JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Tables and Graphs

Marie Diener-West, PhD
Johns Hopkins University

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

An Introduction to Tables and Graphs

Tables and Graphs

- Data may be summarized and presented in tables
- Data may be displayed in graphs

Suggestions for Presenting Data

- In **tables**
  - Round numbers or use significant figures
  - Use summary values (averages or totals)
  - Pay attention to order, spacing, and layout
  - Provide clear labels for titles and column/row headings
- In **graphs**
  - Show the data in a clear fashion
  - Avoid distorting the data
  - Do not change the scale mid-axis
  - Use precise labels for titles, axes, legends, and footnotes

Purposes of Graphing

- Visually explore the data
- Identify trends in the data
  - Linear trends
  - Exponential trends

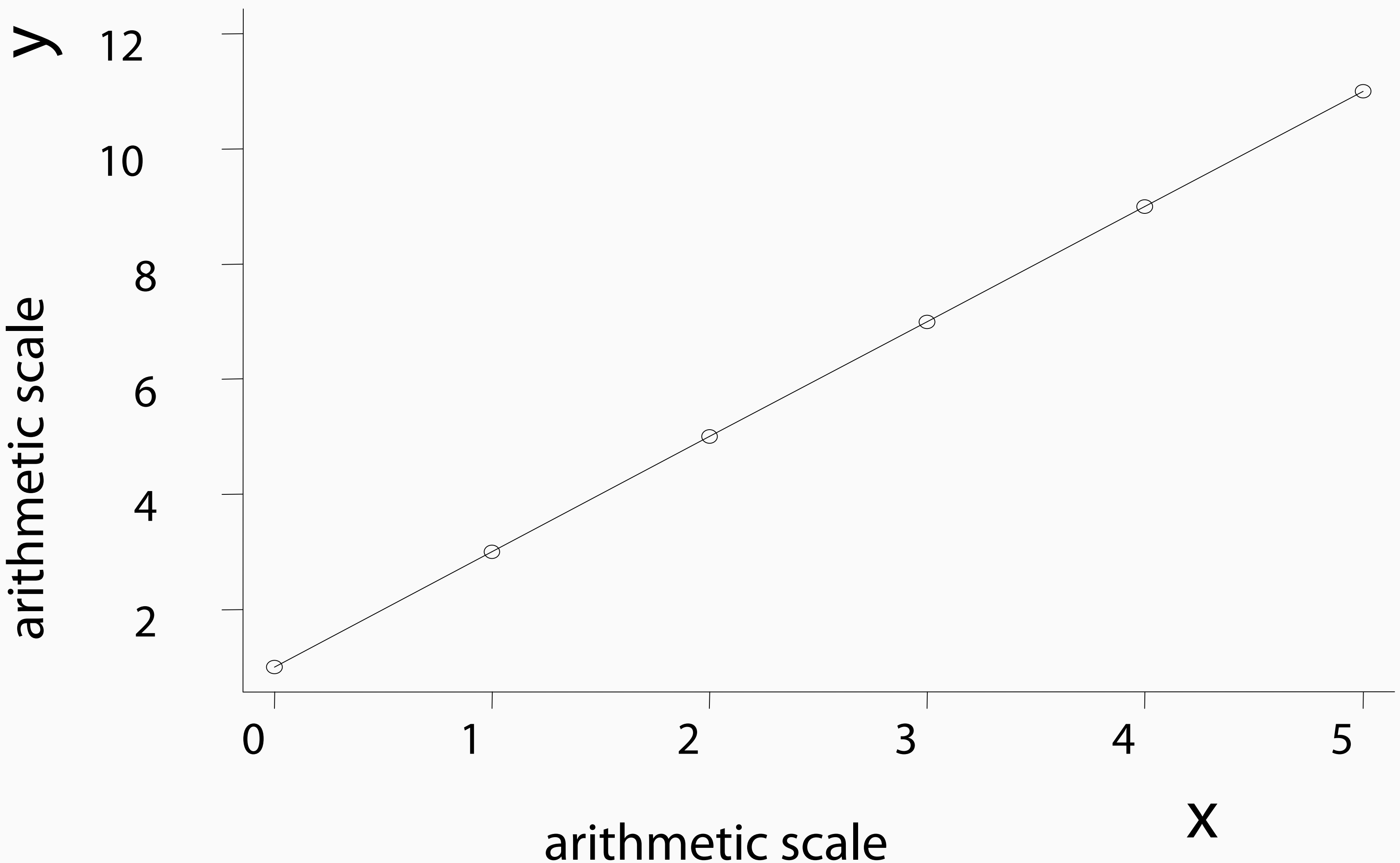
Graphing on Arithmetic Paper

- The formula for a straight line on arithmetic paper is **$y = ax + b$**
- The **slope (a)**
  - Is the change in y divided by the change in x
- The **y-intercept (b)**
  - Is the value of y when $x = 0$

Linear Trend

- A linear relationship produces a straight line on arithmetic paper (indicating an increase by the same number in y per unit increase in x)
- Each increment represents change by a constant amount

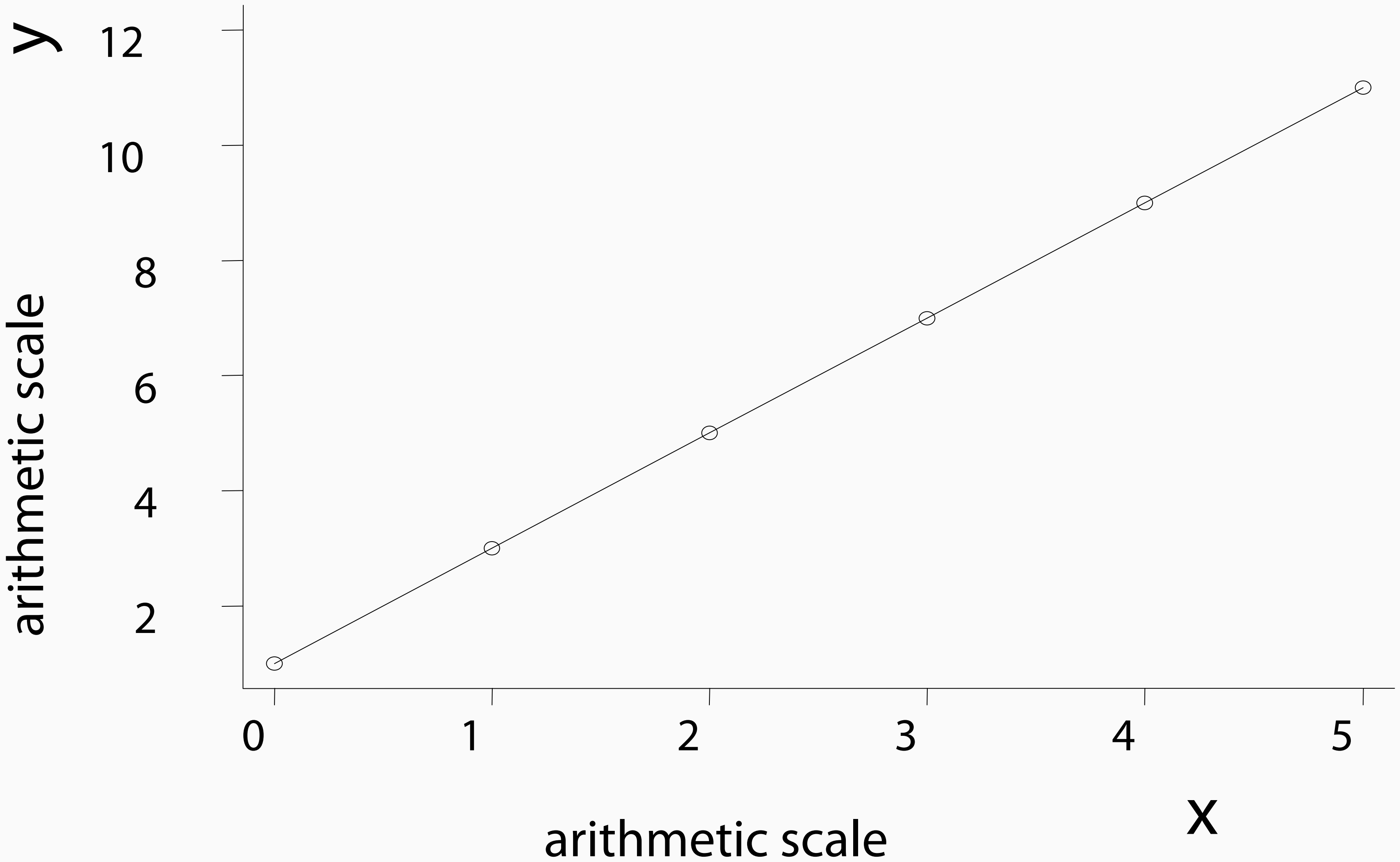
Simple Example of 5 Data Points



Example 1

- Suppose there were five points (x,y):
  (1,3) (2,5) (3,7) (4,9) (5,11)
- The slope can be calculated as 2
  - [e.g., (5 – 3) / (2 – 1)]
- The y-intercept is 1 (y=1 when x=0)
- The line can be written $y = 2x + 1$
- An equation of the form $y = 2x + 1$ indicates a linear relationship between x and y
- Plotting on arithmetic paper (arithmetic scale for y, arithmetic scale for x) shows this trend

