FML_Assignment1_Urooj

2023-09-08

#1. The dataset used in this Assignment is PayEquity.

#2.The data is imported from https://github.com/davidcaughlin/R-Tutorial-Data-Files/blob/master/PayEquity.csv

```
#install.packages("readr")    #"readr" package is installed to read the files
library(readr)                #library is used to load the package
setwd("C:\\Users\\user\\Desktop\\Masters\\Assignments\\FML Assignments\\Urooj_FML_Assignment")
read.csv(file="PayEquity.csv",header=TRUE,sep=",") #raw data file is imported using the read.csv function
```

```
##       emp_id    job_title pay_grade    status base_pay job_tenure
## 1    ee20097 technician 1       e07 nonexempt    51179        4.7
## 2    ee20032 technician 1       e07 nonexempt    46039        2.0
## 3    ee20050 technician 1       e07 nonexempt    50414        3.2
## 4    ee20019 technician 2       e07 nonexempt    56442        4.5
## 5    ee20026 technician 2       e07 nonexempt    57211        8.1
## 6    ee20034 technician 2       e07 nonexempt    53129        7.0
## 7    ee20089 technician 1       e07 nonexempt    46826        3.8
## 8    ee20020 technician 2       e07 nonexempt    53705        5.0
## 9    ee20036 technician 2       e07 nonexempt    55382        3.3
## 10   ee20022 technician 2       e07 nonexempt    58551        5.7
## 11  ee200121 technician 1       e07 nonexempt    50823        4.2
## 12   ee20030 technician 3       e07 nonexempt    59767        2.3
## 13   ee20057 technician 1       e07 nonexempt    52726        7.0
## 14  ee200117 technician 1       e07 nonexempt    51462        7.8
## 15   ee20066 technician 1       e07 nonexempt    49228        0.9
## 16   ee20078 technician 3       e07 nonexempt    61627        6.4
## 17   ee20045 technician 2       e07 nonexempt    54347        7.5
## 18   ee20074 technician 1       e07 nonexempt    52078        3.8
## 19   ee20079 technician 3       e07 nonexempt    59031        7.2
## 20   ee20085 technician 3       e07 nonexempt    60688        4.2
## 21  ee200115 technician 1       e07 nonexempt    45978        2.4
## 22   ee20068 technician 1       e07 nonexempt    52167        5.8
## 23    ee2001 technician 1       e07 nonexempt    52016        6.9
## 24   ee20059 technician 1       e07 nonexempt    49496        5.1
## 25   ee20082 technician 3       e07 nonexempt    61932       10.0
## 26  ee200101 technician 1       e07 nonexempt    52909        4.1
## 27   ee20033 technician 1       e07 nonexempt    52878        6.3
## 28   ee20018 technician 2       e07 nonexempt    53695        4.8
## 29   ee20070 technician 1       e07 nonexempt    42084        0.2
## 30   ee20083 technician 2       e07 nonexempt    56414        6.1
## 31  ee200111 technician 2       e07 nonexempt    56216        4.6
## 32   ee20042 technician 3       e07 nonexempt    64095        5.5
## 33   ee20011 technician 3       e07 nonexempt    67753        7.4
## 34   ee20061 technician 2       e07 nonexempt    59509        3.9
## 35   ee20025 technician 1       e07 nonexempt    50519        5.4
## 36  ee200123 technician 2       e07 nonexempt    56670        2.6
## 37   ee20038 technician 2       e07 nonexempt    56970        6.0
## 38   ee20099 technician 2       e07 nonexempt    54384        3.1
## 39    ee2007 technician 2       e07 nonexempt    58828        6.9
## 40   ee20027 technician 2       e07 nonexempt    57004        5.6
## 41   ee20062 technician 1       e07 nonexempt    48234        6.9
## 42   ee20031 technician 2       e07 nonexempt    54431        2.8
## 43   ee20043 technician 2       e07 nonexempt    55689        5.6
## 44    ee2002 technician 2       e07 nonexempt    54331        4.4
## 45   ee20014 technician 1       e07 nonexempt    50086        7.4
## 46   ee20076 technician 2       e07 nonexempt    59909        9.8
## 47   ee20081 technician 1       e07 nonexempt    52635        5.2
## 48   ee20040 technician 1       e07 nonexempt    47598        5.6
## 49   ee20069 technician 1       e07 nonexempt    51417        5.6
## 