An Introduction to Visualization

J Pollock

Adam A Miller
CIERA/Adler Planetarium

LSSTC DSFP Session 4
17 Sep 2017

What is Visualization?

What is Visualization?

Curse of dimensionality

From Wikipedia, the free encyclopedia

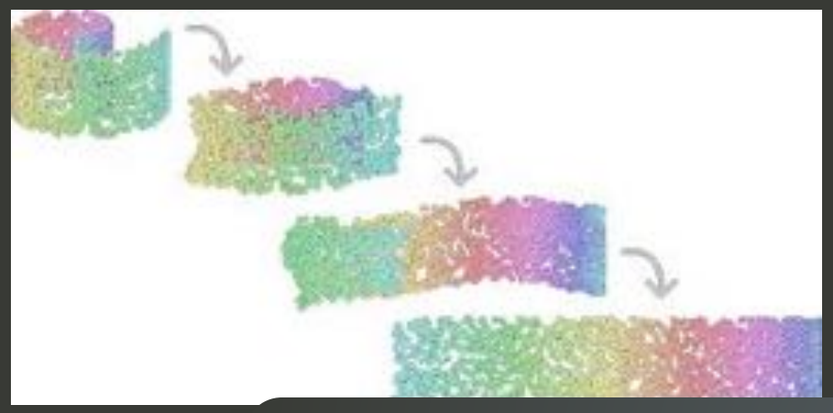
The **curse of dimensionality** refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces (often with hundreds or thousands of dimensions) that do not occur in low-dimensional settings such as the three-dimensional physical space of everyday experience. The expression was coined by Richard E. Bellman when considering problems in dynamic optimization.[1][2]

There are multiple phenomena referred to by this name in domains such as numerical analysis, sampling, combinatorics, machine learning, data mining, and databases. The common theme of these problems is that when the dimensionality increases, the volume of the space increases so fast that the available data become sparse. This sparsity is problematic for any method that requires statistical significance. In order to obtain a statistically sound and reliable result, the amount of data needed to support the result often grows exponentially with the dimensionality. Also, organizing and searching data often relies on detecting areas where objects form groups with similar properties; in high dimensional data, however, all objects appear to be sparse and dissimilar in many ways, which prevents common data organization strategies from being efficient.

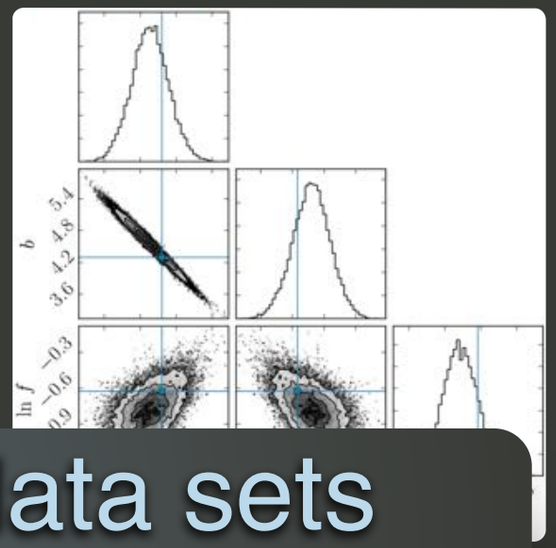
What is Visualization?

