Level difficulty in Candy Crush

Tong Shen

February 6, 2019

```
# Loading in packages
library(tidyverse)

# Reading in the data
data <- read_csv("candy_crush.csv")
```

```
# Printing out the first couple of rows
head(data)
```

```
## # A tibble: 6 x 5
##   player_id                        dt         level num_attempts num_success
##   <chr>                            <date>     <int>        <int>       <int>
## 1 6dd5af4c7228fa353d505767143f5~ 2014-01-04     4            3           1
## 2 c7ec97c39349ab7e4d39b4f74062e~ 2014-01-01     8            4           1
## 3 c7ec97c39349ab7e4d39b4f74062e~ 2014-01-05    12            6           0
## 4 a32c5e9700ed356dc8dd5bb3230c5~ 2014-01-03    11            1           1
## 5 a32c5e9700ed356dc8dd5bb3230c5~ 2014-01-07    15            6           0
## 6 b94d403ac4edf639442f93eeffdc7~ 2014-01-01     8            8           1
```

```
summary(data)
```

```
##   player_id               dt                  level
##  Length:16865       Min.   :2014-01-01   Min.   : 1.000
##  Class :character   1st Qu.:2014-01-02   1st Qu.: 6.000
##  Mode  :character   Median :2014-01-04   Median : 9.000
##                     Mean   :2014-01-04   Mean   : 9.287
##                     3rd Qu.:2014-01-06   3rd Qu.:14.000
##                     Max.   :2014-01-07   Max.   :15.000
##   num_attempts      num_success
##  Min.   :  0.000   Min.   : 0.0000
##  1st Qu.:  1.000   1st Qu.: 0.0000
##  Median :  3.000   Median : 1.0000
##  Mean   :  5.535   Mean   : 0.6272
##  3rd Qu.:  7.000   3rd Qu.: 1.0000
##  Max.   :258.000   Max.   :55.0000
```

```
#checking in the dataset
total_player<-length(unique(data$player_id))
range<-range(data$dt)
paste("The total number of players is", total_player)
```

```
## [1] "The total number of players is 6814"
```

```
paste("The period for which we have the data is from", min(range), "to", max(range))
```

