Agenda

- Warm Up
- Checkup
  - *Check, copies*
- Correlation coefficient "r"   15 min
- Make-a-list   10 min
- Who Dun It?   20 min
  - *Update heights*
  - *Copies of handprint*
- Regression line *y=a+bx*   15 min
- Project time?
  - *Reg project ideas due*
- Exit Pass   10 min
- *Check housing survey*

Warm Up

For each of the following, <u>briefly</u> (~5 words) describe the suspected relationship using DFS. Skip context.

1. Age at death and weekly cigarette consumption
2. Distance to destination and cost of flight
3. Number of letters in a person's name and their shoe size
4. Miles driven and price for used cars
5. Price and number of pages of a textbook

Checkup time

Correlation coefficient "r"

- Measures the **direction** and **strength** of the association between two quantitative variables

  $-1 \leq r \leq 1$

    - *-1 → Perfect negative*

    - *0 → No association at all*

    - *1 → Perfect positive*

- **Warnings:**

    – Not resistant to outliers.

    – Does not guarantee linearity. Use only with a graph.

    – Does not imply causation.

Window/Door

For each of the following pairs of variables, guess the correlation coefficient (r).

1. Age at death and weekly cigarette consumption
2. Distance to destination and cost of flight
3. Number of letters in a person's name and their shoe size
4. Miles driven and price for used cars
5. Price and number of pages of a textbook

Calculating *r*

$$r = \frac{\sum_{i=1}^{n}(z_x)(z_y)}{n}$$

$$r = \frac{\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{s_x}\right)\left(\frac{y_i - \overline{y}}{s_y}\right)}{n}$$

Calculators and r

- 2nd: Catalog: DiagnosticOn (Enter)
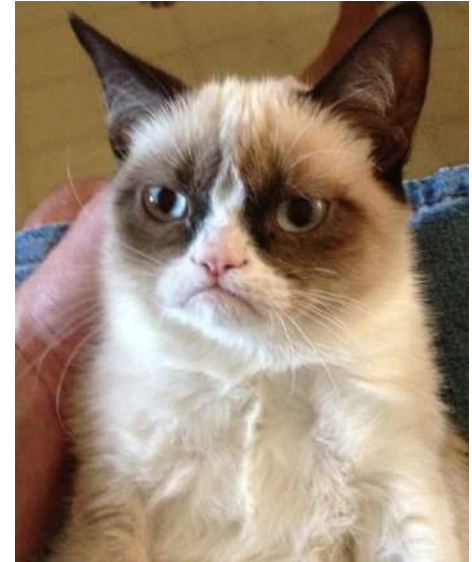- STAT → CALC → 4:LinReg(ax+b)

Explanatory	Response
1	5
2	6
3	3
3	5
7	1

Make-a-list

- Pairs of variables with a strong positive association.

- Pairs of variables with a strong negative association.

- Pairs of variables with no association.

Who dun it?

| TEACHER	HEIGHT (in.)
Saepan	67
Cole	64
VanBuskirk	74
Frantz	70
Ryan	65
Hook	69
Tan	61
Ceo	68.5
Hwang	61
Colligan	75.5
Burton	64
Okita	66

Who dun it? (P.2)

1. Use a calculator to create a scatterplot of this class' heights and hand-spans. Interpret with DFS.

2. Predict the murderer's identity. Justify your answer.

Handspan	Height
11	60
10	71
9.5	75
9.5	74
9	72
8.5	68
8.5	71
8	66
8	67
7.5	63
7.5	61.5
7.5	64
7	63
7	65
7	60
6.5	62
6.5	60
6	65
6	63
6	64

