Tanmay Sonawane
DSC 101
12/12/22
Donghui Yan

Analysis of Annual Rainfall in Massachusetts

Table of Contents:

Purpose:
To study rainfall patterns (spatial variation) in the state of Massachusetts by creating maps and graphs using the programming language R.

Abstract:

In this report, annual rainfall data in Massachusetts was examined and any trends/ patterns found in the graphs were reported. Precipitation data was collected from the website mass.gov (an official website for the commonwealth of Massachusetts). The data was given for several years between the middle 1800s to 2019, however, data only for the year 2018 was used. The data contained the amount of rainfall (in inches) reported by all the weather stations for every month of 2018 in certain cities in Massachusetts. Using this data, the average rainfall for cities in Massachusetts in the year 2018 was calculated. This was then used to calculate the average rainfall for all the counties. All the calculations were performed in Excel sheets.

R, a programming language, was used to display the rainfall data on a county map of the United States zooming in only on Massachusetts. Data showed that nearby counties had similar amounts of rainfall. It showed that counties like Hudson and Hampshire had the heaviest rainfall in 2018 whereas counties like Barnstable and Dukes had the lowest rainfall in that year. A bar graph was made using the average rainfall per month for 2018. Data in the bar graph displayed that rainfall was the highest in Massachusetts around the months of March and November. It also showed little variance in the amount of rainfall throughout the year. Several uses of this data were found. This data and the graphs could be used to predict the weather/ rainfall pattern for future years. Predicting rainfall a year ahead would be useful for people planning a vacation, for farmers who have to plan when and where certain types of crops they can grow. Similar data could be used to predict what time of the year and in which regions floods would occur due to heavy rainfall, allowing people to prepare for such events well in advance. Physicists wanting to see astronomical phenomena would also use this data for future planning.

A map was used to see geographically how close counties were and to explore the relation between this distance between counties and the rainfall they received.

Limitations:

Some of the limitations were that R did not contain abbreviations of counties and therefore county labels had to be added manually. The limitations in data collection were that the excel sheet only had data up until September of 2019, therefore data for 2018, the next most recent year, that contained data for all months January through December, had to be chosen to be used. Data on mass.gov on rainfall for more recent years was only given for certain regions in the state, not all. Another limitation was that every city in Massachusetts did not have a weather station and hence the rainfall data for that particular city was not provided. Therefore the map had to be a county map and not a city map. Mass.gov was used due to its trustworthiness. Other sources either asked for money to see the data or were less trustworthy.

Introduction:

Rainfall occurs every year in the US. Several weather stations record many types of data about the weather including the amount of rainfall for its various uses (mentioned briefly in 'abstract' and in detail in 'discussion'). In this study, the average rainfall for the year 2018 for each county in the state of Massachusetts was reported and analyzed.

Methodology:

	A	B	C	D
1	fips	state	county	rainfall
2	25001	MA	Barnstable County	4.32
3	25003	MA	Berkshire County	5.43
4	25005	MA	Bristol County	5.62
5	25007	MA	Dukes County	4.32
6	25009	MA	Essex County	4.62
7	25011	MA	Franklin County	5.2
8	25013	MA	Hampden County	5.62
9	25015	MA	Hampshire County	5.61
10	25017	MA	Middlesex County	4.71
11	25019	MA	Nantucket County	4.13
12	25021	MA	Norfolk County	4.32
13	25023	MA	Plymouth County	5.38
14	25025	MA	Suffolk County	4.32
15	25027	MA	Worcester County	5.53
16				

(For the expanded data set see references (1))
Data was obtained from Mass.gov in the form of an excel sheet. The website says that the stations in Massachusetts use the online tool NOWData to record weather changes. The data set is available to the general public and is used to find data on monthly weather conditions, precipitation levels, and drought levels. The excel sheet contained multiple columns starting with, 'stations' which had the name of the weather station that collected the data, 'year', which was 2018, a column for each of the months of the year, and one column with the name of the city where the weather station is based. The last two columns showed what region the station was in and its basin name.

Procedure:

To be able to plot the collected data accurately onto a map, I explored the package tidycensus in

R for data collection. ACS (American Census Survey) was a dataset inside of tidycensus. ACS was already provided in R which contained data provided by the american census survey. It included many statistics for example geography, homelessness rate, unemployment rate, population, etc. However, I was unable to find data on rainfall in that package.

