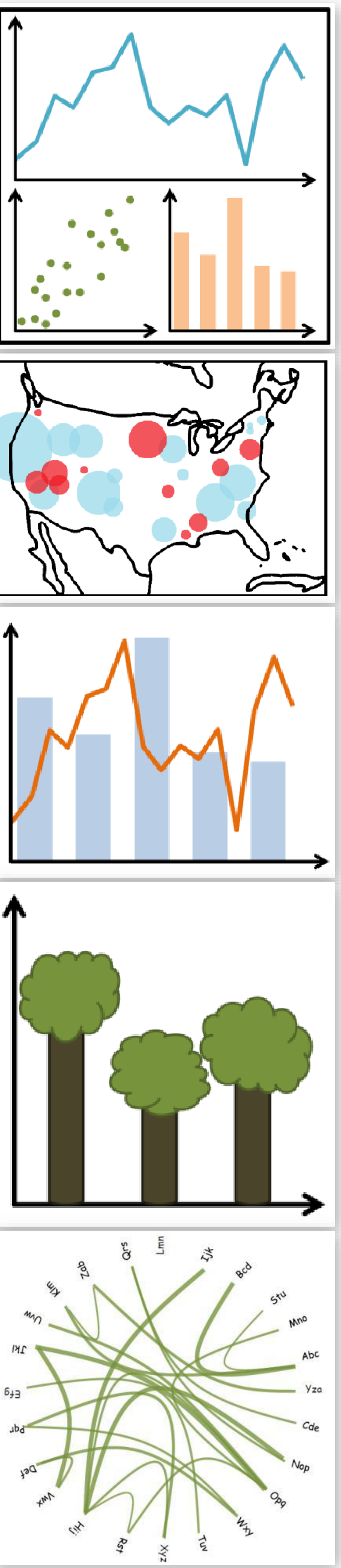


Design Rules of Thumb — Continued 2

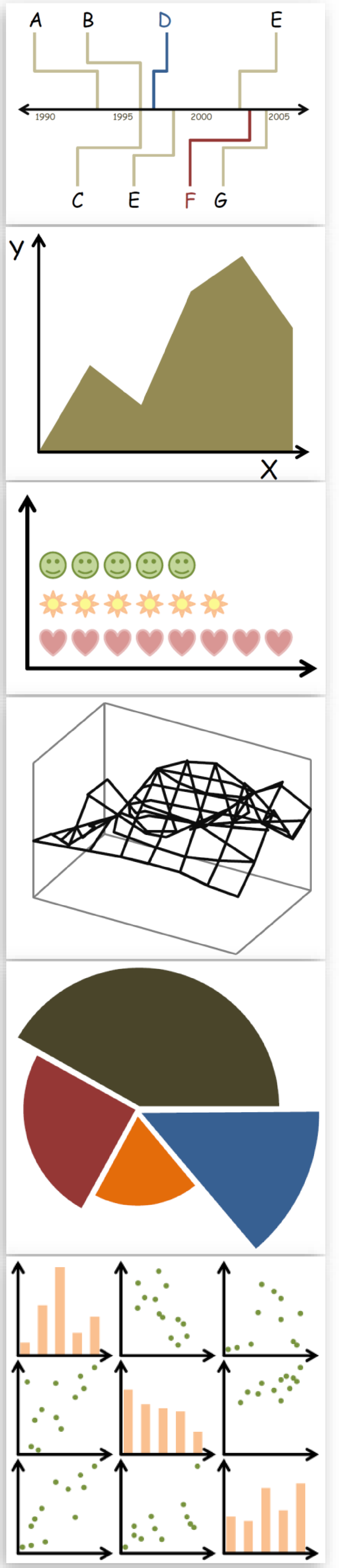
CS 7250

SPRING 2020

Prof. Cody Dunne

NORTHEASTERN UNIVERSITY

Slides and inspiration from Michelle Borkin, Krzysztof Gajos, Hanspeter Pfister, Miriah Meyer, Jonathan Schwabish, and David Sprague



Previously, on CS 7250…

JS, D3

In-Class Programming – D3 Line Chart (Part 1)

Now, on CS 7250…

In-Class Programming — D3 Line Chart (Part 2)

Previously, on CS 7250…

“Graphical Integrity”

“Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.”

(Axes and axis labels, titles, annotations, legends, etc.)

Tufte, “Visual Display of Quantitative Information” (1983)

“Graphical Integrity”

“The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities measured.”

Tufte, “Visual Display of Quantitative Information” (1983)

NOW, ON CS 7250…

Design & Rules of Thumb

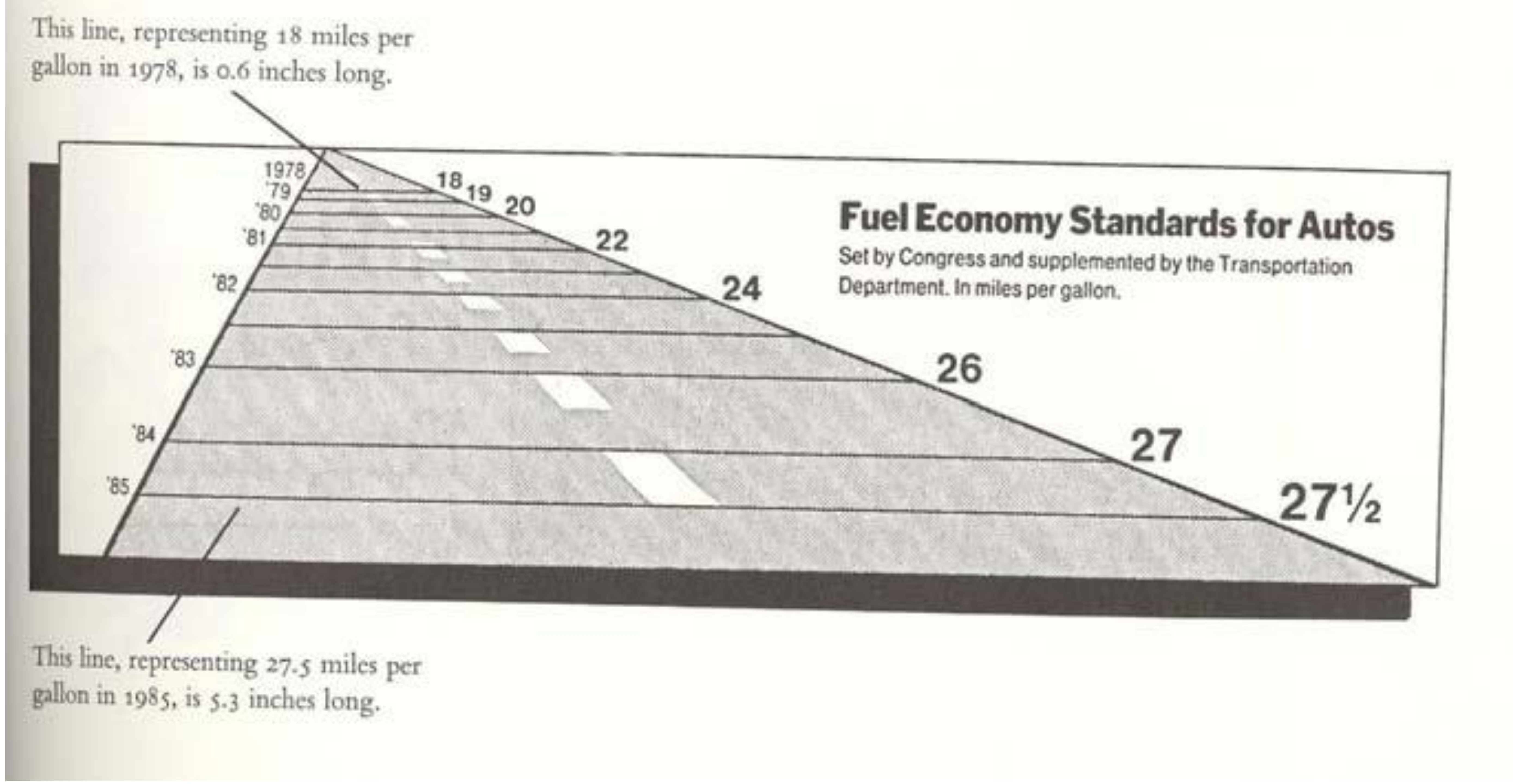
Lie Factor

Lie Factor = $\frac{\text{(Size of effect in graphic)}}{\text{(Size of effect in data)}}$

Lie Factor = >1, overstating

Lie Factor = 1, accurate :)

Lie Factor = <1, understating



"The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities measured."

Tufte, "Visual Display of Quantitative Information" (1983)

Lie Factor

Lie Factor = $\frac{\text{(Size of effect in graphic)}}{\text{(Size of effect in data)}}$

