Data Visualization

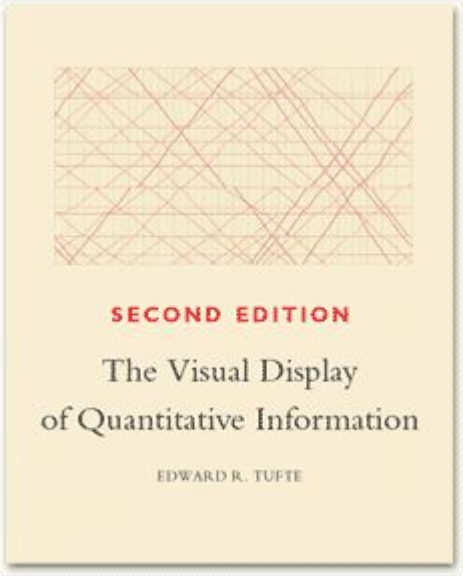
2017-01-23

Agenda

- Examples
- Software
- Theory
- Making good visualizations
- Preattentive cognition

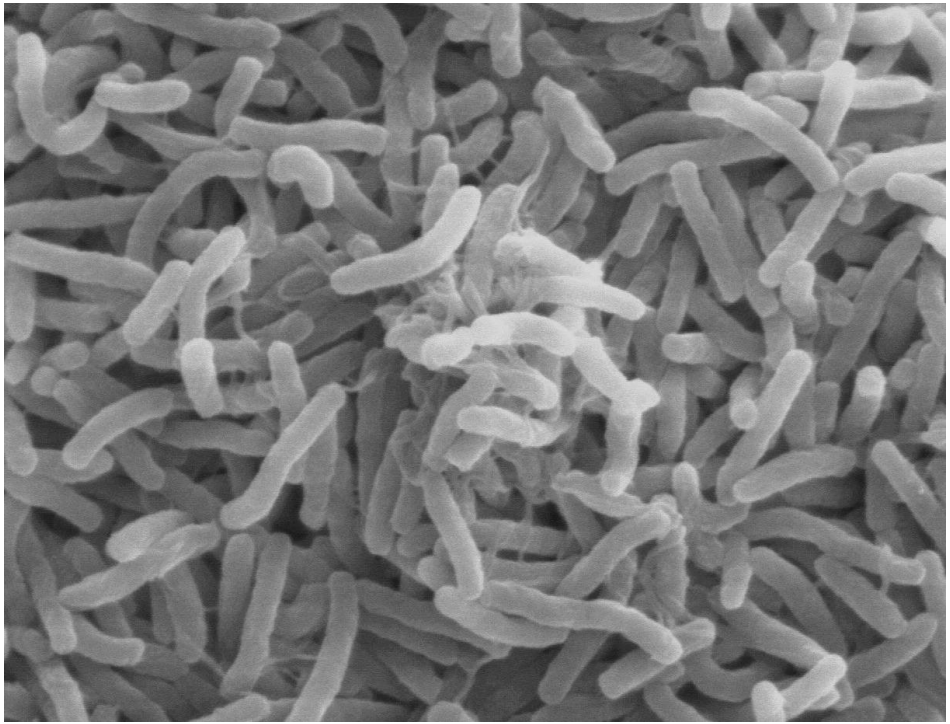
See Also

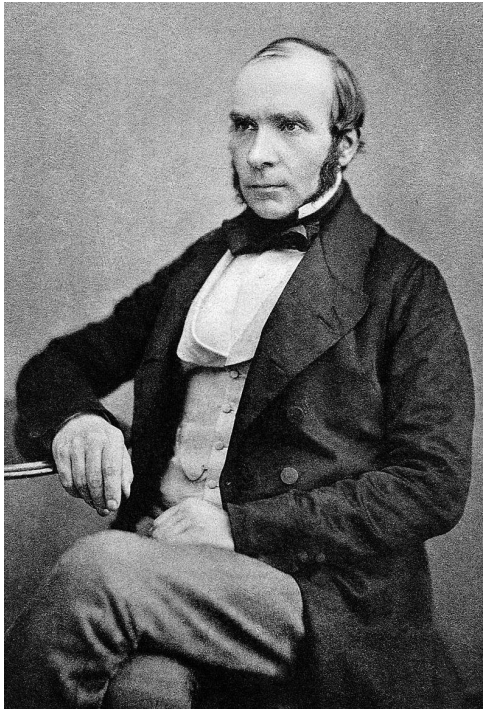
The Visual Display of Quantitative Information by Edward Tufte



