**Project 1: Team A: Maria-Jose, Brian, Natasha**

1. **Data Cleaning and Variable Creation**

Data cleaning was presented clearly. Reviewing code, several steps were similar and reported metrisl that included/confounded by another variables. For example, some tables included RTs for congruent trials that averaged across correct and incorrect trials and some summarized correct trials for congruent and incongruent trials. Summaries could have been combined with grouping that split the data so that they were not confounded, for example, the 2 x 2 complement of data grouping by is\_congruent and is\_accurate. This would provide both simpler code and would provide cleaner estimates for the design.

Overall, clear communication on data cleaning and variable creation. One limitation was inclusion of RTs for inaccurate trials. In order to obtain appropriate estimates, remove RTs for incorrect trials before averaging. One typo on data cleaning side; looks like copy and paste error.

1. **Data Summary: Story Telling with Tables or Numbers**

Summary metrics communicated clearly; RT and SIE seem to be confounded by accuracy; remove rts for inaccurate trials. Tables and plots presented important details clearly. Good use of confidence intervals; shows no overlap. Sample was communicated well. Based on rmd files, group-level SIE appears to be calculated by grouping by id and summarizing and then summarizing without id. Although you will obtain an average this way, you have to manage multiple data frames. Alternatively, creation of respondent level sie could be accomplished with tidyr::pivot\_wider() to pivot the data frame wide and mutate a difference score for the congruent and incongruent rts and accuracy rates. For example:

tidyr::pivot\_wider(.,

names\_from = trialtype,

values\_from = rt,

) %>%

mutate(., SIE = Incongruent - Congruent)

A data frame then contains the respondent-level SIE which can be averaged for group-level SIE.

The team went above and beyond in creative ways by creating novel metrics (although the streak/runs would be correlated with accuracy) and functions as well as creating nice plots corresponding to spectrum and distance metrics. They also discuss the implications of colors and interference across the colors.

1. **Data Visualizations: Story Telling with Pictures**

Performance metrics communicated appropriately; excellent use of plots; there was a comprehensive story about the data which was engaging for the audience

1. **Presentation Characteristics**

Group was clear in their presentation, thoughtful in their analysis, and prepared for telling story. Organization was logical and appropriate. Team took audience on a nice exploratory journey of the data showing nice “above and beyond” characteristics.

1. **Weekly Progress Report Logs**

Weekly progress logs were not always submitted or submitted on time weekly as indicated in assignment. As such there was lack of clarity on what personal goals were or whether they were achieved in a weekly manner.

1. **Reproducibility**

Rmd files allowed for a clear and reproducible product. File was well documented and easy to follow in addition to the slides.

**Project 1: Team Kayleah, Rebecca, Janey**

1. **Data Cleaning and Variable Creation**

Data cleaning was presented clearly. The team did remove inaccurate trials from the RT data, which is essential to obtaining clear estimates of performance. Reviewing code, this is all clear and easy to follow. Lagging can create problems though you worked around it here. I do recommend either pulling out the separate vectors in order to create a new data frame and then perform difference scores or using pivot\_wider() on the existing data frame. Cleaning did not appear to include removal of long response trials, something Rebecca astutely mentioned in class during the first class working session. I provided some suggestions using z scores or other trimming but long responses are retained in the data. Such responses greatly affect mean metrics (less so medians – a different solution). I’m unsure why the team didn’t pursue this further even after seeing the plots with an extreme outlier. Compared with other teams, you can see the differences in SIE based on the outlier data. Overall, clear communication on data cleaning and variable creation.

1. **Data Summary: Story Telling with Tables or Numbers**

Some data visuals on “mean” accuracy rate and response time on slides 14 and 15 appear to be out of bounds. I’m unsure how mean rts would be in the 200,000 range for some participants. RTs were in milliseconds, so average should be around 2k not 200k. Accuracy also is reported above 150 so I’m confused by the slides.

Summary metrics communicated clearly; RT and SIE seem to be confounded by accuracy; remove rts for inaccurate trials. Tables and plots presented important details clearly. Good use of confidence intervals; shows no overlap. Sample was communicated well. Based on rmd files, group-level SIE appears to be calculated by grouping by id and summarizing and then summarizing without id. Although you will obtain an average this way, you have to manage multiple data frames. Alternatively, creation of respondent level sie could be accomplished with tidyr::pivot\_wider() to pivot the data frame wide and mutate a difference score for the congruent and incongruent rts and accuracy rates. For example:

tidyr::pivot\_wider(.,

names\_from = trialtype,

values\_from = rt,

) %>%

mutate(., SIE = Incongruent - Congruent)

A data frame then contains the respondent-level SIE which can be averaged for group-level SIE.

1. **Data Visualizations: Story Telling with Pictures**

Performance metrics communicated appropriately; good use of plots; there was a comprehensive story about the data. The team went above and beyond in creative ways by creating novel plots of to communicate elements of the data.

1. **Presentation Characteristics**

Slide scheme was thoughtfully choses. Group was clear in their presentation, thoughtful in their analyses, and prepared for telling story. Some slides are questionable. The code sections were exceptionally clear and nicely presented in a way that was easy to follow. Organization was logical and appropriate as a story about data.

1. **Weekly Progress Report Logs**

Weekly progress logs were not always submitted or submitted on time weekly as indicated in assignment. As such there was lack of clarity on what personal goals were or whether they were achieved in a weekly manner.

1. **Reproducibility**

Rmd files allowed for a clear and reproducible product. File was well documented and easy to follow in addition to the slides.

**Project 1: Team Ming, Jenna, Catherine**

1. **Data Cleaning and Variable Creation**

Data cleaning was presented clearly. Good use of Tukey’s hinges, bringing in new methods to the project. I commend you for using an approach to take care of outliers this way. However, keep in mind that you arbitrarily removed data from some participants more than others because of where they fell in the data range. Consider doing this metric at the participant level. Overall, clear communication on data cleaning and variable creation.

1. **Data Summary: Story Telling with Tables or Numbers**

Summary metrics communicated clearly; tables showed differences in across participants, though a plot might have been a simpler and more consumable approach.

Alternatively, creation of respondent level sie could be accomplished with tidyr::pivot\_wider() to pivot the data frame wide and mutate a difference score for the congruent and incongruent rts and accuracy rates. For example:

tidyr::pivot\_wider(.,

names\_from = trialtype,

values\_from = rt,

) %>%

mutate(., SIE = Incongruent - Congruent)

A data frame then contains the respondent-level SIE which can be averaged for group-level SIE.

The team went above and beyond in creative ways by creating novel metrics (although the streak/runs would be correlated with accuracy) and functions as well as creating nice plots corresponding to spectrum and distance metrics. They also discuss the implications of colors and interference across the colors.

1. **Data Visualizations: Story Telling with Pictures**

Performance metrics communicated appropriately; excellent use of plots; there was a comprehensive story about the data which was engaging for the audience

1. **Presentation Characteristics**

Group was clear in their presentation, thoughtful in their analysis, and prepared for telling story. Organization was logical and appropriate. Team took audience on a nice exploratory journey of the data showing nice “above and beyond” characteristics.

1. **Weekly Progress Report Logs**

Weekly progress logs were not always submitted or submitted on time weekly as indicated in assignment. As such there was lack of clarity on what personal goals were or whether they were achieved in a weekly manner.

1. **Reproducibility**

Rmd files allowed for a clear and reproducible product. File was well documented and easy to follow in addition to the slides.