CIS663ProjectSP25

Vlad Krotov

2025-02-26

## Abstract

A bias or discrimination against persons based on their age is a significant issue in the workplace. This project aims to analyze the impact of age on job satisfaction and compensation in the tech industry. We investigate whether such discrimination exists by analyzing data from the Stack Overflow Annual Developer Survey for 2024, focusing on U.S.-based developers only. Two hypotheses are tested: (1) older developers tend to earn lower salaries, and (2) they report lower job satisfaction compared to their younger counterparts. The analysis will focus on understanding how age influences job satisfaction and compensation, while also considering other factors such as education level, years of professional coding experience, and employment status. By examining these relationships, we aim to provide insights into the potential biases faced by older developers in the tech industry.The entire project is conducted using R for data preprocessing, regression analysis as well as visualizations to support the findings. Key variables such as age, education level, employment type, developer type, job satisfaction, compensation, and experience were numerically encoded for the purpose of this analysis. Multicollinearity is addressed using variance inflation factors (VIF), and a stepwise regression is used to determine significant predictors of job satisfaction and yearly compensation. The results suggest age has an inverse relationship with job satisfaction and compensation, indicating potential bias against older developers. The results also assert that, compensation, years of professional coding experience, and full-time employment status are positively and significantly associated with job satisfaction. Furthermore, job satisfaction, age, education level, and full-time employment status significantly predict yearly compensation. The findings highlight the need for policy makers to strengthen enforcement of the Age Discrimination in Employment Act (ADEA).

## Introduction

Ageism is broadly defined as discrimination against individuals based on age, has become an increasingly concerning issue in the technology industry. Unlike in other industries where the signs of age-related bias typically emerge around the age of 41, IT professionals may begin facing such discrimination as early as age 29, according to a recent CWJobs survey (Sevilla, 2019). This early onset of discrimination reflects a deep-rooted preference for youth within the tech industry, where innovation is often equated with being young and agile. The tech industry is notorious for its fast-paced environment, where the latest trends and technologies are constantly evolving. This rapid change can create a perception that older workers may not be as adaptable or up-to-date with the latest skills, leading to age-related biases in hiring and promotion decisions.This bias is reinforced by both cultural narratives and industry practices like the idiomatic expression “you can’t teach an old new tricks”, which generally means that it is difficult for older people to learn new things. Popular media spotlights like “30 under 30” lists celebrate youth as a hallmark of success, inadvertently sidelining older professionals (Sevilla, 2019). Controversial statements by technology leaders such as “young people are just smarter” by Mark Zuckerberg reinforce ageism as well (Kotler, 2015). Addressing ageism in tech is critical for several reasons. First, It is illegal under US law to discriminate based on age, just like any other form of discrimination. It can also lead to a loss of valuable talent, experience and mentoring as older workers may be pushed out of the industry or discouraged from pursuing tech careers altogether. Ageism can also create a toxic work environment that stifles collaboration and innovation, as younger and older workers may be pitted against each other rather than working together to solve problems. Furthermore, ageism can have serious consequences for the mental health and well-being of older workers, who may feel undervalued and marginalized in their workplaces.Finally, there is a strong ethical imperative to ensure fairness and equal opportunity in a field that is otherwise dominated by young professionals. Promoting age inclusivity is not just a legal obligation but buildnig a just and equitable workplace culture.

## Literature Review

Ageism, or discrimination based on age, is widely recognized as a violation of human rights. Article 26 of the International Covenant on Civil and Political Rights mandates equal protection against all forms of discrimination, including age (UN General Assembly, 1966). Discrimination is the unfair or prejudicial treatment of people and groups based on characteristics such as race, gender, age, or sexual orientation(“Discrimination: What It Is and How to Cope,” 2024). In the context of the workplace environment, age discrimination can be direct, such as refusing to hire older candidates, or indirect, where seemingly neutral policies disproportionately disadvantage older workers (Stanford Encyclopedia of Philosophy, 2020). AN example is “crunch-time” culture, which is prevalent in the tech industry and often requires employees to work long hours, making it difficult for older workers with family responsibilities to keep up. This can lead to a perception that older workers are less committed or less capable of handling the demands of the job, further perpetuating age-related biases (Sevilla, 2019). Additionally, the burden of proof for age discrimination cases is often placed on the employee, making it difficult to prove discrimination occurred (Stanford Encyclopedia of Philosophy, 2020). Despite these challenges, there have been some positive developments in recent years. For example, some companies have implemented diversity and inclusion initiatives that specifically address ageism in the workplace. These initiatives aim to create a more inclusive work environment by promoting intergenerational collaboration and mentorship programs (Sevilla, 2019).

## Theory

If ageism exists in the technology industry, the following hypothesis are expected to be true:

H1: Older developers are likely to earn a lower salary in comparison to younger employees

H2: Older developers are likely to report lower job satisfaction than younger employees

## Data

The data used in this project is sourced from the Stack Overflow Annual Developer Survey for 2024. The survey is a comprehensive questionnaire that collects information on various aspects of the developer community, including demographics, job satisfaction, compensation, and work experience. The dataset includes about 70,000 responses from developers worldwide, but for this analysis, we will focus specifically on U.S.-based developers. The survey data is publicly available and can be accessed through the Stack Overflow website (<https://survey.stackoverflow.co/2024/>). The dataset contains a wide range of variables, including age, education level, employment status, developer type, job satisfaction, compensation, and years of professional coding experience. These variables will be used to analyze the relationship between age and job satisfaction/compensation. The data is in CSV format and can be easily imported into R for analysis. The survey data is structured in a way that allows for easy manipulation and analysis using R’s data manipulation libraries.

## Reading Data from StackOverflow Files

survey\_data <- read.csv("survey\_results\_public.csv")  
schema\_data <- read.csv("survey\_results\_schema.csv")  
save(survey\_data, schema\_data, file = "survey\_data.RData")

