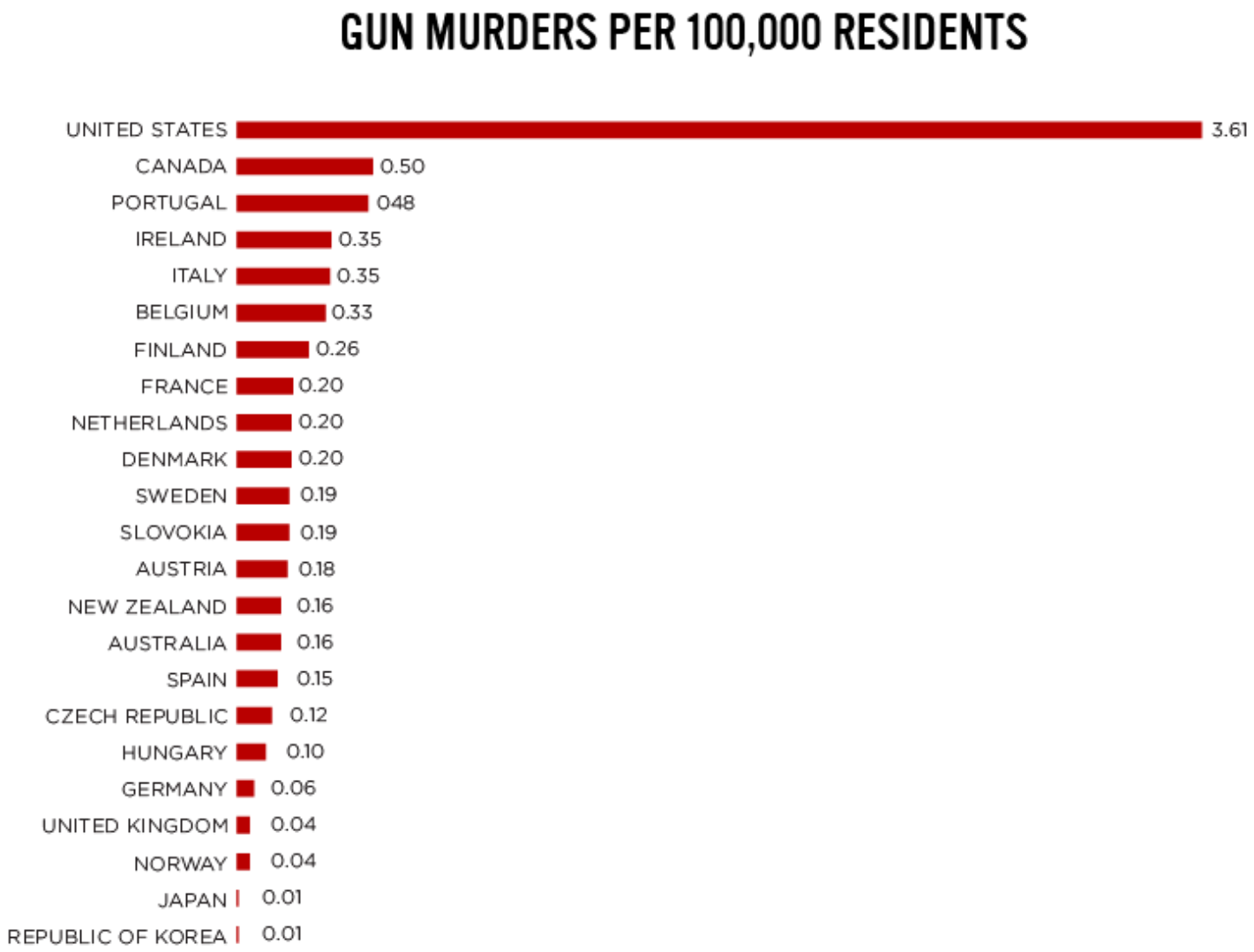
Homework 1: US gun murders

Jenny Wang (71401898)

September 4, 2019

Due: September 27th by 11:59pm

Several of your friends live in Europe and are offered jobs at a US company with many locations all across the country. The job offers are great but news with headlines such as **America is one of 6 countries that make up more than half of guns deaths worldwide** (https://www.vox.com/2018/8/29/17792776/us-gun-deaths-global) have them worried. Charts like this make them worry even more:



