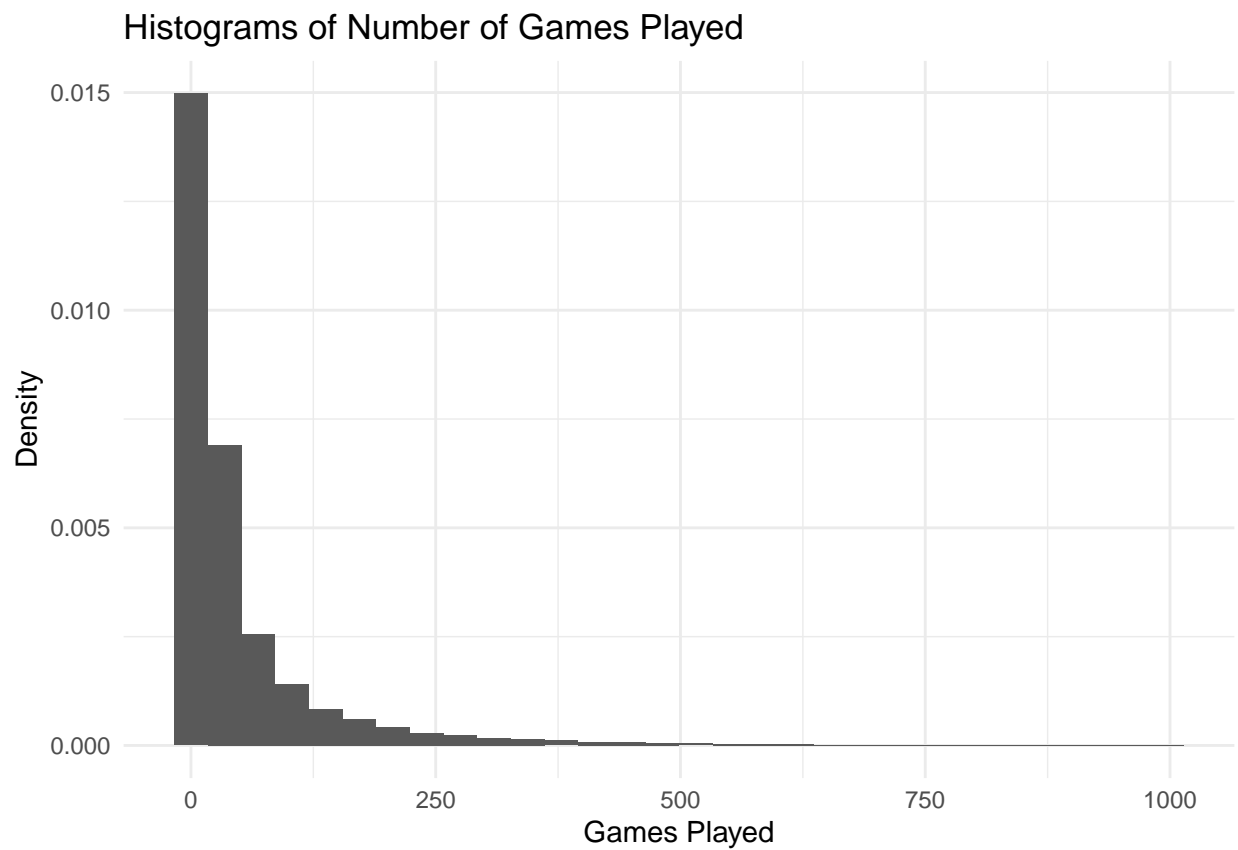# Data project 2: Game developer A/B testing

Young Woong Min SSID (3034519151)

