To Draw a Tree

Pat Hanrahan

Computer Science Department
Stanford University

Motivation

Hierarchies

- File systems and web sites
- Organization charts
- Categorical classifications
- Similiarity and clustering

Branching processes

- Genealogy and lineages
- Phylogenetic trees

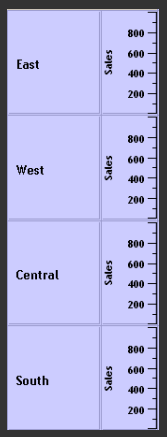
Decision processes

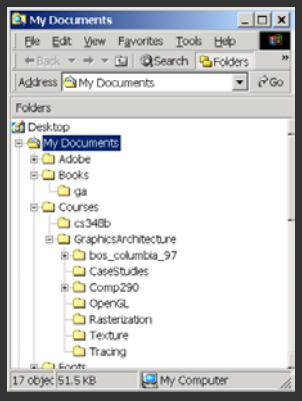
- Indices or search trees
- Decision trees
- Tournaments

Tree Drawing

Simple Tree Drawing

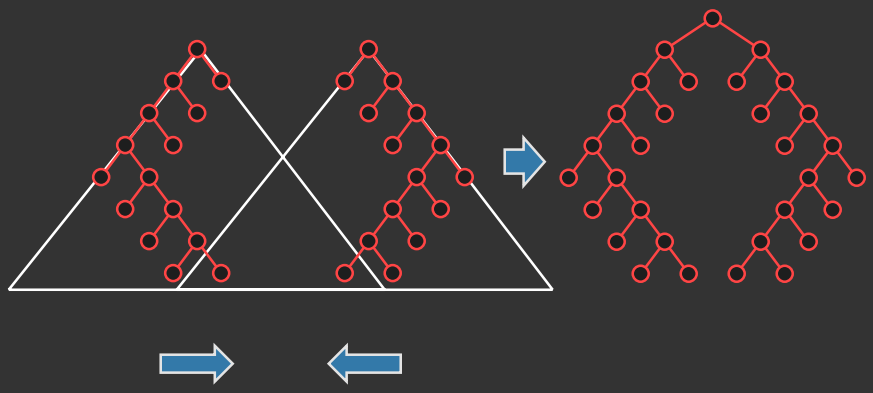
Preorder or inorder traversal





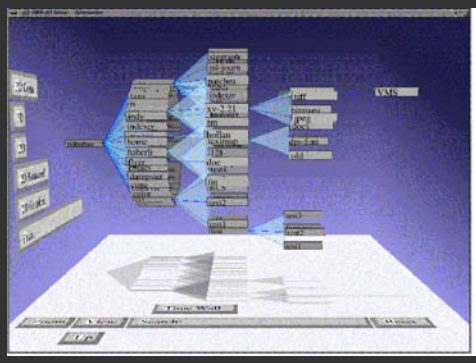
East				West				Cent	
Coffee	Espresso	Herbal Te	Tea	Coffee	Espresso	Herbal Te	Tea	Coffee	Espresso

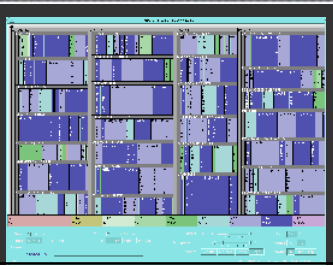
Rheingold-Tilford Algorithm



Information Visualization

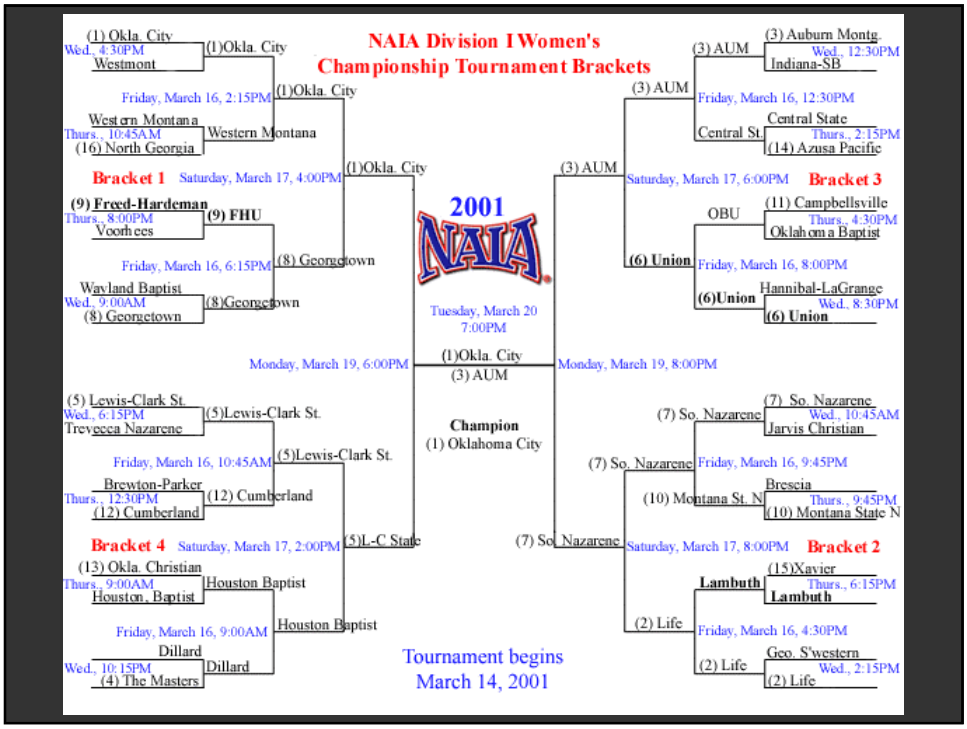
Tree Representations





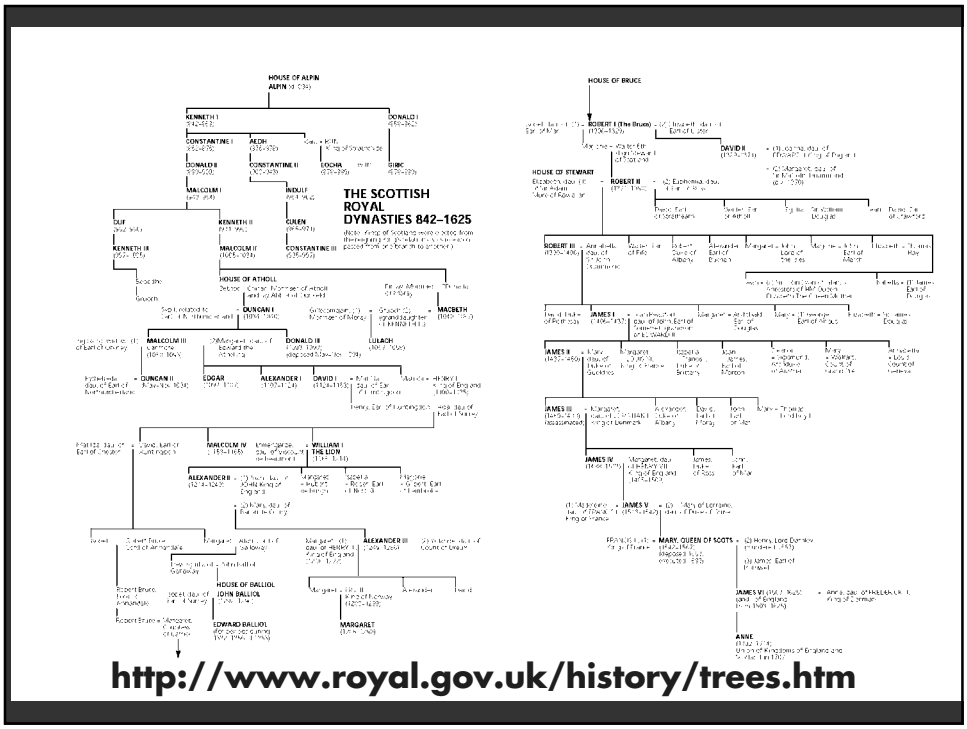
Most Common ...

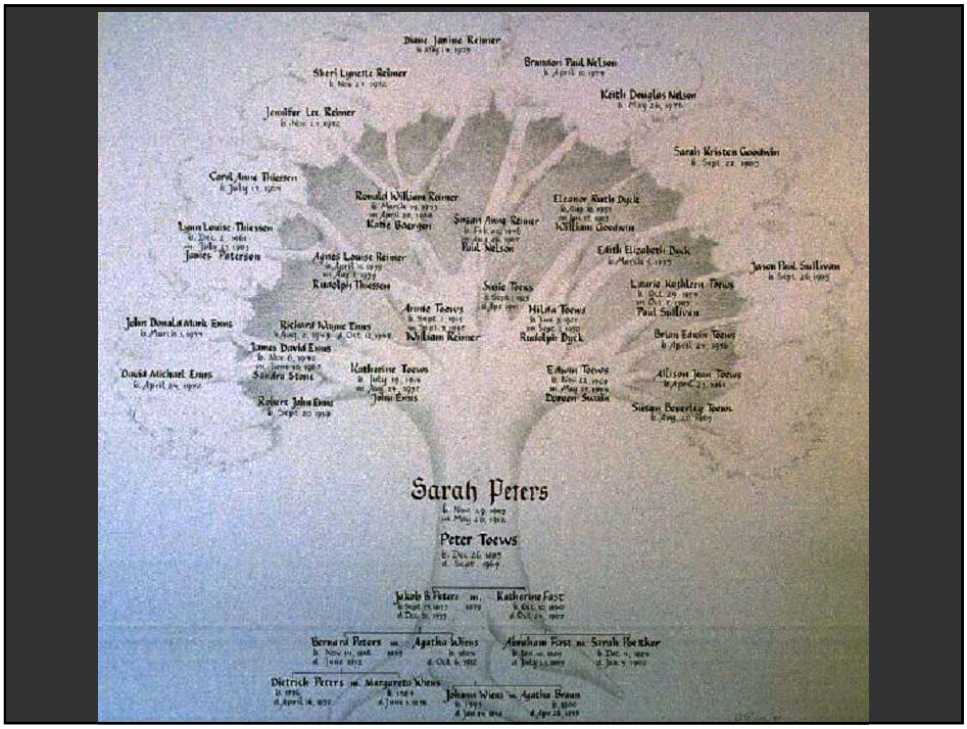
Tournaments!



