Dallas ISD School Attendance Zone Demographic Analysis

GISC 7364 Final Project

Rachel Brasier

11 May 2021

Project Purpose

- Investigate the demographic composition of attendance zones for schools in Dallas ISD
- Determine which schools should expect increased enrollment due to young population

Data Sources and Processes

- Population data (U.S. Census Bureau, 2015-2019 American Communities Survey)
  - Population estimates by sex and age group
  - Via CensusAPI in R
- County and Census Tract Boundaries (U.S. Census Bureau)
- Dallas Independent School District (DISD)
  - Spatial polygons of school attendance zones (Elementary, Middle, High School)
  - Spatial points of campus locations

Population Data Process

- Compile population data using Census API

```
library(censusapi)
names2019_all <- as.data.frame(listCensusMetadata(name = "2019/acs/acs5", type = "variables"))
names2019_geo <- as.data.frame(listCensusMetadata(name = "2019/acs/acs5", type = "geography"))
View(names2019_all[grepl("^SEX BY AGE", names2019_all$concept) == TRUE,])
# "B01001": Sex by Age, Sex by Age by Race
# "B05013": Sex by Age for the Foreign-Born Population
names2019 <- names2019_all[grepl("NAME|B01001|B05013", names2019_all$name) == TRUE,
                           c("name", "label", "concept")]
pop2019_1 <- makeVarlist("2019/acs/acs5", find = "B01001", varsearch = "name")
pop2019_2 <- makeVarlist("2019/acs/acs5", find = "B05013", varsearch = "name")
pop2019 <- getCensus(name = "2019/acs/acs5",
                             vars = c("NAME", pop2019_1, pop2019_2),
                             region = "tract:*",
                             regionin = "state:48+county:113")
row.names(pop2019) <- paste("tract_", pop2019$tract, sep="")
pop2019 <- select(pop2019, -c(1:4)) %>% t() %>% as.data.frame()
pop2019 <- merge(names2019, pop2019, by.x = "name", by.y = 0, all = TRUE)
```

Population Data Process

- Merge with Dallas County Census tracts

```
row.names(pop2019) <- pop2019$name
pop2019.t <- t(pop2019) %>% as.data.frame()
names(pop2019.t) <- gsub("\\W+","_",pop2019.t[c("label","concept"),])
pop2019.t <- pop2019.t[-c(1:3),]
row.names(pop2019.t) <- gsub("tract_", "", row.names(pop2019.t))
rnames <- row.names(pop2019.t)
pop2019.t <- lapply(pop2019.t, as.numeric) %>% as.data.frame()
row.names(pop2019.t) <- rnames

tracts <- st_read(dsn = "data", "tl_2019_48_tract")
tractsDal <- tracts[tracts$COUNTYFP == "113",]
tractsDal <- merge(tractsDal, pop2019.t, by.x = "TRACTCE", by.y = 0)
```

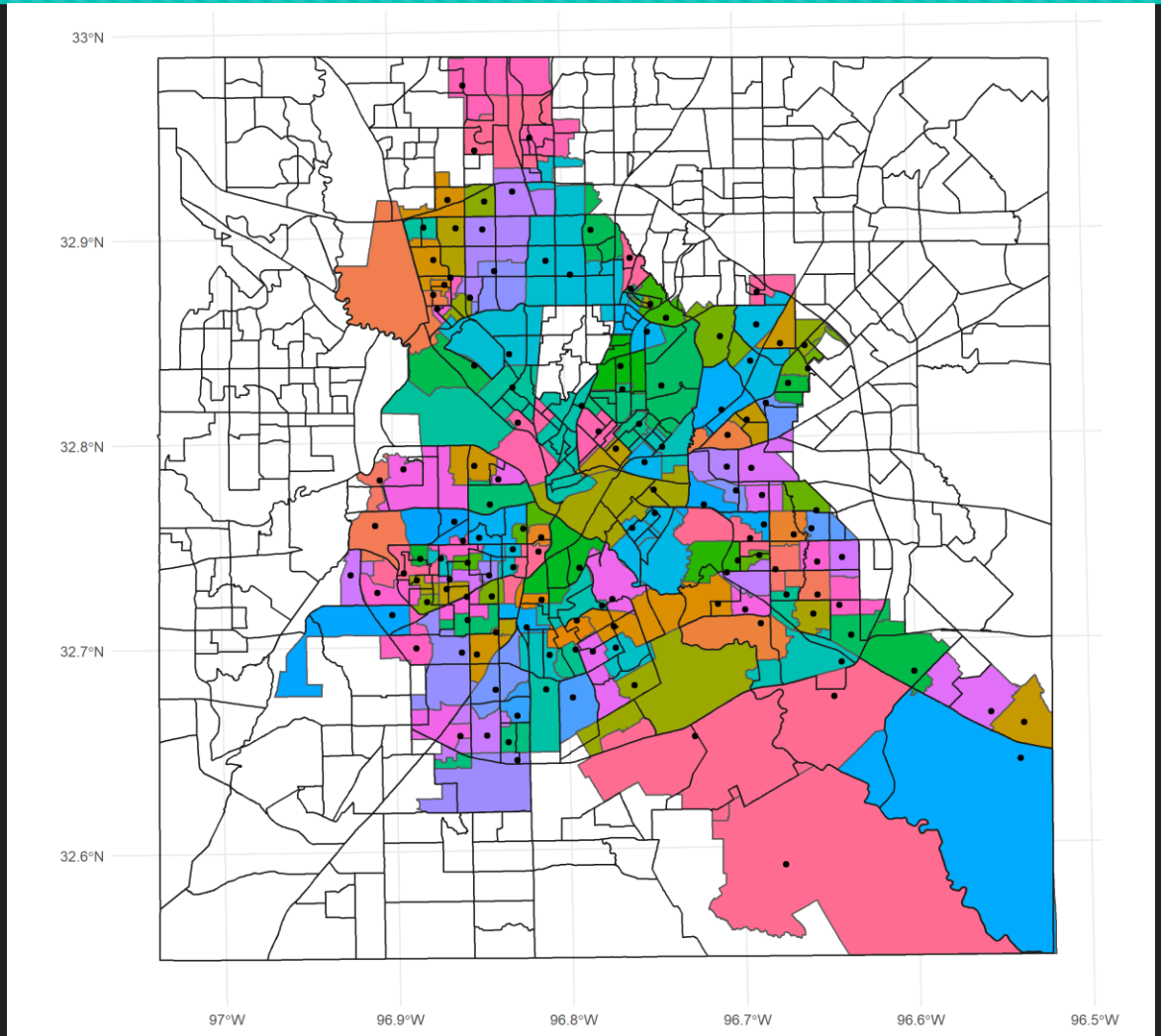
Dallas ISD Shapefiles

- Aggregate within school attendance zones

```
zones_tractsElem <- st_join(zonesElem, tractsDal)
zones_tractsElem$SchoolID <- as.factor(zones_tractsElem$SLN)
zones_tractsElem$WhitePct <- zones_tractsElem$c_Estimate_Total_SEX_BY_AGE_WHITE_ALONE_NOT_HISPANIC_OR_LATINO_/
  zones_tractsElem$c_Estimate_Total_SEX_BY_AGE_

zones_aggElem <- zones_tractsElem[,which(grepl("^SchoolID|^c_", names(zones_tractsElem)) == TRUE)] %>%
  mutate_at(vars(matches("c_")), as.numeric) %>%
  group_by(SchoolID) %>%
  summarise(across(where(is.numeric), list(sum = sum)))
zones_infoElem <- as.data.frame(zonesElem[,2:7])
zones_schoolsElem <- merge(zones_infoElem, zones_aggElem, by.x = "SLN", by.y = "SchoolID")

campuses <- st_read(dsn = "data", "DISD_Campuses")
campuses <- st_transform(campuses, crs = st_crs(zonesElem))
campusesElem <- campuses[campuses$LEVEL == "Elementary",]

# Repeated for middle and high schools (not shown)
```

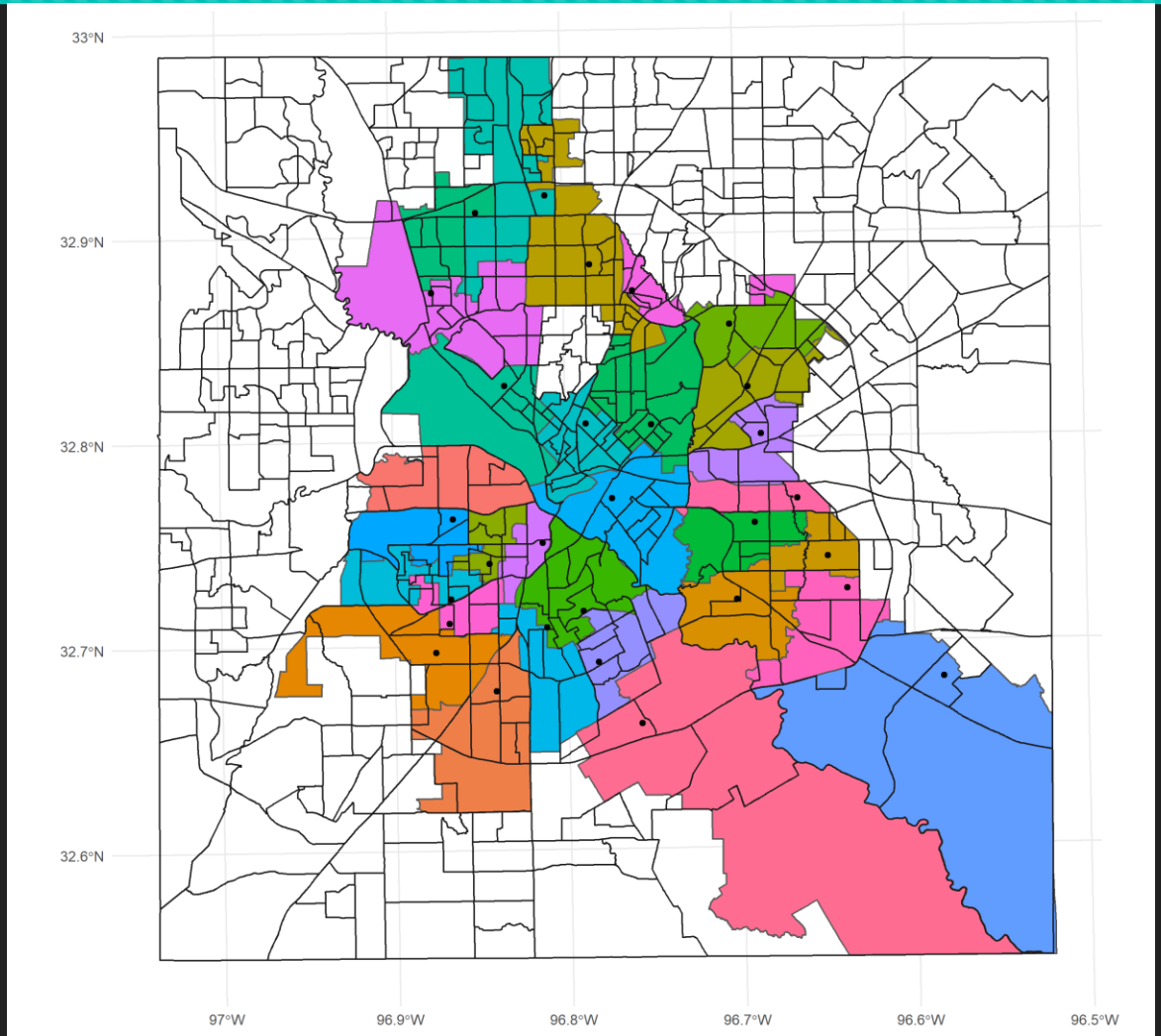
Census Tracts and School Attendance Zones: Elementary Schools

```
ggplot() +
  geom_sf(
    mapping = aes(
      fill = SchoolID),
    data = zones_tractsElem,
    show.legend = FALSE
  ) +
  geom_sf(
    fill = NA,
    color = "grey10",
    data = tractsDal,
    show.legend = FALSE
  ) +
  geom_sf(
    data = campusesElem
  ) +
  theme_minimal()
```



