Introduction to Data Visualization

Administrative Items

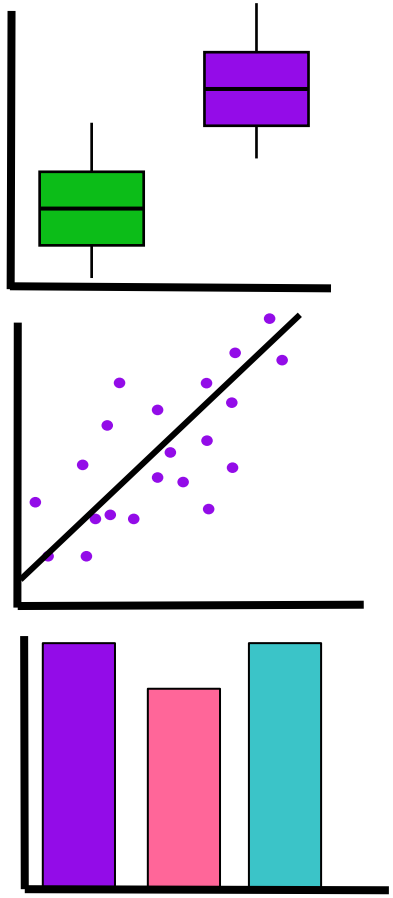
1. Late submission deadline has passed for assignment 1.
    a. For questions with grading, email cogs108@gmail.com.

2. Homework 2 was posted over the weekend
    a. Due date is February 3rd at 11:59
    b. Section this week will be used for working on this assignment

Data Visualization Types

Graphs

- require at least two scales
- visualize variables and their relationships



Examples: box plots, scatterplots, bar plots

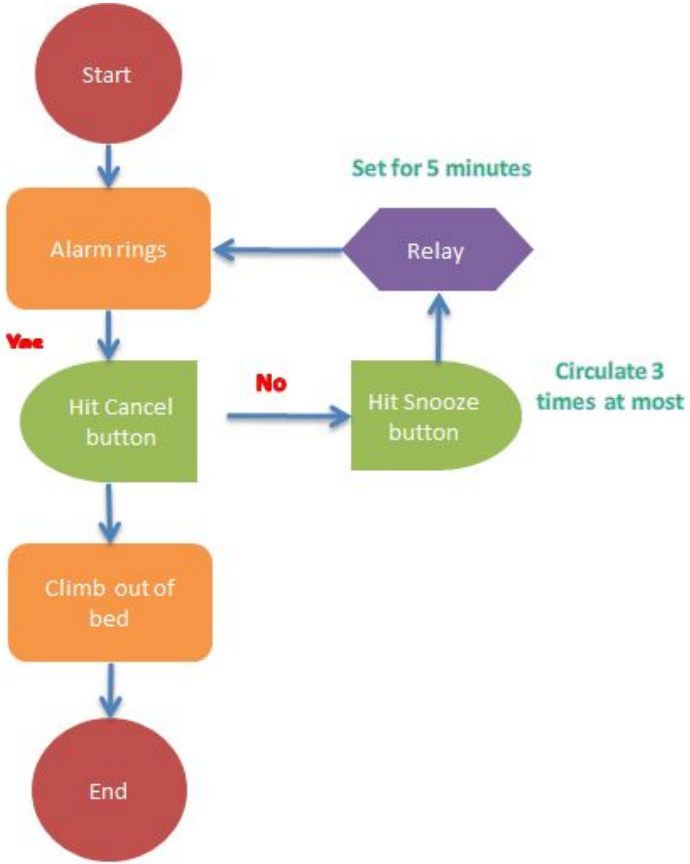
Tables

- display summaries of data
- tabular format

	mean	median
height	68 in	70 in
weight	180 lbs	175 lbs
age	50 yrs	51 yrs

Charts

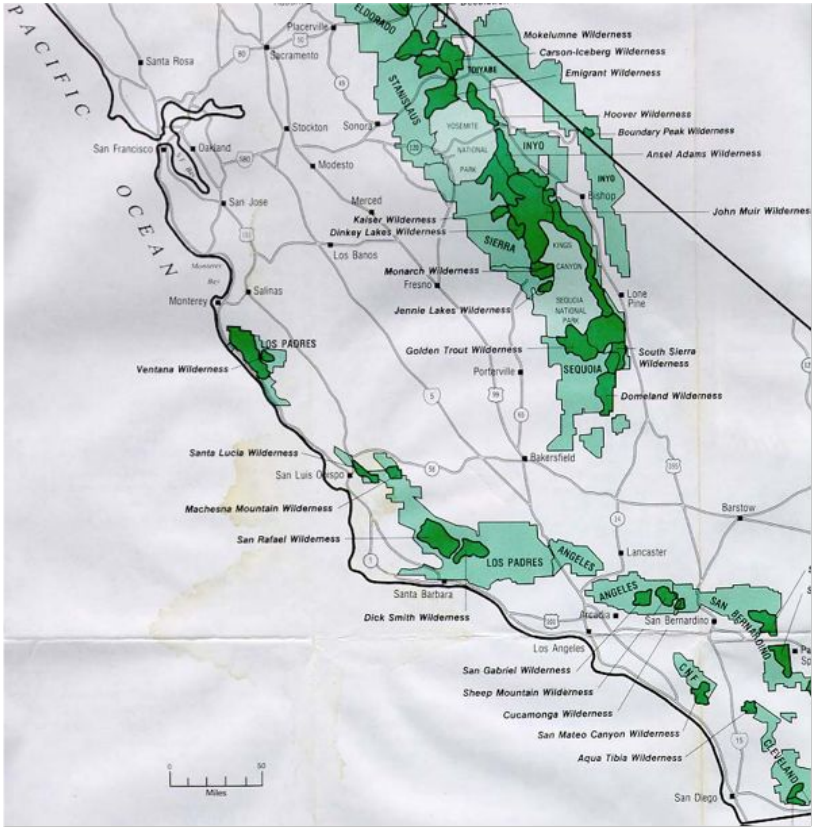
- structure demonstrates relationship
- lines serve as links



Examples: flow charts, family trees, network diagrams

Maps

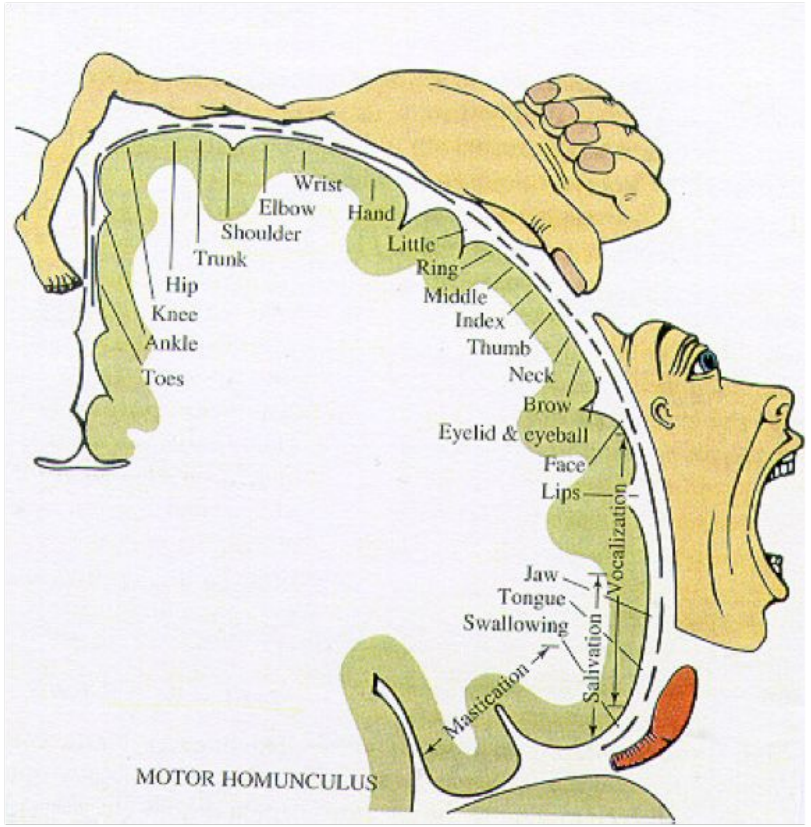
- internal relations determined (in part) by the spatial relations of what is pictured
- labels paired with locations



Examples: physical maps, topographic maps, political maps, maps of census data

Diagrams

- schematic pictures of objects or entities
- parts are *symbolic* (unlike photographs)



Examples: how-to illustrations, figures in a manual

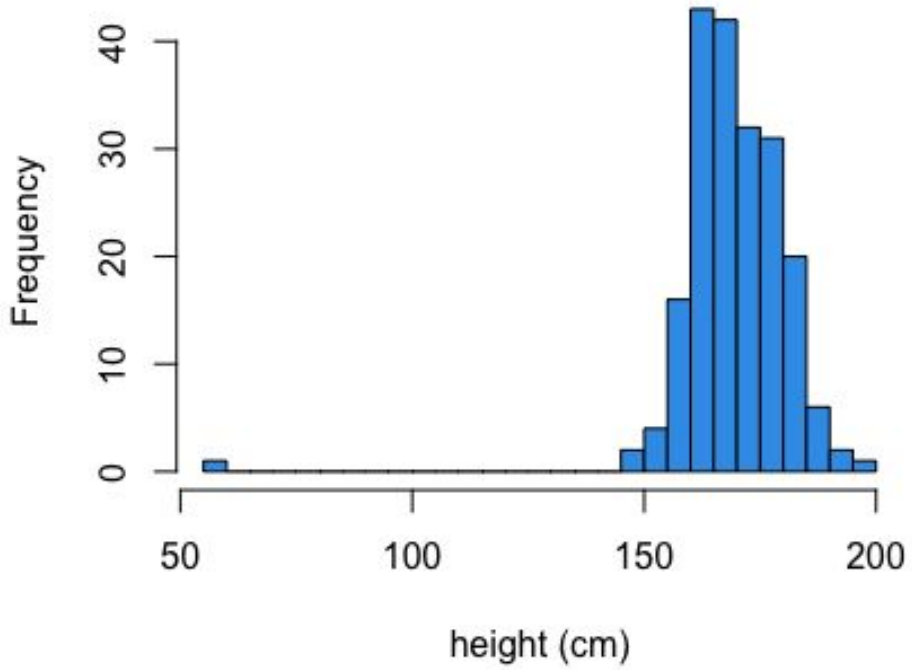
Data Visualization Types: Graphs & Tables

A portion of the Davis dataset we'll be working with

	sex	weight	height	repwt	repht
1	M	77	182	77	180
2	F	58	161	51	159
3	F	53	161	54	158
4	M	68	177	70	175
5	F	59	157	59	155
6	M	76	170	76	165
7	M	76	167	77	165
8	M	69	186	73	180
9	M	71	178	71	175
10	M	65	171	64	170
11	M	70	175	75	174
12	F	51	161	52	158
13	F	64	168	64	165
14	F	52	163	57	160
15	F	65	166	66	165
16	M	92	187	101	185
17	F	62	168	62	165
18	M	76	197	75	200
19	F	61	175	61	171
20	M	119	180	124	178
21	F	61	170	61	170
22	M	65	175	66	173
23	M	66	173	70	170
24	F	54	171	59	168

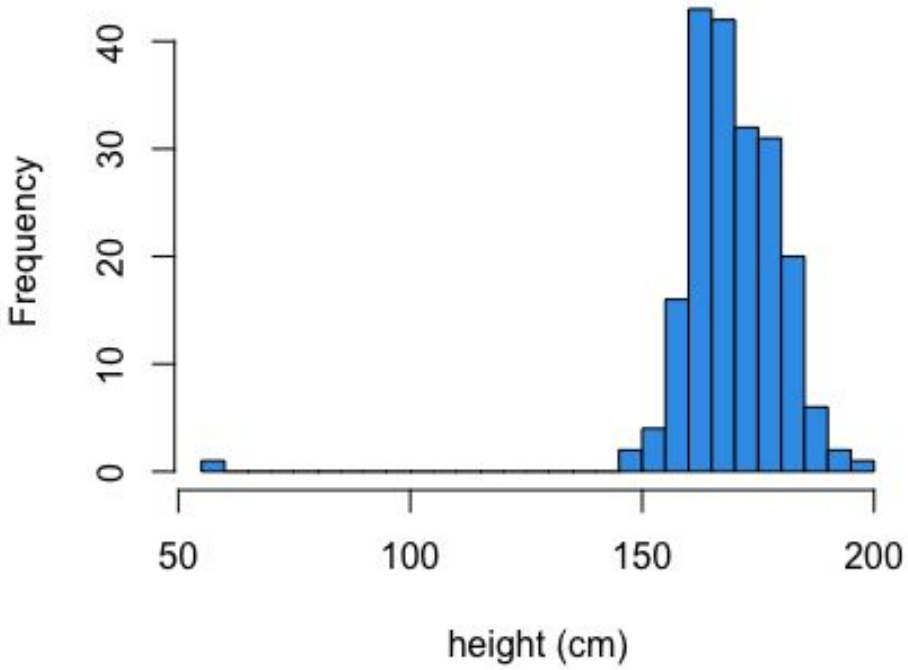
Histograms

Information about a single set of numbers



Histograms

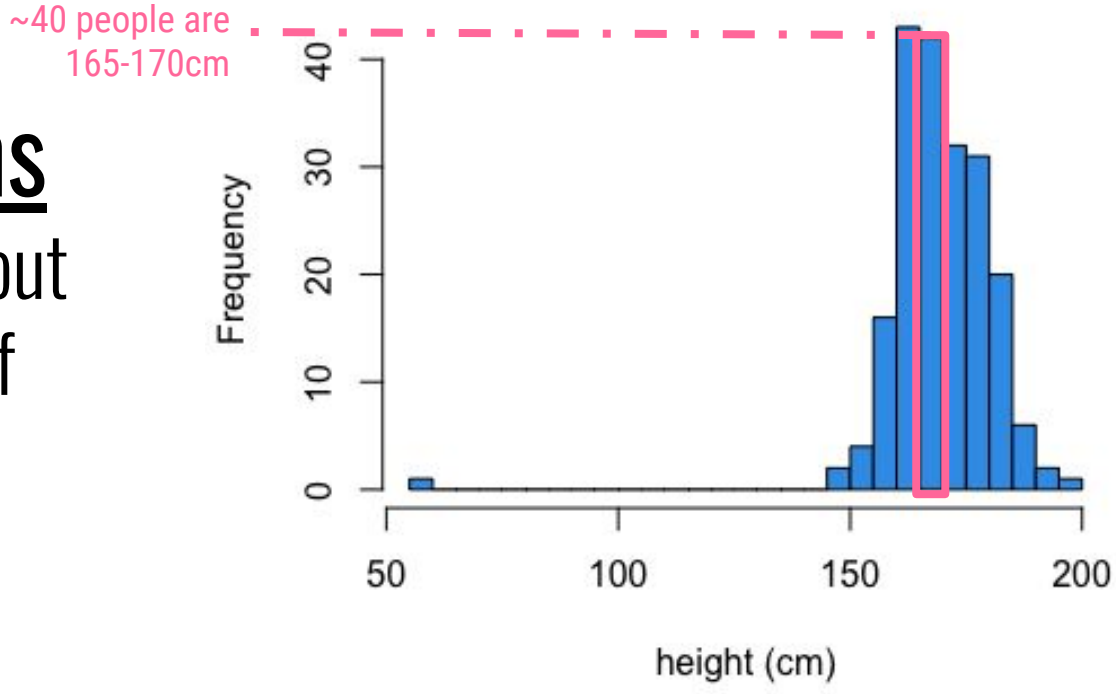
Information about a single set of numbers



Range of possible height values is easily visualized

Histograms

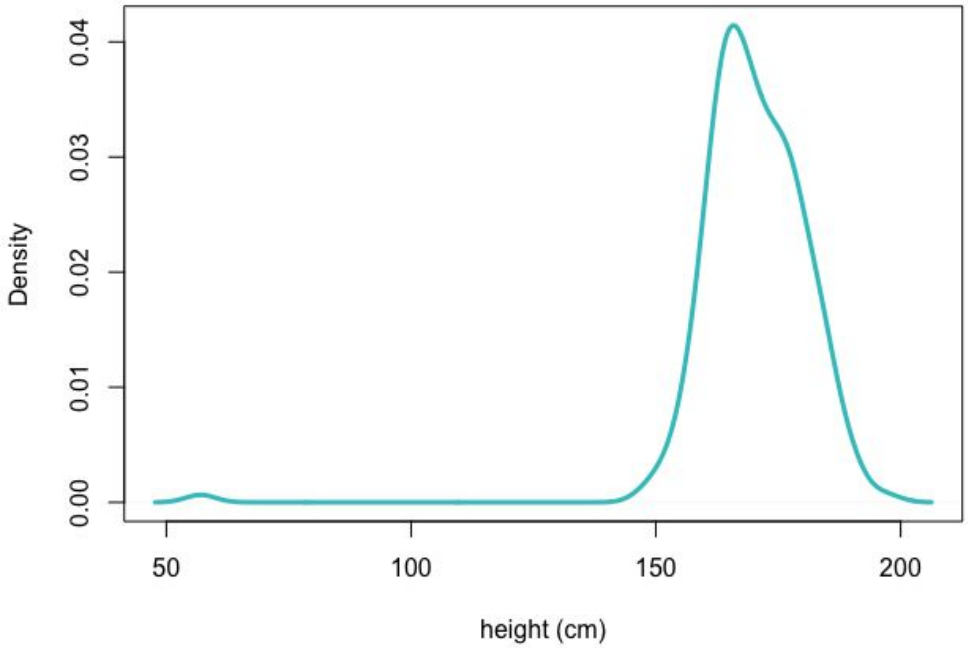
Information about a single set of numbers



Range of possible height values is easily visualized

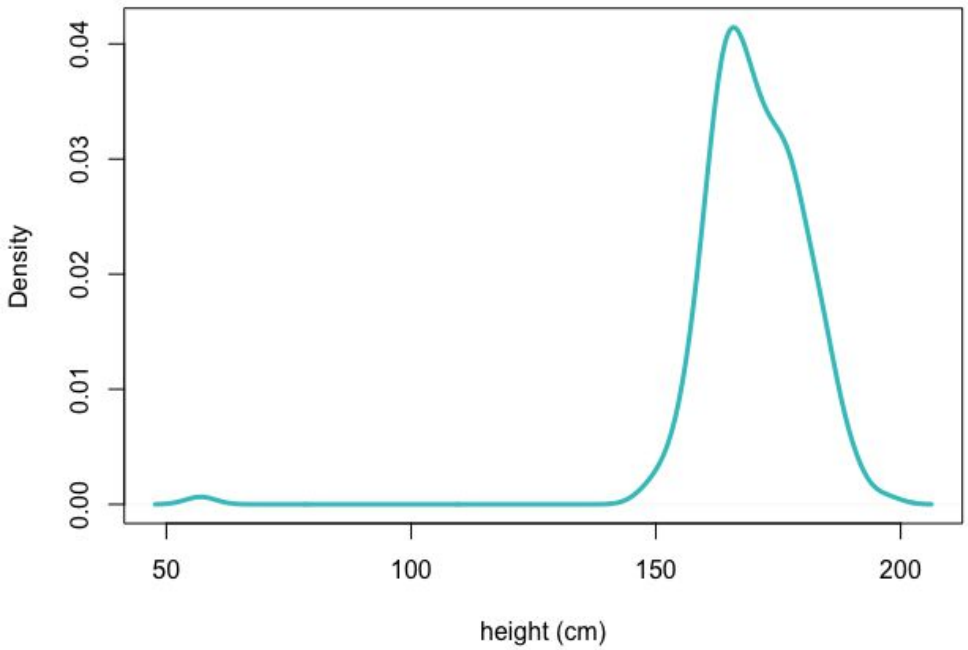
Densityplot

Information about a single set of numbers



Densityplot

Information about a single set of numbers



Demonstrates the *distribution* of the data
(A smoothed version of a histogram)

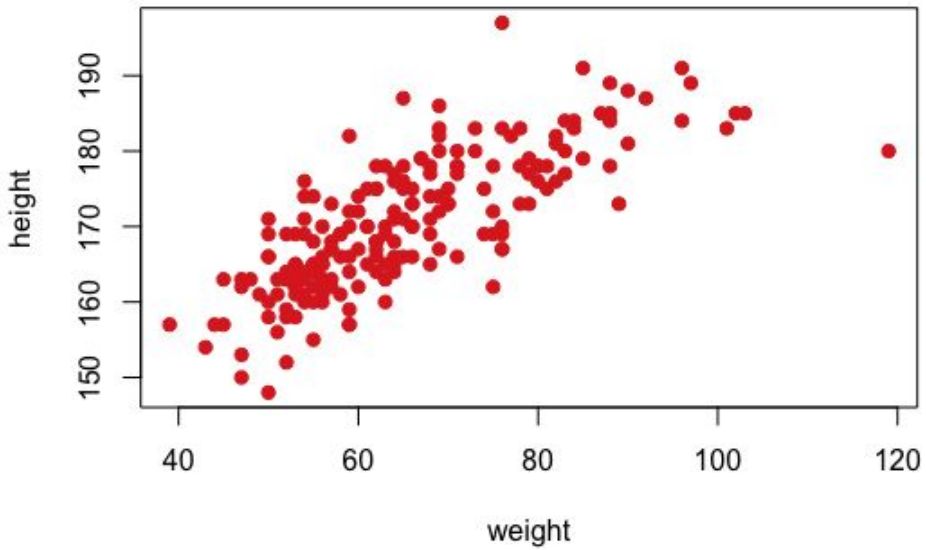
Densityplot

Information about a single set of numbers

Demonstrates the *distribution* of the data
And helps to identify extreme values

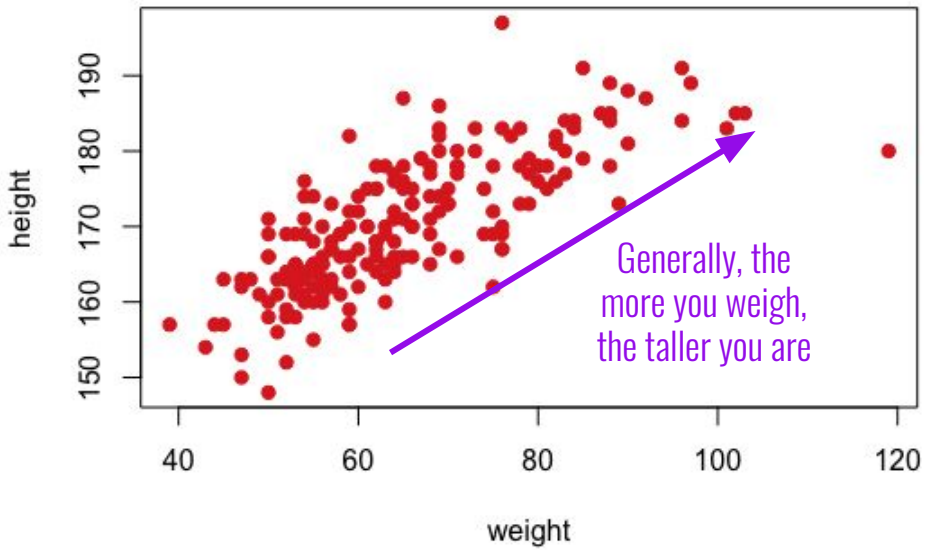
Scatterplot

Relationship between two numerical variables



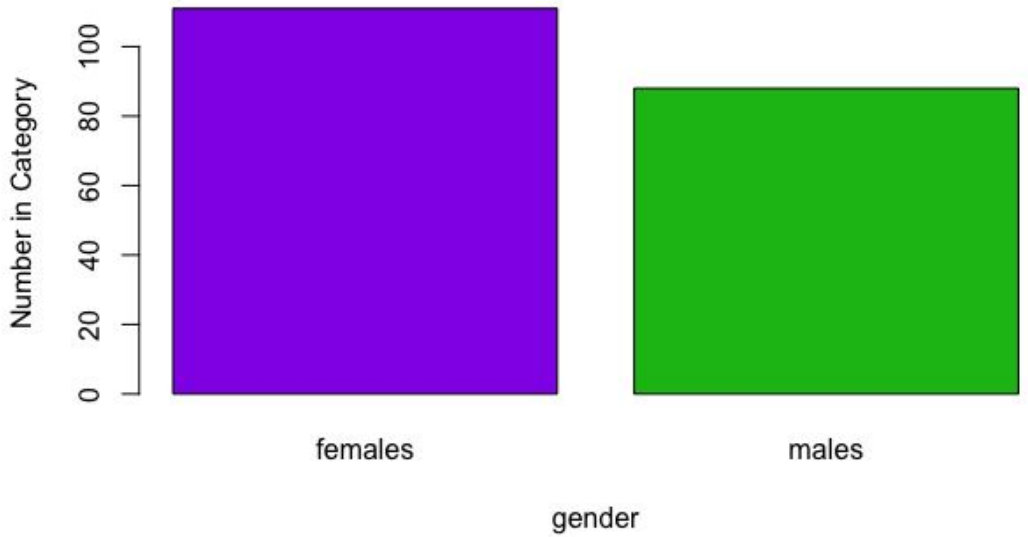
Scatterplot

Relationship between two numerical variables



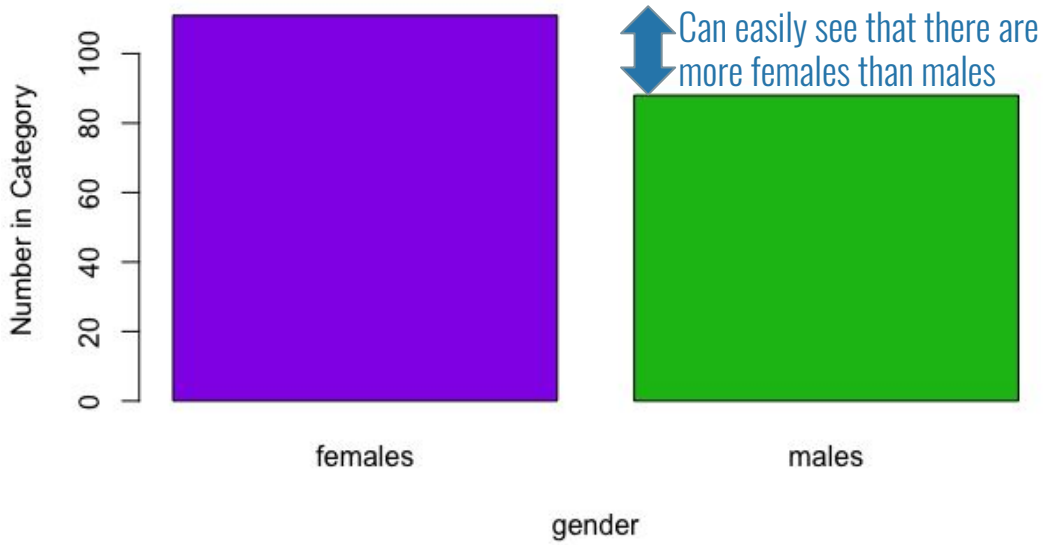
Barplot

Count of levels within a categorical variable



Barplot

Count of levels within a categorical variable



N=54

5 sets of 10 values were plotted

- 20 graphs (10 pie, 10 bar)

