GEOG 111A Final Synthesis Lab

Fall 2022

Your Name: Tyler Padua

When you knit the .Rmd to a PDF for the first time, you may need to run `tinytex::install_tinytex()` in the console.

Load the packages and data

The packages and data are loaded for you. At this point, you should be quite familiar with the CHTS data. You can look up for variables in the `DataDictionary.xlsx`.

Load the packages

```
library(tidyverse)
library(maps)
```

Load the data

```
PersonData <- readRDS('PersonData_111A.Rds')
HHData <- readRDS('HHData_111A.Rds')
hh_bgDensity <- readRDS('hh_bgDensity.Rds')
personHHData <- left_join(PersonData, HHData) %>% left_join(hh_bgDensity)
```

Questions

Answer the following questions with code and texts (if asked). There are many different ways to write code to achieve the same goal, you are supposed to provide only ONE code solution. Only the first solution will be graded if you include multiple solutions.

1. What are the average PMT and number of trips for students? (5 points)

```
PersonData %>%
  filter(Student == 1) %>%
  summarise(Avg_PMT = mean(Sum_PMT),
            Avg_Trips = mean(Sum_Trips))
```

```
## # A tibble: 1 x 2
##   Avg_PMT Avg_Trips
##     <dbl>     <dbl>
## 1    20.6      3.04
```

Average PMT for students: 20.64522

Average number of trips for students: 3.038292

2. Write the code to return the number of employed males who work from home. (4 points)

```
PersonData %>%
  filter(Male == 1, WorkHome == 1, Employed == 1) %>%
  summarise(How_many = n())
```

```
## # A tibble: 1 x 1
##   How_many
##      <int>
## 1     4293
```

The number of employed males are working from home is 4293.

3. Write the code that outputs four columns: household id, person number, number of bike trips, and a column that suggests whether or not the person made at least one bike trip on the survey day. The value would be TRUE if the person made at least one bike trip and would be FALSE otherwise. (4 points)

```
PersonData %>%
  summarise(Household_id = hhid,
            Person_number = pnum,
            Bike_Trips = Bike_trips,
            At_least_one_trip = Bike_trips >= 1)
```

```
## # A tibble: 94,901 x 4
##    Household_id Person_number Bike_Trips At_least_one_trip
##           <dbl>         <dbl>      <dbl> <lgl>
##  1      1031985             1          0 FALSE
##  2      1031985             2          0 FALSE
##  3      1032036             1          0 FALSE
##  4      1032036             2          0 FALSE
##  5      1032036             3          0 FALSE
##  6      1032036             4          0 FALSE
##  7      1032036             5          0 FALSE
##  8      1032053             1          0 FALSE
##  9      1032053             2          0 FALSE
## 10      1032053             3          0 FALSE
## # ... with 94,891 more rows
```

4. How many people work from home and have a fixed schedule? How many people don't work from home and have a flexible schedule. Your code should return both results in **one output**. (6 points)

```
PersonData %>%
  filter(WorkHome ==1 & FlexSched == 0 | WorkHome == 0 & FlexSched == 1) %>%
  group_by(Work_From_Home = WorkHome == 1, Flexible_Schedule = FlexSched == 1) %>%
  summarise(How_many = n())
```

```
## `summarise()` has grouped output by 'Work_From_Home'. You can override using
## the `.groups` argument.
```

```
## # A tibble: 2 x 3
## # Groups:   Work_From_Home [2]
##   Work_From_Home Flexible_Schedule How_many
##   <lgl>          <lgl>                <int>
## 1 FALSE          TRUE                 22447
## 2 TRUE           FALSE                 1007
```

The number of people who work from home and have a fixed schedule: 1007

The number of people who don't work from home and have a flexible schedule: 22447

5. Which county has the highest average number of trips by bike? Indicate the county name and its average number of trips by bike in the answer. (5 points)

```
county_bike_trips <- personHHData %>%
  group_by(CTFIP, County) %>%
  summarise(Average_Bike_Trips = mean(Bike_trips))
```

```
## `summarise()` has grouped output by 'CTFIP'. You can override using the
## `.groups` argument.
```

```
view(county_bike_trips)
```

Yolo has the highest average number of trips by bike and it is 0.280858676.

