

Data Visualization

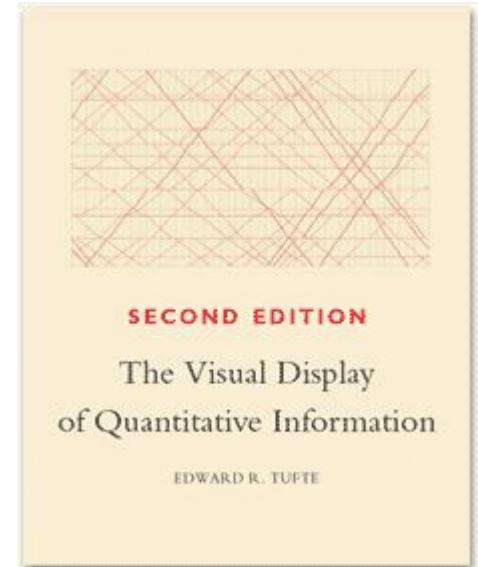
2017-06-01

Agenda

- Examples
- Software
- Theory
- Making good visualizations
- Assignment
- Preattentive cognition
- Using color

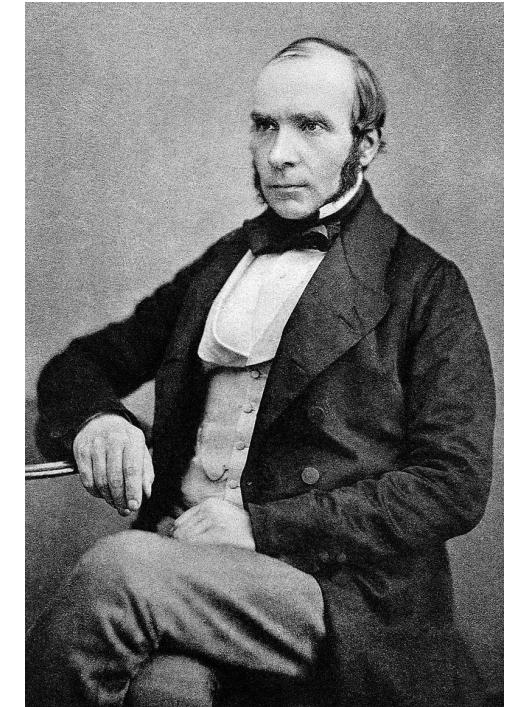
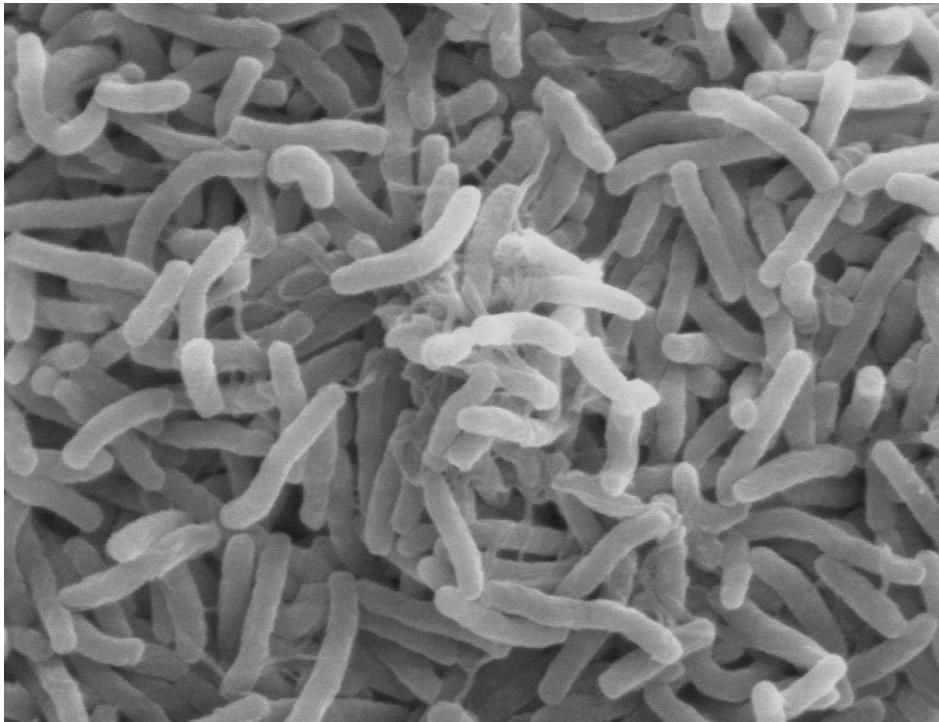
See Also

The Visual Display of Quantitative Information by Edward Tufte



Examples

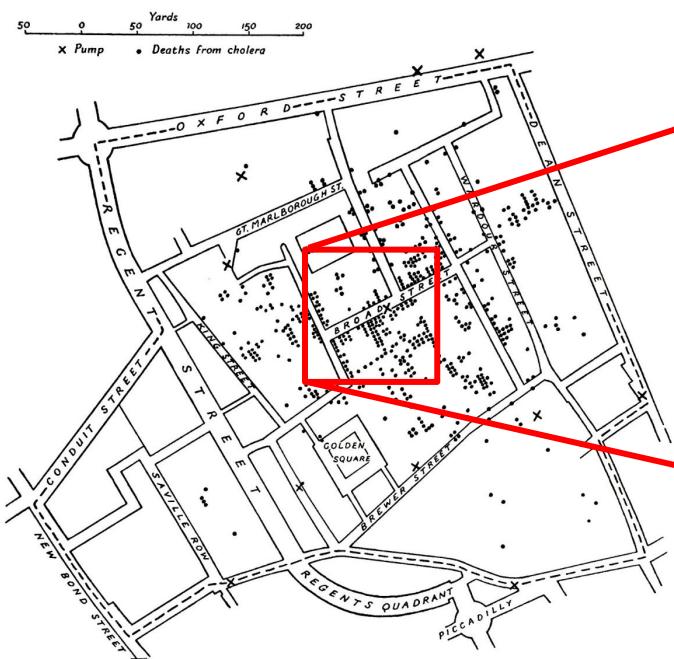
John Snow and Cholera



John Snow and Cholera



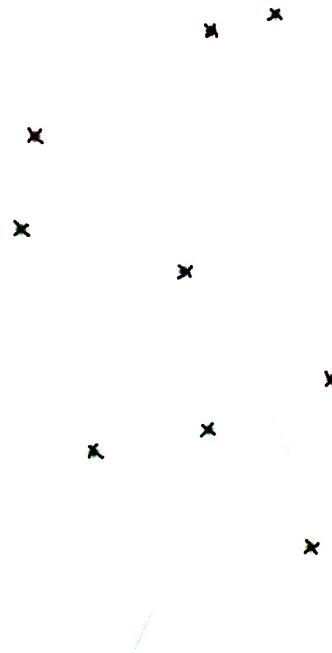
John Snow and Cholera



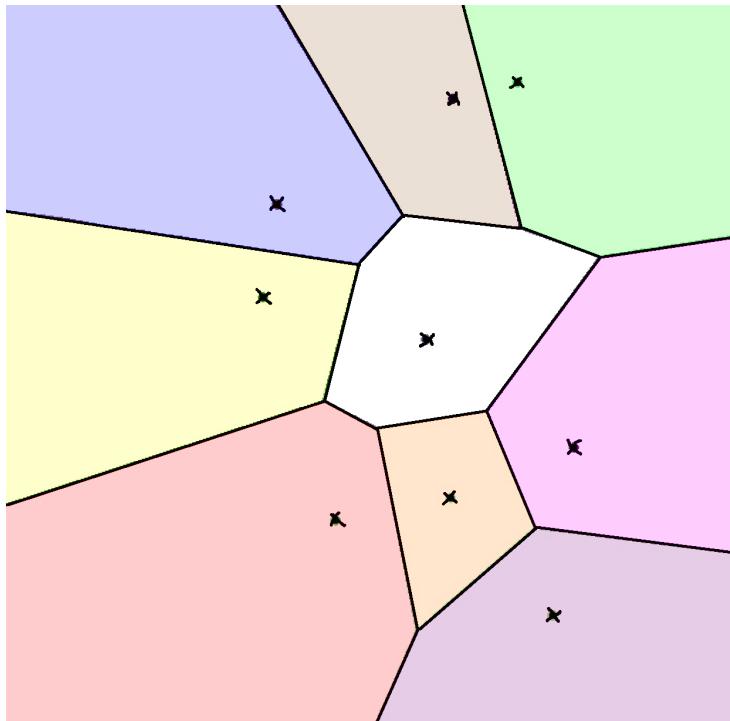
John Snow and Cholera



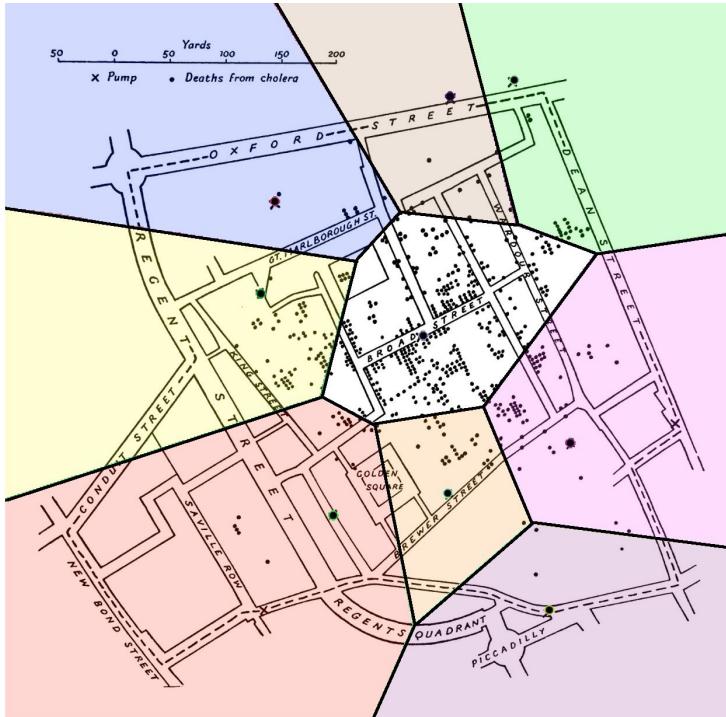
John Snow and Cholera



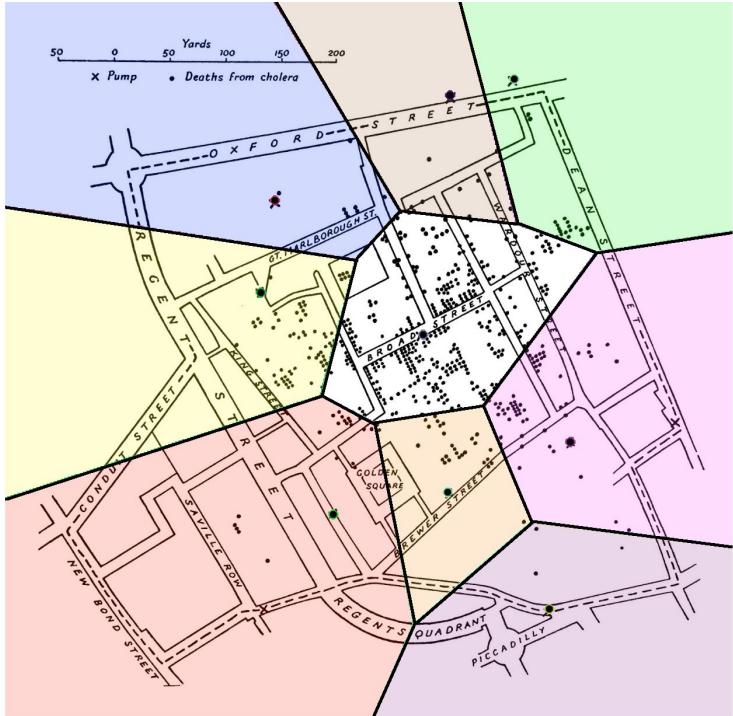
John Snow and Cholera



John Snow and Cholera



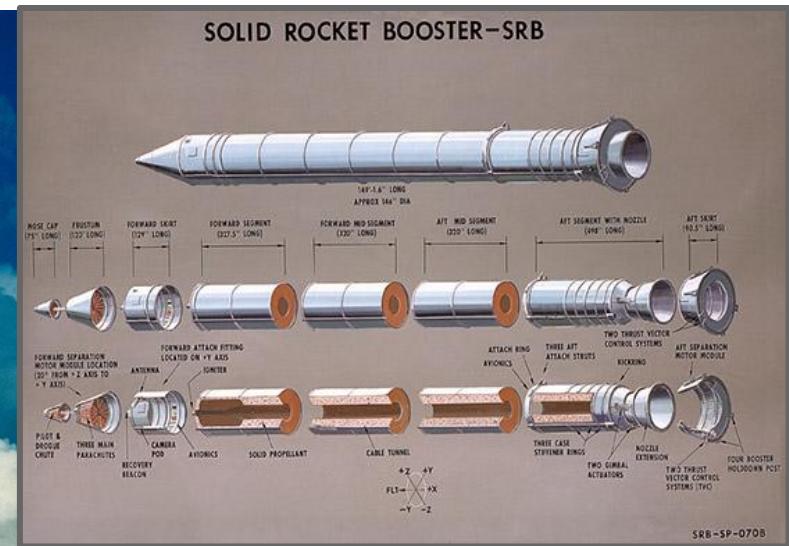
John Snow and Cholera



Morton Thiokol and Solid Rocket Boosters



Morton Thiokol and Solid Rocket Boosters

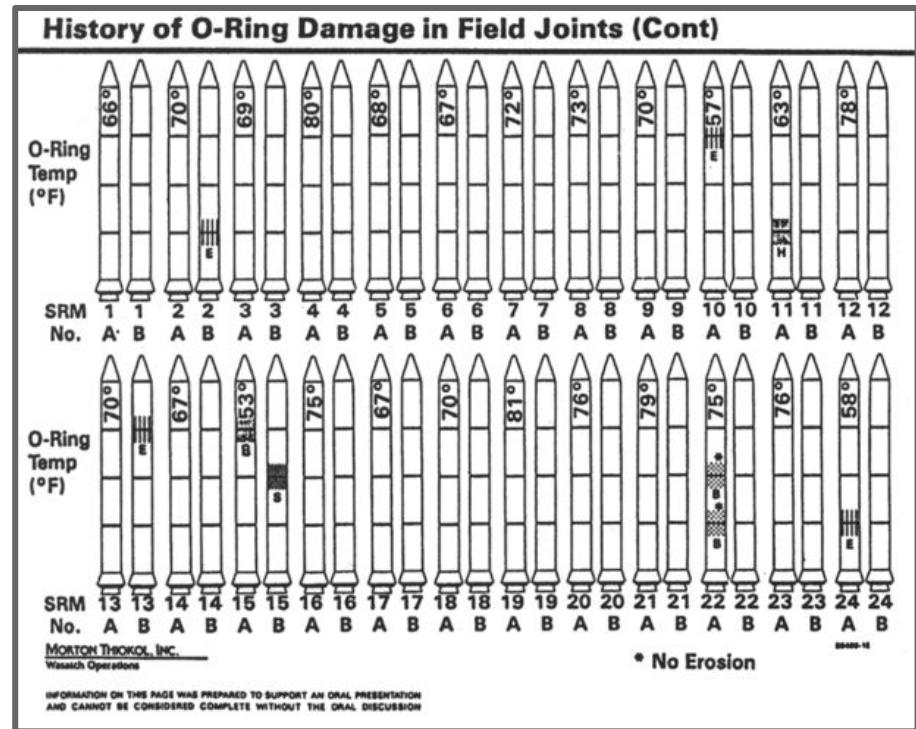


Morton Thoikol and Solid Rocket Boosters

- 1982: Evidence of O-Ring erosion
- 1985: Presentation in Washington
- 1986 (27 January): Teleconference
- 1986 (28 January): Launch

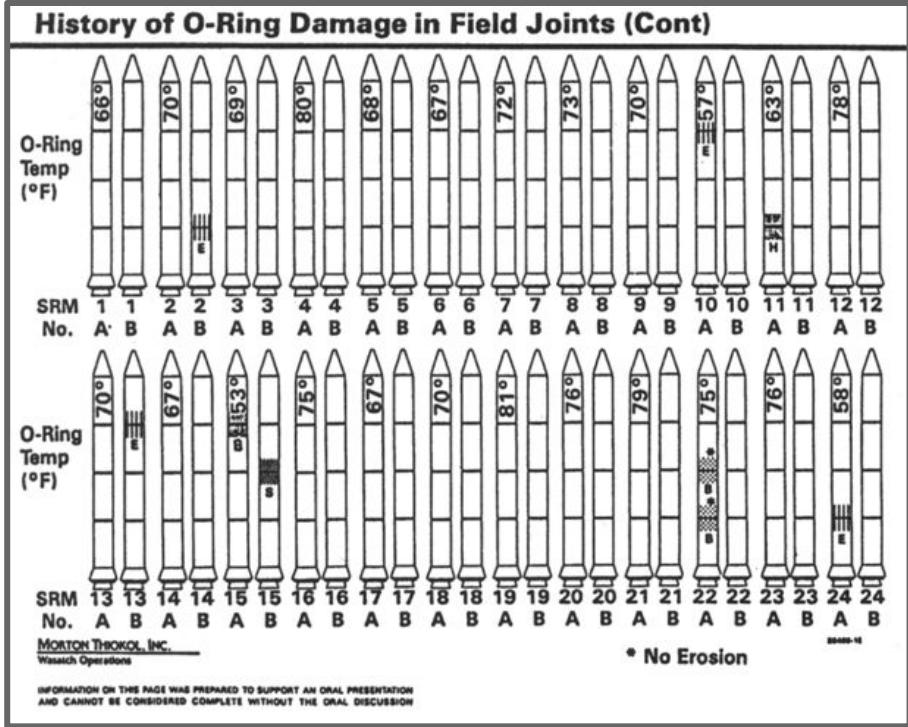
Morton Thoikol and Solid Rocket Boosters

- 1982: Evidence of O-Ring erosion
- 1985: Presentation in Washington
- 1986 (27 January): Teleconference
- 1986 (28 January): Launch



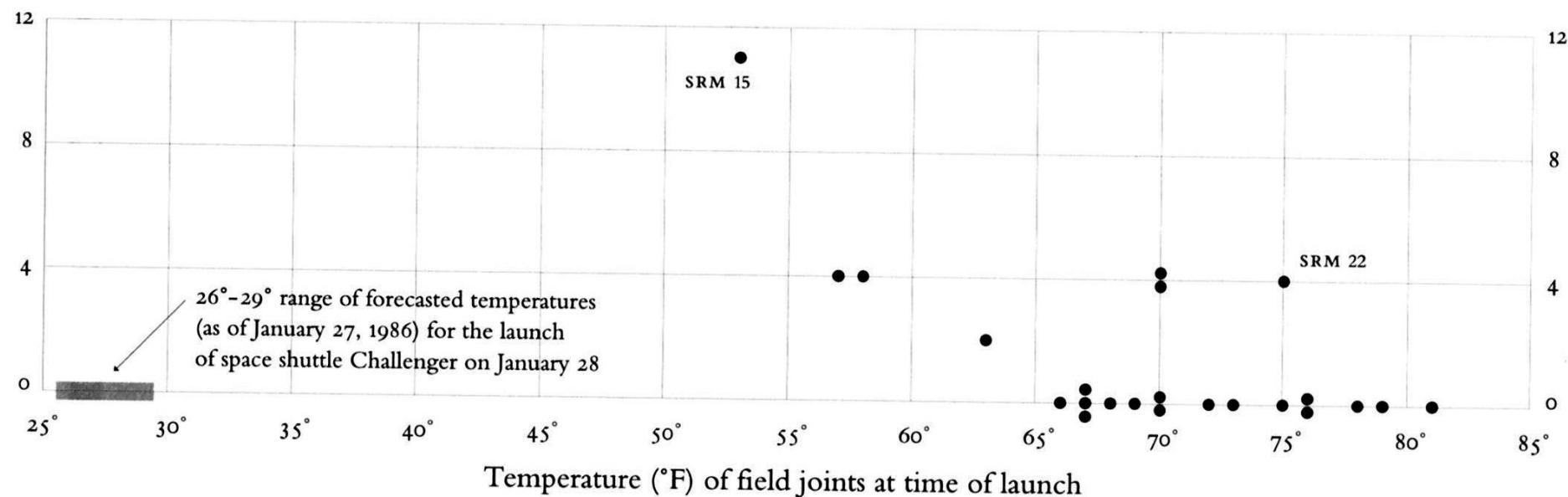
id Rocket Boosters

on
on
nce



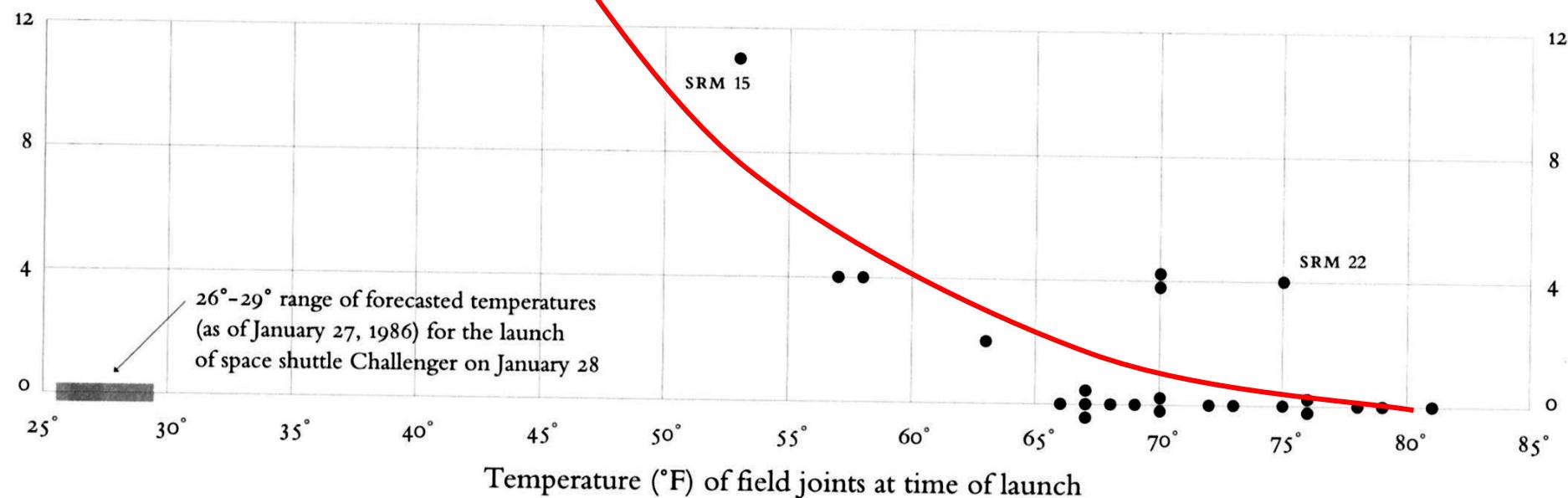
Morton Thoikol and Solid Rocket Boosters

O-ring damage
index, each launch



Morton Thoikol and Solid Rocket Boosters

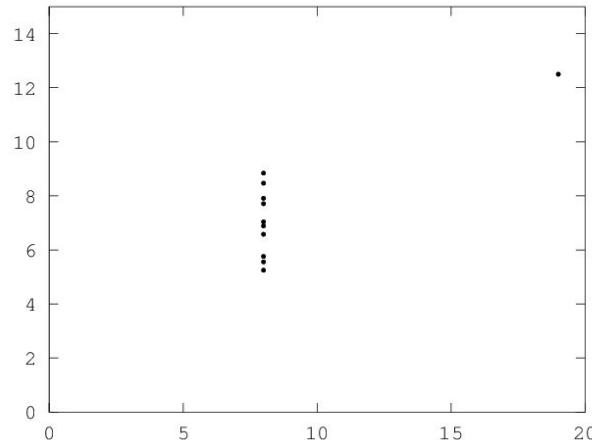
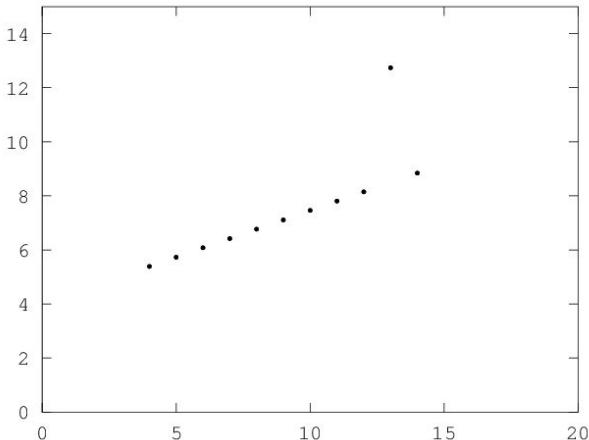
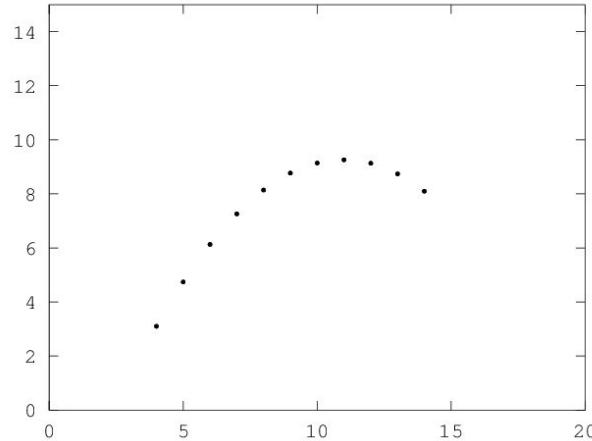
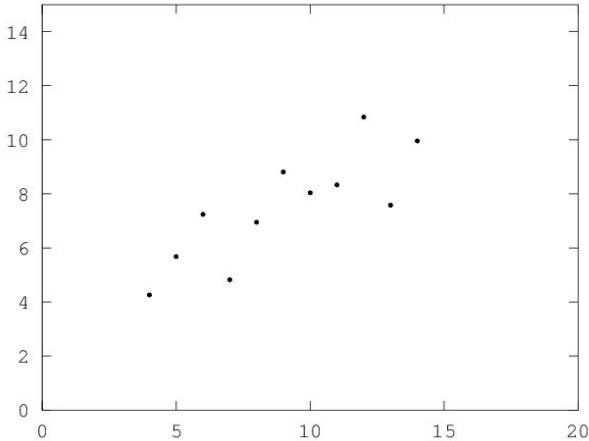
O-ring damage
index, each launch



