

Final Project Milestone 1: Proposal

Adam Aleksic

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1
```

```
## v ggplot2 3.3.2    v purrr  0.3.4
## v tibble  3.0.3    v dplyr  1.0.2
## v tidyr   1.1.2    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflic
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
```

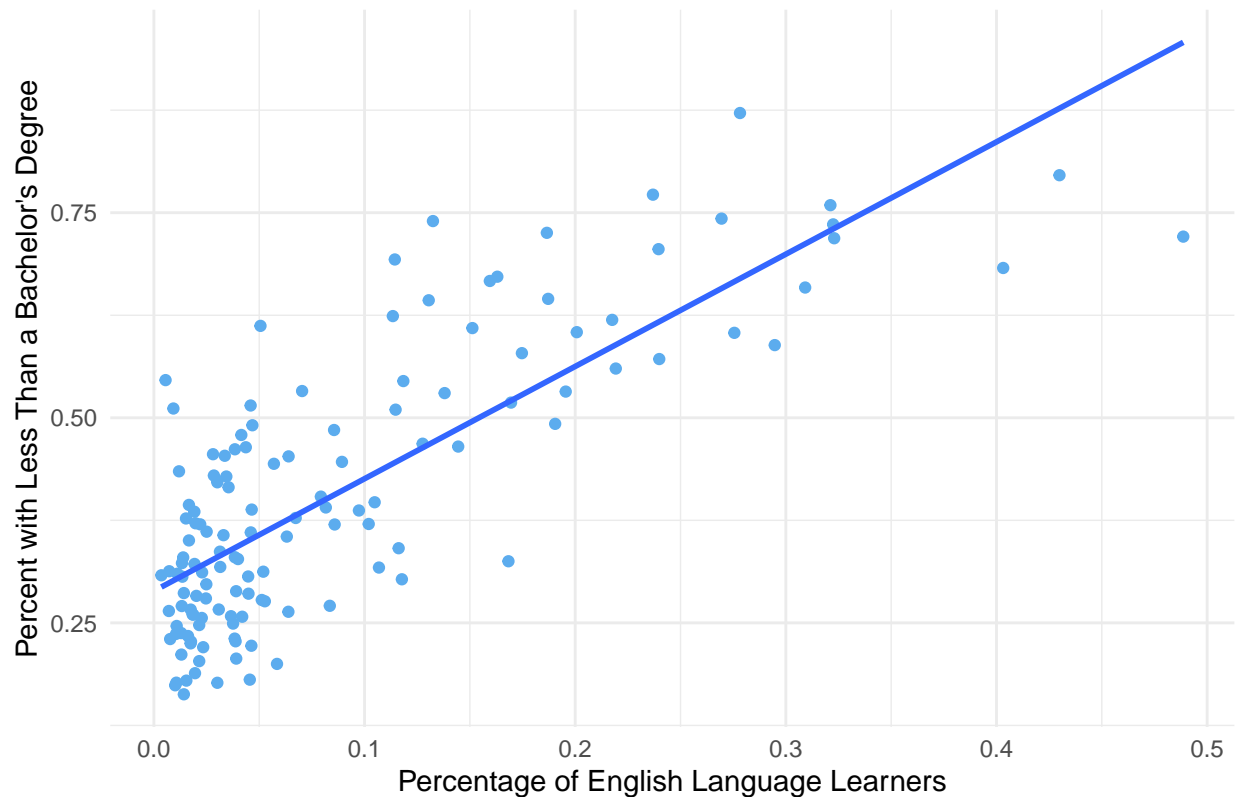
```
seattle <- read.csv("~/Downloads/Gov/Racial_and_Social_Equity_Composite_Index.csv") %>%
  na.omit()
```

```
# setting up my data
```

```
education <- seattle %>%
  ggplot(aes(x = PCT_ENGLISH_LESSTHAN_VERY_WELL, y = PCT_LESS_BACHELOR_DEGREE)) +
  geom_point(color = "steelblue2") +
  labs(title = "Lack of English Proficiency versus Lack of Education",
       x = "Percentage of English Language Learners",
       y = "Percent with Less Than a Bachelor's Degree") +
  theme_minimal() +
  geom_smooth(method = lm, se = FALSE)
education
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