Related to several topics

Dimensionality Reduction

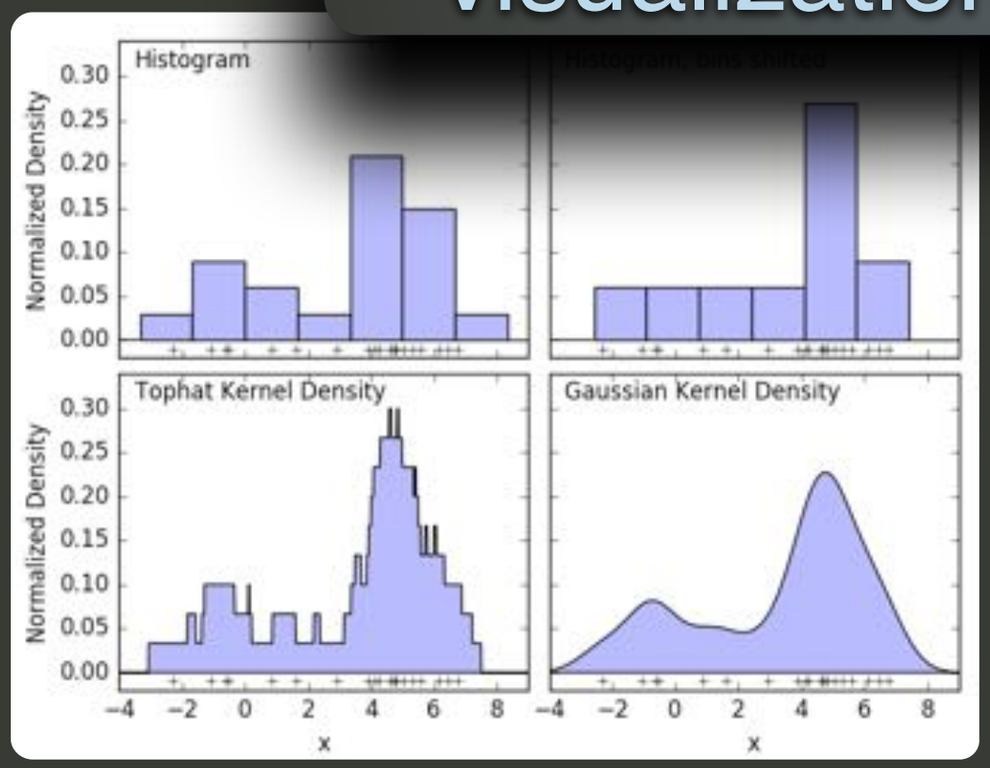


Data Exploration

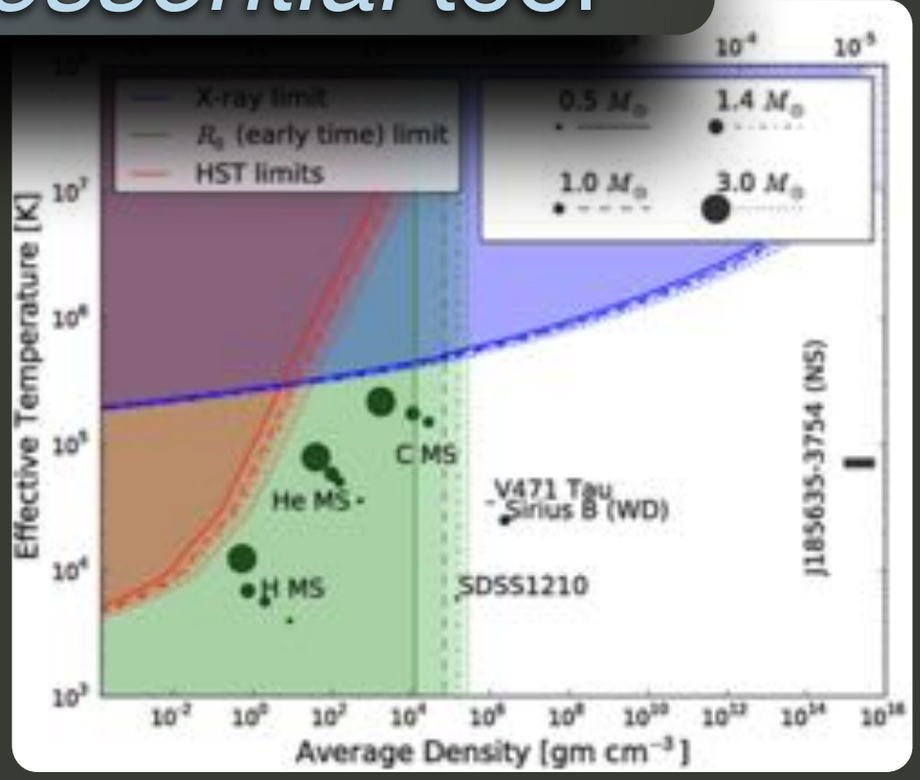


In the era of large-D data sets visualization is an *essential* tool

