



908647 Group Project Report

Joseph Moore, Jimmy Barnett, Maximiliano Arguindegui



Abstract

Our presentation will discuss the changes in demographics over the last 5 years in the four largest public universities in Texas: Texas A&M, Texas Tech, Texas, and Texas State. We created histograms, as well as a spatial weighted mean center graph to show how demographics have changed. The histograms are specific to each school, while the spatial weighted mean center takes into account every school. We found in our histograms that each school in Texas has white students as their largest demographic, however the percentage of white students is dropping across the board. We also found that hispanic student populations are rising more than any other. In our spatial weighted mean graph, we found, for example, that international students more likely to attend Texas Tech, while Asian students are more likely to attend Texas.



Intro

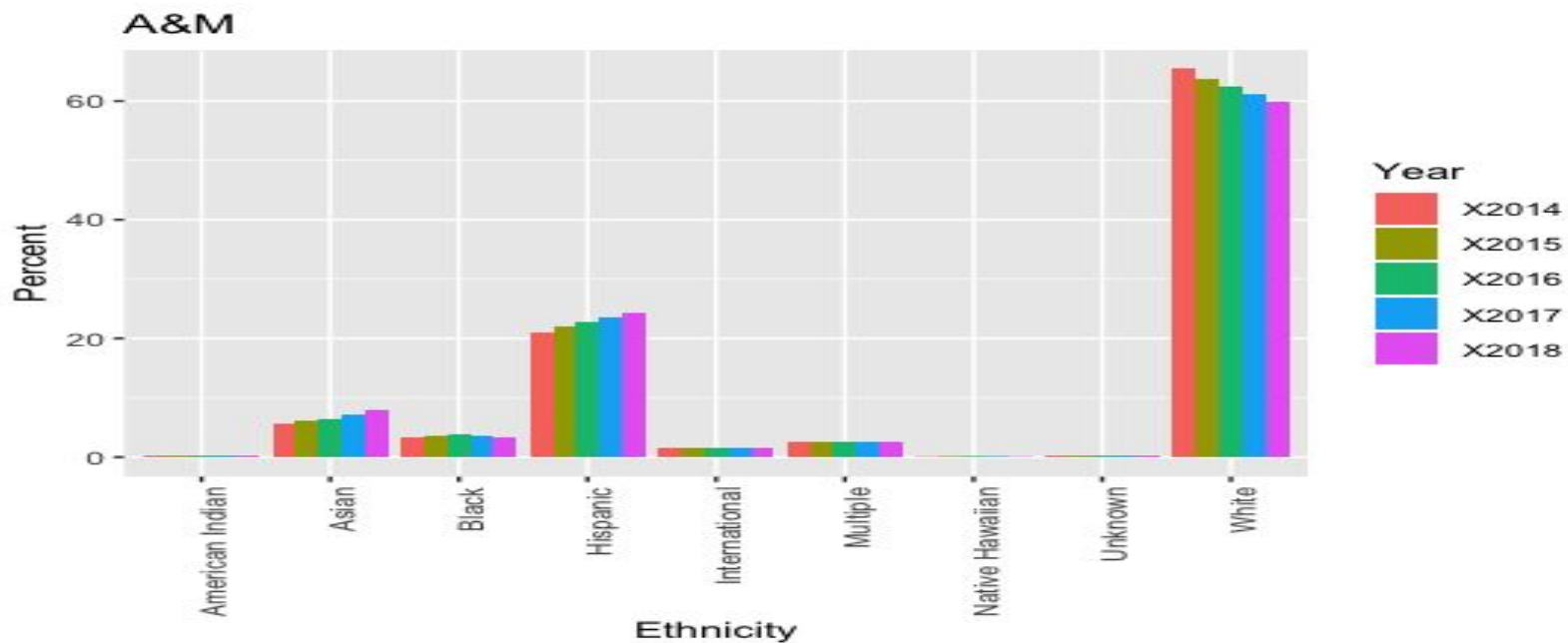
Our project will be comparing the demographics of the four largest public universities in Texas over the last 5 years. This idea comes from a study called “Texas Demographics Trends, Characteristics, and Projections” done by the Department of Savings and Mortgage Lending. We will be using histograms and weighted mean center to determine how the demographics have changed. We found our data on the websites of Texas A&M University, University of Texas, Texas Tech University, and Texas State University. Our Null hypothesis is that schools have become more diverse over time, specifically higher Hispanic populations and UT Austin being more diverse than the other schools..



