

# Multidimensional Data

Steven Braun

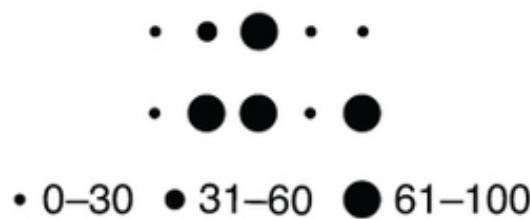
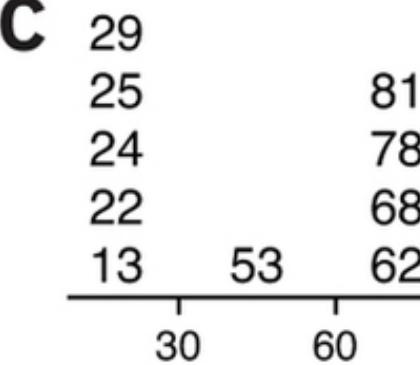
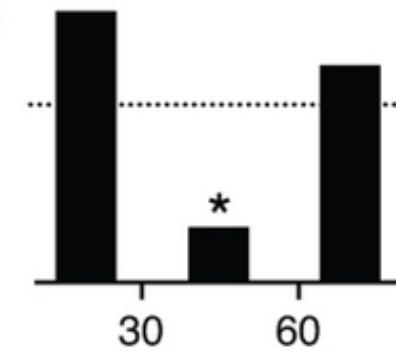
*Data Analytics and Visualization Specialist*

May 9, 2017

**Data are manifold**, and our choices of representation have a direct impact on their interpretation and use

**a**

13 53 81 29 25  
22 68 62 24 78

**b****c****d**

# Working with Multidimensional Data

What does multidimensionality look like in your data and research?

What kinds of problems do you encounter when working with multidimensional data?

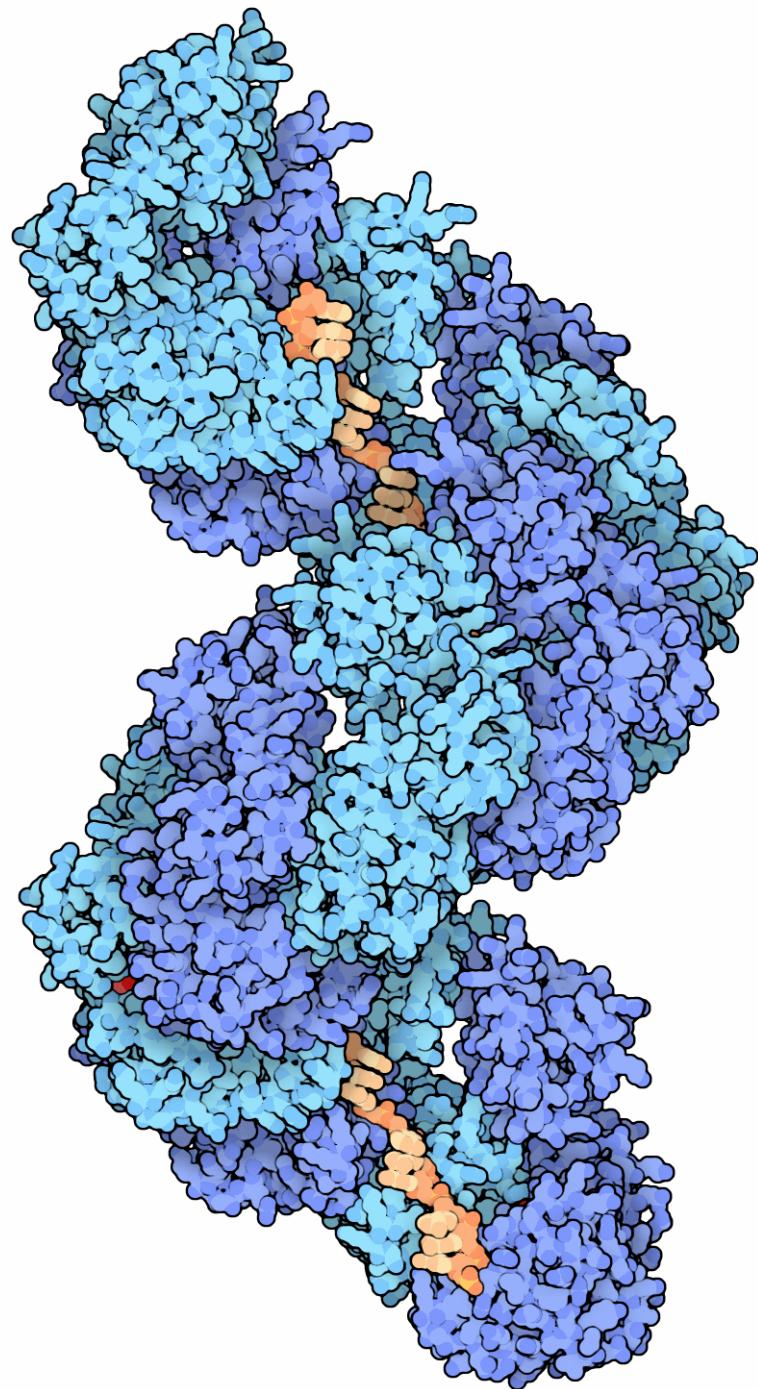
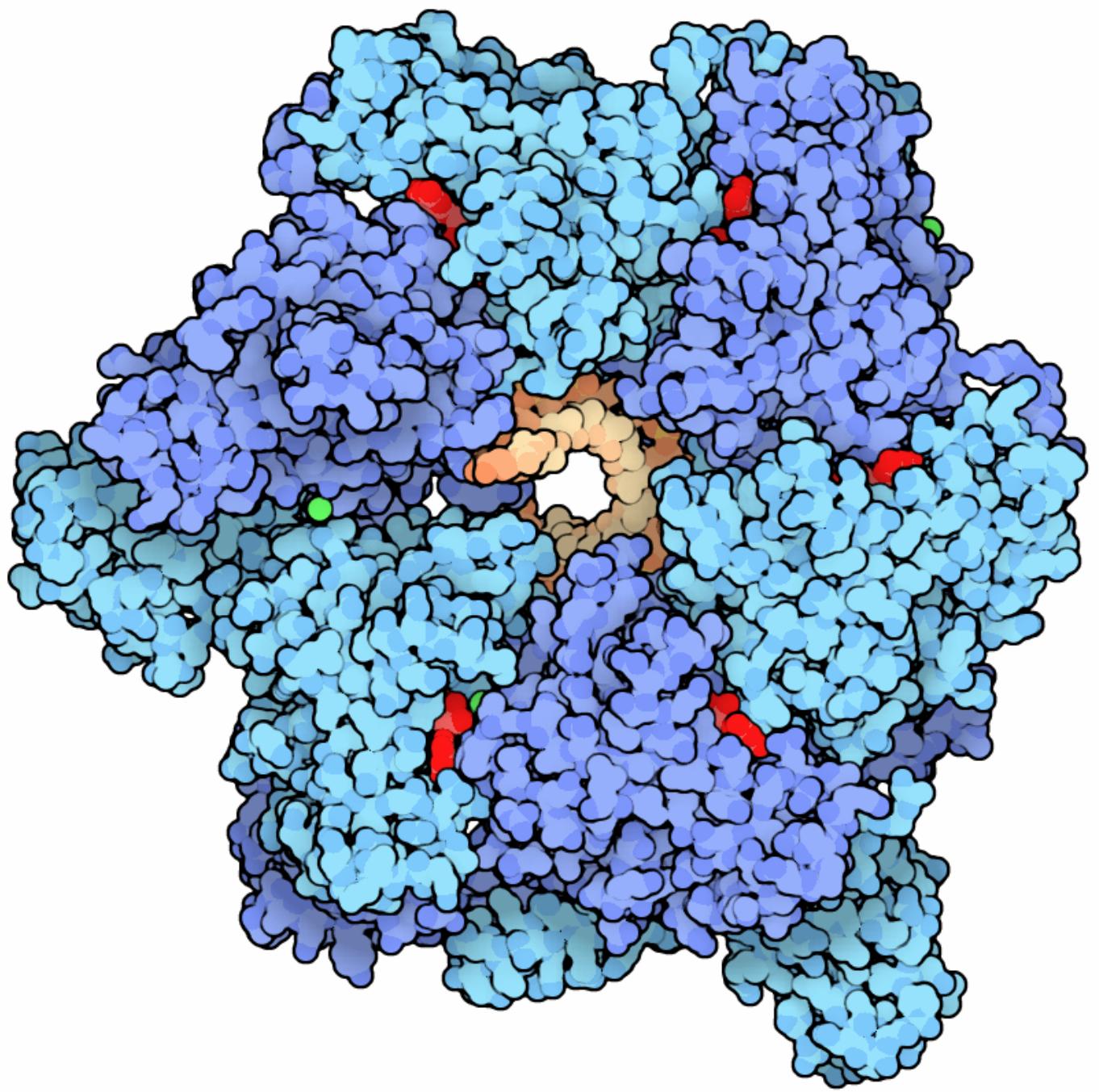
What strategies do you employ to meet those challenges?

