14. Time- Series data visualization

Prof. Tulasi Prasad Sariki
SCSE, VIT, Chennai
www.learnersdesk.weebly.com

Overview

- What is forecasting

- Time series & its components

- Smooth a data series

  - Moving average

  - Exponential smoothing

- Forecast using trend models

  - Simple Linear Regression

  - Auto-regressive

- Examples of Time Series Visualizations

What Is Forecasting?

- Time series forecasting is performed in nearly every organization that works with quantifiable data.

- Retail stores forecast sales.

- Energy companies forecast reserves, production, demand, and prices.

- Educational institutions forecast enrollment.

- Governments forecast tax receipts and spending.

- International financial organizations forecast inflation and economic activity.

- Forecasting is a fundamental analytic process in every organization.

Process of predicting a future event

# Forecasting Approaches

Qualitative Methods

- Used when situation is vague & little data exist
  - New products
  - New technology

- Involve intuition, experience

- e.g., forecasting sales on Internet

Quantitative Methods

- Used when situation is 'stable' & historical data exist
  - Existing products
  - Current technology

- Involve mathematical techniques

- e.g., forecasting sales of color televisions

Quantitative Forecasting Methods
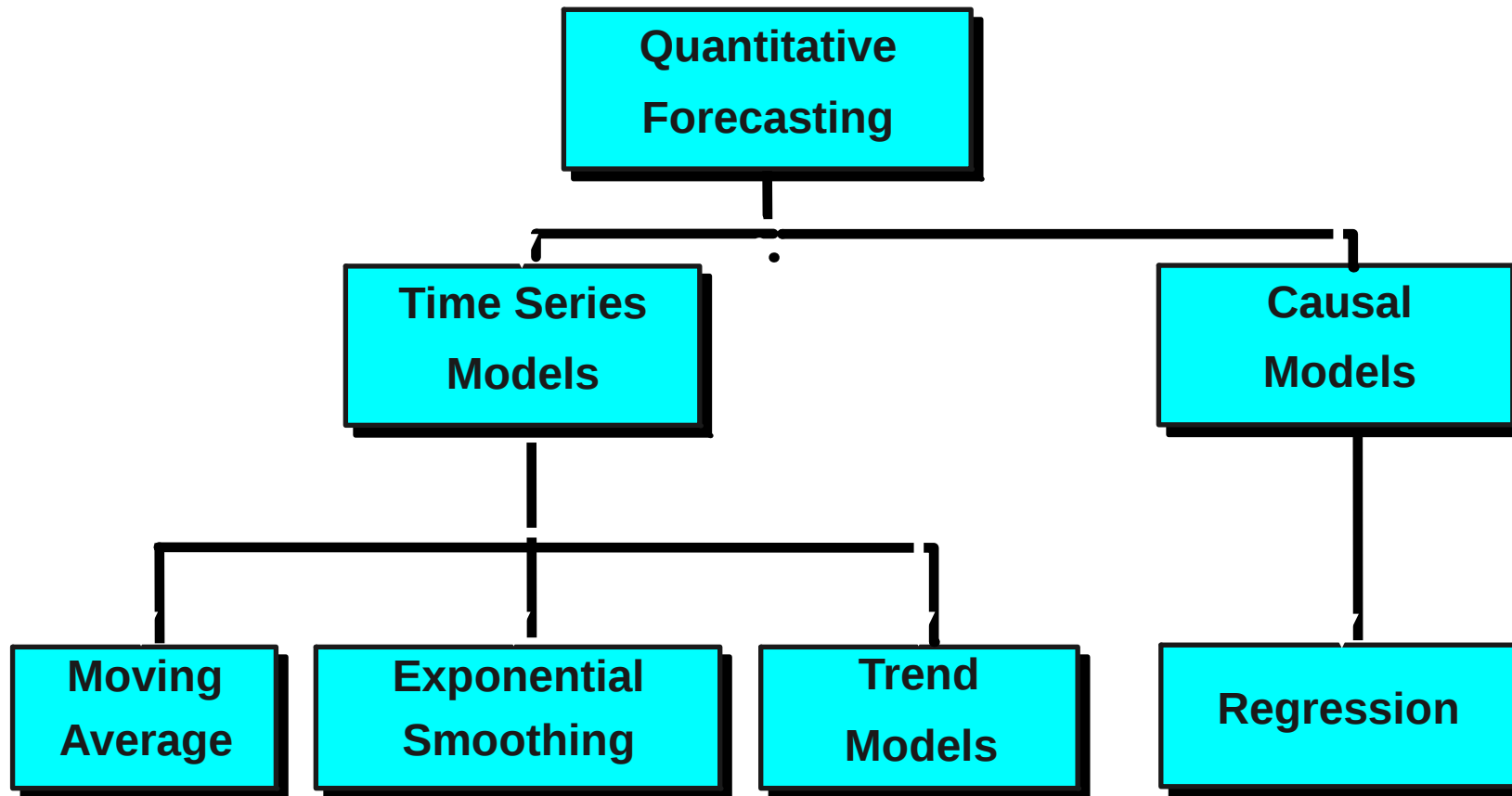
What is a Time Series?