Examples

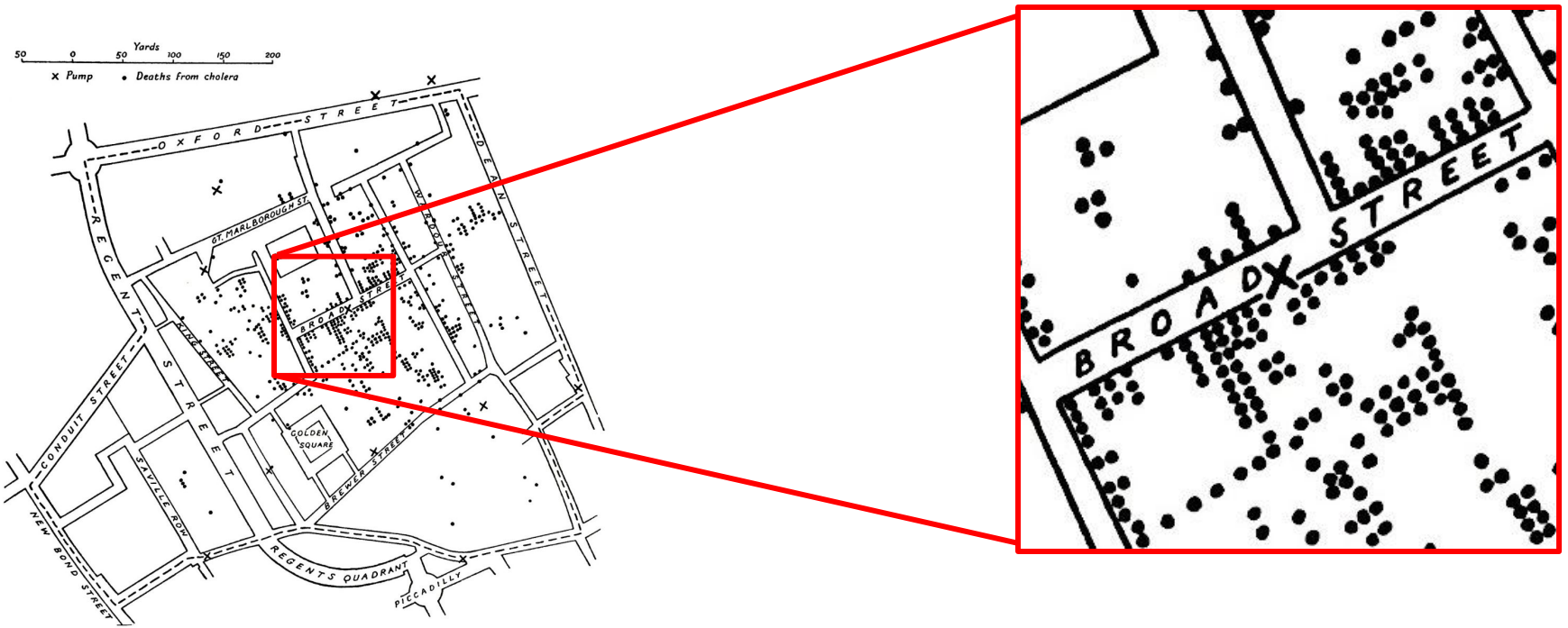
John Snow and Cholera



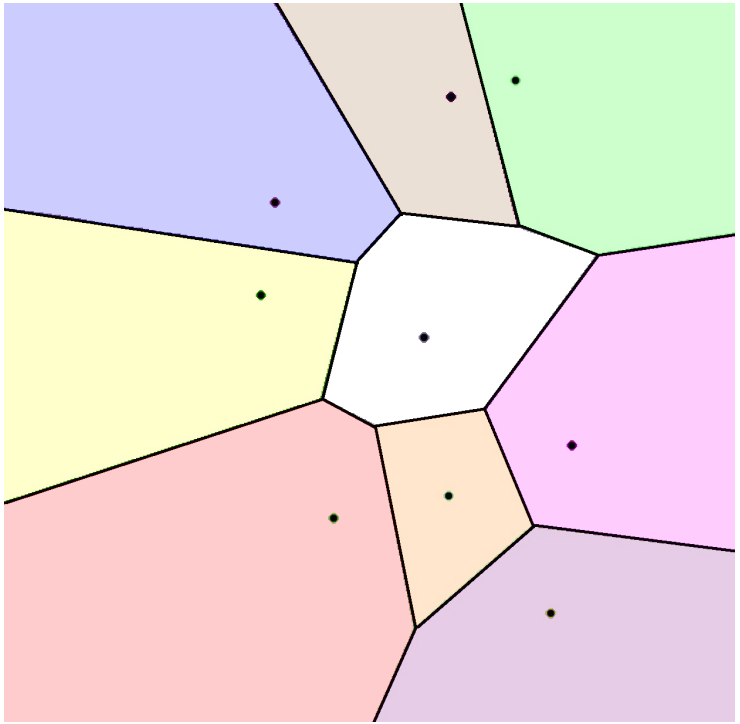


John Snow and Cholera

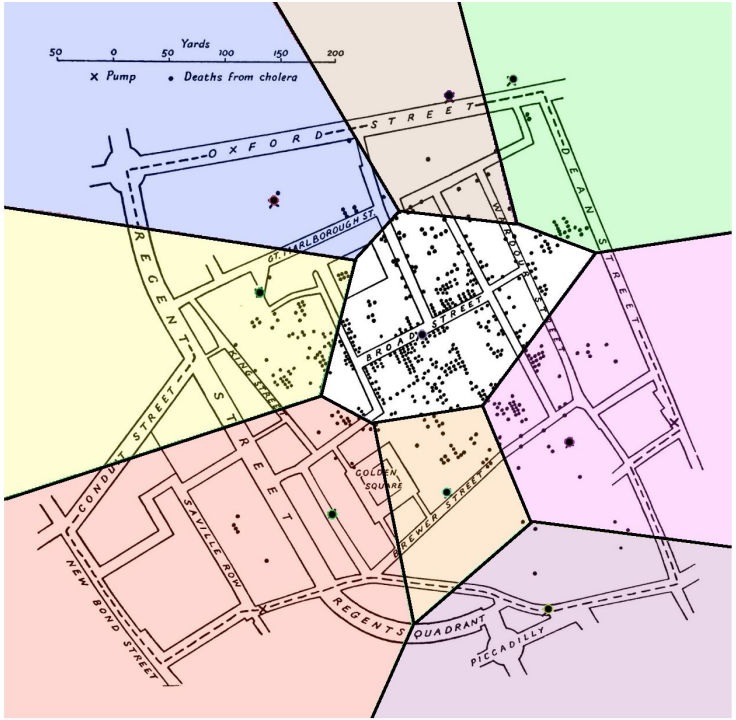
John Snow and Cholera



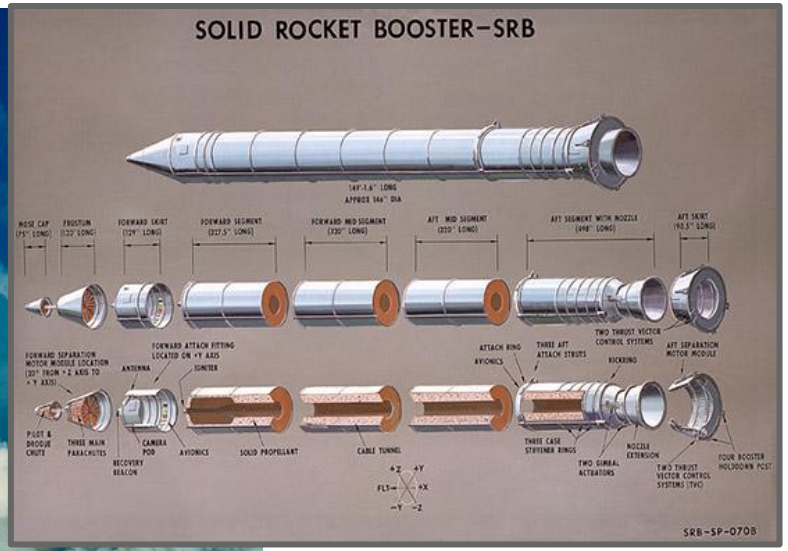
John Snow and Cholera



John Snow and Cholera

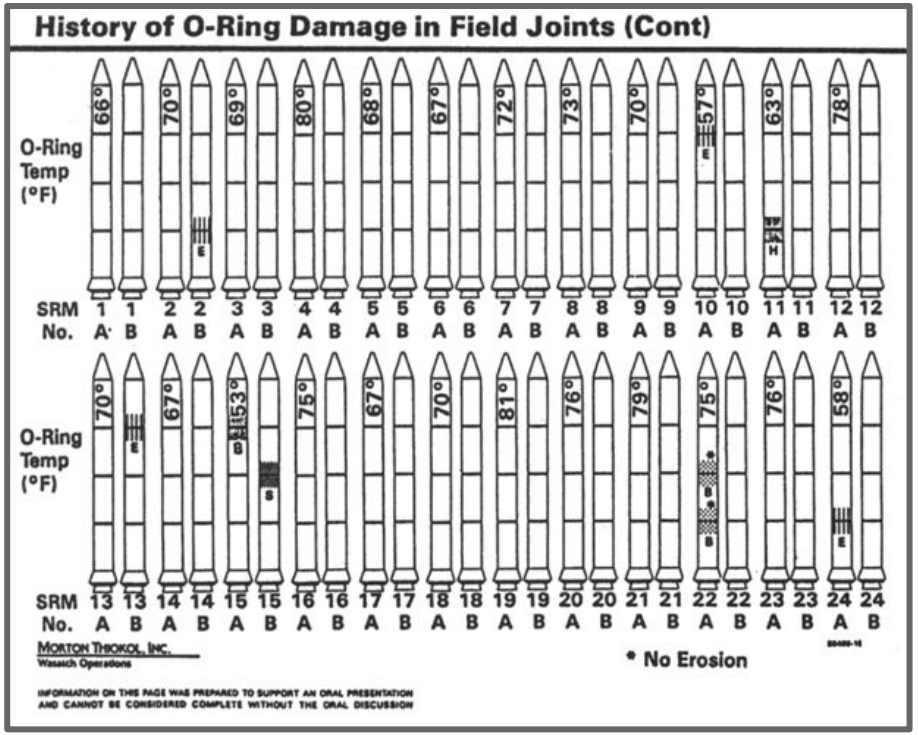


Morton Thoikol and Solid Rocket Boosters

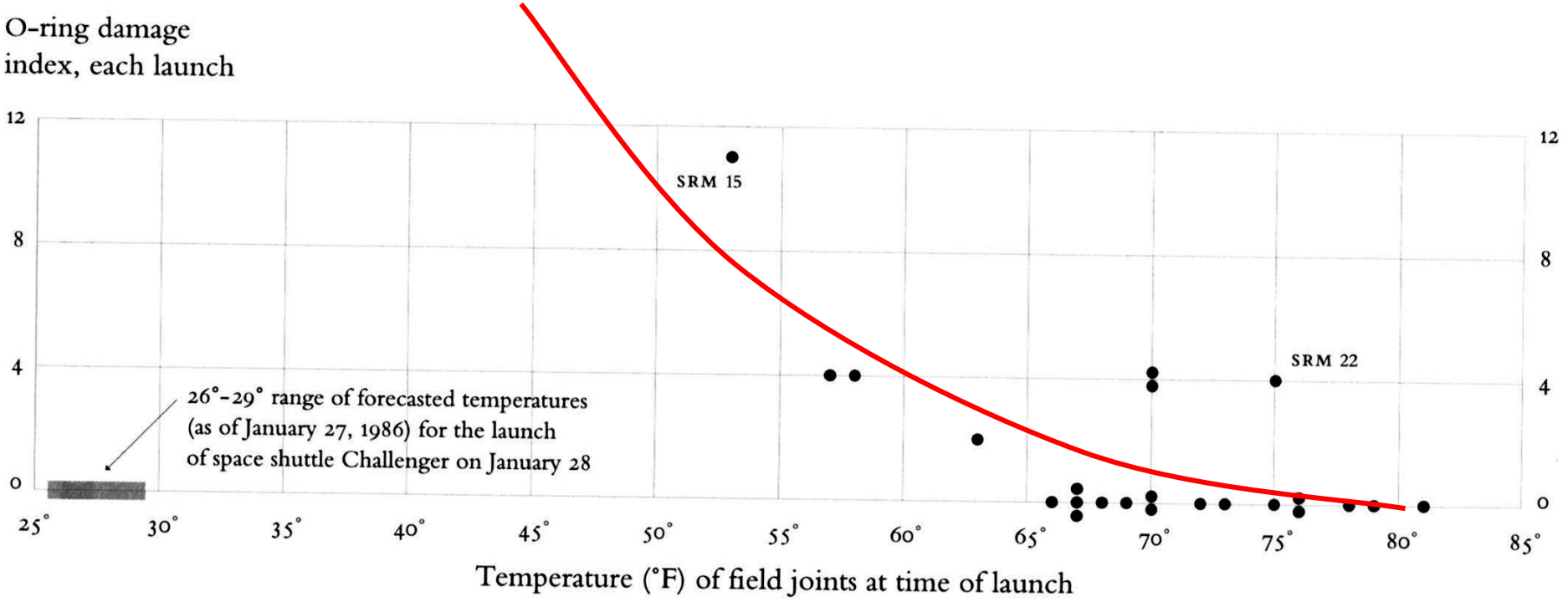


id Rocket Boosters

on
on
nce



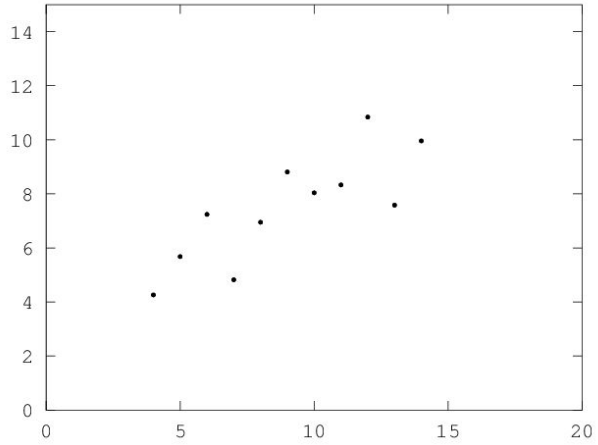
Morton Thoikol and Solid Rocket Boosters

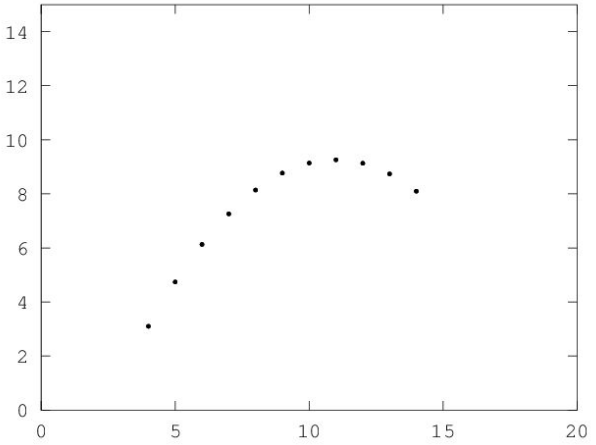


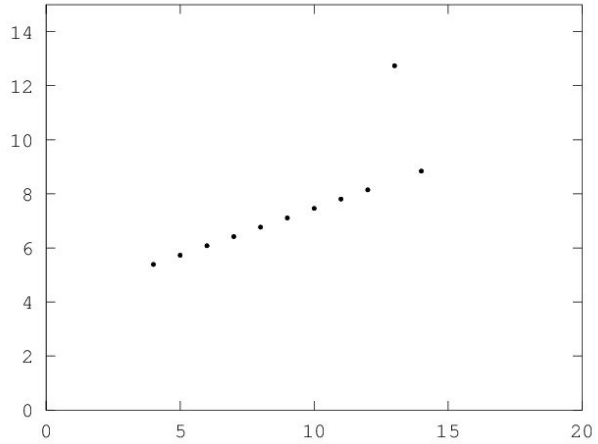
Anscombe's Quartet

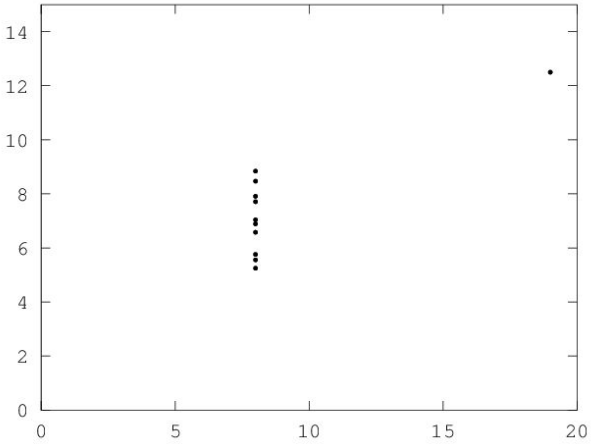
x1	y1	x2	y2	x3	y3	x4	y4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
5	4.26	5	3.10	5	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Anscombe's Quartet









Purposes of Visualization

- Supporting exploratory data analysis (exploratory)
- Explaining or supporting presentation (explanatory)

Visualization Software

Visualization Software

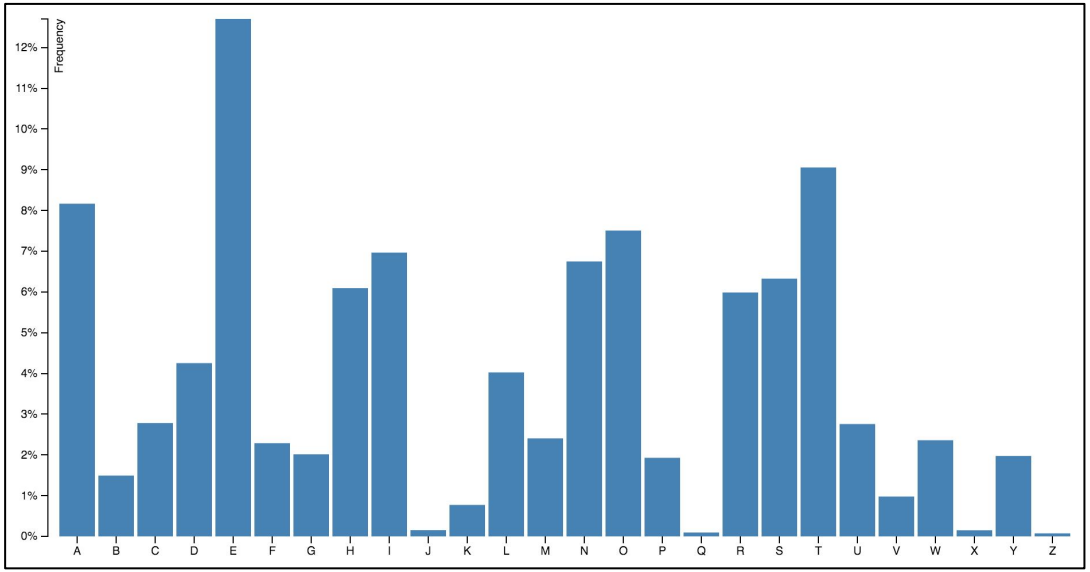
- MATLAB (Octave/gnuplot)
- Python
  - matplotlib
  - Seaborn
  - ggplot
- Python and Web
  - Bokeh
  - plotly
- Web
  - D3
- On the extremes:
  - OpenGL / WebGL
  - Tableau

D3.js

- Data Driven Documents
- Grew out of Mike Bostock's graduate work
- Lots of examples
- YouTube tutorial (from Galvanize!)

D3 Hello World

- Let’s Make a Bar Chart



D3 Bar Chart Example

```
<!DOCTYPE html>
<meta charset="utf-8">
<style>
.bar {
  fill: steelblue;
}
.bar:hover {
  fill: brown;
}
.axis {
  font: 10px sans-serif;
}
.axis path,
.axis line {
  fill: none;
  stroke: #000;
  shape-rendering: crispEdges;
}
.x.axis path {
  display: none;
}
</style>
<body>
<script src="//d3js.org/d3.v3.min.js"></script>
<script>
var margin = {top: 20, right: 20, bottom: 30, left: 40},
    width = 960 - margin.left - margin.right,
    height = 500 - margin.top - margin.bottom;
var x = d3.scale.ordinal()
    .rangeRoundBands([0, width], .1);
var y = d3.scale.linear()
    .range([height, 0]);
var xAxis = d3.svg.axis()
    .scale(x)
    .orient("bottom");
var yAxis = d3.svg.axis()
    .scale(y)
    .orient("left")
    .ticks(10, "%");
var svg = d3.select("body").append("svg")
    .attr("width", width + margin.left + margin.right)
    .attr("height", height + margin.top + margin.bottom)
  .append("g")
    .attr("transform", "translate(" + margin.left + "," + margin.top + ")");

d3.tsv("data.tsv", type, 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; })]);
  svg.append("g")
      .attr("class", "x axis")
      .attr("transform", "translate(0," + height + ")")
      .call(xAxis);
  svg.append("g")
      .attr("class", "y axis")
      .call(yAxis)
    .append("text")
      .attr("transform", "rotate(-90)")
      .attr("y", 6)
      .attr("dy", ".71em")
      .style("text-anchor", "end")
      .text("Frequency");
  svg.selectAll(".bar")
      .data(data)
    .enter().append("rect")
      .attr("class", "bar")
      .attr("x", function(d) { return x(d.letter); })
      .attr("width", x.rangeBand())
      .attr("y", function(d) { return y(d.frequency); })
      .attr("height", function(d) { return height - y(d.frequency); });
});

function type(d) {
  d.frequency = +d.frequency;
  return d;
}

</script>
```

D3, HTML, CSS, SVG

```
<!DOCTYPE html>
<meta charset="utf-8">
<style>
.bar {
  fill: steelblue;
}
.bar:hover {
  fill: brown;
}
.axis {
  font: 10px sans-serif;
}
.axis path,
.axis line {
  fill: none;
  stroke: #000;
  shape-rendering: crispEdges;
}
.x.axis path {
  display: none;
}
</style>
<body>
<script src="//d3js.org/d3.v3.min.js"></script>
<script>
var margin = {top: 20, right: 20, bottom: 30, left: 40},
    width = 960 - margin.left - margin.right,
    height = 500 - margin.top - margin.bottom;
var x = d3.scale.ordinal()
    .rangeRoundBands([0, width], .1);
var y = d3.scale.linear()
    .range([height, 0]);
var xAxis = d3.svg.axis()
    .scale(x)
    .orient("bottom");
var yAxis = d3.svg.axis()
    .scale(y)
    .orient("left")
    .ticks(10, "%");
var svg = d3.select("body").append("svg")
    .attr("width", width + margin.left + margin.right)
    .attr("height", height + margin.top + margin.bottom)
  .append("g")
    .attr("transform", "translate(" + margin.left + "," + margin.top + ")");

d3.tsv("data.tsv", type, 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; })]);
  svg.append("g")
      attr("class", "x axis")
      attr("transform", "translate(0," + height + ")")
      call(xAxis);
  svg.append("g")
      attr("class", "y axis")
      call(yAxis)
    .append("text")
      attr("transform", "rotate(-90)")
      attr("y", 6)
      attr("dy", ".71em")
      style("text-anchor", "end")
      text("Frequency");
  svg.selectAll(".bar")
      data(data)
    .enter().append("rect")
      attr("class", "bar")
      attr("x", function(d) { return x(d.letter); })
      attr("width", x.rangeBand())
      attr("y", function(d) { return y(d.frequency); })
      attr("height", function(d) { return height - y(d.frequency); });
});

function type(d) {
  d.frequency = +d.frequency;
  return d;
}
</script>
```

Theory and Vocabulary

Data Taxonomy and Visual Encodings

Relational Data Model

- *Database* is a collection of *tables*
- *Tables* are a list of *records*
- *Records* are datapoints giving *values* for *attributes*

Name	*Color*	*Mass*	*Kingdom*	*Taste Rating*	...
Apple	Red	400g	Plantae	Okay	
Banana	Yellow	800g	Plantae	Good	
Morel	Greenish	350g	Fungi	Bad	
Cow	Black/White	250,000g	Anamalia	Excellent	
...					...

