

Is Jazz Dead? An analysis of restaurants, bars, and coffee shops that host live jazz music

Lucas McLaughlin

Saturday, November 21, 2015

Introduction

Jazz music is one of the few global styles of music that developed in the United States of America. As an amateur jazz musician, I have an interest in seeing jazz music continue to thrive in the U.S. Unfortunately, most statistics seem to show that consumption of jazz music is decreasing over time.

Something that particularly interests me is the businesses that continue to host live jazz music today. Where can one go to hear live jazz music? What are the unique characteristics of these businesses? How does the jazz scene vary from city to city?

To answer these questions, I use the Yelp dataset, which is available [here](#). My primary question of interest is this: are the businesses that host jazz music considered to be classy, as in the 1960s swing era, or are they considered to be “looser,” as in the 1920s prohibition era.

Methods and Data

Getting the Data

The Yelp dataset is comprised of five JSON files which can be read about in detail [here](#). They can be downloaded and unzipped using simple commands. To stream them into R I used the `stream_in` function from the `jsonlite` package. Only the checkin data, review data, and business data were used (fig 1).

The criterion I used for narrowing down which businesses hosted live jazz music was to search through reviews for instances of the word “jazz” and grab the business ID that corresponded to that review. This also grabs businesses for which the word “jazz” was used in a review in another context, but after manually reading through a subset of the reviews, I found this happened minimally.

After some initial exploration, I chose to retain the data pertaining to [r]estaurants, [b]ars, and [c]offee shops, creating the `rbc` dataset. This covered approximately 90% of the businesses (fig 2). After ordering the checking data in chronological order, it was subset by the same criteria (fig 3).

Cleaning the Data

Categorizing by jazz/not jazz

The most important variable I added was a categorical variable that identified the business as hosting live jazz music or not.

Modifying Opening and Closing Times

Next, opening and closing times were converted to POSIXct format in order to use them for time-series data. Special care was used for businesses whose closing times were after midnight, as these times actually occur on the next day. In the end, I calculated average closing times for weekdays (considered Sun-Thu in this analysis) and weekends (Fri-Sat).

Grouping Suburbs together

The Yelp data centers around 10 major cities; six in the U.S., two in Canada, one in the U.K., and one in Germany. The reality, though, is that many more than 10 cities are present in the dataset, because of suburbs that are located near the major metropolitan areas. In order to simplify the analysis, I took the most frequently non-metropolitan cities and recategorized them by their nearby metropolitan center.

```
levels(rbc$City_Region)
```

```
## [1] "Phoenix"      "Las Vegas"    "Montreal"     "Charlotte"    "Edinburgh"
## [6] "Pittsburgh"   "Madison"      "Karlsruhe"    "Champaign"    "Waterloo"
```

Results

When jazz music was still developing in the 1920s and 1930s, it had become associated with irreputable establishments. The [wikipedia article on jazz](#) states “From 1920 to 1933 Prohibition in the United States banned the sale of alcoholic drinks, resulting in illicit speakeasies which became lively venues of the ‘Jazz Age’...”. Later on in the 1960s when swing bands were popular, jazz became more associated with classy venues. So, which of these stereotypes holds true today? Let’s find out.

Hours Visited

My first theory was that people visited jazz venues later at night than venues that don’t host jazz music. This appears to be true. Within each daily cycle there are two peaks: one around lunchtime and one around dinnertime (fig 4). The peaks for jazz establishments are consistently lower around lunch and higher around dinner, with an especially noticable difference on Friday.

Business Attributes

Next, let’s look at some business attributes to determine the character of these venues. The categories I’ve chosen to investigate are whether or not the venue serves alcohol, whether it’s categorized as “good for kids”, and whether it’s described as either classy or upscale (fig 5). We see that jazz establishments on average more often alcohol, are less kid friendly, and are more upscale/classy.

Geographic Distribution

Let’s also investigate which cities in the U.S. and abroad have the highest percentage of jazz establishments (fig 6). Since jazz developed in the U.S. it’s not surprising to see lower rates abroad, with the one notable exception of Montreal.