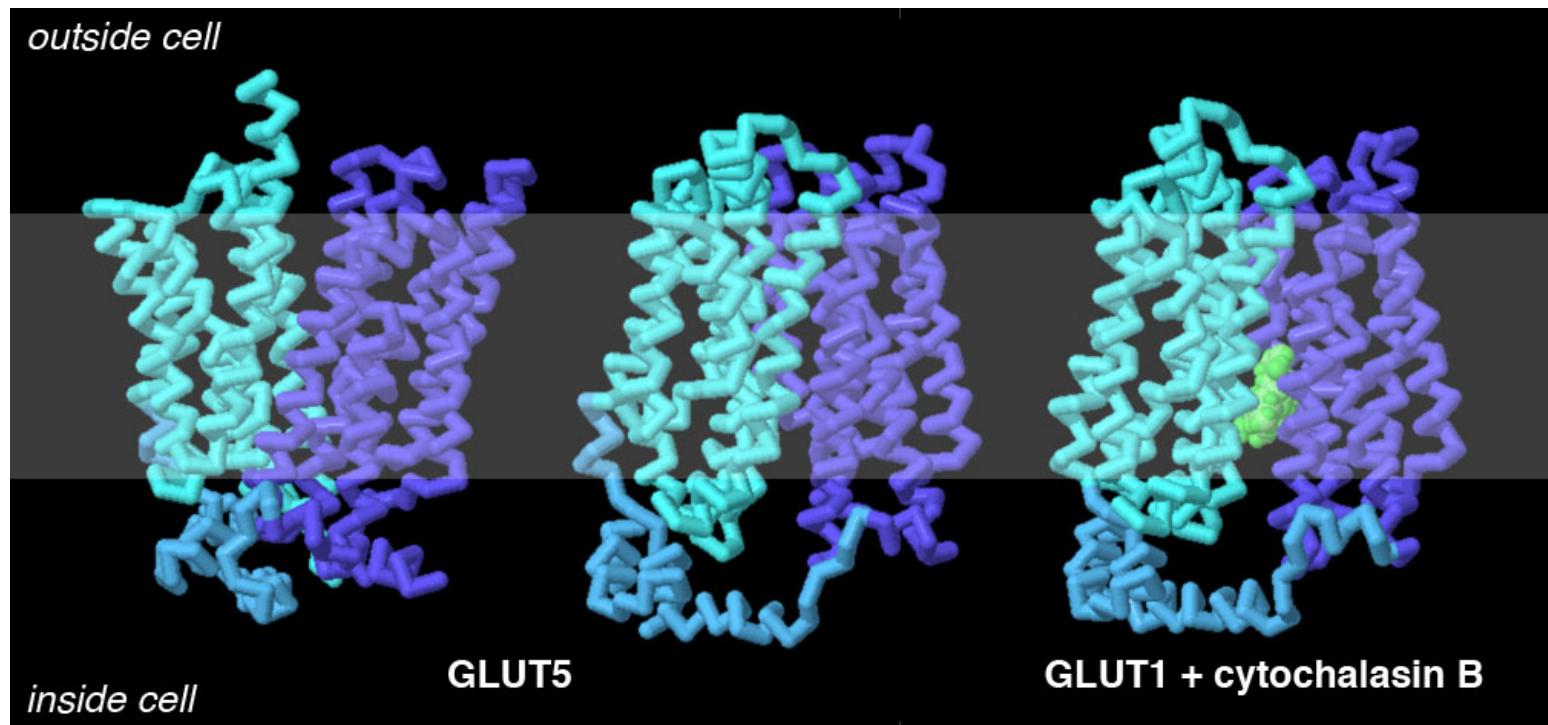
How do you choose what data to present and what to keep out?

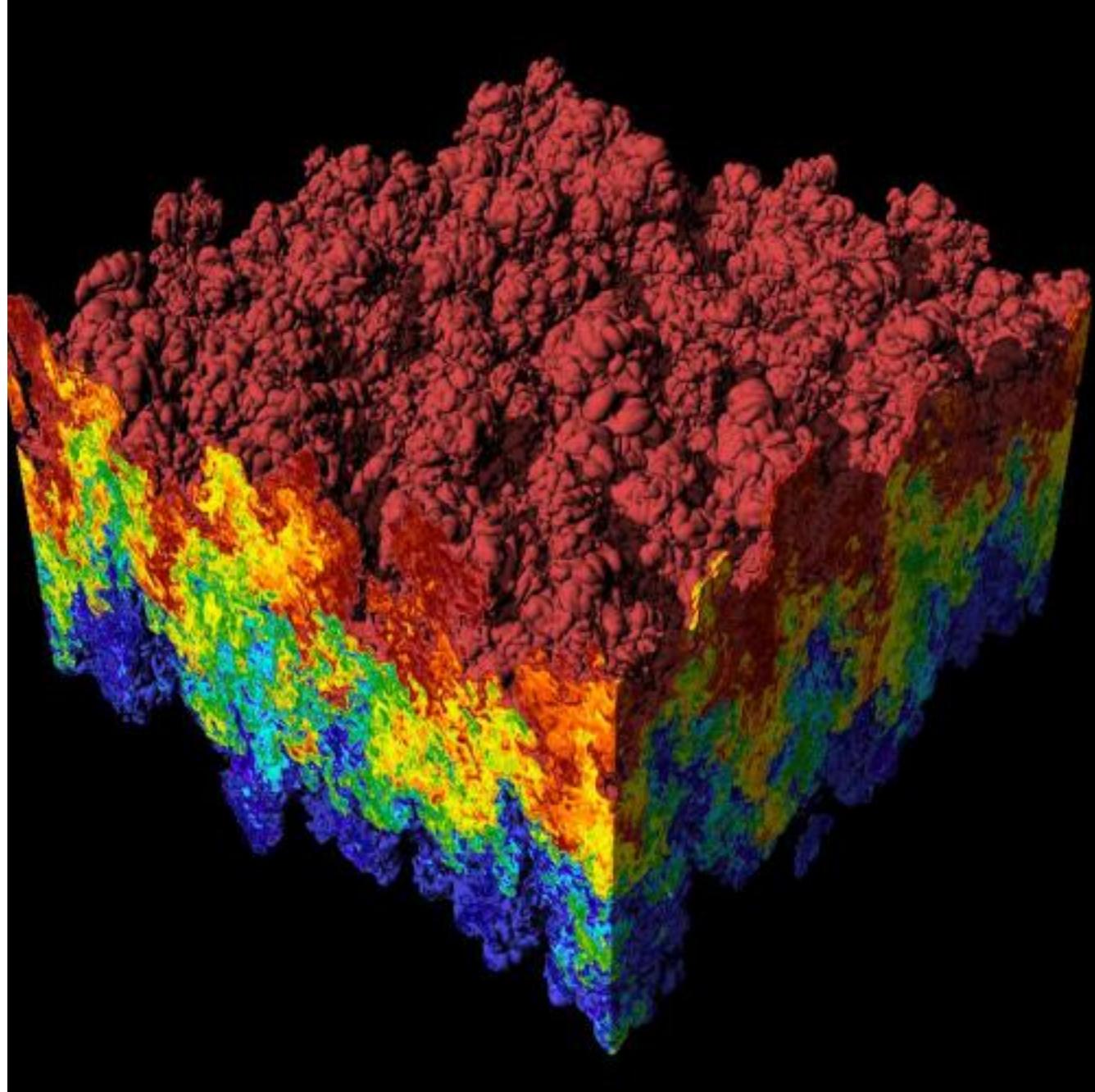
When is it appropriate to create a  
visualization in 3D?

When is it appropriate to create a  
visualization in 3D?

Only when the 3<sup>rd</sup> dimension is inherently spatial  
(rarely ever)







Two-dimensional representations of  
three-dimensional data use  
**visual depth cues**  
to represent the third dimension

