

# FROM DATA TYPES TO VISUAL CUES

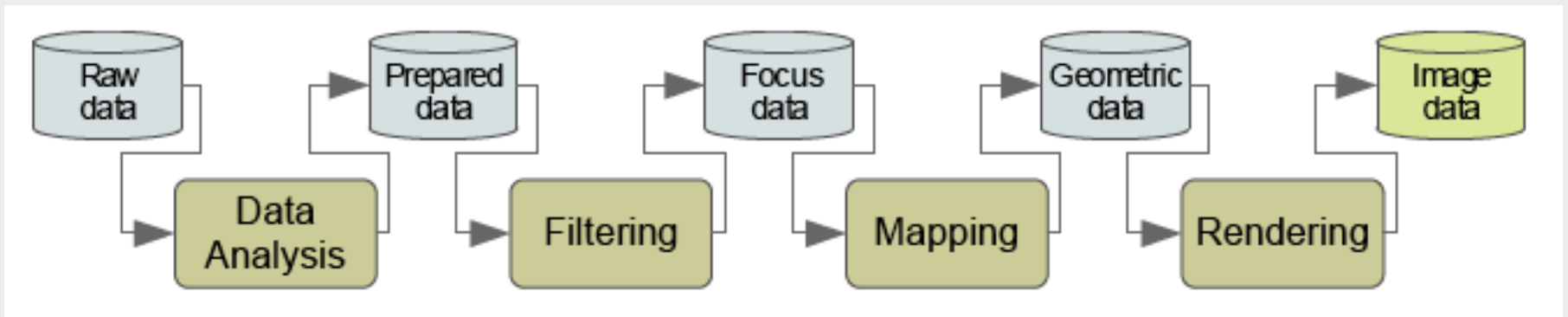
# VISUALIZATION PIPELINE

## TODAY'S MENU:

Data

Visual cues

Mapping between them



# DATA SOURCES

## REAL WORLD DATA

Physics, Biology, Chemistry  
Economy, Sociology, Psychology

## SYNTHETIC DATA

Simulations

## DERIVED DATA

From processing real-world data

---

# EXAMPLE OF A SIMPLE DATA STRUCTURE

## CAR TRAFFIC DATA:

Day	(string from a set)
Town name	(arbitrary string)
Road number	(integer)
Number of cars / hour	(integer)
Outside temperature	(float)

---

# EXAMPLE OF A SIMPLE DATA STRUCTURE

## CAR TRAFFIC DATA:

Day	nominal, ordinal
Town name	nominal, categorical
Road number	discrete, categorical
Number of cars / hour	discrete, ordinal
Outside temperature	continuous, ordinal

---

# SIMPLE DATA

## NUMERICAL

Numbers, identifiers, ranges, intervals

## NOMINAL

Names, codes

## ORDINAL

Can be ordered

## CATEGORICAL

Groups with no meaningful ordering

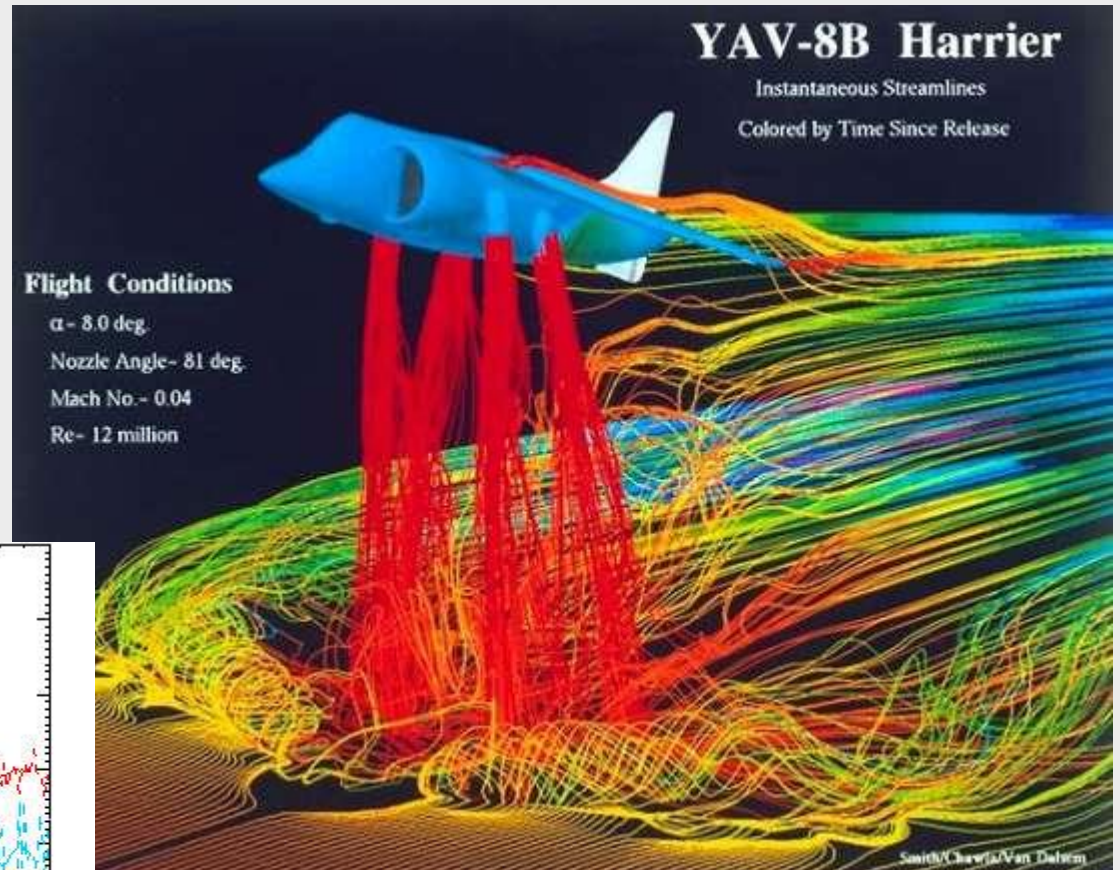
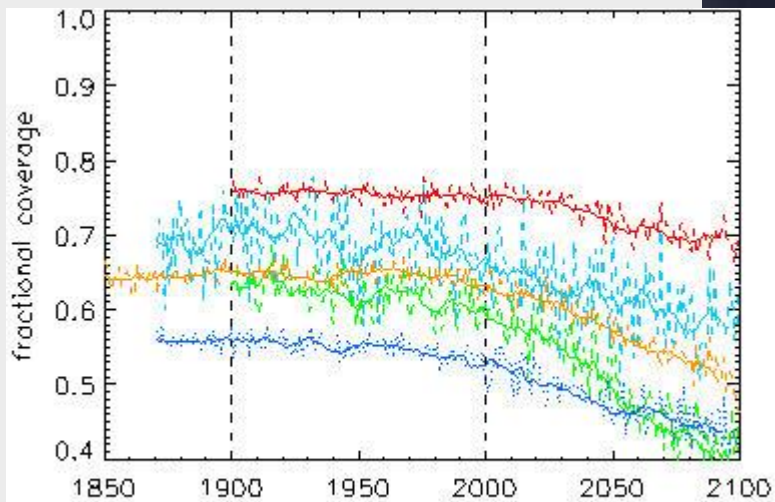
## CONTINUOUS / DISCRETE

Arbitrary / finite set of values

---

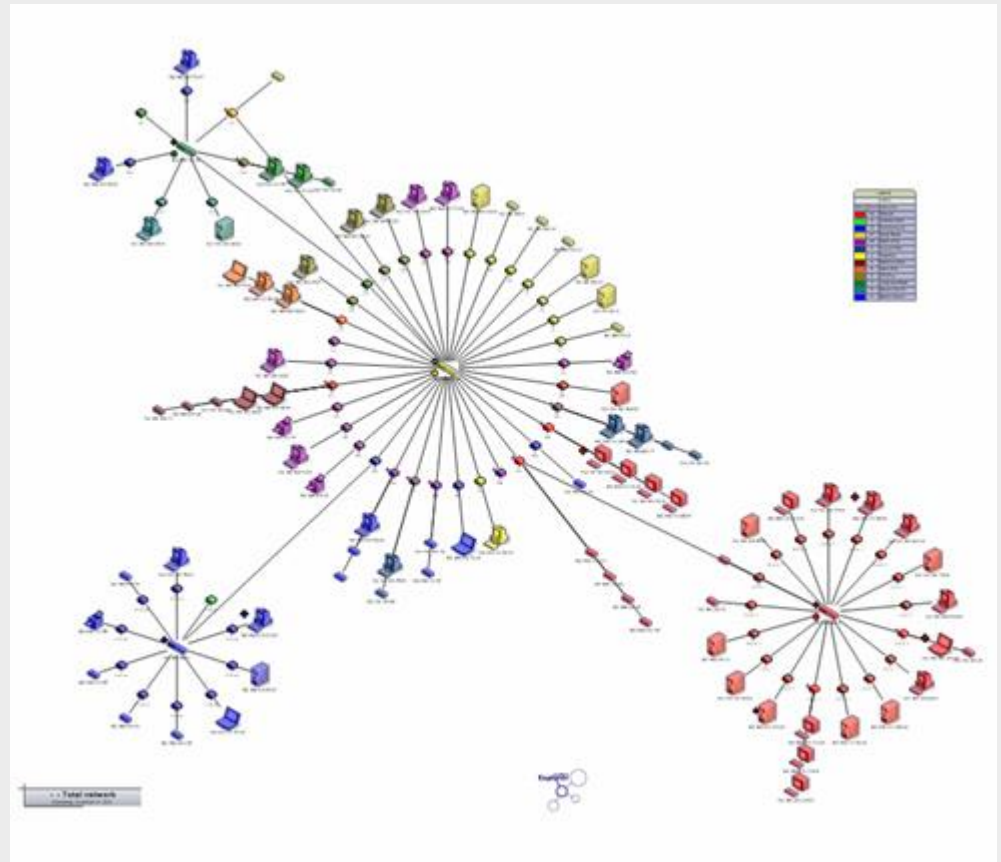
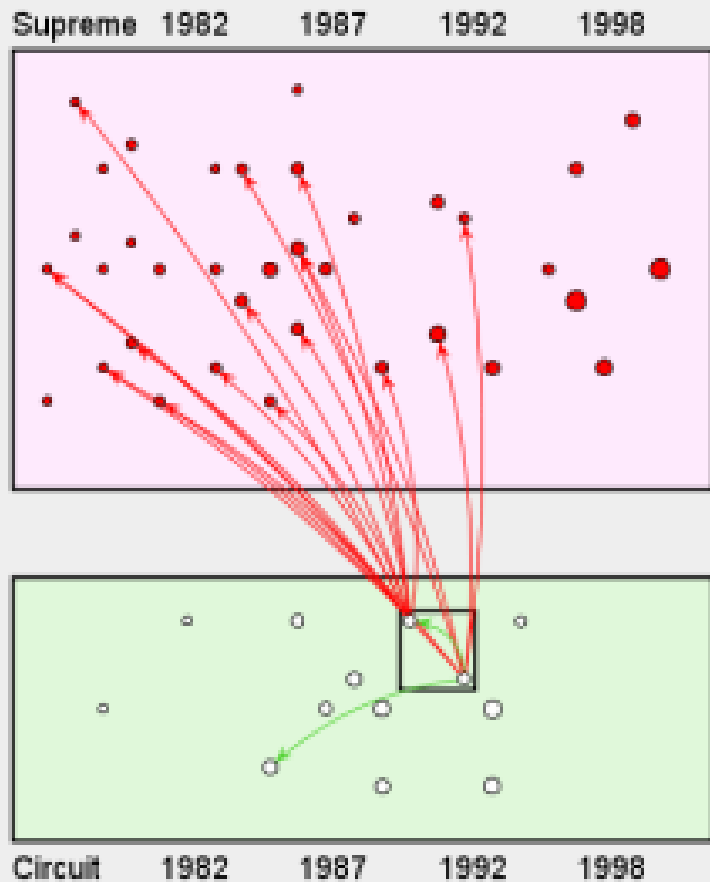
# COMPOUND DATA