Task: judge which segment or bar is largest

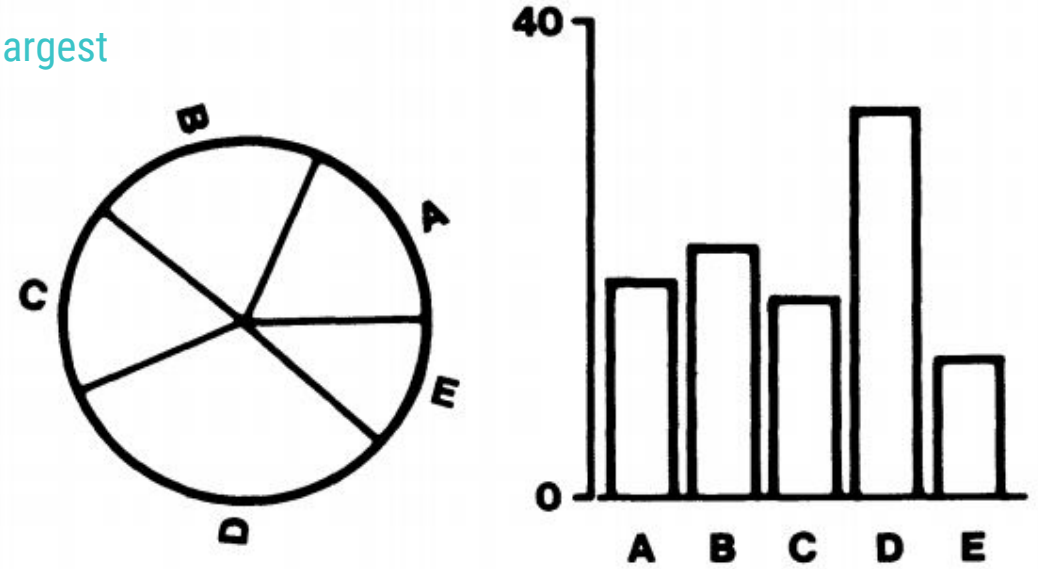
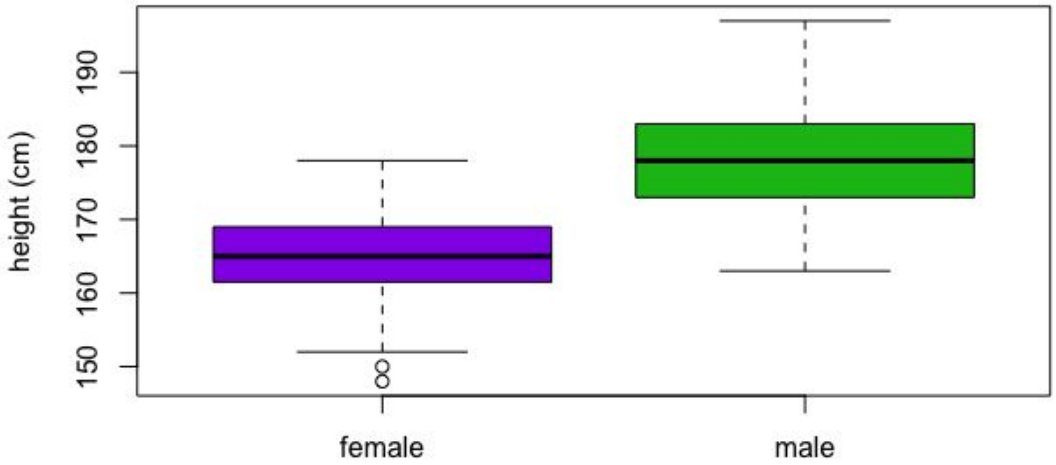


Figure 3. Graphs from position–angle experiment.

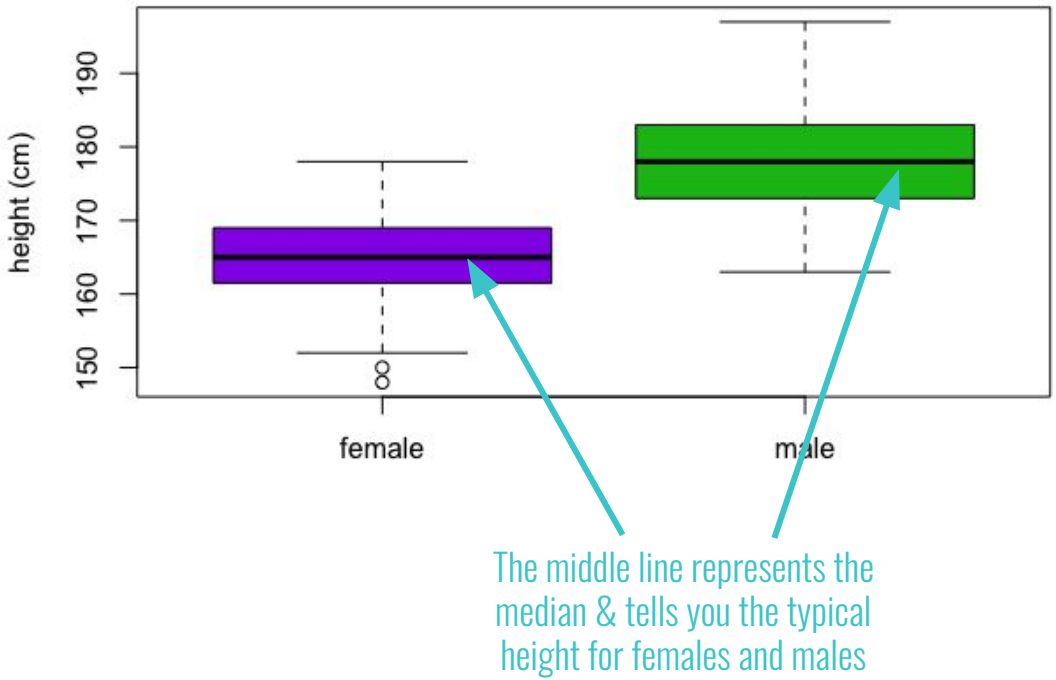
Boxplot

Summary of numerical values across categories



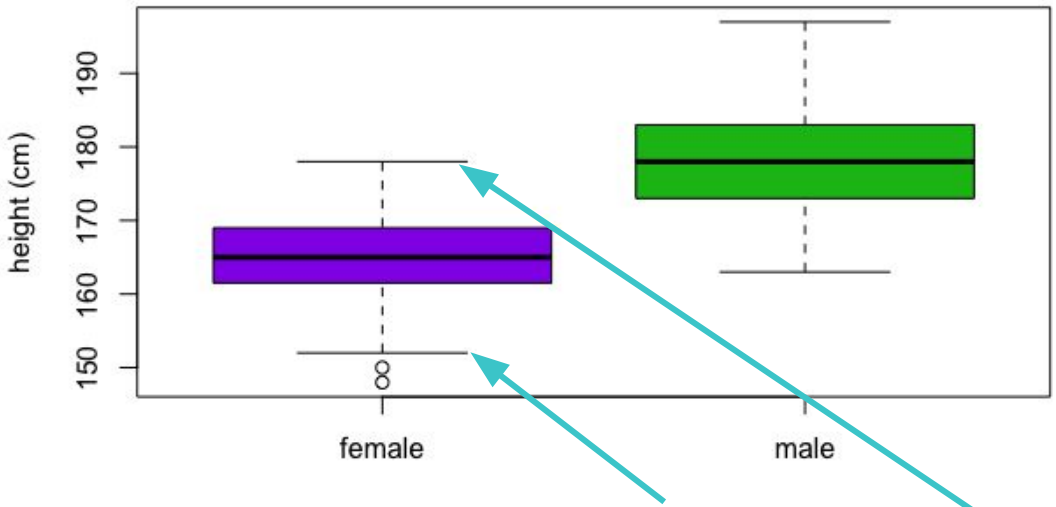
Boxplot

Summary of numerical values across categories



Boxplot

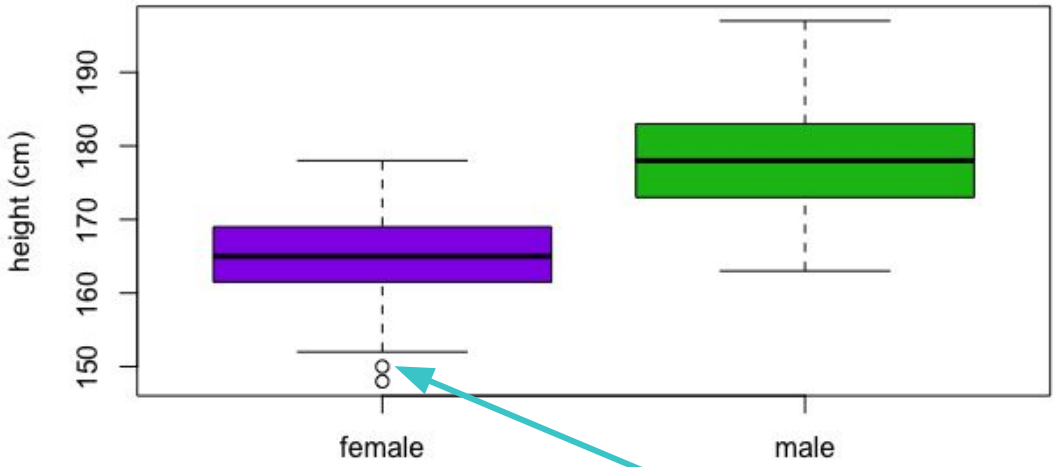
Summary of numerical values across categories



The lines give you an idea of the typical range of values for each category

Boxplot

Summary of numerical values across categories



Values outside the typical range are shown as circles. These are known as **outliers**.

Tables

Effective ways to display data summaries

	Number	Percentage (%)
male	111	55.8
female	88	44.2

Table 1: Gender Breakdown Across Davis Dataset

Tables count as data visualizations!

Visualization Best Practices

Saliency

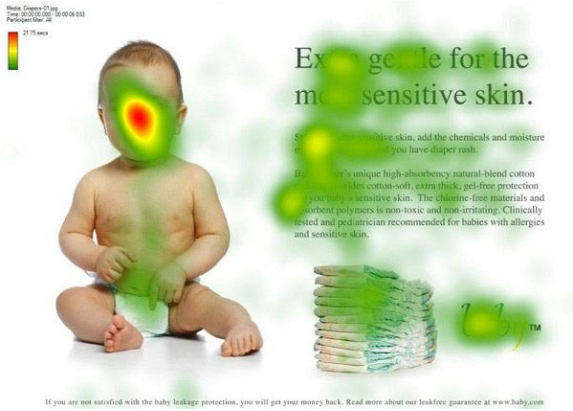

21.75 secs
sensitive skin.
sensitive skin, add the chemicals and moisture
you have diaper rash.
's unique high-absorbency natural-blend cotton
ides cotton-soft, extra thick, gel-free protection
you baby's sensitive skin. The chlorine-free materials and
sorbent polymers is non-toxic and non-irritating. Clinically
tested and pediatrician recommended for babies with allergies
and sensitive skin.
If you are not satisfied with the baby leakage protection, you will get your money back. Read more about our leakfree guarantee at www.baby.com

Saliency

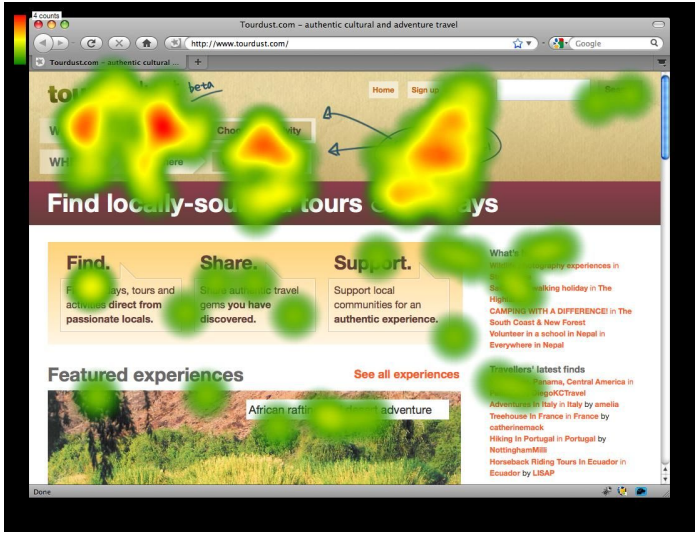

sensitive skin.
If you are not satisfied with the baby leakage protection, you will get your money back. Read more about our leakfree guarantee at www.baby.com


NEW
GRILL
GRILL

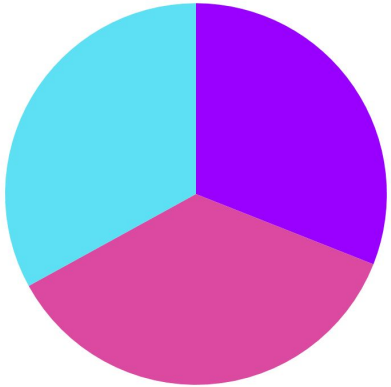
Saliency

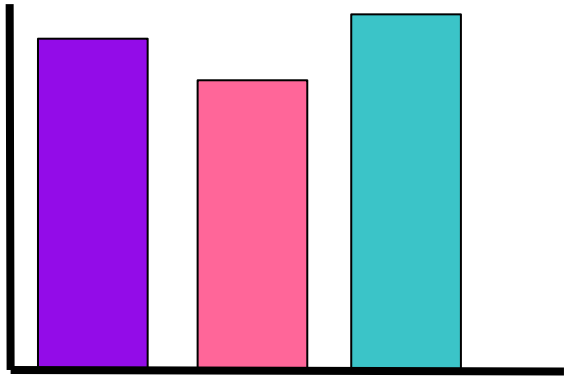

sensitive skin.
If you are not satisfied with the baby leakage protection, you will get your money back. Read more about our leakfree guarantee at www.baby.com


NEW
GRILL
GRILL


Tourdust.com – authentic cultural and adventure travel
http://www.tourdust.com/
Find.
Share.
Support.
Support local communities for an authentic experience.
Featured experiences
See all experiences

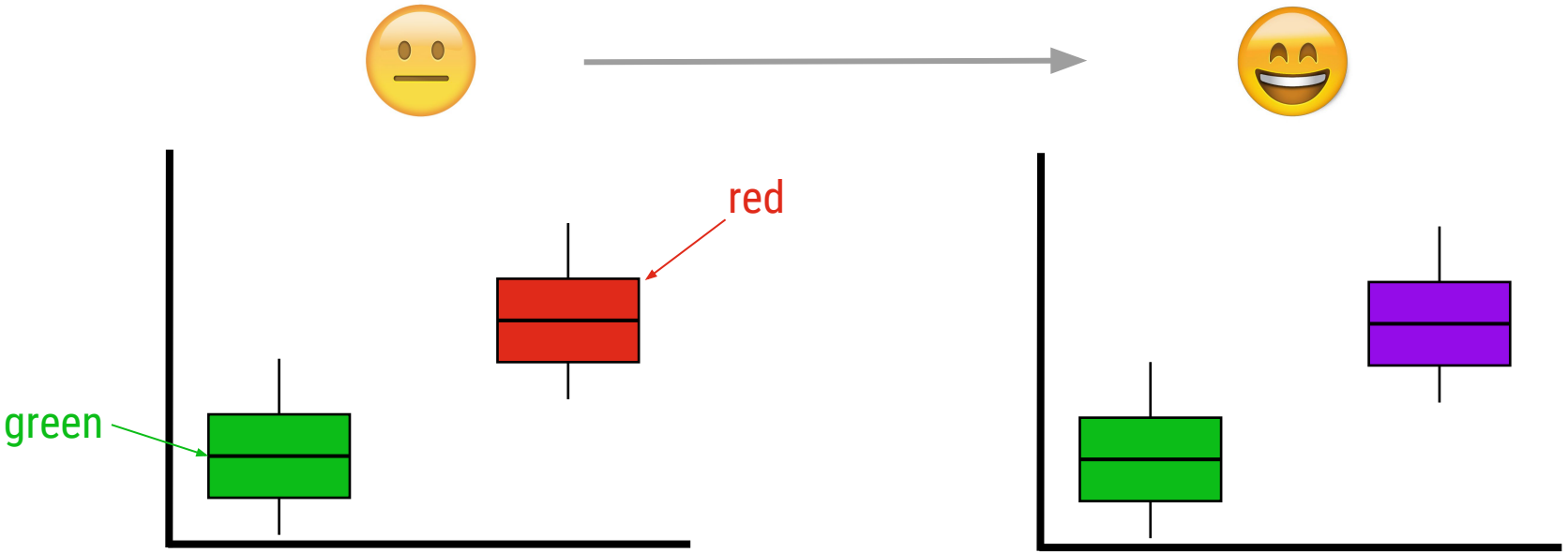
Choose the right type of plot





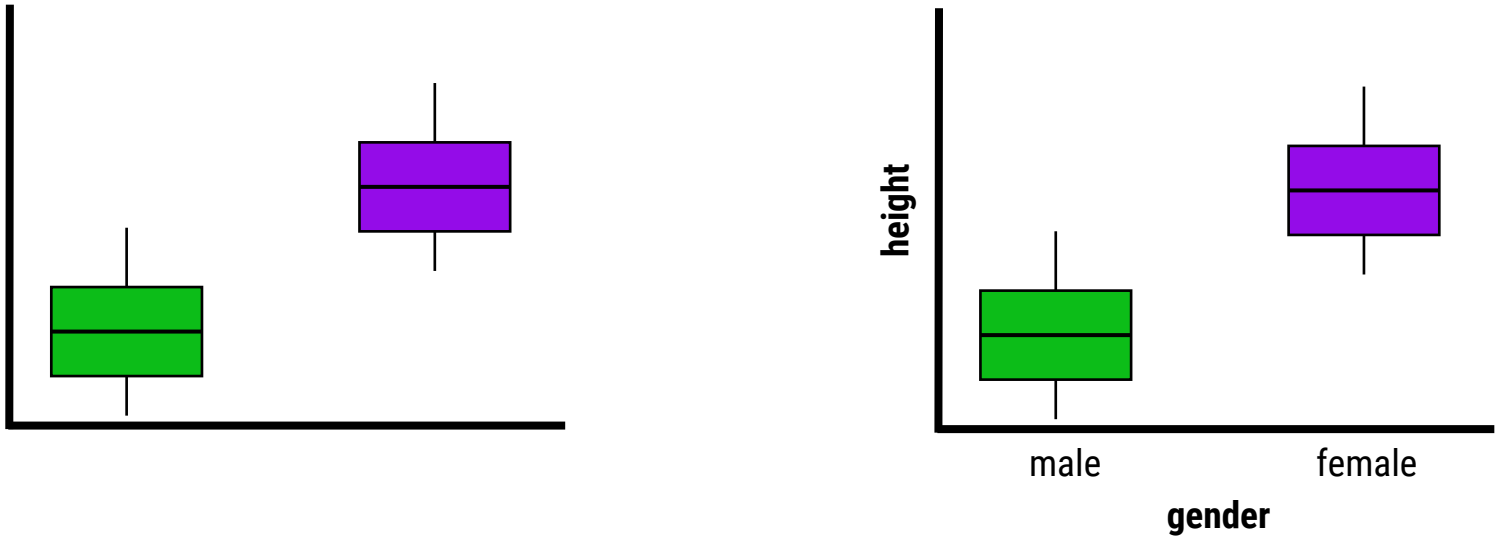
When looking at values, bar charts make it much easier to see the difference between groups!

Be mindful when choosing colors

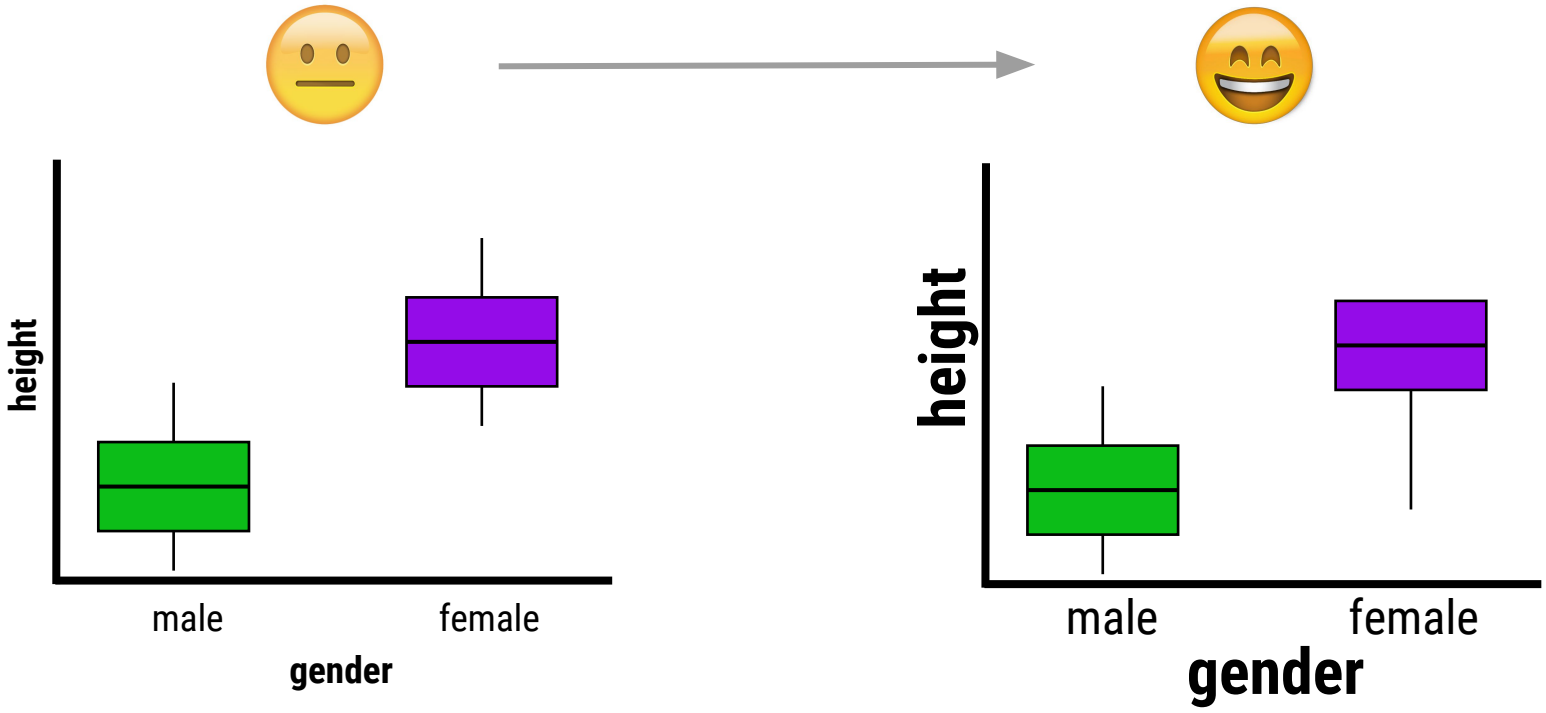


Many color-blind individuals cannot see the difference between red and green.

Label your axes!

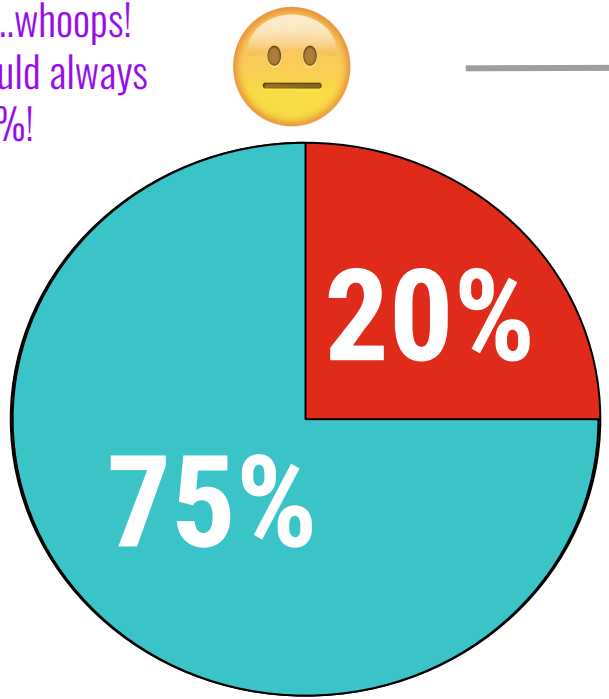


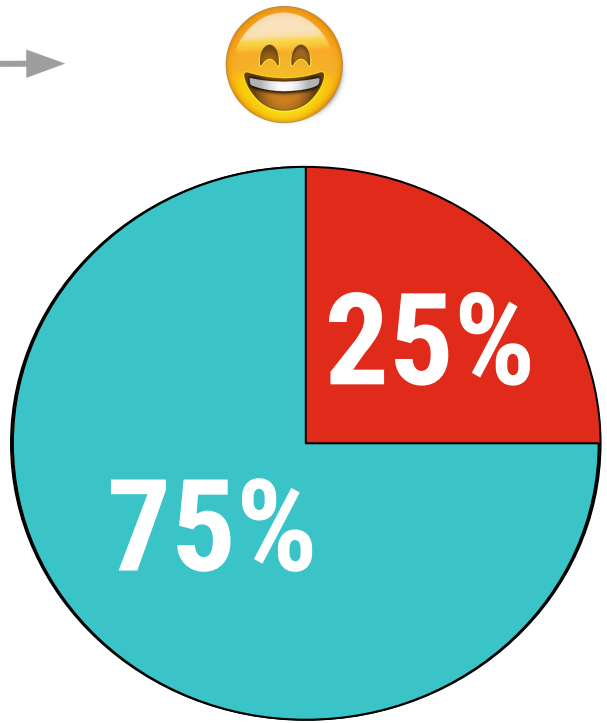
Make sure the text size is big enough!



