

# Table 7 -> 8

*Han Nguyen*

*February 26, 2018*

```
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.4.3
## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 2.2.1      v purrr  0.2.4
## v tibble  1.4.2      v dplyr  0.7.4
## v tidyr   0.8.0      v stringr 1.2.0
## v readr   1.1.1      v forcats 0.2.0

## Warning: package 'ggplot2' was built under R version 3.4.3
## Warning: package 'tibble' was built under R version 3.4.3
## Warning: package 'tidyr' was built under R version 3.4.3
## Warning: package 'readr' was built under R version 3.4.3
## Warning: package 'purrr' was built under R version 3.4.3
## Warning: package 'dplyr' was built under R version 3.4.3
## Warning: package 'stringr' was built under R version 3.4.2
## Warning: package 'forcats' was built under R version 3.4.3

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

bb <- read_csv("billboard.csv")

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   artist.inverted = col_character(),
##   track = col_character(),
##   time = col_time(format = ""),
##   genre = col_character(),
##   date.entered = col_date(format = ""),
##   date.peaked = col_date(format = ""),
##   x66th.week = col_character(),
##   x67th.week = col_character(),
##   x68th.week = col_character(),
##   x69th.week = col_character(),
##   x70th.week = col_character(),
##   x71st.week = col_character(),
##   x72nd.week = col_character(),
##   x73rd.week = col_character(),
##   x74th.week = col_character(),
##   x75th.week = col_character(),
##   x76th.week = col_character()
```

```
## )
## See spec(...) for full column specifications.
bb.1 <- bb %>% gather(key = "week", value = "rank", -year, -artist.inverted, -track, -time, -genre, -date)
bb.1
```

```
## # A tibble: 24,092 x 9
##   year artist.inverted track time genre date.entered date.peaked week
##   <int> <chr>          <chr> <tim> <chr> <date>      <date>      <chr>
## 1 2000 Destiny's Child Inde~ 03:38 Rock 2000-09-23 2000-11-18 x1st~
## 2 2000 Santana        Mari~ 04:18 Rock 2000-02-12 2000-04-08 x1st~
## 3 2000 Savage Garden I Kn~ 04:07 Rock 1999-10-23 2000-01-29 x1st~
## 4 2000 Madonna        Music 03:45 Rock 2000-08-12 2000-09-16 x1st~
## 5 2000 Aguilera, Chris~ Come~ 03:38 Rock 2000-08-05 2000-10-14 x1st~
## 6 2000 Janet          Does~ 04:17 Rock 2000-06-17 2000-08-26 x1st~
## 7 2000 Destiny's Child Say ~ 04:31 Rock 1999-12-25 2000-03-18 x1st~
## 8 2000 Iglesias, Enriq~ Be W~ 03:36 Latin 2000-04-01 2000-06-24 x1st~
## 9 2000 Sisqo          Inco~ 03:52 Rock 2000-06-24 2000-08-12 x1st~
## 10 2000 Lonestar      Amaz~ 04:25 Coun~ 1999-06-05 2000-03-04 x1st~
## # ... with 24,082 more rows, and 1 more variable: rank <chr>
```

```
bb.2 <- bb.1 %>% select(year, artist = artist.inverted, time, track, date = date.entered, week, rank)
bb.2
```

```
## # A tibble: 24,092 x 7
##   year artist          time track          date          week rank
##   <int> <chr>          <time> <chr>          <date>      <chr> <chr>
## 1 2000 Destiny's Child 03:38 Independent W~ 2000-09-23 x1st.~ 78
## 2 2000 Santana        04:18 Maria, Maria 2000-02-12 x1st.~ 15
## 3 2000 Savage Garden 04:07 I Knew I Love~ 1999-10-23 x1st.~ 71
## 4 2000 Madonna        03:45 Music          2000-08-12 x1st.~ 41
## 5 2000 Aguilera, Christina 03:38 Come On Over ~ 2000-08-05 x1st.~ 57
## 6 2000 Janet          04:17 Doesn't Reall~ 2000-06-17 x1st.~ 59
## 7 2000 Destiny's Child 04:31 Say My Name 1999-12-25 x1st.~ 83
## 8 2000 Iglesias, Enrique 03:36 Be With You 2000-04-01 x1st.~ 63
## 9 2000 Sisqo          03:52 Incomplete 2000-06-24 x1st.~ 77
## 10 2000 Lonestar      04:25 Amazed 1999-06-05 x1st.~ 81
## # ... with 24,082 more rows
```

```
bb.3 <- bb.2 %>% arrange(track)
```

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

```
bb.3
```

```
## # A tibble: 24,092 x 7
##   year artist time track          date          week rank
##   <int> <chr> <time> <chr>          <date>      <chr> <chr>
## 1 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x1st.w~ 100
## 2 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x2nd.w~ 99
## 3 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x3rd.w~ 96
## 4 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x4th.w~ 76
## 5 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x5th.w~ 55
## 6 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x6th.w~ 37
## 7 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x7th.w~ 24
```

