

Investigating the Relationship Between Ticket Prices & Seats Sold for The Lion King & Hamilton*

Reem Alasadi

20 April 2021

Abstract

A linear regression model is used to analyze the relationship between ticket prices and seats sold. Results show that ...

*Code and data are available at: github.com/alasadir/broadway

Table 1:

week_ending	week_number	weekly_gross_overall	show	theatre	weekly_gross	potential_g
1997-10-19	20	10541144	The Lion King	Minskoff Theatre	630828	69

1 Introduction

I am a modest musical fan. I know a good number of musicals thanks to Disney and my theater-fanatic friend from high school, Hadeel. My 2 favorite theatrical musicals are The Lion King and Hamilton. While browsing Alex Cookson’s blog (Cookson, n.d.), his take on the most successful Broadway show of all time grabbed my attention. He analyzes box office returns to determine the highest grossing show. I will build on Alex’s work and will try to determine whether the average ticket price should remain as-is, or can go up as the top ticket price. I will run a linear regression model to see how ticket prices correlate with the number of seats sold during performances. Linear regression should quantify the strength of the relationship between the ticket price and the seats sold.

2 Data

The dataset includes information about many Broadway shows from the very first week they started showing until the latest performance on March 8, 2020 (all shows were stopped on that day because of COVID-19 pandemic). The dataset includes information like the number of weeks the show has been running, the overall gross, the theater the show was performed at, the number of seats sold, average and top ticket prices, etc. The information has been scraped off of Playbill website, which uses data from The Broadway League’s official website. The Broadway League is the official entity that oversees all theatrical shows on Broadway.

The data will be analyzed using **R** (R Core Team 2020), a statistical programming language, and packages that allow easier data manipulation: **tidyverse** (Wickham et al. 2019), **lubridate** (Grolemund and Wickham 2011), **knitr** (Xie 2021). Data will be visualized using the **ggplot2** (Wickham 2016) package.

The original dataset contains 47,524 observations and 14 variables. Since my focus will be on The Lion King and Hamilton shows only, the dataset has been filtered and now has 1,407 observations and 14 variables. Here’s a preview of how it looks like:

Below is a list of the variable names:

```
## [1] "week_ending"      "week_number"      "weekly_gross_overall"
## [4] "show"             "theatre"          "weekly_gross"
## [7] "potential_gross"  "avg_ticket_price" "top_ticket_price"
## [10] "seats_sold"       "seats_in_theatre" "pct_capacity"
## [13] "performances"     "previews"
```

The Lion King is considered one of the longest and most successful Broadway shows in history [insert citation here]. It has been running since July 1997 (over 20 years) and grossed over \$2 billion USD. Hamilton, on the other hand, first debuted in October 2015, but has gained more popularity than The Lion King. In fact, in the 6 years of running, Hamilton grossed over \$500 million USD. That is 25% of what The Lion King made in X years. One might think that Hamilton made more money due to performing more times a week to compensate for the 18 year difference, but figure 1 exhibits that both shows have consistently maintained 8 performances per week since their first debut:

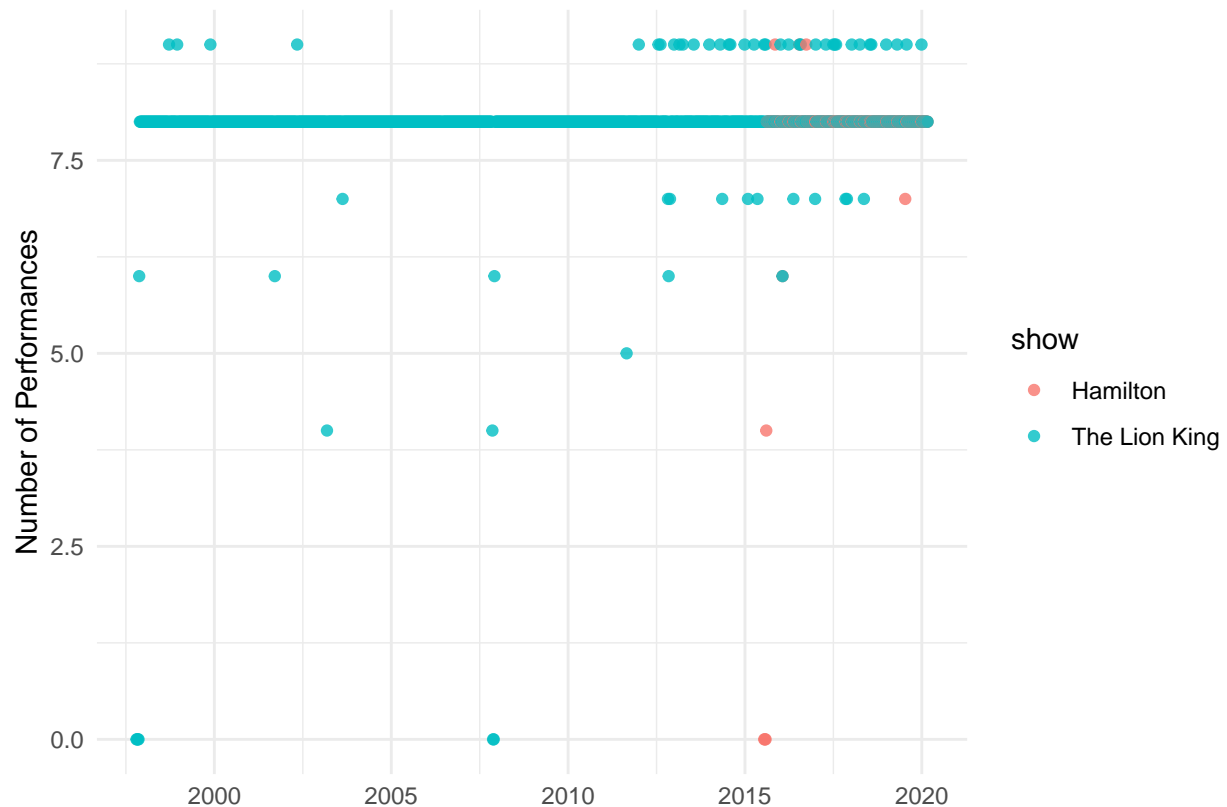


Figure 1: Performances per week

Another assumption could be that Hamilton performed at a bigger theater and therefore sold more seats. However, figure 2 shows that that is not the case. Hamilton has consistently sold 10,000 seats at average ticket price and The Lion King sold between 13 - 15 thousand seats.