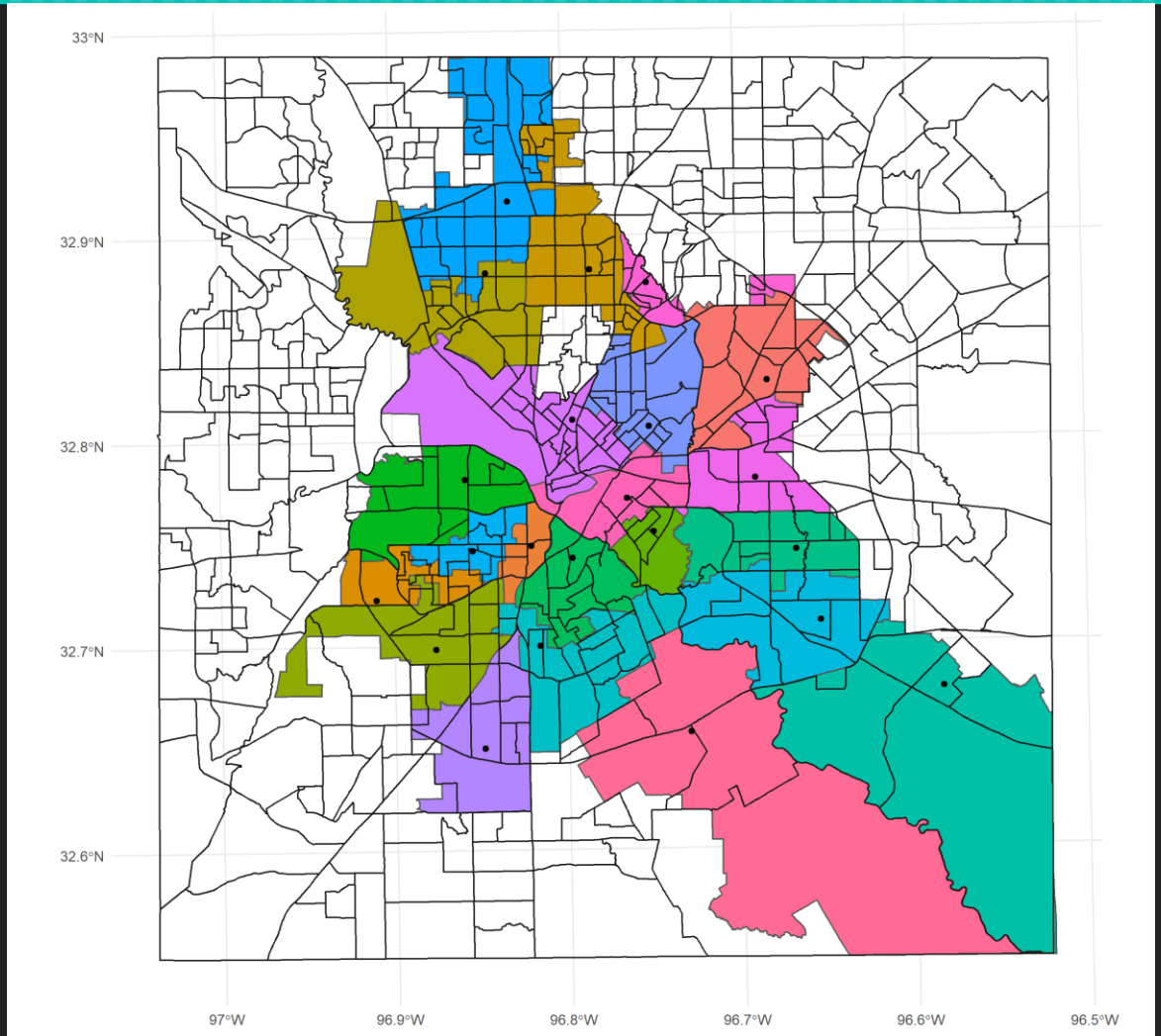
Census Tracts and School Attendance Zones: Middle Schools

```
ggplot() +
  geom_sf(
    mapping = aes(
      fill = SchoolID),
    data = zones_tractsMid,
    show.legend = FALSE
  ) +
  geom_sf(
    fill = NA,
    color = "grey10",
    data = tractsDal,
    show.legend = FALSE
  ) +
  geom_sf(
    data = campusesMid
  ) +
  theme_minimal()
```



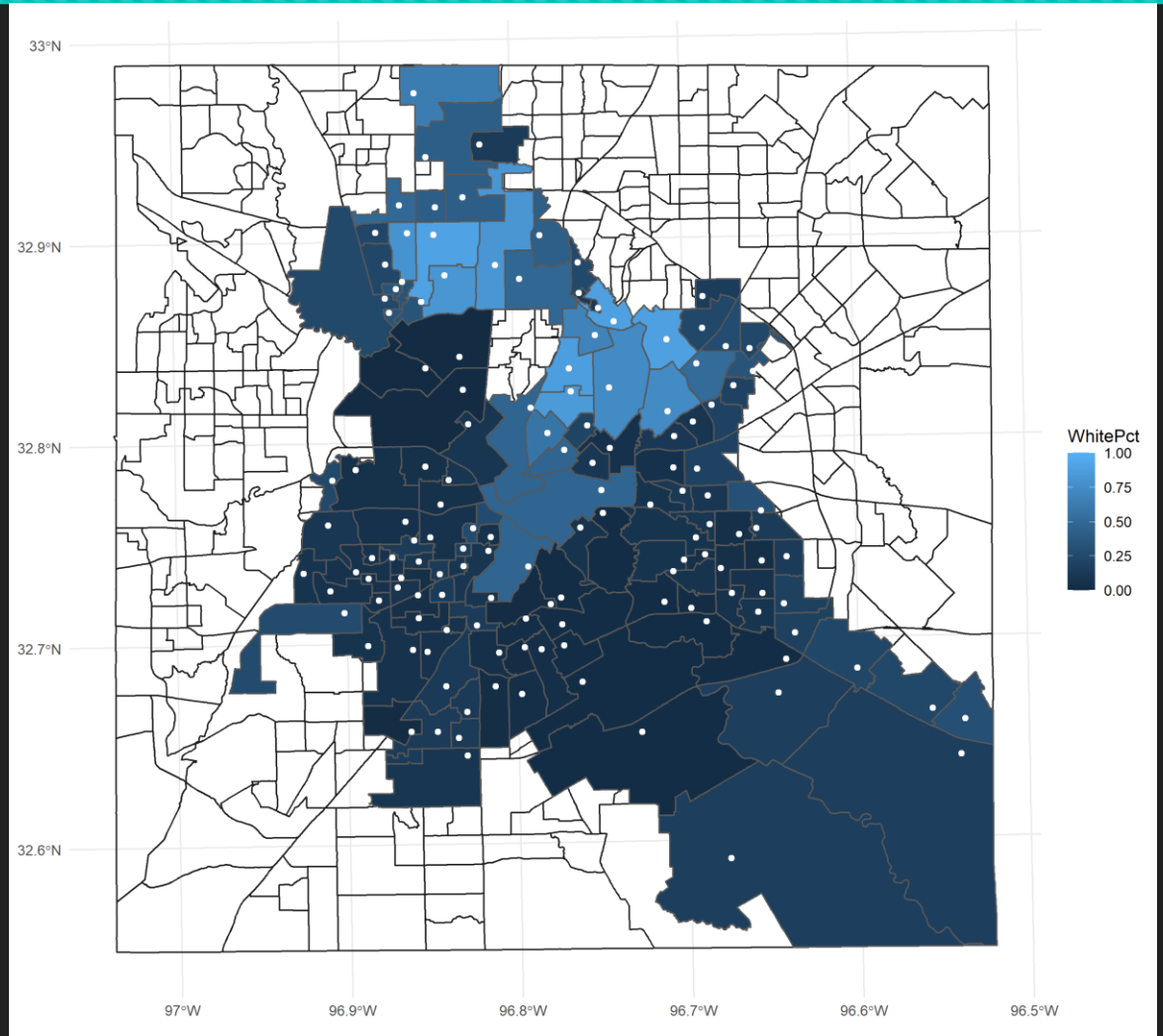
Census Tracts and School Attendance Zones: High Schools

```
ggplot() +
  geom_sf(
    mapping = aes(
      fill = SchoolID),
    data = zones_tractsHigh,
    show.legend = FALSE
  ) +
  geom_sf(
    fill = NA,
    color = "grey10",
    data = tractsDal,
    show.legend = FALSE
  ) +
  geom_sf(
    data = campusesHigh
  ) +
  theme_minimal()
```



