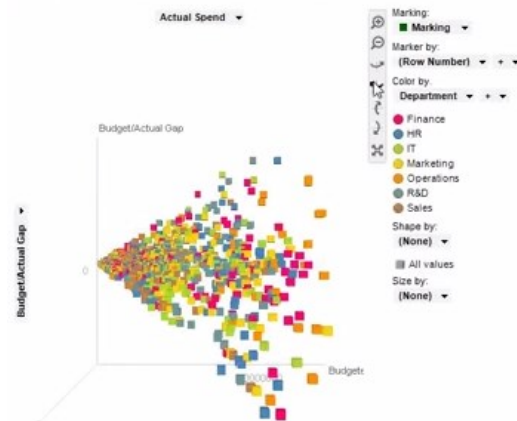


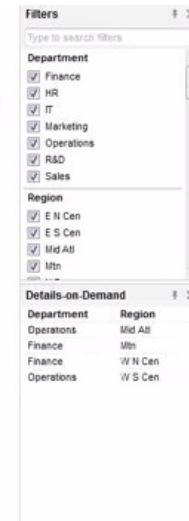
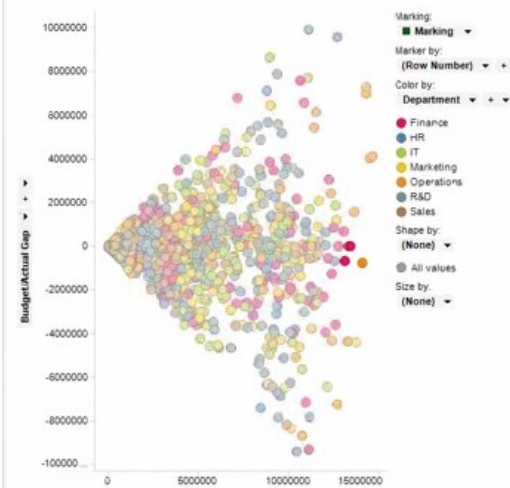


Data

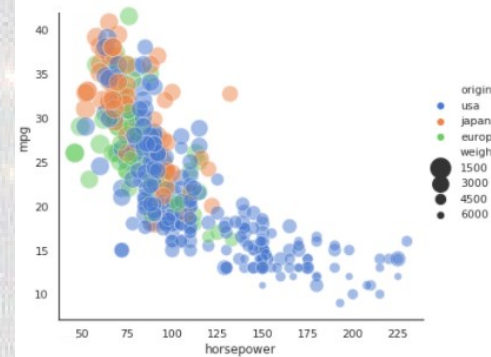
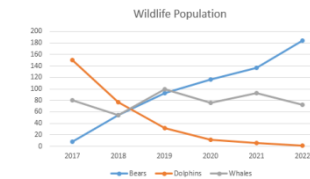
Actual Spend vs. Budgeted Spend and Budget/Actual Gap



