What's Visualisation and why do it?

Visualisation (Vis) allows people to analyse data when they don't know exactly what question to ask in advance

Based on Chapter 1 of "Visualisation, Analysis and Design"

Back one step…

What do we mean by data?

Back one step…

What do we mean by data?

More complex than one might think!

Back one step…

What do we mean by data?

More complex than one might think!

Big Data, Little Data, No Data, Christine Borgman

Back one step…

What do we mean by data?

More complex than one might think!

Big Data, Little Data, No Data, Christine Borgman

“Data” includes, at a minimum, digital observation, scientific monitoring, data from sensors, metadata, model output and scenarios, qualitative or observed behavioural data, visualizations, and statistical data collected for administrative or commercial purposes. Data are generally viewed as input to the research process.

Fox and Harris Data Science Journal Vol 12, 10 Feb 2013

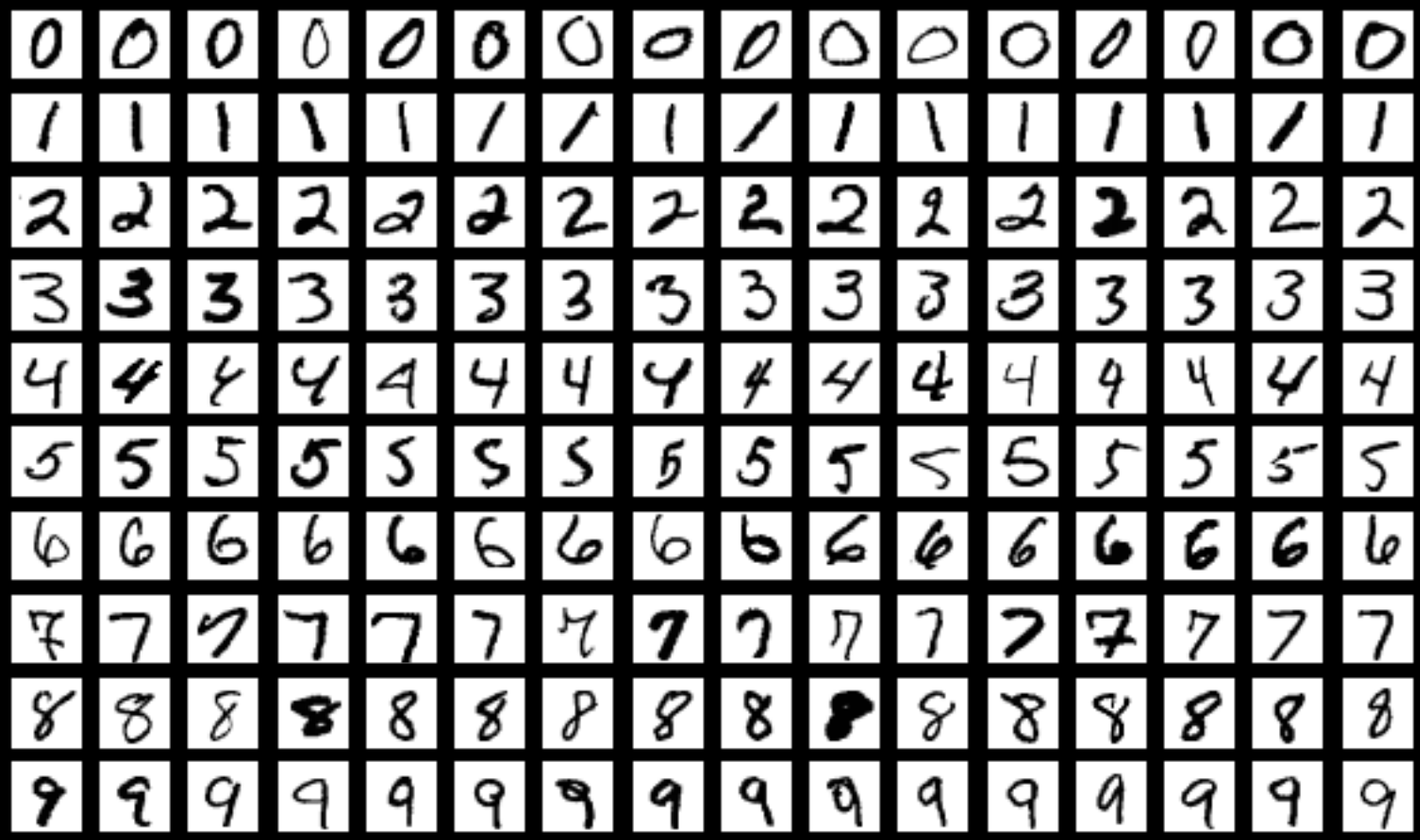
More colloquially

Data (for visualisation) is that which can be stored electronically….

