Sara Alotaibi
Daniel Song

Milestone #4

Description

The main goal of Code For PDX's Expungement Project is to build a web service that can automate the expungement analysis for all Oregonians. The project addresses the long and complex process to determine which records are eligible, and there can be a lot of errors if done by hand.

The part of the system we worked on is taking a collection of criminal records data in JSON pulled from the public Oregon Database, and creating a relational database that will easily enable sharing statistical crime data to the public.

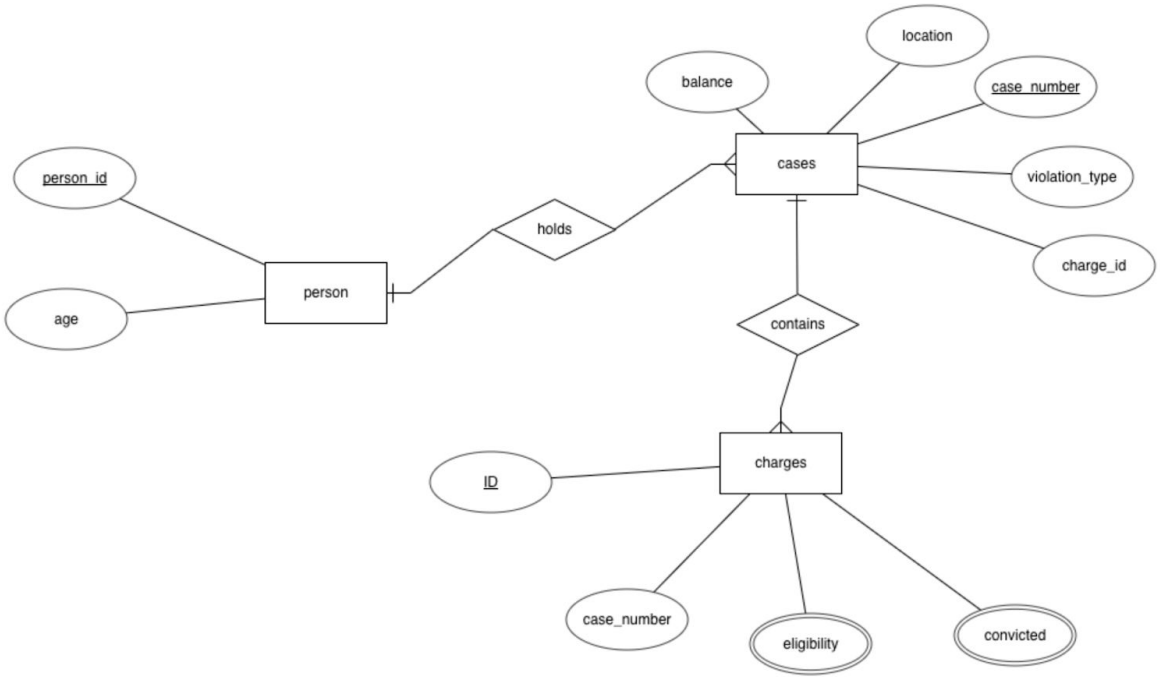
For our database design, we tried to emphasize data privacy by excluding names such as birthdate, and citation number, all while including whatever we thought would be useful to answer our questions. For our questions we tried to focus on understanding the dataset, and also the demographics of the people. For understanding our dataset, we asked questions like - "How many felonies are there? How many misdemeanors are there? How many are eligible for expungement?" And to understand the demographics of the people, we asked questions like - "What is the age of the youngest person with most criminal records?"

Due to our misunderstanding of the data from the beginning of the term, like not realizing that we can't tell if there is a record that has been expunged, we ended up using maybe half of what we had in Milestone #2. We also had to leave out some of the attributes from Milestone #2, name of violation (DUII, Trespassing, Assault, etc) because our database was getting too complex. Since it was our first time experimenting, we wanted to keep it simple. Overall, we thought some questions sounded redundant, but ultimately, we asked questions that answered what we really wanted to know about the dataset.

We also realized that our design was not perfect. For example, person and cases did NOT need to be a many-to-many relationship. But by the time we realized, it was too late. That being said, we think it was an important lesson, as we learned how crucial the design part is.