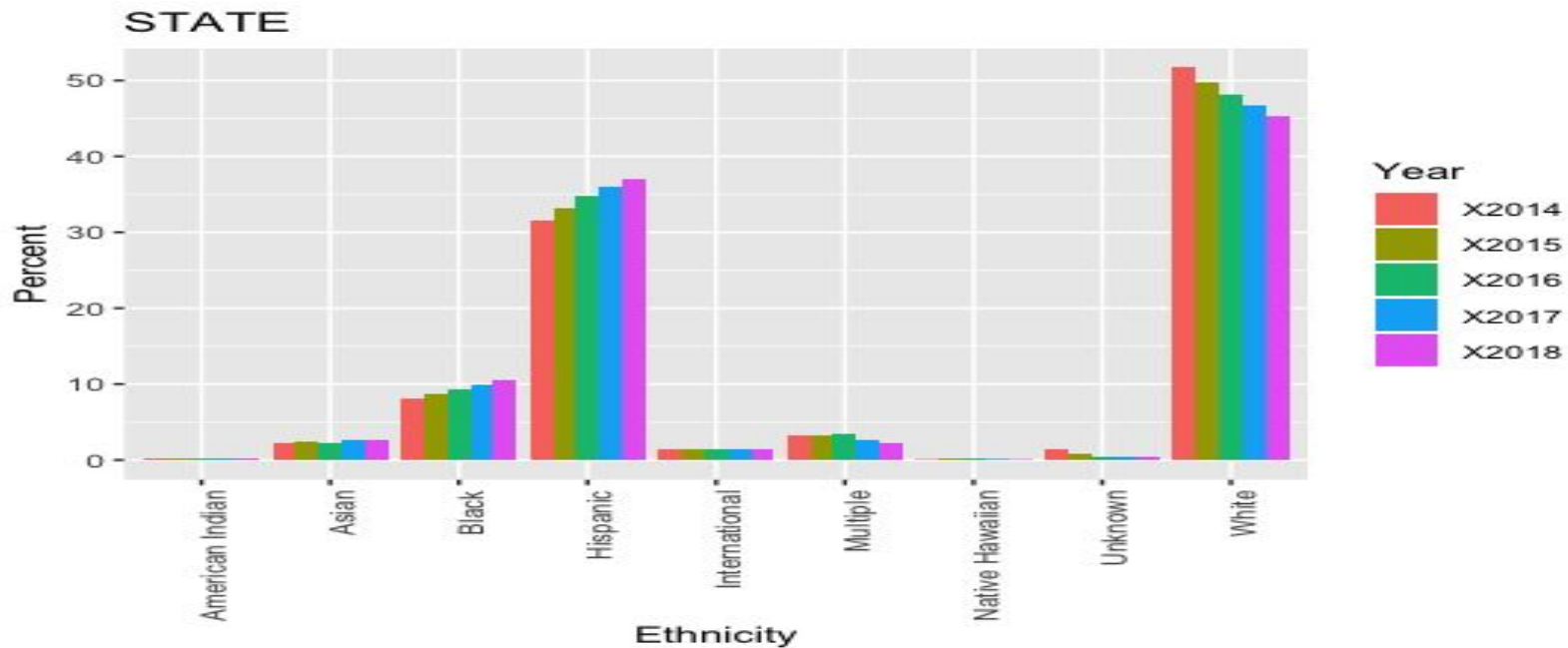
Methodology

We used our data taken from various University demographics trackers in a csv format so we could better use it in R. We took the data and compared the change in percentages of each ethnicity over time in our histograms. It took another package in R, called ggplot, and the end result was much better than could be done without, because we wanted to display the changes per ethnicity by year on a single plot per school for clarity and this allowed us to do so. We thought this help better show change in makeup by university. In addition, we used a mean center, and weighted mean center in our spatial analysis of the data. We initially plotted the school's coordinates on an xy coordinate grid, and then used simple averages of coordinates to find the mean center of the data. After this, we used the weighted mean center formula to determine where the weighted mean centers are across Texas, using different colors to represent each ethnicities' WMC. This can be used to display which ethnicities have larger populations at which schools, and therefore help us draw conclusions about population distributions across Texas as a whole.