Next I explored the geom_sf function in R. This function also comes with preloaded geographical data from the US Census. Using this I was able to plot the map for US cities in a state. The issue however, was that I could not insert my own data in it. US Census data contained within it did not have rainfall data and I could not find a way to put in my own, and hence I had to disregard the idea.

Using the excel sheet, I calculated the average rainfall for each month for 2018 in the state of MA. This was put on a bar plot (see Figure 1) (to see where the code was borrowed from, see references (3))

For the map, using the excel sheet, the rainfall for every city was averaged. This was done for all the cities which contained at least one weather station. Then, looking at a map of the counties in Massachusetts, the cities and their information were rearranged according to the county they were in. Then the average rainfall for each county was calculated by averaging the rainfall of all the cities in that respective county.

To plot this on the county map of the US, code was borrowed from github's website (see references (2)). To understand how the code worked, the data it took in had to be looked at. I used the function head(data file name) to see the contents of the file. This was helpful and using it I could make my own identical data file containing rainfall data.

The packages usmaps and ggplot2 were loaded into R. Then data had to be arranged a certain way for R to be able to properly read the file. The data set made in excel contained fips code in one column (so that R knows which county certain data is being assigned to). It has the state in the next column, the county name which corresponds to its fips code, and finally the average rainfall was in the last column. This file would not be read by R as an excel file and so it had to be saved as a .csv (comma separated values) file. Then, using R studio, the file was imported into R. In the code, red color was assigned to higher rainfall values, and yellow color was assigned to places where the average rainfall was low. A title was given to the graph.

R code:

Map:

(See References (2))

```
CountyFipsPopData <-
read.csv("C:/Users/tanma/Downloads/CountyFipsPopData.csv")

library(usmap)
library(ggplot2)

plot_usmap(data = CountyFipsPopData, values = "rainfall",
```

```
include = c("MA"), color = "black")
+ scale_fill_continuous(low = "yellow", high = "red",
name = "Legend", label = scales::comma)
+ labs(title = "Massachusetts Annual Rainfall",
subtitle = "Annual rainfall for MA counties in 2018")
+ theme(legend.position = "right")
```

CountyFipsPopData is the data file containing fips code for counties, their names and their respective average rainfall. The packages us map and ggplot2 were loaded for the map to load. 'Values = rainfall' takes in the numbers in the rainfall column and plots them onto the county map of Massachusetts. The line 'include c = ("MA")' is so that it only displays Massachusetts

Barplot:

(see References (3))

```
numbers = c(3.66,    3.29,    4.03,    3.87,    3.72,
          3.78,    3.70,    3.88,    3.83,    3.87,
          4.12, 3.93);

colors = c("green", "green", "green", "green",
          "green", "green", "green", "green",
          "green", "green", "green", "green");
cats = c("JAN","FEB", "MAR", "APR", "MAY", "JUNE",
      "JULY", "AUG", "SEP", "OCT", "NOV", "DEC");

barplot(numbers,names.arg=cats, col=colors, cex.lab=1.5,
      cex.axis=1.5, cex.names=1.5, font=2, font.main=3)
```

Results:

Figure 1: Annual rainfall for MA counties in 2018 (in inches).
From this graph it can be seen that the rainfall increases as one moves away from the shore. Rainfall is the lowest (comparatively) in the east, in counties like Barnstable and highest (comparatively) in the west in counties like Hampshire. The highest recorded average rainfall in any county is close to 5.6 and the lowest 4.3. All the neighboring counties, namely Berkshire, Franklin, Worcester, Hampshire, and Hampden have very similar rainfall patterns. One reason why the rainfall is low at the coastline and high in the center and the west of the continent could be that the coast has beaches where it does not rain much. Whereas other parts of the continent would have more trees and mountains which would cause higher amounts of rainfall. As the water over the oceans evaporates, the winds would carry it towards the center of the continent where it turns into rain water.