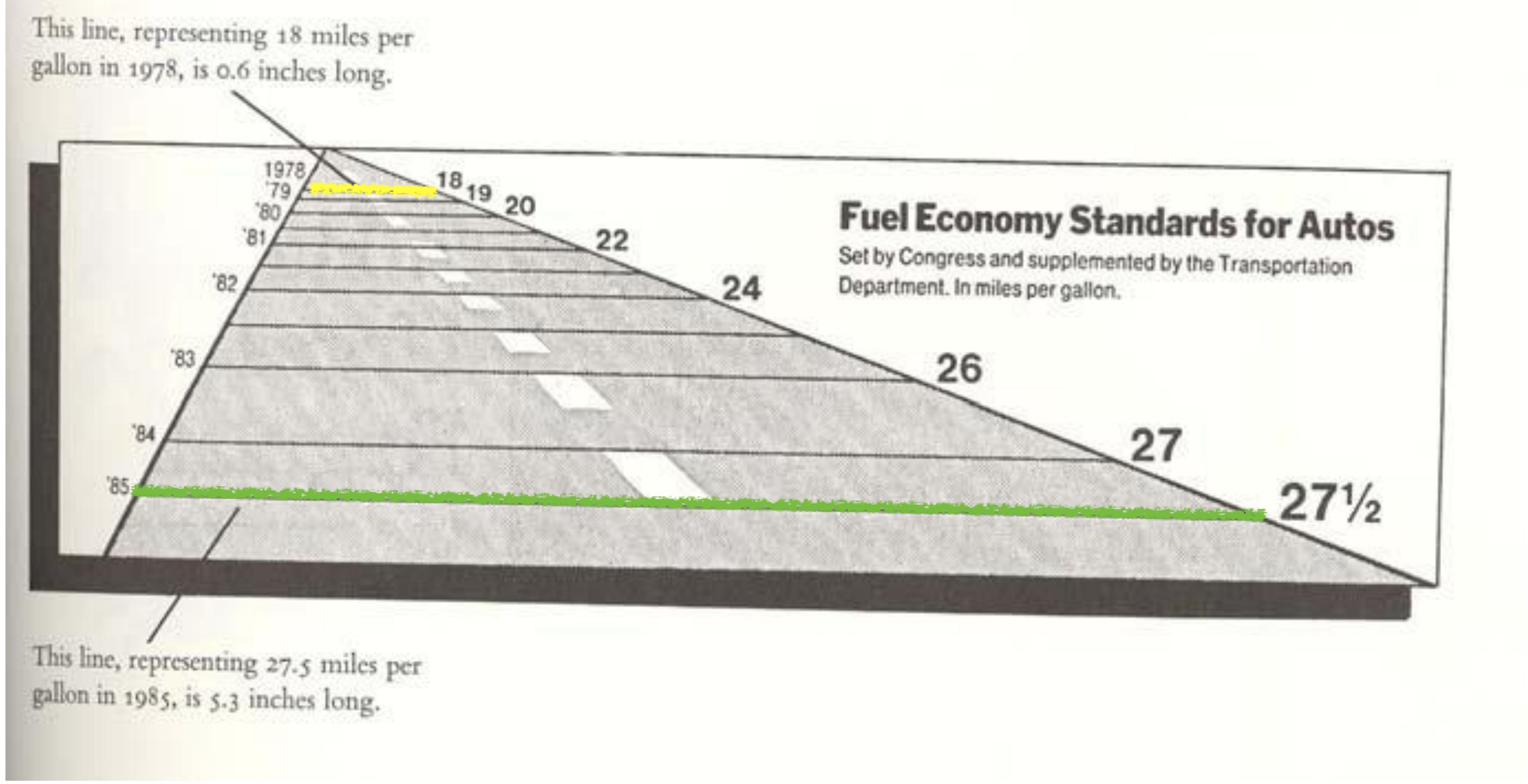


Image = $\frac{5.3'' - 0.6''}{0.6''}$ = 7.83 = 783%

Data = $\frac{27.5 - 18}{18}$ = 0.53 = 53%

Lie Factor = $\frac{783\%}{53\%}$ = 14.8

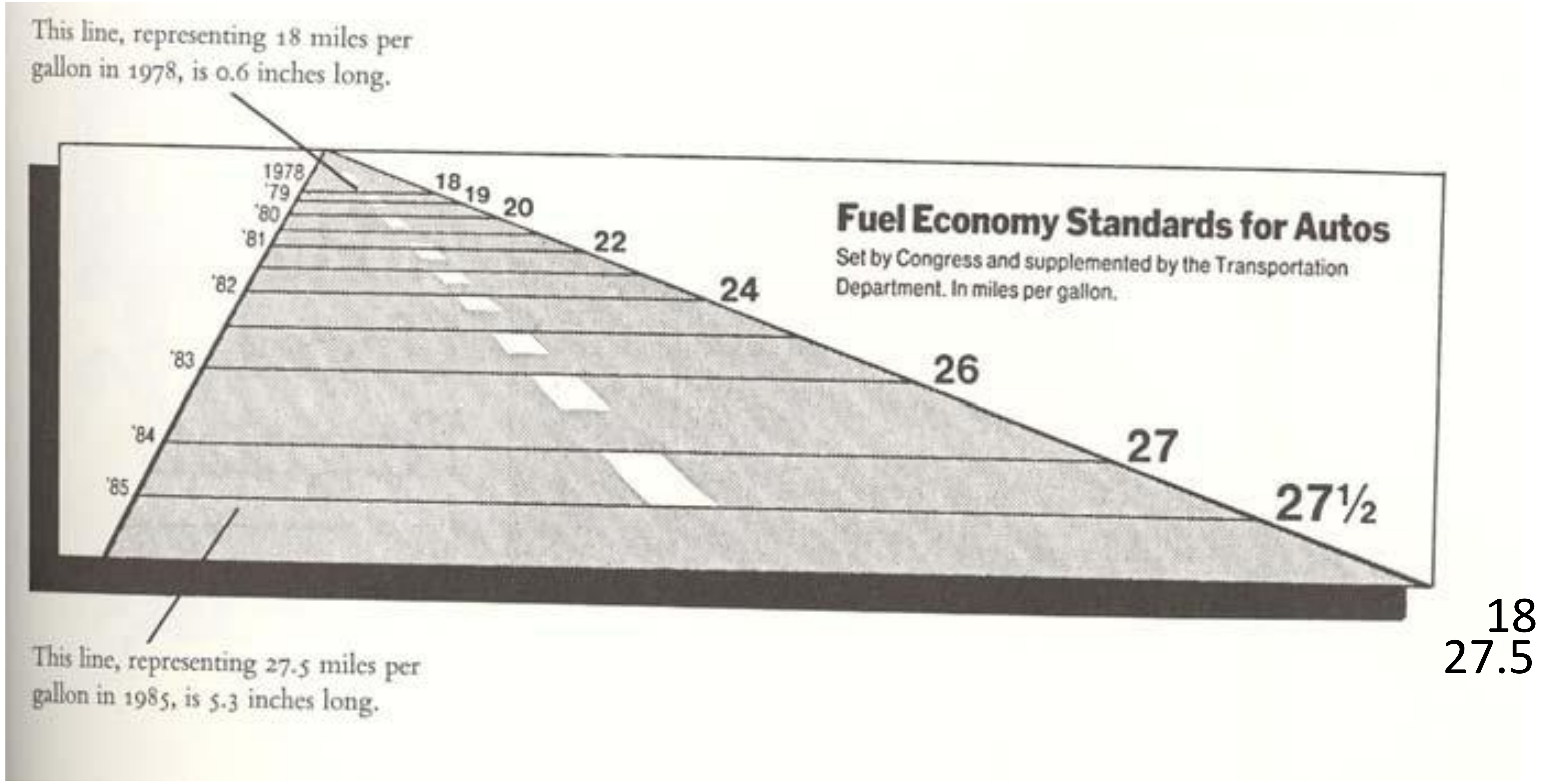
Lie Factor = >1, overstating

"The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities measured."

Tufte, "Visual Display of Quantitative Information" (1983)

Lie Factor

$$\text{Lie Factor} = \frac{\text{(Size of effect in graphic)}}{\text{(Size of effect in data)}}$$



$$\text{Image} = \frac{5.3'' - 0.6''}{0.6''} = 7.83 = 783\%$$

$$\text{Data} = \frac{27.5 - 18}{18} = 0.53 = 53\%$$

$$\text{Lie Factor} = \frac{783\%}{53\%} = 14.8$$

Lie Factor = >1, overstating

18

27.5

"The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities measured."

Tufte, "Visual Display of Quantitative Information" (1983)

IN-CLASS ACTIVITY:
Calculate for yourself!

Lie Factor

Data = $\frac{2 - 1}{1} = 1 = 100\%$

Lie Factor = $\frac{\text{(Size of effect in graphic)}}{\text{(Size of effect in data)}}$

Make sure *area* is proportional to data!

Don't use 3D bar charts!

Image = $\frac{2 - 1}{1} = 1 = 100\%$

Lie Factor = $\frac{100\%}{100\%} = 1$

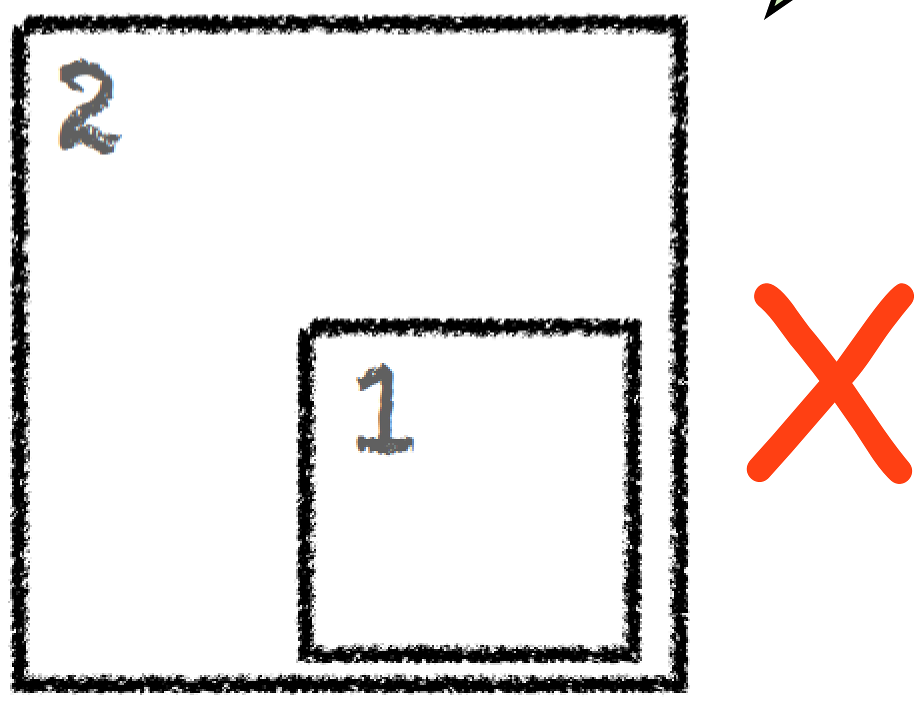


Image = $\frac{2^2 - 1^2}{1^2} = 3 = 300\%$

Lie Factor = $\frac{300\%}{100\%} = 3$

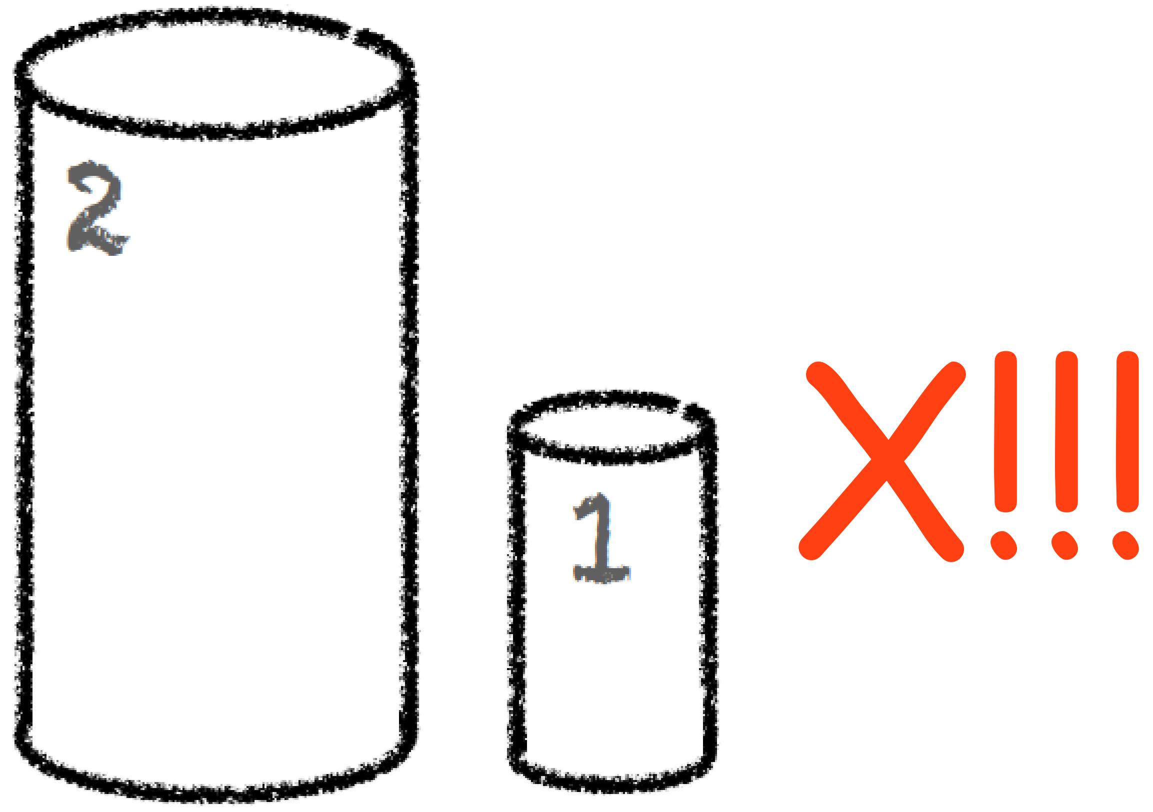


Image = $\frac{2*\pi 1^2 - 1*\pi 0.5^2}{1*\pi 0.5^2} = 7 = 700\%$

Lie Factor = $\frac{700\%}{100\%} = 7$

"The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities measured."

Tufte, "Visual Display of Quantitative Information" (1983)

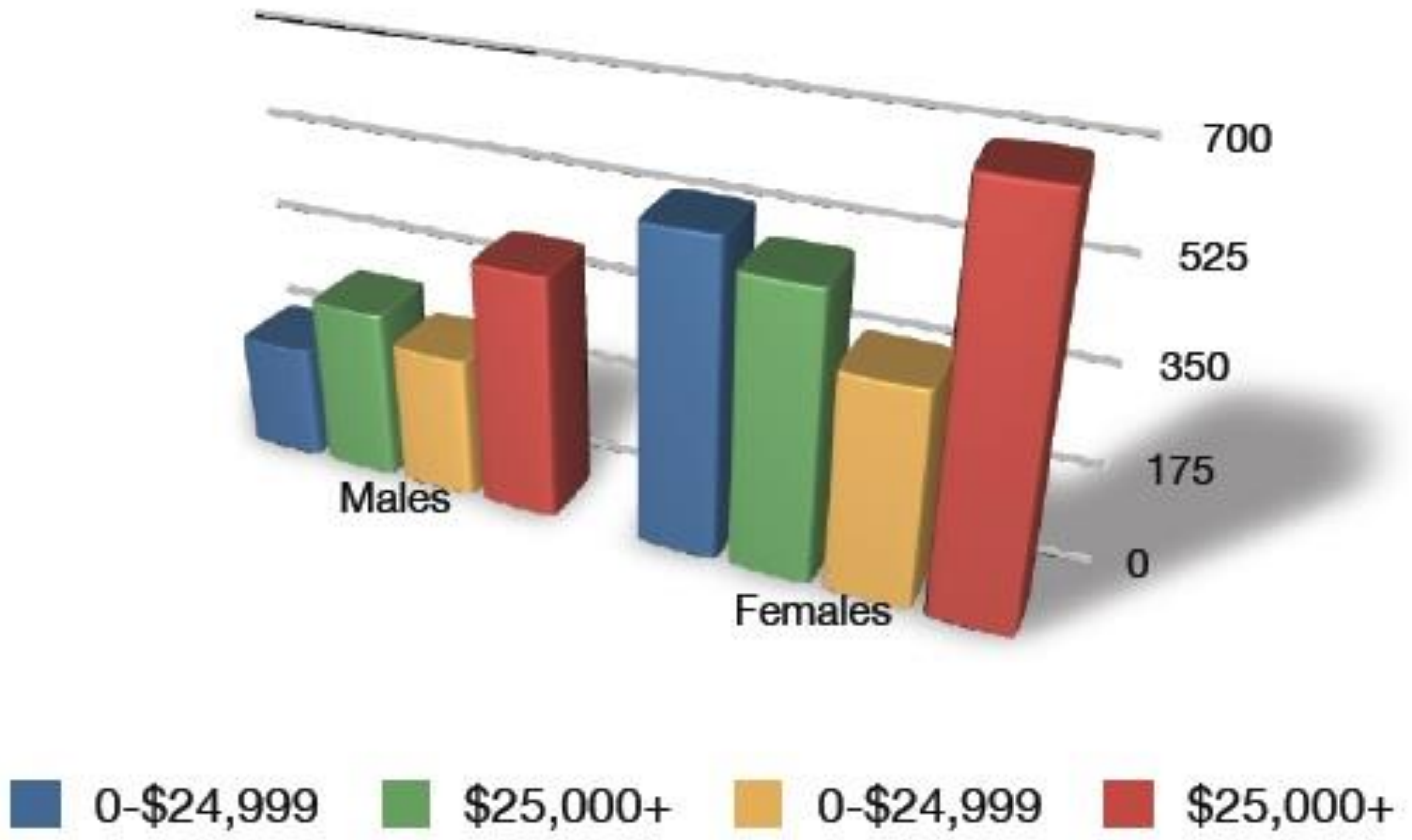
“Graphical Integrity”

