



Easy tools for processing and exploring data

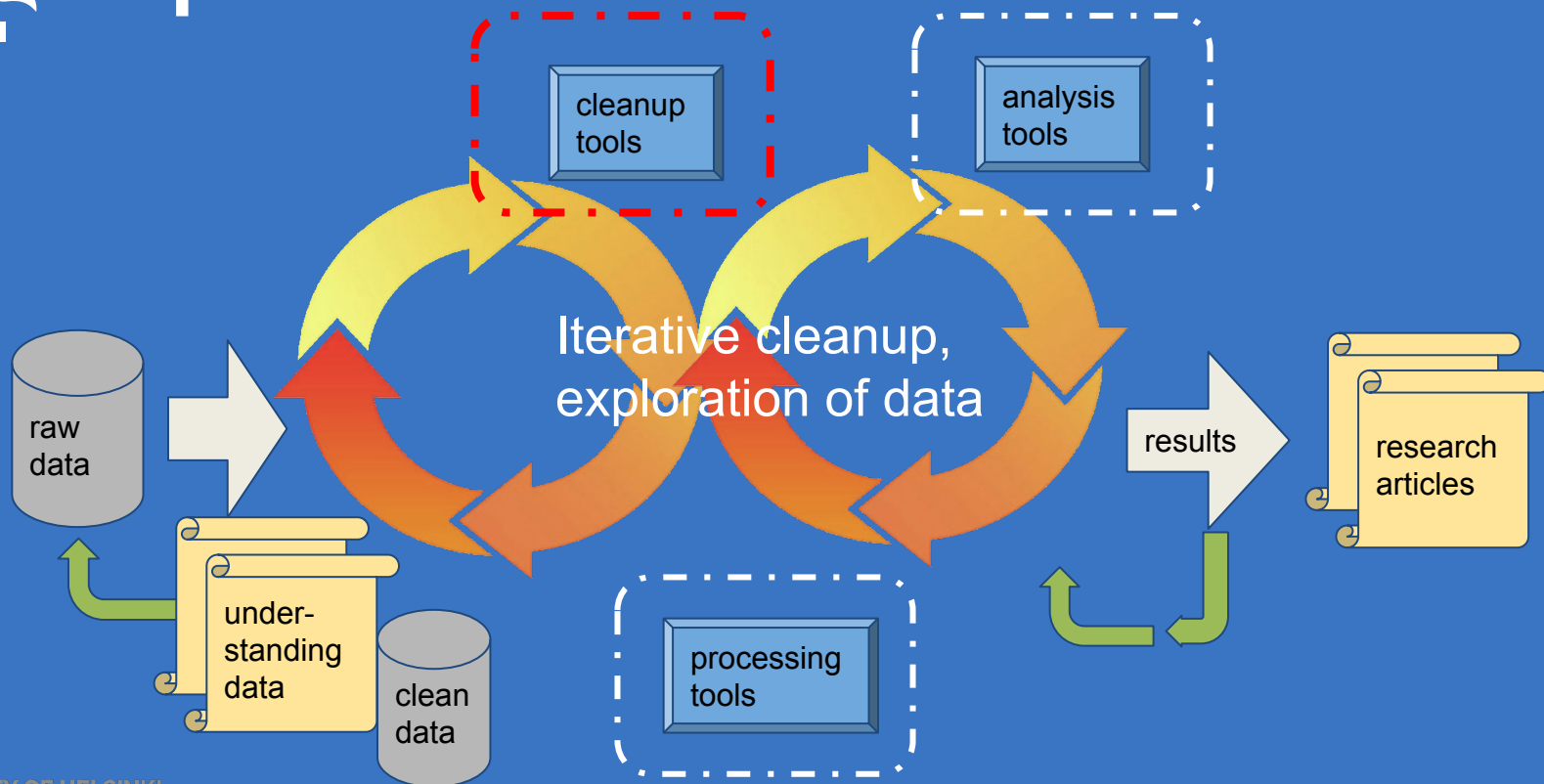
Eetu Mäkelä, D.Sc.

Professor (tenure track) in Humanities–Computing Interaction /
University of Helsinki

Docent (Adjunct Professor) in Computer Science / Aalto University



Digital humanities research process





OpenRefine

- “A free, open source, powerful tool for working with messy data”
- <http://openrefine.org/>

Cluster & Edit column "Column"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method nearest neighbor

Distance Function levenshtein

Radius 1.0

Block Chars 6

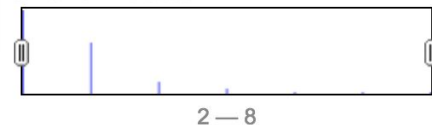
195 clusters found

		consciousne (1 rows)		
2	5	- development (4 rows) - developmen (1 rows)	☐	development
2	3	- practice (2 rows) - practise (1 rows)	☐	practice
2	2	- coherence (1 rows) - coherency (1 rows)	☐	coherence
2	2	- govermm (1 rows) - govern (1 rows)	☐	governm
2	2	- standardisation (1 rows) - standardization (1 rows)	☐	standardisation
2	3	- competition (2 rows) - ompetition (1 rows)	☐	competition
2	2	- registe (1 rows) - regist (1 rows)	☐	registe

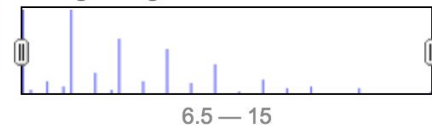
Choices in Cluster



Rows in Cluster



Average Length of Choices



Length Variance of Choices



Select All

Unselect All

Export Clusters

Merge Selected & Re-Cluster

Merge Selected & Close

Close

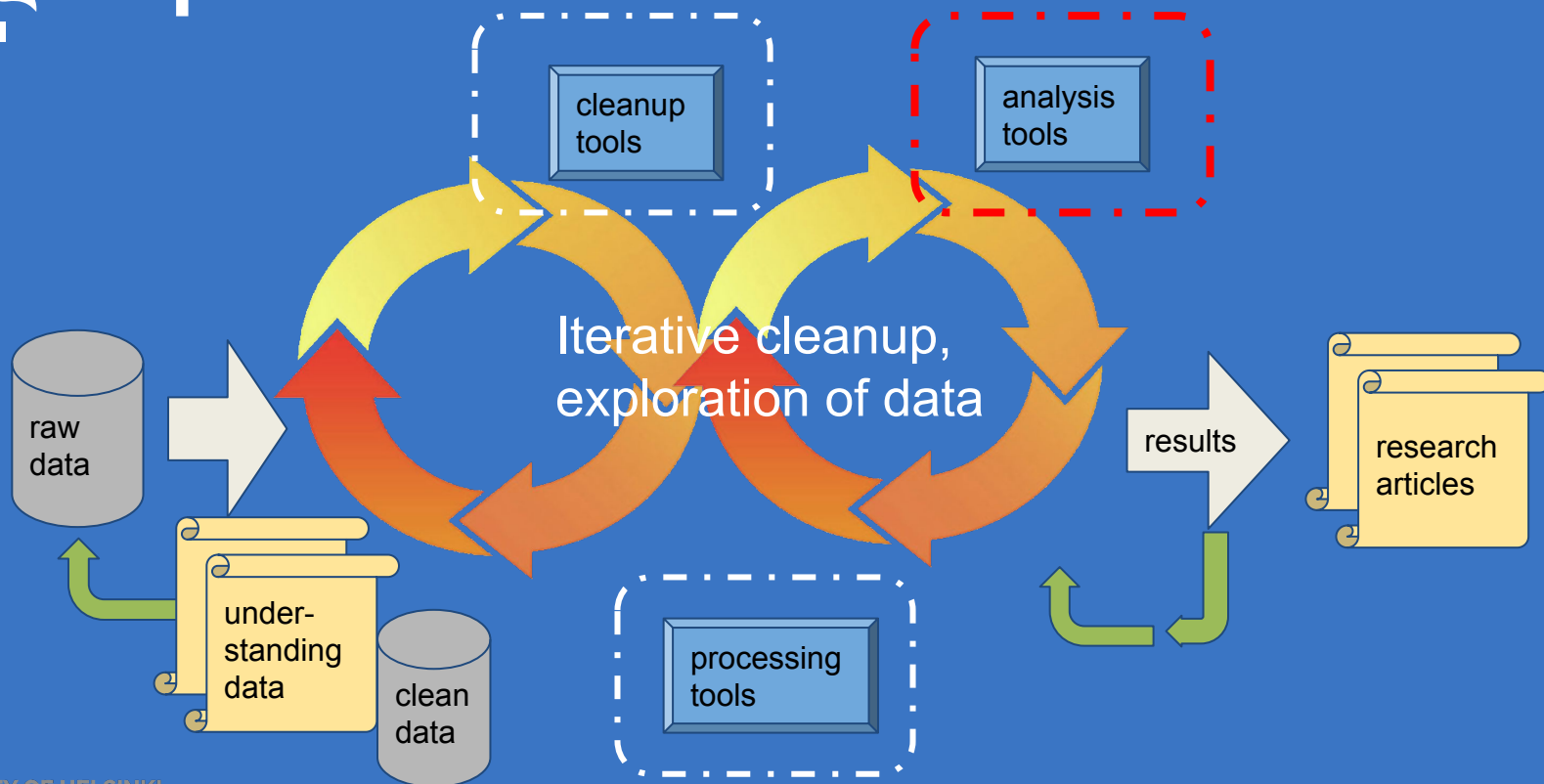


Homework

- Go through the OpenRefine tutorial at <http://freeyourmetadata.org/cleanup/>



Digital humanities research process





Different modes of engagement with data

N:o 10.

Toimittajat: J. PÄIVÄRINTA (päätoimittaja); K. LEHTONEN. — Toimituspaikka ja konttori lehden kustantaja O. B. BLOMFELT'in talossa.

1888.

SUOMI,

sanomalehti uskonnollisten, valtiollisten, kasvatustaloudellisten ja taloudellisten asiain käsittelemistä varten tarjotaan täten tilattavaksi.

Uskonnon asioissa tulemme pitämään lutherilaisen tunnustuksen-raamatullista kantaa.

Mitä valtiollisiin tulee, aivomme puollustaa perustuslaillisen edustuslaitoksemme kehittämistä kansanvaltaiseen suuntaan, ja siis kieli-asiassa jakaumatonta suomenmielisyyttä.

Kirjallisuudessa vaadimme kristilliss-siveellistä henkeä.

Kotimaan ja paikkakunnan uutisia seurataan tarkkuudella.

Ulkomaan valtiollisia ja kirkollisia asioita käsitellään niin tarkkaan, kuin tila myöntää.

Annamme mielellämme tilaa keskusteluille lehden vaikutukseen kuuluvissa asioissa, ja otamme niinkään vastaan julkaistaviksi lähetettyjä kirjoituksia.

Lehti ilmestyy tämän kokoisena joka Keskiviikkona ja Lauvantaina kello 2 päivällä ja jaetaan:

Lääketohtori W. Timgrén tavataan joka päivä kello 11—12 ja 4—5 päivällä kauppias A. Mannisen talossa Raatihuoneen- ja Mikonkadun nukkassa. Wiskaali W. Dunajeff tavataan kotona joka arkipäivä kello 9—10 ja 3—4 päivällä. Kaupungin rahaston hoitaja tavataan kotona tehtailia M. Paasosen talossa Ylikolainkadun varrella arkipäivinä kello 9—10 ja 3—4 päivällä. Suomen Pankin konttori avoinna joka arkipäivä kello 10—2 päivällä. Mikkelin läänin säästöpankki avoinna G. Laasosen perillisten talossa N:o 174 Mikonkadun varrella joka keskiviikkona ja lauantaina kello 10—11 päivällä. Yhdyspankki kello 9—11 talossa N:o 76 Nikolainkadun varrella.

Sähkölennätinasema auki toistaiseksi.

Arki- ja juhlapäivinä kello 8—2 ja 5—7. Sunnuntaina kello 9—11 ja 5—7.

Mikkelin postikonttori pidetään auki

Lähtöpostia varten:

Sunnuntaina ja juhlapäivinä kello 8—11 päivällä. Maanantaina kello 9—10, 11—1 ja 2—5 päivällä. Tiistaina kello 9—1 ja 2—4 päivällä. Keskiviikkona kello 9—2 ja 3—4 päivällä. Torstaina kello 8—12 ja 2—4 päivällä. Perjantaina kello 8—10 ja 11—3 päivällä. Lauantaina kello 9—2 ja 3—4 päivällä.

Talonestus vasten:

KOULUJEN VIRSIKIRJA,

vähintään 50 uudesta virsikirjasta valittua virren osaa koulun rukoushetkiä varten, tarjotaan täten tilattavaksi.

Hinta 25 p. nidottuna,

50 „ sidottuna paperikansiin.

Joka tilaa 10 kpl. saapi 1:n päälliseksi.

Mikkelissä 27 p. Marrask. 1888.

J. Päivärinta.

A. Westerholm

sulkee kunnioitettavan yleisön suosioon hyvin lajitellun

Vaate-, Koru-, Viini-, Herkku- ja Seka-tavara varastonsa.

A. Westerholm.

Lääkäri.

Allekirjoittanut vastaan

Suomen kaupunkien
PALOAPU-YHTIÖ

Hinnat konttuiliseti



Different modes of engagement with data

N:o 1.

