Game of Thrones Survival Analysis

Shannon Houser

9/27/2021

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1
```

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

```r
library(survival)
library(survminer)
```

```
## Warning: package 'survminer' was built under R version 4.1.1
```

```
## Loading required package: ggpubr
```

```
## Warning: package 'ggpubr' was built under R version 4.1.1
```
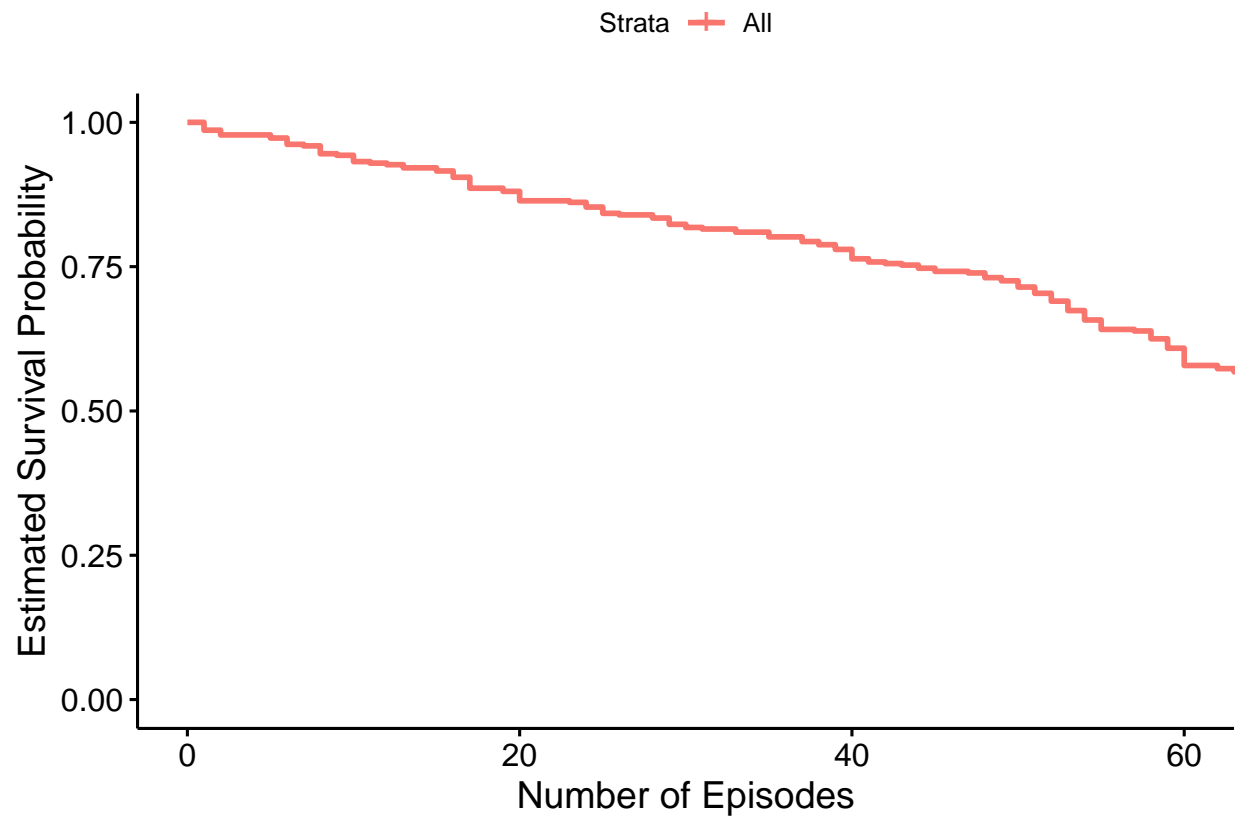
```r
got <- read_csv("210927_got.csv")
```