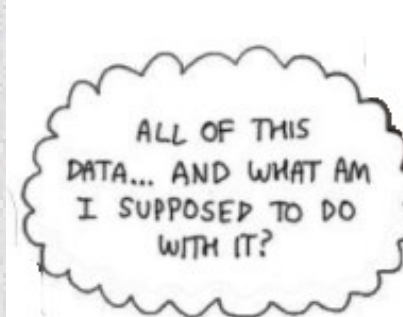
Budget/Actual Gap vs. Budgeted Spend



40	102	0.50	0	0.0	0	7.4199	0.0009	0.009	0	1.200	0.015	1.169	2.07
67	201	0.77	0	0.0	0	8.7669	0.0018	0.022	0	1.166	0.018	1.130	1.78
88	201	3.12	0	0.0	0	11.8279	0.0062	0.072	0	0.421	0.015	0.490	0.85
52	183	-0.15	0	0.0	0	8.2485	0.0011	0.015	0	0.425	0.007	0.490	2.60
02	115	0.28	3	0.0	0	11.0625	0.0035	0.035	0	1.567	0.197	1.490	0.56
25	151	1.82	0	0.0	0	6.6890	0.0006	0.007	0	0.955	0.005	0.940	4.38
48	100	-0.51	0	0.0	0	8.6208	0.0015	0.011	0	0.516	0.015	0.590	1.65
50	105	2.69	0	0.0	0	7.6923	0.0012	0.010	0	1.438	0.017	1.430	1.65
81	203	-0.58	0	0.0	0	8.8563	0.0017	0.024	0	0.962	0.014	0.950	1.68
38	106	1.01	0	0.0	0	7.6626	0.0015	0.011	0	0.456	0.015	0.530	1.92
48	144	0.70	0	0.0	0	9.1657	0.0015	0.017	0	0.520	0.015	0.600	1.30
54	198	7.07	0	0.0	0	6.4211	0.0006	0.008	0	0.763	0.003	0.800	2.87
63	135	-0.37	0	0.0	0	9.2438	0.0018	0.020	0	0.535	0.020	0.610	1.46
86	188	-1.27	0	0.0	0	9.4598	0.0026	0.026	0	0.671	0.017	0.730	1.35
65	184	-0.61	0	0.0	0	8.9385	0.0013	0.016	0	0.468	0.013	0.540	1.66
98	180	-0.35	0	0.0	0	9.3189	0.0019	0.024	0	1.041	0.020	1.010	1.44
40	157	-1.40	0	0.0	0	8.3069	0.0011	0.012	0	0.142	0.010	0.160	2.43
60	87	2.24	0	0.0	0	7.7648	0.0020	0.014	0	1.453	0.010	1.460	1.66
72	165	-0.74	0	0.0	0	9.1842	0.0022	0.024	0	0.840	0.017	0.860	1.33
32	158	0.32	0	0.0	0	8.2076	0.0011	0.010	0	0.466	0.015	0.540	2.00
30	111	6.90	0	0.0	0	6.5466	0.0009	0.007	0	0.514	0.005	0.590	3.55
56	138	2.04	0	0.0	0	9.2353	0.0016	0.020	0	0.916	0.004	0.960	1.20
38	104	-0.25	0	0.0	0	7.8503	0.0015	0.009	0	1.150	0.014	1.110	1.90
93	137	4.22	0	0.0	0	10.5345	0.0037	0.035	0	0.391	0.045	0.460	0.67
55	129	-0.53	0	0.0	0	8.7854	0.0019	0.020	0	0.778	0.015	0.810	1.81
38	114	0.97	0	0.0	0	7.5669	0.0010	0.010	0	0.475	0.009	0.550	2.10
09	134	5.96	1	0.0	0	10.3360	0.0055	0.082	2	0.570	0.042	0.640	0.58
64	177	-0.96	0	0.0	0	8.7201	0.0018	0.021	0	0.026	0.013	0.040	1.75
43	106	0.07	0	0.0	0	8.2910	0.0011	0.010	0	0.387	0.012	0.450	1.68
28	157	1.36	0	0.0	0	6.3456	0.0006	0.007	0	1.032	0.005	1.000	4.16
41	114	1.71	0	0.0	0	8.0825	0.0012	0.009	0	1.104	0.018	1.070	1.57
87	136	-1.30	1	0.0	0	9.7391	0.0025	0.028	0	0.800	0.000	0.800	1.07
50	170	1.94	0	0.0	0	8.7350	0.0013	0.017	0	1.035	0.016	1.010	1.39
69	164	1.13	0	0.0	0	10.8416	0.0044	0.044	0	1.150	0.020	1.250	0.80
44	170	1.65	0	0.0	0	7.4752	0.0008	0.009	0	1.164	0.008	1.130	2.70
68	129	0.76	0	0.0	0	9.6389	0.0020	0.018	0	0.450	0.030	0.520	1.22
34	257	2.90	0	0.0	0	6.6231	0.0006	0.010	0	0.616	0.005	0.680	4.70
65	89	1.42	0	0.0	0	9.1049	0.0022	0.014	0	0.906	0.024	0.910	1.11



Data
visualization?