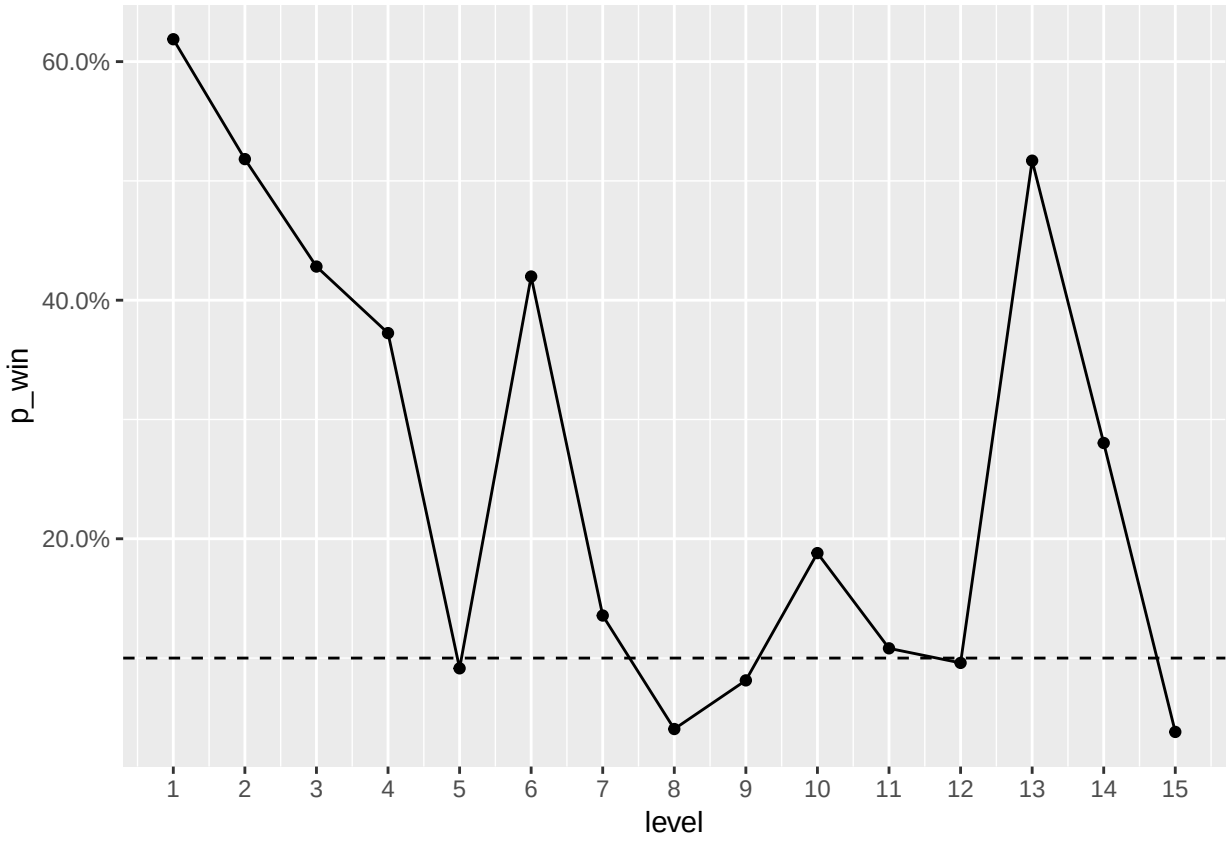
```
## [1] "The period for which we have the data is from 2014-01-01 to 2014-01-07"
```

```
# Calculating level difficulty
difficulty <- data %>%
  group_by(level) %>%
```

```
  summarise(attempts= sum(num_attempts),total_win= sum(num_success)) %>%
  mutate(p_win = total_win/attempts)
difficulty
```

```
## # A tibble: 15 x 4
##    level attempts total_win  p_win
##    <int>    <int>     <int>  <dbl>
##  1     1     1322       818 0.619
##  2     2     1285       666 0.518
##  3     3     1546       662 0.428
##  4     4     1893       705 0.372
##  5     5     6937       634 0.0914
##  6     6     1591       668 0.420
##  7     7     4526       614 0.136
##  8     8    15816       641 0.0405
##  9     9     8241       670 0.0813
## 10    10     3282       617 0.188
## 11    11     5575       603 0.108
## 12    12     6868       659 0.0960
## 13    13     1327       686 0.517
## 14    14     2772       777 0.280
## 15    15    30374      1157 0.0381
```

```
# Plotting the level difficulty profile with points and a 10% dashed line
difficulty %>% ggplot(aes(x = level, y = p_win))+
 geom_line()+
scale_x_continuous(breaks = 1:15) +
scale_y_continuous(labels = scales::percent) +
geom_point()+
geom_hline(yintercept = 0.1,linetype = 2)
```

