



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

WINTER SEMESTER 2021

PROJECT REPORT
CSE3020 – DATA VISUALIZATION

ANALYSIS OF 2018 STACK OVERFLOW DEVELOPER SURVEY	
TEAM NO: 18	
19BCE0623	KABILAN V

Submitted to
Dr. Archana Tamizharasan
SCOPE
Vellore Institute of Technology, Vellore

Project at a glance

No of objectives considered: 4	
1. General analysis of Coding Habits, Developer Types and Job Status 2. Analysis of Languages, IDEs and Frameworks Used and Coding Experience 3. Salary Analysis on various factors 4. Network Analysis of Developer Types, IDEs and Languages	
Language used: R	
Statistical Measured used:	
Correlation	
Median	
Average	
Library(s) used:	
tidyverse knitr highcharter plotly viridis wesanderson countrycode visNetwork randomcoloR	
Total no of Visuals created:	46
No of Individual Chart types used:	14
Bar chart	12
Grouped Bar chart	5
Pie chart	6
Histogram	2
Lollipop chart	2
Bubble Chart	2
Boxplot	1
Violin Plot	1
Line Chart	1
Area Chart	1
Heatmap	2
Treemap	6
Geomap	2
Network	3
Total Charts in project	46

1.1. Project Statement

To analyse the 2018 Stack Overflow Developer Survey using various visualization methods.

1.2. Project Objective

With this project, I plan to use the knowledge gained during the Data Visualization Course to analyze the 2018 Stack Overflow Developer Survey results using various visualization methods. By doing this project, I hope to implement the various visualization techniques learnt to gain insights and a better understanding of the Developer Community around the world. The aim is to analyze their coding habits, their jobs, their coding preferences and their salaries with respect to a variety of parameters. I believe that this project will help provide a better understanding of the Developer Community around the world using the power of Data Visualization.

1.3. Modules

I plan to do this project in 4 modules

PART 1: General analysis of coding as a hobby, contribution to Open-Source projects, Education, Developer Types and Job Status.

PART 2: Analysis of Language Preferences, Languages to work with in the future, IDE Preferences and Framework Preferences and Coding Experience

PART 3: Salary analysis on different parameters

PART 4: Network analysis on IDE Preferences, Language Desire and Developer Types.

1.4. Code with Visuals

R Notebook

```
library(tidyverse)
library(knitr)
library(highcharter)
library(plotly)
library(viridis)
library(wesanderson)
library(plotly)
library(countrycode)
library(visNetwork)
library(randomcoloR)

survey <- read.csv("F:\\DataViz\\Project\\survey_results_public.csv")
schema <- read.csv("F:\\DataViz\\Project\\survey_results_schema.csv")
```

The datasets have been imported. survey represents the results of the survey schema represents the questions asked during the survey.

Before exploring the results of the survey, let us take a look at the questions asked during the survey.

kable(schema)

Column	QuestionText
Respondent	Randomized respondent ID number (not in order of survey response time)
Hobby	Do you code as a hobby?
OpenSource	Do you contribute to open source projects?
Country	In which country do you currently reside?
Student	Are you currently enrolled in a formal, degree-granting college or university program?
Employment	Which of the following best describes your current employment status?
FormalEducation	Which of the following best describes the highest level of formal education that you've completed?
UndergradMajor	You previously indicated that you went to a college or university. Which of the following best describes your main field of study (aka 'major')
CompanySize	Approximately how many people are employed by the company or organization you work for?
DevType	Which of the following describe you? Please select all that apply.
YearsCoding	Including any education, for how many years have you been coding?
YearsCodingProf	For how many years have you coded professionally (as a part of your work)?
JobSatisfaction	How satisfied are you with your current job? If you work more than one job, please answer regarding the one you spend the most hours on.
CareerSatisfaction	Overall, how satisfied are you with your career thus far?
HopeFiveYears	Which of the following best describes what you hope to be doing in five years?
JobSearchStatus	Which of the following best describes your current job-seeking status?
LastNewJob	When was the last time that you took a job with a new employer?
AssessJob1	Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging

the choices up and down), where 1 is the most important and 10 is the least important. The industry that I'd be working in

AssessJob2

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging the choices up and down), where 1 is the most important and 10 is the least important. The financial performance or funding status of the company or organization

AssessJob3

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging the choices up and down), where 1 is the most important and 10 is the least important. The specific department or team I'd be working on

AssessJob4

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging the choices up and down), where 1 is the most important and 10 is the least important. The languages, frameworks, and other technologies I'd be working with

AssessJob5

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging the choices up and down), where 1 is the most important and 10 is the least important. The compensation and benefits offered

AssessJob6

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging the choices up and down), where 1 is the most important and 10 is the least important. The office environment or company culture

AssessJob7

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging the choices up and down), where 1 is the most important and 10 is the least important. The opportunity to work from home/remotely

AssessJob8

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging

the choices up and down), where 1 is the most important and 10 is the least important. Opportunities for professional development

AssessJob9

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging the choices up and down), where 1 is the most important and 10 is the least important. The diversity of the company or organization

AssessJob10

Imagine that you are assessing a potential job opportunity. Please rank the following aspects of the job opportunity in order of importance (by dragging the choices up and down), where 1 is the most important and 10 is the least important. How widely used or impactful the product or service I'd be working on is

AssessBenefits1

Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Salary and/or bonuses

AssessBenefits2

Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Stock options or shares

AssessBenefits3

Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Health insurance

AssessBenefits4

Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Parental leave

AssessBenefits5

Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is

AssessBenefits6	<p>most important and 11 is least important. Fitness or wellness benefit (ex. gym membership, nutritionist)</p> <p>Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Retirement or pension savings matching</p>
AssessBenefits7	<p>Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Company-provided meals or snacks</p>
AssessBenefits8	<p>Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Computer/office equipment allowance</p>
AssessBenefits9	<p>Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Childcare benefit</p>
AssessBenefits10	<p>Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Transportation benefit (ex. company-provided transportation, public transit allowance)</p>
AssessBenefits11	<p>Now, imagine you are assessing a job's benefits package. Please rank the following aspects of a job's benefits package from most to least important to you (by dragging the choices up and down), where 1 is most important and 11 is least important. Conference or education budget</p>
JobContactPriorities1	<p>Imagine that a company wanted to contact you about a job that is a good fit for you. Please rank your preference in how you are contacted (by dragging the choices up and down), where 1 is the most preferred and 5 is the least preferred. Telephone call</p>

JobContactPriorities2	Imagine that a company wanted to contact you about a job that is a good fit for you. Please rank your preference in how you are contacted (by dragging the choices up and down), where 1 is the most preferred and 5 is the least preferred. Email to my private address
JobContactPriorities3	Imagine that a company wanted to contact you about a job that is a good fit for you. Please rank your preference in how you are contacted (by dragging the choices up and down), where 1 is the most preferred and 5 is the least preferred. Email to my work address
JobContactPriorities4	Imagine that a company wanted to contact you about a job that is a good fit for you. Please rank your preference in how you are contacted (by dragging the choices up and down), where 1 is the most preferred and 5 is the least preferred. Message on a job site
JobContactPriorities5	Imagine that a company wanted to contact you about a job that is a good fit for you. Please rank your preference in how you are contacted (by dragging the choices up and down), where 1 is the most preferred and 5 is the least preferred. Message on a social media site
JobEmailPriorities1	Imagine that same company decided to contact you through email. Please rank the following items by how important it is to include them in the message (by dragging the choices up and down), where 1 is the most important and 7 is the least important. Details on the company I'd be working for
JobEmailPriorities2	Imagine that same company decided to contact you through email. Please rank the following items by how important it is to include them in the message (by dragging the choices up and down), where 1 is the most important and 7 is the least important. Details on the specific department I'd be working for or product I'd be working on
JobEmailPriorities3	Imagine that same company decided to contact you through email. Please rank the following items by how important it is to include them in the message (by dragging the choices up and down), where 1 is the most important and 7 is the least important. Specifics of why they think I'd be a good fit for the role (ex. my prior work history, projects on GitHub)
JobEmailPriorities4	Imagine that same company decided to contact you through email. Please rank the following items by how

	important it is to include them in the message (by dragging the choices up and down), where 1 is the most important and 7 is the least important. Details of which technologies I'd be working with
JobEmailPriorities5	Imagine that same company decided to contact you through email. Please rank the following items by how important it is to include them in the message (by dragging the choices up and down), where 1 is the most important and 7 is the least important. An estimate of the compensation range
JobEmailPriorities6	Imagine that same company decided to contact you through email. Please rank the following items by how important it is to include them in the message (by dragging the choices up and down), where 1 is the most important and 7 is the least important. Information on the company's hiring process
JobEmailPriorities7	Imagine that same company decided to contact you through email. Please rank the following items by how important it is to include them in the message (by dragging the choices up and down), where 1 is the most important and 7 is the least important. Details on the company's product development process
UpdateCV	Think back to the last time you updated your resumé, CV, or an online profile on a job site. What is the main reason that you did so?
Currency	Which currency do you use day-to-day? If your answer is complicated, please pick the one you're most comfortable estimating in.
Salary	What is your current gross salary (before taxes and deductions), in <div> <div></div> <div> <div></div> <div></div> </div> </div> ? Please enter a whole number in the box below, without any punctuation. If you are paid hourly, please estimate an equivalent weekly, monthly, or yearly salary. If you prefer not to answer, please leave the box empty.
SalaryType	Is that salary weekly, monthly, or yearly?
ConvertedSalary	Salary converted to annual USD salaries using the exchange rate on 2018-01-18, assuming 12 working months and 50 working weeks.
CurrencySymbol	Three digit currency abbreviation.

CommunicationTools	Which of the following tools do you use to communicate, coordinate, or share knowledge with your coworkers? Please select all that apply.
TimeFullyProductive	Suppose a new developer with four years of experience, including direct experience working with your company's main technical stack, joined your team tomorrow. All other things being equal, how long would you expect it to take before they were fully productive and contributing at a typical level to your main code base?
EducationTypes	Which of the following types of non-degree education have you used or participated in? Please select all that apply.
SelfTaughtTypes	You indicated that you had taught yourself a programming technology without taking a course. What resources did you use to do that? If you've done it more than once, please think about the most recent time you've done so. Please select all that apply.
TimeAfterBootcamp	You indicated previously that you went through a developer training program or bootcamp. How long did it take you to get a full-time job as a developer after graduating?
HackathonReasons	You indicated previously that you had participated in an online coding competition or hackathon. Which of the following best describe your reasons for doing so?
AgreeDisagree1	To what extent do you agree or disagree with each of the following statements? I feel a sense of kinship or connection to other developers
AgreeDisagree2	To what extent do you agree or disagree with each of the following statements? I think of myself as competing with my peers
AgreeDisagree3	To what extent do you agree or disagree with each of the following statements? I'm not as good at programming as most of my peers
LanguageWorkedWith	Which of the following programming, scripting, and markup languages have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the language and want to continue to do so, please check both boxes in that row.)
LanguageDesireNextYear	Which of the following programming, scripting, and markup languages have you done extensive

	development work in over the past year, and which do you want to work in over the next year? (If you both worked with the language and want to continue to do so, please check both boxes in that row.)
DatabaseWorkedWith	Which of the following database environments have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the database and want to continue to do so, please check both boxes in that row.)
DatabaseDesireNextYear	Which of the following database environments have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the database and want to continue to do so, please check both boxes in that row.)
PlatformWorkedWith	Which of the following platforms have you done extensive development work for over the past year? (If you both developed for the platform and want to continue to do so, please check both boxes in that row.)
PlatformDesireNextYear	Which of the following platforms have you done extensive development work for over the past year? (If you both developed for the platform and want to continue to do so, please check both boxes in that row.)
FrameworkWorkedWith	Which of the following libraries, frameworks, and tools have you done extensive development work in over the past year, and which do you want to work in over the next year?
FrameworkDesireNextYear	Which of the following libraries, frameworks, and tools have you done extensive development work in over the past year, and which do you want to work in over the next year?
IDE	Which development environment(s) do you use regularly? Please check all that apply.
OperatingSystem	What is the primary operating system in which you work?
NumberMonitors	How many monitors are set up at your workstation?
Methodology	Which of the following methodologies do you have experience working in?
VersionControl	What version control systems do you use regularly? Please select all that apply.

CheckInCode	Over the last year, how often have you checked-in or committed code?
AdBlocker	Do you have ad-blocking software installed on any computers you use regularly?
AdBlockerDisable	In the past month, have you disabled your ad blocker for any reason, even temporarily or for a specific website?
AdBlockerReasons	What are the reasons that you have disabled your ad blocker in the past month? Please select all that apply.
AdsAgreeDisagree1	To what extent do you agree or disagree with the following statements: Online advertising can be valuable when it is relevant to me
AdsAgreeDisagree2	To what extent do you agree or disagree with the following statements: I enjoy seeing online updates from companies that I like
AdsAgreeDisagree3	To what extent do you agree or disagree with the following statements: I fundamentally dislike the concept of advertising
AdsActions	Which of the following actions have you taken in the past month? Please select all that apply.
AdsPriorities1	Please rank the following advertising qualities in order of their importance to you (by dragging the choices up and down), where 1 is the most important, and 7 is the least important. The advertisement is relevant to me
AdsPriorities2	Please rank the following advertising qualities in order of their importance to you (by dragging the choices up and down), where 1 is the most important, and 7 is the least important. The advertisement is honest about its goals
AdsPriorities3	Please rank the following advertising qualities in order of their importance to you (by dragging the choices up and down), where 1 is the most important, and 7 is the least important. The advertisement provides useful information
AdsPriorities4	Please rank the following advertising qualities in order of their importance to you (by dragging the choices up and down), where 1 is the most important, and 7 is the least important. The advertisement seems trustworthy
AdsPriorities5	Please rank the following advertising qualities in order of their importance to you (by dragging the choices up and down), where 1 is the most important, and 7 is the

	least important. The advertisement is from a company that I like
AdsPriorities6	Please rank the following advertising qualities in order of their importance to you (by dragging the choices up and down), where 1 is the most important, and 7 is the least important. The advertisement offers something of value, like a free trial
AdsPriorities7	Please rank the following advertising qualities in order of their importance to you (by dragging the choices up and down), where 1 is the most important, and 7 is the least important. The advertisement avoids fluffy or vague language
AlDangerous	What do you think is the most dangerous aspect of increasingly advanced AI technology?
AlInteresting	What do you think is the most exciting aspect of increasingly advanced AI technology?
AlResponsible	Whose responsibility is it, primarily, to consider the ramifications of increasingly advanced AI technology?
AlFuture	Overall, what's your take on the future of artificial intelligence?
EthicsChoice	Imagine that you were asked to write code for a purpose or product that you consider extremely unethical. Do you write the code anyway?
EthicsReport	Do you report or otherwise call out the unethical code in question?
EthicsResponsible	Who do you believe is ultimately most responsible for code that accomplishes something unethical?
EthicalImplications	Do you believe that you have an obligation to consider the ethical implications of the code that you write?
StackOverflowRecommend	How likely is it that you would recommend Stack Overflow overall to a friend or colleague? Where 0 is not likely at all and 10 is very likely.
StackOverflowVisit	How frequently would you say you visit Stack Overflow?
StackOverflowHasAccount	Do you have a Stack Overflow account?
StackOverflowParticipate	How frequently would you say you participate in Q&A on Stack Overflow? By participate we mean ask, answer, vote for, or comment on questions.
StackOverflowJobs	Have you ever used or visited Stack Overflow Jobs?
StackOverflowDevStory	Do you have an up-to-date Developer Story on Stack Overflow?

StackOverflowJobsRecommend	How likely is it that you would recommend Stack Overflow Jobs to a friend or colleague? Where 0 is not likely at all and 10 is very likely.
StackOverflowConsiderMember	Do you consider yourself a member of the Stack Overflow community?
HypotheticalTools1	Please rate your interest in participating in each of the following hypothetical tools on Stack Overflow, where 1 is not at all interested and 5 is extremely interested. A peer mentoring system
HypotheticalTools2	Please rate your interest in participating in each of the following hypothetical tools on Stack Overflow, where 1 is not at all interested and 5 is extremely interested. A private area for people new to programming
HypotheticalTools3	Please rate your interest in participating in each of the following hypothetical tools on Stack Overflow, where 1 is not at all interested and 5 is extremely interested. A programming-oriented blog platform
HypotheticalTools4	Please rate your interest in participating in each of the following hypothetical tools on Stack Overflow, where 1 is not at all interested and 5 is extremely interested. An employer or job review system
HypotheticalTools5	Please rate your interest in participating in each of the following hypothetical tools on Stack Overflow, where 1 is not at all interested and 5 is extremely interested. An area for Q&A related to career growth
WakeTime	On days when you work, what time do you typically wake up?
HoursComputer	On a typical day, how much time do you spend on a desktop or laptop computer?
HoursOutside	On a typical day, how much time do you spend outside?
SkipMeals	In a typical week, how many times do you skip a meal in order to be more productive?
ErgonomicDevices	What ergonomic furniture or devices do you use on a regular basis? Please select all that apply.
Exercise	In a typical week, how many times do you exercise?
Gender	Which of the following do you currently identify as? Please select all that apply. If you prefer not to answer, you may leave this question blank.
SexualOrientation	Which of the following do you currently identify as? Please select all that apply. If you prefer not to answer, you may leave this question blank.

EducationParents	What is the highest level of education received by either of your parents? If you prefer not to answer, you may leave this question blank.
RaceEthnicity	Which of the following do you identify as? Please check all that apply. If you prefer not to answer, you may leave this question blank.
Age	What is your age? If you prefer not to answer, you may leave this question blank.
Dependents	Do you have any children or other dependents that you care for? If you prefer not to answer, you may leave this question blank.
MilitaryUS	Are you currently serving or have you ever served in the U.S. Military?
SurveyTooLong	How do you feel about the length of the survey that you just completed?
SurveyEasy	How easy or difficult was this survey to complete?
<pre>glimpse(survey) ## Rows: 98,855 ## Columns: 129 ## \$ Respondent 8... ## \$ Hobby ... ## \$ OpenSource S... ## \$ Country U... ## \$ Student e... ## \$ Employment ,... ## \$ FormalEducation E... ## \$ UndergradMajor l... ## \$ CompanySize e... ## \$ DevType n... ## \$ YearsCoding r... ## \$ YearsCodingProf 1... ## \$ JobSatisfaction i... ## \$ CareerSatisfaction</pre>	
	<pre><int> 1, 3, 4, 5, 7, 8, 9, 10, 11, 16, 17, 1 <fct> Yes, Yes, Yes, No, Yes, Yes, Yes, Yes, ... <fct> No, Yes, Yes, No, No, No, Yes, Yes, Ye ... <fct> Kenya, United Kingdom, United States, ... <fct> "No", "No", "No", "No", "Yes, part-tim ... <fct> Employed part-time, Employed full-time ,... <fct> "Bachelorâ\200\231s degree (BA, BS, B. ... <fct> "Mathematics or statistics", "A natura ... <fct> "20 to 99 employees", "10,000 or more ... <fct> "Full-stack developer", "Database admi ... <fct> 3-5 years, 30 or more years, 24-26 yea ... <fct> 3-5 years, 18-20 years, 6-8 years, 12- ... <fct> Extremely satisfied, Moderately dissat ... <fct> Extremely satisfied, Neither satisfied</pre>

```

...
## $ HopeFiveYears      <fct> Working as a founder or co-founder of
m...
## $ JobSearchStatus    <fct> "Iâ\200\231m not actively looking, but
...
## $ LastNewJob         <fct> Less than a year ago, More than 4 year
S...
## $ AssessJob1         <int> 10, 1, NA, NA, 8, 8, 5, 6, 6, NA, NA,
N...
## $ AssessJob2         <int> 7, 7, NA, NA, 5, 5, 3, 5, 3, NA, NA, N
A...
## $ AssessJob3         <int> 8, 10, NA, NA, 7, 4, 9, 4, 7, NA, NA,
N...
## $ AssessJob4         <int> 1, 8, NA, NA, 1, 9, 4, 2, 4, NA, NA, N
A...
## $ AssessJob5         <int> 2, 2, NA, NA, 2, 1, 1, 7, 1, NA, NA, N
A...
## $ AssessJob6         <int> 5, 5, NA, NA, 6, 3, 8, 8, 5, NA, NA, N
A...
## $ AssessJob7         <int> 3, 4, NA, NA, 4, 6, 2, 10, 10, NA, NA,
...
## $ AssessJob8         <int> 4, 3, NA, NA, 3, 2, 7, 1, 8, NA, NA, N
A...
## $ AssessJob9         <int> 9, 6, NA, NA, 10, 10, 10, 9, 9, NA, NA
,...
## $ AssessJob10        <int> 6, 9, NA, NA, 9, 7, 6, 3, 2, NA, NA, N
A...
## $ AssessBenefits1    <int> NA, 1, NA, NA, 1, 1, 1, 1, 1, NA, NA,
N...
## $ AssessBenefits2    <int> NA, 5, NA, NA, 10, 3, 3, 3, 3, NA, NA,
...
## $ AssessBenefits3    <int> NA, 3, NA, NA, 2, 4, 2, 5, 2, NA, NA,
N...
## $ AssessBenefits4    <int> NA, 7, NA, NA, 4, 10, 9, 7, 9, NA, NA,
...
## $ AssessBenefits5    <int> NA, 10, NA, NA, 8, 9, 11, 6, 11, NA, N
A...
## $ AssessBenefits6    <int> NA, 4, NA, NA, 3, 2, 4, 2, 5, NA, NA,
N...
## $ AssessBenefits7    <int> NA, 11, NA, NA, 11, 6, 8, 11, 8, NA, N
A...
## $ AssessBenefits8    <int> NA, 9, NA, NA, 7, 5, 6, 9, 4, NA, NA,
N...
## $ AssessBenefits9    <int> NA, 6, NA, NA, 5, 11, 7, 4, 10, NA, NA
,...
## $ AssessBenefits10   <int> NA, 2, NA, NA, 9, 7, 10, 10, 7, NA, NA
,...
## $ AssessBenefits11   <int> NA, 8, NA, NA, 6, 8, 5, 8, 6, NA, NA,
N...
## $ JobContactPriorities1 <int> 3, 3, NA, NA, 2, 4, 3, 1, 5, NA, NA, N