P.2
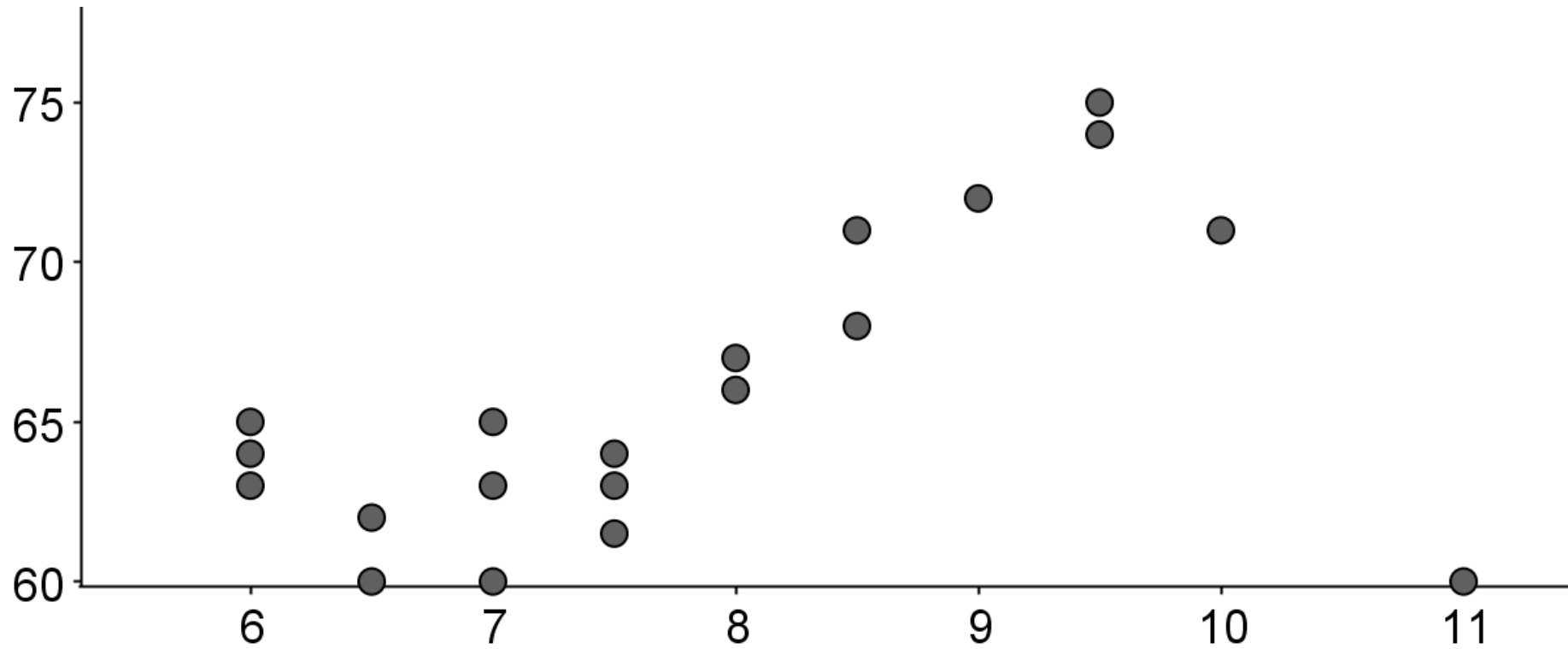
P.2
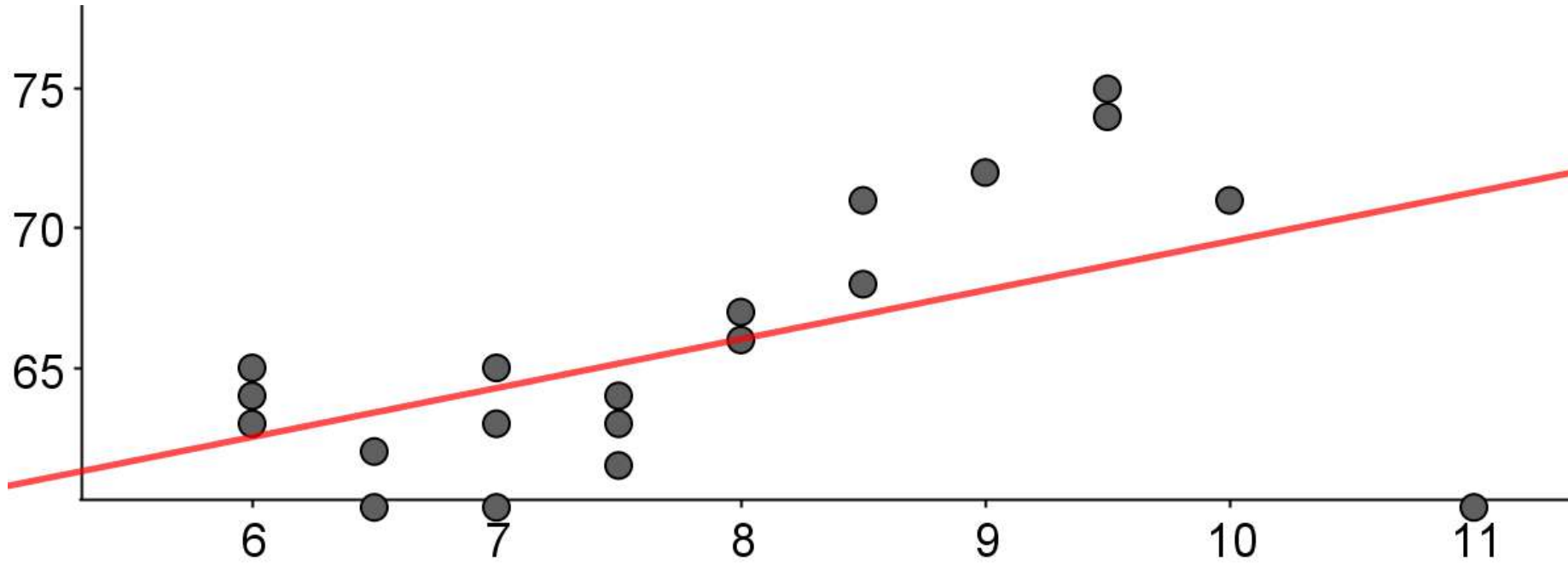
Regression Line

