Interactive Visual Analysis of Epidemiological Data

Tutorial: Interactive Visual Analysis of Scientific Data

Steffen Oeltze

Outline

- Introduction
- Motivation of IVA
- Study of Health in Pomerania (SHiP)
- Norwegian Cognitive Aging Study
- Summary

What is Epidemiology?

- Study of *patterns*, causes and effects of health and disease conditions in *defined populations (cohort)*
  - Socio-demographic as well as biological factors and their correlations indicate per subject the absolute risk of getting a disease

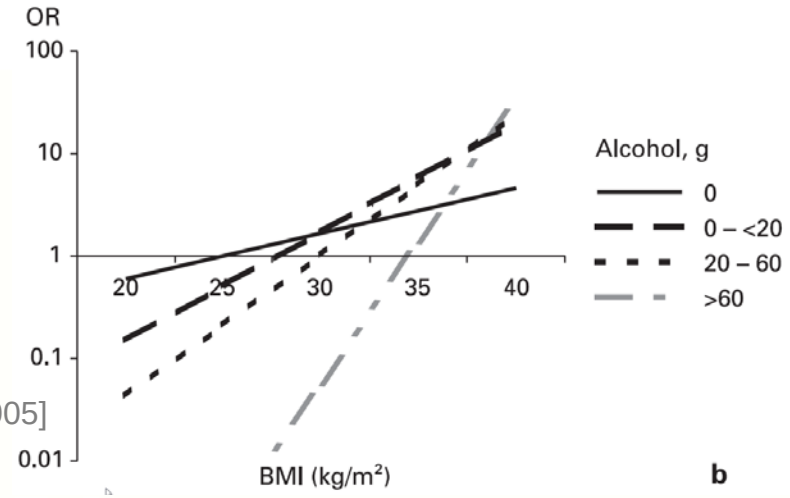
- Has huge consequences
- Establishing, evaluating, terminating preventive measures:
  - passive smoking protection,
  - recommendations for vaccinations,
  - early cancer detection strategies,…

What is Epidemiology?

- Longitudinal studies, carried out over years, comprise thousands of individuals
  - Medical image data is recently included (MRI)
- Massive amount of heterogeneous data
  - Imaging (possibly multiple modalities) and non-imaging data, e.g., questionnaires, blood tests, DNA sequencing results, ECG, biopsy results,…
  - Different domains and aggregation levels
  - (Different) parameters per individual, organ, organ part
  - Time-dependent and static parameters
  - Missing and unreliable data
  - Geographic information

Traditional Analysis in Epidemiology

- Hypothesis-driven
  - Targeted selection of parameters from the data pool
  - Confirmation or rejection of a priori hypothesis based on statistical analysis of these parameters
- Detection of inter-parameter correlations
- Identification of risk factors and *interaction terms* (relations between risk factors)



[Völzke et al.,2005]

Table 4. Independent risk factors for cholelithiasis in men*

	OR	95%-CI	p*
Age, per 10-year increase	1.61	1.50–1.73	<0.05
Body mass index, per 5 kg/m²	1.30	1.22–1.97	<0.05
HDL cholesterol, per 1 mmol/l	0.66	0.44–1.00	0.05
LDL cholesterol, per 1 mmol/l	0.88	0.77–0.99	<0.05
Daily alcohol consumption (ref. 0 g)			
>0–20 g	0.77	0.54–1.09	0.14
>20–60 g	0.72	0.51–1.03	0.07
>60 g	0.42	0.23–0.78	<0.05
Cups of coffee per day (ref. 0 cups)			
1–2	1.56	1.01–2.40	<0.05
3–4	2.42	1.57–3.72	<0.05
≥ 5	1.71	0.98–2.97	0.06
Use of tea	1.44	1.03–2.03	<0.05
Use of glucocorticoids	2.19	0.96–5.02	0.06

Potential Role of IVA

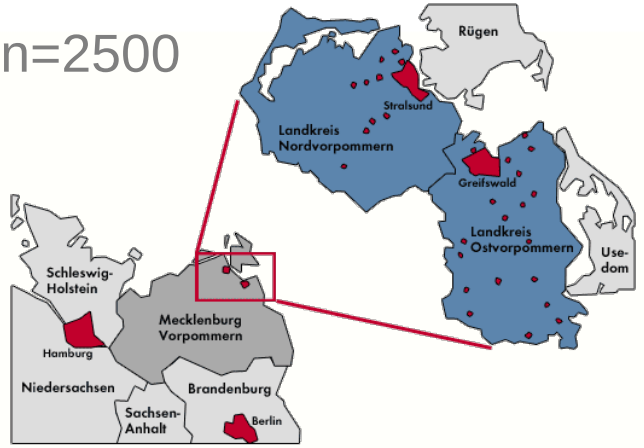
- **Generation** and testing of new hypotheses…
- …in an IVA framework:
  - Manual/automatic definition (brushing) of interesting parameters and ranges of values in attribute views
  - Linking of attributes views for identifying relations
  - Analysis across aggregation levels, parameters, subjects
  - Definition of groups by means of complex brushes or semi-automatically by means of clustering
  - Visual queries and direct feedback enable easy exploration as compared to spreadsheet-based inspection
- Due to the complex nature of cohort data, an efficient and flexible database is required.

The Study of Health in Pomerania (SHiP)

- In the past century, mortality in West Germany decreased similar as in other Western nations
- Changes less pronounced in East Germany
- After reunification in 1990, lack of scientific data from East Germany to explain the difference ➔ SHiP[Völzke,2011]
- Study is being accomplished in three waves:
  - SHiP-0: 1997-2001, 4308 adults (f/m), age: 20-79
  - SHiP-1: 5-year-follow up, n=3300
  - SHiP-2: currently running 12-year follow-up, n=2500

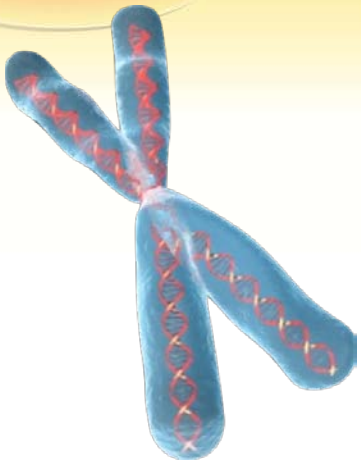
    SHiP-Trend: new sample drawn, n=8016

http://www.medizin.uni-greifswald.de/cm/fv/english/ship_en.html

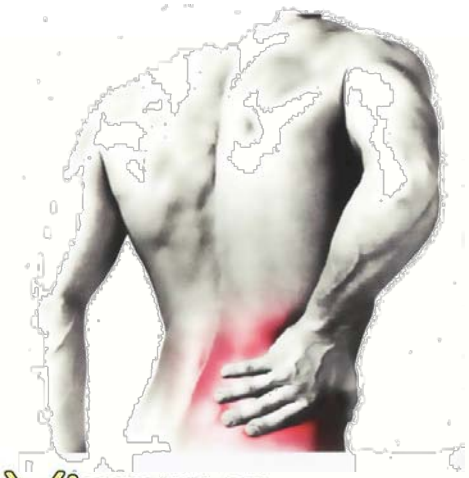


The Study of Health in Pomerania



- Interviews, laboratory test, simple-medical/dental/ skin examinations, cardiopulmonary exercise testing sleep monitoring, Ultrasound, **Whole-body-MRI**, …