6. Make a stacked bar chart that shows how mode share based on the number of trips differs between male and female. Make a stacked bar chart of mode share based on the number of trips, with one bar corresponding to male and one to female. Add an appropriate title and legend, and briefly discuss the results. (You do not need to attach your final plots, as long as we can run your code and get the plot ourselves) (10 points)

```
cPlot2data <- personHHData %>% select(Male, ends_with('trips'))
names(cPlot2data)
```

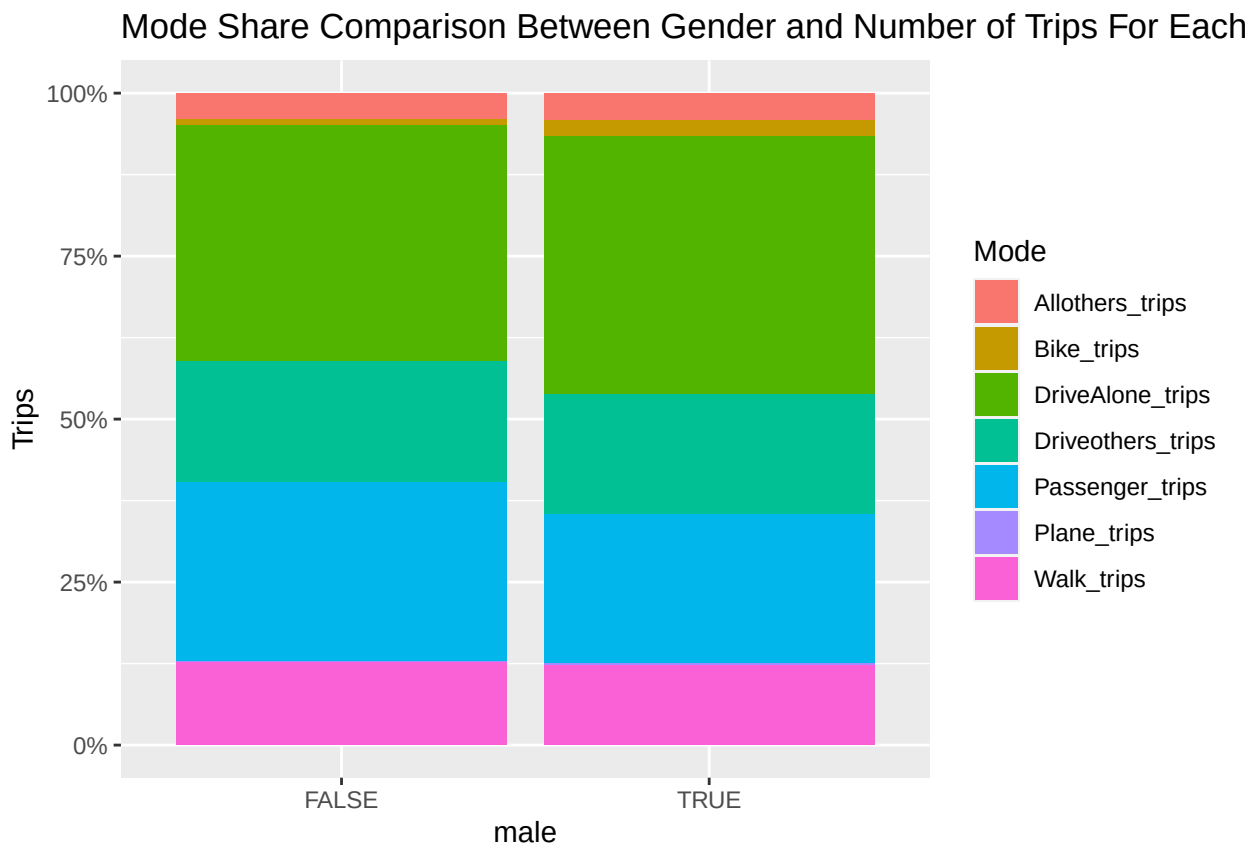
```
##  [1] "Male"              "Walk_trips"        "Bike_trips"
##  [4] "DriveAlone_trips"  "Driveothers_trips" "Passenger_trips"
##  [7] "Plane_trips"       "Allothers_trips"   "Sum_Trips"
## [10] "HH_nTrips"
```

```
cPlot2data <- cPlot2data %>% select(-Sum_Trips, -HH_nTrips)
names(cPlot2data)
```

```
## [1] "Male"              "Walk_trips"        "Bike_trips"
## [4] "DriveAlone_trips"  "Driveothers_trips" "Passenger_trips"
## [7] "Plane_trips"       "Allothers_trips"
```

```
cPlot2newvar <- cPlot2data %>% mutate(male = Male == 1,
                                      female = Male == 0)

cPlot2dataSum <- cPlot2newvar %>% group_by(male) %>%
```

```
  summarize_at(vars(ends_with('trips')),sum)

view(cPlot2dataSum)

cPlot2dataStacked <- cPlot2dataSum %>% gather('Mode', 'Trips', -male)

view(cPlot2dataStacked)

cPlot2 <- cPlot2dataStacked %>% ggplot(aes(x = male, y = Trips, fill = Mode))

cPlot2 + geom_bar(stat = 'identity', position = 'Fill') +
  scale_y_continuous(labels = scales::percent) +
  ggtitle('Mode Share Comparison Between Gender and Number of Trips For Each Mode')
```