Second Most Common ...

Lineages

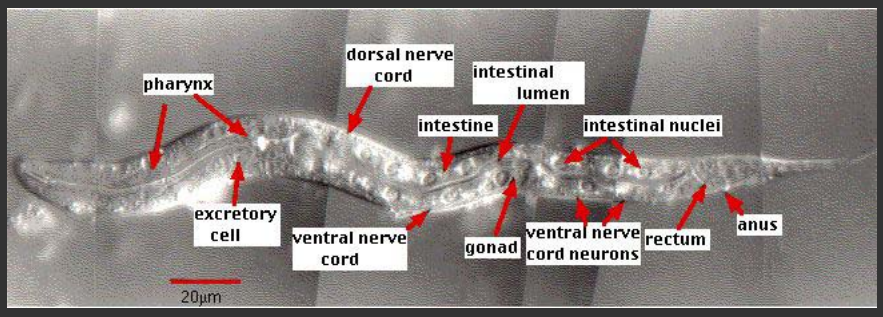

THE SCOTTISH ROYAL DYNASTIES 842–1625
HOUSE OF ALPIN
HOUSE OF BRUCE
HOUSE OF STEWART
HOUSE OF ATHOLL
HOUSE OF BALLIOL
http://www.royal.gov.uk/history/trees.htm


