Bootstrap Methods and Applications

A data-based journey through two U.S. sitcoms

Jacob Forbes

Friday, November 18, 2022

1 Data exploration

2 Statistical analysis

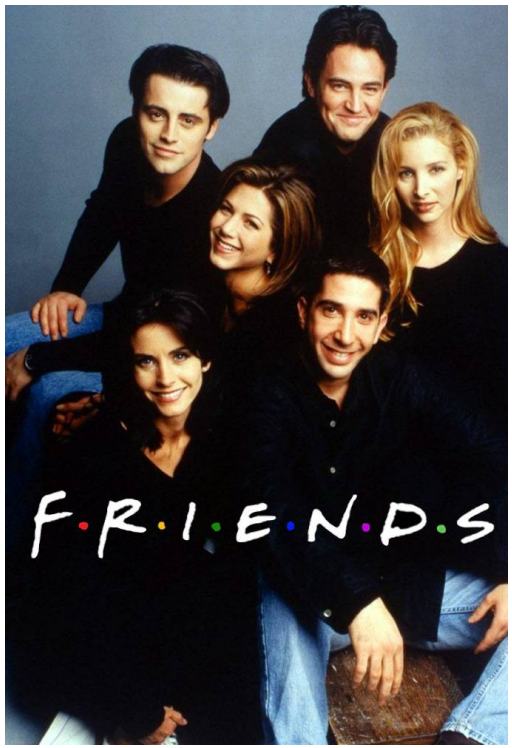
Data exploration

Two TV shows

We're going to explore data from two famous U.S. sitcoms that together aired from 1994 to 2013. Any guesses?

Two TV shows

We're going to explore data from two famous U.S. sitcoms that together aired from 1994 to 2013. Any guesses?



The data

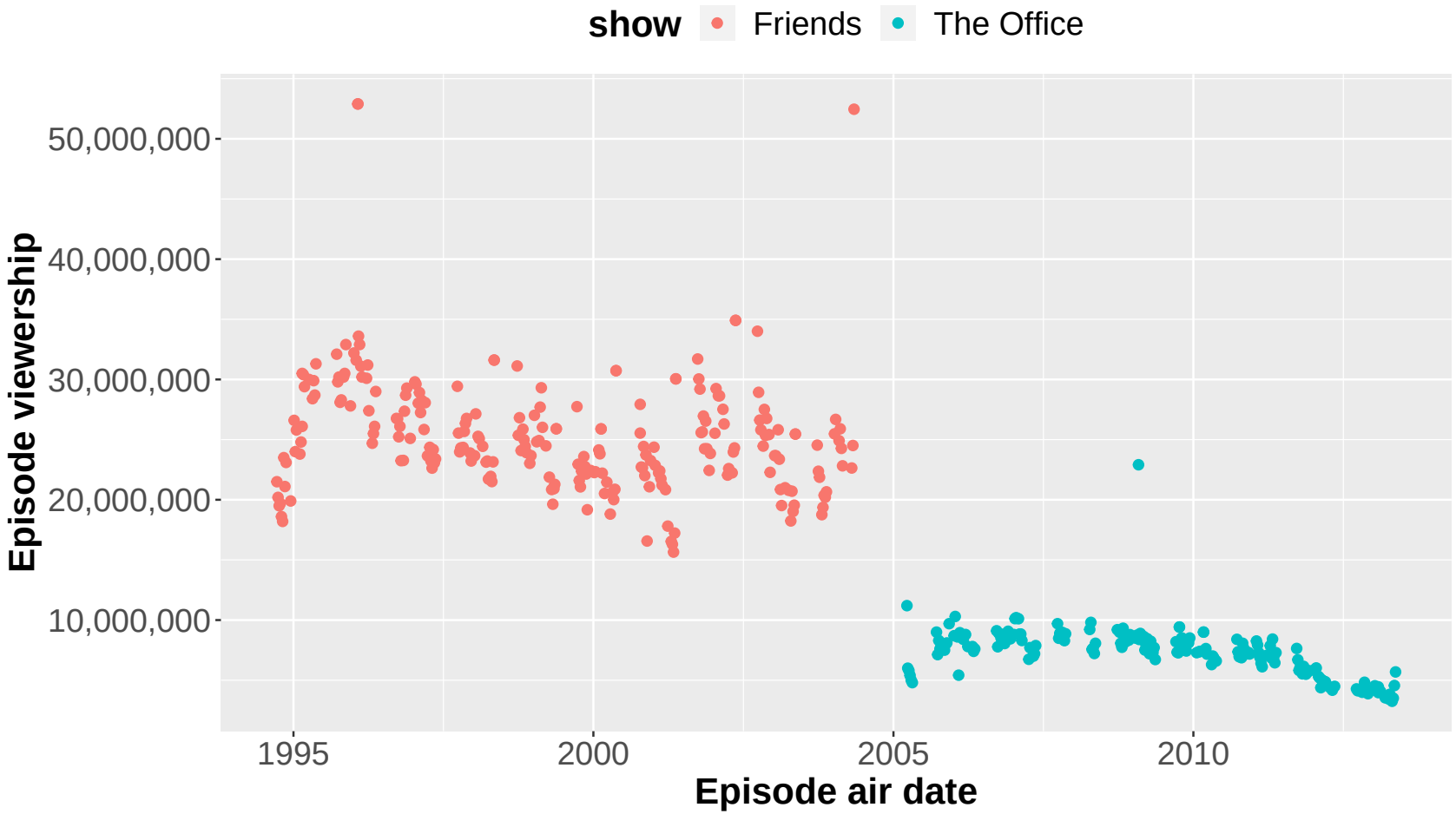
Our data come from the Friends Series Dataset and The Office Episodes Data both available on Kaggle. The data were combined into a `.csv` then read in as a data frame in `R` for analysis. Let's take a look at the columns we have available to analyze.