```


A...	
## \$ JobContactPriorities2	<int> 1, 1, NA, NA, 1, 2, 1, 3, 1, NA, NA, N
A...	
## \$ JobContactPriorities3	<int> 4, 5, NA, NA, 4, 5, 5, 2, 2, NA, NA, N
A...	
## \$ JobContactPriorities4	<int> 2, 2, NA, NA, 5, 1, 4, 4, 3, NA, NA, N
A...	
## \$ JobContactPriorities5	<int> 5, 4, NA, NA, 3, 3, 2, 5, 4, NA, NA, N
A...	
## \$ JobEmailPriorities1	<int> 5, 1, NA, NA, 7, 2, 1, 2, 3, NA, NA, N
A...	
## \$ JobEmailPriorities2	<int> 6, 3, NA, NA, 3, 6, 5, 6, 7, NA, NA, N
A...	
## \$ JobEmailPriorities3	<int> 7, 4, NA, NA, 6, 7, 3, 1, 2, NA, NA, N
A...	
## \$ JobEmailPriorities4	<int> 2, 5, NA, NA, 2, 3, 4, 3, 4, NA, NA, N
A...	
## \$ JobEmailPriorities5	<int> 1, 2, NA, NA, 1, 1, 2, 7, 1, NA, NA, N
A...	
## \$ JobEmailPriorities6	<int> 4, 6, NA, NA, 4, 5, 6, 5, 6, NA, NA, N
A...	
## \$ JobEmailPriorities7	<int> 3, 7, NA, NA, 5, 4, 7, 4, 5, NA, NA, N
A...	
## \$ UpdateCV	<fct> My job status or other personal status
...	
## \$ Currency	<fct> NA, British pounds sterling (£), NA,
U...	
## \$ Salary	<fct> NA, 51000, NA, NA, 260000, 30000, 1200
0...	
## \$ SalaryType	<fct> Monthly, Yearly, NA, NA, Yearly, NA, Y
e...	
## \$ ConvertedSalary	<dbl> NA, 70841, NA, NA, 21426, 41671, 12000
0...	
## \$ CurrencySymbol	<fct> KES, GBP, NA, NA, ZAR, GBP, USD, NA, U
S...	
## \$ CommunicationTools	<fct> "Slack", "Confluence;Office / producti
v...	
## \$ TimeFullyProductive	<fct> One to three months, One to three mont
h...	
## \$ EducationTypes	<fct> "Taught yourself a new language, frame
w...	
## \$ SelfTaughtTypes	<fct> "The official documentation and/or sta
n...	
## \$ TimeAfterBootcamp	<fct> NA, NA, NA, NA, NA, NA, NA, Immediatel
y...	
## \$ HackathonReasons	<fct> "To build my professional network", NA
,...	
## \$ AgreeDisagree1	<fct> Strongly agree, Agree, NA, Disagree, S
t...	
## \$ AgreeDisagree2	<fct> Strongly agree, Agree, NA, Disagree, A

g...	
## \$ AgreeDisagree3	<fct> Neither Agree nor Disagree, Neither Ag
r...	
## \$ LanguageWorkedWith	<fct> JavaScript;Python;HTML;CSS, JavaScript
;	
## \$ LanguageDesireNextYear	<fct> JavaScript;Python;HTML;CSS, Go;Python,
...	
## \$ DatabaseWorkedWith	<fct> "Redis;SQL Server;MySQL;PostgreSQL;Ama
Z...	
## \$ DatabaseDesireNextYear	<fct> "Redis;SQL Server;MySQL;PostgreSQL;Ama
Z...	
## \$ PlatformWorkedWith	<fct> AWS;Azure;Linux;Firebase, Linux, NA, A
Z...	
## \$ PlatformDesireNextYear	<fct> AWS;Azure;Linux;Firebase, Linux, NA, A
Z...	
## \$ FrameworkWorkedWith	<fct> Django;React, Django, NA, NA, NA, Angu
l...	
## \$ FrameworkDesireNextYear	<fct> Django;React, React, NA, Angular;.NET
C...	
## \$ IDE	<fct> Komodo;Vim;Visual Studio Code, IPython
...	
## \$ OperatingSystem	<fct> Linux-based, Linux-based, NA, Windows,
...	
## \$ NumberMonitors	<fct> 1, 2, NA, 2, 2, 2, 2, 1, 1, NA, NA, NA
,	
## \$ Methodology	<fct> Agile;Scrum, NA, NA, Agile;Kanban;Scru
m...	
## \$ VersionControl	<fct> Git, Git;Subversion, NA, Git, Zip file
...	
## \$ CheckInCode	<fct> Multiple times per day, A few times pe
r...	
## \$ AdBlocker	<fct> Yes, Yes, NA, Yes, No, Yes, Yes, Yes,
N...	
## \$ AdBlockerDisable	<fct> No, Yes, NA, Yes, NA, Yes, Yes, No, NA
,	
## \$ AdBlockerReasons	<fct> NA, The website I was visiting asked m
e...	
## \$ AdsAgreeDisagree1	<fct> Strongly agree, Somewhat agree, NA, Ne
i...	
## \$ AdsAgreeDisagree2	<fct> Strongly agree, Neither agree nor disa
g...	
## \$ AdsAgreeDisagree3	<fct> Strongly agree, Neither agree nor disa
g...	
## \$ AdsActions	<fct> Saw an online advertisement and then r
e...	
## \$ AdsPriorities1	<int> 1, 3, NA, NA, 2, 1, 1, NA, 1, NA, NA,
N...	
## \$ AdsPriorities2	<int> 5, 5, NA, NA, 3, 3, 4, NA, 3, NA, NA,
N...	
## \$ AdsPriorities3	<int> 4, 1, NA, NA, 4, 4, 2, NA, 5, NA, NA,

N...	
## \$ AdsPriorities4	<int> 7, 4, NA, NA, 6, 2, 5, NA, 4, NA, NA,
N...	
## \$ AdsPriorities5	<int> 2, 6, NA, NA, 1, 7, 3, NA, 2, NA, NA,
N...	
## \$ AdsPriorities6	<int> 6, 7, NA, NA, 7, 5, 7, NA, 7, NA, NA,
N...	
## \$ AdsPriorities7	<int> 3, 2, NA, NA, 5, 6, 6, NA, 6, NA, NA,
N...	
## \$ AIDangerous	<fct> Artificial intelligence surpassing hum
a...	
## \$ AIInteresting	<fct> Algorithms making important decisions,
...	
## \$ AIResponsible	<fct> The developers or the people creating
t...	
## \$ AIFuture	<fct> "I'm excited about the possibilities m
O...	
## \$ EthicsChoice	<fct> No, Depends on what it is, NA, No, No,
...	
## \$ EthicsReport	<fct> "Yes, and publicly", "Depends on what
i...	
## \$ EthicsResponsible	<fct> Upper management at the company/organi
Z...	
## \$ EthicalImplications	<fct> Yes, Yes, NA, Yes, Yes, Unsure / I don
'...	
## \$ StackOverflowRecommend	<fct> 10 (Very Likely), 10 (Very Likely), NA
,...	
## \$ StackOverflowVisit	<fct> Multiple times per day, A few times pe
r...	
## \$ StackOverflowHasAccount	<fct> Yes, Yes, NA, Yes, Yes, Yes, Yes, Yes,
...	
## \$ StackOverflowParticipate	<fct> I have never participated in Q&A on St
a...	
## \$ StackOverflowJobs	<fct> "No, I knew that Stack Overflow had a
j...	
## \$ StackOverflowDevStory	<fct> "Yes", "No, I have one but it's out of
...	
## \$ StackOverflowJobsRecommend	<fct> NA, 7, NA, 8, NA, 8, 7, NA, 7, NA, NA,
...	
## \$ StackOverflowConsiderMember	<fct> Yes, Yes, NA, Yes, Yes, No, No, Yes, N
O...	
## \$ HypotheticalTools1	<fct> Extremely interested, A little bit int
e...	
## \$ HypotheticalTools2	<fct> Extremely interested, A little bit int
e...	
## \$ HypotheticalTools3	<fct> Extremely interested, A little bit int
e...	
## \$ HypotheticalTools4	<fct> Extremely interested, A little bit int
e...	
## \$ HypotheticalTools5	<fct> Extremely interested, A little bit int

```

e...
## $ WakeTime          <fct> Between 5:00 - 6:00 AM, Between 6:01 -
...
## $ HoursComputer     <fct> 9 - 12 hours, 5 - 8 hours, NA, 9 - 12
h...
## $ HoursOutside      <fct> 1 - 2 hours, 30 - 59 minutes, NA, Less
...
## $ SkipMeals         <fct> Never, Never, NA, 3 - 4 times per week
,...
## $ ErgonomicDevices  <fct> Standing desk, Ergonomic keyboard or m
O...
## $ Exercise          <fct> 3 - 4 times per week, Daily or almost
e...
## $ Gender            <fct> "Male", "Male", NA, "Male", "Male", "M
a...
## $ SexualOrientation <fct> Straight or heterosexual, Straight or
h...
## $ EducationParents  <fct> "Bachelorâ\200\231s degree (BA, BS, B.
E...
## $ RaceEthnicity     <fct> Black or of African descent, White or
O...
## $ Age               <fct> 25 - 34 years old, 35 - 44 years old,
N...
## $ Dependents        <fct> Yes, Yes, NA, No, Yes, No, No, No, Yes
,...
## $ MilitaryUS        <fct> NA, NA, NA, No, NA, NA, No, NA, No, NA
,...
## $ SurveyTooLong     <fct> The survey was an appropriate length,
T...
## $ SurveyEasy        <fct> Very easy, Somewhat easy, NA, Somewhat
...

```

From the glimpse, we can see that there are some null values in some questions. This is due to the fact that most of the questions are optional in the survey. For the sake of exploration of this dataset, we will not be considering null values but it would be interesting to explore why a few questions have been answered null by the respondents.

Part I : General exploration of survey results

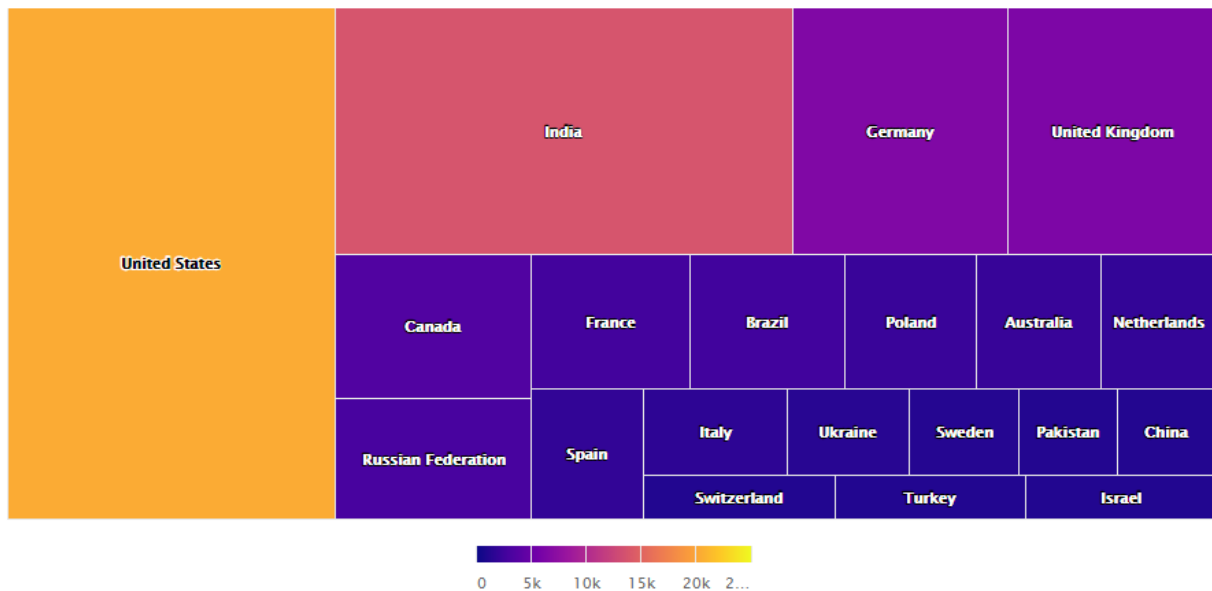
Let's see which countries have the maximum respondents of the survey.

```

survey %>%
  group_by(Country) %>%
  count() %>%
  arrange(desc(n)) %>%
  head(20) %>%
  hchart('treemap', hcaes(x=Country, value=n, color=n)) %>%
  hc_colorAxis(stops=color_stops(colors=plasma(10))) %>%
  hc_title(text="Top 20 Countries of respondents of survey")

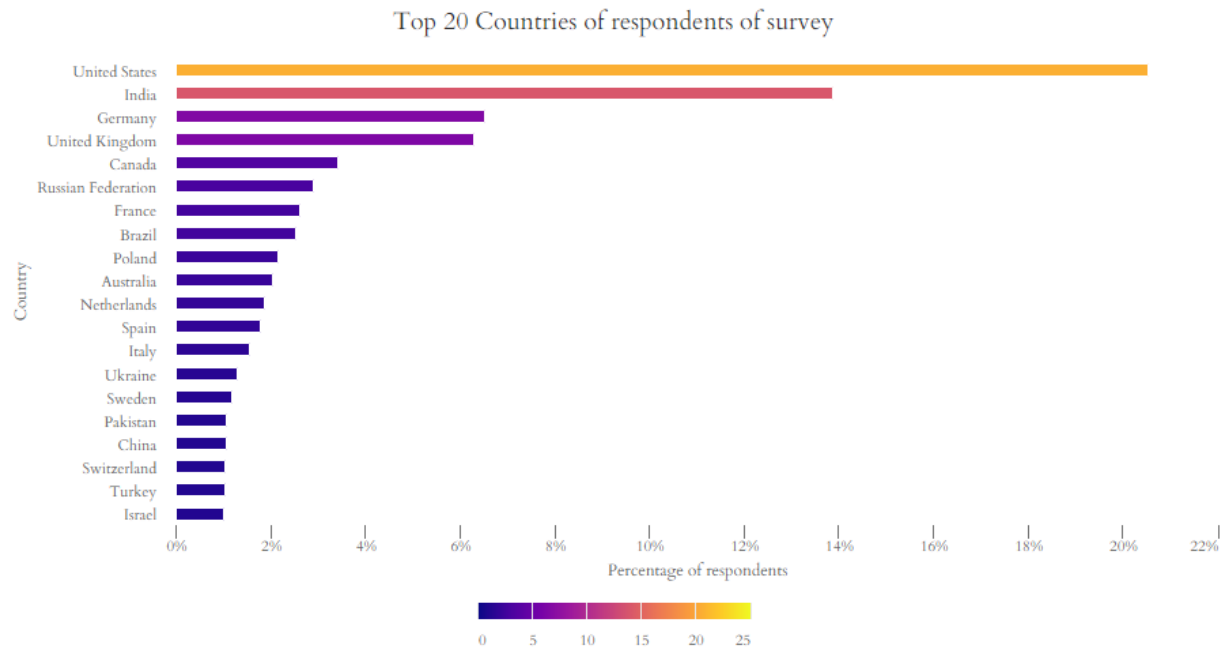
```

Top 20 Countries of respondents of survey



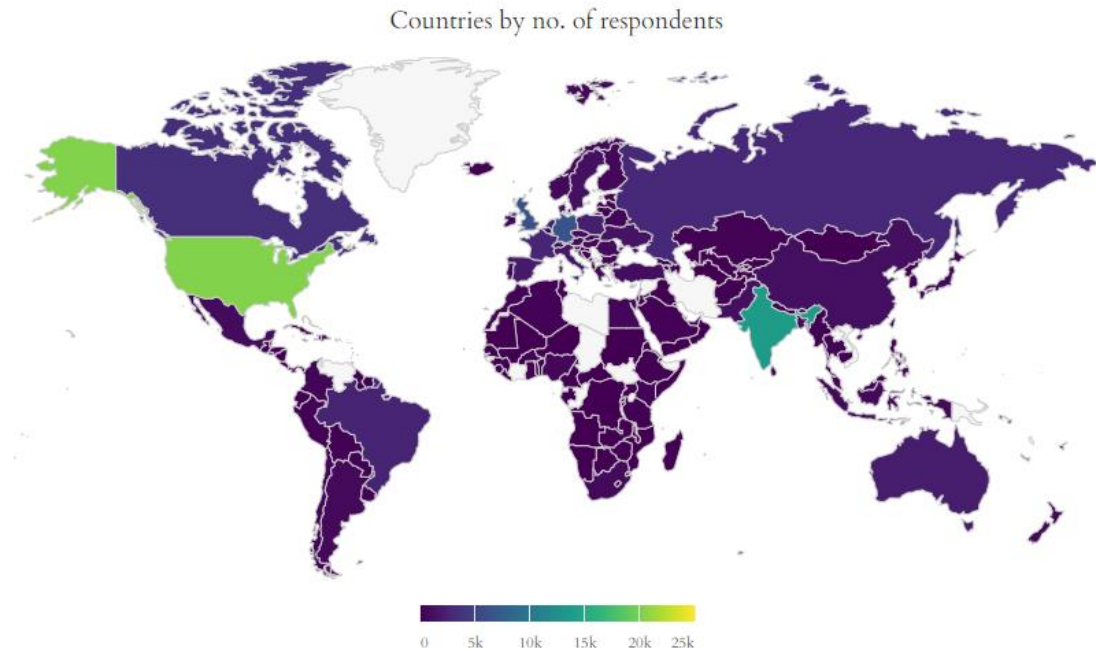
As we can see from the above tree map, the country with the maximum number of respondents come from United States with India at second position while Germany, United Kingdom and Canada make up the top 5. Let us also take a look at the percentage distribution of respondents

```
size=nrow(survey)
survey %>%
  group_by(Country) %>%
  count() %>%
  arrange(desc(n)) %>%
  head(20) %>%
  hchart('bar', hcaes(x=Country, y=round((n/size)*100,2))) %>%
  hc_colorAxis(stops=color_stops(colors=plasma(10))) %>%
  hc_title(text="Top 20 Countries of respondents of survey") %>%
  hc_yAxis(title=list(text="Percentage of respondents"), labels = list(format
= "{value}%")) %>%
  hc_add_theme(hc_theme_tufte())
```



From the bar chart above, we can see that : 20.54% of respondents are from the United States 13.88% of respondents are from India Germany(6.53%),United Kingdom(6.29%) and Canada(3.43%) have the next 3 highest respondents.

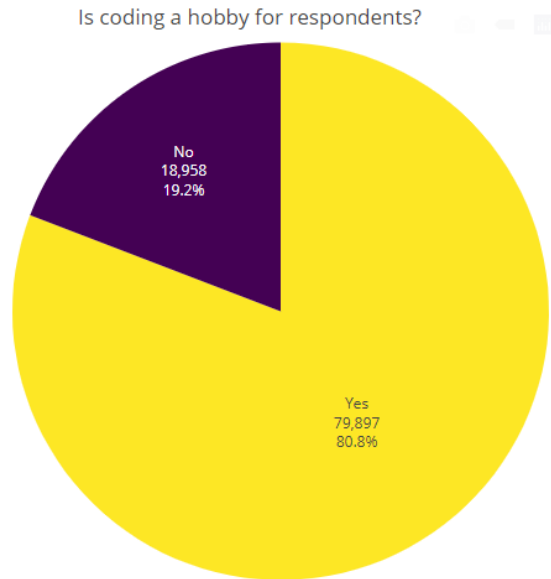
```
survey %>%
  group_by(Country) %>%
  count() -> country
country$Country <- as.character(country$Country)
country[country=="United States"] <- "United States of America"
country[country=="Russian Federation"] <- "Russia"
highchart() %>%
  hc_add_series_map(worldgeojson,country,value='n',joinBy=c('name','Country'
')) %>%
  hc_colorAxis(stops=color_stops(colors=viridis(10))) %>%
  hc_title(text="Countries by no. of respondents") %>%
  hc_add_theme(hc_theme_tufte())
```



The above map is a better visualization of all the respondents by country.

Now, let us look at how many respondents take up coding as a hobby.

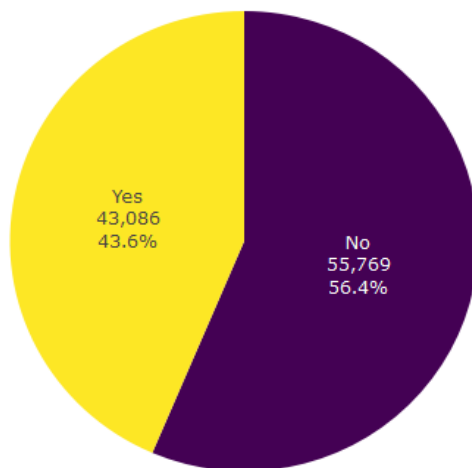
```
survey %>%  
  group_by(Hobby) %>%  
  count() %>%  
  plot_ly(type="pie",  
          labels=~Hobby,  
          values=~n,  
          textposition="inside",  
          textinfo='label+percent+value',  
          showlegend=FALSE,  
          marker=list(colors = viridis(2))  
  )%>%  
  layout(title="Is coding a hobby for respondents?")
```



From the above piechart, we can see that more than 80% of the respondents take up coding as a hobby.

```
survey %>%  
  group_by(OpenSource) %>%  
  count() %>%  
  plot_ly(type="pie",  
    labels=~OpenSource,  
    values=~n,  
    textposition="inside",  
    textinfo='label+percent+value',  
    showlegend=FALSE,  
    marker=list(colors = viridis(2))  
  ) %>%  
  layout(title="Do respondents contribute to OpenSource Projets?")
```

Do respondents contribute to OpenSource Projets?



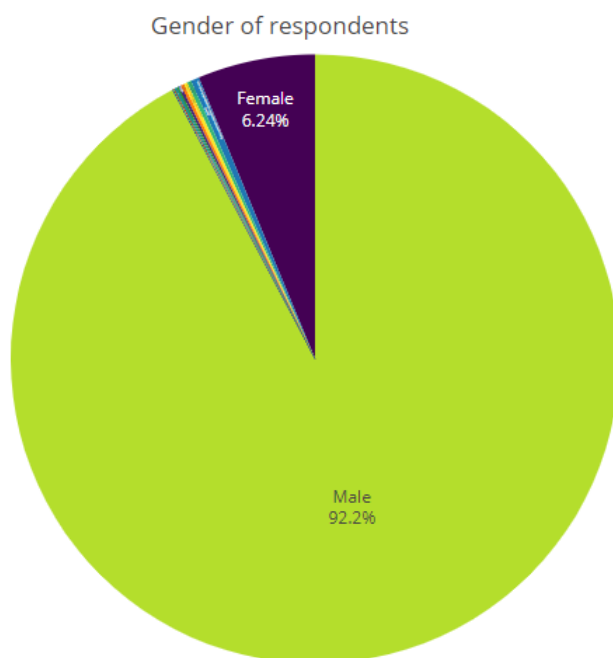
From the above piechart, we can see that around 56% of respondents contribute to open source projects.