Make sure your numbers add up!

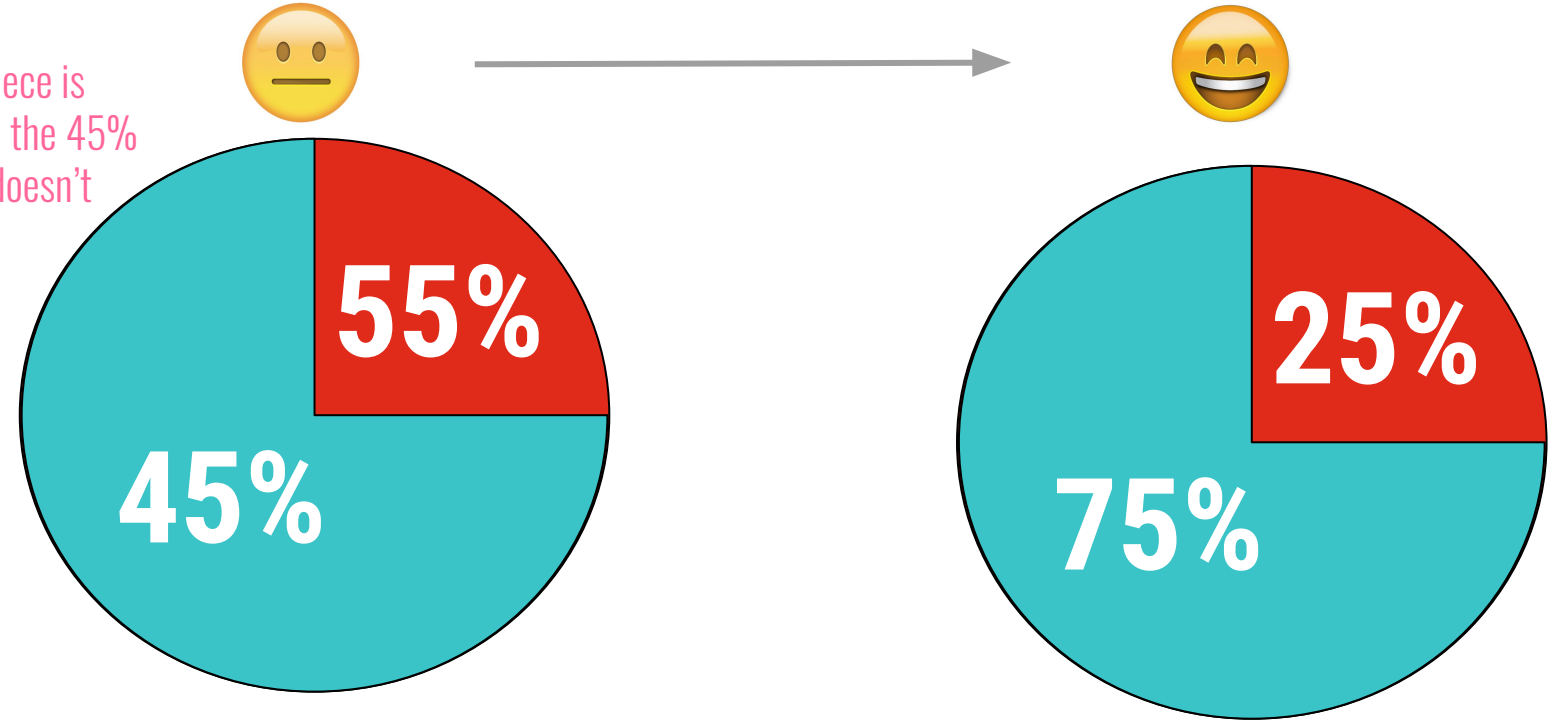
75 + 20 = 95 ...whoops!
Pie charts should always add up to 100%!





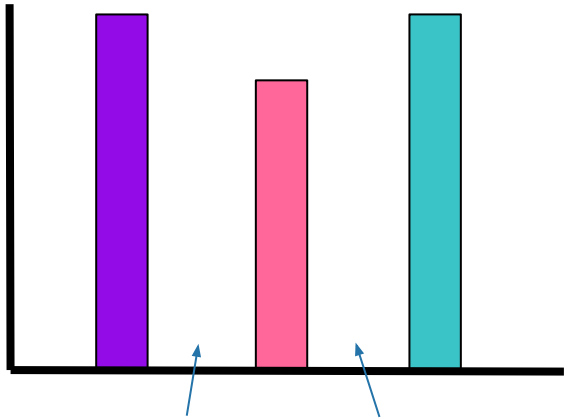
Make sure the numbers and graphic represent the data

That 55% piece is smaller than the 45% piece. That doesn't make sense!

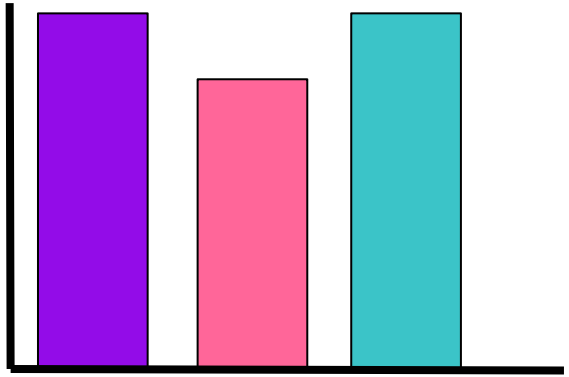


Make comparisons easy on your readers

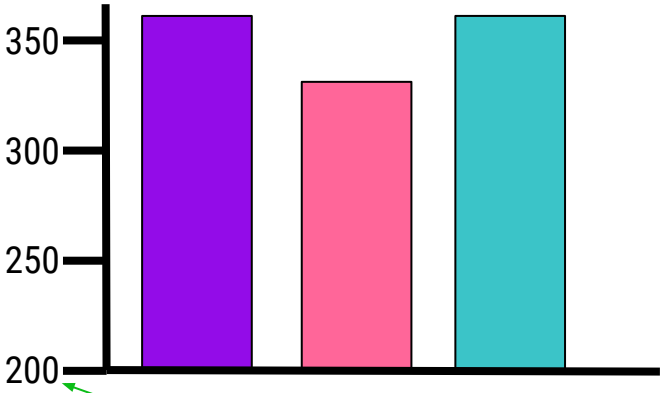
Avoid unnecessary whitespace



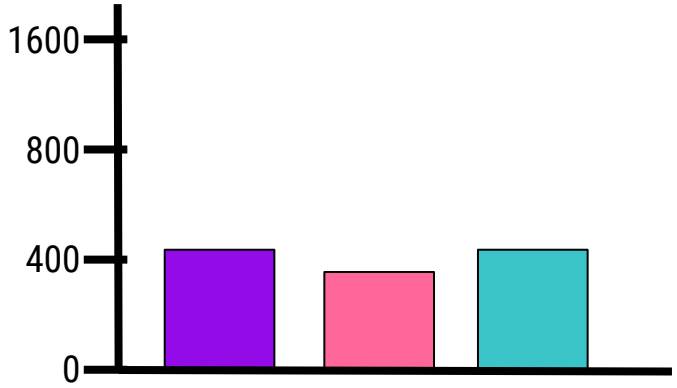
When the bars are spread out,
there is unnecessary whitespace



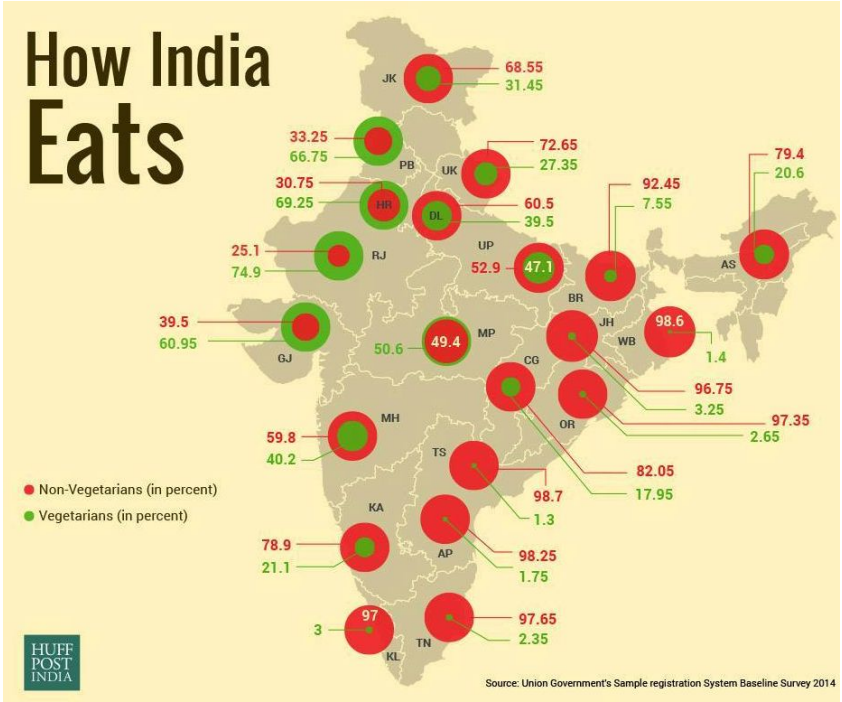
Use y-axes that start at 0 for barplots



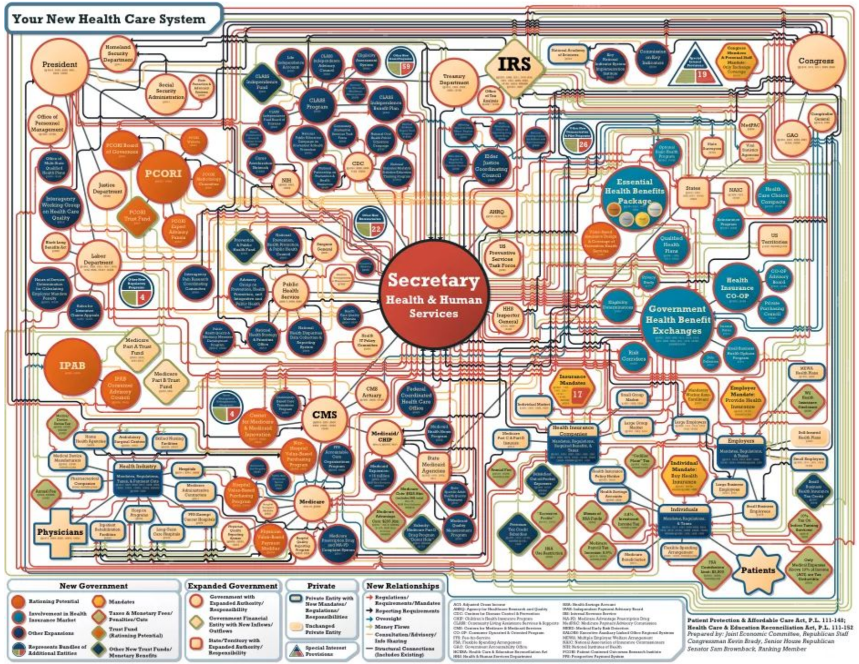
The y-axis starts at 200.
This is misleading & makes differences look larger than they actually are



Keep it Simple



Keep it Simple



https://www.jec.senate.gov/public/index.cfm/republicans/committeenews?ID=bb302d88-3d0d-4424-8e33-3c5d2578c2b0

Allow viewer to make comparison top to bottom

Airport Name	Cleveland Hopkins Intl	William P Hobby	Metropolitan Oakland Intl	San Francisco Intl	Bellingham Intl
Airport Code	CLE	HOU	OAK	SFO	BLI
Mean Arrival Delay	26.15	10.25	10.07	8.86	8.67

Avoid comparisons across rows

Airport Name	**Airport Code**	**Mean Arrival Delay**
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

Compare down columns

Order rows logically

Airport Name	Airport Code	Mean Arrival Delay
Metropolitan Oakland Intl	OAK	10.07
Bellingham Intl	BLI	8.67
	CLE	26.15
San Francisco Intl	SFO	8.86
William P Hobby	HOU	10.25

Airport Name	Airport Code	Mean Arrival Delay
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

longest delay

shortest delay

Order columns logically

Mean Arrival Delay	Airport Code	Airport Name
26.15	CLE	
10.25	HOU	William P Hobby
10.07	OAK	Metropolitan Oakland Intl
8.86	SFO	San Francisco Intl
8.67	BLI	Bellingham Intl

Order sensically from left to right

Airport Name	Airport Code	Mean Arrival Delay
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

Limit the number of rows and columns

	UrbanPop	Rape
Alabama	58	21.2
Alaska	48	44.5
Arizona	80	31.0
Arkansas	50	19.5
California	91	40.6
Colorado	78	38.7
Connecticut	77	11.1
Delaware	72	15.8
Florida	80	31.9
Georgia	60	25.8
Hawaii	83	20.2
Idaho	54	14.2
Illinois	83	24.0
Indiana	65	21.0
Iowa	57	11.3
Kansas	66	18.0
Kentucky	52	16.3
Louisiana	66	22.2
Maine	51	7.8
Maryland	67	27.8
Massachusetts	85	16.3
Michigan	74	35.1
Minnesota	66	14.9
Mississippi	44	17.1
Missouri	70	28.2
Montana	53	16.4
Nebraska	62	16.5
Nevada	81	46.0
New Hampshire	56	9.5
New Jersey	89	18.8
New Mexico	70	32.1
New York	86	26.1
North Carolina	45	16.1
North Dakota	44	7.3
Ohio	75	21.4
Oklahoma	68	20.0
Oregon	67	29.3
Pennsylvania	72	14.9
Rhode Island	87	8.3
South Carolina	48	22.5
South Dakota	45	12.8
Tennessee	59	26.9
Texas	80	25.5
Utah	80	22.9
Vermont	32	11.2
Virginia	63	20.7
Washington	73	26.2
West Virginia	39	9.3
Wisconsin	66	10.8
Wyoming	60	15.6

Too many rows!

Airport Name	Airport Code	Mean Arrival Delay
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

Include informative labels

Needs better labels!

AN	AC	MAD
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

Airport Name	Airport Code	Mean Arrival Delay
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

Be mindful of significant digits

Airport Name	Airport Code	Mean Arrival Delay
	CLE	26.15145
William P Hobby	HOU	10.258923
Metropolitan Oakland Intl	OAK	10.07987
San Francisco Intl	SFO	8.8688173
Bellingham Intl	BLI	8.671234

Too many digits

Airport Name	Airport Code	Mean Arrival Delay
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

Format table so it can be quickly understood

Table 1: Airports with the longest delay in arrival time for planes leaving the Pacific Northwest in 2014

Airport Name	**Airport Code**	**Mean Arrival Delay**
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

Include a good caption

Table 1: Airports with the longest delay in arrival time for planes leaving the Pacific Northwest in 2014

Airport Name	**Airport Code**	**Mean Arrival Delay**
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

Include the source of the data

Table 1: Airports with the longest delay in arrival time for planes leaving the Pacific Northwest in 2014

Airport Name	**Airport Code**	**Mean Arrival Delay**
	CLE	26.15
William P Hobby	HOU	10.25
Metropolitan Oakland Intl	OAK	10.07
San Francisco Intl	SFO	8.86
Bellingham Intl	BLI	8.67

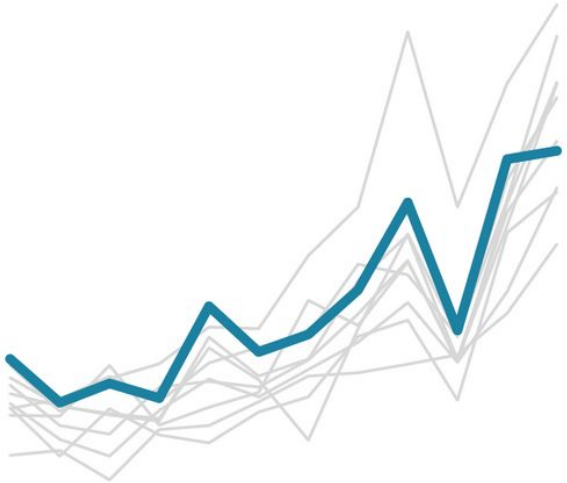
Source: Chester Ismay's `pnwflights14` R package

Color Considerations

Gray is your friend.

NOT IDEAL

BETTER

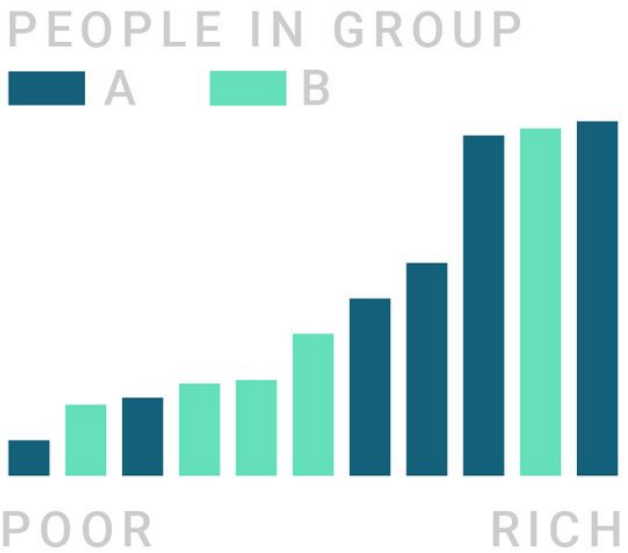


https://blog.datawrapper.de/colors/

Let the plot highlight values. Use colors for categories.

NOT IDEAL

BETTER



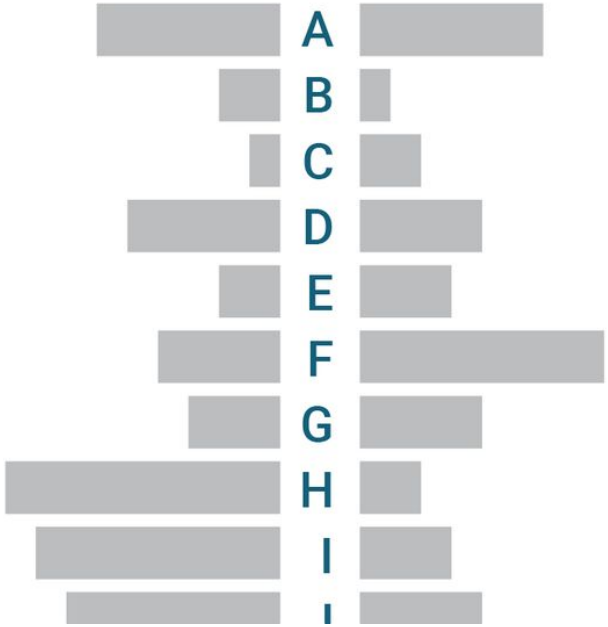
Gradients are great for patterns (i.e. population on a map), but aren't effective at highlighting your point. s. Consider showing your most important values with bars, position (like in a dot plot) or even areas, and to use colors to only show categories. Consider using

https://blog.datawrapper.de/colors/

More than 7 colors suggests you need a different chart.

NOT IDEAL

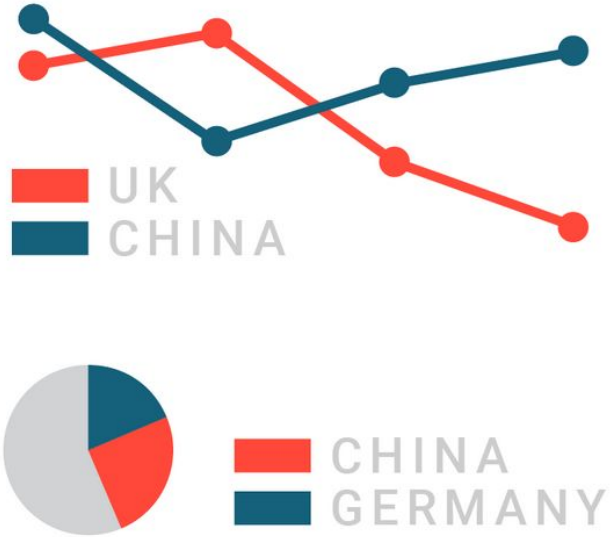
BETTER



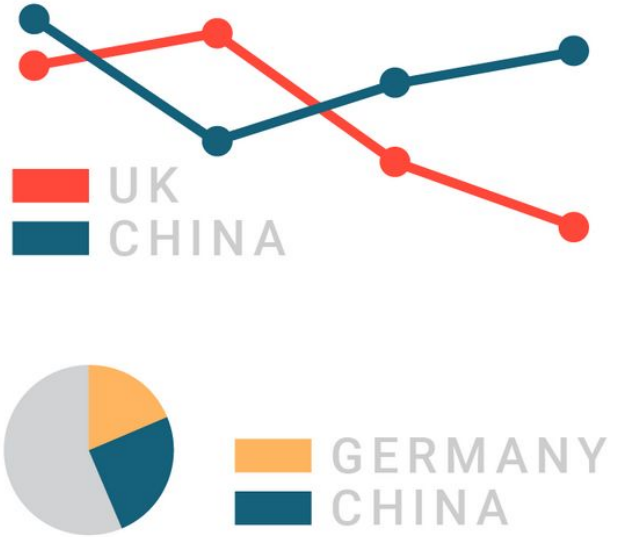
https://blog.datawrapper.de/colors/

Color consistency helps viewers.

NOT IDEAL



BETTER



https://blog.datawrapper.de/colors/

Always ensure that viewers know what the colors represent.

Choose intuitive colors

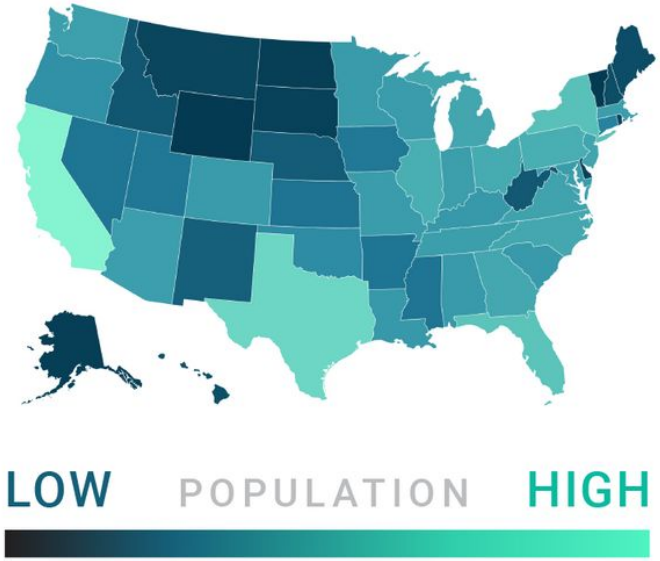
NOT IDEAL

BETTER

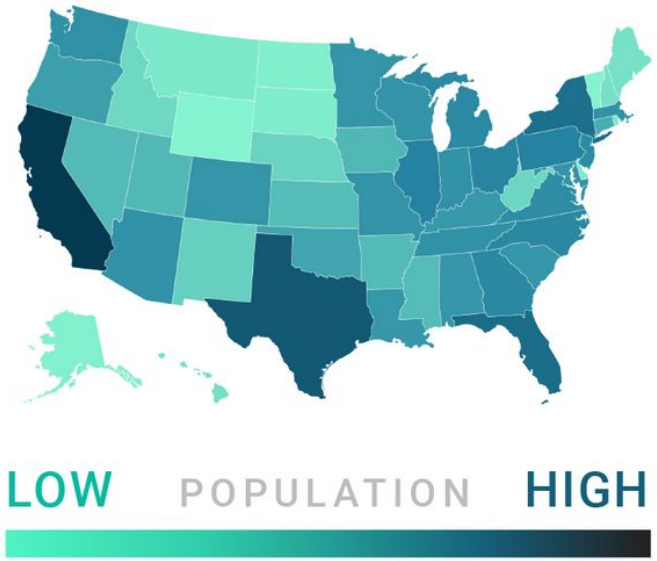
https://blog.datawrapper.de/colors/

Use light colors for low values. Dark colors for high values.