The data

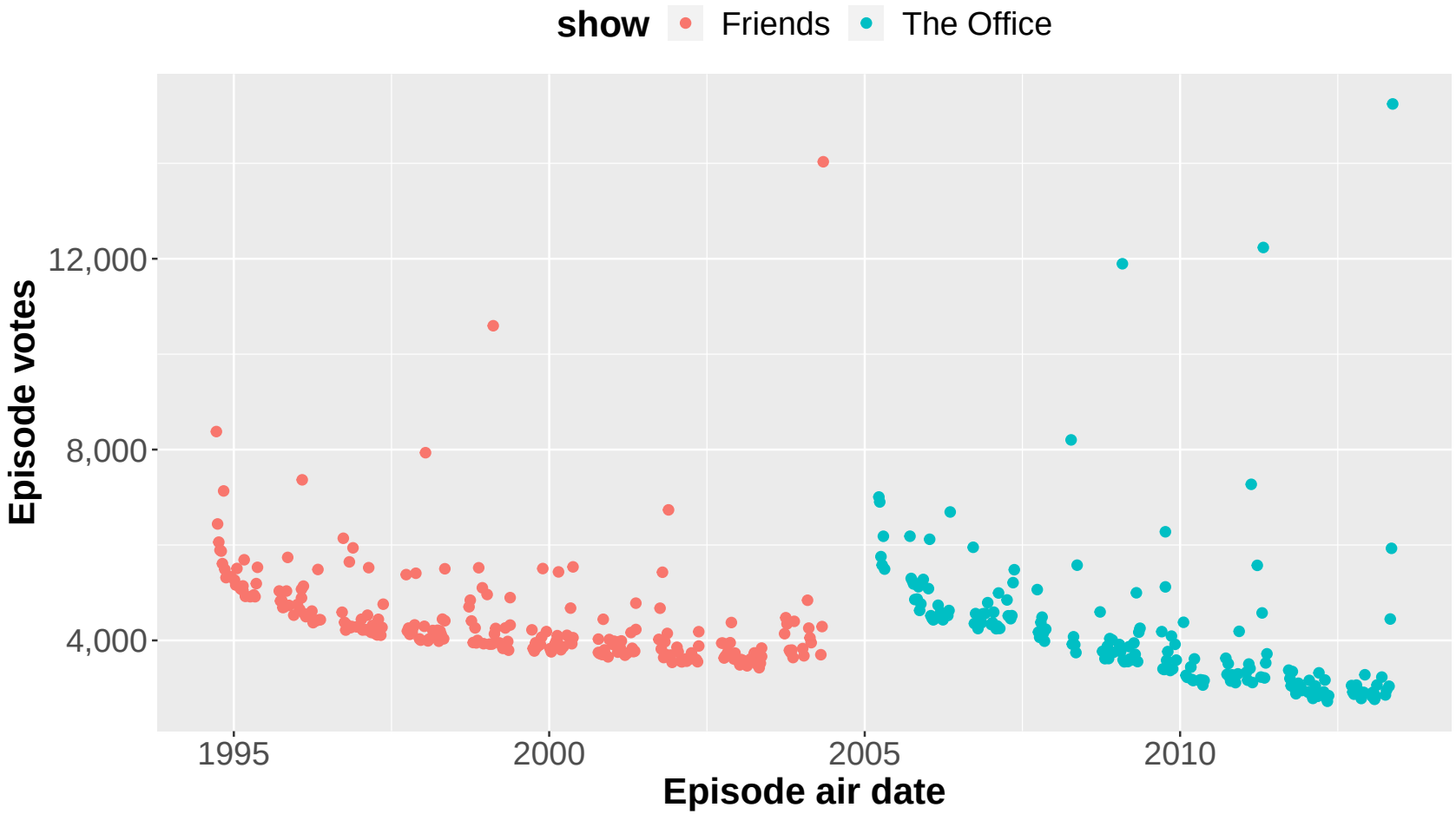
Our data come from the Friends Series Dataset and The Office Episodes Data both available on Kaggle. The data were combined into a `.csv` then read in as a data frame in `R` for analysis. Let's take a look at the columns we have available to analyze.

```
## 'data.frame':    423 obs. of  9 variables:
##  $ show       : Factor w/ 2 levels "Friends","The Office": 2 2 2 ...
##  $ season     : int  1 1 1 ...
##  $ episode    : int  1 2 3 ...
##  $ title      : chr  "Pilot" ...
##  $ us_viewers : int  11200000 6000000 5800000 ...
##  $ air_date   : Date, format: "2005-03-24" ...
##  $ imdb_rating: num  7.4 8.3 7.7 ...
##  $ total_votes: int  7006 6902 5756 ...
##  $ description: chr  "The premiere episode introduces the boss and st"| __truncated__ ...
```