Discussion: Males have a larger number of trips made by bike and plane. Females have a larger number of trips made by walking, driving alone, driving others, as passengers, and other all other trips.

7. What is the absolute difference in the average travel distance by walking between the employed and the unemployed? (4 points)

```
PersonData %>%
  group_by(Employed) %>%
  summarise(Average_walk_distance = mean(Walk_Dist))
```

```
## # A tibble: 3 x 2
```

```
##   Employed Average_walk_distance
##      <dbl>                 <dbl>
## 1        0                 0.217
## 2        1                 0.194
## 3       NA                 0.234
```

```
0.1938877 - 0.2172929
```

```
## [1] -0.0234052
```

The absolute difference is -0.0234052.

8. Which day of the week has the highest average miles-per-trip? Write the code to compute average miles-per-trip for each day of the week and find the day that has the highest value. (Reminder: you will exclude the people who didn't travel because they had zero trips) (6 points)

```
personHHData %>%
  group_by(DOW) %>%
  summarise(Average_miles_per_trip = mean(Sum_PMT) / mean(Sum_Trips > 0))
```

```
## # A tibble: 7 x 2
##   DOW       Average_miles_per_trip
##   <chr>                      <dbl>
## 1 Friday                      36.8
## 2 Monday                      32.6
## 3 Saturday                    42.2
## 4 Sunday                      38.0
## 5 Thursday                    32.0
## 6 Tuesday                     31.6
## 7 Wednesday                   33.3
```

Saturday has the highest average miles-per-trip.

9. Write the code that creates a table showing the average distance traveled by each of the following modes: Drive Alone, Passenger, and Walking on each day of the week for people with disability. (5 points)

```
personHHData %>%
  filter(Disability == 1) %>%
  group_by(DOW) %>%
  summarise(mean_Drive_Alone_Distance = mean(DriveAlone_Dist),
            mean_Passenger_Distance = mean(Passenger_Dist),
            mean_Walking_Distance = mean(Walk_Dist))
```