Lack of English Proficiency versus Lack of Education



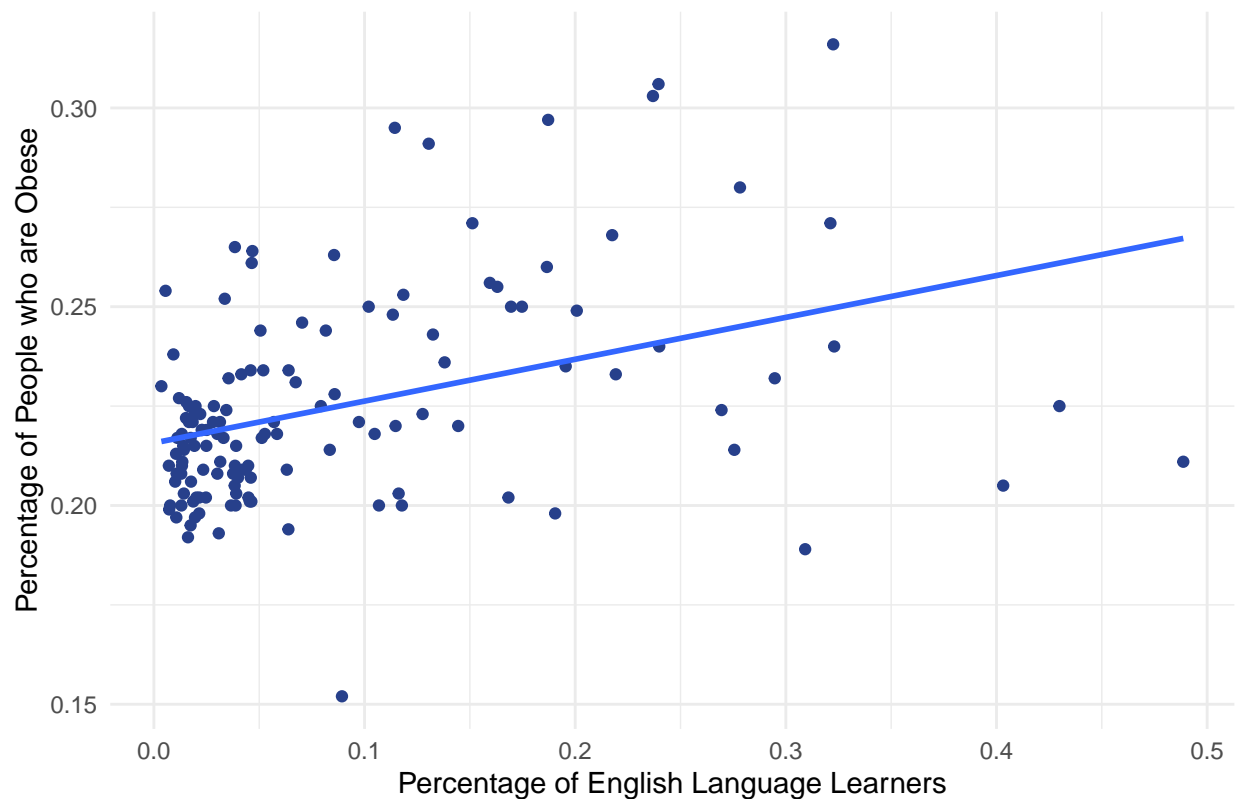
```
education_fit <- lm(PCT_ENGLISH_LESSTHAN_VERY_WELL ~ PCT_LESS_BACHELOR_DEGREE, data = seattle)
education_fit_sum <- summary(education_fit)
education_fit_sum$r.squared
```

```
## [1] 0.6236412
```

```
obese <- seattle %>%
  ggplot(aes(x = PCT_ENGLISH_LESSTHAN_VERY_WELL, y = PCT_ADULT_OBESE)) +
  geom_point(color = "royalblue4") +
  labs(title = "Lack of English Proficiency versus Obesity",
       x = "Percentage of English Language Learners",
       y = "Percentage of People who are Obese") +
  theme_minimal() +
  geom_smooth(method = lm, se = FALSE)
obese
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

Lack of English Proficiency versus Obesity



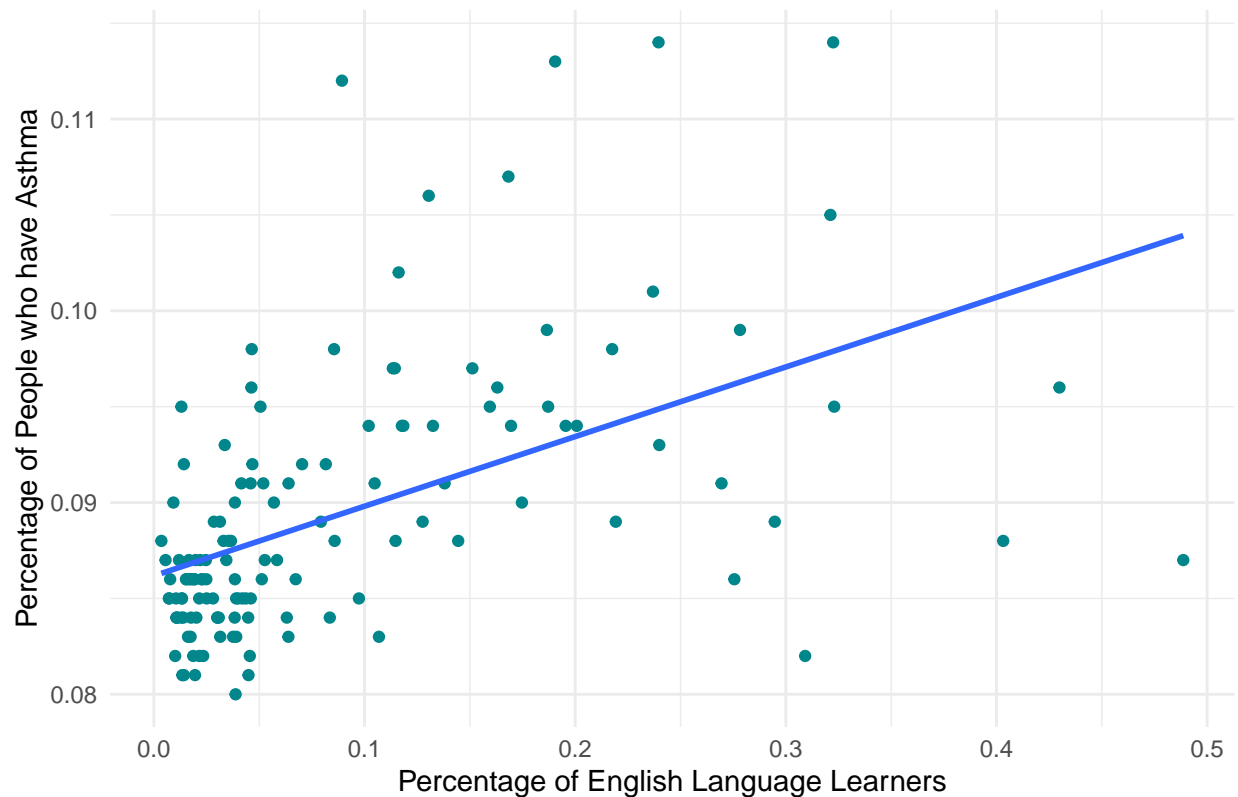
```
obese_fit <- lm(PCT_ADULT_OBESE ~ PCT_LESS_BACHELOR_DEGREE, data = seattle)
obese_fit_sum <- summary(obese_fit)
obese_fit_sum$r.squared
```