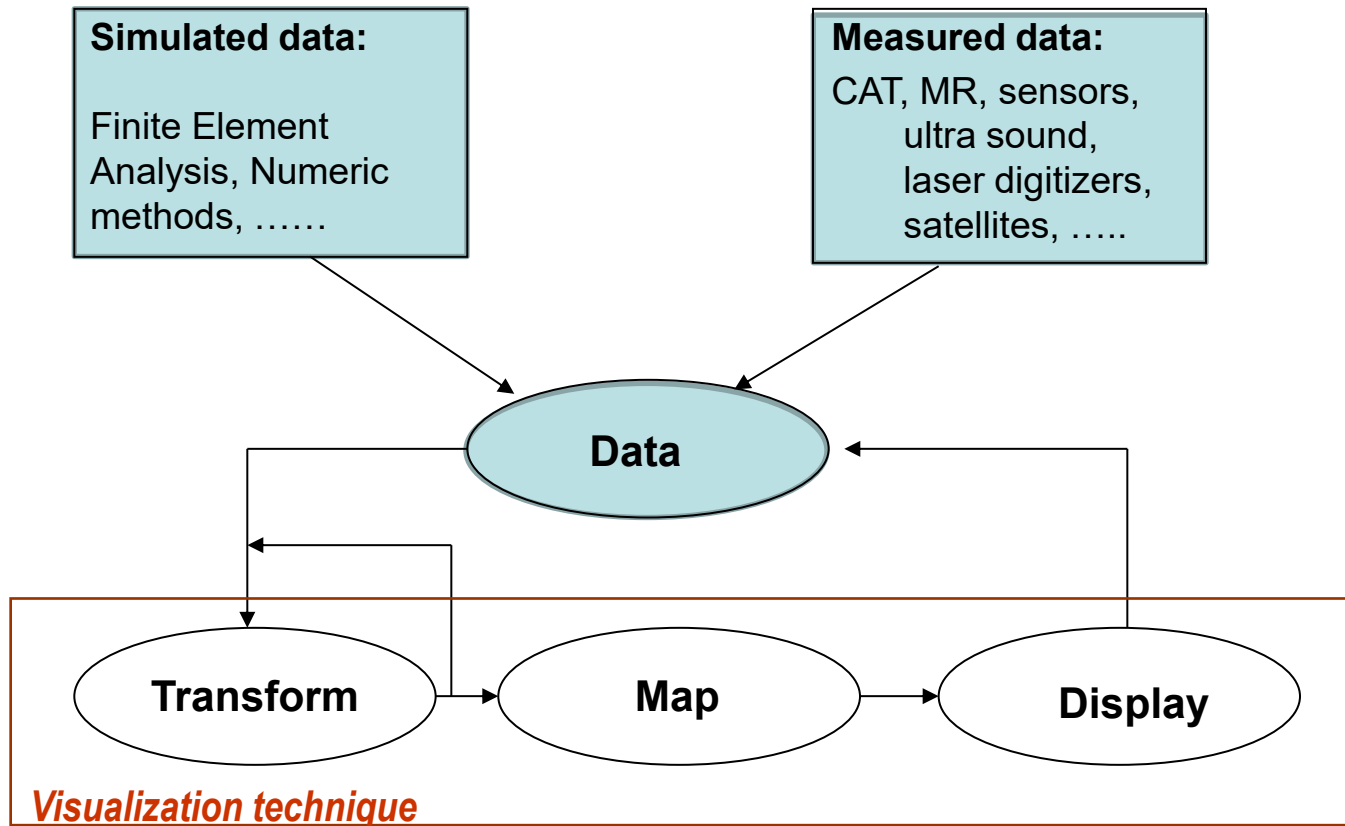


Machine learning?

Statistics?



Scientific Visualization reference model



(adapted from Schroeder et al., 2006)

Adequate data pre-processing is vital!



Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Data Characteristics

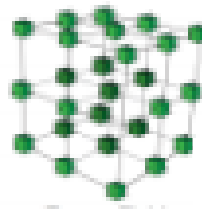
- Data may have a lot of different forms and there are many techniques and systems to visualize them
- A data classification is important to:
 - predict what visualization techniques are adequate
 - make easier the communication about the data
 - allow a more systematic approach to Visualization
 -

Data Abstraction

name	rank	gender	year
Jacob	1	boy	2010
Isabella	1	girl	2010
Ethan	2	boy	2010
Sophia	2	girl	2010
Michael	3	boy	2010

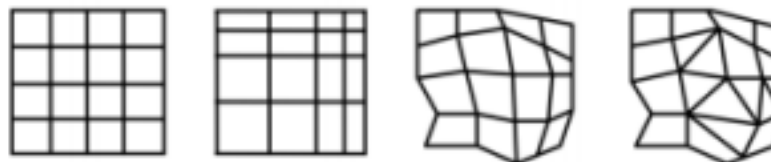
- Four basic dataset types:

- Tables
- Networks
- Fields
- Geometry



- Five basic datatypes

- Items
- Attributes
 - Categorical
 - Ordered
 - Ordinal
 - Quantitative
- Links
- Positions
- Grids



- **Data representation level:**
 - Qualitative (or categorical)
 - Quantitative (or numeric)
- **Data nature:**
 - Continuous
 - Discrete
- **Measuring scale:**
 - Nominal
 - Ordinal
 - Interval
 - Ratio