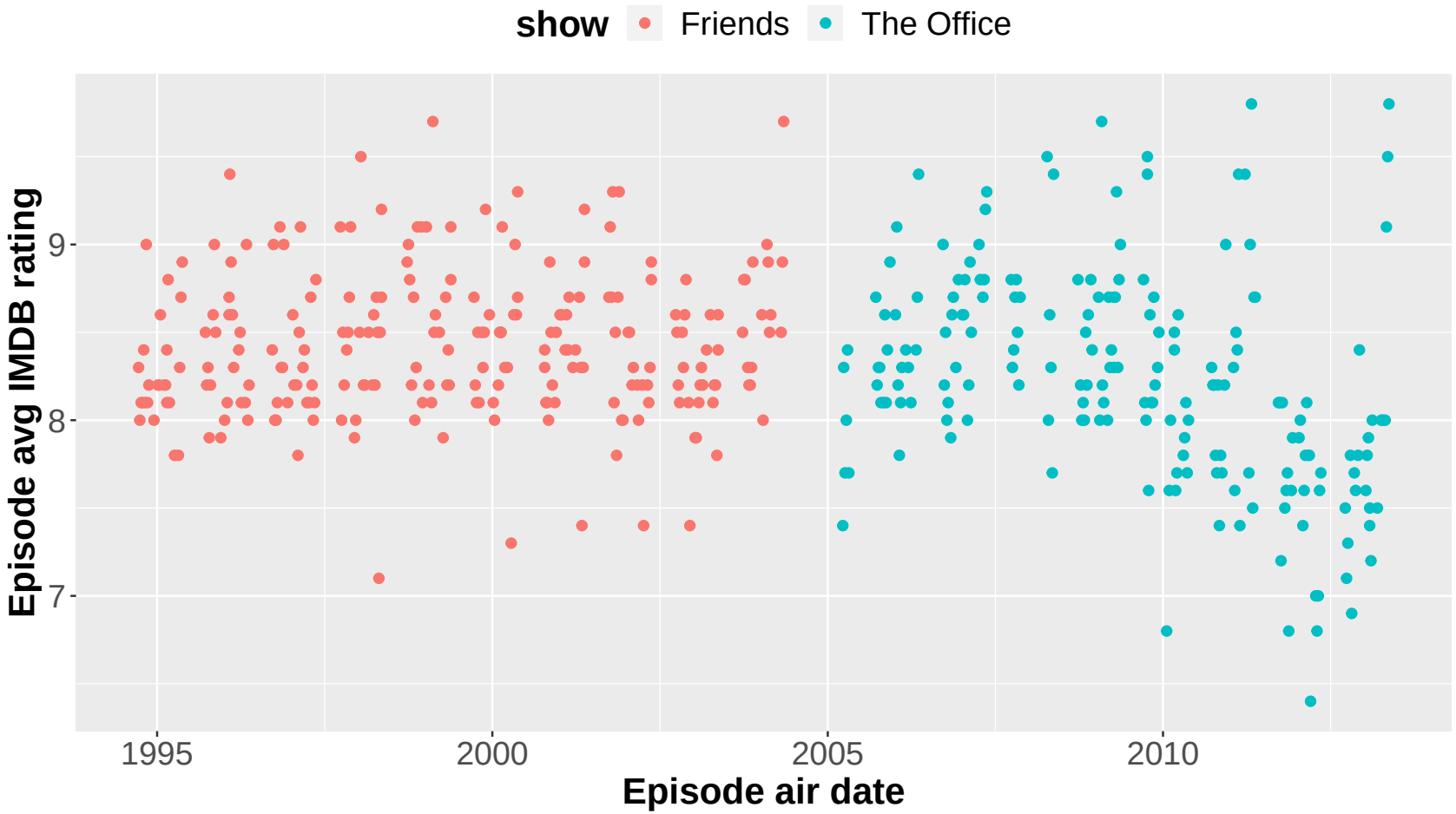
U.S. viewership per episode



Number of IMDB votes per episode



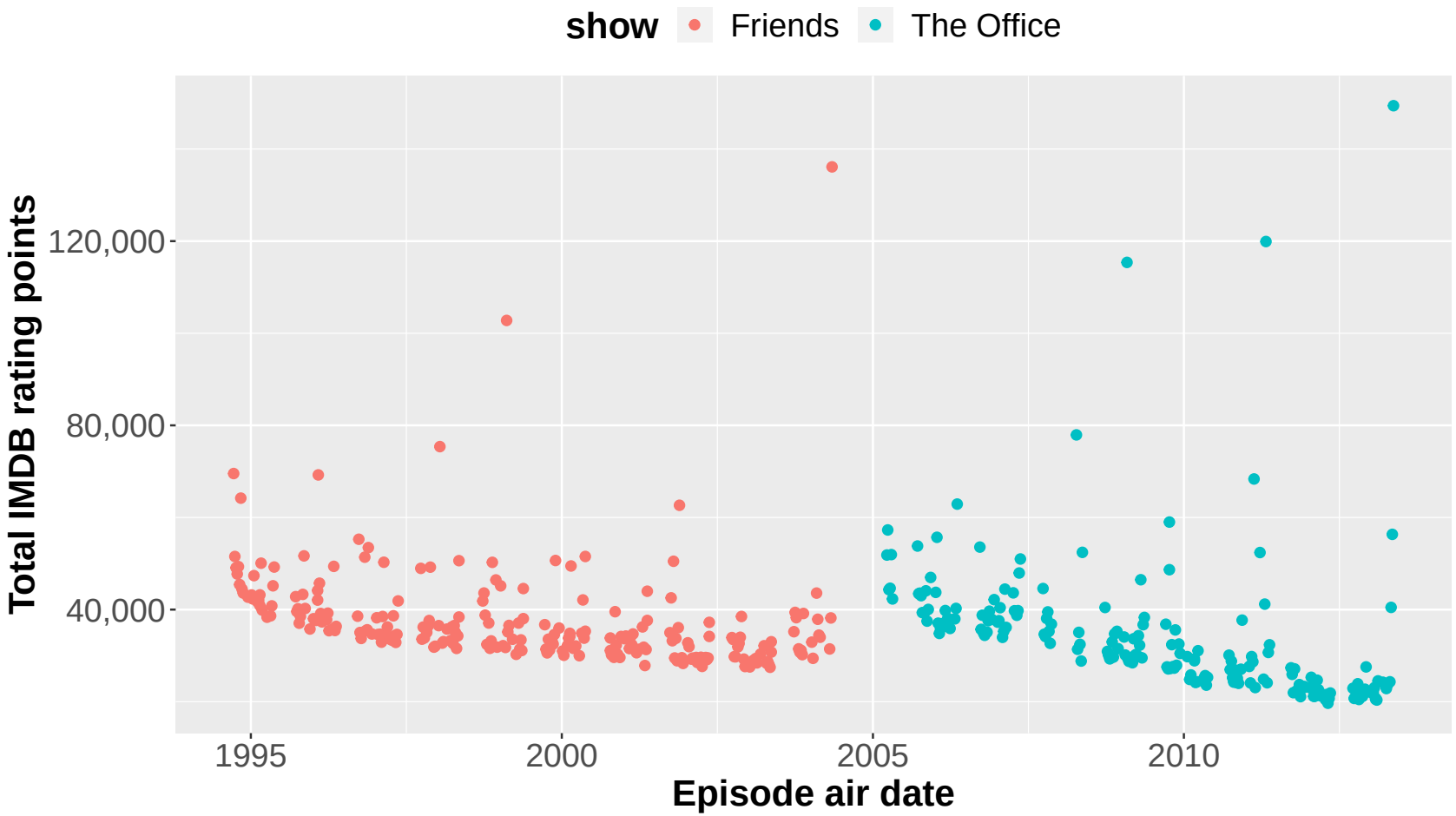