```
## [1] 0.4831654
```

```
asthma <- seattle %>%
  ggplot(aes(x = PCT_ENGLISH_LESSTHAN_VERY_WELL, y = PCT_ADULT_WITH_ASTHMA)) +
  geom_point(color = "turquoise4") +
  labs(title = "Lack of English Proficiency versus Asthma",
       x = "Percentage of English Language Learners",
       y = "Percentage of People who have Asthma") +
  theme_minimal() +
  geom_smooth(method = lm, se = FALSE)
asthma
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

Lack of English Proficiency versus Asthma



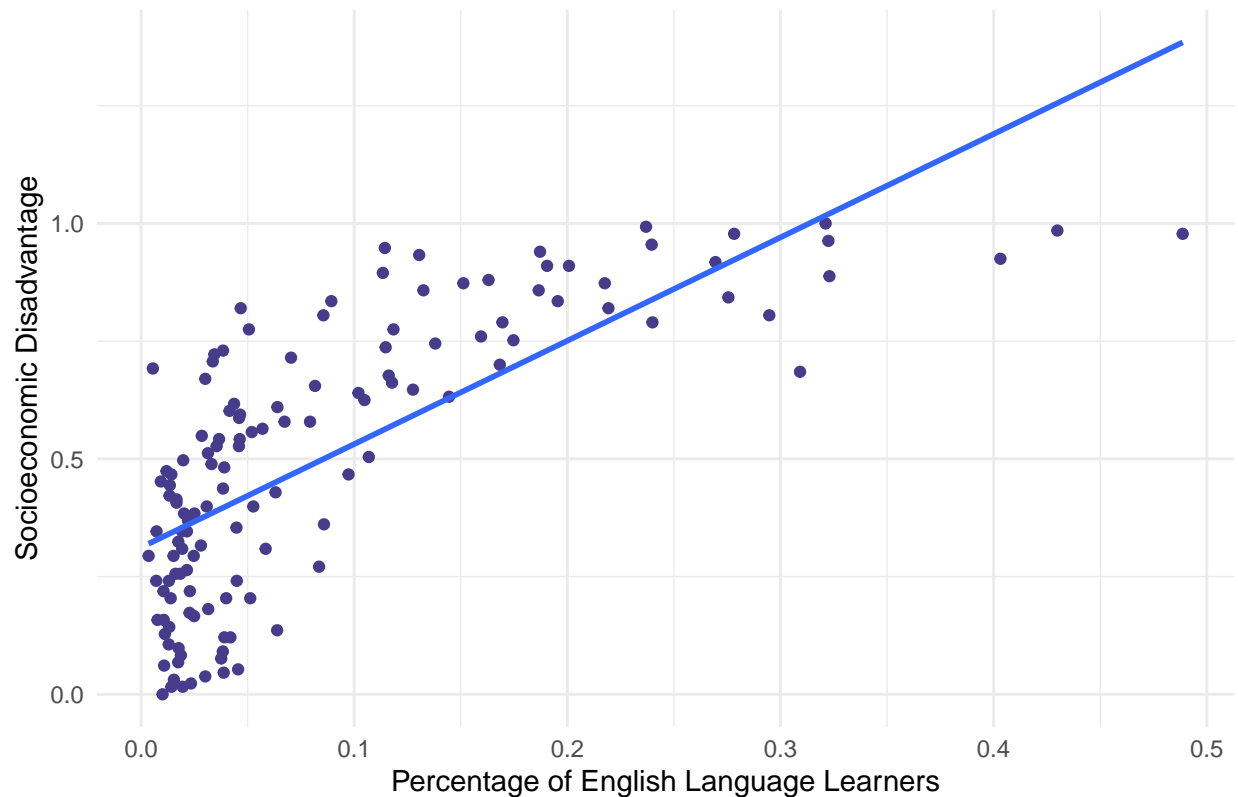
```
asthma_fit <- lm(PCT_ADULT_WITH_ASTHMA ~ PCT_LESS_BACHELOR_DEGREE, data = seattle)
asthma_fit_sum <- summary(asthma_fit)
asthma_fit_sum$r.squared
```

```
## [1] 0.3863937
```

```
