**Project Instructions**

Dear Hackathon Team,

We’re excited to have you with us!

This document contains all the information you’ll need for your project. Please feel free to reach out with any questions, suggestions, or ideas.

Thank you for your help, and good luck!

Reut and Noa

**Project Title**

Data Analysis for Longitudinal Research: Imputation and Visualization

**Study Description**

Irritability (i.e., proneness to anger) and aggressive behavior are of the most common reasons for referrals to mental health treatment. It is a common and consequential domain of psychopathology in youth and adults and associated with functional impairment and increased risk for depression, anxiety, and suicidality. Youth with severe, impairing, and developmentally-atypical irritability exhibit sustained angry mood and frequent and intense aggressive outbursts. Although pediatric irritability is prevalent and impairing, little is known about its mechanisms and treatment. This study posits core mechanism for irritability, particularly impaired inhibitory control as a marker surrounded the symptomatology of

irritability that may serve as a potential target for intervention. Inhibitory control reflects the ability to suppress behaviors that are maladaptive or incompatible with current goals (e.g., the ability to stop an automatic or prepotent behavior). Behaviorally, irritability manifests as aggressive behaviors such as yelling, hitting, throwing items, which can be conceptualized as an inability to inhibit a response. Therefore, it is plausible that impairments in inhibitory control, manifesting as difficulties in stopping impulses, would result in aggressive behaviors in those youth with significant irritability. Yet, deficits in inhibitory control have yet to be tested in irritability. The heart of this work aims to address this gap exploring the dynamics between inhibitory control and irritability, first- in an analog sample of adults. Alterations in inhibitory control are theorized to be central to irritability. We will also examine the role of elicit arousal in the association between impaired inhibitory control and clinical symptoms.

**Short Description of the Project**

This project focuses on reconstructing datasets and developing group-level visualizations from a six-day longitudinal study. Due to natural limitations (e.g., human error, low-quality responses), there is a small proportion of missing data that needs to be imputed. First, students will explore and apply imputation techniques, taking into account both individual trends and each participant’s position relative to the full sample. Second, students will produce publication-ready visualizations of individual and group-level trajectories over time.

**Sample Details**

This is a subsample from a larger-scale study, consisting of 50/51 participants who first visited our lab for a “baseline” session, followed by five days of remote “home” sessions.

During the baseline session, participants completed a series of questionnaires and performed two cognitive tasks while wearing an Empatica watch for physiological measurements. In the home sessions, they performed only the tasks using their personal devices.

**Types of data**

* **Cognitive Data:**  
  Collected from two tasks — *Stop-it* and *Tap-it*. Each participant is expected to have data from 6 sessions per task.
* *Stop-it* indices: SSRT, Post-Error Efficiency
* *Tap-it* indices: D’context, A-cue bias, PBI\_composite  
   Refer to the code\_booksheet for variable descriptions.
* **Physiological Data:**Collected during the lab session and divided into four stages: baseline (pre-tasks), first task, second task, and recovery (post-tasks).

Variables include: HR, HRV, BVP, EDA, Temperature, and ACC. Main focus: HRV and HR.

* **Clinical Questionnaires:**

ARI, AQ-BUSS, RPQ (and also BIS-BAS and DAR-5).

**Your Task**

1. Explore and apply imputation techniques to the cognitive data.

Some participants are missing one or more of the six time points, either due to skipping a session or being excluded for low task performance.

Your goal is to find an appropriate method for reconstructing these missing values, taking into account: (a) The general trend across the full sample; (b) The individual's own performance trajectory across sessions. Feel free to explore imputation techniques and justify your choice.

1. Create visualizations that will represent the longitudinal structure of the data.  
   Your visualizations should clearly communicate individual and group trends across time. Here are two suggested approaches to get you started:

* Graph A - Plot each participant’s scores (e.g., SSRT) over the six time points. *Note:* Your code should be modular and flexible to allow plotting of different measures easily - one graph per measure.
* Graph B - Split the sample into high vs. low fluctuation groups (based on standard deviation median across sessions), and compare them on other variables of interest - such as clinical questionnaire scores or baseline physiological measures. *Same note applies here too* 🙂
* Have your own idea? Great! We’d love to see any creative visualizations that help convey meaningful patterns in the data.