



Docente: Dr. Gómez Carranza Juan Carlos.

Alumno: Juan Trujillo Estrada.

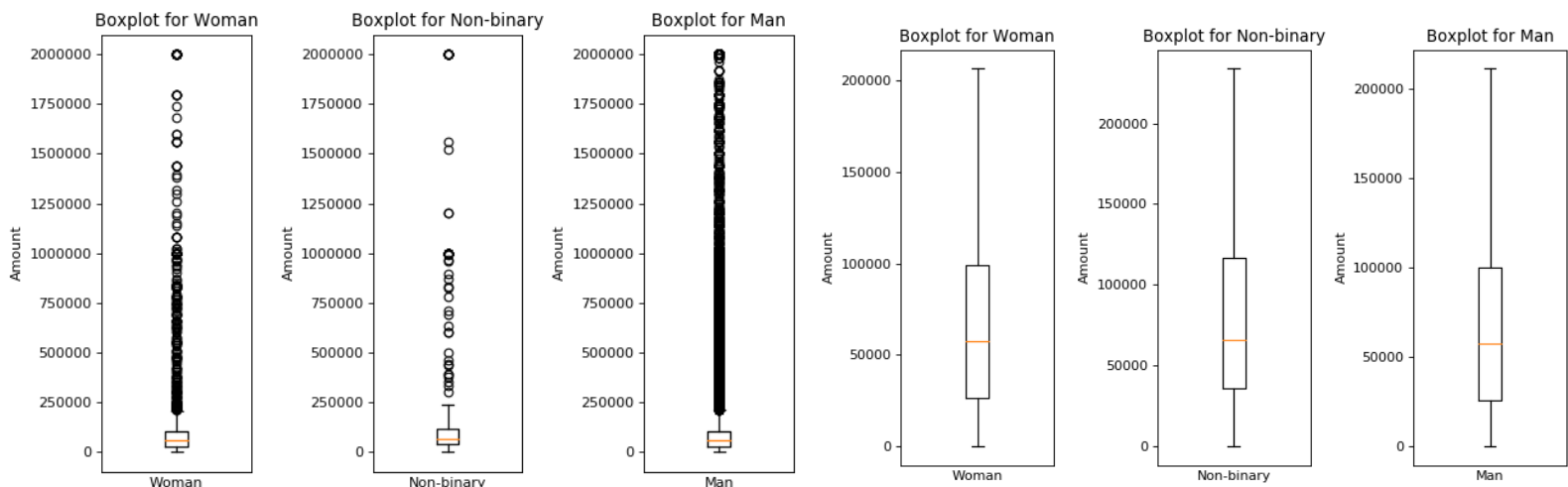
Materia: Minería de Datos.

Asignación: Reporte de práctica 1

The report (a Word file) must describe the analysis of the data and its statistics. Do not describe how you implemented your code but rather describe with your own words what perspectives and/or insights can you extract from the analysis of the statistics. For example: what programming language has a better salary? What developer job is more popular? How many women are registered as back-end developers? Etc

1. Compute the five-number summary, the boxplot, the mean, and the standard deviation for the annual salary per gender.

```
Woman |Minimo = 0.00, Maximo = 2000000.00, Q1 = 26124.00, Q2 = 57287.00, Q3 = 99000.00, Std = 255575.45, Mean = 114066.78|
Non-binary |Minimo = 0.00, Maximo = 2000000.00, Q1 = 36126.00, Q2 = 65653.00, Q3 = 116929.75, Std = 328791.45, Mean = 154890.65|
Man |Minimo = 0.00, Maximo = 2000000.00, Q1 = 25656.00, Q2 = 57254.00, Q3 = 100000.00, Std = 284521.50, Mean = 127346.66|
<Figure size 640x400 with 0 Axes>
```



Del lado izquierdo se encuentra la gráfica sin ignorar los outliers, y en la derecha se encuentra una vista más extendida de lo que es el boxplot, como se puede observar el tope de salario es el mismo para los tres géneros, aunque analizando los cuartiles, el género que tuvo el 3er cuartil mas alto fue el género no binario, el género que tuvo mayos número de personas, fue el masculino eso se logra apreciar en las gráficas donde se ven los outliers, y en conclusión hay una mayor concentración de hombres dentro del sector analizado, pero el sector mejor pagado es el no binario.