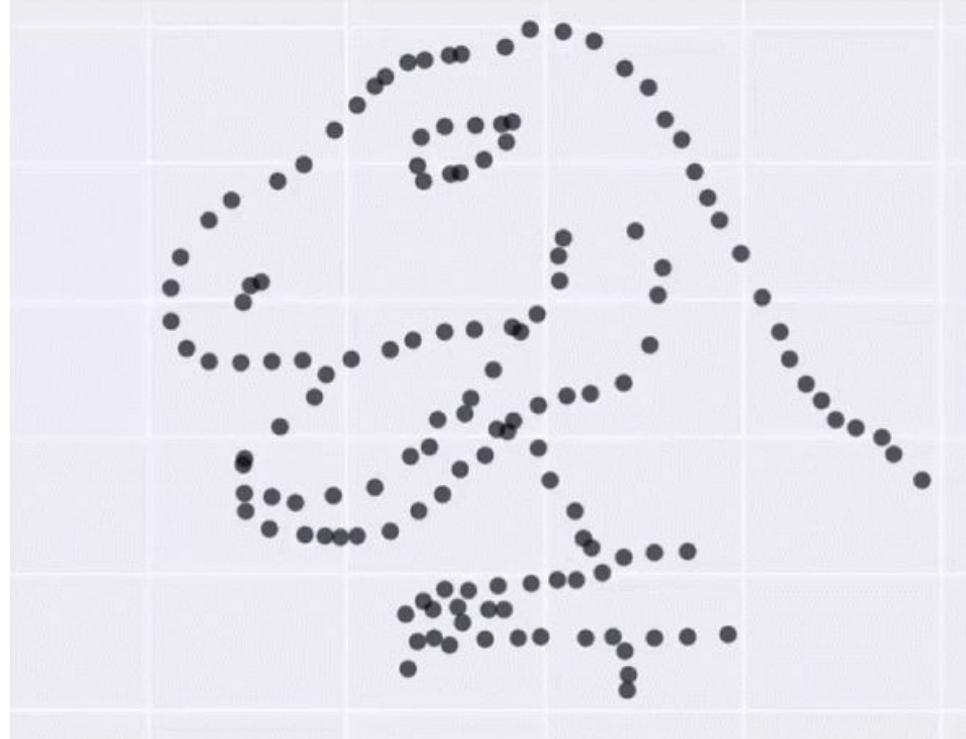
Anscombe's Quartet

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



The Datasaurus



<https://twitter.com/albertocairo/status/765167969139765250/photo/1>

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+Web
 - Bokeh / plotly / pygal
- Web
 - D3
- On the extremes:
 - OpenGL / WebGL
 - Tableau

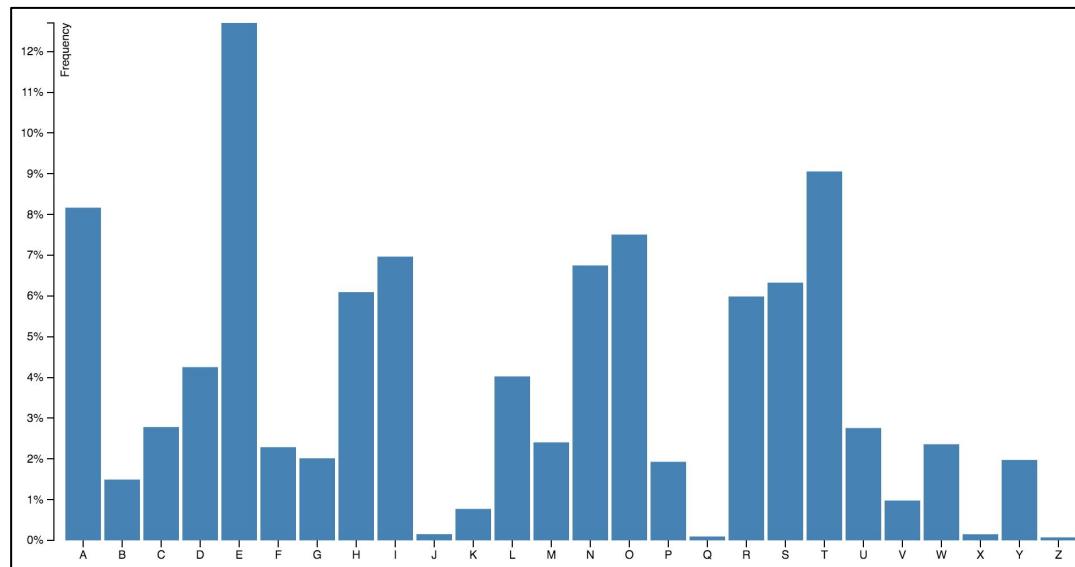


D3.js

- 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");
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); });
});</script>
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, "g");
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)
    .text("Frequency");
  style("text-anchor", "end")
  .attr("dx", -5)
  .attr("dy", 5);
  svg.selectAll("bar")
    .data(data)
    .enter()
    .append("rect")
    .attr("class", "bar")
    .attr("fill", 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); });
});</script>
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 ($=$, \neq)
 - Types and categories (mathematical set)

Attribute Domain Taxonomy

- Nominal ($=, \neq$)
 - Types and categories (mathematical *set*)
- Ordinal ($=, \neq, \leq$)
 - Has an order (mathematical *set* with *order relation*)

Attribute Domain Taxonomy

- Nominal ($=, \neq$)
 - Types and categories (mathematical *set*)
- Ordinal ($=, \neq, \leq$)
 - Has an order (mathematical *set* with *order relation*)
- Interval ($=, \neq, \leq, +, -$)
 - Has a meaningful difference between values (mathematical *group*)

Attribute Domain Taxonomy

- Nominal ($=, \neq$)
 - Types and categories (mathematical *set*)
- Ordinal ($=, \neq, \leq$)
 - Has an order (mathematical *set* with *order relation*)
- Interval ($=, \neq, \leq, +, -$)
 - Has a meaningful difference between values (mathematical *group*)
- Ratio ($=, \neq, \leq, +, -, \times, \div$)
 - Has a meaningful one and zero point and ratio between values (mathematical *field*)

Attribute Domain Taxonomy

- Nominal ($=, \neq$)
 - Types and categories (mathematical *set*)
- Ordinal ($=, \neq, \leq$)
 - Has an order (mathematical *set* with *order relation*)
- Quantitative
 - Interval ($=, \neq, \leq, +, -$)
 - Has a meaningful difference between values (mathematical *group*)
 - Ratio ($=, \neq, \leq, +, -, \times, \div$)
 - Has a meaningful one and zero point and ratio between values (mathematical *field*)

Attribute Domain Taxonomy

- Nominal ($=, \neq$)
 - Types and categories (mathematical *set*)
- Ordinal ($=, \neq, \leq$)
 - Has an order (mathematical *set* with *order relation*)
- Quantitative
 - Interval ($=, \neq, \leq, +, -$)
 - Has a meaningful difference between values (mathematical *group*)
 - E.g: Dates, location, geometric points, temperature (C and F)
 - Ratio ($=, \neq, \leq, +, -, \times, \div$)
 - 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 ($=, \neq$)
 - Types and categories (mathematical *set*)
- Ordinal ($=, \neq, \leq$)
 - Has an order (mathematical *set* with *order relation*)
- Quantitative
 - Interval ($=, \neq, \leq, +, -$)
 - Has a meaningful difference between values (mathematical *group*)
 - E.g: Dates, location, geometric points, temperature (C and F)
 - Ratio ($=, \neq, \leq, +, -, \times, \div$)
 - 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

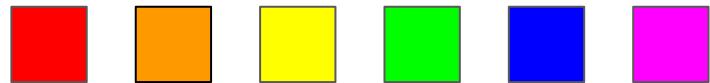
Visual Encodings of **Nominal** Data

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



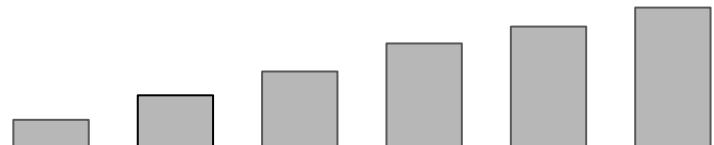
Color luminance



Area, Volume



Position

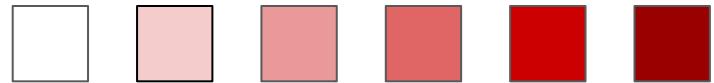


Time, Animation

Visual Encodings of **Quantitative** Data

Visual Encodings of Quantitative Data

Color saturation



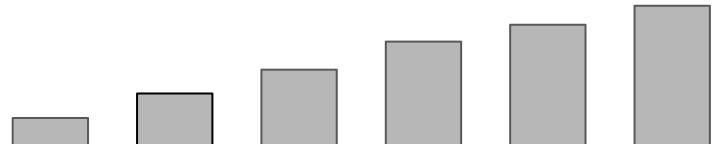
Color luminance



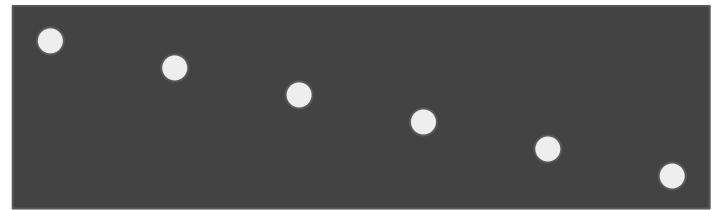
Position



Length, Area, Volume



Angle



Time, Animation



Charts

Chart: *noun*

1. A common pattern for combining visual encodings of data.

Charts

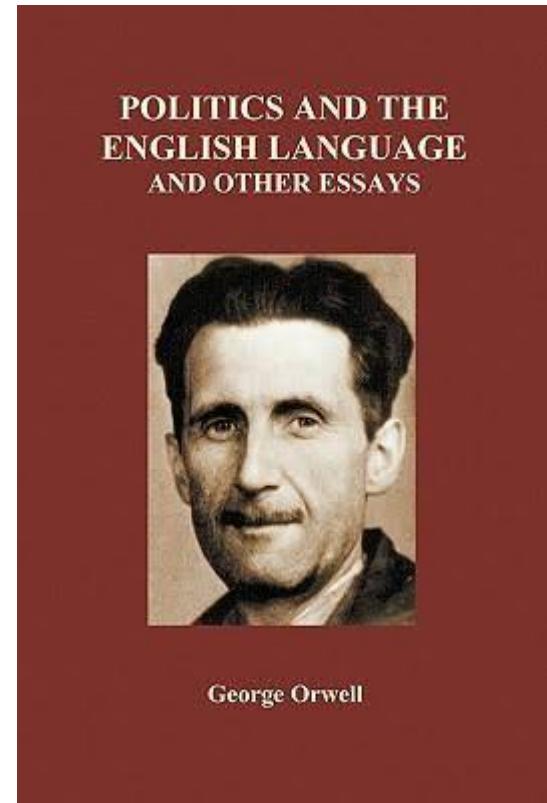
Chart: *noun*

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

Charts

Chart: *noun*

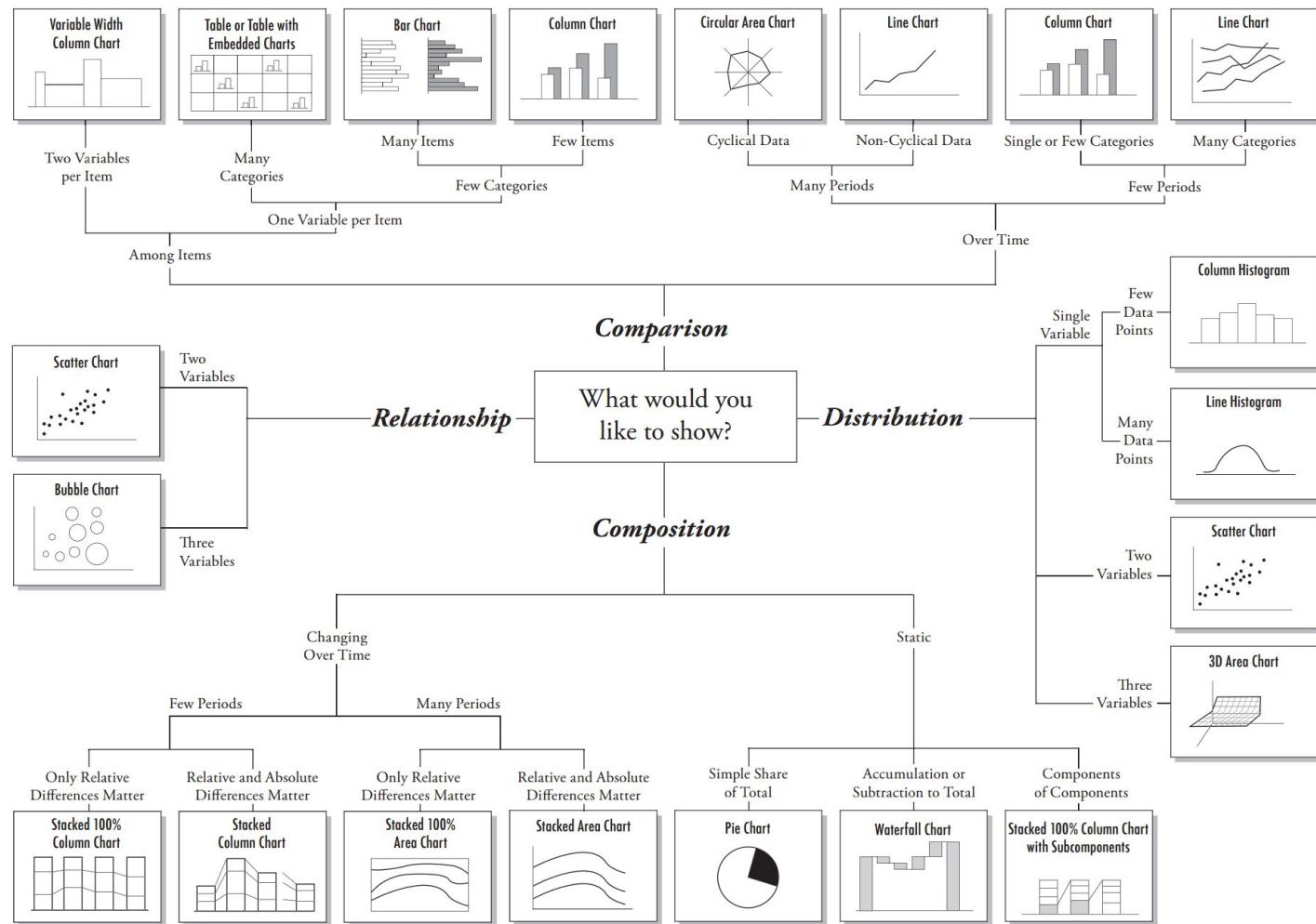
1. A common pattern for combining visual encodings of data.
2. A visual sentence constructed with data encodings as words.



George Orwell

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>

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

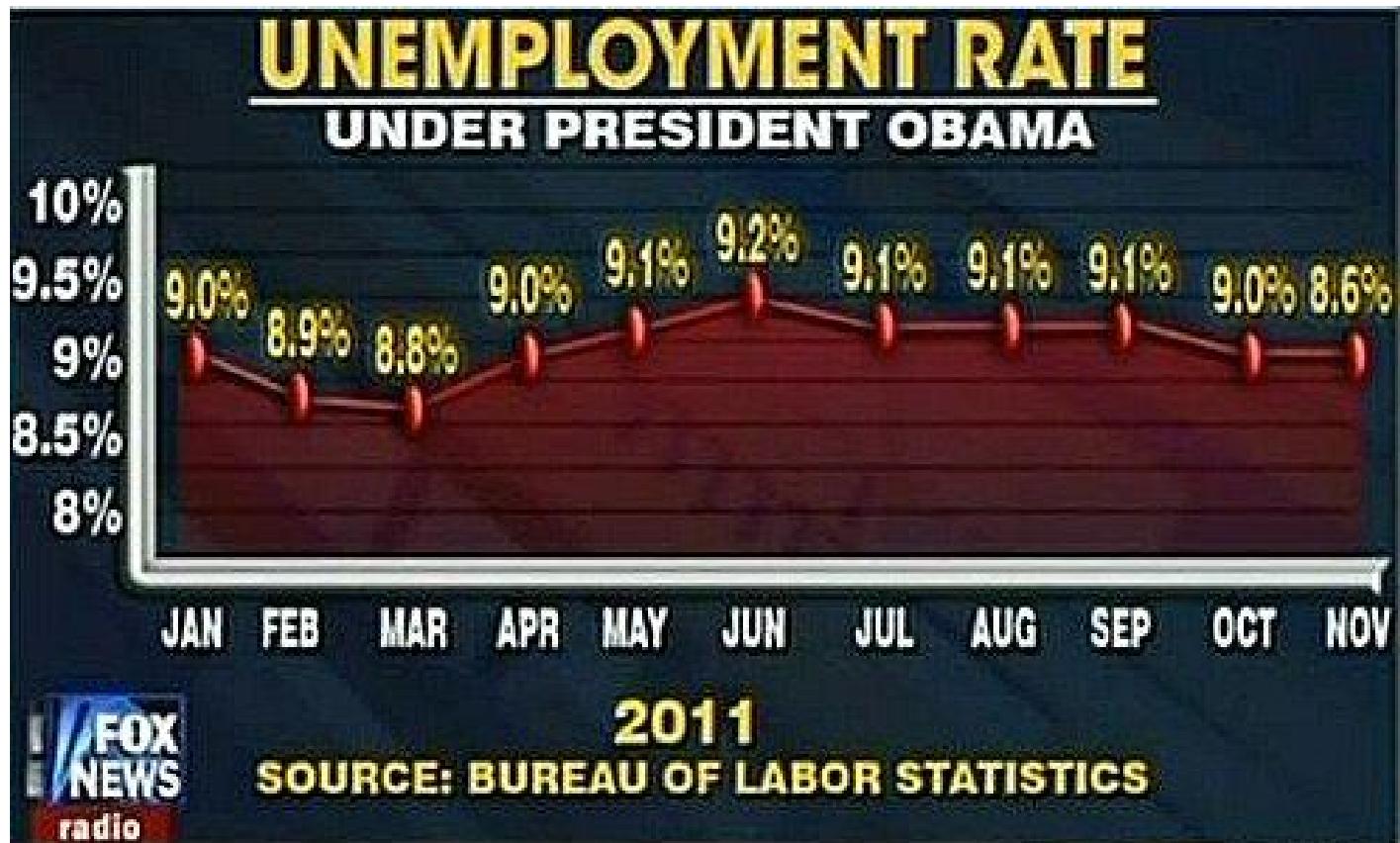
Making Good Visualizations

Not Making Bad Visualizations

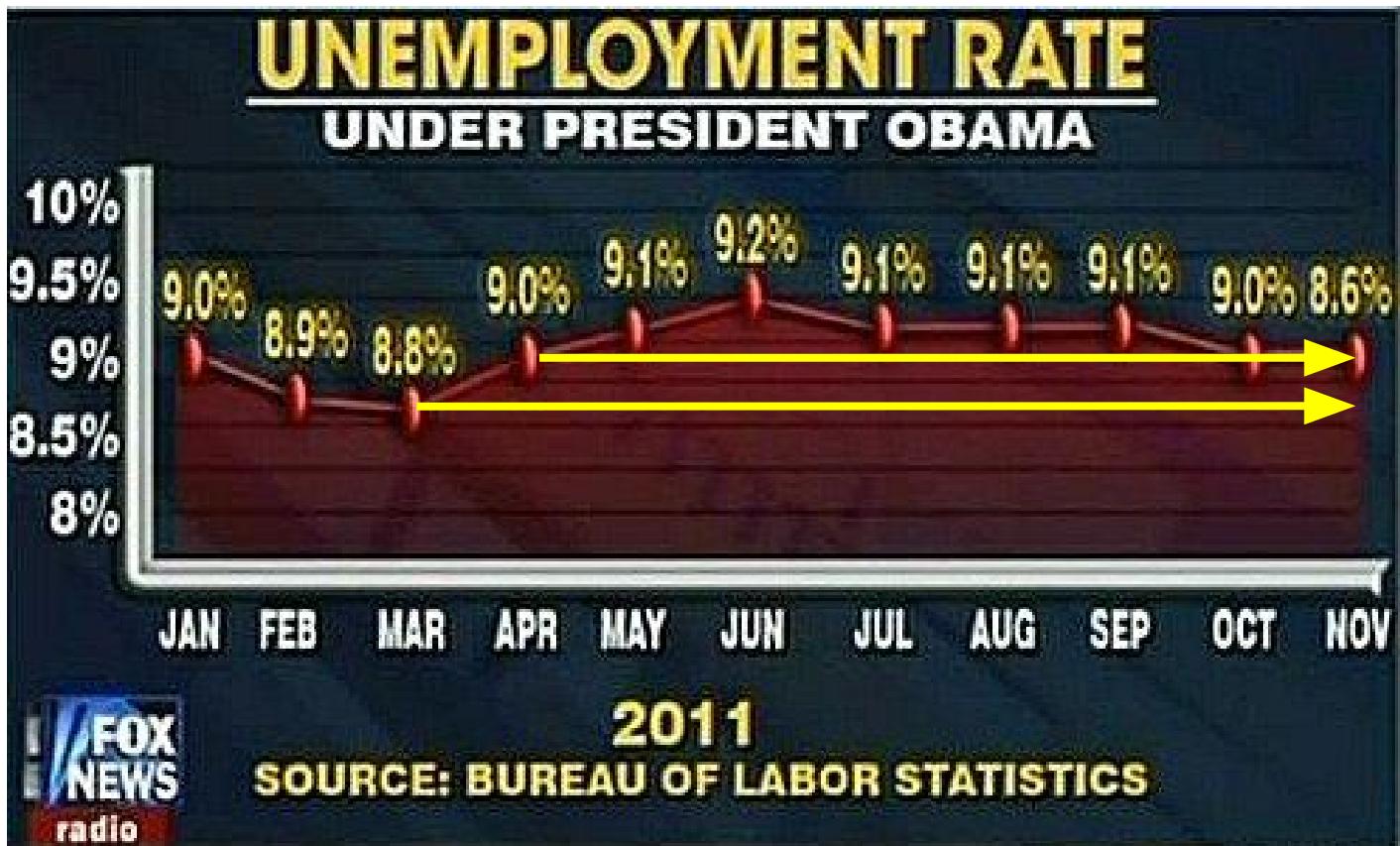
Graphical Integrity

Accurately Representing Data

Accurately Representing Data



Accurately Representing Data



Graphical Integrity: Principles

- Proportionality
- Matching dimensions
- Providing context

Proportionality

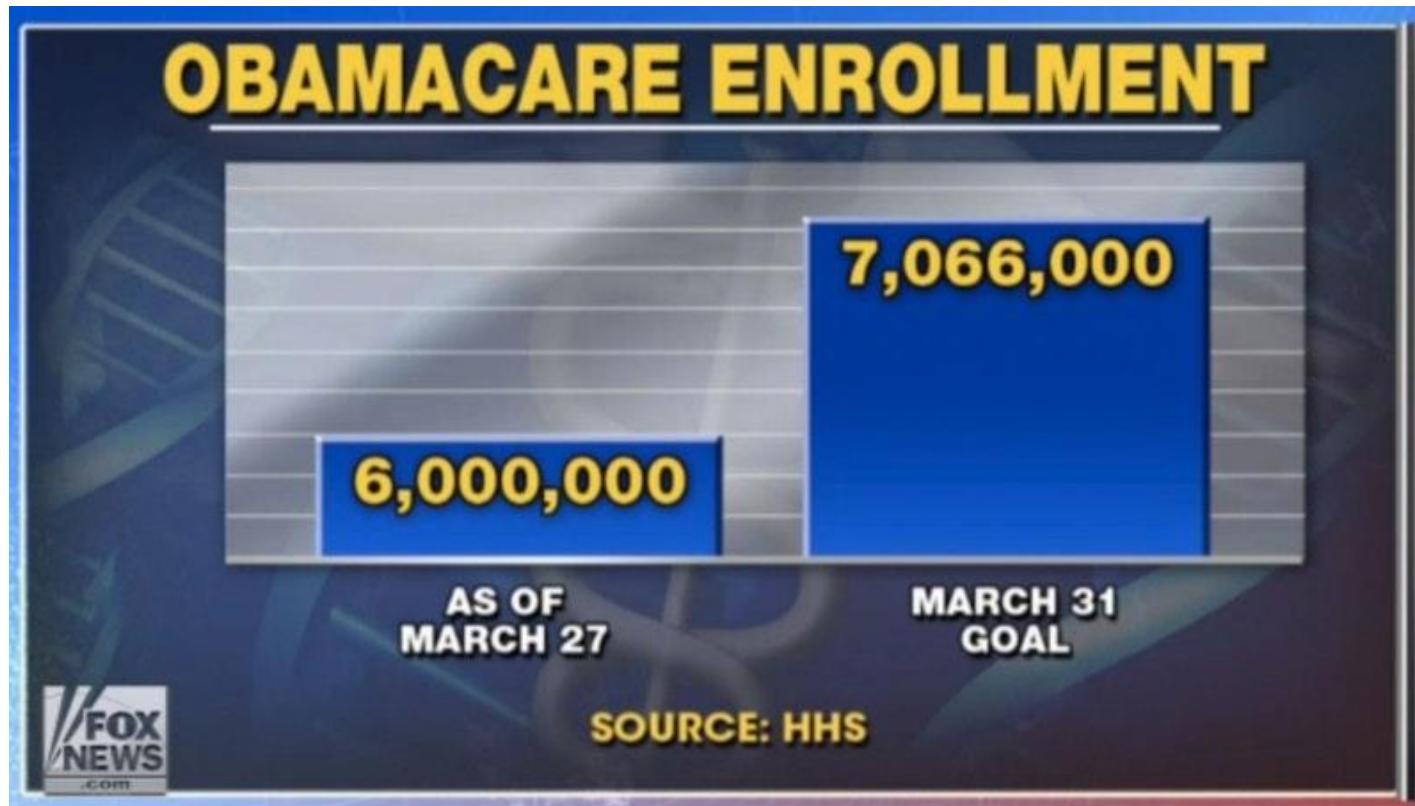
The *physical measurements of the representation of the data* should be proportional to the data itself

Proportionality

The *physical measurements of the representation of the data* should be proportional to the data itself