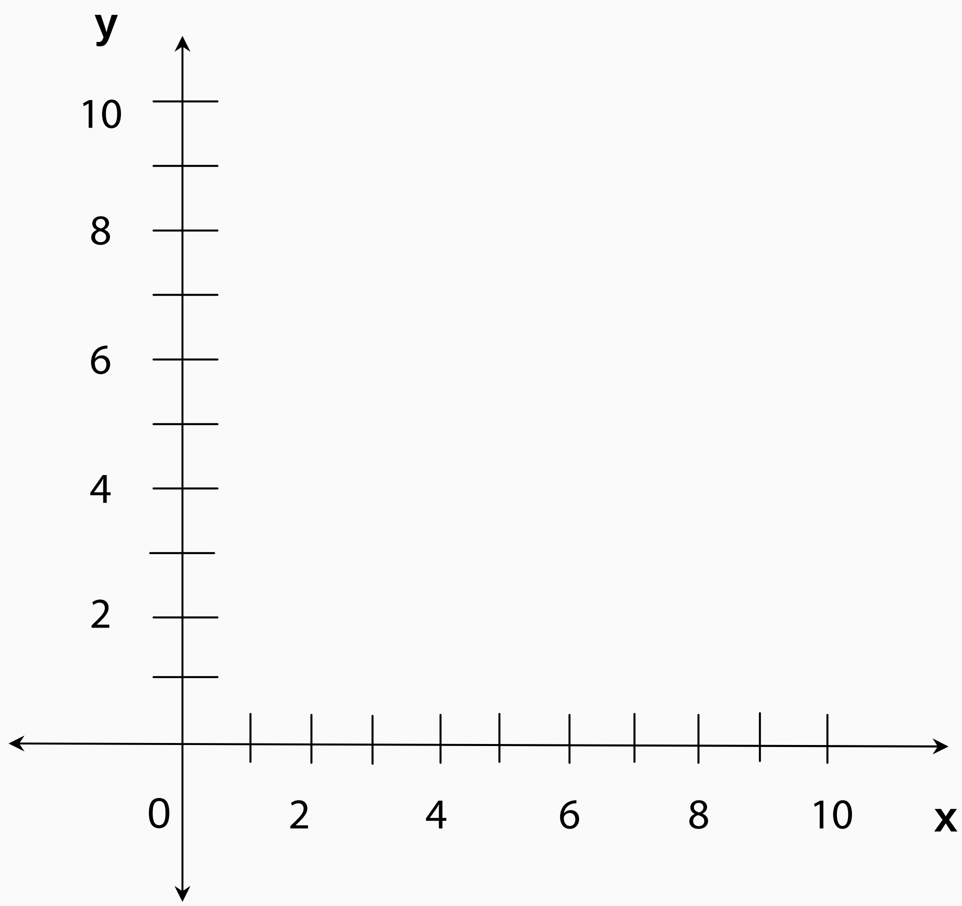
Linear Trend on an Arithmetic Scale for Y



Example 1

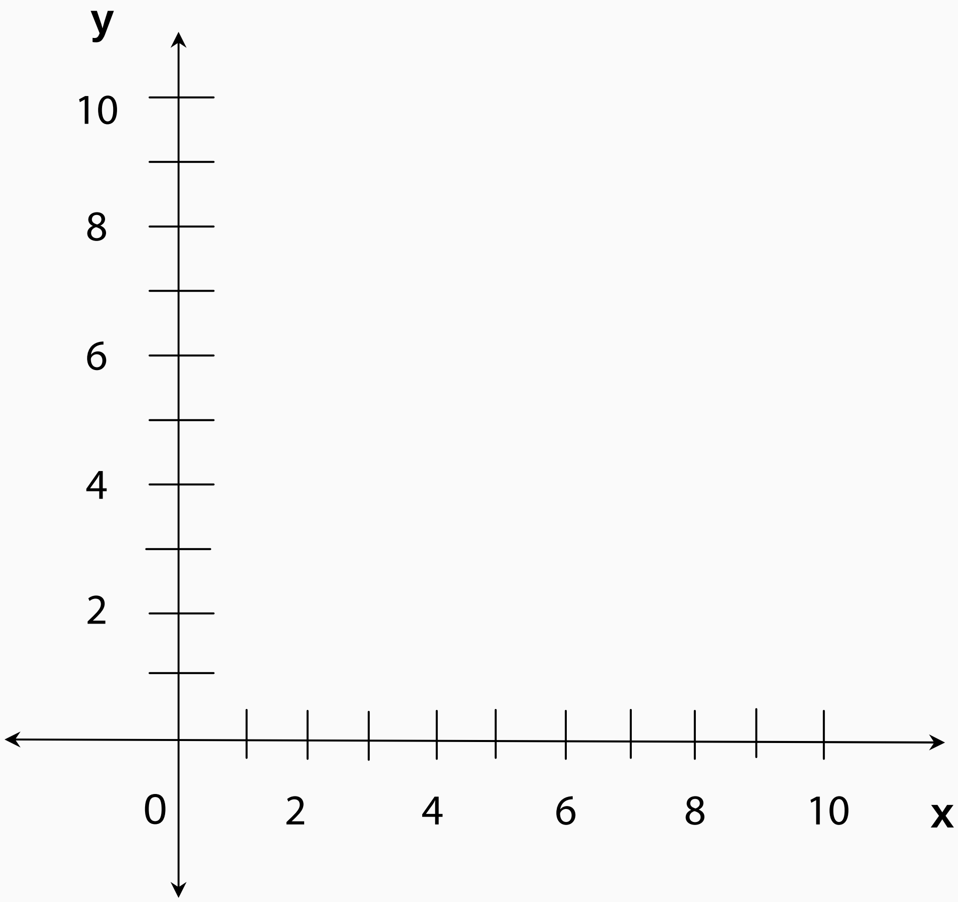
- The straight line relationship with a slope of 2 indicates that for every increase of one unit in x, y increases by two units

Determining the Slope of a Straight Line



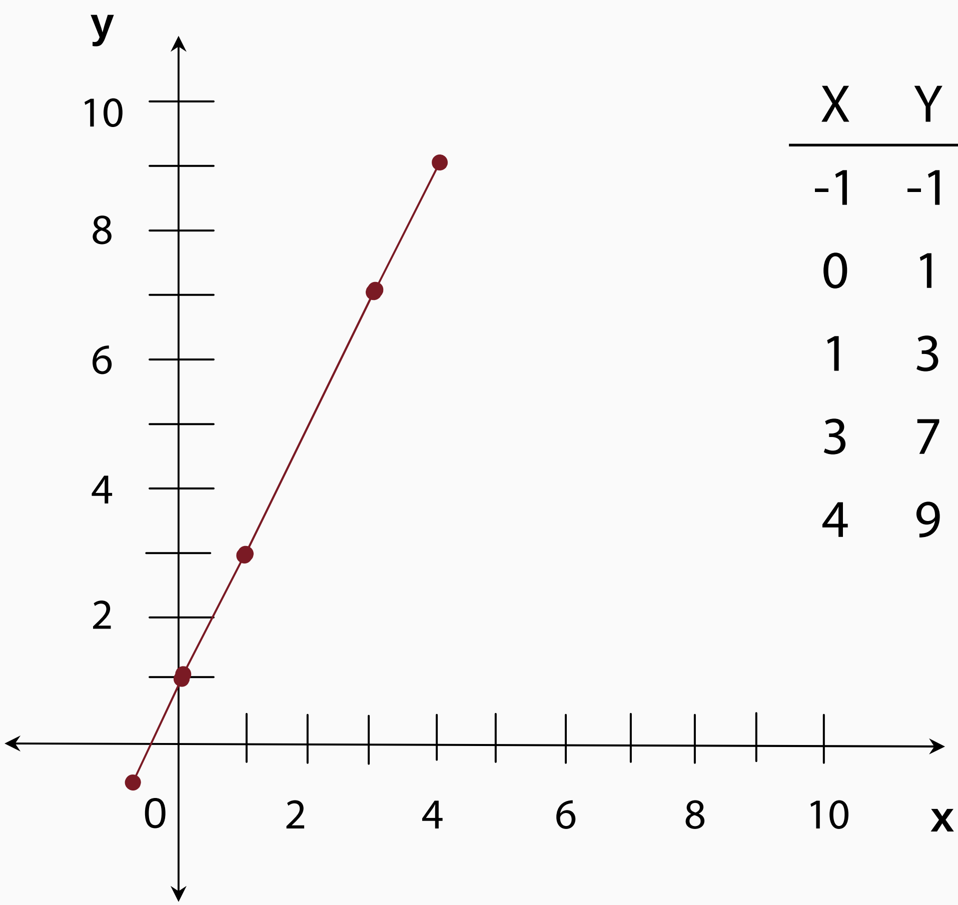
Determining the Slope of a Straight Line

