

Language Learning Platforms

Scenario

Students in a foreign language course choose to use Duolingo or Anki to help study for the final exam. Compare number of hours using the app against final exam score across app type.

Read in data

```
languages = read.table("languages.csv", header = TRUE, sep = ",")
duolingo = languages[which(languages$app == "Duolingo"),]
anki = languages[which(languages$app == "Anki"),]
colnames(anki)
```

```
## [1] "hours" "final" "app"
```

```
colnames(duolingo)
```

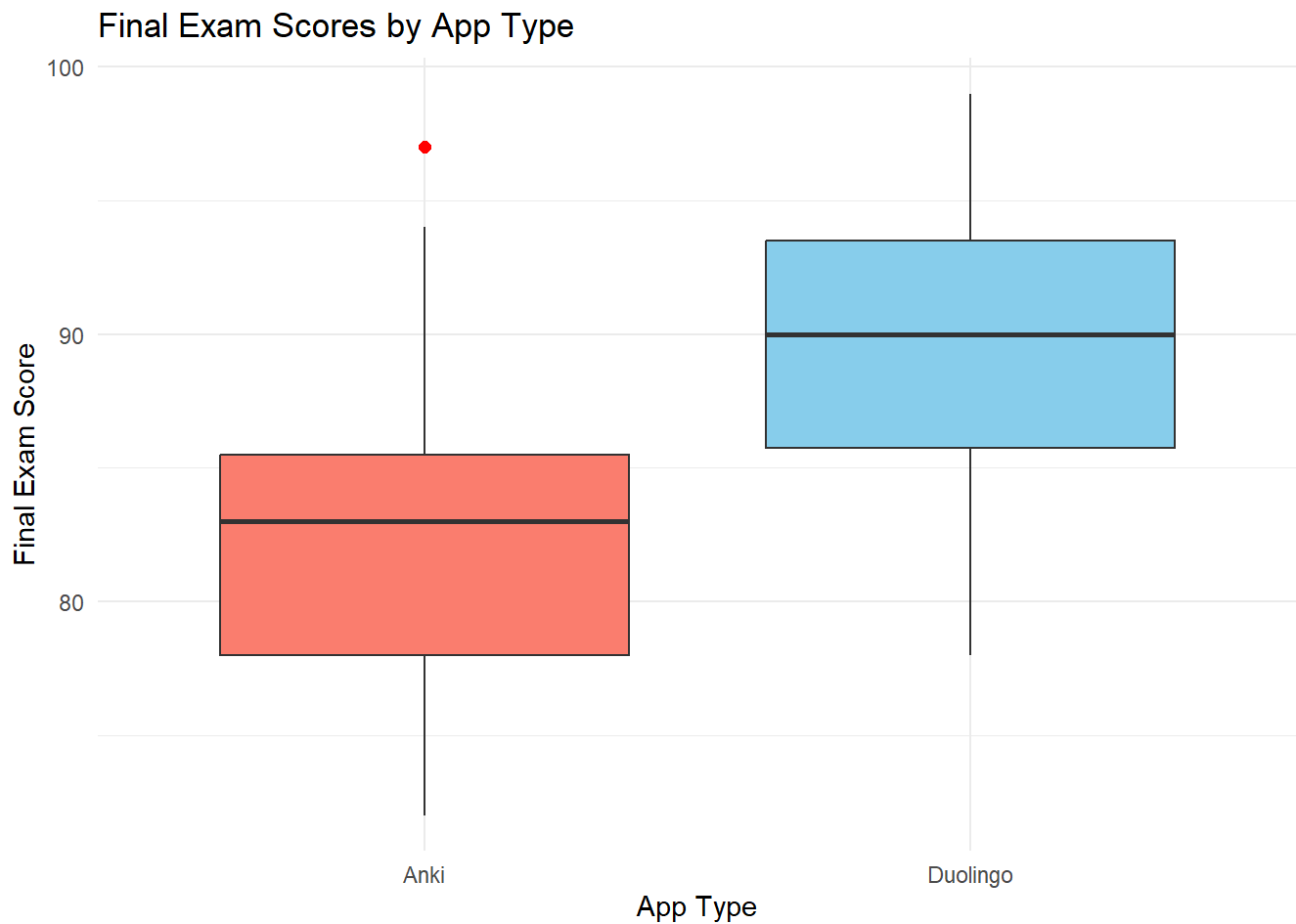
```
## [1] "hours" "final" "app"
```

Boxplots and two-sample t-test

```
# Load necessary libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
# Create a more attractive boxplot
ggplot(languages, aes(x = app, y = final, fill = app)) +