socioeconomic <- seattle %>%
ggplot(aes(x = PCT_ENGLISH_LESSTHAN_VERY_WELL, y = SOCIOECONOMIC_PERCENTILE)) +
  geom_point(color = "slateblue4") +
  labs(title = "Lack of English Proficiency vs Socioeconomic Disadvantage",
       x = "Percentage of English Language Learners",
       y = "Socioeconomic Disadvantage") +
  theme_minimal() +
  geom_smooth(method = lm, se = FALSE)
socioeconomic
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

Lack of English Proficiency vs Socioeconomic Disadvantage



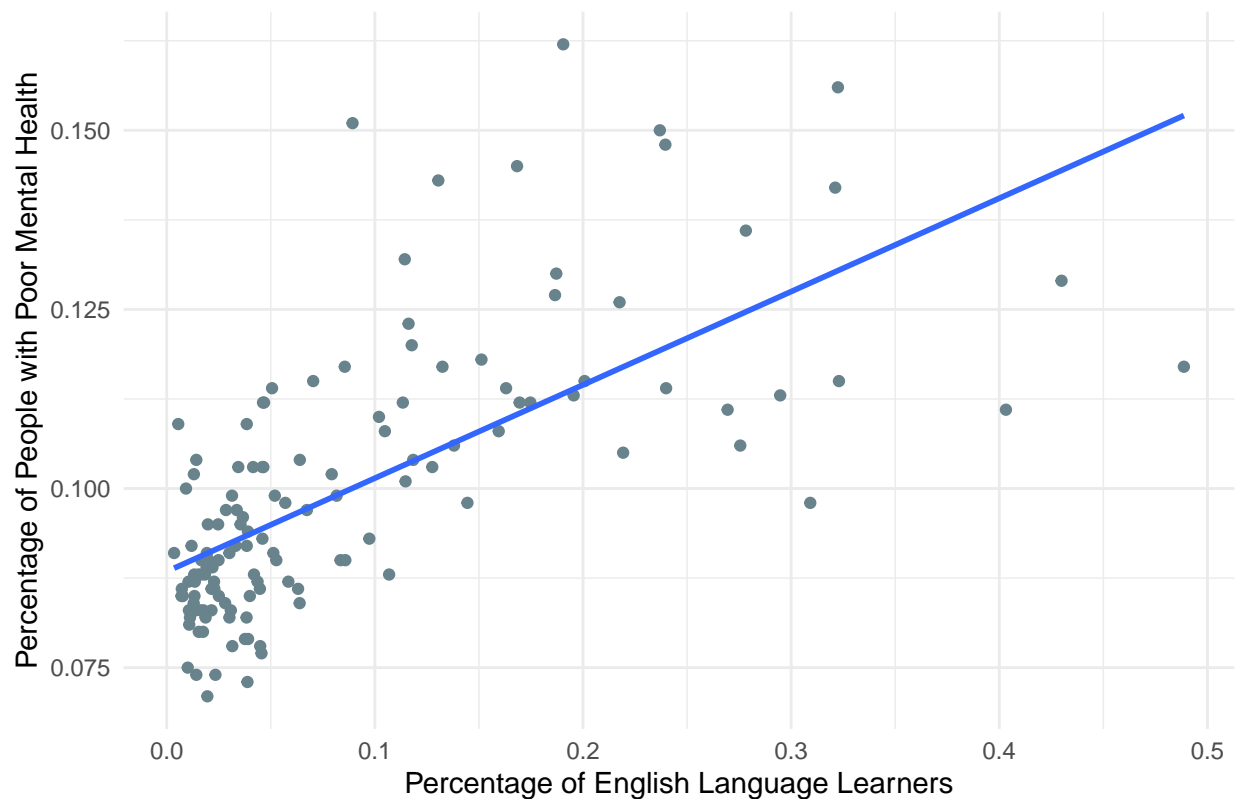
```
socioeconomic_fit <- lm(SOCIOECONOMIC_PERCENTILE ~ PCT_LESS_BACHELOR_DEGREE, data = seattle)
socioeconomic_fit_sum <- summary(socioeconomic_fit)
socioeconomic_fit_sum$r.squared
```