- The formula for determining the slope of a straight line is $y = ax + b$ where a = slope and b = intercept



Determining the Slope of a Straight Line

- If you graph the following set of points, you can determine both the intercept and the slope

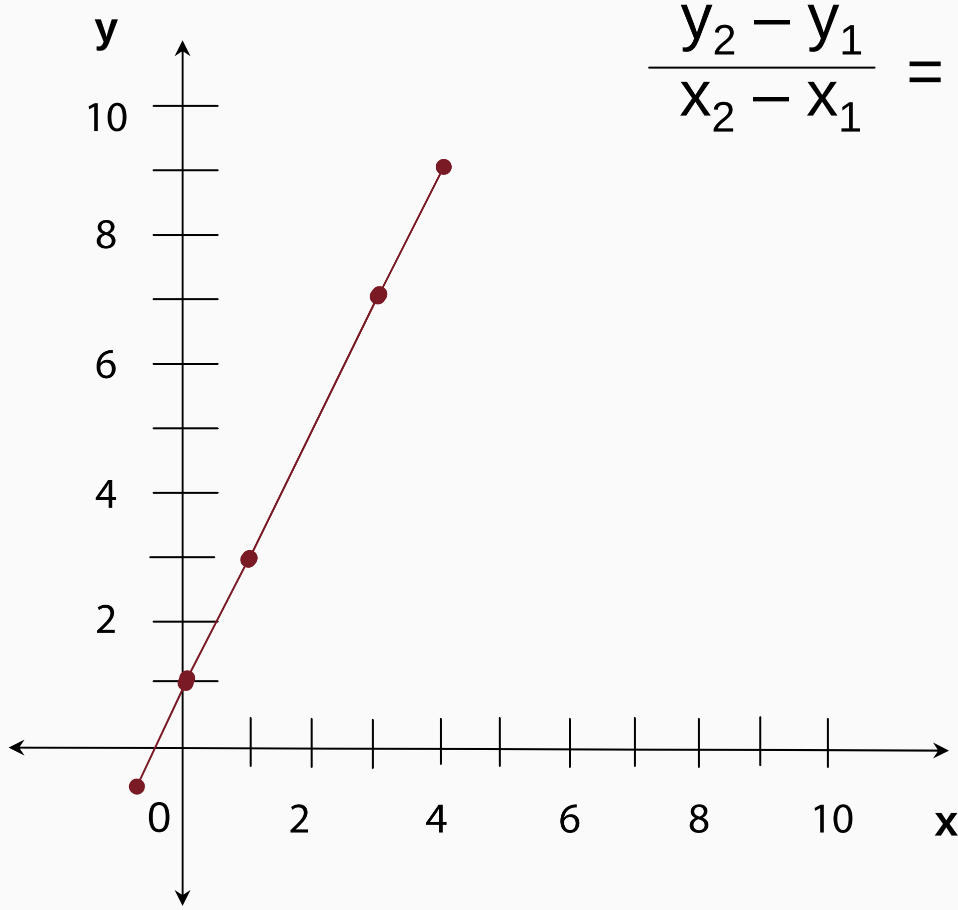


X	Y
-1	-1
0	1
1	3
3	7
4	9

Determining the Slope of a Straight Line

- You can determine the slope of the line with the following equation

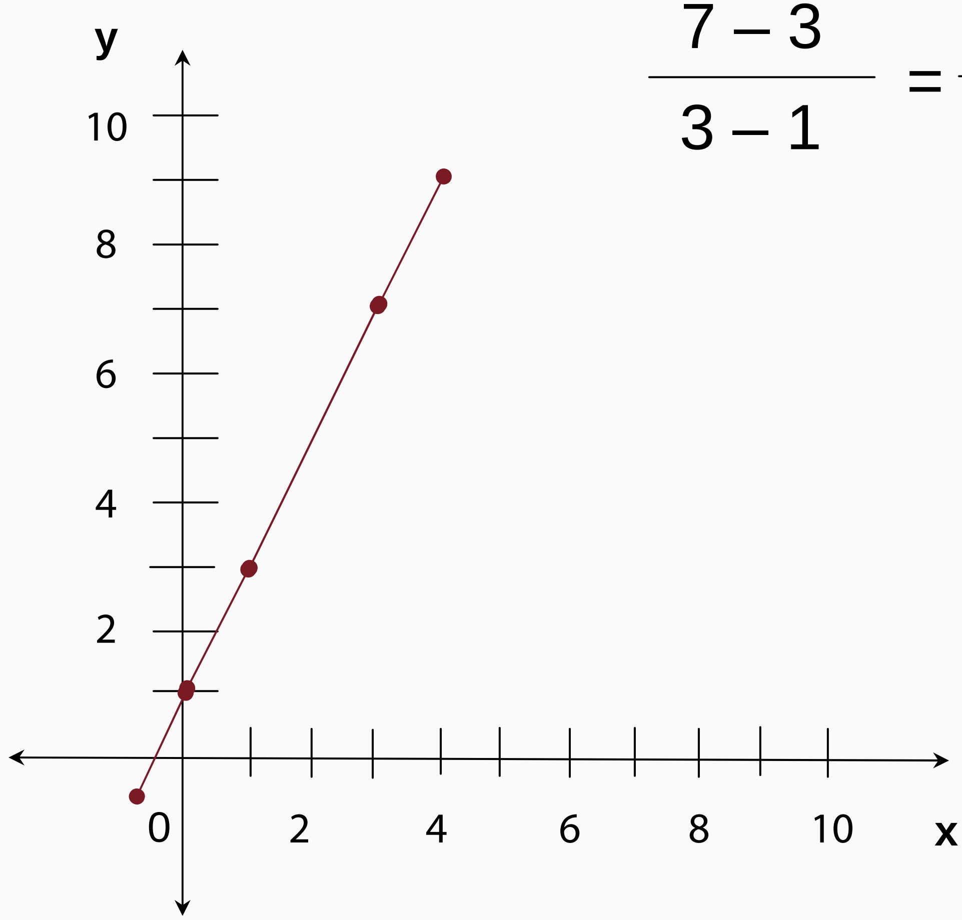
$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{\Delta y}{\Delta x}$$



Determining the Slope of a Straight Line

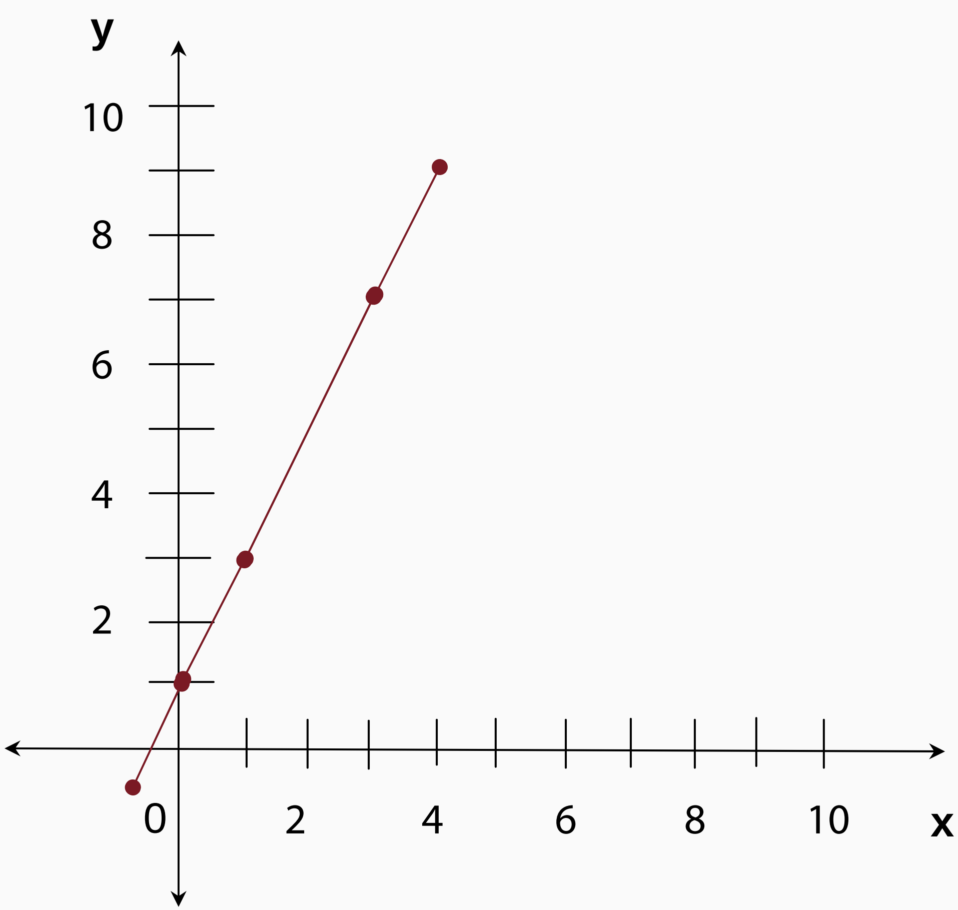
- You can determine the slope of the line with the following equation

