Harrison Montoya Assignments

Assignment 1

Due on Canvas on Monday 9/20 before class at 10:15 am

Problem 1

Install the datasets package on the console below using `install.packages("datasets")`. Now load the library.

```
USArrests
```

```
##                Murder Assault UrbanPop Rape
## Alabama          13.2     236       58 21.2
## Alaska           10.0     263       48 44.5
## Arizona           8.1     294       80 31.0
## Arkansas          8.8     190       50 19.5
## California        9.0     276       91 40.6
## Colorado          7.9     204       78 38.7
## Connecticut       3.3     110       77 11.1
## Delaware          5.9     238       72 15.8
## Florida          15.4     335       80 31.9
## Georgia          17.4     211       60 25.8
## Hawaii            5.3      46       83 20.2
## Idaho             2.6     120       54 14.2
## Illinois         10.4     249       83 24.0
## Indiana           7.2     113       65 21.0
## Iowa              2.2      56       57 11.3
## Kansas            6.0     115       66 18.0
## Kentucky          9.7     109       52 16.3
## Louisiana        15.4     249       66 22.2
## Maine             2.1      83       51  7.8
## Maryland         11.3     300       67 27.8
## Massachusetts     4.4     149       85 16.3
## Michigan         12.1     255       74 35.1
## Minnesota         2.7      72       66 14.9
## Mississippi      16.1     259       44 17.1
## Missouri          9.0     178       70 28.2
## Montana           6.0     109       53 16.4
## Nebraska          4.3     102       62 16.5
## Nevada           12.2     252       81 46.0
## New Hampshire     2.1      57       56  9.5
## New Jersey        7.4     159       89 18.8
## New Mexico       11.4     285       70 32.1
## New York         11.1     254       86 26.1
## North Carolina   13.0     337       45 16.1
## North Dakota      0.8      45       44  7.3
## Ohio              7.3     120       75 21.4
```

```
## Oklahoma          6.6     151       68 20.0
## Oregon            4.9     159       67 29.3
## Pennsylvania      6.3     106       72 14.9
## Rhode Island      3.4     174       87  8.3
## South Carolina   14.4     279       48 22.5
## South Dakota      3.8      86       45 12.8
## Tennessee        13.2     188       59 26.9
## Texas            12.7     201       80 25.5
## Utah              3.2     120       80 22.9
## Vermont           2.2      48       32 11.2
## Virginia          8.5     156       63 20.7
## Washington        4.0     145       73 26.2
## West Virginia     5.7      81       39  9.3
## Wisconsin         2.6      53       66 10.8
## Wyoming           6.8     161       60 15.6
```

Load the USArrests dataset and rename it `dat`. Note that this dataset comes with R, in the package datasets, so there's no need to load data from your computer. Why is it useful to rename the dataset?

```
dat<-USArrests
dat.USArrests <- dat
```

Answer: It is useful to rename the dataset for two reasons. First, it will help you keep track of your work and not confuse it with other generic-looking names of other datasets. Second, it will allow you to keep an original copy of the file while creating a new file with all of the changes you are currently making on it.

