**Quiz 9, Version 1**

Q1. In the figure below, the box-and-whisker plot and the violin plot show the same data. Which of these statements is **false**?

A close up of a map

Description automatically generated

1. The interquartile range for Group B is 11.
2. The boxplot is misleading because it seems like most of the values in Group B should be centered around the median (around 15), but the values in Group B actually have a bimodal distribution with values distributed in 2 groups around y=18 and y=13.
3. The boxplot is misleading because it doesn’t show all of the data points, and thus we can’t see important information, like that Group C has a much smaller sample size than the other groups.
4. Group A has five potential outliers.

Q2. In the figure below, three different types of error bars are plotted on the same data (sepal lengths on flowers of three different iris species). Which of these statements is **false**?

A screenshot of a cell phone

Description automatically generated

1. We can be 95% confident that the true mean sepal length for *versicolor* irises is 6.
2. One should always report what type of error bars are being used in a figure in the figure caption.
3. A violin plot or a box-and-whisker plot with points overlain would show the natural variability in these data better than the “dynamite plots” that show the means and error bars.
4. If we wanted to compare the mean sepal lengths amongst the three iris species shown in this plot, we would use an ANOVA analysis.

Q3. Which of these is **not** good advice for choosing colors for data visualization?

1. Grayscale is always the best color option.
2. When using light-to-dark color scales, low values should be given lighter colors and high values should be given darker colors.
3. If you use a color scale to encode a variable, your figure caption or a figure legend should explain your color scale.
4. You should use intuitive colors (e.g., red=Republican, blue=Democrat).

Q4. Graphs with two different Y axes can be misleading and hard to read, and thus they should be avoided.

1. TRUE
2. FALSE

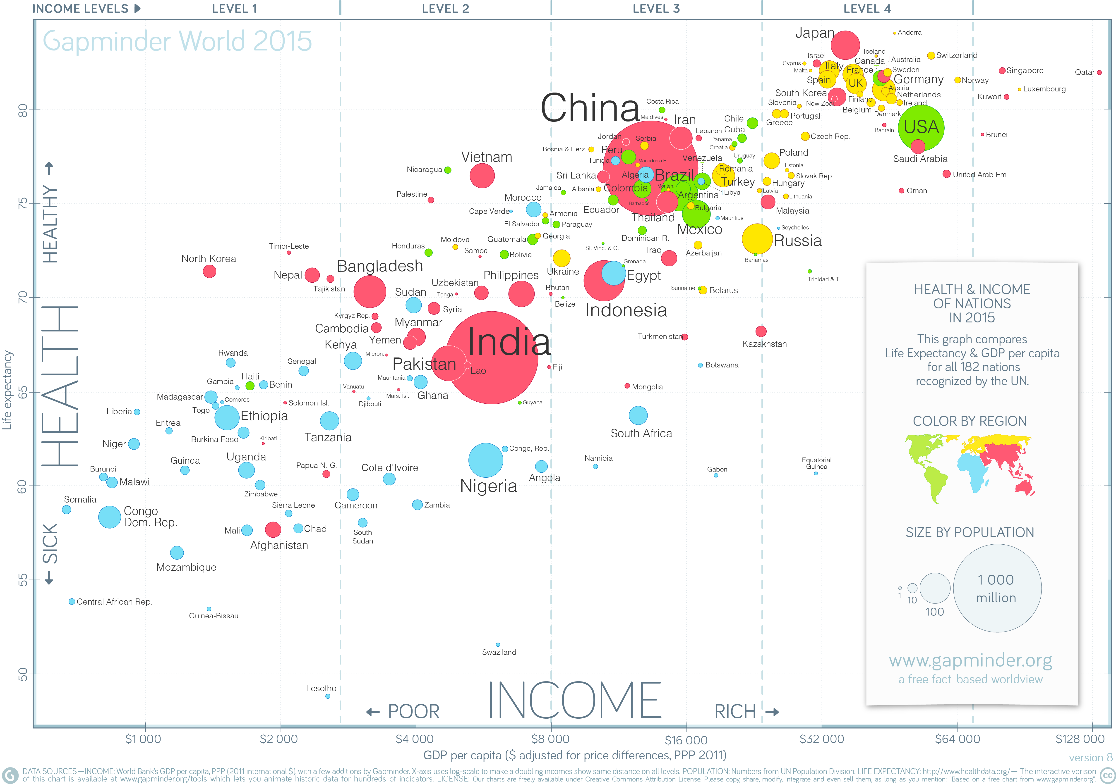
Q5. Which of these is **not** a problem in the following graph?