Toimittajat: J. PÄIVÄRINTA (päätoimittaja); K. LEHTONEN. — Toimituspaikka ja konttori lehden kustantaja O. B. BLOMFELT'in talossa.

1888.

SUOMI,

sanomalehti uskonnollisten, valtiollisten, kasvatusopillisten ja taloudellisten asiain käsittelemistä varten tarjotaan täten tilattavaksi.

Uskonnon asioissa tulemme pitämään lutherilaisen tunnustuksen-raamatullista ka-

Mitä valtiollisiin tulee, aivustussalissa edustuslaitoksemme taiseen suuntaan, ja siis kieliä menmielisyyttä.

Kirjallisuudessa vaadimme keä.

Kotimaan ja paikkakunnan kuudella.

Ulkomaan valtiollisia ja lään niin tarkkaan, kuin tila n

Annamme mielellämme ti vaikutukseen kuuluvissa asioissa vastaan julkaistaviksi lähetetty.

Lehti ilmestyy tämän kokon ja Lauvantaina kello 2 päivällä

Lääketohtori W. Timgrén tavataan joka päivä klo 11—12 ja 4—5 päivällä kauppias A. Mannisen talossa Raati-huoneen- ja Mikonkadun nuckassa. Wiskaali W. Dunajeff tavataan kotona joka arkipäivä klo 9—10 ja 3—4 päivällä. Kaupungin rahaston hoitaja tavataan kotona tehtailia M. Paasosen talossa Ylikolainkadun varrella arkipäivinä klo 9—10 ja 3—4 päivällä.

Suomen — Tiedon konttori — asunoo joka

KOULUJEN VIRSIKIRJA,

vähintään 50 uudesta virsikirjasta valittua virren osaa koulun rukoushetkiä varten, tarjotaan täten tilattavaksi.

Hinta 25 p. nidottuna,

50 „ sidottuna paperikansiin.

Joka tilaa 10 kpl saani 1 n päälliseksi.

- <String ID="P1_ST00017" WIDTH="103" HEIGHT="33" VPOS="666" HPOS="172" CC="93307" WC="0.48" CONTENT="Nuden"/>
- <SP ID="P1_SP00010" WIDTH="23" VPOS="699" HPOS="275"/>
- <String ID="P1_ST00018" WIDTH="112" HEIGHT="31" VPOS="666" HPOS="298" CC="005637" WC="0.61" CONTENT="muoden"/>
- <SP ID="P1_SP00011" WIDTH="21" VPOS="699" HPOS="410"/>
- <String ID="P1_ST00019" WIDTH="92" HEIGHT="41" VPOS="665" HPOS="431" CC="06706" WC="0.97" CONTENT="Herra"/>
- <SP ID="P1_SP00012" WIDTH="24" VPOS="706" HPOS="523"/>
- <String ID="P1_ST00020" WIDTH="95" HEIGHT="38" VPOS="666" HPOS="547" CC="700671" WC="0.61" CONTENT="annoi;/>
- </TextLine>
- <TextLine ID="P1_TL00009" WIDTH="417" HEIGHT="40" VPOS="714" HPOS="171">
- <String ID="P1_ST00021" WIDTH="90" HEIGHT="29" VPOS="717" HPOS="171" CC="67251" WC="0.97" CONTENT="Uuden"/>
- <SP ID="P1_SP00013" WIDTH="24" VPOS="746" HPOS="261"/>
- <String ID="P1_ST00022" WIDTH="201" HEIGHT="29" VPOS="717" HPOS="285" CC="60454031450" WC="0.66" CONTENT="armo-muoden"/>
- <SP ID="P1_SP00014" WIDTH="21" VPOS="746" HPOS="486"/>
- <String ID="P1_ST00023" WIDTH="81" HEIGHT="40" VPOS="714" HPOS="507" CC="2760" WC="0.97">



Different modes of engagement with data

N:o 10. Toimittajat: J. PÄIVÄRINTA (päätoimittaja); K. LEHTONEN. — Toimituspaikka ja konttori lehden kustantaja O. B. BLOMFELT'in talossa. **1888.**

SUOMI,

sanomalehti uskonnollisten, valtiollisten, kasvatusopillisten ja taloudellisten asiain käsittelemistä varten tarjotaan täten tilattavaksi.

Uskonnon asioissa tulemme pitämään lutherilaisen tunnustuksen-raamatullista ka...

Lääketohtori W. Timgrén tavataan joka päivä klo 11—12 ja 4—5 päivällä kauppias A. Mannisen talossa Raati- huoneen- ja Mikonkadun nuckassa. Wiskaa! W. Dunajeff tavataan kotona joka arkipäivä klo 9—10 ja 3—4 päivällä. Kaupungin rahaston hoitaja tavataan kotona tehtailia M. Paasosen talossa Yrjökolainkadun varrella arkipäivinä klo 9—10 ja 3—4 päivällä.

KOULUJEN VIRSIKIRJA,

vähintään 50 uudesta virsikirjasta valittua virren osaa koulun rukoushetkiä varten, tarjotaan täten tilattavaksi.

Hinta 25 p. nidottuna,
50 " sidottuna paperikansiin.

Joka tilaa 10 kpl saani 1 n päälliseksi.

16 DIGI.KANSALLISKIRJASTO.FI ALPHA

ETUSIVU SANOMALEHDET AIKAKAUSLEHDET PIENPAINATTEET MUUT AINEISTOT

HAKU LEIKKEET LEHDET PÄIVÄN LEHDET ARTIKKELIT

Hakulause: historia]

☒ vaadi kaikki hakusanat ☐ sumea haku

Aikaväli: pp.kk.vvvv 31.12.1910

Julkaisupaikka: kaikki

Lehti: kaikki

Kieli: kaikki **Sivut:**

352628 hakutulosta

Osuimmat ensin

Lounas nro.83, 18.10.1887, sivu 2 : Matti Kauppinen , Pori

...Tämä on *historian* meno teillä ei useinkaan ole tätä miirystä. la mitä erittä...

<String ID="P1_ST00017" WIDTH="103" HEIGHT="33" VPOS="666" HPOS="172" CC="93307" WC="0.48"

"699" HPOS="275"/>
GHT="31" VPOS="666" HPOS="298" CC="005637" WC="0.61"

"699" HPOS="410"/>
GHT="41" VPOS="665" HPOS="431" CC="06706" WC="0.97"

"706" HPOS="523"/>
GHT="38" VPOS="666" HPOS="547" CC="700671" WC="0.61"

HEIGHT="40" VPOS="714" HPOS="171">
GHT="29" VPOS="717" HPOS="171" CC="67251" WC="0.97"

"746" HPOS="261"/>
GHT="29" VPOS="717" HPOS="285" CC="60454031450"

"746" HPOS="486"/>
GHT="40" VPOS="714" HPOS="507" CC="2760" WC="0.97"



Search interfaces

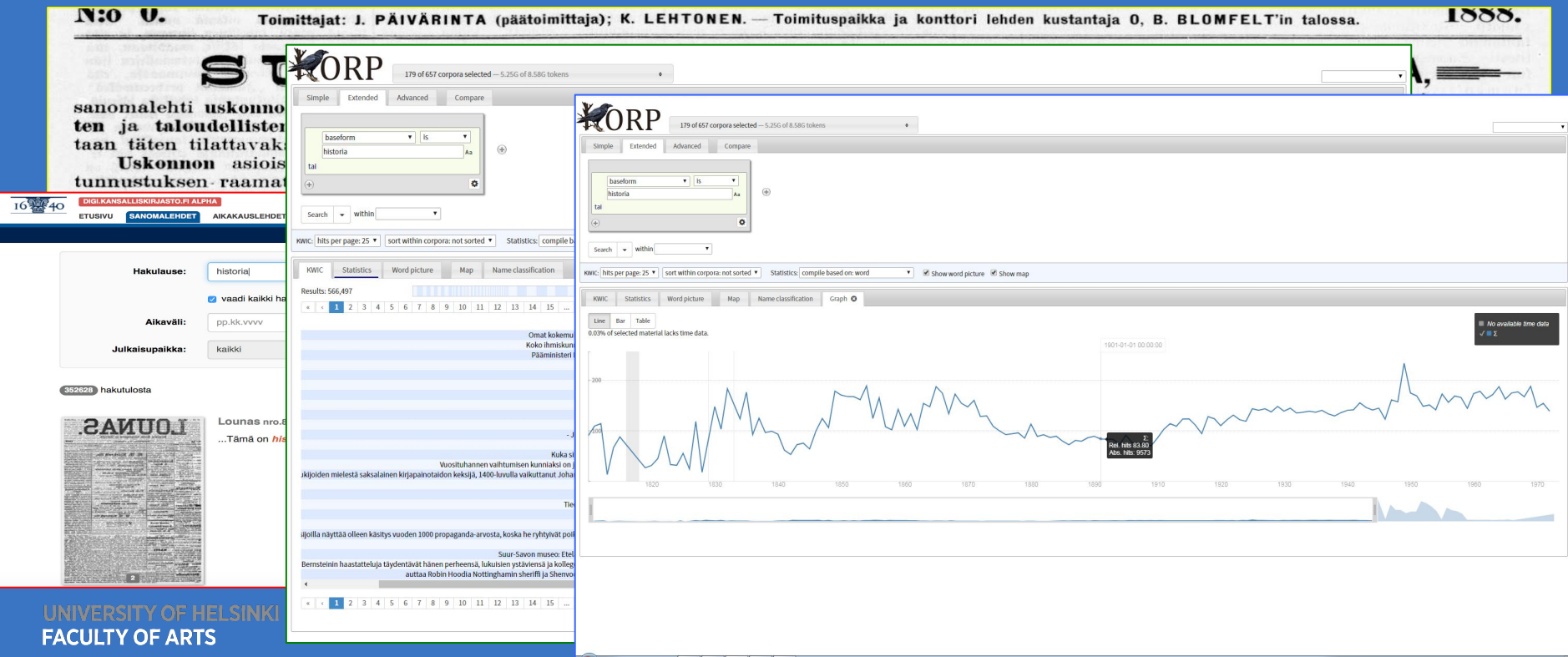
- Texts:
 - [Eighteenth Century Collections Online, ECCO-TCP](#)
 - [Digi - Kansalliskirjaston digitoidut aineistot](#)
 - ...
- Other cultural heritage:
 - [Europeana](#)
 - [Digital Public Library of America](#)
 - [British Museum](#)
 - [Finnish National Gallery](#)
 - [Early Modern Letters Online](#)
 - ...



1888.



Different modes of engagement with data





corpus.byu.edu

Corpus	# words	language/dialect	time period
NOW Corpus	2.8 billion+	20 countries / Web	2010-yesterday
Global Web-Based English (GloWbE)	1.9 billion	20 countries / Web	2012-13
Wikipedia Corpus	1.9 billion	English	-2014
Hansard Corpus (British Parliament)	1.6 billion	British	1803-2005
Corpus of Contemporary American English (COCA)	520 million	American	1990-2015
Corpus of Historical American English (COHA)	400 million	American	1810-2009
TIME Magazine Corpus	100 million	American	1923-2006
Corpus of American Soap Operas	100 million	American	2001-2012
British National Corpus (BYU-BNC)*	100 million	British	1980s-1993
Strathy Corpus (Canada)	50 million	Canadian	1970s-2000s
CORE Corpus NEW	50 million	Web registers	-2014



KORP (at csc.fi)

E-thesis	Abstracts of doctoral theses
FinnTreeBank	Legal texts (EU) and example sentences from VISK
KLK	Newspapers 1820->2000
Literature	Classics of Finnish Literature, Aleksis Kivi, Project Gutenberg (Finnish), SKVR
Legal texts	Finlex etc.
Internet discussion corpus	Suomi24, Ylilauta
Magazines from 1990-2000	Scientific journals and some other periodicals
Other texts	Presidential New Year's speeches, Finnish-Swedish parallel corpus

Spoken language (transcriptions)	Syntactic Archives, a collection of proverbs, samples of dialectal Finnish, Digital morphological archives
Finnish as 2nd language corpus	[permission required]
Corpus of Old Finnish	Written Finnish from 1543 to 1809
Corpus of Early Modern Finnish	Written Finnish from 1809 to 1899

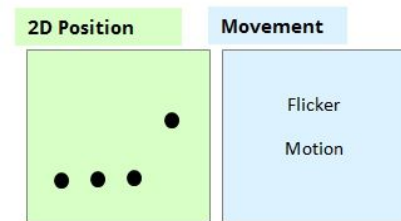
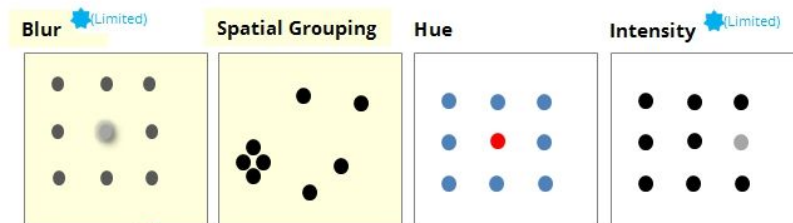
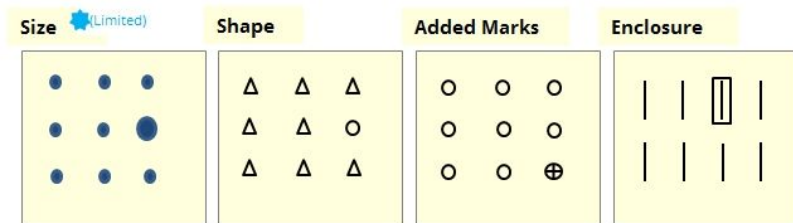
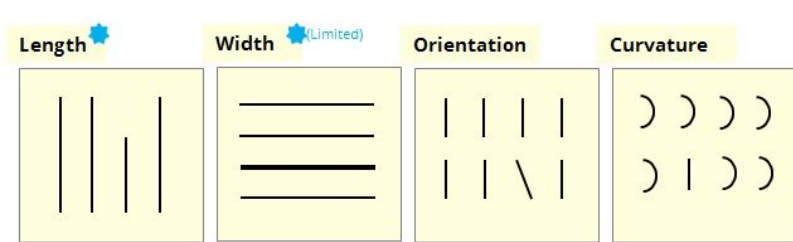