Problem 1: Exploring the data

Part a

I removed all rows from game_data with games_played >= 1000 because extreme outliers in the right tail are messing up the visualization by making the range of values on the x-axis too large.
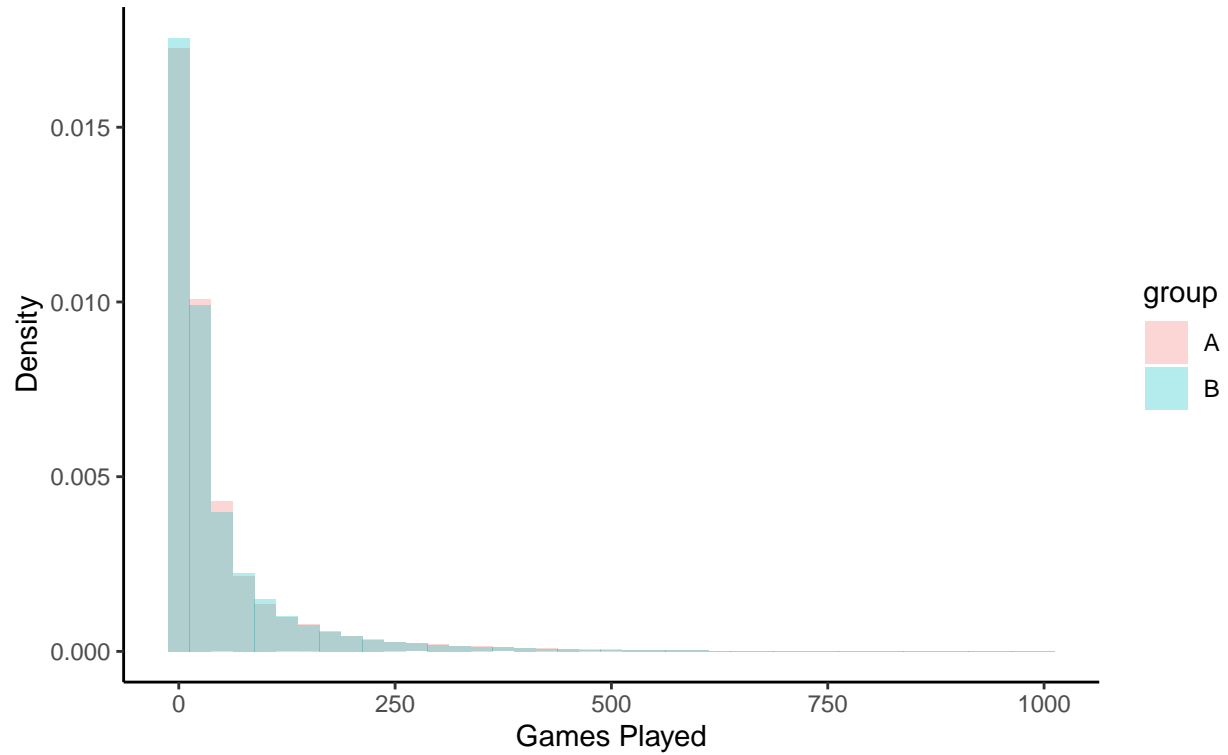
Histograms of Number of Games Played

Part b

We can observe that the number of games played across each group doesn't differ. Based on my plot, I would not recommend one version of the game over the other.

Histograms of Number of Games Played in Group A and Group B
The histograms are near perfectly superimposed on each other

Part c

```r
groupA <- game_data %>%
    filter(group == "A")
groupB <- game_data %>%
    filter(group == "B")
prop1A <- mean(groupA$retained_1)
prop1B <- mean(groupB$retained_1)
prop7A <- mean(groupA$retained_7)
prop7B <- mean(groupB$retained_7)
```

The proportion of players in groups A and B that continued to play the game 1 day after installation was 0.4481879 and 0.4422827 respectively. The proportion of players in groups A and B that continued to play the game 7 days after installation was 0.1902013 and 0.182 respectively.

