DIVISIÓN DE INGENIERÍAS CAMPUS IRAPUATO - SALAMANCA

Redes de computadoras I

Práctica 3

Adán Hernández Baena : 146062

April 30, 2020

1   StackOverflow users' data

The data consists of a bunch of some users' information from StackOverflow website. The data is stored on a csv file with **88883 rows** and **10 columns** where each row means a user's answers (the data were collected from a survey) and the columns indicate the questions. The columns headers are Country, Education Level, Developer type, Years of coding, Salary in USD, Worked hours per week, Languages that they have used, Age, Gender and Ethnicity. After a superficial look to the csv, a realized that the answers are too homogeneous, then I was not going to need to do much data cleaning (as lowercase converting, remove excess of blanks values, etc) the "only problem" were the multiple answers and some dispersed null values. As first step I considered the visualization of the data distribution of the salary according other variable (Education level,).

2   Salaries distribution

Observing the salaries respect to developer type, gender, country, educational level, etc, (Figures 1 to 3) do not influence much of the distributions of these, the salaries are concentrated around 40K and 50K USD per year. Therefore, any variable does not influence on salary distribution, you can starve of death or you could earn until 2 millions USD regardeless your education, age, gender, age, etc. Another thing that made me interesting to analyze was whether the amount of programming language that someone has worked with influence in the salaries. The results are shown on Figure 4 to 7, where it could be observed how the more programming languages means in less programmers and the average increase a bit.
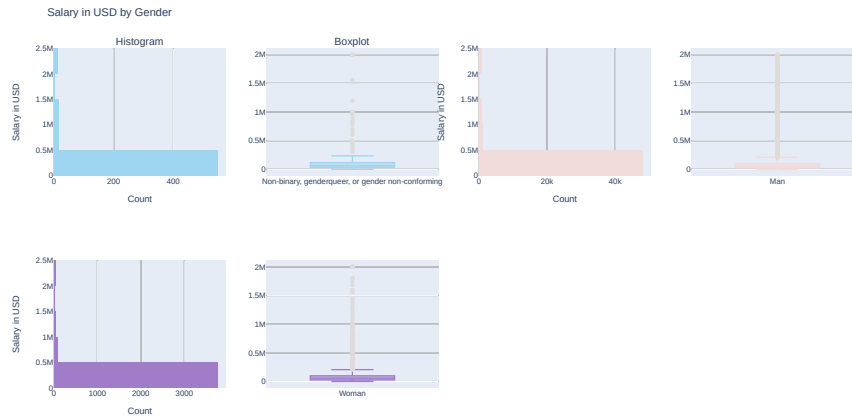

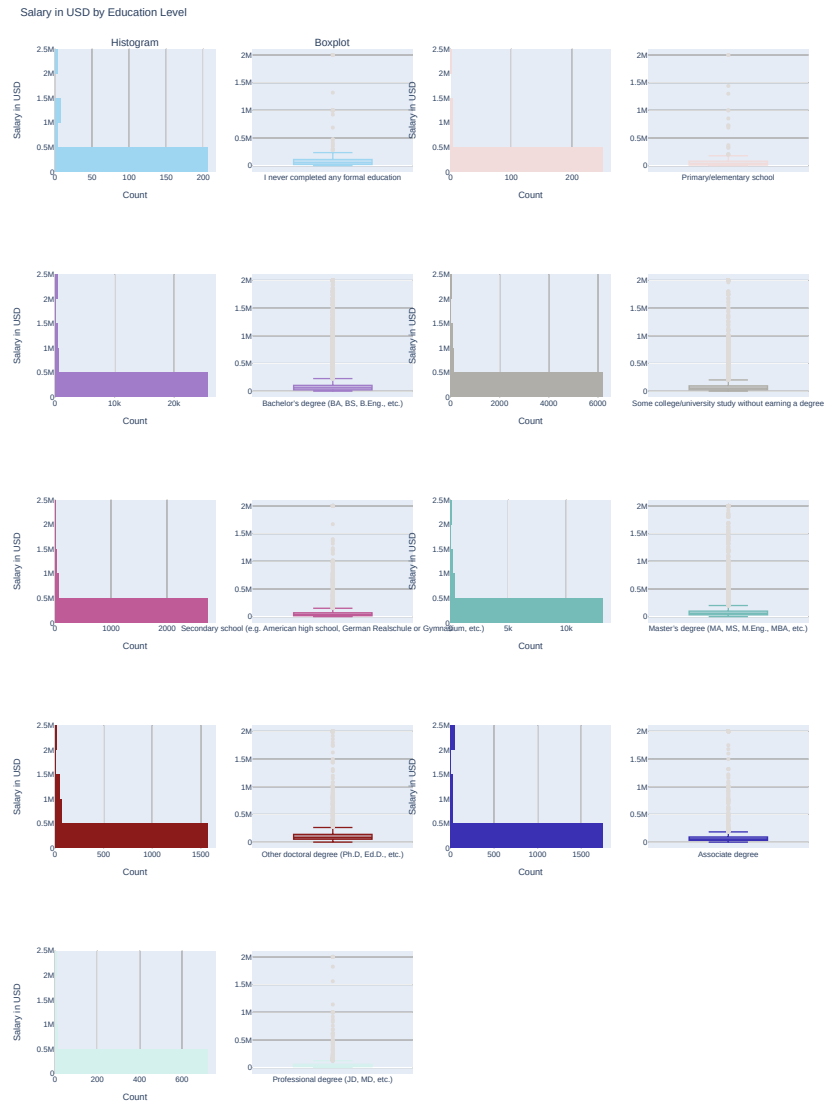Figure 1: Distribution of Salary in USD by Gender

