

Bradley Voytek, Ph.D.
UC San Diego

Department of Cognitive Science
Halıcıoğlu Data Science Institute
Neurosciences Graduate Program

bvoytek@ucsd.edu
voyteklab.com

UC San Diego

So data are often people

What can we do with those data?

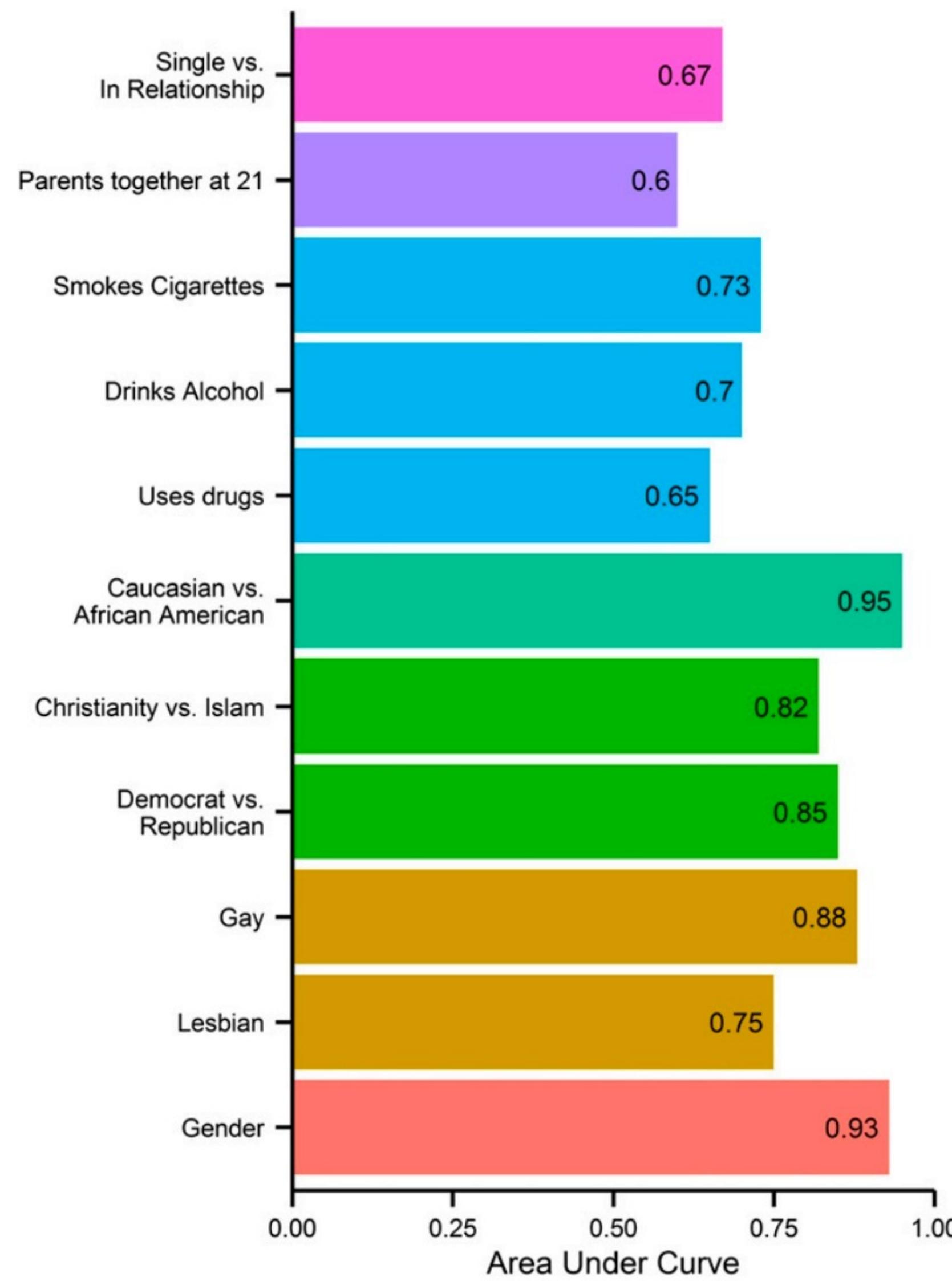


Fig. 2. Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.

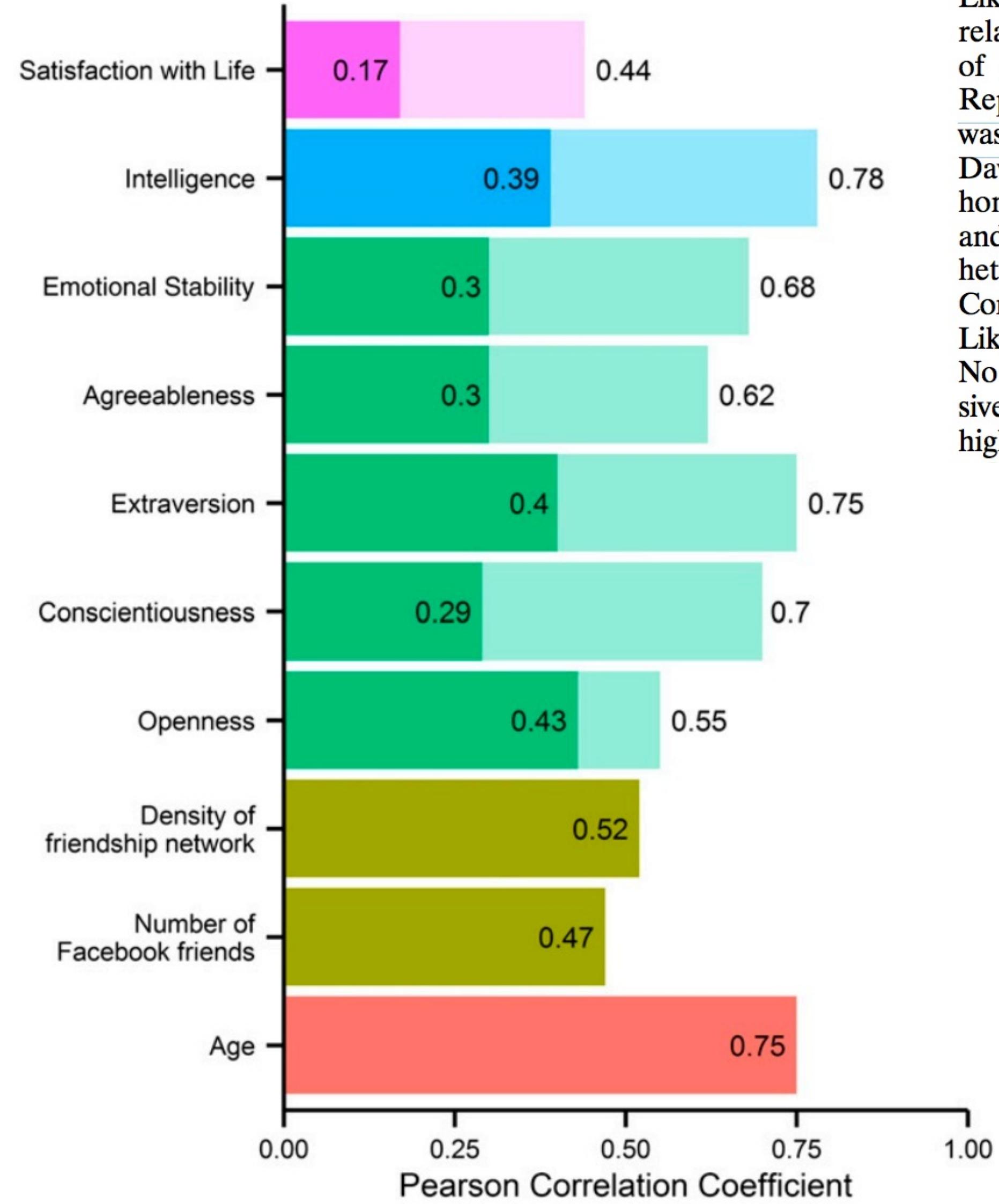
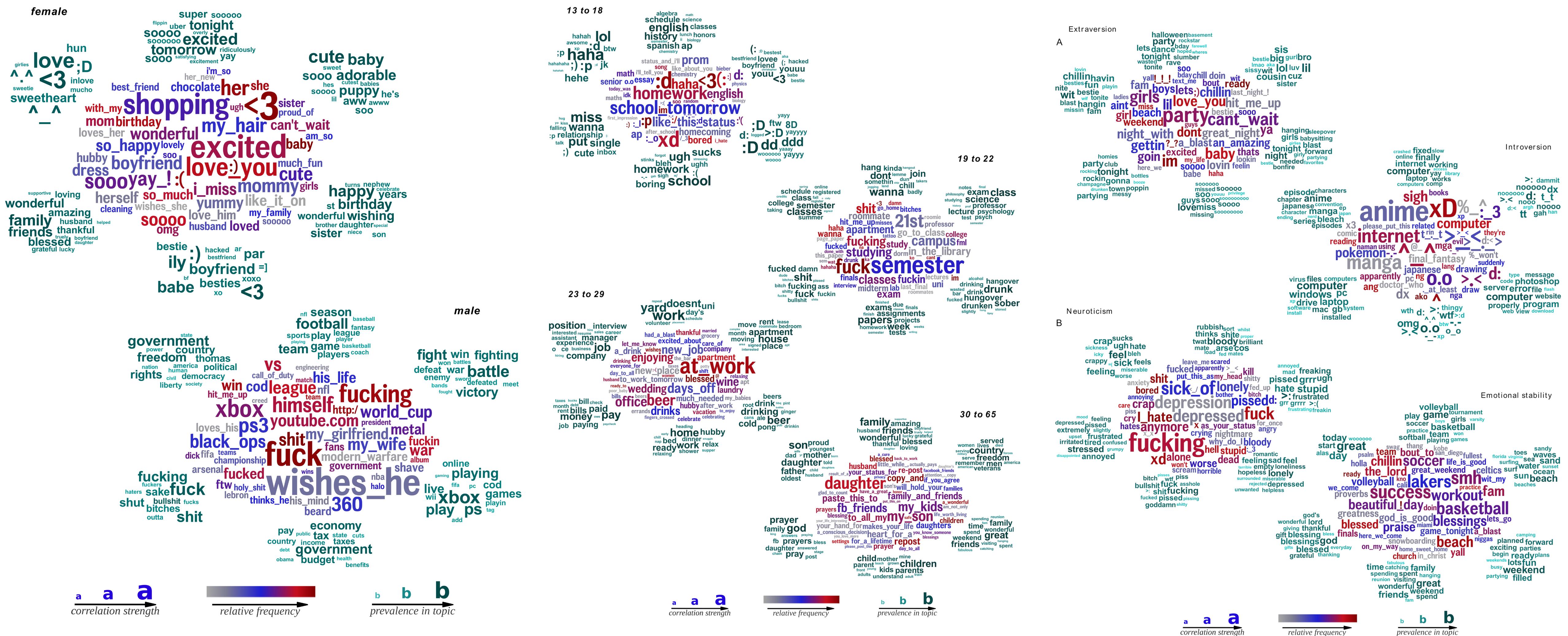


Fig. 3. Prediction accuracy of regression for numeric attributes and traits expressed by the Pearson correlation coefficient between predicted and actual attribute values; all correlations are significant at the $P < 0.001$ level. The transparent bars indicate the questionnaire's baseline accuracy, expressed in terms of test-retest reliability.

Predictive Power of Likes. Individual traits and attributes can be predicted to a high degree of accuracy based on records of users' Likes. **Table S1** presents a sample of highly predictive Likes related to each of the attributes. For example, the best predictors of high intelligence include "Thunderstorms," "The Colbert Report," "Science," and "Curly Fries," whereas low intelligence was indicated by "Sephora," "I Love Being A Mom," "Harley Davidson," and "Lady Antebellum." Good predictors of male homosexuality included "No H8 Campaign," "Mac Cosmetics," and "Wicked The Musical," whereas strong predictors of male heterosexuality included "Wu-Tang Clan," "Shaq," and "Being Confused After Waking Up From Naps." Although some of the Likes clearly relate to their predicted attribute, as in the case of No H8 Campaign and homosexuality, other pairs are more elusive; there is no obvious connection between Curly Fries and high intelligence.

Facebook

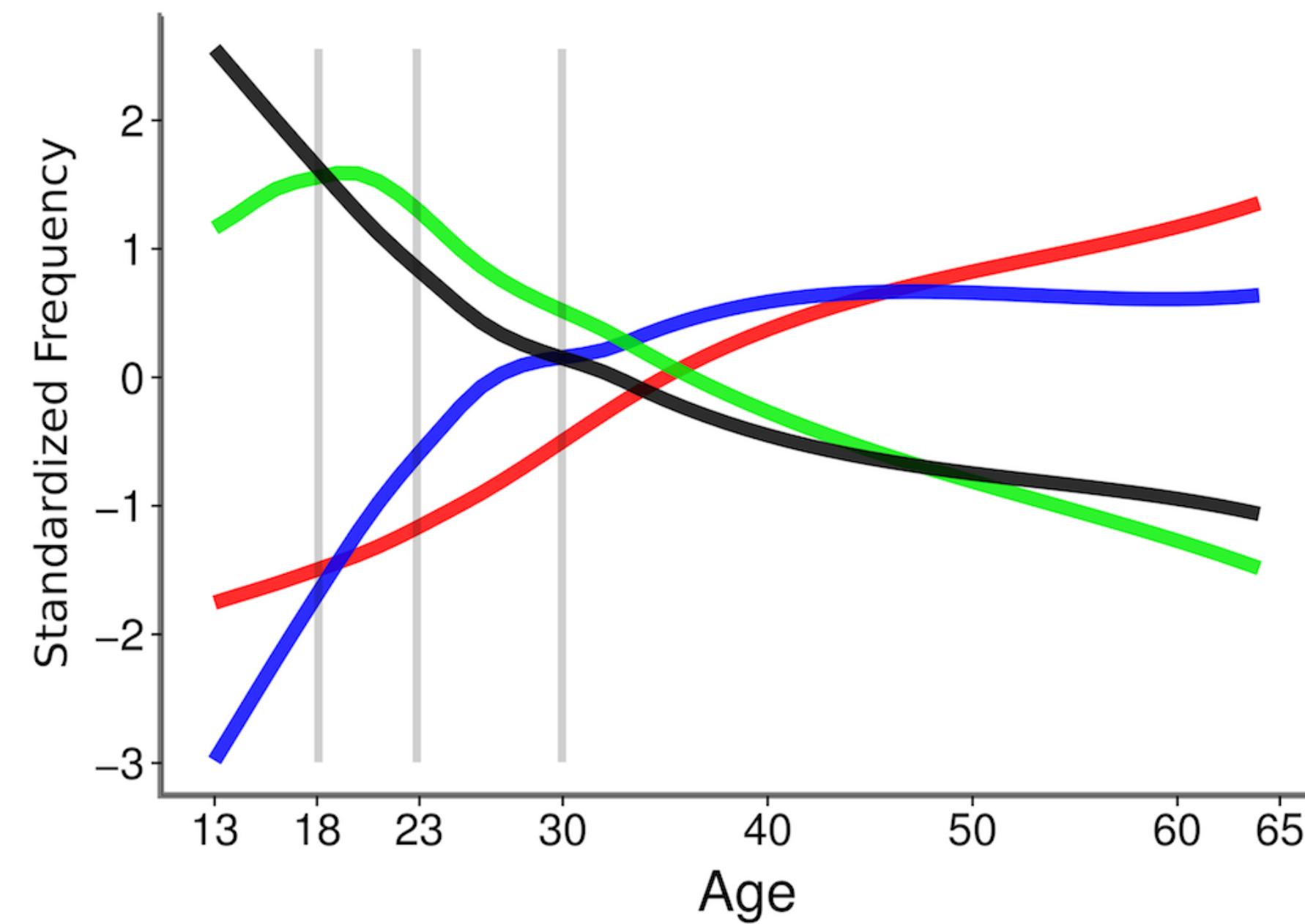


Source: Schwartz et al., PLoS ONE 2013

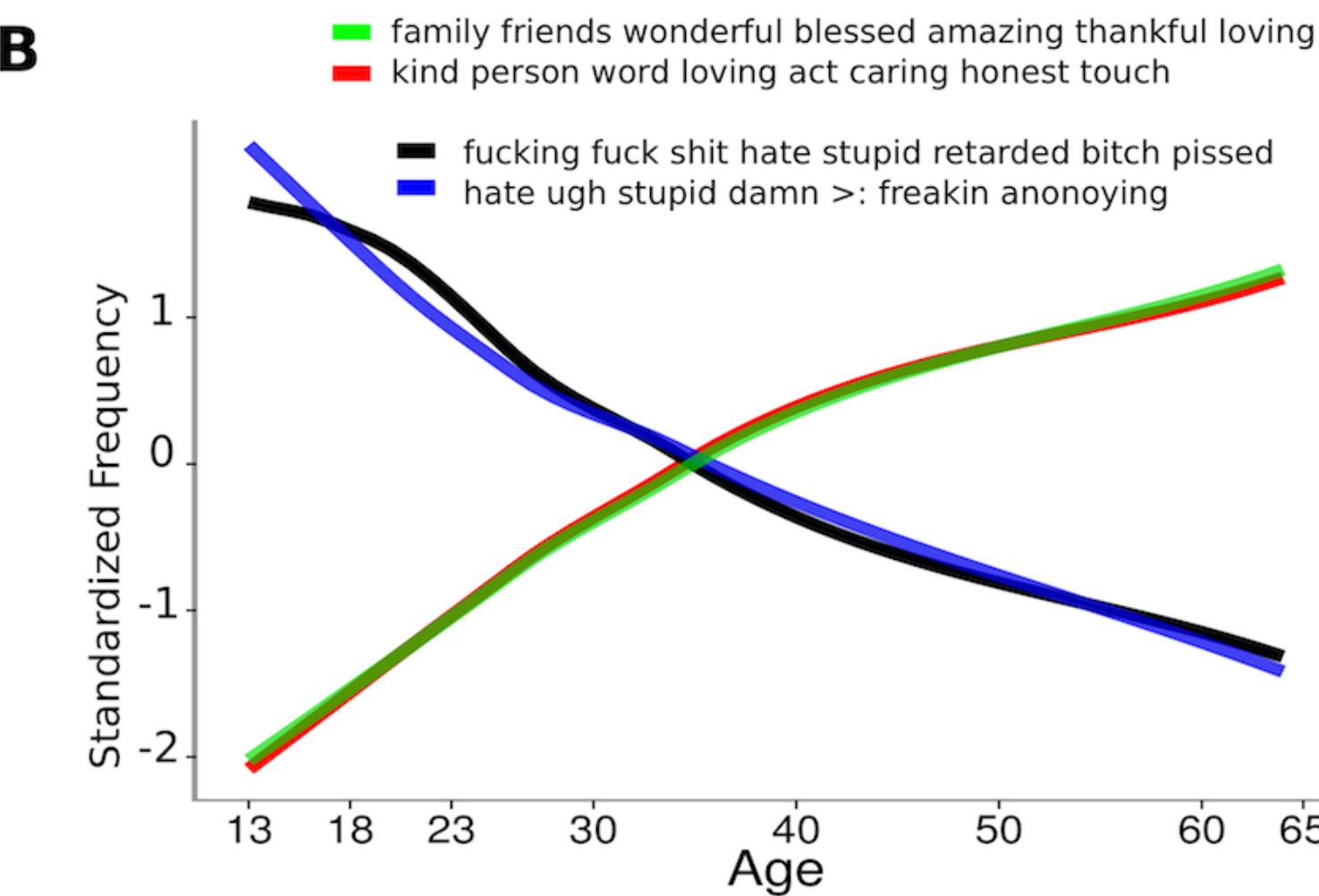
Facebook identification

A

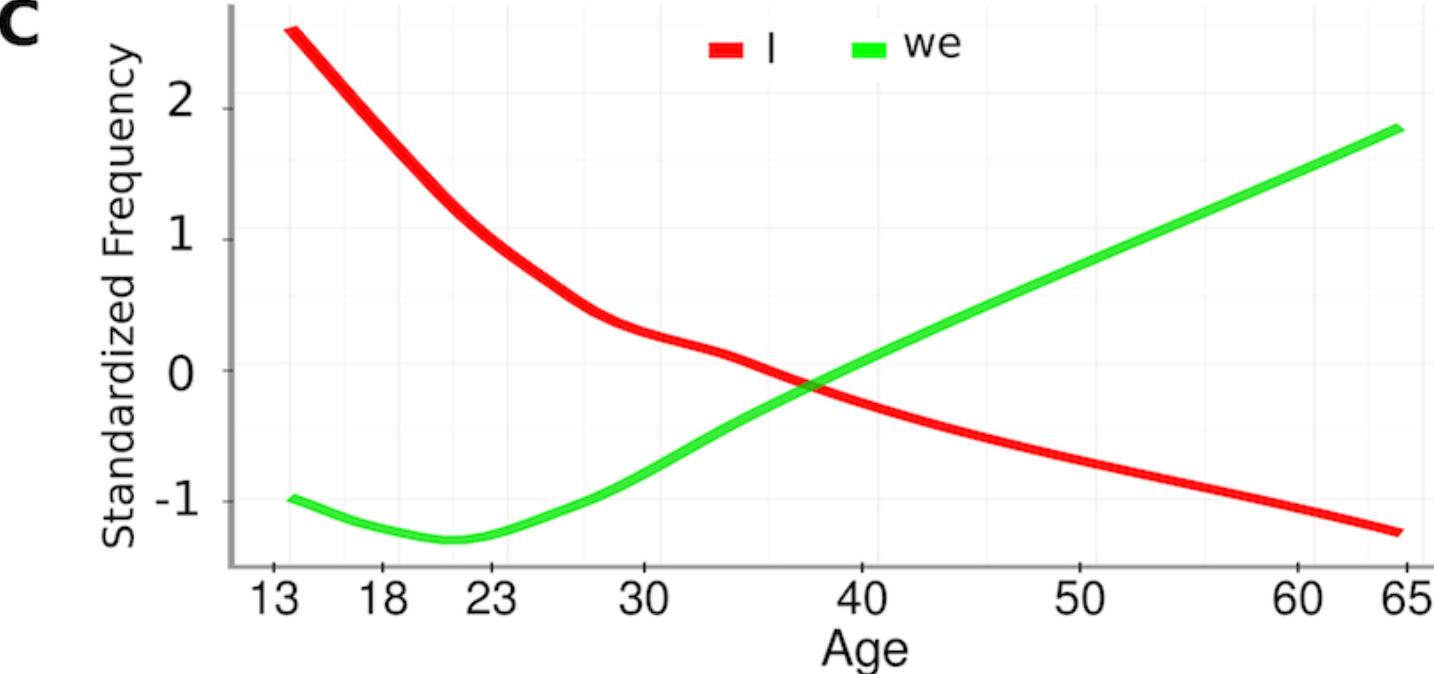
- (30 to 65) son daughter father mother proud oldest data youngest
(23 to 29) job position company manager interview experience office assistant
(19 to 22) classes semester class college schedule summer registered taking
(13 to 18) haha lol :p :D ; hehe jk ;p



B



C



When are we learning something new?

When are we amplifying biases?

When are we stereotyping?

So that's silly, but maybe not

harmful

Data Science can also be quite
powerful

Neural data science

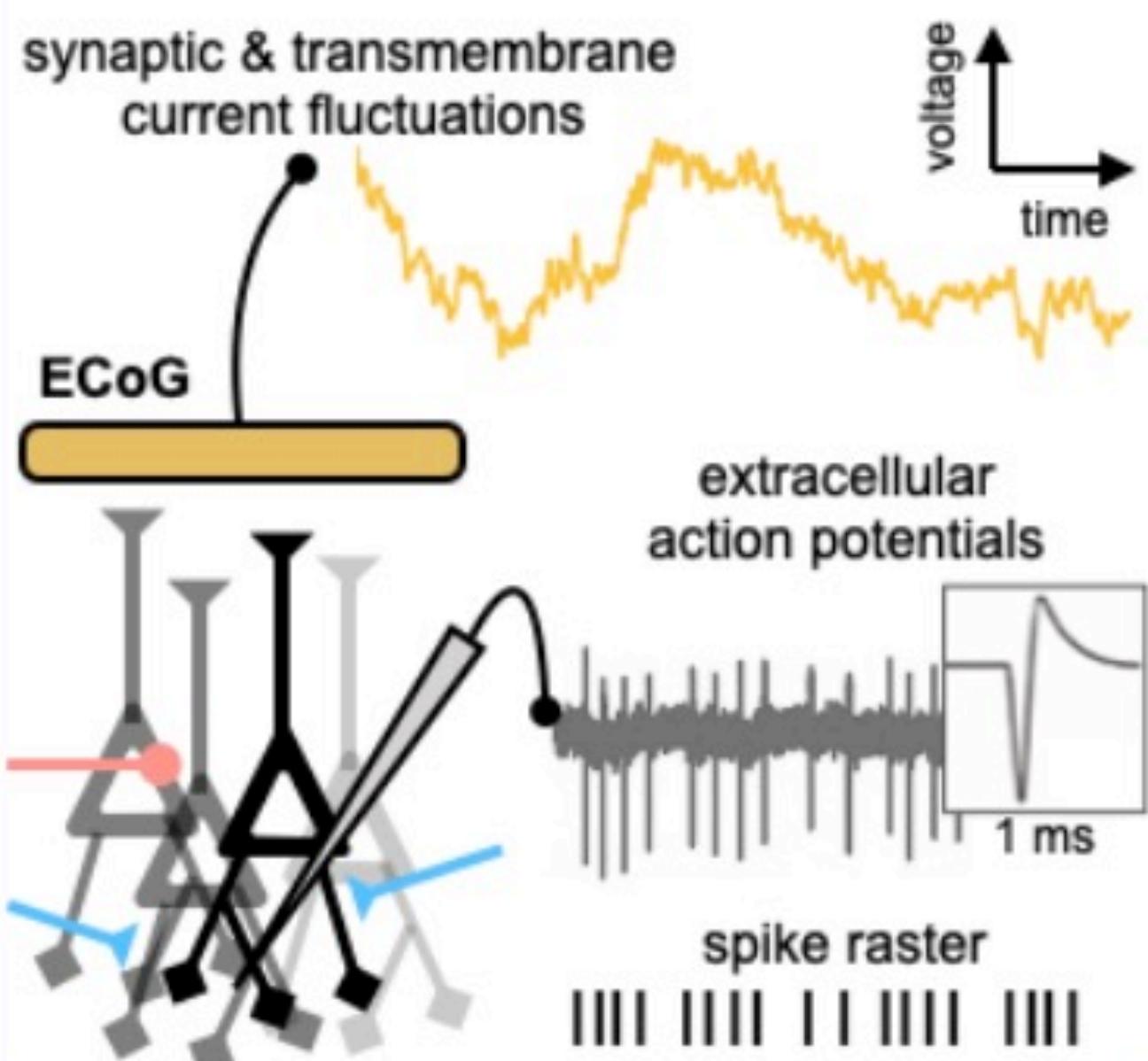
(COGS 138!)

Integrating tons of neural data

a

neural dynamics bridge structure and function

cellular & synaptic physiology
macaque ECoG & spiking



article roadmap
Fig. 2
Fig. 3
Fig. 4

estimating timescales
 τ

anatomical hierarchy
HCP T1w/T2w

neural dynamic timescale
human iEEG/ECoG

modulation of timescale
aging & working memory

transcriptomic basis
Allen Brain Atlas

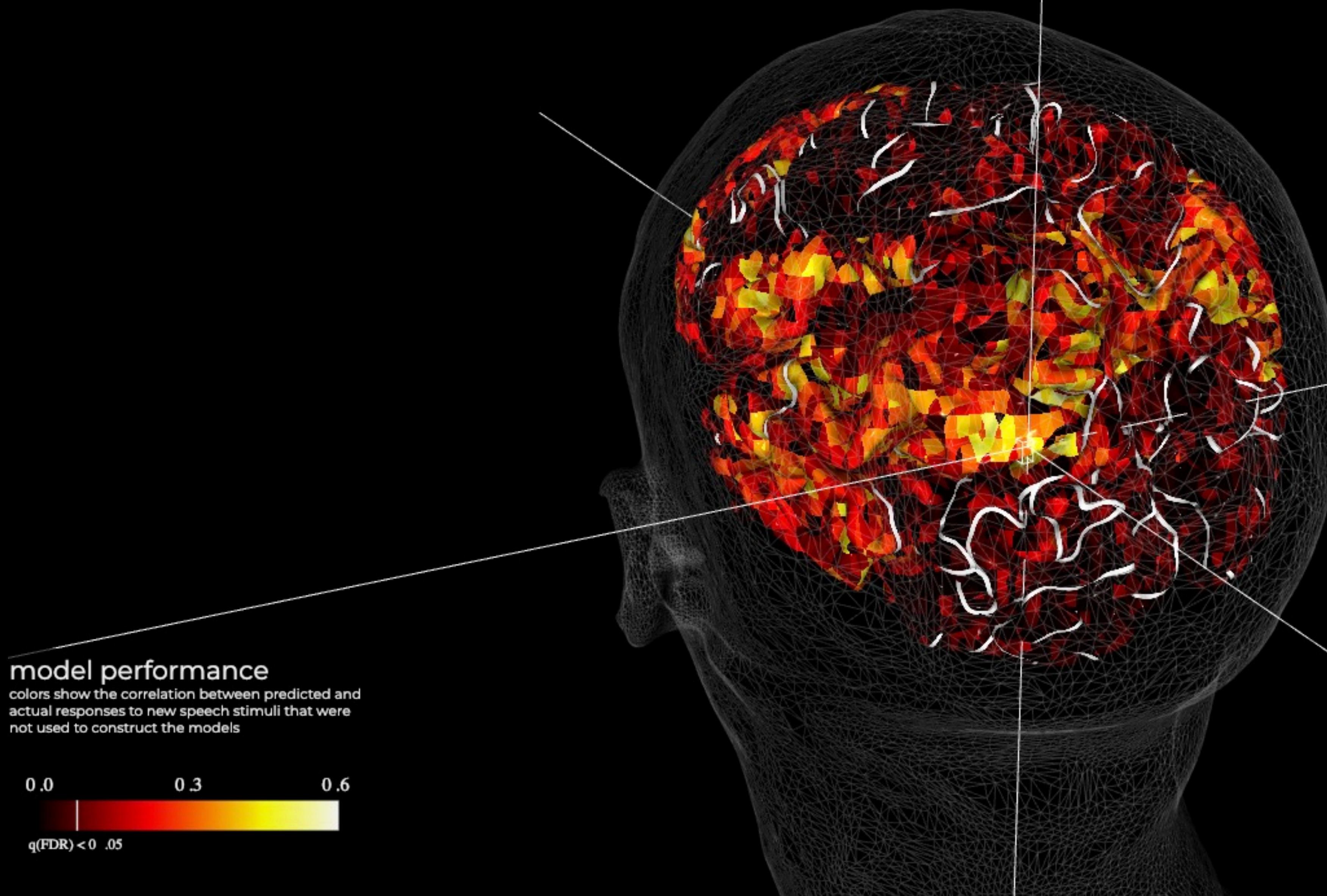
function
dynamics
structure

Richard Gao



• • • • • (show tour)

- camera
- data layers
 - selectivity
 - performance
 - performance_thr
 - PrAGMATIC
- surface
- Close Controls



**We can look for connections
between all of these—cell types,
gene expression, function,
oscillations, connectivity—to see
how they all interrelate.**

**And we can find gaps—missing
connections where topics *should* be
related but aren’t.**

We can mine through these
connections to find potential
missing links.

**That is, we can use the data
themselves to generate new
hypotheses *for us*.**

**But before we get to that
we need to be able to make
discoveries from our data.**

HOW?

