LO5

Data Visualization

Objective

After attending this session, you should know

- Three types of data visualization
- Selecting optimal graphics
- Communicating with color and context
- Creating standard data graphics
- Defining elements of plot
- Plot formatting
- Creating label and annotations
- Visualizing time series
- Creating statistical data graphics

Data visualization types

- Data Storytelling - for presentation to organisational decision maker

- Data showcasing – for presentation to analysts, scientist, mathematicians, and engineers

- Data art – for presentations to activities or to general public

Diving into Data Storytelling

- Make it easy for the audience to get the point

  (understand that point within the first 10 to 15 seconds of looking at

  your graphic, otherwise they lose their attention.)

- For these reasons, your data visualization needs to be

  - clutter-free

  - highly organized

- Intended audience
  - Non analysts
  - Nontechnical business managers

- Product types:
  - Static images
  - Simple interactive dashboards

Demonstrating data storytelling

Bloomberg the Company | Bloomberg Anywhere Login

Bloomberg Business

What's Really Warming the World?

By Eric Roston and Blacki Migliozzi | June 24, 2015

Skeptics of manmade climate change offer various natural causes to explain why the Earth has warmed 1.4 degrees Fahrenheit since 1880. But can these account for the planet's rising temperature? Scroll down to see show how much different factors, both natural and industrial, contribute to global warming, based on findings from NASA's Goddard Institute for Space Studies.

Hotter
+2°F

Observed

1880-1910
Average

This line shows the measured, or "observed," land-ocean temperature

1880-2014

-2°F
Colder

Diving into Data Showcasing

- Showcase lots of data so your audience members can think for themselves

- Opposite to Data storytelling, the audience is actually looking to draw their own conclusions in data showcasing

- Your data visualization should be:

  - High contextual

  - Open ended

- Intended audience:

- Analysts, quants

- Engineers, mathematicians, scientist

(Basically, people with an analytical mindset, background, and bent, because

this is the kind of thing that's going to be of value to them)

- Product types:

  - Static images

  - Interactive dashboards

Demonstrating data showcasing
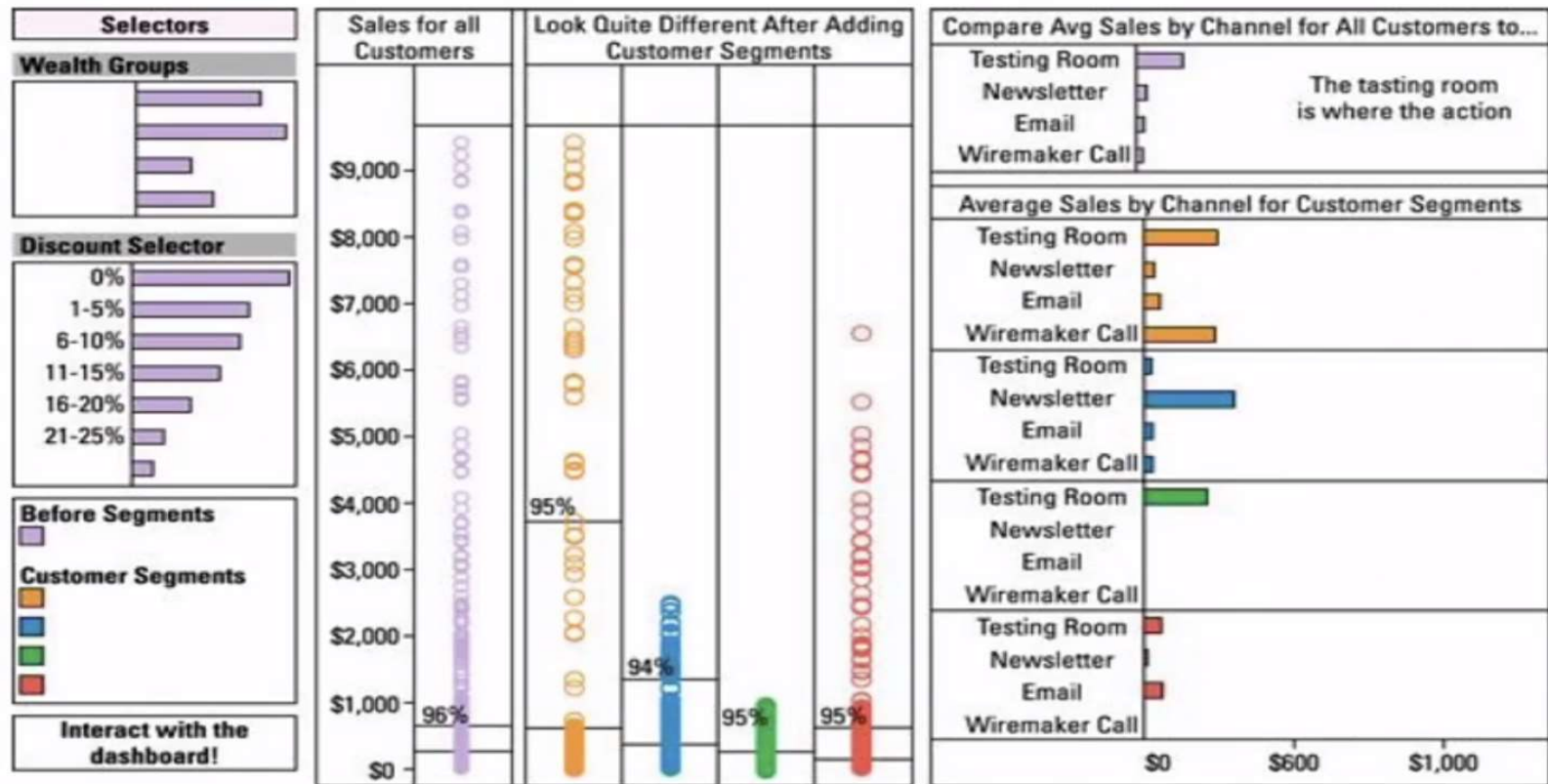
Displaying Data Art

