

EDA LA-2 REPORT

Amogh Javali USN:1NT21IS026

2024-01-27

Stack Overflow Developer Survey Dataset

Load a dataset:

```
my_data <- read.csv("C:/Users/amogh/Desktop/AMOGH/Book1.csv")
```

Quickly Exploring Data

summary statistics

```
# Display summary statistics for the entire dataset
summary(my_data)
```

##	ResponseId	Q120	MainBranch	Age
##	Min. : 1.00	Length:60	Length:60	Length:60
##	1st Qu.:15.75	Class :character	Class :character	Class :character
##	Median :30.50	Mode :character	Mode :character	Mode :character
##	Mean :30.50			
##	3rd Qu.:45.25			
##	Max. :60.00			
##				
##	Employment	RemoteWork	CodingActivities	EdLevel
##	Length:60	Length:60	Length:60	Length:60
##	Class :character	Class :character	Class :character	Class :character
##	Mode :character	Mode :character	Mode :character	Mode :character
##				
##				
##				
##				
##	LearnCode	LearnCodeOnline	LearnCodeCoursesCert	YearsCode
##	Length:60	Length:60	Length:60	Min. : 1.00
##	Class :character	Class :character	Class :character	1st Qu.: 5.50
##	Mode :character	Mode :character	Mode :character	Median :10.00
##				Mean :12.47
##				3rd Qu.:17.00
##				Max. :35.00

```

##                                     NA's      :1
##   YearsCodePro      DevType      OrgSize      PurchaseInfluence
##   Min.      : 1.000      Length:60      Length:60      Length:60
##   1st Qu.: 4.000      Class :character      Class :character      Class :character
##   Median : 9.000      Mode  :character      Mode  :character      Mode  :character
##   Mean    : 9.203
##   3rd Qu.:13.000
##   Max.    :30.000
##   NA's    :1
##   TechList          BuyNewTool          Country          Currency
##   Length:60          Length:60          Length:60          Length:60
##   Class :character      Class :character      Class :character      Class :character
##   Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##   CompTotal          LanguageHaveWorkedWith LanguageWantToWorkWith
##   Min.      : 5555      Length:60          Length:60
##   1st Qu.: 64500      Class :character      Class :character
##   Median : 102500      Mode  :character      Mode  :character
##   Mean    : 227580
##   3rd Qu.: 151500
##   Max.    :2000000
##   NA's    :16
##   DatabaseHaveWorkedWith DatabaseWantToWorkWith PlatformHaveWorkedWith
##   Length:60          Length:60          Length:60
##   Class :character      Class :character      Class :character
##   Mode  :character      Mode  :character      Mode  :character
##
##
##
##   PlatformWantToWorkWith WebframeHaveWorkedWith WebframeWantToWorkWith
##   Length:60          Length:60          Length:60
##   Class :character      Class :character      Class :character
##   Mode  :character      Mode  :character      Mode  :character
##
##
##
##   MiscTechHaveWorkedWith MiscTechWantToWorkWith ToolsTechHaveWorkedWith
##   Length:60          Length:60          Length:60
##   Class :character      Class :character      Class :character
##   Mode  :character      Mode  :character      Mode  :character
##
##
##
##   ToolsTechWantToWorkWith NEWCollabToolsHaveWorkedWith
##   Length:60          Length:60
##   Class :character      Class :character
##   Mode  :character      Mode  :character
##

```

```

##
##
##
## NEWCollabToolsWantToWorkWith OpSysPersonal.use OpSysProfessional.use
## Length:60 Length:60 Length:60
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
## OfficeStackAsyncHaveWorkedWith OfficeStackAsyncWantToWorkWith
## Length:60 Length:60
## Class :character Class :character
## Mode :character Mode :character
##
##
##
## OfficeStackSyncHaveWorkedWith OfficeStackSyncWantToWorkWith
## Length:60 Length:60
## Class :character Class :character
## Mode :character Mode :character
##
##
##
## AISearchHaveWorkedWith AISearchWantToWorkWith AIDevHaveWorkedWith
## Length:60 Length:60 Length:60
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
## AIDevWantToWorkWith NEWSOSites SOVisitFreq SOAccount
## Length:60 Length:60 Length:60 Length:60
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## SOPartFreq SOComm SOAI AISelect
## Length:60 Length:60 Length:60 Length:60
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## AISent AIAcc AIBen
## Length:60 Length:60 Length:60
## Class :character Class :character Class :character

```

```

## Mode :character Mode :character Mode :character
##
##
##
##
## AIToolInterested.in.Using AIToolCurrently.Using AIToolNot.interested.in.Using
## Length:60 Length:60 Length:60
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
##
## AINextVery.different AINextNeither.different.nor.similar
## Mode:logical Mode:logical
## NA's:60 NA's:60
##
##
##
##
## AINextSomewhat.similar AINextVery.similar AINextSomewhat.different
## Mode:logical Mode:logical Mode:logical
## NA's:60 NA's:60 NA's:60
##
##
##
##
## TBranch ICorPM WorkExp Knowledge_1
## Length:60 Length:60 Min. : 2.00 Length:60
## Class :character Class :character 1st Qu.: 5.00 Class :character
## Mode :character Mode :character Median : 8.00 Mode :character
## Mean :10.19
## 3rd Qu.:13.25
## Max. :39.00
## NA's :24
## Knowledge_2 Knowledge_3 Knowledge_4 Knowledge_5
## Length:60 Length:60 Length:60 Length:60
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
##
## Knowledge_6 Knowledge_7 Knowledge_8 Frequency_1
## Length:60 Length:60 Length:60 Length:60
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
##
## Frequency_2 Frequency_3 TimeSearching TimeAnswering

```

```
## Length:60      Length:60      Length:60      Length:60
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## ProfessionalTech Industry      SurveyLength      SurveyEase
## Length:60      Length:60      Length:60      Length:60
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## CCY
## Min. : 14
## 1st Qu.: 53873
## Median : 76507
## Mean : 93960
## 3rd Qu.:122500
## Max. :285000
## NA's :16
```

```
# Display summary statistics for numerical variables only
summary(my_data[, sapply(my_data, is.numeric)])
```