Let's take a look at the gender of respondents.

```
S <- sum(is.na(survey$Gender))
S
## [1] 34386
round((S/size)*100,2)
## [1] 34.78
```

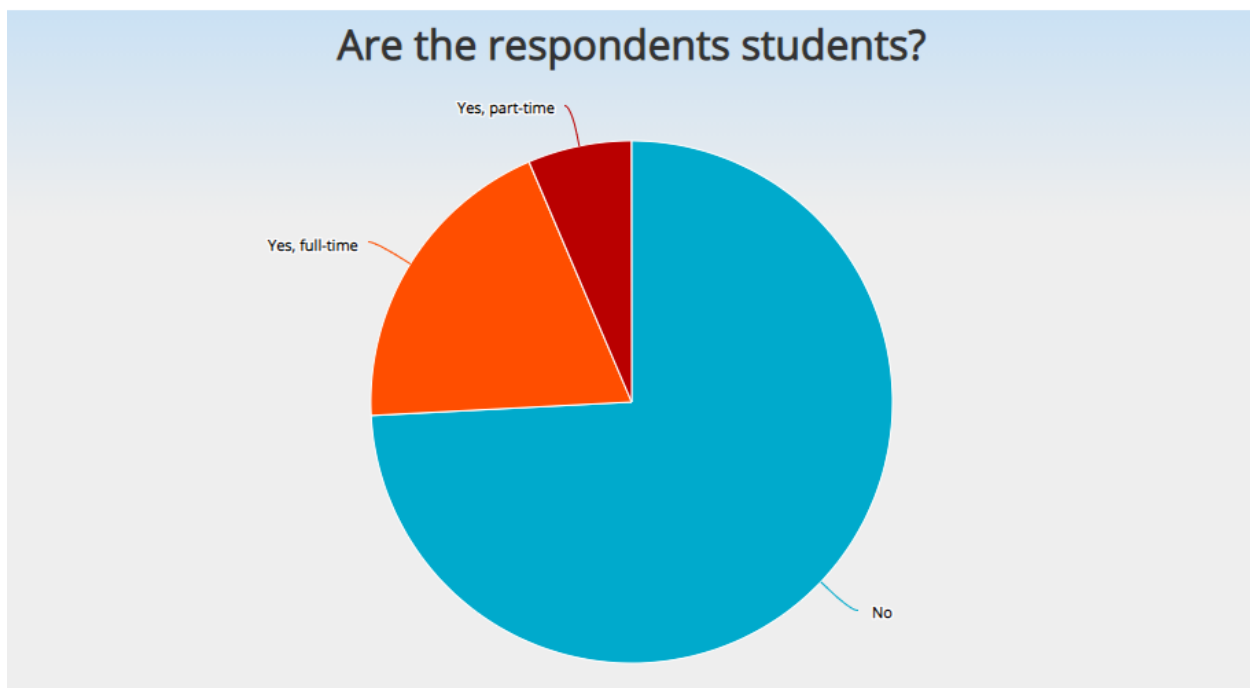
We see that almost 34386(34.78%) respondents have not disclosed their gender in the survey. It would be interesting to know the reason behind this.

```
survey %>%
  filter(!is.na(Gender)) %>%
  group_by(Gender) %>%
  count() %>%
  plot_ly(type="pie",
          labels=~Gender,
          values=~n,
          textposition="inside",
          textinfo='label+percent',
          showlegend=FALSE,
          marker=list(colors = viridis(10))
  ) %>%
  layout(title="Gender of respondents")
```



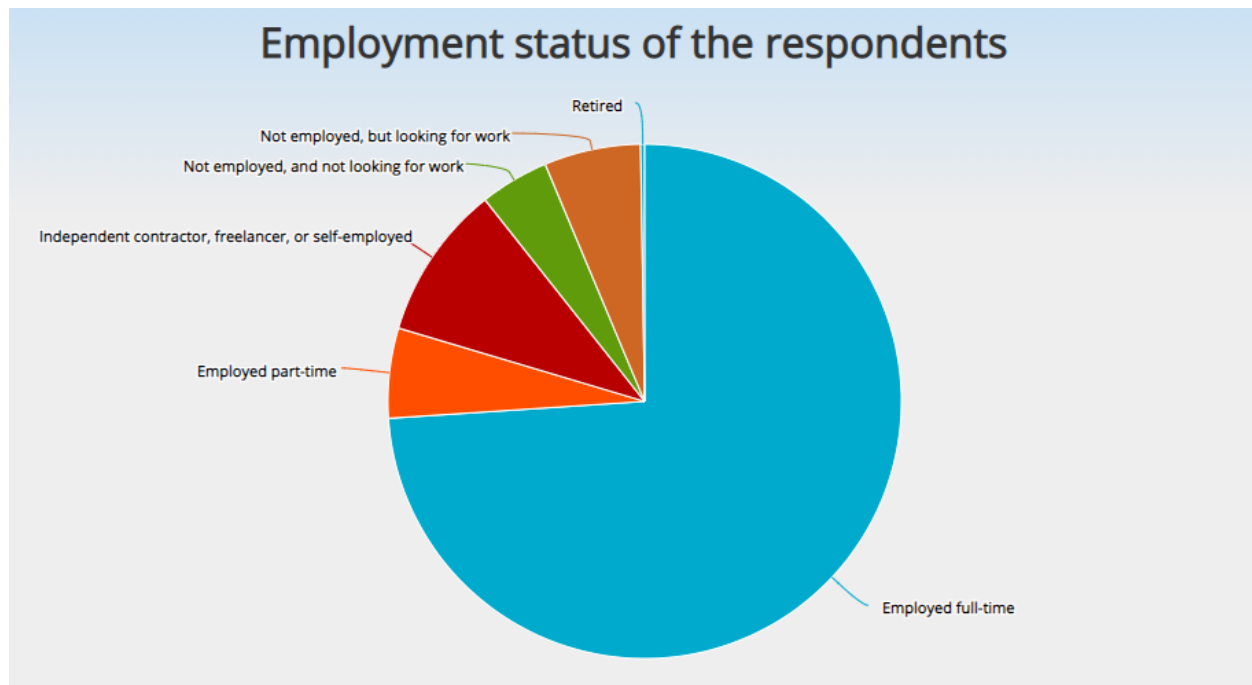
The results are very shocking. We can see that 92.2% respondents who gave their gender are male, 6.24% of the respondents are female while the rest are transgender, non-binary or gender non-conforming. The results maybe misleading due to the fact that 34.78% of the respondents have not disclosed their gender. The reasons for this is not known, but the above chart shows us that there exists a huge gender gap in the StackOverflow Community. Now, let's take a look at whether the respondents are students and their current employment status.

```
survey %>%
  filter(!is.na(Student)) %>%
  group_by(Student)%>%
  count() %>%
  hchart("pie",hcaes(x=Student,y=round((n/size)*100,2))) %>%
  hc_add_theme(hc_theme_ffx()) %>%
  hc_title(text="Are the respondents students?",align="center",style = list(fontWeight="bold", fontSize="30px"))
```



From the above pie chart, we can see that more than 71.21% of the respondents are not students, 18.61% are full time students while 6.18 are part-time students.

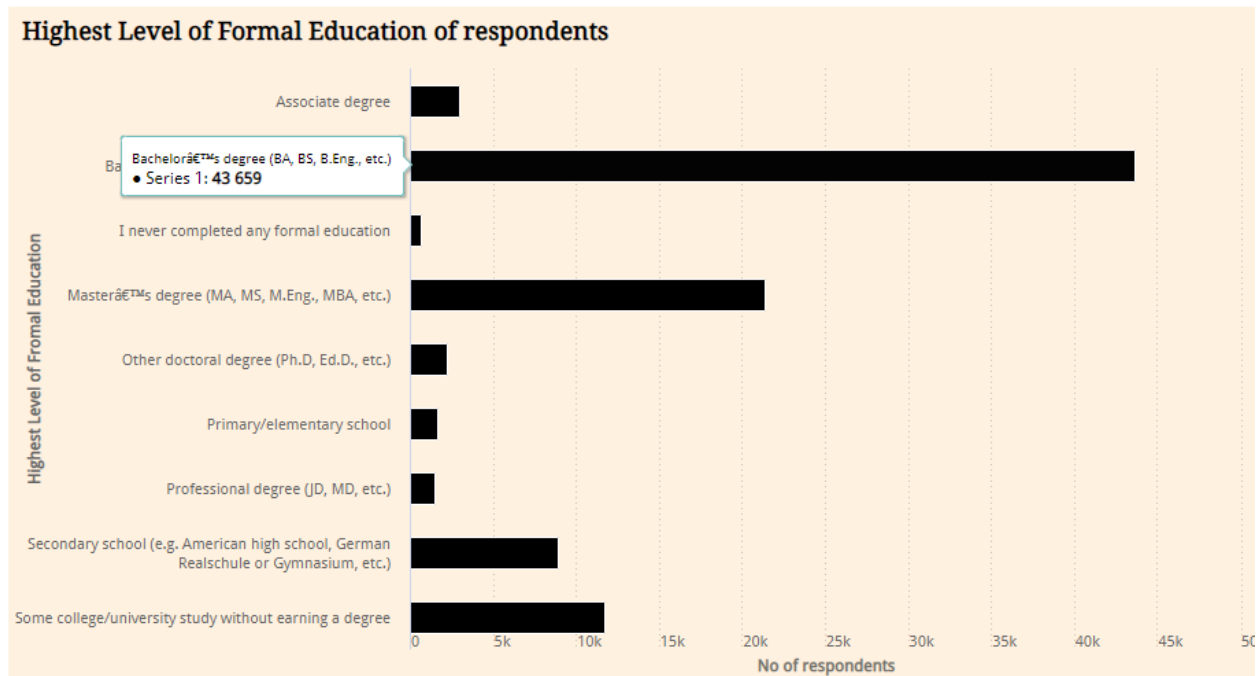
```
survey %>%
  filter(!is.na(Employment)) %>%
  group_by(Employment)%>%
  count() %>%
  hchart("pie",hcaes(x=Employment,y=round((n/size)*100,2))) %>%
  hc_add_theme(hc_theme_ffx()) %>%
  hc_title(text="Employment status of the respondents",align="center",style = list(fontWeight="bold", fontSize="30px"))
```



From, the above pie chart, we can see that 71.31% are employed full time, 5.44% are employed part time, 9.39% are self-employed or freelancers while 20.05% are not employed. It is refreshing to see that full time employees take time to contribute to the StackOverflow Community.

Now, let us take a look at the education of the respondents.

```
survey %>%
  filter(!is.na(FormalEducation)) %>%
  group_by(FormalEducation) %>%
  count() %>%
  hchart("bar", hcaes(x=FormalEducation, y=n)) %>%
  hc_title(text="Highest Level of Formal Education of respondents", style=list(
fontWeight="bold")) %>%
  hc_yAxis(title=list(text="No of respondents", style=list(fontWeight="bold")))
) %>%
  hc_xAxis(title=list(text="Highest Level of Fromal Education", style=list(fon
tWeight="bold"))) %>%
  hc_add_theme(hc_theme_ft()) %>%
  hc_colors("black")
```



We can see that majority of respondents have received a bachelor's degree or a master's degree. Only a very few students have not completed any formal education.

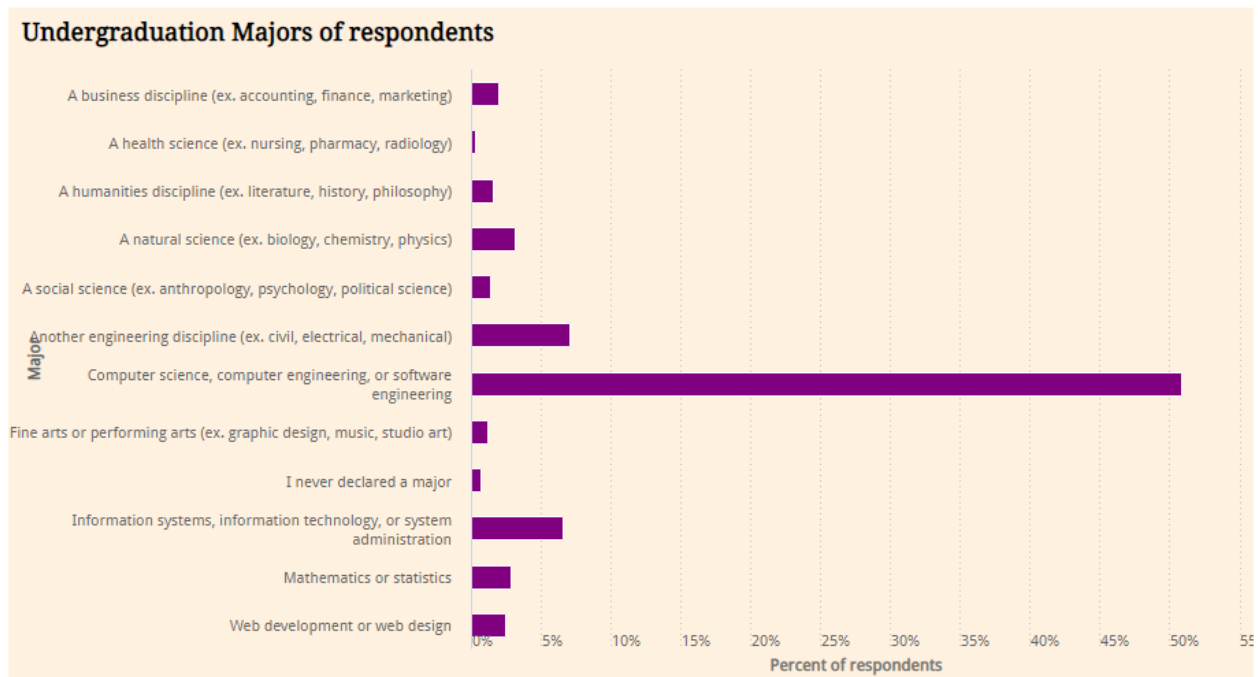
Let's take a look at what the respondents have majored in during their undergraduation.

```
S <- sum(is.na(survey$UndergradMajor))
S
## [1] 19819
round((S/size)*100,2)
## [1] 20.05
```

As we can see above, around 20% of respondents have not disclosed their undergrad major. This can be explained by the fact that almost 20% of students have either not received an undergrad degree as they have only completed their schooling, haven't had any formal education or have studied in college without earning a degree. This can be seen in the previous barplot.

```
survey %>%
  filter(!is.na(UndergradMajor)) %>%
  group_by(UndergradMajor) %>%
  count() %>%
  hchart("bar", hcaes(x=UndergradMajor, y=round((n/size)*100,2))) %>%
  hc_title(text="Undergraduation Majors of respondents", style=list(fontWeight="bold")) %>%
  hc_yAxis(title=list(text="Percent of respondents", style=list(fontWeight="bold")), labels = list(format = "{value}%")) %>%
  hc_xAxis(title=list(text="Major", style=list(fontWeight="bold"))) %>%
```

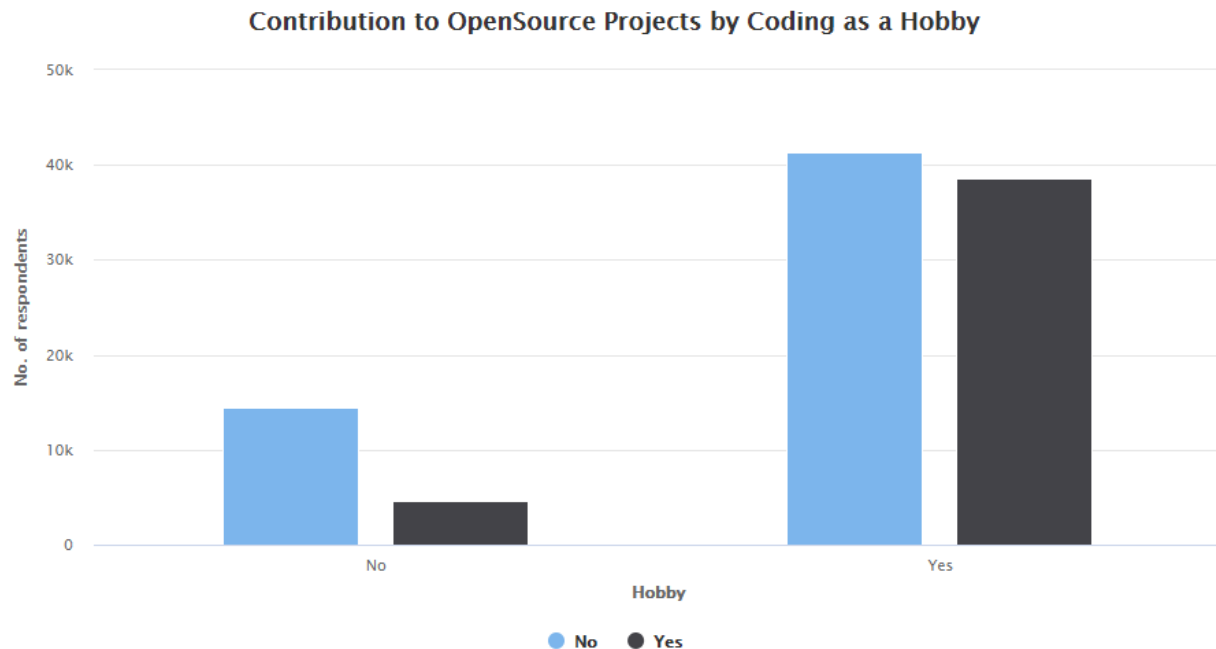
```
hc_add_theme(hc_theme_ft()) %>%
hc_colors("purple")
```



The results are obvious as about 50.92% of respondents have a major in Computer Science stream, 7.03% are from another engineering discipline, 6.58% are from the IT stream.

Now that we've generally explored a few responses from the survey, let us try to find a few relationships between these responses

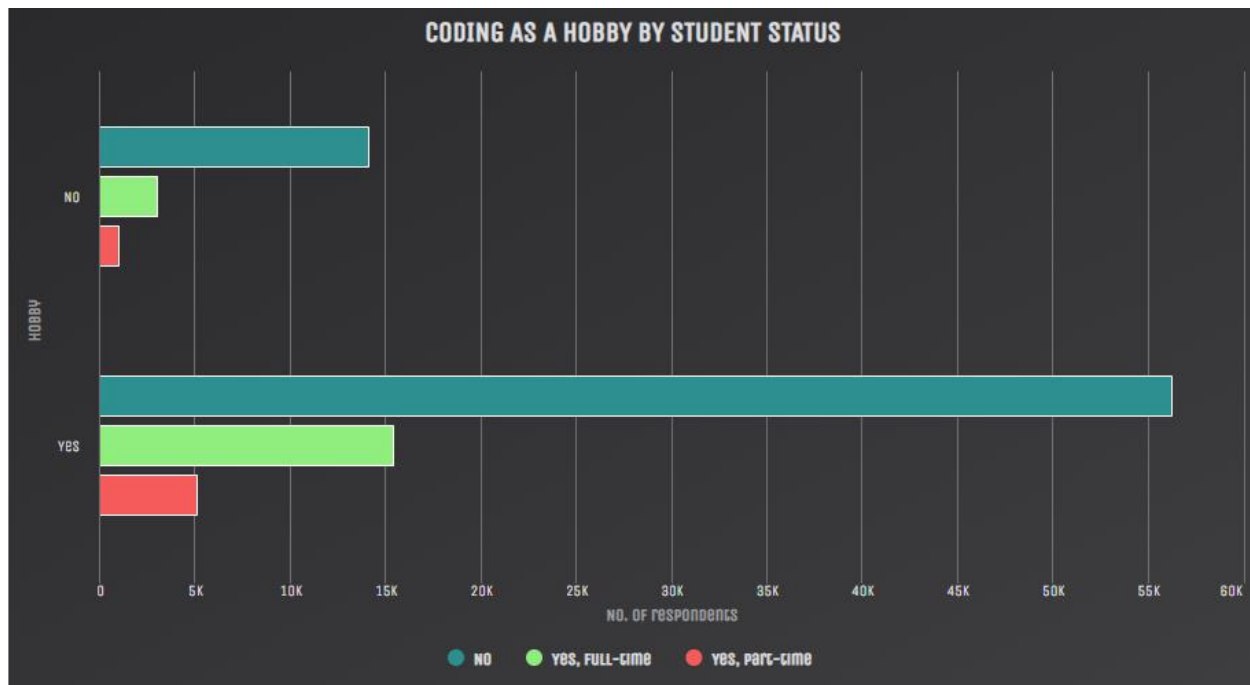
```
survey %>%
  count(Hobby,OpenSource) %>%
  hchart("column",hcaes(x=Hobby,y=n,group=OpenSource)) %>%
  hc_title(text="Contribution to OpenSource Projects by Coding as a Hobby",style=list(fontWeight="bold")) %>%
  hc_yAxis(title=list(text="No. of respondents",style=list(fontWeight="bold"))) %>%
  hc_xAxis(title=list(style=list(fontWeight="bold")))
```



It's pretty clear that people who code as a hobby tend to contribute more to opensource projects.