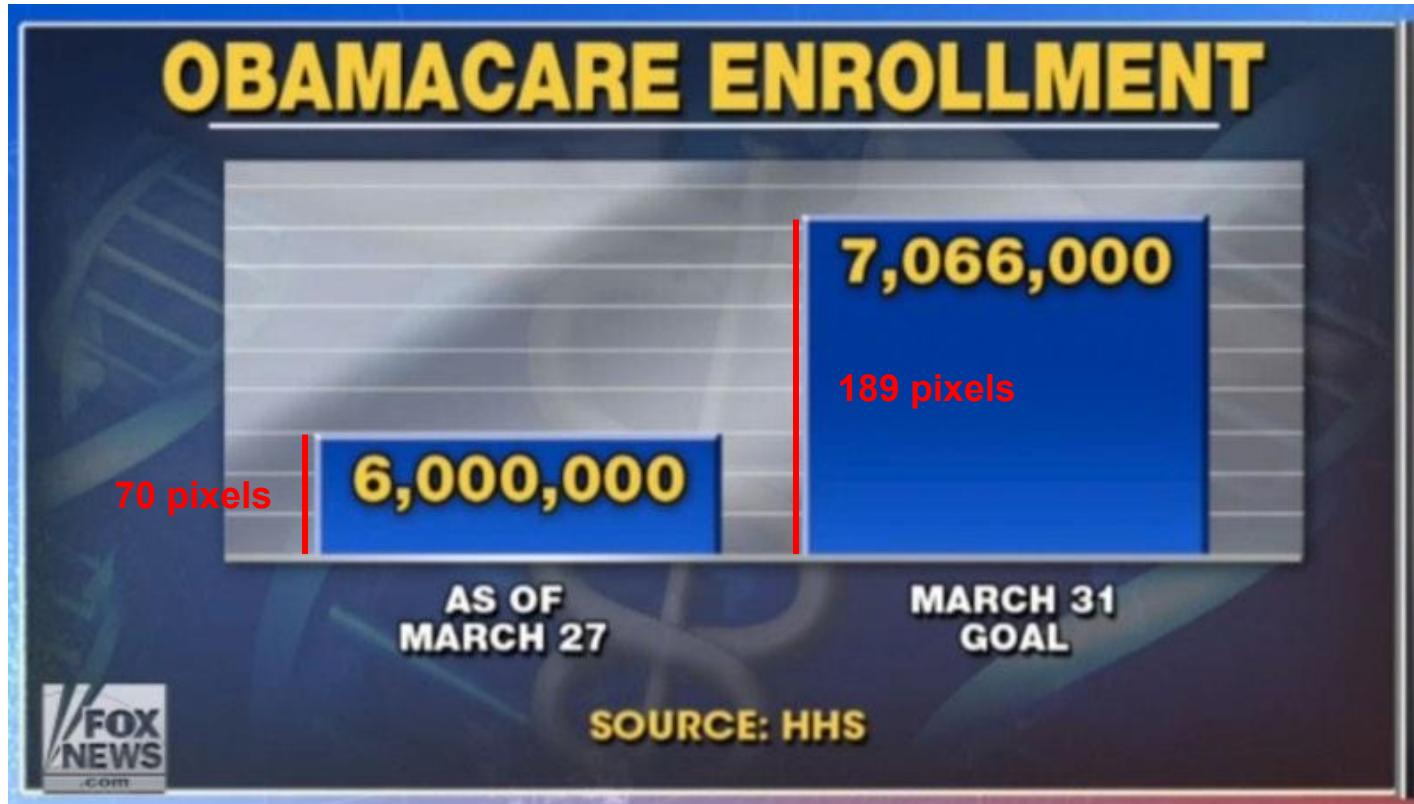
$$\text{Lie factor} = \frac{\text{size of effect shown in graph}}{\text{size of effect in the data}}$$

Proportionality



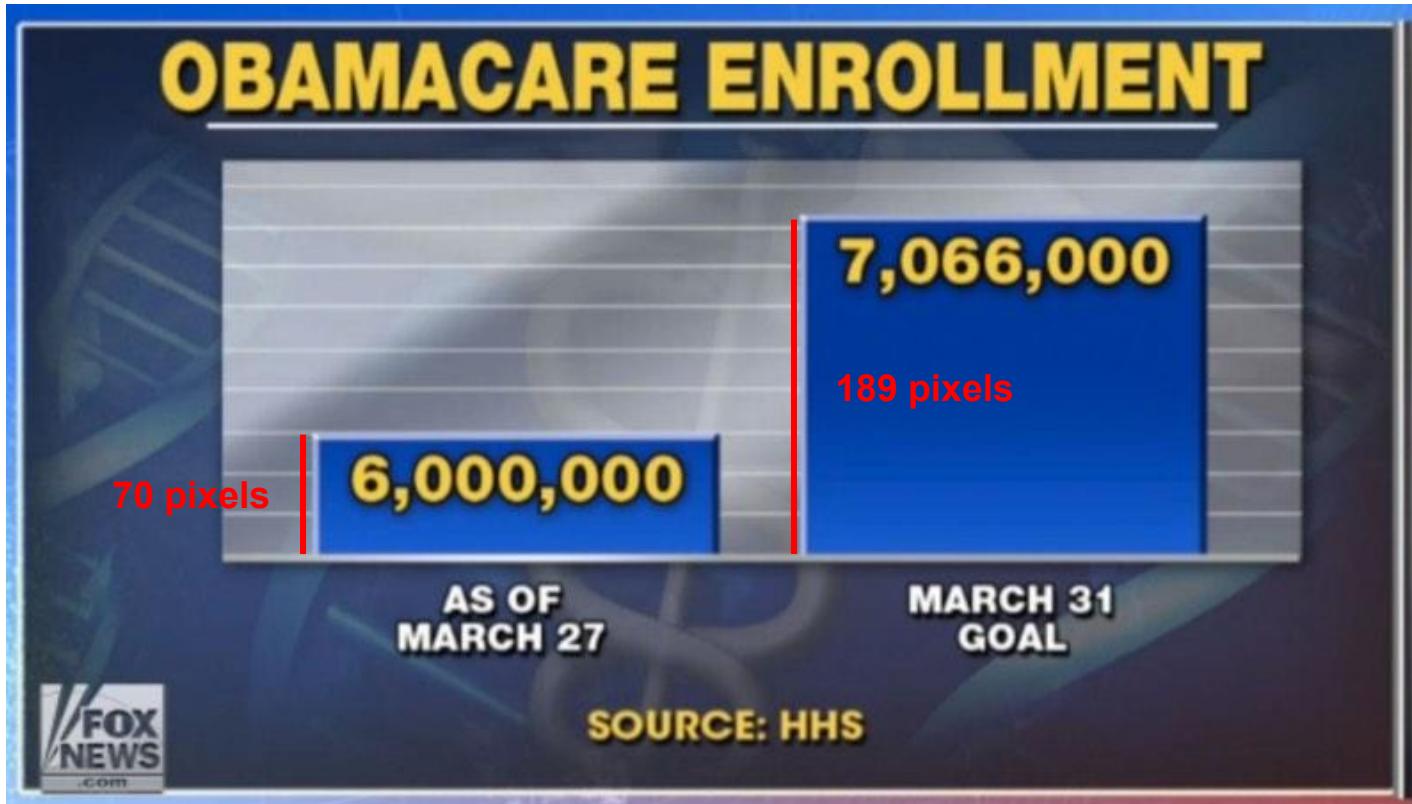
Proportionality

Graphical Ratio: $189\text{px} / 70\text{px} = 2.7$
Data Ratio: $6\text{m} / 7.1\text{m} = 1.8$

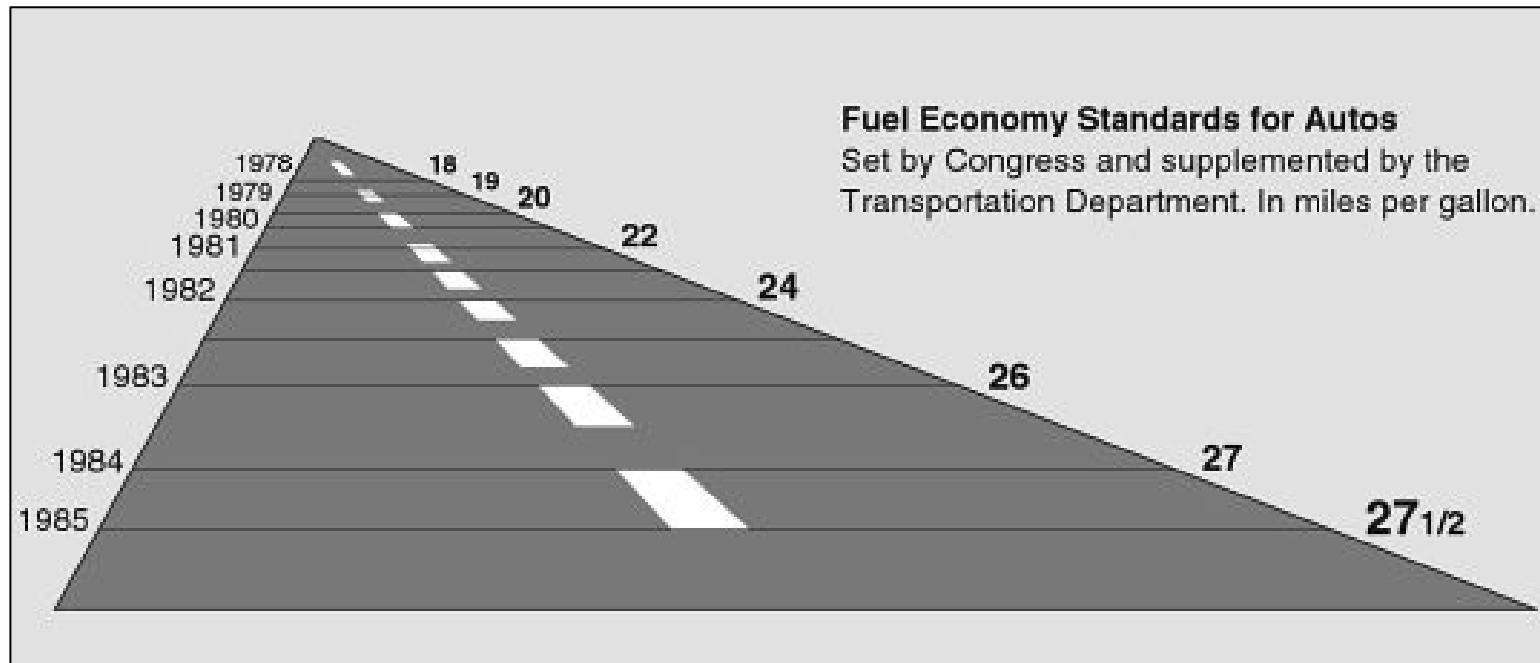


Proportionality

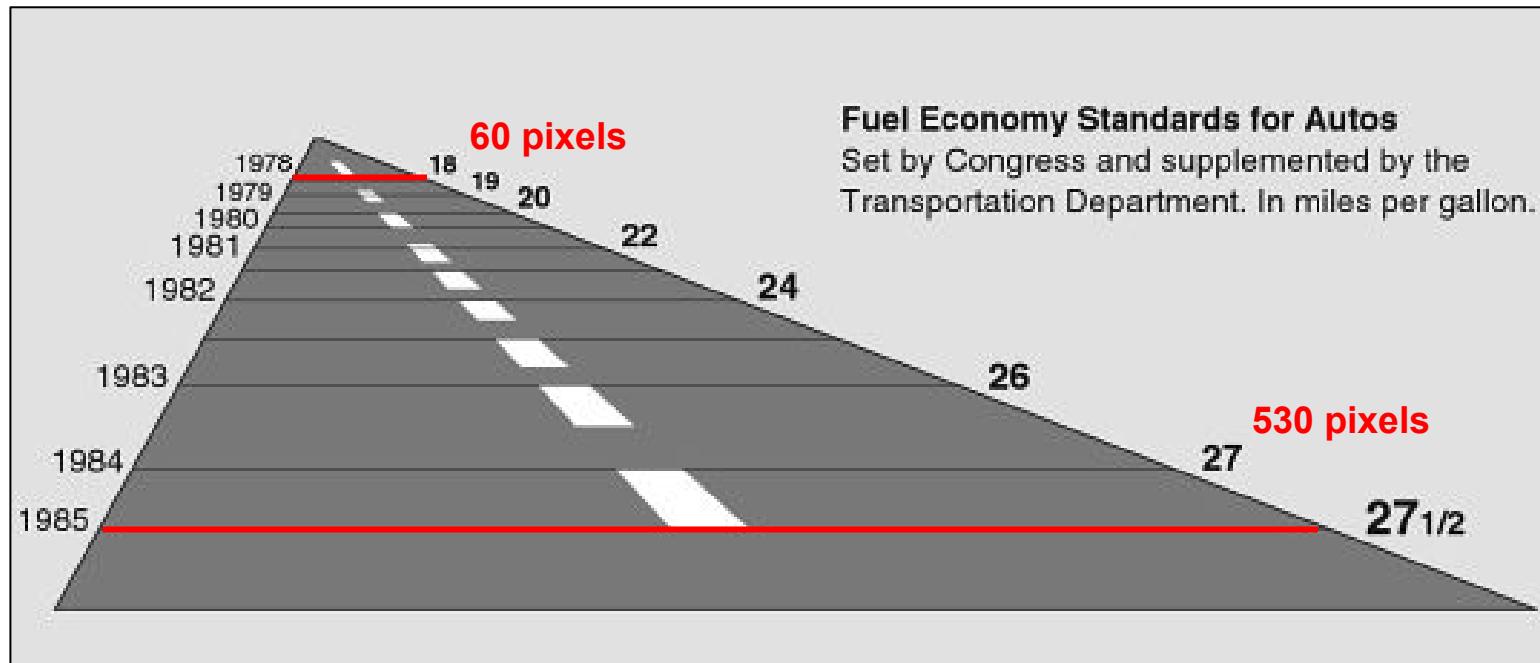
Graphical Ratio: $189\text{px} / 70\text{px} = 2.7 = 2.3$
Data Ratio: $6\text{m} / 7.1\text{m} = 1.8$



Proportionality

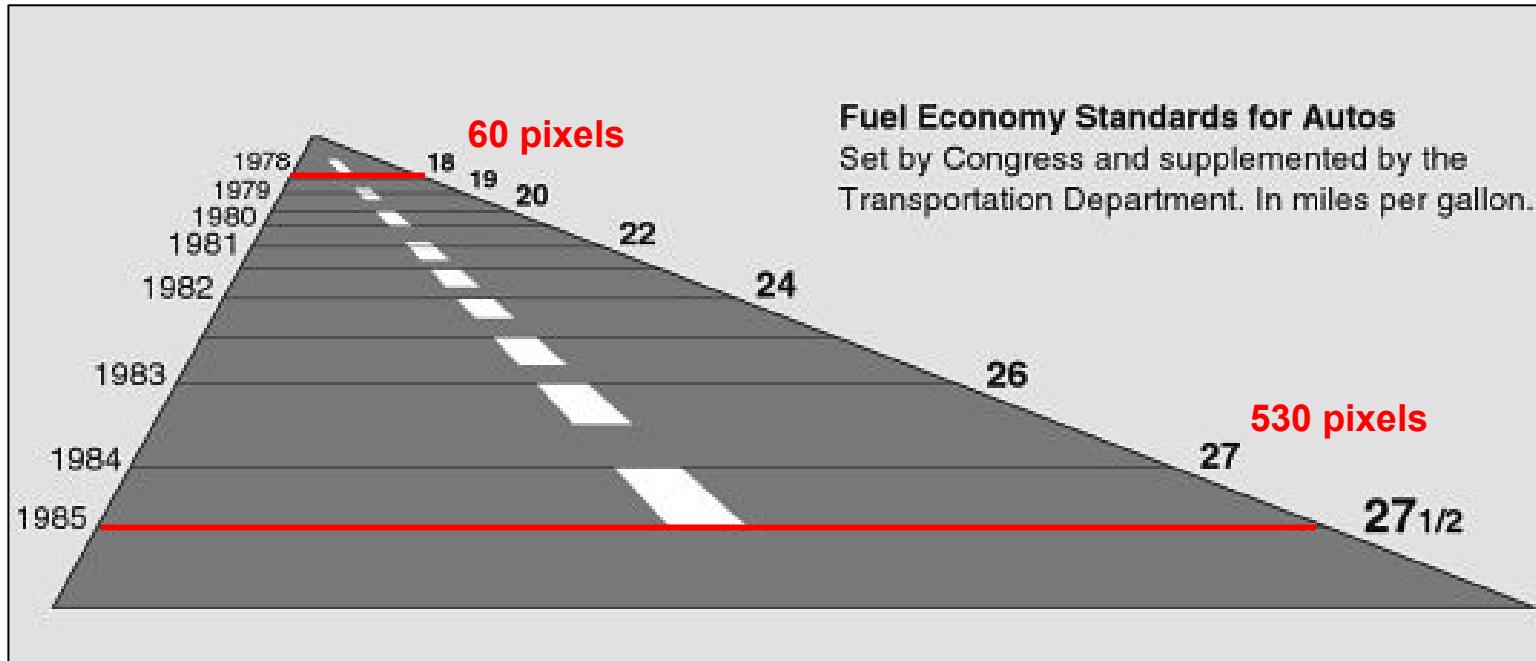


Proportionality



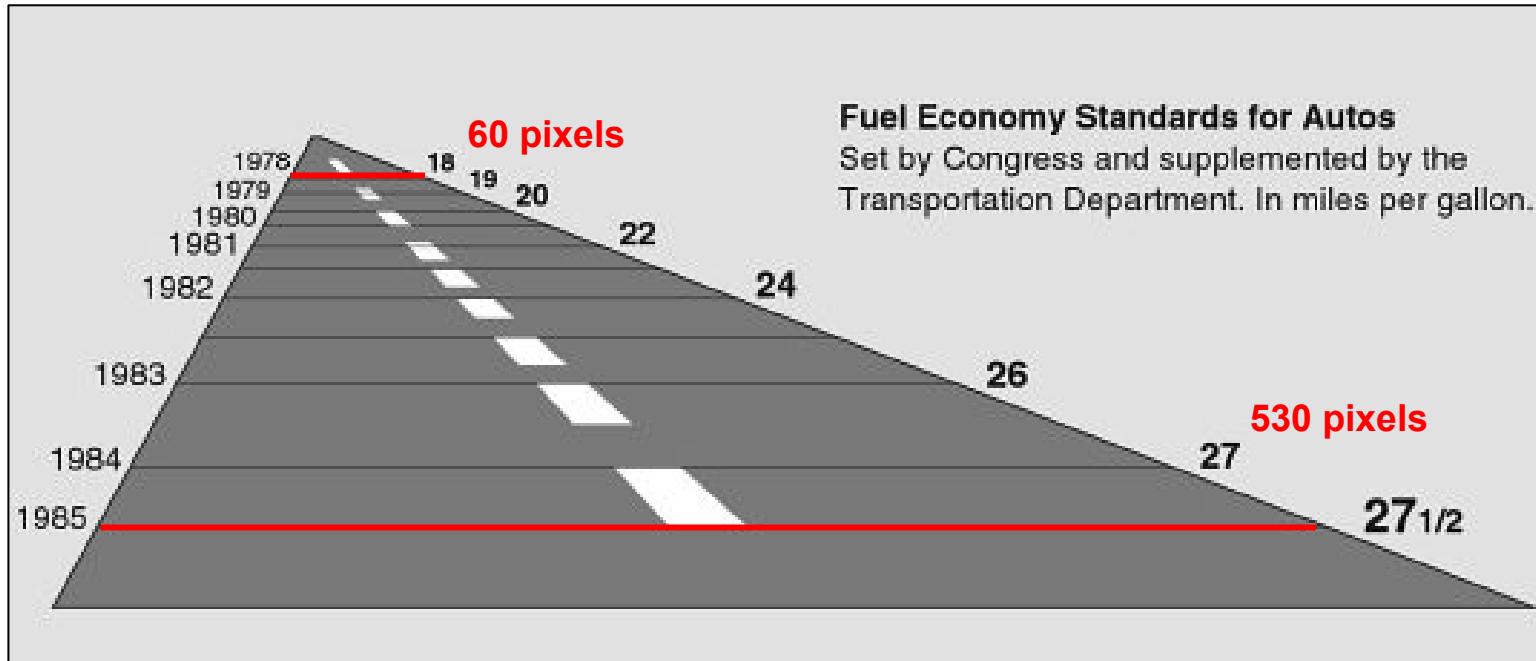
Proportionality

Graphical Ratio: $530\text{px} / 60\text{px} = 8.8$
Data Ratio: $27.5\text{mpg} / 18\text{mpg} = 1.5$

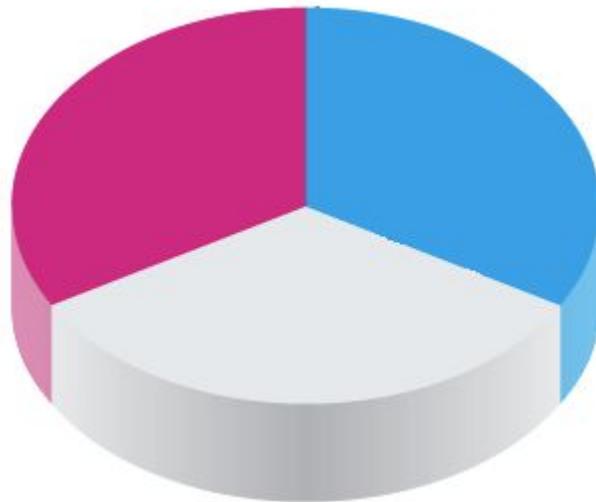


Proportionality

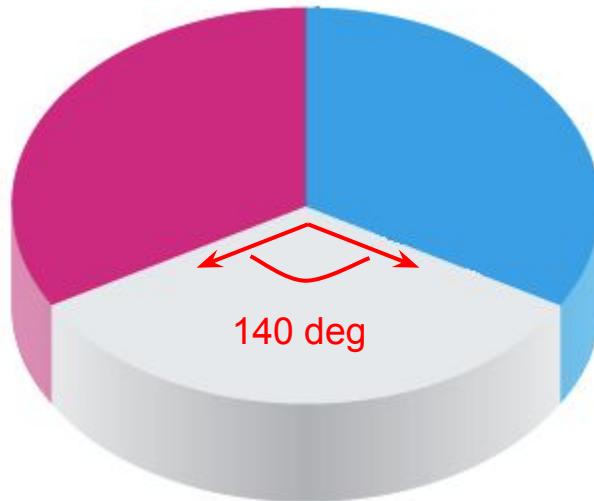
Graphical Ratio: $530\text{px} / 60\text{px} = 8.8$ Data Ratio: $27.5\text{mpg} / 18\text{mpg} = 1.5$ = 5.86



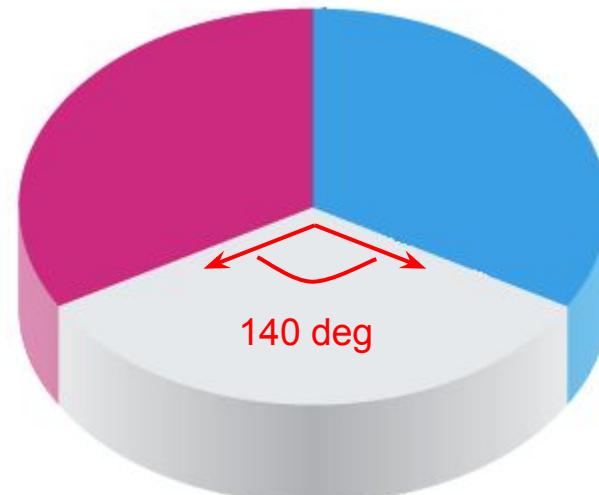
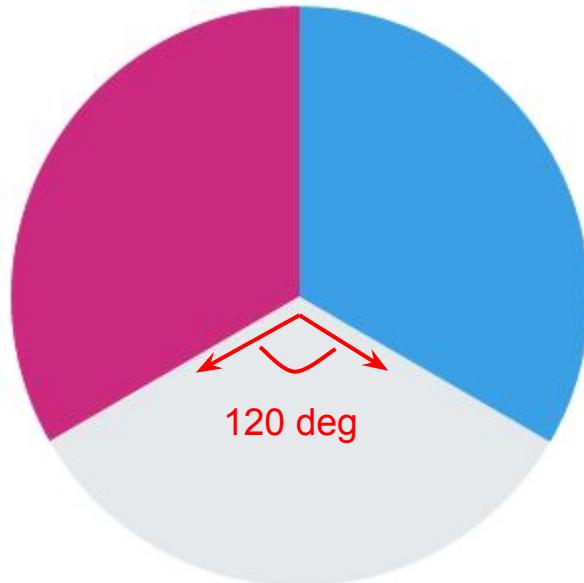
Proportionality & Perspective



