Práctica 2 - Tipología y ciclo de vida de los datos

Autores: Luis Arnaldo Torres González, Gabriel Patricio Bonilla Sanchez

Junio - 2021

Índice general


```
library(kableExtra)
```

Introducción

En esta práctica se elabora un caso práctico orientado a aprender a identificar los datos relevantes para un proyecto analítico y usar las herramientas de integración, limpieza, validación y análisis de las mismas. Para hacer esta práctica tendréis que trabajar en grupos de 2 personas. Tendréis que entregar un solo archivo con el enlace Github (https://github.com) donde se encuentren las soluciones incluyendo los nombres de los componentes del equipo. Podéis utilizar la Wiki de Github para describir vuestro equipo y los diferentes archivos que corresponden a vuestra entrega. Cada miembro del equipo tendrá que contribuir con su usuario Github. Aunque no se trata del mismo enunciado, los siguientes ejemplos de ediciones anteriores os pueden servir como guía:

- Ejemplo: https://github.com/Bengis/nba-gap-cleaning
- Ejemplo complejo (archivo adjunto).

Competencias

En esta práctica se desarrollan las siguientes competencias del Máster de Data Science: * Capacidad de analizar un problema en el nivel de abstracción adecuado a cada situación y aplicar las habilidades y conocimientos adquiridos para abordarlo y resolverlo. * Capacidad para aplicar las técnicas específicas de tratamiento de datos (integración, ransformación, limpieza y validación) para su posterior análisis.

Objetivos

Los objetivos concretos de esta práctica son: * Aprender a aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios o multidisciplinares. * Saber identificar los datos relevantes y los tratamientos necesarios (integración, limpieza y validación) para llevar a cabo un proyecto analítico. * Aprender a analizar los datos adecuadamente para abordar la información contenida en los datos. * Identificar la mejor representación de los resultados para aportar conclusiones sobre el problema planteado en el proceso analítico. * Actuar con los principios éticos y legales relacionados con la manipulación de datos en función del ámbito de aplicación. * Desarrollar las habilidades de aprendizaje que les permitan continuar estudiando de un modo que tendrá que ser en gran medida autodirigido o autónomo. * Desarrollar la capacidad de búsqueda, gestión y uso de información y recursos en el ámbito de la ciencia de datos.

1 Descripción del dataset

En este apartado resummiremos los pasos necesarios para la preparación del dataset final realizado en la PRA 1:

El dataset seleccionado ha sido obtenido desde el siguiente enlace: https://www.kaggle.com/aitzaz/stack-overflow-developer-survey-2020. Este juego de datos contiene los resultados de la Encuesta Anual a Desarrolladores StackOverflow 2020. Se obtuvo alrededor de 65000 participaciones de programadores y desarrolladores de 180 paises. La encuesta aborda varios ámbitos, tanto a nivel de experiencia, formación académica y skills (habilidades técnicas) en diferentes tecnologías que el encuestado ha ido adquiriendo a lo largo del tiempo.

Esta encuesta anual ha recolectado datos sobre 61 variables que se pasan a detallar a continuación:

- *1) Respondent:* número de identificación del encuestado aleatorizado (no en orden de tiempo de respuesta de la encuesta)
- *2) MainBranch:* ¿Cuál de las siguientes opciones te describe mejor hoy?
- *3) Hobbyist:* ¿Desarrollas como pasatiempo?
- *4) Age:* ¿Cuál es su edad (en años)?
- *5) Age1stCode:* ¿A qué edad escribiste tu primera línea de código o programa?
- *6) CompFreq:* ¿Esa compensación es semanal, mensual o anual?
- *7) CompTotal:* ¿Cuál es su compensación total actual (salario, bonificaciones y beneficios, antes de impuestos y deducciones), en "CurrencySymbol"?. Número entero.
- *8) ConvertedComp:* Salario anual en USD, utilizando el tipo de cambio del 19 de febrero de 2020, asumiendo 12 meses laborales y 50 semanas laborales".
- *9) Country:* País dónde vive.
- *10) CurrencyDesc:* ¿Qué moneda utiliza a diario? Descripción.
- *11) CurrencySymbol:* ¿Qué moneda usa a diario? Forma abreviada.
- *12) DatabaseDesireNextYear:* ¿En qué entornos de base de datos desea trabajar durante el próximo año?
- *13) DatabaseWorkedWith:* ¿En qué entornos de base de datos ha realizado un trabajo de desarrollo extenso durante el año pasado?
- *14) DevType:* ¿Cuál de los siguientes lo describe?
- *15) EdLevel:* ¿Cuál de las siguientes opciones describe mejor el nivel más alto de educación formal que ha completado?
- *16) Employment:* ¿cuál de las siguientes opciones describe mejor su situación laboral actual?

- *17) Ethnicity:* ¿Cuál de los siguientes grupos étnicos lo describe?
- *18) Gender:* ¿Cuál de las siguientes opciones de sexo lo describe?
- *19) JobFactors:* Para el caso de decidiendo entre dos ofertas de trabajo con la misma compensación, beneficios y ubicación. ¿Qué factores son los más importantes para usted?
- *20) JobSat:* ¿Qué tan satisfecho está con su trabajo actual?
- *21) JobSeek:* ¿Cuál de las siguientes opciones describe mejor su estado actual de búsqueda de empleo?
- *22) LanguageDesireNextYear:* "¿En qué lenguajes de programación, scripting y marcado desea trabajar durante el próximo año?.
- *23) LanguageWorkedWith:* ¿En qué lenguajes de programación, scripting y marcado ha realizado un trabajo de desarrollo extenso durante el año pasado?.
- *24) MiscTechDesireNextYear:* ¿En qué otros frameworks, bibliotecas y herramientas desea trabajar durante el próximo año?.
- *25) MiscTechWorkedWith:* ¿En qué otros frameworks, bibliotecas y herramientas ha realizado un trabajo de desarrollo extenso durante el año pasado?.
- *26) NEWCollabToolsDesireNextYear:* ¿En qué herramientas de colaboración desea trabajar durante el próximo año?
- *27) NEWCollabToolsWorkedWith:* ¿En qué herramientas de colaboración ha realizado un trabajo de desarrollo extenso durante el año pasado?
- *28) NEWDevOps:* ¿Su empresa tiene una persona dedicada a DevOps?
- *29) NEWDevOpsImpt:* ¿Qué importancia tiene la práctica de DevOps para escalar el desarrollo de software?
- *30) NEWEdImpt:* ¿Qué importancia tiene una educación formal, como un título universitario en ciencias de la computación, para su carrera?
- *31) NEWJobHunt:* En general, ¿Cuáles son las motivaciones que lo impulsan a buscar un nuevo trabajo?.
- *32) NEWJobHuntResearch:* Cuando busca trabajo, ¿cómo puede obtener más información sobre una empresa?
- *33) NEWLearn:* ¿Con qué frecuencia aprende un nuevo lenguaje o marco?
- *34) NEWOffTopic:* ¿Crees que Stack Overflow debería relajar las restricciones sobre lo que se considera "fuera de tema"?
- *35) NEWOnboardGood:* ¿Cree que su empresa tiene un buen proceso de incorporación? (Por incorporación, nos referimos al proceso estructurado para que se adapte a su nuevo puesto en una empresa)
- *36) NEWOtherComms:* ¿Es miembro de alguna otra comunidad de desarrolladores en línea?
- *37) NEWOvertime:* ¿Con qué frecuencia trabaja horas extraordinarias o más allá de las expectativas formales de su trabajo?
- *38) NEWPurchaseResearch:* Al comprar una nueva herramienta o software, ¿cómo descubre e investiga las soluciones disponibles?
- *39) NEWPurpleLink:* Busca una solución de codificación en línea y el primer enlace de resultado es violeta porque ya lo visitó. ¿Cómo se siente?
- *40) NEWSOSites:* ¿Cuál de los siguientes sitios de Stack Overflow ha visitado?
- *41) NEWStuck:* ¿Qué hace cuando se queda atascado en un problema?
- *42) OpSys:* ¿Cuál es el sistema operativo principal en el que trabaja?
- *43) OrgSize:* Aproximadamente, ¿cuántas personas emplea la empresa u organización para la que trabaja actualmente?
- *44) PlatformDesireNextYear:* ¿En qué plataformas desea trabajar durante el próximo año?
- *45) PlatformWorkedWith:* ¿En qué plataformas ha realizado un trabajo de desarrollo extenso durante el año pasado?
- *46) PurchaseWhat:* ¿Qué nivel de influencia tiene usted, personalmente, sobre las compras de nueva tecnología en su organización?
- *47) Sexuality:* ¿Cuál de los siguientes lo describe a usted sobre su sexualidad?.
- *48) SOAccount:* ¿Tiene una cuenta de Stack Overflow?
- *49) SOComm:* ¿Te consideras miembro de la comunidad de Stack Overflow?
- *50) SOPartFreq:* ¿Con qué frecuencia diría que participa en preguntas y respuestas en Stack Overflow? Por participar nos referimos a preguntar, responder, votar o comentar preguntas.
- *51) SOVisitFreq:* ¿Con qué frecuencia visita Stack Overflow?
- *52) SurveyEase:* ¿Qué tan fácil o difícil fue completar esta encuesta?
- *53) SurveyLength:* ¿Qué opina de la duración de la encuesta este año?
- *54) Trans:* ¿Eres transgénero?

- *55) UndergradMajor:* ¿Cuál fue su campo de estudio principal?
- *56) WebframeDesireNextYear:* ¿En qué frameworks web desea trabajar durante el próximo año?
- *57) WebframeWorkedWith:* ¿En qué frameworks web ha realizado un extenso trabajo de desarrollo durante el año pasado?
- *58) WelcomeChange:* En comparación con el año pasado, ¿qué tan bienvenido se siente en Stack Overflow?
- *59) WorkWeekHrs:* En promedio, ¿cuántas horas por semana trabaja?
- *60) YearsCode:* Incluyendo cualquier educación, ¿cuántos años ha estado programando en total?
- *61) YearsCodePro:* NO incluye educación, ¿cuántos años ha programado profesionalmente (como parte de su trabajo)?

Las capacidades análiticas del dataset, que se tomaron en cuenta para elegirlo son:

- Cuenta con una cantidad suficientes variables, tanto numéricas, categóricas. Las variables categóricas también pueden volverse a convertir a variables numéricas. Esto permitiría aplicar algoritmos supervisados y no supervisados, donde se puede clasificar a los programadores o desarrolladores según la experticia actual.
- También permite agregar nuevas variables númericas que representen el número de tecnologías que domina cada encuestado.
- Al incluir las tecnologías usadas por desarrolladores en: base de datos, lenguages de programación, frameworks y demás herramientas, permite tener una gran cantidad de preferencias de las que se puede extraer reglas de asociación interesantes sobre las tecnologías más usadas entre los distintos tipos de desarrolladores.
- Cuenta con variables que pueden discretizarse y otras donde se puede aplicar tareas de limpieza y preparación previa antes de aplicar los distintos métodos.

Sin embargo, para efectos del análisis, del dataset original, se excluirán las siguientes variables:

1. Age
2. ConvertedComp
3. Country
4. DatabaseWorkedWith
5. DevType
6. EdLevel
7. Employment
8. LanguageWorkedWith
9. MiscTechWorkedWith
10. NEWCollabToolsWorkedWith
11. NEWLearn
12. NEWOvertime
13. OpSys
14. PlatformWorkedWith
15. SOPartFreq
16. UndergradMajor
17. WebframeWorkedWith
18. WorkWeekHrs
19. YearsCodePro

Muchos de estos campos no son relevantes para el alcance de la Práctica #1 y #2; otros reflejan deseos de los programadores respecto a tecnologías, para lo cual solo tomaremos los datos que reflejan la experiencia actual del programador.

Vamos a trabajar preliminarmente con 19 variables propias del dataset original, de las cuales 4 son númericas (Age, ConvertedComp, WorkWeekHrs y YearsCodePro). También tenemos variables no numéricas, las cuales vamos a realizar un análisis más detallado posteriormente, generando variables númericas a partir de ellas, las cuales son:

- DatabaseWorkedWith
- LanguageWorkedWith
- MiscTechWorkedWith
- NEWCollabToolsWorkedWith
- PlatformWorkedWith
- WebframeWorkedWith

Estas nuevas variables numéricas a generarse posteriormente servirán principalmente cuando se defina el número total de tecnologías conocidas y usadas por los encuestados.

Este nuevo dataset de variables numéricas, ayudará a dar respuesta a las siguientes preguntas:

REVISAR...