$$\frac{7-3}{3-1} = \frac{4}{2} = 2$$



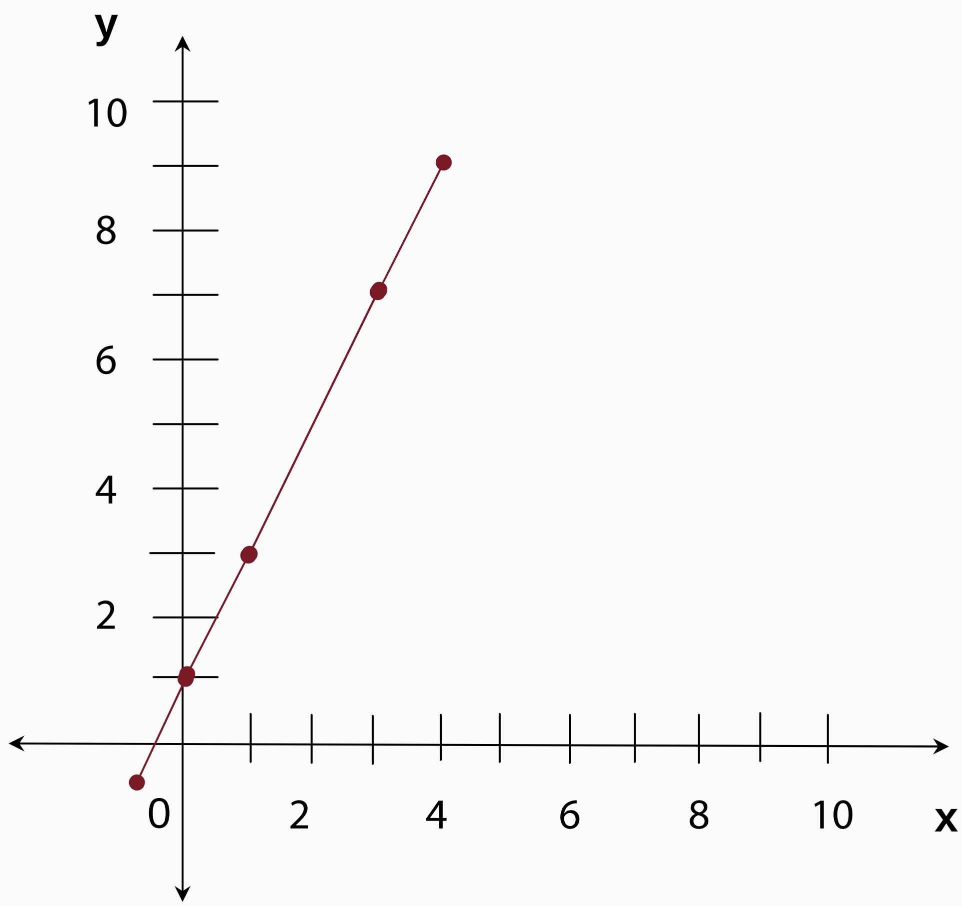
Determining the Slope of a Straight Line

- The intercept is the value of y when x = 0
- Intercept = b = 1



Determining the Slope of a Straight Line

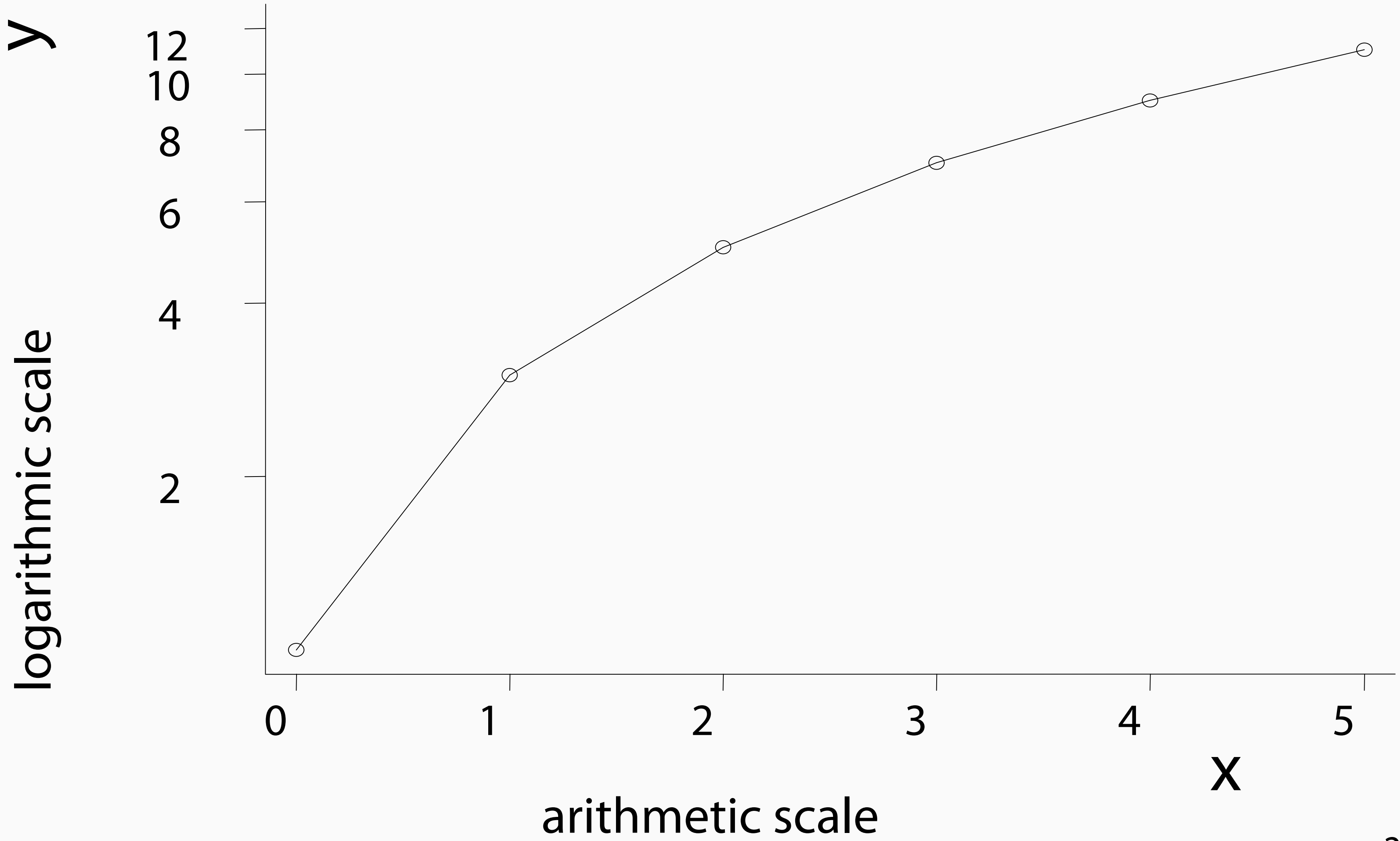
- The resulting equation in the format y = ax + b, looks like this: **y = 2x + 1**



Example 1

- Plotting the same data points on semi-logarithmic paper (logarithmic scale for y, arithmetic scale for x) does not produce a straight line

