EDAV Test #1

October 26, 2020

Instructions

- You have 75 minutes to complete this test. You are NOT permitted any outside material or assistance to complete this exam.
- On your bubble sheet, if you haven't already, fill in your first name, last name, UNI, and ZIP ID. Then bubble in the numbers of your ZIP ID.
- We will have limited ability to answer questions during the exam. If you have an urgent matter you may send me a private message in the Zoom chat. For clarification on exam questions, explain the issue on the back of the bubble sheet and we will take it into account while grading. Indicate the question number clearly.

Start and End Times

Students with electronic version of test

5:30pm Start

6:35pm "10 minutes left" announcement

6:45pm End (pencils down)

6:55pm Scan and upload deadline

Students with printed version of test

5:40pm Start (wait until you are instructed to begin)

6:45pm "10 minutes left" announcement

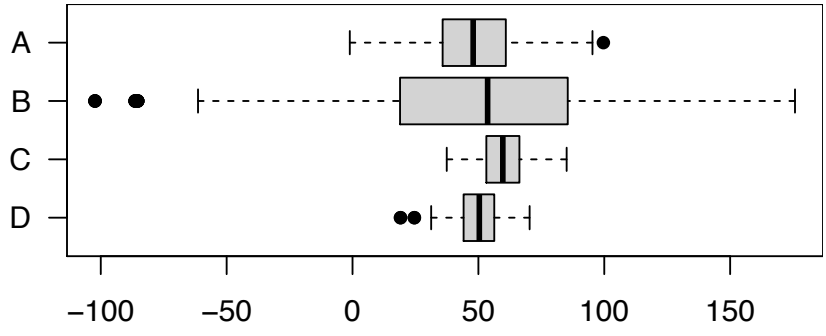
6:55pm End (pencils down)

7:05pm Scan and upload deadline

1. Which of the following will aid in identifying whether the data displayed in a histogram is rounded to the nearest `0.5`?

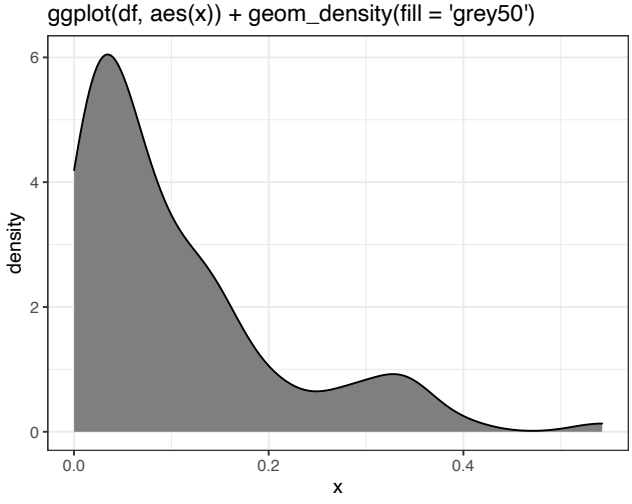
a) Set the binwidth to `0.25`
b) Set the binwidth to `0.5`
c) Set the binwidth to `1`
d) None of the above

2. Which of the following can be deduced from the graph below?



a) Group B has more data values than Group A
b) The mean of Group B is higher than the mean of Group A
c) Group C is bimodal
d) None of the above