Data Ink = the ink used to show data

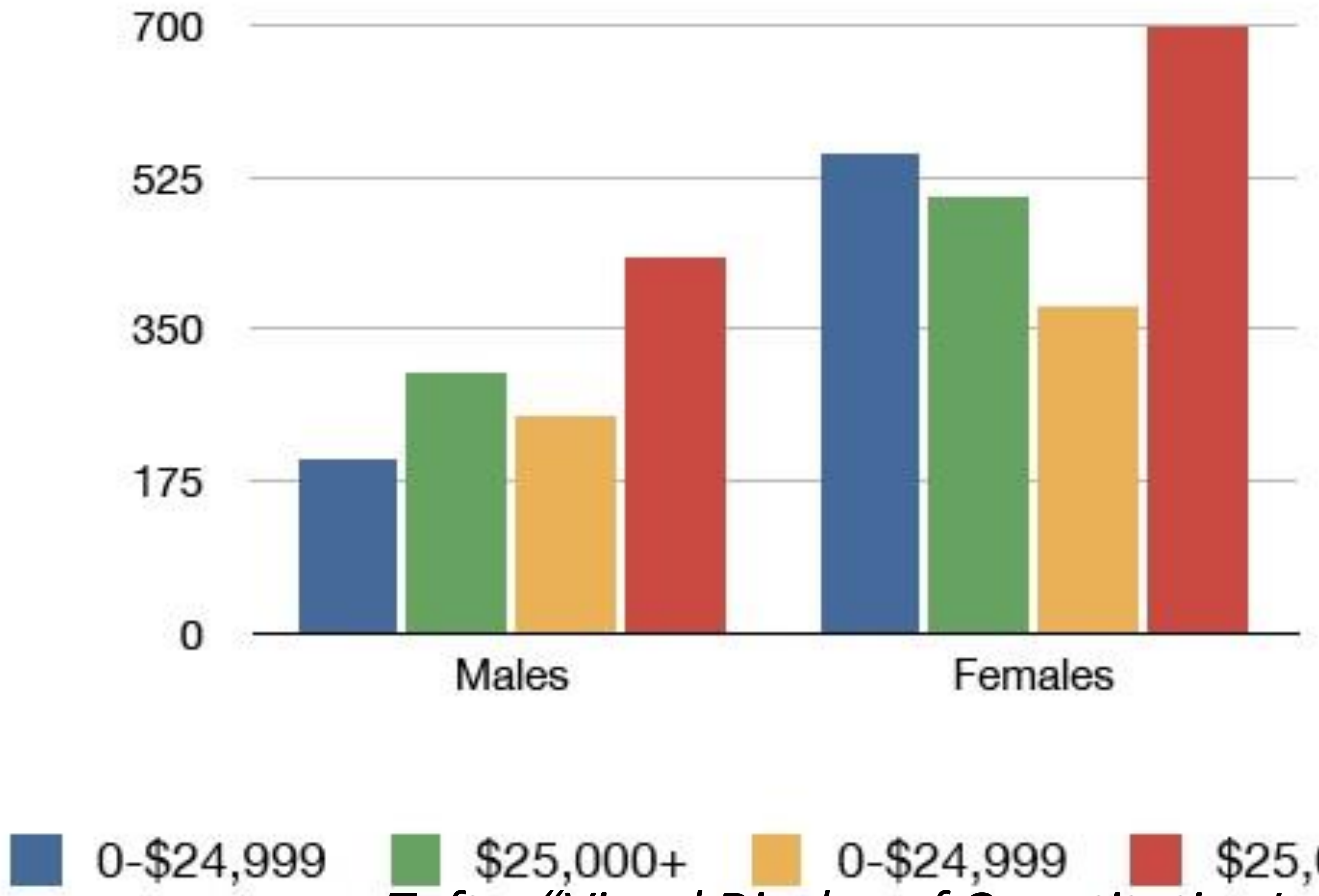
Data Ink Ratio = $\dfrac{\text{data-ink}}{\text{total ink in graphic}}$

Tufte: maximize the data ink ratio

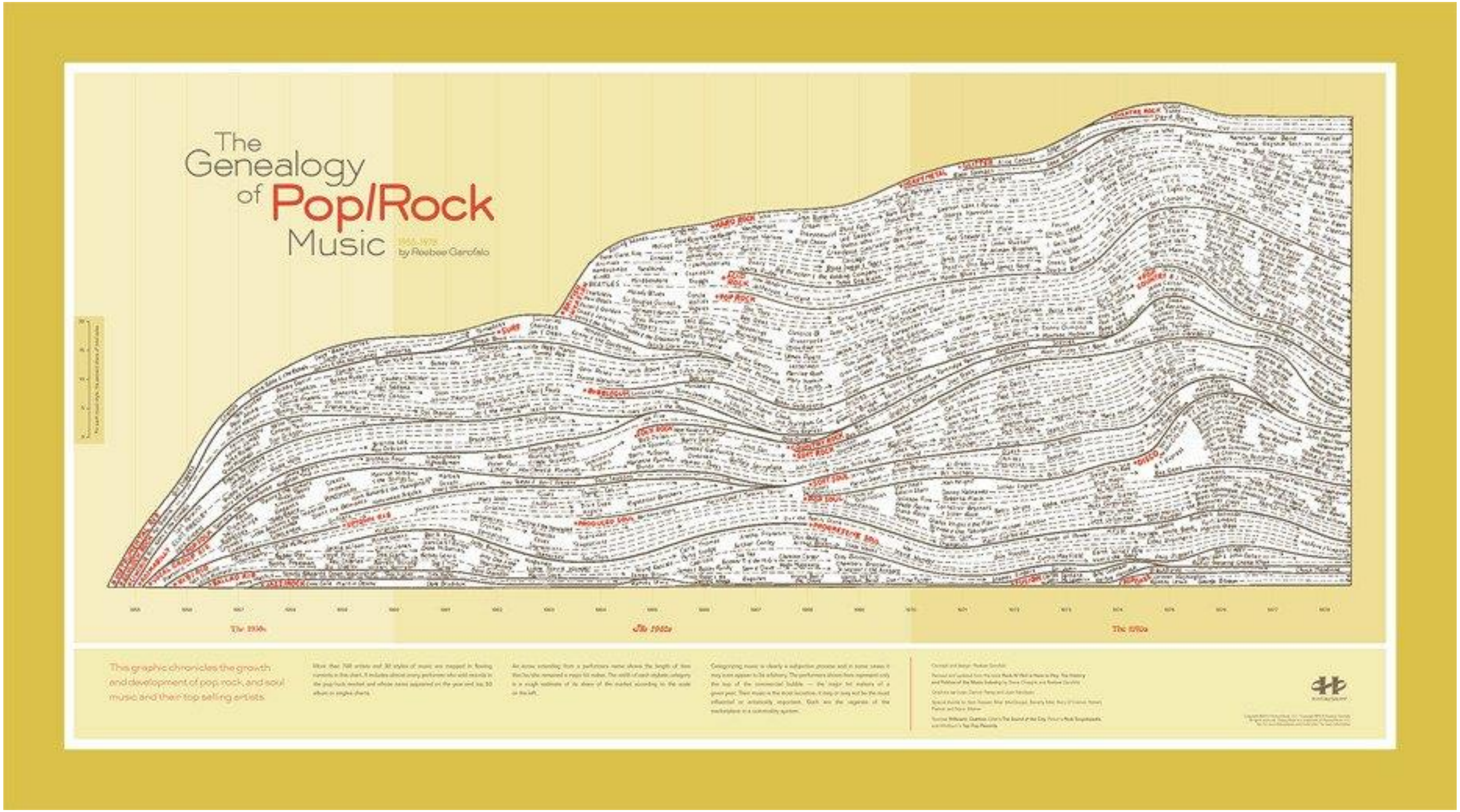
Low Data Ink Ratio



High Data Ink Ratio



Tufte, “Visual Display of Quantitative Information” (1983)



High Data Ink Ratio

[Reebee Garofalo, Genealogy of Pop/Rock Music]

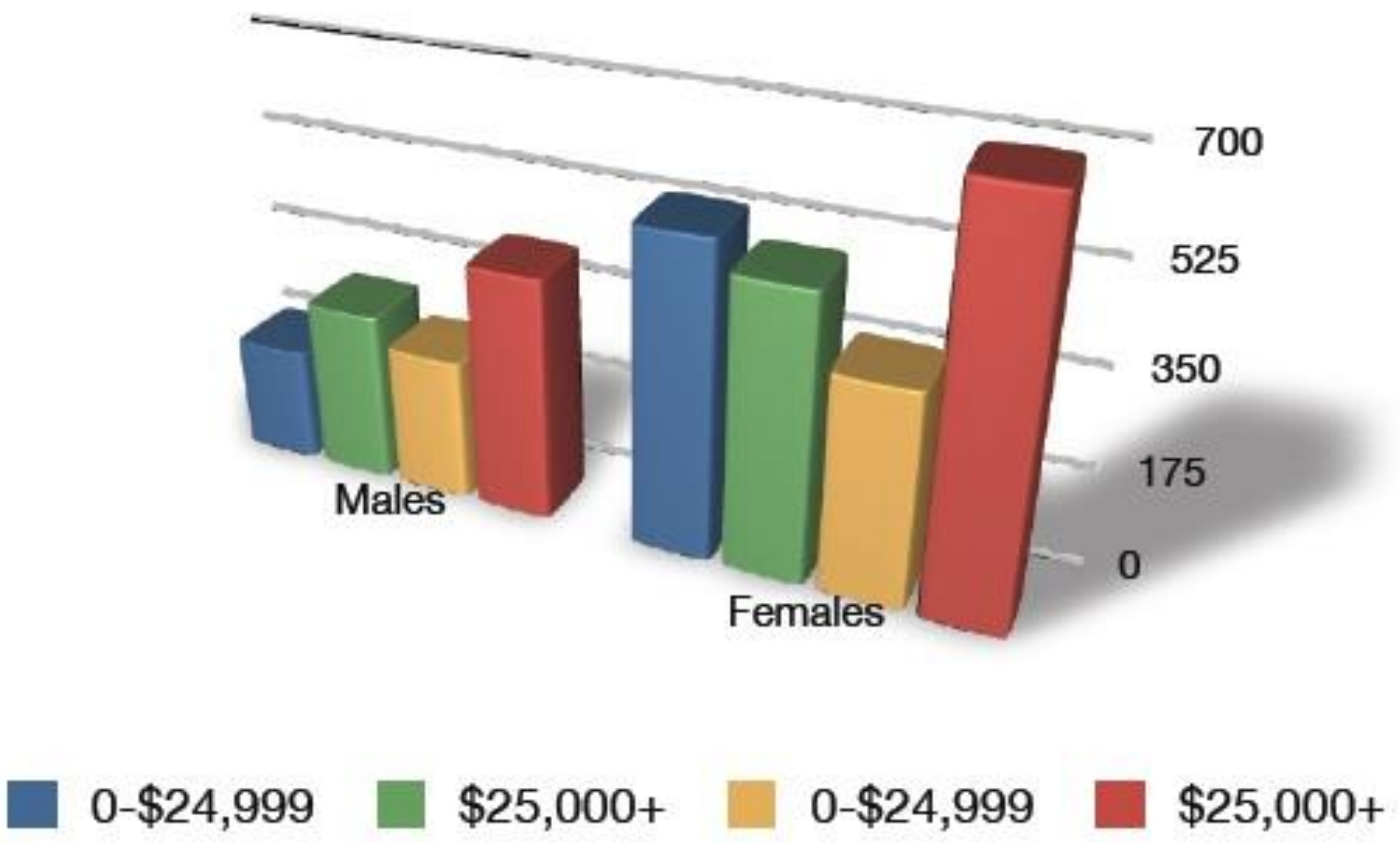
“Graphical Integrity”