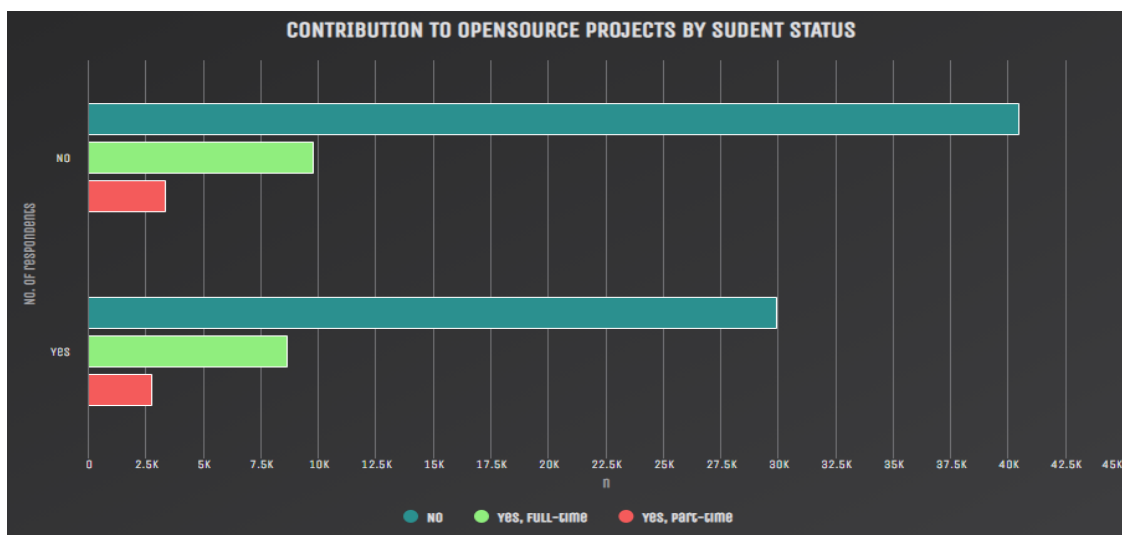
Let's first look at whether respondents code as a hobby and contribute to open source according to their student and employment status

```
survey %>%
  filter(!is.na(Student)) %>%
  count(Hobby,Student) %>%
  hchart("bar",hcaes(x=Hobby,y=n,group=Student)) %>%
  hc_title(text="Coding as a hobby by Student Status",style=list(fontWeight="bold")) %>%
  hc_yAxis(title=list(text="No. of respondents",style=list(fontWeight="bold"))) %>%
  hc_xAxis(title=list(style=list(fontWeight="bold"))) %>%
  hc_add_theme(hc_theme_darkunica())
```



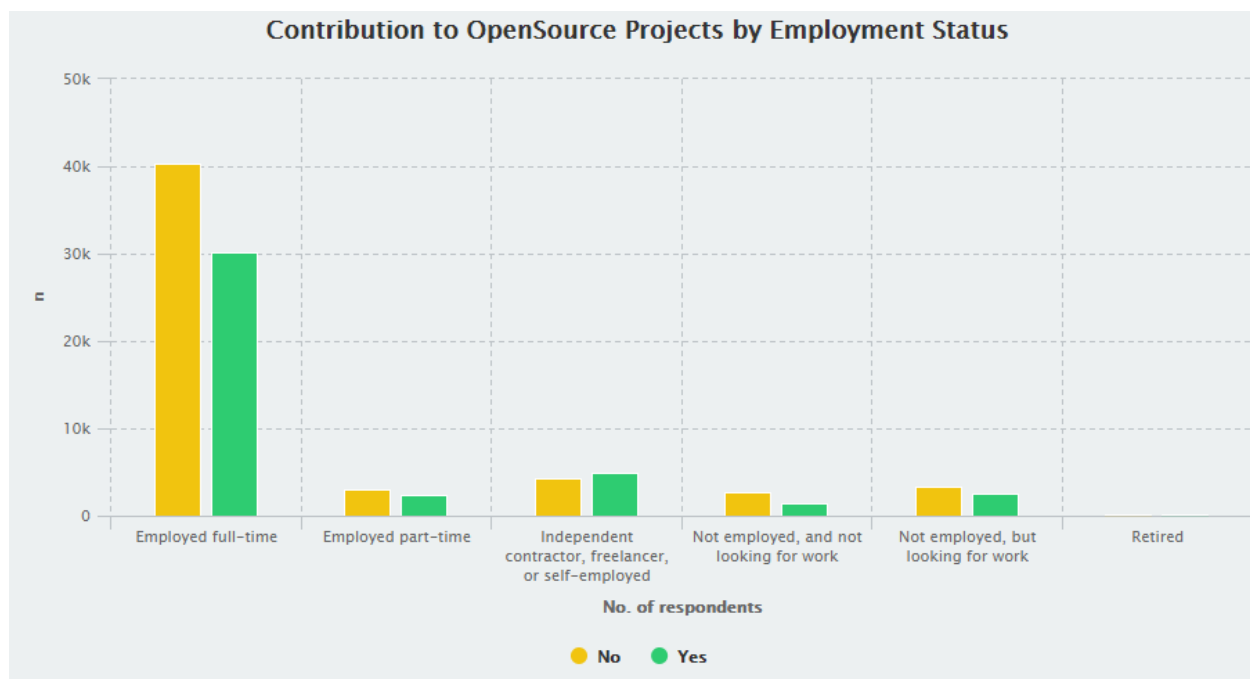
It is good to see that more number of students are coding as a hobby rather than studying it for the sake of university.

```
survey %>%
  filter(!is.na(Student)) %>%
  count(OpenSource, Student) %>%
  hchart("bar", hcaes(x=OpenSource, y=n, group=Student)) %>%
  hc_title(text="Contribution to OpenSource Projects by Student Status", style=
list(fontWeight="bold")) %>%
  hc_xAxis(title=list(text="No. of respondents", style=list(fontWeight="bold")))
) %>%
  hc_yAxis(title=list(style=list(fontWeight="bold"))) %>%
  hc_add_theme(hc_theme_darkunica())
```



It can be seen that most people who are not students contribute to open source projects, while full time students contribute to open source projects the least. This is worrying to see as contributing to open source projects may be a great way to learn for students.

```
survey %>%
  filter(!is.na(Employment)) %>%
  count(OpenSource, Employment) %>%
  hchart("column", hcaes(x=Employment, y=n, group=OpenSource)) %>%
  hc_title(text="Contribution to OpenSource Projects by Employment Status", style=list(fontWeight="bold")) %>%
  hc_xAxis(title=list(text="No. of respondents", style=list(fontWeight="bold"))) %>%
  hc_yAxis(title=list(style=list(fontWeight="bold"))) %>%
  hc_add_theme(hc_theme_flat())
```



The results are pretty clear that respondents who are employed full time do not contribute more to Open Source Projects while Freelancers and Independent contractors tend to contribute more to Open Source Projects. However, the results are puzzling that respondents who are employed part time or not employed do not contribute more to Open Source Projects.

Next, Let us take a look at the current Employment Status vs where respondents hope to be in five years.

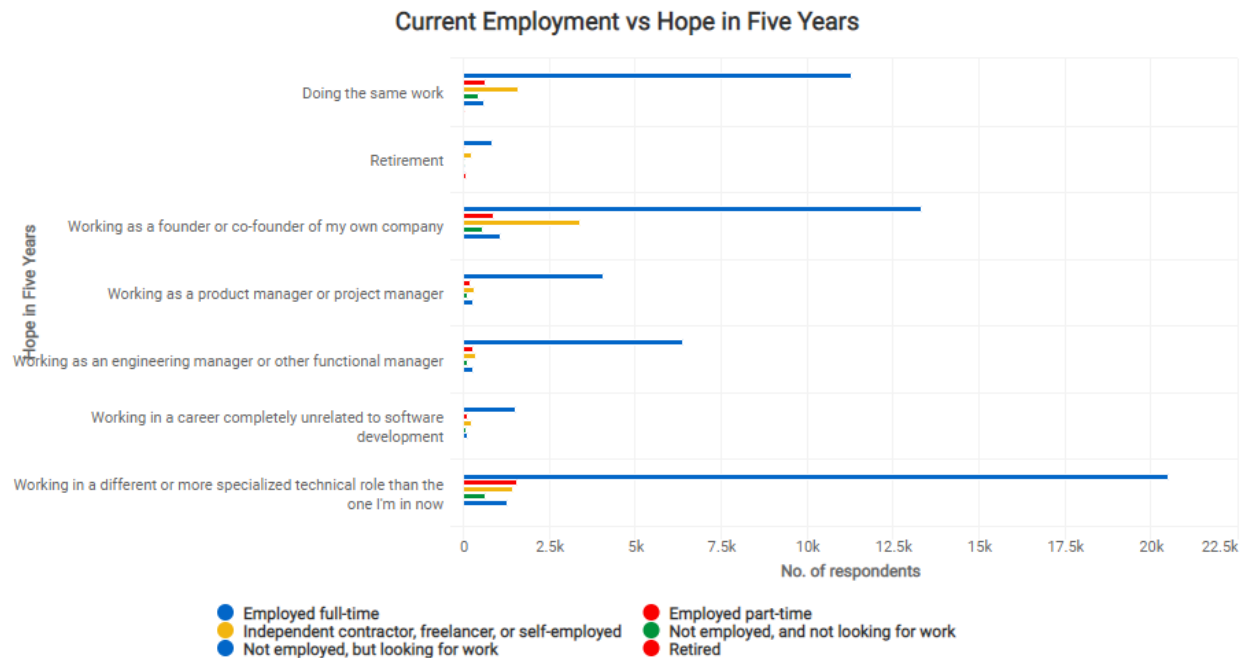
```
survey %>%
  filter(!is.na(Employment)) %>%
  filter(!is.na(HopeFiveYears)) %>%
  count(Employment, HopeFiveYears) %>%
  hchart("bar", hcaes(x=HopeFiveYears, y=n, group=Employment)) %>%
  hc_title(text="Current Employment vs Hope in Five Years", style=list(fontWeight="bold"))
```



```

ght="bold")) %>%
  hc_yAxis(title=list(text="No. of respondents",style=list(fontWeight="bold")) %>%
)) %>%
  hc_xAxis(title=list(text="Hope in Five Years",style=list(fontWeight="bold")) %>%
)) %>%
  hc_add_theme(hc_theme_google())

```



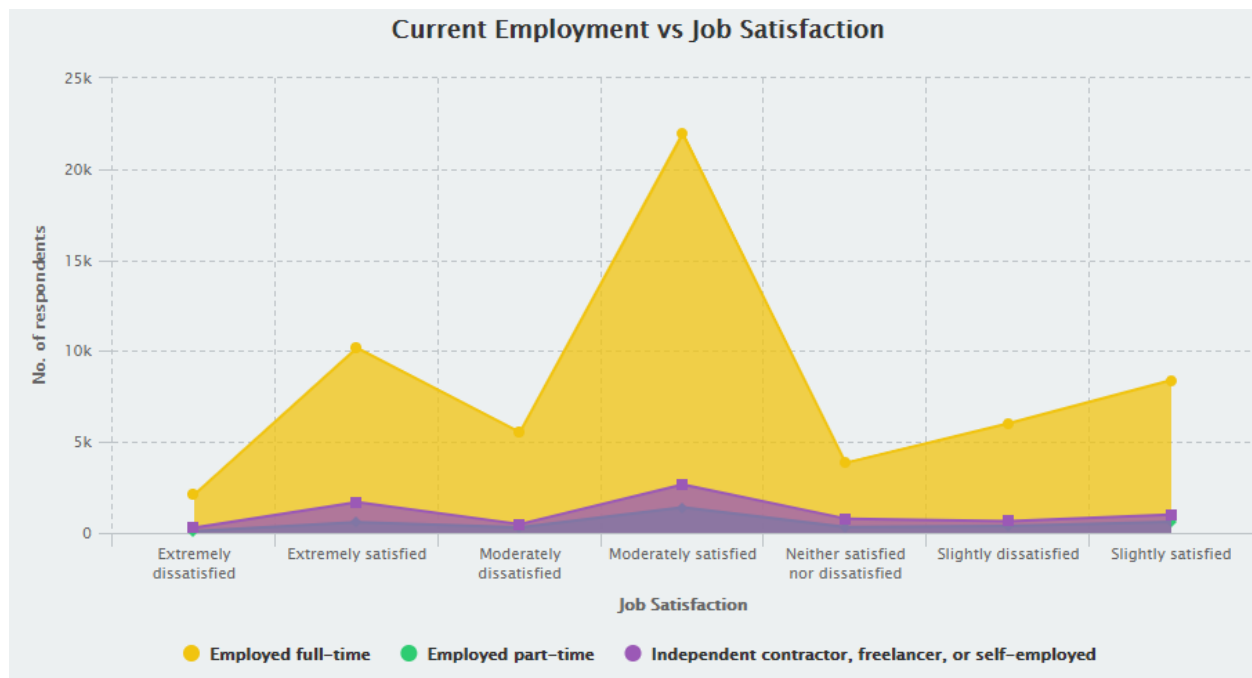
Inferences from the above chart: 1. Most people who are currently employed full time or part time hope to either work in a different/more specialized role in five years or want to find their own company. 2. Most people who are self-employed or independent contractors hope to be a founder/co-founder of their own company or expect to be doing the same work in five years, which shows that people who are self-employed or independent contractors are more satisfied with their job. 3. It is surprising that people who are currently not looking for work hope to be working in five years.

Let us look at the job-satisfaction according to current employment Status

```

survey %>%
  filter(!is.na(Employment)) %>%
  filter(!is.na(JobSatisfaction)) %>%
  count(Employment,JobSatisfaction) %>%
  hchart("area",hcaes(x=JobSatisfaction,y=n,group=Employment)) %>%
  hc_title(text="Current Employment vs Job Satisfaction",style=list(fontWeight="bold")) %>%
  hc_yAxis(title=list(text="No. of respondents",style=list(fontWeight="bold")) %>%
)) %>%
  hc_xAxis(title=list(text="Job Satisfaction",style=list(fontWeight="bold"))) %>%
  hc_add_theme(hc_theme_flat())

```



From the above area chart, we can see that majority of the respondents are either extremely or moderately satisfied with their jobs, while only a very little percentage of respondents are extremely dissatisfied with their job.

```
survey %>%
  filter(!is.na(HopeFiveYears)) %>%
  filter(!is.na(JobSatisfaction)) %>%
  count(HopeFiveYears, JobSatisfaction) %>%
  hchart("heatmap", hcaes(x=JobSatisfaction, y=HopeFiveYears, value=n)) %>%
  hc_title(text="Hope in Five Years vs Job Satisfaction", style=list(fontWeight="bold")) %>%
  hc_yAxis(title=list(text="Hope in Five years", style=list(fontWeight="bold"))) %>%
  hc_xAxis(title=list(text="Job Satisfaction", style=list(fontWeight="bold"))) %>%
  hc_add_theme(hc_theme_google())
```

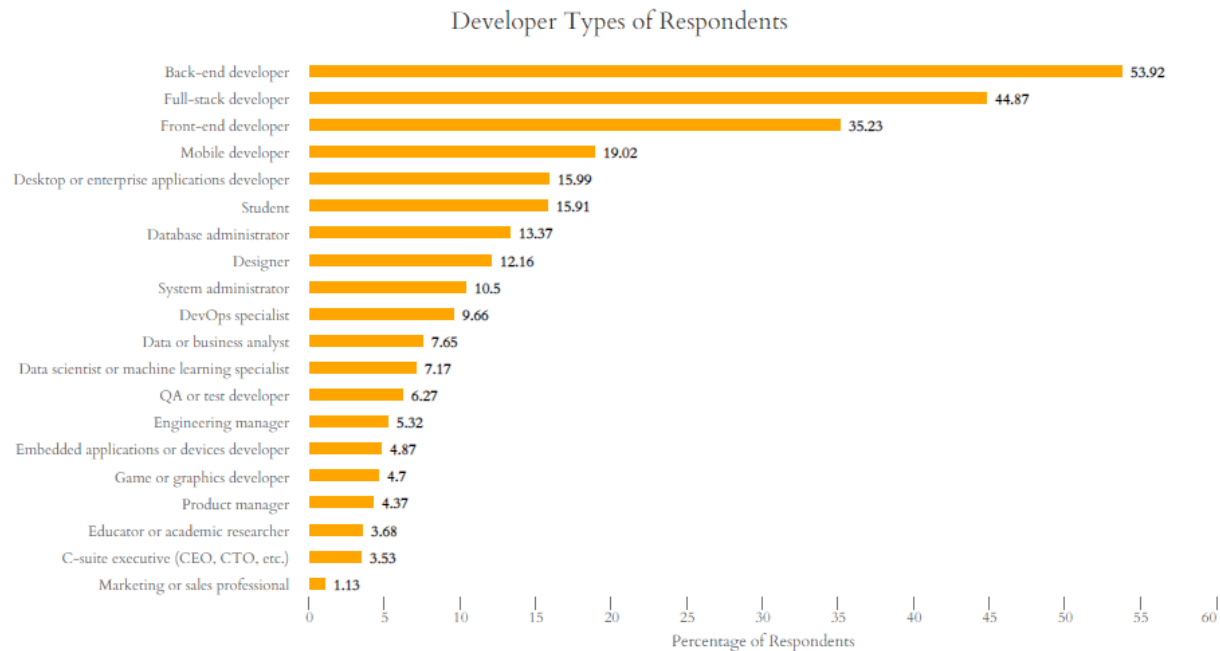


The above heat map shows us that most people who are currently satisfied with their jobs expect to be in a higher role or start their own companies.

Part II: Developer Types, Years spent coding, Preferred Languages, IDEs, Frameworks and Company Size.

Let's take a look at the type of Developers who took part in the survey

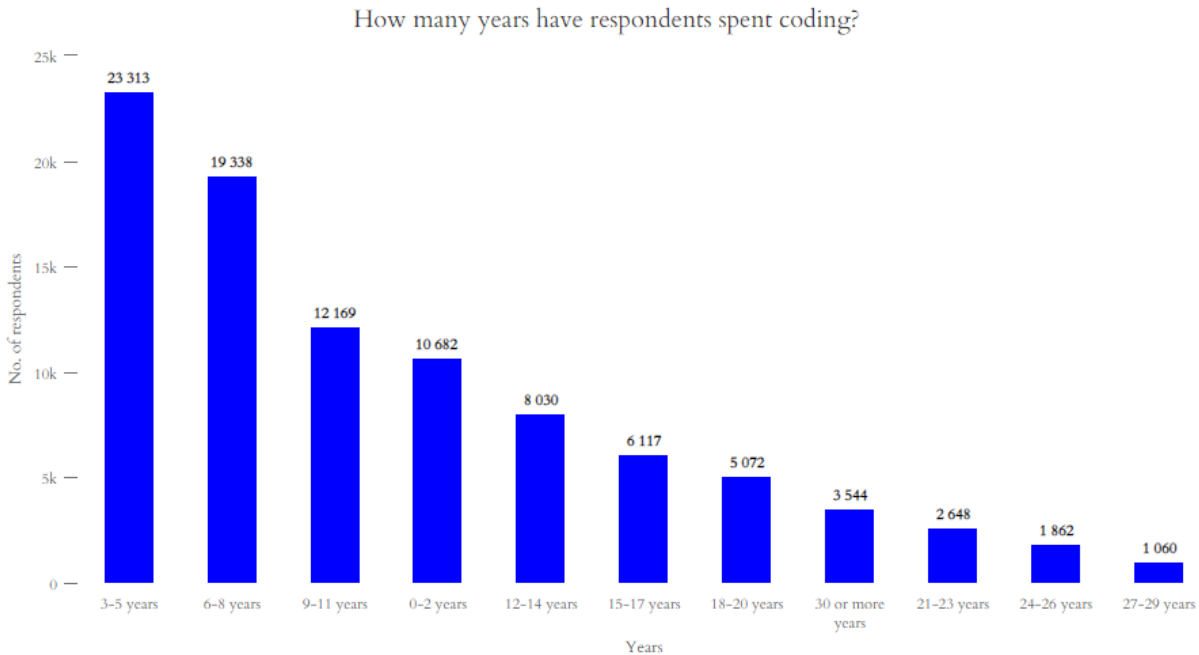
```
survey %>%
  filter(!is.na(DevType)) %>%
  select(DevType)%>%
  mutate(DevType=str_split(DevType,";")) %>%
  unnest(DevType) %>%
  group_by(DevType) %>%
  count() %>%
  arrange(desc(n))->t
t$n<- round((t$n/size)*100,2)
highchart() %>%
  hc_add_series(t,type = "bar", hcaes(x = "DevType", y = "n"), dataLabels = list(enabled = TRUE)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_title(text="Developer Types of Respondents") %>%
  hc_xAxis(categories=t$DevType) %>%
  hc_yAxis(title=list(text="Percentage of Respondents")) %>%
  hc_add_theme(hc_theme_tufte()) %>%
  hc_colors("orange")
```



We can see that a staggering 53.92% of respondents are Back-end Developers and 44.87% of respondents are Full-Stack Developers, whereas Front-end developers are third most with 35.23%. While Data Scientists and Data analysts are considered the hottest prospect of jobs, it is a bit surprising to see that only 7.17% and 7.65% of them have taken the survey.

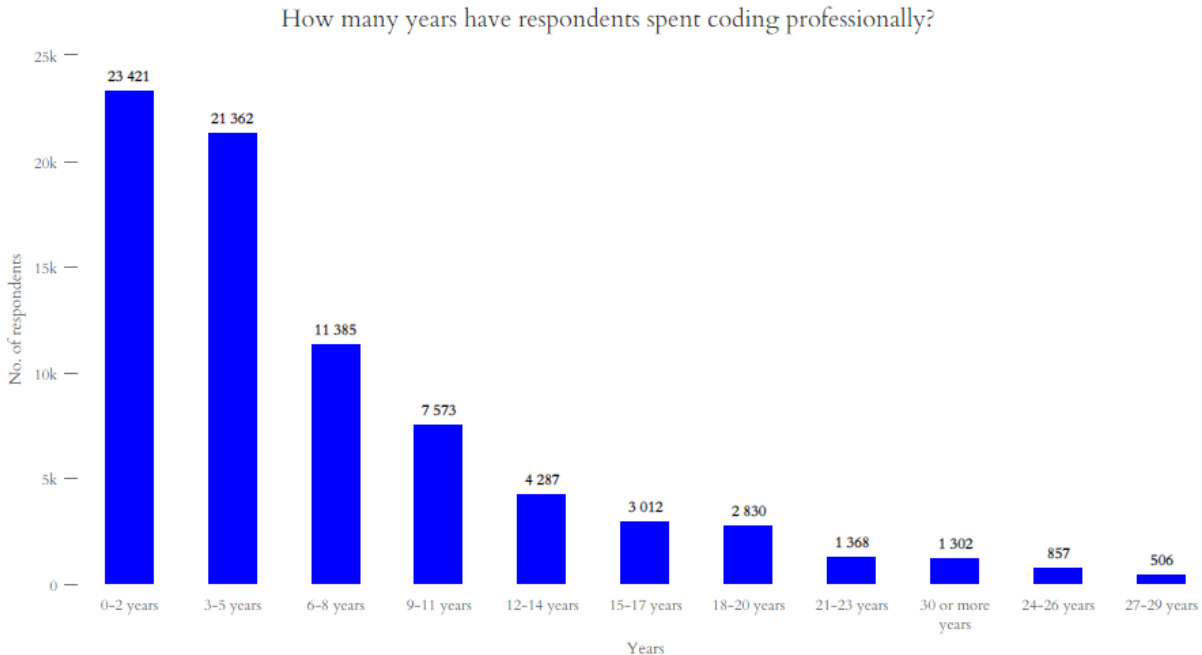
Let's take a look at how many years the respondents have spent coding.

```
survey %>%
  filter(!is.na(YearsCoding)) %>%
  group_by(YearsCoding) %>%
  count() %>%
  arrange(desc(n)) -> t
highchart() %>%
  hc_add_series(t, type = "column", hcaes(x = "YearsCoding", y = "n"), dataLabels = list(enabled = TRUE)) %>%
  hc_legend(enabled = FALSE) %>%
  hc_title(text = "How many years have respondents spent coding?") %>%
  hc_xAxis(title = list(text = "Years"), categories = t$YearsCoding) %>%
  hc_yAxis(title = list(text = "No. of respondents")) %>%
  hc_add_theme(hc_theme_tufte()) %>%
  hc_colors("blue")
```



Most of the respondents have spent coding for 3-5 years while the second and third-most are respondents who have spent coding for 6-8 years and 9-11 years. What is most surprising to see is that more than 8000 respondents (~9%) have coded for more than 20 years and still contribute to the StackOverflow Community.

