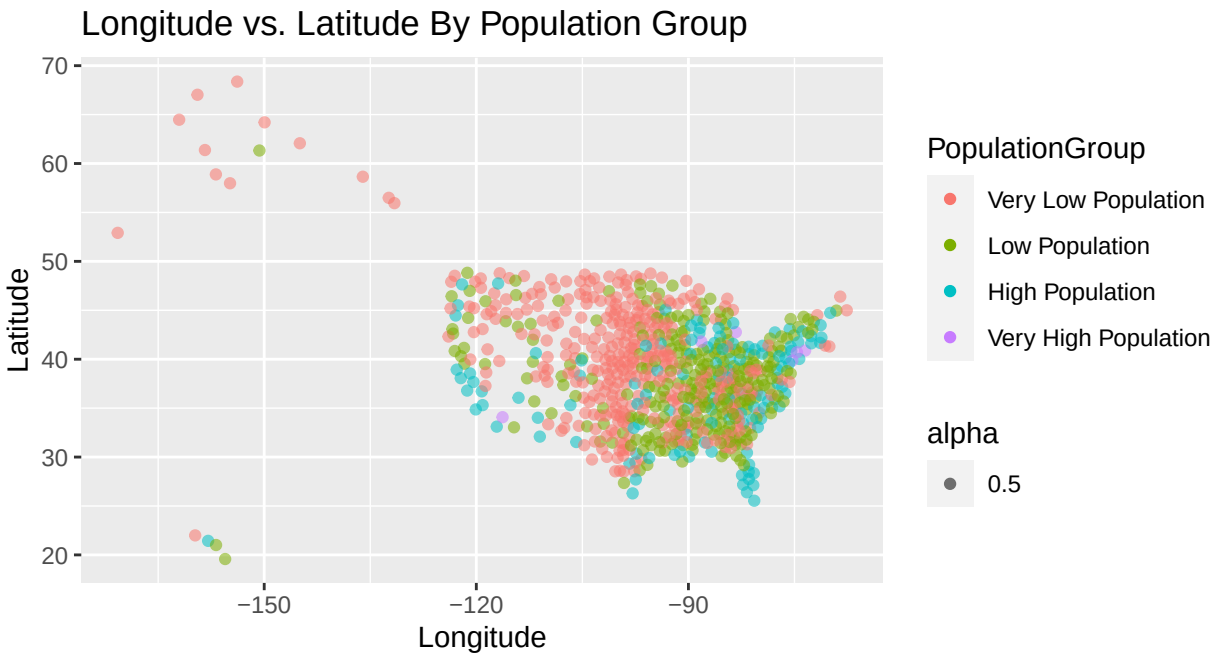
Assignment 2 Question 1

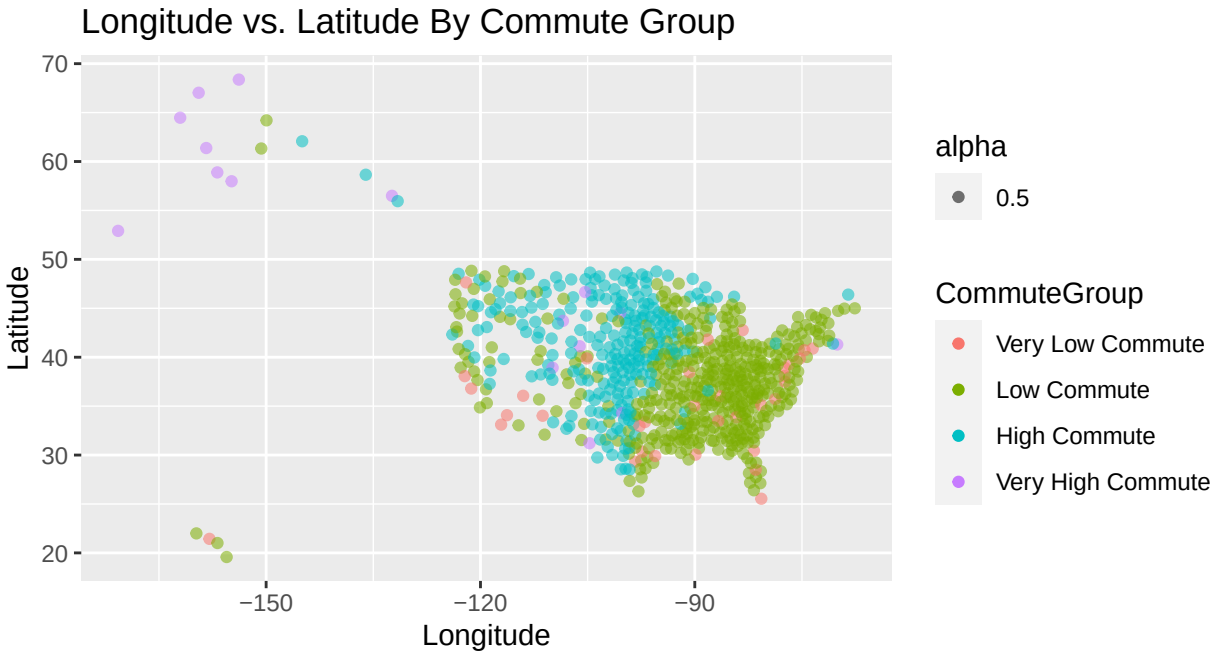
Sheen Thusoo

Part a)

```
library("ggplot2")
library("gridExtra")

data <- read.csv("EconomicMobility.csv")
data$PopulationGroup <- cut(data$Population,
breaks=c(0, 100000, 500000, 5000000, Inf), labels=c("Very Low Population","Low Population",
                                                    "High Population", "Very High Population"))

data$CommuteGroup <- cut(data$Commute,
breaks=c(0, 0.25, 0.5, 0.75, Inf), labels=c("Very Low Commute","Low Commute",
                                            "High Commute", "Very High Commute"))
```

```
plot1 <- ggplot(data) +
  geom_point(
    aes(x = Longitude,
        y = Latitude,
        color = PopulationGroup,
        alpha = 0.5)
  ) +
  labs(
    title = "Longitude vs. Latitude By Population Group",
    x = "Longitude",
    y = "Latitude"
  )

plot2 <- ggplot(data) +
  geom_point(
    aes(x = Longitude,
        y = Latitude,
        color = CommuteGroup,
        alpha = 0.5)
  ) +
  labs(
    title = "Longitude vs. Latitude By Commute Group",
    x = "Longitude",
    y = "Latitude"
  )

grid.arrange(plot1, plot2, nrow = 2)
```





Comment on the patterns that you observe in each plot. Reading these plots, do you observe any relationship between the two variables *commute* **and** *population***?**

The two plots are identical in shape and outliers as the variates we are plotting (longitude vs. latitude) are the same. This shape is similar to the country of the United States of America. In the Population Group plot, the high populations are grouped along the west coast and east coast as well. There also seems to be some high populations near the borders of the south. There seems to be a very population in the center and north border of the country. In the Commute Group plot, there seems to be a low commute time for most of the East side of the country and a higher commute for the center to center west of the country. However, the west coast seems to have a lower commute time as well. It seems that the relationship between the two variables *commute* and *population* are inversely proportional. So, areas where there is a lower commute seem to have higher populations and areas where there is a higher commute seem to have lower populations.

Thus, as the population increases, the fraction of workers with a commute time of less than 15 minutes decreases and as the population decreases, the fraction of workers with a commute time of less than 15 minutes increases.