Set of evenly spaced numerical data

  Obtained by observing response variable at regular time periods
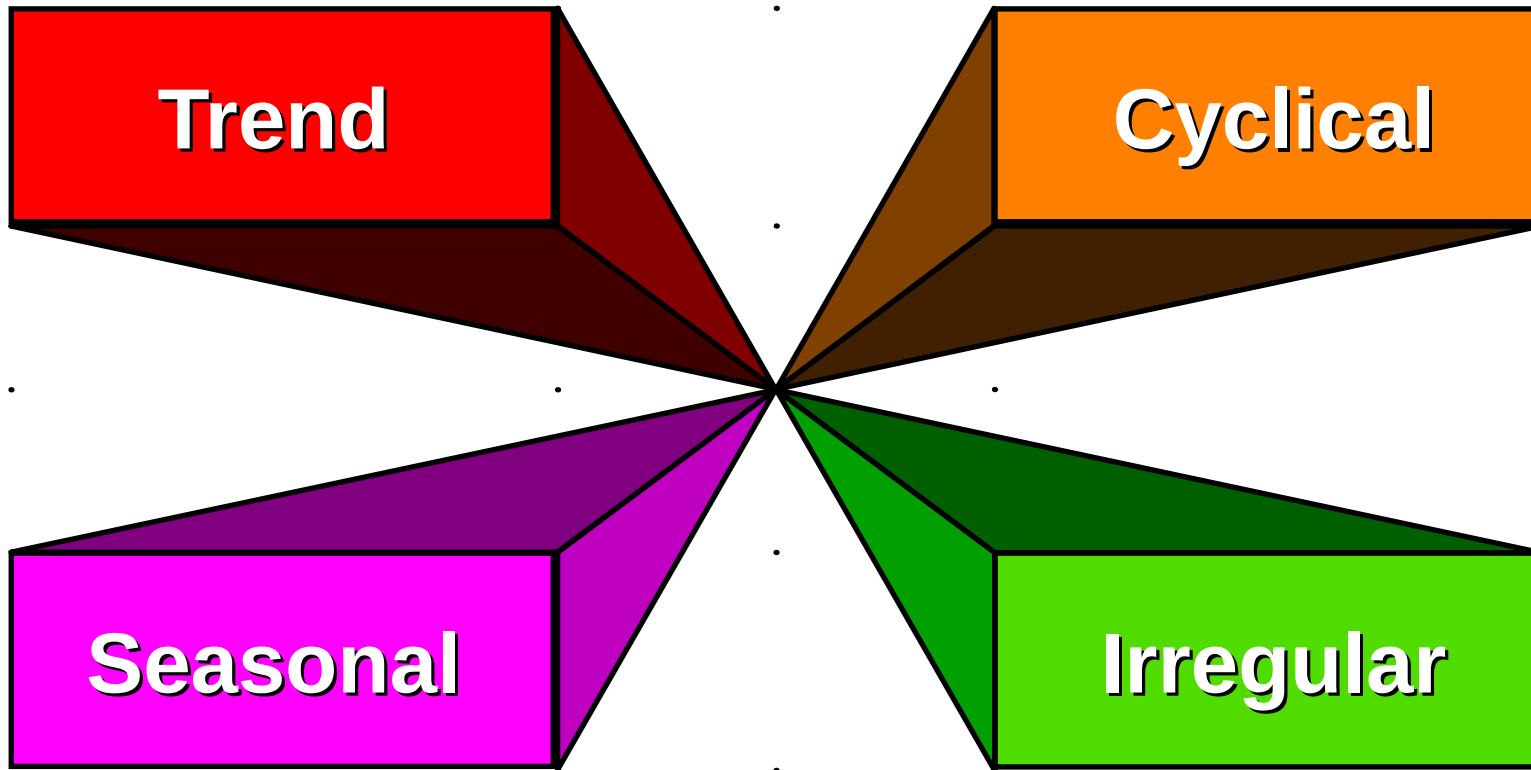
Forecast based only on past values

  Assumes that factors influencing past, present, & future will continue

Example

  Year: 1995   1996   1997   1998   1999

  Sales: 78.7   63.5   89.7   93.2   92.1
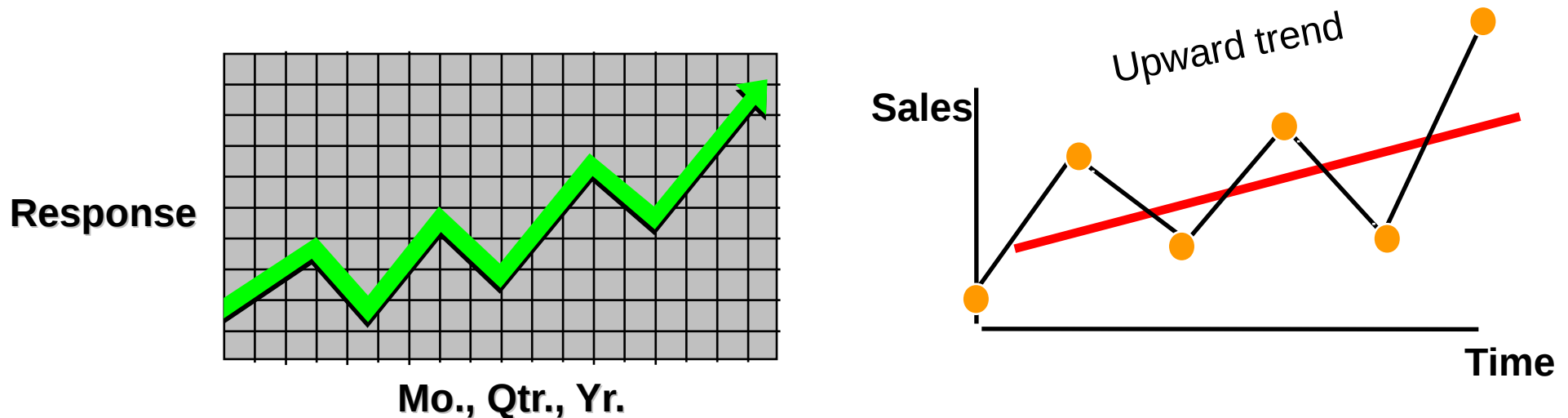
Time Series Components

Trend Component

Persistent, overall upward or downward pattern

Due to population, technology etc.