> ¿Hay relación directa entre el número de tecnologías que domina el programador y su sueldo anual, en Ecuador? ¿Hay relación directa entre el número de años de experiencia que tiene el programador y el número de tecnologías que domina, en Ecuador? ¿En Ecuador, influye el número de años de experiencia del programador con el número de tecnologías que domina o conoce? ¿Qué relación hay entre el número de años de experiencia del programador y el sueldo que percibe anualmente en el mercado ecuatoriano? ¿Como afecta el número de horas trabajadas a la semana sobre el sueldo que percibe anualmente el programador ecuatoriano? ¿Hay relación directa entre el número de años de experiencia que tiene el programador y la edad del mismo en Ecuador?

```
# Cargamos el dataset completo
data_devs <- read.csv('../data/survey_results_public.csv', sep=",", encoding = "UTF-8")

# Resumen del dataset original
head(data_devs)
```

```
##   Respondent                                                MainBranch Hobbyist
## 1          1                          I am a developer by profession      Yes
## 2          2                          I am a developer by profession       No
## 3          3                             I code primarily as a hobby      Yes
## 4          4                          I am a developer by profession      Yes
## 5          5 I used to be a developer by profession, but no longer am      Yes
## 6          6                          I am a developer by profession       No
##   Age Age1stCode CompFreq CompTotal ConvertedComp            Country
## 1  NA         13  Monthly        NA            NA            Germany
## 2  NA         19     <NA>        NA            NA     United Kingdom
## 3  NA         15     <NA>        NA            NA Russian Federation
## 4  25         18     <NA>        NA            NA            Albania
## 5  31         16     <NA>        NA            NA      United States
## 6  NA         14     <NA>        NA            NA            Germany
##     CurrencyDesc CurrencySymbol DatabaseDesireNextYear
## 1  European Euro            EUR   Microsoft SQL Server
## 2 Pound sterling            GBP                   <NA>
## 3          <NA>           <NA>                   <NA>
## 4   Albanian lek            ALL                   <NA>
## 5          <NA>           <NA>       MySQL;PostgreSQL
## 6  European Euro            EUR                   <NA>
##                          DatabaseWorkedWith
## 1 Elasticsearch;Microsoft SQL Server;Oracle
## 2                                      <NA>
## 3                                      <NA>
## 4                                      <NA>
## 5             MySQL;PostgreSQL;Redis;SQLite
## 6                                      <NA>
##                                                DevType
```

```
## 1 Developer, desktop or enterprise applications;Developer, full-stack
## 2                           Developer, full-stack;Developer, mobile
## 3                                                              <NA>
## 4                                                              <NA>
## 5                                                              <NA>
## 6                     Designer;Developer, front-end;Developer, mobile
##                                                                                  EdLevel
## 1                                    Master's degree (M.A., M.S., M.Eng., MBA, etc.)
## 2                                       Bachelor's degree (B.A., B.S., B.Eng., etc.)
## 3                                                                               <NA>
## 4                                    Master's degree (M.A., M.S., M.Eng., MBA, etc.)
## 5                                       Bachelor's degree (B.A., B.S., B.Eng., etc.)
## 6 Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)
##                                          Employment
## 1 Independent contractor, freelancer, or self-employed
## 2                                 Employed full-time
## 3                                               <NA>
## 4                                               <NA>
## 5                                 Employed full-time
## 6                                 Employed full-time
##                     Ethnicity Gender
## 1 White or of European descent    Man
## 2                         <NA>   <NA>
## 3                         <NA>   <NA>
## 4 White or of European descent    Man
## 5 White or of European descent    Man
## 6 White or of European descent    Man
##
## 1               Languages, frameworks, and other technologies I'd be working with;Remote work optio
## 2
## 3
## 4                                 Flex time or a flexible schedule;Office environment or company cultu
## 5
## 6 Diversity of the company or organization;Languages, frameworks, and other technologies I'd be worki
##                 JobSat
## 1    Slightly satisfied
## 2     Very dissatisfied
## 3                  <NA>
## 4 Slightly dissatisfied
## 5                  <NA>
## 6    Slightly satisfied
##                                                    JobSeek
## 1               I am not interested in new job opportunities
## 2               I am not interested in new job opportunities
## 3                                                      <NA>
## 4 I'm not actively looking, but I am open to new opportunities
## 5                                                      <NA>
## 6               I am not interested in new job opportunities
##     LanguageDesireNextYear       LanguageWorkedWith
## 1   C#;HTML/CSS;JavaScript   C#;HTML/CSS;JavaScript
## 2             Python;Swift         JavaScript;Swift
## 3 Objective-C;Python;Swift Objective-C;Python;Swift
## 4                     <NA>                     <NA>
## 5          Java;Ruby;Scala        HTML/CSS;Ruby;SQL
## 6 HTML/CSS;Java;JavaScript HTML/CSS;Java;JavaScript
##             MiscTechDesireNextYear MiscTechWorkedWith
## 1                .NET Core;Xamarin     .NET;.NET Core
## 2 React Native;TensorFlow;Unity 3D       React Native
```

```
## 3                            <NA>              <NA>
## 4                            <NA>              <NA>
## 5                   Ansible;Chef            Ansible
## 6                            <NA>              <NA>
##            NEWCollabToolsDesireNextYear
## 1 Microsoft Teams;Microsoft Azure;Trello
## 2                          Github;Slack
## 3                                  <NA>
## 4                                  <NA>
## 5  Github;Google Suite (Docs, Meet, etc)
## 6                          Github;Slack
##                                        NEWCollabToolsWorkedWith NEWDevOps
## 1                Confluence;Jira;Slack;Microsoft Azure;Trello        No
## 2                         Confluence;Jira;Github;Gitlab;Slack      <NA>
## 3                                                         <NA>      <NA>
## 4                                                         <NA>        No
## 5 Confluence;Jira;Github;Slack;Google Suite (Docs, Meet, etc)      <NA>
## 6                           Confluence;Github;Slack;Trello  Not sure
##       NEWDevOpsImpt                          NEWEdImpt
## 1 Somewhat important                 Fairly important
## 2              <NA>                 Fairly important
## 3              <NA>                             <NA>
## 4              <NA> Not at all important/not necessary
## 5              <NA>                   Very important
## 6              <NA>                 Fairly important
##                                                                NEWJobHunt
## 1                                                                    <NA>
## 2                                                                    <NA>
## 3                                                                    <NA>
## 4 Curious about other opportunities;Wanting to work with new technologies
## 5                                                                    <NA>
## 6                                                                    <NA>
##   NEWJobHuntResearch      NEWLearn NEWOffTopic NEWOnboardGood NEWOtherComms
## 1              <NA>   Once a year    Not sure           <NA>            No
## 2              <NA>   Once a year    Not sure           <NA>            No
## 3              <NA> Once a decade        <NA>           <NA>            No
## 4              <NA>   Once a year    Not sure            Yes           Yes
## 5              <NA>   Once a year          No           <NA>           Yes
## 6              <NA>   Once a year          No             No            No
##                                        NEWOvertime
## 1                      Often: 1-2 days per week or more
## 2                                              <NA>
## 3                                              <NA>
## 4 Occasionally: 1-2 days per quarter but less than monthly
## 5                                              <NA>
## 6                                             Never
##
## 1
## 2
## 3
## 4
## 5 Start a free trial;Ask developers I know/work with;Visit developer communities like Stack Overflow;
## 6
##       NEWPurpleLink
## 1            Amused
## 2            Amused
## 3              <NA>
## 4              <NA>
```

```
## 5 Hello, old friend
## 6            Amused
##
## 1
## 2
## 3
## 4
## 5 Stack Overflow (public Q&A for anyone who codes);Stack Exchange (public Q&A for a variety of topics
## 6
##
## 1                                                                    Visit Stack Overflow;Go for a walk o
## 2                                                                                                      Vi
## 3
## 4
## 5 Call a coworker or friend;Visit Stack Overflow;Watch help / tutorial videos;Do other work and come
## 6                                                                     Play games;Visit Stack Overflow;Wat
##         OpSys                  OrgSize
## 1     Windows         2 to 9 employees
## 2       MacOS 1,000 to 4,999 employees
## 3 Linux-based                     <NA>
## 4 Linux-based       20 to 99 employees
## 5     Windows                     <NA>
## 6     Windows                     <NA>
##                             PlatformDesireNextYear
## 1    Android;iOS;Kubernetes;Microsoft Azure;Windows
## 2                        iOS;Kubernetes;Linux;MacOS
## 3                                             <NA>
## 4                                             <NA>
## 5 Docker;Google Cloud Platform;Heroku;Linux;Windows
## 6                                          Android
##               PlatformWorkedWith                       PurchaseWhat
## 1                        Windows                               <NA>
## 2                            iOS    I have little or no influence
## 3                           <NA>                               <NA>
## 4                           <NA> I have a great deal of influence
## 5 AWS;Docker;Linux;MacOS;Windows                               <NA>
## 6       Android;Docker;WordPress            I have some influence
##                 Sexuality SOAccount          SOComm
## 1 Straight / Heterosexual        No  No, not at all
## 2                    <NA>       Yes Yes, definitely
## 3                    <NA>       Yes   Yes, somewhat
## 4 Straight / Heterosexual       Yes Yes, definitely
## 5 Straight / Heterosexual       Yes   Yes, somewhat
## 6 Straight / Heterosexual       Yes   Yes, somewhat
##                            SOPartFreq                     SOVisitFreq
## 1                                <NA>          Multiple times per day
## 2 Less than once per month or monthly          Multiple times per day
## 3     A few times per month or weekly           Daily or almost daily
## 4     A few times per month or weekly          Multiple times per day
## 5 Less than once per month or monthly A few times per month or weekly
## 6     A few times per month or weekly            A few times per week
##                   SurveyEase          SurveyLength Trans
## 1 Neither easy nor difficult Appropriate in length    No
## 2                       <NA>                 <NA>  <NA>
## 3 Neither easy nor difficult Appropriate in length  <NA>
## 4                       <NA>                 <NA>    No
## 5                       Easy            Too short    No
## 6 Neither easy nor difficult Appropriate in length  <NA>
```

```
##                                                    UndergradMajor
## 1 Computer science, computer engineering, or software engineering
## 2 Computer science, computer engineering, or software engineering
## 3                                                            <NA>
## 4 Computer science, computer engineering, or software engineering
## 5 Computer science, computer engineering, or software engineering
## 6                                                            <NA>
##   WebframeDesireNextYear   WebframeWorkedWith
## 1           ASP.NET Core ASP.NET;ASP.NET Core
## 2                   <NA>                 <NA>
## 3                   <NA>                 <NA>
## 4                   <NA>                 <NA>
## 5   Django;Ruby on Rails        Ruby on Rails
## 6              React.js                 <NA>
##                              WelcomeChange WorkWeekHrs YearsCode YearsCodePro
## 1  Just as welcome now as I felt last year          50        36           27
## 2 Somewhat more welcome now than last year          NA         7            4
## 3 Somewhat more welcome now than last year          NA         4         <NA>
## 4 Somewhat less welcome now than last year          40         7            4
## 5  Just as welcome now as I felt last year          NA        15            8
## 6                                   <NA>          NA         6            4
```

2 Integración y selección de los datos de interés a analizar

```
# Creamos un juego de datos resumido
data_so <- data_devs[, c(4, 8, 9, 13, 15, 16, 23, 25, 27, 33, 37, 42, 45, 50, 55, 57, 59, 61)]

head(data_so)
```

```
##   Age ConvertedComp            Country
## 1  NA            NA            Germany
## 2  NA            NA     United Kingdom
## 3  NA            NA Russian Federation
## 4  25            NA            Albania
## 5  31            NA      United States
## 6  NA            NA            Germany
##                          DatabaseWorkedWith
## 1 Elasticsearch;Microsoft SQL Server;Oracle
## 2                                      <NA>
## 3                                      <NA>
## 4                                      <NA>
## 5            MySQL;PostgreSQL;Redis;SQLite
## 6                                      <NA>
##                                                                   EdLevel
## 1                                  Master's degree (M.A., M.S., M.Eng., MBA, etc.)
## 2                                     Bachelor's degree (B.A., B.S., B.Eng., etc.)
## 3                                                                     <NA>
## 4                                  Master's degree (M.A., M.S., M.Eng., MBA, etc.)
## 5                                     Bachelor's degree (B.A., B.S., B.Eng., etc.)
## 6 Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)
##                                          Employment       LanguageWorkedWith
## 1 Independent contractor, freelancer, or self-employed   C#;HTML/CSS;JavaScript
## 2                                Employed full-time         JavaScript;Swift
## 3                                              <NA> Objective-C;Python;Swift
## 4                                              <NA>                     <NA>
```

```
## 5                                  Employed full-time        HTML/CSS;Ruby;SQL
## 6                                  Employed full-time HTML/CSS;Java;JavaScript
##   MiscTechWorkedWith
## 1     .NET;.NET Core
## 2       React Native
## 3               <NA>
## 4               <NA>
## 5            Ansible
## 6               <NA>
##                                    NEWCollabToolsWorkedWith      NEWLearn
## 1               Confluence;Jira;Slack;Microsoft Azure;Trello   Once a year
## 2                       Confluence;Jira;Github;Gitlab;Slack   Once a year
## 3                                                     <NA> Once a decade
## 4                                                     <NA>   Once a year
## 5 Confluence;Jira;Github;Slack;Google Suite (Docs, Meet, etc)   Once a year
## 6                             Confluence;Github;Slack;Trello   Once a year
##                                           NEWOvertime       OpSys
## 1                        Often: 1-2 days per week or more     Windows
## 2                                                  <NA>       MacOS
## 3                                                  <NA> Linux-based
## 4 Occasionally: 1-2 days per quarter but less than monthly Linux-based
## 5                                                  <NA>     Windows
## 6                                                 Never     Windows
##               PlatformWorkedWith                          SOPartFreq
## 1                        Windows                                <NA>
## 2                            iOS Less than once per month or monthly
## 3                           <NA>     A few times per month or weekly
## 4                           <NA>     A few times per month or weekly
## 5 AWS;Docker;Linux;MacOS;Windows Less than once per month or monthly
## 6       Android;Docker;WordPress     A few times per month or weekly
##                                                    UndergradMajor
## 1 Computer science, computer engineering, or software engineering
## 2 Computer science, computer engineering, or software engineering
## 3                                                            <NA>
## 4 Computer science, computer engineering, or software engineering
## 5 Computer science, computer engineering, or software engineering
## 6                                                            <NA>
##     WebframeWorkedWith WorkWeekHrs YearsCodePro
## 1 ASP.NET;ASP.NET Core          50           27
## 2                 <NA>          NA            4
## 3                 <NA>          NA         <NA>
## 4                 <NA>          40            4
## 5        Ruby on Rails          NA            8
## 6                 <NA>          NA            4
```

3 Limpieza de los datos

3.1 Tratamiento de valores nulos o ceros

Se ha revisado que existen valores nulos en todas las variables seleccionadas, por lo que se procederá a filtrar los registros que no contengan valores *NA* en ninguna de las variables.

