figure_assignment

Sihan Yang

Original Paper

- link
  - paper
  - data
- introduction
- main figure
  - explanation
  - strengths
  - weakness

Reproduction of Main Figures

Dara Preprocessing

While the authors have provided the preprocessed data for further analysis in R, here we started with the raw data to examine whether their data are processed correctly. (Since the raw empathy inference response are not provided, we will use directly used the accuracy metrics (Pearson correlation score and RMSE) provided by the authors).

```
library(tidyverse)

data_path <- 'data/N709_EmpathicAccuracyTaskDat.csv'
loaded <- read.csv(data_path)
```

```
# get rid of redundant columns
cleared_loaded <- loaded |>
  select(obsID, movie, gender, age, obsRace, Ideology,
         stimID, stimRace, storyteller_label_attn_check, cond, visit, EAcorr, EArmse,
         compassion, RemoveMisLabeled1)

# remove those with missing data
cleared_loaded <- cleared_loaded |>
  filter(!RemoveMisLabeled1)
cleared_loaded <- cleared_loaded |> select(-RemoveMisLabeled1)
```

Separate the subject's empathy data and other data

```
subject_info <- cleared_loaded |>
  select(obsID, movie, gender, age, obsRace, Ideology) |>
  distinct(obsID, .keep_all = TRUE)

survey_data <- cleared_loaded |>
  select(obsID, stimID, storyteller_label_attn_check, cond, visit, EAcorr, EArmse, compassion

head(subject_info)
```

```
  obsID      movie gender age obsRace     Ideology
1   271  Moneyball  woman  29   White      Liberal
2   274 Concussion    man  50   White      Liberal
3   276  Moneyball    man  65   White   OtherRight
4   284 Just Mercy  woman  49   White      Liberal
5   286 Concussion  woman  64   White      Liberal
6   289 Just Mercy    man  56   Asian Conservative
```

Compute each subjects' average inference accuracy and compassion in two visits

```
empathy_collapsed <- survey_data |>
  group_by(obsID, visit, storyteller_label_attn_check) |>
  summarize (
    compassion = mean(compassion, na.rm=TRUE),
    EAcorr=mean(EAcorr, na.rm=TRUE),
    EArmse=mean(EArmse, na.rm=TRUE)
  ) |> ungroup()
```

`summarise()` has grouped output by 'obsID', 'visit'. You can override using the `.groups` argument.

```
# leave out those who do not have both types of story-teller in both visits
# i.e. visit 1/2 x story-telley prisonser/student
empathy_collapsed <- empathy_collapsed |>
  group_by(obsID) |>
  filter(n() == 4)  |>
  ungroup()

nrow(empathy_collapsed)
```

```
[1] 2672
```

```
head(empathy_collapsed)
```

```
# A tibble: 6 x 6
  obsID visit storyteller_label_attn_check compassion  EAcorr EArmse
  <int> <int> <chr>                             <dbl>   <dbl>  <dbl>
1   271     1 Formerly Incarcerated              86.8  0.626    17.7
2   271     1 Student                            94.2  0.354    25.2
3   271     2 Formerly Incarcerated              88.3  0.138    29.2
4   271     2 Student                            92.8  0.461    26.2
5   273     1 Formerly Incarcerated              56.7  0.645    17.0
6   273     1 Student                            43   -0.0404   29.2
```

Fitting

```
[1] 4
```

Visualization