Dens



credit: sklearn



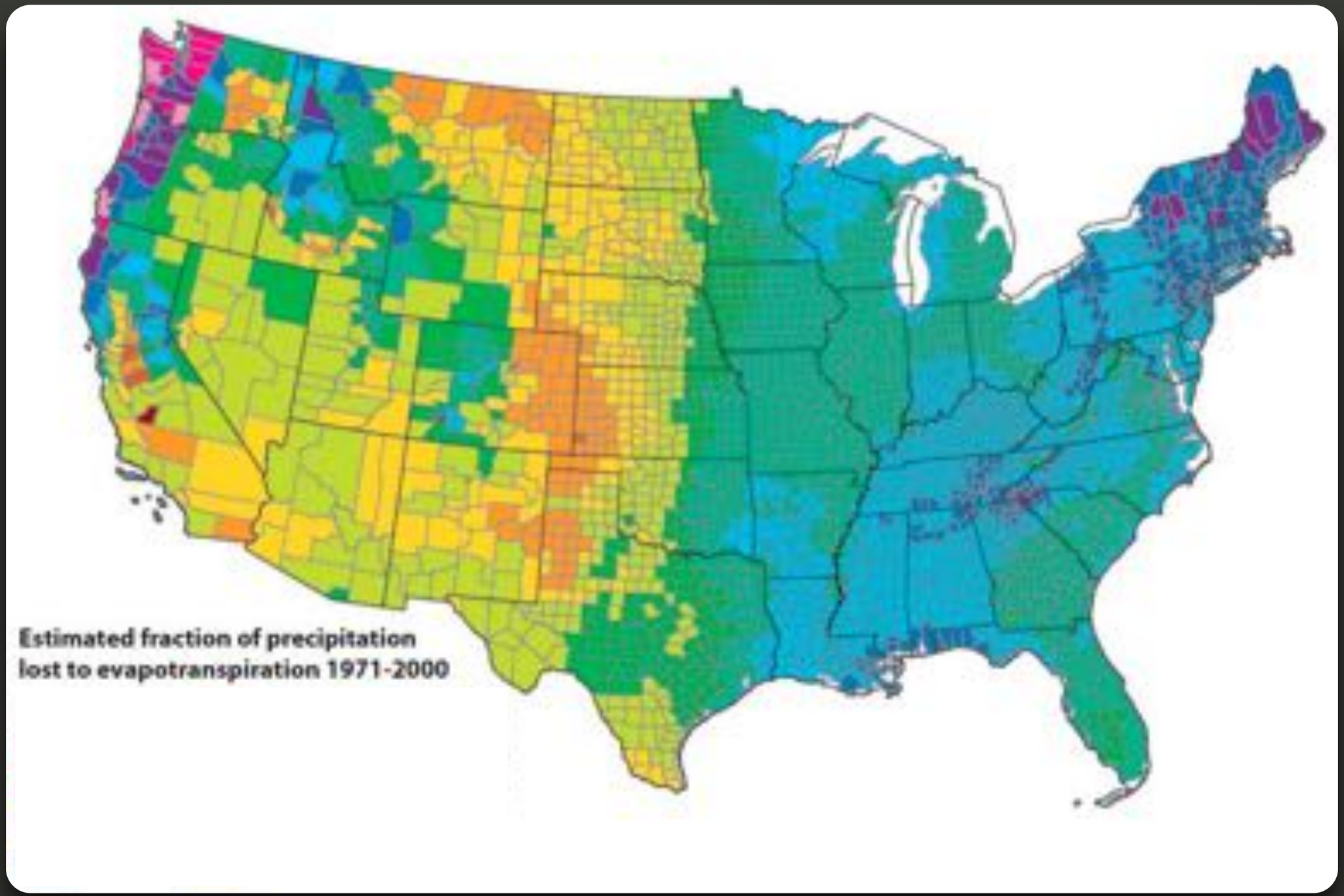
D. Foreman-Mackey

Bloom+12

Visualization Bugaboos

Choice of color is essential

Rainbow colormaps are EVIL



