DATA SCIENCE INITIATIVE
BROWN UNIVERSITY

5/10/2022

Prediction of Fatal Shootings using Time Series and Mixed Effect Model

DATA2020 Project, Spring 2022 at Brown University

Supervised by **Prof. Alice Paul**

Hanjun Wei, Keying Gong, Yurui Zhang

GitHub Repository: https://github.com/yuruizhang9734/DATA2020FinalProject

Introduction

Motivation

After Michael Brown, an unarmed Black man, was killed in 2014 by police in Ferguson, Mo., a Post investigation found that the FBI **undercounted fatal police shootings** by **more than half**.

Report fatal shootings by police departments is **voluntary** and many departments **fail to do so**

Problem Statement

- What will be the number of monthly fatal shootings by on duty police officers all over the united states next quarter?
- What is the expected number of fatal shootings by demographic groups?

Data Source

- Fatal Shootings:
  - 7951 rows
  - 17 columns
- State Race:
  - 51 rows
  - 7 columns
  - Different race
- Personal Income:
  - 51 rows
  - 28 columns
  - 2015-2021

Data Clean Process

- Check Missing Values: Online in Longitude, Latitude
- Armed -> Different armed levels
- Age -> Different age groups
- State -> Different State regions

EDA

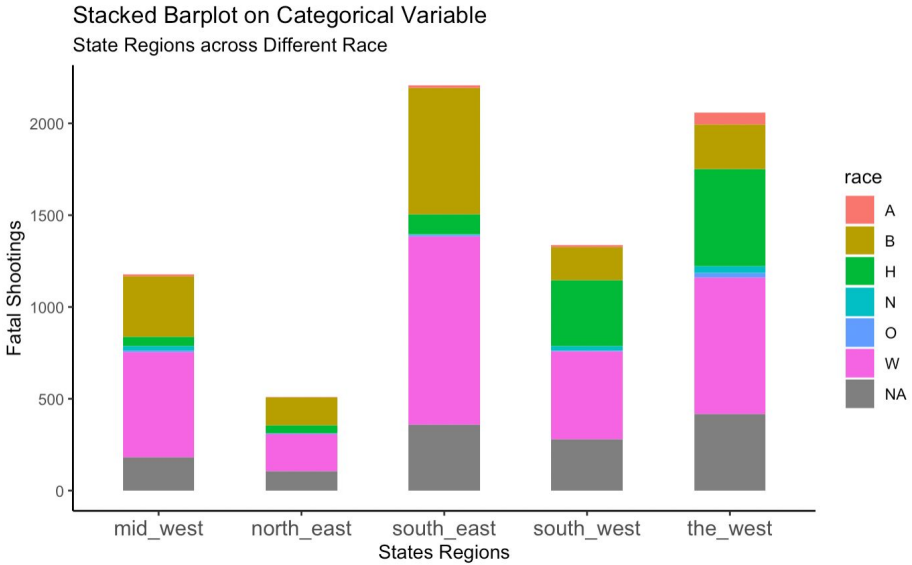


Fig 1. Stacked Barplot of Total Fatal shootings on Categorical Variables State Regions across All races.

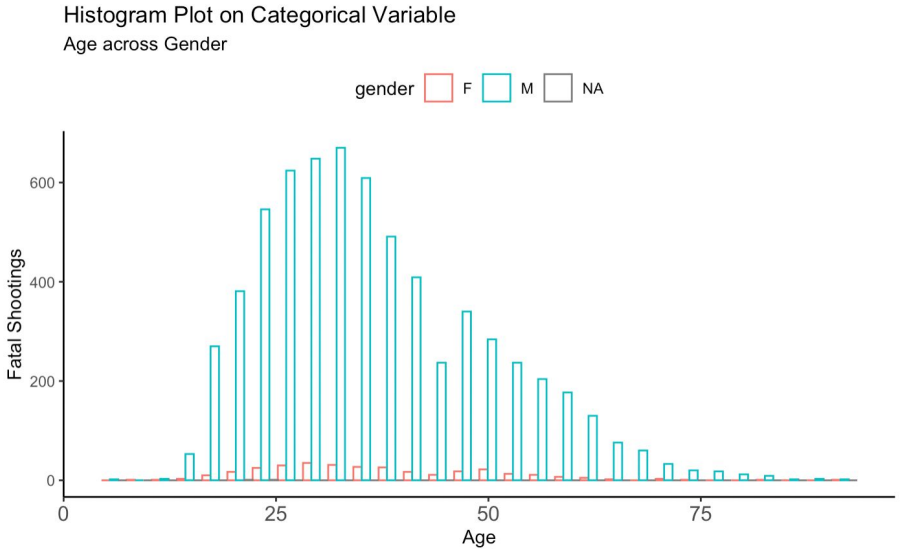


Fig 2. Histogram Plot of Total Fatal shootings on Gender

EDA (State)