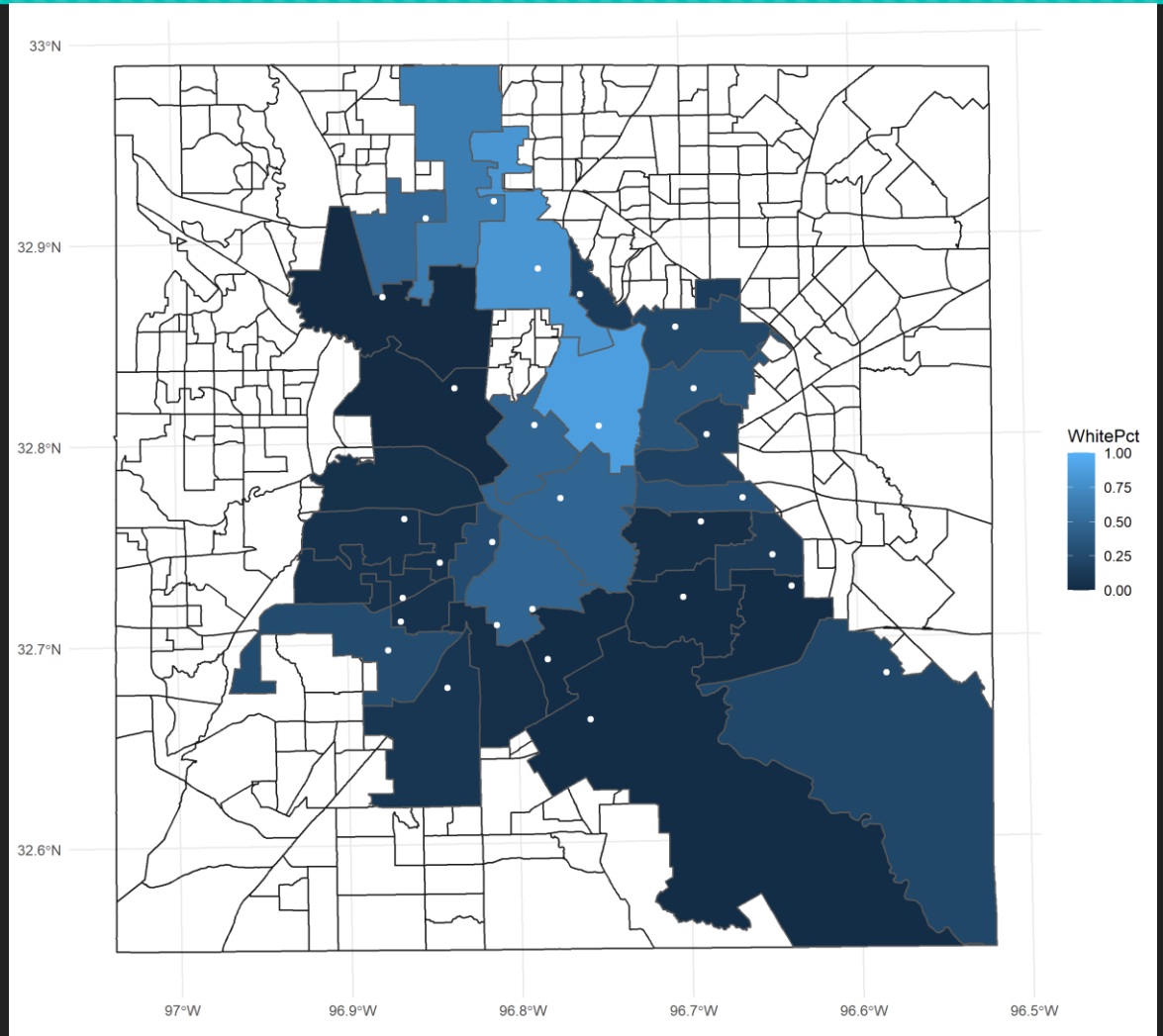
White Population of School Attendance Zones: Elementary Schools

```
ggplot() +
  geom_sf(
    fill = NA,
    color = "grey10",
    data = tractsDal,
    show.legend = FALSE
  ) +
  geom_sf(
    mapping = aes(
      fill = WhitePct),
    data = zones_tractsElem,
    show.legend = TRUE
  ) +
  geom_sf(
    color = "white",
    data = campusesElem
  ) +
  theme_minimal()
```



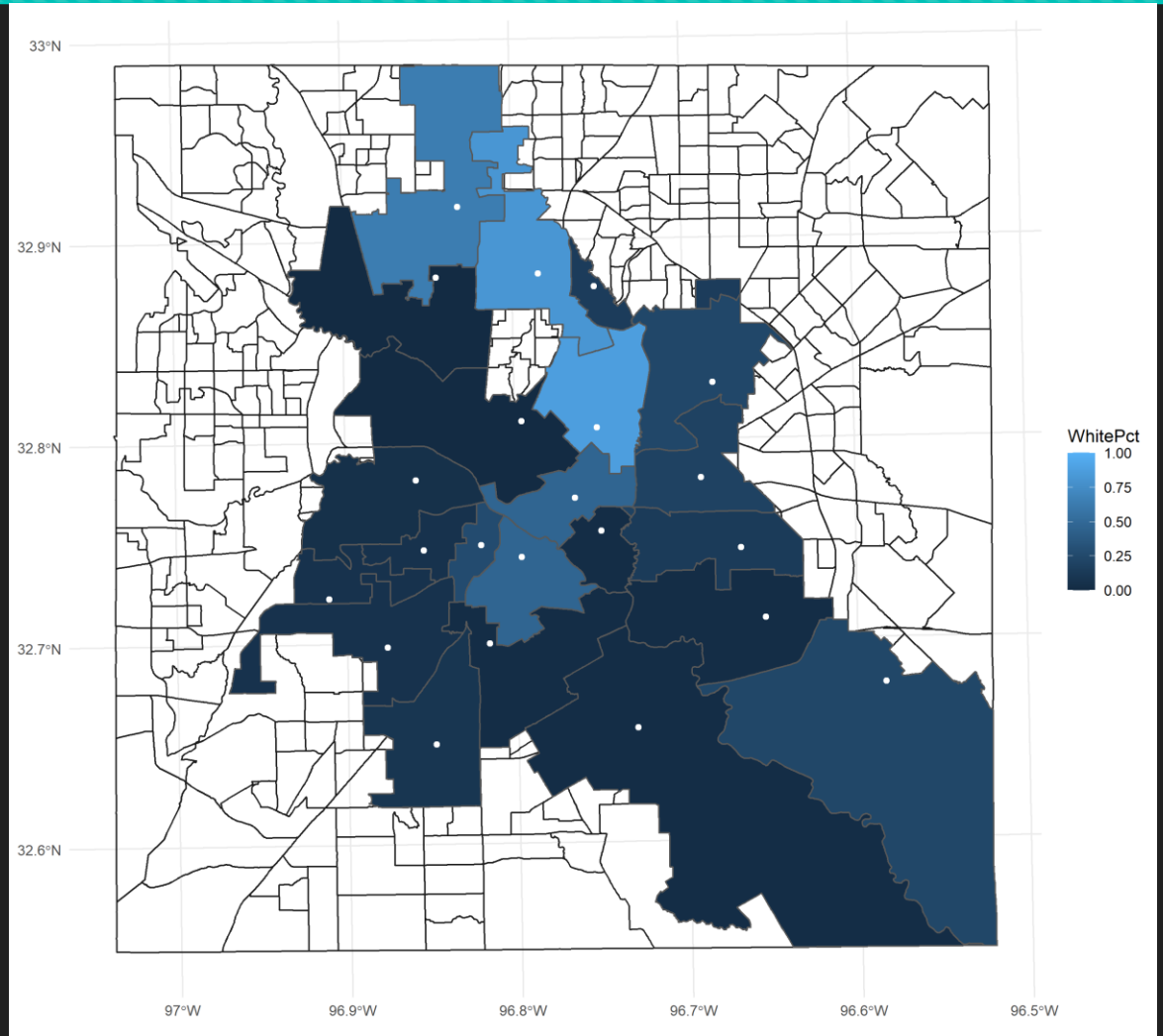
White Population of School Attendance Zones: Middle Schools

```
ggplot() +
  geom_sf(
    fill = NA,
    color = "grey10",
    data = tractsDal,
    show.legend = FALSE
  ) +
  geom_sf(
    mapping = aes(
      fill = WhitePct),
    data = zones_tractsMid,
    show.legend = TRUE
  ) +
  geom_sf(
    color = "white",
    data = campusesMid
  ) +
  theme_minimal()
```



White Population of School Attendance Zones: High Schools

```
ggplot() +
  geom_sf(
    fill = NA,
    color = "grey10",
    data = tractsDal,
    show.legend = FALSE
  ) +
  geom_sf(
    mapping = aes(
      fill = WhitePct),
    data = zones_tractsHigh,
    show.legend = TRUE
  ) +
  geom_sf(
    color = "white",
    data = campusesHigh
  ) +
  theme_minimal()
```



Index of Dissimilarity

- Measures the levels of residential segregation of racial and ethnic groups across urban areas
- 0 indicates proportional distribution of groups within parcel; 1 indicates that no groups share a parcel
- Formula:
  - $D = \frac{1}{2}\sum_{i=1}^{I}\left|\frac{n_{ij}}{Nj} - \frac{n_{ik}}{Nk}\right|,$

    Where n_{ij} is the number of White (non-Hispanic) residents in each attendance zone; N_j is the total number of White residents in the DISD area; n_{ik} is the number of non-White residents in each attendance zone; and N_k is the total number of non-White residents in the DISD area.

```
dissimilarity <- cbind(
  white = zones_schoolsElem$ c_Estimate_Total_SEX_BY_AGE_WHITE_ALONE_NOT_HISPANIC_OR_LATINO_,
  total = zones_schoolsElem$c_Estimate_Total_SEX_BY_AGE_
) %>% as.data.frame()
dissimilarity["nonwhite"] <- dissimilarity["total"] - dissimilarity["white"]
totalwhite <- colSums(dissimilarity["white"])
totalnonwhite <- colSums(dissimilarity["nonwhite"])
D <- 0.5*sum(abs(dissimilarity["white"]/totalwhite - dissimilarity["nonwhite"]/totalnonwhite))
```

Source: *Encyclopedia of Social Measurement* (White & Kim, 2005).