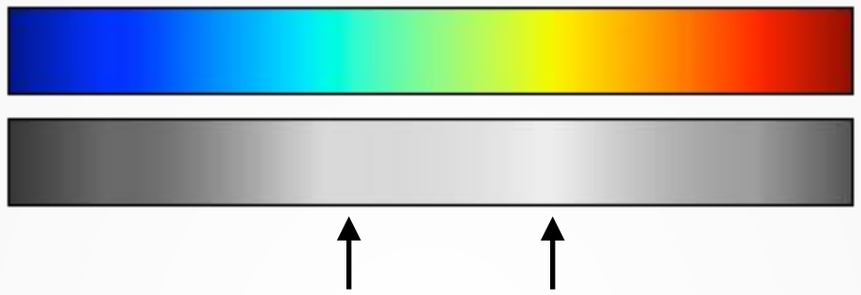
Sanford & Selnick 12

Visualization Bugaboos

Choice of color is essential

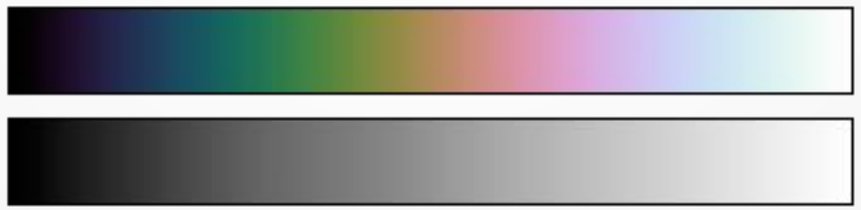
Rainbow colormaps are EVIL

jet



