Data collected and combined into a csv file, to be loaded into an R dataframe.

Dataframe Columns:

1) **id:** a unique, anonymized string ID generated by ANONYMIZED referring to an individual attendee

2) **regionId:** a unique, anonymized string ID generated by ANONYMIZED referring to an event location

3) **startsAt:** a string representation of the date of an event

4) **gender:** voluntarily self-reported by attendee (male, female, nonBinary, unknown)

5) **ethnicity:** voluntarily self-reported by attendee (black, white, asian, latinx, other, unknown)

6) **codingInterest:** voluntarily self-reported by attendee (more, low, none, unknown)

7) **presented:** voluntarily self-reported by attendee (yes, no, unknown)

8) **newRegionId:** integer in range 0-27, assigned by researchers to replace the 28 original unique regionId string identifiers

9) **newEventId:** integer in range 0-6, assigned by researchers to replace the 7 original unique eventID string identifiers (dates)

10) **outcome:** voluntarily self-reported persistence in coding by attendee, via follow-up survey (yes, no, unknown)

11) **my_gender:** calculated value -- the number of attendees of same self-reported gender at the same date/region event as current attendee

12) **my_ethnicity:** calculated value -- the number of attendees of same self-reported ethnicity at the same date/region event as current attendee

13) **num_participants:** calculated value -- the total number of attendees at the same date/region event as current attendee

14) **gender_pct:** calculated from my_gender/num_participants

15) **ethnicity_pct:** calculated from my_ethnicity/num_participants