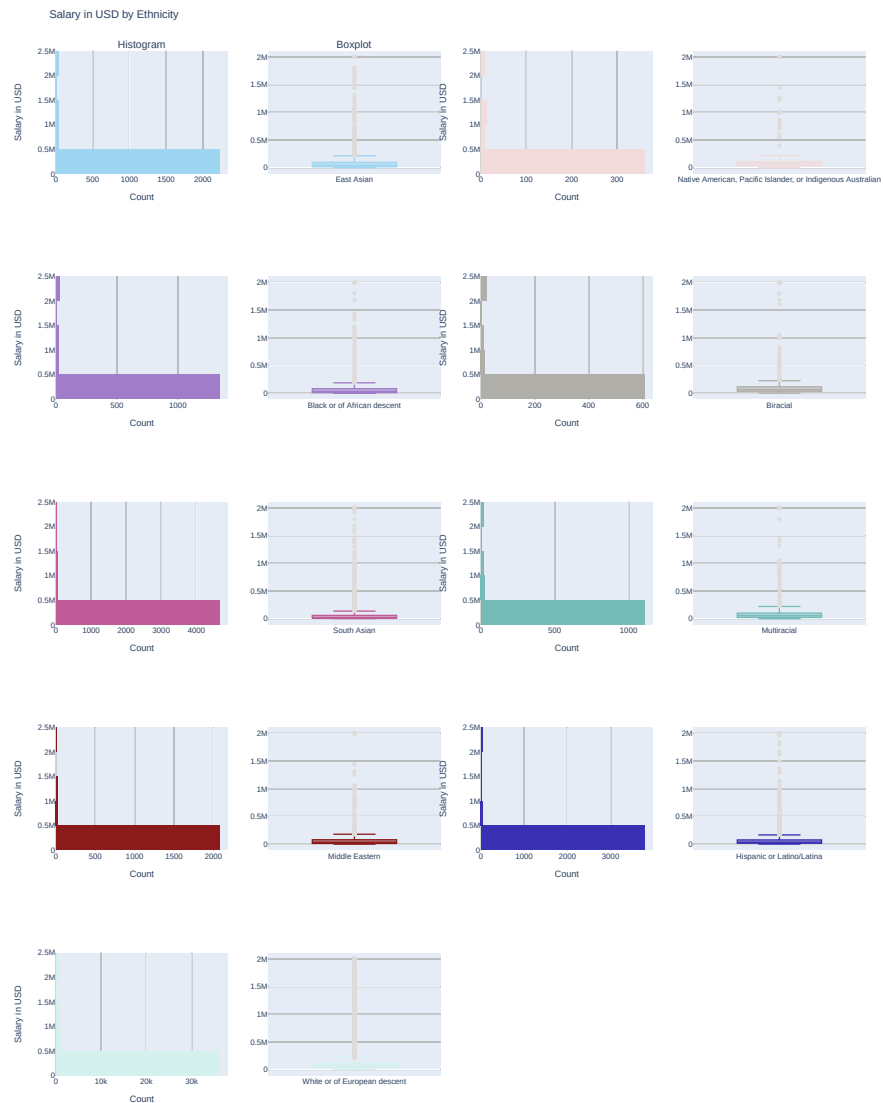Figure 2: Distribution of Salary in USD by Education Level
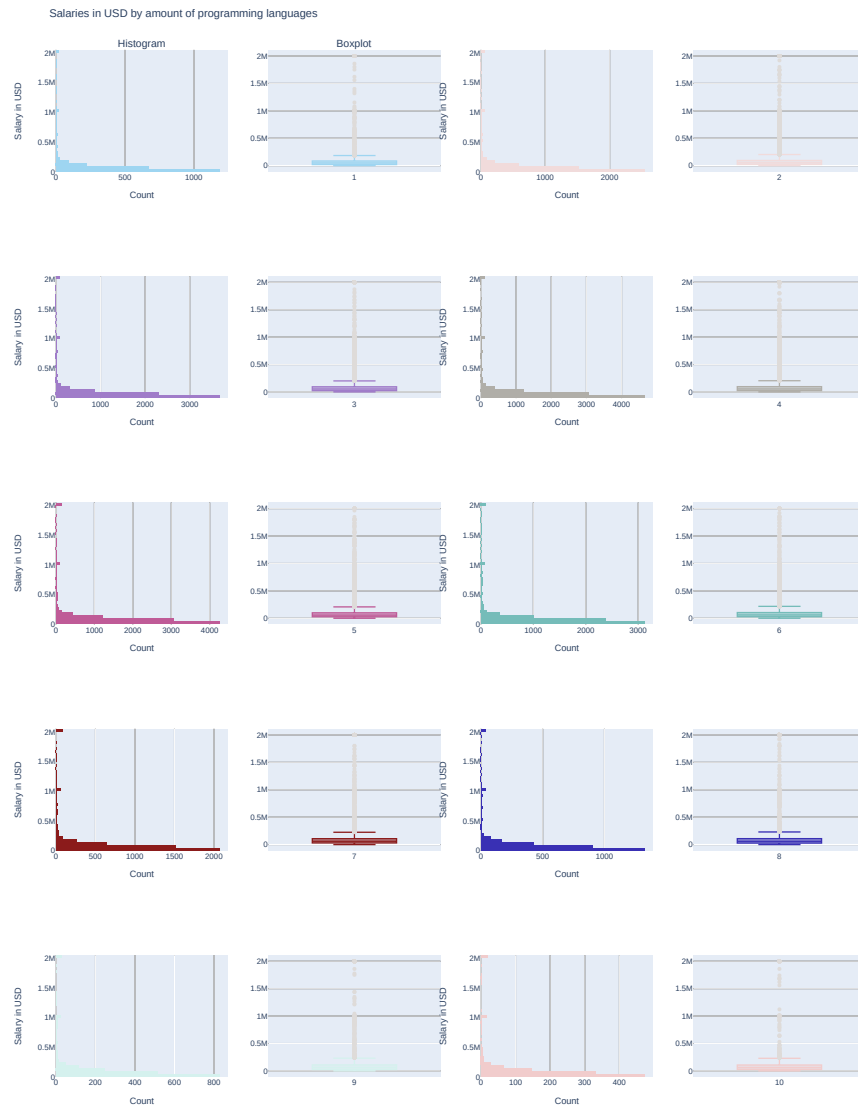
Figure 3: Distribution of Salary in USD by Ehnicity

