

TablesHWK

Anthea Yichen Li

March 2, 2018

translate table 4 to table 6

```
library(foreign)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.2.1 --
## √ ggplot2 2.2.1      √ purrr  0.2.4
## √ tibble  1.4.2      √ dplyr  0.7.4
## √ tidyr   0.8.0      √ stringr 1.2.0
## √ readr   1.1.1      √ forcats 0.3.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(stringr)
library(plyr)

## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize

## The following object is masked from 'package:purrr':
##
##   compact

library(reshape2)

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##   smiths

source("xtable.r")

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

```

# Load data -----

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

## re-encoding from CP1252

## Warning in `levels<~`(`*tmp*`, value = if (nl == nL) as.character(labels)
## else paste0(labels, : duplicated levels in factors are deprecated

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", "$40-50k", "$50-75k", "$75-100k", "$100-150k", ">150k", "Don't know/refused"))

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

raw <- dcast(counts, religion ~ income)

## Using freq as value column: use value.var to override.
head(raw, 15)

```

```

##           religion <$10k $10-20k $20-30k $30-40k $40-50k $50-75k
## 1           Agnostic    27     34      60     81      76     137
## 2            Atheist    12     27      37     52      35      70
## 3           Buddhist    27     21      30     34      33      58
## 4           Catholic  418    617     732    670     638    1116
## 5  Don't know/refused    15     14      15     11      10      35
## 6    Evangelical Prot  575    869    1064    982     881    1486
## 7             Hindu      1      9       7      9      11      34
## 8  Historically Black Prot 228    244     236    238     197     223
## 9      Jehovah's Witness   20     27      24     24      21      30
## 10             Jewish    19     19      25     25      30      95
## 11           Mainline Prot 289    495     619    655     651    1107

```

```
## 12      Mormon      29      40      48      51      56      112
## 13      Muslim       6       7       9      10       9       23
## 14      Orthodox     13      17      23      32      32      47
## 15      Other Christian  9       7      11      13      13      14
##      $75-100k $100-150k >150k Don't know/refused
## 1      122      109      84                      96
## 2       73       59      74                      76
## 3       62       39      53                      54
## 4      949      792     633                    1489
## 5       21       17      18                      116
## 6      949      723     414                    1529
## 7       47       48      54                      37
## 8      131      81      78                      339
## 9       15       11       6                      37
## 10      69       87     151                    162
## 11     939      753     634                    1328
## 12      85       49      42                      69
## 13      16        8       6                      22
## 14      38       42      46                      73
## 15      18       14      12                      18
```

```
new <- raw %>% gather(key = "income", value = "freq", -religion) %>% arrange(religion)
head(new, 15)
```

```
##      religion      income freq
## 1 Agnostic      <$10k   27
## 2 Agnostic     $10-20k  34
## 3 Agnostic     $20-30k  60
## 4 Agnostic     $30-40k  81
## 5 Agnostic     $40-50k  76
## 6 Agnostic     $50-75k 137
## 7 Agnostic     $75-100k 122
## 8 Agnostic     $100-150k 109
## 9 Agnostic      >150k   84
## 10 Agnostic Don't know/refused 96
## 11 Atheist      <$10k   12
## 12 Atheist     $10-20k  27
## 13 Atheist     $20-30k  37
## 14 Atheist     $30-40k  52
## 15 Atheist     $40-50k  35
```

translate table 7 to table 8

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

```
##
## 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")

raw <- read.csv("billboard.csv")
raw <- raw[, c("year", "artist.inverted", "track", "time", "date.entered",
              "x1st.week", "x2nd.week", "x3rd.week", "x4th.week", "x5th.week",
              "x6th.week", "x7th.week", "x8th.week", "x9th.week", "x10th.week",
              "x11th.week", "x12th.week", "x13th.week", "x14th.week", "x15th.week", "x16th.week", "x17th.week")]
names(raw)[2] <- "artist"