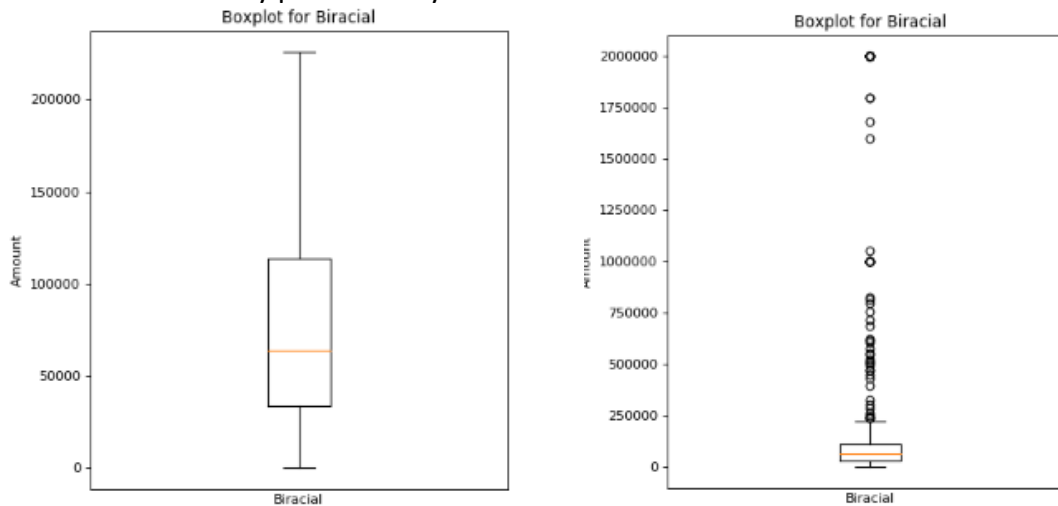
Docente: Dr. Gómez Carranza Juan Carlos.

Alumno: Juan Trujillo Estrada.

Materia: Minería de Datos.

Asignación: Reporte de práctica 1

2. Compute the five-number summary, the boxplot, the mean, and the standard deviation for the annual salary per ethnicity.



En promedio las personas blancas o de descendencia Europea tienen el sueldo más grande, aparte de que su mediana el cuartil dos se encuentra también por encima del resto de etnias, cabe mencionar que claro su mayor cuantitativo es mayor a comparación de las otras etnias, por ejemplo el de las bi-raciales o nativos americanos las cuales son las dos etnias con menor número de personas en nuestras estadísticas, en el caso de los nativos americanos la mediana es mayor que el de la gente blanca, y en el caso de los bi-raciales su mediana es menor por muy poco que la de la gente blanca así que en conclusión la gente blanca por ser mas no es mejor pagada que los demás, por cuestión de media y mediana los nativos americanos son los mejor pagados.



Docente: Dr. Gómez Carranza Juan Carlos.

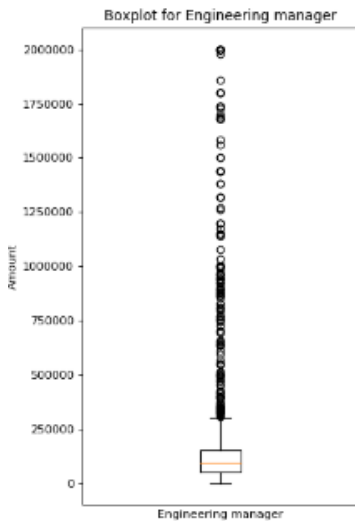
Alumno: Juan Trujillo Estrada.

Materia: Minería de Datos.

Asignación: Reporte de práctica 1

3. Compute the five-number summary, the boxplot, the mean, and the standard deviation for the annual salary per developer type.

El developer full – stack es el devtype que predomina en cuestión de población por persona dentro de los datos recopilados, el Engineering manager es el grupo con la mediana más grande así como el 3er cuartil y una media de casi 180,000 hablando de que su mediana es de 95,000 con 2968 personas me parece que es el sector con mejor paga por persona.



4. Compute the median, mean and standard deviation of the annual salary per country.

El país con la mayor cantidad de personas dentro de nuestro censo es USA, el cual también contiene la mediana más grande, es evidente que puede que este país contenga muchas etnias dentro de sí como ya las analizamos en los problemas anteriores, cabe destacar que su mediana es casi 6 veces y media la de México por lo consiguiente USA es el país mejor pagado en este ámbito.



