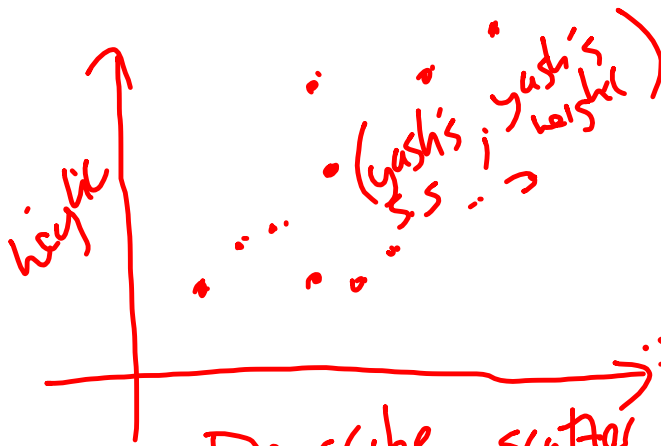


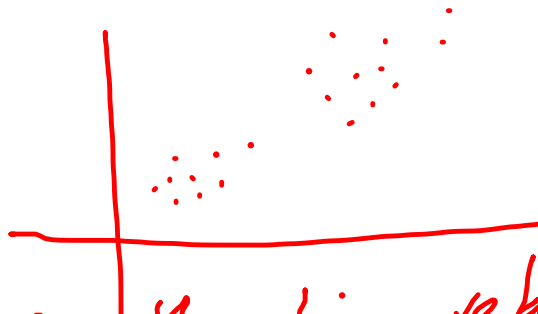
Scatterplots



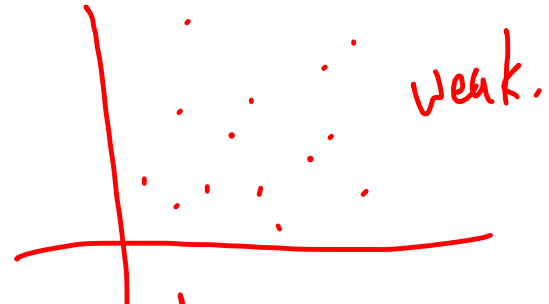
bivariate data
2-variables

Describe scatter plot

- ① positive/negative, association (slope)
- ② clumping



- ③ how strong is the linear relationship



r -correlation coefficient
- measures strength of linear relationship

properties of r

- ① r is unitless.
- ② r is always between $[-1, 1]$
- ③ $-0.5 < r < 0.5$ weak linear relationship
- ④ $0.5 < r < 0.8$ moderate linear relationship
- ⑤ $-0.8 < r < -0.5$
- ⑤ $r = 1$ means all data is collinear w/ positive slope.

when r is near 1 or -1

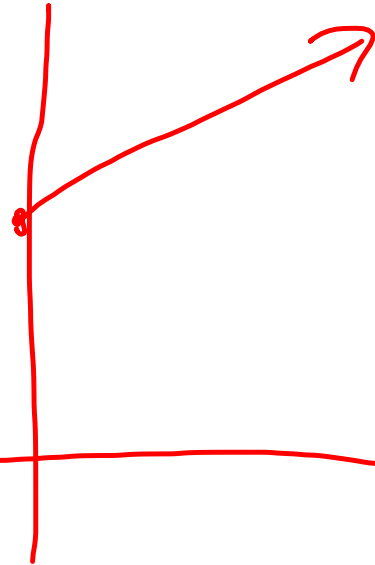
we say x and y are highly
correlated contrast with causation

Ex Hours spent studying a week vs GPA.

Generate Least Squares. Line.

$$\hat{y} = a + bx$$

\uparrow \uparrow
 y-int slope.



1. seeing general trend of data
2. strength of linearity
3. Create a model of some phenomena.

\hat{y} is a predicted value from the model $\hat{y} = a + bx$

2 questions on scatterplots