Average IMDB rating per episode



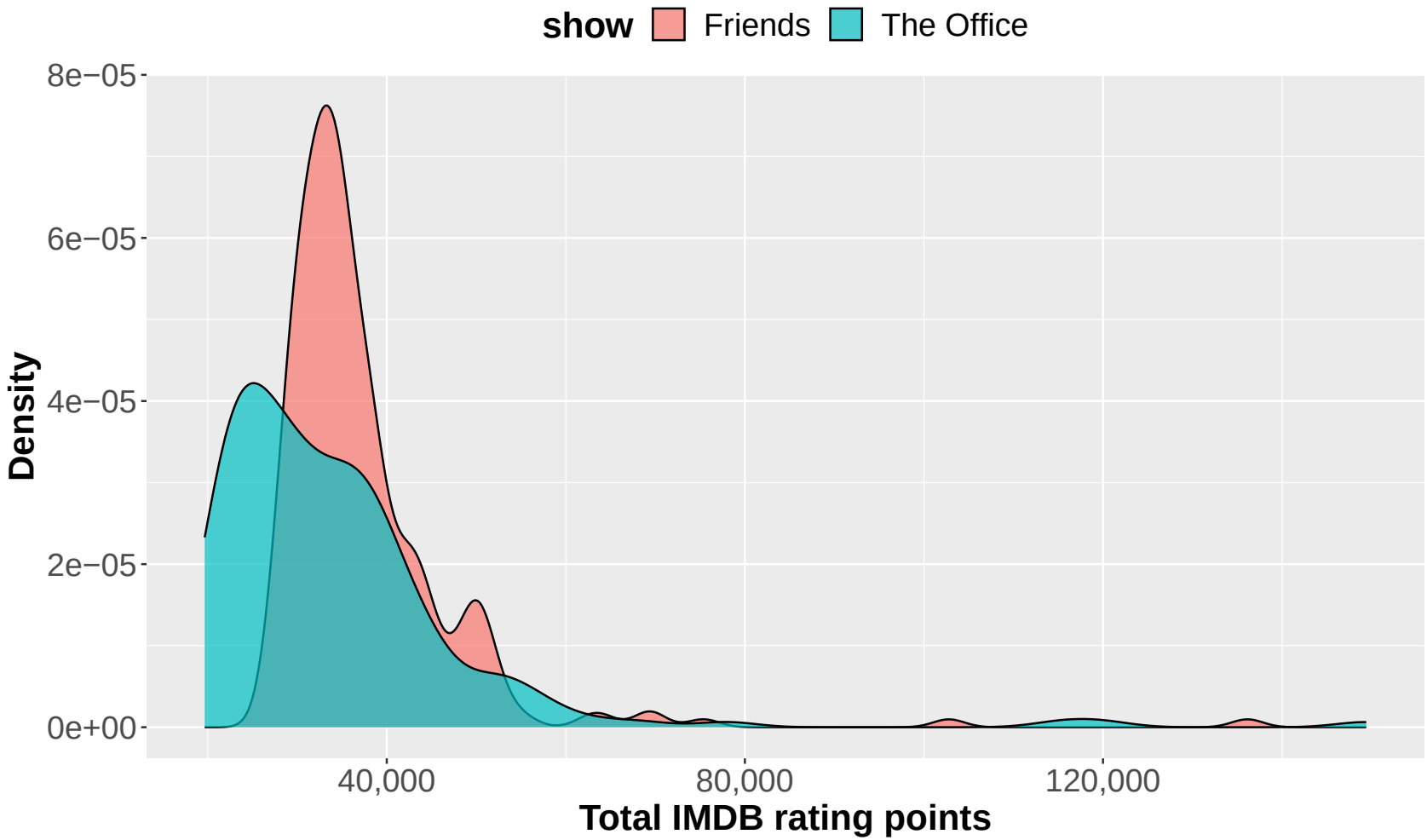
Relationships between viewership, votes, and ratings