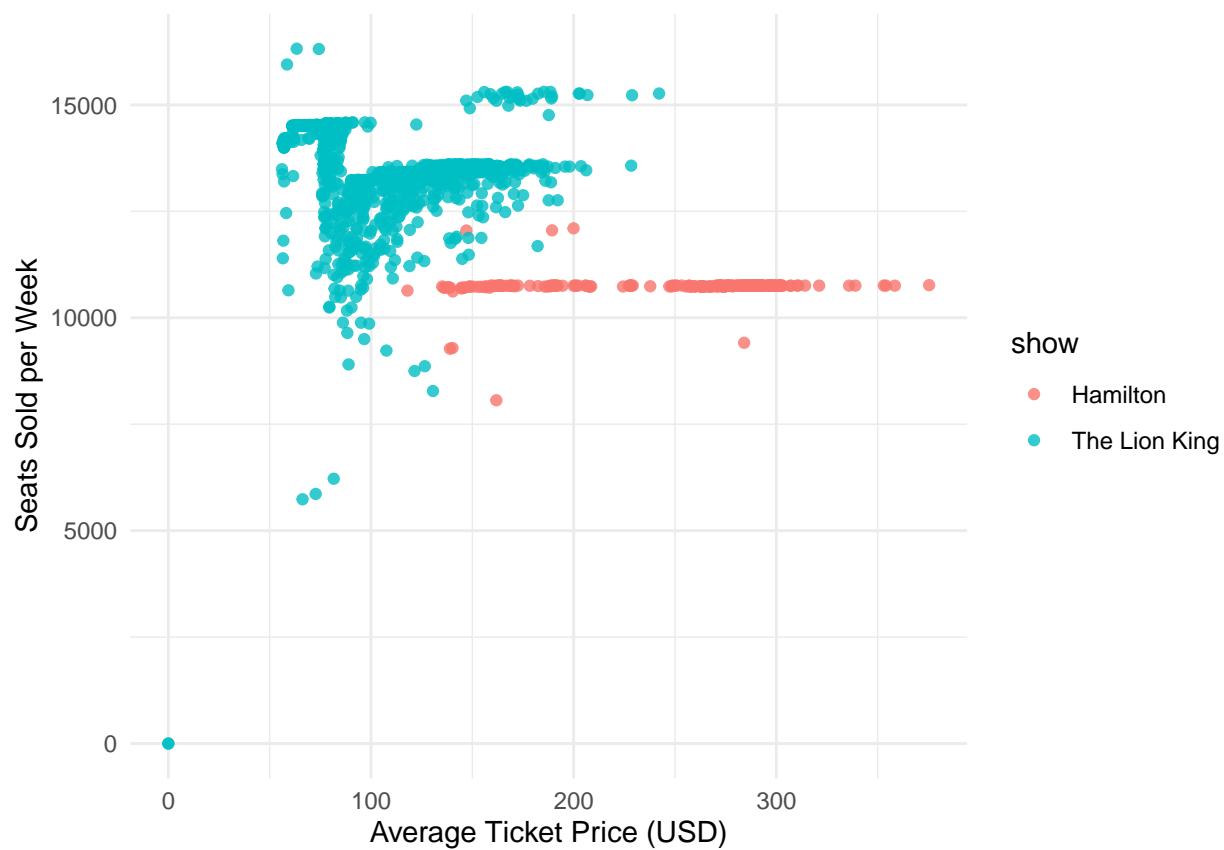


Figure 2: Seats sold per week vursus average ticket prices

So the number of seats sold for Hamilton is less than The Lion King's, but what about the theater's capacity?

