Memo: April 19

Data on Federal and State Prison Population, 1926-1986

Nathan Alexander

Title

```
# read in data and update
read_csv("/Users/nathanalexander/Dropbox/Projects/prisons/data/fed_state_prison_pop_1926_1986.csv",
         col_types = cols(Year = col_date(format = "%Y"),
                          Total = col_number(),
                          `Total Percentage` = col_number(),
                          `Prison Type` = col_factor(levels = c("State + Federal", "State", "Federal")),
                          `White Percentage` = col_number(),
                          `Black Percentage` = col_number())) -> fed_state_prison_pop_1926_1986
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
problems(fed_state_prison_pop_1926_1986) # identify issues with df
```

```
## # A tibble: 4 x 5
##     row   col expected actual file
##   <int> <int> <chr>    <chr>  <chr>
## 1    25     5 a number -      ""
## 2    25     6 a number -      ""
## 3    46     5 a number -      ""
## 4    46     6 a number -      ""
```

```
df <- fed_state_prison_pop_1926_1986
df %>%
  rename(Count = `Total`,
         TotalPct = `Total Percentage`,
         Type = `Prison Type`,
         WhitePct = `White Percentage`,
         BlackPct = `Black Percentage`) %>%
  relocate(Year, Type) -> df
```

```
# subset data
df_state = df %>% filter(Type == "State")
df_state
```

```
## # A tibble: 41 x 6
##    Year       Type  Count TotalPct WhitePct BlackPct
##    <date>     <fct> <dbl>    <dbl>    <dbl>    <dbl>
##  1 1926-01-01 State 38318      100       75       23
##  2 1927-01-01 State 39041      100       77       22
##  3 1928-01-01 State 42642      100       NA       NA
##  4 1929-01-01 State 49172      100       76       23
##  5 1930-01-01 State 56213      100       75       24
##  6 1931-01-01 State 60905      100       76       23
##  7 1932-01-01 State 57825      100       76       23
##  8 1933-01-01 State 54468      100       74       25
##  9 1934-01-01 State 52976      100       73       26
## 10 1935-01-01 State 53886      100       72       27
## # i 31 more rows
```

```
df_federal = df %>% filter(Type == "Federal")
df_federal
```