# Load the dataset  
load("survey\_data.RData")  
  
# Check what objects were loaded  
#ls()  
  
# View dataset structure  
#str(survey\_data)   
  
# View first few rows  
head(survey\_data)

## ResponseId MainBranch Age  
## 1 1 I am a developer by profession Under 18 years old  
## 2 2 I am a developer by profession 35-44 years old  
## 3 3 I am a developer by profession 45-54 years old  
## 4 4 I am learning to code 18-24 years old  
## 5 5 I am a developer by profession 18-24 years old  
## 6 6 I code primarily as a hobby Under 18 years old  
## Employment RemoteWork Check  
## 1 Employed, full-time Remote Apples  
## 2 Employed, full-time Remote Apples  
## 3 Employed, full-time Remote Apples  
## 4 Student, full-time <NA> Apples  
## 5 Student, full-time <NA> Apples  
## 6 Student, full-time <NA> Apples  
## CodingActivities  
## 1 Hobby  
## 2 Hobby;Contribute to open-source projects;Other (please specify):  
## 3 Hobby;Contribute to open-source projects;Other (please specify):  
## 4 <NA>  
## 5 <NA>  
## 6 <NA>  
## EdLevel  
## 1 Primary/elementary school  
## 2 Bachelor’s degree (B.A., B.S., B.Eng., etc.)  
## 3 Master’s degree (M.A., M.S., M.Eng., MBA, etc.)  
## 4 Some college/university study without earning a degree  
## 5 Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)  
## 6 Primary/elementary school  
## LearnCode  
## 1 Books / Physical media  
## 2 Books / Physical media;Colleague;On the job training;Other online resources (e.g., videos, blogs, forum, online community)  
## 3 Books / Physical media;Colleague;On the job training;Other online resources (e.g., videos, blogs, forum, online community);School (i.e., University, College, etc)  
## 4 Other online resources (e.g., videos, blogs, forum, online community);School (i.e., University, College, etc);Online Courses or Certification  
## 5 Other online resources (e.g., videos, blogs, forum, online community)  
## 6 School (i.e., University, College, etc);Online Courses or Certification  
## LearnCodeOnline  
## 1 <NA>  
## 2 Technical documentation;Blogs;Books;Written Tutorials;Stack Overflow  
## 3 Technical documentation;Blogs;Books;Written Tutorials;Stack Overflow;Coding sessions (live or recorded);Social Media  
## 4 Stack Overflow;How-to videos;Interactive tutorial  
## 5 Technical documentation;Blogs;Written Tutorials;Stack Overflow;Social Media;Video-based Online Courses  
## 6 <NA>  
## TechDoc  
## 1 <NA>  
## 2 API document(s) and/or SDK document(s);User guides or README files found in the source repository;First-party knowledge base;Traditional public search engine  
## 3 API document(s) and/or SDK document(s);User guides or README files found in the source repository;Traditional public search engine  
## 4 <NA>  
## 5 API document(s) and/or SDK document(s);User guides or README files found in the source repository;First-party knowledge base;Traditional public search engine  
## 6 <NA>  
## YearsCode YearsCodePro DevType OrgSize PurchaseInfluence  
## 1 <NA> <NA> <NA> <NA> <NA>  
## 2 20 17 Developer, full-stack <NA> <NA>  
## 3 37 27 Developer Experience <NA> <NA>  
## 4 4 <NA> Developer, full-stack <NA> <NA>  
## 5 9 <NA> Developer, full-stack <NA> <NA>  
## 6 10 <NA> Student <NA> <NA>  
## BuyNewTool BuildvsBuy TechEndorse  
## 1 <NA> <NA> <NA>  
## 2 <NA> <NA> <NA>  
## 3 <NA> <NA> <NA>  
## 4 <NA> <NA> <NA>  
## 5 <NA> <NA> <NA>  
## 6 <NA> <NA> <NA>  
## Country Currency CompTotal  
## 1 United States of America <NA> NA  
## 2 United Kingdom of Great Britain and Northern Ireland <NA> NA  
## 3 United Kingdom of Great Britain and Northern Ireland <NA> NA  
## 4 Canada <NA> NA  
## 5 Norway <NA> NA  
## 6 United States of America <NA> NA  
## LanguageHaveWorkedWith  
## 1 <NA>  
## 2 Bash/Shell (all shells);Go;HTML/CSS;Java;JavaScript;Python;TypeScript  
## 3 C#  
## 4 C;C++;HTML/CSS;Java;JavaScript;PHP;PowerShell;Python;SQL;TypeScript  
## 5 C++;HTML/CSS;JavaScript;Lua;Python;Rust  
## 6 Bash/Shell (all shells);HTML/CSS;Java;JavaScript;Python;Rust;Swift;TypeScript  
## LanguageWantToWorkWith  
## 1 <NA>  
## 2 Bash/Shell (all shells);Go;HTML/CSS;Java;JavaScript;Kotlin;Python;TypeScript  
## 3 C#  
## 4 HTML/CSS;Java;JavaScript;PowerShell;Python;SQL;TypeScript  
## 5 C++;HTML/CSS;JavaScript;Lua;Python  
## 6 Bash/Shell (all shells);HTML/CSS;Java;JavaScript;Python;Swift;TypeScript  
## LanguageAdmired  
## 1 <NA>  
## 2 Bash/Shell (all shells);Go;HTML/CSS;Java;JavaScript;Python;TypeScript  
## 3 C#  
## 4 HTML/CSS;Java;JavaScript;PowerShell;Python;SQL;TypeScript  
## 5 C++;HTML/CSS;JavaScript;Lua;Python  
## 6 Bash/Shell (all shells);HTML/CSS;Java;JavaScript;Python;Swift;TypeScript  
## DatabaseHaveWorkedWith DatabaseWantToWorkWith  
## 1 <NA> <NA>  
## 2 Dynamodb;MongoDB;PostgreSQL PostgreSQL  
## 3 Firebase Realtime Database Firebase Realtime Database  
## 4 MongoDB;MySQL;PostgreSQL;SQLite MongoDB;MySQL;PostgreSQL  
## 5 PostgreSQL;SQLite PostgreSQL;SQLite  
## 6 Cloud Firestore Cloud Firestore  
## DatabaseAdmired PlatformHaveWorkedWith  
## 1 <NA> <NA>  
## 2 PostgreSQL Amazon Web Services (AWS);Heroku;Netlify  
## 3 Firebase Realtime Database Google Cloud  
## 4 MongoDB;MySQL;PostgreSQL Amazon Web Services (AWS);Fly.io;Heroku  
## 5 PostgreSQL;SQLite <NA>  
## 6 Cloud Firestore Cloudflare  
## PlatformWantToWorkWith  
## 1 <NA>  
## 2 Amazon Web Services (AWS);Heroku;Netlify  
## 3 Google Cloud  
## 4 Amazon Web Services (AWS);Vercel  
## 5 <NA>  
## 6 Cloudflare  
## PlatformAdmired  
## 1 <NA>  
## 2 Amazon Web Services (AWS);Heroku;Netlify  
## 3 Google Cloud  
## 4 Amazon Web Services (AWS)  
## 5 <NA>  
## 6 Cloudflare  
## WebframeHaveWorkedWith WebframeWantToWorkWith  
## 1 <NA> <NA>  
## 2 Express;Next.js;Node.js;React Express;Htmx;Node.js;React;Remix  
## 3 ASP.NET CORE ASP.NET CORE  
## 4 jQuery;Next.js;Node.js;React;WordPress jQuery;Next.js;Node.js;React  
## 5 <NA> <NA>  
## 6 Node.js Node.js  
## WebframeAdmired EmbeddedHaveWorkedWith EmbeddedWantToWorkWith  
## 1 <NA> <NA> <NA>  
## 2 Express;Node.js;React <NA> <NA>  
## 3 ASP.NET CORE Rasberry Pi Rasberry Pi  
## 4 jQuery;Next.js;Node.js;React Rasberry Pi <NA>  
## 5 <NA> CMake;Cargo;Rasberry Pi CMake;Rasberry Pi  
## 6 Node.js Rasberry Pi Rasberry Pi  
## EmbeddedAdmired MiscTechHaveWorkedWith  
## 1 <NA> <NA>  
## 2 <NA> <NA>  
## 3 Rasberry Pi .NET (5+) ;.NET Framework (1.0 - 4.8);.NET MAUI  
## 4 <NA> NumPy;Pandas;Ruff;TensorFlow  
## 5 CMake;Rasberry Pi <NA>  
## 6 Rasberry Pi <NA>  
## MiscTechWantToWorkWith  
## 1 <NA>  
## 2 <NA>  
## 3 .NET (5+) ;.NET Framework (1.0 - 4.8);.NET MAUI  
## 4 <NA>  
## 5 <NA>  
## 6 <NA>  
## MiscTechAdmired  
## 1 <NA>  
## 2 <NA>  
## 3 .NET (5+) ;.NET Framework (1.0 - 4.8);.NET MAUI  
## 4 <NA>  
## 5 <NA>  
## 6 <NA>  
## ToolsTechHaveWorkedWith  
## 1 <NA>  
## 2 Docker;Homebrew;Kubernetes;npm;Vite;Webpack  
## 3 MSBuild  
## 4 Docker;npm;Pip  
## 5 APT;Make;npm  
## 6 Docker;Homebrew;npm;Pip;pnpm  
## ToolsTechWantToWorkWith  
## 1 <NA>  
## 2 Docker;Homebrew;Kubernetes;npm;Vite;Webpack  
## 3 MSBuild  
## 4 Docker;Kubernetes;npm  
## 5 APT;Make  
## 6 Homebrew;npm;Pip;pnpm  
## ToolsTechAdmired  
## 1 <NA>  
## 2 Docker;Homebrew;Kubernetes;npm;Vite;Webpack  
## 3 MSBuild  
## 4 Docker;npm  
## 5 APT;Make  
## 6 Homebrew;npm;Pip;pnpm  
## NEWCollabToolsHaveWorkedWith NEWCollabToolsWantToWorkWith  
## 1 <NA> <NA>  
## 2 PyCharm;Visual Studio Code;WebStorm PyCharm;Visual Studio Code;WebStorm  
## 3 Visual Studio Visual Studio  
## 4 <NA> <NA>  
## 5 Vim Vim  
## 6 Nano;Vim;Visual Studio Code;Xcode Nano;Vim;Visual Studio Code;Xcode  
## NEWCollabToolsAdmired OpSysPersonal.use  
## 1 <NA> <NA>  
## 2 PyCharm;Visual Studio Code;WebStorm MacOS;Windows  
## 3 Visual Studio Windows  
## 4 <NA> <NA>  
## 5 Vim Other (please specify):  
## 6 Nano;Vim;Visual Studio Code;Xcode iOS;MacOS;Ubuntu  
## OpSysProfessional.use  
## 1 <NA>  
## 2 MacOS  
## 3 Windows  
## 4 <NA>  
## 5   
## 6   
## OfficeStackAsyncHaveWorkedWith  
## 1 <NA>  
## 2 <NA>  
## 3 <NA>  
## 4 <NA>  
## 5 GitHub Discussions;Markdown File;Obsidian;Stack Overflow for Teams  
## 6 Markdown File  
## OfficeStackAsyncWantToWorkWith  
## 1 <NA>  
## 2 <NA>  
## 3 <NA>  
## 4 <NA>  
## 5 GitHub Discussions;Markdown File;Obsidian  
## 6 Markdown File  
## OfficeStackAsyncAdmired  
## 1 <NA>  
## 2 <NA>  
## 3 <NA>  
## 4 <NA>  
## 5 GitHub Discussions;Markdown File;Obsidian  
## 6 Markdown File  
## OfficeStackSyncHaveWorkedWith OfficeStackSyncWantToWorkWith  
## 1 <NA> <NA>  
## 2 Microsoft Teams;Slack Slack  
## 3 Google Chat;Google Meet;Microsoft Teams;Zoom Google Chat;Google Meet;Zoom  
## 4 <NA> <NA>  
## 5 Discord;Whatsapp Discord;Whatsapp  
## 6 Discord Discord  
## OfficeStackSyncAdmired AISearchDevHaveWorkedWith  
## 1 <NA> <NA>  
## 2 Slack <NA>  
## 3 Google Chat;Google Meet;Zoom <NA>  
## 4 <NA> <NA>  
## 5 Discord;Whatsapp <NA>  
## 6 Discord ChatGPT;GitHub Copilot;OpenAI Codex  
## AISearchDevWantToWorkWith AISearchDevAdmired  
## 1 <NA> <NA>  
## 2 <NA> <NA>  
## 3 <NA> <NA>  
## 4 <NA> <NA>  
## 5 <NA> <NA>  
## 6 ChatGPT;GitHub Copilot ChatGPT;GitHub Copilot  
## NEWSOSites  
## 1 I have never visited Stack Overflow or the Stack Exchange network  
## 2 Stack Overflow for Teams (private knowledge sharing & collaboration platform for companies);Stack Overflow;Stack Exchange  
## 3 Stack Overflow;Stack Exchange;Stack Overflow Blog or Podcast  
## 4 Stack Overflow  
## 5 Stack Overflow for Teams (private knowledge sharing & collaboration platform for companies);Stack Overflow;Stack Exchange  
## 6 Stack Overflow;Stack Exchange  
## SOVisitFreq SOAccount SOPartFreq  
## 1 <NA> <NA> <NA>  
## 2 Multiple times per day Yes Multiple times per day  
## 3 Multiple times per day Yes Multiple times per day  
## 4 Daily or almost daily No <NA>  
## 5 Multiple times per day Yes Multiple times per day  
## 6 Multiple times per day Yes Multiple times per day  
## SOHow  
## 1 <NA>  
## 2 Quickly finding code solutions;Finding reliable guidance from community-vetted answers;Showcase expertise with code solutions;Engage with community by commenting on questions and answers or voting on questions and answers  
## 3 Quickly finding code solutions;Finding reliable guidance from community-vetted answers;Showcase expertise with code solutions;Engage with community by commenting on questions and answers or voting on questions and answers;Learning new-to-me technology/techniques;Learning new-to-everyone technology/techniques  
## 4 Quickly finding code solutions  
## 5 Quickly finding code solutions;Engage with community by commenting on questions and answers or voting on questions and answers;Learning new-to-me technology/techniques;Other (please specify):  
## 6 Quickly finding code solutions;Finding reliable guidance from community-vetted answers;Showcase expertise with code solutions;Engage with community by commenting on questions and answers or voting on questions and answers  
## SOComm AISelect AISent  
## 1 <NA> Yes Very favorable  
## 2 Yes, definitely No, and I don't plan to <NA>  
## 3 Yes, definitely No, and I don't plan to <NA>  
## 4 No, not really Yes Very favorable  
## 5 Yes, definitely No, and I don't plan to <NA>  
## 6 Yes, definitely Yes Favorable  
## AIBen  
## 1 Increase productivity  
## 2 <NA>  
## 3 <NA>  
## 4 Increase productivity;Greater efficiency;Improve collaboration;Speed up learning;Improve accuracy in coding  
## 5 <NA>  
## 6 Increase productivity;Greater efficiency;Improve accuracy in coding  
## AIAcc AIComplex  
## 1 <NA> <NA>  
## 2 <NA> <NA>  
## 3 <NA> <NA>  
## 4 Somewhat trust Bad at handling complex tasks  
## 5 <NA> <NA>  
## 6 Somewhat trust Good, but not great at handling complex tasks  
## AIToolCurrently.Using  
## 1 <NA>  
## 2 <NA>  
## 3 <NA>  
## 4 Learning about a codebase;Project planning;Writing code;Documenting code;Debugging and getting help;Deployment and monitoring;Search for answers;Generating content or synthetic data  
## 5 <NA>  
## 6 Writing code;Debugging and getting help  
## AIToolInterested.in.Using  
## 1 <NA>  
## 2 <NA>  
## 3 <NA>  
## 4 Testing code;Committing and reviewing code;Predictive analytics  
## 5 <NA>  
## 6 Documenting code;Testing code;Committing and reviewing code;Deployment and monitoring;Predictive analytics;Search for answers;Generating content or synthetic data  
## AIToolNot.interested.in.Using  
## 1 <NA>  
## 2 <NA>  
## 3 <NA>  
## 4   
## 5 <NA>  
## 6 Learning about a codebase;Project planning  
## AINextMuch.more.integrated  
## 1 <NA>  
## 2 <NA>  
## 3 <NA>  
## 4 Learning about a codebase;Project planning;Writing code;Documenting code;Debugging and getting help;Deployment and monitoring;Search for answers;Generating content or synthetic data  
## 5 <NA>  
## 6   
## AINextNo.change AINextMore.integrated  
## 1 <NA> <NA>  
## 2 <NA> <NA>  
## 3 <NA> <NA>  
## 4   
## 5 <NA> <NA>  
## 6 Writing code;Debugging and getting help   
## AINextLess.integrated AINextMuch.less.integrated AIThreat  
## 1 <NA> <NA> <NA>  
## 2 <NA> <NA> <NA>  
## 3 <NA> <NA> <NA>  
## 4 No  
## 5 <NA> <NA> <NA>  
## 6 No  
## AIEthics  
## 1 <NA>  
## 2 <NA>  
## 3 <NA>  
## 4 Circulating misinformation or disinformation;Missing or incorrect attribution for sources of data  
## 5 <NA>  
## 6 Circulating misinformation or disinformation;Missing or incorrect attribution for sources of data;Replacing jobs without options for new employment opportunities  
## AIChallenges TBranch ICorPM WorkExp  
## 1 <NA> No <NA> NA  
## 2 <NA> Yes Individual contributor 17  
## 3 <NA> No <NA> NA  
## 4 Don’t trust the output or answers <NA> <NA> NA  
## 5 <NA> <NA> <NA> NA  
## 6 <NA> <NA> <NA> NA  
## Knowledge\_1 Knowledge\_2 Knowledge\_3 Knowledge\_4 Knowledge\_5  
## 1 <NA> <NA> <NA> <NA> <NA>  
## 2 Agree Disagree Agree Agree Agree  
## 3 <NA> <NA> <NA> <NA> <NA>  
## 4 <NA> <NA> <NA> <NA> <NA>  
## 5 <NA> <NA> <NA> <NA> <NA>  
## 6 <NA> <NA> <NA> <NA> <NA>  
## Knowledge\_6 Knowledge\_7 Knowledge\_8 Knowledge\_9 Frequency\_1  
## 1 <NA> <NA> <NA> <NA> <NA>  
## 2 Neither agree nor disagree Disagree Agree Agree <NA>  
## 3 <NA> <NA> <NA> <NA> <NA>  
## 4 <NA> <NA> <NA> <NA> <NA>  
## 5 <NA> <NA> <NA> <NA> <NA>  
## 6 <NA> <NA> <NA> <NA> <NA>  
## Frequency\_2 Frequency\_3 TimeSearching TimeAnswering Frustration  
## 1 <NA> <NA> <NA> <NA> <NA>  
## 2 <NA> <NA> <NA> <NA> <NA>  
## 3 <NA> <NA> <NA> <NA> <NA>  
## 4 <NA> <NA> <NA> <NA> <NA>  
## 5 <NA> <NA> <NA> <NA> <NA>  
## 6 <NA> <NA> <NA> <NA> <NA>  
## ProfessionalTech ProfessionalCloud ProfessionalQuestion Industry  
## 1 <NA> <NA> <NA> <NA>  
## 2 <NA> <NA> <NA> <NA>  
## 3 <NA> <NA> <NA> <NA>  
## 4 <NA> <NA> <NA> <NA>  
## 5 <NA> <NA> <NA> <NA>  
## 6 <NA> <NA> <NA> <NA>  
## JobSatPoints\_1 JobSatPoints\_4 JobSatPoints\_5 JobSatPoints\_6 JobSatPoints\_7  
## 1 NA NA NA NA NA  
## 2 0 0 0 0 0  
## 3 NA NA NA NA NA  
## 4 NA NA NA NA NA  
## 5 NA NA NA NA NA  
## 6 NA NA NA NA NA  
## JobSatPoints\_8 JobSatPoints\_9 JobSatPoints\_10 JobSatPoints\_11  
## 1 NA NA NA NA  
## 2 0 0 0 0  
## 3 NA NA NA NA  
## 4 NA NA NA NA  
## 5 NA NA NA NA  
## 6 NA NA NA NA  
## SurveyLength SurveyEase ConvertedCompYearly JobSat  
## 1 <NA> <NA> NA NA  
## 2 <NA> <NA> NA NA  
## 3 Appropriate in length Easy NA NA  
## 4 Too long Easy NA NA  
## 5 Too short Easy NA NA  
## 6 Appropriate in length Easy NA NA

