

Data Science & Visualization

Introduction to Data Visualization

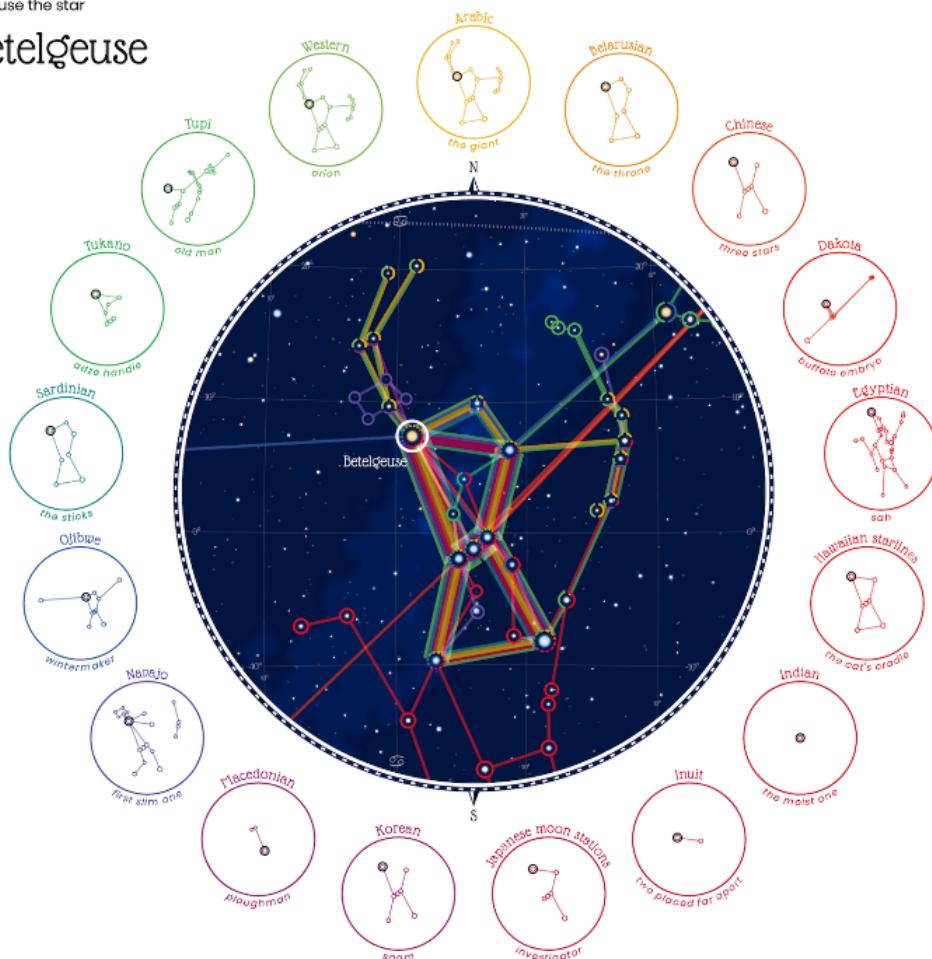
Gabriela Molina León
molina@uni-bremen.de

Institute for Information Management Bremen
Information Management Group (AGIM)



The cultures & constellations
that use the star

Betelgeuse



Source: [Nadieh Bremer](#)

Sessions & Deliverables

Date	Lecture (10:00-11:30)	Practical (11:45-13:15)
08.04.24	Introduction to Data Science	Python Introduction
15.04.24	Basic Statistics & Supervised Learning	Practical Statistics + Supervised
22.04.24	Unsupervised Learning	Practical Unsupervised
29.04.24	Introduction to Data Visualization	Guest Lecture on NLP by Oxana Vitman
06.05.24	Exploratory Data Analysis	Data Science & Vis Presentation

19.07.24 Deadline Final Report

Note that for the slots **marked red**, you are expected to prepare presentations.
For the slots **marked blue**, you are expected to bring a computer.
In the slots **marked purple**, you are not expected to come to the classroom .

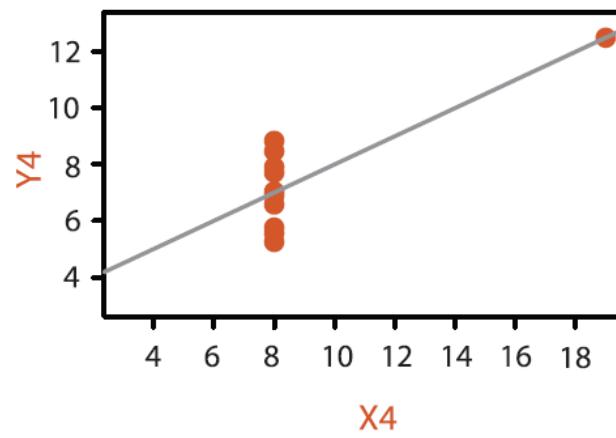
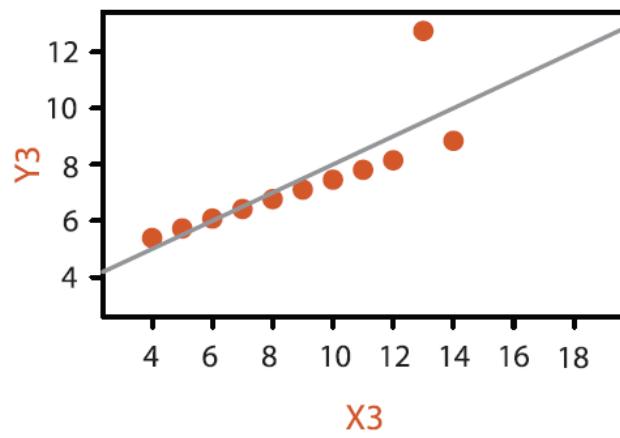
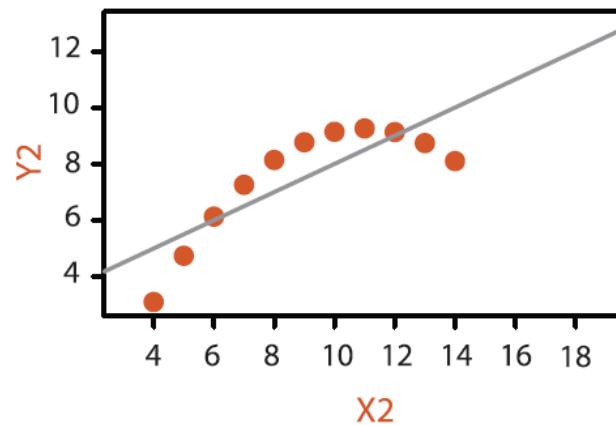
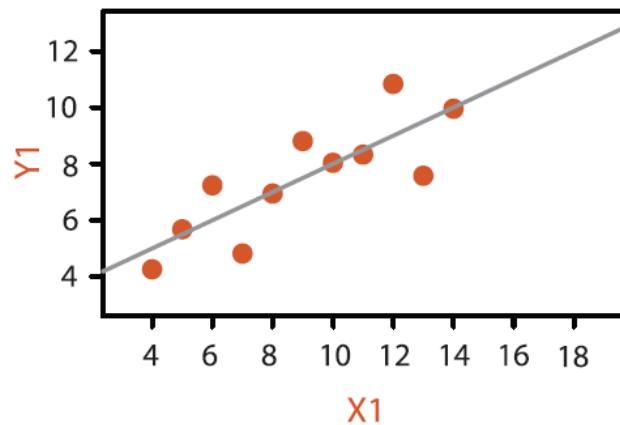


Why Visualization?

Anscombe's Quartet: Raw Data

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

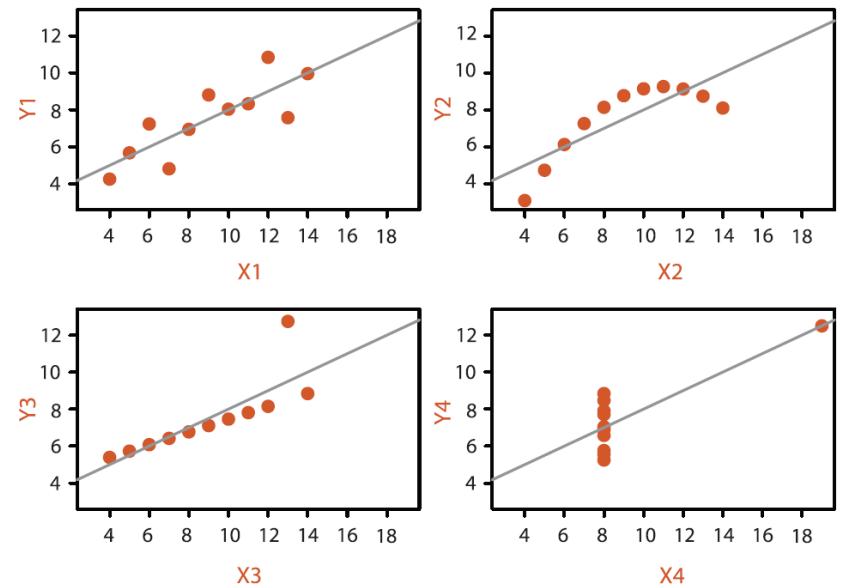
Why Visualization?



Why Visualization?

Anscombe's Quartet: Raw Data

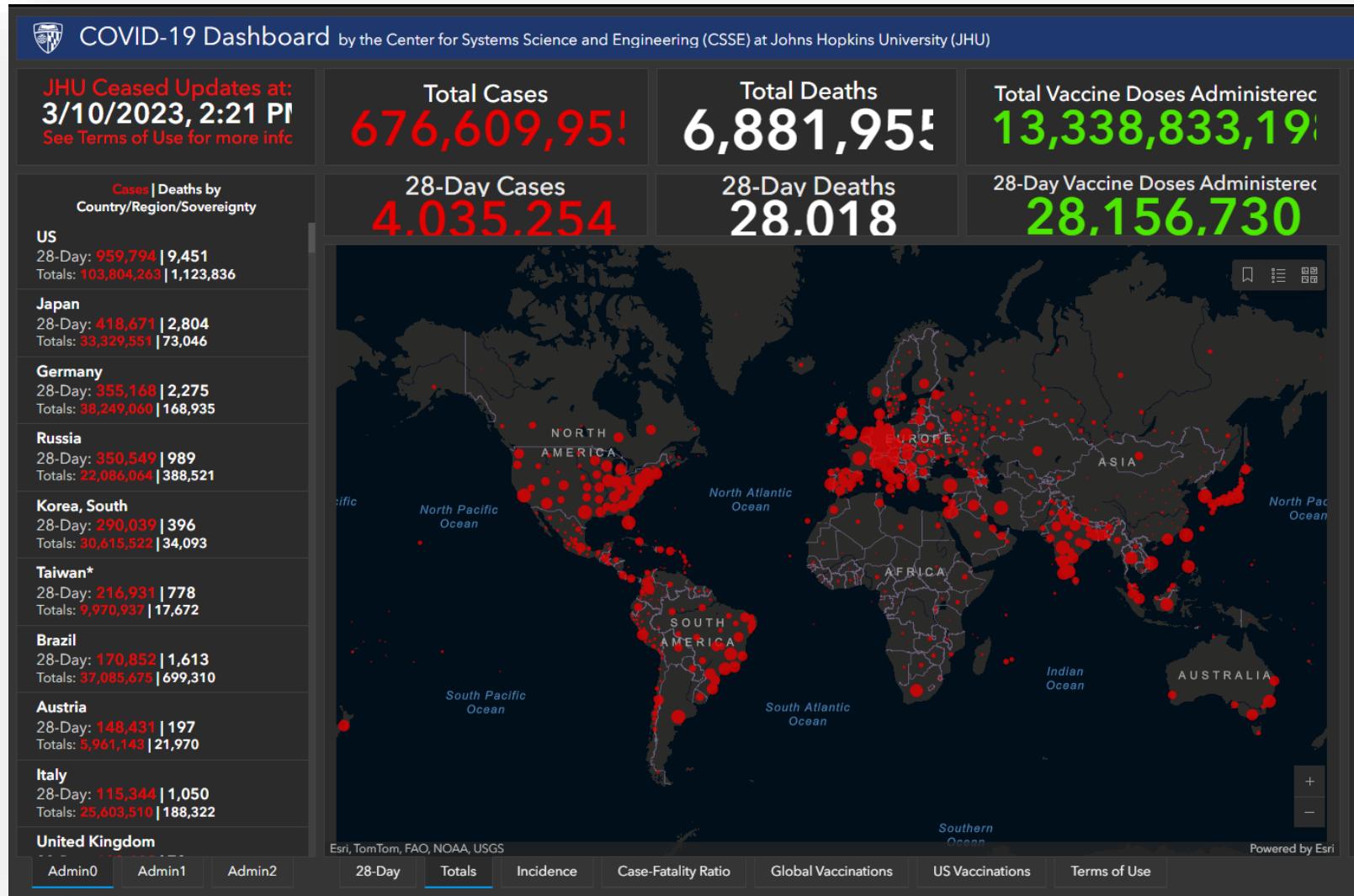
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