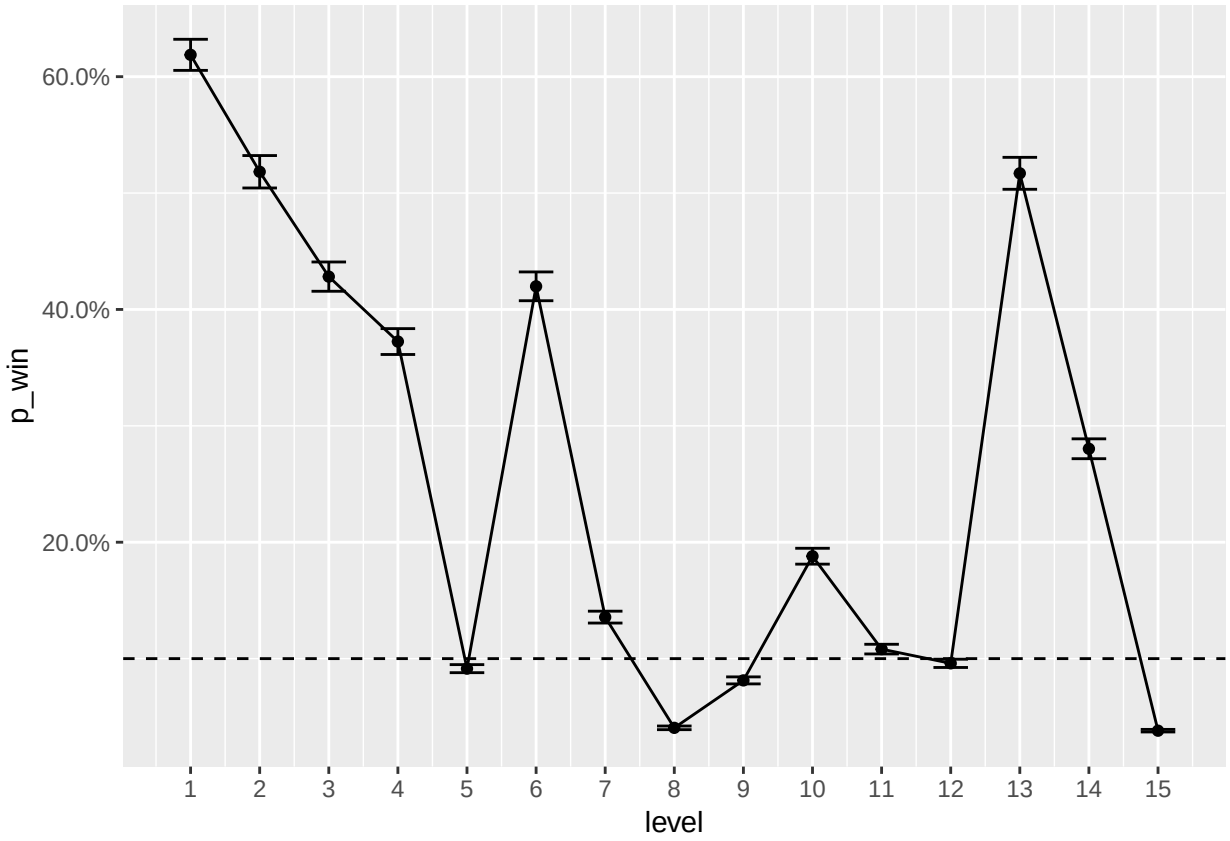



```
# Computing the standard error of p_win for each level
difficulty <- difficulty %>%
    mutate(error = sqrt(p_win*(1-p_win)/attempts))
difficulty
```

```
## # A tibble: 15 x 5
##    level attempts total_win  p_win   error
##    <int>    <int>     <int>  <dbl>   <dbl>
##  1     1     1322       818 0.619  0.0134
##  2     2     1285       666 0.518  0.0139
##  3     3     1546       662 0.428  0.0126
##  4     4     1893       705 0.372  0.0111
##  5     5     6937       634 0.0914 0.00346
##  6     6     1591       668 0.420  0.0124
##  7     7     4526       614 0.136  0.00509
##  8     8    15816       641 0.0405 0.00157
##  9     9     8241       670 0.0813 0.00301
## 10    10     3282       617 0.188  0.00682
## 11    11     5575       603 0.108  0.00416
## 12    12     6868       659 0.0960 0.00355
## 13    13     1327       686 0.517  0.0137
## 14    14     2772       777 0.280  0.00853
## 15    15    30374      1157 0.0381 0.00110
```

```
# Adding standard error bars
difficulty %>% ggplot(aes(x = level, y = p_win))+
 geom_line()+
```

```
scale_x_continuous(breaks = 1:15) +
scale_y_continuous(labels = scales::percent) +
geom_point()+
geom_hline(yintercept = 0.1,linetype = 2) +
geom_errorbar(aes(ymin = p_win - error, ymax = p_win + error), width = 0.5)
```




```
# The probability of completing the episode without losing a single time
prod(difficulty$p_win)
```

```
## [1] 9.447141e-12
```

This means that the game designer does not need to worry that the player might complete the episode in one attempt.