Assignment Report

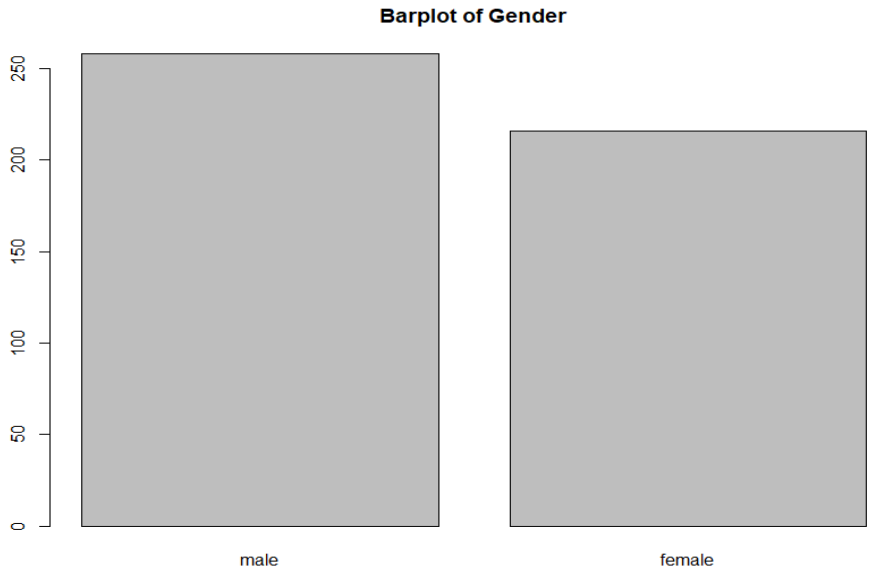
Legkas Sotiris

1.

2. There are 258 men according to the frequency table. The most appropriate graph for a categorical variable like gender is a bar plot.

total N=474 valid N=474 mean=1.46 sd=0.50

Value	N	Raw %	Valid %	Cum. %
male	258	54.43	54.43	54.43
female	216	45.57	45.57	100.00

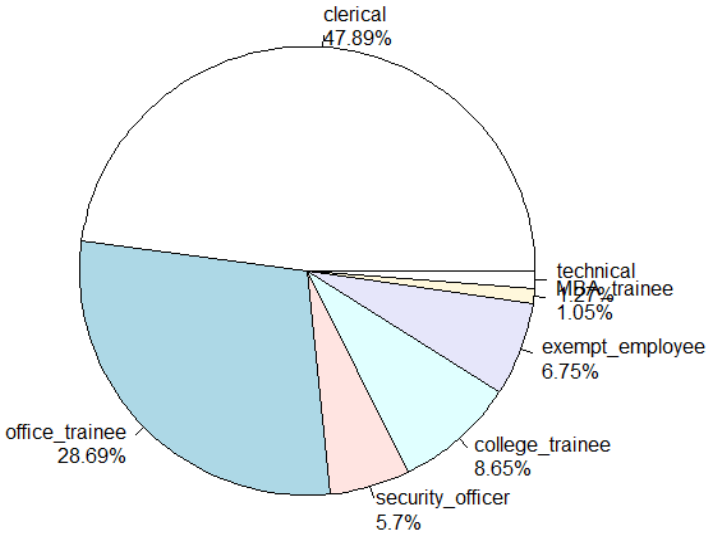


3. The most common job according to frequency table is clerical with a percentage of 47.89%

total N=474 valid N=474 mean=2.06 sd=1.41

Value	N	Raw %	Valid %	Cum. %
clerical	227	47.89	47.89	47.89
office_trainee	136	28.69	28.69	76.58
security_officer	27	5.70	5.70	82.28
college_trainee	41	8.65	8.65	90.93
exempt_employee	32	6.75	6.75	97.68
MBA_trainee	5	1.05	1.05	98.73
technical	6	1.27	1.27	100.00

4.