Problem 2: Bootstrapped estimates

Part a

A theoretical 95% confidence interval should be 50.6 to 53.1.
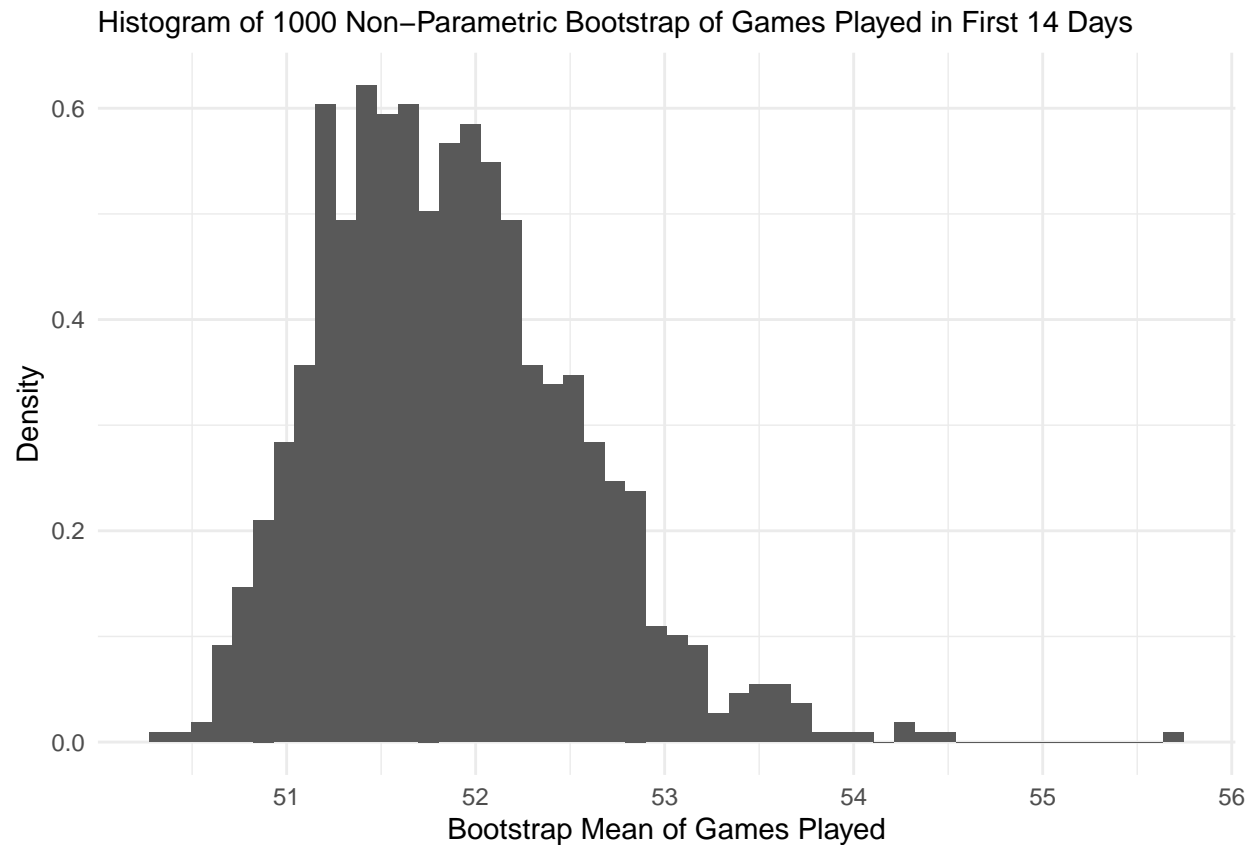
```
# 95% ci for average number of games played in the first 14 days ignoring group
sample_mean <- mean(game_data$games_played)
sample_sd <- sd(game_data$games_played)/(dim(game_data)[1]^0.5)
ci_95 <- c(sample_mean - qnorm(0.975) * sample_sd, sample_mean + qnorm(0.975) * sample_sd)
ci_95
```

```
## [1] 50.59948 53.14543
```

Part b

```
np_boot_mean_df <- map_df(1:1000, function(i) {
    # sample from the data with replacement
    bootstrap_data <- sample(game_data$games_played, dim(game_data)[1], replace = TRUE)
    # compute the sample mean of the bootstrap sample
    data.frame(boot_mean = mean(bootstrap_data))
})

# plot the histogram of the nonparametric bootstrap
ggplot(np_boot_mean_df, aes(boot_mean)) + geom_histogram(aes(y = ..density..), bins = 50) +
    labs(title = "Histogram of 1000 Non-Parametric Bootstrap of Games Played in First 14 Days") +
    xlab("Bootstrap Mean of Games Played") + ylab("Density") + theme(plot.title = element_text(size = 1
```