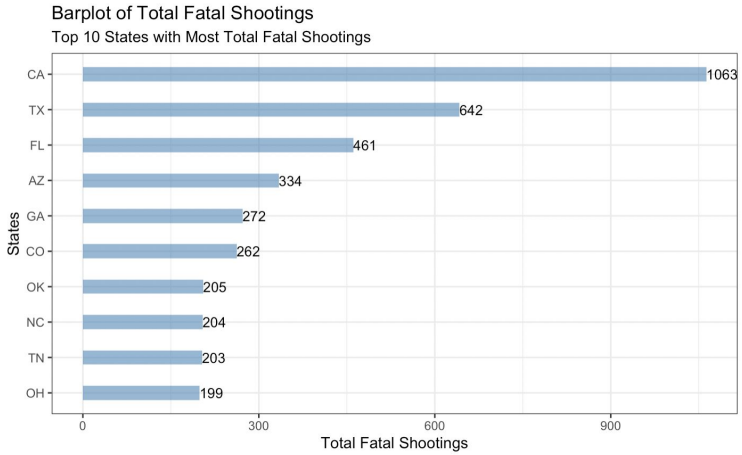


Fig 3. Barplot of the Top 10 States with the most Total Fatal Shootings

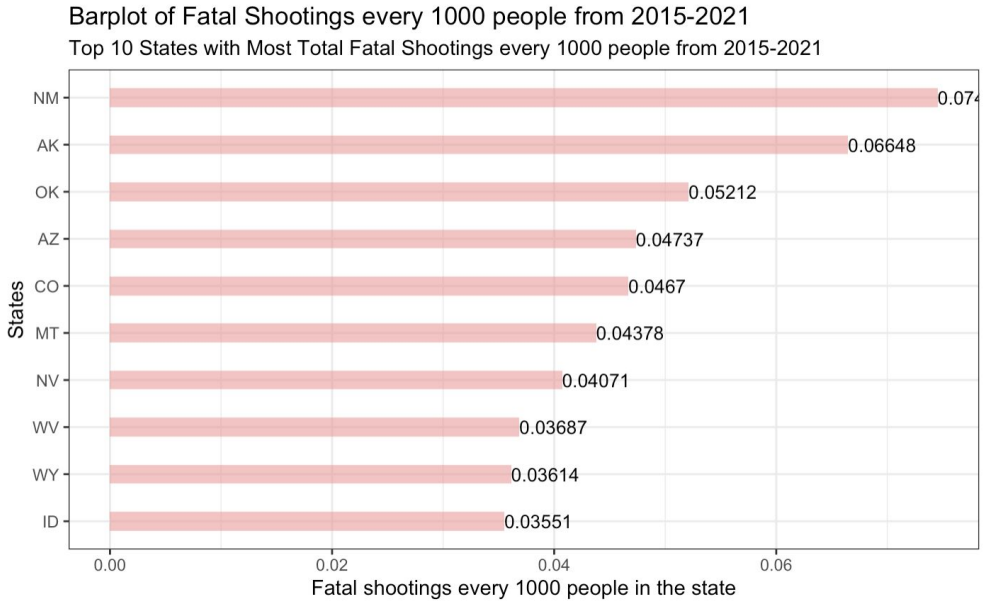


Fig 4. Barplot of the Top 10 States with the most Total Fatal Shootings every 1000 people from 2015-2021

EDA (US Map)

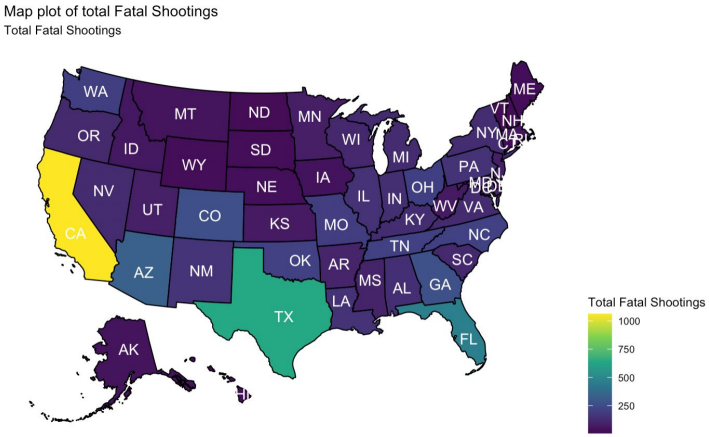


Fig 5. US Map plot of the Total Fatal Shootings

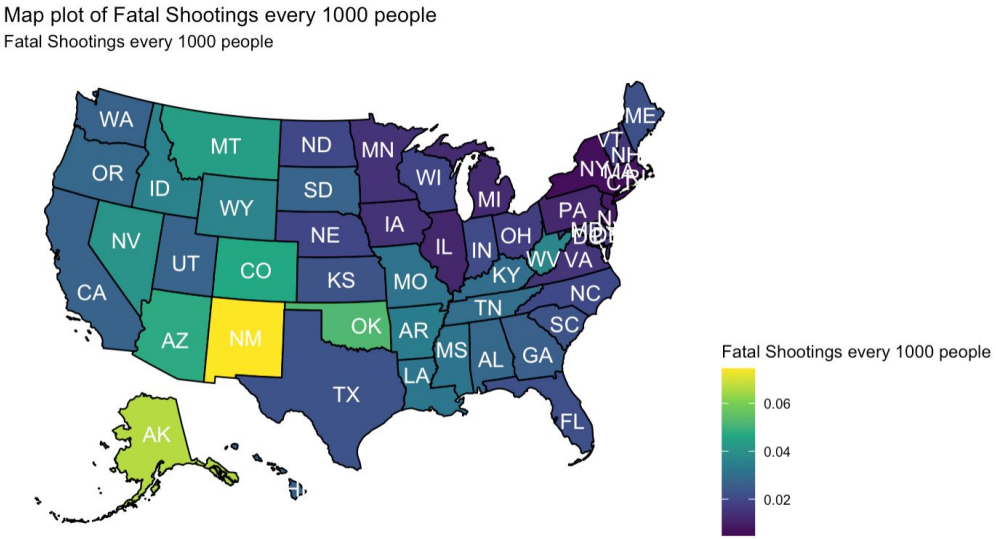


Fig 6. US Map plot of the Total Fatal Shootings every 1000 people from 2015-2021

EDA (Race)

