Exam 1 - Parts 1

Due Sep 26 at 11:59pm **Points** 20 **Questions** 20
Available Sep 26 at 12am - Sep 26 at 11:59pm 23 hours and 59 minutes
Time Limit 120 Minutes **Allowed Attempts** 3

Instructions

Class -

Welcome to Exam 1 - Part 1

Part 1 is a set of 20 questions and it will be auto-graded.

Good luck.

AI

Take the Quiz Again

Attempt History

	Attempt	Time	Score
KEPT	Attempt 2	17 minutes	16.88 out of 20
LATEST	Attempt 2	17 minutes	16.88 out of 20
	Attempt 1	32 minutes	12.96 out of 20

ⓘ Correct answers will be available on Oct 3 at 12am.

Score for this attempt: **16.88** out of 20
Submitted Sep 26 at 12:18pm
This attempt took 17 minutes.

Question 1 1 / 1 pts

The most common approach to visualizing amounts (numerical values of categories) is using vertical or horizontal bars. TRUE/FALSE?

- (●) True
- ( ) False

Question 2

1 / 1 pts

Select all that apply when it comes to visualizing amounts

- [ ] Violin plots cannot be used to represent all the information
- [x] Box plots can give a summary of the data but they do not show modality
- [x] The most common approach is to use horizontal or vertical bars
- [ ] Heatmaps must not be used
- [x] Dots can also be placed at the location where corresponding bars would end
- [x] If there are two or more sets of categories for which we want to show amounts, we can group or stack the bars

Partial

Question 3

0.67 / 1 pts

Which of these are true? Select all that apply.

- [x] Cumulative density plots can represent the data but may be difficult to interpret
- [ ] Q-Q plots are always misleading
- [ ] Histograms and density plots both require parameter choices
- [x] Histograms and density plots provide the most intuitive visualizations of a variable distribution
- [ ] If we require overall shifts among variable distributions, we can always use pie charts

Question 4

1 / 1 pts

The violin plot and box plot, when conditioned on a categorical variable, do not allow us to visualize subsets of a variable at the same time. TRUE/FALSE

- ( ) True
- (•) False

Question 5

1 / 1 pts

We can use the kernel density estimate (kde) plot to refine insights gained from the histogram of the same variable. TRUE/FALSE?

- (●) True
- ( ) False

Question 6

1 / 1 pts

Which of these is true?

- ( ) Violin plots combine the histogram and the boxplot
- (●) Stacked histograms and overlapping densities allow a more in-depth comparison of a smaller number of variable distributions
- ( ) Boxplots are not useful when we want to visualize many distributions at once
- ( ) For violin plots, the hue argument in Seaborn is used to condition the plot on the first categorical variable
- ( ) The pairplot always gives perfect information about the relationships between two variables

Partial

Question 7

0.5 / 1 pts

Why should we employ data visualization to communicate with stakeholders? Select all that apply.

- [ ] Human brains always suppress numerical data and look for visuals
- [x] Visualizations always employ complex tools to communicate findings
- [ ] It is always best to restrict messages to numbers
- [x] Data visualization can be compact, accessible to stakeholders
- [x] The human brain can spot tiny clues about speed and distance, and can distill that information from what the eyes see almost instantaneously
- [x] To be taken seriously by stakeholders, in addition to presenting the data or insights to support communication, we need to show them the data to support our message
- [ ] Human brains are wired for seeing patterns and differences and also for understanding spatial relationships from patterns and differences

Question 8

1 / 1 pts

Visualizations are useful in the early stages of statistical analysis. It is good to generate many charts for the benefit of your analysis and stakeholders. TRUE/FALSE?

- (●) True
- ( ) False

Question 9

1 / 1 pts

Anscombe's quartet is an example of plots which show a dataset with exactly the same statistics but which mask great differences. In a pairplot, we may see a different picture from calculating the correlation between two variables. TRUE/FALSE?

- (●) True
- ( ) False

Question 10

1 / 1 pts

Gelman and Unwin (2012) note the following as the core motivations for data visualization. Select all that apply.