Histogram of 1000 Non−Parametric Bootstrap of Games Played in First 14 Days


```
bootstrap_ci_95 <- c(mean(np_boot_mean_df$boot_mean) - qnorm(0.975) * sd(np_boot_mean_df$boot_mean),
    mean(np_boot_mean_df$boot_mean) + qnorm(0.975) * sd(np_boot_mean_df$boot_mean))

bootstrap_ci_95
```

```
## [1] 50.55252 53.19369
```
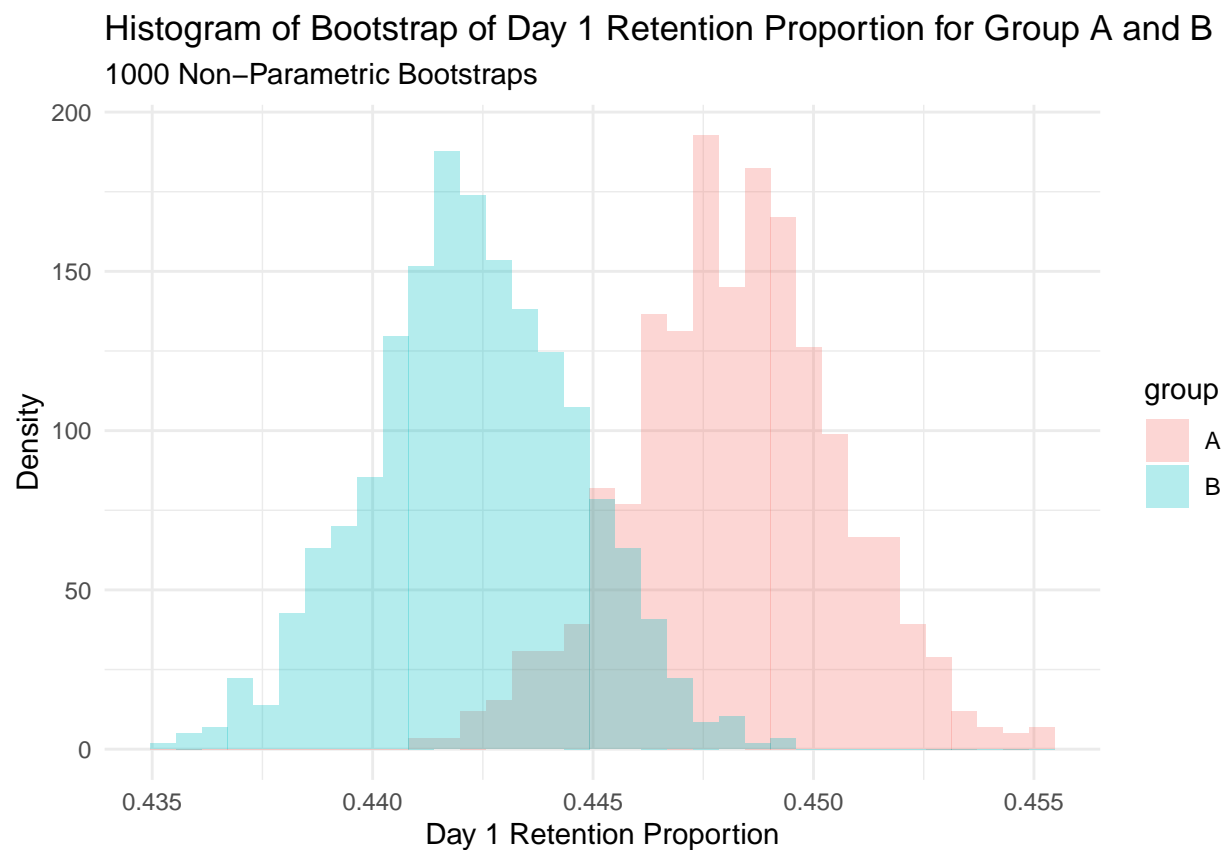
Part c

```
groupA <- game_data %>%
    filter(group == "A")

groupB <- game_data %>%
    filter(group == "B")


np_boot_mean_df <- map_df(1:1000, function(i) {
    # sample from the data with replacement
    bootstrap_dataA <- sample(groupA$retained_1, dim(groupA)[1], replace = TRUE)
    bootstrap_dataB <- sample(groupB$retained_1, dim(groupB)[1], replace = TRUE)
    # compute the sample mean of the bootstrap sample
    data.frame(boot_mean = c(mean(bootstrap_dataA), mean(bootstrap_dataB)), group = c("A",
        "B"))
```

```
})
```

```
# plot the histogram of the nonparametric bootstrap
ggplot(np_boot_mean_df, aes(x = boot_mean, fill = group)) + geom_histogram(position = "identity",
    aes(y = ..density..), alpha = 0.3, bins = 35) + labs(title = "Histogram of Bootstrap of Day 1 Reten
    subtitle = "1000 Non-Parametric Bootstraps") + xlab("Day 1 Retention Proportion") +
    ylab("Density") + theme(plot.title = element_text(size = 11), plot.subtitle = element_text(size = 9)
    theme_minimal()
```