```
## Rows: 368 Columns: 68
```

```
## -- Column specification ----------------------------------------------------------
## Delimiter: ","
## chr (15): character_name, gender, house, spouse, parents, siblings, parent_o...
## dbl (53): royal, kingsguard, s1_episodes, s1_screenTime, s1_numOfCharactersI...
```
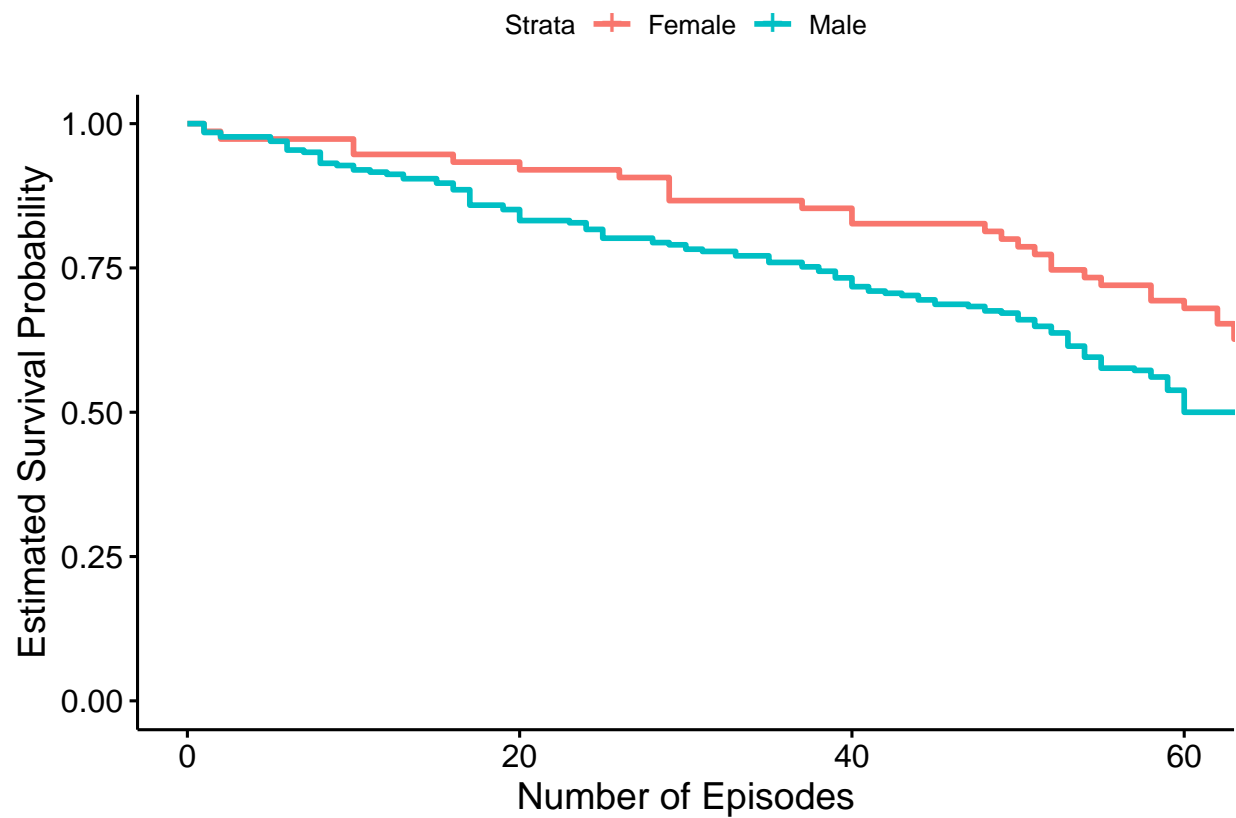
```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
got2 <- got %>%
  select(duration_in_episodes, is_dead, character_name, royal, house, gender)
```
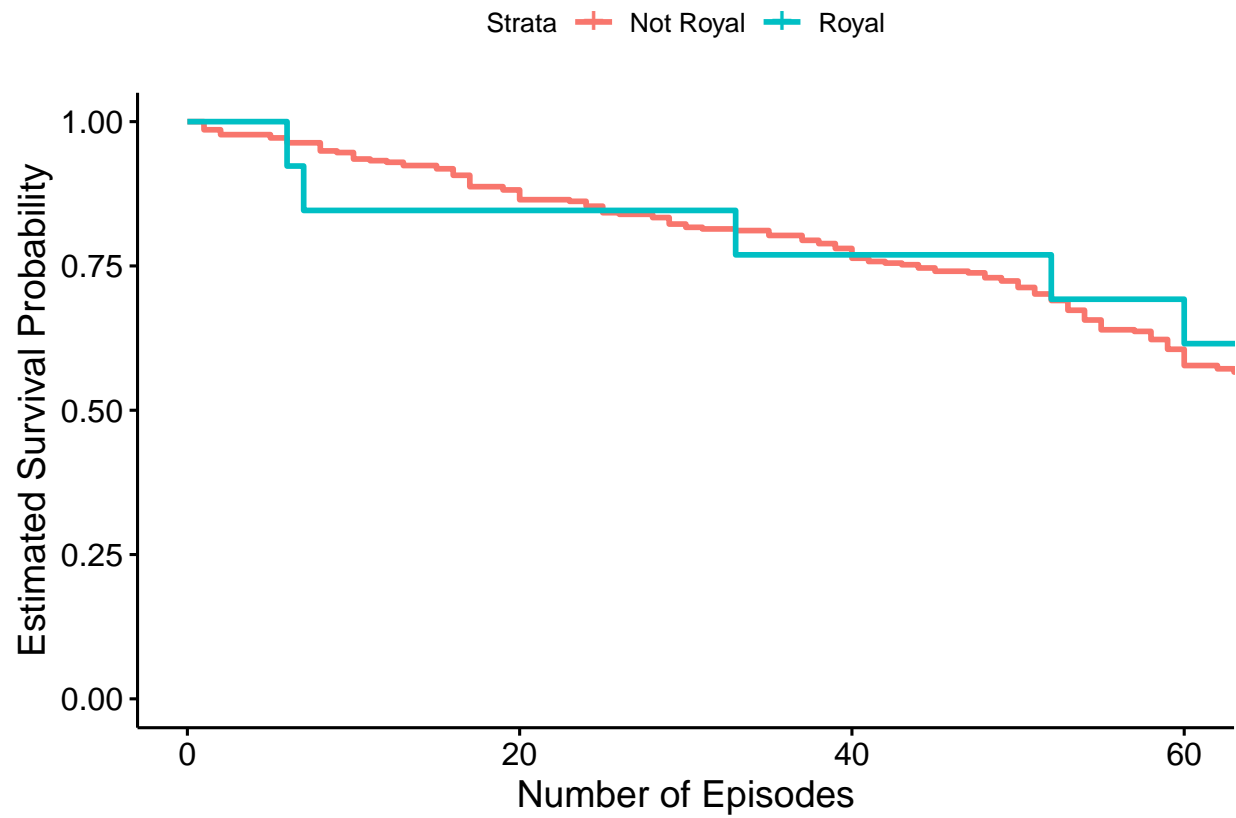
```r
ggsurvplot(survfit(Surv(duration_in_episodes, is_dead) ~ 1, data = got2),
           xlab = "Number of Episodes",
           ylab = "Estimated Survival Probability",
           conf.int = F)
```



