Visualization

Scott Donaldson | Adv. User Interface Software | CMU HCII 2017

1. What is it?
2. How did we get here?
3. How do we do it today?

Data Visualization

WHAT IS IT?

DATA VISUALIZATION

DATA VISUALIZATION

Statistics
Mathematics
Scientific Research
Census Data
Political/Economic

DATA VISUALIZATION

Graphical Representation
Visual Variables
(A Much Newer Notion)

Data Visualization

HOW DID WE GET HERE?

Early Visualization Methods

Symbolic Representation

Lists & Tables

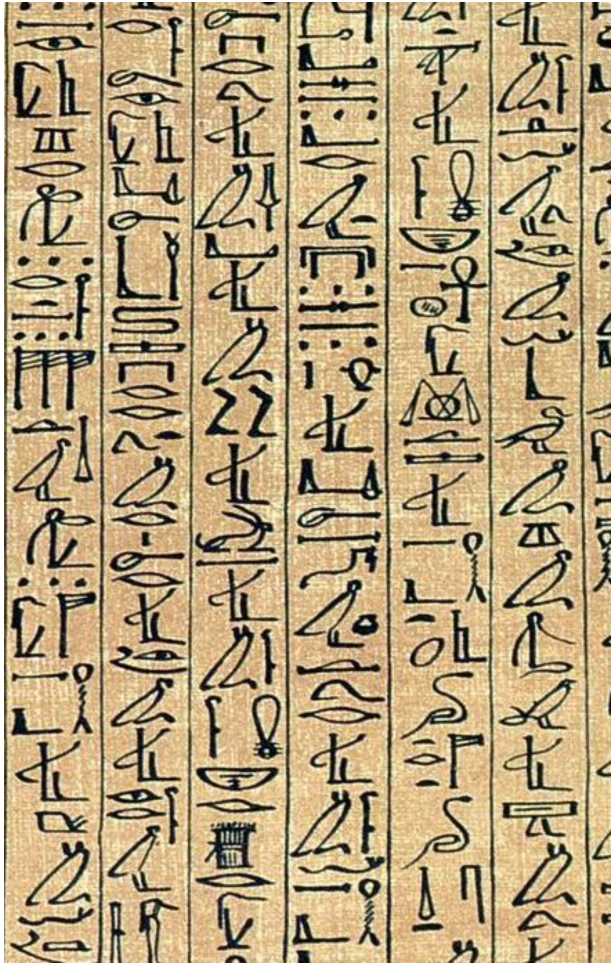
2-Dimensional Graphing

Cartography

Symbolic Representation

First step toward creating data visualization is defining visual variables (numbers, letters, icons) that communicate quantities, qualities

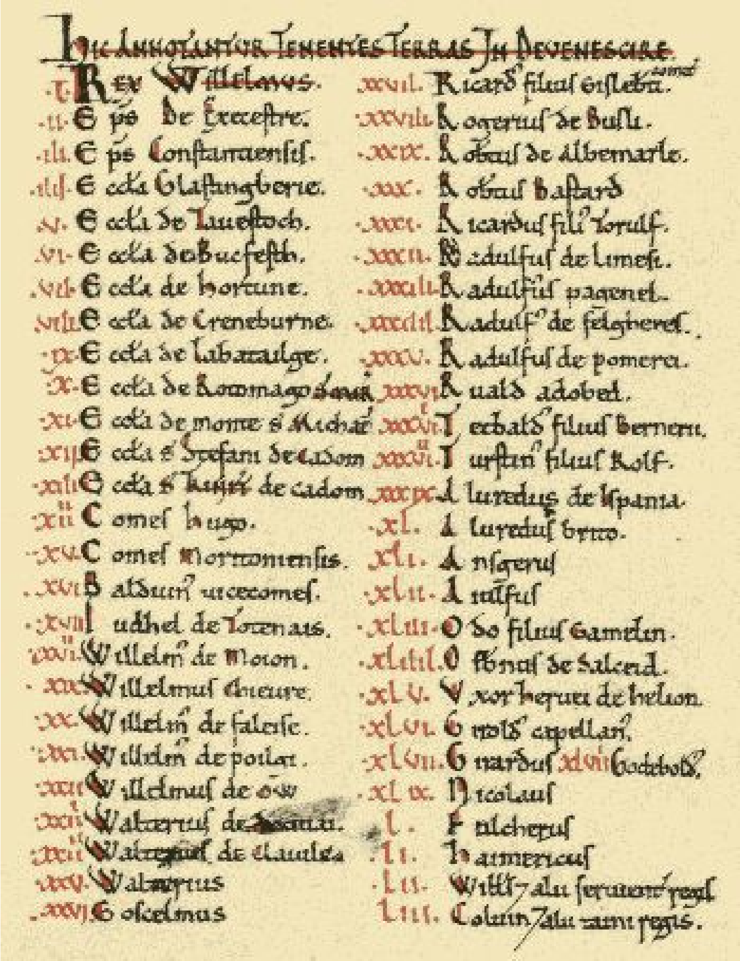
Ex. Hieroglyphics



Lists & Tables

Information displayed according to a single dimension (alphabetical order, index, size, location)

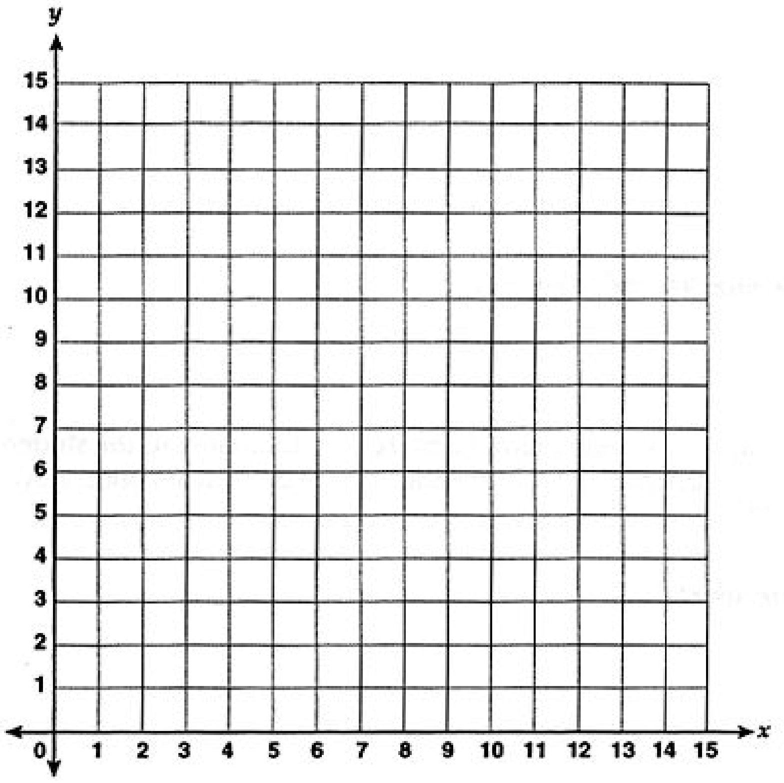
Ex. Domesday Book (1086): “Great Survey” of England



