Research Papers and Process

Tamara Munzner Department of Computer Science University of British Columbia CPSC 547, Information Visualization 30 November 2022

http://www.cs.ubc.ca/~tmm/courses/547-22

Idiom pitfalls

Unjustified Visual Encoding

- -should justify why visual encoding design choices appropriate for problem - prerequisite: clear statement of problem and encoding!
- Hammer In Search of Nail
- -should characterize capabilities of new technique if proposed in paper

Color Cacophony

- -avoid blatant disregard for basic color perception issues
- · huge areas of highly saturated color • categorical color coding for 15+ category levels
- red/green without luminance differences
- encoding 3 separate attributes with RGB
- Rainbows Just Like In The Sky
- -avoid hue for ordered attribs, perceptual nonlinearity along rainbow gradient
- Later pitfalls: Tactics
- Stealth Contributions - don't leave them implicit, it's your job to tell reader explicitly!
- -consider carefully, often different from original project goals I Am So Unique
- -don't ignore previous work
- -both on similar problems and with similar solutions Enumeration Without Justification
- -"X did Y" not enough

Final pitfalls: Submission

Slimy Simultaneous Submission

- must say why previous work doesn't solve your problem
- -what limitations of their does your approach fix? I Am Utterly Perfect
- -no you're not; discussion of limitations makes paper stronger!

Straw Man Comparison

Today

• papers & research: pitfalls & process

- review reading, review writing, conference talks

• incl. course paper vs research paper differences

- making research available, reproducible, replicable

• What I Did Over My Summer Vacation

-don't be too low level, it's not a manual

-bonus points: new name for old technique

-don't focus on effort rather than contribution

-avoid tiny increment beyond (your own) previous work

- don't cram in so much content that can't explain why/what/how

-writing infovis research papers

· course endgame expectations

-final presentations

Later pitfalls: Strategy

Least Publishable Unit

Dense As Plutonium

Bad Slice and Dice

• fails reproducibility test

-two papers split up wrong

Later pitfalls: Results

• Unfettered By Time

- neither is standalone, yet both repeat

-final report

[evaluations]

open science

- -head-to-head hardware is best (re-run benchmarks yourself, all on same machine)
- Tiny Toy Datasets
- -asking labmates not convincing if target audience is domain experts

- Generality
- instant dual rejection, often multi-conference blacklist
- Resubmit Unchanged

- often detected when same reviewer for both

- respond to previous reviews: often get reviewer overlap, irritated if ignored

· strategy: all research tactics: all research · results: visualization specific

· encoding: visualization specific

style: all research, except

-Story-Free Captions, My Picture Speaks For Itself

Writing InfoVis Papers

Later pitfalls: Tactics Stealth Contributions

- don't leave them implicit, it's your job to tell reader explicitly! - consider carefully, often different from original project goals

Final pitfalls: Style

· Deadly Detail Dump

Story-Free Captions

Grammar Is Optional

Mistakes Were Made

- -choose level of detail for performance numbers - detailed graphs for technique papers, high-level for design & eval papers
- compare appropriately against state-of-the-art algorithms
- compare against state-of-the-art dataset sizes for technique (small ok for eval)
- But My Friends Liked It
- Unjustified Tasks
- -use ecologically valid user study tasks: convincing abstraction of real-world use

- don't use passive voice, leaves ambiguity about actor • your research contribution or done by others?

My Picture Speaks For Itself

- optimize for flip-through-pictures skimming

- explicitly walk them through images with discussion

- explain how only **after** what and why; provide high-level framing before low-level

-good low-level flow is necessary (but not sufficient), native speaker check good if ESL

Research Process & Pitfalls

- rare: insufficient background to judge worth - if reviewer didn't get your point, many readers won't

Pitfalls

writing infovis papers: pitfalls to avoid

Contributions in research papers

determines everything

often not obvious

Final pitfalls: Style 2

Nonspecific Use Of Large

• Jargon Attack

what are your research contributions?

-what can we do that wasn't possible before?

- diverged from original goals, in retrospect

-avoid where you can, define on first use • all acronyms should be defined

-quantify! hundreds? 10K? 100K? millions? billions?...

-how can we do something better than before?

- what do we know that was unknown or unclear before?

-from high-level message to which details worth including

state them explicitly and clearly in the introduction

-don't leave unsaid should be obvious after close reading of previous work

-goal is clarity, not overselling (limitations typically later, in discussion section)

-don't hope reviewer or reader will fill them in for you

Springer LNCS Volume 4950, p 134-153, 2008.