Our visual capabilities can reveal patterns very quickly

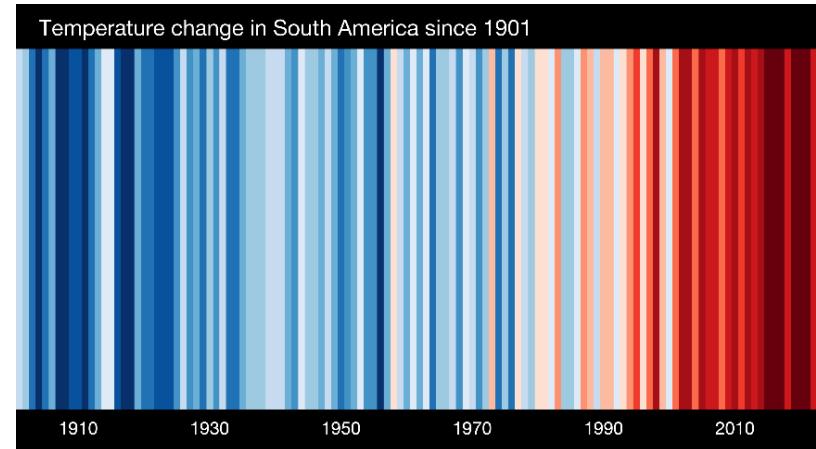
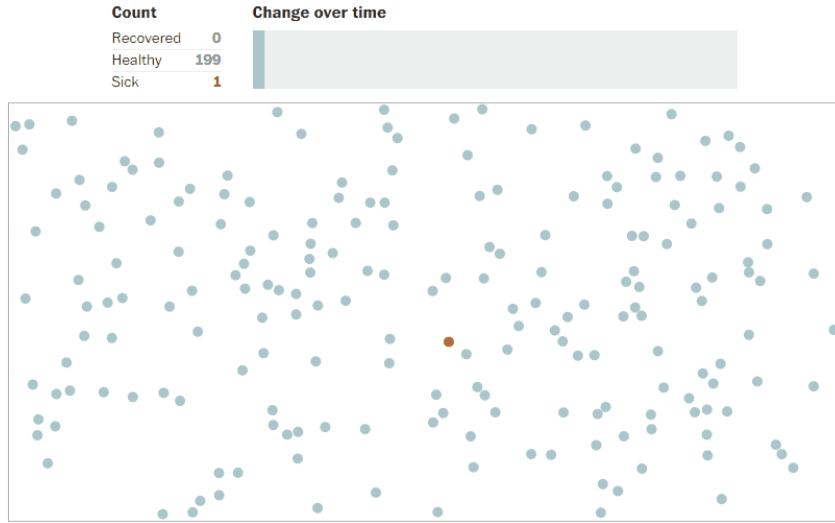


Powerful tool to help us understand





Powerful tool to help us understand



Sources: [Johns Hopkins University](#), [The Washington Post](#), [Ed Hawkins](#)

Gestalt Principles

Law of Proximity

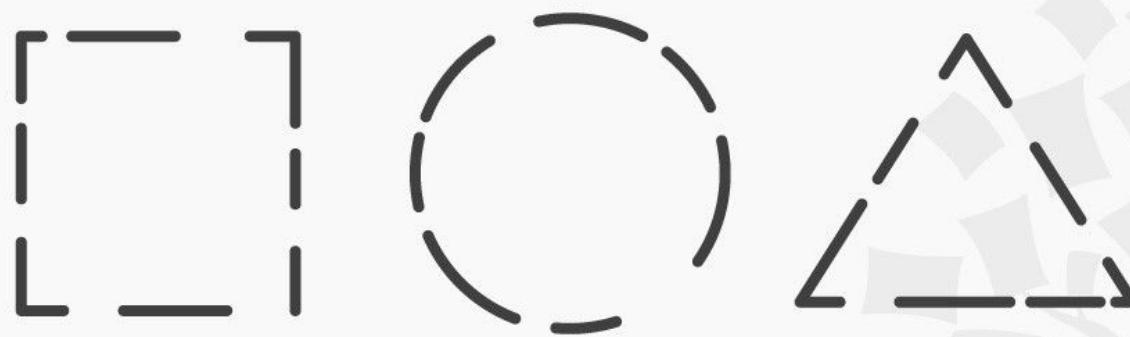


INTERACTION DESIGN
FOUNDATION

INTERACTION-DESIGN.ORG

Gestalt Principles

Law of Closure



INTERACTION DESIGN
FOUNDATION

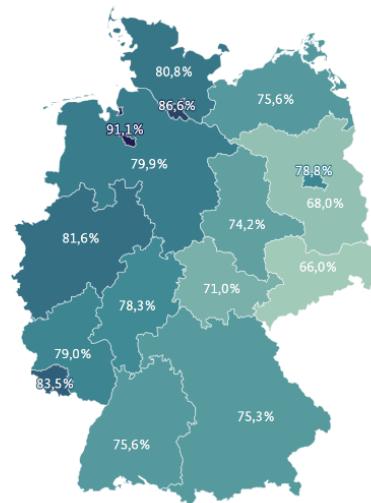
INTERACTION-DESIGN.ORG

[More on Gestalt Principles](#)

What is Data Visualization?

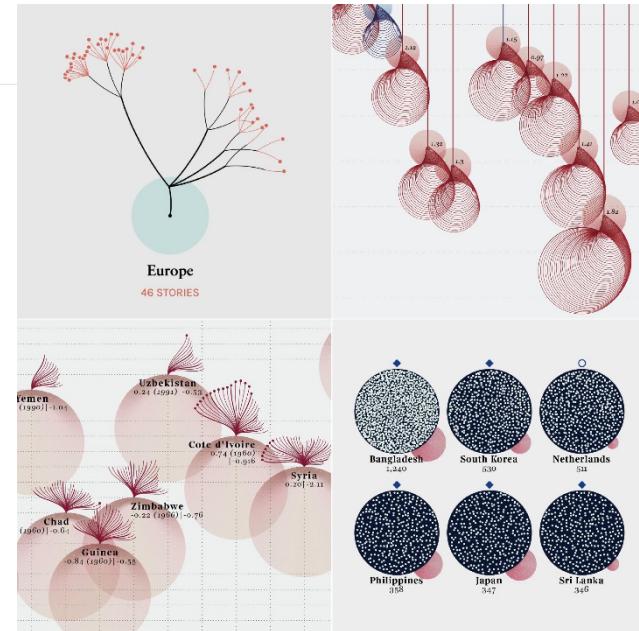
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”
(Munzner, 2014)

Impfdashboard.de Stand: 04. Juli 2022, 10:13



Zusammen gegen Corona

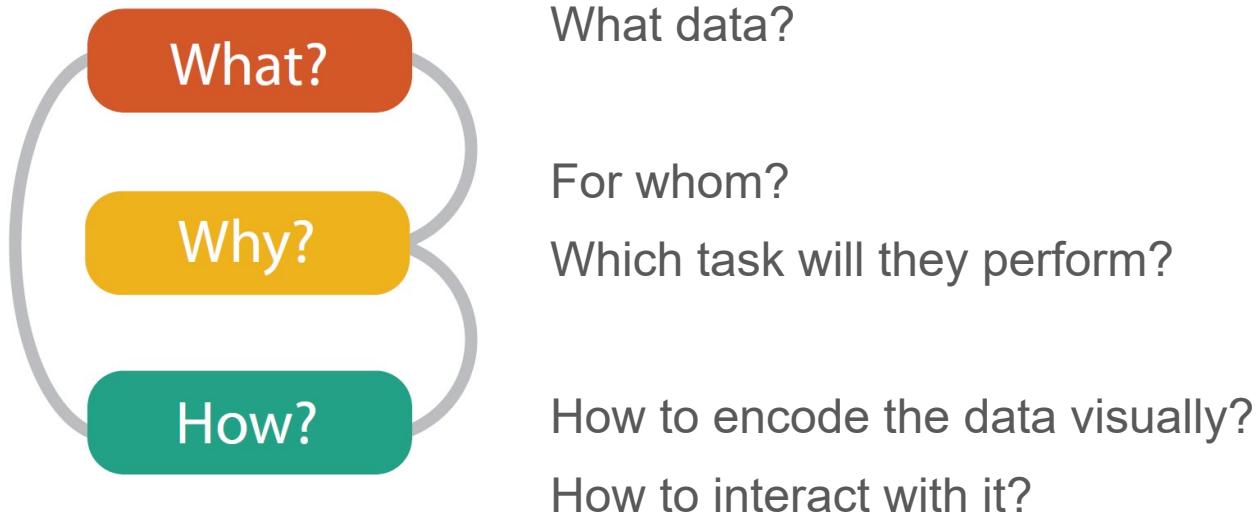
Bremen	91,1%
Hamburg	86,6%
Saarland	83,5%
Nordrhein-Westfalen	81,6%
Schleswig-Holstein	80,8%
Niedersachsen	79,9%
Rheinland-Pfalz	79,0%
Berlin	78,8%
Hessen	78,3%
Baden-Württemberg	75,6%
Mecklenburg-Vorpommern	75,6%
Bayern	75,3%
Sachsen-Anhalt	74,2%
Thüringen	71,0%
Brandenburg	68,0%
Sachsen	66,0%



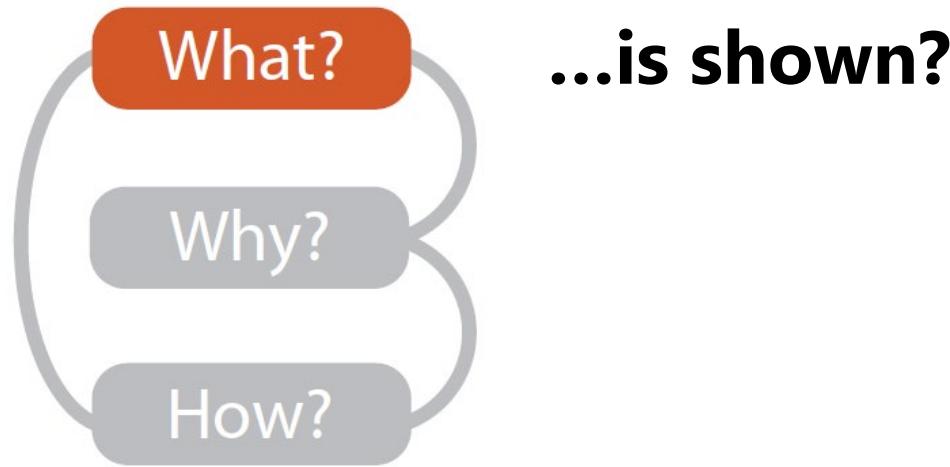
[Moritz Stefaner](#)

[Federica Fragapane](#)

Munzner's framework



Data



Data

Numbers? Text?