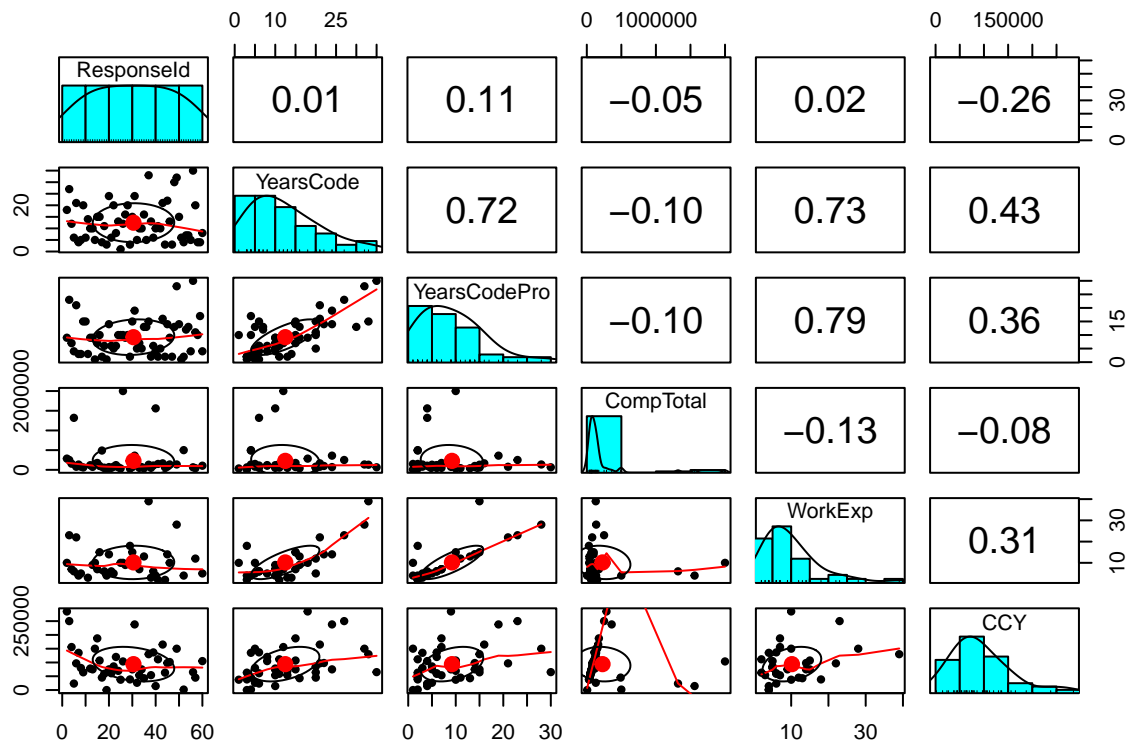
```
## ResponseId      YearsCode      YearsCodePro      CompTotal
## Min. : 1.00      Min. : 1.00      Min. : 1.000      Min. : 5555
## 1st Qu.:15.75      1st Qu.: 5.50      1st Qu.: 4.000      1st Qu.: 64500
## Median :30.50      Median :10.00      Median : 9.000      Median : 102500
## Mean :30.50      Mean :12.47      Mean : 9.203      Mean : 227580
## 3rd Qu.:45.25      3rd Qu.:17.00      3rd Qu.:13.000      3rd Qu.: 151500
## Max. :60.00      Max. :35.00      Max. :30.000      Max. :2000000
## NA's :1          NA's :1          NA's :16
## WorkExp          CCY
## Min. : 2.00      Min. : 14
## 1st Qu.: 5.00      1st Qu.: 53873
## Median : 8.00      Median : 76507
## Mean :10.19      Mean : 93960
## 3rd Qu.:13.25      3rd Qu.:122500
## Max. :39.00      Max. :285000
## NA's :24          NA's :16
```

Pair Plots

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.3.2
```

```
pairs.panels(my_data[, sapply(my_data, is.numeric)])
```



Bar Graphs

Bar Plot

```
library(ggplot2)
```

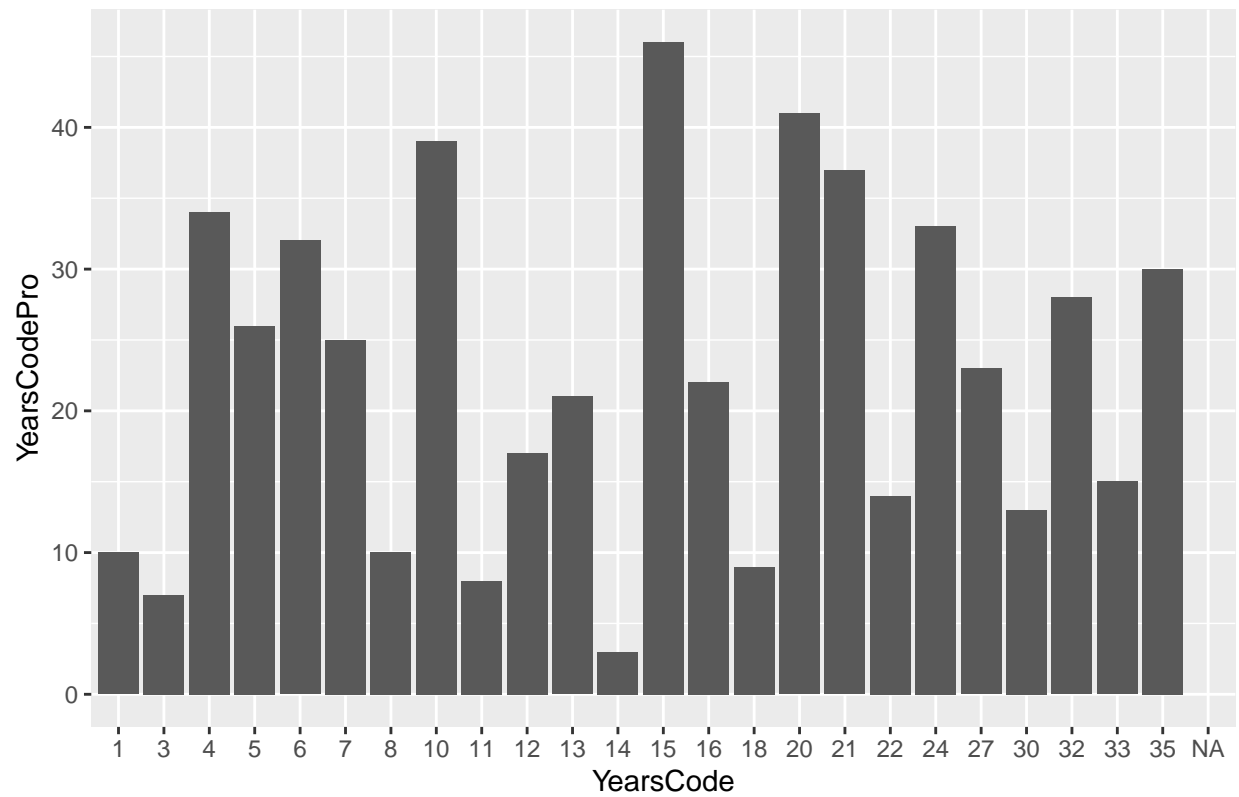
```
##
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
##
## %+%, alpha
```