strange luminescence spikes

cubehelix

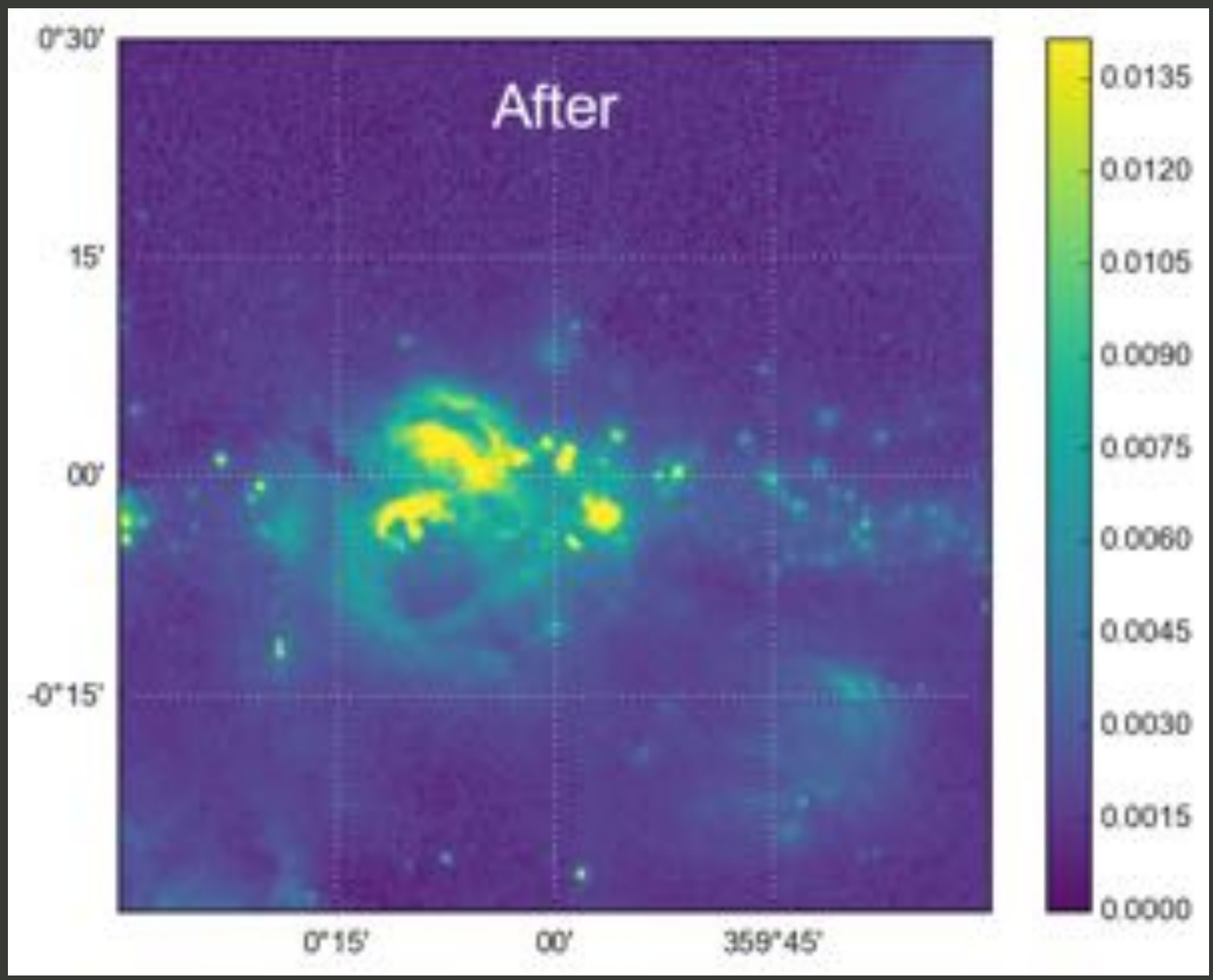


Visualization Bugaboos

Choice of color is essential

Rainbow colormaps are EVIL

Jet (rainbow) used to be matplotlib default!