US gun homicides chart

You want to convince your friends that the US is a large and diverse country with 50 very different states as well as the District of Columbia (DC). You want to recommend some states for each friend knowing that some like hiking, while others would like to be close to several large cosmopolitan cities. Use data from the US murders data set (https://www.rdocumentation.org/packages/dslabs/versions/0.7.1/topics/murders):

```
library(dslabs)
data(murders)
```

Question 1

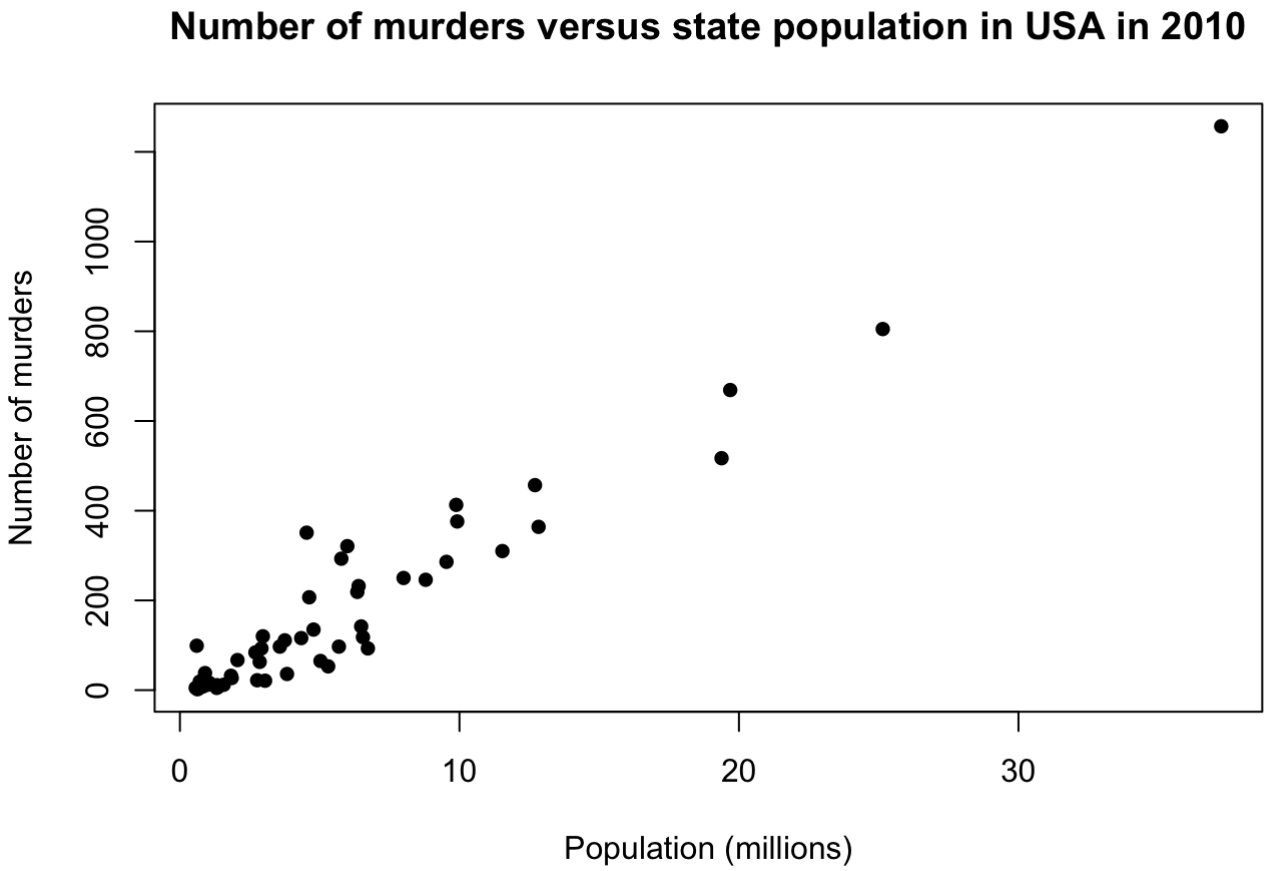
What is the state with the most murders? Would you say this is the most dangerous state? Hint: Make a plot showing the relationship between population size and number of murders.

```
library(dplyr)
```

```
murders %>%
  filter(total==max(total)) %>%
  select(state)
```

```
##        state
## 1 California
```

```
plot(murders$population/10^6, murders$total, pch=16,
     xlab="Population (millions)", ylab="Number of murders",
     main="Number of murders versus state population in USA in 2010")
```