Sarah Peters
Peter Toews

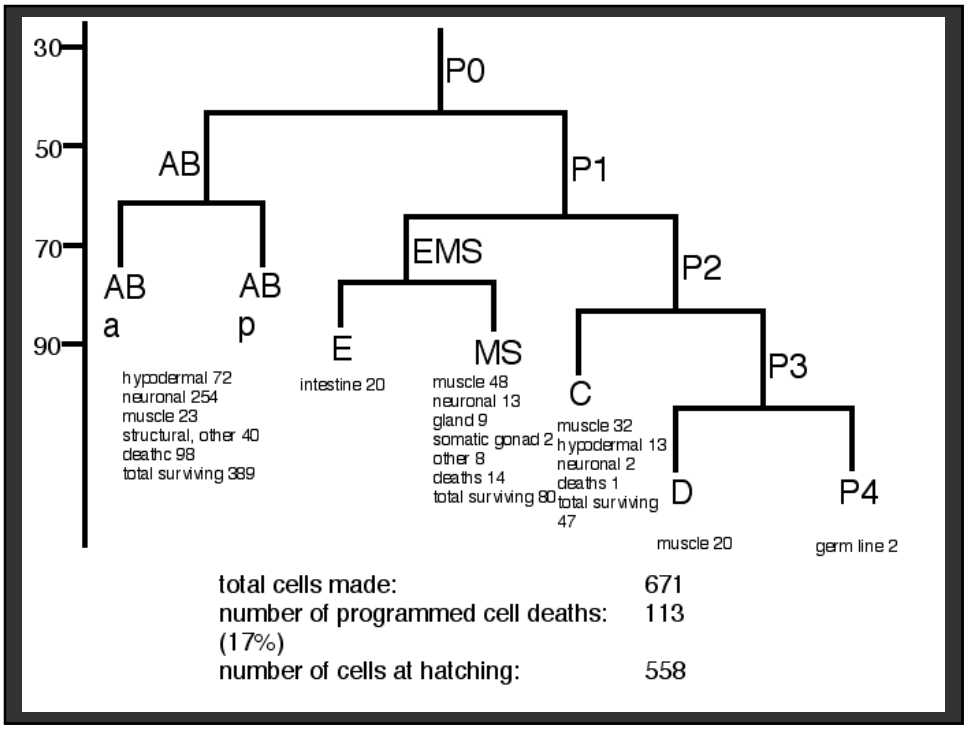
Demonstration

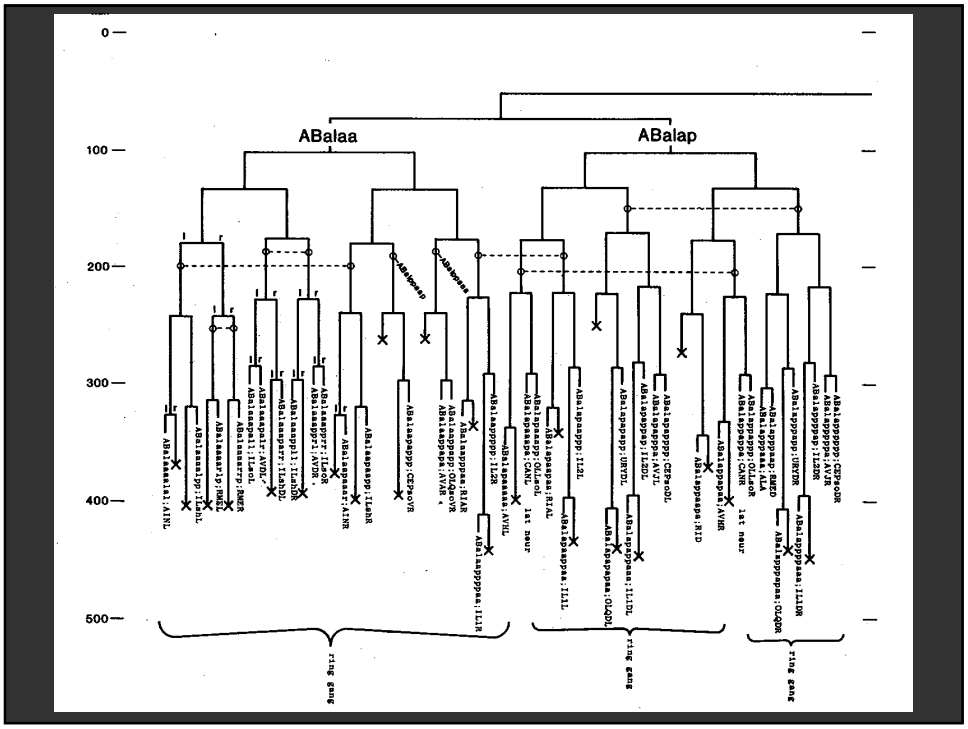
Saito-Sederberg Genealogy Viewer

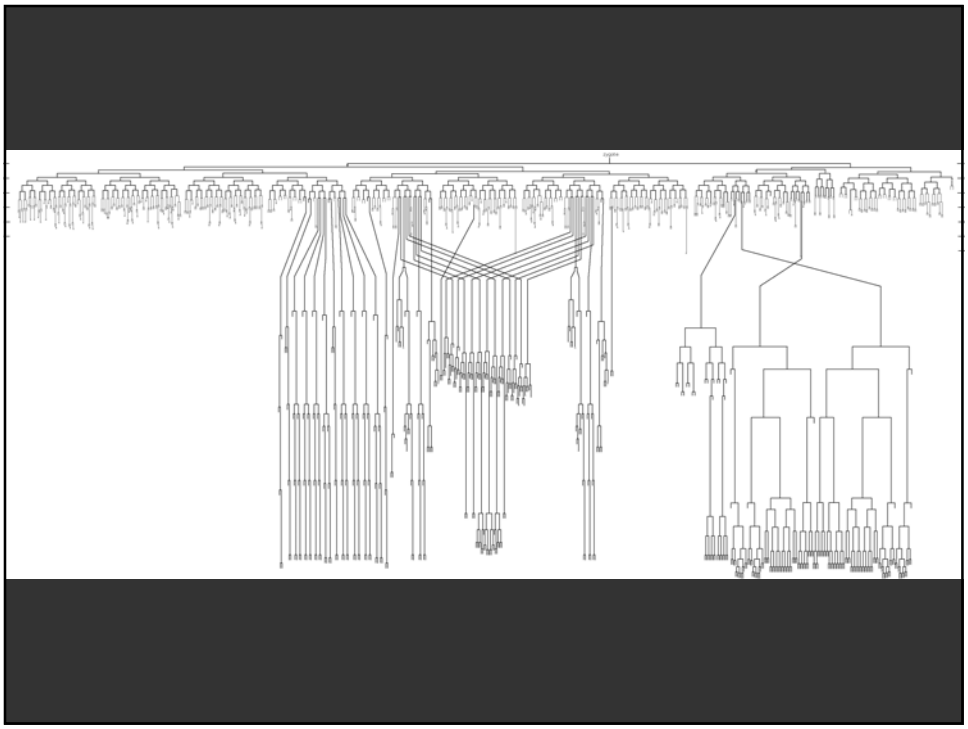
C. Elegans Cell Lineage

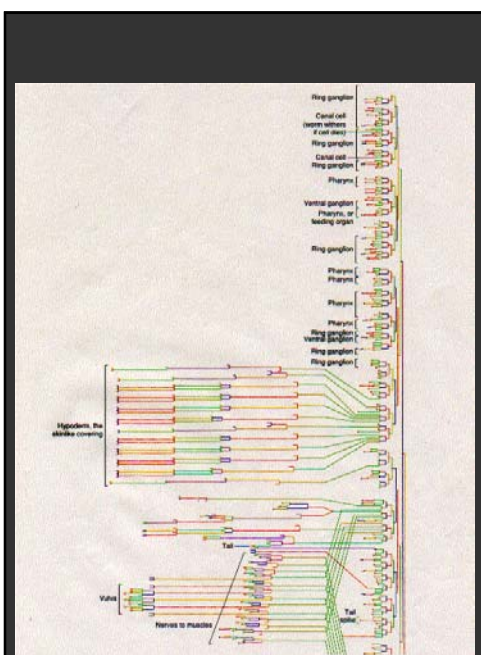


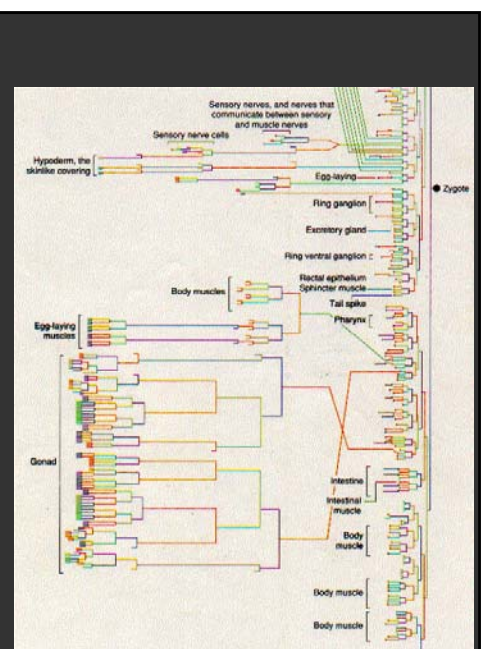
[Sulston]

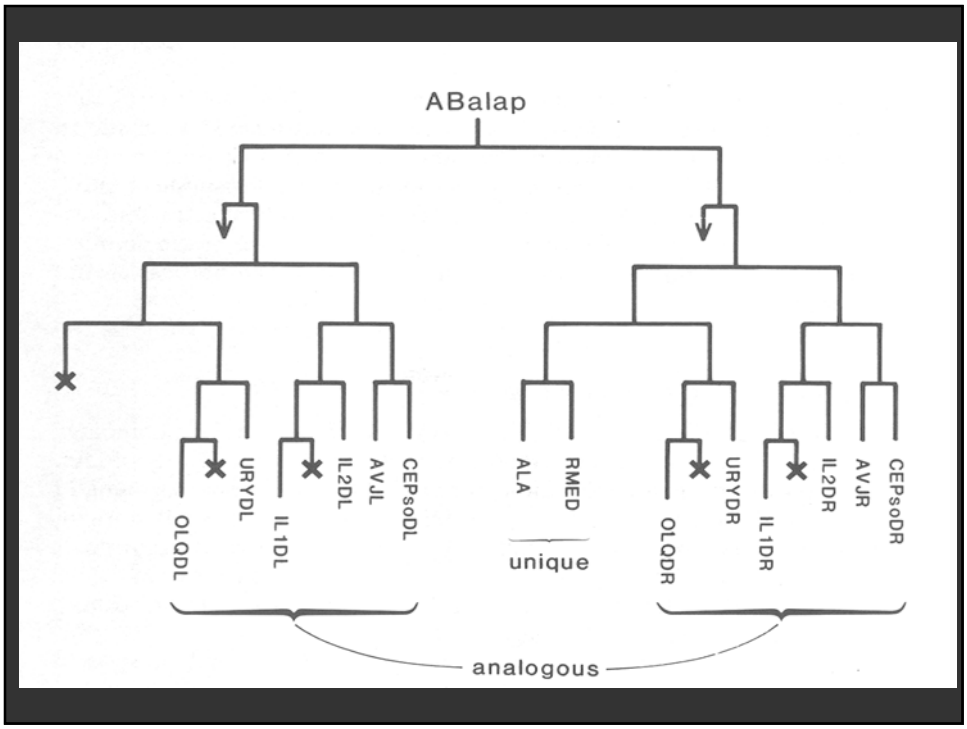

30
50
70
90
P0
AB
P1
EMS
P2
P3
AB a
AB p
E
MS
C
D
P4
hypodermal 72
neuronal 254
muscle 23
structural, other 40
deathc 98
total surviving 389
intestine 20
muscle 48
neuronal 13
gland 9
somatic gonad 2
other 8
deaths 14
total surviving 80
muscle 32
hypodermal 13
neuronal 2
deaths 1
total surviving 47
muscle 20
germ line 2
total cells made: 671
number of programmed cell deaths: 113
(17%)
number of cells at hatching: 558

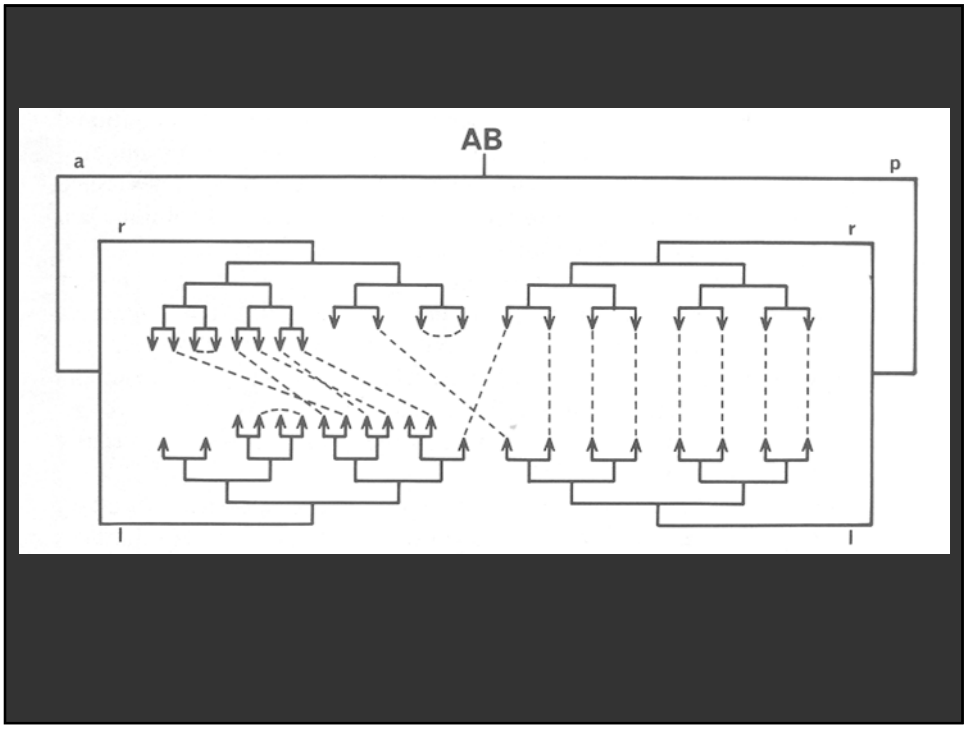

0
100
200
300
400
500
ABalaa
ABalap
ring gang
ring gang
ring gang



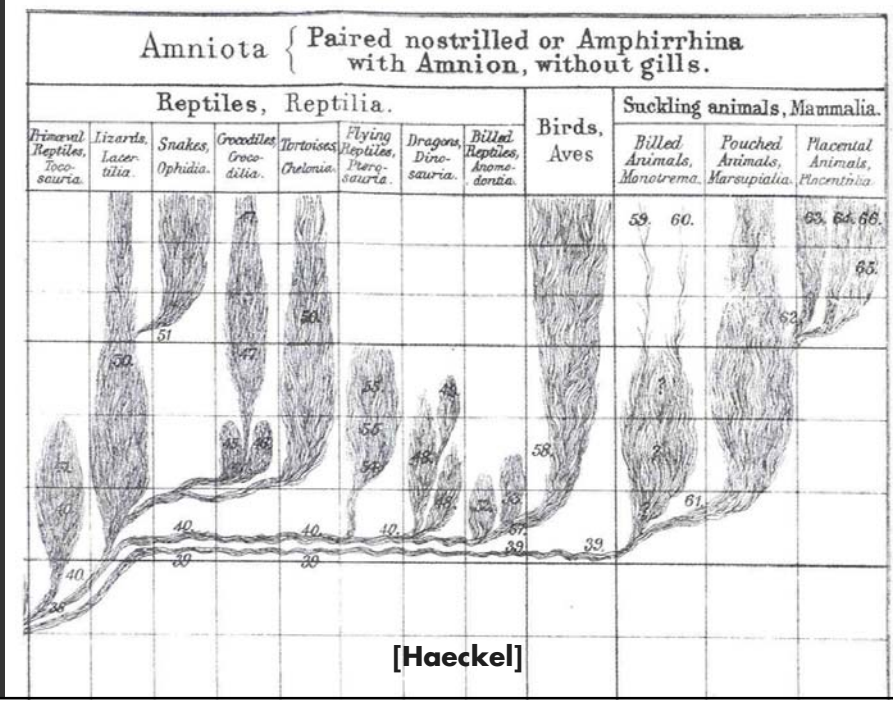

Ring ganglion
Canal cell (worm withers if cell dies)
Ring ganglion
Canal cell
Ring ganglion
Pharynx
Ventral ganglion
Pharynx, or feeding organ
Ring ganglion
Pharynx
Pharynx
Pharynx
Pharynx
Ring ganglion
Ventral ganglion
Ring ganglion
Ring ganglion
Hypoderm, the skinlike covering
Tail
Vulva
Nerves to muscles
Tail spike