Proportionality & Perspective



Proportionality & Perspective



THE SHRINKING FAMILY DOCTOR

In California

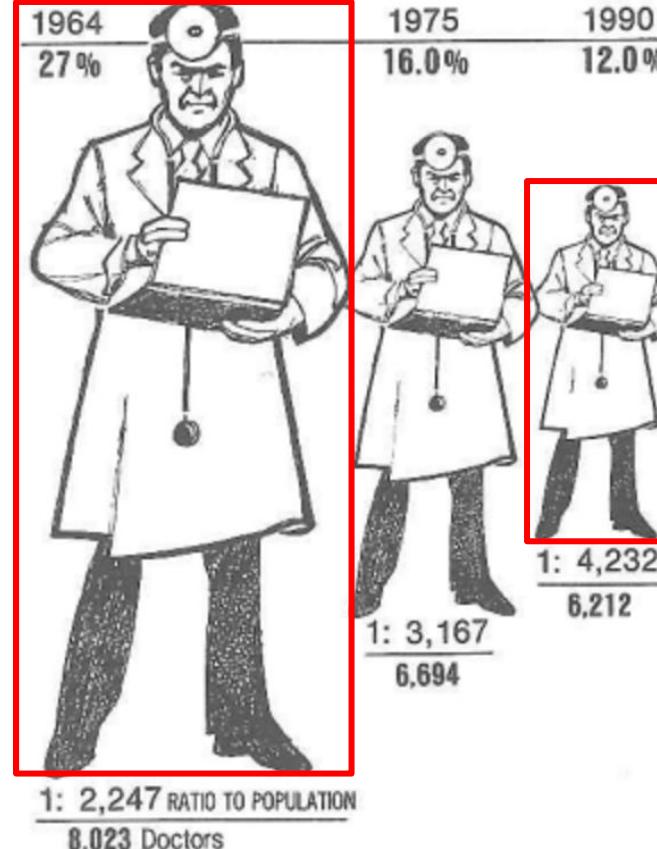
Match Dimensions

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

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

686 pixels tall

281x686 = 192k pixels



Match Dimensions

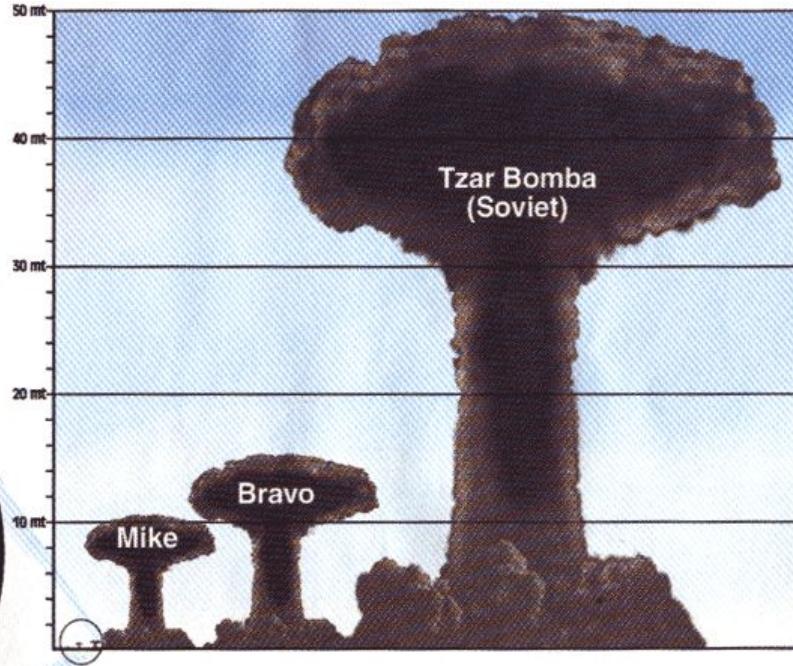
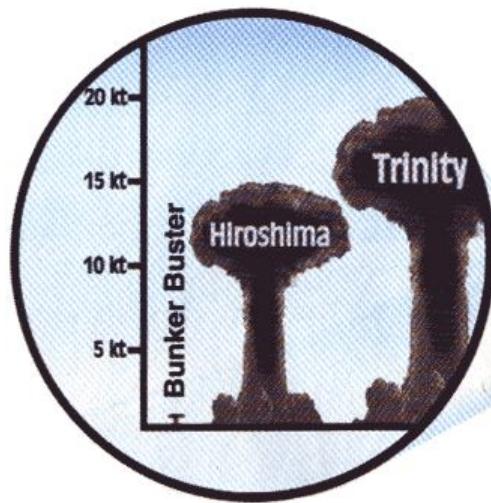
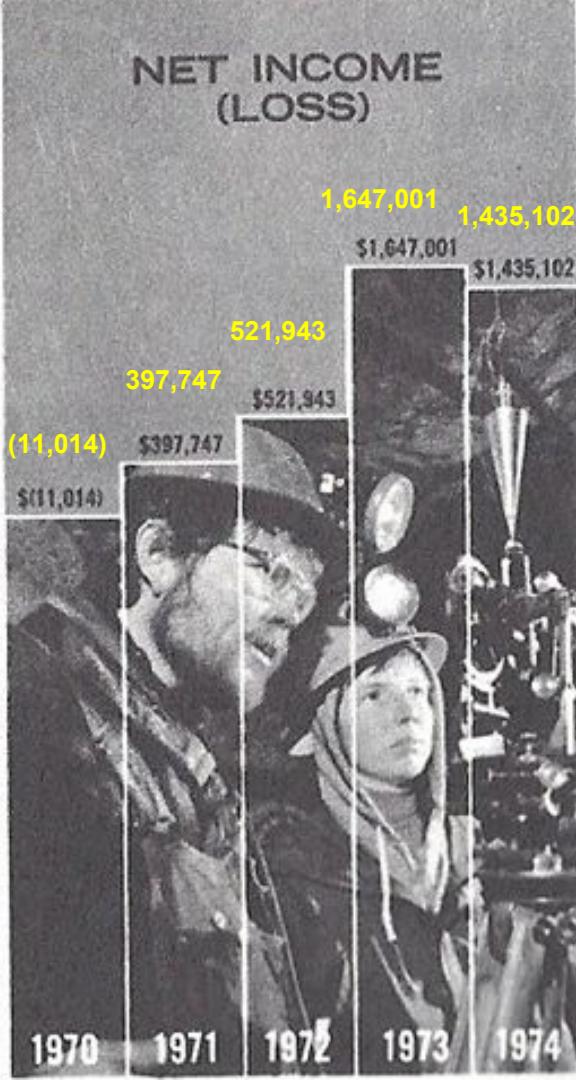
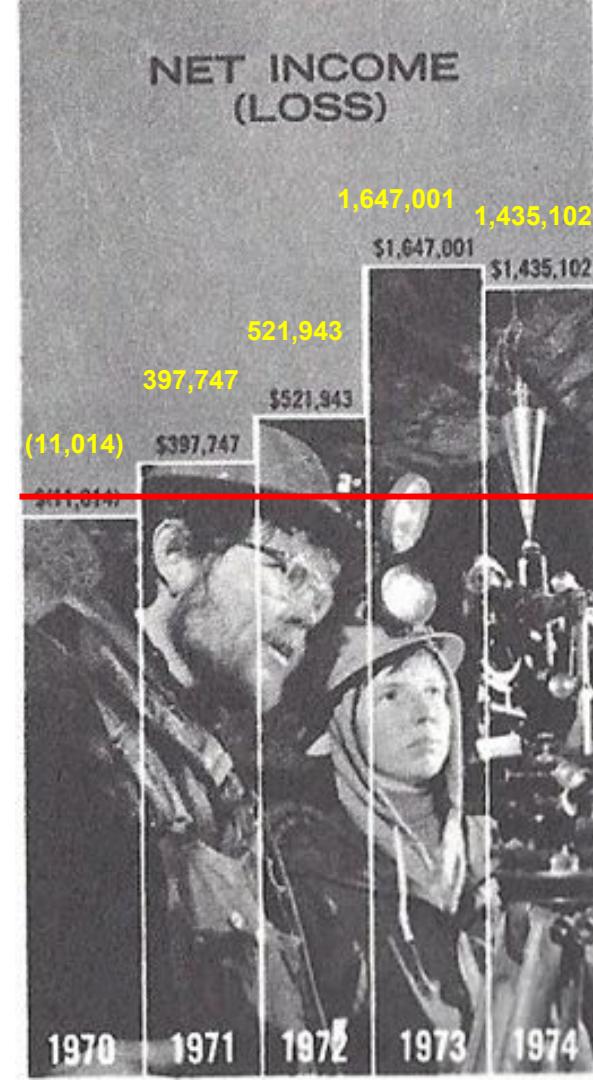


Illustration From October 2002
Issue of "Popular Mechanics" (pg. 69)

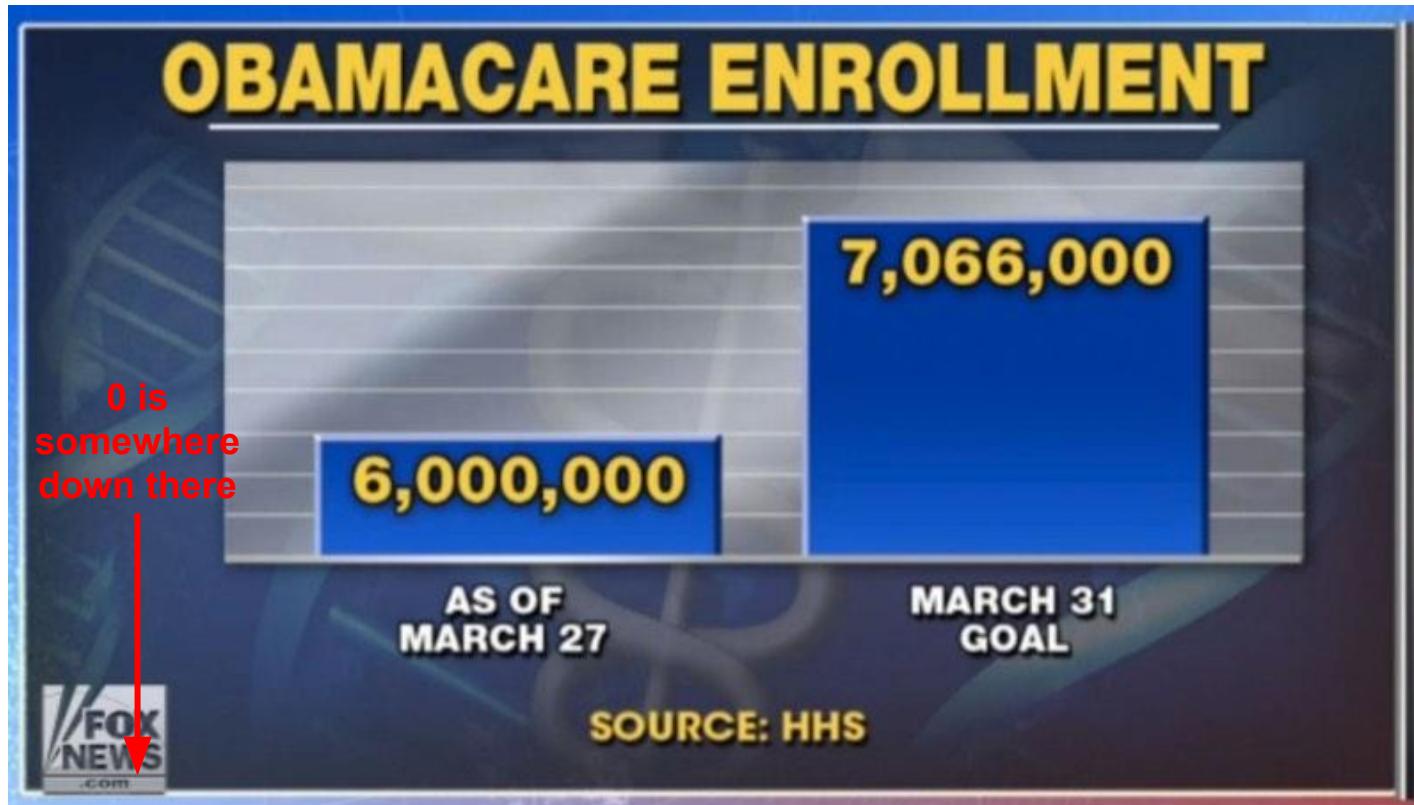
Providing Context



Providing Context



Providing Context



Graphical Integrity: Principles

- Proportionality
 - The *physical measurements of the representation of the data* should be proportional to the data itself
- Matching dimensions
 - Beware of pitfalls of using area, volume, and perspective
- Providing context
 - Anchor the audience

Maximizing Impact

(don't waste pixels)

Data ink

- The ink (pixels) devoted to representing data

Data ink Ratio

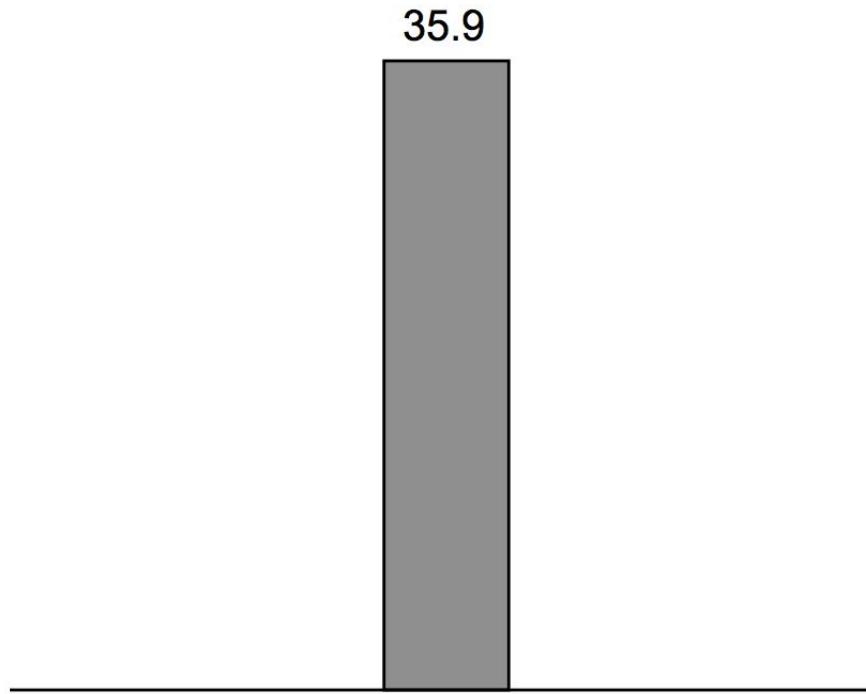
$$\text{Data ink ratio} = \frac{\text{Data ink}}{\text{Total ink}}$$