5. The percentage of people who are either clerical workers or security officers is 53.58%.

6.

total N=474 valid N=474 mean=2.38 sd=1.72

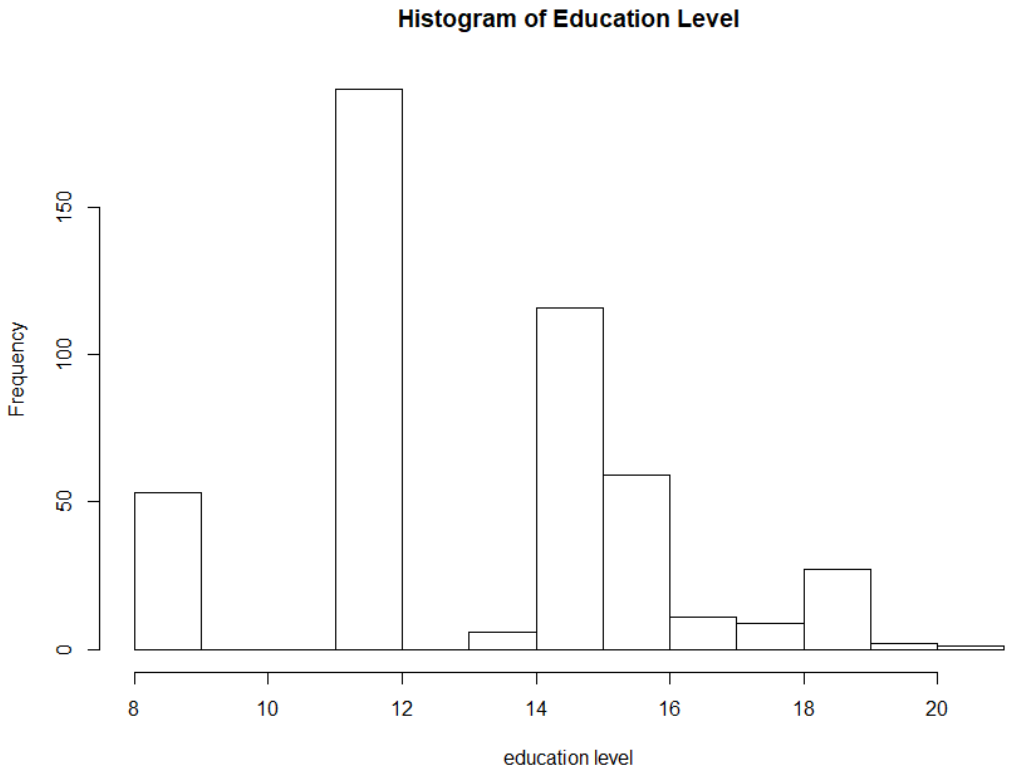
Value	N	Raw %	Valid %	Cum. %
[6300,10275]	160	33.76	33.76	33.76
(10275,14250]	186	39.24	39.24	73.00
(14250,18225]	47	9.92	9.92	82.91
(18225,22200]	25	5.27	5.27	88.19
(22200,26175]	20	4.22	4.22	92.41
(26175,30150]	19	4.01	4.01	96.41
(30150,34125]	7	1.48	1.48	97.89
(34125,38100]	4	0.84	0.84	98.73
(38100,42075]	4	0.84	0.84	99.58
(42075,46050]	1	0.21	0.21	99.79
(46050,50025]	0	0.00	0.00	99.79
(50025,54000]	1	0.21	0.21	100.00

7. The highest salary is 54000 and only one person makes that salary.

8. The mean salary is 13767.83 and the median is 11550.

9. The 10% of the highest paid employees has a salary of at least 23575.

10. The shape of the distribution cannot be described precisely from the histogram. The frequency of education level of 9, 10 ,11 and 13 is zero, which makes the distribution difficult to describe. The findings of the descriptive statistics are also indicating this complexity. The mean is 13.49, which is higher than the median which is 12. This indicates that the distribution is right skewed. On the contrary, the skew is slightly negative, which means that the distribution is left skewed. The standard deviation equals 2.88. Also, the maximum education level is 21 and the minimum 8.



vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	474	13.49	2.88	12	13.54	4.45	8	21	13	-0.11	-0.29	0.13

11.

