

Young_DSC 640_Assignment 4.2_R

Bret Young

10/20/2020

```
# load data
file = 'crimerates-by-state-2005.csv'
data = read.delim(file, header = TRUE, sep = ',')
```

```
# Print data
head(data)
```

state <fctr>	mur... <dbl>	forcible_rape <dbl>	robbery <dbl>	aggravated_assault <dbl>	burglary <dbl>	larceny_t <dbl>
1 United States	5.6	31.7	140.7	291.1	726.7	22
2 Alabama	8.2	34.3	141.4	247.8	953.8	26
3 Alaska	4.8	81.1	80.9	465.1	622.5	25
4 Arizona	7.5	33.8	144.4	327.4	948.4	29
5 Arkansas	6.7	42.9	91.1	386.8	1084.6	27
6 California	6.9	26.0	176.1	317.3	693.3	19

6 rows | 1-8 of 10 columns

```
# Load library for transformation
library(dplyr)

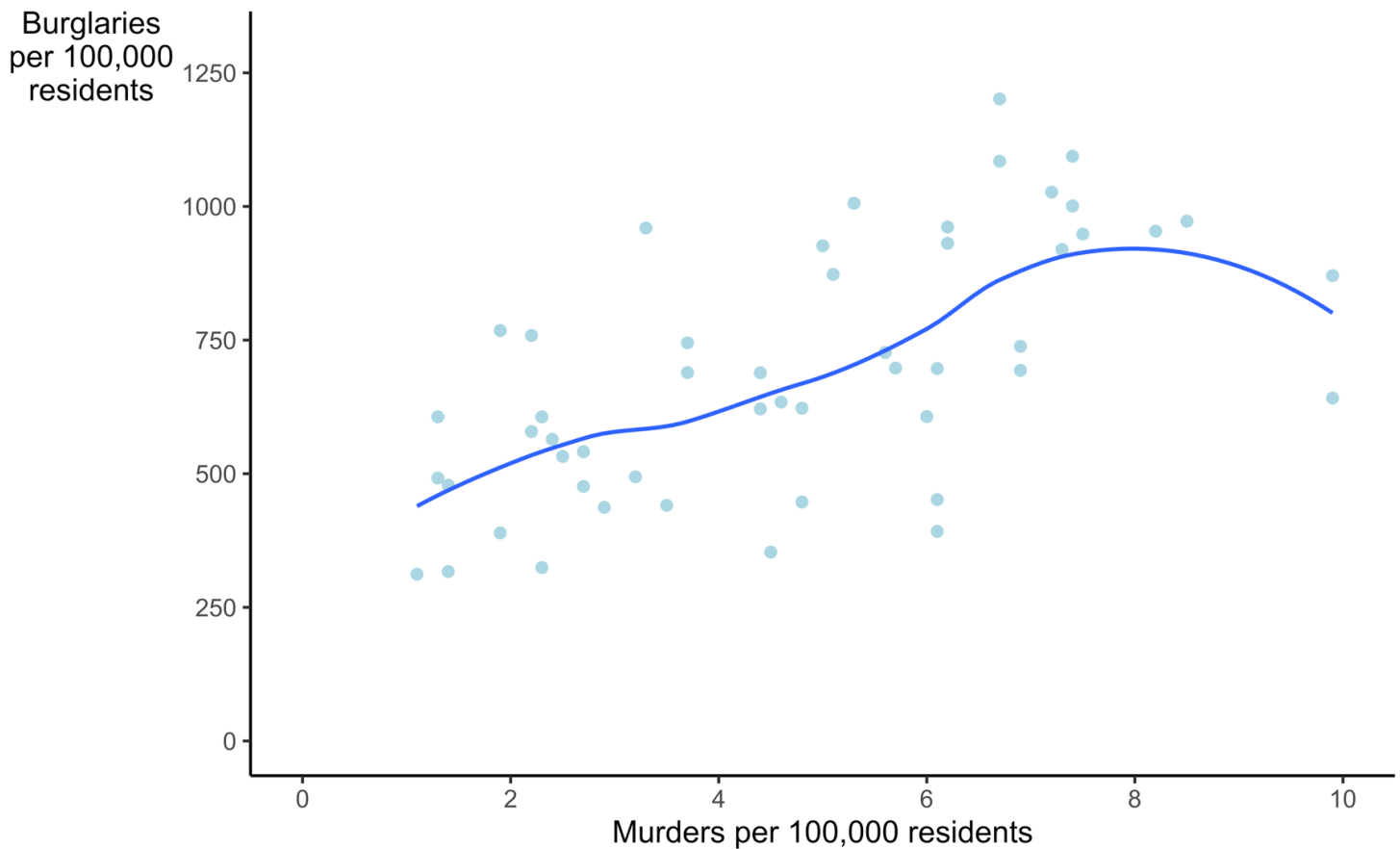
# remove outlier; District of Columbia
data_filter = data %>% filter(state != 'District of Columbia')
```

```
# Load library for plotting data
library(ggplot2)

# Create Plot and custom features
ggplot(data_filter, aes(murder, burglary)) +
  geom_point(color = 'light blue') +
  geom_smooth(method = "loess", size = .7, se = FALSE) +
  ggtitle("Murders Vs Burglaries For States in U.S.",
    subtitle = "Higher murder rates are usually associated with higher burglary rates.") +
  labs(caption = "Source: Data Collected By Nathan Yau from U.S. Census Bureau",
    x = "Murders per 100,000 residents",
    y = "Burglaries per 100,000 residents") +
  theme_classic() +
  theme(plot.title = element_text(face = "bold", size = 18),
    plot.subtitle = element_text(color = "light gray"),
    plot.caption = element_text(color = "light gray"),
    axis.title.y = element_text(angle = 0),
    ) +
  scale_x_continuous(breaks=seq(0, 10, 2), limits = c(0, 10)) +
  scale_y_continuous(breaks=seq(0, 1300, 250), limits = c(0, 1300))
```