- Process and Pitfalls in Writing Information Visualization Research Papers.

Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, Chris North, eds.

Tamara Munzner. In: Information Visualization: Human-Centered Issues and Perspectives.

-your job: rewrite so clearly that nobody can misunderstand Reviewers Were Threatened By My Brilliance

Review reading pitfalls

• Reviewers Were Idiots

- seldom: unduly harsh since intimately familiar with area

• I Just Know Person X Wrote This Review

- sometimes true, sometimes false -don't get fixated, try not to take it personally

• It's The Writing Not The Work

- sometimes true: bad writing can doom good work (good writing may save borderline)

-sometimes false: weak work common! reinvent the wheel worse than previous one

Review writing pitfalls Uncalibrated Dismay - remember you've only read the best of the best! - most new reviewers are overly harsh It's Been Done, Full Stop - you must say who did it in which paper, full citation is best You Didn't Cite Me - stop and think whether it's appropriate - be calm, not petulant You Didn't Channel Me - don't compare against paper you would have written • review the paper they submitted	 Conference talk pitfalls Results As Dessert don't save until the end as a reward for the stalwart! showcase early to motivate A Thousand Words, No Pictures aggressively replace words with illustrations most slides should have a picture Full Coverage Or Bust cannot fit all details from paper communicate big picture talk as advertising: convince them it's worth their time to read paper! 	Paper writing process suggestions • pre-paper talk - write and give talk first, as if presenting at conference - iterate on talk slides to get structure, ordering, arguments right - then create paper outline from final draft of slides • encourages concise explanations of critical ideas, creation of key diagrams • avoids wordsmithing digressions and ratholes • easier to cut slides than prose you agonized over • pre-paper/practice talk feedback session: at least 2-3x talk length - global comments, then slide by slide detailed discussion - nurture culture of internal critique (build your own critique group if necessary) • have non-authors read paper before submitting - internal review can catch many problems - ideally group feedback session as above	Course Endgame
Logistics Assignments: Final Presentations on Canvas -upload due Wed Dec 14 noon (2 hrs before session) required & posted: slides (Project Final Presentation Slides, PDF) optional & posted: video (Project Final Presentation Video, mp4) Assignments: Final Report on Canvas -upload due Fri Dec 16 8pm (PST) required & posted: report (Project Final Report, PDF) required & posted: showcase image (Project Teaser Image, png) required but not posted: code incl README (Project Source Code and Other Materials, zip) encouraged & posted: live demo URL (include in code README) encouraged & posted: video (include in code zip *only* if different from final present video)	Final Presentations	Final presentations: Wed Dec 14 2-5:15 pm • length (14 projects) - presentation (live or prerecorded): 10 min for groups, 8 min for solo - Q&A live: 2 min per project • session structure - order reverse alphabetical by first name, from bottom up on project page - 2 breaks, between each set of 5-6 presentations - CS dept (fac / grads) & infovis group invited, friends/others very welcome! - refreshments served	Final presentations, cont • presentation structure - content: motivation/framing, project, results, critique/limitation • standalone: don't assume audience has read proposal or updates (or remembers your pitch) - slides (& slide numbers) mandatory for main part - demo strongly encouraged, either live or prerecorded - format is up to you: live presentation or prerecorded video or a mix • slides/video upload - upload to Canvas Assignments: Final Slides (mandatory), Final Video (optional) - by noon Wed Dec 14 • note: code freeze after presentations! - no additional work on project allowed after presentation deadline - additional two days to get it all written down coherently for final report
Final Presentations Schedule 2.00-2:10 Yaman Sanobar. The Use of Data Visualization in E-Commerce: A Review. 2.10-2:20 Rosalyn Carr. Modified VAST Challenge with Applications to Data Breaches. 2.20-2:30 Matas I.B. Oddo. B-Martix Network Visualization. 2.230-2:42 Madonna Huang, Michael Yin. A Visualization Tool for Global Wastewater Treatment Plants. 2.42-2:52 Jianhao Cao. TableRepo Viz: Visualizing Tabular Data Repositories for Facilitating Descriptive Tag Augmentation. 2.52-3:05 BREAK 3.05-3:17 Ian Hill, Matthew Tang. GraceFall: Visualizar for Diverse Stress Test Degradation Data Spanning Multiple Time Scales. 3:17-3:29 Han Wang, Xin Wang. AsylumLoupe: EU Asylum Demographics and Movement Information Visualization.	Final presentations marking • template (may change) - Intro/Framing: 20% - Main: 30% - Limitations/Critique/Lessons: 10% - Slides: 10% - Presentation/Video Style: 10% - Demo: 10% (or N/A) - Question Handling: 10% • marking by buckets - great 100% - good 89% - ok 78% - poor 67% - zero 0%	Marking: Course overall • 50% Project, summative assessment at end - 15% Final Presentation - 25% Final Report - 60% Content - (Milestones pass/fail, penalty only if missed or unacceptable) • pitch 5%, proposal 10%, update 10% • 14% Sync: In-Class Participation - 12 sessions, 1% per session - 2% final presentations - (most got full credit)	Final Reports
Final reports • PDF, use InfoVis templates http://junctionpublishing.org/vgtc/Tasks/camera_tvcg.html -your choice to use Latex/Word/whatever • no length cap: illustrate freely with screenshots! - design study / technique: aim for at least 6-8 pages - analysis / survey: aim for at least 15-20 pages • strongly encouraged to re-use text from proposal & update writeups • encourage looking at my writing correctness and style guidelines - https://www.cs.ubc.ca/~tmm/writing.html • strongly encourage looking at previous examples - https://www.cs.ubc.ca/~tmm/courses/547-22/projectdesc.html#examp - Example Past Projects (curated list) - direct links to all project pages to browse, 2021-2003	 Course requirements vs research paper standards research novelty not required mid-level discussion of implementation is required - part of my judgement is about how much work you did - high level: what toolkits etc did you use - medium level: what pre-existing features did you use/adapt - low level not required: manual of how to use, data structure details design justification is required - (unless analysis/survey project) - different in flavour between design study projects and technique projects - description of design what you did is necessary but not sufficient publication-level validation not required - user studies, extensive computational benchmarks, utility to target audience 	Report structure: General • low level: necessary but not sufficient - correct grammar/spelling - sentence flow - ideal: formal technical voice, not conversational style • medium level: order of explanations - build up ideas - ideal: carefully structured, not stream-of-consciousness infodump • high through low level: why/what before how - paper level • motivation: why should I (reader) care • overview: what did you (writer) do • details: how did you (writer) do it - section level • overview then details - sometimes subsection or paragraph level	Sample outlines: Design study • https://www.cs.ubc.ca/~tmm/courses/547-22/projectdesc.html#outlines • Abstract - concise summary of your project - do not include citations • Introduction - give big picture, establish scope, some background material might be appropriate • Related work - include both work aimed at similar problems & similar solutions - no requirement for research novelty, but still frame how your work relates to prev - cover both academic & relevant non-academic work - (you could reorder to have this section later)