Histogram of Bootstrap of Day 1 Retention Proportion for Group A and B
1000 Non–Parametric Bootstraps

Based on my plot, I would recommend version A of the game because the average day 1 retention proportion of group A appears to be larger than group B's.

Part d

```
groupA <- game_data %>%
    filter(group == "A")

groupB <- game_data %>%
    filter(group == "B")


np_boot_mean_df <- map_df(1:1000, function(i) {
    # sample from the data with replacement
    bootstrap_dataA <- sample(groupA$retained_7, dim(groupA)[1], replace = TRUE)
```
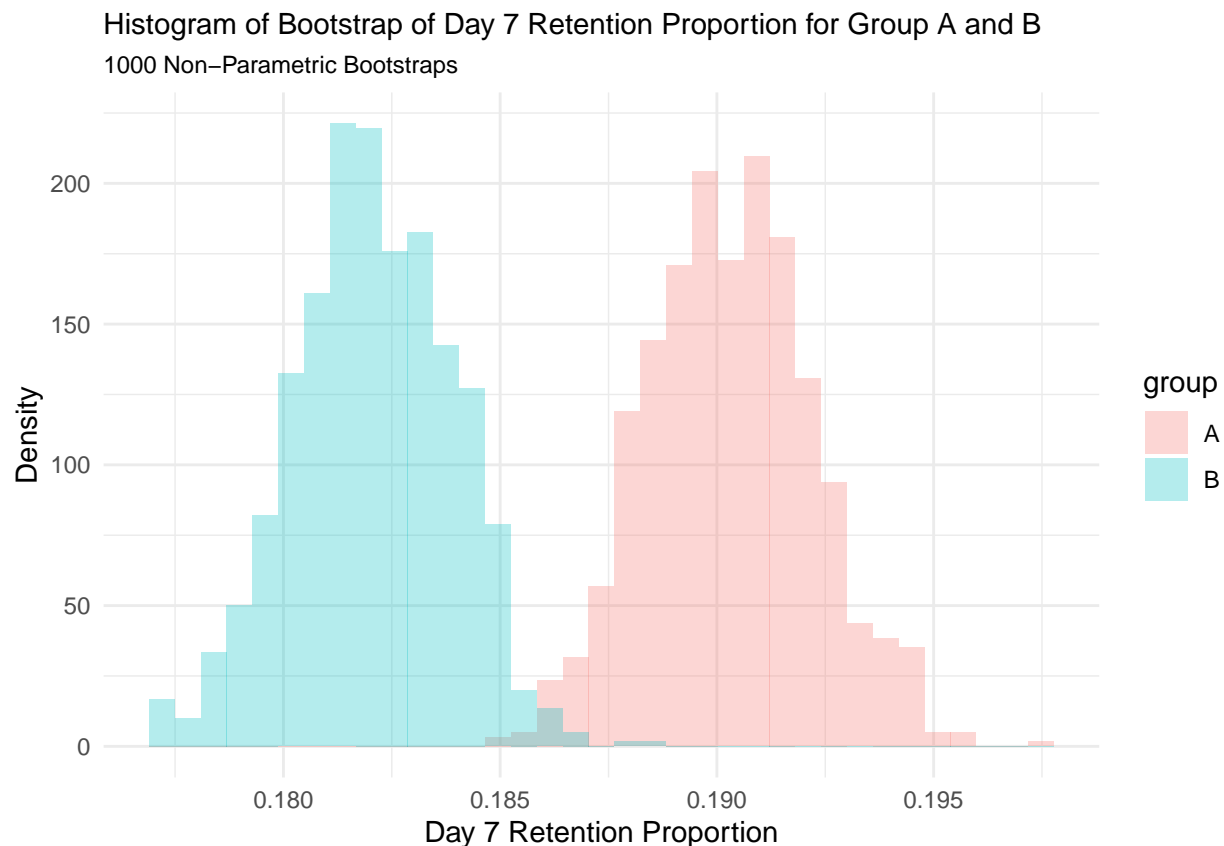
```
    bootstrap_dataB <- sample(groupB$retained_7, dim(groupB)[1], replace = TRUE)
    # compute the sample mean of the bootstrap sample
    data.frame(boot_mean = c(mean(bootstrap_dataA), mean(bootstrap_dataB)), group = c("A",
        "B"))
})

# plot the histogram of the nonparametric bootstrap
ggplot(np_boot_mean_df, aes(x = boot_mean, fill = group)) + geom_histogram(position = "identity",
    aes(y = ..density..), alpha = 0.3, bins = 35) + labs(title = "Histogram of Bootstrap of Day 7 Reten
    subtitle = "1000 Non-Parametric Bootstraps") + xlab("Day 7 Retention Proportion") +
    ylab("Density") + theme(plot.title = element_text(size = 11), plot.subtitle = element_text(size = 9)
```