  geom_boxplot(outlier.colour = "red", outlier.shape = 16, outlier.size = 2) +
  labs(title = "Final Exam Scores by App Type",
       x = "App Type",
       y = "Final Exam Score") +
  scale_fill_manual(values = c("Duolingo" = "skyblue", "Anki" = "salmon")) +
  theme_minimal() +
  theme(legend.position = "none")
```



T Test

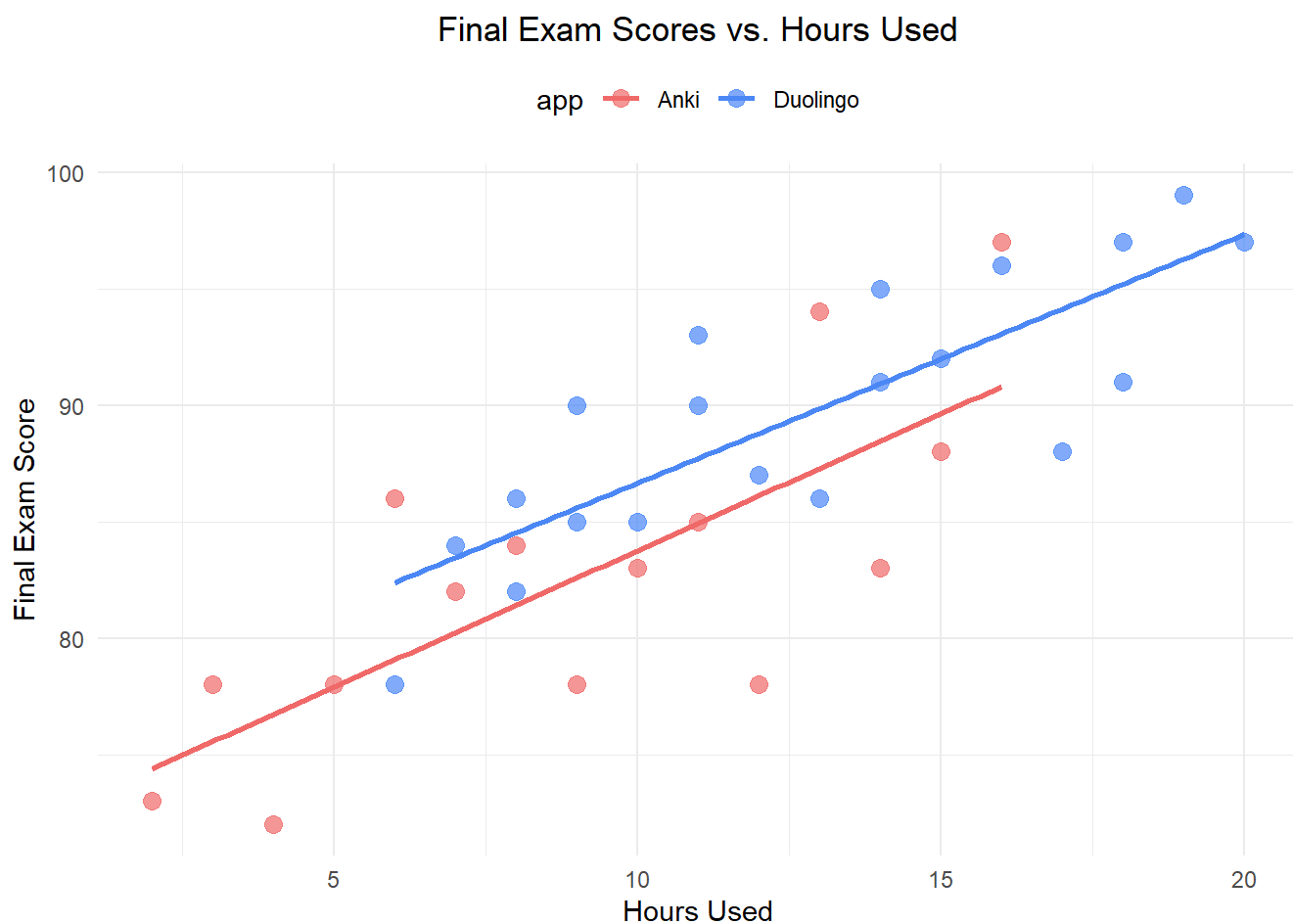
```
t.test(duolingo$final, anki$final, alternative = "two.sided", var.equal = FALSE)
```

```
##  
##  Welch Two Sample t-test  
##  
## data:  duolingo$final and anki$final  
## t = 3.2044, df = 26.279, p-value = 0.003534  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
##   2.512089 11.487911  
## sample estimates:  
## mean of x mean of y  
##      89.6      82.6
```