Sample outlines: Design study II Data and Task Abstractions analyze your domain problem according to book framework (what/why) include both domain-language descriptions and abstract versions could split into data vs task, then domain vs abstract - or vice versa! typically data first then task, so that can refer to data abstr within task abstr must have tight connections between data & task abstr Solution describe your solution idiom (visual encoding and interaction) analyze it according to book framework (how) only for custom encodings, no need to repeat book material for standard chart types justify your design choices as solutions to problem set up w/ data/task abstractions provide rationale, discuss choices with respect to alternatives if significant algorithm work, discuss algorithm and data structures	 Implementation medium-level implementation description specifics of what you wrote vs what existing libraries/toolkits/components do Milestones breakdown of who did what work remember to update milestones: add actual hours/date to estimated hours/date totals required Results include scenarios of use, extensively illustrated with multiple screenshots of your software walk reader through exactly how your interface succeeds (or falls short) of solving intended problem report on evaluation, if you did any (eg deployment to target users, computational benchmarks) screenshots should be png (lossless compression) not jpg (lossy compression)! Discussion / Future Work reflect on your approach: strengths, weaknesses, limitations lessons learned: what do you know now that you didn't when you started? future work: what would you do if you had more time? 	 Conclusions summarize what you've done different than abstract since reader has seen all the details Bibliography note format is numerical & alphabetical use citation manager / bibtex! make sure to use real references for work that's been published academically not just URL check arxiv papers, some have link to final publication venue, also search on titles! check carefully to ensure consistency & nothing mangled or missing most online sources require cleanup see guidance at html#refs	Marking • design study & technique & explainer • 12.5% each for - intro - related work - abstractions - solution - implementation/milestones - results - discussion - 10% style, 2.5% bibliography
Sample outlines: Technique (diffs) Abstract, Introduction (same as above) Related Work - big focus on similar solutions, some discussion of similar problems (same task/data combo) Data and Task Abstractions - much shorter than the corresponding one for design studies, framing context not core contrib Solution - describing proposed idiom exactly, not justifying its use for particular domain problem - as above, analyze in terms of design choices, justify why appropriate vs alternatives Implementation/Milestones (same as above) Results - less emphasis on scenarios with particular target users - more emphasis on characterizing the breadth of possible uses - still definitely include screenshots of the system in action Discussion / Future Work, Conclusions, Bibliography (same as above)	 Sample outlines: Survey (diffs) Abstract (same as above) Introduction discuss the scope of what you're covering, why it's interesting/reasonable partition compared to visualization as a whole Related Work only previous surveys focus on how your work is similar to or different from them, especially wrt coverage Main break up into sections based on your own synthesis of themes of work covered you might want a Background section at the start if domain-focused survey where there's important vocabulary/ideas to establish before diving into main discussion analyze visualizations proposed in these papers in terms of what/why/how framework (if applicable) include images from papers Milestones, Discussion / Future Work, Conclusions, Bibliography (same as above) marking: intro (10%), relwork (10%), main (60%), milestones/discussion (10%), style (10%) 	 Sample outlines: Implementation (diffs) Abstract, Introduction (same as above) Related Work - paper you're reimplementing, maybe other closely related work for framing context - much shorter than other project types Scope - big picture of what you did, esp. only a subset of original paper or covering multiple papers - nice to have somewhat comprehensible & standalone document but no need to explain in full • ok to discuss similarities and differences assuming familiarity with goals of original work Implementation - detailed implementation discussion: much more than other project types - as above, include specifics of what you build on vs what you coded yourself - issues that arose: choices unclear in original, subtleties and nuances you discovered along the way, challenges in adapting toolkit capabilities 	Sample outlines: Implementation (diffs) Results - as above, should include screenshots of your software that illustrate scenarios of how to use it • but less emphasis particular target users in scenarios - definitely include computational benchmarks to evaluate your work • Milestones, Discussion / Future Work, Conclusions, Bibliography (same as above) • marking: intro (10%), relwork (10%), main (60%), milestones/discussion (10%), style (10%)
Report marking • required: at least material I've listed —you may include more material —you may choose alternate orderings • reminder: project content is 60% of entire project mark —report is 25%, presentation is 15% • you'll get detailed written feedback —combined: final presentation, final report, project content —in some cases, next steps	• required: submit your code - so I can see what you've done, but I will not post - include README.txt file at root with brief roadmap/overview of organization • which parts are your code vs libraries • how to compile and run - but I do not necessarily expect your code compiles on my machine • no need to submit data if it's huge • encouraged but not required - submit live demo URL (provide in README.txt file) - open-source your code (if so, fine to just send me that URL) - submit supporting video (if different from final presentation) • with or without voiceover • voiceover is very very nice to have later, software bitrot makes demos not last forever!	Showcase image • showcase image for projects page - 300x300 image - call it showcase.png - required	Course Evaluations (link on Canvas)
Open Science: Available, Reproducible, & Replicable Research	Making your research available & reproducible: why bother? • moral high ground -for Science! • enlightened self-interest -make your own life easier -you'll be cited more often by academics -your work more likely to be used by industry	Making the world care about your research! Increasing the Impact of Visualization Research panel, VIS 2017 Krist Wongsuphasawat, Data Visualization Scientist, Twitter WHAT LEADS TO SUCCESSFUL ADOPTION! WHAT LEADS TO SUCCESSFUL ADOPTION! The standard of the second source of the second source and	Disseminating research • paper page for each paper - everything! PDF, supplemental materials, videos, software/demos, talk slides, figures, - examples: • Table Scraps, http://www.cs.ubc.ca/group/infovis/pubs/2020/table-scraps/ • TimeLineCurator, http://www.cs.ubc.ca/group/infovis/pubs/2020/table-scraps/ • Write blog post to accompany each paper - very high-impact bang for the time buck • Multiple Views: Visualization Research Explained umbrella blog https://medium.com/multiple-views-visualization-research-explained • UW IDL individual lab blog - Surprise Maps: Showing the Unexpected https://medium.com/@uwdata/surprise-maps-showing-the-unexpected-e92b67398865 - Bayesian Surprise Maps http://idl.cs.washington.edu/papers/surprise-maps/