2-Dimensional Graphing

New dimension allows greater freedom in representing data — independent axis corresponds to list, data measured by dependent axis

Ex. Cartesian coordinate system
(René Descartes, 1637)



Cartography

Adding ‘metadata’ to maps like Cartesian coordinates — creates higher dimensions to display more complex information

Ex. Edmund Halley’s map of magnetic variations (1702)

Development of Foundational Visualization Techniques

Bar Chart

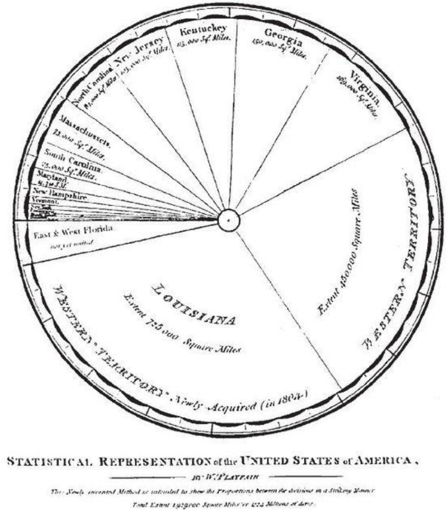
Pie Chart

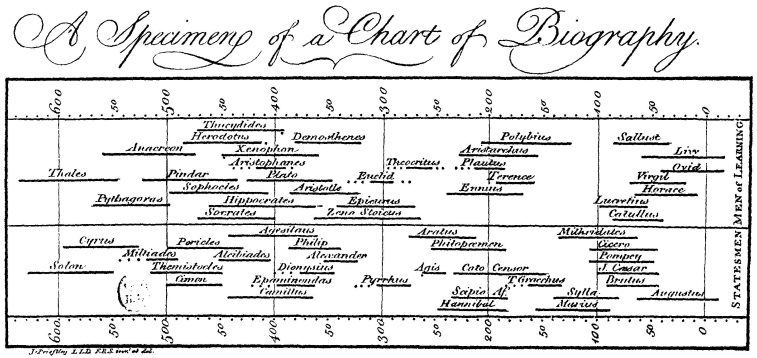
Timeline

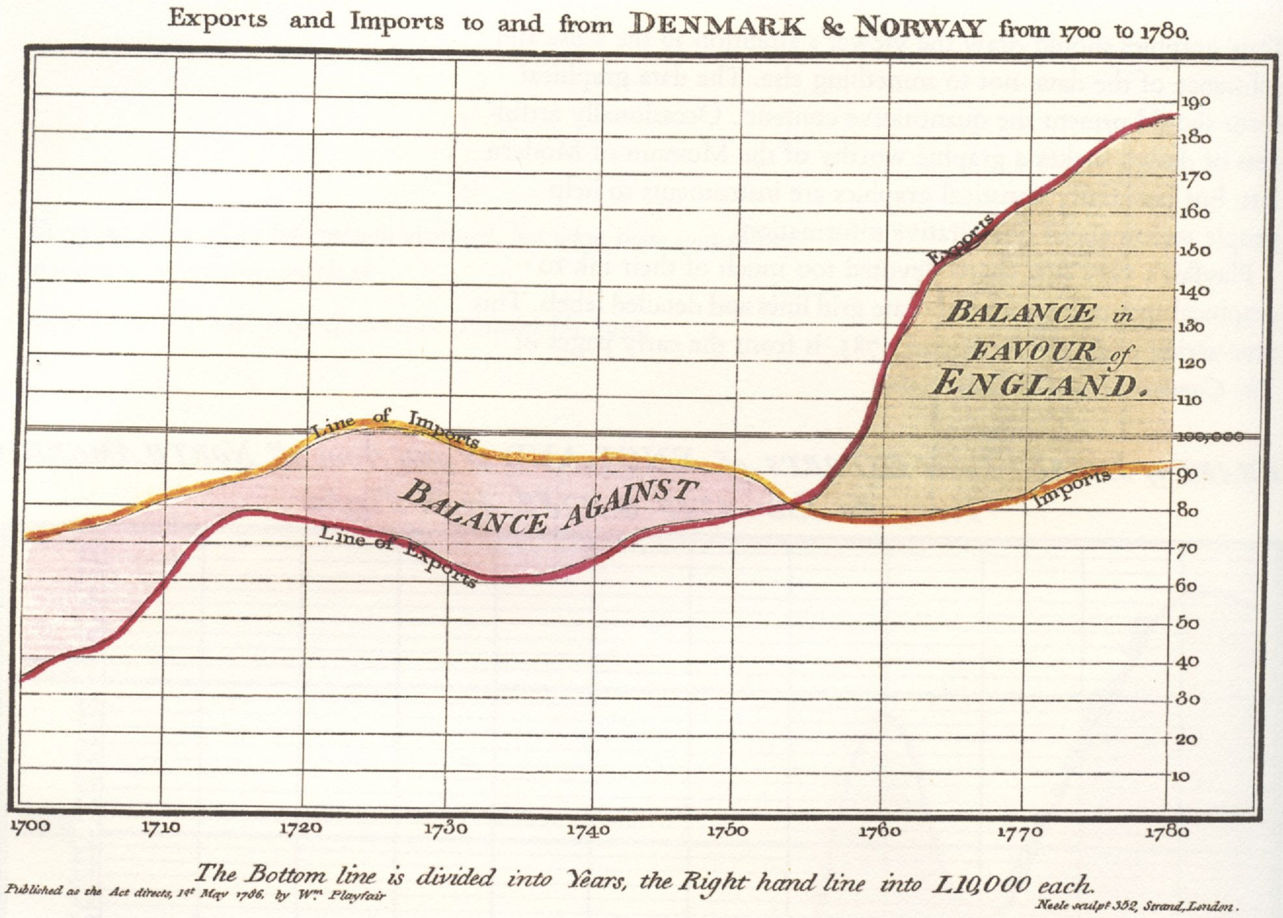
More Complex Representations

Pie Chart, Bar Chart, Timeline

Almost all basic formats of visualization invented by William Playfair (1759-1823), Scottish engineer, political economist