50  ee200114 technician 1       e07 nonexempt    51006        3.4
## 51   ee20096 technician 1       e07 nonexempt    47615        2.8
## 52   ee20016 technician 3       e07 nonexempt    68335        7.2
## 53  ee200112 technician 1       e07 nonexempt    43873        1.5
## 54    ee2006 technician 3       e07 nonexempt    58607        2.7
## 55   ee20055 technician 2       e07 nonexempt    57380        6.3
## 56   ee20054 technician 2       e07 nonexempt    55159        3.7
## 57   ee20039 technician 2       e07 nonexempt    55087        3.0
## 58   ee20060 technician 1       e07 nonexempt    48632        2.0
## 59   ee20046 technician 3       e07 nonexempt    60503        7.0
## 60  ee200104 technician 1       e07 nonexempt    52751        3.3
## 61  ee200118 technician 1       e07 nonexempt    47352        1.6
## 62   ee20088 technician 2       e07 nonexempt    58200        2.8
## 63  ee200102 technician 3       e07 nonexempt    59442        5.8
## 64  ee200113 technician 3       e07 nonexempt    63956        5.7
## 65   ee20090 technician 2       e07 nonexempt    56096       10.1
## 66    ee2009 technician 2       e07 nonexempt    53413        3.7
## 67   ee20094 technician 3       e07 nonexempt    71166        6.4
## 68   ee20049 technician 2       e07 nonexempt    54204        5.8
## 69  ee200108 technician 3       e07 nonexempt    59964        6.1
## 70  ee200105 technician 1       e07 nonexempt    40000        1.4
## 71   ee20091 technician 1       e07 nonexempt    52786        3.5
## 72   ee20047 technician 2       e07 nonexempt    57329        6.3
## 73   ee20037 technician 1       e07 nonexempt    52504        6.4
## 74   ee20064 technician 1       e07 nonexempt    51863        5.4
## 75   ee20080 technician 1       e07 nonexempt    48203        7.6
## 76   ee20058 technician 1       e07 nonexempt    52089        5.4
## 77   ee20093 technician 1       e07 nonexempt    50716        4.4
## 78   ee20021 technician 1       e07 nonexempt    54170        2.8
## 79   ee20013 technician 2       e07 nonexempt    55523        3.7
## 80    ee2008 technician 2       e07 nonexempt    52729        5.0
## 81   ee20056 technician 1       e07 nonexempt    48805        4.7
## 82  ee200119 technician 1       e07 nonexempt    52565        5.6
## 83   ee20075 technician 3       e07 nonexempt    64392        7.2
## 84   ee20067 technician 1       e07 nonexempt    47928        4.6
## 85   ee20086 technician 2       e07 nonexempt    59392        7.3
## 86   ee20077 technician 3       e07 nonexempt    61966        9.5
## 87  ee200100 technician 1       e07 nonexempt    50532        3.9
## 88  ee200116 technician 3       e07 nonexempt    63208        5.1
## 89   ee20051 technician 1       e07 nonexempt    46026        2.2
## 90   ee20010 technician 1       e07 nonexempt    49983        6.1
## 91   ee20023 technician 1       e07 nonexempt    53091        7.8
## 92   ee20087 technician 1       e07 nonexempt    53268        3.2
## 93   ee20028 technician 1       e07 nonexempt    51527        3.4
## 94  ee200106 technician 2       e07 nonexempt    59403        6.4
## 95   ee20017 technician 2       e07 nonexempt    57692        7.4
## 96   ee20048 technician 3       e07 nonexempt    62452        6.5
## 97    ee2005 technician 1       e07 nonexempt    46194        2.5
## 98  ee200110 technician 1       e07 nonexempt    51210        4.1
## 99  ee200120 technician 3       e07 nonexempt    63756        4.3
## 100  ee20095 technician 2       e07 nonexempt    57234        7.8
## 101 ee200107 technician 3       e07 nonexempt    65288        4.2
## 102 ee200125 technician 2       e07 nonexempt    55685        6.7
## 103   ee2003 technician 1       e07 nonexempt    53776        0.3
## 104 ee200109 technician 2       e07 nonexempt    59141        7.8
## 105   ee2004 technician 1       e07 nonexempt    54257        5.0
## 106  ee20044 technician 1       e07 nonexempt    51242        7.0
## 107  ee20063 technician 3       e07 nonexempt    62811        8.2
## 108  ee20012 