- Data art to make a statement

- Most data art is used in something like data journalism, or if you're trying to get your community, or a community of activists, to be engaged and taking action

- Your data visualization should be:

  - Attention getting

  - Creative, controversial (Basically, the more controversial is the more action you're going to get)

- Intended audience

  - Idealists, dreamers, artists

  - Social activists

- Product types:

  - Static images (Sometimes it is also produced as an interactive piece on the web)

U.S. GUN KILLINGS IN 2010 2013

9,595
PEOPLE KILLED

415,128
STOLEN YEARS

AGE 0 ... 100

An example from Periscopic: http://www.periscopic.com

- on the left, you've got an account for the number of people killed (9595 people, highlighted in orange)

- on the right side, counted the number of years that were lost from those lives, from the people that were killed

Choosing graphics for data storytelling

Choropleth map

Choropleth map shows area boundaries

Point map

Choosing Graphics for data showcasing

Area Charts

Bar Charts

Line Charts

Cloropleths

Point Maps

Histograms

Scatter Plots

Scatter Plot Matrices

Raster Maps

Raster map

- a picture of a Doppler weather radar on top of a geographic map

- it's a raster file that's actually made up of an X and Y grid, which is filled in with a variety of values. The XY grid is then filled in with values that are colored according to count
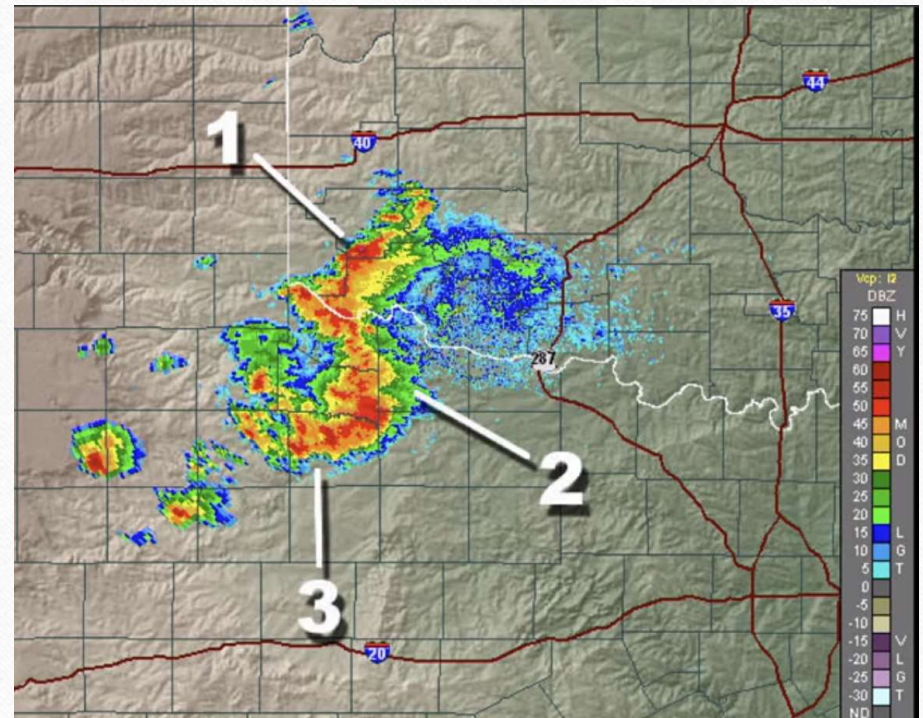
Choosing graphics for data art

Line Charts

Graph Networks

Cloropleths

Something Weird and Artistic…

Four steps to choosing data graphics

1. Make a list of the questions that your data visualization is meant to answer.

2. Is your data visualization type data storytelling, data showcasing or data art?

3. What data graphics types are preferable for that type of data visualization?

4. Test out different types of data graphics with your data. Which graphic type displays the most clear and obvious answers to your questions?

Testing out the graphics

- The two data graphics below represent the same statistic. Notice how the data graphic on the right does a much better job of visually emphasizing the difference in values?