Data Visualization & Communication

Section Learning Objective

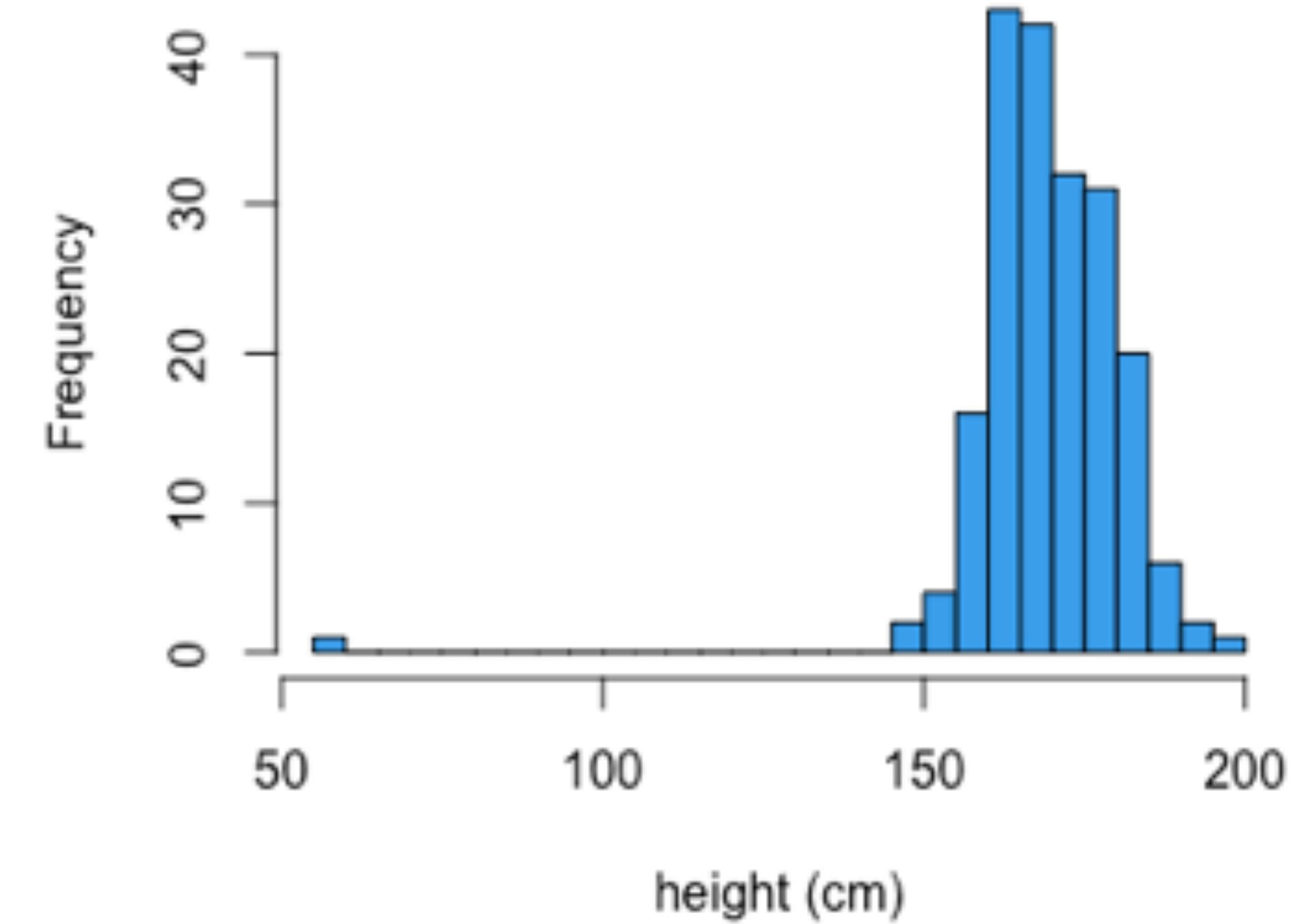
- *Explain appropriate visualization given data*
- *Interpret graphs and tables effectively*
- *Diagnose issues with problematic visualizations*

**There are many ways to
visualize data.**

**Some ways are more effective
than others.**

Histogram

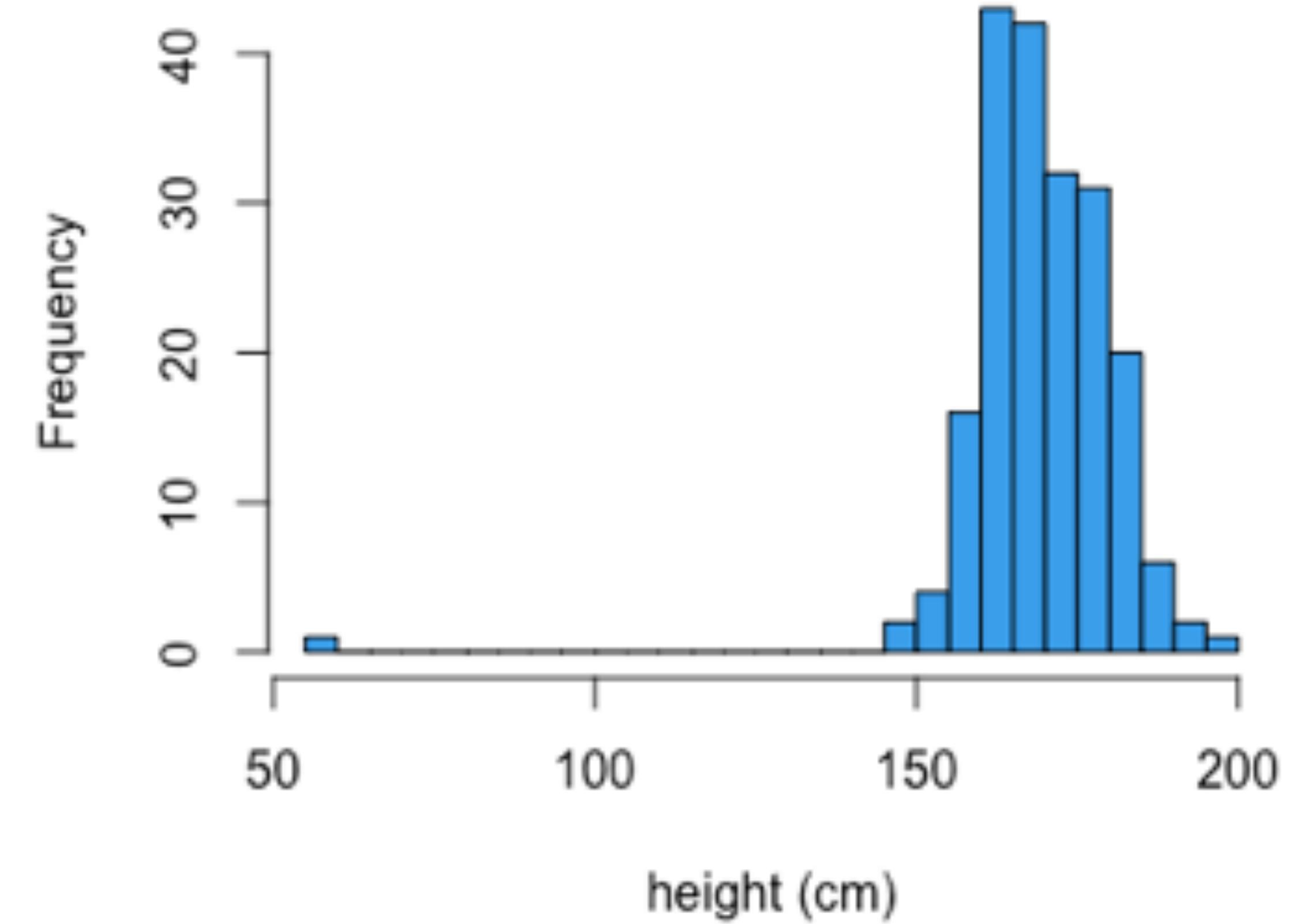
- Distribution of counts or proportions about a single set of numbers



Histogram

- Distribution of counts or proportions about a single set of numbers

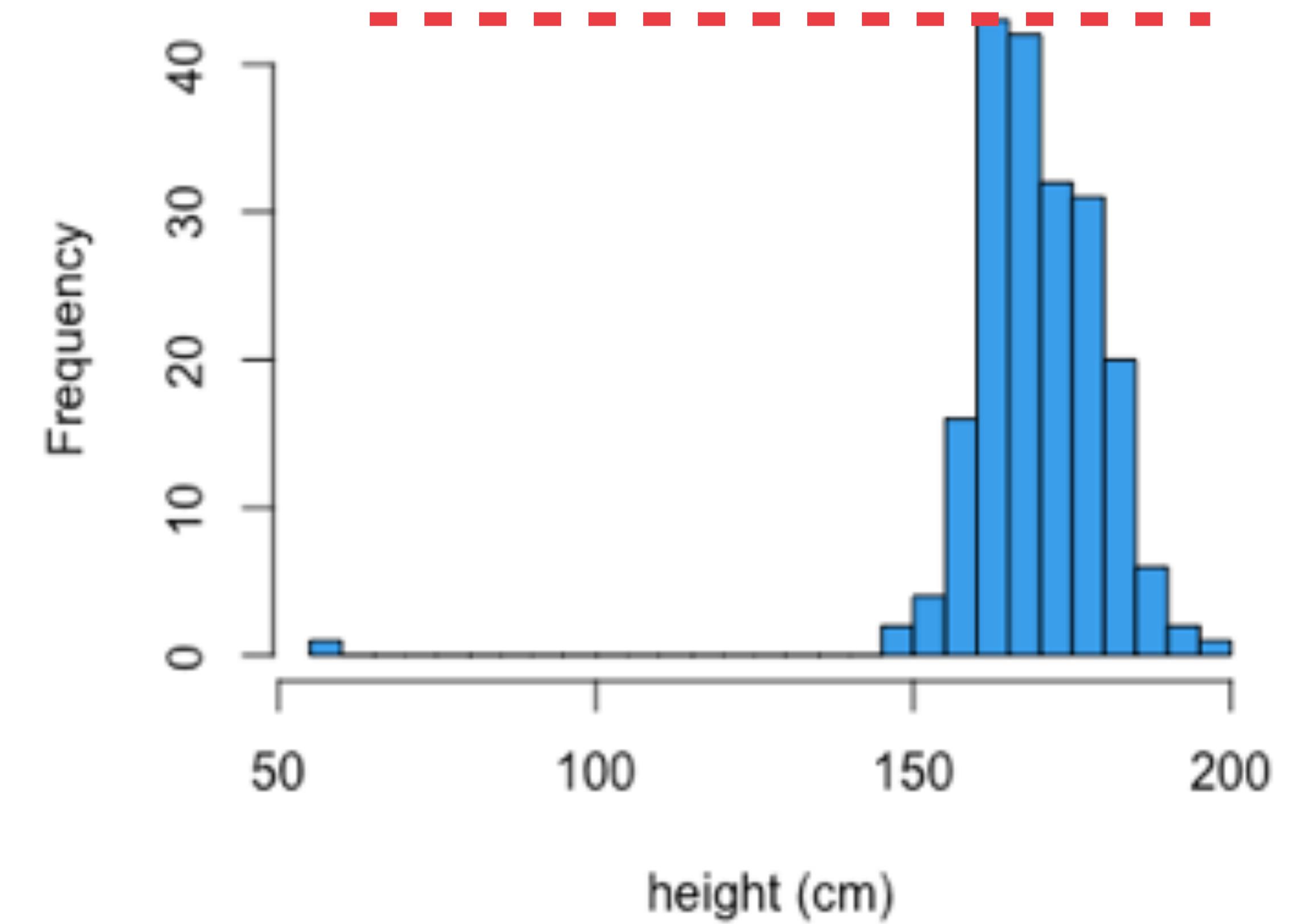
range of values easily visualized



Histogram

- Distribution of counts or proportions about a single set of numbers

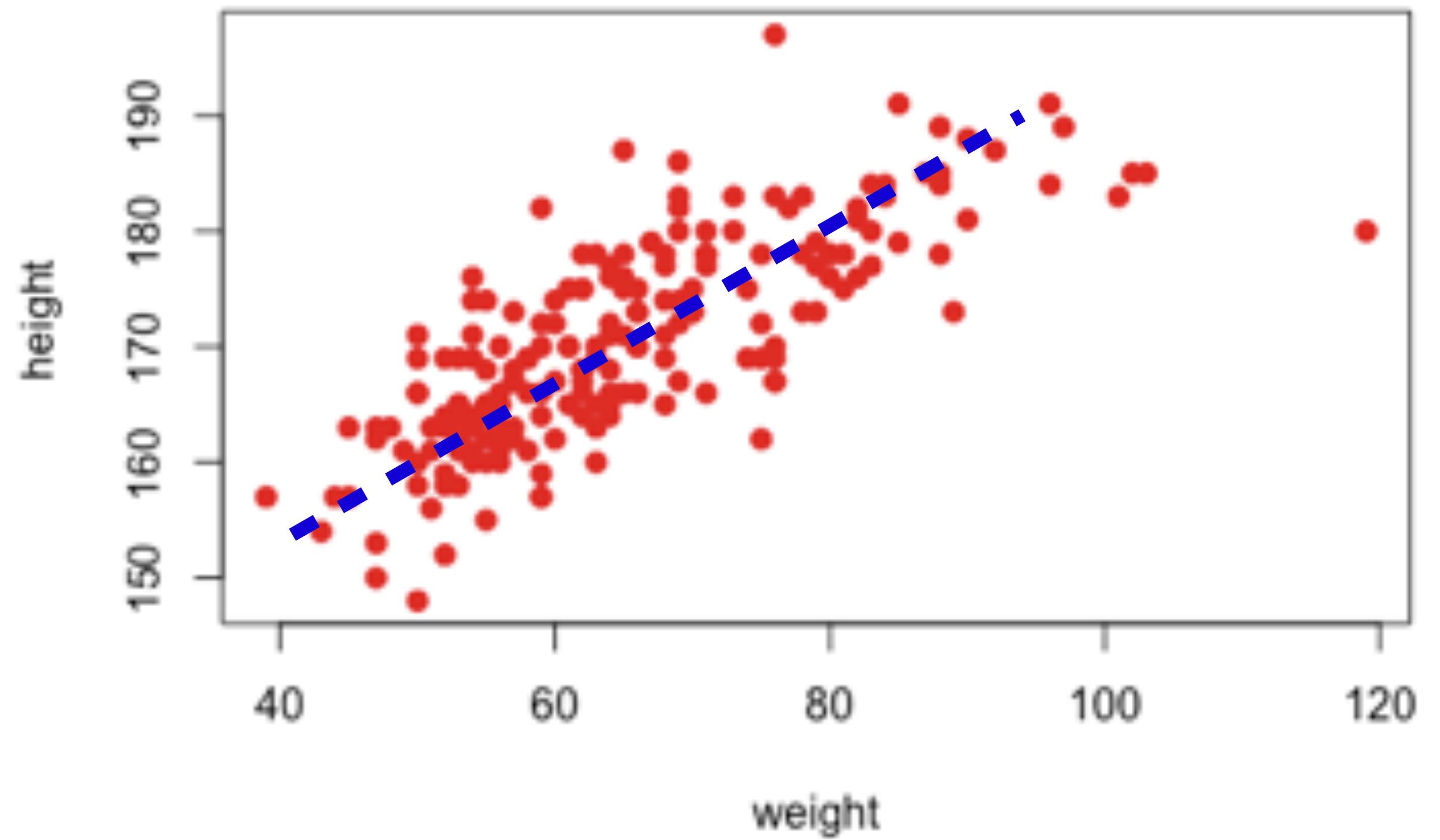
range of values easily visualized



Scatterplot

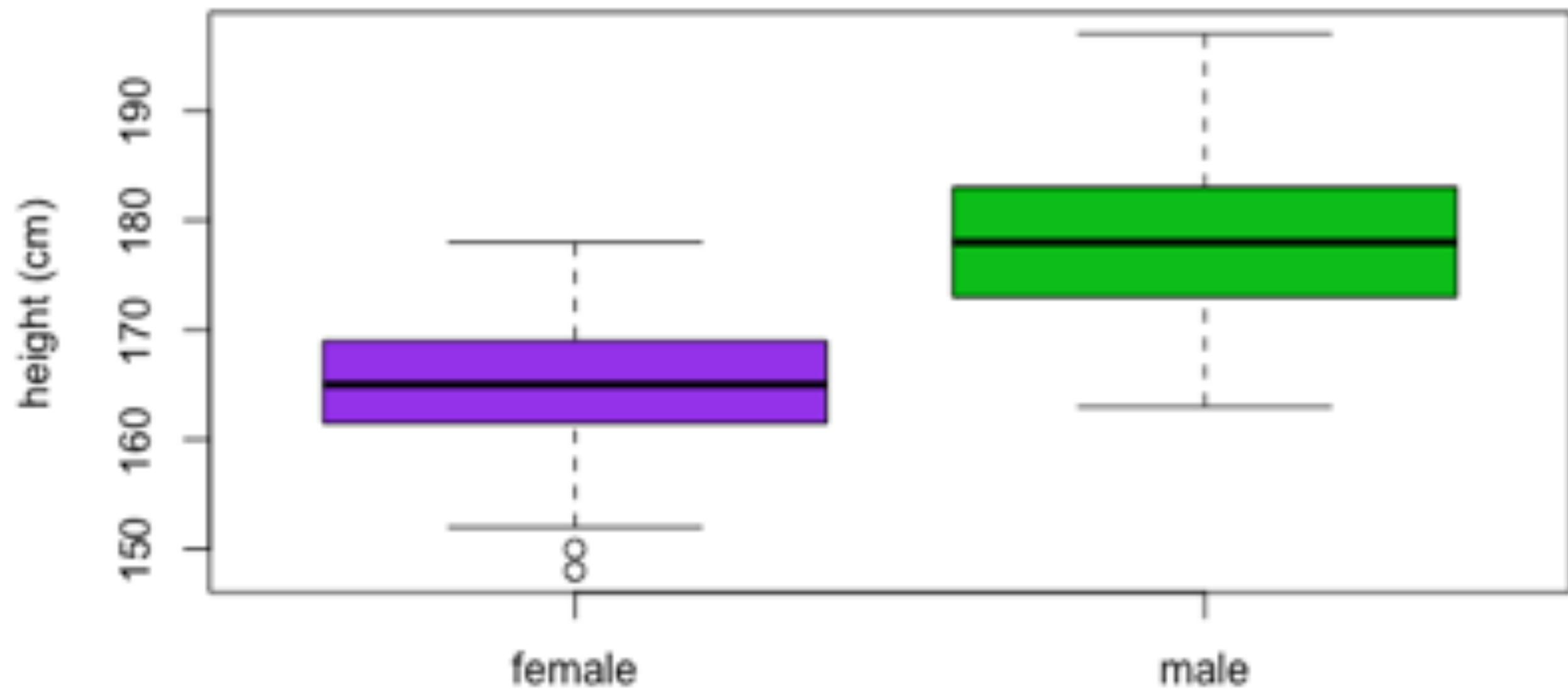
- Relationship between two numerical variables

*generally the more you weigh,
the taller you are*



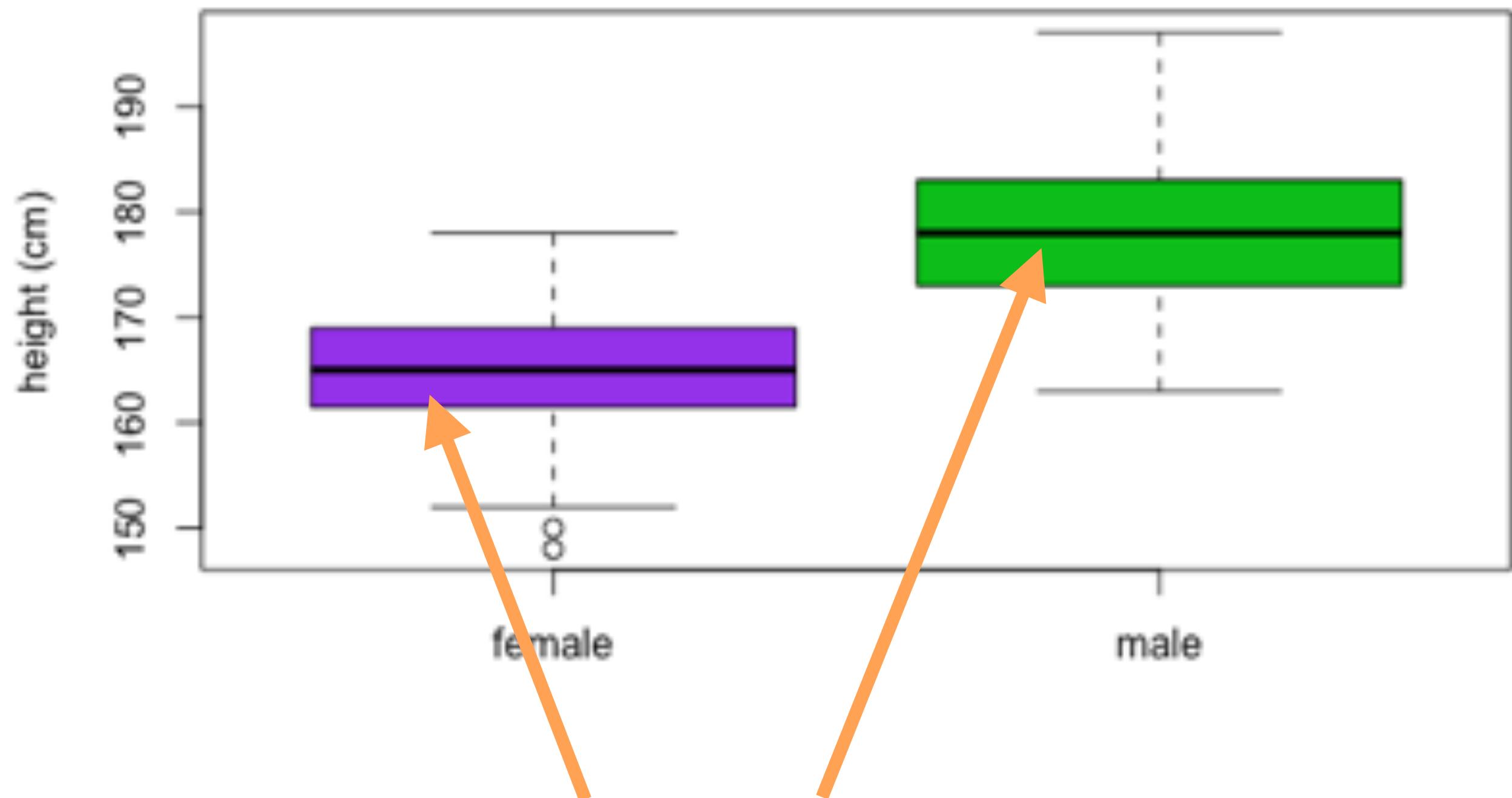
Boxplot

- Summary of numerical values across categories



Boxplot

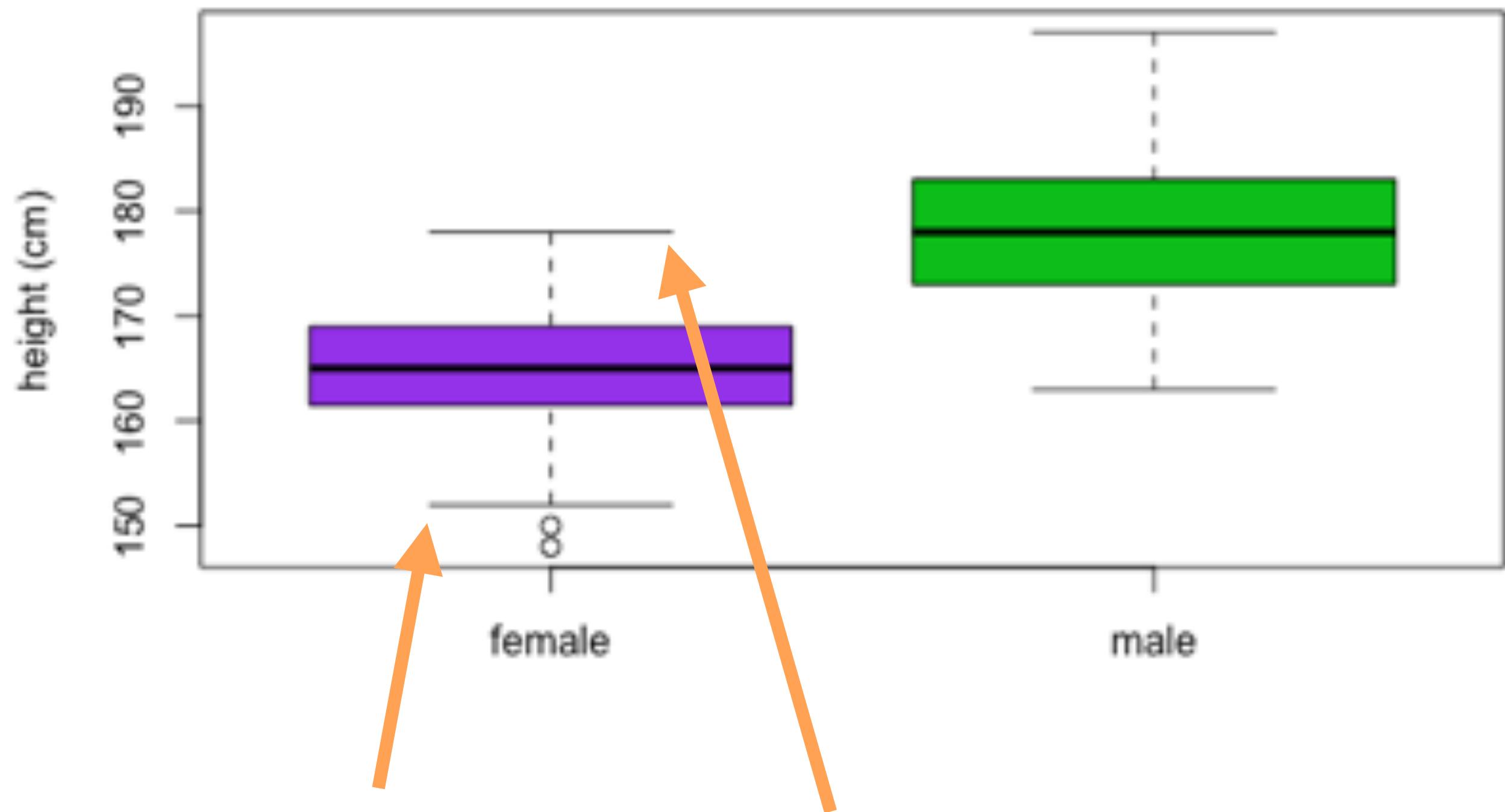
- Summary of numerical values across categories



The middle line represents the *central tendency*

Boxplot

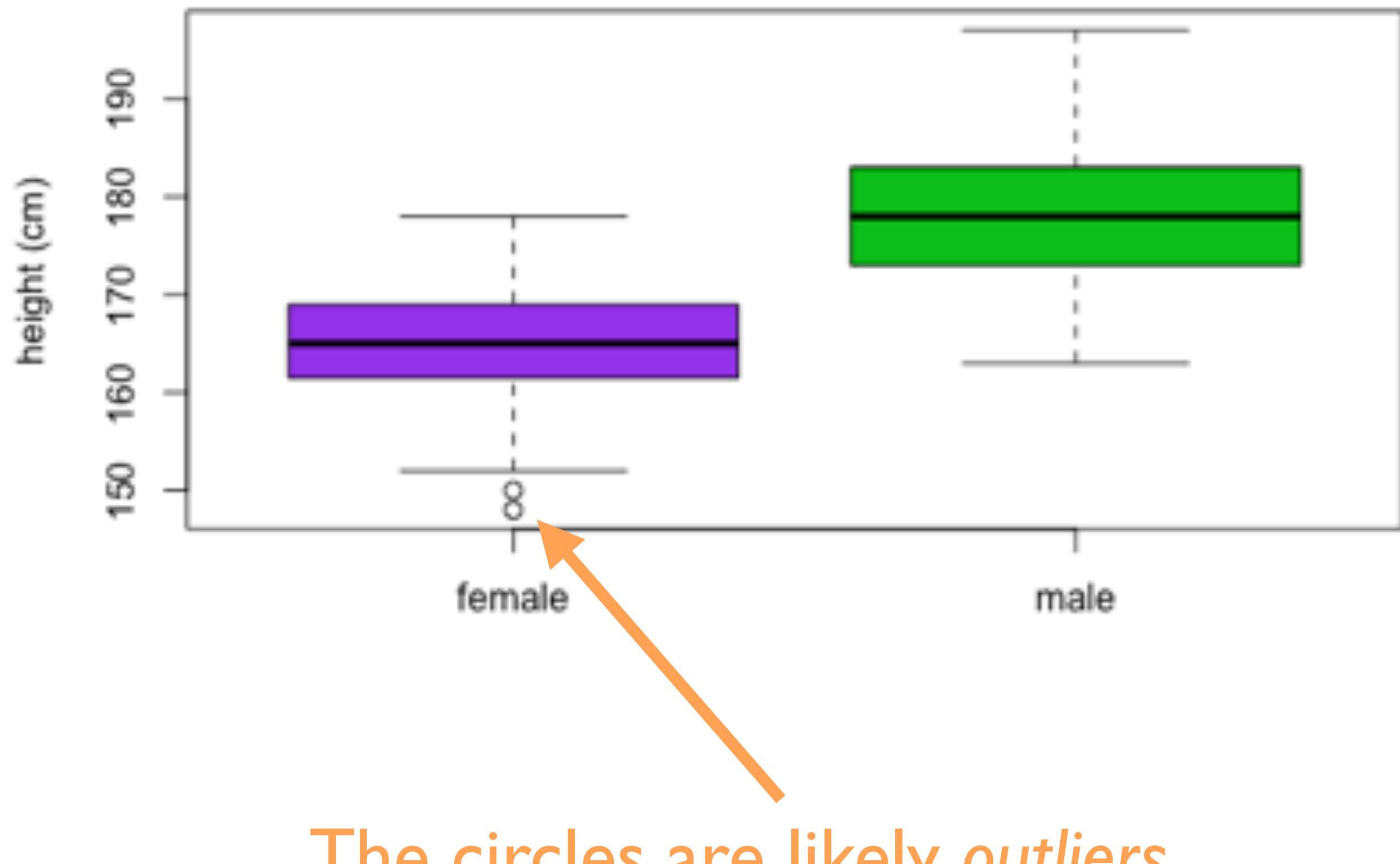
- Summary of numerical values across categories



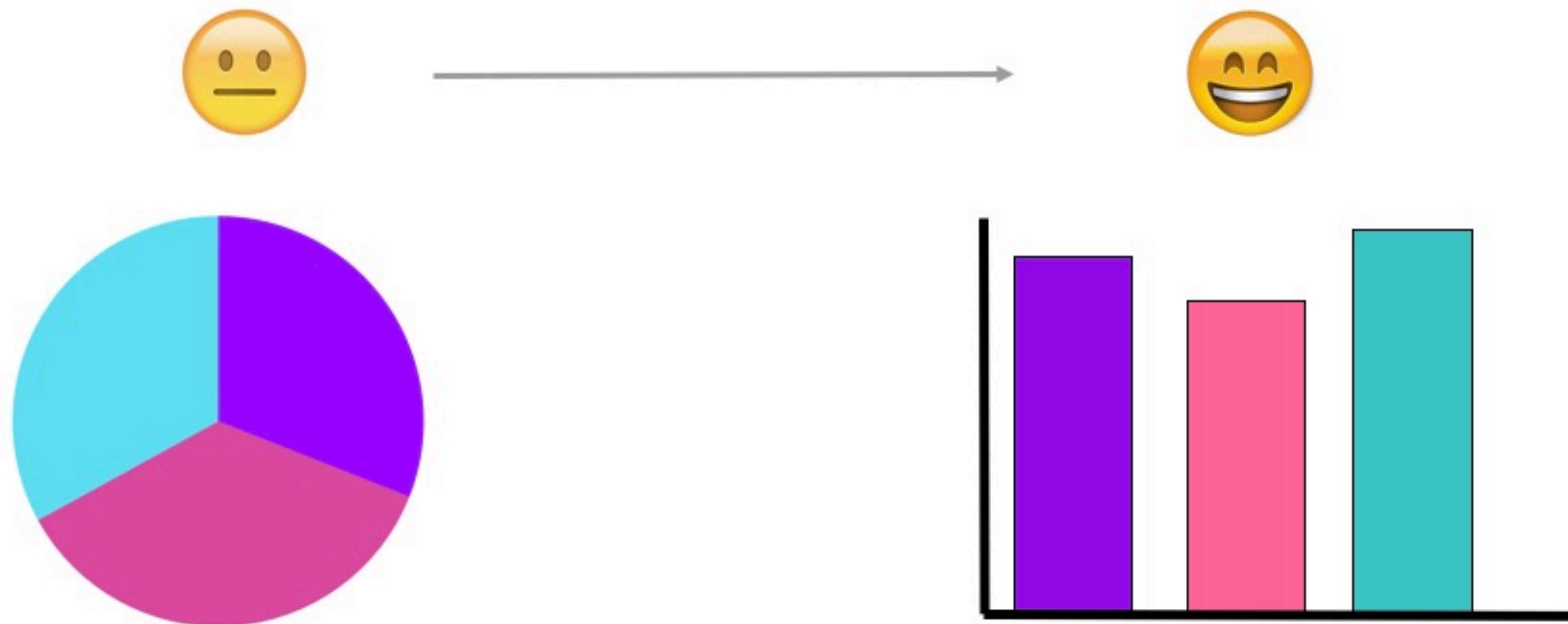
The *whiskers* give you an idea
of the typical *range*

Boxplot

- Summary of numerical values across categories

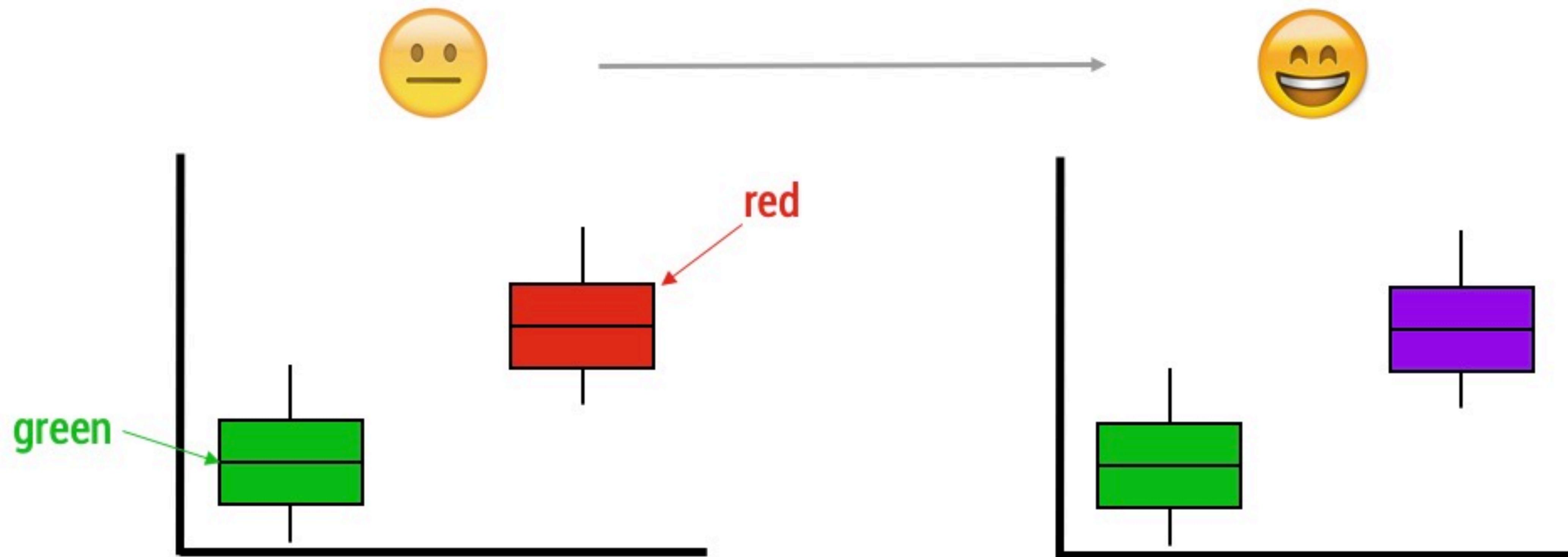


Choose the right visualization



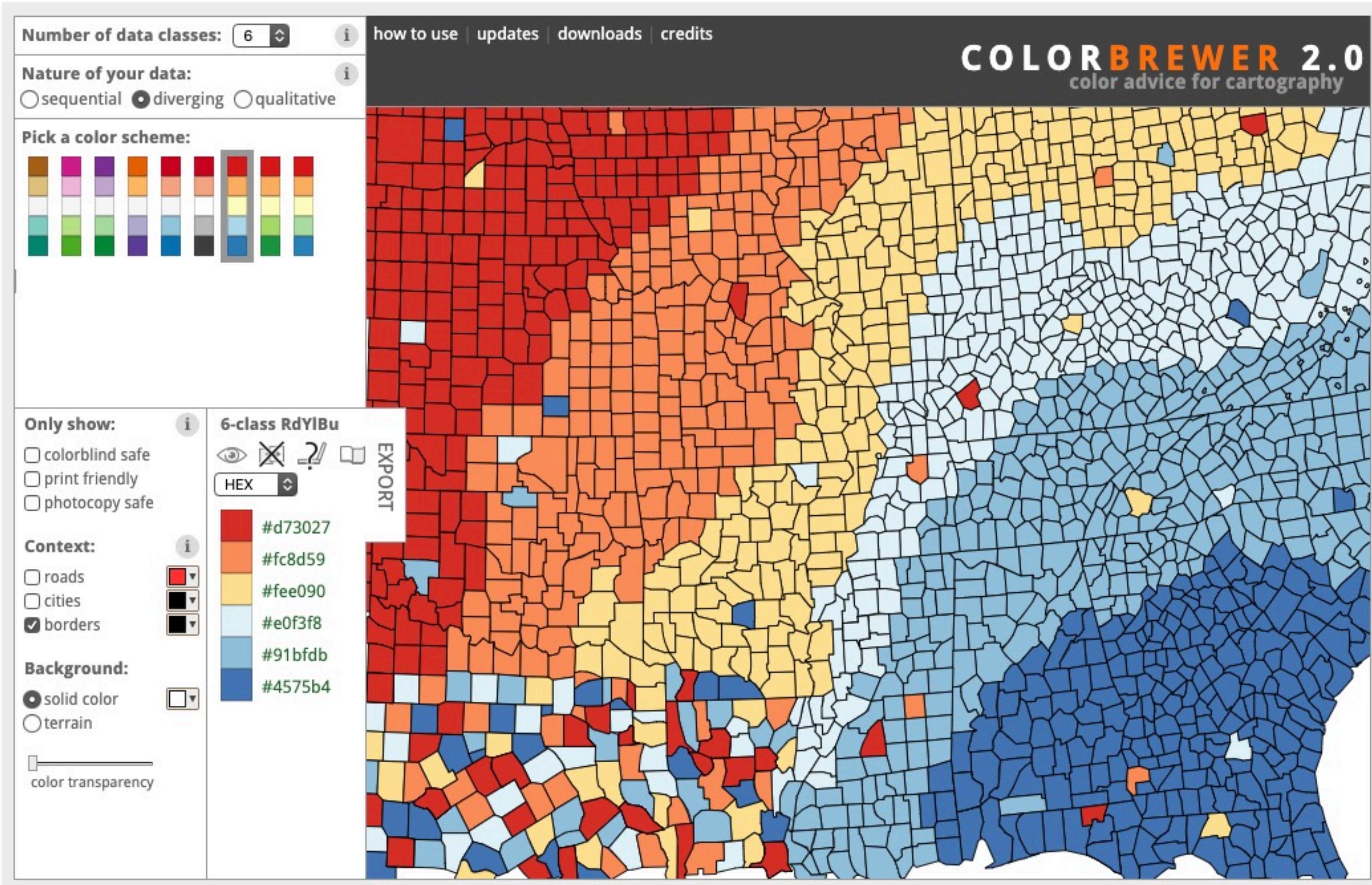
When looking at values, bar charts make it much easier to see the difference between groups!

