Daniel Hanks Jr
IST565 – Assignment One

For this assignment I decided to explore what information I could gather about the students location in this class. My focus would be on determining what time zone is the dominant time zone. This could be particualarly helpful to a Professor who wants to know the best time to setup class meetings and wants to choose the best time for differences in time zone. As a student, maybe this could show what other students are in the same area to make group projects easier.

I chose to do this project in R, using the following packages:

Ggmap: mapping package for R.
Zipcode: This package contains city, state, latitude and longitude information for US ZIP codes. It also has built in features that will clean up the data, which is my main reason for using it.

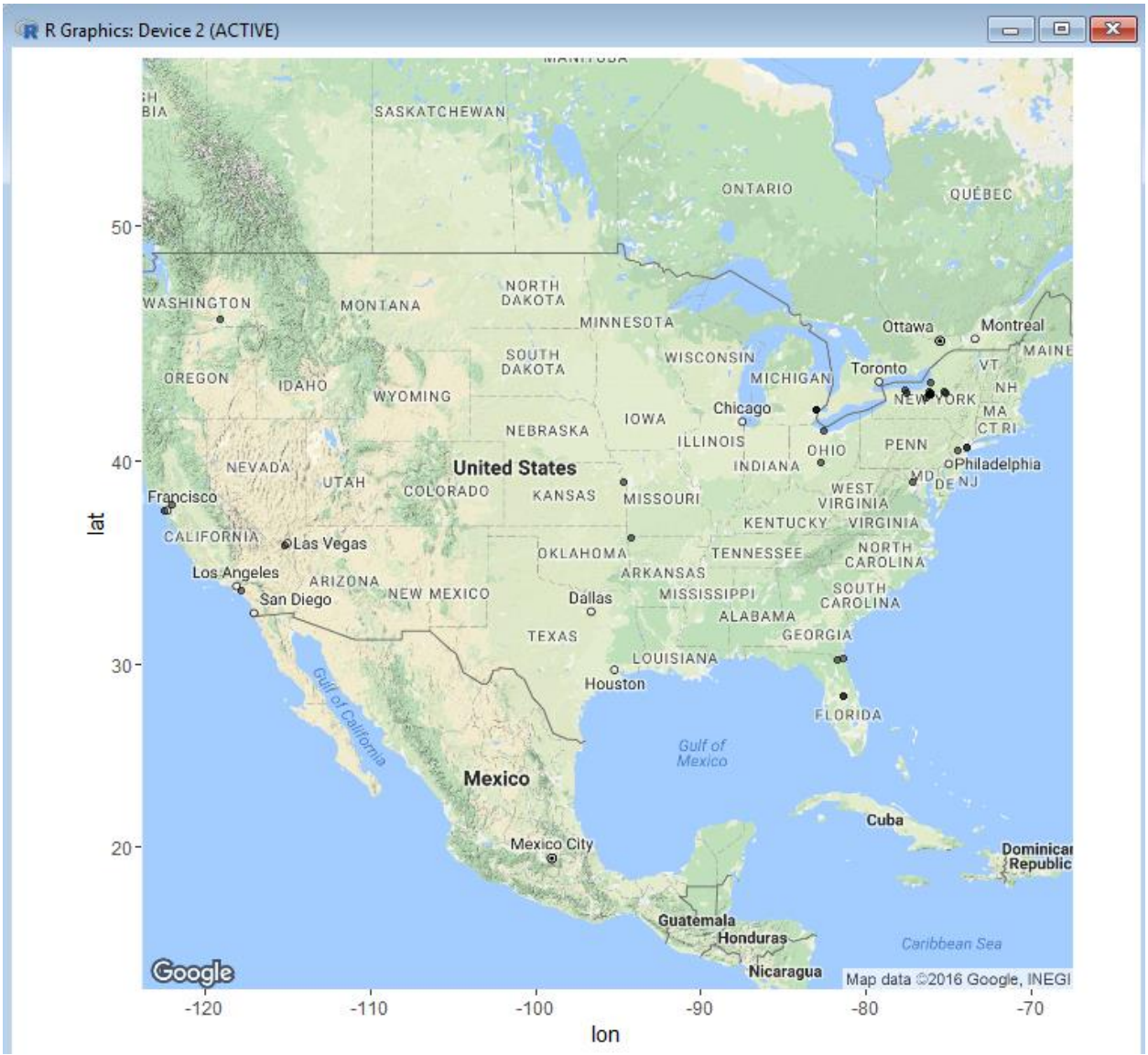
To get my results I loaded the csv file as shown here.

```
> df <- read.csv("IST565.Fall2016.withtext.csv", header = TRUE)
> df
         startDate         endDate        status       ipAddress progress
1       Start Date        End Date Response Type      IP Address Progress
2    9/3/2016 0:52   9/3/2016 0:56    IP Address     76.14.45.37      100
3    9/2/2016 0:45   9/2/2016 0:49    IP Address   73.171.205.37      100
4   9/1/2016 19:55  9/1/2016 20:04    IP Address    45.47.100.57      100
5   9/1/2016 12:10  9/1/2016 12:13    IP Address 206.211.153.213      100
6  8/30/2016 20:28 8/30/2016 21:03    IP Address   65.28.101.212      100
7  8/30/2016 19:28 8/30/2016 19:31    IP Address   73.192.120.99      100
8  8/29/2016 20:20 8/29/2016 20:25    IP Address    100.1.75.118      100
9  8/29/2016 19:44 8/29/2016 20:03    IP Address     45.47.15.19      100
10 8/29/2016 16:11 8/29/2016 16:15    IP Address      8.24.70.17      100
11 8/28/2016 20:50 8/28/2016 20:53    IP Address   97.100.86.207      100
12 8/28/2016 20:48 8/28/2016 20:50    IP Address   97.100.86.207      100
13 8/28/2016 13:09 8/28/2016 13:17    IP Address  104.230.227.44      100
14  8/28/2016 5:47  8/28/2016 5:50    IP Address   72.240.15.254      100
15  8/28/2016 1:20  8/28/2016 1:28    IP Address  71.115.167.109      100
16  6/5/2016 11:46  6/5/2016 11:51    IP Address  128.230.38.126      100
17   6/5/2016 9:54  6/5/2016 10:36    IP Address   35.16.234.120      100
18  5/30/2016 0:02  5/30/2016 0:05    IP Address    104.0.231.93      100
19 5/29/2016 20:42 5/29/2016 20:51    IP Address   24.92.242.146      100
20 5/28/2016 22:10 5/28/2016 22:24    IP Address   71.176.84.198      100
21 5/28/2016 14:23 5/28/2016 14:24    IP Address   35.16.232.149      100
```