- [x] To attract attention and stimulate interest
- [x] To communicate findings
- [x] To show the scale and complexity of the data
- [x] To allow exploration of the data
- [ ] To demonstrate expertise of the presenter

- [ ] To disprove the simplicity of numbers
- [x] To give an overview
- [x] To tell a story

Question 11

1 / 1 pts

The default kernel for Seaborn kde plot is Gaussian. TRUE/FALSE?

- (•) True
- ( ) False

Partial

Question 12

0.71 / 1 pts

Select the true statements from the list below

- [x] The kernel for a kde plot is a function that helps estimate a known probability density function
- [x] The Shade=True option in Seaborn provides a way to shade the area underneath the kernel density estimate plot

- [x] Bivariate kernel density estimate plots reveal the joint probability distribution for two random variables

- [x] Kernel density estimate plots are used to show what the underlying distribution of numerical variables are

- [ ] Visualizing statistical relationships use only one common approach: scatterplots

- [x] Data visualization involves visualizing distributions of variables and possible statistical relationships between variables

- [x] The kernel for a Kde plot helps estimate an unknown probability density function

- [ ] Kde kernels cannot be normal

- [ ] Kernel density estimate (kde plot) kernel options include Gaussian, triangular, cosine, exponential

- [x] Statistical analysis is a process of understanding how variables in a dataset relate to each other and how those relationships depend on other variables

Question 13

1 / 1 pts

Infographics are NOT data visualizations because they do not represent numbers in any way except writing them out in text. TRUE/FALSE?

- (●) True
- ( ) False

Question 14

1 / 1 pts

Why is it important to sometimes layout information from simple to complex when presenting information?

- ( ) It prevents you from having to tell the story yourself
- ( ) It provides just enough imagery to support messages and make data comparisons
- (●) It allows stakeholders to drill down as far as they want
- ( ) It allows you show as many variables as in many charts as possible
- ( ) It provides the best way to avoid criticism of your work

Question 15

1 / 1 pts

Which of the following are good questions to ask yourself as you begin work on a visualization? Select all that apply.

- [x] What stakeholder questions am I trying to answer?
- [ ] What stakeholder representatives need to be in the room?
- [x] What is the message is the chart providing?
- [ ] Is this chart complex enough?
- [ ] How many stakeholders do I have?
- [x] Do I/we know how to do this, or can we learn from it/adapt someone else's work or do we need to hire in?
- [x] What other parts need to be shown for contrast/context?
- [x] What is the context or overall environment for the visualization?
- [x] What parts of the data are evidence for the message?

Incorrect

Question 16

0 / 1 pts

Which of these show us an example of why data visualizations are important?

- ( ) Frank Anscombe's artificial datasets, which all have the same means, standard deviations, and correlations would have been misleading without the production of charts for each of the datasets
- ( ) Cumulative density plots are always easy to interpret relative to histograms and density plots

- [x] Statisticians have no way to communicate their findings without charts
- [ ] Gelman and Unwin (2012) did not give sufficient reasons why statisticians would want to create visualizations of data
- [ ] Human brains are not very good at spotting anomalies

Question 17

1 / 1 pts

For two or more sets of categories for which we want to show amounts, we can map the categories on the horizontal and vertical axes and show amounts by color. TRUE/FALSE?

- [x] True
- [ ] False

Incorrect

Question 18

0 / 1 pts

Why can histograms and density plots be misleading? Select all that apply.

- [x] They can require arbitrary parameters
- [x] They can be unfriendly to the human eye
- [x] They can be hard to present

- [ ] They are not easy to interpret without color
- [ ] They do not require arbitrary parameters
- [ ] They do not allow an in-depth comparison of distributions across categories
- [ ] They are not friendly to the human eye

Question 19

1 / 1 pts

The rule that most normally distributed variables follow is called? Select all that apply.

- [ ] The central dispersion rule
- [ ] The box and whisker rule
- [ ] The Gaussian estimate
- [ ] The proportion rule
- [x] The 68-95-99.7 percent rule
- [x] The empirical rule
- [ ] The law of small numbers
- [ ] The kde plot estimate

Question 20

1 / 1 pts

What effective options can a statistician use to visualize amounts? Select all that apply.

- [x] Side-by-side bar charts
- [ ] Stacked correlogram chart
- [x] Stacked density charts
- [x] Stacked bar charts
- [x] Pie charts
- [ ] None of these answers

Quiz Score: **16.88** out of 20