“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

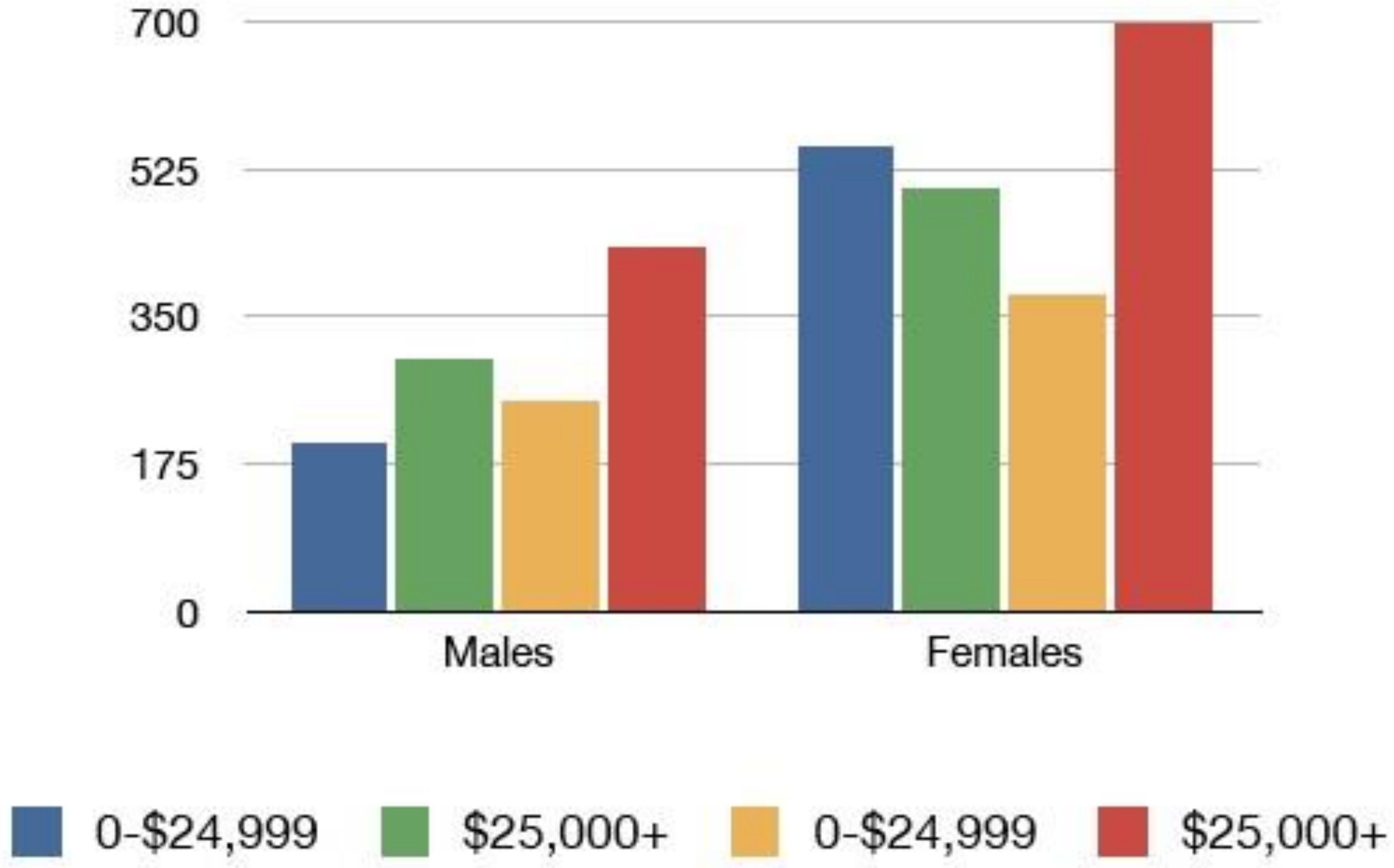
Tufte, “Visual Display of Quantitative Information” (1983)

“No Unjustified 3D”

Dimensions in data: 3
Dimensions in plot: 4

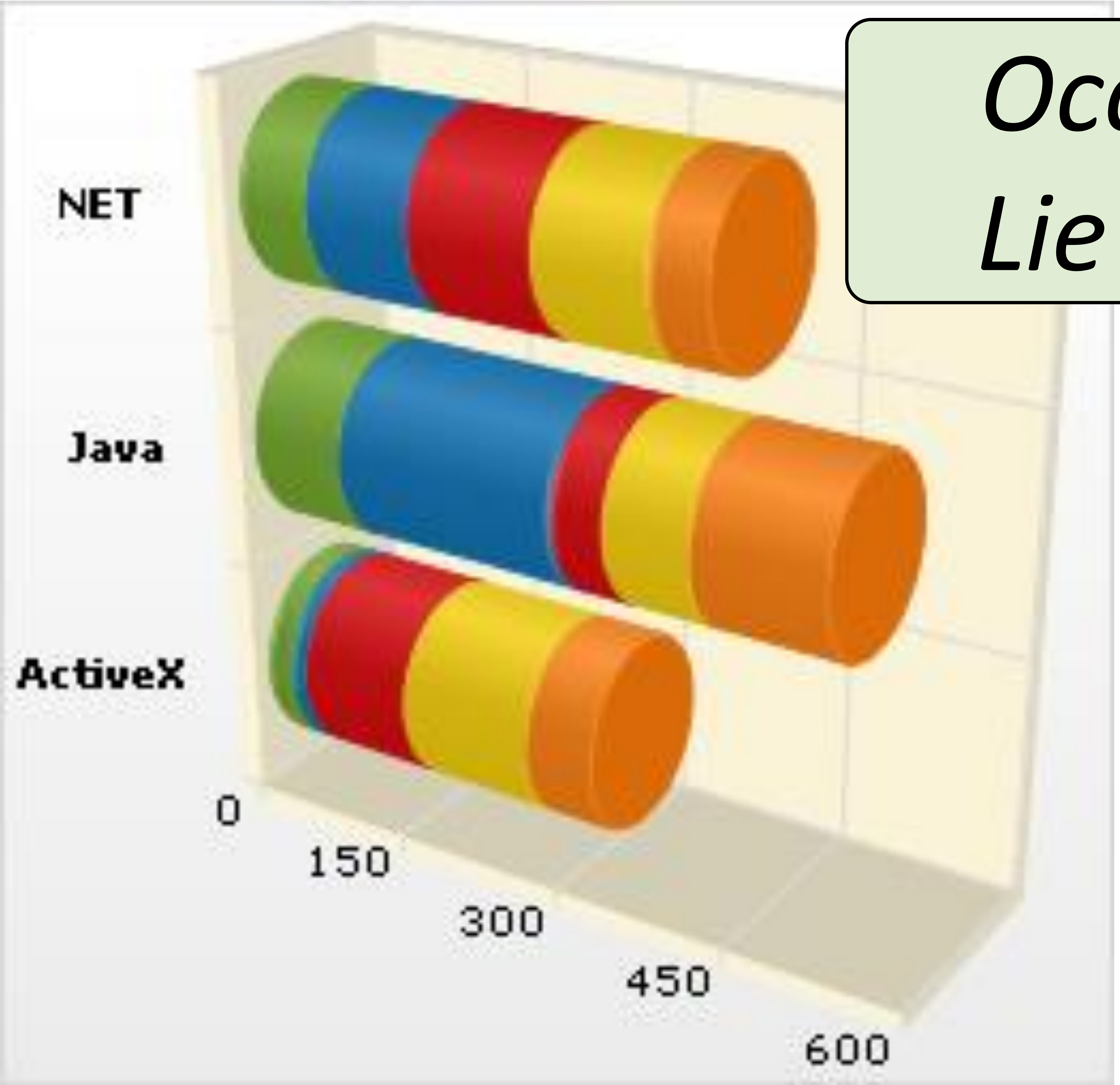


Dimensions in data: 3
Dimensions in plot: 3

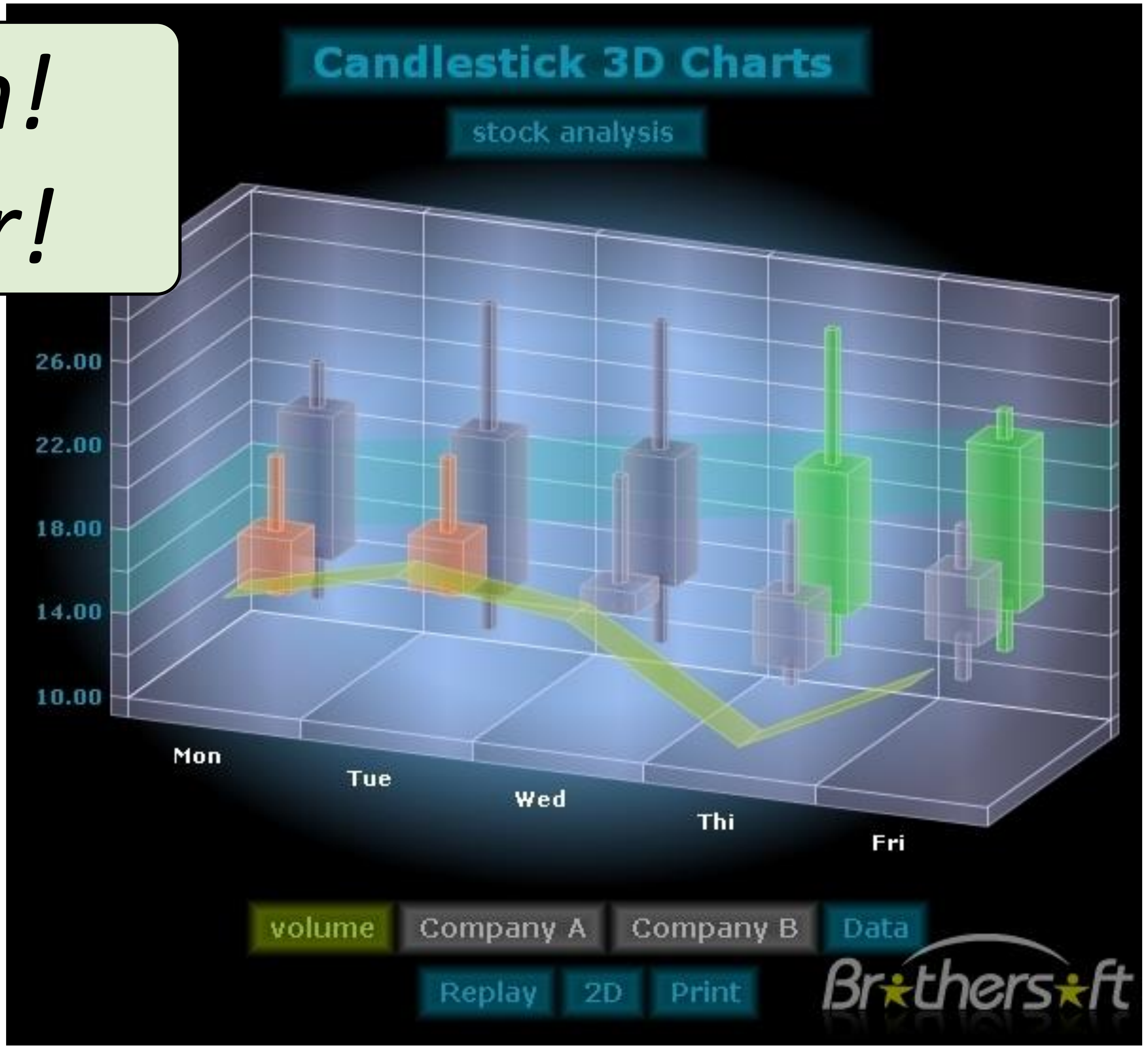


“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

“No Unjustified 3D”



Occlusion!
Lie Factor!



http://help.infragistics.com/Help/Doc/WinForms/2014.2/CLR4.0/html/Images/Chart_Bar_Chart_03.png

http://img.brothersoft.com/screenshots/softimage/0/3d_charts-171418-1269568478.jpeg