California has the most murders. However, it is **not** the most dangerous state because it has more people to begin with. To find the most dangerous state, we would need to look for the state with a large number of murders on top of a small population.

Question 2

Add a column to the murder data table called `murder_rate` with each state's murder rate.

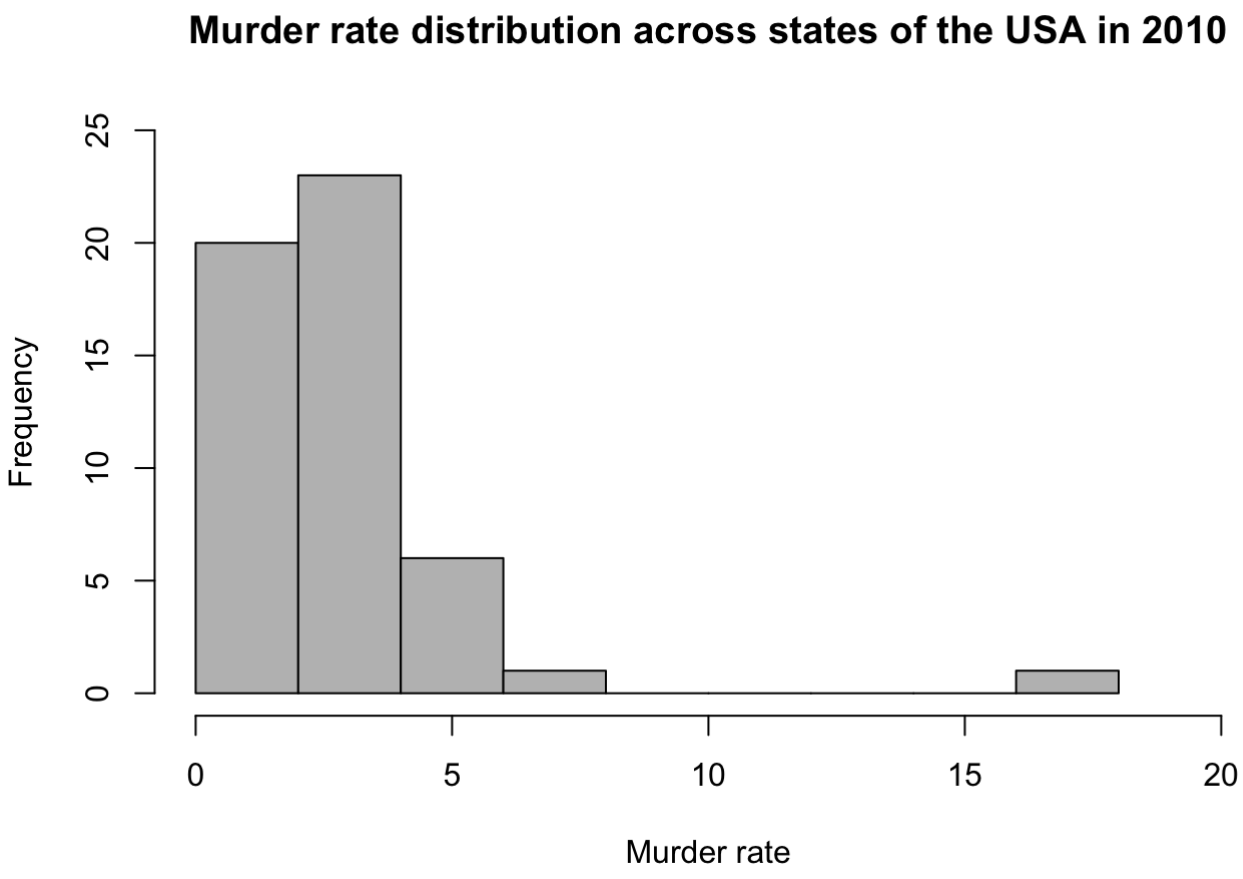
```
murders$murder_rate <- murders$total / murders$population*100000

head(murders)
```

```
##        state abb region population total murder_rate
## 1    Alabama  AL  South    4779736   135    2.824424
## 2     Alaska  AK   West     710231    19    2.675186
## 3    Arizona  AZ   West    6392017   232    3.629527
## 4   Arkansas  AR  South    2915918    93    3.189390
## 5 California  CA   West   37253956  1257    3.374138
## 6   Colorado  CO   West    5029196    65    1.292453
```