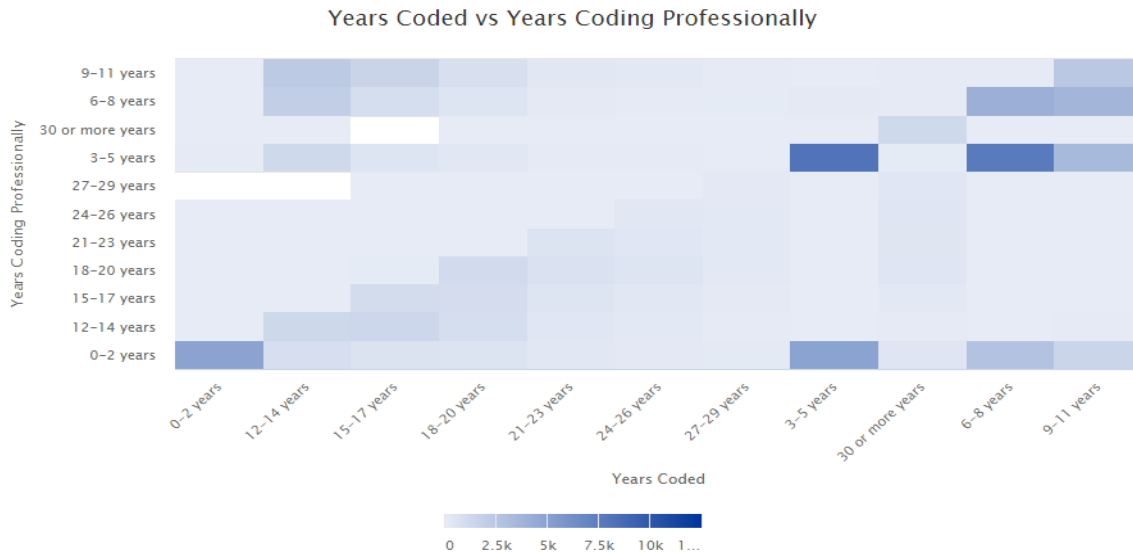
```
survey %>%
  filter(!is.na(YearsCodingProf)) %>%
  group_by(YearsCodingProf) %>%
  count() %>%
  arrange(desc(n)) -> t
highchart() %>%
  hc_add_series(t, type = "column", hcaes(x = "YearsCodingProf", y = "n"), data
Labels = list(enabled = TRUE)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_title(text="How many years have respondents spent coding professionally?") %>%
  hc_xAxis(title=list(text="Years"), categories=t$YearsCodingProf) %>%
  hc_yAxis(title=list(text="No. of respondents")) %>%
  hc_add_theme(hc_theme_tufte()) %>%
  hc_colors("blue")
```



While most respondents have coded for 3-5 years and 6-8 years as seen in the previous plot, a majority of them have coded professionally for only less than 5 years. This can indicate one of the following things:

1. A lot of people start coding during their college years and continue to code professionally.
2. A lot of people start coding only when they get a job. Let us take a look at the following plot to understand more.

```
survey %>%  
  count(YearsCoding, YearsCodingProf) %>%  
  hchart("heatmap", hcaes(x=YearsCoding, y=YearsCodingProf, value=n)) %>%  
  hc_title(text="Years Coded vs Years Coding Professionally") %>%  
  hc_xAxis(title=list(text="Years Coded")) %>%  
  hc_yAxis(title=list(text="Years Coding Professionally"))
```



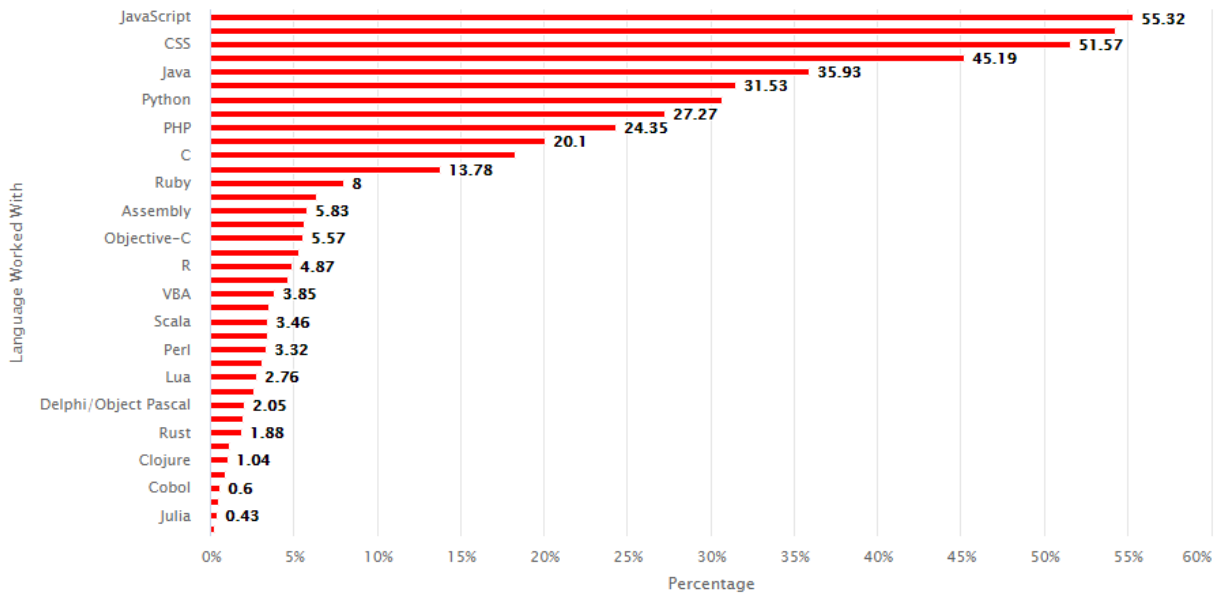
As we can see, 1. Most people who have coded for 0-2 years have coded only professionally. 2. Majority of people who have coded for 3-5 have also coded professionally for 3-5 years. 3. A lot of people who have coded for 6-8 years have also coded professionally for 6-8 years. All of this indicates that a majority of people have only ever coded professionally which means they have begun coding only after getting or starting a job.

We can also see the next majority of people who have coded for 3-5 years, 6-8 years and 9-11 years have coded professionally for 0-2 years, 3-5 years and 6-8 years professionally. This most probably indicates that respondents start coding while in college and start coding professionally immediately after graduation.

Now, Let us take a look at the Languages, Frameworks and IDEs used by Developers.

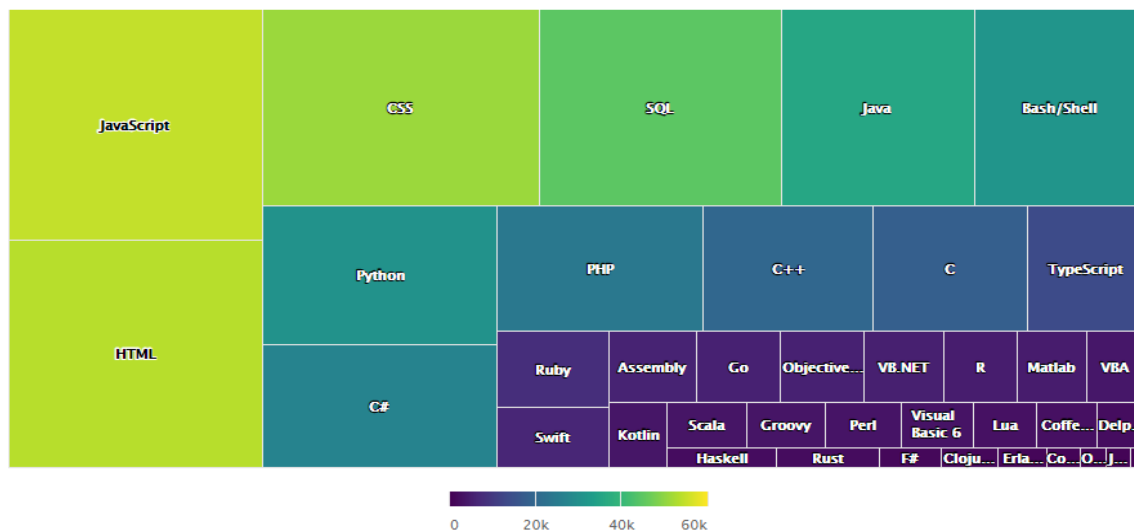
```
survey %>%
  filter(!is.na(LanguageWorkedWith)) %>%
  select(LanguageWorkedWith) %>%
  mutate(LanguageWorkedWith = str_split(LanguageWorkedWith, pattern = ";")) %>%
  unnest(LanguageWorkedWith) %>%
  group_by(LanguageWorkedWith) %>%
  count() %>%
  arrange(desc(n)) -> t
t$n <- round((t$n/size)*100,2)
highchart() %>%
  hc_add_series(t, type="bar", hcaes(x="LanguageWorkedWith", y="n"), dataLabels = list(enabled = TRUE)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_colors("red") %>%
  hc_title(text = 'Languages Developers Have Worked With') %>%
  hc_xAxis(title=list(text="Language Worked With"), categories=t$LanguageWorkedWith) %>%
  hc_yAxis(title=list(text="Percentage"), labels = list(format = "{value}%"))
```

Languages Developers Have Worked With



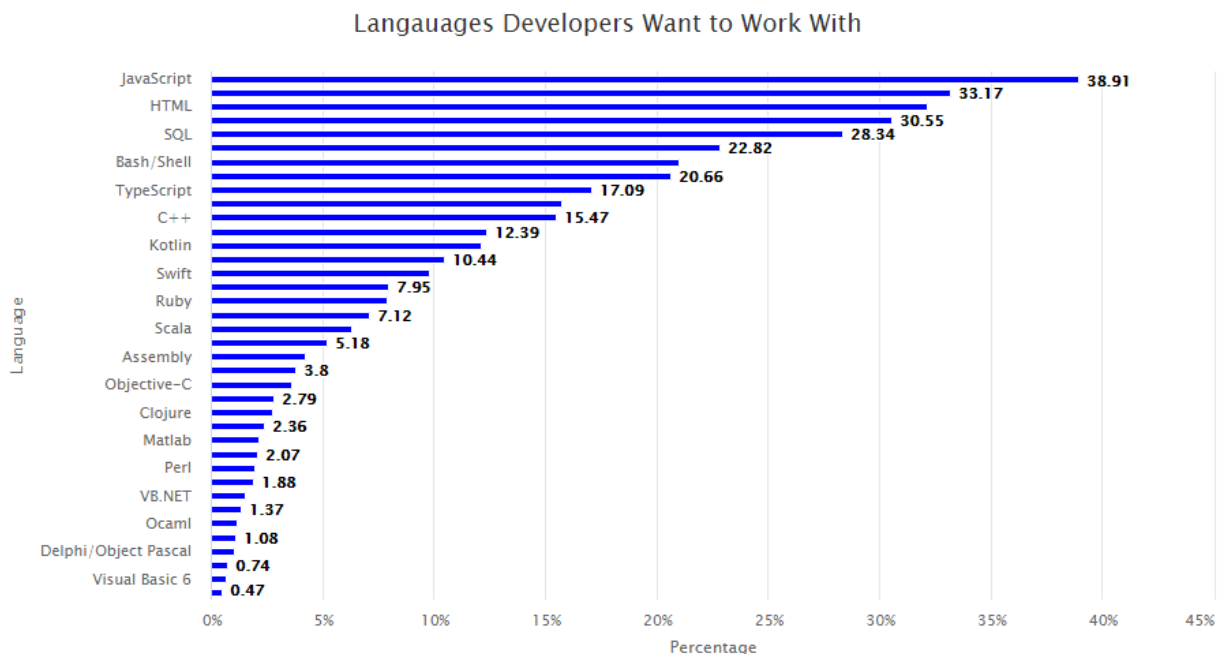
```
survey %>%
  filter(!is.na(LanguageWorkedWith)) %>%
  select(LanguageWorkedWith) %>%
  mutate(LanguageWorkedWith = str_split(LanguageWorkedWith, pattern = ";")) %
  >%
  unnest(LanguageWorkedWith) %>%
  group_by(LanguageWorkedWith) %>%
  count() %>%
  arrange(desc(n)) %>%
  hchart("treemap", hcaes(x=LanguageWorkedWith, value=n, color=n)) %>%
  hc_colorAxis(stops=color_stops(colors=viridis(10))) %>%
  hc_title(text="Languages Developers Have Worked With")
```

Languages Developers Have Worked With



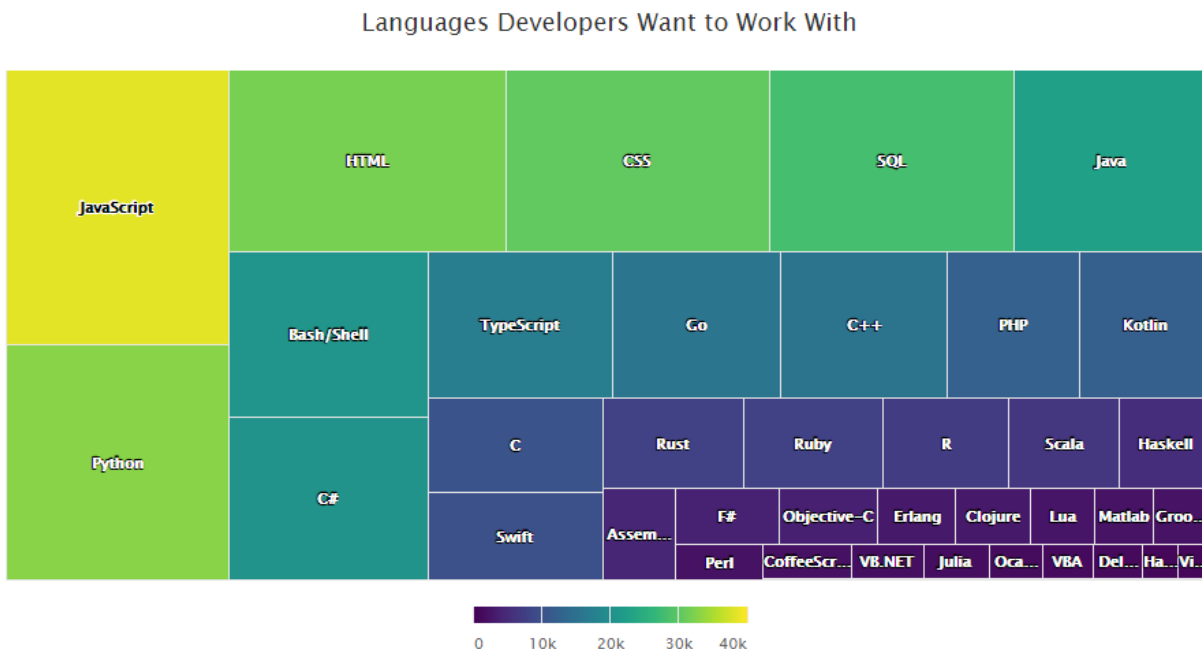
Top Languages Developers have worked with: JavaScript : 55.32% HTML : 54.25% CSS : 51.57% SQL : 45.19% Java : 35.93% Bash/Shell : 31.53% Python : 30.71 %

```
survey %>%
  filter(!is.na(LanguageDesireNextYear)) %>%
  select(LanguageDesireNextYear) %>%
  mutate(LanguageDesireNextYear = str_split(LanguageDesireNextYear, pattern =
";")) %>%
  unnest(LanguageDesireNextYear) %>%
  group_by(LanguageDesireNextYear) %>%
  count() %>%
  arrange(desc(n)) ->t
t$n <- round((t$n/size)*100,2)
highchart() %>%
  hc_add_series(t,type="bar",hcaes(x="LanguageDesireNextYear",y="n"),dataLabels
= list(enabled = TRUE)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_colors("blue") %>%
  hc_title(text = 'Langauages Developers Want to Work With') %>%
  hc_xAxis(title=list(text="Language"),categories=t$LanguageDesireNextYear) %
>%
  hc_yAxis(title=list(text="Percentage"),labels = list(format = "{value}%"))
```



```
survey %>%
  filter(!is.na(LanguageDesireNextYear)) %>%
  select(LanguageDesireNextYear) %>%
  mutate(LanguageDesireNextYear = str_split(LanguageDesireNextYear, pattern =
";")) %>%
  unnest(LanguageDesireNextYear) %>%
  group_by(LanguageDesireNextYear) %>%
```

```
count() %>%
arrange(desc(n)) %>%
hchart("treemap", hcaes(x=LanguageDesireNextYear, value=n, color=n)) %>%
hc_colorAxis(stops=color_stops(colors=viridis(10))) %>%
hc_title(text="Languages Developers Want to Work With")
```



Most of developers want to continue working with JavaScript, HTML and CSS, SQL and Bash/Shell. 33.17% of Developers would like to work with Python. Languages that developers currently don't work with a lot but would like to work with are C++, C#, Kotlin, Swift.

The tree maps give a better visualization of Languages Worked With by developers and Languages they desire to work with.

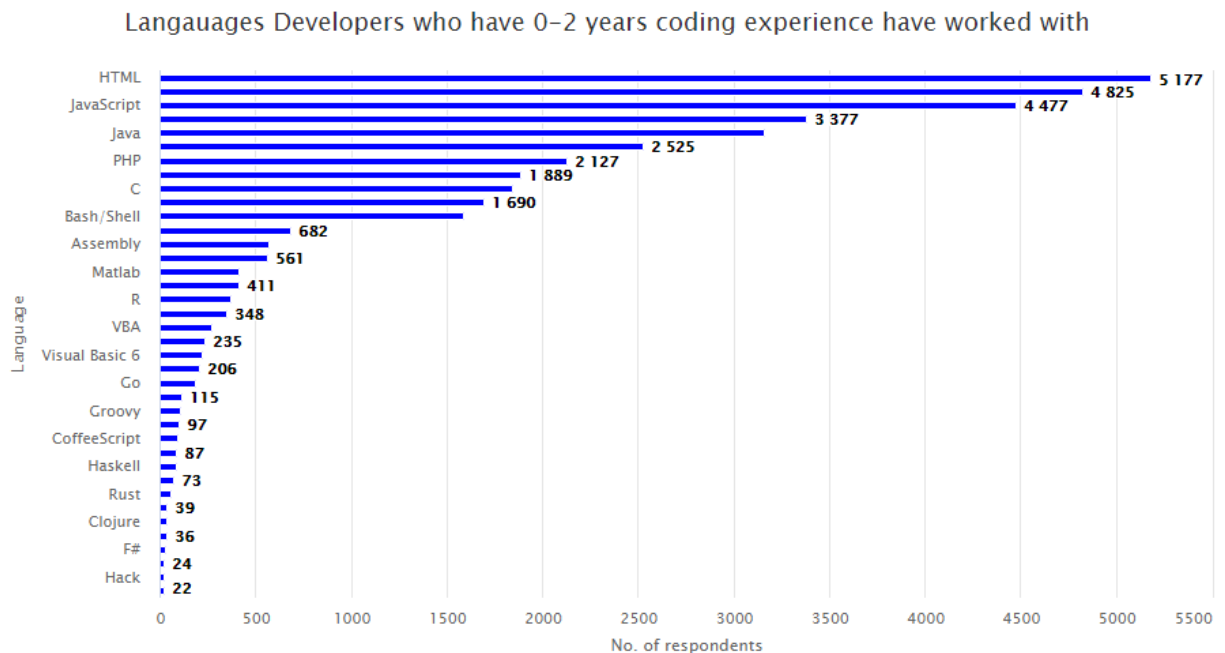
Now, let us take a look at what languages people who have just begun coding learn.

```
survey %>%
  filter(!is.na(LanguageWorkedWith)) %>%
  filter(!is.na(YearsCoding), YearsCoding %in% c("0-2 years")) %>%
  select(LanguageWorkedWith) %>%
  mutate(LanguageWorkedWith = str_split(LanguageWorkedWith, pattern = ";")) %>%
  unnest(LanguageWorkedWith) %>%
  group_by(LanguageWorkedWith) %>%
  count() %>%
  arrange(desc(n)) -> t
highchart() %>%
hc_add_series(t, type="bar", hcaes(x="LanguageWorkedWith", y="n"), dataLabels = list(enabled = TRUE)) %>%
hc_legend(enabled=FALSE) %>%
```

```

hc_colors("blue") %>%
hc_title(text = 'Langauages Developers who have 0-2 years coding experience
have worked with') %>%
hc_xAxis(title=list(text="Language"),categories=t$LanguageWorkedWith) %>%
hc_yAxis(title=list(text="No. of respondents"))

```



This shows that a lot of people who have just begun coding start with HTML/CSS , JavaScript, SQL and Java.

