Final Project

Samruddhi Shinde

5/1/2021

Loading packages

library(here)

## here() starts at /Users/SammyShinde/Desktop/GitHub/hw-samruddhis/final\_project

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.3 ✓ purrr 0.3.4  
## ✓ tibble 3.0.6 ✓ dplyr 1.0.3  
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

Loading in Dataset

covid\_prison\_cases <- read\_csv(here("data", "covid\_prison\_cases.csv"))

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## name = col\_character(),  
## abbreviation = col\_character(),  
## staff\_tests = col\_double(),  
## staff\_tests\_with\_multiples = col\_double(),  
## total\_staff\_cases = col\_double(),  
## staff\_recovered = col\_double(),  
## total\_staff\_deaths = col\_double(),  
## prisoner\_tests = col\_double(),  
## prisoner\_tests\_with\_multiples = col\_double(),  
## total\_prisoner\_cases = col\_double(),  
## prisoners\_recovered = col\_double(),  
## total\_prisoner\_deaths = col\_double(),  
## as\_of\_date = col\_character(),  
## notes = col\_character()  
## )

This dataset explores COVID-19 data in prisons across the United States.

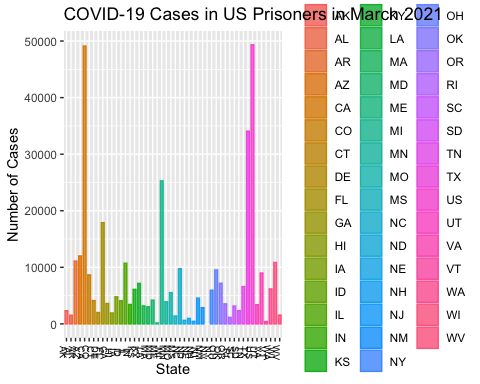
Data Set of Covid-19 Cases for every State in March 2021

march2021\_covid\_cases <- covid\_prison\_cases %>%  
 filter(as\_of\_date == "03/01/2021"|   
 as\_of\_date =="03/02/2021"|   
 as\_of\_date =="03/03/2021"|   
 as\_of\_date == "03/04/2021"|   
 as\_of\_date =="03/05/2021")

Bar plot Covid-19 Cases for every State in March 2021

march2021\_covid\_cases %>%  
 ggplot(aes  
 (x = abbreviation,   
 y = total\_prisoner\_cases,  
 color = abbreviation,  
 fill = abbreviation)) +  
 geom\_bar(stat="identity", alpha=.8, width=.6) +  
 theme(axis.text.x = element\_text(angle=-90, hjust=0.5, size=7,colour="black")) +  
 ggtitle("COVID-19 Cases in US Prisoners in March 2021") +  
 ylab("Number of Cases") +  
 xlab("State") +  
 theme(legend.title = element\_blank())

## Warning: Removed 1 rows containing missing values (position\_stack).

 This bar graph shows that California and federal prisons had the highest cases of COVID-19 in prisoners in March 2021.

us\_covid\_stats <- march2021\_covid\_cases %>%  
 summarize(Average\_Staff\_Cases = mean(total\_staff\_cases, na.rm = TRUE),  
 Min\_Staff\_Cases = min(total\_staff\_cases, na.rm = TRUE),  
 Max\_Staff\_Cases = max(total\_staff\_cases, na.rm = TRUE),  
 Average\_Staff\_Deaths = mean(total\_staff\_deaths, na.rm = TRUE),  
 Min\_Staff\_Deaths = min(total\_staff\_deaths, na.rm = TRUE),  
 Max\_Staff\_Deaths = max(total\_staff\_deaths, na.rm = TRUE),  
 Average\_Prisoner\_Cases = mean(total\_prisoner\_cases, na.rm = TRUE),  
 Min\_Prisoner\_Cases = min(total\_prisoner\_cases, na.rm = TRUE),  
 Maz\_Prisoner\_Cases = max(total\_prisoner\_cases, na.rm = TRUE),  
 Average\_Prisoner\_Deaths = mean(total\_prisoner\_deaths, na.rm = TRUE),  
 Min\_Prisoner\_Deaths = min(total\_prisoner\_deaths, na.rm = TRUE),  
 Max\_Prisoner\_Deaths = max(total\_prisoner\_deaths, na.rm = TRUE))  
  
tibble(us\_covid\_stats)

## # A tibble: 1 x 12  
## Average\_Staff\_C… Min\_Staff\_Cases Max\_Staff\_Cases Average\_Staff\_D…  
## <dbl> <dbl> <dbl> <dbl>  
## 1 2217. 57 15810 4.75  
## # … with 8 more variables: Min\_Staff\_Deaths <dbl>, Max\_Staff\_Deaths <dbl>,  
## # Average\_Prisoner\_Cases <dbl>, Min\_Prisoner\_Cases <dbl>,  
## # Maz\_Prisoner\_Cases <dbl>, Average\_Prisoner\_Deaths <dbl>,  
## # Min\_Prisoner\_Deaths <dbl>, Max\_Prisoner\_Deaths <dbl>

This table shows that on average, prisoners have higher case & death rates compared to staff.