Murders Vs Burglaries For States in U.S.

Higher murder rates are usually associated with higher burglary rates.



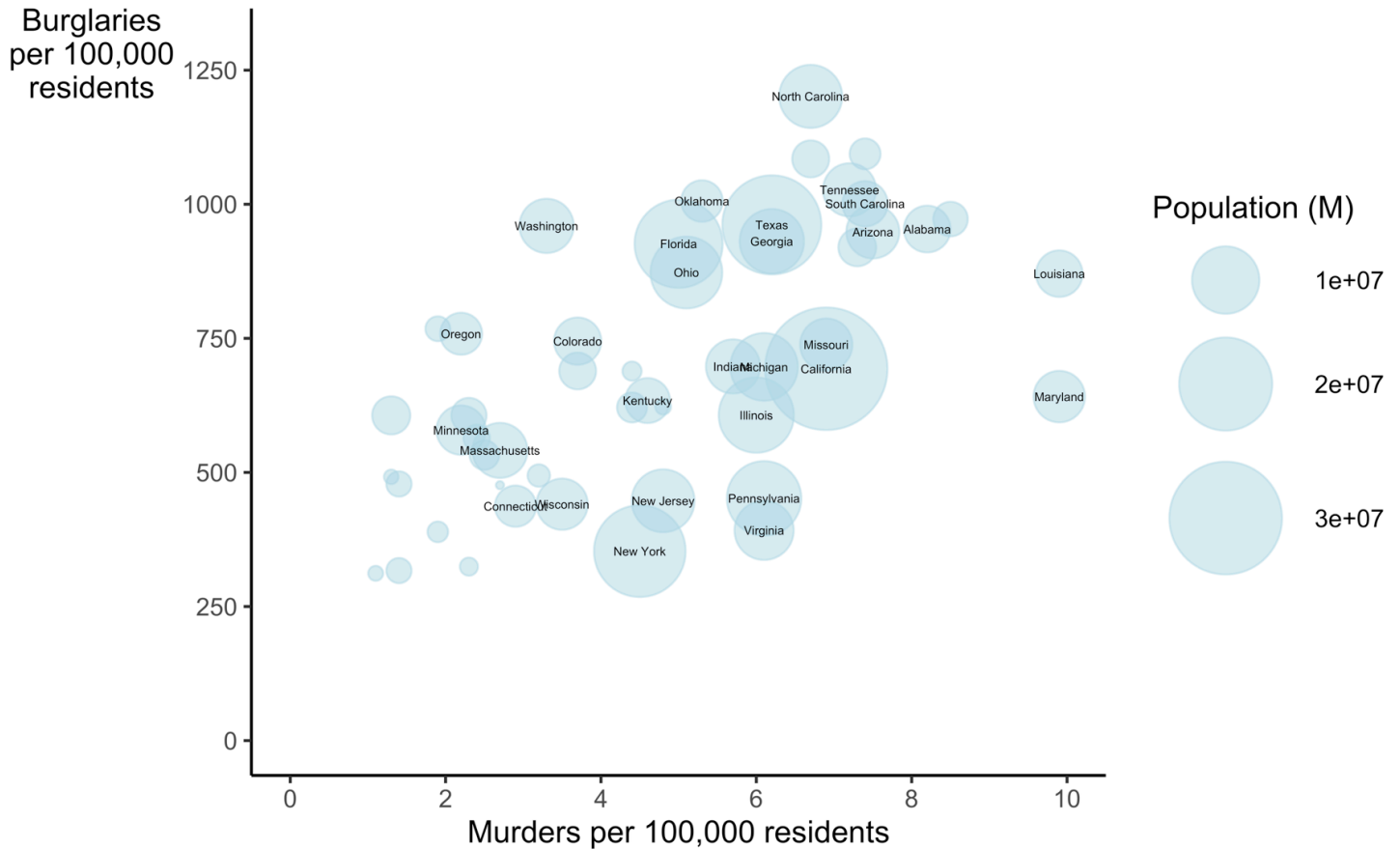
Source: Data Collected By Nathan Yau from U.S. Census Bureau

```
# remove outlier; District of Columbia  
data_filter_US = data_filter %>% filter(state != 'United States')
```

```
# Create Plot and custom features
ggplot(data_filter_US, aes(murder, burglary)) +
  geom_point(aes(size = population), color = 'light blue', alpha = 0.5) +
  geom_text(data = data_filter_US %>% filter(population >= 3000000), aes(label = state), size = 1.5) +
  ggtitle("Murders Vs Burglaries For States in U.S.",
    subtitle = "Higher murder rates are usually associated with higher burglary rates.") +
  labs(caption = "Source: Data Collected By Nathan Yau from U.S. Census Bureau",
    x = "Murders per 100,000 residents",
    y = "Burglaries\nper 100,000\nresidents") +
  theme_classic() +
  theme(plot.title = element_text(face = "bold", size = 18),
    plot.subtitle = element_text(color = "light gray"),
    plot.caption = element_text(color = "light gray"),
    axis.title.y = element_text(angle = 0),
    ) +
  scale_x_continuous(breaks=seq(0, 10, 2), limits = c(0, 10)) +
  scale_y_continuous(breaks=seq(0, 1300, 250), limits = c(0, 1300)) +
  scale_size(range = c(1, 20), name = "Population (M)")
```