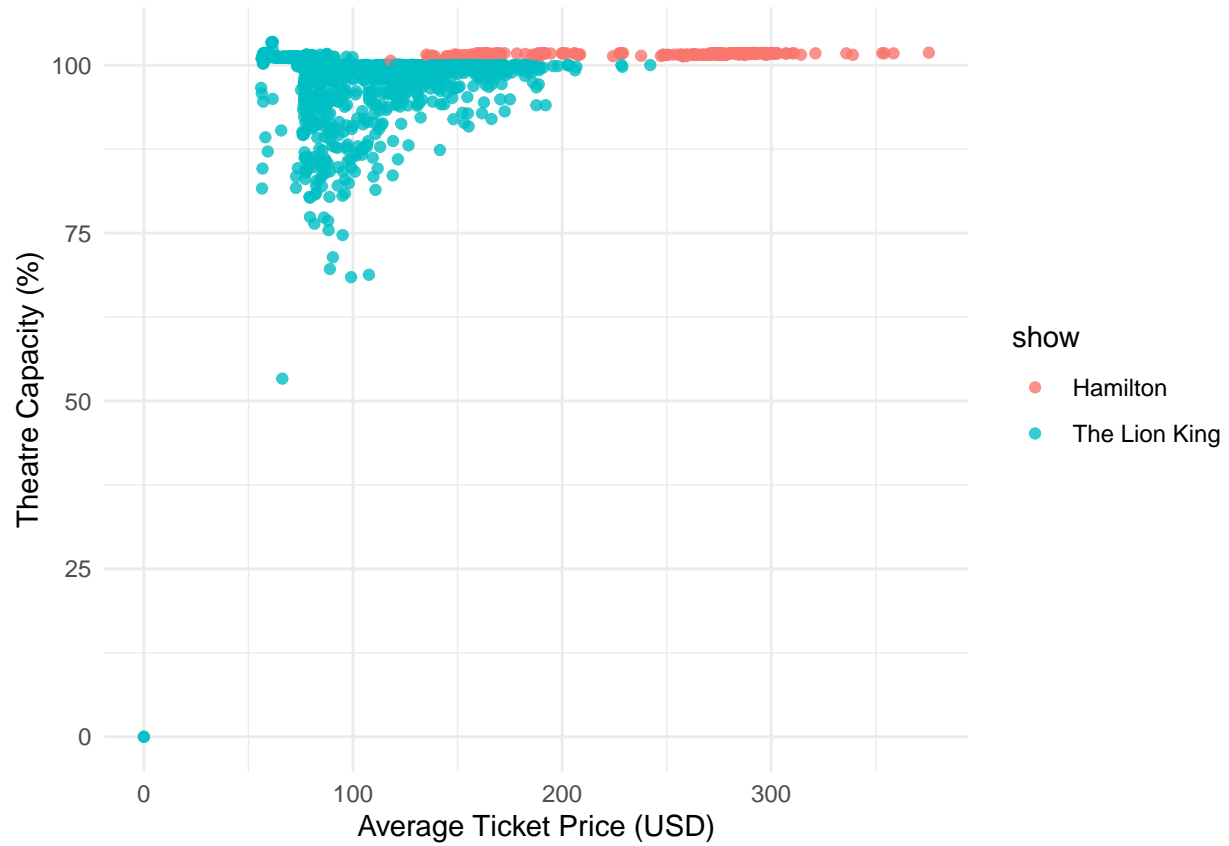


Figure 3: Percentage of theater capacity sold vursus average ticket price

Figure 3 shows that when the average ticket price for The Lion King was between about \$50 - \$200 dollars, the theater was at 100% capacity, meaning it sold out. The demand for Hamilton was stronger; average tickets sold out when they ranged between \$150 - \$350 dollars. Some shows even had over 100% capacity. These are instances of standing room tickets.

Time to explore the ticket prices.