Sensory nerves, and nerves that communicate between sensory and muscle nerves
Sensory nerve cells
Hypoderm, the skinlike covering
Egg-laying
Zygote
Ring ganglion
Excretory gland
Ring ventral ganglion
Rectal epithelium
Sphincter muscle
Body muscles
Tail spike
Pharynx
Egg-laying muscles
Gonad
Intestine
Intestinal muscle
Body muscle
Body muscle
Body muscle

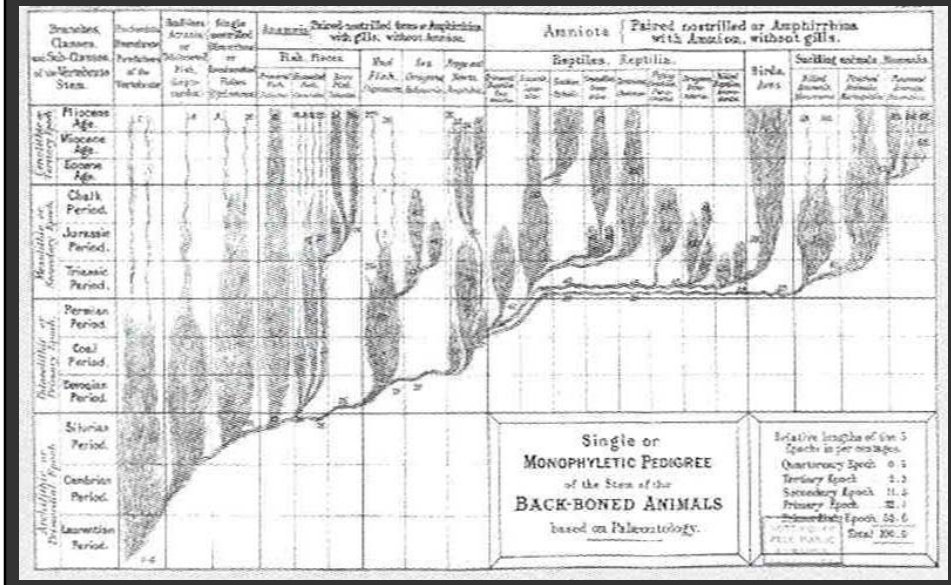

ABalap
OLQDL
URYDL
IL1DL
IL2DL
AVJL
CEPsoDL
ALA
RMED
unique
OLQDR
URYDR
IL1DR
IL2DR
AVJR
CEPsoDR
analogous

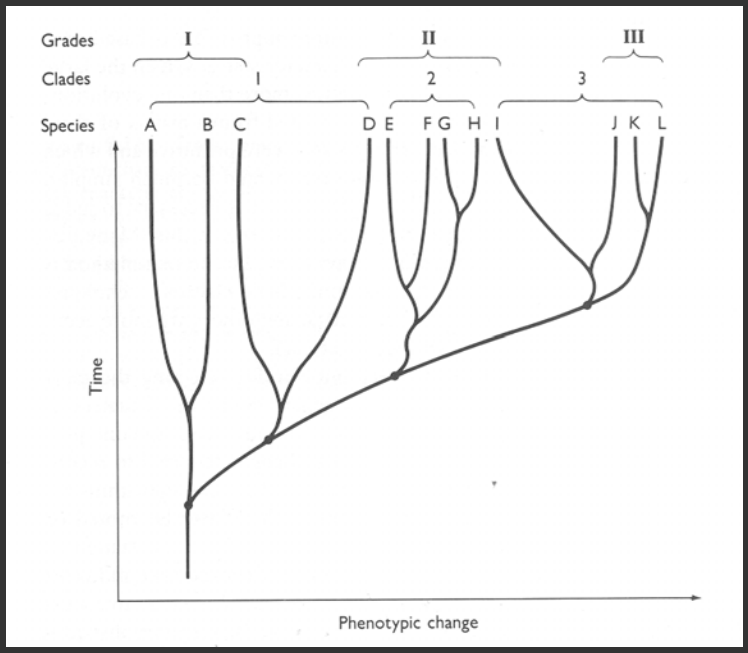

AB
a
p
r
r
l
l

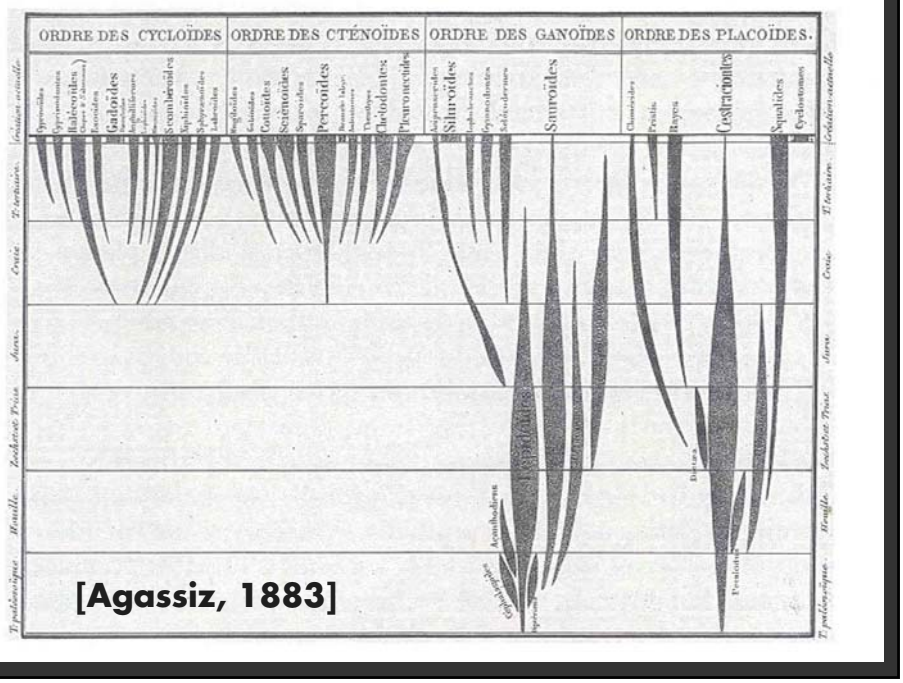
Evolutionary Trees



[Haeckel]


Single or
MONOPHYLETIC PEDIGREE
of the Stem of the
BACK-BONED ANIMALS
based on Palæontology.


Grades
I
II
III
Clades
1
2
3
Species
A
B
C
D
E
F
G
H
I
J
K
L
Time
Phenotypic change