Problem 2

Use this command to make the state names into a new variable called State.

```
dat.USArrests$state <- tolower(rownames(USArrests))
dat.USArrests
```

```
##                Murder Assault UrbanPop Rape          state
## Alabama          13.2     236       58 21.2        alabama
## Alaska           10.0     263       48 44.5         alaska
## Arizona           8.1     294       80 31.0        arizona
## Arkansas          8.8     190       50 19.5       arkansas
## California        9.0     276       91 40.6     california
## Colorado          7.9     204       78 38.7       colorado
## Connecticut       3.3     110       77 11.1    connecticut
## Delaware          5.9     238       72 15.8       delaware
## Florida          15.4     335       80 31.9        florida
## Georgia          17.4     211       60 25.8        georgia
## Hawaii            5.3      46       83 20.2         hawaii
## Idaho             2.6     120       54 14.2          idaho
## Illinois         10.4     249       83 24.0       illinois
## Indiana           7.2     113       65 21.0        indiana
## Iowa              2.2      56       57 11.3           iowa
## Kansas            6.0     115       66 18.0         kansas
## Kentucky          9.7     109       52 16.3       kentucky
## Louisiana        15.4     249       66 22.2      louisiana
## Maine             2.1      83       51  7.8          maine
## Maryland         11.3     300       67 27.8       maryland
## Massachusetts     4.4     149       85 16.3  massachusetts
## Michigan         12.1     255       74 35.1       michigan
## Minnesota         2.7      72       66 14.9      minnesota
```

```
## Mississippi      16.1     259       44 17.1    mississippi
## Missouri          9.0     178       70 28.2       missouri
## Montana           6.0     109       53 16.4        montana
## Nebraska          4.3     102       62 16.5       nebraska
## Nevada           12.2     252       81 46.0         nevada
## New Hampshire     2.1      57       56  9.5  new hampshire
## New Jersey        7.4     159       89 18.8     new jersey
## New Mexico       11.4     285       70 32.1     new mexico
## New York         11.1     254       86 26.1       new york
## North Carolina   13.0     337       45 16.1 north carolina
## North Dakota      0.8      45       44  7.3   north dakota
## Ohio              7.3     120       75 21.4           ohio
## Oklahoma          6.6     151       68 20.0       oklahoma
## Oregon            4.9     159       67 29.3         oregon
## Pennsylvania      6.3     106       72 14.9   pennsylvania
## Rhode Island      3.4     174       87  8.3   rhode island
## South Carolina   14.4     279       48 22.5 south carolina
## South Dakota      3.8      86       45 12.8   south dakota
## Tennessee        13.2     188       59 26.9      tennessee
## Texas            12.7     201       80 25.5          texas
## Utah              3.2     120       80 22.9           utah
## Vermont           2.2      48       32 11.2        vermont
## Virginia          8.5     156       63 20.7       virginia
## Washington        4.0     145       73 26.2     washington
## West Virginia     5.7      81       39  9.3  west virginia
## Wisconsin         2.6      53       66 10.8      wisconsin
## Wyoming           6.8     161       60 15.6        wyoming
```

This dataset has the state names as row names, so we just want to make them into a new variable. We also make them all lower case, because that will help us draw a map later - the map function requires the states to be lower case.

List the variables contained in the dataset `USArrests`.

```
names(dat.USArrests)
```

```
## [1] "Murder"   "Assault"  "UrbanPop" "Rape"     "state"
```

Answer: The variables include Murder, Assault, Rape, Urban Population, and State.

Problem 3

What type of variable (from the DVB chapter) is `Murder`?

Answer: In the DVB chapter, "Murder" would be considered a qualitative, or categorical, variable.

What R Type of variable is it?

Answer: "Murder" is considered a character in R.

Problem 4

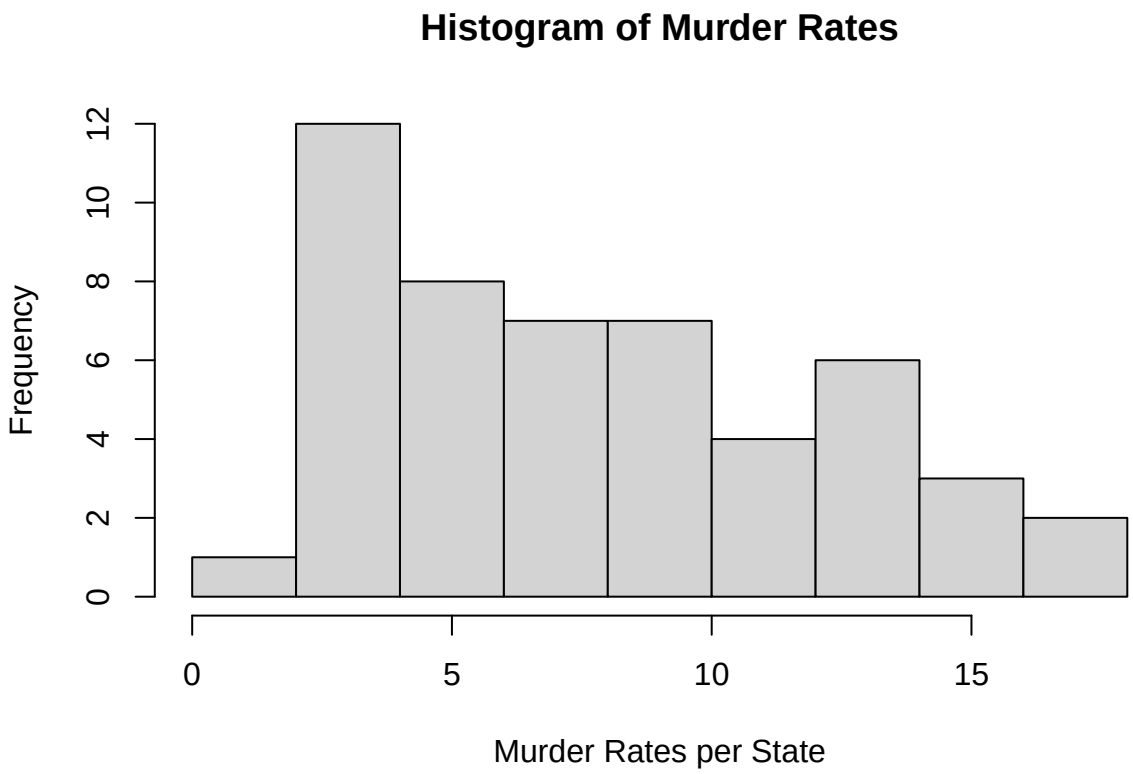
What information is contained in this dataset, in general? What do the numbers mean?