Question 3

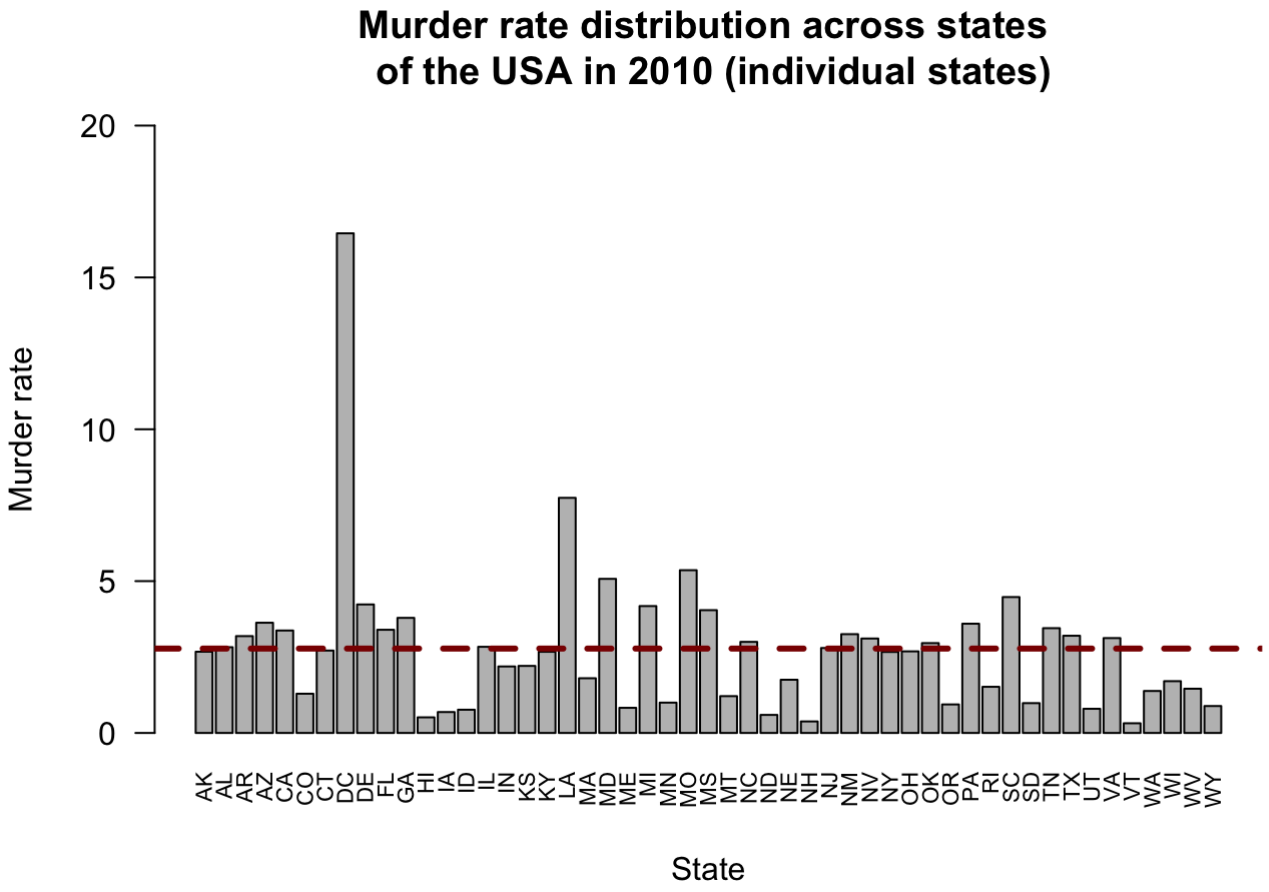
Describe the distribution of murder rates across states. How similar are states? How much do murder rates vary by geographical regions?

```
# Histogram
hist(murders$murder_rate, col="grey",
     xlim=c(0, 20), ylim=c(0, 25),
     xlab="Murder rate",
     main="Murder rate distribution across states of the USA in 2010")
```