Debriefing homework: Visualisation

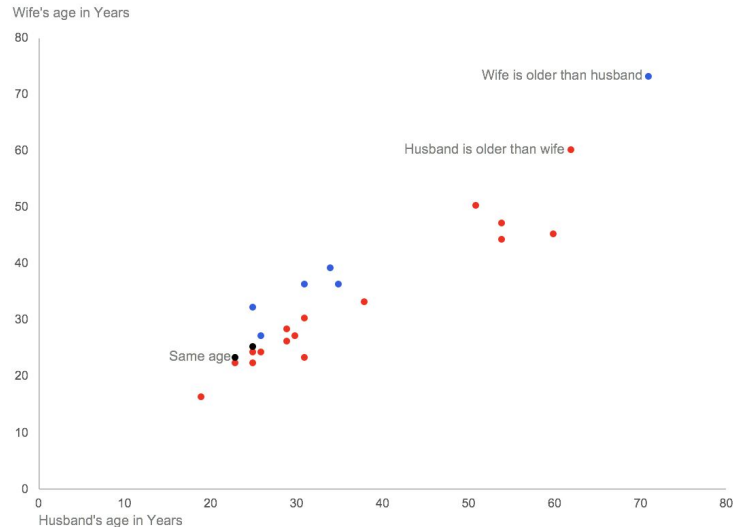
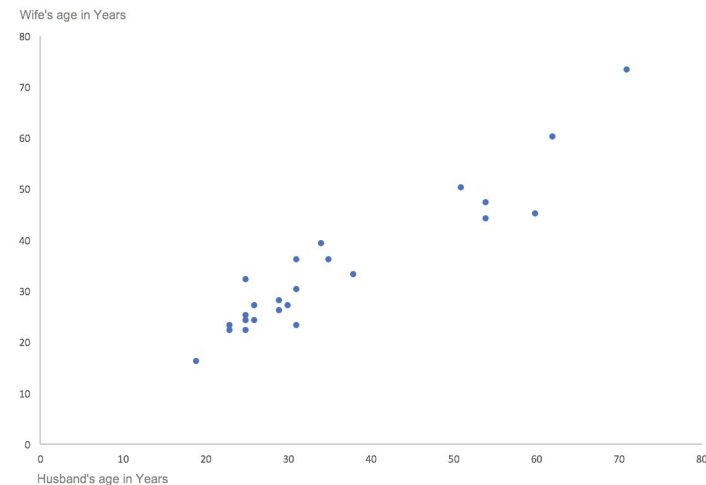


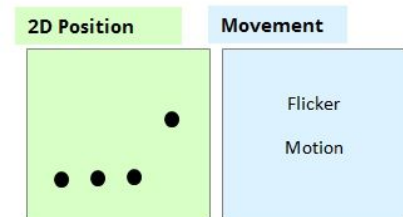
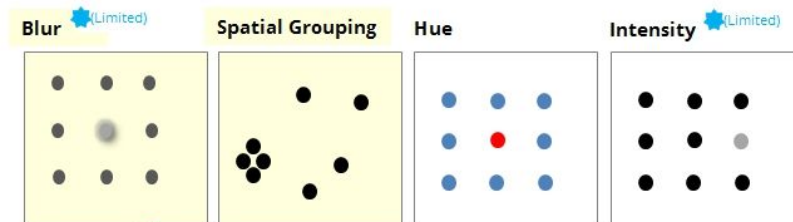
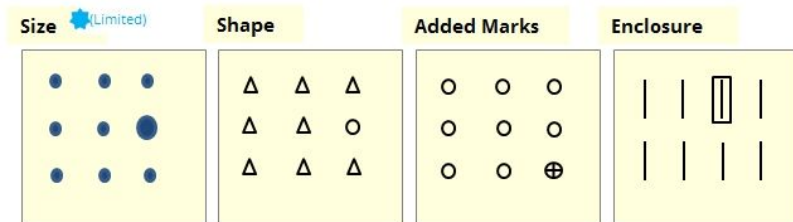
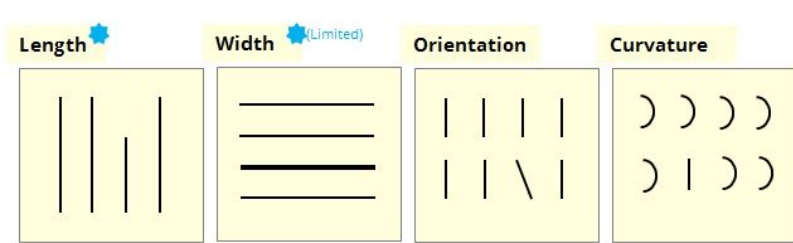
Research process - Magic (?)

- Hedge magic (spreadsheets, Excel graphs)
- Common ritual magic (statistics: correlation, regression, ANOVA, PCA)
 - Relatively simple, commonly understood formulae you could mostly go through with pen and paper if you wanted to
- Higher ritual magic (SVM, LSA, LDA, SnE)
 - More complex, harder to follow formulae, impossible to work through manually
- Black magic (most machine learning, neural networks)
 - You feed the machine both an input and a desired output, it derives a mostly unintelligible black box that links the two
- Flashy magic (proper visualizations)

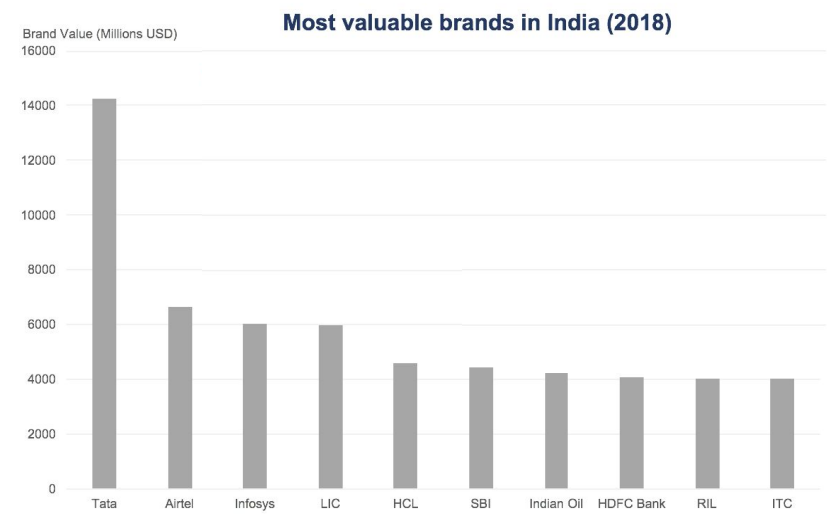
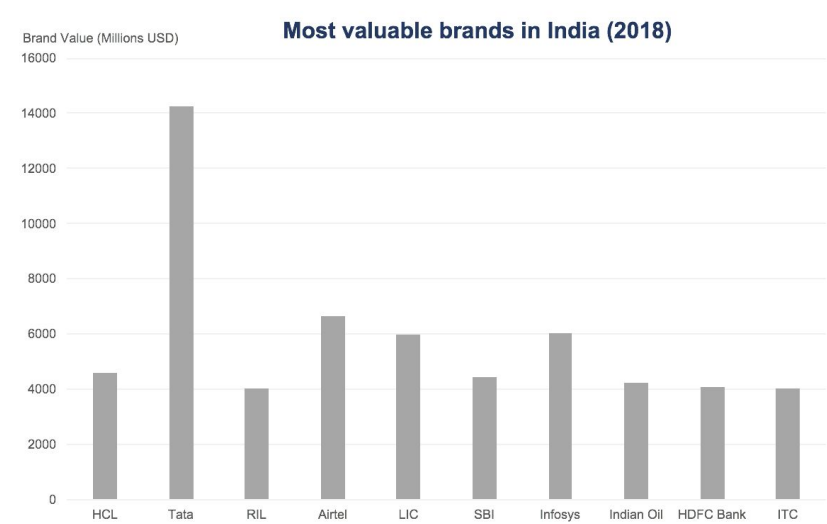


Preattentive processing
of visual information





Preattentive processing of visual information





Academy grants key project funding to 101 projects

**30 million euros, i.e.
around**

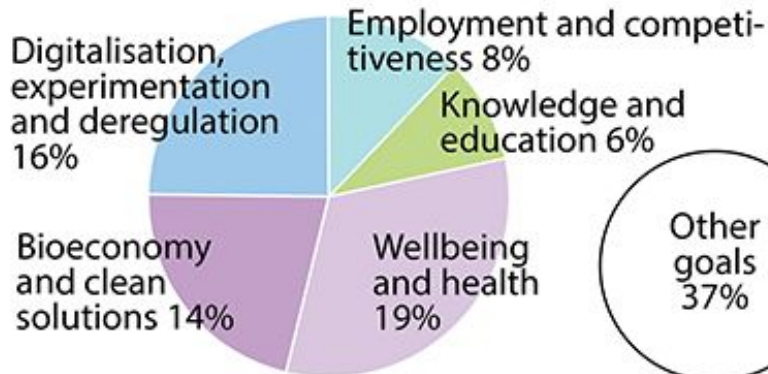
**300,000
euros per project**

598 applications
13 review panels
138 Finnish and foreign
experts reviewed the
applications
101 funded projects
(17%)

Funding recipients:

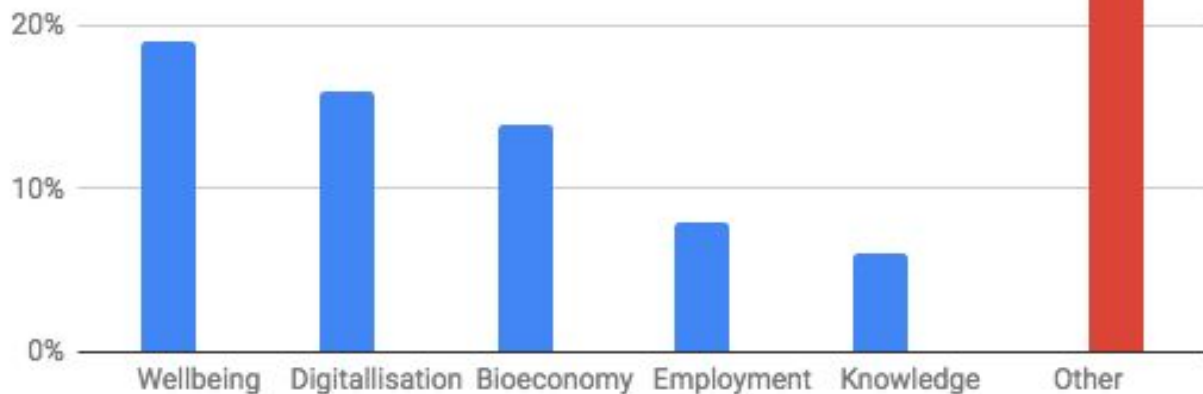
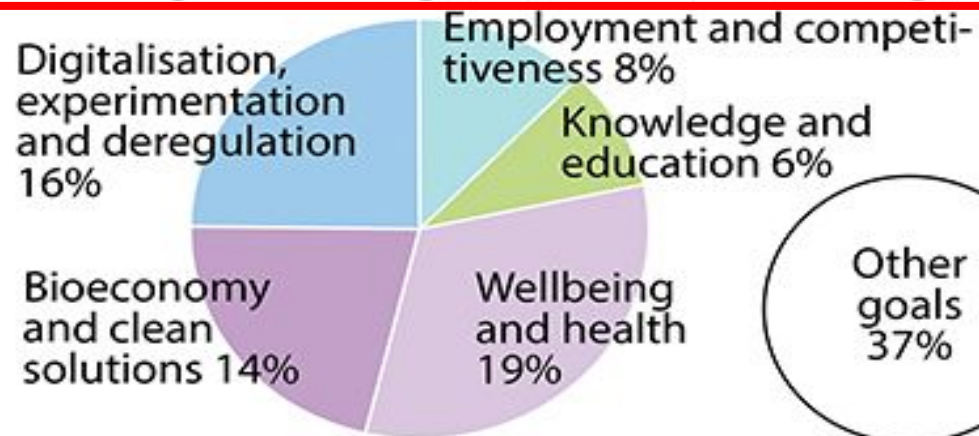


The funding matches the goals of the Government Programme



**Focus on
early-career
researchers**

The funding matches the goals of the Government Programme



Satisfaction loss by things gone wrong

	Less than "Very Satisfied"	"Very Satisfied"
Steering wheel vibrates excessively while driving	73%	
Steering wheel has too much play, feels loose	71%	
Point is uneven: has sags, runs, overspray, thin spots, etc.	71%	
Vehicle vibrates excessively when driving (not braking)	70%	
Abnormal squeaks/rattles/noises coming from front doors	69%	
Vehicle vibrates excessively at idle	69%	
Capet has gaps/poor fit	67%	
Fuel consumption is worse than expected	67%	
Hands-free phone/bluetooth system doesn't work	67%	
Abnormal noises coming from under vehicle	67%	
Lumbar support adjustment controls are difficult	67%	
Engine loses power when air conditioner is on	66%	
Foot pedals poorly located; too close/far from other pedals	66%	
Shift between gears is rough/harsh under normal driving	65%	
Brakes vibrate/shudder excessively when applied	65%	
Brakes do not have enough stopping power	65%	
Heater doesn't get hot enough/fast enough	65%	
Squeaks/rattles/abnormal noises from rear cargo shelf	64%	
"Check Engine" light indicated trouble	64%	
Moonroof controls don't work	64%	

The **top 2 issues** with the highest percentage of non-complete satisfaction are **both steering wheel related**.

