

Analytics of Business Intelligence - Fall 2019, Practice Set # 4

1. The table below is named `dt`, write the code for the correct categorical column.

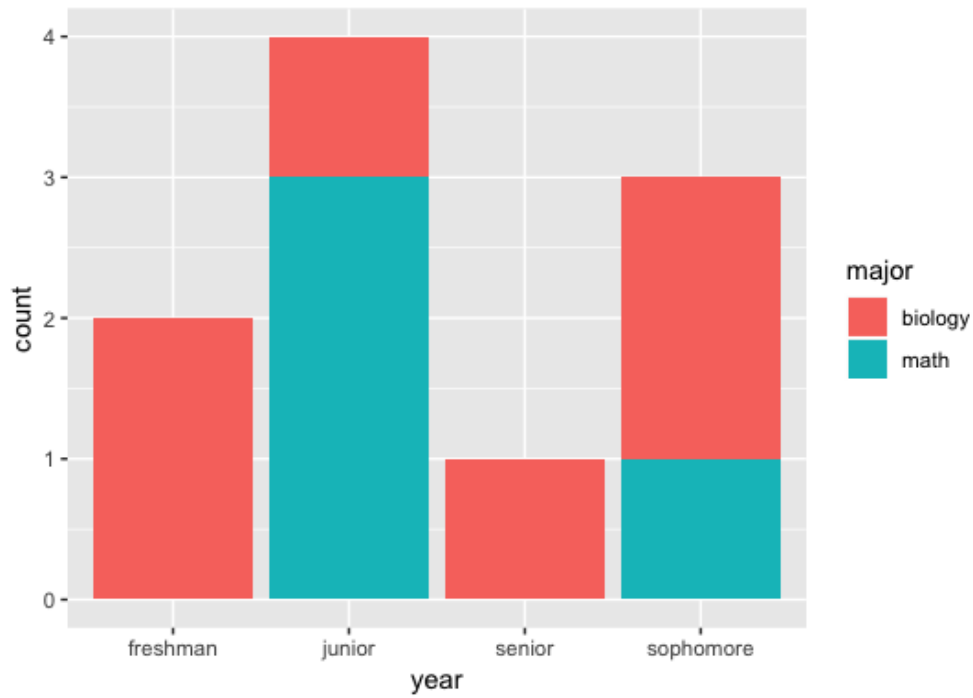
```
Classes 'data.table' and 'data.frame': 172 obs. of 7 variables:
 $ google_adwords : num  65.7 39.1 174.8 34.4 78.2 ...
 $ facebook       : num  47.9 55.2 52 62 40.9 ...
 $ twitter        : num  52.5 77.4 68 86.9 30.4 ...
 $ marketing_total: num  166 172 295 183 150 ...
 $ revenues       : num  39.3 38.9 49.5 40.6 40.2 ...
 $ employees      : int   5 7 11 7 9 3 10 6 6 4 ...
 $ pop_density    : chr   "High" "Medium" "Medium" "High" ...
- attr(*, ".internal.selfref")=<externalptr>
```

2. Which data is the closest to being normally distributed?

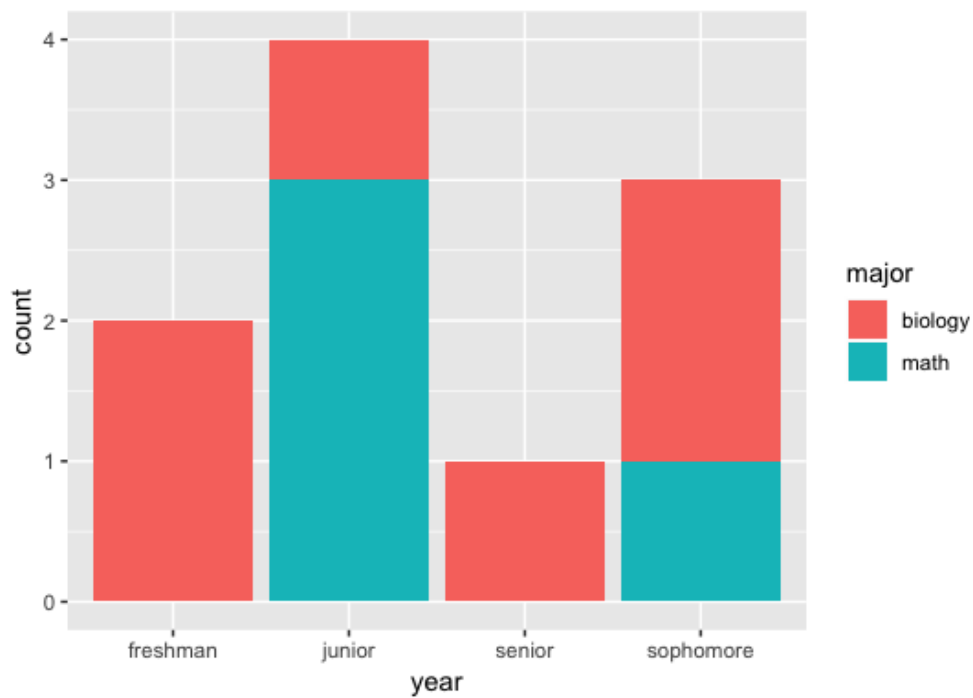
A.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	1.00	25.75	49.00	49.01	75.00	100.00
B.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	2.00	30.50	55.50	53.49	82.00	100.00
C.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	1.00	26.75	50.00	49.85	75.25	100.00
D.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	7.00	24.75	59.50	203.17	74.75	1000.00