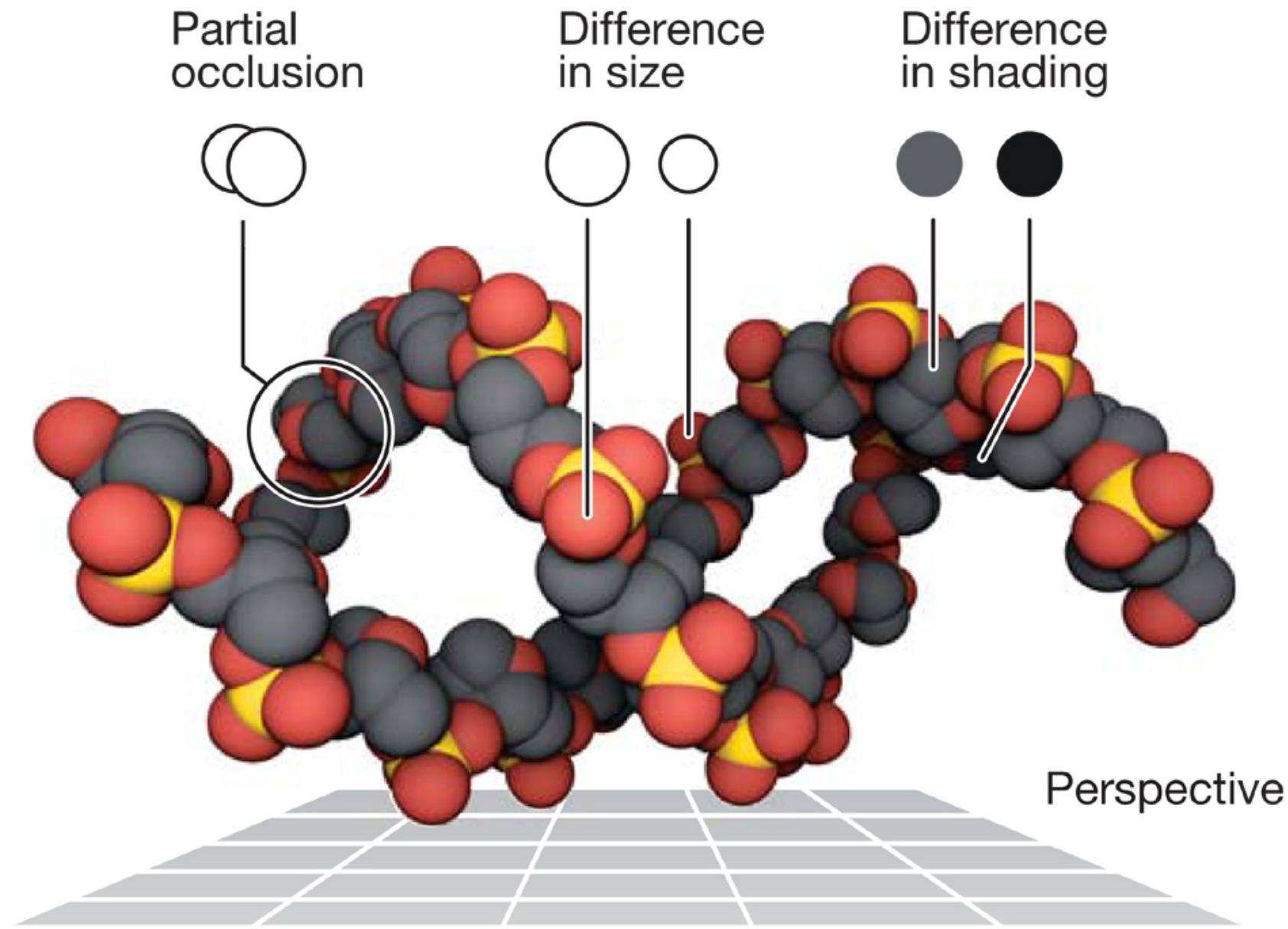
## VISUAL DEPTH CUES

Partial Occlusion

Shading

Color

Size



When plotting data in 3D space,  
visual cues for indicating depth

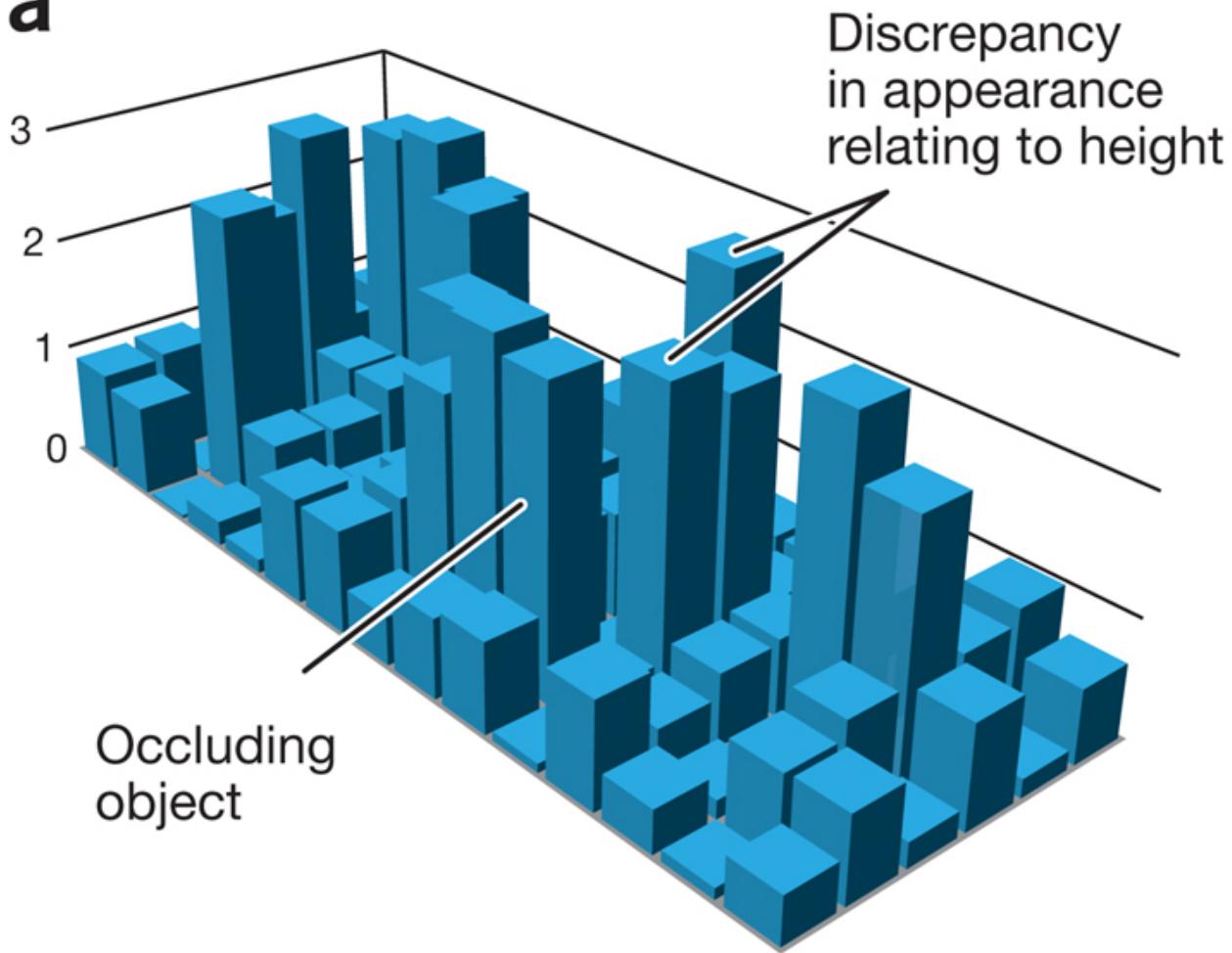
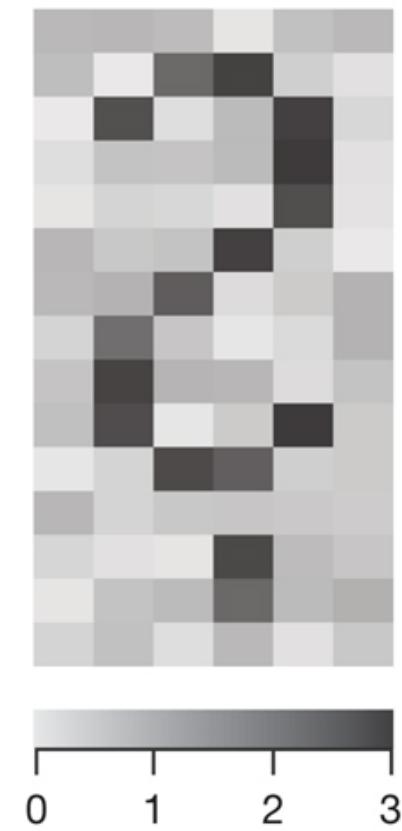
**can interfere**

with other common visual encodings

Sometimes, visual depth cues can  
produce unwanted

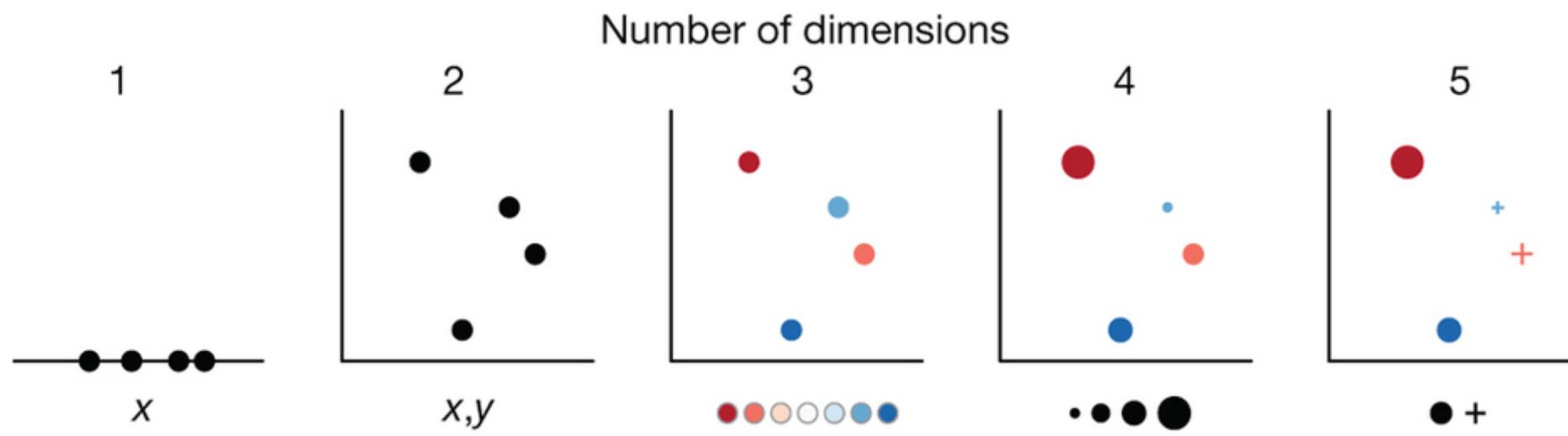
**perceptual artifacts**

if not used carefully

**a****b**

## VISUAL ENCODINGS

Color		Sequence		Length	
Value/Gradation		Size + Scale		Area	
Texture		Orientation		Proportion	
Symbol		Proximity/Density		Count	