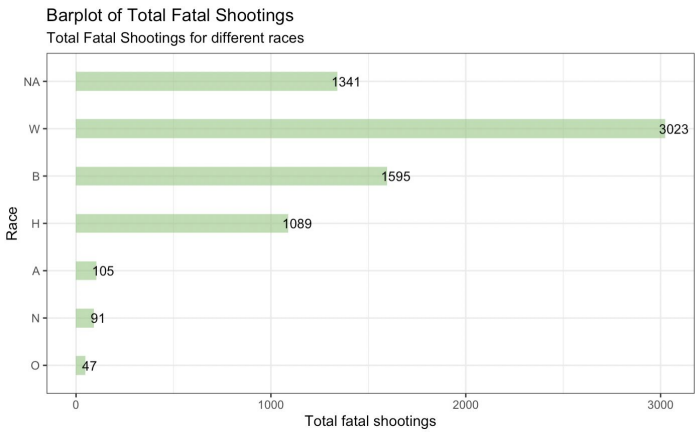


Fig 7. Barplot of the Total Fatal Shootings of all races

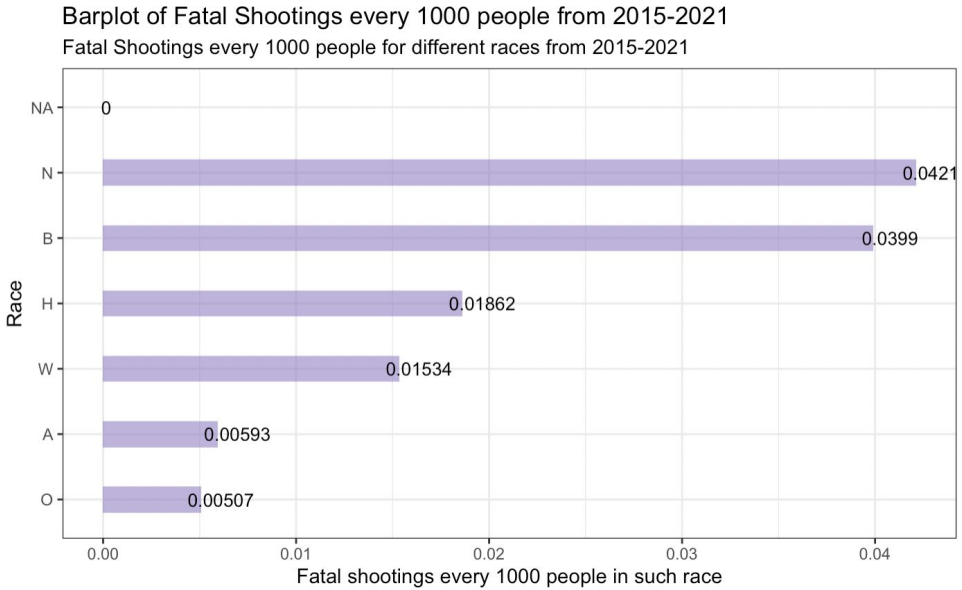


Fig 8. Barplot of the most Total Fatal Shootings every 1000 people of all races from 2015-2021

* Further information needed for how the Washington Post categorize Hispanic and Native Americans
* One mistake in the Washington Post article, they used the population of non-hispanic black to represent the total black (B)

EDA

Fig 9. Scatter Plot of Different States with Population over Per Capita Mean Annual Personal income

Findings

- States with lower Per Capita Mean Annual Personal income has a higher counts per 1000 people from 2015-2021
- The group of south_east and south_west states lie on the left part of the plot, while the north_east lies of the right

Time Trend by Race

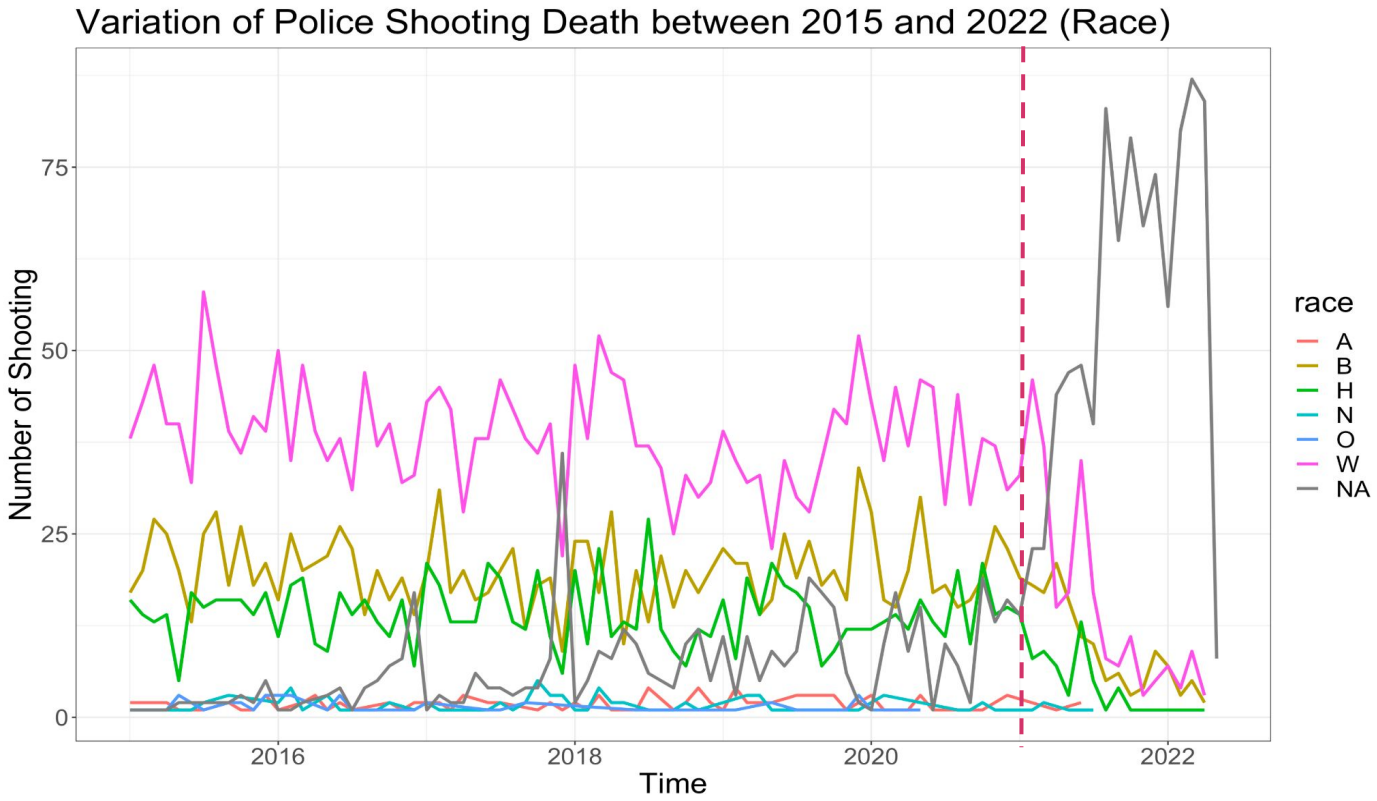


Fig 10.

- Too many NA values after the end of 2020, so we build our model for 2015~2020.

Time Series Analysis and Prediction (All Races)