We used Python and Psycopg2 to create and populate our tables (Which was more painful than we expected). All of our questions were answered using SQL queries. Charts were created using Python and Plotly

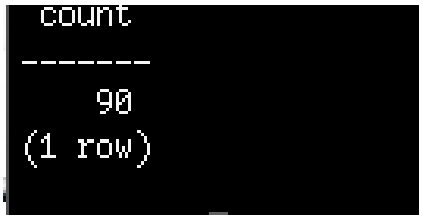
Our ERD


location
balance
case_number
cases
violation_type
person_id
holds
charge_id
person
age
contains
charges
ID
case_number
eligibility
convicted

Our 20 questions answered:

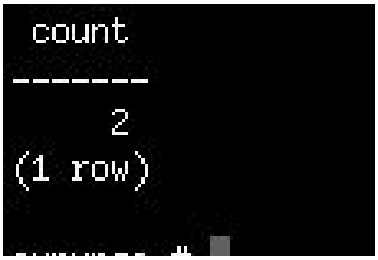
1. How many people who have at least 1 charge eligible for expungement now?

   ```
   SELECT COUNT(DISTINCT(p.person_id))
   FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
           h.case_number = c.case_number JOIN charges ch ON c.case_number =
           ch.case_number
   WHERE ch.eligibility = 'Eligible';
   ```

2. What are the names of clients who have at least 1 misdemeanor eligible for expungement and have a balance over $1000?

   ```
   SELECT COUNT(DISTINCT(p.person_id))
   FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
           h.case_number = c.case_number JOIN charges ch ON c.case_number =
           Ch.case_number
   WHERE c.violation_type = 'Offense Misdemeanor' AND ch.eligibility = 'Eligible' AND
   c.balance > 1000;
   ```

3. How many people have at least 1 felony and a total balance of over $1000?

```
SELECT COUNT(DISTINCT(p.person_id))
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
        h.case_number = c.case_number JOIN charges ch ON c.case_number =
        ch.case_number
WHERE c.violation_type = 'Offense Felony' AND c.balance > 1000;
```

```
 count
-------
    22
(1 row)
```

4. What are the distinct counties clients got their violation in?

```
SELECT DISTINCT(c.location)
FROM cases c;
```

```
  location
------------
 Washington
 Clatsop
 Yamhill
 Jackson
 Douglas
 Linn
 Curry
 Marion
 Deschutes
 Klamath
 Benton
 Clackamas
 Lane
 Josephine
 Lake
 Multnomah
 Coos
 Jefferson
(18 rows)
```

5. How many cases took place for each county in descending order?

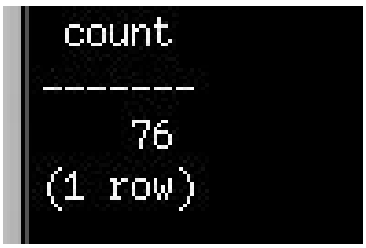
```
SELECT location, COUNT(case_number)
FROM cases
GROUP BY location;
ORDER BY COUNT(case_number) DESC;
```

location	count
Jackson	343
Josephine	81
Deschutes	29
Multnomah	24
Douglas	23
Coos	17
Lane	14
Linn	13
Marion	8
Clackamas	8
Klamath	7
Benton	6
Washington	3
Yamhill	3
Clatsop	2
Jefferson	2
Lake	1
Curry	1

(18 rows)

6. How many clients have at least 1 felony that is eligible for expungement?

```
SELECT COUNT(DISTINCT(p.person_id))
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
        h.case_number = c.case_number JOIN charges ch ON c.case_number =
        ch.case_number
WHERE c.violation_type = 'Offense Felony' AND ch.eligibility = 'Eligible'
```



count
76

(1 row)

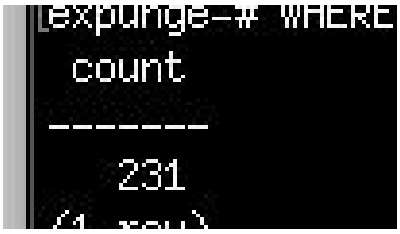
7. How many felony cases are there?

```
SELECT COUNT(violation_type)
FROM cases
WHERE violation_type = 'Offense Felony'
GROUP BY violation_type;
```

```
 violation_type | count
----------------+-------
 Offense Felony |   250
```