```
ggplot(my_data, aes(x = as.factor(YearsCode), y = YearsCodePro)) +
  geom_bar(stat = "identity") +
  ggtitle("Bar Plot") +
  xlab("YearsCode") + ylab("YearsCodePro")
```

```
## Warning: Removed 1 rows containing missing values ('position_stack()').
```

Bar Plot

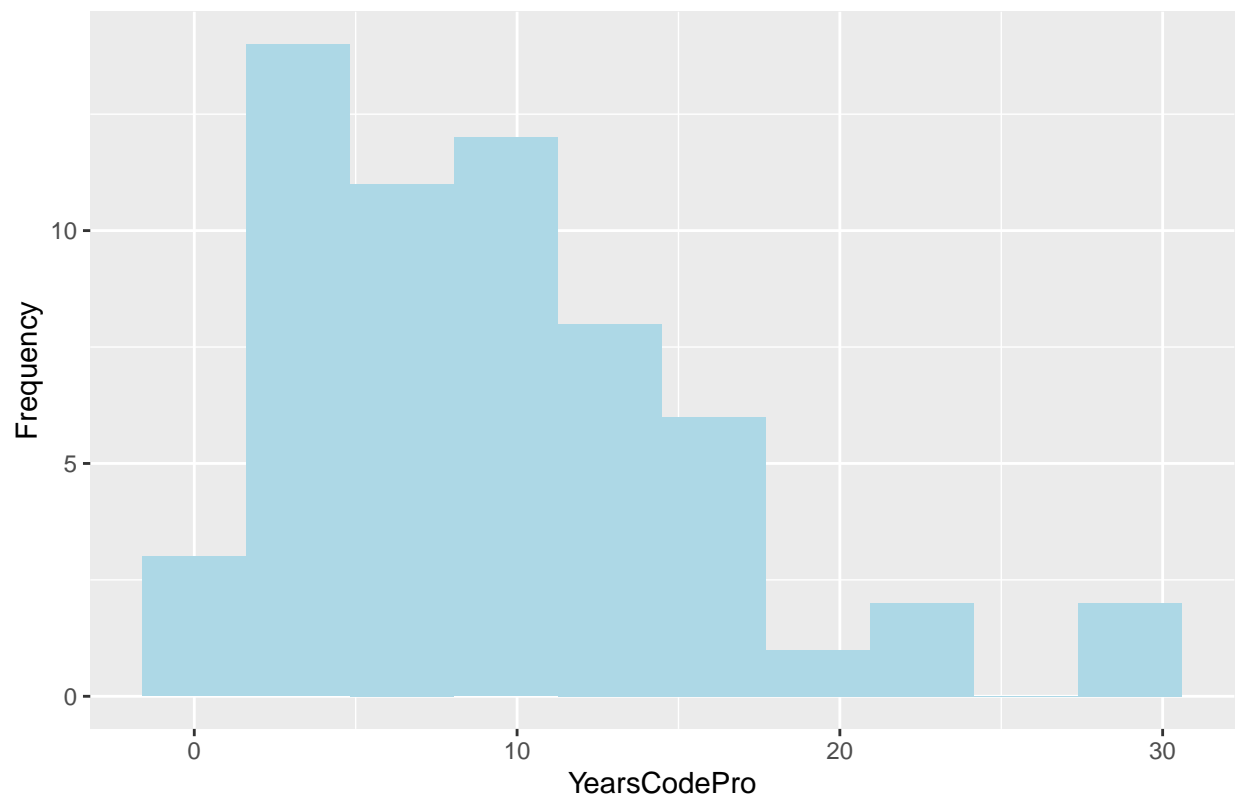


Histogram

```
ggplot(my_data, aes(x = YearsCodePro )) +
  geom_histogram(fill = "lightblue", bins = 10) +
  ggtitle("Histogram") +
  xlab("YearsCodePro") + ylab("Frequency")
```

Warning: Removed 1 rows containing non-finite values ('stat_bin()').

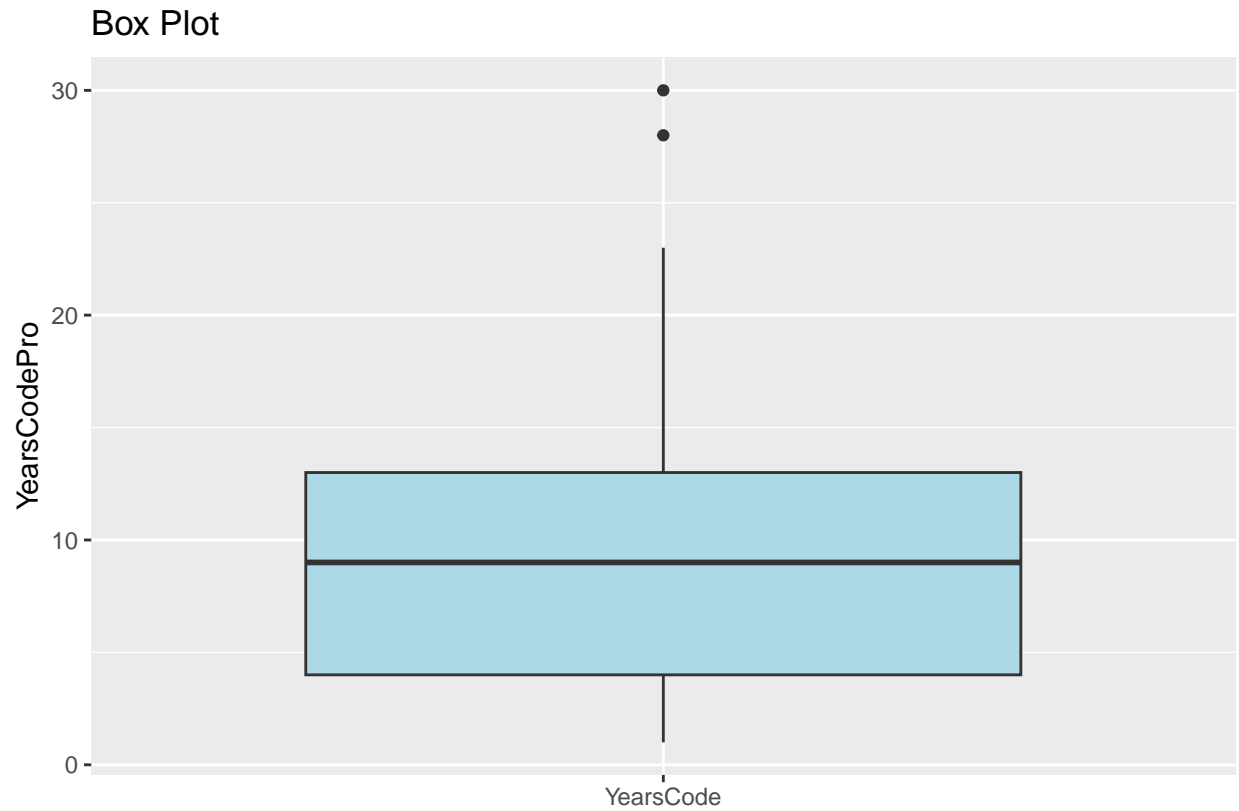
Histogram



Box plot