Scatterplot

```
ggplot(languages, aes(x = hours, y = final, color = app)) +
  geom_point(size = 3, alpha = 0.7) + # Adjust point size and transparency
  geom_smooth(method = "lm", se = FALSE, aes(color = app), linetype = "solid") + # Add solid regression lines
  labs(title = "Final Exam Scores vs. Hours Used",
        x = "Hours Used",
        y = "Final Exam Score") +
  scale_color_manual(values = c("Duolingo" = "#4C8BF9", "Anki" = "#F26B6B")) + # Softer colors
  theme_minimal() +
  theme(legend.position = "top", # Move the legend to the top
        plot.title = element_text(hjust = 0.5)) # Center the title
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Final exam score - Test for appropriateness of ANACOVA

```
languages$duolingo = ifelse(languages$app == "Duolingo", 1, 0)
full = lm(final ~ hours + duolingo + hours*duolingo, data = languages)
reduced = lm(final ~ hours + duolingo, data = languages)
anova(full)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
hours	1	1142.865956	1142.865956	73.1432913	1.166507e-09
duolingo	1	57.445279	57.445279	3.6764913	6.443980e-02
hours:duolingo	1	1.712905	1.712905	0.1096257	7.427984e-01
Residuals	31	484.375860	15.625028	NA	NA
4 rows					

```
anova(reduced)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
hours	1	1142.86596	1142.86596	75.236692	6.524577e-10
duolingo	1	57.44528	57.44528	3.781714	6.064903e-02
Residuals	32	486.08877	15.19027	NA	NA
3 rows					

Final exam score - Test for equality of adjusted means

```
full = lm(final ~ hours + duolingo, data = languages)
reduced = lm(final ~ hours, data = languages)
anova(full)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
hours	1	1142.86596	1142.86596	75.236692	6.524577e-10
duolingo	1	57.44528	57.44528	3.781714	6.064903e-02
Residuals	32	486.08877	15.19027	NA	NA
3 rows					

```
anova(reduced)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
hours	1	1142.866	1142.86596	69.3877	1.272165e-09
Residuals	33	543.534	16.47073	NA	NA
2 rows					

Results

After an initial analysis Duolingo users appear to perform better on average on the test, Duolingo users also studied for more hours on average. This longer time studying could be a result of the more enjoyable gamification of learning that Duolingo implements, and be a testament to their systems. Going deeper to discover if Duolingo still outperforms Anki on an hour basis we conduct a test. After adjusting the mean scores of both groups (Duolingo and Anki), Duolingo users still significantly outperform Anki Users. This indicates that on a per hour return basis Duolingo is superior to Anki, and overall the interface and learning curriculum implemented by Duolingo could contribute to users spending more time overall studying. These 2 effects combined show the superiority of Duolingo's platform to learning languages.