Nonlinear Trend on a Logarithmic Scale for Y



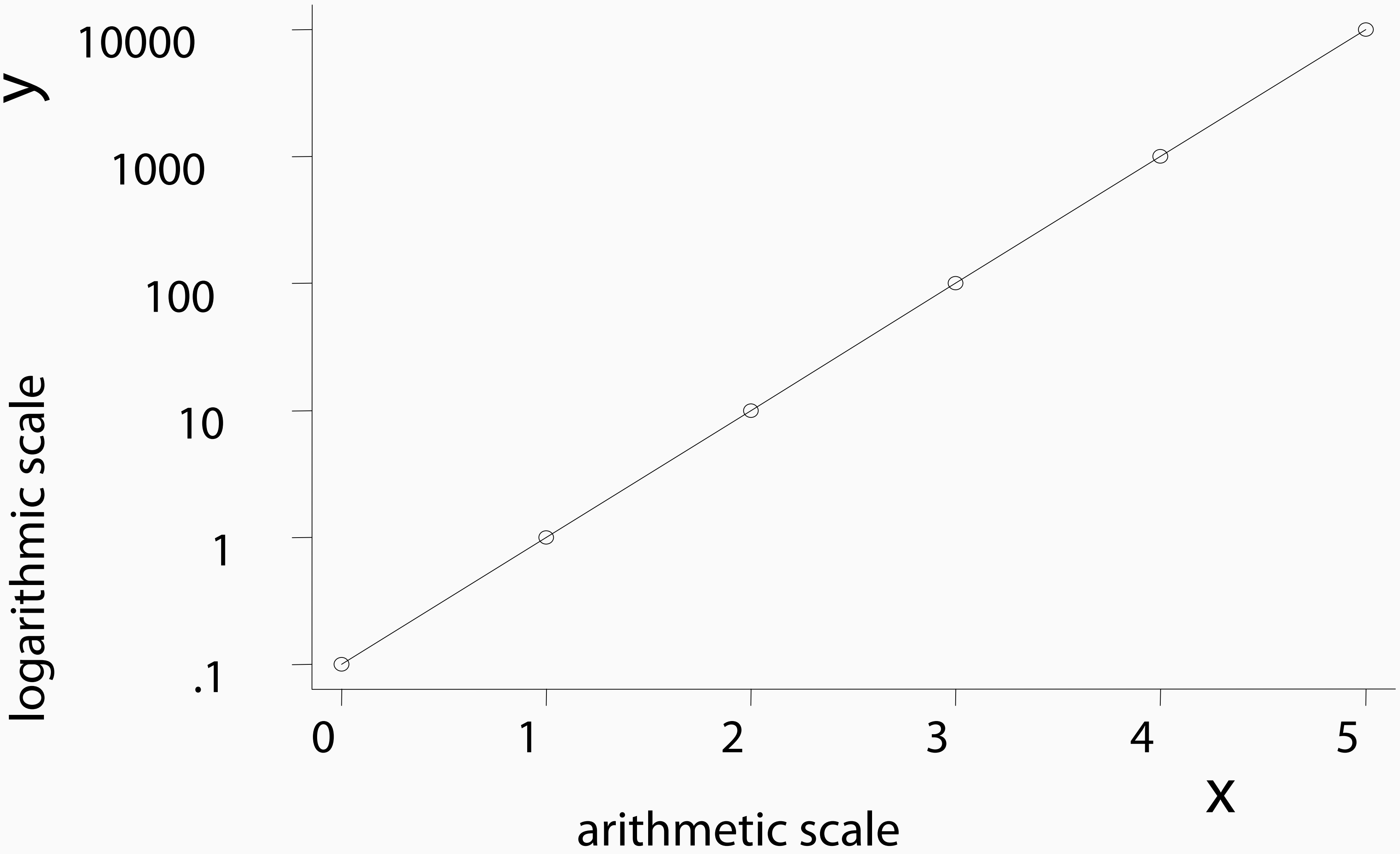
Exponential Trend

- A relationship that is linear on the log scale produces a straight line on semi-logarithmic paper (indicating an increase in y by the same proportion per unit increase in x)
- Each increment represents change by a constant factor

Example 2

- Suppose there were six points (x,y):
  (0, 0.1) (1, 1) (2,10) (3,100) (4,1000) and (5,10000)
- Plotting these points on semi-logarithmic paper results in a straight line
- For every increase of one unit in x, y increases by a constant factor of 10