From the histogram, most states have a murder rate of less than or equal to 5, meaning on average, there are 5 murders per 100,000 people in these states. One state/region (Washington DC) has the largest murder rate in the country, more than 3 times the national average. This gives the histogram a large skew to the right.

```
# Barplot
par(las=2)
with(murders, barplot(murder_rate ~ abb,
                      ylim=c(0, 20), cex.names=0.7,
                      xlab="State", ylab="Murder rate",
                      main=paste("Murder rate distribution across states", "\n", "of the
USA in 2010 (individual states)")))
abline(h=mean(murders$murder_rate), col="red4", lty=2, lwd=3)
```




```
# Number of states above the mean
murders %>%
  filter(murder_rate > mean(murder_rate)) %>%
  select(state) %>%
  nrow()
```

```
## [1] 24
```

```
# Number of states below the mean
murders %>%
  filter(murder_rate < mean(murder_rate)) %>%
  select(state) %>%
  nrow()
```

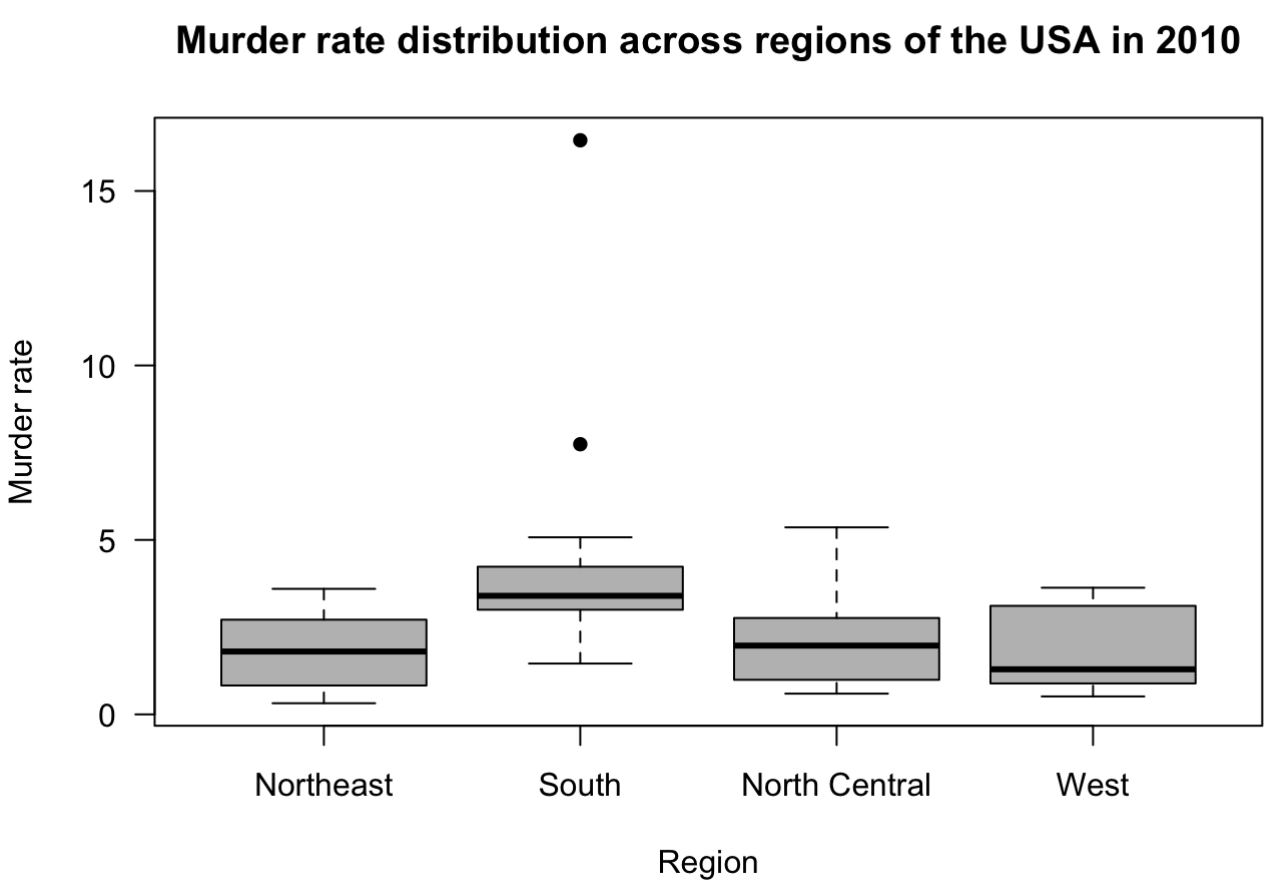
```
## [1] 27
```

```
# Less than or equal to 0.71
murders %>%
  filter(murder_rate <= 0.71) %>%
  select(state)
```

```
##           state
## 1        Hawaii
## 2          Iowa
## 3 New Hampshire
## 4  North Dakota
## 5       Vermont
```