VECTORS  
TENSORS  
TIME SERIES



# ABSTRACT DATA

## CONNECTIONS, RELATIONSHIPS





# ABSTRACT DATA – TYPES OF CONNECTIONS

TREES, NETWORKS

EDGE TYPES

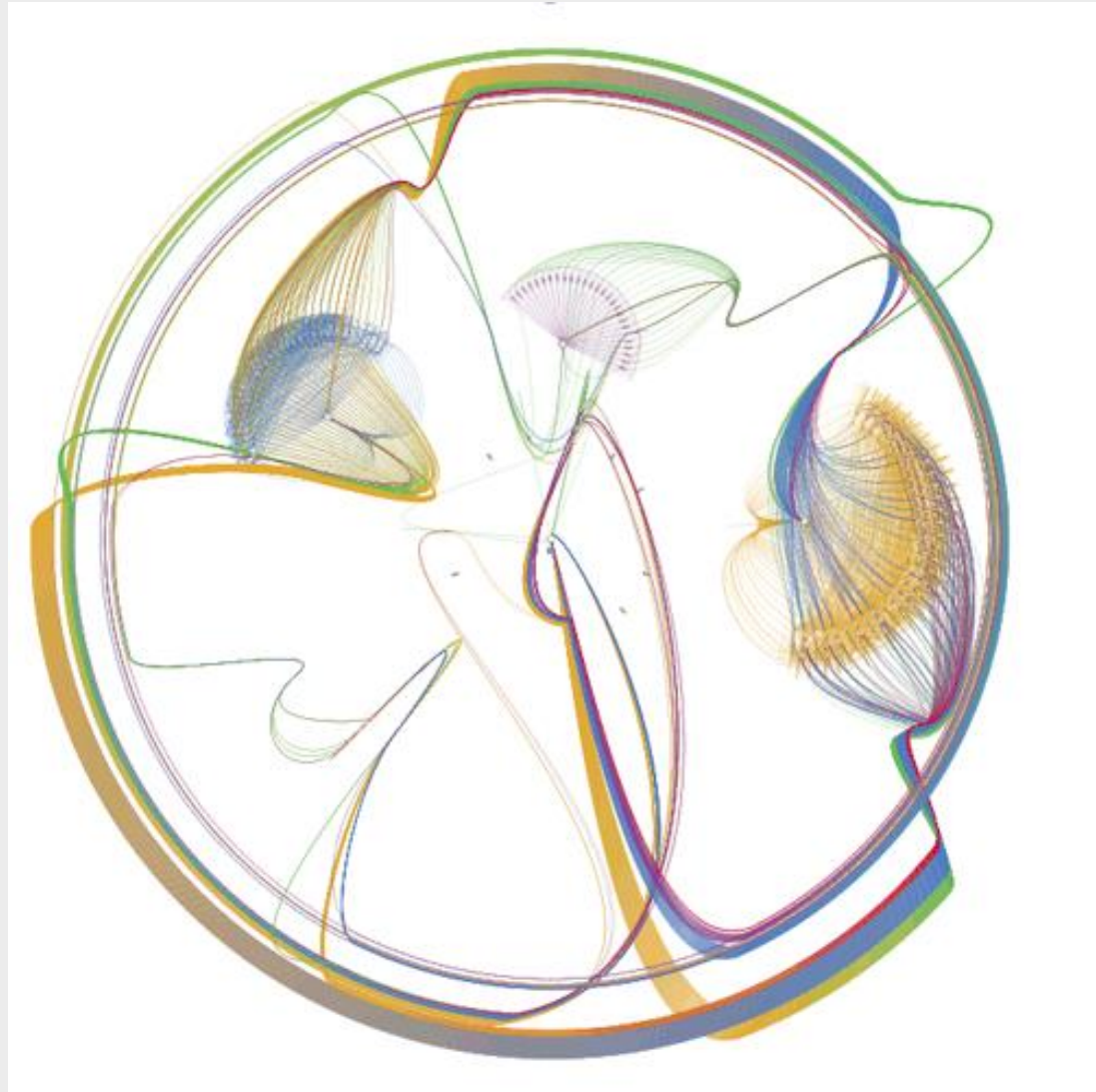
Orientation

Type of relation

ENTITIES TYPES

Nodes

Leaves



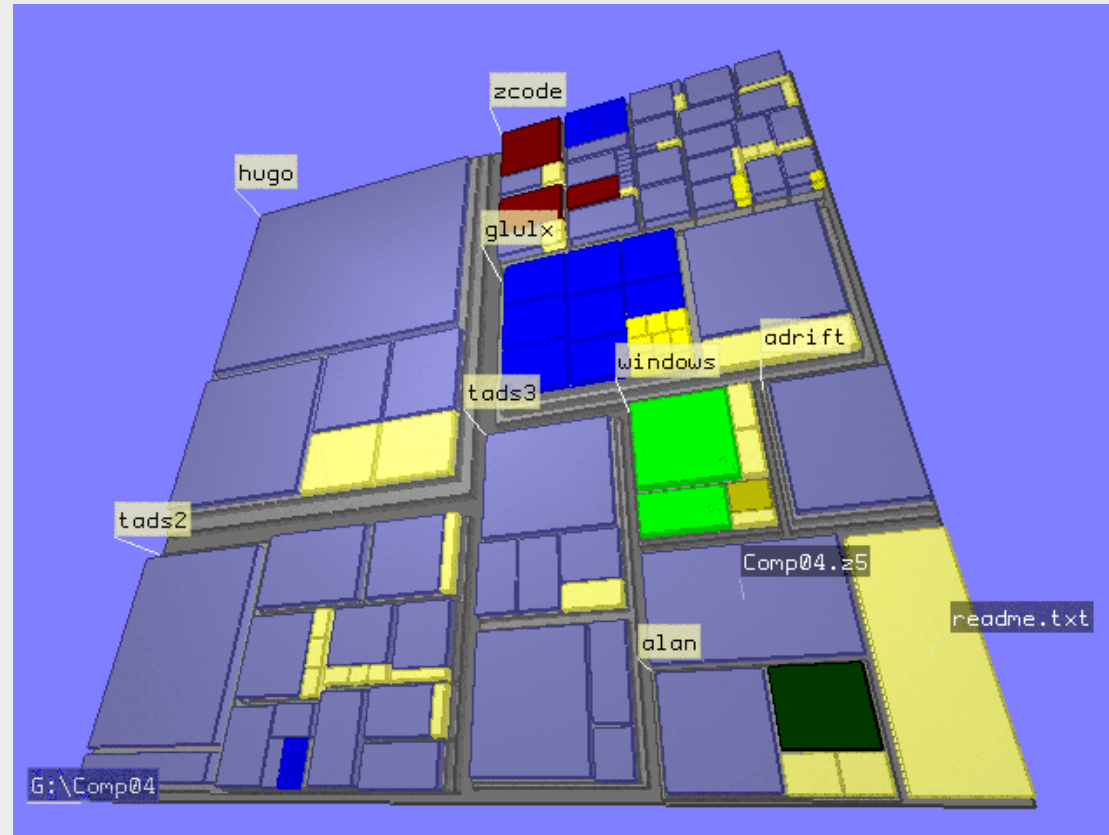
# ABSTRACT DATA – TYPES OF STRUCTURES

HIERARCHY

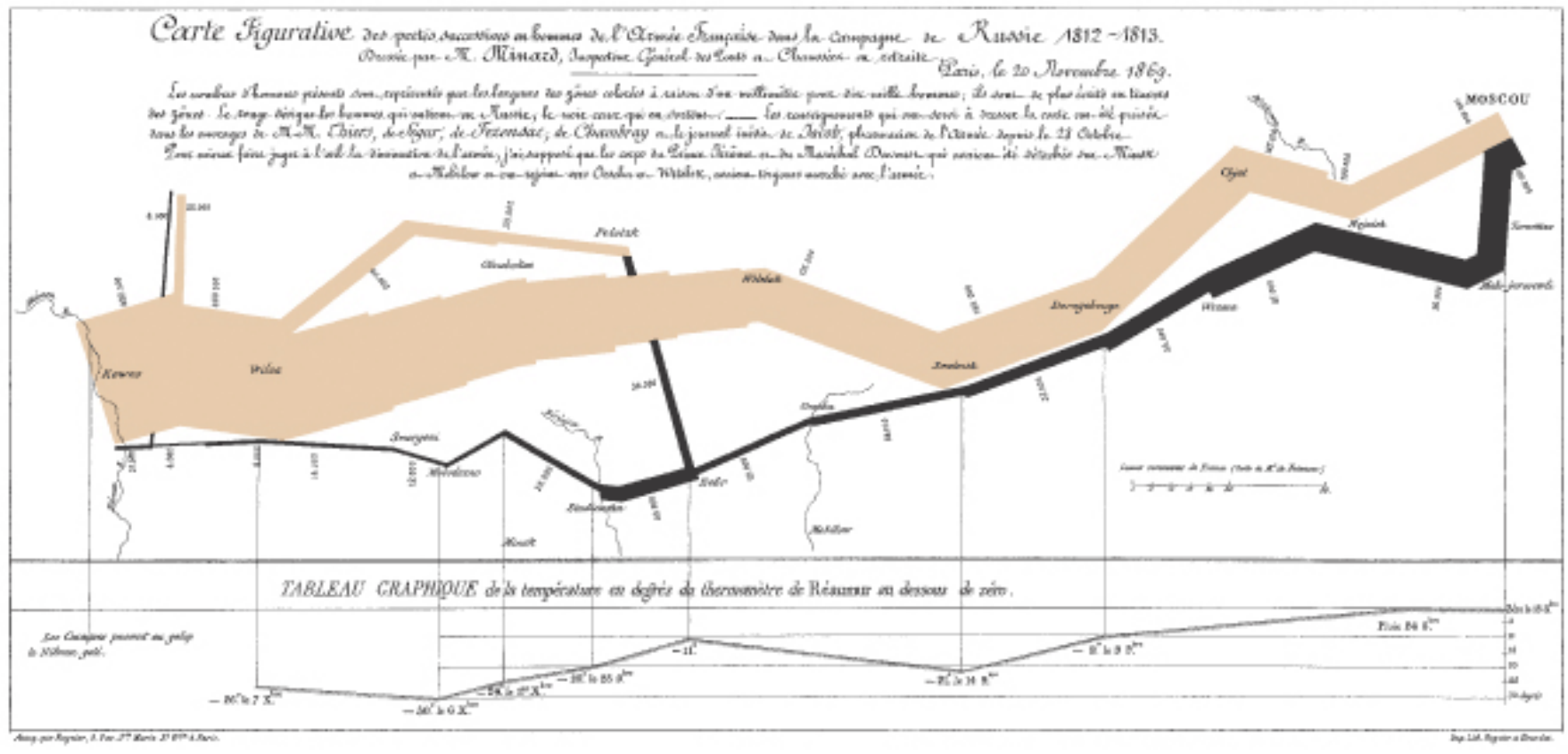
ORIENTED  
RELATIONSHIPS

CAUSALITY

DEPENDENCY



# ABSTRACT DATA – TEMPORAL ASPECT



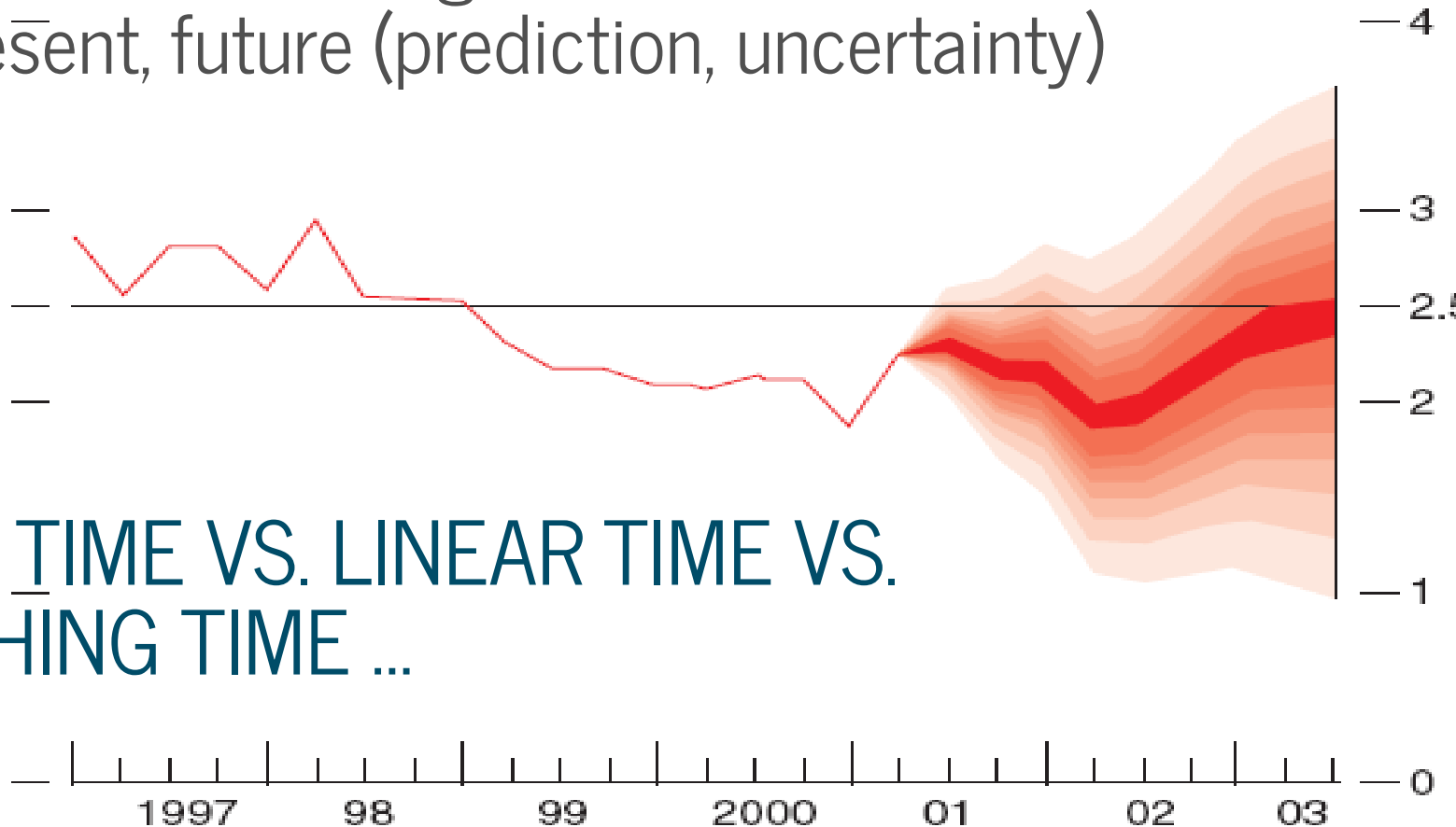
# ABSTRACT DATA – TYPES OF TIME

## TIME-DEPENDENT DATA

Snapshots in time (e.g. anonymous censuses)

Evolution of entities (e.g. website traversal)

Past, present, future (prediction, uncertainty)



CYCLIC TIME VS. LINEAR TIME VS.  
BRANCHING TIME ...