## Methodology

On importing the Data in R, the first step is to select relevant columns based on the project focus. The selected columns include Age, Country, EdLevel, Employment, DevType, YearsCodePro, ConvertedCompYearly, JobSat, and WorkExp. The next step is to check the number of missing values per column and drop all missing values. After that, we filter for only United States-based developers and drop the Country column since it’s no longer needed. The next step is to convert YearsCodePro, JobSat, ConvertedCompYearly and WorkExp to numeric values. We also drop records with ConvertedCompYearly less than 5000.This was done because some observations had single digit values as yearly income and this was not realistic. The cut-off value of 5000 minimum yearly compensation was arrived at taking into consideration the state with the lowest minimum wage (Georgia and Wyoming) at $5.15 and minimum working hours of 20 per week. The next step is to clean the DevType column by normalizing separators and whitespace, splitting it into a list column, and trimming whitespace for individual dev types. We also create a dummy variable where dev type containing “Developer” is 1 and all others are 0. The next step is to clean the Employment column by dropping certain values and creating binary columns for each employment type. We also rename columns to remove spaces and special characters. Finally, we drop unnecessary columns and encode Education Level and Age as ranked numeric variables.

# Selecting relevant columns based on project focus  
selected\_columns <- c("Age", "Country", "EdLevel", "Employment",   
 "DevType", "YearsCodePro", "ConvertedCompYearly", "JobSat","WorkExp")  
survey\_data\_selected <- survey\_data[selected\_columns]

# Checking data after selection  
str(survey\_data\_selected)

## 'data.frame': 65437 obs. of 9 variables:  
## $ Age : chr "Under 18 years old" "35-44 years old" "45-54 years old" "18-24 years old" ...  
## $ Country : chr "United States of America" "United Kingdom of Great Britain and Northern Ireland" "United Kingdom of Great Britain and Northern Ireland" "Canada" ...  
## $ EdLevel : chr "Primary/elementary school" "Bachelor’s degree (B.A., B.S., B.Eng., etc.)" "Master’s degree (M.A., M.S., M.Eng., MBA, etc.)" "Some college/university study without earning a degree" ...  
## $ Employment : chr "Employed, full-time" "Employed, full-time" "Employed, full-time" "Student, full-time" ...  
## $ DevType : chr NA "Developer, full-stack" "Developer Experience" "Developer, full-stack" ...  
## $ YearsCodePro : chr NA "17" "27" NA ...  
## $ ConvertedCompYearly: int NA NA NA NA NA NA NA NA NA NA ...  
## $ JobSat : int NA NA NA NA NA NA NA NA NA NA ...  
## $ WorkExp : int NA 17 NA NA NA NA NA NA NA NA ...

head(survey\_data\_selected)

## Age Country  
## 1 Under 18 years old United States of America  
## 2 35-44 years old United Kingdom of Great Britain and Northern Ireland  
## 3 45-54 years old United Kingdom of Great Britain and Northern Ireland  
## 4 18-24 years old Canada  
## 5 18-24 years old Norway  
## 6 Under 18 years old United States of America  
## EdLevel  
## 1 Primary/elementary school  
## 2 Bachelor’s degree (B.A., B.S., B.Eng., etc.)  
## 3 Master’s degree (M.A., M.S., M.Eng., MBA, etc.)  
## 4 Some college/university study without earning a degree  
## 5 Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)  
## 6 Primary/elementary school  
## Employment DevType YearsCodePro ConvertedCompYearly  
## 1 Employed, full-time <NA> <NA> NA  
## 2 Employed, full-time Developer, full-stack 17 NA  
## 3 Employed, full-time Developer Experience 27 NA  
## 4 Student, full-time Developer, full-stack <NA> NA  
## 5 Student, full-time Developer, full-stack <NA> NA  
## 6 Student, full-time Student <NA> NA  
## JobSat WorkExp  
## 1 NA NA  
## 2 NA 17  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA

# Checking the number of missing values per column  
colSums(is.na(survey\_data\_selected))

## Age Country EdLevel Employment   
## 0 6507 4653 0   
## DevType YearsCodePro ConvertedCompYearly JobSat   
## 5992 13827 42002 36311   
## WorkExp   
## 35779

# Drop all missing values  
survey\_data\_selected <- survey\_data\_selected %>% drop\_na()  
# Check the number of missing values again  
colSums(is.na(survey\_data\_selected))

## Age Country EdLevel Employment   
## 0 0 0 0   
## DevType YearsCodePro ConvertedCompYearly JobSat   
## 0 0 0 0   
## WorkExp   
## 0

# Filter for only united states  
survey\_data\_selected <- survey\_data\_selected %>% filter(Country == "United States of America")  
# Drop the Country column since it's no longer needed  
survey\_data\_selected <- survey\_data\_selected %>% select(-Country)

# Convert YearsCodePro to numeric  
survey\_data\_selected$YearsCodePro <- as.integer(gsub("[^0-9]", "", survey\_data\_selected$YearsCodePro))  
  
# Convert JobSat to numeric  
survey\_data\_selected$JobSat <- as.integer(gsub("[^0-9]", "", survey\_data\_selected$JobSat))  
  
# Convert ConvertedCompYearly to numeric  
survey\_data\_selected$ConvertedCompYearly <- as.integer(gsub("[^0-9]", "", survey\_data\_selected$ConvertedCompYearly))  
# Drop records with ConvertedCompYearly = 1  
survey\_data\_selected <- survey\_data\_selected %>% filter(ConvertedCompYearly > 5000)  
  
  
# Convert WorkExp to numeric  
survey\_data\_selected$WorkExp <- as.integer(gsub("[^0-9]", "", survey\_data\_selected$WorkExp))  
  
# Check the structure of the dataset after conversion  
str(survey\_data\_selected)

## 'data.frame': 3196 obs. of 8 variables:  
## $ Age : chr "25-34 years old" "45-54 years old" "25-34 years old" "35-44 years old" ...  
## $ EdLevel : chr "Some college/university study without earning a degree" "Some college/university study without earning a degree" "Bachelor’s degree (B.A., B.S., B.Eng., etc.)" "Bachelor’s degree (B.A., B.S., B.Eng., etc.)" ...  
## $ Employment : chr "Employed, full-time;Student, part-time" "Employed, full-time" "Employed, full-time" "Employed, full-time" ...  
## $ DevType : chr "Student" "Developer, full-stack" "Engineer, site reliability" "Developer, full-stack" ...  
## $ YearsCodePro : int 7 30 11 23 18 10 18 26 40 4 ...  
## $ ConvertedCompYearly: int 110000 195000 230000 85000 160000 115000 300000 140000 200000 93000 ...  
## $ JobSat : int 10 5 8 8 7 6 1 8 8 8 ...  
## $ WorkExp : int 8 30 15 25 20 10 28 23 40 4 ...