Coordinates? Categories?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	country	countryname	oecdmember	eumember	edate	date	party	partyname	partyabbrev	parfam	coderid	manual	coderyear	testresult	testeditsim	pervote	voteest	presvote	abseat	totseats	progtpe	datasetorig	corpusversict
2	11	Sweden	0	0	17/09/1944	194409	11220	Communist	ISKP	20	117	0	1983			10.3	0	15	230	1	41		
3	11	Sweden	0	0	17/09/1944	194409	11320	Social Demo	SAP	30	117	0	1983			46.5	0	115	230	1	41		
4	11	Sweden	0	0	17/09/1944	194409	11420	Peopleâ€™s FP		40	117	0	1983			12.9	0	26	230	1	41		
5	11	Sweden	0	0	17/09/1944	194409	11620	Right Party		60	117	0	1983			15.8	0	39	230	1	41		
6	11	Sweden	0	0	17/09/1944	194409	11810	Agrarian Party		80	117	0	1983			13.6	0	35	230	1	41		
7	11	Sweden	0	0	19/09/1948	194809	11220	Communist	ISKP	20	117	0	1983			6.312	0	8	230	1	10		
8	11	Sweden	0	0	19/09/1948	194809	11320	Social Demo	SAP	30	117	0	1983			46.132	0	112	230	1	10		
9	11	Sweden	0	0	19/09/1948	194809	11420	Peopleâ€™s FP		40	117	0	1983			22.749	0	57	230	1	10		
10	11	Sweden	0	0	19/09/1948	194809	11620	Right Party		60	117	0	1983			12.343	0	23	230	1	10		
11	11	Sweden	0	0	19/09/1948	194809	11810	Agrarian Party		80	117	0	1983			12.385	0	30	230	1	10		
12	11	Sweden	0	0	21/09/1952	195209	11220	Communist	ISKP	20	117	0	1983			4.34	0	5	230	1	10		
13	11	Sweden	0	0	21/09/1952	195209	11320	Social Demo	SAP	30	117	0	1983			46.047	0	110	230	1	10		
14	11	Sweden	0	0	21/09/1952	195209	11420	Peopleâ€™s FP		40	117	0	1983			24.442	0	58	230	1	10		
15	11	Sweden	0	0	21/09/1952	195209	11620	Right Party		60	117	0	1983			14.373	0	31	230	1	10		
16	11	Sweden	0	0	21/09/1952	195209	11810	Agrarian Party		80	117	0	1983			10.735	0	26	230	1	10		
17	11	Sweden	0	0	26/09/1956	195609	11220	Communist	ISKP	20	117	0	1983			5.001	0	6	231	1	10		
18	11	Sweden	0	0	26/09/1956	195609	11320	Social Demo	SAP	30	117	0	1983			44.581	0	106	231	1	10		
19	11	Sweden	0	0	26/09/1956	195609	11420	Peopleâ€™s FP		40	117	0	1983			23.807	0	58	231	1	10		
20	11	Sweden	0	0	26/09/1956	195609	11620	Right Party		60	117	0	1983			17.108	0	42	231	1	10		
21	11	Sweden	0	0	26/09/1956	195609	11810	Agrarian Party		80	117	0	1983			9.45	0	19	231	1	10		
22	11	Sweden	0	0	1/6/1958	195806	11220	Communist	ISKP	20	117	0	1983			3.364	0	5	231	1	10		
23	11	Sweden	0	0	1/6/1958	195806	11320	Social Demo	SAP	30	117	0	1983			46.216	0	111	231	1	10		
24	11	Sweden	0	0	1/6/1958	195806	11420	Peopleâ€™s FP		40	117	0	1983			18.209	0	38	231	1	10		
25	11	Sweden	0	0	1/6/1958	195806	11620	Right Party		60	117	0	1983			19.518	0	45	231	1	10		
26	11	Sweden	0	0	1/6/1958	195806	11810	Centre Party	CP	80	117	0	1983			12.662	0	32	231	1	10		
27	11	Sweden	0	0	18/09/1960	196009	11220	Communist	ISKP	20	117	0	1983			4.479	0	5	232	1	10		
28	11	Sweden	0	0	18/09/1960	196009	11320	Social Demo	SAP	30	117	0	1983			47.789	0	114	232	1	10		
29	11	Sweden	0	0	18/09/1960	196009	11420	Peopleâ€™s FP		40	117	0	1983			17.492	0	40	232	1	10		
30	11	Sweden	0	0	18/09/1960	196009	11620	Right Party		60	117	0	1983			16.557	0	39	232	1	10		
31	11	Sweden	0	0	18/09/1960	196009	11810	Centre Party	CP	80	117	0	1983			13.611	0	34	232	1	10		
32	11	Sweden	10	0	20/09/1964	196409	11220	Communist	ISKP	20	117	0	1983			5.223	0	8	233	1	10		
33	11	Sweden	10	0	20/09/1964	196409	11320	Social Demo	SAP	30	117	0	1983			47.269	0	113	233	1	10		
34	11	Sweden	10	0	20/09/1964	196409	11420	Peopleâ€™s FP		40	117	0	1983			16.975	0	42	233	1	10		
35	11	Sweden	10	0	20/09/1964	196409	11620	Right Party		60	117	0	1983			13.722	0	32	233	1	10		
36	11	Sweden	10	0	20/09/1964	196409	11810	Centre Party	CP	80	117	0	1983			13.181	0	33	233	1	10		
37	11	Sweden	10	0	15/09/1968	196809	11220	Left Commu	VPK	20	117	0	1983			3.006	0	3	233	1	10		
38	11	Sweden	10	0	15/09/1968	196809	11320	Social Demo	SAP	30	117	0	1983			50.116	0	125	233	1	10		
39	11	Sweden	10	0	15/09/1968	196809	11420	Peopleâ€™s FP		40	117	0	1983			14.256	0	32	233	1	10		
40	11	Sweden	10	0	15/09/1968	196809	11620	Right Party		60	117	0	1983			12.859	0	29	233	1	10		

Data types

→ Categorical



Apples, Oranges, Kiwis

Liberal, Conservative

→ Ordered

→ *Ordinal*



S, M, L

3.0, 2.3, 1.0

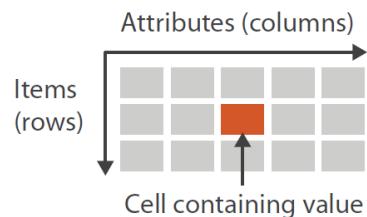
→ *Quantitative*



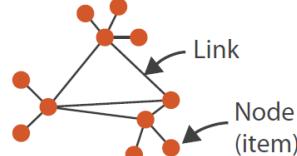
0, 0.5, 1...

Dataset types

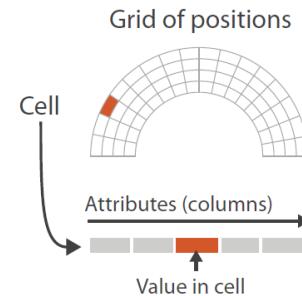
→ Tables



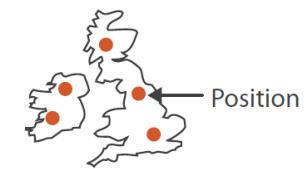
→ Networks



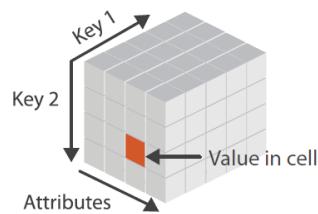
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table

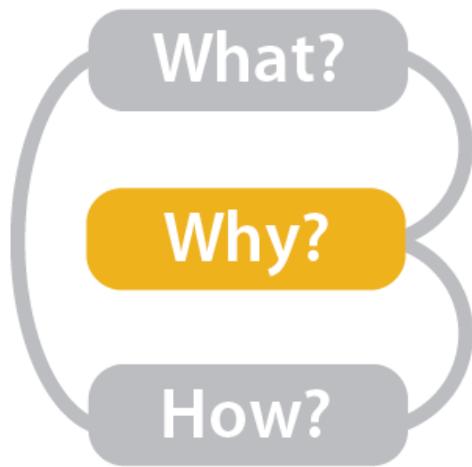


→ Trees



[Figures from VAD book](#)

Task abstraction



... is the user looking at it?

Task abstraction

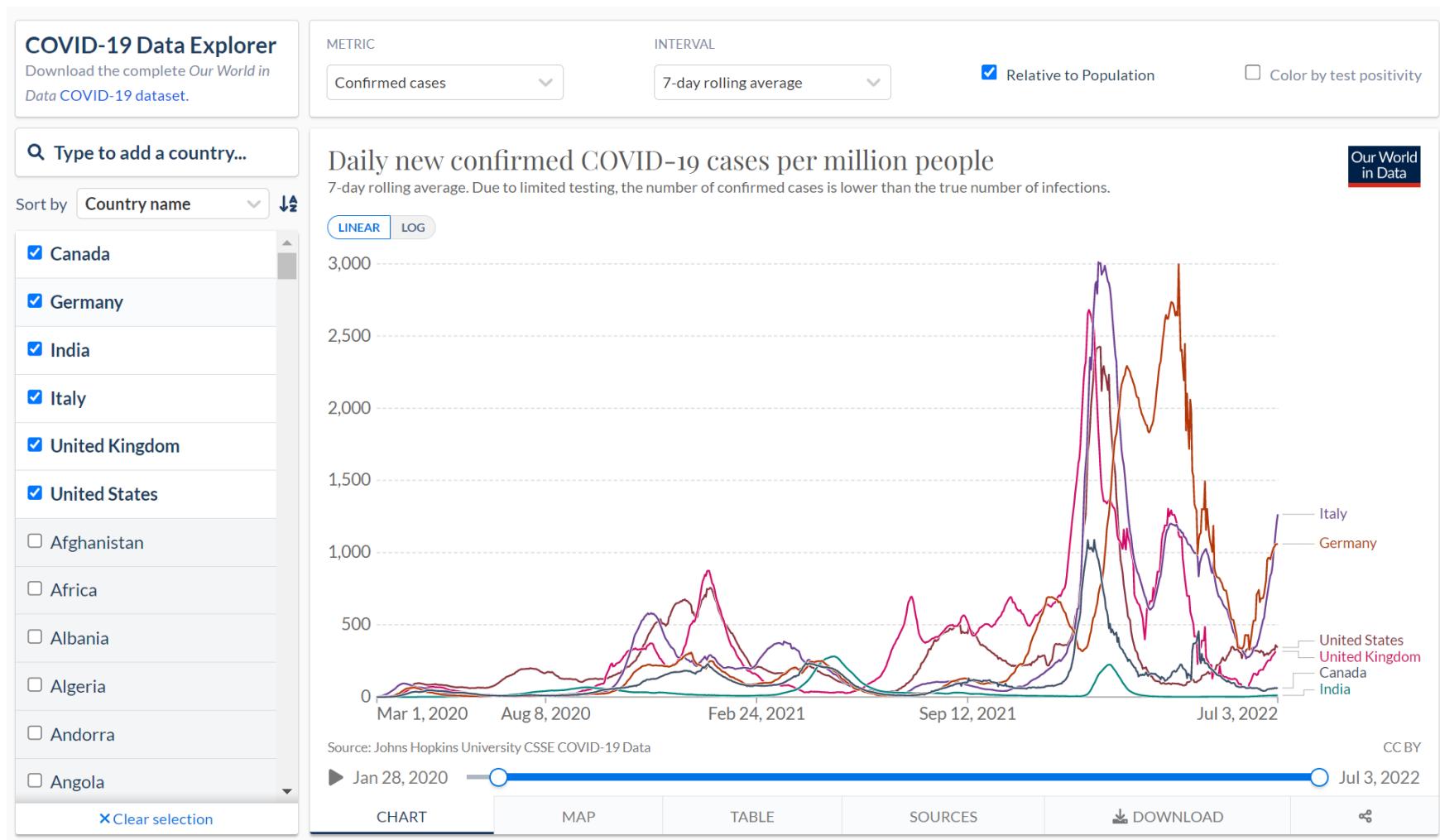


Task abstraction

Data analysis → Exploratory

- You don't know what you are looking for
- You want to know what questions to ask!