```
ggplot(my_data, aes(x = "YearsCode", y = YearsCodePro)) +  
  geom_boxplot(fill = "lightblue") +  
  ggtitle("Box Plot") +  
  xlab("") + ylab("YearsCodePro")
```

Warning: Removed 1 rows containing non-finite values ('stat_boxplot()').



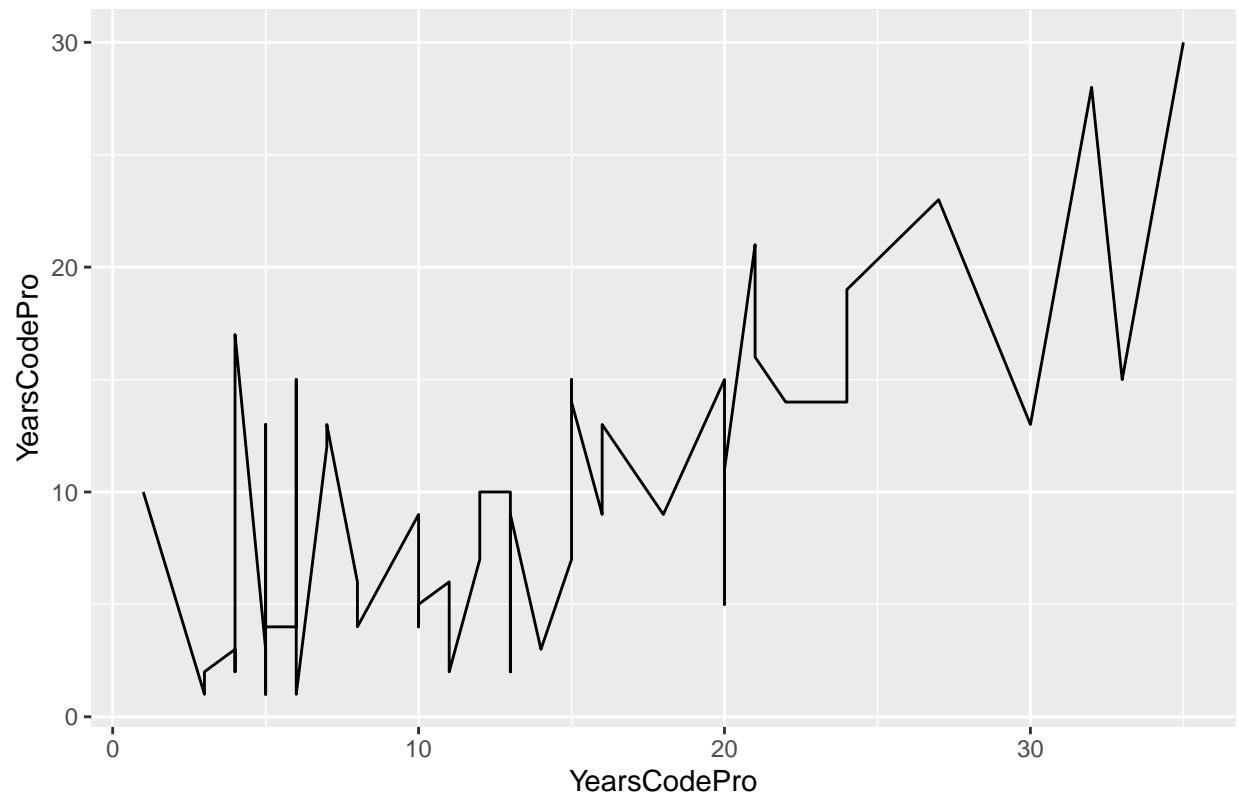
Line Graphs

Line plot

```
ggplot(my_data, aes(x = YearsCode , y = YearsCodePro)) +  
  geom_line() +  
  ggtitle("Line Plot") +  
  xlab("YearsCodePro") + ylab("YearsCodePro")
```