Satisfaction loss by things gone wrong

	Less than "Very Satisfied"	"Very Satisfied"
Steering wheel vibrates excessively while driving	73%	
Steering wheel has too much play, feels loose	71%	
Point is uneven: has sags, runs, overspray, thin spots, etc.	71%	
Vehicle vibrates excessively when driving (not braking)	70%	
Abnormal squeaks/rattles/noises coming from front doors	69%	
Vehicle vibrates excessively at idle	69%	
Capet has gaps/poor fit	67%	
Fuel consumption is worse than expected	67%	
Hands-free phone/bluetooth system doesn't work	67%	
Abnormal noises coming from under vehicle	67%	
Lumbar support adjustment controls are difficult	67%	
Engine loses power when air conditioner is on	66%	
Foot pedals poorly located; too close/far from other pedals	66%	
Shift between gears is rough/harsh under normal driving	65%	
Brakes vibrate/shudder excessively when applied	65%	
Brakes do not have enough stopping power	65%	
Heater doesn't get hot enough/fast enough	65%	
Squeaks/rattles/abnormal noises from rear cargo shelf	64%	
"Check Engine" light indicated trouble	64%	
Moonroof controls don't work	64%	

Satisfaction loss by things gone wrong

	Less than "Very Satisfied"	"Very Satisfied"
Steering wheel vibrates excessively while driving	73%	
Steering wheel has too much play, feels loose	71%	
Point is uneven: has sags, runs, overspray, thin spots, etc.	71%	
Vehicle vibrates excessively when driving (not braking)	70%	
Abnormal squeaks/rattles/noises coming from front doors	69%	
Vehicle vibrates excessively at idle	69%	
Capet has gaps/poor fit	67%	
Fuel consumption is worse than expected	67%	
Hands-free phone/bluetooth system doesn't work	67%	
Abnormal noises coming from under vehicle	67%	
Lumbar support adjustment controls are difficult	67%	
Engine loses power when air conditioner is on	66%	
Foot pedals poorly located; too close/far from other pedals	66%	
Shift between gears is rough/harsh under normal driving	65%	
Brakes vibrate/shudder excessively when applied	65%	
Brakes do not have enough stopping power	65%	
Heater doesn't get hot enough/fast enough	65%	
Squeaks/rattles/abnormal noises from rear cargo shelf	64%	
"Check Engine" light indicated trouble	64%	
Moonroof controls don't work	64%	

Comment about **vibration**...

Satisfaction loss by things gone wrong

	Less than "Very Satisfied"	"Very Satisfied"
Steering wheel vibrates excessively while driving	73%	
Steering wheel has too much play, feels loose	71%	
Point is uneven: has sags, runs, overspray, thin spots, etc.	71%	
Vehicle vibrates excessively when driving (not braking)	70%	
Abnormal squeaks/rattles/noises coming from front doors	69%	
Vehicle vibrates excessively at idle	69%	
Capet has gaps/poor fit	67%	
Fuel consumption is worse than expected	67%	
Hands-free phone/bluetooth system doesn't work	67%	
Abnormal noises coming from under vehicle	67%	
Lumbar support adjustment controls are difficult	67%	
Engine loses power when air conditioner is on	66%	
Foot pedals poorly located; too close/far from other pedals	66%	
Shift between gears is rough/harsh under normal driving	65%	
Brakes vibrate/shudder excessively when applied	65%	
Brakes do not have enough stopping power	65%	
Heater doesn't get hot enough/fast enough	65%	
Squeaks/rattles/abnormal noises from rear cargo shelf	64%	
"Check Engine" light indicated trouble	64%	
Moonroof controls don't work	64%	

The **top 2 issues** with the highest percentage of non-complete satisfaction are **both steering wheel related**.

Satisfaction loss by things gone wrong

	Less than "Very Satisfied"	"Very Satisfied"
Steering wheel vibrates excessively while driving	73%	
Steering wheel has too much play, feels loose	71%	
Point is uneven: has sags, runs, overspray, thin spots, etc.	71%	
Vehicle vibrates excessively when driving (not braking)	70%	
Abnormal squeaks/rattles/noises coming from front doors	69%	
Vehicle vibrates excessively at idle	69%	
Capet has gaps/poor fit	67%	
Fuel consumption is worse than expected	67%	
Hands-free phone/bluetooth system doesn't work	67%	
Abnormal noises coming from under vehicle	67%	
Lumbar support adjustment controls are difficult	67%	
Engine loses power when air conditioner is on	66%	
Foot pedals poorly located; too close/far from other pedals	66%	
Shift between gears is rough/harsh under normal driving	65%	
Brakes vibrate/shudder excessively when applied	65%	
Brakes do not have enough stopping power	65%	
Heater doesn't get hot enough/fast enough	65%	
Squeaks/rattles/abnormal noises from rear cargo shelf	64%	
"Check Engine" light indicated trouble	64%	

Comment about **noises**...

Satisfaction loss by things gone wrong

	Less than "Very Satisfied"	"Very Satisfied"
Steering wheel vibrates excessively while driving	73%	
Steering wheel has too much play, feels loose	71%	
Point is uneven: has sags, runs, overspray, thin spots, etc.	71%	
Vehicle vibrates excessively when driving (not braking)	70%	
Abnormal squeaks/rattles/noises coming from front doors	69%	
Vehicle vibrates excessively at idle	69%	
Capet has gaps/poor fit	67%	
Fuel consumption is worse than expected	67%	
Hands-free phone/bluetooth system doesn't work	67%	
Abnormal noises coming from under vehicle	67%	
Lumbar support adjustment controls are difficult	67%	
Engine loses power when air conditioner is on	66%	
Foot pedals poorly located; too close/far from other pedals	66%	
Shift between gears is rough/harsh under normal driving	65%	
Brakes vibrate/shudder excessively when applied	65%	
Brakes do not have enough stopping power	65%	
Heater doesn't get hot enough/fast enough	65%	
Squeaks/rattles/abnormal noises from rear cargo shelf	64%	
"Check Engine" light indicated trouble	64%	
Moonroof controls don't work	64%	

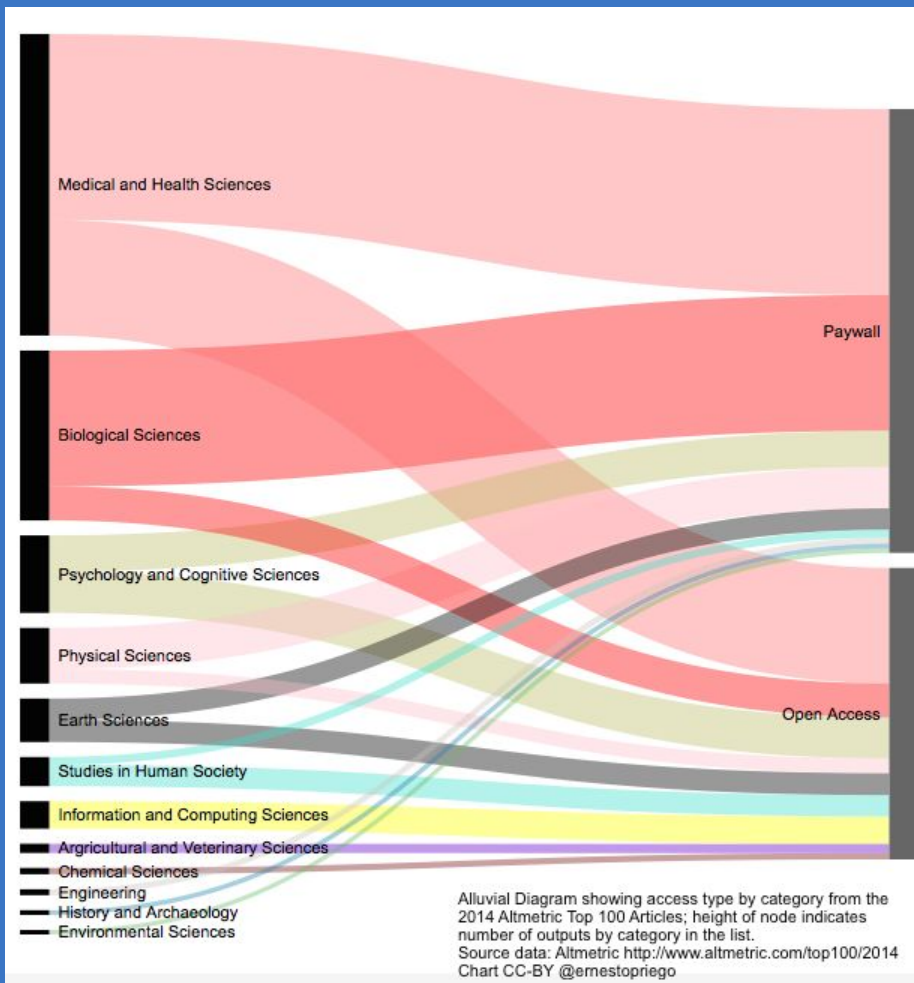


RAW

- <https://github.com/jiemakel/dhintro/blob/dad05a321dee7ab37ec4269555de428d6baa358c/socialist-frequent-words-filtered-timeseries.csv>
- <https://raw.githubusercontent.com/jiemakel/dhintro/dad05a321dee7ab37ec4269555de428d6baa358c/socialist-frequent-words-filtered-timeseries.csv>