Data exploration



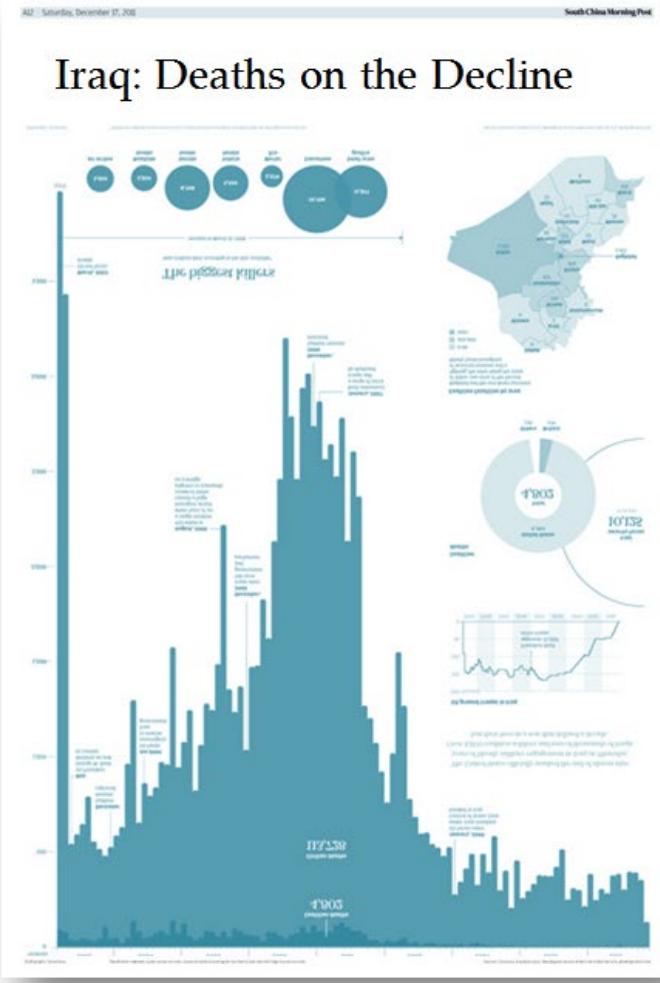
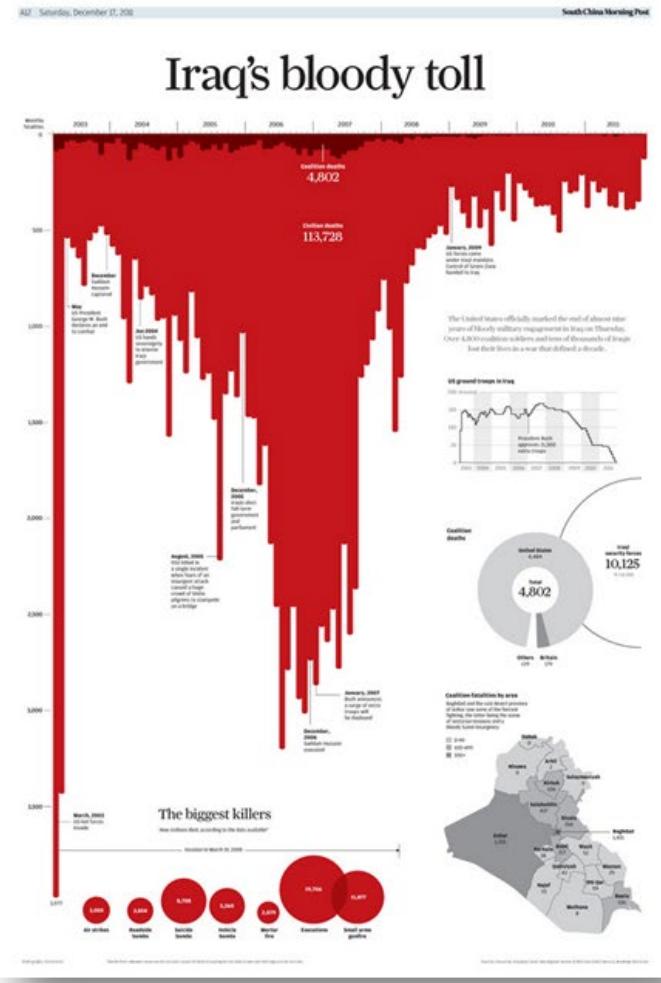
[COVID-19 Data Explorer](#)

Task abstraction

Data presentation → Explanatory

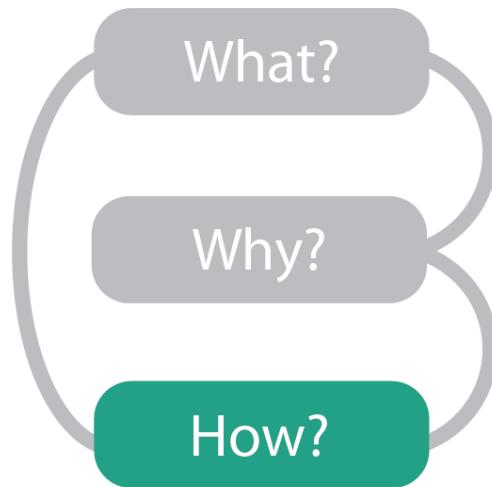
- You know what you want to show
 - Focus/highlight
 - Provide evidence
 - Simplify
 - Influence and persuade

Data presentation



Iraq's bloody toll

Visual encoding



...is it shown?

Visual encoding

Marks

→ Points



→ Lines



→ Areas



Channels

→ Position



→ Color



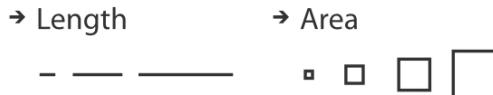
→ Shape



→ Tilt



→ Size



→ Volume



Visual encoding

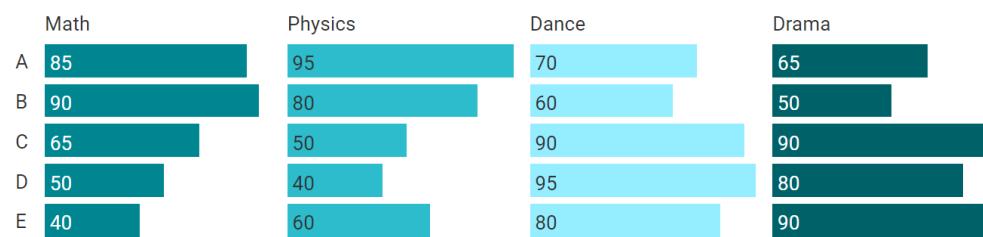
Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Visual encoding

Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

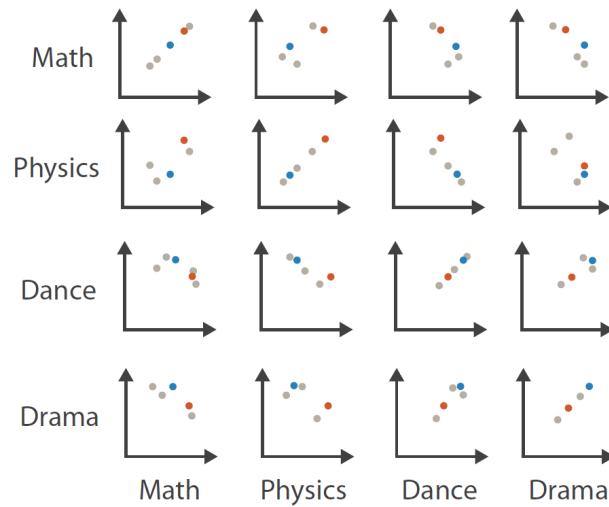


Visual encoding

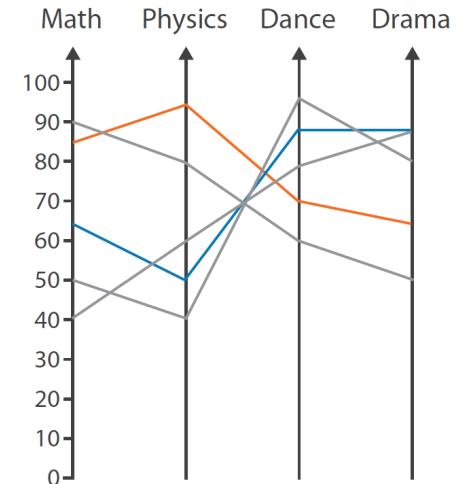
Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Scatterplot Matrix

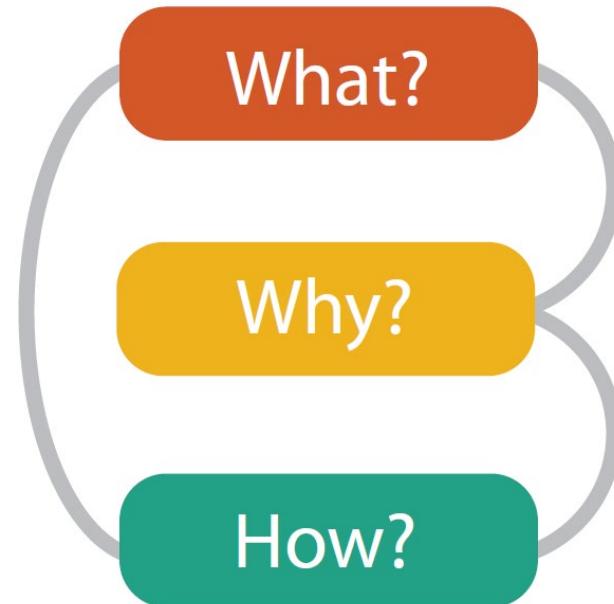


Parallel Coordinates



Take aways

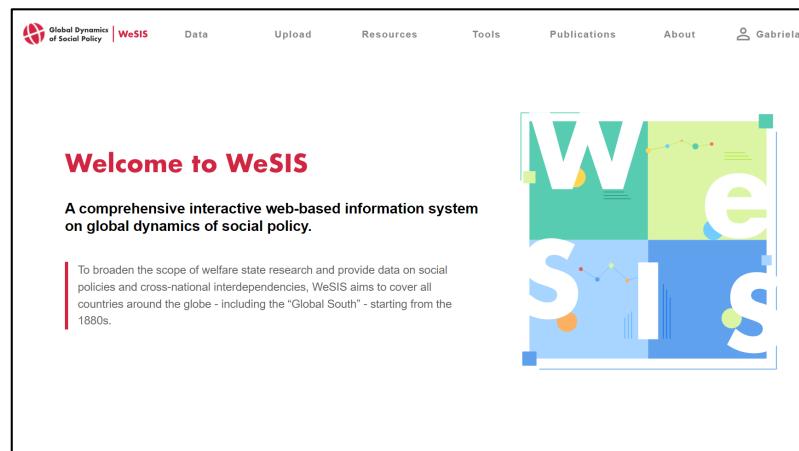
- ✓ Use Munzner's analysis framework
- ✓ Take a look at the data
- ✓ Think about the task of the potential user
- ✓ How could data visualization help?
- ✓ Then, try compatible visual encodings
- ✓ You can add interactivity after deciding on the visual encoding



Research Example

Computational Social Science Research CRC 1342: Global Dynamics of Social Policy

- CRCs are 3 x 4-year German research programmes
- Our partner: SOCIUM (Research Center on Inequality and Social Policy)
- Collaboration of computer scientists with political scientists, sociologists and geographers
- Goal: Analyze the global dynamics of public social policy