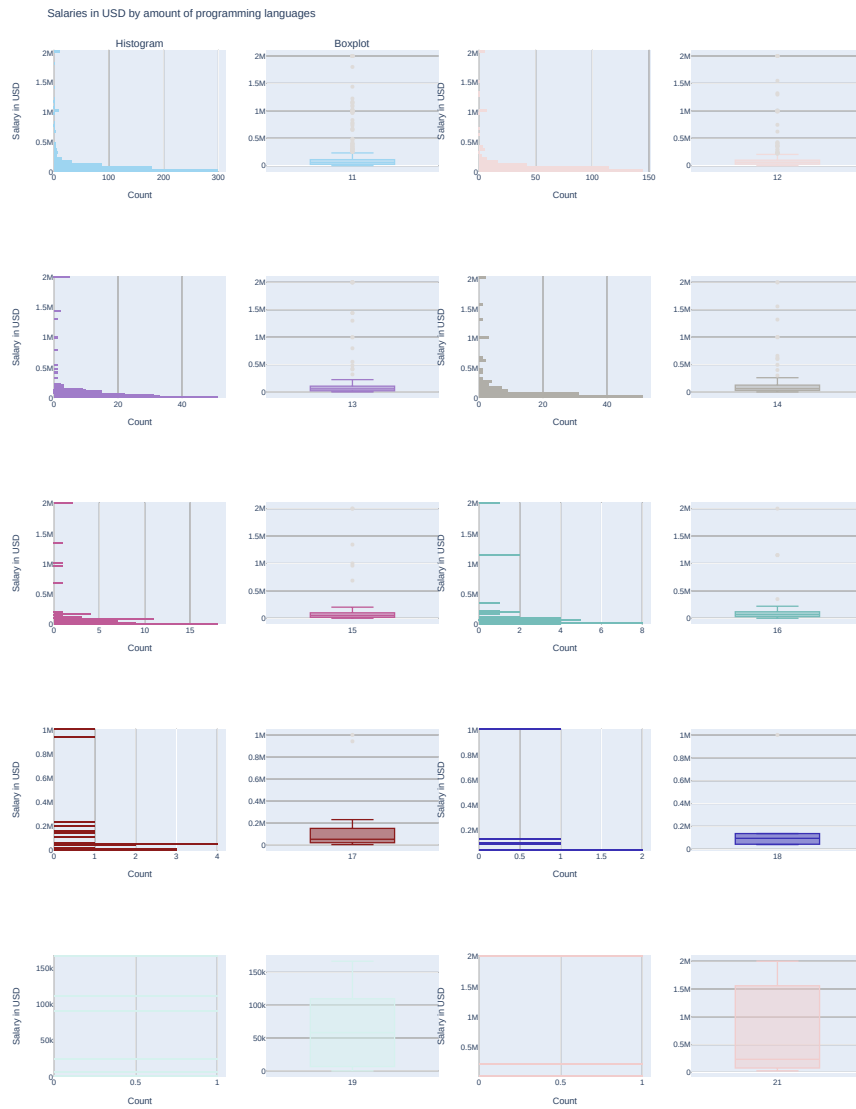Figure 4: Distribution of Salary in USD by amount of languages (Part 1/3)