Now - viridis

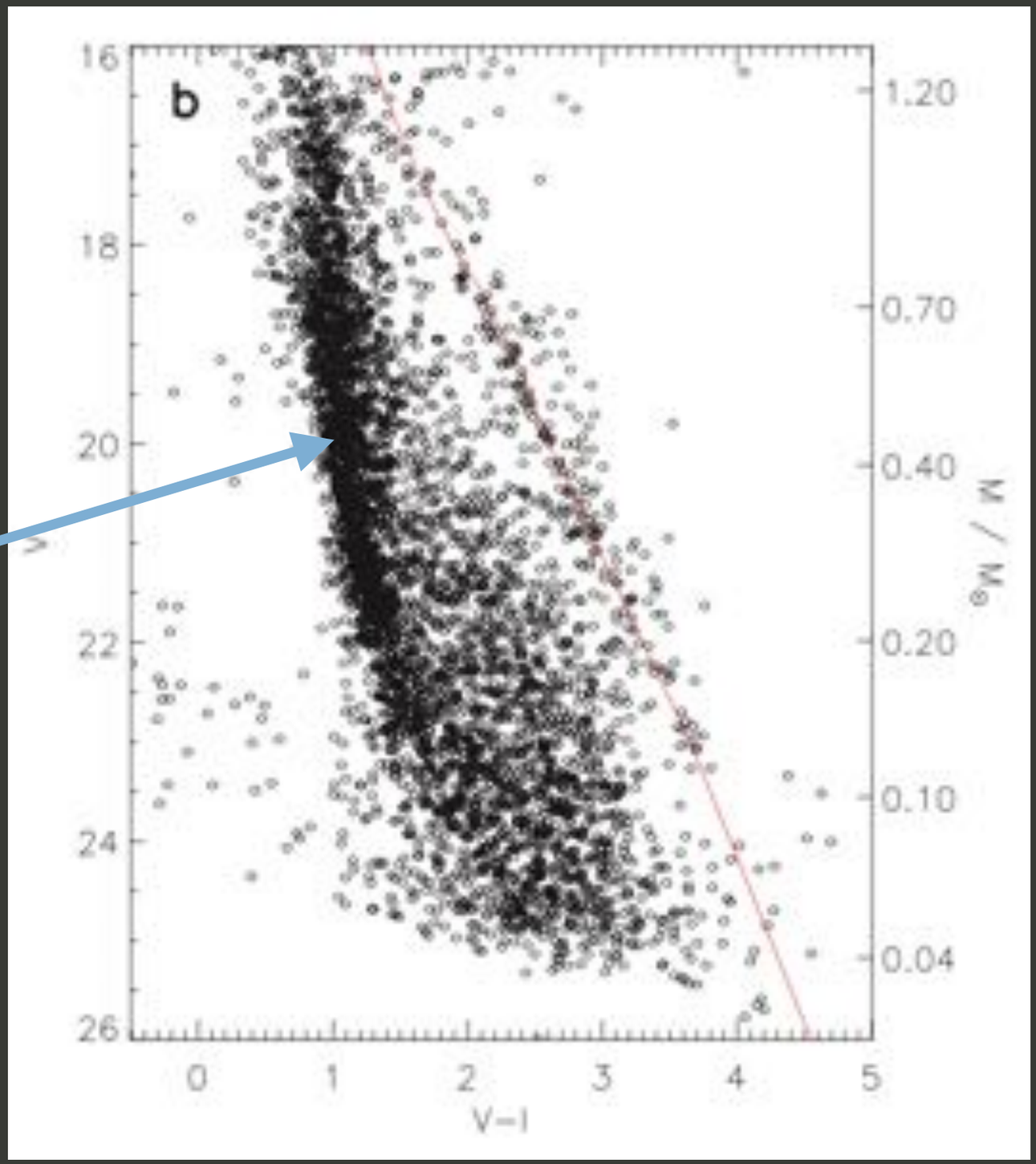


credit: T. Robitaille

Visualization Bugaboos

Too many points hide the truth

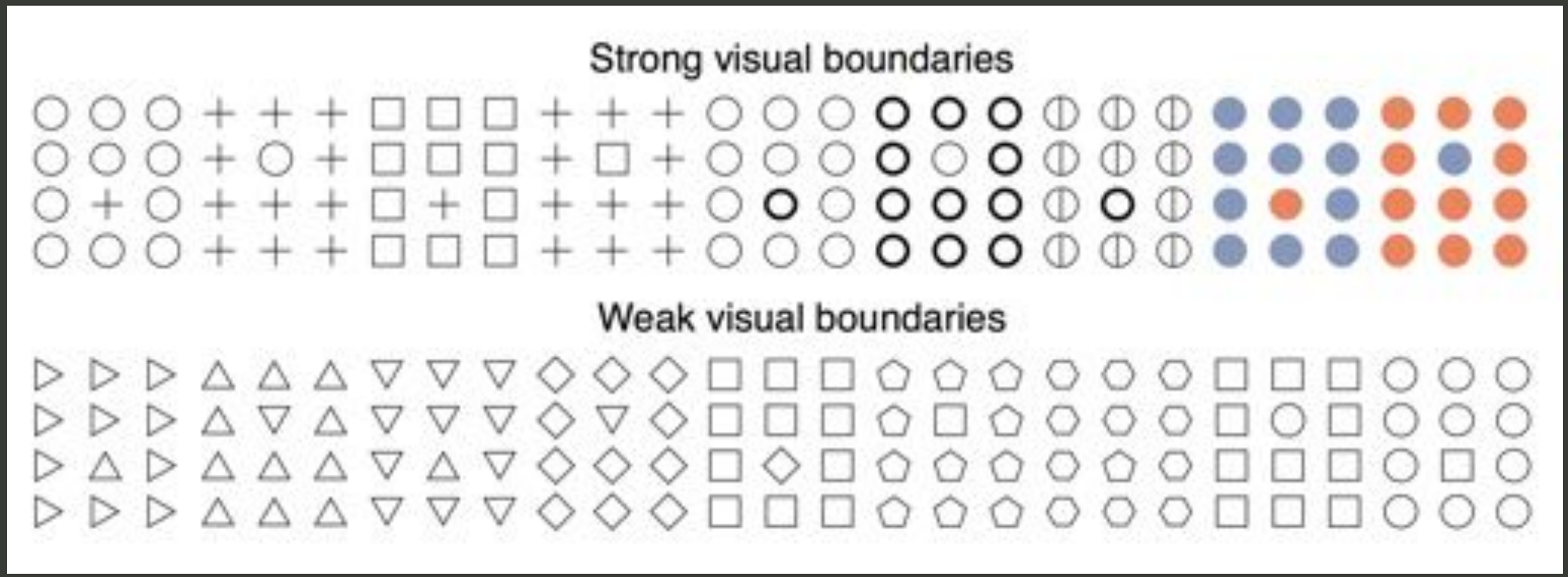
WTH is happening here?