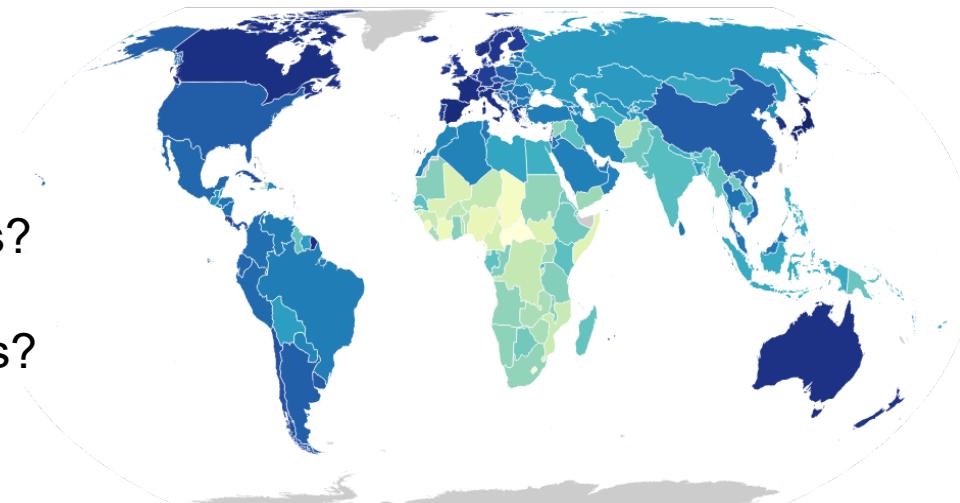
Mapping the Global South: Equal -Area Projections for Choropleth Maps

Gabriela Molina León, Michael Lischka, Andreas Breiter

Published at IEEE VIS 2020

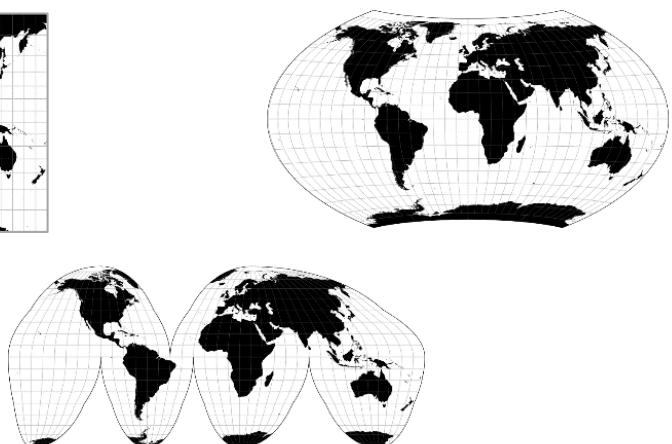
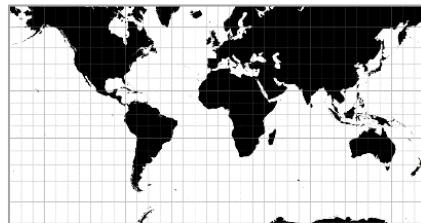
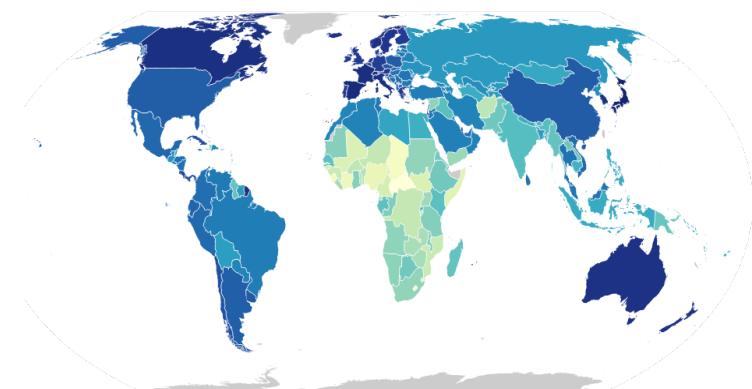
Research Questions

1. How do domain tasks influence the design choices for choropleth maps?
2. What equal-area projection is preferred for world choropleth maps?



Motivation

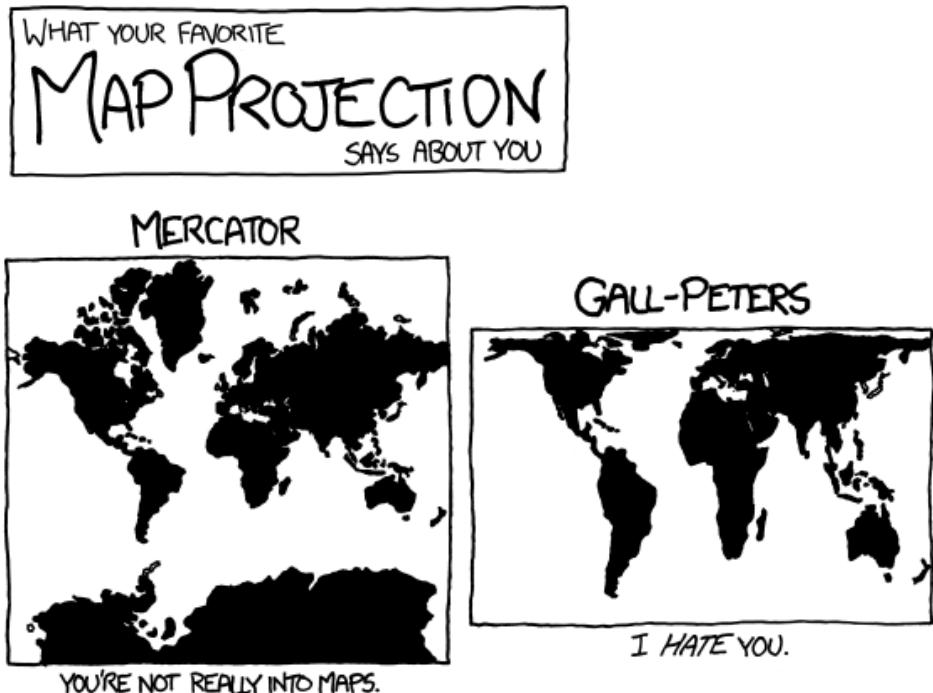
- Choropleth maps are one of the most common visualization techniques used to present geographical data
- Many different map projections to choose from
- Choropleth maps require equal-area projections to visually compare densities correctly



Projections from [D3-geo-projection](#)

Motivation

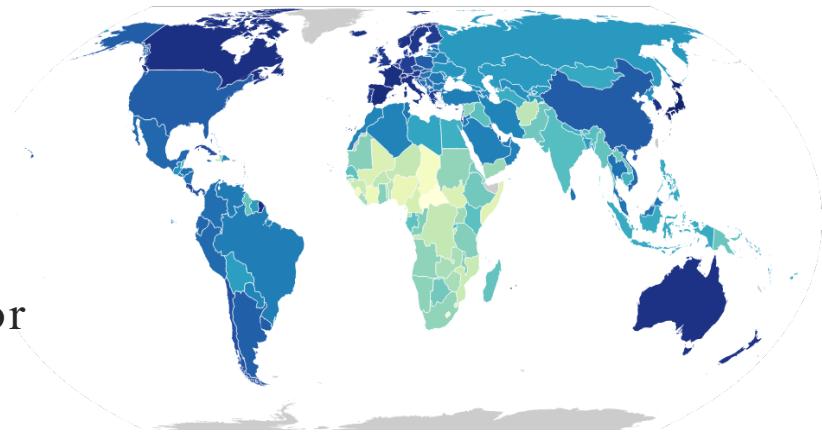
- There is no clear criteria to select one equal-area projection
- Ongoing debate on how projections influence our view of the world
 - Mercator is often used wrongly
 - Gall-Peters has become popular although it severely distorts tropical regions
- Visualization research has not looked into projection selection for choropleth maps yet



Extracts from [xkcd's “Map projections“ comic](#)

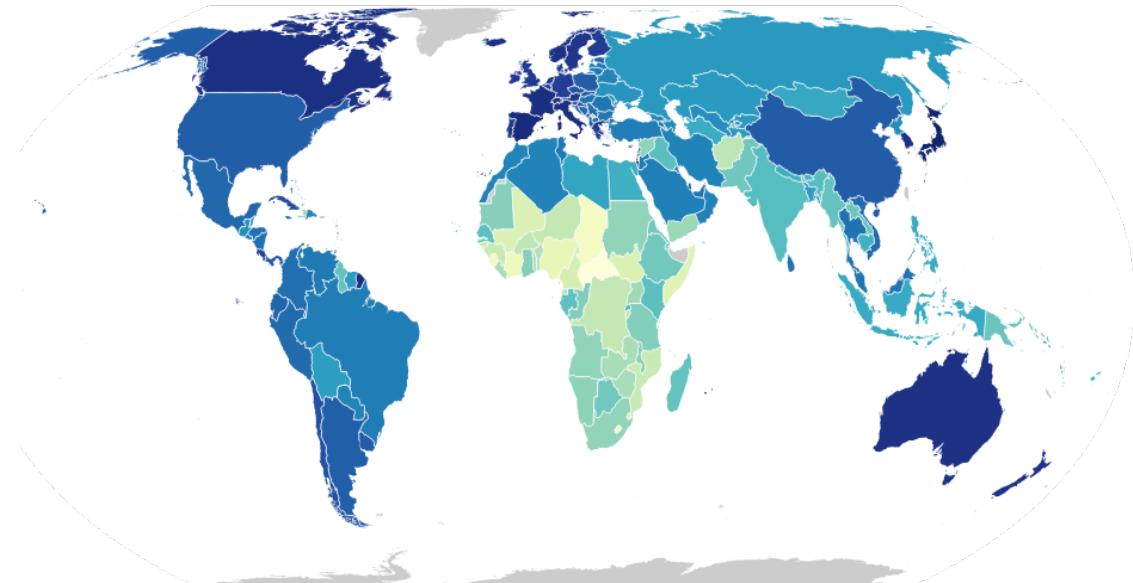
Case Study

- We worked together with social scientists who conduct research on the Global South
 - “Third World”
 - Low- and middle-income countries
- They visualize data with choropleth maps
- They asked us for
 - A map projection different than Mercator
 - World maps not centered in Europe
- Social science data portals often use a non-equal-area projection