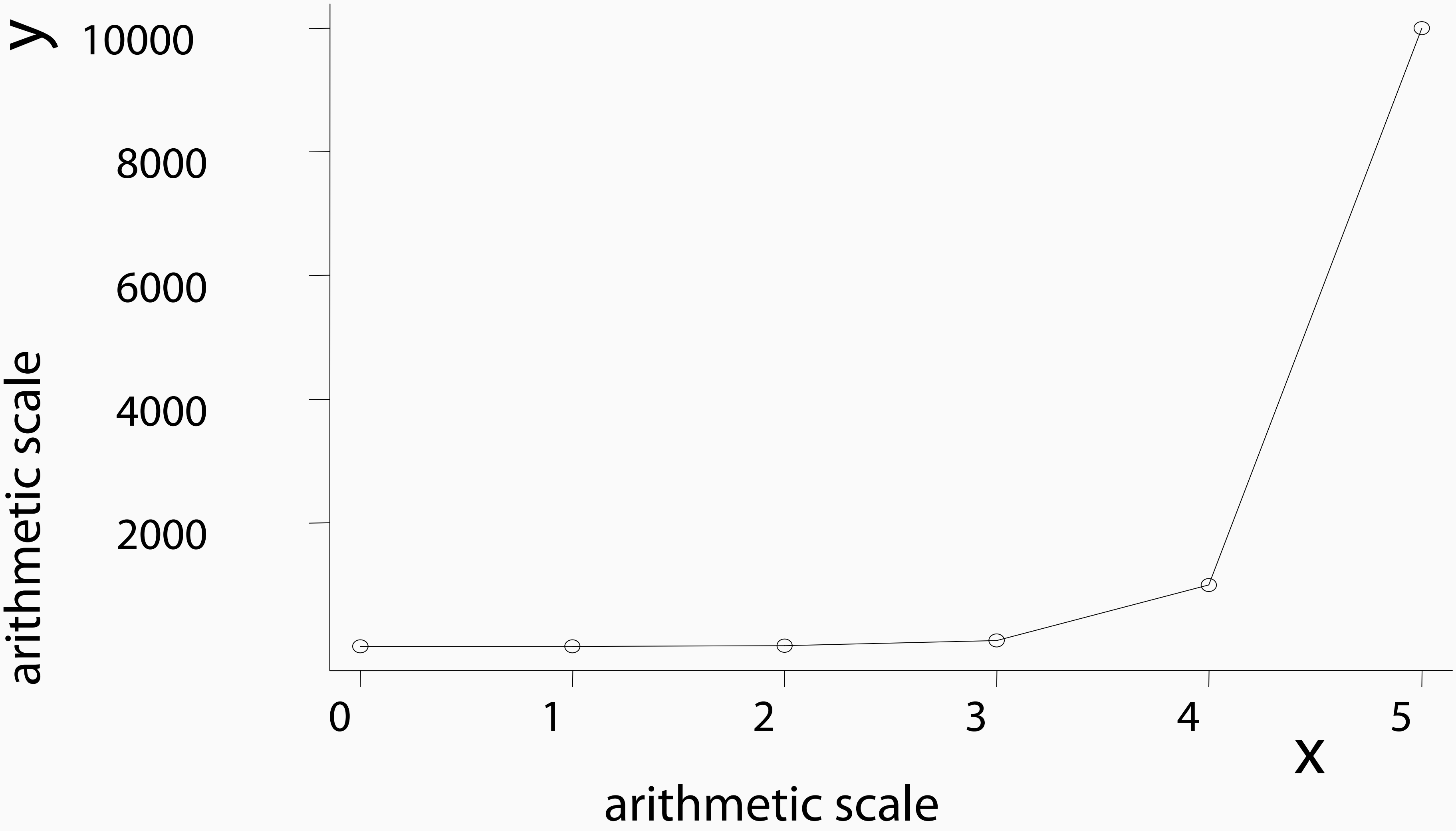
Linear Trend on a Log Scale for Y



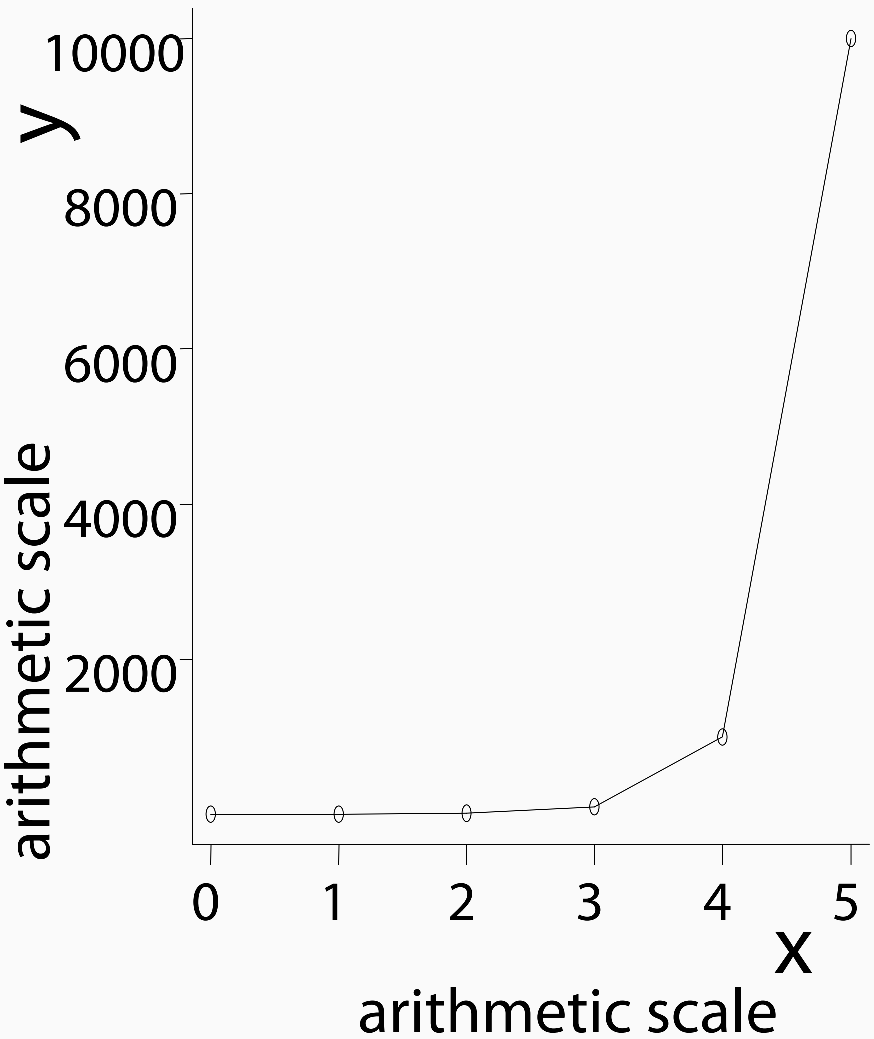
Example 2

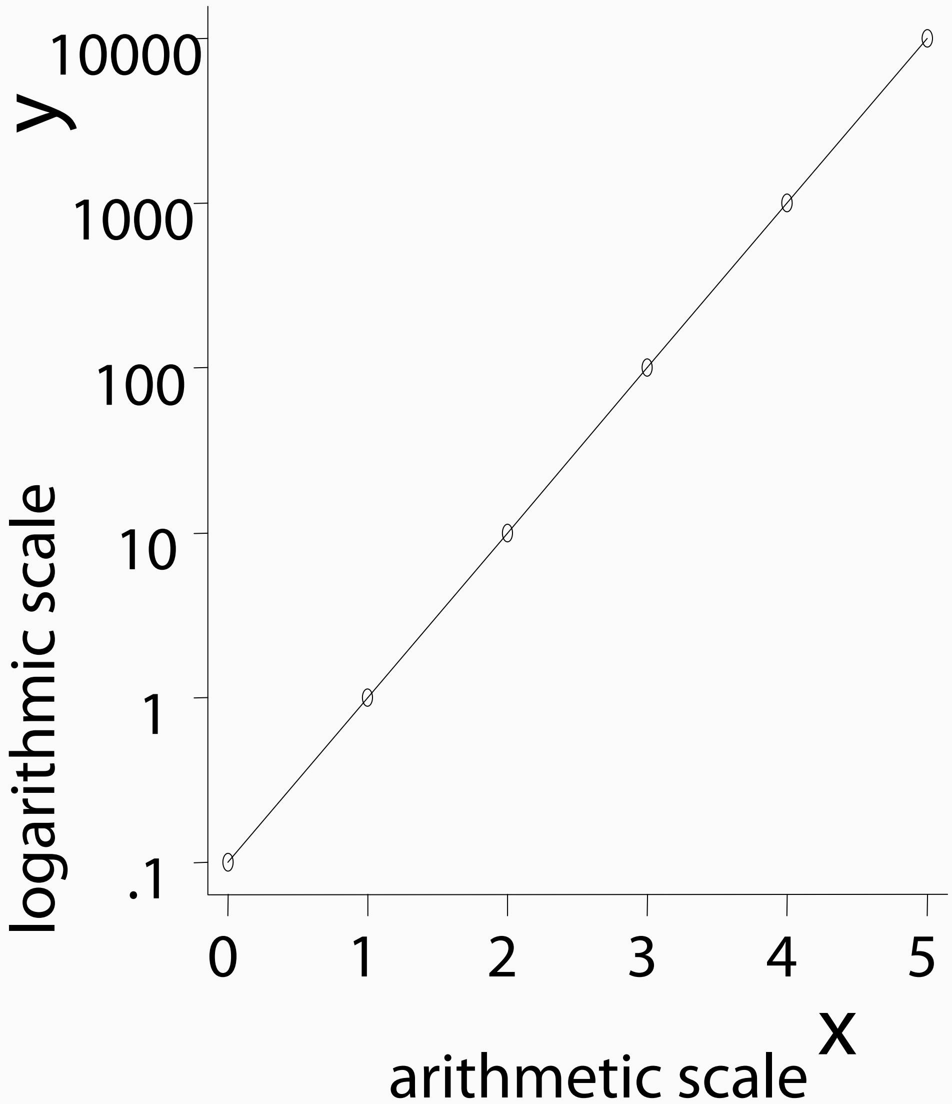
- Plotting the same data points on arithmetic paper reveals a non-linear trend

Nonlinear Trend on an Arithmetic Scale for Y



Quick Check: Plotting on Different Scales





Graphing on Semi-Logarithmic Paper

- Allows plotting of numbers of different magnitudes on the same graph
- Describes certain biological relationships
  - For example, exponential growth
- Aids in exploratory data analysis
- The logarithmic scale typically is based on the logarithm to base 10 (the common log)
  - $x = 10^y$ or $\log_{10}(x) = y$

Graphing on Semi-Logarithmic Paper

- $\log_{10}(1) = 0$ since $10^0 = 1$
- $\log_{10}(10) = 1$ since $10^1 = 10$
- $\log_{10}(100)=2$ since $10^2 = 100$

Graphing on Semi-Logarithmic Paper

- The x-axis is on an arithmetic scale
- The y-axis is on a logarithmic scale (logarithm to base 10)

Cautions in Interpreting Graphs

- Relationships displayed in a graph may be influenced by the method of data collection, changes in calendar time or definitions, and other factors
- Correlation or association that is displayed in a graph does not imply causation

JOHNS HOPKINS BLOOMBERG SCHOOL *of* PUBLIC HEALTH

Section B

Graphing Example

Number of Deaths in Baltimore City 1950–1980

Year	Malignant Neoplasms	Tuberculosis
1950	1,582	535
1960	1,856	163
1970	2,018	94
1980	2,054	12

Comparisons, Trends?

- Is it possible to make comparisons or find trends in death rates over time?
- No, one needs to adjust for the total population at each time period

Total Population and Deaths

Year	Total Population	Malignant Neoplasms	Tuberculosis
1950	949,708	1,582	535
1960	939,024	1,856	163
1970	901,582	2,018	94
1980	741,865	2,054	12

Calculating Death Rates

Year	Total Population	Malignant Neoplasms	Tuberculosis
1950	949,708	1,582	535
1960	939,024	1,856	163
1970	901,582	2,018	94
1980	741,865	2,054	12

- Example: 1950 death rate due to malignant neoplasms
  Deaths/population = 1,582 deaths /949,708 * 100,000
  Death rate = 170 deaths per 100,000 population

Death Rates Per 100,000 Population

- Death rates (log_{10} death rates)

Year	Malignant Neoplasms	Tuberculosis
1950	170 (2.2)	56 (1.8)
1960	200 (2.3)	17 (1.2)
1970	220 (2.3)	10 (1.0)
1980	280 (2.5)	2 (0.3)

Conclusions

- Conclusions based on data presented in an **arithmetic graph:**
  - The cancer death rate has increased over time
  - The tuberculosis death rate has decreased over time but is smaller in magnitude than the cancer death rate

- Conclusions based on data presented in a **semi-logarithmic graph:**
  - The proportional decrease in tuberculosis death rate is greater than the proportional increase in cancer death rate