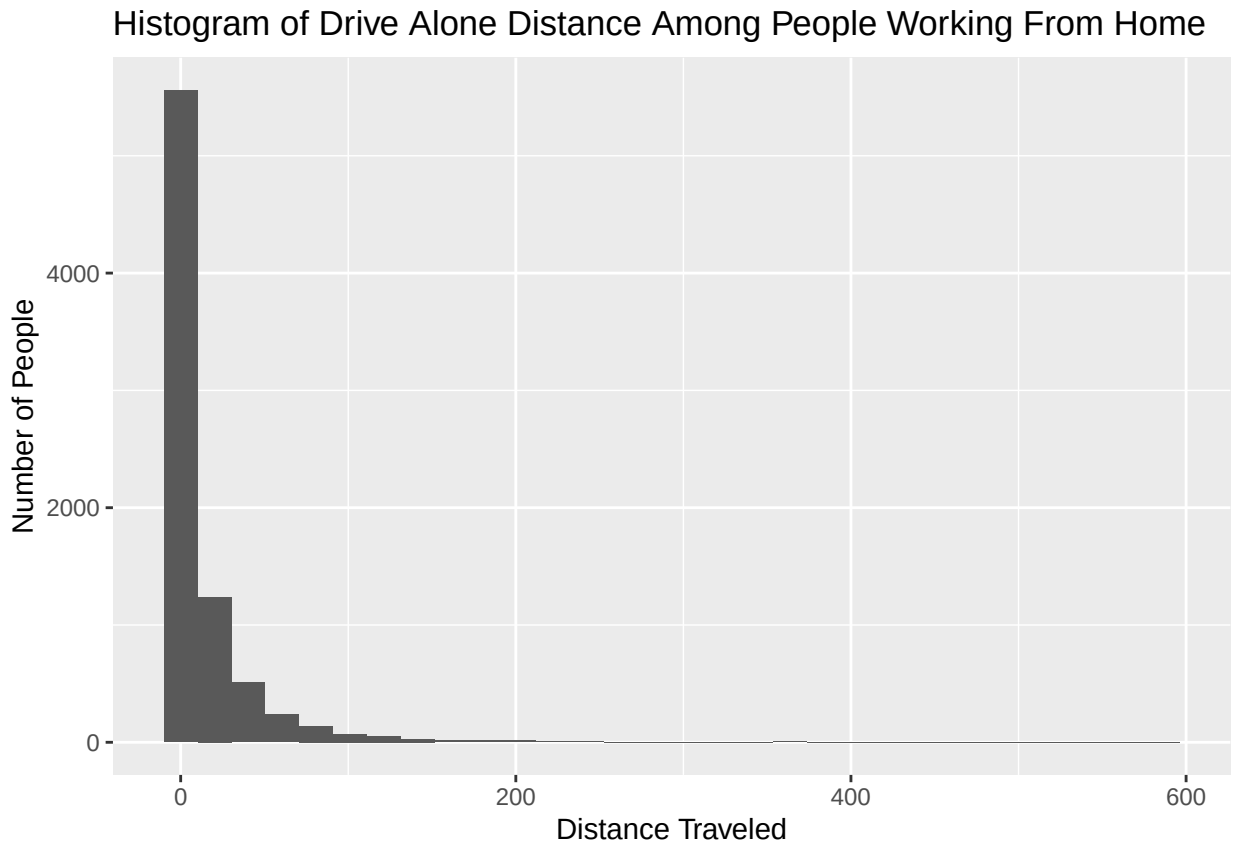
```
## # A tibble: 7 x 4
##   DOW       mean_Drive_Alone_Distance mean_Passenger_Distance mean_Walking_Dis~1
##   <chr>                         <dbl>                   <dbl>              <dbl>
## 1 Friday                         4.37                    7.04              0.151
## 2 Monday                         5.18                    5.49              0.172
## 3 Saturday                       4.25                    7.43              0.189
## 4 Sunday                         2.23                    6.82              0.133
```

```
## 5 Thursday                    5.91                  5.15              0.208
## 6 Tuesday                     5.52                  5.33              0.233
## 7 Wednesday                   5.66                  5.63              0.169
## # ... with abbreviated variable name 1: mean_Walking_Distance
```