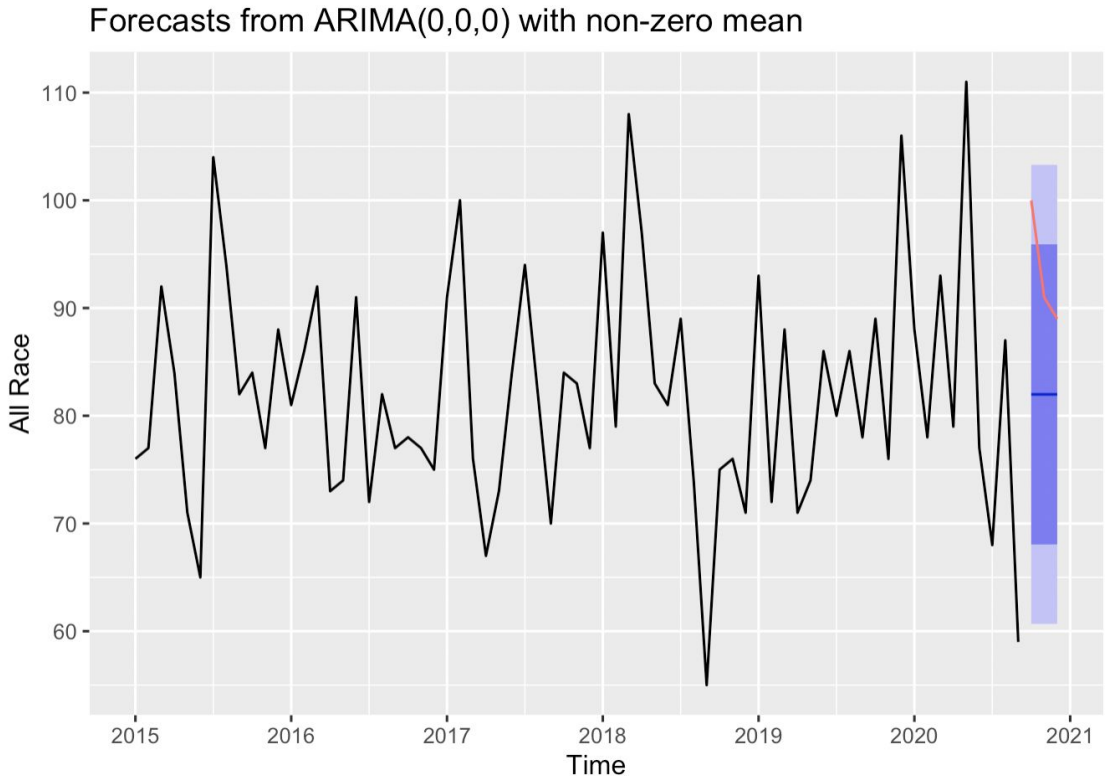


Fig 11.

Diagnostic plots

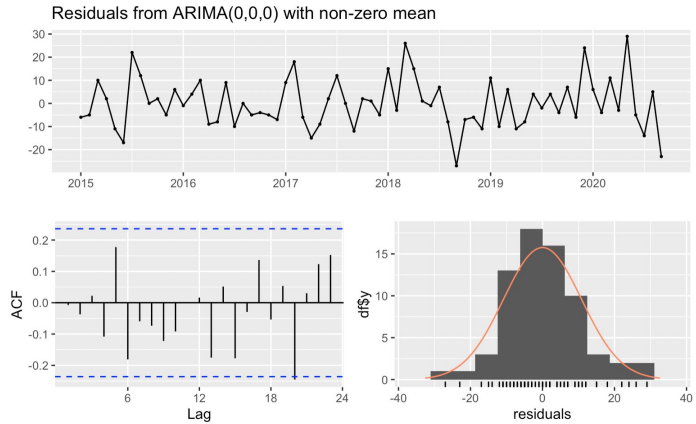


Fig 12.

- The data is composed of the constant mean and white noise.
- The fitted ARIMA(0,0,0) model predicts the mean for the following 3 months.
- According to the residual plots, the fitted model is appropriate.
- The actual shooting counts (2020-10 ~ 2020-12) fall into the 95% prediction interval

Time Series Analysis and Prediction (White)

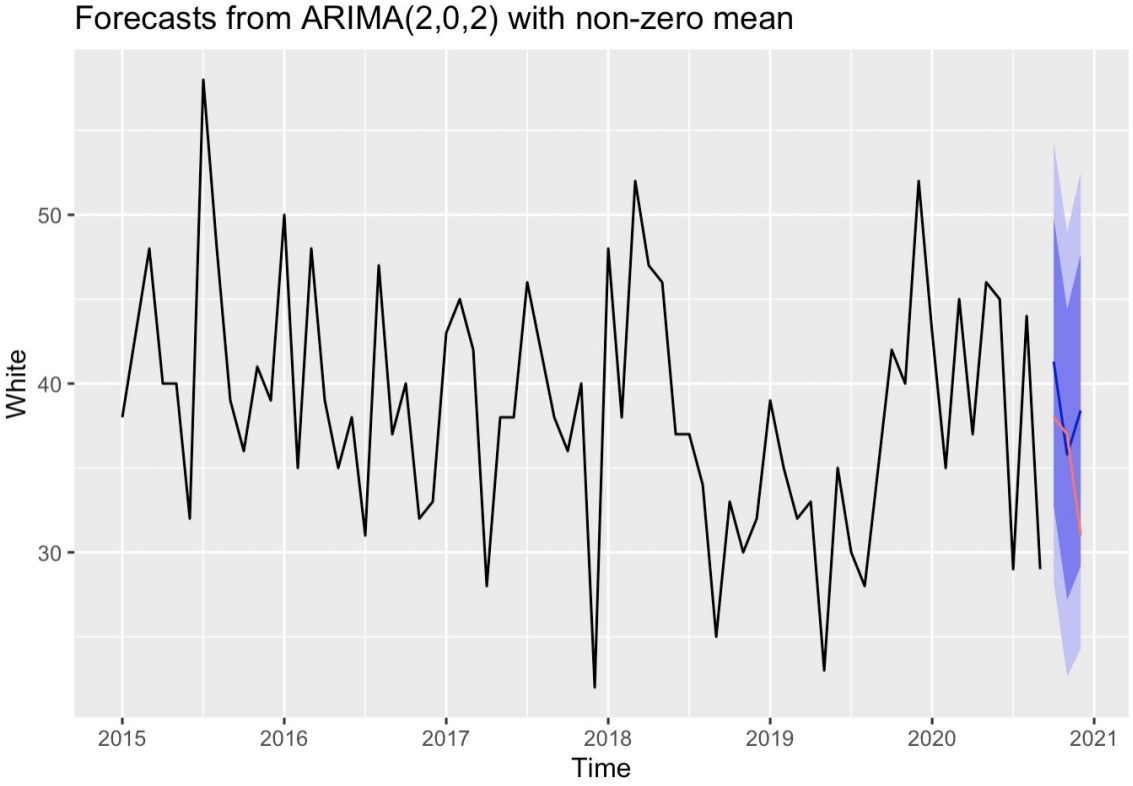


Fig 13.

p = 2 lags of previous values (autoregressive terms)
d = 0
q = 2 lags of errors (moving average terms)

```
Coefficients:
          ar1      ar2     ma1     ma2     mean
      -0.3423  -0.3929  0.4799  0.8333  38.5461
s.e.   0.2107   0.1713  0.1481  0.1130   1.0243
```