```
## Warning: Removed 1 row containing missing values ('geom_line()').
```

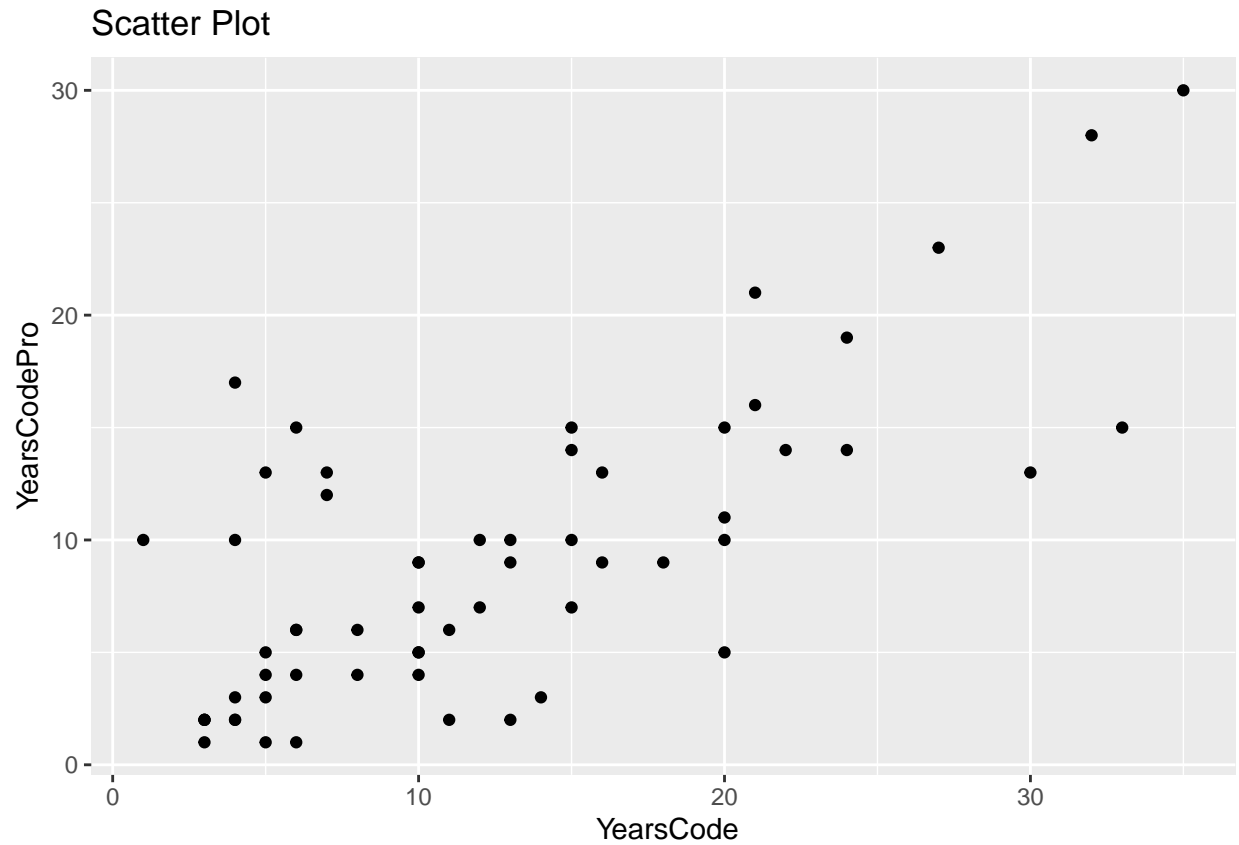
Line Plot



Scatter Plot

```
ggplot(my_data, aes(x = YearsCode, y = YearsCodePro)) +  
  geom_point() +  
  ggtitle("Scatter Plot") +  
  xlab(" YearsCode") + ylab("YearsCodePro")
```

```
## Warning: Removed 1 rows containing missing values ('geom_point()').
```

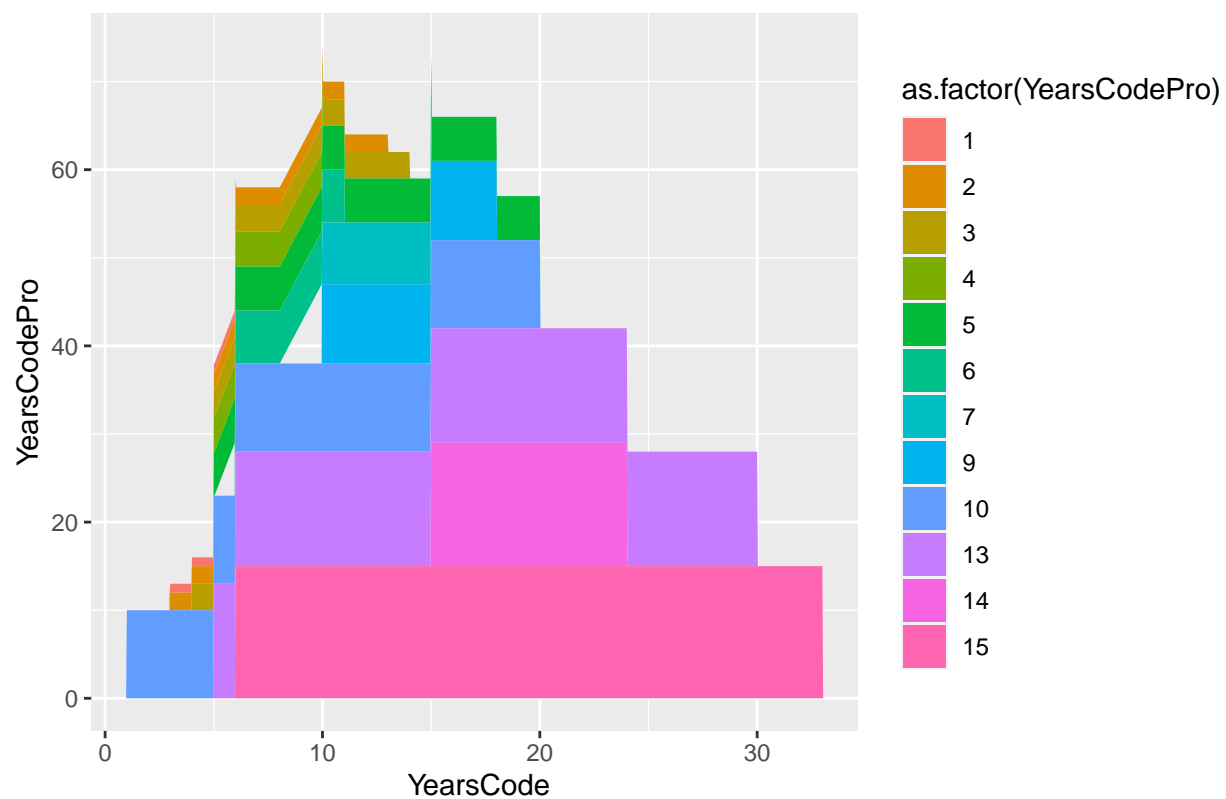


Area plot

