Assignment 4 - Tidy Data Article

Kevin Chan

February 28 2018

Converting Table 4 to Table 6.

```
library(foreign)
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 3.4.3
```

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.4.3
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.4.3
```

```
source("xtable.r")
```

```
## Warning: package 'xtable' was built under R version 3.4.3
```

```
# Data from http://pewforum.org/Datasets/Dataset-Download.aspx

# Load data ---------------------------------------------------------------

pew <- read.spss("pew.sav")
```

```
## re-encoding from CP1252
```

```
## Warning in read.spss("pew.sav"): Undeclared level(s) 2, 3, 4, 9 added in
## variable: density3
```

```
## Warning in read.spss("pew.sav"): Duplicated levels in factor denom:
## Electronic ministries
```

```
## Warning in read.spss("pew.sav"): Undeclared level(s) 1, 2, 3, 4, 5, 6, 7,
## 8, 9, 10, 11, 12, 14, 16, 23, 33 added in variable: children
```

```
## Warning in read.spss("pew.sav"): Undeclared level(s) 18, 19, 20, 21, 22,
## 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41,
## 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
## 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79,
## 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96 added in
## variable: age
```

```
pew <- as.data.frame(pew)


religion <- pew[c("q16", "reltrad", "income")]
religion$reltrad <- as.character(religion$reltrad)
religion$reltrad <- str_replace(religion$reltrad, " Churches", "")
religion$reltrad <- str_replace(religion$reltrad, " Protestant", " Prot")
religion$reltrad[religion$q16 == " Atheist (do not believe in God) "] <- "Atheist"
religion$reltrad[religion$q16 == " Agnostic (not sure if there is a God) "] <- "Agnostic"
```

```
religion$reltrad <- str_trim(religion$reltrad)
religion$reltrad <- str_replace_all(religion$reltrad, " \\(.*?\\)", "")

religion$income <- c("Less than $10,000" = "<$10k",
  "10 to under $20,000" = "$10-20k",
  "20 to under $30,000" = "$20-30k",
  "30 to under $40,000" = "$30-40k",
  "40 to under $50,000" = "$40-50k",
  "50 to under $75,000" = "$50-75k",
  "75 to under $100,000" = "$75-100k",
  "100 to under $150,000" = "$100-150k",
  "$150,000 or more" = ">150k",
  "Don't know/Refused (VOL)" = "Don't know/refused")[religion$income]

religion$income <- factor(religion$income, levels = c("<$10k", "$10-20k", "$20-30k", "$30-40k", "$40-50
  "$75-100k", "$100-150k", ">150k", "Don't know/refused"))

table6 <- count(religion, c("reltrad", "income"))
names(table6)[1] <- "religion"

xtable(table6[1:10, ], file = "pew-clean.tex")
View(table6)

# Convert into the form in which I originally saw it -------------------------

table4 <- dcast(table6, religion ~ income)
```

```
## Using freq as value column: use value.var to override.
```

```
xtable(table4[1:10, 1:7], file = "table4.tex")
View(table4)
```

Converting Table 7 to Table 8.

```
options(stringsAsFactors = FALSE)
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.4.3
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:plyr':
##
##     here
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(reshape2)
library(stringr)
library(plyr)
source("xtable.r")
```

```
table7 <- read.csv("billboard.csv")
table7 <- table7[, c("year", "artist.inverted", "track", "time", "date.entered", "x1st.week", "x2nd.week
names(table7)[2] <- "artist"

table7$artist <- iconv(table7$artist, "MAC", "ASCII//translit")
table7$track <- str_replace(table7$track, " \\(.*?\\)", "")
names(table7)[-(1:5)] <- str_c("wk", 1:76)
table7 <- arrange(table7, year, artist, track)

long_name <- nchar(table7$track) > 20
table7$track[long_name] <- paste0(substr(table7$track[long_name], 0, 20), "...")


xtable(table7[c(1:3, 6:10), 1:8], "table7.tex")
View(table7)

clean <- melt(table7, id = 1:5, na.rm = T)
clean$week <- as.integer(str_replace_all(clean$variable, "[^0-9]+", ""))
clean$variable <- NULL

clean$date.entered <- ymd(clean$date.entered)
clean$date <- clean$date.entered + weeks(clean$week - 1)
clean$date.entered <- NULL
clean <- rename(clean, c("value" = "rank"))
clean <- arrange(clean, year, artist, track, time, week)
clean <- clean[c("year", "artist", "time", "track", "date", "week", "rank")]

table8 <- mutate(clean,
  date = as.character(date))
xtable(table8[1:15, ], "table8.tex")
View(table8)
```