

# Practice 1 - Basic concepts of probability.

## Environment statistics

Mayra Cristina Berrones Reyes

September 7, 2020

### 1 Data description

The data used for this work comes from a government page called INEGI [1] that according to their web site is an “autonomous public body responsible for regulating and coordinating the National System of Statistical and Geographical Information, as well as for capturing and disseminating information on Mexico regarding the territory, resources, population and economy, which allows to publicize the characteristics of our country and help decision making” [1]. Here we selected the subject of environmental data to draw some important information with basic concepts of probability.

Among all of the subjects inside the environmental material, we decided to work with the topic of water[3], and see how it is used across all 32 federative states encompassed in Mexico.

We took the tables in “Origen y Destino”[3] to review how the water resources in Mexico are distributed. The name of the columns in the file of destination were displayed as shown in Table 1.

Name of the federative state contains the name of the 32 states in Mexico. Here we only had to use a small shortcut in the Sublime Text app to add the quotation marks to each name. Total, Domestic, Comercial, Industrial, Public service and Pipe are numbers that represent the volume of thousands of cubic meters.

Name of the federative state	Total	Domestic	Comercial	Industrial	Public services	Pipe
------------------------------	-------	----------	-----------	------------	-----------------	------

## 2 Working with the data

For the propose of this practice we used the proگرامing language of R, version 3.4.2. First we took all the columns in the document and turned them in 7 different variable names. Using the variables of *Total*, *Domestic*, *Comercial*, *Industrial*, *Public services* and *Pipe* we constructed a boxplot to show how the numbers are distributed in all of this variables from all 32 federative states. Figure 1 we can see the boxplot of all the columns.

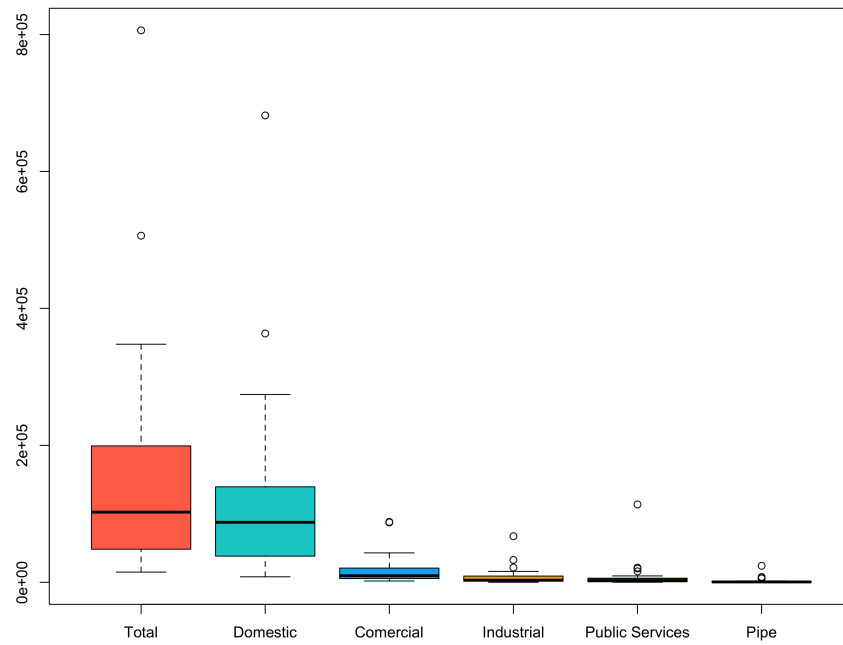


Figure 1

Because the *Total* and *Domestic* values are significantly bigger than the other variables, the rest are barely visible. To fix this, and be able to appreciate better all the variables and outliers, in Figure 2 we change the way we display each box plot, giving each one its individual box.