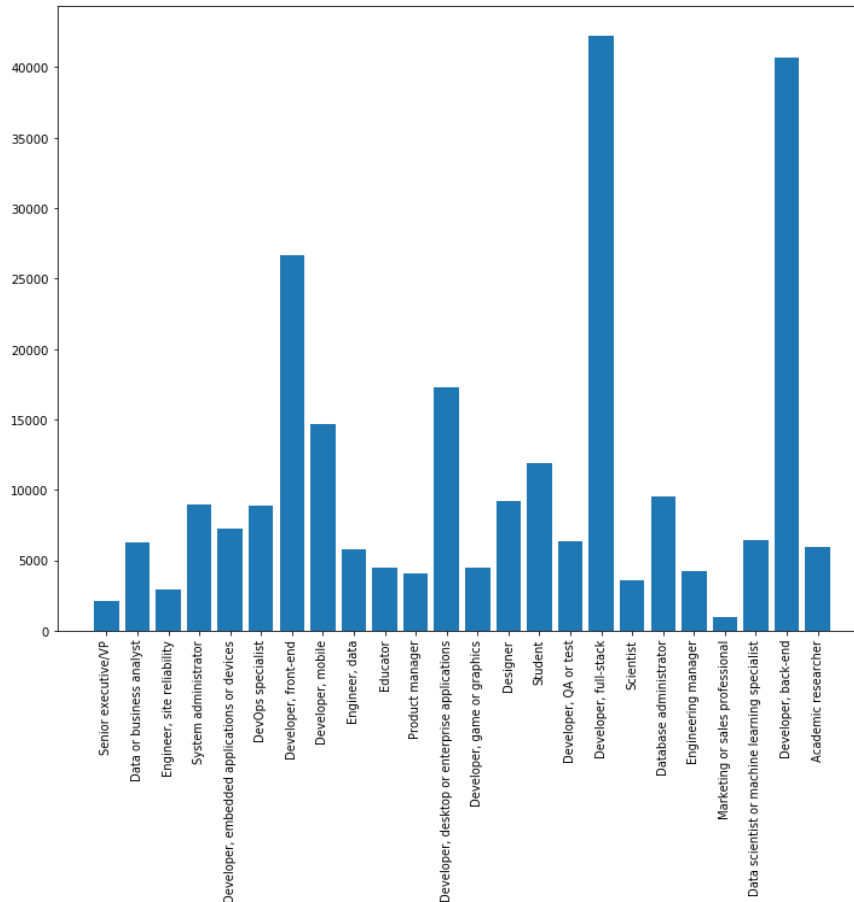
Docente: Dr. Gómez Carranza Juan Carlos.

Alumno: Juan Trujillo Estrada.

Materia: Minería de Datos.

Asignación: Reporte de práctica 1

5. Obtain a bar plot with the frequencies of responses for each developer type.



Como ya se había mencionado anteriormente el developer full-stack es el que se encuentra con mayor numero por consiguiente es la barra más grande dentro de nuestra gráfica de frecuencias.



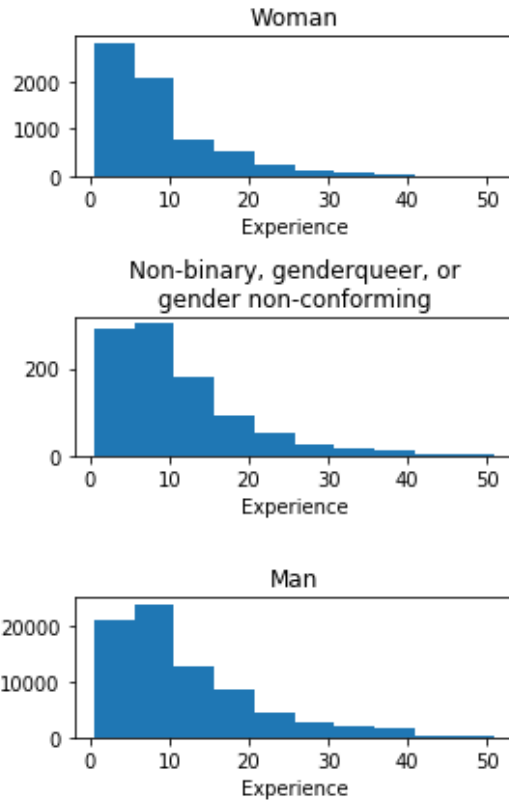
Docente: Dr. Gómez Carranza Juan Carlos.