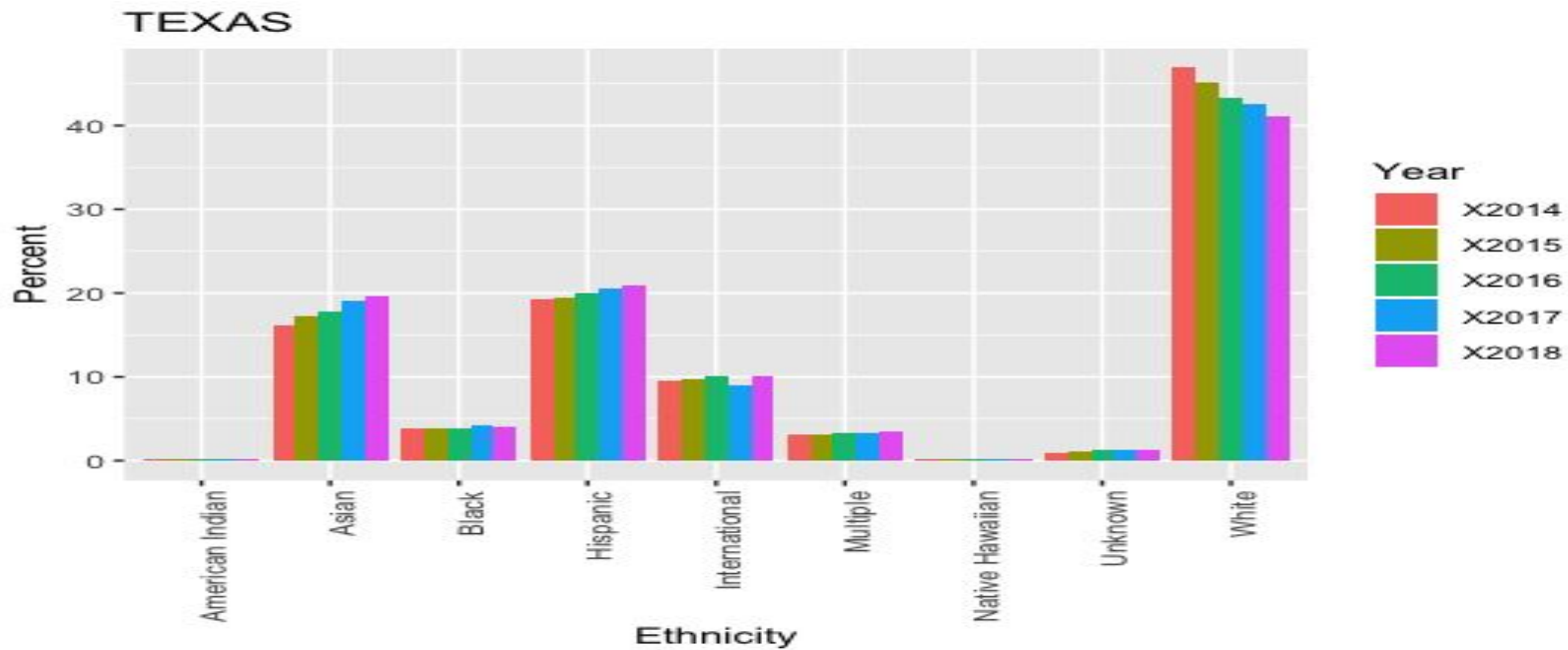
Texas A&M



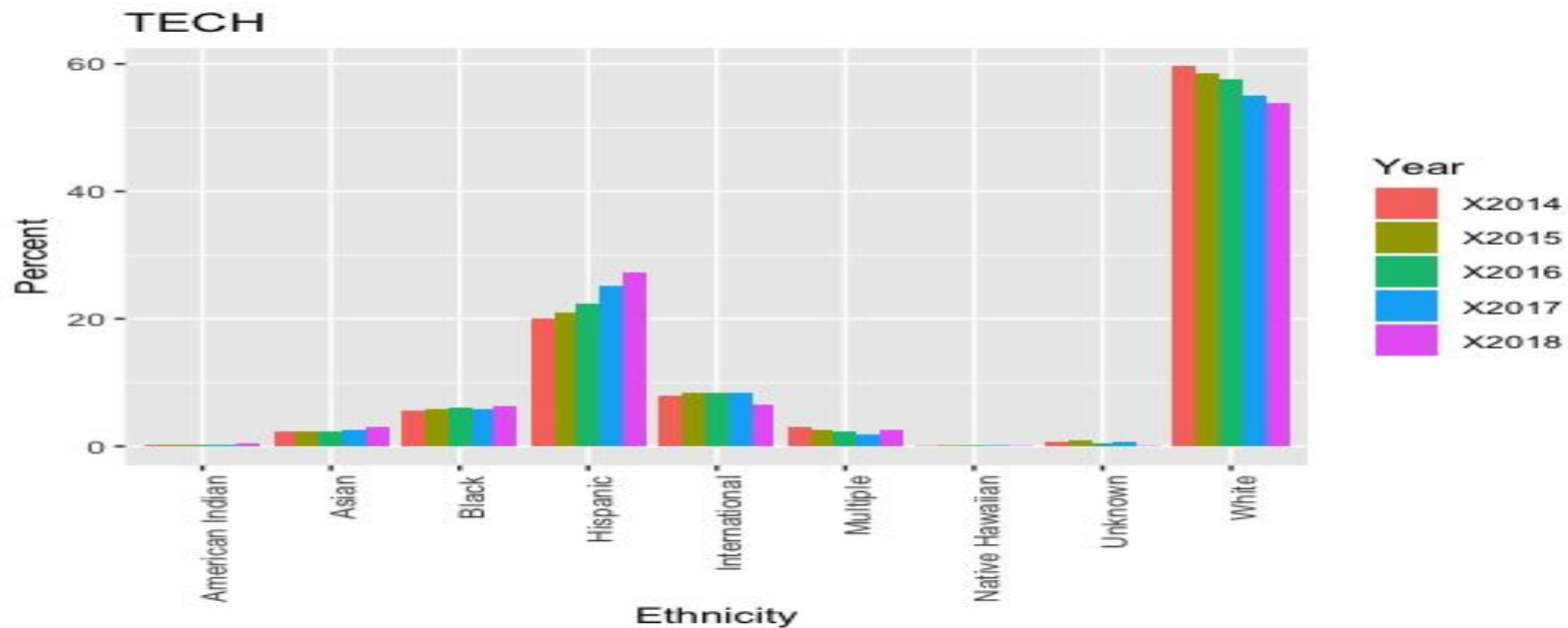
Texas State



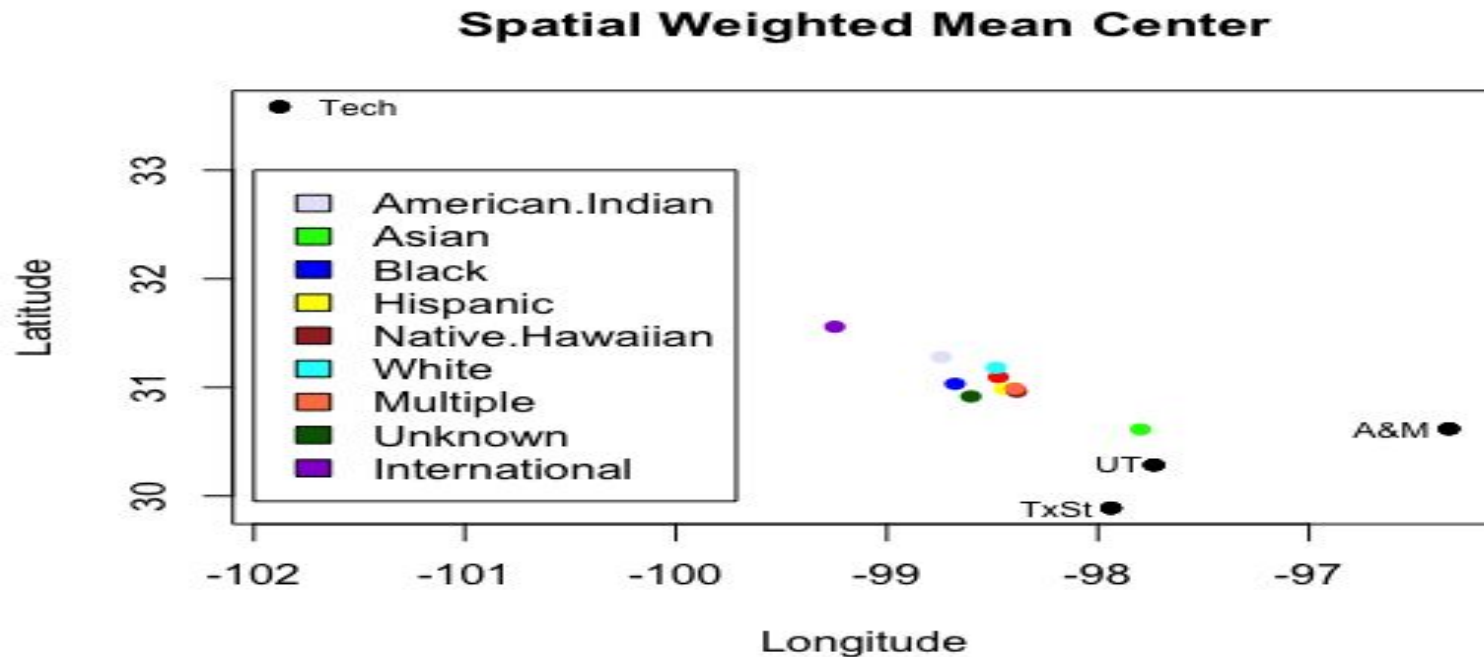
Texas



Texas Tech



Spatial Weighted Mean Center





Results pt. 1

For our first analysis, four histograms were made displaying the percent of each race present in the four schools observed. The x axis of the bar chart indicates the race while the y axis of the chart shows the percent of each race attending each given school. There are five different colored bars displayed with each given race, each color represents the year that the data corresponds with. Right away, it is easily seen that white students have the largest percent for each school from 2014-2018. On the other hand, Native Hawaiian and American Indian students have the least percent of students attending.



Results pt.2

The spatial statistic conducted was the spatial weighted mean center of each race in relation to each of the schools observed. The weighted mean center is weighted based upon sample values. The x axis displays longitude and the y axis displays latitude. The plot has four black dots labeled with the four schools observed at the coordinates of each school, and the colored dots represent the weighted mean center of the 9 observed races at each school. The most evident example would be that the weighted mean center indicates that there are more Asian students present near the University of Texas. This does not mean there are not Asian students attending