

Unified App - Data Preparation

Justin Baldwin

July 24, 2017

The goal of this document is to produce a clean data frame of all data, one chapter per app upon each knit. Last, it should also produce a table of all topics that are ordered by category and perform the column selection.

1. Demographics

1.1. Data Cleaning

```
setwd("C:/Users/Justin/Desktop/Research/SEIGMA/unified")
Dem_data <- read.csv(file="../demographics/demographics/demodata.csv")
colnames(Dem_data)[12:37] <- c("Age_under_5_Pct", "Margin_Error_under_5_Pct", "Age_5-9_Pct", "Margin_Error_5-9_Pct",
                              "Age_10-14_Pct", "Margin_Error_10-14_Pct", "Age_15-19_Pct", "Margin_Error_15-19_Pct",
                              "Age_20-24_Pct", "Margin_Error_20-24_Pct", "Age_25-34_Pct", "Margin_Error_25-34_Pct",
                              "Age_35-44_Pct", "Margin_Error_35-44_Pct", "Age_45-54_Pct", "Margin_Error_45-54_Pct",
                              "Age_55-59_Pct", "Margin_Error_55-59_Pct", "Age_60-64_Pct", "Margin_Error_60-64_Pct",
                              "Age_65-74_Pct", "Margin_Error_65-74_Pct", "Age_75-84_Pct", "Margin_Error_75-84_Pct",
                              "Age_85+Pct", "Margin_Error_85+_Pct")
colnames(Dem_data)[54:59] <- c("Age_under_20_Pct_plot", "Age_20-34_Pct_plot", "Age_35-54_Pct_plot", "Age_55-64_Pct_plot",
                              "Age_65-74_Pct_plot", "Age_75+Pct_plot")
Dem_data$Year <- as.numeric(substr(Dem_data$Five_Year_Range, 1, 4))+2
write.csv(Dem_data, "unified.csv", row.names = F)
```

1.2. Topics table

```
demvars <- names(Dem_data)[6:59][!grep(names(Dem_data)[6:59], pattern = "Error")]

topics <- data.frame("App"="Demographics",
                    "Variables"=demvars,
                    "Labels"=gsub(gsub(demvars, pattern = "_", replacement = " "),
                                   pattern = "Pct", replacement = "%"),
                    "x"=c(-2:0, 2:14, 16:23, 25:30),
                    "y"=c(rep(4,3), rep(3,13), rep(2,8), rep(1,6))
                    )
write.csv(topics, "topics.csv", row.names = F)
```