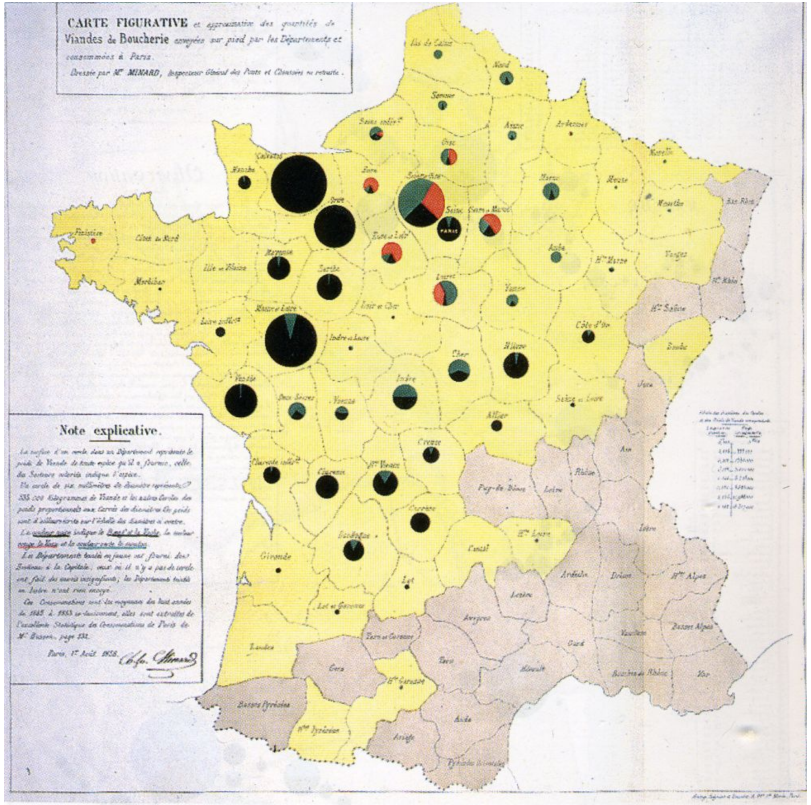


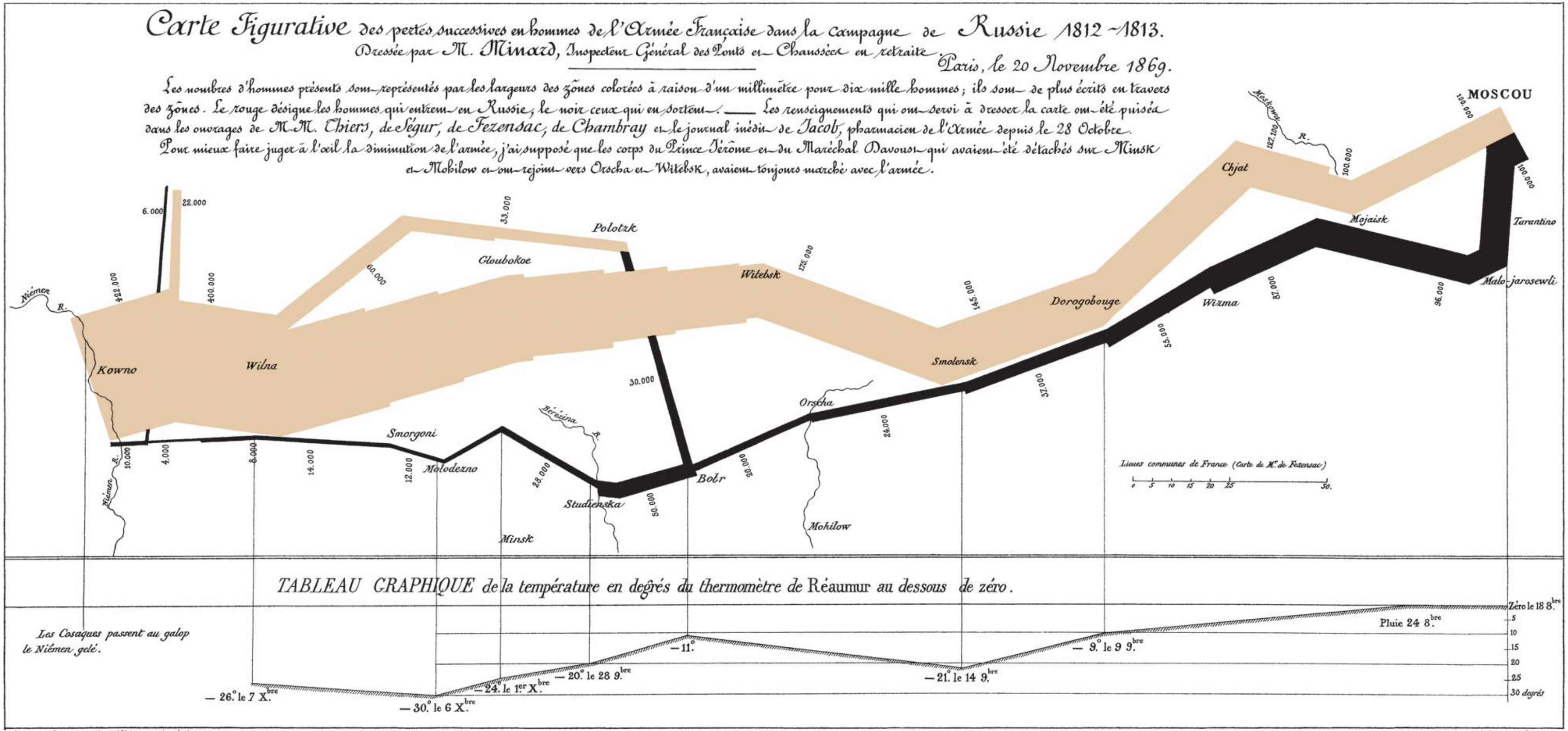
Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.
190
180
170
160
150
140
130
120
110
100,000
90
80
70
60
50
40
30
20
10
1700
1710
1720
1730
1740
1750
1760
1770
1780
Line of Imports
Line of Exports
BALANCE AGAINST
Exports
Imports
BALANCE in FAVOUR of ENGLAND.
The Bottom line is divided into Years, the Right hand line into L10,000 each.
Published as the Act directs, 14th May 1786, by Wm. Playfair
Neele sculpt 352, Strand, London.

Building on Foundations

After Playfair, others began combining types of graphics into more complex visualizations