Sometimes don’t need Vis

Many cases where a statistical analysis or using Machine Learning is sufficient

Example MNIST Data Set



By Josef Steppan - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=64810040

But often you do need Vis

But often you do need Vis

analysis problems are ill specified

But often you do need Vis

analysis problems are ill specified

people don’t know how to approach the problem

But often you do need Vis

analysis problems are ill specified

people don’t know how to approach the problem

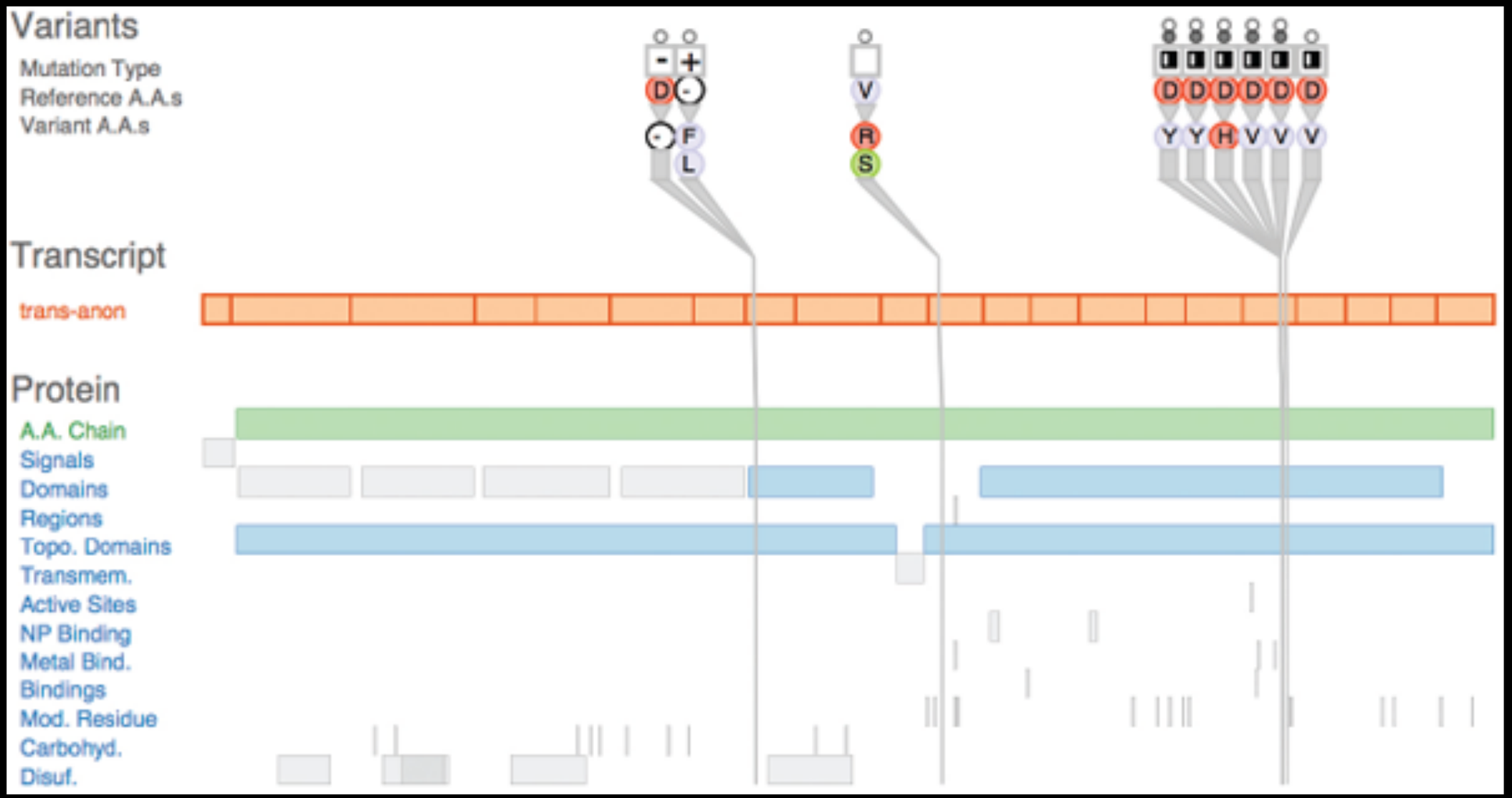
Augment human capabilities rather than replace humans

Vis can be..

For long term use - visualisations that are updated and used again and again

For short term/one off use - understanding something about the data; telling a story

Long term use



Genomic sequences - showing variation

Joel A. Ferstay, Cydney B. Nielsen, and Tamara Munzner. "Variant View: Visualizing Sequence Variants in Their Gene Context." *IEEE Transactions on Visualization Computer Graphics (Proc. InfoVis 13)* 19:12 (2013), 2546–2555. *(page 4)*

Why Computational tools?