Overall Upward or Downward Movement
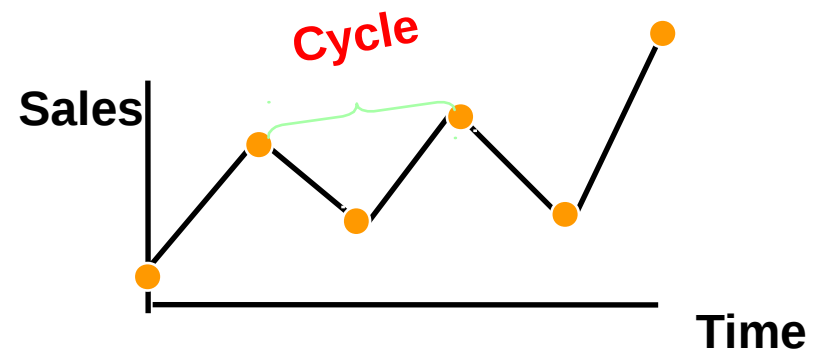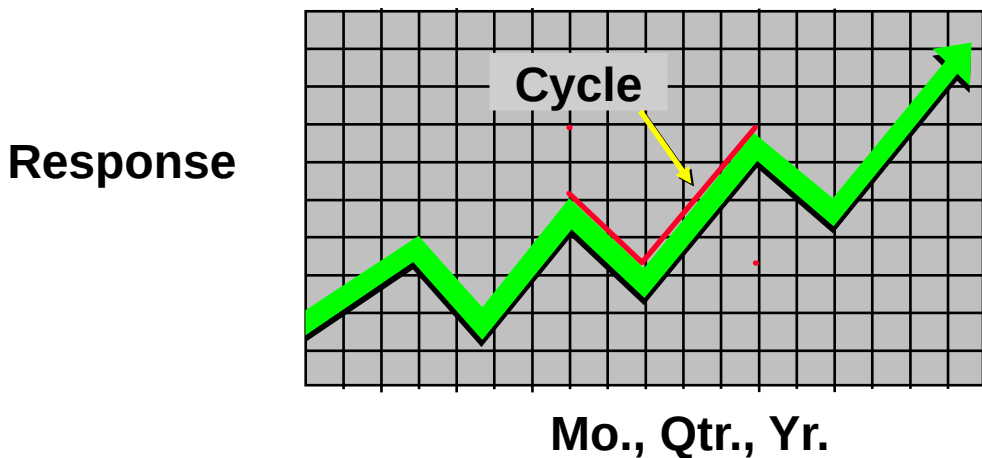
Data Taken Over a Period of Years

Cyclical Component

Repeating up & down movements.
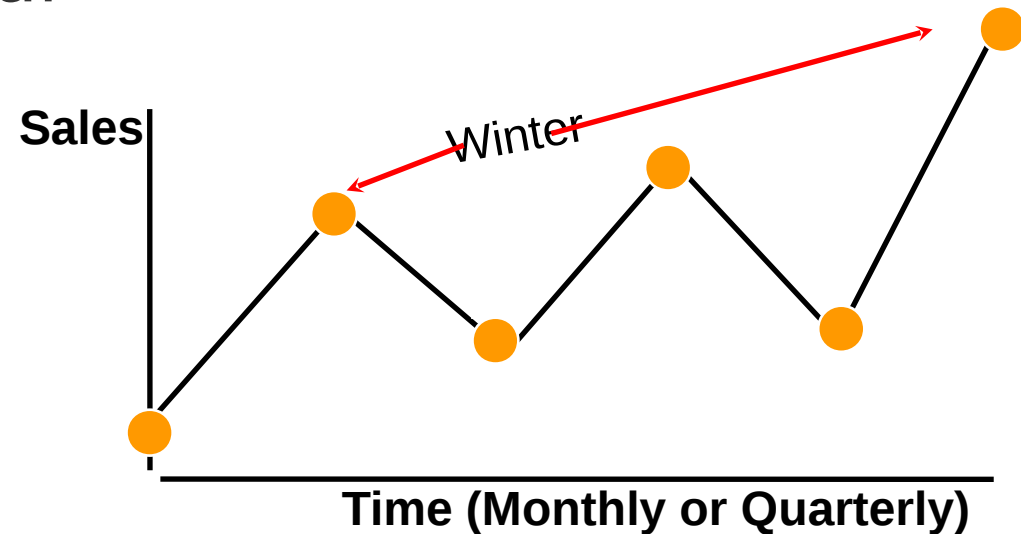
Due to interactions of factors influencing economy

Usually 2-10 years duration.

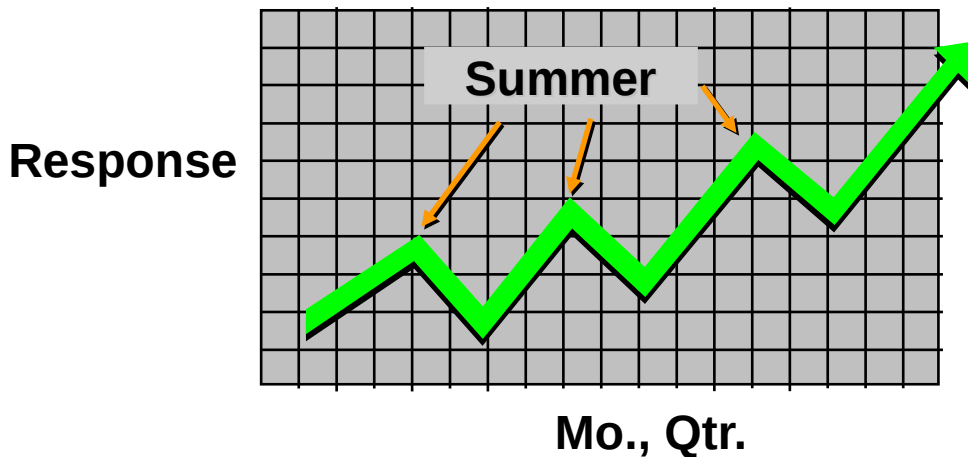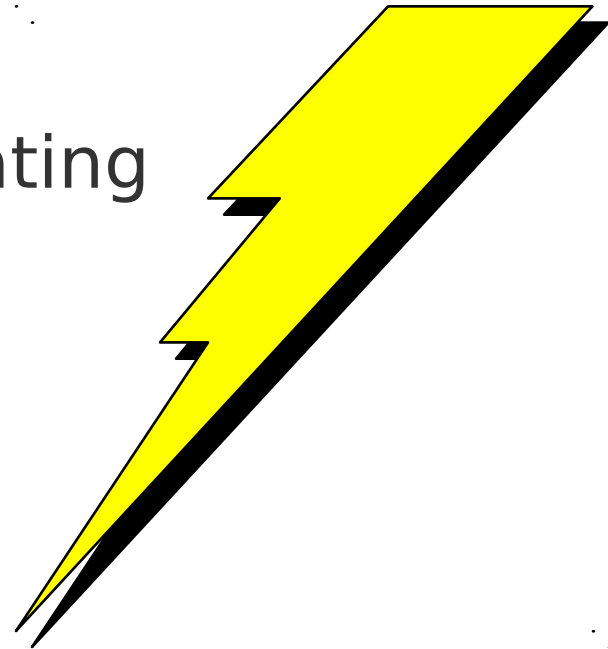May Vary in Length; Usually Lasts 2 - 10 Year

Seasonal Component

- Regular pattern of up & down fluctuations

- Due to weather, customs etc.