Entropy Index

- Measures the extent of racial segregation across schools
- Values close to 0 indicate that the average diversity of categories is similar to the total diversity level (i.e. less segregation)
  - $H = \frac{E^* - E}{E^*}$,
  - $E^* = (-1)\sum_{k=1}^{K} P_k(\ln(P_k))$,
  - $E = (-1)\sum_{i=1}^{I} \frac{n_i}{N} \sum_{k=1}^{K} P_{ik}(\ln(P_{ik}))$,

    Where P_k is the proportion of White (non-Hispanic) residents in the population; P_{ik} is the proportion of White residents in each attendance zone; n_i is the total population in each attendance zone; and N is the total population.

```
P_k <- totalwhite/(totalwhite+totalnonwhite)
P_ik <- dissimilarity["white"]/dissimilarity["total"] %>% as.vector
N <- totalwhite + totalnonwhite
n_i <- dissimilarity["total"]
E_1 <- (-1)*sum(P_k*log(P_k))
logPik <- P_ik*log(P_ik); logPik[is.na(logPik)] = 0
E_2 <- (-1)*sum(n_i/N*logPik)
H.Elem <- (E_1-E_2)/E_1
```

Source: *Encyclopedia of Social Measurement* (White & Kim, 2005).

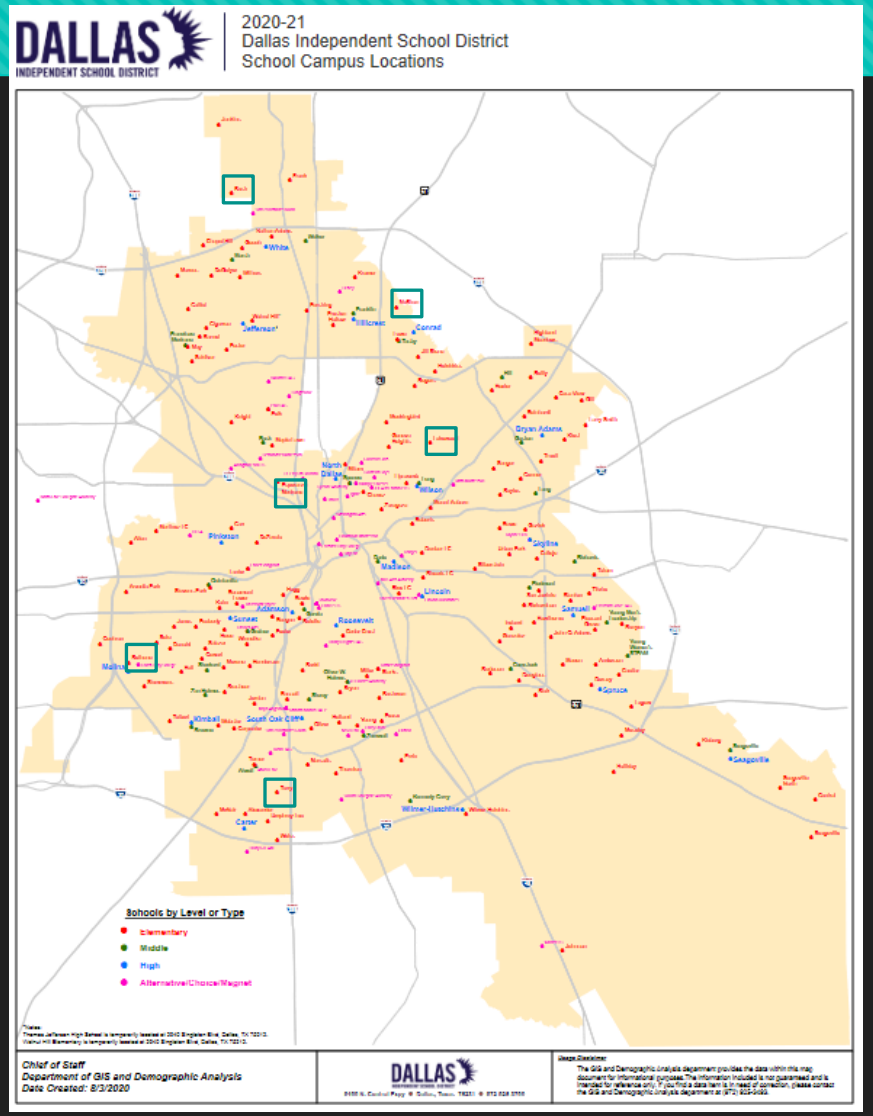
Index Results

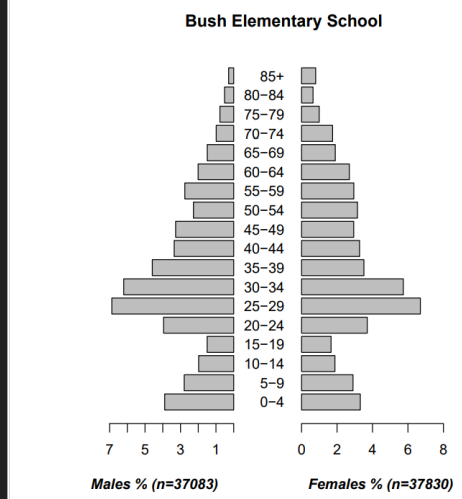
	Index of Dissimilarity	Entropy Index
Dallas County Census Tracts	0.45	0.24
DISD Elementary Schools Attendance Zones	0.47	0.24
DISD Middle Schools Attendance Zones	0.46	0.20
DISD High Schools Attendance Zones	0.45	0.19

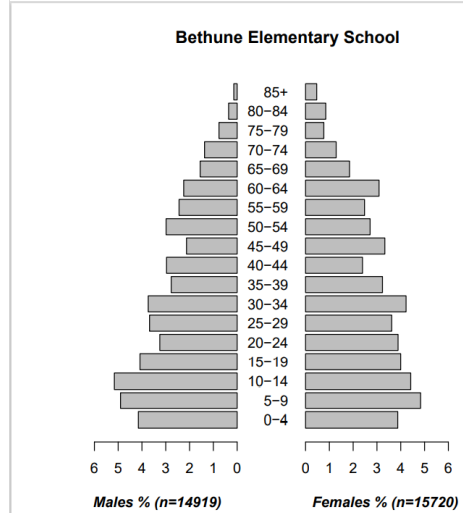
Population Pyramids

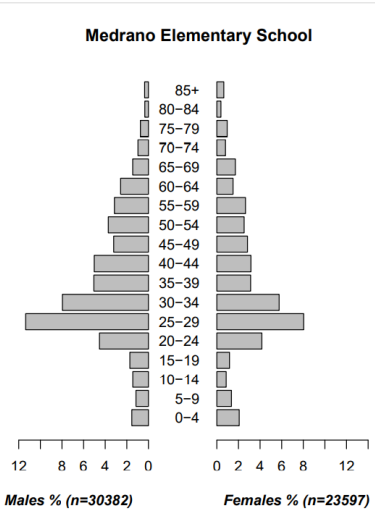
```
library(epiDisplay)
zonesList <- c("Total Dallas ISD Zone", zones_schoolsElem[,2])
zones <- zones_schoolsElem[,-c(1:7,375)] %>% as.matrix()
row.names(zones) <- zonesList
zonesTotal <- colSums(zones)
zones <- rbind(zonesTotal, zones)
pdf(file="DISDElementary_Pyramids.pdf", width = 6, height = 6, onefile = TRUE, version="1.2")
for(j in 1:nrow(zones)){
  temp <- cbind(zones[j,3:25],zones[j,27:49])
  temp[5,] <- temp[4,]+temp[5,]
  temp[8,] <- temp[6,]+temp[7,]+temp[8,]
  temp[17,] <- temp[16,]+temp[17,]
  temp[19,] <- temp[18,]+temp[19,]
  temp <- temp[-c(4,6,7,16,18),]
  malesum <- sum(temp[ , 1])
  femsum <-  sum(temp[ , 2])
  colnames(temp) <- c(paste("Males % (n=",malesum,")",sep=""), paste("Females % (n=",femsum,")",sep=""))
  row.names(temp) <- c("0-4","5-9","10-14","15-19","20-24","25-29","30-34","35-39","40-44","45-49","50-54","55-
    59","60-64","65-69","70-74","75-79","80-84","85+")
  pyramid(inputTable=temp[,1:2], percent="total", font.lab=4, main=zonesList[j])
}
dev.off()
```