```r
ggsurvplot(survfit(Surv(duration_in_episodes, is_dead) ~ gender, data = got2),
           xlab = "Number of Episodes",
           ylab = "Estimated Survival Probability",
           conf.int = F,
           legend.labs = c("Female", "Male"))
```
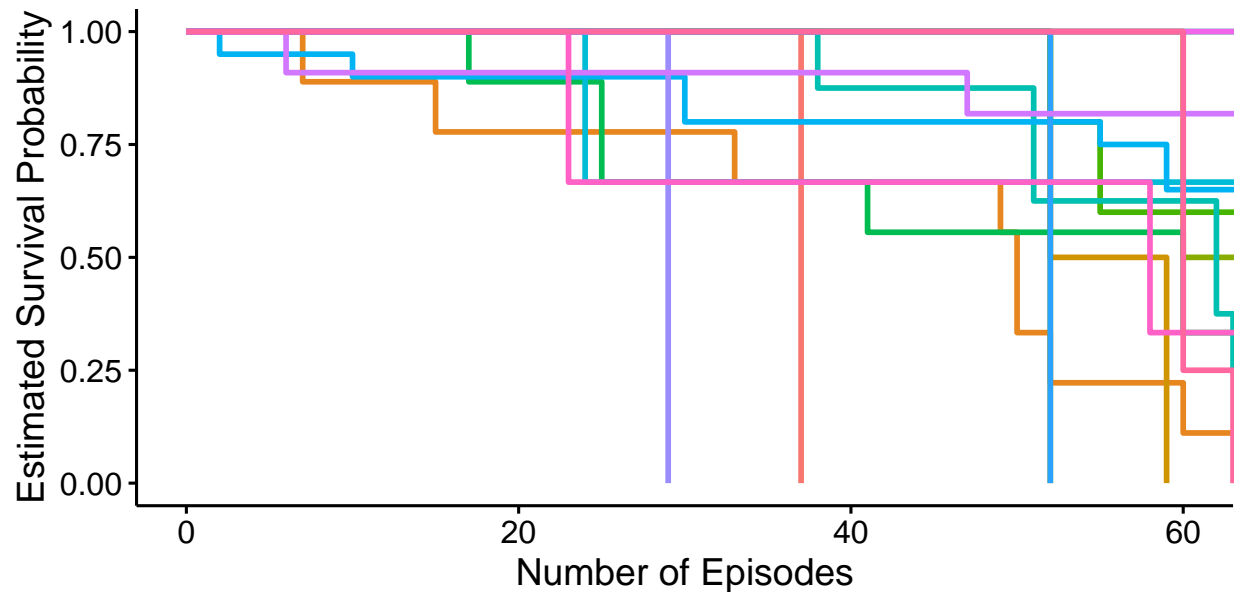
```
ggsurvplot(survfit(Surv(duration_in_episodes, is_dead) ~ royal, data = got2),
          xlab = "Number of Episodes",
          ylab = "Estimated Survival Probability",
          conf.int = F,
          legend.labs = c("Not Royal", "Royal"))
```