Histogram of Bootstrap of Day 7 Retention Proportion for Group A and B
1000 Non–Parametric Bootstraps

Based on the plot, I would recommend version A of the game because the mean day 7 proportion for group A is larger than that of group B. My recommendation agrees with my conclusion in part c of the question. In fact, I believe that these findings are more compelling than my 1-day findings because if we look at the histogram, there is less overlap between the Group A and Group B Day-7 Retention Proportion histograms than the Day-1 Retention Proportion histograms. As such, we could be more confident that the difference between group A and group B day-7 retention proportion is not due to chance because there is very little overlap between group A and B's histograms. To confirm this, we would run a hypothesis test.

Problem 3: Hypothesis tests

Part a

The null hypothesis is that the average player plays 52 games in the 14 days after they install the app. The alternative hypothesis is that the number of games played is less than 52 games in the 14 days after they install the app. Since the p-value is essentially zero, it is less than 0.05. As such, we reject the null hypothesis. There is evidence that the average player plays less than 52 games in the 14 days after they installed the app.

```
mu_0 = 52
xbar = mean(game_data$games_played)
n = dim(game_data)[1]
s = sd(game_data$games_played)
alpha = 0.05

z = (xbar - mu_0)/(s/(n^1/2))

pnorm(z) <= 0.05
```

```
## [1] TRUE
```

Part b

I would tell my coworker to interpret the average game play data as "inflated" since the distribution of the number of players resembled an exponential distribution. Since the data was strongly right tailed, the mean is not the best measure of "center" for the distribution, as extreme values would "inflate" the average number of games played.

As for the result of the hypothesis test, our null hypothesis was that the average player plays 52 games in the 14 days after they install the app. The alternative hypothesis was that the number of games played is less than 52 games in the 14 days after they install the app. Our significance level is 0.05, which we can choose ourselves. However, 0.05 is conventional so we used 0.05. Then, we calculated the p-value, which is the probability of observing a test statistic that is equal to or more extreme than our z-score under than null hypothesis. Our p-value was essentially 0, so we have enough evidence to reject the null hypothesis that the average player plays 52 games in the 14 days after they install the app. We have evidence to accept our alternative hypothesis, which states that the average player plays less than less 52 games in the 14 day timespan. Additionally, our results are statistically significant.