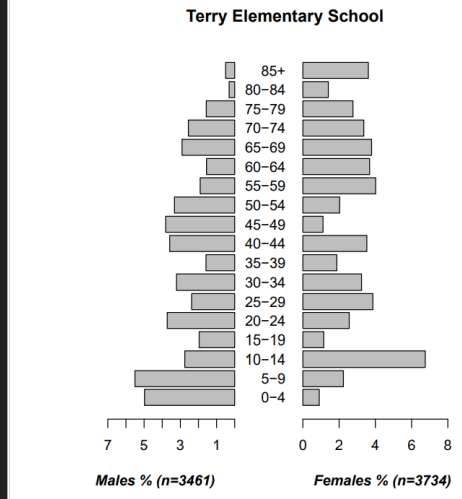
Population Pyramids

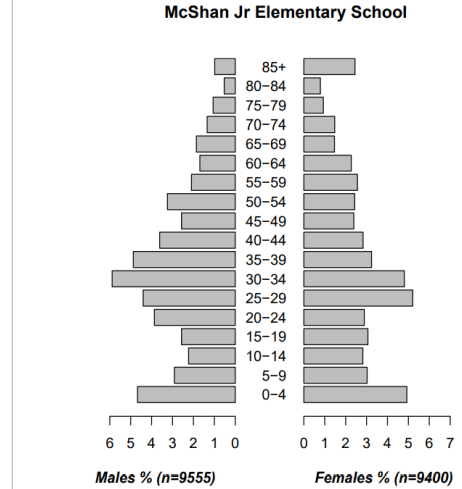


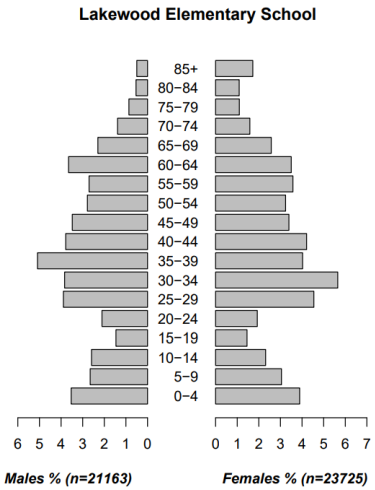




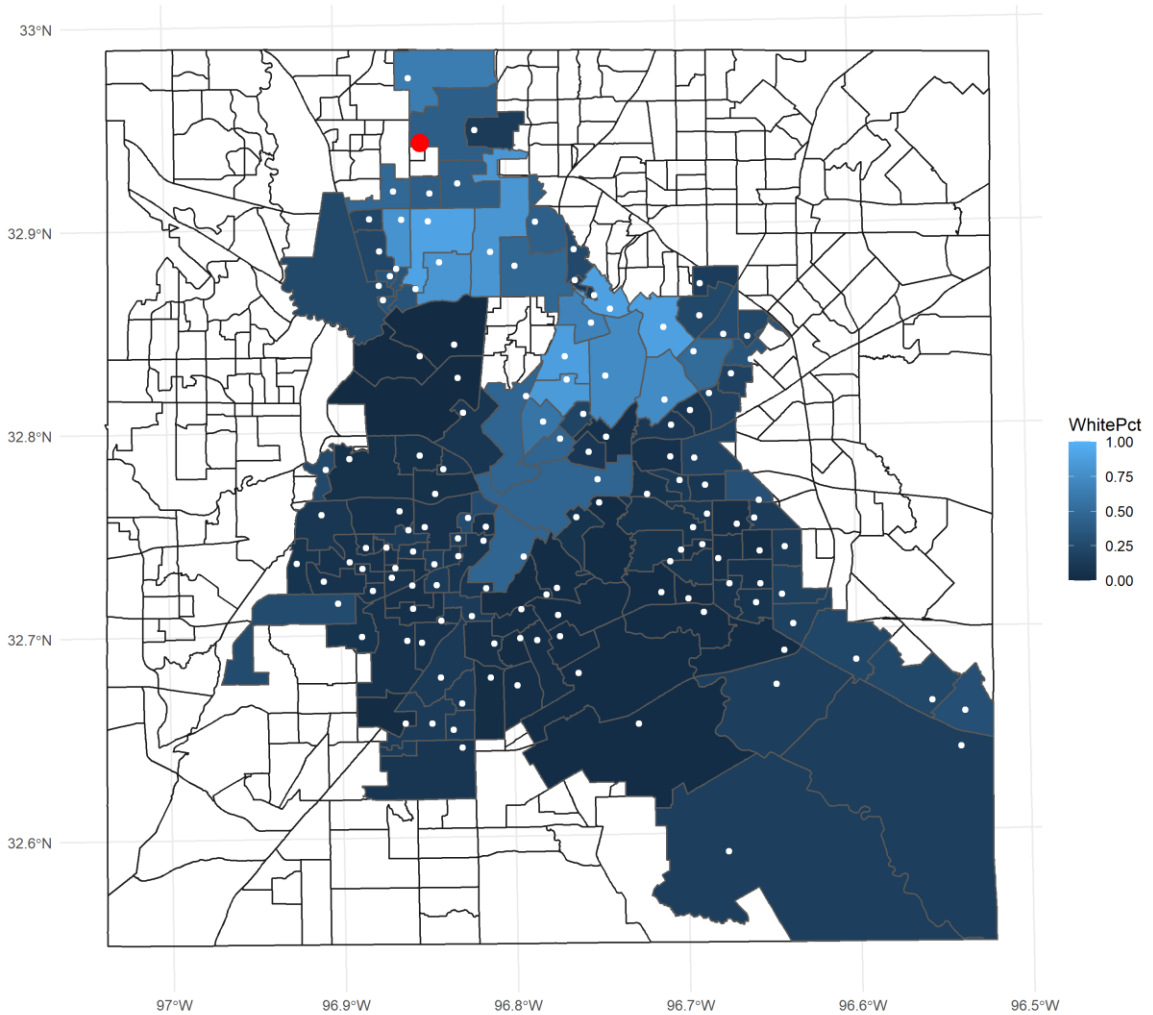


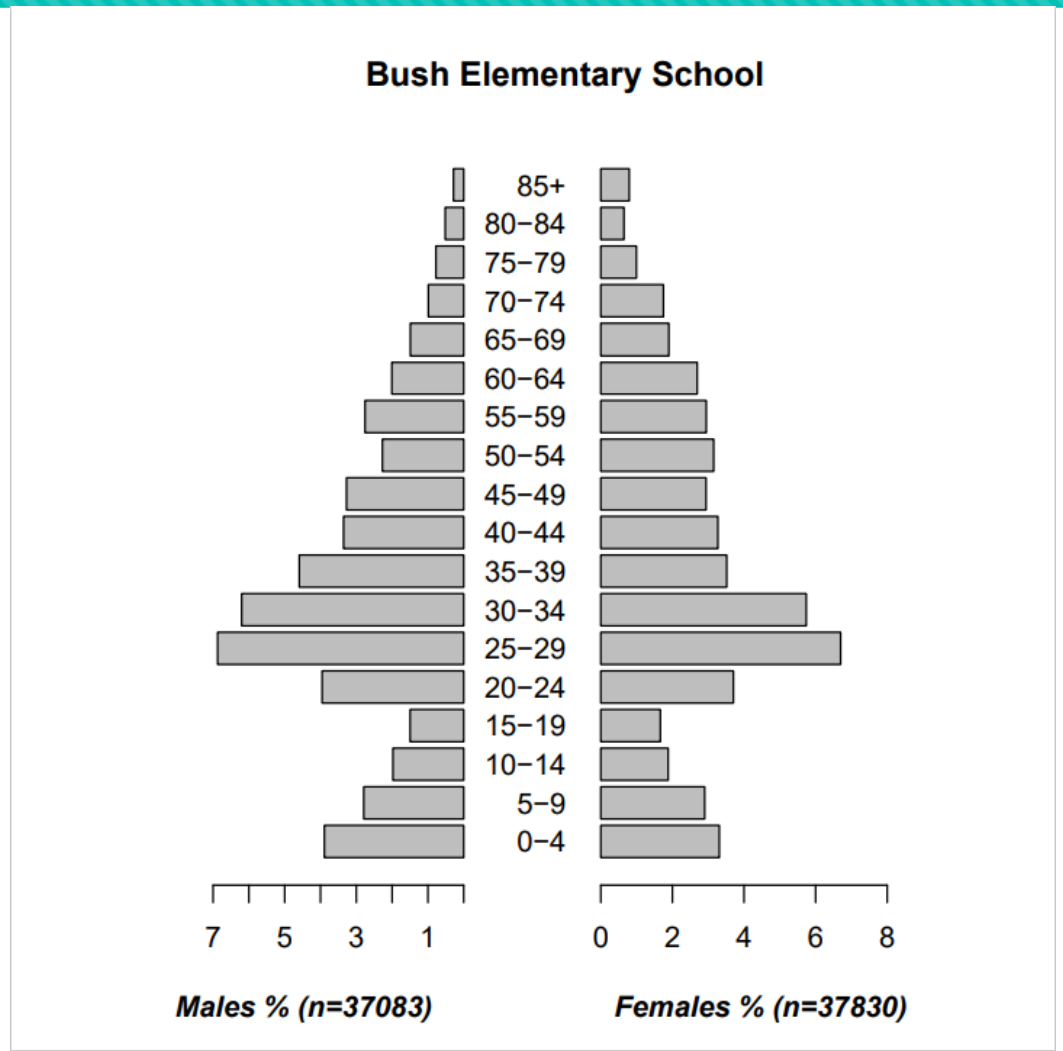




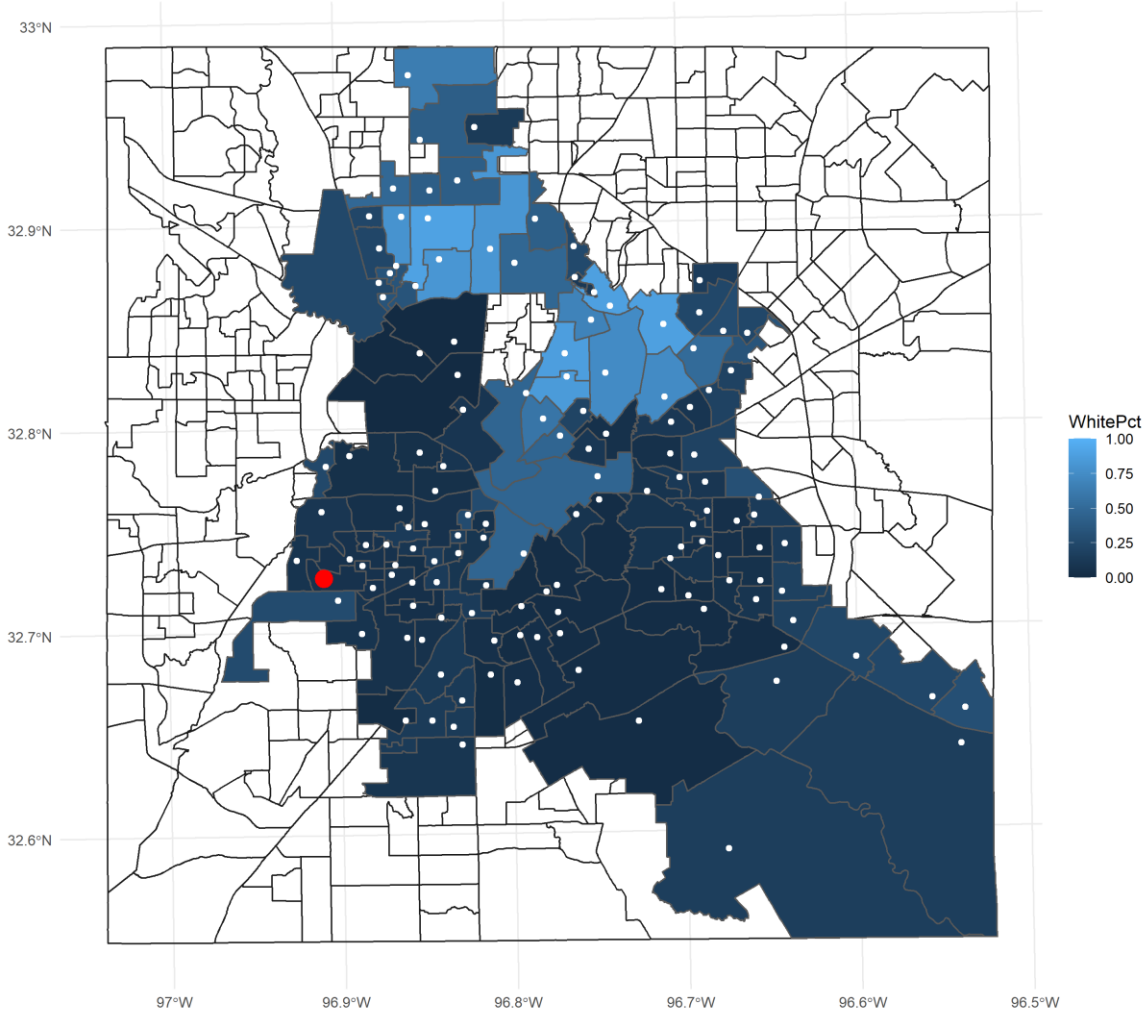