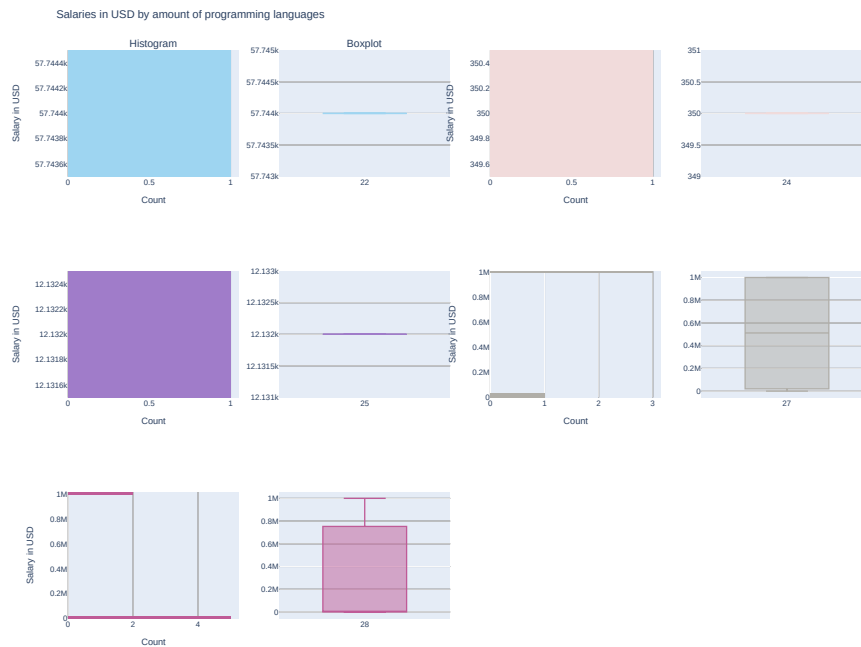Figure 5: Distribution of Salary in USD by amount of languages (Part 2/3)

