

# William's Defense

*Caleb Dowdy*

*January 26, 2017*

## Data

data obtained from the Offender Based Transaction System (OBTS) was obtained for cases heard in the Florida Twelfth Circuit via a Freedom Of Information Act request.

The raw data consisted of a SQL dump file. Recovering the data followed the typical procedure for restoring from a dump.<sup>1</sup>

A separate communication provided a data dictionary (OBTS Criminal Justice data Element Dictionary, July 1997).

In order to retrieve the subset of data related to Judge Williams and Judge Haworth from the database the following SQL query was run:<sup>2</sup>

```
SELECT *
  FROM obts_data
 WHERE "SP_JudgeatSentencing" iLike 'williams%'
 OR "SP_JudgeatSentencing" iLike 'haworth%';
```

Just as a sanity check I'll include a list of what names were included in our `iLike` query:

```
SELECT "SP_JudgeatSentencing",
       count(*) as "Frequency"
  FROM obts_data
 WHERE "SP_JudgeatSentencing" iLike 'williams%'
 GROUP BY 1;
```

SP_JudgeatSentencing	Frequency
WILLIAMS, CHARLES	18674
WILLIAMS, CHARLES E	20587

```
SELECT "SP_JudgeatSentencing",
       count(*) as "Frequency"
  FROM obts_data
 WHERE "SP_JudgeatSentencing" iLike 'haworth%'
 GROUP BY 1;
```

SP_JudgeatSentencing	Frequency
HAWORTH	2
HAWORTH, LEE	18243
HAWORTH, LEE E	10780

As you can see there are a number of variants present in our dataset.

Thus, the raw data consists of 68286 observations each exhibiting 105 features.

<sup>1</sup>For more info on restoring SQL dump please refer to the following docs <https://www.postgresql.org/docs/9.1/static/backup-dump.html>

<sup>2</sup>More on exporting query results can be found in the postgres docs here: <https://www.postgresql.org/docs/9.2/static/sql-copy.html>

The quality of the data leaves much to be desired. Our first course of action is to investigate to what extent each row gives a unique case.

According to the data dictionary, there are fields recording OBTS and case number.

## Data Quality

Below is a table of the features that contain NULL entries along with the total number of complete cases.

BP_BirthDate	BP_USCitizen	BP_FBINumber	IP_ChargeQualifierUnit
68116	673	0	0

PP_ChargeQualifierUnit	CP_ChargeQualifierUnit	TRIAL	PLEA
0	0	61892	60036

sp_days_suspended
0

## Data Subsetting

We subset the data to only allow cases where the prosecutor pursues charges:<sup>3</sup>

```
data_reduced <- data %>% filter((PROSECUTOR_ACTION == "N") | (PROSECUTOR_ACTION == "C"))
```

- The code C implies that the charge is being prosecuted under another court docket number, and the existing court docket number will be eliminated. This will not include situations where cases are remaining separate but being heard simultaneously.
- The code N implies that the charge not disposed of at the prosecutor's level (filed for action by the court). The disposal for the charge would be reflected in the court phase.

Restricting on the cases above reduces the total number of observations by 10199.

Next, we subset the data to remove any juvenile cases:<sup>4</sup>

```
data2 <- data_reduced %>% filter((COURT_DESIGNATOR != "J"))
```

- J corresponds to Juvenile Court

Further restricting our previous subset on the cases above reduces the total number of observations by another 8217. In total we have lost 18416 observations from our raw dataset.

We subset once again to only include cases who are tried as felons:<sup>5</sup>

```
data3 <- data2 %>% filter((COURT_CHARGE_LEVEL == "F"))
```

- F corresponds to any offense punishable by death or imprisonment for a term of one year or more. Further restricting our previous subset on the cases above reduces the total number of observations by another 9808. In total we have lost 28224 observations from our raw dataset.

<sup>3</sup>This field will represents the final action taken by the prosecutor or grand jury

<sup>4</sup>This field will identify the court responsible for the charge.

<sup>5</sup>Used to identify the seriousness of the offense as identified by the arresting/initiating agency.