```
ggplot(my_data, aes(x = YearsCode, y = YearsCodePro , fill = as.factor( YearsCodePro ))) +  
  geom_area() +  
  ggtitle("Area Plot") +  
  xlab(" YearsCode") + ylab(" YearsCodePro ")
```

Warning: Removed 1 rows containing non-finite values ('stat_align()').

Area Plot

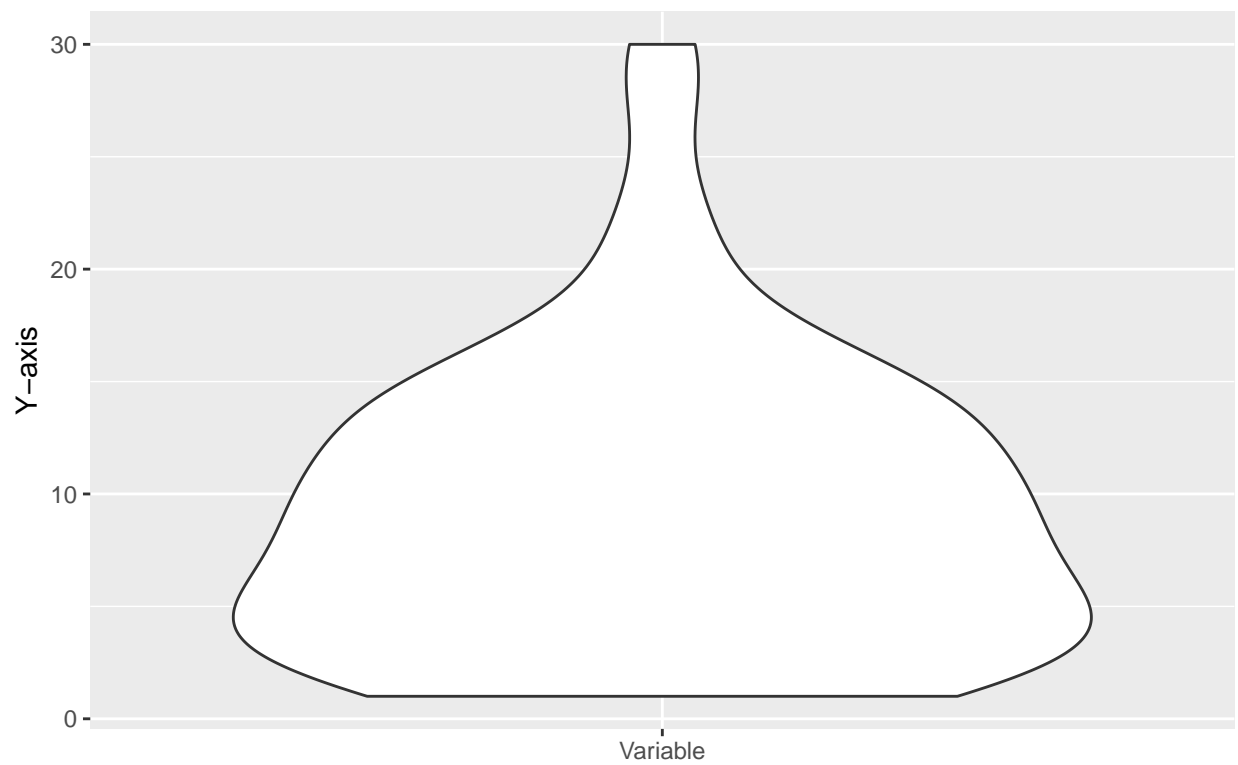


Violin