Interactive Visual Analysis of Lumbar Spine Canal Variability

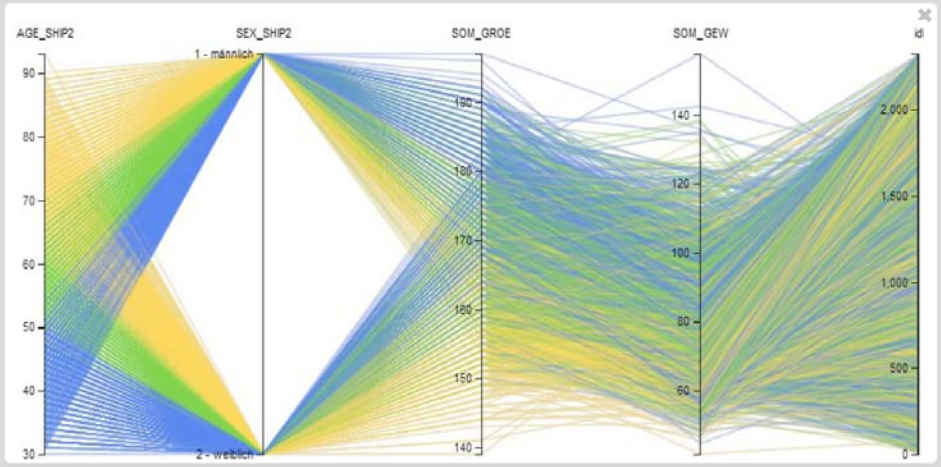
P. Klemm, K. Lawonn, M. Rak, B. Preim, K.D. Toennies, K. Hegenscheid, H. Völzke, S. Oeltze: Visualization and Analysis of Lumbar Spine Canal Variability in Cohort Study Data. Proc. of VMV (121-128), 2013.
P. Klemm, S. Oeltze, K. Hegenscheid, H. Völzke, K.D. Toennies, B. Preim: Visualization and Exploration of Shape Variance for the Analysis of Cohort Study Data. Proc. of VMV (221-222), 2012.
M. Rak, K. Engel, K.D. Tönnies: Closed-Form Hierarchical Finite Element Models for Part-Based Object Detection. Proc. of VMV (137-144), 2013.

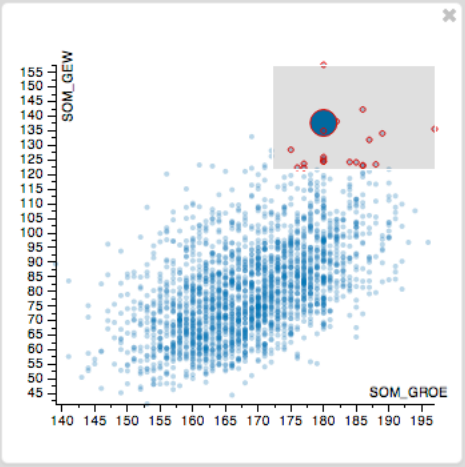
Analysis of Non-Image Data

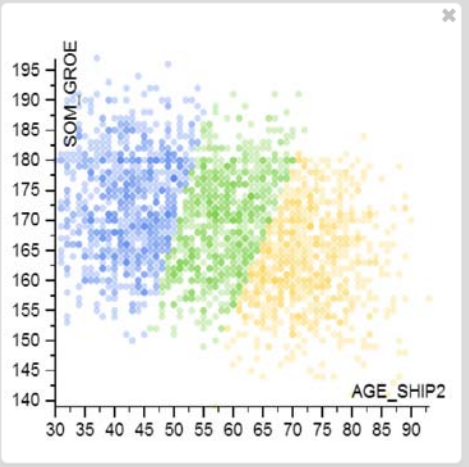
Cargo Add Visualization Contact

Showing all 21 rows

S2_SCHM...	S2_SOZI...	S2_SOZI...	S2_SOZI...	S2_SOZI...	S2_SOZI...	S2_SOZI...	BIA_KOE...	BIA_KOE...	SOM_GR...	SOM_GEW	SOM_TAIL	SOM_HU...	GES_BL_...	GES_BL_...	GES_BL_...	id	se
4 - Ja, Aus...	1 - Ja	1 - Vollzeit...	9 - noData	9 - noData	Fahrdienst...	4 - 1300 -...	53.1	38.4	182	138.2	135	144	3 - Mäßig	3 - Mäßig	1 - Gar nic...	399	
1 - Nein	2 - Nein	9 - noData	3 - arbeitsl...	9 - noData	Elektronik...	2 - 500 - ...	43.7	32.6	189	134.1	120.2	124.1	3 - Mäßig	3 - Mäßig	3 - Mäßig	479	
4 - Ja, Aus...	2 - Nein	9 - noData	7 - in Vorr...	1 - Wegen...	Berufskraf...	4 - 1300 -...	99983 - e...	NaN	188	123.5	122.1	116.2	4 - Stark	3 - Mäßig	4 - Stark	749	
1 - Nein	**2 - Nein**	**9 - noData**	**7 - in Vo...**	**1 - Wege...**	**Agrotech...**	**4 - 1300...**	**47.1**	**34.2**	**180**	**137.7**	**141.9**	**139.6**	**4 - Stark**	**3 - Mäßig**	**3 - Mäßig**	**771**	
2 - Ja, Aus...	2 - Nein	9 - noData	3 - arbeitsl...	9 - noData	Dachdecke...	3 - 900 - ...	44.5	35.7	180	124.5	120.5	125.6	3 - Mäßig	3 - Mäßig	3 - Mäßig	1046	
2 - Ja, Aus...	2 - Nein	9 - noData	7 - in Vorr...	1 - Wegen...	Dispetsche...	4 - 1300 -...	49.9	39.9	180	125.2	135.5	139.6	4 - Stark	1 - Gar nic...	2 - Kaum	1121	
9 - noData	1 - Ja	1 - Vollzeit...	9 - noData	9 - noData	Ausbildun...	7 - 2800 -...	37.7	30.7	186	123	115	112	2 - Kaum	4 - Stark	3 - Mäßig	1160	
1 - Nein	1 - Ja	1 - Vollzeit...	9 - noData	9 - noData	Fallmanager	8 - 3300 -...	38.3	29	187	131.9	113	114.5	2 - Kaum	1 - Gar nic...	2 - Kaum	1162	
9 - noData	1 - Ja	1 - Vollzeit...	9 - noData	9 - noData	Diplom-In...	9 - noData	39	31.4	185	124.2	118.3	116.5	2 - Kaum	2 - Kaum	1 - Gar nic...	1303	