# Check the first few rows  
head(survey\_data\_selected)

## Age  
## 1 25-34 years old  
## 2 45-54 years old  
## 3 25-34 years old  
## 4 35-44 years old  
## 5 35-44 years old  
## 6 45-54 years old  
## EdLevel  
## 1 Some college/university study without earning a degree  
## 2 Some college/university study without earning a degree  
## 3 Bachelor’s degree (B.A., B.S., B.Eng., etc.)  
## 4 Bachelor’s degree (B.A., B.S., B.Eng., etc.)  
## 5 Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)  
## 6 Associate degree (A.A., A.S., etc.)  
## Employment DevType  
## 1 Employed, full-time;Student, part-time Student  
## 2 Employed, full-time Developer, full-stack  
## 3 Employed, full-time Engineer, site reliability  
## 4 Employed, full-time Developer, full-stack  
## 5 Employed, full-time Developer, full-stack  
## 6 Employed, full-time Developer, full-stack  
## YearsCodePro ConvertedCompYearly JobSat WorkExp  
## 1 7 110000 10 8  
## 2 30 195000 5 30  
## 3 11 230000 8 15  
## 4 23 85000 8 25  
## 5 18 160000 7 20  
## 6 10 115000 6 10

# Save cleaned dataset  
save(survey\_data\_selected, file = "survey\_data\_cleaned.RData")

# Load the cleaned dataset  
load("survey\_data\_cleaned.RData")  
# Check the structure of the cleaned dataset  
str(survey\_data\_selected)

## 'data.frame': 3196 obs. of 8 variables:  
## $ Age : chr "25-34 years old" "45-54 years old" "25-34 years old" "35-44 years old" ...  
## $ EdLevel : chr "Some college/university study without earning a degree" "Some college/university study without earning a degree" "Bachelor’s degree (B.A., B.S., B.Eng., etc.)" "Bachelor’s degree (B.A., B.S., B.Eng., etc.)" ...  
## $ Employment : chr "Employed, full-time;Student, part-time" "Employed, full-time" "Employed, full-time" "Employed, full-time" ...  
## $ DevType : chr "Student" "Developer, full-stack" "Engineer, site reliability" "Developer, full-stack" ...  
## $ YearsCodePro : int 7 30 11 23 18 10 18 26 40 4 ...  
## $ ConvertedCompYearly: int 110000 195000 230000 85000 160000 115000 300000 140000 200000 93000 ...  
## $ JobSat : int 10 5 8 8 7 6 1 8 8 8 ...  
## $ WorkExp : int 8 30 15 25 20 10 28 23 40 4 ...

# Check the first few rows  
head(survey\_data\_selected)

## Age  
## 1 25-34 years old  
## 2 45-54 years old  
## 3 25-34 years old  
## 4 35-44 years old  
## 5 35-44 years old  
## 6 45-54 years old  
## EdLevel  
## 1 Some college/university study without earning a degree  
## 2 Some college/university study without earning a degree  
## 3 Bachelor’s degree (B.A., B.S., B.Eng., etc.)  
## 4 Bachelor’s degree (B.A., B.S., B.Eng., etc.)  
## 5 Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)  
## 6 Associate degree (A.A., A.S., etc.)  
## Employment DevType  
## 1 Employed, full-time;Student, part-time Student  
## 2 Employed, full-time Developer, full-stack  
## 3 Employed, full-time Engineer, site reliability  
## 4 Employed, full-time Developer, full-stack  
## 5 Employed, full-time Developer, full-stack  
## 6 Employed, full-time Developer, full-stack  
## YearsCodePro ConvertedCompYearly JobSat WorkExp  
## 1 7 110000 10 8  
## 2 30 195000 5 30  
## 3 11 230000 8 15  
## 4 23 85000 8 25  
## 5 18 160000 7 20  
## 6 10 115000 6 10

# clean devtype column  
survey\_data\_selected <- survey\_data\_selected %>%  
 mutate(  
 # Normalize separators and whitespace  
 DevType = str\_replace\_all(DevType, "\\s\*;\\s\*", ";"),  
 DevType = str\_trim(DevType),  
 # Split into list column  
 DevType\_split = strsplit(DevType, ";")  
 ) %>%  
 # Trim whitespace for individual dev types  
 mutate(DevType\_split = map(DevType\_split, ~ str\_trim(.x)))  
# Get unique dev types  
dev\_types <- survey\_data\_selected$DevType\_split %>%  
 unlist() %>%  
 unique() %>%  
 sort()  
# Drop I prefer not to say from dev types because it is vague and only represents small portion of data  
dev\_types <- dev\_types[dev\_types != "I prefer not to say"]  
# Create one dummy variable where dev type containing "Developer" is 1 and all others are 0  
survey\_data\_selected$DevType\_Developer <- as.integer(  
 sapply(survey\_data\_selected$DevType\_split, function(x) any(grepl("Developer", x)))  
)  
# drop intermediate columns  
survey\_data\_selected <- survey\_data\_selected %>%  
 select(c(-DevType\_split, -DevType))

# Clean the Employment column  
# list unique values in Employment column  
unique(survey\_data\_selected$Employment)

## [1] "Employed, full-time;Student, part-time"   
## [2] "Employed, full-time"   
## [3] "Employed, full-time;Student, full-time"   
## [4] "Not employed, but looking for work"   
## [5] "Independent contractor, freelancer, or self-employed"   
## [6] "Employed, full-time;Independent contractor, freelancer, or self-employed"   
## [7] "Employed, full-time;Independent contractor, freelancer, or self-employed;Student, part-time"   
## [8] "Student, full-time;Employed, part-time"   
## [9] "Employed, full-time;Retired"   
## [10] "Not employed, but looking for work;Independent contractor, freelancer, or self-employed"   
## [11] "Independent contractor, freelancer, or self-employed;Employed, part-time"   
## [12] "Employed, part-time"   
## [13] "Student, part-time;Employed, part-time"   
## [14] "Employed, full-time;Employed, part-time"   
## [15] "Employed, full-time;Student, full-time;Independent contractor, freelancer, or self-employed"   
## [16] "Employed, full-time;Not employed, and not looking for work"   
## [17] "Student, full-time;Independent contractor, freelancer, or self-employed;Employed, part-time"   
## [18] "Employed, full-time;Student, full-time;Student, part-time;Employed, part-time"   
## [19] "Independent contractor, freelancer, or self-employed;Student, part-time;Employed, part-time"   
## [20] "Employed, full-time;Not employed, but looking for work"   
## [21] "Employed, full-time;Independent contractor, freelancer, or self-employed;Employed, part-time"   
## [22] "Independent contractor, freelancer, or self-employed;Retired"   
## [23] "Independent contractor, freelancer, or self-employed;Student, part-time"   
## [24] "Employed, full-time;Student, full-time;Employed, part-time"   
## [25] "Not employed, but looking for work;Independent contractor, freelancer, or self-employed;Employed, part-time"

# drop "Employed, full-time;Not employed, but looking for work" and "Employed, full-time;Not employed, and not looking for work" in Employment column  
survey\_data\_selected <- survey\_data\_selected %>%  
 filter(!Employment %in% c(  
 "Employed, full-time;Not employed, but looking for work",  
 "Employed, full-time;Not employed, and not looking for work"  
 ))  
  
# Create binary columns for each employment type  
# Normalize Employment column  
survey\_data\_selected$Employment <- str\_replace\_all(survey\_data\_selected$Employment, "\\s\*;\\s\*", ";")  
survey\_data\_selected$Employment <- str\_trim(survey\_data\_selected$Employment)  
  
# Split Employment column into a list of employment types  
survey\_data\_selected$Employment\_split <- strsplit(survey\_data\_selected$Employment, ";")  
# Trim whitespace for individual employment types  
survey\_data\_selected$Employment\_split <- lapply(survey\_data\_selected$Employment\_split, function(x) {  
 x <- str\_trim(x)  
 x[x != ""] # Remove empty strings  
})  
  
# get unique employment types  
employment\_types <- unique(unlist(survey\_data\_selected$Employment\_split))  
# Create binary columns for each employment type  
for (type in employment\_types) {  
 survey\_data\_selected[[type]] <- as.integer(  
 sapply(survey\_data\_selected$Employment\_split, function(x) type %in% x)  
 )  
}  
  
  
# Rename columns to remove spaces and special characters  
colnames(survey\_data\_selected) <- gsub(" ", "\_", colnames(survey\_data\_selected))  
colnames(survey\_data\_selected) <- gsub(",", "", colnames(survey\_data\_selected))  
colnames(survey\_data\_selected) <- gsub("-", "\_", colnames(survey\_data\_selected))  
  
  
  
# Drop the intermediate Employment\_split column  
survey\_data\_selected <- survey\_data\_selected %>%  
 select(-Employment\_split)  
# Drop Employment column since it's no longer needed  
survey\_data\_selected <- survey\_data\_selected %>%  
 select(-Employment)  
# Drop Student\_full\_time column, Student\_part\_time column and Retired since they are not needed  
survey\_data\_selected <- survey\_data\_selected %>%  
 select(-Student\_full\_time, -Student\_part\_time, -Retired, -Not\_employed\_but\_looking\_for\_work, -Employed\_part\_time, -Independent\_contractor\_freelancer\_or\_self\_employed)

# List unique values in EdLevel column  
unique(survey\_data\_selected$EdLevel)

## [1] "Some college/university study without earning a degree"   
## [2] "Bachelor’s degree (B.A., B.S., B.Eng., etc.)"   
## [3] "Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)"  
## [4] "Associate degree (A.A., A.S., etc.)"   
## [5] "Professional degree (JD, MD, Ph.D, Ed.D, etc.)"   
## [6] "Master’s degree (M.A., M.S., M.Eng., MBA, etc.)"   
## [7] "Primary/elementary school"   
## [8] "Something else"

# Drop records with Something else in EdLevel  
survey\_data\_selected <- survey\_data\_selected %>%  
 filter(EdLevel != "Something else")  
  
# Encode Education Level as ranked numeric variable  
survey\_data\_selected <- survey\_data\_selected %>%  
 mutate(EdLevel = as.integer(factor(EdLevel, levels = c(  
 "Primary/elementary school",  
 "Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)",  
 "Some college/university study without earning a degree",  
 "Associate degree (A.A., A.S., etc.)",  
 "Bachelor’s degree (B.A., B.S., B.Eng., etc.)",  
 "Master’s degree (M.A., M.S., M.Eng., MBA, etc.)",  
 "Professional degree (JD, MD, Ph.D, Ed.D, etc.)"  
 ), ordered = TRUE)))

# drop Age that is prefer not to say since its vague and only represents small portion of data  
survey\_data\_selected <- survey\_data\_selected %>%  
 filter(Age != "Prefer not to say")  
  
#Encode Age as ranked numeric variable  
survey\_data\_selected <- survey\_data\_selected %>%  
 mutate(Age = as.integer(factor(Age, levels = c(  
 "Under 18 years old",  
 "18-24 years old",  
 "25-34 years old",  
 "35-44 years old",  
 "45-54 years old",  
 "55-64 years old",  
 "65 years or older"  
 ), ordered = TRUE)))

The final dataset was checked for structure and the first few rows were displayed to ensure that the data was cleaned and encoded correctly. The cleaned and encoded dataset was then saved for further analysis.