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

long_name <- nchar(raw$track) > 20
raw$track[long_name] <- paste0(substr(raw$track[long_name], 0, 20), "...")
head(raw, 10)
```

##	year	artist	track	time	date.entered	wk1	wk2								
## 1	2000	2 Pac	Baby Don't Cry	4:22	2000-02-26	87	82								
## 2	2000	2Ge+her	The Hardest Part Of ...	3:15	2000-09-02	91	87								
## 3	2000	3 Doors Down	Kryptonite	3:53	2000-04-08	81	70								
## 4	2000	3 Doors Down	Loser	4:24	2000-10-21	76	76								
## 5	2000	504 Boyz	Wobble Wobble	3:35	2000-04-15	57	34								
## 6	2000	98~0	Give Me Just One Nig...	3:24	2000-08-19	51	39								
## 7	2000	A*Teens	Dancing Queen	3:44	2000-07-08	97	97								
## 8	2000	Aaliyah	I Don't Wanna	4:15	2000-01-29	84	62								
## 9	2000	Aaliyah	Try Again	4:03	2000-03-18	59	53								
## 10	2000	Adams, Yolanda	Open My Heart	5:30	2000-08-26	76	76								
##	wk3	wk4	wk5	wk6	wk7	wk8	wk9	wk10	wk11	wk12	wk13	wk14	wk15	wk16	wk17
## 1	72	77	87	94	99	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 2	92	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 3	68	67	66	57	54	53	51	51	51	51	47	44	38	28	22
## 4	72	69	67	65	55	59	62	61	61	59	61	66	72	76	75
## 5	25	17	17	31	36	49	53	57	64	70	75	76	78	85	92
## 6	34	26	26	19	2	2	3	6	7	22	29	36	47	67	66
## 7	96	95	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 8	51	41	38	35	35	38	38	36	37	37	38	49	61	63	62
## 9	38	28	21	18	16	14	12	10	9	8	6	1	2	2	2
## 10	74	69	68	67	61	58	57	59	66	68	61	67	59	63	67
##	wk18	wk19	wk20	wk21	wk22	wk23	wk24	wk25	wk26	wk27	wk28	wk29	wk30	wk31	
## 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 3	18	18	14	12	7	6	6	6	5	5	4	4	4	4	
## 4	67	73	70	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 5	96	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 6	84	93	94	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 8	67	83	86	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	

## 9	2	3	4	5	5	6	9	13	14	16	23	22	33	36
## 10	71	79	89	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
##	wk32	wk33	wk34	wk35	wk36	wk37	wk38	wk39	wk40	wk41	wk42	wk43	wk44	wk45
## 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 3	3	3	3	4	5	5	9	9	15	14	13	14	16	17
## 4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 9	43	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
##	wk46	wk47	wk48	wk49	wk50	wk51	wk52	wk53	wk54	wk55	wk56	wk57	wk58	wk59
## 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 3	21	22	24	28	33	42	42	49	NA	NA	NA	NA	NA	NA
## 4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 9	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
##	wk60	wk61	wk62	wk63	wk64	wk65	wk66	wk67	wk68	wk69	wk70	wk71	wk72	wk73
## 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 9	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
##	wk74	wk75	wk76											
## 1	NA	NA	NA											
## 2	NA	NA	NA											
## 3	NA	NA	NA											
## 4	NA	NA	NA											
## 5	NA	NA	NA											
## 6	NA	NA	NA											
## 7	NA	NA	NA											
## 8	NA	NA	NA											
## 9	NA	NA	NA											
## 10	NA	NA	NA											

```

raw$artist <- iconv(raw$artist, "MAC", "ASCII//translit")

raw$track <- str_replace(raw$track, " \\(.*?\\)", "")
names(raw)[-1:5] <- str_c("wk", 1:76)
raw <- arrange(raw, year, artist, track)