Be mindful of colors

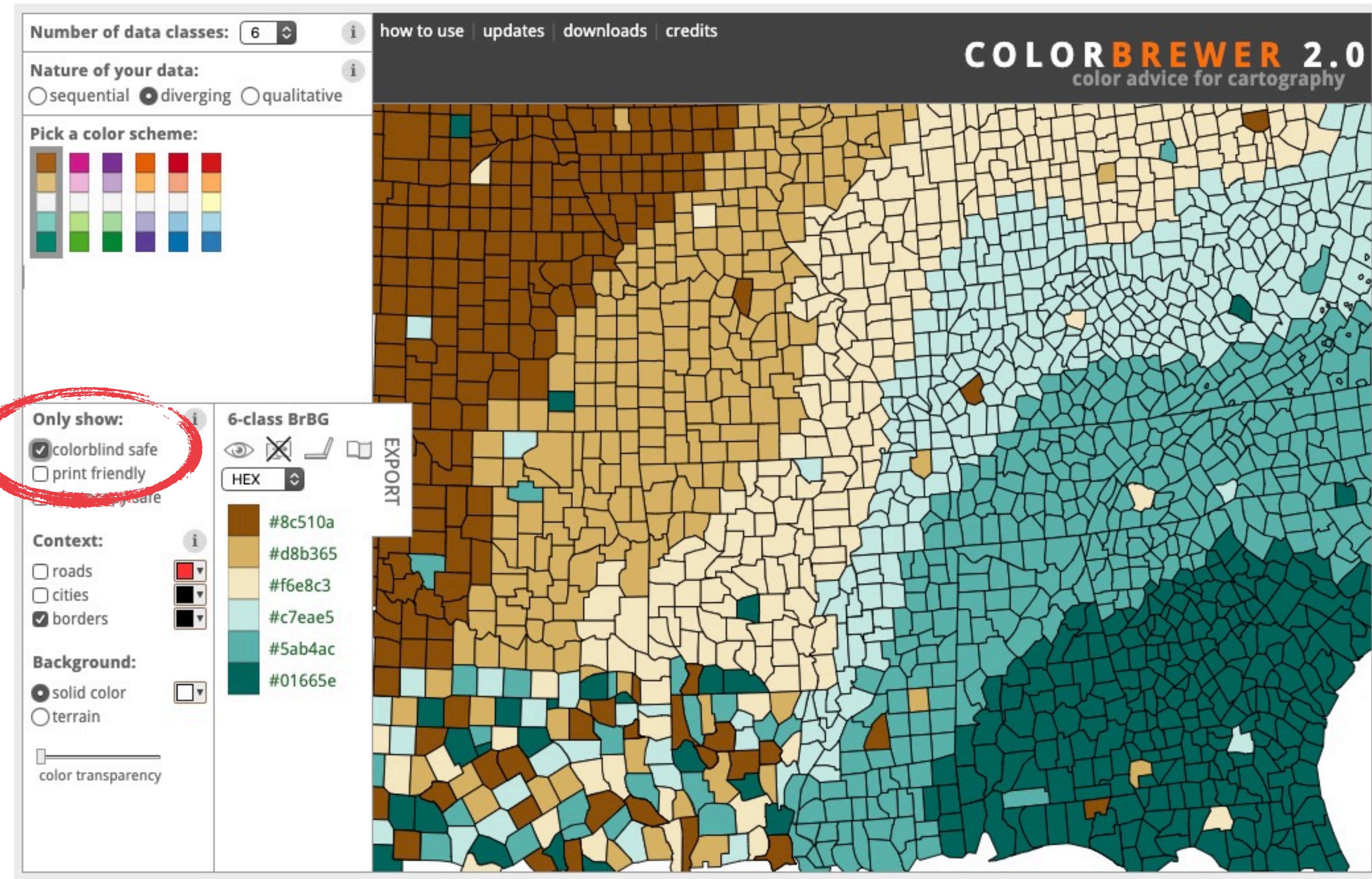


Many color-blind individuals cannot see
the difference between red and green.

Be mindful of colors



Be mindful of colors



**What are tricks
that help make visualizations
more effective and memorable?**

Data-to-ink Ratio

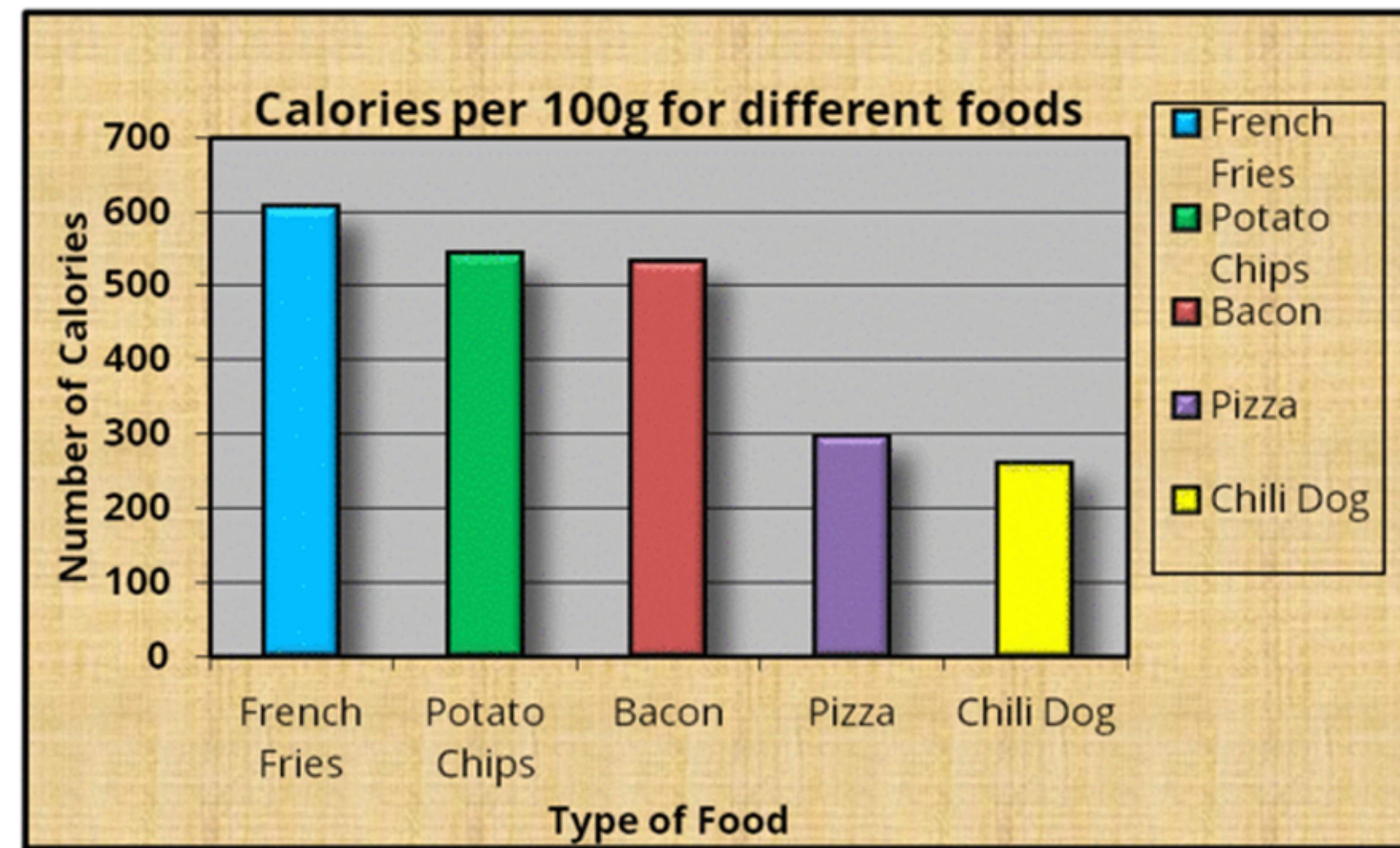
Remove
to improve
(the **data-ink** ratio)

