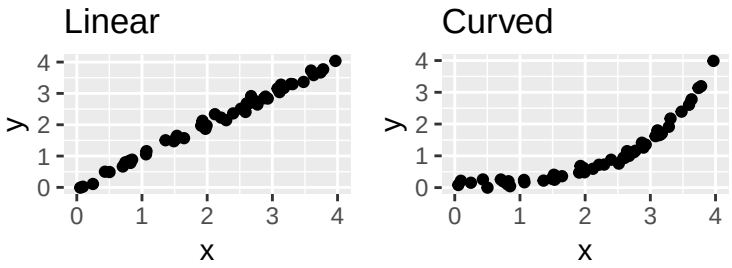
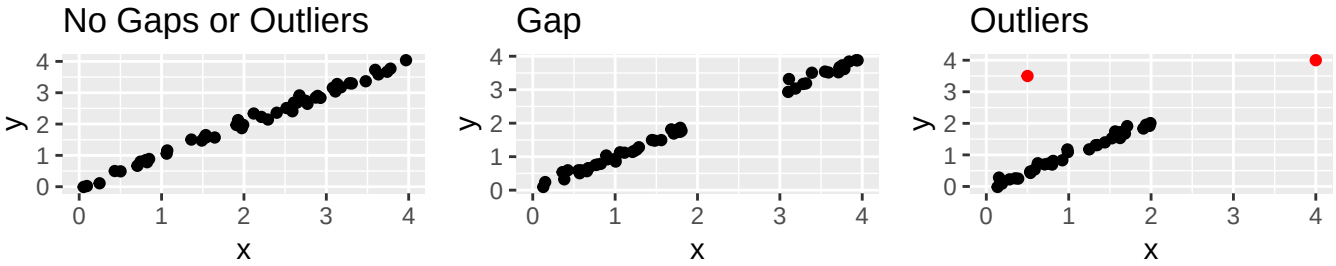
Scatter Plots and Correlation: Summary

Describing the Relationship Between 2 Quantitative Variables

A. Shape (linear or curved)



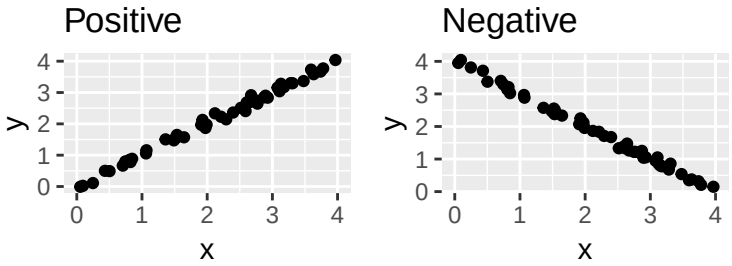
B. Are there any gaps or outliers?



C. Direction (positive or negative association)

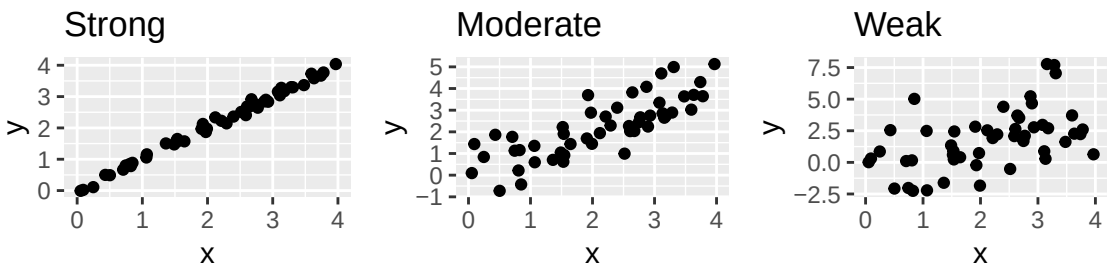
Positive association: as one variable increases, the other also tends to increase.

Negative association: as one variable increases, the other tends to decrease.



D. Strength (weak, moderate, or strong)

The relationship is strong if the points fall close to the trend, and weak if they do not.



Correlation

The correlation is **only useful if the relationship is linear and there are no outliers**.

If those conditions are met, it summarizes:

- **Strength**:
  - Strong relationship if correlation is close to 1 or -1
  - Weak relationship if correlation is close to 0
- **Direction**:
  - Positive association if correlation is positive
  - Negative association if correlation is negative.

Example of Calculation in R

Suppose we want to calculate the correlation between `petal_length` and `petal_width` in the following data set with measurements on 150 iris flowers:

```
head(iris)
```

```
##   sepal_length sepal_width petal_length petal_width species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

We can do this with the following code:

```
iris %>%
  select(petal_length, petal_width) %>%
  cor()
```

```
##              petal_length petal_width
## petal_length    1.0000000   0.9628654
## petal_width     0.9628654   1.0000000
```

The correlation between `petal_length` and `petal_width` is about 0.963. This indicates a strong, positive association between these variables. We'd need to check a scatter plot to be sure it was a useful summary of the relationship though!