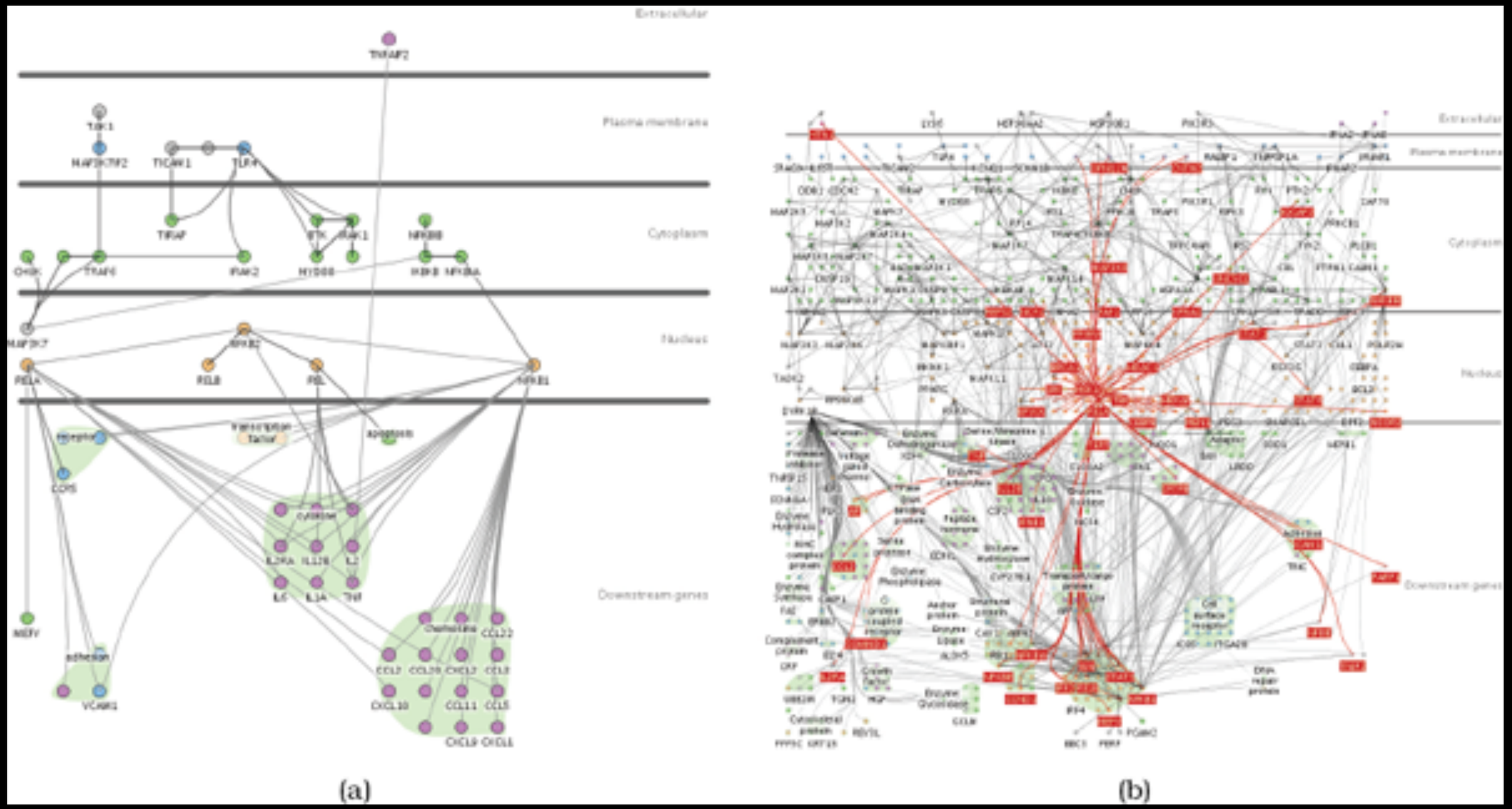
Mona Chalabi and Charles Minard did it by hand!

Need to visualise

Large/Complex data sets

Variable data sets

Quickly and in an exploratory fashion



Could be done by hand

Too hard to do by hand!

Aaron Barsky, Jennifer L. Gardy, Robert E. Hancock, and Tamara Munzner. “Cerebral: A Cytoscape Plugin for Layout of and Interaction with Biological Networks Using Subcellular Localization Annotation.” *Bioinformatics* 23:8 (2007), 1040–1042. *(pages 5, 6, 295)*

Why Vision?

Able to absorb large amounts of information in parallel.

Why Vision?

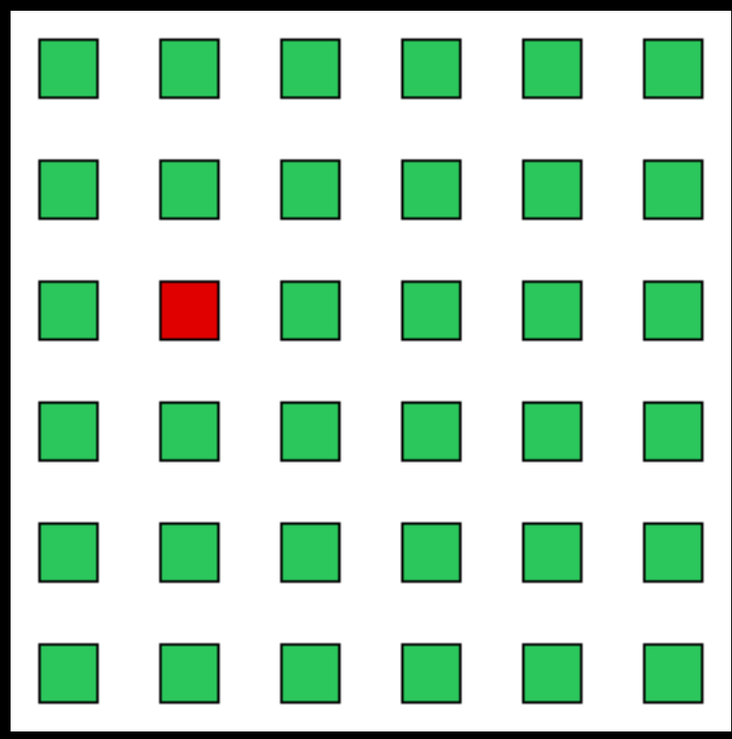
Able to absorb large amounts of information in parallel.

