Wrapup: Research Papers and Process

Tamara Munzner
Department of Computer Science
University of British Columbia

CPSC 547, Information Visualization
1 December 2021

http://www.cs.ubc.ca/~tmm/courses/547-21
Today

• papers & research: pitfalls & process
  – writing infovis research papers
  – review reading, review writing, conference talks

• course endgame expectations
  – final presentations
  – final report
    • incl. course paper vs research paper differences

• [evaluations]

• open science
  – making research available, reproducible, replicable

• next steps
  – ways to continue on with visualization
Writing InfoVis Papers
Pitfalls

• writing infovis papers: pitfalls to avoid
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 attributes, perceptual nonlinearity along rainbow gradient
Later pitfalls: Strategy

• What I Did Over My Summer Vacation
  – don’t focus on effort rather than contribution
  – don’t be too low level, it’s not a manual

• Least Publishable Unit
  – avoid tiny increment beyond (your own) previous work
  – bonus points: new name for old technique

• Dense As Plutonium
  – don’t cram in so much content that can’t explain why/what/how
    • fails reproducibility test

• Bad Slice and Dice
  – two papers split up wrong
  – neither is standalone, yet both repeat
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
Contributions in research papers

• what are your research contributions?
  – what can we do that wasn’t possible before?
  – how can we do something better than before?
  – what do we know that was unknown or unclear before?

• determines everything
  – from high-level message to which details worth including

• often not obvious
  – diverged from original goals, in retrospect

• state them explicitly and clearly in the introduction
  – don’t hope reviewer or reader will fill them in for you
  – don’t leave unsaid should be obvious after close reading of previous work
  – goal is clarity, not overselling (limitations typically later, in discussion section)
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
  – 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!
Later pitfalls: Results

• Unfettered By Time
  – choose level of detail for performance numbers
  – detailed graphs for technique papers, high-level for design & eval papers

• Straw Man Comparison
  – compare appropriately against state-of-the-art algorithms
  – head-to-head hardware is best (re-run benchmarks yourself, all on same machine)

• Tiny Toy Datasets
  – compare against state-of-the-art dataset sizes for technique (small ok for eval)

• But My Friends Liked It
  – asking labmates not convincing if target audience is domain experts

• Unjustified Tasks
  – use ecologically valid user study tasks: convincing abstraction of real-world use
Final pitfalls: Style

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

- **Story-Free Captions**
  - optimize for flip-through-pictures skimming

- **My Picture Speaks For Itself**
  - explicitly walk them through images with discussion

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

- **Mistakes Were Made**
  - don’t use passive voice, leaves ambiguity about actor
    - *your research contribution or done by others?*
Final pitfalls: Style 2

• Jargon Attack
  – avoid where you can, define on first use
    • all acronyms should be defined

• Nonspecific Use Of Large
Final pitfalls: Submission

• Slimy Simultaneous Submission
  – often detected when same reviewer for both
  – instant dual rejection, often multi-conference blacklist

• Resubmit Unchanged
  – respond to previous reviews: often get reviewer overlap, irritated if ignored
Generality

- encoding: visualization specific
- strategy: all research
- tactics: all research
- results: visualization specific
- style: all research, except
  - Story-Free Captions, My Picture Speaks For Itself
Research Process & Pitfalls
Review reading pitfalls

• Reviewers Were Idiots
  – rare: insufficient background to judge worth
  – if reviewer didn’t get your point, many readers won’t
  – your job: rewrite so clearly that nobody can misunderstand

• Reviewers Were Threatened By My Brilliance
  – 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 15 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 17 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 15 2-5pm

• 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 alphabetical by first name, as on project page
  – 2 breaks, between each set of 5-6 presentations
  – CS dept (fac / grads) & infovis group invited, friends/others very welcome!