Created by Darkhorse Analytics

www.darkhorseanalytics.com

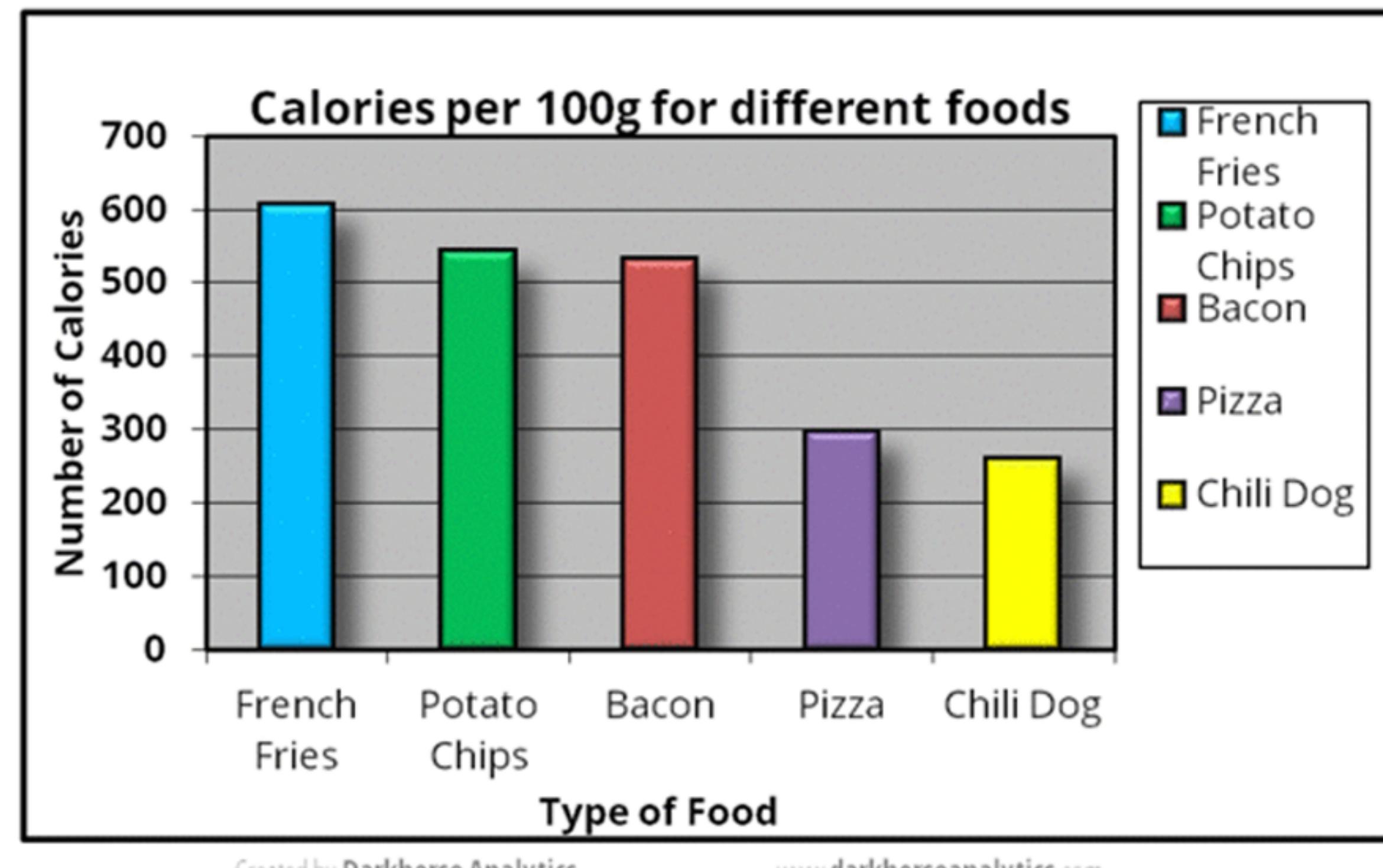
Data-to-ink Ratio

Remove backgrounds



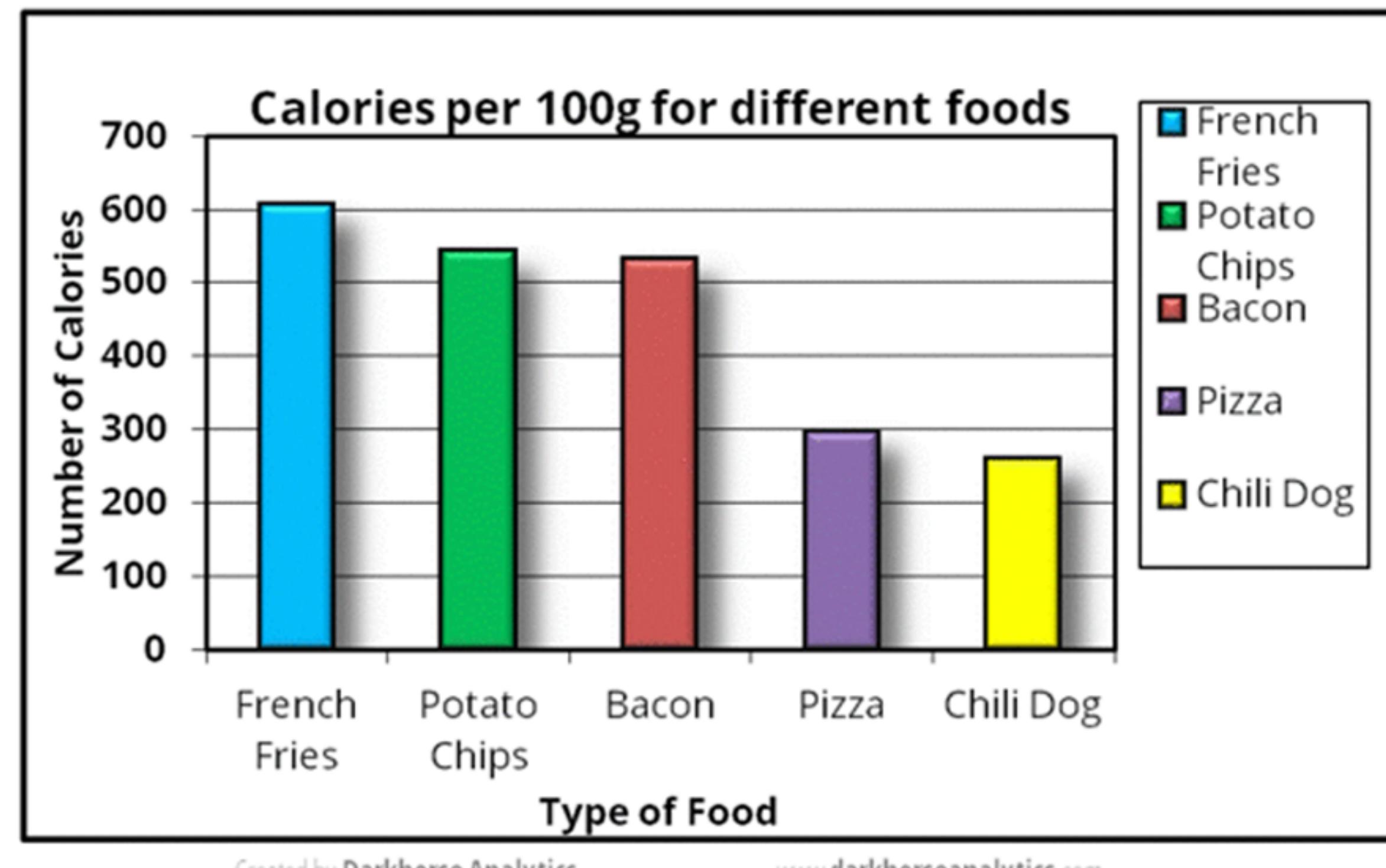
Data-to-ink Ratio

Remove backgrounds



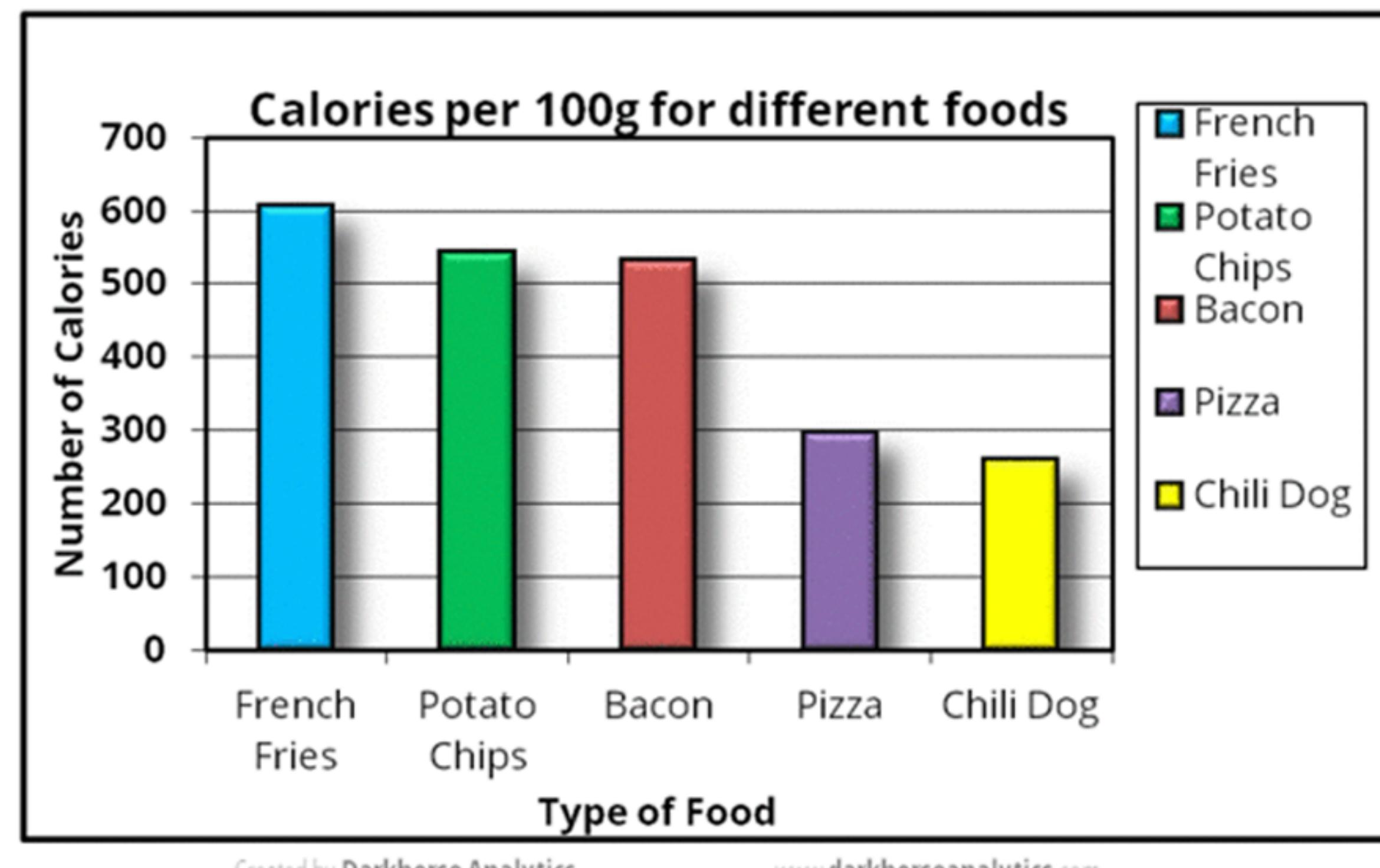
Data-to-ink Ratio

Remove backgrounds



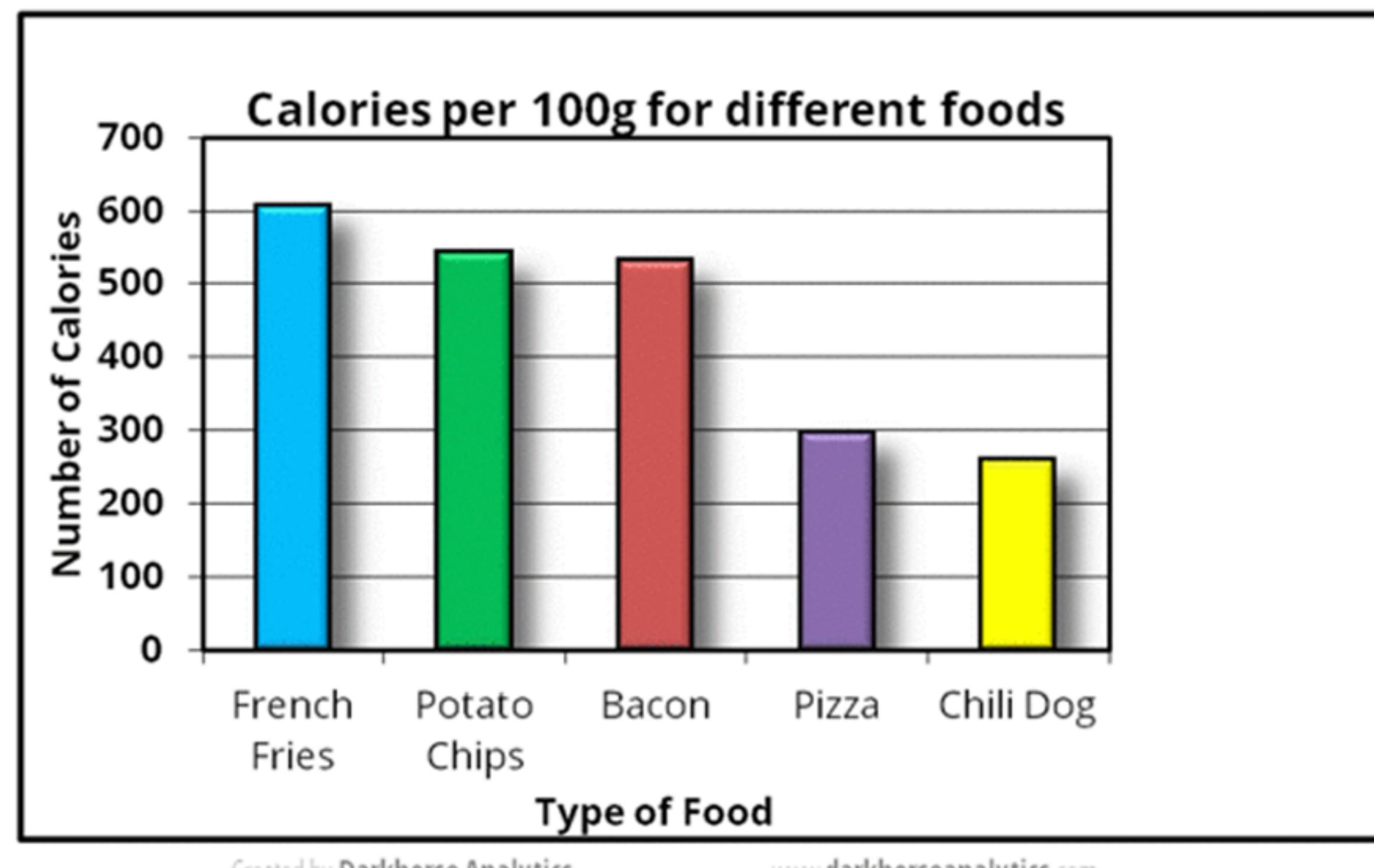
Data-to-ink Ratio

Remove redundant labels



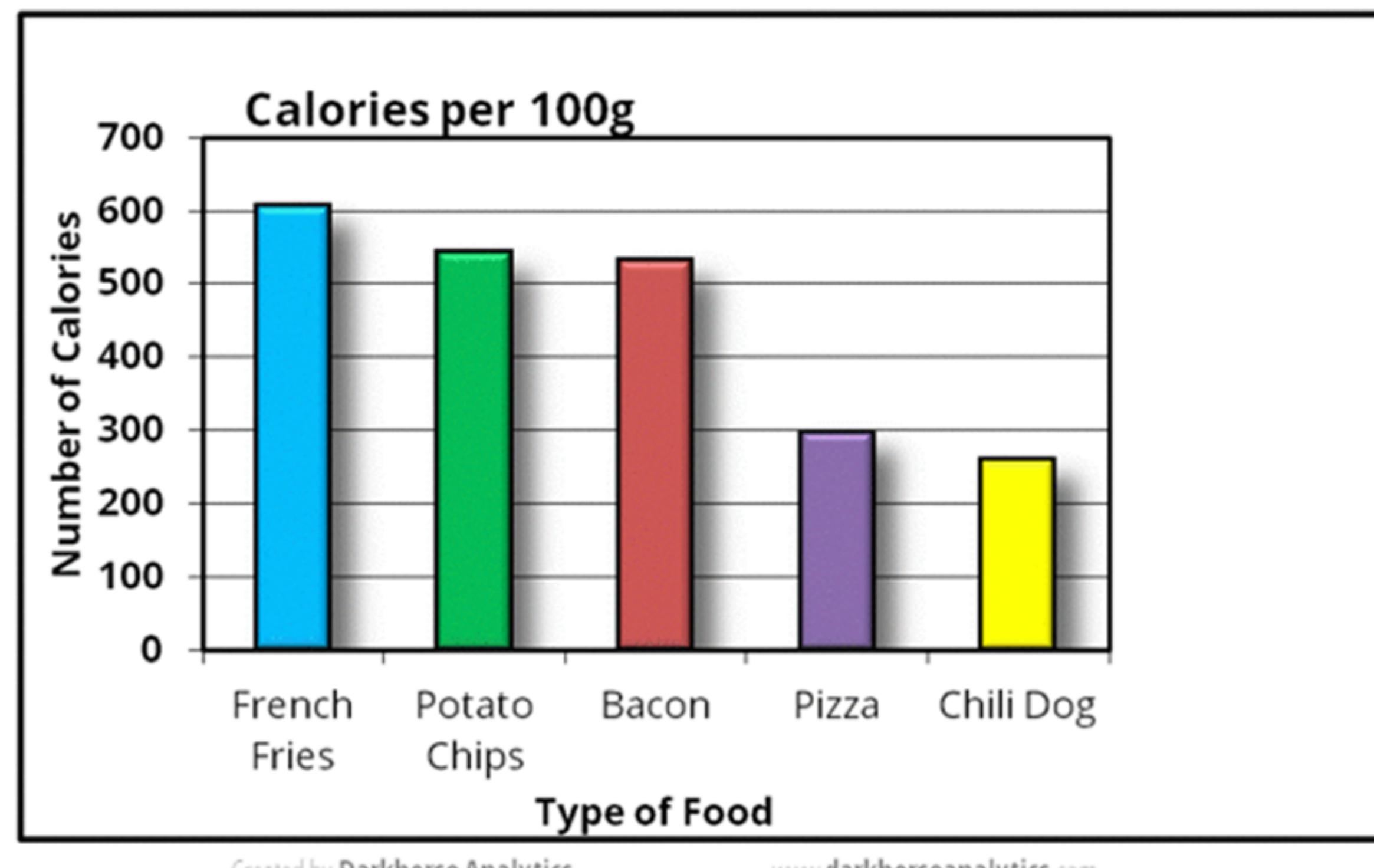
Data-to-ink Ratio

Remove redundant labels



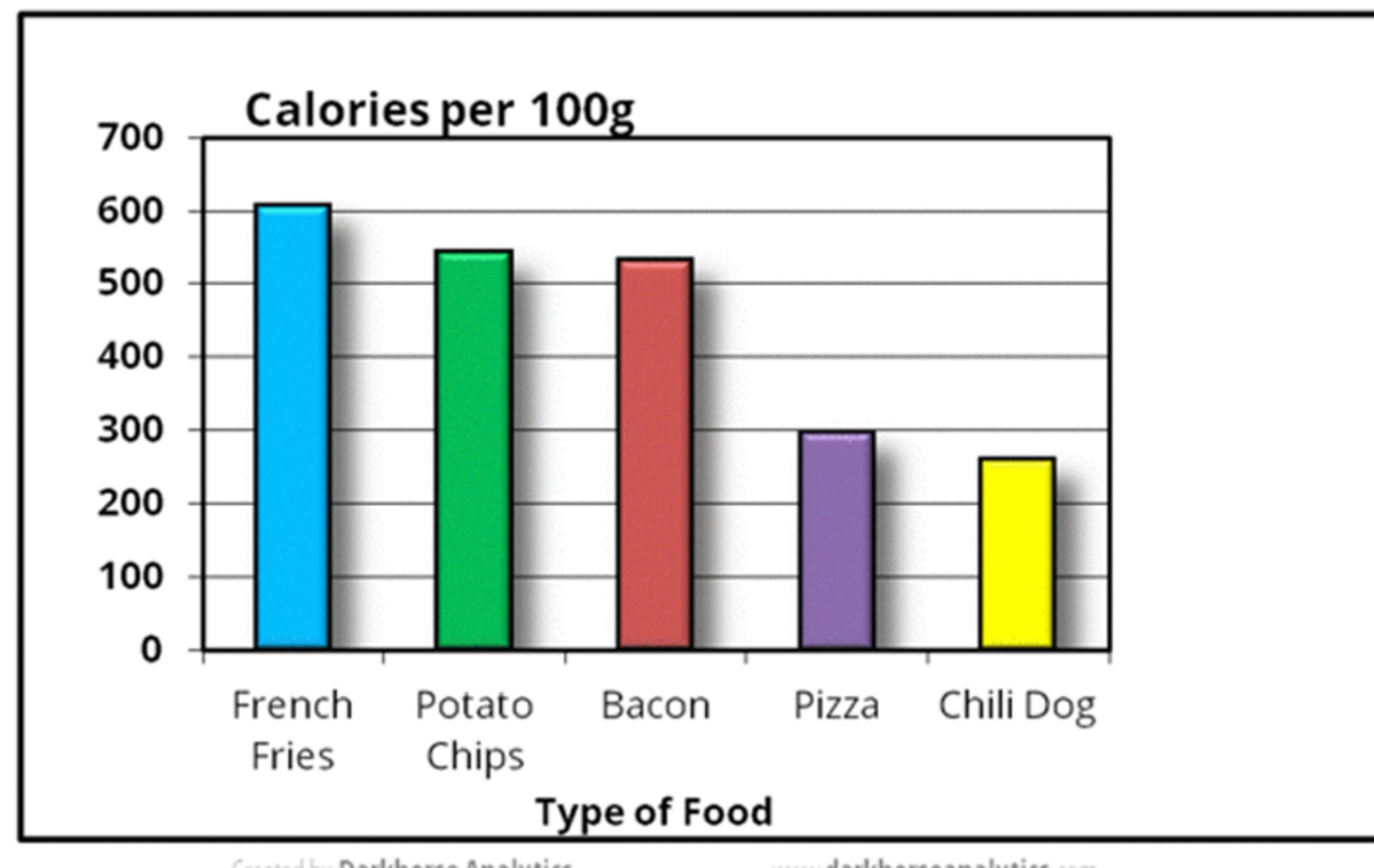
Data-to-ink Ratio

Remove redundant labels



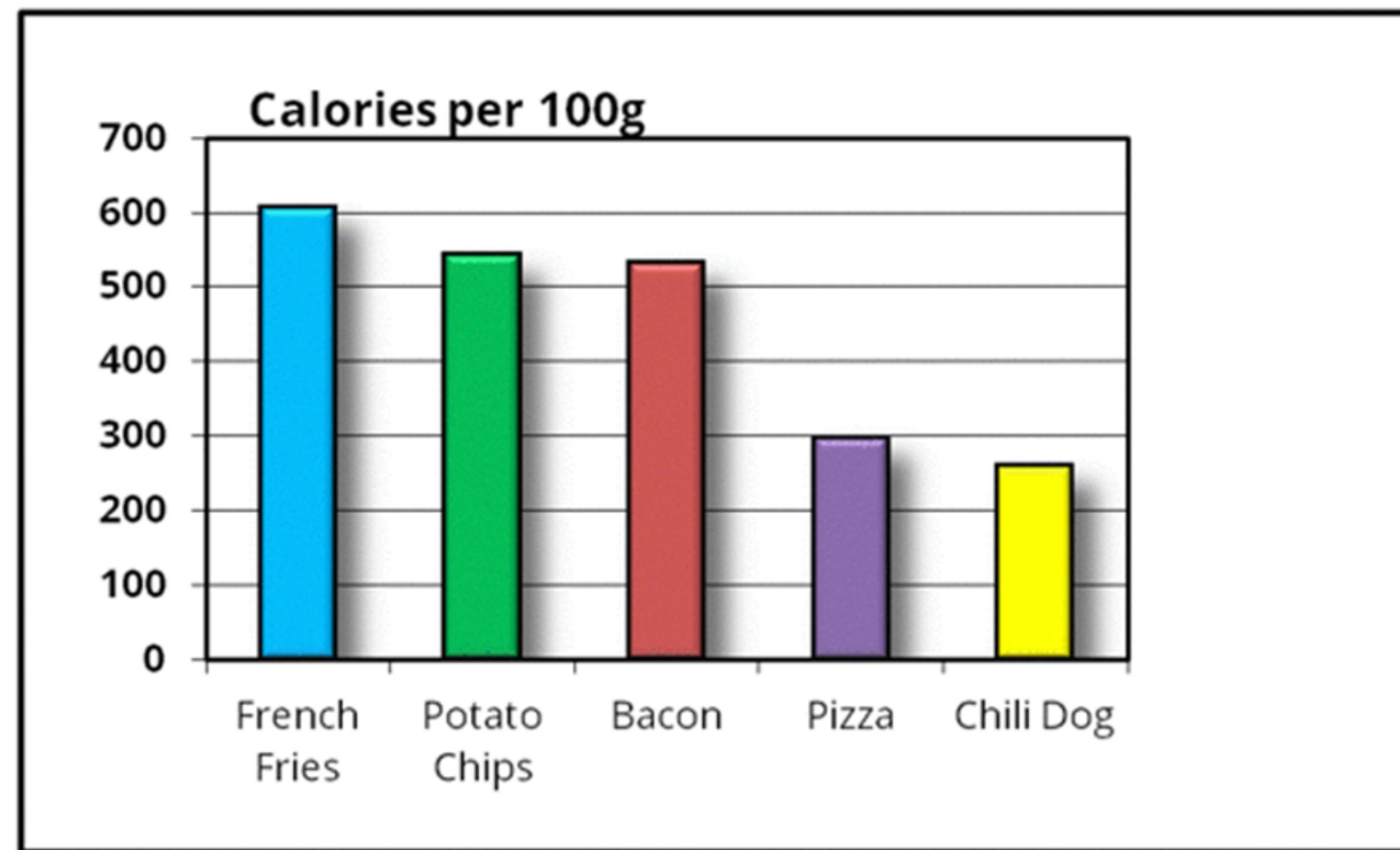
Data-to-ink Ratio

Remove redundant labels



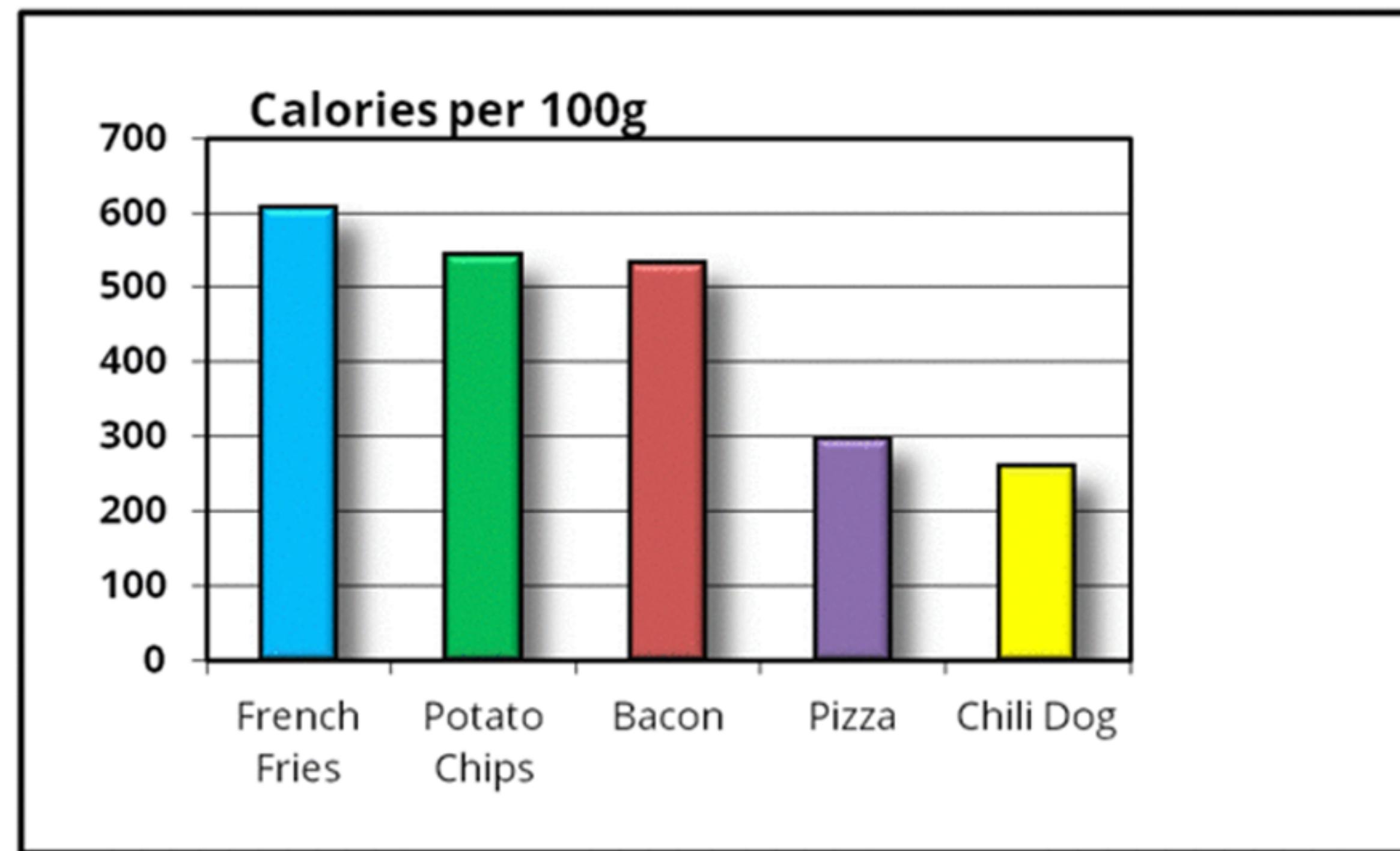
Data-to-ink Ratio

Remove redundant labels



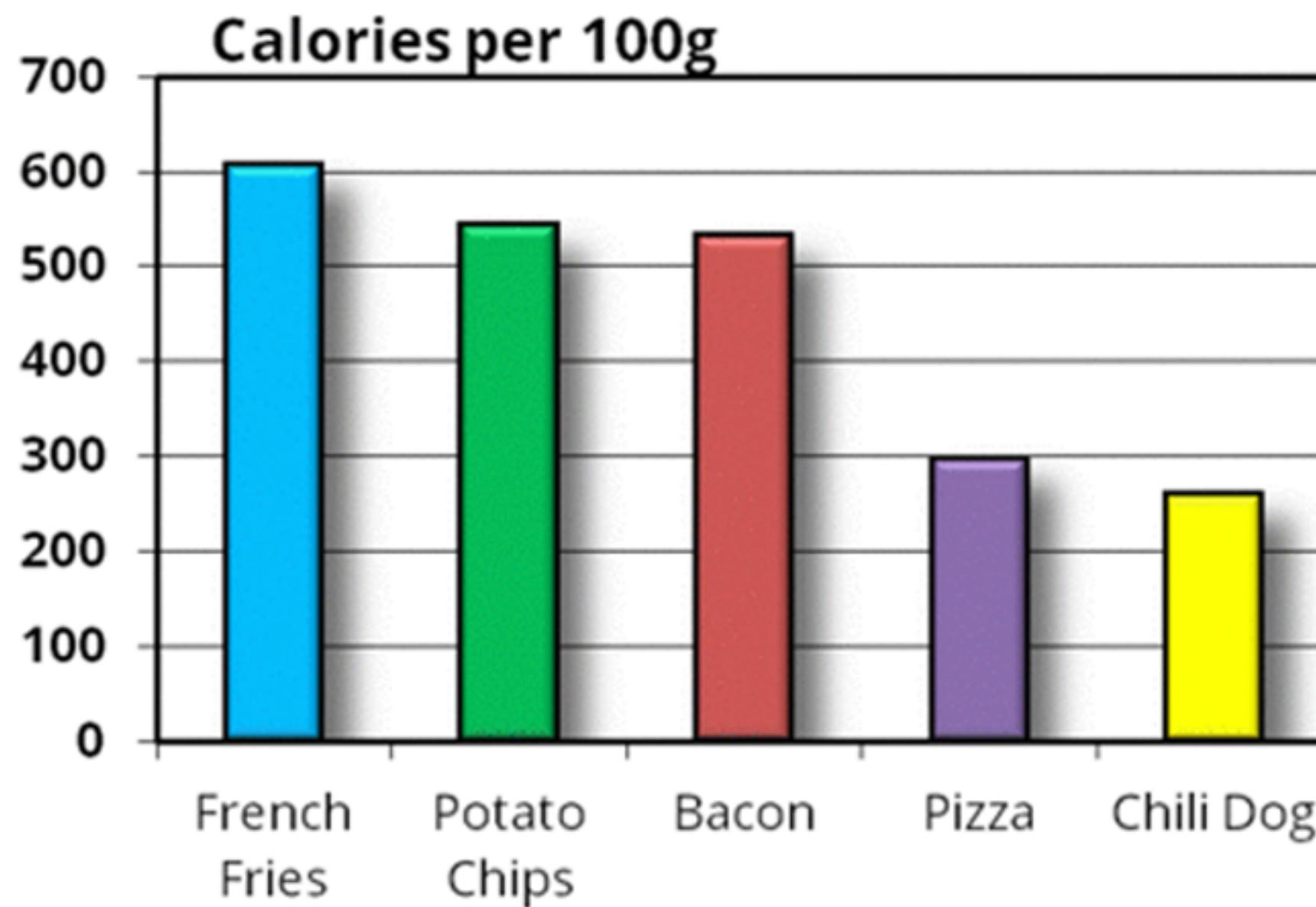
Data-to-ink Ratio

Remove borders



Data-to-ink Ratio

Remove borders

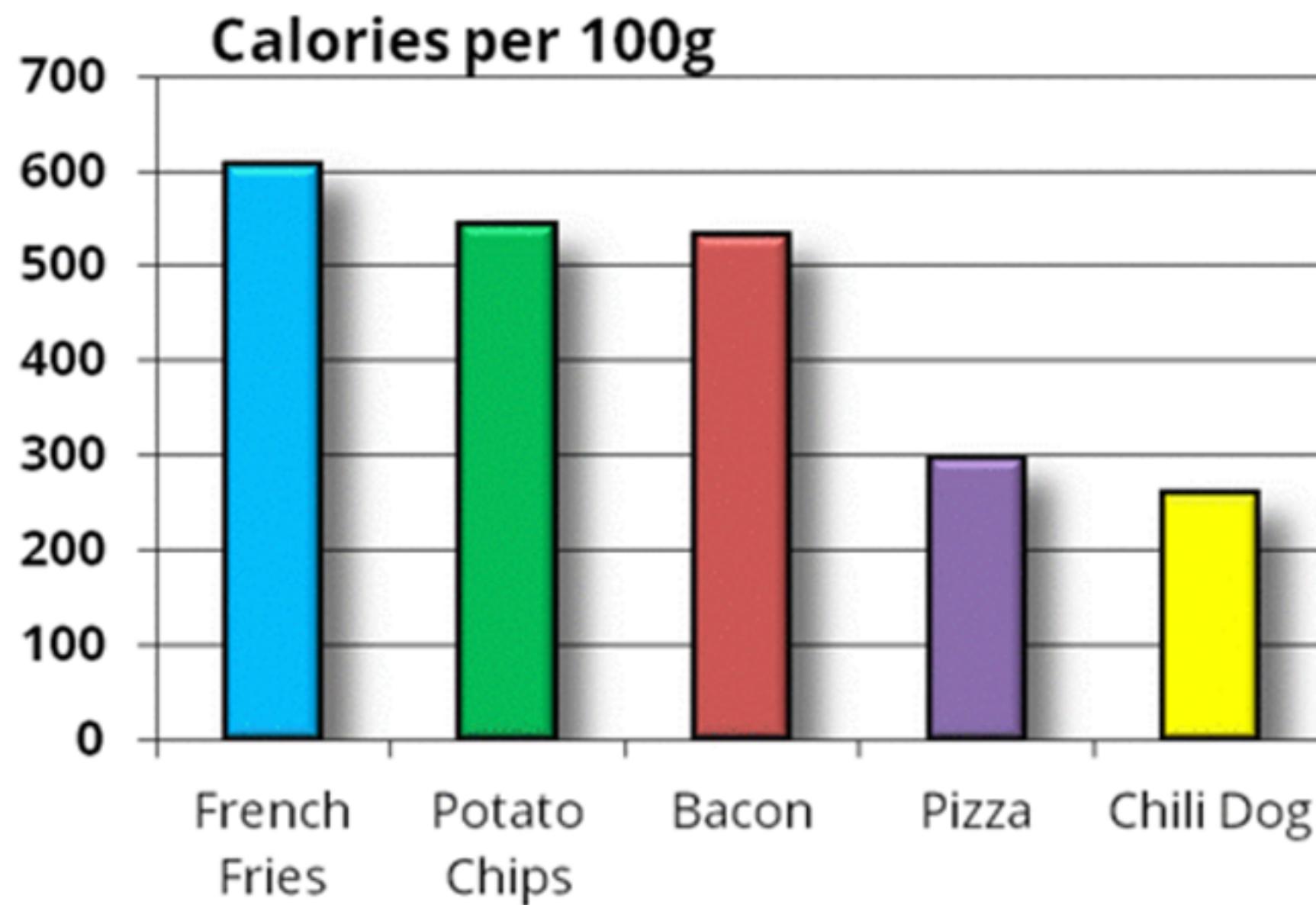


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Remove borders

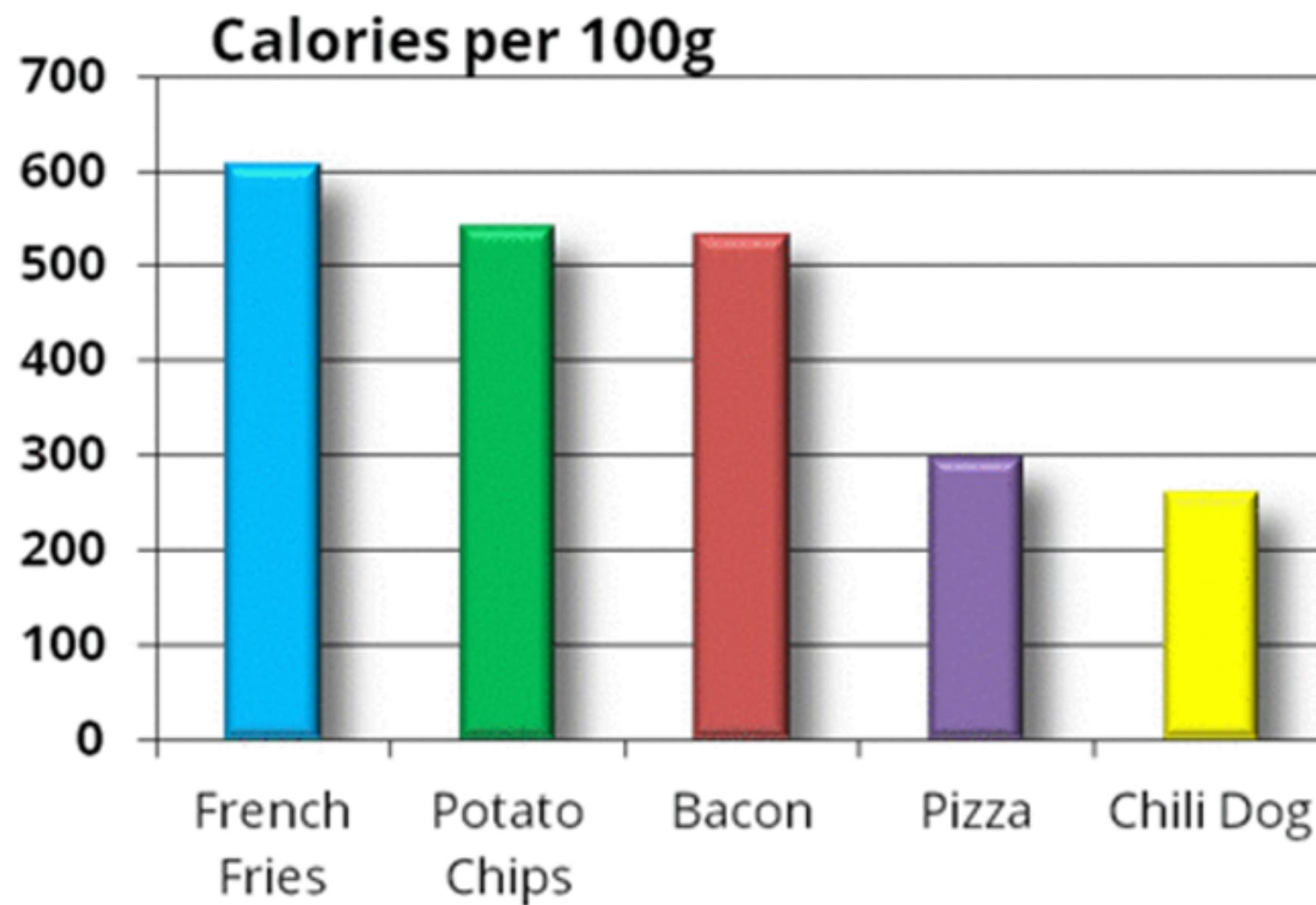


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Remove borders

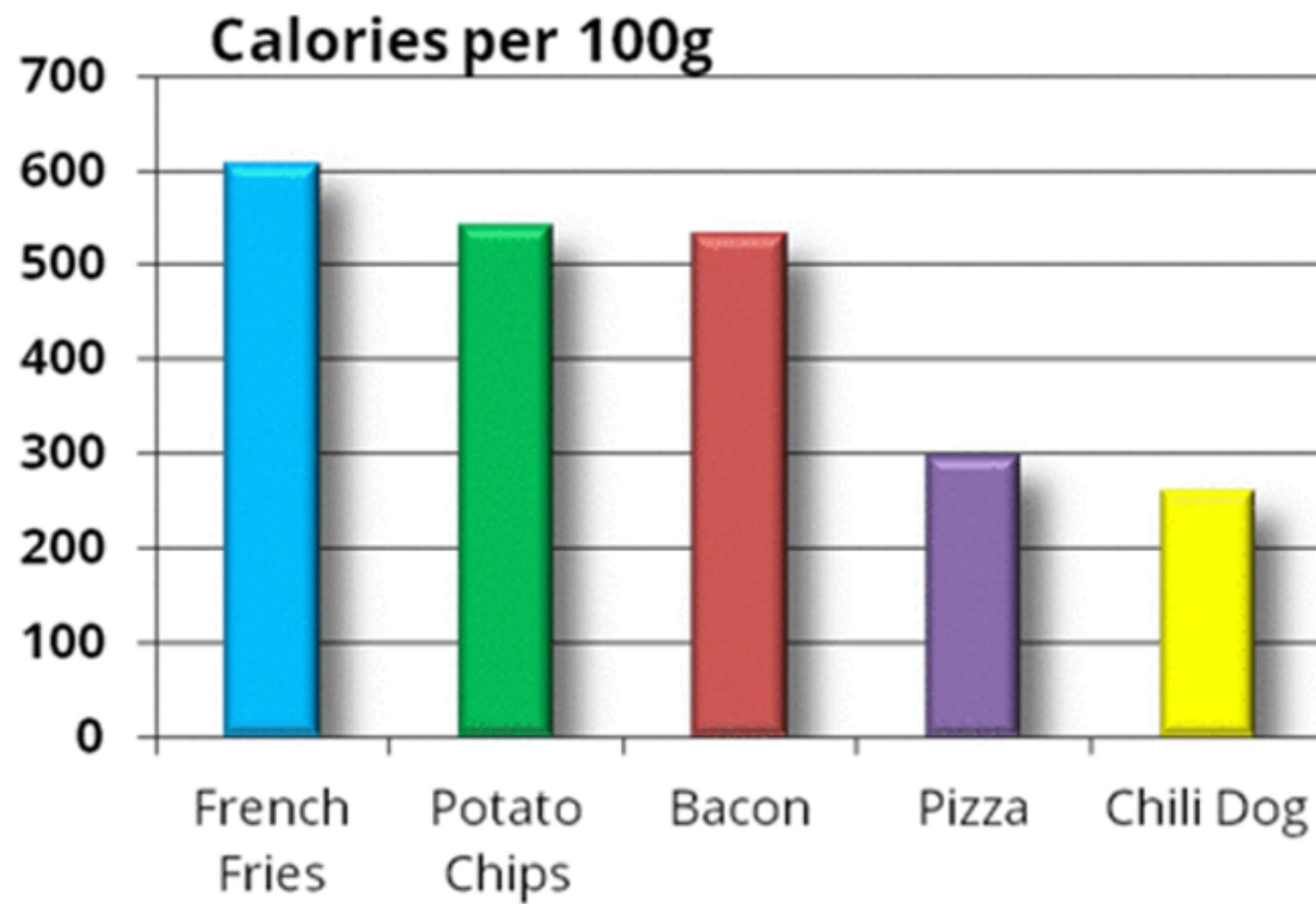


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Reduce colors

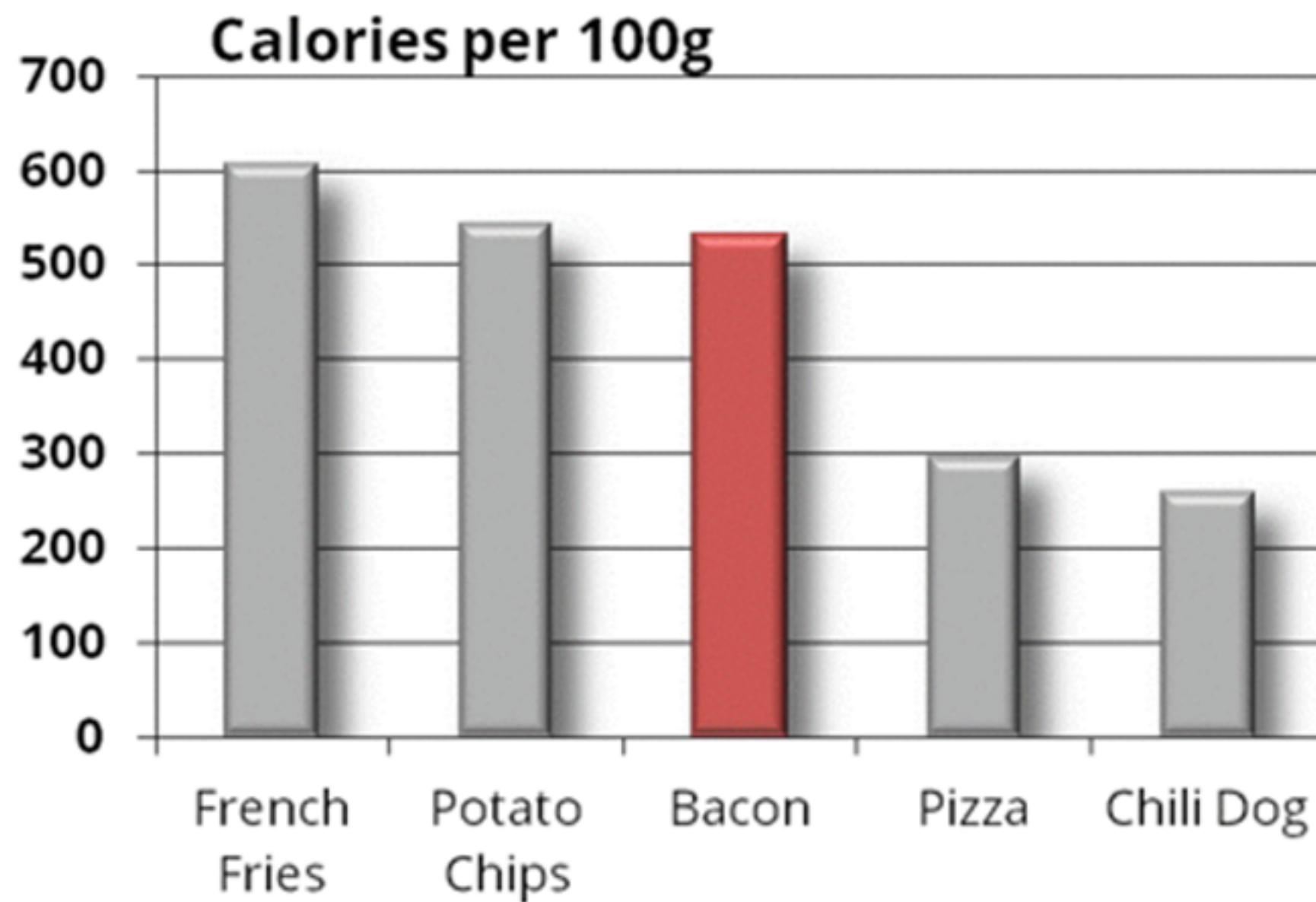


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Reduce colors

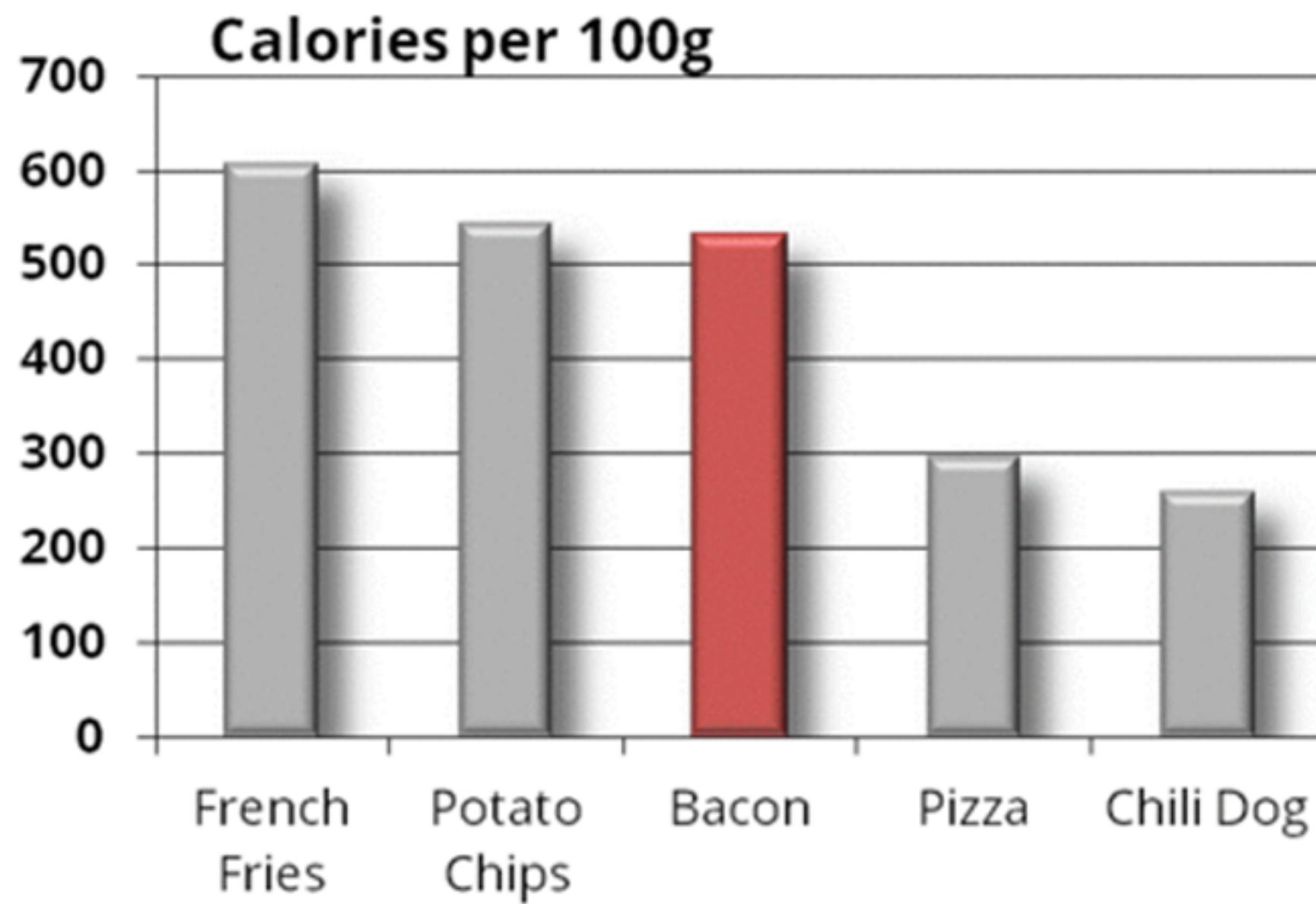