Non-data ink

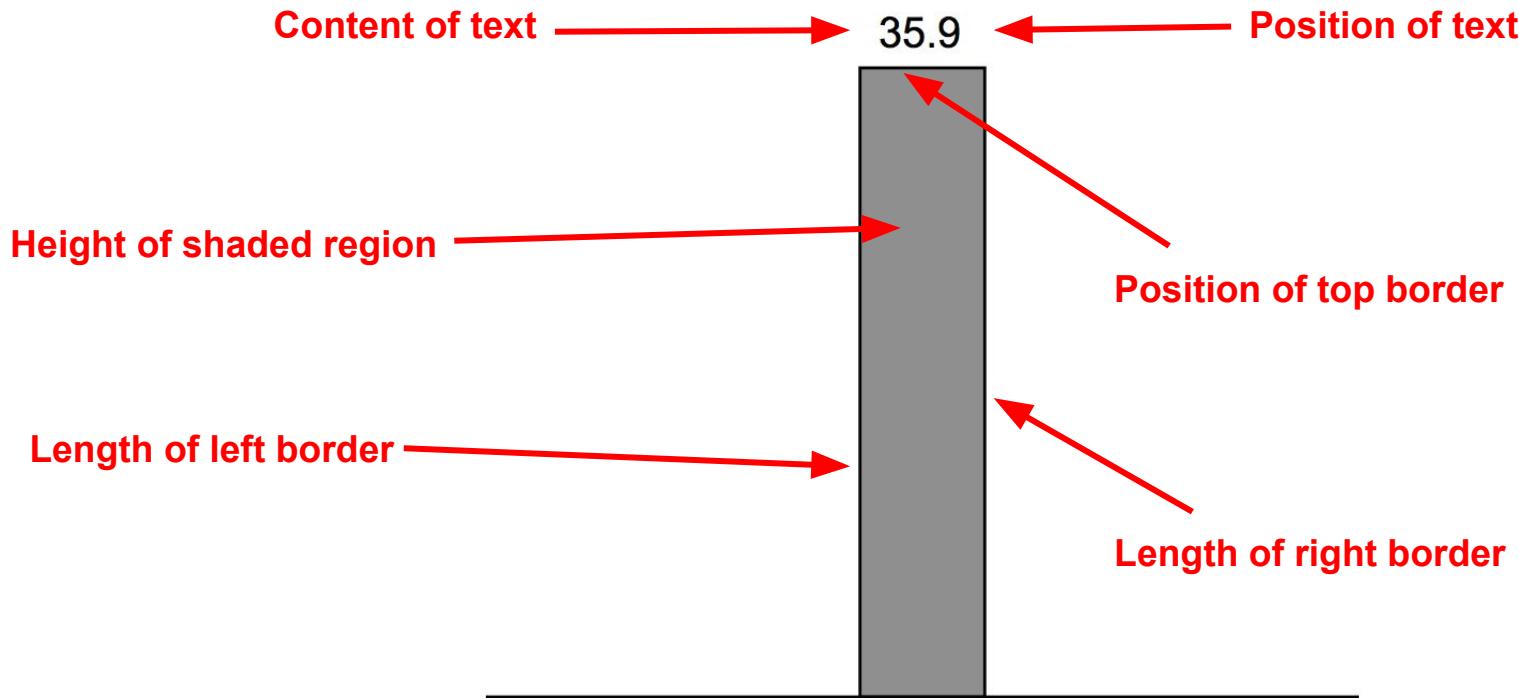
- Redundant data
- Metadata
- Chartjunk

Redundant Data

Redundant Data

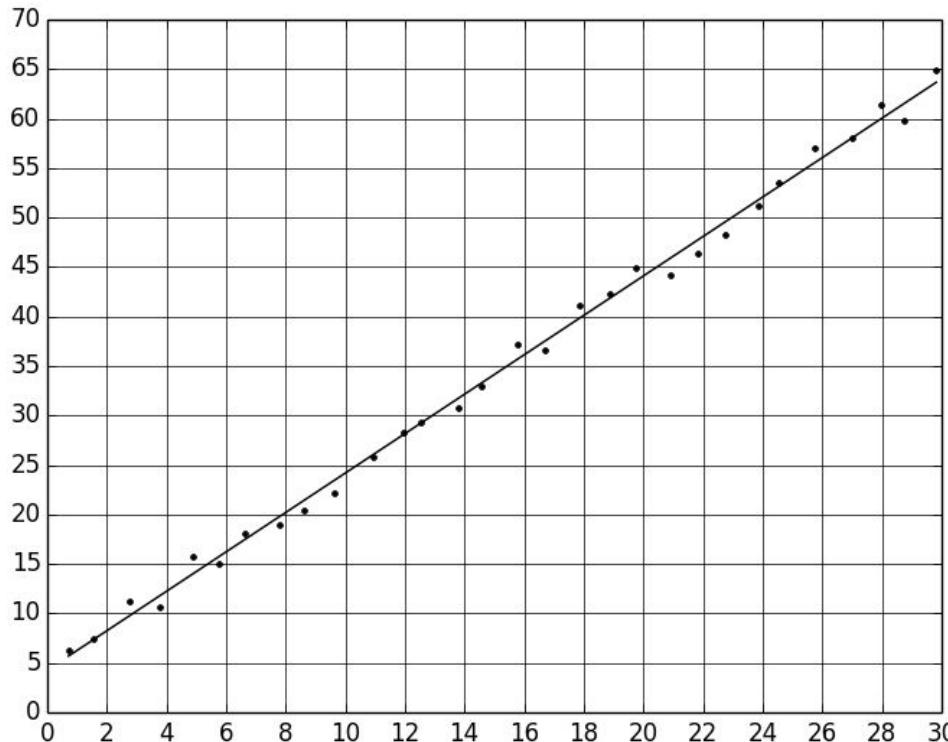


Redundant Data

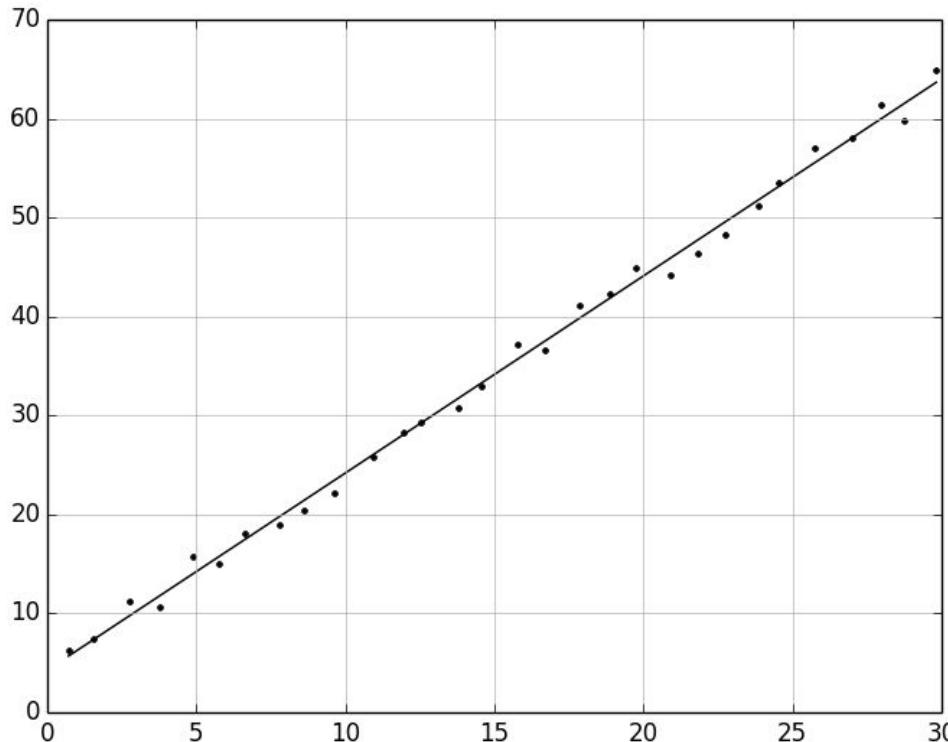


Metadata

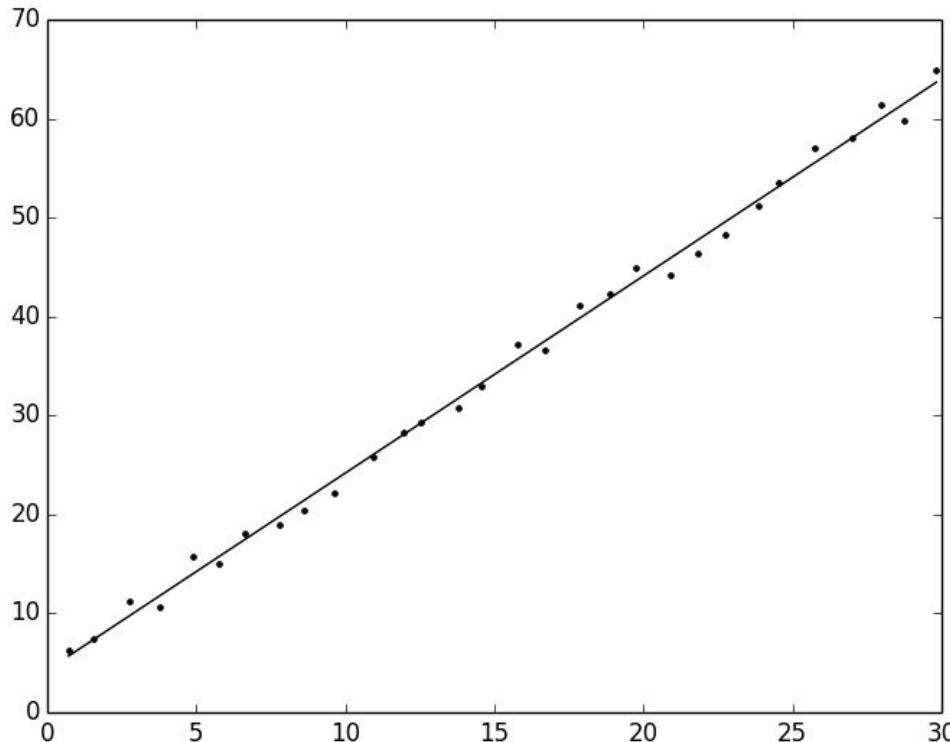
Metadata



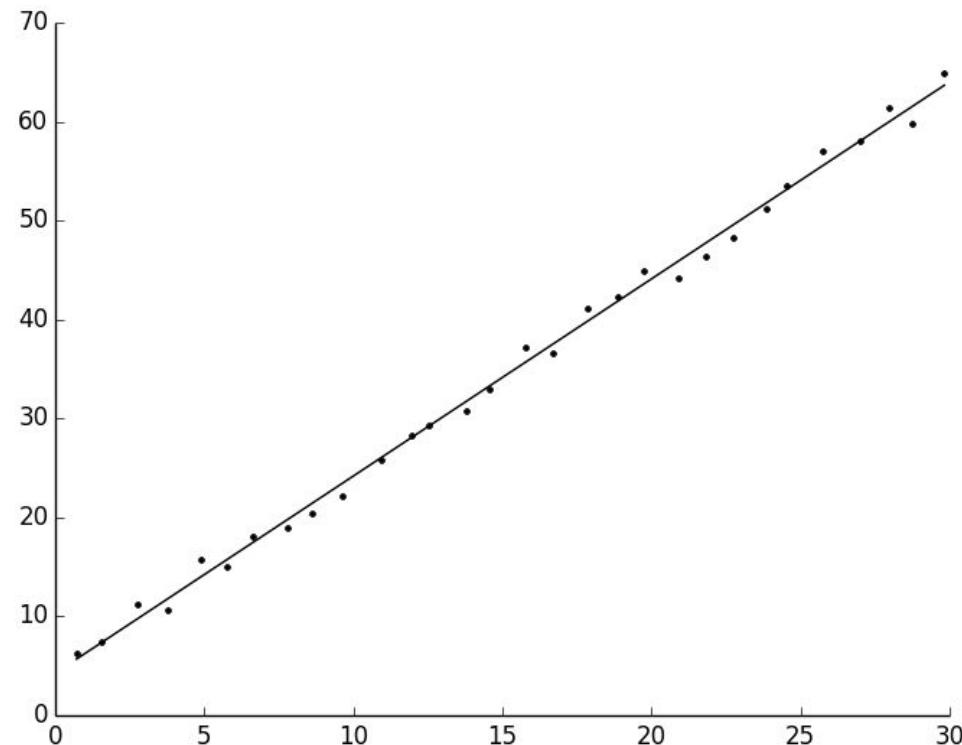
Metadata



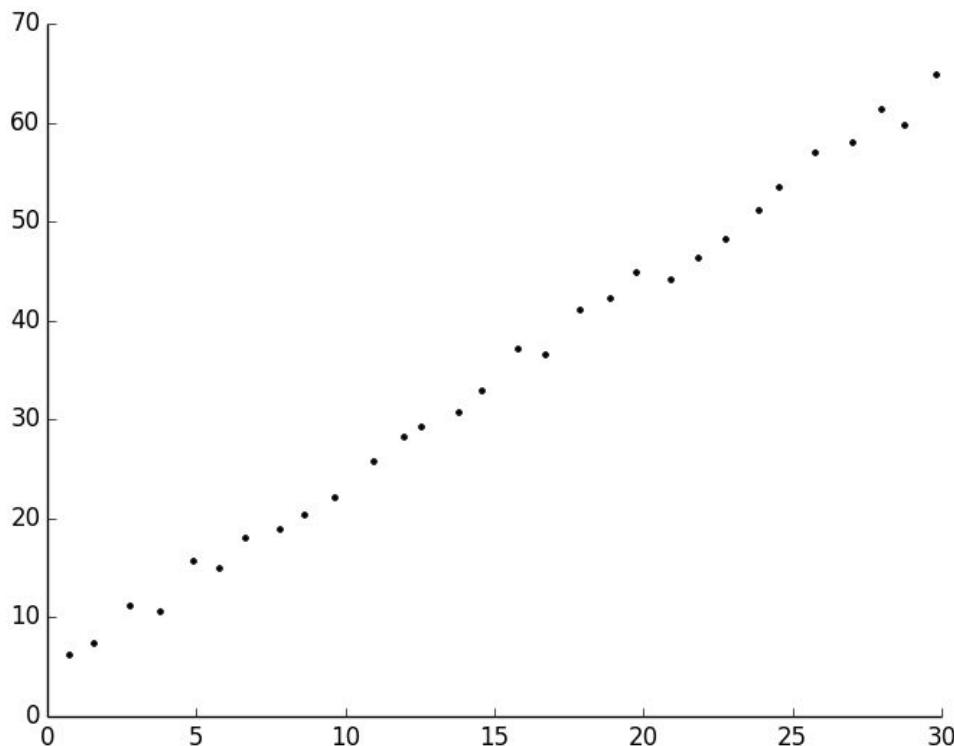
Metadata



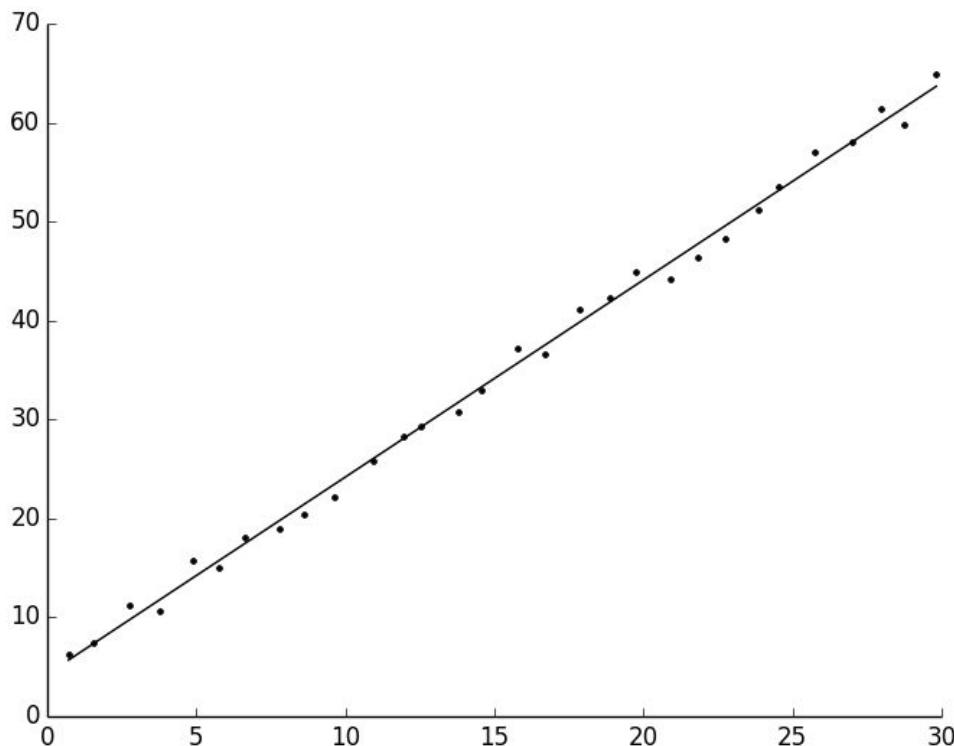
Metadata



Metadata

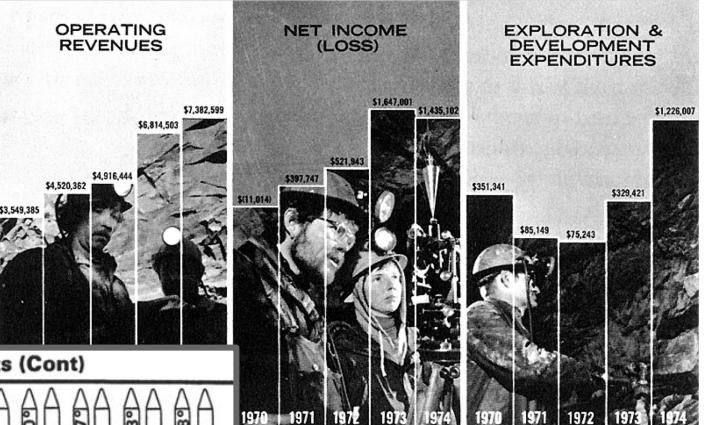
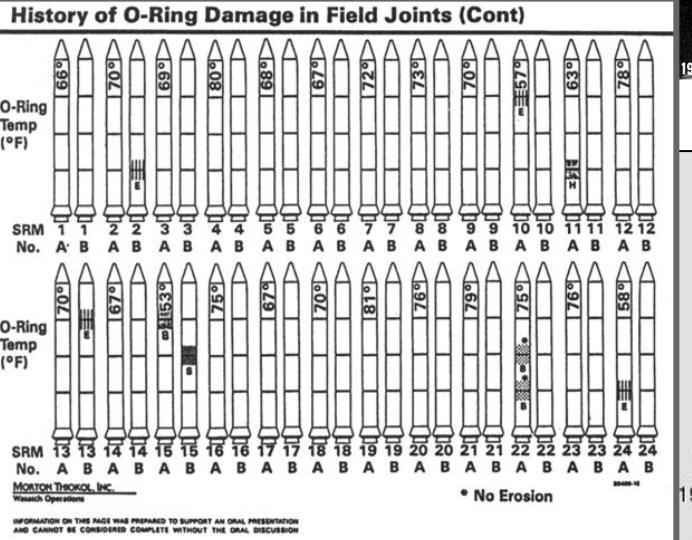


Metadata



Chartjunk

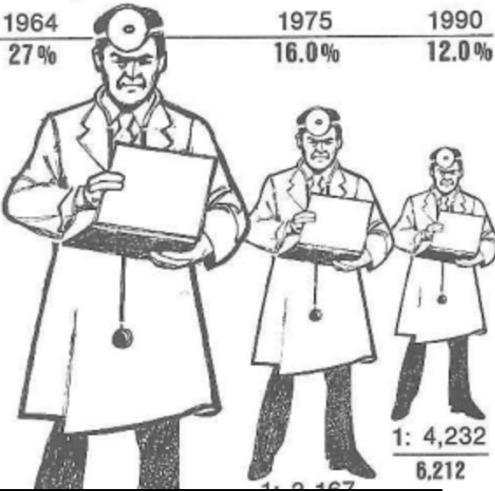
Chartjunk



THE SHRINKING FAMILY DOCTOR In California

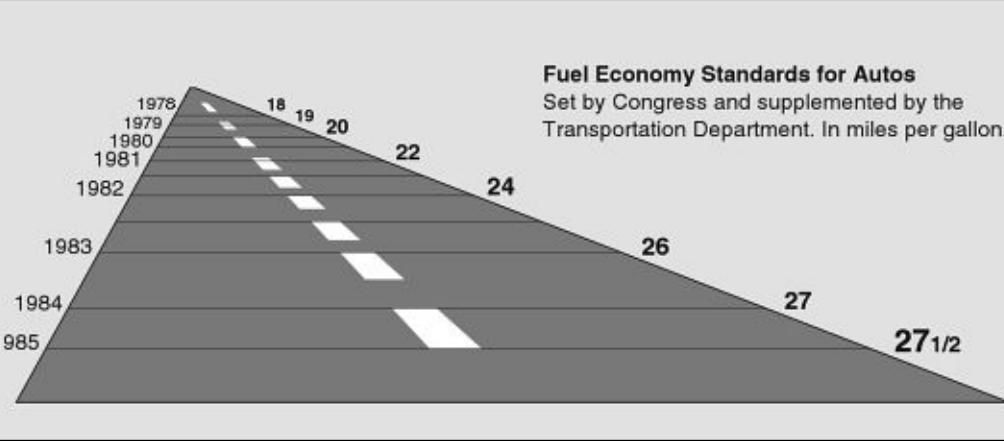
Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%



Fuel Economy Standards for Autos

Set by Congress and supplemented by the Transportation Department. In miles per gallon.



Improving charts

- Erase redundant data
 - Representing the same data multiple times
 - Within reason
- Erase metadata
 - Pixels giving context to the data
 - Within reason
- Iterate design
- Avoid chartjunk
 - Not necessary to understand the data
 - Distract the viewer

Orwell's Six Rules

“Never use a metaphor, simile, or other figure of speech which you are used to seeing in print.”

“Never use a long word where a short one will do.”

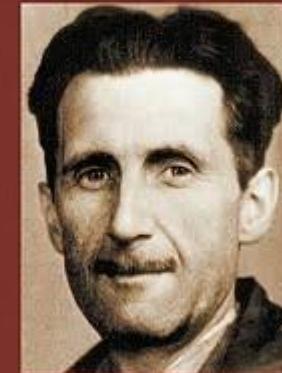
“If it is possible to cut a word out, always cut it out.”

“Never use the passive where you can use the active.”

“Never use a foreign phrase, a scientific word, or a jargon word if you can think of an everyday English equivalent.”

“Break any of these rules sooner than say anything outright barbarous.”

POLITICS AND THE
ENGLISH LANGUAGE
AND OTHER ESSAYS



George Orwell

Orwell's Six Rules

~~"Never use a metaphor, simile, or other figure of speech which you are used to seeing in print."~~

"Never use a long word where a short one will do."

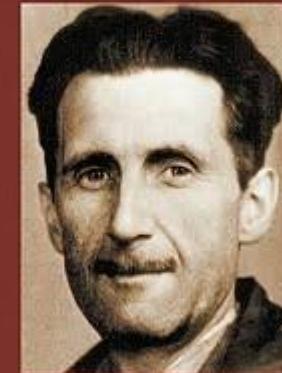
"If it is possible to cut a word out, always cut it out."

"Never use the passive where you can use the active."

"Never use a foreign phrase, a scientific word, or a jargon word if you can think of an everyday English equivalent."

"Break any of these rules sooner than say anything outright barbarous."

POLITICS AND THE
ENGLISH LANGUAGE
AND OTHER ESSAYS



George Orwell

Orwell's Six Rules

~~"Never use a metaphor, simile, or other figure of speech which you are used to seeing in print."~~

"Never use a long word where a short one will do."

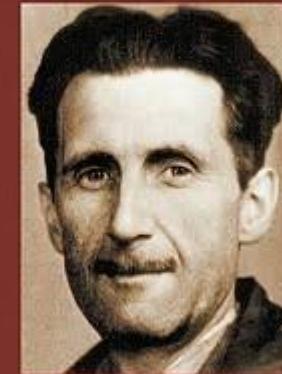
"If it is possible to cut a word out, always cut it out."

~~"Never use the passive where you can use the active."~~

"Never use a foreign phrase, a scientific word, or a jargon word if you can think of an everyday English equivalent."

"Break any of these rules sooner than say anything outright barbarous."

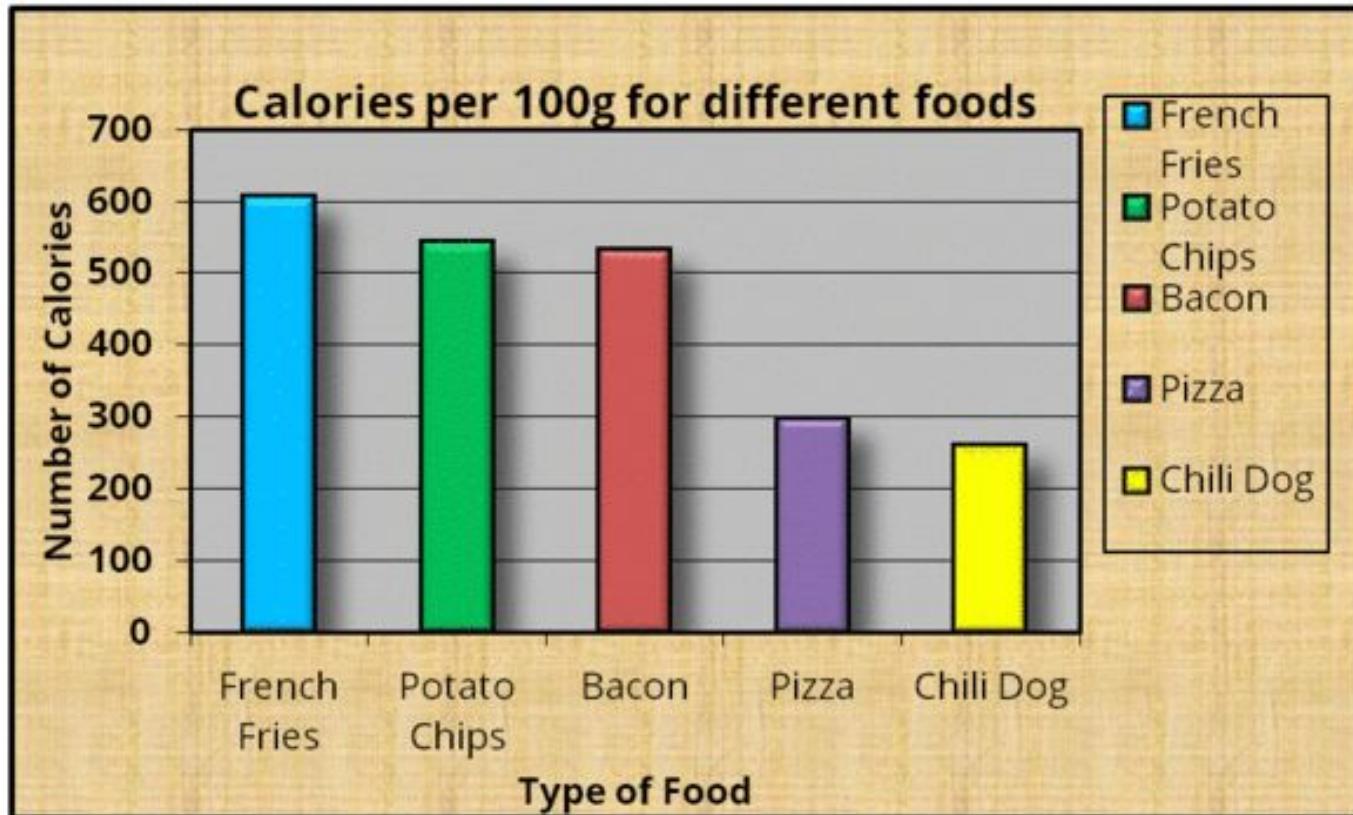
POLITICS AND THE
ENGLISH LANGUAGE
AND OTHER ESSAYS