Miller+08

Visualization Considerations

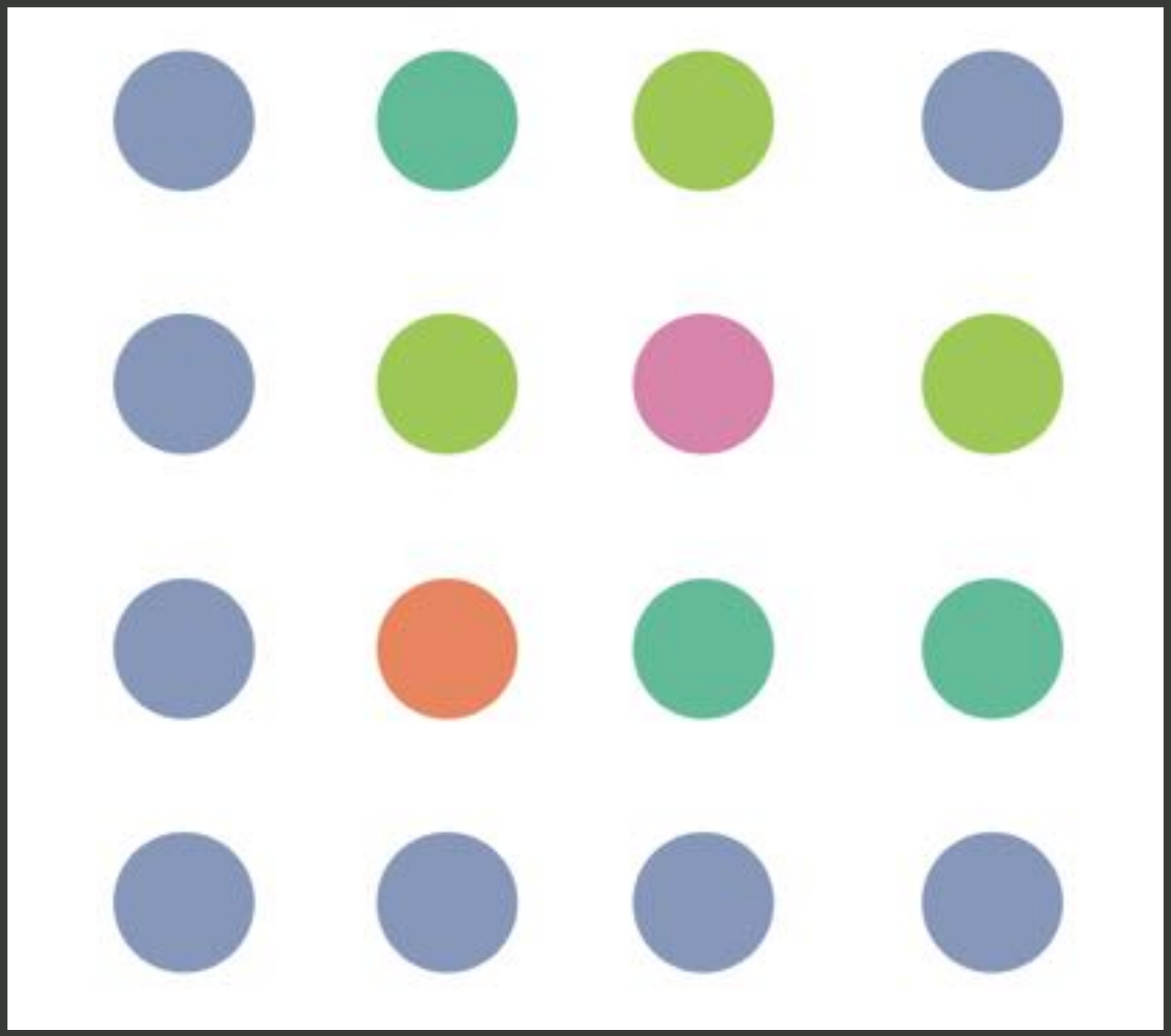
Choose strong visual boundaries



Krzywinski & Wong, Nature Methods, 2013

Visualization Considerations

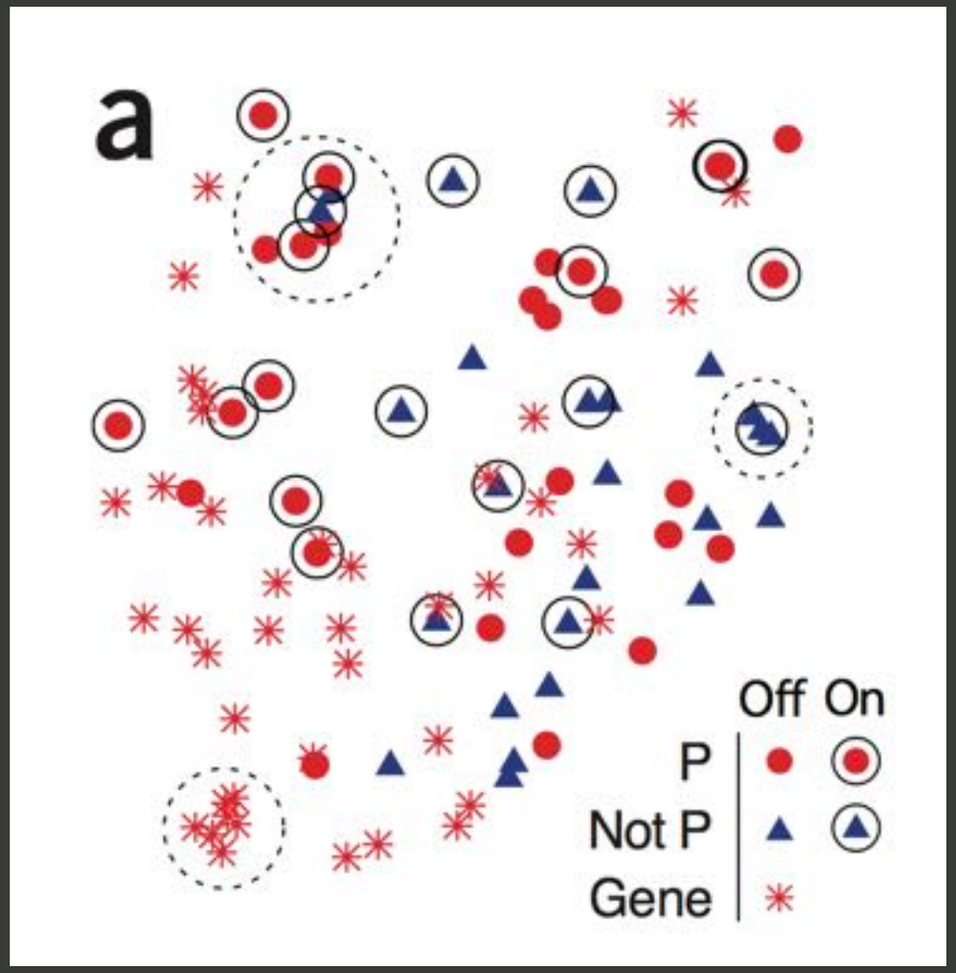