Research Questions

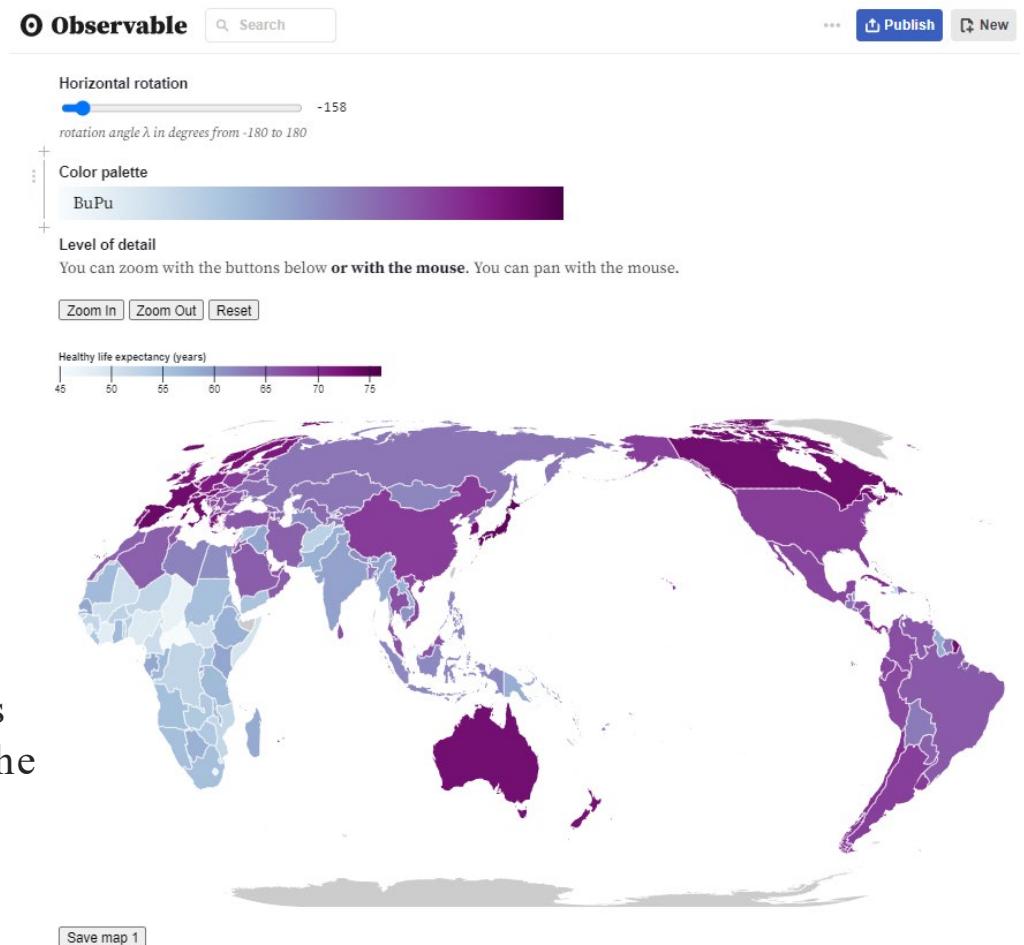
1. How do domain tasks influence the design choices for choropleth maps?
2. What equal-area projection is preferred for world choropleth maps?
 - Eckert IV
 - Equal Earth
 - Mollweide
 - Hammer
 - Galls-Peters





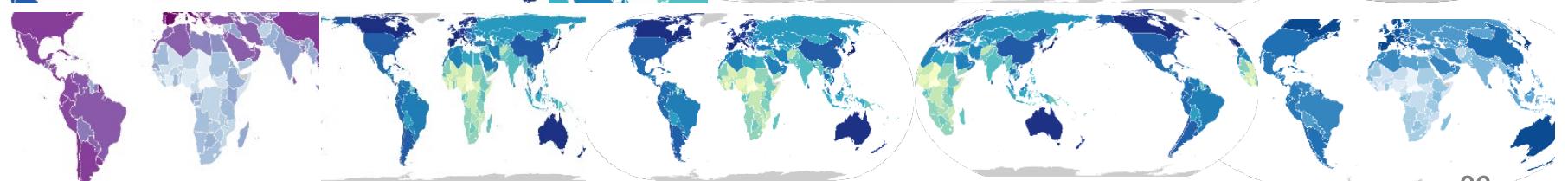
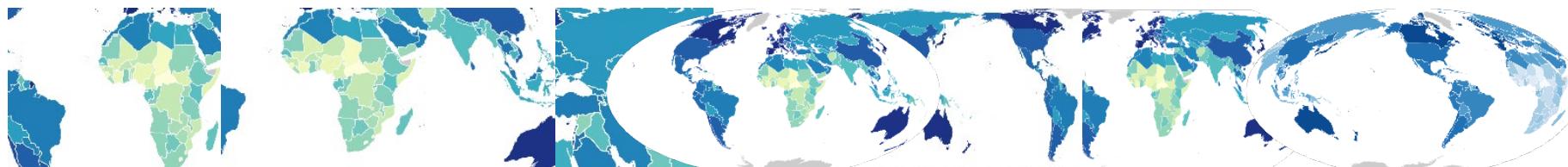
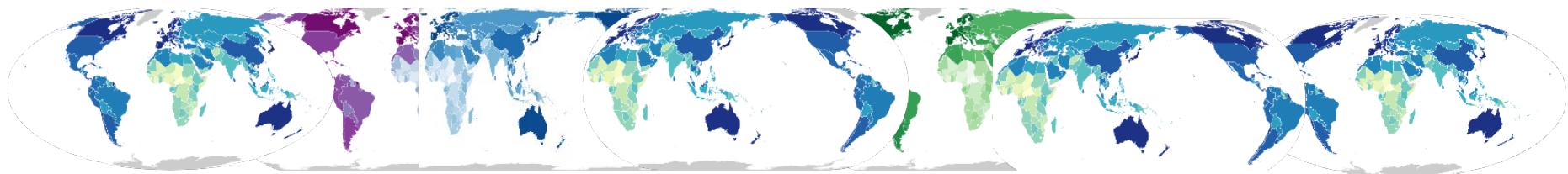
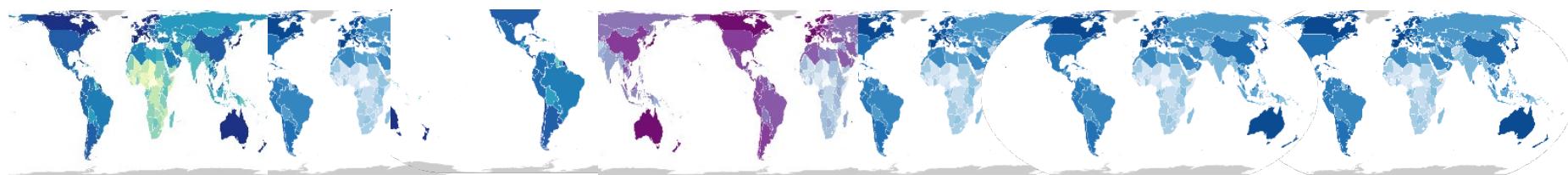
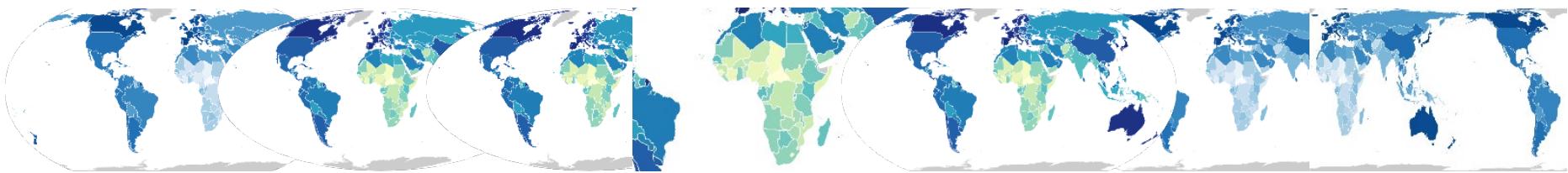
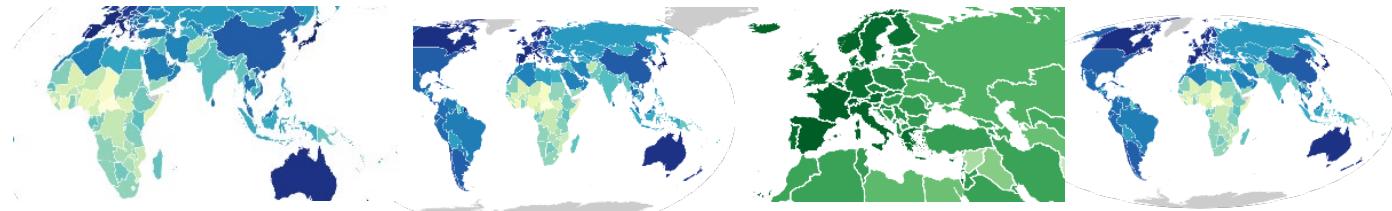
Survey

- We asked 20 social scientists to design 2 world choropleth maps
- Tasks in computational notebooks
- Design parameters
 - Projection
 - Center
 - Scale
 - Color scheme
- Participants ranked the parameters according to their importance for the design



[Observable notebook example](#)

Results



Results

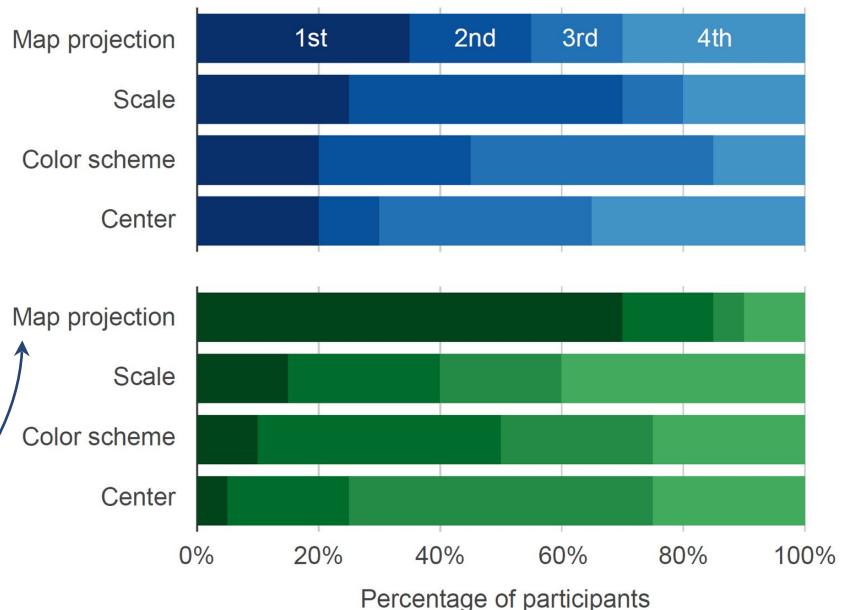
1. How do domain tasks influence the design choices for choropleth maps?

The design choices were influenced by the research goals and tasks

- 17 of 20 participants chose 2 different projections for each task
- Researchers interested in a specific region ranked the scale and center as more important
- In Task 2, Friedman's test revealed a significant effect of the parameters in the ranking

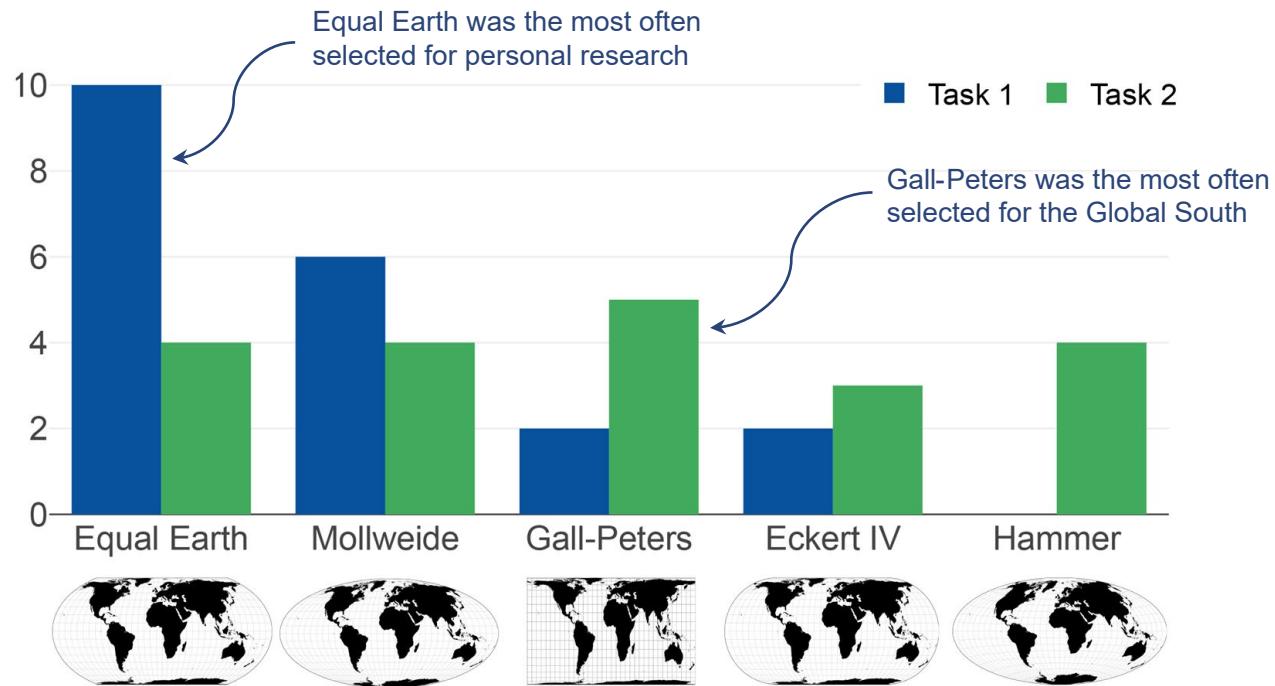
$$\chi^2(3) = 15, p = 0.002$$

The projection was significantly more important in Task 2



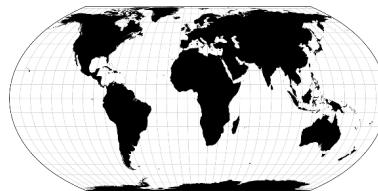
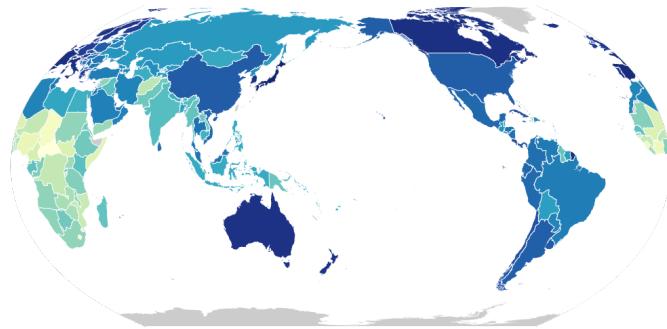
Results