```
## [1] 0.7973289
```

```
mental <- seattle %>%
  ggplot(aes(x = PCT_ENGLISH_LESSTHAN_VERY_WELL, y = PCT_ADULTMENTALHEALTHNOTGOOD)) +
  geom_point(color = "lightblue4") +
  labs(title = "Lack of English Proficiency versus Poor Mental Health",
       x = "Percentage of English Language Learners",
       y = "Percentage of People with Poor Mental Health") +
  theme_minimal() +
  geom_smooth(method = lm, se = FALSE)
mental
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

Lack of English Proficiency versus Poor Mental Health



```
mental_fit <- lm(PCT_ADULTMENTALHEALTHNOTGOOD ~ PCT_LESS_BACHELOR_DEGREE, data = seattle)
mental_fit_sum <- summary(mental_fit)
mental_fit_sum$r.squared
```

```
## [1] 0.583541
```

Stuff from Milestone 1

Data Set

Data set: <https://catalog.data.gov/dataset/racial-and-social-equity-composite-index-a44fc>

Understanding the variables: https://data-seattlecitygis.opendata.arcgis.com/datasets/225a4c2c50e94f2cb548a046217f49f7_0?geometry=-122.509%2C47.574%2C-122.164%2C47.655

This data set examines linguistic, racial, ethnic, income, education, and health statistics for census tracts in Seattle.

I want to examine how the percent of English language learners in census tracts correlates with factors like obesity, poverty, education level, asthma, and mental health.

From what I've seen of it, this data set is excellent because it has a lot of information I can draw from and is "clean" so I don't need to recode too much. It is also well under the recommended cap of 50 MB.

Proposal

Research question: How much of a socioeconomic advantage do English speakers have over non-English speakers in Seattle? In this study, I plan to use public information released by the city of Seattle to visualize the connection between ELL status and certain health, education, and income variables. I hypothesize that, because of the institutional advantages of knowing English in America, English speakers will tend to be healthier, more educated, and have higher-paying jobs. The explanatory variable in this experiment is the amount of ELL people living in census tracts (represented in the data with a percentage, but I might break this up into a few bins) and the dependent variable is the variation in education level, income, asthma rates, mental health, and diabetes rates (also measured with percentages). I will test my hypothesis by trying to make all other factors equal and then plotting ELL percentages against the other variables I'm studying. If, after controlling for all those other factors, the lower-income census tracts have statistically significant worse health, education, and income, the pattern would support my hypothesis.

I plan to do this project by myself.