- Regular Patterns

- Observed Within One Year

Irregular Component

- Erratic, unsystematic, 'residual' fluctuations
- Due to random variation or unforeseen events
  - Union strike
  - War
- Short duration & non-repeating

Time Series Forecasting
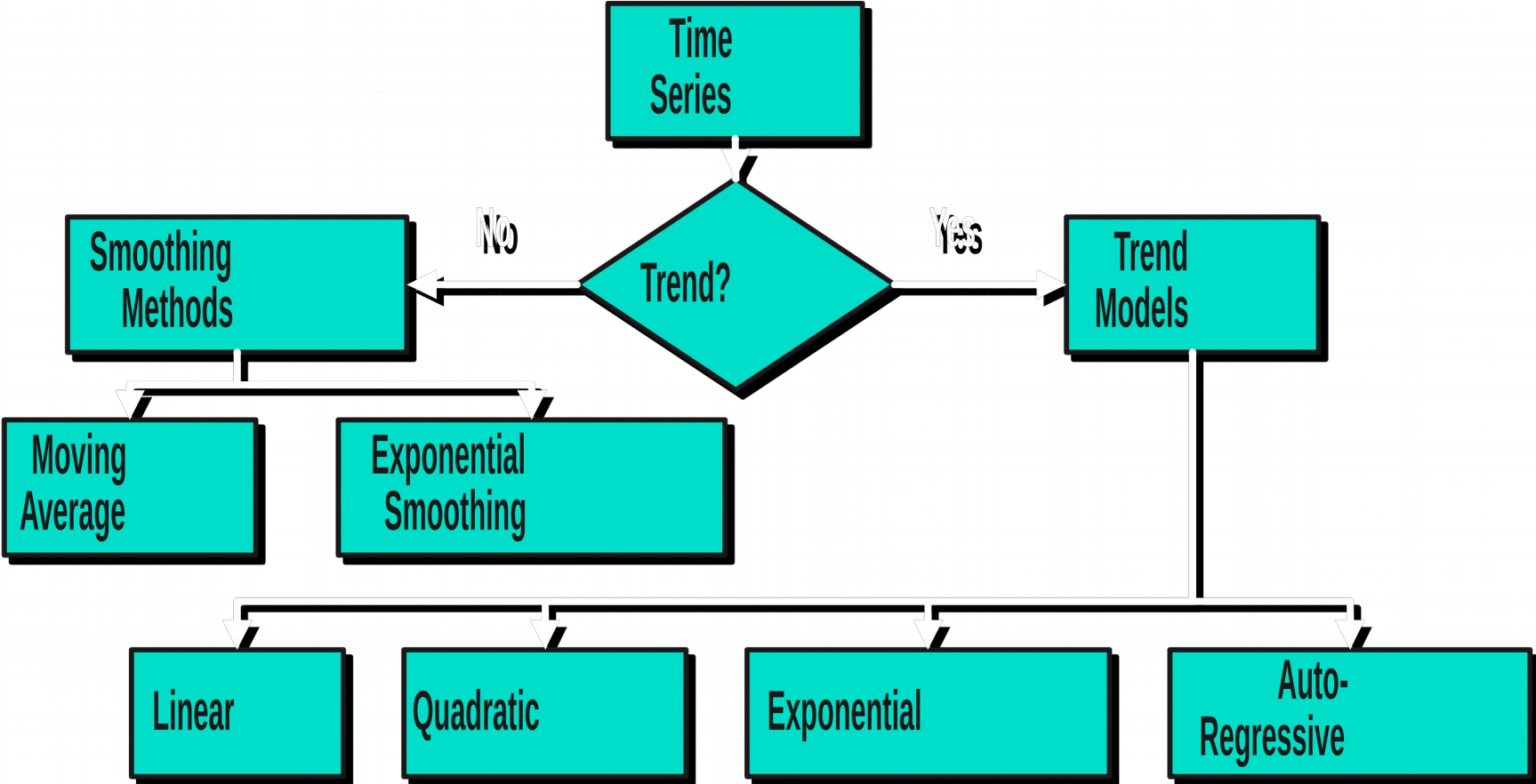
Moving Average Method

You work for Firestone Tire. You want to smooth random fluctuations using a 3-period moving average.

Year	Sales	MA(3) in 1,000
1995	20,000	NA
1996	24,000	(20+24+22)/3 = 22
1997	22,000	(24+22+26)/3 = 24
1998	26,000	(22+26+25)/3 = 24
1999	25,000	NA

Exponential Smoothing

You want to forecast attendance for 1998 using exponential smoothing(α = .20).

$$E_i = W \cdot Y_i + (1 - W) \cdot E_{i-1}$$

Time	Y_i	Smoothed Value, E_i (W = .2)	Forecast Y_{i+1}
1995	4	4.0	NA
1996	6	(.2)(6) + (1-.2)(4.0) = 4.4	4.0
1997	5	(.2)(5) + (1-.2)(4.4) = 4.5	4.4
1998	3	(.2)(3) + (1-.2)(4.5) = 4.2	4.5
1999	7	(.2)(7) + (1-.2)(4.2) = 4.8	4.2
2000	NA	NA	4.8

Time Series Plot - Line Graph

- A line graph is the simplest way to represent time series data.

- It is intuitive, easy to create, and helps the viewer get a quick sense of how something has changed over time.

- A line graph uses points connected by lines (also called trend lines) to show how a dependent variable and independent variable changed.

- For temporal visualizations, time is always the independent variable, which is plotted on the horizontal axis. Then the dependent variable is plotted on the vertical axis.