3. What is the approximate area of the shaded portion of the graph below?

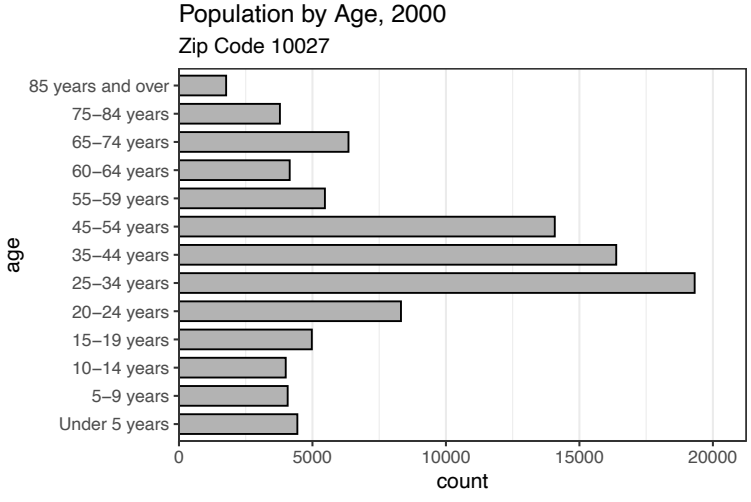


a) 4
b) 5.5
c) 2
d) None of the above

4. In a mosaic plot, the dependent variable should be the _____ cut.

   a) next to last
   b) last
   c) second
   d) first

5. Based on the bar chart below, what is your best estimate of the ratio of 27 year olds to 21 year olds in 2010?



   a) 2:1
   b) 4:1
   c) 1:1
   d) 8:1

6. Suppose you have the following data for fifty movies: Critics rating (1-100), Audience rating (1-100), and Box Office Sales (in $). Which type of graph is best for check if there are any movies that top the charts in all three indicators?

   a) ridgeline plots
   b) mosaic plot
   c) scatterplot matrix
   d) parallel coordinate plots

7. Which of the following is/are true?

a) Topcoding is acceptable if the number of values in the top category or bin is small.
b) Topcoding is acceptable if the number of values in the top category or bin is large.
c) Topcoding is discouraged due to privacy concerns.
d) None of the above

8. Ridgeline plots show _____ better than multiple boxplots.

a) outliers
b) modes
c) means
d) medians

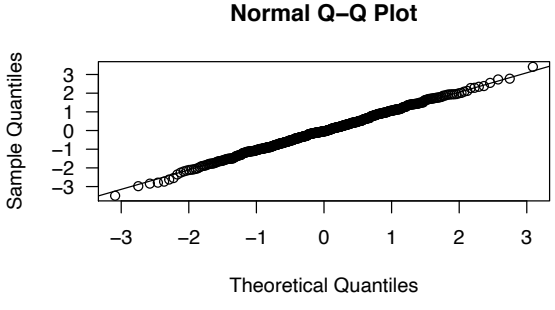
9. How should the axes in a parallel coordinate plot be ordered?

a) by frequency of observations
b) by experimentation for best view
c) by order of factor levels
d) by decreasing median

10. A diverging stacked bar chart is designed to be used with

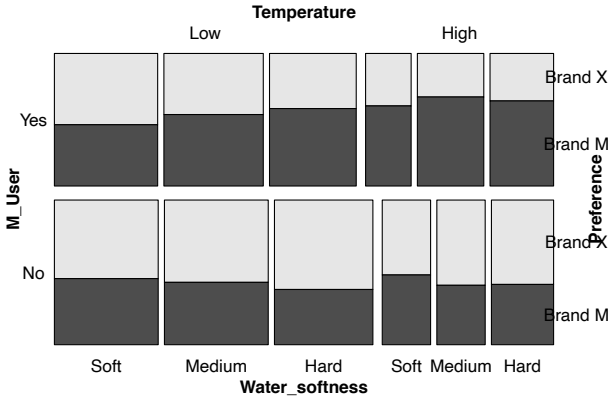
a) time series data
b) nominal categorical data with great disparities in frequency counts
c) continuous data with many outliers
d) ordinal categorical data in which the first and last categories are polar opposites

11. What can be concluded from plot below?



a) The data appear to be uniformly distributed
b) The data appear to be positively skewed
c) The data appear to be bimodal
d) The data appear to be normally distributed

12. What is the order in which cuts were made to construct the mosaic plot below?



a) `Temperature, M_User, Preference, Water_softness`
b) `M_User, Temperature, Water_softness, Preference`
c) `M_User, Water_softness, Temperature, Preference`
d) `Temperature, Preference, Water_softness, M_User`

13. The Shapiro-Wilk normality test is performed on a set of data. The p-value is .65. At a significance level of .05, what can we conclude?

a) We cannot reject the assumption that the data is normally distributed.
b) There is a 65% chance that the data is normally distributed.
c) The data is not normally distributed.
d) There is a 65% chance that the data is not normally distributed.