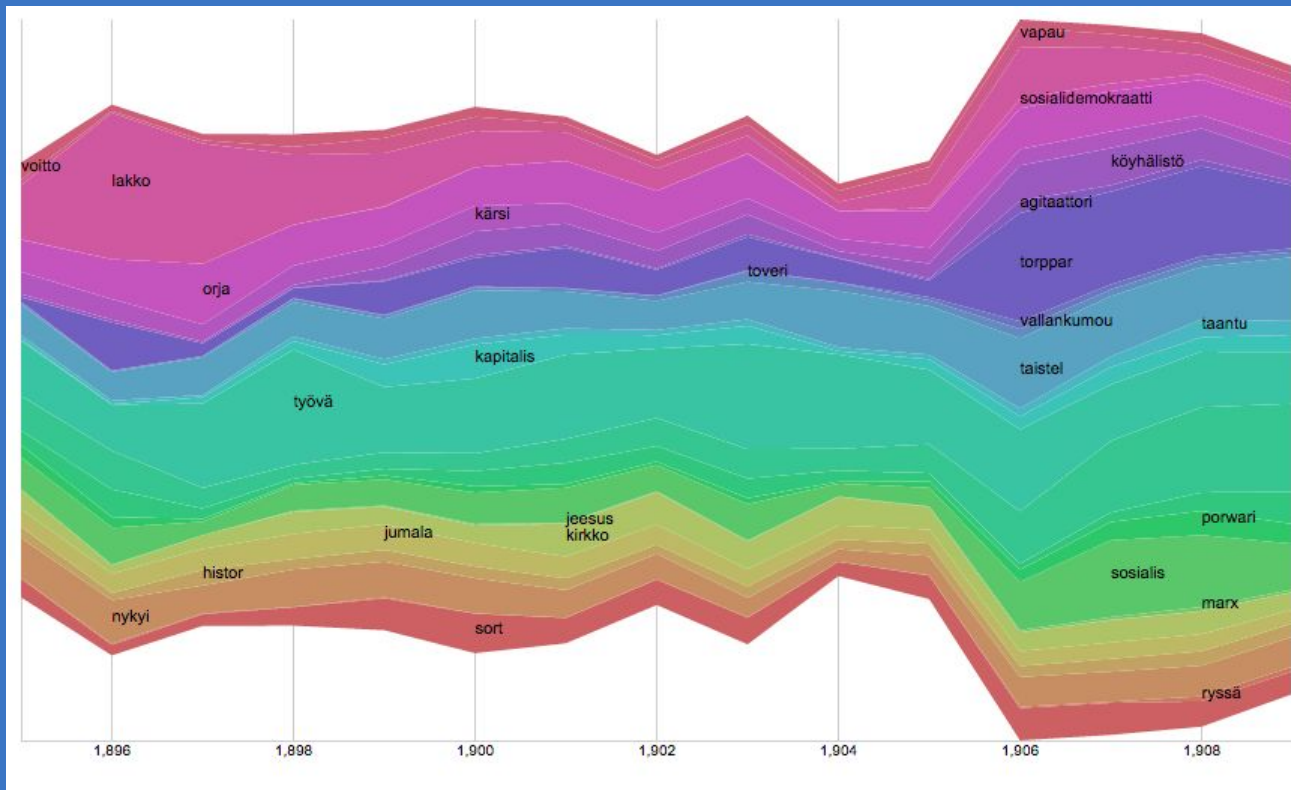


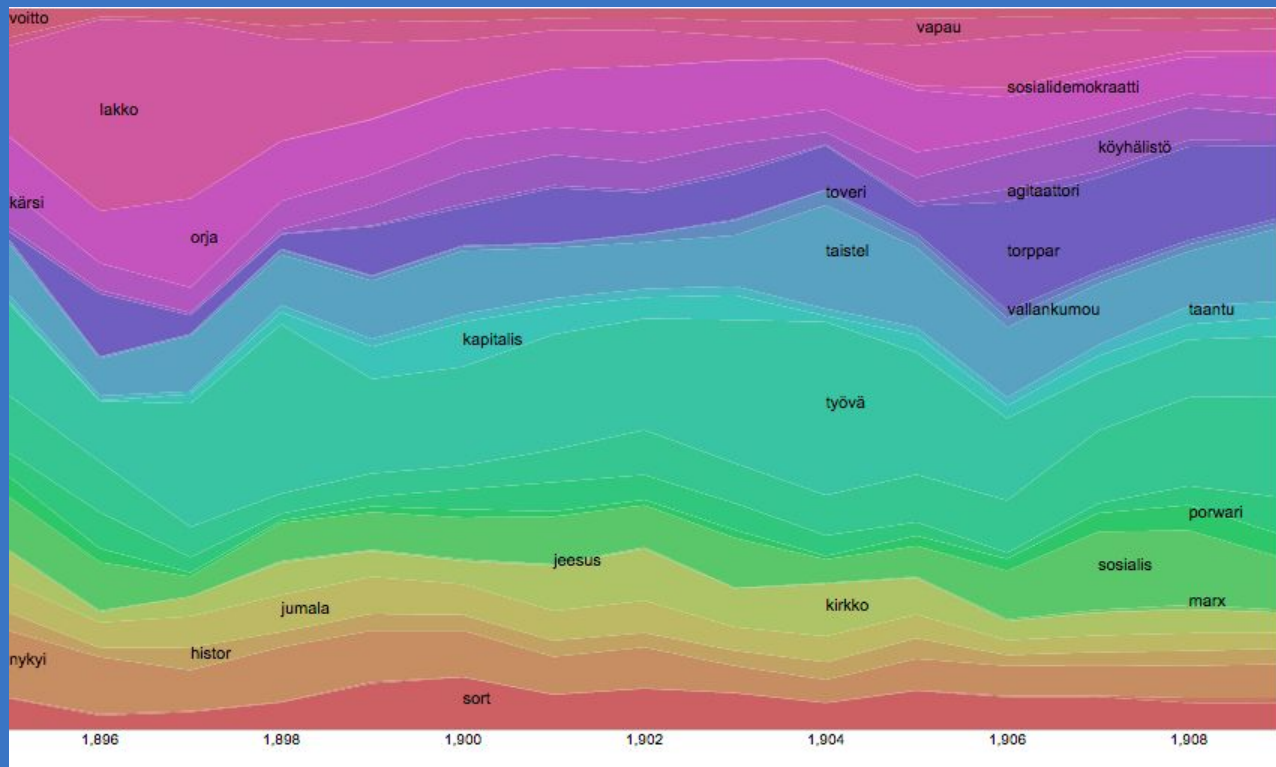
Alluvial diagram

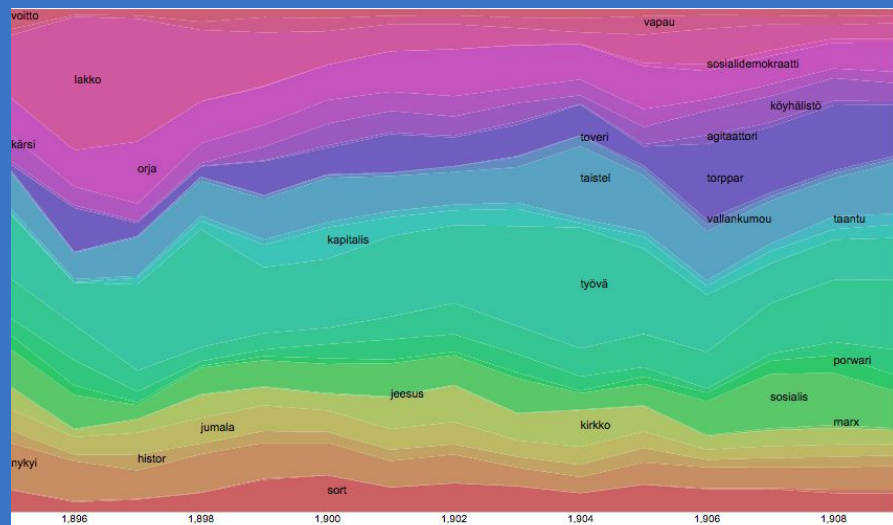
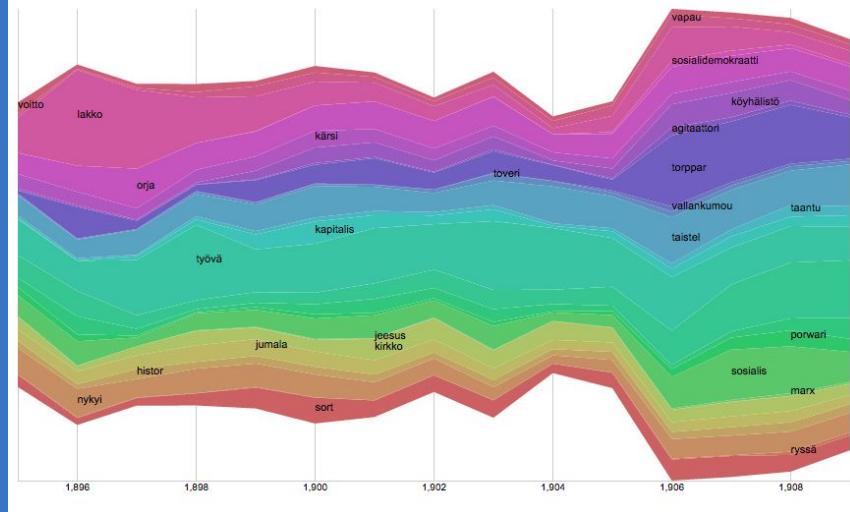




Stream-graph

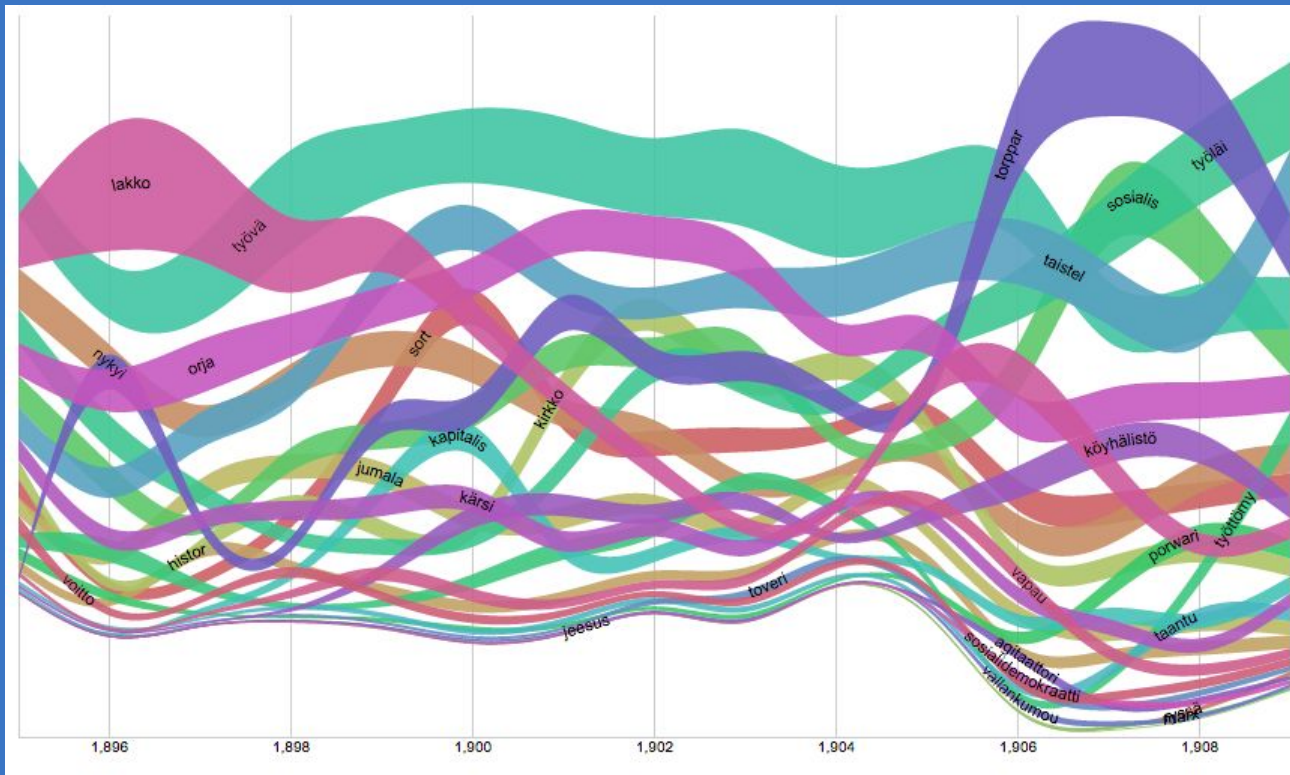


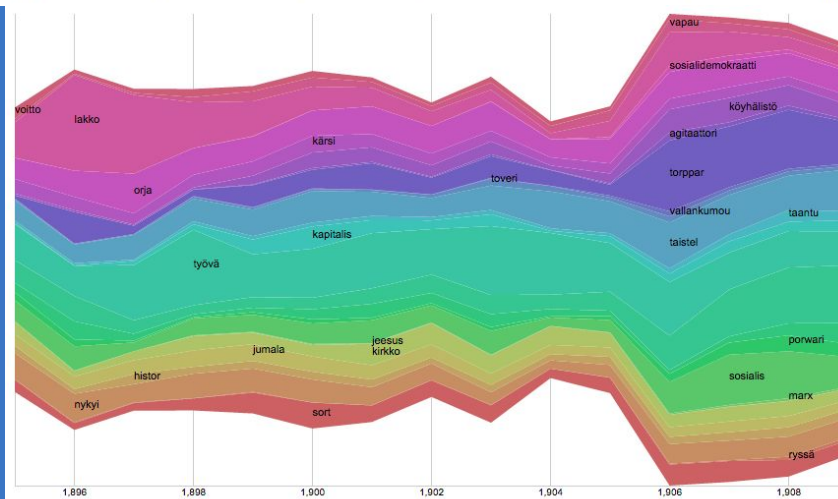
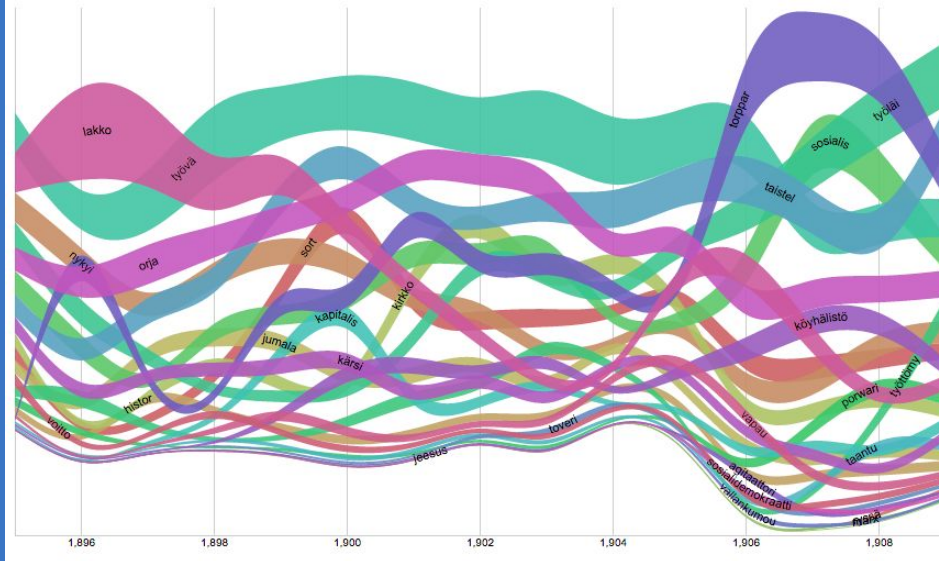