10. Makes a histogram that shows how drive alone distance traveled varies over the people who work from home. Add an appropriate title, and briefly discuss the results. (You do not need to attach your final plots, as long as we can run your code and get the plot ourselves) (6 points)

```
drive_alone_dist <- personHHData %>%
  filter(WorkHome == 1) %>%
  summarise(drive_alone_dist = DriveAlone_Dist)

plot1var <- drive_alone_dist %>%
  ggplot(aes(x = drive_alone_dist))

plot1var + geom_histogram() + xlab('Distance Traveled') +
  ylab('Number of People') +
  ggtitle('Histogram of Drive Alone Distance Among People Working From Home')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Discussion: The histogram shows a large peak to the left side of the center which means that as the distance traveled increases, the number of people who traveled that far decreases. We can see that over 4000 people who work from home did not travel at all. Over half of the people that work from home traveled less than 100 miles.

11. Which area (among urban, suburb, exurb, and rural) do you see the lowest average PMT for employed people who are between 35-55 years old (including 35 and 55)? (5 points)

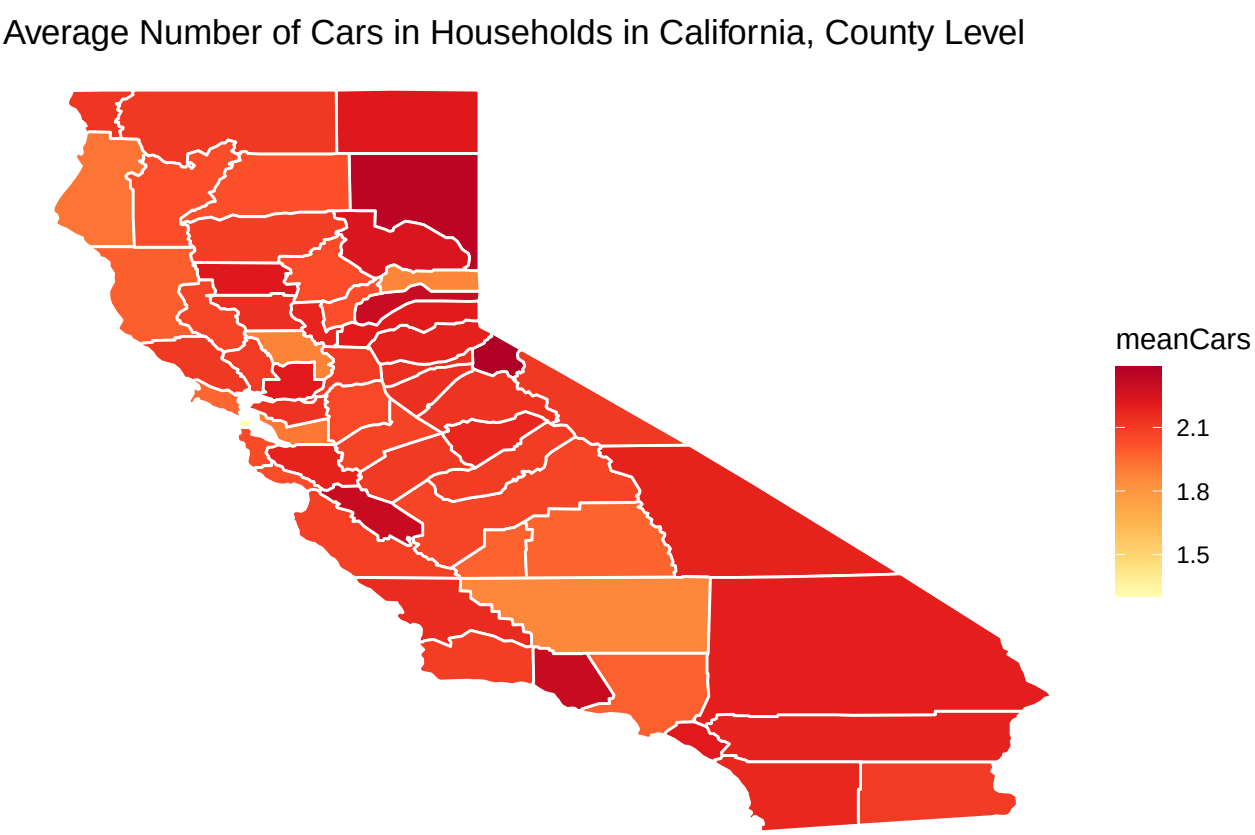
```
personHHData %>%
  filter(Age > 34 & Age < 56 & Employed == 1) %>%
  group_by(bg_group) %>%
  summarise(mean_PMT = mean(Sum_PMT))
```

```
## # A tibble: 4 x 2
##   bg_group mean_PMT
##   <fct>       <dbl>
## 1 Urban        29.4
## 2 Suburban     35.1
## 3 Exurban      41.2
## 4 Rural        41.8
```

Urban has the lowest average PMT for employed people who are between 35-55 years old (including 35 and 55).

12. Complete the following code to make a static map of the average number of cars in a household at county level. Remove background grid, add an appropriate title and choose appropriate color palette, and briefly discuss the results. (7 points)

```
county <- ggplot2::map_data("county", region = "california")

prhh_aggreg <- personHHData %>%
  group_by(County, CTFIP) %>%
  summarise(meanCars = mean(HH_nCars))

county_prhh <- prhh_aggreg %>%
  mutate(subregion = tolower(County)) %>%
  full_join(county, by = "subregion")

ggplot(county_prhh) +
  geom_polygon(aes(x = long, y = lat, group = subregion, fill = meanCars), colour = 'white') +
  scale_fill_distiller(palette = 'YlOrRd', direction = 1) +
  ggtitle('Average Number of Cars in Households in California, County Level') +
  theme_void()
```



Discussion: The average number of cars in California households ranges from about 1.3 to 2.4. The county on average with the highest number of cars per household is Alpine. The county on average with the smallest number of cars per household is San Francisco.

13. Do not include ANY irrelevant code! You can only provide ONE code solution to each question. (1 point)