```
## 8 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x8th.w~ 24
## 9 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x9th.w~ 30
## 10 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x10th.~ 36
## # ... with 24,082 more rows
```

```
bb.4 <- bb.3 %>% filter(!is.na(rank))
```

```
bb.4
```

```
## # A tibble: 5,307 x 7
##   year artist time track date week rank
##   <int> <chr> <time> <chr> <date> <chr> <chr>
## 1 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x1st.w~ 100
## 2 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x2nd.w~ 99
## 3 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x3rd.w~ 96
## 4 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x4th.w~ 76
## 5 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x5th.w~ 55
## 6 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x6th.w~ 37
## 7 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x7th.w~ 24
## 8 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x8th.w~ 24
## 9 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x9th.w~ 30
## 10 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 x10th.~ 36
## # ... with 5,297 more rows
```

```
bb.5 <- bb.4 %>% separate(week, into=c("A", "B", "C"), sep=c(1,-8), convert = TRUE)
```

```
bb.5
```

```
## # A tibble: 5,307 x 9
##   year artist time track date A B C rank
##   <int> <chr> <time> <chr> <date> <chr> <int> <chr> <chr>
## 1 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 1st.w~ 100
## 2 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 2nd.w~ 99
## 3 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 3rd.w~ 96
## 4 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 4th.w~ 76
## 5 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 5th.w~ 55
## 6 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 6th.w~ 37
## 7 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 7th.w~ 24
## 8 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 8th.w~ 24
## 9 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x NA 9th.w~ 30
## 10 2000 Nelly 04:17 (Hot S**t) Cou~ 2000-04-29 x 1 0th.w~ 36
## # ... with 5,297 more rows
```

```
bb.6 <- bb.5 %>% select(-A, -C)
```

```
bb.6
```

```
## # A tibble: 5,307 x 7
##   year artist time track date B rank
##   <int> <chr> <time> <chr> <date> <int> <chr>
## 1 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 NA 100
## 2 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 NA 99
## 3 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 NA 96
## 4 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 NA 76
## 5 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 NA 55
## 6 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 NA 37
## 7 2000 Nelly 04:17 (Hot S**t) Country Grammar 2000-04-29 NA 24
```

```
## 8 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 24
## 9 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 30
## 10 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 1 36
## # ... with 5,297 more rows
```

```
bb.7 <- bb.6 %>% dplyr::rename(week = B)
```

```
bb.7
```

```
## # A tibble: 5,307 x 7
##   year artist time track date week rank
##   <int> <chr> <time> <chr> <date> <int> <chr>
## 1 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 100
## 2 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 99
## 3 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 96
## 4 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 76
## 5 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 55
## 6 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 37
## 7 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 24
## 8 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 24
## 9 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 NA 30
## 10 2000 Nelly 04:17 (Hot S**) Country Grammar 2000-04-29 1 36
## # ... with 5,297 more rows
```

```
bb.8 <- bb.7 %>% arrange(artist, track)
```

```
bb.8
```

```
