

# Mervar - Lab 1

Alexander Mervar

2022-08-31

1. Install R on your machine by visiting <https://www.r-project.org/>

**Done**

2. Most likely, you'll want to install R studio as well. It is a useful interactive environment for R programming that includes windows for code, instantaneously executed commands, variables, and plots. Download R Studio at <https://www.rstudio.com/products/rstudio/download2/>

**Done**

3. Read in the data file using the command: `cities<-read.table("/Users/rgoldsto/Desktop/top200cities.txt", header = TRUE)` replacing the path with the one that leads to where you put the file.

```
cities<-read.table("top200cities.txt", header = TRUE)
```

4. Cities has the following variables: rank, city [the city's name], state, pop2015 [population in 2015], pop2010, pctchange [% change to population from 2010 to 2015], landsqmi [area of city in square miles], landsqkm, popdensity [population density in square miles], popdensitysqkm, latitude [in degrees; larger values are more to the North], longitude [in degrees; larger values are to the West]. Report these names with the command: `names(cities)`

```
names(cities)
```

```
## [1] "rank"          "city"          "state"         "pop2015"
## [5] "pop2010"       "pctchange"     "landsqmi"      "landsqkm"
## [9] "popdensitysqmi" "popdensitysqkm" "latitude"      "longitude"
```

5. Are there more cities to the East or West of the midpoint of the continental U.S., which is located at 98.35 degrees? Show the R commands you used to answer this question. You shouldn't have to count them by hand.

```
midpointLong<-98.35
numOfWest<-nrow(cities[cities$longitude>midpointLong,])
numOfEast<-nrow(cities[cities$longitude<midpointLong,])
numOfWest
```

```
## [1] 81
```

```
numOfEast
```

```
## [1] 119
```

Looking at the data, there are more cities on the East side of the midpoint of the continental U.S., which is located at 98.35 degrees.

6. Are there more cities to the South or North of the midpoint of the continental U.S., which is located at 39.50 degrees? Show the R commands you used to answer this question.

```
midpointLat<-39.5
numOfNorth<-nrow(cities[cities$latitude>midpointLat,])
numOfSouth<-nrow(cities[cities$latitude<midpointLat,])
numOfNorth
```

```
## [1] 55
```

```
numOfSouth
```

```
## [1] 145
```

Looking at the data, there are more cities to the South side of the midpoint of the continental U.S., which is located at 39.50 degrees.

7. Show the R command that let's you answer the question: what is the population density of Miami?

```
cities$popdensqmi[cities$city=="Miami"]
```

```
## [1] 11539
```

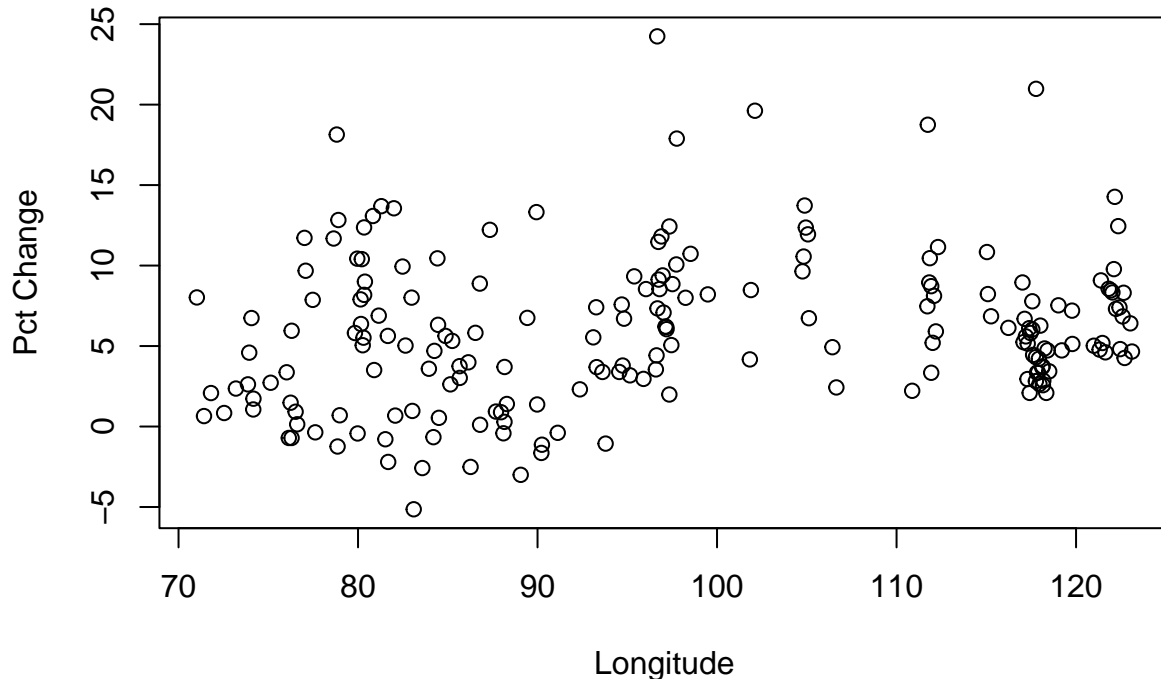
8. There's one city that has a percentage growth more than 50% from 2010 to 2015. Show the R command that finds and prints the name of this city.

```
cities$city[cities$pctchange>50]
```

```
## [1] "Macon"
```

