---

title: "datathon"

author: "Monique Sierra"

date: "10/28/2023"

output:

word\_document: default

html\_document: default

pdf\_document: default

---

```{r setup, include=FALSE}

library(dplyr)

library(readxl)

library(ggplot2)

leaCharCopy <- LEA.Characteristics

edgeCopy <- EDGE\_GEOCODE\_PUBLICLEA\_1718

advMathCopy <- Advanced.Mathematics

ussd17Copy <- ussd17

tractCopy <- grf17\_lea\_tract

addressBlockCountList <- tract\_kfr\_rP\_gP\_pall

```

```{r setup, include=FALSE}

difference1 <- setdiff(leaCharCopy$LEAID, edgeCopy$LEAID)

completeDifferenecs <- (leaCharCopy$LEAID == difference1)

head(completeDifferenecs)

```

# minor data improvements

```{r}

# adding first two numbers to match on IDs

ussd17Copy$completeAID <- paste(ussd17Copy$`State FIPS Code`, ussd17Copy$`District ID`, sep = "")

```

```{r}

# 3. Plot the number of students per school district vs. the number of students

# in poverty for all school districts in the US. Explore how all subject matters,

# which you have now imported into PostgreSQL, are related to poverty. Do

# data aggregation and query in PostgreSQL. Plot in your preferred

# programming language.

# the number of students/school district =

```

## joining to solve Q3

```{r}

# joining together the poverty and ID from edge and lea. doing these first bc their longer IDs match. edge has abbrev one added for future join

joiningOnNameForPoverty <- inner\_join(edgeCopy, leaCharCopy, by = "LEAID")

# joining together on shortened ID with other two tables

threeTables <- left\_join(ussd17Copy, joiningOnNameForPoverty, by = c("completeAID"= "LEAID"))

print(head(threeTables))

```

A graph showing a number of colored dots

Description automatically generatedA graph showing a line of poverty

Description automatically generated with medium confidence