The actual shooting counts (2020-10 ~ 2020-12) fall into the 80% prediction interval

Time Series Analysis and Prediction (Hispanic)

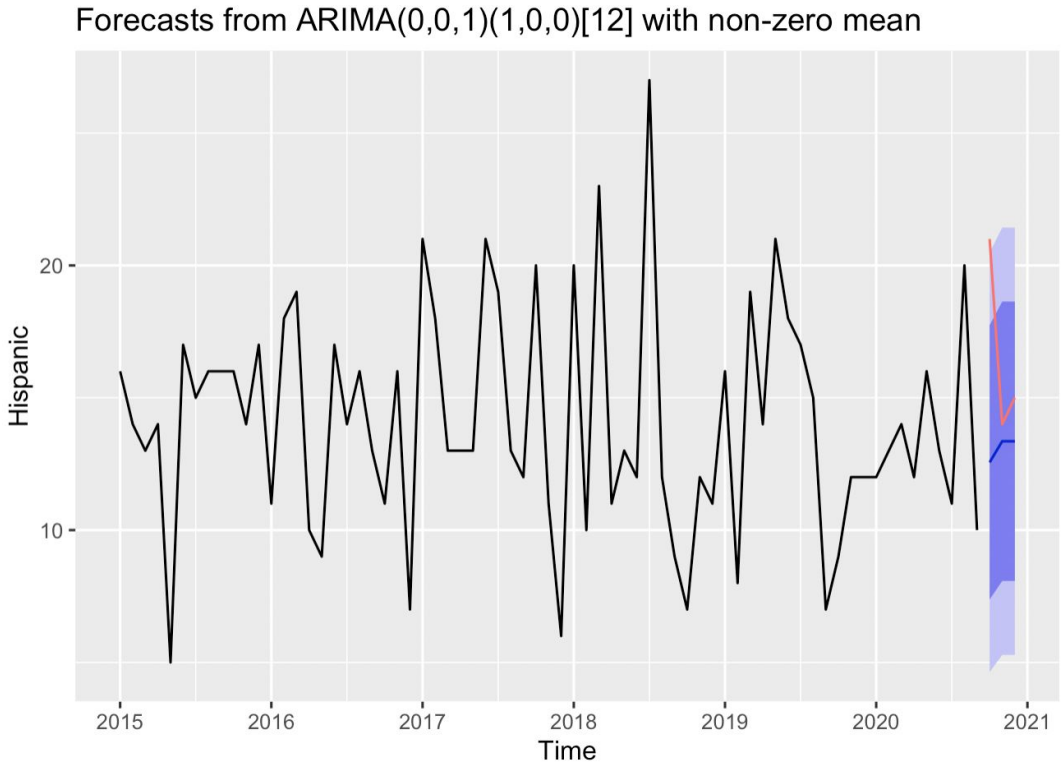


Fig 14.

Non-seasonal trend:
MA(1) lag of errors

Seasonal trend:
AR(1) uses x_{t-12} to predict x_t

	ma1	sar1	mean
	-0.1938	0.3188	13.9910
s.e.	0.1271	0.1142	0.5251

Time Series Analysis and Prediction (Black)

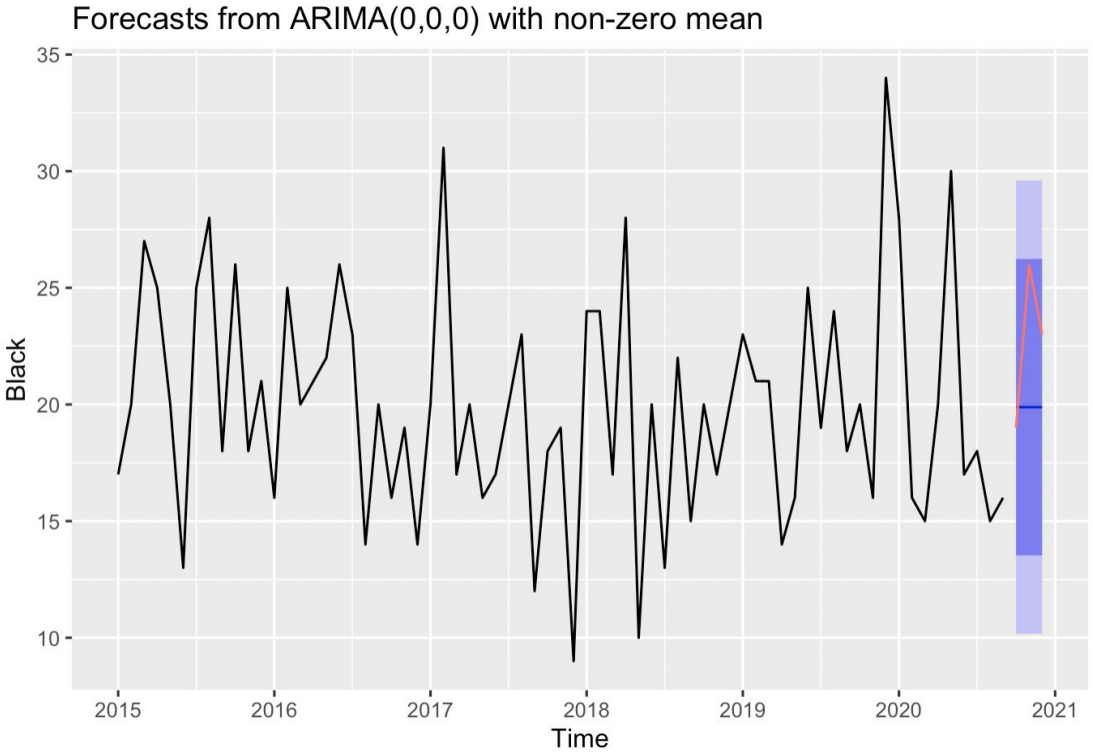


Fig 15.