14. What information do variable width boxplots provide that equal width boxplots do not?

a) group sizes
b) group means
c) group ranges
d) group medians

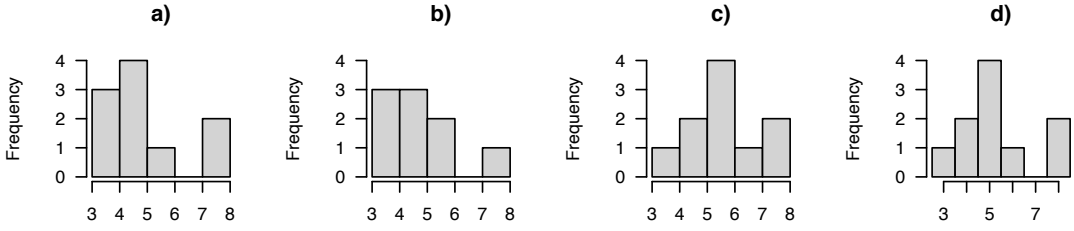
15. If the data in the table below were turned into a tidy data frame, how many rows and columns would it have?

	satisfaction			
income	Very Dissatisfied	Dissatisfied	Satisfied	Very Satisfied
< 15k	1	3	10	6
15-25k	2	3	10	7
25-40k	1	6	14	12
> 40k	0	1	9	11

a) 4 rows, 3 columns
b) 4 rows, 2 columns
c) 16 rows, 2 columns
d) 16 rows, 3 columns

16. Which of the following could NOT be a histogram of the given data?

```
x = c(3, 4, 4, 5, 5, 5, 5, 6, 8, 8)
```



17. Based on the plot below, which of the following is true?

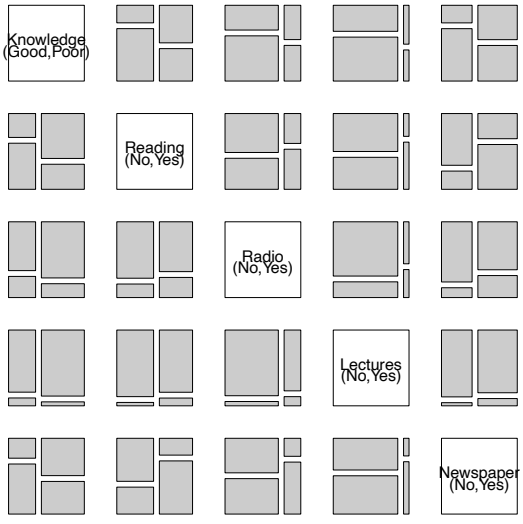
```
The decimal point is 2 digit(s) to the right of the |

2 | 5
3 | 55555
4 | 00000000000000555555555555
5 | 00000000000000000000000000000055555555555555555
6 | 0000000000000055
7 | 00
8 | 0
```

a) The data is approximately symmetric
b) The data shows rounding patterns
c) The data is unimodal
d) All of the above

18. A perceptually uniform colorspace is one in which

a) equal differences in data values are perceived as equal differences in color values.
b) only one color is employed, for example, a color scale that ranges from pale blue to deep blue.
c) only full opacity colors are employed (i.e., alpha = 1).
d) only one level of brightness is employed, for example, all colors are very light or very dark.

19. Based on the graph, which variables display the strongest association?



a) `Knowledge` and `Reading`
b) `Reading` and `Lectures`
c) `Knowledge` and `Radio`
d) `Radio` and `Lectures`

20. Boxplots are NOT a good choice for

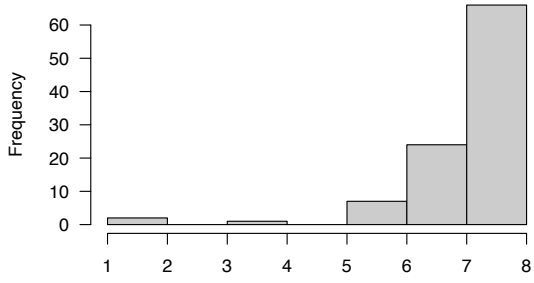
a) summarizing Likert data
b) comparing ranges
c) large datasets
d) viewing outliers

21. If a boxplot were drawn in R based on the data listed below, which of the values, if any, would be shown as outliers?

Values: `-11, -8, 3, 5, 6, 8, 9, 12, 13, 15, 28, 32, 35`

a) `-11, -8, 35`
b) `-11, 35`
c) `-11, -8, 28, 32, 35`
d) `-11, 32, 35`

22. Which of the following five number summaries could have been calculated with the same data as shown in the histogram below?



a) 1.2 1.4 2.6 4.3 8.0
b) 1.6 2.2 3.5 7.8 8.0
c) 1.6 6.3 7.0 8.0 9.1
d) 1.2 6.7 7.4 7.8 8.0

23. All of the following are techniques to reduce overplotting in a scatterplot EXCEPT

a) alpha blending
b) sampling
c) increasing point size
d) jittering points

24. Experiments in graphical perception by Cleveland and others demonstrate that we are most accurate judging:

a) area
b) position along a common scale
c) length
d) color