Ex. Charles J. Minard (1781-1870), French civil engineer

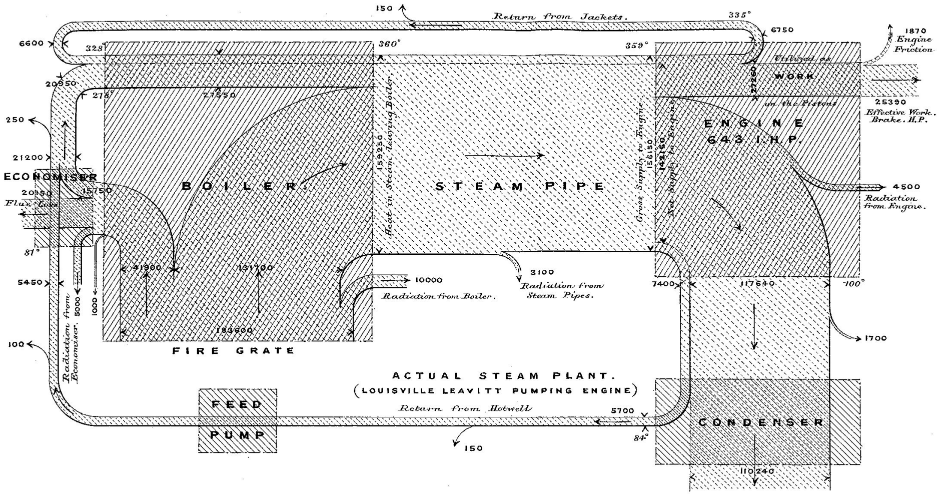


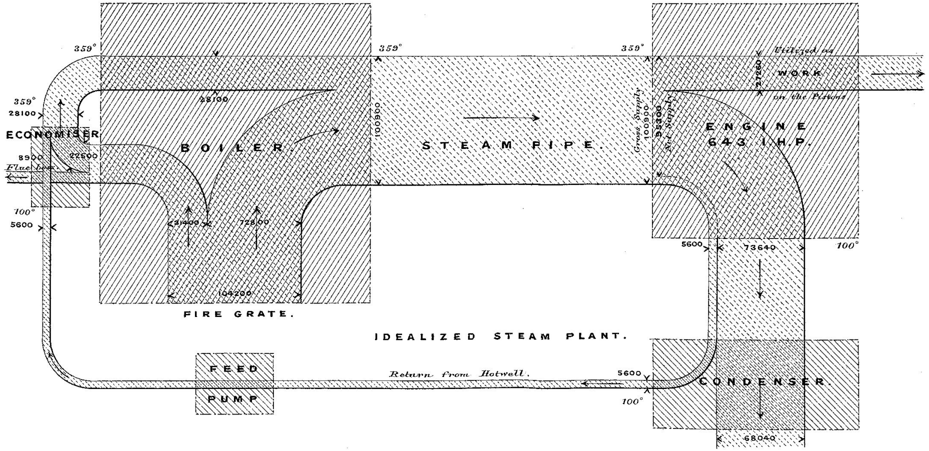


Autog. par Regnier, 8. Pas. Ste Marie St Gain à Paris.

Imp. Lith. Regnier et Dourdet.

THE THERMAL EFFICIENCY OF STEAM-ENGINES.





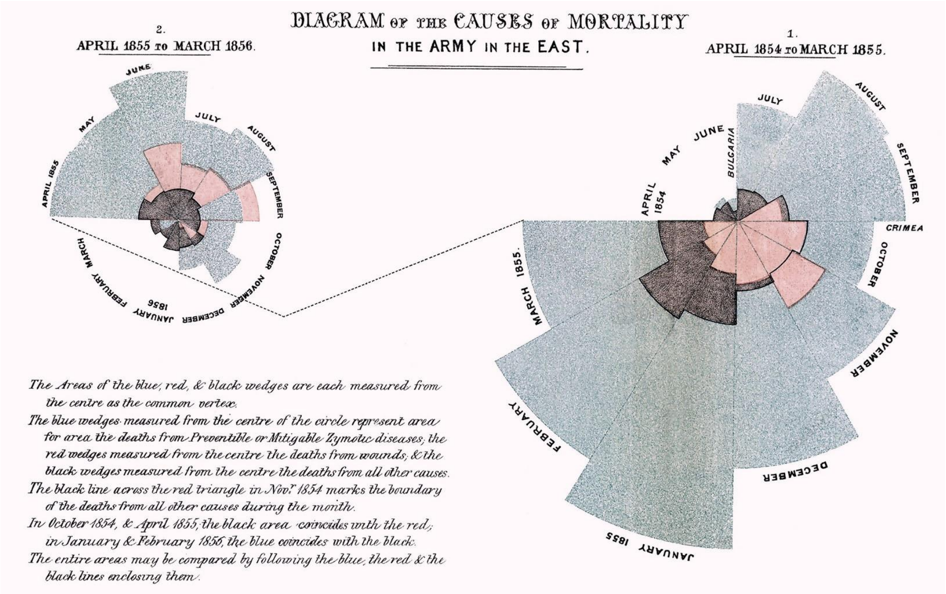
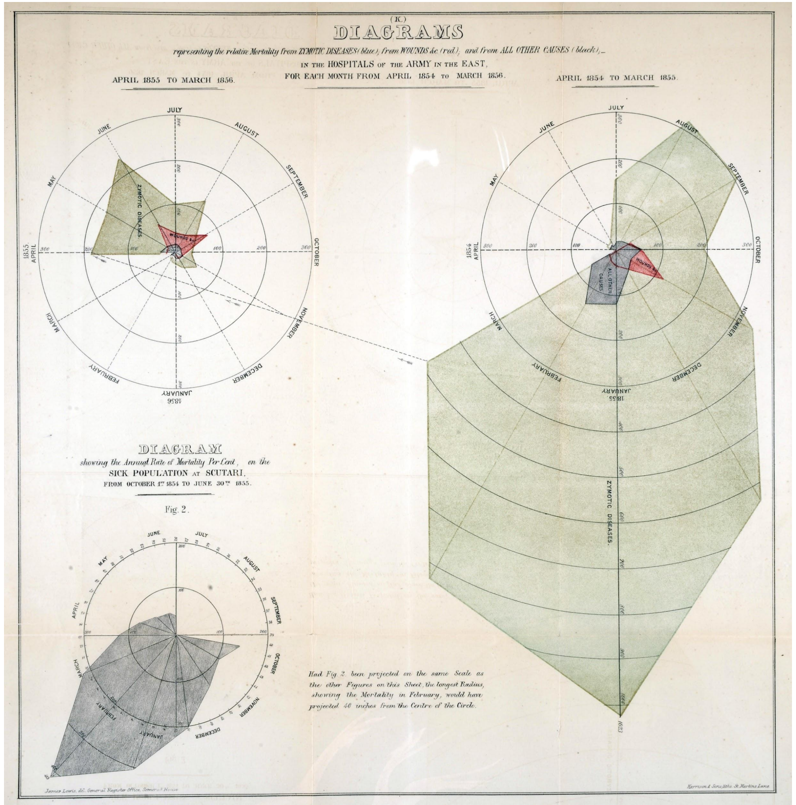


DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.
2.
APRIL 1855 TO MARCH 1856.
1.
APRIL 1854 TO MARCH 1855.
APRIL 1855
MAY
JUNE
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVEMBER
DECEMBER
JANUARY 1856
FEBRUARY
MARCH
APRIL 1854
MAY
JUNE
BULGARIA
JULY
AUGUST
SEPTEMBER
CRIMEA
OCTOBER
NOVEMBER
DECEMBER
JANUARY 1855
FEBRUARY
MARCH 1855.
The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Novr. 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.



