

Correlation_Plots

Melody He

2022-06-17

```
library(sf)
library(rgdal)
library(ggpubr)
library(ggplot2)
library(data.table)
library(dplyr)
library(gridExtra)

df <- fread("C:\\Users\\melod\\OneDrive\\Desktop\\STAT 405\\Crime.csv")
## secondary dataset
data1 <- fread("C:\\Users\\melod\\Downloads\\poverty.csv")

df <- df[!is.na(df$Longitude) & !is.na(df$Latitude),]
## obtain subset
sub <- df[df$Year >= 2008 & df$Year <= 2012,]

## import shape file for community areas
nbounds <- read_sf("C:\\Users\\melod\\Downloads\\Boundaries - Neighborhoods\\neighborhoods.shp")
## longitude and latitude coordinates
pnts <- data.frame(sub$Longitude, sub$Latitude)
colnames(pnts) <- c("long", "lat")
## convert to spatial data
pnts_sf <- st_as_sf(pnts, coords = c('long', 'lat'), crs = st_crs(nbounds))

## find intersection in neighborhood bounds
pnts <- pnts_sf %>% mutate(
  intersection = as.integer(st_intersects(geometry, nbounds))
  ,neighborhood = if_else(is.na(intersection), '', nbounds$pri_neigh[intersection]))

## add neighborhood column to dataset
sub <- cbind(sub, pnts$neighborhood)
colnames(sub)[colnames(sub) == 'V2'] <- 'neighborhood'

## subset
sub <- sub[sub$neighborhood == data1$`COMMUNITY AREA NAME`,]

## frequency table dataframe
neigh.freq <- data.frame(table(sub$neighborhood))
colnames(neigh.freq) <- c("neighborhood", "Frequency")

## merge frequency dataframe with secondary dataset
comm <- merge(neigh.freq, data1, by.x = "neighborhood", by.y = "COMMUNITY AREA NAME")
```

```

colors = c("red", "orange", "green", "blue", "magenta", "purple")

## correlation plots with correlation coefficient and p-values
p1 <- ggplot(data = comm, aes(x = Frequency, y = `PERCENT HOUSEHOLDS BELOW POVERTY`)) +
  geom_point(col = colors[1]) +
  geom_smooth(method='lm', formula= y~x, size = 0.25, col = "black", se = FALSE) +
  xlab("Crime Frequency")+ ylab("% Households Below Poverty") +
  stat_cor(label.x = 1000, label.y = 55)

p2 <- ggplot(data = comm, aes(x = Frequency, y = `PERCENT AGED 16+ UNEMPLOYED`)) +
  geom_point(col = colors[2]) +
  geom_smooth(method='lm', formula= y~x, size = 0.25, col = "black", se = FALSE) +
  xlab("Crime Frequency") + ylab("% unemployed (Ages 16+)") +
  stat_cor(label.x = 1000, label.y = 55)

p3 <- ggplot(data = comm, aes(x = Frequency, y = `PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA`)) +
  geom_point(col = colors[3]) +
  geom_smooth(method='lm', formula= y~x, size = 0.25, col = "black", se = FALSE) +
  xlab("Crime Frequency")+ ylab("% Without High School Diploma (Ages 25+)") +
  stat_cor(label.x = 1000, label.y = 55)

p4 <- ggplot(data = comm, aes(x = Frequency, y = `PER CAPITA INCOME`, )) +
  geom_point(col = colors[4]) +geom_smooth(method='lm', formula= y~x, size = 0.25, col = "black", se = FALSE) +
  xlab("Crime Frequency") +
  ylab("Per Capita Income") +
  stat_cor(label.x = 1000, label.y = 70000)

p5 <- ggplot(data = comm, aes(x = Frequency, y = `HARDSHIP INDEX`)) +
  geom_point( col = colors[5]) +
  geom_smooth(method='lm', formula= y~x, size = 0.25, col = "black", se = FALSE) +
  xlab("Crime Frequency") + ylab("Hardship Index (0-100)") +
  stat_cor(label.x = 1000, label.y = 100)

p6 <- ggplot(data = comm, aes(x = Frequency, y = `PERCENT OF HOUSING CROWDED`)) +
  geom_smooth(method='lm', formula= y~x, size = 0.25, col = "black", se = FALSE) +
  geom_point( col = colors[6]) +
  xlab("Crime Frequency") + ylab("% Crowded Housing") +
  stat_cor(label.x = 1100, label.y = 50)

grid.arrange(p1, p2, p3, p4, p5, p6, nrow = 2)

```

