Wrapup: Research Papers and Process

Tamar Munzner
Department of Computer Science
University of British Columbia

CPSC 547 Information Visualization
26 November 2019
http://www.cs.ubc.ca/~tmm/courses/547-19

Sample outlines: Design study III
• implementation
  – medium-level implementation description
  – specifics of what you wrote vs what existing libraries/components do
  – common pitfalls of what did work
  – results
  – report on evaluation you did (eg deployment to target users, computational benchmarks)
  – requirements of your visual interface
  – should you be try a user study?
  – discuss the scope of what you're covering
  – why it's interesting/reasonable partition compared to what existing libraries do
  – what did work

Sample outlines: Survey (diffs)
• Abstract (same as above)
• Introduction
  – background section as the start if domain-focused survey
  – what's important in your visual domain
  – how to set up your study
  – design
  – user interface
  – how to test your design
  – results
  – discussion

Sample outlines: Analysis (diffs)
• Abstract, same as above
• Introduction
  – background section as the start if domain-focused survey
  – design
  – user interface
  – how to test your design
  – results
  – discussion

Sample outlines: Other types
• see page for implementation project types

Today
• finalize final presentation slot
• presentations
• final papers and final presentations
• course paper vs research paper expectations
• writing infols: papers: pitfalls to avoid
  – Process and Prifles in Writing Information Visualization Research Papers.
  – Tamar Munzner: In Visualization Information: Human-Centered Issues and Perspectives.
  – Andreas Kerren, John T. Dabbs, Jean-Daniel Fekete, Chris North, eds.
• other research pitfalls and process
• reproducible and repeatable research

Final papers and Presentations
• PDF, use InfoVis templates http://junctionpublishing.org/vgtc/Tasks/camera_tvcg.html
• your choice to use Latex/Word/whatever
• Example Past Projects
• browse 2015, 2014,… reports

Report structure: General
• low level necessary but not sufficient
• medium level: order of explanations
• 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 than details
  – sometimes subsection or paragraph level

Wrapup:
Research Papers and Process

Tamara Munzner
Department of Computer Science
University of British Columbia

CPSC 547 Information Visualization
26 November 2019
http://www.cs.ubc.ca/~tmm/courses/547-19

Sample outlines: Design study III
• implementation
  – medium-level implementation description
  – specifics of what you wrote vs what existing libraries/components do
  – common pitfalls of what did work
  – results
  – report on evaluation you did (eg deployment to target users, computational benchmarks)
  – requirements of your visual interface
  – should you be try a user study?
  – discuss the scope of what you're covering
  – why it's interesting/reasonable partition compared to what existing libraries do
  – what did work

Sample outlines: Survey (diffs)
• Abstract (same as above)
• Introduction
  – background section as the start if domain-focused survey
  – design
  – user interface
  – how to test your design
  – results
  – discussion

Sample outlines: Analysis (diffs)
• Abstract, same as above
• Introduction
  – background section as the start if domain-focused survey
  – design
  – user interface
  – how to test your design
  – results
  – discussion

Sample outlines: Other types
• see page for implementation project types

Report marking
• required at least material I've listed
• you may include more material, you may choose alternate orderings
• probable marking scheme (may change)
  – design study & technique: 12.5% each for
    – intro, related work, abstraction, solution, implementation, results, discussion, style
      – style: 10% more, 2.5% bibliography
    – survey: intro (10%), related work (10%), main (60%), style (20%
    – analysis: intro/domain (8%), abstr (8%), related work (8%), methods/tools (8%), analysis
      (52%), discussion (8%), style (8%)
  – reminder: project content is 60% of entire project mark

report is 23%, presentation is 15%
Contributions in research papers
• what are your research contributions?
– what can we do that wasn't possible... reading of previous work
– goal is clarity, not overselling (limitations typically later, in discussion section)

Encouraged but not required
• submit live demo URL
– open-source your code (if so, fine to just send me that URL)
– submit supporting video
– with or without voiceover
– very nice to have later, software bistro makes demos last forever!
– can be same or different from what you show in final presentation

Code/Video
• required: submit your code
– so I can see what you've done, but I will not post
– include README file
– includes questions for 1 min (brief questions only)
– submission structure
– no prerequisites files, just code
– must be open source, includes minimal code to reproduce experiment
– can be required vs. optional: I will run it, you show me what you have
– code freeze

Demo/Video
– submit live demo URL
– should contain questions: aim for 1 min (brief questions only)
– …

Idiomatic
• Unjustified Visual Encoding:
– should justify why visual encoding design choices appropriate for problem
– preeminate clear statement of problem and encoding!
• Hammer in Search of Nails:
– 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
– redundant without limitations differences
– encoding 3 attributes with RGB
– Rainbows Just Like In The Sky
– avoid hue for ordered attribs, perceptual nonlinearity along rainbow gradient

Process & Pitfalls for InfoVis 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
– general 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 will read it
– don’t leave unclear how you contributed

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

Later pitfalls: Results
• Unfettered By Time
– choose level of detail for performance numbers
– detailed graphs for technical 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 listeners not convinced that original audience is domain experts

Later pitfalls: Style
• Deadly Detail Dump
– explain how only after what and why; provide high-level framing before details
• Story-Free Captions
– optimize for flow-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

Final presentations
• context
– CS department will be invited, also feel free to invite others
– refreshments will be served, two short breaks
– order alphabetical by first name
• code freeze
– no additional work on project after presentation deadline
– additional three days to get it all written down coherently for final report

Final presentations: Tue Dec 13 3-7 (?) FSC 2300A
• length (19 projects)
– 14 min for 3-person teams, 10 min for 2-person teams, 10 min for 1-person teams
– includes questions for 1 min (brief questions only)
• session structure
– order alphabetically first name as an project page (with or without)
– 2 breaks, between each set of presentations
– dinner hosted, friends welcome, refreshments served