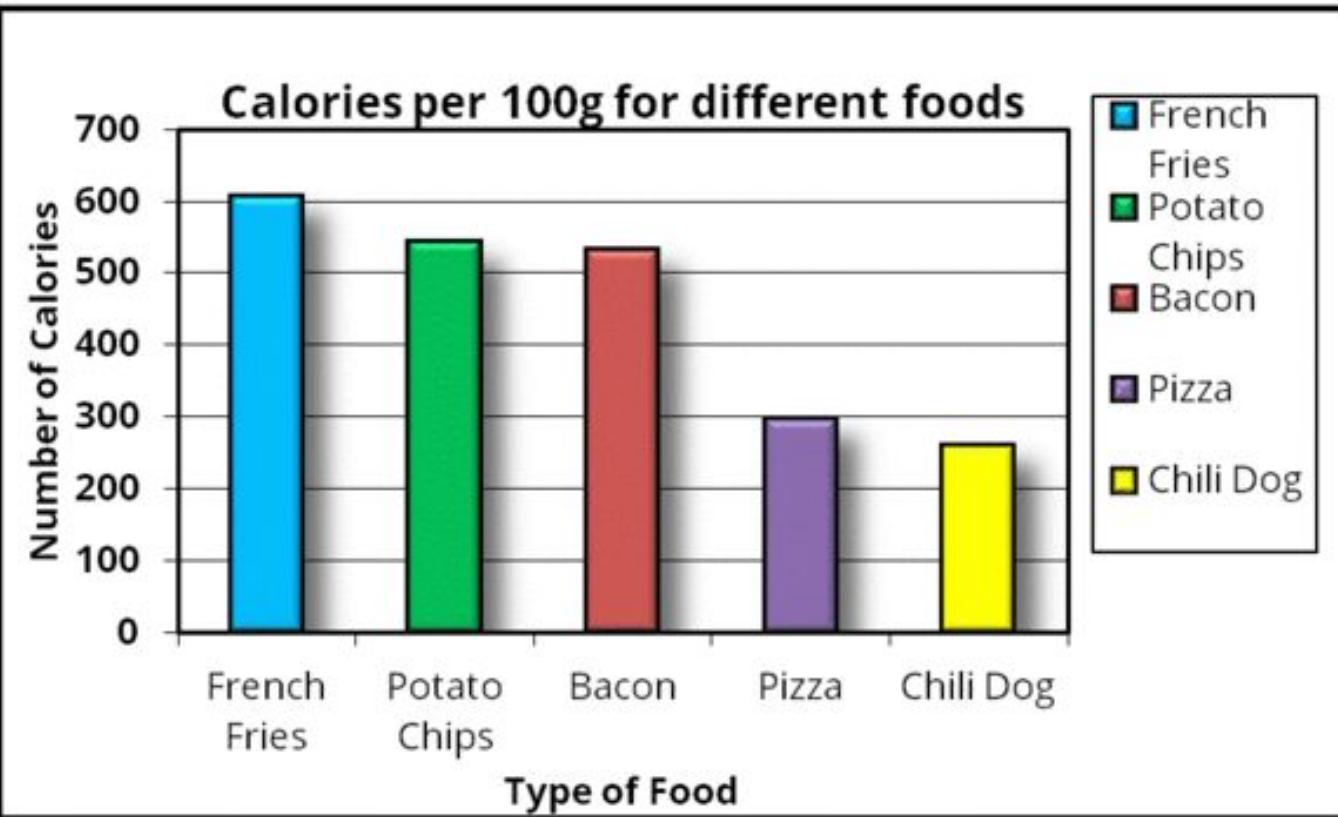
George Orwell

Example of Iteration

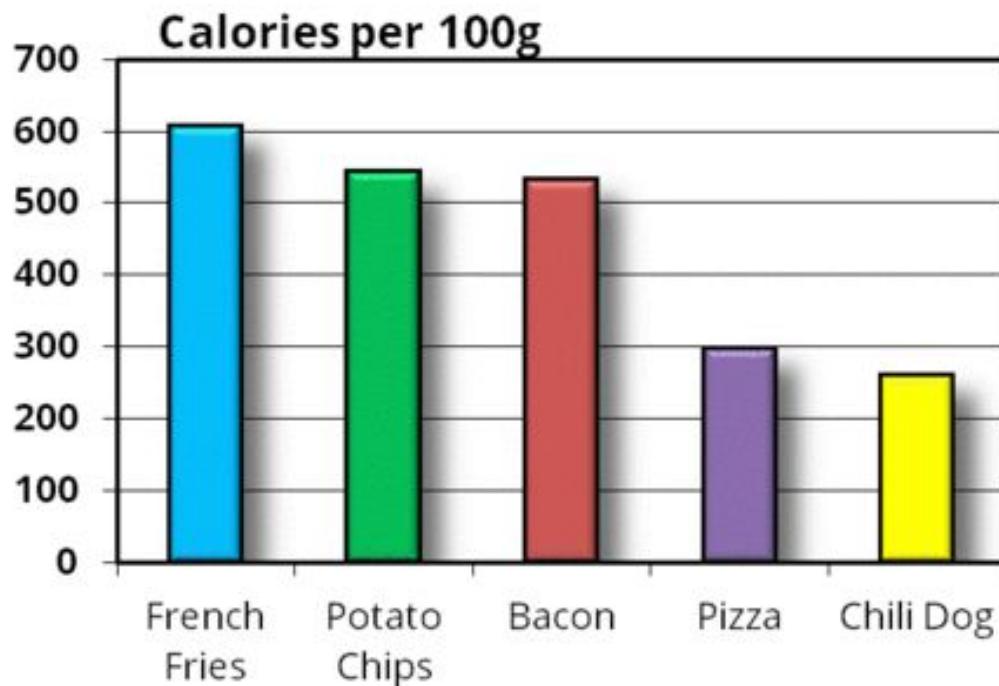
Story: Putting Bacon in Context



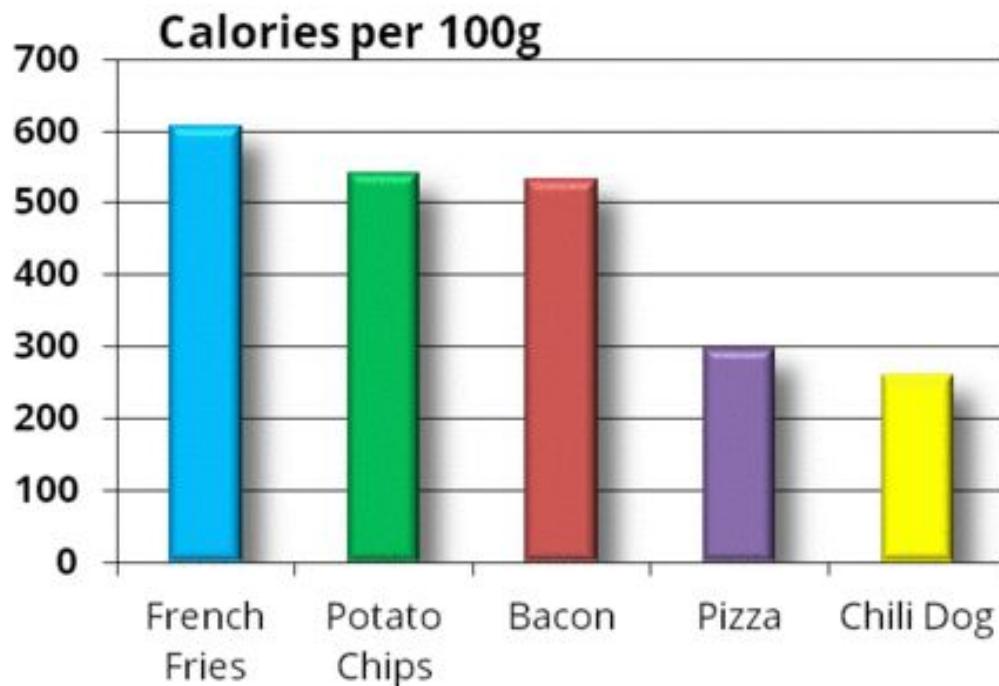
Remove backgrounds



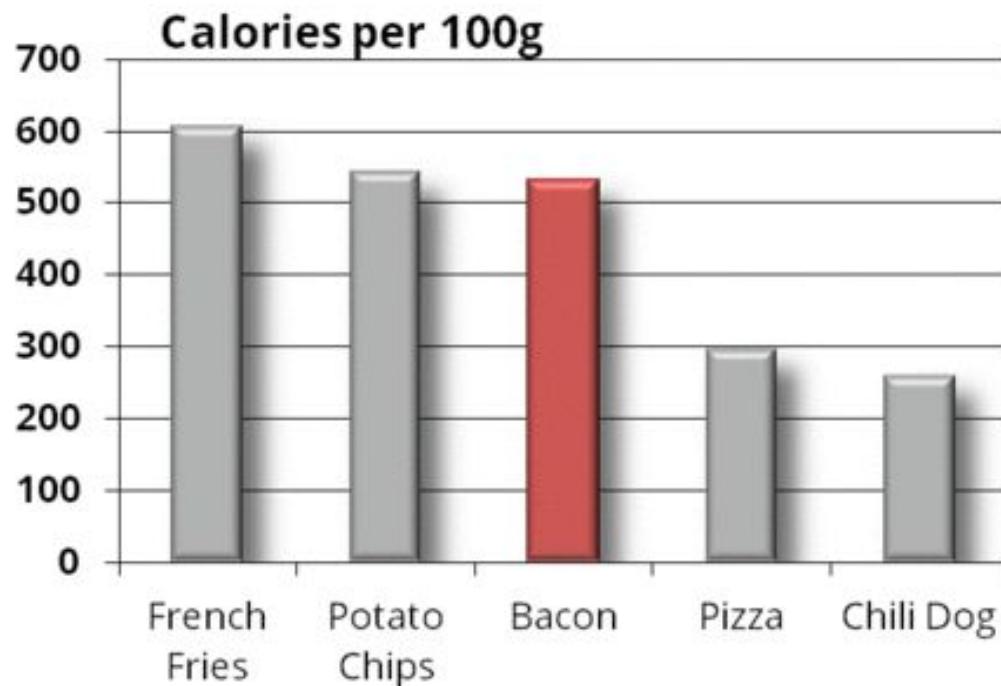
Remove redundant labels



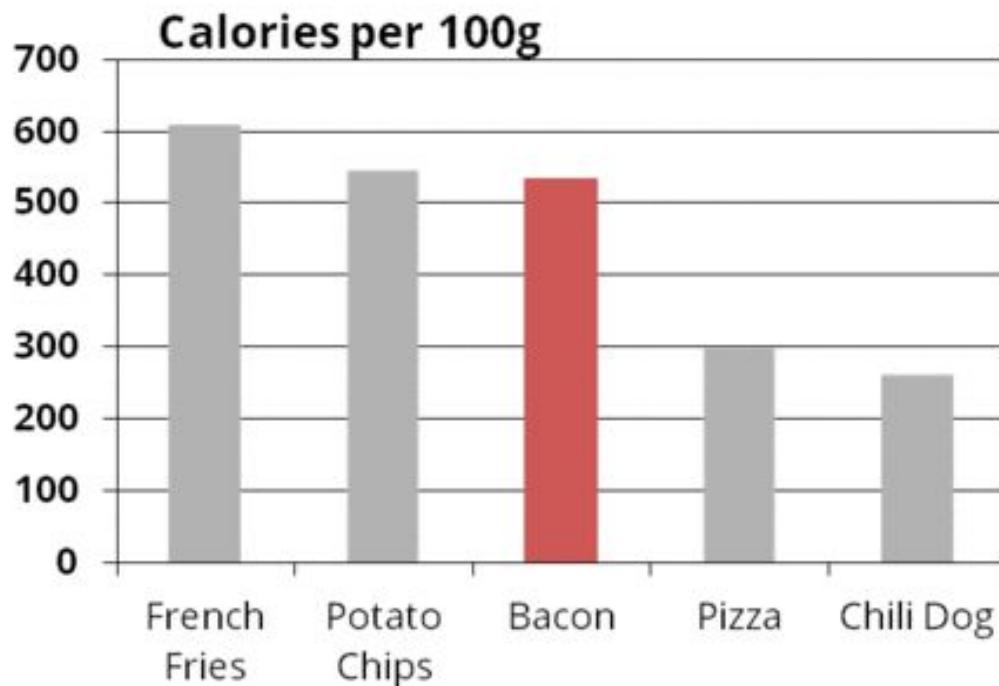
Remove borders



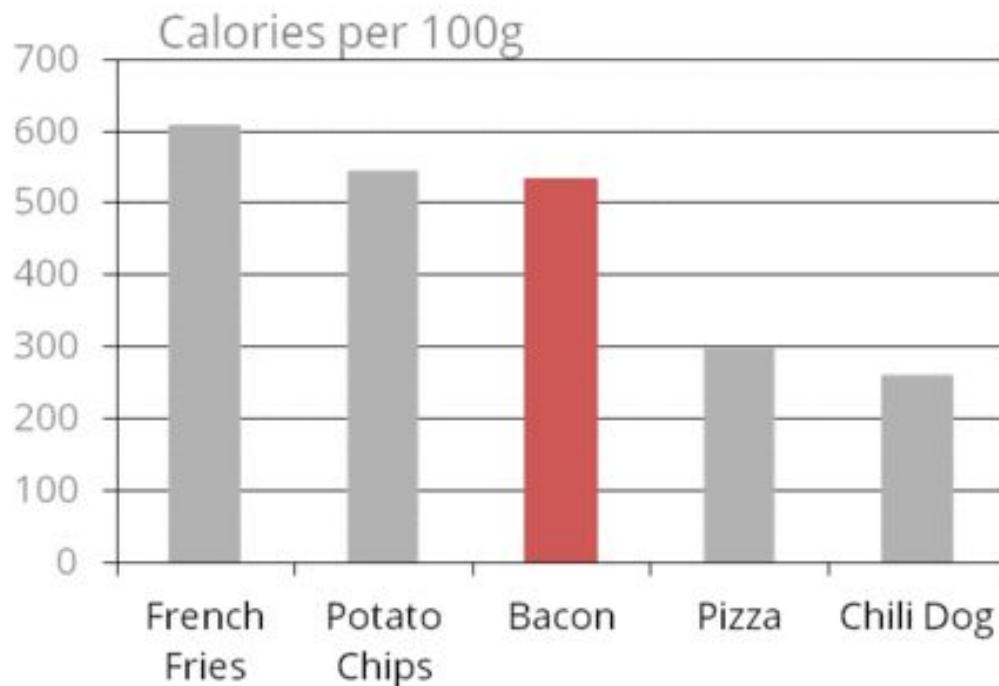
Reduce colors



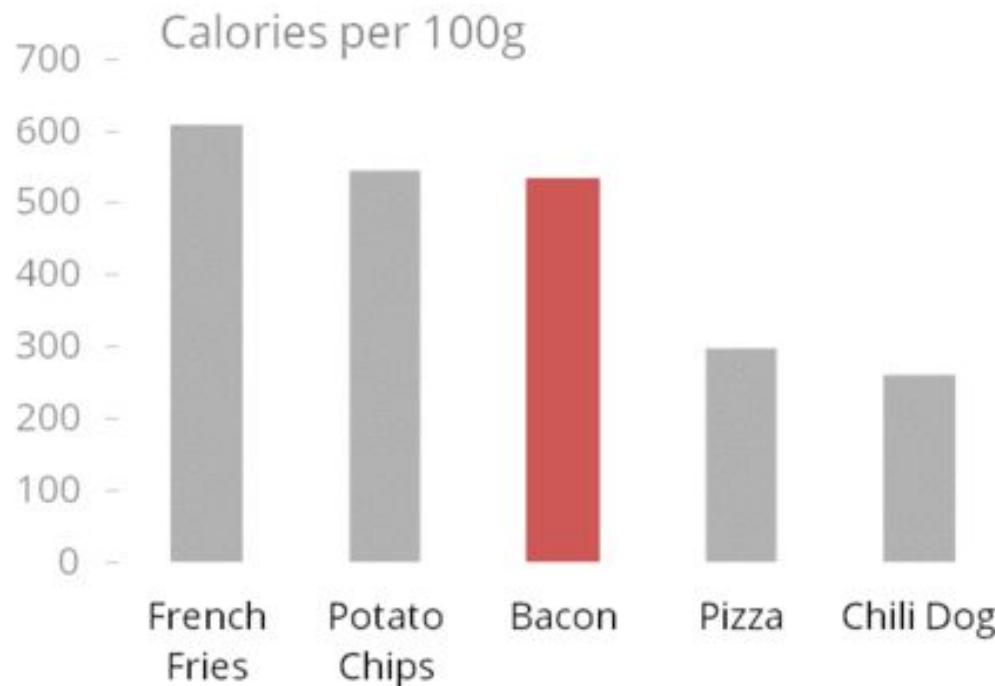
Remove special effects



Lighten labels

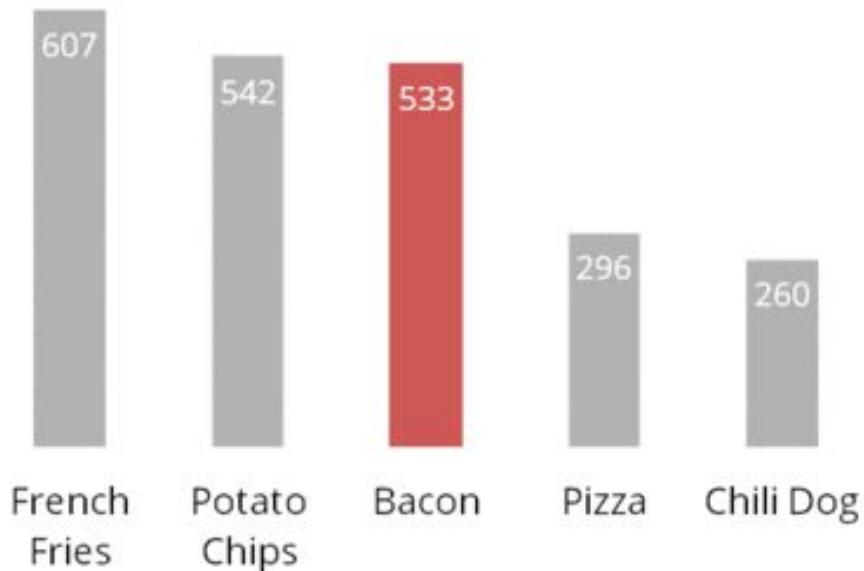


Or remove lines

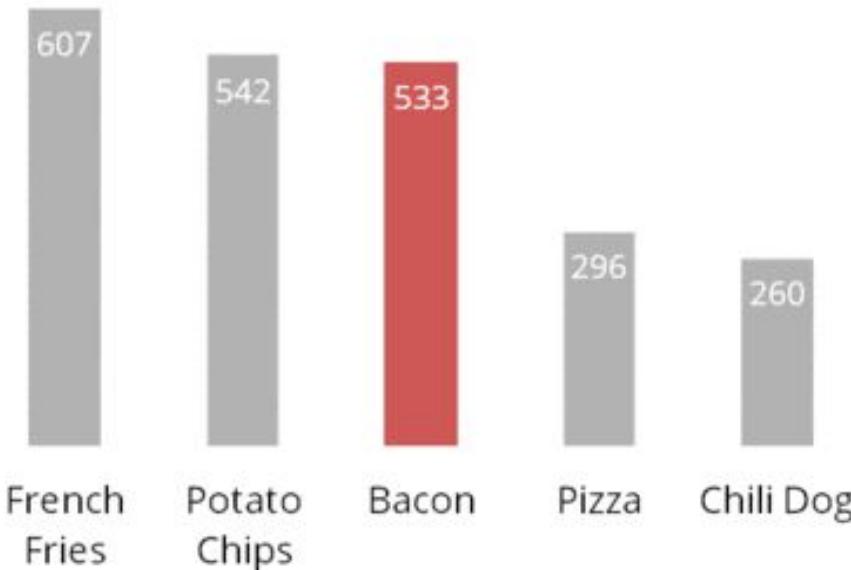


Direct label

Calories per 100g



Calories per 100g



Using Attentive Cognition

Attentive Processing

1904727116848316516

0806174557061387374

1548311125468098808

9323343870208212744

1713102036938742890

Preattentive Processing

1904727116848316516

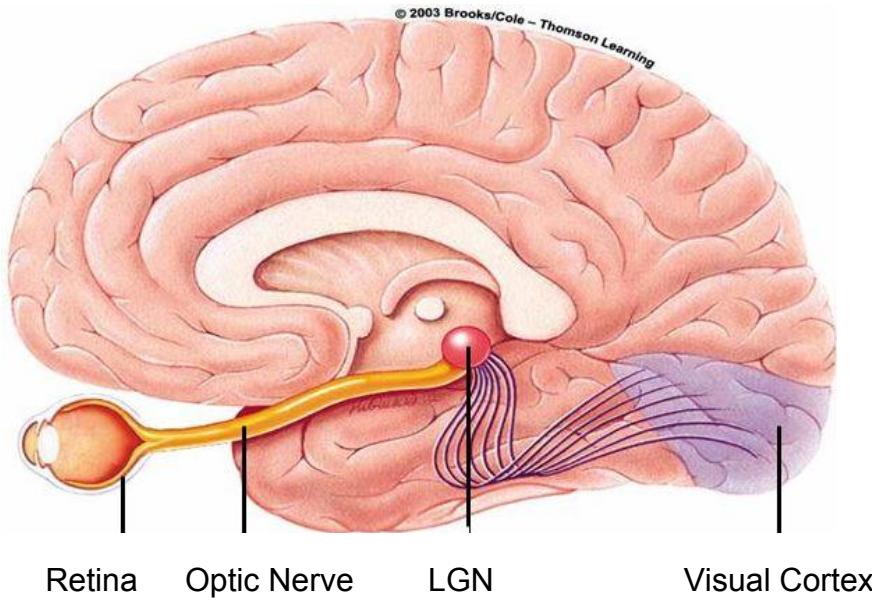
0806174557061387374

1548311125468098808

9323343870208212744

1713102036938742890

Visual Processing System

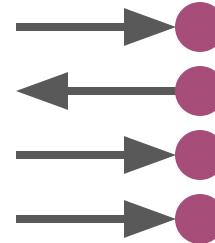


Preattentive Stimuli

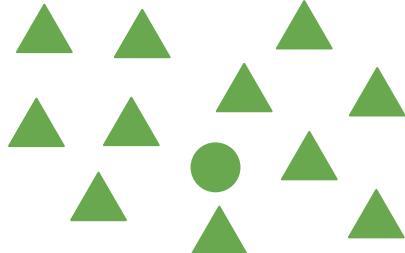
Color



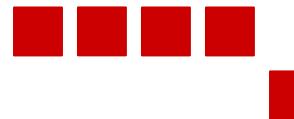
Movement



Form



Position



Preattentive Stimuli

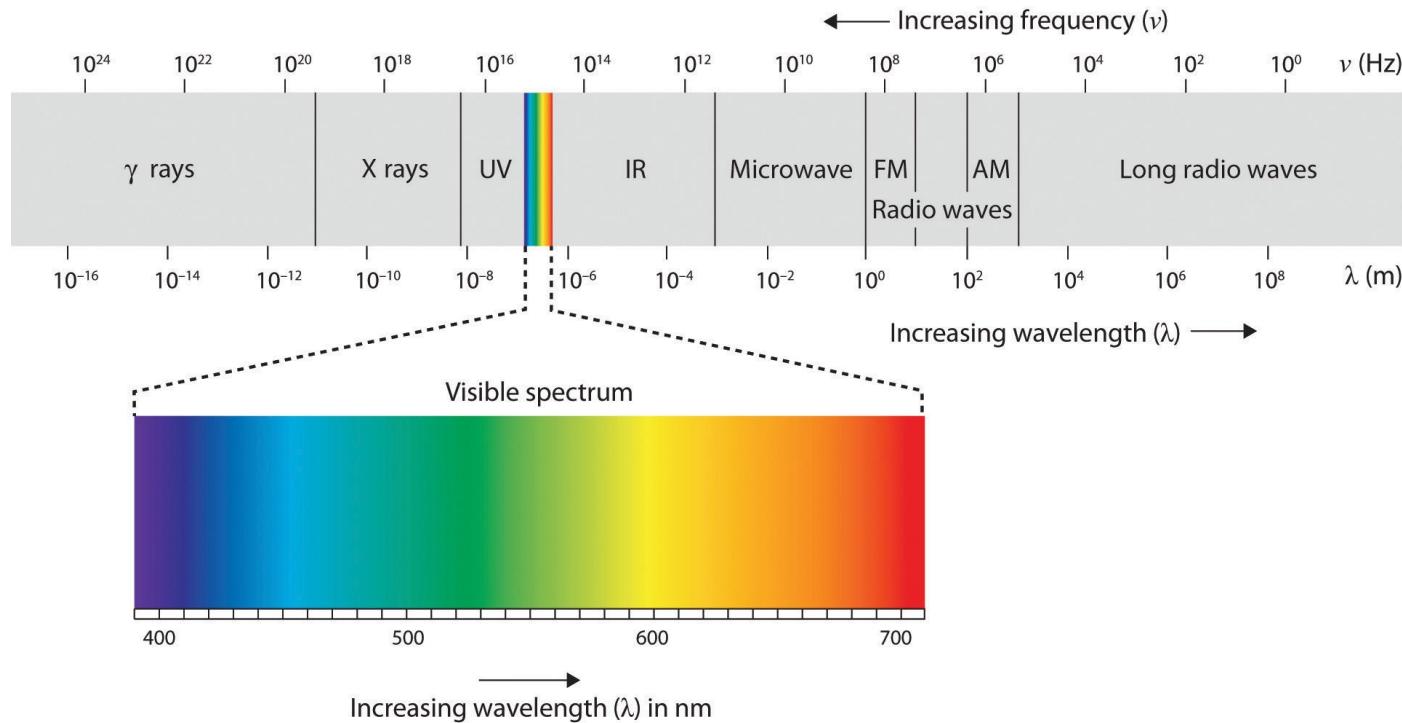
- Interactive Applications
 - <http://learnforeverlearn.com/preattentive/>
- More on theory and types
 - http://www.perceptualedge.com/articles/ie/visual_perception.pdf
- Applications to computer vision
 - <https://people.eecs.berkeley.edu/~malik/papers/malik-perona90.pdf>

Post-Preattentive Takeaways

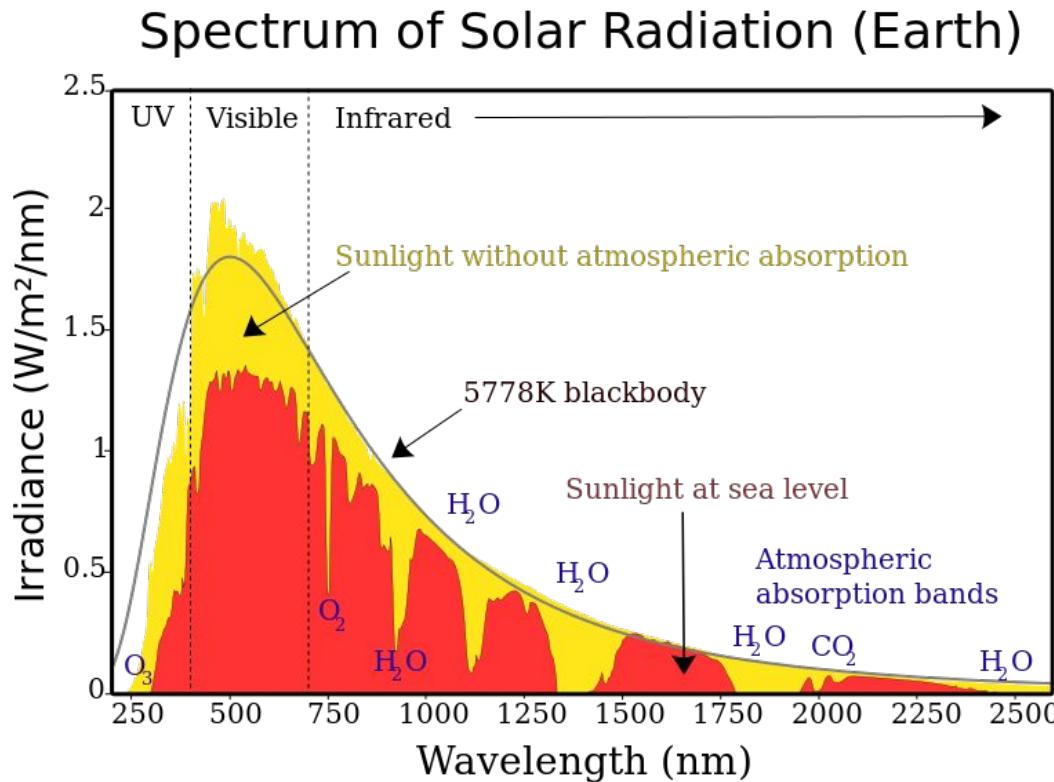
- Draw the viewer's attention
- Don't distract the viewer

Visual Encoding Using Color

Physics of Color



Physics of Color



What Not To Do With Color: Physics

Elevation is quantitative (-4000m to 1500m) ...

Wavelength is quantitative (400nm to 750nm)

So...

What Not To Do With Color: Physics

Elevation is quantitative (-4000m to 1500m) ...

Wavelength is quantitative (400nm to 750nm)

So...

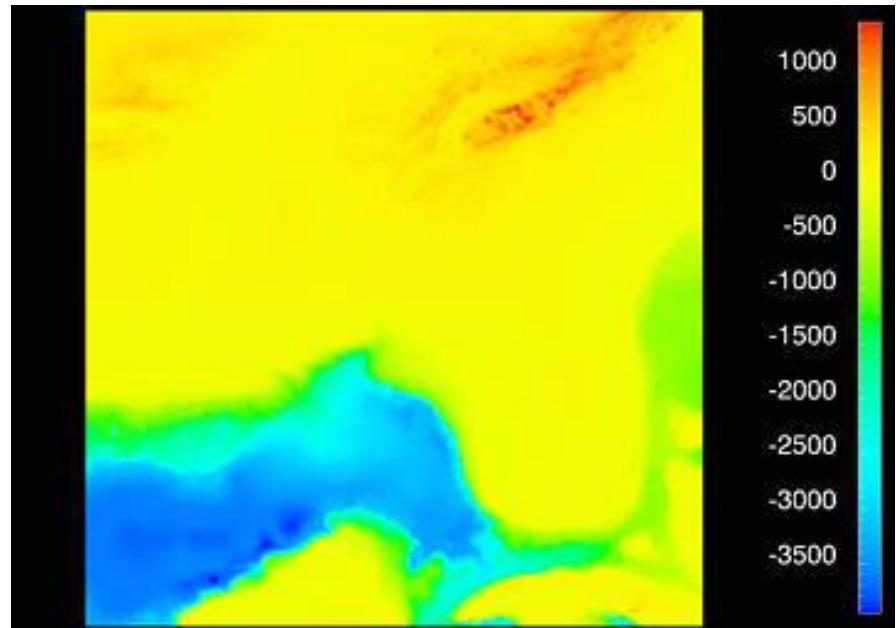


What Not To Do With Color: Physics

Elevation is quantitative (-4000m to 1500m) ...

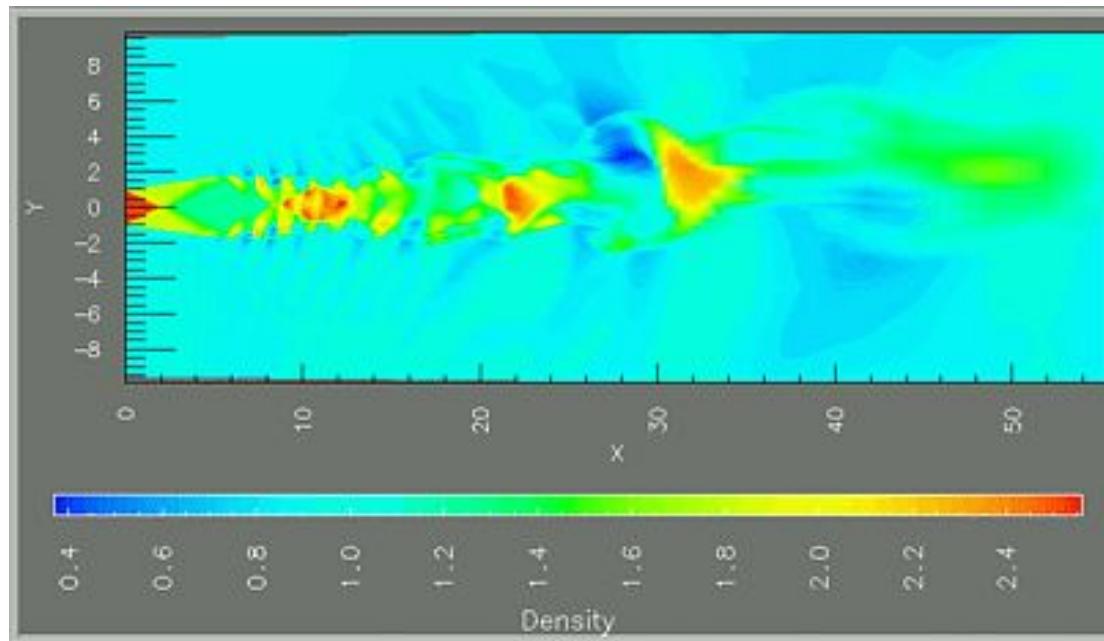
Wavelength is quantitative (400nm to 750nm)

So...



Jet / Rainbow Colormap Considered Harmful

Yet it's the default behavior for matplotlib

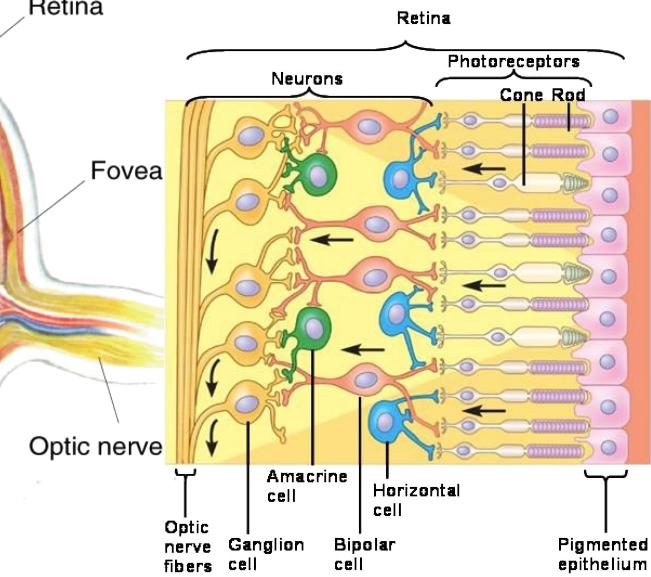
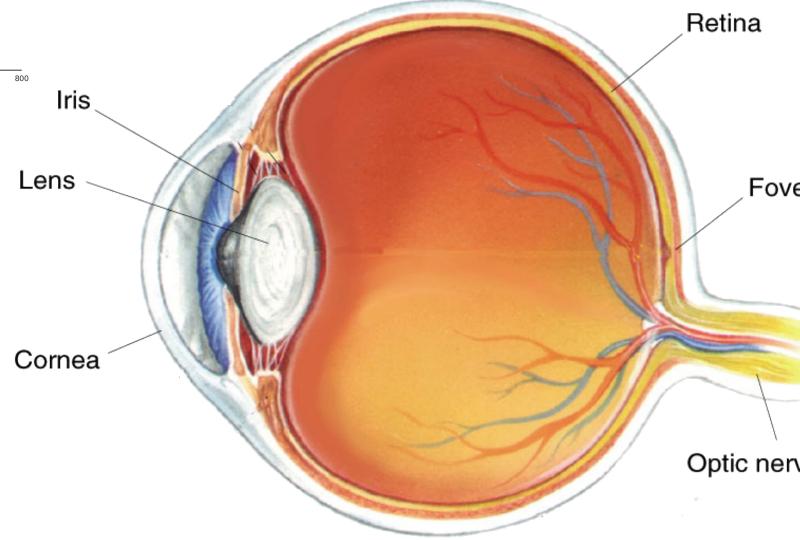
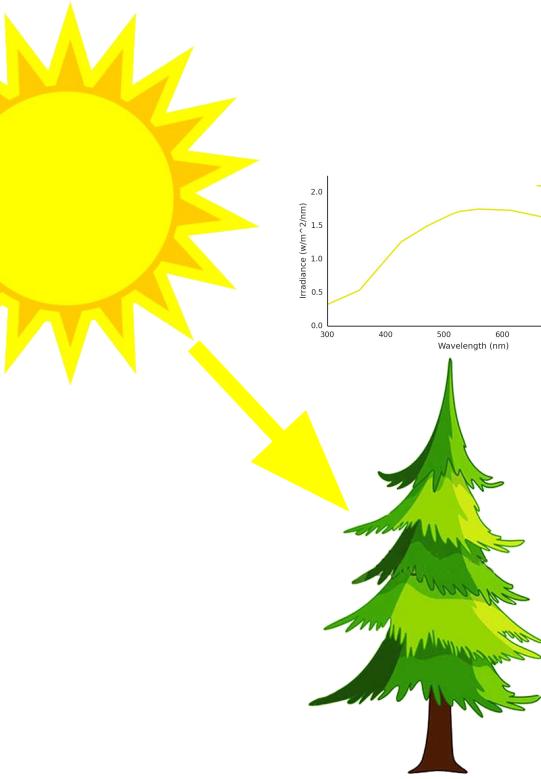


Perception

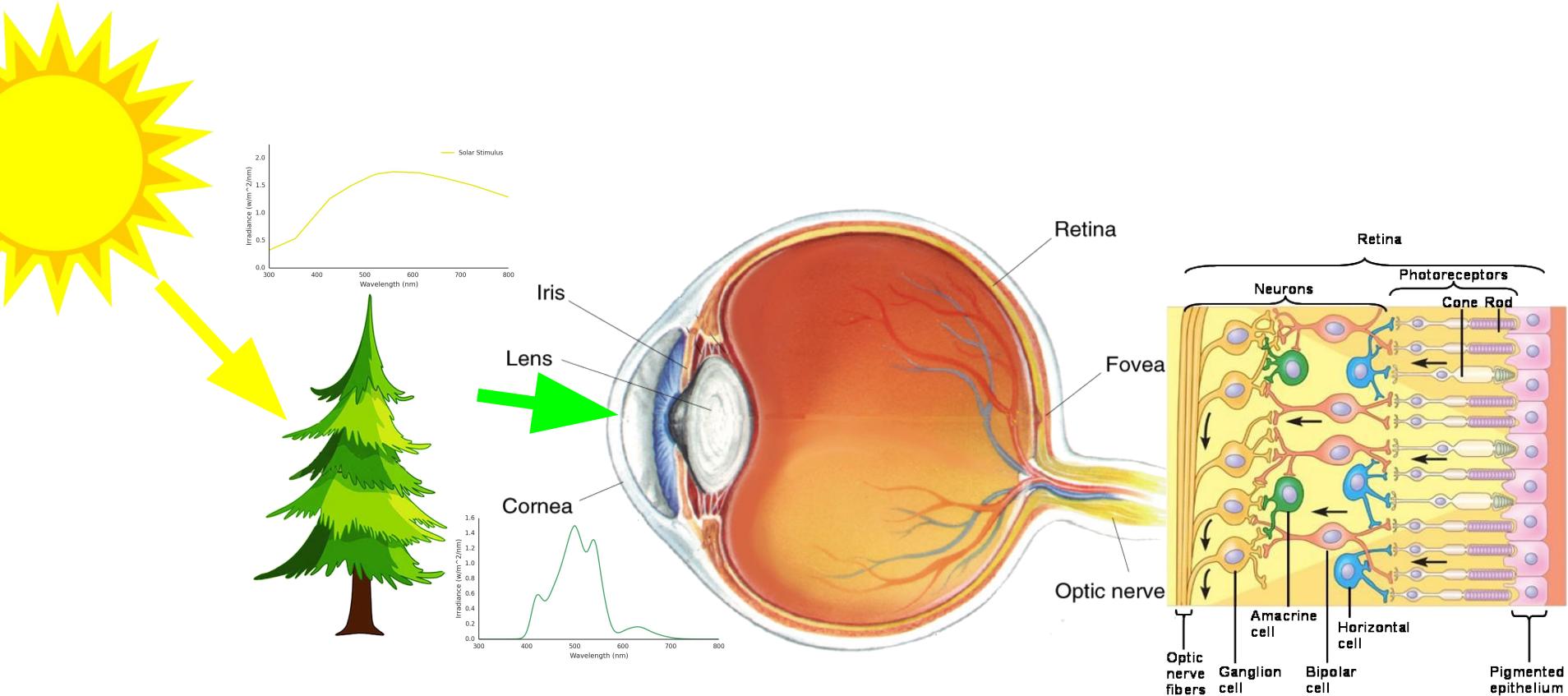
(it's better than reality)