Texas Tech or any other schools, but the weighted mean center is dependent on sample values. There are clearly more Asian students in the UT Austin area. Additionally, it leads one to believe there is more of a percentage of international students at Tech comparatively.



Conclusion pt. 1

In regards to the four histograms displaying the percent of each race attending each school, it was evident that white students were the largest group across all four schools observed from 2014-2018. One interesting observation is that Hispanic students have the second largest percent across all schools. However, the University of Texas has a significantly large percent of Asian students and has nearly as many Asian students as Hispanic students for the 5 years observed. The University of Texas also had the most international students from 2014-2018. Texas State had the least amount of Asians, and Texas A&M had the least amount of Black students.



Conlcusion pt. 2

The plot showing the weighted mean center gave some very insightful data. One important observation worth noting is the large distance between the Texas Tech point and the rest of the colored points. A possible explanation for this is the large amount of racially diverse students in the UT and Texas State. Both of these schools are relatively close to each other, have a large number of students, and are the most racially diverse. The nature of the spatial statistic conducted will cause the weighted mean centers of each race to be pulled near the UT and Texas State area. Texas A&M also has a diverse student body, so this is another reason that the colored points are further away from Texas Tech.

R code Histogram

```
# Semester Project
library(ggplot2)
#specify directory containing csv tables:
fdir =
"/Users/trejimmy5562/Desktop/GEOG312_2/Project/"
#list all of the csv files in the dir:
fnames = list.files(fdir, pattern = ".csv")
#paste name on the end of the path:
fpaths = paste0(fdir, fnames)
#make table from the csv file:
tab = read.csv(fpaths, header = T)
#separate data into yearly chunks for plotting
attach(tab)
"brown", "cyan", "coral", "darkgreen",
"darkviolet"), breaks=9)
color2 = c("red", "green", "blue", "yellow",
"brown", "cyan", "coral", "darkgreen",
"darkviolet")
#CHANGE THIS TO CHANGE SCHOOLV
fourteen_am = data.frame(tab[28:36,1:7])
```

```
#THIS IS HOW YOU STACK DATA IN NEW TABLE
data = data.frame()
year = 2014
for(i in 1:4){
  data1 = data.frame(fourteen_am$Ethnicity,
rep(paste0("X", toString(year)), 9),
fourteen_am[i+1])
  data2 = data.frame(fourteen_am$Ethnicity,
rep(paste0("X", toString(year+1)), 9),
fourteen_am[i+2])
  data1 <- setNames(data1,
c("Ethnicity", "Year", "Percent"))
  data2 <- setNames(data2,
c("Ethnicity", "Year", "Percent"))
  data <- rbind(data, data1, data2)
  #Fourteen Fifteen Graph
  year <- year+1
}
ggplot(data, aes(fill=Year, y=Percent,
x=Ethnicity)) +
  geom_bar(position="dodge", stat="identity")+
  theme(axis.text.x = element_text(angle = 90,
hjust = 1)) + ggtitle("STATE")
```