Example - Visual Pop-out

Why Vision?

Able to absorb large amounts of information in parallel.

Example - Visual Pop-out



By Head 21:20, 24 August 2007 (UTC) (self-drawn) [Public domain], via Wikimedia Commons

Other senses?

Sound - sequence-based, not parallel

Listen to individual voices in a choir (for example)

Why do we need to look at data in detail?

Why do we need to look at data in detail?

Even “simple” data can mislead us

Why do we need to look at data in detail?

Even “simple” data can mislead us

Need to check if summary statistics are right.

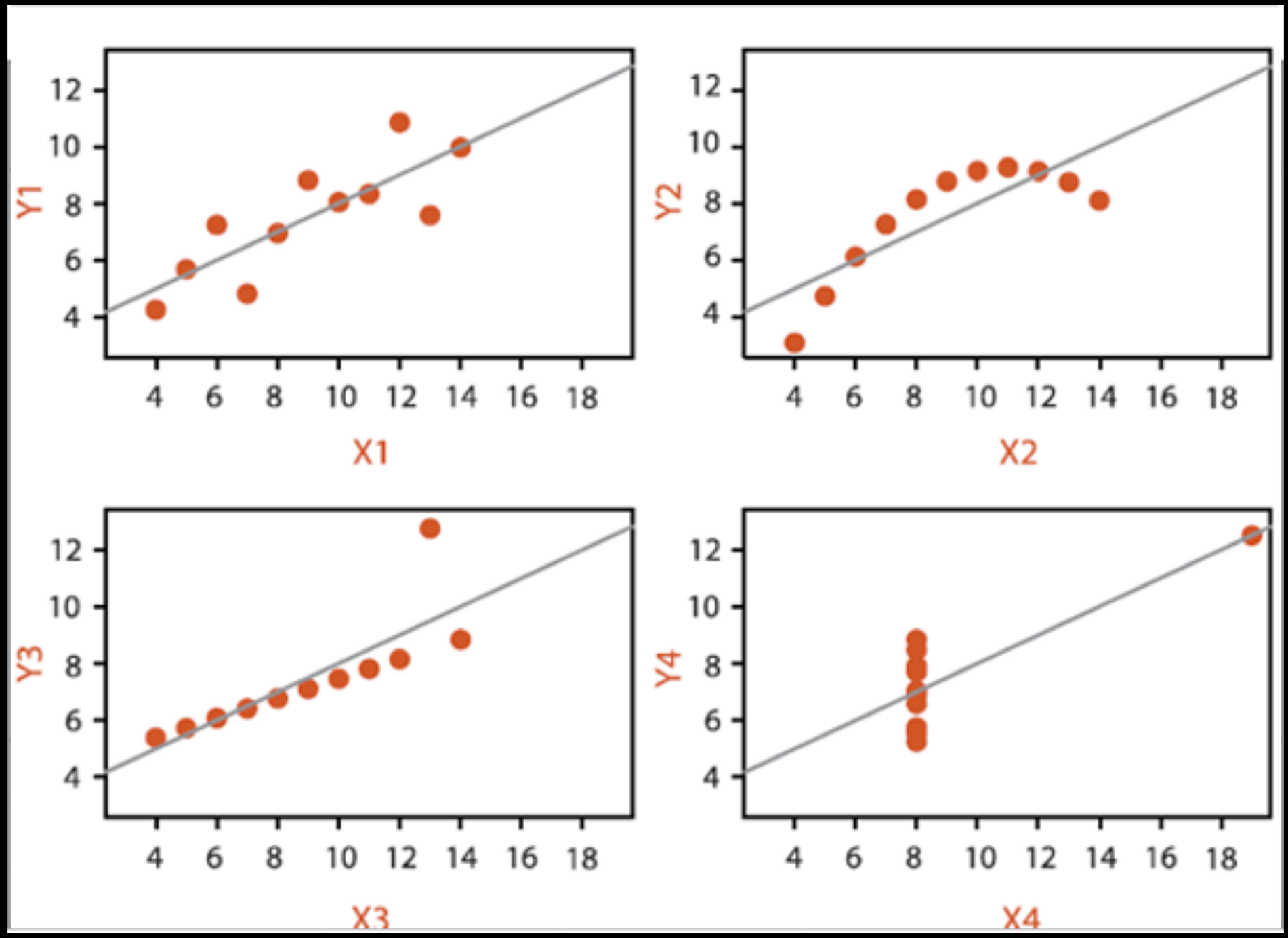
Anscombe's Quartet

Anscombe's Quartet: Raw Data

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

F.J. Anscombe. "Graphs in Statistical Analysis." *American Statistician* 27 (1973), 17–21. *(pages 7, 8, 19)*