Filtering the covid\_prison\_cases dataset for California statistics Also only looking at staff & prisoner cases/deaths

cali\_covid\_cases <- covid\_prison\_cases %>%  
 filter(name == "California") %>%  
 select(total\_staff\_cases, total\_staff\_deaths, total\_prisoner\_cases, total\_prisoner\_deaths, as\_of\_date)

Summary table of average & sd of staff cases/deaths and prisoner cases/deaths

cali\_covid\_stats <- cali\_covid\_cases %>%  
 summarize(Average\_Staff\_Cases = mean(total\_staff\_cases, na.rm = TRUE),  
 Min\_Staff\_Cases = min(total\_staff\_cases, na.rm = TRUE),  
 Max\_Staff\_Cases = max(total\_staff\_cases, na.rm = TRUE),  
 Average\_Staff\_Deaths = mean(total\_staff\_deaths, na.rm = TRUE),  
 Min\_Staff\_Deaths = min(total\_staff\_deaths, na.rm = TRUE),  
 Max\_Staff\_Deaths = max(total\_staff\_deaths, na.rm = TRUE),  
 Average\_Prisoner\_Cases = mean(total\_prisoner\_cases, na.rm = TRUE),  
 Min\_Prisoner\_Cases = min(total\_prisoner\_cases, na.rm = TRUE),  
 Maz\_Prisoner\_Cases = max(total\_prisoner\_cases, na.rm = TRUE),  
 Average\_Prisoner\_Deaths = mean(total\_prisoner\_deaths, na.rm = TRUE),  
 Min\_Prisoner\_Deaths = min(total\_prisoner\_deaths, na.rm = TRUE),  
 Max\_Prisoner\_Deaths = max(total\_prisoner\_deaths, na.rm = TRUE))  
  
tibble(cali\_covid\_stats)

## # A tibble: 1 x 12  
## Average\_Staff\_C… Min\_Staff\_Cases Max\_Staff\_Cases Average\_Staff\_D…  
## <dbl> <dbl> <dbl> <dbl>  
## 1 5035. 9 15810 8.41  
## # … with 8 more variables: Min\_Staff\_Deaths <dbl>, Max\_Staff\_Deaths <dbl>,  
## # Average\_Prisoner\_Cases <dbl>, Min\_Prisoner\_Cases <dbl>,  
## # Maz\_Prisoner\_Cases <dbl>, Average\_Prisoner\_Deaths <dbl>,  
## # Min\_Prisoner\_Deaths <dbl>, Max\_Prisoner\_Deaths <dbl>

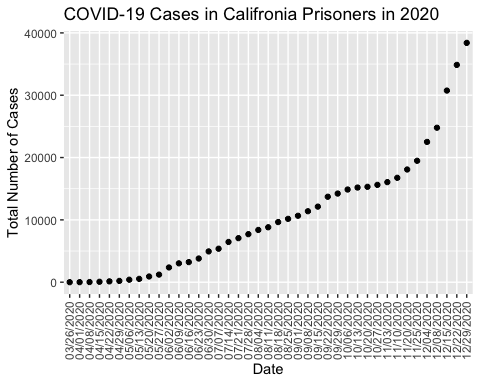
This table shows that on average, prisoners in California have higher case & death rates compared to staff.

Filtering data to only look at 2020 cases

cali\_2020covid\_cases <- cali\_covid\_cases[-c(1:9),]

Scatter Plot for total prisoner cases per date

cali\_2020covid\_cases %>%  
 ggplot(aes  
 (x = as\_of\_date,  
 y = total\_prisoner\_cases)) +  
 geom\_point() +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 ggtitle("COVID-19 Cases in Califronia Prisoners in 2020") +  
 ylab("Total Number of Cases") +  
 xlab("Date")



This scatter plot shows the trend of increasing COVID-19 cases in prisoners during 2020.