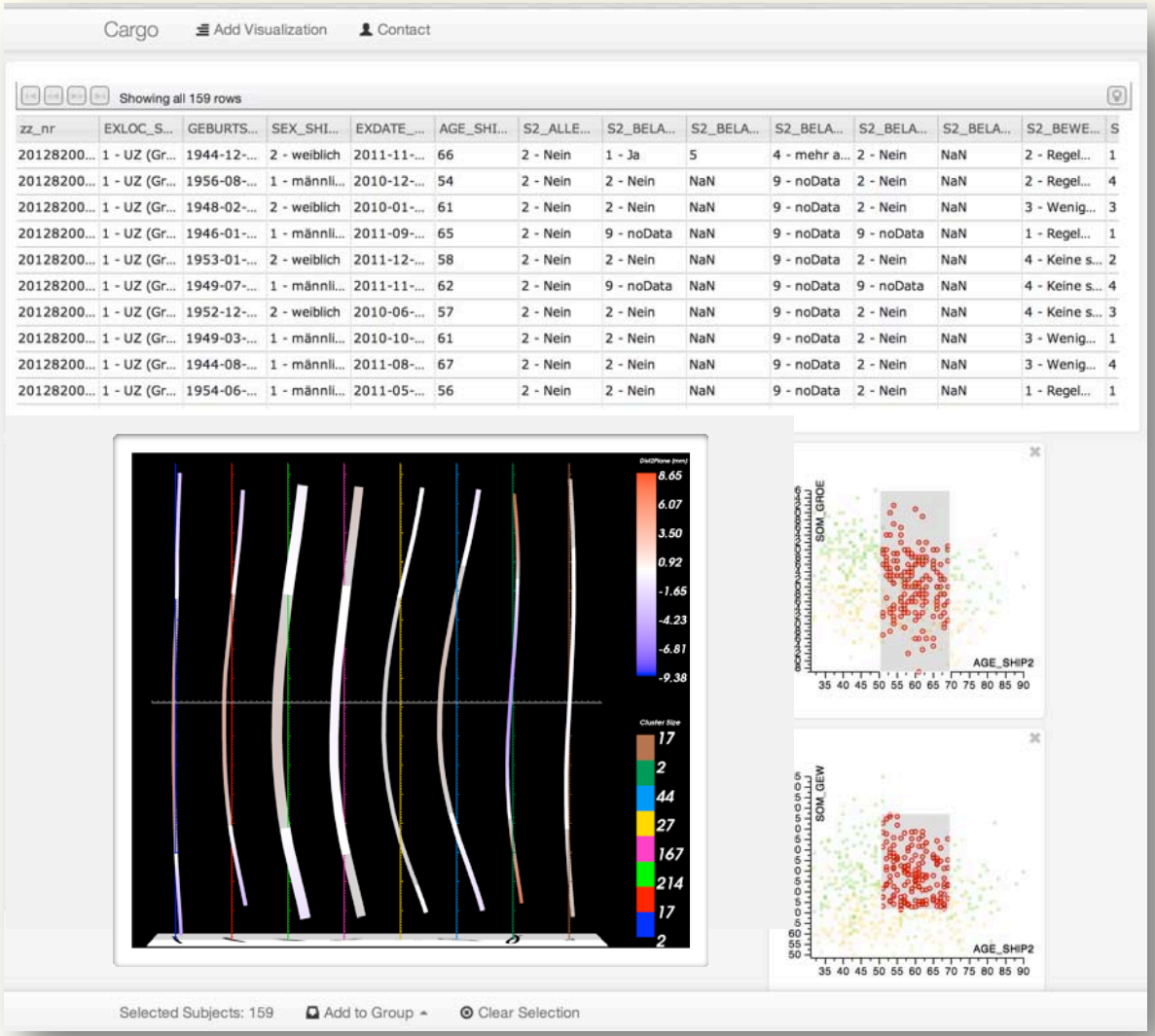




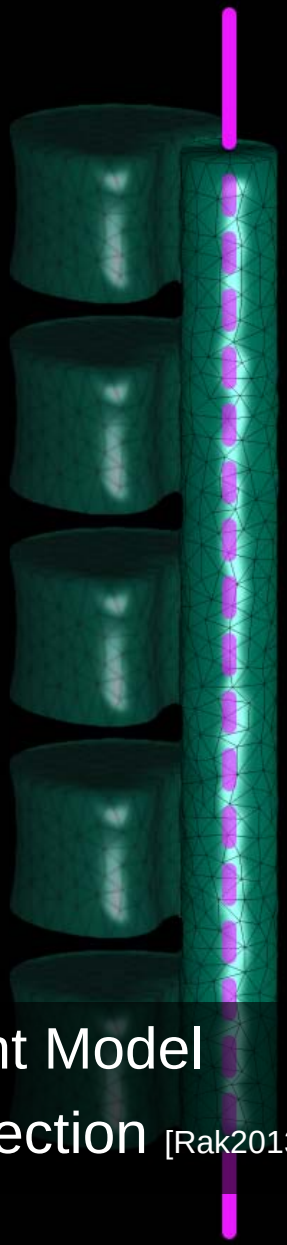


Linking to Image Data Analysis

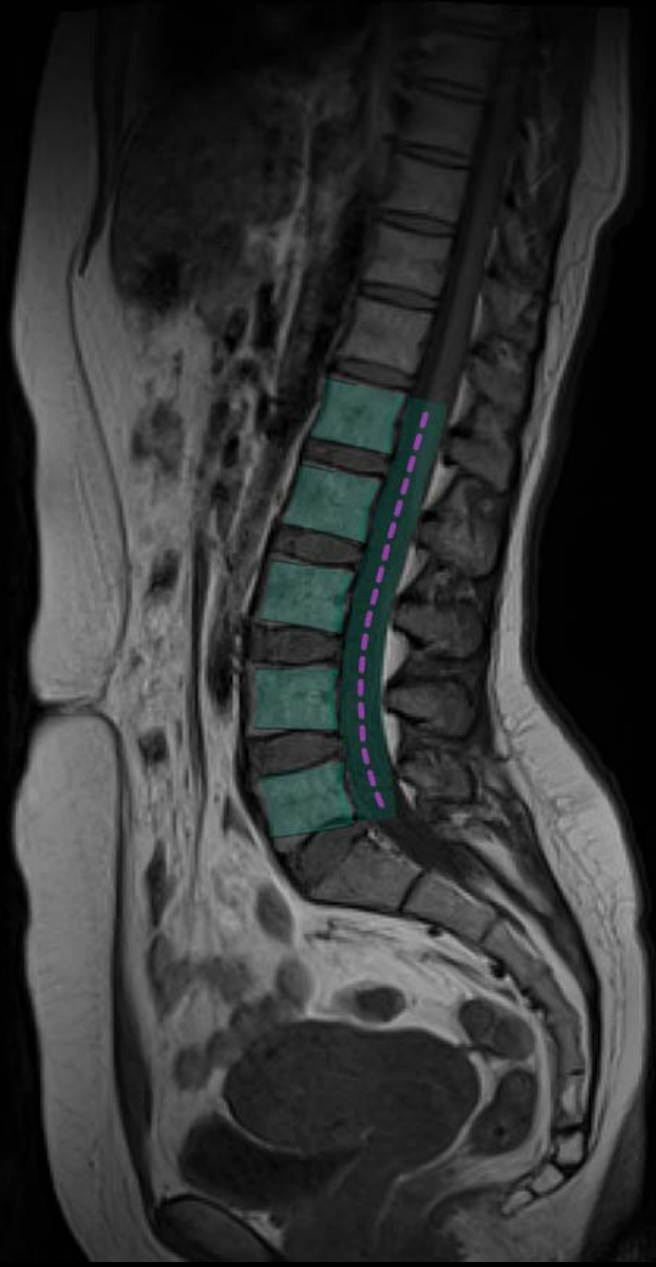
- Definition of groups in attribute views
- Parameters derived from image data of the group can be inspected in physical views
- Selection in physical views to refine the group



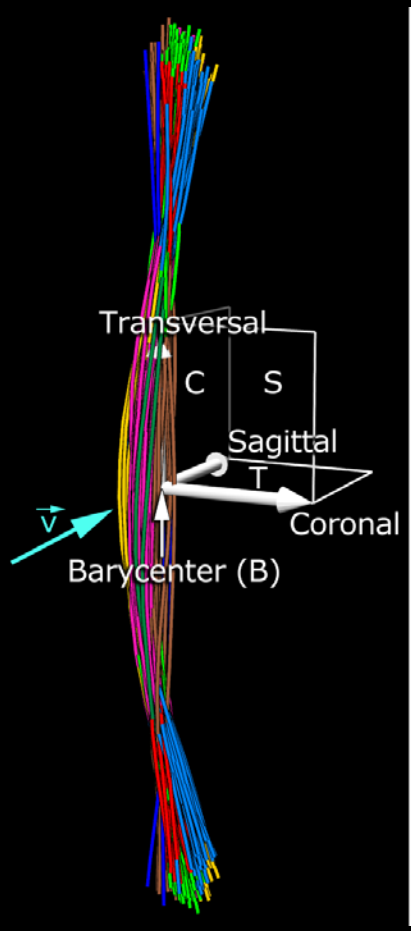
IVA of Lumbar Spine Canal Variability

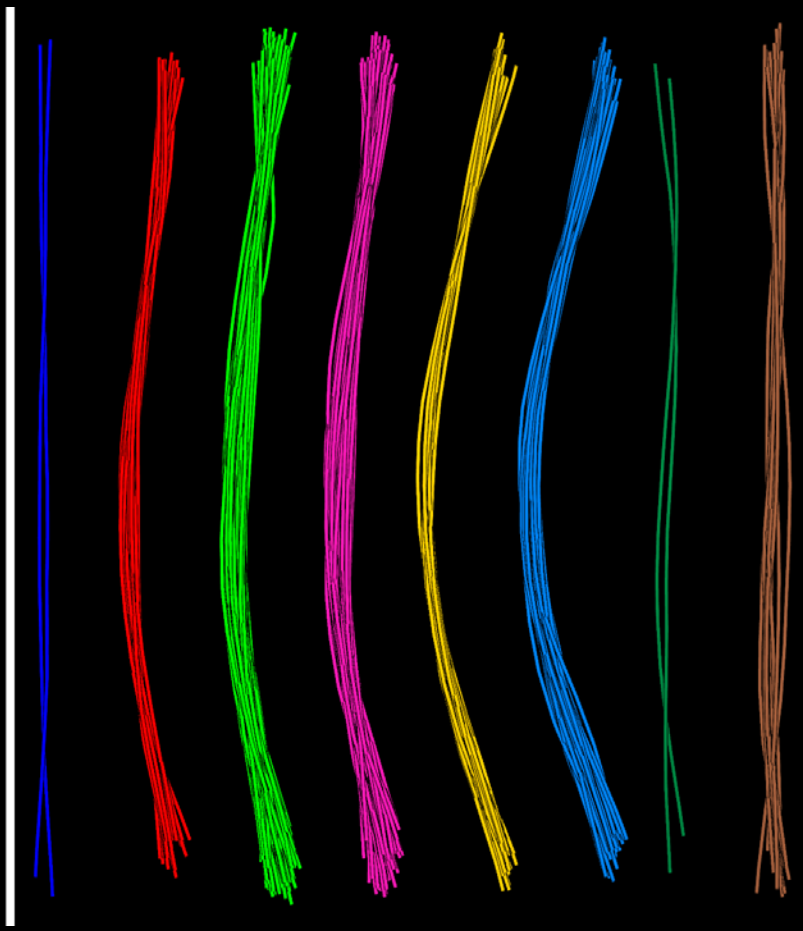


Finite-Element Model
for Spine Detection [Rak2013]