Anscombe's Quartet



F.J. Anscombe. "Graphs in Statistical Analysis." *American Statistician* 27 (1973), 17–21. *(pages 7, 8, 19)*

Interaction

Interaction

Same data set - many different ways to visualise

Interaction

Same data set - many different ways to visualise

Tools to analyse particular type of data

Interaction

Same data set - many different ways to visualise

Tools to analyse particular type of data

Variant View - discussed above

Interaction

Same data set - many different ways to visualise

Tools to analyse particular type of data

Variant View - discussed above

But can also interact by writing new code

Vis Idiom

A Vis idiom is a distinct way to creating and manipulating visual representations.

Vis Idiom

A Vis idiom is a distinct way to creating and manipulating visual representations.

Examples of idioms

Scatter plots

Bar plots

Pie Charts (please don’t)

Vis Idiom

A Vis idiom is a distinct way to creating and manipulating visual representations.

Examples of idioms

Scatter plots

Bar plots

Pie Charts (please don’t)

Huge possible number of Idioms

See for example Information is Beautiful

Tasks

No one visualisation does everything

Task - what is the purpose of the visualisation?

Journalism or telling a story?

Exploring data?

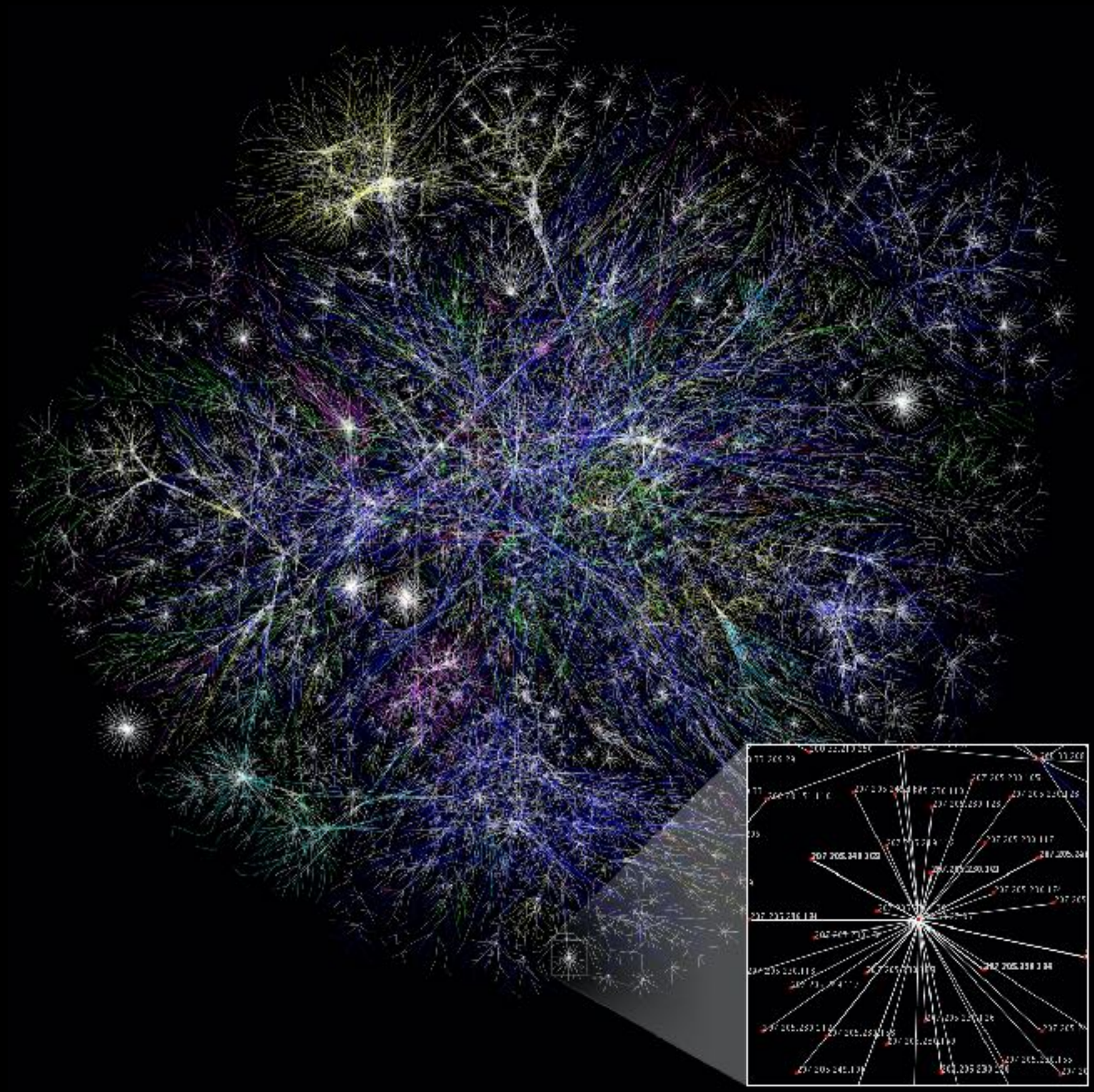
Testing hypotheses?