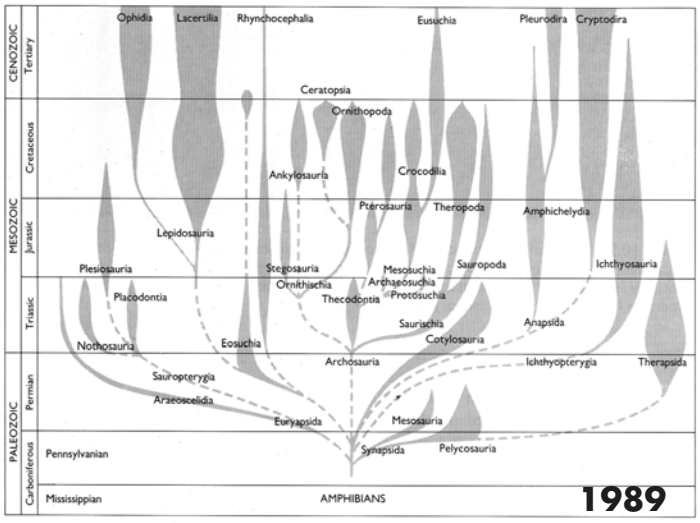
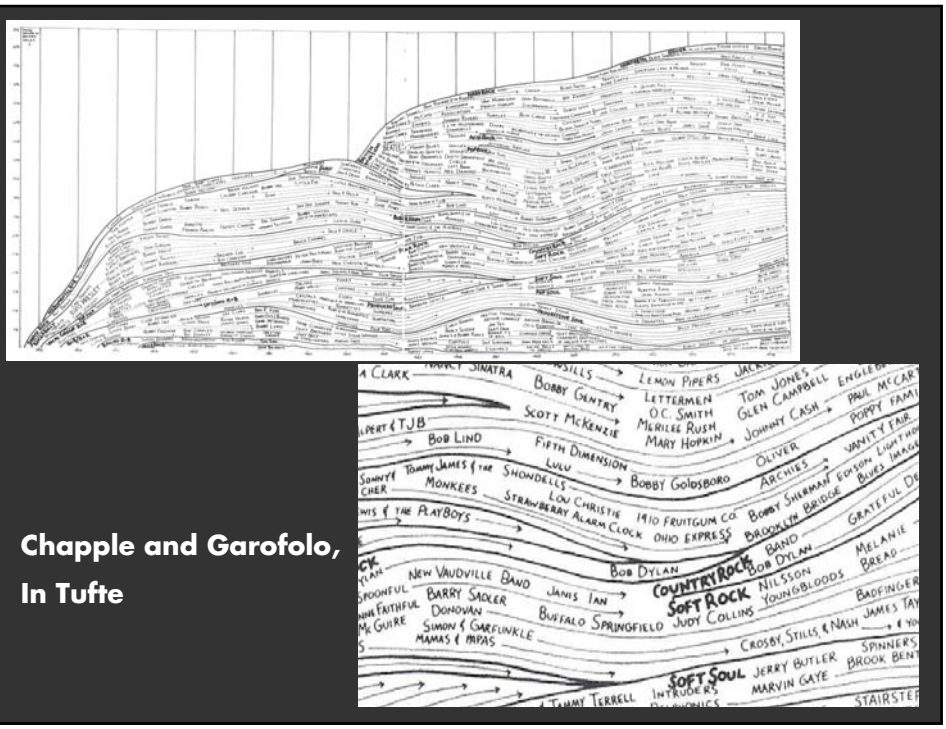
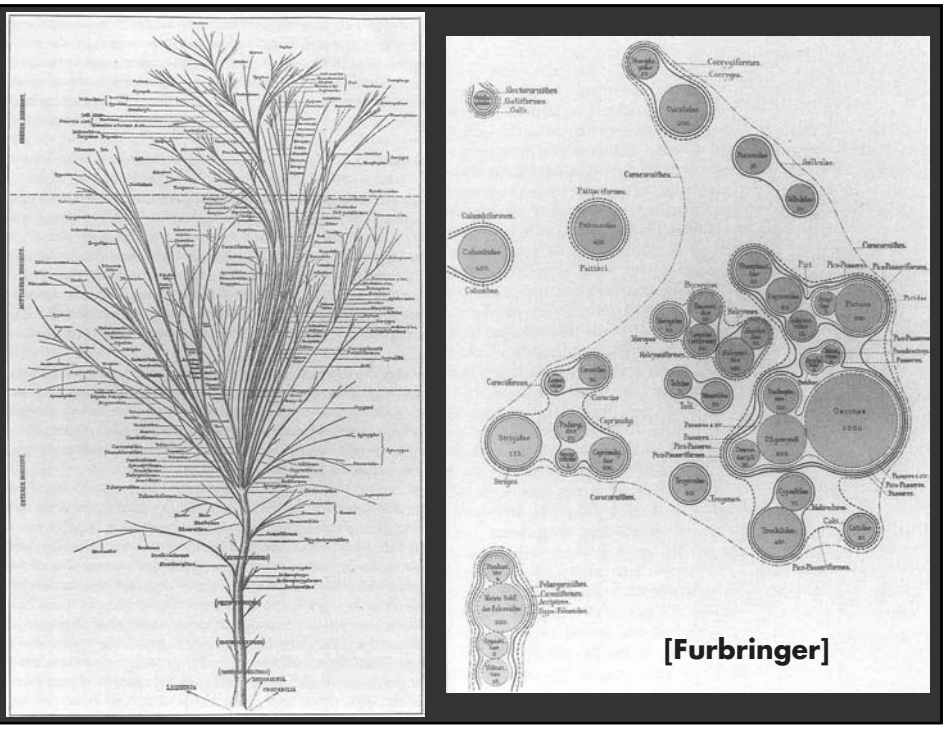
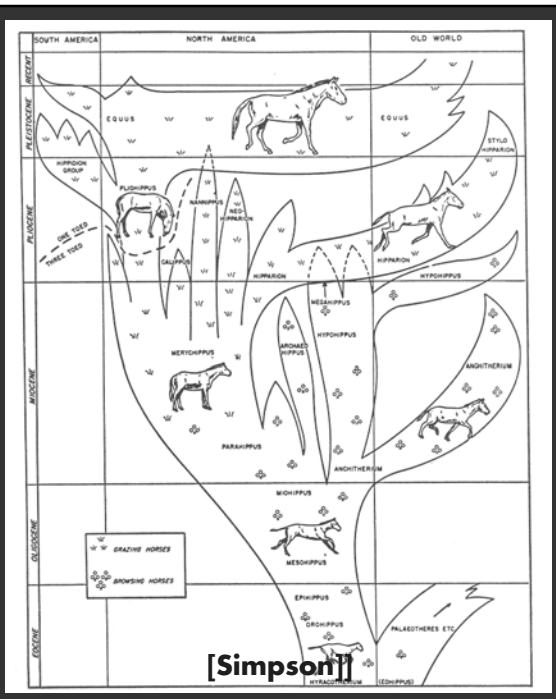


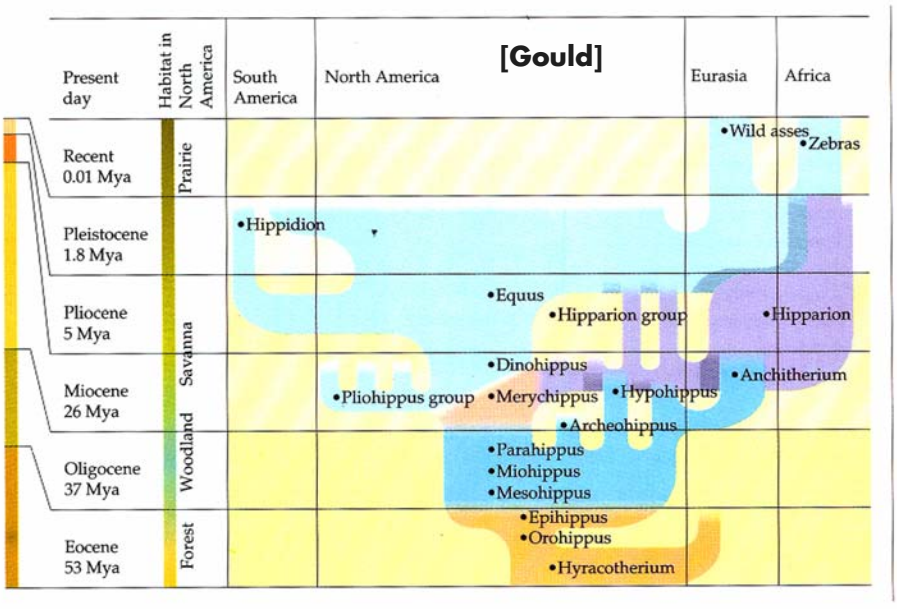
Figure 10.25. Suggested evolutionary history of the reptiles from their origin in the Carboniferous Period, through their greatest diversification in the Mesozoic Era, to the remaining living orders. Note particularly the extinctions at the end of the Triassic and end of the Cretaceous periods. The width of a line or band indicates the relative size of the group at the corresponding geologic interval. Dashed lines indicate uncertainties about relationship or time of origin. (From J. W. Valentine, "The Evolution of Multicellular Plants and Animals." Copyright © 1978 by Scientific American, Inc. All rights reserved)


Chapple and Garofolo,
In Tufte
A CLARK
SINATRA
STILLS
LEMON PIPERS
BOBBY GENTRY
LETTERMEN
O.C. SMITH
MCRILEE RUSH
MARY HOPKIN
TOM JONES
GLEN CAMPBELL
JOHNNY CASH
SCOTT MCKENZIE
BOB LIND
FIFTH DIMENSION
LULU
BOBBY GOLDSBORO
OLIVER
ARCHIES
TOMMY JAMES & THE SHONDELLS
MONKEES
LOU CHRISTIE
STRAWBERRY ALARM CLOCK
1910 FRUITGUM CO.
OHIO EXPRESS
BOBBY SHERMAN
BROOKLYN BRIDGE
THE PLAYBOYS
BOB DYLAN
BAND
NEW VAUDVILLE BAND
JANIS IAN
BARRY SADLER
DONOVAN
BUFFALO SPRINGFIELD
COUNTRY ROCK
SOFT ROCK
NILSSON
YOUNGBLOODS
JUDY COLLINS
MELANIE
BREAD
SIMON & GARFUNKLE
MAMAS & PAPAS
CROSBY, STILLS, & NASH
SOFT SOUL
JERRY BUTLER
MARVIN GAYE
TAMMY TERRELL
INTRUDERS
STAIRSTEP