IVA of Lumbar Spine Canal Variability

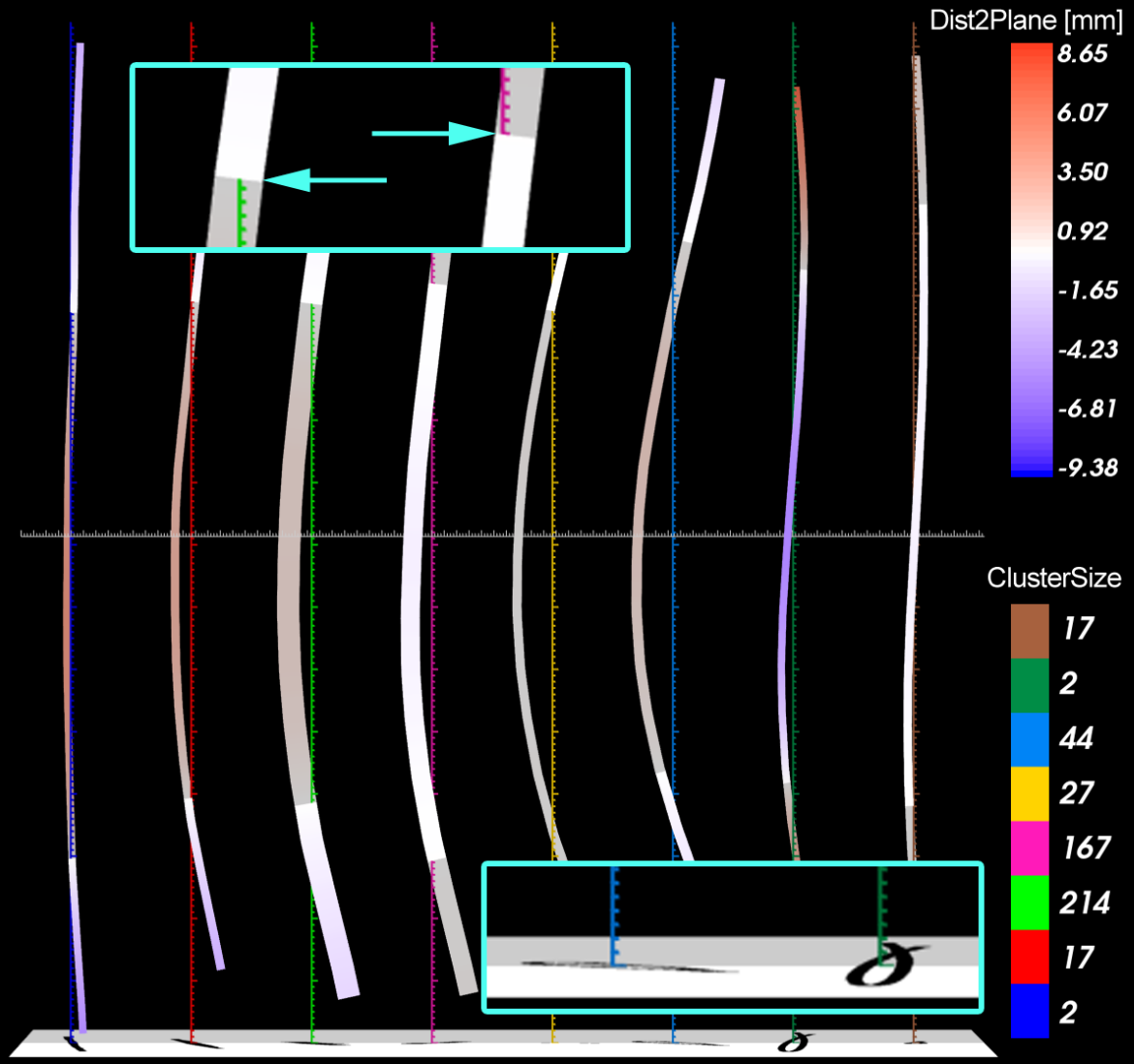


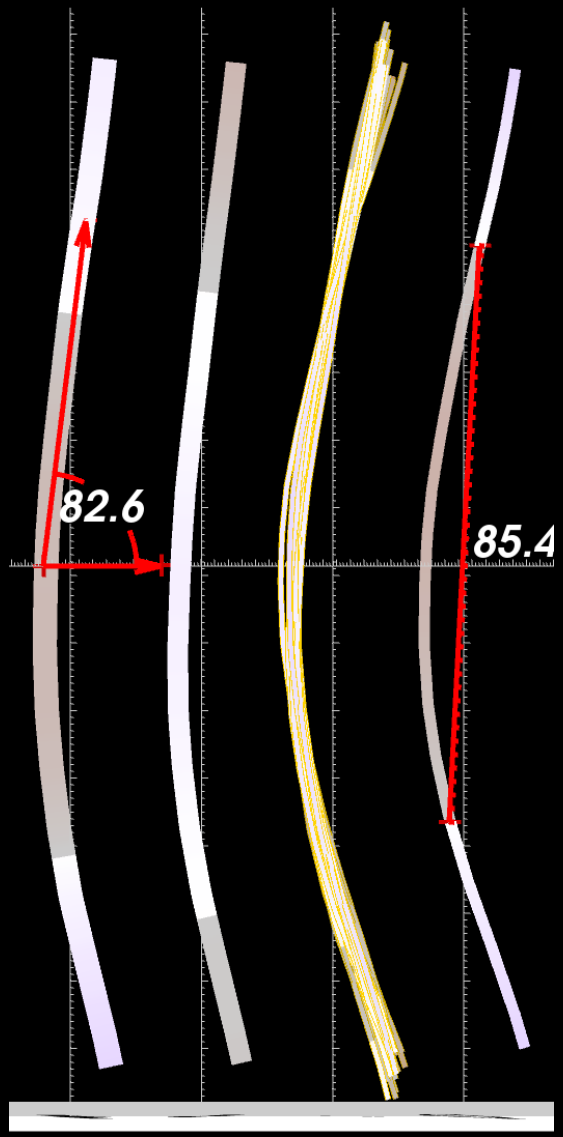


Clustering spine canal centerlines

- Analyze shape and bending of lumbar spine
- Relation of resulting clusters to parameters of the other cohort variables

IVA of Lumbar Spine Canal Variability





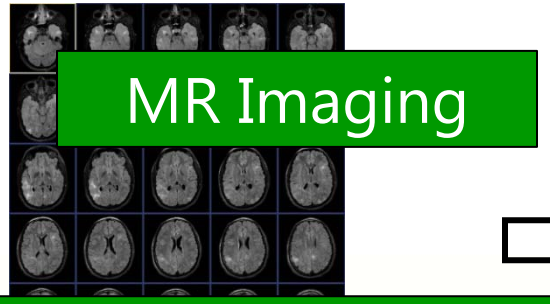
Analyze representative centerlines of clusters

Open Problems

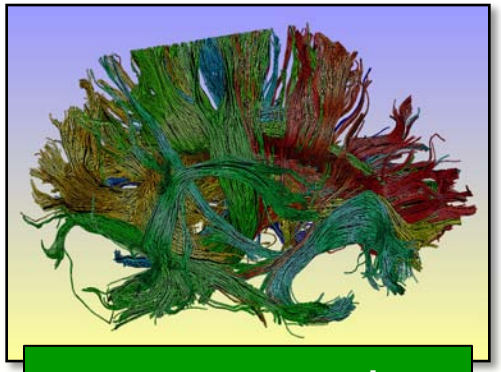
- Clustering and visualization not yet fast enough to enable direct feedback during group refinement based on attributes
- Feed back derived data, e.g., spine canal measurements, into analysis of non-image data
- Mixture of categorical and numerical attributes has impact on:
  - Design of attribute views
  - Attribute clustering
- How to cope with missing data?

