December 7, 2021 Joint Methods & CI Working Group Call

Methods/CI November 9, 2021 Notes HERE

Attendees: Carl Boettiger, Libby Mohr, Jody Peters, Kelly Heilman, Hassan Moustahfid, Ben Toh, David LeBauer, Rob Kooper
Regrets: Cee Nell, Quinn Thomas

Agenda/Notes:

1. Poll for January to May calls: http://whenisgood.net/ixtw9n9

2. Visualization/Decision Support Tools, User Interface Task View Updates and Next Steps
    a. Updates on visuals for the Static Viz section
    b. Go over Section 1 - Core Principle with Jessica if she is available.
    c. Section 12-14 - how to incorporate into Section 1
    d. Update from Cee:
        i. I did populate the gif/animation section somewhat with R- and CLI-based tools that can be used to make animations, as well as some guidance about how to make them accessible. I haven't gotten to it yet, but after talking to python users, I would like to add ImageIO to that list as a complement to the R-based packages.
        ii. I left this section fairly barebones because I wasn't sure how in depth we wanted to go, or if the goal was to provide a list of tools and leave it at that. If we wanted to go more in depth, I listed some applications where animation is useful - HOPs, timeseries, movement data, etc. **We could include an example using one of these, maybe using gganimate since it is very simple to use. Previously there was mention of locating datasets to use for creating examples, but I am not sure what the outcome of that was.**
        iii. I left a question re: **is the section specific to gifs and video, or is it animation in general? One piece that may be missing here are javascript-based animations, using libraries like gsap and D3.** This may be more than we want to get into here, but D3/javascript are definitely worth mentioning for interactive visualizations.
    e. Hassan's text could be moved to the beginning for the pitch/introduction - we are making visualizations but for what. Think about whether to expand to apply to end users/stakeholders
    f. Dealing with large data Section 10 - addresses a different problem. It deals with optimizing data. Think it fits better with the Data Ingest, Cleaning, Management Task View

   i. Ben's R Shiny Seminar
   ii. Making heat maps
   iii. Google maps and zooming in
   iv. Vismet - has hpc backend a
   v. Panopoli - visualizes images from a server and does automatic downscaling so doesn't bring in all the data
   vi. Use Big data instead of Large data?
      1. Using GPU systems are growing faster than CPU. Have many cores of CPU, but need GPU to do better visualization
      2. Big data is defined in the computing resources available. If plot takes 5 minutes to render, then want to think of a better way to get the information. If it takes 5 minutes to render then probably will show a cloud of points/data that don't provide details.
      3. Research development setting rather than the operations
      4. Big data - talking about compute power
g. Examples Update
   i. Libby and Matt worked to make examples for all the static plots listed in Section 3. That is in this GitHub repo and saved as an Rmd that can be transferred over easily to the Task View bookdown once we get the Google doc finalized
   ii. Libby would be willing to do the animated examples too. Jody will put her in touch with Cee
   iii. Viz examples were the target data from the Forecast Challenge or the Palmer Penguins Package data.
   iv. Will be good to have Static, Animation, and Interaction section. Headings are now updated
   v. Rob will ask about any new libraries for Animation - wonder if there are new libraries other than D3. D3 is hard to work with. People have written large wrappers around it. Rob will ask what they will recommend. Update from 12-8-21: Rob's colleagues pointed him towards this post about D3 which includes a pros/cons list of D3, matplotlib, seaborn, plotly, and ggplot2
   vi. Go with Cee's plan to mention javascript and libraries but don't go into too much detail
h. Tableau - need to have license for this. Do we want to avoid and stick to open resources
   i. open source alternative https://superset.apache.org/
i. ArcGIS is the same way
j. Do we want to mention at the beginning that we are focusing on open source only and say there are commercial options.
   i. Compromise by listing both open source and the ones that require license. Be clear what requires license and the alternative open source

resources and make a note that not all features described with the licensed version wudl be available in the open source option.
    ii. ArcGIS and QGIS example - we are talking more about the tool
k. Leave out Matlab. But do we want to talk about excel?
    i. Don't need to get into presentations, but thinking about diagrams
    ii. Do we want to include something about flowcharts
    iii. Excel - Rob uses all the time to make a quick dirty plot of data
    iv. Spreadsheets - provide basic plotting functionality (spreadsheets then apply to Excel, Google sheets, Libre Office, etc).
l. Interactive Spatial Visualization vs Animated
    i. Animated is static but moving. Interactive lets users change the content and make decisions.
m. Merge Statics Spatial With the Static Section
n. Merge the Interactive Spatial with the Interactive Section
o. Make the Uncertainty and Big Data there own sections at the end
p. Create a Geospatial Section Under the How to Make Visuals Header
q. Jody will rearrange the document to follow the outline below (the top two sections alone). On the next call look at the Linking Visualizations to Science and Policy sections
    i. New Outline Headers
        1. How to Make Visuals
            a. Static
            b. Animated
            c. Interactive
            d. Geospatial
                i. Static
                ii. Interactive
        2. Uncertainty Visualization
        3. Dealing with Big Data
        4. Linking Visualizations to Science and Policy
            a. User Interfaces and Dashboards
            b. Environmental Decision Support
            c. Connecting Dashboards to models/forecasts

3. Uncertainty Quantification & Propagation, Modeling & Stats and Workflow Task Views to use as Reference: https://projects.ecoforecast.org/taskviews/

4. Jody is leaving in as a placeholder. This would be good to discuss when Mike is able to join the call. Any more gaps people thought of in response to Mike's prompt for the November call?
    a. Mike's prompt: Discuss next steps beyond the current task views (2 done, 1 in progress, 1 on deck). In particular, I think we've completed enough of the task

views that we could have some useful discussions about identifying gaps and bottlenecks in current methods and tools and ways we might think about moving forward on identifying and addressing those (e.g. surveys, workshops, hackathons, group projects, virtual panels / webinars, etc). This is also motivated by some recent interagency (NOAA, NASA, CDC) roundtables that EFI has been sponsoring where standards and community CI are emerging as priority items for the federal agencies (caveat, that doesn't mean there's any funds). It is also relevant to the next round of the EFI NEON challenge and the upcoming EFI2022 "all-hands" conference

5. Data Ingest, Cleaning, Management
    a. Placeholder until we are further along with the other Task Views or have an identified leader for this

6. NEON Ecological Forecast Challenge CI Update