Project Proposal Revised

due October 20, 2021 by 11:59 PM

Isabelle Xiong and Maya Ghanem

10/17/2021

Load Data

```
X500_cities_diabetes <- read_excel("~/ Stats 198 Project/data/500_Cities_diabetes.xlsx")
X500_cities_health_insurance <- read_excel("~/ Stats 198 Project/data/500_cities_health_insurance.xlsx")
X500_cities_heart_diseases <- read_excel("~/ Stats 198 Project/data/500_cities_heart_diseases.xlsx")
X500_cities_mental_health <- read_excel("~/ Stats 198 Project/data/500_cities_mental_health.xlsx")
```

Wrangling Data

```
Insurance_CDC_Cities <- X500_cities_health_insurance %>%
  rename(percent_lack_insurance = Data_Value) %>%
  filter(StateDesc %in% c("North Carolin", "South Carolin", "California", "Alabama", "New York", "Ohio")
  filter(UniqueID %in% c("3702140-37021000100", "3710740-37037020701", "3712000-37119000100", "3714100-3
  select(CityName, percent_lack_insurance)
Mental_Health_CDC_Cities <- X500_cities_mental_health %>%
  rename(percent_mental_health = Data_Value) %>%
  filter(StateDesc %in% c("North Carolin", "South Carolin", "California", "Alabama", "New York", "Ohio")
  filter(UniqueID %in% c("3702140-37021000100", "3710740-37037020701", "3712000-37119000100", "3714100-3
  select(CityName, percent_mental_health)
Diabetes_CDC_Cities <- X500_cities_diabetes %>%
  rename(percent_diabetes = Data_Value) %>%
  filter(StateDesc %in% c("North Carolin", "South Carolin", "California", "Alabama", "New York", "Ohio")
  filter(UniqueID %in% c("3702140-37021000100", "3710740-37037020701", "3712000-37119000100", "3714100-3
  select(CityName, percent_diabetes)
Heart_Disease_CDC_Cities <- X500_cities_heart_diseases %>%
  rename(percent_heart_disease = Data_Value) %>%
  filter(StateDesc %in% c("North Carolin", "South Carolin", "California", "Alabama", "New York", "Ohio")
  filter(UniqueID %in% c("3702140-37021000100", "3710740-37037020701", "3712000-37119000100", "3714100-3
  select(StateDesc, CityName, GeoLocation, UniqueID, percent_heart_disease)
```

```
left_join_1 <- Mental_Health_CDC_Cities %>%
  left_join(Insurance_CDC_Cities)
```

```
left_join_2 <- left_join_1 %>%
  left_join(Diabetes_CDC_Cities)
```

```
final_CDC_data <- left_join_2 %>%
  left_join(Heart_Disease_CDC_Cities)
```

Introduction and Data, including Research Questions

Motivation for research:

Research Questions:

Glimpse

Data Analysis Plan

For the research questions, the explanatory and response variables are as follows:

The data analysis methodology is as follows:

Step 1: Contextualize and Situate the Data

Step 2: Conduct Statistical Tests

Step 3: Display Results of Statistical Tests

Step 4: Draw Conclusions

References