# Checking final structure  
str(survey\_data\_selected)

## 'data.frame': 3185 obs. of 8 variables:  
## $ Age : int 3 5 3 4 4 5 5 5 6 3 ...  
## $ EdLevel : int 3 3 5 5 2 4 7 5 6 4 ...  
## $ YearsCodePro : int 7 30 11 23 18 10 18 26 40 4 ...  
## $ ConvertedCompYearly: int 110000 195000 230000 85000 160000 115000 300000 140000 200000 93000 ...  
## $ JobSat : int 10 5 8 8 7 6 1 8 8 8 ...  
## $ WorkExp : int 8 30 15 25 20 10 28 23 40 4 ...  
## $ DevType\_Developer : int 0 1 0 1 1 1 1 1 1 1 ...  
## $ Employed\_full\_time : int 1 1 1 1 1 1 1 1 1 1 ...

head(survey\_data\_selected)

## Age EdLevel YearsCodePro ConvertedCompYearly JobSat WorkExp DevType\_Developer  
## 1 3 3 7 110000 10 8 0  
## 2 5 3 30 195000 5 30 1  
## 3 3 5 11 230000 8 15 0  
## 4 4 5 23 85000 8 25 1  
## 5 4 2 18 160000 7 20 1  
## 6 5 4 10 115000 6 10 1  
## Employed\_full\_time  
## 1 1  
## 2 1  
## 3 1  
## 4 1  
## 5 1  
## 6 1

# Saving cleaned encoded data  
save(survey\_data\_selected, file = "survey\_data\_cleaned\_encoded.RData")

Variance Inflation Factor (VIF) and correlation matrices were used to check for multicollinearity using job satisfaction as the dependent variable before final regression modeling to ensure the reliability of the results. The VIF values were plotted on a bar plot to visualize the multicollinearity. A VIF value greater than 5 indicates a potential multicollinearity issue. The correlation matrix was also displayed as a table using the stargazer package for better readability.

# Load the cleaned and encoded dataset  
load("survey\_data\_cleaned\_encoded.RData")  
  
# check for multicollinearity  
library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

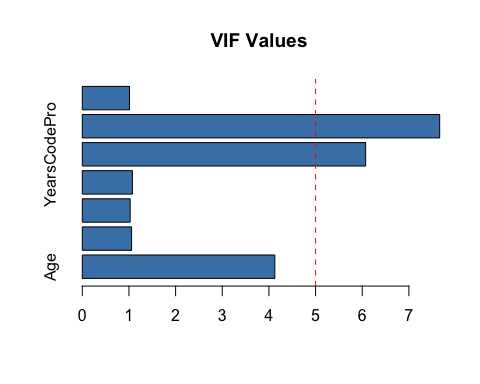
## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

# Check for multicollinearity using VIF  
vif\_model <- lm(JobSat ~ Age + EdLevel + DevType\_Developer + ConvertedCompYearly + YearsCodePro + WorkExp + Employed\_full\_time, data = survey\_data\_selected)  
vif(vif\_model)

## Age EdLevel DevType\_Developer ConvertedCompYearly   
## 4.126664 1.056655 1.026062 1.074054   
## YearsCodePro WorkExp Employed\_full\_time   
## 6.073114 7.660026 1.011128

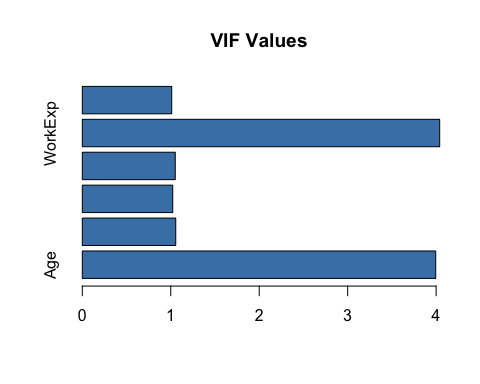
# barplot for VIF values  
vif\_values <- vif(vif\_model)  
# Create a barplot for VIF values  
barplot(vif\_values, main = "VIF Values", horiz = TRUE, col = "steelblue")  
# Add a vertical line at VIF = 5  
abline(v = 5, col = "red", lty = 2)



vif\_model2 <- lm(JobSat ~ Age + EdLevel + DevType\_Developer + ConvertedCompYearly + WorkExp + Employed\_full\_time, data = survey\_data\_selected)  
vif(vif\_model2)

## Age EdLevel DevType\_Developer ConvertedCompYearly   
## 3.995599 1.056471 1.022619 1.050742   
## WorkExp Employed\_full\_time   
## 4.040328 1.011055

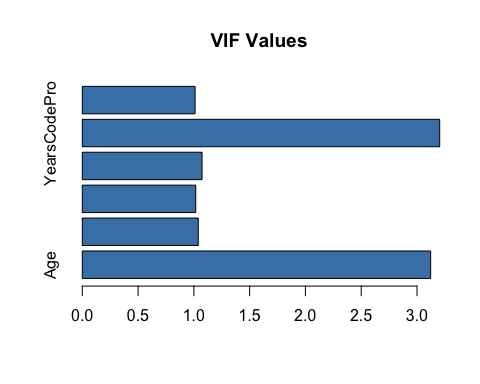
# barplot for VIF values  
vif\_values2 <- vif(vif\_model2)  
# Create a barplot for VIF values  
barplot(vif\_values2, main = "VIF Values", horiz = TRUE, col = "steelblue")  
# Add a vertical line at VIF = 5  
abline(v = 5, col = "red", lty = 2)



vif\_model3 <- lm(JobSat ~ Age + EdLevel + DevType\_Developer + ConvertedCompYearly + YearsCodePro + Employed\_full\_time, data = survey\_data\_selected)  
vif(vif\_model3)

## Age EdLevel DevType\_Developer ConvertedCompYearly   
## 3.122377 1.039304 1.016838 1.073444   
## YearsCodePro Employed\_full\_time   
## 3.203302 1.010924

# barplot for VIF values  
vif\_values3 <- vif(vif\_model3)  
# Create a barplot for VIF values  
barplot(vif\_values3, main = "VIF Values", horiz = TRUE, col = "steelblue")  
# Add a vertical line at VIF = 5  
abline(v = 5, col = "red", lty = 2)



# Check for multicollinearity using correlation matrix  
correlation\_matrix <- cor(survey\_data\_selected[, c("Age", "EdLevel", "DevType\_Developer", "ConvertedCompYearly", "YearsCodePro", "WorkExp")], use = "pairwise.complete.obs")  
  
# Display the correlation matrix as table with stargazer  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

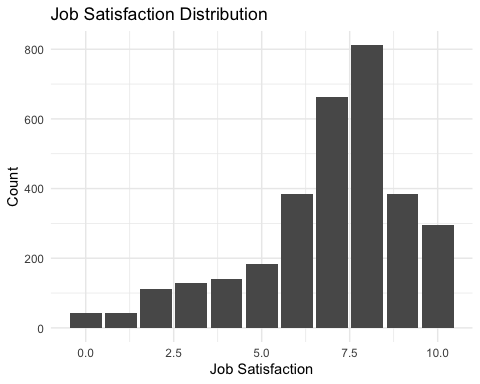
stargazer(correlation\_matrix, type = "text", title = "Correlation Matrix", digits = 4)

##   
## Correlation Matrix  
## ==============================================================================================  
## Age EdLevel DevType\_Developer ConvertedCompYearly YearsCodePro WorkExp  
## ----------------------------------------------------------------------------------------------  
## Age 1 0.0560 -0.0273 0.1585 0.8213 0.8615   
## EdLevel 0.0560 1 -0.1198 0.0911 -0.0048 -0.0273  
## DevType\_Developer -0.0273 -0.1198 1 -0.0532 -0.0261 -0.0544  
## ConvertedCompYearly 0.1585 0.0911 -0.0532 1 0.2264 0.1842   
## YearsCodePro 0.8213 -0.0048 -0.0261 0.2264 1 0.9085   
## WorkExp 0.8615 -0.0273 -0.0544 0.1842 0.9085 1   
## ----------------------------------------------------------------------------------------------

In the initial model,some VIF values were above the commonly accepted threshold of 5, indicating potential multicollinearity issues (YearsCodePro and WorkExp). The model was refined by dropping the YearsCodePro first and subsequently the WorkExp variable, which resulted in lower VIF values. The latter variable resulted in lowest VIF values, hence WorkExp was dropped from the analysis. The correlation matrix was also checked to ensure that there were no strong correlations between the independent variables. Overall, the VIF values and correlation matrix indicated that multicollinearity was not a significant issue in the final mode and all the selected predictors were appropriate to be included in the models. The final model included Age, EdLevel, DevType\_Developer, ConvertedCompYearly, and Employed\_full\_time as independent variables.

The distribution of job satisfaction was explored using a bar plot. The plot showed the frequency of each job satisfaction level, providing insights into the overall job satisfaction levels among developers. To identify the significant factors influencing job satisfaction, a stepwise regression analysis was performed. The stepwise regression model was built using the lm() function in R, with JobSat as the dependent variable and all other variables as independent variables. The stepAIC() function from the MASS package was used to perform the stepwise regression, which automatically selects the best model based on AIC criteria. The summary of the final model was displayed using the stargazer package for better readability.