## # A tibble: 5,307 x 7
##   year artist time track date week rank
##   <int> <chr> <time> <chr> <date> <int> <chr>
## 1 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 87
## 2 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 82
## 3 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 72
## 4 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 77
## 5 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 87
## 6 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 94
## 7 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 99
## 8 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 NA 91
## 9 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 NA 87
## 10 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 NA 92
## # ... with 5,297 more rows
```

```
bb.9 <- bb.8 %>% group_by(artist)
```

```
bb.9
```

```
## # A tibble: 5,307 x 7
## # Groups:   artist [228]
##   year artist time track date week rank
##   <int> <chr> <time> <chr> <date> <int> <chr>
## 1 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 87
## 2 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 82
## 3 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 72
## 4 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 77
## 5 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 87
## 6 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 94
```

```
## 7 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 NA 99
## 8 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 NA 91
## 9 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 NA 87
## 10 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 NA 92
## # ... with 5,297 more rows
```

```
bb.10 <- bb.9 %>% mutate(week = 1:n())
```

```
bb.10
```

```
## # A tibble: 5,307 x 7
## # Groups:   artist [228]
##   year artist time track date week rank
##   <int> <chr> <time> <chr> <date> <int> <chr>
## 1 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 1 87
## 2 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 2 82
## 3 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 3 72
## 4 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 4 77
## 5 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 5 87
## 6 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 6 94
## 7 2000 2 Pac 04:22 Baby Don't Cry (Keep Ya He~ 2000-02-26 7 99
## 8 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 1 91
## 9 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 2 87
## 10 2000 2Ge+her 03:15 The Hardest Part Of Breaki~ 2000-09-02 3 92
## # ... with 5,297 more rows
```

```
bb.11 <- bb.10 %>% mutate(data = date + (week-1)*7)
```

```
bb.11
```

```
## # A tibble: 5,307 x 8
## # Groups:   artist [228]
##   year artist time track date week rank data
##   <int> <chr> <time> <chr> <date> <int> <chr> <date>
## 1 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 1 87 2000-02-26
## 2 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 2 82 2000-03-04
## 3 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 3 72 2000-03-11
## 4 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 4 77 2000-03-18
## 5 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 5 87 2000-03-25
## 6 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 6 94 2000-04-01
## 7 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 7 99 2000-04-08
## 8 2000 2Ge+her 03:15 The Hardest Par~ 2000-09-02 1 91 2000-09-02
## 9 2000 2Ge+her 03:15 The Hardest Par~ 2000-09-02 2 87 2000-09-09
## 10 2000 2Ge+her 03:15 The Hardest Par~ 2000-09-02 3 92 2000-09-16
## # ... with 5,297 more rows
```

```
bb.12 <- bb.11 %>% mutate(rank = as.integer(rank))
```

```
bb.12
```

```
## # A tibble: 5,307 x 8
## # Groups:   artist [228]
##   year artist time track date week rank data
##   <int> <chr> <time> <chr> <date> <int> <int> <date>
## 1 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 1 87 2000-02-26
## 2 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 2 82 2000-03-04
## 3 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 3 72 2000-03-11
```

```
## 4 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 4 77 2000-03-18
## 5 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 5 87 2000-03-25
## 6 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 6 94 2000-04-01
## 7 2000 2 Pac 04:22 Baby Don't Cry ~ 2000-02-26 7 99 2000-04-08
## 8 2000 2Ge+her 03:15 The Hardest Par~ 2000-09-02 1 91 2000-09-02
## 9 2000 2Ge+her 03:15 The Hardest Par~ 2000-09-02 2 87 2000-09-09
## 10 2000 2Ge+her 03:15 The Hardest Par~ 2000-09-02 3 92 2000-09-16
## # ... with 5,297 more rows
```