Forenic on Gacha Game

Uyen Doan

2025-06-24

Contents

Set up

```
#make sure code chunk is rendered with code
knitr::opts_chunk$set(echo = TRUE, comment = "")
```

Load Data

The data set used for this is named: "gacha database - 2024.csv"

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------------------------------------------
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.2     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.4
-- Conflicts ------------------------------------------------------------------------------------
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(patchwork)
library(dplyr)

Gacha <- read.csv("Gacha Database - 2024.csv", stringsAsFactors=T)
```

Goal(s):

To test the accuracy of the notions that: Gacha games developed based on established IPs (Intellectual Properties) have shorter lifespan than stanalone gacha games.

About the data set:

- The data set has 9 variables, but notables are:
    1. Title: The name of the game. All using English names or the romanization of the Languages.
    2. Time: Total time the game had been in service. Counted in years.
    3. Region: The Region the game has servers in.
    4. IP: Whether or not the game was using/based on an already existed IP (Intellectual Property)
    5. Franchise: If the answers to Variable #4 is Y (Yes), what franchise does the game belonged to.
- All chosen games are those that announced End of Service (EoS) in 2024.

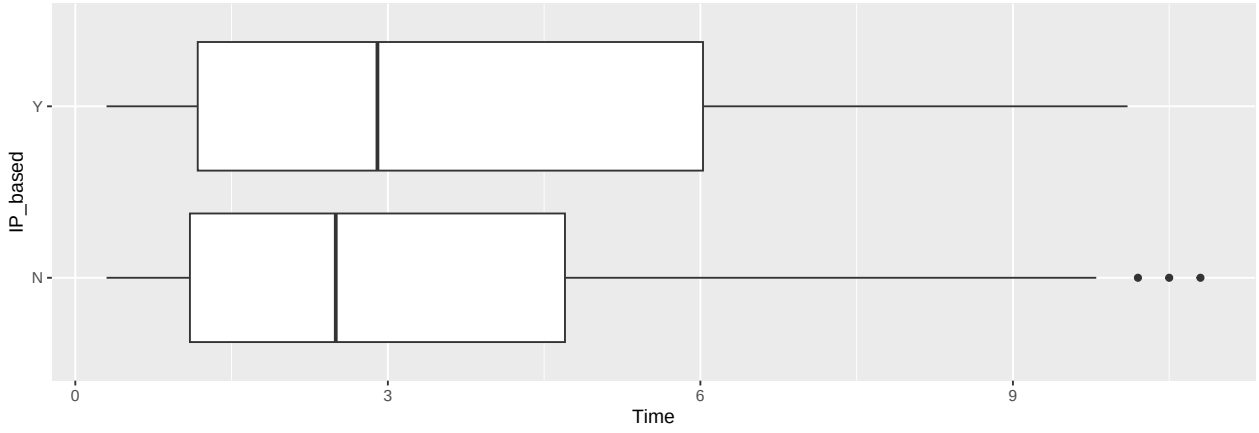
Plotting:

The dataset - due to the nature of live-service, multiserver games - contain multiple duplicates as certain games may have different servers open up as different time. The approach chosen for this is to compare:

- The original with duplicates
- The one grouped together using the mean time of all group's elements
- The one grouped together using the max time of all group's elements

```
Gacha <- as.data.frame(Gacha)
Gacha$Time <- as.numeric(Gacha$Time)

og <- (ggplot(Gacha, aes(x=Time, y=IP_based)) + geom_boxplot())
og
```

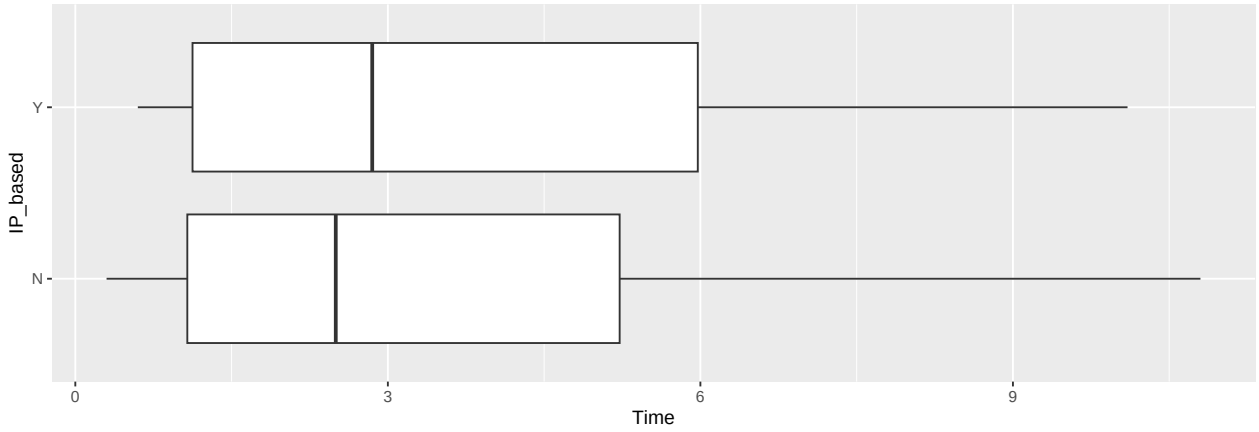