**VISUAL LANGUAGE  
IS BUILT FROM  
WORDS**

**WORDS ARE BUILT  
FROM LETTERS**

**VISUAL CUES**

# SOME BASIC VISUAL ATTRIBUTES

LENGTH

AREA / VOLUME

COLOR

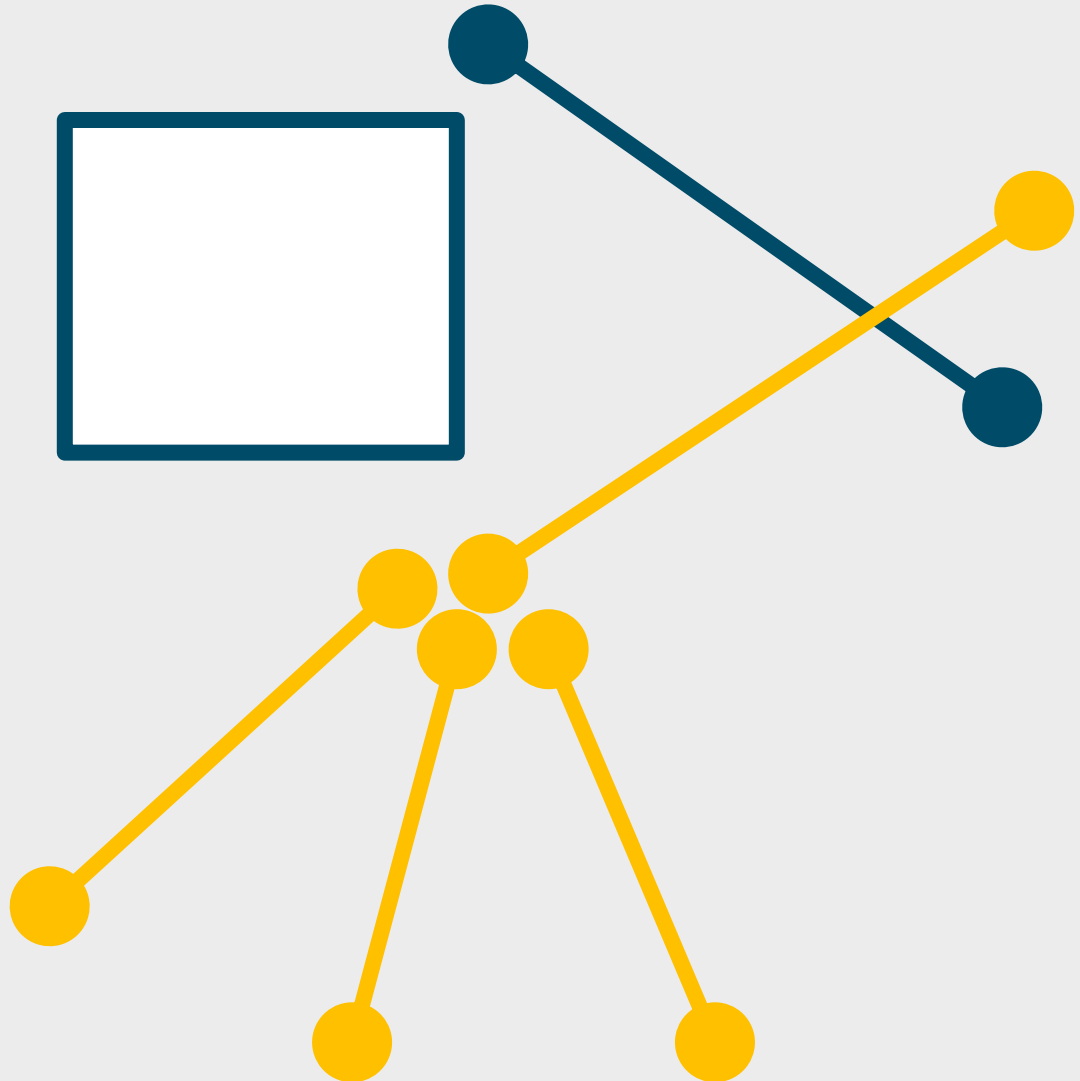
ANGLE

CONNECTIVITY

HIERARCHY

DISTANCE

DENSITY



# QUICK RECOLLECTION

## PRE-ATTENTIVE CUES:

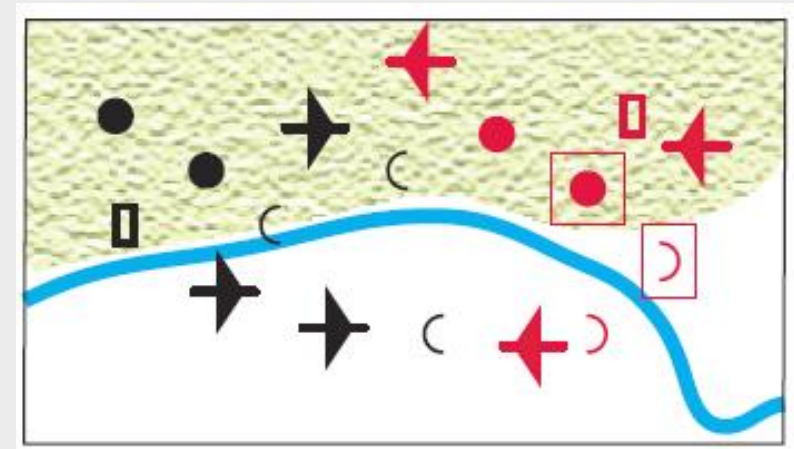
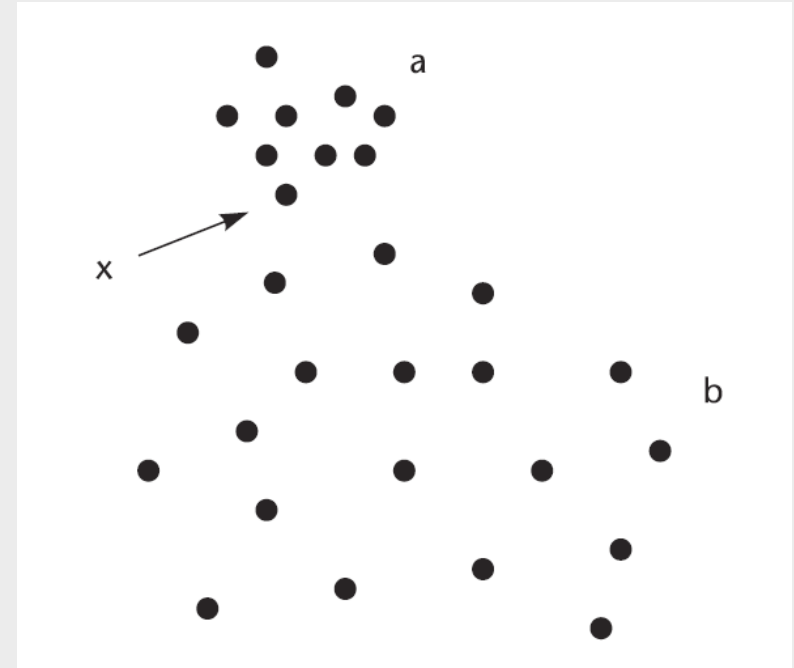
Color, form, orientation,...

## SYMBOLIC CUES

Learned symbols

## PATTERNS:

Connectivity, continuity,  
Symmetry, lines, edges, ...



“MERE COLOR, UNSPOILED BY  
MEANING, AND UNALLIED WITH  
DEFINITE FORM, CAN SPEAK TO  
THE SOUL IN A THOUSAND  
DIFFERENT WAYS.”

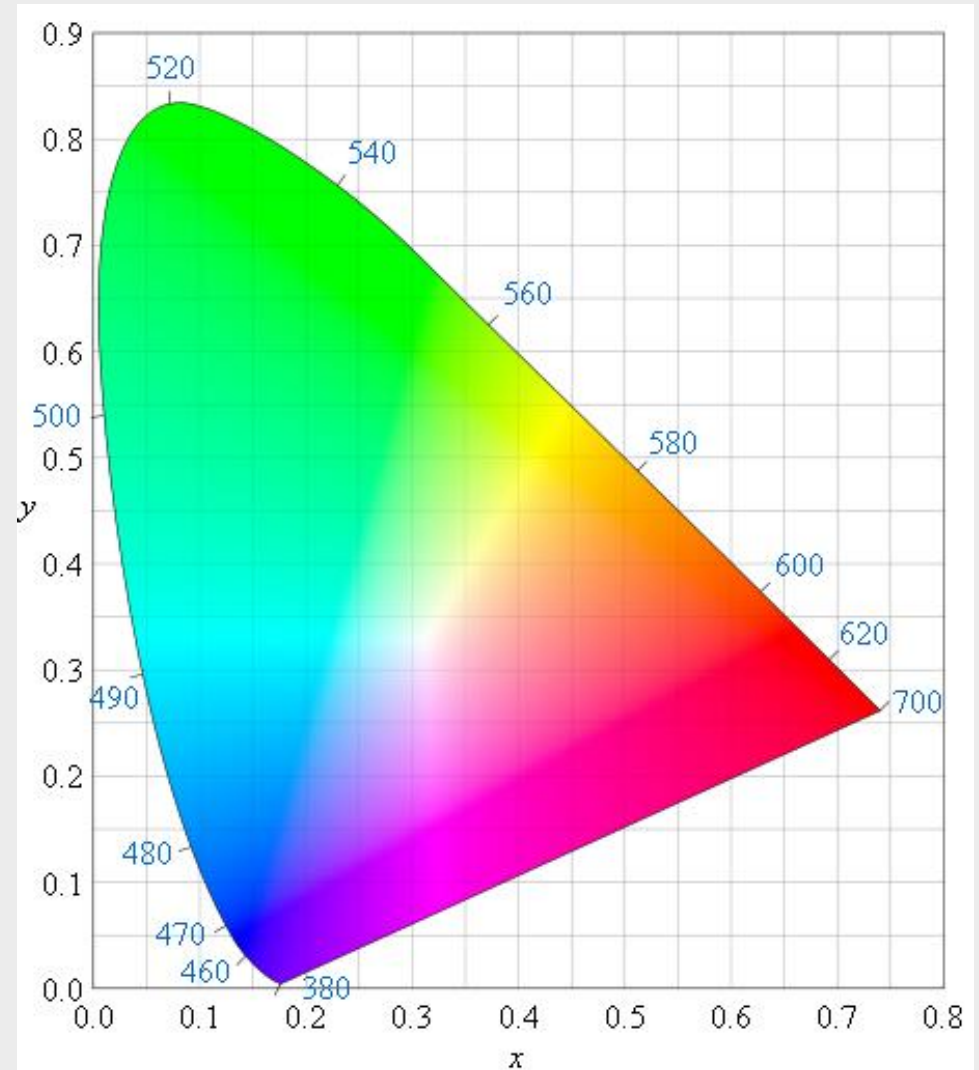
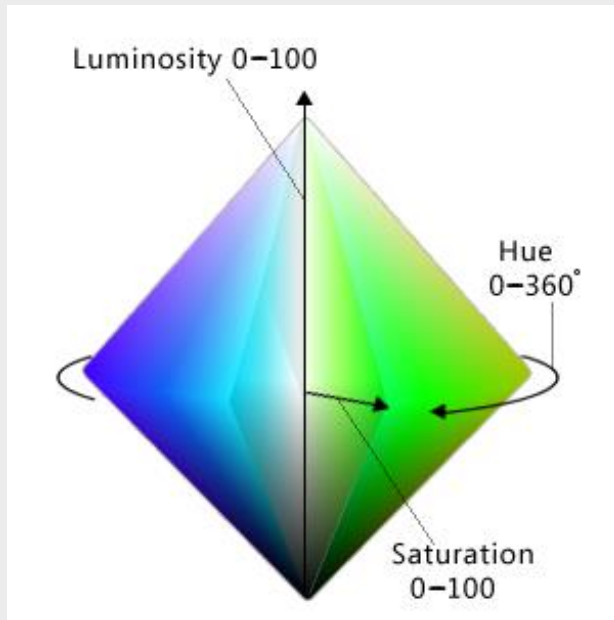
OSCAR WILDE

# COLOR IN VISUALIZATION



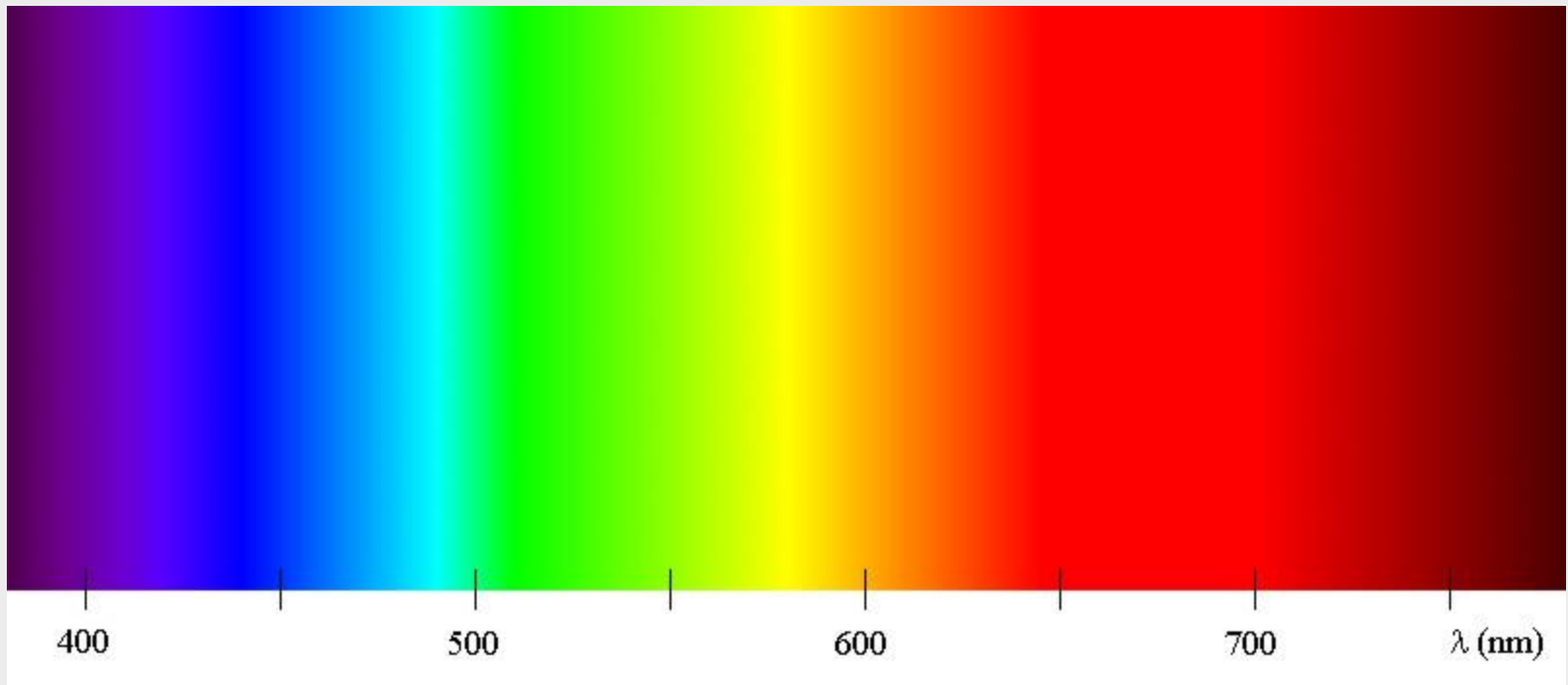
# COLOR MODELS IN COMPUTER GRAPHICS

RGB, CMYK  
CIEXYZ, CIELUV  
HSV, HSL,



# COLOR IN VISUALIZATION

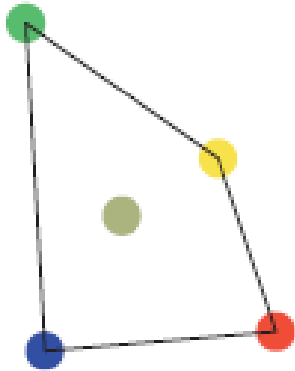
VISIBLE LIGHT = INFINITE NUMBER OF COLORS  
MONITOR = ~MILLIONS OF COLORS  
OUR PERCEPTION = ? OF COLORS