9. Make a new data frame called `mostcities` that is the same as `cities` but excluding cities with growth more than 30% or longitudes > 140 degrees (Hawaii and Alaska). Now show the scattergram that plots (using “plot”) the percentage change in `mostcities` as a function of longitude. The plot should have one circle for each of the remaining 196 cities. Looking at this plot, describe how recent growth of cities in America varies with their East-West location.

```
mostcities<-cities[cities$pctchange<30,]
mostcities<-mostcities[mostcities$longitude<140,]
plot(mostcities$longitude,mostcities$pctchange,xlab = "Longitude",ylab = "Pct Change")
```



Looking at the plot, it seems that the West side of the country has a major packet of cities at longitude 118 that have strong amount of growth ~3-7%. The East side of the country does not seem to have this kind of strong packet of cities. Instead, it has a lot of smaller packets with a cluster of a handful of cities.

10. Make a new data frame called `continguoUS` that is the same as `cities` but excludes cities in Hawaii and Alaska (which have longitudes > 140 degrees). Show a plot of where the cities are that looks like plot below. Note: because increasing longitudes go from East to West, you'll need to come up with some way to reverse the longitude values so that the map doesn't come out reversed.

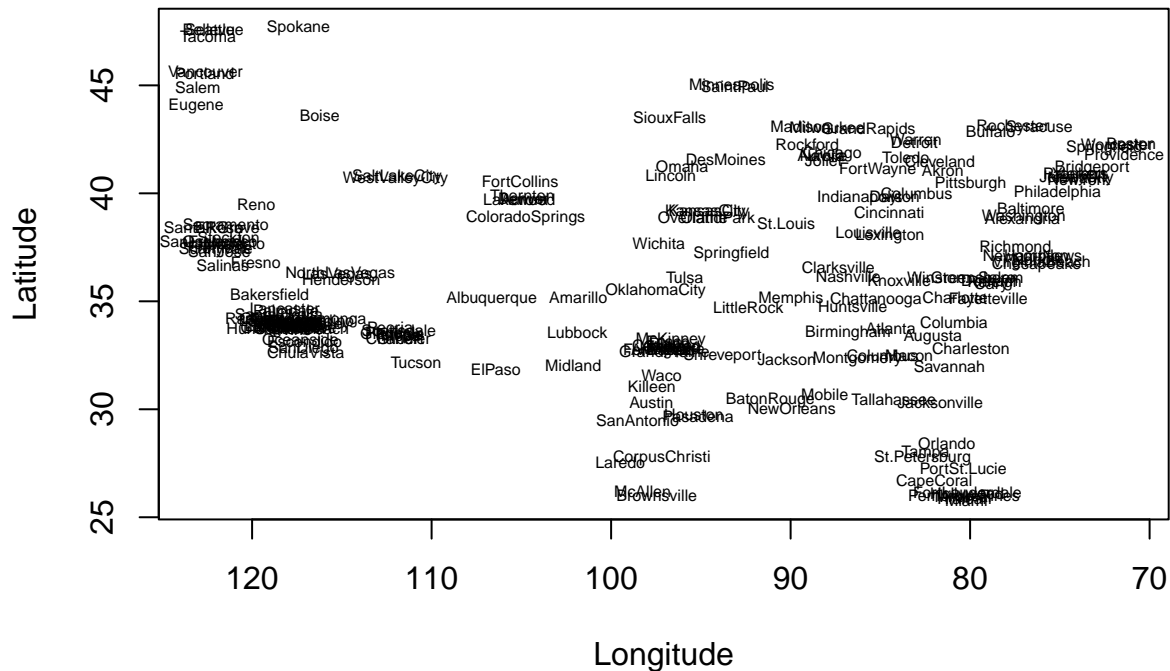
```
continguoUS <- cities[cities$longitude<140,]
plot(xlim = rev(range(continguoUS$longitude)),continguoUS$latitude, type="n", xlab = "Longitude", ylab = "Latitude")
text(continguoUS$longitude,continguoUS$latitude, labels=continguoUS$city,cex= 0.5)

## Warning in text.default(continguoUS$longitude, continguoUS$latitude, labels =
## continguoUS$city, : conversion failure on 'Winston-Salem' in 'mbcsToSbcs': dot
## substituted for <e2>

## Warning in text.default(continguoUS$longitude, continguoUS$latitude, labels =
## continguoUS$city, : conversion failure on 'Winston-Salem' in 'mbcsToSbcs': dot
## substituted for <80>

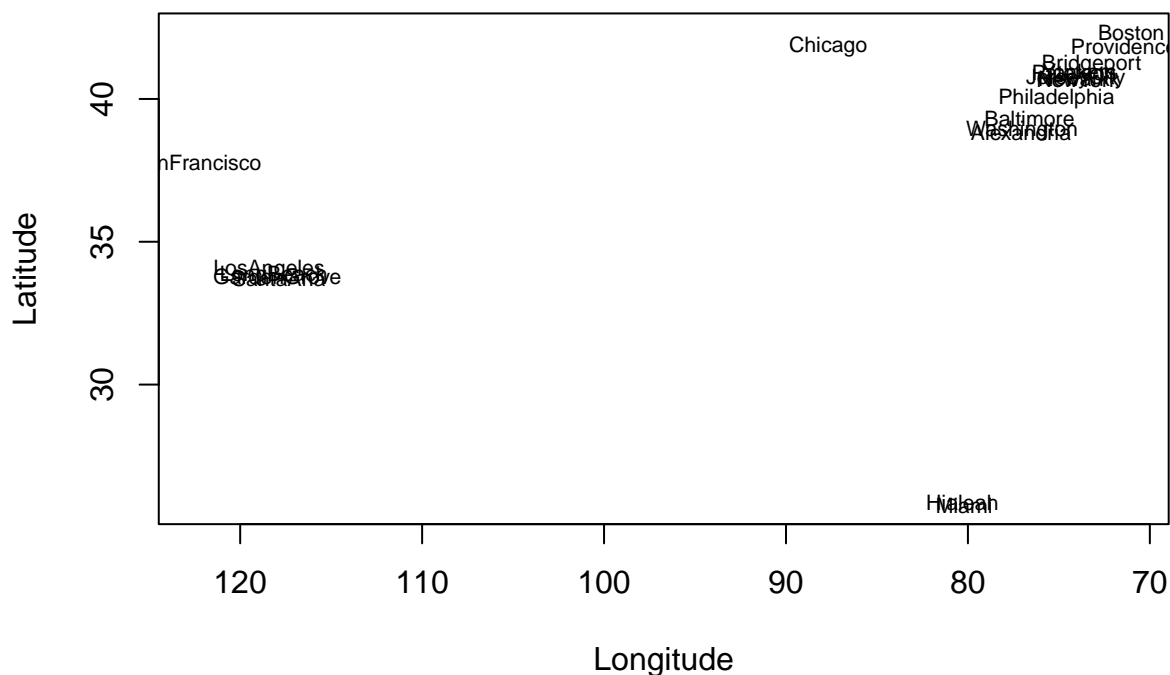