technician 2       e07 nonexempt    58524        4.3
## 109  ee20084 technician 3       e07 nonexempt    60352        9.7
## 110  ee20024 technician 2       e07 nonexempt    58876        6.0
## 111 ee200103 technician 2       e07 nonexempt    58623        3.8
## 112  ee20035 technician 2       e07 nonexempt    55168       11.4
## 113  ee20015 technician 3       e07 nonexempt    59837        5.2
## 114  ee20065 technician 1       e07 nonexempt    53365        5.0
## 115  ee20072 technician 3       e07 nonexempt    63586        5.2
## 116  ee20041 technician 2       e07 nonexempt    58376        5.8
## 117  ee20098 technician 2       e07 nonexempt    53144        6.2
## 118  ee20052 technician 2       e07 nonexempt    56782        7.2
## 119  ee20092 technician 3       e07 nonexempt    64215        3.3
## 120  ee20053 technician 1       e07 nonexempt    48814        1.5
## 121 ee200122 technician 1       e07 nonexempt    52649        2.6
## 122  ee20073 technician 3       e07 nonexempt    61561        8.5
## 123  ee20071 technician 2       e07 nonexempt    57811        7.2
## 124  ee20029 technician 1       e07 nonexempt    48050        1.3
## 125 ee200124 technician 2       e07 nonexempt    54924        9.5
##     pay_grade_tenure org_tenure  age educ perf    sex  race data_retrieval_date
## 1               11.0       13.4 44.8    1  9.4 female asian          2019-04-01
## 2                2.0        2.0 32.6    3  9.6 female black          2019-04-01
## 3                3.2        3.2 38.5    1  9.6 female asian          2019-04-01
## 4                6.6       10.9 31.5    4 10.0 female black          2019-04-01
## 5                9.0       10.1 34.7    4  3.4 female white          2019-04-01
## 6                7.8        7.8 36.4    2  8.6 female asian          2019-04-01
## 7                3.8        3.8 36.4    5  8.7 female white          2019-04-01
## 8                9.5        9.5 39.6    3  6.8 female asian          2019-04-01
## 9                3.3        3.3 47.8    5  3.8   male asian          2019-04-01
## 10               7.0        7.0 45.7    1  6.1 female white          2019-04-01
## 11               4.6        4.6 29.7    4  5.8 female black          2019-04-01
## 12              10.0       10.0 37.0    1  7.2   male white          2019-04-01
## 13               9.8        9.8 47.3    4  7.6 female black          2019-04-01
## 14               8.5       11.6 38.7    4  6.2 female black          2019-04-01
## 15               1.8        1.8 31.5    2  8.6 female black          2019-04-01
## 16              10.3       13.5 38.9    3 10.0   male black          2019-04-01
## 17              16.6       16.6 40.1    4  6.6 female black          2019-04-01
## 18               3.8        3.8 38.2    2  9.5 female black          2019-04-01
## 19              14.4       14.4 43.5    3 10.0   male black          2019-04-01
## 20               6.0       11.3 41.0    4  9.5   male asian          2019-04-01
## 21               2.4        2.4 36.4    3  9.1 female asian          2019-04-01
## 22               7.2        7.2 43.6    3  7.4 female black          2019-04-01
## 23               6.9        6.9 45.7    1  8.0 female black          2019-04-01
## 24               5.4        5.4 34.4    1  5.6 female asian          2019-04-01
## 25              12.2       13.8 45.9    3  9.2   male white          2019-04-01
## 26               4.1        4.1 33.0    2  6.6 female white          2019-04-01
## 27              13.4       13.4 42.4    5 10.0   male black          2019-04-01
## 28               7.5        7.5 37.7    3  9.8 female black          2019-04-01
## 29               8.9        8.9 31.2    2  9.5 female black          2019-04-01
## 30              11.1       11.1 40.7    4  7.8   male black          2019-04-01
## 31               8.6        8.6 34.0    1  8.5 female white          2019-04-01
## 32              12.7       12.7 37.2    3 10.0   male white          2019-04-01
## 33               7.4        7.4 39.1    3  