Visualization and Data Presentation 1

Jordi Bieger

Reykjavik University
School of Computer Science
Center for Analysis and Design of Intelligent Agents

jbieger@gmail.com

September 29, 2014

Overview

Research

Critical!

A critical part of research is *communicating* your findings to an *audience*.

Communication Methods

- Text
- Math / Logic / Code
- **Tables**
- **Graphs**
- Diagrams
- Illustrations
- Animation
- ...

Communication Methods

- Text
- Math / Logic / Code
- Tables
- Graphs
- Diagrams
- Illustrations
- Animation
- ...

Communication Methods

- Text
- Math / Logic / Code
- **Tables**
- **Graphs**
- Diagrams
- Illustrations
- Animation
- ...

Communication Methods

- Text
- Math / Logic / Code
- **Tables**
- **Graphs**
- Diagrams
- Illustrations
- Animation
- ...

Communication Methods

- Text
- Math / Logic / Code
- **Tables**
- **Graphs**
- Diagrams
- Illustrations
- Animation
- ...

Communication Methods

- Text
- Math / Logic / Code
- **Tables**
- **Graphs**
- Diagrams
- Illustrations
- Animation
- ...

Communication Methods

- Text
- Math / Logic / Code
- **Tables**
- **Graphs**
- Diagrams
- Illustrations
- Animation
- ...

Communication Methods

- Text
- Math / Logic / Code
- **Tables**
- **Graphs**
- Diagrams
- Illustrations
- Animation
- ...

General remarks

Style

Always check the journal / conference and author instructions for the general style of tables and figures.

Captions

Captions should make it possible to understand completely what a table or figure shows.

General remarks - cont.

Completeness

By highlighting and discussing the important parts of tables and figures, the text should be understandable just by reading the text.

Early

Consider the kind of visualizations you want to use when you are designing your experiment.

Variable Types

Experimental variables:

1. **Independent**: Variable that is not changed by the other variables (e.g. age).
2. **Dependent**: Measured variable that is affected by others (e.g. cancer risk).

Data types:

1. **Nominal / Categorical**: Discrete data that cannot be ordered (e.g. people).
2. **Ordinal**: Quantities with a natural order (e.g. Likert questions).
3. **Interval**: Ordinal + the interval between each value is equal (e.g. Fahrenheit).
4. **Ratio**: Interval + a natural zero point (e.g. elevation).

Tables

Table: Caption (often above table).

Stub	**Column heading**	**Column heading**
Row variable 1	x%	x%
Row variable 2	x%	x%
Row variable 3	x%	x%
Row variable 4	x%	x%
Total	x%	x%

Multivariate table

Attitude towards uranium mining by age and gender (hypothetical data)

	Number of respondents												
	<25		25–34		35–44		45–54		55+		Total		
Attitude towards uranium mining	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**T**
Strongly favourable	0	0	1	1	3	1	5	2	3	–	12	4	16
Favourable	0	0	1	2	3	2	3	1	0	0	7	5	12
Uncertain	0	0	0	0	1	1	2	2	0	0	3	3	6
Unfavourable	1	1	4	3	1	0	0	0	0	0	6	4	10
Strongly unfavourable	4	8	17	7	8	7	2	3	0	0	31	25	56
Total	5	9	23	13	16	11	12	8	3	0	59	41	100

Figure: Table 16.4 from the book.

Confusion matrix

		Predicted	
		True	*False*
Actual	*True*	tp	fn
	False	fp	tn

	A	*B*	*C*	*D*
A				
B				
C				
D				

Table usage

Use a table when:

- Detailed data and text
- Large volume*
- No trend or pattern

2D Chart Anatomy

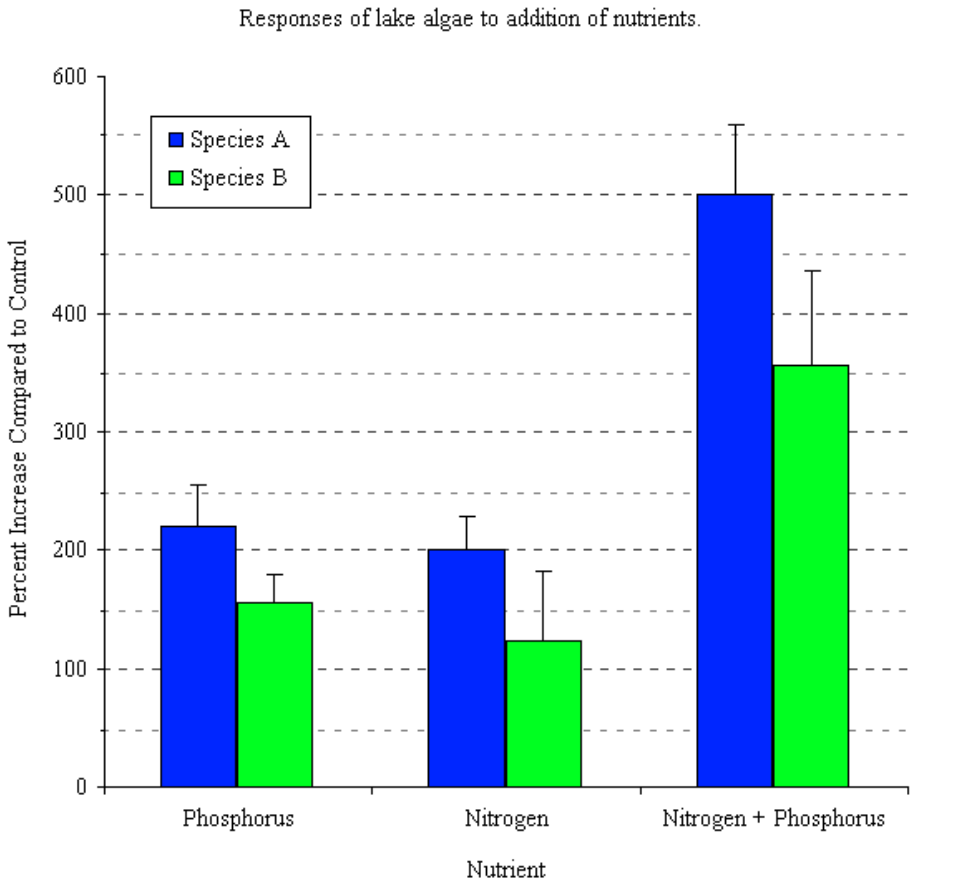


Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

N = 50 lakes.

Figure: Adapted from **University of Wisconsin-La Crosse (2001)**.

2D Chart Anatomy

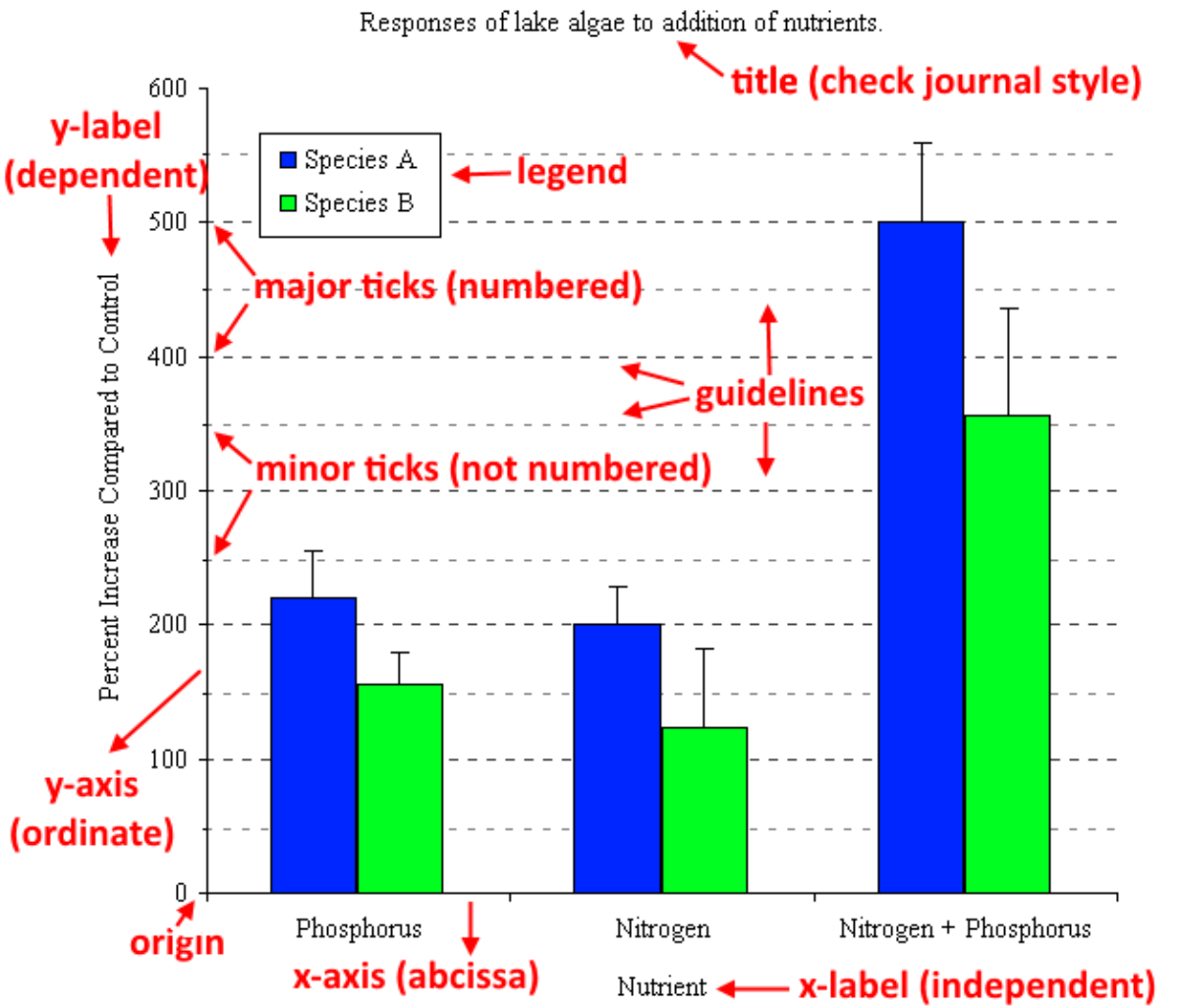


Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

N = 50 lakes.

