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
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.
Building on Foundations

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

Ex. Charles J. Minard (1781-1870), French civil engineer
Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Peinte par M. Minard, Ingénieur Général des Ponts et Chaussées en retraite

Paris, le 20 Novembre 1869.

Les nombres d'hommes présentés sont représentés par les longueurs des zones ombreés à raison d'une millimètre pour dix mille hommes; ils seront plus éloignés ou suivant des zones. Le rouge désigne les hommes qui sont restés en Russie; le noir ceux qui en sortirent; les zones ombreés qui ont tenu à revoir la carte ont été pris dans les proportions de L. M. Chevaillier, le général de Lescure, et de D. de Tourterelle, de Chambrey et le journal de Musset, photographe de l'Armée depuis le 18 Octobre.

Comme mieux faire juger à l'œil la diminution de l'Armée, j'ai soumis que le corps du Prince Napoléon et du Maréchal Davoust, qui avaient été attachés aux Neuf-Rossiennes, à la région de Cour. Il est de même, nous avons toujours marché avec l'Armée.

TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

[Diagramme avec des zones ombrées représentant les pertes successives en hommes de l'Armée Française, avec des températures indiquées sur le bas de la page.]
Diagram of the Causes of Mortality in the Army in the East.

The areas of the blue, red, and black wedges are each measured from the centre as the common section. The blue wedge measured from the centre of the circles represents the area of the deaths from preventable or mitigable Zymotic diseases; the red wedge measured from the centre the deaths from wounds; and the black wedge measured from the centre the deaths from all other causes.

The black line across the red triangle in April 1854 marks the boundary of the deaths from all other causes during that month.

In October 1854, a red line was drawn from the black area to the red area in January 1855 the blue wedge coincides with the black.

The entire areas may be compared by following the blue, the red, and the black lines enclosing them.
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.

Creating the History of Visualization


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

<table>
<thead>
<tr>
<th>Proprietary, Business-Oriented</th>
<th>Open Source, General Purpose</th>
</tr>
</thead>
</table>
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)
Sweden is the highest achieving country due to the optimal national infrastructure and support for students. A similarly strong performance is seen in Taiwan, and Singapore. This is followed by the United Kingdom, Switzerland, and the United States. The performance of Brazil, Argentina, and Spain is considerably lower in comparison.
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

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

var g = svg.append("g")
    .attr("class", "axis axis--y")
    .call(d3.axisLeft(y).ticks(10, "s"));
    .append("text")
```