$$= \quad +$$

- Also called "least-squares line", "line of best fit"
  - Minimizes the "sum of the squared deviations" of predicted and observed values

- Describes how a response variable *y* changes as an explanatory variable *x* changes Identify *x* and *y*!
  - Used to predict *y*, given a value for *x*
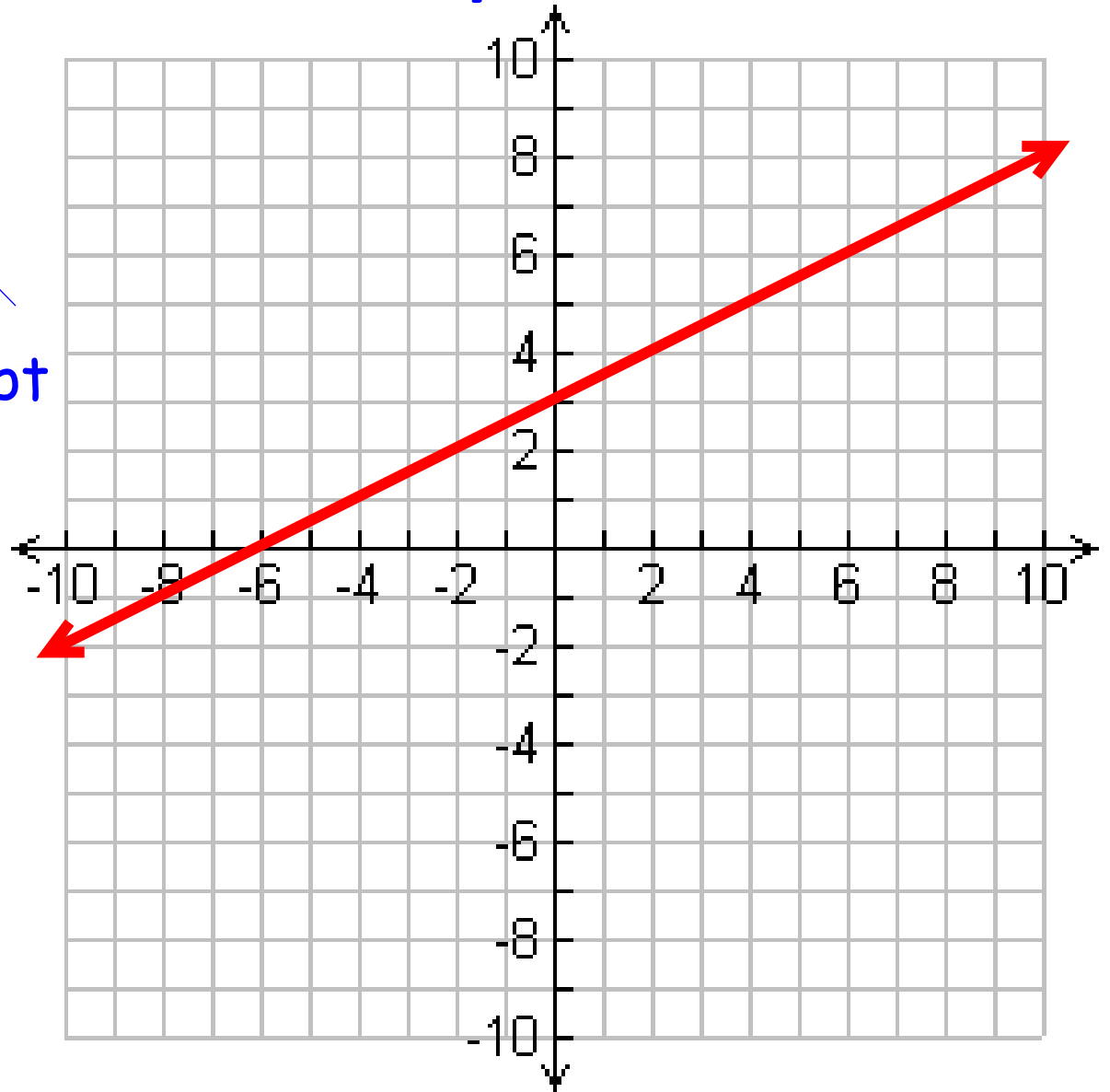
It's... *Slope*-*Intercept* *Form!*