Archival dissemination: what to provide	Supplemental materials: provide as much as possible	
 paper post it online at non-paywalled site algorithm document well in paper itself document further with code code make available as open source (github.com) pick right spot on continuum of effort involved, from minimal to massive just put it up warts and all, minimal documentation well documented and tested (build a whole community - not the common case) supplemental materials 	 demo videos: show interactive look & feel data for computational benchmarks & case studies tricky issue in visualization: data might not be yours to release! qualitative work: thematic analysis raw & intermediate materials quant experimental stimuli: full set of images, not just a few examples quant evaluation: data analysis code/scripts evaluation: detailed study results advance planning: ethics approval! if sanitize PII (personally identifiable information) technique refinement: previous iterations parameters: how exactly to regenerate/produce figures, tables additional case studies, screenshots, other exposition surveys / design spaces: interactive faceted browser examples: treevis.net, dashboarddesignpatterns.github.io 	
Replication: crisis in psychology, medicine, etc early rumblings left me with (ignorable) qualms - papers: Is most published research false?, Storks Deliver Babies (p= 0.008), The Earth is spherical (p < 0.05), False-Positive Psychology groundswell of change for what methods are considered legitimate - out: questionable research practices (QRPs) p-hacking / p-value fishing / data dredging Hypothesizing After Results are Known (HARKing) - in replication pre-registration: avoid "garden of forking paths" & motivated reasoning - brouhaha with bimodal responses some people doubling down and defending previous work many willing to repudiate (their own) earlier styles of working	Remarkable introspection on methods • psych: thoughtful willingness to change standards of field - Andrew Gelman's commentary on the Susan Fiske article • http://andrewgelman.com/2016/09/21/what-has-happened-down-here-is-the-winds-have-changed/ - Simine Vazire's entire Sometimes I'm Wrong blog • http://sometimesimwrong.typepad.com/ • especially posts on topic Scientific Integrity - Joe Simmons Data Colada blog post What I Want Our Field to Prioritize • http://datacolada.org/53/ - Dana Carvey's brave statement on her previous power pose work • http://faculty.haas.berkeley.edu/dana_carney/pdf_My%20position%20on%20power%20poses.pdf • vis:	
Next week: Research guests & more • Steve Kasica (UBC)	Come talk! • encourage meeting with me to get advice/feedback before final present	