Time Series Plot - Line Graph

Legend: Ireland (green), Europe (blue)

Y-axis (left): Population of Ireland (millions) — 0 to 9

Y-axis (right): Population of Europe (millions) — 0 to 900

X-axis: 1740, 1780, 1820, 1860, 1900, 1940, 1980
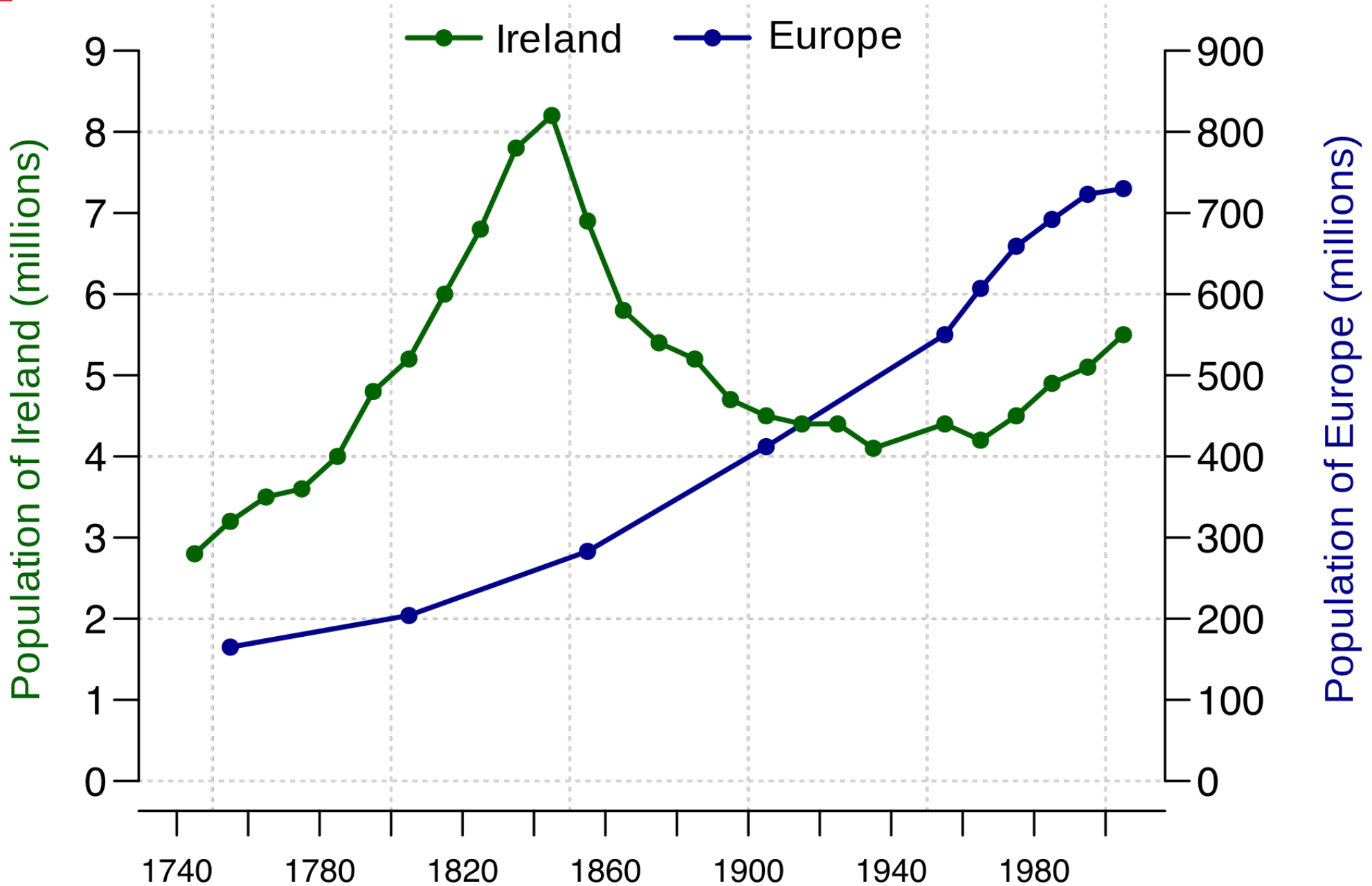
Time Series Plot - Line Graph

- This graph captures the population growth in Europe and Ireland from 1740 to around 2010.

- It clearly highlights the sudden drop in Ireland's population in the 1840s(resultt of the devastating Irish Potato Famine, a period of mass starvation, disease, and emigration in Ireland between 1845 and 1852).

- Note that this graph uses different y-axis scales for the populations of Europe and Ireland. If the viewer doesn't pay attention to the difference in the scales, they could be led to the conclusion that until about 1920, Ireland's population was greater than that of Europe!

Time Series Plot - Stacked Area Chart

- An area chart is similar to a line chart in that it has points connected by straight lines on a two-dimensional chart.

- It also puts time as the independent variable on the x-axis and the dependent variable on the y-axis.

- However, in an area chart, multiple variables are "stacked" on top of each other, and the area below each line is colored to represent each variable.

Time Series Plot - Stacked Area Chart

- This is a stacked area chart showing time series data of student enrollments in India from 2001-10.

Student enrollments in India (2001-10)

Time Series Plot - Stacked Area Chart

- Stacked area charts are useful to show how both a cumulative total and individual components of that total changed over time.

- The order in which we stack the variables is crucial because there can sometimes be a difference in the actual plot versus human perception.
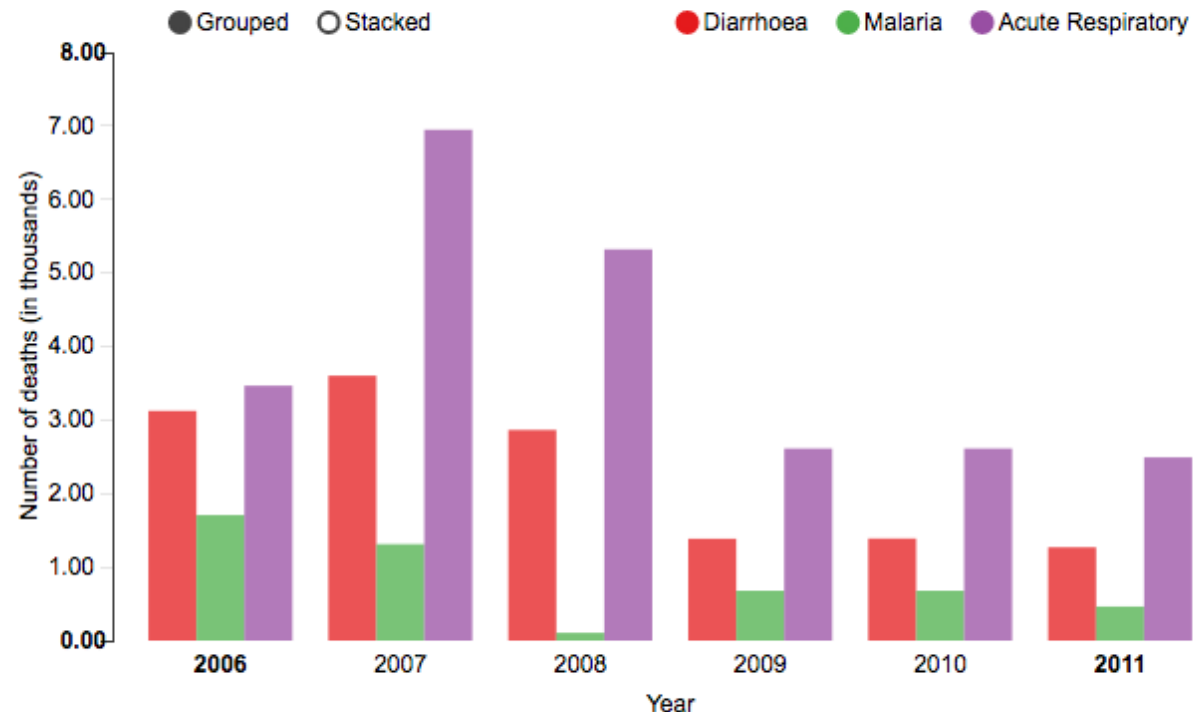
Time Series Plot - Bar Chart