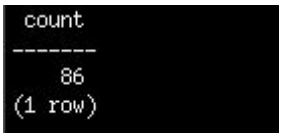
8. How many felonies are eligible for expungement?

```
SELECT COUNT(DISTINCT(p.person_id))
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
        h.case_number = c.case_number JOIN charges ch ON c.case_number =
        ch.case_number
WHERE c.violation_type = 'Offense Felony' AND ch.eligibility = 'Eligible'
```

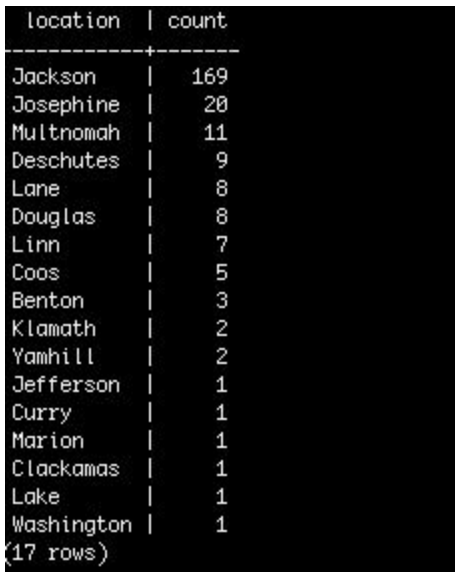


9. How many clients have no cases that are eligible for expungement?

```
SELECT COUNT(DISTINCT(p.person_id))
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
        h.case_number = c.case_number JOIN charges ch ON c.case_number =
        ch.case_number
WHERE ch.eligibility ='Ineligible';
```



10. List all counties with corresponding number of felonies in descending order

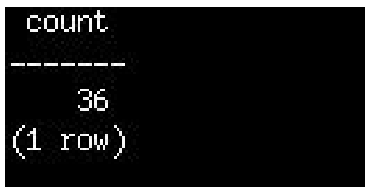


location	count
Jackson	169
Josephine	20
Multnomah	11
Deschutes	9
Lane	8
Douglas	8
Linn	7
Coos	5
Benton	3
Klamath	2
Yamhill	2
Jefferson	1
Curry	1
Marion	1
Clackamas	1
Lake	1
Washington	1

(17 rows)

11. How many clients at least 30 years old, and have misdemeanors that are eligible for expungement?

```
SELECT COUNT(p.person_id)
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
        h.case_number = c.case_number JOIN charges ch ON c.case_number =
        ch.case_number
WHERE p.age > 30 AND
        c.violation_type = 'Offense Misdemeanor' AND
        ch.eligibility = "Eligible";
```

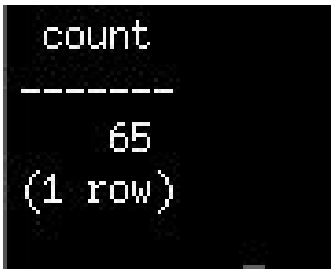


count
36

(1 row)

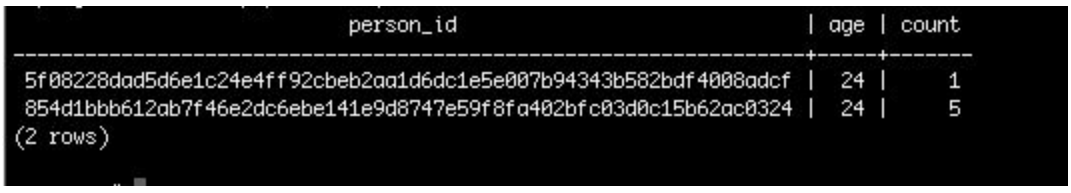
12. How many clients older than 30 years old, have at least 1 felony?

```
SELECT COUNT(DISTINCT(p.person_id))
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
        h.case_number = c.case_number
WHERE p.age > 30 AND
         c.violation_type = 'Offense Felony'
```



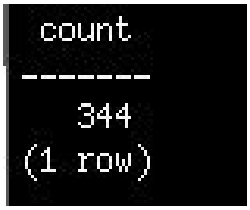
13. How many felonies and misdemeanors does the youngest person (or people) have?

```
SELECT p.person_id, p.age, COUNT(*)
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
                h.case_number = c.case_number
WHERE p.age = (SELECT MIN(age)
                    FROM person) AND
                (c.violation_type = 'Offense Felony' or
                c.violation_type = 'Offense Misdemeanor')
GROUP BY p.person_id;
```