Salaries in USD by amount of programming languages

Figure 6: Distribution of Salary in USD by amount of languages (Part 3/3)

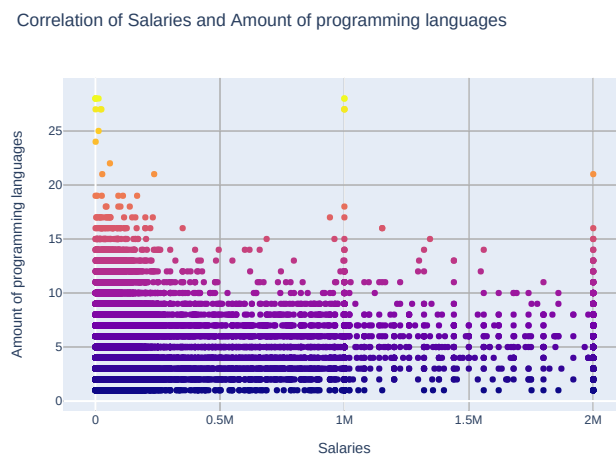Correlation of Salaries and Amount of programming languages

Figure 7: Sactter plot of salary in USD by amount of languages

3   Programmers

Almost 30 different programming languages were captured by the survey where JavaScript head the list (Figure 8) with almost 60K users, closely followed by HTML/CSS with 55K users. Teh five first languages are JavaScript, HTML/CSS, SQL, Python and Java. The languages with the less users are Erlang, F#, WebAssembly, Clojure and Elixir.
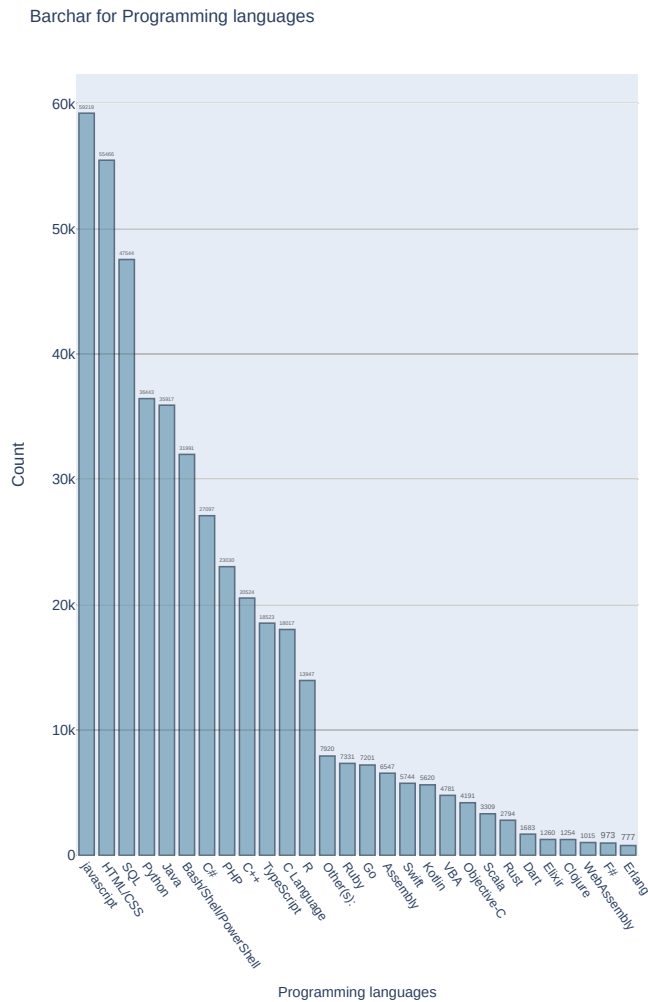

Figure 8: Used programming languages

As an extra analyze I decided to calculate some statistical info about the salaries and the amount of user per programming language. For this calculation