I then loaded the ggplot2 library and installed and loaded the zipcode package and library. I did a simple copy and paste of the zip code data into a data frame in R and cleaned the data with the built-in features included with the zipcode package. I also had to merge the survey data into zipcode package data. This is shown here: (note the postal column was cleaned up and merged into the zip column.)

```
> zdata
     zip postal              city state latitude  longitude
1  08901   8901      New Brunswick    NJ 40.48830  -74.44775
2  11248  11248           Brooklyn    NY 40.64510  -73.94503
3  11248  11248           Brooklyn    NY 40.64510  -73.94503
4  13090  13090          Liverpool    NY 43.15202  -76.22068
5  13142  13142            Pulaski    NY 43.55993  -76.13619
6  13152  13152        Skaneateles    NY 42.89717  -76.38963
7  13206  13206           Syracuse    NY 43.06987  -76.10724
8  13206  13206           Syracuse    NY 43.06987  -76.10724
9  13208  13208           Syracuse    NY 43.07457  -76.14747
10 13208  13208           Syracuse    NY 43.07457  -76.14747
11 13210  13210           Syracuse    NY 43.03717  -76.12653
12 13210  13210           Syracuse    NY 43.03717  -76.12653
13 13210  13210           Syracuse    NY 43.03717  -76.12653
14 13212  13212           Syracuse    NY 43.12853  -76.13931
15 13219  13219           Syracuse    NY 43.04157  -76.22072
16 13219  13219           Syracuse    NY 43.04157  -76.22072
17 13244  13244           Syracuse    NY 43.03772  -76.13965
18 13244  13244           Syracuse    NY 43.03772  -76.13965
19 13244  13244           Syracuse    NY 43.03772  -76.13965
20 13244  13244           Syracuse    NY 43.03772  -76.13965
21 13244  13244           Syracuse    NY 43.03772  -76.13965
```

I then installed and used the ggmap package and library to create a map of the United States and plot the coordinates from the data.



Conclusion: Simply observing the table and a quick glance of the map shows most of the students are located on the east coast and mostly in New York itself. This would indicate the Eastern Time zone would be the dominant time zone. If the data sample was larger I would choose to use a heat map to give a better representation of student location. The small size of this survey allows me make my conclusions pretty easily just by looking at the table itself along with observing the heavy focus of pins in NY alone, let alone the east coast.

R code used:

```
#set working directory
setwd("C:/IST565/hw1")

#read csv file
df <- read.csv("IST565.Fall2016.withtext.csv", header = TRUE)

#install zipcode package
install.packages('zipcode')

#load zipcode
library(zipcode)
data(zipcode)

#load zipcodes into data frame
zdata = data.frame(postal=c(94014,
32250,13090,92866,66203,32210,8901,13219,13403,34741,34741,43068,44839,13210,13244,48202,72712,13208
,13152,48202,48202,13244,89139,13502,13210,20854,13208,99301,13142,13206,13206,13210,14626,13244,146
18,13244,13244,33-0006,13212,11248,11248,581-0085,13219,94521))

#clean and merge data
zdata$zip = clean.zipcodes(zdata$postal)
data(zipcode)
zdata=merge(zdata, zipcode, by.x='zip', by.y='zip')
zipcode$region = substr(zipcode$zip, 1, 1)

#create map-install and load ggmap package
install.packages('ggmap')
library(ggmap)

map<-get_map(location='united states', zoom=4, maptype = "terrain", source='google',color='color')

ggmap(map) + geom_point(
aes(x=longitude, y=latitude, show_guide = TRUE),
data=zdata, alpha=.5, na.rm = T)  + scale_color_gradient(low="beige", high="blue")
```