## EXAMPLES OF MULTIDIMENSIONAL PLANE VISUALIZATIONS

Heatmap

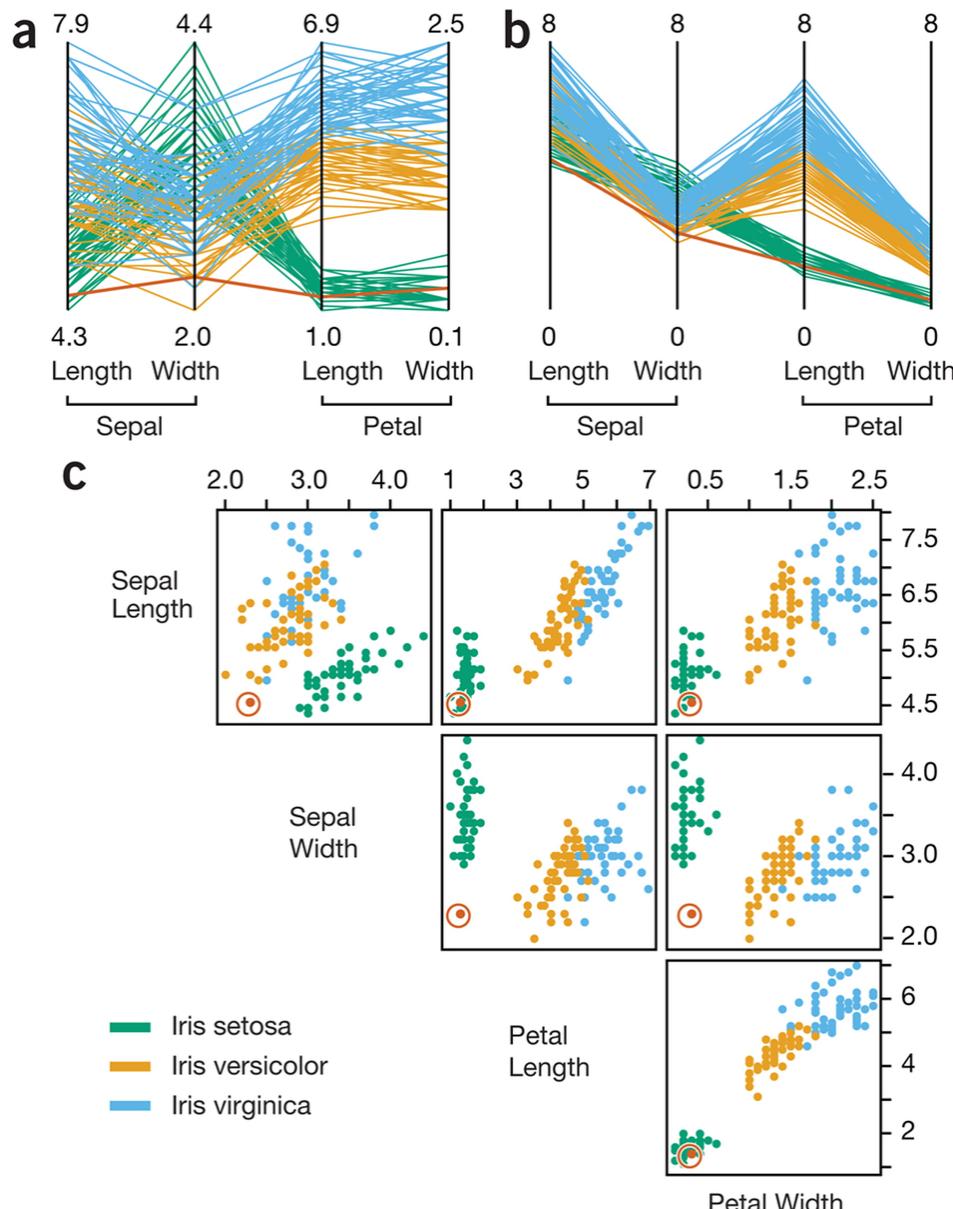
Scatter plot matrix

Parallel coordinates

Network

Trellis

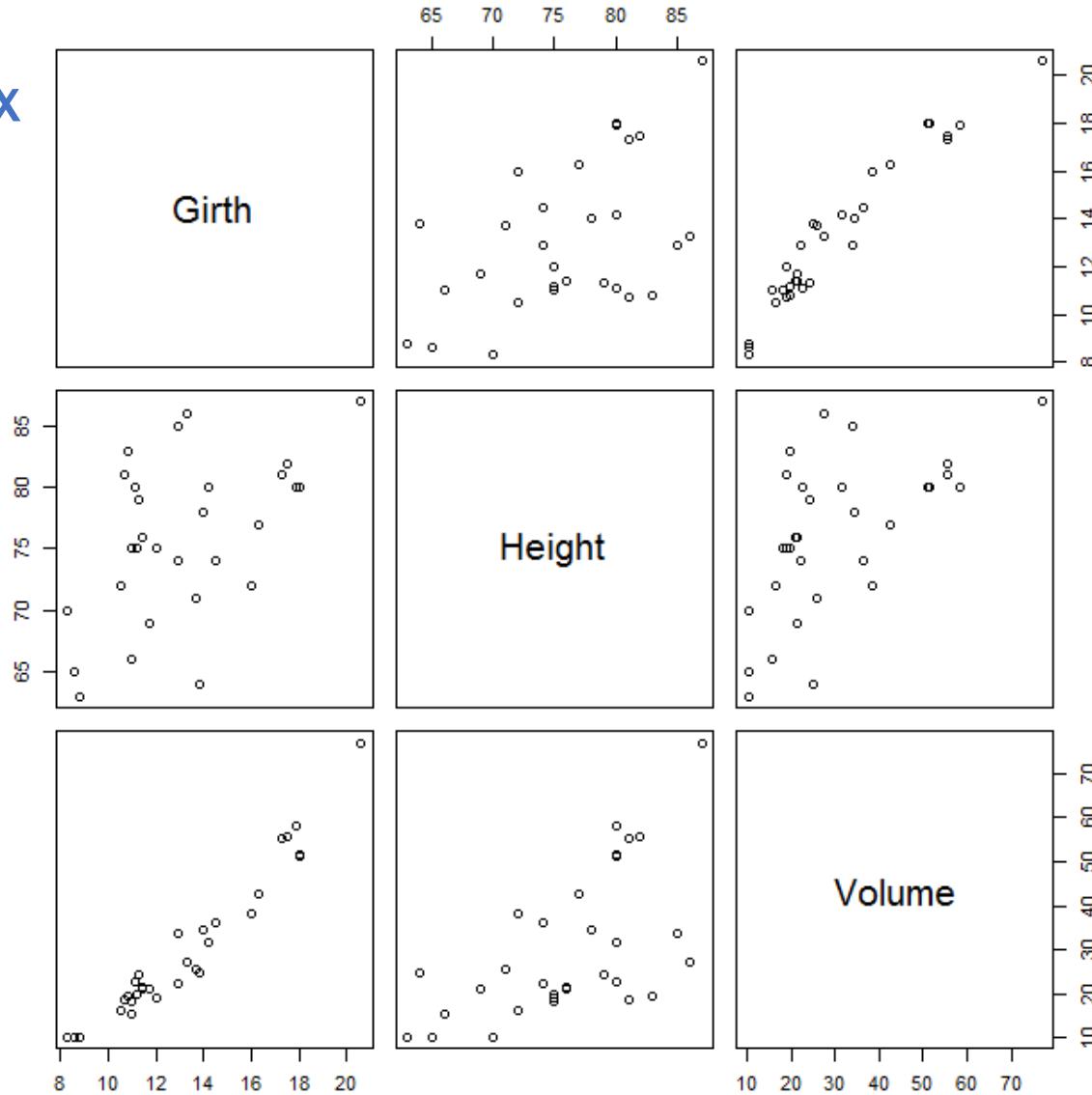
Small multiples



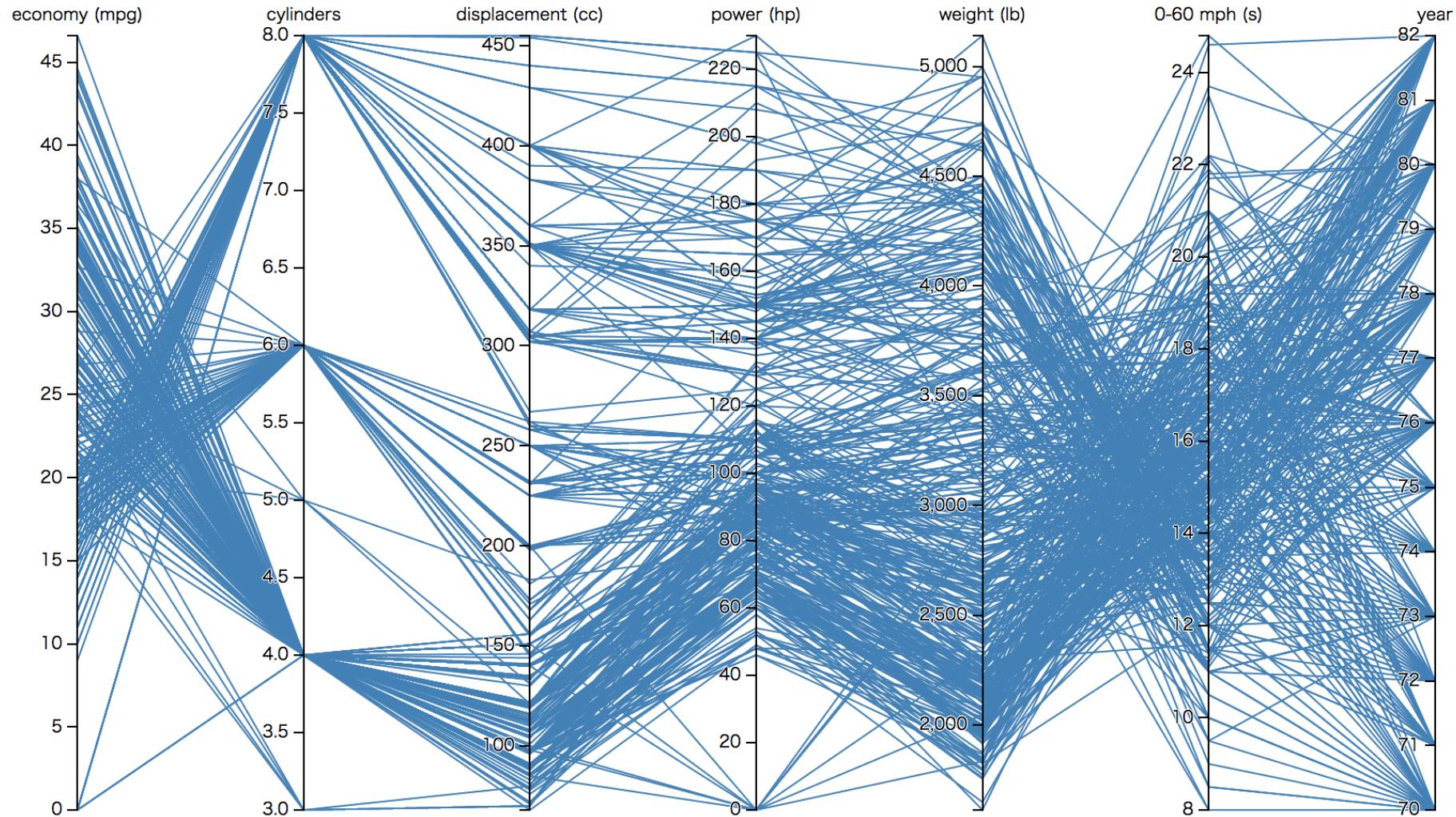
Parallel coordinate plots split out dimensions across several axes joined together by paths