NOT IDEAL



BETTER



https://blog.datawrapper.de/colors/

Gradients are for continuous values. Distinct colors are for categories

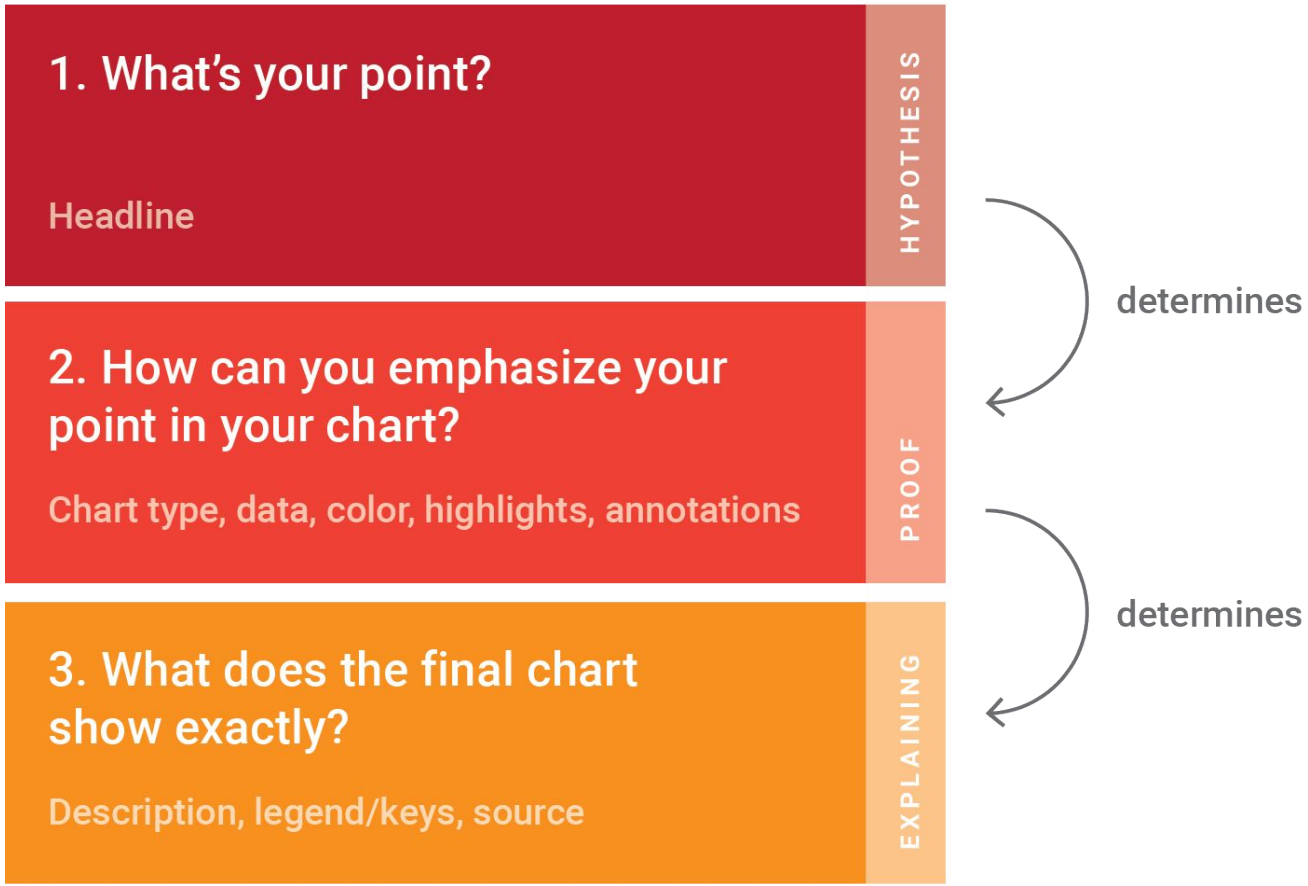
NOT IDEAL

BETTER

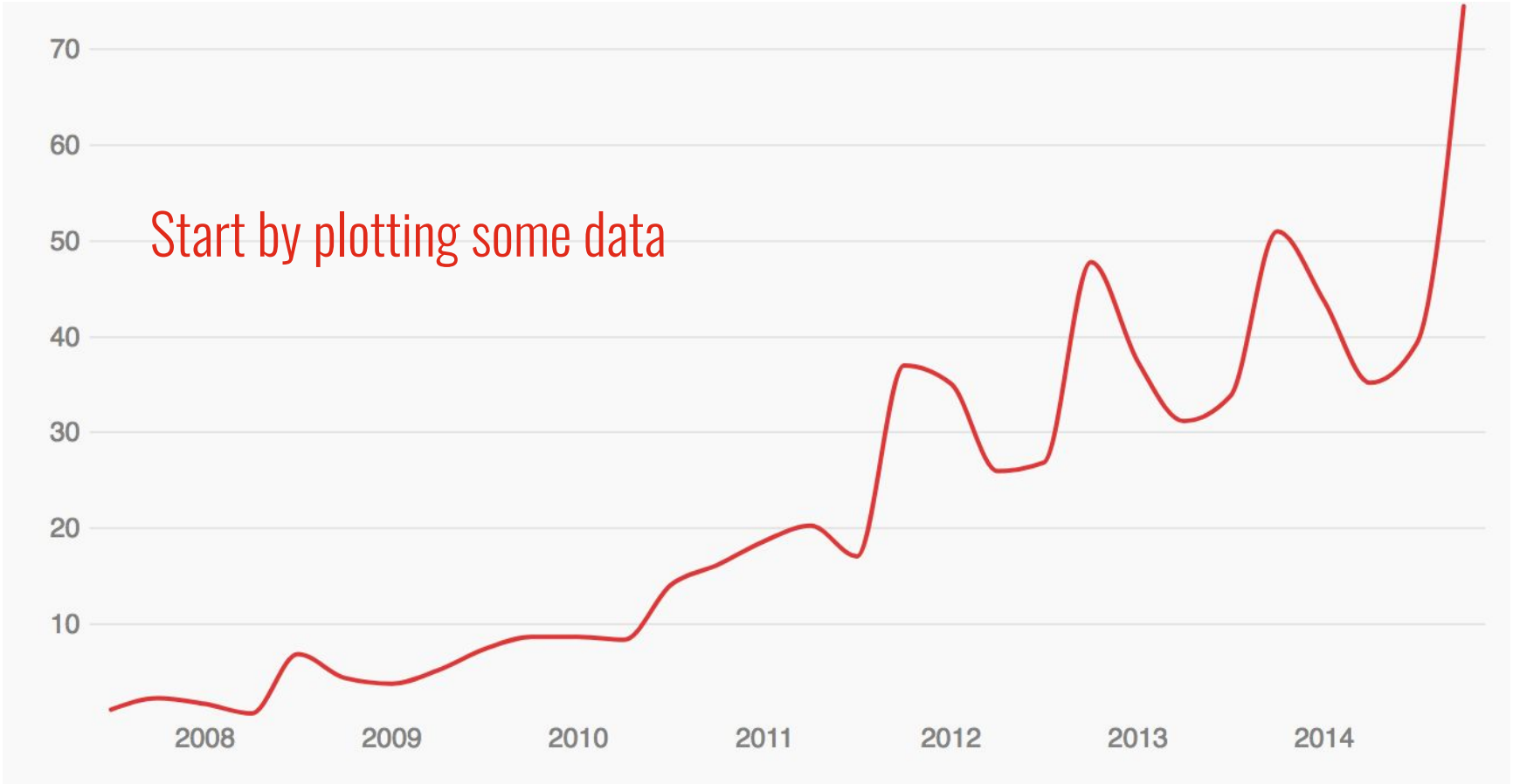
https://blog.datawrapper.de/colors/

Iterative Improvement

Three questions for creating a chart

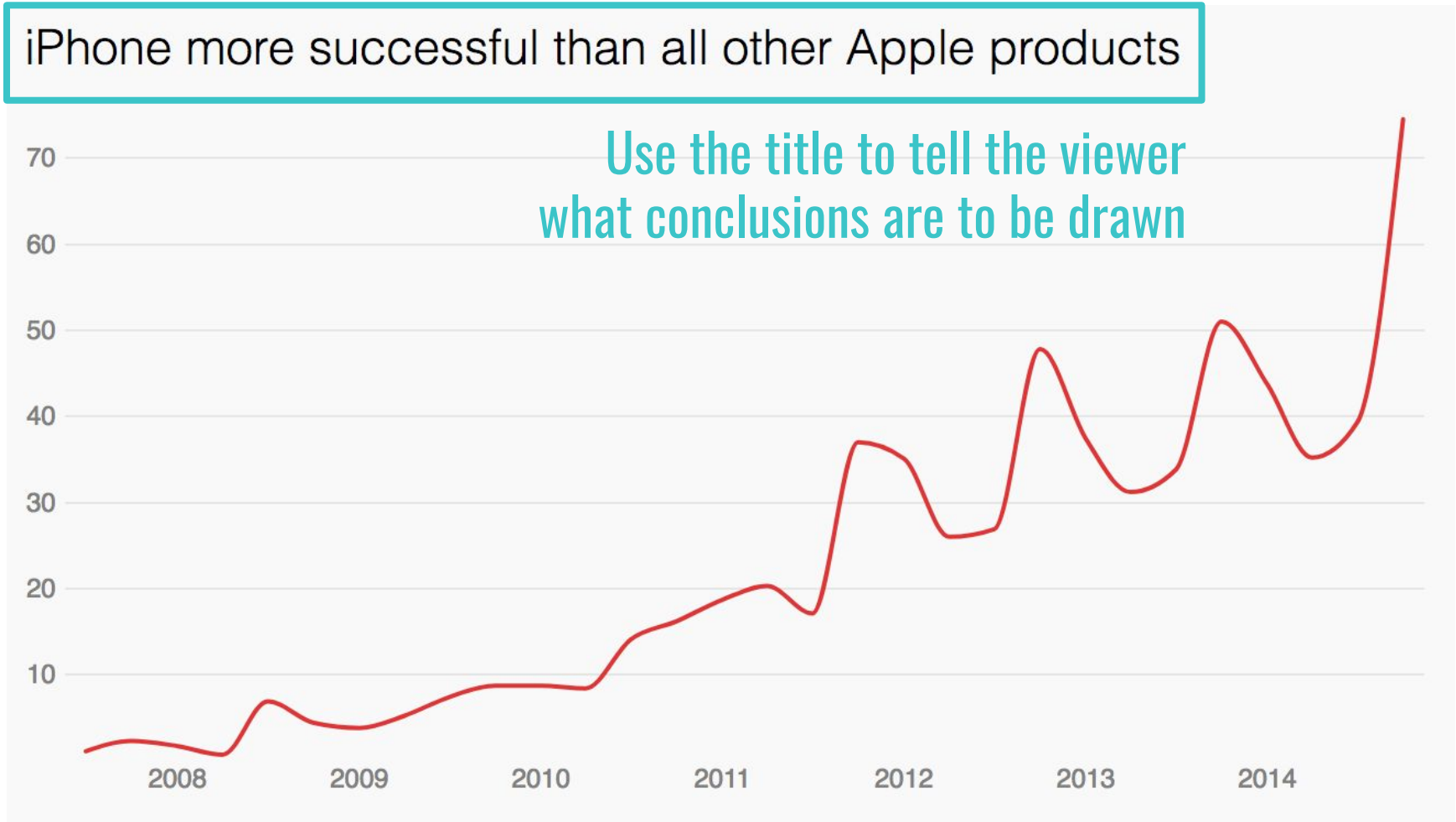


https://blog.datawrapper.de/better-charts/

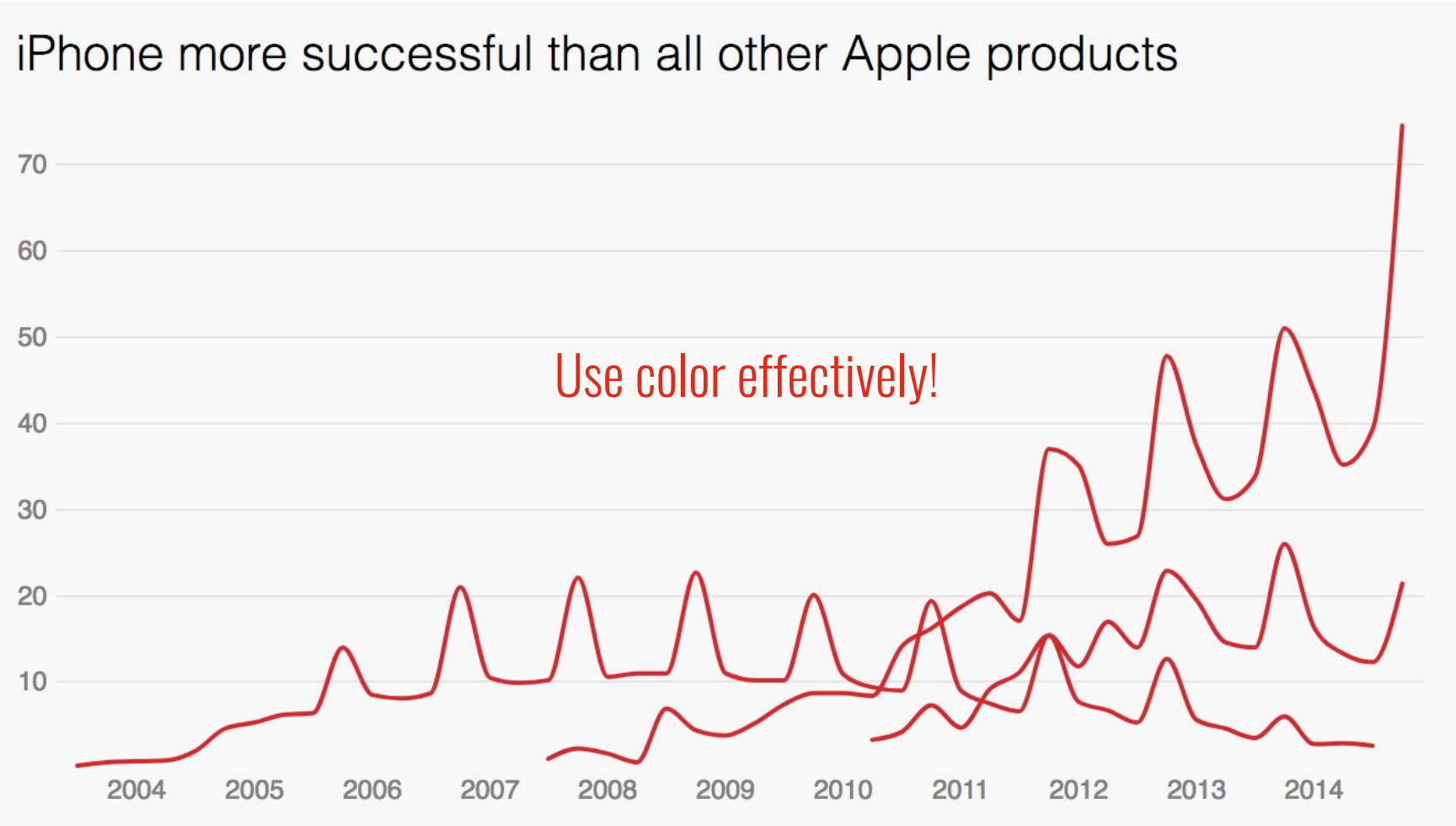


https://blog.datawrapper.de/better-charts/

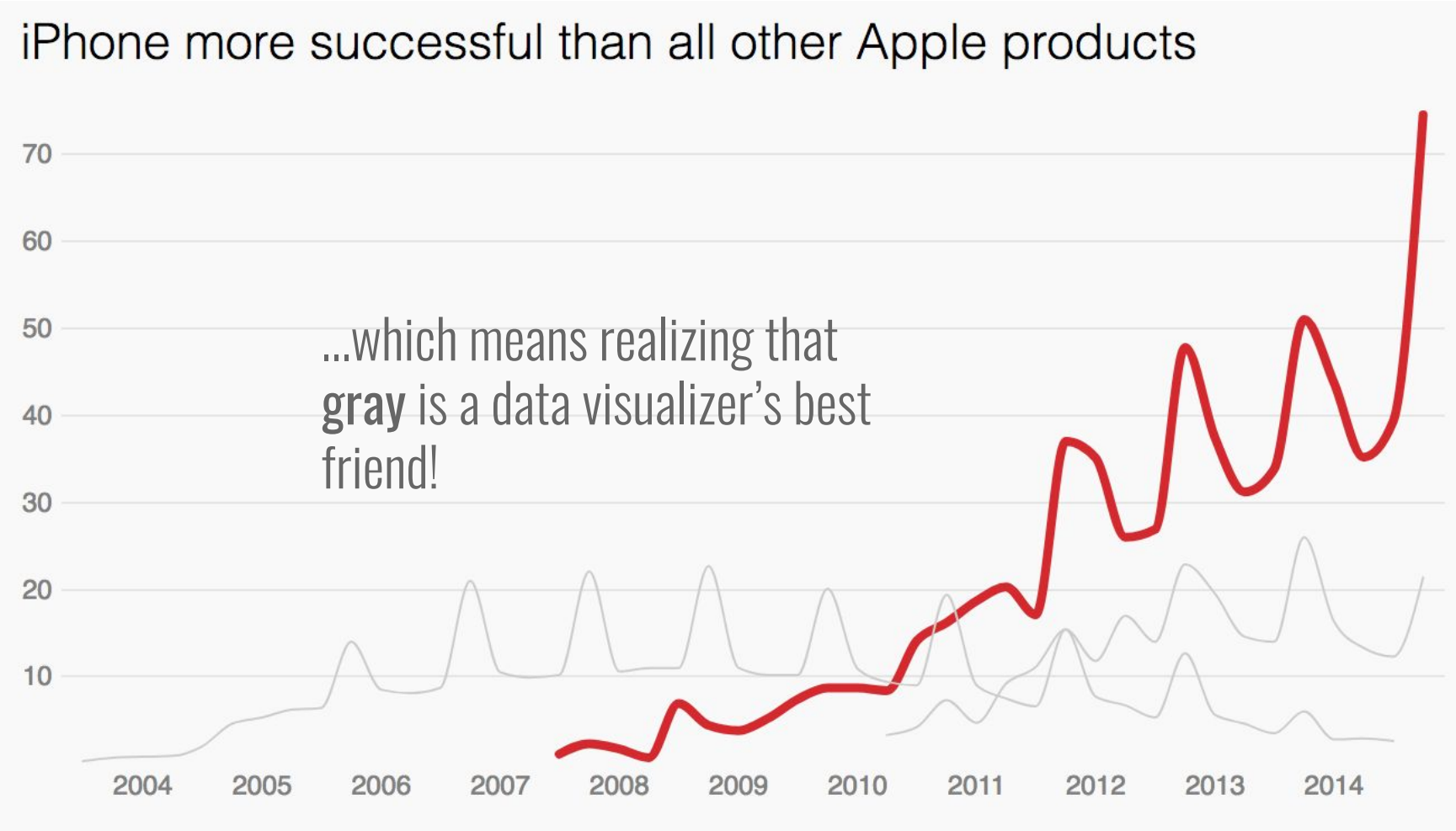
iPhone more successful than all other Apple products



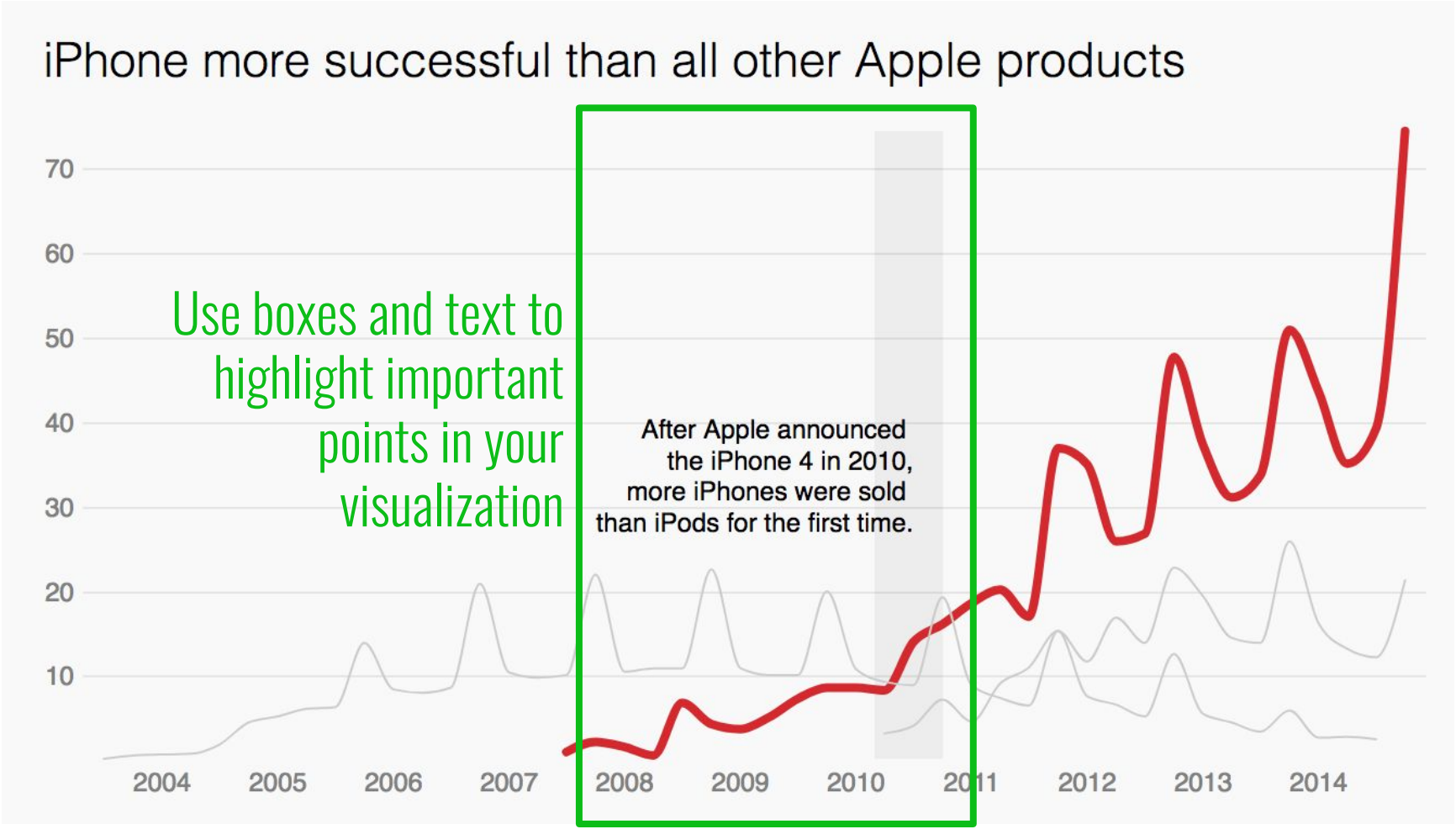
https://blog.datawrapper.de/better-charts/



https://blog.datawrapper.de/better-charts/



https://blog.datawrapper.de/better-charts/

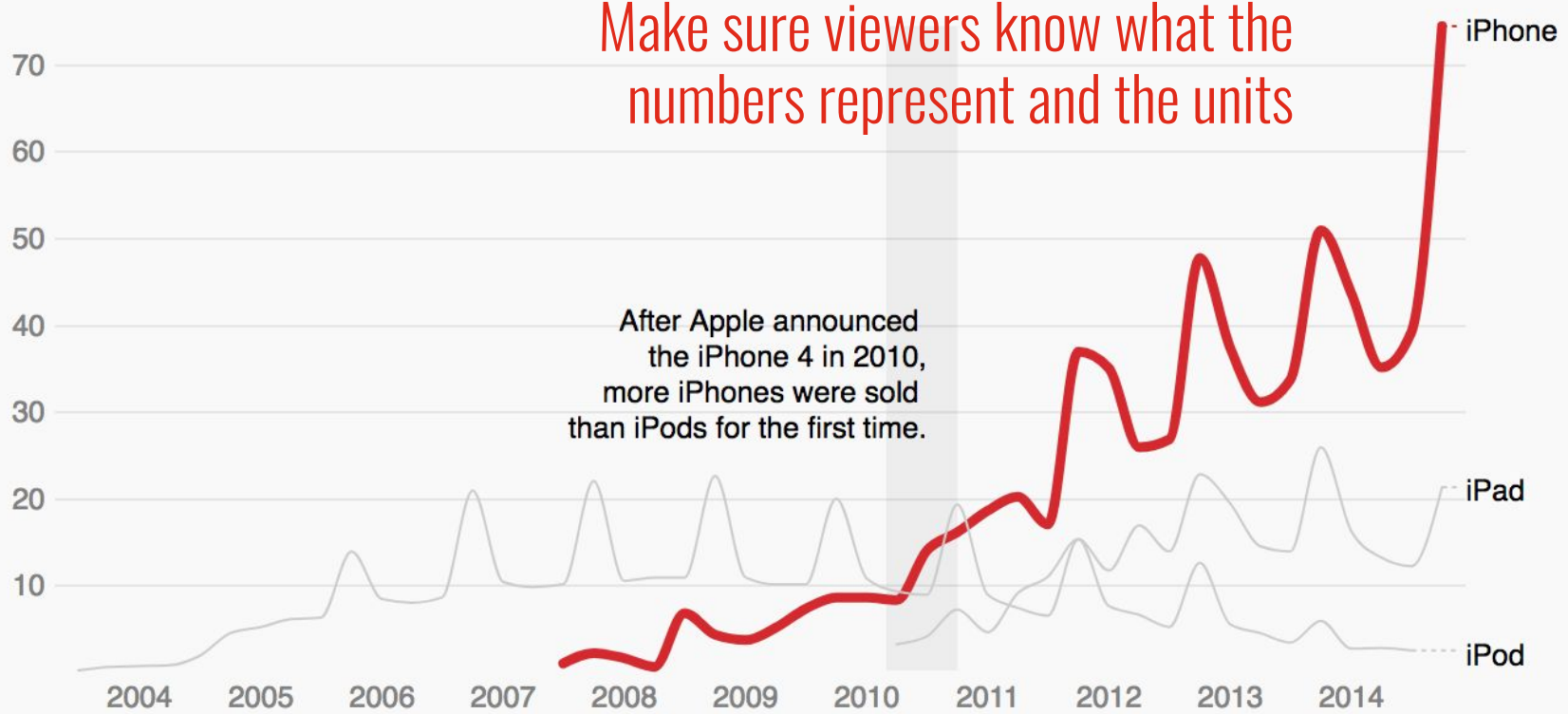


https://blog.datawrapper.de/better-charts/

iPhone more successful than all other Apple products

Worldwide sales of selected Apple products in million, by fiscal quarter, 2000 to 2014