Ratings

At the end of the day, one of the most important things this dataset can tell us is how well-rated these places are. I’ve created a linear model that describes rating by the jazz/not jazz variable (fig 7). This tells us essentially what we’d know from simply calculating averages, but also gives us a p value for significance. The average rating for jazz establishments is 3.76 while the average star rating for non-jazz establishments is 3.48. The p-value is 0

Discussion

In conclusion, it should be noted that the question of whether or not modern-day establishments that host jazz music are more like speakeasies from the '20s or classy venues from the '60s is only somewhat answerable. The reality is that there are many different kinds of restaurants, bars, and coffee shops in the U.S. and abroad, many with their own unique atmosphere.

I do believe, though, that the data allows us to sketch a picture (albiet with somewhat broad strokes) of what these places are like. We know they are described as classier, we know they more often serve alcohol, and we know they are frequented later at night. This leads me to believe jazz music is more associated with high-class living today, and that these venues are more akin to the venues from the 1960s.

Appendix

Figure 1

```
file_URL <- "https://d396qusza40orc.cloudfront.net/dsscystone/dataset/yelp_dataset_challenge_academic_
download.file(file_URL, destfile = "zipped_data.zip")
unzip("zipped_data.zip")

check <- stream_in(file("./data/yelp_academic_dataset_checkin.json"))
check <- flatten(check) # Nested data frame with same dims
bus <- stream_in(file("./data/yelp_academic_dataset_business.json"))
bus <- flatten(bus) # Nested data frame with same dims
review <- stream_in(file("./data/yelp_academic_dataset_review.json"))
```

Figure 2

```
rest.index <- sapply(bus$categories, is.element, el = "Restaurants")
bar.index <- sapply(bus$categories, is.element, el = "Bars")
coffee.index <- sapply(bus$categories, is.element, el = "Coffee & Tea")
rbc <- bus[rest.index | bar.index | coffee.index, ]
```