$$y = \tfrac{1}{2}x + 3$$

slope

y-intercept

1/2

3

1. What **is** slope?

2. What **is** the y-intercept?

Regression Line

$$ \boxed{ = \qquad + }$$

- Also called "Least-squares line", "line of best fit"
  - Minimizes the "sum of the squared deviations" of predicted and observed values

- Describes how a response variable *y* changes as an explanatory variable *x* changes

  Identify *x* and *y*!
  - Used to predict *y*, given a value for *x*

- *a* → *slope or "constant"*

  - "For every 1 [x], [y] goes up by _____."

- *b* → *y-intercept or "amount"*

  - "When [x] is 0, [y] is _____."

- EXAMPLE. Handspan and height (Geogebra)

P.2
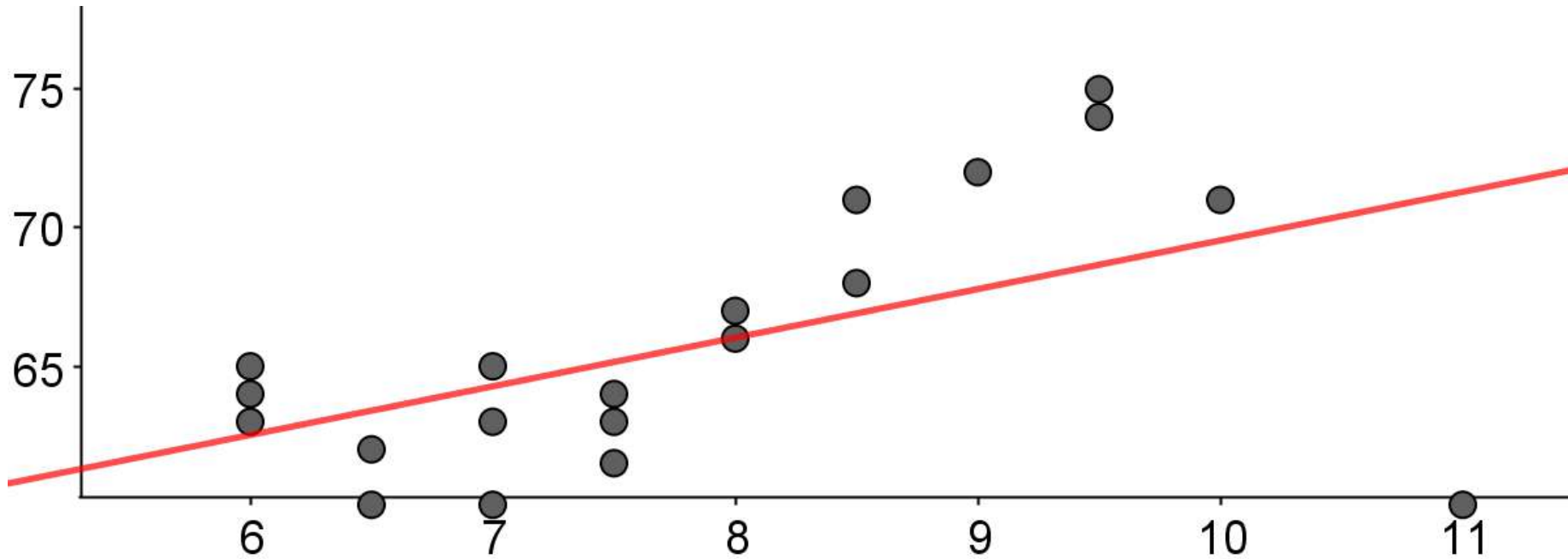
$y = 1.75x + 52$

Who dun it?