Además, en este apartado se ha decidido filtrar los valores que no representan una realidad en el estudio de la variable en cuestión. Por ejemplo un valor de 0 en el sueldo anual (*variable ConvertedComp*) de un desarrollador no representa la realidad. A pesar de ser un valor que puede ser tratado como un valor extremo, para esta ocasión se lo tratará como no válido.

```r
# Eliminar filas con valores nulos
data_wo_na <- data_so[!is.na(data_so$Age), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$ConvertedComp), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$Country), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$DatabaseWorkedWith), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$EdLevel), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$Employment), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$LanguageWorkedWith), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$MiscTechWorkedWith), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$NEWCollabToolsWorkedWith), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$NEWLearn), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$NEWOvertime), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$OpSys), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$PlatformWorkedWith), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$SOPartFreq), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$UndergradMajor), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$WebframeWorkedWith), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$WorkWeekHrs), ]
data_wo_na <- data_wo_na[!is.na(data_wo_na$YearsCodePro), ]

# Campos con valores cero
data_wo_na <- data_wo_na[data_wo_na$ConvertedComp != 0, ]

head(data_wo_na)
```

```
##    Age ConvertedComp        Country                DatabaseWorkedWith
## 8   36        116000  United States               MariaDB;MySQL;Redis
## 10  22         32315 United Kingdom             Microsoft SQL Server
## 11  23         40070 United Kingdom Firebase;MongoDB;PostgreSQL;SQLite
## 14  27         66000  United States      Firebase;Microsoft SQL Server
## 19  24         83400  United States       MariaDB;Microsoft SQL Server
## 33  39         63564        Belgium    MongoDB;MySQL;PostgreSQL;SQLite
##                                          EdLevel         Employment
## 8     Bachelor's degree (B.A., B.S., B.Eng., etc.) Employed full-time
## 10 Master's degree (M.A., M.S., M.Eng., MBA, etc.) Employed full-time
## 11    Bachelor's degree (B.A., B.S., B.Eng., etc.) Employed full-time
## 14             Associate degree (A.A., A.S., etc.) Employed full-time
## 19    Bachelor's degree (B.A., B.S., B.Eng., etc.) Employed full-time
## 33    Bachelor's degree (B.A., B.S., B.Eng., etc.) Employed full-time
##                                           LanguageWorkedWith
## 8                                                 Python;SQL
## 10                         HTML/CSS;Java;JavaScript;Python;SQL
## 11                                      C#;JavaScript;Swift
## 14                          HTML/CSS;JavaScript;SQL;TypeScript
## 19 Bash/Shell/PowerShell;C#;HTML/CSS;JavaScript;SQL;TypeScript
## 33                             HTML/CSS;JavaScript;TypeScript
##        MiscTechWorkedWith
## 8                 Ansible
## 10                 Pandas
## 11                Node.js
## 14                Node.js
## 19 .NET;.NET Core;Node.js
## 33   Node.js;React Native
##
## 8                                                 Confluence;Jira;Github;Slack;Microsoft Teams;Googl
## 10
## 11 Confluence;Jira;Github;Slack;Microsoft Teams;Microsoft Azure;Trello;Google Suite (Docs, Meet, etc)
```

```
## 14                                                                                     Confluence;Ji
## 19                                                                                              Con
## 33                                 Confluence;Jira;Github;Gitlab;Slack;Google Suite (Docs, Meet, etc)
##            NEWLearn                                              NEWOvertime
## 8       Once a year Occasionally: 1-2 days per quarter but less than monthly
## 10 Every few months                        Often: 1-2 days per week or more
## 11 Every few months                       Rarely: 1-2 days per year or less
## 14 Every few months Occasionally: 1-2 days per quarter but less than monthly
## 19      Once a year Occasionally: 1-2 days per quarter but less than monthly
## 33      Once a year       Sometimes: 1-2 days per month but less than weekly
##          OpSys
## 8  Linux-based
## 10     Windows
## 11     Windows
## 14     Windows
## 19     Windows
## 33       MacOS
##                                                        PlatformWorkedWith
## 8                                                                  Docker
## 10                                 Android;Linux;Raspberry Pi;Windows
## 11                                                         AWS;Heroku;iOS
## 14                                          Google Cloud Platform;Windows
## 19                                                                Windows
## 33 AWS;Docker;Google Cloud Platform;Heroku;iOS;Kubernetes;MacOS;Raspberry Pi
##                                                  SOPartFreq
## 8                  Less than once per month or monthly
## 10                              Multiple times per day
## 11 I have never participated in Q&A on Stack Overflow
## 14                                A few times per week
## 19                 Less than once per month or monthly
## 33                 Less than once per month or monthly
##                                                       UndergradMajor
## 8  Computer science, computer engineering, or software engineering
## 10                                       Mathematics or statistics
## 11 Computer science, computer engineering, or software engineering
## 14 Computer science, computer engineering, or software engineering
## 19 Computer science, computer engineering, or software engineering
## 33 Computer science, computer engineering, or software engineering
##                          WebframeWorkedWith WorkWeekHrs YearsCodePro
## 8                                     Flask          39           13
## 10                             Flask;jQuery          36            4
## 11         Angular;Angular.js;Django;React.js          40            2
## 14                             Angular;Vue.js          40            1
## 19           Angular;Angular.js;ASP.NET Core          35            3
## 33 Angular;Angular.js;Django;Express;React.js          40           14
```

Analizamos el tipo de dato de las columnas antes mencionadas

```r
library(dplyr)
#Vemos el tipo de dato de las variables
glimpse(data_wo_na)
```

```
## Rows: 12,973
## Columns: 18
## $ Age                     <dbl> 36, 22, 23, 27, 24, 39, 34, 35, 32, 22, 53...
## $ ConvertedComp           <dbl> 116000, 32315, 40070, 66000, 83400, 63564,...
## $ Country                 <chr> "United States", "United Kingdom", "United...
```

```
## $ DatabaseWorkedWith       <chr> "MariaDB;MySQL;Redis", "Microsoft SQL Serv...
## $ EdLevel                  <chr> "Bachelor's degree (B.A., B.S., B.Eng., et...
## $ Employment               <chr> "Employed full-time", "Employed full-time"...
## $ LanguageWorkedWith       <chr> "Python;SQL", "HTML/CSS;Java;JavaScript;Py...
## $ MiscTechWorkedWith       <chr> "Ansible", "Pandas", "Node.js", "Node.js",...
## $ NEWCollabToolsWorkedWith <chr> "Confluence;Jira;Github;Slack;Microsoft Te...
## $ NEWLearn                 <chr> "Once a year", "Every few months", "Every ...
## $ NEWOvertime              <chr> "Occasionally: 1-2 days per quarter but le...
## $ OpSys                    <chr> "Linux-based", "Windows", "Windows", "Wind...
## $ PlatformWorkedWith       <chr> "Docker", "Android;Linux;Raspberry Pi;Wind...
## $ SOPartFreq               <chr> "Less than once per month or monthly", "Mu...
## $ UndergradMajor           <chr> "Computer science, computer engineering, o...
## $ WebframeWorkedWith       <chr> "Flask", "Flask;jQuery", "Angular;Angular....
## $ WorkWeekHrs              <dbl> 39, 36, 40, 40, 35, 40, 40, 40, 37, 35, 40...
## $ YearsCodePro             <chr> "13", "4", "2", "1", "3", "14", "3", "12",...
```

```
# Otra forma de ver el tipo de dato de cada columna
sapply(data_wo_na, class)
```

```
##                       Age             ConvertedComp                   Country
##                 "numeric"                 "numeric"               "character"
##       DatabaseWorkedWith                   EdLevel                Employment
##               "character"               "character"               "character"
##       LanguageWorkedWith        MiscTechWorkedWith NEWCollabToolsWorkedWith
##               "character"               "character"               "character"
##                  NEWLearn               NEWOvertime                     OpSys
##               "character"               "character"               "character"
##       PlatformWorkedWith                SOPartFreq           UndergradMajor
##               "character"               "character"               "character"
##       WebframeWorkedWith               WorkWeekHrs              YearsCodePro
##               "character"                 "numeric"               "character"
```

Para la variable YearsCodePro observamos que el tipo de datos es "character", por lo que contiene valores cualitativos: *More than 50 years* y *Less than 1 year*, por lo que vamos a transformarlos en valores numéricos o cuantitativos

```
# Convertimos los valores categoricos en numéricos para la variable YearsCodePro
data_temp <- data_wo_na[data_wo_na$YearsCodePro %in% c("More than 50 years", "Less than 1 year"), ]

dim(data_temp)
```

```
## [1] 375  18
```

```
# datosEcuador$YearsCode[datosEcuador$YearsCode=="Less than 1 year"] <- 1
```

Existen 372 valores con valores cualitativos que deben ser transformados a valores cuantitativos

```
# Convertimos los valores categoricos en numéricos para la variable YearsCodePro
data_wo_na$YearsCodePro[data_wo_na$YearsCodePro=="Less than 1 year"] <- 1
data_wo_na$YearsCodePro[data_wo_na$YearsCodePro=="More than 50 years"] <- 50


# Finalmente convertimos dicha columna en número ica
data_wo_na$YearsCodePro <- as.numeric(data_wo_na$YearsCodePro)

filas_df=dim(data_wo_na)[1]
```

Verificamos nuevamente el tipo de dato de la columna YearsCodePro para validar que se convirtió correctamente de *character* a *numeric*

```
# Otra forma de ver el tipo de dato de cada columna
sapply(data_wo_na, class)
```

```
##                       Age             ConvertedComp                   Country
##                 "numeric"                 "numeric"               "character"
##       DatabaseWorkedWith                   EdLevel                Employment
##                "character"               "character"               "character"
##       LanguageWorkedWith        MiscTechWorkedWith NEWCollabToolsWorkedWith
##                "character"               "character"               "character"
##                  NEWLearn               NEWOvertime                     OpSys
##                "character"               "character"               "character"
##       PlatformWorkedWith                SOPartFreq           UndergradMajor
##                "character"               "character"               "character"
##       WebframeWorkedWith               WorkWeekHrs             YearsCodePro
##                "character"                 "numeric"                 "numeric"
```

Ahora vamos a agregar nuevas variables que contabilizan el número de tecnologías o herramientas de: bases de datos, lenguajes de programación, de colaboración, entre otros. Primero para la base de datos, vamos a usar la columna *DatabaseWorkedWith*. La variable a crearse será **db_techs**:

```
data_wo_na$db_techs <- 0

for(i in 1:filas_df) {
  if (is.na(data_wo_na$DatabaseWorkedWith[i])) {
    data_wo_na$db_techs[i] <- 0
  } else {
    longitud <- sapply(strsplit(data_wo_na$DatabaseWorkedWith[i], ";"), length)
    data_wo_na$db_techs[i] <- longitud
  }
}
```

Para agregar una nueva variable que represente el número de lenguajes de programación que usa. Este dato se basa en la experiencia ya adquirida y no en los deseos para usar o aprender el siguiente año. Para esto usaremos la columna *LanguageWorkedWith*. La variable a crearse será **prog_langs**:

```
data_wo_na$prog_langs <- 0

for(i in 1:filas_df) {
  if (is.na(data_wo_na$LanguageWorkedWith[i])) {
    data_wo_na$prog_langs[i] <- 0
  } else {
    longitud  <- sapply(strsplit(data_wo_na$LanguageWorkedWith[i], ";"), length)
    data_wo_na$prog_langs[i] <- longitud
  }
}
```

Ahora vamos a agregar una nueva variable para el número de frameworks, librerías y demás herramientras que usa el desarrollador. Este dato se basa en la experiencia ya adquirida y no en los deseos para usar o aprender el siguiente año. Para esto usaremos la columna *MiscTechWorkedWith*. La variable a crearse será **misc_techs**:

```
data_wo_na$misc_techs <- 0

for(i in 1:filas_df) {
  if (is.na(data_wo_na$MiscTechWorkedWith[i])) {
    data_wo_na$misc_techs[i] <- 0
  } else {
    longitud  <- sapply(strsplit(data_wo_na$MiscTechWorkedWith[i], ";"), length)
    data_wo_na$misc_techs[i] <- longitud
  }
}
```