R Code Spatial

```
# Semester Project
#####
#
#library(ggplot2)
#specify directory containing csv tables:
fdir = "/Users/trejimmy5562/Desktop/GEOG312_2/Project/plots/"
#list all of the csv files in the dir:
fnames = list.files(fdir, pattern = ".csv")
#paste name on the end of the path:
fpaths = paste0(fdir, fnames)
#make table from the csv file:
tab = read.csv(fpaths, header = T)
#separate data into yearly chunks for plotting
#attach(tab)
#specify directory containing csv tables:
fdir = "/Users/trejimmy5562/Desktop/GEOG312_2/Project/new/"
#list all of the csv files in the dir:
fnames = list.files(fdir, pattern = ".csv")
#paste name on the end of the path:
fpaths = paste0(fdir, fnames)
#make table from the csv file:
tab2 = read.csv(fpaths, header = T)
#separate data into yearly chunks for plotting
attach(tab2)
#set columns equal to variables
x = tab$lon
y = tab$lat
#plot it for mean center/std. dist
#plot(x,y, pch = 3)
#spatial centers
xc = mean(x)
yc = mean(y)
```

```
#points(xc,yc, pch=16)
#standard distance
#st1 = sqrt(sum((x - xc)^2) + sum((y - yc)^2))
#plot on top
#par(new=T)
#plot the stand distance as a circle:
#symbols(xc, yc, circles = st1, xlim = range(x), ylim =
range(y), xlab = "", ylab = "", main = fnames)
ff = names(tab2)
f1 = ff[4:12]
lat_weightedmean = vector()
lon_weightedmean = vector()
colr = c("lavender", "green", "blue", "yellow", "brown",
"cyan", "coral", "darkgreen", "darkviolet")
plot(tab2$Lon, tab2$Lat, cex=1, pch=16, main = "Spatial
Weighted Mean Center", xlab = "Longitude", ylab = "Latitude")
am_lon = tab2$Lon[1]
texas_lon = tab2$Lon[6]
tech_lon = tab2$Lon[11]
state_lon = tab2$Lon[16]
am_lat = tab2$Lat[1]
texas_lat = tab2$Lat[6]
tech_lat = tab2$Lat[11]
state_lat = tab2$Lat[16]
text(am_lon,am_lat,labels="A&M", adj = 1.25, cex = .75)
text(texas_lon,texas_lat,labels="UT", adj = 1.25, cex = .75)
text(tech_lon,tech_lat,labels="Tech", adj = -.5, cex = .75)
text(state_lon,state_lat,labels="TxSt", adj = 1.25, cex = .75)
points(xc, yc, col="red", pch=16, cex=1)
for(j in 4:12){
  lat_weightedmean[j] = sum(tab2$Lat*tab2[,j])/sum(tab2[,j])
  lon_weightedmean[j] = sum(tab2$Lon*tab2[,j])/sum(tab2[,j])
  points(lon_weightedmean[j], lat_weightedmean[j],
col=colr[j-3], pch=16, cex=1)
}
legend(x=-102, y=33, legend = f1, fill = colr)
```



References

https://demographics.texas.gov/Resources/Presentations/OSD/2017/2017_09_14_DepartmentofSavingsandMortgageLending.pdf

<https://datausa.io/profile/university/texas-tech-university/>

<http://techdata.irs.ttu.edu/Factbook/Enrollment/ENRETHCLASS.aspx>

<https://accountability.tamu.edu/All-Metrics/Mixed-Metrics/Student-Demographics>

<https://www.ir.txstate.edu/reports-projects/highlights/highlights-demographics>

<https://datausa.io/profile/university/texas-state-university/>

<https://datausa.io/profile/university/the-university-of-texas-at-austin/>

<https://www.utexas.edu/about/facts-and-figures>