```
ggsurvplot(survfit(Surv(duration_in_episodes, is_dead) ~ house,
                   data = got2),
           xlab = "Number of Episodes",
           ylab = "Estimated Survival Probability",
           conf.int = F)
```

```
survdiff(Surv(duration_in_episodes, is_dead) ~ gender,
                  data = got2)
```

```
## Call:
## survdiff(formula = Surv(duration_in_episodes, is_dead) ~ gender,
##     data = got2)
##
## n=337, 31 observations deleted due to missingness.
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## gender=female    75       28     40.8      4.03      5.44
## gender=male     262      137    124.2      1.33      5.44
##
##  Chisq= 5.4  on 1 degrees of freedom, p= 0.02
```

```
survdiff(Surv(duration_in_episodes, is_dead) ~ royal,
                  data = got2)
```

```
## Call:
## survdiff(formula = Surv(duration_in_episodes, is_dead) ~ royal,
##     data = got2)
##
##            N Observed Expected (O-E)^2/E (O-E)^2/V
## royal=0  355      160   159.08    0.0053      0.15
```

```
## royal=1  13        5     5.92    0.1423       0.15
##
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.7
```

```
survdiff(Surv(duration_in_episodes, is_dead) ~ house,
                data = got2)
```

```
## Call:
## survdiff(formula = Surv(duration_in_episodes, is_dead) ~ house,
##     data = got2)
##
## n=86, 282 observations deleted due to missingness.
##
##                           N Observed Expected (O-E)^2/E (O-E)^2/V
## house=Arryn;Tully          1       1    0.190     3.451     3.508
## house=Baratheon            9       8    2.944     8.684     9.657
## house=Bolton               2       2    0.882     1.416     1.482
## house=Bolton;Frey          1       1    0.388     0.967     0.998
## house=Frey                 6       3    4.447     0.471     0.547
## house=Greyjoy              5       2    3.254     0.483     0.540
## house=Lannister            9       6    4.316     0.657     0.756
## house=Lannister;Baratheon  1       0    0.815     0.815     0.871
## house=Martell              8       6    4.642     0.398     0.459
## house=Mormont              3       1    1.727     0.306     0.332
## house=Stark               20       8   12.362     1.539     2.160
## house=Stark;Targaryen      1       1    0.388     0.967     0.998
## house=Stark;Tully          1       1    0.136     5.497     5.580
## house=Targaryen           11       3    7.590     2.776     3.466
## house=Tarly                1       0    0.815     0.815     0.871
## house=Tully                3       2    1.339     0.326     0.348
## house=Tyrell               4       4    2.766     0.551     0.620
##
##  Chisq= 32.3  on 16 degrees of freedom, p= 0.009
```

It appears that gender and house have a differential survival, but not royalty. We can see that females have
a smaller probability of being killed than males. We can also see that although the probability of dying is
not statistically significant between royals and non-royals, we can see that the plots are very different. It
seems that when royals die, they die in groups, whereas non-royals die much more sporadically. This might
make sense considering royals would probably only die when another group infiltrated their castle/kingdom
or during war or other events where many royals would die together. Lastly, house seems to have an effect
on survival. This probably has confounding factors because those in the same house are most often together
in location and have common enemies, so they are more likely to be killed together. Furthermore, many of
the characters did not have a house assigned, so we are completing this analysis on less than 100 characters.