Human Color Vision: Tristimulus Theory

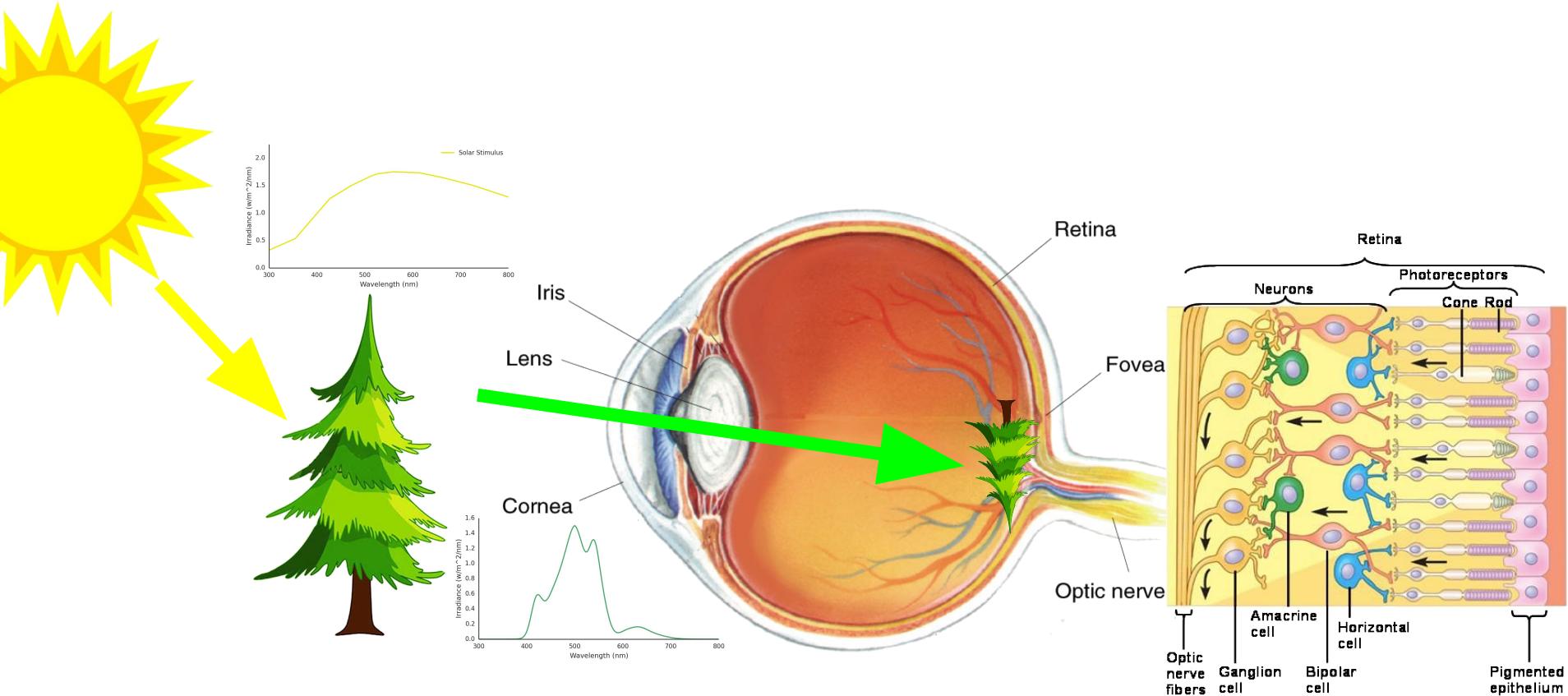
What To Do With Color: Perception

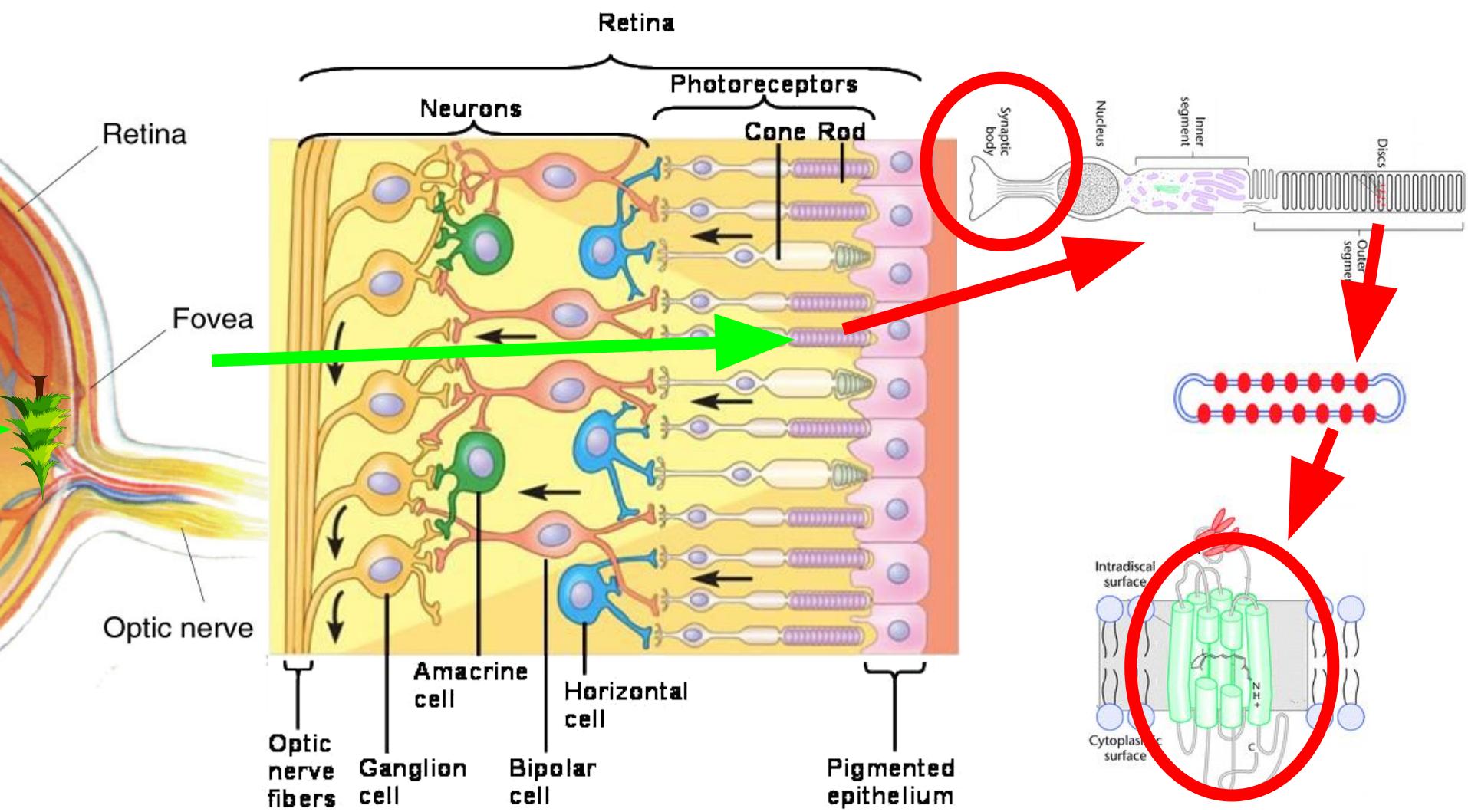


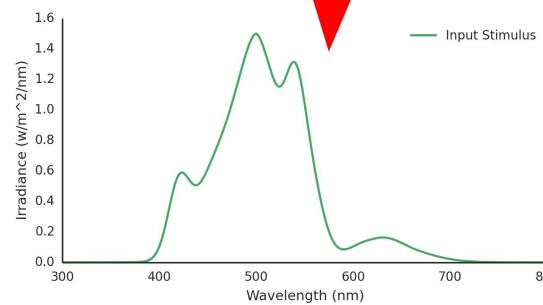
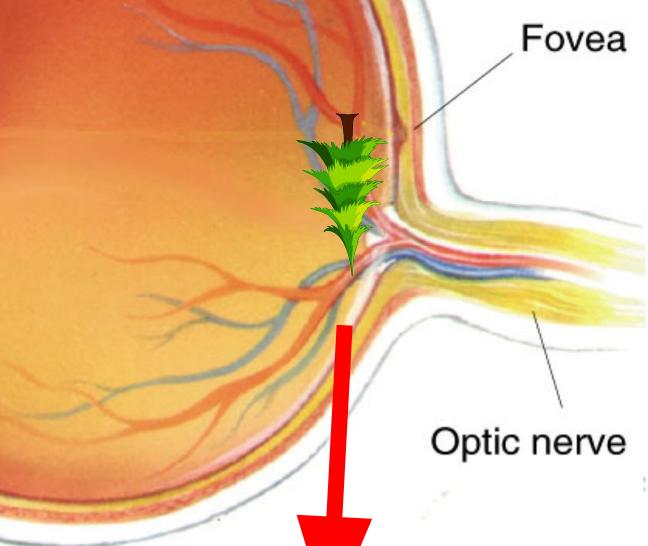
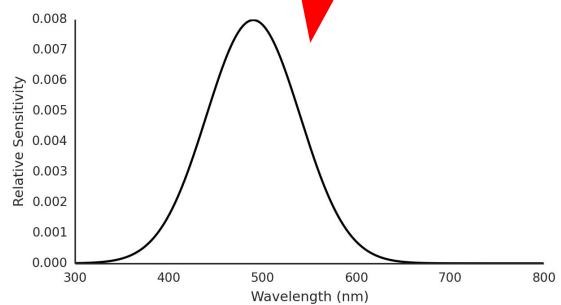
What To Do With Color: Perception



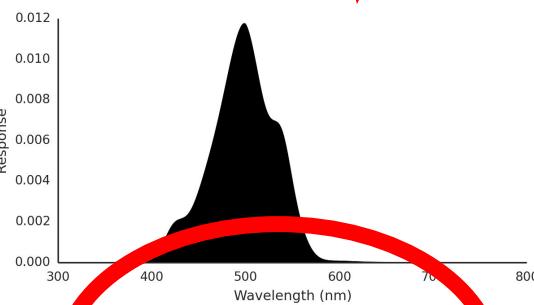
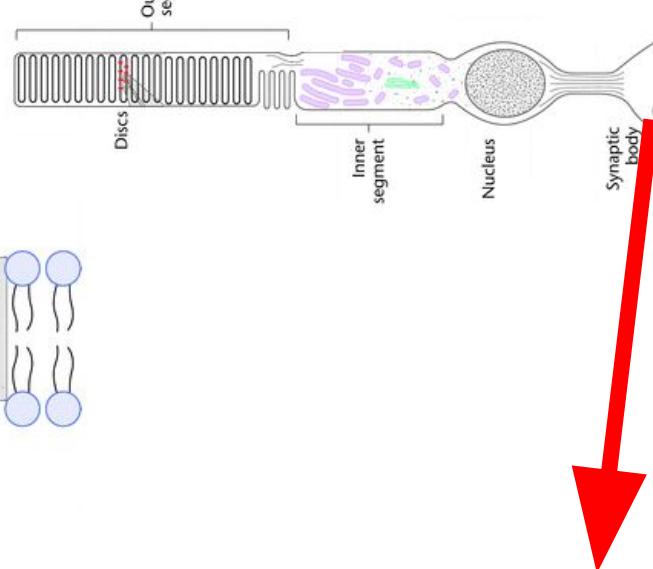
What To Do With Color: Perception






 Φ

 R

Stimulus Value



$$\int_{\Lambda} R \Phi$$

Photoreceptors

Rods (RHO): 490 nm
Chromosome 3

S cones (OPN1SW): 420 nm
Chromosome 7

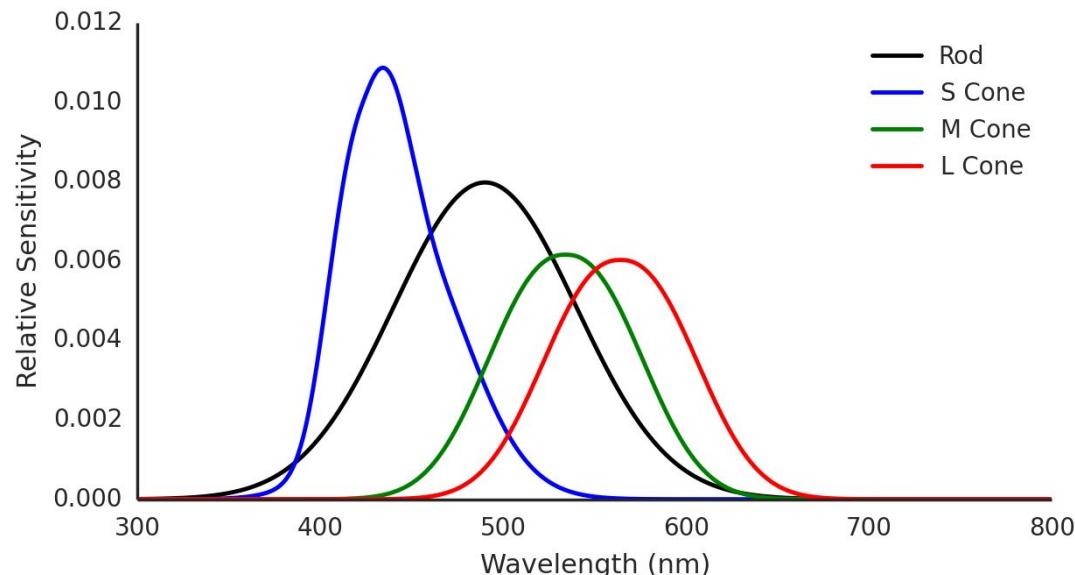
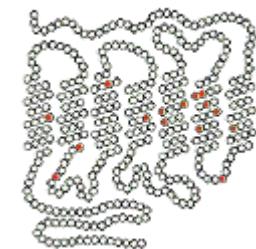
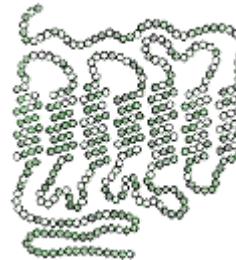
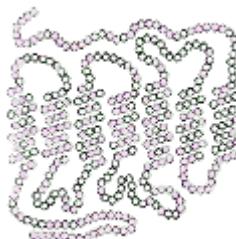
M cones (OPN1MW): 534 nm
Chromosome 23 (X only)

L cones (OPN1LW): 564 nm
Chromosome 23 (X only)

S

M

L



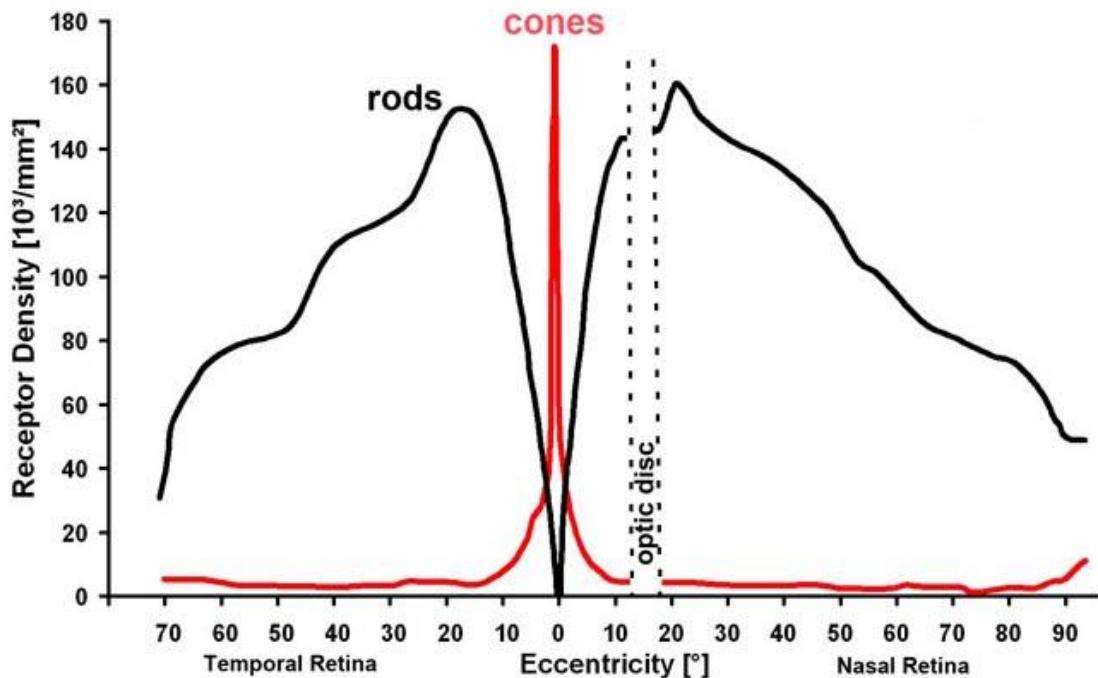
Rod and Cone Domains

Cones

- High light intensity
- Near the fovea

Rods:

- Low light intensity
- Away from the fovea



Cone Stimulation

S cones: 420 nm

Chromosome 7

M cones: 534 nm

Chromosome 23 (X only)

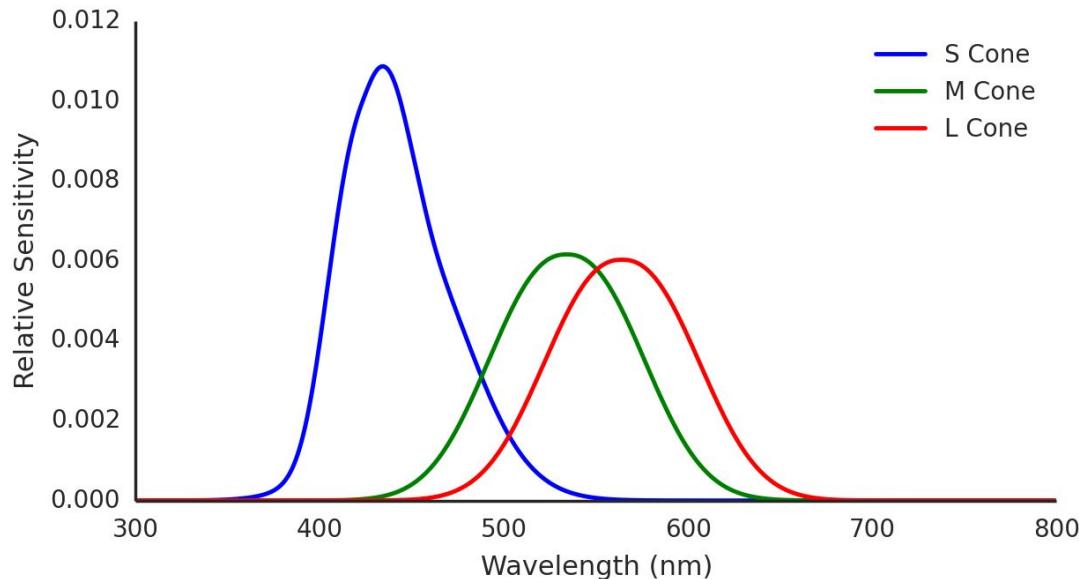
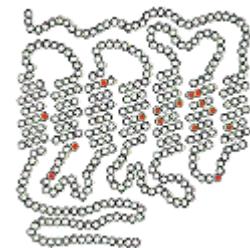
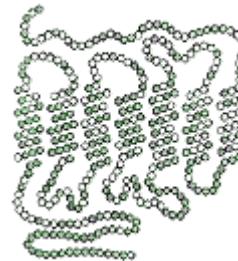
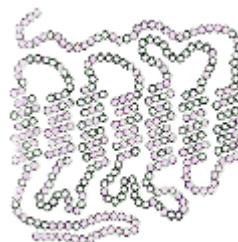
L cones: 564 nm

Chromosome 23 (X only)

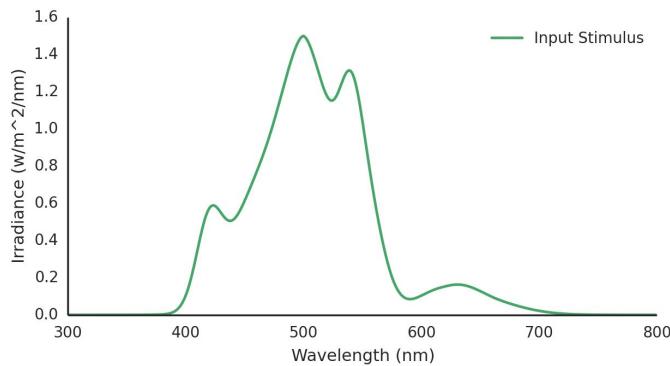
S

M

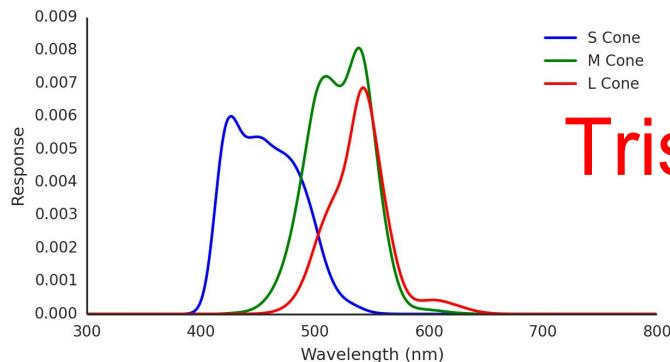
L



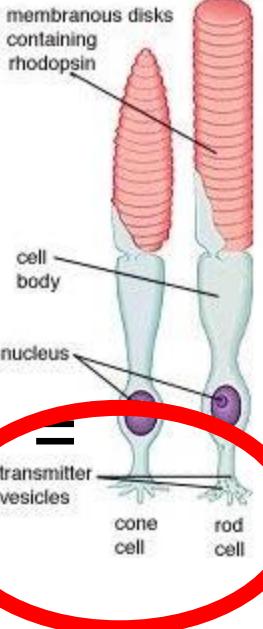
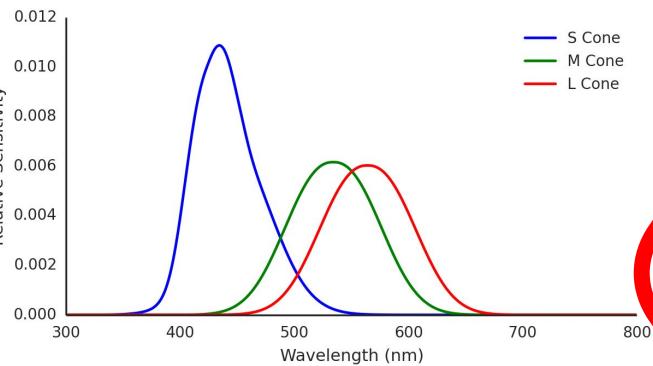
Tristimulus (LMS) Values