# Explore job satisfaction  
library(ggplot2)  
ggplot(survey\_data\_selected, aes(x = JobSat)) +  
 geom\_bar() +  
 labs(title = "Job Satisfaction Distribution", x = "Job Satisfaction", y = "Count") +  
 theme\_minimal()



# perform regression analysis with JobSat as dependent variable using stepwise regression  
library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

# Perform stepwise regression with both directions  
stepwise\_model\_both <- stepAIC(lm(JobSat ~ ., data = survey\_data\_selected), direction = "both")

## Start: AIC=5050.7  
## JobSat ~ Age + EdLevel + YearsCodePro + ConvertedCompYearly +   
## WorkExp + DevType\_Developer + Employed\_full\_time  
##   
## Df Sum of Sq RSS AIC  
## - EdLevel 1 2.108 15477 5049.1  
## - DevType\_Developer 1 4.769 15479 5049.7  
## - WorkExp 1 6.003 15481 5049.9  
## <none> 15475 5050.7  
## - Employed\_full\_time 1 25.199 15500 5053.9  
## - Age 1 25.549 15500 5053.9  
## - YearsCodePro 1 38.456 15513 5056.6  
## - ConvertedCompYearly 1 56.636 15531 5060.3  
##   
## Step: AIC=5049.13  
## JobSat ~ Age + YearsCodePro + ConvertedCompYearly + WorkExp +   
## DevType\_Developer + Employed\_full\_time  
##   
## Df Sum of Sq RSS AIC  
## - DevType\_Developer 1 5.696 15482 5048.3  
## - WorkExp 1 7.065 15484 5048.6  
## <none> 15477 5049.1  
## + EdLevel 1 2.108 15475 5050.7  
## - Employed\_full\_time 1 24.421 15501 5052.2  
## - Age 1 28.576 15505 5053.0  
## - YearsCodePro 1 38.225 15515 5055.0  
## - ConvertedCompYearly 1 55.166 15532 5058.5  
##   
## Step: AIC=5048.3  
## JobSat ~ Age + YearsCodePro + ConvertedCompYearly + WorkExp +   
## Employed\_full\_time  
##   
## Df Sum of Sq RSS AIC  
## - WorkExp 1 6.127 15489 5047.6  
## <none> 15482 5048.3  
## + DevType\_Developer 1 5.696 15477 5049.1  
## + EdLevel 1 3.034 15479 5049.7  
## - Employed\_full\_time 1 24.445 15507 5051.3  
## - Age 1 27.900 15510 5052.0  
## - YearsCodePro 1 40.046 15522 5054.5  
## - ConvertedCompYearly 1 53.488 15536 5057.3  
##   
## Step: AIC=5047.56  
## JobSat ~ Age + YearsCodePro + ConvertedCompYearly + Employed\_full\_time  
##   
## Df Sum of Sq RSS AIC  
## <none> 15489 5047.6  
## + WorkExp 1 6.127 15482 5048.3  
## + DevType\_Developer 1 4.758 15484 5048.6  
## + EdLevel 1 4.084 15484 5048.7  
## - Age 1 21.780 15510 5050.0  
## - Employed\_full\_time 1 23.915 15512 5050.5  
## - ConvertedCompYearly 1 52.408 15541 5056.3  
## - YearsCodePro 1 123.375 15612 5070.8

# Display the summary of the both direction stepwise model  
summary(stepwise\_model\_both)

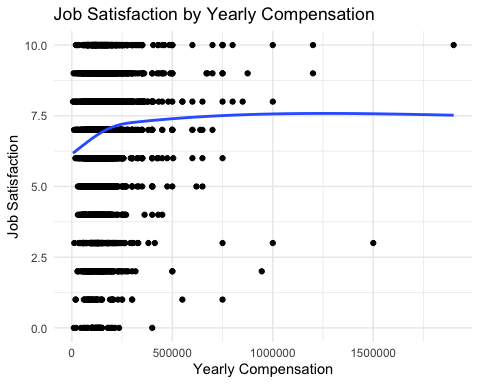
##   
## Call:  
## lm(formula = JobSat ~ Age + YearsCodePro + ConvertedCompYearly +   
## Employed\_full\_time, data = survey\_data\_selected)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.2086 -0.8788 0.3717 1.3684 3.9082   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.287e+00 2.671e-01 23.542 < 2e-16 \*\*\*  
## Age -1.345e-01 6.362e-02 -2.115 0.03454 \*   
## YearsCodePro 3.659e-02 7.271e-03 5.033 5.1e-07 \*\*\*  
## ConvertedCompYearly 1.236e-06 3.768e-07 3.280 0.00105 \*\*   
## Employed\_full\_time 4.681e-01 2.112e-01 2.216 0.02677 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.207 on 3180 degrees of freedom  
## Multiple R-squared: 0.02056, Adjusted R-squared: 0.01933   
## F-statistic: 16.69 on 4 and 3180 DF, p-value: 1.516e-13

stargazer(stepwise\_model\_both, type = "text", align=TRUE)

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## JobSat   
## -----------------------------------------------  
## Age -0.135\*\*   
## (0.064)   
##   
## YearsCodePro 0.037\*\*\*   
## (0.007)   
##   
## ConvertedCompYearly 0.00000\*\*\*   
## (0.00000)   
##   
## Employed\_full\_time 0.468\*\*   
## (0.211)   
##   
## Constant 6.287\*\*\*   
## (0.267)   
##   
## -----------------------------------------------  
## Observations 3,185   
## R2 0.021   
## Adjusted R2 0.019   
## Residual Std. Error 2.207 (df = 3180)   
## F Statistic 16.690\*\*\* (df = 4; 3180)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# Explore the relationship between JobSat and ConvertedCompYearly  
ggplot(survey\_data\_selected, aes(x = ConvertedCompYearly, y = JobSat)) +  
 geom\_point() +  
 geom\_smooth(method = "auto", se = FALSE) +  
 labs(title = "Job Satisfaction by Yearly Compensation", x = "Yearly Compensation", y = "Job Satisfaction") +  
 theme\_minimal()

## `geom\_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'



# Explore the relationship between JobSat and YearsCodePro using histogram  
ggplot(survey\_data\_selected, aes(x = YearsCodePro, y = JobSat)) +  
 geom\_point() +  
 geom\_smooth(method = "auto", se = FALSE) +  
 labs(title = "Job Satisfaction by Years of Professional Coding Experience", x = "Years of Professional Coding Experience", y = "Job Satisfaction") +  
 theme\_minimal()

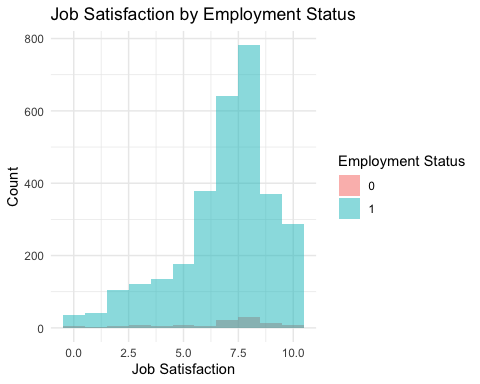
## `geom\_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'



# Explore the relationship between JobSat and Age with barplot  
ggplot(survey\_data\_selected, aes(x = Age, fill = as.factor(JobSat))) +  
 geom\_bar(position = "dodge") +  
 labs(title = "Job Satisfaction by Age", x = "Age", y = "Count") +  
 theme\_minimal() +  
 scale\_fill\_discrete(name = "Job Satisfaction")



# Explore the relationship between JobSat and Employment\_full\_time with histogram  
ggplot(survey\_data\_selected, aes(x = JobSat, fill = as.factor(Employed\_full\_time))) +  
 geom\_histogram(binwidth = 1, position = "identity", alpha = 0.5) +  
 labs(title = "Job Satisfaction by Employment Status", x = "Job Satisfaction", y = "Count") +  
 theme\_minimal() +  
 scale\_fill\_discrete(name = "Employment Status")

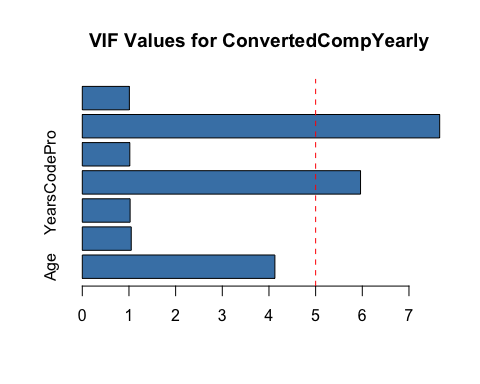


For the second model, the same process was followed to check for multicollinearity using VIF and correlation matrices with ConvertedCompYearly as the dependent variable. The VIF values were plotted on a bar plot to visualize the multicollinearity. A VIF value greater than 5 indicates a potential multicollinearity issue. The correlation matrix was also displayed as a table using the stargazer package for better readability.

# perform regression analysis with ConvertedCompYearly as dependent variable  
  
# Check for multicollinearity using VIF  
vif\_model\_comp <- lm(ConvertedCompYearly ~ Age + EdLevel + DevType\_Developer + YearsCodePro + JobSat + WorkExp + Employed\_full\_time, data = survey\_data\_selected)  
vif(vif\_model\_comp)

## Age EdLevel DevType\_Developer YearsCodePro   
## 4.126974 1.048665 1.024638 5.962195   
## JobSat WorkExp Employed\_full\_time   
## 1.018186 7.658221 1.008895

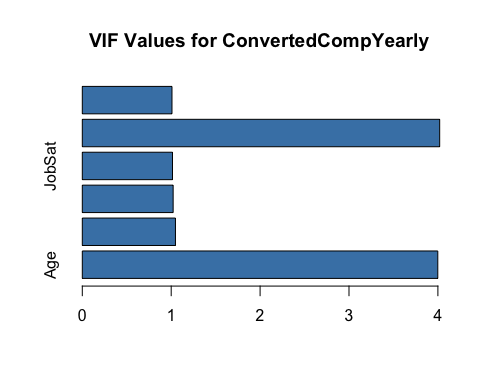
# barplot for VIF values  
vif\_values\_comp <- vif(vif\_model\_comp)  
# Create a barplot for VIF values  
barplot(vif\_values\_comp, main = "VIF Values for ConvertedCompYearly", horiz = TRUE, col = "steelblue")  
# Add a vertical line at VIF = 5  
abline(v = 5, col = "red", lty = 2)



# drop YearsCodepro and check for multicollinearity  
vif\_model\_comp2 <- lm(ConvertedCompYearly ~ Age + EdLevel + DevType\_Developer + JobSat + WorkExp + Employed\_full\_time, data = survey\_data\_selected)  
vif(vif\_model\_comp2)

## Age EdLevel DevType\_Developer JobSat   
## 3.998945 1.047914 1.021907 1.014618   
## WorkExp Employed\_full\_time   
## 4.020137 1.008893

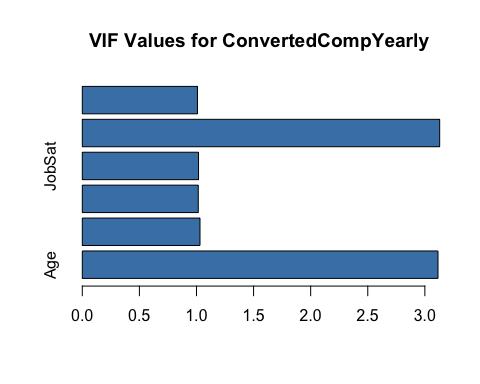
# barplot for VIF values  
vif\_values\_comp2 <- vif(vif\_model\_comp2)  
# Create a barplot for VIF values  
barplot(vif\_values\_comp2, main = "VIF Values for ConvertedCompYearly", horiz = TRUE, col = "steelblue")  
# Add a vertical line at VIF = 5  
abline(v = 5, col = "red", lty = 2)