Haremos lo mismo para el número de herramientras colaborativas que usa el desarrollador, según el contenido de la columna *NEWCollabToolsWorkedWith*. Este dato se basa en la experiencia ya adquirida y no en los deseos para usar o aprender el siguiente año. La variable a crearse será **collab_techs**:

```
data_wo_na$collab_techs <- 0

for(i in 1:filas_df) {
  if (is.na(data_wo_na$NEWCollabToolsWorkedWith[i])) {
    data_wo_na$collab_techs[i] <- 0
  } else {
    longitud  <- sapply(strsplit(data_wo_na$NEWCollabToolsWorkedWith[i], ";"), length)
    data_wo_na$collab_techs[i] <- longitud
  }
}
```

También vamos a agregar una variable para el número de plataformas que usa el desarrollador. Este dato se basa en la experiencia ya adquirida y no en los deseos para usar o aprender el siguiente año. Usaremos el contenido de la columna *PlatformWorkedWith*. La variable a crearse será **plat_techs**:

```
data_wo_na$plat_techs <- 0

for(i in 1:filas_df) {
  if (is.na(data_wo_na$PlatformWorkedWith[i])) {
    data_wo_na$plat_techs[i] <- 0
  } else {
    longitud  <- sapply(strsplit(data_wo_na$PlatformWorkedWith[i], ";"), length)
    data_wo_na$plat_techs[i] <- longitud
  }
}
```

Finalmente, agregaremos una variable para el número de *frameworks web* que usa el desarrollador. Este dato se basa en la experiencia ya adquirida y no en los deseos para usar o aprender el siguiente año. Para esto usaremos la columna *WebframeWorkedWith*. La variable a crearse será **web_techs**:

```
data_wo_na$web_techs <- 0

for(i in 1:filas_df) {
  if (is.na(data_wo_na$WebframeWorkedWith[i])) {
    data_wo_na$web_techs[i] <- 0
  } else {
    longitud  <- sapply(strsplit(data_wo_na$WebframeWorkedWith[i], ";"), length)
    data_wo_na$web_techs[i] <- longitud
  }
}
```

Eliminamos las variables cualitativas previamente tratadas

```
data_wo_na$DatabaseWorkedWith = NULL
data_wo_na$LanguageWorkedWith = NULL
data_wo_na$MiscTechWorkedWith = NULL
data_wo_na$NEWCollabToolsWorkedWith = NULL
data_wo_na$PlatformWorkedWith = NULL
data_wo_na$WebframeWorkedWith = NULL
```

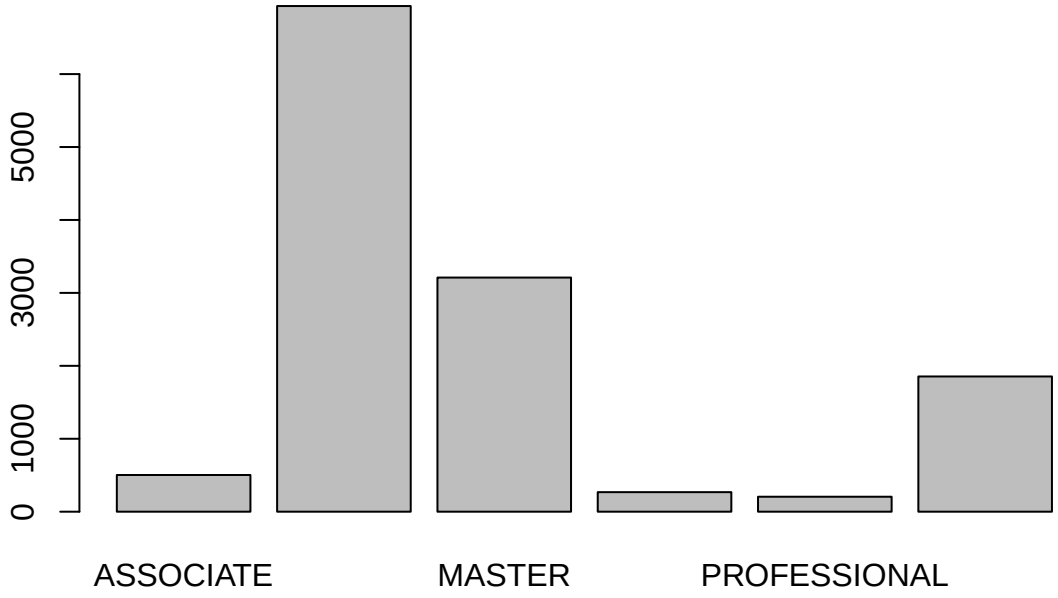
La variable *EdLevel* tiene valores como cadenas de texto muy extensas que dificultan su legibilidad, para lo cual vamos a realizar una reasignación de valores, con nomenclaturas más cortas, para lo cual procederemos de la siguiente manera:

- "I never completed any formal education" -> "NEVER"
- "Primary/elementary school" -> "PRIMARY"
- "Secondary school (e.g. American high school, German Realschule or Gymnasium, etc.)" -> "SECONDARY"
- "Some college/university study without earning a degree" -> "SOME_STUDY_WITHOUT_DEGREE"
- "Associate degree (A.A., A.S., etc.)" -> "ASSOCIATE"
- "Bachelor's degree (B.A., B.S., B.Eng., etc.)" -> "BACHELOR"
- "Master's degree (M.A., M.S., M.Eng., MBA, etc.)" -> "MASTER"
- "Professional degree (JD, MD, etc.)" -> "PROFESSIONAL"
- "Other doctoral degree (Ph.D., Ed.D., etc.)" -> "OTHER_PHD"

```
# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="I never completed any formal education"] <- 'NEVER'

# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="Primary/elementary school"] <- 'PRIMARY'

# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="Secondary school (e.g. American high school, German Realschule o

# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="Some college/university study without earning a degree"] <- 'SOM

# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="Associate degree (A.A., A.S., etc.)"] <- 'ASSOCIATE'

# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="Bachelor's degree (B.A., B.S., B.Eng., etc.)"] <- 'BACHELOR'

# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="Master's degree (M.A., M.S., M.Eng., MBA, etc.)"] <- 'MASTER'

# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="Professional degree (JD, MD, etc.)"] <- 'PROFESSIONAL'

# Reasigamos el valor
data_wo_na$EdLevel[data_wo_na$EdLevel=="Other doctoral degree (Ph.D., Ed.D., etc.)"] <- 'OTHER_PHD'


table(as.factor(data_wo_na$EdLevel))
```

```
##
##                 ASSOCIATE                  BACHELOR                    MASTER
##                       503                      6933                      3210
##                 OTHER_PHD             PROFESSIONAL SOME_STUDY_WITHOUT_DEGREE
##                       268                       205                      1854
```

```r
plot(as.factor(data_wo_na$EdLevel))
```



De la misma manera procedemos con la variable *Employment*, para lo cual vamos a realizar una reasignación de valores, con nomenclaturas más cortas, para lo cual procederemos de la siguiente manera:

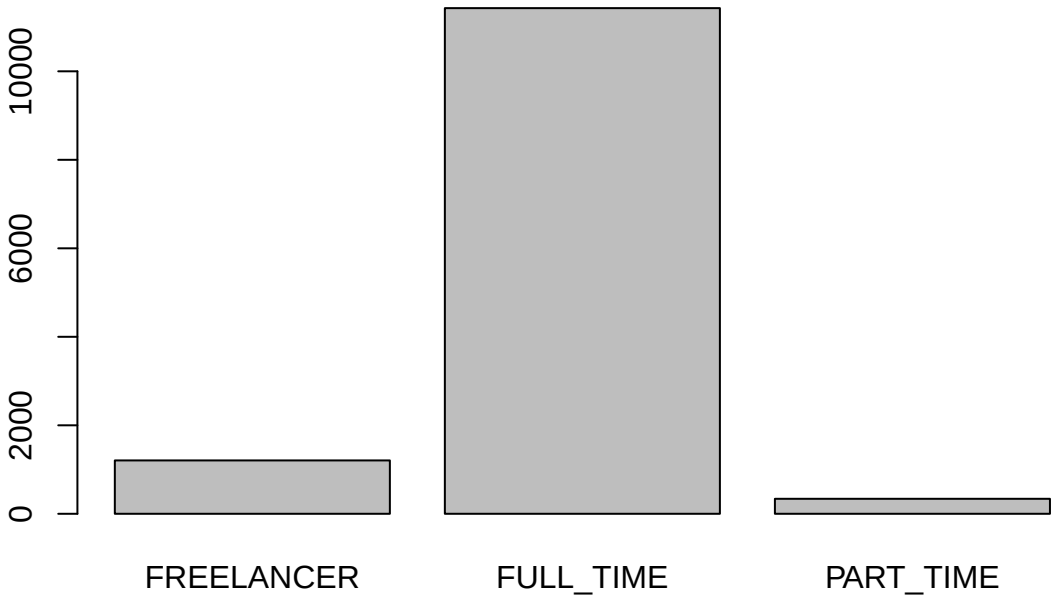
- "Employed full-time" -> "FULL_TIME"
- "Employed part-time" -> "PART_TIME"
- "Independent contractor, freelancer, or self-employed" -> "FREELANCER"
- "Not employed, but looking for work" -> "NOT_EMPLOYED_LOOKING_FOR"
- "Not employed, and not looking for work" -> "NOT_EMPLOYED_NOT_LOOKING_FOR"
- "Student" -> "STUDENT"
- "Retired" -> "RETIRED"

```r
# Reasigamos el valor
data_wo_na$Employment[data_wo_na$Employment=="Employed full-time"] <- 'FULL_TIME'

# Reasigamos el valor
data_wo_na$Employment[data_wo_na$Employment=="Employed part-time"] <- 'PART_TIME'

# Reasigamos el valor
data_wo_na$Employment[data_wo_na$Employment=="Independent contractor, freelancer, or self-employed"] <-

# Reasigamos el valor
data_wo_na$Employment[data_wo_na$Employment=="Not employed, but looking for work"] <- 'NOT_EMPLOYED_LOOK

# Reasigamos el valor
data_wo_na$Employment[data_wo_na$Employment=="Not employed, and not looking for work"] <- 'NOT_EMPLOYED_

# Reasigamos el valor
```

```r
data_wo_na$Employment[data_wo_na$Employment=="Student"] <- 'STUDENT'

# Reasigamos el valor
data_wo_na$Employment[data_wo_na$Employment=="Retired"] <- 'RETIRED'

table(as.factor(data_wo_na$Employment))
```

```
##
## FREELANCER  FULL_TIME  PART_TIME
##       1206      11428        339
```

```r
plot(as.factor(data_wo_na$Employment))
```



Para la variable *NEWOvertime* vamos a realizar una reasignación de valores, con valores más cortos, para lo cual procederemos de la siguiente manera:

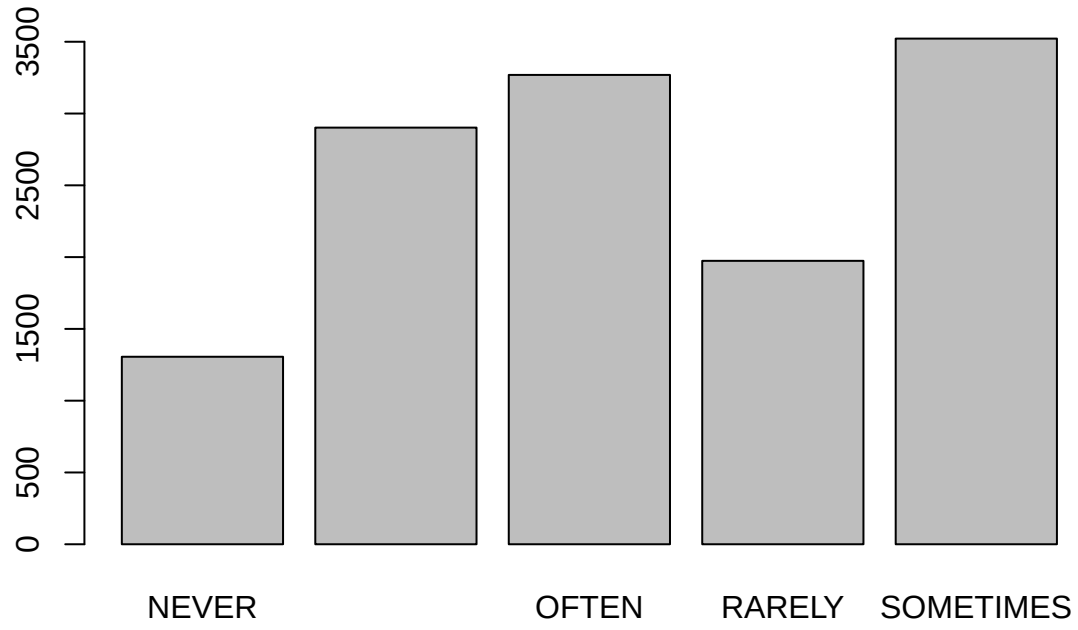
- "Never" -> "NEVER"
- "Occasionally: 1-2 days per quarter but less than monthly" -> "OCCASIONALLY"
- "Often: 1-2 days per week or more" -> "OFTEN"
- "Rarely: 1-2 days per year or less" -> "RARELY"
- "Sometimes: 1-2 days per month but less than weekly" -> "SOMETIMES"

```r
# Reasigamos el valor
data_wo_na$NEWOvertime[data_wo_na$NEWOvertime=="Never"] <- 'NEVER'

# Reasigamos el valor
data_wo_na$NEWOvertime[data_wo_na$NEWOvertime=="Occasionally: 1-2 days per quarter but less than monthly
```

```r
# Reasigamos el valor
data_wo_na$NEWOvertime[data_wo_na$NEWOvertime=="Often: 1-2 days per week or more"] <- 'OFTEN'

# Reasigamos el valor
data_wo_na$NEWOvertime[data_wo_na$NEWOvertime=="Rarely: 1-2 days per year or less"] <- 'RARELY'

# Reasigamos el valor
data_wo_na$NEWOvertime[data_wo_na$NEWOvertime=="Sometimes: 1-2 days per month but less than weekly"] <-

table(as.factor(data_wo_na$NEWOvertime))
```

```
##
##         NEVER OCCASIONALLY         OFTEN       RARELY    SOMETIMES
##          1306          2902          3269          1974          3522
```

```r
plot(as.factor(data_wo_na$NEWOvertime))
```