A screenshot of a cell phone

Description automatically generated

1. The legend is unnecessary.
2. The graph uses unnecessary 3D.
3. The labels on the graph are hard to read.
4. The grid lines create unnecessary clutter.

Q6. What type of graph is this?



1. Bubble plot
2. Box-and-whisker plot
3. Density plot
4. Violin plot

**Quiz 9, Version V2**

Q1. In the figure below, the box-and-whisker plot and the violin plot show the same data. Which of these statements is **false**?

A close up of a map

Description automatically generated

1. All groups (Groups A-D) have normal distributions.
2. Boxplots do not show individual data points and thus they hide the sample size of each group.
3. Boxplots do not show individual data points and thus they hide the data distributions in each group.
4. The median for Group A is 10.

Q2. In the figure below, three different types of error bars are plotted on the same data (sepal lengths on flowers of three different iris species). Which of these statements is **false**?

A screenshot of a cell phone

Description automatically generated

1. If we wanted to compare the mean sepal lengths amongst the three iris species shown in this plot, we would use a classical linear regression analysis.
2. The error bars showing the 95% confidence intervals are roughly twice as wide as the error bars showing +/- one SE.
3. We cannot see the distributions of sepal lengths measured for individual flowers by looking at this graph.
4. There was less variability in sepal length in the sampled *setosa* irises than in the sampled *virginica* irises.

Q3. Which of these is **not** good advice for choosing colors for data visualization?

1. Red and green are great colors for data visualization because red means “bad” and green means “good”.
2. Dark text should go on light backgrounds and light text should go on dark backgrounds.
3. If you have multiple graphs, your color scheme should be consistent for all graphs.
4. If you use a color scale to encode a variable, your figure caption or a figure legend should explain your color scale.

Q4. The creators of ggplot made it very difficult to make graphs with two different Y axes because they don’t want users to make graphs with two Y axes. Which of these statements is **false**?

1. Using two Y axes is OK if both axes start at zero.
2. People usually use graphs with two different Y axes to compare two trends with each other, and other types of graphs could accomplish this same goal.
3. On graphs with two different Y axes, the two Y axes could be plotted using any arbitrary scale, and this could accidentally or purposefully mislead readers about the relationships between the trends being plotted.
4. Instead of using graphs with two Y axes, the same trends could be plotted in two different side-by-side plots.

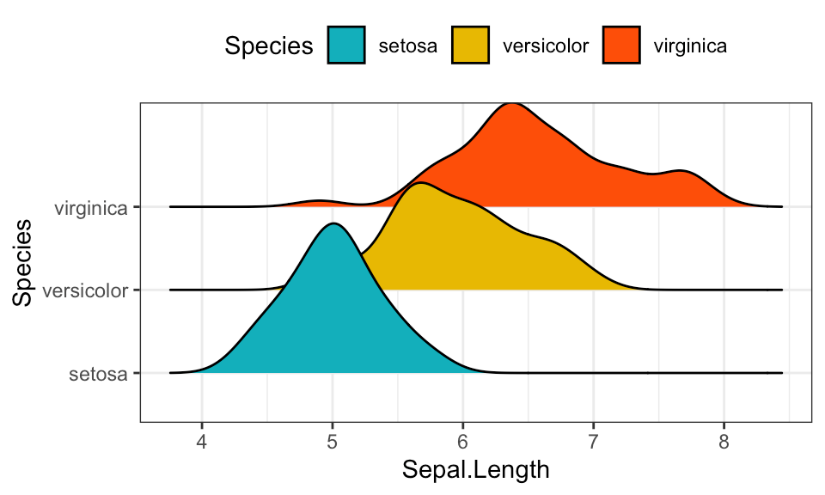
Q5. Which of these would **not** improve the following graph?

A screenshot of a cell phone

Description automatically generated

1. Reduce the size of all fonts.
2. Shorten the X labels and make them horizontal.
3. Use bigger intervals for the Y axis so that fewer labels are necessary.
4. Remove unnecessary 3D and grid lines.

Q6. What type of graph is this?



1. Density plot
2. Bubble plot
3. Box-and-whisker plot
4. Violin plot