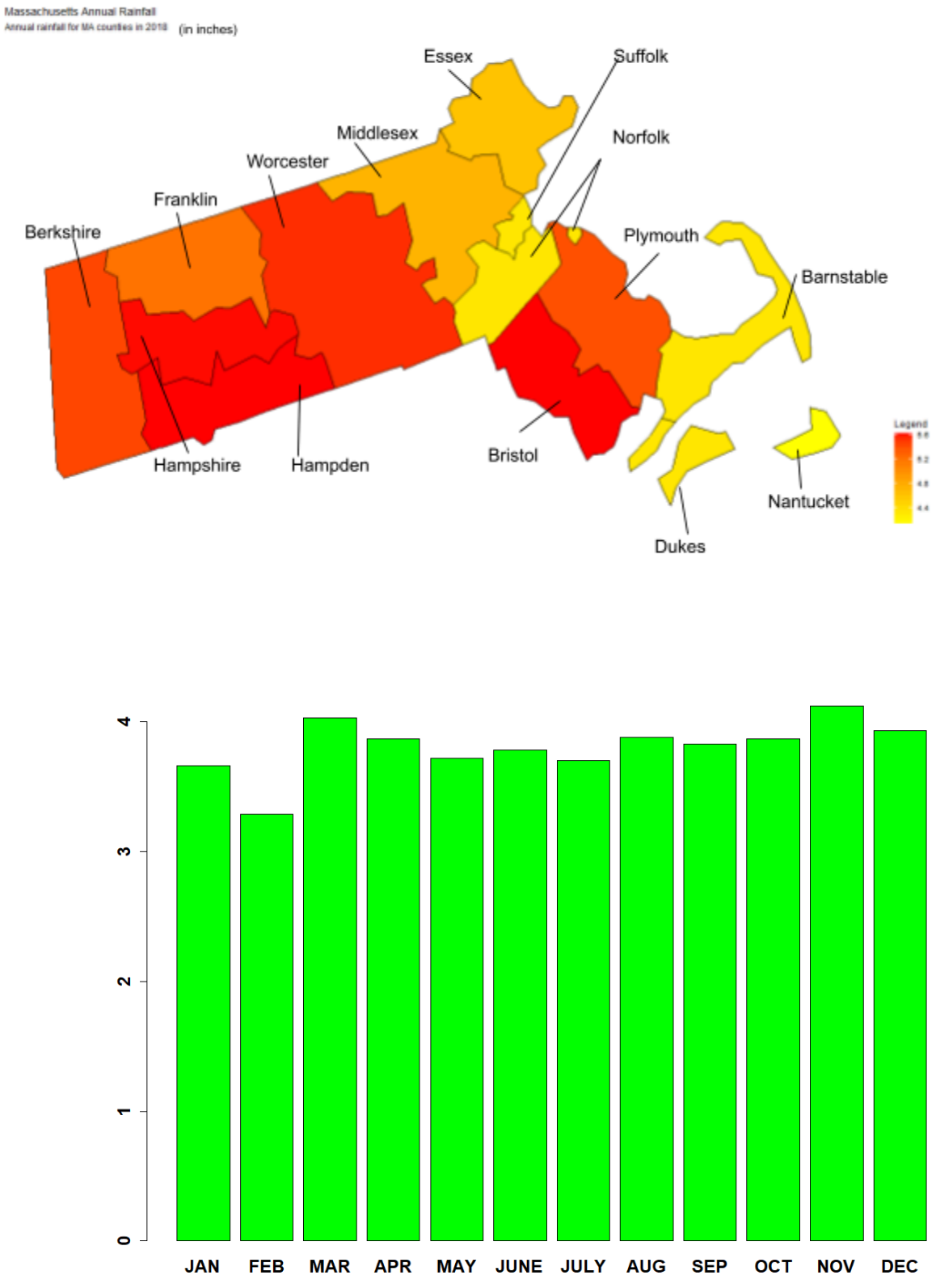
Massachusetts Annual Rainfall
Annual rainfall for MA counties in 2018 (in inches)
Essex
Suffolk
Middlesex
Norfolk
Worcester
Franklin
Berkshire
Plymouth
Barnstable
Hampshire
Hampden
Bristol
Nantucket
Dukes
Legend
0
1
2
3
4
JAN
FEB
MAR
APR
MAY
JUNE
JULY
AUG
SEP
OCT
NOV
DEC

Figure 2

Interpreting the bar graph, one can see that the average rainfall does not differ much per month. It is stable throughout the year. Despite there being three different seasons, winter, spring, and summer, it rains evenly in each season in Massachusetts. The exceptions would be the months of January and February where rainfall lessens. The rainfall is the lowest in February and highest in November.

