

Keegan Moynahan

### Week 3 Questions

- 1) The types of plots that show every data point are scatterplots, Cleveland dotplot, Coplot, and QQ plots
- 2) The types of plots that show aggregated or summarized data are histograms, boxplots
- 3) In terms of the definition of a conditional variable in context of graphical data exploration, I'd describe a conditional variable as a dependent variable, that produces different results due to the variable changing. In the reading by Zuur, Zuur kept using conditional as a standalone word (no variable). Zuur used this word to describe different aspects of plots, "Panels B and C show the conditional histograms for the GSI index conditional on sex" (Zuur, Ch 4 page 28). This shows that the word conditional is being used to describe the different histogram made from the variable sex.
- 4) Three common measures of spread are variance, standard deviation, and range
- 5) Range is used to measure the range of variables in your data. You would take the min and max and find the difference between the two. This type of spread shows the absolute range of values. Variance is the squared deviation from the mean or expected value. This spread is used to show the deviation from the mean.
- 6) I haven't worked with many datasets where I needed to plot and create statistics, so I'm going to use a potential dataset I could use. I could get data from recent wildfires to study elevation trends that could help identify potential areas that have a high risk of wildfires. The way I would explore this data would be done through graphical data exploration. The first reason is, I'd want to find where the data is center, how is the data spread, and to find out if the data is symmetric, skewed, or bimodal. I'd answer these questions by creating a histogram, for elevation I could have ranges of elevation in the histogram that group the data together. This would show me ranges in which the elevation of most wildfires happens at, it would also tell me how accurate my data is. The second reason would be to find outliers, I could also use a histogram for this, but I think a box plot would work better. This would show me the range of the elevations and any outliers out of the range would stick out. This would help provide an index of elevation risks.