Attribute Domain Taxonomy

- Nominal (=, ≠)
  - Types and categories (mathematical *set*)

Attribute Domain Taxonomy

- Ordinal (=, ≠, ≤)
  - Has an order (mathematical *set* with *order relation*)

Attribute Domain Taxonomy

- Interval (=, ≠, ≤, +, -)
  - Has a meaningful difference between values (mathematical *group*)

Attribute Domain Taxonomy

- Ratio (=, ≠, ≤, +, -, x, ÷)
  - Has a meaningful one and zero point and ratio between values (mathematical *field*)

Attribute Domain Taxonomy

- Nominal (=, ≠)
  - Types and categories (mathematical *set*)
- Ordinal (=, ≠, ≤)
  - Has an order (mathematical *set* with *order relation*)
  - E.g: Rankings, grades
- Quantitative
  - Interval (=, ≠, ≤, +, -)
    - Has a meaningful difference between values (mathematical *group*)
    - E.g: Dates, location, geometric points, temperature (C and F)
  - Ratio (=, ≠, ≤, +, -, x, ÷)
    - Has a meaningful one and zero point and ratio between values (mathematical *field*)
    - E.g: Distance, mass, temperature (K), time, counts

Attribute Domain Taxonomy

- Nominal (=, ≠)
  - Types and categories (mathematical *set*)
- Ordinal (=, ≠, ≤)
  - Has an order (mathematical *set* with *order relation*)
  - E.g: Rankings, grades
- Quantitative
  - Interval (=, ≠, ≤, +, -)
    - Has a meaningful difference between values (mathematical *group*)
    - E.g: Dates, location, geometric points, temperature (C and F)
  - Ratio (=, ≠, ≤, +, -, x, ÷)
    - Has a meaningful one and zero point and ratio between values (mathematical *field*)
    - E.g: Distance, mass, temperature (K), time, counts
- Topological
  - Connectivity, inclusion

Visual Encodings

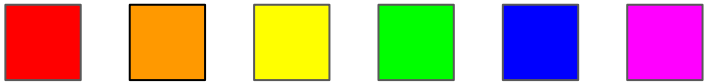
Visual Encodings of **Nominal** Data

Visual Encodings of **Nominal** Data

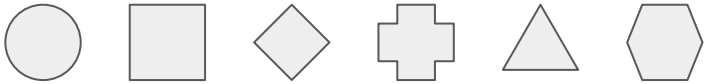
Text

A B C D E F

Color hue



Shape

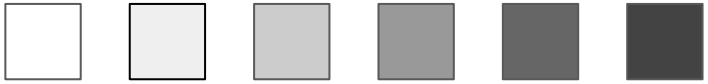


Visual Encodings of **Ordinal** Data

Visual Encodings of **Ordinal** Data

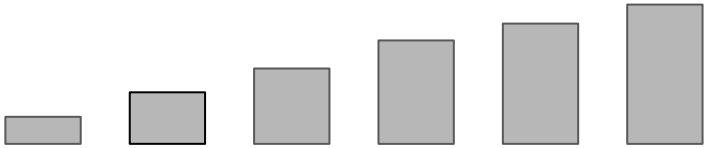
Color saturation

Color luminance



Area, Volume

Position



Time, Animation

Visual Encodings of **Quantitative** Data

Visual Encodings of **Quantitative** Data

Color saturation

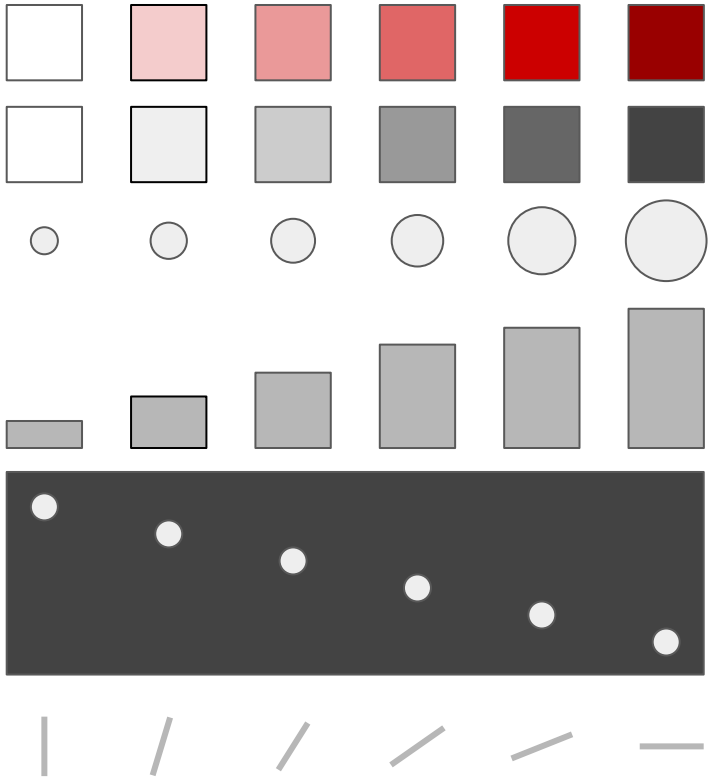
Color luminance

Position

Length, Area, Volume

Angle

Time, Animation



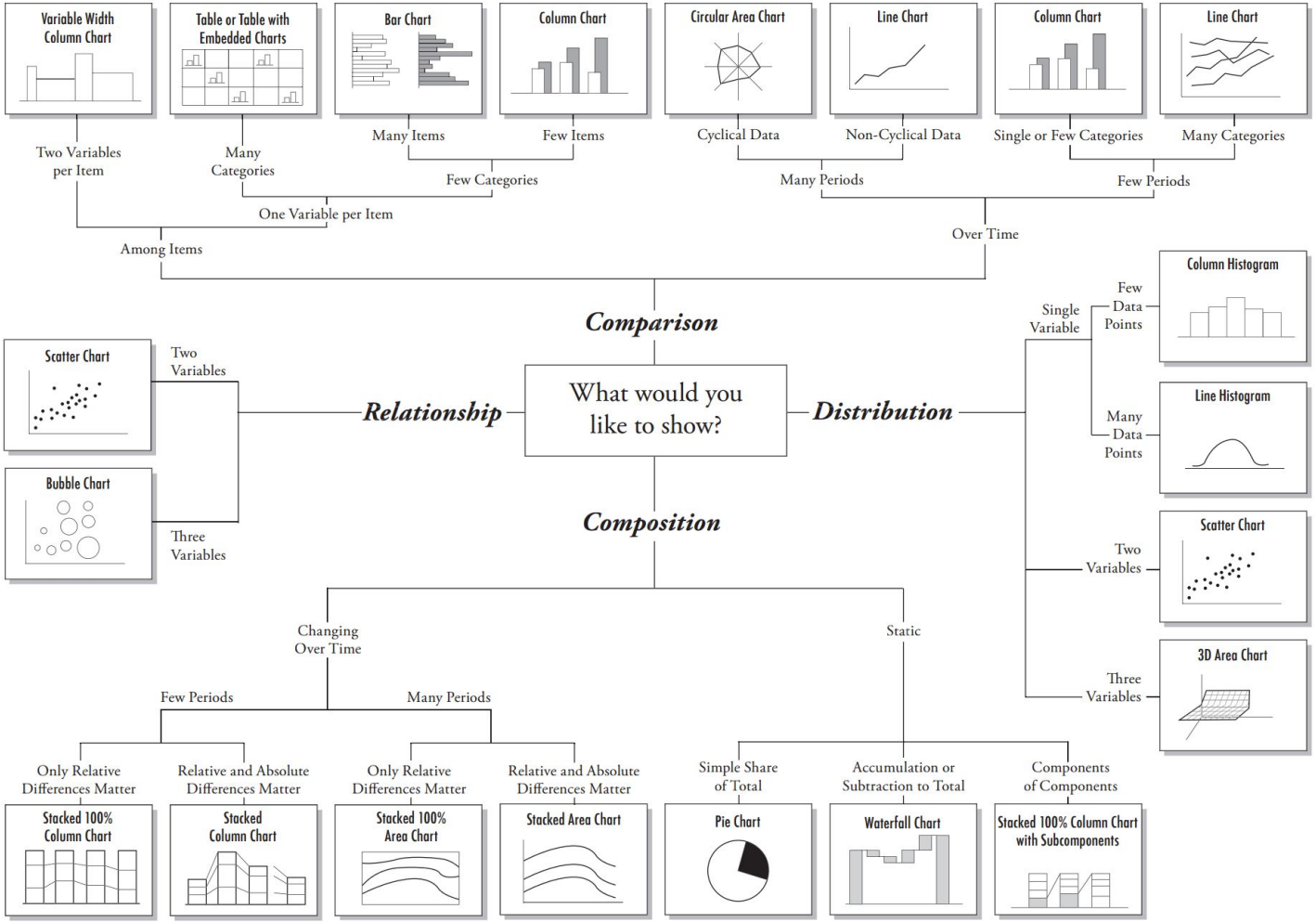
Charts

Chart: *noun*

1. A common pattern for combining visual encodings of data.
2. A visual sentence constructed with data encodings as words.

Chart Suggestions—A Thought-Starter

http://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf



Applying the Theory

- http://www.nytimes.com/interactive/2012/05/17/business/dealbook/how-the-facebook-offering-compares.html

Company	*IPO Year*	*IPO Value*	*1st Day Value*	*3 Year Value*
Apple	1980	$3.4B	$4.5B	$2.6B
Microsoft	1986	$1.1B	$1.5B	$4.1B
Google	2004	$28B	$33B	$140B