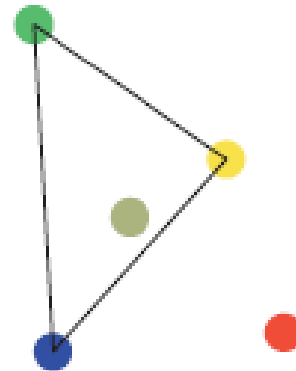
# PRE-ATTENTIVE COLOR DISTINCTION

## CONVEX HULLS IN CIE SPACE

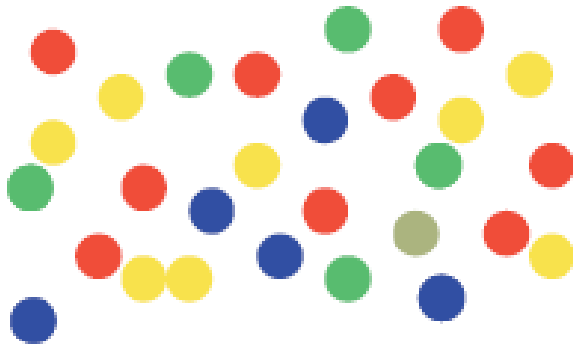
a



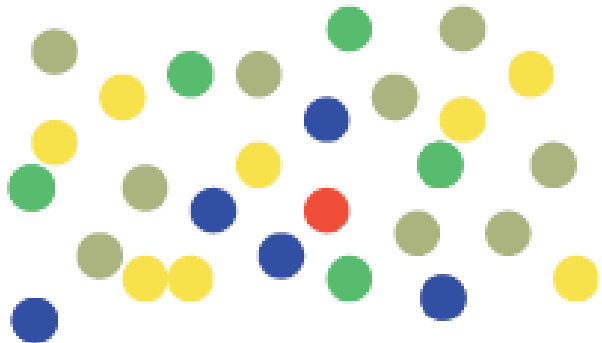
b



c



d



# USING DISTINCT COLORS

## MOST DISTINCT COLORS

Black, white

Red

Yellow, green

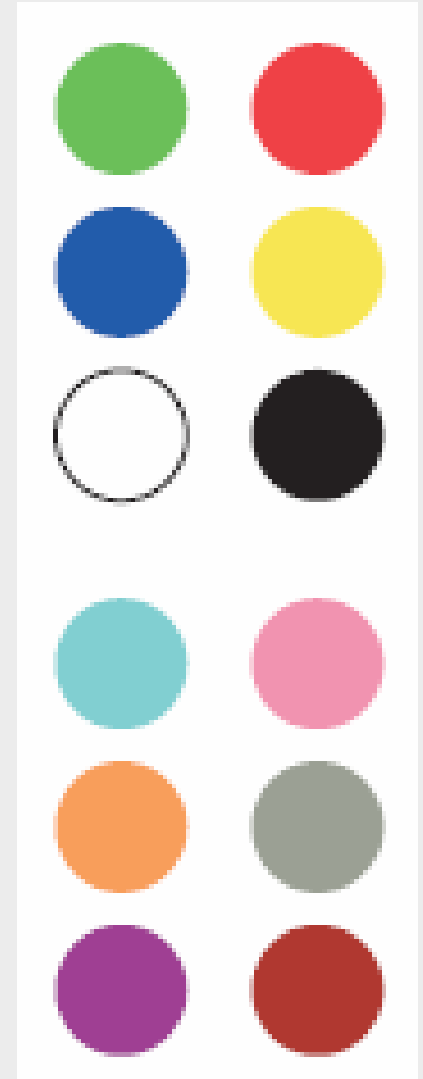
Blue

## UNIQUE HUES

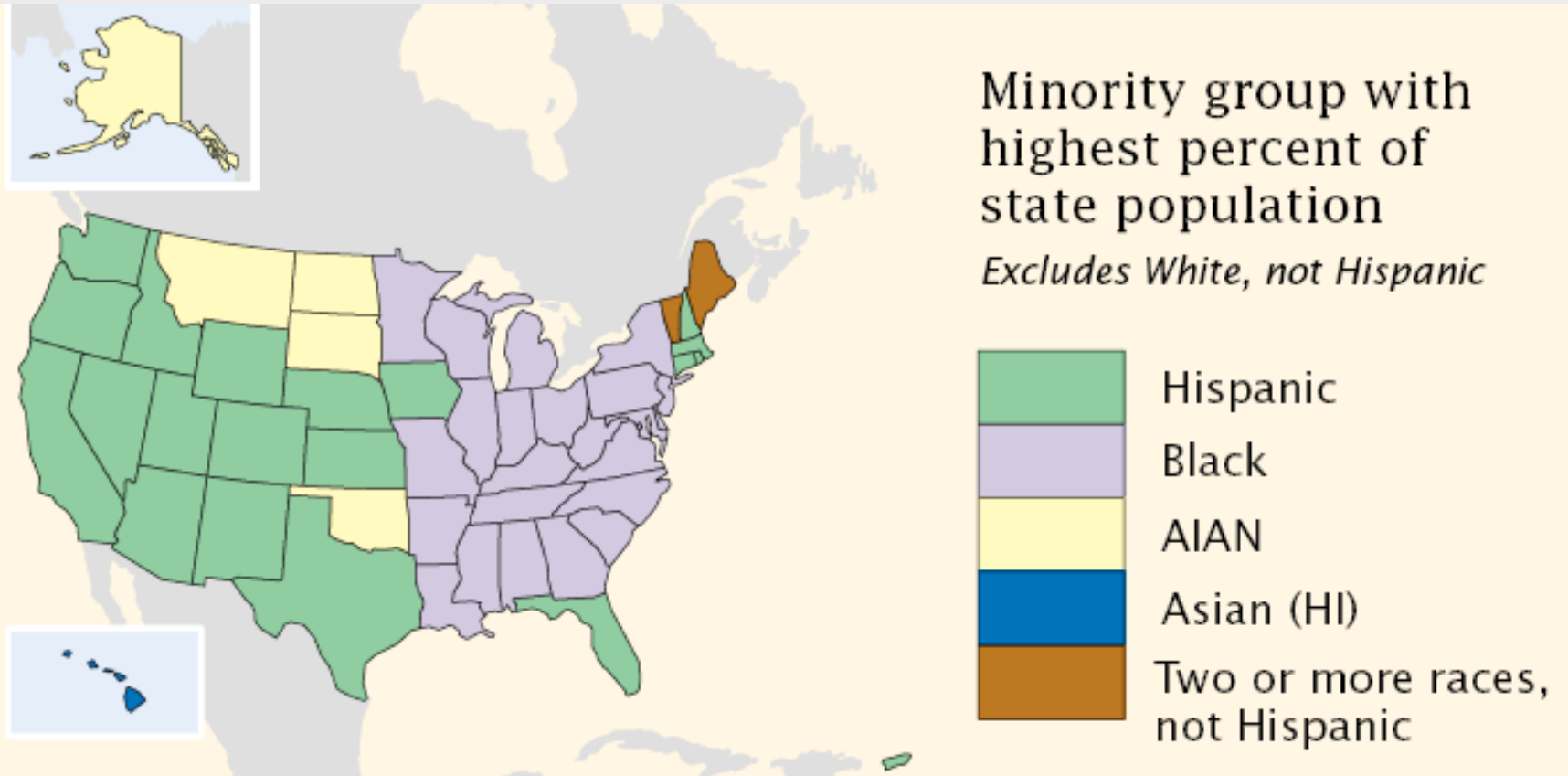
## USAGE:

Labeling, grouping

Categorical values

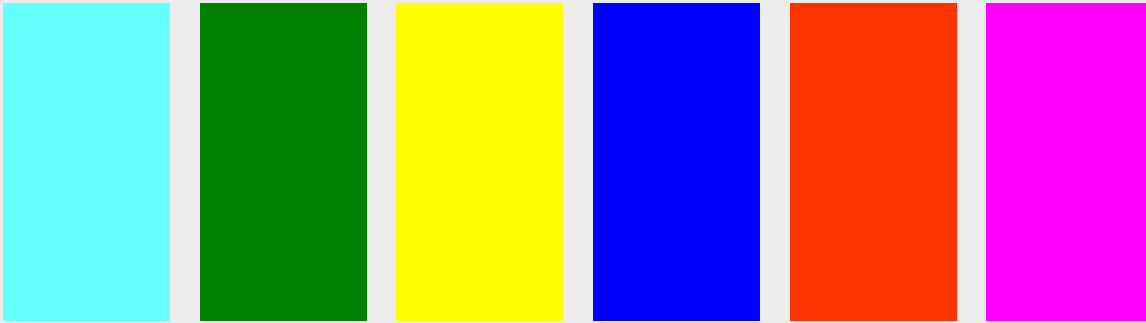


# EXAMPLE – CATEGORICAL VALUES

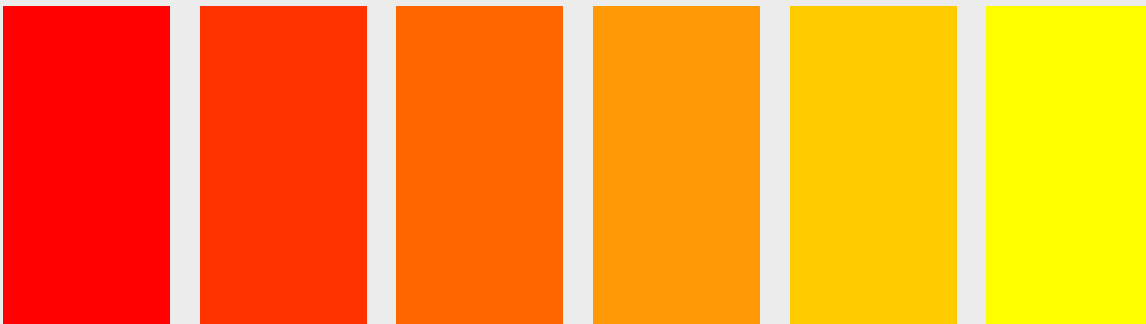


# CAN WE USE COLOR FOR ORDINAL VALUES?

ALSO FOR ORDINAL VALUES ?