6.6   male white          2019-04-01
## 34               7.3        7.3 41.0    3  8.3 female black          2019-04-01
## 35               9.3        9.3 40.8    4  5.4 female black          2019-04-01
## 36               2.6        2.6 40.2    4  3.0   male white          2019-04-01
## 37              14.3       14.3 36.5    3  8.0 female asian          2019-04-01
## 38              11.1       11.1 37.8    2 10.0   male white          2019-04-01
## 39              13.4       13.4 42.9    5 10.0   male white          2019-04-01
## 40               5.6        5.6 43.8    3  8.4   male black          2019-04-01
## 41              14.7       14.7 45.6    2  4.4 female asian          2019-04-01
## 42               4.0        5.1 42.4    4  6.4 female asian          2019-04-01
## 43              14.5       14.5 41.0    3  5.9   male white          2019-04-01
## 44               4.4        4.4 31.6    2  6.2   male black          2019-04-01
## 45              13.9       13.9 43.7    5  6.8 female black          2019-04-01
## 46              11.9       11.9 45.1    2  6.8   male black          2019-04-01
## 47               6.6        6.6 50.1    1  6.8 female asian          2019-04-01
## 48              10.1       10.1 40.5    4  3.8   male asian          2019-04-01
## 49              13.8       13.8 45.1    5  6.4 female black          2019-04-01
## 50               3.4        3.4 31.3    4  6.0 female asian          2019-04-01
## 51               7.1        7.1 41.1    2  7.9 female asian          2019-04-01
## 52              11.3       11.3 37.0    4  8.2   male black          2019-04-01
## 53               7.0        7.0 33.6    2  3.1 female white          2019-04-01
## 54               8.8        8.8 38.6    1  7.8 female white          2019-04-01
## 55               7.8        7.8 28.8    2  8.9 female white          2019-04-01
## 56               8.8       10.4 31.9    3  8.9   male white          2019-04-01
## 57               3.0        3.0 35.7    1  3.5   male black          2019-04-01
## 58               2.0        2.0 25.3    1 10.0 female black          2019-04-01
## 59               7.0        7.0 43.1    5 10.0   male white          2019-04-01
## 60               6.2        6.2 39.4    3  7.0   male black          2019-04-01
## 61               5.9        5.9 30.0    3  4.9 female black          2019-04-01
## 62               4.7        4.7 39.9    2  8.8 female white          2019-04-01
## 63               7.7        7.7 31.0    4 10.0   male black          2019-04-01
## 64              11.1       11.1 41.7    2  9.1   male white          2019-04-01
## 65              11.1       11.1 36.4    3  8.6 female black          2019-04-01
## 66              12.3       12.3 38.2    3  9.2   male black          2019-04-01
## 67              13.6       13.6 44.2    3  7.7   male black          2019-04-01
## 68              11.2       14.7 38.6    5  7.4 female asian          2019-04-01
## 69               8.4        8.4 45.5    2  8.1   male black          2019-04-01
## 70               8.6        8.6 37.7    1  4.1 female black          2019-04-01
## 71               5.1        5.1 45.4    4  7.8   male black          2019-04-01
## 72              12.9       12.9 35.8    5  5.8   male asian          2019-04-01
## 73              13.3       13.3 39.5    5  4.2   male black          2019-04-01
## 74              15.0       15.0 38.2    5  5.3 female white          2019-04-01
## 75               9.5        9.5 44.1    2  8.9 female asian          2019-04-01
## 76               6.2        6.2 40.8    5  7.4 female asian          2019-04-01
## 77               6.7        6.7 38.4    5  5.1 female black          2019-04-01
## 78               9.5        9.5 33.0    5  8.9 female black          2019-04-01
## 79              12.9       12.9 31.0    4 10.0   male white          2019-04-01
## 80               8.1        8.1 38.3    4  6.8 female black          2019-04-01
## 81               5.9        5.9 24.5    3  6.7 female white          