(K.)
DIAGRAMS
representing the relative Mortality from ZYMOTIC DISEASES (blue), from WOUNDS &c (red), and from ALL OTHER CAUSES (black),—
IN THE HOSPITALS OF THE ARMY IN THE EAST,
FOR EACH MONTH FROM APRIL 1854 TO MARCH 1856.
APRIL 1855 TO MARCH 1856.
APRIL 1854 TO MARCH 1855.
ZYMOTIC DISEASES
WOUNDS &c
ALL OTHER CAUSES
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVEMBER
DECEMBER
JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
DIAGRAM
showing the Annual Rate of Mortality Per Cent, on the
SICK POPULATION AT SCUTARI,
FROM OCTOBER 1ST 1854 TO JUNE 30TH 1855.
Fig. 2.
Had Fig. 2. been projected on the same Scale as the other Figures on this Sheet, the longest Radius, showing the Mortality in February, would have projected 40 inches from the Centre of the Circle.

Modern Developments

Computers and Graphical User Interface

Visualization as Autonomous

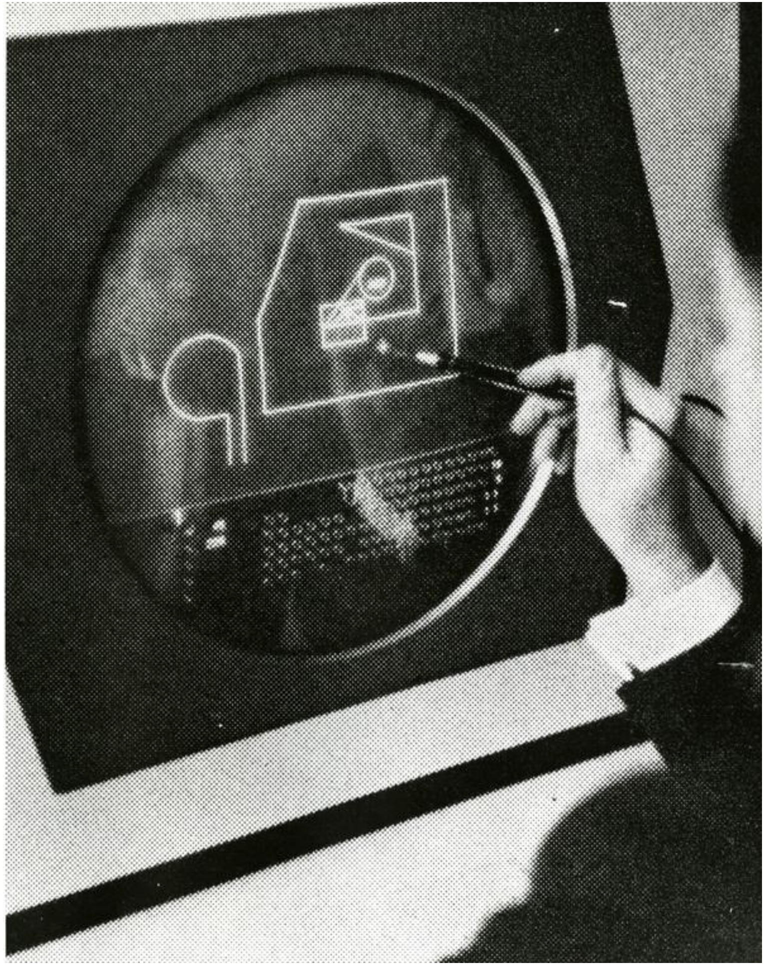
Building on This History

Computation and Visualization

Pre-1960s, computers but no graphical user interfaces

Ivan Sutherland, Sketchpad, 1962-63

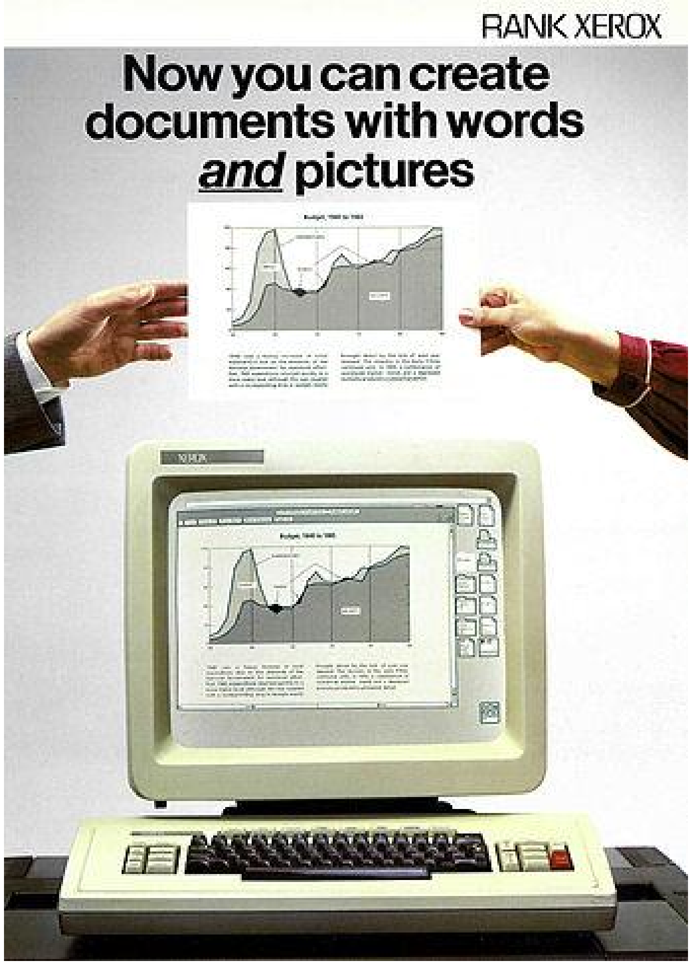
Demonstration video



Developing the Graphical User Interface

GUIs introduced in the late ‘60s and ‘70s but did not become widespread until adoption of PCs in workplaces in the ‘80s