- The errors are not correlated across time. The fitted ARIMA(0,0,0) model predicts the mean for the following 3 months.
- Since we have few observations for “Native American” and “Other”, we could not fit a meaningful time series model

Shooting Counts Dataset

ID <fctr>	date <date>	race <fctr>	state_loc <fctr>	income <dbl>	counts <dbl>
11	2015-02-01	H	mid_west	43877.08	1
15	2015-02-01	H	north_east	51967.89	1
12	2015-02-01	H	south_east	41030.49	2
13	2015-02-01	H	south_west	42639.66	7
14	2015-02-01	H	the_west	49820.83	3
16	2015-02-01	N	mid_west	43877.08	0

- Grouping Police Shooting dataset and Combine external dataset
- **Goal 1**: Exploring the relationship between average income and monthly death counts
- **Goal 2**: Creating a mixed effect model to predict the death count of each race in each state

Distribution Assumption for Counts

- **Assumption 1**: Counts are Poisson Distributed
  - Appealing when dealing with rare events
  - Constrain: Variance and Mean are equivalent
  - Plot indicates counts of death fits the poisson distribution poorly
- **Assumption 2**: Counts are Negative Binomial Distributed
  - A generalization of Poisson regression
  - Update: loosens the restrictive assumption that the variance is equal to the mean
  - Plot indicates Negative Binomial is more suitable in our case

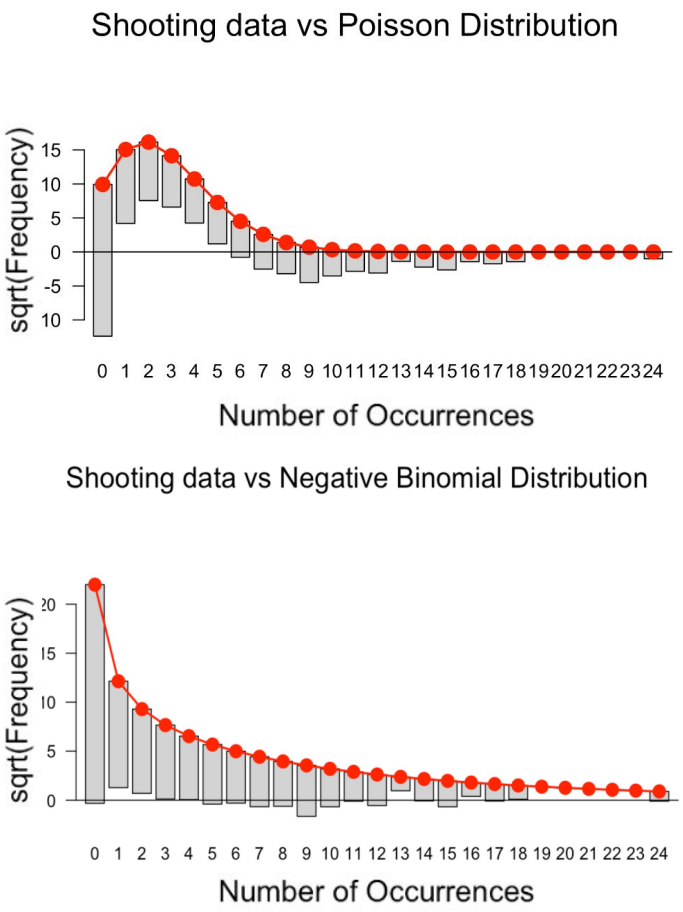
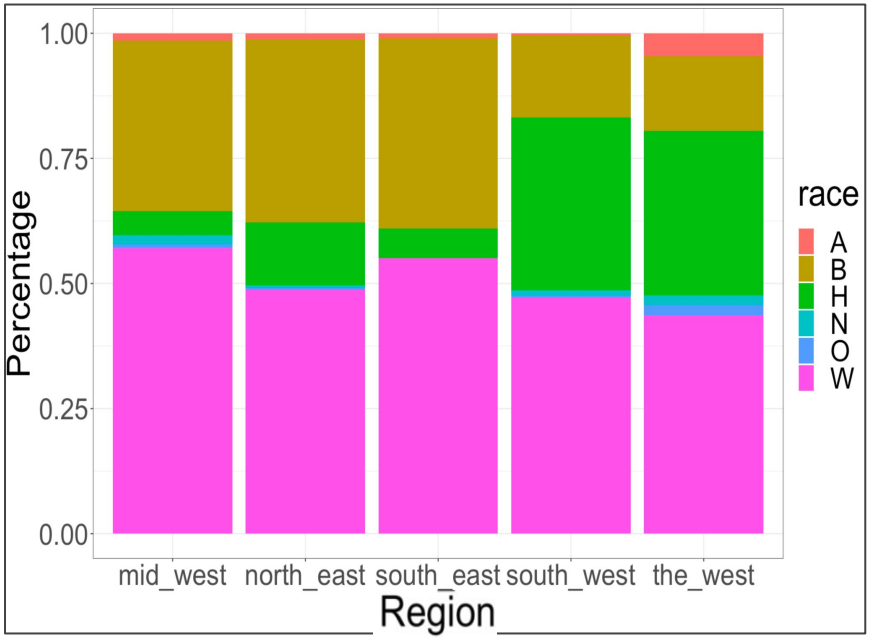


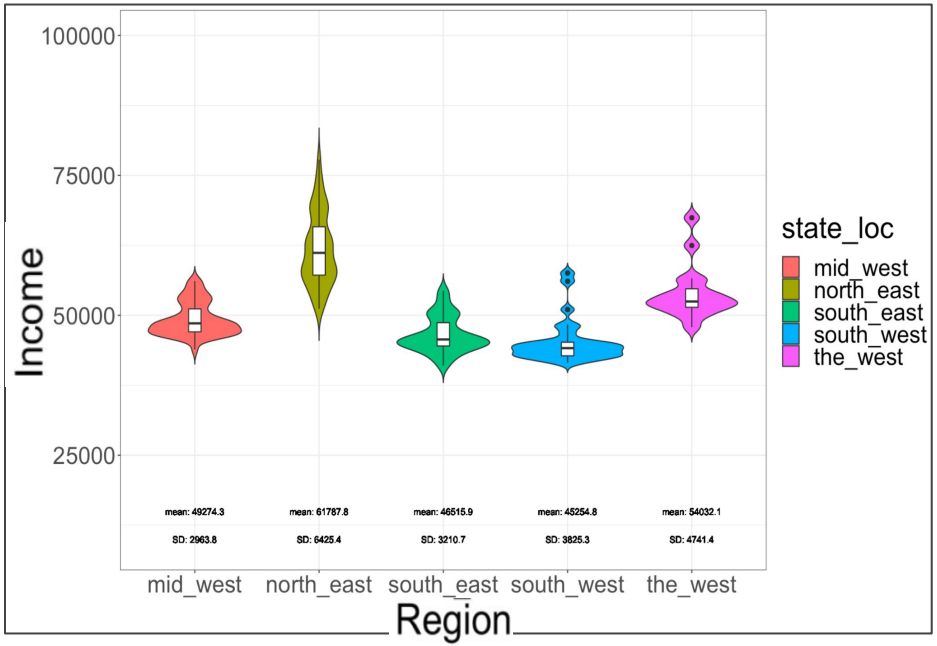
Fig 16.