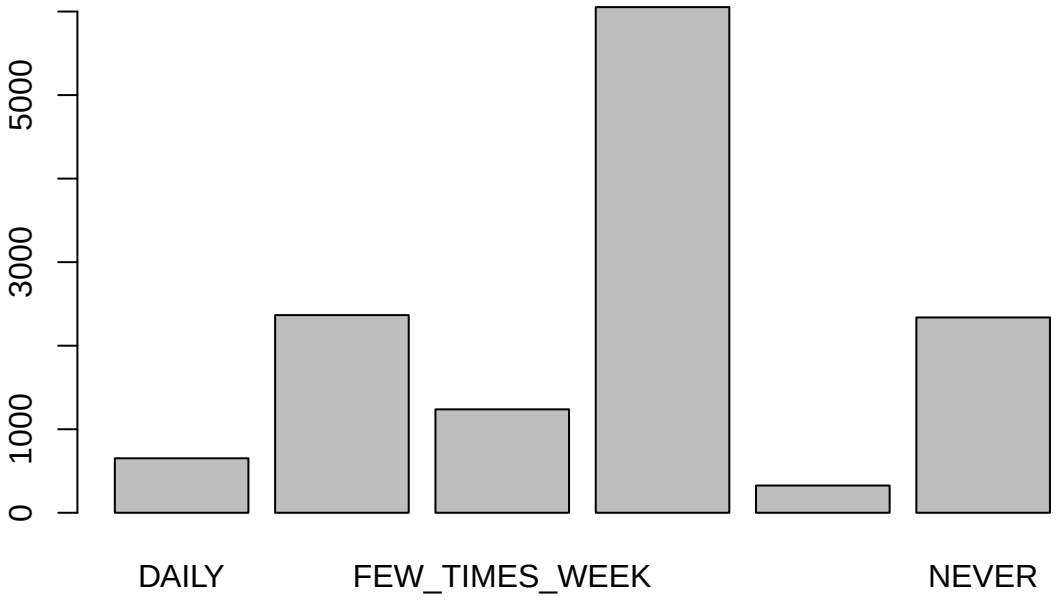
Para la variable *SOPartFreq* vamos a realizar una reasignación de valores, con valores más cortos, para lo cual procederemos de la siguiente manera:

- "Less than once per month or monthly" -> "LESS_ONCE_MONTH"
- "A few times per month or weekly" -> "FEW_TIMES_MONTH"
- "Multiple times per day" -> "MULTIPLE_TIMES_DAY"
- "I have never participated in Q&A on Stack Overflow" -> "NEVER"
- "A few times per week" -> "FEW_TIMES_WEEK"
- "Daily or almost daily" -> "DAILY"

```
# Reasigamos el valor
data_wo_na$SOPartFreq[data_wo_na$SOPartFreq=="Less than once per month or monthly"] <- 'LESS_ONCE_MONTH'

# Reasigamos el valor
data_wo_na$SOPartFreq[data_wo_na$SOPartFreq=="A few times per month or weekly"] <- 'FEW_TIMES_MONTH'

# Reasigamos el valor
data_wo_na$SOPartFreq[data_wo_na$SOPartFreq=="Multiple times per day"] <- 'MULTIPLE_TIMES_DAY'

# Reasigamos el valor
data_wo_na$SOPartFreq[data_wo_na$SOPartFreq=="I have never participated in Q&A on Stack Overflow"] <- 'N

# Reasigamos el valor
data_wo_na$SOPartFreq[data_wo_na$SOPartFreq=="A few times per week"] <- 'FEW_TIMES_WEEK'

# Reasigamos el valor
data_wo_na$SOPartFreq[data_wo_na$SOPartFreq=="Daily or almost daily"] <- 'DAILY'

table(as.factor(data_wo_na$SOPartFreq))
```

```
##
##              DAILY    FEW_TIMES_MONTH     FEW_TIMES_WEEK    LESS_ONCE_MONTH
##                652               2366               1238               6053
## MULTIPLE_TIMES_DAY              NEVER
##                326               2338
```

```
plot(as.factor(data_wo_na$SOPartFreq))
```



Para la variable *UndergradMajor* vamos a realizar una reasignación de valores, con valores más cortos, para lo cual procederemos de la siguiente manera:

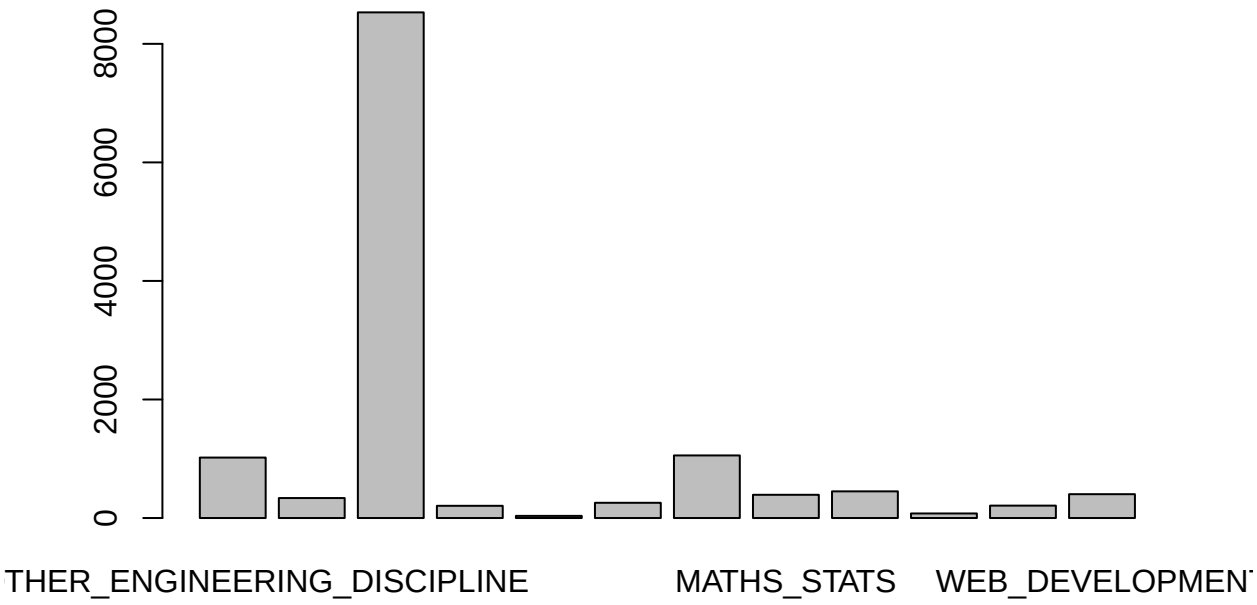
- “Computer science, computer engineering, or software engineering” -> “COMPUTER_SCIENCE”
- “Web development or web design” -> “WEB_DEVELOPMENT”
- “Information systems, information technology, or system administration” -> “INFORMATION_SYSTEMS”
- “Mathematics or statistics” -> “MATHS_STATS”
- “Another engineering discipline (such as civil, electrical, mechanical, etc.)” -> “ANOTHER_ENGINEERING_DISCIPL
- “A business discipline (such as accounting, finance, marketing, etc.)” -> “BUSINESS”
- “A health science (such as nursing, pharmacy, radiology, etc.)” -> “HEALTH”
- “A humanities discipline (such as literature, history, philosophy, etc.)” -> “HUMANITIES”
- “A natural science (such as biology, chemistry, physics, etc.)” -> “NATURAL_SCIENCE”
- “A social science (such as anthropology, psychology, political science, etc.)” -> “SOCIAL_SCIENCE”
- “Fine arts or performing arts (such as graphic design, music, studio art, etc.)” -> “FINE_ARTS”
- “I never declared a major” -> “NEVER_MAJOR”

```
# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="Computer science, computer engineering, or softwar

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="Web development or web design"] <- 'WEB_DEVELOPMEN

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="Information systems, information technology, or sy

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="Mathematics or statistics"] <- 'MATHS_STATS'

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="Another engineering discipline (such as civil, ele

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="A business discipline (such as accounting, finance

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="A health science (such as nursing, pharmacy, radio

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="A humanities discipline (such as literature, histo

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="A natural science (such as biology, chemistry, phy

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="A social science (such as anthropology, psychology

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="Fine arts or performing arts (such as graphic desi

# Reasigamos el valor
data_wo_na$UndergradMajor[data_wo_na$UndergradMajor=="I never declared a major"] <- 'NEVER_MAJOR'


table(as.factor(data_wo_na$UndergradMajor))
```

```
##
## ANOTHER_ENGINEERING_DISCIPLINE                      BUSINESS
##                           1020                           337
##               COMPUTER_SCIENCE                    FINE_ARTS
##                           8531                           206
```

```
##                       HEALTH                   HUMANITIES
##                           36                          257
##        INFORMATION_SYSTEMS                  MATHS_STATS
##                         1056                          392
##            NATURAL_SCIENCE                  NEVER_MAJOR
##                          450                           77
##             SOCIAL_SCIENCE                WEB_DEVELOPMENT
##                          209                          402
```

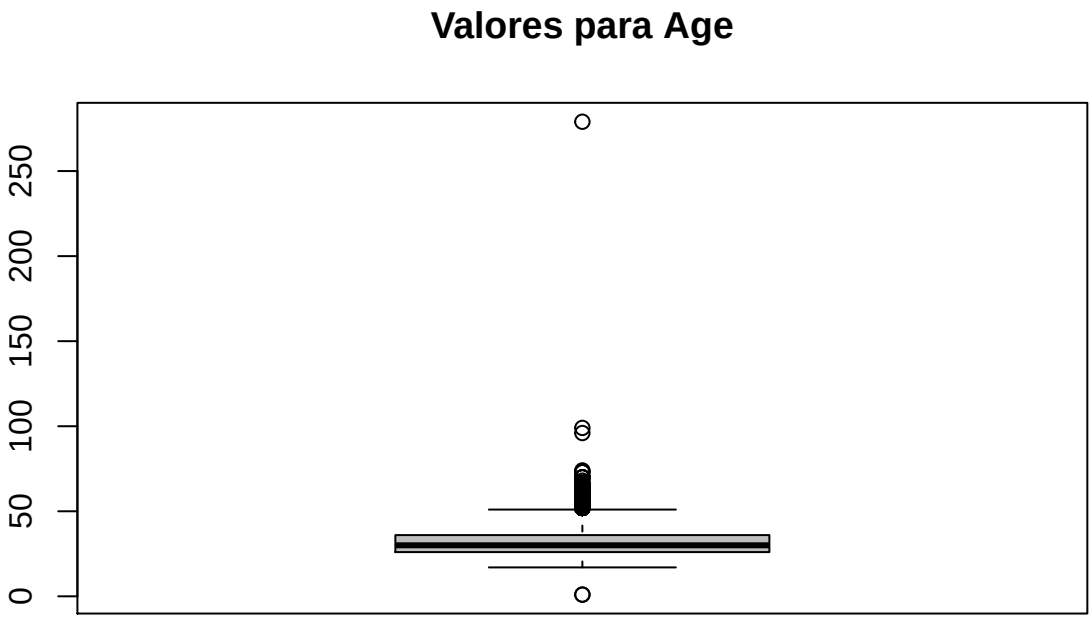
```r
plot(as.factor(data_wo_na$UndergradMajor))
```



3.2 Identificación y tratamiento de valores extremos.

- Para la variable **Age**

```r
boxplot(as.numeric(data_wo_na$Age), main="Valores para Age", col="gray")
```



Para la variable Age los valores extremos los obtenenmos de la siguiente manera:

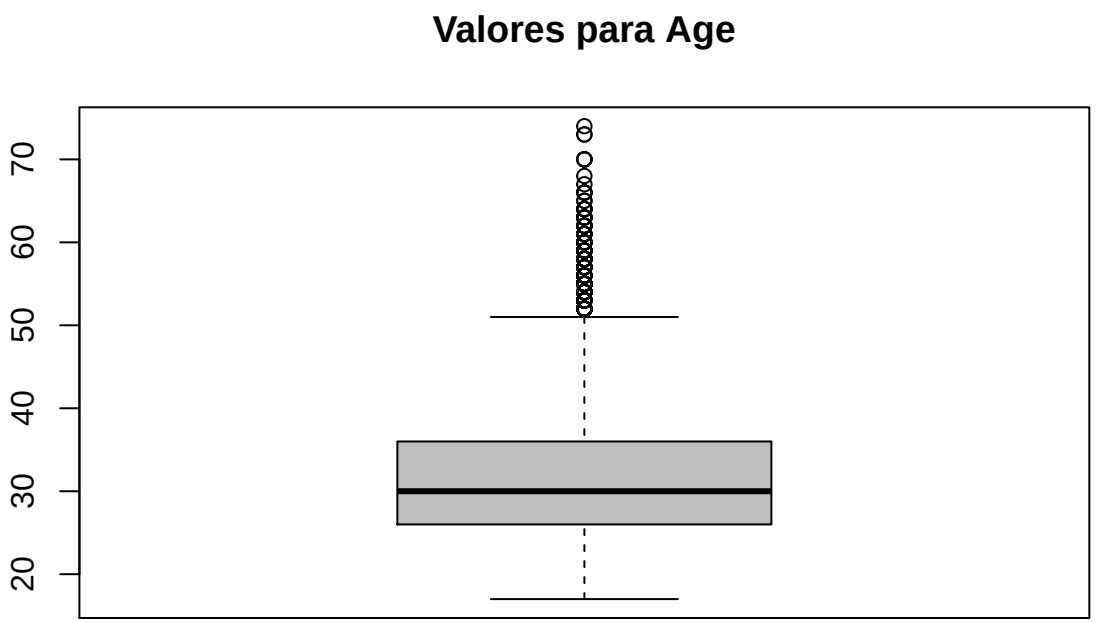
```
atipicos_age <- boxplot.stats(as.numeric(data_wo_na$Age))

sort(atipicos_age$out, decreasing = TRUE)
```

```
##   [1] 279  99  96  74  73  73  70  70  70  70  70  68  67  66  66  66  66  66
##  [19]  65  65  64  64  64  64  64  63  63  63  63  63  63  63  63  63  62  62
##  [37]  62  62  62  62  62  62  62  61  61  61  61  61  60  60  60  60  60  60
##  [55]  60  60  60  60  60  60  60  60  60  60  59  59  59  59  59  59  59  59
##  [73]  59  59  59  59  59  59  59  59  59  59  59  59  59  59  58  58  58  58  58
##  [91]  58  58  58  58  58  58  58  58  58  58  58  58  57  57  57  57  57  57  57
## [109]  57  57  57  57  57  57  57  57  57  57  57  57  56  56  56  56  56  56  56
## [127]  56  56  56  56  56  56  56  56  56  56  56  56  56  56  56  56  56  56  56
## [145]  56  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55
## [163]  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55  55
## [181]  55  55  55  55  54  54  54  54  54  54  54  54  54  54  54  54  54  54  54
## [199]  54  54  54  54  54  54  54  54  54  54  54  54  54  54  54  53  53  53  53
## [217]  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53
## [235]  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53  53
## [253]  53  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52
## [271]  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52
## [289]  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52  52
## [307]  52  52  52  52  52   1   1
```

Al ordenar los valores y analizarlos encontramos que los valores máximos que podrían considerarse outliers son los mayores o iguales a 96 y los menores o iguales a 1.

Por lo tanto, procedemos a borrar esas observaciones que los contienen

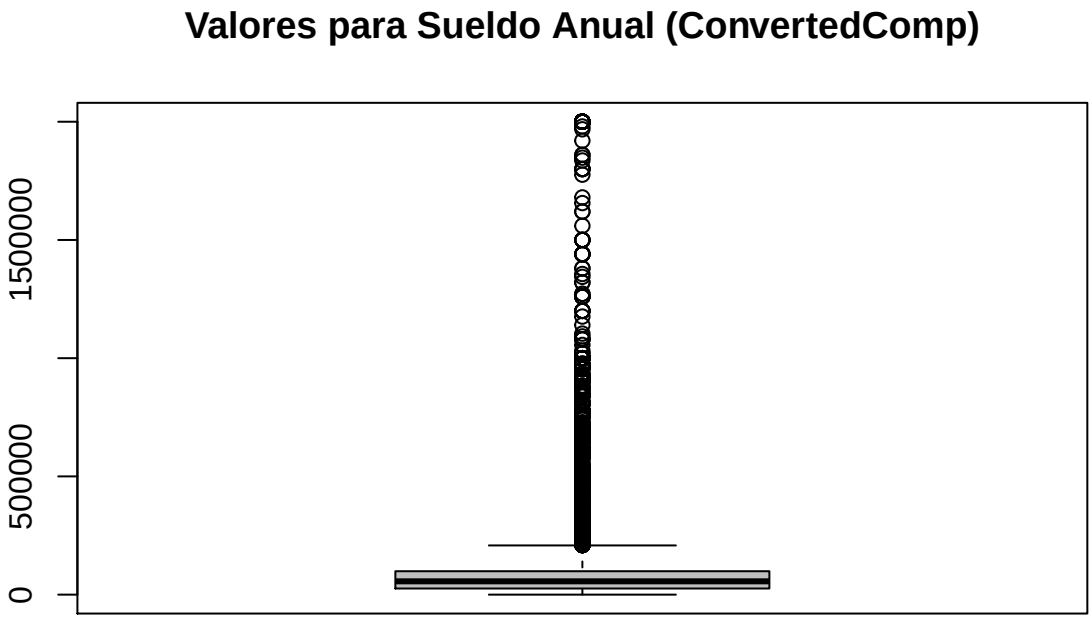
```r
data_wo_na <- data_wo_na[data_wo_na$Age < 96, ]
data_wo_na <- data_wo_na[data_wo_na$Age > 1, ]

# volvemos a revisar los valores outliers
boxplot(as.numeric(data_wo_na$Age), main="Valores para Age", col="gray")
```



A pesar de encontrarse unos valores superiores, por fuera del tercer cuartil (75%) aún se consideran válidos.

- Para la variable **ConvertedComp**

```r
boxplot(as.numeric(data_wo_na$ConvertedComp), main="Valores para Sueldo Anual (ConvertedComp)", col="gra
```



Para la variable Age los valores extremos los obtenenmos de la siguiente manera:

```
atipicos_sueldo_anual <- boxplot.stats(as.numeric(data_wo_na$ConvertedComp))

sort(atipicos_sueldo_anual$out, decreasing = TRUE)
```

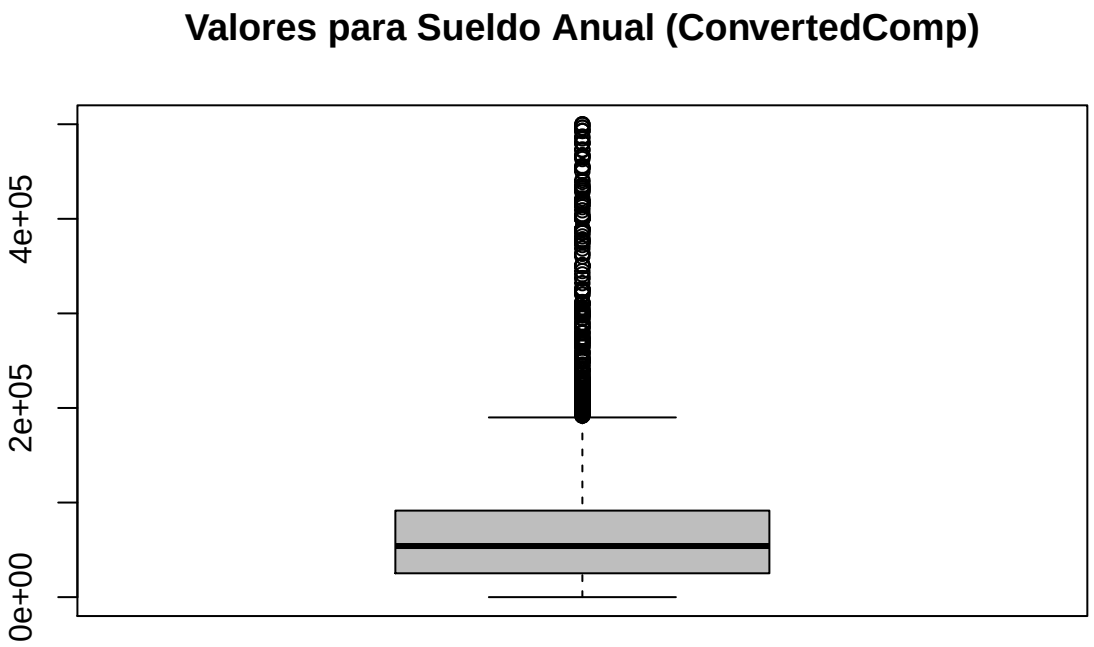
```
##   [1] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [10] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [19] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [28] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [37] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [46] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [55] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [64] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [73] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [82] 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000 2000000
##  [91] 2000000 2000000 2000000 2000000 2000000 2000000 1980000 1968000 1920000
## [100] 1860000 1860000 1848000 1836000 1800000 1800000 1800000 1800000 1800000
## [109] 1776000 1680000 1656000 1620000 1620000 1560000 1500000 1500000 1500000
## [118] 1500000 1500000 1500000 1440000 1440000 1440000 1440000 1440000 1440000
## [127] 1440000 1380000 1380000 1356000 1344000 1320000 1320000 1272000 1272000
## [136] 1260000 1260000 1260000 1260000 1260000 1200000 1200000 1200000 1200000
## [145] 1200000 1176000 1140000 1104000 1092000 1080000 1080000 1080000 1080000
## [154] 1056000 1032000 1020000 1020000 1000000 1000000 1000000 1000000 1000000
## [163] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
## [172] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
## [181] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
## [190] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
## [199] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
## [208] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
```

```
## [217] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
## [226] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
## [235] 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000
## [244] 1000000 1000000 1000000 1000000 1000000 1000000 1000000  998832  977196
## [253]  977196  972888  972888  972888  972888  972888  969444  963504  962064
## [262]  960708  960000  959916  958584  953424  953244  936000  930672  930672
## [271]  930672  924000  921972  921000  920016  920016  919008  917280  908028
## [280]  908028  907572  900000  900000  899640  899640  884136  881892  877884
## [289]  876384  869112  868620  862200  856140  853116  852336  843168  843168
## [298]  843168  841584  841584  840186  840000  840000  840000  837600  817224
## [307]  817224  816000  811236  806580  806580  806580  804252  801720  781668
## [316]  780000  780000  778308  778308  778308  778308  778308  778308  778308
## [325]  778308  775560  775560  775560  775560  775560  775560  775560  774348
## [334]  768216  761628  753936  753936  753612  752364  744540  739392  729000
## [343]  726420  726060  720000  720000  713448  713448  712236  712104  701448
## [352]  700476  698086  698004  698004  698004  698004  694704  687504  687504
## [361]  687504  677448  674532  674532  674532  673272  672000  670164  666984
## [370]  660324  660000  651468  650000  648588  648588  648588  648588  648588
## [379]  647424  643716  643716  635616  629136  622644  620448  620448  620448
## [388]  620448  620448  620448  620448  620448  620448  613656  613344  608448
## [397]  603192  601284  600000  600000  600000  597180  596700  596700  594539
## [406]  594539  593268  591132  589428  589428  589428  587076  586392  583728
## [415]  583728  583728  583728  583728  583728  583728  583728  581667  581664
## [424]  578325  576000  570756  558396  553620  551304  544812  544812  544812
## [433]  542892  542892  542892  540490  540450  529128  528000  525048  525000
## [442]  518868  518868  518868  518868  518868  518868  518868  511872  511872
## [451]  511872  505896  505668  504000  502620  502620  500000  500000  499164
## [460]  496356  496356  492924  492924  492924  487000  485000  480840  480000
## [469]  479952  479952  473376  472500  466980  466980  466980  465336  465336
## [478]  463188  456000  455000  454008  454008  454008  450000  441036  438960
## [487]  435600  434316  434250  432392  432000  430000  430000  428064  421584
## [496]  418848  418800  418800  417972  415092  415092  415092  412324  409000
## [505]  405000  403284  402420  402420  400704  400000  400000  400000  400000
## [514]  400000  400000  389153  389148  389148  389148  389148  389148  387780
## [523]  385344  385000  381372  378343  377700  377484  376176  375528  375000
## [532]  372264  368592  363204  363204  360000  351319  350232  350232  350232
## [541]  350000  350000  350000  350000  344736  341244  337260  337260  331992
## [550]  326000  325728  325728  325596  324528  324300  323149  323000  320000
## [559]  320000  312000  311328  311328  310224  310224  310000  310000  310000
## [568]  306500  305000  303934  302472  301572  301572  300000  300000  300000
## [577]  300000  300000  300000  300000  300000  298356  298356  298356  297828
## [586]  295000  295000  294876  290000  290000  290000  288000  285384  285384
## [595]  285384  284371  280000  280000  279204  279204  275000  275000  275000
## [604]  273600  272412  272412  270250  270000  270000  268248  268068  267239
## [613]  265000  264960  264000  259440  259440  258519  258500  256332  255000
## [622]  253000  252000  252000  251316  251316  251316  250140  250000  250000
## [631]  250000  250000  250000  250000  250000  250000  250000  250000  250000
## [640]  250000  250000  250000  250000  250000  250000  250000  250000  250000
## [649]  250000  250000  250000  250000  250000  249000  245868  245000  244000
## [658]  240423  240000  240000  240000  240000  240000  240000  237000  235970
## [667]  235000  235000  232668  232000  231144  230000  230000  230000  230000
## [676]  230000  230000  230000  230000  230000  230000  230000  230000  230000
## [685]  230000  227497  227006  226894  226574  226204  226200  225000  225000
## [694]  225000  225000  225000  225000  225000  225000  225000  225000  224484
## [703]  224250  224136  223627  220000  220000  220000  220000  220000  220000
## [712]  220000  220000  220000  220000  220000  220000  220000  220000  220000
## [721]  219741  219741  219000  219000  218500  217156  217100  217000  216960
## [730]  216515  216196  216000  215830  215820  215000  215000  215000  215000
```

```
## [739]  215000  215000  215000  213463  213000  212750  212625  212269  212000
## [748]  211104  211000  210187  210132  210132  210050  210000  210000  210000
## [757]  210000  210000  210000  210000  210000  210000  210000  210000  210000
```

Al ordenar los valores y analizarlos encontramos que los valores máximos que podrían considerarse outliers son los mayores o iguales a 208000. Sin embargo, para efectos del análisis vamos a considerar como valores a los no sobrepasan el límite de 500K. Por lo tanto, procedemos a borrar esas observaciones que contienen los valores superiores a el límite.