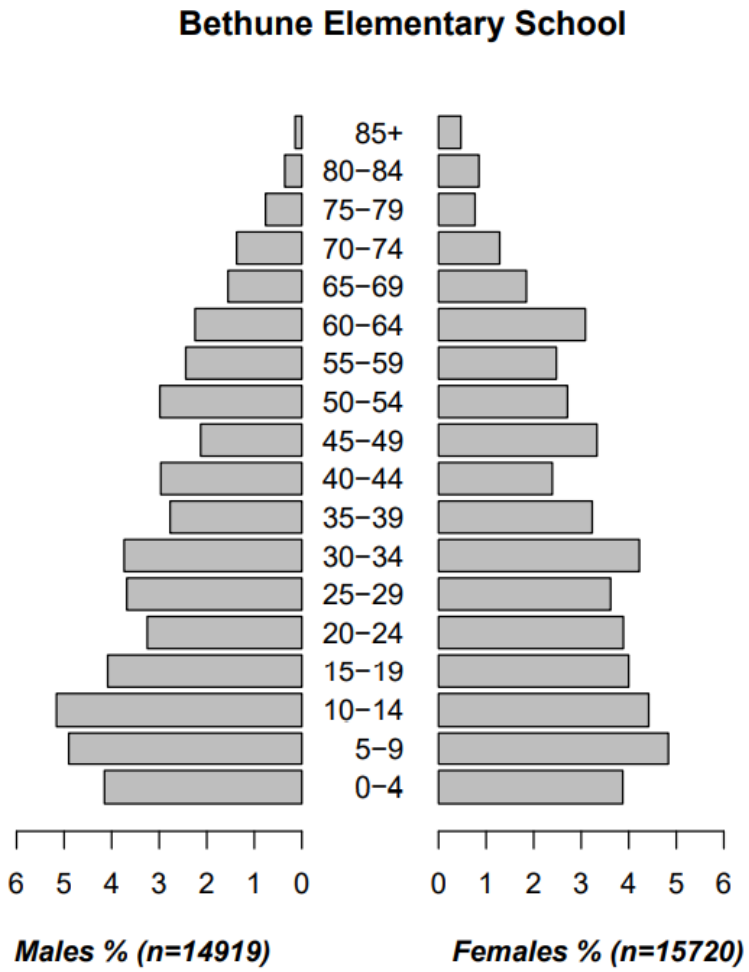
Population Pyramids



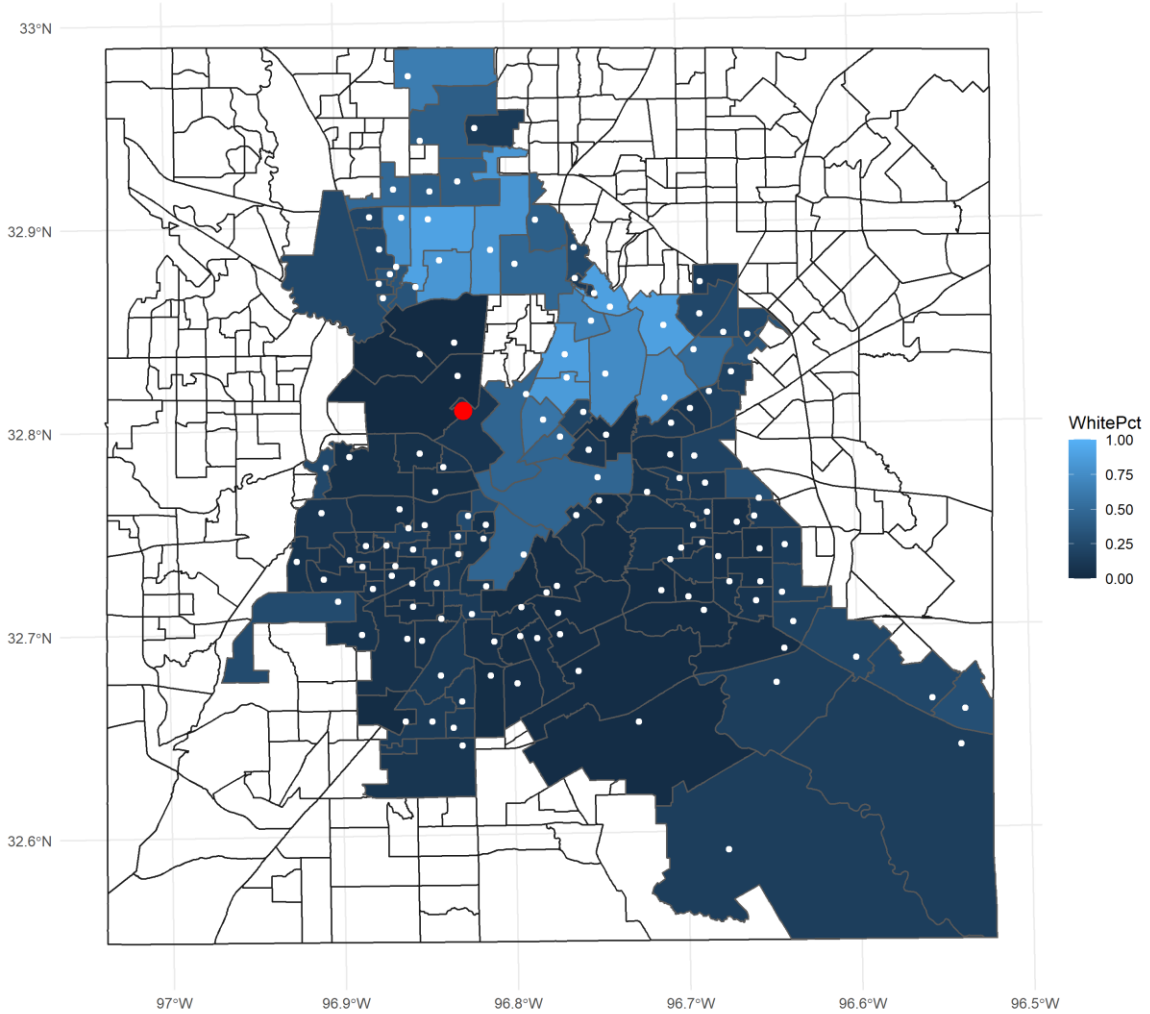


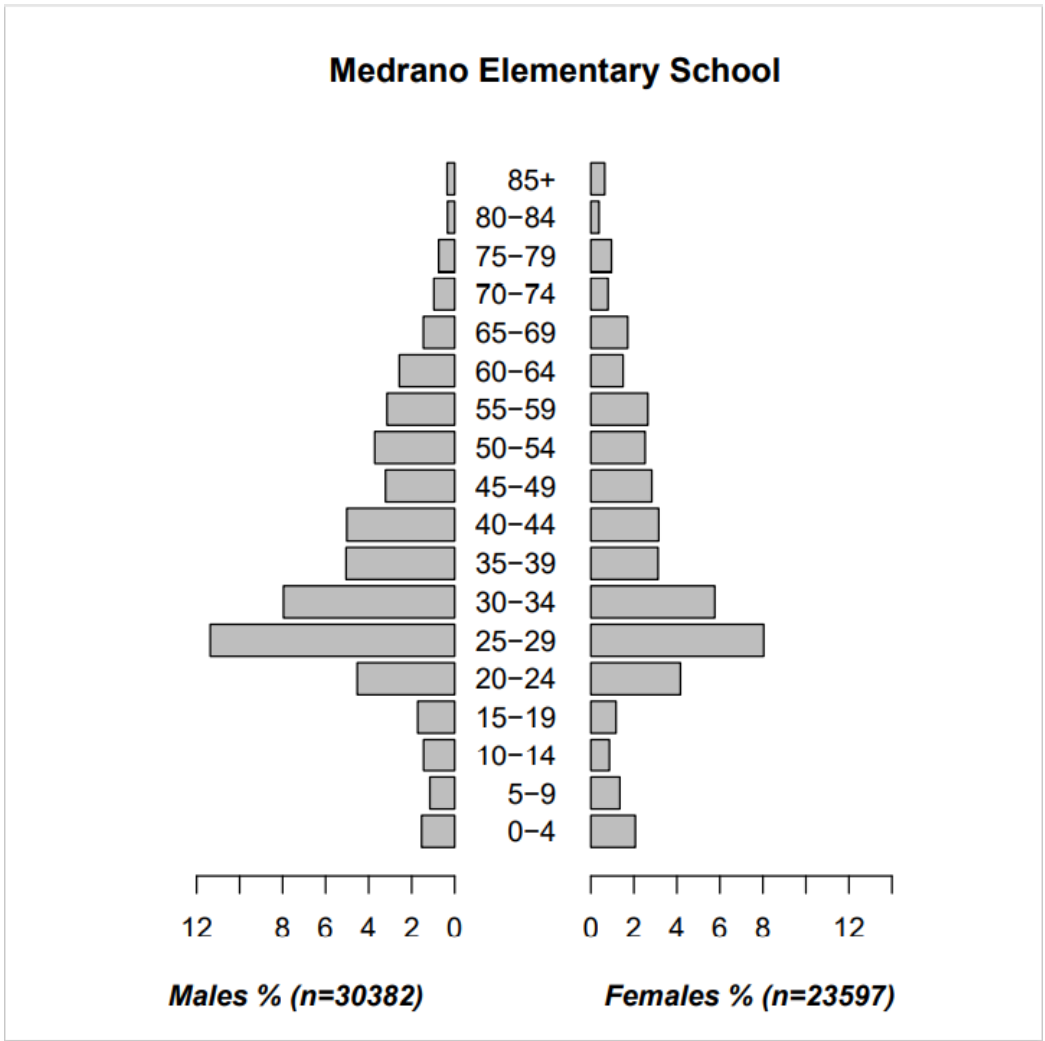
Population Pyramids



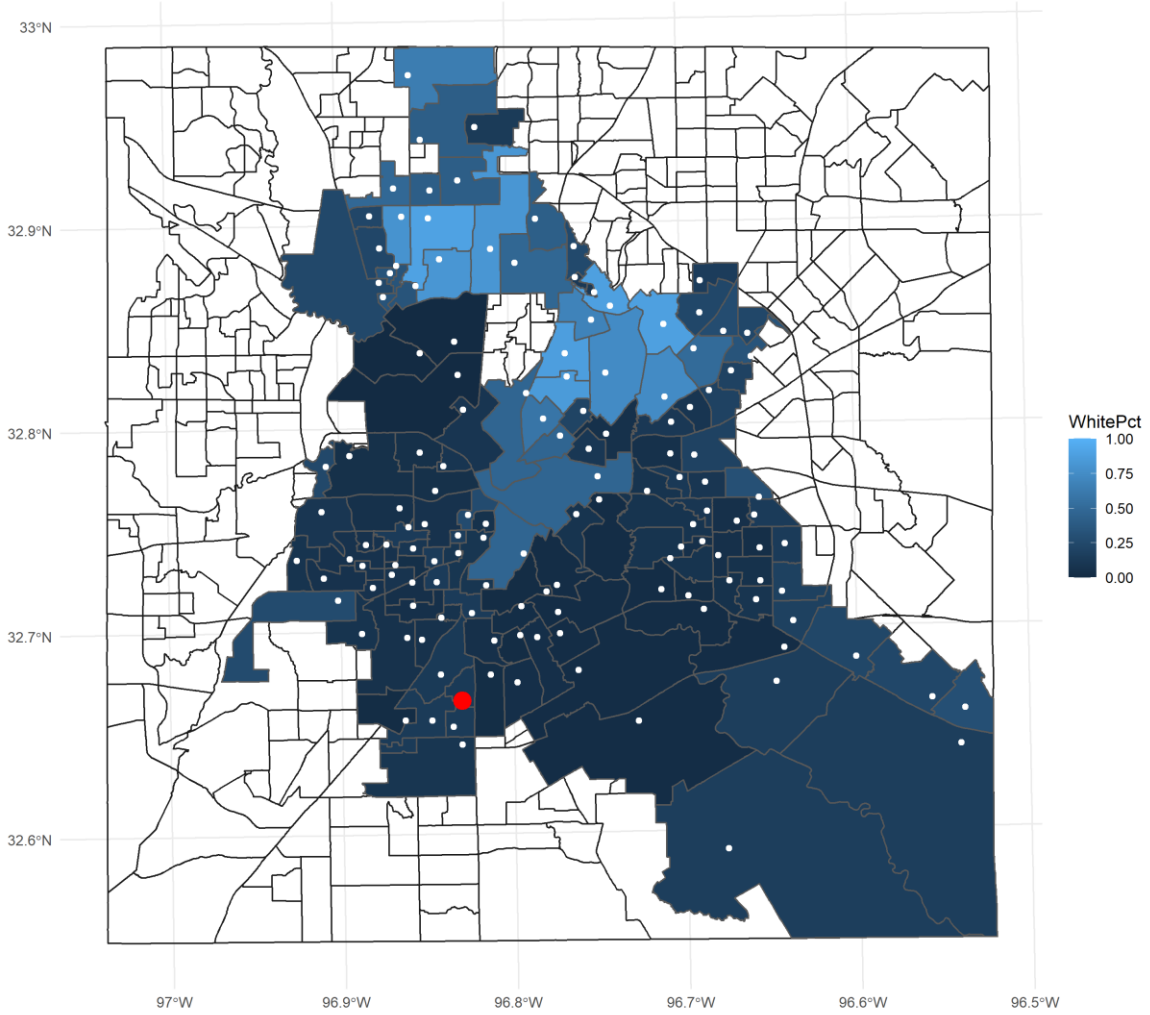


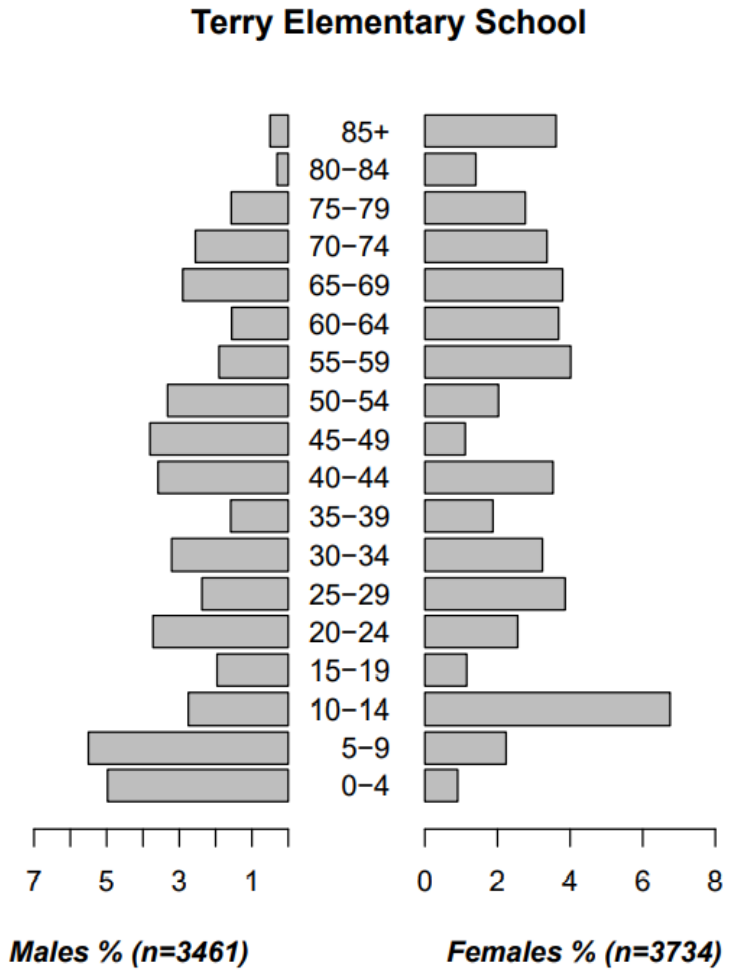
Population Pyramids