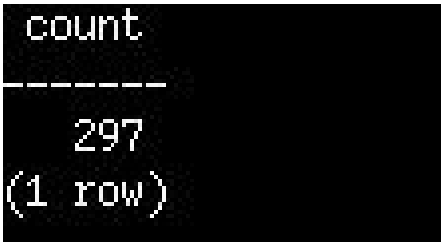
14. How many expungeable misdemeanors and felonies are there in the dataset?

```
SELECT COUNT(DISTINCT(c.case_number))
FROM cases c JOIN charges ch ON c.case_number = ch.case_number
WHERE c.violation_type = 'Offense Felony' OR
        c.violation_type = 'Offense Misdemeanor' AND
         ch.eligibility = 'Eligible';
```



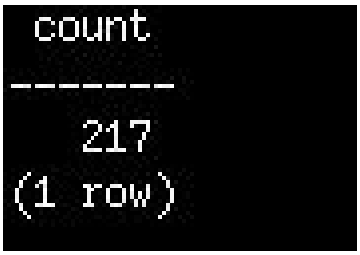
15. How many non expungeable misdemeanors and felonies are there?

```
SELECT COUNT(DISTINCT(c.case_number))
FROM cases c JOIN charges ch ON c.case_number = ch.case_number
WHERE c.violation_type = 'Offense Felony' or c.violation_type = 'Offense Misdemeanor'
AND ch.eligibility = ‘Ineligible’;
```



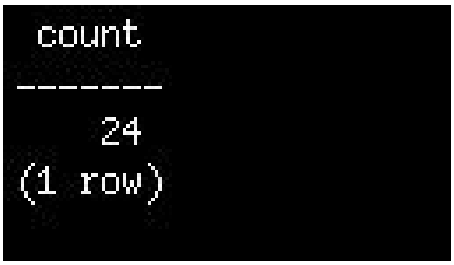
16. How many cases were convicted but still are expungeable?

```
SELECT COUNT(DISTINCT(c.case_number))
FROM cases c JOIN charges ch ON c.case_number = ch.case_number
WHERE ch.convicted = 'Convicted' AND ch.eligibility = 'Eligible';
```



17. How many people owe more than $1000?

```
SELECT COUNT(DISTINCT(p.person_id))
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
        h.case_number = c.case_number
WHERE c.balance > 1000;
```



18. How many felonies and misdemeanors do the oldest people have and owe less than $1000?

```
SELECT p.person_id, COUNT(*)
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
h.case_number = c.case_number
WHERE p.age = (SELECT MAX(age)
                    FROM person) AND
                c.balance < 1000 AND
               (c.violation_type = 'Offense Felony' or c.violation_type = 'Offense
               Misdemeanor')
GROUP BY p.person_id;
```

```
                          person_id                          | count
-------------------------------------------------------------+-------
 342ceba5a33b03f7109f4b42c05d61647db22344261016cbdb864e073d18d4f2 |     1
 8f5a64676ba1ecd9a077be47ed12011cb34da7ac033a6e4c15984a381063fd21 |     2
(2 rows)
```