(Figure 9) I decided to delimited the salaries greater than 1k and less than 2 millions to remove a little bit of outliers. The graphic shows a small relation between the salaries and the number of programmers that use certain language. For example, the five programming languages with the less number of users have better earnings that the five top ones. The interesting thing is that Golang with just 10 years of life have programmers with the best salaries above other long-lived languages like Python, Java or C.
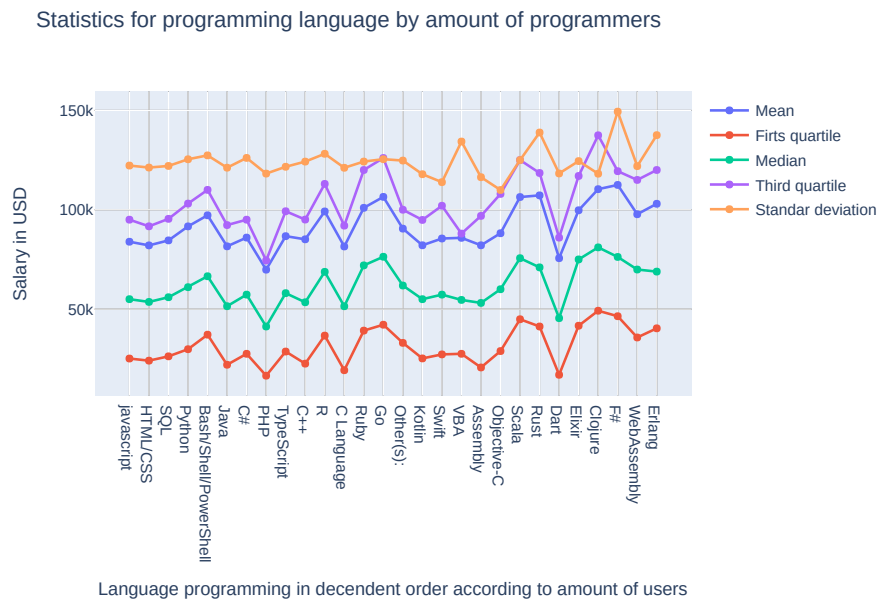
Statistics for programming language by amount of programmers

Figure 9: Statistics per programming languages

4 Education

As I mentioned before, the distribution of the salaries respect to other variables does not change, however, an expected characteristic of educations is shown in Figure 10. This graphic shows the relation of salaries respect teh education level of the programmers and even though anybody can have a job with 2 Millions USD of salary, it is true that the programmers who have less education, their salaries are concentrated much lower than the ones with higher education.

Other thing about education that I have always heard it is that with a degree you will work less, so I also studied that aspect of the education and worked hours per week. Figure 12 shows how hours behave respect the education of the programmers. The graphic (It is delimited between 6 hours per day and 16 hour per day. With this delimitation I expected to focus the analyze in full
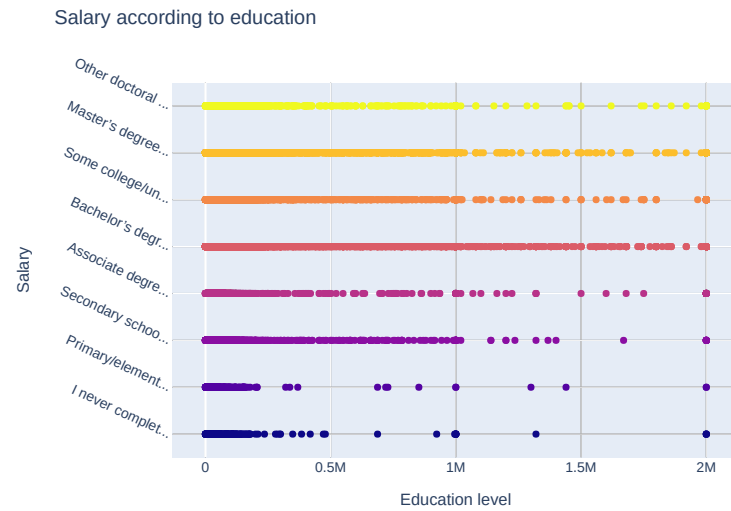
Figure 10: Scatter plot of Salaries respect education level

time programmers or almost full time programmers) support the hypothesis
that with higher education the spent time while programming is less than when
your do not have a degree.
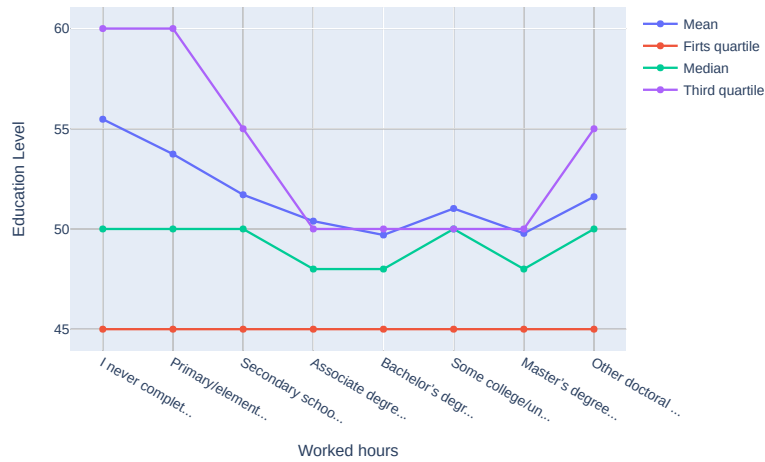
Statistics for worked hours per week by education