```
ggplot(my_data, aes(x = "Variable", y = YearsCodePro)) +  
  geom_violin() +  
  ggtitle("Violin Plot") +  
  xlab("") + ylab("Y-axis")
```

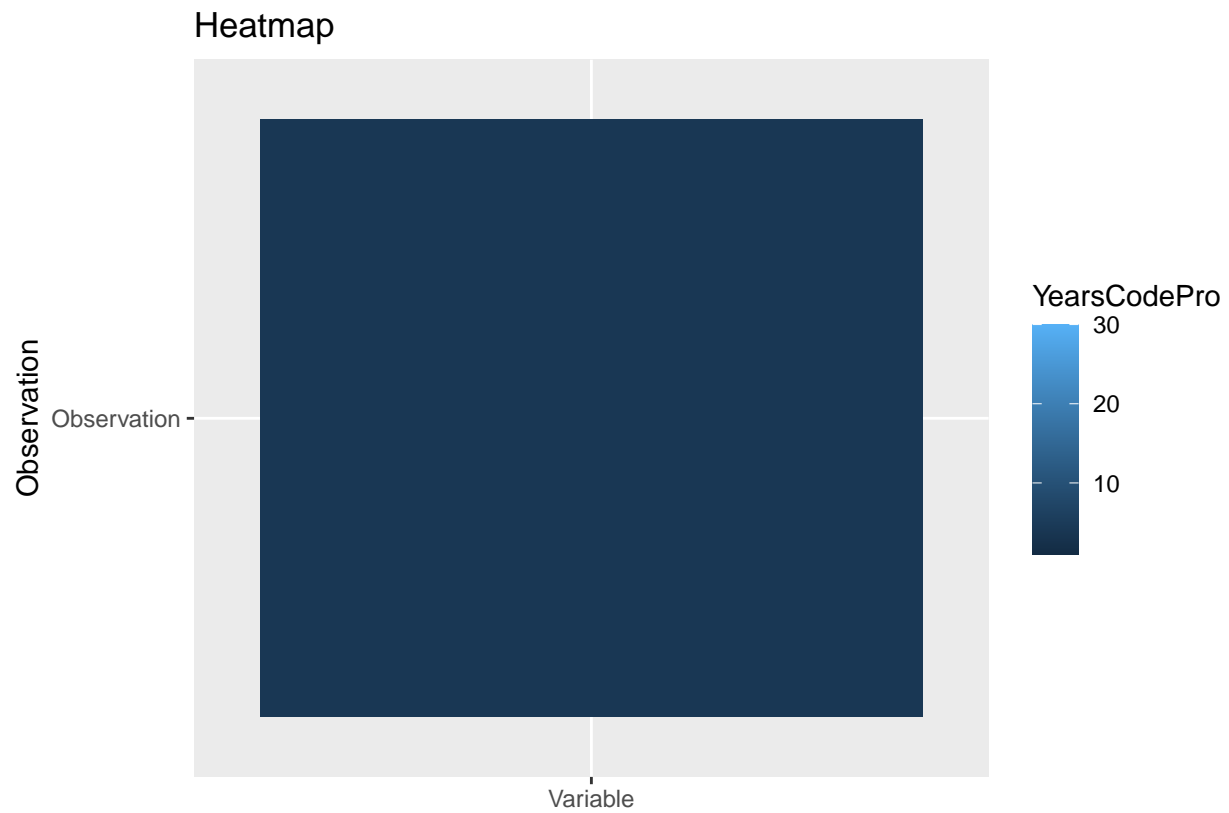
Warning: Removed 1 rows containing non-finite values ('stat_ydensity()').

Violin Plot



Heatmap

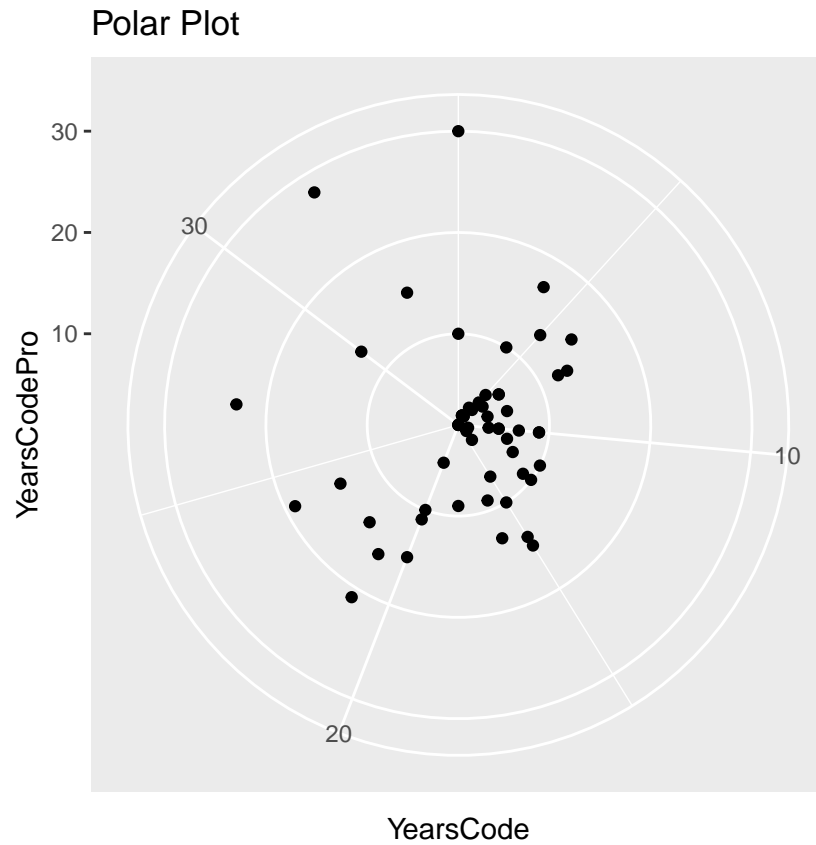
```
ggplot(my_data, aes(x = "Variable", y = "Observation", fill = YearsCodePro)) +  
  geom_tile() +  
  ggtitle("Heatmap") +  
  xlab("") + ylab("Observation")
```



Polar plot

```
ggplot(my_data, aes(x = YearsCode, y = YearsCodePro)) +  
  geom_point() +  
  coord_polar() +  
  ggtitle("Polar Plot") +  
  xlab("YearsCode") + ylab("YearsCodePro")
```

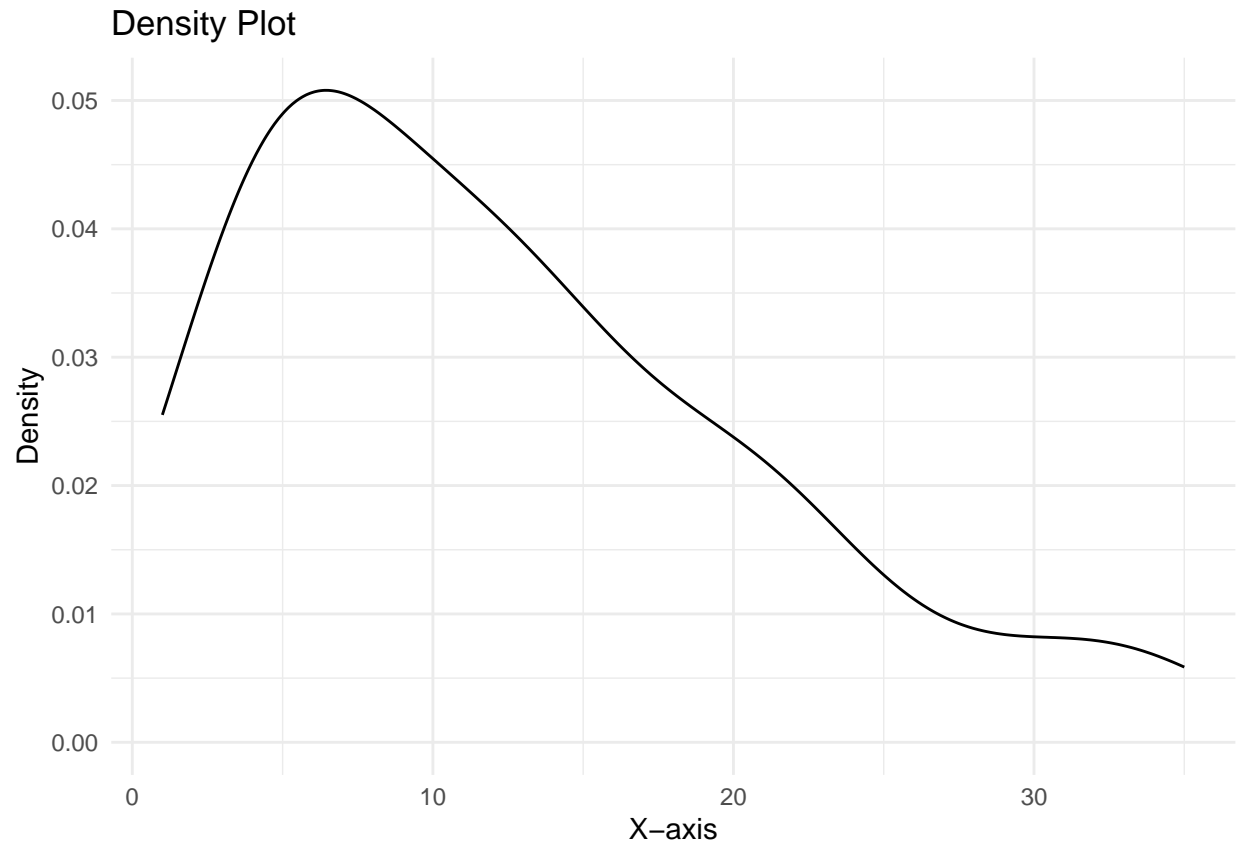
```
## Warning: Removed 1 rows containing missing values ('geom_point()').
```



density curve