- You should always test different data graphics to see its effectiveness to present your data

Creating context with color

- Color should be used:
  - Strategically
  - Sparingly
  - Consistently
- You want to use color to draw attention to the parts of the visualization that matter, and away from the parts that don't

Choose consistent colors

- The first rule is the color shades you choose are all from the same color formula.

- The options include monochromatic, analogous, complimentary, split complimentary, triadic, tetradic-based on color theory

EFEAEB 997BB0 4B4F69 C24B5A 31374B

AI-driven color match tool - colormind, http://colormind.io

Annotation as context

- Provide your audience some information about why data is as it is

- Useful in data storytelling and data showcasing
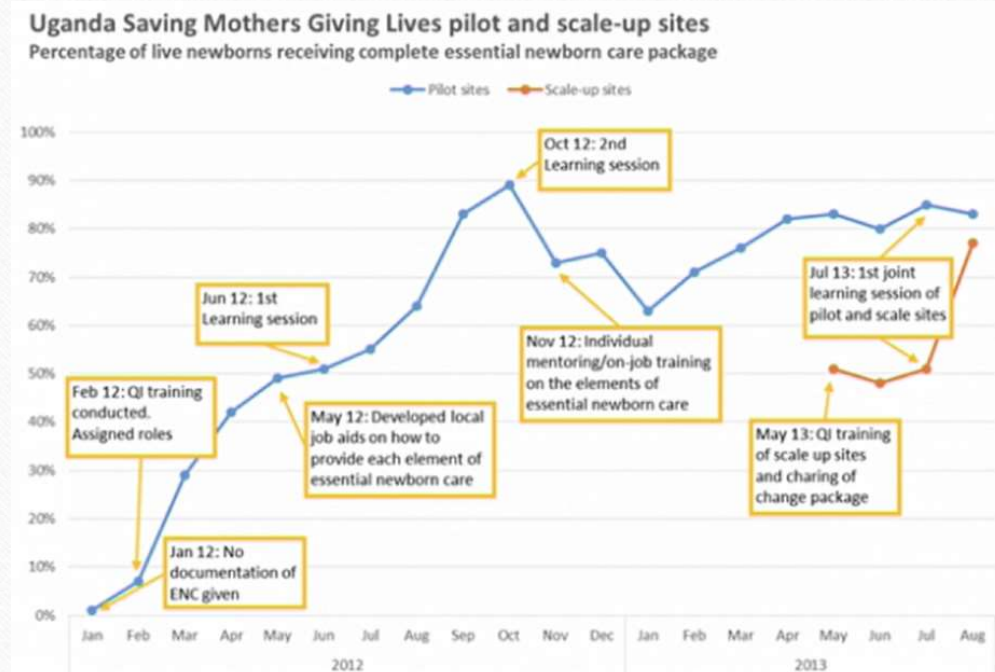
Graphics as context

- Help your audience understand the relative significance of the data, you are showing them

- Trendlines, single-value alerts, and benchmarks
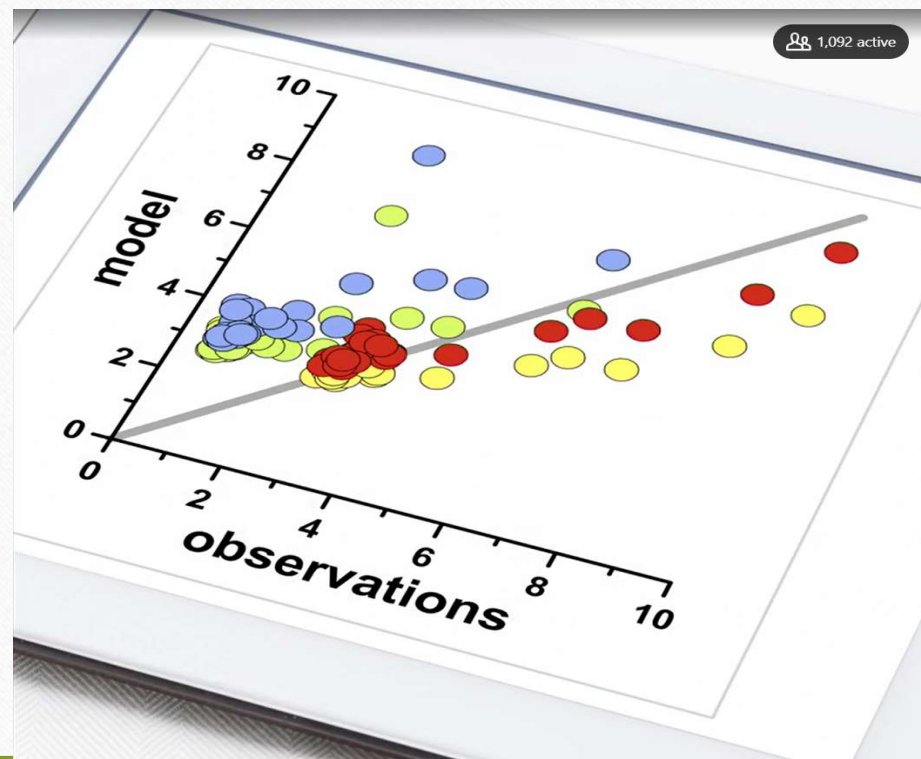
Creating context