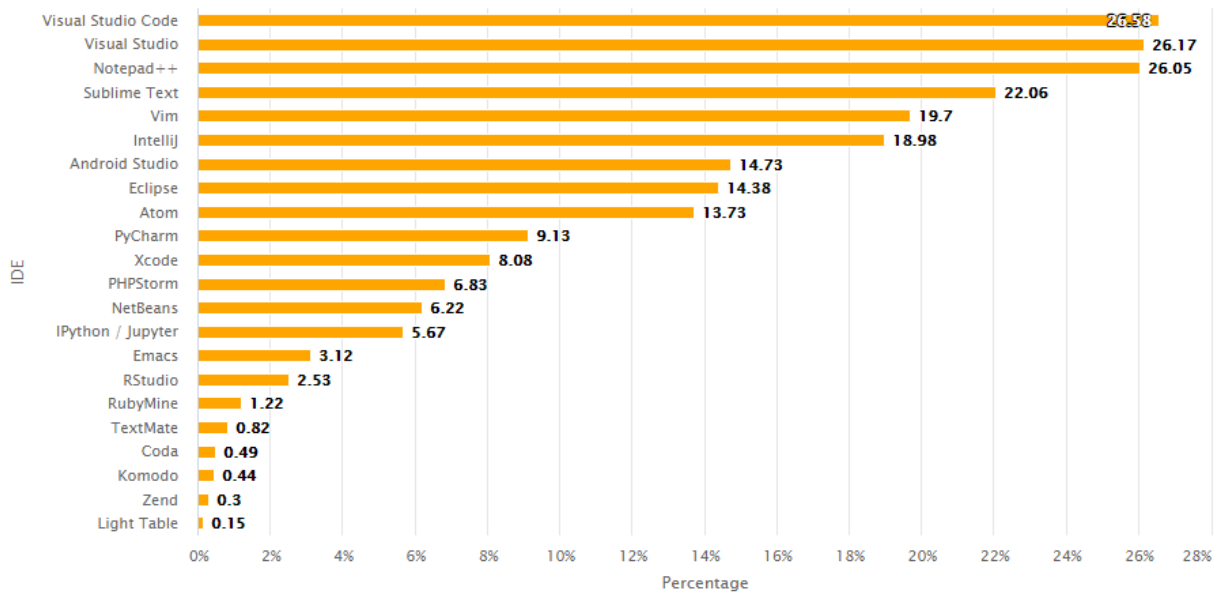
Let us now take a look at IDE Preferences of Developers.

```

survey %>%
  filter(!is.na(IDE)) %>%
  select(IDE) %>%
  mutate(IDE = str_split(IDE, pattern = ";")) %>%
  unnest(IDE) %>%
  group_by(IDE) %>%
  count() %>%
  arrange(desc(n)) -> t
t$n <- round((t$n/size)*100,2)
highchart() %>%
hc_add_series(t,type="bar",hcaes(x="IDE",y="n"),dataLabels = list(enabled = T
RUE)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_colors("orange") %>%
  hc_title(text = 'IDE Preferences of Developers') %>%
  hc_xAxis(title=list(text="IDE"),categories=t$IDE) %>%
  hc_yAxis(title=list(text="Percentage"),labels = list(format = "{value}%"))

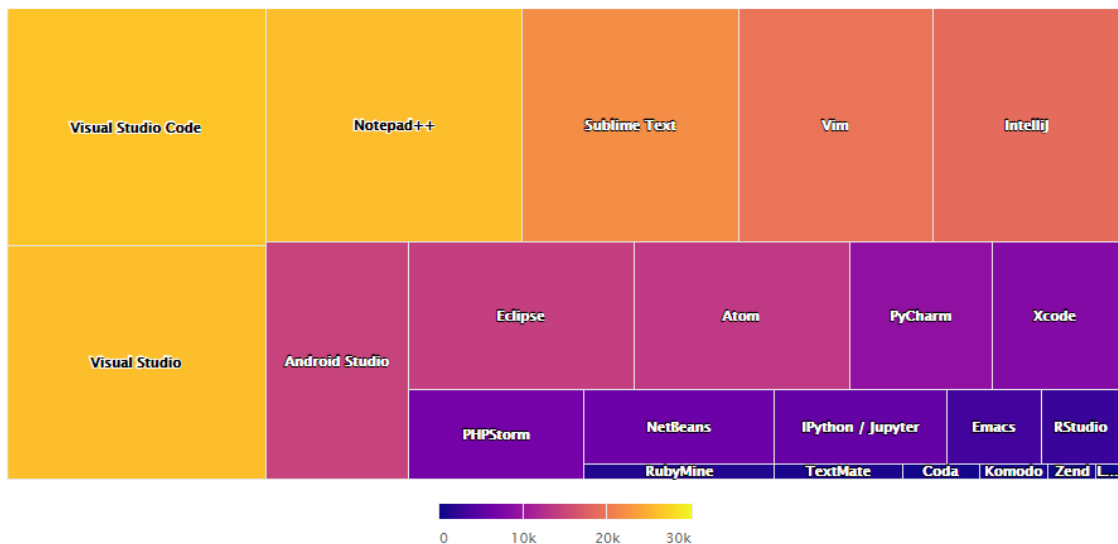
```

IDE Preferences of Developers



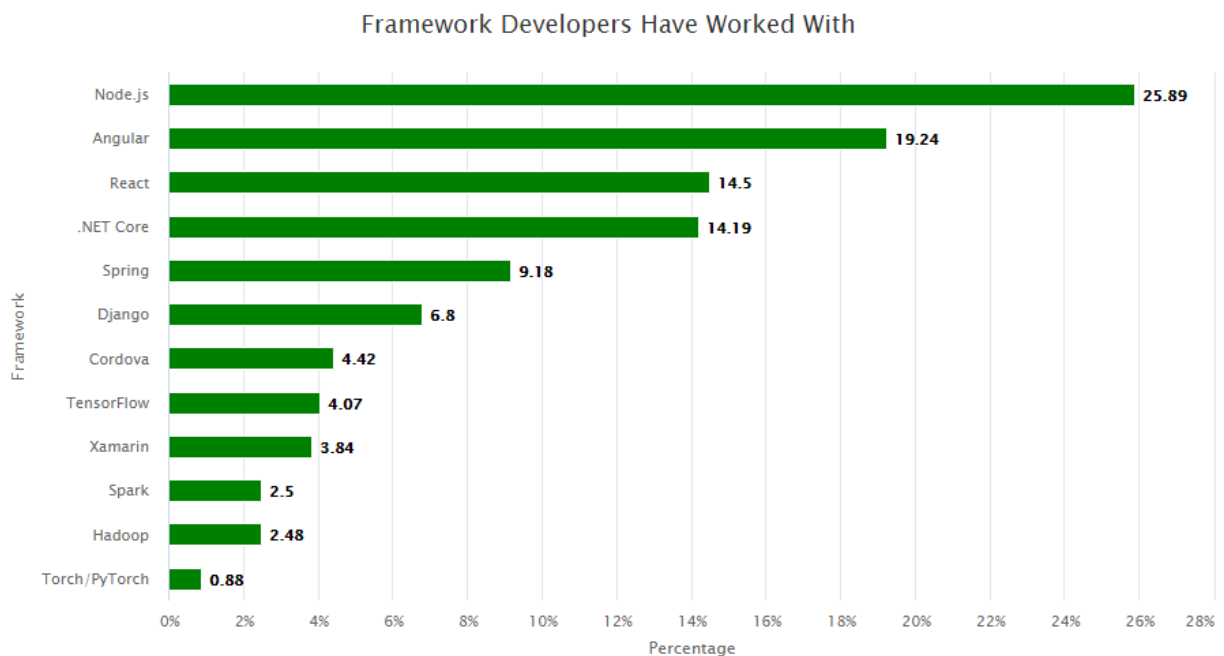
```
survey %>%
  filter(!is.na(IDE)) %>%
  select(IDE) %>%
  mutate(IDE = str_split(IDE, pattern = ";")) %>%
  unnest(IDE) %>%
  group_by(IDE) %>%
  count() %>%
  arrange(desc(n)) %>%
  hchart("treemap", hcaes(x=IDE, value=n, color=n)) %>%
  hc_colorAxis(stops=color_stops(colors=plasma(10))) %>%
  hc_title(text="IDE Preferences of Developers")
```

IDE Preferences of Developers



Majority of Developers prefer Visual Studio Code, Visual Studio and Notepad++ while Sublime Text, Vim and IntelliJ remain popular options.

```
survey %>%
  filter(!is.na(FrameworkWorkedWith)) %>%
  select(FrameworkWorkedWith) %>%
  mutate(FrameworkWorkedWith = str_split(FrameworkWorkedWith, pattern = ";"))
%>%
  unnest(FrameworkWorkedWith) %>%
  group_by(FrameworkWorkedWith) %>%
  count() %>%
  arrange(desc(n)) -> t
t$n <- round((t$n/size)*100,2)
highchart() %>%
  hc_add_series(t,type="bar",hcaes(x="FrameworkWorkedWith",y="n"),dataLabels =
list(enabled = TRUE)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_colors("green") %>%
  hc_title(text = 'Framework Developers Have Worked With') %>%
  hc_xAxis(title=list(text="Framework"),categories=t$FrameworkWorkedWith) %>%
  hc_yAxis(title=list(text="Percentage"),labels = list(format = "{value}%"))
```

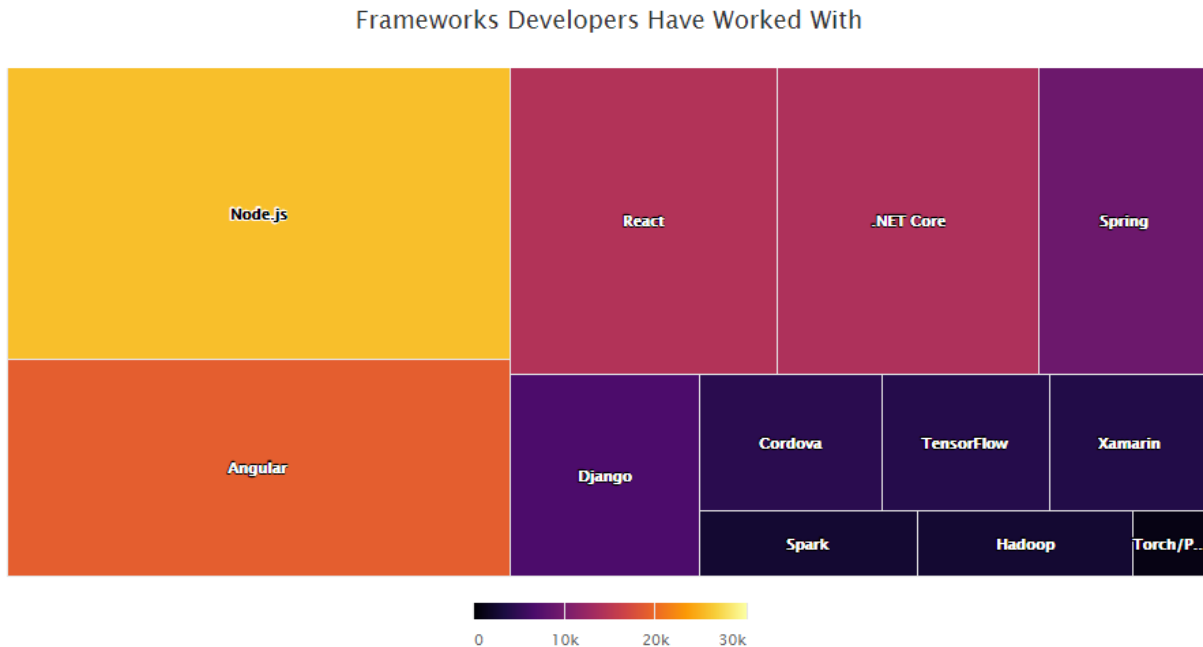


```
survey %>%
  filter(!is.na(FrameworkWorkedWith)) %>%
  select(FrameworkWorkedWith) %>%
  mutate(FrameworkWorkedWith = str_split(FrameworkWorkedWith, pattern = ";"))
%>%
  unnest(FrameworkWorkedWith) %>%
  group_by(FrameworkWorkedWith) %>%
  count() %>%
```

```

arrange(desc(n)) %>%
hchart("treemap", hcaes(x=FrameworkWorkedWith, value=n, color=n)) %>%
hc_colorAxis(stops=color_stops(colors=inferno(10))) %>%
hc_title(text="Frameworks Developers Have Worked With")

```



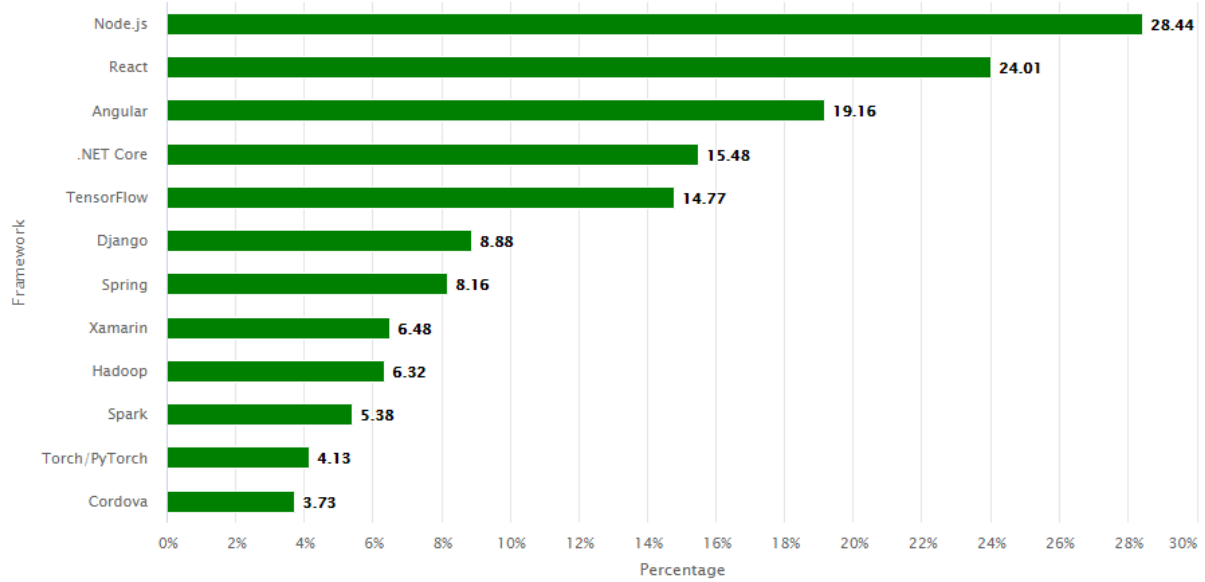
Considering that most people have worked with JavaScript,HTML and CSS, Node.js, Angular and React are the most popular Frameworks developers have worked with as they are all related to web development.

```

survey %>%
  filter(!is.na(FrameworkDesireNextYear)) %>%
  select(FrameworkDesireNextYear) %>%
  mutate(FrameworkDesireNextYear = str_split(FrameworkDesireNextYear, pattern
= ";")) %>%
  unnest(FrameworkDesireNextYear) %>%
  group_by(FrameworkDesireNextYear) %>%
  count() %>%
  arrange(desc(n)) ->t
t$n <- round((t$n/size)*100,2)
highchart() %>%
hc_add_series(t, type="bar", hcaes(x="FrameworkDesireNextYear", y="n"), dataLabel
s = list(enabled = TRUE)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_colors("green") %>%
  hc_title(text = 'Framework Developers Want to Work With') %>%
  hc_xAxis(title=list(text="Framework"), categories=t$FrameworkDesireNextYear)
%>%
  hc_yAxis(title=list(text="Percentage"), labels = list(format = "{value}%"))

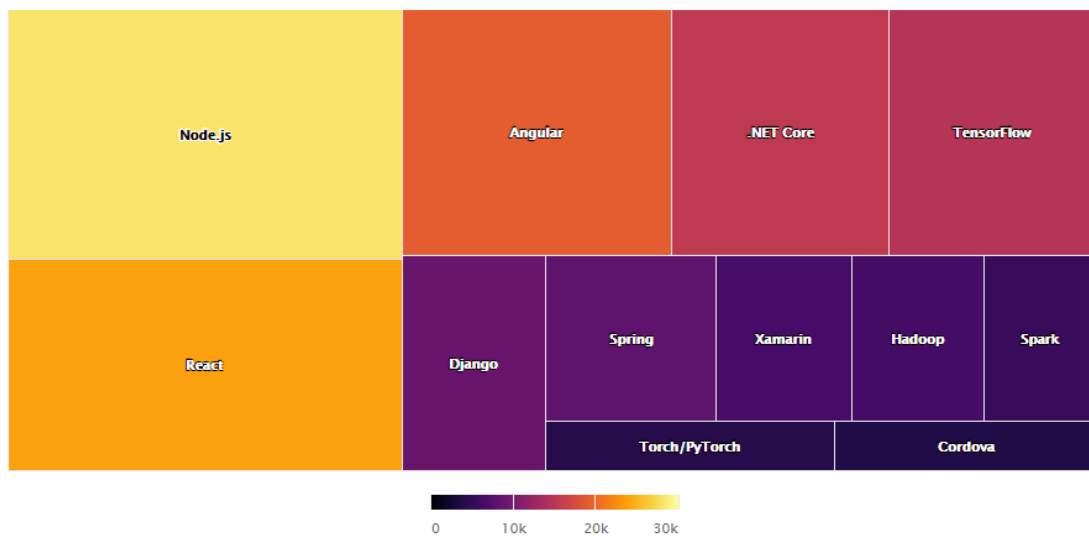
```

Framework Developers Want to Work With



```
survey %>%
  filter(!is.na(FrameworkDesireNextYear)) %>%
  select(FrameworkDesireNextYear) %>%
  mutate(FrameworkDesireNextYear = str_split(FrameworkDesireNextYear, pattern
= ";")) %>%
  unnest(FrameworkDesireNextYear) %>%
  group_by(FrameworkDesireNextYear) %>%
  count() %>%
  arrange(desc(n)) %>%
  hchart("treemap", hcaes(x=FrameworkDesireNextYear, value=n, color=n)) %>%
  hc_colorAxis(stops=color_stops(colors=inferno(10))) %>%
  hc_title(text="Frameworks Developers want to Work With")
```

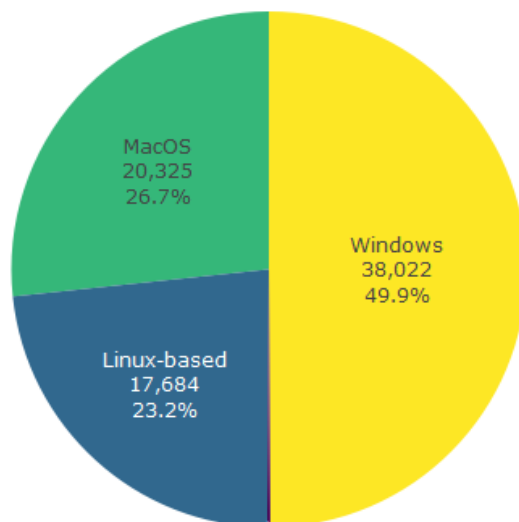
Frameworks Developers want to Work With



Node.js still remains at the top, but more developers want to explore React and Angular. TensorFlow is the Framework that a lot of people haven't worked with yet, but would like to work with a lot. I believe this is due to increasing ML research and opportunities.

```
survey %>%
  filter(!is.na(OperatingSystem)) %>%
  select(OperatingSystem) %>%
  mutate(OperatingSystem = str_split(OperatingSystem, pattern = ";")) %>%
  unnest(OperatingSystem) %>%
  group_by(OperatingSystem) %>%
  count() %>%
  plot_ly(type="pie",
          labels=~OperatingSystem,
          values=~n,
          textposition="inside",
          textinfo='label+percent+value',
          showlegend=FALSE,
          marker=list(colors = viridis(4)))
)%>%
  layout(title="Operating System Preference of Developers")
```

Operating System Preference of Developers



Windows still remains the popular Operating System Developers work with.

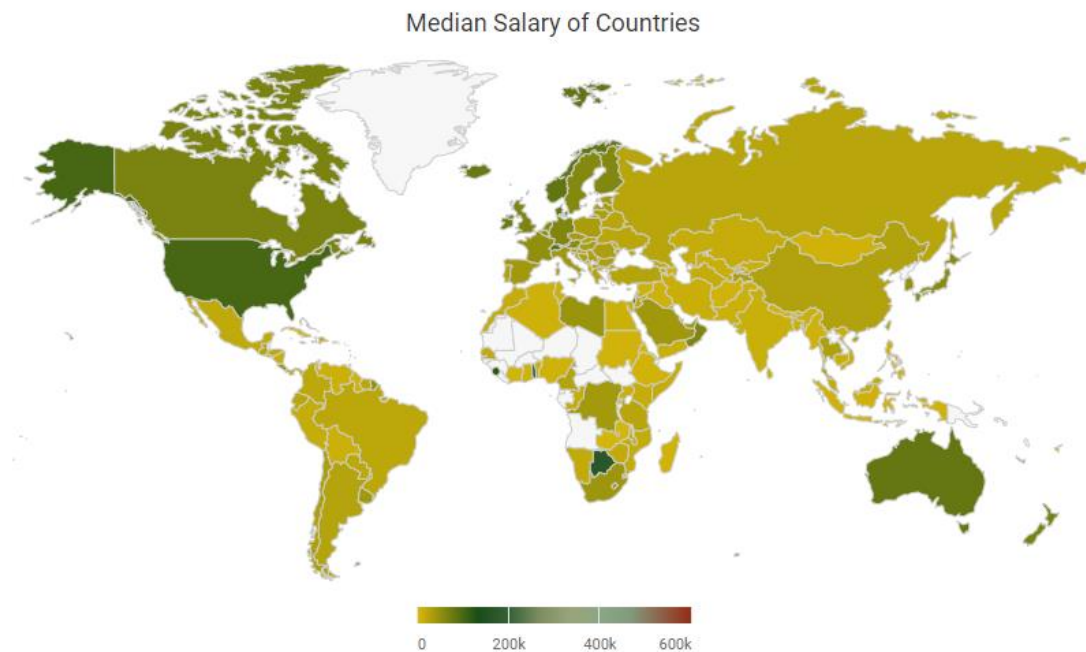
Part III: Salary Analysis

```
survey %>%
  filter(!is.na(Country)) %>%
  filter(!is.na(ConvertedSalary)) %>%
  group_by(Country) %>%
  summarise(Median_salary=median(ConvertedSalary,na.rm = TRUE)) ->t
code <- countrycode(t$Country, 'country.name', 'iso3c')
```

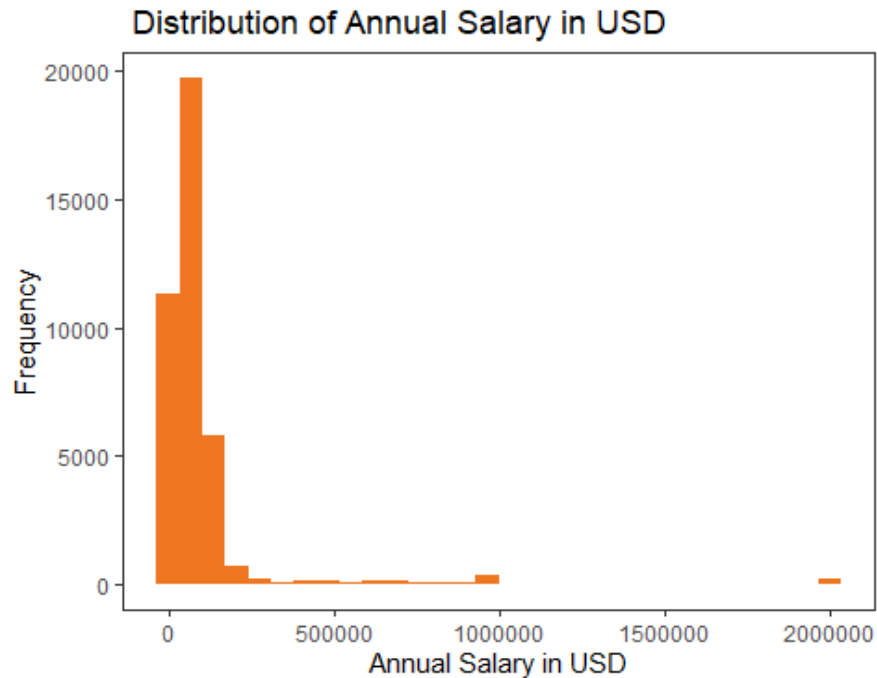


```
## Warning in countrycode(t$Country, "country.name", "iso3c"): Some values we
re not matched unambiguously: Other Country (Not Listed Above)

t$iso3 <- code
highchart() %>%
  hc_add_series_map(worldgeojson, t, value = "Median_salary", joinBy = "iso
3") %>%
  hc_colorAxis(stops=color_stops(colors=wes_palette("Cavalcanti1", 10, type
= "continuous"))) %>%
  hc_title(text="Median Salary of Countries") %>%
  hc_add_theme(hc_theme_google())
```

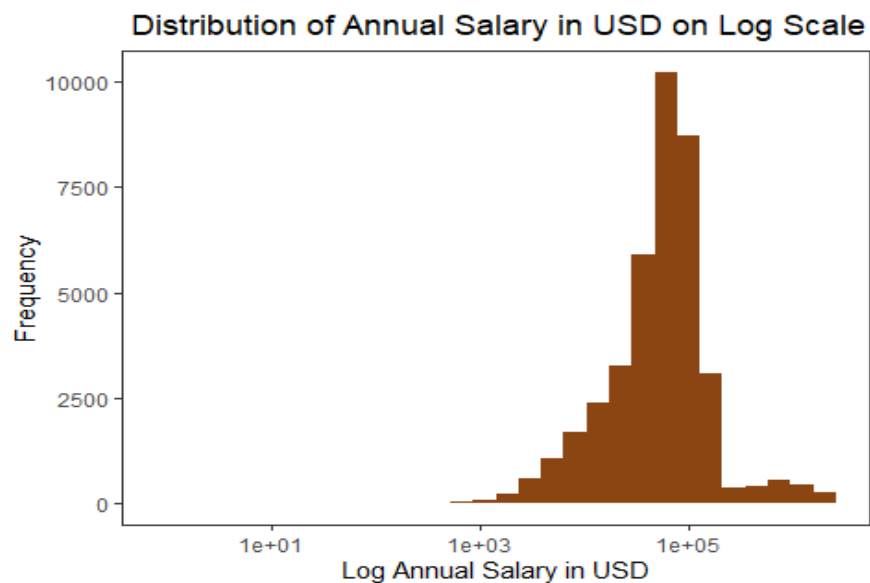


```
survey %>%
  filter(Employment %in% 'Employed full-time') %>%
  ggplot() +
  geom_histogram(aes(ConvertedSalary), fill = "chocolate2") +
  labs(x = "Annual Salary in USD", y = "Frequency",
       title = " Distribution of Annual Salary in USD") + theme_test() -> sal
sal
```



Note : Only the Salary of full time employees have been taken for the above distribution.