- The length of each bar is proportional to the value of the variable at that point in time.

- A bar chart is the right choice for you when you wish to look at how the variable moved over time or when you wish to compare variables versus each other.

- Grouped or stacked bar charts help you combine both these purposes in one chart while keeping your visualization simple and intuitive.

- To avoid clutter and confusion, make sure to not use more than 3 variables in a stacked or group bar chart.

- It is also a good practice to use consistent bold colors and leave appropriate space between two bars in a bar chart.

Time Series Plot - Bar Chart

Number of deaths by type of diseases in India (2006-2011)

- To avoid clutter and confusion, make sure to not use more than 3 variables in a stacked or group bar chart.

Time Series Plot - Bar Chart
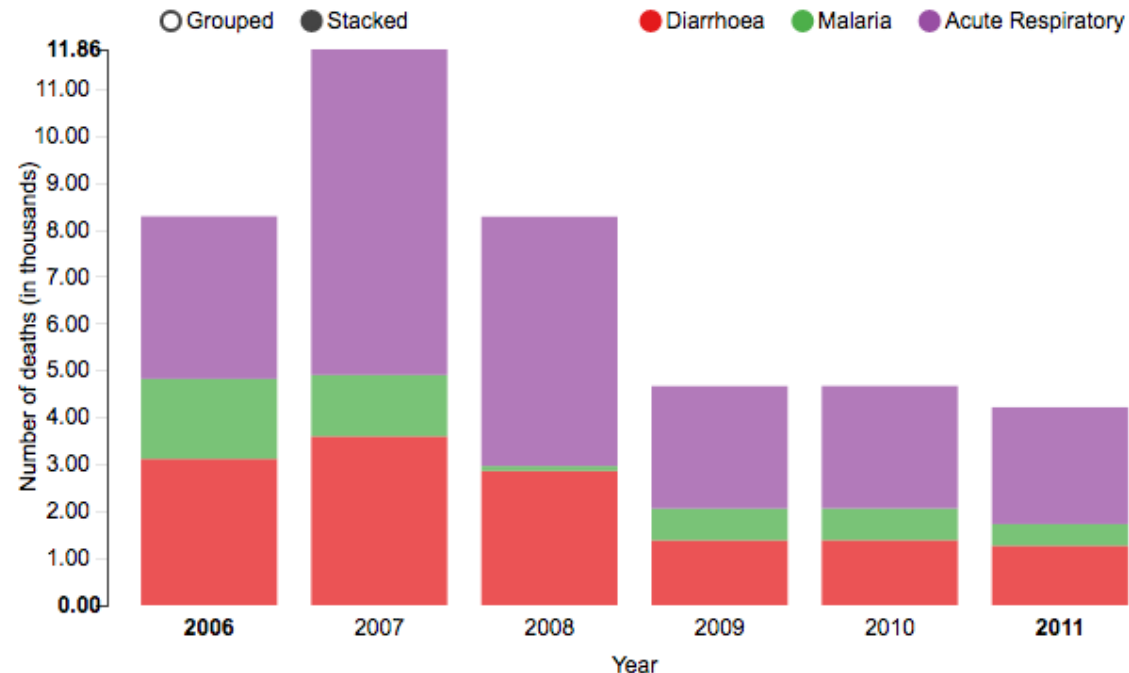
Number of deaths by type of diseases in India (2006-2011)

○ Grouped   ● Stacked   ● Diarrhoea   ● Malaria   ● Acute Respiratory

Number of deaths (in thousands)