3. What graph would you use to visualize the relationship between numerical and categorical data?

4. Using the graph below, how many math majors are there?



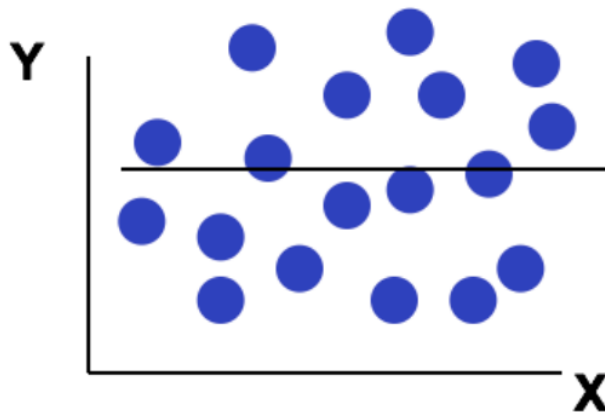
5. Using the graph below, how many sophomores are biology majors?



6. Using the graph in #5 and the table below, how would you show the same graph split by sex?

	year	major	sex
1	junior	math	Female
2	sophomore	biology	Female
3	junior	math	Female
4	senior	biology	Male
5	junior	biology	Female
6	freshman	biology	Male
7	sophomore	biology	Female
8	freshman	biology	Male
9	sophomore	math	Male
10	junior	math	Male

7. What is correlation of the below graph?



8. Which L.I.N.E. assumption is violated, if the scatterplot of the independent and dependent variables have a curved line?
9. Which L.I.N.E. assumption is violated, if the histogram of the residuals are skewed to the side?
10. Which L.I.N.E. assumption is violated, if the scatterplot of the residuals and residual Z-scores are curved to the side?
11. Which L.I.N.E. assumption is violated, if the scatterplot of the residuals and predicted values show a pattern?

12. A linear regression model predicting the linear effects of the amount of water used in growing almonds is developed below:

$$y = 56 + 10x$$

The water data has a range of data from 5 to 90, and is in units of 100 gallons, and the almond data is in pounds. According to the model, how much pounds of almonds will grow, if no water is used?

13. Using question # 12, how many acres will grow if 1000 gallons is used?
14. Using question # 12, how many acres will grow if 400 gallons is used?
15. Using question # 12, an increase of 100 gallons, increases almond production by how many pounds?
16. A linear regression model predicting the linear effects of the amount of water used in growing almonds is developed below:

$$y = 56 + 10x$$

The water coefficient has a t-statistic of 5.7 with a p-value of .000000000000002. Do you reject the null hypothesis for the water variable?

17. A linear regression model predicting the linear effects of the amount of water and the amount of sunlight used in growing almonds is developed below:

$$y = 56 + 10x$$

The data had 200 rows of data. The water coefficient has a t-statistic of 5.7 with a p-value of .000000000000002, the sunlight coefficient has a t-statistic of 12.6 with a p-value of .68. The model has a F-statistic of 5. Do you reject the null hypothesis for the model?

18. Using question #17, can you improve the model?
19. Using questions #17, if the adjusted R-square ratio was .05, what does that mean?
20. What is M3[,sum(3)]?

```
1 M3=data.table(x=1:5,y=seq(2,10,2),z=7:3)
```

21. What is sum(M2[,2])?

```
1 M3=data.table(x=1:5,y=seq(2,10,2),z=7:3)
```

22. What is sum(M2+rep(2,5))?

```
1 M3=data.table(x=1:5,y=seq(2,10,2),z=7:3)
```

23. What is M3[y<6,x]?

```
1 M3=data.table(x=1:5,y=seq(2,10,2),z=7:3)
```

24. What is `sum(M3[,2]+M3[,x])`?

```
1 M3=data.table(x=1:5,y=seq(2,10,2),z=7:3)
```

25. What is `sum(sum(giants)*eagles)`?

```
1 giants=c(2:5)
2 cowboys=rep(3,4)
3 eagles=c(6:giants[cowboys[2]])
```

26. What is `sum(giants + c(eagles,cowboys[giants[1]]))`?

```
1 giants=c(2:5)
2 cowboys=rep(3,4)
3 eagles=c(6:giants[cowboys[2]])
```

27. What is `a`?

```
1 a = 6:1
2 a[c(2,4)]=c(10,20)
```

28. What is `a`?

```
1 a = 6:1
2 a[2:4]=rep(2,3)
```

29. What is `mm`?

```
1 mm = c(1:9)
2 mm[2:4]=c(12,13)
```

30. What is `mm`?

```
1 mm = c(9:1)
2 mm[c(7,3:2)]=100:1
```