## Warning in text.default(continguoUS$longitude, continguoUS$latitude, labels =
## continguoUS$city, : conversion failure on 'Winston-Salem' in 'mbcsToSbcs': dot
## substituted for <93>
```

```
## Warning in text.default(continguoUS$longitude, continguoUS$latitude, labels =
## continguoUS$city, : font metrics unknown for Unicode character U+2013
```



11. Make a plot of the most densely populated 20 cities in the contiguous US. You'll probably want to first (reverse) order the cities in terms of population density and then extract a subset of 20 of the cities. It should look something like:

```
denselyPop <- continguoUS[rev(order(continguoUS$popdensitysqmi))[1:20], ]
plot(xlim = rev(range(denselyPop$longitude)),denselyPop$latitude, type="n", xlab = "Longitude", ylab = "Latitude")
text(denselyPop$longitude,denselyPop$latitude, labels=denselyPop$city,cex= 0.7)
```



12. **Extra Credit:** According to Zipf's Law, there is a fairly strict relation between the rank order of a city (1 = New York, 2 = Los Angeles, etc.) in terms of its population and its actual population. In particular:

$$Population \propto Rank^p$$

where  $p = -1$  and  $\propto$  means "proportional to." This law predicts that if the log of population is plotted against the log of rank, then the data will be well fit by a straight line, and furthermore, the slope of that line should be -1. Show what this plot looks like. Does a straight line fit the log transformed data well (you could use the "lm" procedure as one way to find the best fitting linear relation between variables)? Is the slope of that line approximately -1?

13. Which of these 200 cities would you most like to live in, and why?

**For me, I firmly believe that once someone emerges into the "real world," it's extremely beneficial to move to the closest and biggest city in order to grow as a young person. Over the Summer, I worked for Abbott and I realized just how much a person can gain from living in a city. Although personally I find myself being less gregarious when left to my own devices, Chicago seems like a place that is too obviously beneficial for me to pass up following my graduation.**