Answer: The dataset includes the arrest rates for murder, assault, and rape per 100,000 residents in each of the US's 50 states. Additionally, the percent of the population living in urban areas is given. Here, then, the numbers mean either the arrest rates for a crime per 100k residents in a state or the percent of residents living in urban spaces in a state.

Problem 5

Draw a histogram of `Murder` with proper labels and title.

```
hist(dat.USArrests$Murder, main="Histogram of Murder Rates", xlab="Murder Rates per State", ylab="Frequ
```



Problem 6

Please summarize `Murder` quantitatively. What are its mean and median? What is the difference between mean and median? What is a quartile, and why do you think R gives you the 1st Qu. and 3rd Qu.?

```
summary(dat.USArrests$Murder)
```

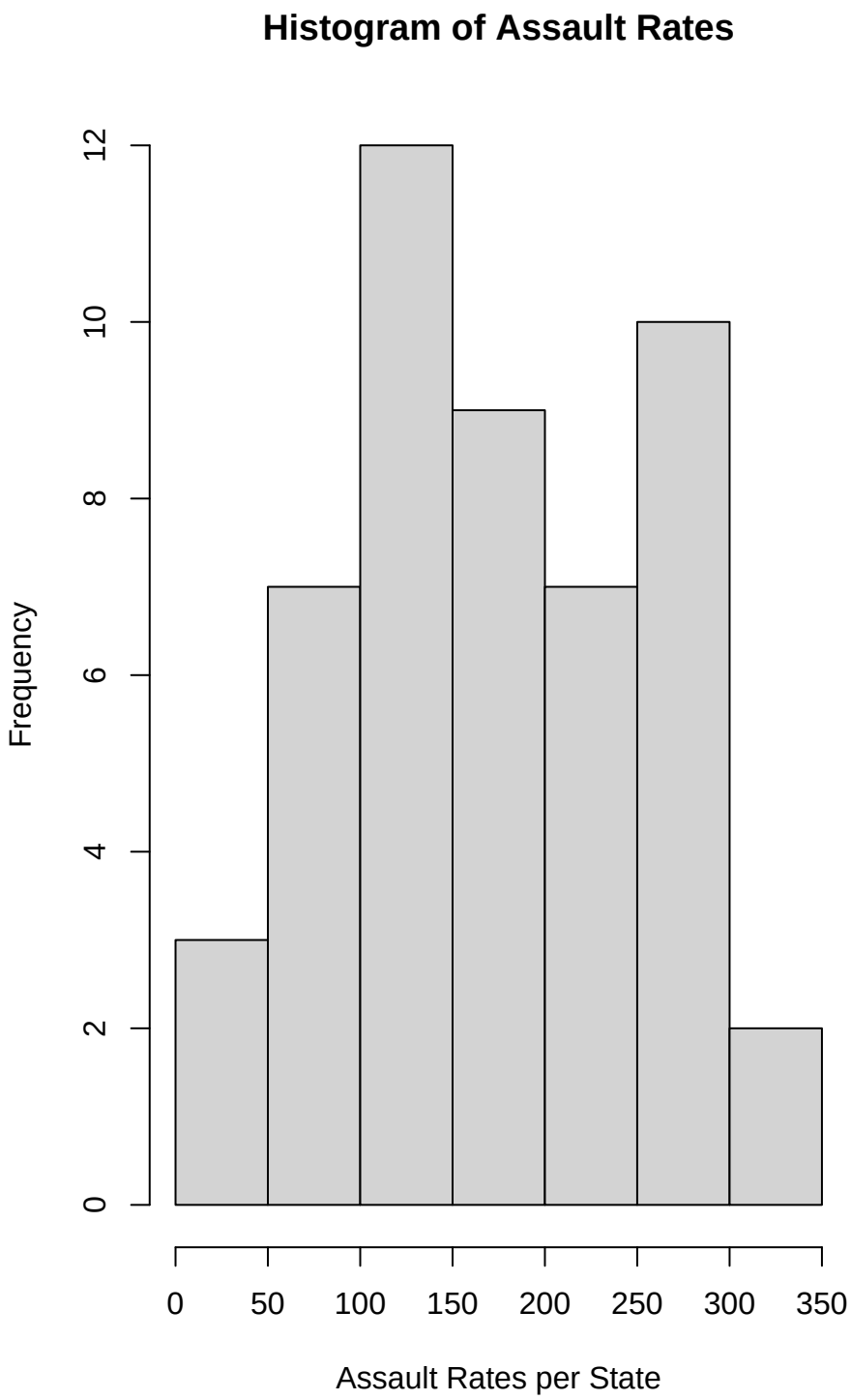
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.800   4.075   7.250   7.788  11.250  17.400
```

Answer: The mean of "Murder" is 7.788, while its median is 7.250. Generally, mean signifies the solution of all of the values added together and then divided by the number of values, while median signifies the middle value when all values are lined up in ascending order. A quartile constitutes one of three values that divides a data distribution into fourths. Lastly, R would provide the first and third quartile in order to help the