Comments on existing study implementation:

* Could not expect to get statistically significant results with the current study design, needs to be altered to make it possible to obtain statistically significant results
* We should remove the more complex graphs
  + Stream graphs & starfish graphs were too difficult to understand and make judgements about. They didn’t really help with the assessment of uncertainty visualization
* “throught” should be spelled “throughout”
* You need to be clear with if you use labels. If the user has to guess what they mean then that confounds the experiment.
  + instead put something like “Question-34: Iraq (IRQ), Indonesia (IDN), Pakistan (PKS) have same uncertainties in marked areas”
* Instead of having hints, you should start each the section with a training session that explains the charts and how the uncertainty will be represented. And shows how the user makes choices.
* The size of the font in the usage chart may be too small for some to read. - Done
* We need to add counterbalancing and more questions for statistical significance
* We need to map the data values to colors (using Viridis color map like VSUP uses)
* Need to have a legend for both the data colors and the amount chromatic aberration
* Need to test it on different window sizes if someone uses a smaller screen
* Will need to try to calibrate how much space is needed for CA before user study. Any perceptual basis for size of CA?

Study Design:

* We need to simplify the study design, given that we found more closely related prior work and some of our experimental designs don’t work as well as others. If we include too many conditions, we will need a larger study with more participants. We should try to follow the user study methodology of the prior work too, specifically [Correll 2018]
* Study overview:
  + a comparison of Chromatic Aberration (CA) with prior work (VSUP) [Correll 2018] with commonly used graphs (bubble chart, grid chart) using real-world (Covid 19) data
  + a within subject experiment:
    - “Having fewer conditions also afforded a within-subjects design that controlled for the variation in the interpersonal differences” [Correll 2018]
  + We have continuous (not binary) data
* Two Independent variables:
  + Uncertainty visualization technique (CA, VSUP)
  + Visualization type (bubble chart, grid chart – you sometimes called this a “usage chart” or “impact chart”)
* Hypotheses (need to check over):
  + Using the existing uncertainty visualization method (VSUP) yields poorer accuracy than Chromatic Aberration
  + Using the existing uncertainty visualization method (VSUP) requires more time to complete identification tasks than Chromatic Aberration
  + Users do not prefer the existing uncertainty visualization method (VSUP) over Chromatic Aberration
  + VSUP eliminates higher uncertainties by considering only single option.
* User Tasks:
  + “Participants should consider both value and uncertainty” at the same time [Correll 2018]
  + Like in [Correll 2018], users will be asked to select elements in the visualization that have a combination of *X data value* and *Y level of uncertainty*
    - It is possible that there are more then one example of each combination
  + Each user will be given 8 of these selection tasks for each of the 4 conditions, for a total of 32 selection tasks
* Counter Balancing
  + We will use a factorial design with two independent variables, which creates 4 conditions:
    - A: CA+bubble = CA into 4 groups (0-100), values (8 groups)
    - B: CA+grid = CA + value represented by VSUP color
    - C: VSUP+bubble = all vsup colors in bubble charts instead of CA
    - D: VSUP+grid = already there in prev implementation
  + need to use counter balancing to counter act any learning effects within each subject
  + use 4 x 4 balances latin squares to counter balance:
    - <https://www.yorku.ca/mack/RN-Counterbalancing.html>

Shape

Description automatically generated with medium confidence

* How many participants?
  + We are aiming for 32 participants in the study
  + 6/8? users in each of the 4 Latin Squares counter balancing sequences
* Introductory training
  + need a short training explanation before each of the four conditions of the user study
* Screening tests:
  + Need to test for Color Blindness
    - “present participants with a set of Ishihara plates [Hardy 1945], and exclude those that misidentified values or who self-reported as having a color vision deficiency” [Correll 2018]
* Measures to collect:
  + Accuracy of the user selections
  + Completion time for the condition (8 trials per condition)
  + Need to give the user a questionnaires after each condition of the user study
    - NASA-TLX questionnaire [Sandra 1988], and
    - SUS [Brooke, 1995]
  + Post session questionnaire
    - At the end of all conditions, obtain information about the users background (age, gender, geographic location, experience level with visualizations)
  + Record computer details (computer, webpage size, etc)
* Analysis of data (need to check over):
  + ANOVA or Paired t-tests
  + For the NASA-TLX and SUS data
    - First use Shapiro-Wilk normality test [Shapiro 1965]
    - If data has normal distribution:
      * Use ANOVA test [Sauro 2016]
    - If data does NOT have a normal distribution:
      * Use Kruskal-Wallis test [Kruskal 1952]
* Potential Risks to the user:
  + note risk of user being upset by data such as death counts, etc. and add warning to user in study solicitation
  + Note potential risk of eye strain in user study due to CA
* Quantization
  + For a more direct comparison with [Correll 2018], we quantize the data values into 8 levels and the uncertainty into 4 levels for our technique (CA)
* Color Map
  + [Correll 2018] uses Viridis [Stéfan 2015] and we also use it as well for data values to facilitate direct comparison
* Legend style
  + For VSUP [Correll 2018] we use their quantized wedge-shaped legend:

Chart, funnel chart, sunburst chart

Description automatically generated

* + For Chromatic Aberration, **in the case of the grid chart**, we require two legends, one for the amount of uncertainty and the other for the data to bubble size,:



A picture containing application

Description automatically generatedA picture containing application

Description automatically generatedA picture containing application

Description automatically generatedA picture containing application

Description automatically generated

* + For Chromatic Aberration, **in the case of the bubble chart**, we require two legends, one for the amount of uncertainty and the other for the data to bubble size:

<need to add bubble legend size>

A picture containing application

Description automatically generatedA picture containing application

Description automatically generatedA picture containing application

Description automatically generatedA picture containing application

Description automatically generated

* Size of Charts:
  + We should reduce the size of the charts, closer to those used in [Correll 2018] which was 5x5 = 25 elements
  + Select randomly from the set of countries, but exclude the outliers (the largest and smallest data values)
  + Bubble Chart
    - Use 25 bubbles, excluding (outliers)
  + Grid Chart
    - How many elements? 5x5 grid elements

References:

[Brooke, 1995] John Brooke. Sus: A quick and dirty usability scale. Usability Eval. Ind., 189, 11 1995.

[Correll 2018] Michael Correll, Dominik Moritz, and Jeffrey Heer. 2018. Value-Suppressing Uncertainty Palettes. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, Paper 642, 1–11.

[Kruskal 1952] William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260):583–621, 1952.

[Hardy 1945] LeGrand H Hardy, Gertrude Rand, and M Catherine Rittler. 1945. Tests for the detection and analysis of color-blindness. I. The Ishihara test: an evaluation. JOSA 35, 4 (1945), 268–275.

[Sandra 1988] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, Human Mental Workload, volume 52 of Advances in Psychology, pages 139–183. North-Holland, 1988.

[Sauro 2016] Jeff Sauro and James R. Lewis. Quantifying the User Experience: Practical Statistics for User Research. 2nd edition. Morgan Kaufmann, 2016.

[Shapiro 1965] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). Biometrika, 52(3-4):591–611, dec 1965.

[Stéfan 2015] Stéfan van der Walt and Nathaniel Smith. 2015. Mpl colormaps. https://bids.github.io/colormap/, (2015).