Figure 3

```
check.rbc <- check[check$business_id %in% rbc$business_id, ]
```

Figure 4

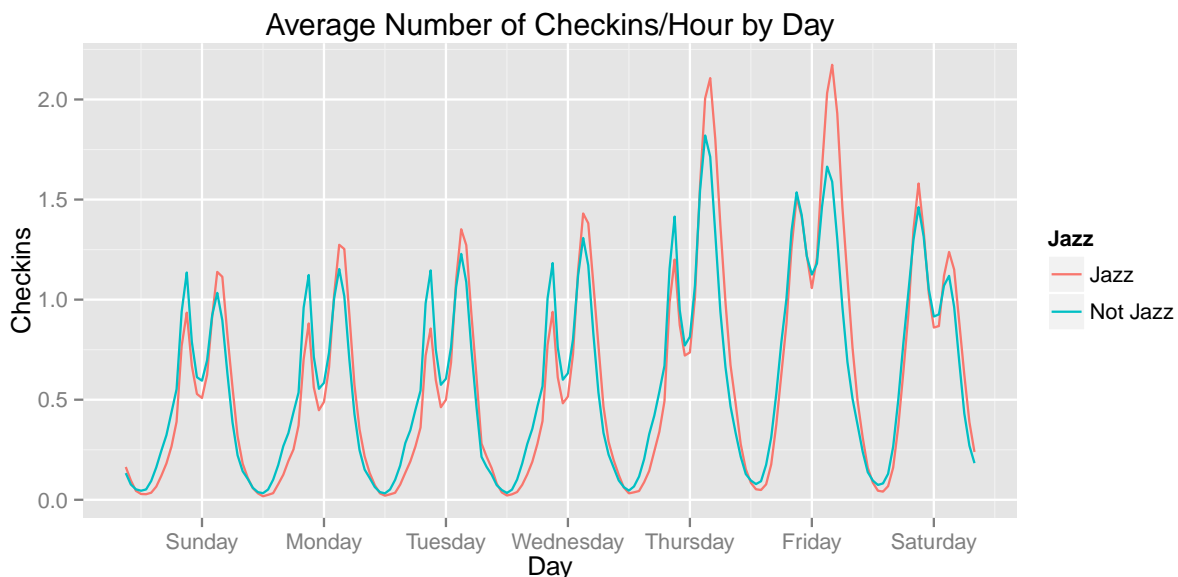


Figure 5

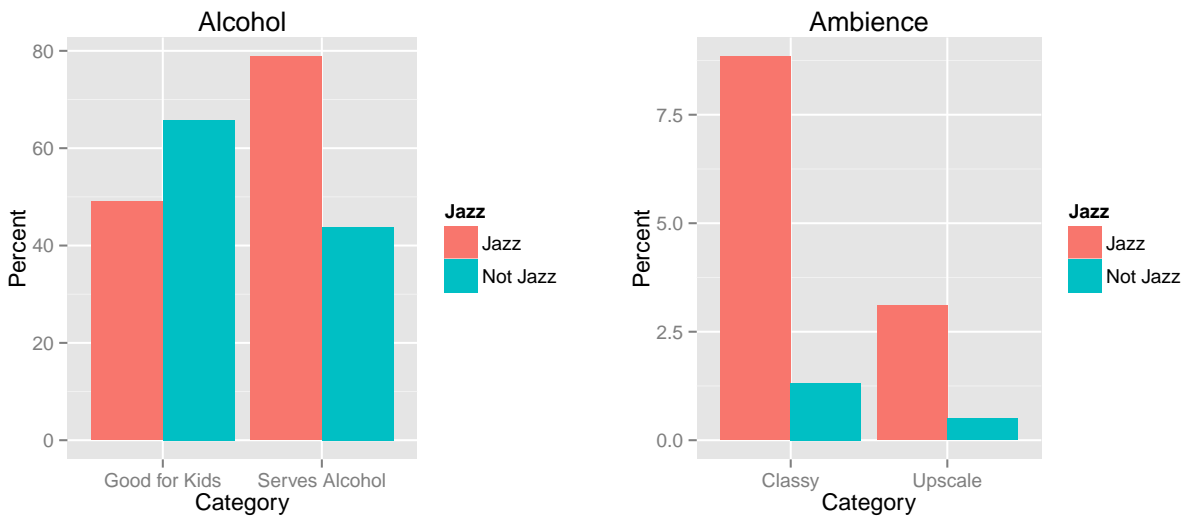


Figure 6

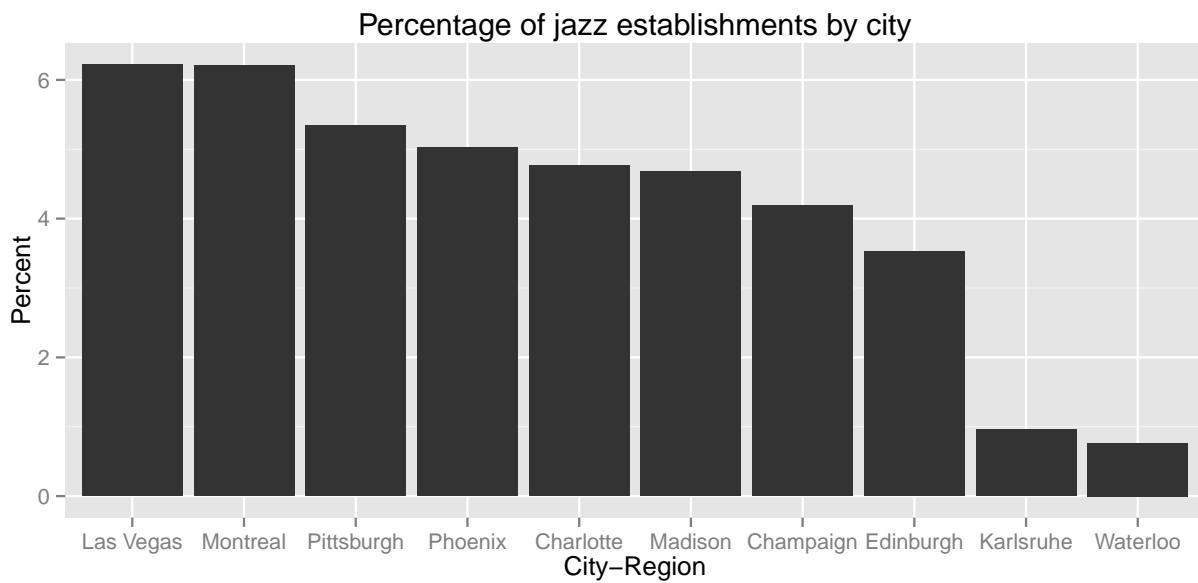


Figure 7

```
fit <- lm(stars ~ Jazz, data = rbc); coef <- summary(fit)$coefficients
```