Correlation Among Predictors



- **Plot 1**: Race v.s. Region
- Sign of Cluster (Random Effect)

Fig 17.



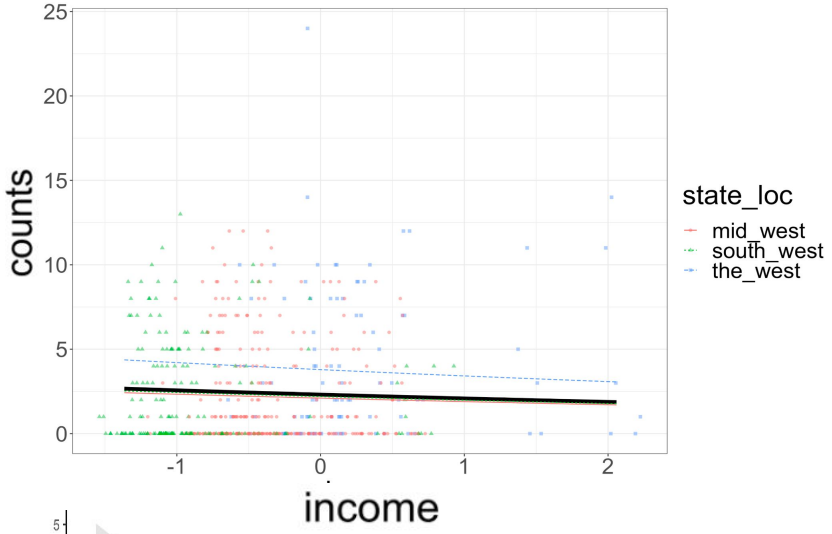
- **Plot 2**: Income v.s. Region
- Sign of Cluster (Random Effect)

Fig 18.

Relationship *Income | Region*

- **Fixed Effect** - Income
- **Random Effect** - Region (state_loc)
- Negative Correlation
- Random Effect for each Region (random Intercept)
- Coefficient income are statistically insignificant

	counts		
Predictors	*Incidence Rate Ratios*	*CI*	*p*
(Intercept)	2.31	1.51 – 3.55	**<0.001**
income	0.90	0.74 – 1.09	0.294
Random Effects			
σ^2	1.35		
$\tau_{00\ state\_loc}$	0.22		
ICC	0.14		
N_{state_loc}	5		
Observations	972		
Marginal R^2 / Conditional R^2	0.007 / 0.146		



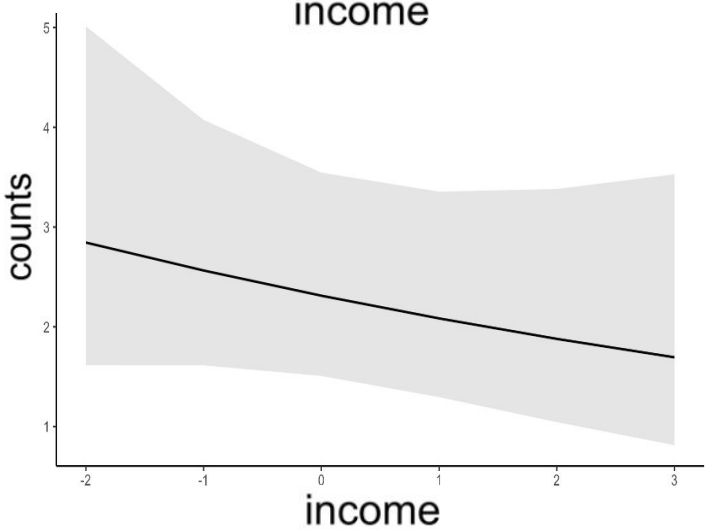


Fig 19.

Prediction Model *Race | Region*

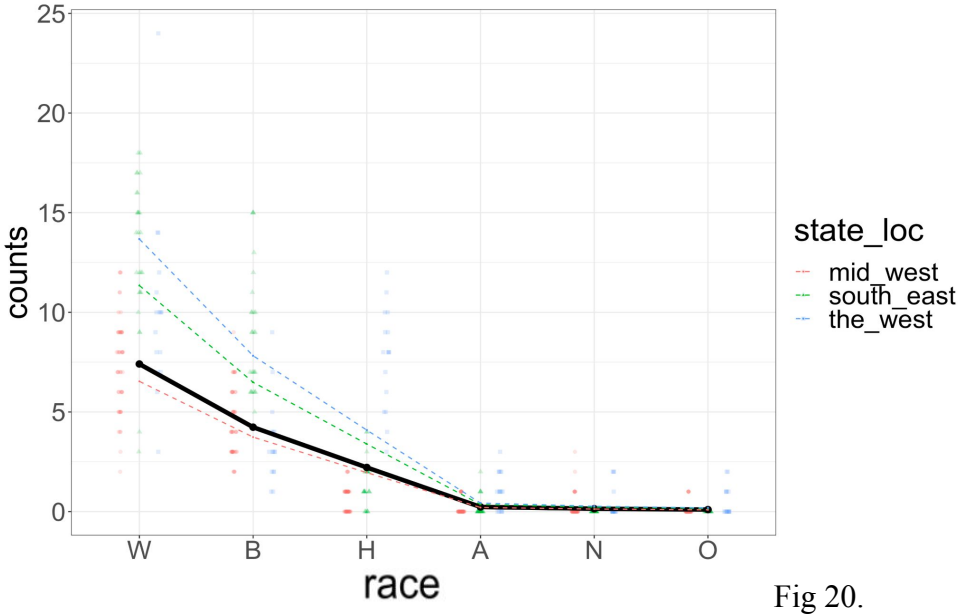


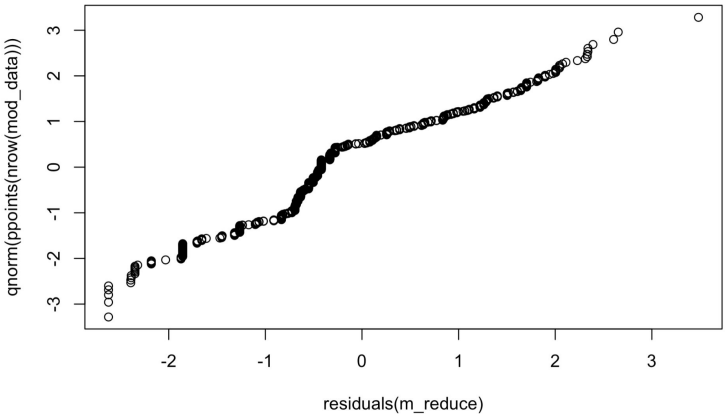
Fig 20.

