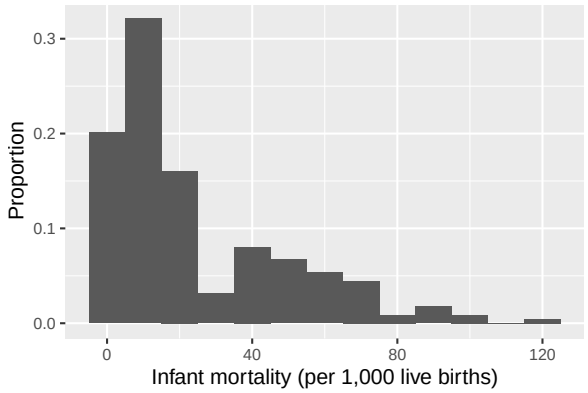
PS 3

1. **Measures of center and spread.** The infant mortality rate is defined as the number of infant deaths per 1,000 live births. This rate is often used as an indicator of the level of health in a country. The relative frequency histogram below (which displays proportions instead of counts) shows the distribution of estimated infant death rates for 224 countries for which such data were available in 2014.

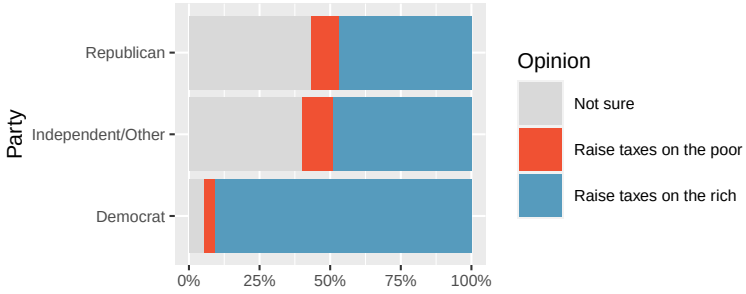


   a. Estimate Q1, the median, and Q3 from the histogram.

   b. Would you expect the mean of this dataset to be smaller or larger than the median? Explain your reasoning.

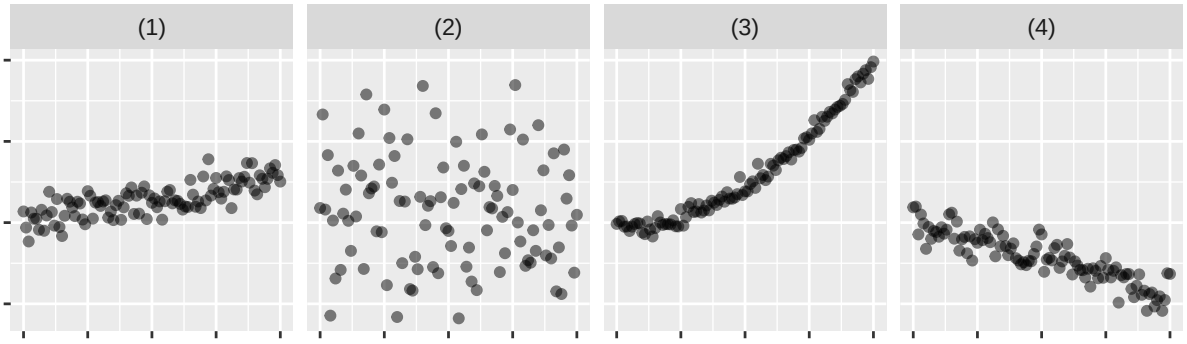
   c. Would the sample standard deviation or the IQR do a better job of measuring the spread of this distribution? Why?

2. **Associations in bar charts.** A random sample of registered voters nationally were asked whether they think it's better to raise taxes on the rich or raise taxes on the poor. The survey also collected information on the political party affiliation of the respondents.

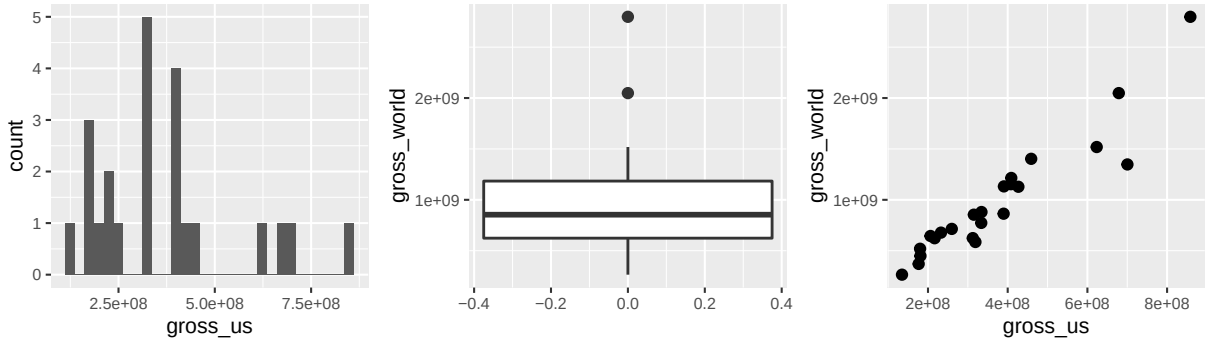


   a. Based on the stacked bar plot shown above, do views on raising taxes and political affiliation appear to be associated? Explain your reasoning.

3. **Associations in scatterplots.** Indicate which of the plots show (a) a positive association, (b) a negative association, or (c) no association. Also determine if the positive and negative associations are linear or nonlinear. Each part may refer to more than one plot.



4. **Practice with ggplot I**. The following three plots come from a data set called `mcu_films` inside the `openintro` package. Please write out the `ggplot2` code that will produce each one.



5. **Practice with ggplot II**.The following three plots come from a data set called `penguins` inside the `palmerpenguins` package. Please write out the `ggplot2` code that will produce each one.