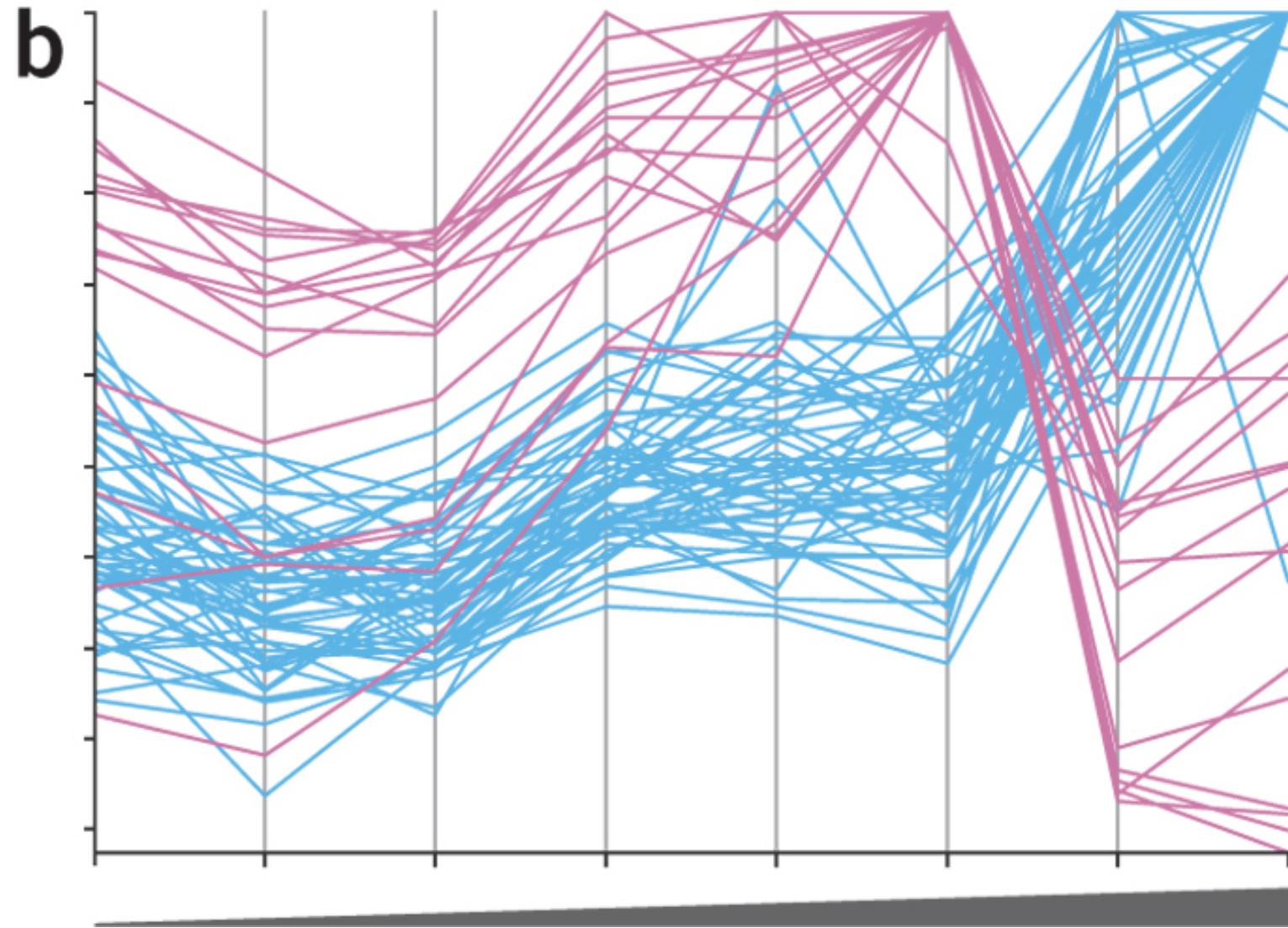
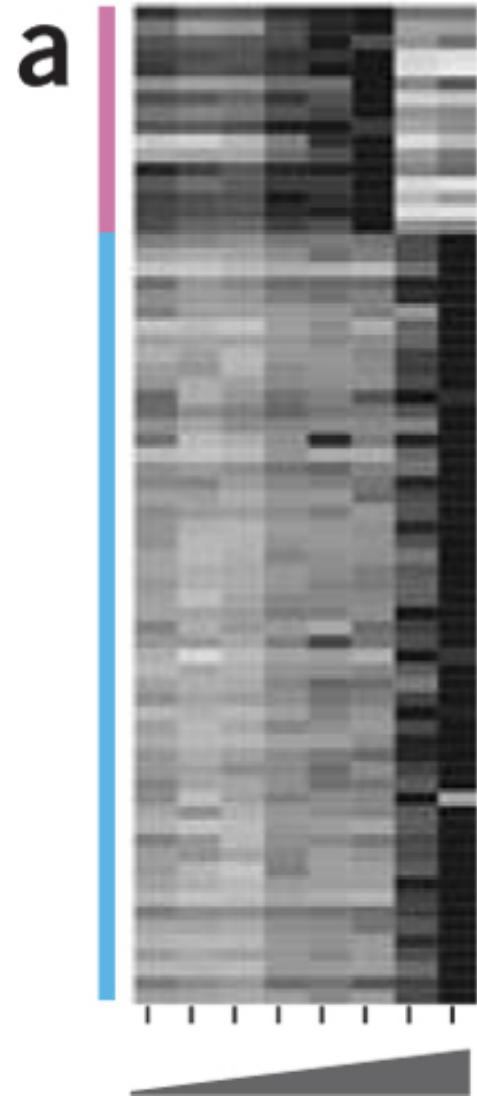
Scatter plot matrices arrange multiple scatter plots to enable cross-comparison across many dimensions