> sapply(numeric_data, mean)

salbeg	time	age	salnow	edlevel	work
6806.43459	81.109705	37.186139	13767.8270	13.491561	7.988608

> sapply(numeric_data, sd)

salbeg	time	age	salnow	edlevel	work
3148.25526	10.060945	11.787242	6830.26458	2.884846	8.715411

> sapply(numeric_data, range)

	salbeg	time	age	salnow	edlevel	work
[1,]	3600	63	23	6300	8	0
[2,]	31992	98	64.5	54000	21	39.67

12.

	Mean	sd	min	max	range
salbeg	0	1	-1.02	8.00	9.02
time	0	1	-1.80	1.68	3.48
age	0	1	-1.20	2.32	3.52
salnow	0	1	-1.09	5.89	6.98
edlevel	0	1	-1.90	2.60	4.51
work	0	1	-0.92	3.64	4.55

13. The average standardized score for education level for females only is -0.3886.

14. The mean, median and standard deviation of the variable “raise” are 6961.39, 5700, 4325.84 respectively.

15. The greatest increase in salary is 30496. The person that had this increase had the ID of 149.

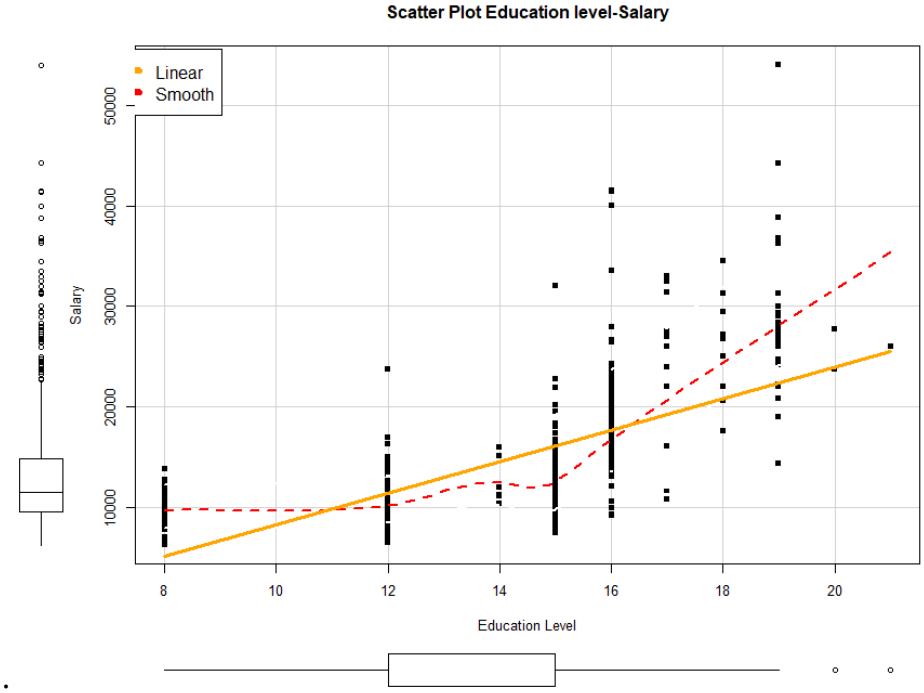
16.

0 = $0-$4,999 1 = $5,000 - $9,999 2 = $10,000 - $14,999 3 = $15,000 - $19,999 4 = $20,000 - $24,999

Value | N | Raw % | Valid % | Cum. %
---
0 | 119 | 25.11 | 25.11 | 25.11
1 | 294 | 62.03 | 62.03 | 87.13
2 | 49 | 10.34 | 10.34 | 97.47
3 | 9 | 1.90 | 1.90 | 99.37
4 | 2 | 0.42 | 0.42 | 99.79
5 | 1 | 0.21 | 0.21 | 100.00

17. 49 (10.34%) people had a salary between of 10000 and 14999.

18.



.

19-20. The relationship between salary and education level is positive, however, it does not seem to be linear. This means that people tend to earn a higher salary when their education level is higher. However, the increase does not seem to be linear, as people with education level between 8 and 12 have almost similar salary. People with an education level above 15 seem to have a linear relationship between their education and salary.