Use color judiciously

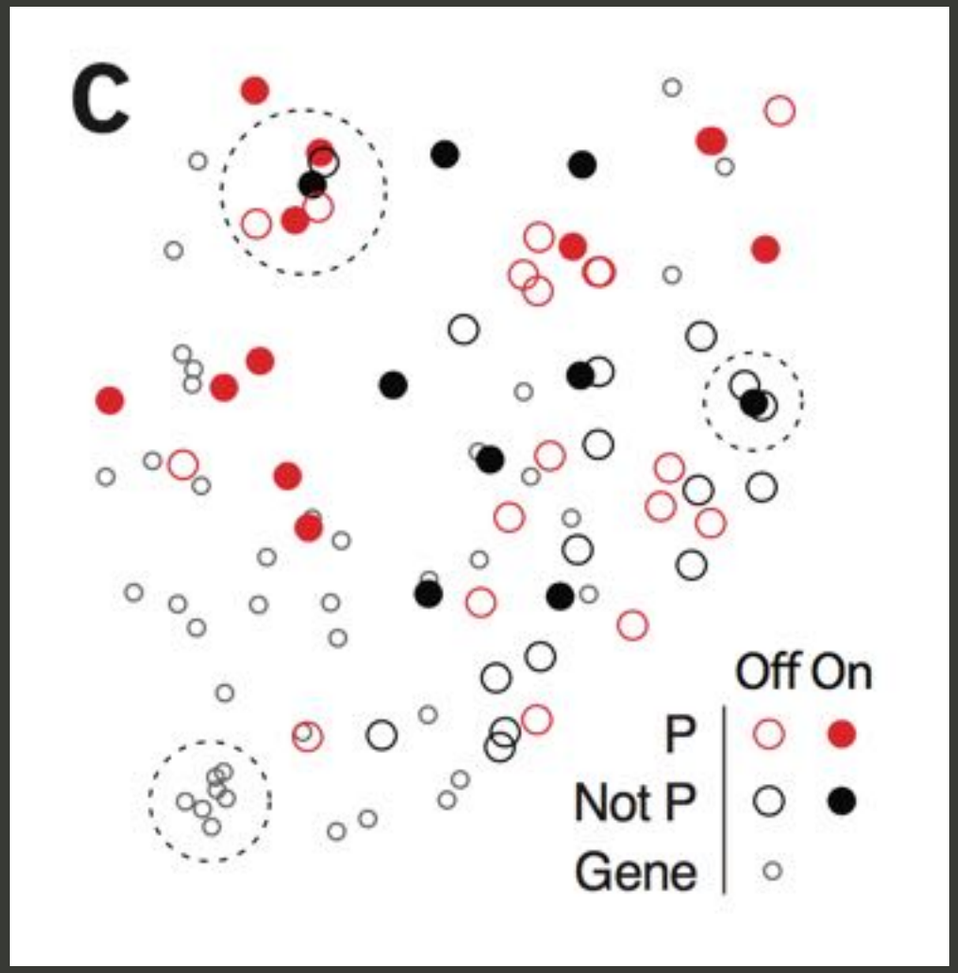


Krzywinski & Wong, Nature Methods, 2013

Visualization Considerations

Represent data hierarchies





Krzywinski & Wong, Nature Methods, 2013