## Scatter plot matrix

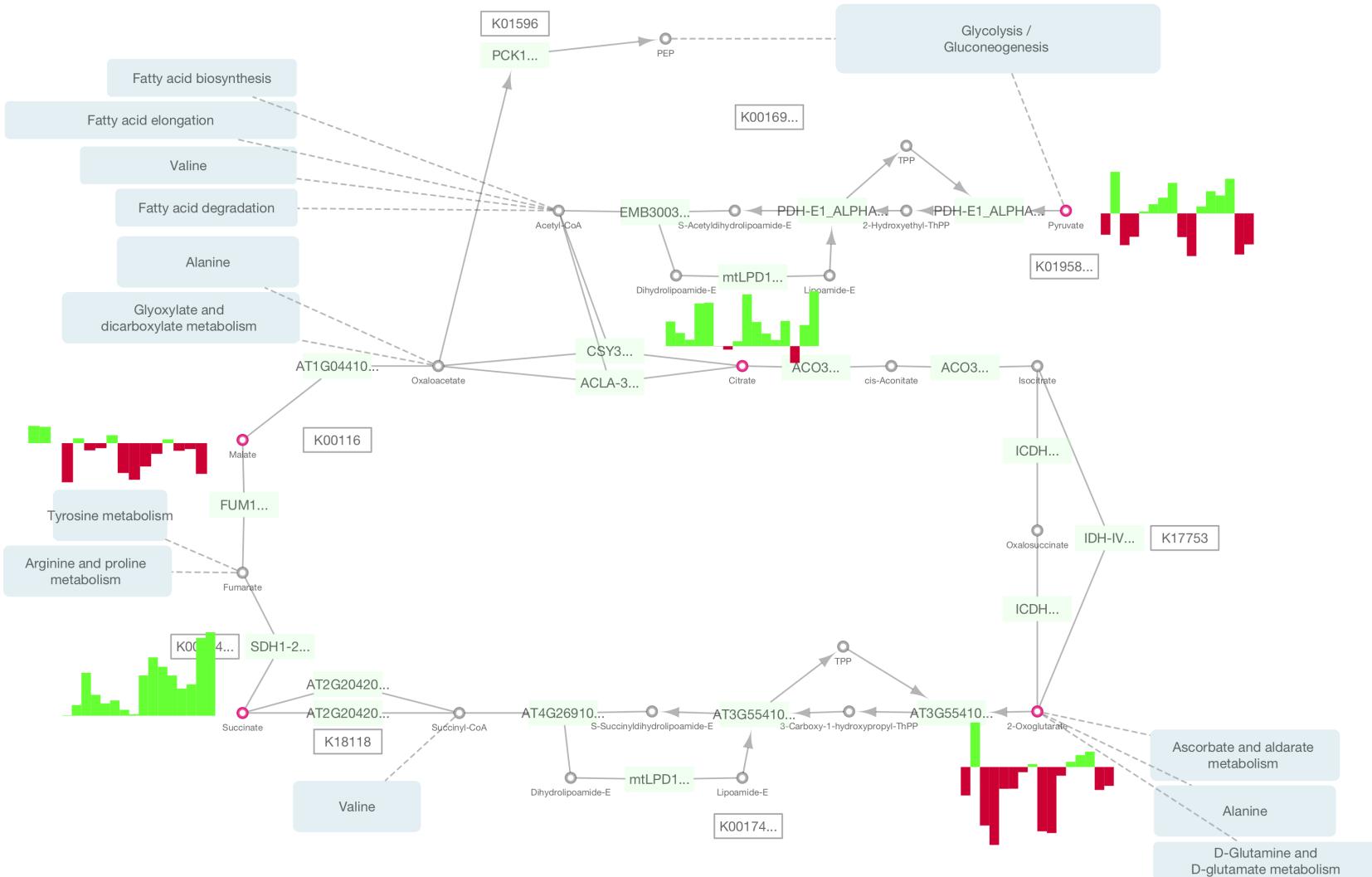


# Parallel Coordinates

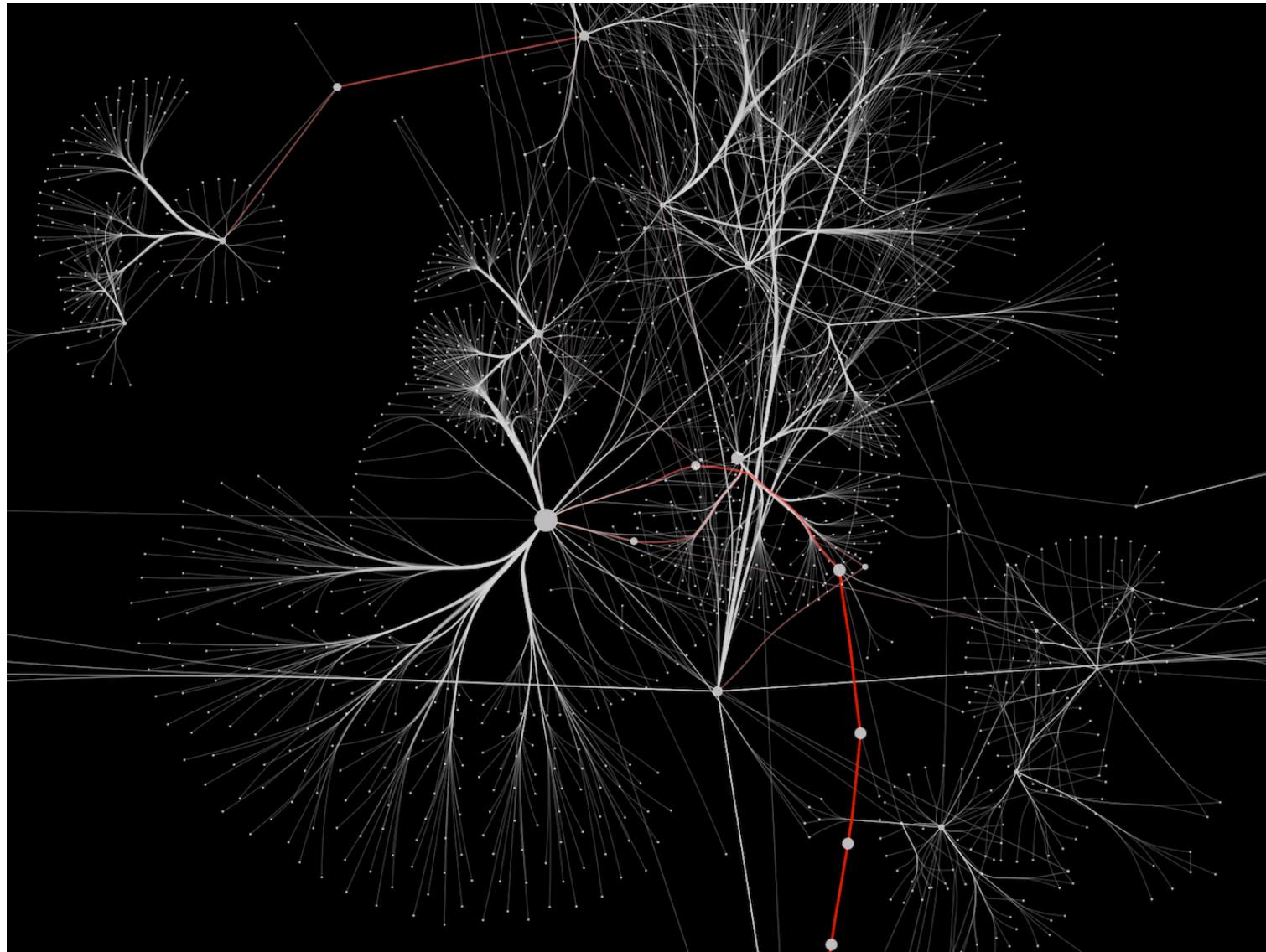




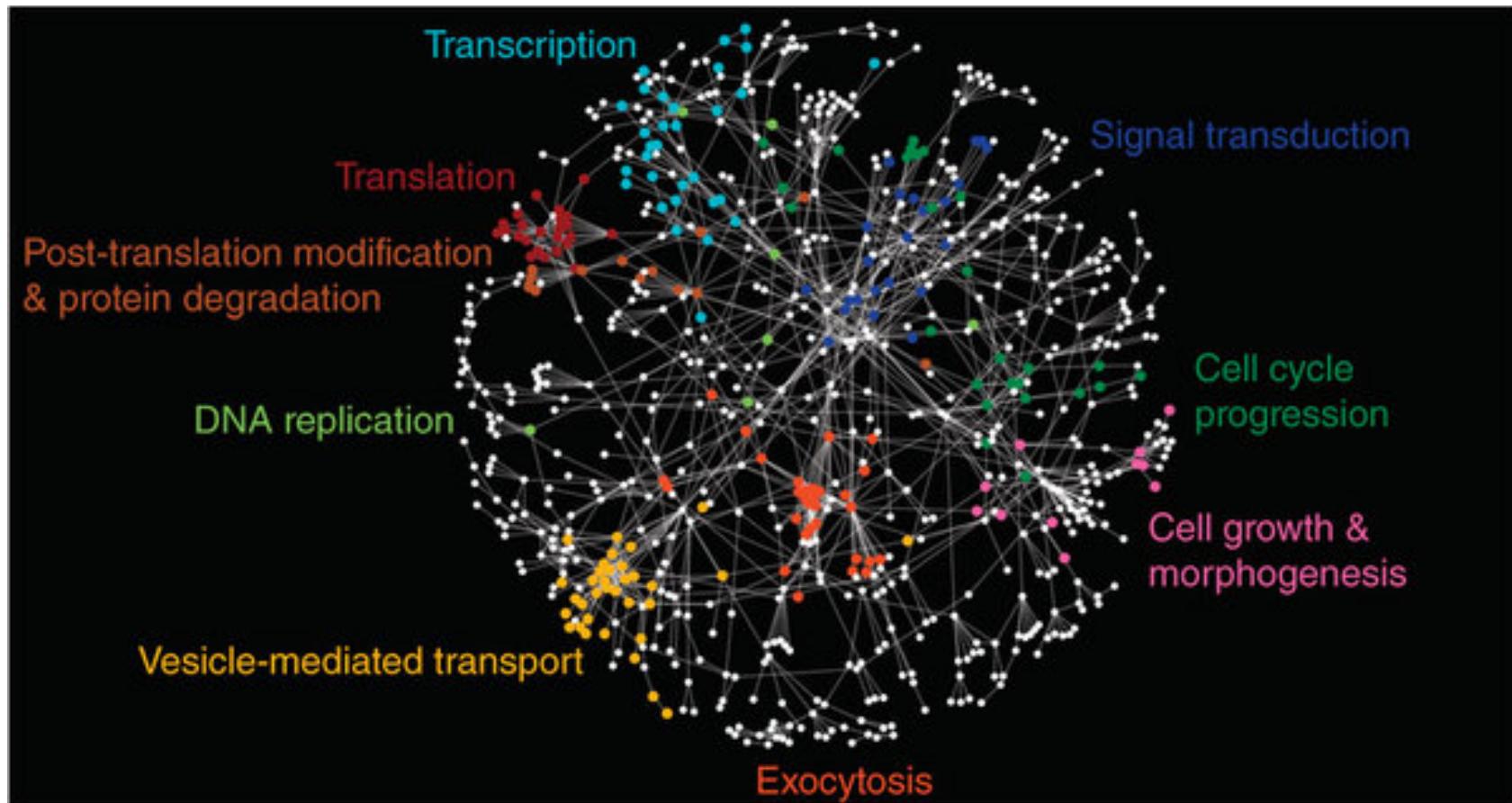
# Cytoscape



# Cytoscape

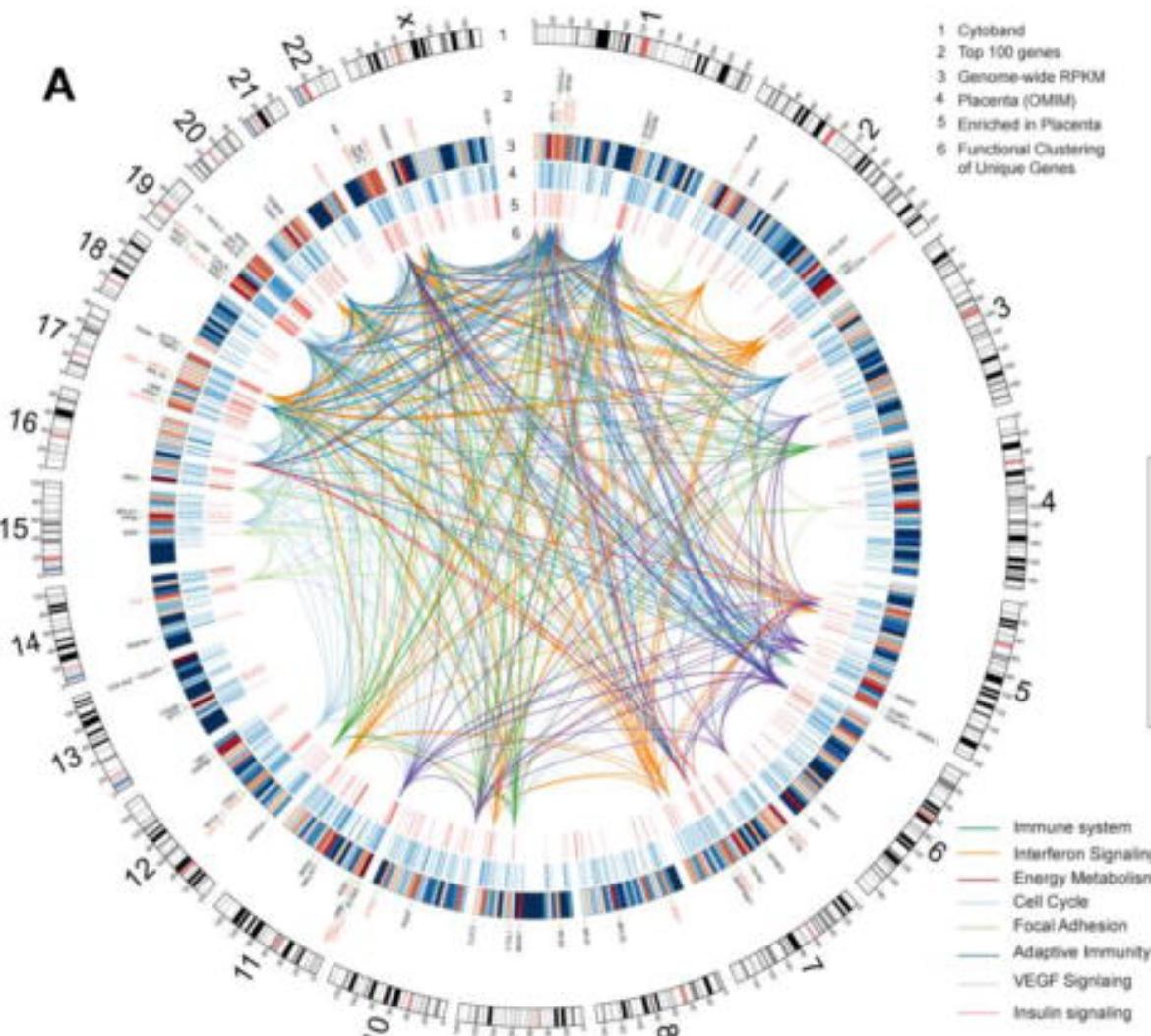


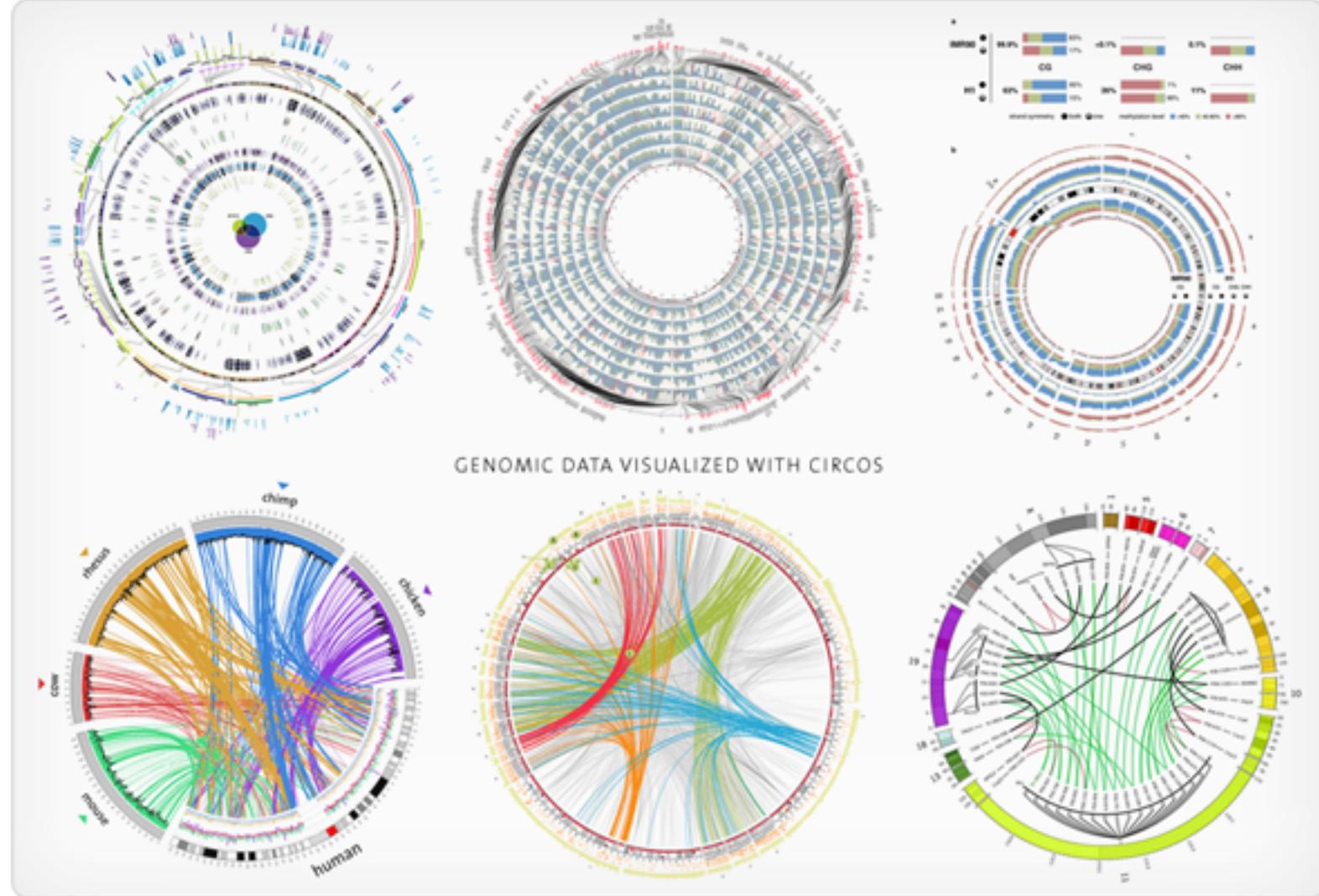
# Network



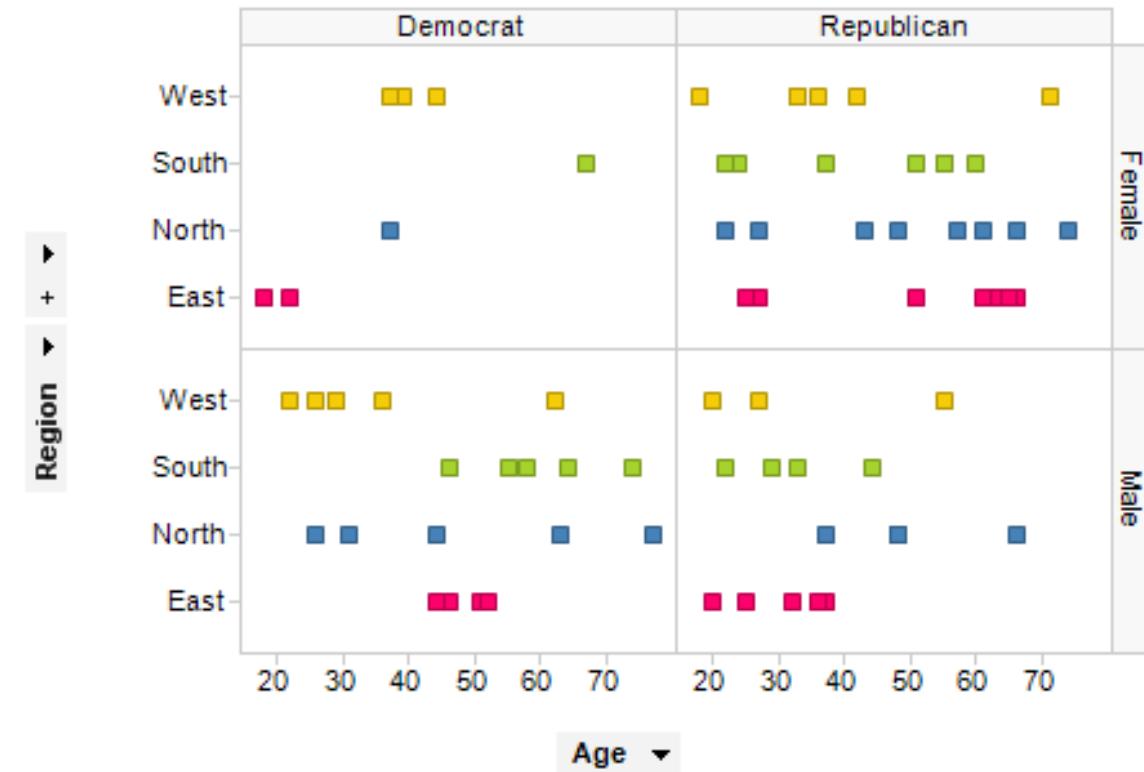
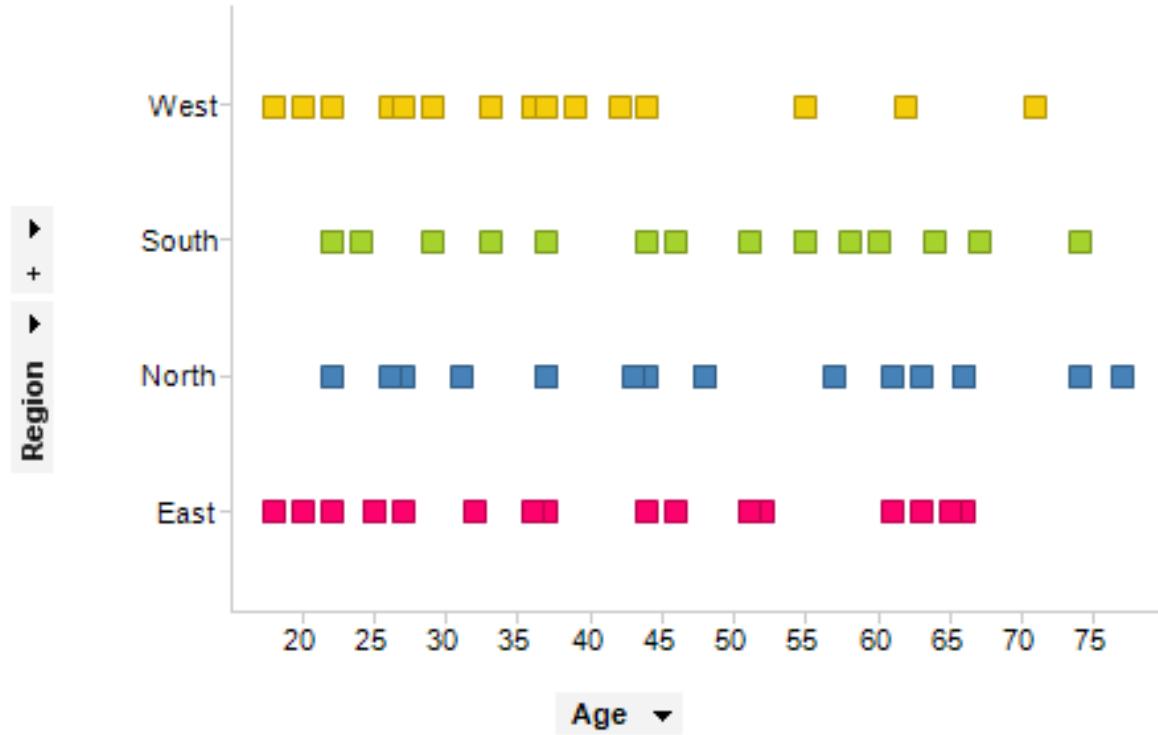
Magtanong et al. Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. *Nature Biotechnology* 29, 505–511 (2011)

# Chord Diagram





# Trellis Display



Encode null values appropriately to aid understanding

## TYPES OF NULL VALUES

Missing value

No answer, “N/A” (survey)

NaN

Type error

No signal

Nulls resulting from data that do not fit schema

# Working with Multidimensional Data

What kinds of null values or missing data do you deal with in your research?

What strategies do you employ to visually represent those values?

## Remember:

Visualizations of multidimensional  
data require more effort and training  
to be read and understood

When working with multidimensional  
data, we can use principles of  
**storytelling** to guide the focus of our  
audience

introduction  