Statistical analysis

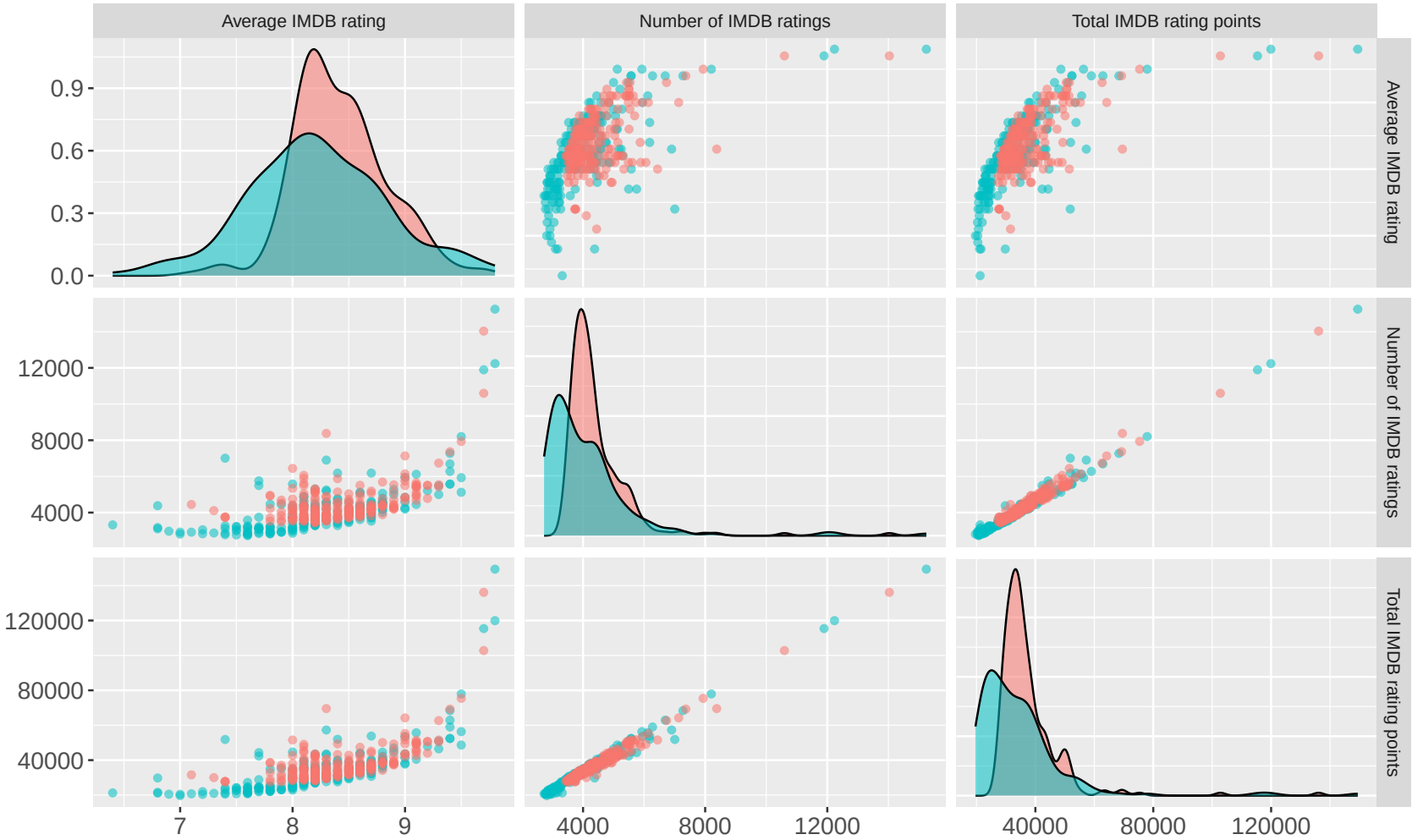
Sandbox total IMDB



