

Exploring Texas Death Row Data

Introduction

In this assignment you are going to explore a dataset about Texas deathrow inmates, I know, kinda morbid. This assignment is meant to get you refamiliarized with the tidyverse! The dataset was taken from a website called Select Star SQL, which is a great place to learn some SQL if you are interested!

```
library("tidyverse")
```

```
-- Attaching packages ----- tidyverse 1.3.2 --
v ggplot2 3.3.6      v purrr   0.3.4
v tibble  3.1.8      v dplyr   1.0.9
v tidyr   1.2.0      v stringr 1.4.1
v readr   2.1.2      v forcats 0.5.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
```

The dataset is contained in the project. Start off by loading in the data using `read_csv`

```
TXDR <- read_csv('tx_deathrow_full.csv')
```

```
Rows: 553 Columns: 18
```

```
-- Column specification -----
Delimiter: ","
chr  (9): Last Name, First Name, Race, County, Eye Color, Height, Native Cou...
dbl  (5): Execution, Highest Education Level, TDCJ
Number, Age at Execution,...
date (4): Date of Birth, Date of Offence, Date Received, Execution Date
```

- i Use ``spec()`` to retrieve the full column specification for this data.
- i Specify the column types or set ``show_col_types = FALSE`` to quiet this message.

Now that we have the data loaded, lets answer some questions.

1. Some counties in Texas are known for executing more inmates than others. For this data, I want you to count how many executions each county has and then arrange the result so that you can see the highest ones.
2. The dataset contains the inmates last statement before they are executed. What proportion of them claim innocence? This will require some string manipulation.
3. I want to know if the executions have been constant over time or have they changed. First, count how many executions happened for each year in the data. Second, plot the number of cumulative executions over time.
4. Lastly, I want you to make up your own question about the data and answer it!

```
TXDR %>%  
  count(County) %>%  
  arrange(desc(n))
```

```
# A tibble: 92 x 2  
  County      n  
  <chr>    <int>  
1 Harris    128  
2 Dallas    58  
3 Bexar     46  
4 Tarrant   41  
5 Nueces    16  
6 Jefferson 15  
7 Montgomery 15  
8 Lubbock   13  
9 Brazos    12  
10 Smith    12  
# ... with 82 more rows
```

Q1: By far the highest county is Harris County with 128 exections, followed by Dalls County with less than half of that, at 58.

```
ClaimsInnocence <- na.omit(str_detect(TXDR$`Last Statement`, "innocent|innocence"))

sum(ClaimsInnocence == TRUE) / sum(ClaimsInnocence == FALSE)
```

[1] 0.0804878

Q2: 33 of the 410 inmates who provided statements claimed to be innocent, based on a search of the keywords “innocent” and “innocence”

```
TXDR$`Execution Date` <- as.Date(TXDR$`Execution Date`, format="%d/%m/%Y")
tab <- table(cut(TXDR$`Execution Date`, 'year'))
dframe <- data.frame(Date=format(as.Date(names(tab)), '%Y'),
                     Frequency=as.vector(tab))
```

```
YearsOfExecution <- ggplot(dframe, aes(x=Date, y=Frequency)) +
  geom_bar(stat="identity", fill="#56CBF9", color="black") + scale_x_discrete(guide = guide_
    plot.title = element_text(face="bold", size=14, hjust=.5),
    plot.subtitle = element_text(hjust=.5))
YearsOfExecution
```



Q3: We can see that the death penalty was most common from 1997 to 2002, more recently we have seen that decline, likely due to controversy surrounding the death penalty as a concept.