caption

2D Chart Anatomy - Axis Offset

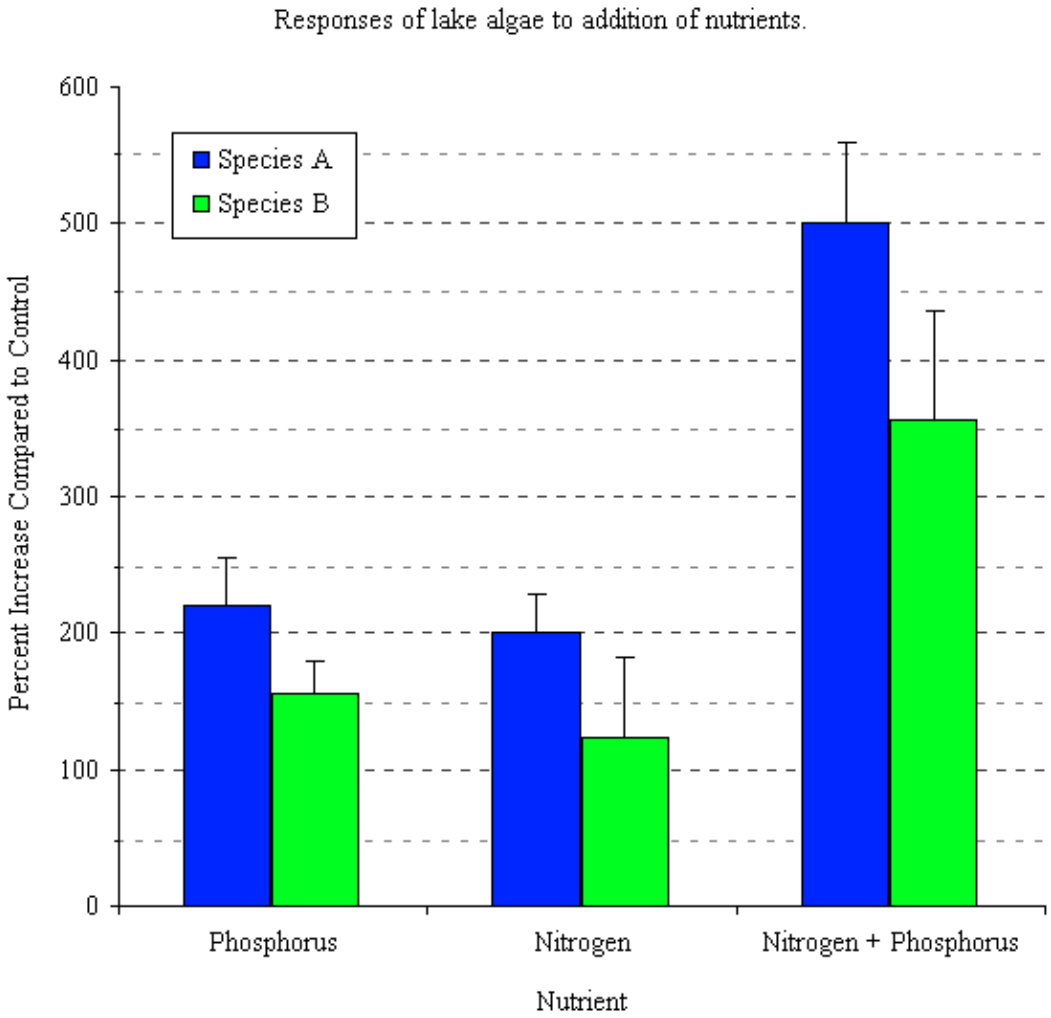


Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

N = 50 lakes.

2D Chart Anatomy - Axis Offset

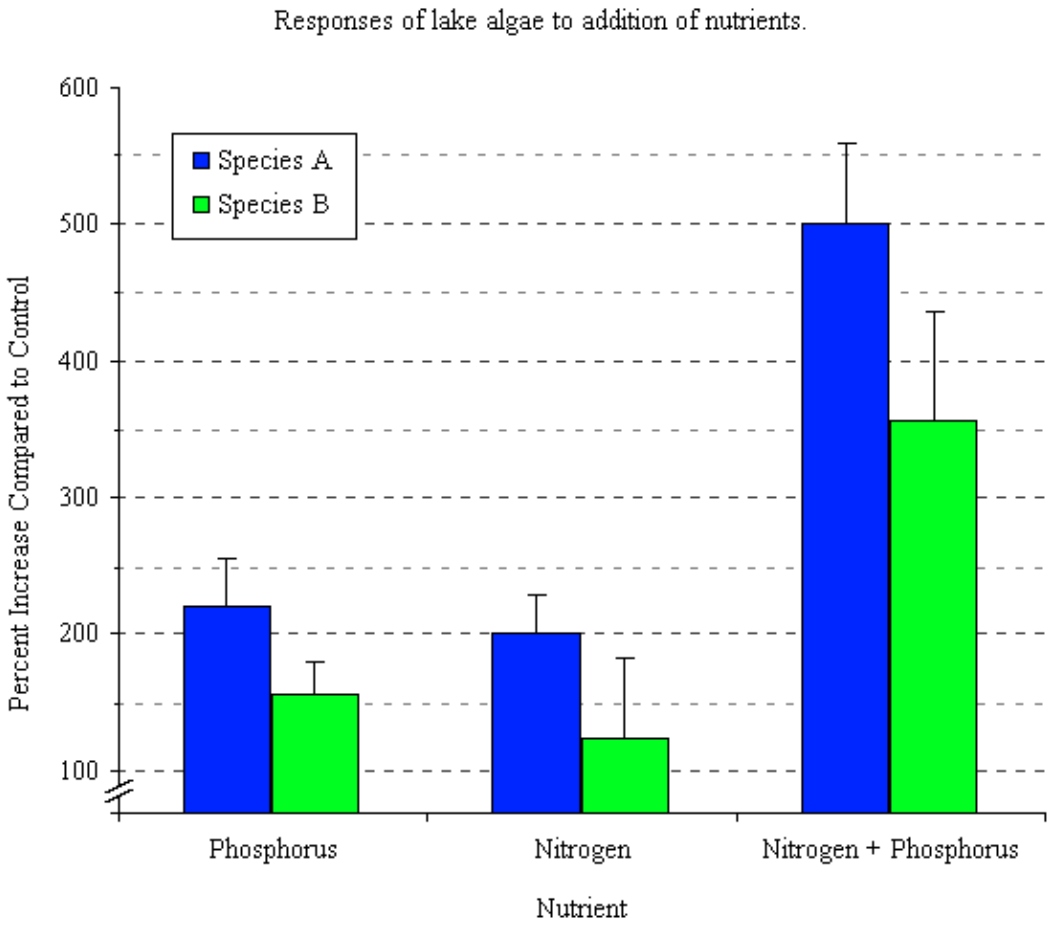


Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

N = 50 lakes.

2D Chart Anatomy - Axis Offset

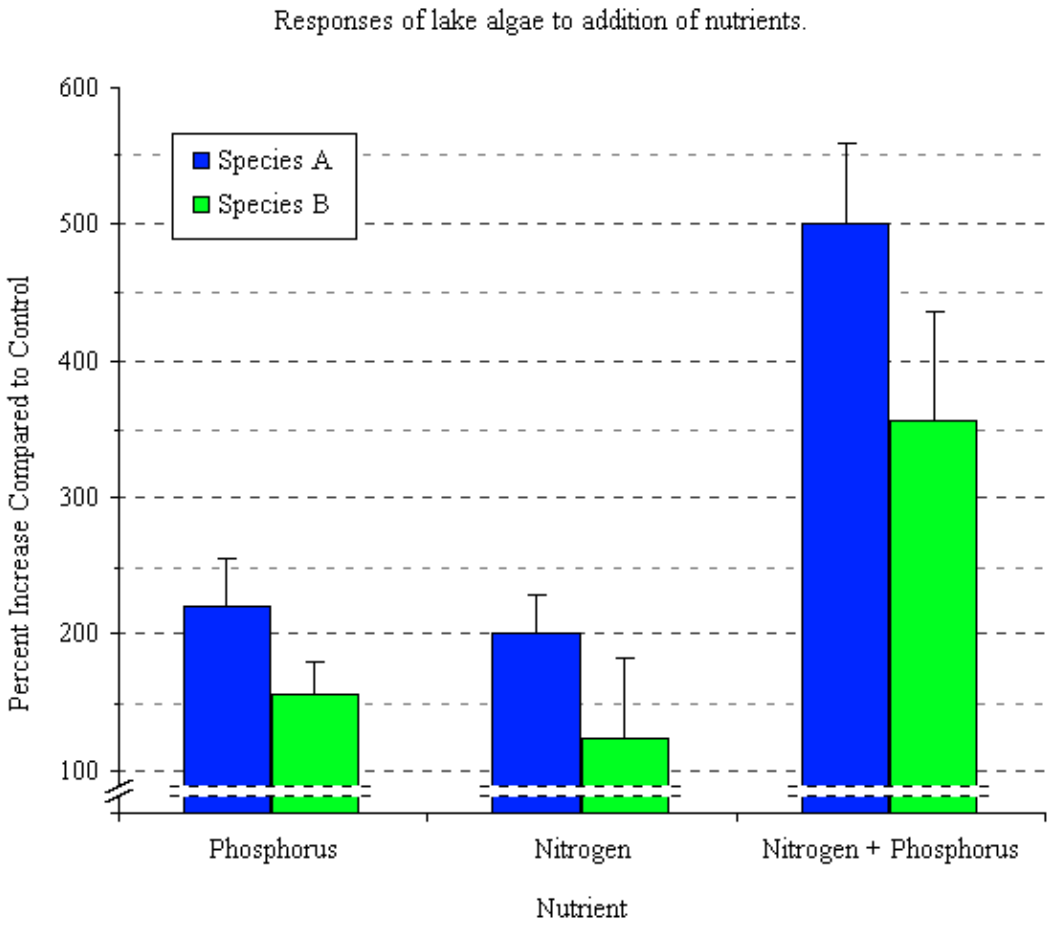
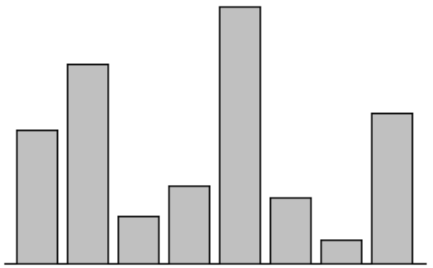


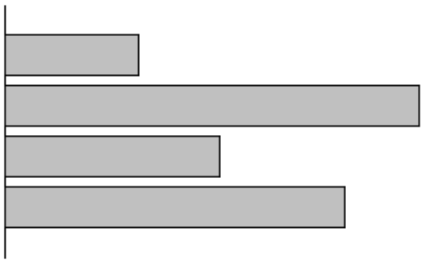
Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

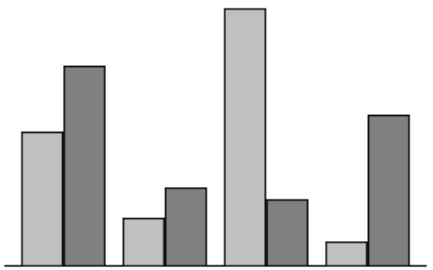
N = 50 lakes.