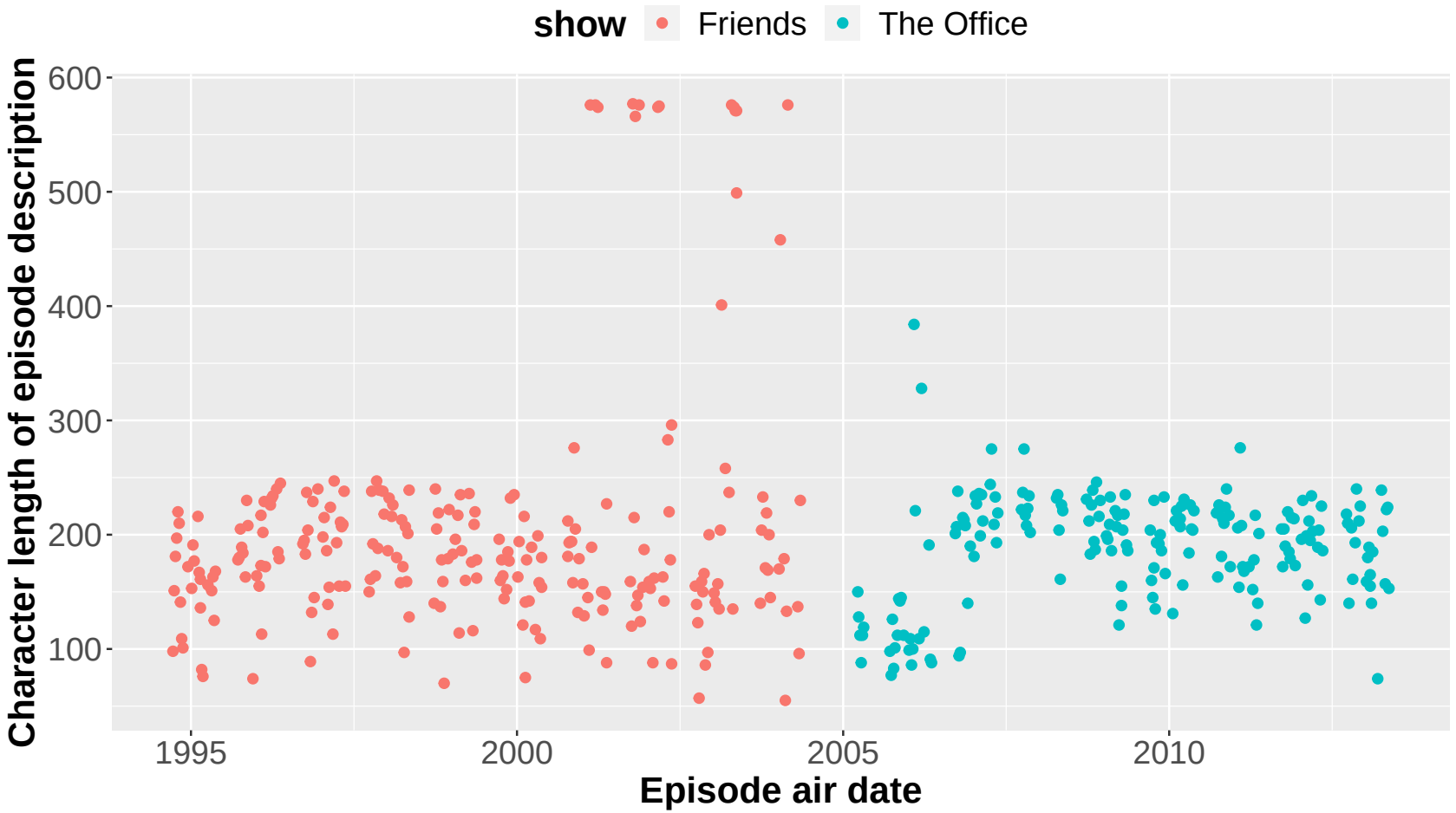
Sandbox total IMDB



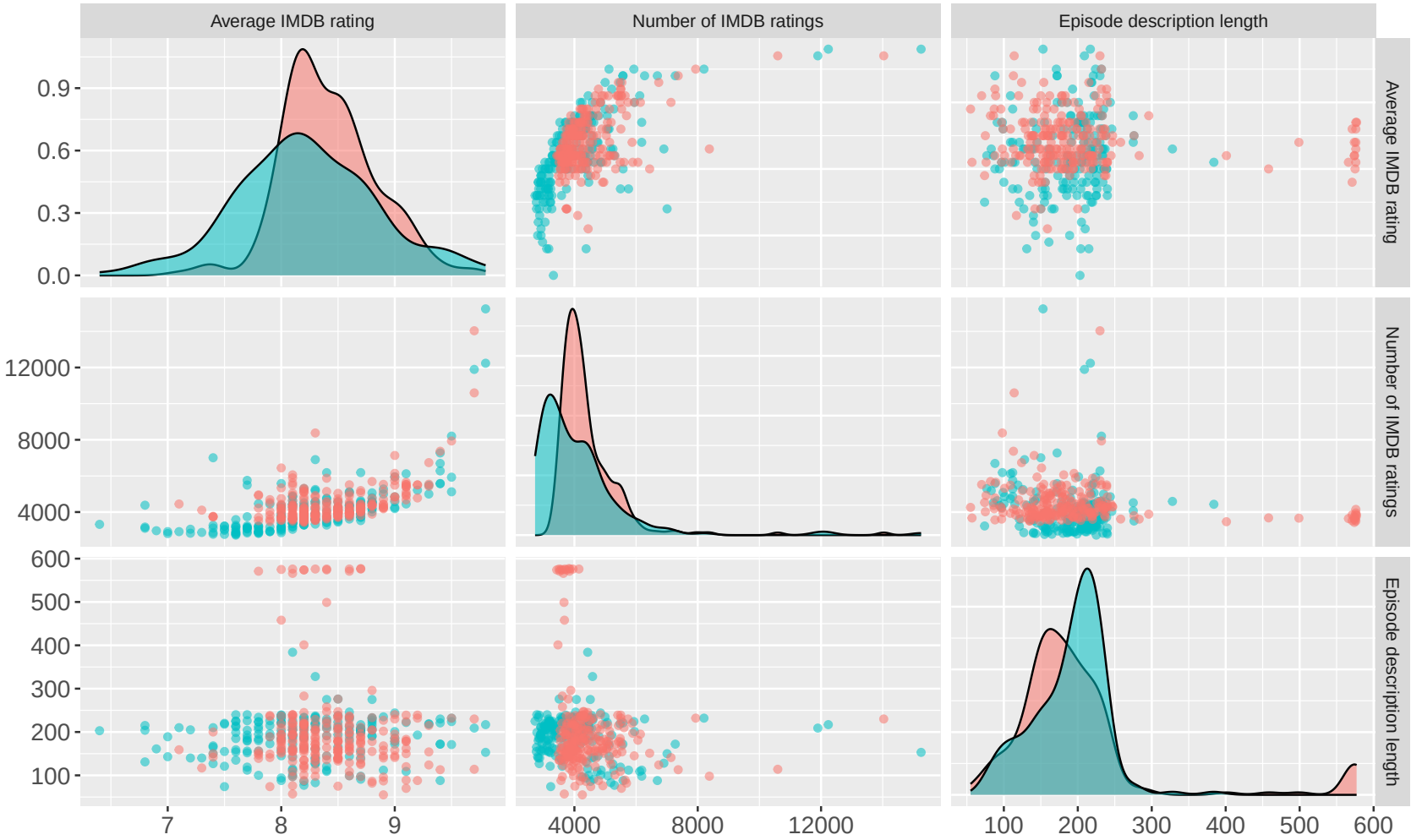