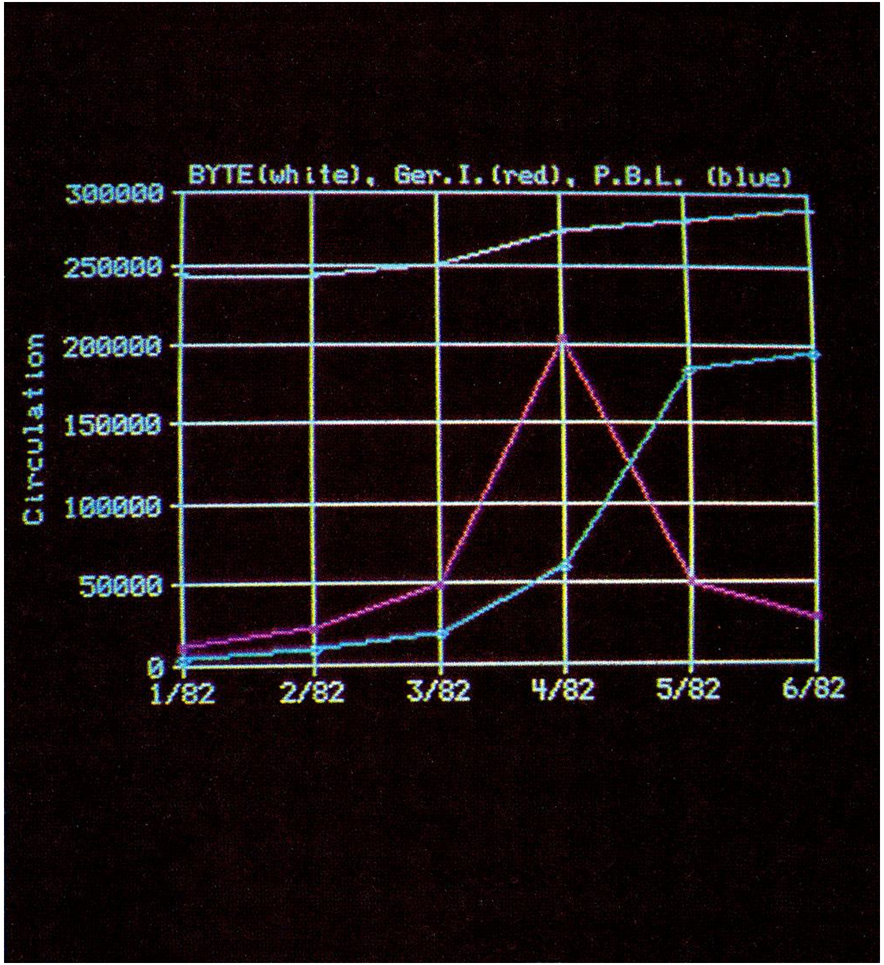
Ex. Xerox Star (1981)



Using Computers to Visualize Data

Digital visualization did not invent a language of representation; inherited it from Playfair, Minard, Nightingale

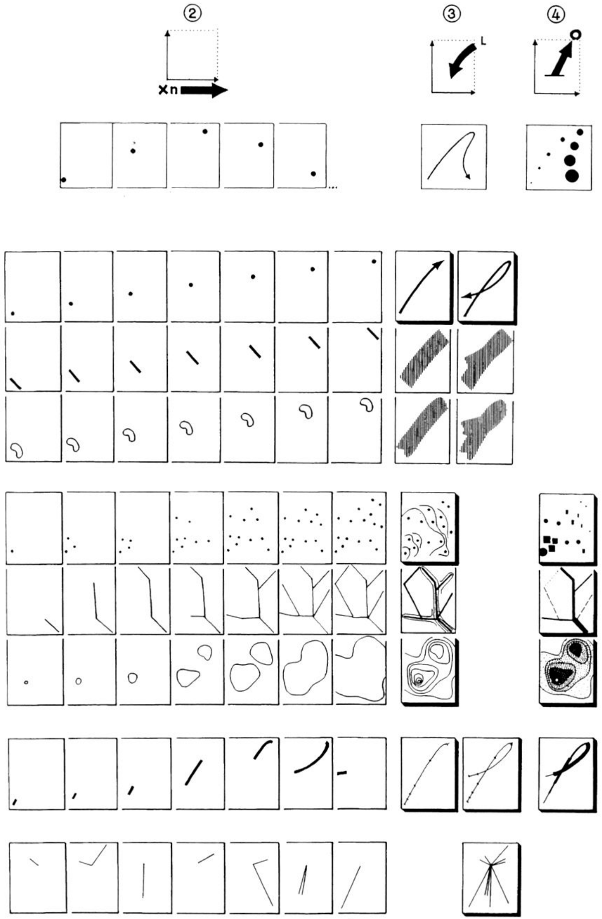
Ex. Lotus 1-2-3 Graph Maker (1983)



Visualization as Autonomous

Simultaneous to GUI development, new theories of visualization apart from the data it represented were being formulated in the academy

Ex. Jacques Bertin, *Semiologie Graphique* (1967); described elements of visualization as parameters, variables (language of computation)

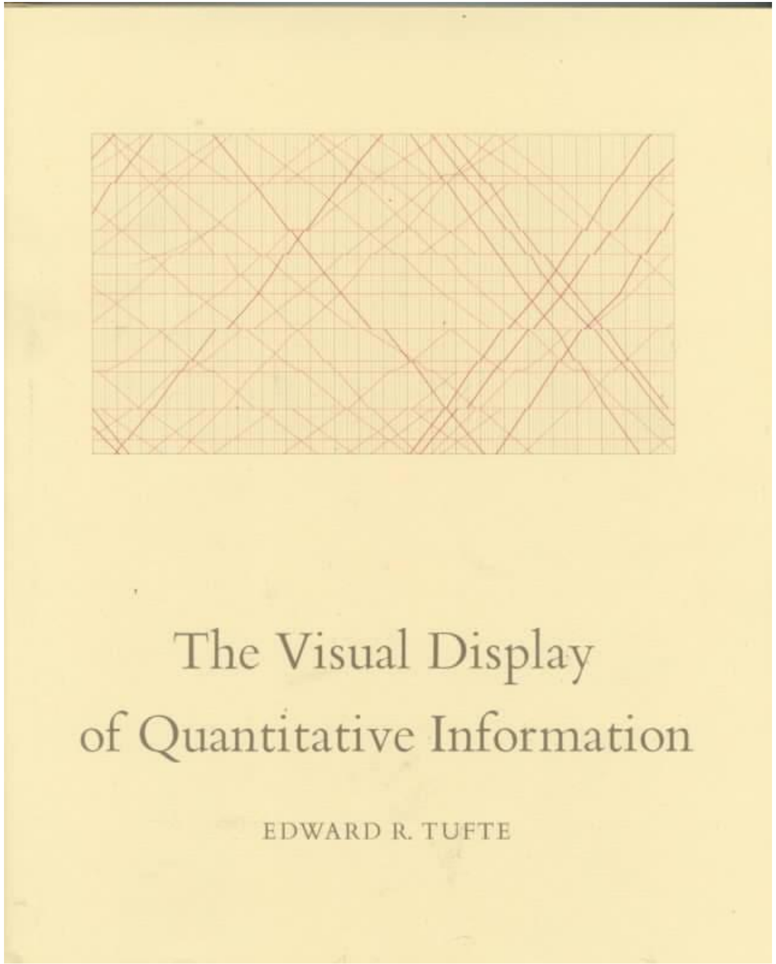


J. Bertin, SÈmiologie graphique, EHESS, 1998, p343

Creating the History of Visualization

Edward Tufte's *The Visual Display of Quantitative Information* (1983) identified the lineage of Playfair, Minard, Nightingale, and others

Made visualization accessible to a broader audience than statisticians/
engineers, to include businesses and the general public



Data Visualization

HOW DO WE DO IT TODAY?