Making Good Visualizations

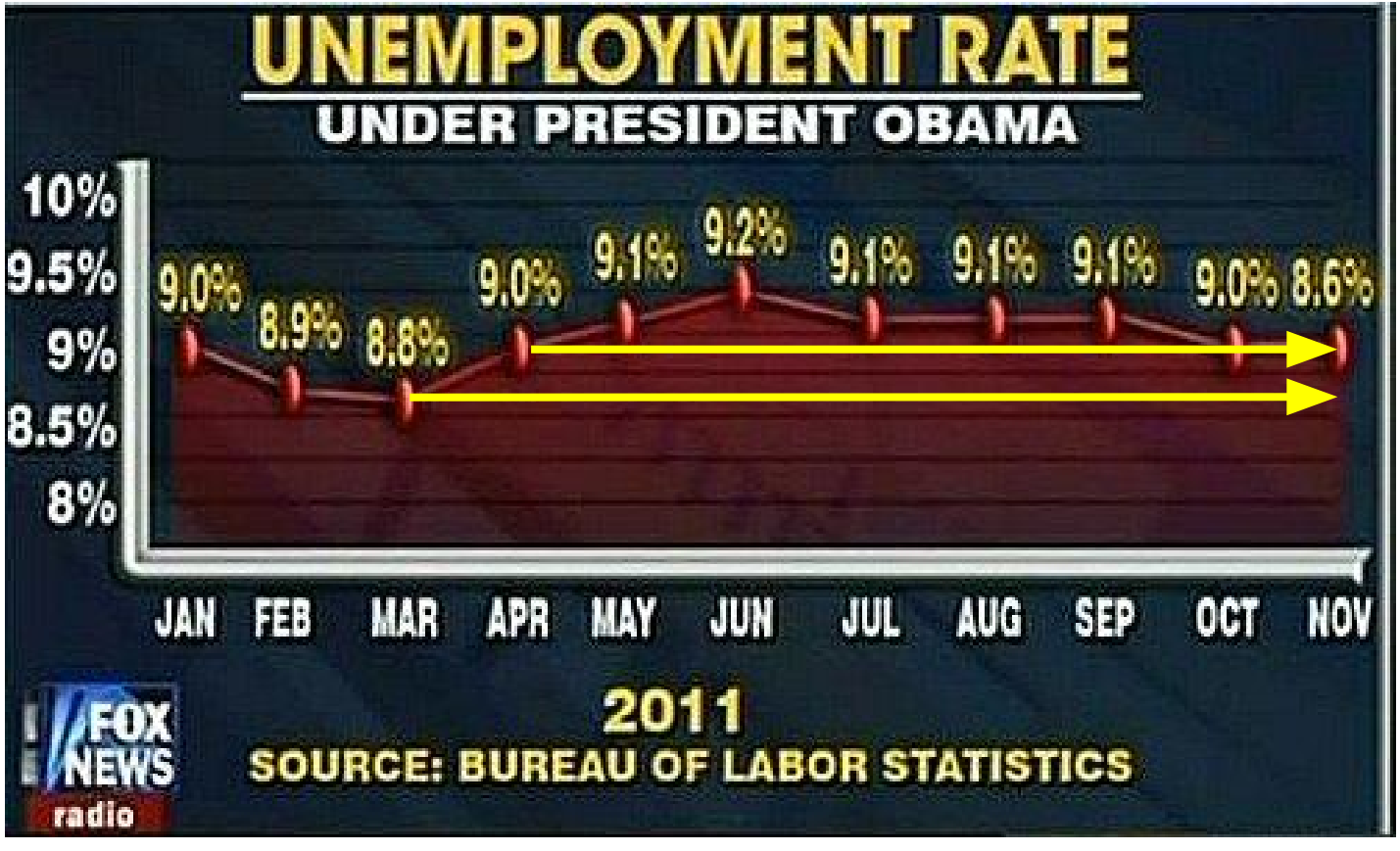
Not Making Bad Visualizations

Graphical Integrity

Graphical Integrity: Principles

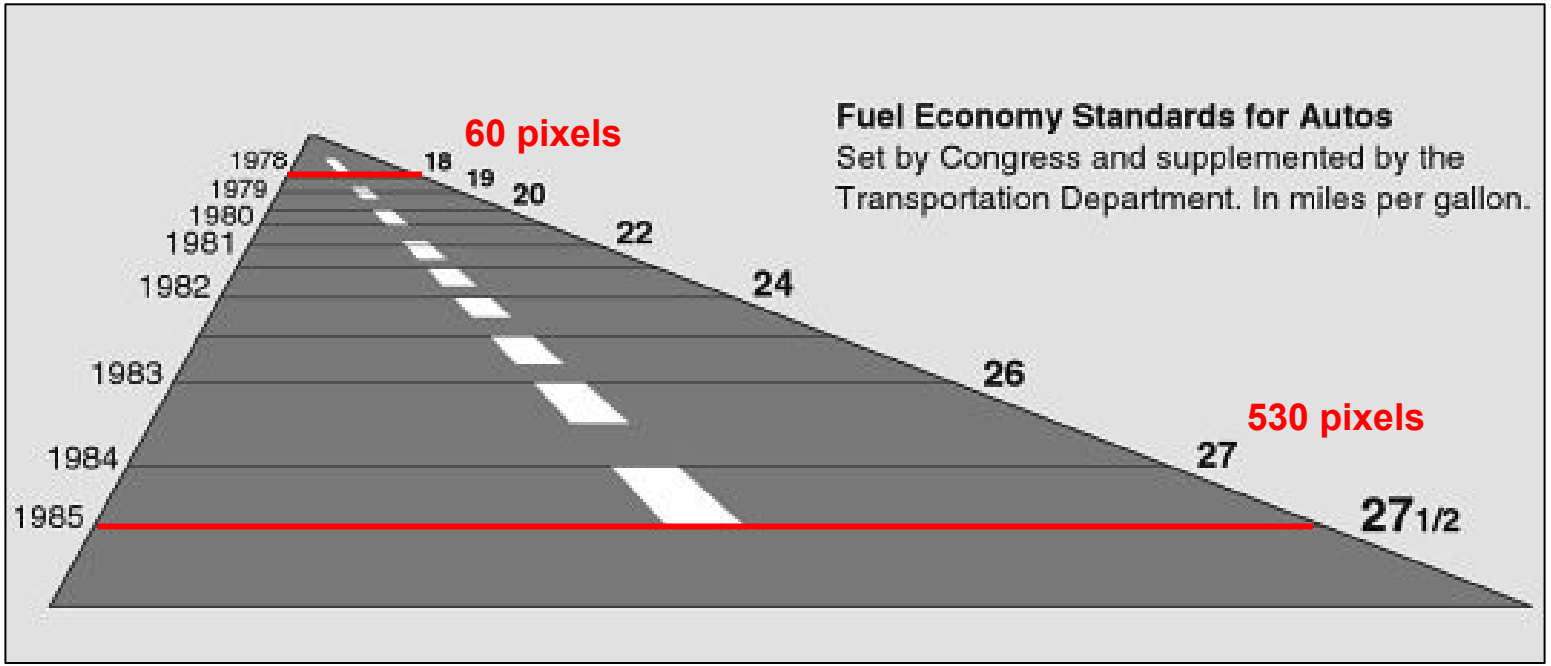
- Proportionality (between data and physical representation)
- Match dimensions (between data and physical representation)
- Provide important context

Proportionality



Proportionality

Graphical Ratio: 530px / 60px = 8.8
Data Ratio: 27.5mpg / 18mpg = 1.5



Proportionality

$$\text{Lie factor} = \frac{\text{size of effect shown in graph}}{\text{size of effect in the data}}$$

Proportionality

What about color?

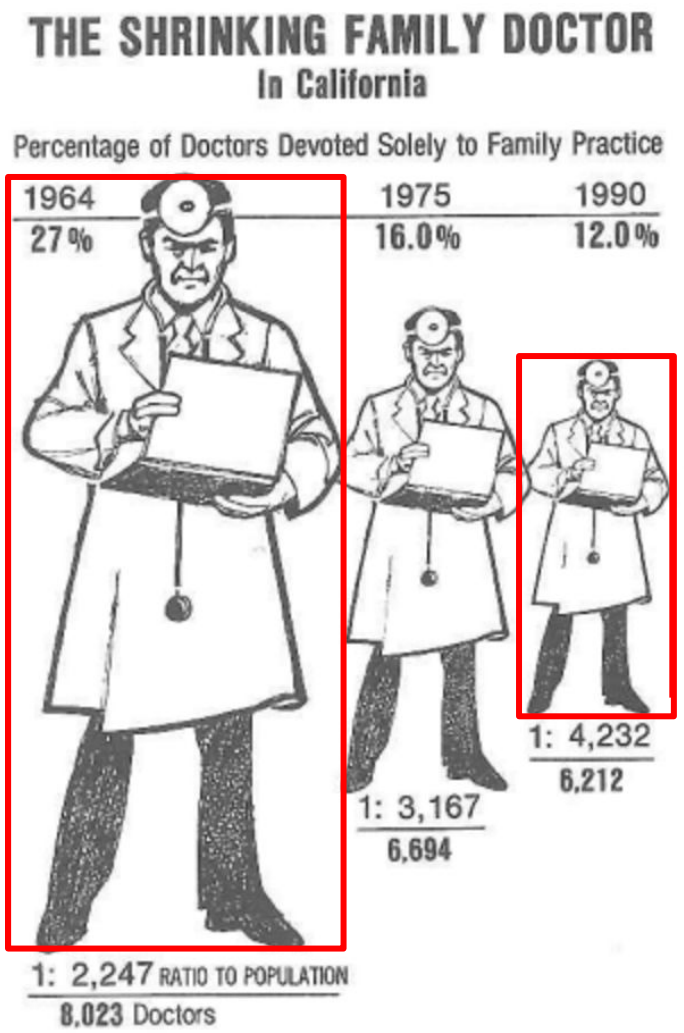
Match Dimensions

686px/315px = 2.2 = 27%/12%
… seems legit …?

192kpx/39kpx = 5 >> 2.2 = 27%/12%
… not so much.