2019-04-01
## 82               7.7        9.8 35.7    1  9.0 female asian          2019-04-01
## 83              16.2       16.2 37.0    4 10.0   male white          2019-04-01
## 84               9.4        9.4 39.0    1  6.2 female white          2019-04-01
## 85              11.6       11.6 49.1    1  5.8 female black          2019-04-01
## 86              18.1       18.1 44.1    1  6.4 female black          2019-04-01
## 87              12.5       12.5 35.0    2 10.0 female black          2019-04-01
## 88               5.1        5.1 35.4    3 10.0   male black          2019-04-01
## 89               2.2        2.2 32.6    2  5.7   male black          2019-04-01
## 90              15.8       15.8 30.6    4  7.6   male black          2019-04-01
## 91              15.4       15.4 43.0    3  4.2 female black          2019-04-01
## 92               7.2        7.2 41.3    2  5.6   male white          2019-04-01
## 93              13.0       13.0 43.6    3  7.3 female asian          2019-04-01
## 94               7.0        7.0 40.2    5  7.5   male black          2019-04-01
## 95              10.1       11.1 45.9    2 10.0   male black          2019-04-01
## 96              15.6       15.6 45.9    2 10.0   male black          2019-04-01
## 97               2.5        2.5 43.4    3  4.3 female asian          2019-04-01
## 98               7.3        7.3 37.5    4  8.3 female asian          2019-04-01
## 99               8.2        8.2 47.6    1  9.2 female black          2019-04-01
## 100             10.0       10.0 39.1    3  6.2   male asian          2019-04-01
## 101             10.9       10.9 44.0    3 10.0   male white          2019-04-01
## 102             13.6       13.6 41.8    3  7.7   male white          2019-04-01
## 103              4.2        4.2 40.6    2  7.4   male black          2019-04-01
## 104              8.9       12.3 35.6    2 10.0 female black          2019-04-01
## 105             10.0       12.1 39.6    3  5.7   male black          2019-04-01
## 106             10.6       10.6 38.0    1  9.3 female black          2019-04-01
## 107              8.2        8.2 40.3    5  7.0   male black          2019-04-01
## 108              5.1        5.1 42.0    2  8.3   male white          2019-04-01
## 109             13.7       13.7 40.7    5  7.3   male white          2019-04-01
## 110              9.8        9.8 53.4    5  9.2 female black          2019-04-01
## 111             10.4       10.4 39.9    3  5.7   male white          2019-04-01
## 112             13.5       13.5 40.4    4  6.6 female white          2019-04-01
## 113             13.0       13.0 48.1    3 10.0 female white          2019-04-01
## 114              9.2        9.2 40.6    1  6.0 female white          2019-04-01
## 115             14.6       14.6 38.4    3  9.7   male white          2019-04-01
## 116              7.8        7.8 45.5    3  8.9 female white          2019-04-01
## 117             14.1       14.1 38.8    5  8.0 female white          2019-04-01
## 118              7.2        7.2 46.1    4  6.6 female white          2019-04-01
## 119             10.5       10.5 35.7    2  8.1   male black          2019-04-01
## 120              6.3        6.3 31.7    2  6.1 female white          2019-04-01
## 121              7.6       12.5 31.2    3 10.0 female black          2019-04-01
## 122              8.5        8.5 40.0    3  5.9   male white          2019-04-01
## 123              9.7        9.7 32.6    3 10.0 female white          2019-04-01
## 124              3.3        4.0 30.4    1  7.3   male black          2019-04-01
## 125              9.5        9.5 43.5    5  5.5 female asian          2019-04-01
```

```
Data<-read.csv(file="PayEquity.csv",header=TRUE,sep=",") #The imported file is assigned to a Data frame "Data"
View(Data)                    #View function is used to display the data
head(Data,5)                  # Used to find the 5 rows from data
```

```