Contemporary Software for Data Visualization

Proprietary, Business-Oriented	Open Source, General Purpose
AnyChart (2003)	R Language (1993)
FusionCharts (2003)	Processing (2001)
Tableau (2005)	Protovis (2009) → D3.js (2011)

AnyChart

Originally Flash, now HTML5 + JavaScript

JavaScript library + API (no GUI) (Documentation)

Charts: Area, Bar, Box, Bubble, Bullet, Column, Doughnut, Error, Funnel, Heat Map, Japanese Candlestick, Jump Line, Line, Marker, OHLC, Pareto, Pie, Polar Plot, Pyramid, Radar, Scatter Plot, Sparkline, Spline, Spline Area, TreeMap, 3D Bar, 3D Doughnut, 3D Pie, and more

AnyChart Example

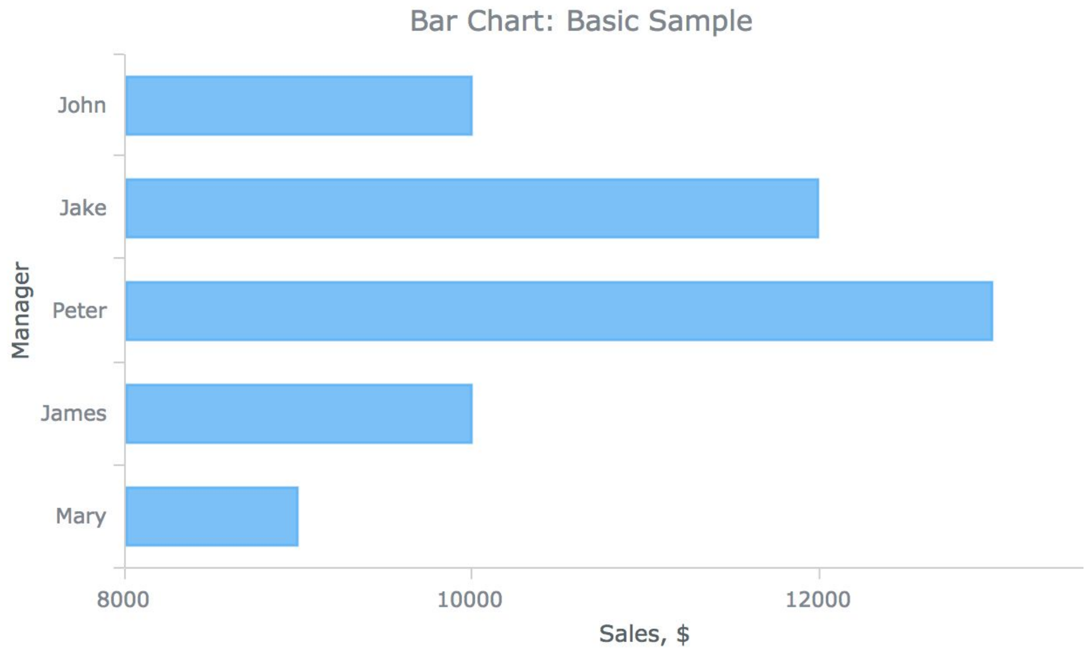
```
// create a data set
var data = anychart.data.set([
  ["John", 10000],
  ["Jake", 12000],
  ["Peter", 13000],
  ["James", 10000],
  ["Mary", 9000]
]);

// create a chart
var chart = anychart.bar();

// create a bar series and set the data
var series = chart.bar(data);

// set the container id
chart.container("container");

// initiate drawing the chart
chart.draw();
```



Tableau

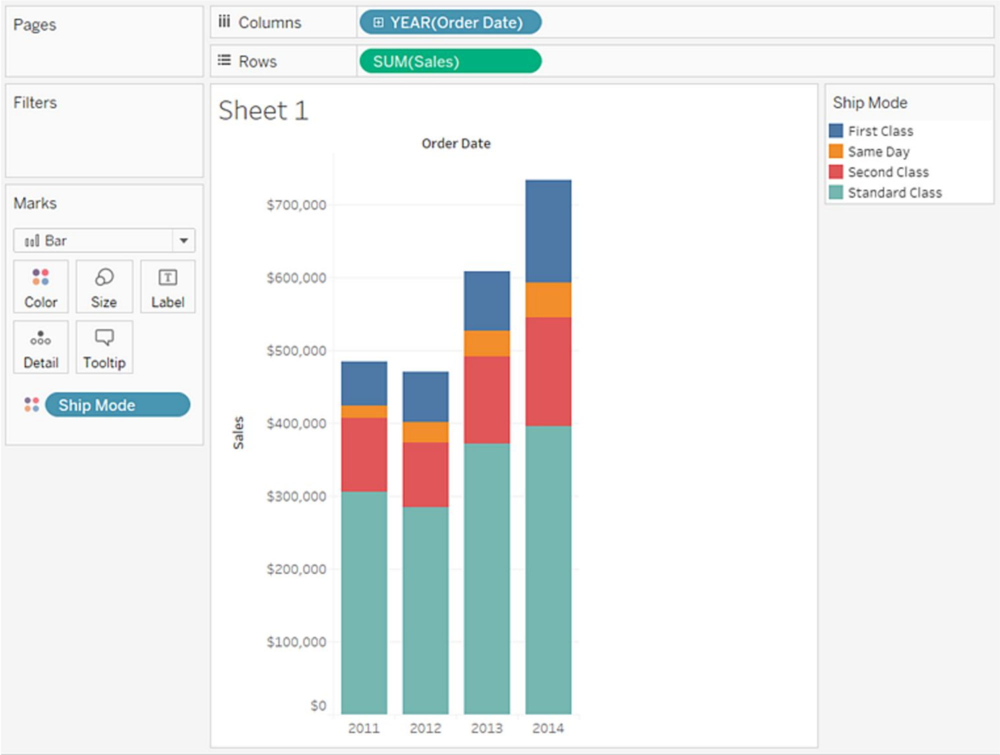
Desktop app/web app ($, but free version with limited features)

Drag and drop GUI, some customization but difficult to adjust fine-grain details

Charts: Bar, Table, Line, Scatter Plot, Heat Map, Histogram, Gantt, Pie, Treemap, Box Plot, Packed Bubble, Map (many more not listed on website)

Tableau Example

1. Connect to data source
2. Drag the **Order Date** dimension to **Columns**
3. Drag the **Sales** measure to **Rows**
4. On the **Marks** card, select **Bar** from the drop-down list
5. Drag the **Ship Mode** dimension to **Color** on the **Marks** card
6. Drag the **Region** dimension to **Rows**, and drop it to the left of **Sales** to produce multiple axes for sales by region



FusionCharts

JavaScript library + extensions for jQuery, AngularJS, PHP, Asp.NET, JSP, Django, React, Ruby on Rails