Norwegian Cognitive Aging Study

- Investigates the effect of healthy aging on cognition
- >100 healthy elderly subjects examined every 3 years
- Image-data: (functional) MRI, Diffusion Tensor Imaging (DTI)
- Non-image data: personal data, neuropsychological tests



7 features computed for each of **45 brain segments**



Anisotropy measures for each fiber

Dual Analysis Method

C. Turkay, A. Lundervold, A.J. Lundervold, H. Hauser: Hypothesis Generation by Interactive Visual Exploration of Heterogeneous Medical Data. Proc of. CHI-KDD (1-12), 2013.

C. Turkay, J. Parulek, H. Hauser: Dual analysis of DNA microarrays. In Proc. of International Conference on Knowledge Management and Knowledge Technologies (p. 26), 2012.

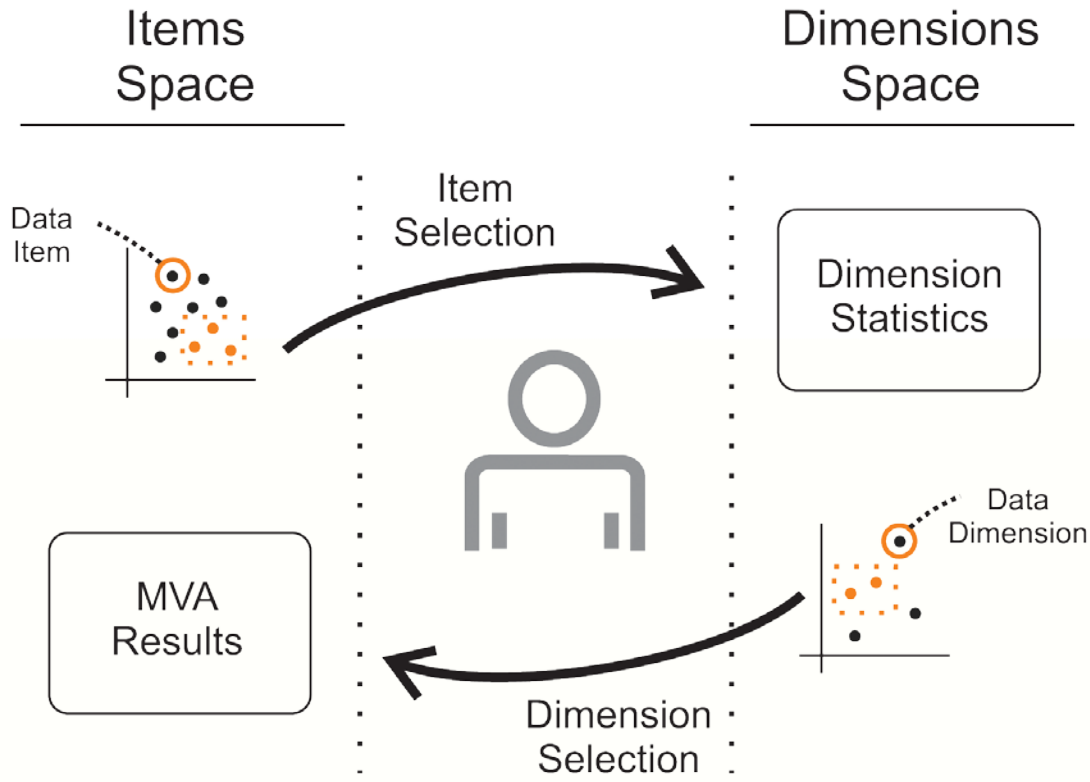
C. Turkay, P. Filzmoser, H. Hauser: Brushing Dimensions-A Dual Visual Analysis Model for High-Dimensional Data. IEEE Trans. Vis. Comput. Graph. 17(12): 2591-2599. (2011)

Dual Analysis Method

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
ID	BirthDate	Gender	Education	cvlt_tri_tc	cvlt_tri_tc	cvlt_sd_fr	cvlt_sd_fr	cvlt_ld_fr	cvlt_ld_fr	cvlt_rec_h	cvlt_rec_h	cvlt_rec_f	cvlt_rec_f	cvlt_tot_d	cvlt_tot_d	cwi_1_col	cwi_1_col	cwi_2_wo
501	1951	2	14	70	73	15	1.5	16	1.5	16	0.5	0	-1	4	1.5	30	10	22
507	1948	2	15	50	51	11	0	9	-1	15	0	0	-1	3.7	1	33	9	20
508	1947	1	15	52	64	11	1	10	0.5	15	0.5	1	-1	3.4	1	27	12	21
510	1956	1	11	56	62	12	1	12	1	13	-0.5	4	0.5	2	-0.5	37	7	28
512	1935	2	17	61	68	15	1.5	16	2	16	0.5	0	-1	4	1.5	34	10	26
514	1930	1	18	40	54	9	0.5	10	1	15	0.5	1	-1	3.4	1.5	37	9	20
517	1949	1	13	54	60	13	1.5	12	1	15	0.5	5	0.5	2.5	0	24	13	20
518	1942	1	19	45	56	9	0	11	1	14	0	1	-1	3	0.5	38	8	20
519	1944	2	14	61	65	13	1	14	1	16	0.5	1	-0.5	3.7	1	38	8	24
520	1946	1	18	33	41	6	-0.5	7	-0.5	11	-2	3	0	1.8	-1	29	12	24
523	1946	1	19	75	87	16	2.5	16	2.5	16	1	0	-1	4	2	29	12	20
524	1950	2	16	63	66	15	1.5	14	1	16	0.5	0	-1	4	1.5	26	12	21
526	1945	1	18	59	71	12	1	11	1	15	0.5	3	0	2.9	0.5	28	12	22
527	1945	2	12	63	66	15	2	14	1	14	-1	1	-0.5	3	0	27	12	18
529	1942	2	19	65	68	14	1.5	13	0.5	14	-1	2	0	2.7	-0.5	28	12	19
530	1949	1	12	50	56	10	0.5	10	0	15	0.5	0	-1	3.7	1.5	24	13	21
532	1949	1	15													34	8	25
533	1948	2	9	58	62	12	1	12	0.5	15	0	0	-1	3.7	1	26	13	21
537	1951	2	15	61	64	12	0.5	16	1.5	16	0.5	0	-1	4	1.5	27	11	20
538	1956	2	16	66	69	15	1.5	16	1.5	16	0.5	0	-1	4	1.5	28	11	19
539	1950	1	15	62	68	15	2	16	2	16	1	0	-1	4	1.5	37	7	26
542	1954	2	15	58	61	11	0	13	0.5	16	0.5	1	-0.5	3.7	1	35	8	21
543	1946	2	18	65	68	15	2	16	1.5	16	0.5	1	-0.5	3.7	1	35	9	23
544	1945	2	12	51	55	10	0	12	0.5	15	0	0	-1	1.7	1	30	11	22
545	1952	2	15	62	65	12	0.5	13	0.5	14	-0.5	0	-1	3.3	0.5	24	13	21
547	1940	2	17	66	70	16	2	16	1.5	15	0	0	-1	3.7	1	24	14	18
549	1951	2	12	75	80	16	2	16	1.5	16	0.5	0	-1	4	1.5	22	13	19
551	1938	2	10	58	62	15	2	15	1.5	15	0	1	-0.5	3.4	0.5	27	12	26
555	1947	1	16	49	60	11	1	14	2	16	1	0	-1	4	2	42	6	29
556	1932	1	15	32	45	6	-0.5	6	-0.5	11	-2	5	0.5	1.5	-1	35	9	23
558	1933	2	18	61	68	14	1.5	14	1	16	0.5	0	-1	4	1.5	32	11	23
559	1944	1	16	50	62	7	-0.5	9	0	14	0	4	0	2.3	-0.5	19	16	17
560	1951	2	20	56	58	10	-0.5	12	0	16	0.5	0	-1	4	1.5	33	9	21
564	1935	2	10	67	74	11	0.5	13	1	15	0	1	-0.5	3.4	0.5			
566	1945	2	10	53	57	13	1	12	0.5	16	0.5	0	-1	4	1.5	38	8	27
567	1949	2	14	69	72	14	1	15	1.5	16	0.5	0	-1	4	1.5	34	8	22
569	1945	2	10	57	61	12	1	13	0.5	16	0.5	8	2	2.5	-1	38	8	30