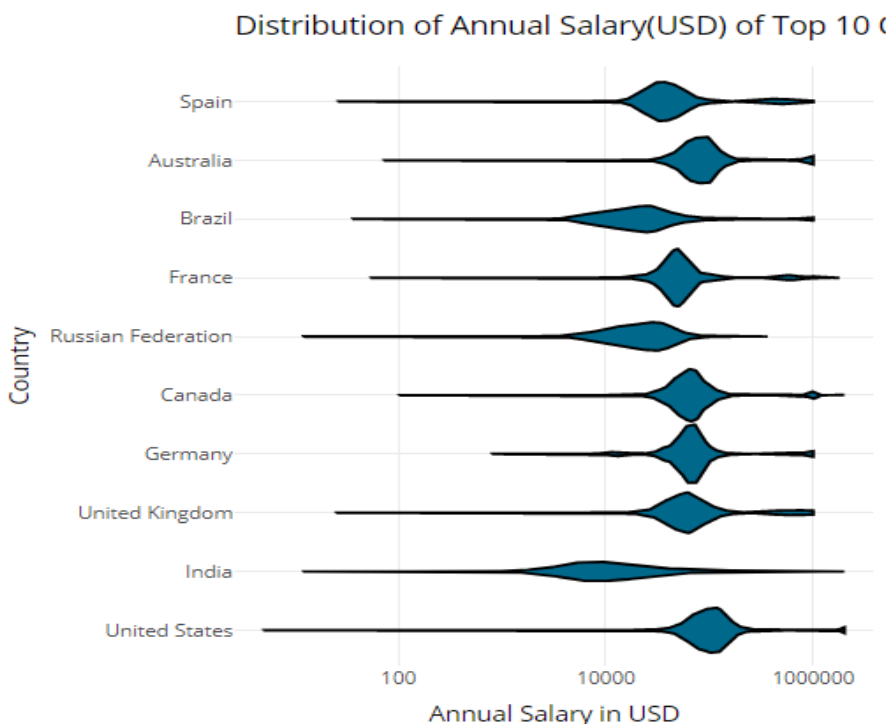
```
survey %>%
  filter(Employment %in% 'Employed full-time') %>%
  ggplot() +
  geom_histogram(aes(ConvertedSalary), fill = "chocolate4") +
  scale_x_log10() +
  labs(x = "Log Annual Salary in USD", y = "Frequency",
       title = "Distribution of Annual Salary in USD on Log Scale") + theme_
test() -> sal
sal
```



As we have seen in the beginning of the project, the most respondents have come from the following countries (Top 10) : United States, India, Germany, United Kingdom, Canada, Russia, France, Brazil, Poland, Australia (in order).

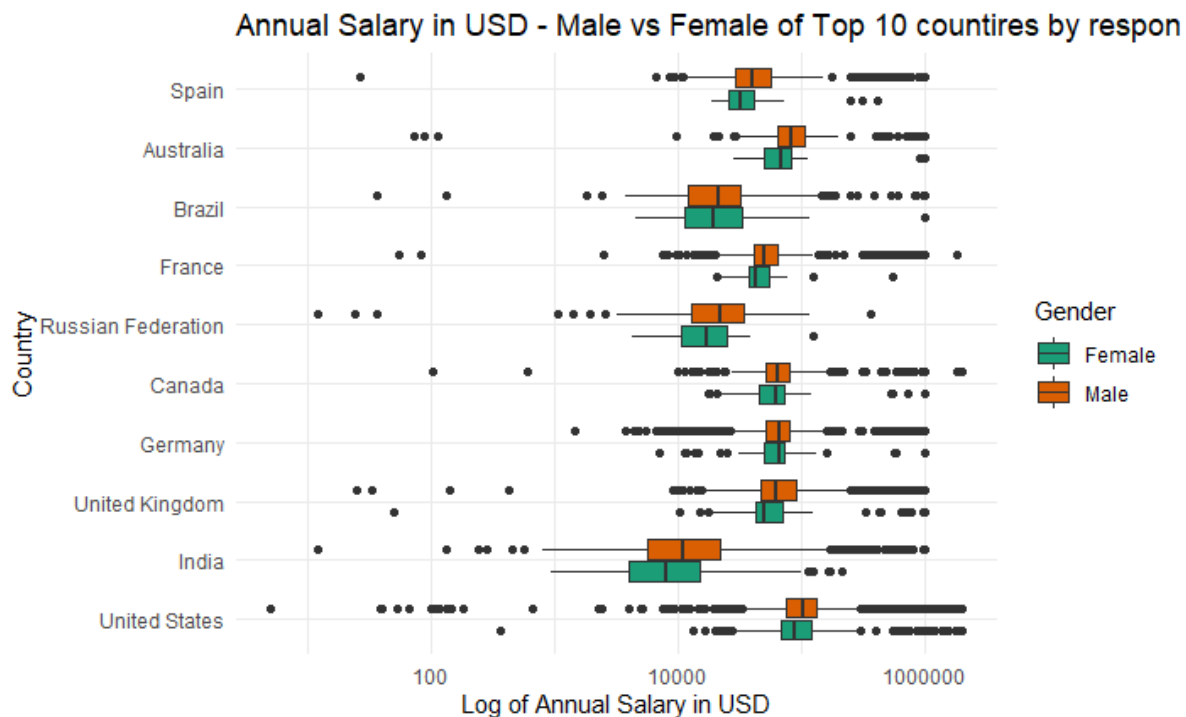
Let us see the distribution of people from these countries alone.

```
survey %>% filter(Employment %in% 'Employed full-time') %>%
  group_by(Country) %>%
  count() %>%
  arrange(desc(n)) %>%
  head(10) %>%
  select(Country) %>% mutate(Country = factor(Country)) -> countries
options(scipen=999)
survey %>% filter(Employment %in% 'Employed full-time') %>%
  group_by(Country) %>%
  mutate(n = n()) %>%
  filter(n > 1355) %>%
  ungroup(Country) %>%
  ggplot() +
  geom_violin(aes(Country,ConvertedSalary), fill= "deepskyblue4", color = "black") + theme_minimal()+
  scale_x_discrete(limits = countries$Country) +
  coord_flip() +
  scale_y_log10() +
  labs(x = "Country", y = "Annual Salary in USD",
       title = "Distribution of Annual Salary(USD) of Top 10 Countries of respondents") -> p
ggplotly(p)
```



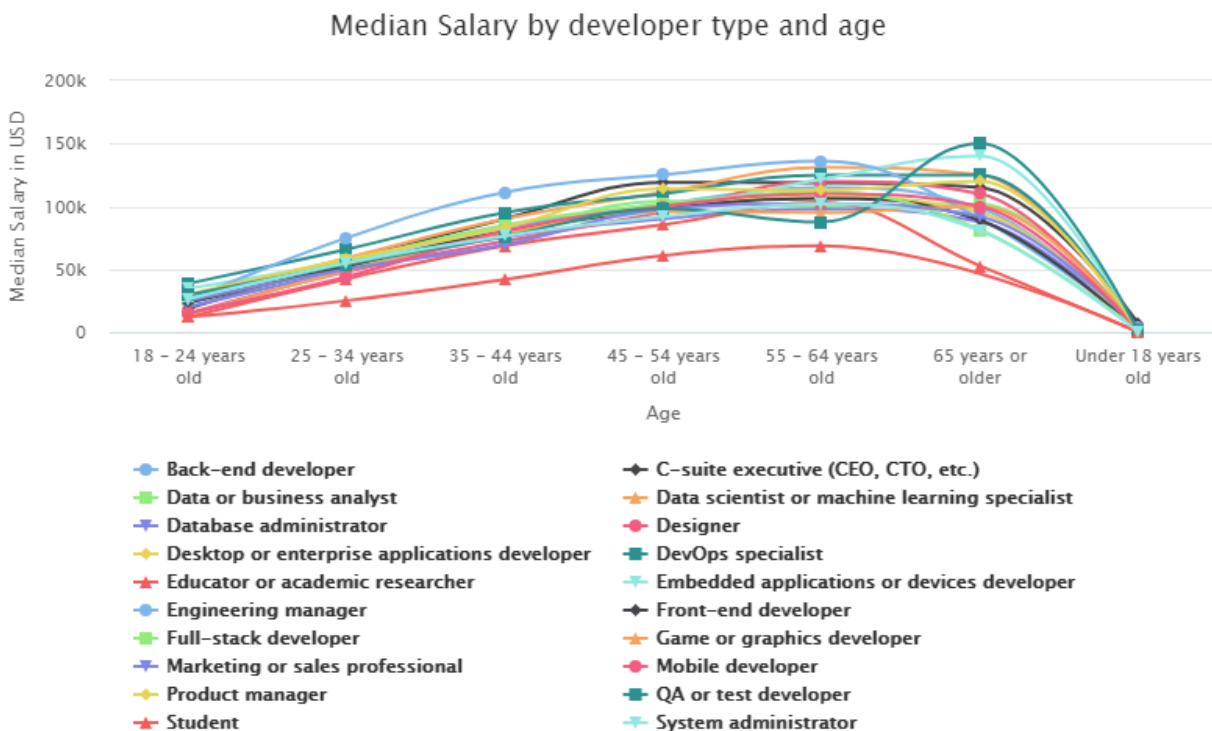
From the above countries, we can see that the median salary for United States is the highest and for India is the lowest. Along with that, India and United States have the highest variance in terms of Annual Salary.

```
options(scipen=999)
survey %>% filter(Employment %in% 'Employed full-time') %>%
  filter(Gender %in% c('Male','Female')) %>%
  group_by(Country) %>%
  mutate(n = n()) %>%
  ungroup(Country) %>%
  ggplot() +
  geom_boxplot(aes(Country,ConvertedSalary, fill = Gender)) + theme_minimal(
) +
  scale_fill_brewer(palette="Dark2") +
  scale_x_discrete(limits = countries$Country) +
  coord_flip() +
  scale_y_log10() +
  labs(x = "Country", y = "Log of Annual Salary in USD",
       title = "Annual Salary in USD - Male vs Female of Top 10 countires by respondents")
```



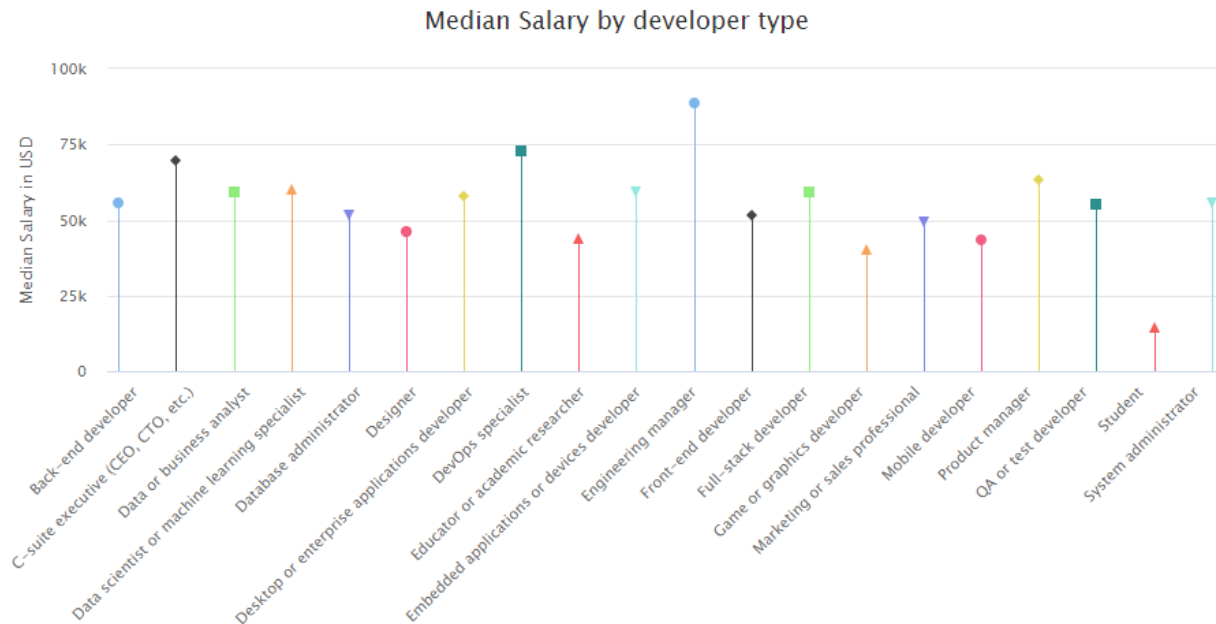
From the above box plot, we can observe similar results as the violin plot above but differences in salaries of genders can be seen.

```
survey%>%
  select(DevType,ConvertedSalary,Age)%>%
  filter(!is.na(DevType),!is.na(ConvertedSalary),!is.na(Age)) %>%
  mutate(DevType=str_split(DevType,";"))%>%
  unnest(DevType)%>%group_by(Age,DevType)%>%
  summarise(avg_salary=round(median(ConvertedSalary),0))%>%
  ungroup()%>%
  hchart("spline",hcaes(x=Age,y=avg_salary,group=DevType)) %>%
  hc_title(text="Median Salary by developer type and age") %>%
  hc_xAxis(title=list(text="Age")) %>%
  hc_yAxis(title=list(text="Median Salary in USD"))
```

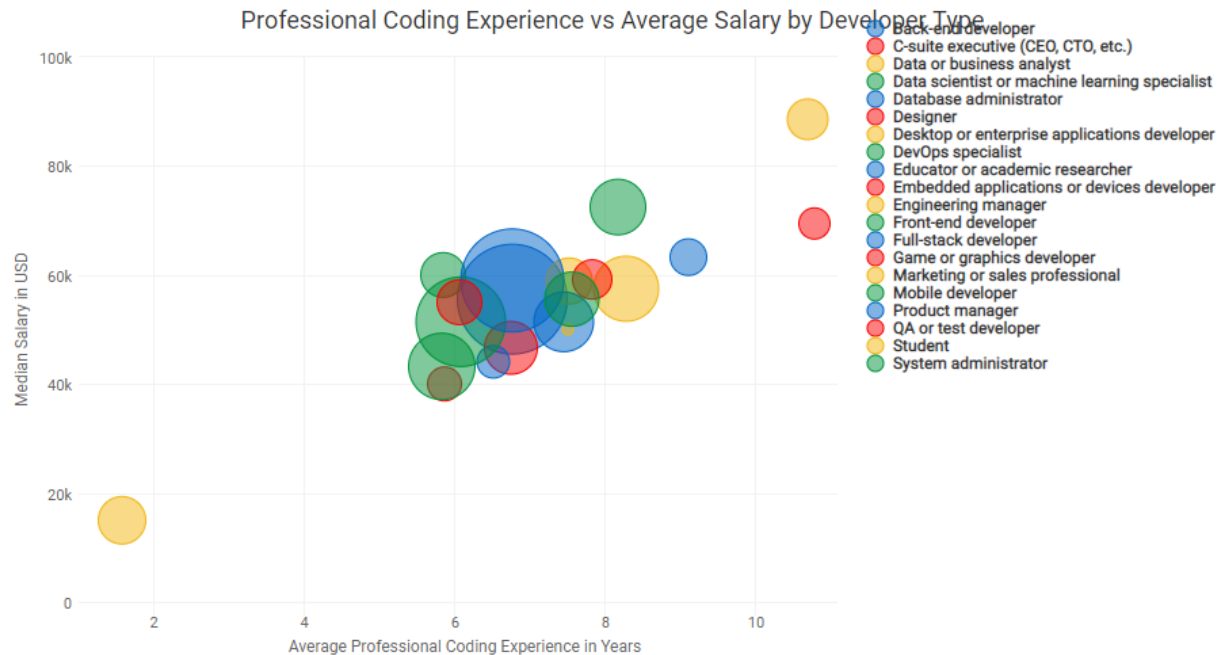


```
survey%>%
  select(DevType,ConvertedSalary)%>%
  filter(!is.na(DevType),!is.na(ConvertedSalary))%>%
  mutate(DevType=str_split(DevType,";"))%>%
  unnest(DevType)%>%
  group_by(DevType)%>%
  summarise(avg_salary=round(median(ConvertedSalary),0))%>%
  arrange(desc(avg_salary))%>%
  ungroup()%>% arrange(desc(avg_salary)) %>%
  hchart("lollipop",hcaes(x=DevType,y=avg_salary,group=DevType,size=avg_salary)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_title(text="Median Salary by developer type") %>%
```

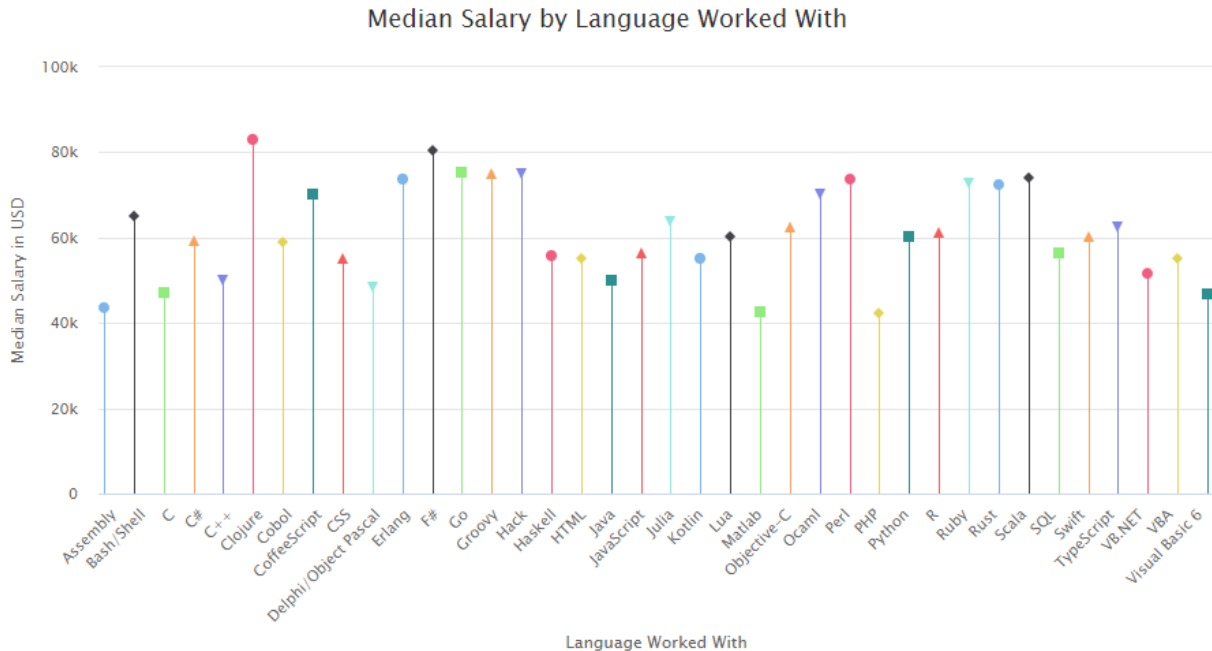
```
hc_xAxis(title=list(text="Developer Type")) %>%
hc_yAxis(title=list(text="Median Salary in USD"))
```