Murders Vs Burglaries For States in U.S.

Higher murder rates are usually associated with higher burglary rates.



Source: Data Collected By Nathan Yau from U.S. Census Bureau

```
# load data
file_2 = "birth-rate.csv"
data_2 = read.delim(file_2, header = TRUE, sep = ',')
```

```
head(data_2)
```

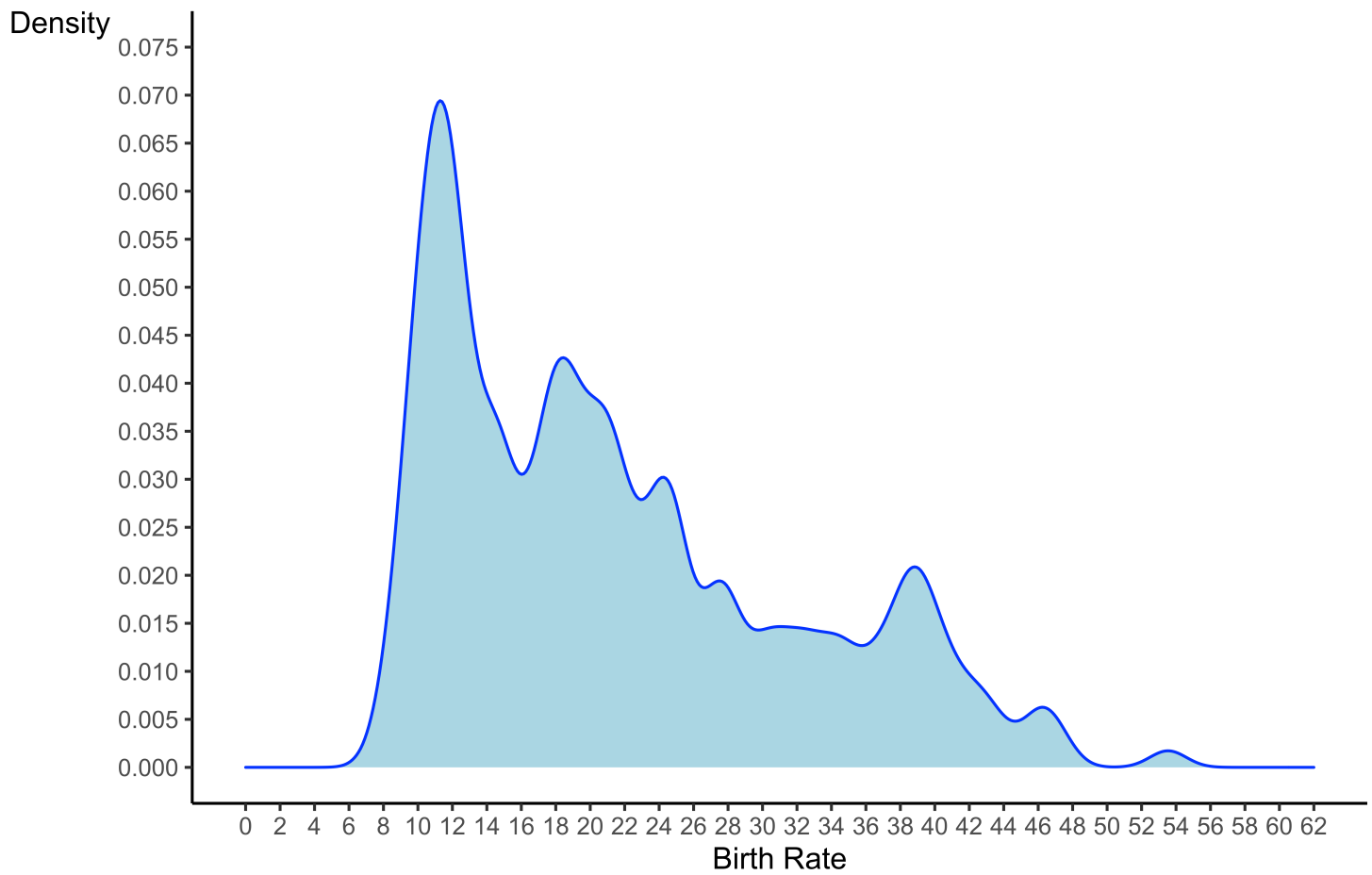
Country <fctr>	X1960 <dbl>	X1961 <dbl>	X1962 <dbl>	X1963 <dbl>	X1964 <dbl>	X1965 <dbl>	X1966 <dbl>	X1967 <dbl>
1 Aruba	36.40000	35.179	33.863	32.459	30.994	29.51300	28.069	26.721
2 Afghanistan	52.20100	52.206	52.208	52.204	52.192	52.16800	52.130	52.076
3 Angola	54.43200	54.394	54.317	54.199	54.040	53.83600	53.585	53.296
4 Albania	40.88600	40.312	39.604	38.792	37.913	37.00800	36.112	35.245
5 Netherlands Antilles	32.32100	30.987	29.618	28.229	26.849	25.51800	24.280	23.173

6 Arab World	47.61122	NA	NA	NA	NA	46.57288	NA	NA
--------------	----------	----	----	----	----	----------	----	----

6 rows | 1-10 of 51 columns

```
# Create Plot and custom features
ggplot(data_2, aes(x = X2008)) +
  geom_density(color = 'blue', fill = 'light blue', adjust = 1/3) +
  ggtitle("Birthrate Density Plot for 2008") +
  labs(caption = "Source: Data Collected By Nathan Yau from World Bank",
       x = "Birth Rate",
       y = "Density") +
  theme_classic() +
  theme(plot.title = element_text(face = "bold", size = 18),
        plot.subtitle = element_text(color = "light gray"),
        plot.caption = element_text(color = "light gray"),
        axis.title.y = element_text(angle = 0),
        ) +
  scale_x_continuous(breaks=seq(0, 62, 2), limits = c(0, 62)) +
  scale_y_continuous(breaks=seq(0, 0.075, .005), limits = c(0, 0.075))
```

Birthrate Density Plot for 2008



Source: Data Collected By Nathan Yau from World Bank