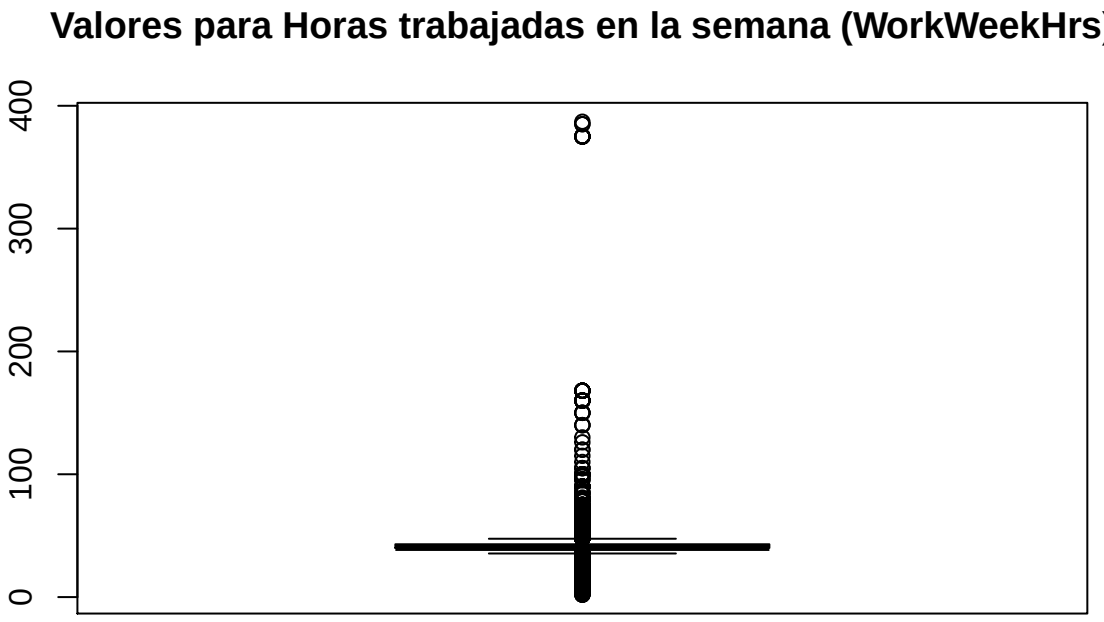
```
data_wo_na <- data_wo_na[data_wo_na$ConvertedComp <= 500000, ]

# volvemos a revisar los valores outliers
boxplot(as.numeric(data_wo_na$ConvertedComp), main="Valores para Sueldo Anual (ConvertedComp)", col="gra
```



A pesar de encontrarse unos valores superiores, por fuera del tercer cuartil (75%) aún se consideran válidos.

- Para la variable **WorkWeekHrs**

```
boxplot(as.numeric(data_wo_na$WorkWeekHrs), main="Valores para Horas trabajadas en la semana (WorkWeekHr
```



Para la variable Age los valores extremos los obtenenmos de la siguiente manera:

```
atipicos_horas_semanales <- boxplot.stats(as.numeric(data_wo_na$WorkWeekHrs))

sort(atipicos_horas_semanales$out, decreasing = FALSE)
```

```
##     [1]   2.00000   2.00000   2.00000   2.00000   3.00000   3.00000   3.00000
##     [8]   3.00000   4.00000   4.00000   4.00000   4.00000   4.00000   4.00000
##    [15]   5.00000   5.00000   5.00000   5.00000   5.00000   5.00000   5.00000
##    [22]   5.00000   5.00000   5.00000   5.00000   5.00000   5.00000   5.00000
##    [29]   5.00000   5.00000   5.00000   5.00000   5.00000   6.00000   6.00000
##    [36]   6.00000   6.00000   6.00000   6.00000   6.00000   6.00000   6.00000
##    [43]   6.00000   6.00000   6.00000   6.00000   6.00000   6.00000   6.00000
##    [50]   6.00000   6.00000   6.00000   6.00000   6.00000   6.00000   6.00000
##    [57]   6.00000   7.00000   7.00000   7.00000   7.00000   7.00000   7.00000
##    [64]   7.00000   7.00000   7.00000   7.00000   7.00000   7.00000   7.00000
##    [71]   7.00000   7.00000   7.00000   7.00000   7.00000   7.00000   7.00000
##    [78]   7.00000   7.00000   7.00000   7.00000   7.00000   7.00000   7.00000
##    [85]   7.00000   7.00000   7.50000   7.50000   8.00000   8.00000   8.00000
##    [92]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##    [99]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [106]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [113]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [120]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [127]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [134]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [141]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [148]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [155]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##   [162]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
```

```
##  [169]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [176]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [183]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [190]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [197]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [204]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [211]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [218]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [225]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [232]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [239]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [246]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [253]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [260]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [267]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [274]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [281]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [288]   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000   8.00000
##  [295]   8.50000   8.50000   8.50000   9.00000   9.00000   9.00000   9.00000
##  [302]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [309]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [316]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [323]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [330]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [337]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [344]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [351]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [358]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [365]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [372]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [379]   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000   9.00000
##  [386]   9.00000   9.15000  10.00000  10.00000  10.00000  10.00000  10.00000
##  [393]  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000
##  [400]  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000
##  [407]  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000
##  [414]  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000
##  [421]  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000
##  [428]  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000  10.00000
##  [435]  10.00000  10.00000  11.00000  11.00000  11.00000  11.00000  11.00000
##  [442]  11.00000  12.00000  12.00000  12.00000  12.00000  12.00000  12.00000
##  [449]  12.00000  12.00000  12.00000  12.00000  12.00000  12.00000  12.00000
##  [456]  12.00000  12.00000  12.00000  12.00000  12.00000  12.00000  12.00000
##  [463]  12.00000  12.00000  12.00000  13.00000  13.00000  13.00000  13.00000
##  [470]  13.00000  13.00000  13.00000  14.00000  14.00000  14.00000  14.00000
##  [477]  14.00000  14.00000  14.00000  15.00000  15.00000  15.00000  15.00000
##  [484]  15.00000  15.00000  15.00000  15.00000  15.00000  15.00000  15.00000
##  [491]  15.00000  16.00000  16.00000  16.00000  16.00000  16.00000  16.00000
##  [498]  16.00000  16.00000  16.00000  16.00000  16.00000  16.00000  17.00000
##  [505]  18.00000  18.00000  18.00000  18.00000  18.00000  18.00000  18.00000
##  [512]  18.00000  19.00000  19.00000  19.00000  20.00000  20.00000  20.00000
##  [519]  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000
##  [526]  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000
##  [533]  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000
##  [540]  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000
##  [547]  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000
##  [554]  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000
##  [561]  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000
##  [568]  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000  20.00000
```

```
##  [575] 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000
##  [582] 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000
##  [589] 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000
##  [596] 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000
##  [603] 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000
##  [610] 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000
##  [617] 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000 20.00000
##  [624] 20.00000 20.00000 20.00000 21.00000 21.00000 21.00000 21.00000
##  [631] 21.00000 21.00000 22.00000 22.00000 22.00000 23.00000 23.00000
##  [638] 24.00000 24.00000 24.00000 24.00000 24.00000 24.00000 24.00000
##  [645] 24.00000 24.00000 24.00000 24.00000 24.00000 24.00000 24.00000
##  [652] 24.00000 24.00000 24.00000 24.00000 24.00000 24.00000 24.00000
##  [659] 24.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [666] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [673] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [680] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [687] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [694] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [701] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [708] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [715] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [722] 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000 25.00000
##  [729] 25.00000 26.00000 26.00000 26.00000 26.00000 27.00000 27.00000
##  [736] 28.00000 28.00000 28.00000 28.00000 28.00000 28.00000 28.00000
##  [743] 28.00000 28.00000 28.00000 28.00000 28.00000 28.00000 29.00000
##  [750] 29.00000 29.40000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [757] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [764] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [771] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [778] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [785] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [792] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [799] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [806] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [813] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [820] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [827] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [834] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [841] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [848] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [855] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [862] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [869] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [876] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [883] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [890] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [897] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [904] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [911] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [918] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [925] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [932] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [939] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [946] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [953] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [960] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [967] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
##  [974] 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000 30.00000
```

```
##  [981]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
##  [988]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
##  [995]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
## [1002]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
## [1009]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
## [1016]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
## [1023]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
## [1030]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
## [1037]  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000  30.00000
## [1044]  31.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1051]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1058]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1065]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1072]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1079]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1086]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1093]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1100]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1107]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1114]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1121]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1128]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1135]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1142]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000
## [1149]  32.00000  32.00000  32.00000  32.00000  32.00000  32.00000  33.00000
## [1156]  33.00000  33.00000  33.00000  33.00000  33.00000  33.00000  33.00000
## [1163]  33.00000  33.00000  33.00000  33.75000  34.00000  34.00000  34.00000
## [1170]  34.00000  34.00000  34.00000  34.00000  34.00000  34.00000  34.00000
## [1177]  34.00000  34.00000  34.00000  34.00000  34.00000  35.00000  35.00000
## [1184]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1191]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1198]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1205]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1212]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1219]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1226]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1233]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1240]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1247]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1254]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1261]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1268]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1275]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1282]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1289]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1296]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1303]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1310]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1317]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1324]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1331]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1338]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1345]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1352]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1359]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1366]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1373]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1380]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
```

```
## [1387]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1394]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1401]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1408]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1415]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1422]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1429]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1436]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1443]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1450]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1457]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1464]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1471]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1478]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1485]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1492]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1499]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1506]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1513]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1520]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1527]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1534]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1541]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1548]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1555]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1562]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1569]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1576]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1583]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1590]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1597]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1604]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1611]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1618]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1625]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1632]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1639]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1646]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1653]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1660]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1667]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1674]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1681]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1688]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1695]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1702]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1709]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1716]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1723]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1730]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1737]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1744]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1751]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1758]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1765]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1772]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1779]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1786]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
```

```
## [1793]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1800]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1807]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1814]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1821]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1828]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1835]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1842]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1849]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1856]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1863]  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000  35.00000
## [1870]  35.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1877]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1884]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1891]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1898]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1905]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1912]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1919]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1926]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1933]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1940]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1947]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1954]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1961]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1968]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1975]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1982]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1989]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [1996]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [2003]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [2010]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [2017]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [2024]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [2031]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [2038]  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000  48.00000
## [2045]  48.00000  48.00000  49.00000  49.00000  49.00000  49.00000  49.00000
## [2052]  49.00000  49.00000  49.00000  49.00000  49.00000  49.00000  50.00000
## [2059]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2066]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2073]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2080]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2087]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2094]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2101]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2108]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2115]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2122]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2129]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2136]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2143]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2150]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2157]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2164]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2171]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2178]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2185]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2192]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
```

```
## [2199]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2206]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2213]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2220]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2227]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2234]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2241]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2248]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2255]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2262]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2269]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2276]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2283]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2290]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2297]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2304]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2311]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2318]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2325]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2332]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2339]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2346]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2353]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2360]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2367]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2374]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2381]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2388]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2395]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2402]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2409]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2416]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2423]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2430]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2437]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2444]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2451]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2458]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2465]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2472]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2479]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2486]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2493]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2500]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2507]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2514]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2521]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2528]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2535]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2542]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2549]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2556]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2563]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2570]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2577]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2584]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2591]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2598]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
```

```
## [2605]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2612]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2619]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2626]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2633]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2640]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2647]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2654]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2661]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2668]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2675]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2682]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2689]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2696]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2703]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2710]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2717]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2724]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2731]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2738]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2745]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2752]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2759]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2766]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2773]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2780]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2787]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2794]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2801]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2808]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2815]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2822]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2829]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2836]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2843]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2850]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2857]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2864]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2871]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2878]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2885]  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000  50.00000
## [2892]  50.00000  50.00000  50.00000  50.11111  51.00000  51.00000  51.00000
## [2899]  52.00000  52.00000  52.00000  52.00000  52.00000  52.00000  52.00000
## [2906]  52.00000  52.00000  52.00000  52.00000  52.00000  52.00000  52.00000
## [2913]  52.00000  53.00000  53.00000  54.00000  54.00000  54.00000  54.00000
## [2920]  54.00000  54.00000  54.00000  54.00000  54.00000  54.00000  54.00000
## [2927]  54.00000  54.00000  54.00000  54.00000  54.00000  54.00000  54.00000
## [2934]  54.00000  54.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2941]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2948]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2955]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2962]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2969]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2976]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2983]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2990]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [2997]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [3004]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
```