Effectiveness

Once you've decided upon the task

Did it do the job?

Visualisations can be beautiful - but not effective!



By The Opte Project [CC BY 2.5 (https://creativecommons.org/licenses/by/2.5)], via Wikimedia Commons

Resource limitation

Resource limitation

Computational limitations

Resource limitation

Computational limitations

e.g. How quickly can a visualisation go through a data set?

Resource limitation

Computational limitations

e.g. How quickly can a visualisation go through a data set?

Human limitations

Resource limitation

Computational limitations

e.g. How quickly can a visualisation go through a data set?

Human limitations

Human memory for items not visible is poor

Resource limitation

Computational limitations

e.g. How quickly can a visualisation go through a data set?

Human limitations

Human memory for items not visible is poor

“Change Blindness” - focus on one thing, ignore others

Resource limitation

Computational limitations

e.g. How quickly can a visualisation go through a data set?

Human limitations

Human memory for items not visible is poor

“Change Blindness” - focus on one thing, ignore others

Information Density

Information Density

Information Density

The information density of a single image is a measure of the amount of information encoded versus the amount of unused space.

Information Density

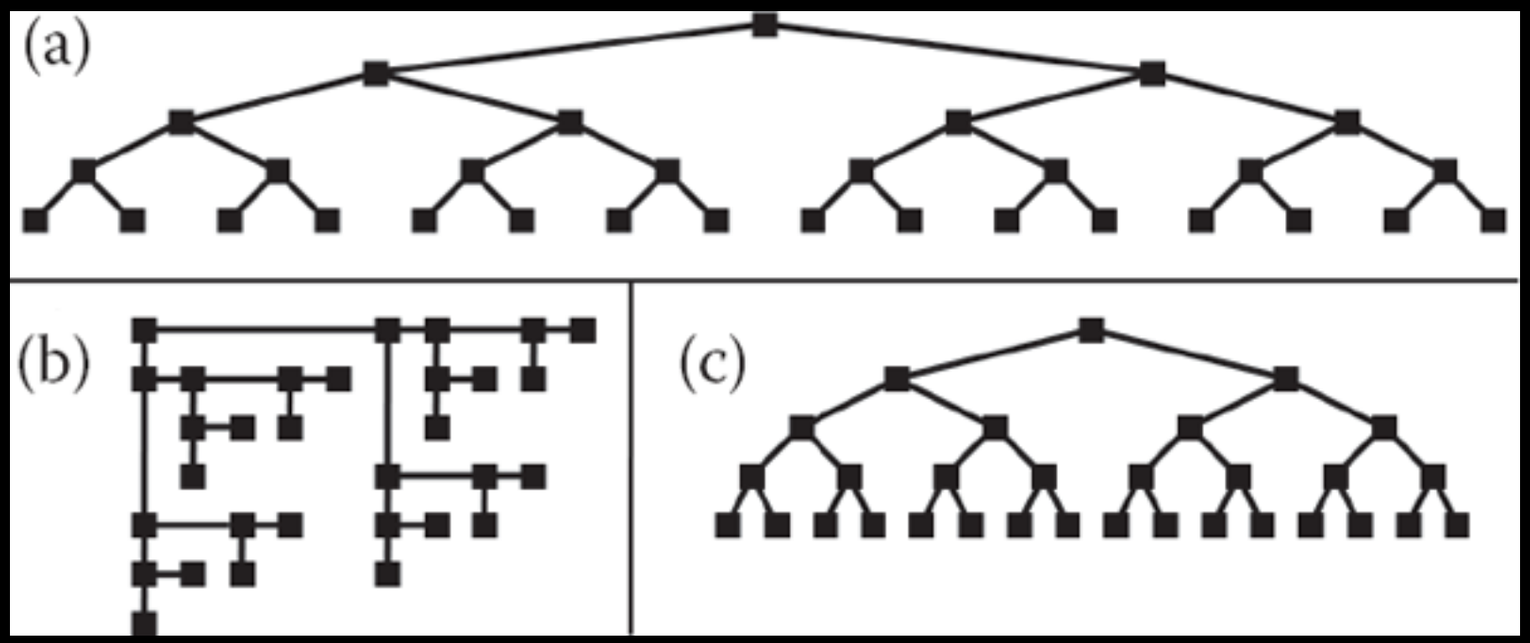
The information density of a single image is a measure of the amount of information encoded versus the amount of unused space.

Information Density also called Graphic Density or Data-Ink Ratio

Information Density

The information density of a single image is a measure of the amount of information encoded versus the amount of unused space.