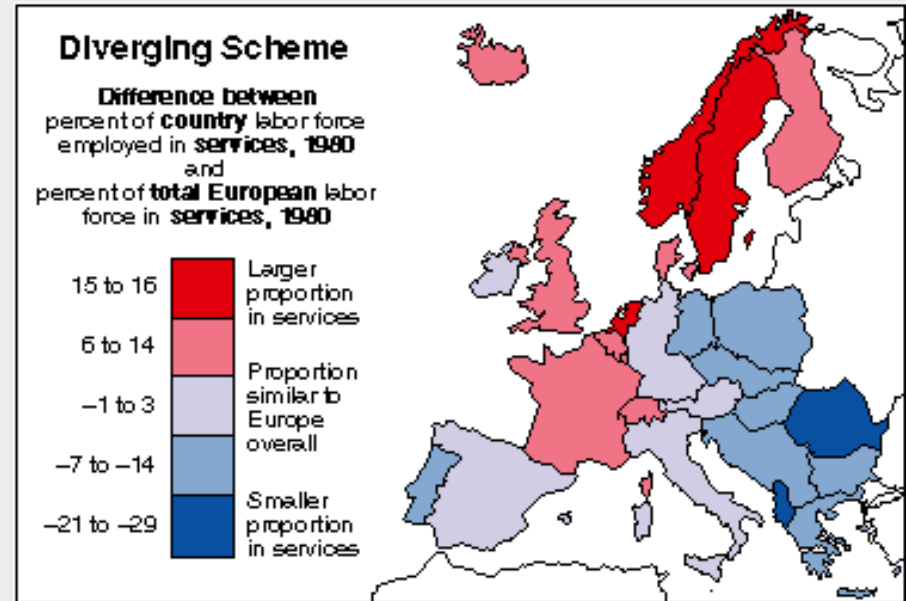
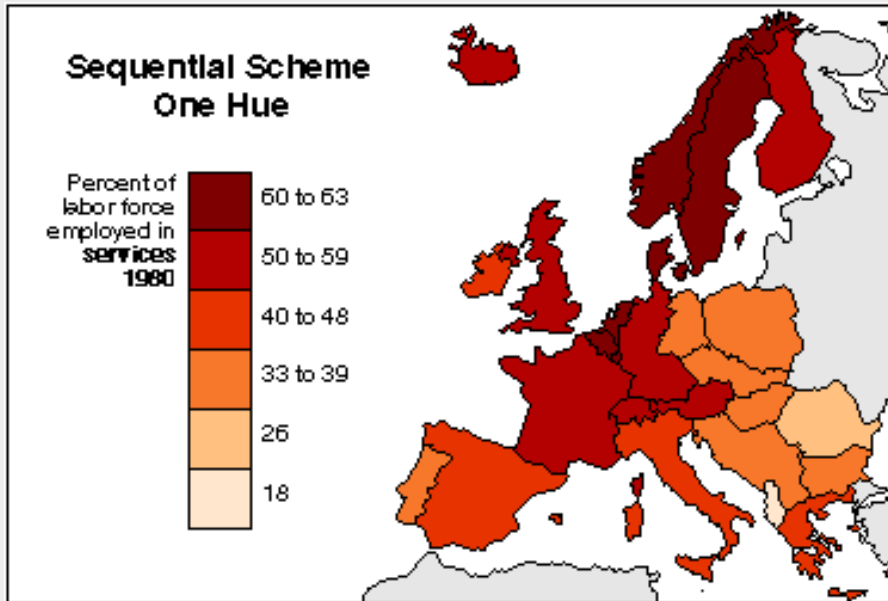


SEQUENCE OBVIOUS WHEN HUES ADJACENT



# USING SIMILAR COLORS

## ORDINAL RATHER THAN CATEGORICAL VALUES



ONLY FEW DISTINCT LEVELS.  
CONTINUOUS VALUES WILL NOT BE READ  
PROPERLY.

# CONCLUSION: USE COLORS FOR ...

## FOR CATEGORIES

Qualitative color schemes

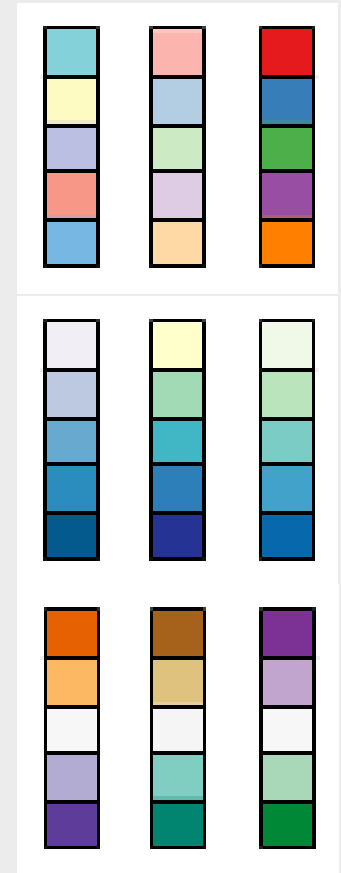
## FOR ORDINAL VALUES

Sequential color schemes

Diverging color schemes

## FOR SMALL NUMBER OF VALUES

Can be increased e.g. by using texture

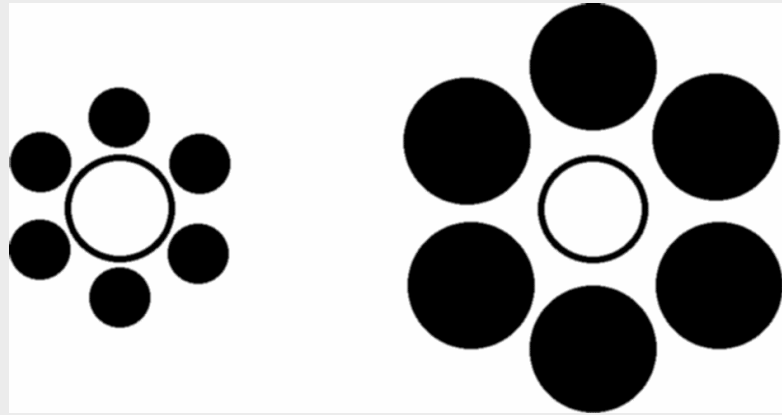
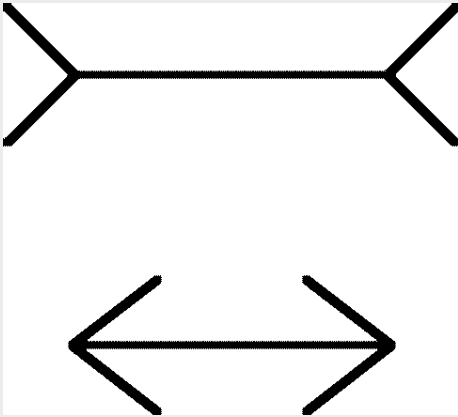




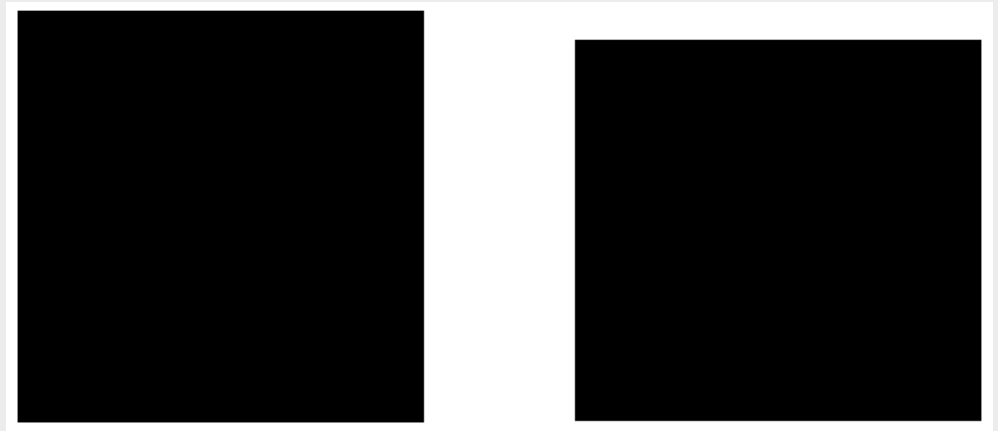
# SHAPES IN VISUALIZATION

# PERCEPTUAL ISSUES RECOLLECTION

PERSPECTIVE COMPENSATION, CONTEXT...



DIFFERENT  
PERCEPTION  
OF WIDTH AND  
HEIGHT



# DATA -> MAPPING -> VISUAL CUE

## DIFFERENT MAPPING TYPES

### LINEAR MAPPING

Length, position on a line, line thickness

Is color (1-dimensional) a linear visual cue?

### QUADRATIC - AREA

### CUBIC - VOLUME

## NO SENSE OF EXACT VALUES, ONLY RATIO

We read values by comparing them

---

# EX.: LIMITATIONS OF LINEAR MAPPING

HIGH VALUES  
EXHAUST THE  
AVAILABLE  
RANGE  
OF VISUAL  
VALUES

**45.000 marriages**

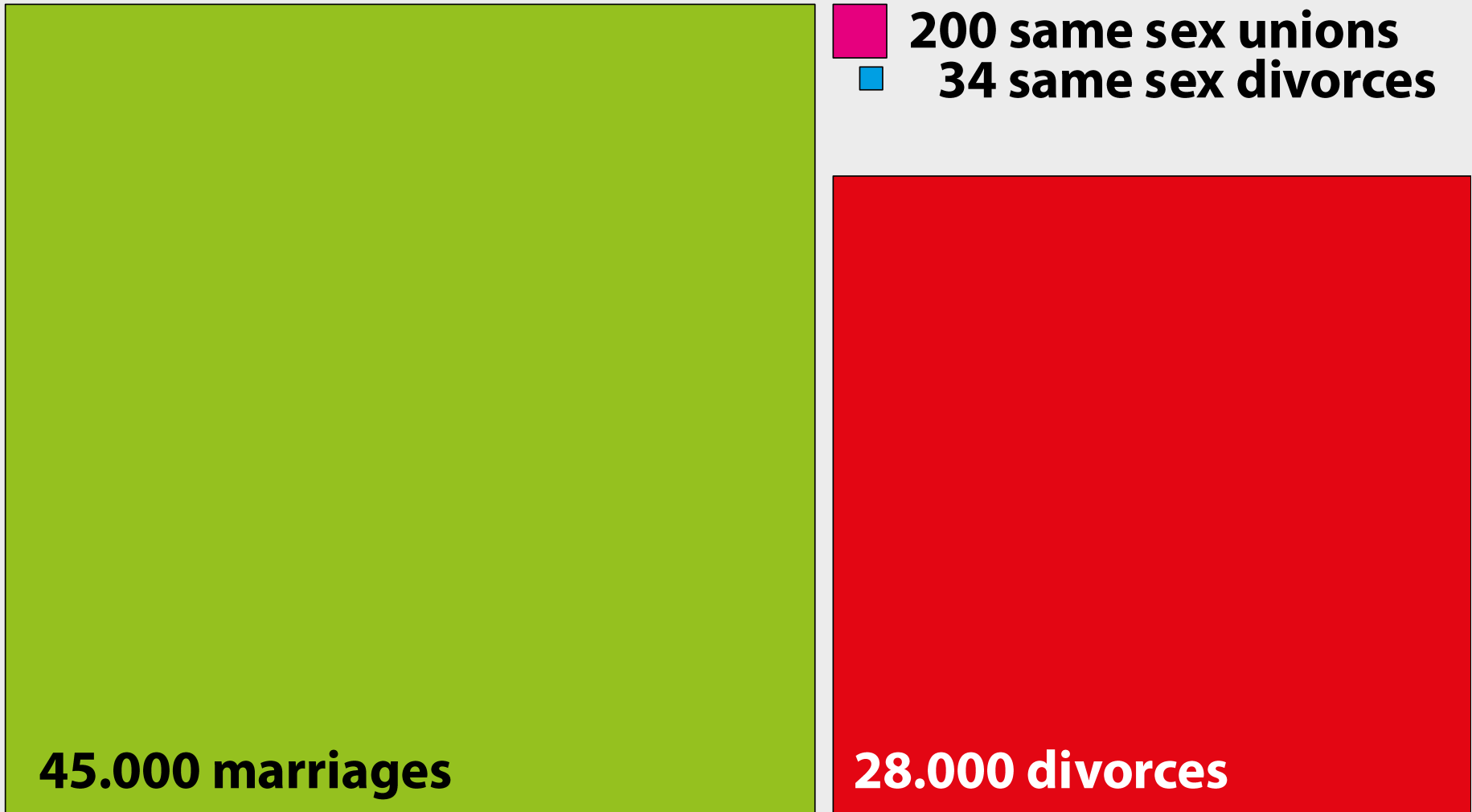
**28.000 divorces**

**200 same sex unions**

**34 same sex divorces**

Annual partnership statistics in Czech republic

# EX.: USING QUADRATIC MAPPING



Annual partnership statistics in Czech republic

# COMMON ELEMENTS OF FORM

LENGTH, POSITION ON A LINE

SHAPE TYPE

Lines, rectangle, polygons, stars, spikes  
Symbols

ANGLE, ORIENTATION

Remember the bad perception of angles, orientation!

GROUPS

Concentration is pre-attentive

---

# MAPPING VALUES TO VISUAL CUES

DATA → GEOMETRY

MAPPING FUNCTION, TRANSFER FUNCTION

Linear, non-linear (logarithmic, quadratic, hyperbolic)

ZERO-LEVEL

LIE FACTOR

FEW GENERAL RULES

Most applications require ad-hoc evaluation

---

# EXAMPLES OF SHAPES IN VISUALIZATION



# ACCEPTABLE USE OF ANGLES

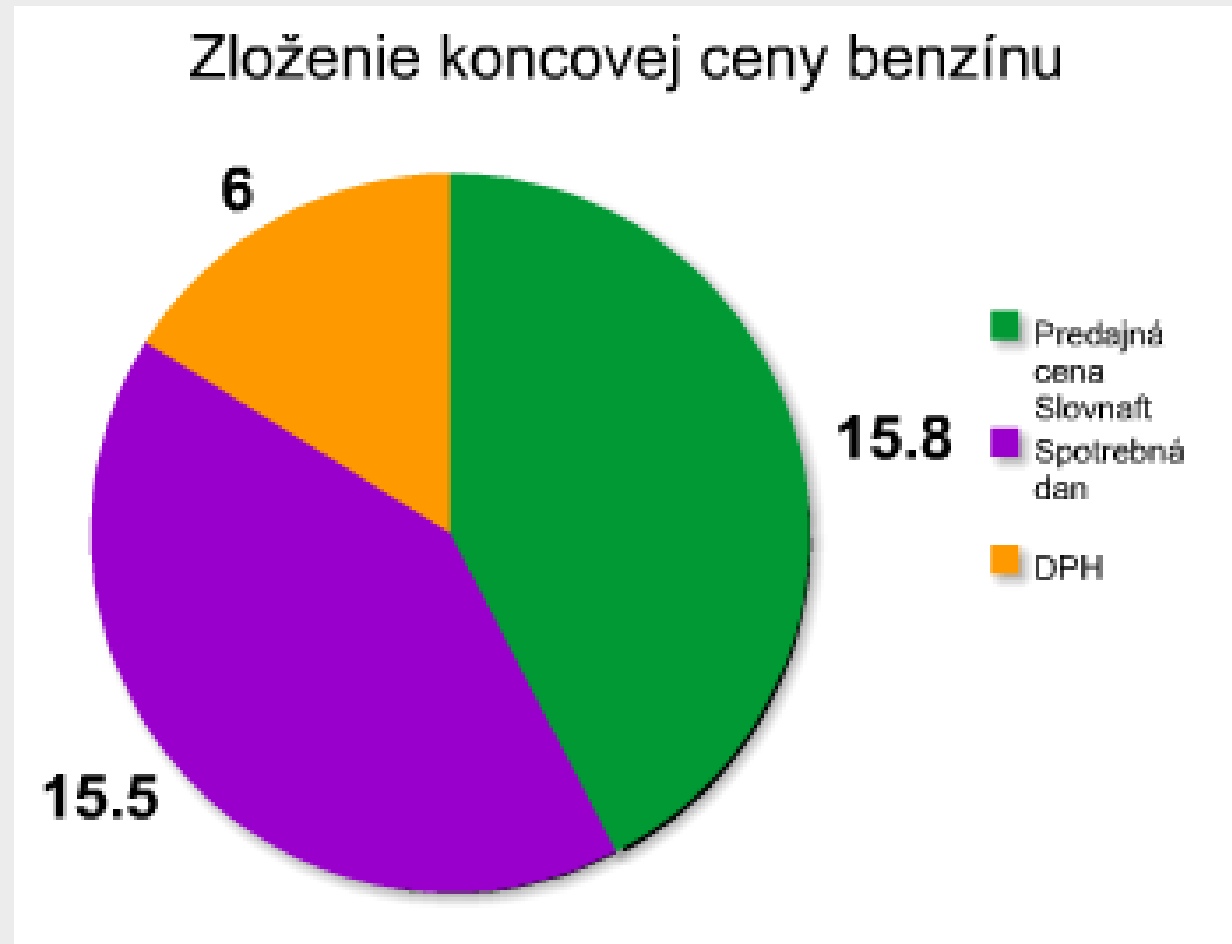
SMALL  
NUMBER OF  
CATEGORIES

GOAL:

- Compare areas

RESULT:

- Mission accomplished



# ACCEPTABLE USE OF ANGLES

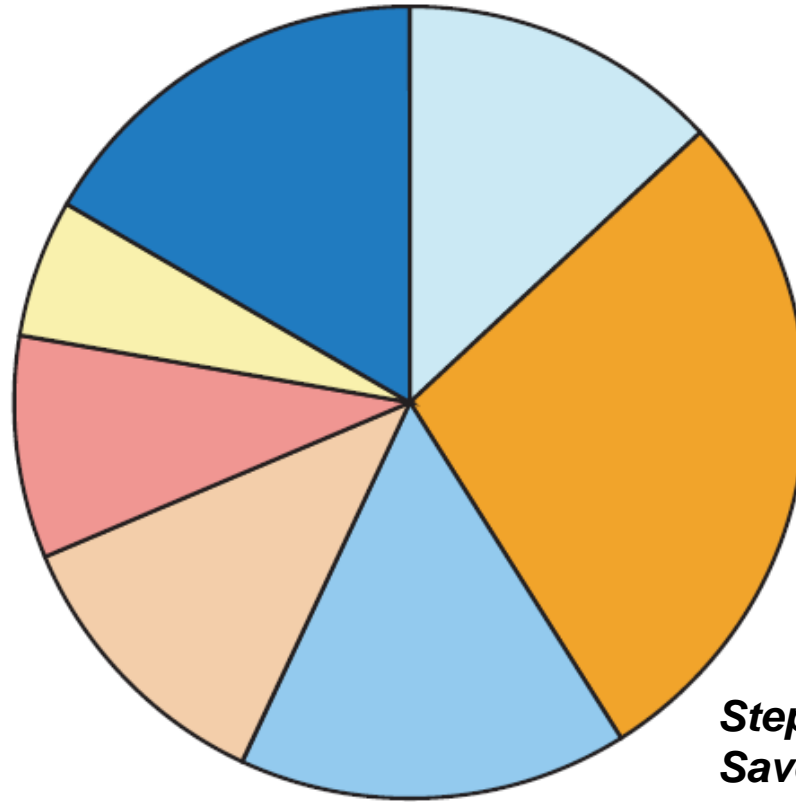
LARGE (?)  
NUMBER OF  
CATEGORIES

GOALS:

- Order by size
- Sum the blue

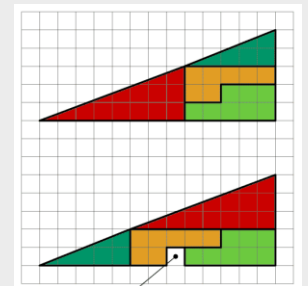
RESULT:

- Mission failed



- Cabernet Sauvignon
- Sangiovese
- Prosecco
- Chardonnay
- Syrah
- Tempranillo
- Pinot Grigio

**Stephen Few**  
***Save the Pies for Dessert***

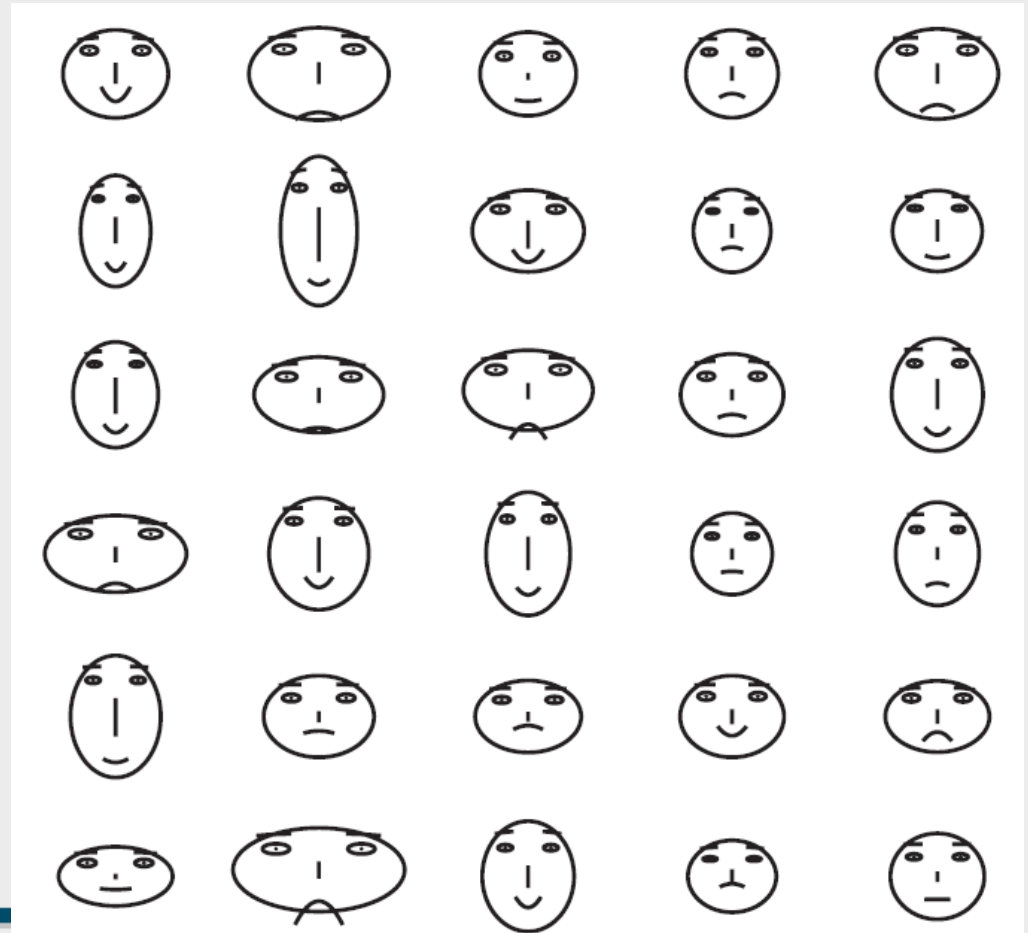


# CHERNOFF FACES

FACE RECOGNITION ABILITIES USED FOR  
PERCEIVING SEVERAL ATTRIBUTES AT THE  
SAME TIME

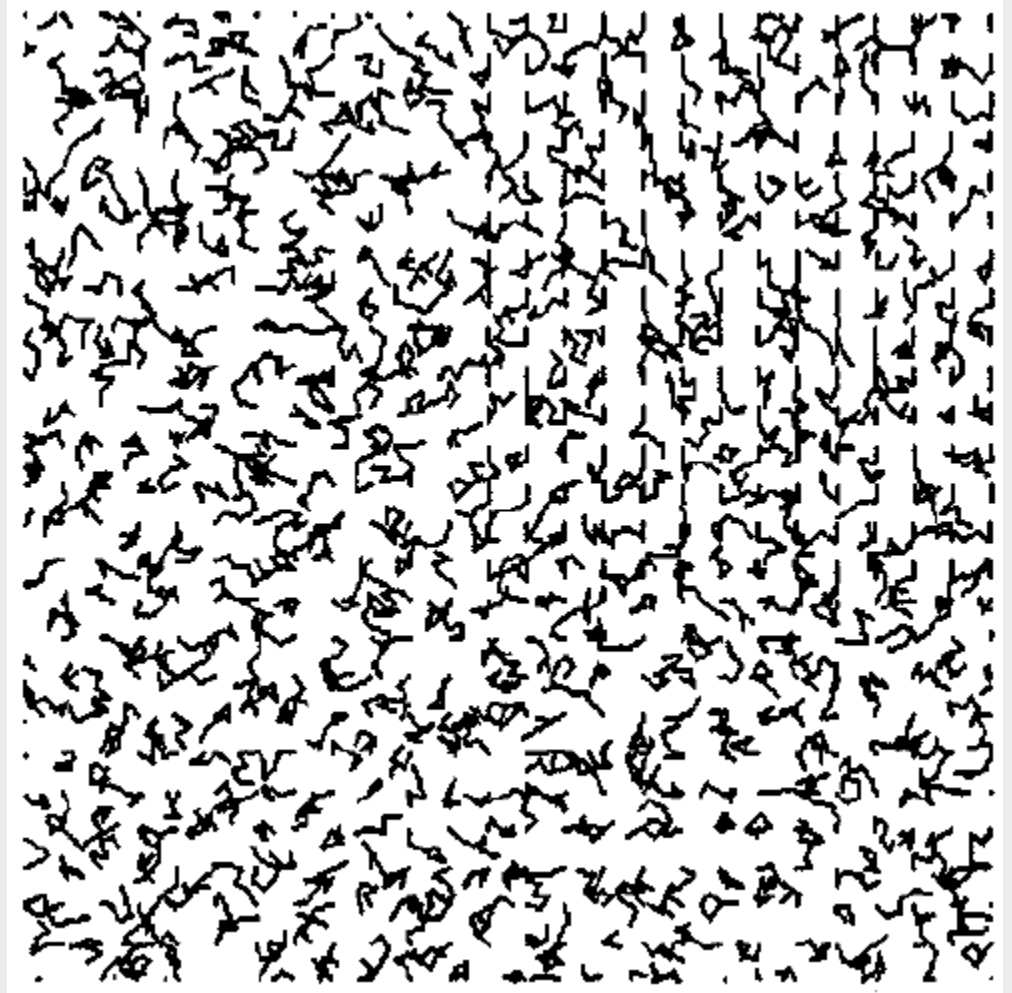
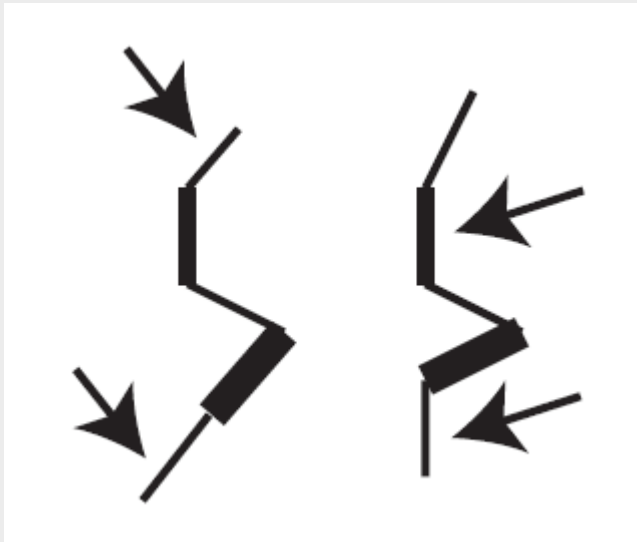
EACH TEXTBOOK  
HAS THEM