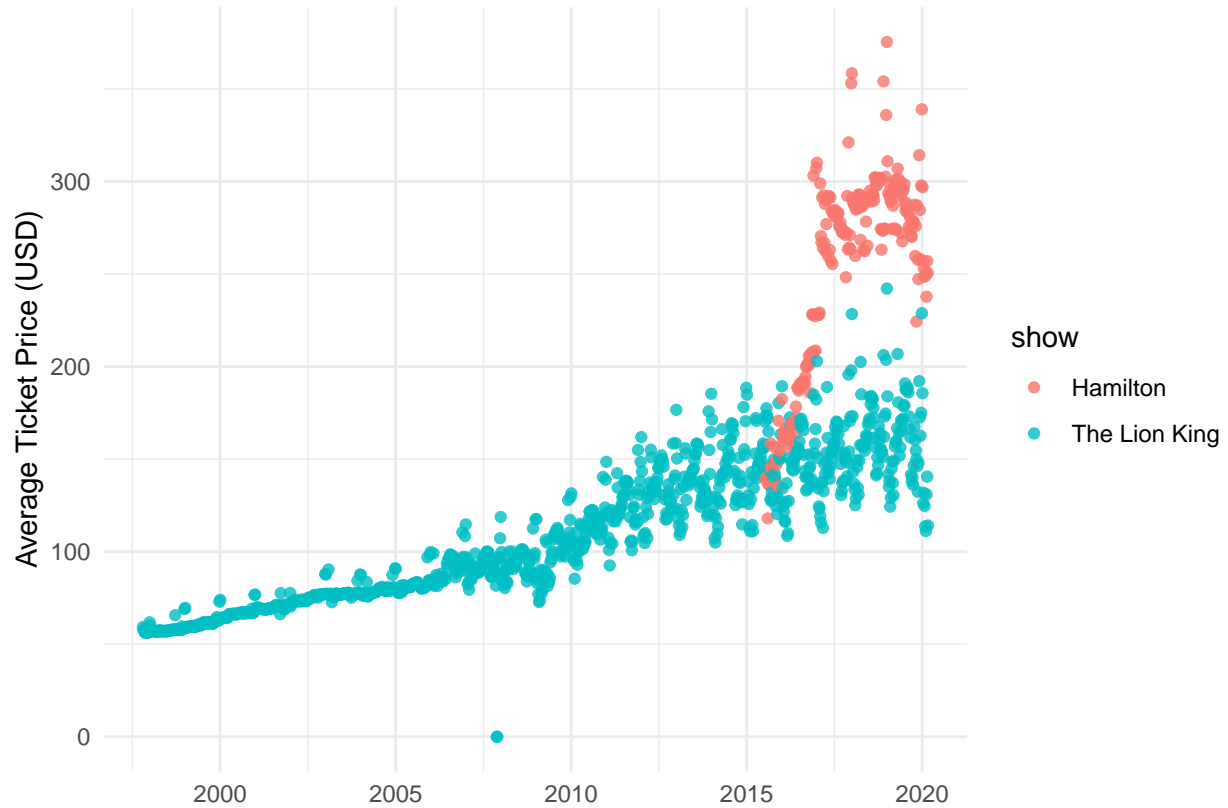


Figure 4: Average ticket prices each week

It is evident in figure 4 that Hamilton tickets are way more expensive than The Lion King's. Hamilton may have started close to the lowest ticket price of The Lion King in 2015, but the average ticket price for Hamilton doubled within the first year and exceeded The Lion King's within the second year. While the average ticket price for The Lion King could reach up to a maximum of \$200 USD, Hamilton's *started* at \$250 USD.

Similarly, the top ticket prices for Hamilton start at the maximum of The Lion King's tickets. It is almost double the price. Figure 5 shows a summary of the top ticket prices:

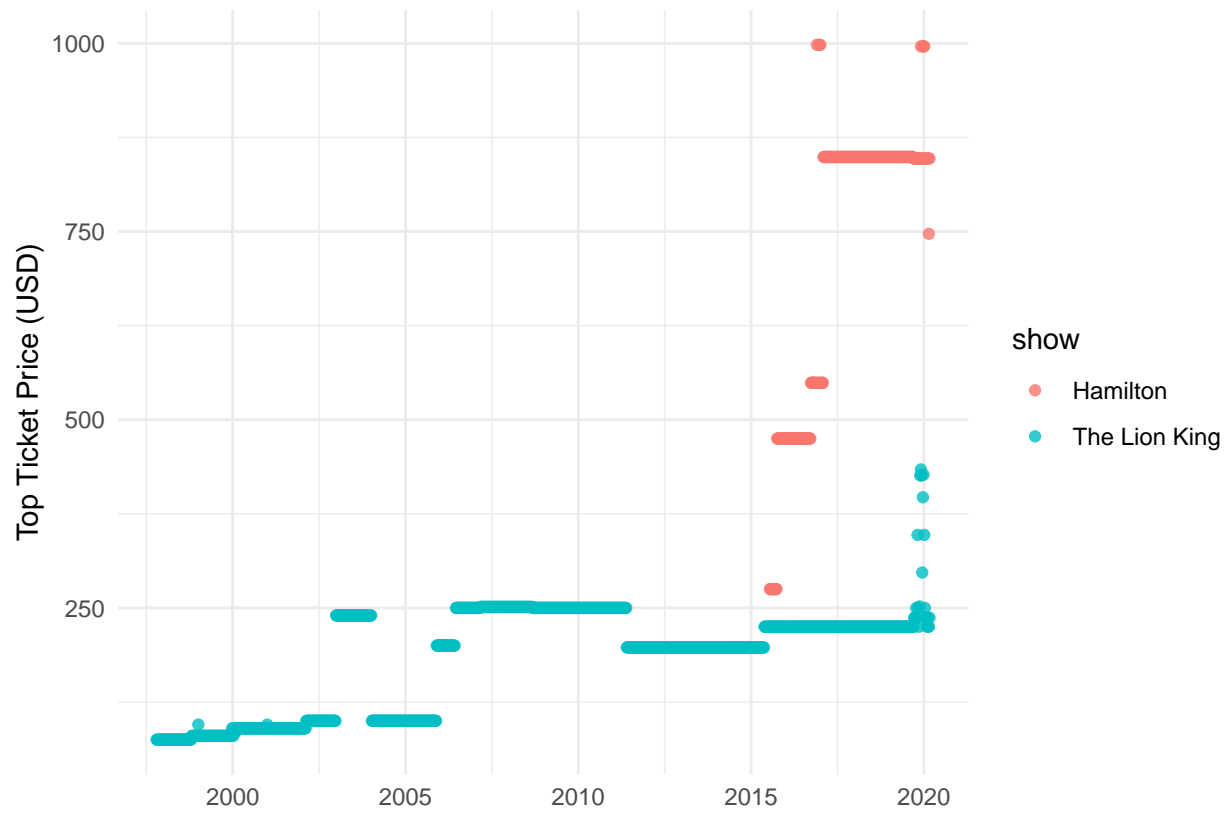


Figure 5: Top ticket prices per week

To put things into perspective, figure 6 compares top ticket price verses average ticket price for both shows over the years. The Lion King's ticket prices had incremental increases in the last 5 years. However Hamilton's ticket prices grew exponentially over the first 2 years, and gradually for the remaining years.