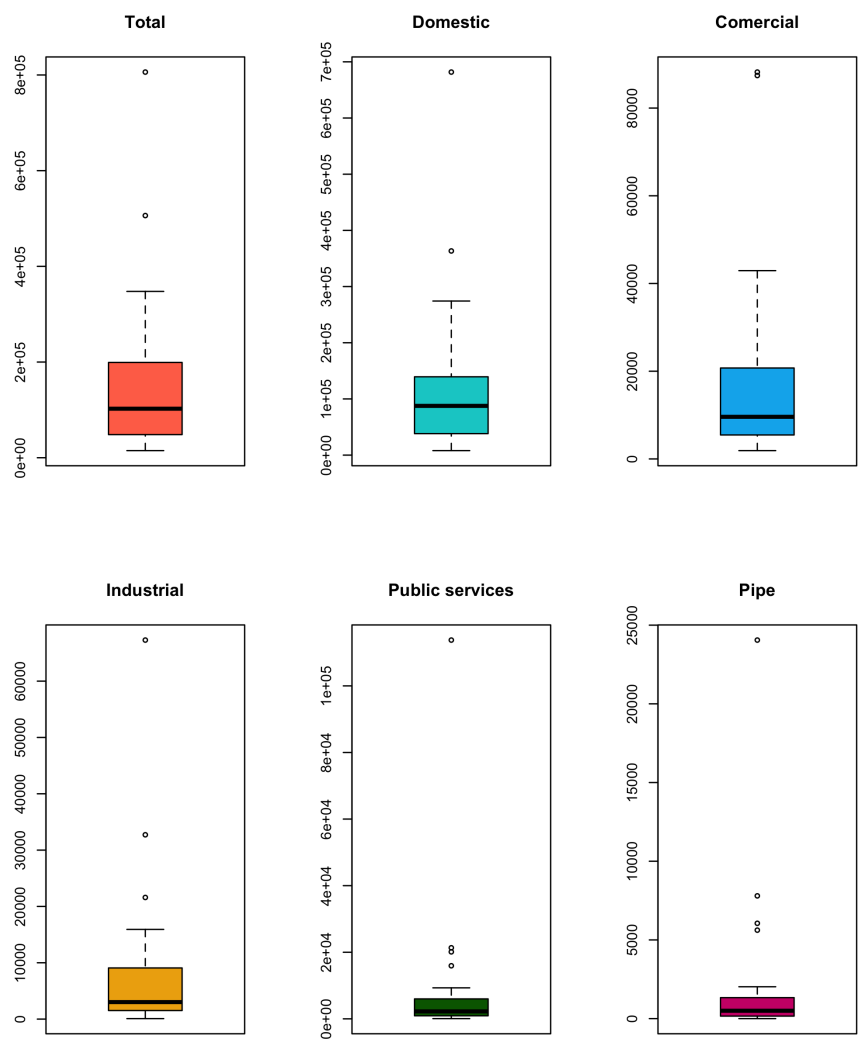


Figure 2

Now that we have a better view of the plots, we can appreciate that the outliers of the *Public services* variable are bigger than *Comercial*, *Industrial* and *Pipe* variables. Knowing this and the big gap between domestic and all the other variables, we want to know which state is the one consuming the most water resources across all of the variables.

To be able to do this, we search in each variable the maximum value using the *max* function, and we save it in another variable as a key number with the function *which.max*, that will allow us to later search it in the variable *States*, which holds a list of all the federative states. In Figure 3 we see a bar graph that show the states that take most of the resources and cause those big outliers.

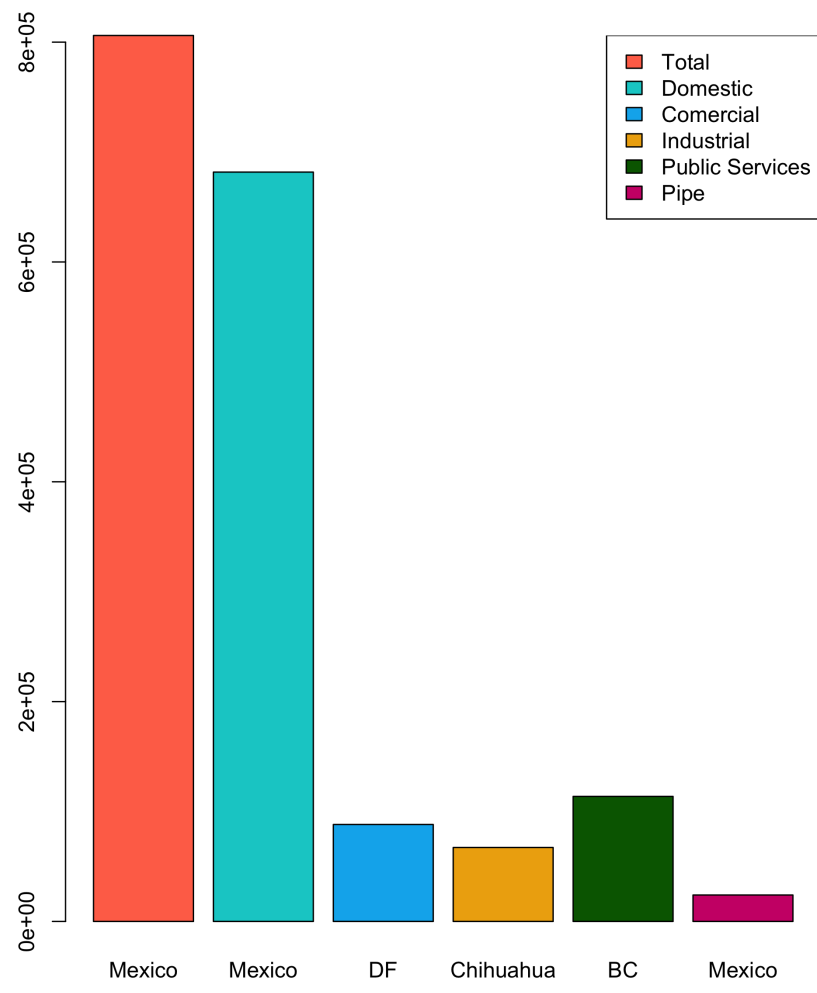


Figure 3

We confirm more visually that the outlier of the *Public services* variable were in fact bigger than the other three.