Information Density also called Graphic Density or Data-Ink Ratio

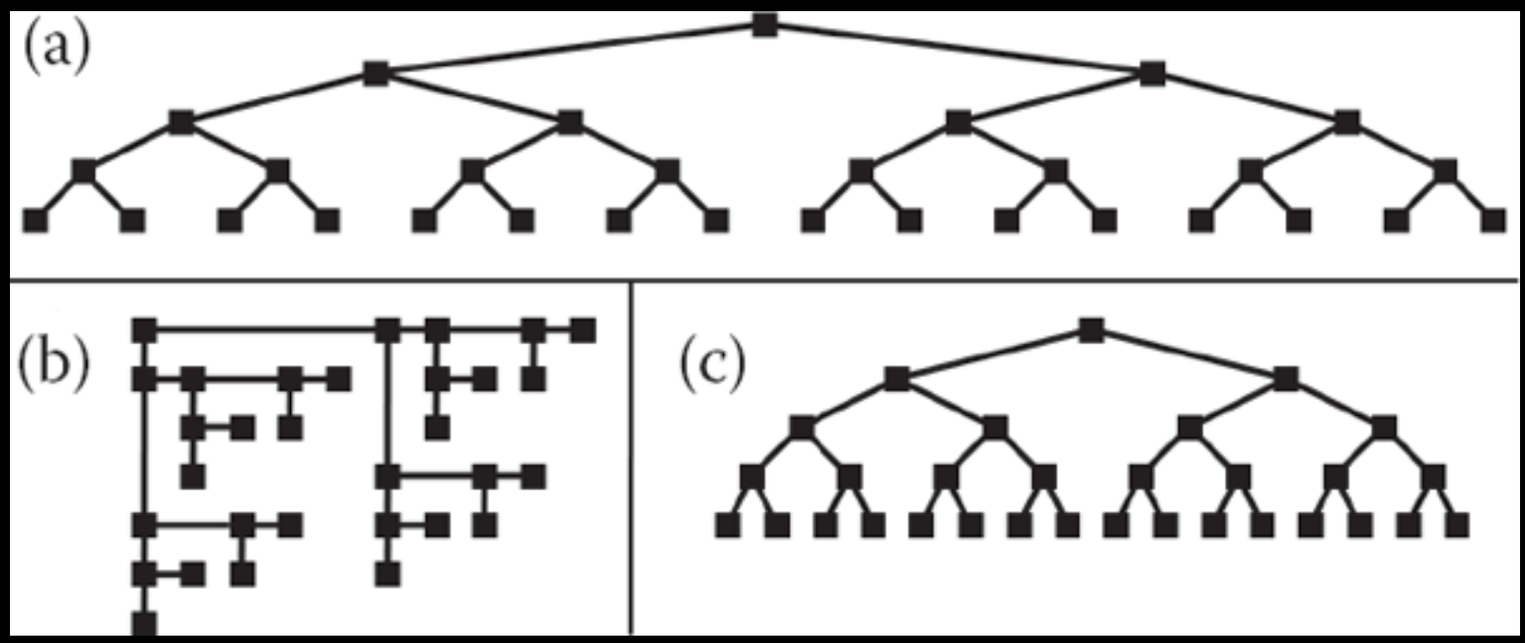


Information Density

The information density of a single image is a measure of the amount of information encoded versus the amount of unused space.