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Remove special effects

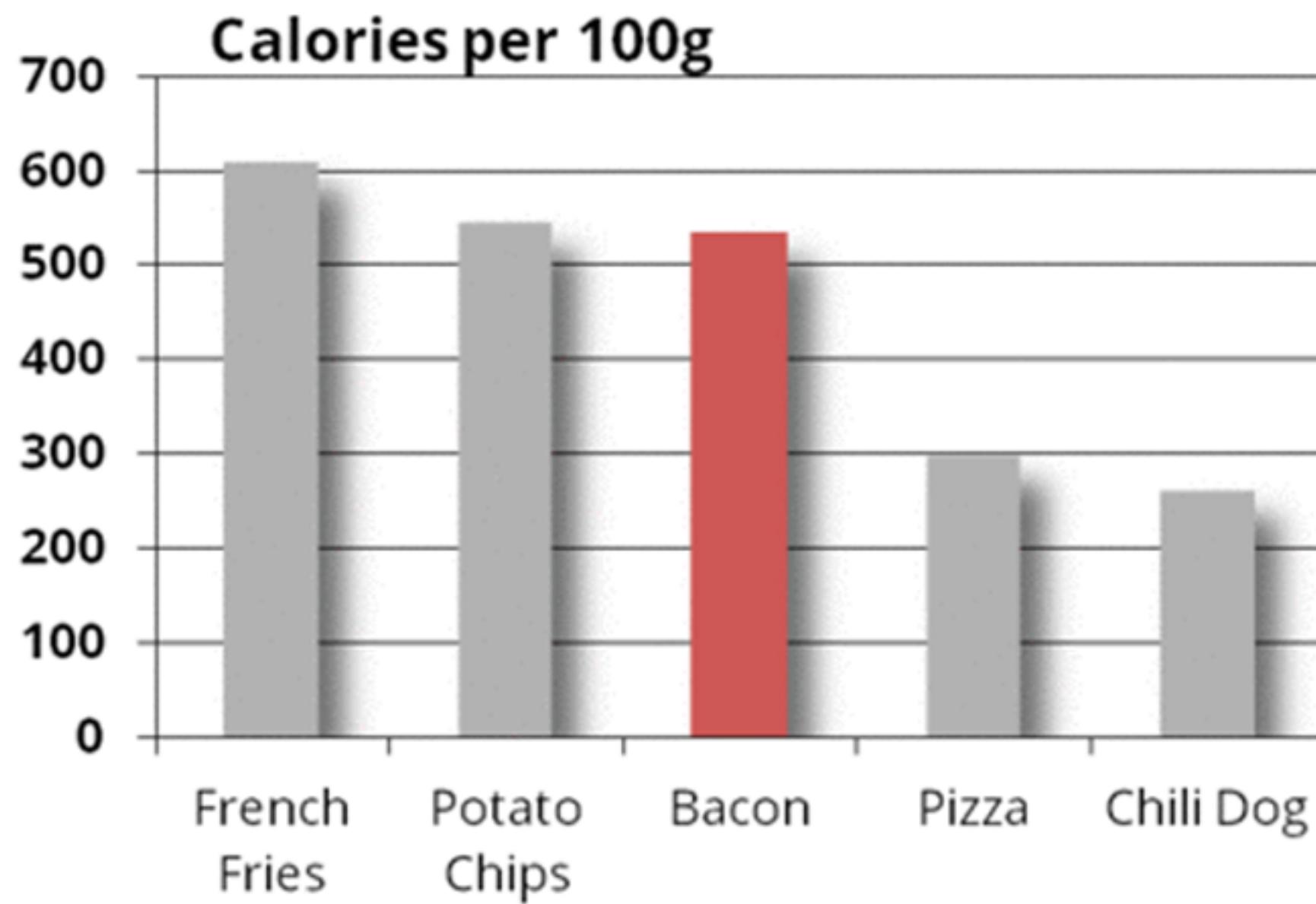


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Remove special effects

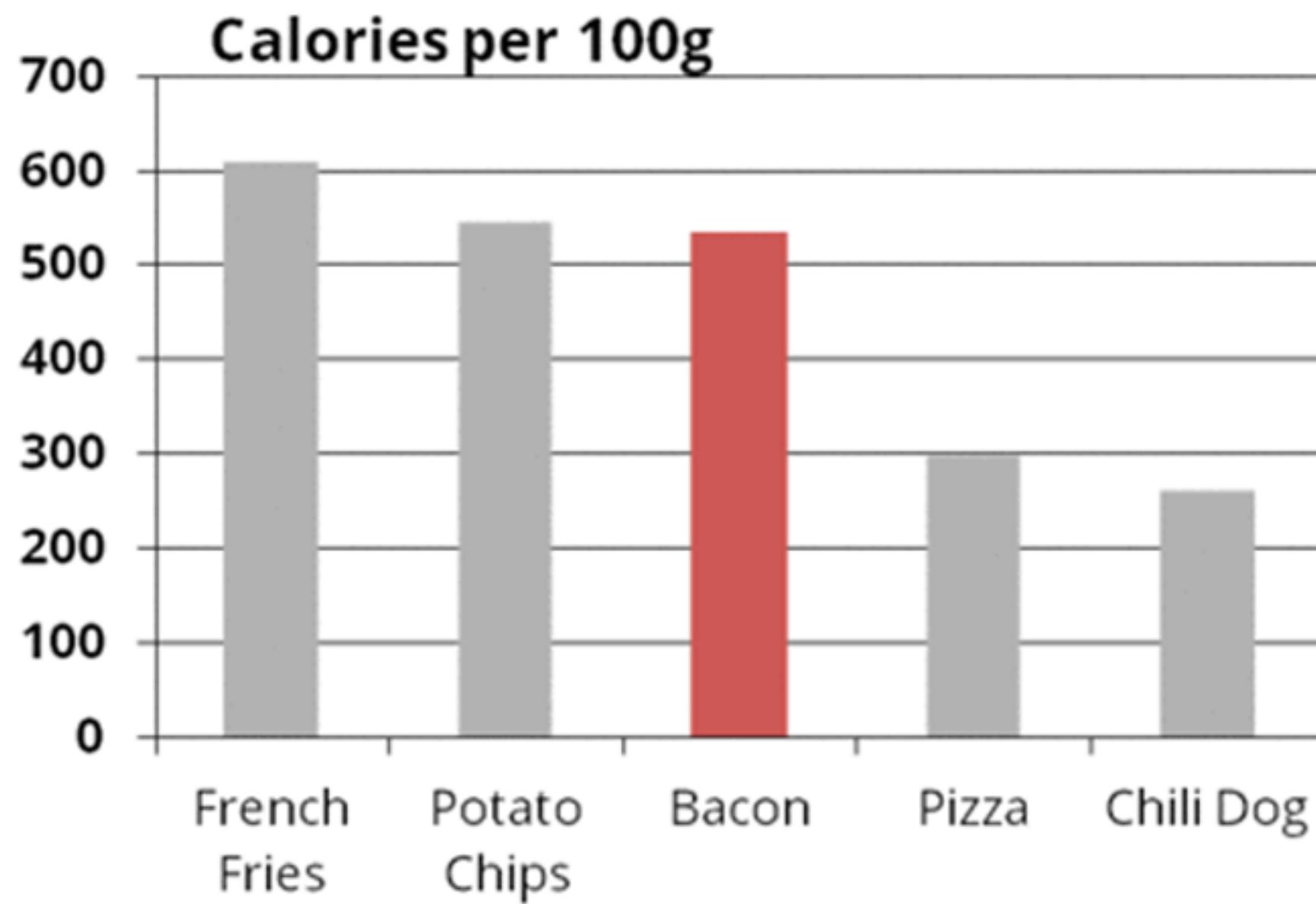


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Remove special effects

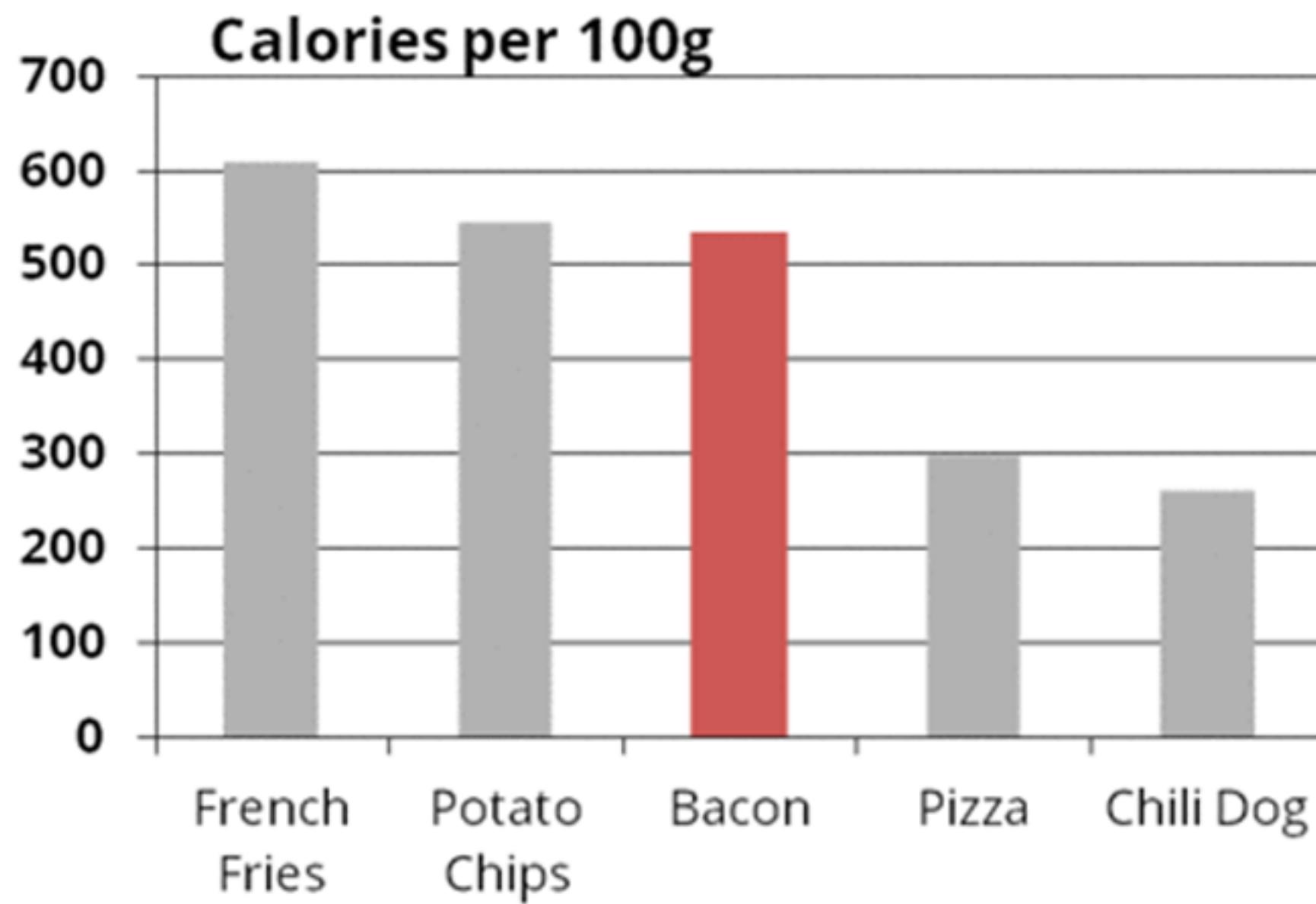


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Remove bolding

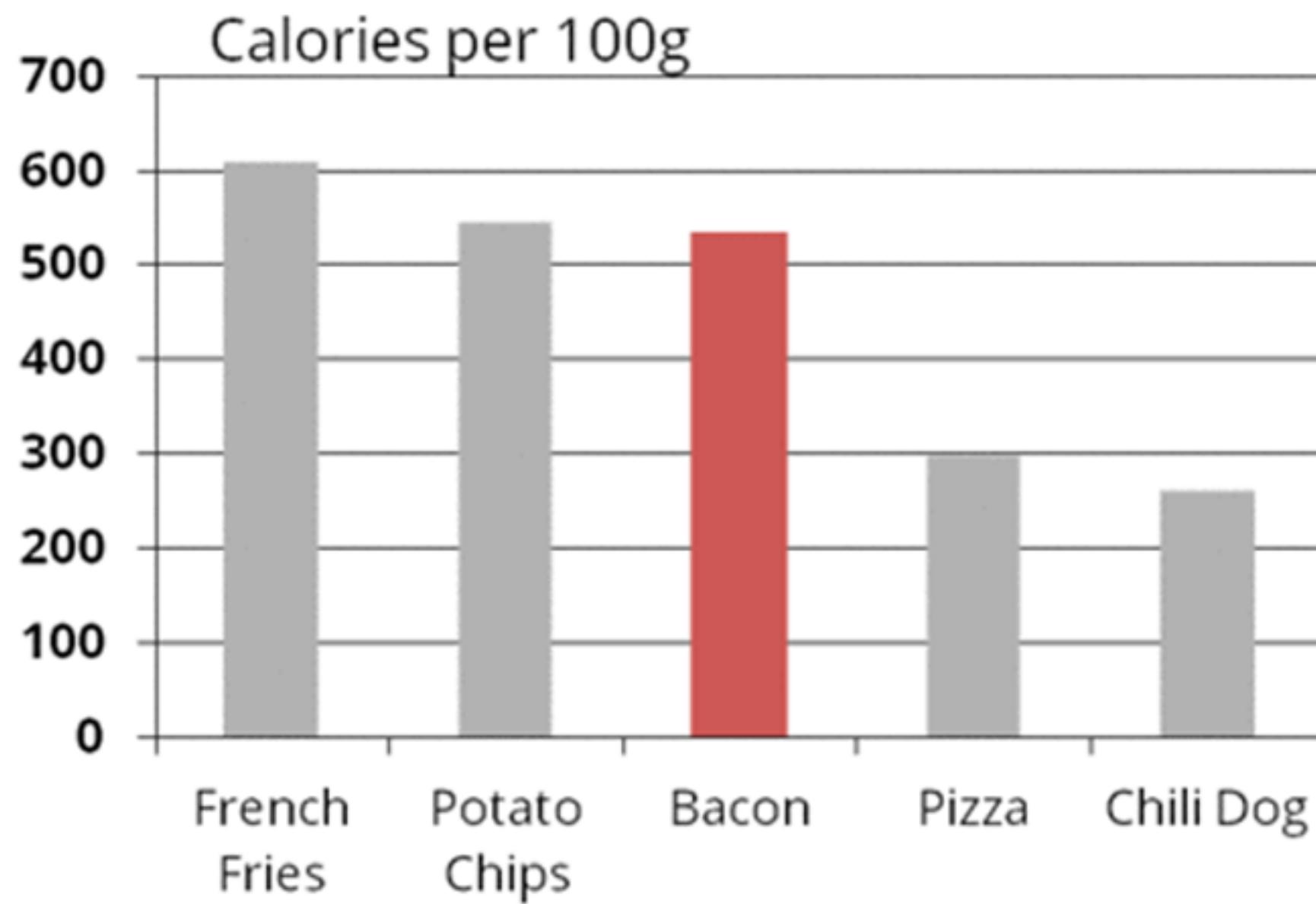


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Remove bolding

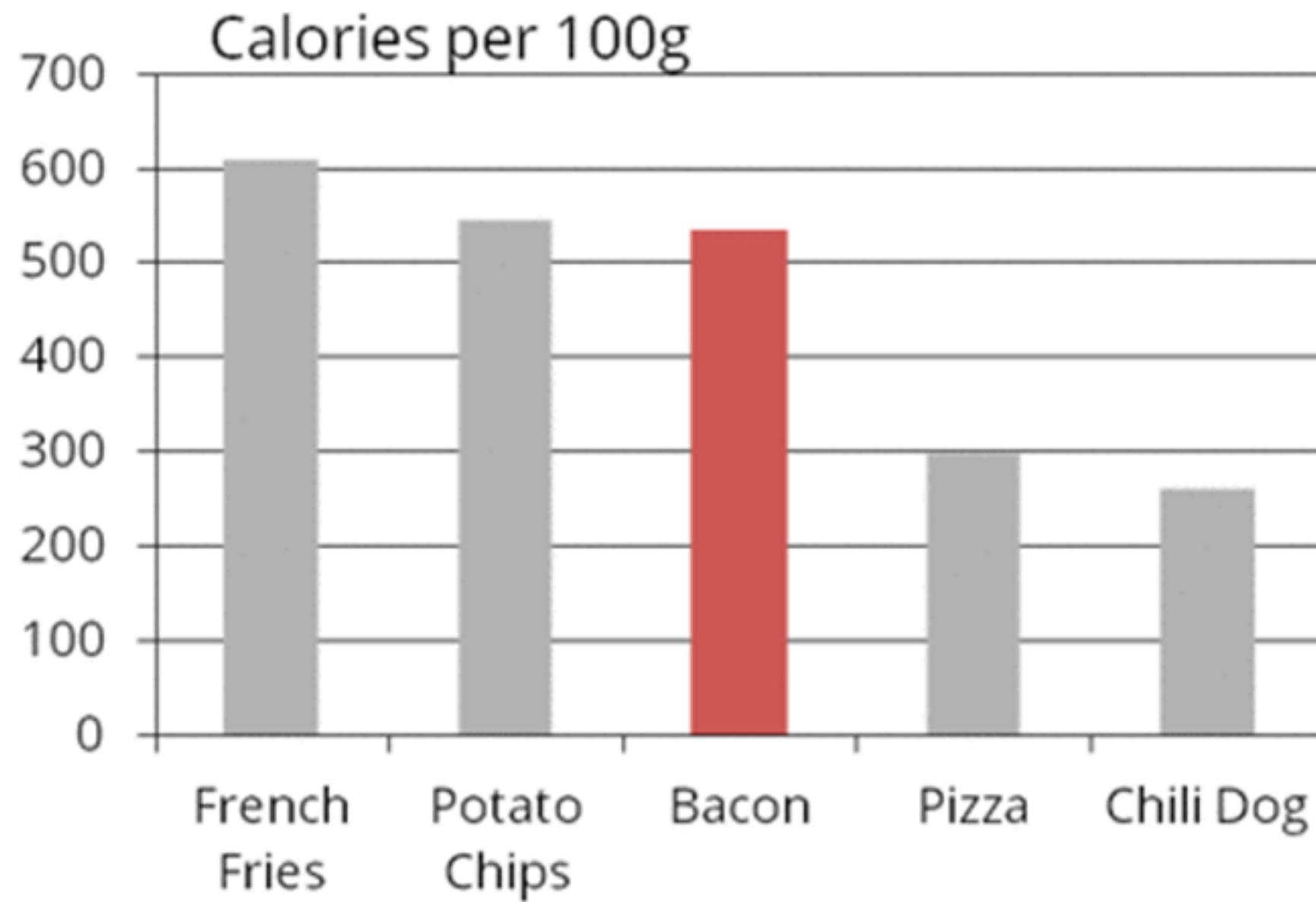


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Remove bolding

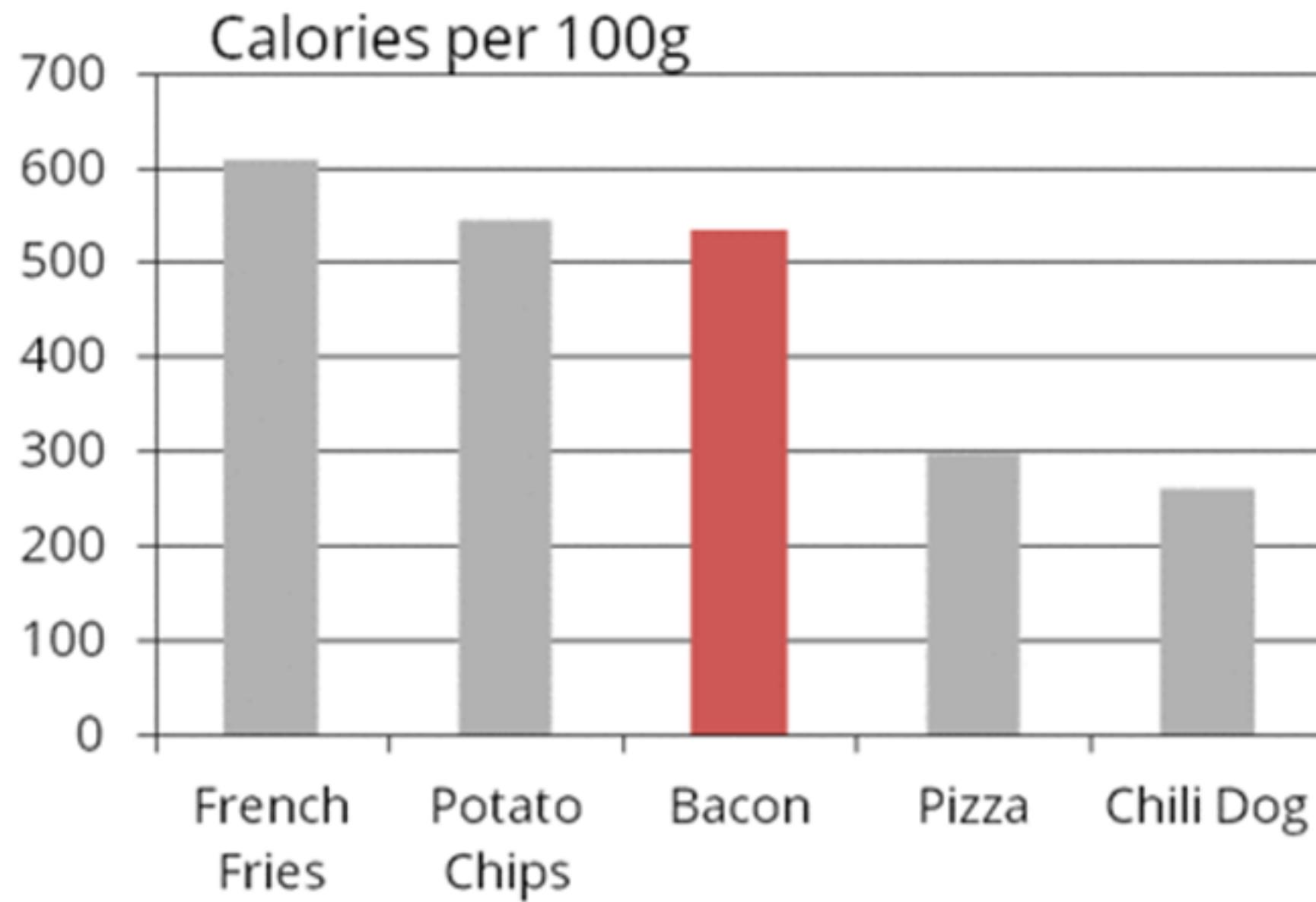


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Lighten labels

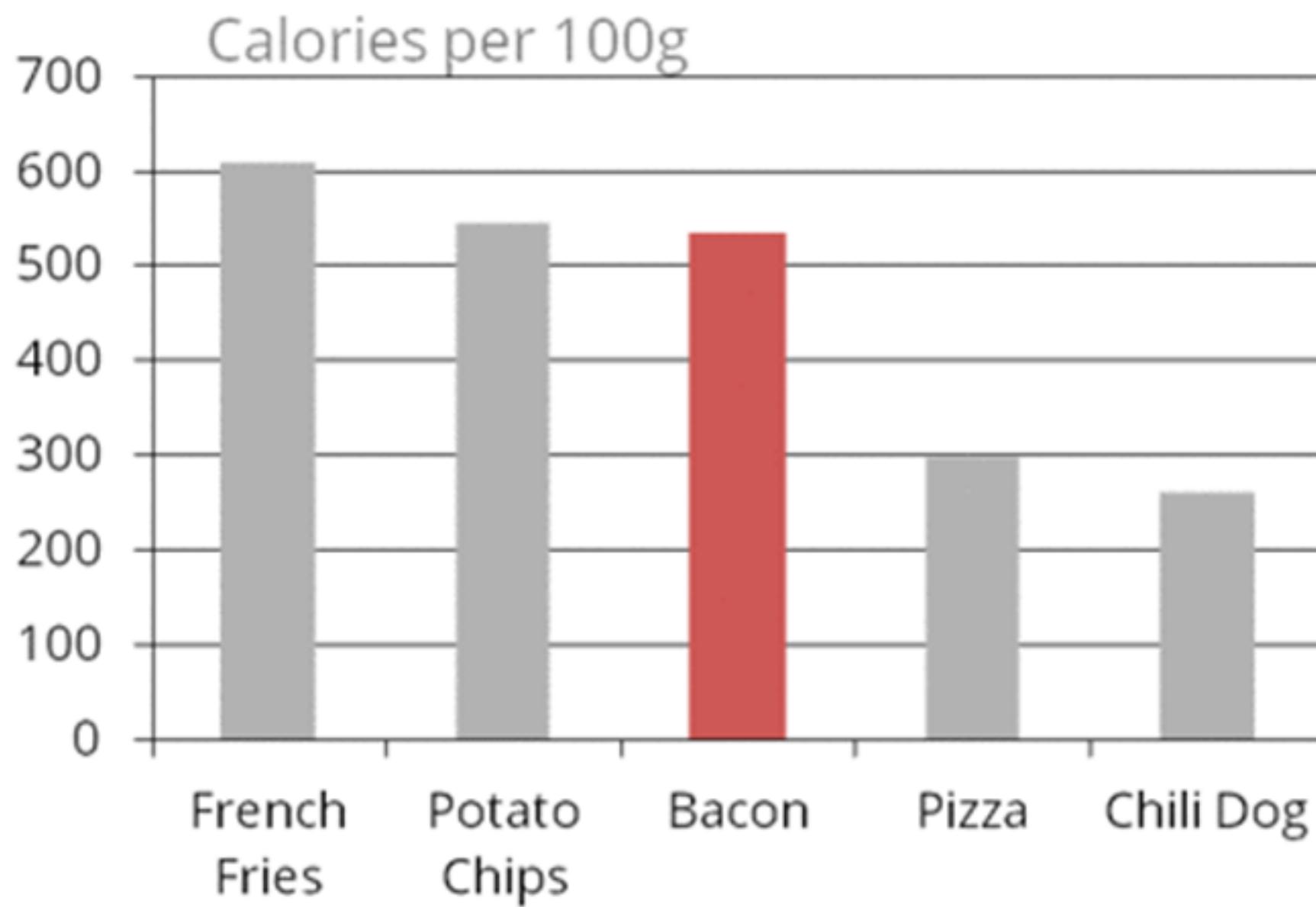


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Lighten labels

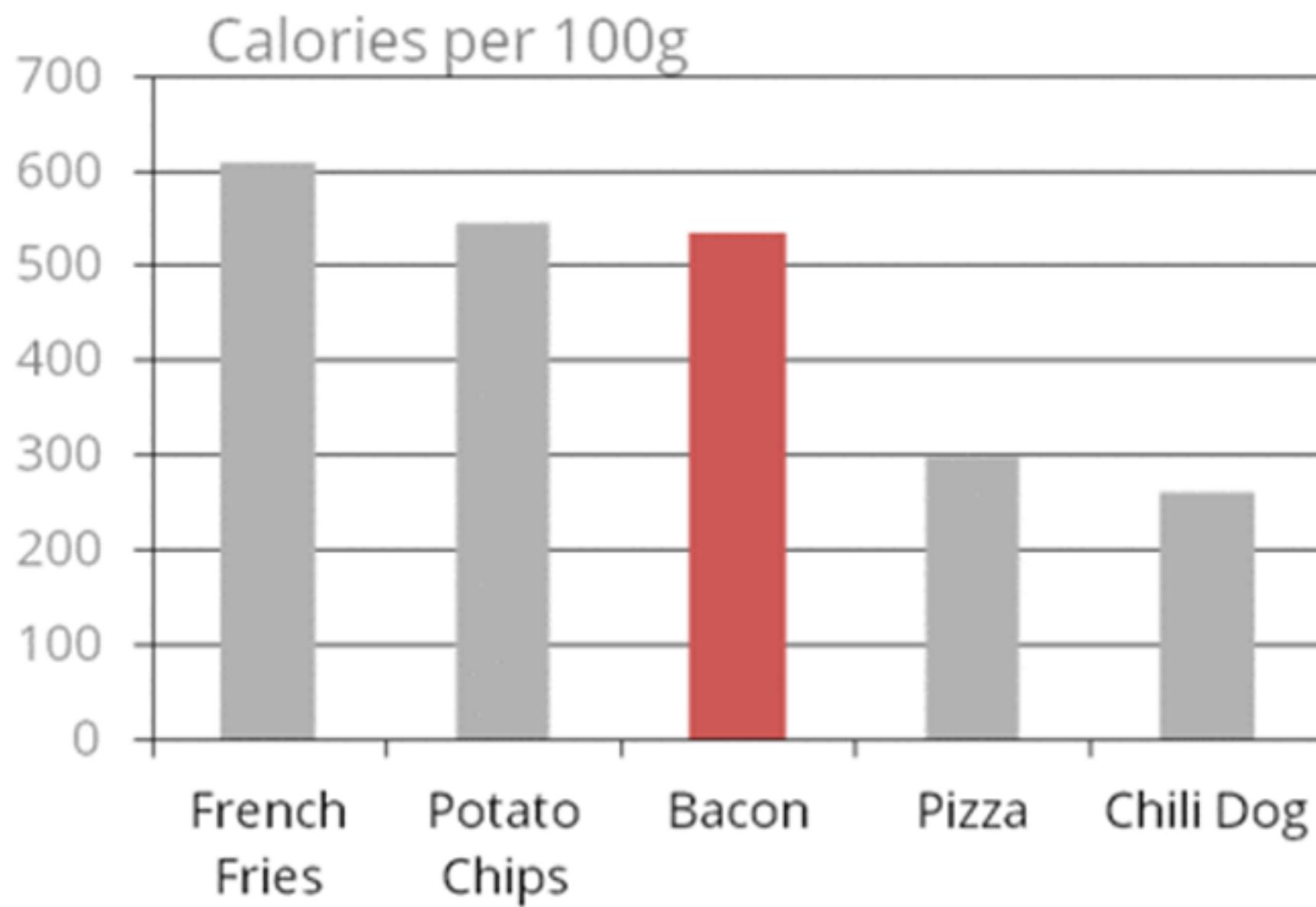


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Lighten labels

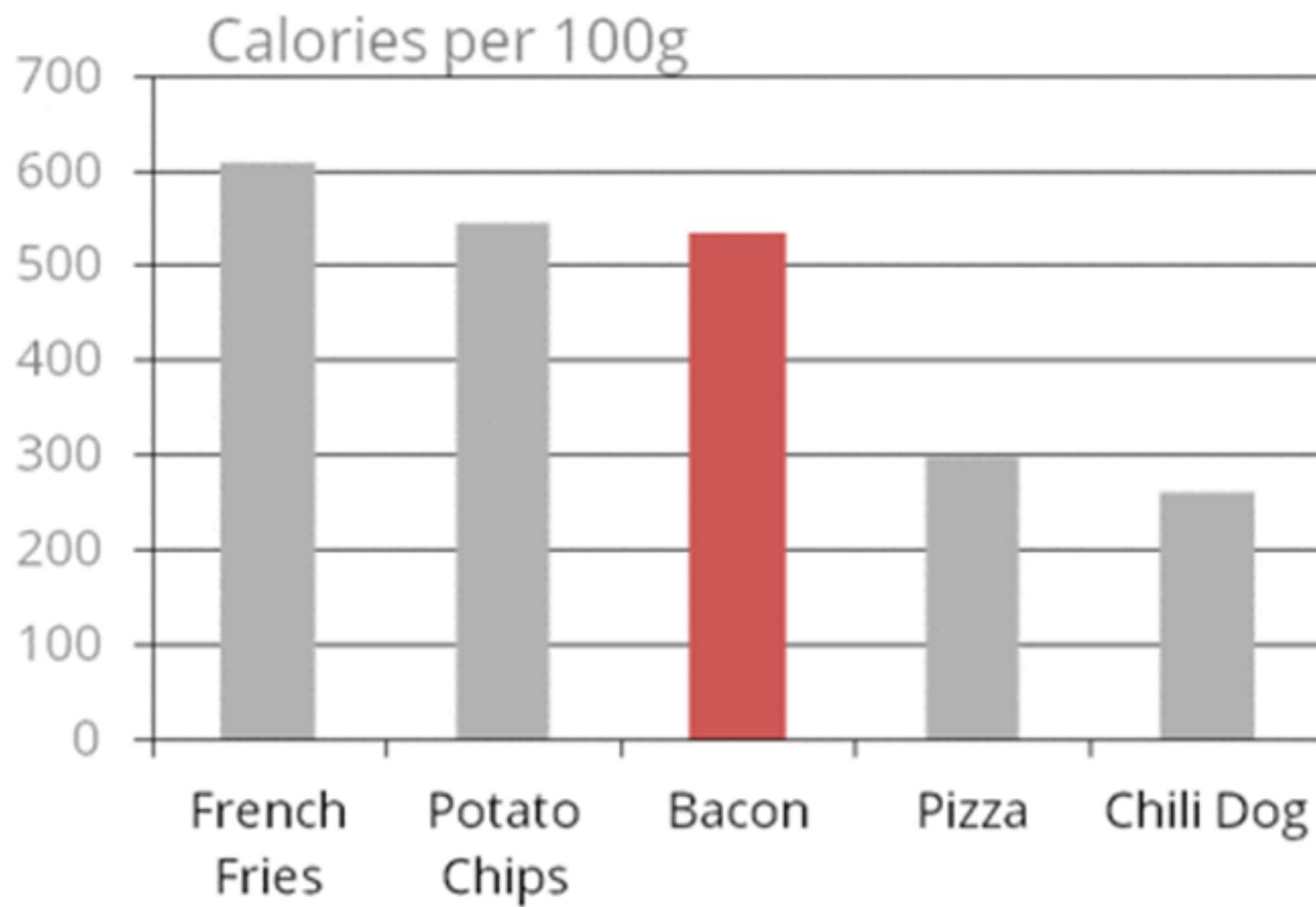


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Lighten lines

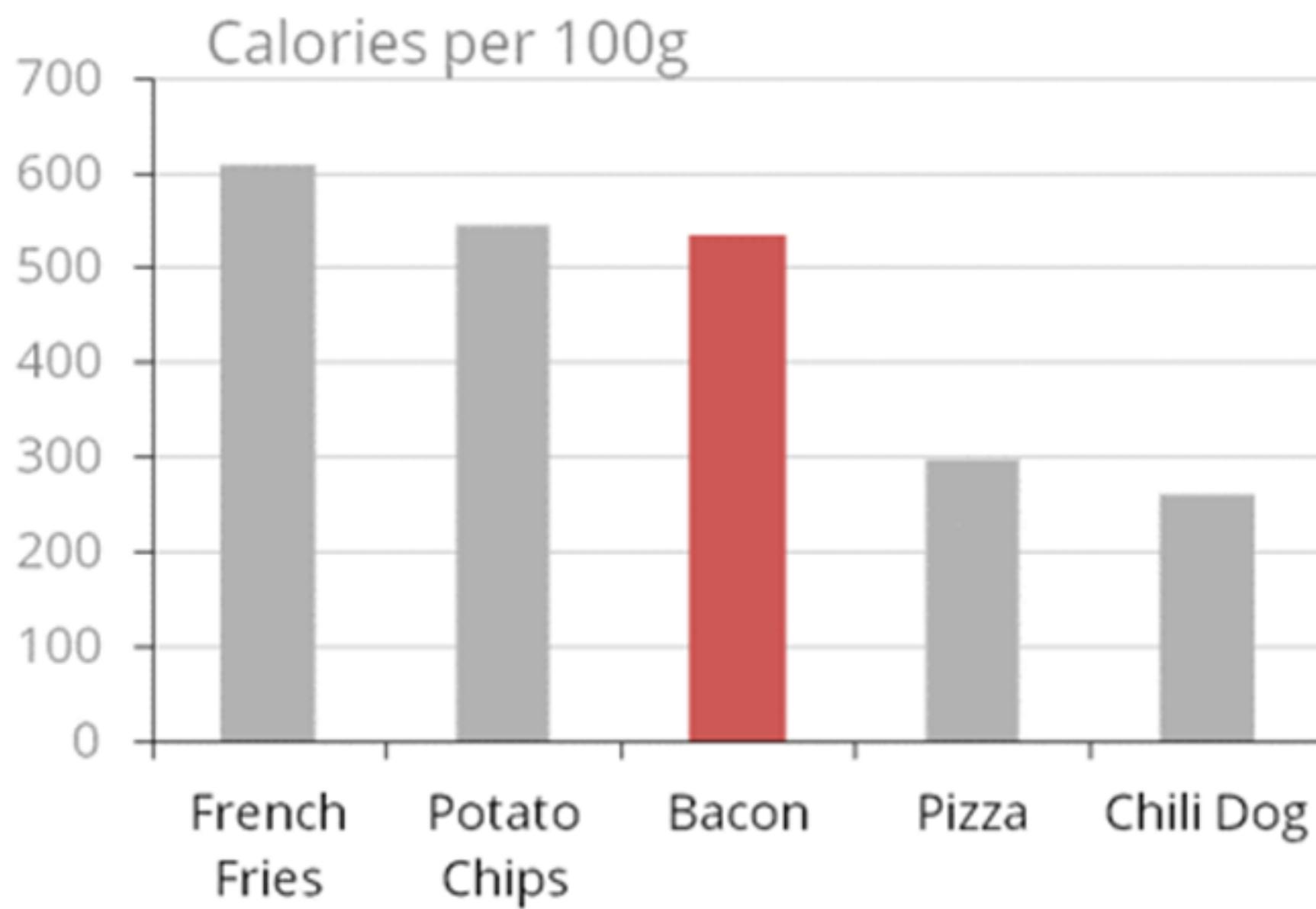


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Lighten lines

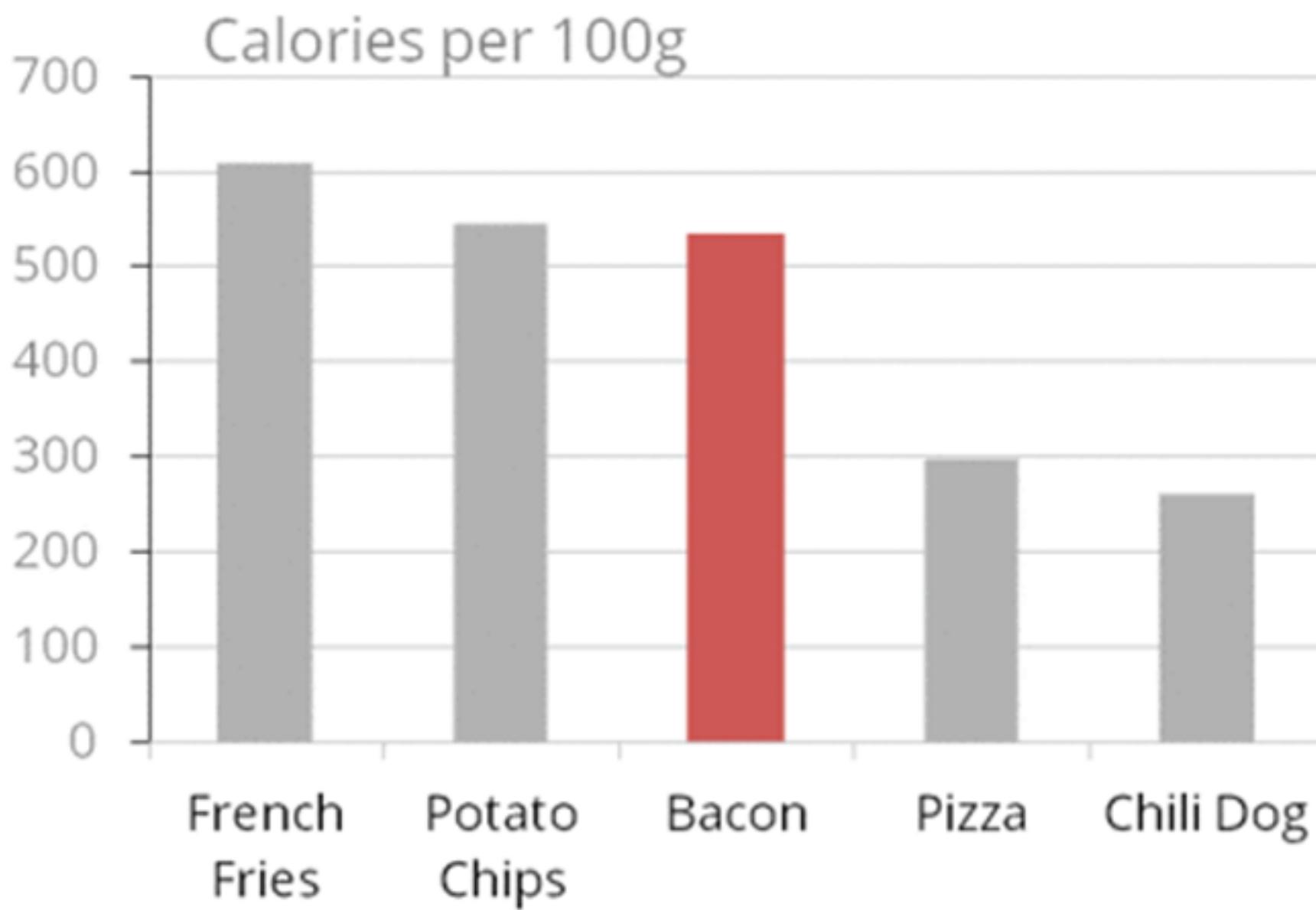


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Lighten lines

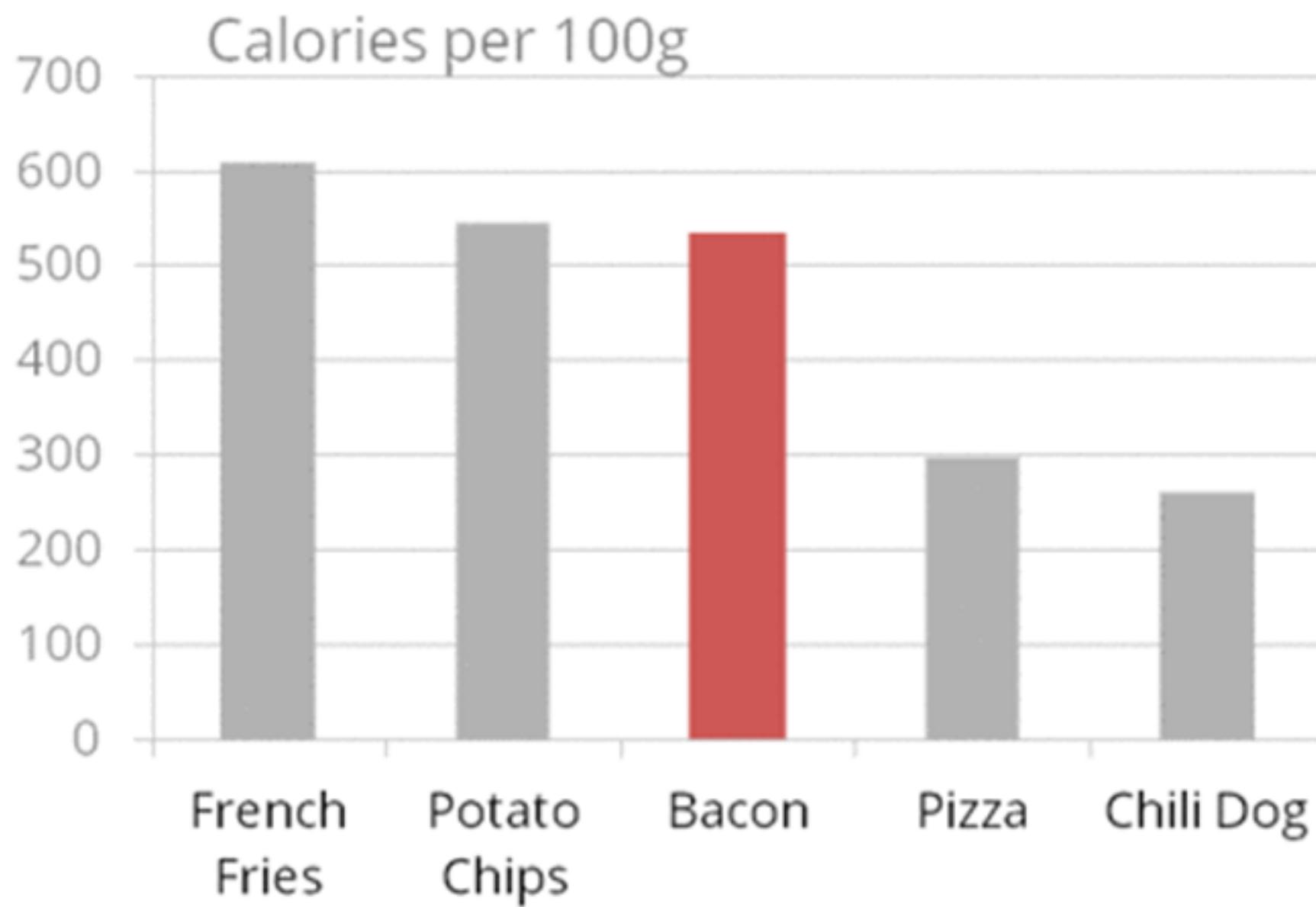


