### **Group 3**

Junseo Choi, Sojeong Lee, Soyeon Park, Ho-Jae Song, Heewon Yang

**4. Pseudo-code of how the modules will be used**

1. **Data preparation Module**

Processing the data - function:

crime dataset <- read\_csv()

select( select useful variable = Date, Block, Primary Type, Description, Location Description, Arrest… )

rename

separate() Date variable into date, month, and year by whitespace

filter()

1. **Exploratory Data Analysis (EDA) Module**

Summary Statistics - summarise(), mutate(), group\_by()

Visualization - ggplot(), geom\_line, geom\_smooth, geom\_boxplot,geom\_bar(), labs()

data <- data %>%

mutate(categorize the months into the 4 seasons)

data(seasonal) <- data(og) %>%

group\_by(seasons) %>%

summarise(sumarise the crimecount)

ggplot(seasonal, aes(x = season, y = crimecount, fill = crime\_type)) +

geom\_bar(stat = "identity") +

labs(x = "Season", y = "Crime Count", title = "Crime Counts by Season")

#Creating boxplot

crime\_data %>%

ggplot() +

geom\_boxplot(put the x and y)

1. **Hypothesis testing Module**

#Regression Analysis -function: lm(incidents ~ Season, data = crime\_data\_selected)

null\_crime <- data(og) %>%

specify(use the crime amount for response variable and can use season and crime\_type for explanatory variable) %>%

hypothesize(use the code"null = "independece") %>%

generate(reps = 1000, type = "permute") %>%

calculate(stat would be means)

#Visualize

null\_crime %>%

ggplot() +

geom\_histogram(using this to visualize and compare the variables)

facet\_wrap(diversify graphs by crime types or seasons)

calculate observed stat

observed\_stat <- data(og) %>%

specify(same as the previous code) %>%

calculate(same as the previous)

crime\_null %>%

get\_p\_value(use the observed\_stat and choose the direction of the test (right,left,both))

crime\_null %>%

visualize() +

shade\_p\_value(same as the get\_p\_value)

1. **Modeling Module**

(1)Regression model

crime\_data\_selected <- crime\_data\_selected %>%

mutate(Season = factor(Season))

#Linear regression model

model <- lm(explanatory variable and response variable, data used)

e.g ) model <- lm(crime\_rate ~ Season + other\_Variables, data=crime\_data\_selected)

#Visualization of R

model %>%

tidy(can choose what outcome to see i.e. intercept)

model %>%

glance(can choose what outcome to see i.e. r.squared, adj.r.squared, sigma)

#create predictions and residuals

crime\_df <- crime\_data %>%

add\_predictions(model) %>%

add\_residuals(model) %>%

#Linearity test

crime\_df %>%

ggplot() +

geom\_point(usually x = pred) +

geom\_abline(make the slope 1 and intercept 0)

#Constant variability test

crime\_df %>%

ggplot() +

geom\_point(usually x is pred and y is residual) +

geom\_hline(y intercept is 1)

#nearly normal residual tets using QQplot

crime\_df %>%

ggplot() +

geom\_qq(using the sample as residual) +

geom\_qq\_line(using the sample as residual)

(2) Prediction model

#crime\_rate = dependent variable, Season + other\_Variables = independent variable

crime\_model\_1 <- glm (crime\_rate ~ Season + other\_Variables , family = "poisson", data = crime\_data\_selected)

# add predictions

crime\_model\_1\_preds <- crime\_model %>%

add\_predictions(model , use type parameter) %>%

mutate(round to predicted outcome)