statistician understand where the majority of values lie (in between the first and third quartile) or what values might be considered outliers (before the first and after the third).

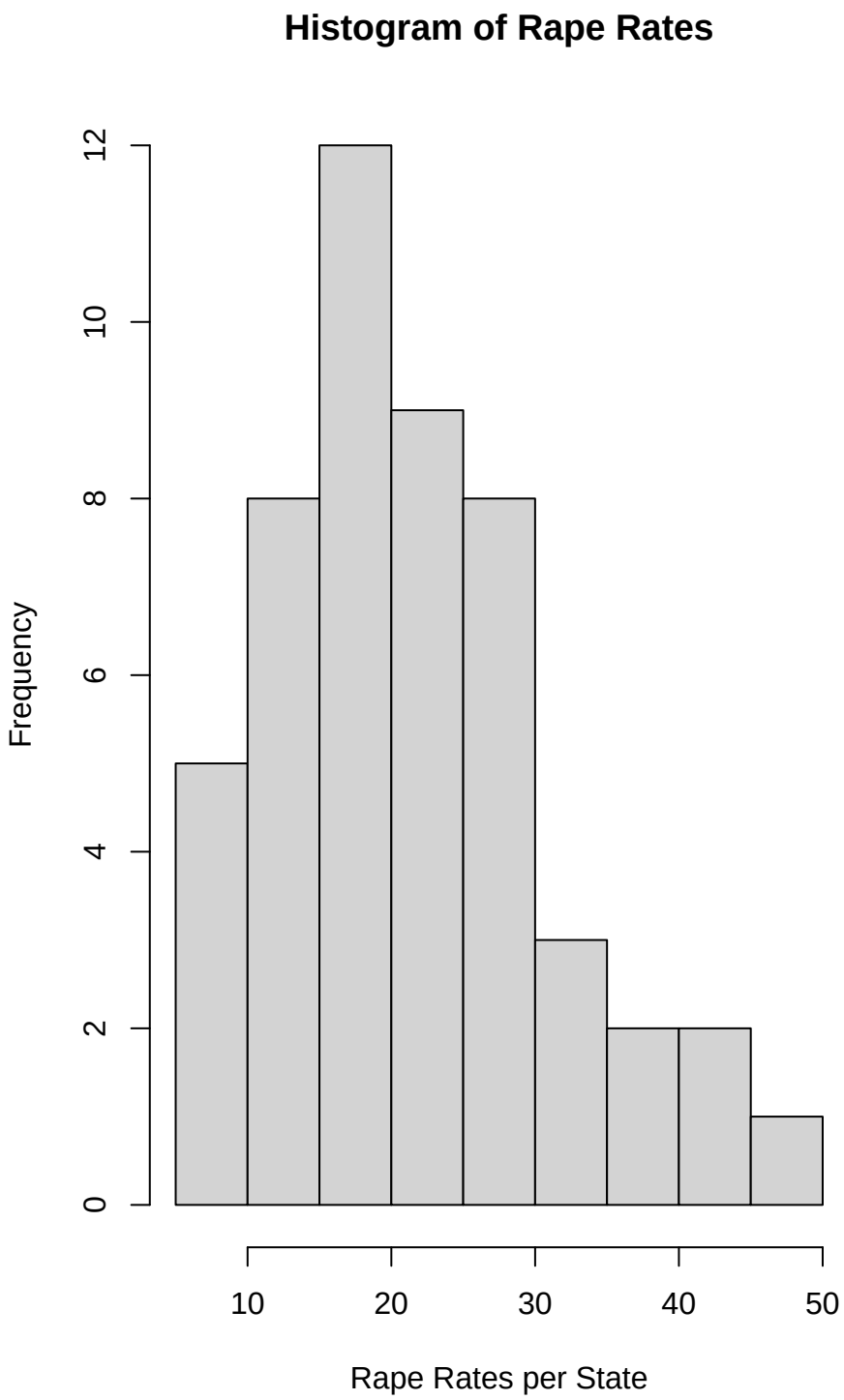
Problem 7

Repeat the same steps you followed for `Murder`, for the variables `Assault` and `Rape`. Now plot all three histograms together. You can do this by using the command `par(mfrow=c(3,1))` and then plotting each of the three.

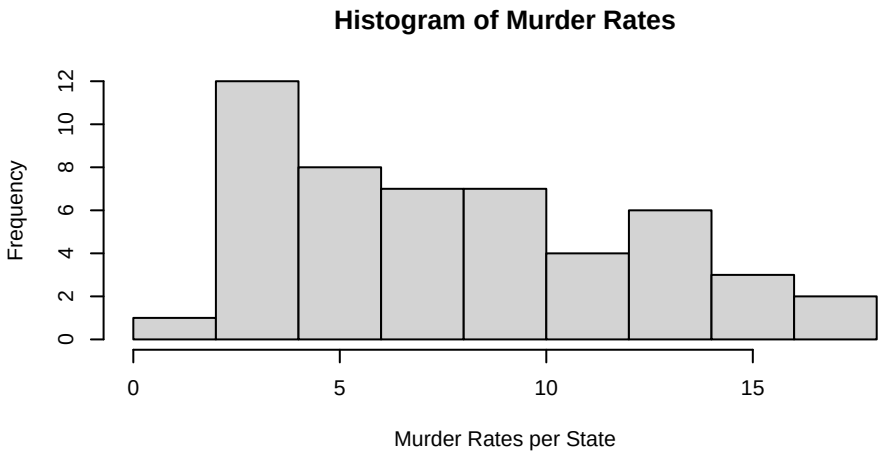
```
hist(dat.USArrests$Assault, main="Histogram of Assault Rates", xlab="Assault Rates per State", ylab="Fr
```

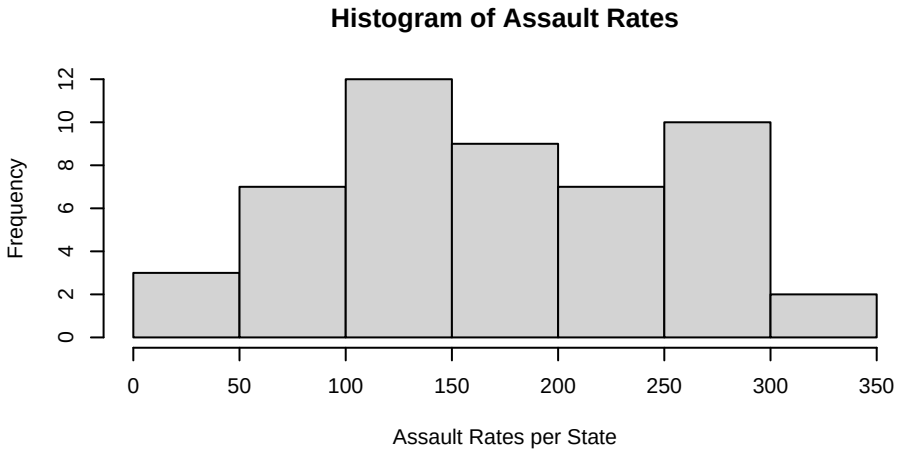


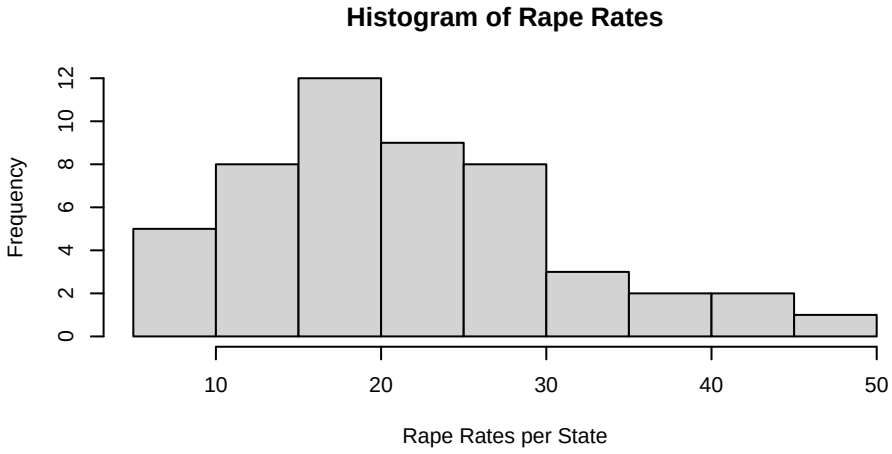