4.1 27 102 3.14
-0.1 16

Numerical data



Categorical data

Monday Wednesday
Tuesday Thursday

Ordinal data

(Spence, 2007)

- Examples of measuring scales and types of data:
 - **nominal** --> car brands, gender, animal species...
 - **ordinal** --> week days, preferences, levels measured in a Likert-type scale
 - **Interval** --> date, IQ, temperatures in °C
 - **Ratio** --> temperatures in °K, weight, height



- The ratio scale represents the **highest level of representation**, has a non-arbitrary zero (unlike the interval scale)
- This is a general classification and might be used to select the statistical methods to use with the data

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

<i>latitude</i>	<i>longitude</i>	<i>d</i>
-----------------	------------------	----------

- Which is the most adequate way to visualize these data?
- If *d* is depth or altitude?

the selected visualization
technique may involve
interpolation

(e.g. isocontours,
isosurfaces, 3D surface)

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

latitude

longitude

d

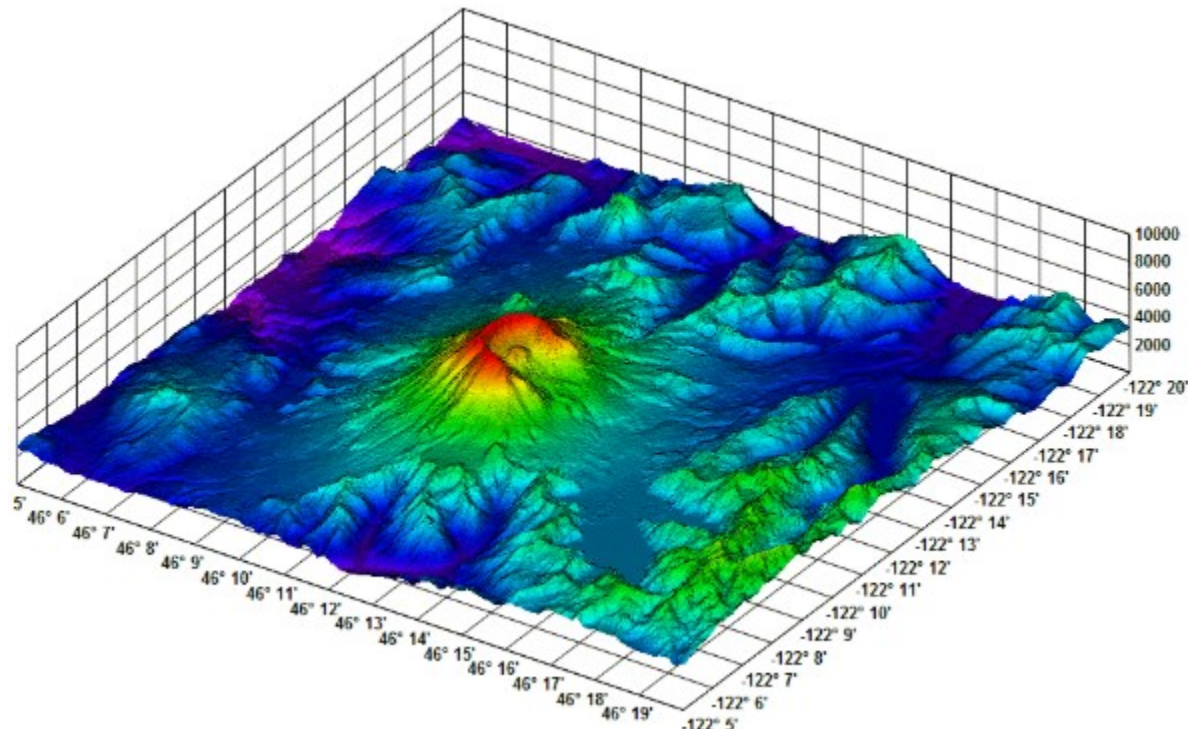


- Which is the most adequate way to visualize these data?

