# Lab 6

Pranav Bhagat

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────────────────────────────────────── tidyverse
1.2.1 ──
```

```
## ✔ ggplot2 3.2.1     ✔ purrr   0.3.2
## ✔ tibble  2.1.3     ✔ dplyr   0.8.3
## ✔ tidyr   1.0.0     ✔ stringr 1.4.0
## ✔ readr   1.3.1     ✔ forcats 0.4.0
```

```
## ── Conflicts ────────────────────────────────────────────────────────────────────────── tidyverse_confl
icts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
horror_movies <- readr::read_csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/
2019/2019-10-22/horror_movies.csv")
```

```
## Parsed with column specification:
## cols(
##   title = col_character(),
##   genres = col_character(),
##   release_date = col_character(),
##   release_country = col_character(),
##   movie_rating = col_character(),
##   review_rating = col_double(),
##   movie_run_time = col_character(),
##   plot = col_character(),
##   cast = col_character(),
##   language = col_character(),
##   filming_locations = col_character(),
##   budget = col_character()
## )
```

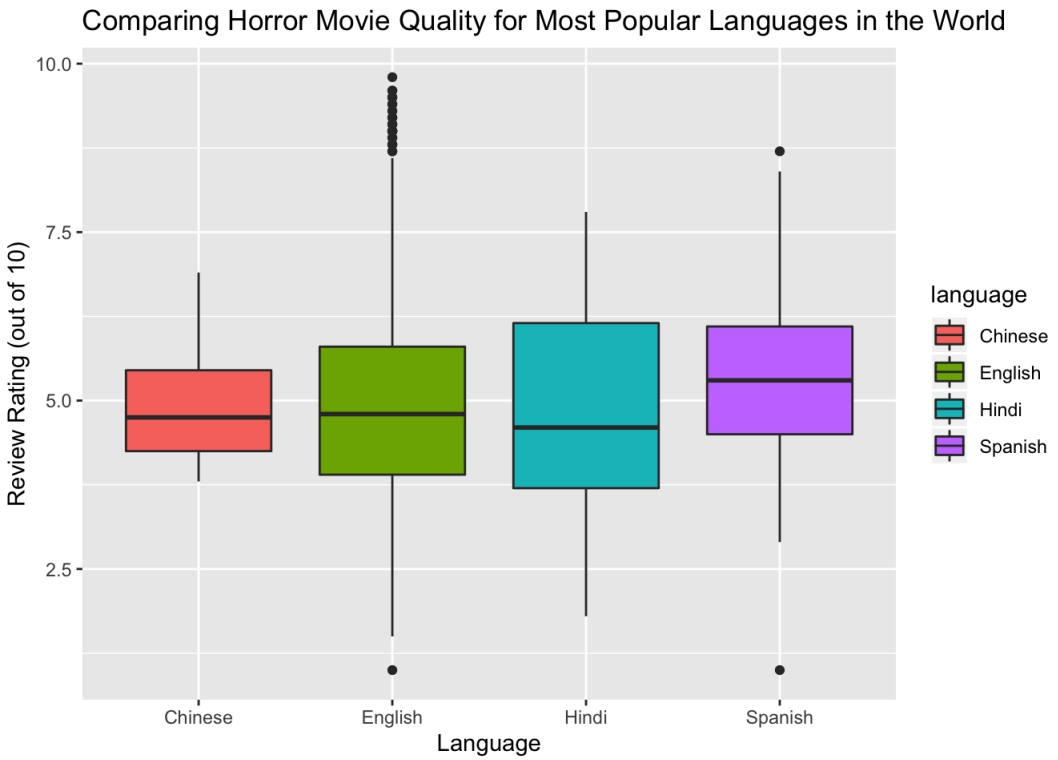
```
head(horror_movies)
```

```
## # A tibble: 6 x 12
##   title genres release_date release_country movie_rating review_rating
##   <chr> <chr>  <chr>        <chr>           <chr>                <dbl>
## 1 Gut … Drama… 26-Oct-12    USA             <NA>                   3.9
## 2 The … Horror 13-Jan-17    USA             <NA>                  NA
## 3 Slee… Horror 21-Oct-17    Canada          <NA>                  NA
## 4 Trea… Comed… 23-Apr-13    USA             NOT RATED              3.7
## 5 Infi… Crime… 10-Apr-15    USA             <NA>                   5.8
## 6 In E… Horro… 2017         UK              <NA>                  NA
## # … with 6 more variables: movie_run_time <chr>, plot <chr>, cast <chr>,
## #   language <chr>, filming_locations <chr>, budget <chr>
```

```
Bhaasha <- filter(horror_movies, language == "English" | language == "Hindi" | language == "Chinese" | langu
age == "Spanish")

ggplot(Bhaasha, aes(x = as.factor(language), y = review_rating, fill = language)) +
  geom_boxplot() +
  ggtitle("Comparing Horror Movie Quality for Most Popular Languages in the World") +
  xlab("Language") +
  ylab("Review Rating (out of 10)")
```

```
## Warning: Removed 211 rows containing non-finite values (stat_boxplot).
```




This graph illustrates the differences in horror movie ratings when we look at the most popular languages in the world. Spanish films had the highest median rating, while Hindi had the lowest, which I'm not super surprised by since being Indian I've watched a lot of Hindi films growing up with my parents and I never really liked any of them. Also, English films and Hindi films have the highest variability, which makes sense due to the presence of Hollywood and Bollywood in the US and India, respectively, leading to a lot of movies being made. And with a large quantity of movies being made, there are a lot of movies that are bad and those that are good, and then most are in between. In addition, the US has a lot of outliers with really high review ratings, indicating that the US makes a lot of really good horror movies.