TEACHER	HEIGHT (in.)
Saepan	67
Cole	64
VanBuskirk	74
Frantz	70
Ryan	65
Hook	69
Tan	61
Ceo	68.5
Hwang	61
Colligan	75.5
Burton	64
Okita	66

All students
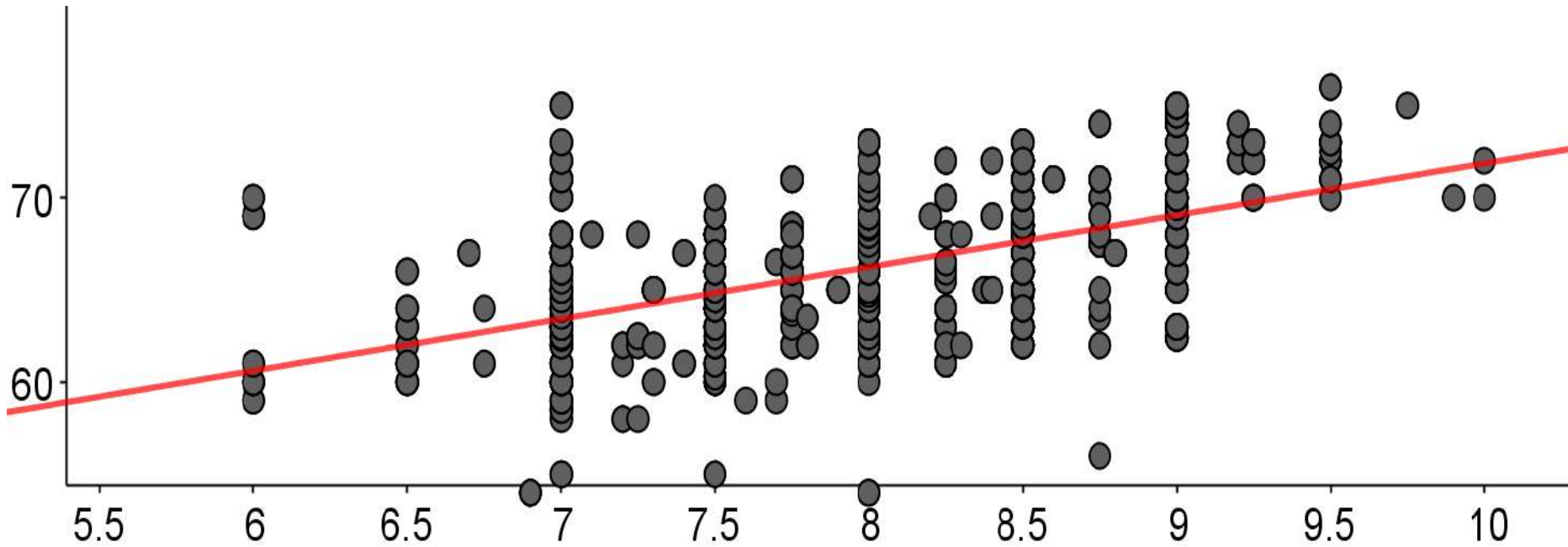
y = 2.81x + 43.118

Who dun it?

TEACHER	HEIGHT (in.)
Saepan	67
Cole	64
VanBuskirk	74
Frantz	70
Ryan	65
Hook	69
Tan	61
Ceo	68.5
Hwang	61
Colligan	75.5
Burton	64
Okita	66

Window/Door

Briefly describe what the <u>slope</u> and <u>y-intercept</u> mean in context.

1. On your recent Unit 2 test, the relationship between your expected and actual scores was y = 0.5781x + 11.007, where expected scores is your explanatory variable.

2. From your survey data, the relationship between hours of sleep and GPA is y = -0.0028x + 3.6526, where sleep is your explanatory variable.

3. From your survey, the relationship between the hours you spent on a computer the last day of summer, and average hours studying per week, is y = 0.4897x + 5.6385, where hours on computer is the explanatory variable.

4. In 2012, the relationship between birth and death rates per 1000 people in all fifty US states was y = -0.3362x + 13.1295, where birth rates is the explanatory variable.

Project #2: Predicting the Uncertain

- **Today 1/31:**
  - Pick 3 project ideas. Why are they good ideas? How will you collect data?

- **Wednesday 2/5:**
  - Collect a sample of 25+ <u>pairs</u> of data. Describe your process.
  - If you collected your own data, describe your process, and how you tried to make the process fair and unbiased.

- **Friday 2/7:**
  - Create numerical summaries for <u>each</u> of your two sets of data.
  - Create 3 separate graphs → 1) a graph for your explanatory variable, 2) a graph for your response variable, 3) a scatterplot for both variables.

- **Monday 2/10:**
  - Write an analysis of your data. Comment on your scatterplot. Identify/ interpret r, and the least-squares equation. Interpret your residual plot. Comment broadly on your data.

You may do a poster instead of a report.

Exit Pass

1. Construct a scatterplot of the data below.

2. Interpret the scatterplot.

SAT Math	700	700	600	660	500	600	730	540
SAT Writing	720	650	600	610	550	690	680	570

tinyurl.com/602housesurvey