“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

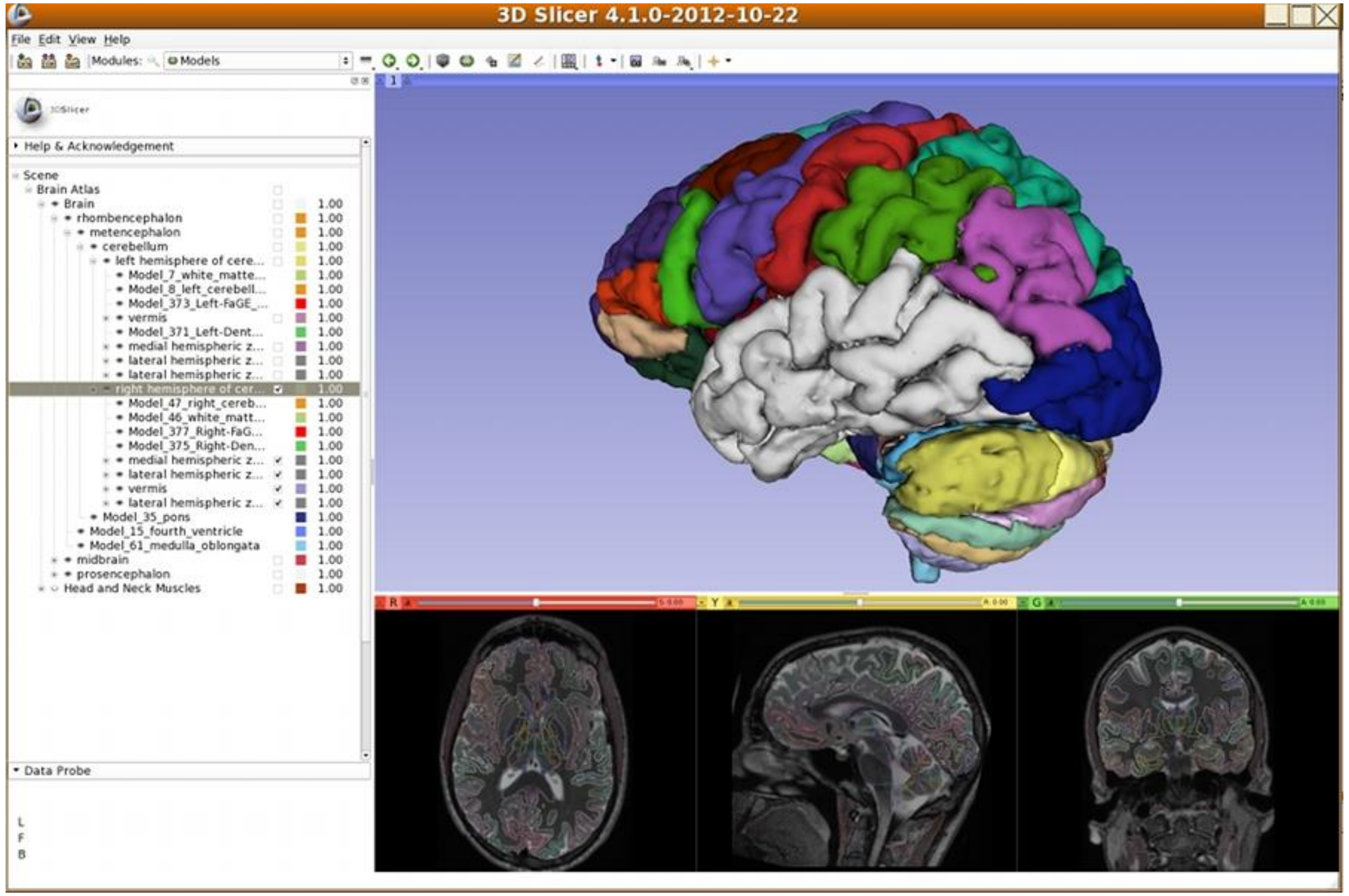
“No Unjustified 3D”

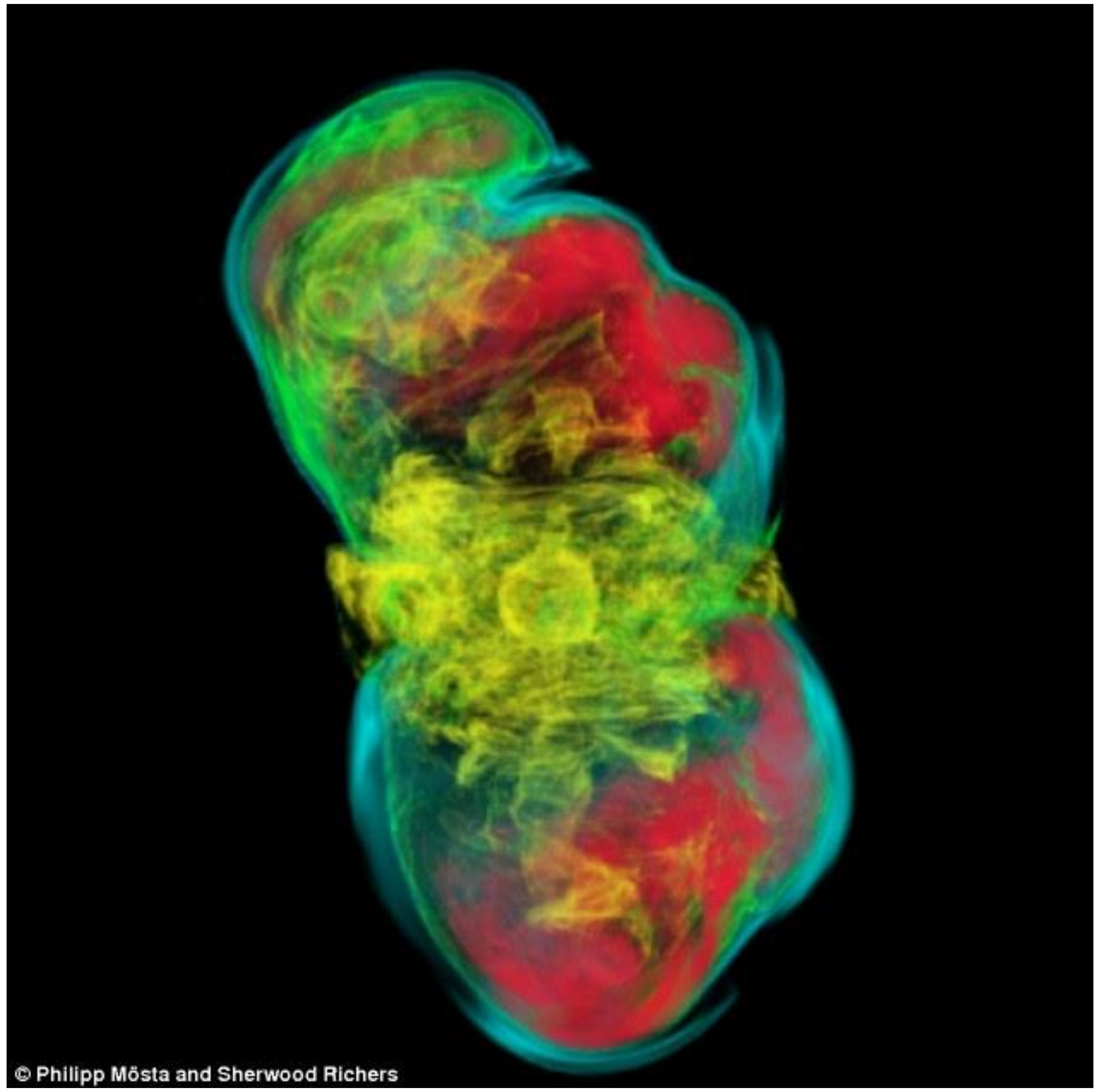
Unjustified 3D!

Lie factor!

http://stats.stackexchange.com/questions/109076/what-is-your-favorite-statistical-graph/109080

“No Unjustified 3D”



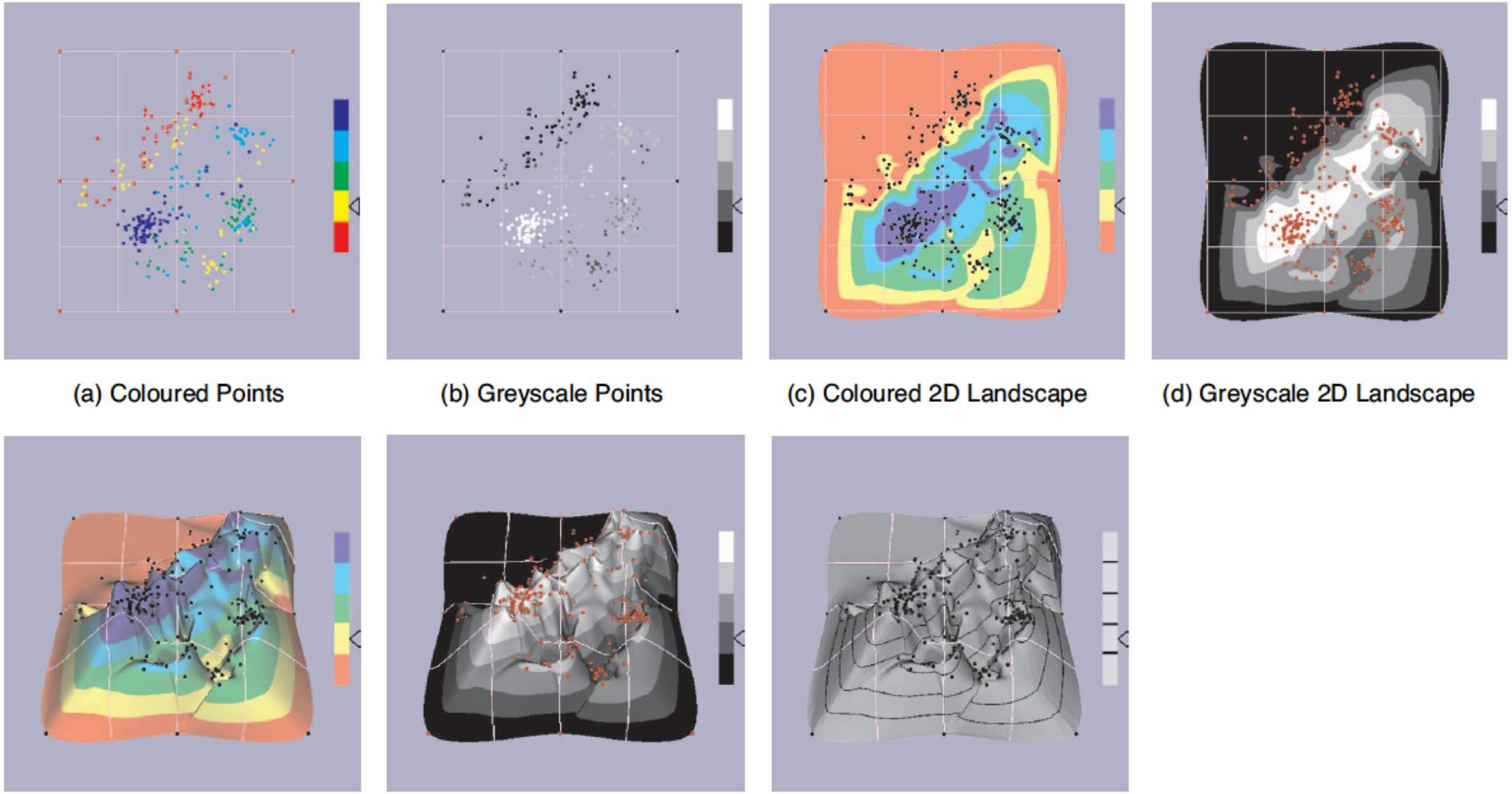


“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

“No Unjustified 3D”

This is not just a design principle, it has lots of experimental and quantitative data to back it up!

“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”



(a) Coloured Points (b) Greyscale Points (c) Coloured 2D Landscape (d) Greyscale 2D Landscape

(e) Coloured 3D Landscape (f) Greyscale 3D Landscape (g) Height-only

Fig. 1 Point-based displays and information landscapes used in our experiment. All displays show the same data.

Dr. David Sprague (Former Lecturer, Khoury)

"The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data."