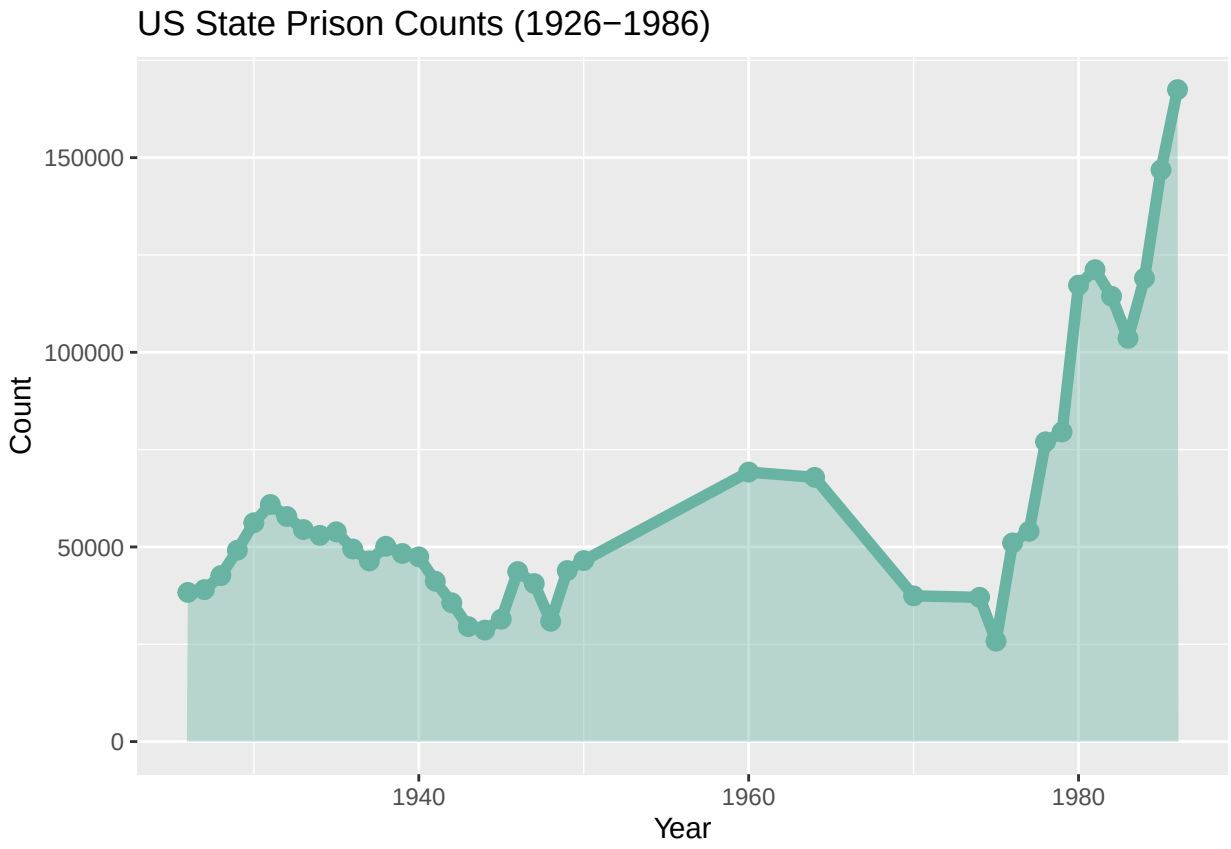
```
## # A tibble: 41 x 6
##    Year       Type    Count TotalPct WhitePct BlackPct
##    <date>     <fct>   <dbl>    <dbl>    <dbl>    <dbl>
##  1 1926-01-01 Federal  5010      100       81       13
##  2 1927-01-01 Federal  5021      100       84       14
##  3 1928-01-01 Federal  5570      100       NA       NA
##  4 1929-01-01 Federal  9734      100       86       12
##  5 1930-01-01 Federal  9800      100       86       12
##  6 1931-01-01 Federal 10615      100       87       11
##  7 1932-01-01 Federal  9652      100       88       10
##  8 1933-01-01 Federal  8333      100       88       10
##  9 1934-01-01 Federal  9275      100       87       11
## 10 1935-01-01 Federal 11837      100       84       14
## # i 31 more rows
```

```
df_both = df %>% filter(Type == "State + Federal")
df_both
```

```
## # A tibble: 41 x 6
##    Year       Type            Count TotalPct WhitePct BlackPct
##    <date>     <fct>           <dbl>    <dbl>    <dbl>    <dbl>
##  1 1926-01-01 State + Federal 43328      100       78       21
##  2 1927-01-01 State + Federal 44062      100       78       21
##  3 1928-01-01 State + Federal 48212      100       78       21
##  4 1929-01-01 State + Federal 58906      100       78       21
##  5 1930-01-01 State + Federal 66013      100       77       22
##  6 1931-01-01 State + Federal 71520      100       77       22
##  7 1932-01-01 State + Federal 67477      100       77       22
##  8 1933-01-01 State + Federal 62801      100       76       23
##  9 1934-01-01 State + Federal 62251      100       75       24
## 10 1935-01-01 State + Federal 65723      100       74       25
## # i 31 more rows
```

```
# basic plot of state population counts
ggplot(df_state, aes(x=Year, y=Count)) +
  geom_area( fill="#69b3a2", alpha=0.4) +
  geom_line(color="#69b3a2", size=2) +
  geom_point(size=3, color="#69b3a2") +
  ggtitle("US State Prison Counts (1926-1986)")
```