Year

Powered By
S

(Data Source)
data.gov in

- It is also a good practice to use consistent bold colors and leave appropriate space between two bars in a bar chart.
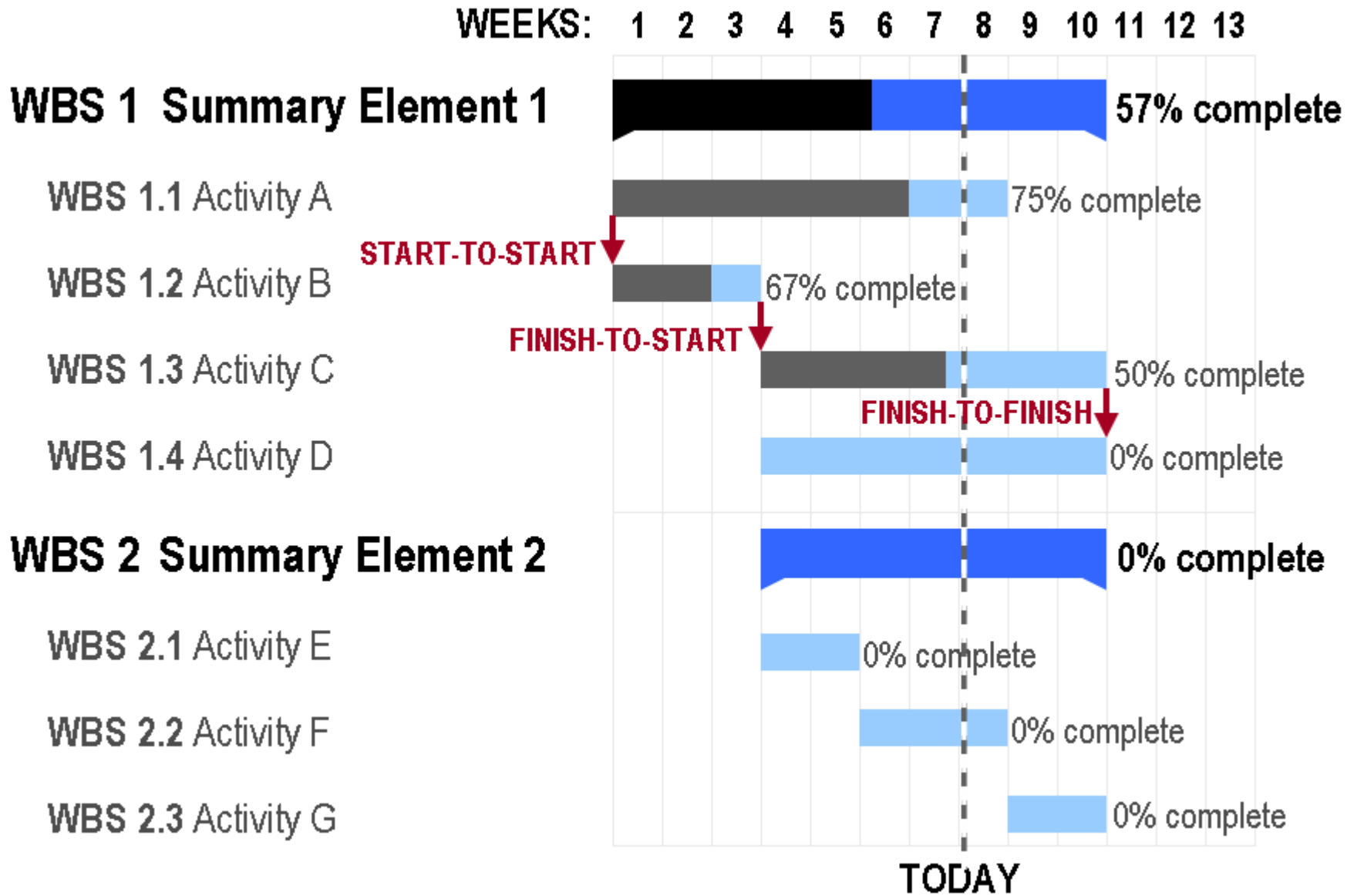
Time Series Plot - Gantt chart

- A Gantt chart is a horizontal bar chart showing work completed in a certain period of time with respect to the time allocated for that particular task.

- Gantt charts are a popular project management tool since they present a concise snapshot of various tasks spread across various phases of the project.

- You can show additional information such as the correlation between individual tasks, resources used in each task, overlapping resources, etc., by the use of colors and placement of bars in a Gantt chart.

Time Series Plot - Gantt chart

WEEKS: 1 2 3 4 5 6 7 8 9 10 11 12 13

WBS 1 Summary Element 1 — 57% complete

WBS 1.1 Activity A — 75% complete

START-TO-START

WBS 1.2 Activity B — 67% complete

FINISH-TO-START

WBS 1.3 Activity C — 50% complete

FINISH-TO-FINISH

WBS 1.4 Activity D — 0% complete

WBS 2 Summary Element 2 — 0% complete

WBS 2.1 Activity E — 0% complete

WBS 2.2 Activity F — 0% complete

WBS 2.3 Activity G — 0% complete

TODAY

Time Series Plot - Gantt chart

Dance Concert Planning

| Task	
Hire choreographer	
Soundtrack selection	
Select dancers	
Choreography	
Prop selection	
Costume selection	
Stage décor	
Final rehearsals	

■ Task duration (in days)

Time Series Plot - Stream Graph

- A stream graph is essentially a stacked area graph, but displaced around a central horizontal axis. The stream graph looks like flowing liquid, hence the name.

- Stream graphs are great to represent and compare time series data for multiple variables.

- Stream graphs are, thus, apt for large data sets. Remember that choice of colors is very important, especially when there are lots of variables.

- Variables that do not have significantly high values might tend to get drowned out in the visualization if the colors are not chosen well

Time Series Plot - Stream Graph

- Below is a stream graph showing a randomly chosen listener's last.fm music-listening habits over time.
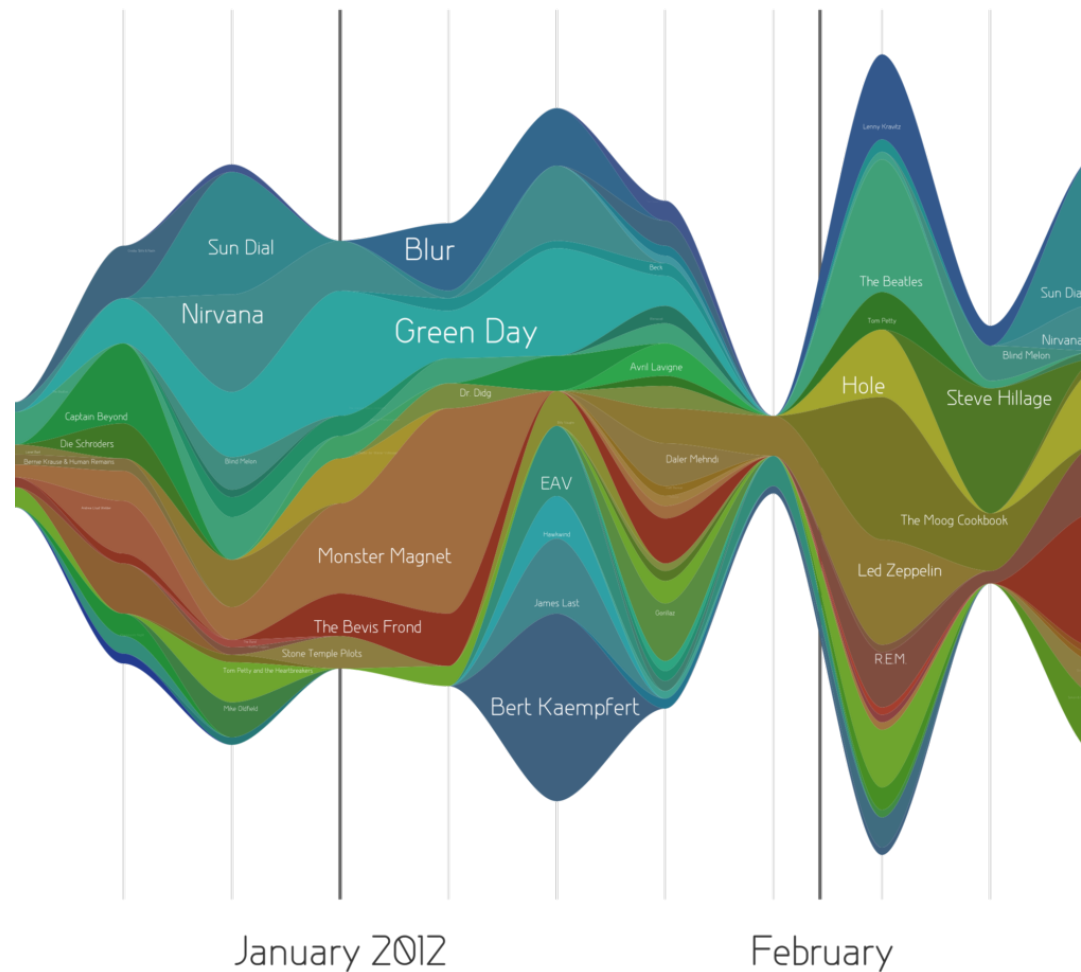
Time Series Plot - Heat Map