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Lighten lines

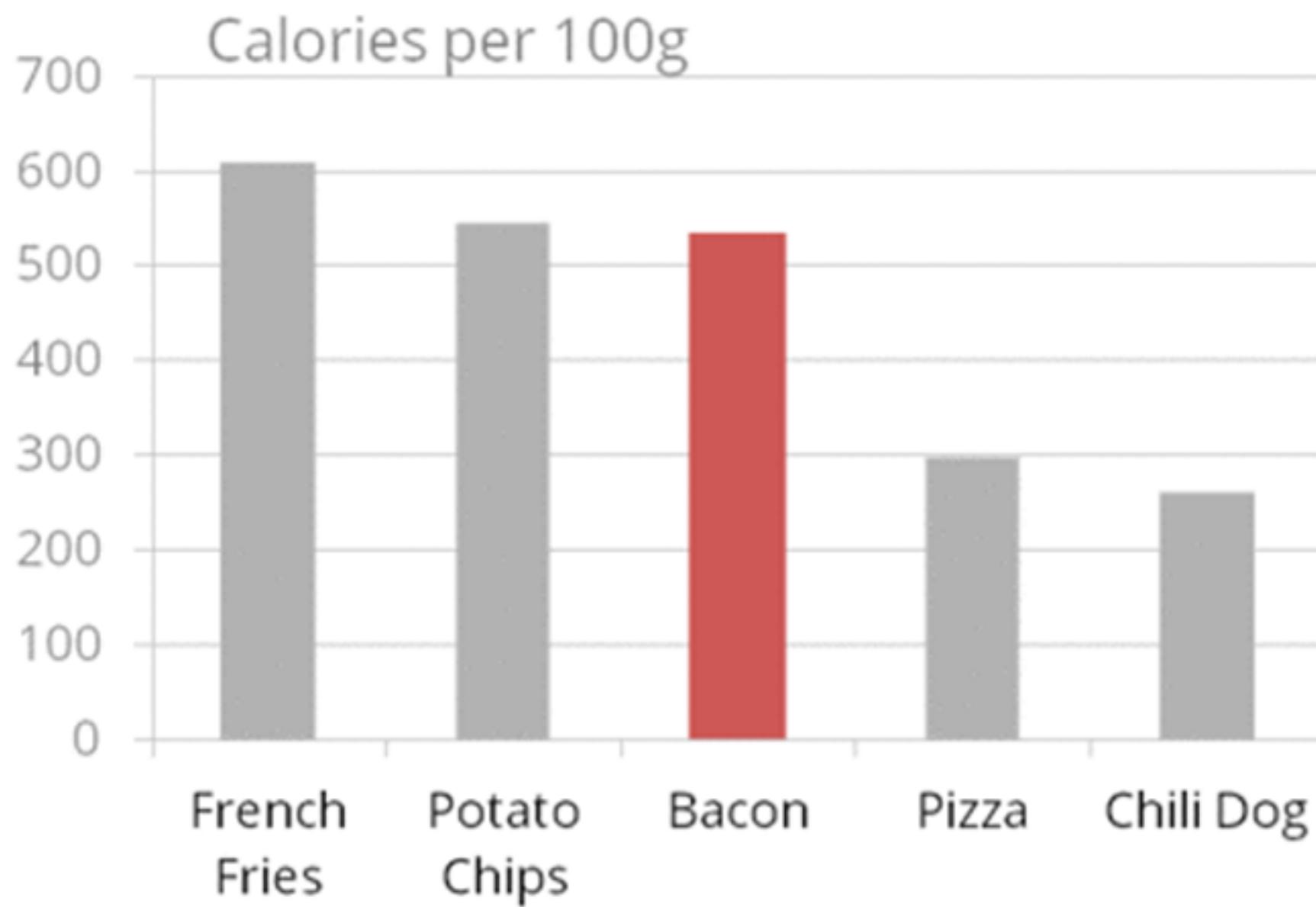


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Or remove lines

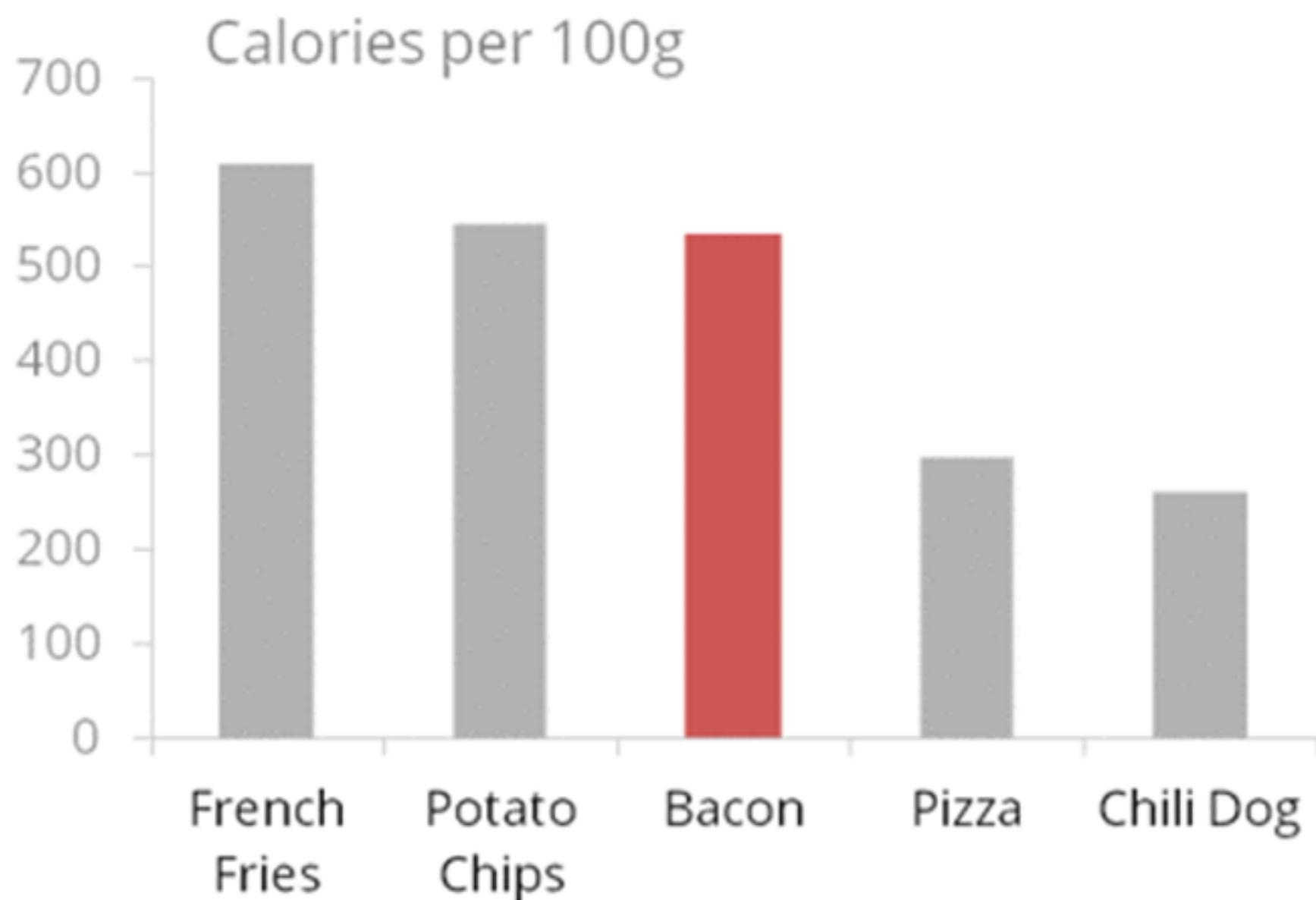


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Or remove lines

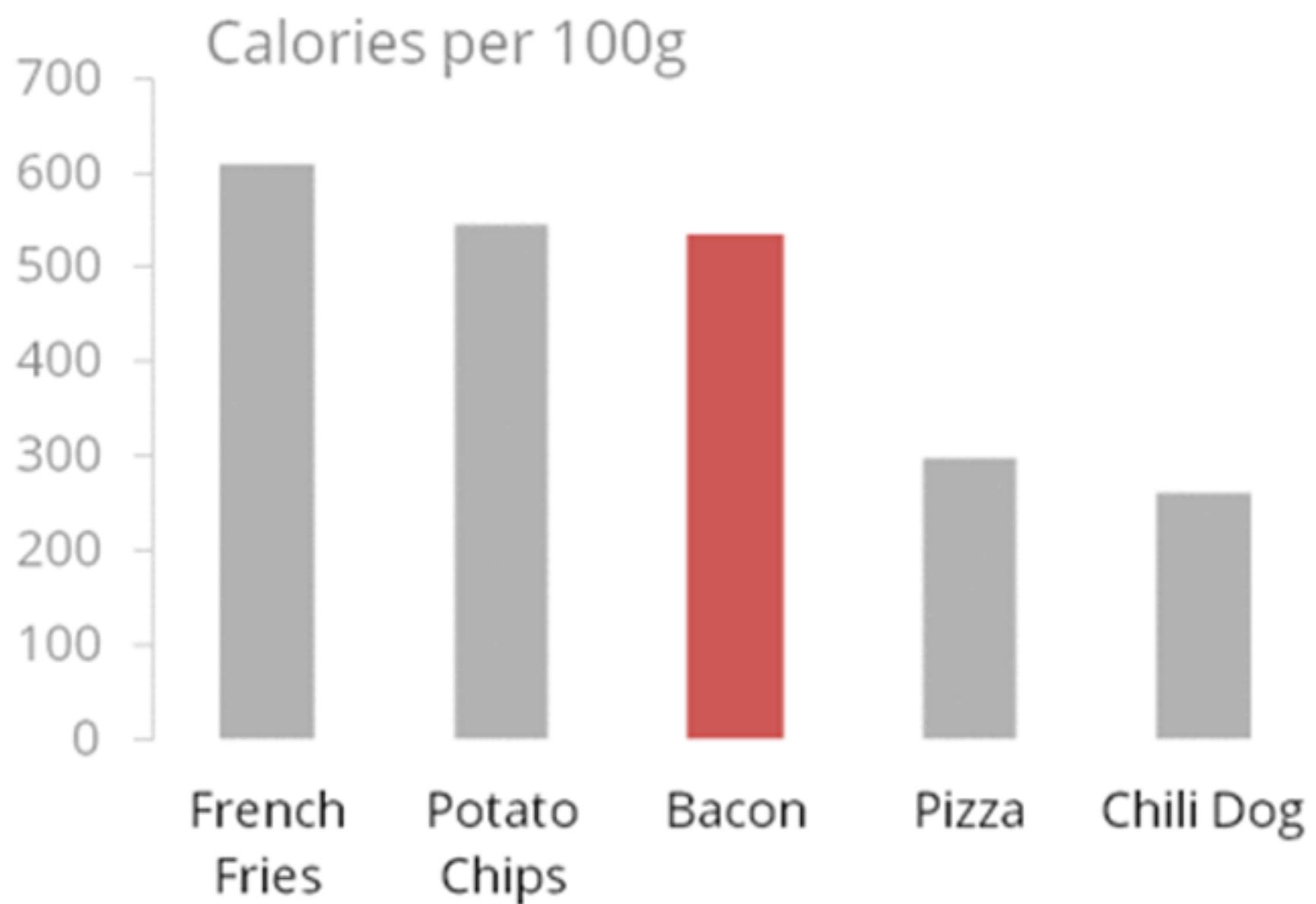


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Or remove lines

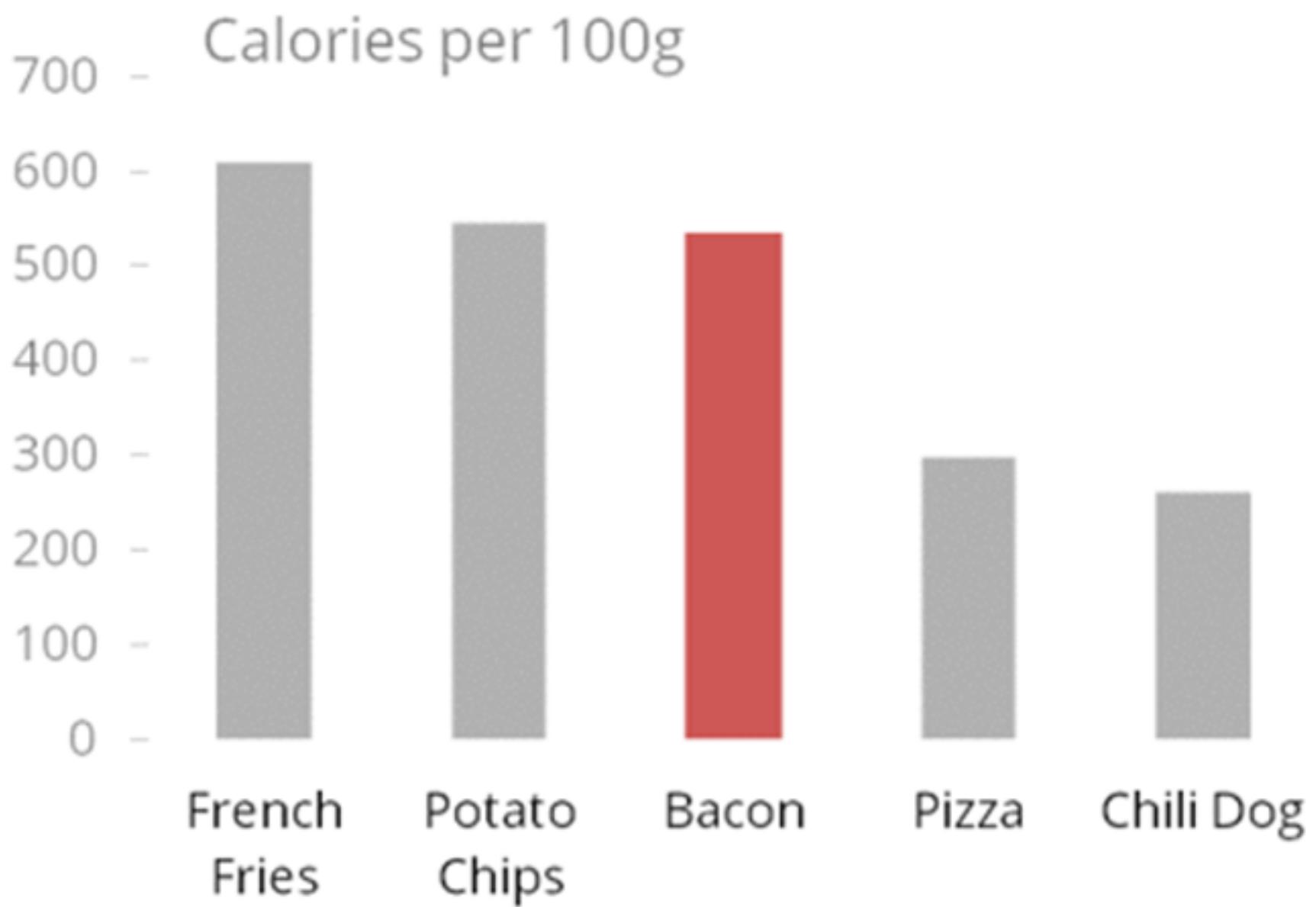


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Or remove lines

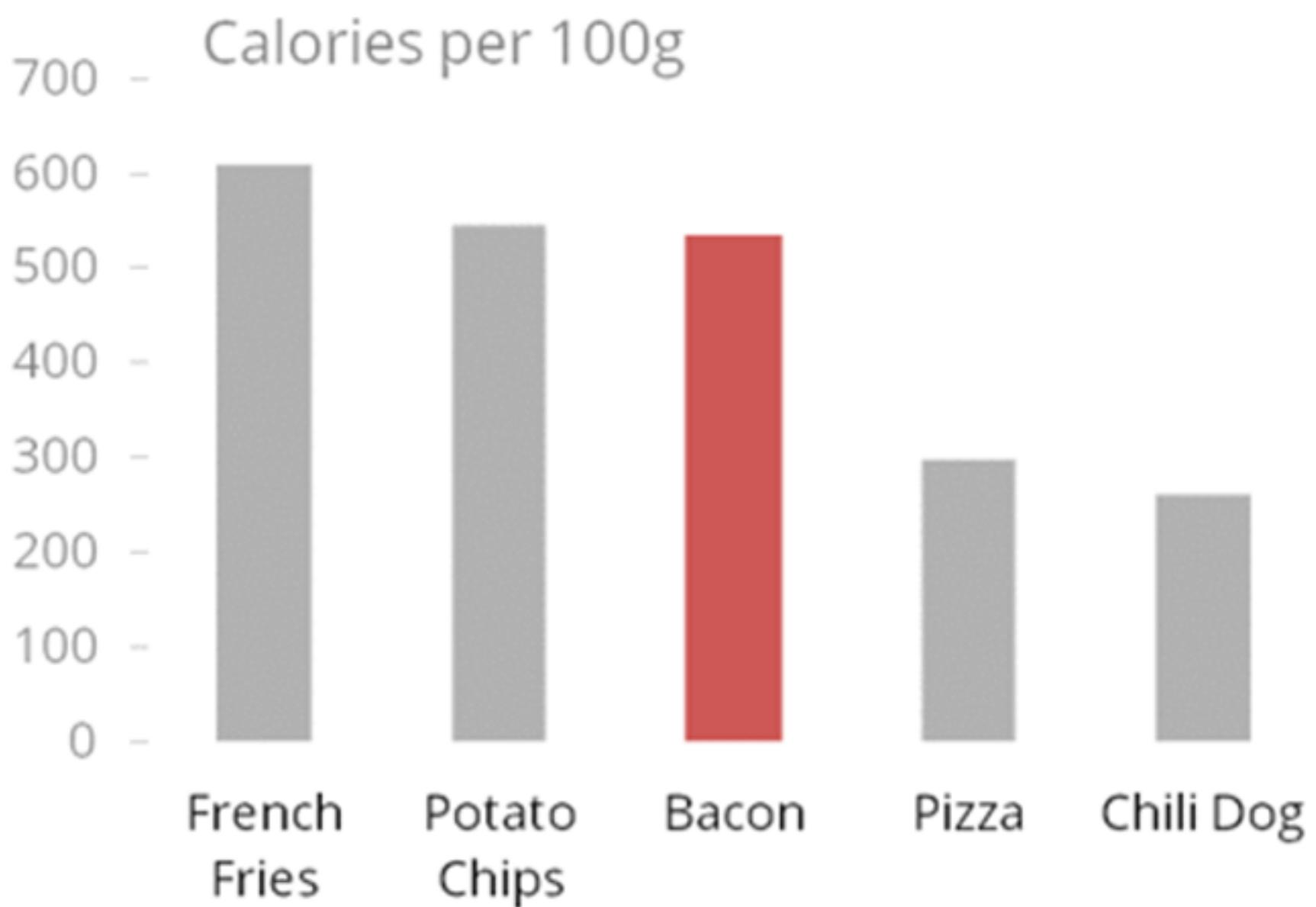


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Direct label



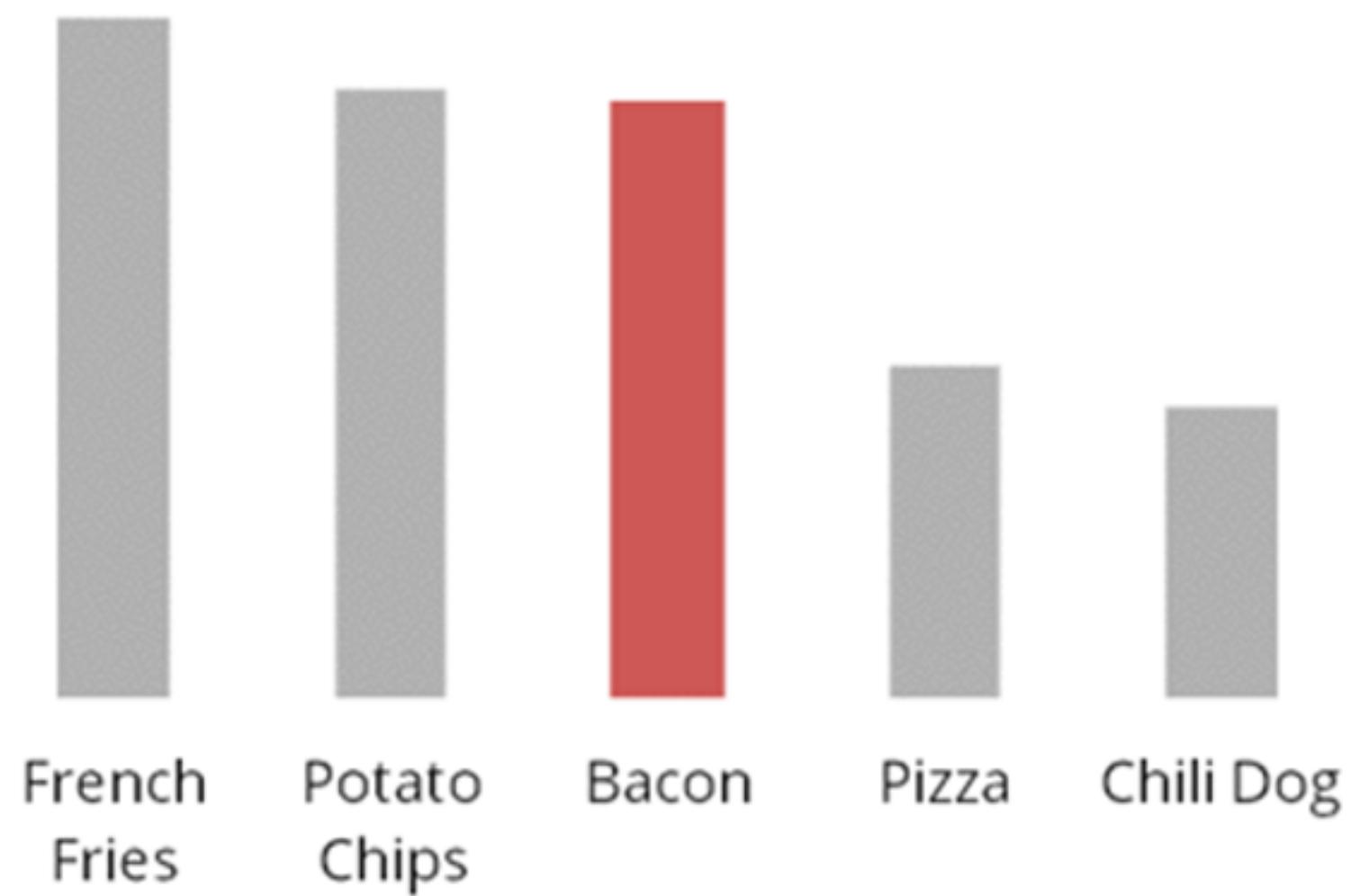
Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Direct label

Calories per 100g



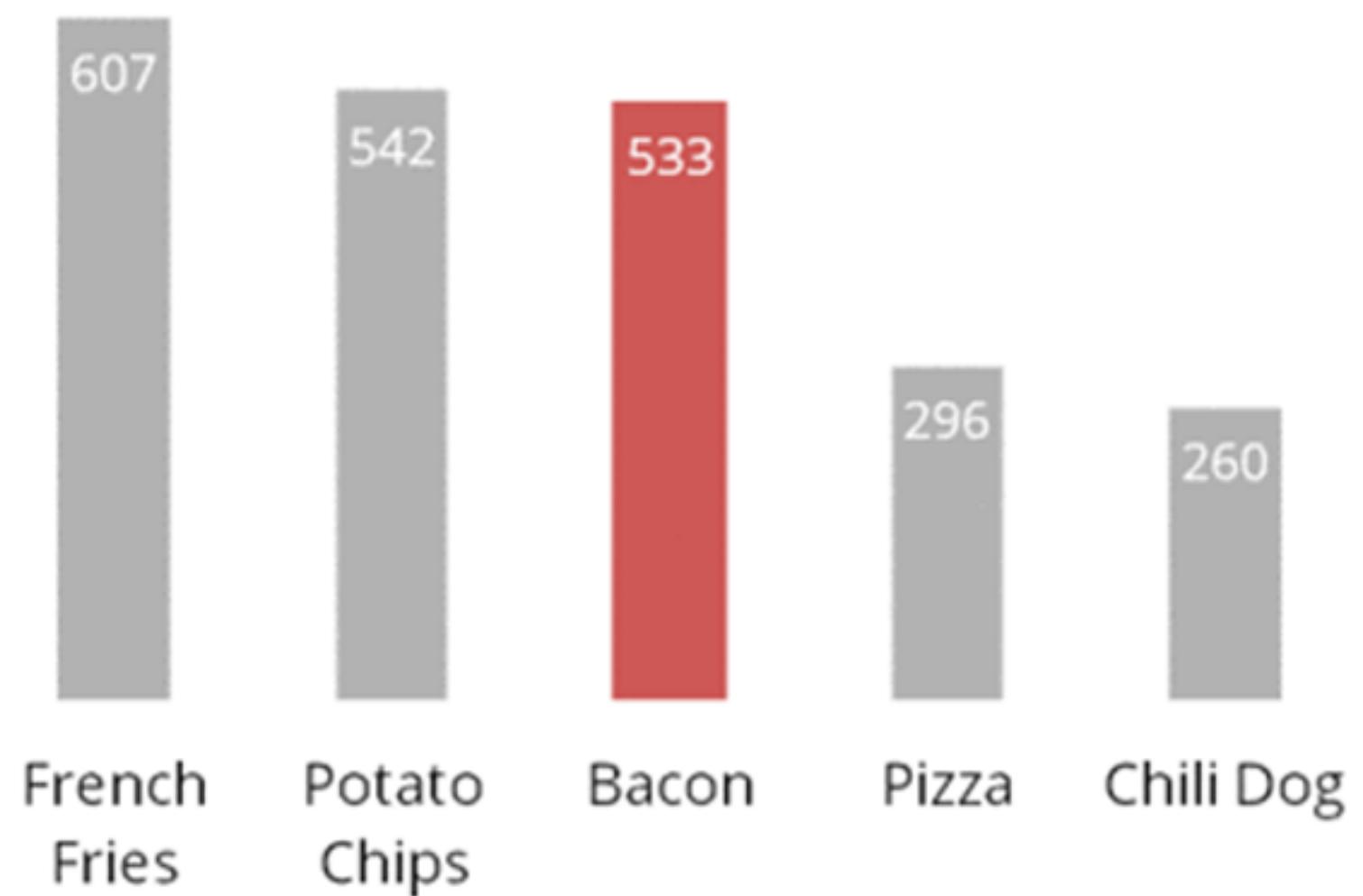
Created by Darkhorse Analytics

www.darkhorseanalytics.com

Data-to-ink Ratio

Direct label

Calories per 100g



Created by Darkhorse Analytics

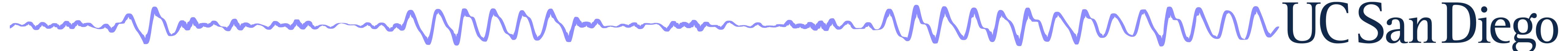
www.darkhorseanalytics.com

Data-to-ink Ratio

Less
is **more**
(effective)

Created by Darkhorse Analytics

www.darkhorseanalytics.com

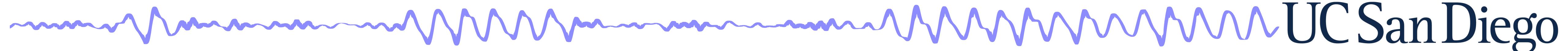


Data-to-ink Ratio

Less
is **more**
(attractive)

Created by Darkhorse Analytics

www.darkhorseanalytics.com

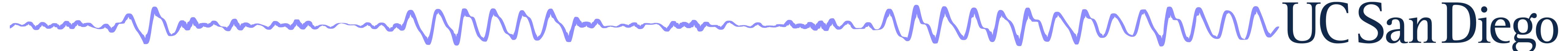


Data-to-ink Ratio

Less
is **more**
(impactive)