686 pixels tall

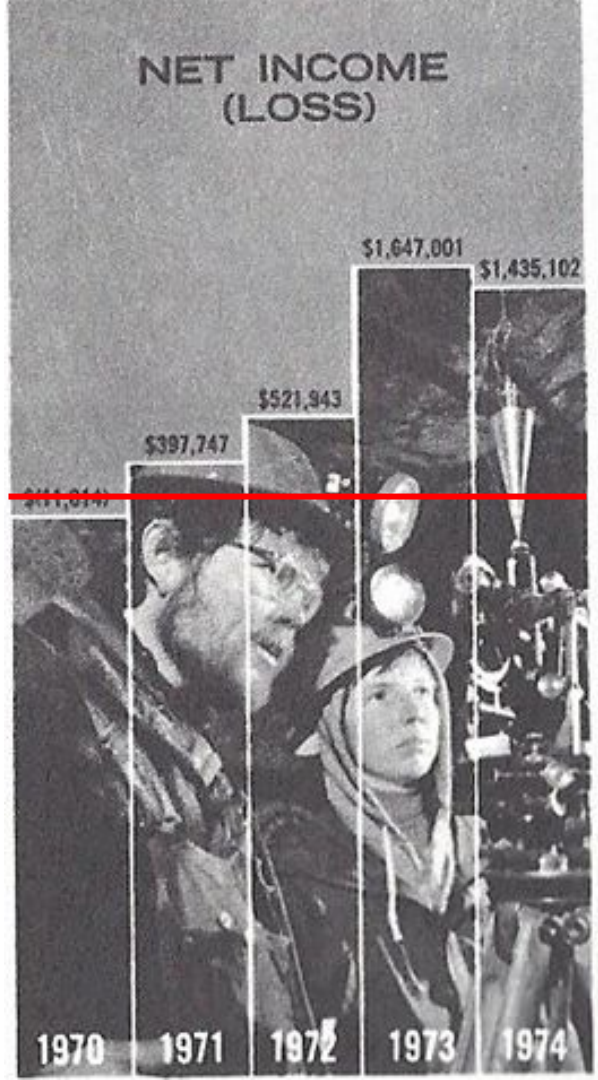
281x686 = 192k pixels



315 pixels tall

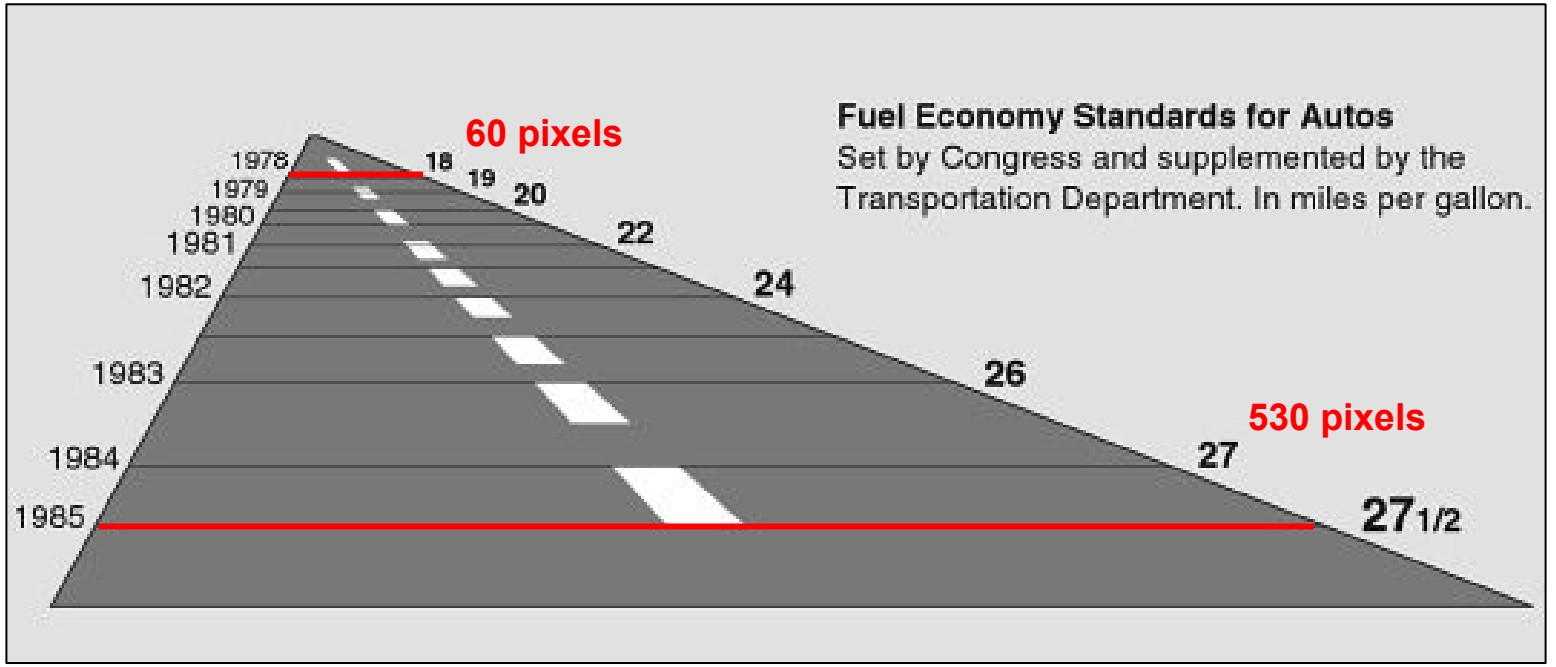
124x315 = 39k pixels

Providing Context



Graphical Integrity

Graphical Ratio: 530px / 60px = 8.8
Data Ratio: 27.5mpg / 18mpg = 1.5



Maximizing Impact

Data ink

- The ink (pixels) devoted to representing data

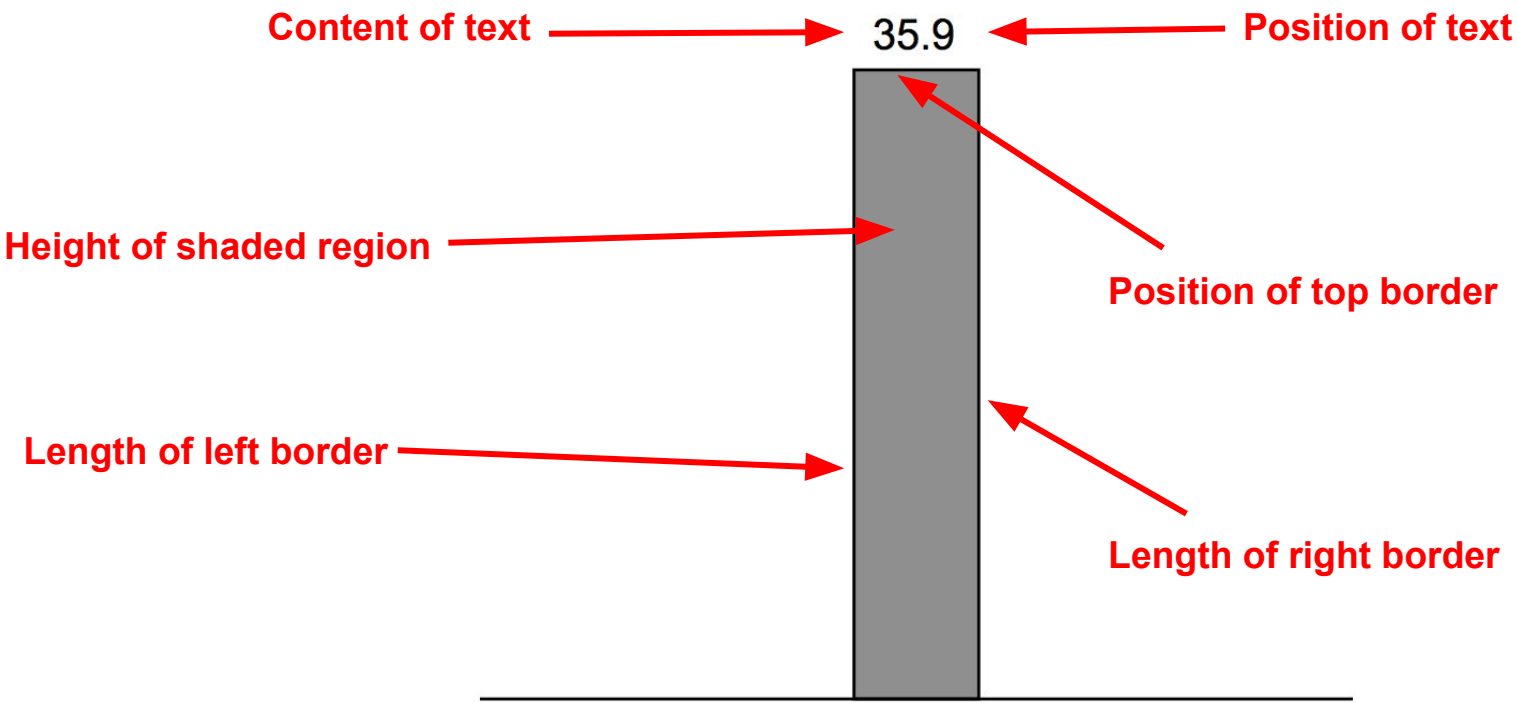
Data ink Ratio

$$\text{Data ink ratio} = \frac{\text{Data ink}}{\text{Total ink}}$$

Non-data ink

- Redundant data

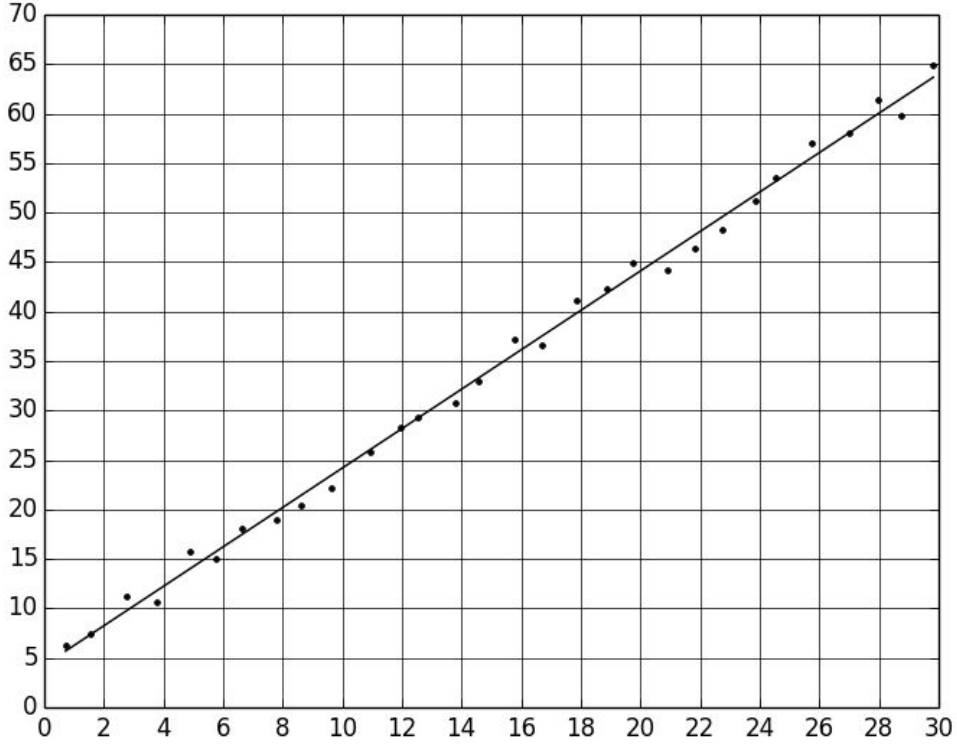
Redundant Data



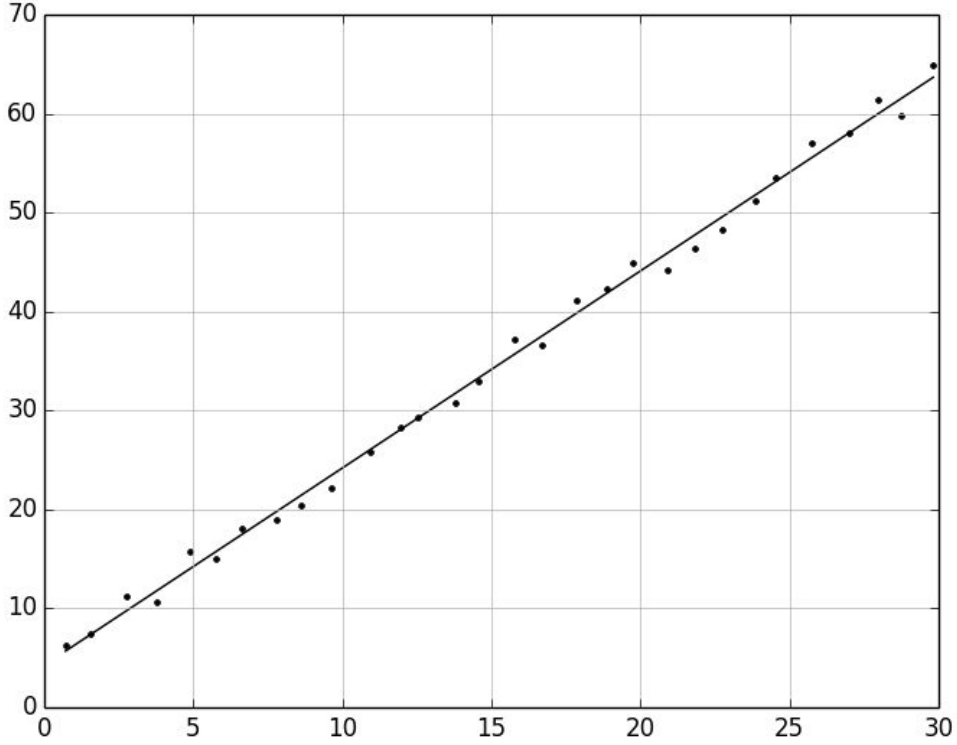
Non-data ink

- Redundant data
- Metadata

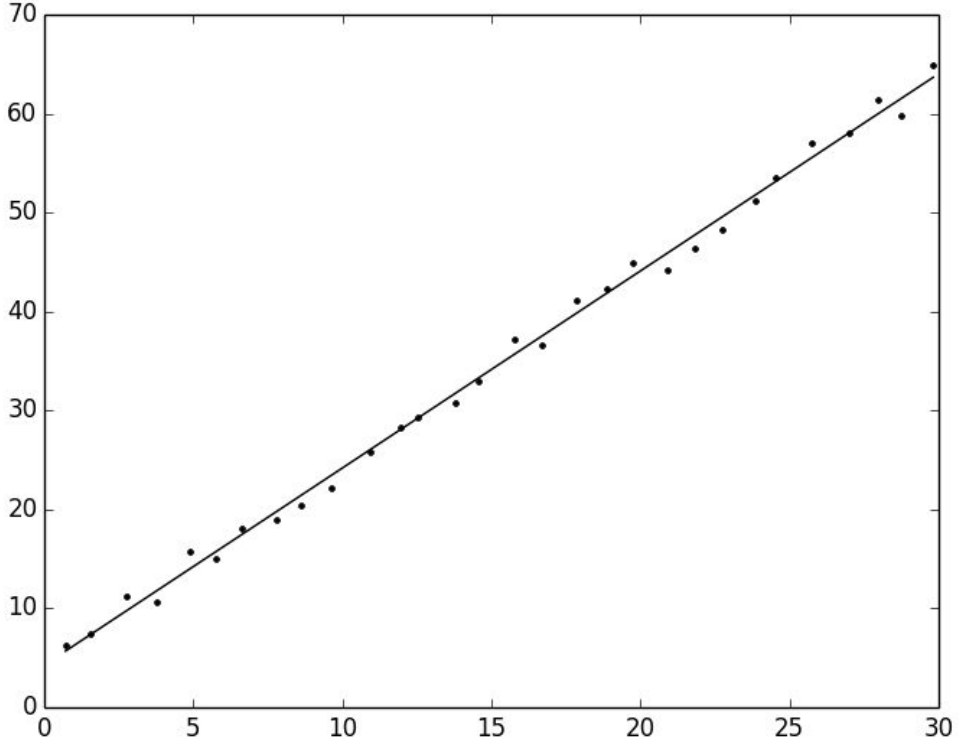
Metadata


70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30

