Data Interpretation

Anthony Chau

7/24/2021 (updated: 2021-07-25)

Data Intrepretation

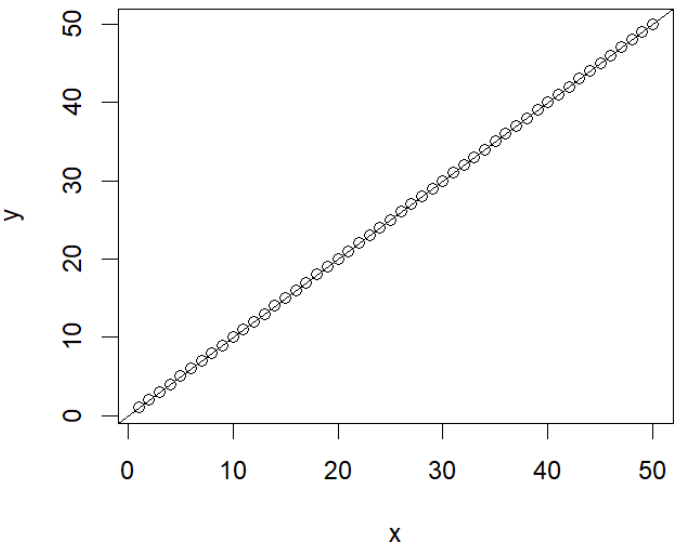
- After visualizing our data, it's useful to provide an interpretation of the visualization
- We'll focus on how to describe the relationship between two quantitative variables
- Three ways to describe relationship: linearity, strength, and direction

Linear relationships

- The relationship between two quantitative variables can be called *linearly related* if the distribution of the variables follow a straight line
- Now, it is rare for real data to have a perfectly linear relationship
- It's more likely that the relationship between two variablesroughly follows a straight line
- The opposite of a linear relationship is a *non-linear* relationship. For example, two variables may be quadratrically related.

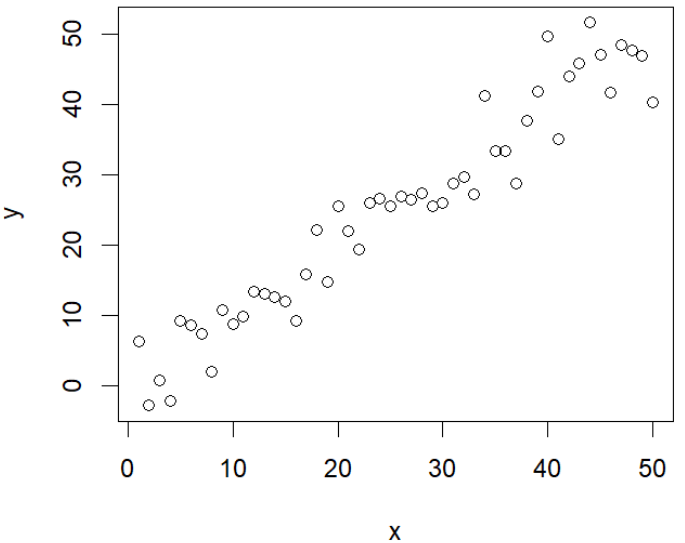
Perfect linearity

```r
x ← c(1:50)
y ← c(1:50)
# plot points
plot(x, y)
# add 45 degree line
abline(a = 0,
       b = 1)
```



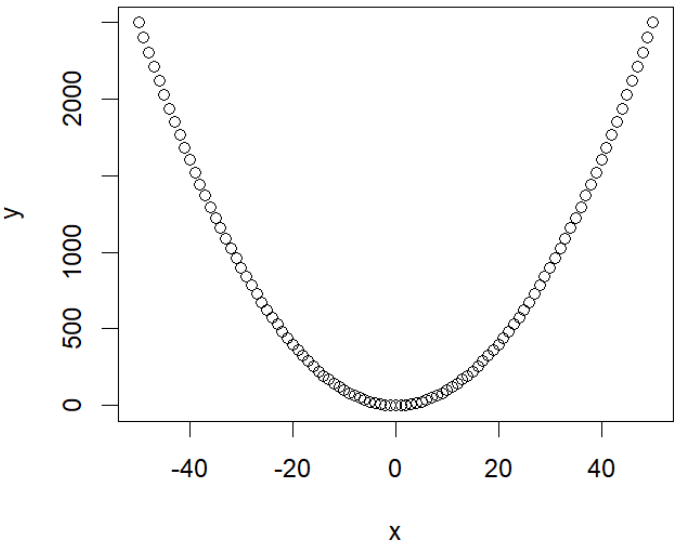
General Linearity

```
x ← c(1:50)
y ← c(1:50) +
  rnorm(n = 50,
        mean = 0,
        sd = 4)
plot(x, y)
```