# drop workexp and check for multicollinearity  
vif\_model\_comp3 <- lm(ConvertedCompYearly ~ Age + EdLevel + DevType\_Developer + JobSat + YearsCodePro + Employed\_full\_time, data = survey\_data\_selected)  
vif(vif\_model\_comp3)

## Age EdLevel DevType\_Developer JobSat   
## 3.115141 1.030775 1.015544 1.017848   
## YearsCodePro Employed\_full\_time   
## 3.129819 1.008619

# barplot for VIF values  
vif\_values\_comp3 <- vif(vif\_model\_comp3)  
# Create a barplot for VIF values  
barplot(vif\_values\_comp3, main = "VIF Values for ConvertedCompYearly", horiz = TRUE, col = "steelblue")  
# Add a vertical line at VIF = 5  
abline(v = 5, col = "red", lty = 2)



# Check for multicollinearity using correlation matrix  
correlation\_matrix\_comp <- cor(survey\_data\_selected[, c("Age", "EdLevel", "JobSat","DevType\_Developer", "YearsCodePro", "WorkExp")], use = "pairwise.complete.obs")  
# Display the correlation matrix as table  
correlation\_matrix\_comp

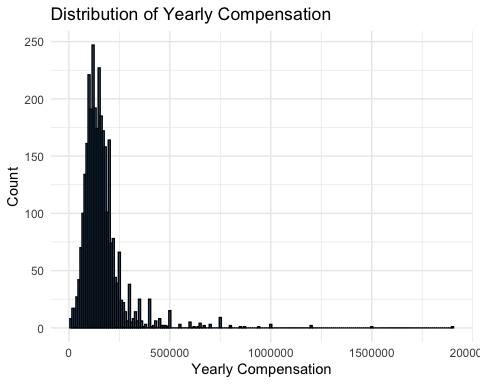
## Age EdLevel JobSat DevType\_Developer  
## Age 1.00000000 0.055974448 0.07487290 -0.02727501  
## EdLevel 0.05597445 1.000000000 -0.01206932 -0.11980331  
## JobSat 0.07487290 -0.012069321 1.00000000 0.01176676  
## DevType\_Developer -0.02727501 -0.119803310 0.01176676 1.00000000  
## YearsCodePro 0.82133104 -0.004788312 0.11792810 -0.02613618  
## WorkExp 0.86147847 -0.027281700 0.10549144 -0.05438975  
## YearsCodePro WorkExp  
## Age 0.821331040 0.86147847  
## EdLevel -0.004788312 -0.02728170  
## JobSat 0.117928095 0.10549144  
## DevType\_Developer -0.026136180 -0.05438975  
## YearsCodePro 1.000000000 0.90848607  
## WorkExp 0.908486070 1.00000000

# Display the correlation matrix as table with stargazer  
library(stargazer)  
stargazer(correlation\_matrix\_comp, type = "text", title = "Correlation Matrix for ConvertedCompYearly", digits = 4)

##   
## Correlation Matrix for ConvertedCompYearly  
## ================================================================================  
## Age EdLevel JobSat DevType\_Developer YearsCodePro WorkExp  
## --------------------------------------------------------------------------------  
## Age 1 0.0560 0.0749 -0.0273 0.8213 0.8615   
## EdLevel 0.0560 1 -0.0121 -0.1198 -0.0048 -0.0273  
## JobSat 0.0749 -0.0121 1 0.0118 0.1179 0.1055   
## DevType\_Developer -0.0273 -0.1198 0.0118 1 -0.0261 -0.0544  
## YearsCodePro 0.8213 -0.0048 0.1179 -0.0261 1 0.9085   
## WorkExp 0.8615 -0.0273 0.1055 -0.0544 0.9085 1   
## --------------------------------------------------------------------------------

The distribution of ConvertedCompYearly was explored using a histogram. The histogram showed the distribution of yearly compensation among developers, providing insights into the overall compensation levels. To identify the significant factors influencing ConvertedCompYearly, a stepwise regression analysis was performed. The stepwise regression model was built using the lm() function in R, with ConvertedCompYearly as the dependent variable and all other variables as independent variables. The stepAIC() function from the MASS package was used to perform the stepwise regression, which automatically selects the best model based on AIC criteria. The summary of the final model was displayed using the stargazer package for better readability.

# Explore ConvertedCompYearly  
ggplot(survey\_data\_selected, aes(x = ConvertedCompYearly)) +  
 geom\_histogram(binwidth = 10000, fill = "steelblue", color = "black") +  
 labs(title = "Distribution of Yearly Compensation", x = "Yearly Compensation", y = "Count") +  
 theme\_minimal()



summary(survey\_data\_selected$ConvertedCompYearly)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 5800 105000 145000 163318 190000 1900000

# Perform regression analysis with ConvertedCompYearly as dependent variable using stepwise regression  
stepwise\_model\_comp <- stepAIC(lm(ConvertedCompYearly ~ ., data = survey\_data\_selected), direction = "both")

## Start: AIC=73541.71  
## ConvertedCompYearly ~ Age + EdLevel + YearsCodePro + JobSat +   
## WorkExp + DevType\_Developer + Employed\_full\_time  
##   
## Df Sum of Sq RSS AIC  
## - WorkExp 1 2.1075e+10 3.3812e+13 73542  
## <none> 3.3791e+13 73542  
## - Age 1 5.3245e+10 3.3844e+13 73545  
## - DevType\_Developer 1 5.7400e+10 3.3848e+13 73545  
## - JobSat 1 1.2367e+11 3.3914e+13 73551  
## - Employed\_full\_time 1 1.2991e+11 3.3920e+13 73552  
## - EdLevel 1 2.6209e+11 3.4053e+13 73564  
## - YearsCodePro 1 7.1417e+11 3.4505e+13 73606  
##   
## Step: AIC=73541.69  
## ConvertedCompYearly ~ Age + EdLevel + YearsCodePro + JobSat +   
## DevType\_Developer + Employed\_full\_time  
##   
## Df Sum of Sq RSS AIC  
## <none> 3.3812e+13 73542  
## + WorkExp 1 2.1075e+10 3.3791e+13 73542  
## - DevType\_Developer 1 5.1491e+10 3.3863e+13 73545  
## - Age 1 1.2133e+11 3.3933e+13 73551  
## - JobSat 1 1.2186e+11 3.3933e+13 73551  
## - Employed\_full\_time 1 1.3169e+11 3.3943e+13 73552  
## - EdLevel 1 2.8676e+11 3.4098e+13 73567  
## - YearsCodePro 1 1.0574e+12 3.4869e+13 73638

# Display the summary of the both direction stepwise model  
summary(stepwise\_model\_comp)

##   
## Call:  
## lm(formula = ConvertedCompYearly ~ Age + EdLevel + YearsCodePro +   
## JobSat + DevType\_Developer + Employed\_full\_time, data = survey\_data\_selected)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -213031 -51451 -18784 21327 1722812   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 68063.7 15765.1 4.317 1.63e-05 \*\*\*  
## Age -10088.1 2987.3 -3.377 0.000742 \*\*\*  
## EdLevel 9095.6 1752.0 5.192 2.22e-07 \*\*\*  
## YearsCodePro 3366.9 337.7 9.969 < 2e-16 \*\*\*  
## JobSat 2800.6 827.5 3.384 0.000722 \*\*\*  
## DevType\_Developer -9259.8 4209.1 -2.200 0.027883 \*   
## Employed\_full\_time 34761.5 9880.6 3.518 0.000441 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 103100 on 3178 degrees of freedom  
## Multiple R-squared: 0.07176, Adjusted R-squared: 0.07001   
## F-statistic: 40.95 on 6 and 3178 DF, p-value: < 2.2e-16

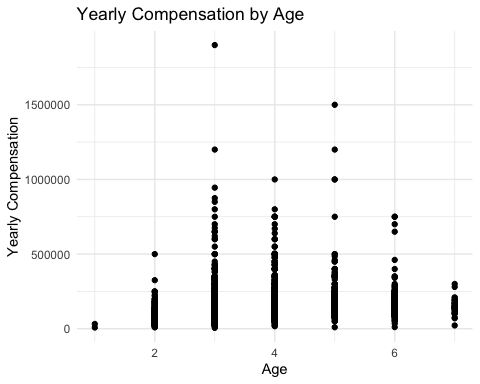
stargazer(stepwise\_model\_comp, type = "text", align=TRUE)

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## ConvertedCompYearly   
## -----------------------------------------------  
## Age -10,088.080\*\*\*   
## (2,987.265)   
##   
## EdLevel 9,095.649\*\*\*   
## (1,751.994)   
##   
## YearsCodePro 3,366.872\*\*\*   
## (337.729)   
##   
## JobSat 2,800.595\*\*\*   
## (827.523)   
##   
## DevType\_Developer -9,259.831\*\*   
## (4,209.130)   
##   
## Employed\_full\_time 34,761.510\*\*\*   
## (9,880.567)   
##   
## Constant 68,063.750\*\*\*   
## (15,765.110)   
##   
## -----------------------------------------------  
## Observations 3,185   
## R2 0.072   
## Adjusted R2 0.070   
## Residual Std. Error 103,146.900 (df = 3178)   
## F Statistic 40.950\*\*\* (df = 6; 3178)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# scatter plot age and compensation  
ggplot(survey\_data\_selected, aes(x = Age, y = ConvertedCompYearly)) +  
 geom\_point() +  
 geom\_smooth(method = "auto", se = FALSE) +  
 labs(title = "Yearly Compensation by Age", x = "Age", y = "Yearly Compensation") +  
 theme\_minimal()

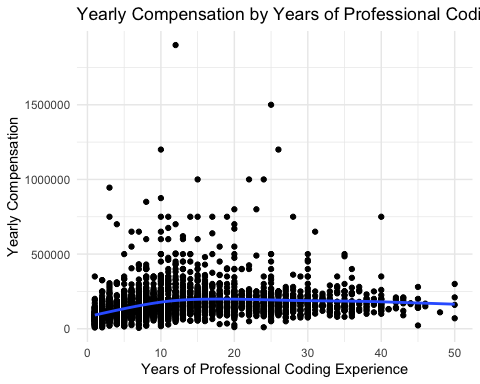
## `geom\_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'

## Warning: Failed to fit group -1.  
## Caused by error in `smooth.construct.cr.smooth.spec()`:  
## ! x has insufficient unique values to support 10 knots: reduce k.