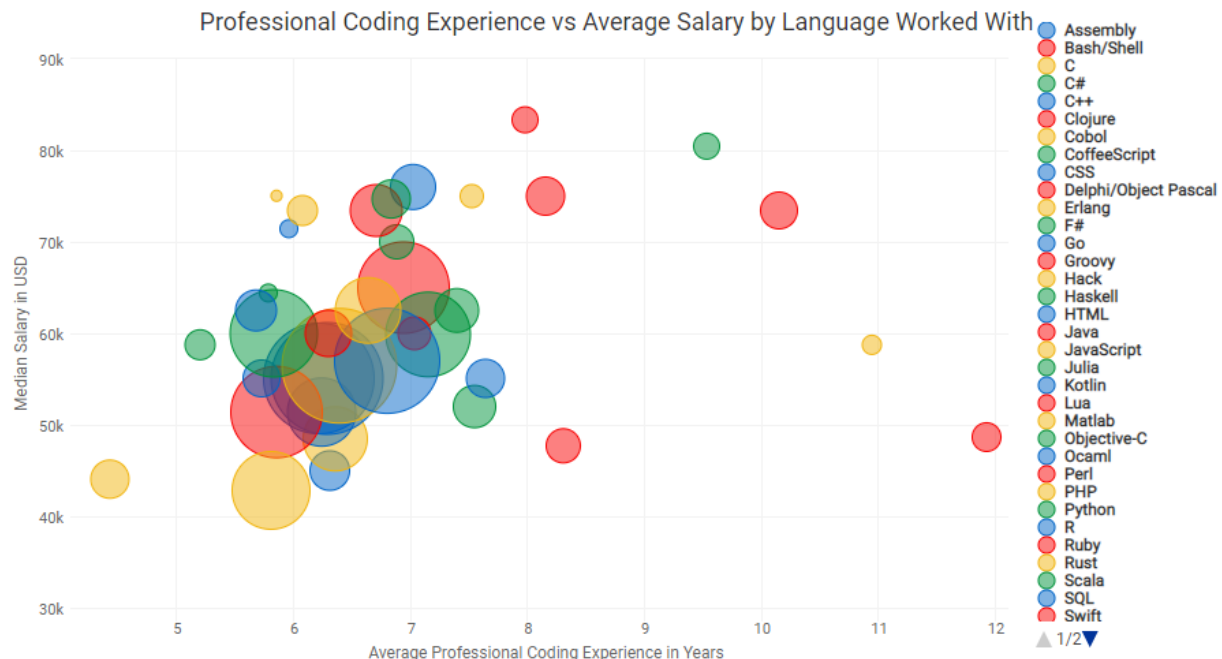
```
survey%>%
  select(DevType,ConvertedSalary,YearsCodingProf)%>%
  filter(!is.na(DevType),!is.na(ConvertedSalary),!is.na(YearsCodingProf))%>%
  mutate(YearsCoding=parse_number(as.character(YearsCodingProf)))%>%
  mutate(DevType=str_split(DevType,";"))%>%
  unnest(DevType)%>%
  group_by(DevType)%>%
  summarise(avg_salary=round(median(ConvertedSalary),0),n=n(),avg_years=mean(
YearsCoding))%>%
  arrange(desc(n))%>%
  ungroup() -> t
hchart(t ,"bubble",hcaes(x=avg_years,y=avg_salary,group=DevType,size=n)) %>%
  hc_title(text="Professional Coding Experience vs Average Salary by Developer Type") %>%
  hc_xAxis(title=list(text="Average Professional Coding Experience in Years"))
) %>%
  hc_yAxis(title=list(text="Median Salary in USD")) %>%
  hc_legend(align = "right", layout = "vertical", verticalAlign = "top") %>%
  hc_add_theme(hc_theme_google())
```



```
survey%>%
  select(LanguageWorkedWith,ConvertedSalary)%>%
  filter(!is.na(LanguageWorkedWith),!is.na(ConvertedSalary))%>%
  mutate(LanguageWorkedWith=str_split(LanguageWorkedWith,";"))%>%
  unnest(LanguageWorkedWith)%>%
  group_by(LanguageWorkedWith)%>%
  summarise(avg_salary=round(median(ConvertedSalary),0))%>%
  arrange(desc(avg_salary))%>%
  ungroup()%>% arrange(desc(avg_salary)) %>%
  hchart("lollipop",hcaes(x=LanguageWorkedWith,y=avg_salary,group=LanguageWor
kedWith,size=avg_salary)) %>%
  hc_legend(enabled=FALSE) %>%
  hc_title(text="Median Salary by Language Worked With") %>%
  hc_xAxis(title=list(text="Language Worked With")) %>%
  hc_yAxis(title=list(text="Median Salary in USD"))
```



```
survey%>%
  select(LanguageWorkedWith,ConvertedSalary,YearsCodingProf)%>%
  filter(!is.na(LanguageWorkedWith),!is.na(ConvertedSalary),!is.na(YearsCodingProf))%>%
  mutate(YearsCoding=parse_number(as.character(YearsCodingProf)))%>%
  mutate(LanguageWorkedWith=str_split(LanguageWorkedWith,";"))%>%
  unnest(LanguageWorkedWith)%>%
  group_by(LanguageWorkedWith)%>%
  summarise(avg_salary=round(median(ConvertedSalary),0),n=n(),avg_years=mean(YearsCoding))%>%
  arrange(desc(n))%>%
  ungroup() -> t
hchart(t,"bubble",hcaes(x=avg_years,y=avg_salary,group=LanguageWorkedWith,size=n)) %>%
  hc_title(text="Professional Coding Experience vs Average Salary by Language Worked With") %>%
  hc_xAxis(title=list(text="Average Professional Coding Experience in Years")) %>%
  hc_yAxis(title=list(text="Median Salary in USD")) %>%
  hc_legend(align="right", layout="vertical", verticalAlign="top") %>%
  hc_add_theme(hc_theme_google())
```

Part IV: Network Analysis

In a Network plot for X, Each node denotes the different X values and the size of node denotes the number of respondents for X. Each connecting edge between any two nodes denotes that the respondents chose both X values. And, the width of the edge denotes the number of users that chose both X values. Let us look at the network plots for Developer Types, Languages developers want to work with and IDEs they use.

```
survey %>% select(Respondent, DevType) -> t1
t1 %>%
  mutate(DevType = strsplit(as.character(DevType), ";"))%>%
  unnest(DevType) -> t2

t2 %>%
  group_by(Respondent)%>%
  filter(n()>=2)%>%
  do(data.frame(t(combn((.)[["DevType"]], 2)), stringsAsFactors=FALSE))%>%
  ungroup()%>%
  rename(source = X1, target = X2)%>%
  select(-Respondent) -> t2_edges

t2_edges %>%
  group_by(source, target)%>%
  summarise(weight=n()) -> t2_edges

## `summarise()` has grouped output by 'source'. You can override using the `
.groups` argument.

names(t2_edges) <- c("from", "to", "weight")
t2_edges$weight <- t2_edges$weight/1500
```

```

t2_edges$width <- 1+t2_edges$weight
t2_edges$smooth <- FALSE
t2_edges$shadow <- FALSE

t2_nodes <- t2 %>% filter(!is.na(DevType)) %>% group_by(DevType) %>% summaris
e(n = n()/750) %>% arrange(desc(n))
names(t2_nodes) <- c("id","size")

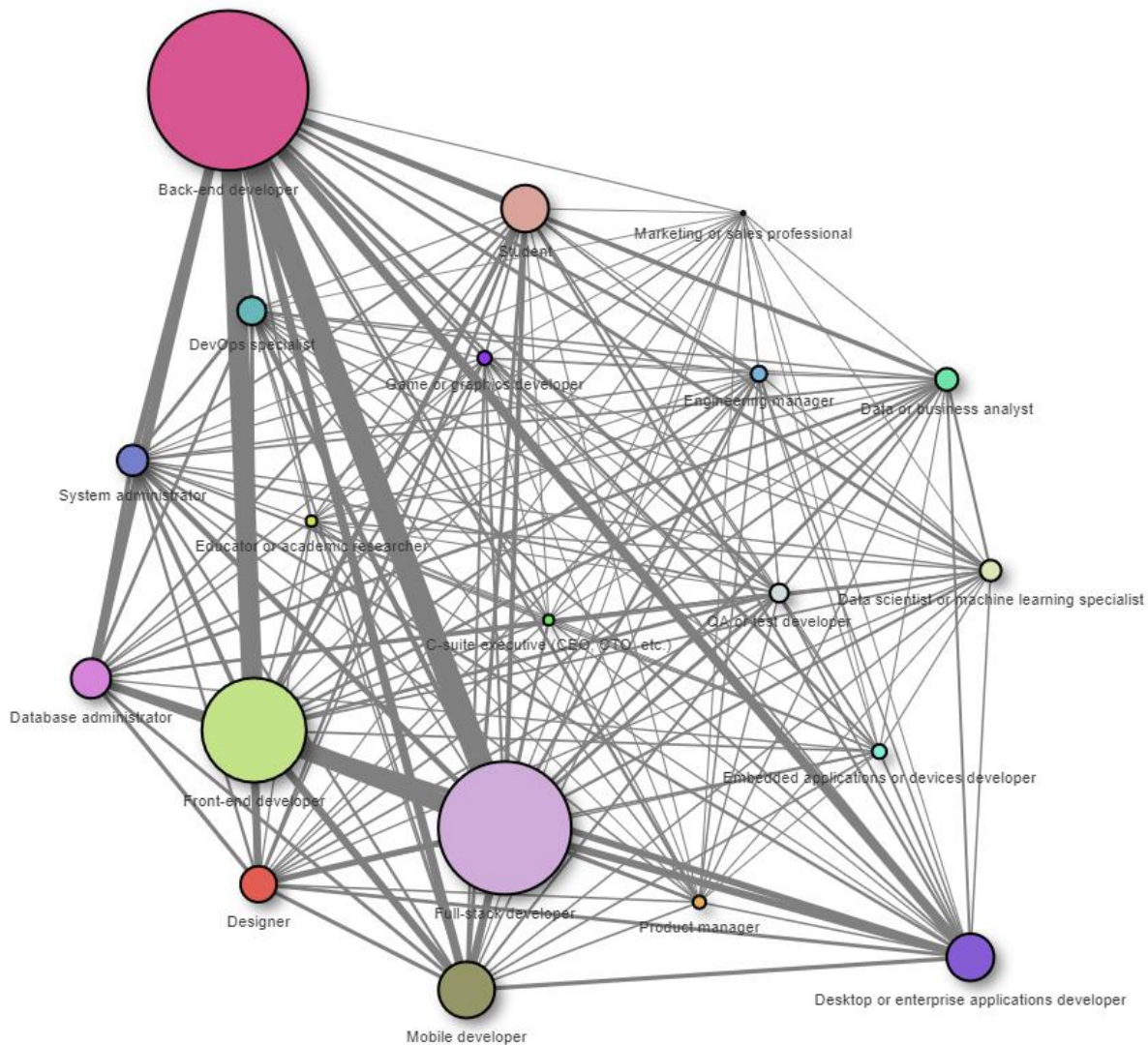
n <- nrow(t2_nodes)
palette <- distinctColorPalette(n)

t2_nodes$shape <- "dot"
t2_nodes$shadow <- TRUE
t2_nodes$title <- t2_nodes$id
t2_nodes$label <- t2_nodes$id
t2_nodes$size <- t2_nodes$size
t2_nodes$borderWidth <- 2

t2_nodes$color.background <- palette[as.numeric(as.factor(t2_nodes$id))]
t2_nodes$color.border <- "black"
t2_nodes$color.highlight.background <- "gold"
t2_nodes$color.highlight.border <- "gold"

visNetwork(t2_nodes, t2_edges,height = "1000px", width="100%") %>% visIgraphL
ayout(layout = "layout_with_lgl") %>%
  visEdges(shadow = TRUE,
           color = list(color = "gray", highlight = "darkorange")) %>% visLeg
end()

```



From the network plot above, it can be seen that, it can be seen that the relationships with the highest correlation are Back-end developer — Full-stack developer Front-end developer — Full-stack developer Back-end developer — Front-end developer Back-end developer — Desktop or enterprise application developer

```
survey %>% select(Respondent, LanguageDesireNextYear) -> t1
t1 %>%
  mutate(LanguageDesireNextYear = strsplit(as.character(LanguageDesireNextYear), ";"))%>%
  unnest(LanguageDesireNextYear) -> t2

t2 %>%
  group_by(Respondent)%>%
  filter(n()>=2)%>%
  do(data.frame(t(combn((.)[["LanguageDesireNextYear"]], 2)), stringsAsFactor = FALSE))%>%
  ungroup()%>%
```

```

    rename(source = X1, target = X2)%>%
    select(-Respondent) -> t2_edges

t2_edges %>%
  group_by(source,target)%>%
  summarise(weight=n()) ->t2_edges

## `summarise()` has grouped output by 'source'. You can override using the `
.groups` argument.

names(t2_edges) <- c("from","to","weight")
t2_edges$weight <- t2_edges$weight/1500

t2_edges$width <- 1+t2_edges$weight
t2_edges$smooth <- FALSE
t2_edges$shadow <- FALSE

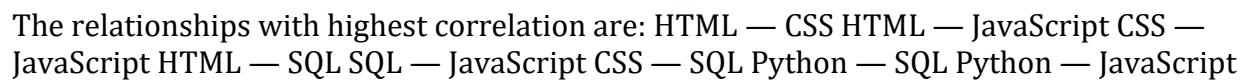
t2_nodes <- t2 %>% filter(!is.na(LanguageDesireNextYear)) %>% group_by(LanguageDesireNextYear) %>% summarise(n = n()/750) %>% arrange(desc(n))
names(t2_nodes) <- c("id","size")

n <- nrow(t2_nodes)
palette <- distinctColorPalette(n)

t2_nodes$shape <- "circle"
t2_nodes$shadow <- TRUE
t2_nodes$title <- t2_nodes$id
t2_nodes$label <- t2_nodes$id
t2_nodes$size <- t2_nodes$size
t2_nodes$borderWidth <- 2

t2_nodes$color.background <- palette[as.numeric(as.factor(t2_nodes$id))]
t2_nodes$color.border <- "black"
t2_nodes$color.highlight.background <- "purple"
t2_nodes$color.highlight.border <- "purple"

visNetwork(t2_nodes, t2_edges,height = "1200px", width="100%") %>% visIgraphLayout(layout = "layout_with_lgl") %>%
  visEdges(shadow = TRUE,
           color = list(color = "gray", highlight = "magenta")) %>% visLegend
()
```



```
survey %>% select(Respondent, IDE) -> t1
t1 %>%
  mutate(IDE = strsplit(as.character(IDE), ";"))%>%
  unnest(IDE) -> t2

t2 %>%
  group_by(Respondent)%>%
  filter(n()>=2)%>%
  do(data.frame(t(combn((.)[["IDE"]], 2)), stringsAsFactors=FALSE))%>%
  ungroup()%>%
  rename(source = X1, target = X2)%>%
  select(-Respondent) -> t2_edges
```



```

t2_edges %>%
  group_by(source,target)%>%
  summarise(weight=n()) ->t2_edges

## `summarise()` has grouped output by 'source'. You can override using the `
.groups` argument.

names(t2_edges) <- c("from","to","weight")
t2_edges$weight <- t2_edges$weight/1500

t2_edges$width <- 1+t2_edges$weight
t2_edges$smooth <- FALSE
t2_edges$shadow <- FALSE

t2_nodes <- t2 %>% filter(!is.na(IDE)) %>% group_by(IDE) %>% summarise(n = n(
)/750) %>% arrange(desc(n))
names(t2_nodes) <- c("id","size")

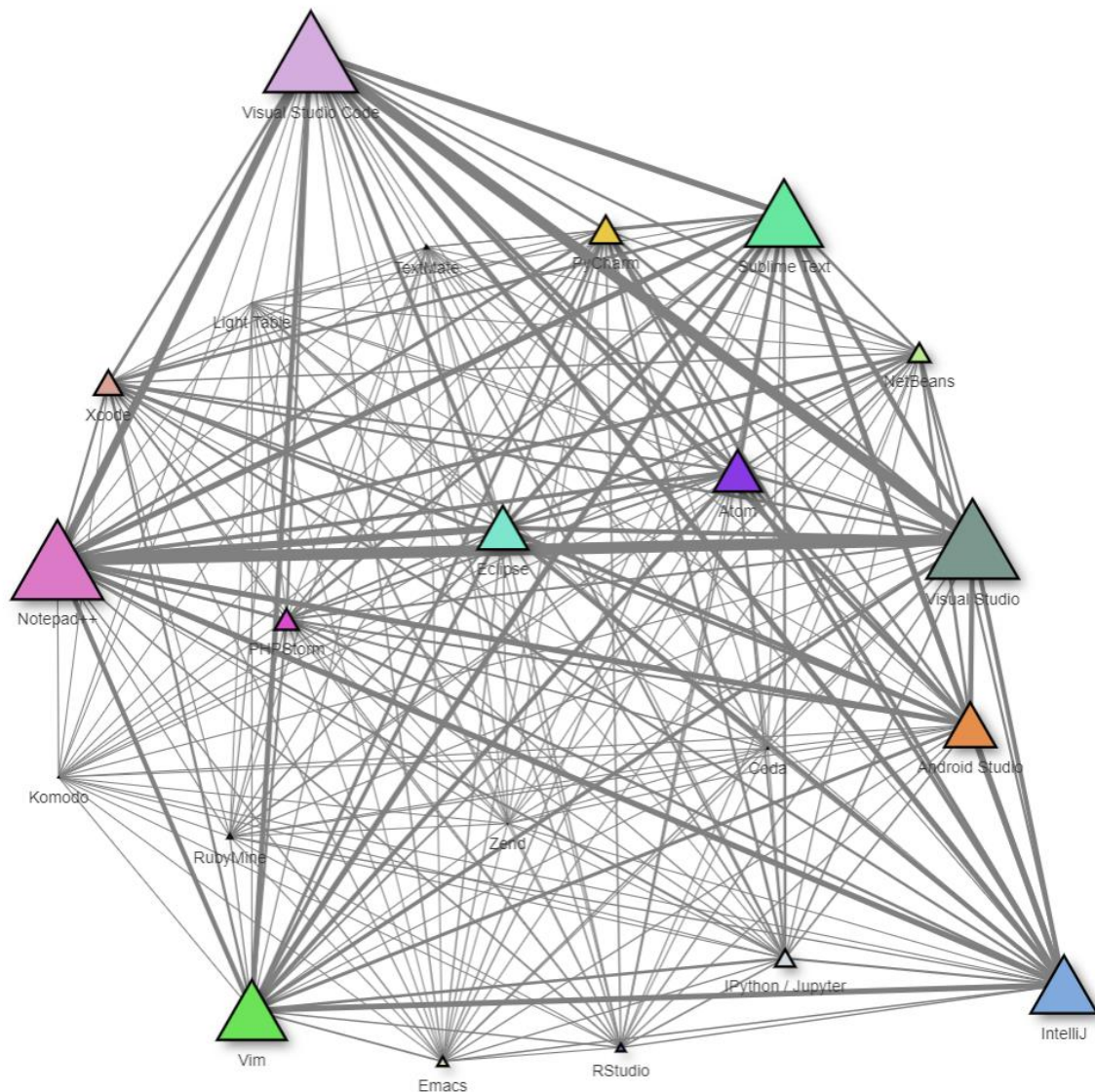
n <- nrow(t2_nodes)
palette <- distinctColorPalette(n)

t2_nodes$shape <- "triangle"
t2_nodes$shadow <- TRUE
t2_nodes$title <- t2_nodes$id
t2_nodes$label <- t2_nodes$id
t2_nodes$size <- t2_nodes$size
t2_nodes$borderWidth <- 2

t2_nodes$color.background <- palette[as.numeric(as.factor(t2_nodes$id))]
t2_nodes$color.border <- "black"
t2_nodes$color.highlight.background <- "chocolate"
t2_nodes$color.highlight.border <- "chocolate"

visNetwork(t2_nodes, t2_edges,height = "1000px", width="100%") %>% visIgraphL
ayout(layout = "layout_with_lgl") %>%
  visEdges(shadow = TRUE,
           color = list(color = "gray", highlight = "brown")) %>% visLegend()

```



1.5.Conclusion

This project has given me some unknown insights into the Developer Community through the many visualizations created. Conclusions for each module are as follows:

PART 1: General analysis of coding as a hobby, contribution to Open-Source projects, Education, Developer Types and Job Status.

- The StackOverflow community has a large gender-gap.
- Most respondents are not students and are employed full time.
- Most people of the developer community have atleast one degree and are majors in the field of computer science

- Most people who are not students do not code as a hobby and do not contribute to open-source projects. A lot of students also code as a hobby but do not contribute to opensource projects.
- Most people are moderately satisfied with their current job but hope to work in a more technical or specialized role in the next five years.

PART 2: Analysis of Language Preferences, Languages to work with in the future, IDE Preferences and Framework Preferences and Coding Experience

- The StackOverflow Community mostly consists of Web Developers which includes Front-End, Back-End and Full-Stack Developers
- JavaScript, HTML and CSS tops the list of most worked-with languages while Python, C#, Kotlin and Swift are gaining popularity as most people want to work with it in the future.
- Visual Studio Code and Visual Studio are the most popular IDEs of developers with Notepad++ being third.
- Django, TensorFlow are among the frameworks developers want to learn most which show increasing popularity of Python and the field of ML.
- Most beginners start out learning web-development
- Most Coders start learning coding either in college or only when they get a job.

PART 3: Salary analysis on different parameters

- United States, Canada and Australia are where developers earn the highest salary.
- United States and India have the highest variance when it comes to range of salaries.
- Young Developers earn significantly less compared to older developers.
- Engineering managers and DevOps specialist are the highest earning jobs
- Clojure and F# are the highest paying languages

PART 4: Network analysis on IDE Preferences, Language Desire and Developer Types.

- Based on Network analysis of Developer Types, it can be concluded that most web developers are very complete in the field, meaning that they have knowledge about almost everything related to web development.
- Based on Network analysis of Languages, we can see that JavaScript, HTML, CSS are part of a package. Most developers who use one, use the other two.
- Based on Network Analysis done on IDEs, Visual Studio, Visual Studio Code and Notepad++ are preferred together by majority of developers.

List of figures

S. No.	Title of Visual	Page No
1.	Top 20 Countries of respondents of survey - Treemap	21
2.	Top 20 Countries of respondents of survey - Barchart	22
3.	Countries by no. of respondents	23
4.	Is coding a hobby for respondents?	24
5.	Do respondents contribute to OpenSource Projects?	24
6.	Gender of respondents	25
7.	Are the respondents students?	26
8.	Employment status of the respondents	27
9.	Highest Level of Fromal Education	28

10.	Undergraduation Majors of respondents	29
11.	Contribution to OpenSource Projects by Coding as a Hobby	30
12.	Coding as a hobby by Student Status	31
13.	Contribution to OpenSource Projects by Student Status	31
14.	Contribution to OpenSource Projects by Employment Status	32
15.	Current Employment vs Hope in Five Years	33
16.	Current Employment vs Job Satisfaction	34
17.	Hope in Five Years vs Job Satisfaction	35
18.	Developer Types of Respondents	36
19.	How many years have respondents spent coding?	37
20.	How many years have respondents spent coding professionally?	38
21.	Years Coded vs Years Coding Professionally	39
22.	Languages Developers Have Worked With - Barchart	40
23.	Languages Developers Have Worked With - Treemap	40
24.	Languages Developers Want to Work With - Barchart	41
25.	Languages Developers Want to Work With - Treemap	42
26.	Languages Developers who have 0-2 years coding experience have worked with	43
27.	IDE Preferences of Developers - Barchart	44
28.	IDE Preferences of Developers - Treemap	44

29.	Framework Developers Have Worked With - Barchart	45
30.	Framework Developers Have Worked With - Treemap	46
31.	Framework Developers Want to Work With - Barchart	47
32.	Framework Developers Want to Work With - Treemap	47
33.	Operating System Preference of Developers	48
34.	Median Salary of Countries	49
35.	Distribution of Annual Salary in USD	50
36.	Distribution of Annual Salary in USD on Log Scale	50
37.	Distribution of Annual Salary(USD) of Top 10 Countries of respondents	51
38.	Annual Salary in USD - Male vs Female of Top 10 countries by respondents	52
39.	Median Salary by developer type and age	53
40.	Median Salary by developer type	54
41	Professional Coding Experience vs Average Salary by Developer Type	55
42.	Median Salary by Language Worked With	56
43.	Professional Coding Experience vs Average Salary by Language Worked With	57
44.	Network of Developer Type	58
45.	Network of Language Desire	59