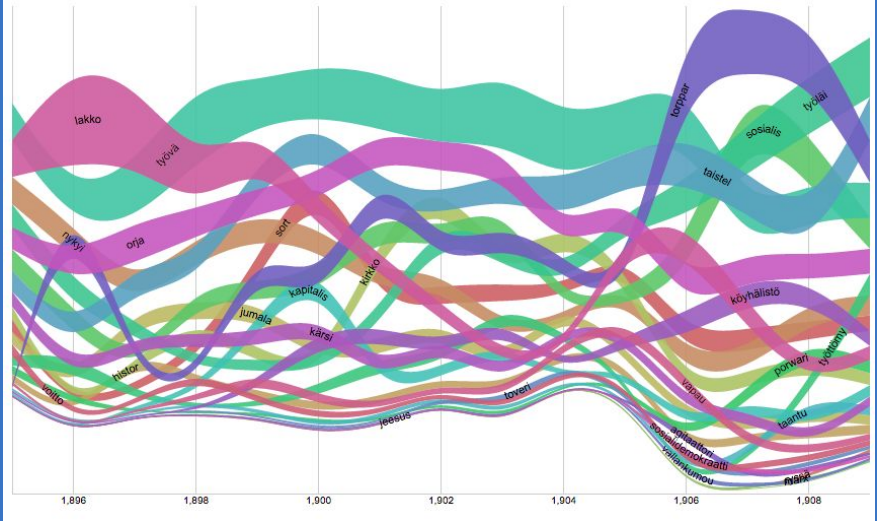




Bump chart

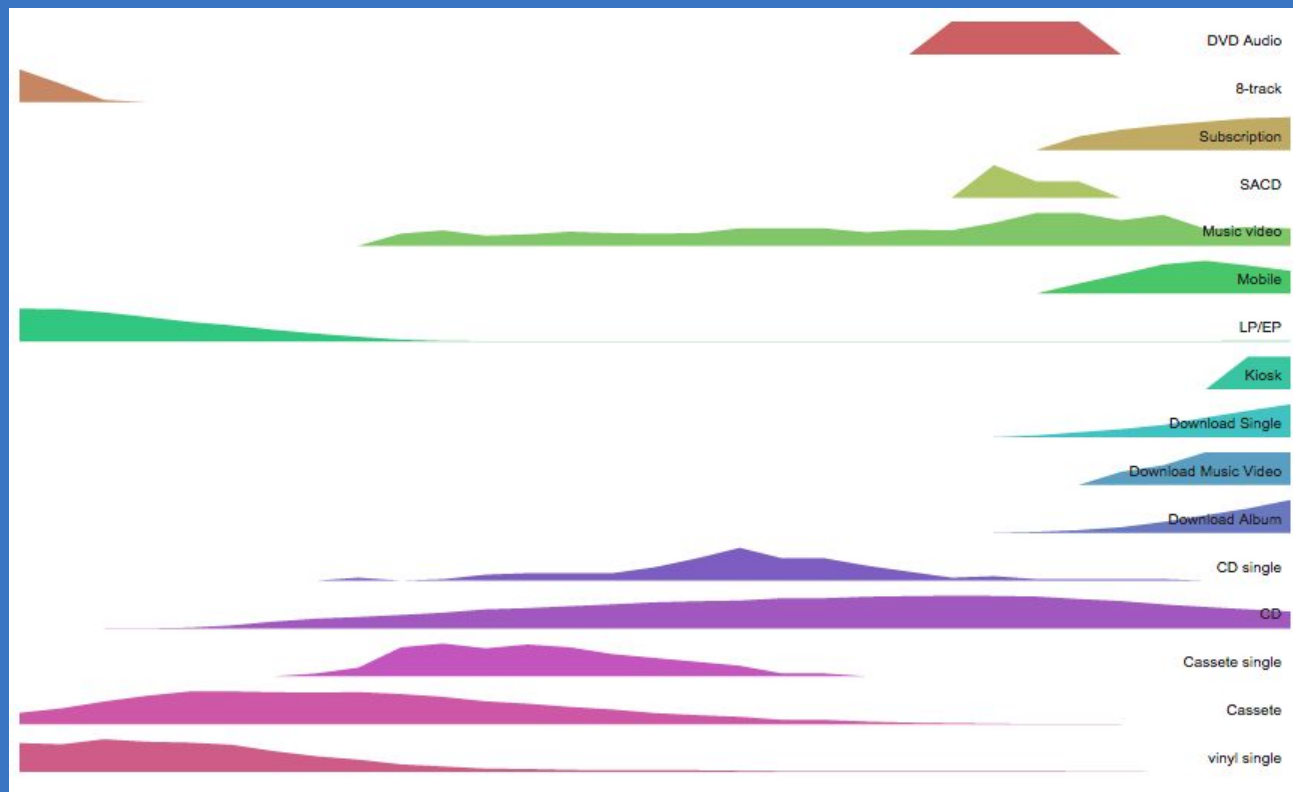






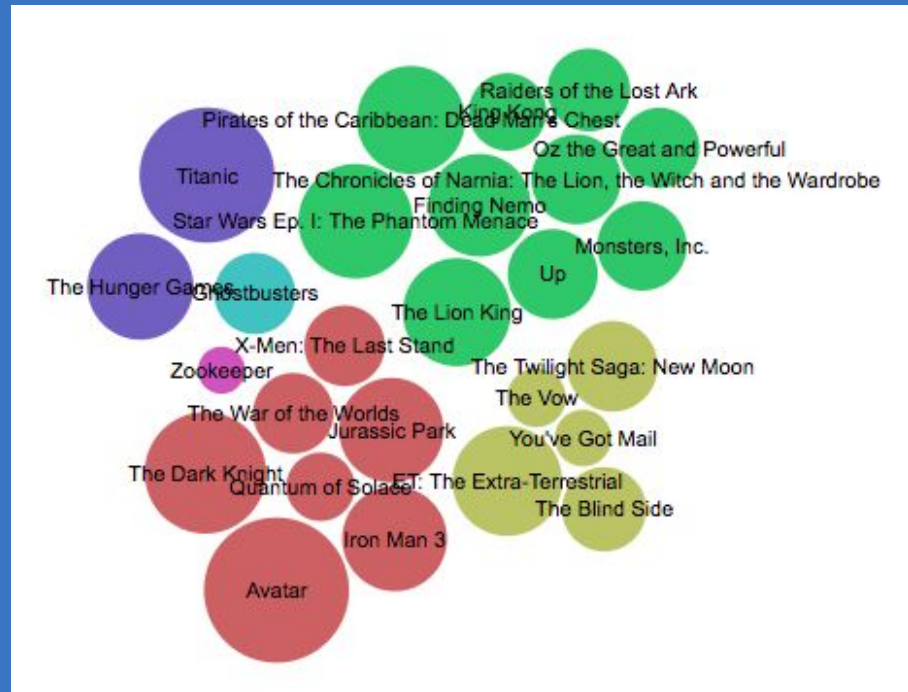


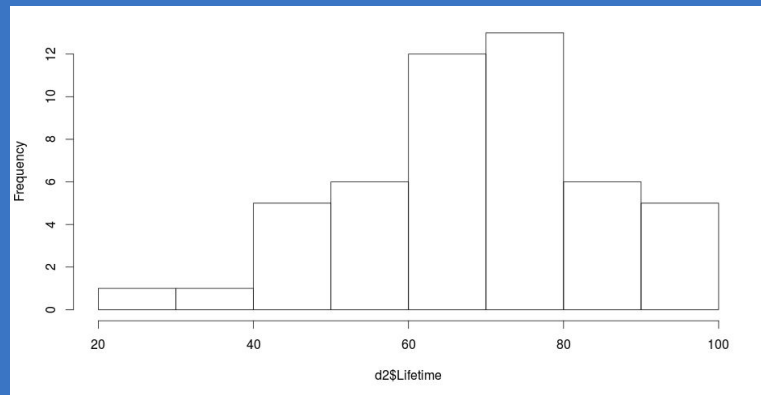
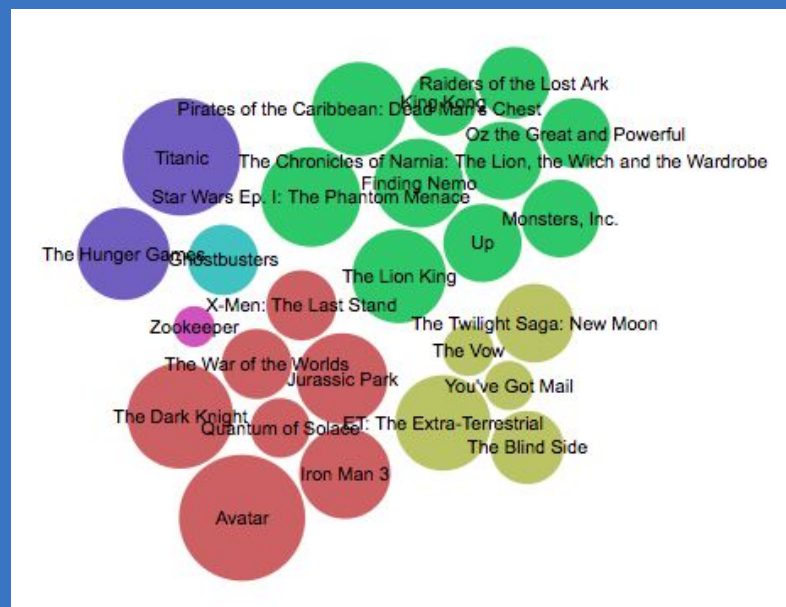
Area graph