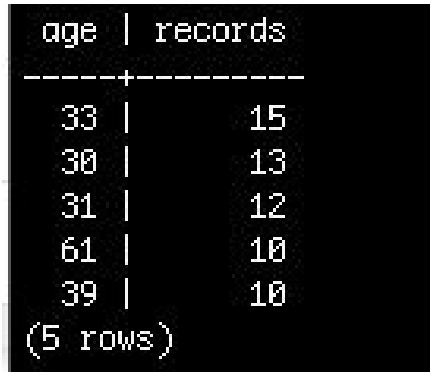
19. Show a list of all clients ordered by their number of felonies and misdemeanors

```
SELECT p.person_id, COUNT(*)
FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
        h.case_number = c.case_number
WHERE c.violation_type = 'Offense Felony'
GROUP BY p.person_id
ORDER BY COUNT(*) DESC
```

```
                          person_id                          | count
-------------------------------------------------------------+-------
 3243f73453594c5e396c7c8d1ac3c87e22ec0297cd546450d3650aa00d50ece7 |    15
 78d62e38d2e4756f58d3957b158481043aea523068184c2d8742b0d63d69ee2b |    13
 35eac9d5e44d02482fccb1d2b2727dd6243021924d0692ead220314e251d610a |    12
 e765b9fcfe40b902cc173690a4a194c202adc881acb700be91abaaf4c7e5129a |    12
 09e6a72f667a71bc1e0d9aef7c20bc75390810813f790cb27cbfe75cd52baf22 |    10
 deb532d89adb72c7b71f9632de249eee7d9c09bb4c8072cea58c105f638cc553 |    10
 4f89bd355600273c0dab1166089a54a97b4c3edef4aef54770904293d0ebd230 |    10
 7a0e60eb2a85d9ae07afcee2f011600a6b40a452959692a988f6bd237443fab2 |     8
 4e7e2e505ccdd1863b3a09be48c330264bdfc9f74d5a996360d5f14eb6371c6f |     8
 27099839d278165540861ace4d56d013fa211769e8bcb2d3ce9ae321b3c67e61 |     7
 c2c4ad978313724bf5bc578e90d87252d7dc4791802dab46ad8b56ad2c970402 |     7
 a35e9055310831ac2bda446da3ccd9c56cbe8935e597bbe32525a2c7207df9e5 |     6
 b0a64ce8f05db99c74c58534f8c983d41912ca003db374ab823b285af7da65b6 |     5
 de384e706ba3efbaeb6c01fc887dae0fce15327c47450b32f7dafa7361f3e67c |     5
 4a5984f348bc4c7230ca1ec26f955d4090cc022f0b51a6613c2780ba69d5c117 |     5
```

20. List age of first 5 people with the highest number of felonies and misdemeanors

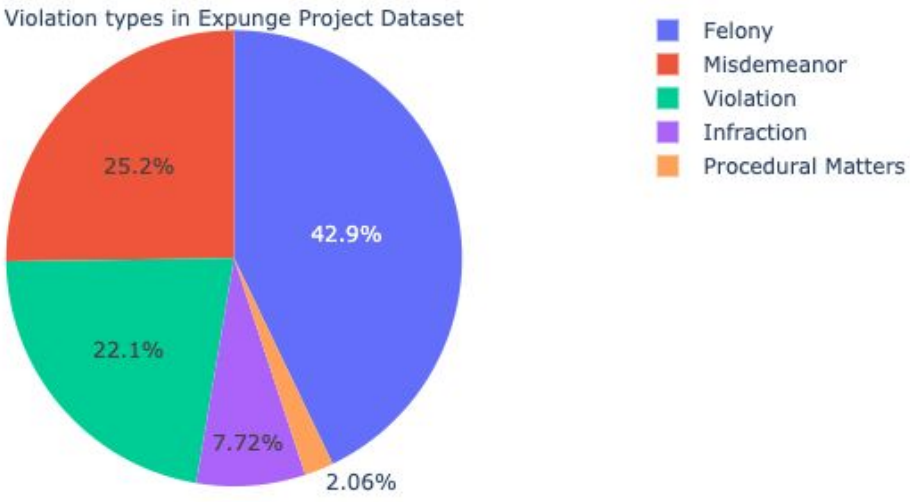
```
SELECT age, MAX(cnt) as records
FROM(SELECT p.person_id, p.age, COUNT(*) AS cnt
        FROM person p JOIN holds h ON p.person_id = h.person_id JOIN cases c ON
                h.case_number = c.case_number
                WHERE c.violation_type = 'Offense Felony'
                GROUP BY p.person_id) AS maxcount
GROUP BY age
ORDER BY records DESC
LIMIT 5;
```