- Geospatial visualizations often use heat maps since they quickly help identify "hot spots" or regions of high concentrations of a given variable.

- When adapted to temporal visualizations, heat maps can help us explore two levels of time in a 2D array.

- This heat map visualizes birthdays for babies born in the United States between 1973 and 1999.

- The vertical axis represents the 31 days in a month while the horizontal axis represents the 12 months in a year.

- This chart quickly helps us identify that a large number of babies were born in the later half of July, August, and September.

DAY JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC

BIRTHDAY RANK

LESS COMMON   MORE COMMON

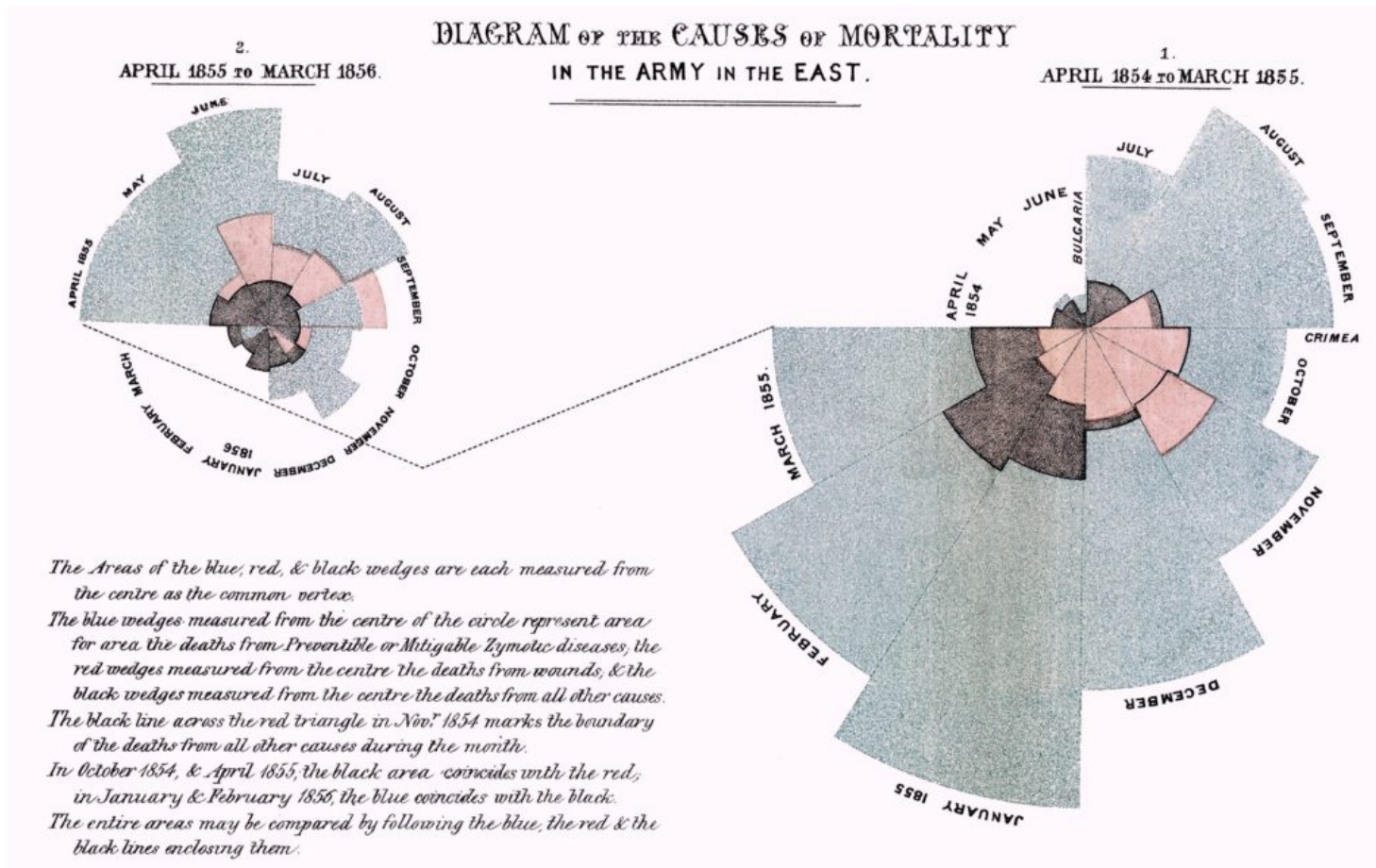# Time Series Plot -
Polar Area Diagram

- Think beyond the straight line! Sometimes, time series data can be cyclical — a season in a year, time of the day, and so on.

- Polar area diagrams help represent the cyclical nature time series data cleanly.

- A polar diagram looks like a traditional pie chart, but the sectors differ from each other not by the size of their angles but by how far they extend out from the centre of the circle.

- This popular polar area diagram created by Florence Nightingale shows causes of mortality among British troops in the Crimean War. Each color in the diagram represents a different cause of death.

Time Series Plot - Polar Area Diagram

Thank You