- How: add data on additional metrics that are relevant to the datasets you are showing, trendlines, colors, and annotations

- Why: meant to give audience some deeper perspective and insight into what's happening

- When: useful in data showcasing

Context is especially useful in data showcasing, because, again, data showcasing pieces are meant for analytical audiences that really want to put in a lot of thought into drawing conclusions from the data.
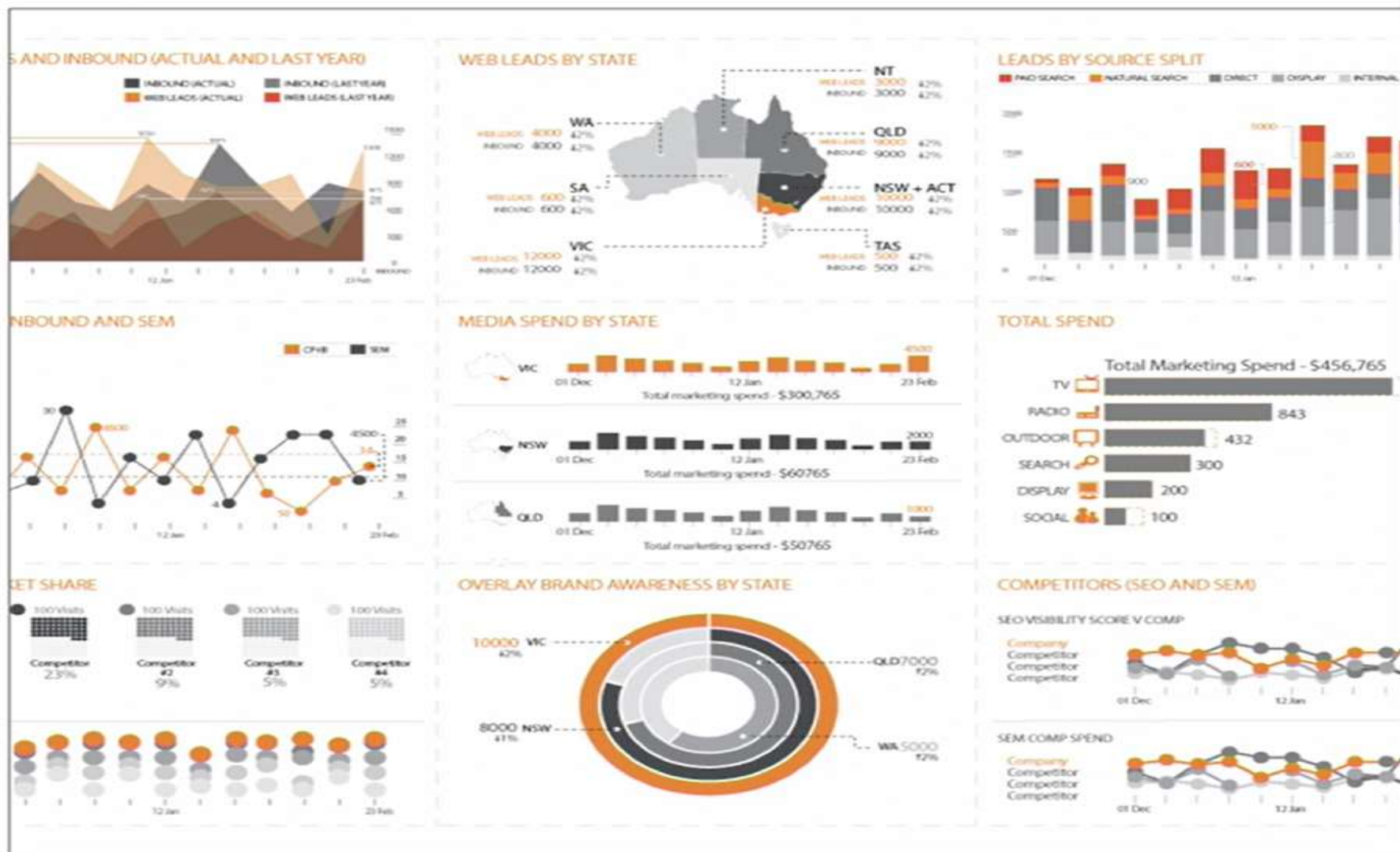
Charts : Line charts

- Show the changes in value of an attribute with respect to a x-variable which is often time