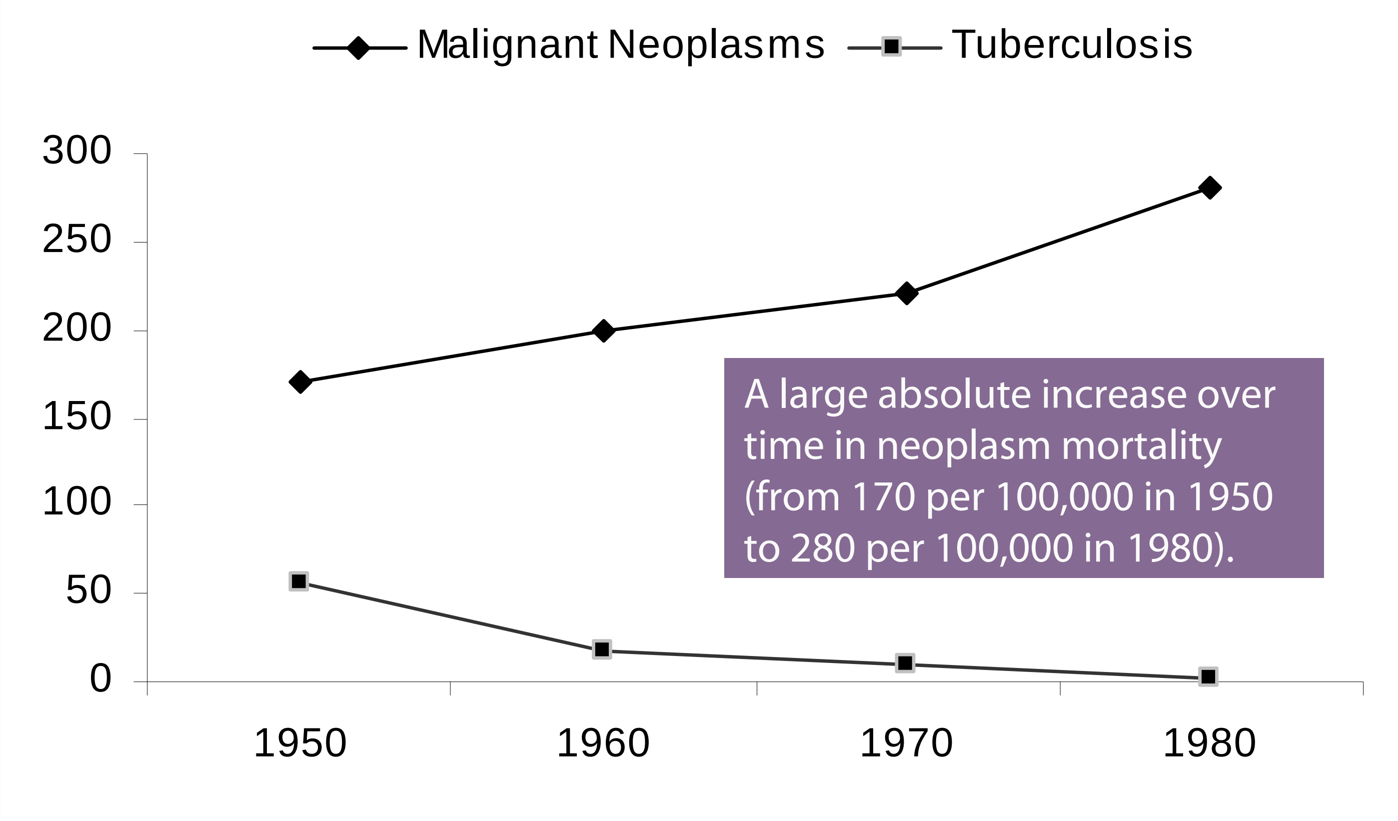
Plotting on Arithmetic and Logarithmic Scales

- Supposing you have a given data set that details the cause-specific death rates over a period of time in Baltimore City…

Year	Malignant Neoplasms	Tuberculosis
1950	170 (2.2)	56 (1.8)
1960	200 (2.3)	17 (1.2)
1970	220 (2.3)	10 (1.0)
1980	280 (2.5)	2 (0.3)

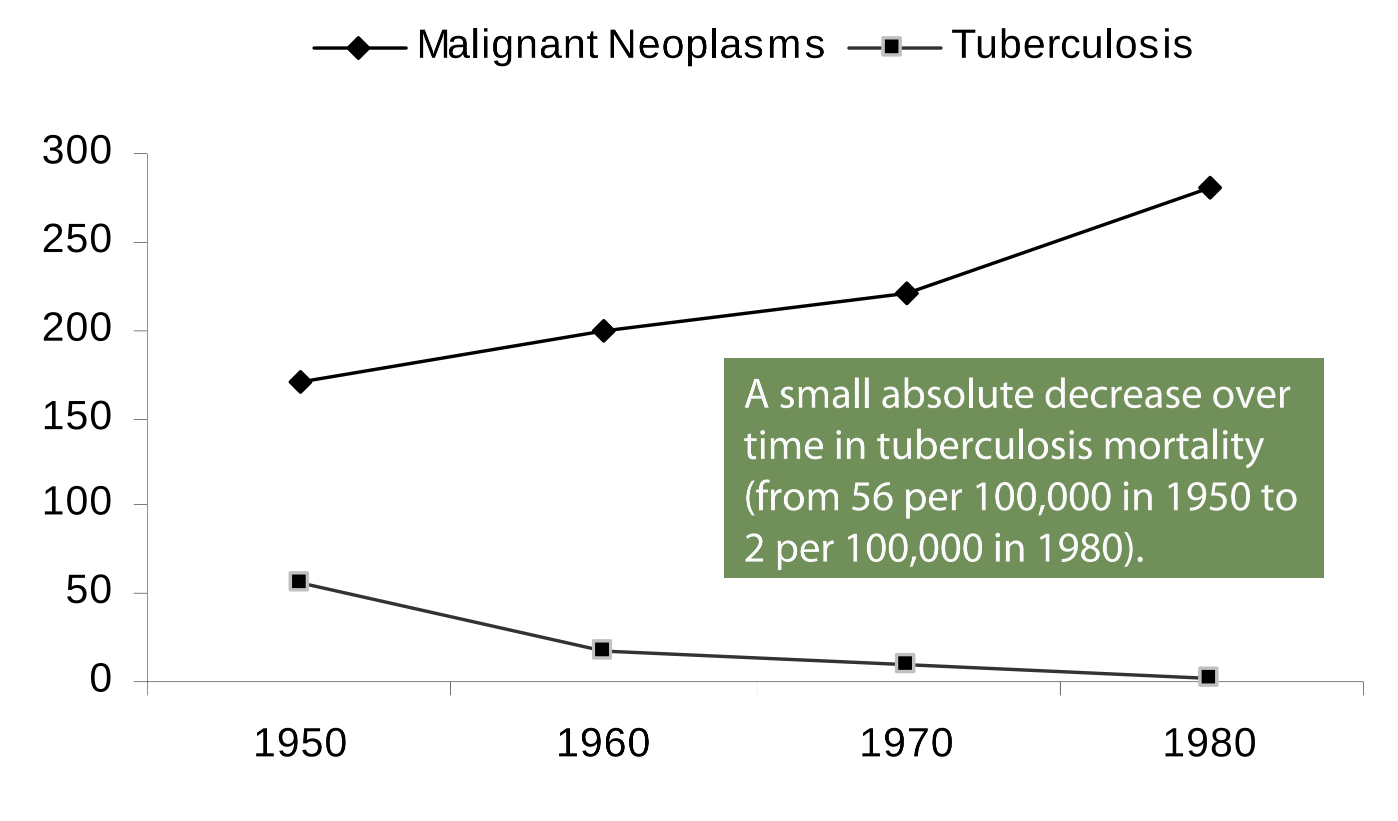
Plotting on Arithmetic and Logarithmic Scales

- A plot of the cause-specific death rates by year on an arithmetic scale shows:



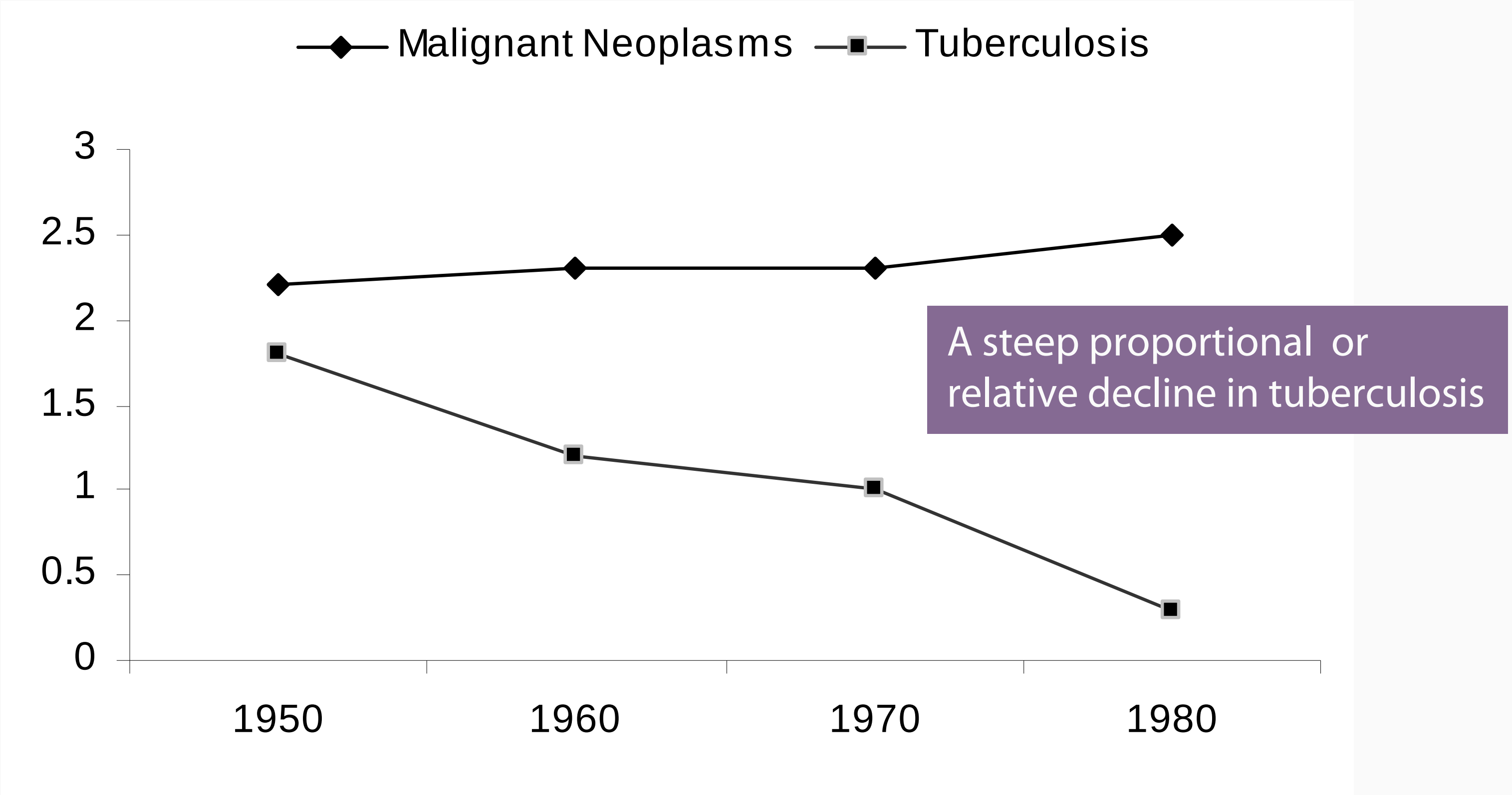
Plotting on Arithmetic and Logarithmic Scales