[Tory, et al. (2007)]

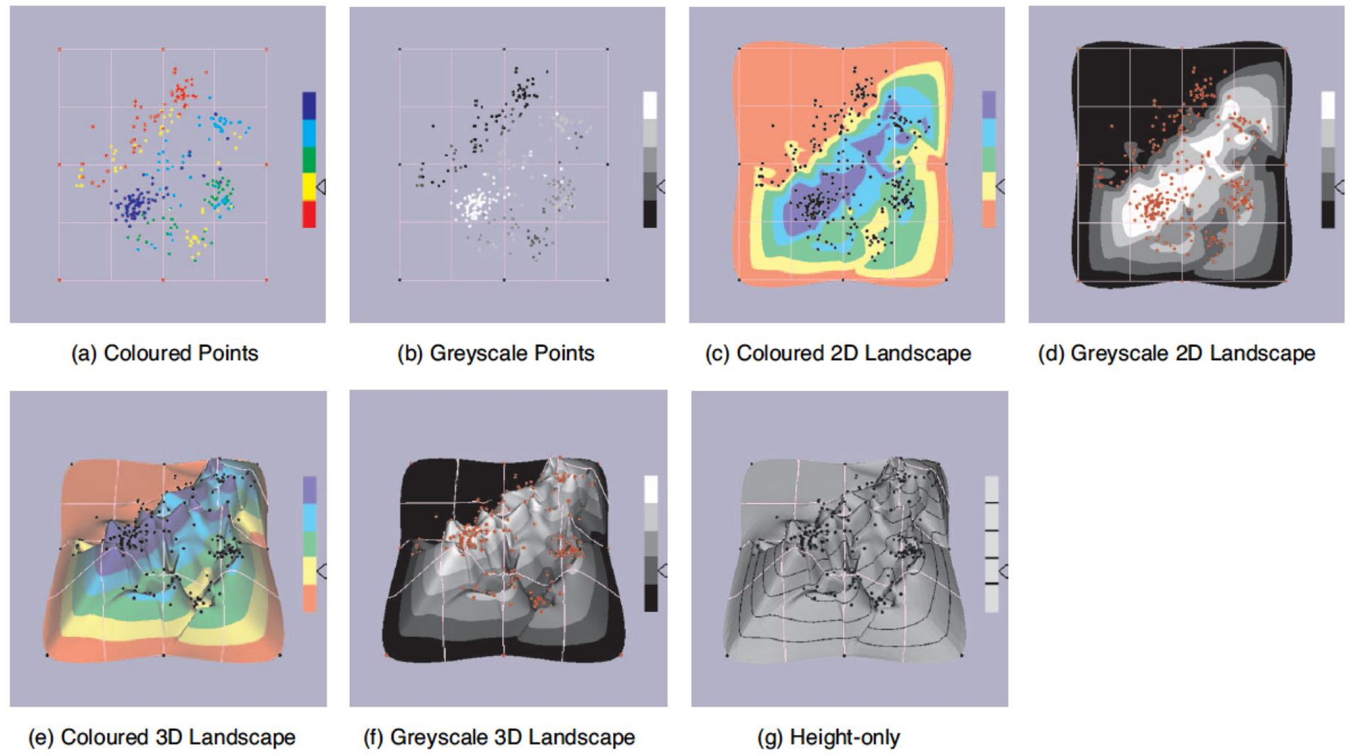
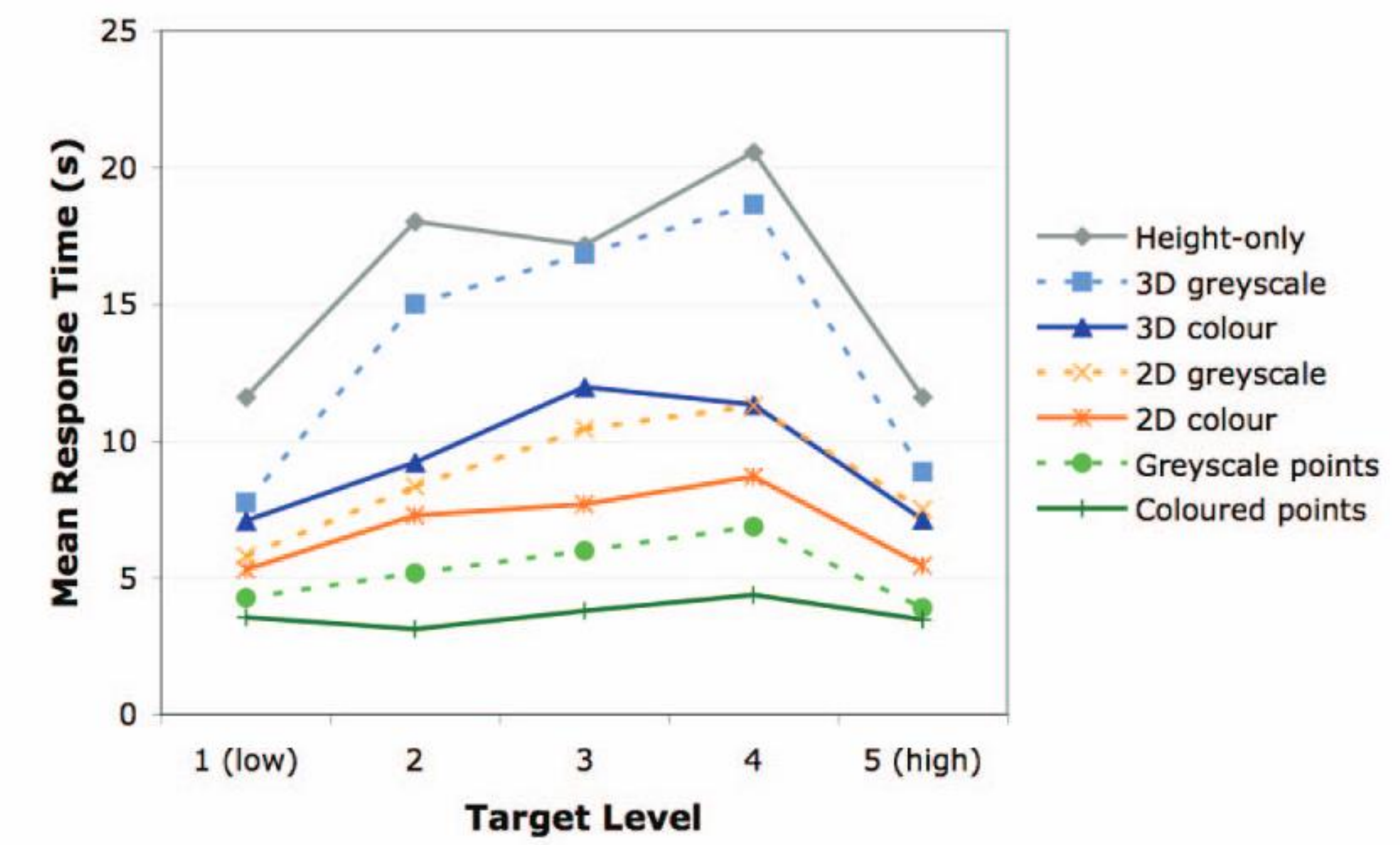
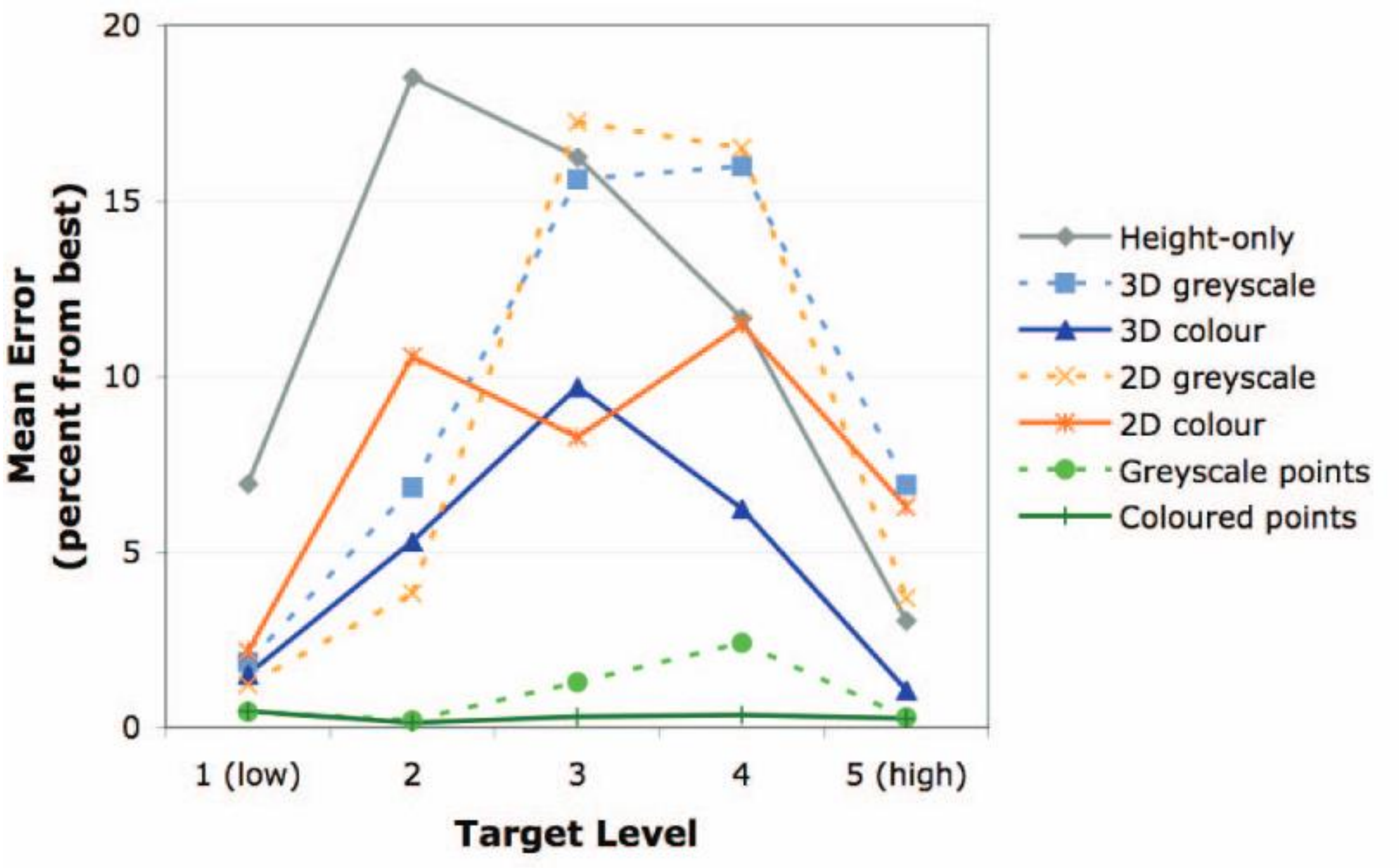


Fig. 1 Point-based displays and information landscapes used in our experiment. All displays show the same data.

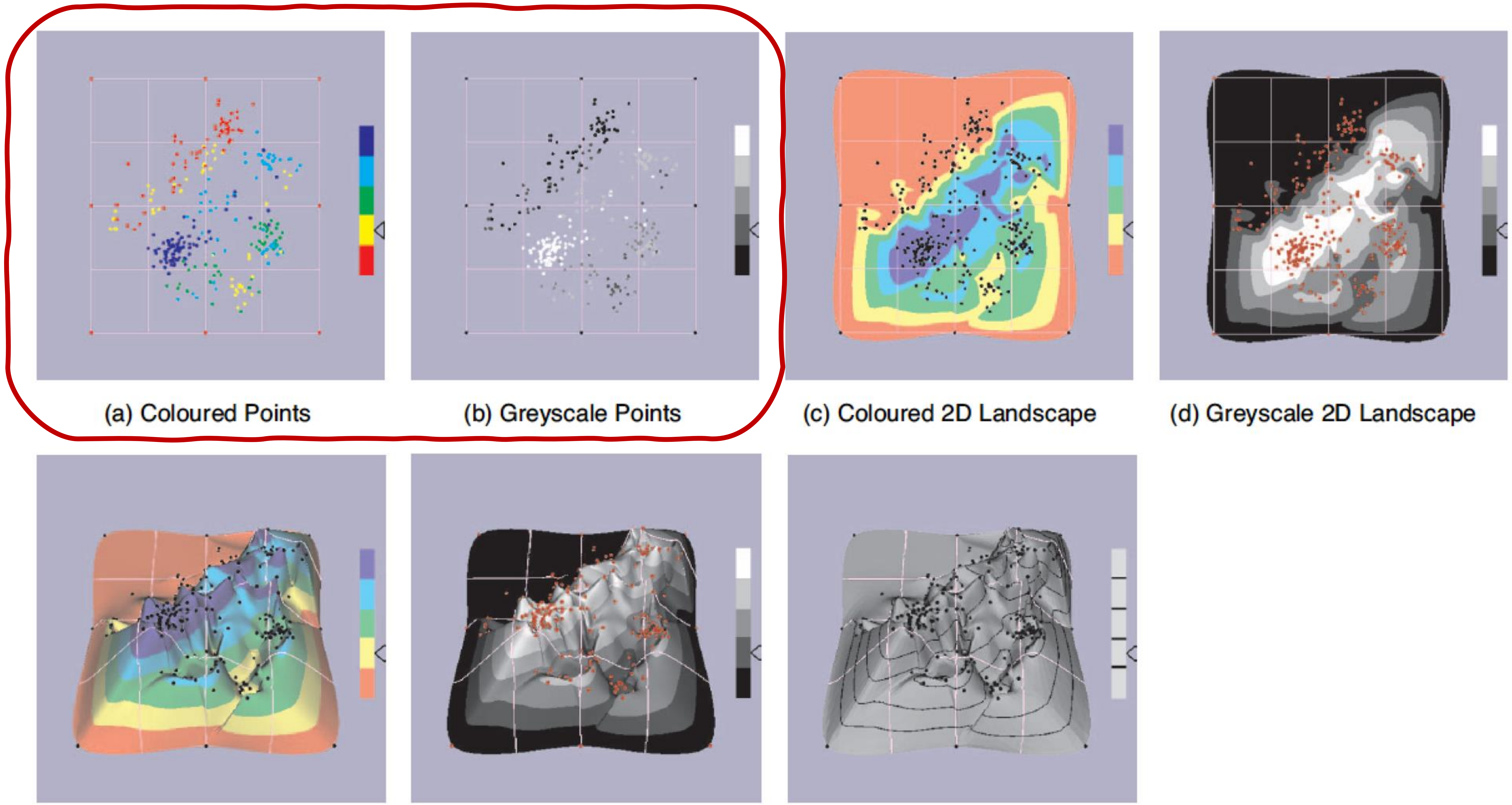
“Which spatial area contained the most points of a specified target value range?”