NOBODY USES  
THEM



# EXVIS GLYPH – PATTERN PERCEPTION

MULTIPLE DATA  
ATTRIBUTES  
MAPPED TO  
GLYPH SHAPE  
+ X,Y POSITION



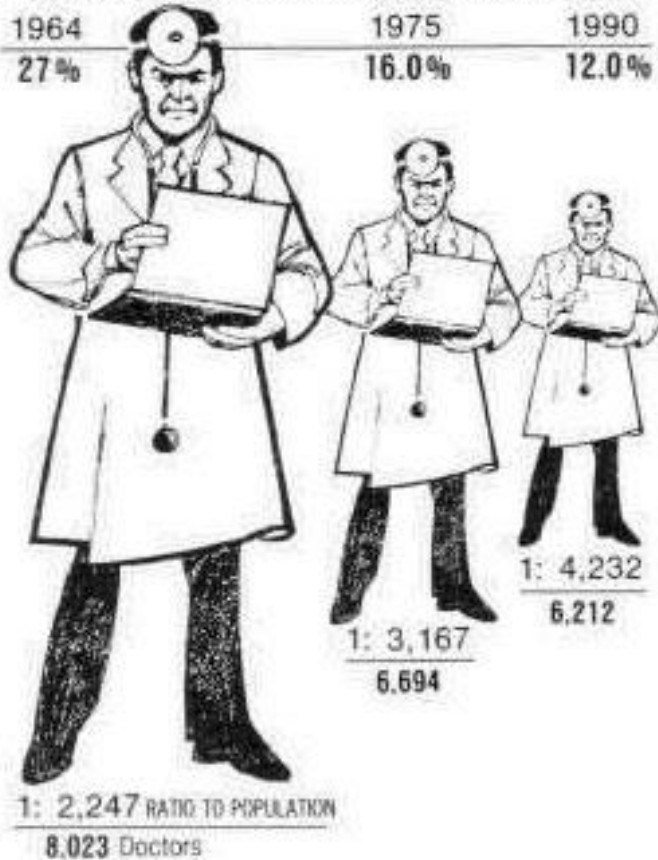
# ERRORS IN VISUALIZATION

# IGNORING THE QUADRATIC NATURE

## THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%



1980 = \$1.00



1985 = \$0.70



1990 = \$0.56



1995 = \$0.50

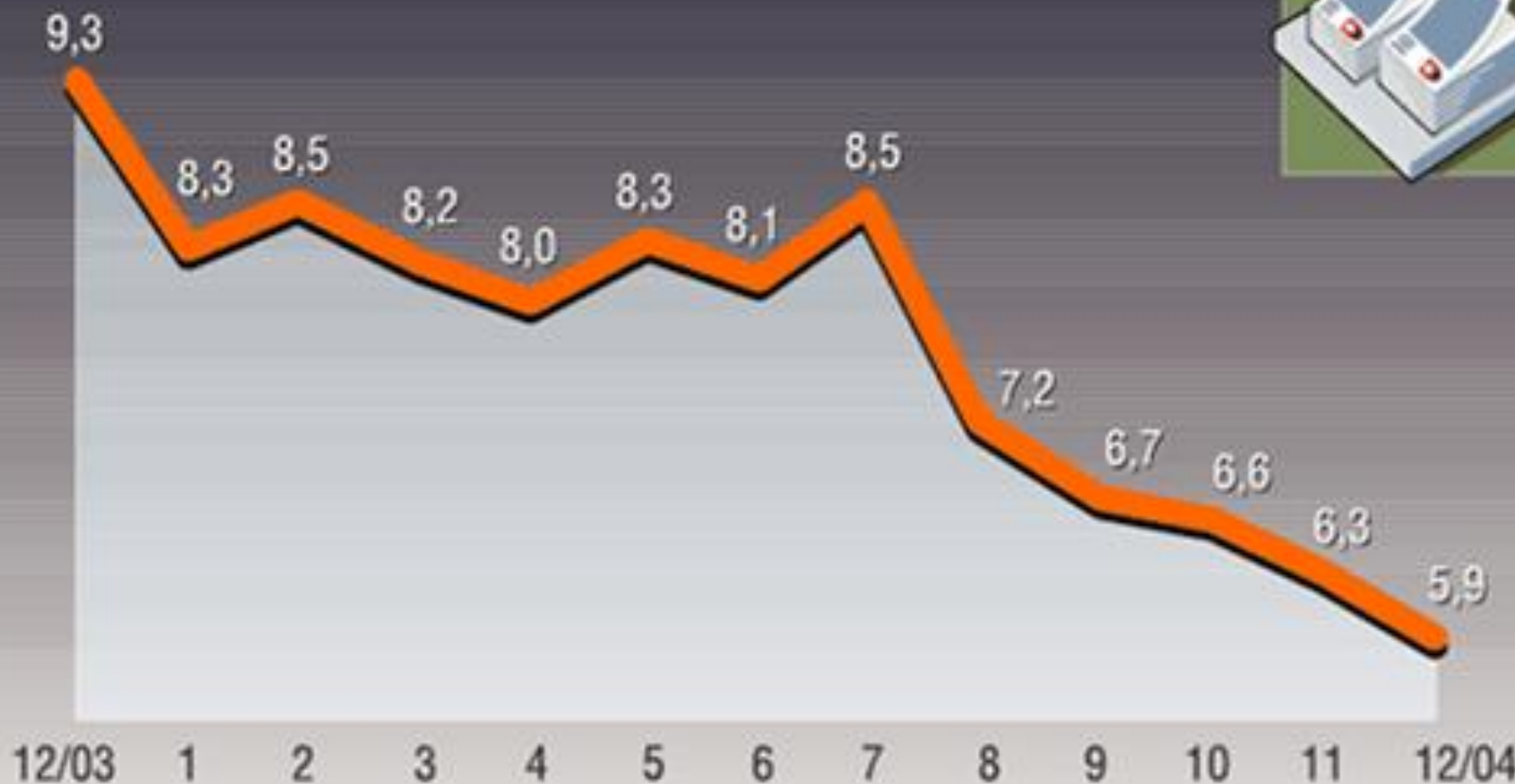


2000 = \$0.46

# ZERO LEVEL

## INFLÁCIA NA SLOVENSKU

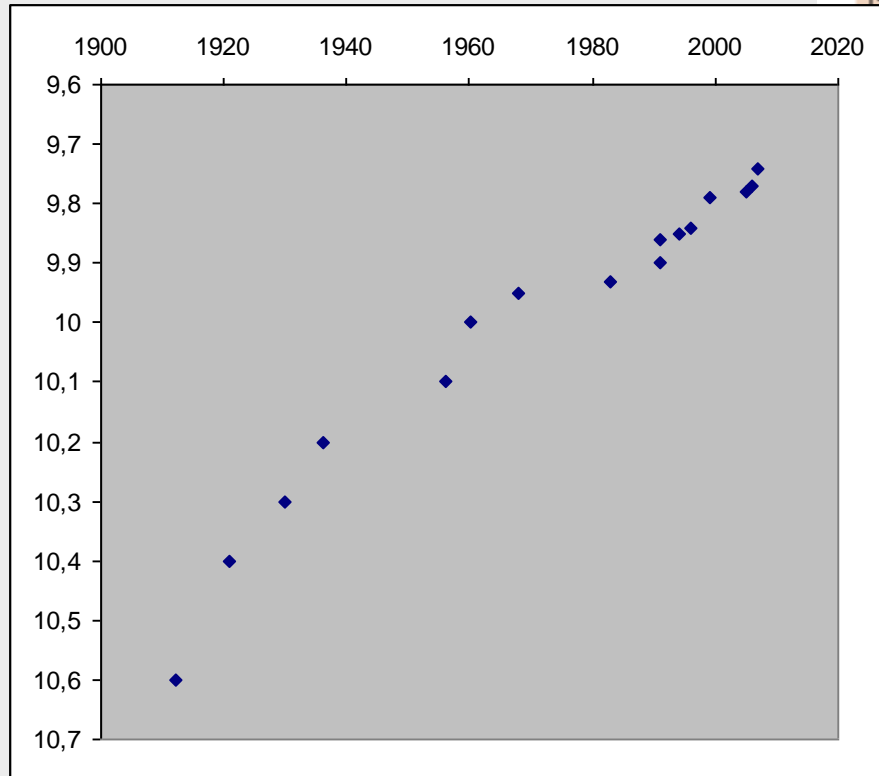
medziročný rast spotrebných cien (v percentách)



zdroj: Štatistický úrad SR, graf: ČTK

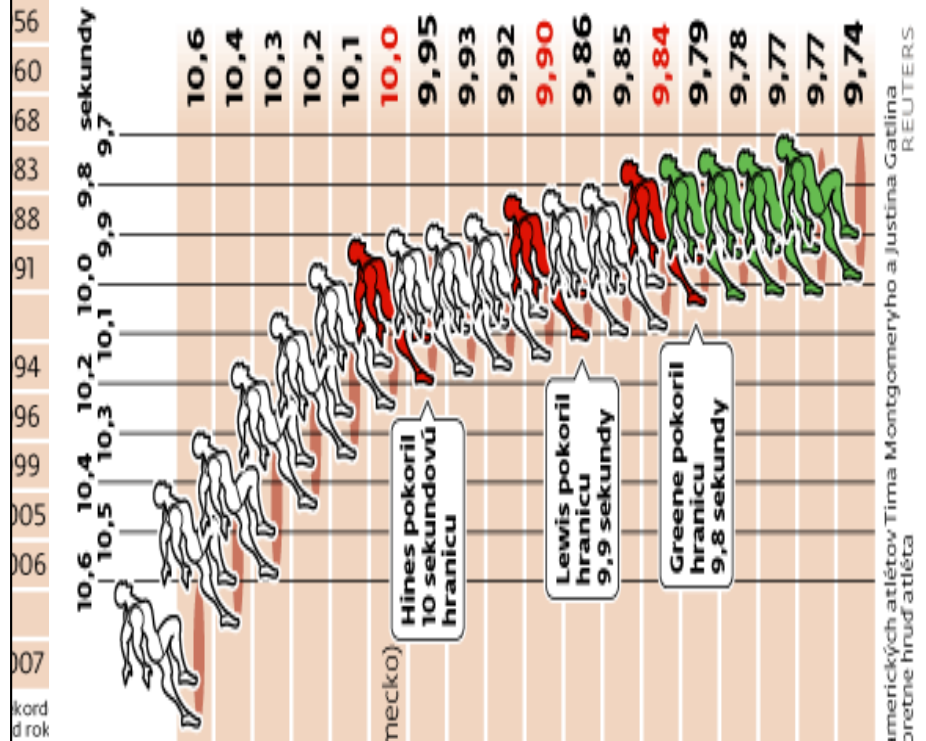
# (THIS ERROR HAS NO ONE-LINE NAME)

## 100M RECORD EVOLUTION



Asafa Powell překonal svoj světový rekord na 100 metrov v nedelju časom 9,74 sekundy.

1912	6. júl	Donald Lippincott (U S A)	10,6
1921	23. apr.	Charles Paddock (U S A)	10,4
1930	9. aug.	Percy Williams (Kanada)	10,3
1936	20. jún	Jesse Owens (U S A)	10,2

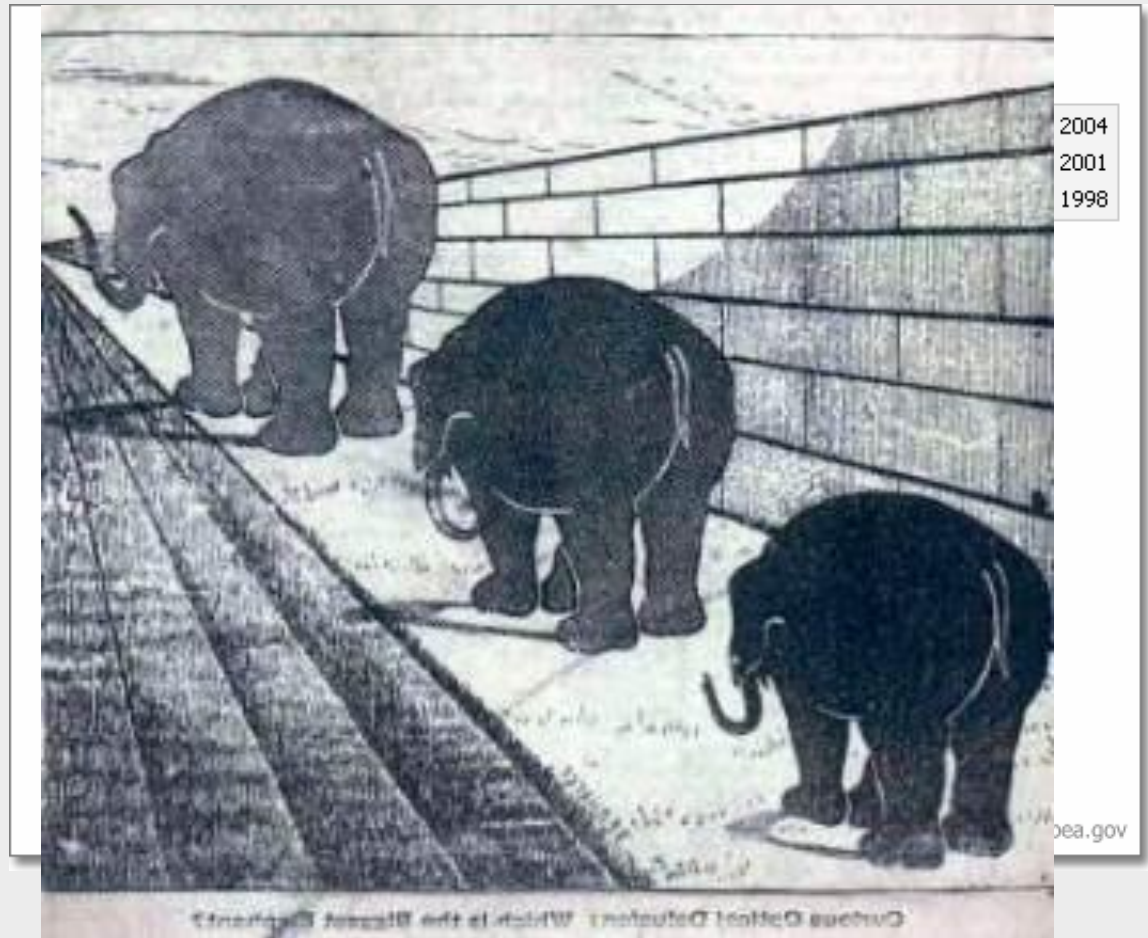
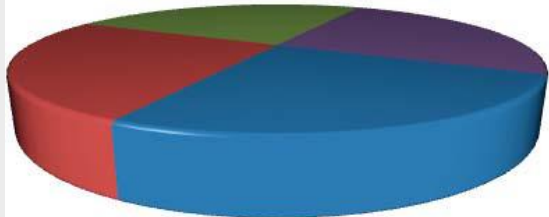
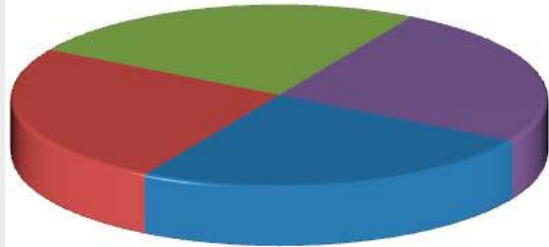
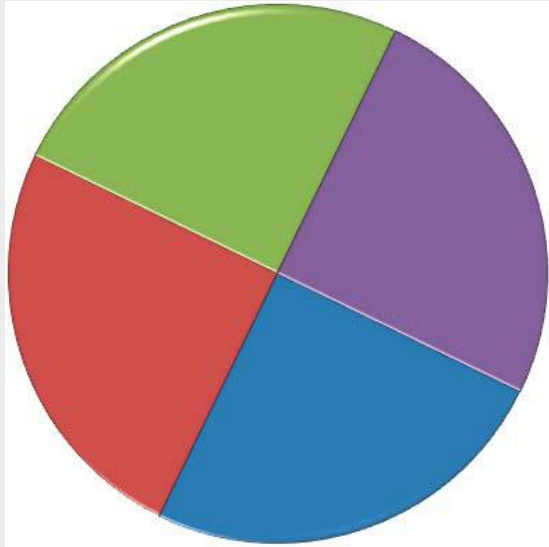