Non-Linear Relationship

```
x ← c(-50:50)
y ← x^2
plot(x, y)
```

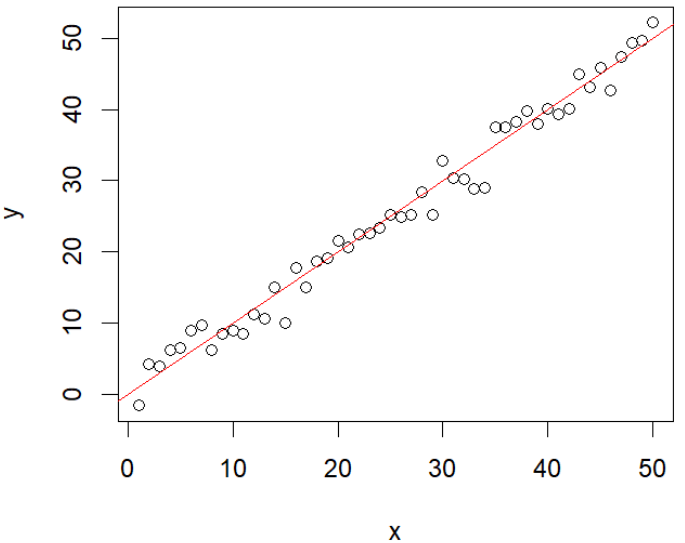


Strength of relationship

- We can also describe the strength of the relationship between two variables
- The strength of the relationship can be thought of as how closely two variables follow each other.
- If two variables line up exactly on a 45-degree line --> *perfect relationship*
- If two variables tend to cluster closely together --> *strong relationship*
- If two variables tend to spread out --> *weak relationship*
- If two variables make a horizontal line --> *no relationship*

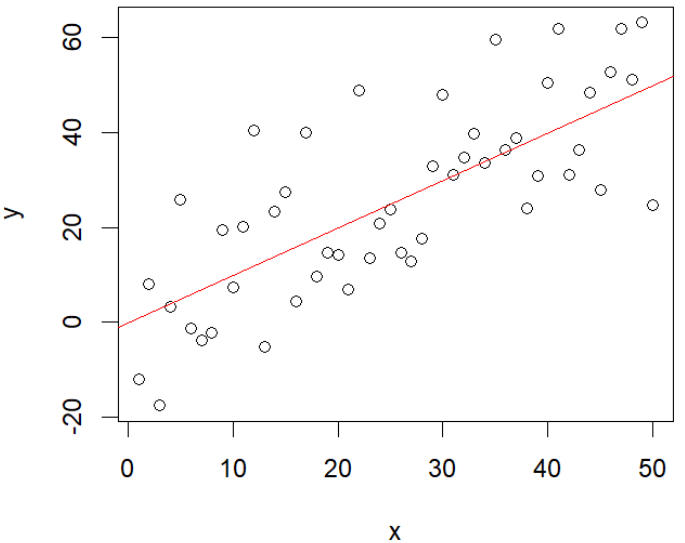
Strong relationship

```r
x ← c(1:50)
y ← c(1:50) +
  rnorm(n = 50,
        mean = 0,
        sd = 2)
plot(x, y)
abline(a = 0,
       b = 1,
       col = "red")
```