- Steve Kasica (UBC)
- qualitative research
- Stephen Kobourov (Arizona)
- techniques & algorithms
- Mara Solen (UBC)
- survey papers

- · design spaces for visualization
- visualizing imperfect models
- next steps

- -chance to get feedback while you can still act on it
- -optional, not mandatory
- -wise to schedule in advance by email
- can't meet with all 14 teams in next week office hours, or in last few days!
- encourage meeting with me to get advice/feedback before final present

Dissemination & reproducibility: motivation & howto

- Open Practices in Vis Research, Steve Haroz
- https://osf.io/8ag3w/download
- Cody Dunne VIS22 panel talk (10 min) https://youtu.be/nPdr7xybUbA?t=260
- why important to host stuff on site that will stay forever (arxiv.org, osf.io)
- vs personal and even research group sites that can disappear
- appropriately enough his slides hosted at https://osf.io/mfk5z
- osf.io
- -great for supplemental materials in addition to paper (vs arxiv focus on paper PDF)
- can create anonymous view-only link for double-blind review
- https://help.osf.io/article/155-create-a-view-only-link-for-a-registration
- -advice: post when you submit, update with camera-ready
- don't wait conference presentation, might not happen! (worse yet: promise will do it soon)
- examples: osf.io/tr3sb, osf.io/uezfk

When and how will this storm hit visualization?

- they're ahead of us (they = psychology)
- they have some paper retractions
- we don't (yet) have any retractions for methodological considerations
- they agonize about difficulty of getting failure-to-replicate papers accepted
- we hardly ever even try to do such work
- they are a much older field
- we're younger: might our power hierarchies thus be less entrenched??...
- -they are higher profile
- we don't have vis research results appear regularly in major newspapers/magazines
- they have rich fabric of blogs as major drivers of discussion
- crosscutting traditional power hierarchies
- we have far fewer active bloggers
- replication crisis was focus of BELIV 2018 workshop at IEEE VIS
- evaluation and BEyond methodoLogIcal approaches for Visualization
- https://beliv-workshop.github.io/2018/

Reproducibility: Levels of effort required

• 5: 15 minutes with free tools

• 3: considerable effort

• 0: cannot be reproduced

• 2: extreme effort

[Vandewalle, Kovacevic and Vetterli.

• 4: 15 minutes with proprietary tools

• I: cannot seem to be reproduced

Reproducible Research in Signal Processing - What, why, and how. IEEE Signal Processing Magazine, 26(3):37-47, May 2009.]

Upcoming