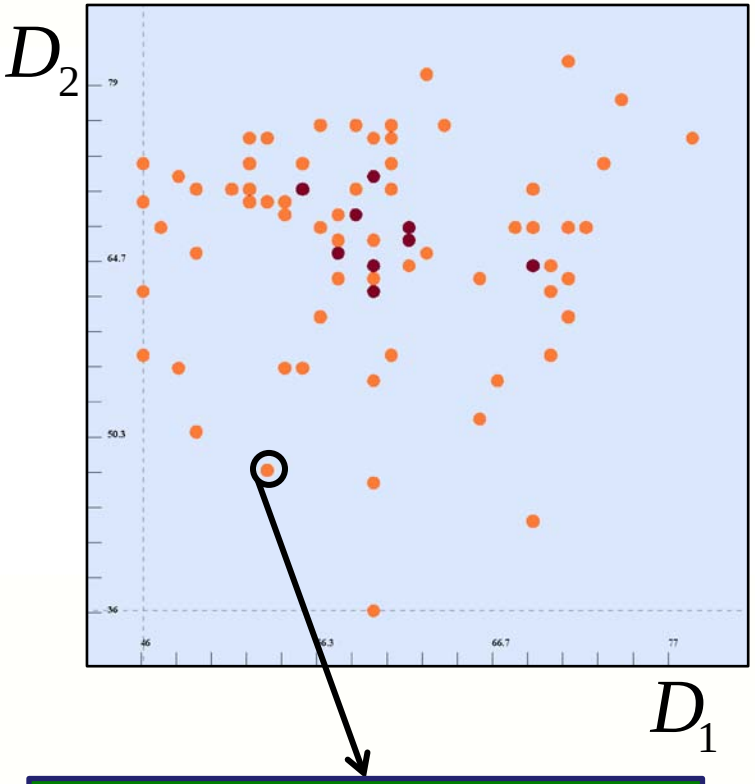
Dual Analysis Method

- Dimensions carry special **characteristics** due to their distribution of items
- Treat dimensions as **first-order analysis objects**
- Interactive visual analysis in two **linked spaces**

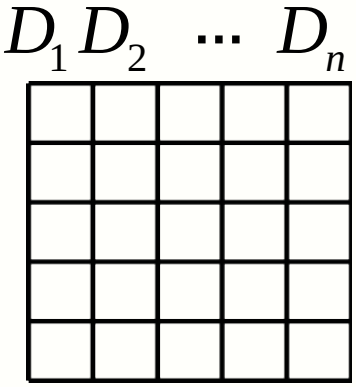


Dual Analysis Method

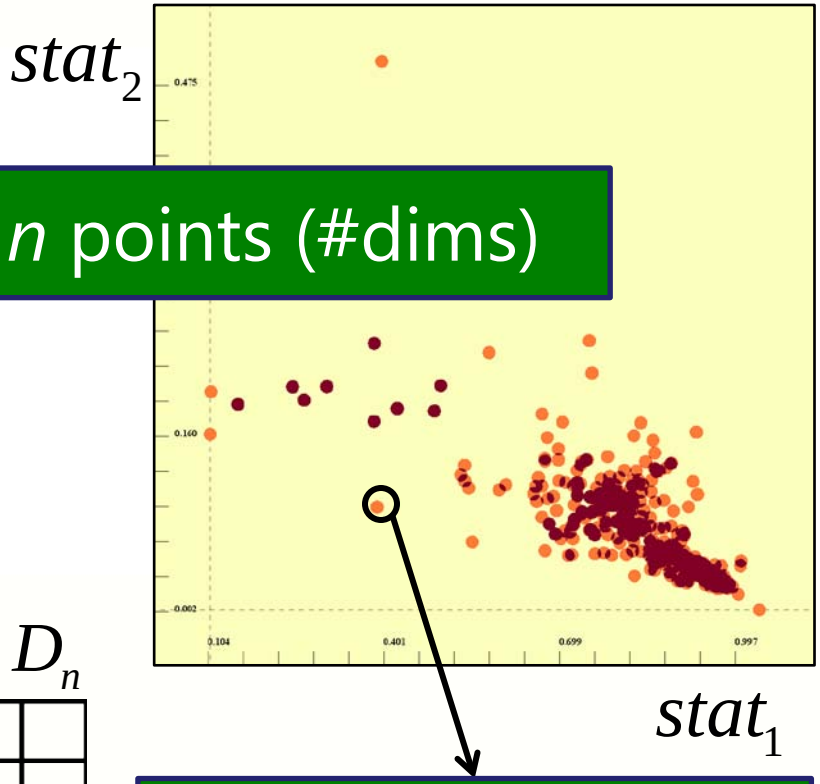
Items



A single data item



Dimensions



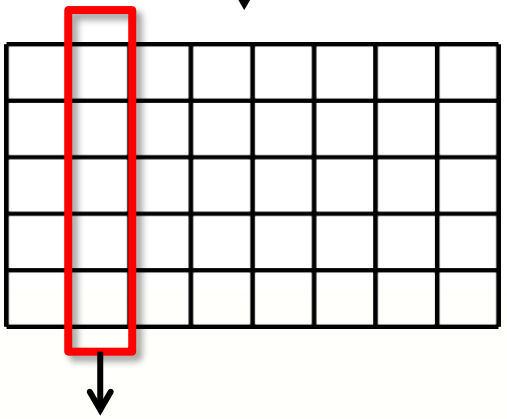
n points (#dims)