Part c

From our visualization, we can observe that distribution of the games player for both group A and group B is roughly normal. It would be reasonable to assume that the variance for both group A and group B's games played is roughly equal. The reasoning for such an assumption is because I am assuming that player preference is still the same regardless of whether they are apart of group A or group B. Additionally, the difference between group A and group B (the location of the challenge) is not a drastic change in the actual game. As such, I will be using a two-sample t-test where the variance is unknown but with pooled variance. Our null hypothesis is that the mean of group A is equal to the mean of group B. Our alternative hypothesis is that the mean of group A is not equal to the mean of group B. Given a significance level of 0.05, since the p-value of 0.3729 is not less than 0.05, we do not have enough evidence to reject the null hypothesis. As such, the mean of group A is equal to the mean of group B. My answer agrees with my conclusion from Problem 1.2 since from the two sample t-test, I would not recommend group A over group B.

```
samp_mean_A <- mean(groupA$games_played)
samp_mean_B <- mean(groupB$games_played)
sd_A <- sd(groupA$games_played)
sd_B <- sd(groupB$games_played)
n <- dim(groupA)[1]
m <- dim(groupB)[1]
sp <- sqrt(((n - 1) * sd_A^2 + (m - 1) * sd_B^2)/(n + m - 2))
df <- n + m - 2
t <- (samp_mean_A - samp_mean_B)/(sp * sqrt(1/n + 1/m))
pvalue <- 2 * (1 - pt(t, df))
pvalue
```

```
## [1] 0.3729087
```

```
t.test(groupA$games_played, groupB$games_played, alternative = "two.sided", var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  groupA$games_played and groupB$games_played
## t = 0.89104, df = 90187, p-value = 0.3729
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.388592  3.703569
## sample estimates:
## mean of x mean of y
##  52.45626  51.29878
```

Part d

Our null hypothesis is that the proportion of players that are retained after 1 day in group A is equal to the proportion of players that are retained after 1 day in group B. The alternative hypothesis is that the proportion of players that are retained after 1 day in group A is NOT equal to the proportion of players that are retained after 1 day in group B. Given a 0.05 significance level, our p value of 0.074 is not less than 0.05. Hence, we do not have enough evidence to reject the null hypothesis. There is not a difference in the proportion of players that are retained after 1 day in group A is equal to the proportion of players that are retained after 1 day in group B.

```
propA <- mean(groupA$retained_1)
propB <- mean(groupB$retained_1)
n <- dim(groupA)[1]
m <- dim(groupB)[1]
p_hat <- (sum(groupA$retained_1) + sum(groupB$retained_1))/(n + m)
sd_null <- sqrt(p_hat * (1 - p_hat) * (1/n + 1/m))
z = (propA - propB)/sd_null
pvalue <- 2 * (1 - pnorm(z))
pvalue
```

```
## [1] 0.07440966
```

Part e

Our null hypothesis is that the proportion of players that are retained after 7 days in group A is equal to the proportion of players that are retained after 7 days in group B. The alternative hypothesis is that the proportion of players that are retained after 7 days in group A is NOT equal to the proportion of players that are retained after 7 days in group B. Given a 0.05 significance level, our p value of 0.001 is not less than 0.05. As such, we have enough evidence to reject the null hypothesis. This means that we have enough evidence to believe that there is a difference in the proportion of players that are retained after 7 days across groups A and B. Additionally, our results are statistically significant.

```
propA <- mean(groupA$retained_7)
propB <- mean(groupB$retained_7)
n <- dim(groupA)[1]
m <- dim(groupB)[1]
p_hat <- (sum(groupA$retained_7) + sum(groupB$retained_7))/(n + m)
sd_null <- sqrt(p_hat * (1 - p_hat) * (1/n + 1/m))
z = (propA - propB)/sd_null
pvalue <- 2 * (1 - pnorm(z))
pvalue
```

```
## [1] 0.00155425
```

Part f

Based on all the analyses I have conducted, I would recommend sticking with version A of the game where the challenge remains at level 10 rather than switched to version B. The reason why is that for almost all situations, the results of version A are identical to version B. For instance, the proportion of players that are retained after day 1 and the average number of games played in 14 days is the same for both version A and B. The only difference is that for the 7-day retention proportion, there is a difference between group A and group B. However, looking at 3d, we can observe that the difference in proportion is roughly 1%. This is a fairly small difference so there are marginal benefits of switching to version B. As such, I recommend that the company should stick with version A of the game.