“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

Tory, et al. (2007)



(a) Coloured Points (b) Greyscale Points (c) Coloured 2D Landscape (d) Greyscale 2D Landscape

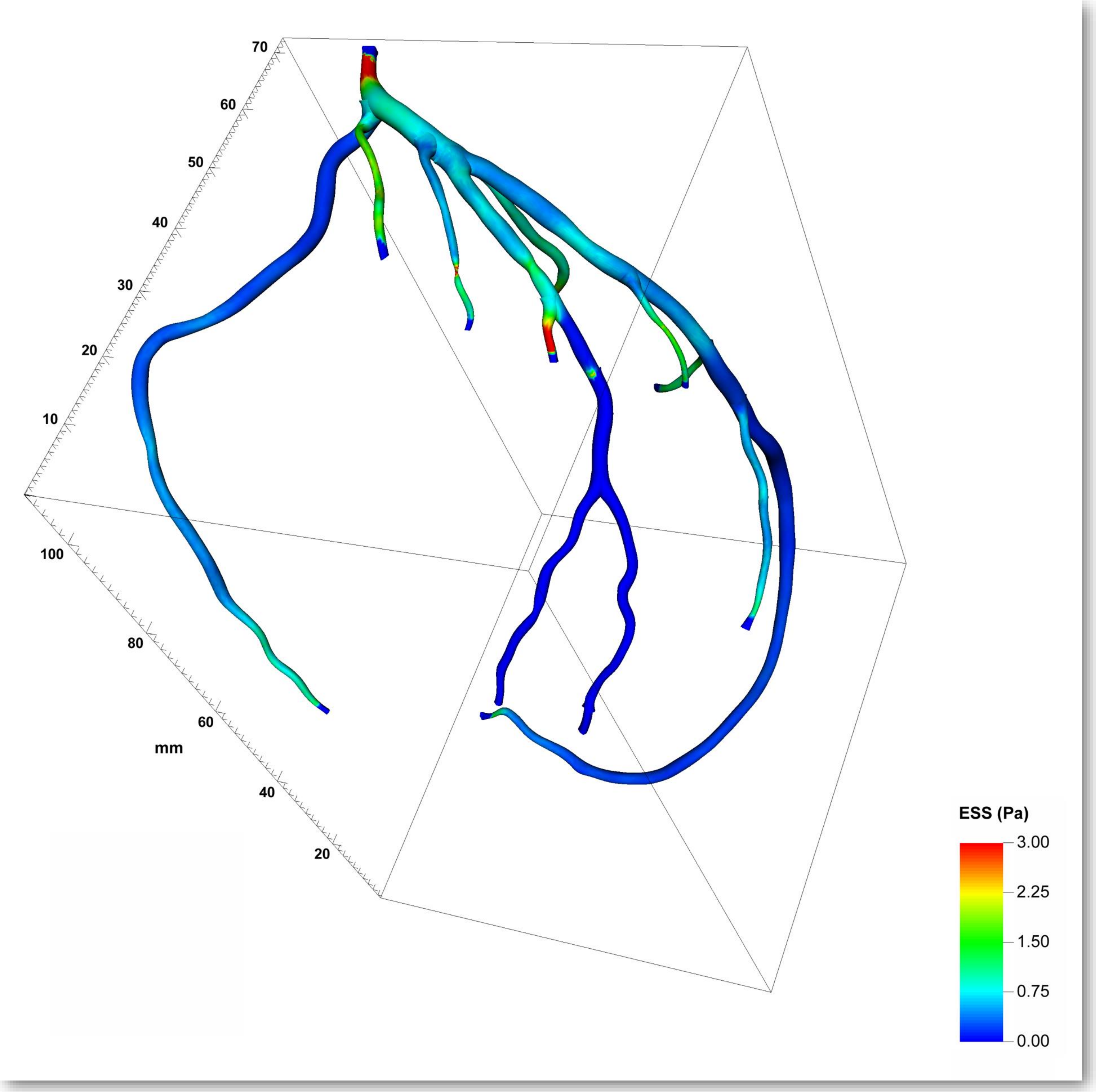
(e) Coloured 3D Landscape (f) Greyscale 3D Landscape (g) Height-only

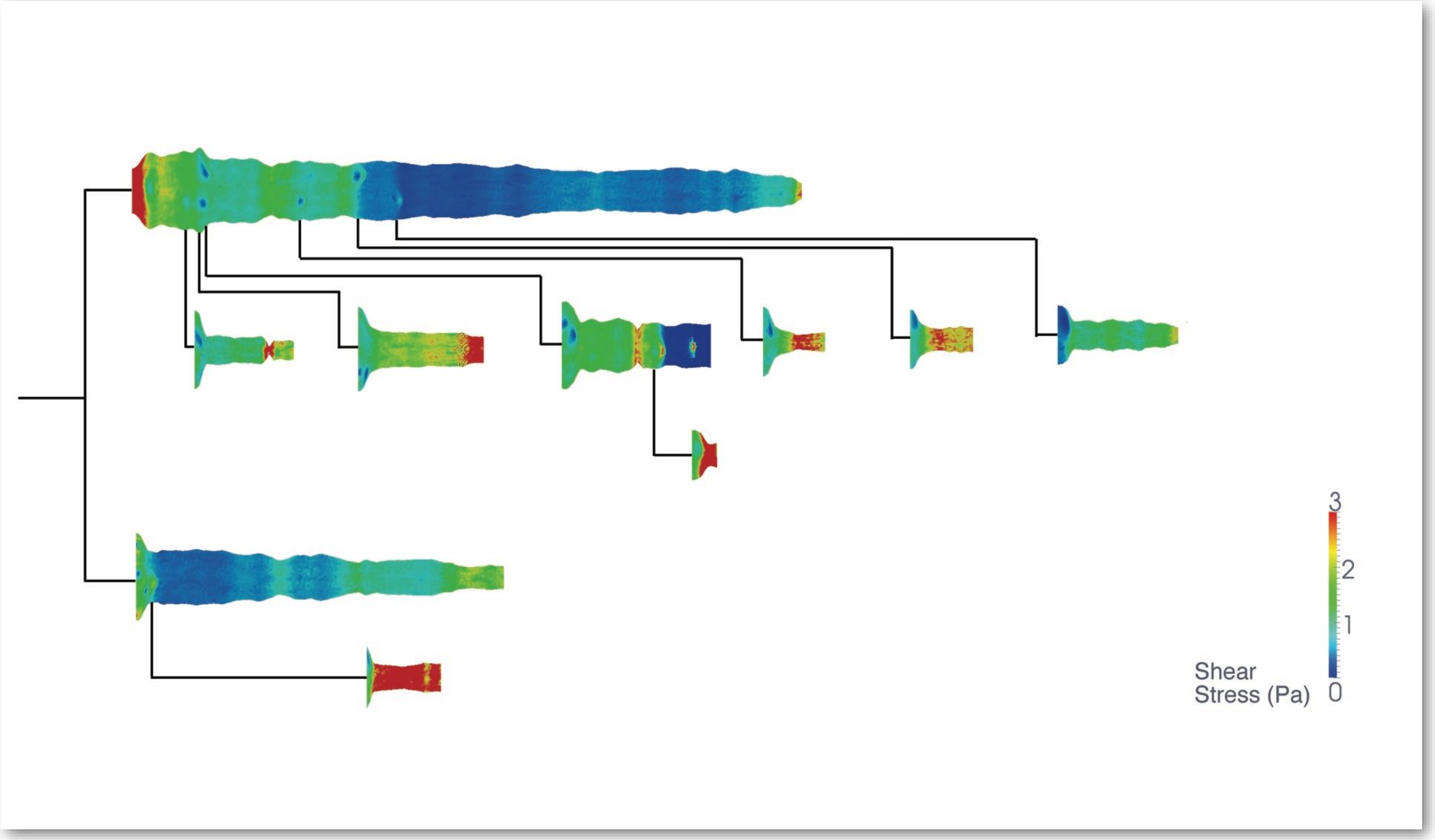
Fig. 1 Point-based displays and information landscapes used in our experiment. All displays show the same data.

"The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data."

Tory, et al. (2007)

“No Unjustified 3D”





“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

Borkin, et al. (2011)