A single dimension

Visualizations in the Dimensions Space

- Dimensions are the main **visual entities !!**

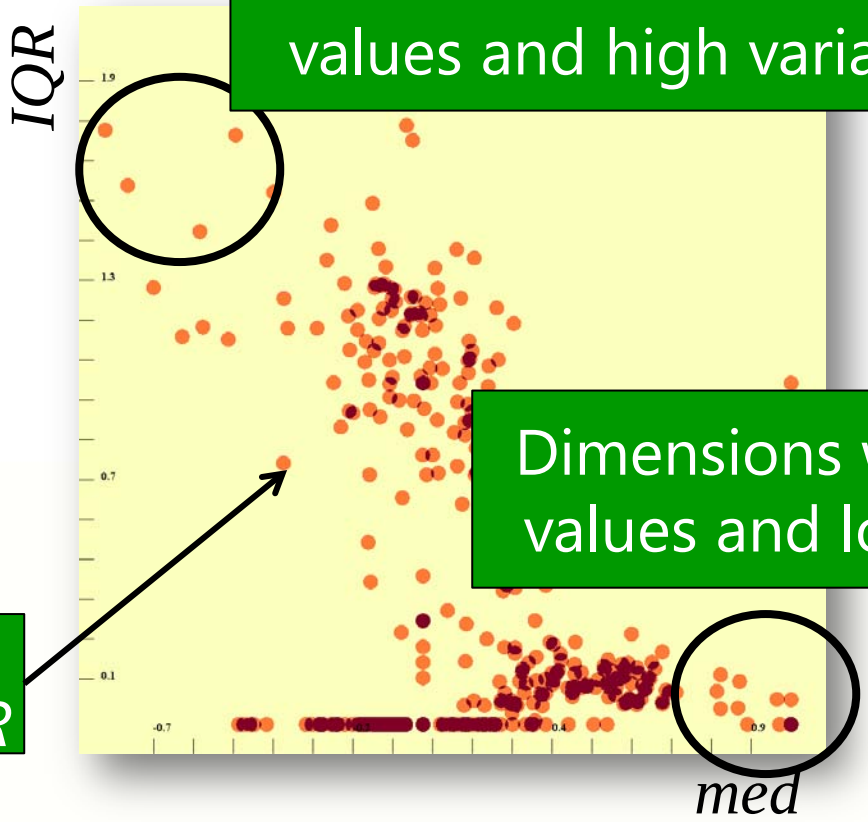
Normalize data first



For each column, compute *med* and *IQR*

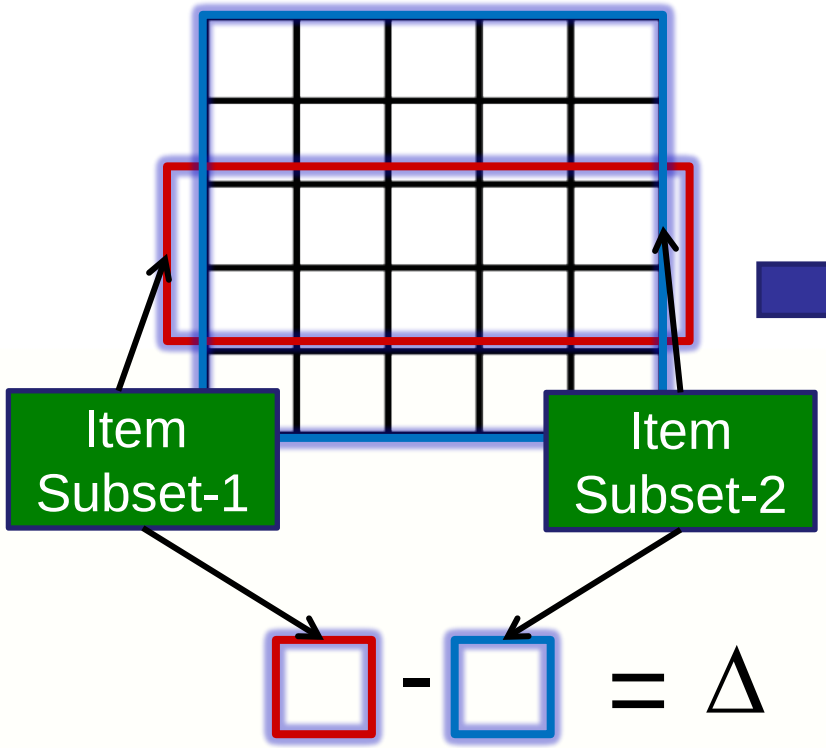
Dimensions with smaller values and high variance