Discussion:

This data is more important for Massachusetts since the weather is unpredictable. It gets hot or cold anytime and therefore having a system that predicts the general climate well in advance is very helpful.
Thunderstorms and floods often take many lives. The data in this study (when combined with other data) will be helpful in predicting these thunderstorms and floods in advance, saving lives. People would clear out themselves before the flood is expected to approach. This would require very accurate predicting since relocating a large group of people, even temporarily, would cost a lot of time and money.

This data can also show which areas might be experiencing water shortage or even droughts due to low amounts of rainfall. It would enable us to prepare in advance to send water to places with water scarcity. For example, if we see that in some months the weather gets very hot, we would store rainwater in the months prior to that month. Eastern parts of the Nantucket do contain drought regions and these areas in particular might get considerably dry in which case water would have to be shipped to those locations.

Data presented in this study is useful for farmers. Most of the farmland in Massachusetts is in counties like Worcester, Franklin, Berkshire, and Plymouth. According to the graphs, all these counties receive plenty of rainfall throughout the year. Using this data they could properly plan where they can plant certain crops. Crops like beans, beets, and carrots that require more water, could be planted in counties like Worcester where rainfall is abundant. Herbs and crops like cow beans and lima beans could be grown in Dukes and Barnstable, where rainfall is less frequent.

Another use: Cloudy weather prevents physicists from observing astronomical phenomena. They could use the data in this study to predict the likeliness of a cloudy sky during a given month and plan their observations accordingly.
Data in this research can also be used by the general public when they are planning for a trip months in advance. Although this data by itself would not be sufficient, when it is combined with other data, it will be useful in accurately predicting future weather, like the likeness of rain or snow.

Conclusion:

Rainfall data helps predict many future weather related phenomena, which aids in planning ahead, not only for the common person but also for people like farmers and physicists. The map and barplot provided valuable insight into Massachusetts rain patterns during the year 2018. We saw that rainfall was heavy in the western counties and low in eastern counties. We also saw that rainfall in Massachusetts had little variation throughout the year.

References:

(1) Commonwealth of Massachusetts. (n.d.). *Precipitation Data*. Mass.gov. https://www.mass.gov/info-details/precipitation-data
(2) *Chapter 41 Different Ways of Plotting U.S. Map in R | Community contributions for EDAV Fall 2019*. (2019, December 13). https://jtr13.github.io/cc19/different-ways-of-plotting-u-s-map-in-r.html
(3) https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnx1bWFzc2Rkc2MxMDF8Z3g6NGJmNzljNzZlMjFhZDIyOA
(4) *Acres – Counties*. (2020, January 27). Center for Agriculture, Food, and the Environment. https://ag.umass.edu/resources/massachusetts-agricultural-data/land-in-farms/acres-counties
(5) *Watering and Irrigation of Fruits and Vegetables*. (2010, December 5). GrowVeg. https://www.growveg.com/guides/watering-and-irrigation-of-fruits-and-vegetables/
(6) Ly, L. (2019, July 18). *Vibrant Gardens Using Drought Tolerant Vegetables*. Gilmour. https://gilmour.com/drought-tolerant-vegetable-garden
(7) *Spatial data in tidycensus*. (n.d.). https://walker-data.com/tidycensus/articles/spatial-data.html

Further Studies:

If given more time there are many ways in which I would want to expand this study. Firstly, I would compare the average rainfall data for each year for the past 100 years to see any trends. I would like to see how the climate has changed and how the amount of rainfall different counties received has changed. I would also like to compare rainfall in Massachusetts to rainfall in other states. Most important of all, I would like to compare rainfall data with data on snow, humidity, water vapor, and many other weather phenomena to have a complete picture of the weather in Massachusetts.

Appendix:

Precipitation: rain that falls to the ground
R code: R is a programming language that is used in fields like Data Science. The code provided can be run in R
Flood: an overflowing of a large amount of water beyond its normal confines, especially over what is normally dry land