$, and

$a = 34.75 - (0.229)(39.75) \doteq 25.636$

giving us the line $\boxed{\hat{y} = 0.229x + 25.636}$.

(c) The desired residual is

$$\varepsilon = 33 - [(0.229)(28) + 25.636] = \boxed{0.952}$$

2. (a) iv (b) i (c) iii (d) ii

3. (a) It is symmetric, bimodal

(b) median

(c) range $\approx 21 - 6 = 15$

5-number summary: 6, 8.5, 10.5, 13, 21

IQR $\approx 13 - 8.5 = 4.5$

4. Sensitive in this list are the mean, the standard deviation, and the range

5. True statements: (i), (iv) and (vi)

6. (a) `sleep = c(8, 6.5, 7.5, 8.5, 8)`

(b) `names(houses)`

(c) `filter(houses, sqFt > 2000)`

(d) `tally(~ dominantHand | sex, data = personalData)`

(e) `gf_histogram(~ sqFt, data = houses)`

(f) `data(package = "mosaicData")`

7. We are given these probabilities, when a random message from the week is selected:

$P(\text{marked as spam}) = 0.127$

$P(\text{"free"}) = 0.058$

$P(\text{marked as spam and "free"}) = 0.0455$

(a) $$P(\text{"free"} \mid \text{marked as spam}) = \frac{P(\text{marked as spam and "free"})}{P(\text{marked as spam})}$$

$$= \frac{0.0455}{0.127} \doteq \boxed{0.358}$$

(b) $$P(\text{marked as spam} \mid \text{"free"}) = \frac{P(\text{marked as spam and "free"})}{P(\text{"free"})}$$

$$= \frac{0.0455}{0.058} \doteq \boxed{0.784}$$

	Spam	Nonspam	
Free	422	116	538
No Free	756	7982	8738
	1178	8098	9276

8. One can check independence of A and B by seeing if any one of these three equations hold:

i. $P(A \mid B) = P(A)$

ii. $P(B \mid A) = P(B)$

iii. $P(A \text{ and } B) = P(A)\,P(B)$

They all hold simultaneously, or none of them do.

As there are 860 people in this group

$$P(A \text{ and } B) = \frac{45}{860} \doteq 0.0523 \quad \text{and} \quad P(A)P(B) = \frac{86}{860} \cdot \frac{450}{860} \doteq 0.0523.$$

So, iii holds, and A, B are independent.

9. (a) The study is observational in nature. No conditions are imposed on participants by the researchers.

(b) The cases are individuals from a single community, presumably all of them adolescents or teens.

(c) Is there an association between time spent watching TV and the number of aggressive acts committed?

(d) Explanatory variable (quantitative): time spent watching TV
Response variable (quantitative): number of aggressive acts committed

(e) Many possibilities here. For instance:

- level of parental supervision
- number of activities in which the individual participates as a teen