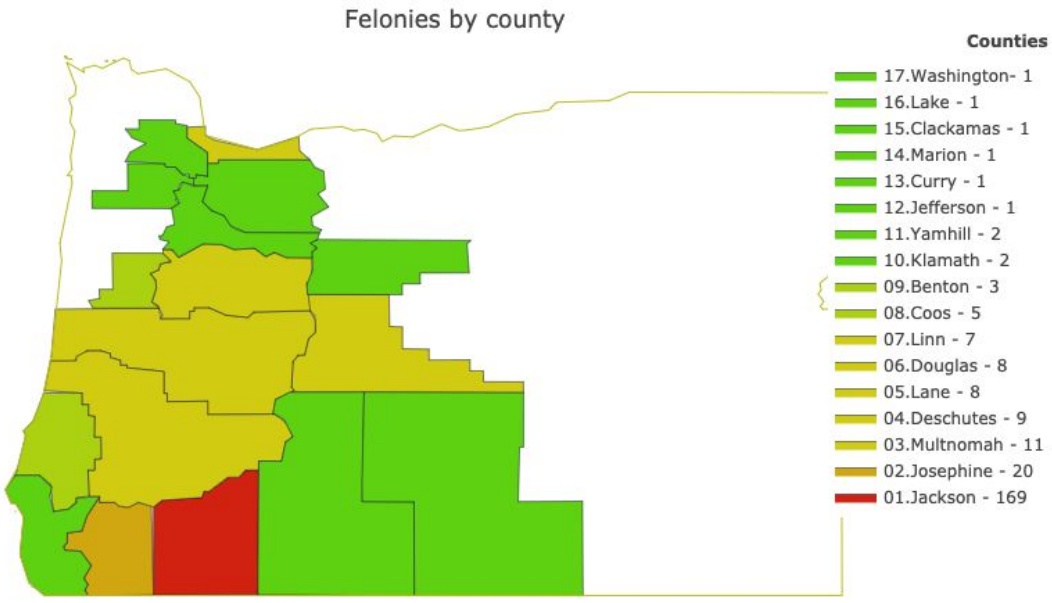
```
 age | records
-----+---------
  33 |      15
  30 |      13
  31 |      12
  61 |      10
  39 |      10
(5 rows)
```

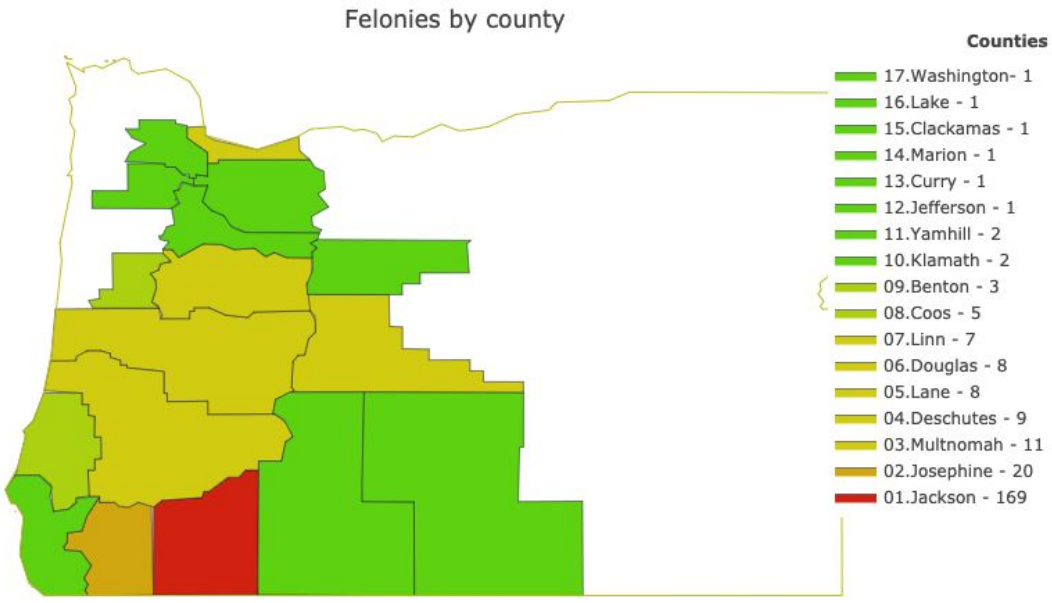
And our attempt at data visualization for extra credit..

*Disclaimer: These were created with very limited dataset, therefore may not be very accurate. Enjoy at your own risk



*Note: Please ignore the numbers in front of the county names. I put them there to keep that order.




Felonies by county
Counties
17.Washington- 1
16.Lake - 1
15.Clackamas - 1
14.Marion - 1
13.Curry - 1
12.Jefferson - 1
11.Yamhill - 2
10.Klamath - 2
09.Benton - 3
08.Coos - 5
07.Linn - 7
06.Douglas - 8
05.Lane - 8
04.Deschutes - 9
03.Multnomah - 11
02.Josephine - 20
01.Jackson - 169