Texas Death Row Inmate Analysis

Chris Selig

October 23, 2017

# Summary

I am curious about what goes through the mind of someone who was about to be executed for the crimes they committed. Are they sad? Angry? Full of remorse? This document attempts to solve some of my curiosity by taking a look at the statements of 544 executed people that were on death row in Texas. This analysis will have three parts: An exploratory analysis of some basic inmate data that was scraped from the [Texas Department of Criminal Justice Website](http://www.tdcj.state.tx.us/death_row/dr_executed_offenders.html), a look at the last statements themselves, and finally a topic modelling exercise.

You can find the code, data, and writeup of this analysis on my [Github page](https://github.com/chrisselig/TexasDeathRow)

### The Data

I scraped basic info from the Texas Department of Criminal Justice website, which included:

1. Execution number
2. Link for offender information(not used)
3. Link for last statement
4. Last Name
5. First Name
6. TDCJ Number
7. Age (at execution)
8. Execution date
9. Race
10. County
11. Last statement

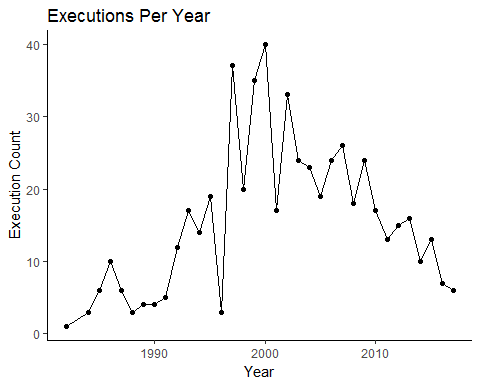
Unfortunately, data about the crimes committed, education level, date of birth, height, weight, age at time of offence, prior occupation, prior record etc., was mostly in jpeg format so I didn't make an attempt to scrape them. Some of that data was in tables, but it was a very small percentage of the total amount.

The second dataset used in this analysis is basic info on the number of executions per year. Columns in this datset include:

1. Year
2. Counts for number executions per race: white, black, hispanic and other

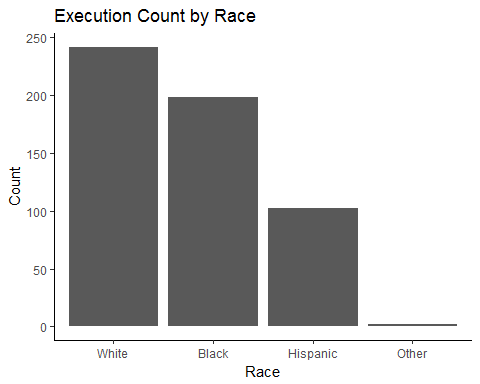
### Exploratory Analysis - General/Inmate Information

A good place to start our analysis is to take a look at the overall number of executions per year.



The late 1990s and early 2000s seemed to be the peak of the number of executions in Texas with a downward slope overall to the current year. I should note though, there are 4 more executions scheduled in 2017, so 2017 will have more executions than 2016, but still fewer than 2015.

Moving on, lets take a look at the breakdown of executions by race.



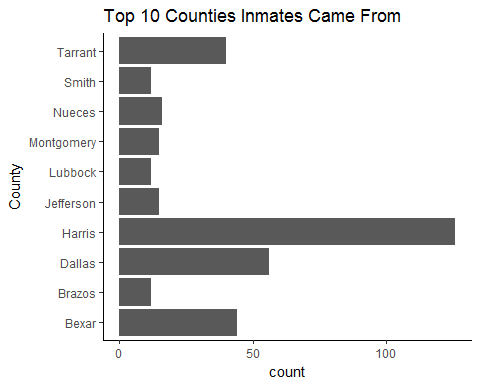
By far, more white people have been executed versus the other races. Interestingly enough, there were very few "other" races executed.

Let's take a look at the demographic datda for Texas to see if the counts make sense. According to the [2010 census](https://en.wikipedia.org/wiki/Demographics_of_Texas), 45.3% of Texans were white, 37.6% were Hispanic, 11.8% were African American. In comparision the proportion of black people executed was 36.4%. That is 3 times the proportion of black people in Texas. Hispanic inmates made up 18.8% of the executed population.

Moving on, let's take a look at the distribution of ages at the time of executions. The majority of executions are for people in the prime of their lives, between the ages of 30 and 40.



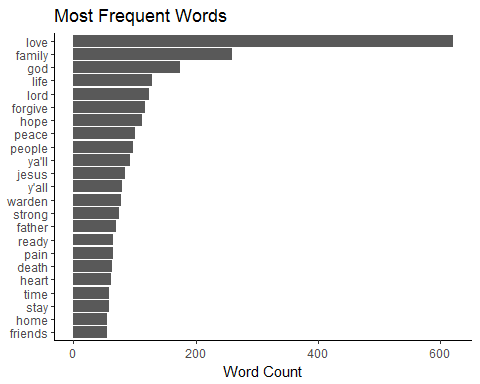
For the final section of this exploratory analysis, we look at the top 10 Counties that the executed inmates came from.



