

Bootstrap Methods and Applications

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

Jacob Forbes

Friday, November 18, 2022

1 Data exploration

Data exploration

Two TV shows

We're going to explore data from a famous U.S. sitcom that together aired from 2005 to 2013. Any guesses?

Two TV shows

We're going to explore data from a famous U.S. sitcom that together aired from 2005 to 2013. Any guesses?



The data

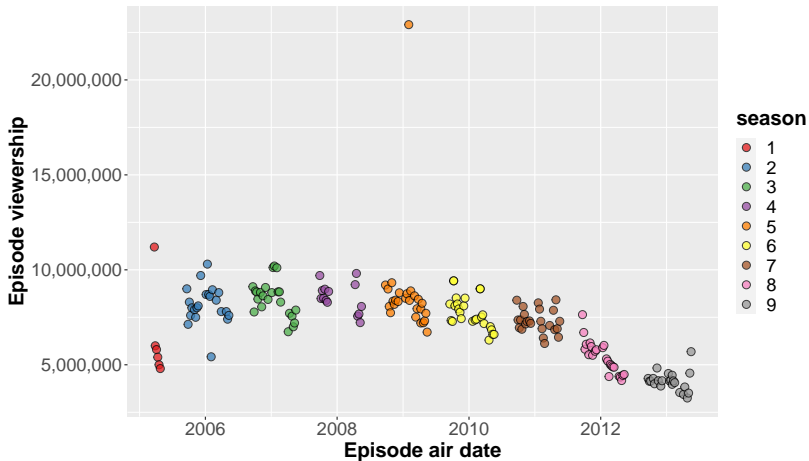
Our data come from the [The Office Episodes Data](#) which is available on Kaggle. The data were read in as a data frame into R for analysis. Let's take a look at the columns we have available to work with.

The data

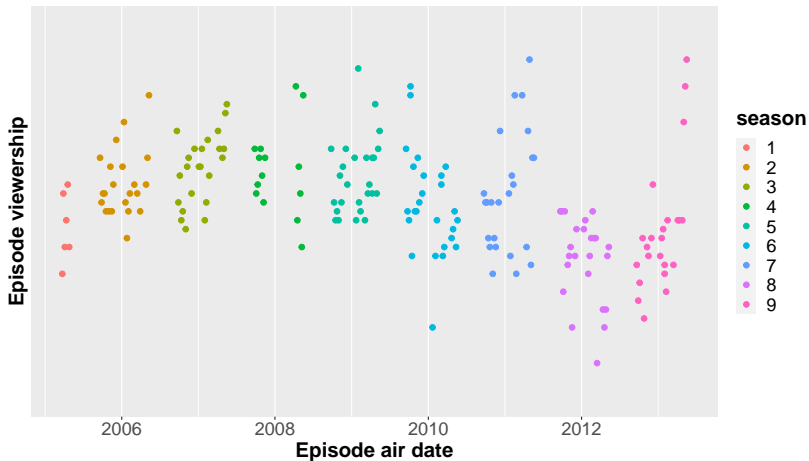
Our data come from the [The Office Episodes Data](#) which is available on Kaggle. The data were read in as a data frame into R for analysis. Let's take a look at the columns we have available to work with.

```
## 'data.frame': 188 obs. of 8 variables:
## $ season : Factor w/ 9 levels "1","2","3","4",...: 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



Average IMDB rating per episode



Relationships between viewership, votes, and ratings