Information Density also called Graphic Density or Data-Ink Ratio

Low info. Density



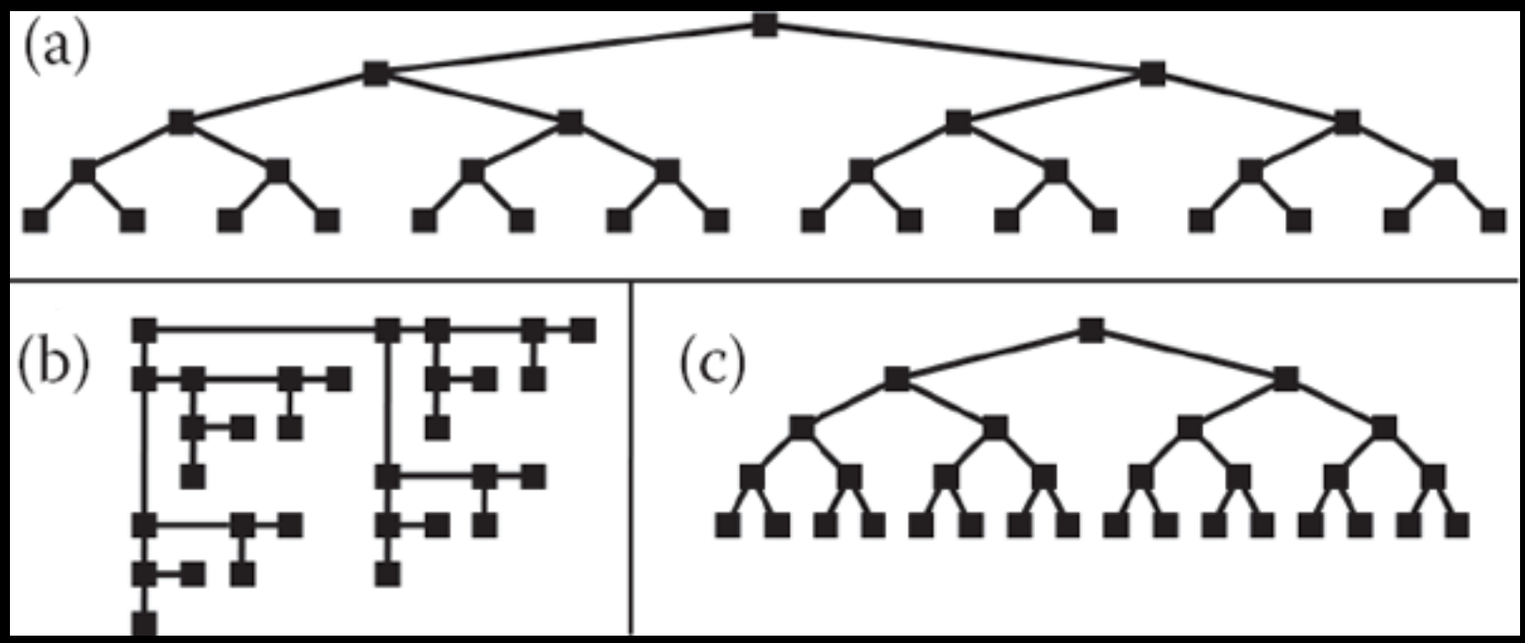
Information Density

The information density of a single image is a measure of the amount of information encoded versus the amount of unused space.

Information Density also called Graphic Density or Data-Ink Ratio

Low info. Density

High Info. Density (but can't tell tree order)



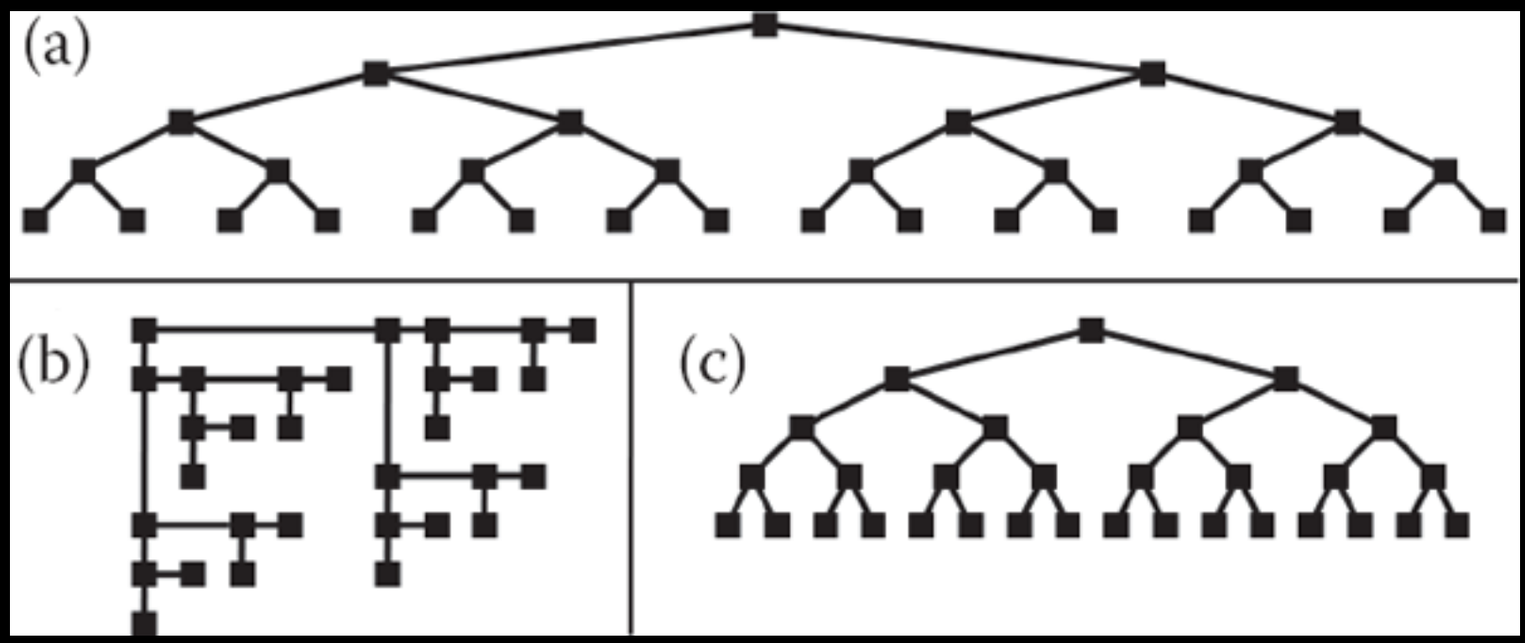
Information Density

The information density of a single image is a measure of the amount of information encoded versus the amount of unused space.

Information Density also called Graphic Density or Data-Ink Ratio

Low info. Density

High Info. Density (but can't tell tree order)



High Info. Density and can Tell tree order