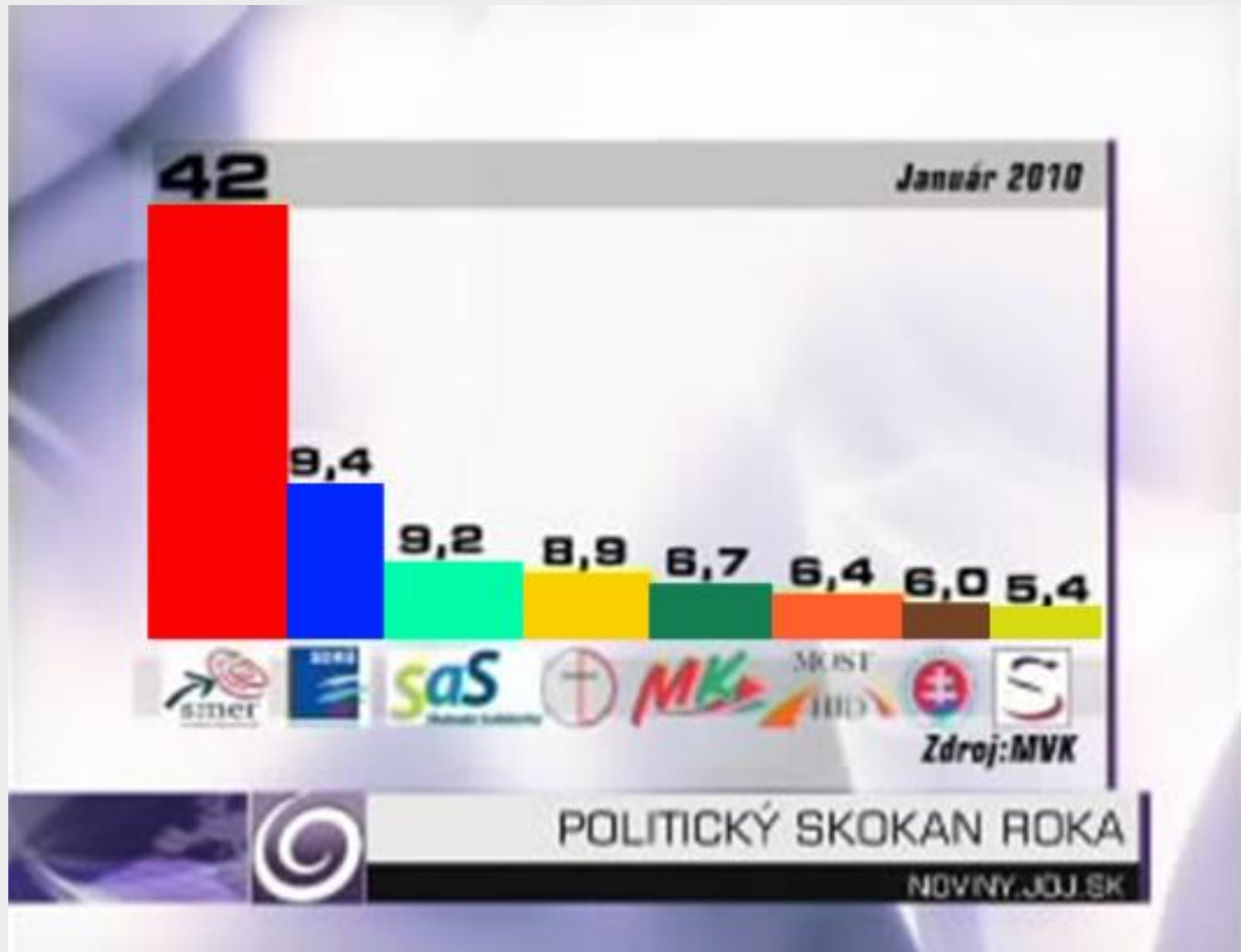
# HOW TO LIE WITH VISUALIZATION



# PERSPECTIVE / 3D IS RISKY



# BAR CHART GONE WILD



# CONCLUSION

SHAPES AND SIMPLE GRAPHICAL CUES CAN BE USED FOR SIMPLE DATA TYPES (~SCALAR VALUES)

ADVANCED DATA REQUIRE PATTERNS AND COMBINATIONS OF VISUAL CUES TO BE VISUALIZED

VICE VERSA:

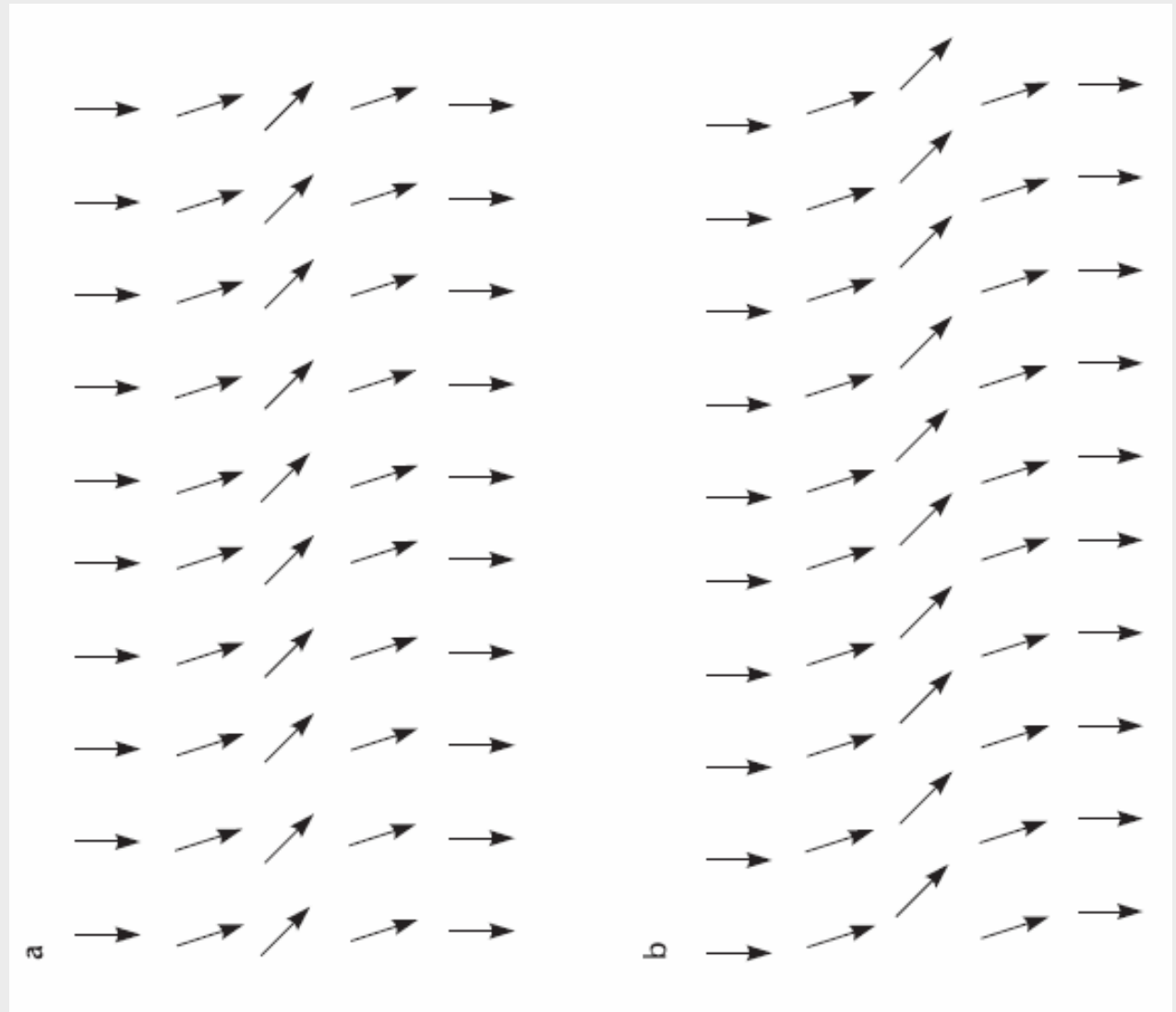
Through pattern perception -> relations and advanced information can be discovered

---

I HAVE A FEELING  
WE'RE NOT IN KANSAS  
ANYMORE

ADVANCED  
DATA

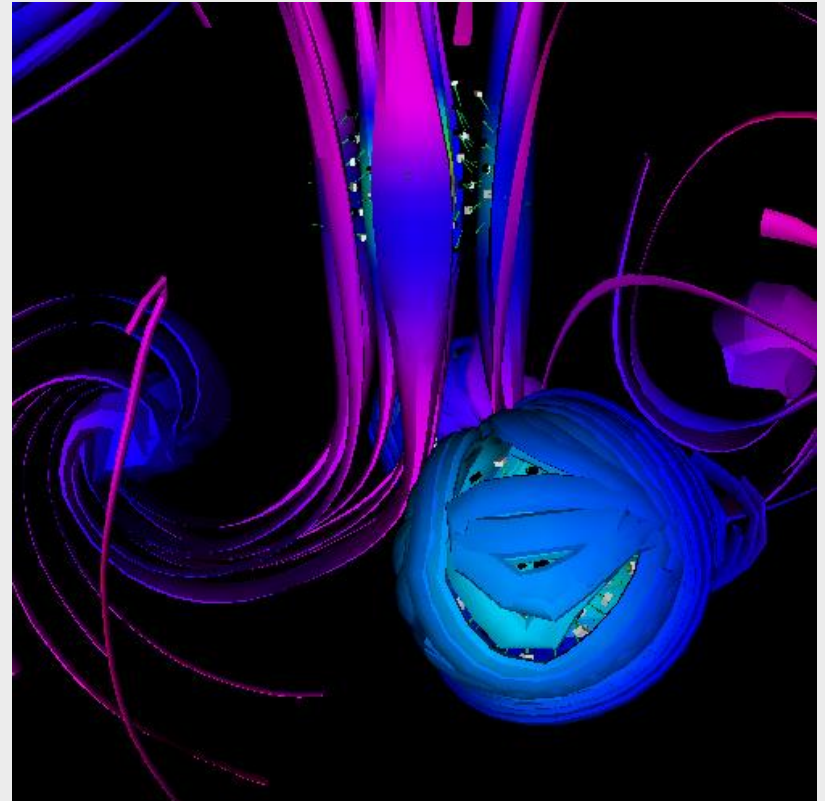
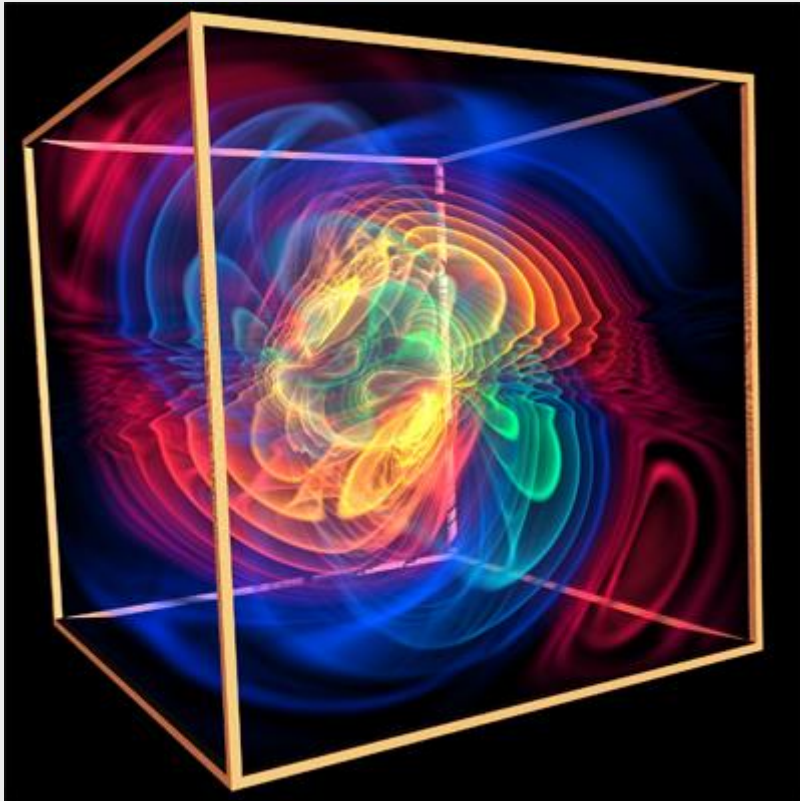
# VECTORS: COMBINE SIZE & ORIENTATION





# TENSORS

- ~ MULTIDIMENSIONAL MATRICES
- ~ VECTOR OPERATIONS



# COMPLEX REAL WORLD DATA

MULTIDIMENSIONAL

MISSING VALUES OR ERRORS IN DATA

ATTRIBUTES OF DIFFERENT TYPES

Numerical, nominal, vectors

...

DESIGNING VISUALIZATION - FIND THE RIGHT  
MAPPING BETWEEN VALUES AND VISUAL CUES

---