long_name <- nchar(raw$track) > 20

```

```
raw$track[long_name] <- paste0(substr(raw$track[long_name], 0, 20), "...")
head(raw,10)
```

##	year	artist	track	time	date.entered	wk1	wk2								
## 1	2000	2 Pac	Baby Don't Cry	4:22	2000-02-26	87	82								
## 2	2000	2Ge+her	The Hardest Part Of ...	3:15	2000-09-02	91	87								
## 3	2000	3 Doors Down	Kryptonite	3:53	2000-04-08	81	70								
## 4	2000	3 Doors Down	Loser	4:24	2000-10-21	76	76								
## 5	2000	504 Boyz	Wobble Wobble	3:35	2000-04-15	57	34								
## 6	2000	98~0	Give Me Just One Nig...	3:24	2000-08-19	51	39								
## 7	2000	A*Teens	Dancing Queen	3:44	2000-07-08	97	97								
## 8	2000	Aaliyah	I Don't Wanna	4:15	2000-01-29	84	62								
## 9	2000	Aaliyah	Try Again	4:03	2000-03-18	59	53								
## 10	2000	Adams, Yolanda	Open My Heart	5:30	2000-08-26	76	76								
##	wk3	wk4	wk5	wk6	wk7	wk8	wk9	wk10	wk11	wk12	wk13	wk14	wk15	wk16	wk17
## 1	72	77	87	94	99	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 2	92	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 3	68	67	66	57	54	53	51	51	51	51	47	44	38	28	22
## 4	72	69	67	65	55	59	62	61	61	59	61	66	72	76	75
## 5	25	17	17	31	36	49	53	57	64	70	75	76	78	85	92
## 6	34	26	26	19	2	2	3	6	7	22	29	36	47	67	66
## 7	96	95	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## 8	51	41	38	35	35	38	38	36	37	37	38	49	61	63	62
## 9	38	28	21	18	16	14	12	10	9	8	6	1	2	2	2
## 10	74	69	68	67	61	58	57	59	66	68	61	67	59	63	67
##	wk18	wk19	wk20	wk21	wk22	wk23	wk24	wk25	wk26	wk27	wk28	wk29	wk30	wk31	
## 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 3	18	18	14	12	7	6	6	6	5	5	4	4	4	4	
## 4	67	73	70	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 5	96	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 6	84	93	94	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 8	67	83	86	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 9	2	3	4	5	5	6	9	13	14	16	23	22	33	36	
## 10	71	79	89	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
##	wk32	wk33	wk34	wk35	wk36	wk37	wk38	wk39	wk40	wk41	wk42	wk43	wk44	wk45	
## 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 3	3	3	3	4	5	5	9	9	15	14	13	14	16	17	
## 4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 9	43	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
##	wk46	wk47	wk48	wk49	wk50	wk51	wk52	wk53	wk54	wk55	wk56	wk57	wk58	wk59	
## 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 3	21	22	24	28	33	42	42	49	NA	NA	NA	NA	NA	NA	
## 4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
## 6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	

```
## 7    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 8    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 9    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 10   NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      wk60 wk61 wk62 wk63 wk64 wk65 wk66 wk67 wk68 wk69 wk70 wk71 wk72 wk73
## 1    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 2    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 3    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 4    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 5    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 6    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 7    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 8    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 9    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## 10   NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##      wk74 wk75 wk76
## 1    NA    NA    NA
## 2    NA    NA    NA
## 3    NA    NA    NA
## 4    NA    NA    NA
## 5    NA    NA    NA
## 6    NA    NA    NA
## 7    NA    NA    NA
## 8    NA    NA    NA
## 9    NA    NA    NA
## 10   NA    NA    NA
```

```
clean <- melt(raw, 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")]

clean_out <- mutate(clean,
                     date = as.character(date))
head(clean_out, 10)
```

##	year	artist	time	track	date	week	rank
## 1	2000	2 Pac	4:22	Baby Don't Cry	2000-02-26	1	87
## 2	2000	2 Pac	4:22	Baby Don't Cry	2000-03-04	2	82
## 3	2000	2 Pac	4:22	Baby Don't Cry	2000-03-11	3	72
## 4	2000	2 Pac	4:22	Baby Don't Cry	2000-03-18	4	77
## 5	2000	2 Pac	4:22	Baby Don't Cry	2000-03-25	5	87
## 6	2000	2 Pac	4:22	Baby Don't Cry	2000-04-01	6	94
## 7	2000	2 Pac	4:22	Baby Don't Cry	2000-04-08	7	99
## 8	2000	2Ge+her	3:15	The Hardest Part Of ...	2000-09-02	1	91
## 9	2000	2Ge+her	3:15	The Hardest Part Of ...	2000-09-09	2	87
## 10	2000	2Ge+her	3:15	The Hardest Part Of ...	2000-09-16	3	92