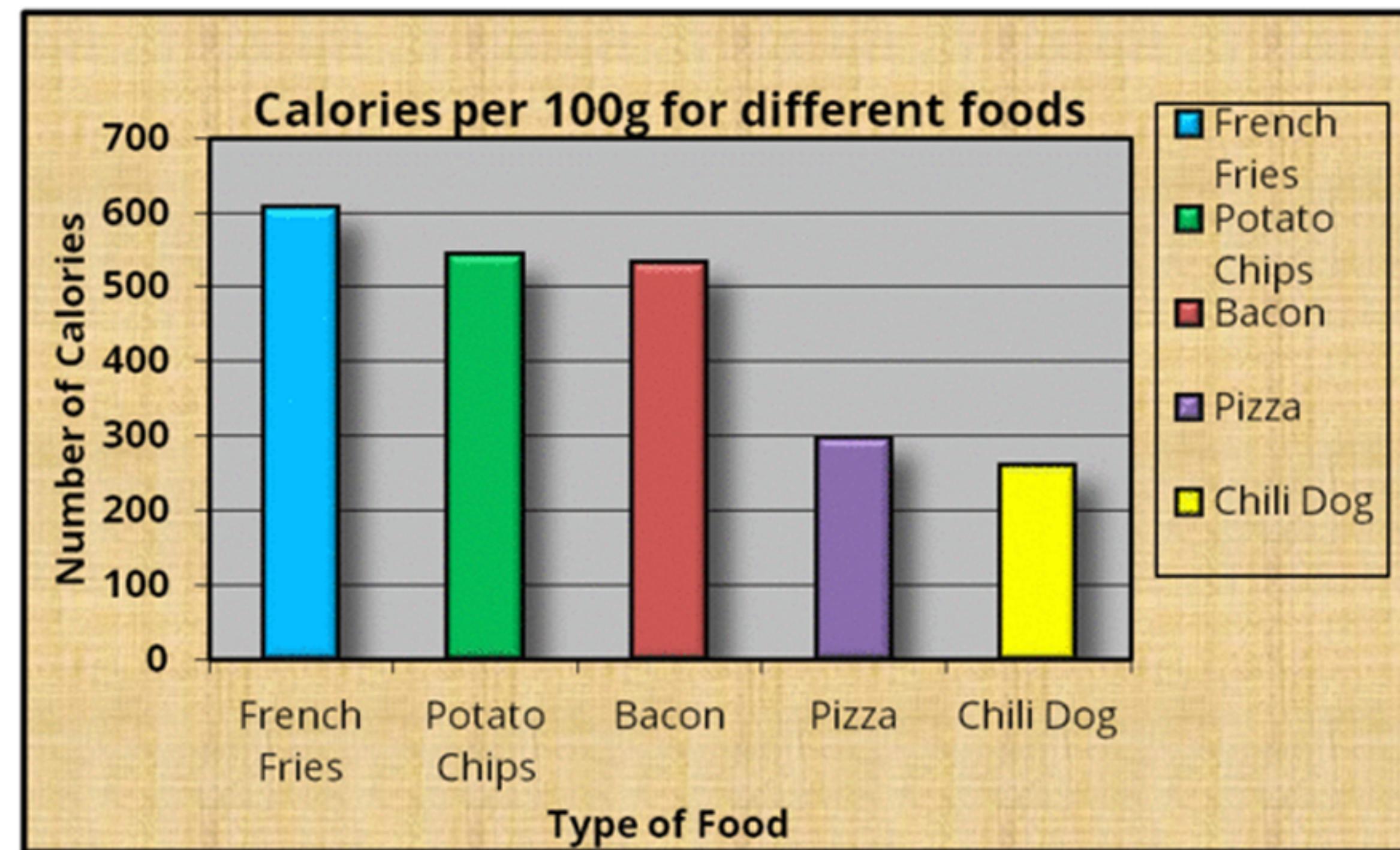
Created by Darkhorse Analytics

www.darkhorseanalytics.com



Data-to-ink Ratio

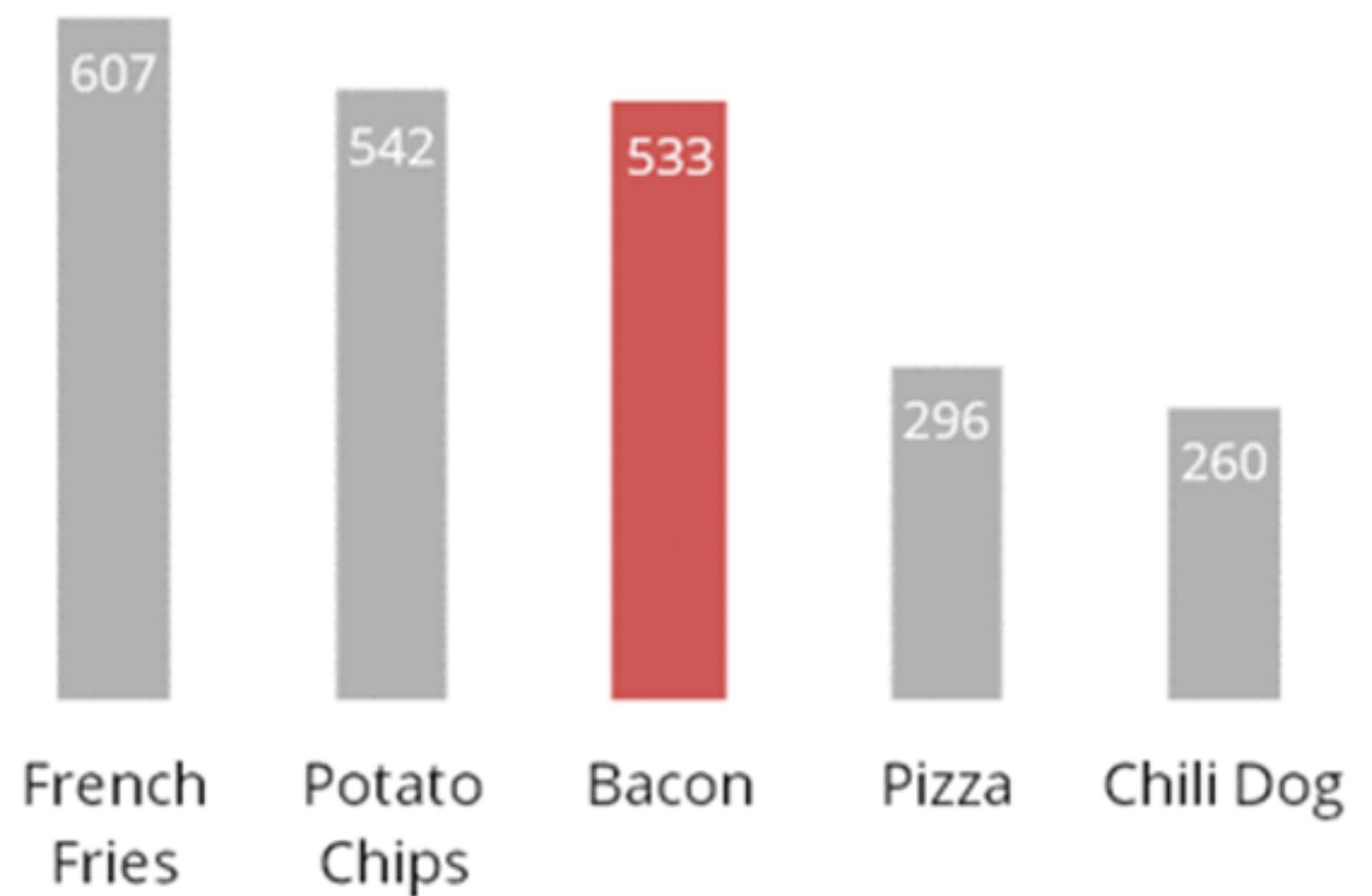
Before



Data-to-ink Ratio

After

Calories per 100g

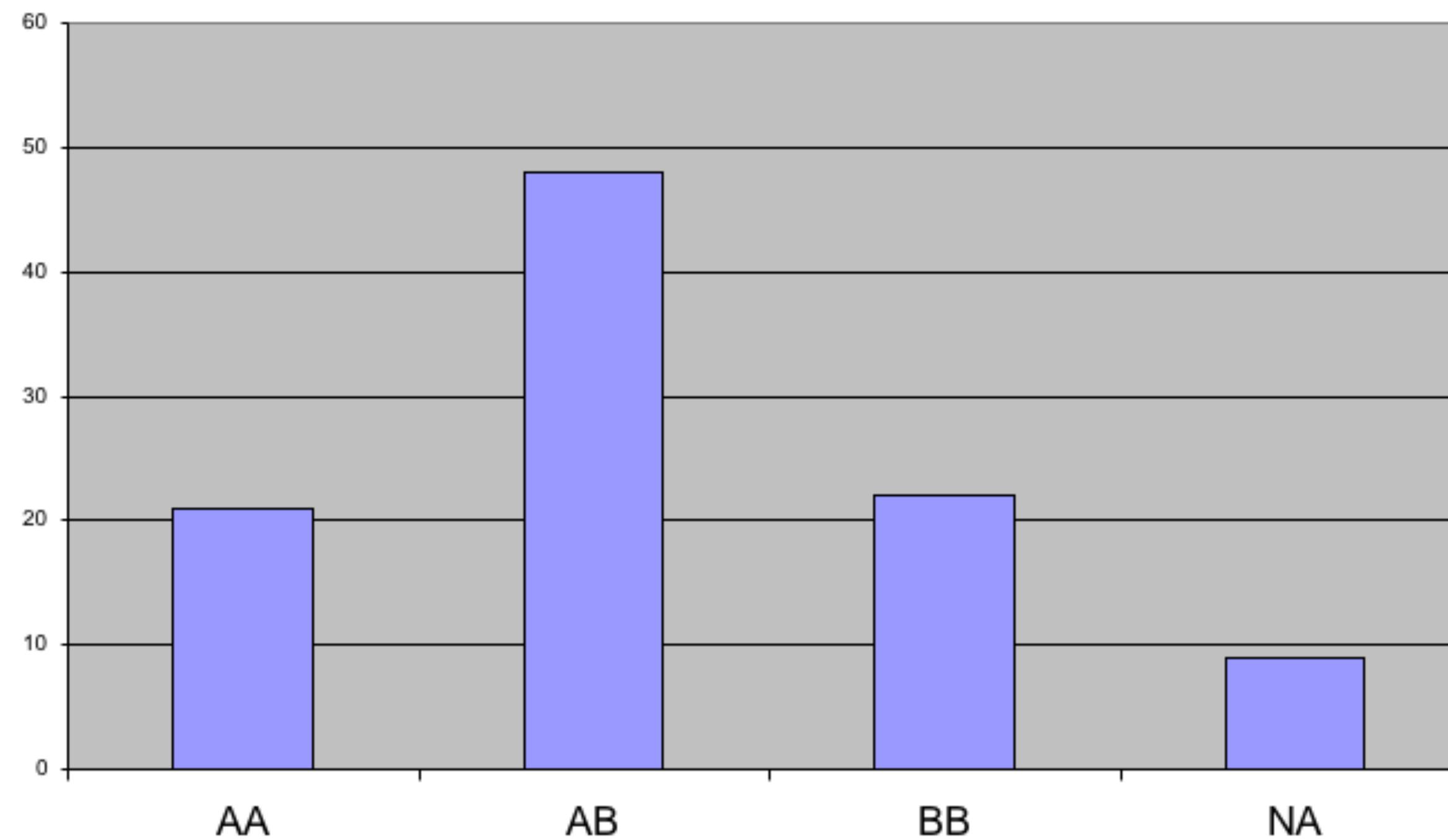


Created by Darkhorse Analytics

www.darkhorseanalytics.com

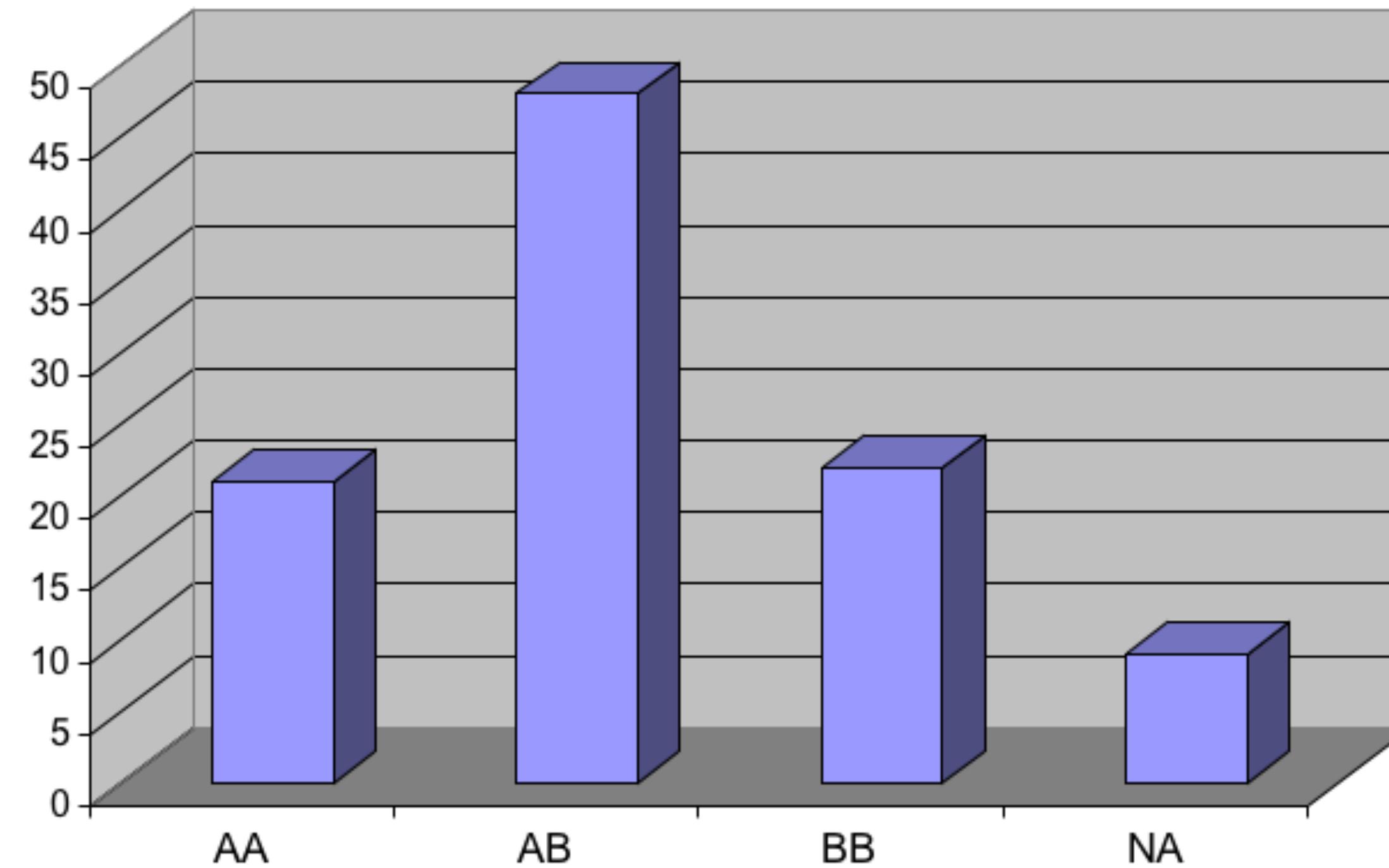
Data-to-ink Ratio

Distribution of genotypes

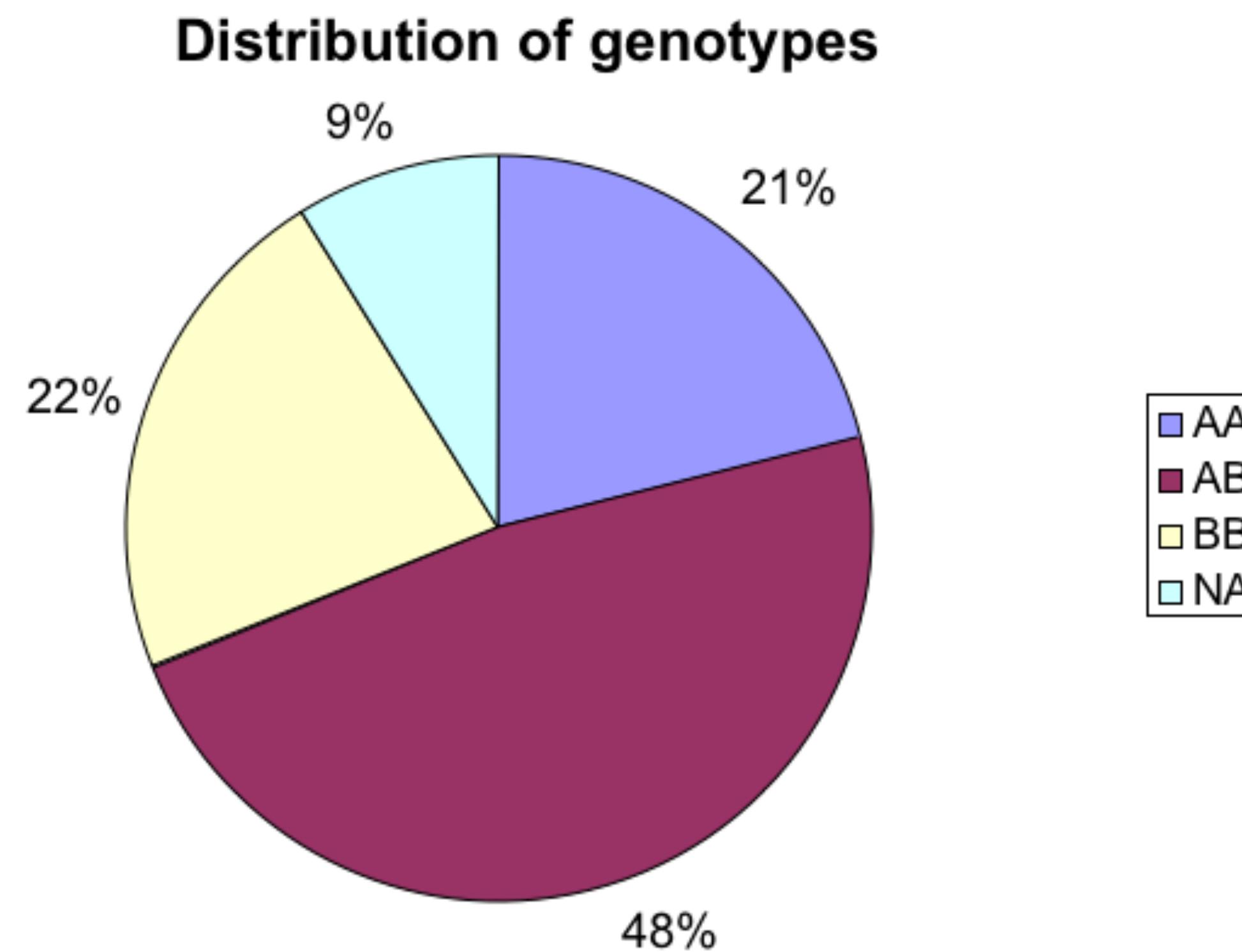


Data-to-ink Ratio

Distribution of genotypes

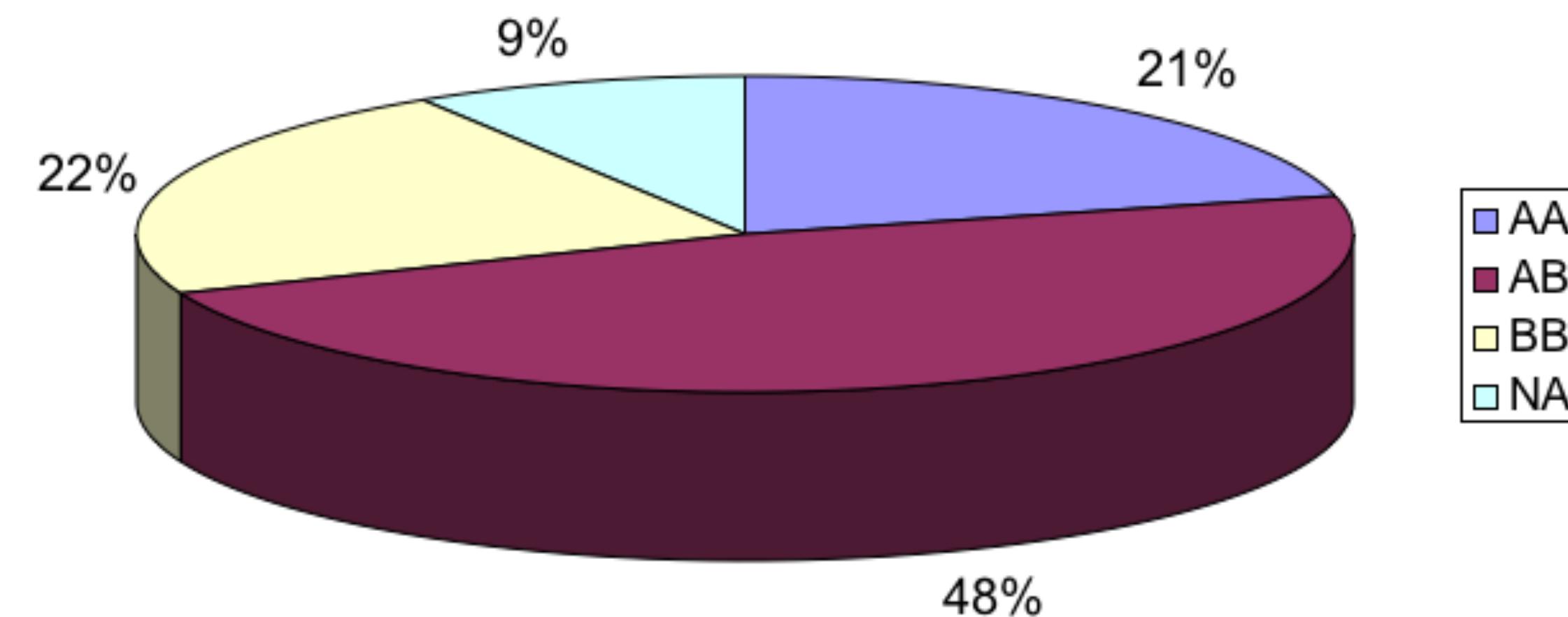


Data-to-ink Ratio



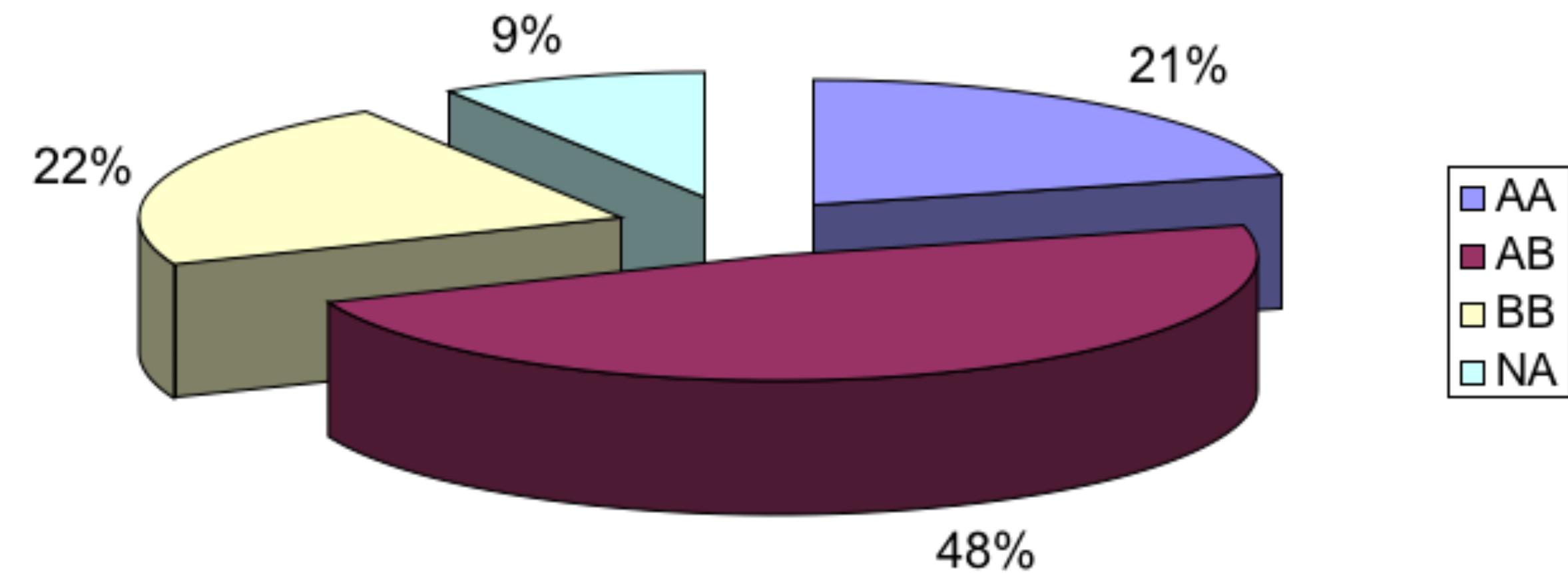
Data-to-ink Ratio

Distribution of genotypes



Data-to-ink Ratio

Distribution of genotypes



Why Visualization?

We're good at pattern recognition!

- scanning
- recognizing
- remembering images

Graphical elements facilitate comparisons:

- length
- shape
- orientation
- texture

Visualization Goals

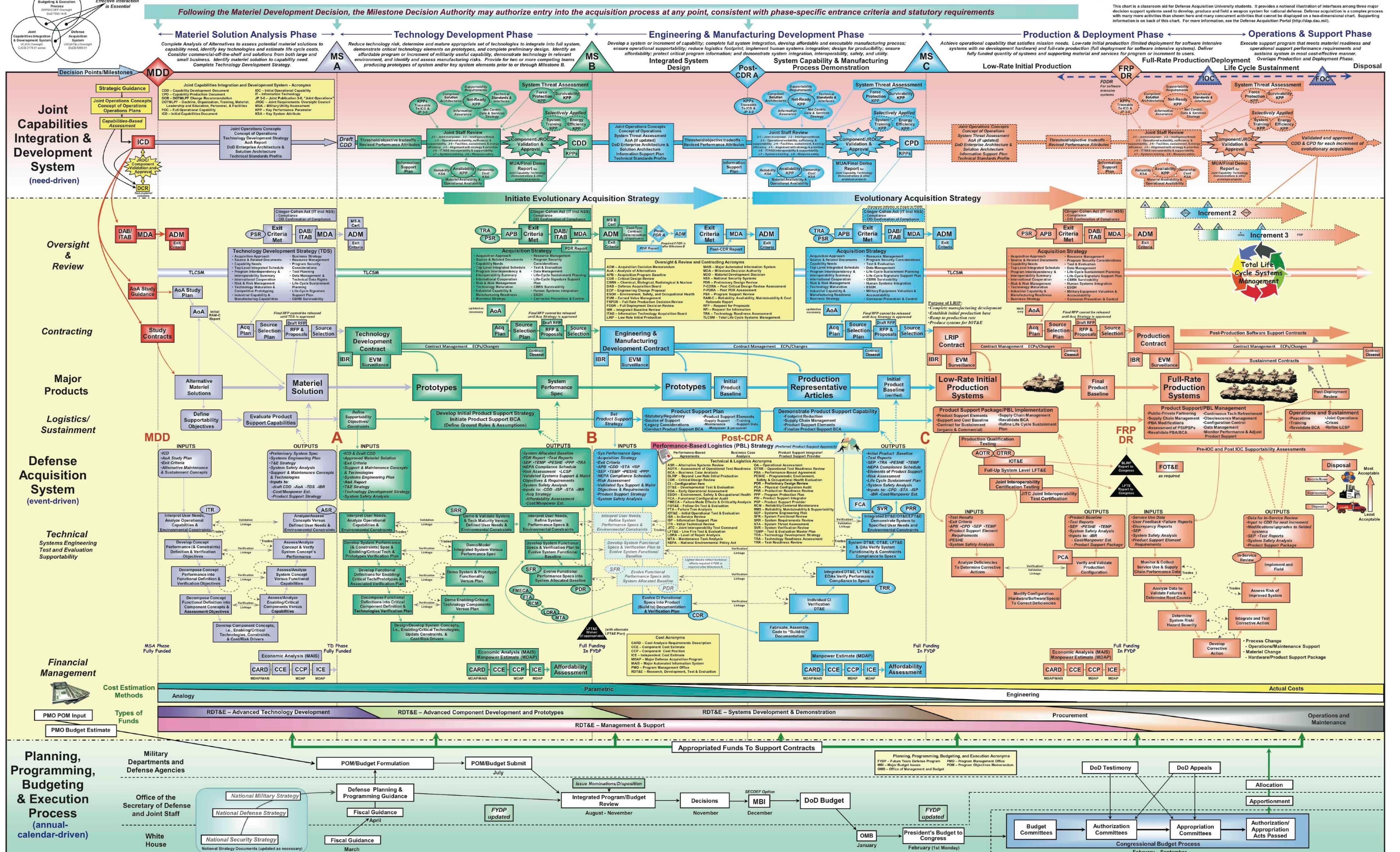
Explore/Calculate

- Analyze
- Reason about Information

Communicate

- Explain
- Make Decisions
- Reason about Information

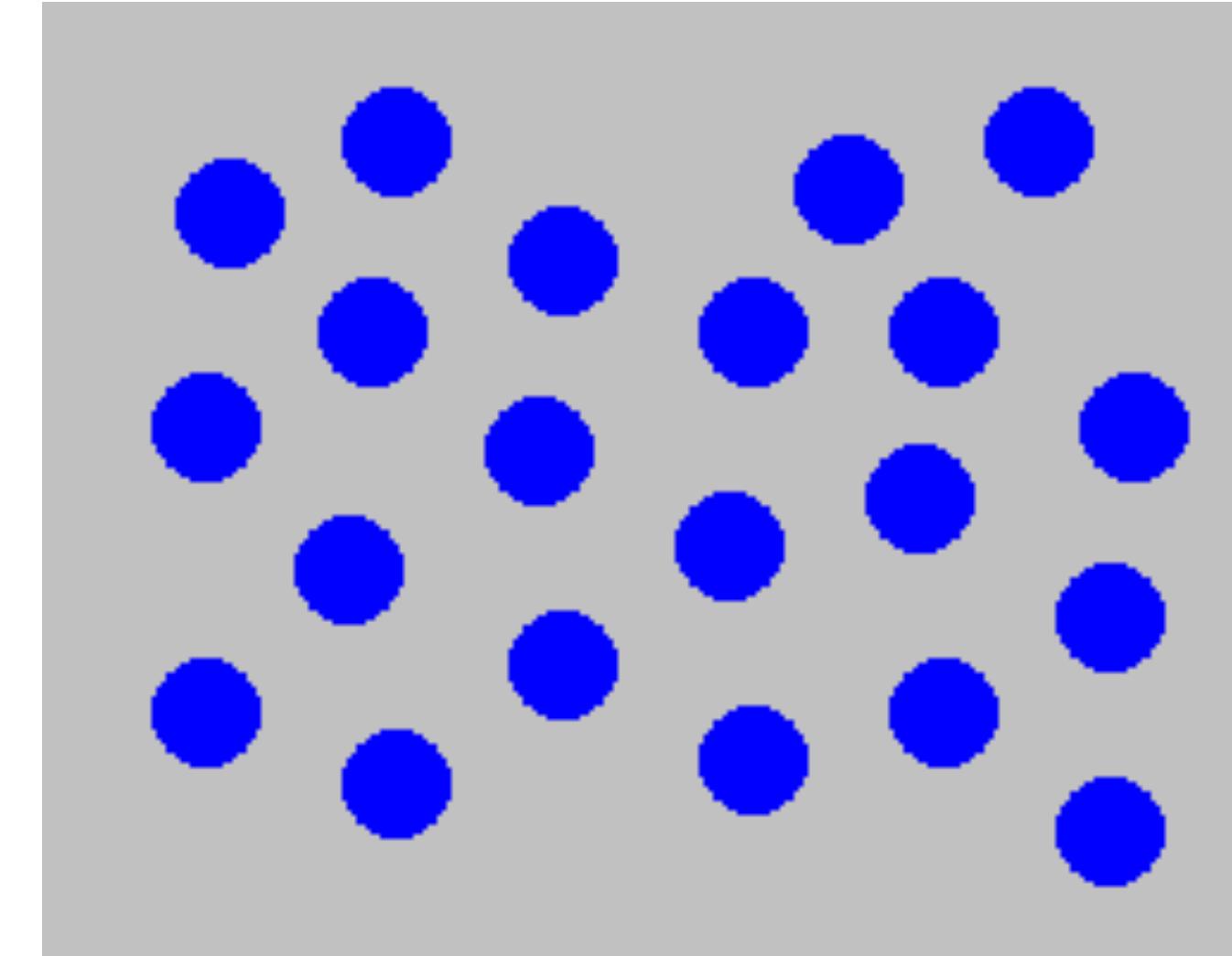
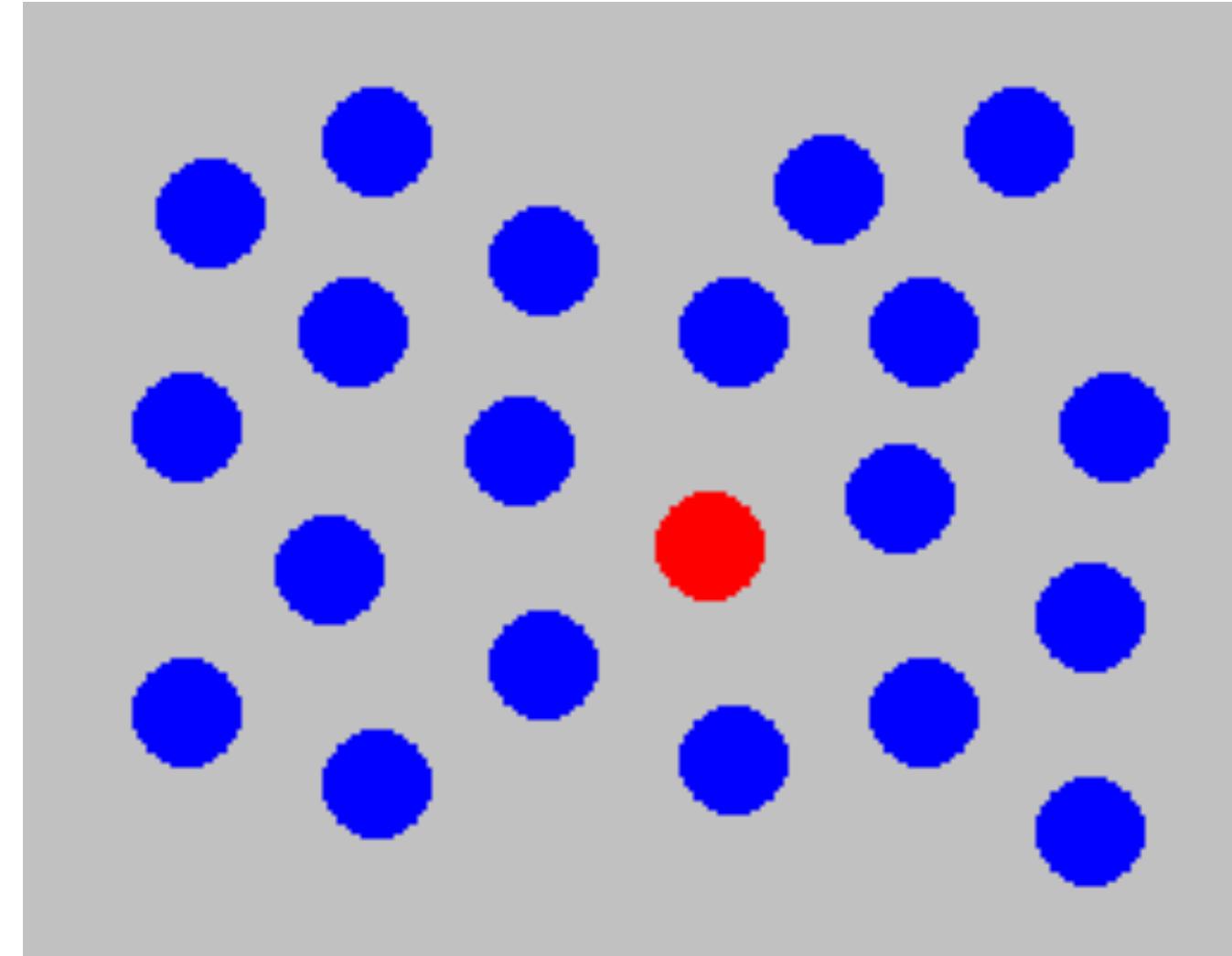
Integrated Defense Acquisition, Technology, and Logistics Life Cycle Management System



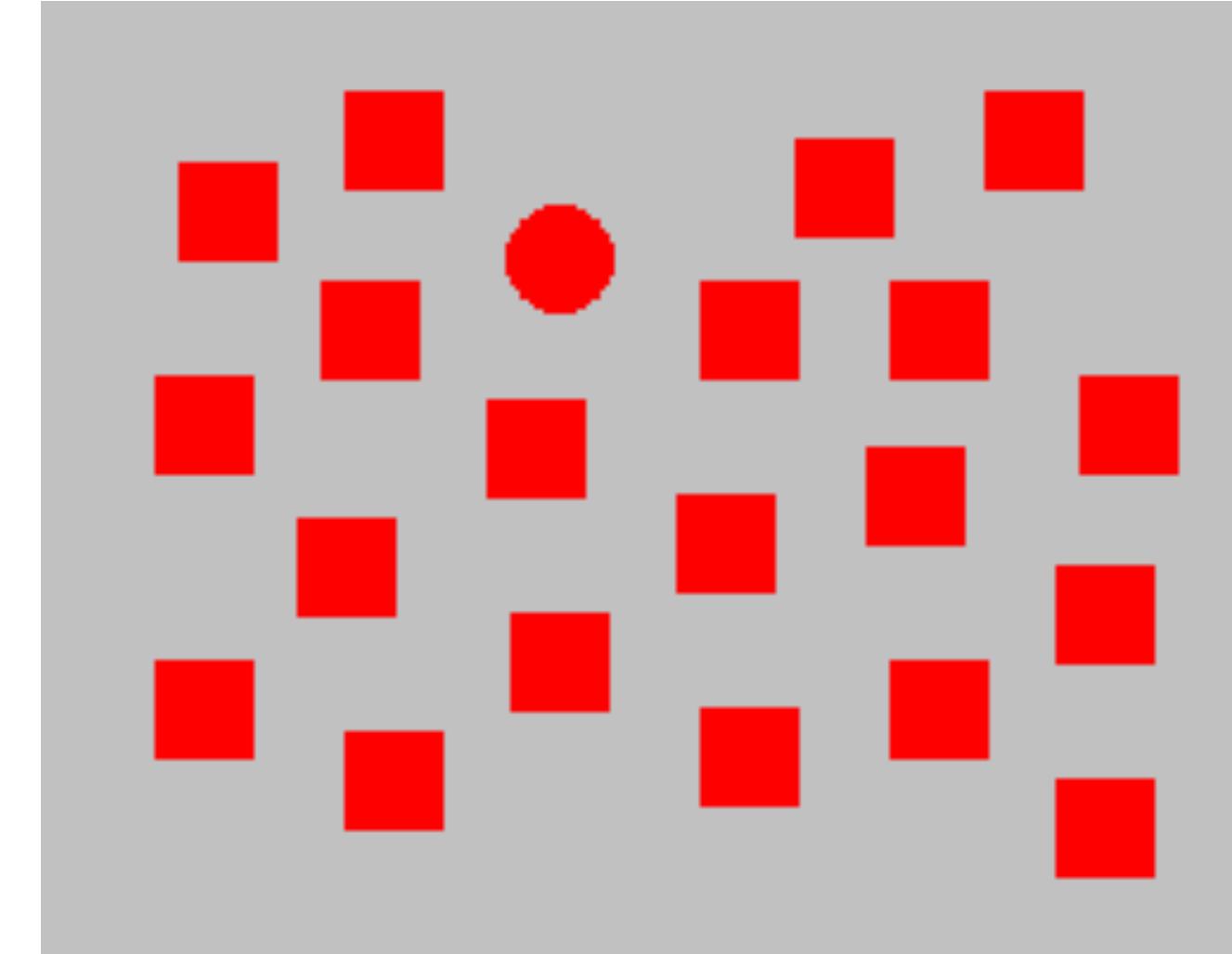
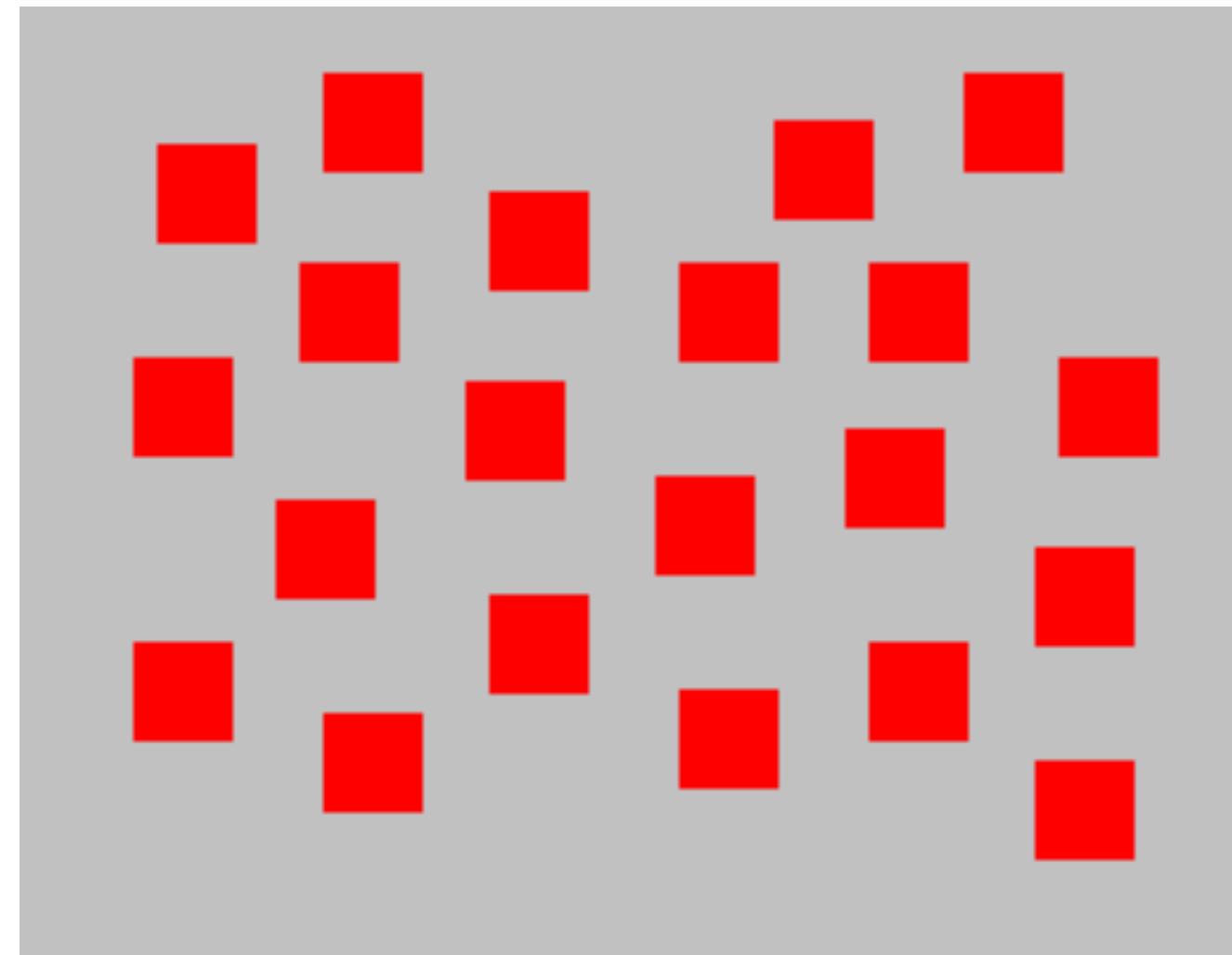
Visualization tips