• presentation structure
  – content: **motivation/framing, project, results, critique/limitation**
    • standalone: don’t assume audience has read proposal or updates (or remembers your pitch)
  – slides (**slate 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
Final presentations, cont

• slides/video upload
  – upload to Canvas Assignments: Final Slides (mandatory), Final Video (optional)
  – by noon Wed Dec 15

• 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

• 2:00-2:12 Abi Kuganesan, Ivan Song, Lufei Liu. 

• 2:12-2:24 Arash Kamyabi, Negar Sadrzadeh. 
  Drinking Behavior Patterns in Dairy Cattle.

• 2:24-2:36 Armita Safa, Janet Li, Neera Patadia. 
  Multiscale Visualization of Pathogenic Structural Variants.

• 2:36-2:48 David Chen, Hongyang Yang, Madison Lore, Niels Semb. 
  A New City Map.

• 2:48-3:00 Deepansha Chhabra, Lucie Polakova, Niloofar Zarif. 
  What Can We Learn from User-Movie Ratings?

• 3:00-3:10 BREAK

• 3:10-3:22 Elizabeth Reid, Mifta Sintaha, Nichole Boufford. 
  SoundMap: A Visualization Tool to Explore Multi-Attribute Sound Data.

• 3:22-3:34 Felipe Gonzalez-Pizarro, Soheil Alavi. 
  MultiModalTopicExplorer: Topic modeling for exploring multi-modal data from asynchronous online conversations.

• 3:34-3:44 Hadi Sinaee. 
  PartViz: Visualizing Graph Partitioners.

• 3:44-3:56 Inna Ivanova, Jonatan Engstad. 
  Explorify: A Personalized Interactive Visualization Tool for Spotify Listening History.

• 3:56-4:08 Jocelyn Minns, Mary Abikoye, Minglong Li. 
  Necklace Maps for COVID-19 Visualization.

• 4:08-4:18 BREAK

• 4:18-4:28 Mara Solen. 
  Visualization Literacy in the Age of Big Data: Vital Skills for Modern Media Consumption.

• 4:28-4:40 Marie Salomon, Noa Heyl, Shizuko Akamoto, ToTo Tokao. 
  Course Friction Explorer: Visualizing and Validating Indicators of Student Struggle.

• 4:40-4:50 Michael Tegegn. 
  Visualizing Android Features Through Time.

• 4:50-5:00 Zainab Saeed Wattoo. 
  Visualizing the Run Time Execution of Command Patterns.
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 up to 25% if missed)*
    • pitch 5%, proposal 10%, update 10%

• 36% Async Discussion
  – 9 weeks, 4% per week
    • 75% own comments, 25% responses
    • *(almost all got full credit)*

• 14% Sync: In-Class Participation
  – 12 sessions, 1% per session
  – 2% final presentations
  – *(almost all got full credit)*
Final Reports
Final reports

• PDF, use InfoVis templates http://junctionpublishing.org/vgtc/Tasks/camera_tvvcg.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

• strongly encourage looking at previous examples
  – www.cs.ubc.ca/~tmm/courses/547-21/projectdesc.html#examp
  – Example Past Projects (curated list)
  – direct links to all project pages to browse 2020-2003
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
  – technique explanation alone is not enough
• 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

• medium level: order of explanations
  – build up ideas

• high through low level: why/what before how
  – paper level
    • motivation: why should I care
    • overview: what did you do
    • details: how did you do it
  – section level
    • overview then details

  – sometimes subsection or paragraph level
Sample outlines: Design study

- [www.cs.ubc.ca/~tmm/courses/547-21/projectdesc.html#examp](http://www.cs.ubc.ca/~tmm/courses/547-21/projectdesc.html#examp)
- **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 it
  - cover both academic & relevant non-academic work
  - you might 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

• 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 with respect to alternatives
  – if significant algorithm work, discuss algorithm and data structures
Sample outlines: Design study III

• 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

• Results
  – include scenarios of use illustrated with multiple screenshots of your software
    • walk reader through how your interface succeeds (or falls short) of solving intended problem
    • report on evaluation you did (e.g., 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?
Sample outlines: Design study IV

• 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 http://www.cs.ubc.ca/~tmm/writing.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: 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
Code / Video

• 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
    • I do not necessarily expect your code compiles on my machine

• 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
    • 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
Evaluations
Open Science: Available, Reproducible, & Replicable Research
Making the world care about your research!

