Final Project:

American Crime and Incarceration

PS I Love Yue: Oliver, Karam, Steven, Emely

```
library(tidyverse)
crime <- read_csv("data/crime_and_incarceration_by_state.csv")
```

Beginning of Project

```
#Seperate data between Federal prisoners and State prisoners
#Federal observations missing a lot of data so maybe best to work with state

federal_crime <- crime %>%
  filter(jurisdiction == "FEDERAL")

state_crime <- crime %>%
  filter(jurisdiction != "FEDERAL")
```

```
#population and prisoner proportions

state_crime <- state_crime %>%
  mutate(prisoner_per_pop = prisoner_count/state_population)

state_crime %>%
  arrange(desc(prisoner_per_pop)) %>%
  select(jurisdiction, year, prisoner_count, state_population, prisoner_per_pop)
```

```
## # A tibble: 800 x 5
##    jurisdiction  year prisoner_count state_population prisoner_per_pop
##    <chr>        <dbl>          <dbl>            <dbl>            <dbl>
##  1 ALASKA        2012           6308           730307          0.00864
##  2 DELAWARE      2001           6841           796599          0.00859
##  3 ALASKA        2011           6216           723860          0.00859
##  4 ALASKA        2014           6323           736732          0.00858
##  5 DELAWARE      2002           6637           805945          0.00824
##  6 DELAWARE      2006           7021           853476          0.00823
##  7 DELAWARE      2007           7110           864764          0.00822
##  8 DELAWARE      2004           6753           830069          0.00814
##  9 DELAWARE      2003           6630           818166          0.00810
## 10 DELAWARE      2005           6788           841741          0.00806
## # ... with 790 more rows
```

```
state_crime %>%
  arrange(state_population) %>%
  select(jurisdiction, year, state_population)
```

```
## # A tibble: 800 x 3
```

```
##    jurisdiction  year state_population
##    <chr>        <dbl>            <dbl>
##  1 WYOMING       2001           493754
##  2 WYOMING       2002           498830
##  3 WYOMING       2003           502111
##  4 WYOMING       2004           505887
##  5 WYOMING       2005           508798
##  6 WYOMING       2006           515004
##  7 WYOMING       2007           522830
##  8 WYOMING       2008           532668
##  9 WYOMING       2009           544270
## 10 WYOMING       2010           564554
## # ... with 790 more rows
```

We can see here that the 25 highest prisoner per population observations are either Alaska or Delaware (two low population states, but not the lowest two).

Texas and Oklahoma are the only other states in the top 50 highest prisoner per population observations.

```
#We can use this new data set to look at the most recent data

recent_state_crime <- state_crime %>%
  filter(year == 2016)

recent_state_crime %>%
  arrange(desc(prisoner_per_pop)) %>%
  select(jurisdiction, prisoner_per_pop)
```

```
## # A tibble: 50 x 2
##    jurisdiction prisoner_per_pop
##    <chr>                   <dbl>
##  1 OKLAHOMA              0.00672
##  2 DELAWARE              0.00665
##  3 ARIZONA               0.00612
##  4 ALASKA                0.00590
##  5 TEXAS                 0.00542
##  6 MISSOURI              0.00532
##  7 ARKANSAS              0.00530
##  8 GEORGIA               0.00518
##  9 ALABAMA               0.00489
## 10 FLORIDA               0.00474
## # ... with 40 more rows
```

```
recent_state_crime %>%
  arrange(desc(prisoner_count)) %>%
  select(jurisdiction, prisoner_count)
```

```
## # A tibble: 50 x 2
##    jurisdiction prisoner_count
##    <chr>                 <dbl>
##  1 TEXAS                151276
##  2 CALIFORNIA           129416
##  3 FLORIDA               98010
##  4 GEORGIA               53433
##  5 OHIO                  52172
##  6 NEW YORK              50611
```

```
##  7 PENNSYLVANIA          48287
##  8 ILLINOIS              43616
##  9 ARIZONA               42248
## 10 MICHIGAN              41122
## # ... with 40 more rows
```

We can see here that in 2016 Oklahoma had the highest proportion of population imprisoned, while Texas had the highest raw number.

We also observe that the only states in the top 10 `prisoners_per_pop` and the top 10 `prisoner_count` are Texas, Florida, Arizona.

```
ggplot(data = state_crime,
       mapping = aes(x = year,
                     y = prisoner_per_pop,
                     color = jurisdiction)) +
  geom_point() +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```