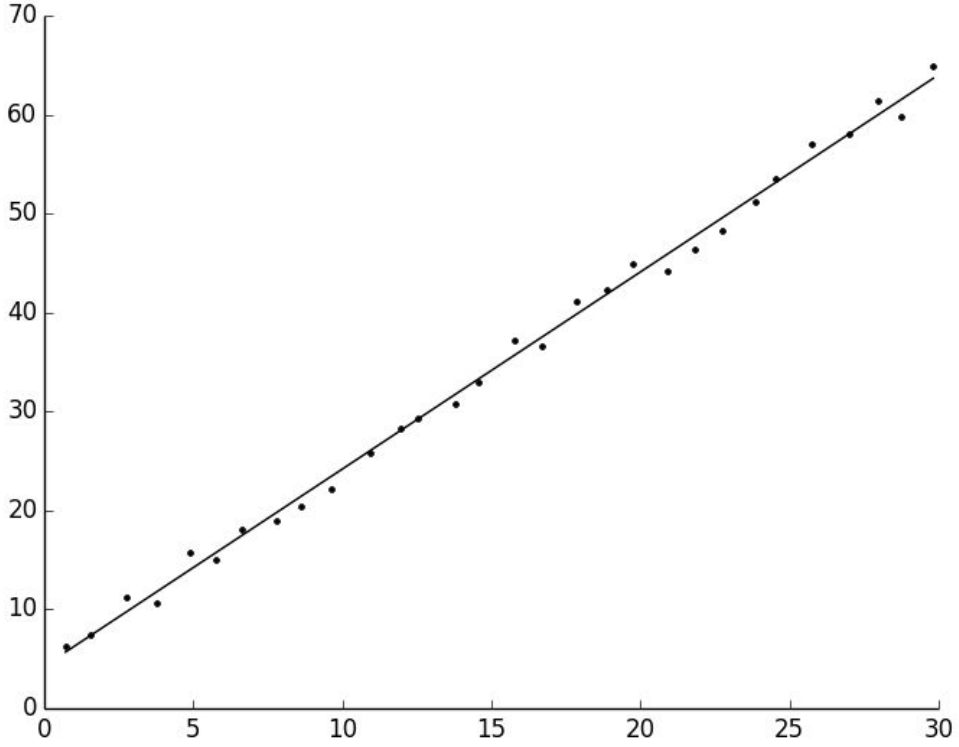
Metadata



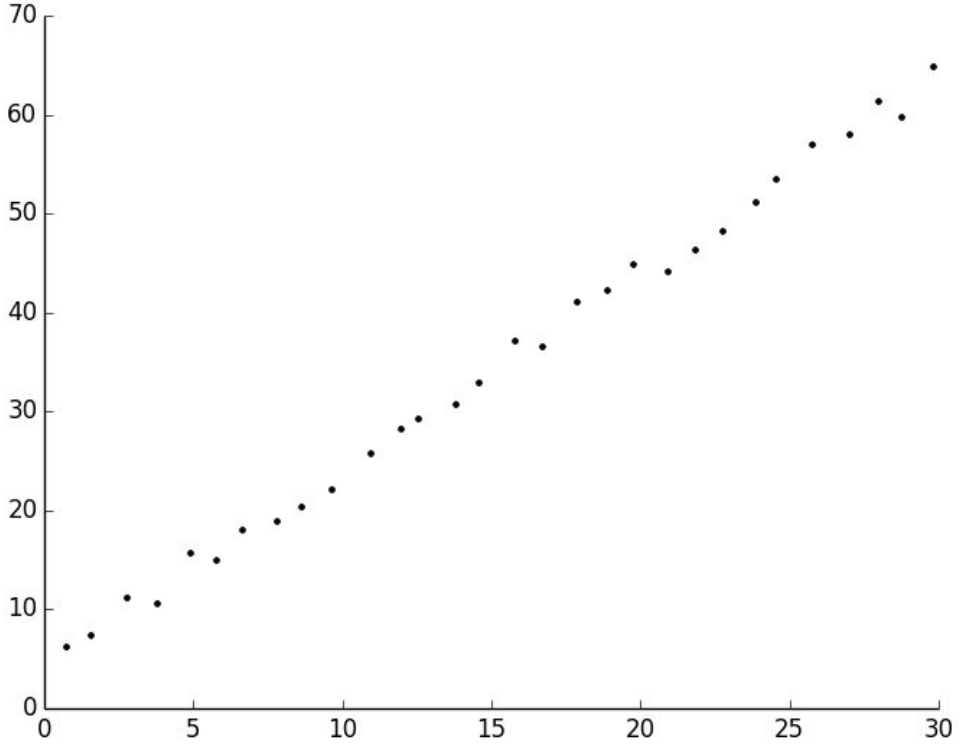
Metadata



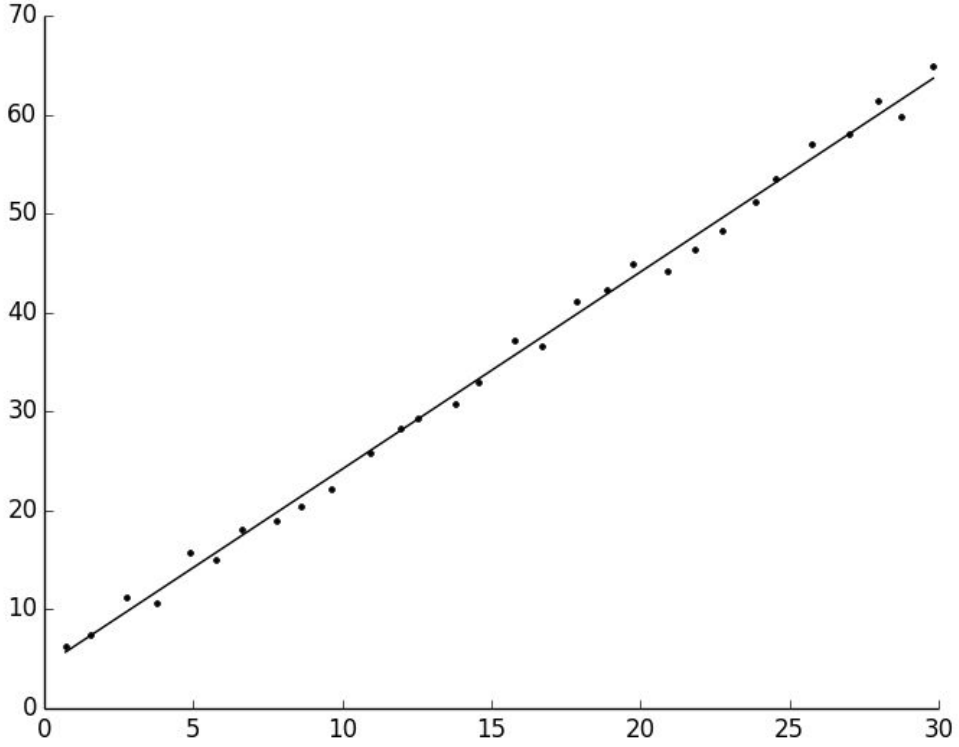
Metadata



Metadata



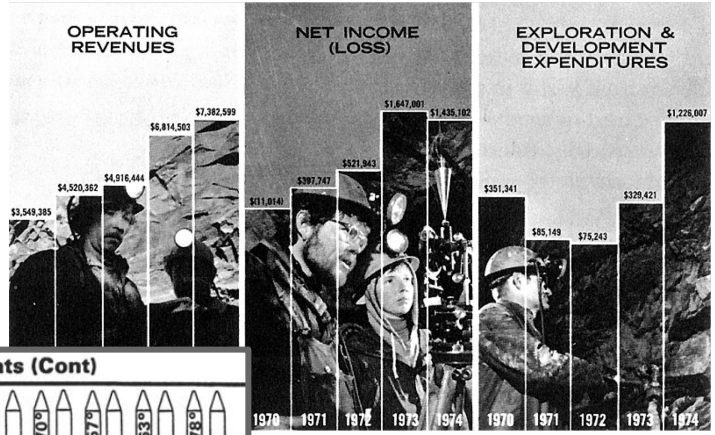
Metadata

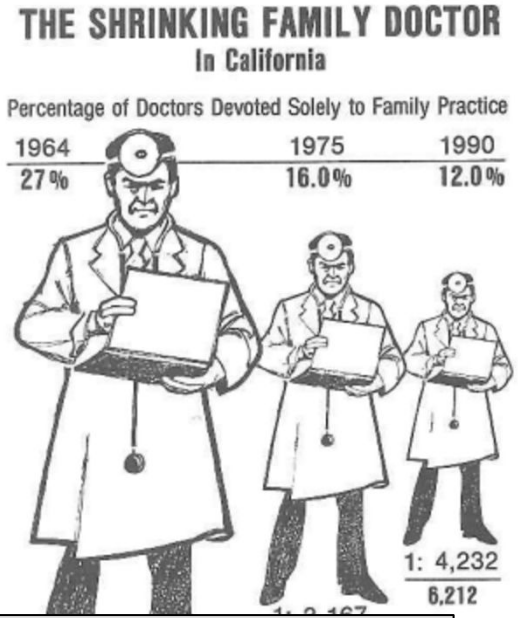


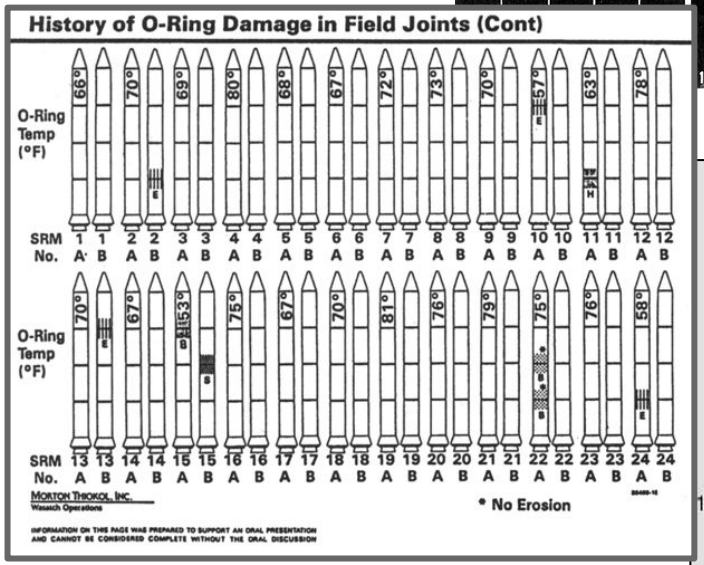
Improving charts

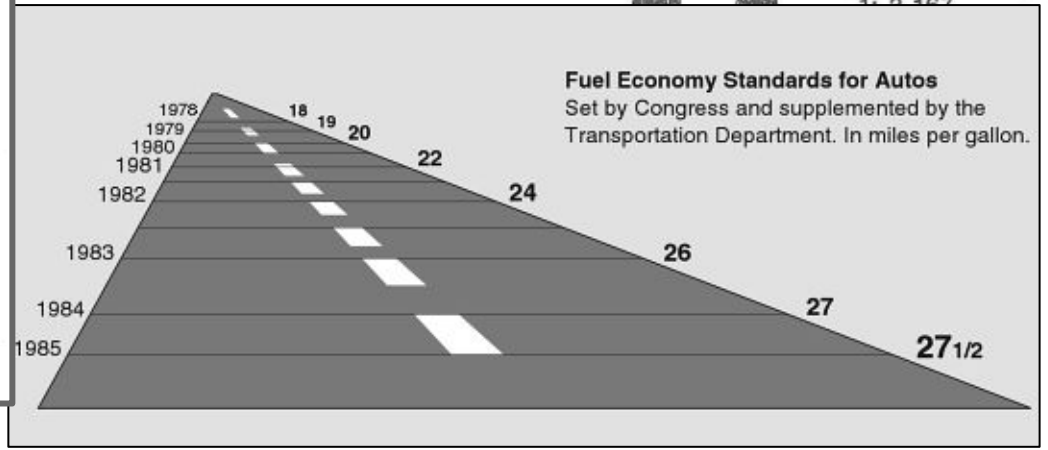
- Erase redundant data (within reason)
- Erase metadata (within reason)
- Iterate design
- Avoid chartjunk

Chartjunk

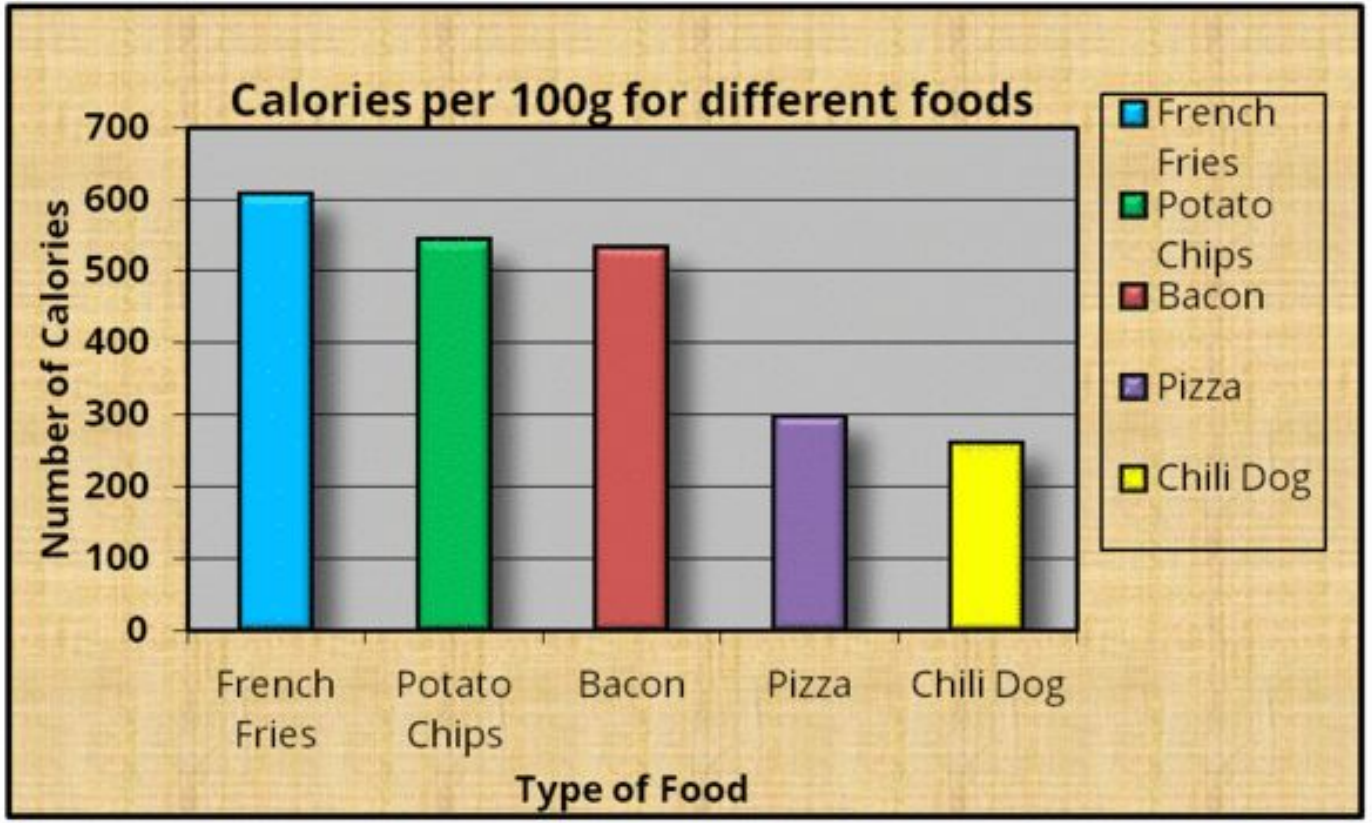

OPERATING REVENUES
$3,549,385
$4,520,362
$4,916,444
$6,814,503
$7,382,599
NET INCOME (LOSS)
$(11,614)
$397,747
$521,943
$1,647,001
$1,435,102
1970
1971
1972
1973
1974
EXPLORATION & DEVELOPMENT EXPENDITURES
$351,341
$85,149
$75,243
$329,421
$1,228,007
1970
1971
1972
1973
1974