question  
conflict  
buildup  
resolution

# WHERE THERE'S SMOKE—THERE'S CANCER

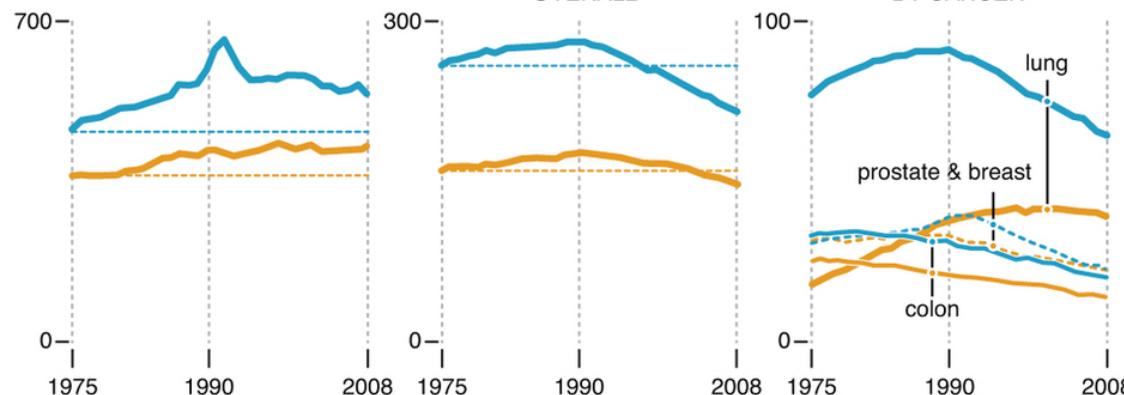
Cancer rates are up, but mortality is down. New diagnostics and treatments are responsible for part of this trend. But the greatest single contributing factor is the decline in smoking—rates are at their lowest level in 50 years.

Men      Women

## 1 Increased incidence

An aging population contributes to rising incidence of cancer.

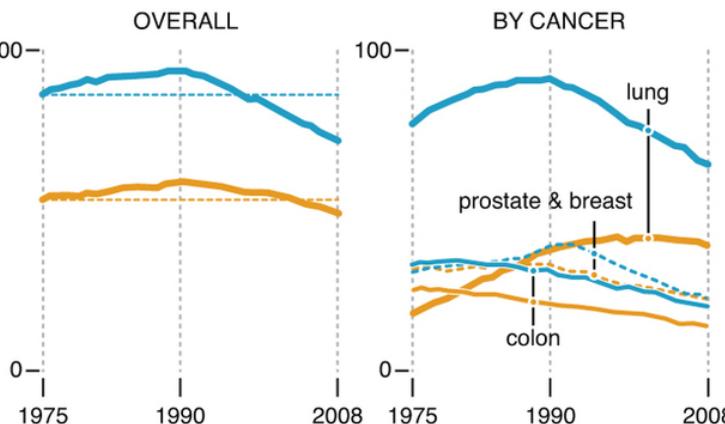
Cancer incidence rates (per 100,000)



## 2 Fewer deaths

Cancer deaths have been dropping since 1991, especially in males.

Cancer death rates (per 100,000)



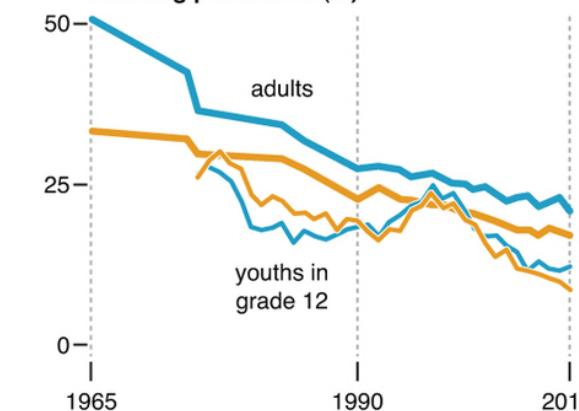
## 3 Decline of lung cancer

Drop in lung cancer deaths in males is the primary reason why death rates are down.

## 4 Decline in smoking

Since the 1964 first Surgeon General's report, smoking rates have been dropping. By 2010, the rate among males was down to 20%, from 50% at its peak. Among youths, rates have been on an even steeper decline since 1997.

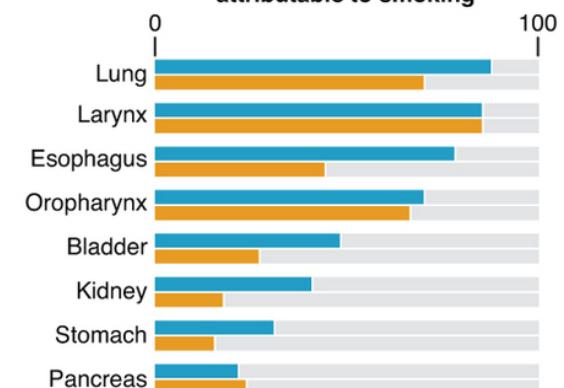
Smoking prevalence (%)



## 5 Impact of smoking on cancer deaths

Smoking is a major risk factor for many types of cancer and significant contributor to cancer-related deaths. It remains the single largest preventable cause of disease and premature death in the US.

Percentage of cancer deaths attributable to smoking



source: American Cancer Society Cancer Statistics 2012; Monitoring the Future (University of Michigan).

# Storytelling with Multidimensional Data

Explore the gallery of data visualizations and data stories for 2016 at the New York Times:

<https://www.nytimes.com/interactive/2016/12/28/us/year-in-interactive-graphics.html>

What are common storytelling elements employed in the examples you see?

# Storytelling with Multidimensional Data

In pairs, identify all the independent dimensions in the data set(s) provided.

Determine a narrative arc in the data to highlight, and using R markdown and ggplot2, create a series of visualizations with annotations that depict that narrative

introduction  
question  
conflict  
buildup  
resolution

Multidimensional data  
benefit from being represented in

**linked views**

## **TECHNIQUES FOR DIMENSIONAL REDUCTION**

Principle Component Analysis (PCA)

Clustering