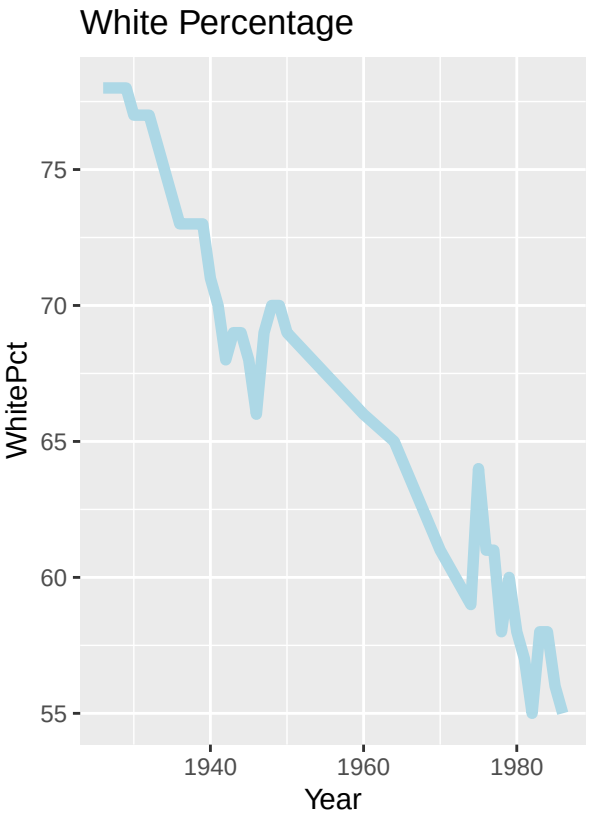
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

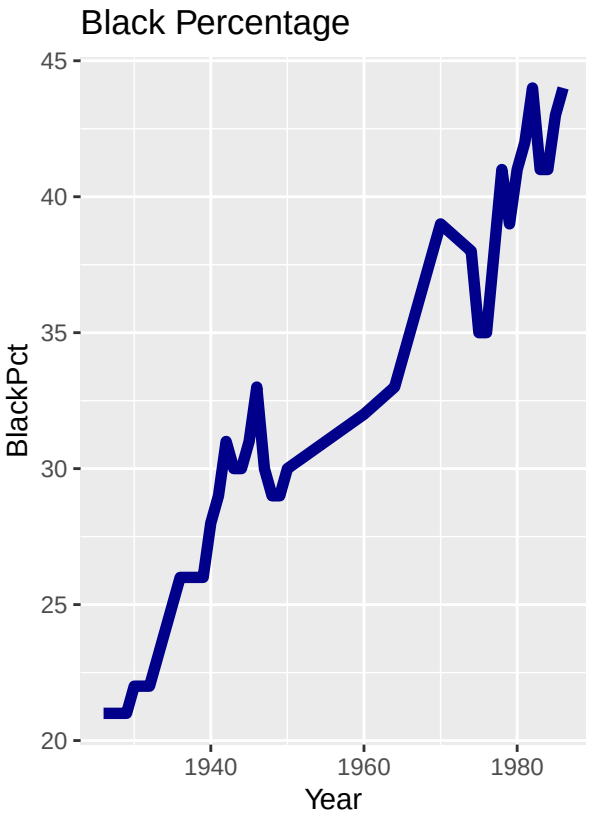



```
# basic plot of white vs. black population counts (state + federal)
p1 <- ggplot(df_both, aes(x=Year, y=WhitePct)) +
  geom_line(color="lightblue", size=2) +
  ggtitle("White Percentage")

p2 <- ggplot(df_both, aes(x=Year, y=BlackPct)) +
  geom_line(color="darkblue",size=2) +
  ggtitle("Black Percentage")
```

```
# Display both charts side by side thanks to the patchwork package
p1 + p2
```


White Percentage
WhitePct
75
70
65
60
55
1940
1960
1980
Year


Black Percentage
BlackPct
45
40
35
30
25
20
1940
1960
1980
Year