X



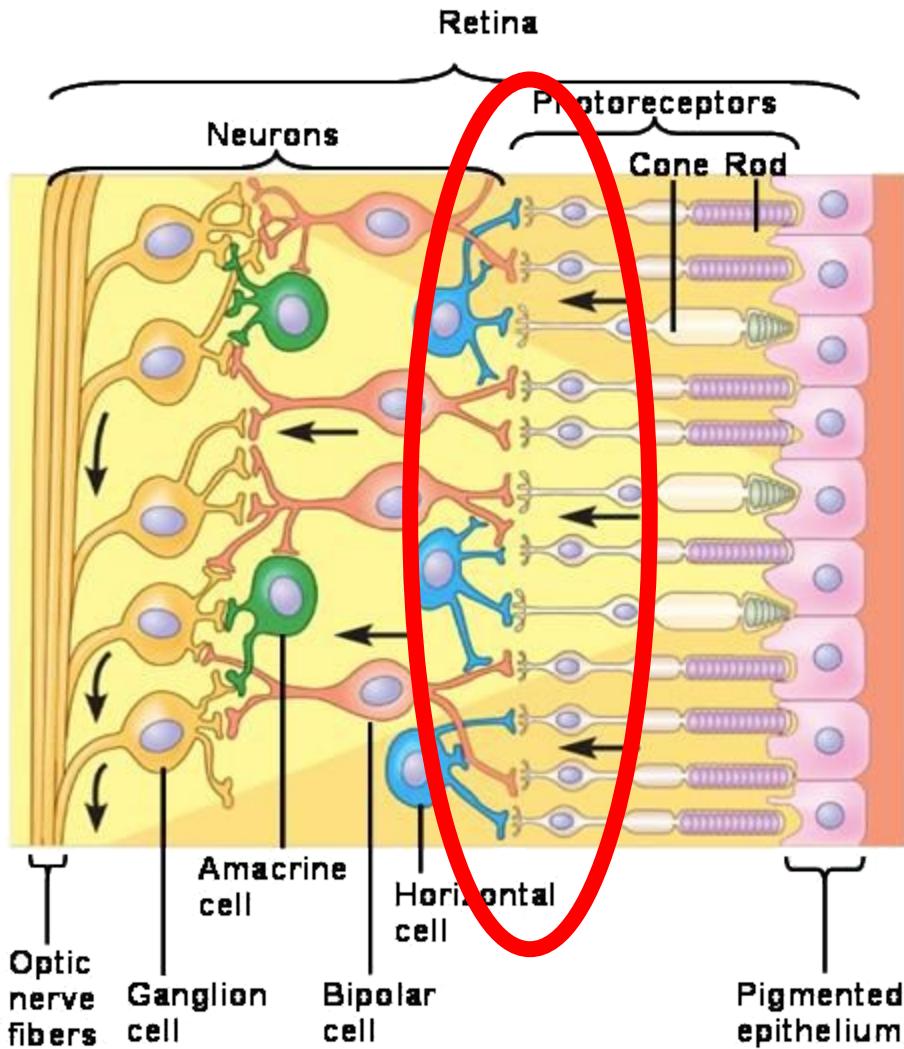
Tristimulus
Value



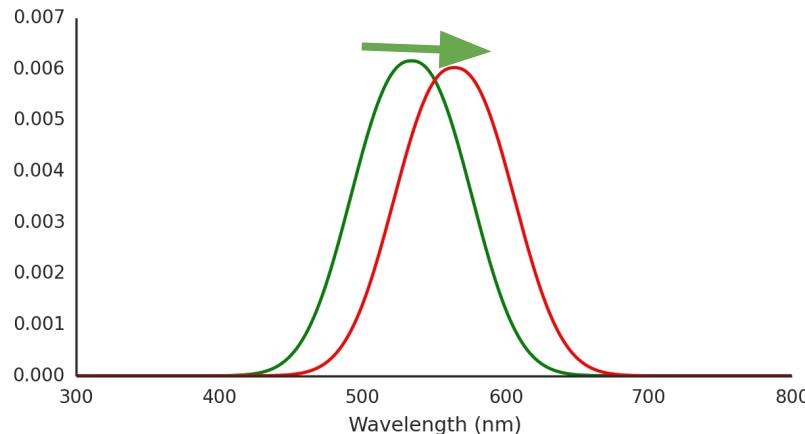
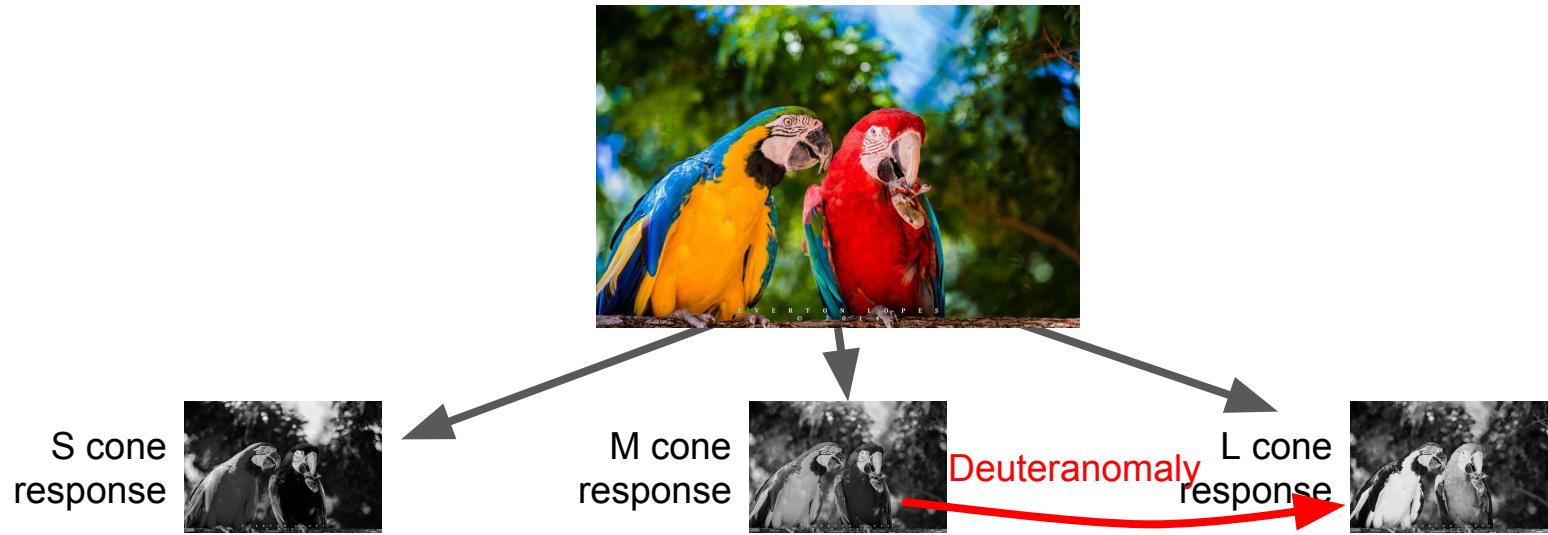
$$S \text{ stimulus} = \int_{\Lambda} S \Phi$$

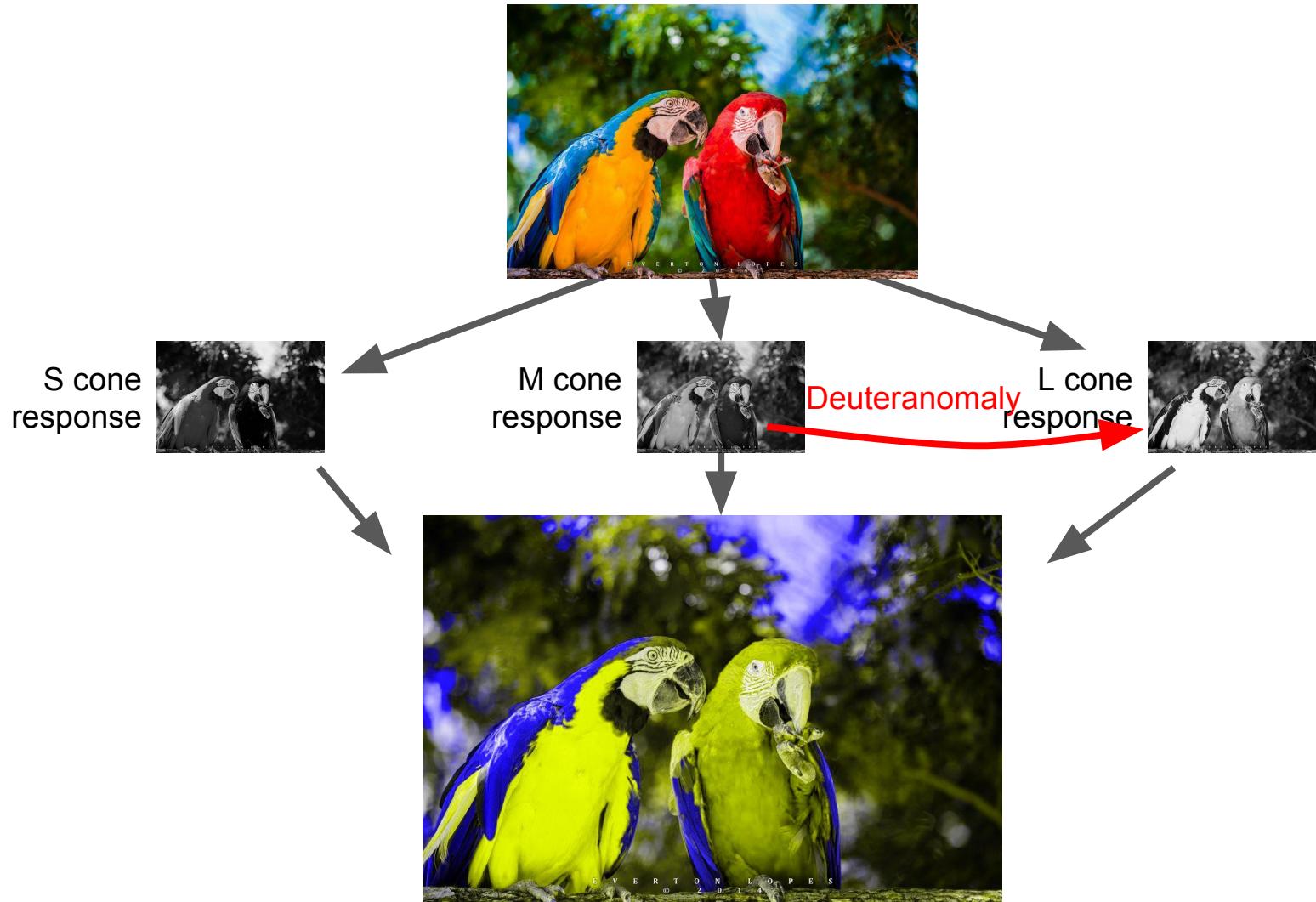
$$M \text{ stimulus} = \int_{\Lambda} M \Phi$$

$$L \text{ stimulus} = \int_{\Lambda} L \Phi$$

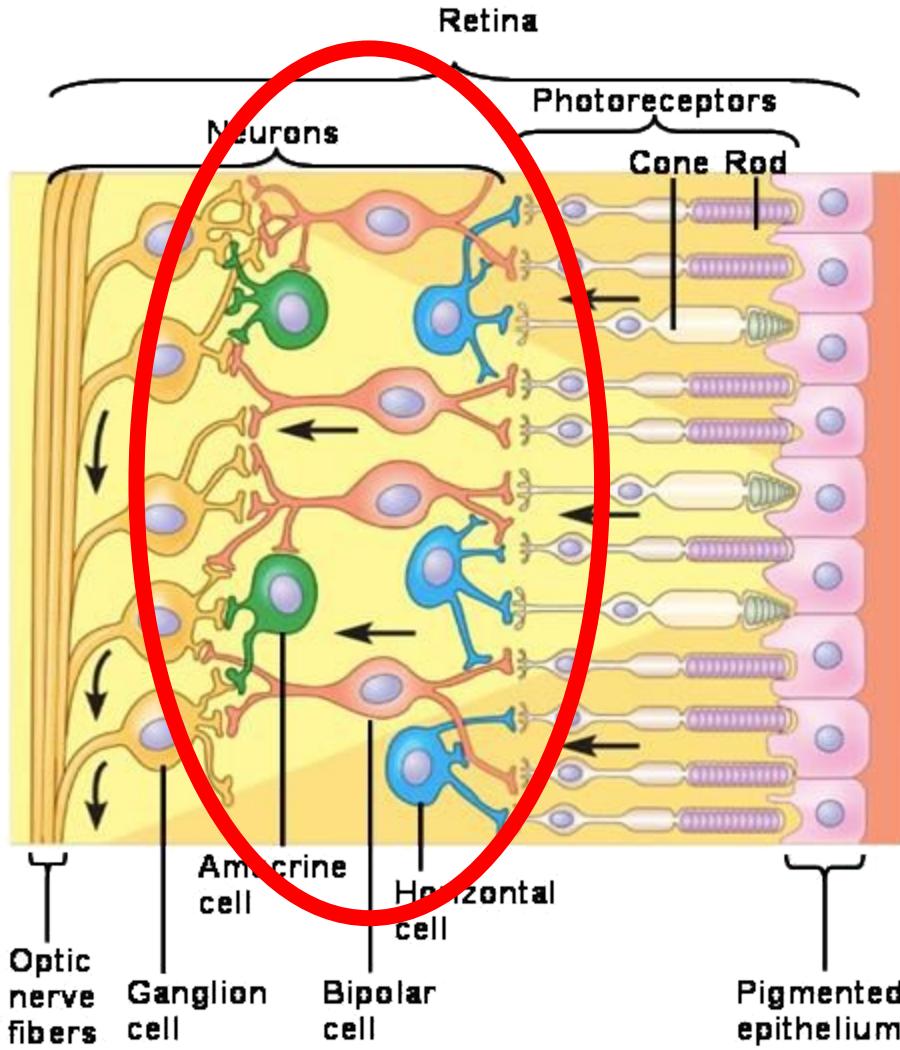


Colorblindness



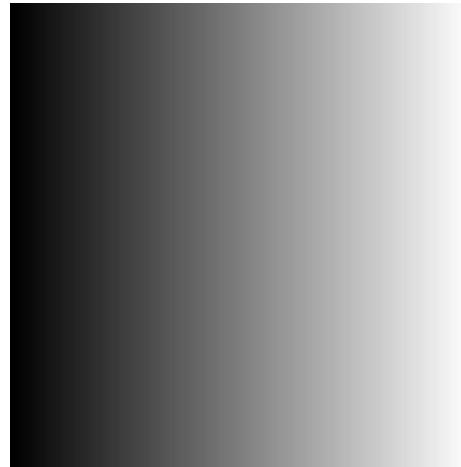


Human Color Vision: Opponent Process Theory



Color Before the Brain

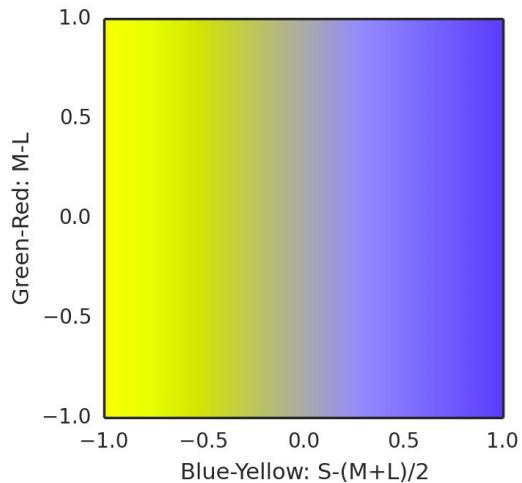
Luminance = S + M + L (+ rods)



Color Before the Brain

Luminance = S + M + L (+ rods)

Blue-Yellow opposition = $S - (M + L)$

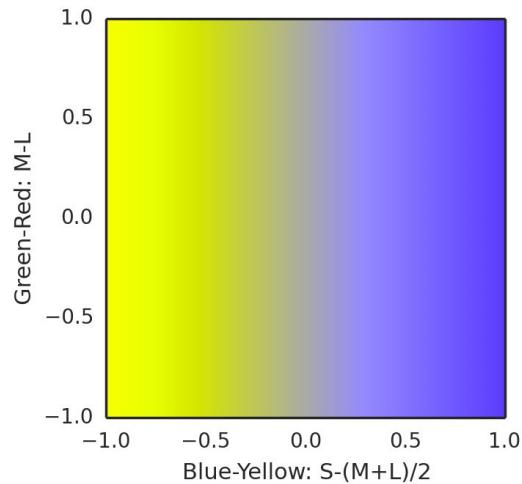
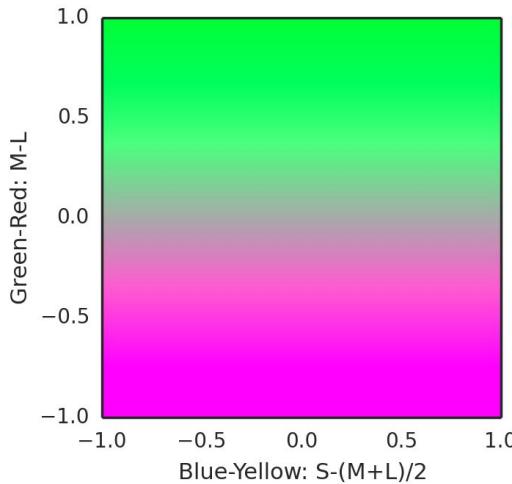


Color Before the Brain

Luminance = S + M + L (+ rods)

Blue-Yellow opposition = $S - (M + L)$

Red-Green opposition = $M - L$

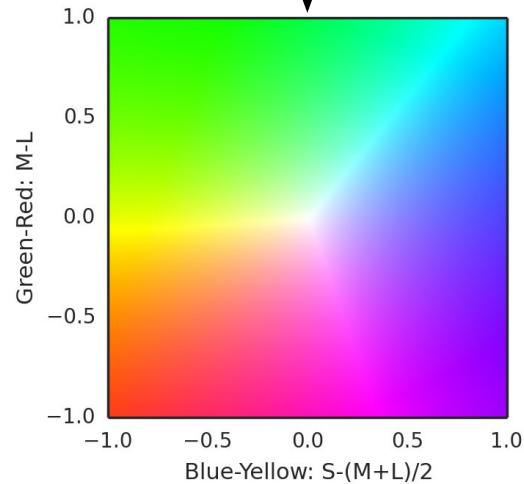
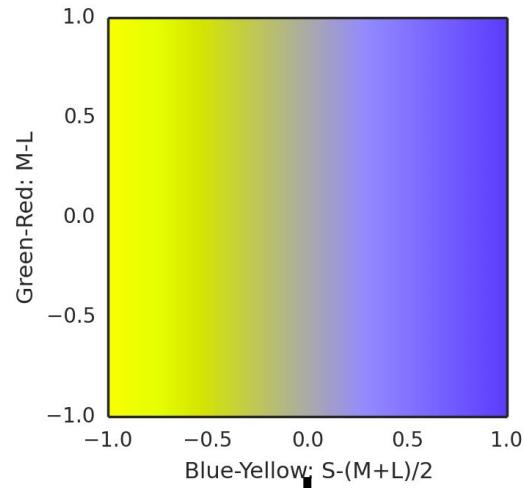
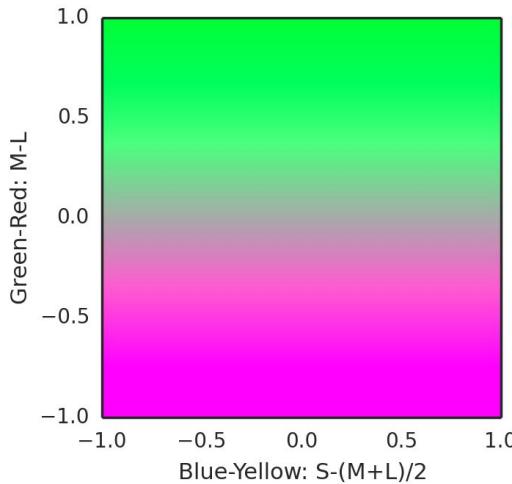


Color Before the Brain

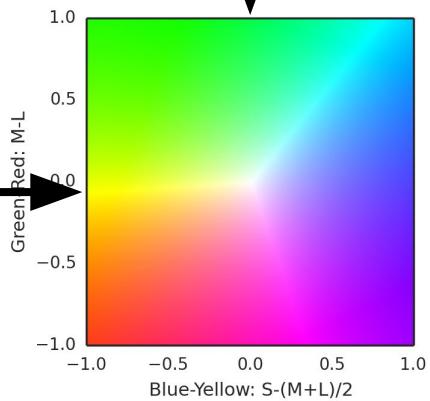
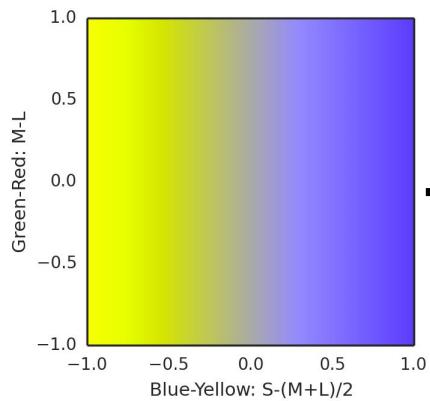
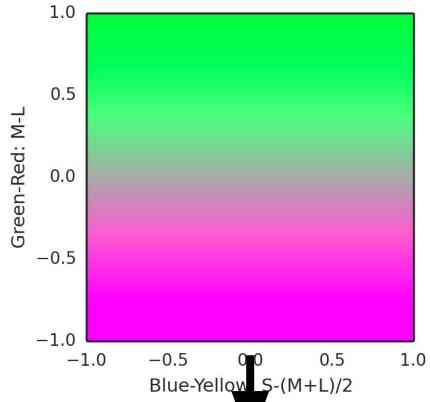
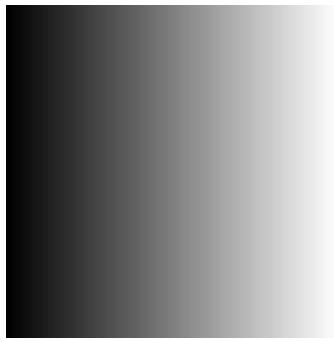
Luminance = S + M + L (+ rods)

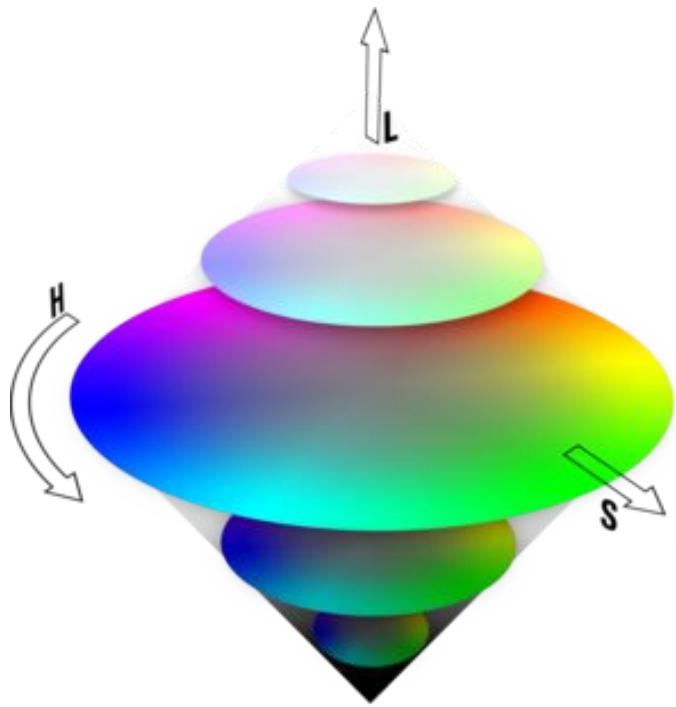
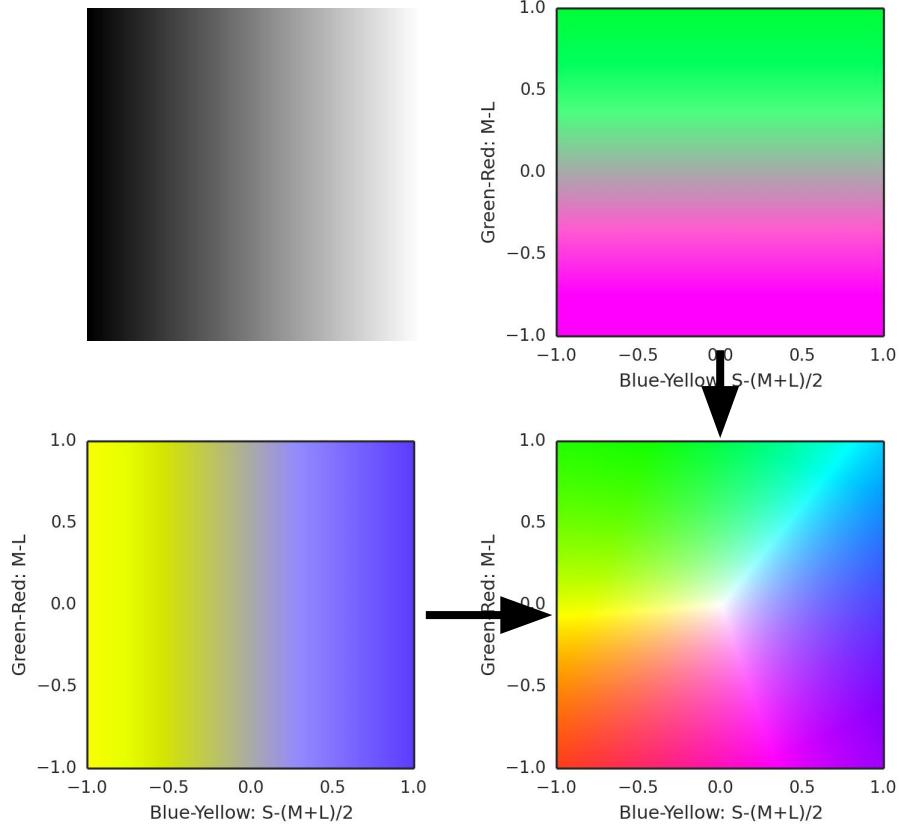
Blue-Yellow opposition = $S - (M + L)$

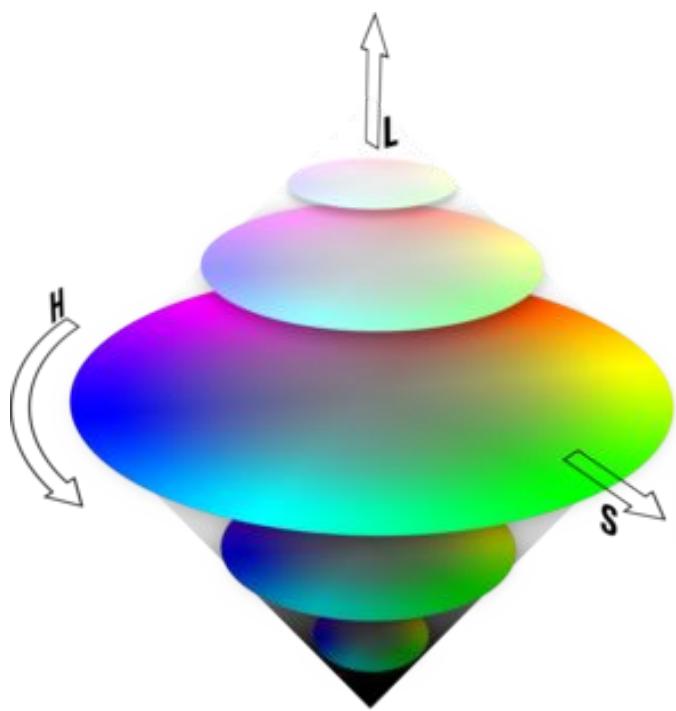
Red-Green opposition = M - L



Encoding Using Hue Saturation and Luminance Space







HSL Encoding

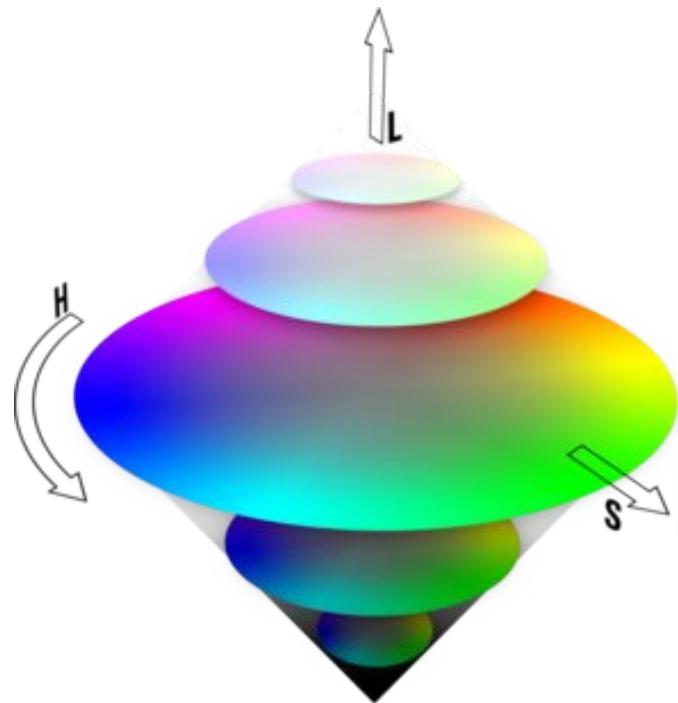
Hue



Saturation (and opposing processes)



Luminance/Value



HSL Encoding

Hue:

Nominal



Saturation (and opposing processes):

Quantitative



Luminance/Value:

Quantitative