```
#Find the mean of the Time
df <- Gacha %>%
  group_by(IP_based) %>%
  summarise (Standard = sd(Time), Time = mean(Time)) %>%
  #sd(Time) must be placed before mean(Time) or it will use the sd(Time) instead
  ungroup()
```

```
df %>% dplyr::select(IP_based, Time, Standard)
```

```
# A tibble: 2 x 3
  IP_based  Time Standard
  <fct>    <dbl>    <dbl>
1 N         3.39     2.96
2 Y         3.81     2.75
```

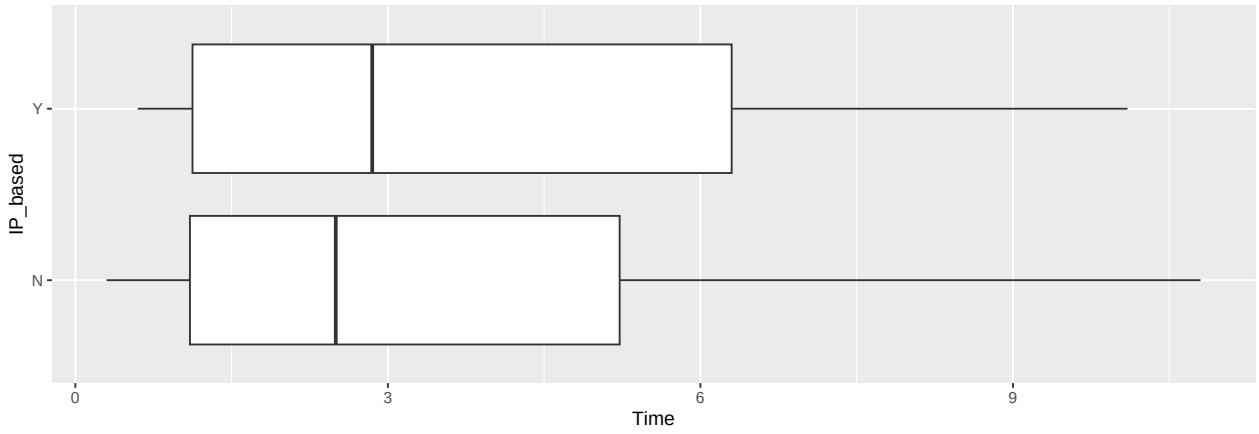
```
df_average <- Gacha %>%
              group_by(Title) %>%
              summarise (Time = mean(Time),
                         IP_based = unique (IP_based)) %>%
              ungroup()
average_plot <- ggplot(df_average, aes(x=Time, y=IP_based)) + geom_boxplot()
average_plot
```




```
df <- df_average %>%
  group_by(IP_based) %>%
  summarise (Standard = sd(Time), Time = mean(Time)) %>%
  ungroup()
df %>% dplyr::select(IP_based, Time, Standard)
```

```
# A tibble: 2 x 3
  IP_based  Time Standard
  <fct>    <dbl>    <dbl>
1 N         3.46     3.20
2 Y         3.73     2.83
```