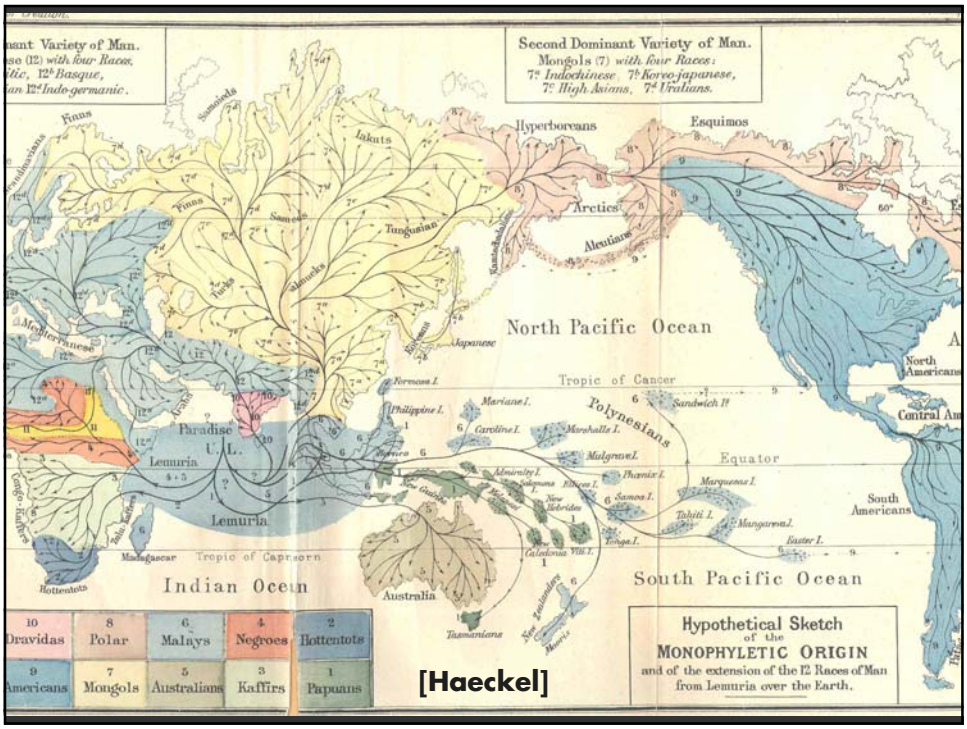

[Furbringer]



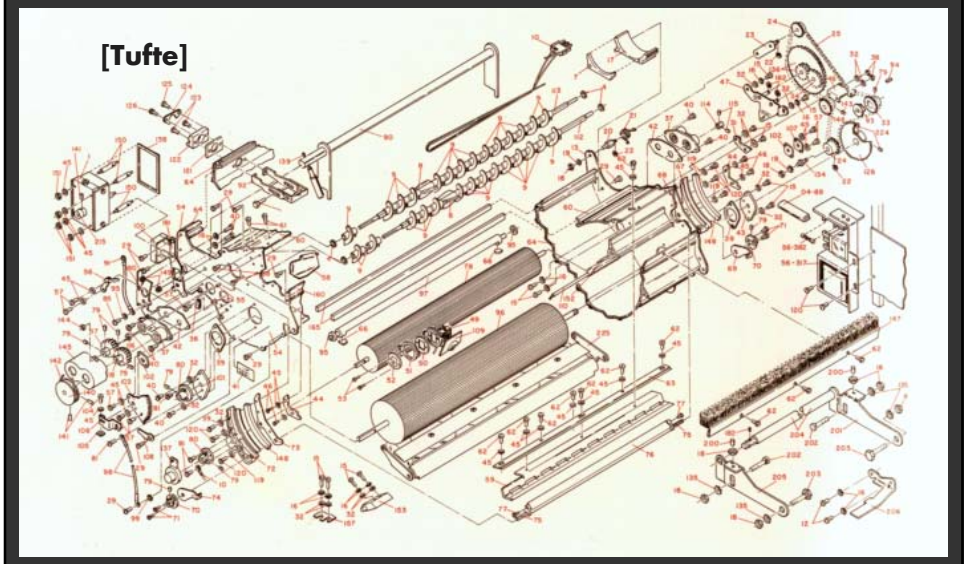
[Simpson]]



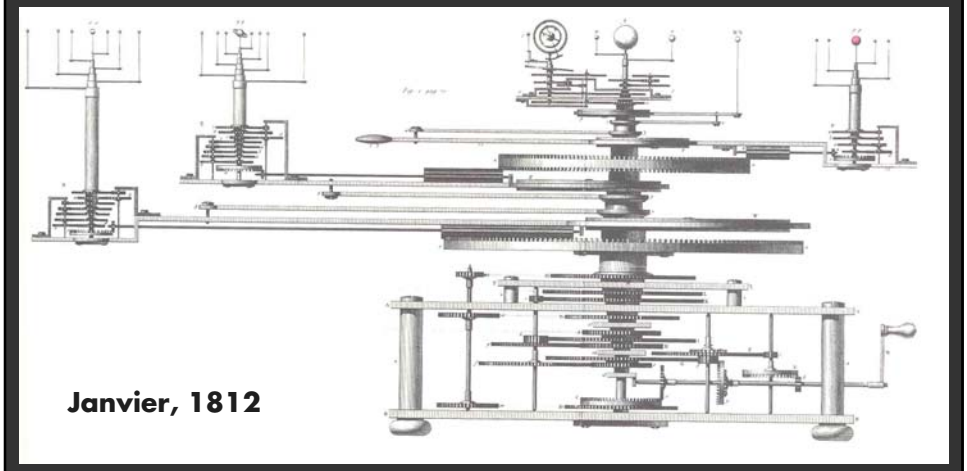
[Gould]



[Haeckel]



[Tufte]

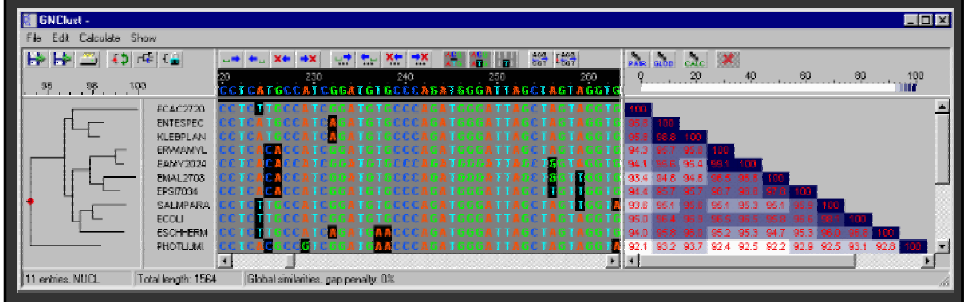


Janvier, 1812

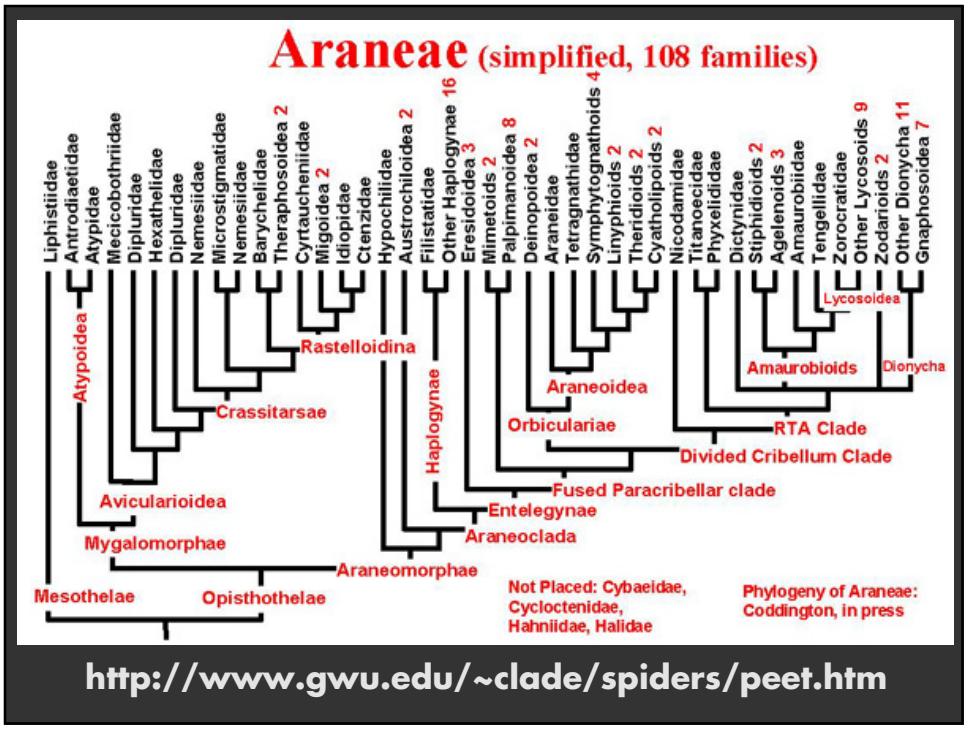
"Graphical Excellence is nearly always multivariate"