##    emp_id    job_title pay_grade    status base_pay job_tenure pay_grade_tenure
## 1 ee20097 technician 1       e07 nonexempt    51179        4.7             11.0
## 2 ee20032 technician 1       e07 nonexempt    46039        2.0              2.0
## 3 ee20050 technician 1       e07 nonexempt    50414        3.2              3.2
## 4 ee20019 technician 2       e07 nonexempt    56442        4.5              6.6
## 5 ee20026 technician 2       e07 nonexempt    57211        8.1              9.0
##   org_tenure  age educ perf    sex  race data_retrieval_date
## 1       13.4 44.8    1  9.4 female asian          2019-04-01
## 2        2.0 32.6    3  9.6 female black          2019-04-01
## 3        3.2 38.5    1  9.6 female asian          2019-04-01
## 4       10.9 31.5    4 10.0 female black          2019-04-01
## 5       10.1 34.7    4  3.4 female white          2019-04-01
```

#3.Descriptive statistics for a selection of quantitative and categorical variables

#3(a)Quantitative variable

```
#install.packages("modeest")
library(modeest)         #estimating mode of distribution
summary(Data$job_tenure) #provides with min, max,mean,median and quartiles of attribute
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.200   3.700   5.200   5.199   6.900  11.400
```

```
mean(Data$job_tenure)    #calculates the average
```

```
## [1] 5.1992
```

```
median(Data$job_tenure)  #calculates the mid value
```

```
## [1] 5.2
```

```
mfv(Data$job_tenure)     #most frequent value function is used to calculate mode
```

```
## [1] 5.6 7.2
```

```
table(Data$job_tenure)   # table function shows all the values with its frequency
```

```
##
##  0.2  0.3  0.9  1.3  1.4  1.5  1.6    2  2.2  2.3  2.4  2.5  2.6  2.7  2.8    3
##    1    1    1    1    1    2    1    2    1    1    1    1    2    1    4    1
##  3.1  3.2  3.3  3.4  3.5  3.7  3.8  3.9  4.1  4.2  4.3  4.4  4.5  4.6  4.7  4.8
##    1    2    3    2    1    3    3    2    2    3    2    2    1    2    2    1
##    5  5.1  5.2  5.4  5.5  5.6  5.7  5.8    6  6.1  6.2  6.3  6.4  6.5  6.7  6.9
##    4    2    3    3    1    5    2    4    2    3    1    3    4    1    1    3
##    7  7.2  7.3  7.4  7.5  7.6  7.8  8.1  8.2  8.5  9.5  9.7  9.8   10 10.1 11.4
##    4    5    1    3    1    1    4    1    1    1    2    1    1    1    1    1
```

```
sd(Data$job_tenure)      # calculates the standard deviation
```

```
## [1] 2.216088
```

```
var(Data$job_tenure)     # calculates the Variance
```

```
## [1] 4.911048
```

```
length(Data$job_tenure)  # calculates the number of rows in column
```

```
## [1] 125
```

```
quantile(Data$job_tenure)# calculates the quantile
```

```
##   0%  25%  50%  75% 100%
##  0.2  3.7  5.2  6.9 11.4
```

```
range(Data$job_tenure)   # range calculates the min and max values
```

```
## [1]  0.2 11.4
```

```
sum(Data$job_tenure)     # calculates the sum of column values
```

```
## [1] 649.9
```

```
min(Data$job_tenure)     # calculates the minimum value in column
```

```
## [1] 0.2
```

```
max(Data$job_tenure)     # calculates the maximum value in column
```

```
## [1] 11.4
```

```
sort(Data$job_tenure)    # ascending order
```

```
##   [1]  0.2  0.3  0.9  1.3  1.4  1.5  1.5  1.6  2.0  2.0  2.2  2.3  2.4  2.5  2.6
##  [16]  2.6  2.7  2.8  2.8  2.8  2.8  3.0  3.1  3.2  3.2  3.3  3.3  3.3  3.4  3.4
##  [31]  3.5  3.7  3.7  3.7  3.8  3.8  3.8  3.9  3.9  4.1  4.1  4.2  4.2  4.2  4.3
##  [46]  4.3  4.4  4.4  4.5  4.6  4.6  4.7  4.7  4.8  5.0  5.0  5.0  5.0  5.1  5.1
##  [61]  5.2  5.2  5.2  5.4  5.4  5.4  5.5  5.6  5.6  5.6  5.6  5.6  5.7  5.7  5.8
##  [76]  5.8  5.8  5.8  6.0  6.0  6.1  6.1  6.1  6.2  6.3  6.3  6.3  6.4  6.4  6.4
##  [91]  6.4  6.5  6.7  6.9  6.9  6.9  7.0  7.0  7.0  7.0  7.2  7.2  7.2  7.2  7.2
## [106]  7.3  7.4  7.4  7.4  7.5  7.6  7.8  7.8  7.8  7.8  8.1  8.2  8.5  9.5  9.5
## [121]  9.7  9.8 10.0 10.1 11.4
```

```
sort(Data$job_tenure, decreasing=TRUE)# descending order
```

```