```
df_max <- Gacha %>%
              group_by(Title) %>%
              summarise (Time = max(Time),
                         IP_based = unique (IP_based)) %>%
              ungroup()
max_plot <- ggplot(df_max, aes(x=Time, y=IP_based)) + geom_boxplot()
max_plot
```



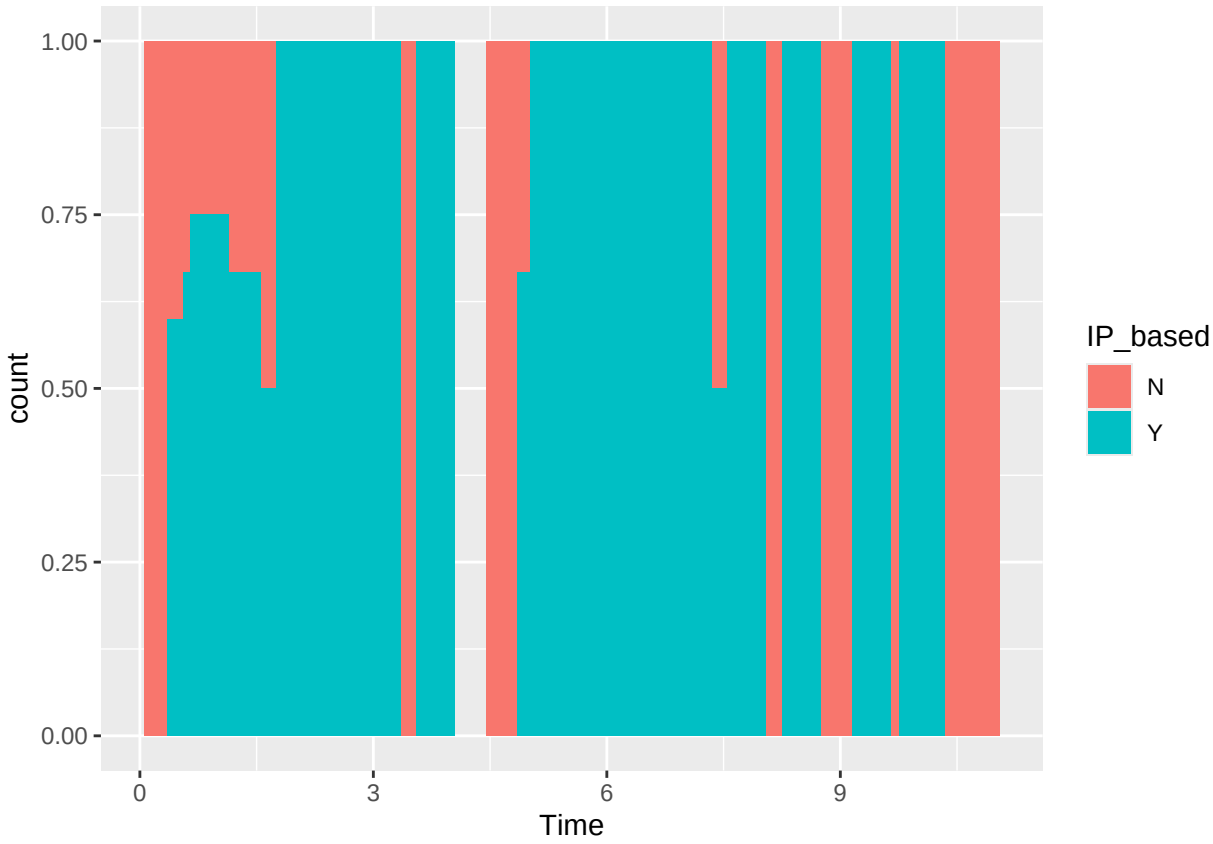

```
df <- df_max %>%
  group_by(IP_based) %>%
  summarise (Standard = sd(Time), Time = mean(Time)) %>%
  ungroup()
df %>% dplyr::select(IP_based, Time, Standard)
```

```
# A tibble: 2 x 3
  IP_based  Time Standard
  <fct>    <dbl>    <dbl>
1 N         3.53     3.21
2 Y         3.80     2.87
```

- We can see that, in contrast to the notion, in all case:
  - IP-based gacha games either have higher or almost the same minimum values.
  - IP-baseds on average have higher life span.
  - IP-baseds have higher median.
- It should be noted that:
  - The original data frame have the 3 outliers to the right on the Standalones.
  - Standalones have much longer Q3 and higher maximum values in all cases.
  - IP-baseds have longer boxes in all 3 cases.
  - All cases, both catagories have almost the same length on Q1.

Proportion plot

```
ggplot(df_average,aes(x = Time,fill = IP_based)) +
    geom_bar(width = 0.5, position = "fill")
```



Hypothesis:

Gacha games developed based on established IPs do not generally have shorter lifespan than stanalone gacha games. This notion may stem from the fact that: Standalone games have a fews with long life span. These games could earn fame throughout their services, leading to some sort of survivorship bias.