From the data and the barplot, 23 states and DC have murder rates above the mean murder rate of the entire country, while 27 states have murder rates that fall below the mean. Five states have a murder rate of less than or equal to 0.71, namely Hawaii, Iowa, New Hampshire, North Dakota, and Vermont.

```
# Boxplot
par(las=1)
boxplot(murder_rate ~ region, data=murders, col="grey", pch=16,
        xlab="Region", ylab="Murder rate",
        main="Murder rate distribution across regions of the USA in 2010")
```



From the boxplot, the region with the lowest median murder rate is the West; however, most states in the West have a murder above the median, giving the distribution an upward skew. The Northest and North Central regions have the second lowest median murder rate, with the North Central region also displaying an upward skew. The highest median murder rate is found in the South, which also has 2 outliers that significantly skew the data upwards.

Question 4

Write a report for your friends reminding them that the US is a large and diverse country with 50 very different states as well as the District of Columbia (DC). Suppose one of your friends loves hiking, one wants to live in a warm climate, and another would like to be close to several large cosmopolitan cities. Recommend a desirable state for each friend. Answers should be a minimum of 1 paragraph and a maximum of 3 paragraphs.

If you like hiking, I would suggest moving to Colorado. Some notable hikes include Pike's Peak, Garden of the Gods, Longs Peak, Devil's Head Lookout, Emerald Lake, and Quandary Peak. Additionally, Colorado's murder rate is well below the national average. However, if this still concerns you, you can always go to New Hampshire. There are beautiful mountains and copious trails, including the famous Appalachin Trail. It is also one of the safest states in the country. The only downside with New Hampshire is that it is cold and a little bit difficult to get around. You might consider getting a car. The Rocky Mountains also span various states. You may want to go to (or visit) Wyoming, Utah, Idaho, and Montana. These are fairly safe states with trails with high elevations, if you are into that.

If you are looking for somewhere warm to live, my first pick would be Hawaii. It is warm all year long, and you will feel like you are on vacation all the time. It is also one of the safest states in the country, benchmarked alongside New Hampshire. However, if you don't want to live too far from the mainland, I

would suggest California. It is true that California has the highest number of murders in the nation, but keep in mind that California is also the most populated state. The murder rate is not as concerning as the raw number of murders. San Diego and the Bay Area are some nice cities to live in.

If you would like to be close to cosmopolitan cities, I would suggest that you come to Massachusetts. Boston is a bumbling city in the heart of Massachusetts, and from here, it is only 1.5 hours plane ride to New York City, Philadelphia, and Washington DC. You could even drive if you want to. It is a little farther form Chicago and the west coast cities, but it is also a lot safer with lower murder rates. Also, compared to Chicago's winter, Boston's winter is a godsend.