##   [1] 11.4 10.1 10.0  9.8  9.7  9.5  9.5  8.5  8.2  8.1  7.8  7.8  7.8  7.8  7.6
##  [16]  7.5  7.4  7.4  7.4  7.3  7.2  7.2  7.2  7.2  7.2  7.0  7.0  7.0  7.0  6.9
##  [31]  6.9  6.9  6.7  6.5  6.4  6.4  6.4  6.4  6.3  6.3  6.3  6.2  6.1  6.1  6.1
##  [46]  6.0  6.0  5.8  5.8  5.8  5.8  5.7  5.7  5.6  5.6  5.6  5.6  5.6  5.5  5.4
##  [61]  5.4  5.4  5.2  5.2  5.2  5.1  5.1  5.0  5.0  5.0  5.0  4.8  4.7  4.7  4.6
##  [76]  4.6  4.5  4.4  4.4  4.3  4.3  4.2  4.2  4.2  4.1  4.1  3.9  3.9  3.8  3.8
##  [91]  3.8  3.7  3.7  3.7  3.5  3.4  3.4  3.3  3.3  3.3  3.2  3.2  3.1  3.0  2.8
## [106]  2.8  2.8  2.8  2.7  2.6  2.6  2.5  2.4  2.3  2.2  2.0  2.0  1.6  1.5  1.5
## [121]  1.4  1.3  0.9  0.3  0.2
```

```
IQR(Data$job_tenure)     # calculates the InterQuartile Range
```

```
## [1] 3.2
```

```
xtabs(~sex+race,data=Data)# Display value in cross tab
```

```
##         race
## sex      asian black white
##   female    20    32    20
##   male       5    27    21
```

#3(b)Categorical variable

```
table(Data$race)  # table function is used to count the frequency of each categorical value
```

```
##
## asian black white
##    25    59    41
```

```
prop.table(table(Data$race))*100 #calculates the proportion of different categories
```

```
##
## asian black white
##  20.0  47.2  32.8
```

```
mfv(Data$race)    # most frequent value function is used to find the highest frequent categorical value in a column
```

```
## [1] "black"
```

```
unique(Data$race) # Unique function is used to  find the unique values in a column by removing the duplicates
```

```
## [1] "asian" "black" "white"
```

#4 Transforming a Quantitative Variable-Zscore, Square root

```
#install.packages("ggplot2") # ggplot2 package is used to visualize data and plot the charts
library(ggplot2)            # ggplot2 is loaded to use for graphs
#install.packages("gridExtra")#gridExtra is used as an extension for grids
library(gridExtra)           # gridExtra is loaded for grids

#z-score
z=(Data$pay_grade_tenure-mean(Data$pay_grade_tenure))/sd(Data$pay_grade_tenure) #calculating the z score
z # The variable will give the output of zscore formula assigned to z
```

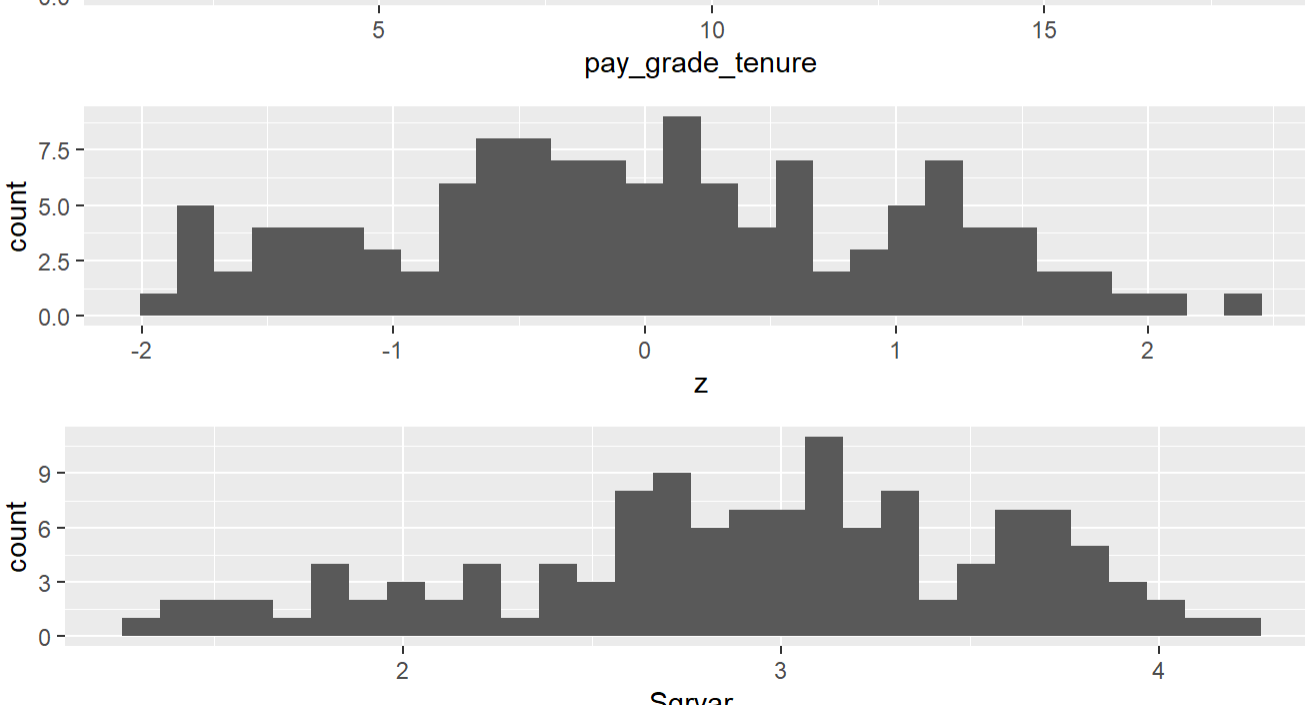
```
##   [1]  0.536500306 -1.844418211 -1.526962409 -0.627504302  0.007407302
##   [6] -0.310048500 -1.368234508  0.139680553 -1.500507759 -0.521685702
##  [11] -1.156597306  0.271953804  0.219044504 -0.124865949 -1.897327512
##  [16]  0.351317755  2.017960716 -1.368234508  1.435958412 -0.786232204
##  [21] -1.738599610 -0.468776401 -0.548140352 -0.944960105  0.853956108
##  [26] -1.288870557  1.171411910 -0.389412451 -0.019047348  0.562954956
##  [31] -0.098411299  0.986229359 -0.415867101 -0.442321751  0.086771253
##  [36] -1.685690310  1.409503762  0.562954956  1.171411910 -0.892050804
##  [41]  1.515322363 -1.315325207  1.462413062 -1.209506607  1.303685161
##  [46]  0.774592158 -0.627504302  0.298408454  1.277230511 -1.474053109
##  [51] -0.495231052  0.615864256 -0.521685702 -0.045501998 -0.310048500
##  [56] -0.045501998 -1.579871709 -1.844418211 -0.521685702 -0.733322903
##  [61] -0.812686854 -1.130142656 -0.336503150  0.562954956  0.562954956
##  [66]  0.880410758  1.224321211  0.589409606 -0.151320599 -0.098411299
##  [71] -1.024324055  1.039138659  1.144957260  1.594686313  0.139680553
##  [76] -0.733322903 -0.601049652  0.139680553  1.039138659 -0.230684550
##  [81] -0.812686854 -0.336503150  1.912142116  0.113225903  0.695228207
##  [86]  2.414780469  0.933320059 -1.024324055 -1.791508911  1.806323515
##  [91]  1.700504914 -0.468776401  1.065593310 -0.521685702  0.298408454
##  [96]  1.753414215 -1.712144960 -0.442321751 -0.204229899  0.271953804
## [101]  0.510045656  1.224321211 -1.262415907 -0.019047348  0.271953804
## [106]  0.430681705 -0.204229899 -1.024324055  1.250775861  0.219044504
## [111]  0.377772405  1.197866561  1.065593310  0.060316602  1.488867713
## [116] -0.310048500  1.356594462 -0.468776401  0.404227055 -0.706868253
## [121] -0.362957801 -0.124865949  0.192589853 -1.500507759  0.139680553
```

```
#Square root
Sqrvar<-sqrt(Data$pay_grade_tenure) # sqrt function applies square root to each value in the column