Alumno: Juan Trujillo Estrada.

Materia: Minería de Datos.

Asignación: Reporte de práctica 1

6. Plot histograms with 10 bins for the years of experience with coding per gender.



El grupo que contiene personas con más experiencia en base a lo mostrado por esta gráfica, es el de los hombres y los no binarios ya que cuentan con cierto número aunque es muy reducido de personas con experiencia de 50 años.



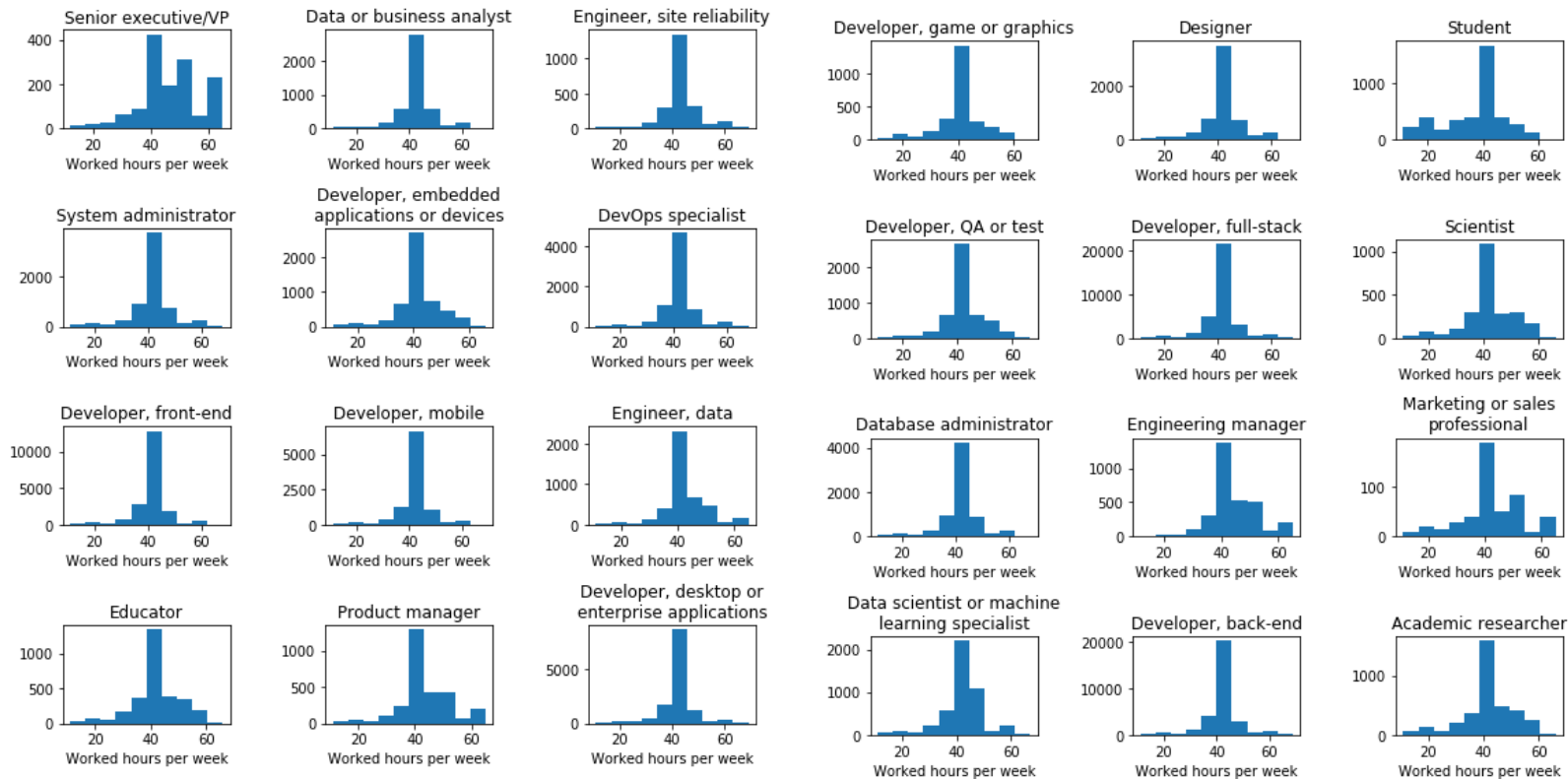
Docente: Dr. Gómez Carranza Juan Carlos.