```
hist(dat.USArrests$Rape, main="Histogram of Rape Rates", xlab="Rape Rates per State", ylab="Frequency")
```




```
par(mfrow=c(3,1))
hist(dat.USArrests$Murder, main="Histogram of Murder Rates", xlab="Murder Rates per State", ylab="Frequ
hist(dat.USArrests$Assault, main="Histogram of Assault Rates", xlab="Assault Rates per State", ylab="Fr
hist(dat.USArrests$Rape, main="Histogram of Rape Rates", xlab="Rape Rates per State", ylab="Frequency")
```







What does the command par do, in your own words (you can look this up by asking R `?par`)?

Answer: The command par enables the statistician to set graphical parameters for data in either a singular graph or multiple graphs.

What can you learn from plotting the histograms together?

Answer: By plotting histograms together, you are able to compare the data between different categories – in this case, comparing the differences in assault and murder rates per state, for example. Additionally, you can

gain a better understand of the data overall by looking at it holistically instead of piece-by-piece.

Problem 8

In the console below (not in text), type `install.packages("maps")` and press Enter, and then type `install.packages("ggplot2")` and press Enter. This will install the packages so you can load the libraries.

Run this code:

```
library('maps')
library('ggplot2')

ggplot(dat, aes(map_id=state, fill=Murder)) +
  geom_map(map=map_data("state")) +
  expand_limits(x=map_data("state")$long, y=map_data("state")$lat)
```

What does this code do? Explain what each line is doing.

Answer: This code is mapping the arrest rates of murder per 100,000 citizens per state. With this, we are able to see the salience and prominence of arrest rates through a colored map of the United States. The first line is using the data groups of "state" and "Murder" to construct aesthetic mapping in a ggplot, filling the map with "Murder" rates. Next, the second line is the direction to map the states, while the third line is expanding the x and y axes, i.e. longitude and latitude, in the graph.