Bar charts

- Use when you want to compare how values of 1 or 2 discrete independent variables affect a numeric dependent variable or count.
- Actual numbers and/or error bars can be added on top of the bars.
- For ordinal data, a histogram may also be used.
- With 2 independent variables, a stacked bar chart can also be used, but this is not recommended for comparisons.

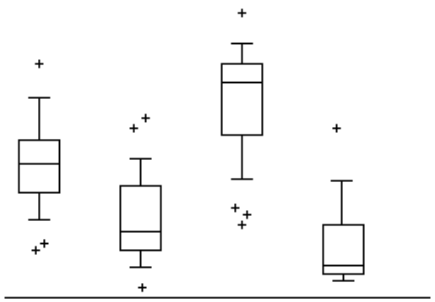






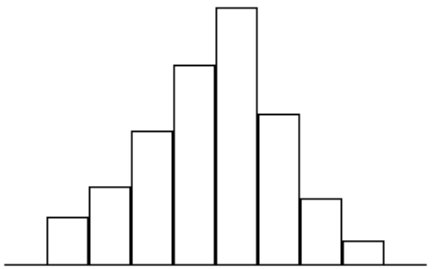
Box plots

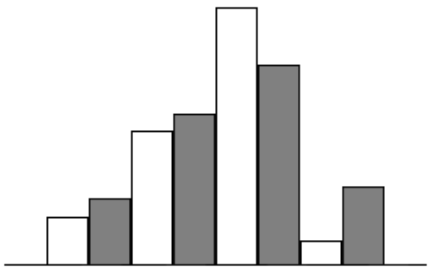
- Box plots are like bar charts with extra information.
- They generally show the 1st, 2nd and 3rd quartile of the data, the range and outliers.



Histograms

- Use when you want to show the distribution of items over a small number of values of a quantitative variable.
- For ordinal data, a bar chart may also be used.
- Ratio or interval data can be divided into buckets/intervals; otherwise you can use a area chart.
- With multiple variables, use a 3D effect or an overlapping area chart.
- The book calls drawing a line over the histogram a "frequency polygon".





Stem-and-leaf display

A stem-and-leaf display let's you show fairly detailed distribution information in the shape of a histogram.

Example (Data)

```
37, 33, 33, 32, 29, 28, 28, 23,
22, 22, 22, 21, 21, 21, 20, 20,
19, 19, 18, 18, 18, 18, 16, 15,
14, 14, 14, 12, 12, 9, 6
```

Example from **Lane @ OnlineStatBook**.

Example (S&L display 1)

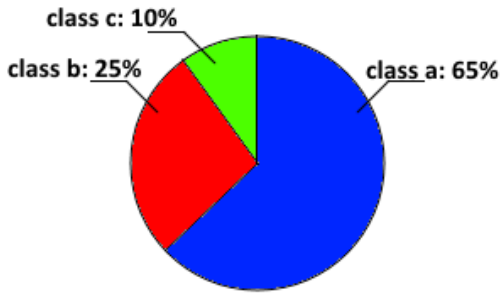
```
3|2337
2|001112223889
1|2244456888899
0|69
```

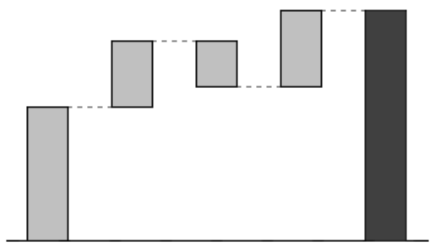
Example (S&L display 2)

```
3|7
3|233
2|889
2|001112223
1|56888899
1|22444
0|69
```

Static composition charts

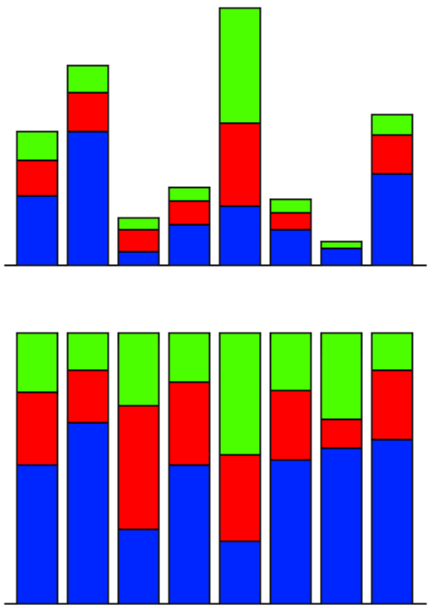
- If the composition doesn't vary, you can use a pie chart.
- If the composition also contains negative items, you can use a waterfall chart.





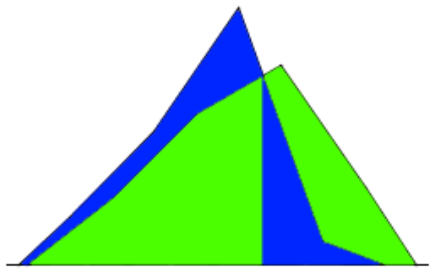
Stacked bar charts

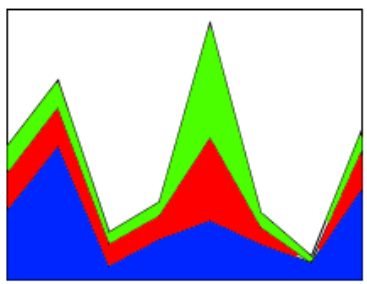
- Use to show the composition of a thing varying along some discrete dimension.
- Use a 100% bar chart if the absolute value doesn't matter.

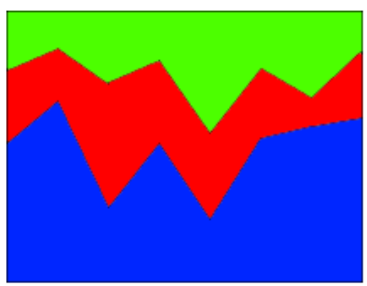


Area charts

- Area charts can be used to show distributions under a continuous independent variable.
- Stacked area charts can also be used to show how compositions vary with such a variable.

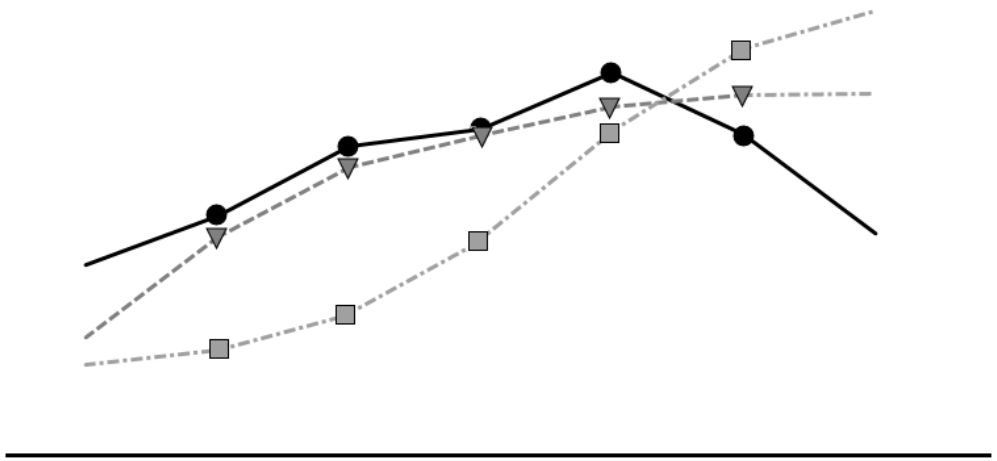






Line charts

Line charts can be used to show how several numeric quantitative variables change with another variable (e.g. time).



Radar charts

Radar or spider charts can be used as cyclical line charts or to show how various items compare to each other on a number of dimensions.

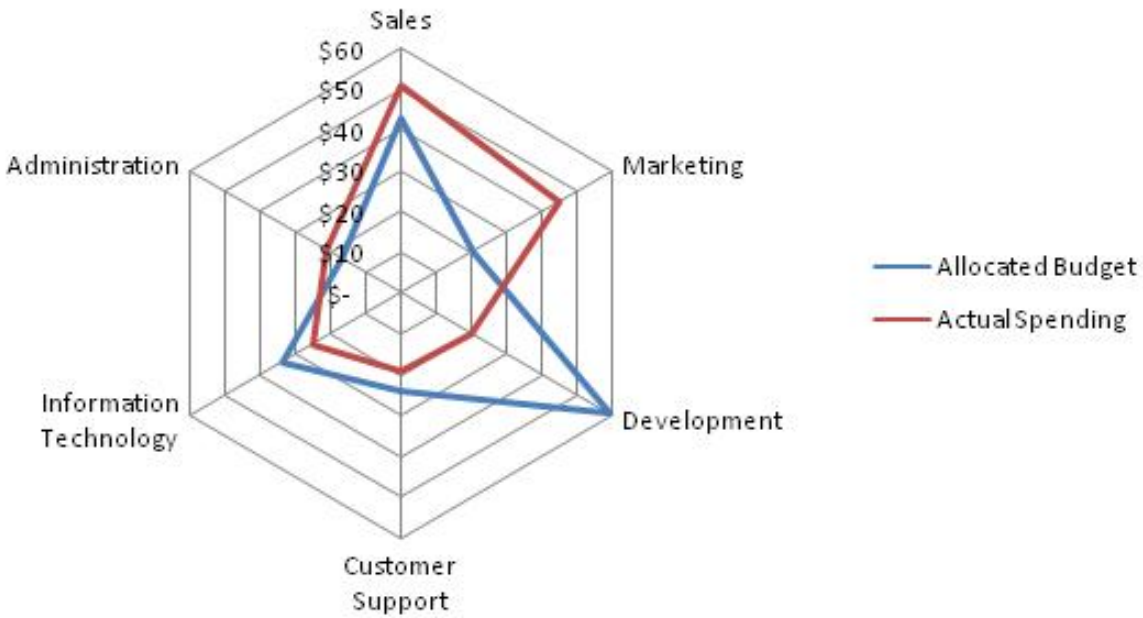
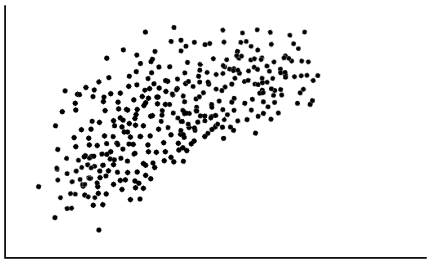
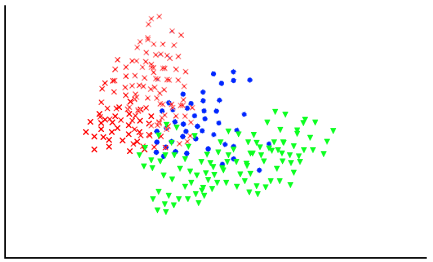