THE SHRINKING FAMILY DOCTOR
In California
Percentage of Doctors Devoted Solely to Family Practice
1964
27%
1975
16.0%
1990
12.0%
1: 4,232
6,212

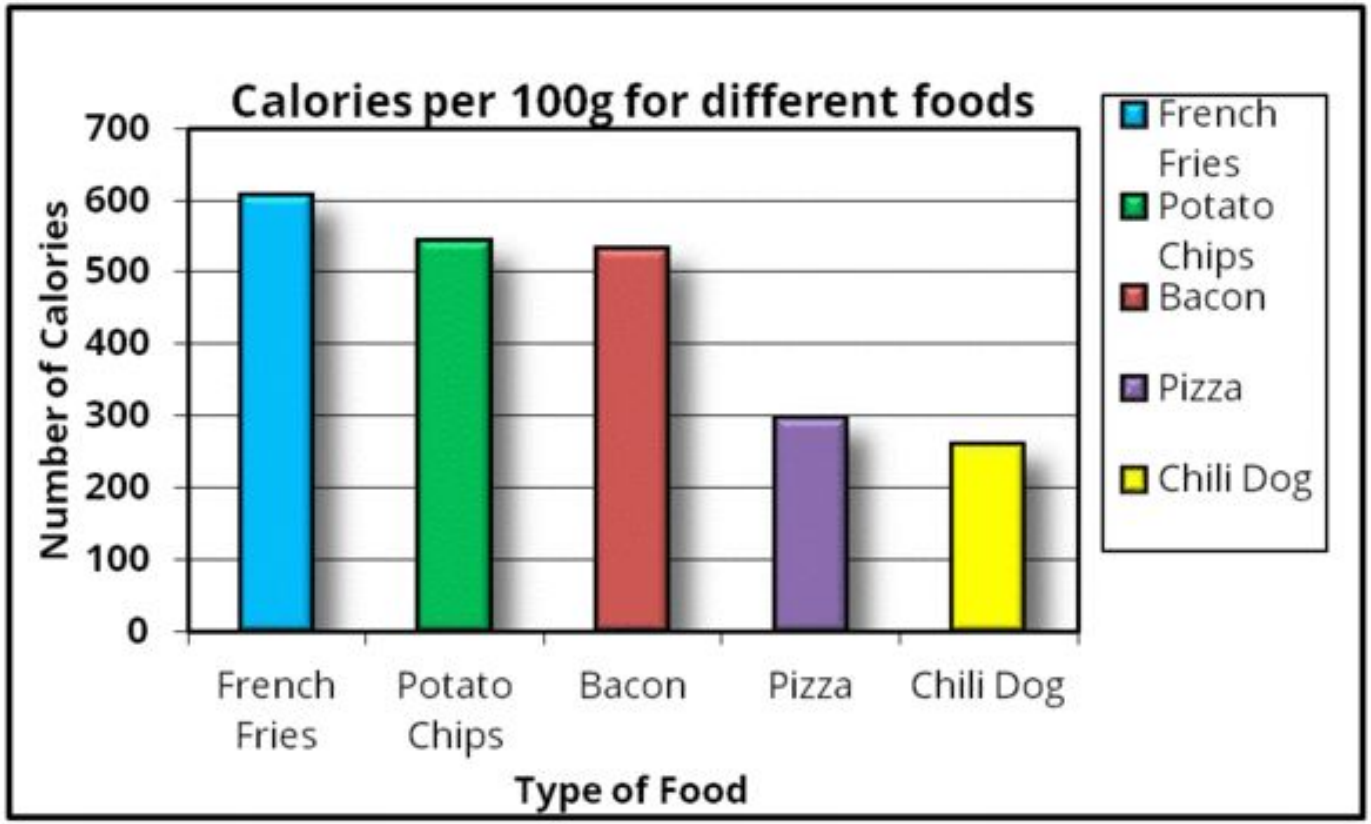

History of O-Ring Damage in Field Joints (Cont)
O-Ring Temp (°F)
66° 70° 69° 80° 68° 67° 72° 73° 70° 57° 63° 78°
SRM No.
1A 1B 2A 2B 3A 3B 4A 4B 5A 5B 6A 6B 7A 7B 8A 8B 9A 9B 10A 10B 11A 11B 12A 12B
O-Ring Temp (°F)
70° 67° 53° 75° 67° 70° 81° 76° 79° 75° 76° 58°
SRM No.
13A 13B 14A 14B 15A 15B 16A 16B 17A 17B 18A 18B 19A 19B 20A 20B 21A 21B 22A 22B 23A 23B 24A 24B
MORTON THIOKOL, INC.
Wasatch Operations
* No Erosion
INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION


Fuel Economy Standards for Autos
Set by Congress and supplemented by the Transportation Department. In miles per gallon.
1978
1979
1980
1981
1982
1983
1984
1985
18
19
20
22
24
26
27
27 1/2