Edward Tufte

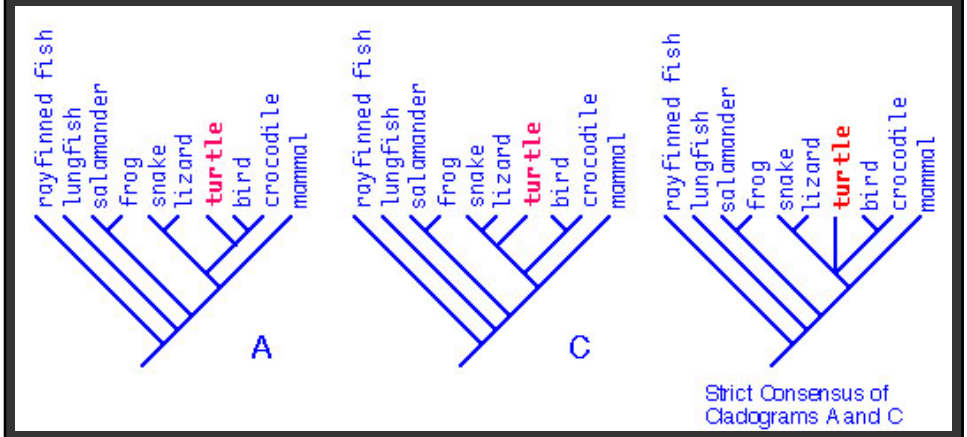
Phenograms to Cladograms

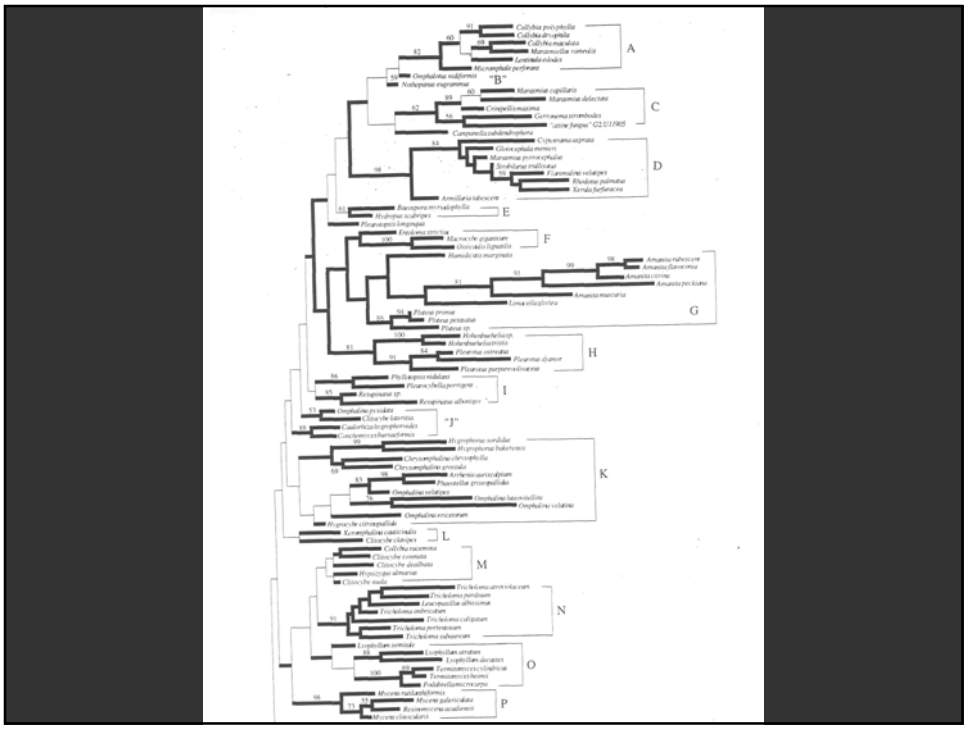


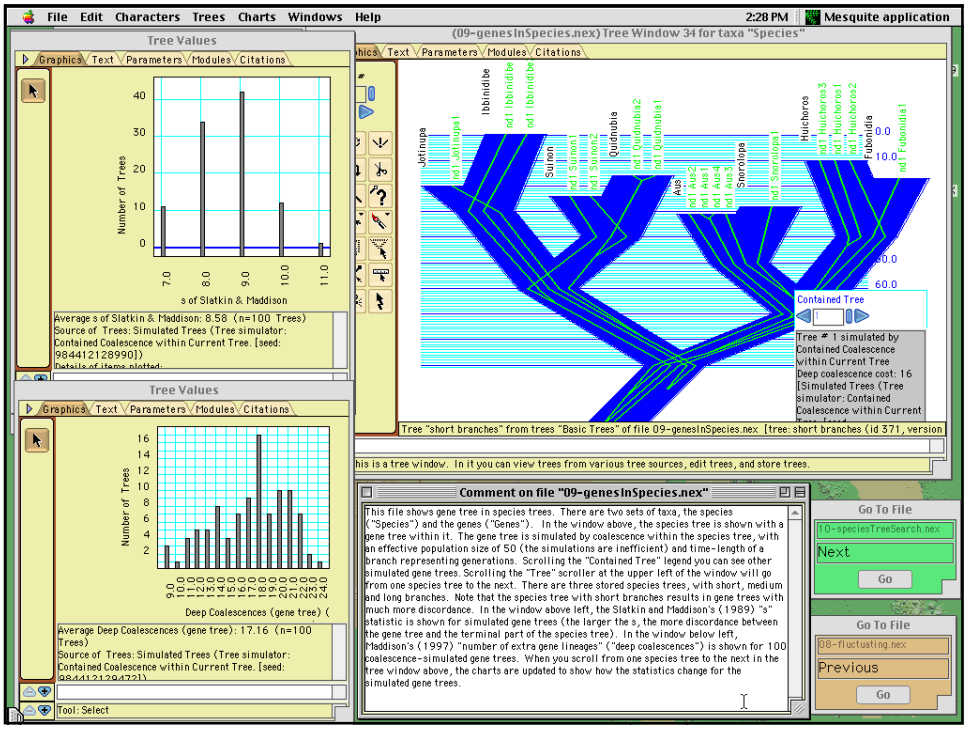
GeneBase

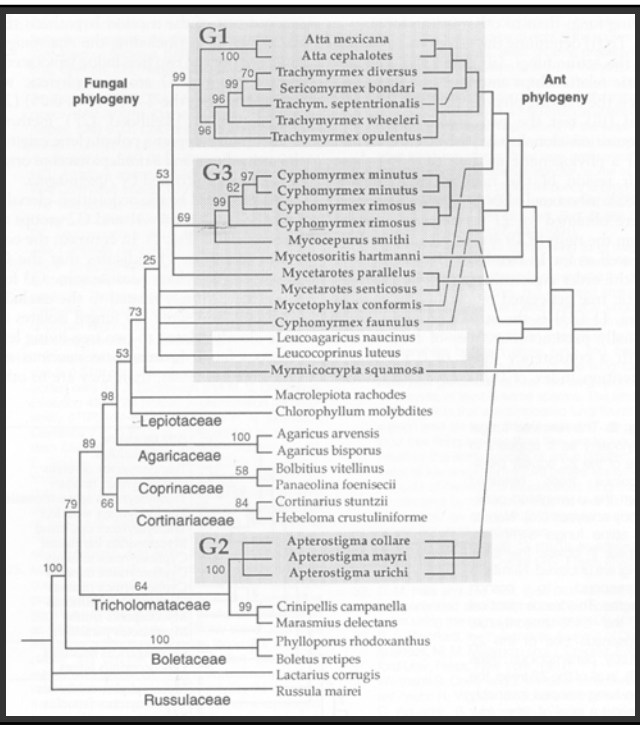


http://www.gwu.edu/~clade/spiders/peet.htm

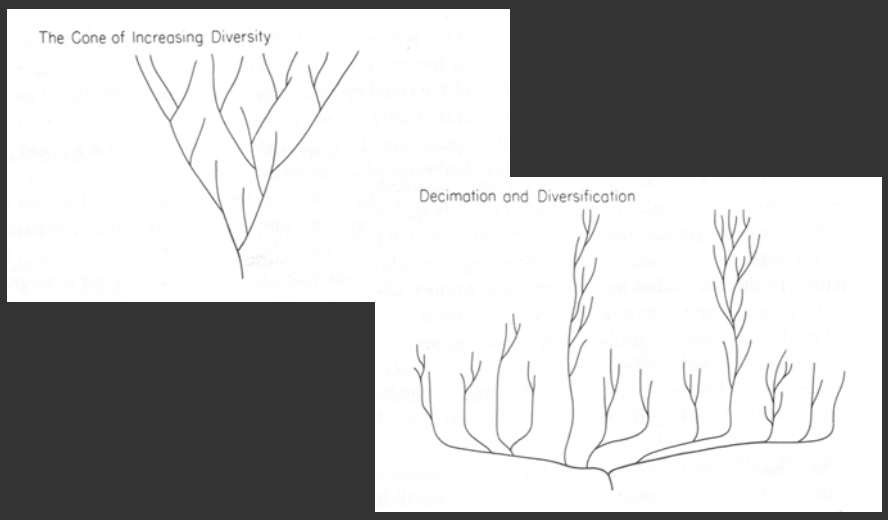




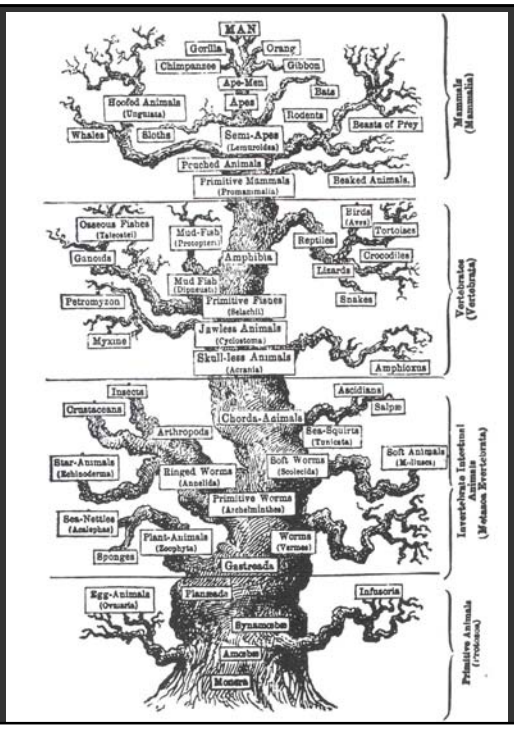

File Edit Characters Trees Charts Windows Help
2:28 PM
Mesquite application
Tree Values
Graphics Text Parameters Modules Citations
Number of Trees
s of Slatkin & Maddison
Average s of Slatkin & Maddison: 8.58 (n=100 Trees)
Source of Trees: Simulated Trees (Tree simulator: Contained Coalescence within Current Tree. [seed: 9844121289901)
Tree Values
Graphics Text Parameters Modules Citations
Number of Trees
Deep Coalescences (gene tree) (
Average Deep Coalescences (gene tree): 17.16 (n=100 Trees)
Source of Trees: Simulated Trees (Tree simulator: Contained Coalescence within Current Tree. [seed:
Tool: Select
(09-genesInSpecies.nex) Tree Window 34 for taxa "Species"
Text Parameters Modules Citations
Jetinupa
Ibbinidibe
Sulton
Quidnubia
Aus
Snorolopa
Huichoros
Fubonidia
Contained Tree
Tree # 1 simulated by Contained Coalescence within Current Tree
Deep coalescence cost: 16
[Simulated Trees (Tree simulator: Contained Coalescence within Current
Tree "short branches" from trees "Basic Trees" of file 09-genesInSpecies.nex [tree: short branches (id 371, version
his is a tree window. In it you can view trees from various tree sources, edit trees, and store trees.
Comment on file "09-genesInSpecies.nex"
This file shows gene tree in species trees. There are two sets of taxa, the species ("Species") and the genes ("Genes"). In the window above, the species tree is shown with a gene tree within it. The gene tree is simulated by coalescence within the species tree, with an effective population size of 50 (the simulations are inefficient) and time-length of a branch representing generations. Scrolling the "Contained Tree" legend you can see other simulated gene trees. Scrolling the "Tree" scroller at the upper left of the window will go from one species tree to the next. There are three stored species trees, with short, medium and long branches. Note that the species tree with short branches results in gene trees with much more discordance. In the window above left, the Slatkin and Maddison's (1989) "s" statistic is shown for simulated gene trees (the larger the s, the more