2. What equal-area projection is preferred for world choropleth maps?

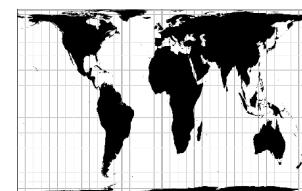


What we concluded

- The design choices changed across tasks
- The projection was considered the most important choice
- More important than the map center for avoiding Eurocentrism
- Equal Earth was most often selected
- Gall-Peters was often associated with the belief of an equal-area projection not being aesthetically appealing



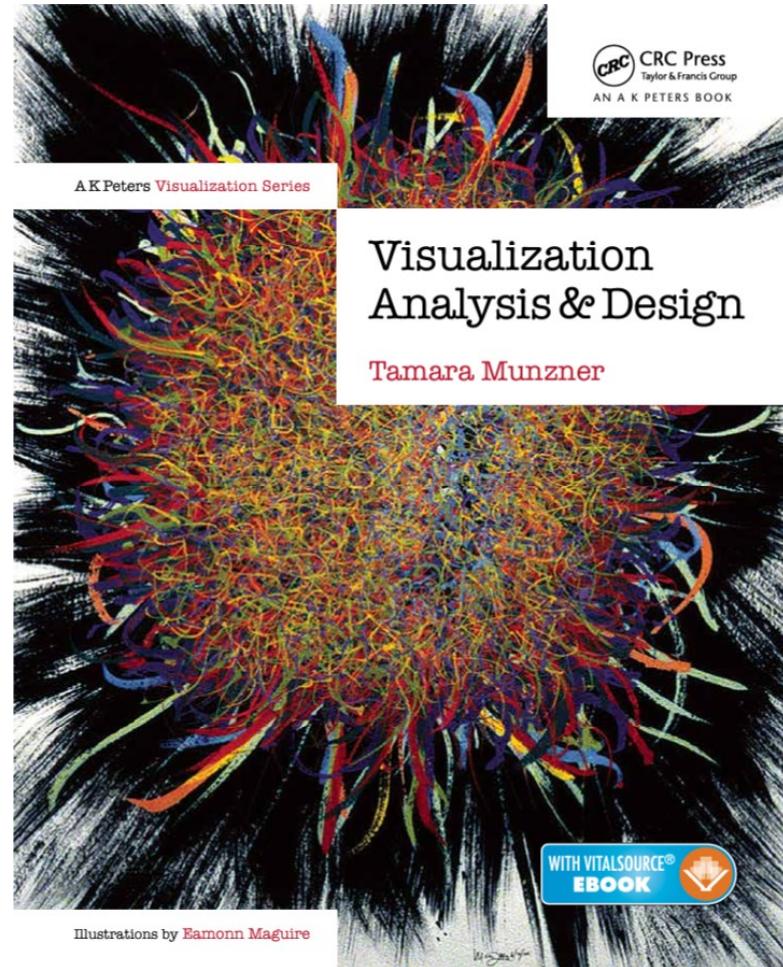
Equal Earth



Gall-Peters

Main textbook

- “Visualization Analysis & Design” by Prof. Tamara Munzner (University of British Columbia)
- It does not require previous knowledge on vis or complex math
- E-book is [available via the library](#)





Next week: Data Science & Vis Presentation (20% of your grade)

- **Every group member needs to present.**
- If you can't make it in person, talk to your team and make it possible in hybrid format.



Data Science & Vis Presentation

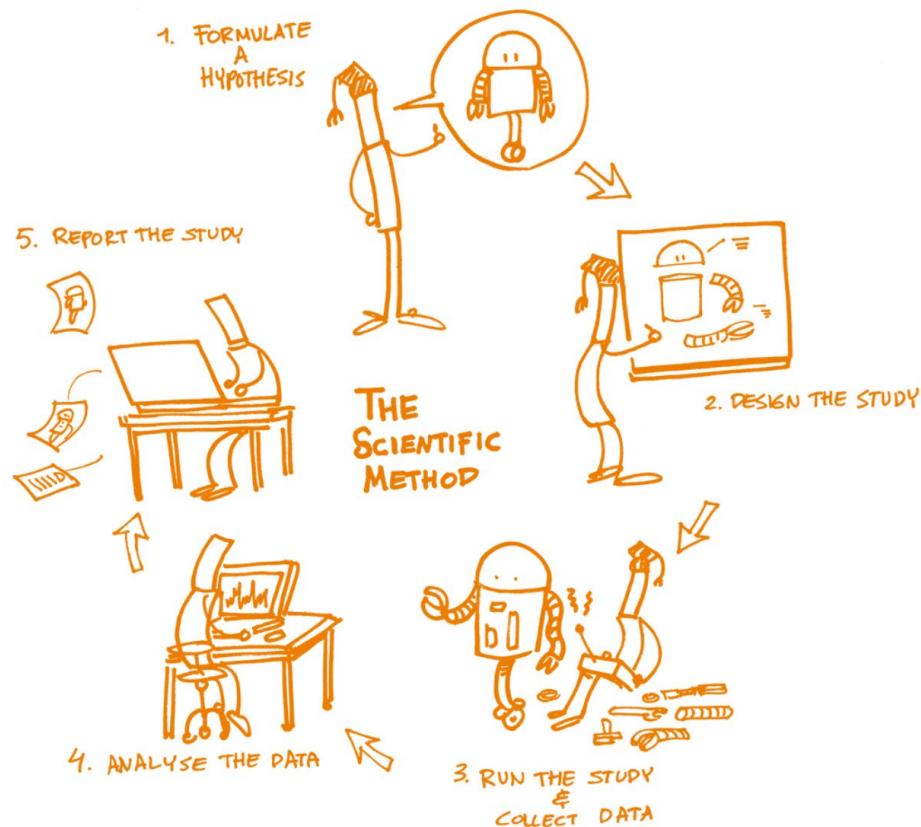
- In groups of 2 or 3, you will present a data science & vis paper
- Goal: Get a shared understanding of data science & vis
- You have to present the content of a paper
- Argue why it is relevant to data science & vis
- Argue why it is relevant to you
- You have 5 minutes (4 minutes if you are group of 2)

Group:
Topic:

- Clear Slides
- Clear Delivery
- Timing
- Clarity of topic / focus
- Command of content
- Critical analysis
- Conclusion
- Questions and Answers
- Overall Performance

Comments:

Data Science & Visualization



Scientific Presentations

Gabriela Molina León
molina@uni-bremen.de

Institute for Information Management Bremen
Information Management Group (AGIM)

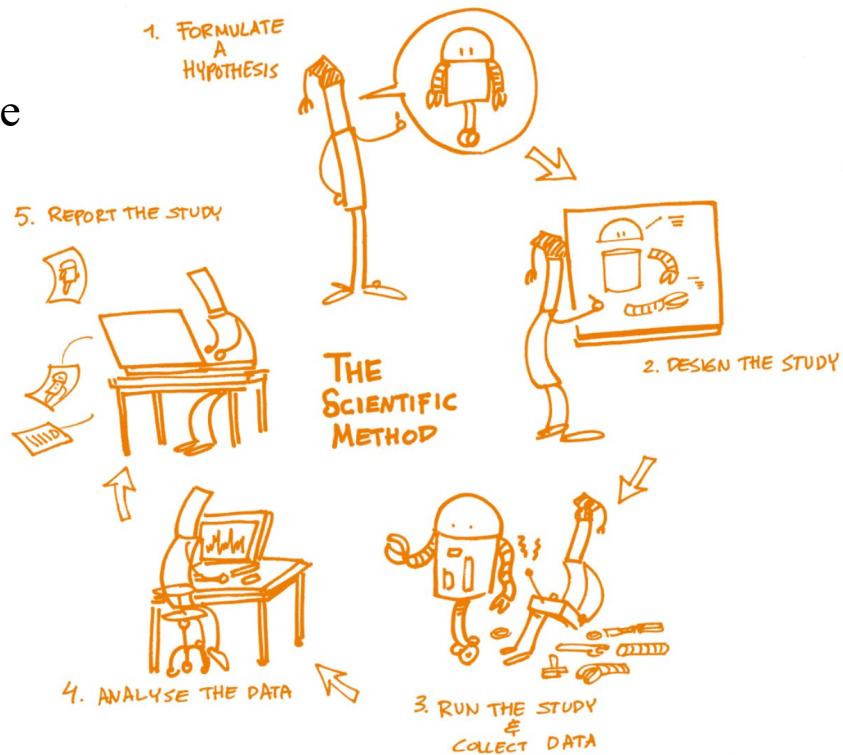
Source: [The Open Science Training Handbook](#)



What is research?

Fundamental or applied research

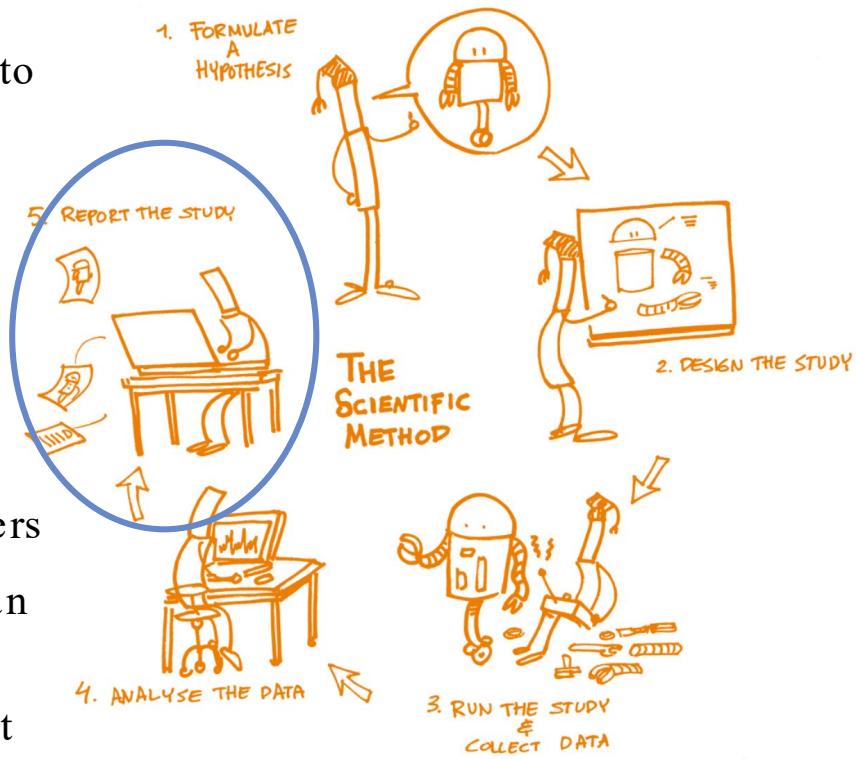
- Research is the creation of new knowledge
- Based on existing knowledge, we can generate new concepts, methodologies, and understandings
- The scientific method is the process of observing, asking questions, and seeking answers through tests and experiments



Papers, why?