Sqrvar # The variable will give the output of square root formula assigned to Sqrvar
```

```
##   [1] 3.316625 1.414214 1.788854 2.569047 3.000000 2.792848 1.949359 3.082207
##   [9] 1.816590 2.645751 2.144761 3.162278 3.130495 2.915476 1.341641 3.209361
##  [17] 4.074310 1.949359 3.794733 2.449490 1.549193 2.683282 2.626785 2.323790
##  [25] 3.492850 2.024846 3.660601 2.738613 2.983287 3.331666 2.932576 3.563706
##  [33] 2.720294 2.701851 3.049590 1.612452 3.781534 3.331666 3.660601 2.366432
##  [41] 3.834058 2.000000 3.807887 2.097618 3.728270 3.449638 2.569047 3.178050
##  [49] 3.714835 1.843909 2.664583 3.361547 2.645751 2.966479 2.792848 2.966479
##  [57] 1.732051 1.414214 2.645751 2.489980 2.428992 2.167948 2.774887 3.331666
##  [65] 3.331666 3.507136 3.687818 3.346640 2.898275 2.932576 2.258318 3.591657
##  [73] 3.646917 3.872983 3.082207 2.489980 2.588436 3.082207 3.591657 2.846050
##  [81] 2.428992 2.774887 4.024922 3.065942 3.405877 4.254409 3.535534 2.258318
##  [89] 1.483240 3.974921 3.924283 2.683282 3.605551 2.645751 3.178050 3.949684
##  [97] 1.581139 2.701851 2.863564 3.162278 3.301515 3.687818 2.049390 2.983287
## [105] 3.162278 3.255764 2.863564 2.258318 3.701351 3.130495 3.224903 3.674235
## [113] 3.605551 3.033150 3.820995 2.792848 3.754997 2.683282 3.240370 2.509980
## [121] 2.756810 2.915476 3.114482 1.816590 3.082207
```

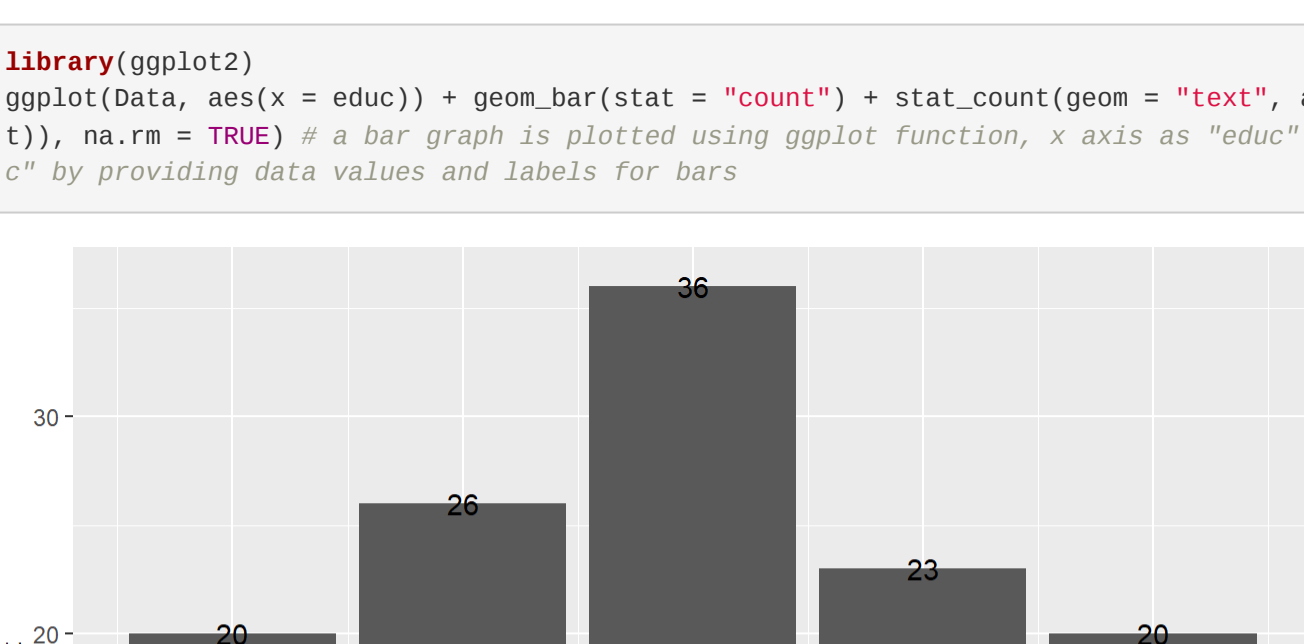
```
# to plot a histogram with pay_grade_tenure column ggplot is used and assigned to Var1 variable
Var1<- ggplot(data= Data, aes(x= pay_grade_tenure)) + geom_histogram()

# to plot a histogram with zscore (Z column) ggplot is used and assigned to Var2 variable
Var2<- ggplot(data= Data, aes(x= z)) + geom_histogram()

# to plot a histogram with square root (Sqrvar column) ggplot is used and assigned to Var3 variable
Var3<- ggplot(data= Data, aes(x= Sqrvar)) + geom_histogram()

# grid arrange is used to assemble the multiple plots in a page either in rows or columns specified
grid.arrange(Var1, Var2, Var3, ncol=1)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

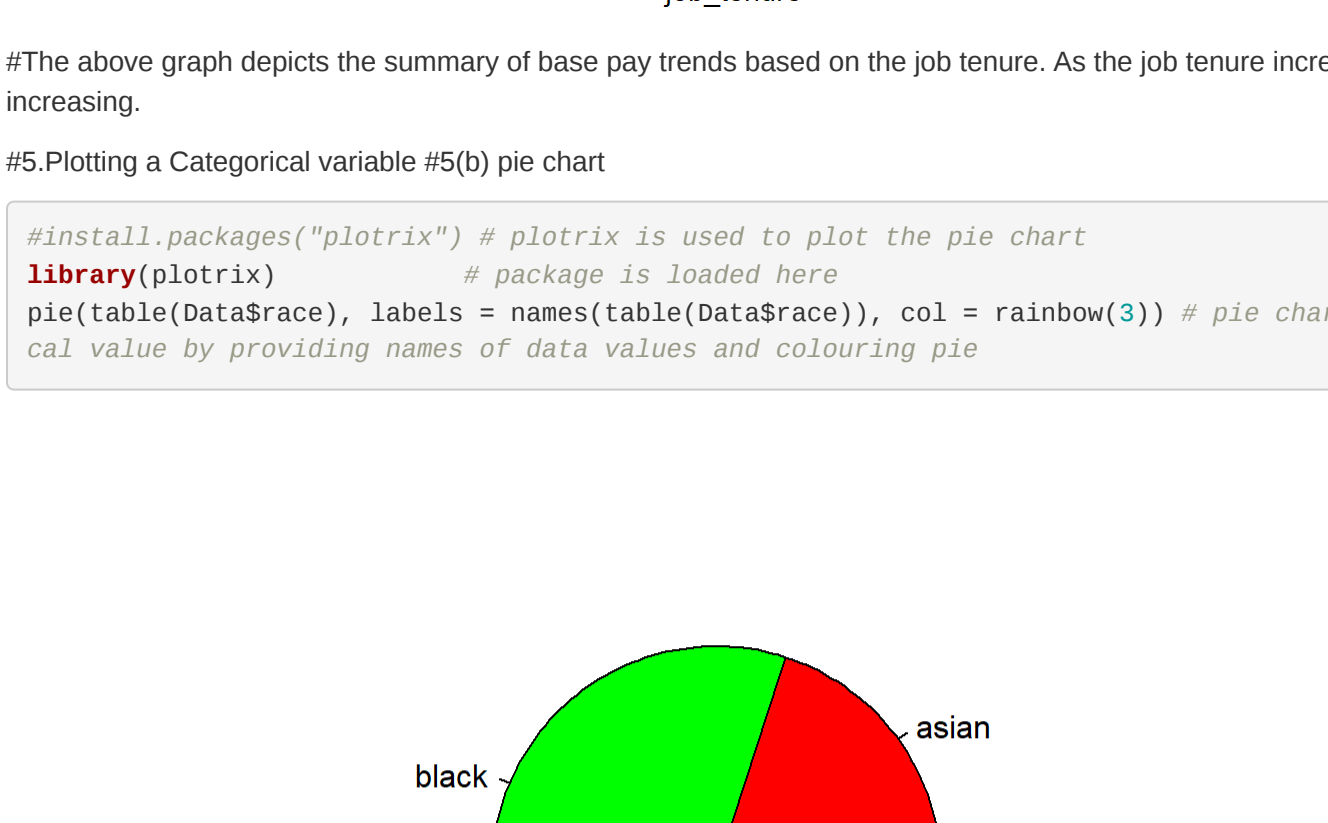


#5.Plotting a quantitative variable, and one scatterplot #5(a)Plotting a quantitative variable

```
library(ggplot2)
ggplot(Data, aes(x = educ)) + geom_bar(stat = "count") + stat_count(geom = "text", aes(label = after_stat(count)), na.rm = TRUE) # a bar graph is plotted using ggplot function, x axis as "educ" column and y as count of "educ" by providing data values and labels for bars
```



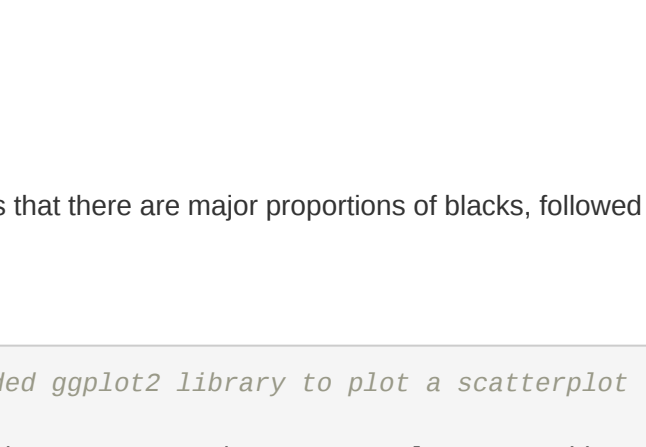

```
ggplot(data = Data, aes(x = job_tenure, y = base_pay)) + geom_line()
```



#The above graph depicts the summary of base pay trends based on the job tenure. As the job tenure increases, the base pay trends are increasing.

#5.Plotting a Categorical variable #5(b) pie chart

```
#install.packages("plotrix") # plotrix is used to plot the pie chart
library(plotrix)            # package is loaded here
pie(table(Data$race), labels = names(table(Data$race)), col = rainbow(3)) # pie chart is drawn with race categorical value by providing names of data values and colouring pie
```



#The above piechart illustrates that there are major proportions of blacks, followed with white people. Asians are the lower population compare to other two groups.

#5(c) Scatter plot

```
library(ggplot2) # loaded ggplot2 library to plot a scatterplot

ggplot(Data, aes(x = job_tenure, y = base_pay, color = race)) + geom_point() #scatter plot is plotted using job_tenure as x, base_pay as y and coloured data points from darker to lighter for higher to lower tenure
```

#The above graph illustrates that the majority of the data points lie between the tenure of 3-7 years of job tenure and their base pay is between 45,000 to 65,000. #There is a high proportions of white and black followed with asian people. #On the other hand, there are very few proportion of audience between 0-3 years and above 9 years and they are whites and blacks.