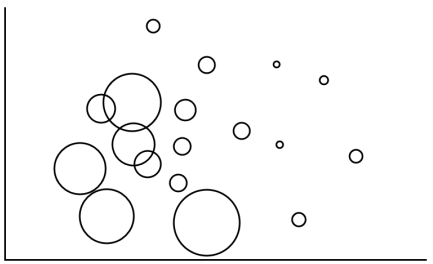
Figure: From **Wikipedia**.

Scatter plots

- Scatter plots are useful for seeing the relationship between two quantitative variables.
- Bubble plots let you add another dimension.





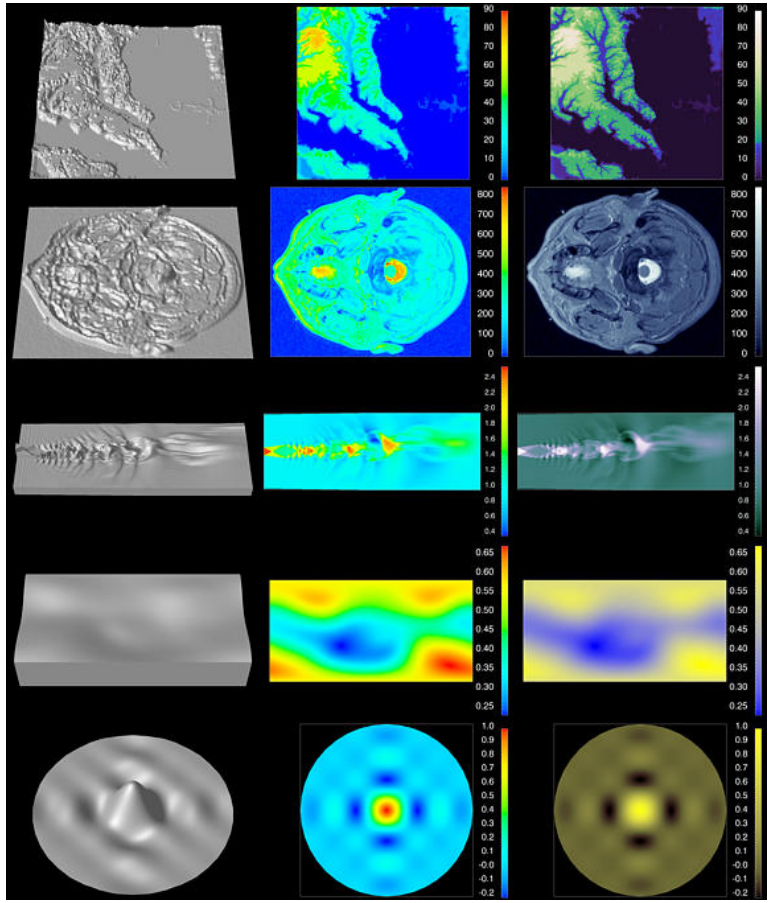


Heat maps

Heat maps can correlate large amounts of data to specific locations.

Image from

Rogowitz & Treinish (1996)



Color

Journal

Always check the style of the journal!

Legibility

Keep everything legible!

Account for B&W

Even if you use color, make sure your figures are interpretable if someone prints them without or is color blind. (Don't refer to the color in the text.)

Bar chart color

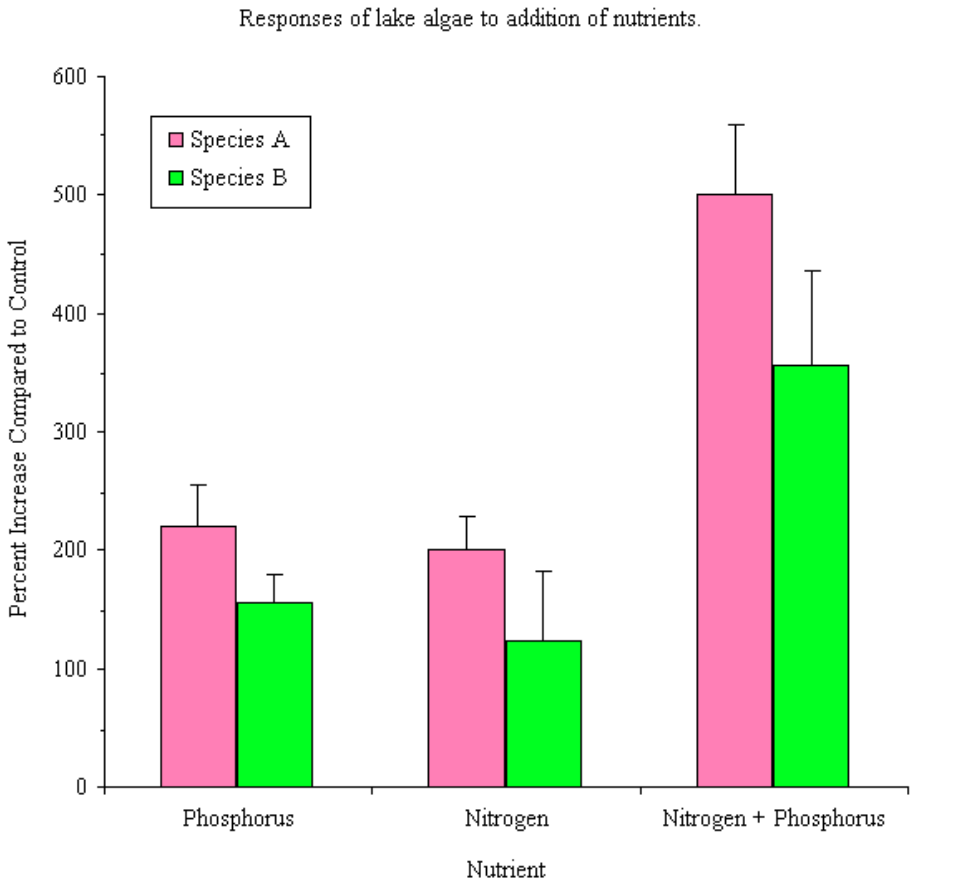


Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

N = 50 lakes.

Figure: Adapted from **University of Wisconsin-La Crosse (2001)**.

Bar chart color

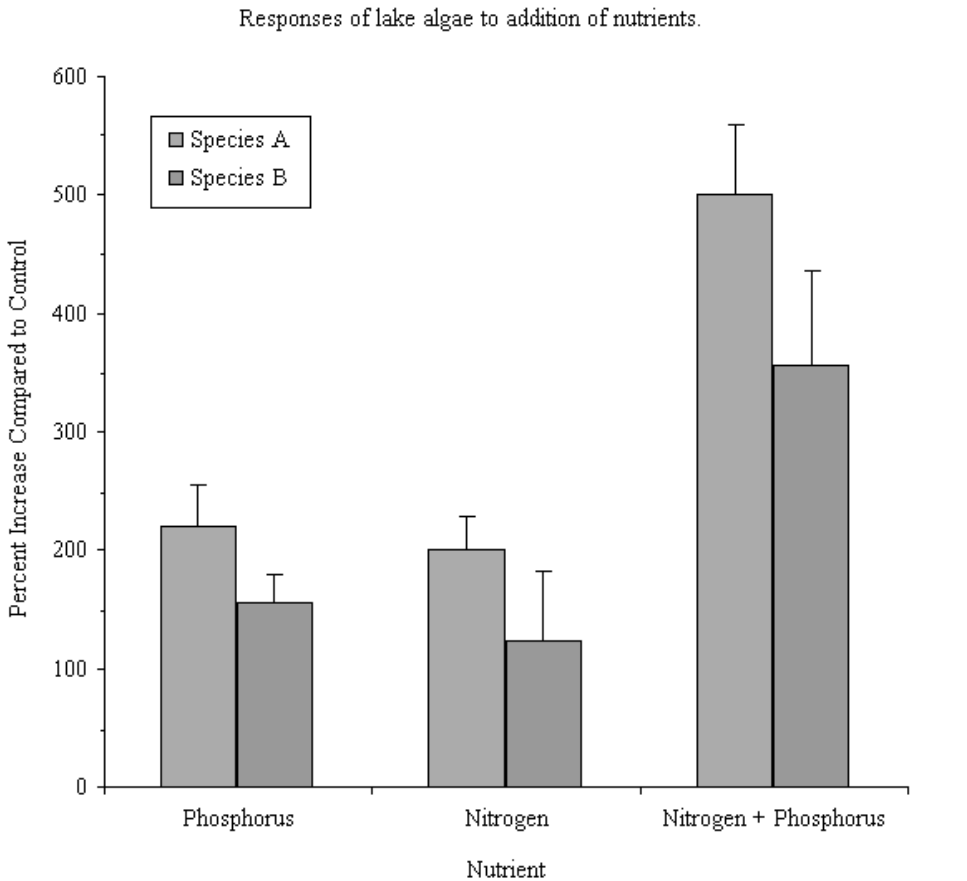


Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

N = 50 lakes.

Figure: Adapted from **University of Wisconsin-La Crosse (2001)**.