Scientific currency

- Academic publications are the main method to share scientific findings
- Publication record is important to get a scientific job
- The audience is academics
 - Reviewers often have at least a Master degree
 - Professors are the most wanted reviewers
- Depending on the conference/journal, you can assume that the reader is....
 - A computer scientist or a social scientist
 - An expert on the main topic, e.g., social network analysis



Structure

Sections

- Usually 4-7 sections
- Standard sections
 - Abstract
 - Introduction
 - Related work
 - ...Design / Methodology / Evaluation / Results ...
 - Discussion
 - Conclusion
 - References

Catchy Titles Are Good: But Avoid Being Cute

Jacob O. Wobbrock
The Information School | DUB Group
University of Washington
Seattle, WA USA 98195
wobbrock@uw.edu

ABSTRACT
The most important rule of Abstracts is that they describe the work, not the paper. Include, at most, one sentence of motivation. Save the rest of your motivation for the introduction. Effective Abstracts focus on two things: (1) Describing what was *done*. (2) Describing what was *found* (key results). Be specific about your key findings. Instead of “many” say “84%”. Keep the Abstract to one paragraph and fewer than 200 words.

Author Keywords
Authors’ choice of terms; separated by semicolons; commas, within terms only; this section is required.

ACM Classification Keywords
Example: H.5.m. Information interfaces and presentation (e.g., HCI); Miscellaneous;
See <http://acm.org/about/class/1998> for the full list of ACM classifiers. This section is required.

INTRODUCTION
The Introduction delivers the motivation for your paper. It explains *why* you did the work you did. This is the primary function of the Introduction.

I have found a 5-point structure for Introductions to be particularly effective. (Here, I build on advice I received as a Ph.D. student from Prof. Scott E. Hudson.)

- State of the world...
- The big BUT...
- Therefore we did...
- The key findings are...
- The contributions of this work are...

The state of the world is a description of issues in whatever “world” is relevant to your topic. Drawing on popular press can be effective here if recent news items or data from articles support your cause.

The big BUT is where a problem is introduced. Or, similarly, a problem can be framed as an opportunity. Whether you are solving a problem or seizing an opportunity, motivate your work by connecting it to *things that matter to people*.

Finally, a good way to end your Introduction is by framing the contributions that your work makes. (See “Seven Research Contributions in HCI” [12] for examples of different research contributions.) I often structure my contributions as a numbered list within a paragraph. “The main contributions of this work are: (1) ...; (2) ...; (3) ...”

Copyright block. For submissions, put simply: "Submitted for review."

Copyright © 2015 Jacob O. Wobbrock – DRAFT v3

Figure 1. Choose a telling figure for your paper that is placed at the top of the right-hand column on the first page. I like to make figure captions (at least 8 pt) so that the figure is the body of the text. I also like to place my figures in-line as single “characters” in their own paragraphs, and captions as their own separate paragraphs, rather than placing either in floating text boxes, which jump around. I give captions 12 pt, after-paragraph line spacing. This figure is from [13].

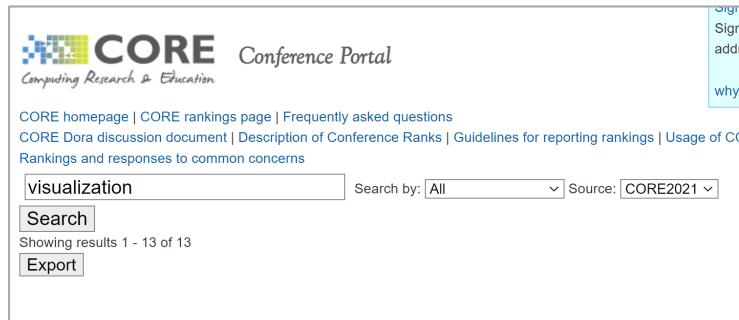
[Wobbrock \(2015\)](#)

Selecting literature

Rule of thumb:

Books > scientific journal papers > conference papers > workshop papers > tutorials > others

- Journals and conferences have different rankings
- Most common source to check is CORE:
<https://www.core.edu.au/conference-portal>
- Most popular search engine for scientific publications is [Google Scholar](#)
- ACM Digital Library and IEEE Xplore for CS publications

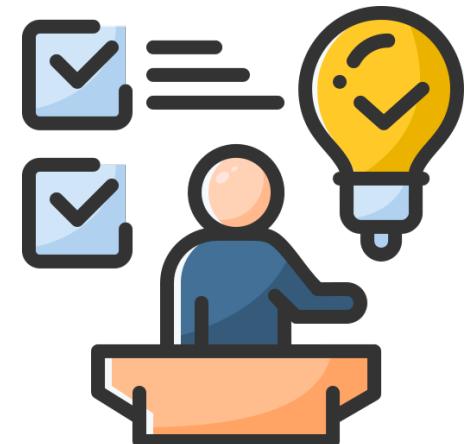


The screenshot shows the CORE Conference Portal interface. At the top, there's a logo for 'CORE Computing Research & Education' and a 'Conference Portal' link. Below that is a search bar with 'visualization' typed in, and dropdown menus for 'Search by: All' and 'Source: CORE2021'. To the right of the search bar are buttons for 'Search', 'Export', and links to 'CORE homepage', 'CORE rankings page', 'Frequently asked questions', 'CORE Dora discussion document', 'Description of Conference Ranks', 'Guidelines for reporting rankings', and 'Usage of CORE Rankings and responses to common concerns'. The main area displays a table of search results:

Title ◇	Acronym ◇	Source ◇	Rank ◇	DBLP ◇	hasData? ◇
IEEE Visualization	IEEE VIS	CORE2021	A	view	No
Eurographics/IEEE Symposium on Visualization	EuroVis	CORE2021	B	view	Yes
Eurographics Symposium on Parallel Graphics and Visualization	EGPGV	CORE2021	B	view	No
IEEE Pacific Visualization Symposium (was APVIS)	PacificVis	CORE2021	B	view	Yes
Cooperative Design, Visualization, and	CDVE	CORE2021	C	view	No

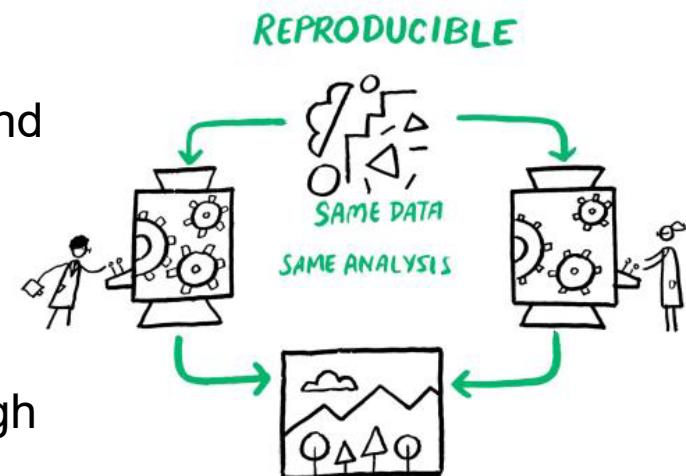
Research Questions

- What is the specific issue or problem the authors focus on ?
 - Helping a specific group of people in a specific task
 - Solving a computational problem faster
- Why is this a relevant topic for research?
 - Why should anyone work on it?
 - It needs to be relevant for others



Methods

- Quantitative, qualitative or mixed methods
 - Quantitative methods require enough data and a representative sample (e.g., statistics)
 - Common challenge with ML
 - Qualitative methods require time to go through the data (e.g., interviews)
- Every decision needs to be carefully considered: sample, variables, etc.



Source: [The Open Science Training Handbook](#)

Research contributions

- **What can we do that wasn't possible before?**
 - Creating a new system, tackling a new task, etc.
- **How can we do something better than before ?**
 - Proposing a new method or improving an existing method
- **What do we know that was unknown or unclear before?**
 - Findings about a specific group of people, find new data, etc.

Oral Presentation

- Scientific presentation is storytelling
 - Coherent story
 - Completeness is not the goal, be selective
- The goal is that your audience understands
 - Adjust complexity to your audience
 - Figure and examples help



Structure of your slides

- **Title slide**
 - Paper title, authors, date, speakers
- **Content slides**
 - Your version of the paper's story
 - Leave enough space between elements
 - Decide what to put on slide and what to say
 - Clarify what is from you and what from others
- **Conclusion slide(s)**
 - Always! Takeaways, outlook



Source: [Plagiarism signature](#)

Talking and timing

- Match words on slides, but complement them
 - No need to learn them by heart
- Speak to your audience, look at them
 - You can have notes, but avoid reading them out loud
- Timing
 - Use your time, but stick with time limit
 - Practice your talk, is your story understandable?



Sessions & Deliverables

Date	Lecture (10:00-11:30)	Practical (11:45-13:15)
08.04.24	Introduction to Data Science	Python Introduction
15.04.24	Basic Statistics & Supervised Learning	Practical Statistics + Supervised
22.04.24	Unsupervised Learning	Practical Unsupervised
29.04.24	Introduction to Data Visualization	Guest Lecture on NLP by Oxana Vitman
06.05.24	Exploratory Data Analysis	Data Science & Vis Presentation

19.07.24 Deadline Final Report

Note that for the slots **marked red**, you are expected to prepare presentations.
For the slots **marked blue**, you are expected to bring a computer.
In the slots **marked purple**, you are not expected to come to the classroom .

Grading Criteria

- Data Science & Vis paper presentation (20%)
- Exposé & exposé presentation (required)
- Interim presentation (required)
- Final presentation (30%)
- Final report (50%)

Credits (ECTS)

6 ECTS == 180 hours

- Lectures: 24h
- Tutorials: 24h
- Learning Python: 6h
- Data Science & Vis Paper Presentation: 12h
- Exposé: 18h
- Progress Presentation: 12h
- Final Presentation: 12h
- Final Report: 24h
- Final Project: 48h

Group Project

- As groups of three people, you will work on a project throughout the course
- Pick a dataset
- Pick a research question
- Pick a suitable method
- Find the best* analysis and visualization techniques for your dataset, question, and method
- Write a report on your findings and motivate your choices



Exposé Presentation

Date:

27.05.2024

- Title clear title that describes your project
- Abstract short summary of approach and main findings
- Introduction a text that introduces somebody new to the topic to what you did and why it is worth doing
- Background correctly cite the tools and approaches you used
- Method
 - Data what was your method?
 - Collection what analysis/vis/ML techniques did you apply?
how did you get the data?
- Results
 - Discussion what are your answers to your research questions?
what are highlights and lowlights?
what other interesting things did you observe?
what didn't work and why?
- Conclusion quick summary of your work in 1-2 paragraphs

Project Report

Deadline:
19.07.2024

- Title clear title that describes your project
- Abstract short summary of approach and main findings
- Introduction a text that introduces somebody new to the topic to what you did and why it is worth doing
- Background correctly cite the tools and approaches you used
- Method
 - Data what was your method?
 - Collection what analysis/vis/ML techniques did you apply?
how did you get the data?
- Results
 - Discussion what are your answers to your research questions?
what are highlights and lowlights?
what other interesting things did you observe?
what didn't work and why?
- Conclusion quick summary of your work in 1-2 paragraphs