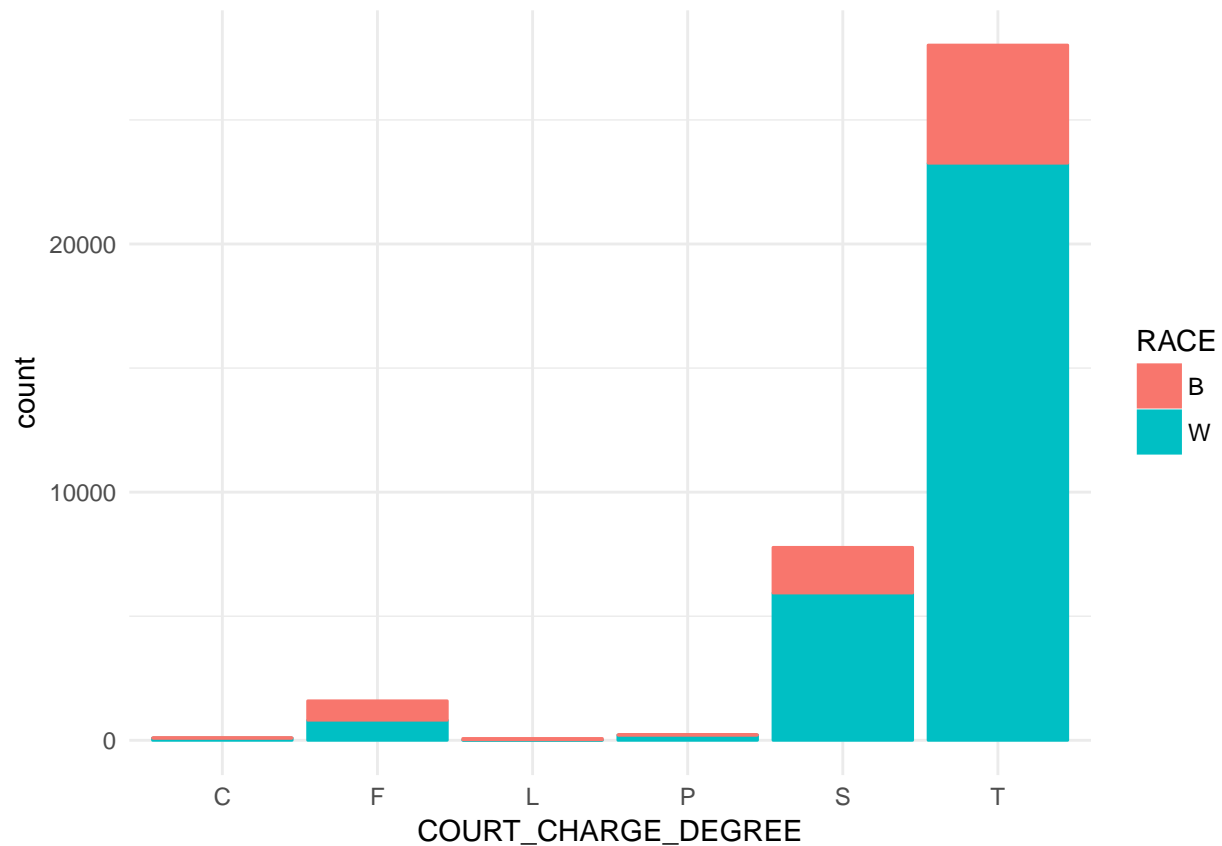
Finally, we subset what is remaining on the subjects final plea:

```
data4 <- data3 %>% filter((TRIAL == "1"))
```

- 1: Not Guilty/Deny
- 2: Guilty (includes Alford Plea)/Admit
- 3: Nolo-Contendere
- 4: Not Guilty by reason of Insanity
- 5: No plea entered (primarily used in conjunction with bond estreatures and dismissals)

```
data4 <- data4 %>% filter((RACE == "B" | RACE == "W"))
```

## Results



- C: Capital
- L: Life
- P: First Degree, punishable by life\*
- F: First Degree
- S: Second Degree
- T: Third Degree
- N: N/A - When the level of the charge is reflected as an infraction, county or municipal.