Figure 11: Statistics of worked hours respect education level.
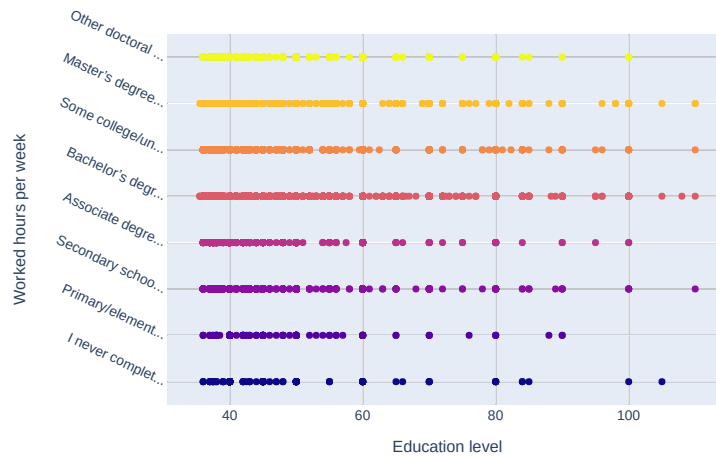
Worked hour per week according to education

Figure 12: Worked hours per week respect education level

5   Gender

The number of non-traditional gender programmers and women programmers are much much lower respect of the number of men programmers (Shown in Figure 13). The interesting thing is that the non-binary gender coders have higher salaries than the men (Almost 7K USD higher as Figure 1 shows). This may be due to the small sample of non-male programmers and how they are spread across all possible wages, from the lowest to the millionaires.
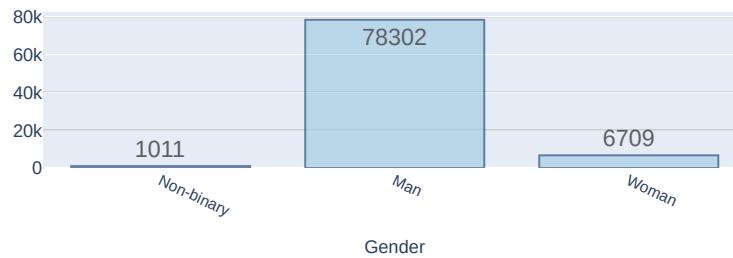
Barchar for Gender

Figure 13: Programmers count by gender
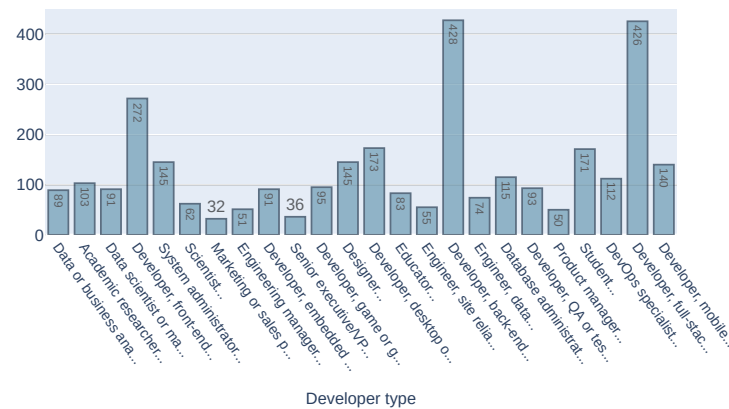
Developer type for non-binary