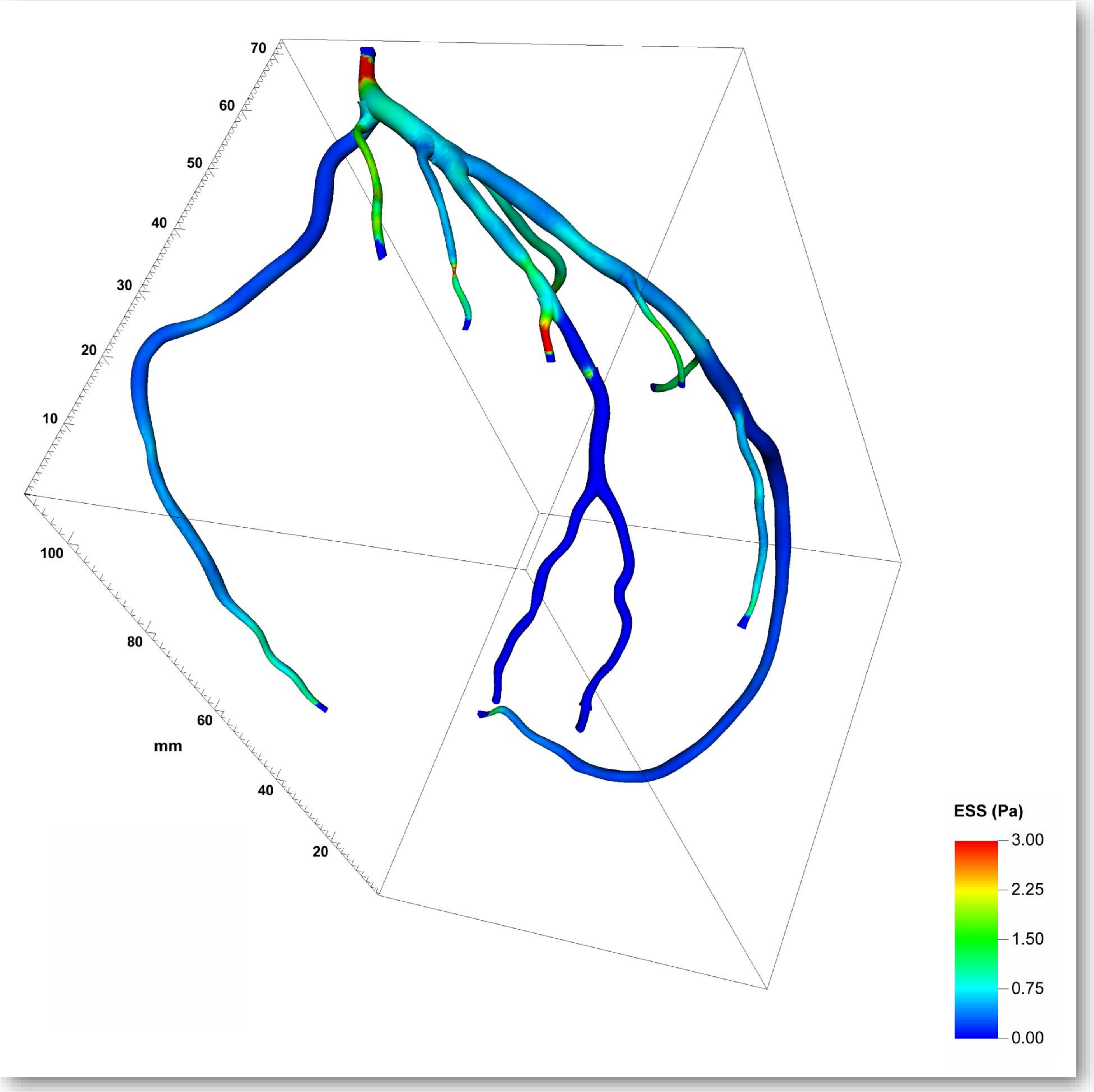
Accuracy

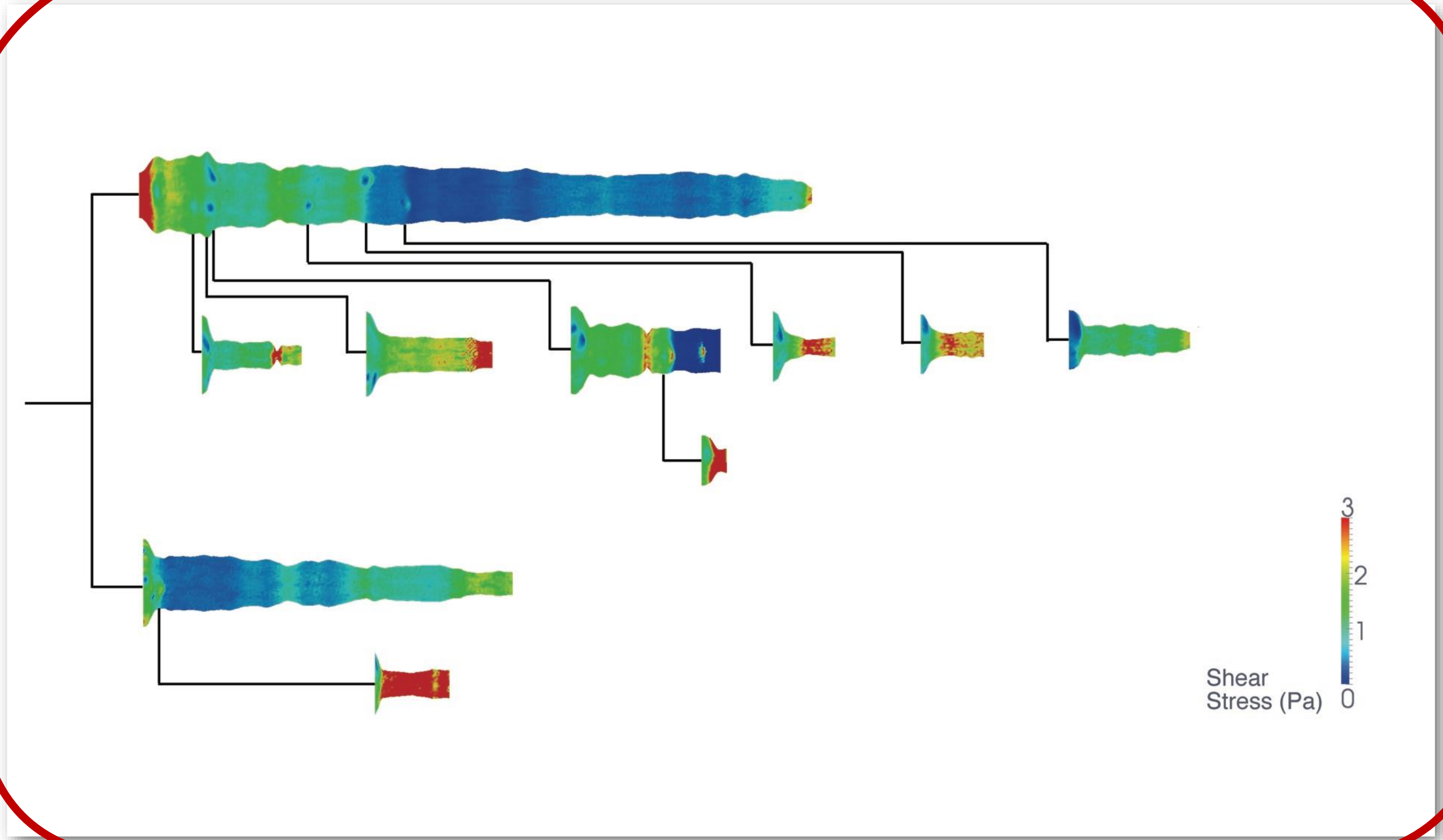
Strong effect of **dimensionality** on accuracy

39%

How many diseased regions found?

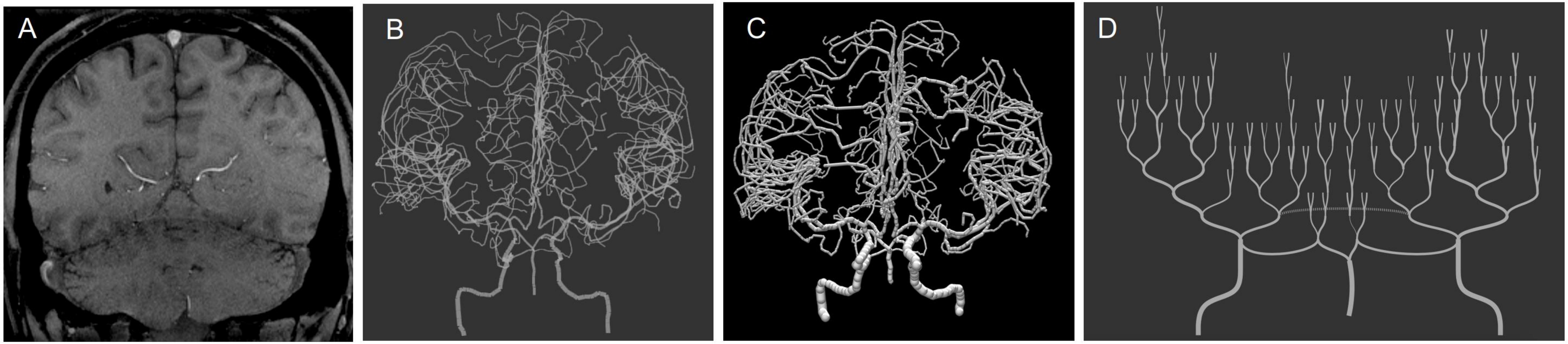
62%





[Borkin, et al. (2011)]

“No Unjustified 3D”

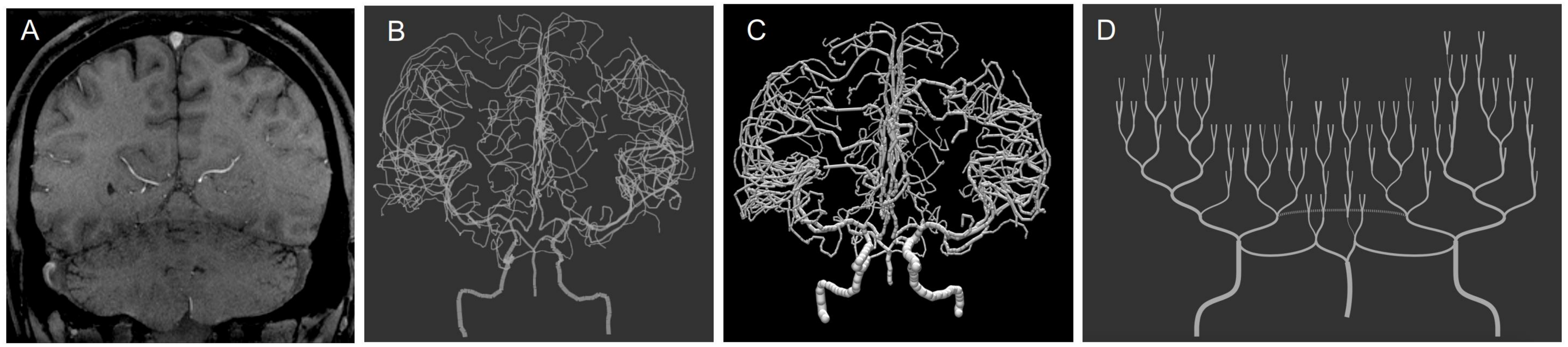


Data Source | **MIPS: 2D projection of 3D Arteries** | **3D Rendering** | **CerebroVis: Novel Network Visualization**

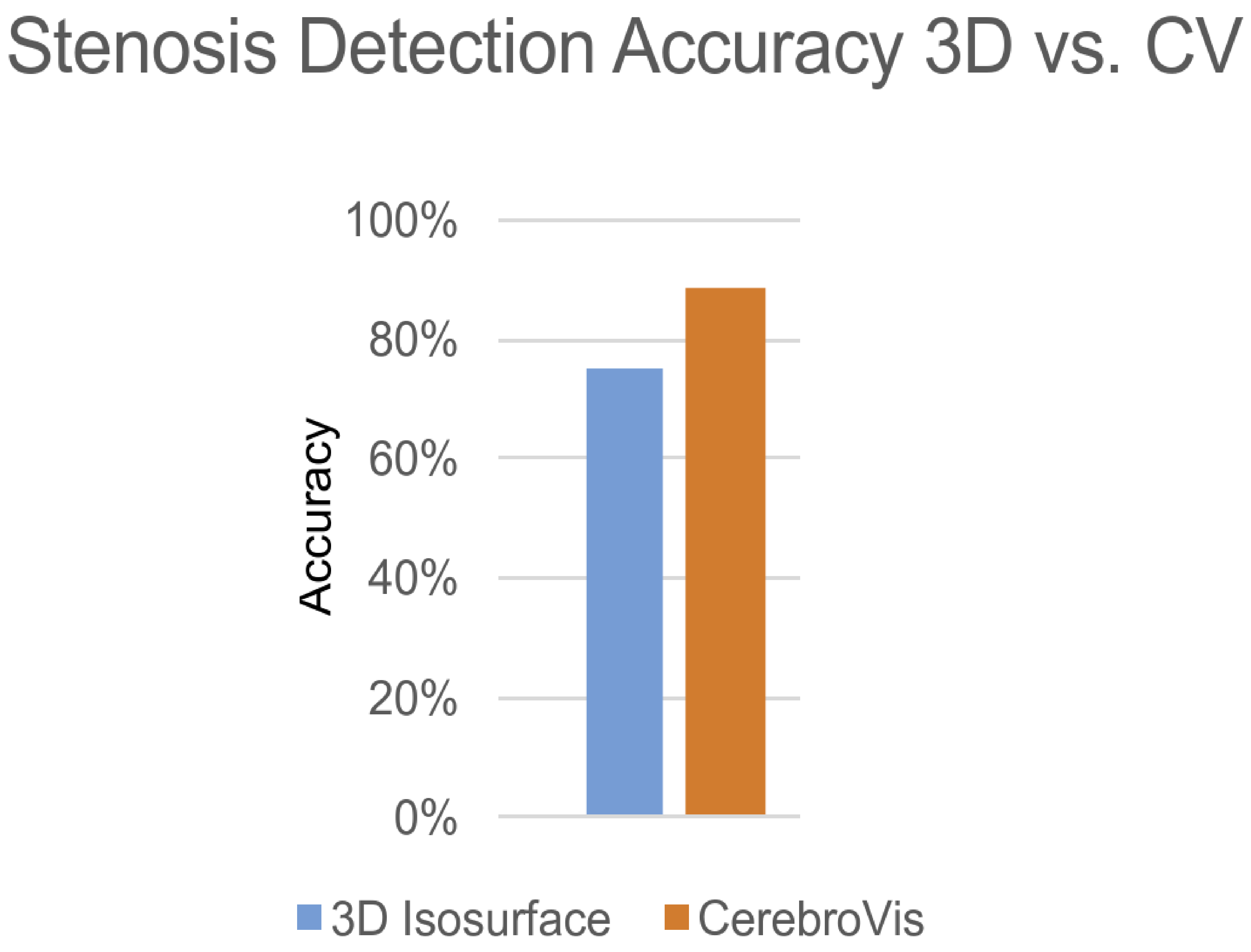
“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

[Pandey et al. VIS 2019]

“No Unjustified 3D”



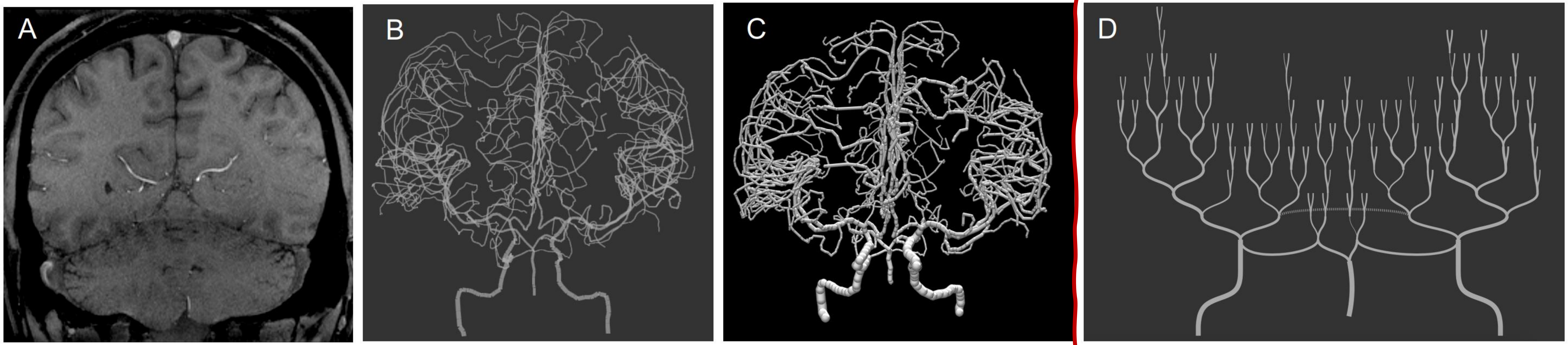
Data Source | **MIPS: 2D projection of 3D Arteries** | **3D Rendering** | **CerebroVis: Novel Network Visualization**



“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

[Pandey et al. VIS 2019]

“No Unjustified 3D”



Data Source | **MIPS: 2D projection of 3D Arteries** | **3D Rendering** | **CerebroVis: Novel Network Visualization**

“The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.”

[Pandey et al. VIS 2019]