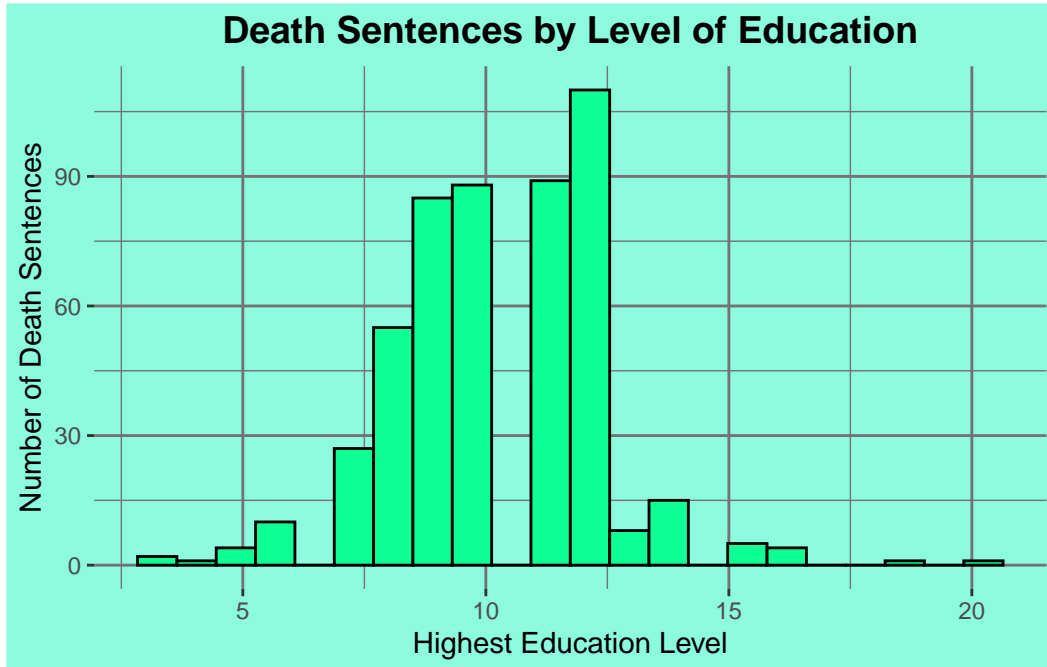
For Question 4, I will investigate how education correlates with death sentences, to see if higher levels of education are less likely among those sentenced.

TXDR

```
# A tibble: 553 x 18
  Execution `Date of Birth` Date of 0~1 Highe~2 Last ~3 First~4 TDCJ\~5 Age a~6
      <dbl> <date>          <date>          <dbl> <chr>    <chr>    <dbl> <dbl>
1      553 1983-09-24      2004-11-21          9 Young   Christ~ 999508    34
2      552 1951-08-28      1979-05-27         12 Bible   Danny ~ 999455    66
3      551 1981-04-03      2003-12-03         10 Castil~ Juan E~ 999502    37
4      550 1987-04-04      2008-04-06         11 Davila   Erick ~ 999545    31
5      549 1980-03-26      2005-09-13         12 Rodrig~ Rosendo 999534    38
6      548 1955-08-02      2001-05-02         12 Battag~ John D~ 999412    62
7      547 1953-05-21      1999-11-16         12 Rayford Willia~ 999371    64
8      546 1962-06-25      1992-04-16         12 Shore    Anthony 999488    55
9      545 1970-04-07      1997-02-22         11 Carden~ Ruben   999275    47
10     544 1979-09-18      1999-12-17          8 Pruett   Robert~ 999411    38
# ... with 543 more rows, 10 more variables: `Date Received` <date>,
#   `Execution Date` <date>, Race <chr>, County <chr>, `Eye Color` <chr>,
#   Weight <dbl>, Height <chr>, `Native County` <chr>, `Native State` <chr>,
#   `Last Statement` <chr>, and abbreviated variable names
#   1: `Date of Offence`, 2: `Highest Education Level`, 3: `Last Name`,
#   4: `First Name`, 5: `TDCJ\nNumber`, 6: `Age at Execution`
```

```
ggplot(TXDR, aes(x=`Highest Education Level`))+
  geom_histogram(color="black", fill="#0FFF95", bins= 22) + labs(y="Number of Death Senten
    plot.title = element_text(face="bold", size=14, hjust=.5),
    plot.subtitle = element_text(hjust=.5))
```

Warning: Removed 48 rows containing non-finite values (stat_bin).



While we can see that the most common level of education among those who recieved the death sentence was actually somewhere around 11-13, this makes sense because it falls around the range of graduating high school but not finishing college. Though it may be expected that more would have dropped out prior to completing high school, this is a small fraction of the US population, so for 6-10 to be so close to 11-13 shows that lower levels of income do correlate with recieving the death sentence.