Mexico city as the place that consumes more water generally and in *Domestic* and *Pipe* aspects comes to no surprise when we examine information outside the given dataset like the population of this state. In the density of population study made by the same organization, we find that the density in Mexico city is 724.2 inhabitants per square kilometer. The second most populated state is Morelos is of 390.2 inhabitants per square kilometer, which is almost half of the population of Mexico city[2].

In the industrial front, Chihuahua as well as other northern states are important points of the industry for Mexico, and in the case of Chihuahua, 95 percent of establishments are immersed in industrial businesses of the state[4]. As for the *Public service*, they do not specify what areas this englobes, so we have shortage of information to make conclusions for this results.

In the Origin file, they only have two columns. The name of the federative state and the volume of water extracted from surface and subsoil mantles represented in thousands cubic meters. With this information, in Figure 4 we put the income of water from all of the surface and subsoil mantles, and we compare it to the *Total* variable, which represents all of the utilization of the water resources in Mexico.

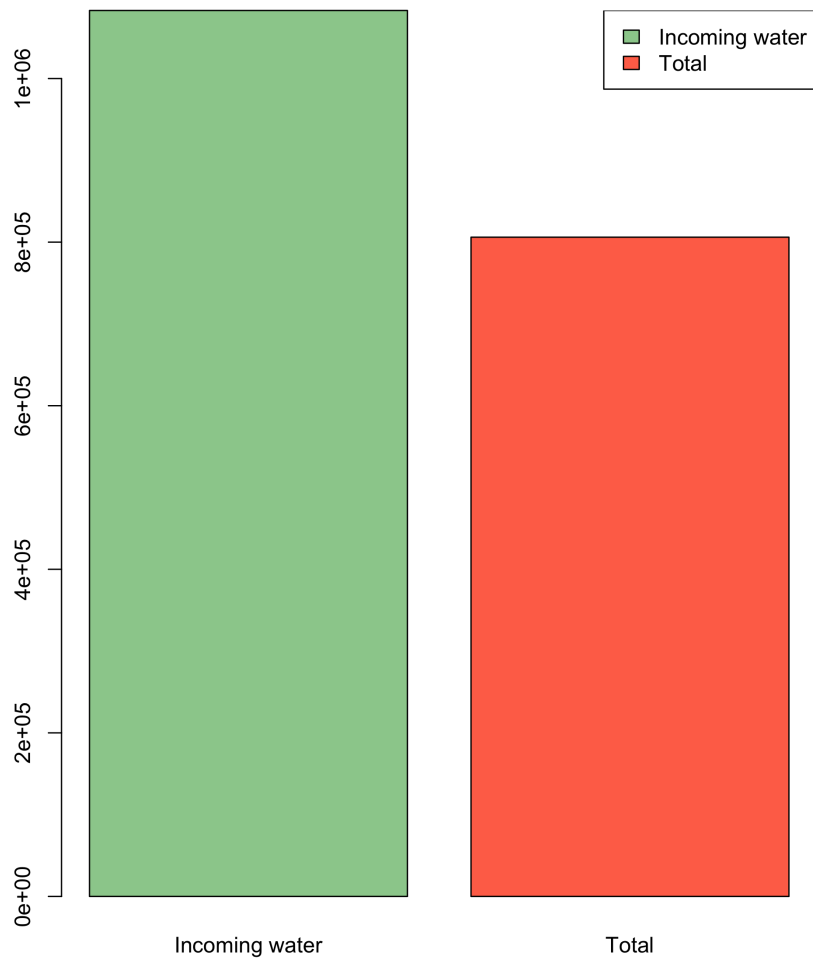


Figure 4: Palindrome example for an array



As we can appreciate in Figure 4 almost a quarter of the income water goes to exportation.

### 3 Conclusions

Choosing the subject of environment was a very quick decision, but now after reviewing the data and the plots made, I see another perspective to a problem faced in my hometown. Here in Tamaulipas, we make use of pipes of water and water tanks, because all year long, except on rainy season, we only have running water very early in the morning. This is, according to my parents, because the two only sources of water that we have are two dams that we share with Nuevo Leon and Texas. Seeing how much water goes to exportation and using it for industrial proposes, now I have a better clue as to why the shortage of water happens in my state.

### References

- [1] INEGI - Quienes somos. <https://www.inegi.org.mx/inegi/contenido/instituto.html#:~:text=Somos%20un%20organismo%20p%C3%ABlico%20aut%C3%B3nomo,nuestro%20pa%C3%ADs%20y%20ayudar%20a> Accessed: 07/09/2020.
- [2] INEGI - Población. <https://www.inegi.org.mx/temas/estructura/default.html#Tabulados>. Accessed: 07/09/2020.
- [3] INEGI - Medio Ambiente. <https://www.inegi.org.mx/temas/agua/default.html#Tabulados>. Accessed: 07/09/2020
- [4] Industria y desarrollo económico y regional. <http://www.chihuahua.gob.mx/attach2/sf/uploads/indtfisc/Plan%20Estatal%20de%20Desarrollo%202004-2010/industria.pdf>. Accessed: 07/09/2020