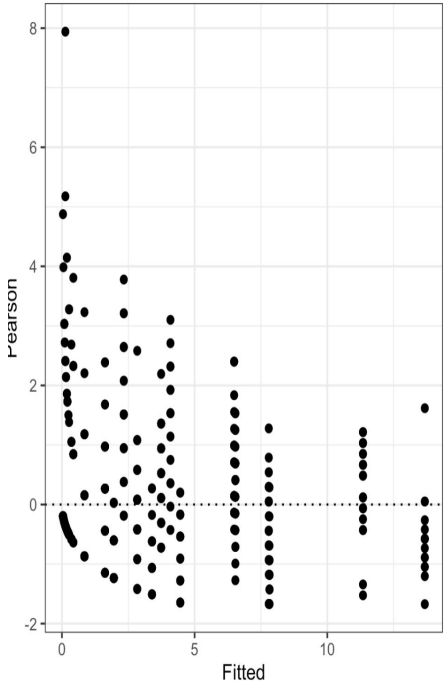
	counts		
Predictors	*Incidence Rate Ratios*	*CI*	*p*
(Intercept)	0.23	0.13 – 0.42	**<0.001**
race [B]	18.22	12.92 – 25.70	**<0.001**
race [H]	9.53	6.71 – 13.53	**<0.001**
race [N]	0.63	0.37 – 1.06	0.082
race [O]	0.43	0.24 – 0.79	**0.006**
race [W]	31.86	22.68 – 44.75	**<0.001**
Random Effects			
σ^2	0.45		
$\tau_{00\ state\_loc}$	0.30		
ICC	0.40		
$N_{\ state\_loc}$	5		
Observations	972		
Marginal R^2 / Conditional R^2	0.794 / 0.877		

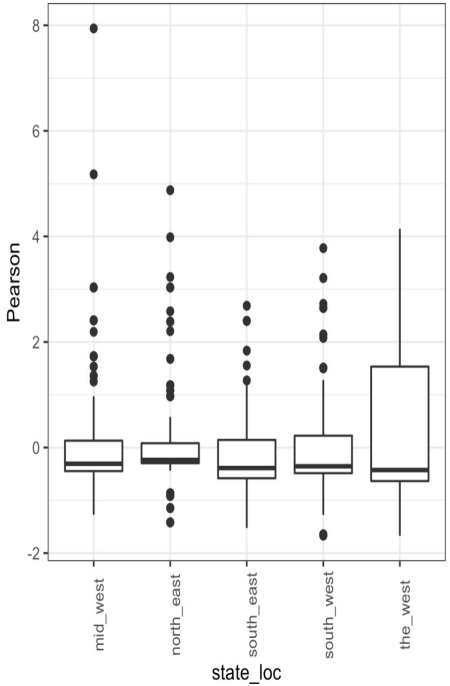
- Select Feature at 0.05 significance level - Race
- 0.40 of variance is explained by Random Effect
- Marginal R square: 0.794
- Conditional R square: 0.877

Diagnostic

- Variation between region indicates random effect is validate
- QQ plot
- Pearson plot







Prediction

predicted counts <dbl>	actual counts <dbl>
0	0
2	1
1	1
0	0
0	0
3	4
0	0
2	1
1	0
0	0

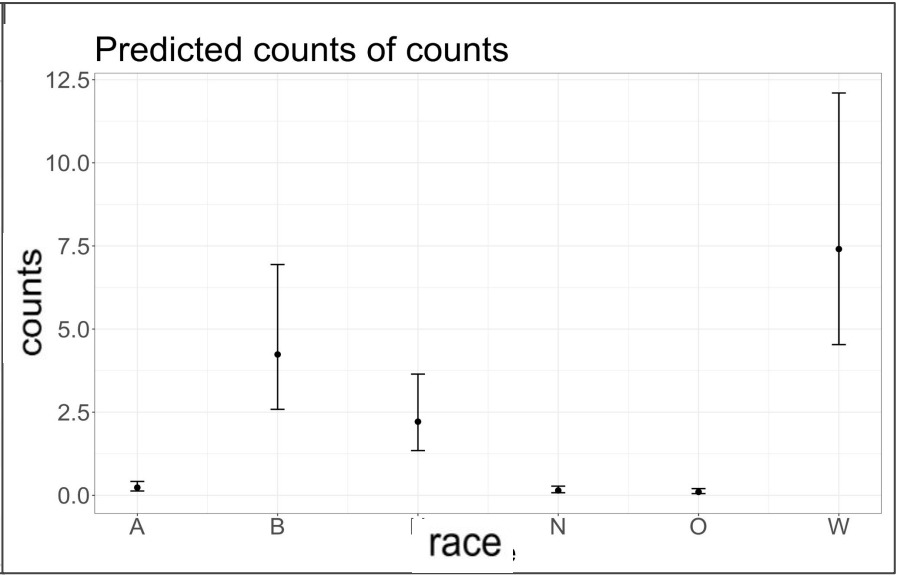


Fig 21.

- Unique prediction interval for each race
- Predict the next 12 unseen observations with MAE of 5

Future Improvement

- Uncategorized Race data from 2022
- More future data to improve model accuracy
- Explore fatal shootings before 2015

- Research and get more precise income data by race within states

- Take more external datasets that relates to both cities and states (e.g. unemployment rate)
- Recheck the state race dataset to match how the Washington Post divide race

Contribution Page

Contribution of each member

- Hanjun Wei:
  - Mixed Effect Model build and evaluate
  - Time series plot Visualization
  - External Data Cleaning
- Keying Gong:
  - Arima Model and evaluate
  - US Map Visualization
  - Mixed Effect Model Improvement
  - PPT Check and proofread
- Yurui Zhang:
  - Data Cleaning
  - External Data Research ,Input and Merge
  - EDA
  - Data Visualization

Citation Page

Dataset and resources

1. 1,050 people have been shot and killed by police in the past year
   Link:https://www.washingtonpost.com/graphics/investigations/police-shootings-database/
2. Police Fatal Shootings:
   Link:https://github.com/washingtonpost/data-police-shootings
3. US States by Race
   Link:https://worldpopulationreview.com/states/states-by-race
4. Regional Economic Accounts: Personal Income
   Link:https://apps.bea.gov/regional/downloadzip.cfm

Thank you and be safe !