Clustered Force Layout

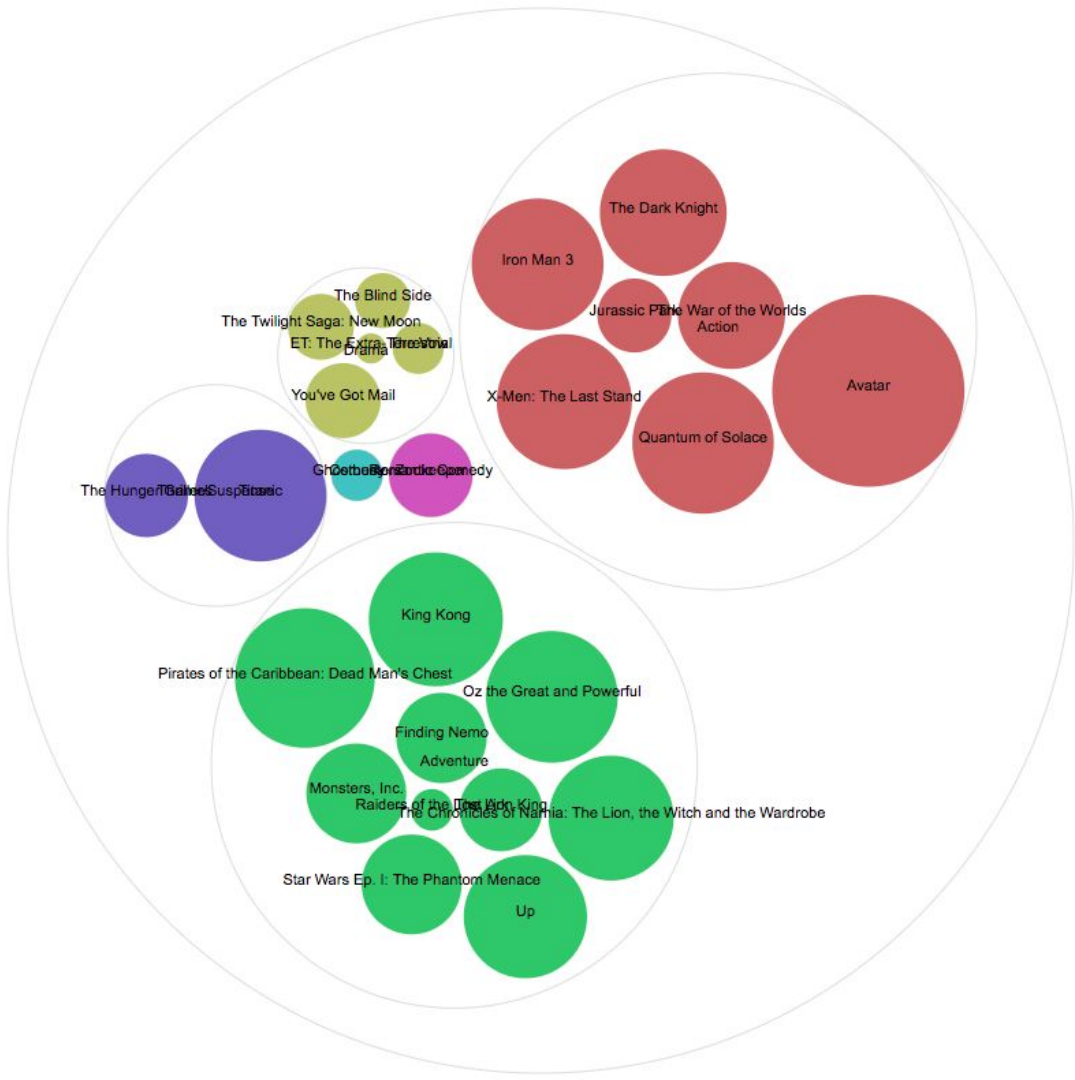






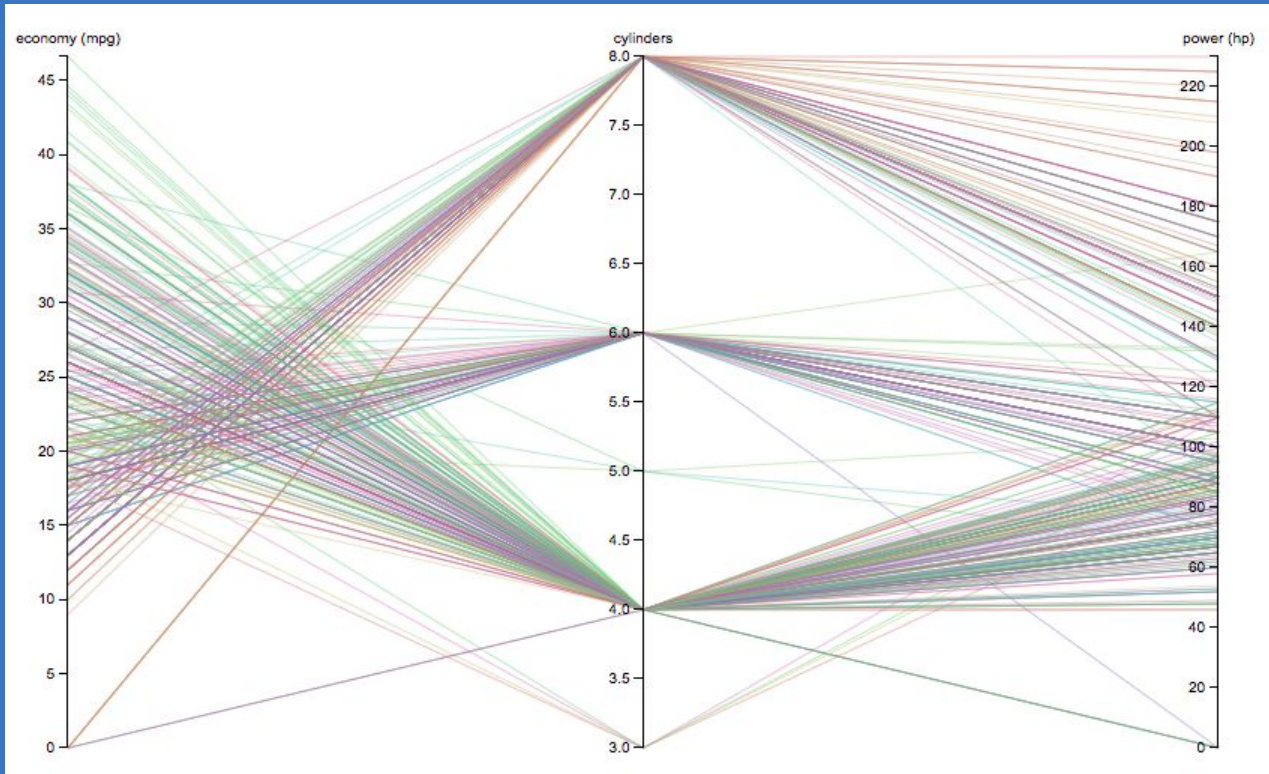
Circle Packing, Treemaps, Voronoi

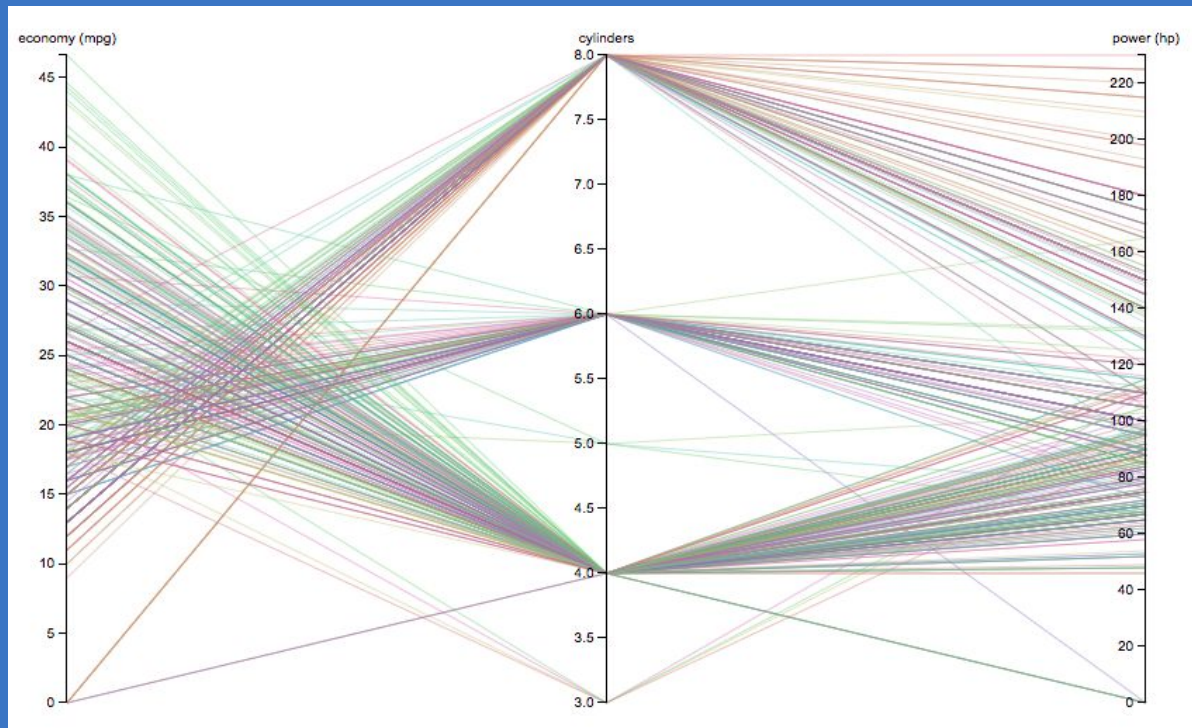
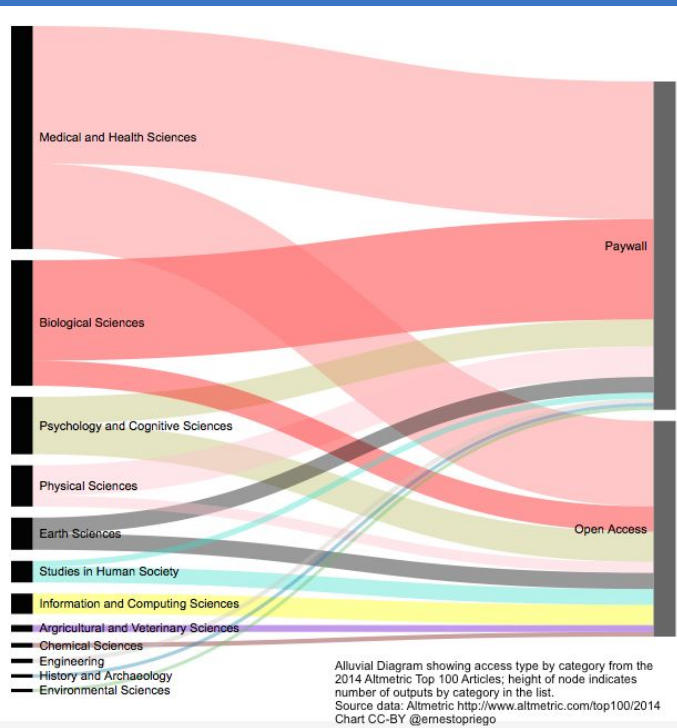
Deep hierarchy example, playground