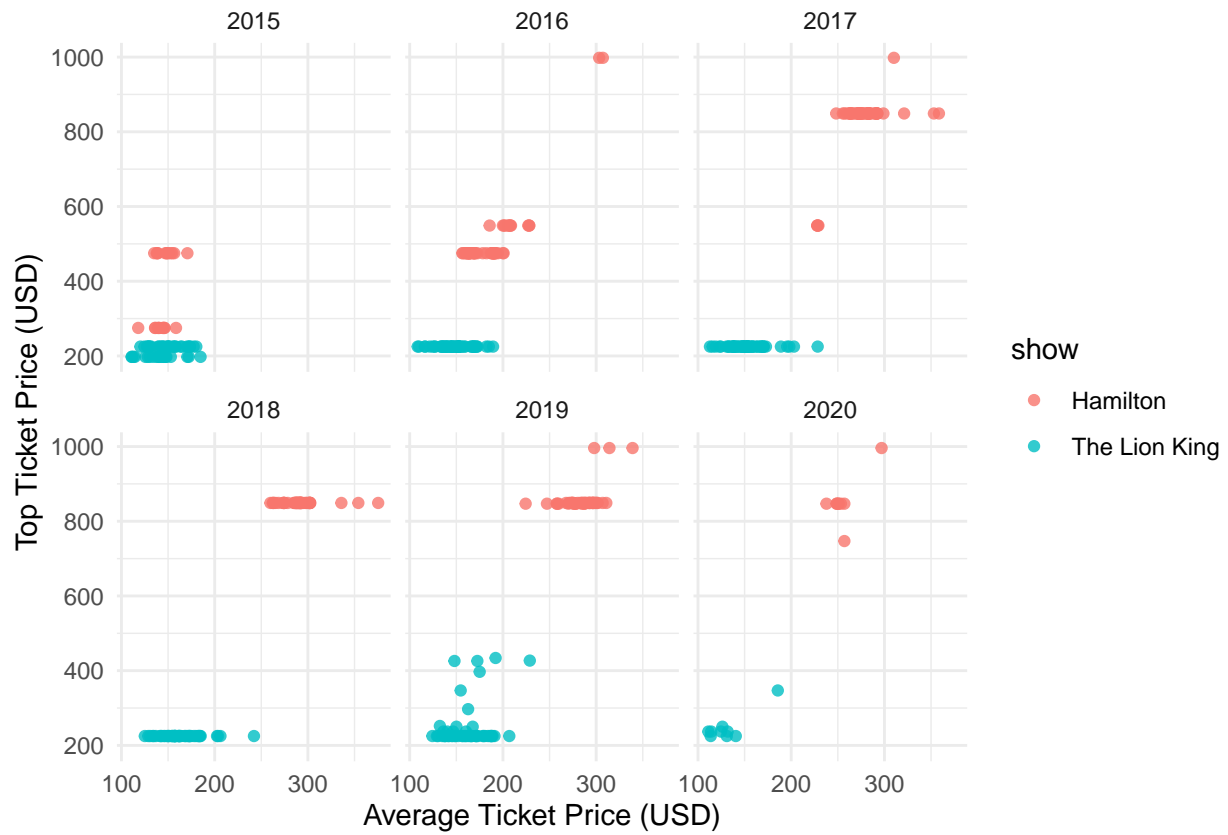


Figure 6: Top ticket prices vursus average ticket prices per year

3 Model

4 Results

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Appendix

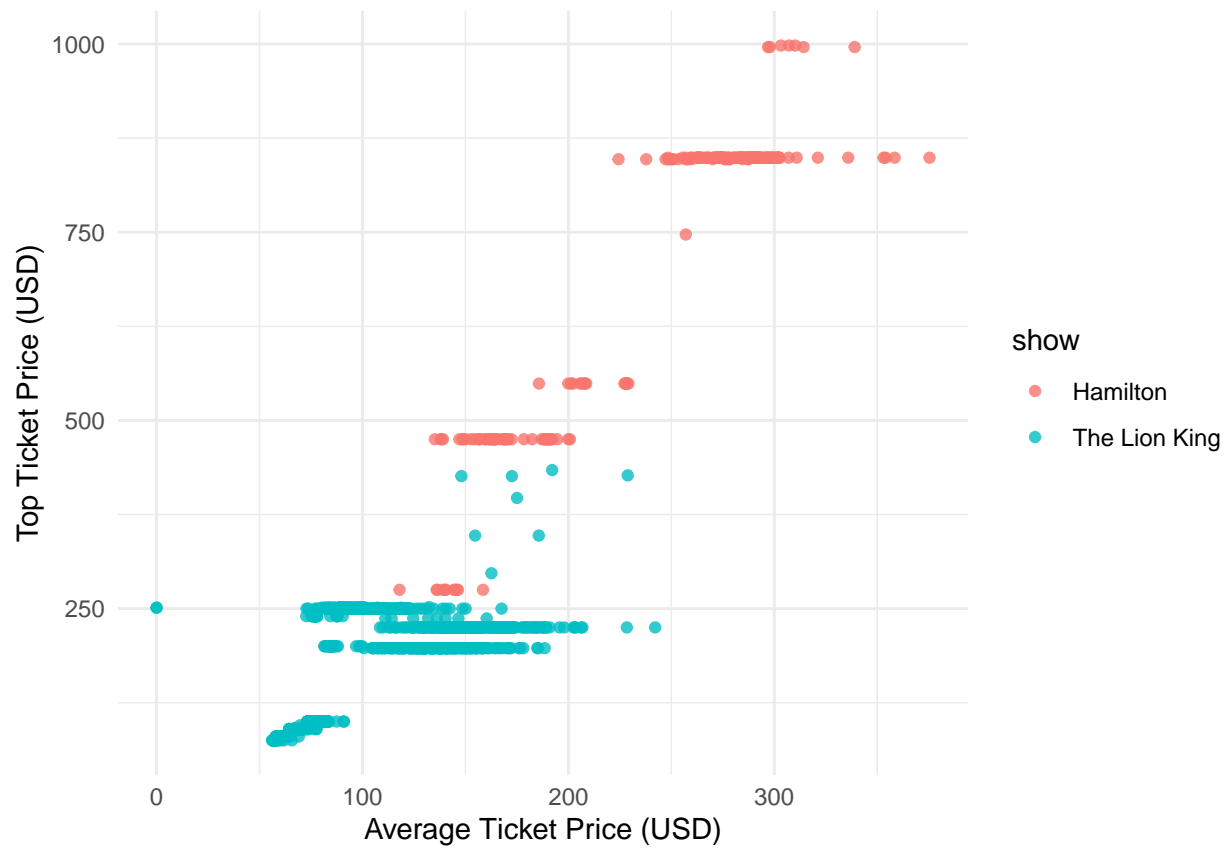


Figure 7: Top ticket prices vursus average ticket prices

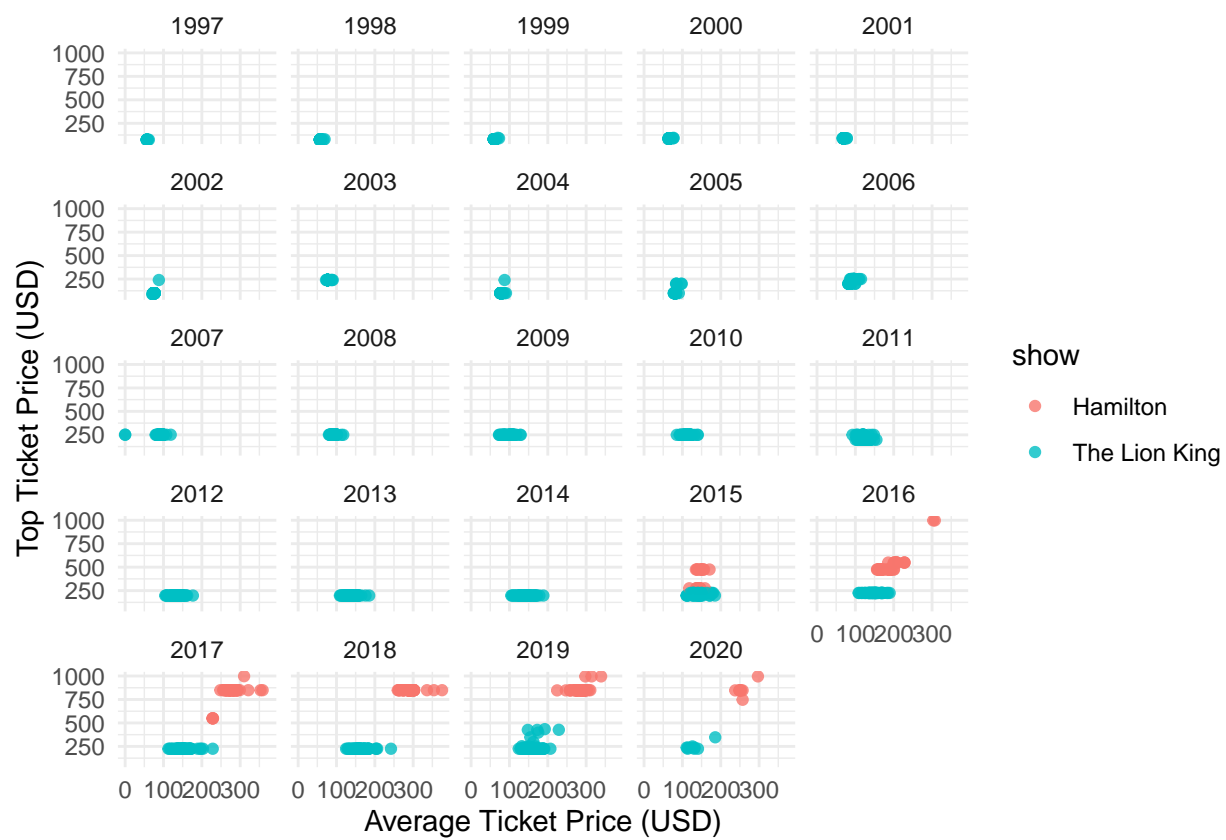


Figure 8: Top ticket prices versus average ticket prices per year

A Additional details

References

- Cookson, Alex. n.d. “What’s the Most Successful Broadway Show of All Time?” *Alex Cookson*. <https://www.alexcookson.com/post/most-successful-broadway-show-of-all-time/>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Xie, Yihui. 2021. “Knitr: A General-Purpose Package for Dynamic Report Generation in R.” <https://yihui.org/knitr/>.