Bar chart color - Hatching

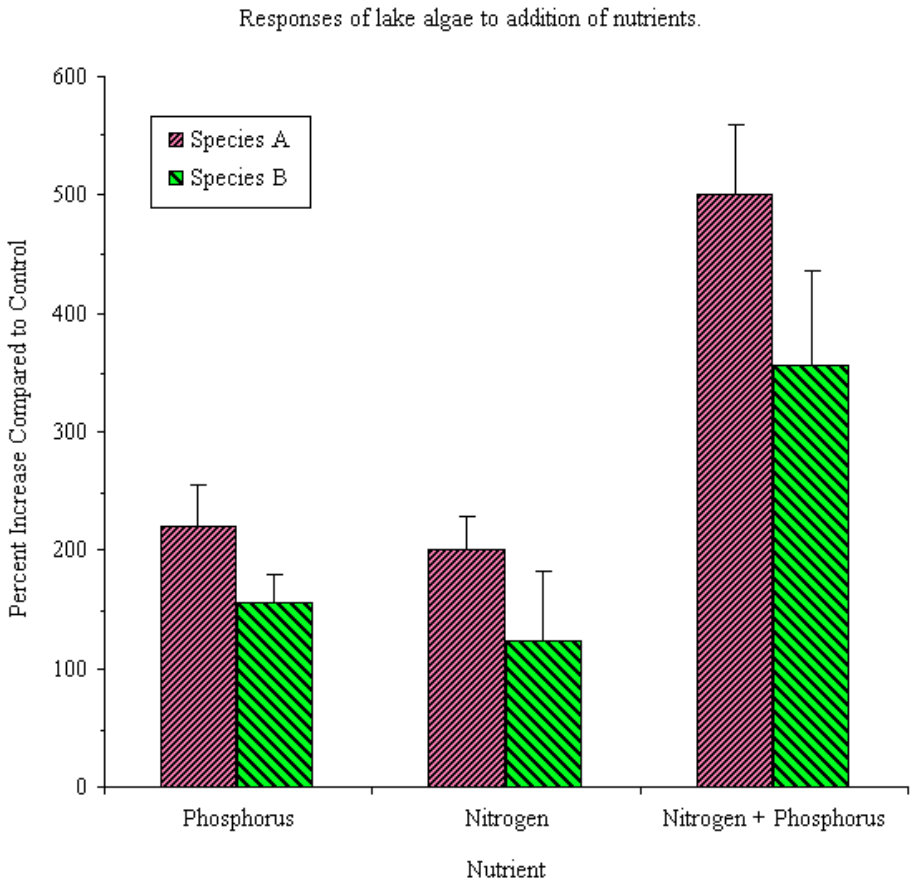


Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

N = 50 lakes.

Figure: Adapted from **University of Wisconsin-La Crosse (2001)**.

Bar chart color - Hatching

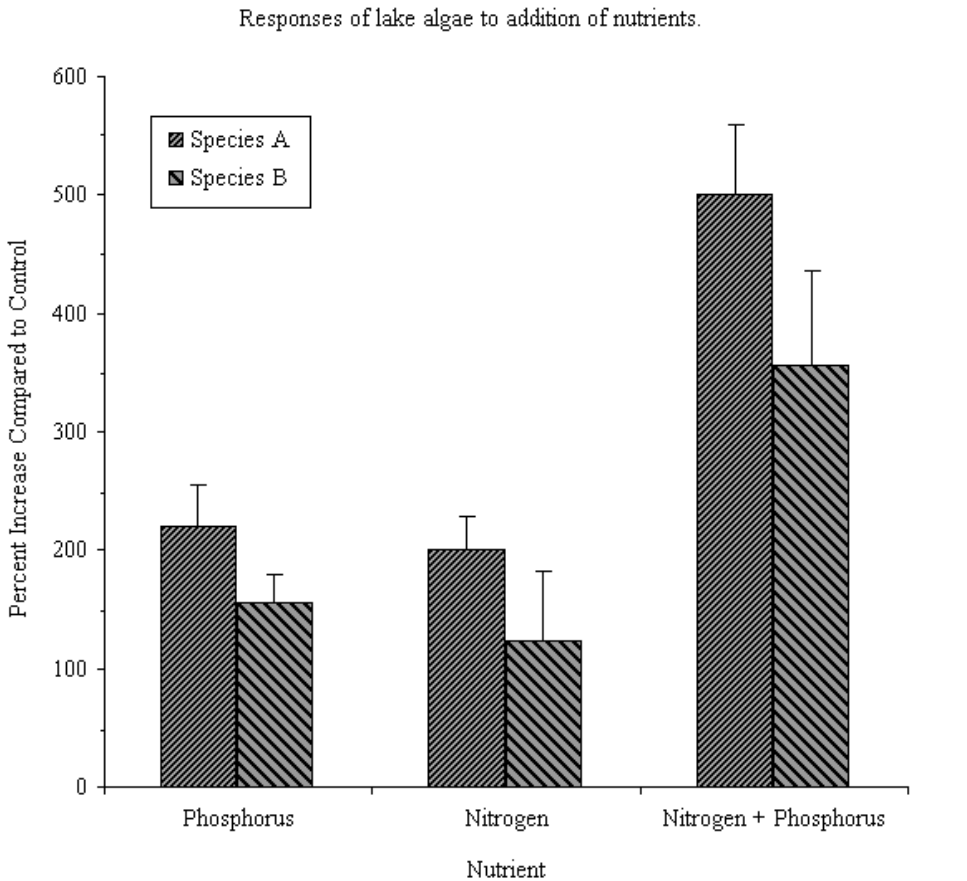


Figure 3. Responses of lake algae to addition of nutrients. Error bars are 95% confidence limits.

N = 50 lakes.

Figure: Adapted from **University of Wisconsin-La Crosse (2001)**.

Line chart color - Notches and Line Types

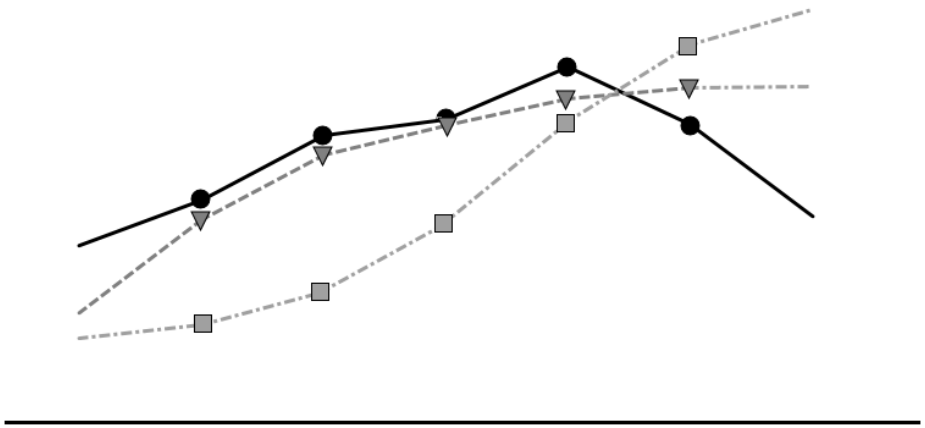


Figure: Use different notches and line types.

Zero crossing

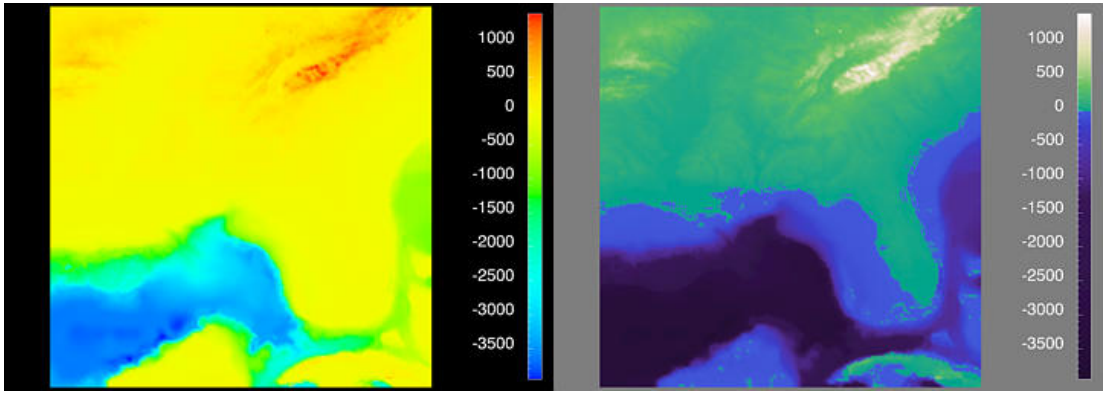


Figure: From **Rogowitz & Treinish (1996)**.

Spatial contrast

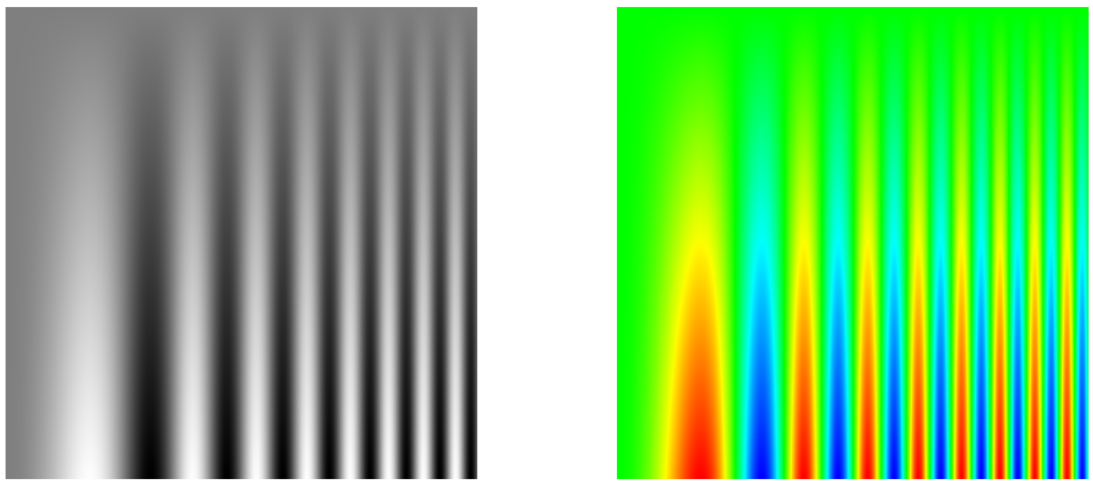


Figure: From **Moreland (2009)**.