Figure 14: Developer type for non-binary

11
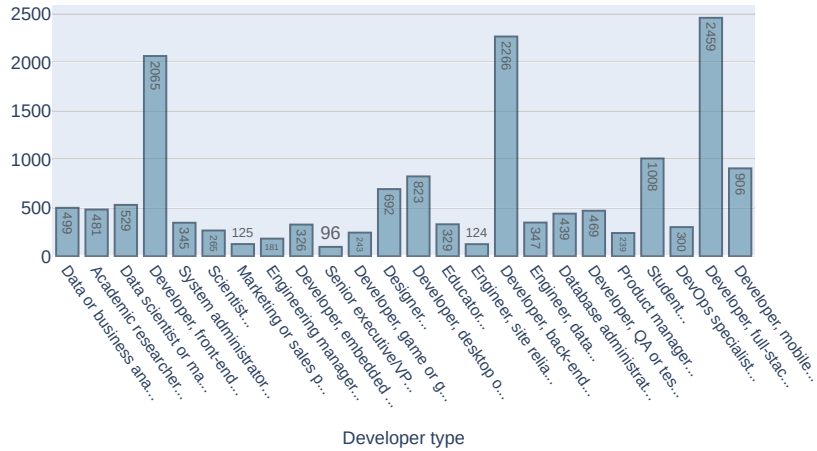
Developer type for women

Figure 15: Developer type for women
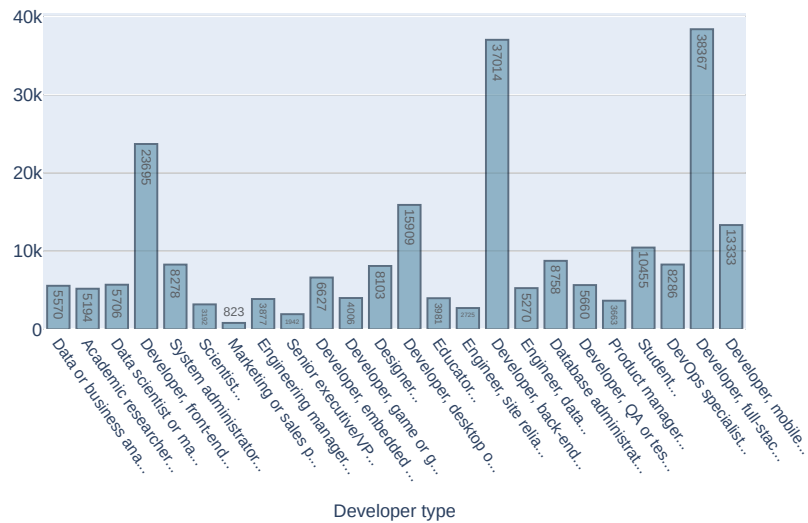
Developer type for men

Figure 16: Developer type for men

Talking about developer types, for the three genders, the web development heads the list of the most popular dev types (Fullstack, backend and frontend) and the less popular are Marketing and senior executive. The not web developers, the more abundant are desktop developer, mobile developer, system administrator and designers. Now, the average age for men is 30 year, for women is 29 and for non binary is 28 (Figure 17).
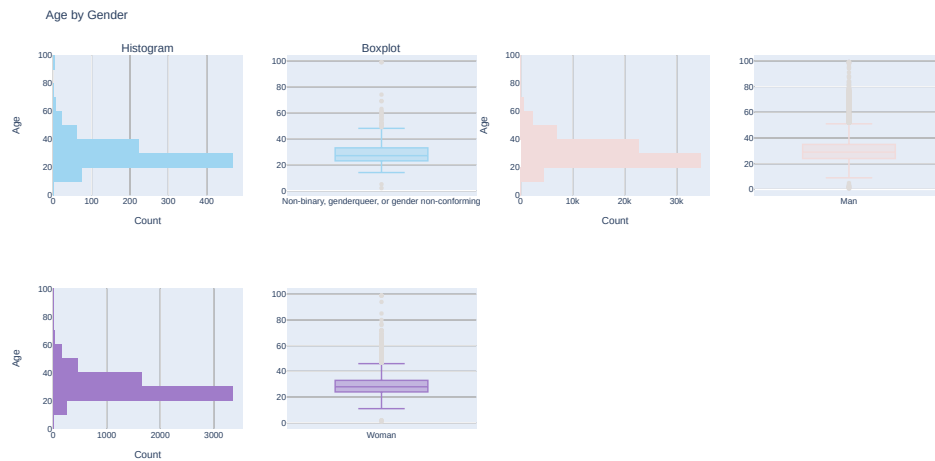

Figure 17: Age by gender