- If *d* is depth or altitude?

the selected visualization technique may involve interpolation

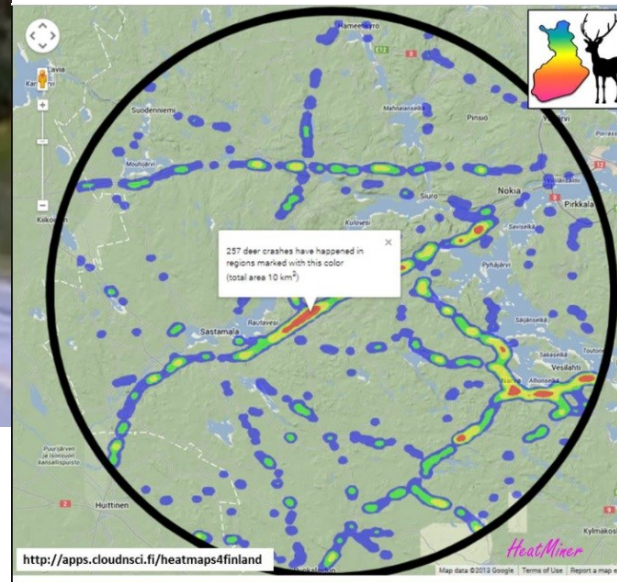
(e.g. isocontours, isosurfaces, 3D surface)



- What if the data represent location and the number of “deer crash” accidents?

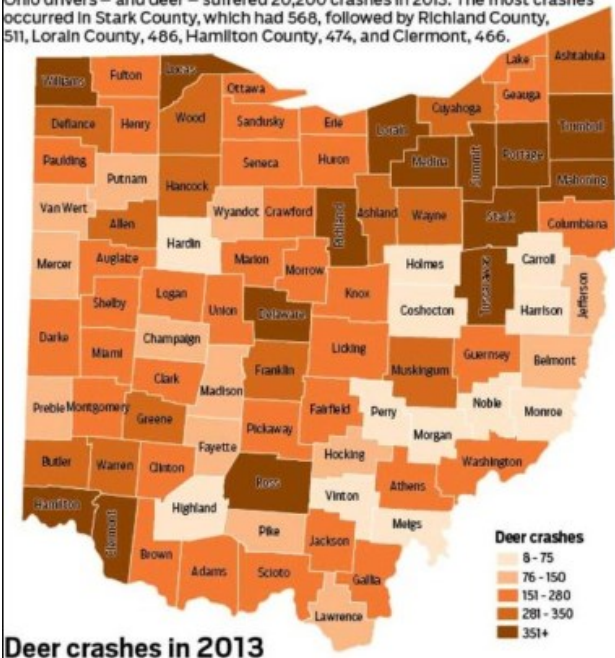


<http://cloudnsi.fi/wiki/index.php?n=Applications.Heatmaps4Finland>



Ohio deer crashes by county in 2013

Ohio drivers – and deer – suffered 20,200 crashes in 2013. The most crashes occurred in Stark County, which had 568, followed by Richland County, 511, Lorain County, 486, Hamilton County, 474, and Clermont, 466.



- Interpolation and contours don't make sense!
Know the data structure is not enough

It is necessary to **know the phenomenon behind the data**



Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Data preparation

- Data preparation is very important and very time consuming
- Several phases and terms:
 - Data pre-processing
 - Data wrangling
 - Data cleaning, Data tidying ...
 - Data transformation

Data integrity becomes more essential when the volume of data increases

“Brilliant visualizations cannot redeem bad data!”

Or

“Garbage in garbage out ...”

Cleansing Data

- Data is dirty: it contains typos, inconsistencies, fails in some way to meet a standard...

Transforming Data

(at the variable level)

- Encoding
- Aggregation
- Derived data
- Removal
- Standardization

Examples:

Cleansing Data

Birth date: Feb/30/2000

Temperature: -300 °K

City: Lixboa

Transforming Data

- Encoding – answers to an open question need to be parsed and coded
- Aggregation – detail may be excessive (age: <18; 18-40; 41-65; >65)
- Derived data – add new relevant variables ($T_{\text{range}} = T_{\text{max}} - T_{\text{min}}$)
- Removal – remove data that are not needed
- Standardization – M/F; °C or °F

Main bibliography

- Camões, J., *Data at Work : Best practices for creating effective charts and information graphics in Microsoft Excel*, Pearson Education, 2016 (Chap.4)
<https://learning.oreilly.com/library/view/data-at-work/9780134268798/title.html>
- Spence, R., *Information Visualization, Design for Interaction*, 2nd ed., Prentice Hall, 2007
- Munzner, T., *Visualization Analysis and Design*, A K Peters, 2014
- Kirk, A., *Data Visualization : a successful design process*. Packt Publishing., 2012 (Chap.3)
https://books.google.pt/books/about/Data_Visualization.html?id=I4qBVLfD3t4C&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false