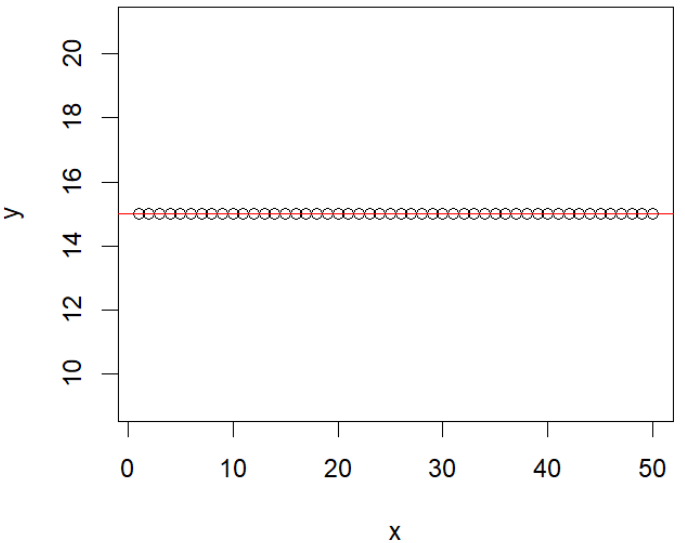
Weak relationship

```r
x ← c(1:50)
y ← c(1:50) +
  rnorm(n = 50,
        mean = 0,
        sd = 13)
plot(x, y)
abline(a = 0,
       b = 1,
       col = "red")
```



No relationship

```r
x ← c(1:50)
y ← rep(15, 50)
plot(x, y)
abline(a = 15,
       b = 0,
       col = "red")
```

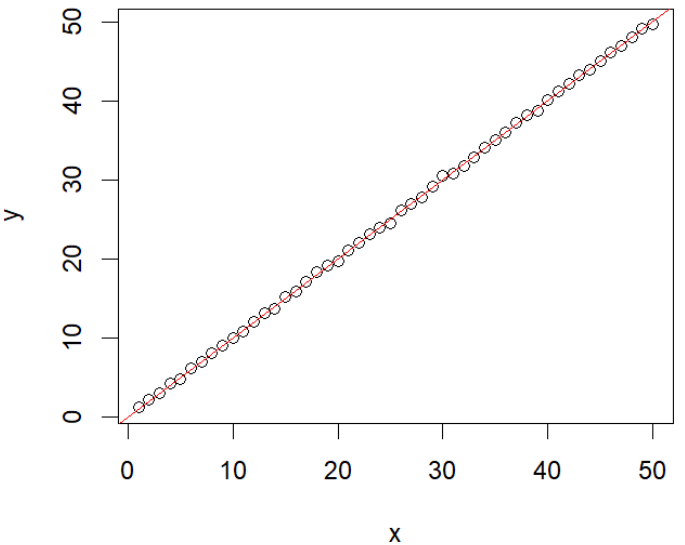


Direction of relationship

- Lastly, we can describe the direction of relationship between two variables
- If one variable increases as the other variable increases --> *positive relationship*
- If one variable decreases as the other variable increases --> *negative relationship*

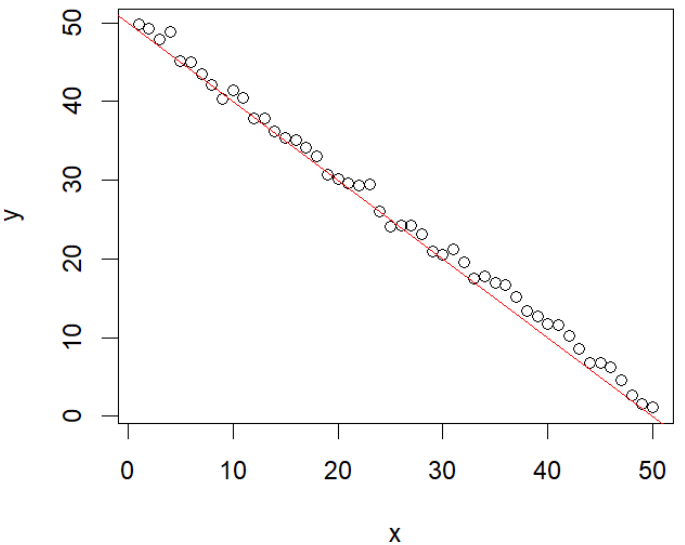
Positive relationship

```r
x ← c(1:50)
y ← c(1:50) +
  rnorm(50,
        mean = 0,
        sd = 0.2)
plot(x, y)
abline(a = 0,
       b = 1,
       col = "red")
```



Negative relationship

```r
x ← c(1:50)
y ← c(50:1) +
  rnorm(n = 50,
        mean = 0,
        sd = 1)
plot(x, y)
abline(a = 50,
       b = -1,
       col = "red")
```



Putting it all together

- We can describe the relationship between two variables by combining descriptions for the linearity, strength, and direction of a relationship
- Example: the relationship between height and weight is a moderate, positive and linear relationship.
- Example: the relationship between temperature and air conditioning use is a strong, negative, and linear relationship.