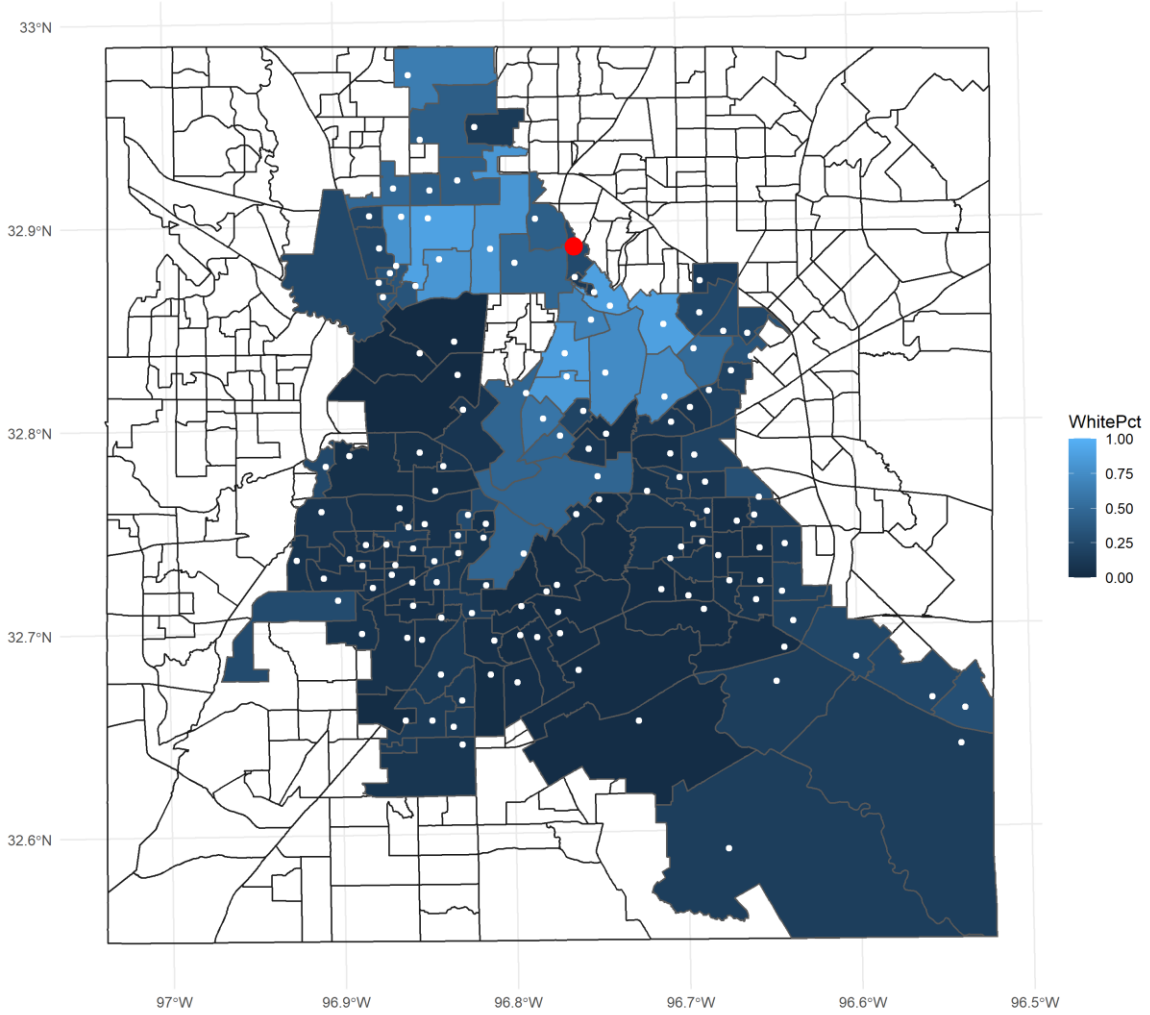


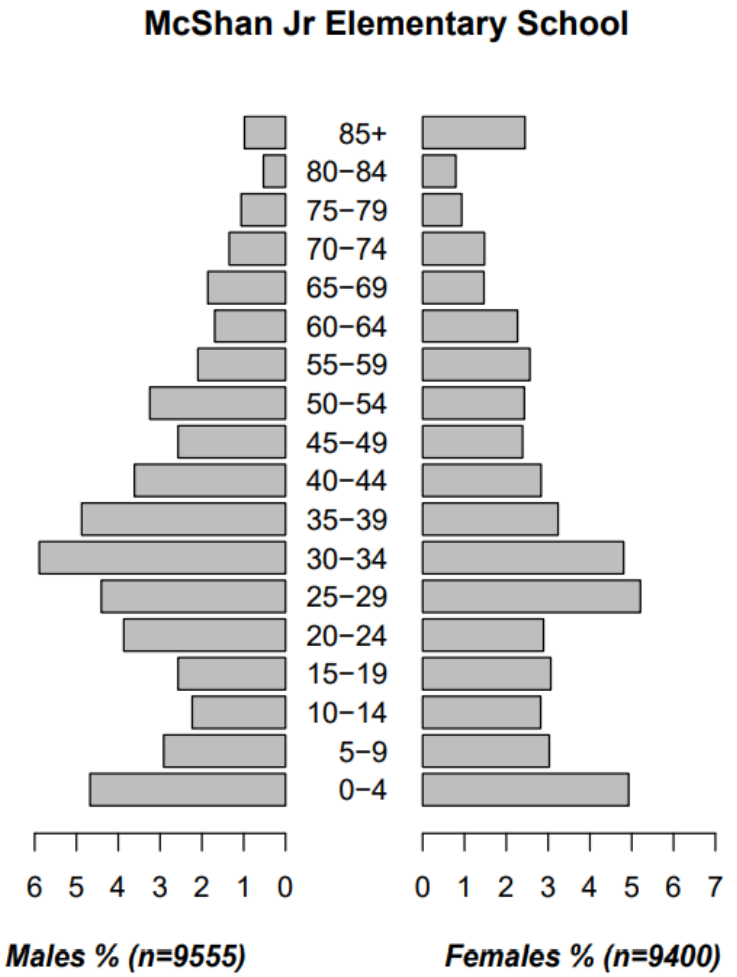
Population Pyramids



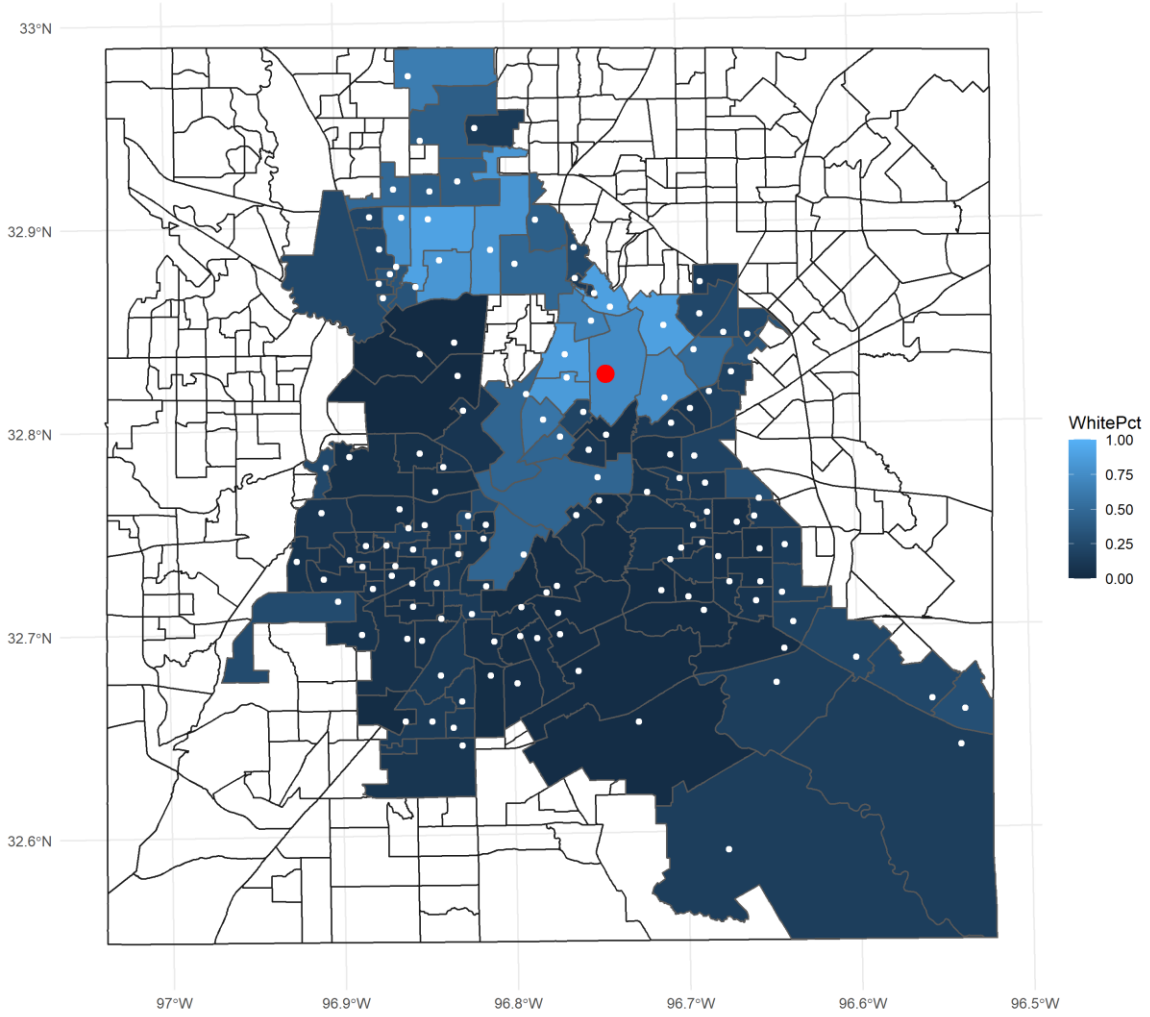


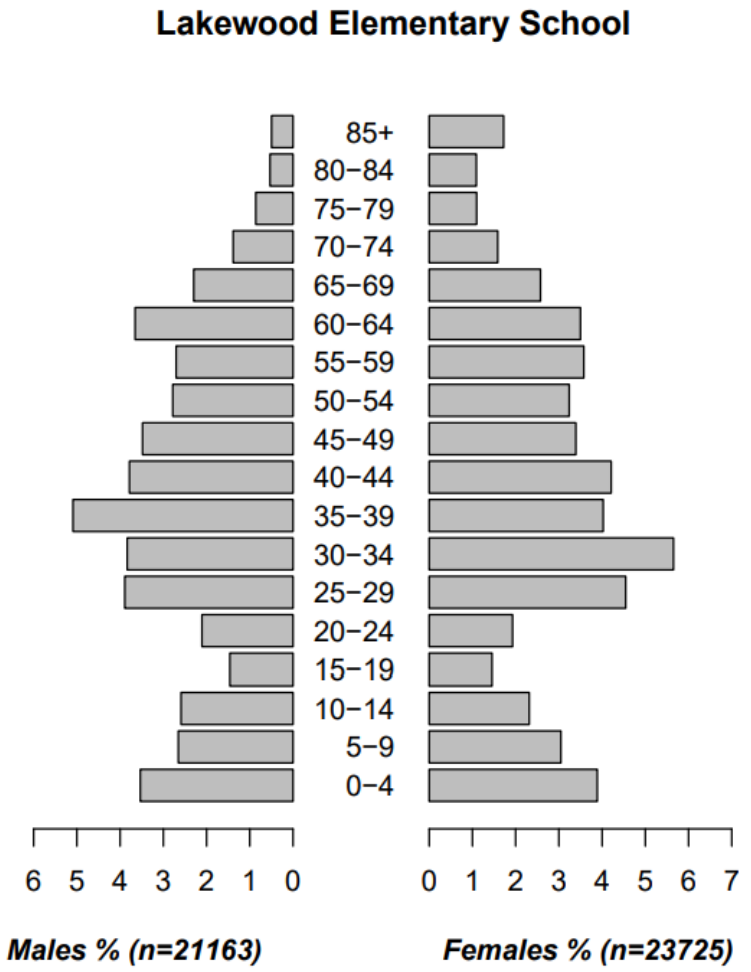
Population Pyramids





Population Pyramids





Conclusion

- There are significant demographic differences between school attendance zones in Dallas ISD
- High school attendance zones are less segregated than elementary schools (re: White (non-Hispanic) vs. non-White population)

Dallas ISD School Attendance Zone Demographic Analysis

GISC 7364 Final Project

Rachel Brasier

11 May 2021