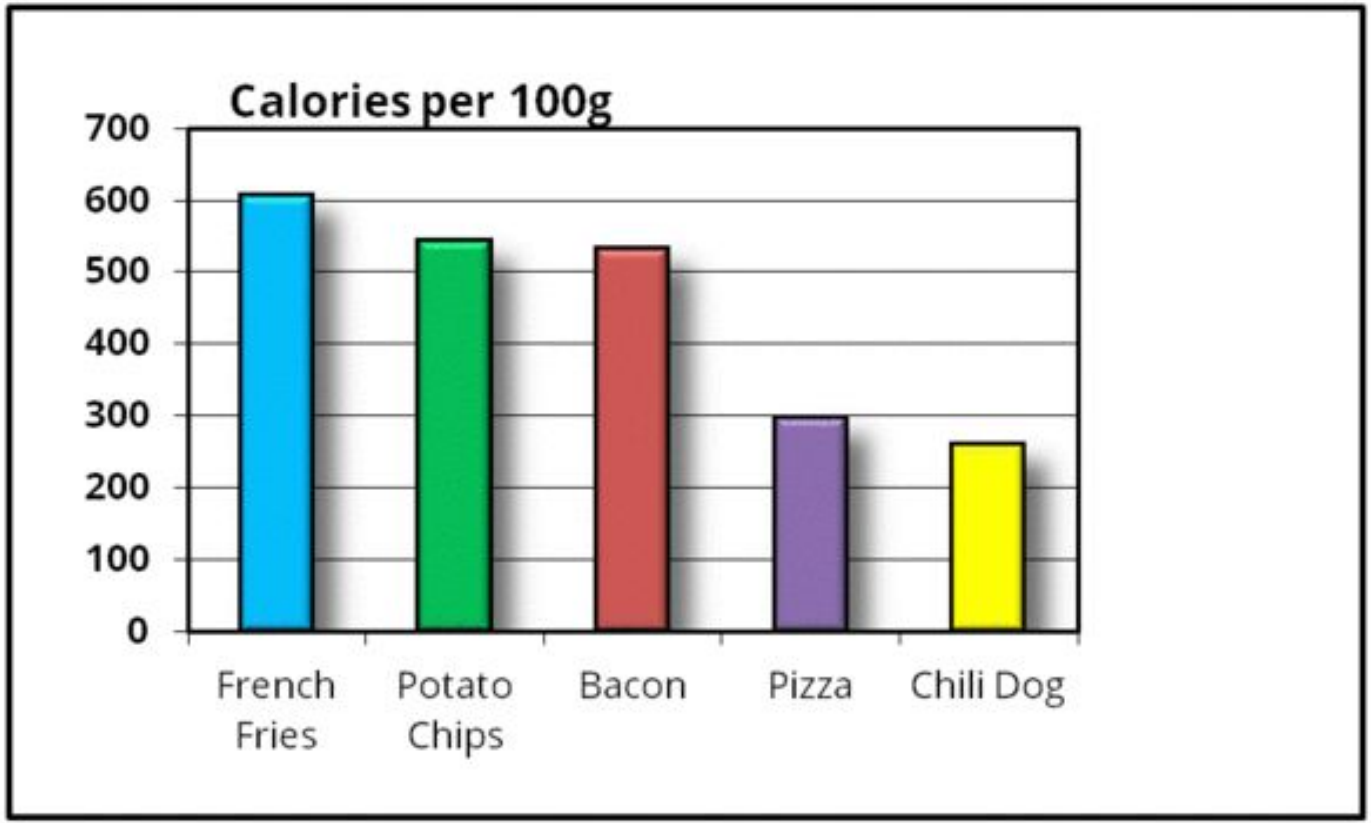
Story: Putting Bacon in Context



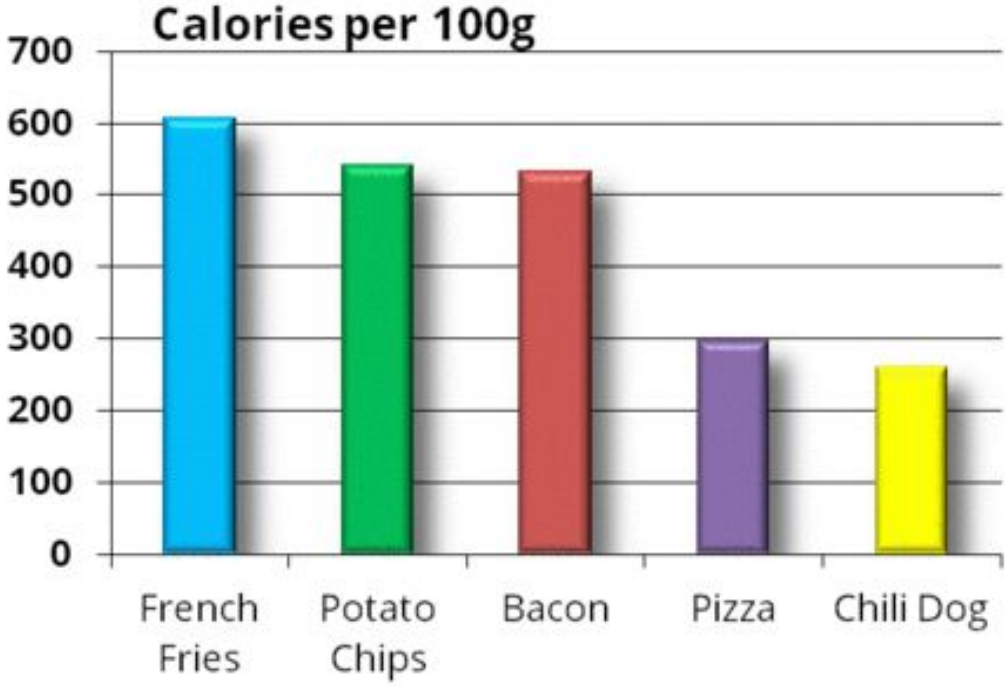
Remove backgrounds



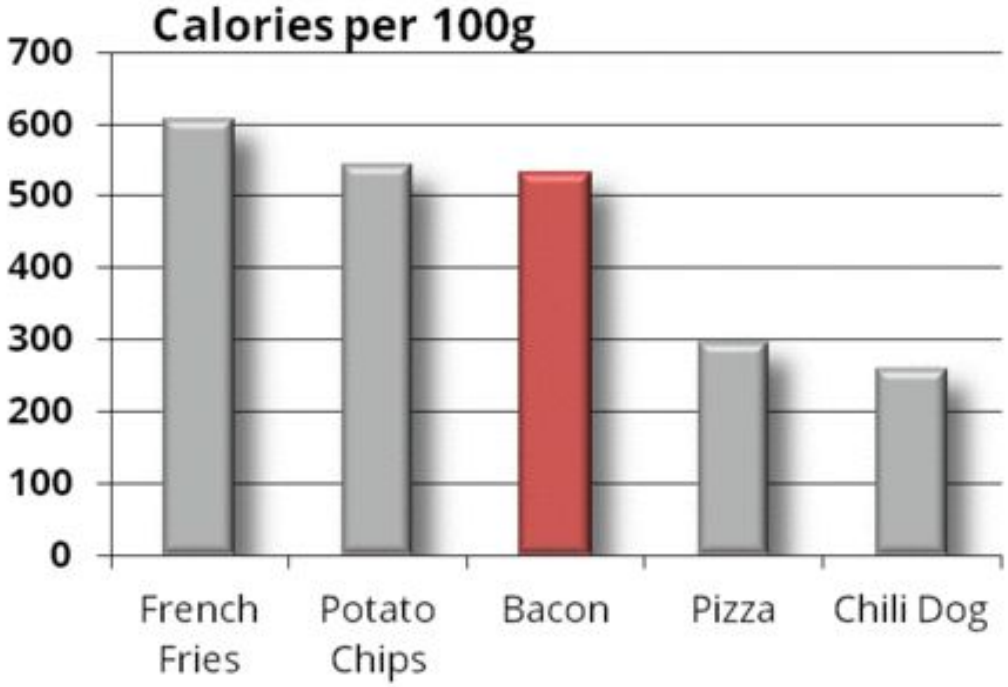
Remove redundant labels



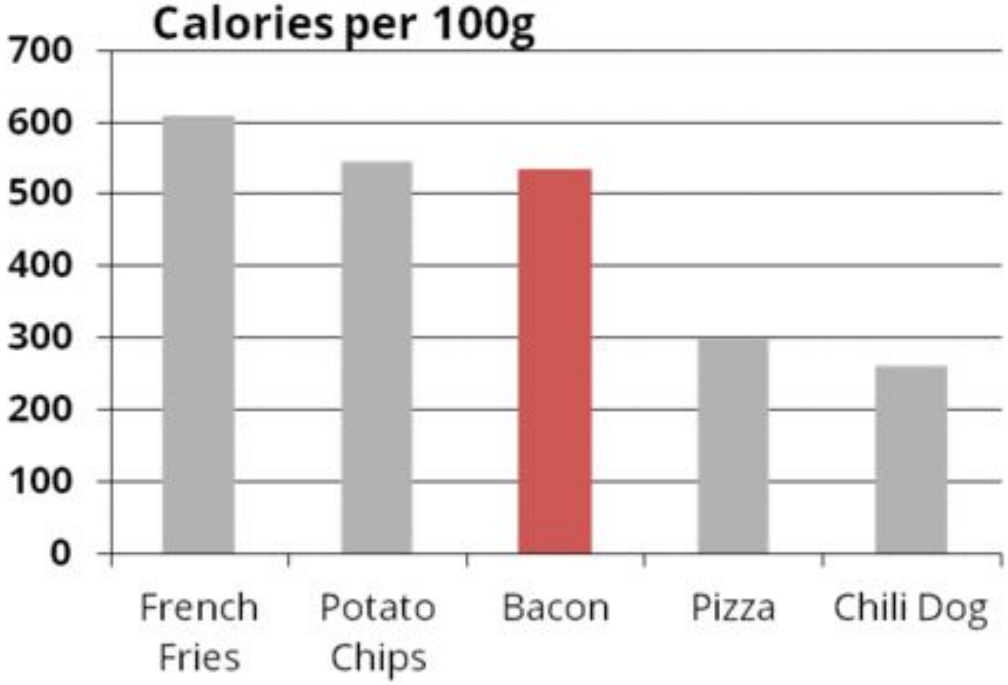
Remove borders



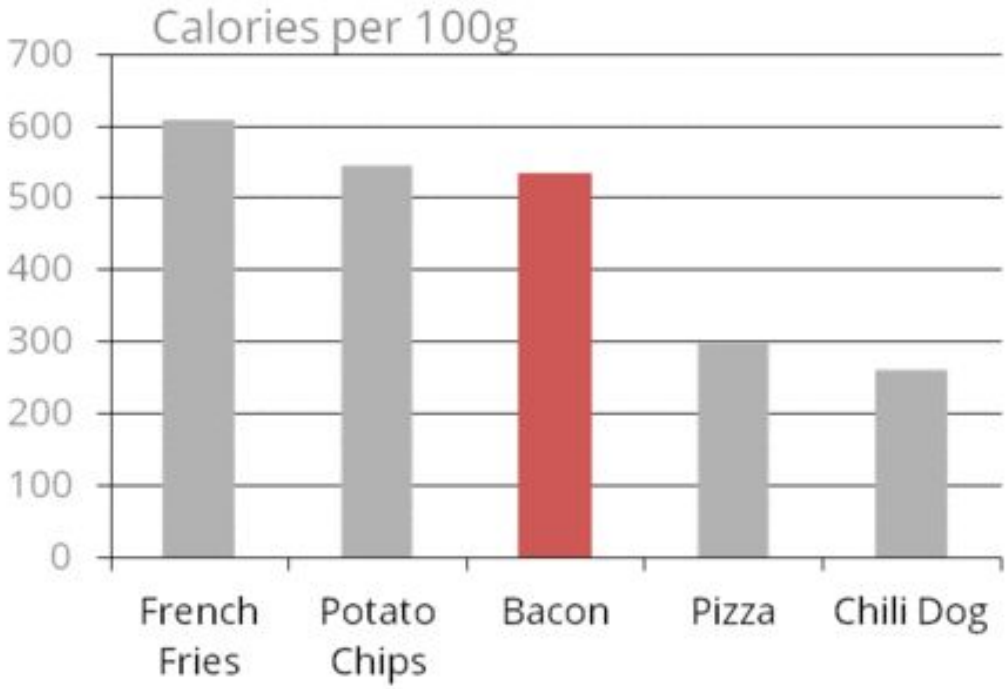
Reduce colors



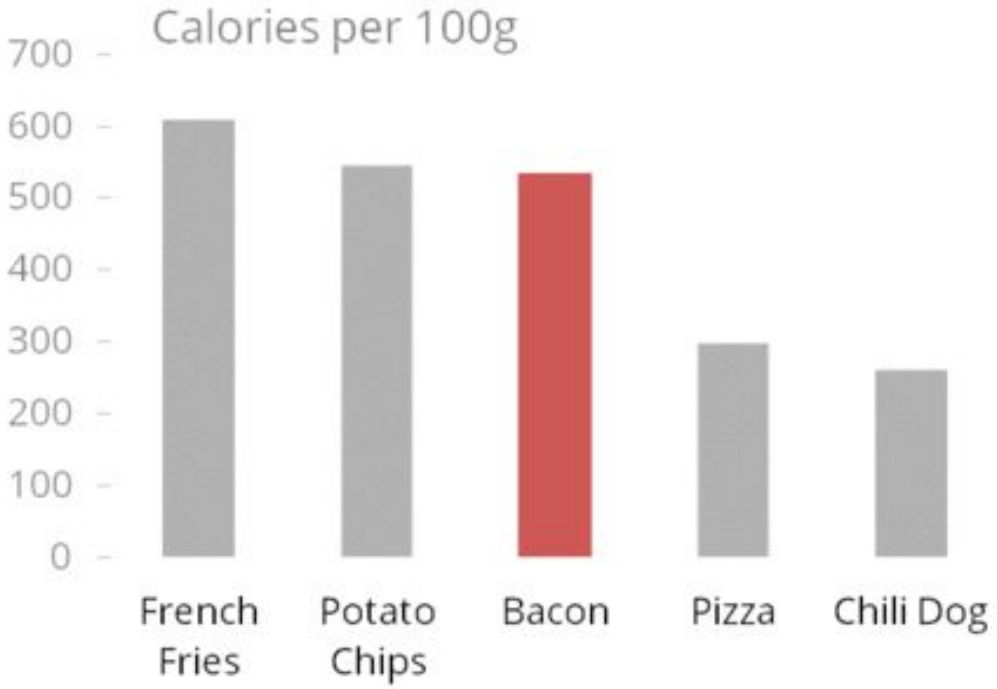
Remove special effects



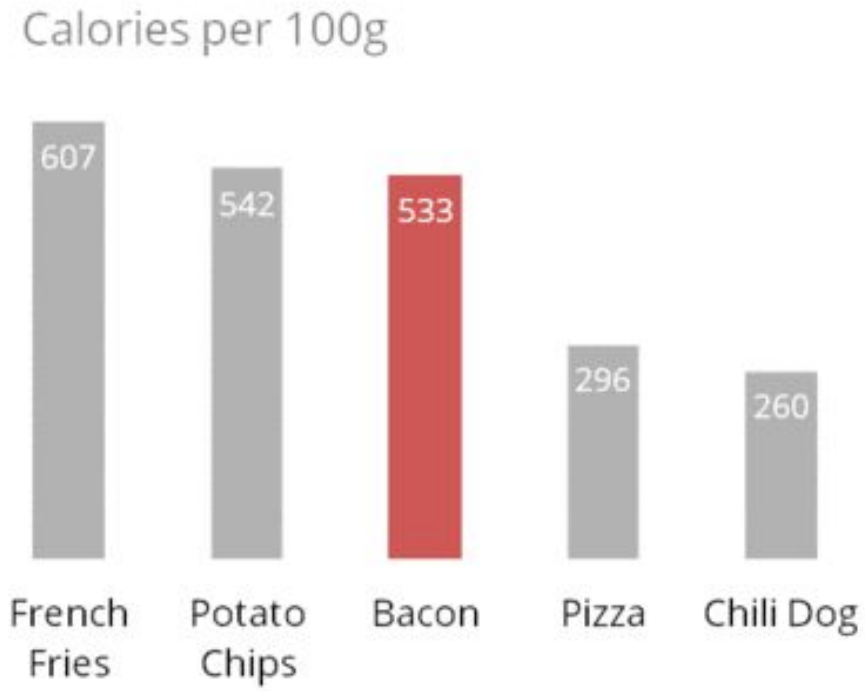
Lighten labels

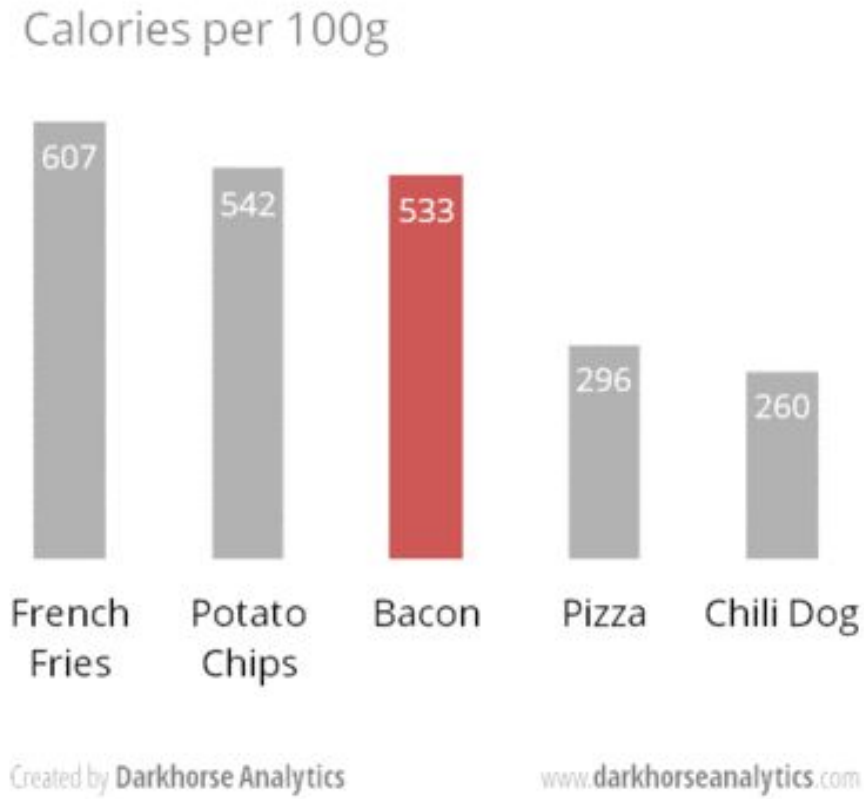


Or remove lines



Direct label




Calories per 100g
607
542
533
296
260
French Fries
Potato Chips
Bacon
Pizza
Chili Dog
Created by Darkhorse Analytics
www.darkhorseanalytics.com

Using Attentive Cognition

Attentive Processing

19047271168483165160806174557061387374154831111254680988089323343870208212744171310203693874289

Preattentive Processing

1904727116848316516

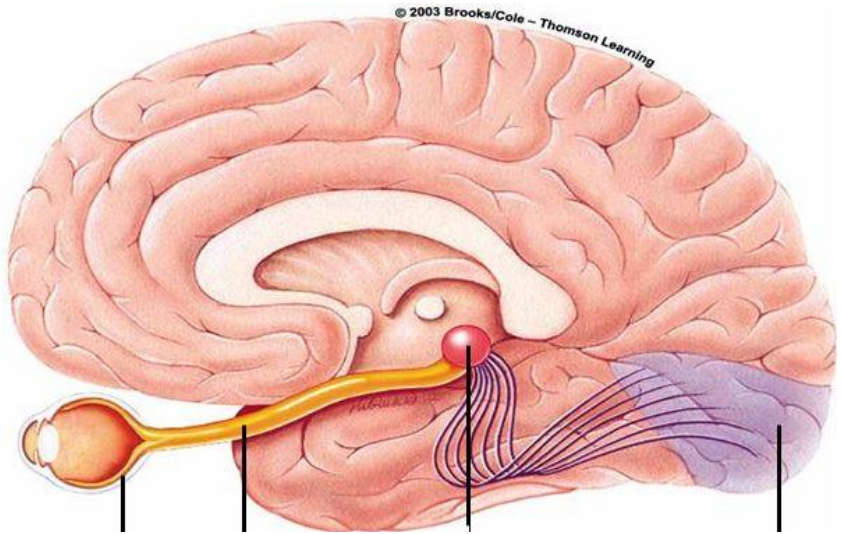
0806174557061387374

1548311125468098808

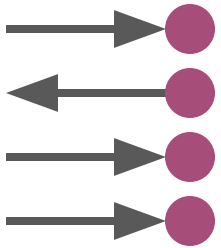
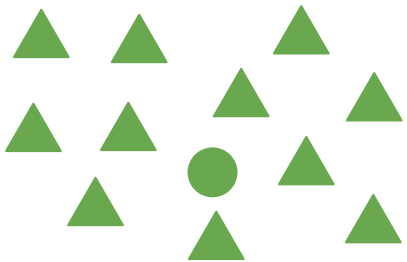
9323343870208212744

1713102036938742890

Visual Processing System



Preattentive Stimuli

Color 	Movement 
Form 	Position 

Preattentive Stimuli

- Interactive Applications
  - http://learnforeverlearn.com/preattentive/
- More on theory and types
  - http://www.perceptualedge.com/articles/ie/visual_perception.pdf
- Applications to computer vision
  - http://mplab.ucsd.edu/~marni/Igert/Malik_Perona_1990.pdf

Post-Preattentive Takeaways

- Draw the viewer’s attention
- Don’t distract the viewer