Color theory - Color wheel

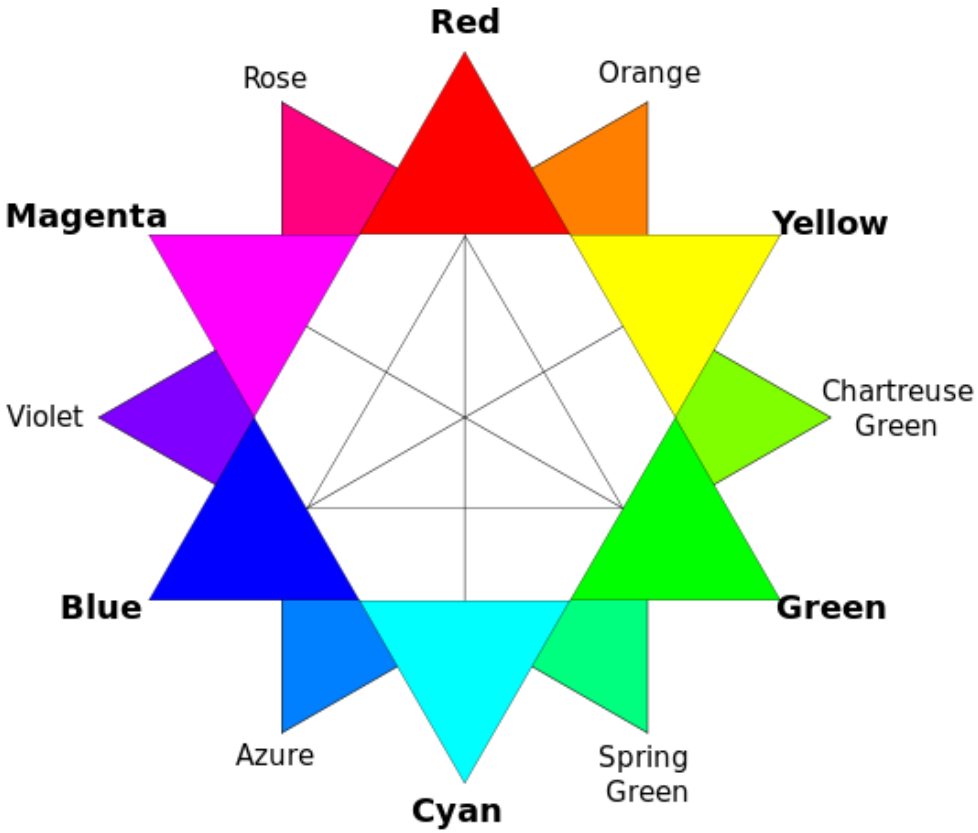


Figure: From **Wikipedia**.

Color theory - Opponent colors

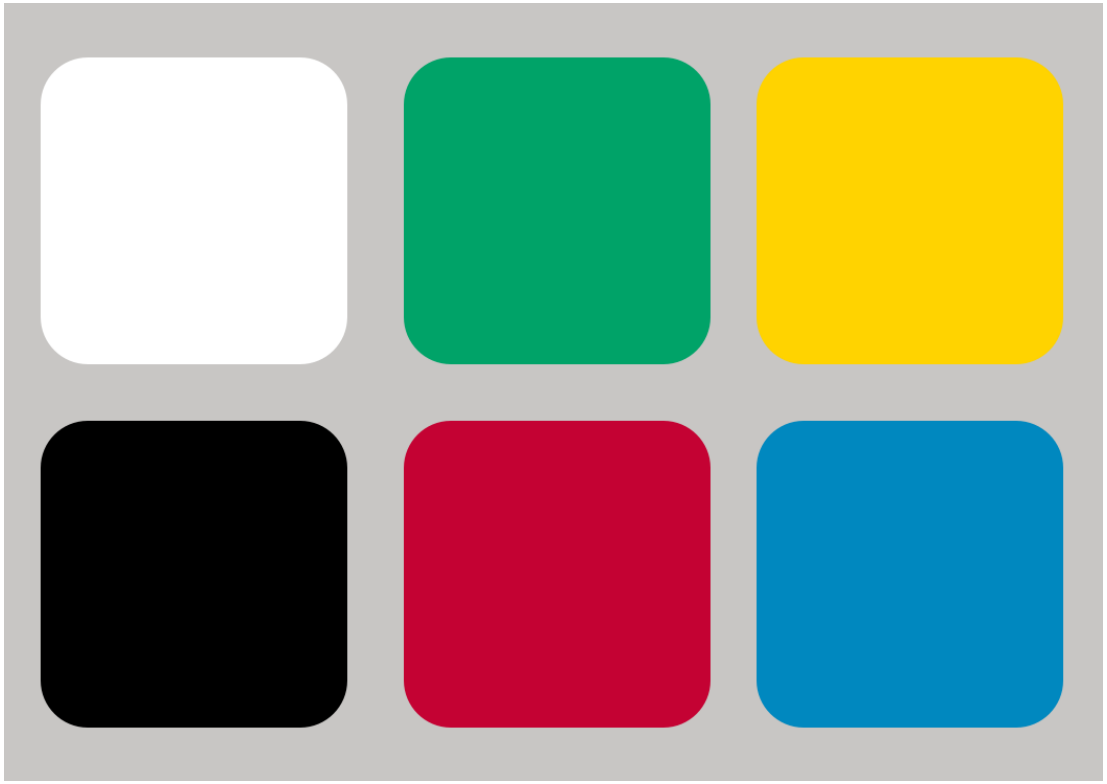


Figure: From **Wikipedia**.

Color theory - HSL

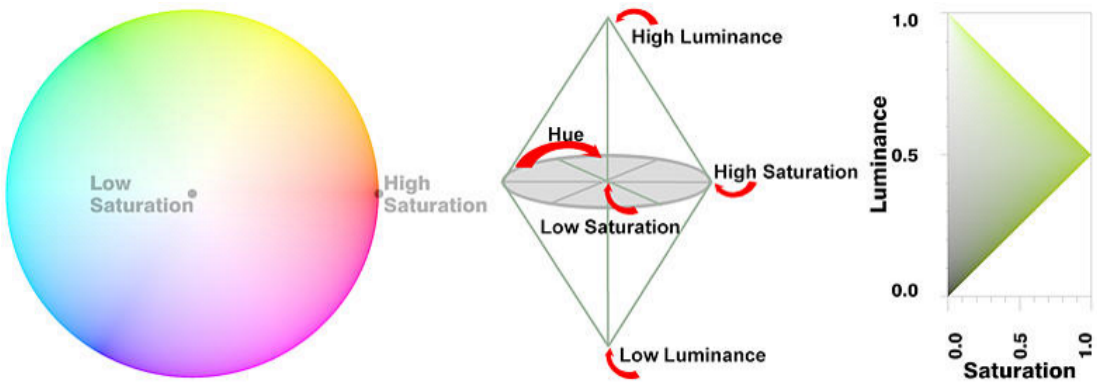


Figure: From **Rogowitz & Treinish (1996)**.

Choosing a color map

- Isomorphic: represent data structure faithfully
- Segmentation: divide data into distinct categories
- Highlighting: call attention to particular features

Choosing a color map - Ratio

Data Type	Spatial Frequency	Representation Task		
		Isomorphic	Segmentation	Highlighting
Ratio (true zero)	Low	*Luminance:* uniform *Hue:* opponent or complementary pairs *Saturation:* monotonically increasing from gray	Even number of segments Many segments OK	Larger range for highlighted features
	High	*Luminance:* monotonically increasing *Hue:* opponent or complementary pairs *Saturation:* monotonically increasing from gray	Even number of segments Fewer segments	Smaller range for highlighted features

Figure: From **Bergman, Rogowitz & Treinish (1995)**.

Choosing a color map - Interval

Data Type	Spatial Frequency	Representation Task: Isomorphic	Representation Task: Segmentation	Representation Task: Highlighting
Interval	Low	*Luminance:* uniform *Hue:* opponent pairs *Saturation:* monotonically increasing from gray	Many segments OK	Larger range for highlighted features
	High	*Luminance:* monotonically increasing *Hue:* uniform or small hue variation *Saturation:* monotonically decreasing	Fewer segments	Smaller range for highlighted features

Figure: From **Bergman, Rogowitz & Treinish (1995)**.

Choosing a color map - Ordinal

Data Type	Spatial Frequency	Representation Task		
		Isomorphic	Segmentation	Highlighting
Ordinal	Low	*Luminance:* uniform *Hue:* variation around hue circle *Saturation:* monotonically decreasing	Fewer segments	Increase luminance of highlighted area
	High	*Luminance:* monotonically increasing *Hue:* variation around hue circle *Saturation:* uniform	More segments	Increase saturation of highlighted area

Figure: From **Bergman, Rogowitz & Treinish (1995)**.

Choosing a color map - Nominal

Data Type	Spatial Frequency	Representation Task		
		Isomorphic	Segmentation	Highlighting
Nominal	Low	*Luminance:* uniform *Hue:* variation around hue circle *Saturation:* uniform	Fewer segments than 7	Increase luminance or saturation of highlighted area
	High			

Figure: From **Bergman, Rogowitz & Treinish (1995)**.

Color maps

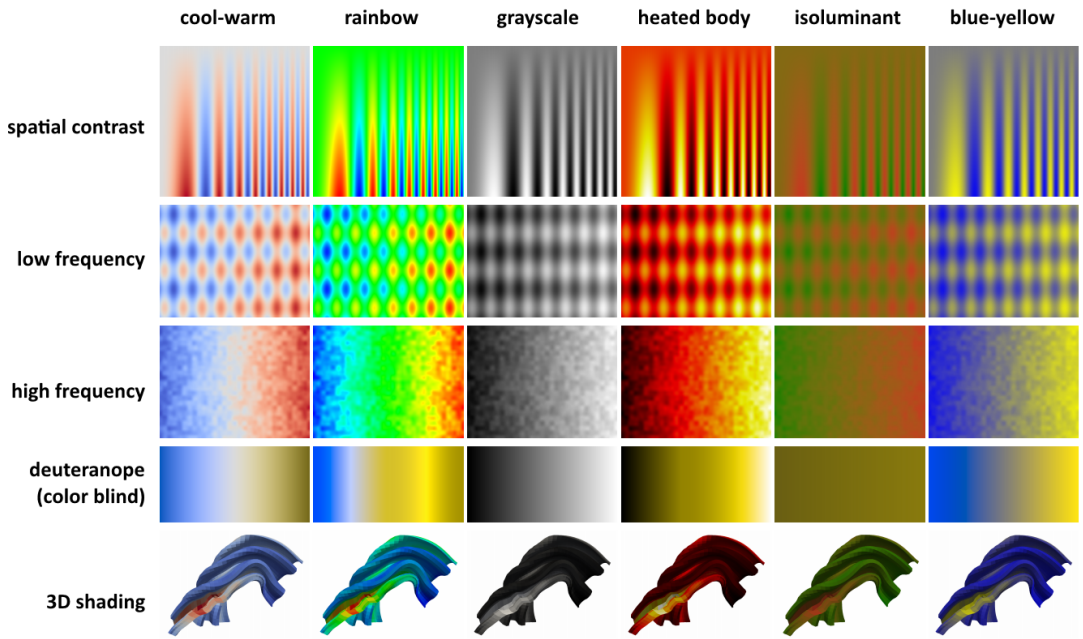


Figure: From **Moreland (2009)**.