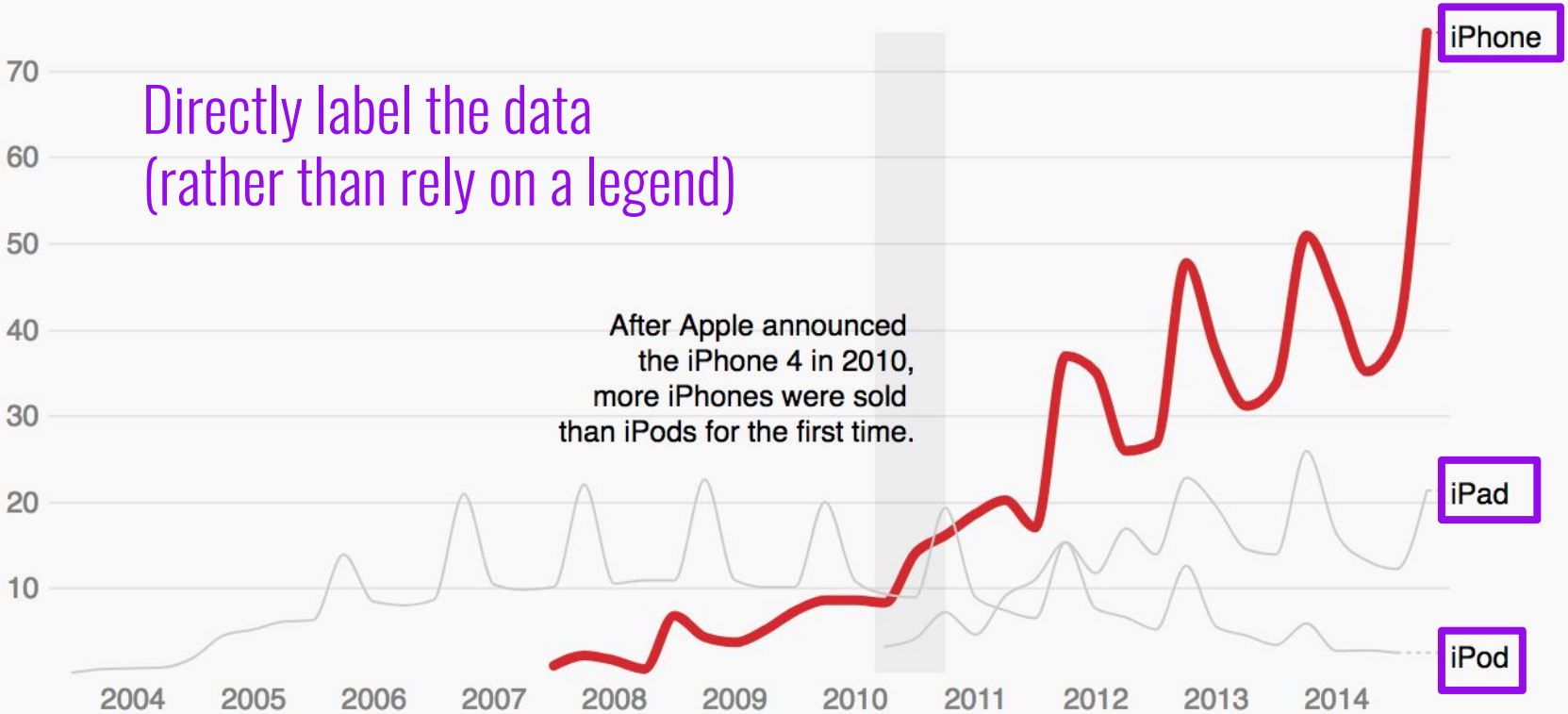
Make sure viewers know what the numbers represent and the units



https://blog.datawrapper.de/better-charts/

iPhone more successful than all other Apple products

Worldwide sales of selected Apple products in million, by fiscal quarter, 2000 to 2014



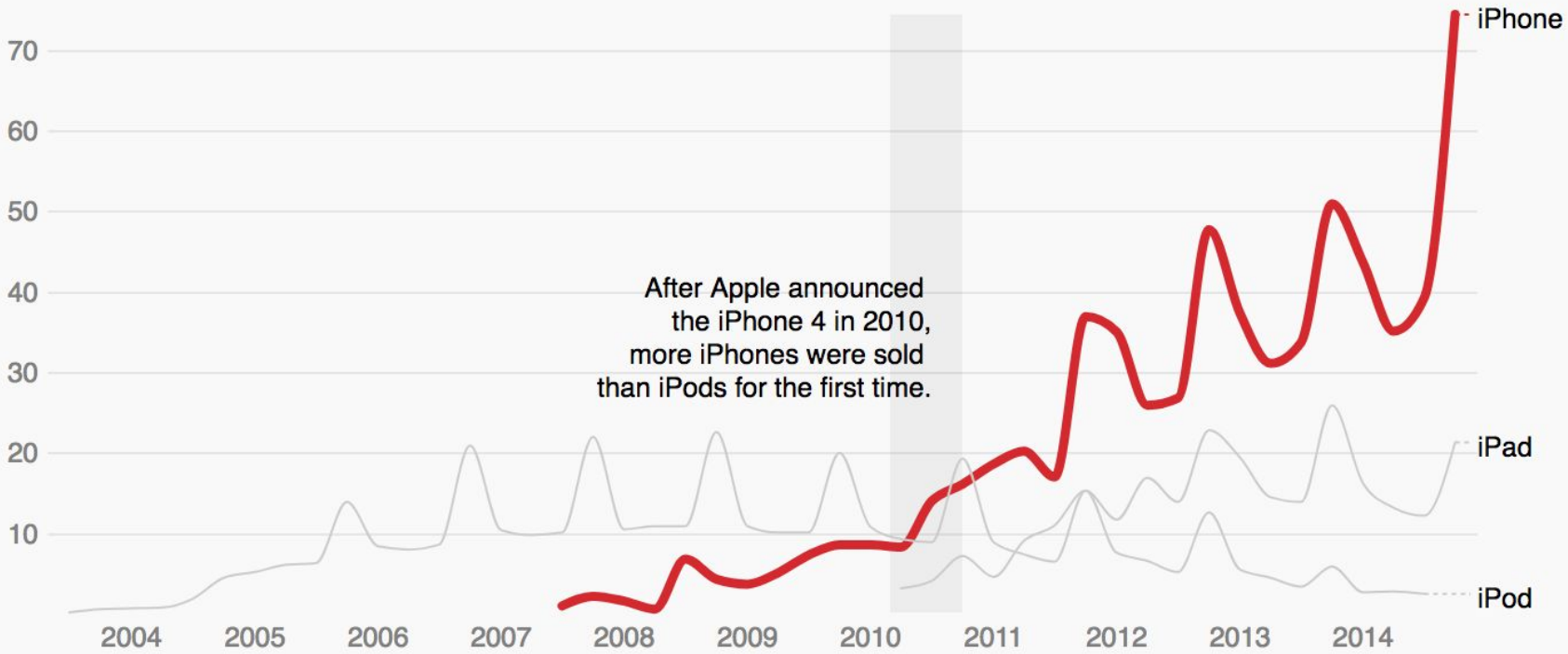
https://blog.datawrapper.de/better-charts/

Always include your source

https://blog.datawrapper.de/better-charts/

iPhone more successful than all other Apple products

Worldwide sales of selected Apple products in million, by fiscal quarter, 2000 to 2014



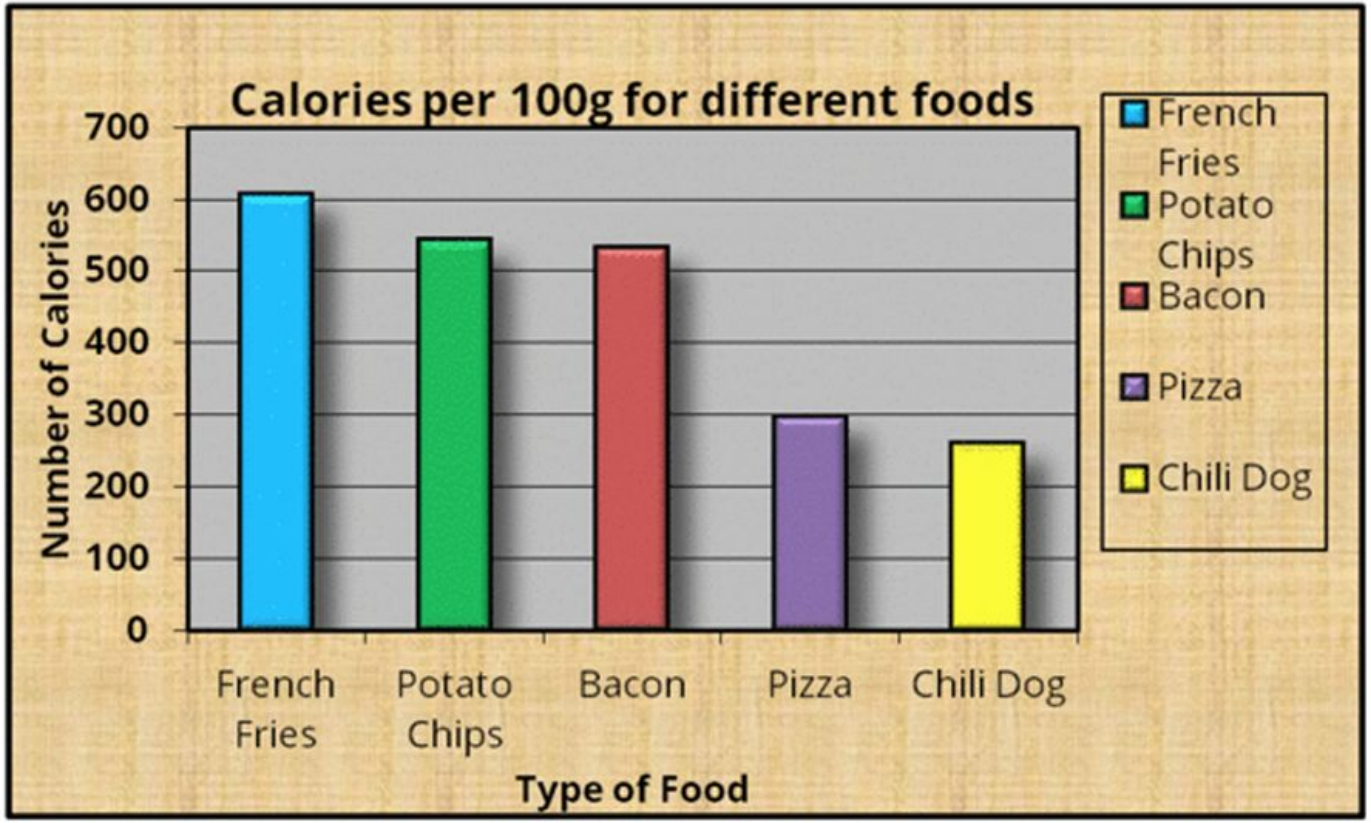
Source: Apple Inc. • Get the data • Created with Datawrapper

https://blog.datawrapper.de/better-charts/

Aim to improve your:

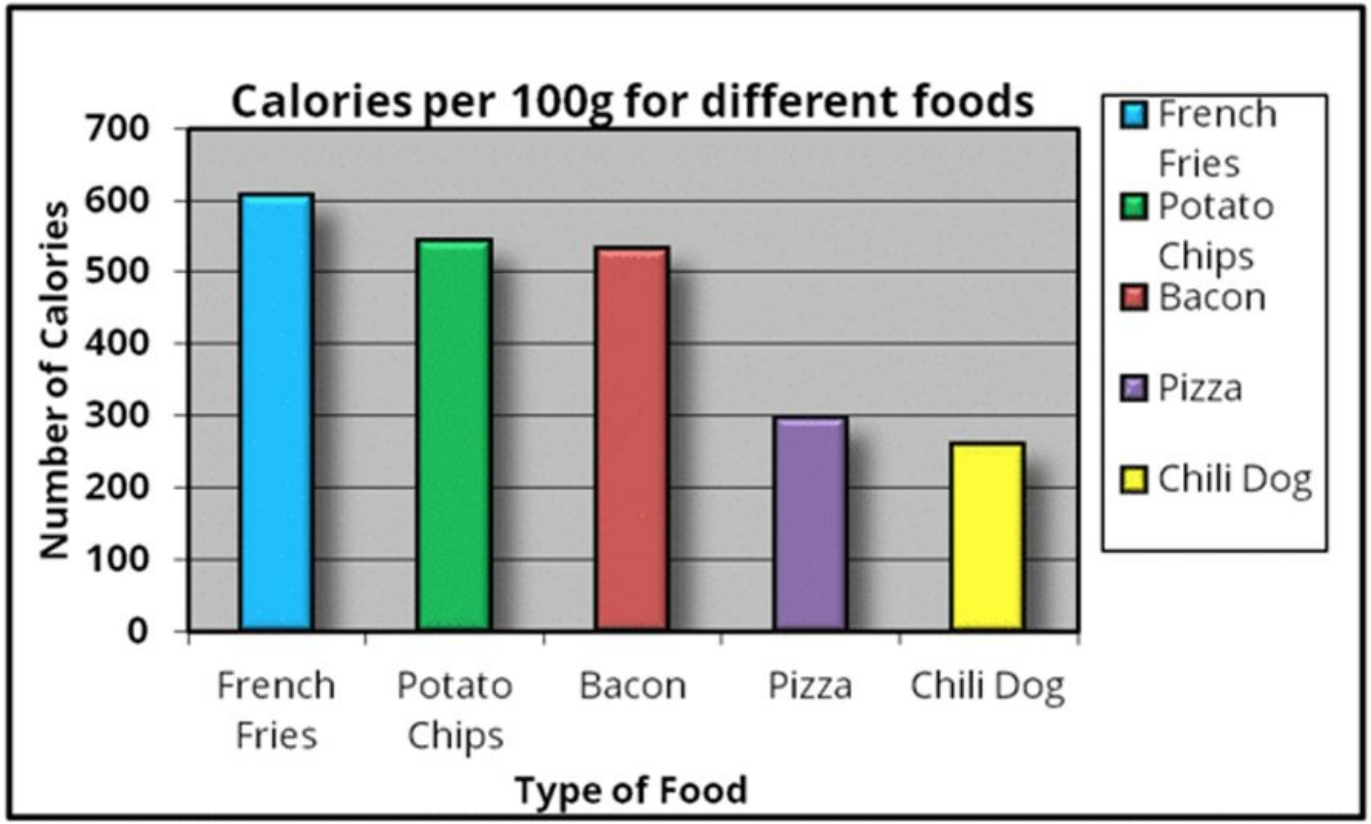
data:ink ratio

Remove backgrounds



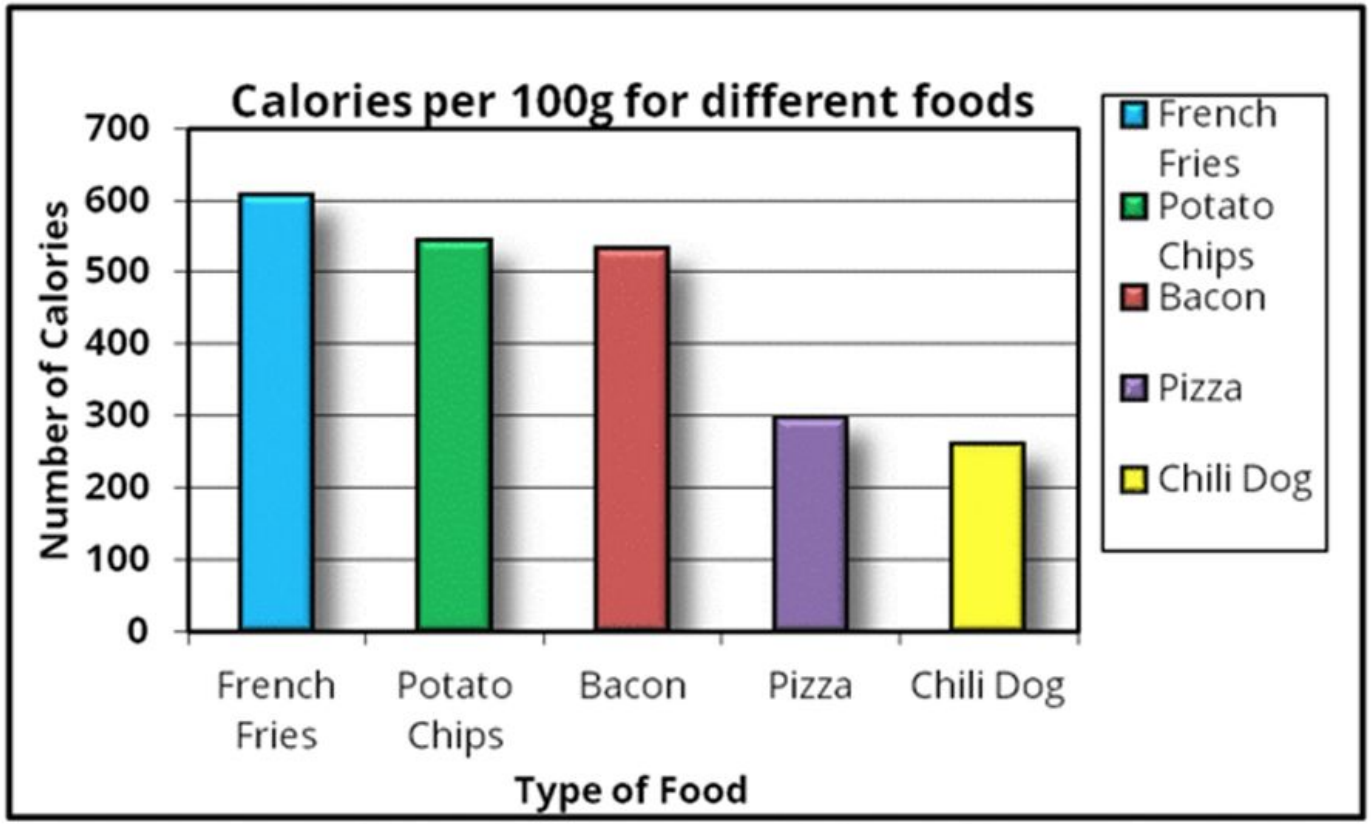
adapted from Brad Voytek

Remove backgrounds



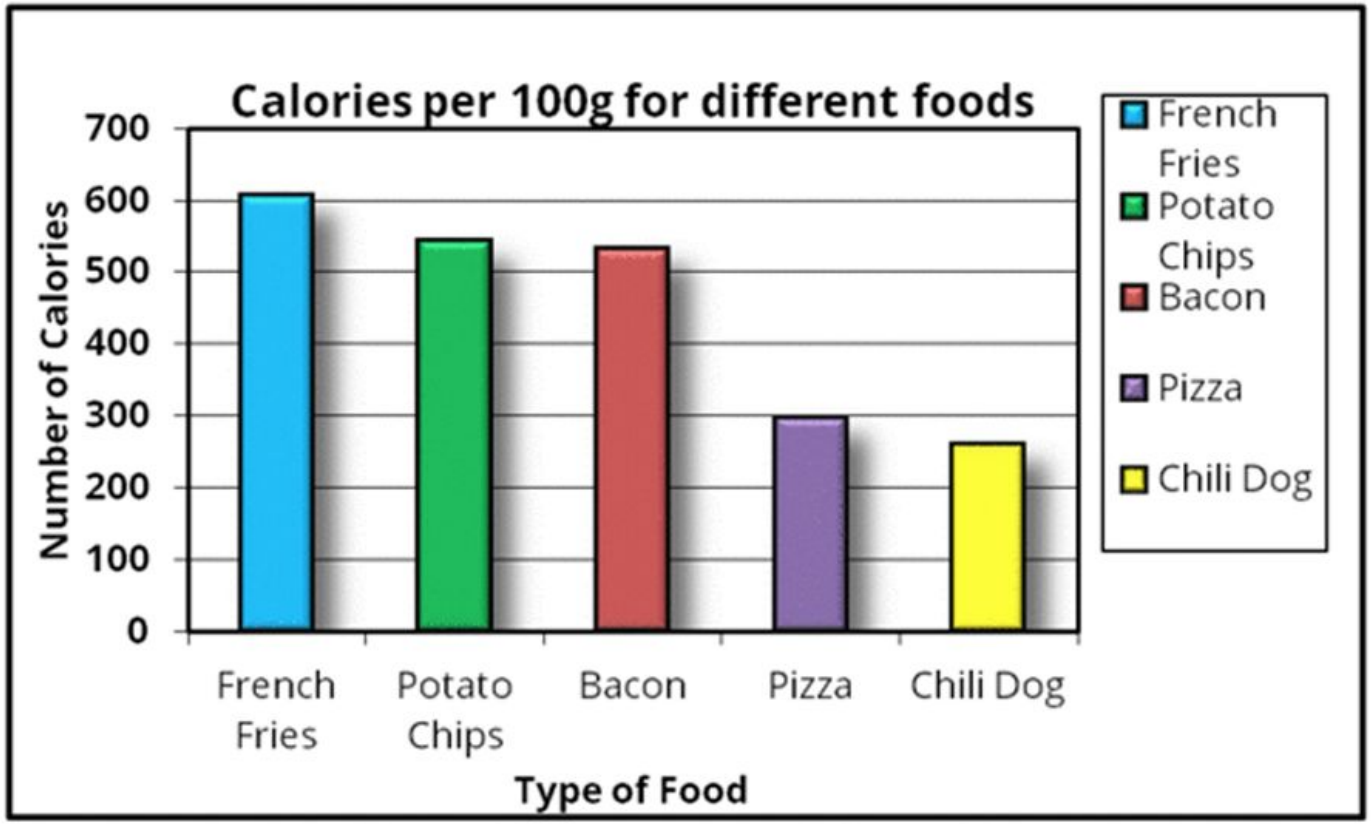
adapted from Brad Voytek

Remove backgrounds



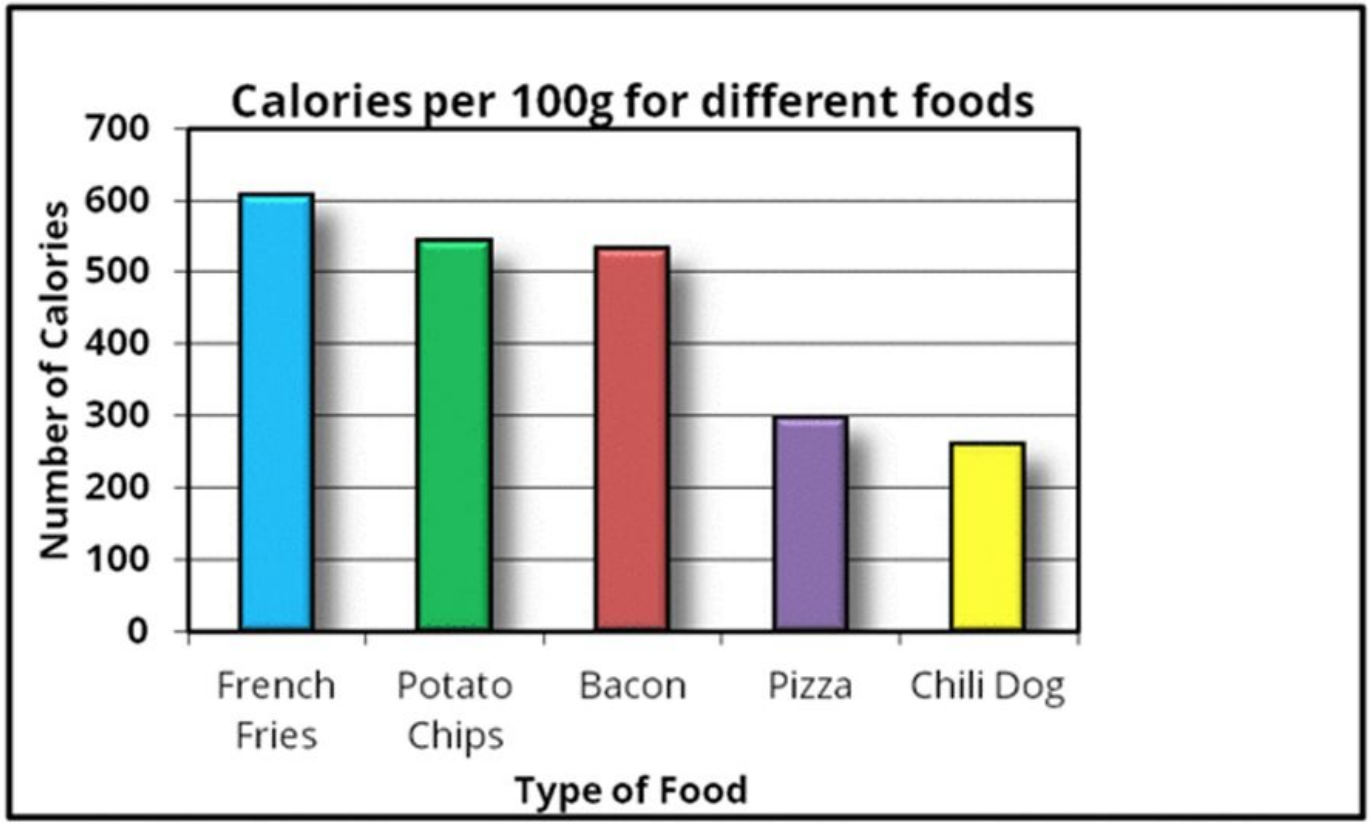
adapted from Brad Voytek

Remove redundant labels



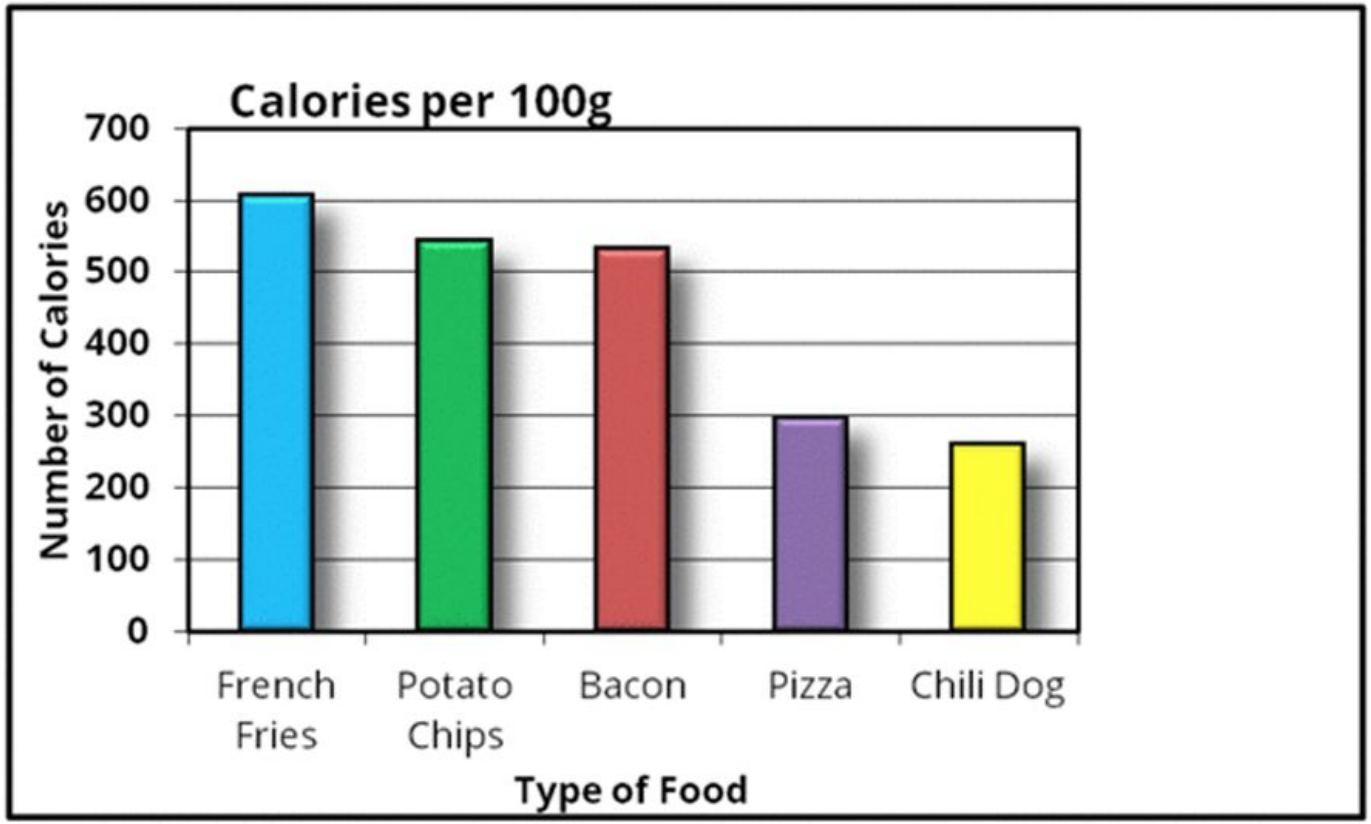
adapted from Brad Voytek

Remove redundant labels



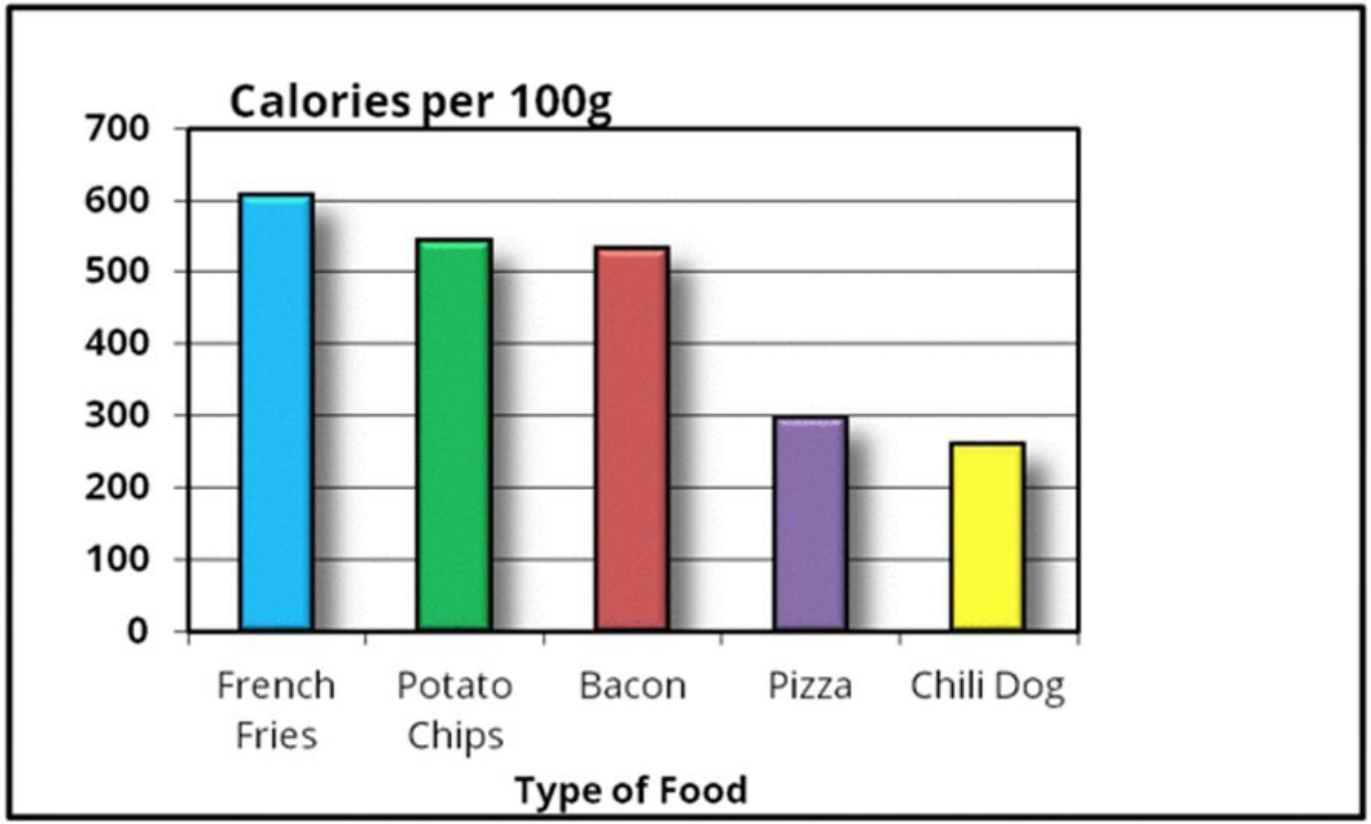
adapted from Brad Voytek

Remove redundant labels



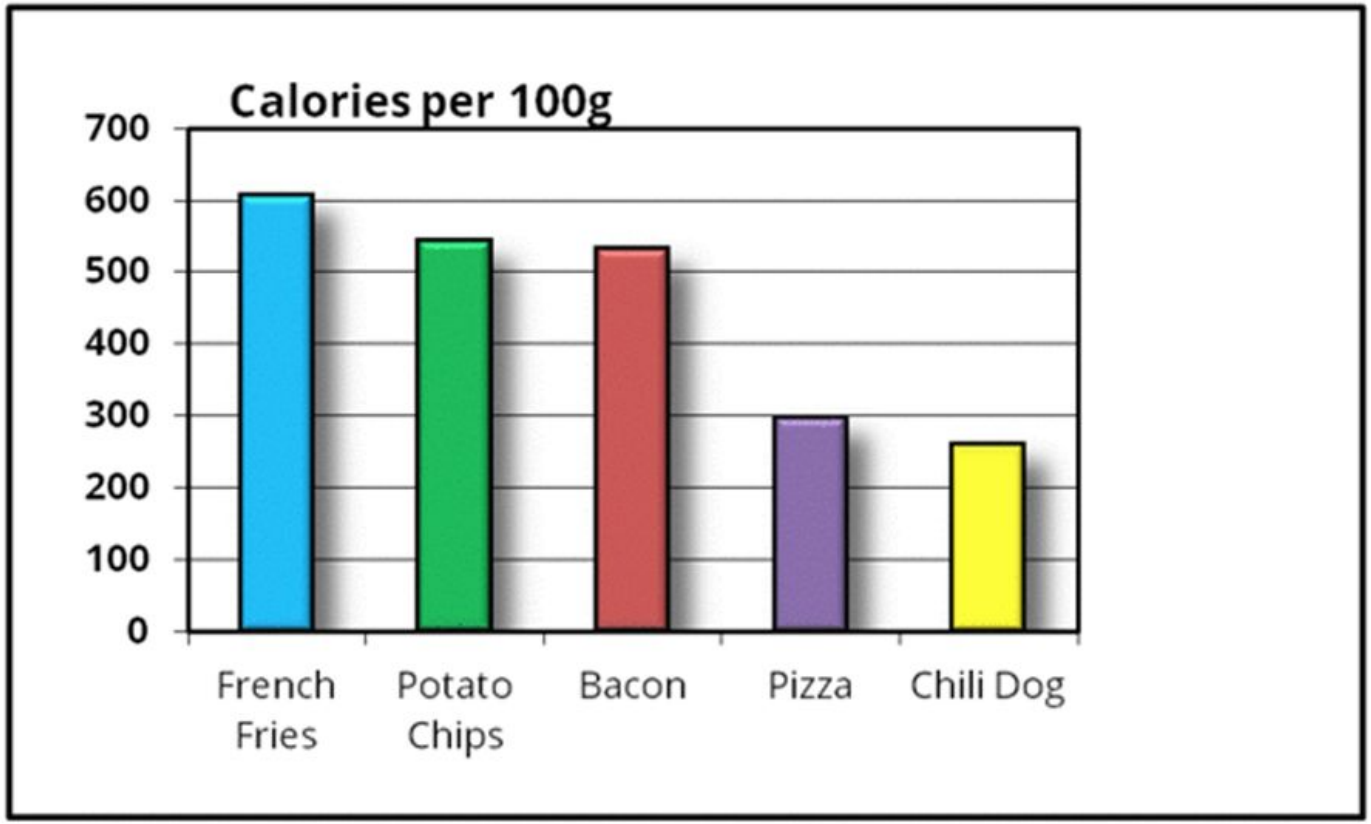
adapted from Brad Voytek

Remove redundant labels



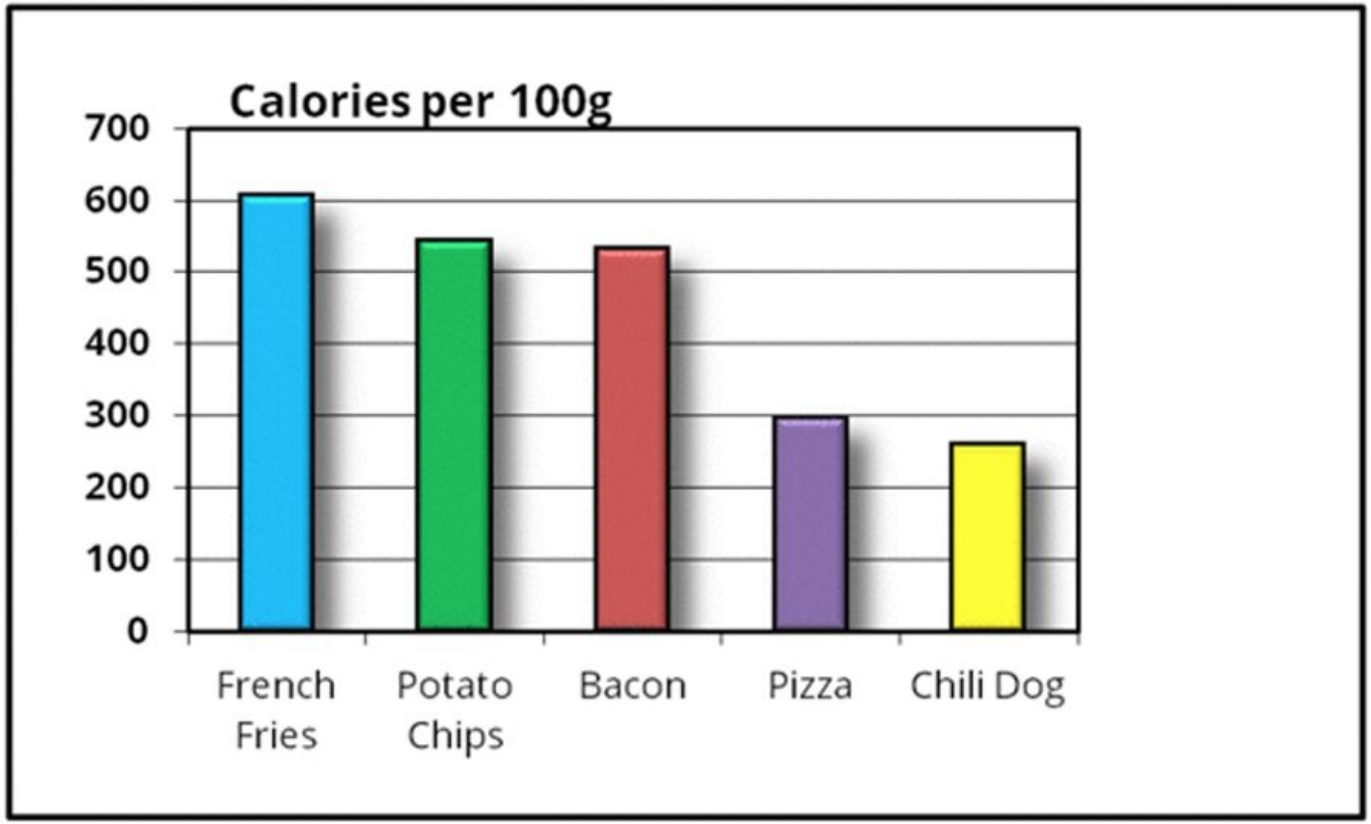
adapted from Brad Voytek

Remove redundant labels



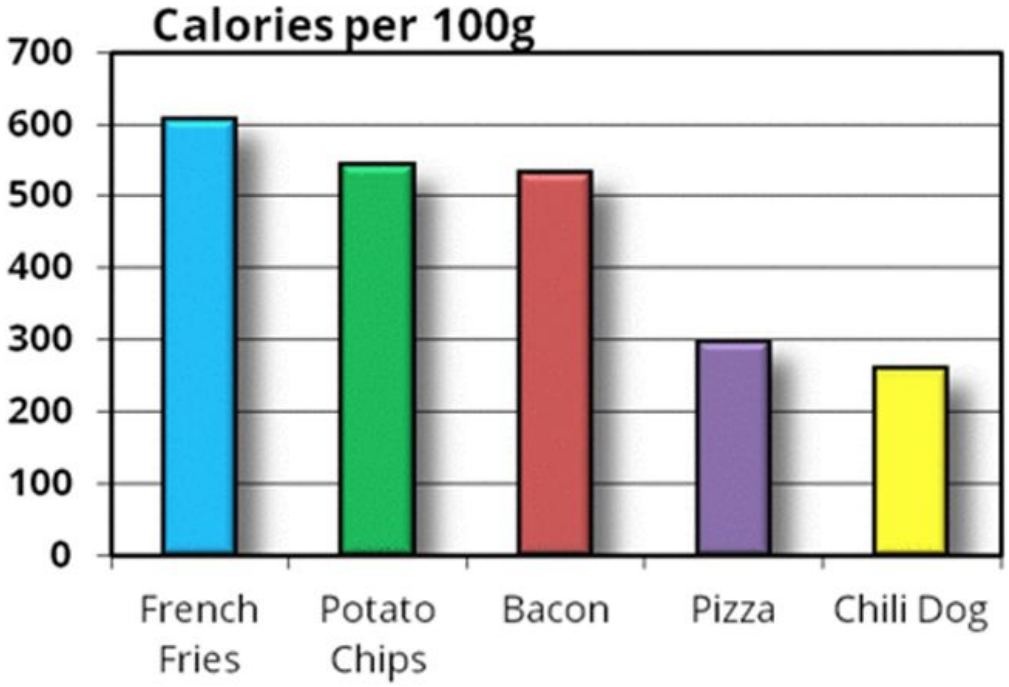
adapted from Brad Voytek

Remove borders



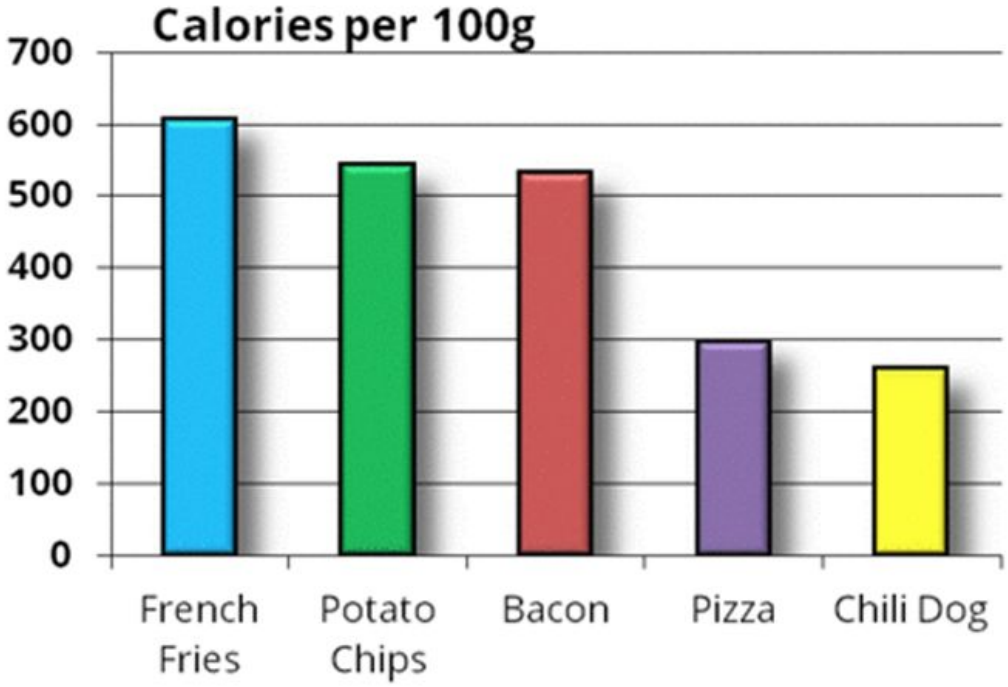
adapted from Brad Voytek

Remove borders



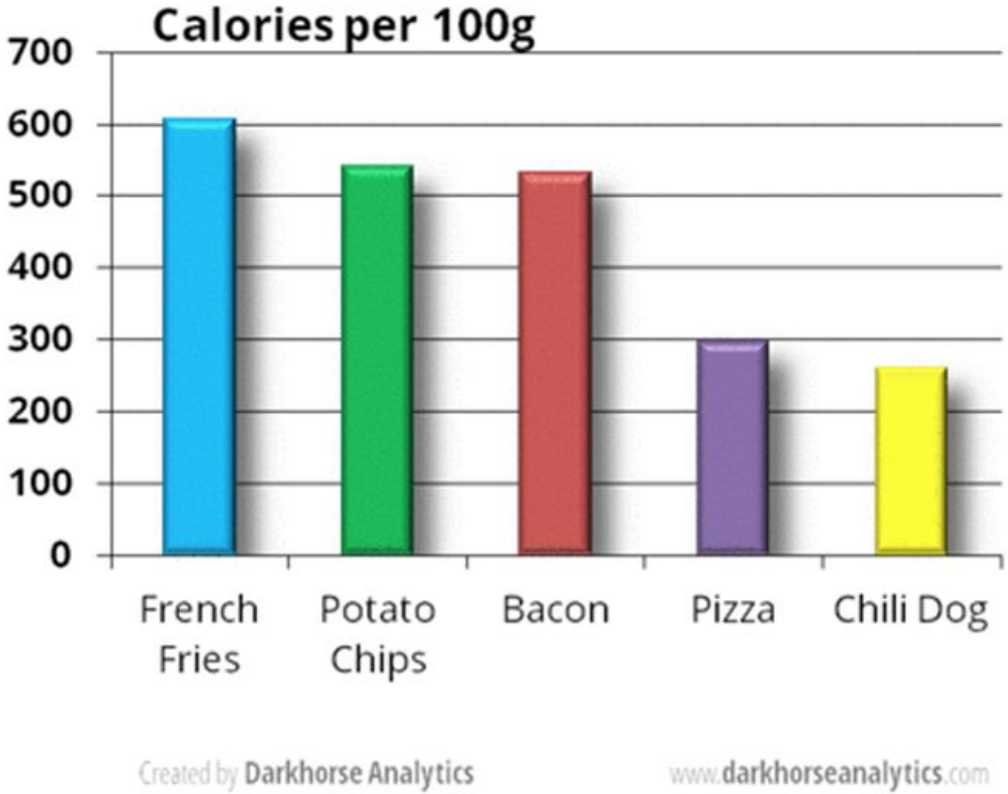
adapted from Brad Voytek

Remove borders



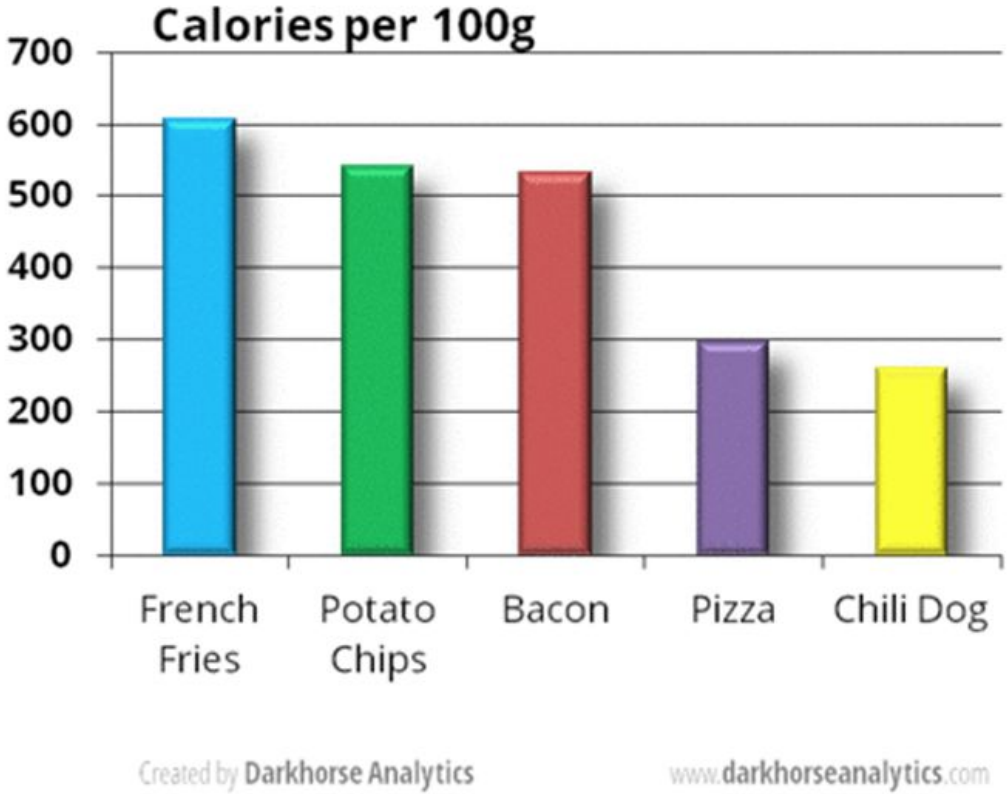
adapted from Brad Voytek

