Assignment 4 Tidy Data

Doug Goon

April 27, 2018

Table 4 -> Table 6

```
pew <- as.tibble(pewdf)
pew%>% gather(-religion, key = "income", value = "frequency") %>% arrange(religion) %>% head(n=20)
```

```
## # A tibble: 20 x 3
##    religion             income frequency
##       <chr>              <chr>     <int>
##  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
## 16  Atheist            $50-75k        70
## 17  Atheist           $75-100k        73
## 18  Atheist          $100-150k        59
## 19  Atheist              >150k        74
## 20  Atheist Don't know/refused        76
```

Table 7 -> Table 8

```
bb <- as.tibble(read.csv("billboard.csv"))
bb.tidy <- bb %>%
  gather(key = "week", value = "rank" , -year, - artist.inverted, -track, -time, -genre,
  -date.entered, ... = -date.peaked)%>%
  select(year, artist=artist.inverted, time, track, date=date.entered, week, rank) %>%
  arrange(track)%>%
  filter(!is.na(rank))%>%
  separate(week, into=c("A","B","C"), sep=c(1:2), convert = F)%>%
  select(-A,-C) %>%
  dplyr::rename(week = B) %>%
  arrange(artist,track)%>%
  mutate(date=as.Date(date) +(as.numeric(week)-1)*7) %>%
  mutate(rank = as.integer(rank))
bb.tidy
```

```
## # A tibble: 5,307 x 7
##     year  artist   time
##    <int>  <fctr> <fctr>
```

```
##  1  2000   2 Pac   4:22
##  2  2000   2 Pac   4:22
##  3  2000   2 Pac   4:22
##  4  2000   2 Pac   4:22
##  5  2000   2 Pac   4:22
##  6  2000   2 Pac   4:22
##  7  2000   2 Pac   4:22
##  8  2000 2Ge+her   3:15
##  9  2000 2Ge+her   3:15
## 10  2000 2Ge+her   3:15
## # ... with 5,297 more rows, and 4 more variables: track <fctr>,
## #   date <date>, week <chr>, rank <int>
```