Truncated Y-Axis

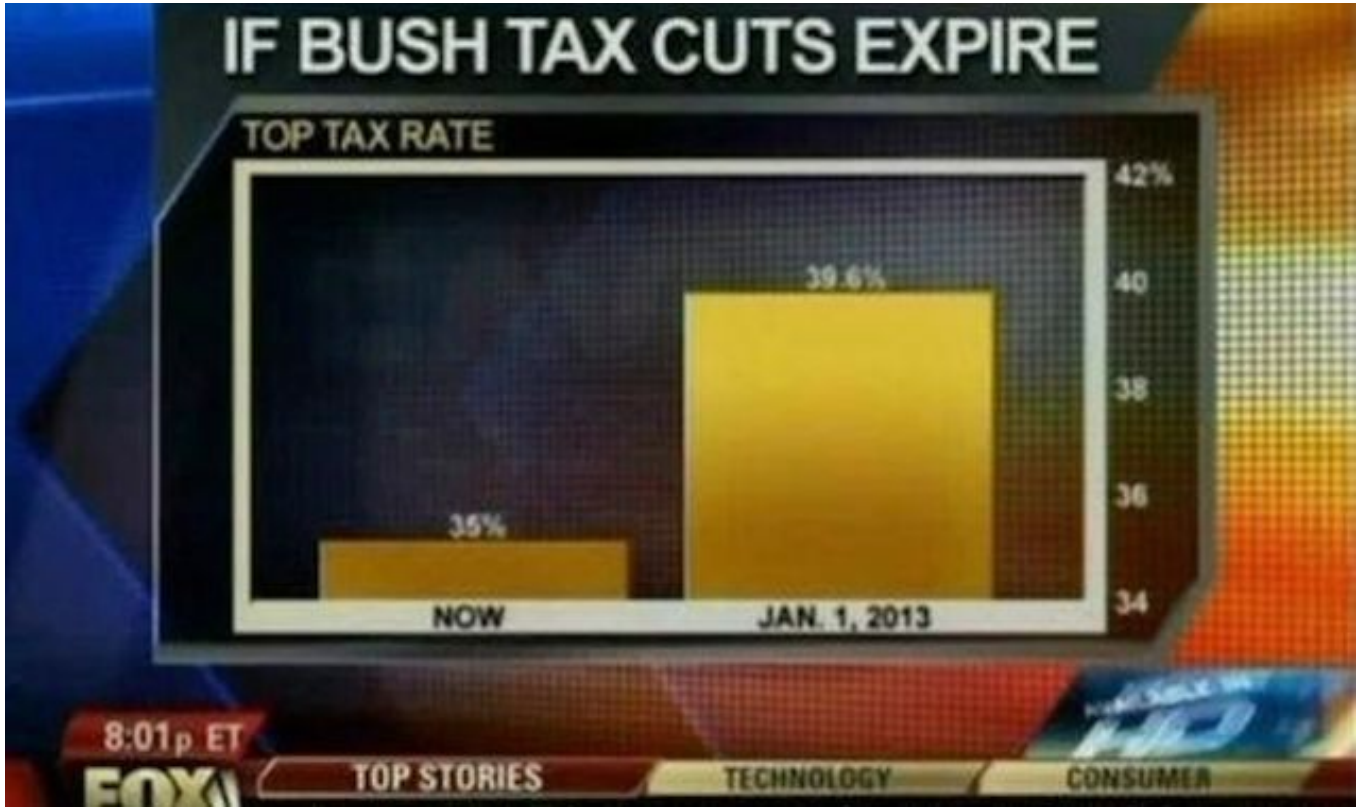


Figure: Via **Parikh @ Gizmodo**.

Different Y-Axis

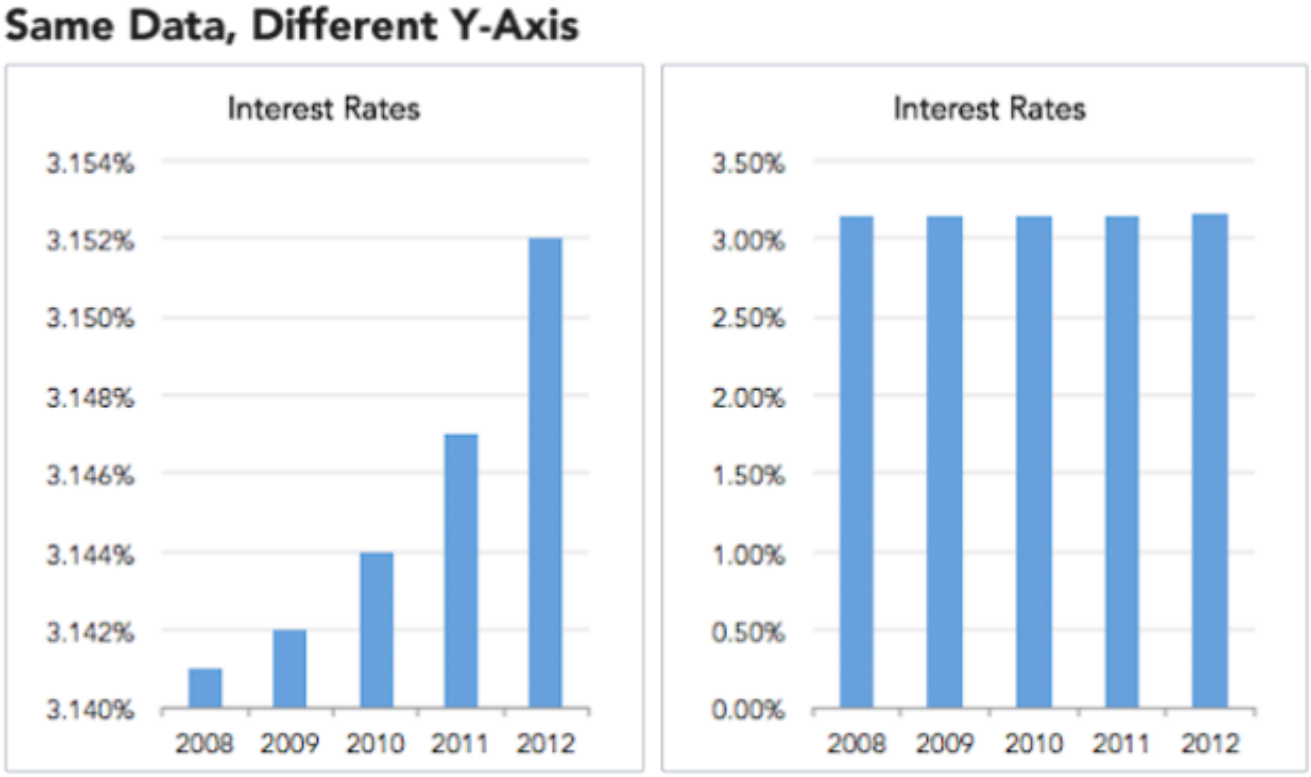


Figure: From **Parikh @ Gizmodo**.

Different Y-Axis

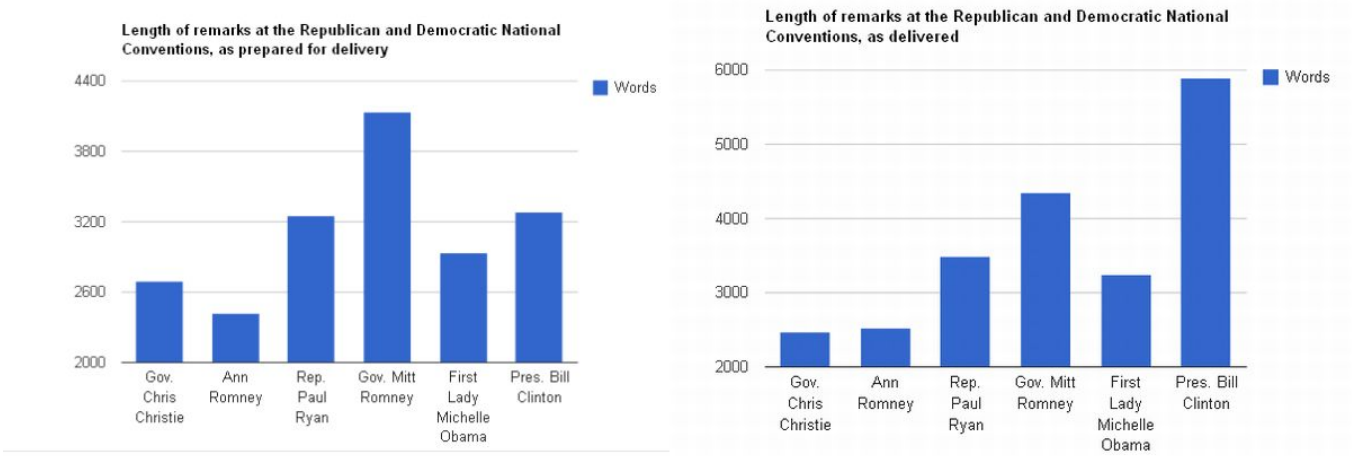


Figure: From **Cliff @ Washington Post**.