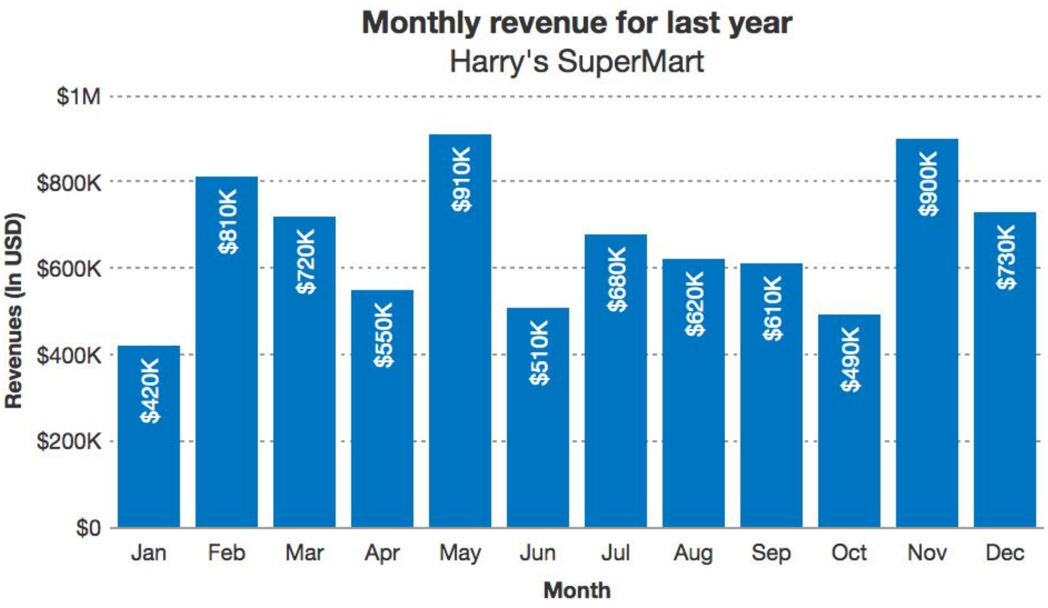
Declarative, not imperative:
Write configuration markup
(JSON, XML, or HTML + JS)

Powers Federal IT Dashboard,
Google Docs chart widget

President Obama is a fan!

FusionCharts Example

```
{
  "chart": {
    "caption": "Monthly revenue for last year",
    "subCaption": "Harry's SuperMart",
    "xAxisName": "Month",
    "yAxisName": "Revenues (In USD)",
    "numberPrefix": "$",
    "theme": "fint"
  },
  "data": [
    {
      "label": "Jan",
      "value": "420000"
    },
    {
      "label": "Feb",
      "value": "810000"
    },
    {
      "label": "Mar",
      "value": "720000"
    },
    {
      "label": "Apr",
      "value": "550000"
    },
```



R Language

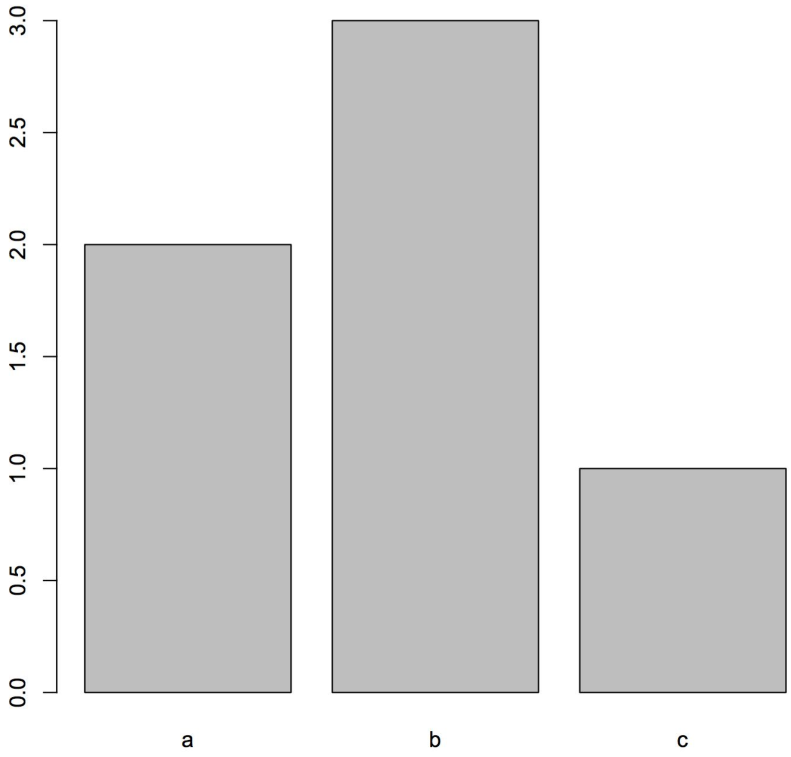
Open source language for statistics, analysis, and visualization

~8,000 packages, many specifically visualization-oriented (ggplot2)

Operates efficiently at scale (millions of data points), favored by data analysts both for computing power and visualization

R Language Example

```r
H <- c("a","a","b","b","b","c")
counts <- table(H)
barplot(counts)
```



Processing

Java-based IDE for “learning how to code within the context of the visual arts”

Libraries for JavaScript, Python

Plugins/libraries for audio, video, 3d visualization, physical devices (Arduino, Raspberry Pi)

Processing Example

```
void barchart(String filename,float yf1, float xf1, float yscale, float xscale,
   //Declare a float variabe for the max y axis value.
   float ymax;

   //Declare a float variable for the minimum y axis value.
   float ymin = 0;

   //Set the stroke color to a medium gray for the axis lines.
   stroke(175);

   //draw the axis lines.
   line(xf1-3,yf1+2,xf1+xscale,yf1+2);
   line(xf1-3,yf1+2,xf1-3,yf1-yscale);

   //turn off strokes for the bar charts.
   noStroke();

   //loop the chart drawing once.
   for (int c1 = 0; c1 < 1; c1++){

   //Set the start x point value.
   float xfstart = xf1;

   //Load the file.
   String[] lines = loadStrings(filename);
```

Sorry, no bar chart to show!

Moral of the story: lots of coding “by hand” to do simple things

D3.js

Acronym for "data-driven documents." JavaScript library for "manipulating documents based on data."

Steep learning curve, but extremely flexible/extensive

Gained popularity through New York Times interactive projects

D3.js Example

```
var svg = d3.select("svg"),
    margin = {top: 20, right: 20, bottom: 30, left: 40},
    width = +svg.attr("width") - margin.left - margin.right,
    height = +svg.attr("height") - margin.top - margin.bottom;

var x = d3.scaleBand().rangeRound([0, width]).padding(0.1),
    y = d3.scaleLinear().rangeRound([height, 0]);

var g = svg.append("g")
    .attr("transform", "translate(" + margin.left + "," + margin.top

d3.tsv("data.tsv", function(d) {
  d.frequency = +d.frequency;
  return d;
}, function(error, data) {
  if (error) throw error;

  x.domain(data.map(function(d) { return d.letter; }));
  y.domain([0, d3.max(data, function(d) { return d.frequency; })]);

  g.append("g")
      .attr("class", "axis axis--x")
      .attr("transform", "translate(0," + height + ")")
      .call(d3.axisBottom(x));

  g.append("g")
      .attr("class", "axis axis--y")
      .call(d3.axisLeft(y).ticks(10, "%"))
    .append("text")
```