By far, most of the inmates came from Harris County. Harris County had a population of 4,092,459 as of the [2010 census](https://en.wikipedia.org/wiki/Harris_County,_Texas), and is the third-most populous county in the United States, and most populous in Texas, so it isn't that surprising that it is at the top of the list. According to [Data USA](https://datausa.io/profile/geo/harris-county-tx/), 16.7% of Harris County lives in poverty. [Children At Risk](https://en.wikipedia.org/wiki/Harris_County,_Texas) estimates that 38% of children in Harris County will not finish high school.

### Exploratory Analysis - Last Statements

Now that we have explored some basic data about the inmates, let's look at the last statements. Out of the 544 executions in this dataset, 356 left a statement. The first step in this part of the analysis was to isolate the last statements, and remove stop (common) words. After you do this, the most common word is "love", followed by "family", "god" and "life". Looking at the most common words, there does not appear to be many words you would not expect, and appears that most inmates have reached at least a form of resignation about their fate.



Next we will look at both bigrams and trigrams. Looking at the bigrams first, there are 2 religious references in the top 5.

|  |  |  |
| --- | --- | --- |
| word1 | word2 | n |
| stay | strong | 48 |
| god | bless | 30 |
| love | ya'll | 23 |
| death | row | 19 |
| lord | jesus | 19 |

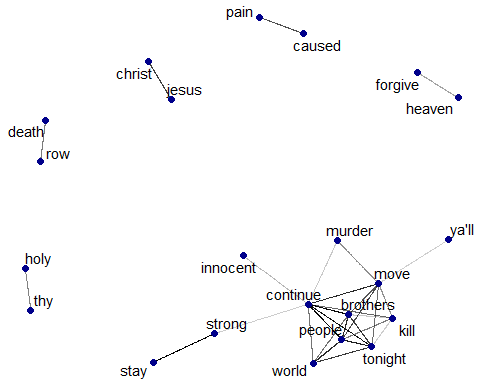
I am curious to know if the inmates were religious before they were placed into prison, but I do not have that data.

The trigrams were mostly uninteresting as they had alot of religious references. So, I decided to take a bit different approach. I was curious to know if the inmates would proclaim their innocence before their execution. So, I searched for the trigrams that included "not" and "innocent" thinking that "not guilty" or "am innocent" might appear frequently.

|  |  |  |  |
| --- | --- | --- | --- |
| word1 | word2 | word3 | n |
| not | going | to | 10 |
| not | into | temptation | 6 |
| innocent | man | i | 4 |
| not | want | he | 4 |
| not | commit | i | 3 |
| not | do | it | 3 |

There are definitely a few references to being innocent, but not as many as I would have thought. Of note, references to the Lord's Prayer is pretty common.

For the last bit of analysis, we will look at the correlations between words. The plot belows shows the correlation between various words. All correlations are greater than or equal to 0.5.



In the chart above, "Jesus" and "Christ" are highly correlated, but not surprising. I find the correlation between "caused" and "pain" interesting as it is a reference, in most cases, the inmates remorse for committing murder.

### Topic Modelling

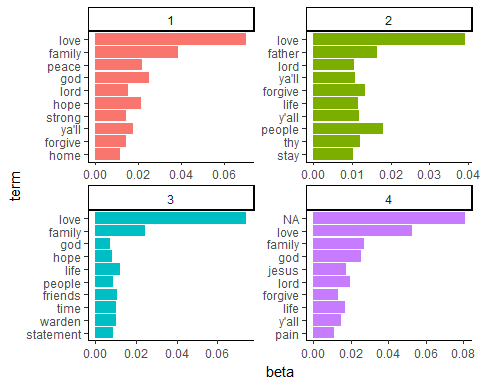
The last section of this analysis is an attempt at topic modelling to see how the inmates are feeling right before their executions. Topic modelling in this exercise will use the [Latent Dirichlet allocation (LDA)](https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation) statistical model. The LDA model treats each document (last statement in our case) as a mixture of topics and each topic as a mixture of words. Documents can overlap each other in terms of content rather than being separated into discrete groups.

The LDA model is guided by two principals.

1. Every document is a mixture of topics: This means that each document may contain words from several topics in particular proportions.
2. Every topic is a mixture of words: For example, if a topic is a financial news source, you could see the words "earnings", "stocks", or "bonds". Also important is that words can be shared across topics.

Therefore, LDA is a method of estimating both at the same time: finding the mixture of words that is associated with a topic, but also determining the mixture of topics that describes each document.

Applying the model, estimating four topics, these are the top 10 terms for each topic, and their associated betas (probabilities).



All four topics look pretty much the same, using the same words. All four mention love, family, and religion. Even after both reducing the number of topics in the model to two, and even increasing to 15, I was left with pretty much the same answer: most inmates were both remorseful, and wanted to wish their loved ones the best. A considerable amount also had religious references, and very few seemed angry.