Dimensions with higher values and low variance



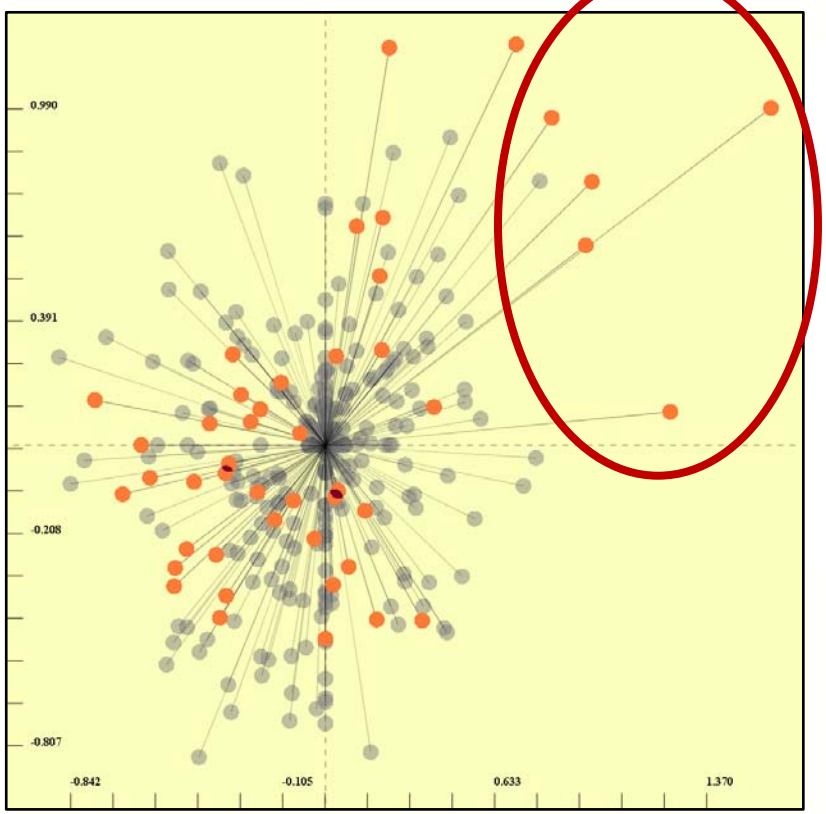
Deviation Plot

Compute *"med"* & "IQR" values using two subsets of items



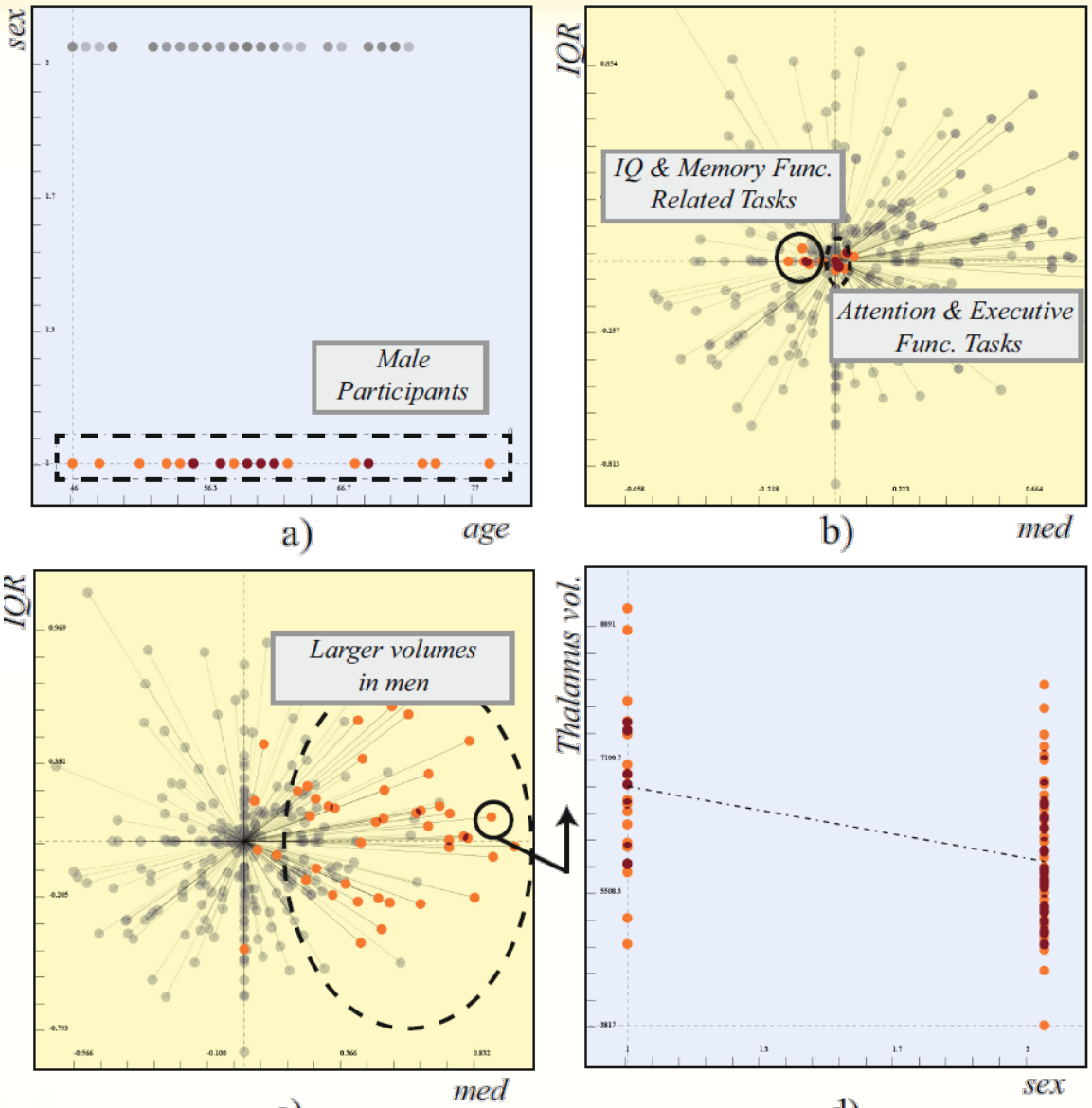
Higher values for the selection

Change in *"IQR"* values

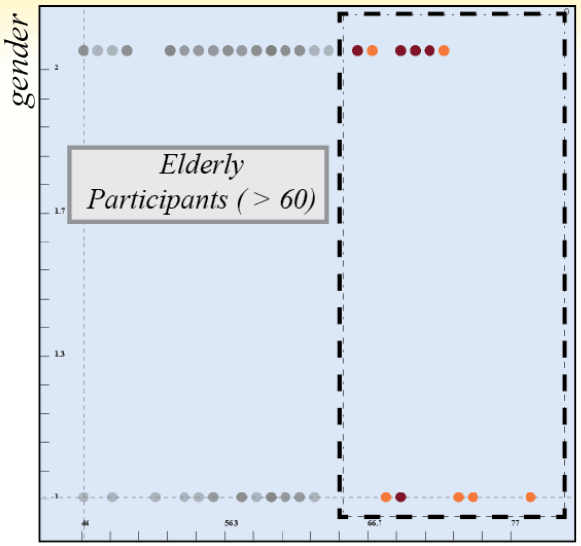


Change in "med" values

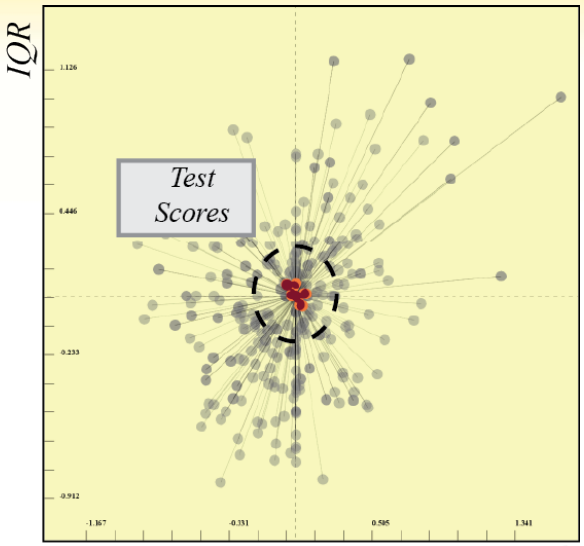
Cognitive Aging – Findings



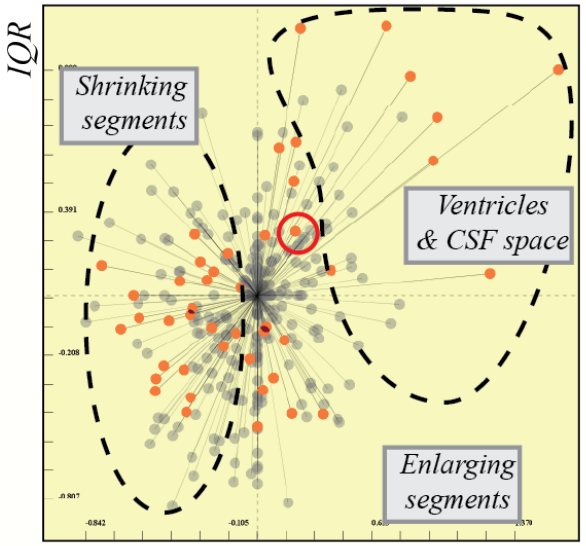
Cognitive Aging – Findings



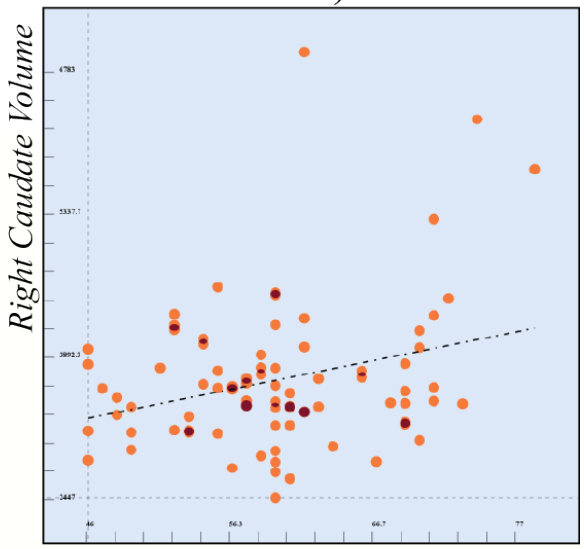
a)



b)



c)



d)

Summary

- IVA supports joint analysis of complex, heterogeneous, time-dependent imaging and non-imaging data
- IVA helps in:
  - **Hypothesis generation**/confirmation
  - Filtering, guiding the user
  - Detecting correlations between attributes
  - Analyzing dimensions and making an informed selection
  - Understanding anatomical shape variances
  - Differentiating tissue types
  - Predicting disease outbreaks
  - Planning surgery
  - Navigating complex anatomical structures
- But:
  - Few documented, IVA-based clinical research results
  - IVA neither part of clinical nor of epidemiological routine

Acknowledgements

- Helmut Doleisch, Philipp Muigg, Wolfgang Freiler
- Helwig Hauser
- Paolo Angelelli (for providing his slides and video of the CohortExplorer)
- Cagatay Turkay (for providing his slides on the dual analysis method)
- Sylvia Glaßer, Paul Klemm, Bernhard Preim
- Lisa Frauenstein
- David Perlich

- We thank all our clinical and epidemiological collaborators for acquiring and providing the data, for fruitful discussions and for the inspiring joint work.

Literature

- P. Angelelli, S. Oeltze, J. Haasz, C. Turkay, E. Hodneland, A. Lundervold, A.J. Lundervold, B. Preim, H. Hauser: *Interactive Visual Analysis of Heterogeneous Cohort Study Data*. Submitted to Computer Graphics & Applications, 2013.
- K. Hegenscheid, J. Kühn, H. Völzke, R. Biffar, N. Hosten, R. Puls: *Whole-Body Magnetic Resonance Imaging of Healthy Volunteers: Pilot Study Results from the Population-Based SHIP Study*. Fortschr Röntgenstr 181(8): 748–59. (2009)
- P. Klemm, K. Lawonn, M. Rak, B. Preim, K.D. Toennies, K. Hegenscheid, H. Völzke, S. Oeltze: *Visualization and Analysis of Lumbar Spine Canal Variability in Cohort Study Data*. Proc. of VMV (121-128), 2013.
- P. Klemm, S. Oeltze, K. Hegenscheid, H. Völzke, K.D. Tönnies, B. Preim: *Visualization and Exploration of Shape Variance for the Analysis of Cohort Study Data*. Proc. of VMV (221-222), 2012.
- M. Rak, K. Engel, K.D. Tönnies: *Closed-Form Hierarchical Finite Element Models for Part-Based Object Detection*. Proc. of VMV (137-144), 2013.
- C. Turkay, A. Lundervold, A.J. Lundervold, H. Hauser: *Hypothesis Generation by Interactive Visual Exploration of Heterogeneous Medical Data*. Proc of. CHI-KDD (1-12), 2013.
- C. Turkay, J. Parulek, H. Hauser: *Dual analysis of DNA microarrays*. In Proc. of International Conference on Knowledge Management and Knowledge Technologies (p. 26), 2012.
- C. Turkay, P. Filzmoser, H. Hauser: *Brushing Dimensions-A Dual Visual Analysis Model for High-Dimensional Data.* IEEE Trans. Vis. Comput. Graph. 17(12): 2591-2599. (2011)
- H. Völzke, D. Alte, C.O. Schmidt et al.: *Cohort Profile: The Study of Health in Pomerania*. International Journal of Epidemiology 40(2): 294-307. (2011)
- H. Völzke, S.E. Baumeister, D. Alte, W. Hoffmann, C. Schwahn, P. Simon, U. John, M.M. Lerch: *Independent Risk Factors for Gallstone Formation in a Region with High Cholelithiasis Prevalence*. Digestion 71(2) :97–105. (2005)