- Can be used to visually compare the values of several related attributes
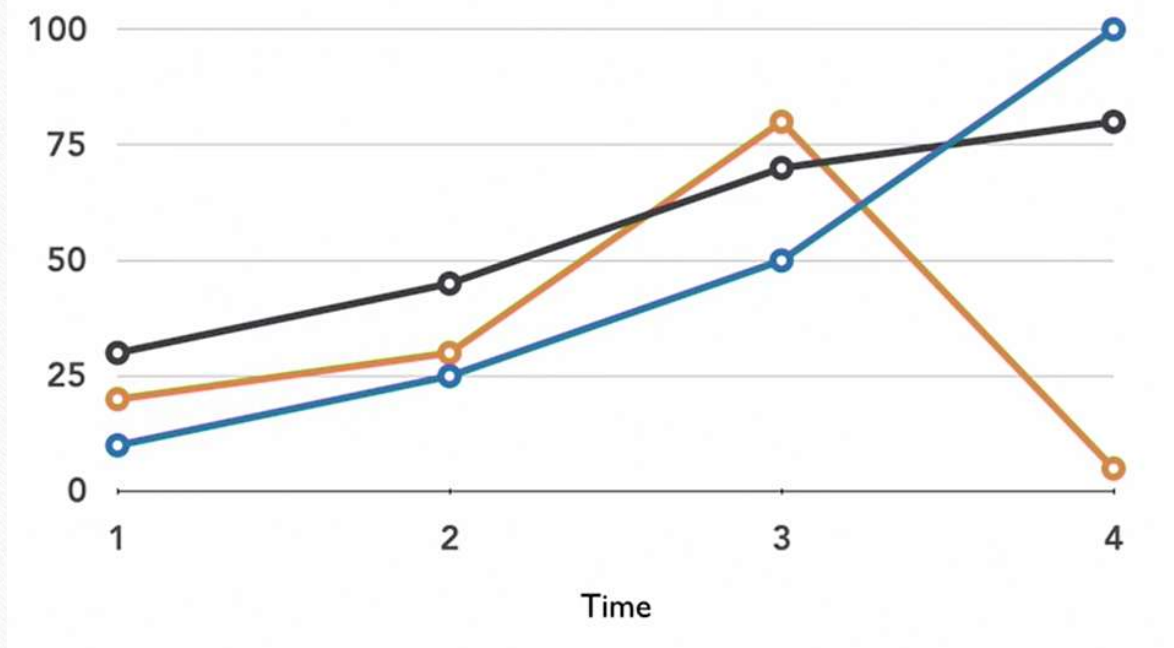
Bar chart

- Represent data attribute values within a particular data category by using bars of differing heights

- Bar charts represent observation counts within categories
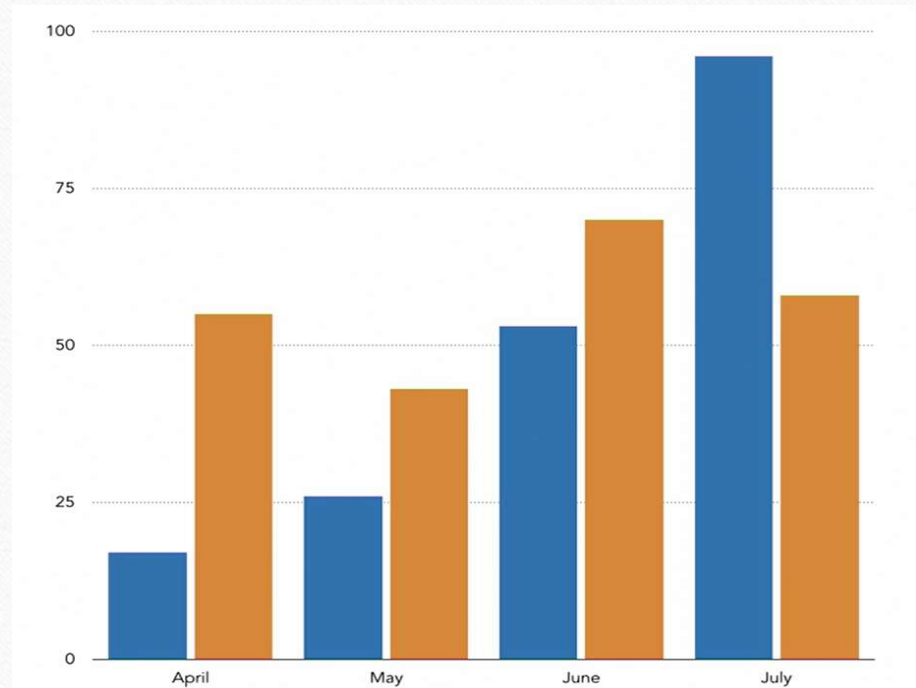
Pie chart

- Represent data attribute values using a circle and slices that comprise it

- A whole and entire set of categorical data is represented by the complete circle