discordance between the gene tree and the terminal part of the species tree). In the window below left, Maddison's (1997) "number of extra gene lineages" ("deep coalescences") is shown for 100 coalescence-simulated gene trees. When you scroll from one species tree to the next in the tree window above, the charts are updated to show how the statistics change for the simulated gene trees.
Go To File
10-speciesTreeSearch.nex
Next
Go
Go To File
08-fluctuating.nex
Previous
Go



The Shape of Trees



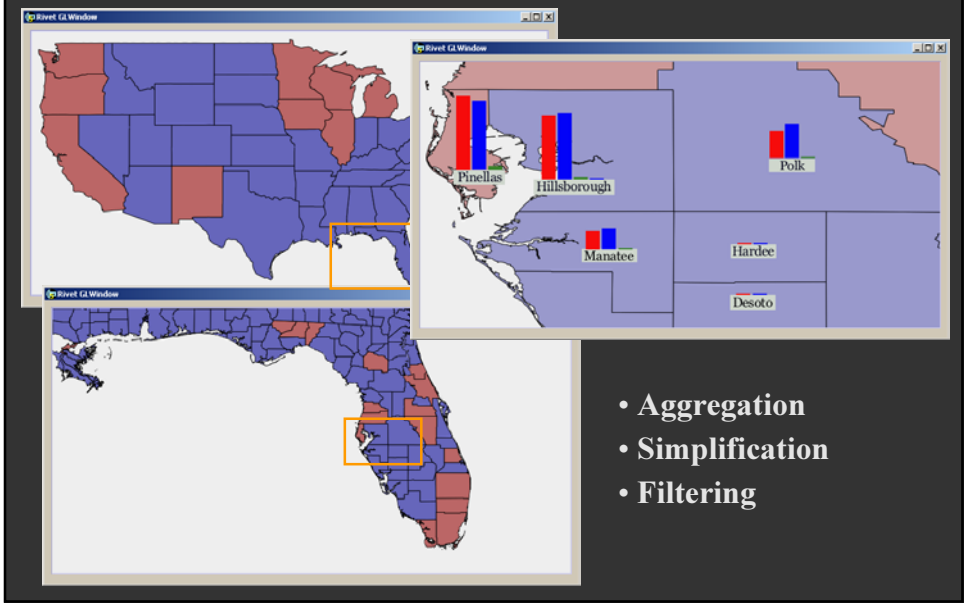
Patterns of Evolution



Hierachical Databases

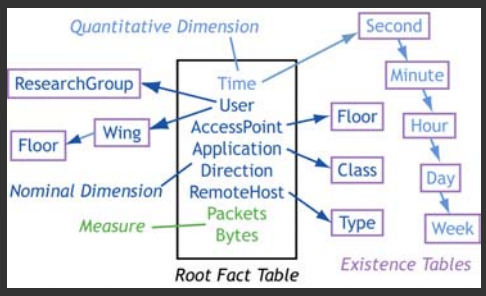
Stolte and Hanrahan, Polaris, InfoVis 2000

Generalization



- Aggregation
- Simplification
- Filtering

Abstraction Hierarchies



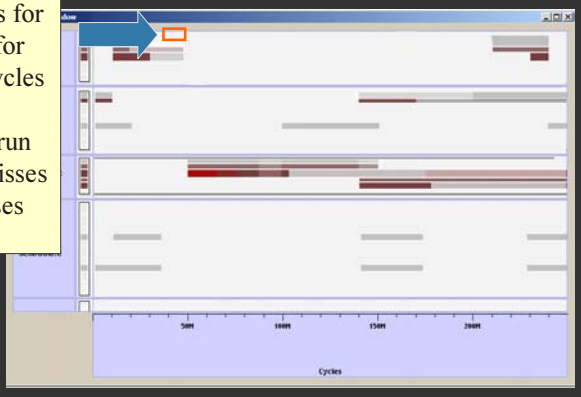
Star and Snowflake Schemes

Datacubes

Memory & Code

Cache misses for a procedure for 10 million cycles

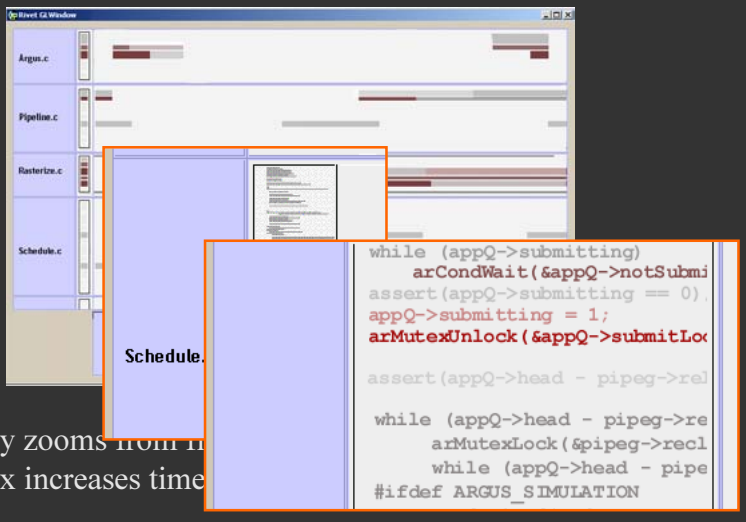
White = not run
Grey = no misses
Red = # misses



y-dimension is source code

x-dimension is cycles (time)

Memory & Code



zooming on y zooms from
zooming on x increases time

Themes

Cognitive Principles for Design

Congruence Principle: The structure and content of the external representation should correspond to the desired structure and content of the internal representation.

Apprehension Principle: The structure and content of the external representation should be readily and accurately perceived and comprehended.

[B. Tversky in press]

Metaphors

- Survey the space of designs
- Best metaphors have been refined over time (and have stood the test of time)
- Need to understand and appreciate what makes metaphors effective
- It takes both creativity and experience, algorithms and aesthetics, to create new metaphors. Often one element is lacking
- "Graphical excellence is multivariate," therefore mix your metaphors

Depiction

How to map information to graphics

- Central problem in visualization
- Formal systems needed
- Automation and composition
- Need high-level tools to make this easy

Key theorists: Bertin, Cleveland, Mackinlay, MacEachren, Wilkinson

Challenge

Best visualizations designed by humans

Computer mediated communication is becoming ubiquitous

Therefore: Visualizations are regressing

Challenge: Develop algorithms that produce high-quality, effective designs

Acknowledgements

Tamara Munzner

Chris Stolte

Diane Tang

Maneesh Agrawala

Barbara Tversky