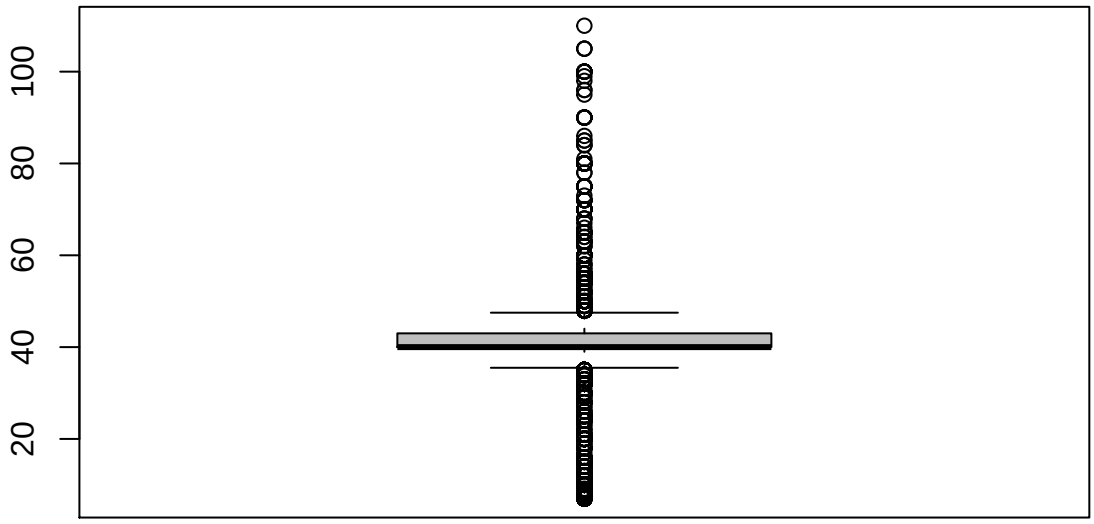
```
## [3011]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [3018]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [3025]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [3032]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [3039]  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000  55.00000
## [3046]  55.00000  55.00000  55.00000  55.00000  56.00000  56.00000  56.00000
## [3053]  56.00000  56.00000  56.00000  56.00000  56.00000  56.00000  56.00000
## [3060]  56.00000  56.00000  56.00000  56.00000  56.00000  56.00000  56.00000
## [3067]  56.00000  56.00000  56.00000  56.00000  56.00000  56.00000  56.00000
## [3074]  56.00000  57.00000  57.00000  58.00000  58.00000  58.00000  59.00000
## [3081]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3088]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3095]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3102]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3109]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3116]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3123]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3130]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3137]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3144]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3151]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3158]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3165]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3172]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3179]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3186]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3193]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3200]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3207]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3214]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3221]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3228]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3235]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3242]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3249]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3256]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3263]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3270]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3277]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3284]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3291]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3298]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3305]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3312]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3319]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3326]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3333]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3340]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3347]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3354]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3361]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3368]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000
## [3375]  60.00000  60.00000  60.00000  60.00000  60.00000  60.00000  62.00000
## [3382]  62.00000  62.00000  63.00000  63.00000  63.00000  63.00000  63.00000
## [3389]  63.00000  63.00000  64.00000  64.00000  65.00000  65.00000  65.00000
## [3396]  65.00000  65.00000  65.00000  65.00000  65.00000  65.00000  65.00000
## [3403]  65.00000  65.00000  65.00000  65.00000  66.00000  67.00000  68.00000
## [3410]  68.00000  68.00000  70.00000  70.00000  70.00000  70.00000  70.00000
```

```
## [3417]  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000
## [3424]  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000
## [3431]  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000
## [3438]  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000
## [3445]  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000
## [3452]  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000
## [3459]  70.00000  70.00000  70.00000  70.00000  70.00000  70.00000  72.00000
## [3466]  72.00000  72.00000  72.00000  72.00000  72.00000  73.00000  75.00000
## [3473]  75.00000  75.00000  75.00000  75.00000  75.00000  75.00000  75.00000
## [3480]  75.00000  75.00000  75.00000  75.00000  78.00000  78.00000  80.00000
## [3487]  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000
## [3494]  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000
## [3501]  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000
## [3508]  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000
## [3515]  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000  80.00000
## [3522]  80.00000  80.00000  81.00000  84.00000  84.00000  84.00000  85.00000
## [3529]  85.00000  86.00000  90.00000  90.00000  90.00000  90.00000  90.00000
## [3536]  90.00000  90.00000  90.00000  90.00000  95.00000  96.00000  96.00000
## [3543]  98.00000  99.00000 100.00000 100.00000 100.00000 100.00000 100.00000
## [3550] 100.00000 100.00000 100.00000 105.00000 105.00000 105.00000 110.00000
## [3557] 115.00000 120.00000 120.00000 126.00000 130.00000 140.00000 140.00000
## [3564] 140.00000 150.00000 150.00000 150.00000 150.00000 160.00000 160.00000
## [3571] 160.00000 160.00000 160.00000 160.00000 160.00000 160.00000 160.00000
## [3578] 160.00000 160.00000 160.00000 168.00000 168.00000 168.00000 168.00000
## [3585] 168.00000 168.00000 168.00000 168.00000 168.00000 168.00000 168.00000
## [3592] 168.00000 375.00000 375.00000 375.00000 375.00000 375.00000 375.00000
## [3599] 375.00000 375.00000 375.00000 385.00000 385.00000 387.00000
```

Al ordenar los valores y analizarlos encontramos que los valores máximos que podrían considerarse outliers son los mayores a 112 horas, ya que trabajar más de 16 horas diarias se encuentra poco probable. De la misma manera, se considera poco probable trabajar menos de una hora diaria (7 horas semanales, incluso en un trabajo a tiempo parcial). Por lo tanto, procedemos a borrar esas observaciones que contienen los valores superiores e inferiores a los límites mencionados.

Cabe recalcar que el valor 112 se obtiene en el supuesto caso que se trabaje 16 horas diarias los 7 dias de la semana, en el caso extremo.

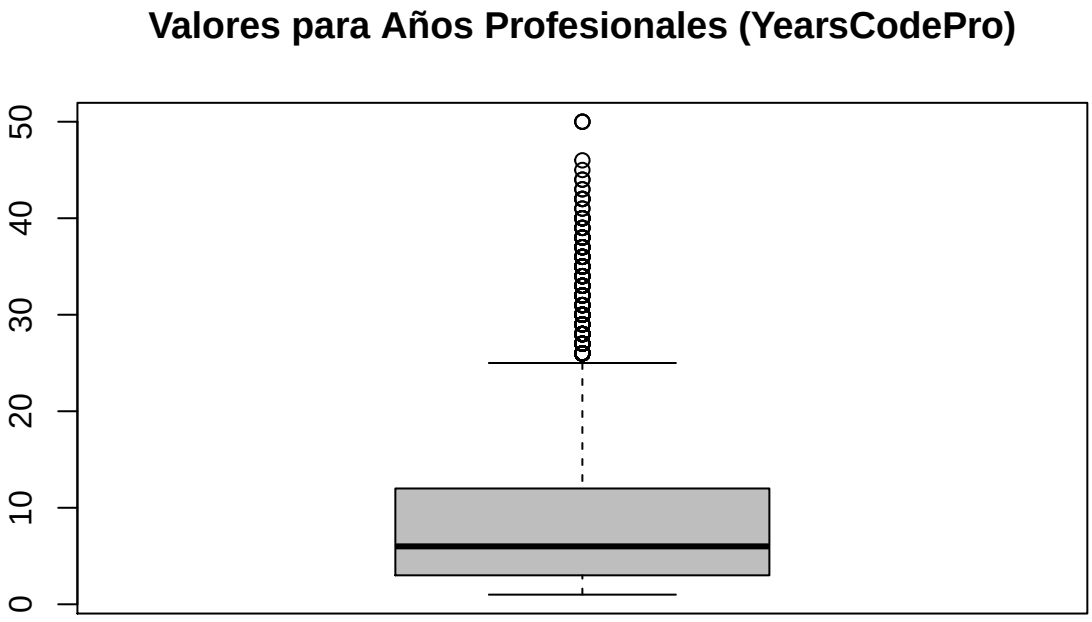
```
data_wo_na <- data_wo_na[data_wo_na$WorkWeekHrs <= 112, ]
data_wo_na <- data_wo_na[data_wo_na$WorkWeekHrs >= 7, ]

# volvemos a revisar los valores outliers
boxplot(as.numeric(data_wo_na$WorkWeekHrs), main="Valores para Horas trabajadas en la semana (WorkWeekHr
```



A pesar de encontrarse unos valores superiores e inferiores, por fuera del tercer(75%) y primer (25%) cuartil, respectivamente; aún se consideran válidos para efectos del análisis que realizaremos.

- Para la variable **YearsCodePro**

```
boxplot(as.numeric(data_wo_na$YearsCodePro), main="Valores para Años Profesionales (YearsCodePro)", col=
```



Para la variable Age los valores extremos los obtenenmos de la siguiente manera:

```
atipicos_years_pro <- boxplot.stats(as.numeric(data_wo_na$YearsCodePro))

sort(atipicos_years_pro$out, decreasing = TRUE)
```

```
##   [1] 50 50 50 46 45 44 44 43 43 42 42 42 42 42 41 41 40 40 40 40 40 40 40 40 40
##  [26] 39 39 39 39 39 38 38 38 38 38 38 38 38 38 38 38 37 37 37 37 37 37 37 37 37 36
##  [51] 36 36 36 36 36 36 36 36 36 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35
##  [76] 35 35 35 35 35 35 35 35 35 35 35 35 34 34 34 34 34 34 34 34 34 34 34 33 33
## [101] 33 33 33 33 33 33 33 33 33 33 33 33 33 33 33 33 33 33 32 32 32 32 32 32 32
## [126] 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 32 31 31 31 31 31
## [151] 31 31 31 31 31 31 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30
## [176] 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30
## [201] 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30
## [226] 30 30 30 30 29 29 29 29 29 29 29 29 29 29 29 29 29 29 28 28 28 28 28 28 28 28
## [251] 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28 28
## [276] 28 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27 27
## [301] 27 27 27 27 27 27 27 27 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26
## [326] 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26
## [351] 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26 26
```

Al ordenar los valores outliers de manera descendente, encontramos los valores en el rango entre 26 y 50 años. Sin embargo, yendo a la lógica actual en los años que podría tener un desarrollador no encontramos como valor atípico una experiencia comprendida entre los mencionados límites. Para verificar de manera más detallada esta variable, vamos a comparar la variable Age vs. la variable YearsCodePro para descartar alguna anormalidad, como que los años de profesional sean mayor a la edad actual o no haya lógica entre la una con la otra.

Para esto vamos a proceder a generar un data frame entre las 2 variables considerando los valores comprendidos en el rango antes mencionado.

```r
data_age_years_pro <- data_wo_na[data_wo_na$YearsCodePro >= 26, c(1, 12)]

# Obtenemos una variable que almacene la diferencia en la edad como profesional y la edad actual
filas_ndf <- dim(data_age_years_pro)[1]

data_age_years_pro$diferencia <- 0

for(i in 1:filas_ndf) {
  data_age_years_pro$diferencia[i] <- data_age_years_pro$Age[i] - data_age_years_pro$YearsCodePro[i]
}


sort(data_age_years_pro$diferencia, decreasing = TRUE)
```

```
##   [1] 42 40 37 32 32 32 32 32 31 31 31 31 31 30 30 30 30 29 29 29 29 29 29 28 28
##  [26] 28 28 28 28 28 28 28 27 27 27 27 27 27 27 27 27 26 26 26 26 26 26 26 26 26 26
##  [51] 26 26 26 26 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25
##  [76] 25 25 25 25 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24 24
## [101] 24 24 24 24 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23 23
## [126] 23 23 23 23 23 23 23 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22
## [151] 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22 22
## [176] 22 22 22 22 22 22 22 22 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21
## [201] 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21
## [226] 21 21 21 21 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20
## [251] 20 20 20 20 20 20 20 20 20 20 20 20 19 19 19 19 19 19 19 19 19 19 19 19 19 19
## [276] 19 19 19 19 19 19 19 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18 18
## [301] 18 18 18 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17 17
## [326] 17 17 17 17 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 16 15 15 15 15 15 15
## [351] 15 15 15 15 15 14 14 14 14 14 14 14 12 12 11 10 10  8
```

Todos los valores en las diferencias obtenidas son lógicos. Sin embargo, a pesar de encontrarse una diferencia mínima de 8 años entre la experiencia como profesional y la edad actual, sabemos de casos extremos de niños que pueden aprender a programar o cualquier otra habilidad a corta edad, por ende no eliminaremos ningún registro.

4 Análisis de los datos

5 Representación de los resultados a partir de tablas y gráficas

6 Resolución del problema

7 Código