- the proportions of observations that fall into the different categories are represented by proportionate pie slices
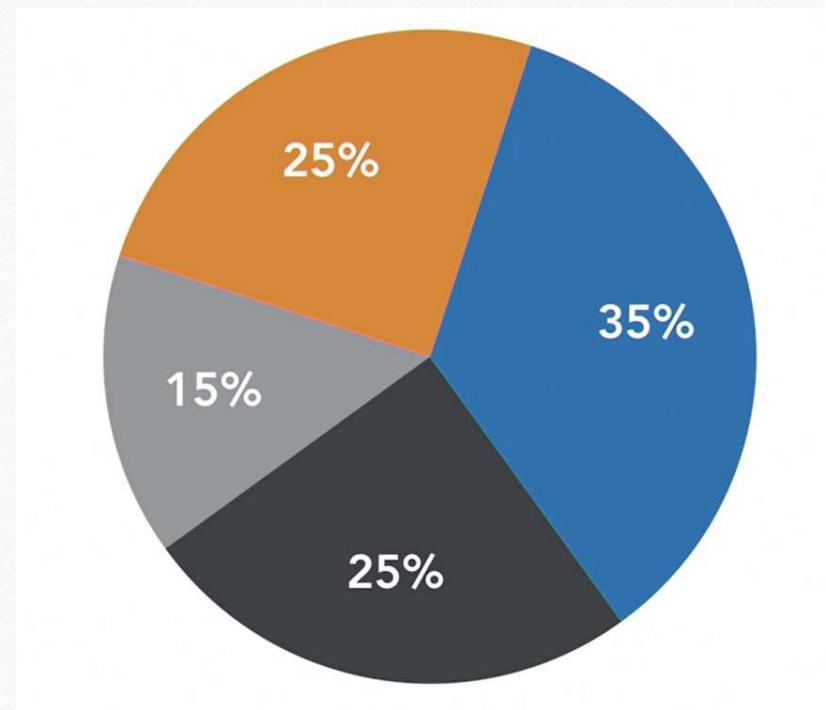
To summarize,

- you use a line chart to show changes overtime,

- you use a bar chart to show changes in categorical data,

- you use a pie chart to show categorical data as proportions of a whole

Two methods for plot building

- Functional method

  build plots by calling the plotting function on a variable or a set of variables

- Object-oriented method

  build a plot by first generating a blank figure object and then populating that object with plot and plot element

- Matplotlib and Seaborn are two most data visualization in python

Defining elements of a plot

- Plot elements add context to your plot, so the plot effectively convey meaning to its viewers

- You can use subplots to visually compare changes in data values under different conditions, like different seasons, different locations, or in different years.

- There are two methods for building plots.

    - functional method

    - object oriented method

Steps in object-oriented plotting

1. Create a blank figure object

2. Add axes to the figure

3. Generate plot(s) within figure object

4. Specify plotting and layout parameters for the plots withing your figure

subplot

- A subplot is plotting figure that contain more than one plot- or subplots

- Easy to generate a subplot in matplotlib with the subplots function
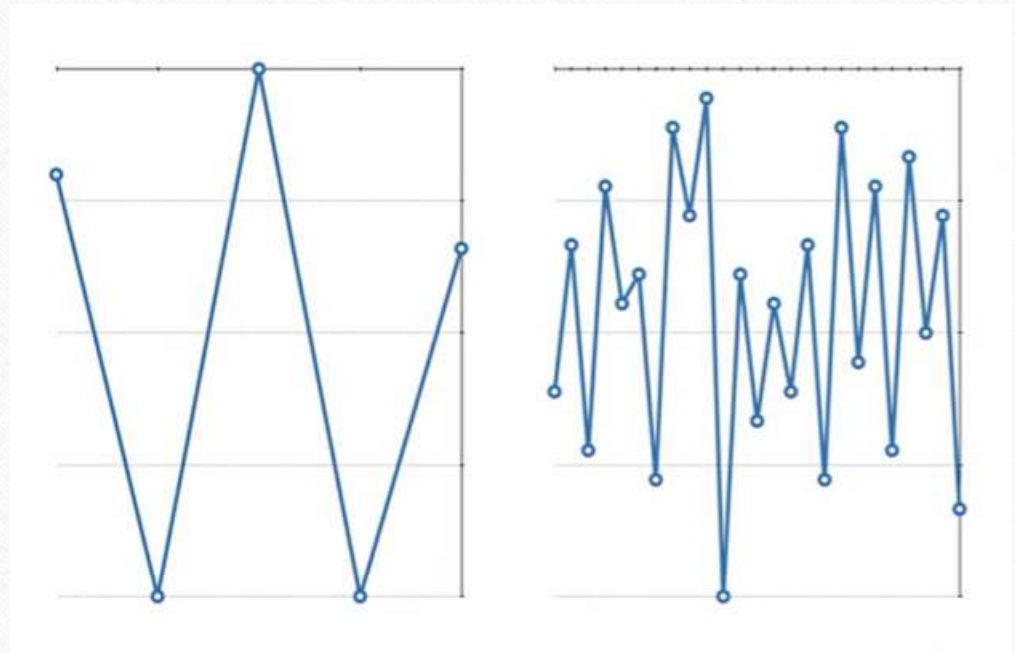
Plot formatting overview

- The reason these things are important is that enhancing chart colors and markers can help you convey a point in a way that is visually easier to understand

- If all the lines in a line chart were the same color and used the same markers, you wouldn't be able to decipher one from the other and the chart would have no meaning