Different Order

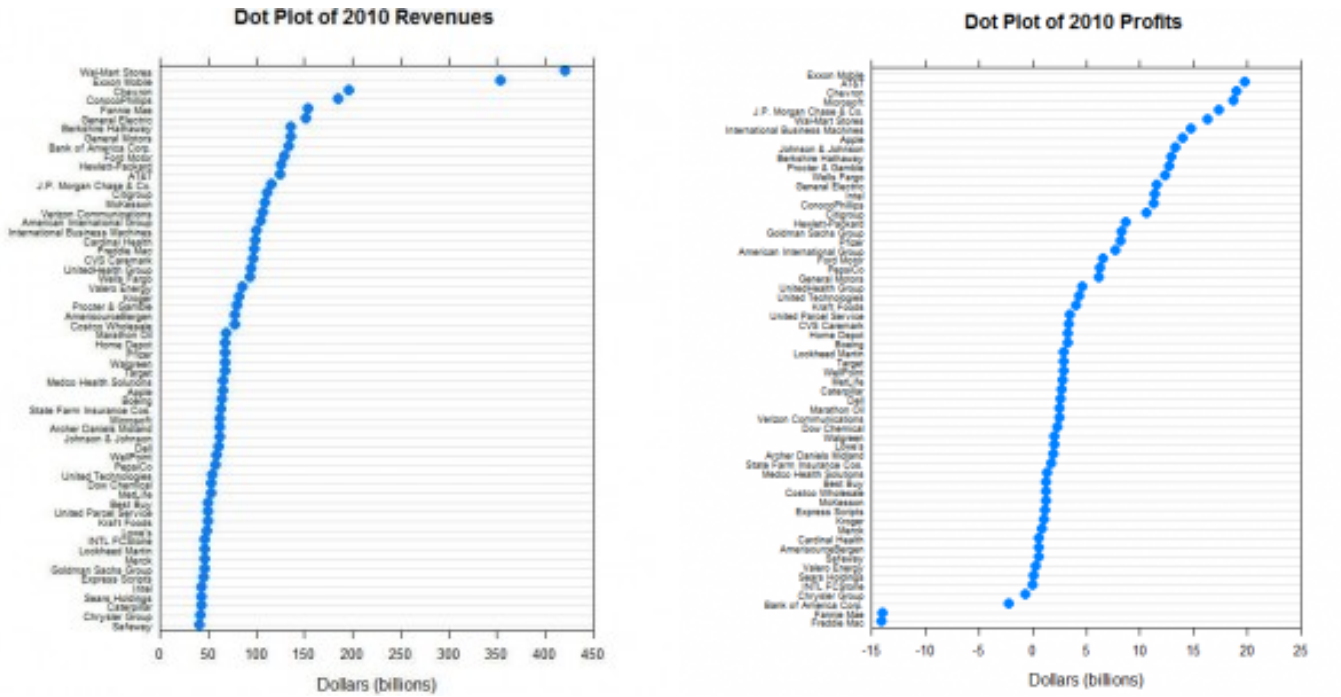


Figure: Via **Robbins @ Forbes**.

Different Order

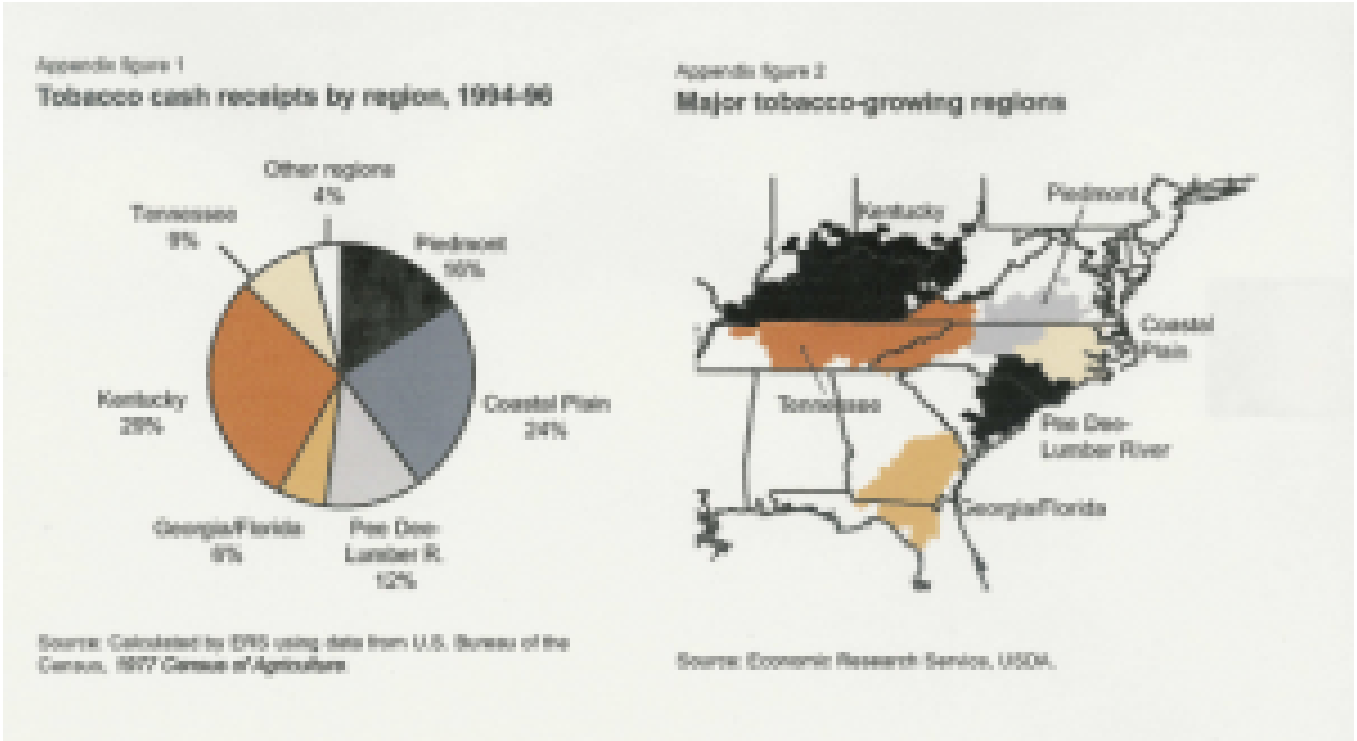


Figure: Via **Robbins @ Forbes**.

Poorly used cumulative graph

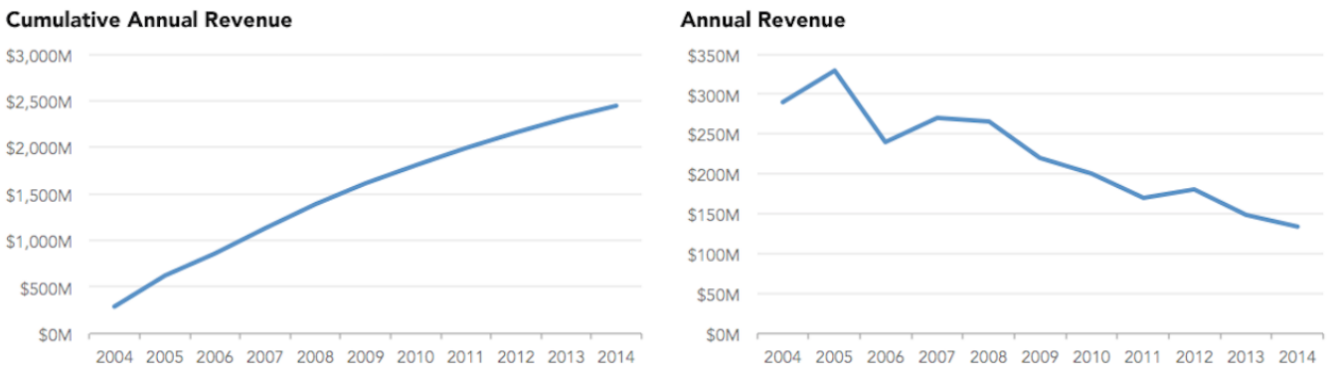


Figure: From **Parikh @ Gizmodo**.

Ignoring conventions and expectations

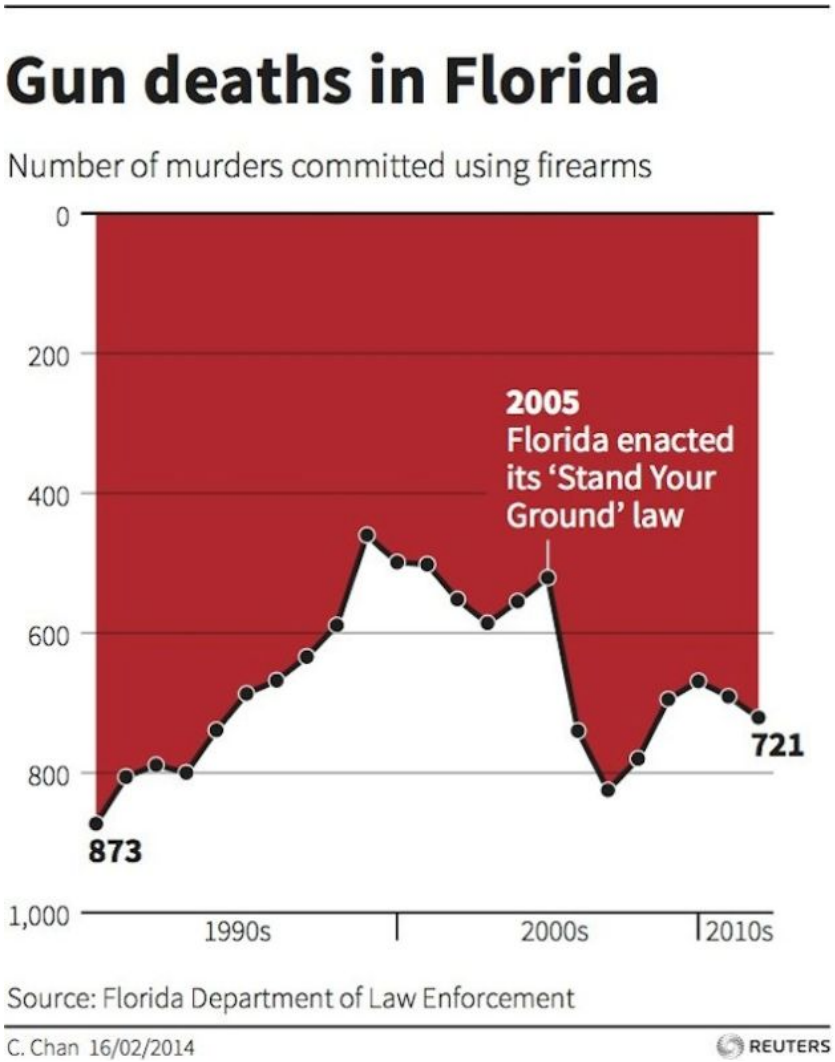


Figure: Via Parikh @ Gizmodo.

etc.

- Poorly used colors.
- "Creative" use of percentages.