Alumno: Juan Trujillo Estrada.

Materia: Minería de Datos.

Asignación: Reporte de práctica 1

7. Plot histograms with 10 bins for the average number of working hours per week, per developer



En todas los devtype analizados se tiene un pico en el rango de entre 40-50 horas de trabajo por semana, el tope de horas que utilizamos fueron 70 y el mínimo 10, por lo consiguiente nuestras gráficas estarán delimitadas en esos rangos, la persona que contiene más picos dentro de nuestros rangos delimitantes es el Senior executive/VP el cual tiene un pico de poco más de 200 personas las cuales trabajan entre 60-65 horas, esto nos dice que en teoría los Senior executive/VP trabajan un poco más de horas por semana en promedio que los demás.



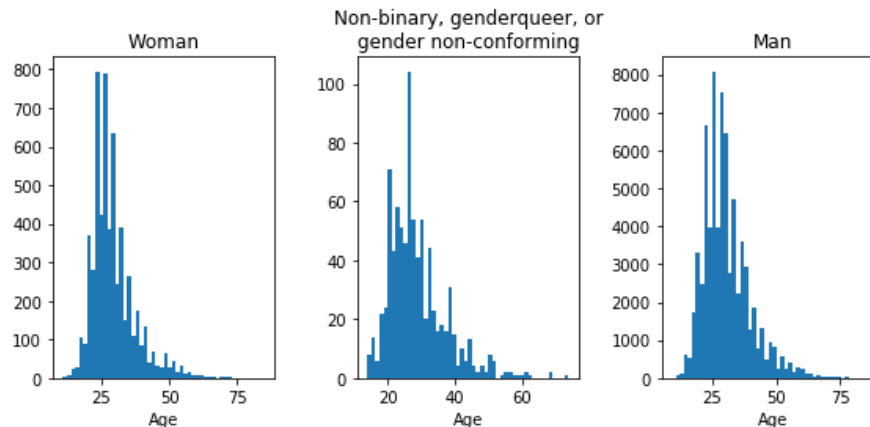
Docente: Dr. Gómez Carranza Juan Carlos.

Alumno: Juan Trujillo Estrada.

Materia: Minería de Datos.

Asignación: Reporte de práctica 1

8. Plot histograms with 10 bins for the age per gender.



En nuestras gráficas se pueden observar diferentes puntos que el mayor número de personas del censo se encuentra entre los 25-30 años en su mayoría oscilan entre esa edad, el género que cuenta con el mayor número de programadores jóvenes es el de los hombres con la mayor cantidad con menos de 25 años, también predomina en el rango de los mayores de 40, en su gran parte se debe a su numeroso grupo de personas.

9. Compute the median, mean and standard deviation of the age per programming language.

El lenguaje con mayor número de adeptos fue Java con una mediana de 28 años y una media de 29 y poquito medio años, lo cual son datos muy parecidos a los de las edades de las personas que programan en Python con una mediana de 28 años y una media de 29 años y medio, los cuales se pueden tomar como lenguajes los cuales son programados por personas jóvenes, en el otro lado de la balanza tenemos web-assembly el cual cuenta con una mediana y una media de 40 años.



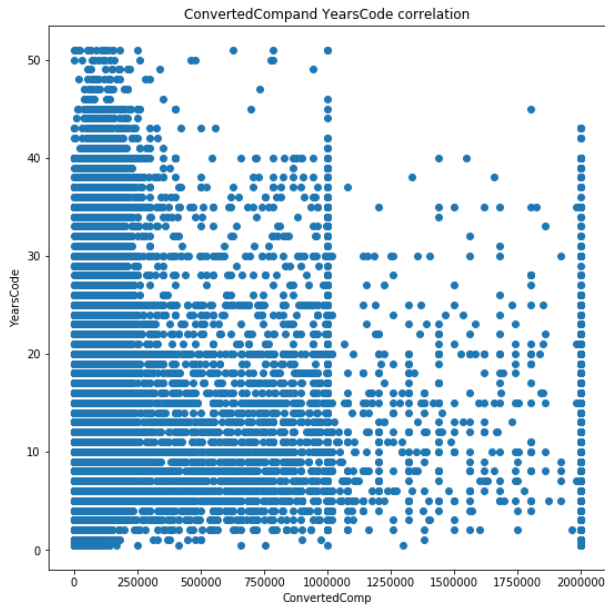
Docente: Dr. Gómez Carranza Juan Carlos.