- Matplotlib offers colors, line styles and marker style options

Label and annotations overview

- Labeling plot feature

- Adding a legend to your plot

- Annotating your plot

- Two methods again
  - Functional method
  - object-oriented method

Adding a legend

- Placing a legend on plot axes:

.legend(label,loc)

- Label= a string, list of strings, or anything printable with '%s' conversion; to be used as variable or category labels

- Loc= int, string, or pair of floats, default: 'upper right' ; the location of the legend

Location String	Location Code
'best'	0
'upper right'	1
'upper left'	2
'lower left'	3
'lower right'	4
'right'	5
'center left'	6
'center right'	7
'lower center'	8
'upper center'	9
'center'	10

Time series

- A time series is a measure of unit change over time for any variable under observation

- The built-in python datetime module supplies classes for manipulating dates and times in both simple and complex ways.
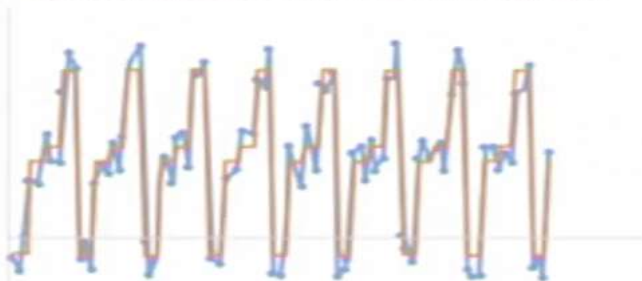
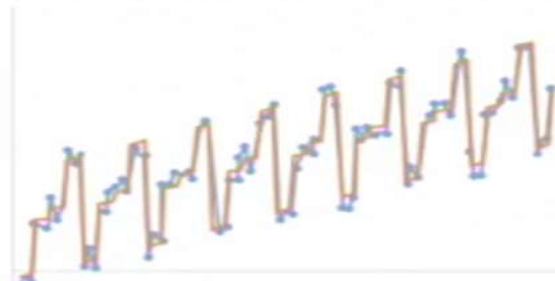# Spotting trend and seasonality

Constant Time Series
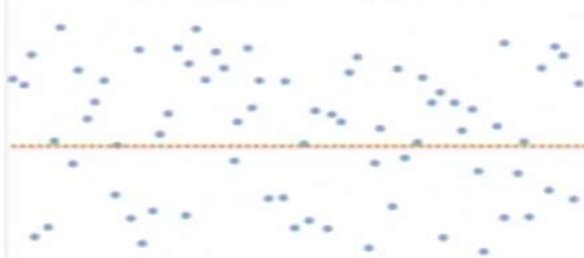
Trended Time Series
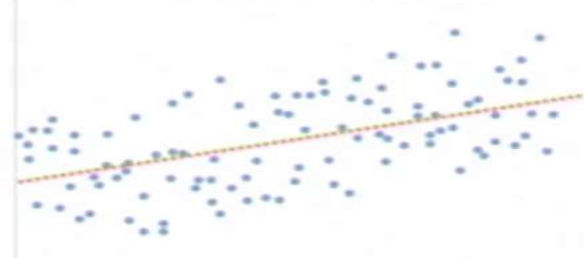
Untrended Seasonal Time Series

Trended Seasonal Time Series

Statistical plots

It is allowed to user:

- Identify outliers

- Visualize distributions

- Deduce variable types

- Discover relationships and core relations between variables in a dataset

Histograms

- It shows a variable's distribution as a set of adjacent rectangles on a data charts.

- It represent counts of data within a numerical range of values
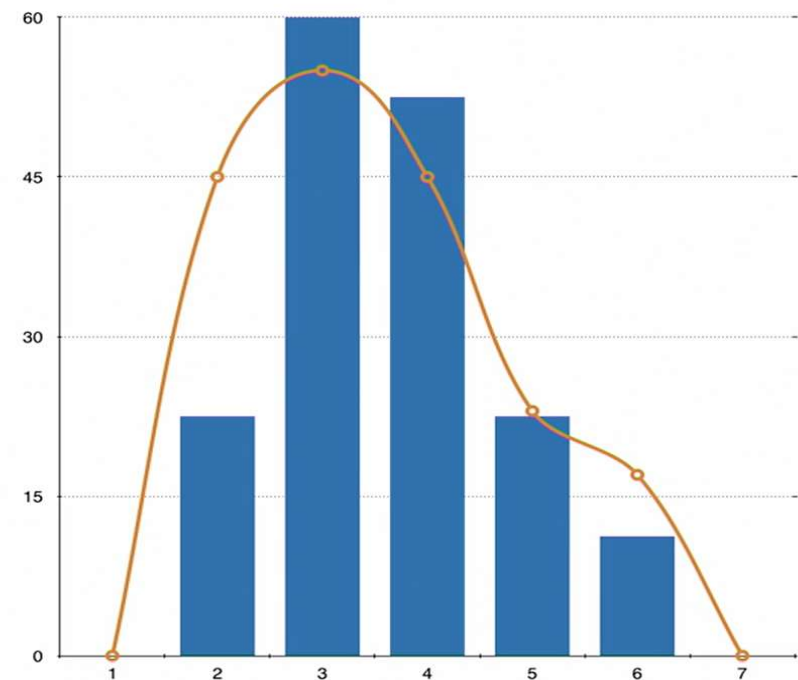
Scatter plots