```
ggplot(my_data, aes(x = YearsCode)) +  
  geom_density(color = "black", alpha = 0.7) +  
  ggtitle("Density Plot") +  
  xlab("X-axis") +  
  ylab("Density") +  
  theme_minimal()
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_density()').
```

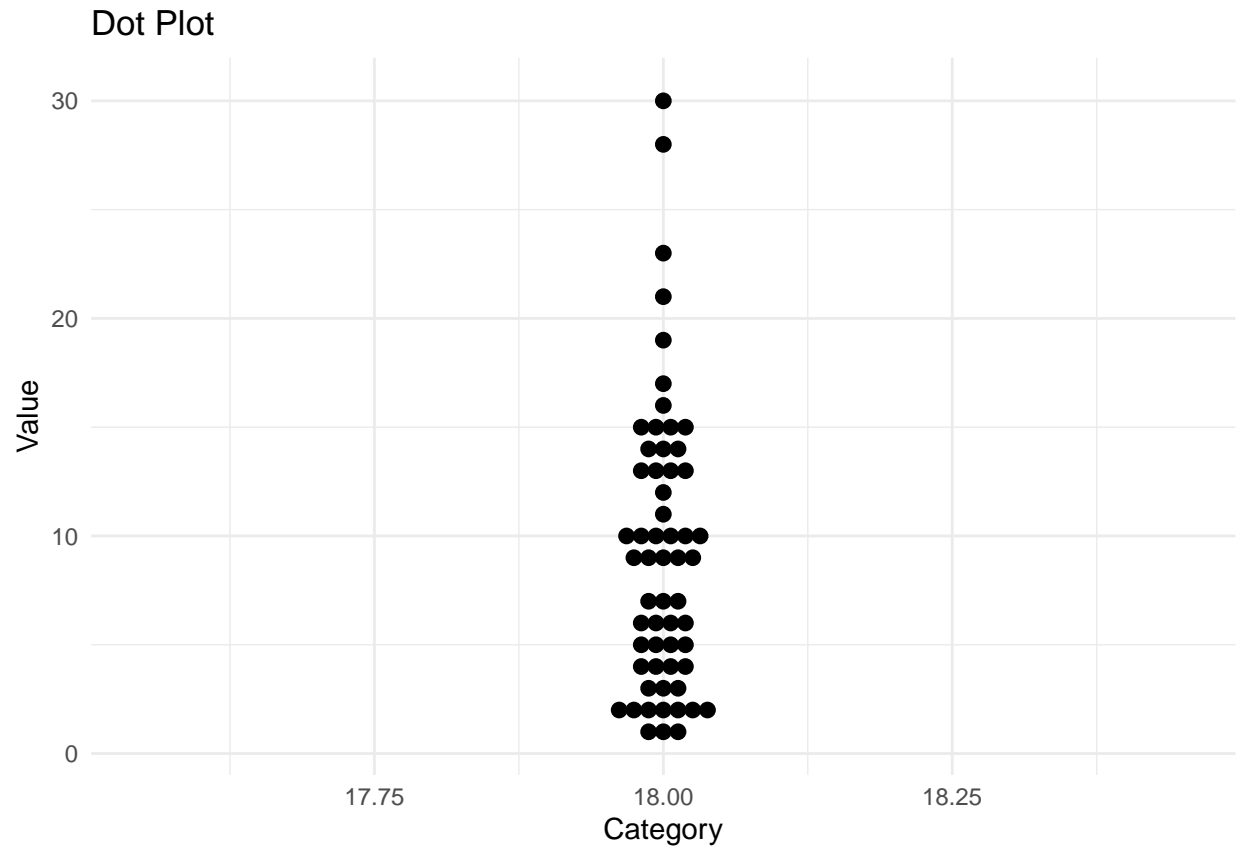


dot plot

```
ggplot(my_data, aes(x = YearsCode, y = YearsCodePro)) +  
  geom_dotplot(binaxis = "y", stackdir = "center", dotsize = 0.7, color = "black") +  
  ggtitle("Dot Plot") +  
  xlab("Category") +  
  ylab("Value") +  
  theme_minimal()
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with  
## 'binwidth'.
```

```
## Warning: Removed 1 rows containing missing values ('stat_bindot()').
```

Summarized Data Distributions

```
summary(my_data[, sapply(my_data, is.numeric)])
```

```
##      ResponseId      YearsCode      YearsCodePro      CompTotal
##  Min.   : 1.00      Min.   : 1.00      Min.   : 1.000      Min.    : 5555
## 1st Qu.:15.75      1st Qu.: 5.50      1st Qu.: 4.000      1st Qu.: 64500
## Median :30.50      Median :10.00      Median : 9.000      Median : 102500
## Mean   :30.50      Mean   :12.47      Mean   : 9.203      Mean   : 227580
## 3rd Qu.:45.25      3rd Qu.:17.00      3rd Qu.:13.000      3rd Qu.: 151500
## Max.   :60.00      Max.   :35.00      Max.   :30.000      Max.   :2000000
##      NA's      :1      NA's      :1      NA's      :16
##      WorkExp      CCY
##  Min.   : 2.00      Min.   : 14
## 1st Qu.: 5.00      1st Qu.: 53873
## Median : 8.00      Median : 76507
## Mean   :10.19      Mean   : 93960
## 3rd Qu.:13.25      3rd Qu.:122500
## Max.   :39.00      Max.   :285000
## NA's   :24      NA's   :16
```