Alumno: Juan Trujillo Estrada.

Materia: Minería de Datos.

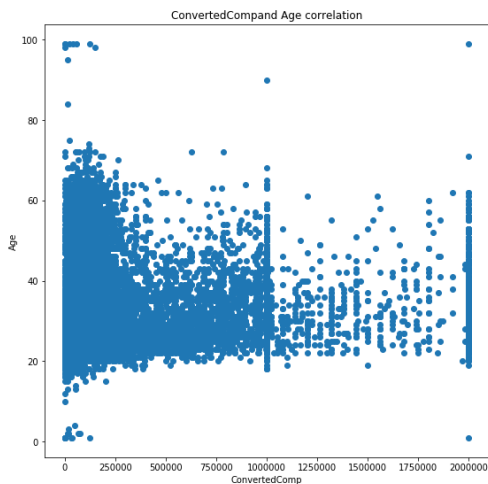
Asignación: Reporte de práctica 1

10. Compute the correlation between years of experience and annual salary.



Lo que nos dice esta gráfica de correlación es que los años de experiencia o los años codificando no te aseguran un gran sueldo debido a que en los puntos con mayores años programando no se encuentra alguien con el salario tope y entre los 0 y 10 años de experiencia se encuentran ciertos puntos en el salario máximo.

11.



Tampoco la edad tiene una relación directa con el salario que vas a adquirir, se puede observar en el 0 absoluto de salario anual a muchas personas las cuales tienen edades mayores a los 50 años que bien se podrían tomar como practicantes.



Docente: Dr. Gómez Carranza Juan Carlos.

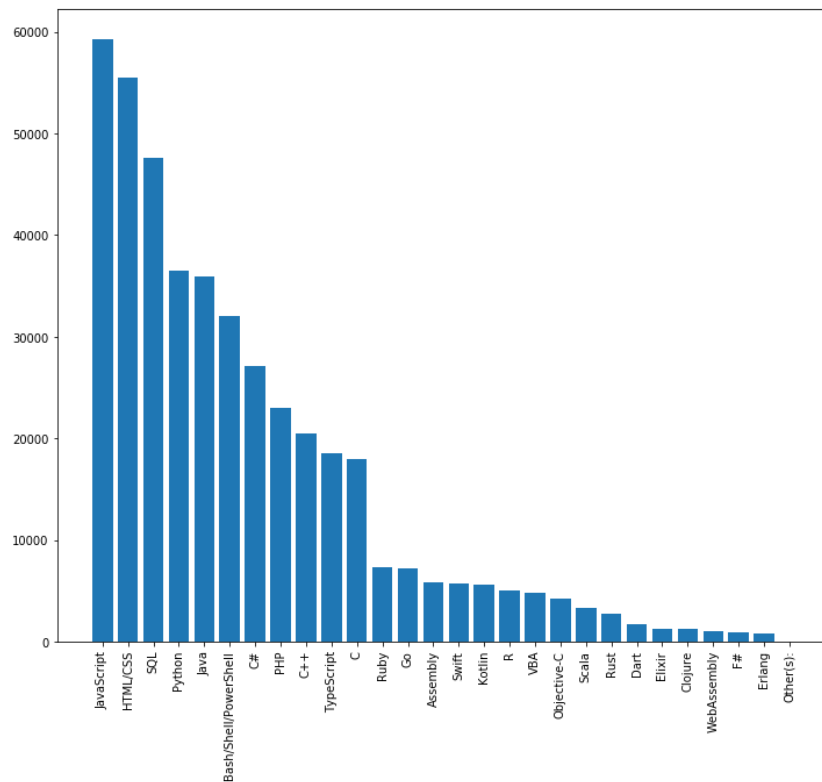
Alumno: Juan Trujillo Estrada.

Materia: Minería de Datos.

Asignación: Reporte de práctica 1

12. No me salió ☹

13. Obtain a bar plot with the frequencies of the different programming languages.



El lenguaje de programación que aparece con mayor frecuencia dentro de nuestro censo es el javascript con poquito menos de 60,000 usuarios o personas que afirmaron usarlo, seguido por HTML/CSS, en los lenguajes con menor frecuencia tenemos F# y Erlang que sinceramente no los conocía.