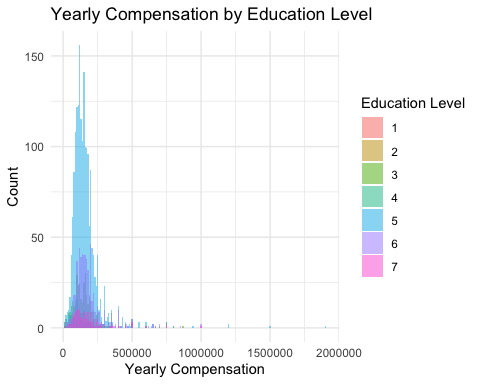


# scatter plot yearscodepro and compensation  
ggplot(survey\_data\_selected, aes(x = YearsCodePro, y = ConvertedCompYearly)) +  
 geom\_point() +  
 geom\_smooth(method = "auto", se = FALSE) +  
 labs(title = "Yearly Compensation by Years of Professional Coding Experience", x = "Years of Professional Coding Experience", y = "Yearly Compensation") +  
 theme\_minimal()

## `geom\_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'



# Explore the relationship between ConvertedCompYearly and EdLevel with histogram  
ggplot(survey\_data\_selected, aes(x = ConvertedCompYearly, fill = as.factor(EdLevel))) +  
 geom\_histogram(binwidth = 10000, position = "identity", alpha = 0.5) +  
 labs(title = "Yearly Compensation by Education Level", x = "Yearly Compensation", y = "Count") +  
 theme\_minimal() +  
 scale\_fill\_discrete(name = "Education Level")



To explore whether the relationship between age and job satisfaction is nonlinear, a quadratic regression model was estimated and compared to a standard linear model. Initial LOESS plots suggested a curved relationship, with job satisfaction increasing up to a point and then declining with age.

# Quadratic regression analysis  
# Example for Job Satisfaction  
ggplot(survey\_data\_selected, aes(x = Age, y = JobSat)) +  
 geom\_point(alpha = 0.3) +  
 geom\_smooth(method = "loess", se = FALSE) +  
 labs(title = "Job Satisfaction vs. Age")

## `geom\_smooth()` using formula = 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : pseudoinverse used at 3

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : neighborhood radius 1

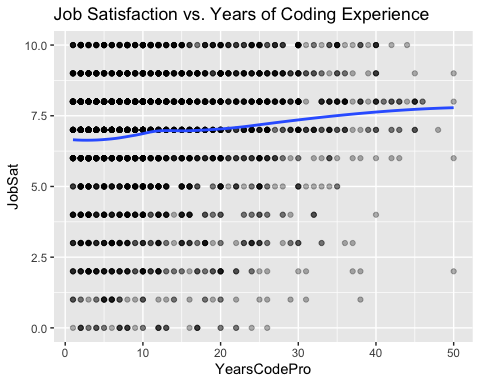
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,  
## : There are other near singularities as well. 1



ggplot(survey\_data\_selected, aes(x = YearsCodePro, y = JobSat)) +  
 geom\_point(alpha = 0.3) +  
 geom\_smooth(method = "loess", se = FALSE) +  
 labs(title = "Job Satisfaction vs. Years of Coding Experience")

## `geom\_smooth()` using formula = 'y ~ x'



# Job Satisfaction ~ Age  
model\_linear <- lm(JobSat ~ Age, data = survey\_data\_selected)  
model\_quadratic <- lm(JobSat ~ Age + I(Age^2), data = survey\_data\_selected)  
  
summary(model\_linear)

##   
## Call:  
## lm(formula = JobSat ~ Age, data = survey\_data\_selected)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.2431 -0.9341 0.2204 1.2204 3.3749   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.31612 0.14218 44.425 < 2e-16 \*\*\*  
## Age 0.15450 0.03647 4.236 2.34e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.223 on 3183 degrees of freedom  
## Multiple R-squared: 0.005606, Adjusted R-squared: 0.005294   
## F-statistic: 17.94 on 1 and 3183 DF, p-value: 2.339e-05

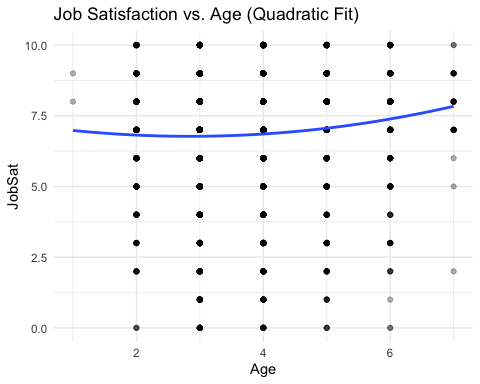
summary(model\_quadratic)

##   
## Call:  
## lm(formula = JobSat ~ Age + I(Age^2), data = survey\_data\_selected)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.3857 -0.8555 0.2255 1.2255 3.2255   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.26764 0.43348 16.766 <2e-16 \*\*\*  
## Age -0.34845 0.21952 -1.587 0.1125   
## I(Age^2) 0.06136 0.02641 2.323 0.0202 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.221 on 3182 degrees of freedom  
## Multiple R-squared: 0.00729, Adjusted R-squared: 0.006666   
## F-statistic: 11.68 on 2 and 3182 DF, p-value: 8.798e-06

anova(model\_linear, model\_quadratic) # Compare fit

## Analysis of Variance Table  
##   
## Model 1: JobSat ~ Age  
## Model 2: JobSat ~ Age + I(Age^2)  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 3183 15725   
## 2 3182 15698 1 26.633 5.3983 0.02022 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Plot the fitted values  
ggplot(survey\_data\_selected, aes(x = Age, y = JobSat)) +  
 geom\_point(alpha = 0.3) +  
 geom\_smooth(method = "lm", formula = y ~ poly(x, 2), se = FALSE) +  
 labs(title = "Job Satisfaction vs. Age (Quadratic Fit)") +  
 theme\_minimal()



## Results

### Job Satisfaction

The stepwise regression analysis for job satisfaction revealed that the final model included Age,YearsCodePro, ConvertedCompYearly, and Employed\_full\_time as significant predictors. The model asserts that age was negatively associated with job satisfaction (B=-0.13,p=0.035) suggesting that older developers tend to report lower job satisfaction levels compared to their younger counterparts. The model also indicated that years of professional coding experience had a positive and significant impact on job satifaction (B=0.037,p<0.001), suggesting that satisfaction increases with professional experience.Yearly compensation also shows a positive association with job satisfaction (B=1.24e-6, p=0.001), supporting the view that higher compensation is associated with greater job satisfaction. Finally, the model indicates that full-time employment status is positively associated with job satisfaction (B=0.47,p=0.027), suggesting that full-time employees tend to report higher job satisfaction levels compared to part-time or contract workers.Overall, the model was statistically significant (F(4,3180) = 16.69, p<0.001), though the explained variance was relatively low (R^2 = 0.021), suggesting that while the identified variables contribute to job satisfaction, other unmeasured factors also play a role. The results suggest that age has an inverse relationship with job satisfaction, indicating potential bias against older developers in the tech industry.

### Yearly Compensation

A stepwise linear regression model was used to identify significant predictors of ConvertedCompYearly (yearly compensation). The final model included Age, Education Level, Years of Professional Coding Experience, Job Satisfaction, Developer Role, and Full-Time Employment Status.Age had a significant negative effect on compensation (β = -10,088, p < 0.001), suggesting that, holding other variables constant, older developers tend to earn less.Education Level and YearsCodePro were both positively associated with salary (β = 9,096 and 3,367 respectively; p < 0.001), consistent with expectations that higher education and experience increase earnings.Job Satisfaction also had a significant positive effect (β = 2,801, p < 0.001), indicating that more satisfied developers tend to earn more. Interestingly, identifying as a developer was negatively associated with compensation (β = -9,260, p = 0.028), possibly due to non-developer roles (e.g., managers, executives) having higher pay scales. And finally, Full-time employment significantly affects compensation (β = 34,761, p < 0.001), confirming that full-time roles pay substantially more than other employment types.The model was statistically significant (F(6, 3178) = 40.95, p < 0.0001), with an R^2 of 0.072, indicating that the model explains about 7.2% of the variance in yearly compensation.These results suggest that while skills, education, and job satisfaction are important drivers of compensation, age still negatively impacts earnings even after controlling for key professional qualifications. This pattern is consistent with broader concerns about ageism in the tech workforce.

### Quadratic Regression

To test for a nonlinear relationship between age and job satisfaction, a quadratic regression model was compared to a standard linear model. The linear model included Age and Years of Professional Coding Experience (YearsCodePro) as predictors of job satisfaction.The quadratic model, which included both Age and Age², provided a significantly better fit than the linear model, as confirmed by the ANOVA test.The quadratic model indicated that job satisfaction increased with age up to a certain point, after which it began to decline. This suggests that while older developers may initially experience higher job satisfaction, this trend reverses at some point in their careers. The model’s coefficients indicated that the relationship between age and job satisfaction is not strictly linear, highlighting the complexity of factors influencing job satisfaction over time.The quadratic term (Age²) captures the inverted-U shape seen in the LOESS plot and the quadratic fit visualization.

## Conclusion

The analysis of the Stack Overflow Annual Developer Survey data for 2024 revealed significant insights into the impact of age on job satisfaction and compensation in the tech industry. The findings suggest that older developers tend to earn lower salaries and report lower job satisfaction compared to their younger counterparts. The results also indicate that years of professional coding experience, education level, and full-time employment status are positively associated with both job satisfaction and compensation. The quadratic regression analysis further suggests a complex relationship between age and job satisfaction, indicating that while older developers may initially experience higher job satisfaction, this trend reverses at some point in their careers. Overall, these findings highlight the need for policy makers to strengthen enforcement of the Age Discrimination in Employment Act (ADEA) and promote age inclusivity in the tech industry.

## References

Discrimination: What it is and how to cope. (2024, May 16). <https://www.apa.org>. <https://www.apa.org/topics/racism-bias-discrimination/types-stress>

Kotler, S. (2015). “Is Silicon Valley Ageist Or Just Smart?”, Forbes, Retrieved from <https://www.forbes.com/sites/stevenkotler/2015/02/14/is-silicon-valley-ageist-or-just-smart/?sh=5fb0b05eed65>

Sevilla, C. (2019). “Everyday ageism in the tech industry”, CWJobs. Retreived from <https://www.cwjobs.co.uk/advice/ageism-in-tech>

Stampler, L (2014). “Ageism in Silicon Valley Is Causing an Uptick in Plastic Surgeries”, Time, Retrieved from <https://time.com/35524/ageism-in-silicon-valley-is-causing-an-uptick-in-plastic-surgeries/>

Stanford Encyclopedia of Philosophy (2020). “Discrimination”. Retrieved from <https://plato.stanford.edu/entries/discrimination/>

* UN General Assembly. (1966). International Covenant on Civil and Political Rights. Retrieved from <https://www.ohchr.org/en/instruments-mechanisms/international-human-rights-instruments/international-covenant-civil-and-political-rights>