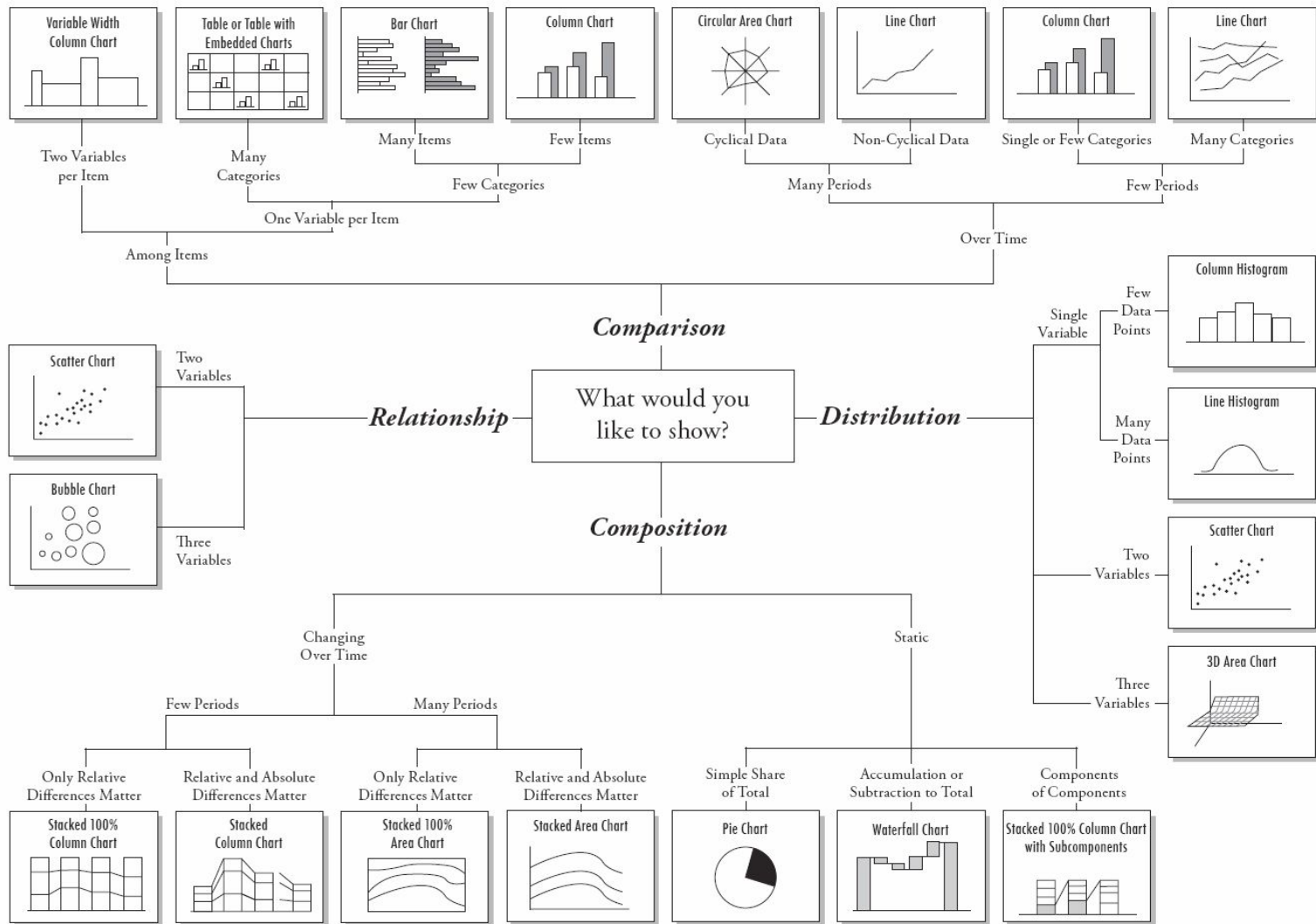




Parallel Coordinates









Voyager

- Interactive environment that suggests visualizations based on your data

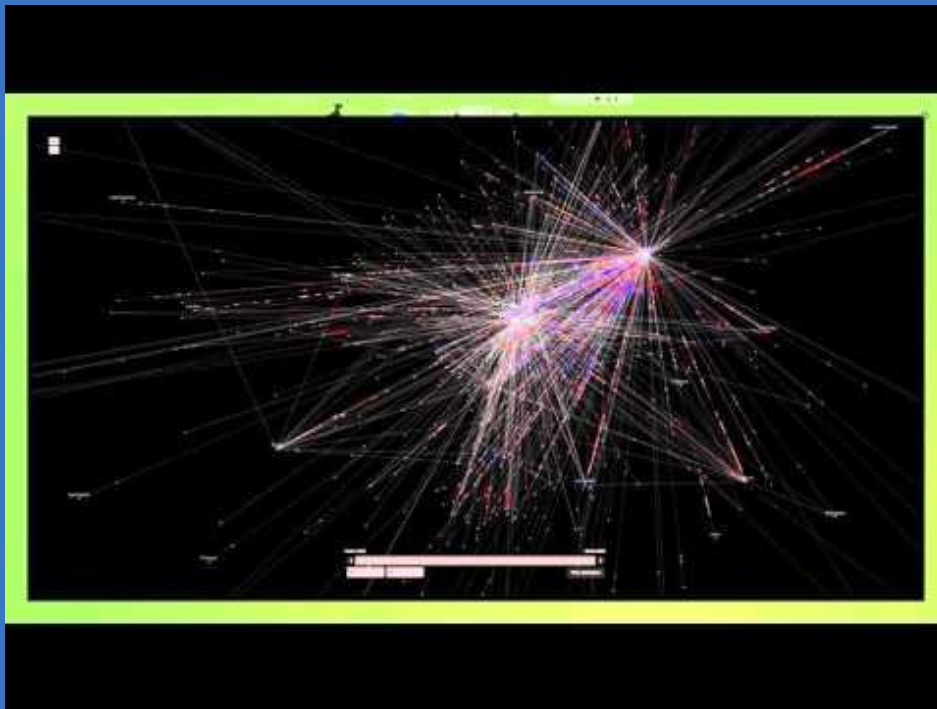


Tableau

<https://public.tableau.com/s/gallery/50-years-crime-us>



NodeGoat





Palladio



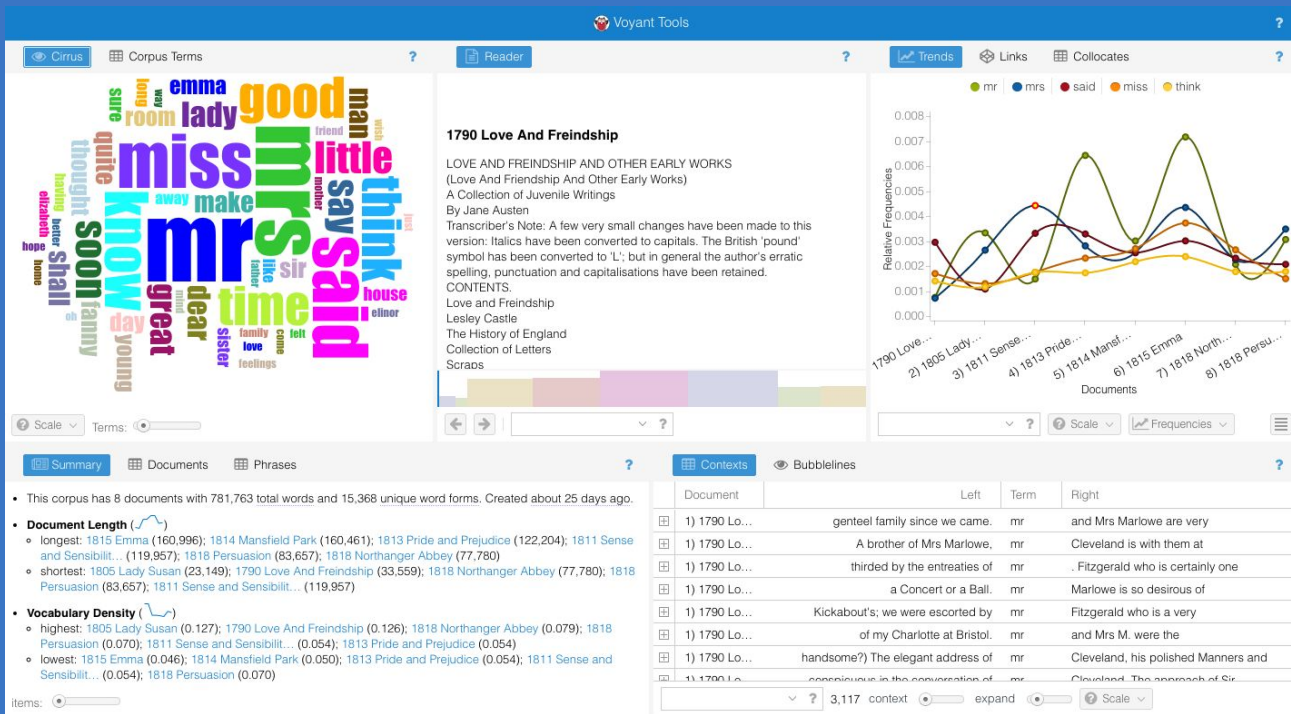


Carto

- <http://goodcitylife.org/smellymaps/>
- <http://brilliantmaps.com/london-smell/>
- https://jiemakel.carto.com/viz/2e88aa9a-167e-11e6-a016-0e3ff518bd15/public_map
- <http://lifewatch.inbo.be/blog/posts/forward-trajectory-visualizations.html>



Voyant tools





Some easy to use end-user data processing and visualization tools

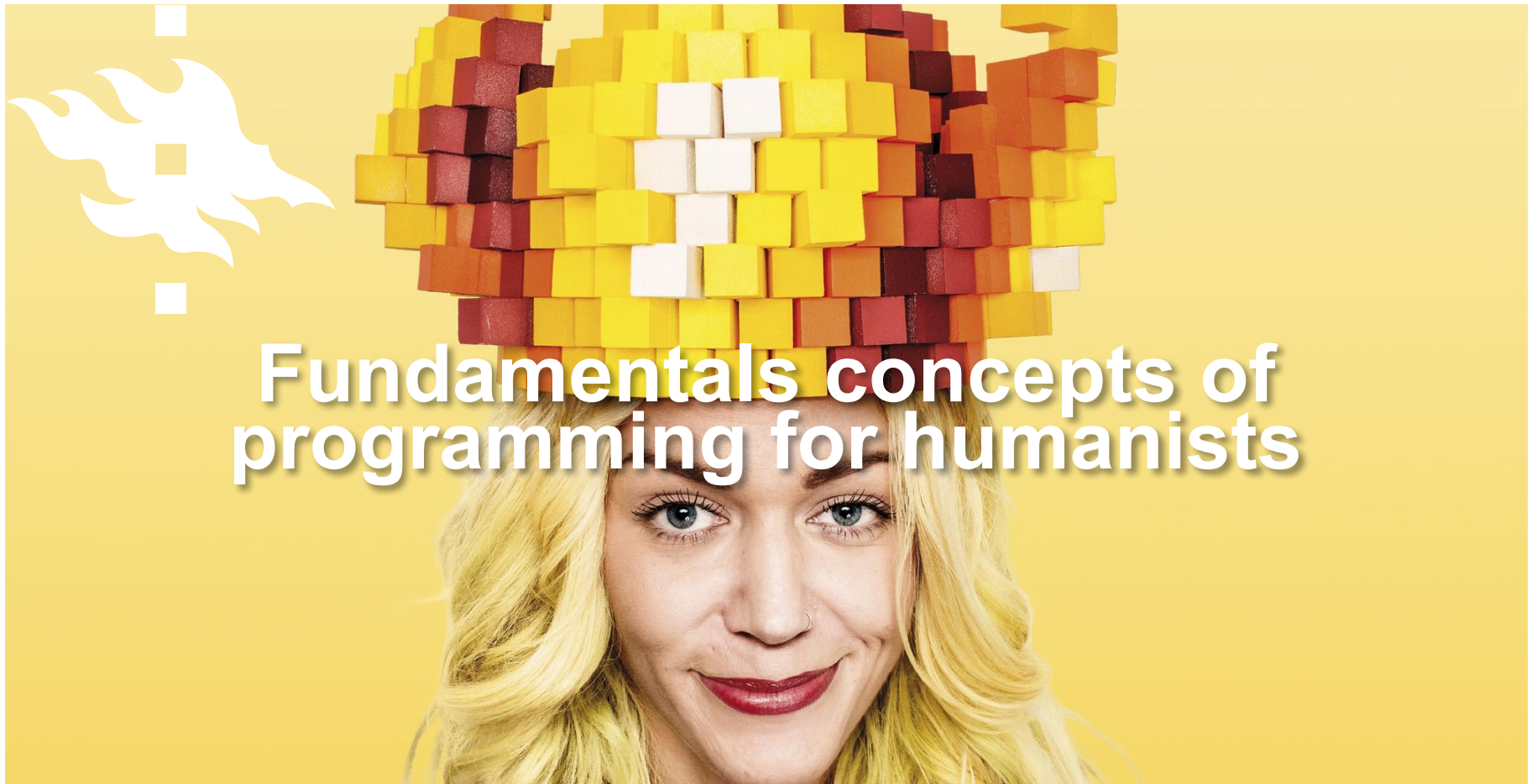
- OpenRefine - tutorial
- AntConc
- Palladio
- RAW
- NodeGoat
- Voyant Tools
- TAPoR
- Paper Machines

Hands-on tutorial of OpenRefine, RAW & Palladio



Homework

- Go through the OpenRefine tutorial at <http://freeyourmetadata.org/cleanup/>
- Experiment with at least one of the tools described in the slides. Post a message on #tools on Slack about your experience with the tool you chose.



Fundamentals concepts of programming for humanists



Knowledge of the fundamentals concepts of programming

- Frees you to process your data more efficiently
- Allows you to more freely apply visualizations etc based on ready libraries and tutorials on the Internet



Crash course into programming

- For everything you want to do, there is a library.
 - e.g. Pandas, Mallet, LDavis, Matplotlib, Requests, tm
- Nowadays, programming is mostly reading up how on to use these libraries from their documentation, and writing glue code to hook them together to perform some useful functionality
- This is mostly done through trial and error, and lots of googling



Literal programming ↔ programming notebooks

- Literal programming: documentation and program code in the same place
- Notebooks: allow partitioning a program to smaller blocks
 - Experimentation
 - Reiteration



Homework

- Complete the OpenRefine tutorial
- Experiment with at least one of the visualization tools described in the slides. Post a message on #tools on Slack about your experience with the tool you chose.
- Work through the fundamental concepts of programming for humanists and the assignments there.



Regular expressions



Regular expressions - a grammar for specifying rules to match text

Aaltio (född Hinderson), Ellen, född i Björneborg 25/B 1887, genomgått 6 klasser i Björneborgs, svenska samskola och en bokföringskurs 1899. -- Kassör vid Wasa-Aktie-Banks filialkontor i Björneborg från % 1904, i 4 års tid under vintermånaderna tjänstgjort vid Björneborgs Sparbank.

Adlercreutz, Herman, född i Kyrkslätt 28/io 1864, stu-. dent 2% 1884, allmän rätts-ex. 81/6 1889, auskultant i Åbo h ofrätt 6/6 s. å. - Kanslist vid Finlands Banks hufvud-kontor sedan 1893. - Vicehäradshöfding 2%2 1892, förste stadsfogde i Helsingfors sedan 1899. -- Deltagit i ridder-skåpet och adelns förhandlingar vid landtdagarne 1897, .1899, 1900, 1904--05, (suppleant i lagutskottet) och-1905--06 (suppleant i fusteringsutskottet).

Aejmelaeus, Otto, född i Paldamo 9/x 1850, genomgått 6 klasser i elementarläroverket i Uleåborg. - Direktör för Nordiska Aktiebankens "för Handel och Industri filialkontor i Jyväskylä från 1896. - Konditionerat hos C. E. Carlström i Kristinestad, Aug. Eklöf i Borgå och i Paul Wahl & C:os trävaruaffär i Jyväskylä. -- Varit ordförande i fattigvårdsnämnden och i drätselkammaren i Jyväskylä.



Regular expressions - a grammar for specifying rules to match text

Aaltio (född Hinderson), **Ellen**, född i **Björneborg** 25/B 1887, genomgått 6 klasser i Björneborgs, svenska samskola och en bokföringskurs 1899. -- Kassör vid Wasa-Aktie-Banks filialkontor i Björneborg från % 1904, i 4 års tid under vintermånaderna tjänstgjort vid Björneborgs Sparbank.

Adlercreutz, **Herman**, född i **Kyrkslätt** 28/io 1864, stu-. dent 2% 1884, allmän rätts-ex. 81/6 1889, auskultant i Åbo h ofrätt 6/6 s. å. - Kanslist vid Finlands Banks hufvud-kontor sedan 1893. - Vicehäradshöfding 2%2 1892, förste stadsfogde i Helsingfors sedan 1899. -- Deltagit i ridder-skåpet och adelns förhandlingar vid landtdagarne 1897, .1899, 1900, 1904--05, (suppleant i lagutskottet) och-1905--06 (suppleant i fusteringsutskottet).

Aejmelaeus, **Otto**, född i **Paldamo** 9/x 1850, genomgått 6 klasser i elementarläroverket i Uleåborg. - Direktör för Nordiska Aktiebankens "för Handel och Industri filialkontor i Jyväskylä från 1896. - Konditionerat hos C. E. Carlström i Kristinestad, Aug. Eklöf i Borgå och i Paul Wahl & C:os trävaruaffär i Jyväskylä. -- Varit ordförande i fattigvårdsnämnden och i drätselkammaren i Jyväskylä.



Aaltio (född Hinderson), Ellen, född i Björneborg 25/B 1887, genomgått 6 klasser i Björneborgs, svenska samskola och en bokföringskurs 1899. -- Kassör vid Wasa-Aktie-Banks filialkontor i Björneborg från % 1904, i 4 års tid under vintermånaderna tjänstgjort vid Björneborgs Sparbank.

- Year: four numbers anywhere
- Last name: letters before a space at the start of the line
- First name: letters following the first , on a line
- Birthplace: letters following “född i”



Aaltio (född Hinderson), Ellen, född i Björneborg 25/B 1887, genomgått 6 klasser i Björneborgs, svenska samskola och en bokföringskurs 1899. -- Kassör vid Wasa-Aktie-Banks filialkontor i Björneborg från % 1904, i 4 års tid under vintermånaderna tjänstgjort vid Björneborgs Sparbank.

- Year: four numbers anywhere
 - [0-9][0-9][0-9][0-9]
 - [0-9]{4}
 - \d{4}
 - 1[89]\d{2}
- Last name: letters before a space at the start of the line
 - ^[A-Z][a-z]+
 - ^[^\s]+
 - ^\w+



Homework

- Complete the OpenRefine tutorial
- Experiment with at least one of the visualization tools described in the slides. Post a message on #tools on Slack about your experience with the tool you chose.
- Work through the fundamental concepts of programming for humanists and the assignments there.
- Work through the regular expression tutorial and assignment



Homework

- Complete the [OpenRefine tutorial](#)
- Experiment with at least one of the visualization tools described in the slides. Post a message on #tools on Slack about your experience with the tool you chose.
- Work through the [fundamental concepts of programming for humanists](#) and the assignments there.
- Work through the [regular expression tutorial and assignment](#)
- In preparation for the lecture on 21.11., read [this research article](#).



eetu.makela@helsinki.fi
<http://iki.fi/eetu.makela>