• Increasing the Impact of Visualization Research panel, VIS 2017
  – Krist Wongsuphasawat, Data Visualization Scientist, Twitter

https://www.slideshare.net/kristw/increasing-the-impact-of-visualization-research
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-scrap/

• 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
Making your research reproducible

• why bother with reproducibility?
  – 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

• reproducibility levels
  – 5: 15 minutes with free tools
  – 4: 15 minutes with proprietary tools
  – 3: considerable effort
  – 2: extreme effort
  – 1: cannot seem to be reproduced
  – 0: cannot be reproduced

Reproducibility: Levels to consider

- **paper**
  - post it online
  - make sure it stays accessible when you move on to new place
  - external archives are better yet ([arxiv.org](http://arxiv.org))

- **algorithm**
  - well documented in paper itself
  - document further with supplemental materials

- **code**
  - make available as open source
  - 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)
Reproducibility: Levels to consider, cont.

• data
  – make available
    • technique/algorithm: data used by system
      – tricky issue in visualization: data might not be yours to release!
    • evaluation: user study results
      – ethics approval possible if PII (personally identifiable information) sanitized, needs advance planning

• parameters
  – how exactly to regenerate/produce figures, tables
  – example: http://www.cs.utah.edu/~gk/papers/vis03/
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: QRPs (questionable research practices)
    • p-hacking / p-value fishing / data dredging
    • Hypothesizing After Results are Known (HARKing)
  – in
    • replication
    • pre-registration

– 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

• 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
When and how will this storm hit visualization?

• they’re ahead of us
  – 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
  – http://beliv.cs.univie.ac.at/
Next Steps
Tools & ideas: Andy Kirk's Visualizing Data

Videos

• many great conferences with free videos online
  – broadly accessible: OpenVisConf, Eyeo, InformationPlus
  – cutting-edge technical research: IEEE VIS
Redesign En Masse: **Makeover Mondays**

- easy entry point (Tableau focus)

---

**Week 14 – Millions of UK workers at risk of being replaced by robots**

Apr 7, 2017

During week 14 we looked at job automation and the potential impact of robots and AI on the UK employment market.

---

**Week 13 – The Secret of Success**

Mar 31, 2017

Week 13 took a look at a Russian survey about the secret of success. Dot plot, bump charts, bar charts, radar charts. This week had it all! Plus seven lessons to take on board.

---

**Week 12 – March Madness**

Mar 24, 2017

We looked at March Madness data for week 12, highlighting the phenomenon that is US college basketball. Quite a few vizzes showed the passion that

---

[http://www.makeovermonday.co.uk/blog/](http://www.makeovermonday.co.uk/blog/)

---
Visual Design Process In Depth: Dear Data

• inspiring celebration of data humanism

http://www.dear-data.com/by-week/

Giorgia Lupi and Stefanie Posavec
Visual Design Process In Depth: **Data Sketches**

- detailed process notes, from sketching through coding

http://www.datasketch.es/

Shirley Wu and Nadieh Brehmer
Pathways for more participation

• join Viz@UBC
  – https://dfp.ubc.ca/initiatives/viz-ubc
  – get on visatubc-announce email list (send mail to vizatubc-info@cs.ubc.ca)
  – talk series

• join Vancouver Visualization meetup
  – https://www.meetup.com/Vancouver-Data-Visualization/
  – 4K members

• join Data Visualization Society
  – https://www.datavisualizationsociety.com/
  – less than three years old, 18K+ members around the world
  – resources, jobs board, super-active Slack incl local groups, challenges, ...
  – articles on highly active blog/journal: Nightingale
Next Week
Come talk!

• encourage meeting with me to get advice/feedback before final present
  – 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 last few days or in Tue office hours!