- A plot of the cause-specific death rates by year on an arithmetic scale shows:



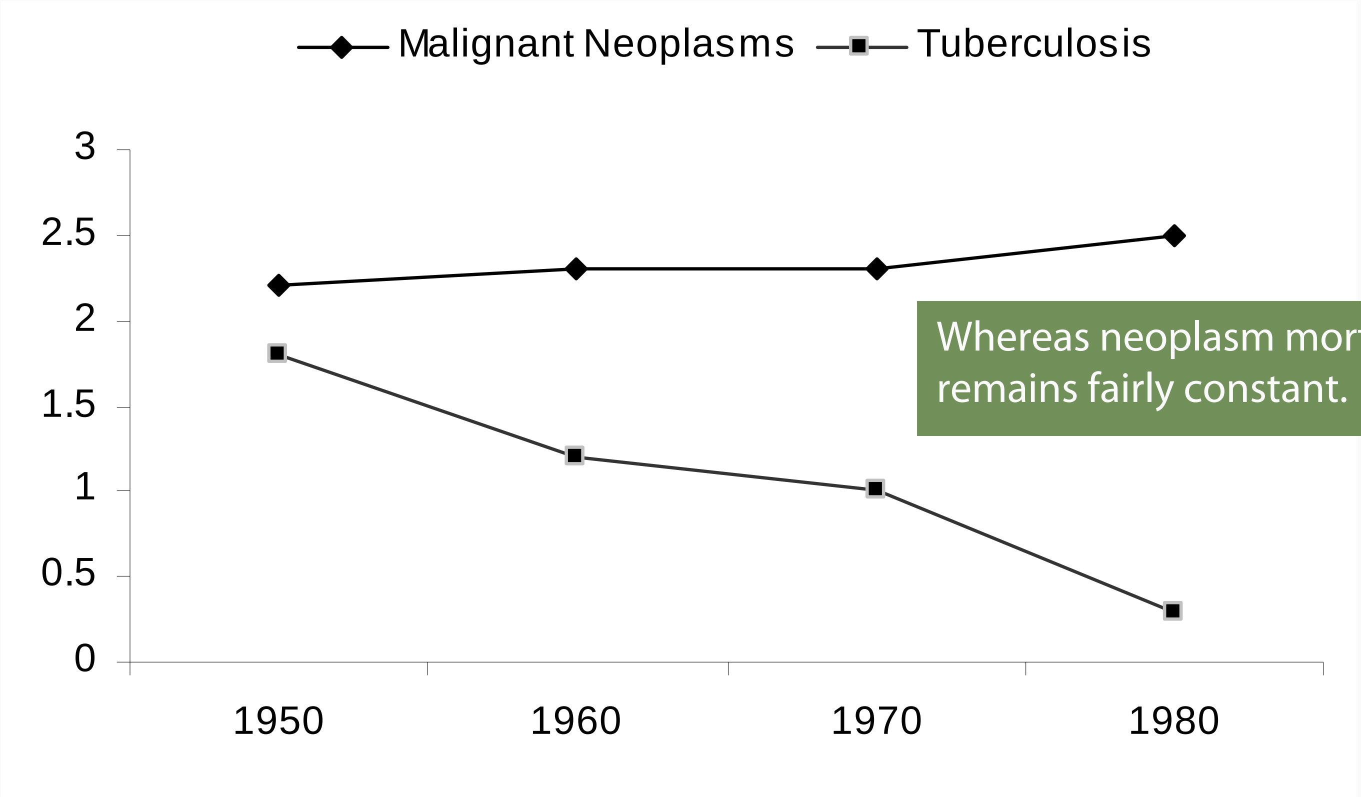
Plotting on Arithmetic and Logarithmic Scales

- A plot of the logs of the death rates shows:



Plotting on Arithmetic and Logarithmic Scales

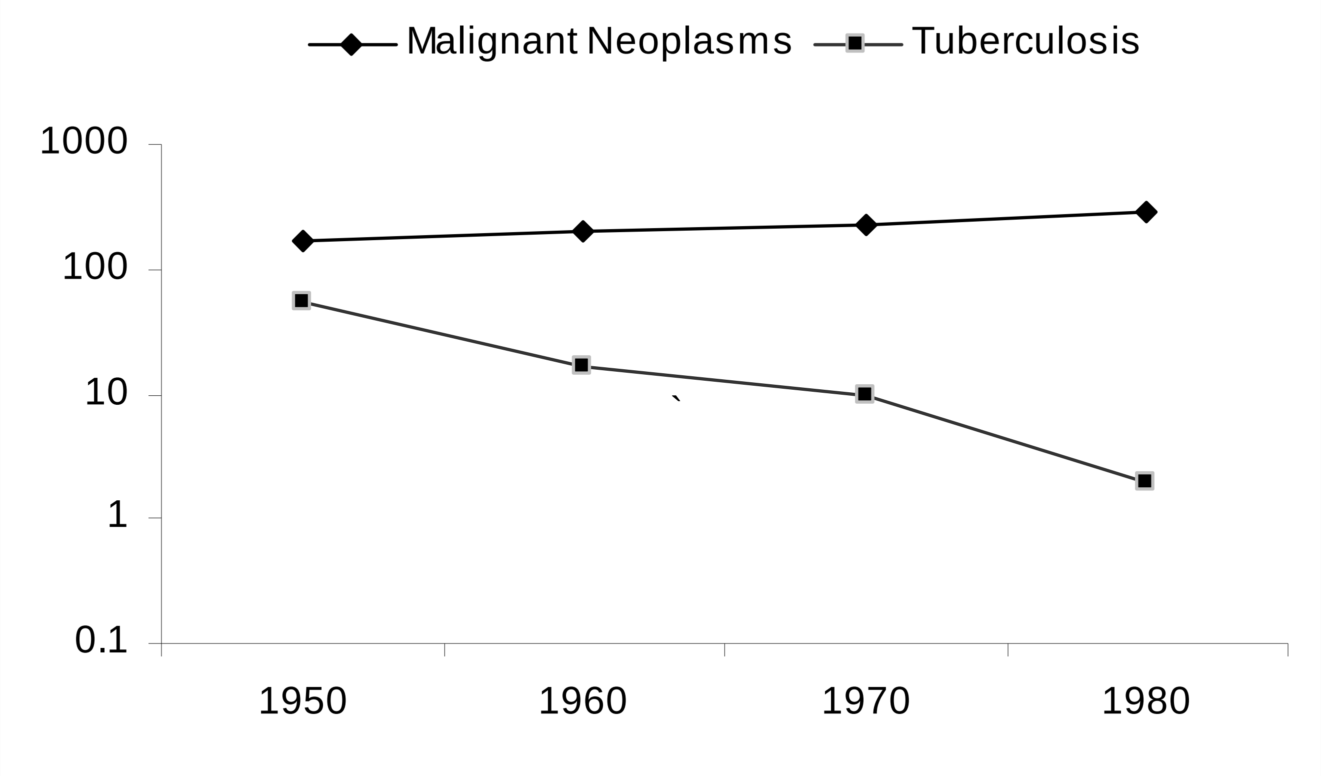
- A plot of the logs of the death rates shows:



Whereas neoplasm mortality remains fairly constant.

Plotting on Arithmetic and Logarithmic Scales

- It is not necessary to first calculate logs of the death rates in order to obtain this plot
- Plotting the actual death rates on a **logarithmic** scale will result in the same visual impression as the previous plot.



Plotting on Arithmetic and Logarithmic Scales

- Plotting the rates on a **logarithmic** scale gives you the same plot as plotting the logs of the rates on an **arithmetic** scale.