- Scatterplots are useful when you want to explore interrelations or dependencies between two different variables

- These data graphics are ideal for visually spotting outliers and trends in data

- Scatterplot matrices use a matrix of scatterplots to visually display correlations between two or more variables
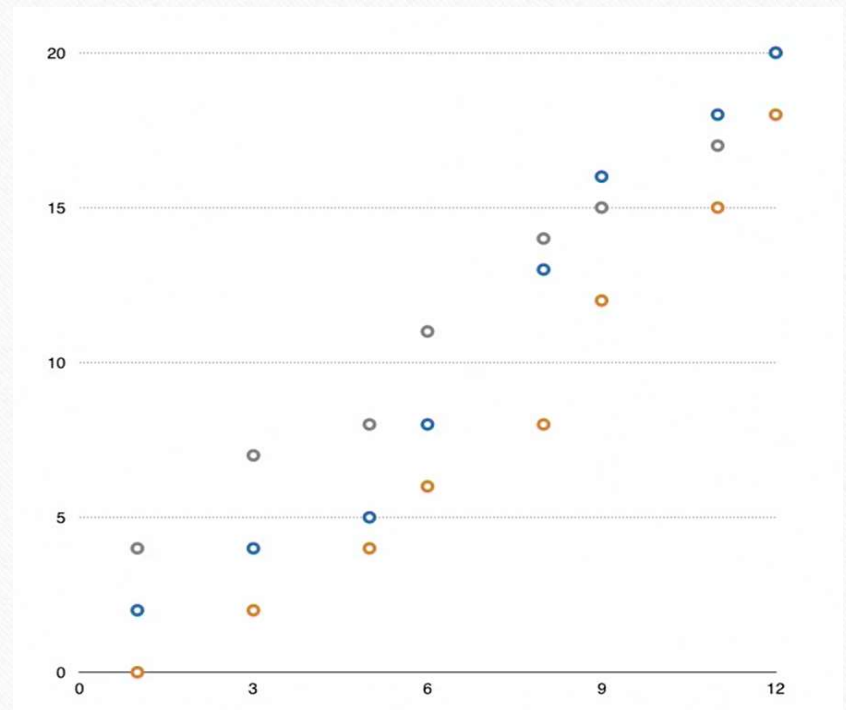
Box plots

- Boxplots are useful for seeing a variable's spread and for detecting outliers
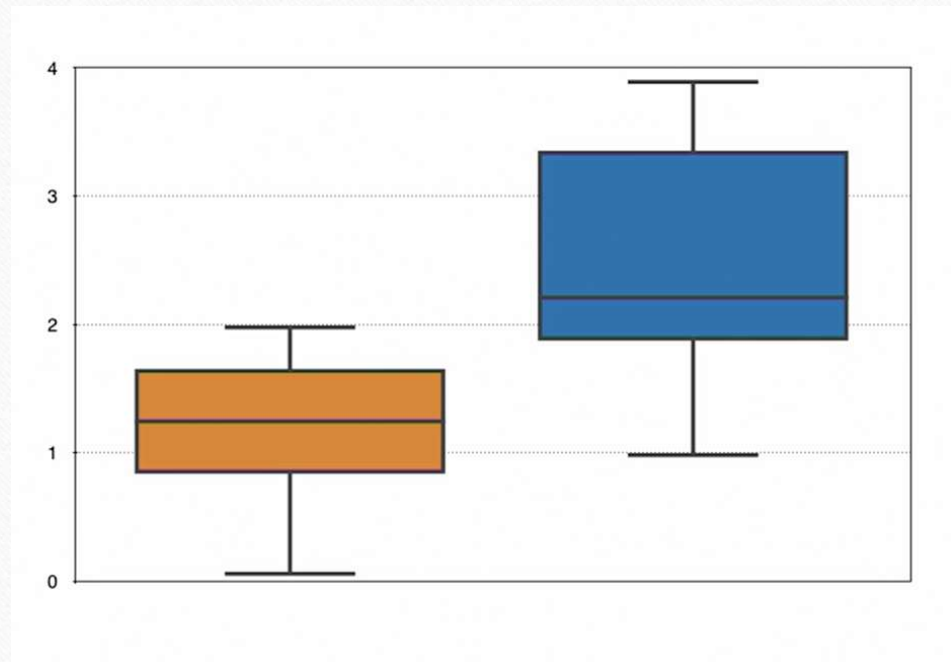
Summary

- Three types of data visualization
- Selecting optimal graphics
- Communicating with color and context
- Creating standard data graphics
- Defining elements of plot
- Plot formatting
- Creating label and annotations
- Visualizing time series
- Creating statistical data graphics

Himanshu Patel, Instructor
Saskatchewan Polytechnic
email: patelh@saskpolytech.ca
Mining Building, phoenix building