Cleaning depression data

Robin Donatello

February 1, 2016

Purpose

The purpose of this file is to clean and recode the depression data set. Each time this file is run it will write a new data set to the hard drive with the current date.

Each time a new recode or edit is added a note should be written about when and why this edit was made, and this file should be compiled.

Import the raw data

```
depress <- read.table("C:/GitHub/MATH456/data/Depress.txt", sep="\t", header=TRUE)
```

Create factor variables.

This section is where catgorical variables are defined as factors and have labels and ordering applied.

- 01-31-16 redefine `MARITAL` and `EDUCAT` as factor variables.

```
depress$MARITAL <- factor(depress$MARITAL,
                          labels = c("Never Married", "Married", "Divorced", "Separated", "Widowed"))
depress$EDUCAT <- factor(depress$EDUCAT,
                          levels = c("<HS", "Some HS", "HS Grad", "Some college", "BS", "MS", "PhD"),
                          labels = c("<HS", "Some HS", "HS Grad", "Some college", "BS", "MS", "PhD"))
```

Edits and recodes

This section is for real changes to the data. Non-trivial edits should include a justification.

- 01-31-16. Fix a typo in `AGE`.

```
depress$AGE[depress$AGE==9] <- 19
```

- 01-31-16. Two values for religion were out of range, so they have been set to missing.

```
depress$RELIG <- ifelse(depress$RELIG == 6, NA, depress$RELIG)
```

Save the cleaned data set with todays date.

The `sys.date()` function takes the current date from your computer. The value is then formatted nicely for human consumption and added (pasted) to the file name before written to the working directory as a new text file.

```
date <- format(Sys.Date(), "%m%d%y")
filename <- paste("C:/GitHub/MATH456/data/Depress_", date, ".txt", sep="")
write.table(depress, filename, sep="\t", row.names=FALSE)
```