Remove borders



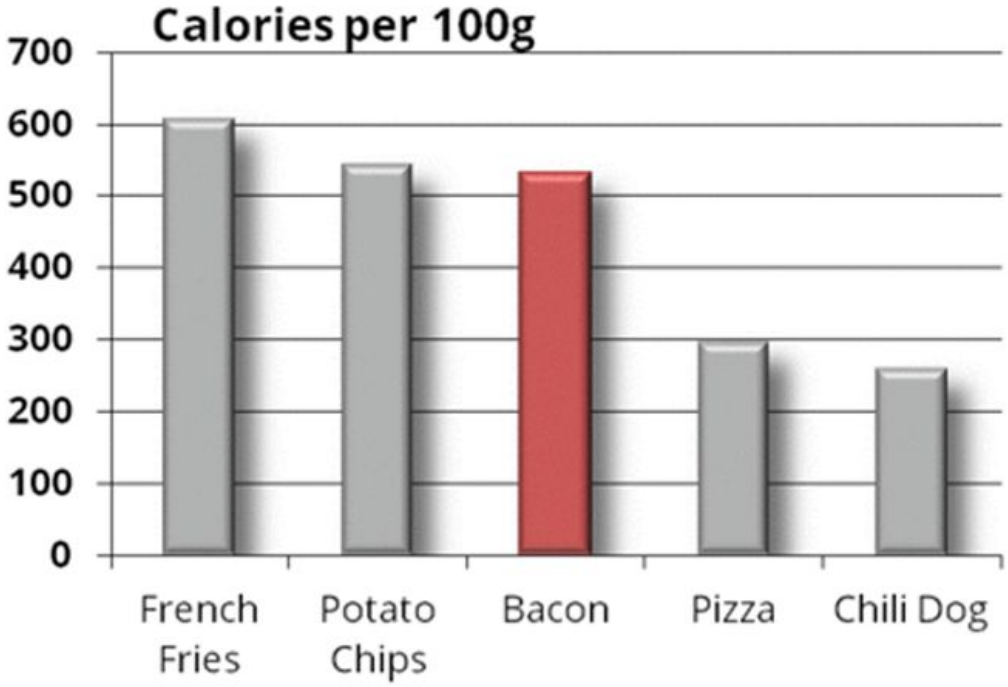
adapted from Brad Voytek

Reduce colors



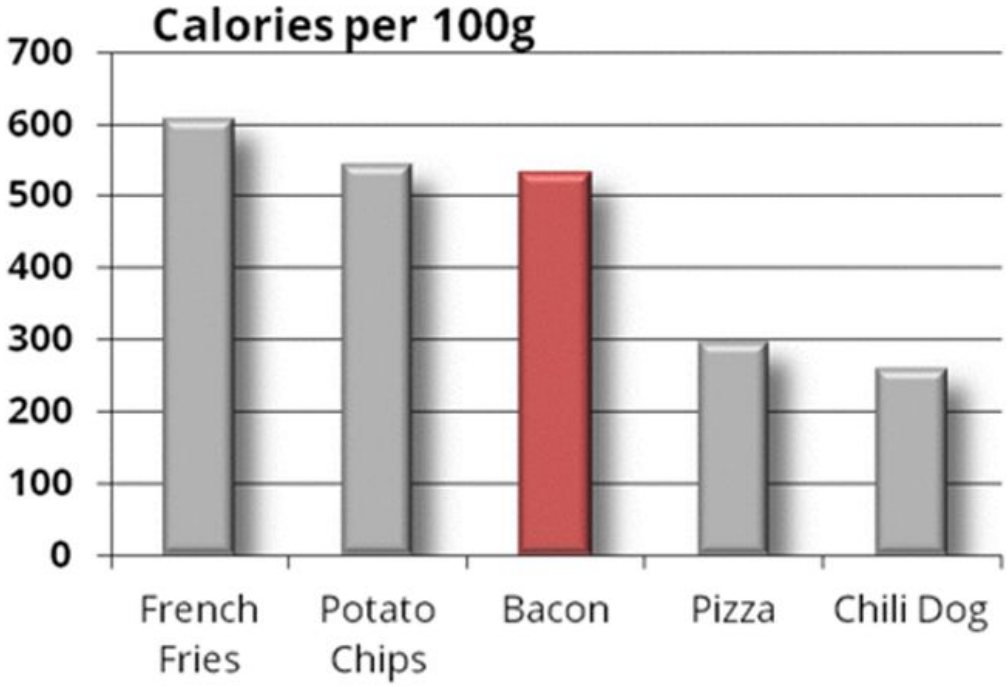
adapted from Brad Voytek

Reduce colors



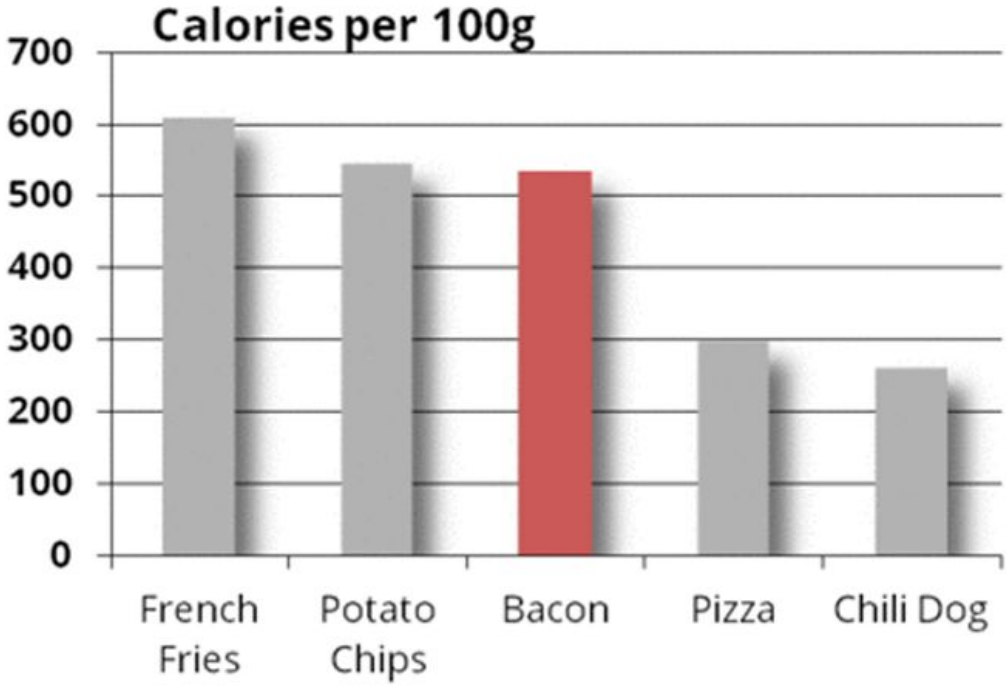
adapted from Brad Voytek

Remove special effects



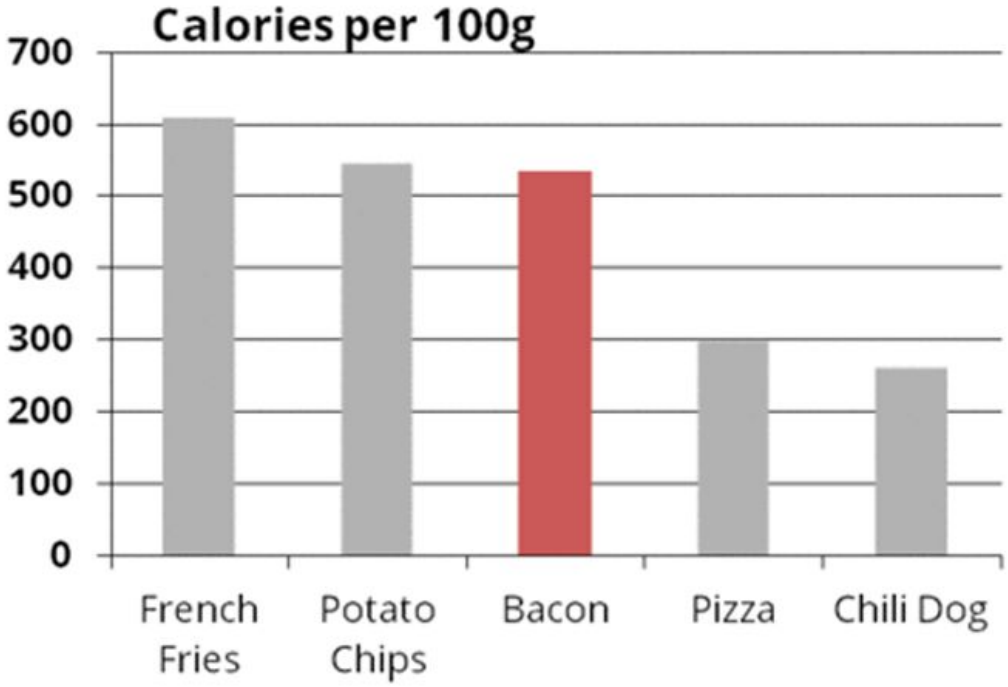
adapted from Brad Voytek

Remove special effects



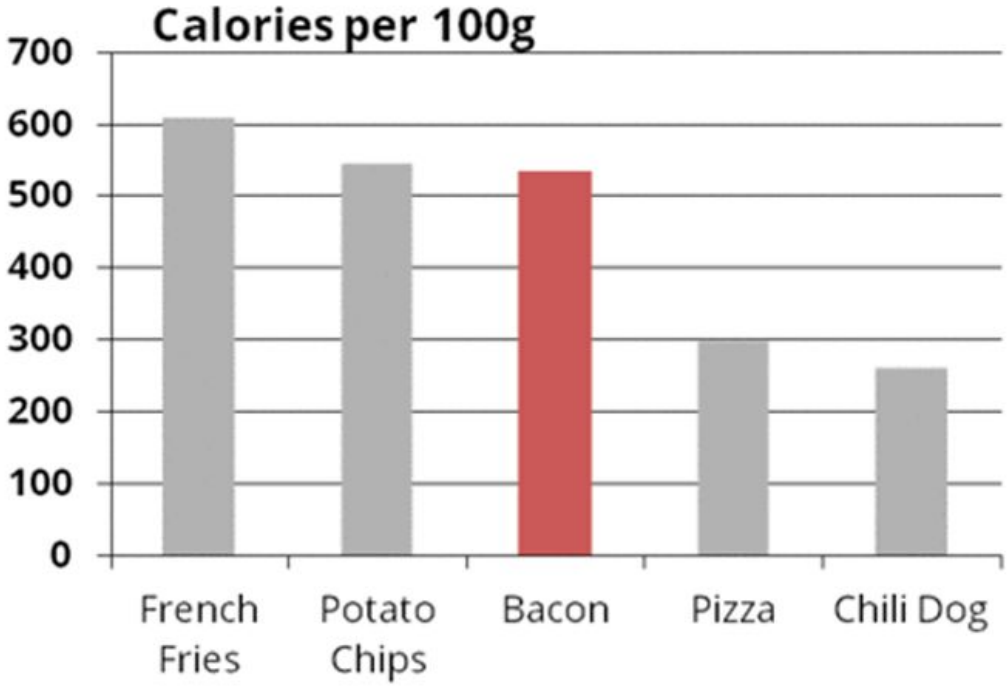
adapted from Brad Voytek

Remove special effects



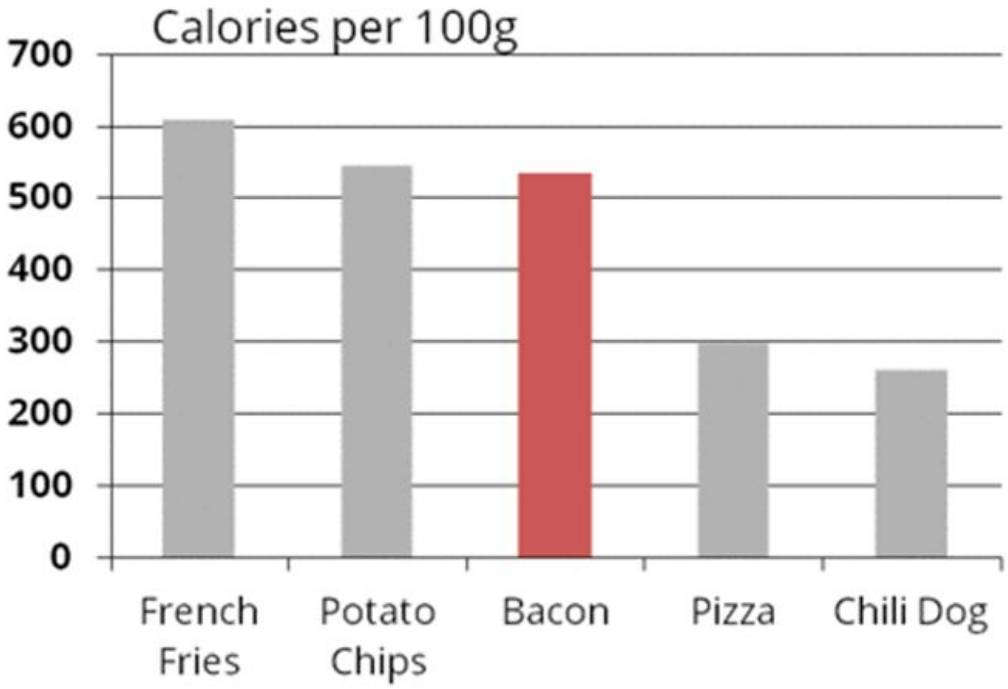
adapted from Brad Voytek

Remove bolding



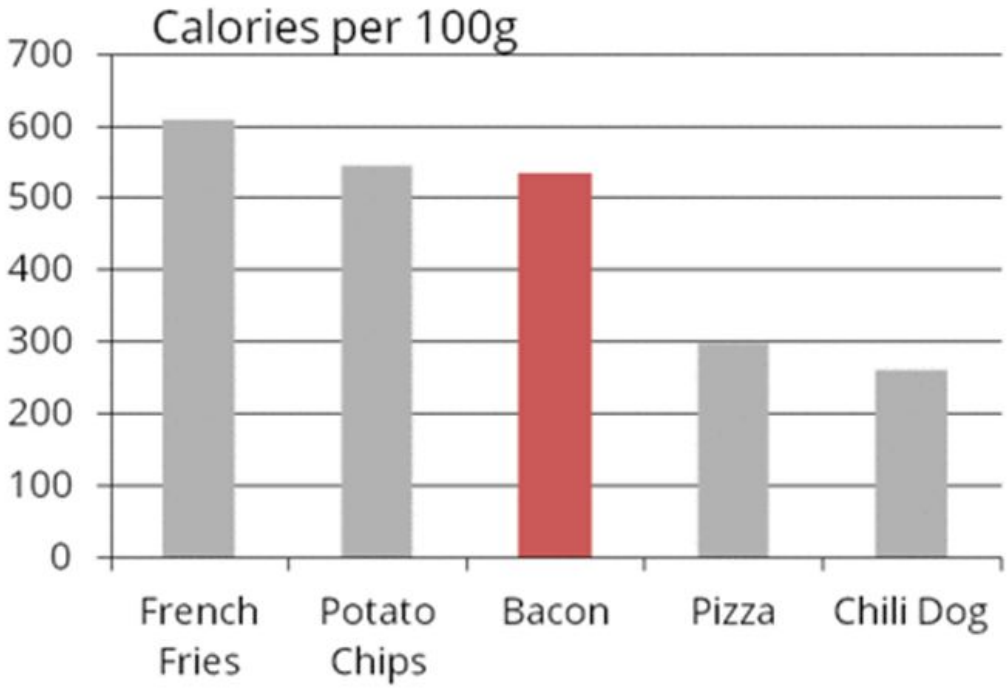
adapted from Brad Voytek

Remove bolding



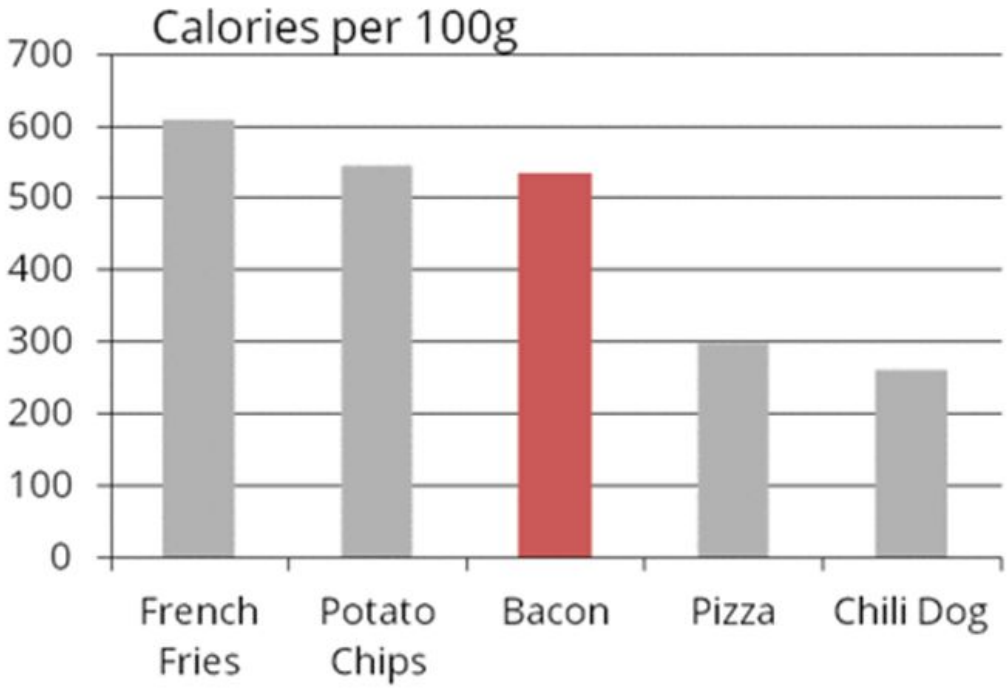
adapted from Brad Voytek

Remove bolding



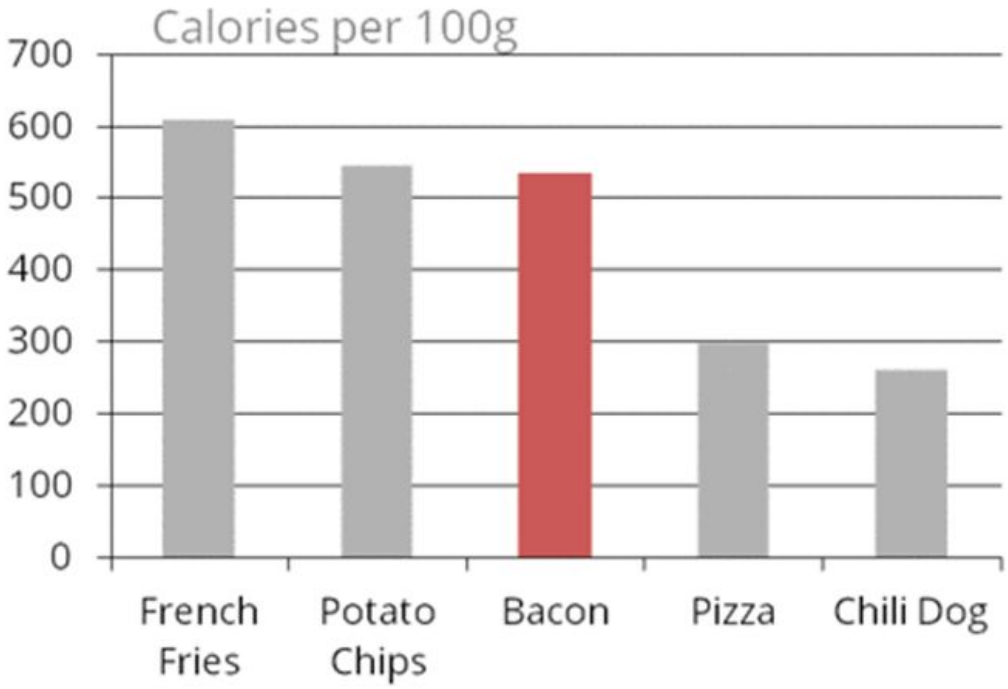
adapted from Brad Voytek

Lighten labels



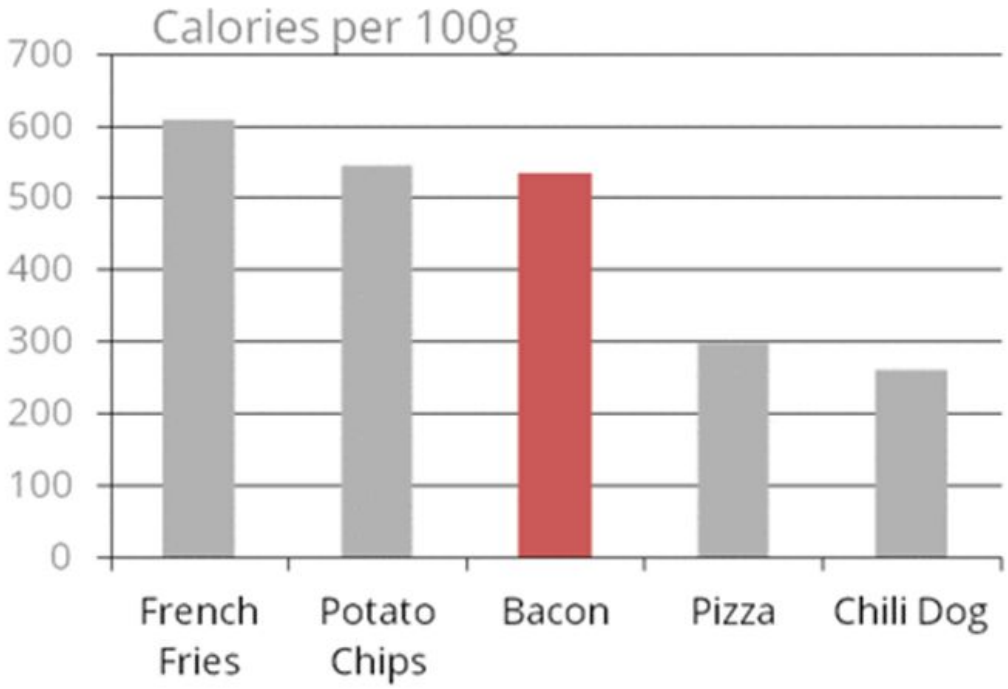
adapted from Brad Voytek

Lighten labels



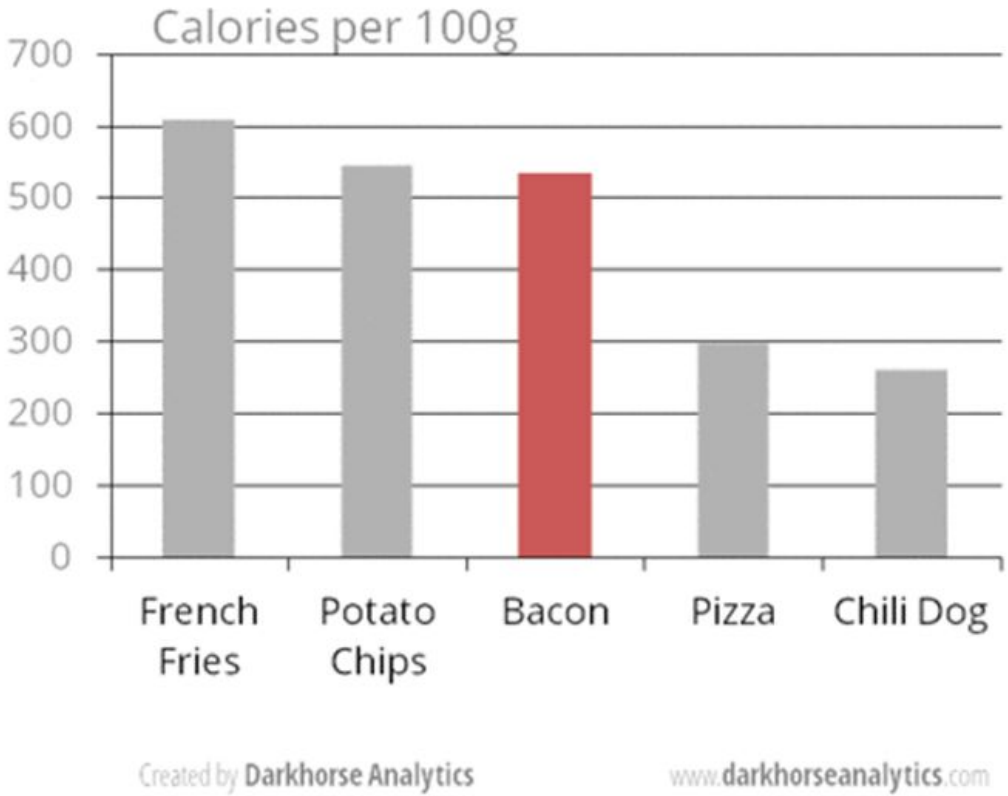
adapted from Brad Voytek

Lighten labels



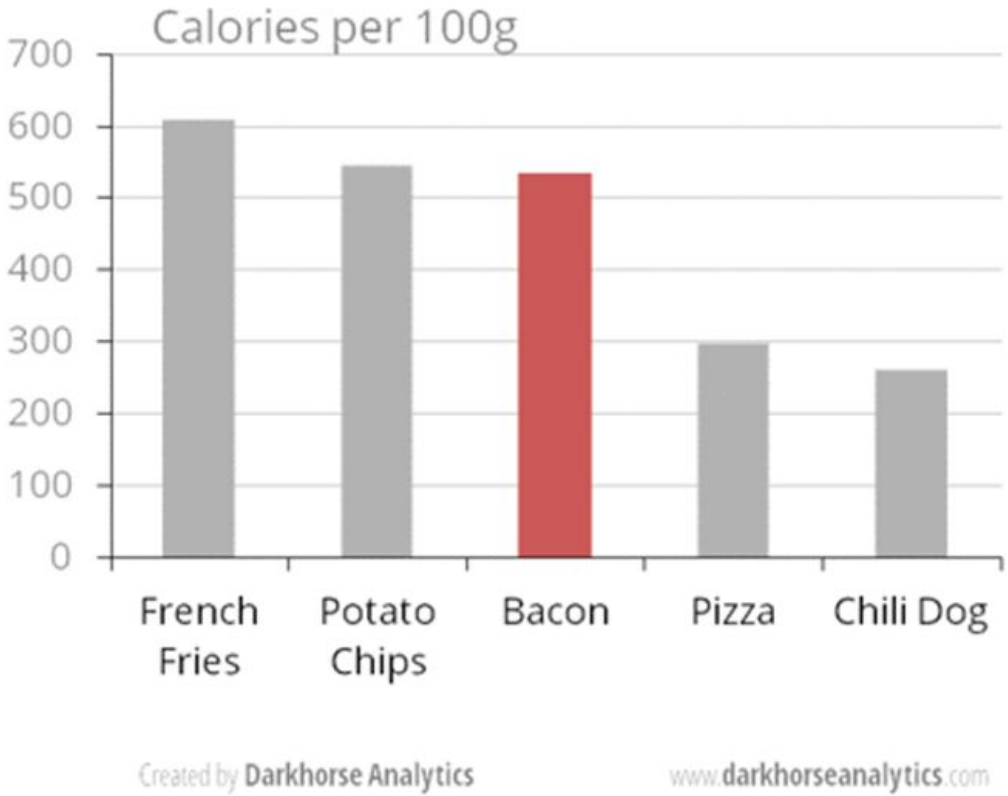
adapted from Brad Voytek

Lighten lines