Sandbox pairs



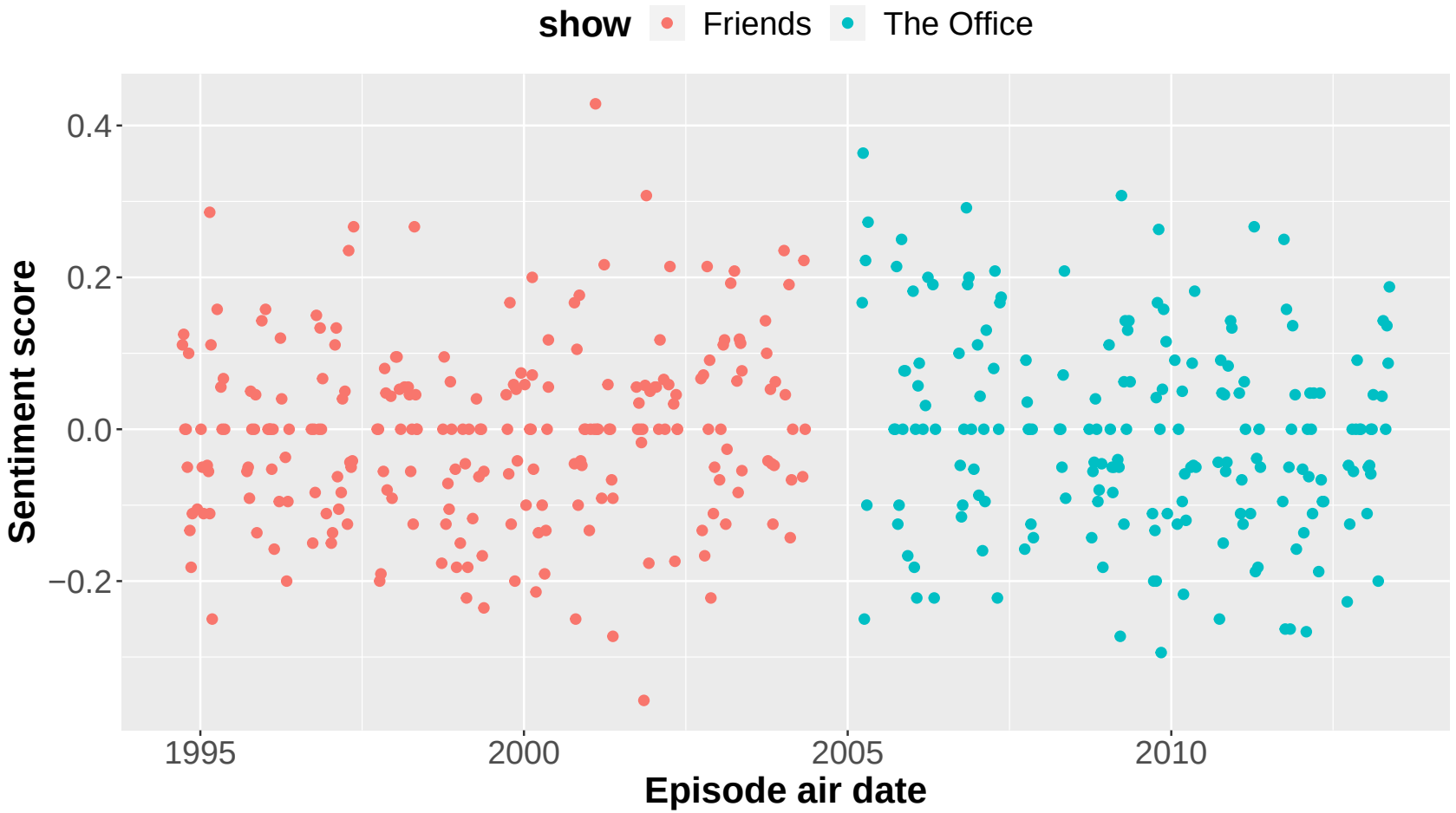
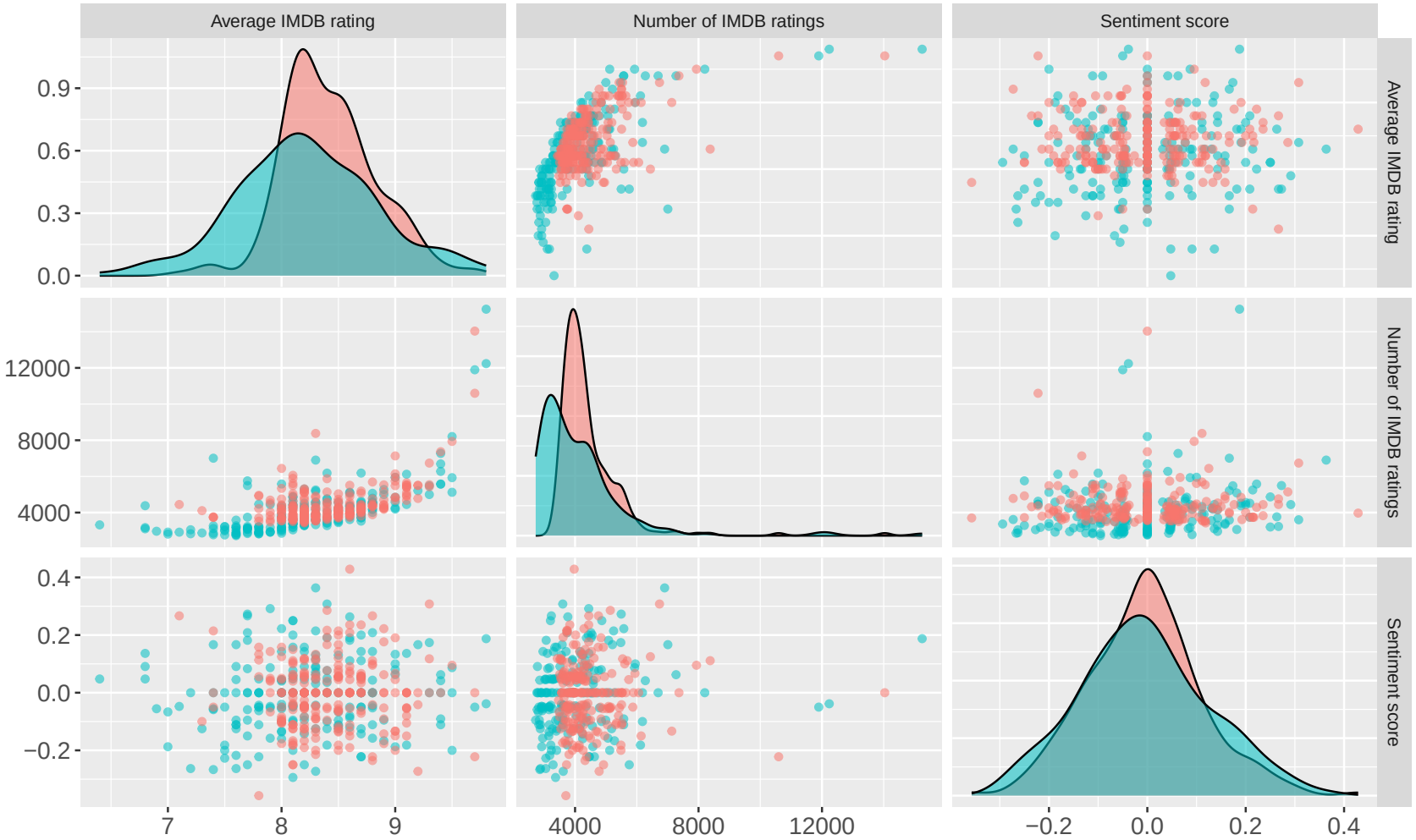
Sandbox description



Sandbox description pairs



Sandbox description



Sandbox sentiment score