Verification Report: data_clean

Ellis Hughes

2021/03/24

Description

This is the verification report for the `data_clean` folder of the `correlates_reporting` project for CoVPN.

In this document, the output of `make_data_proc.R` is compared against the output of `process_data_raw.R`. The two scripts use the same base mock data. The two datasets generated by the two scripts will be compared with each other to confirm they contain the same values.

`process_data_raw.R` was independently double programmed based on the specifications found in `dat_clean_specifications.pdf`. This script outputs its processing to `data_clean/verification/verification_output/data_clean_verification_output.csv`

The file `data_clean/verification/verification_input/practice_data.csv` was provided by the original programmer to the tester for verification purposes of `make_data_proc.R`. Its md5 hash is 5974e0af96a582a930e3190795862094.

Load Data

```
original_data <- read_csv(
  here("data_clean/verification/verification_input","practice_data.csv"),
  guess_max = 30000
  )

verification_data <- read_csv(
  here("data_clean/verification/verification_output/data_clean_verification_output.csv"),
  guess_max = 30000
  )
```

Verification

```
data_clean_comparison <- compare_datasets(
  cols =  colnames(original_data), index = "Ptid",
  ds1 = original_data, ds2 = verification_data
)
```

```
## There are 0 mismatched fields of 86.
```

Output of make_data_proc.R is equivalent to the output of process_data_raw.R, and passes verification.

Signatures

Role	Name	Signature	Date
Tester	Ellis Hughes		