- Some properties are processed pre-attentively (without need for focusing attention)
- This is important for know for the design of visualizations (COGSCI!)
- What can be perceived immediately
- What properties are good discriminators
- What can mislead viewers

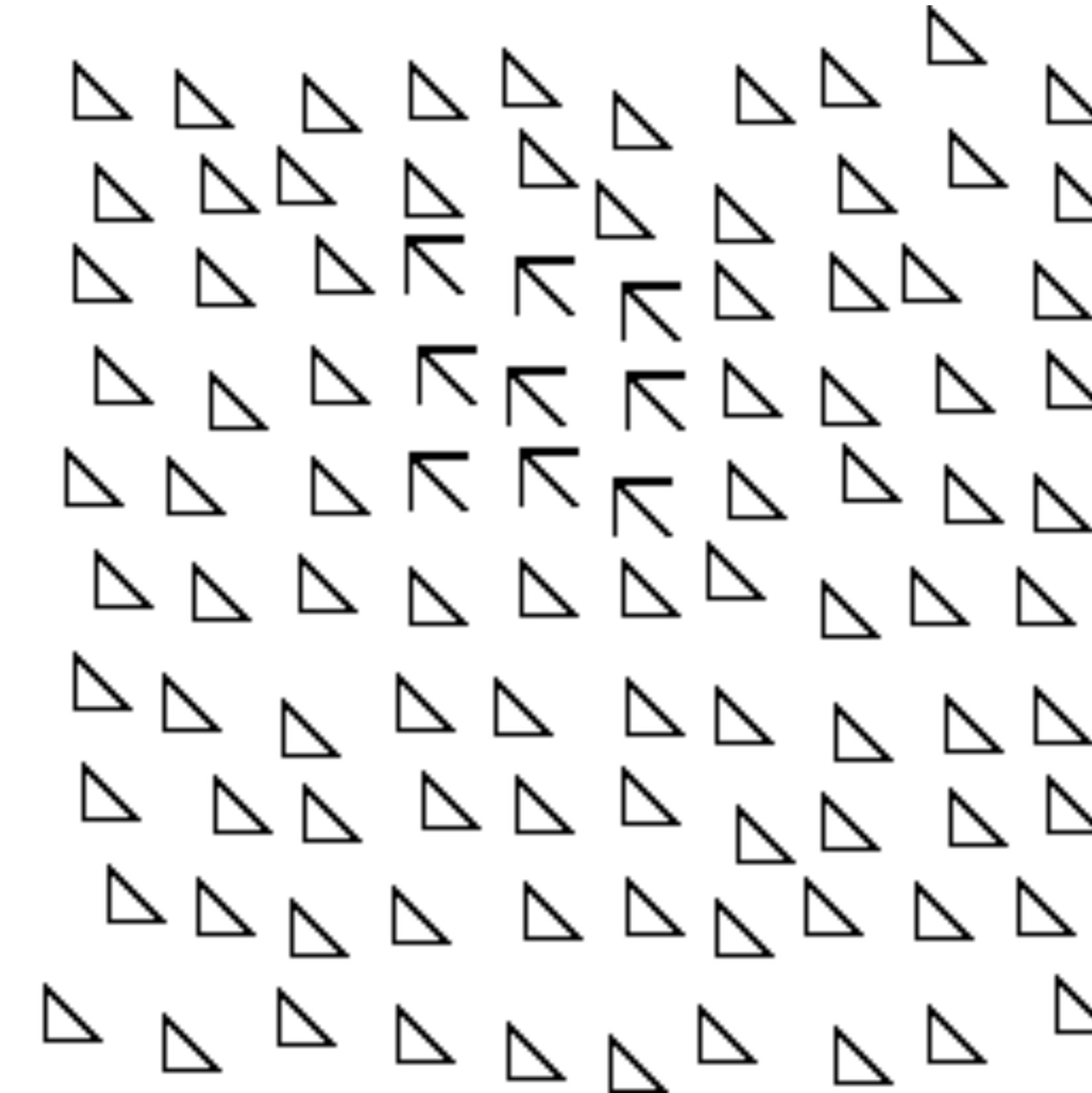
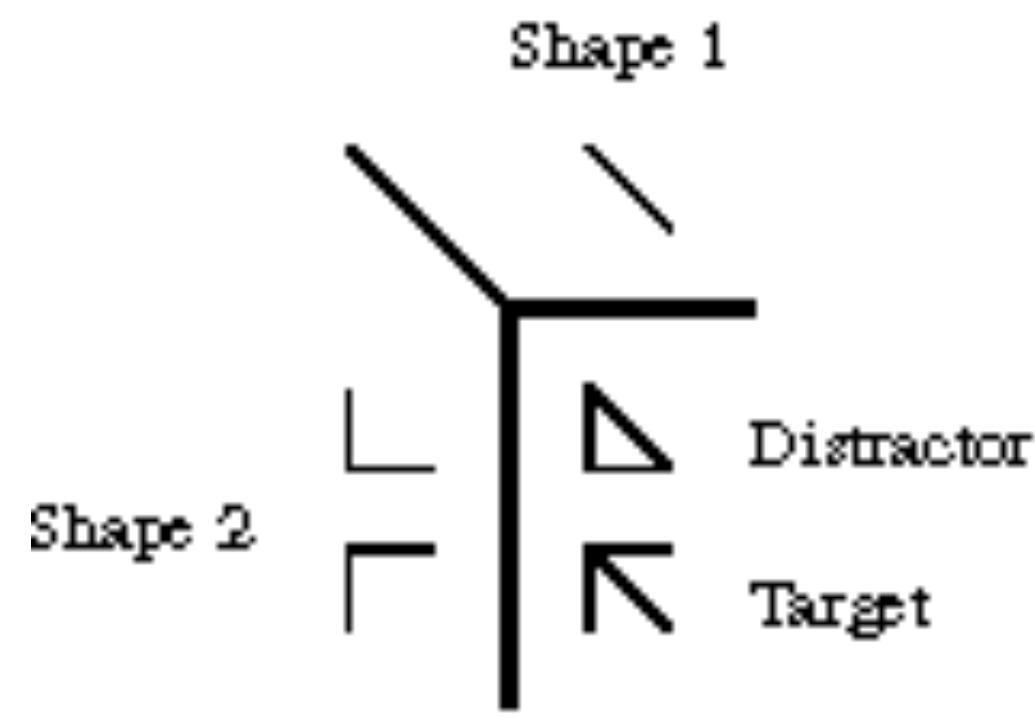
Pre-attentive Processing



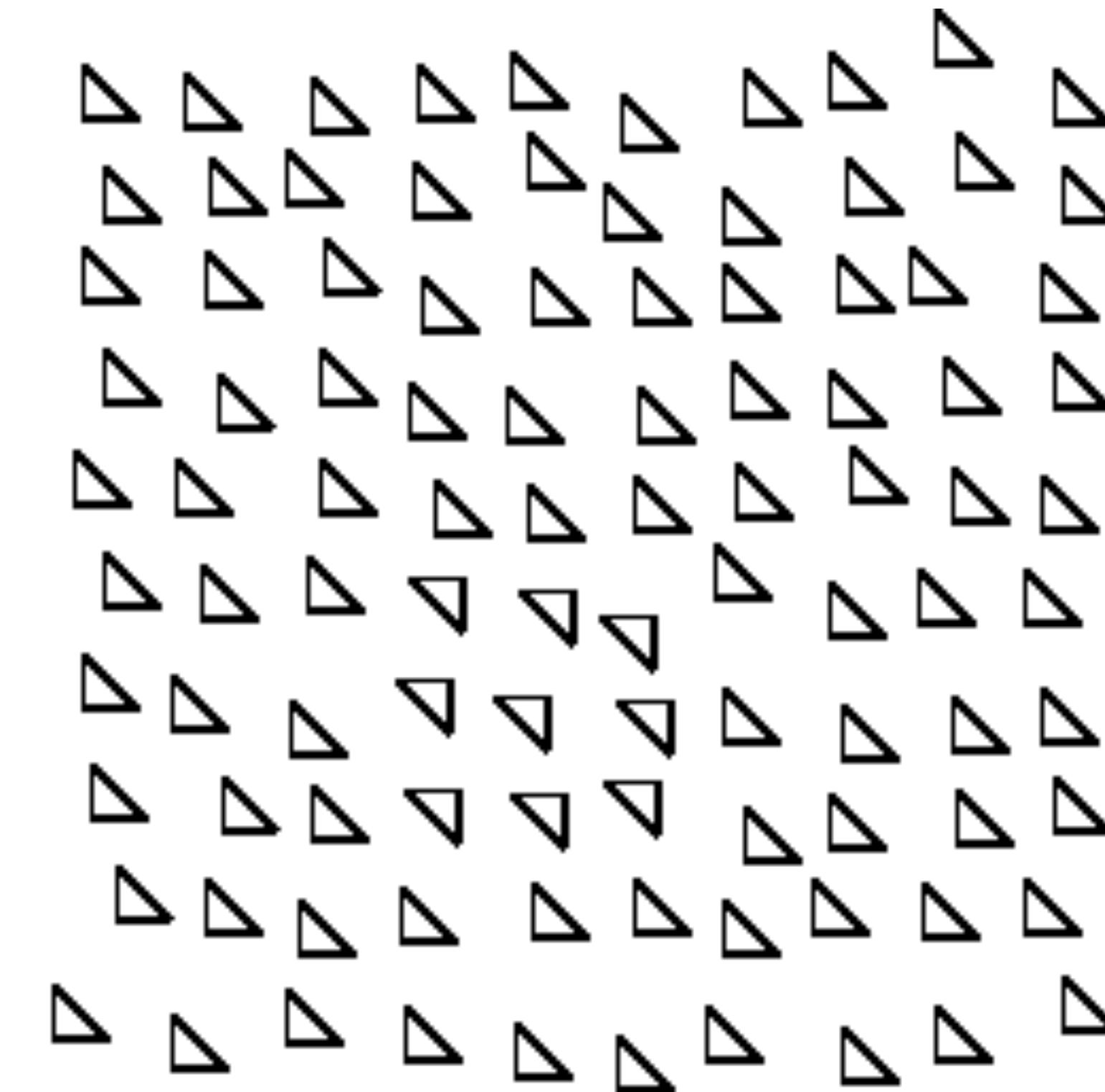
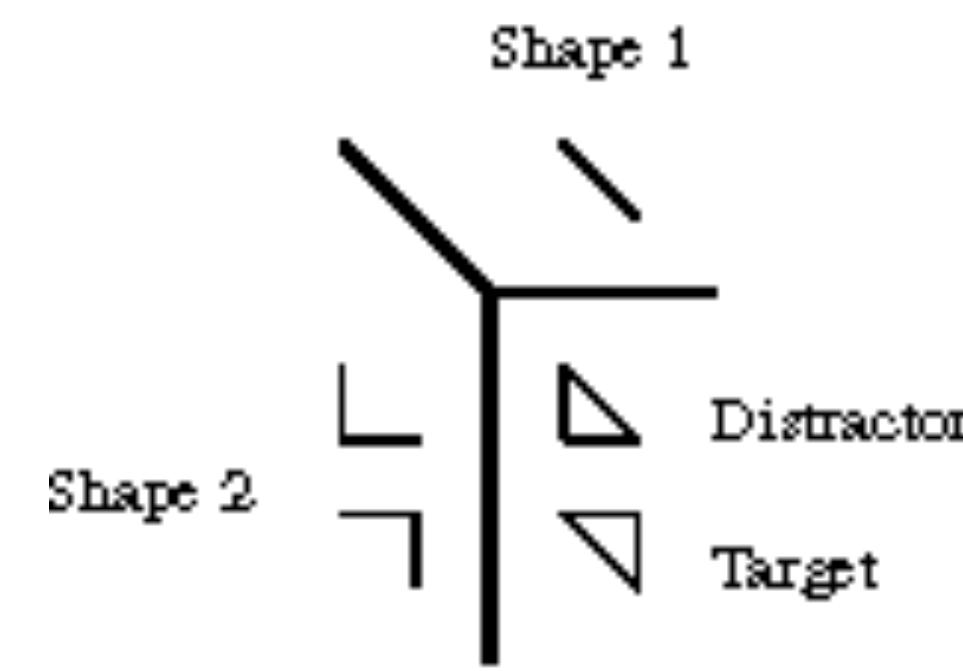
Pre-attentive Processing



Pre-attentive Processing

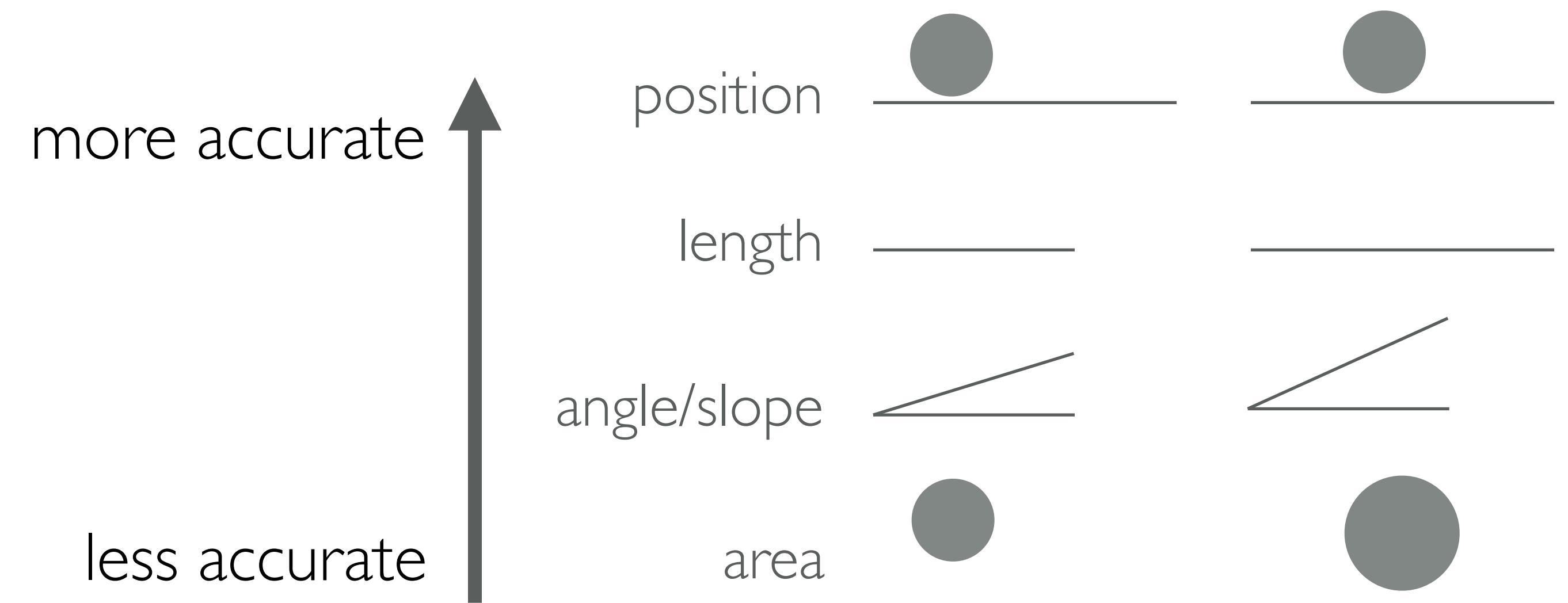


Pre-attentive Processing

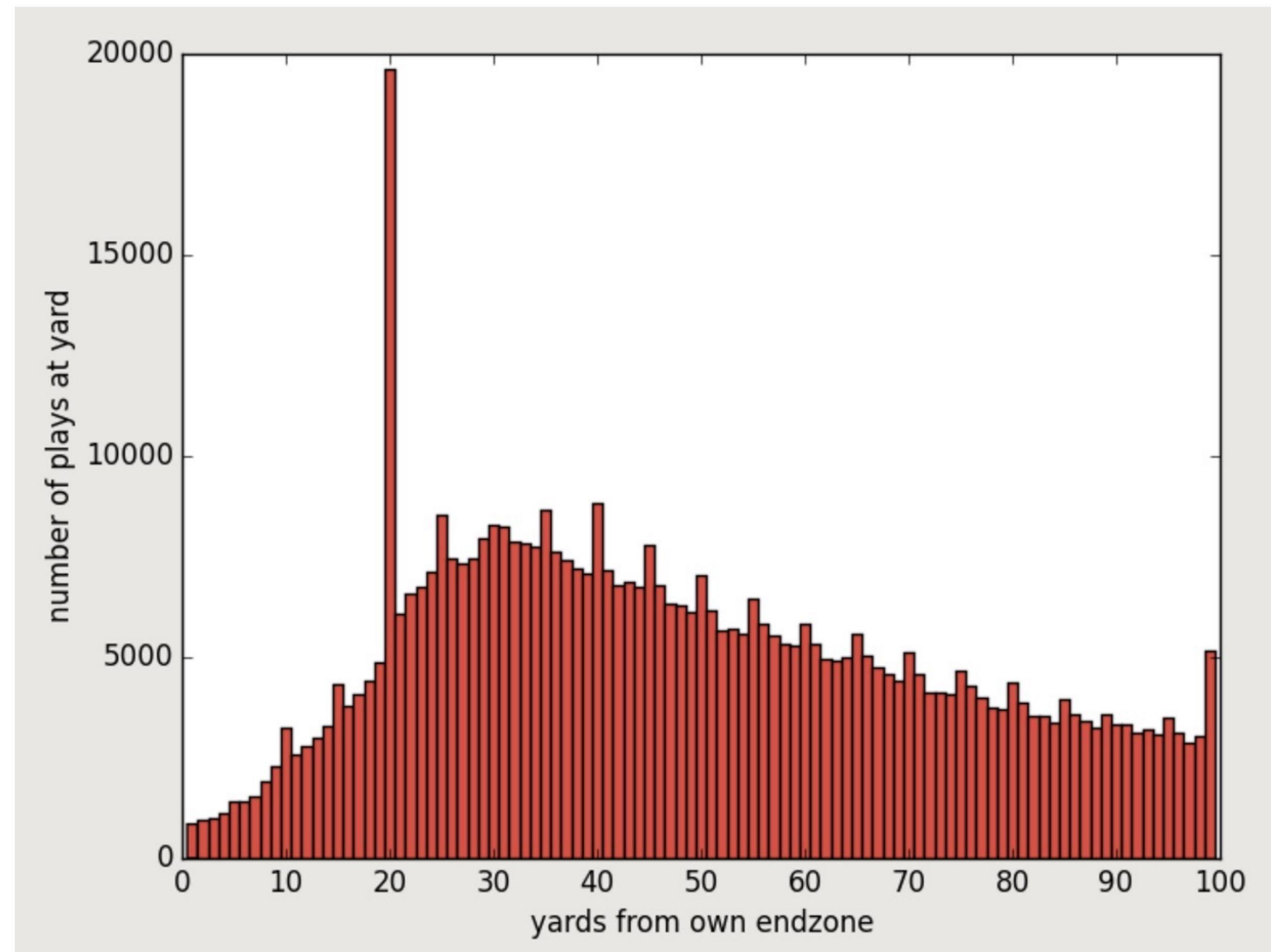


Pre-attentive Processing

- Accuracy ranking of quantitative perceptual tasks

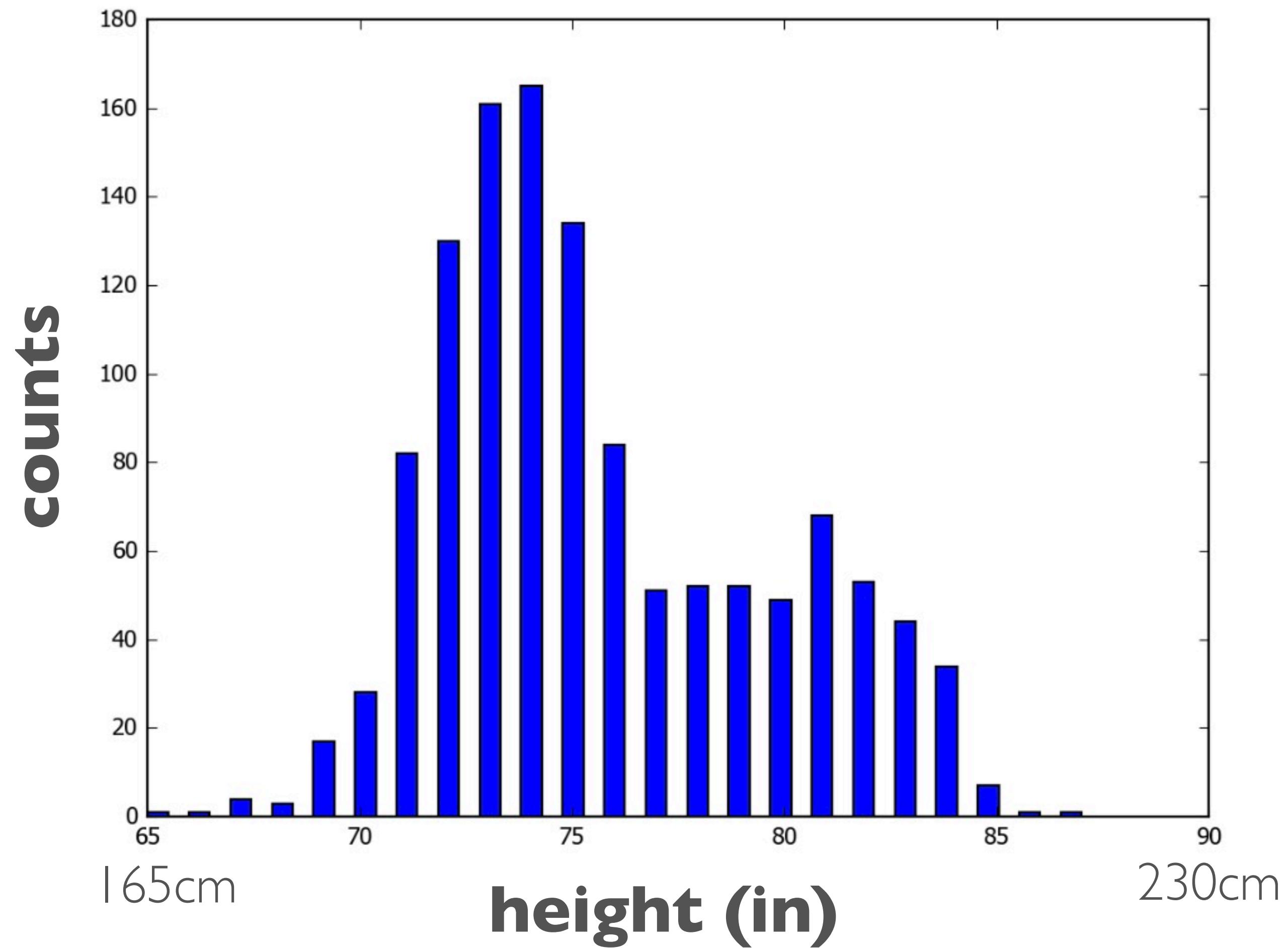


Data intuition



source: <https://gutterstats.wordpress.com/2015/11/03/are-nfl-officials-biased-with-their-ball-placement/>

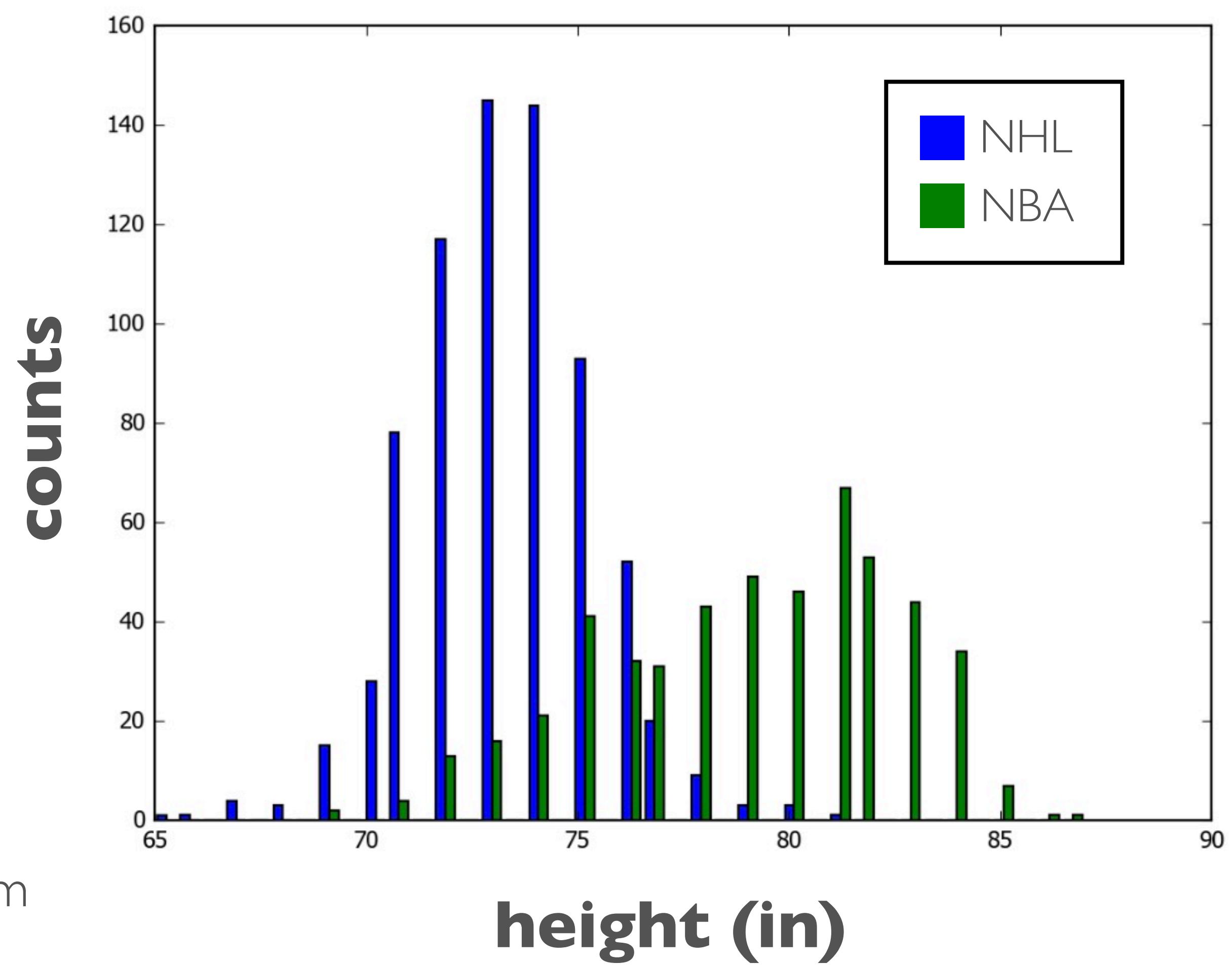
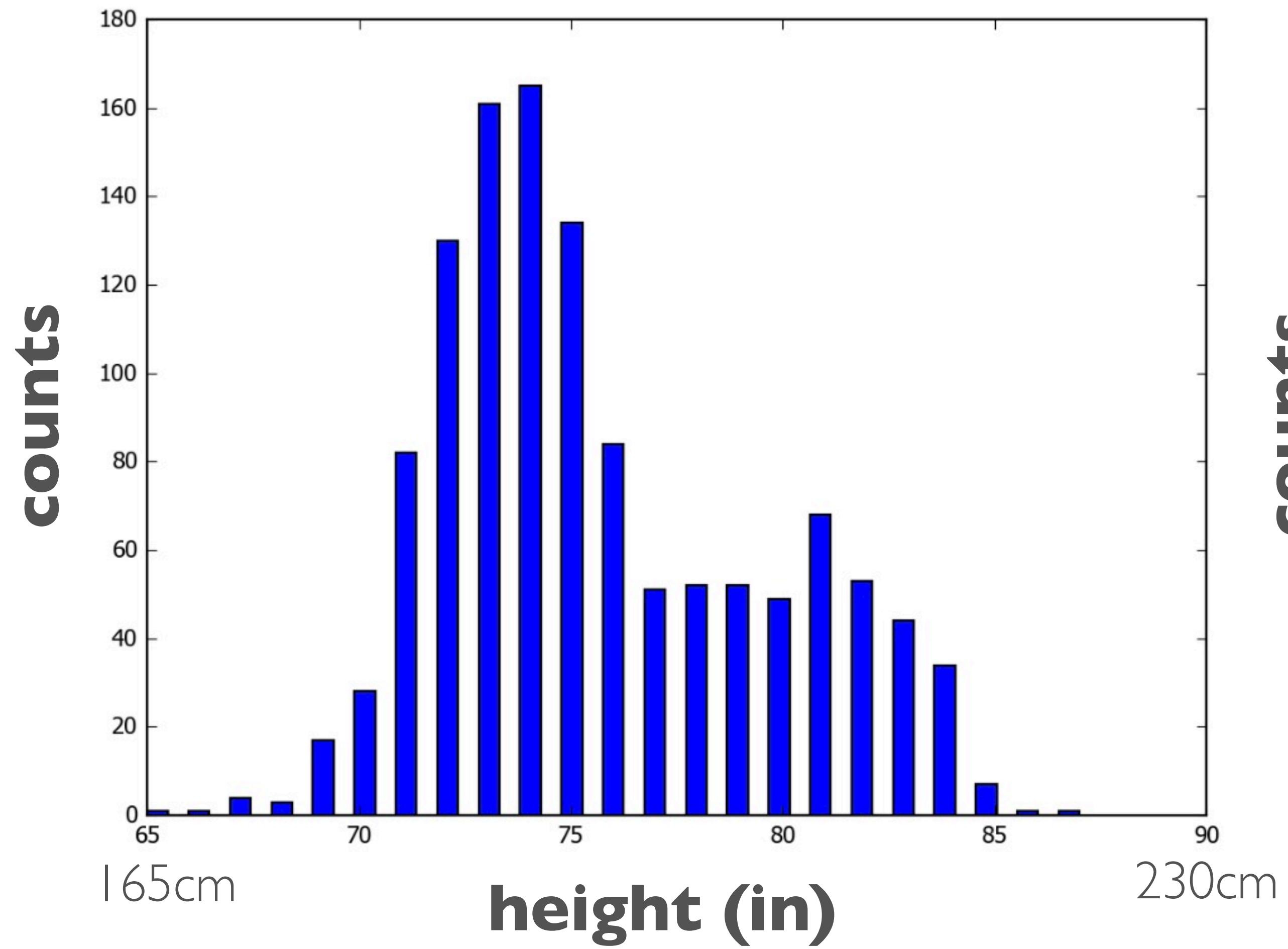
Data intuition



| 65cm

230cm

Data intuition



CURRY!

Stephen Curry's 3-Point Record in Context: Off the Charts

By GREGOR AISCH and KEVIN QUEALLY APRIL 16, 2016

This chart contains 752 lines — one for each N.B.A. player who finished in the top 20 in 3-point attempts made in each season since 1980. Sitting atop it is the Golden State Warriors' Stephen Curry, who finished the regular season with a record 402 3-pointers.

The record is an outlier that defies most comparisons, but here is one: It is the equivalent of hitting 103 home runs in a Major League Baseball season.

The colors show a clear progression toward more 3-pointers. In the 1979-80 N.B.A. season, the first to feature the 3-pointer, making just 21 was good enough to put a player among the league's top 20. On Feb. 27, Curry made 12 3-pointers in a single game.

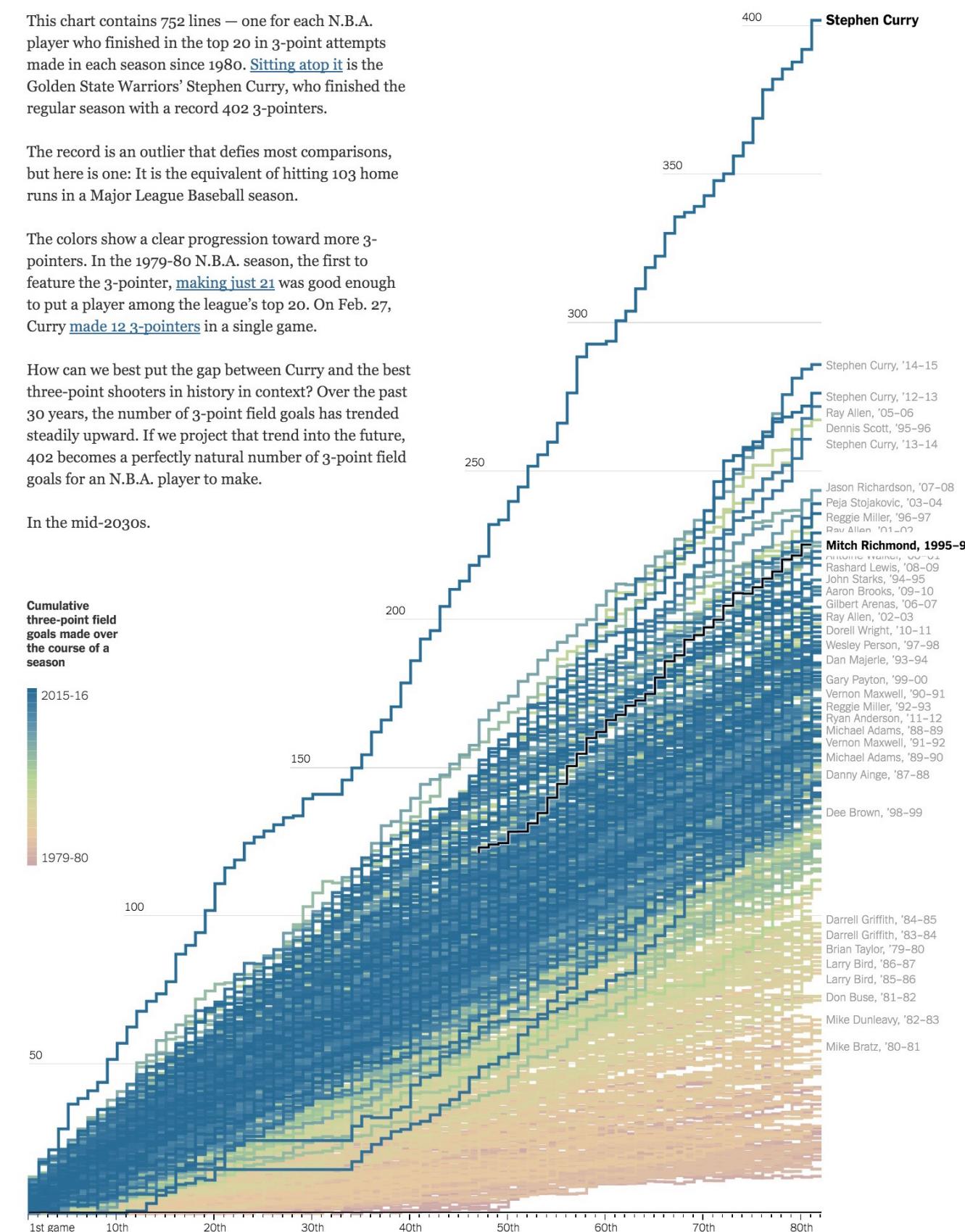
How can we best put the gap between Curry and the best three-point shooters in history in context? Over the past 30 years, the number of 3-point field goals has trended steadily upward. If we project that trend into the future, 402 becomes a perfectly natural number of 3-point field goals for an N.B.A. player to make.

In the mid-2030s.

Cumulative
three-point field
goals made over
the course of a
season

2015-16

1979-80



UC San Diego

source: The New York Times - <http://www.nytimes.com/interactive/2016/04/16/upshot/stephen-curry-golden-state-warriors-3-pointers.html>