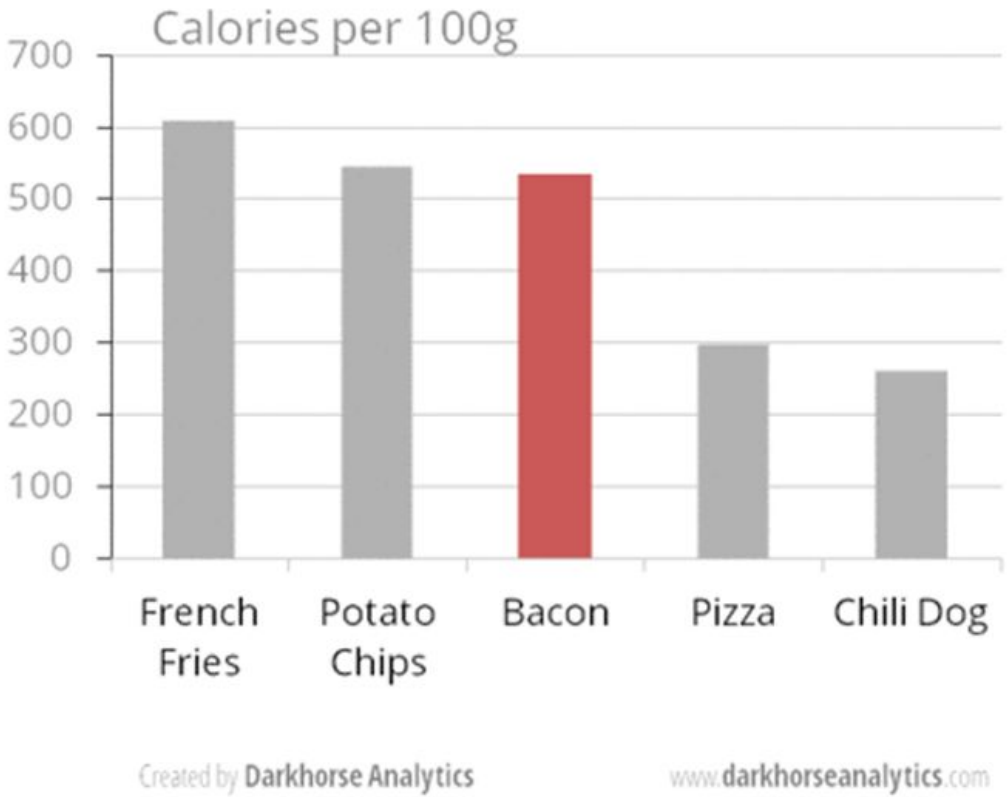
adapted from Brad Voytek

Lighten lines



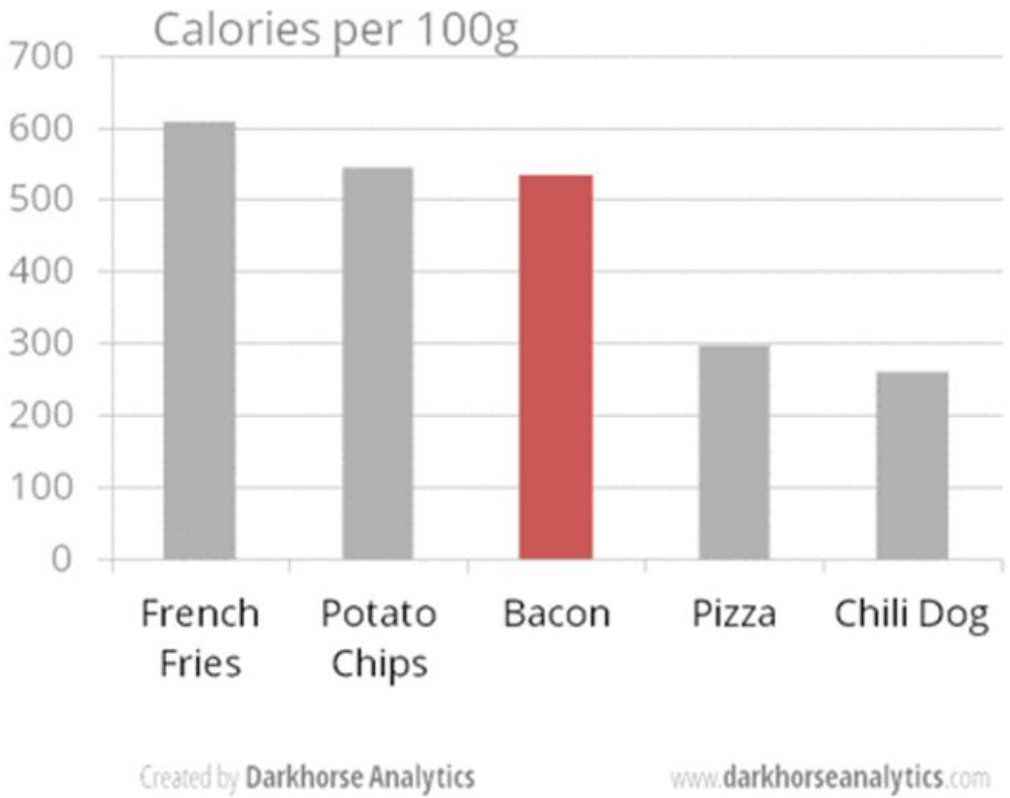
adapted from Brad Voytek

Lighten lines



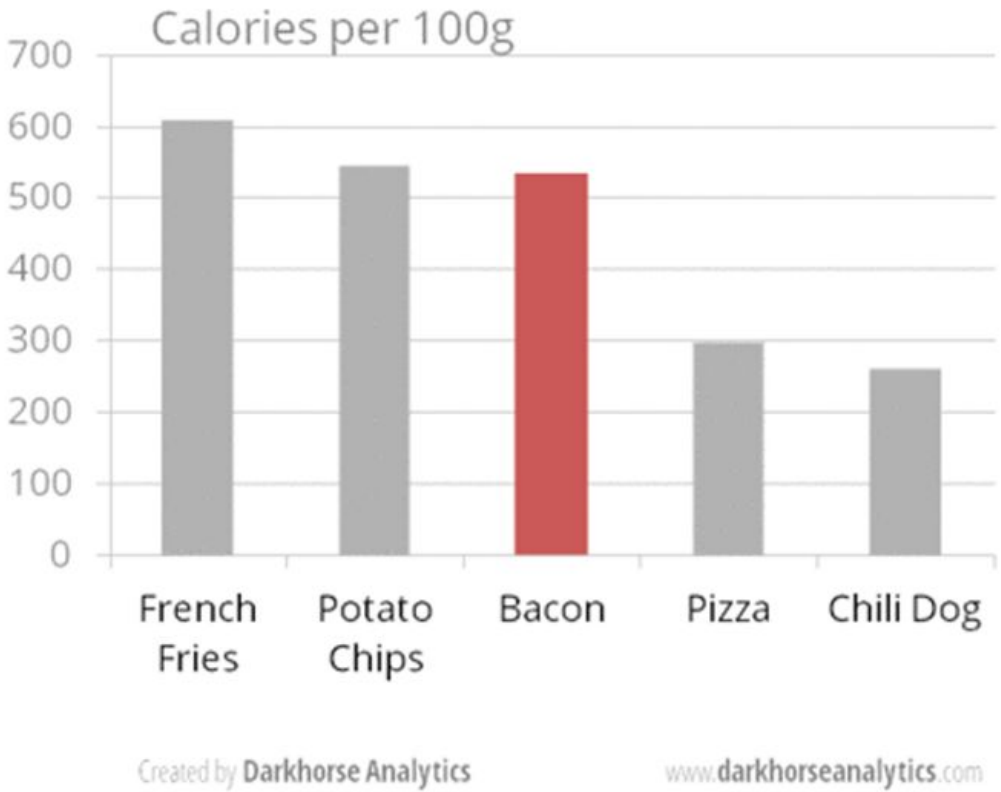
adapted from Brad Voytek

Lighten lines



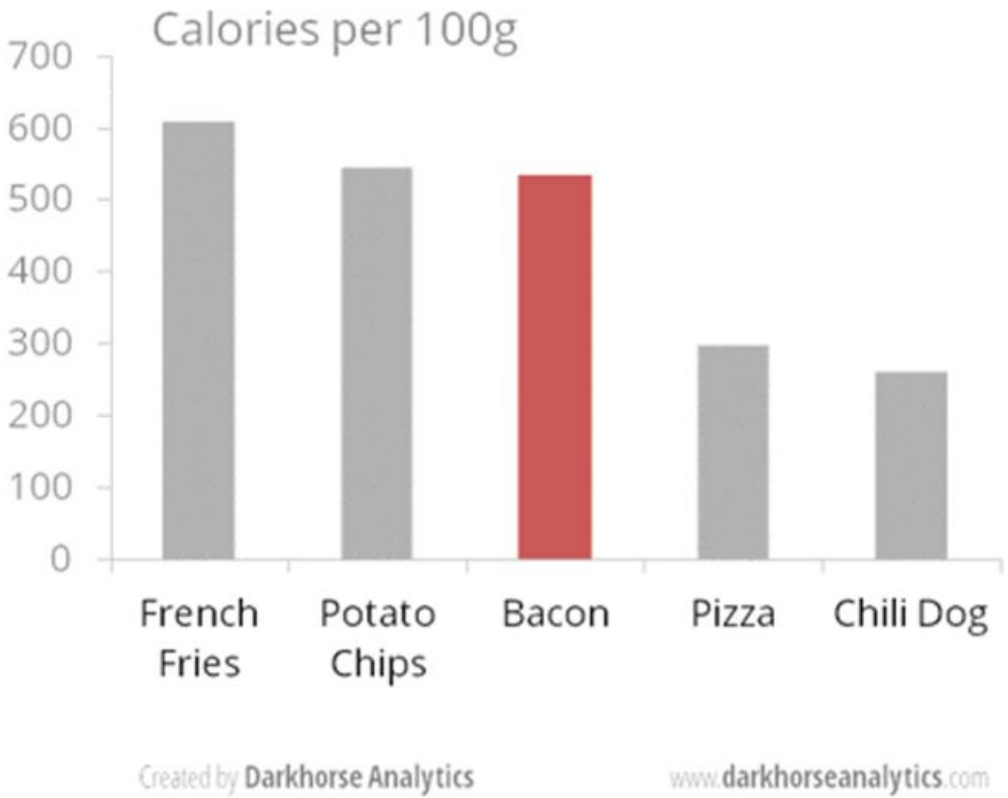
adapted from Brad Voytek

Or remove lines



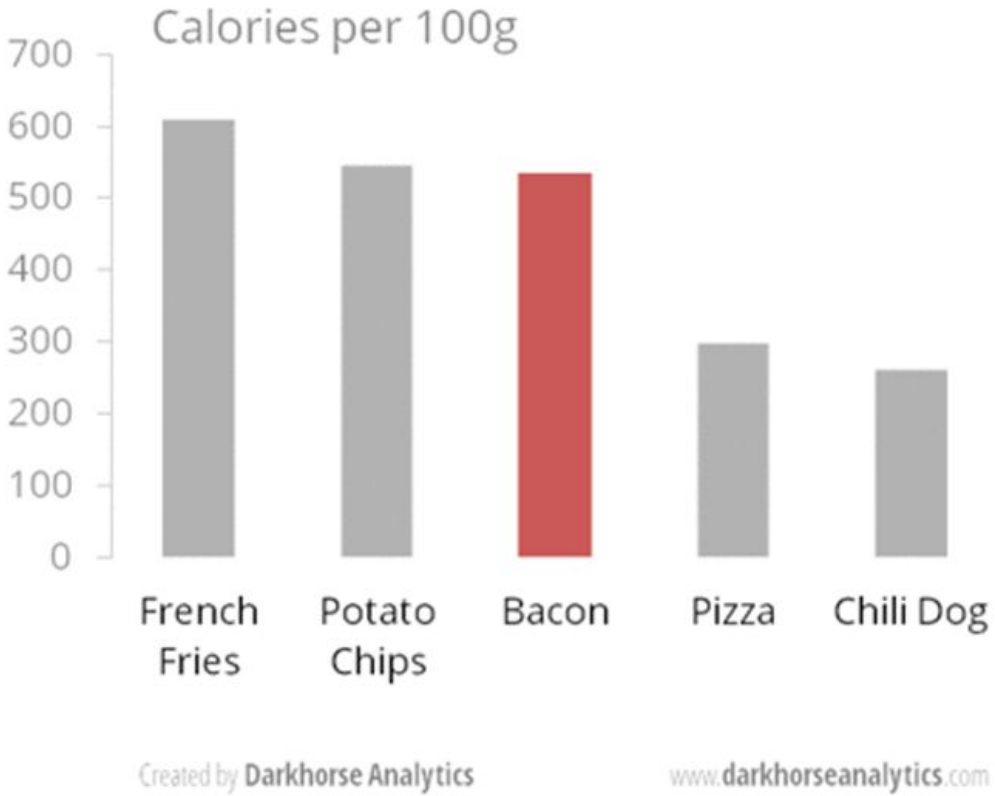
adapted from Brad Voytek

Or remove lines



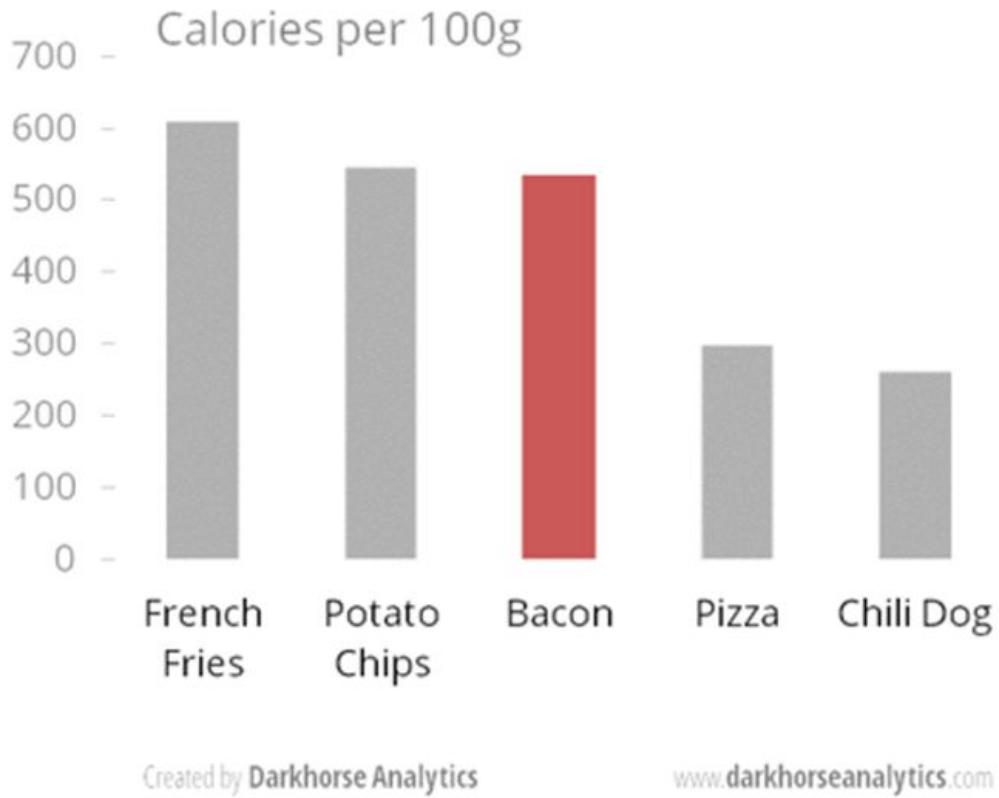
adapted from Brad Voytek

Or remove lines



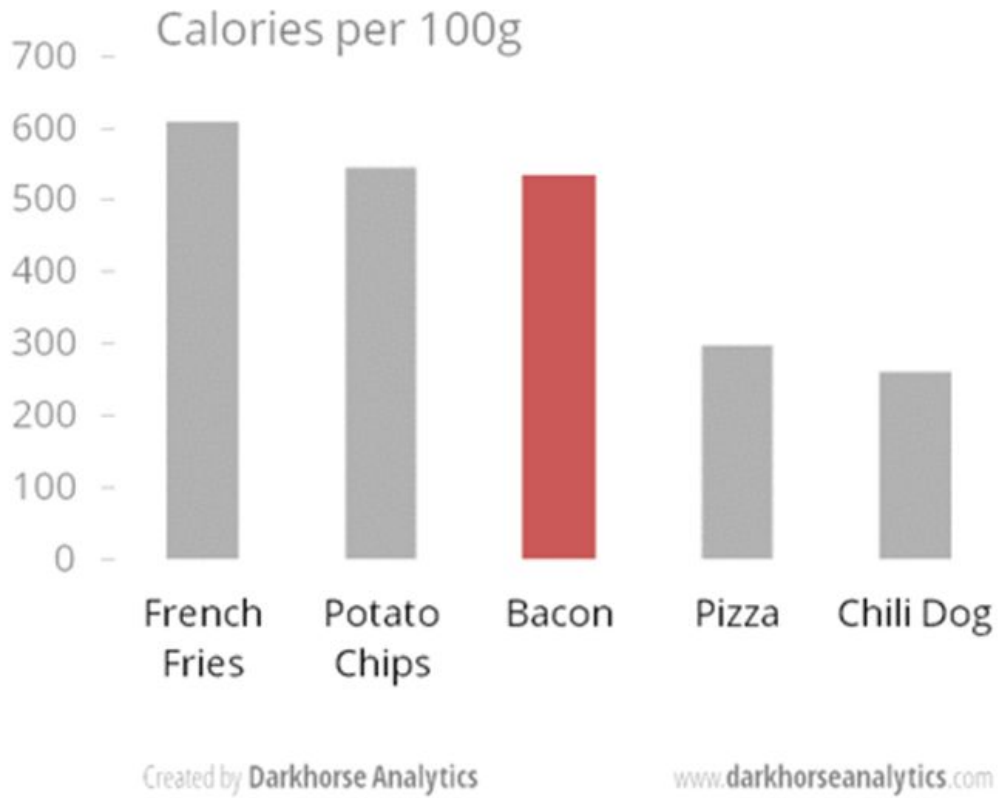
adapted from Brad Voytek

Or remove lines



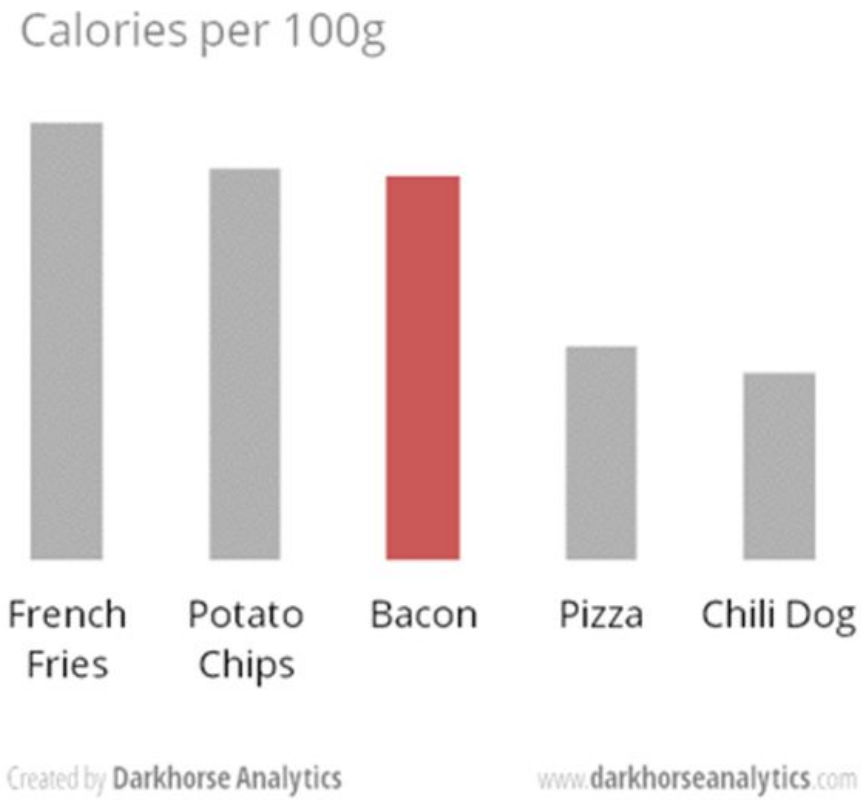
adapted from Brad Voytek

Direct label



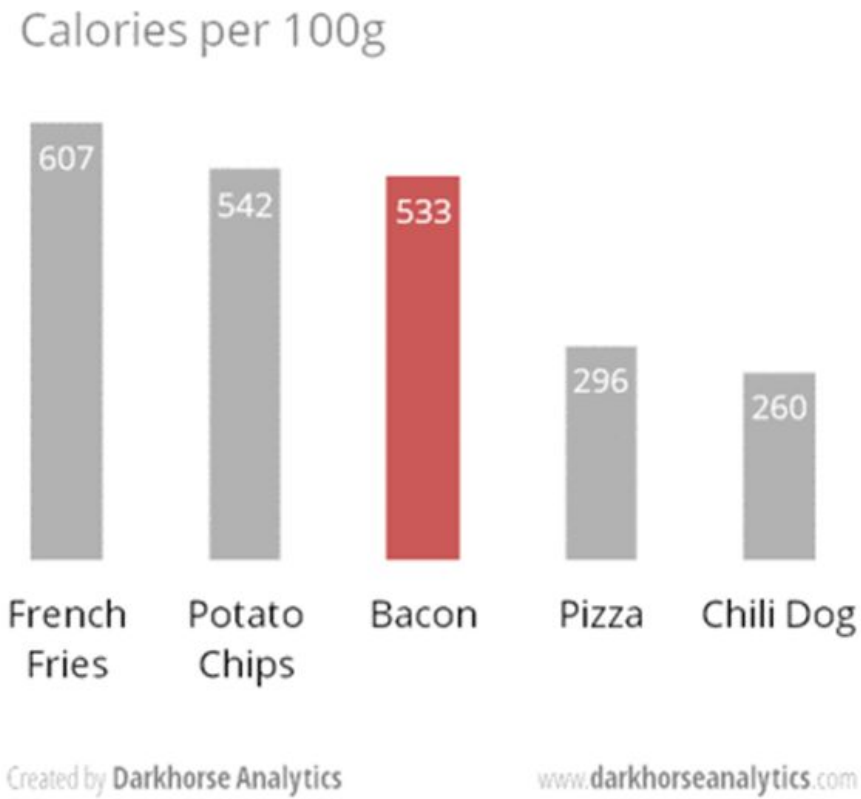
adapted from Brad Voytek

Direct label



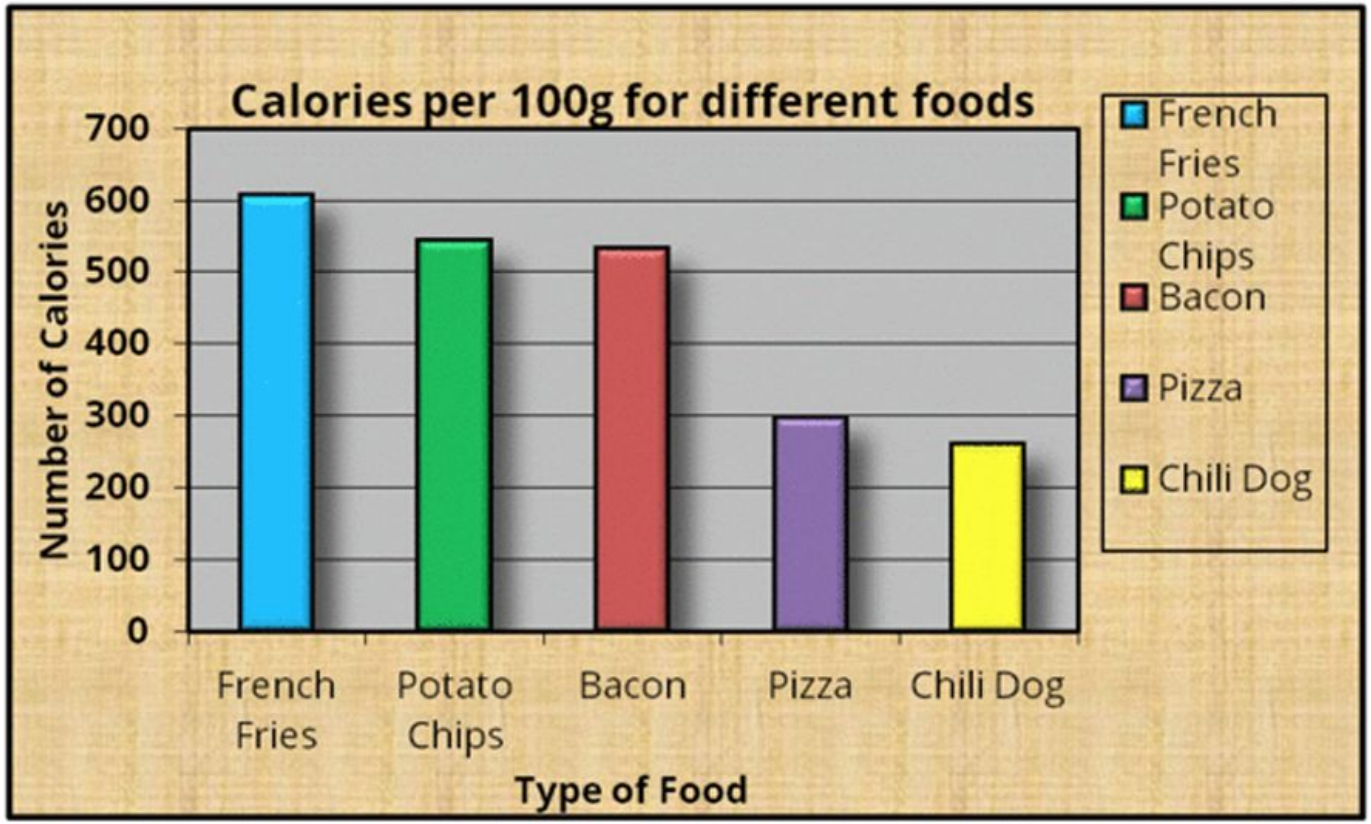
adapted from Brad Voytek

Direct label



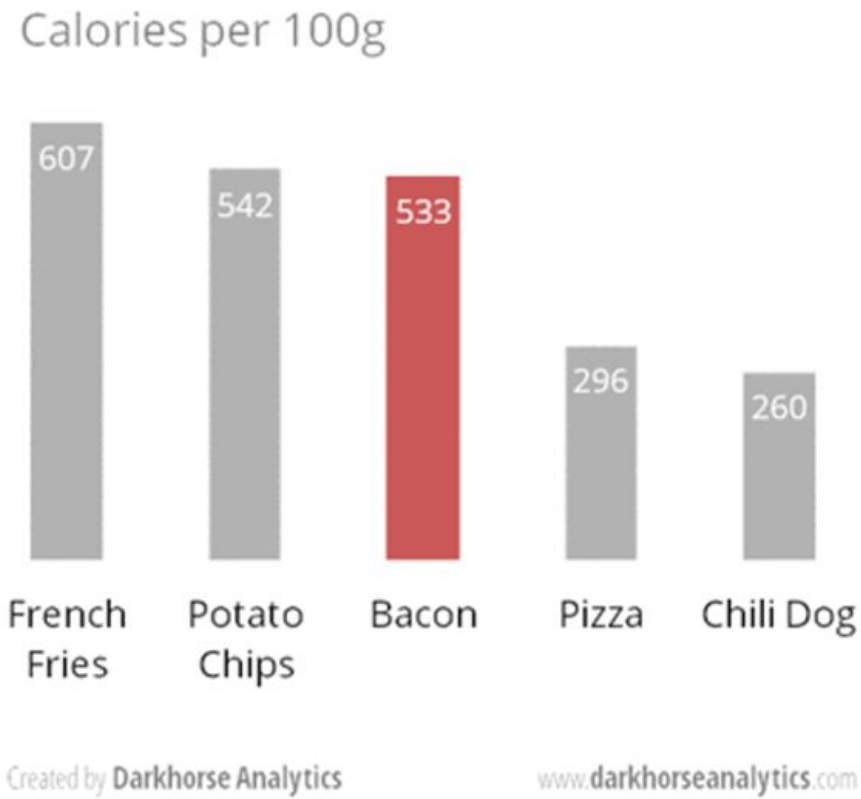
adapted from Brad Voytek

Before



adapted from Brad Voytek

After



adapted from Brad Voytek

less is more

(effective)
(attractive)
(impactive)

adapted from Brad Voytek