Bootstrap Methods and Applications

A data-based journey through a U.S. sitcom

Jacob Forbes

Sobes Research Group

The sitcom

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

The sitcom

We're going to explore data from a famous U.S. sitcom that 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
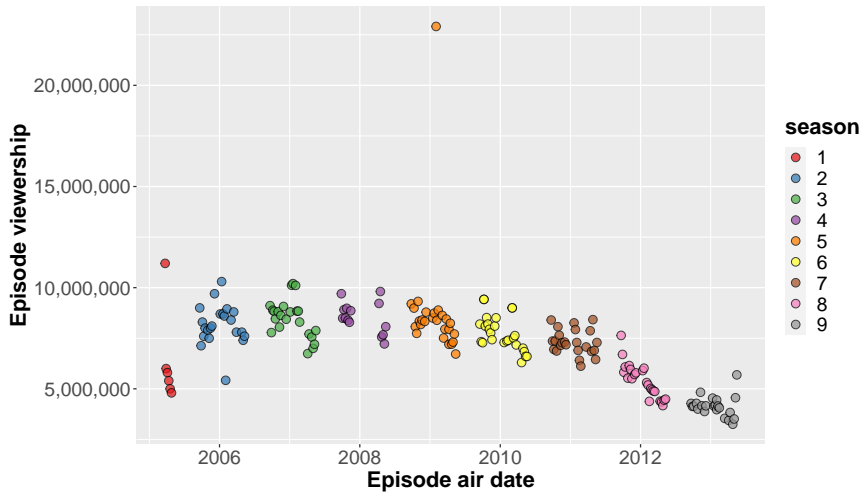columns we have available to work with.

```
## 'data.frame':    188 obs. of  8 variables:
##  $ season      : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 ...
##  $ episode     : int  1 2 3 4 ...
##  $ title       : chr  "Pilot" ...
##  $ us_viewers  : int  11200000 6000000 5800000 5400000 ...
##  $ air_date    : Date, format: "2005-03-24" ...
##  $ imdb_rating : num  7.4 8.3 7.7 8 ...
##  $ total_votes : int  7006 6902 5756 5579 ...
##  $ description : chr  "The premiere episode introduces the boss and staff of the Dunder-Mifflin Paper "| __truncated__ ...
```
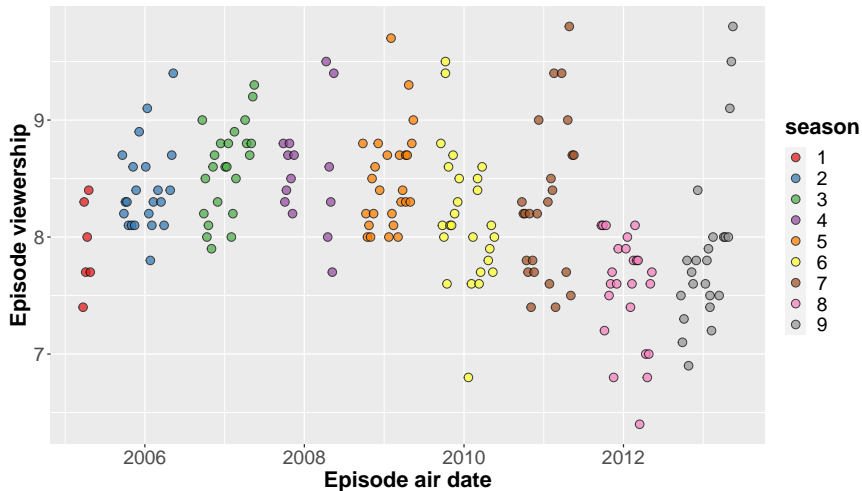
The data

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 1 ...
## $ episode    : int  1 2 3 4 ...
## $ title      : chr  "Pilot" ...
## $ us_viewers : int  11200000 6000000 5800000 5400000 ...
## $ air_date   : Date, format: "2005-03-24" ...
## $ imdb_rating: num  7.4 8.3 7.7 8 ...
## $ total_votes: int  7006 6902 5756 5579 ...
## $ description: chr  "The premiere episode introduces the boss and staff of the Dunder-Mifflin Paper "| __truncated__ ...
```

U.S. viewership per episode

Average IMDB rating per episode

Table of Contents

Data exploration