Marking: Course overall
• 50% Project, summative assessment at end
– 15% Final Presentation
– 25% Final Report
• 40% Written Questions
– 6 weeks, 10% each
• 40% In-Class Discussion & Group Work (pass/fail)
– 4 weeks, 10% each
• marking by buckets
– 100% good
– 98% ok
– 67% poor
– 0% zero

Showcase image
• showcase image for projects page
– 300x300 image
– call it showcase.png or showcase.jpg

Logistics
• Assignments: Final Presentations on Canvas
– upload due Fri Dec 10 6pm
– (upload due 1 hr before presentations if using my laptop)
• Assignments: Final Report on Canvas
– upload Fri Dec 13 11:59pm
– required & posted report, showcase image
– required but not posted code including README
– encouraged live demo URL/video

Come talk!
• encourage meeting with me to get advice/breakdown before final present
– chance to get feedback while you can act on it
– optional, not mandatory
– do send email to schedule, can’t meet with all 19 teams in last few days!

Idiomatic
• Unjustified Visual Encoding
– should justify why visual encoding design choices appropriate for problem
– preeminate clear statement of problem and encoding!
• Hammer in Search of Nails
– 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
– redundant without limitations differences
– encoding 3 attributes with RGB
• Rainbows Just Like In The Sky
– avoid hue for ordered attribs, perceptual nonlinearity along rainbow gradient

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
– general 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 will read it
– don’t leave unclear how you contributed

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
• Idiom pitfalls
• Unjustified Visual Encoding
– should justify why visual encoding design choices appropriate for problem
– preeminate clear statement of problem and encoding!
• Hammer in Search of Nails
– 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
– redundant without limitations differences
– encoding 3 attributes with RGB
• Rainbows Just Like In The Sky
– avoid hue for ordered attribs, perceptual nonlinearity along rainbow gradient

Process & Pitfalls for InfoVis 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
– general 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 will read it
– don’t leave unclear how you contributed

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
• Idiom pitfalls
• Unjustified Visual Encoding
– should justify why visual encoding design choices appropriate for problem
– preeminate clear statement of problem and encoding!
• Hammer in Search of Nails
– 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
– redundant without limitations differences
– encoding 3 attributes with RGB
• Rainbows Just Like In The Sky
– avoid hue for ordered attribs, perceptual nonlinearity along rainbow gradient

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
– general 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 will read it
– don’t leave unclear how you contributed

Earlier 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
• Idiom pitfalls
• Unjustified Visual Encoding
– should justify why visual encoding design choices appropriate for problem
– preeminate clear statement of problem and encoding!
• Hammer in Search of Nails
– 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
– redundant without limitations differences
– encoding 3 attributes with RGB
• Rainbows Just Like In The Sky
– avoid hue for ordered attribs, perceptual nonlinearity along rainbow gradient

Later pitfalls: Results
• Unfettered By Time
– choose level of detail for performance numbers
– detailed graphs for technical 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 listeners not convinced that original audience is domain experts
• Unjustified Tasks
– avoid mentioning non-convincing user studies (convince at least 25%)

Later pitfalls: Style
• Deadly Detail Dump
– explain how only after what and why; provide high-level framing before details
• Story-Free Captions
– optimize for flow-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

Final presentations
• context
– CS department will be invited, also feel free to invite others
– refreshments will be served, two short breaks
– order alphabetical by first name
• code freeze
– no additional work on project after presentation deadline
– additional three days to get it all written down coherently for final report
Reproducibility: Levels to consider
- paper
  - post it online
  - make sure it stays accessible when you move on to new place
- code
  - make available as open source
  - pick right spot on continuum of effort involved, from minimal to massive
- data
  - make available
  - technical/algorithmic data used by system
  - richness in visualization data might not be pure to release
- parameters
  - how exactly to generate/produce figures, tables
  - example: [http://www.cs.utah.edu/~gk/papers/vis03/](http://www.cs.utah.edu/~gk/papers/vis03/)

View from industry
- Increasing the Impact of Visualization Research panel, VIS 2017
  - Krist Wiekpuphasawa, Data Visualization Scientist, Twitter
- View from industry
- Replication: crisis in psychology, medicine, etc
  - early rumblings left me with (ignorable) qualms
  - papers are most published research field, etc
  - data: is spherical (p < 0.05), False-Positive Psychology
  - ethics approval possible if PII (personally identifiable information) sanitized, needs advance planning
  - ethical approval possible if PII (personally identifiable information) sanitized, needs advance planning
- Replication: crisis in psychology, medicine, etc
  - early rumblings left me with (ignorable) qualms
  - papers are most published research field, etc
  - data: is spherical (p < 0.05), False-Positive Psychology
  - ethics approval possible if PII (personally identifiable information) sanitized, needs advance planning
- Replication: crisis in psychology, medicine, etc
  - early rumblings left me with (ignorable) qualms
  - papers are most published research field, etc
  - data: is spherical (p < 0.05), False-Positive Psychology
  - ethics approval possible if PII (personally identifiable information) sanitized, needs advance planning
- Replication: crisis in psychology, medicine, etc
  - early rumblings left me with (ignorable) qualms
  - papers are most published research field, etc
  - data: is spherical (p < 0.05), False-Positive Psychology
  - ethics approval possible if PII (personally identifiable information) sanitized, needs advance planning
When and how will this storm hit visualization?

- they're ahead of us
  - they have some paper retractions
  - 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
    - are younger, might our power hierarchies thus be less entrenched?...
    - they are higher profile
  - we don't have rich fabric of blogs as major drivers of discussion
  - 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/